Please cite this article in press as: Jermusyk 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), https://doi.org/10.1016/j.ajhg.2021.09.002

ARTICLE

A 584 bp deletion in *CTRB2* inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer

Ashley Jermusyk,[1,13] Jun Zhong,[1,13] Katelyn E. Connelly,[1,13] Naomi Gordon,[1,13] Sumeth Perera,[2] Ehssan Abdolalizadeh,[1] Tongwu Zhang,[3] Aidan O'Brien,[1] Jason W. Hoskins,[1] Irene Collins,[1] Daina Eiser,[1] Chen Yuan,[4] PanScan Consortium, PanC4 Consortium, Harvey A. Risch,[5] Eric J. Jacobs,[6] Donghui Li,[7] Mengmeng Du,[8] Rachael Z. Stolzenberg-Solomon,[3] Alison P. Klein,[9,10] Jill P. Smith,[11] Brian M. Wolpin,[4] Stephen J. Chanock,[3] Jianxin Shi,[3] Gloria M. Petersen,[12] Christopher J. Westlake,[2] and Laufey T. Amundadottir[1,*]

Summary

Genome-wide association studies (GWASs) have discovered 20 risk loci in the human genome where germline variants associate with risk of pancreatic ductal adenocarcinoma (PDAC) in populations of European ancestry. Here, we fine-mapped one such locus on chr16q23.1 (rs72802365, $p = 2.51 \times 10^{-17}$, OR = 1.36, 95% CI = 1.31–1.40) and identified colocalization (PP = 0.87) with aberrant exon 5–7 *CTRB2* splicing in pancreatic tissues ($p_{GTEx} = 1.40 \times 10^{-69}$, $\beta_{GTEx} = 1.99$; $p_{LTG} = 1.02 \times 10^{-30}$, $\beta_{LTG} = 1.99$). Imputation of a 584 bp structural variant overlapping exon 6 of *CTRB2* into the GWAS datasets resulted in a highly significant association with pancreatic cancer risk ($p = 2.83 \times 10^{-16}$, OR = 1.36, 95% CI = 1.31–1.42), indicating that it may underlie this signal. Exon skipping attributable to the deletion (risk) allele introduces a premature stop codon in exon 7 of *CTRB2*, yielding a truncated chymotrypsinogen B2 protein that lacks chymotrypsin activity, is poorly secreted, and accumulates intracellularly in the endoplasmic reticulum (ER). We propose that intracellular accumulation of a nonfunctional chymotrypsinogen B2 protein leads to ER stress and pancreatic inflammation, which may explain the increased pancreatic cancer risk in carriers of *CTRB2* exon 6 deletion alleles.

Introduction

Pancreatic cancer is currently the third leading cause of cancer-related deaths in the United States and fourth in Europe.[1,2] Unlike many other cancers, it is increasing in incidence and mortality and is predicted to become the second leading cause of cancer-related deaths in the US by 2030.[3] Both genetic and environmental factors contribute to pancreatic cancer risk.[4] Genetic risk factors include rare variants with moderate to large effects often seen as part of familial cancer syndromes and common variants with smaller magnitudes of association identified through genome-wide association studies (GWASs).[5] To date, 20 pancreatic-cancer-risk loci, containing 23 independent signals, have been identified through GWASs performed in individuals of European descent.[6–11] Most of these loci are believed to influence disease risk through noncoding variants that lie within enhancers or promoters and affect gene expression in *cis* or *trans*.[12]

One of the 20 currently known pancreatic-cancer-risk loci from GWASs is located at chr16q23.1, near genes that encode chymotrypsinogen precursors (*CTRB1* and *CTRB2*) and a signal transduction scaffold protein (*BCAR1*).[6,7] We used fine-mapping, splicing quantitative trait locus (sQTL) analysis, and functional approaches to explore the molecular changes that might underlie this risk signal.

Material and methods

Fine-mapping

Pancreatic cancer case (9,013) and control (12,452) subjects were drawn from four GWASs conducted in individuals of European ancestry by the Pancreatic Cancer Cohort Consortium (PanScan I–III)[7–9] and the Pancreatic Cancer Case Control Consortium (PanC4).[11] The PanC4 GWAS genotype dataset was obtained from the Database of Genotype and Phenotype (dbGaP) via controlled access (dbGaP: phs000648.v1.p1). Genotype quality

[1]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA; [2]Laboratory of Cell and Developmental Signaling, Center for Cancer Research, National Cancer Institute, NIH, Frederick, MD 21702, USA; [3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA; [4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA; [5]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06520, USA; [6]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA 30303, USA; [7]Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [8]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10017, USA; [9]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD 21231, USA; [10]Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA; [11]Department of Medicine, Georgetown University, Washington, DC 20057, USA; [12]Department of Quantitative Health Sciences, College of Medicine, Mayo Clinic, Rochester, MN 55905, USA
[13]These authors contributed equally
*Correspondence: amundadottirl@nih.gov
https://doi.org/10.1016/j.ajhg.2021.09.002.

control, imputation (with the 1000G Phase 3 EUR dataset as a reference panel and IMPUTE2[13]), and association analysis were performed assuming an additive genetic model adjusting for age, sex, geographic region, and eigenvectors via SNPTEST[14] as previously described.[6,7] We also applied a recessive model for the 15 potentially functional variants and the *CTRB2* deletion variant by using SNPTEST.[14] Meta-analyses for PanScan and PanC4 were performed with METAL.[15]

All studies obtained informed consent from participants and institutional review board (IRB) approvals including IRB certifications permitting data sharing in accordance with the NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWASs). The PanScan study was also approved by the NCI Special Studies Institutional Review Board.

We used two approaches to determine the most likely set of potential functional variants at this risk locus. We first calculated the likelihood ratio of all GWAS-significant variants relative to the most significant SNP by dividing the maximized likelihood for the tested SNP to the maximized likelihood for rs72802365[16] and used a threshold of likelihood ratio (LR) $\geq$ 1:1,000 to select SNPs (n = 15) to take forward for functional analysis. We also used the SuSiE (sum of single effects) method that uses an iterative Bayesian stepwise selection (IBSS) procedure to identify credible sets of variants within which there is a high probability of capturing at least one functional variant.[17] This yielded nine SNPs in one credible set at this locus (all of which had LR $\geq$ 1:1,000). To be comprehensive, we included all 15 SNPs in functional analysis.

Splicing quantitative trait loci (sQTLs) analysis

DNA isolation protocols and sample quality-control metrics for the Genotype-Tissue Expression (GTEx, version 7) project have been described elsewhere.[18] Genotypes derived from whole-genome sequencing for the 635 individuals included in GTEx were obtained via controlled access from dbGaP (dbGaP: phs000424.v7.p2) and the junction read counts from RNA sequencing (RNA-seq) were downloaded from the GTEx portal. Samples with <80% European ancestry were excluded on the basis of analysis using the genotyping library and utilities (GLU) *struct.admix* module,[7] resulting in a total of 174 individuals with both genetic data and gene expression data for pancreatic tissue samples available for analyses. DNA isolation protocols, sample quality-control metrics, imputation of genotype, and RNA-seq for the NCI Laboratory of Translational Genomics (LTG) pancreatic tissue dataset (n = 95 samples with European ancestry) have been previously described.[19] Five samples with low expression of *CTRB2* (<0.1 transcripts per million [TPM]) were excluded for this analysis. Genotype and RNA-seq data for the NCI LTG expression quantitative trait locus (eQTL) dataset are available from dbGAP (dbGaP: phs001776.v1.p1). Variable splicing events from RNA-seq data were identified by Leafcutter via a Dirichlet-multinomial generalized linear model,[20] and sQTL mapping was performed by FastQTL with linear regression.[21] Covariates used in the analysis of GTEx data were downloaded from the GTEx portal and included three PCs for genotype data, 30 probabilistic estimation of expression residuals (PEER) factors for RNA-seq data, sex, and platform (HiSeq2000 versus HiSeq2500). Covariates used for the LTG dataset included the top five PCs for genotype data, 15 PEER factors for RNA-seq data,[22] as well as sex. Visualization of sQTLs was conducted by IGV.[23]

eQTL analysis

RNA samples were isolated from 95 histologically normal fresh frozen tissue samples from healthy (not with a diagnosis of cancer) donors or from normal tissue adjacent to pancreatic tumors as previously described.[19] RNA-seq was performed with Illumina HiSeq2000, and DNA genotyping was performed with Illumina OmniExpress or Omni1M arrays.[19] Read alignment, quantification, and eQTL analysis was performed as previously described.[19]

GTEx eQTLs were assessed via the "test your own eQTL calculator" in the GTEx v.7 browser.[18] GTEx eQTL datasets were downloaded from the GTEx portal for further analysis.

Imputing the CTRB2 deletion and CTRB1/CTRB2 inversion variants

We purchased DNA samples from the 1000 Genomes (1000G) European populations (TSI, FIN, GBR, and IBS populations and HapMap CEU, n = 438) from the Coriell Institute for Medical Research (Coriell) and used them to genotype the *CTRB2* exon 6 deletion and *CTRB1/CTRB2* inversion variants (Tables S1 and S2). We used two sets of primers (Table S3) to genotype the deletion (as described by Rosendahl et al.[24]) and designed two sets of primers to genotype the inversion variant (Table S3). The minor allele frequencies of the deletion were very similar to those of the most significant GWAS SNP and ranged from approximately 6% in the TSI samples to ~13% in the IBS samples (Table S2). The deletion and inversion genotypes were then included with the 1000G genotypes for chr16q23.1 (1000G, Phase3, version 80) and used as a reference panel for imputation of PanScan I–III and PanC4 GWAS data across this locus (Hg19: chr16: 72,738,615–77,739,199 bp). Genotype and sample quality thresholds were applied for minor allele frequency (MAF > 0.01) and sample completion (>90%). Post-imputation variants with imputation quality score (R^2) < 0.5 or duplicated variants were removed by PLINK. Association models were as described above for fine-mapping with SNPTEST.[14]

We validated the imputation accuracy for the *CTRB2* deletion variant (73%–97% accuracy) and *CTRB1/CTRB2* inversion variant (88%–100% accuracy) genotypes by PCR genotyping 96 DNA samples from PanScan (Table S4). We also directly genotyped DNA samples from 493 PDAC-affected individuals and 502 control individuals from PanScan for the *CTRB2* exon 6 deletion and rs72802365 to assess imputation accuracy (95% and 98%, respectively). Performing the association analysis in this sample set yielded very similar results as the imputation ($p_{Genotyped} = 3.93 \times 10^{-3}$ versus $p_{Imputed} = 3.89 \times 10^{-3}$ for the deletion; $p_{Genotyped} = 1.17 \times 10^{-3}$ versus $p_{Imputed} = 1.29 \times 10^{-3}$ for rs72802365). MAFs were also similar between the 995 imputed and directly genotyped samples ($MAF_{Genotyped} = 9.05\%$ versus $MAF_{Imputed} = 8.29\%$ for the deletion; $MAF_{Genotyped} = 8.80\%$ versus $MAF_{Imputed} = 9.20\%$ for rs72802365). Finally, when directly genotyped samples were included in the full analysis of all GWAS phases, the p values were nearly the same as before ($p = 2.83 \times 10^{-16}$ before versus 5.41×10^{-16} after correcting for genotyped samples for the deletion; $p = 2.51 \times 10^{-17}$ before versus 2.11×10^{-17} after correcting for genotyped samples for the deletion for rs72802365). We therefore conclude that the imputation for the *CTRB2* deletion variant and rs72802365 is highly accurate.

We performed Sanger sequencing on DNA samples from (1) a 1000G EUR individual homozygous for the *CTRB2* exon 6 deletion, in order to assess the boundaries of the *CTRB2* exon 6 deletion (Figure S6A), and (2) eight subjects from the LTG eQTL dataset to show that *CTRB2* exon 6 is not present in samples where

splicing between exons 5–6 and 6–7 is not seen and only exon 5–7 splicing is seen (Figure S6B). We amplified the *CTRB2* exon 5–7 region from genomic DNA from these eight human subjects via primers previously published in Rosendahl et al. (Table S3) and described above to detect the *CTRB2* exon 6 deletion. PCR products were run on a 2% agarose gel. Bands were excised and DNA was isolated with a QIAGEN DNA extraction kit. Purified DNA was sequenced with Big Dye Terminator (Thermo Fisher Scientific) on a 3500×l genomic analyzer (Thermo Fisher Scientific). Sequences were aligned to the reference *CTRB2* genomic sequence.

We formally calculated the likelihood for the *CTRB2* exon 6 deletion variant and the most significant SNP from the fine-mapping, rs72802365, as functional variants underlying the risk signal at 16q23.1. Consider two genetic variants: a deletion and a SNP (rs72802365) with linkage disequilibrium (LD) estimated to be $r = 0.95$ in our data on the basis of Pearson correlation of imputed genotype dosages from the PanScan I–III and PanC4 GWAS data. The deletion is imputed with imputation quality $R_1^2 = 0.85$; the SNP is imputed with imputation quality $R_2^2 = 0.99$. The score statistics for the two variants are $z = (8.10, 8.49)$, which corresponds to the two p values 5.41×10^{-16} and 2.11×10^{-17}. Consider model 1 that the deletion is causal. Let ξ denote the noncentrality parameter (NCP) of the score statistic for the deletion if deletion is genotyped without error. Then, the NCP for the imputed deletion is $R_1\xi$ and the NCP for the imputed SNP is $R_2 r\xi$. The likelihood of obtaining $z = (8.10, 8.49)$ is given by

$$L_1(z;\xi) = MVN\left(z; \begin{pmatrix} R_1\xi \\ R_2 r\xi \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

which leads to an estimate $\widehat{\xi} = 8.93$. Here, MVN stands for multivariable normal distribution.

Consider model 2 that the SNP is causal. Let η denote the NCP of the score statistic for the SNP if it is genotyped without error. Then, the NCP for the imputed deletion is $R_1 r\eta$ and the NCP for the imputed SNP is $R_2\eta$. The likelihood of obtaining $z = (8.10, 8.49)$ is given by

$$L_2(z;\eta) = MVN\left(z; \begin{pmatrix} R_1 r\eta \\ R_2\eta \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

which leads to an estimate $\widehat{\eta} = 8.10$. The likelihood ratio of the two models is calculated as

$$\frac{L_1\left(z;\widehat{\xi}\right)}{L_2(z;\widehat{\eta})} = 5.75.$$

Thus, we conclude that the deletion variant is more likely to be causal as compared to rs72802365.

Colocalization

Colocalization analysis for the GWAS and eQTL or sQTL signals was performed with the coloc package (coloc.abf) in the R Project for statistical computing.[25] Posterior probabilities (PPs) for shared (PP4) or different (PP3) functional variants underlying the GWAS and eQTL or GWAS and sQTL signals are presented.

Expression and pathway enrichment analyses

Differential expression analysis was conducted with RNA-seq data from normal human pancreatic tissue samples generated by the GTEx project (n = 174) via the EdgeR package in R.[26] The genotype for the *CTRB2* exon 6 deletion variant was determined via exon 6–7, exon 5–6, and exon 5–7 junction read counts. We used only samples that were homozygous noncarriers for the *CTRB1/CTRB2* exon 1 inversion allele (i.e., homozygous for the ancestral allele) for the analysis (inferred by selecting only samples with CC at rs1808427, a proxy SNP for the inversion based on PCR genotyping 438 1000G EUR samples as described above, $r^2 = 0.59$) to minimize any possible confounding effect that the inversion allele (the derived allele and the human reference genome, Hg19, allele) may have on gene expression (as the deletion only occurs on the ancestral non-inverted allele). Only genes with reads in at least 80% of the samples were used. We then scaled counts for sequencing depth and RNA composition to yield normalized counts (CPM). These pseudo-counts were analyzed for differential expression with respect to number of copies of the deletion variant as a continuous variable (0, 1, 2 copies) via the glmQLFit (quasi-likelihood negative binomial generalized linear model) method in EdgeR.[27] Differential expression ($\log_2(\beta)$ and p values—where β is per copy of the deletion) was then calculated with an exact statistic.

Pathway analysis was performed through the use of both ingenuity pathway analysis (IPA, QIAGEN[28]) and gene set enrichment analysis (GSEA) software and Molecular Signature Database (MSigDB).[29] Genes differentially expressed at p value < 0.05 (n = 1,401) were analyzed by IPA on the basis of p value (those with positive and negative fold change were analyzed separately). All genes were analyzed via GSEAPreranked with a ranked list based on signed p value of differential expression for enrichment in gene sets: Biocarta, Reactome, KEGG, biological process (BP, from Gene Ontology), and molecular function (MF, from Gene Ontology).

Cell lines

Human pancreatic cancer cell lines MIA PaCa-2 and PANC-1 and human embryonic kidney cells 293T (HEK293T) were purchased from ATCC and maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) (PANC-1 and HEK293T) or 10% FBS and 2.5% horse serum (MIA PaCa-2). AR42J rat acinar cells were purchased from ATCC and maintained in F-12K medium supplemented with 20% FBS and antibiotics. The human pancreatic cell line HPDE-E6E7 (H6c7)[30] is a normal pancreatic-duct-derived immortalized cell line that was kindly provided by Dr. Ming Tsao at the Princess Margaret Cancer Center at the Ontario Cancer Institute, Toronto, Ontario, Canada. HPDE cells were maintained in Keratinocyte-SFM media with 5 ng/mL human recombinant epidermal growth factor (EGF) and 50 μg/mL bovine pituitary extract. All cell lines were tested routinely for mycoplasma and were always found to be negative. Cell lines were tested for authentication with a panel of short tandem repeats (STRs) via the Identifiler kit (Life Technologies) and compared with ATCC and DSMZ (German Collection of Microorganisms and Cell Cultures) STR profile datasets. All cell lines with profiles in the databases matched.

Chymotrypsin secretion

The full-length *CTRB2* open reading frame (ORF) with a Myc-FLAG (Myc-DDK) tag was purchased from Origene (cat. # RC209957). The *CTRB2* exon 6 deletion ORF was synthesized at ATUM.bio, and entry clones were generated via Reverse BP Clonase reactions into pDonr-255. The full-length and exon 6 deletion clones for *CTRB2* with a Myc-FLAG tag at the 3′ end were cloned into pDest-371 via the Gateway system followed by

sequence verification. The expression clones with C-terminal eGFP and SNAP tags were cloned into pDest-377. CTRB2 expression plasmids were transiently transfected into HEK293T cells via Lipofectamine 2000 (Thermo Fisher Scientific) at approximately 50% confluency. Cell lysates and conditioned media were harvested 48 h after transfection. Conditioned media was concentrated with a SpeedVac to approximately half of the original volume (3 h at room temperature). Cell lysates were harvested with a NP40 deoxycholate (DOC) lysis buffer (1 mL). Immunoprecipitation was carried out with a mouse anti-FLAG antibody (2 μg, Sigma Aldrich F1804) and Dynabeads Protein G (Thermo Fisher Scientific 10004D) on concentrated conditioned media and 10 μL or 100 μL of lysate. Immunoprecipitated samples from lysate and media were run on the same blot and stained with a mouse anti-FLAG antibody (Sigma Aldrich F1804) and goat anti-mouse IgG, Fcγ (heavy chain) fragment specific, horseradish peroxidase (HRP) conjugate (Jackson ImmunoResearch Laboratories 115-035-164). Blots were imaged via a ChemiDoc Touch Imaging System (Bio-Rad) and bands quantified with the ImageLab 5.2.1 software (Bio-Rad).

