Supporting Information

Fløyel et al. 10.1073/pnas.1402571111

SI Materials and Methods

Microarray Analysis. For the Type 1 Diabetes (T1D) Genetics Consortium (T1DGC) B-lymphoblastoid cell lines (BLCLs), transcriptome-wide gene expression (47,392 transcripts) was evaluated using Illumina HumanHT-12 v4 Expression BeadChip Arrays (Illumina). Cells were resuspended in RNA*Later* (Qiagen), and RNA was extracted using TRIzol (Invitrogen). RNA quality was examined using a 2100 Bioanalyzer (Agilent). Samples with RNA integrity number ≥ 8 were used to prepare biotin-labeled cRNA using the Illumina TotalPrep RNA Amplification Kit (Ambion) as described previously (1). The samples were hybridized onto the BeadChips, which were scanned by Beadarray Reader (Illumina). Beadarray summary data produced in Illumina BeadStudio were analyzed in R using the beadarray package (2). For each BeadChip, the summarized expression level of each probe was log2-transformed and then normalized using the quantile method. The mean expression value for the duplicates was taken for subsequent analysis. The diagnostic plots confirmed the data integrity and quality control. Data from the two probes, ILMN_2390853 and ILMN_1752451, that interrogate the cathepsin H (*CTSH*) expression were then extracted.

For the human pancreatic islets, transcriptome-wide gene expression was analyzed using Affymetrix Human Genome U133 Plus 2.0 Arrays (Affymetrix). RNA was extracted using TRIzol (Invitrogen), and RNA quality was examined using a 2100 Bioanalyzer (Agilent). Samples with RNA integrity number ≥ 8 were used. Labeling was done using the One-Cycle Target Labeling Kit (Affymetrix), and the arrays were scanned by a GeneChip Scanner 3000 (Affymetrix). Expression levels were normalized using the robust multiarray analysis method, and probes were annotated using an updated probe set definition (3). Data from the *CTSH* probe, 202295_s_at, were then extracted.

Real-Time Quantitative PCR. For HapMap BLCLs, RNA was extracted using TRIzol (Invitrogen), cDNA was prepared using TaqMan Reverse Transcription reagents (Applied Biosystems), and real-time quantitative PCR (qPCR) was performed using TaqMan Assays and an ABI 7900HT System (Applied Biosystems). The expression level of *CTSH* was evaluated using the $2^{-\Delta\Delta Ct}$ method (4) normalized to the geometric mean of the three housekeeping genes *GAPDH*, *ACTB*, and *UBC*.

For the pancreatic tissue from T1D patients, RNA was extracted using the RNAqueous-Micro Kit (Ambion), cDNA was prepared using SuperScript II Reverse Transcriptase with random primers (Invitrogen), and real-time qPCR was performed as for the HapMap BLCLs but run on a LightCycler 480 (Roche). The *CTSH* expression was evaluated using the $2^{-\Delta\Delta Ct}$ method normalized against *ACTB*.

For purified primary rat β-cells, RNA extraction and reverse transcription were performed as described previously (5). Expression of *Ctsh* was evaluated by real-time qPCR using SYBR Green (5) and comparison with a standard curve (6). Expression values of *Ctsh* were normalized to *Gapdh*, and the values obtained were divided by the highest value of each experiment considered as one. The primer sequences used are *Gapdh* forward primer: 5′-AGTTCAACGGCACAGTCAAG-3′ and reverse primer: 5′-TACTCAGCACCAGCATCACC-3′; *Ctsh* forward primer: 5′-AGATGAGAATGCCGCCCTGGA-3′ and reverse primer: 5′-GCGGGCTTGTTGCTGGTCAAGT-3′.

