Loss-of-function mutations in *SLC30A8* protect against risk of type 2 diabetes

Supplementary Information

Jason Flannick, Gudmar Thorleifsson, Nicola L. Beer, Suzanne B. R. Jacobs, Niels Grarup, Noël P. Burtt, Anubha Mahajan, Christian Fuchsberger, Gil Atzmon, Rafn Benediktsson, John Blangero, Don W. Bowden, Ivan Brandslund, Julia Brosnan, Frank Burslem, John Chambers, Yoon Shin Cho, Cramer Christensen, Desirée A. Douglas, Ravindranath Duggirala, Zachary Dymek, Yossi Farjoun, Timothy Fennell, Pierre Fontanillas, Tom Forsén, Stacey Gabriel, Benjamin Glaser, Daniel F. Gudbjartsson, Craig Hanis, Torben Hansen, Astradur B. Hreidarsson, Kristian Hveem, Erik Ingelsson, Bo Isomaa, Stefan Johansson, Torben Jørgensen, Marit Eika Jørgensen, Sekar Kathiresan, Augustine Kong, Jaspal Kooner, Jasmina Kravic, Markku Laakso, Jong-Young Lee, Lars Lind, Cecilia M Lindgren, Allan Linneberg, Gisli Masson, Thomas Meitinger, Karen L Mohlke, Anders Molven, Andrew P. Morris, Shobha Potluri, Rainer Rauramaa, Rasmus Ribel-Madsen, Ann-Marie Richard, Tim Rolph, Veikko Salomaa, Ayellet V. Segrè, Hanna Skärstrand, Valgerdur Steinthorsdottir, Heather M. Stringham, Patrick Sulem, E Shyong Tai, Yik Ying Teo, Tanya Teslovich, Unnur Thorsteinsdottir, Jeff K. Trimmer, Tiinamaija Tuomi, Jaakko Tuomilehto, Fariba Vaziri-Sani, Benjamin F. Voight, James G. Wilson, Michael Boehnke, Mark I. McCarthy, Pål Njølstad, Oluf Pedersen, the Go-T2D consortium, the T2D-GENES consortium, Leif Groop, David R. Cox, Kari Stefansson, and David Altshuler

Gene	> 10x	Gene	> 10x	Gene	> 10x	Gene	> 10x	Gene	> 10x
ABCC8	98.48%	*CEL*	75.13%	*HNF4A*	99.27%	*NBPF7*	99.79%	*SLC22A18*	62.47%
ABCG5	96.93%	*COPG2*	98.34%	*IDE*	96.78%	*NEUROD1*	100.%	*SLC30A8*	99.89%
ABCG8	98.24%	*CREB5*	99.57%	*IGF2BP2*	97.77%	*NOTCH2*	99.15%	*SLC36A4*	94.31%
ACADS	93.34%	*CRHBP*	98.74%	*INS*	46.42%	*NUDT5*	99.77%	*SV2B*	99.77%
ADAM30	99.88%	*DHTKD1*	98.19%	*INS-IGF2*	13.51%	*OASL*	99.89%	*SYN2*	79.15%
ADAMTS9	99.72%	*DPY19L4*	95.95%	*INSR*	97.26%	*P2RX4*	96.6%	*TAX1BP1*	96.08%
AGGF1	96.29%	*DYNC2LI1*	98.49%	*INTS8*	96.79%	*P2RX7*	99.47%	*TCF7L2*	98.3%
ANAPC5	99.39%	*E2F3*	84.96%	*JAZF1*	98.76%	*PDE8B*	92.09%	*THADA*	99.5%
BCL11A	96.93%	*EIF2AK3*	90.89%	*KCNJ11*	98.6%	*PDX1*	48.32%	*TIMP4*	84.2%
BLM	98.39%	*ESRP1*	99.38%	*KCNQ1*	61.86%	*PHGDH*	97.62%	*TP53INP1*	99.92%
C12orf43	98.13%	*EXOC6*	98.63%	*KIF11*	98.8%	*PHLDA2*	49.25%	*TRA2B*	99.79%
C8orf37	99.63%	*F2RL1*	97.81%	*KLF11*	97.07%	*PLAGL1*	99.94%	*TRPM5*	51.15%
C8orf38	81.1%	*FAT3*	99.65%	*KLF14*	60.87%	*PLEKHF2*	99.88%	*TSEN2*	99.88%
CABP1	50.56%	*FES*	89.42%	*LGR5*	99.52%	*PLEKHH2*	98.39%	*TSGA13*	99.78%
CAMK1D	98.81%	*FURIN*	90.31%	*LIPH*	99.59%	*PPARG*	99.86%	*TSGA14*	99.77%
CAMKK2	97.22%	*GCK*	91.59%	*LMNA*	81.13%	*PPM1B*	99.83%	*TSPAN8*	97.32%
CCNE2	96.86%	*HDDC3*	76.95%	*LOC728819*	99.81%	*PRC1*	99.69%	*TSSC4*	73.19%
CD81	25.7%	*HHEX*	73.47%	*LRPPRC*	95.69%	*PTF1A*	39.94%	*UNC119B*	68.67%
CDC123	98.88%	*HIBADH*	95.68%	*MAN2A2*	96.64%	*RCCD1*	63.63%	*UNC45A*	94.05%
CDKAL1	99.68%	*HMGA2*	81.84%	*MEST*	99.05%	*REG4*	99.93%	*UPF2*	99.74%
CDKN1C	16.02%	*HMGCS2*	99.83%	*MLEC*	83.81%	*S100Z*	98.79%	*VPS33B*	98.76%
CDKN2A	95.48%	*HNF1A*	93.54%	*MTNR1B*	94.52%	*SEC61A2*	99.29%	*WFS1*	90.12%
CDKN2B	91.46%	*HNF1B*	98.05%	*NAP1L4*	99.59%	*SENP2*	97.22%	*ZBED3*	7.7%

Supplementary Table 1: **Genes targeted in initial sequencing experiment.** For initial sequencing, we selected 115 genes that either (i) lie within a genomic region associated with risk of T2D prior to 2008 (within 350kb of the SNP with strongest reported association), or (ii) contain variants reported to cause monogenic forms of diabetes. The table shows the name of each gene, as well as the fraction of targeted bases sequenced to mean 10x coverage (e.g., average, over individuals, of number of bases with at least ten reads aligned).

Genes with...	# genes	% genes
>99% bases >10x	37	32.2%
>95% bases >10x	77	67.0%
>80% bases >10x	95	82.6%
>50% bases >10x	107	93.0%
>25% bases >10x	112	97.4%

Supplementary Table 2: **Sequencing coverage of targeted genes.** Shown are the number and percentage of genes targeted in the initial sequencing experiment meeting various sequencing coverage thresholds, defined as the fraction of bases exceeding mean 10x coverage (e.g., average, over individuals, of number of bases with at least ten reads aligned). As genotypes were only analyzed at sites with at least 10x coverage, the mean number of bases sequenced to 10x estimates the sensitivity to detect a variant observed in a single individual.

Ethnciity	Origin	Cohort	T2D status	N	Female (%)	Age (yr)	BMI (kg/m^2)
European	Finland	Botnia	Case	121	41.3%	58.2 $\pm$ 10.3	24.8 $\pm$ 2.1
			Control	180	57.2%	57.4 $\pm$ 9.9	29.5 $\pm$ 3.7
European	Sweden	Malmo	Case	231	51.6%	50.5 $\pm$ 8.2	22.9 $\pm$ 1.6
			Control	226	38.2%	68.5 $\pm$ 5.4	35.4 $\pm$ 2.2

Supplementary Table 3: **Characteristics of individuals selected for the initial sequencing experiment.** Individuals selected for the initial sequencing experiment were drawn from Finnish and Swedish population based cohorts. To increase power to detect variants of moderate to large effect on T2D risk [1], cases were selected as young and lean (low environmental risk for T2D) and controls were selected as old and obese (high environmental risk for T2D). Shown are, for each cohort, the country of origin, the proportion of studied individuals who are female, the mean and standard deviation (s.d.) of individual ages, and the mean and s.d. of individual BMI values. Statistics are stratified by phenotype.