Untagged full-length and truncated *CTRB2* cDNA constructs were cloned, sequence verified, transfected, and harvested as described above for tagged constructs. Concentrated conditioned media and cell lysate of equal volumes were run on the same blot and stained with a mouse anti-CTRB1/2 antibody (Santa Cruz Biotechnology sc-393414) and rabbit anti-γ-tubulin (Sigma T5192) primary antibodies and goat anti-mouse IgG Fcγ specific and goat anti-rabbit (Abcam, ab205722) HRP conjugate antibodies, respectively. Blots were imaged and quantitated as described above.

Chymotrypsin activity assays

Transfection of *CTRB2* expression plasmids (FLAG-tagged and untagged) was performed as described above for IP-Western analysis of CTRB2 secretion. HEK293T and AR42J (un-transfected) cell lysates were harvested 48 h after transfection and analyzed with a fluorometric Chymotrypsin Assay Kit (Abcam, ab234051) according to the manufacturer's protocol. Briefly, cells were harvested in assay buffer and homogenized on ice with a 19 g needle (10×). Proteins in the total cell lysate were cleaved (activated) by a chymotrypsin activator (i.e., trypsin) and chymotrypsin activity assessed with a synthetic fluorogenic substrate measured by fluorescence (Ex/Em 380/460) every 30 s for 2.5 h in a white-bottom plate via a Tecan Infinite M200. We also ran control wells containing cell lysates without chymotrypsin activator (trypsin) to assess background non-specific activity. Samples were analyzed in triplicate (quadruplicate for untagged *CTRB2* transfections) with 20 μL lysate. Activity was quantified across three time points within linear regions of the fluorescence curve and activity determined with a coumarin standard curve. Lysates were also assayed for CTRB2 protein abundance as described above with a donkey anti-mouse IgG, HRP conjugate (Abcam, ab7061) secondary antibody or donkey anti-mouse IgG Fcγ specific HRP conjugate antibody. Chymotrypsin activity was normalized by protein content as measured by the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific 23227).

Assessment of endoplasmic reticulum stress

Transient transfection of *CTRB2* expression plasmids was performed as described above in HEK293T, MIA PaCa-2, and PANC-1 cells. We harvested RNA 48 h after transfection by using the QIAGEN RNeasy Plus Mini Kit with on-column DNase digestion via the RNase Free DNase Set (QIAGEN). cDNA was synthesized with SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies). Gene expression levels were quantified by quantitative real-time PCR with TaqMan assays for *GAPDH* (cat. # 4352934E), *DDIT3* (Hs00358796_g1), *HsPA5* (Hs00607129_gH), *CALR* (Hs00189032_m1), *EIF2AK3* (Hs00984003_m1), *CTRB2* (Hs 00157181_m1), and *PPIA* (Hs99999904_m1) from Life Technologies. Expression of spliced *XBP1* (forward primer, 5′-GCTGAGTCC GCAGCAGGT-3′; reverse primer, 5′-CCAACAGGATATCAGACTCT GAATCT-3′; probe, CAGAACATCTCCCCATGGA) and unspliced *XBP1* (forward primer, 5′-GCTGAGTCCGCAGCACTCA-3′; reverse primer, 5′-CCAACAGGATATCAGACTCTGAATCT-3′ probe, CAGA ACATCTCCCCATGGA) was assayed with custom Taqman assays as described with fluorescein (FAM) probes. Gene expression levels were normalized to *PPIA*, while expression of *PPIA* was normalized to *GAPDH* and fold change relative to full-length *CTRB2* was calculated via the delta-delta Ct (DDCt) method. Each experiment was performed in triplicate and repeated four times. Significance was assessed with a Student's two-tailed t-statistic.

Endoplasmic reticulum (ER) stress markers were assessed at the protein level by immunoblot analysis. PANC-1 and HEK293T cells were plated 24 h prior to transfection. Cells were transfected with equivalent amounts of *CTRB2* expression plasmids (untagged *CTRB2* constructs) with Lipofectamine 2000 (Thermo Fisher Scientific). After 72 h, transfected cells were harvested for cell lysate as described above. Fifty micrograms of lysate was run on two blots and probed for CTRB2 (Santa Cruz Biotechnology, sc-393414), CALR (Santa Cruz Biotechnology, sc-373863), PERK (Santa Cruz Biotechnology, sc-377400), BiP/Grp78 (Abcam, ab21685), XBP1 (Cell Signaling Technology, 40435S), and GAPDH (Abcam, ab125247). Donkey anti-mouse IgG, HRP conjugate (Abcam, ab7061) and donkey anti-rabbit IgG, HRP conjugate (Abcam, ab205722) secondary antibodies were used. Blots were quantitated via BioRad Image Lab software and normalized to GAPDH loading control prior to fold change calculations. Significance was assessed via a Student's two-tailed t-statistic.

Immunofluorescence and time-lapse microscopy

HPDE and PANC-1 cells were transfected (X-tremeGENE 9 from Roche) with expression plasmids containing either full-length *CTRB2* or exon-6-deleted *CTRB2* ORFs with C-terminal GFP tags. After 48 h, cells were fixed and processed for immunostaining and indirect immunofluorescence as previously described.[31] Briefly, cells were fixed in 4% paraformaldehyde in PBS (pH 7.4), washed, permeabilized in 0.1% Triton X-100, and blocked in 3% bovine serum albumin (BSA) with 0.1% Triton X-100 for 1 h at room temperature. Primary antibodies (rabbit anti-GRP78/BiP, Abcam ab21685, 5 μg/mL; rabbit anti-calnexin, Sigma #C4731, 1:200; mouse anti-GM130, BD Biosciences #610823, 1:1,000; and sheep anti-TGN46, AbD Serotec #AHP500, 1:200) were incubated in the blocking buffer overnight at 4°C, rinsed, and incubated with fluorophore conjugated secondary antibodies (donkey anti-rabbit [Alexa568 #A10042], donkey anti-mouse [Alexa647 #A31571], and donkey anti-sheep [Alexa568 #A21099]; all from Life Technologies) for 1 h at room temperature.

Imaging was performed with a Marianas (Intelligent Imaging Innovations) inverted microscope and a Coolsnap HQ2 camera (Photometrics) for epifluorescence imaging with a 63× 1.4NA objective or a sCMOS camera (Hamamatsu) for spinning disk confocal (SDC) imaging with a 63× 1.4NA oil objective. Nearest-neighbor deconvolution was applied with SlideBook software

Table 1. Association results for main risk variants at the pancreatic-cancer-risk locus on chr16q23.1

			Additive model		Recessive model		Association analysis conditioned on		
Germline variant	**Position (bp)**	**MAF in controls**[a]	**p value**[b]	**OR (95% CI)**[b]	**p value**[b]	**OR (95% CI)**[b]	**rs72802365 p value**	***CTRB2* deletion p value**	***CTRB1/CTRB2* inversion p value**
rs13337397	75,295,639	0.102	2.01×10^{-13}	1.28 (1.23–1.32)	1.18×10^{-3}	1.51 (1.33–1.72)	0.73	0.11	2.23×10^{-14}
rs72802365	75,246,035	0.082	2.51×10^{-17}	1.36 (1.31–1.40)	4.80×10^{-5}	1.86 (1.60–2.17)	1	0.007	1.89×10^{-18}
CTRB2 deletion	75,238,615[c]	0.073	2.83×10^{-16}	1.36 (1.31–1.42)	1.77×10^{-4}	1.91 (1.61–2.28)	0.12	1	3.14×10^{-17}
CTRB1/CTRB2 inversion	75,240,098[d]	0.180	0.26	1.03 (1.00–1.06)	0.19	1.11 (1.03–1.21)	0.01	0.17	1

Results from unconditional and conditional logistic regression analysis of GWAS genotypes generated in PanScan I, PanScan II, PanScan III, and PanC4 GWASs.
[a]*CTRB2* deletion variant start and stop position: chr16: 75,238,615–75,239,199 ± 5 bp (see Figure S6A).
[b]Unconditional p value and odds ratios (OR). Position of SNP or structural variant in NCBI genome build 37 (Hg37). OR, odds ratio (and 95% confidence interval [CI]).
[c]MAF for each variant in control samples for the combined GWAS datasets.
[d]*CTRB1/CTRB2* inversion variant start and stop position: chr16: 75,240,098–75,256,658 bp.

when indicated. To quantify the percentage of HPDE cells with elevated BiP antibody staining, we imaged ten random fields for CTRB2-full-length-GFP-expressing cells and truncated CTRB2 (exon 6 deleted)-GFP-expressing cells on a similar plane by SDC. Counts were taken for the number of cells showing BiP (Alexa568)-GFP colocalizations and of the elevated BiP staining intensities for each field via SlideBook. Percentages of cells with elevated BiP antibody staining were calculated against the total GFP cells in triplicate.

For ER trafficking studies, HPDE cells expressing GFP-tagged full-length CTRB2 or exon-6-deleted CTRB2 were incubated with ER-Tracker Red (BODIPY TR Glibenclamide, Thermo Fisher Scientific #E34250, working conc. 250 nM) for 15 min following manufacturer's instructions and washed five times in PBS, and live cells in complete media were used for SDC time-lapse microscopy as previously described[32] at 37°C.

Results

Fine-mapping a pancreatic-cancer-risk locus on chr16q23.1

To identify candidate functional variants at a pancreatic-cancer-risk locus on chromosome 16q23.1 initially described in 2014,[7] we performed fine-mapping by using GWAS data from our recent meta-analysis of 9,013 affected individuals and 12,452 control individuals of European ancestry.[6] The most significant genotyped variant at 16q23.1 was rs13337397 (OR = 1.28, 95% confidence interval [CI] 1.23–1.32, $p = 2.01 \times 10^{-13}$, MAF = 0.103 in the 1000G European EUR samples). Imputation (with the 1000G EUR reference panel) identified variants with lower p values, and rs72802365 was most significant (OR = 1.36, 95% CI 1.31–1.40, $p = 2.51 \times 10^{-17}$, MAF = 0.083; $r^2 = 0.77$, $D' = 0.99$ in the 1000G EUR samples). After conditioning the analysis on rs72802365, the signal for rs13337397 was no longer significant (p = 0.73). The same held true for rs7190458 (p = 0.22), the tag SNP initially described at this locus in a subset of the GWAS samples.[7] However, when conditioning the analysis on rs13337397, the signal for rs72802365 remained significant ($p = 8.0 \times 10^{-4}$), suggesting that the latter is an improved tag SNP for the same signal (Table 1, Figure S1). This variant lies in an approximately 12 kb genomic region between the transcriptional start sites of two chymotrypsinogen precursor genes, namely *CTRB1* and *CTRB2*, both of which encode digestive enzyme serine proteases. On the basis of LRs[16] compared to the most significant variant (rs72802365), 15 highly correlated variants were considered candidate functional variants (LR > 1:1,000) potentially underlying the association signal at this locus (Figure 1, Table S5). These variants, all highly correlated to rs72802365 ($r^2 > 0.78$ in 1000G EUR samples), are distributed over a 52 kb region on chr16q23.1 from 75,234,273 to 75,286,484 bp (Hg19). A Bayesian fine-mapping approach (SuSiE[17]) yielded one credible set at this locus with nine variants (all with LR ≥ 1:3) (Table S5). To be comprehensive, we

Please cite this article in press as: Jermusyk 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), https://doi.org/10.1016/j.ajhg.2021.09.002

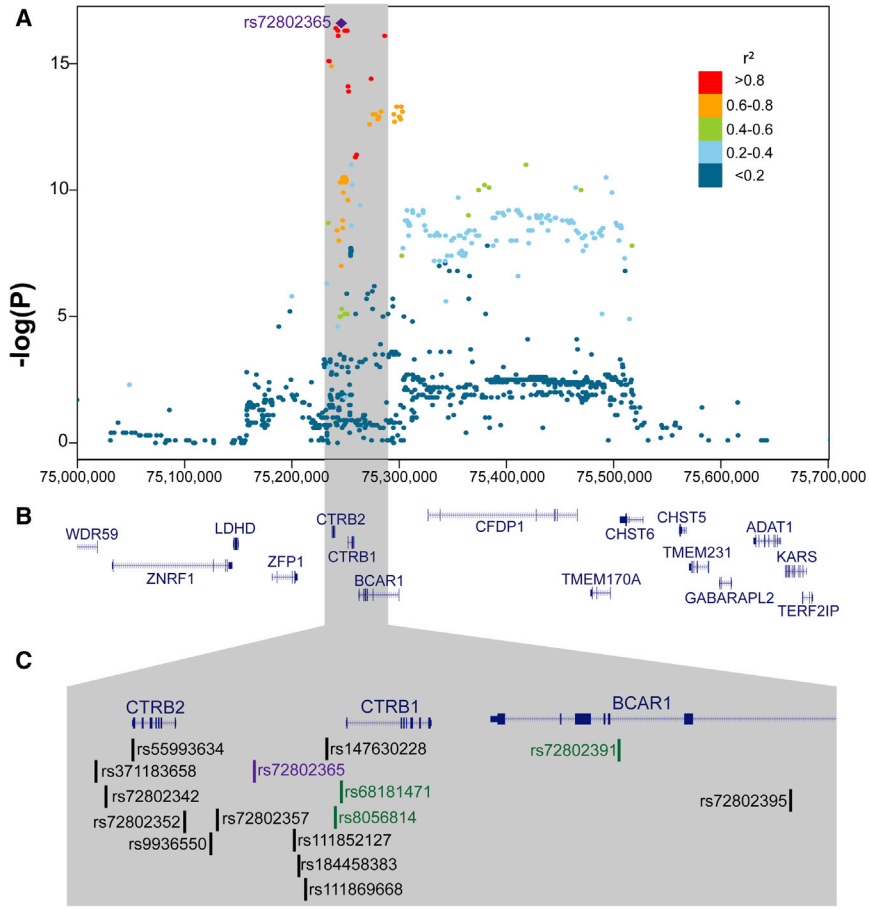

Figure 1. A Genomic map of the chr16q23.1 pancreatic risk locus

(A) Association results from a meta-analysis[6] of PanScan I, II, and III[7–10] and PanC4[11] with variants colored indicating linkage disequilibrium (LD) to tag SNP rs72802365 (red, $r^2 \geq 0.8$; yellow, $0.8 > r^2 \geq 0.6$; green, $0.6 > r^2 \geq 0.4$; light blue, $0.4 > r^2 \geq 0.2$; dark blue, $r^2 < 0.2$ in 1000G EUR).

(B) NCBI RefSeq genes within the chr16q23.1 risk locus as visualized with the UCSC genome browser.

(C) The fifteen candidate functional variants at the 16q23.1 risk locus (LR > 1:100 in black, LR > 1:1,000 in green, tag SNP rs7280265 in purple) distributed over a 52 kb region ranging from chr16: 75,234,273 to 75,286,484 (NCBI GRCh37/Hg19). Results after conditioning the association analysis on rs13337397, rs72802365, and the 584 bp *CTRB2* deletion variant are shown in Figure S1.

moved the 15 SNPs with LR > 1:1,000 forward to functional analysis.

Colocalization of the 16q23.1 pancreatic-cancer-risk locus with aberrant CTRB2 splicing

This association signal lies within a large topologically associating domain (TAD) (as visualized for pancreas tissue in the 3D Interaction Viewer and Database[33]) on 16q23.1 that spans 1.4 Mb and includes 25 genes (Figure S2). Because the 15 candidate functional variants are all noncoding, our initial hypothesis was that one or more of these variants influenced the strength of a noncoding gene regulatory element(s) and expression of one or more of the genes within the TAD in an allele-specific manner. To identify the gene(s) underlying this signal, we performed expression quantitative trait locus (eQTL) analysis for rs72802365 and expression of the 25 genes within the TAD by using RNA-seq and genotype data generated in histologically normal pancreatic tissue samples from two independent datasets (GTEx v.7, n = 212; LTG, n = 95).[18,19] Nominally significant eQTLs were noted for *CHST6* ($p_{GTEx} = 0.013$, $p_{LTG} = 0.114$) and *TMEM170A* ($p_{GTEx} = 0.78$, $p_{LTG} = 0.044$). Additional eQTLs were observed in other normal derived tissues (GTEx) for *BCAR1* (lowest $p = 3.9 \times 10^{-5}$ in whole blood), *TMEM170A* (lowest $p = 9.9 \times 10^{-9}$ in sun-exposed skin), and *CFDP1* (lowest $p = 2.0 \times 10^{-7}$ in visceral adipose tissue). However, colocalization analysis[25,34,35] indicated that these eQTL signals are not likely to share the same causal variant as the GWAS signal (posterior probability that the signals reflect different causal variants, PP = 0.69–1.00) (Table S6).

We next conducted mRNA splicing QTL analysis (sQTL) with the GTEx[18] and LTG[19] pancreatic datasets. In GTEx, the risk-increasing allele at rs72802365 (C) was strongly associated with decreased *CTRB2* exon 5–6 ($p = 5.30 \times 10^{-62}$, $\beta = -2.30$) and exon 6-7 ($p = 3.03 \times 10^{-47}$, $\beta = -1.97$) junction reads, while increased exon 5–7 junction reads were observed ($p = 1.40 \times 10^{-69}$, $\beta = 1.99$) (Figure 2A). This finding replicated in the LTG dataset (*CTRB2* exon 5–6: $p = 5.58 \times 10^{-7}$, $\beta = -1.10$; exon 6–7: $p = 6.81 \times 10^{-24}$, $\beta = -2.16$; exon 5–7: $p = 1.02 \times 10^{-30}$, $\beta = 1.99$) (Figures S3 and S4). Colocalization analysis[25,34,35] showed that the sQTL and GWAS signals are highly likely to share the same causal variant (posterior probability of same causal variants for exon 5–7 sQTL and GWAS signal, $PP_{GTEx} = 0.87$; $PP_{LTG} = 0.87$; Table S7).