For WT and $Ctsh^{-/-}$ mice, the excised pancreatic tissues were perfused with RNA*Later* (Qiagen) (7), placed in cryovials, snap-frozen in dry ice and ethanol, and stored at −80 °C until RNA extraction. Tissue homogenization was performed in Qiazol (Qiagen) with 1.4-mm ceramic beads (Mo Bio Laboratories) using the FastPrep FP120 Cell Disruptor (Bio-101, Thermo Savant; Qbiogene). RNA was extracted using the miRNeasy Mini Kit (Qiagen), cDNA was prepared using the iScript cDNA Synthesis Kit (Bio-Rad), and real-time qPCR was performed using TaqMan Assays (Applied Biosystems) and run on a CFX384 system (Bio-Rad).

For INS-1 cells, RNA extraction, cDNA synthesis, and qPCR were performed as described for the $Ctsh^{-/-}$ mice. The mRNA expressions were evaluated using the $2^{-\Delta\Delta Ct}$ method normalizing against *Hprt1* or *Actb*.

Immunoblotting. Immunoblotting was performed as previously described (8). The following antibodies were used: anti-CTSH (PAB8634; Abnova), anti-IκBα (sc-371; Santa Cruz), antiinducible nitric oxide synthase (610332; BD Bioscience), antiphospho-JNK (Thr183/Tyr185; 9251; Cell Signaling), antiphospho–c-Jun/JunD (Ser73/Ser100; 9164; Cell Signaling), antiphospho-signal transducer and activator of transcription 1 (Tyr701; 9171; Cell Signaling), antiphospho-p38 (Thr180/Tyr182; 9211; Cell Signaling), anti-p38 (9212; Cell Signaling), antiphospho-Erk1/2 (Thr202/Tyr204; 9101; Cell Signaling), anti-Erk1/2 (9102; Cell Signaling), anti–β-actin mouse (ab6276; Abcam), anti–β-actin rabbit (4967; Cell Signaling), anti–β-tubulin (5346; Cell Signaling), anti-goat IgG (sc-2354; Santa Cruz), anti-rabbit IgG (7074; Cell Signaling), and anti-mouse IgG (7076; Cell Signaling). In HapMap BLCLs, the CTSH/β-actin ratios were calculated and correlated to a reference sample loaded on each gel. The signaling proteins were correlated to β-actin and/or β-tubulin.

Immunostaining. Immunoperoxidase and double immunofluorescence staining of human pancreas ($n = 4$) were performed on 5-μm sections of paraffin-embedded, formaldehyde-fixed tissue according to standard protocols using Antigen Unmasking Solution pH 6 (Vector Labs) and Animal-Free Blocker (Vector Labs). For immunoperoxidase staining, the antibodies were anti-CTSH (1:200; HPA003524; Sigma) and anti-rabbit IgG (1:200; P0448; DAKO). Visualization involved the Vector VIP Peroxidase Substrate Kit (Vector Labs), Mayer's hematoxylin (Ampliqon), and Pertex Mounting Media (Histolab). For immunofluorescence staining, the antibodies were anti-CTSH (1:200; HPA003524; Sigma), antiinsulin (1:2,000; ABM-816; Nordic Biosite), antiglucagon (1:2,000; ab10988; Abcam), anti-rabbit IgG (1:200; FI-1000; Vector Labs), and anti-mouse IgG (1:200; TI-2000; Vector Labs). Sections were mounted in Prolong antifade with DAPI (Molecular Probes; Invitrogen).

Double immunofluorescence staining of pancreas from three WT and two $Ctsh^{-/-}$ mice was performed on 4-μm sections of formalin-fixed, paraffin-embedded tissue according to standard protocol using Animal-Free Blocker (Vector Labs). For each WT mouse, we analyzed 4 sections from three different levels separated by 100 μm (12 sections per animal), and for each $Ctsh^{-/-}$ mouse, we analyzed 32 sections per animal ∼40 μm apart. Antibodies were antiinsulin (1:4,000; A0564; DAKO), antiglucagon (1:2,000; ab10988; Abcam), anti-guinea pig IgG (1:200; A11073; Invitrogen), and anti-mouse IgG (1:200; A11032; Invitrogen). The intensity of the insulin staining was evaluated in 320 islets from WT mice and 328 islets from $Ctsh^{-/-}$ mice by two observers, and one of the observers was unaware of the sample identity. The islets were ranked from one to seven, where one and seven indicate islets with the highest and lowest

insulin intensities, respectively. Sections were mounted in Prolong antifade with DAPI (Molecular Probes; Invitrogen). The staining was examined in a light or fluorescent microscope using the cellSens Imaging Software (Olympus).