Ethnciity	Origin	Cohort	T2D status	N	Female (%)	Age (yr)	BMI (kg/m^2)
European	Finland	Botnia	Case	3,805	42.8%	65.8 $\pm$ 10.8	30.3 $\pm$ 5.4
			Control	5,308	53.5%	50.2 $\pm$ 15.5	26.3 $\pm$ 4.3
European	Sweden	Malmo	Case	6,729	40.7%	59.8 $\pm$ 12.6	29.7 $\pm$ 6.8
			Control	5,254	51.2%	57.7 $\pm$ 6.9	25.4 $\pm$ 3.5

Supplementary Table 4: **Characteristics of individuals genotyped for protein-truncating variants.** Six predicted protein-truncating variants were genotyped in additional individuals drawn from the same cohorts used for the initial sequencing experiment (see Supplementary Methods). Shown are characteristics of genotyped individuals from each cohort.

Gene	Codon change	Protein change	Consequence	1000G	PhyloP	MAF	Case	Ctrl
ABCC8	c.1879dupC	p.His627Profs*2	frameshift_variant	No	2.0	0.14%	0	2
ABCG5	c.575dupG	p.Ile193Hisfs*5	frameshift_variant	No	-1.4	0.69%	5	5
ABCG8	c.809_810insC	p.Gln271Profs*5	frameshift_variant	No	4.6	0.066%	0	1
ABCG8	c.1083G>A	p.Trp361*	stop_gained	No	5.3	0.6%	6	3
ACADS	c.563_566delATGC	p.Asn188Thrfs*75	frameshift_variant	No	4.6	0.067%	0	1
ADAM30	c.1837G>T	p.Gly613*	stop_gained	No	2.3	0.066%	1	0
ADAMTS9	c.116-2A>C	NA	splice_acceptor_variant	No	5.0	0.066%	1	0
ADAMTS9	c.637dupC	p.Gln213Profs*12	frameshift_variant	No	0.5	0.13%	1	1
ADAMTS9	c.2556+1G>A	NA	splice_donor_variant	No	4.2	0.066%	1	0
ADAMTS9	c.3226delC	p.Gln1076Argfs*11	frameshift_variant	No	4.9	0.066%	1	0
ADAMTS9	c.3804C>A	p.Tyr1268*	stop_gained	No	1.8	0.066%	0	1
ADAMTS9	c.4727_4734delGGTACCGC	p.Arg1576Lysfs*7	frameshift_variant	No	1.1	0.066%	1	0
ADAMTS9	c.5198-2A>G	NA	splice_acceptor_variant	No	3.0	0.066%	1	0
AGGF1	c.176_189delACGCAGAAAGCAAC	p.Asn59Lysfs*13	frameshift_variant	No	1.4	0.067%	1	0
AGGF1	c.987_988insC	p.Lys332Glnfs*7	frameshift_variant	No	-0.0	0.14%	1	1
AGGF1	c.1716+1G>A	NA	splice_donor_variant	No	5.5	0.072%	1	0
AGGF1	c.1728C>A	p.Tyr576*	stop_gained	Yes	0.1	0.069%	1	0
ANAPC5	c.145G>T	p.Glu49*	stop_gained	No	1.5	0.066%	1	0
ANAPC5	c.1482_1485delGGAA	p.Lys494Asnfs*26	frameshift_variant	No	0.6	0.067%	0	1
BCL11A	c.357_361delAATTT	p.Ile120Profs*9	frameshift_variant	No	2.9	0.066%	1	0
BCL11A	c.385_385+1insT	p.Asp129Valfs*2	frameshift_variant	No	6.2	0.066%	1	0
BCL11A	c.770_773delGGCG	p.Gly257Alafs*22	frameshift_variant	No	0.8	0.066%	1	0
BLM	c.1642C>T	p.Gln548*	stop_gained	No	2.3	0.066%	1	0
BLM	c.2092_2093insGTTA	p.Tyr699*	frameshift_variant	No	4.6	0.066%	1	0
C8orf38	c.222_223insTGCTCCCTGC	p.Glu82Alafs*16	frameshift_variant	Yes	0.3	0.066%	0	1
CAMK1D	c.499_500insT	p.Gly167Valfs*25	frameshift_variant	No	5.7	0.066%	1	0
CAMKK2	c.996dupT	p.Val333Cysfs*4	frameshift_variant	No	4.1	0.066%	1	0
CAMKK2	c.1560dupA	p.Pro521Thrfs*6	frameshift_variant	No	0.4	0.13%	2	0
CEL	c.1682delC	p.Pro562Leufs*145	frameshift_variant	No	-3.6	0.07%	1	0
CRHBP	c.958_959insG	p.Ser320Cysfs*16	frameshift_variant	No	0.6	0.067%	1	0
DPY19L4	c.747C>G	p.Tyr249*	stop_gained	No	0.7	0.066%	1	0
DPY19L4	c.1356_1357delTG	p.Val453Tyrfs*3	frameshift_variant	No	0.5	0.067%	0	1
DYNC2LI1	c.232-1G>A	NA	splice_acceptor_variant	No	5.0	0.066%	1	0
E2F3	c.913C>T	p.Arg305*	stop_gained	No	4.5	0.066%	0	1
E2F3	c.1201_1207delGCCTCCC	p.Ala401Glnfs*16	frameshift_variant	No	2.3	0.066%	0	1
E2F3	c.1225C>T	p.Gln409*	stop_gained	No	1.5	0.066%	1	0
EIF2AK3	c.634-1G>T	NA	splice_acceptor_variant	No	5.7	0.066%	1	0
EIF2AK3	c.1306+1G>A	NA	splice_donor_variant	No	5.8	0.066%	1	0
EIF2AK3	c.1538dupA	p.Asp514Glyfs*43	frameshift_variant	Yes	0.6	0.066%	1	0
ESRP1	c.267delC	p.Gln90Serfs*3	frameshift_variant	No	4.4	0.066%	1	0
ESRP1	c.1370_1371insA	p.Ile458Asnfs*2	frameshift_variant	No	1.5	0.066%	1	0
ESRP1	c.1847_1848delGC	p.Gly616Valfs*7	frameshift_variant	No	5.4	0.066%	0	1
EXOC6	c.2110G>T	p.Glu704*	stop_gained	No	5.6	0.066%	0	1
F2RL1	c.307C>T	p.Arg103*	stop_gained	Yes	1.5	0.066%	0	1
FAT3	c.835G>T	p.Glu279*	stop_gained	No	2.8	0.066%	1	0
FAT3	c.1135G>T	p.Glu379*	stop_gained	No	4.2	0.066%	1	0
FAT3	c.9695_9696insC	p.Val3235Cysfs*3	frameshift_variant	No	4.9	0.066%	1	0
FES	c.579delG	p.Gln194Serfs*30	frameshift_variant	No	5.3	0.067%	1	0
FES	c.1547_1566delAAGGGGAAGGCTTTCCTAGC	p.Glu516Aspfs*37	frameshift_variant	No	5.6	0.068%	0	1
FURIN	c.1648G>T	p.Glu550*	stop_gained	No	5.3	0.069%	0	1
GCK	c.45+1G>T	NA	splice_donor_variant	No	3.3	0.066%	1	0
HDDC3	c.113-2A>G	NA	splice_acceptor_variant	No	3.5	0.067%	1	0
HDDC3	c.358_359insCT	p.Leu120Profs*9	frameshift_variant	No	4.1	0.066%	1	0
HIBADH	c.853-2A>G	NA	splice_acceptor_variant	No	4.7	0.066%	0	1
HMGCS2	c.1169_1185delTGGCCTCGCTTCTGTCC	p.Leu390Profs*63	frameshift_variant	No	2.3	0.066%	1	0
HNF1A	c.770_771delAC	p.Asn257Thrfs*59	frameshift_variant	No	4.6	0.066%	0	1
HNF1B	c.1561dupC	p.Gln521Profs*30	frameshift_variant	No	5.1	1.2%	13	5
IDE	c.85G>T	p.Glu29*	stop_gained	No	0.3	0.11%	1	0
IDE	c.1435delG	p.Ala479Profs*2	frameshift_variant	No	5.8	0.066%	1	0
IDE	c.2209-1G>A	NA	splice_acceptor_variant	No	5.2	0.066%	0	1
INSR	c.591delC	p.Ala198Profs*84	frameshift_variant	No	0.1	0.066%	1	0