Structural variants at the *CTRB1-CTRB2* locus and implications for pancreatic cancer risk

We hypothesized that the sQTL results could be due to the effects of one or more of the 15 candidate functional variants on *CTRB2* splicing or due to structural variant(s) not captured in the GWAS or the imputation panels used in the fine-mapping analysis described above. We did not find support for the first hypothesis because none of the candidate causal variants were predicted to influence splicing by bioinformatic analysis (data not shown). Structural variants, however, have previously been described in

Please cite this article in press as: Jermusyk 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), https://doi.org/10.1016/j.ajhg.2021.09.002

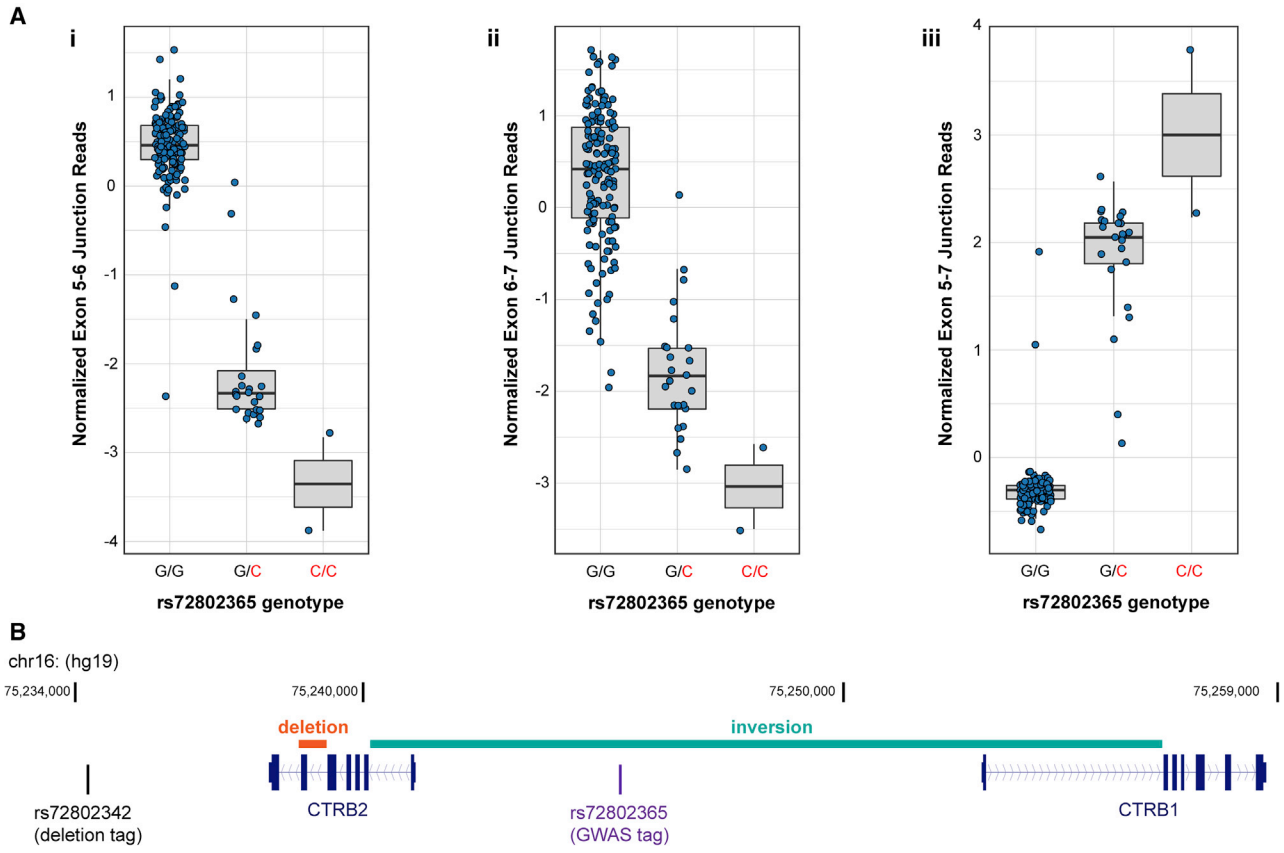

Figure 2. *CTRB2* splicing QTL at the chr16q23.1 pancreatic-cancer-risk locus
(A) Boxplots showing normalized *CTRB2* exon junction reads (Leafcutter[20]) in GTEx normal pancreas samples for *CTRB2* exon 5–6 (Ai), exon 6–7 (Aii), and exon 5–7 (Aiii) based on rs72802365 genotype (risk allele = C as indicated in red text). The number of samples per genotype are G/G, n = 148; G/C, n = 24; C/C, n = 2.
(B) Genomic map of the 584 bp *CTRB2* exon 6 deletion (orange) and 16.6 kb *CTRB2-CTRB1* inversion (teal) variants as well as the deletion tag SNP, rs72802342, and GWAS tag SNP, rs72802365. The location of the initial most significant genotyped SNP, rs13337397, is approximately 37 kb downstream of the start of *CTRB1* (chr16: 75,295,639) and is not shown here. Splicing QTLs for the LTG dataset are shown in Figure S3.

the *CTRB1*/*CTRB2* genomic region (Figure 2B), including a 584 bp deletion variant that completely overlaps exon 6 of *CTRB2* (as well as parts of introns 5 and 6).[36] This deletion ranged widely in frequency across populations (from 0%–47.2% in 871 HGDP-CEPH samples from 57 populations) in the study by Pang et al.,[36] however, it is not included in dbVar or 1000G. The deletion variant would be expected to result in a *CTRB2* mRNA that lacks exon 6 and could therefore explain the observed sQTL. To test this hypothesis, we PCR genotyped (see material and methods) the 1000G European ancestry samples (n = 438, average MAF for the 584 bp *CTRB2* deletion was 8.2%, range 6.1% in TSI to 12.5% in IBS, Tables S1 and S2) and noted high LD between the 584 bp *CTRB2* deletion variant and the top GWAS SNP rs72802365 (r^2 = 0.70, D′ = 0.85). The best tag SNP for the deletion was rs72802342 (r^2 = 0.81, D′ = 0.91, GWAS p value = 7.1×10^{-16}, OR = 1.34, 95% CI 1.29–1.39, LR = 1:27), one of the 15 candidate functional variants at 16q23.1 (Table S5). After imputing the 584 bp *CTRB2* deletion variant into the GTEx dataset, we noted very similar sQTL results to those described above for the most significant GWAS SNP (*CTRB2* exon 5–6 splicing: p = 3.67×10^{-72}, β = −2.35; exon 6–7 splicing: p = 4.88×10^{-54}, β = −2.03; and exon 5–7 splicing: p = 5.37×10^{-85}, β = 2.04) (Figure S5A). Likewise, similar results were seen in the LTG dataset (*CTRB2* exon 5–6 splicing: p = 6.23×10^{-9}, β = −1.25; exon 6–7 splicing: p = 5.65×10^{-30}, β = −2.29; and exon 5–7 splicing: p = 4.59×10^{-44}, β = 2.13) (Figure S5B). This suggests that the *CTRB2* exon 6 genomic deletion variant most likely underlies the sQTL results.

To determine whether the 584 bp *CTRB2* deletion variant (Figures S6A and S6b) was associated with pancreatic cancer risk, we imputed this variant into the pancreatic cancer GWAS datasets (PanScan I–III and PanC4)[7–9,11] by using the 1000G EUR reference panel augmented with the deletion genotypes (Tables S1, S2, and S4). A meta-analysis of these datasets showed that the deletion variant was associated with pancreatic cancer risk (OR = 1.36, p = 2.83×10^{-16}, LR = 1:11) to a similar degree as the most significant variant from the fine-mapping (rs72802365: OR = 1.36, p = 2.51×10^{-17}). The frequency of the deletion allele (0.073) is slightly lower than that for the most significant imputed SNP (rs72802365, 0.082). Conditioning the

analysis on either rs72802365 or the 584 bp *CTRB2* deletion variant resulted in a near complete loss of the association signal (after conditioning on rs72802365, $p_{584\ bp\ CTRB2\ deletion} = 0.12$; after conditioning on the 584 bp *CTRB2* deletion variant, $p_{rs72802365} = 0.007$), indicating that the structural variant marks the same association signal as rs72802365 (Table 1). Due to the lower imputation quality score for the deletion variant ($R^2 = 0.85$) as compared to rs72802365 ($R^2 = 0.99$), we applied noncentrality parameter (NCP) analysis to determine which of these variants is more likely to be causal (see material and methods) and observed greater support for the *CTRB2* deletion variant as a functional variant underlying the association signal (LR = 5.75).

In addition to the 584 bp deletion variant in *CTRB2*, a second structural variant at chr16q23.1 was described by Pang et al.[36] This variant, a 16.6 kb inversion (Figure 2B), spans the genomic region between the first introns of *CTRB2* and *CTRB1* and thereby swaps the first exons of the two chymotrypsinogen B genes without affecting the reading frame of the two chymotrypsinogen precursor proteins. The human reference allele for this inversion is the derived (i.e., inverted) allele as compared to the ancestral chimpanzee reference sequence (Figure S6C). The 584 bp deletion variant is only found on the background of the ancestral allele of this variant at 16q23.1, indicating that the *CTRB1*/*CTRB2* inversion is a more recent event.[36] Subsequent work by Rosendahl et al. described this inversion variant as a functional variant underlying a pancreatitis GWAS signal on chr16q23.1.[24] Because pancreatitis is a known risk factor for pancreatic cancer,[37,38] we sought to determine whether the *CTRB1*/*CTRB2* inversion variant was associated with risk of pancreatic cancer. We PCR genotyped this variant in the 1000G EUR samples ($r^2 = 0.017$, $D' = 1.0$ with both the most significant GWAS SNP, rs72802365, and the 584 bp *CTRB2* deletion; MAF for the inversion variant ranges from 9.1% in the CEU to 26.4% in the GBR 1000G European ancestry samples, Table S2) and imputed it into the pancreatic cancer GWAS datasets. Unlike the deletion, the inversion variant was not significantly associated with pancreatic cancer risk ($p = 0.26$). Furthermore, after conditioning the GWAS analysis on the 16.6 kb inversion variant, the association signals for the 584 bp *CTRB2* exon 6 genomic deletion variant and rs72802365 were not significantly affected ($p = 3.14 \times 10^{-17}$ and $p = 1.89 \times 10^{-18}$, respectively, Table 1). We therefore conclude that the inversion does not explain the underlying biology of this pancreatic-cancer-risk signal on chr16q23.1.

Because a genomic deletion of an entire exon may have a large effect on gene function and therefore pancreatic cancer risk, particularly in individuals homozygous for the *CTRB2* exon 6 deletion, we tested the association with pancreatic cancer risk by using a recessive model. This resulted in a larger estimated OR for rs72802365 ($OR_{Recessive} = 1.86$, 95% CI 1.60–2.17, $p = 4.80 \times 10^{-5}$) and the 584 bp *CTRB2* deletion variant ($OR_{Recessive} = 1.91$, 95% CI 1.61–2.28, $p = 1.77 \times 10^{-4}$) as compared to the additive model (Table 1). The percentage of control individuals included in the GWAS datasets who were homozygous for the deletion ranged from 0.42%–0.63%.

The *CTRB2* exon 6 deletion impairs chymotrypsin B2 function and localization

The 584 bp exon 6 deletion in *CTRB2* is expected to result in a shorter chymotrypsinogen B2 precursor (pre-proenzyme) protein (166aa instead of 263aa) because of a premature stop codon early in exon 7 (the last exon). The resulting protein lacks the third peptide chain (C chain) of chymotrypsin B2, including one of the three amino acids in the catalytic triad of the full-length chymotrypsinogen B2 digestive enzyme precursor, serine-213 (Figure 3A).[39] This prompted an assessment of chymotrypsin function and secretion for the full-length and truncated CTRB2 proteins. We cloned both versions of the *CTRB2* ORFs into expression plasmids and assessed the effects of the deletion on chymotrypsin activity in transiently transfected HEK293T cells. We observed a 7.6-fold lower chymotrypsin activity for the truncated as compared to the full-length CTRB2 protein (80 ± 5 μU/mg versus 606 ± 73 μU/mg, respectively, $p = 0.017$, t-statistic). In fact, the activity for the former was at background levels (93 ± 10 μU/mg) consistent with a total loss of chymotrypsin activity when exon 6 is missing from *CTRB2* (Figure 3B, Figure S7).

As the CTRB2 protein is activated in the small intestine after being secreted from pancreatic acinar cells, we sought to determine whether the deletion influenced secretion of the CTRB2 protein by comparing protein levels in cell lysates and conditioned media after transient transfections in HEK293T cells. Although the full-length CTRB2 protein was predominantly seen in the media, the opposite was noted for the truncated protein, indicating intracellular accumulation (Figure 3C, Figure S8). The amount of protein secreted (observed in the media as compared to the lysate) was significantly reduced for the truncated (18% ± 5%) as compared to the full-length CTRB2 protein (92% ± 3%, $p = 3.7 \times 10^{-4}$) (Figure 3D). Strong reductions in chymotrypsin activity and secretion were seen with both FLAG-tagged (Figures 3C and 3D, Figure S7 and S8) and nontagged (Figure S9) expression clones.

ER stress pathways are upregulated in pancreatic tissues from carriers of the *CTRB2* exon 6 deletion

To start to assess the biological consequences of the 584 bp deletion allele and the presence of a nonfunctional and poorly secreted CTRB2 protein in the pancreas, we compared gene expression in histologically normal human pancreatic tissue samples (GTEx v.7) with 0, 1, and 2 copies of the deletion variant by using a generalized linear model. We observed 950 genes positively correlated ($p < 0.05$) and 494 genes negatively correlated ($p < 0.05$) with the number of copies of the *CTRB2* exon 6 deletion. Gene set and pathway enrichment analyses yielded significant enrichment (false discovery rate $< 1 \times 10^{-4}$) for pathways related

Please cite this article in press as: Jermusyk 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), https://doi.org/10.1016/j.ajhg.2021.09.002

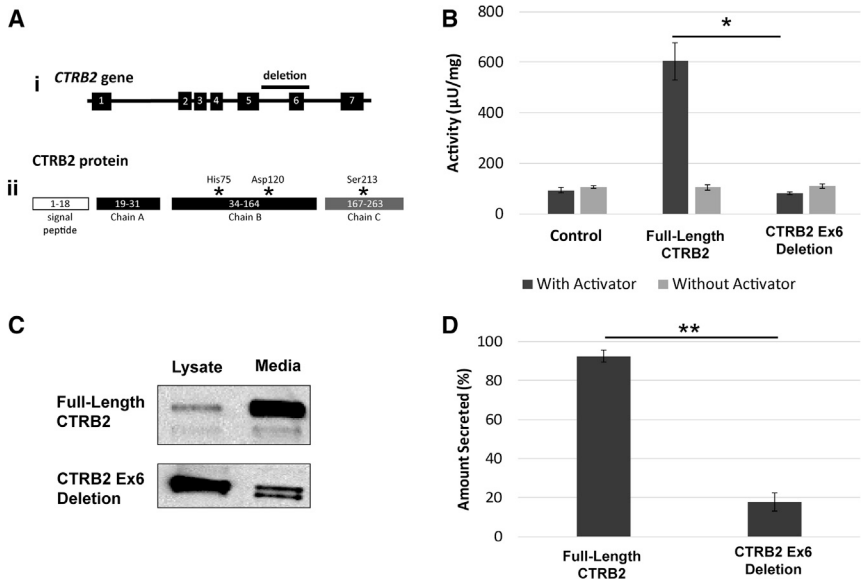

Figure 3. Effect of the exon 6 deletion on chymotrypsin B2 activity and localization

(A) (Ai) Schematic figure of the *CTRB2* gene showing the 584 bp deletion that overlaps all of exon 6 as well as parts of introns 5 and 6. (Aii) Full-length chymotrypsin B2 (CTRB2) protein contains a cleaved signal peptide and three peptide chains (A, B, and C) that are held together by disulfide bonds. When the exon 6 deletion is present, a premature stop codon is introduced (at aa 166 of the truncated protein; at aa 212 as compared to the full-length CTRB2 protein) resulting in the loss of chain C (amino acids 167–263) including one of the three amino acids forming the catalytic triad (denoted by *, Ser213 is lost).

(B) Chymotrypsin activity (μU/mg) calculated from triplicate experiments (* denotes $p < 0.05$) with (dark gray) and without (light gray; activity due to other proteases besides chymotrypsin) chymotrypsin activator (trypsin), error bars represent standard error of the mean (SEM).

(C) Amount of CTRB2 protein in cell lysates and media assessed by immunoprecipitation-immunoblot analysis in HEK293T cells transfected with plasmids expressing full-length and truncated (lacking exon 6) *CTRB2* (FLAG-tagged).

(D) The immunoblot image in (C) was taken with a ChemiDoc Touch Imaging System and bands were quantified with the ImageLab software (see material and methods). The amount of secreted proteins is summarized from triplicate experiments (** denotes $p < 0.001$, error bars represent SEM). The full immunoblot is shown in Figure S8. The effects of the exon 6 deletion on chymotrypsin activity and secretion via nontagged expression clones are shown in Figure S9.

to ER stress (GSEA: GO BP, protein localization to endoplasmic reticulum, establishment of protein localization to endoplasmic reticulum, and response to endoplasmic reticulum stress; IPA canonical pathways, unfolded protein response and endoplasmic reticulum stress pathway), translation (GSEA: Reactome, 3′ UTR-mediated translation regulation, translation, peptide chain elongation, signal-recognition particle [SRP] dependent co-translational protein targeting to membrane; and KEGG, ribosome), and EIF2 signaling (IPA) (Tables S8 and S9). Thus, individuals carrying the *CTRB2* 584 bp genomic deletion allele may exhibit increased levels of ER stress and altered translational activity in the pancreas.

To further study the effect of the *CTRB2* deletion variant on ER stress pathways, we examined the expression of eight canonical ER stress markers (Table S10). Two of these genes, *HSPA5* (encoding BiP) and *EIF2AK3* (encoding PERK), exhibited significantly increased expression in carriers of the *CTRB2* exon 6 deletion allele in the GTEx dataset ($p < 0.05$, $\beta > 1$).

Functional ER stress validation

The mechanism by which ER stress may be induced by the presence of the *CTRB2* deletion allele was examined in the HPDE pancreatic cell line, derived from normal pancreas, transfected with the GFP-tagged *CTRB2* exon 6 deletion and wild-type constructs. Cells transfected with the wild-type *CTRB2* construct showed minimal levels of the protein in the ER (Figure 4Ai) but some localization to the Golgi (Figure 4Bi, Figure S10), consistent with rapid intracellular transport and secretion. On the other hand, the truncated protein showed strong immunostaining in the ER (Figure 4Aii) and little to no colocalization with the Golgi (Figure 4Bii), indicating that it is retained in the ER. Live imaging validated these findings, showing rapid transport of full-length CTRB2 protein from the E, but not of the truncated CTRB2 protein (Figure S11). This results in an increase in ER stress, as demonstrated by an increase in BiP (endoplasmic reticulum chaperone BiP, also named GRP-78 and HSP70 family protein 5; gene name *HSPA5*) protein levels in cells expressing the deletion as compared to wild-type CTRB2 (Figure 4C and 4D, $p = 0.0002$). This was validated in PANC-1 cells (derived from pancreatic adenocarcinoma) (Figure S12).