Apoptosis. Apoptosis was determined using the Cell Death Detection ELISA (Roche) as previously described (8). The values were normalized against the DNA content, which was measured using the QuantiFluor dsDNA System (Promega) after sonication of the lysate for 15 s. Caspase-3/7 activity was determined using the Caspase-Glo 3/7 assay (Promega) according to the protocol by the manufacturer. Absorbance, fluorescence and luminescence were measured using an Infinite M200 Plate Reader (Tecan).

Insulin Measurements. Insulin was measured using the Mouse Insulin ELISA Kit (Mercodia) or the High-Range Rat Insulin ELISA Kit according to the manufacturer's protocols. For accumulated insulin release, cells/islets (10–15 islets per condition in duplicate) were incubated in fresh media for 24 h. The accumulated insulin was normalized against the DNA content as described above. For glucose-stimulated insulin secretion, INS-1 cells ($n = 4$) were incubated in RPMI1640 without FBS for 2 h, washed in Krebs Ringer buffer (KRB) without glucose, incubated in KRB without glucose for 30 min, washed again, incubated in KRB with 2 mM glucose for 30 min, and incubated in KRB with 20 mM glucose for 30 min. Finally, cells were lysed in 0.2 M HCl in ethanol for 20 min on ice.

Expression Quantitative Trait Locus Analysis. We used the GGTools package and hmceuB36 data (Bioconductor; http://www.bioconductor.org/) in R (http://www.r-project.org/) to analyze SNPs for *cis*-expression quantitative trait locus (*cis*-eQTL) signals (9, 10) within the associated linkage disequilibrium (LD) region surrounding *CTSH* (11). We used the rtracklayer package in R to add the eQTL results to the UCSC Genome Browser (http://genome.ucsc.edu/) as custom tracks to illustrate the pattern of eQTL signals across the LD region. The eQTL signals were confirmed in other datasets (12–14).

HapMap qPCR data were analyzed for eQTLs using linear regression with ΔCt of *CTSH* as the dependent variable and SNP genotype for the top 10 SNPs from the GGTools analysis as the explanatory variables. Individual eQTL effects were tested by forward stepwise regression (additive model with genotypes coded zero, one, or two for the number of minor alleles). To investigate whether other SNPs within the associated LD region had eQTL effects independent of rs3825932, we used Tagger (15) (http://www.broadinstitute.org/mpg/tagger/) in Haploview (16) (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview) to select SNPs that tagged the full genetic variation ($r^2 > 0.7$). Each of 34 tag SNPs was analyzed individually in a linear regression model for eQTL in *CTSH*. Any SNP that was significantly associated with *CTSH* expression was included in a regression model using forward stepwise regression to examine the SNPs for eQTL effects independently of rs3825932.

For the T1DGC BLCLs, *CTSH* expression level was correlated to genotypes in an eQTL analysis using PLINK (version 1.07) (http://pngu.mgh.harvard.edu/~purcell/plink/) (17). The association was evaluated using linear regression with *CTSH* expression (log2-transformed) for each of two probes separately as a quantitative variable depending on rs3825932 genotypes coded as an additive variable. Sex and affection status were included as covariates in the model but were not associated with *CTSH* expression.

Genotyping. Genotyping of rs3825932 in the T1DGC BLCLs and the pancreatic tissue samples derived from donors with T1D was performed using TaqMan SNP Genotyping Assays (Applied Biosystems).