INSR	c.2278_2287delAAACGCAGGT	p.Lys760Profs*8	frameshift_variant	No	1.7	0.066%	1	0
INSR	c.3061G>T	p.Glu1021*	stop_gained	No	4.9	0.066%	0	1
INTS8	c.518+2T>A	NA	splice_donor_variant	No	4.8	0.066%	1	0
KCNJ11	c.78C>A	p.Tyr26*	stop_gained	No	1.9	0.066%	1	0
KLF14	c.701_710delAGAAGTTTAC	p.Lys234Serfs*34	frameshift_variant	No	0.2	0.067%	0	1
LGR5	c.1876C>T	p.Arg626*	stop_gained	No	2.0	0.066%	0	1
LIPH	c.940_944delACGAA	p.Thr314Glyfs*4	frameshift_variant	No	0.1	0.066%	1	0
LMNA	c.1886_1887insG	p.Ser632Glnfs*72	frameshift_variant	No	1.5	0.21%	2	1
LMNA	c.1970_1971insC	p.Gln659Profs*45	frameshift_variant	No	-1.0	0.066%	1	0
MAN2A2	c.359_360insG	p.Arg123Profs*19	frameshift_variant	No	2.3	0.58%	5	3
MAN2A2	c.3353_3360delCCTCCTTG	p.Thr1118Asnfs*14	frameshift_variant	No	2.9	0.066%	0	1
NAP1L4	c.670dupA	p.Thr224Asnfs*10	frameshift_variant	No	2.4	0.066%	1	0
NEUROD1	c.243_244delCA	p.Lys82Glufs*11	frameshift_variant	No	5.0	0.066%	1	0
NEUROD1	c.243delC	p.Lys82Argfs*10	frameshift_variant	No	5.0	0.066%	1	0
NEUROD1	c.454delG	p.Ala152Leufs*110	frameshift_variant	No	6.0	0.066%	1	0
NOTCH2	c.17_18delCC	p.Pro6Argfs*27	frameshift_variant	No	0.7	47.%	135	196
OASL	c.826delT	p.Tyr276Thrfs*55	frameshift_variant	No	0.8	0.066%	1	0
OASL	c.1178delA	p.Gln393Argfs*21	frameshift_variant	No	-0.3	0.066%	1	0
P2RX4	c.998_999insC	p.Thr335Hisfs*20	frameshift_variant	No	4.4	0.066%	1	0
P2RX4	c.1126_1127insA	p.Asp377Argfs*8	frameshift_variant	No	3.1	0.067%	1	0
P2RX7	c.125+1G>T	NA	splice_donor_variant	Yes	2.7	0.66%	4	6
P2RX7	c.225_229delCGTGG	p.Val76Glufs*12	frameshift_variant	No	-0.1	0.066%	1	0
P2RX7	c.534-1G>A	NA	splice_acceptor_variant	No	4.9	0.13%	2	0
PDX1	c.54C>A	p.Cys18*	stop_gained	No	1.5	0.076%	1	0
PLEKHH2	c.2248G>T	p.Glu750*	stop_gained	No	5.0	0.066%	0	1
PLEKHH2	c.3544G>T	p.Gly1182*	stop_gained	No	5.7	0.066%	0	1
PPM1B	c.1007delC	p.His337Metfs*34	frameshift_variant	No	6.2	0.066%	0	1
PRC1	c.970+1G>T	NA	splice_donor_variant	No	6.1	0.067%	0	1
SEC61A2	c.1292_1293insA	p.Arg433Lysfs*51	frameshift_variant	No	-1.2	0.066%	1	0
SEC61A2	c.1400_1403delAAGT	p.Glu467Valfs*60	frameshift_variant	No	6.1	0.066%	0	1
SLC22A18	c.659_660insC	p.Arg222fs	frameshift_variant	No	-0.4	0.25%	2	1
SLC22A18	c.1129C>T	p.Arg377*	stop_gained	No	0.9	0.067%	0	1
SLC22A18	c.1145_1146delCG	p.Thr382fs	frameshift_variant	No	3.3	0.067%	0	1
SLC22A18	c.1182_1183insT	p.Val395fs	frameshift_variant	No	-0.9	0.066%	1	0
SLC30A8	c.412C>T	p.Arg138*	stop_gained	No	1.5	0.13%	1	1
SLC30A8	c.456G>A	p.Trp152*	stop_gained	No	4.3	0.066%	0	1
SLC36A4	c.179+1G>T	NA	splice_donor_variant	No	3.8	0.066%	0	1
TAX1BP1	c.1598delA	p.Asp533Alafs*32	frameshift_variant	No	3.7	0.067%	1	0
TP53INP1	c.474-1G>C	NA	splice_acceptor_variant	No	5.0	0.066%	0	1
TRA2B	c.37-1G>A	NA	splice_acceptor_variant	No	5.0	0.2%	0	3
TRA2B	c.40delT	p.Ser14Profs*112	frameshift_variant	No	2.5	0.2%	0	3
TSGA14	c.926dupA	p.Ile310Aspfs*16	frameshift_variant	No	2.2	0.066%	1	0
TSSC4	c.485G>A	p.Trp162*	stop_gained	No	4.3	0.071%	0	1
UNC45A	c.927delC	p.Gln310Lysfs*7	frameshift_variant	No	2.5	0.066%	0	1
UNC45A	c.1738-1G>T	NA	splice_acceptor_variant	Yes	6.2	0.33%	3	2
UPF2	c.200dupA	p.Glu68Glyfs*79	frameshift_variant	No	1.6	0.066%	0	1
WFS1	c.1080_1081delCA	p.Thr361Profs*181	frameshift_variant	No	5.5	0.066%	1	0
WFS1	c.1970delT	p.Met657Argfs*8	frameshift_variant	No	2.1	0.067%	0	1
WFS1	c.1990_1994delCTGAC	p.Thr665Alafs*45	frameshift_variant	No	-1.0	0.066%	1	0

Supplementary Table 5: **List of predicted protein truncating variants identified in initial sequencing experiment.** We annotated potential protein truncating variants using the Variant Effect predictor (VEP) [2]. Analysis focused on nonsense SNPs, frameshifting indels, or splice site variants. Shown in this table are all such variants identified in the intiial sequencing experiment. The Consequence field was produced by the VEP; minor allele frequency (MAF) was computed based on all sequenced individuals. The 1000G field indicates whether the variant was observed in the 1000 genomes project [3], while the PhyloP column shows a quantiative measure of conservation across an alignment of vertebrate species (positive values are conserved). Case and Ctrl show the number of observations in individuals with and without T2D respectively.