This effect of the *CTRB2* exon 6 deletion on the ER stress pathway was further assessed via examination of gene expression (by qRT-PCR) of several ER-stress-related genes in multiple cell lines (MIA PaCa-2, PANC-1, and HEK293T) transfected with full-length and truncated *CTRB2* constructs (see material and methods). Expression of several key ER-pathway-related genes was increased (spliced *XBP1*, 1.78- to 2.59-fold, $p = 0.012$–0.035; *HSPA5/BiP*, 1.86- to 2.76-fold, $p = 0.007$–0.035; *CALR*, 1.49- to 2.12-fold, $p = 0.00$–0.228; and *EIF2AK3/PERK*, 1.36- to 1.50-fold, $p = 0.026$–0.230) in cells transfected with the *CTRB2* exon 6 deletion construct compared to those transfected with the wild-type *CTRB2* construct (Figure 4E). This was confirmed at the protein level for spliced XBP1 (1.52- to 1.60-fold increase, $p = 0.012$), HSPA5/BiP (2.26- to 2.59-fold, $p = 0.011$), CALR (1.31- to 2.99-fold, $p = 0.22$), and EIF2AK3/PERK (1.52- to 2.52-fold, $p = 0.021$) in PANC-1 and HEK293T cells expressing either truncated or full-length *CTRB2* constructs (Figure S13).

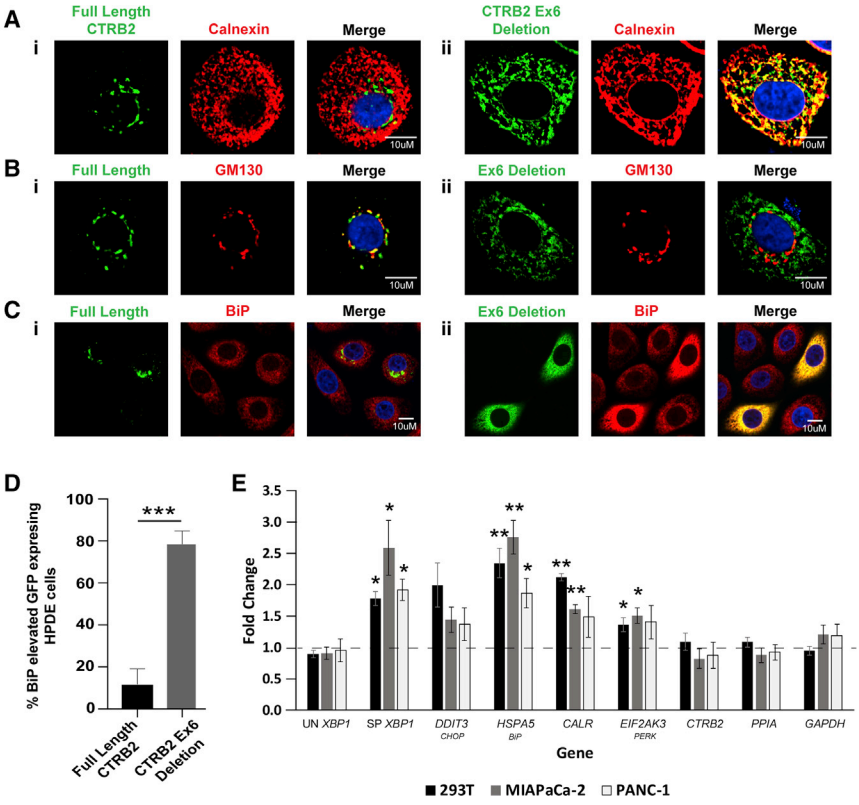

Figure 4. Effect of the truncated CTRB2 protein on ER stress

(A and B) Human normal pancreatic-derived HPDE cells transiently expressing GFP-tagged (green) full-length CTRB2 (Ai and Bi) and truncated (with exon 6 deleted) CTRB2 proteins (Aii and Bii) were immunostained with antibodies for the (A) endoplasmic reticulum marker calnexin (red) or (B) *cis*-Golgi marker GM130 (red). Representative single xy images from a *Z* stack processed with deconvolution software (Slidebook) are shown, taken with a spinning disk confocal microscope with a 63× oil objective and a sCMOS camera.

(C) Transiently transfected HPDE cells were immunostained for BiP (*HSPA5*), a marker of ER stress (red), and imaged by epifluorescence microscopy with a 63× oil objective. Representative images (Ci and Cii) from ten fields are shown. DAPI-stained nuclei shown in blue. Scale bar represents 10 μM.

(D) Quantification of elevated BiP (*HSPA5*) protein levels in GFP-expressing HPDE cells imaged as described in (C) (>40 GFP-expressing cells were analyzed in each of three independent experiments) (*** denotes $p < 0.005$, error bars represent SEM).

(E) Fold change increase in mRNA expression for ER stress pathway genes in cells expressing the truncated as compared to the full-length CTRB2 proteins. Expression was assessed by qRT-PCR analysis following transient overexpression of constructs containing full-length versus truncated (exon 6 deleted) *CTRB2* in HEK293T (black), MIA PaCa-2 (gray), and PANC-1 (white) cells. The data are calculated from four replicate experiments (* denotes $p < 0.05$, ** denotes $p < 0.01$, error bars represent SEM). Note that the fold change increase in protein expression for ER stress genes is shown in Figure S13.

The CTRB2 deletion variant and association with risk of diabetes mellitus

A shared type I diabetes (T1D)- and type II (T2D) diabetes-mellitus-risk locus has been reported on chr16q23.1.[40] Because a diagnosis of T2D increases risk of pancreatic cancer, we investigated whether the diabetes and pancreatic cancer GWAS signals colocalize. Both the T1D and T2D signals[41,42] were found to colocalize[25,34,35] with the pancreatic cancer GWAS signal ($PP_{T1D} = 1.00$; $PP_{T2D} = 0.97$) and with the *CTRB2* sQTL ($PP_{Average} = 0.84$, range 0.49–1.00, Table S11). This indicates that pancreatic cancer, T1D, and T2D share a functional variant at 16q23.1 and that the deletion variant in *CTRB2* may underlie the GWAS signals for these three diseases.

Discussion

We have fine-mapped and functionally characterized a pancreatic-cancer-risk locus on chr16q23.1. The 15 genotyped or imputed variants that mark this signal span a 52 kb region that overlaps three genes: *CTRB2*, *CTRB1* and *BCAR1*. The first two genes encode chymotrypsinogen preproenzymes, which are serine protease digestive enzyme precursors produced in large amounts by the exocrine pancreas and activated in the gut.[43] The third gene, *BCAR1*, encodes breast cancer anti-estrogen resistance protein 1 (also named p130Cas), a scaffold signal transduction protein that is important for multiple cellular pathways, including development, tissue homeostasis, cell cycle control, and motility,[44] and is overexpressed in multiple cancers.[44] Although the 15 potential functional variants underlying this risk locus were all noncoding, we did not observe eQTLs that colocalized with the GWAS signal. However, we identified an association with aberrant splicing of the *CTRB2* mRNA between exons 5 and 7, thus skipping exon 6. This led us to consider a previously reported but unannotated 584 bp structural deletion variant overlapping exon 6 (as well as parts of the adjacent introns) in *CTRB2*.[36] Genotyping and imputing this variant into our pancreatic cancer GWAS datasets showed that the deletion allele was associated with a substantially increased risk of pancreatic cancer, particularly in individuals homozygous for deletion alleles, and that the sQTL and GWAS signals colocalized.

The full-length chymotrypsinogen B2 precursor is a 263aa protein where the first 18aa comprise a signal peptide that is cleaved off the zymogen upon entry into the ER. Active chymotrypsin is formed by cleavage by trypsin at a single peptide bond (Arg33-Ile34) followed by chymotrypsin autolysis (at Tyr164-Asp165) yielding a protein containing three peptide chains (A, B, and C) linked by disulfide bonds.[39] Chain C (aa167–263) contains many of the amino acids responsible for substrate recognition as

well as Ser213, which is part of the catalytic triad.[39] The deletion allele of the 584 bp exon 6 structural variant creates a premature stop codon yielding a 166aa protein (148aa after the signal peptide is removed) lacking chain C. Functional assessment of this truncated CTRB2 protein is consistent with complete loss of chymotrypsin activity. Furthermore, we found that the truncated CTRB2 protein was poorly secreted *in vitro* and accumulated intracellularly in the ER, resulting in ER stress. This effect is likely to be tissue specific because chymotrypsin expression is almost exclusively localized to the exocrine pancreas.

Germline mutations in the *PRSS1*/*PRSS2* (encoding trypsinogen), *CFTR* (cystic fibrosis transmembrane conductance regulator), *SPINK1* (serine protease inhibitor Kazal-type 1), *CTRC* (chymotrypsinogen-C), and *CPA1* (carboxypeptidase A1) genes cause inherited forms of pancreatitis.[24,45,46] Most known mutations in these genes lead to an increase in intrapancreatic trypsin activation or a reduction in its degradation, which can lead to acute or chronic inflammation of the pancreas, loss of pancreatic function, and tissue damage.[46] A subset of mutations in two of these genes, *PRSS1* (cationic trypsinogen) and *CPA1* (carboxypeptidase A1), have been shown to result in misfolding of these digestive enzyme proteins, loss of secretion,[45] and an increase in ER stress.[45] Because pancreatitis is a risk factor for pancreatic cancer,[4] similar mechanisms that contribute to ER-stress-induced pancreatitis may promote pancreatic tumorigenesis.

Previous characterization of pancreatic-cancer-risk loci from GWASs has identified functional variants within enhancers that regulate gene expression in *cis*.[47,48] For instance, our work on a multi-cancer-risk locus on chr5p15.33 identified allele-specific binding of the ZNF148 transcription factor to rs36115365, resulting in altered regulation of expression for the nearby gene, *TERT*.[48] Similar work identified a 30 bp deletion variant in a regulatory element within a pancreatic-cancer-risk locus on chr13q22.1. This indel variant is bound in an allele-specific manner by transcription factors of the TCF/LEF family, physically interacts with a region upstream of *DIS3* (570 kb away), and modulates *DIS3* expression.[47] Such a mechanism is believed to underly most common risk loci identified through GWASs.[49] However, the risk locus at chr16q23.1 described here is strikingly different and does not appear to involve noncoding gene regulatory elements but rather a structural variant that results in aberrant skipping of *CTRB2* exon 6, intracellular accumulation of a truncated and nonfunctional chymotrypsin B2 protein, and ER stress.

This risk locus on chr16q23.1 has previously been reported in GWASs for both T1D and T2D.[41,42] Interestingly, although the T1D and T2D signals at this locus colocalize, suggesting a common causal variant, the risk alleles are opposite for these two diseases.[40] On the basis of our colocalization analysis, it is likely that the *CTRB2* exon 6 deletion variant underlies risk at this locus for pancreatic cancer as well as for T1D and T2D. This suggests that the expression of the truncated and nonfunctional *CTRB2* protein may increase risk for pancreatic cancer and T1D but lower risk for T2D. It is possible that the effect of the *CTRB2* exon 6 deletion variant on T1D risk could be mediated through ER stress and inflammation that may lead to the production of neoantigen peptides that promote autoimmunity.[50] However, the protective effect on T2D seems less clear, in part because T2D is an epidemiologic risk factor for pancreatic cancer.[51]

Our study has important strengths. The size of the GWAS datasets allowed fine-mapping of the chr16q23.1 risk signal to a set of only 15 candidate functional variants. Likewise, we were able to validate the *CTRB2* splicing QTL in two separate normal-derived pancreatic tissue eQTL datasets. Limitations to our study include the lack of non-European populations in the GWAS. Furthermore, the lack of human pancreatic cell line models with detectable levels of endogenous CTRB2 called for *in vitro* overexpression of *CTRB2* for most functional experiments. Going forward, human tissue samples and animal models will be used to assess the effects of the *CTRB2* exon 6 deletion on ER stress, pancreatic inflammation, and carcinogenesis in an *in vivo* setting.

In summary, our work has uncovered a structural variant overlapping *CTRB2* exon 6 as a functional variant underlying risk for pancreatic cancer at chr16q23.1. This 584 bp genomic deletion results in intracellular accumulation of a nonfunctional *CTRB2* protein leading to ER stress. The ensuing inflammation may be key for the increased risk of pancreatic cancer in carriers of this germline variant. Our work represents one of a few studies that have led to the identification of functional structural variants underlying GWAS signals that disrupt protein function in a striking manner.[52,53]

Data and code availability

All programs used for the analysis are listed in the material and methods and web resources sections. The PanScan and PanC4 GWAS data are available through dbGaP under accession numbers dbGaP: phs000206.v5.p3 and phs000648.v1.p1, respectively. The GTEx and NCI/LTG expression QTL datasets are available through dbGAP under accession numbers dbGaP: phs000424.v7.p2 and phs001776.v1.p1, respectively. All unique reagents generated in this study are available from Laufey T. Amundadottir, National Cancer Institute, National Institutes of Health, Bethesda, MD.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.09.002.

Consortia

The members of the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (PanC4) are Demetrius Albanes, Alan A. Arslan, Aurelio Barricarte Gurrea, Laura Beane-Freeman, Paige M. Bracci, Bas Bueno-de-Mesquita, Julie Buring, Federico Canzian, Mengmeng Du, Stephen Gallinger, J. Michael Gaziano, Graham G. Giles, Phyllis J. Goodman, Eric J. Jacobs, Mattias Johansson, Charles Kooperberg,

Loic LeMarchand, Nuria Malats, Rachel E. Neale, Salvatore Panico, Ulrike Peters, Francisco X. Real, Xiao-Ou Shu, Malin Sund, Marc Thornquist, Anne Tjønneland, Ruth C. Travis, Stephen K. Van Den Eeden, Kala Visvanathan, Wei Zheng, Donghui Li, Harvey A. Risch, Stephen J. Chanock, Peter Kraft, Brian M. Wolpin, Gloria M. Petersen, Rachael Z. Stolzenberg-Solomon, Alison P. Klein, and Laufey T. Amundadottir.

Acknowledgments

We thank Dominic Esposito at the Protein Expression Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD, and his team for cloning full-length and truncated *CTRB2* expression clones. We thank Sara Olson, now retired but formerly at the Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center in New York, NY, USA, for her collaboration and support in providing tissue samples for the LTG eQTL dataset. We thank Kevin Brown, Laboratory of Translational Genomics, DCEG, NCI, NIH for careful reading of the manuscript and helpful suggestions. We also thank Kari Rabe and Jennifer Brooks for coordinating Mayo Clinic patient biospecimen and data contributions to this study. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH, Bethesda, MD, USA. This study was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health. Additional funding information is listed in the supplemental information. The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

Declaration of interests

The authors declare no competing interests.

Received: October 1, 2020
Accepted: September 1, 2021
Published: September 23, 2021

Web resources

Biowulf, https://hpc.nih.gov/
Coriell, https://www.coriell.org/
dbGAP, https://www.ncbi.nlm.nih.gov/gap/
EdgeR, https://www.bioconductor.org/packages/release/bioc/html/edgeR.html
GSEA, https://www.gsea-msigdb.org/gsea/index.jsp
GTEx, https://gtexportal.org/home/
IPA, https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
MSigDB, https://www.gsea-msigdb.org/gsea/index.jsp
R Project, https://cran.r-project.org/web/packages/coloc/index.html

References

1. Siegel, R.L., Miller, K.D., and Jemal, A. (2019). Cancer statistics, 2019. CA Cancer J. Clin. *69*, 7–34.
2. Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., and Bray, F. (2018). Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur. J. Cancer *103*, 356–387.
3. Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., and Matrisian, L.M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. *74*, 2913–2921.
4. Stolzenberg-Solomon, R.Z., and Amundadottir, L.T. (2015). Epidemiology and Inherited Predisposition for Sporadic Pancreatic Adenocarcinoma. Hematol. Oncol. Clin. North Am. *29*, 619–640.
5. Amundadottir, L.T. (2016). Pancreatic cancer genetics. Int. J. Biol. Sci. *12*, 314–325.
6. Klein, A.P., Wolpin, B.M., Risch, H.A., Stolzenberg-Solomon, R.Z., Mocci, E., Zhang, M., Canzian, F., Childs, E.J., Hoskins, J.W., Jermusyk, A., et al. (2018). Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat. Commun. *9*, 556.
7. Wolpin, B.M., Rizzato, C., Kraft, P., Kooperberg, C., Petersen, G.M., Wang, Z., Arslan, A.A., Beane-Freeman, L., Bracci, P.M., Buring, J., et al. (2014). Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat. Genet. *46*, 994–1000.
8. Amundadottir, L., Kraft, P., Stolzenberg-Solomon, R.Z., Fuchs, C.S., Petersen, G.M., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., Helzlsouer, K., Jacobs, E.J., et al. (2009). Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat. Genet. *41*, 986–990.
9. Petersen, G.M., Amundadottir, L., Fuchs, C.S., Kraft, P., Stolzenberg-Solomon, R.Z., Jacobs, K.B., Arslan, A.A., Bueno-de-Mesquita, H.B., Gallinger, S., Gross, M., et al. (2010). A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat. Genet. *42*, 224–228.
10. Zhang, M., Wang, Z., Obazee, O., Jia, J., Childs, E.J., Hoskins, J., Figlioli, G., Mocci, E., Collins, I., Chung, C.C., et al. (2016). Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget *7*, 66328–66343.
11. Childs, E.J., Mocci, E., Campa, D., Bracci, P.M., Gallinger, S., Goggins, M., Li, D., Neale, R.E., Olson, S.H., Scelo, G., et al. (2015). Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat. Genet. *47*, 911–916.
12. Tak, Y.G., and Farnham, P.J. (2015). Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin *8*, 57.
13. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
14. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. *39*, 906–913.
15. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
16. Spencer, A.V., Cox, A., and Walters, K. (2014). Comparing the efficacy of SNP filtering methods for identifying a single

causal SNP in a known association region. Ann. Hum. Genet. *78*, 50–61.

17. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping (J. R. Stat. Soc. Ser. B Stat. Methodol).
18. Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B., et al.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; and NIH/NCI (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.
19. Zhang, M., Lykke-Andersen, S., Zhu, B., Xiao, W., Hoskins, J.W., Zhang, X., Rost, L.M., Collins, I., Bunt, M.V., Jia, J., et al. (2018). Characterising *cis*-regulatory variation in the transcriptome of histologically normal and tumour-derived pancreatic tissues. Gut *67*, 521–533.
20. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. *50*, 151–158.
21. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics *32*, 1479–1485.
22. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. *6*, e1000770.
23. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.
24. Rosendahl, J., Kirsten, H., Hegyi, E., Kovacs, P., Weiss, F.U., Laumen, H., Lichtner, P., Ruffert, C., Chen, J.M., Masson, E., et al. (2018). Genome-wide association study identifies inversion in the *CTRB1-CTRB2* locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. Gut *67*, 1855–1863.
25. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
26. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.
27. Lun, A.T.L., Chen, Y., and Smyth, G.K. (2016). It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. Methods Mol. Biol. *1418*, 391–416.
28. Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. Bioinformatics *30*, 523–530.
29. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.
30. Liu, N., Furukawa, T., Kobari, M., and Tsao, M.S. (1998). Comparative phenotypic studies of duct epithelial cell lines derived from normal human pancreas and pancreatic carcinoma. Am. J. Pathol. *153*, 263–269.
31. Westlake, C.J., Baye, L.M., Nachury, M.V., Wright, K.J., Ervin, K.E., Phu, L., Chalouni, C., Beck, J.S., Kirkpatrick, D.S., Slusarski, D.C., et al. (2011). Primary cilia membrane assembly is initiated by Rab11 and transport protein particle II (TRAPPII) complex-dependent trafficking of Rabin8 to the centrosome. Proc. Natl. Acad. Sci. USA *108*, 2759–2764.
32. Insinna, C., Lu, Q., Teixeira, I., Harned, A., Semler, E.M., Stauffer, J., Magidson, V., Tiwari, A., Kenworthy, A.K., Narayan, K., et al. (2019). Investigation of F-BAR domain PACSIN proteins uncovers membrane tubulation function in cilia assembly and transport. Nat. Commun. *10*, 1–17.
33. Yang, D., Jang, I., Choi, J., Kim, M.S., Lee, A.J., Kim, H., Eom, J., Kim, D., Jung, I., and Lee, B. (2018). 3DIV: A 3D-genome Interaction Viewer and database. Nucleic Acids Res. *46* (D1), D52–D57.
34. Wallace, C. (2013). Statistical testing of shared genetic control for potentially related traits. Genet. Epidemiol. *37*, 802–813.
35. Plagnol, V., Smyth, D.J., Todd, J.A., and Clayton, D.G. (2009). Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. Biostatistics *10*, 327–334.
36. Pang, A.W.C., Migita, O., Macdonald, J.R., Feuk, L., and Scherer, S.W. (2013). Mechanisms of formation of structural variation in a fully sequenced human genome. Hum. Mutat. *34*, 345–354.
37. Raimondi, S., Lowenfels, A.B., Morselli-Labate, A.M., Maisonneuve, P., and Pezzilli, R. (2010). Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. Best Pract. Res. Clin. Gastroenterol. *24*, 349–358.
38. Duell, E.J., Lucenteforte, E., Olson, S.H., Bracci, P.M., Li, D., Risch, H.A., Silverman, D.T., Ji, B.T., Gallinger, S., Holly, E.A., et al. (2012). Pancreatitis and pancreatic cancer risk: a pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4). Ann. Oncol. *23*, 2964–2970.
39. Graf, L., and Szilagyi, L. (2013). Venekei, I (Chymotrypsin. In Handbook of Proteolytic Enzymes), pp. 2626–2633.
40. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N., and Gaulton, K.J. (2018). Shared genetic risk contributes to type 1 and type 2 diabetes etiology. Hum. Mol. Genet. https://doi.org/10.1093/hmg/ddy314.
41. Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., et al. (2009). Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat. Genet. *41*, 703–707.
42. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. *50*, 1505–1513.
43. Honey, N.K., Sakaguchi, A.Y., Lalley, P.A., Quinto, C., MacDonald, R.J., Craik, C., Bell, G.I., Rutter, W.J., and Naylor, S.L. (1984). Chromosomal assignments of genes for trypsin, chymotrypsin B, and elastase in mouse. Somat. Cell Mol. Genet. *10*, 377–383.
44. Camacho Leal, Mdel.P., Sciortino, M., Tornillo, G., Colombo, S., Defilippi, P., and Cabodi, S. (2015). p130Cas/BCAR1 scaffold protein in tissue homeostasis and pathogenesis. Gene *562*, 1–7.
45. Sahin-Tóth, M. (2017). Genetic risk in chronic pancreatitis: the misfolding-dependent pathway. Curr. Opin. Gastroenterol. *33*, 390–395.

46. Hegyi, E., and Sahin-Tóth, M. (2017). Genetic Risk in Chronic Pancreatitis: The Trypsin-Dependent Pathway. Dig. Dis. Sci. *62*, 1692–1701.
47. Hoskins, J.W., Ibrahim, A., Emmanuel, M.A., Manmiller, S.M., Wu, Y., O'Neill, M., Jia, J., Collins, I., Zhang, M., Thomas, J.V., et al. (2016). Functional characterization of a chr13q22.1 pancreatic cancer risk locus reveals long-range interaction and allele-specific effects on DIS3 expression. Hum. Mol. Genet. *25*, 4726–4738.
48. Fang, J., Jia, J., Makowski, M., Xu, M., Wang, Z., Zhang, T., Hoskins, J.W., Choi, J., Han, Y., Zhang, M., et al. (2017). Functional characterization of a multi-cancer risk locus on chr5p15.33 reveals regulation of TERT by ZNF148. Nat. Commun. *8*, 15034.
49. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet. *102*, 717–730.
50. Thomaidou, S., Zaldumbide, A., and Roep, B.O. (2018). Islet stress, degradation and autoimmunity. Diabetes Obes. Metab. *20* (*Suppl 2*), 88–94.
51. De Souza, A., Irfan, K., Masud, F., and Saif, M.W. (2016). Diabetes Type 2 and Pancreatic Cancer: A History Unfolding. JOP *17*, 144–148.
52. Long, J., Delahanty, R.J., Li, G., Gao, Y.T., Lu, W., Cai, Q., Xiang, Y.B., Li, C., Ji, B.T., Zheng, Y., et al. (2013). A common deletion in the APOBEC3 genes and breast cancer risk. J. Natl. Cancer Inst. *105*, 573–579.
53. Prokunina-Olsson, L., Muchmore, B., Tang, W., Pfeiffer, R.M., Park, H., Dickensheets, H., Hergott, D., Porter-Gill, P., Mumy, A., Kohaar, I., et al. (2013). A variant upstream of IFNL3 (IL28B) creating a new interferon gene IFNL4 is associated with impaired clearance of hepatitis C virus. Nat. Genet. *45*, 164–171.

The American Journal of Human Genetics, Volume *108*

Supplemental information

A 584 bp deletion in *CTRB2* inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer

Ashley Jermusyk, Jun Zhong, Katelyn E. Connelly, Naomi Gordon, Sumeth Perera, Ehssan Abdolalizadeh, Tongwu Zhang, Aidan O'Brien, Jason W. Hoskins, Irene Collins, Daina Eiser, Chen Yuan, PanScan Consortium, PanC4 Consortium, Harvey A. Risch, Eric J. Jacobs, Donghui Li, Mengmeng Du, Rachael Z. Stolzenberg-Solomon, Alison P. Klein, Jill P. Smith, Brian M. Wolpin, Stephen J. Chanock, Jianxin Shi, Gloria M. Petersen, Christopher J. Westlake, and Laufey T. Amundadottir

SUPPLEMENTARY MATERIAL

Consortia authors for the Pancreatic Cancer Cohort Consortium (PanScan) and the Pancreatic Cancer Case Control Consortium (PanC4).

Demetrius Albanes[1], Alan A. Arslan[2], Aurelio Barricarte Gurrea[3], Laura Beane-Freeman[4], Paige M. Bracci[5], Bas Bueno-de-Mesquita[6], Julie Buring[7], Federico Canzian[8], Mengmeng Du[9], Stephen Gallinger S[10], J. Michael Gaziano[11], Graham G. Giles[12], Phyllis J. Goodman[13], Eric Jacobs[14], Mattias Johansson[15], Charles Kooperberg[16], Loic LeMarchand[17], Nuria Malats[18], Rachel E. Neale[19], Salvatore Panico[20], Ulrike Peters[16], Francisco X. Real[21], Xiao-Ou Shu[22], Malin Sund[23], Marc Thornquist[16], Anne Tjønneland[24], Ruth C. Travis[25], Stephen K. Van Den Eeden[26], Kala Visvanathan K[27], Wei Zheng[22], Donghui Li[28], Harvey A. Risch[29], Stephen J. Chanock S[30], Peter Kraft P[31], Brian M. Wolpin[32], Gloria M. Petersen[33], Rachael Z. Stolzenberg-Solomon[30], Alison P. Klein[34], Laufey T. Amundadottir[35]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD,USA, [2]Department of Obstetrics and Gynecology, Department of Environmental Medicine, and Department of Population Health, New York University School of Medicine, New York, NY, USA, [3]Instituto de Salud Pública de Navarra, Pamplona, Spain, [4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [5]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA,USA, [6]Department for Determinants of Chronic Diseases (DCD), National Institute for Public Health and the Environment (RIVM), Utrecht, The Netherlands, [7]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA, [8]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, [9]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA, [10]Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada, [11]Departments of Medicine, Brigham and Women's Hospital, VA Boston, and Harvard Medical School, Boston, MA, USA., [12]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia.; Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia.; Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, VIC, Australia, [13]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA or Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [14]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA, [15]International Agency for Research on Cancer (IARC), 150 cours Albert Thomas, Lyon, France, [16]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [17]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, [18]Genetic and Molecular Epidemiology Group, Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain, [19]Population Health Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia, [20]Dipartimento di Medicina Clinica e Chirugia, Federico II University, Naples, Italy, [21]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), 28029 Madrid, Spain., Epithelial Carcinogenesis Group, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain., Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08193 Barcelona, Spain. [22]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, [23]Department of Surgical and Perioperative Sciences, Umeå University, Umeå, Sweden, [24]Danish Cancer Society Research Center and Department of Public Health, University of Copenhagen, Copenhagen, Denmark, [25]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 0NR, UK, [26]Kaiser Permanente Division of Research, Oakland, CA, USA, [27]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [28]Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX,USA, [29]Department of Epidemiology and Public Health, Yale School of Public Health, New Haven, CT, USA, [30]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Rockville, MD, USA, [31]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA, [32]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, USA, [33]Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, [34]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, USA, [35]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

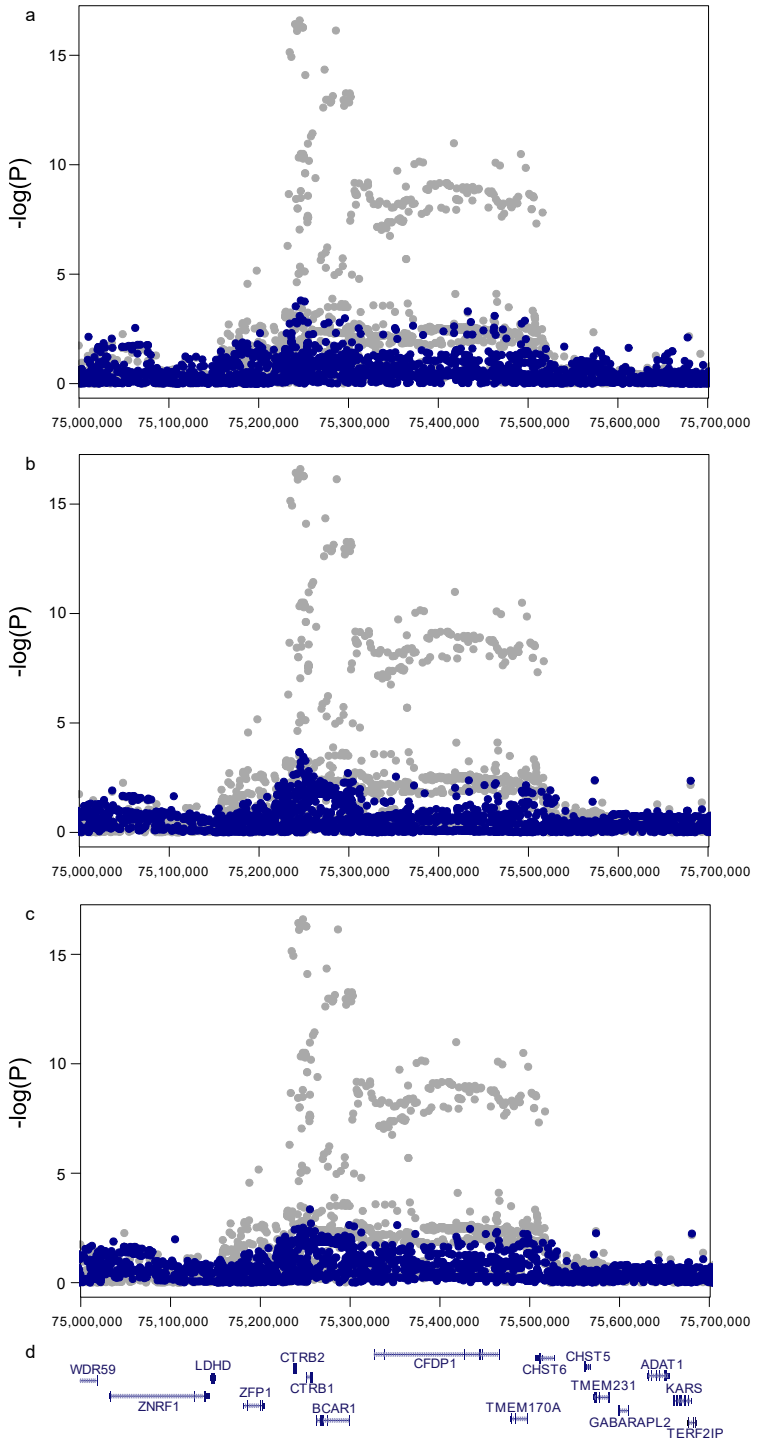

Fig. S1: Results from the association analysis at the 16q23.1 pancreatic cancer risk locus after conditioning the analysis on select variants. The association results from the meta-analysis of PanScan I, II, III, and PanC4 before (gray) and after conditioning (blue) on: (**a**) the original marker SNP rs13337397, (**b**) the most significant tag SNP after imputation rs72802365, and, (**c**) the 584 bp genomic insertion/deletion variant in *CTRB2*. NCBI RefSeq genes within the chr16q23.1 TAD as visualized using UCSC genome browser (Hg19) are shown in (**d**).

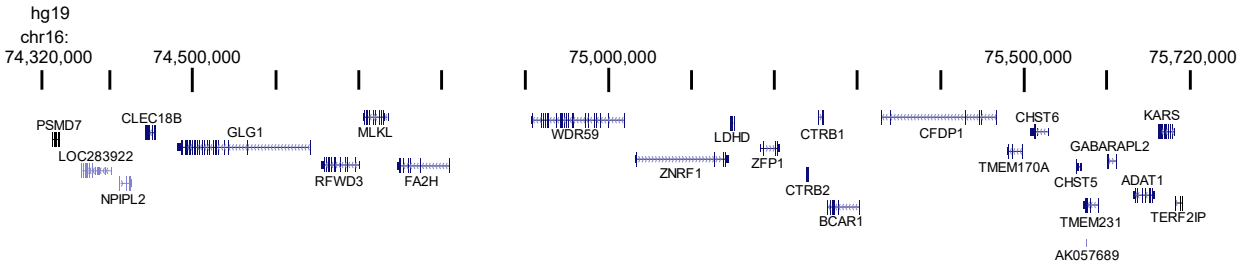

Fig. S2: The genomic region of the topologically associated domain (TAD) at the chr16q23.1 pancreatic cancer risk locus. NCBI RefSeq genes within the chr16q23.1 TAD are visualized using the UCSC genome browser (Hg19).

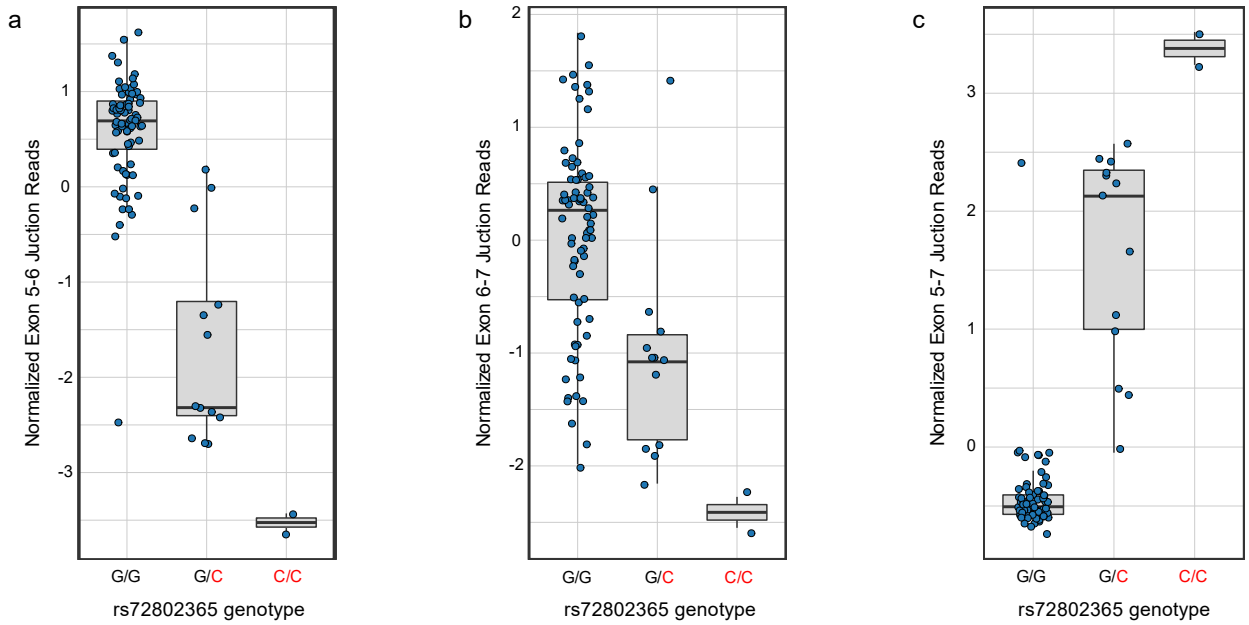

Fig. S3: *CTRB2* splicing QTLs at the chr16q23.1 pancreatic cancer risk locus for rs72802365 in the LTG (histologically normal) pancreatic QTL dataset. Boxplots show normalized *CTRB2* junction reads generated by Leafcutter in LTG pancreas samples for *CTRB2* Exon 5-6 (**a**), Exon 6-7 (**b**), and Exon 5-7 (**c**). The risk increasing allele (C) is indicated in red. Genotype counts are: G/G n=75, G/C n=13, C/C n=2.

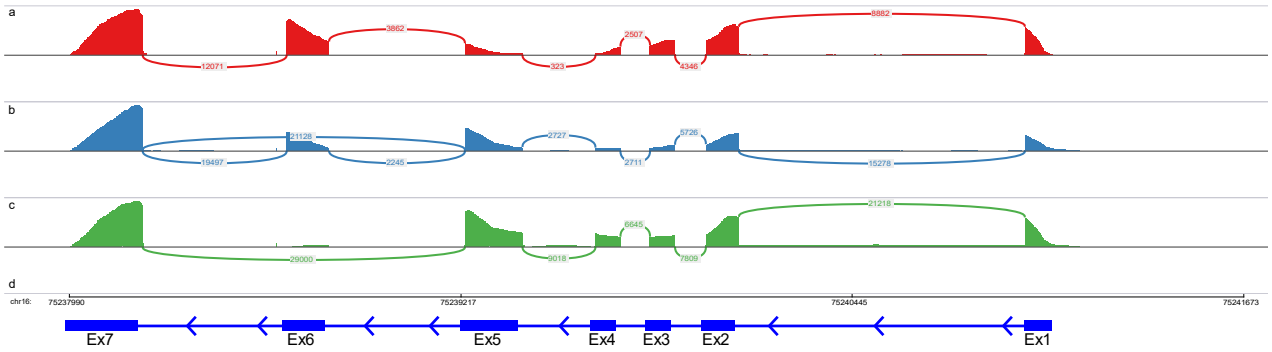

Fig. S4: A Sashimi plot showing *CTRB2* splicing in individuals with different genotypes for the *CTRB2* exon 6 indel variant. Sashimi plot of raw RNA-seq exonic read densities from representative individuals showing junction reads with: (**a**) no copies, (**b**) one copy, and (**c**) two copies, of the *CTRB2* exon 6 genomic deletion as visualized using the IGV browser[1]. Note the difference in RNA-seq read density over exon 6 in samples with 0 (a), 1 (b) and 2 (c) copies of the deletion allele. (**d**), Genomic coordinates (Hg19) and the exon/intron structure of the *CTRB2* gene are shown.