For the two study populations, genotyping of rs3825932 was done using an in-house KASPar SNP Genotyping System at KBioscience. Success rate for genotyping of rs3825932 was 96.2%, with an error rate of 0.0% (estimated from 92 duplicate samples). The genotype obeyed the Hardy–Weinberg equilibrium in the analyzed samples ($P > 0.05$). In the Hvidoere Study Group, genotyping of the rs3825932 variant was successfully carried out in 251 patients. Genotyping of the *HLA DRB1* locus was performed according to the International Histocompatibility Working Group (18).

1. Morahan G, Peeva V, Mehta M, Williams R (2008) Systems genetics can provide new insights in to immune regulation and autoimmunity. *J Autoimmun* 31(3):233–236.
2. Dunning MJ, Smith ML, Ritchie ME, Tavaré S (2007) beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* 23(16):2183–2184.
3. Sandberg R, Larsson O (2007) Improved precision and accuracy for microarrays using updated probe set definitions. *BMC Bioinformatics* 8:48.
4. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402–408.
5. Rasschaert J, et al. (2005) Toll-like receptor 3 and STAT-1 contribute to double-stranded RNA+ interferon-gamma-induced apoptosis in primary pancreatic beta-cells. *J Biol Chem* 280(40):33984–33991.
6. Overbergh L, Valckx D, Waer M, Mathieu C (1999) Quantification of murine cytokine mRNAs using real time quantitative reverse transcriptase PCR. *Cytokine* 11(4): 305–312.
7. Griffin M, Abu-El-Haija M, Abu-El-Haija M, Rokhlina T, Uc A (2012) Simplified and versatile method for isolation of high-quality RNA from pancreas. *Biotechniques* 52(5):332–334.
8. Berchtold LA, et al. (2011) Huntingtin-interacting protein 14 is a type 1 diabetes candidate protein regulating insulin secretion and beta-cell apoptosis. *Proc Natl Acad Sci USA* 108(37):E681–E688.
9. Stranger BE, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315(5813):848–853.
10. Stranger BE, et al. (2007) Population genomics of human gene expression. *Nat Genet* 39(10):1217–1224.
11. Barrett JC, et al.; Type 1 Diabetes Genetics Consortium (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 41(6):703–707.
12. Dimas AS, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325(5945):1246–1250.
13. Moffatt MF, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448(7152):470–473.
14. Dixon AL, et al. (2007) A genome-wide association study of global gene expression. *Nat Genet* 39(10):1202–1207.
15. de Bakker PI, et al. (2005) Efficiency and power in genetic association studies. *Nat Genet* 37(11):1217–1223.
16. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.
17. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
18. van der Zwan A, Griffith B, Rozemuller E, Williams T, Tilanus MGJ (2000) Genomic analysis of the human MHC: DNA based typing for HLA alleles and linked polymorphisms. *IHWG Technical Manual*, ed Tilanus MGJ (International Histocompatibility Working Group, Fred Hutchinson Cancer Research Center, Seattle).