Gene	Codon change	Protein change	Origin	MAF	N		Carriers		Carrier frequency		OR (95% CI)	P
					Case	Ctrl	Case	Ctrl	Case	Ctrl		
SLC30A8	c.412C>T	p.Arg138*	Finland	0.0026	3727	5440	9	39	0.0012	0.0036	0.46 (0.25 - 0.84)	0.012
			Sweden	0.0002	6960	5480	2	3	0.00014	0.00027		
BLM	c.1642C>T	p.Gln548*	Finland	0.	2816	4785	0	0	0.	0.	0.25 (0.05 - 1.33)	0.1
			Sweden	0.00041	5387	4296	2	6	0.00019	0.0007		
F2RL1	c.307C>T	p.Arg103*	Finland	6.6e-05	2816	4785	1	0	0.00018	0.	0.53 (0.13 - 2.19)	0.38
			Sweden	0.00031	5387	4297	2	4	0.00019	0.00047		
UNC45A	c.1738-1G>T	NA	Finland	0.00013	2815	4784	1	1	0.00018	0.0001	1.63 (0.50 - 5.37)	0.42
			Sweden	0.0013	5387	4296	16	10	0.0015	0.0012		
P2RX7	c.125+1G>T	NA	Finland	0.0042	2815	4784	20	44	0.0036	0.0046	0.96 (0.68 - 1.36)	0.83
			Sweden	0.006	5387	4297	70	46	0.0065	0.0054		
ABCG8	c.1083G>A	p.Trp361*	Finland	0.0042	3381	5025	27	44	0.004	0.0044	0.98 (0.82 - 1.18)	0.84
			Sweden	0.0016	6612	5089	22	15	0.0017	0.0015		

Supplementary Table 6: **Association statistics from additional genotyping of variants from the initial sequencing experiment.** Predicted protein truncating variants identified in the initial sequencing experiment and also present on the Illumina Exome Array were genotyped in additional individuals from Sweden and Finland (drawn from the same cohorts used for the initial sequencing experiment). Frequency, counts, and statistics were computed as defined in Supplementary Methods. Based on the observed association statistics in this genotyping, the p.Arg138* variant was further genotyped in additional individuals from Sweden, Finland, and Denmark (from independent cohorts); the sample and association statistics resulting from that analysis are shown in Supplementary Table 4 and Table 1.

Ethnciity	Origin	Cohort	T2D status	N	Female (%)	Age (yr)	BMI (kg/m^2)
European	Denmark	Danish	Case	3,889	37.6%	51.8 $\pm$ 12.6	27.0 $\pm$ 5.2
			Control	7,869	56.7%	46.8 $\pm$ 10.4	25.3 $\pm$ 4.2
European	Finland	Finnish	Case	4,050	32.9%	62.0 $\pm$ 8.0	30.4 $\pm$ 5.2
			Control	8,696	25.7%	59.0 $\pm$ 7.4	26.6 $\pm$ 3.8
European	Sweden	PIVUS/ULSAM	Case	271	16.9%	70.7 $\pm$ 0.7	28.5 $\pm$ 4.6
			Control	1,791	24.2%	70.6 $\pm$ 0.6	26.4 $\pm$ 3.8

Supplementary Table 7: **Characteristics of additional individuals genotyped for the p.Arg138* variant.** The p.Arg138* variant was genotyped in additional individuals from several European countries, including independent Finnish, Swedish, and Danish cohorts. Additional individuals genotyped from Poland and Britain (see Supplementary Methods) were not included in phenotypic analysis given the small number of carrier observations from either cohort. Shown are characteristics of genotyped individuals from each cohort included in phenotypic association analysis.

Ethnciity	Origin	Cohort	T2D status	N	Female (%)	Age (yr)	BMI (kg/m^2)
European	Iceland	deCODE	Case	3,327	39.4%	65.6 $\pm$ 3.8	30.7 $\pm$ 5.8
			Control	78,102	39.6%	50.0 $\pm$ 23.2	26.9 $\pm$ 5.0
European	Norway	HUNT2	Case	1,645	53.4%	68.4 $\pm$ 12.1	29.2 $\pm$ 4.8
			Control	4,069	50.9%	58.1 $\pm$ 17.9	26.6 $\pm$ 4.1

Supplementary Table 8: **Characteristics of individuals genotyped for p.Lys34Serfs*50.** The p.Lys34Serfs*50 variant was genotyped in additional individuals from Iceland and Norway. Shown are characteristics of genotyped individuals from each cohort.

Phenotype	MAF	N cases		N controls		DirImp carriers		OR (95% CI)	P
		Total	DirImp	Total	DirImp	Cases	Controls		
T2D	0.0018	3,463	2,953	79,649	67,919	2	248	0.18 (0.06 - 0.58)	0.0041
T2D Early onset	0.0018	485	409	82,935	69,939	0	239	0.02 (0.00 - 1.82)	0.087
T2D Retinopathy	0.0018	1,746	1,269	94,284	68,526	0	229	0.01 (0.00 - 0.23)	0.0026

Supplementary Table 9: **Association statistics for p.Lys34Serfs*50 variant in Icelandic individuals.** We tested for association between p.Lys34Serfs*50 and type 2 diabetes (T2D), early onset T2D, and T2D retinopathy as described in Supplementary Methods. The number of directly imputed cases and controls is indicated (DirImp); the total list includes in addition individuals imputed using familial information.

Ethncity	Origin	Cohort	T2D status	N	Female (%)	Age (yr)	BMI (kg/m^2)
African-American	USA	JHS	Case	534	59.9%	63.8 ± 9.1	29.3 ± 6.6
			Control	536	56.0%	51.0 ± 11.5	30.0 ± 7.1
African-American	USA	WFS	Case	502	66.3%	58.0 ± 10.4	33.3 ± 6.4
			Control	527	63.4%	56.0 ± 11.3	31.9 ± 6.4
East Asian	Korea	KARE	Case	530	45.7%	53.8 ± 7.5	25.7 ± 3.3
			Control	567	58.2%	63.3 ± 3.6	23.7 ± 3.1
East Asian	Singapore	Singapore Chinese	Case	487	51.9%	58.0 ± 9.3	25.6 ± 3.8
			Control	592	61.3%	58.3 ± 7.0	22.8 ± 3.4
European	Finland	Botnia	Case	130	60.8%	59.0 ± 11.1	26.0 ± 3.0
			Control	62	29.0%	63.6 ± 11.4	28.7 ± 4.0
European	Finland	FUSION	Case	485	41.9%	57.6 ± 7.9	30.9 ± 5.5
			Control	477	44.9%	62.9 ± 7.2	27.9 ± 3.9
European	Finland	METSIM	Case	487	0.0%	60.4 ± 6.7	30.6 ± 5.1
			Control	500	0.0%	54.8 ± 4.5	25.8 ± 3.1
European	Germany	KORA	Case	102	44.1%	61.2 ± 8.3	28.3 ± 2.8
			Control	95	65.3%	69.7 ± 5.6	34.5 ± 3.5
European	Sweden	Malmo	Case	39	56.4%	56.1 ± 12.2	24.7 ± 2.4
			Control	13	38.5%	67.5 ± 4.6	36.2 ± 2.1
European	UK	UKT2D	Case	324	46.0%	50.0 ± 8.4	26.8 ± 2.7
			Control	321	82.9%	60.7 ± 10.0	30.6 ± 5.9
European	USA	Ashkenazim	Case	509	46.9%	65.8 ± 8.7	27.4 ± 3.2
			Control	354	56.5%	78.4 ± 13.8	25.2 ± 4.1
Hispanic	USA	San Antonio	Case	246	57.3%	57.6 ± 12.3	32.0 ± 6.4
			Control	184	58.1%	50.7 ± 14.8	30.3 ± 6.5
Hispanic	USA	Starr County	Case	760	59.9%	56.3 ± 11.9	31.8 ± 6.4
			Control	711	72.0%	39.2 ± 9.9	30.2 ± 6.2
South Asian	Singapore	Singapore Indians	Case	563	44.4%	60.8 ± 9.7	26.9 ± 5.1
			Control	586	49.3%	56.1 ± 10.1	25.3 ± 4.8
South Asian	UK	LOLIPOP	Case	531	14.1%	52.8 ± 5.6	26.6 ± 2.9
			Control	540	15.7%	63.4 ± 9.2	27.2 ± 3.5