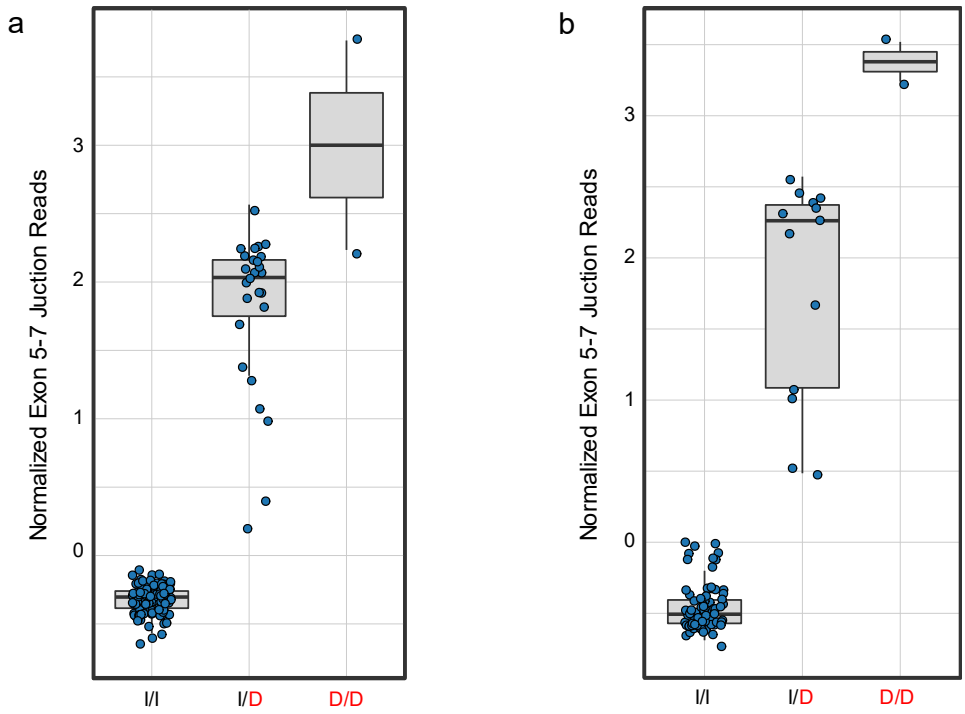

Fig. S5: Splicing QTLs for *CTRB2* and the 584 bp *CTRB2* genomic insertion/deletion variant in the LTG and GTEx histologically normal) pancreatic QTL datasets. The boxplots show normalized *CTRB2* junction reads from Leafcutter in GTEx samples (**a**, I/I n=148, I/D n=24, D/D n=2) and LTG pancreas samples (**b**, I/I n =75, I/D n=13, D/D n=2). I: insertion allele, D: deletion allele. The risk increasing allele (deletion) is indicated in red).

a

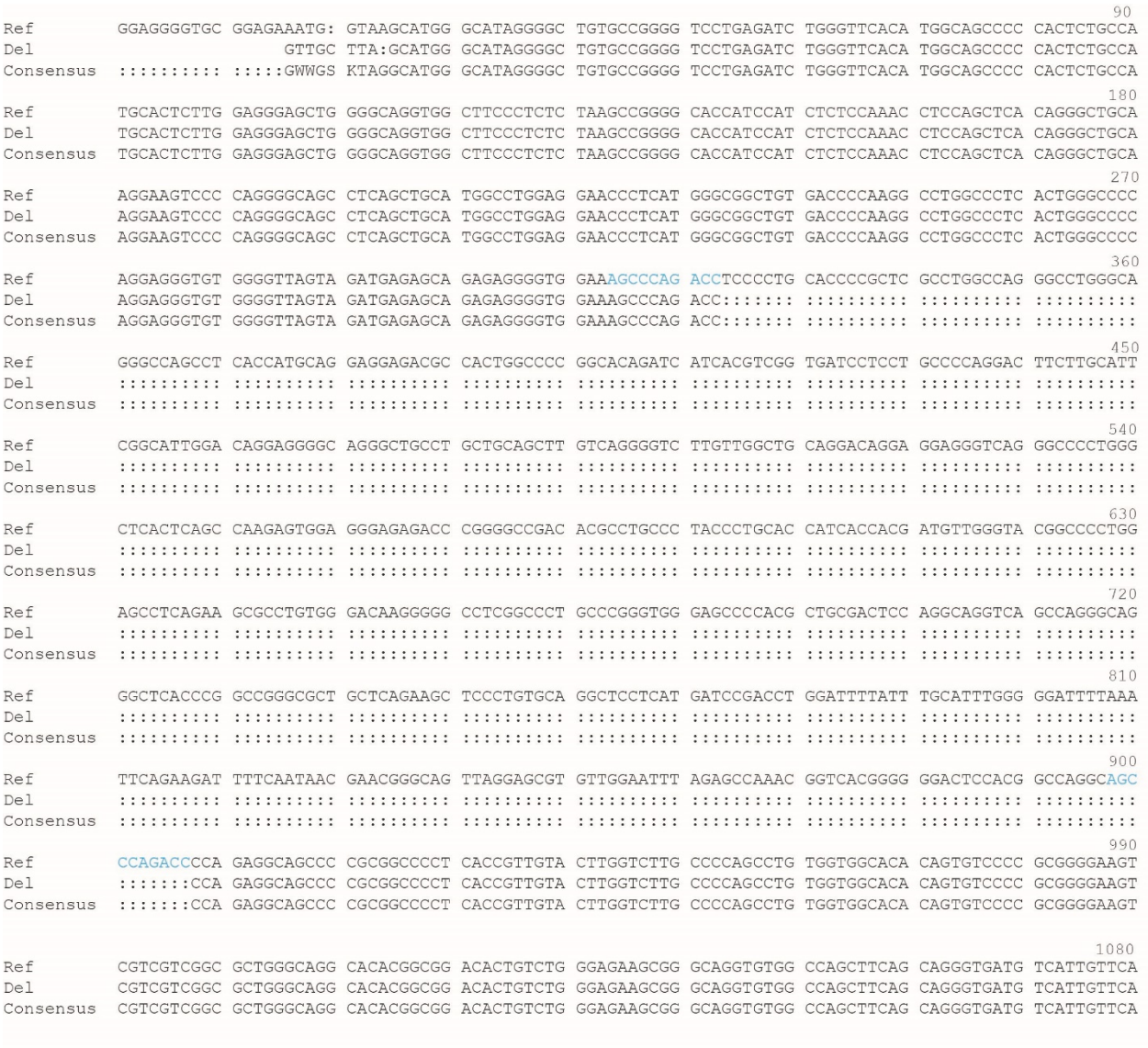

Fig. S6a: Nucleotide sequence of the *CTRB2* exon 6 deletion variant in a 1000G EUR sample. Nucleotide sequence of a 1000 Genomes subject identified to be homozygous for the *CTRB2* Ex6 deletion aligned to the reference sequence reveals a 584 bp deletion. The coordinates for the 584 bp deletion are ambiguous due to a repeated sequence at the beginning and end of the deleted sequence (marked in blue letters) and could range from Hg19: chr16:75,238,611-75,239,194 to chr16:75,238,621-75,239,204. The coordinates in the manuscript are listed as Hg19: chr16:75,238,616-75,239,199 +/- 5bp to reflect this ambiguity.

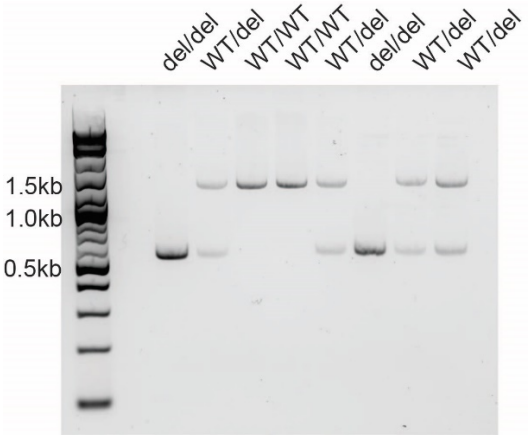

Fig S6b: Nucleotide sequence of the *CTRB2* exon 6 deletion variant in LTG eQTL samples. PCR was performed on genomic DNA from 8 human samples from the LTG eQTL dataset to amplify the *CTRB2* ex5-ex7 region (primers are listed in **Table S3**). The upper band represents the insertion allele with a 1112 bp product; the lower band represents the *CTRB2* exon 6 deletion allele with a 528 bp product (top figure). The two *CTRB2* exon 6 homozygous del/del samples are in lanes 1 and 6. Bands were excised and purified for sequencing (see nucleotide sequence in bottom part of Figure S6b below).

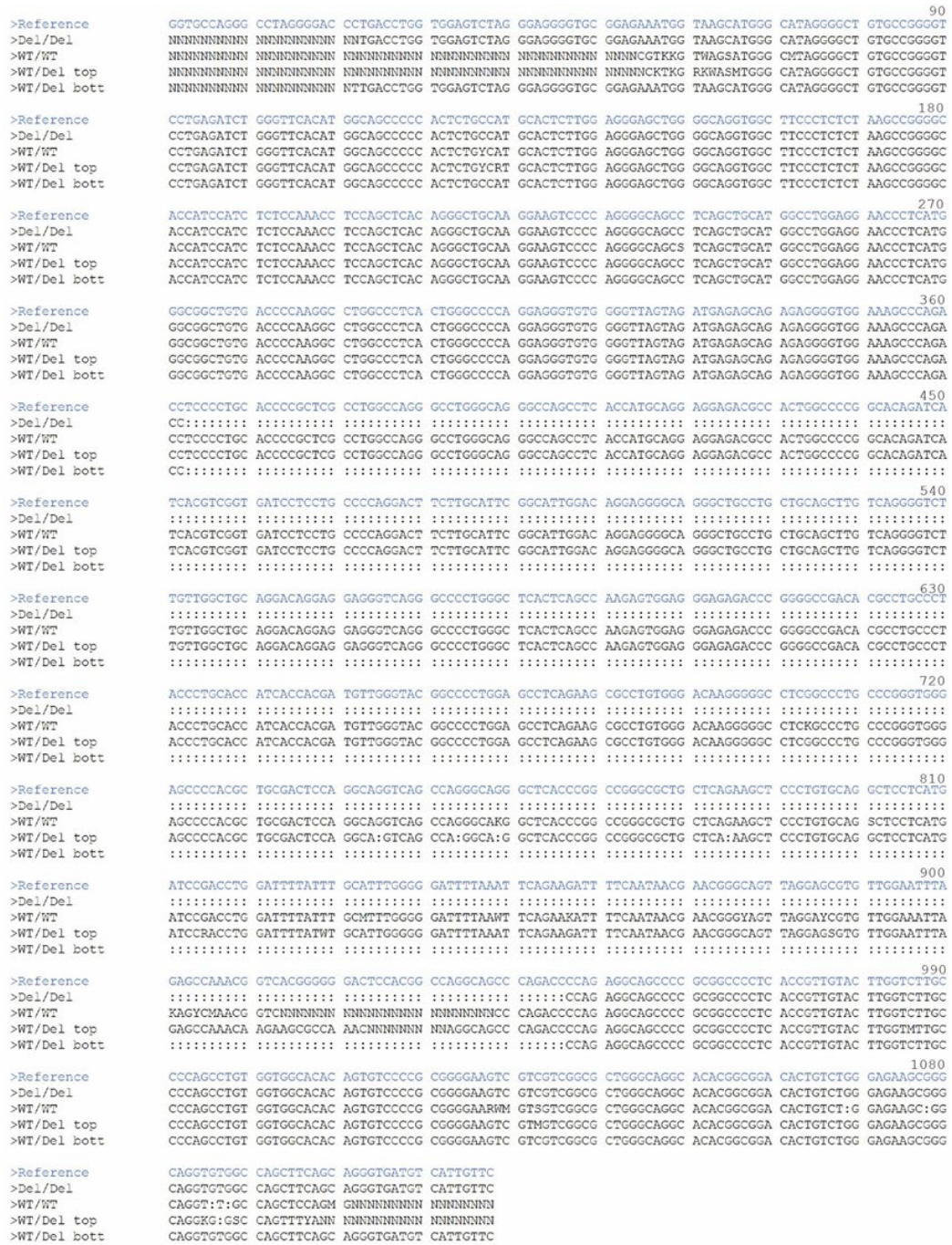

Fig S6b (continued): Bands from the gel above were excised and purified for sequencing. Top line: "Reference" represents the Human Genome Reference sequence (in blue text); "Del/Del" represents one of the homozygous *CTRB2* exon 6 deletion carriers; "WT/WT" represents a sample with no copies of the deletion; "WT/Del top" represents the top band from the PCR for a heterozygous WT/del carrier; "WT/Del bott" represents the lower band from a heterozygous WT/del carrier. Note the missing 584 bps corresponding to the *CTRB2* deletion allele in the Del/Del carrier and the bottom band from the WT/del carrier. "N" indicates missing sequencing from shorter reads, and non-A, T, C, G, N indicates ambiguous results from the sequencing.

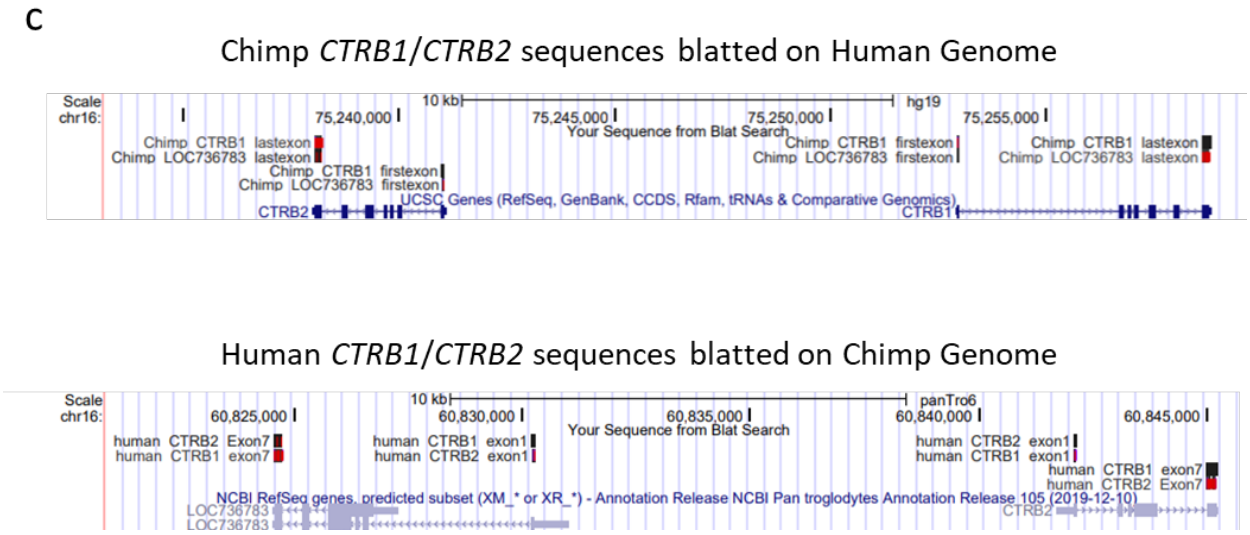

Fig. S6c: Ancestral and derived alleles for the *CTRB1*-exon1/*CTRB2*-exon1 inversion variant. Ancestral and derived alleles for the *CTRB1*/*CTRB2* inversion variant. The nucleotide sequence of human *CTRB1* and *CTRB2* genes on chr16q23.1 is identical apart from exons 1 and 7. To determine if the Human Reference allele is the inverted or non-inverted allele, we compared exons 1 and 7 between human and chimpanzee (*Pan troglodytes*) by BLAT in the UCSC Genome Browser. Note that while the chimp *CTRB1* gene (gene on right side in lower panel) is named *CTRB2*, it is listed as "Also known as *CTRB1*" in NCBI (https://www.ncbi.nlm.nih.gov/gene/736467). Similarly, while the chimp *CTRB2* gene (gene on left side of lower panel) is named LOC736783, it is listed as "Chymotrypsinogen B" and "*CTRB2*" in NCBI (https://www.ncbi.nlm.nih.gov/gene/736783) under General Protein Information. Red tick-marks within the aligned segments of the BLAT track represent substitutions with respect to the reference sequence. The **top panel** shows the BLAT alignment of exons 1 and 7 of chimp *CTRB2* (LOC736783) and *CTRB1* to the hg19 Human Reference genome in UCSC Genome Browser. Chimp *CTRB1* exon 1 aligns best with human *CTRB2* exon 1 and chimp *CTRB2* (LOC736783) exon 1 aligns best to human *CTRB1*. The last exon of chimp *CTRB1* and *CTRB2* (LOC736783) align best to their respective human genes. The **bottom panel** shows the BLAT alignment of human *CTRB2* and *CTRB1* exon 1 and exon 7 to the chimp reference genome. This shows that human *CTRB1* exon 1 aligns best with chimp *CTRB2* (LOC736783), and human *CTRB2* exon 1 aligns best with chimp *CTRB1* (listed as *CTRB2* in chimp browser but is *CTRB1* according to NCBI). Also note the large intron 1 of human *CTRB1* (top panel) and chimp *CTRB2* (LOC736783, bottom panel). The better alignment of human exon 1 to the opposite chimp chymotrypsinogen precursor gene, and the fact that the large intron 1 is part of *CTRB1* in the Human Reference genome but *CTRB2* (LOC736783) in the Chimp Reference genome, indicates that the human alternate allele is ancestral while the human reference allele is the inverted (and derived) allele.