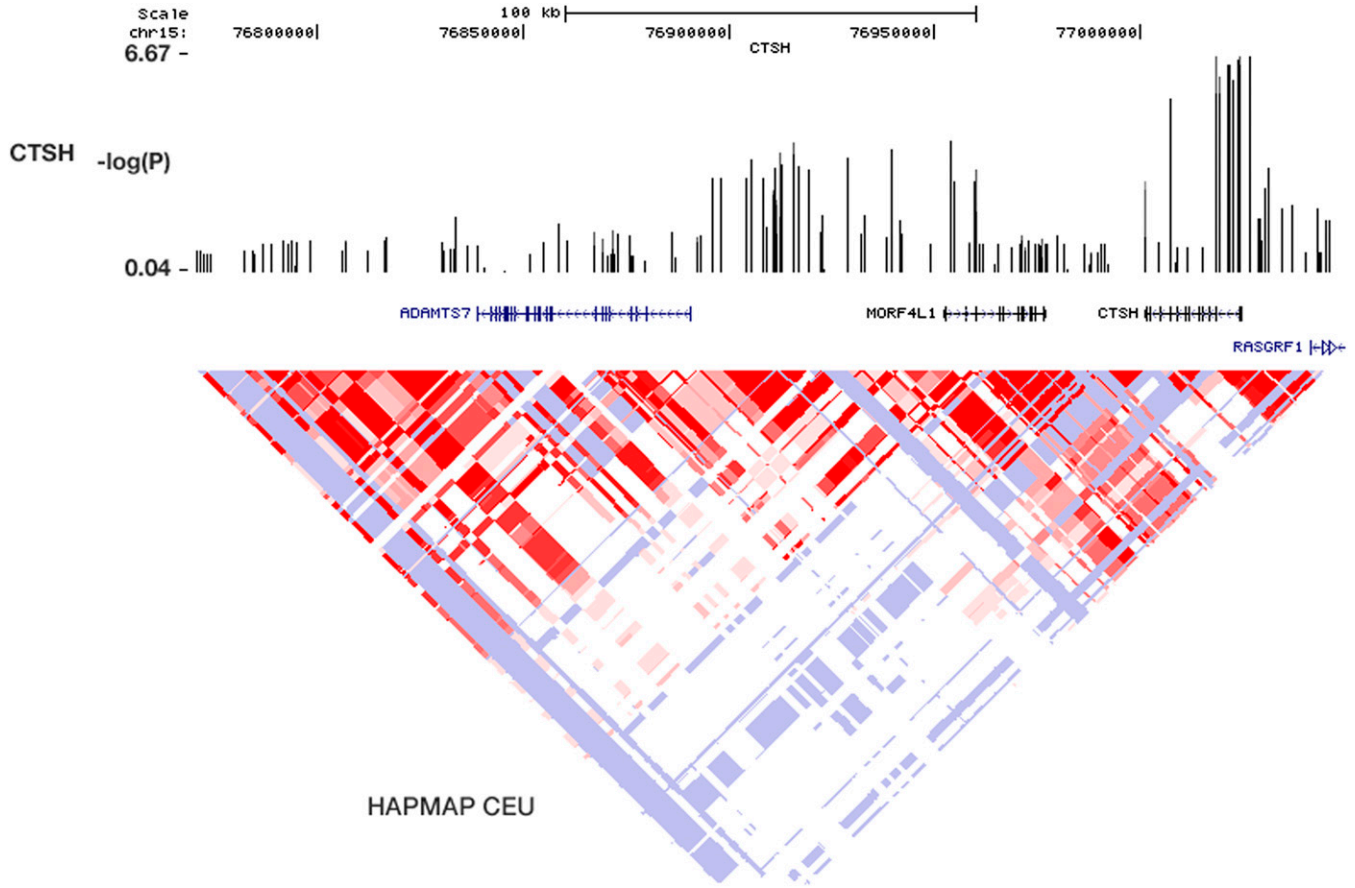

Fig. S1. The eQTL signals and LD region in *CTSH* in the HapMap Utah residents with ancestry from northern and central Europe (CEPH, CEU) population (from the UCSC Genome Browser GRCh37/hg18 with eQTL signals from the GGTools analysis added as custom tracks).

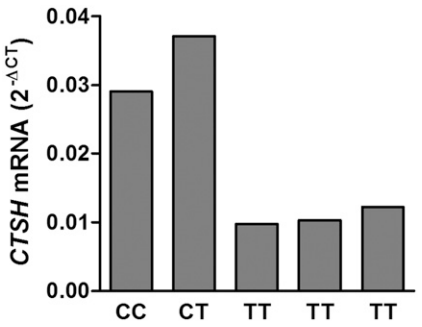

Fig. S2. rs3825932 affects *CTSH* expression in human pancreas tissue. *CTSH* expression was examined in pancreas tissue from five donors with T1D by qPCR and normalized to *ACTB*. Data are presented as relative expression ($2^{-\Delta CT}$) for five individuals with various rs3825932 genotypes.