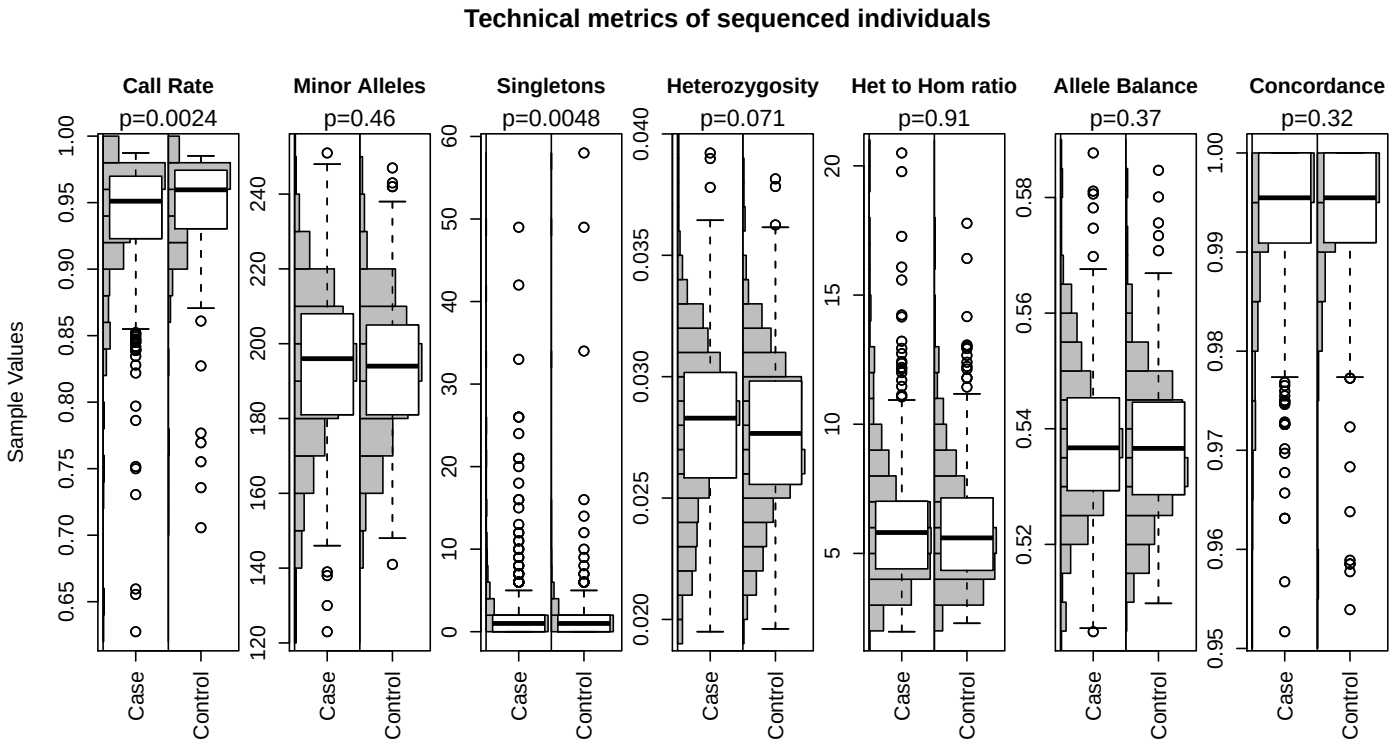
Supplementary Table 10: **Characteristics of individuals selected for additional sequencing of *SLC30A8*.** As part of whole exome or whole-genome sequencing experiments, the exons of *SLC30A8* were sequenced in additional individuals from multiple ethnicities (European, African-American, South Asian, East Asian, and Hispanic). For each ethnicity, individuals were drawn from multiple cohorts (see Supplementary Methods). Shown are characteristics of sequenced individuals from each cohort.

Disease/Trait	Origin	Reference	MAF	OR/Beta	P-value
T2D (DIAGRAM v3)	European	[4]	N/A	0.86	2.2e-11
T2D (DIAGRAM v3 + Metabochip)	European	[4]	0.33	0.88	1.3e-21
Fasting glucose	European	[5]	0.32	-0.027	5.5e-10
Fasting insulin	European	[5]	0.32	0.0036	0.44
Fasting proinsulin	European	[6]	0.24	-0.069	4.9e-11
HOMA-IR	European	[5]	0.32	0.0002	0.97
HOMA-B	European	[5]	0.33	0.016	2.4e-05
2 hr glucose, BMI-adjusted	European	[7]	N/A	-0.088	7.3e-05
2 hr insulin, BMI-adjusted	European	[7]	N/A	0.012	0.47
Hb1Ac	European	[8]	N/A	-0.015	0.051
Body Mass Index (BMI)	European	[9]	0.32	0.012	0.024
Waist-hip ratio, BMI-adjusted	European	[10]	0.25	N/A	0.051
Coronary artery disease (CARDIoGRAM)	European	[11]	0.32	0.99	0.45
Chronic kidney disease	European	[12]	0.25	+	0.0088
eGFR-cys	European	[12]	0.25	-	0.046
eGFR-creat	European	[12]	0.25	-	0.27
Microalbumin	European	[13]	0.25	+	0.32
Urinary albumin/creatinine	European	[13]	0.25	+	0.35
LDL-Cholesterol	European	[14]	N/A	-2.4	0.018
HDL-Cholesterol	European	[14]	N/A	1.3	0.18
Total-Cholesterol	European	[14]	N/A	-2.3	0.022
Triglycerides	European	[14]	N/A	-2.5	0.011
Diastolic Blood Pressure	European	[15]	0.3	-0.041	0.57
Systolic Blood Pressure	European	[15]	0.3	-0.086	0.45