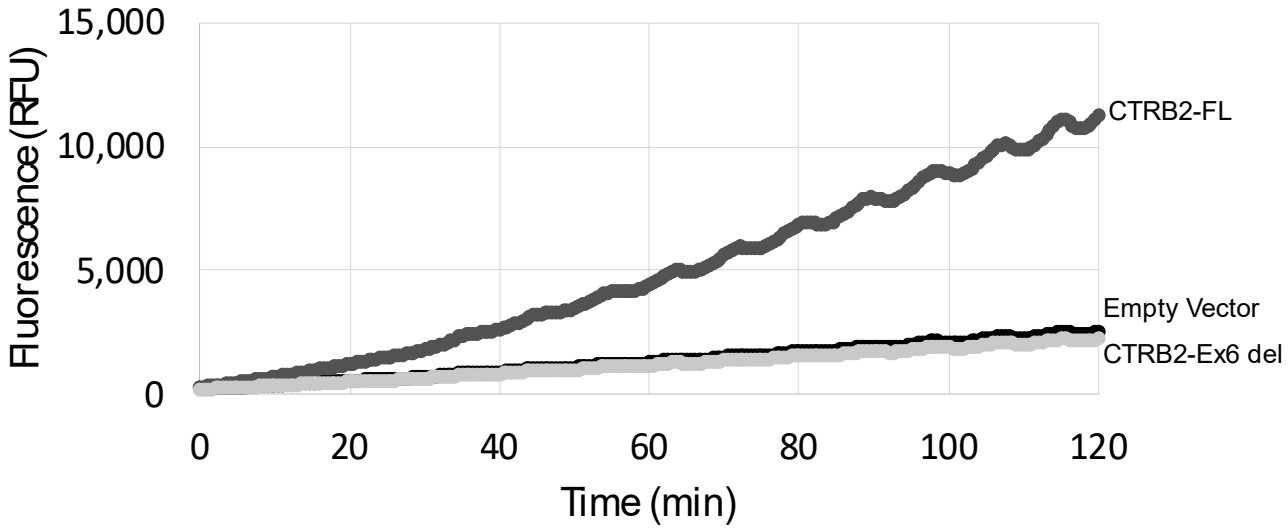

Fig. S7: Effect of the *CTRB2* exon 6 deletion on chymotrypsin reaction kinetics. A representative experiment showing fluorescence (RFU) as a measure of chymotrypsin activity on a synthetic fluorogenic substrate in HEK293T cell lysates after transient transfection with empty vector control (black), FLAG-tagged full-length *CTRB2* (CTRB2-FL, dark gray), and FLAG-tagged *CTRB2* with the exon-6 deletion (CTRB2-Ex6 del, light gray) plasmids. The linear change in fluorescence over a given time period was used to calculate activity, quantification of activity across triplicate experiments is shown in Fig. 3b.

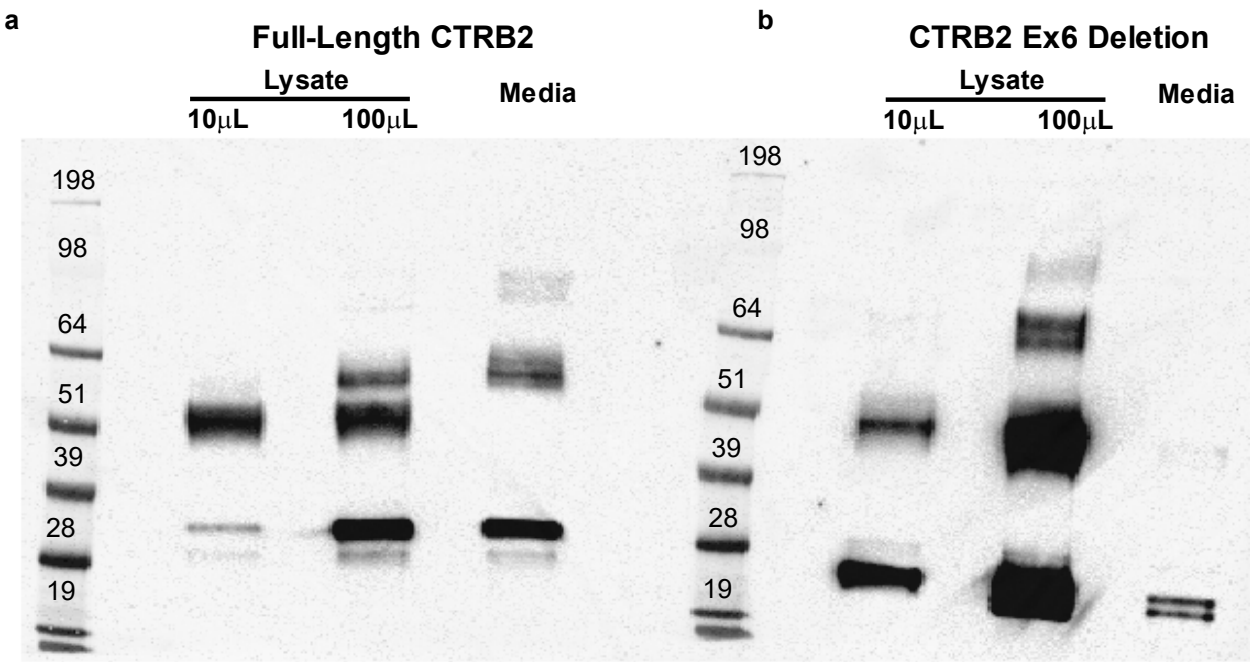

Fig. S8: Full western blot images related to Figure 3c for FLAG-tagged CTRB2 expression constructs. Western blot images corresponding to Fig. 3c-d showing: (**a)** full-length CTRB2 protein (~28 kDa), and (**b)** truncated CTRB2 protein (with exon-6 deletion, ~18 kDa). Immunoprecipitation (IP) for intracellular proteins was performed using either 10μL (first lane) or 100μL (second lane) of lysate; IP for secreted proteins was performed using conditioned media after concentration. Protein secretion was assessed by densitometric quantification of bands on the gel followed by comparing the amount of protein in the cell lysate (first or second lane, band of ~26 kDa) to the media (third lane, bands of ~16 kDa). The trend of less secretion for the truncated as compared to the full length CTRB2 proteins was similar for both lysate-media comparisons. Bands at ~26 kDa in (**a)**, and ~16 kDa in (**b)**, correspond to the full-length and truncated CTRB2 proteins with the 2 kDa signal peptide removed, respectively. Higher bands correspond to the heavy chain of IgG (~50 kDa) from the IP pull-down.

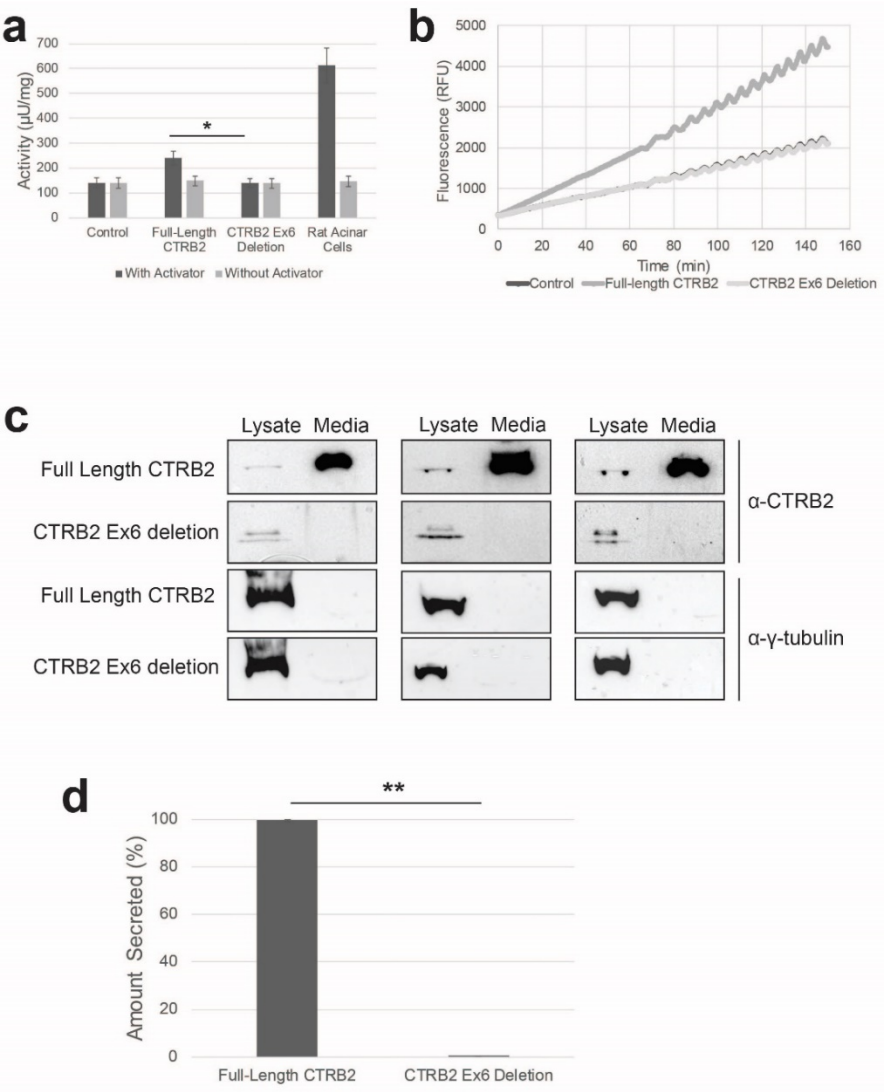

Fig S9. Activity and secretion for untagged CTRB2 expression constructs. (a) Chymotrypsin activity (µg/mg) calculated for the untagged full-length and Ex6 deletion constructs (n=4) and AR42J Rat acinar cells (n=3) (* denotes P <0.05) with (dark grey) and without (light grey) chymotrypsin activator (i.e. trypsin), error bars represent standard error of the mean (SEM). **(b)** Representative chymotrypsin activity time-course measured by fluorescence (RFU) for untagged full-length (dark grey), Ex6 deletion (light grey), and control (darkest grey). The activity of the full length CTRB2 protein is 1.75-fold that of the truncated protein (P=0.023), and the latter is at background levels. (**c**) Amount of CTRB2 protein in cell lysates and media assessed by western blot analysis in HEK293T cells transfected with plasmids expressing full-length and Ex6 deletion CTRB2 (three separate experiments). (**d**) Western blot images in (**c**) were taken using a ChemiDoc Touch Imaging System and bands quantified using ImageLab software. The amount of secreted proteins is summarized from triplicate experiments (** denotes P <0.001, error bars represent SEM). Almost all of the full length CTRB2 protein is detected in the media but almost none of the truncated protein (99.4% vs 0.1%, P=2.7x10^{-6}).

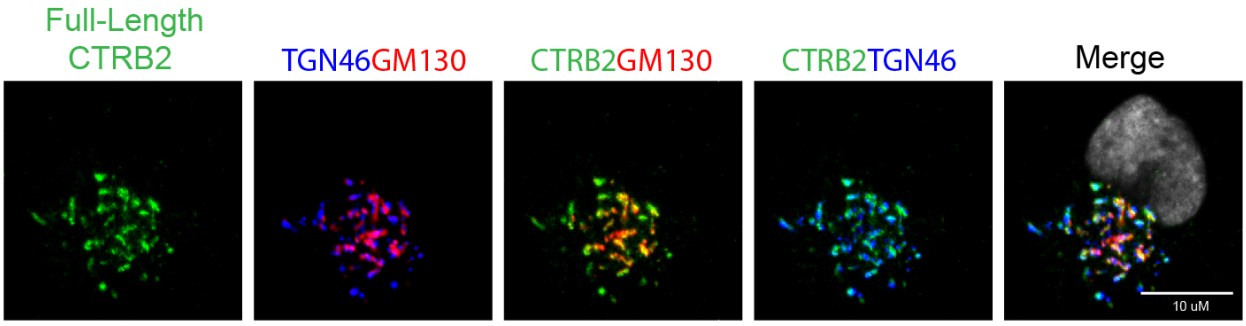

Fig. S10: The full length CTRB2 protein localizes to the Golgi. Normal pancreas-derived HPDE cells transiently expressing GFP-tagged full-length CTRB2 protein (green) immunostained with Golgi markers TGN46 (blue) and GM130 (red) show colocalization of the full length CTRB2 protein with the Golgi. Representative single xy images are shown taken using a spinning disk confocal microscope using and a 63x oil objective and a sCMOS camera. A z-stack was captured and processed with deconvolution software (Slidebook). DAPI stained nuclei shown in grey, scale bar = 10 μM.

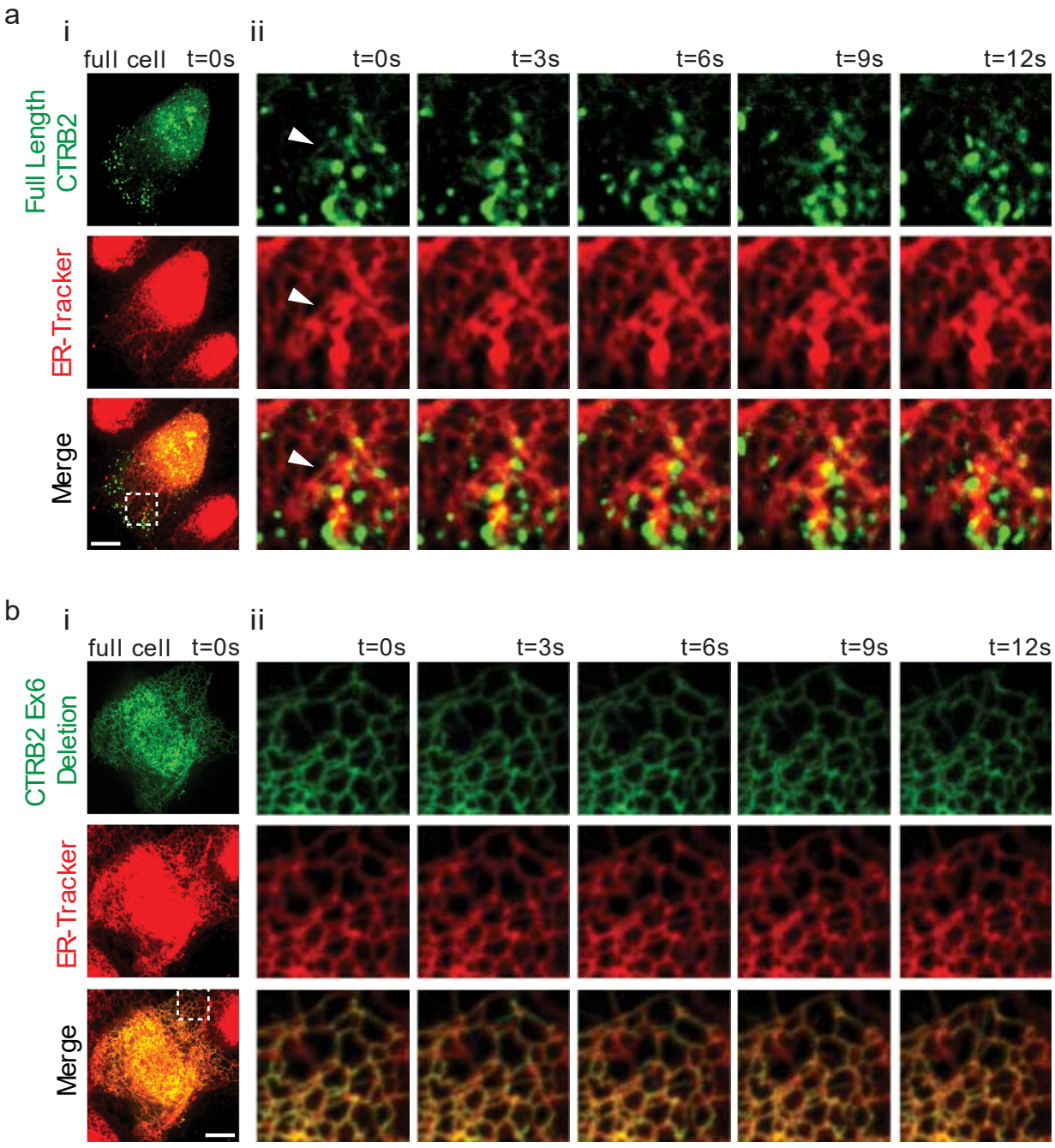

Fig. S11: Live imaging of CTRB2 protein localization in HPDE cells. HPDE pancreatic cells transiently expressing (**a**) GFP-tagged (green) full-length CTRB2 protein, and (**b**) truncated CTRB2 protein (with exon 6 deleted), for 48 hrs were incubated with ER-TrackerTM (red) for 15 min and followed by live spinning disk confocal microscopy with a 63x oil objective every 3 seconds at 37°C in 5% CO_2. (ii) shows an enlarged region (white square) from (i) for each panel. Scale bar = 10um. The full-length CTRB2 protein shows vesicular localization and only little ER overlap (white arrows), whereas the truncated CTRB2 protein completely colocalizes with the ER stain.

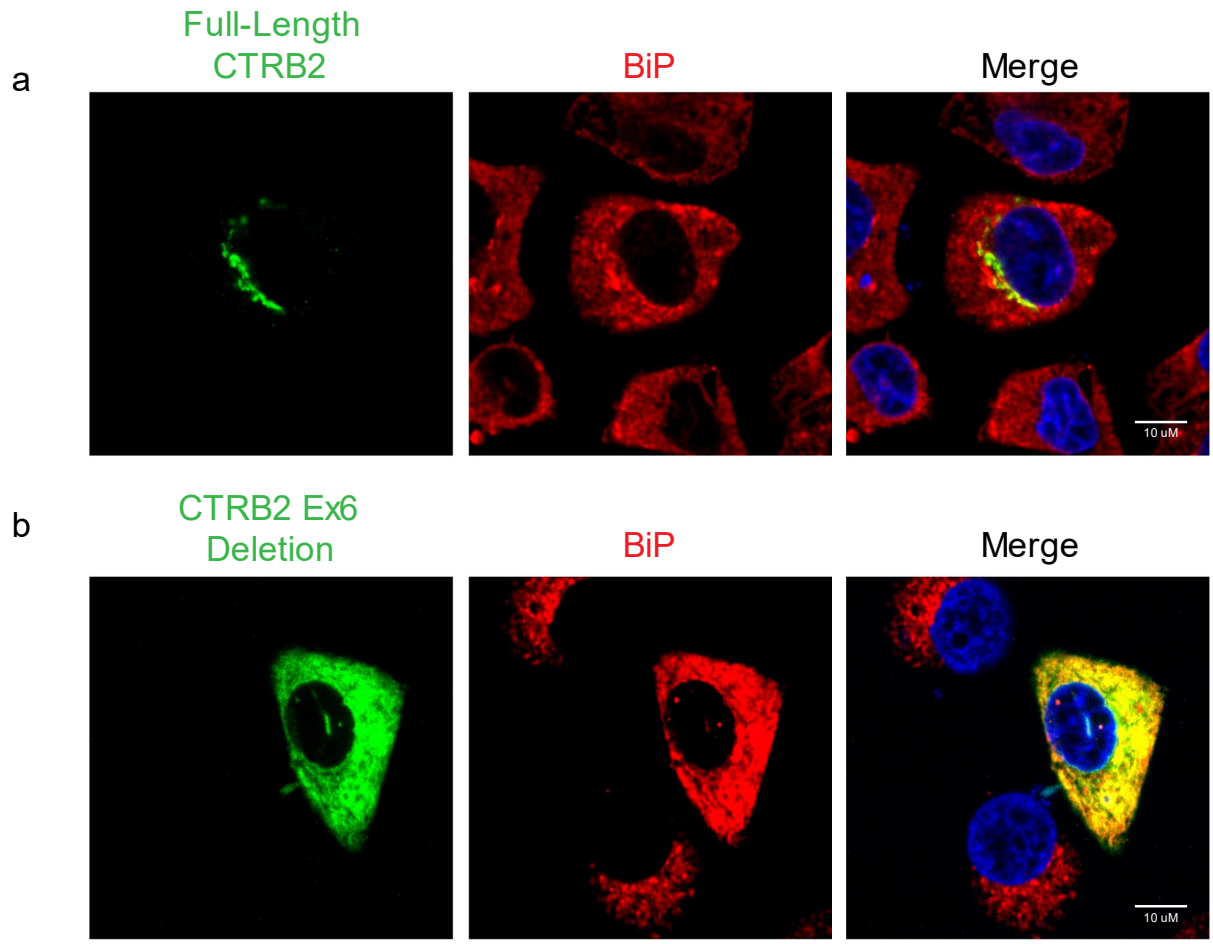

Fig. S12: The truncated CTRB2 protein induces Endoplasmic Reticulum (ER) stress. PANC-1 pancreatic cancer cells transiently expressing GFP-tagged (green) (**a**) full length CTRB2, and (**b**) truncated CTRB2 protein (exon 6 deleted), for 48 hrs were paraformaldehyde fixed and immuno-stained with a marker for ER-stress, BiP (red). Elevated levels of BiP (HSPA5, red) were seen in cells expressing the truncated CTRB2 protein (green in **b**), but not in cells expressing full length CTRB2 (green in **a**). Representative single xy images are shown taken using a spinning disk confocal microscope using a 63x oil objective and a sCMOS camera. A z-stack was captured and processed with Slidebook deconvolution software. Scale bar = 10 μM.