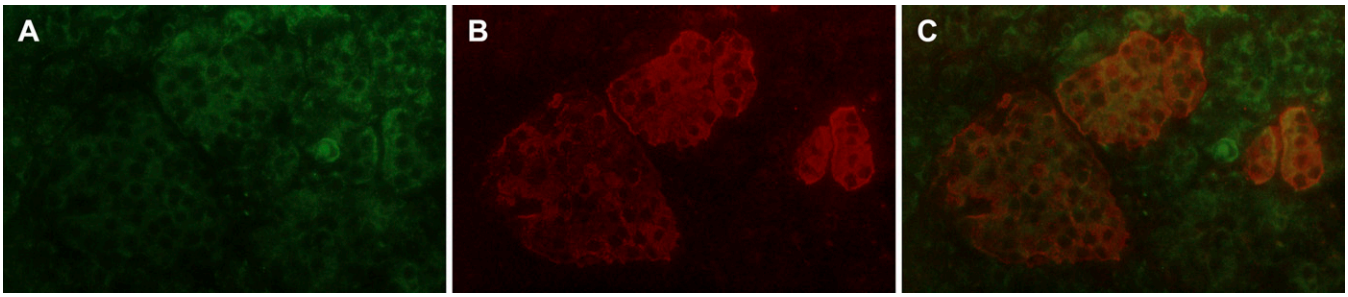

Fig. S3. CTSH is expressed in human β-cells. Section through human nondiabetic pancreas showing a representative example of pancreatic islets and exocrine tissue stained by double immunofluorescence for (*A*) CTSH (green) and (*B*) insulin (red). *C* is the merged image of *A* and *B*. The yellow color indicates co-localization between CTSH and insulin. (Magnification: 63×.)

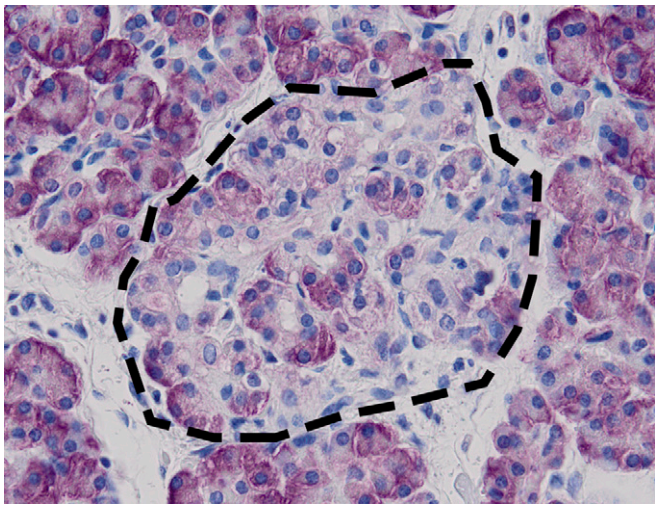

Fig. S4. CTSH in human pancreatic islets and exocrine tissue. Representative immunoperoxidase staining of human nondiabetic pancreas with anti-CTSH antibody (violet). Hematoxylin (blue) marks nuclei. The dotted line marks the periphery of an islet. (Magnification: 40×.)