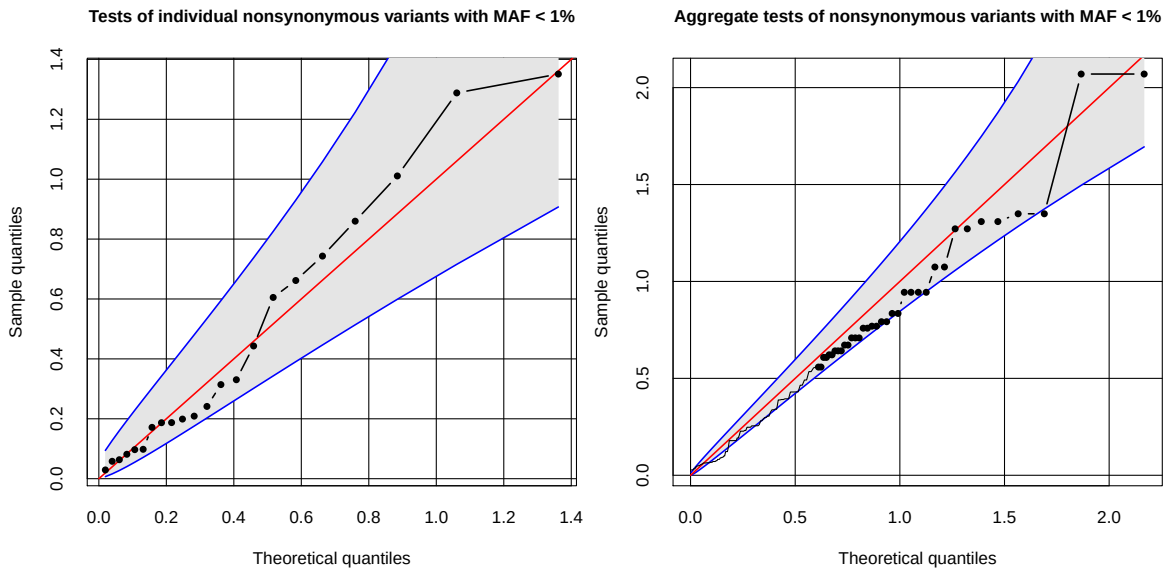
Supplementary Table 11: **Association of p.Trp325Arg variant with metabolic and cardiovascular traits.** The table shows published associations between the common variant in *SLC30A8* (rs13266634) and metabolic traits. All associations are computed with the effect allele equal to the minor allele (W). The OR/Beta/Z-score column shows odds ratios (OR) for T2D, beta estimates for the quantitative traits other than lipids, and Z-scores for lipids. Beta estimates are scaled to the units used in the original publications. Where estimates are not available, $+$ or $-$ indicates increasing or decreasing.

Phenotype	Variant	Cohort	Individually				Combined			
			N	Carriers	Beta (95% CI)	P	N	Carriers	Beta (95% CI)	P
Non-fasting Glucose	p.Lys34Serfs*50	deCODE	60,854	182	−0.17 (−0.27 − −0.08)	0.00046	60,854	182	−0.17 (−0.27 − −0.08)	0.00046
1-hr Glucose	p.Lys34Serfs*50	deCODE	1,602	4	−0.73 (−1.62 − 0.17)	0.05	1,602	4	−0.73 (−1.62 − 0.17)	0.05
Fasting Insulin	Multiple	T2D-GENES	3,560	8	0.68 (−0.01 − 1.38)	0.054	11,326	52	0.24 (0.52 − −0.04)	0.088
	p.Arg138*	Botnia	5,584	39	0.19 (−0.12 − 0.49)	0.23				
	p.Lys34Serfs*50	deCODE	2,182	5	0.05 (−0.74 − 0.84)	0.9				
Fasting Plasma Glucagon	p.Arg138*	Botnia	1,374	7	−0.49 (−1.24 − 0.25)	0.19	1,374	7	−0.49 (−1.24 − 0.25)	0.19
Fasting Glucose	Multiple	T2D-GENES	3,889	9	0.44 (−0.21 − 1.09)	0.18	45,514	146	−0.10 (−0.24 − 0.05)	0.2
	p.Arg138*	Botnia	5,643	39	−0.22 (−0.53 − 0.10)	0.18				
	p.Arg138*	Danish	5,546	7	−0.12 (−0.86 − 0.62)	0.75				
	p.Lys34Serfs*50	deCODE	30,436	91	−0.10 (−0.26 − 0.06)	0.23				
Waist-Hip Ratio	Multiple	T2D-GENES	6,374	10	0.25 (−0.37 − 0.87)	0.43	12,976	50	0.15 (0.44 − −0.15)	0.33
	p.Arg138*	Botnia	6,602	40	0.12 (−0.19 − 0.43)	0.45				
Proinsulin	p.Arg138*	Botnia	4,540	37	−0.14 (−0.47 − 0.18)	0.39	4,540	37	−0.14 (−0.47 − 0.18)	0.39
Fasting Serum Glucagon	p.Arg138*	Botnia	3,187	29	0.16 (−0.21 − 0.52)	0.4	3,187	29	0.16 (−0.21 − 0.52)	0.4
2-hr Glucose	Multiple	T2D-GENES	1,930	1	0.53 (−1.11 − 2.17)	0.53	8,712	10	−0.20 (−0.88 − 0.48)	0.56
	p.Arg138*	Danish	5,285	7	−0.09 (−0.83 − 0.65)	0.81				
	p.Lys34Serfs*50	deCODE	1,497	2	−0.89 (−2.10 − 0.32)	0.24				
HDL	Multiple	T2D-GENES	9,116	17	0.13 (−0.35 − 0.60)	0.6	69,176	215	0.05 (0.21 − −0.11)	0.58
	p.Arg138*	Botnia	8,319	42	−0.05 (−0.35 − 0.25)	0.73				
	p.Lys34Serfs*50	deCODE	51,741	156	0.07 (−0.10 − 0.23)	0.44				
BMI	Multiple	T2D-GENES	10,212	20	0.01 (−0.43 − 0.45)	0.95	94,034	269	−0.04 (−0.20 − 0.12)	0.61
	p.Arg138*	Botnia	9,861	48	0.10 (−0.19 − 0.38)	0.51				
	p.Lys34Serfs*50	deCODE	73,961	201	−0.10 (−0.27 − 0.07)	0.26				
Diastolic Blood Pressure	Multiple	T2D-GENES	9,129	17	−0.01 (−0.49 − 0.47)	0.97	39,642	115	−0.05 (−0.27 − 0.17)	0.67
	p.Arg138*	Botnia	7,344	41	−0.04 (−0.35 − 0.27)	0.78				
	p.Lys34Serfs*50	deCODE	23,169	57	−0.06 (−0.34 − 0.21)	0.65				
LDL	Multiple	T2D-GENES	8,879	17	0.33 (−0.14 − 0.81)	0.17	70,921	220	−0.02 (−0.13 − 0.08)	0.68
	p.Arg138*	Botnia	8,253	42	−0.20 (−0.49 − 0.10)	0.19				
	p.Lys34Serfs*50	deCODE	53,789	161	−0.03 (−0.20 − 0.14)	0.74				
Systolic Blood Pressure	Multiple	T2D-GENES	9,129	17	−0.12 (−0.60 − 0.35)	0.61	39,646	116	−0.05 (−0.27 − 0.17)	0.68
	p.Arg138*	Botnia	7,345	41	−0.03 (−0.33 − 0.28)	0.86				
	p.Lys34Serfs*50	deCODE	23,172	58	−0.04 (−0.31 − 0.24)	0.8				
Fasting C-peptide Levels	p.Arg138*	Botnia	2,016	17	0.07 (−0.39 − 0.53)	0.76	2,016	17	0.07 (−0.39 − 0.53)	0.76
Triglycerides	p.Arg138*	Botnia	8,359	42	−0.00 (−0.31 − 0.30)	0.98	56,953	190	−0.03 (−0.19 − 0.14)	0.76
	p.Lys34Serfs*50	deCODE	48,594	148	−0.03 (−0.20 − 0.14)	0.7				
Proinsulin:Fasting Insulin	p.Arg138*	Botnia	4,523	37	−0.03 (−0.35 − 0.29)	0.86	4,523	37	−0.03 (−0.35 − 0.29)	0.86
Total Cholesterol	Multiple	T2D-GENES	9,131	17	0.33 (−0.15 − 0.81)	0.18	71,682	219	0.01 (0.16 − −0.14)	0.89
	p.Arg138*	Botnia	8,365	42	−0.18 (−0.48 − 0.12)	0.24				
	p.Lys34Serfs*50	deCODE	54,186	160	0.03 (−0.11 − 0.16)	0.71				