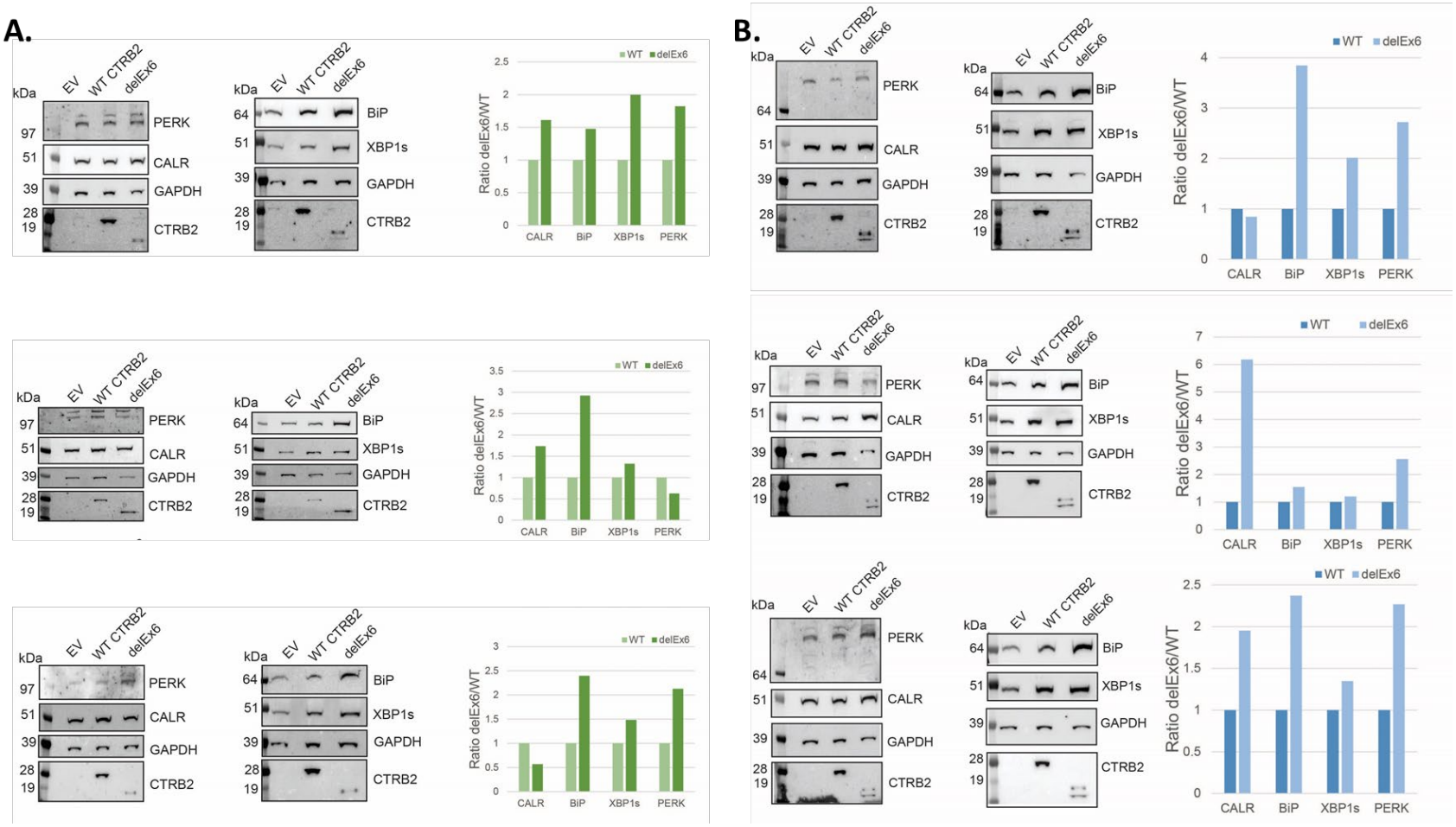

Figure S13: Overexpression of truncated (exon 6 deleted) CTRB2 leads to an increase in ER stress proteins. HEK293T (A) and PANC-1 (B) cells transfected with equal DNA amounts of empty vector, full length CTRB2, and truncated (exon 6 deleted) CTRB2 constructs for 72 hrs. reveal an increase in ER stress proteins in the delEx6 transfections as compared to full length CTRB2. Three biological replicates are shown.

Table S2: Minor allele frequency (MAF) for 16q23.1 GWAS SNPs, *CTRB2* Exon 6 insertion/deletion, and *CTRB2*/*CTRB1* inversion variants found in Europeans via PCR genotyping.

Genotype	EUR	CEU	FIN	GBR	IBS	TSI
rs72802365	0.084	0.089	0.057	0.104	0.130	0.047
Deletion	0.082	0.089	0.068	0.071	0.125	0.061
rs72802342	0.081	0.000	0.068	0.088	0.120	0.047
Inversion*	0.158	0.091	0.104	0.264	0.140	0.160

* The inversion is the minor (derived) allele

Table S3: Primers used for genotyping the *CTRB2* Exon 6 genomic insertion/deletion and the *CTRB2*/*CTRB1* genomic inversion variants.

Structural Variant Tested	Forward Primer (5'-3')	Reverse Primer (5'-3')	Annealing Temperature	Products
Deletion	TGACCTGGTGGAGTCTAGGG	TCAGCATTCTGACCGTGAAC	56.6°C	Reference: 1112bp product/Deletion: 528bp product
	TGACCTGGTGGAGTCTAGGG	TGGCGTCTCCTCCTGCATG	62°C	Reference: 409bp product/Deletion: no product
Inversion	CCTGCTCACTCTACCAAACC	TGGAATTTTGCAACAAGGCG	56°C	Reference (inversion): no product/ Ancestral (noninversion): 1974bp product
	GGGAACGTTTGAGCCTAGAG	TGGAATTTTGCAACAAGGCG	56°C	Reference (inversion): 1728bp/ Ancestral (noninversion): no product

Table S4: Imputation accuracy for two structural variants in the *CTRB2* and *CTRB1* gene region on chr16q23.1.

Genotype (584 bp indel in *CTRB2*)	No. of Individuals		% Accuracy
	Imputed	Genotyped	
Insertion/Insertion	72	70	97%
Insertion/Deletion	15	19	73%
Deletion/Deletion	9	7	78%
Genotype (*CTRB1*/*CTRB2* inversion variant)	**No. of Individuals**		**% Accuracy**
	Imputed	**Genotyped**	
Noninversion/Noninversion	55	52	95%
Noninversion/Inversion	26	29	88%
Inversion/Inversion	14	14	100%

A total of 96 DNA samples from PanScan III were genotyped by PCR for the *CTRB2* exon 6 deletion and *CTRB1/CTRB2* exon 1 inversion (see Methods).

Table S5: Association results for the 15 most likely functional variants at chr16q23.1 (LR < 1:1,000). A Meta-analysis of PanScan I, II, III and PanC4 GWAS data after imputation with the 1000G European reference panel was performed.

SNP	Location (bp)[^]	*P*-value	Likelihood Ratio (LR)*	Posterior Inclusion Probability (PIP)[%]	Odds Ratio OR (95% CI)	r^2 to best GWAS SNP[+]	r^2 to 584 bp *CTRB2* exon 6 deletion variant	EUR MAF[#]	Conditional Analysis	
									P-value conditioned on rs72802365	*P*-value conditioned on deletion
rs72802365	75,246,035	2.51E-17	1.0	0.13	1.36 (1.31-1.40)	1.00	0.69	0.083	1.000	0.007
rs72802352	75,240,883	3.72E-17	1.5	0.15	1.35 (1.31-1.40)	0.99	0.68	0.082	0.589	0.011
rs9936550	75,242,850	4.77E-17	1.9	0.06	1.35 (1.30-1.40)	1.00	0.69	0.082	0.974	0.013
rs111852127	75,249,170	5.11E-17	2.0	0.09	1.35 (1.30-1.40)	1.00	0.69	0.083	0.375	0.011
rs147630228	75,251,551	5.16E-17	2.0	0.09	1.35 (1.30-1.41)	1.00	0.69	0.083	0.628	0.011
rs184458383	75,249,552	5.33E-17	2.1	0.09	1.35 (1.30-1.40)	1.00	0.69	0.083	0.350	0.012
rs111869668	75,249,791	5.33E-17	2.1	0.09	1.35 (1.30-1.41)	1.00	0.69	0.083	0.350	0.012
rs72802395	75,286,484	7.30E-17	2.9	0.09	1.35 (1.30-1.40)	0.99	0.68	0.083	0.552	0.008
rs72802357	75,243,142	7.62E-17	3.0	0.18	1.35 (1.30-1.40)	1.00	0.69	0.082	0.619	0.018
rs72802342	75,234,872	7.13E-16	27.1	NA	1.34 (1.29-1.39)	0.84	0.81	0.081	0.615	0.105
rs371183658	75,234,273	7.78E-16	29.6	NA	1.33 (1.30-1.39)	0.82	0.82	0.081	0.602	0.107
rs55993634	75,236,763	1.17E-15	44.2	NA	1.32 (1.28-1.37)	0.78	0.79	0.089	0.618	0.123
rs72802391	75,273,829	4.46E-15	164.9	NA	1.35 (1.30-1.39)	0.85	0.57	0.075	0.923	0.073
rs8056814	75,252,327	7.92E-15	290.3	NA	1.31 (1.27-1.36)	0.94	0.64	0.088	0.068	0.132
rs68181471	75,252,820	1.14E-14	415.8	NA	1.31 (1.26-1.35)	0.93	0.63	0.091	0.192	0.121

[^] in human GRCh37/hg19 build coordinates

* Likelihood Ratio (LR) is given as 1:LR to the most significant SNP, rs72802365

[%] Posterior Inclusion Probability (PIP) for each SNP using SuSiE's Iterative Bayesian Stepwise Selection. A single credible set was observed at this locus. NA indicates SNPs not included in the credible set

[+] Linkage disequilibrium (r^2) to best GWAS SNP rs72802365

[#] Minor Allele Frequency in the 1000G EUR population

Table S6: Colocalization of expression QTLs (eQTLs) and pancreatic cancer GWAS signals.

Dataset	Tissue	Gene	Probability of Different Causal Variants	Probability of Same Causal Variant
GTEx	Skin - Sun Exposed	*TMEM170A*	1.0	<0.01
GTEx	Skin - Not Sun Exposed	*TMEM170A*	1.0	<0.01
LTG	Pancreas	*TMEM170A*	0.92	0.08
GTEx	Adipose -Visceral	*CFDP1*	1.0	<0.01
GTEx	Adipose - Subcutaneous	*CFDP1*	1.0	<0.01
GTEx	Skin - Sun Exposed	*CFDP1*	1.0	<0.01
GTEx	Whole Blood	*BCAR1*	0.69	0.31
GTEx	Pancreas	*CHST6*	0.87	0.13
TCGA	Pancreas	*CHST6*	0.90	0.08
GTEx	Pancreas	*CTRB1*	0.99	<0.01
GTEx	Pancreas	*CTRB2*	1.0	<0.01

Colocalization was performed using the coloc package in R.

Table S7: Colocalization of *CTRB2* splicing QTLs (sQTLs) and pancreatic cancer GWAS signals.

Dataset	*CTRB2* Splicing	Probability Different Causal Variants	Probability Same Causal Variant
LTG	exon 5-6	0.11	0.89
LTG	exon 6-7	0.54	0.46
LTG	exon 5-7	0.13	0.87
GTEx	exon 5-6	0.09	0.91
GTEx	exon 6-7	<0.01	0.94
GTEx	exon 5-7	0.13	0.87

Colocalization was performed using the coloc package in R.

Table S10: Differential expression of Endoplasmic Reticulum (ER) stress markers in pancreatic tissues from individuals with an increasing number of copies of the *CTRB2* Exon 6 deletion allele.

	GTEx		LTG	
Gene	**logβ**	**P-Value**	**logβ**	**P-Value**
HSPA5 (BiP)	0.335	0.022	-0.103	0.421
EIF2AK3 (PERK)	0.231	0.049	-0.183	0.264
ERN1 (IRE1-alpha)	0.230	0.066	-0.391	0.074
P4HB (PDI)	0.167	0.084	-0.246	0.300
CANX (Calnexin)	0.098	0.211	-0.151	0.118
DDIT3 (CHOP)	0.189	0.218	-0.065	0.681
ERO1A (ERO1-L-alpha)	0.087	0.255	-0.156	0.171
CALR	0.094	0.448	-0.140	0.159
ATF6	0.083	0.537	-0.355	0.049
XBP1	0.042	0.711	-0.375	0.108

Table S11: Colocalization of *CTRB2* splicing QTLs and diabetes GWAS signals.

Diabetes Dataset	sQTL Dataset	*CTRB2* Splicing	Probability Different Causal Variant	Probability Same Causal Variant
T1D - Barrett	GTEx	exon 5-6	<0.01	1.0
T1D - Barrett	GTEx	exon 6-7	<0.01	1.0
T1D - Barrett	GTEx	exon 5-7	<0.01	1.0
T1D - Barrett	LTG	exon 5-6	<0.01	1.0
T1D - Barrett	LTG	exon 6-7	0.36	0.64
T1D - Barrett	LTG	exon 5-7	<0.01	1.0
T2D	GTEx	exon 5-6	0.05	0.96
T2D	GTEx	exon 6-7	0.03	0.98
T2D	GTEx	exon 5-7	0.08	0.92
T2D	LTG	exon 5-6	0.34	0.66
T2D	LTG	exon 6-7	0.51	0.49
T2D	LTG	exon 5-7	0.50	0.50

Colocalization was performed using the coloc package in R.

Supplemental References

1. Robinson, James T.; Thorvaldsdottir, Helga; Winckler, Wendy; Guttman, Mitchell; Lander, Eric S.; Getz, Gad; Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Funding

The work conducted at NCI was supported by the Intramural Research Program (IRP) of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health (NIH).

The Melbourne Collaborative Cohort Study cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Long%20List.pdf"

Cancer incidence data for CLUE were provided by the Maryland Cancer Registry, Center for Cancer Surveillance and Control, Department of Health and Mental Hygiene, 201 W. Preston Street, Room 400, Baltimore, MD 21201, http://phpa.dhmh.maryland.gov/cancer, 410-767-

4055. We acknowledge the State of Maryland, the Maryland Cigarette Restitution Fund, and the National Program of Cancer Registries of the Centers for Disease Control and Prevention for the funds that support the collection and availability of the cancer registry data." We thank all the CLUE participants.

The NYU study was funded by NIH R01 CA098661, UM1 CA182934 and center grants P30 CA016087 and P30 ES000260.

The Physicians' Health Study was supported by research grants CA-097193, CA-34944, CA40360, HL-26490, and HL-34595 from the National Institutes of Health, Bethesda, MD USA.

The Women's Health Study was supported by research grants CA-047988, HL-043851, HL080467, and HL-099355 from the National Institutes of Health, Bethesda, MD USA.

Health Professionals Follow-up Study is supported by NIH grant UM1 CA167552 from the National Cancer Institute, Bethesda, MD USA.

Nurses' Health Study is supported by NIH grants UM1 CA186107, and R01 CA49449 from the National Cancer Institute, Bethesda, MD USA.

The PANKRAS II Study in Spain was supported by research grants from Instituto de Salud Carlos III-FEDER, Spain: Fondo de Investigaciones Sanitarias (FIS) ((#PI95/0017, #PI12/00815, #PI13/00082 and #PI15/01573), Red Temática de Investigación Cooperativa en Cáncer (#RD12/0036/0050), and CIBER de Epidemiología (CIBERESP); Ministerio de Ciencia y Tecnología (CICYT SAF 2000-0097); Generalitat de Catalunya (CIRIT - SGR), Spain.

The IARC/Central Europe study was supported by a grant from the US National Cancer Institute at the National Institutes of Health (R03 CA123546-02) and grants from the Ministry of Health of the Czech Republic (NR 9029-4/2006, NR9422-3, NR9998-3, MH CZDRO-MMCI 00209805).

The National Familial Pancreas Tumor Registry at Johns Hopkins University was supported by the NCI Grants P50CA062924 and R01CA97075. Additional support was provided by the Lustgarten Foundation, Susan Wojcicki and Dennis Troper and the Sol Goldman Pancreas Cancer Research Center. The PANC4 GWAS was supported by RO1 CA154823 and federal funds from the National Cancer Institute (NCI), US National Institutes of Health (NIH) under contract number HHSN261200800001E.

The Mayo Clinic Biospecimen Resource for Pancreas Research study is supported by the Mayo Clinic SPORE in Pancreatic Cancer (P50 CA102701).

Funding at Memorial Sloan Kettering was supported by the National Cancer Institute of the National Institutes of Health grant number P30 CA008748.

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA37429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson) for SELECT.

The PACIFIC Study was supported by RO1CA102765, Kaiser Permanente and Group Health Cooperative.

The Queensland Pancreatic Cancer Study was supported by a grant from the National Health and Medical Research Council of Australia (NHMRC) (Grant number 442302). RE Neale is supported by a NHMRC Senior Research Fellowship (#1060183).

The UCSF pancreas study was supported by NIH-NCI grants (R01CA1009767, R01CA109767-S1 and R0CA059706) and the Joan Rombauer Pancreatic Cancer Fund. Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program; the NCI's SEER Program under contract HSN261201000140C awarded to CPIC; and the CDC's National Program of Cancer Registries, under agreement #U58DP003862-01 awarded to the California Department of Public Health.

The Yale (CT) pancreas cancer study is supported by National Cancer Institute at the U.S. NIH, grant 5R01CA098870. The cooperation of 30 Connecticut hospitals, including Stamford Hospital, in allowing patient access, is gratefully acknowledged. The Connecticut Pancreas Cancer Study was approved by the State of Connecticut Department of Public Health Human Investigation Committee. Certain data used in that study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. The authors assume full responsibility for analyses and interpretation of these data.

The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.