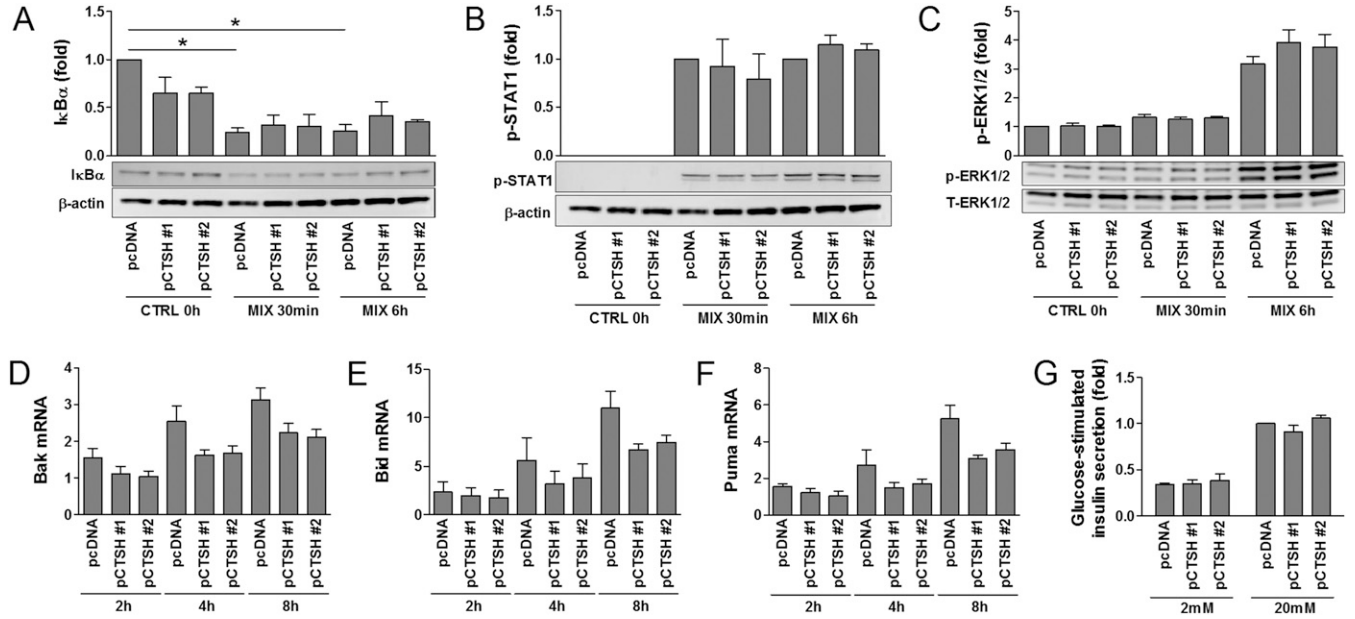

Fig. S5. Effects of CTSH overexpression on cytokine signaling, proapoptotic factors, and stimulated insulin release. INS-1 cells were stably transfected with an empty control vector (pcDNA) or a vector encoding *CTSH* (plasmid encoding *CTSH* #1 (pCTSH #1) and pCTSH #2). (*A–C*) Immunoblotting was performed after stimulation with IL-1β (150 pg/mL) and IFN-γ (5 ng/mL) (MIX) for 0 h, 30 min, or 6 h. Data are means ± SEM (n = 3–4). CTRL, control; p-STAT1, phosphosignal transducer and activator of transcription 1. *$P < 0.05$. (*D–F*) qPCR was performed after stimulation with IL-1β (150 pg/mL) and IFN-γ (5 ng/mL) for 2, 4, or 8 h, and data were normalized to *Actb* and presented as fold induction. Data are means ± SEMs (n = 4). (*G*) Glucose-stimulated insulin secretion in response to 2 and 20 mM glucose. Data are presented as fold vs. pcDNA with mean ± SEM (n = 4).

Table S1. The 10 SNPs with the most significant eQTL signals

SNP	Position	*P* value	r^2
rs3825932	79235446	1.34e-07	1.0
rs1036938 (C→S)	79237247	3.927e-07	0.855
rs1369324	79239719	3.93e-07	0.855
rs11855406	79234864	2.04e-06	0.961
rs11638844	79231518	2.27e-06	0.924
rs1036937	79237180	2.50e-06	0.817*
rs2870085	79234268	3.86e-07	0.962
rs10400902	79231616	5.02e-06	0.925
rs11072817	79236815	6.81e-06	0.851
rs11072818	79236933	6.81e-06	0.851

Shown are SNP rs number, chromosomal position, *P* value for eQTL analysis when using our own *CTSH* expression data on HapMap BLCLs, and LD r^2 values relative to rs3825932. D′ is one for all SNPs except where noted. The position of the SNPs is derived from UCSC Genome Browser GRCh37/hg19. *D′ = 0.96.