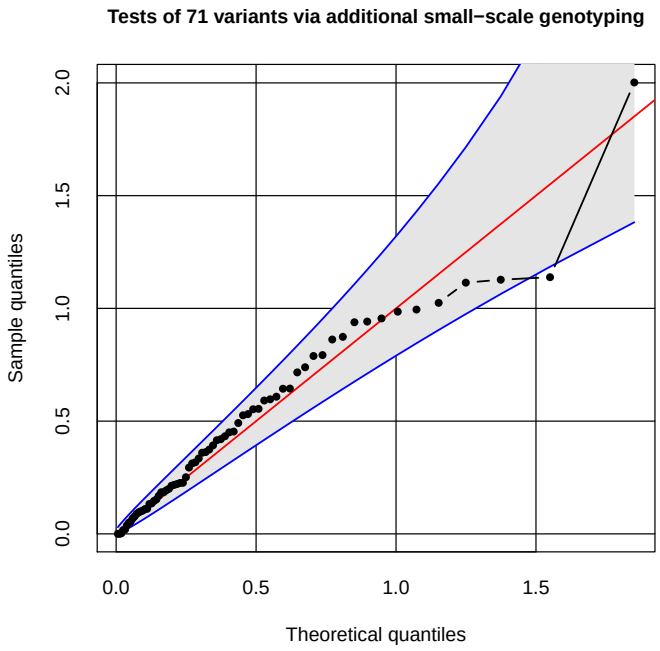
Supplementary Table 12: **Association statistics of *SLC30A8* protein truncating variants and secondary traits.** Association was tested between the 12 identified protein truncating variants and a variety of quantitative traits, as described in Supplementary Methods. Analysis was performed separately for p.Arg138* (in the Botnia or Danish cohorts), p.Lys34Serfs*50 (in the deCODE cohort), or the remaining variants in aggregate (from the cohorts in the additional sequencing experiment, labeled here as T2D-GENES); for most traits, only a subset of the cohorts were included in analysis (due to either missing phenotypic information or a small number of observed carriers). Effect size estimates and statistical significance are shown separately for each analysis in the columns entitled 'Individually'; the columns entitled 'Combined' show the combined association across all variants, obtained via a random-effects meta-analysis as described in Supplementary Methods. When trait information was available for only one variant, the association in the 'Combined' columns is identical to that in the 'Individually' columns. Effect sizes are given in standardized units (e.g., the analyzed trait has zero mean and unit variance).



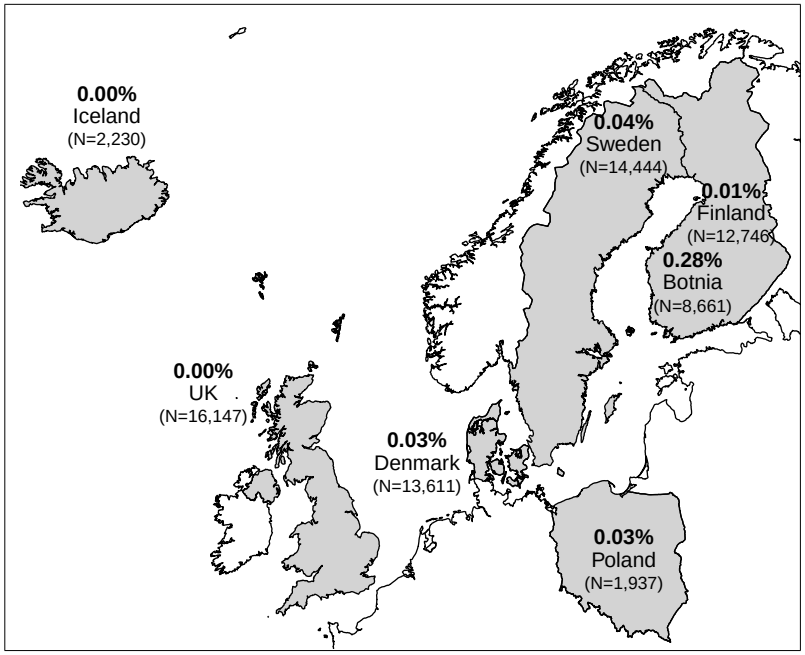
Supplementary Figure 1: **Technical quality control metrics of targeted sequencing.** We computed various metrics to evaluate the quality of the initial sequencing experiment. Shown from the left are the distributions (over all sequenced samples) of the number of variant sites with 10x sequence coverage and hence nonmissing genotypes (Call Rate), the number of minor alleles in genotypes called across all variant sites (Minor Alleles), the number of minor alleles at sites where no other samples have minor alleles (Singletons), the fraction of genotypes called heterozygous (Heterozygosity), the ratio of the number of heterozygous sites to the number of minor allele homozygous sites (Het to Hom ratio), the fraction of sequence reads with the minor allele (averaged over all heterozygous sites; Allele Balance) and the fraction of minor allele genotypes identical to those called from Exome Chip or Metabochip genotyping (Concordance). Absent differential technical artifacts or population structure, distributions are expected to be similar between cases and controls. Statistical comparison between case and control distributions was performed via a Kruskal-Wallis one-way analysis of variance test.



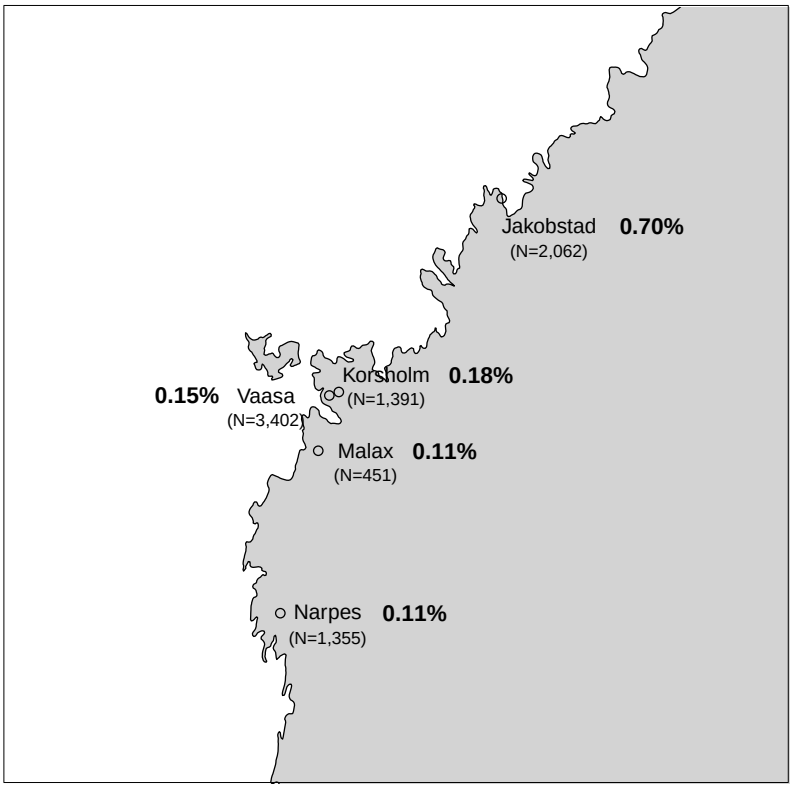
Supplementary Figure 2: **Association with T2D of nonsynonymous variants from the initial sequencing experiment.** Based on data from the initial sequencing experiment, we performed two types of association tests for all low-frequency (below 1%) nonsynonymous variants. First, variants were tested individually using a linear mixed model. Second, variants were collapsed within each gene and tested for aggregate association using the same mixed model approach. Shown are QQ plots of association for each approach: the x-axis plots the expected distribution of $(-\log_{10})$ P-values under the null model whereby no association exists across the entire experiment (e.g. the uniform distribution); the y-axis plots the observed distribution of $(-\log_{10})$ p-values. The blue lines show estimated 95% confidence intervals for observed distributions under the null model. For each plot, only variants (or, analogously, genes) for which ten or more carriers are observed are plotted: for variants with small numbers of counts, the uniform distribution is not a good approximation to the P-value distribution expected under the null model.