Table S2. The insulin intensity distribution in islets from WT and *Ctsh*$^{-/-}$ mice

	1	2	3	4	5	6	7	Total no. of islets
WT	205	29	43	27	13	3	0	320
KO	29	28	94	78	69	21	9	328

In total, 320 islets from WT and 328 islets from *Ctsh*$^{-/-}$ (KO) mice (8–11 wk of age) were evaluated for insulin intensity after immunofluorescence staining of pancreatic tissue sections. Numbers 1–7 indicate the level of intensity; 1 refers to the highest intensity, and 7 refers to the lowest intensity. The difference in insulin intensity between islets from WT and *Ctsh*$^{-/-}$ mice was highly significant ($P < 2.2 \times 10^{-16}$).

Table S3. Anthropometric euglycemic–hyperinsulinemic clamp and i.v. glucose tolerance test measures in healthy subjects

	CC	CT	TT	Effect size (%)	*P* value
BMI (kg/m^2)	24.4 (3.5)	23.9 (2.8)	23.7 (3)	0.97	0.08
Waist-to-hip ratio	0.83 (0.09)	0.84 (0.09)	0.86 (0.05)	0.99	0.28
Total fat (%)	23.2 (7.2)	22.6 (7.4)	18.1 (4.2)	0.99	0.70
VO_{2max} (ml/min per kilogram)	38.8 (7.9)	38.6 (7.2)	43.3 (8.7)	1.00	0.96
M (mg/kg per minute)*	12.3 (3.8)	11.3 (2.7)	11.8 (3.8)	0.96	0.41
$AUC_{ins0–10}$[†]	3,527 (2,390)	2,942 (1,682)	2,778 (1,448)	0.96	0.67
$incAUC_{ins0–10}$[‡]	3,189 (2,323)	2,586 (1,624)	2,431 (1,299)	0.96	0.66
$Peak_{ins}$[§]	510 (322)	441 (266)	436 (222)	0.97	0.71
Ins_{incr}[¶]	23.1 (22)	25 (35.1)	20.5 (18.1)	0.99	0.95
Phi[‖]	2.55E-08 (1.72E-08)	2.18E-08 (1.27E-08)	2.00E-08 (1.05E-08)	0.96	0.66
Di**	2.60E-07 (1.61E-07)	2.04E-07 (9.93E-08)	1.79E-07 (9.20E-08)	0.87	0.15

Data presented as means (SDs). *P* values were calculated using the mathematical model Ln(response variable) = age + sex + VO_{2max} + total fat percentage + rs3825932 genotype. AUC, area under the curve; BMI, body mass index.

*Insulin-stimulated glucose disposal (M) was defined as the glucose infusion rate during the last 30 min of the steady state.

[†]The total insulin AUC was calculated for the initial 10-min period ($AUC_{ins0–10}$).

[‡]The incremental insulin AUC ($incAUC_{ins1–10}$) is a measure of first-phase insulin secretion.

[§]$Peak_{ins}$ is defined as the highest insulin value during the i.v. glucose tolerance test.

[¶]Incremental insulin response (Ins_{incr}) is a measure of first- and second-phase insulin secretions and defined as $(Ins_{30} - Ins_0)/(Glu_{30} - Glu_0)$.

[‖]Insulin response in relation to the glucose concentration (Phi) was calculated as $Phi = AUC_{ins0–10}/AUC_{glu0–10}$.

**Disposition index (Di) is an estimate of the insulin secretion capacity in relation to insulin sensitivity and was calculated as $Di = Phi \times M$.