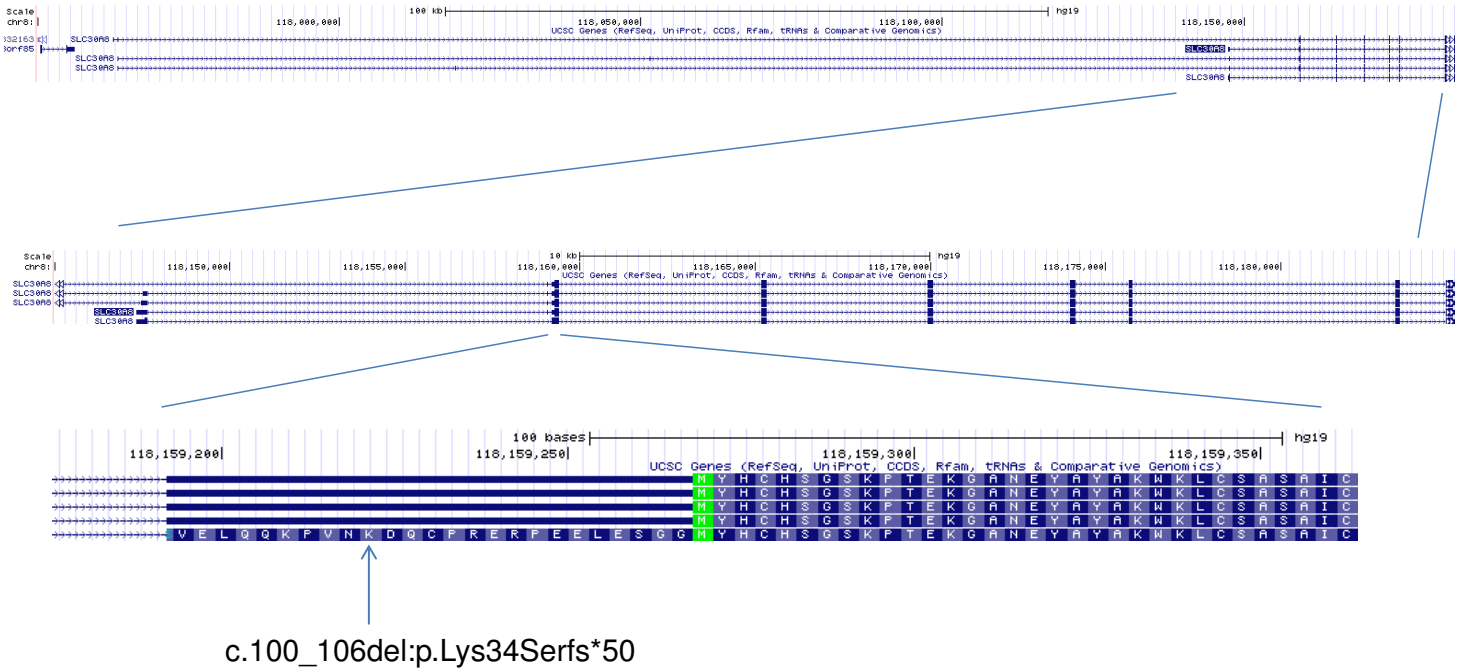
Supplementary Figure 3: **Association with T2D of 71 variants from small-scale genotyping.** We genotyped select variants from the initial sequencing experiment in up to 13,884 additional Finnish and Swedish individuals. Shown is a QQ plot of associations (with the same format as in Supplementary Figure 2), tested via a logistic regression.



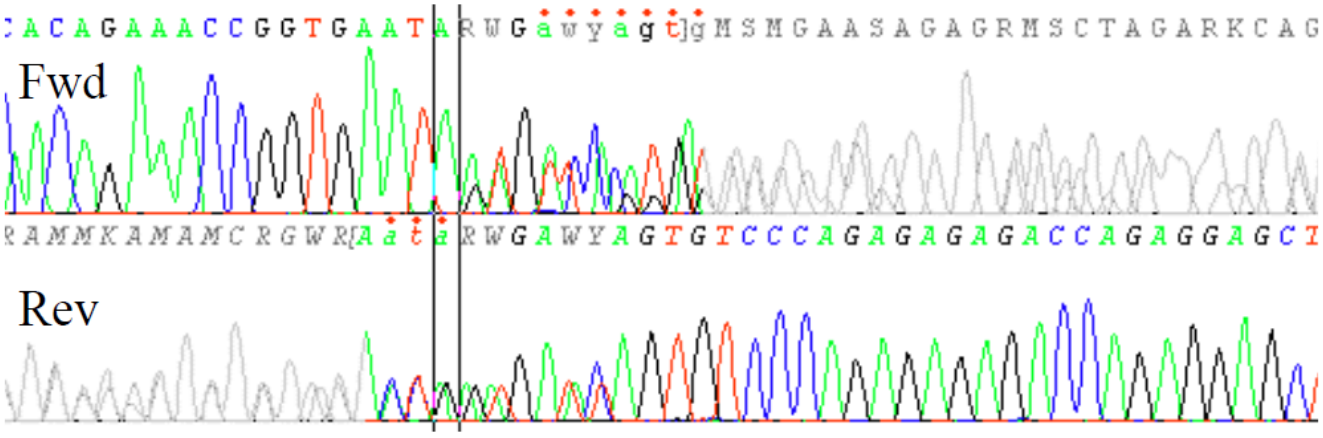
Supplementary Figure 4: **Frequency of the p.Arg138* variant in Europe.** The p.Arg138* variant was genotyped in individuals from Finland, Sweden, Denmark, the UK, and Poland (through the Illumina Human Exome Array or custom Sequenom genotyping), as well as Iceland (through whole genome sequencing). Shown are the observed frequencies of the variant in each country; frequency estimates from Finland are stratified by the Botnia region (individuals from the Botnia cohort) and other regions of Finland. N indicates the number of individuals genotyped.



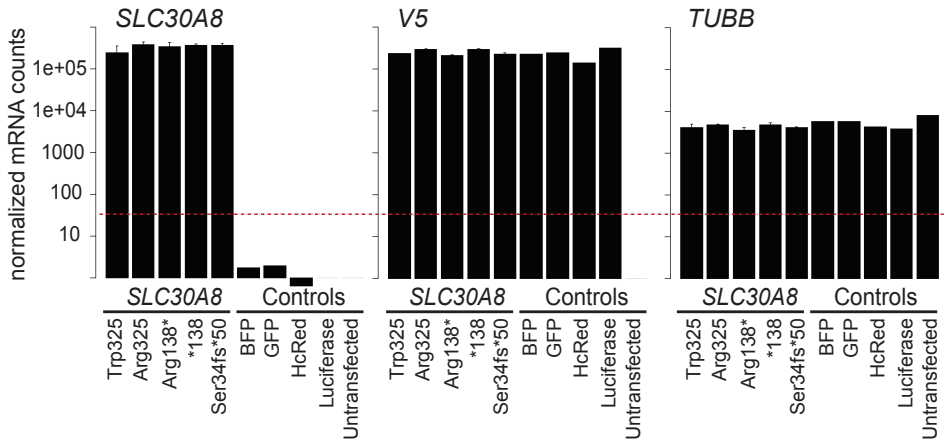
Supplementary Figure 5: **Frequency of the p.Arg138* variant in Botnia.** Frequency estimates for Botnia (as computed in Supplementary Figure 4) were further stratified by town in Botnia. Information on the center of sample selection was available for all samples from the Botnia region of Finland. Individuals were grouped by study center, and frequency estimates were computed for each center separately. N indicates the number of individuals genotyped.



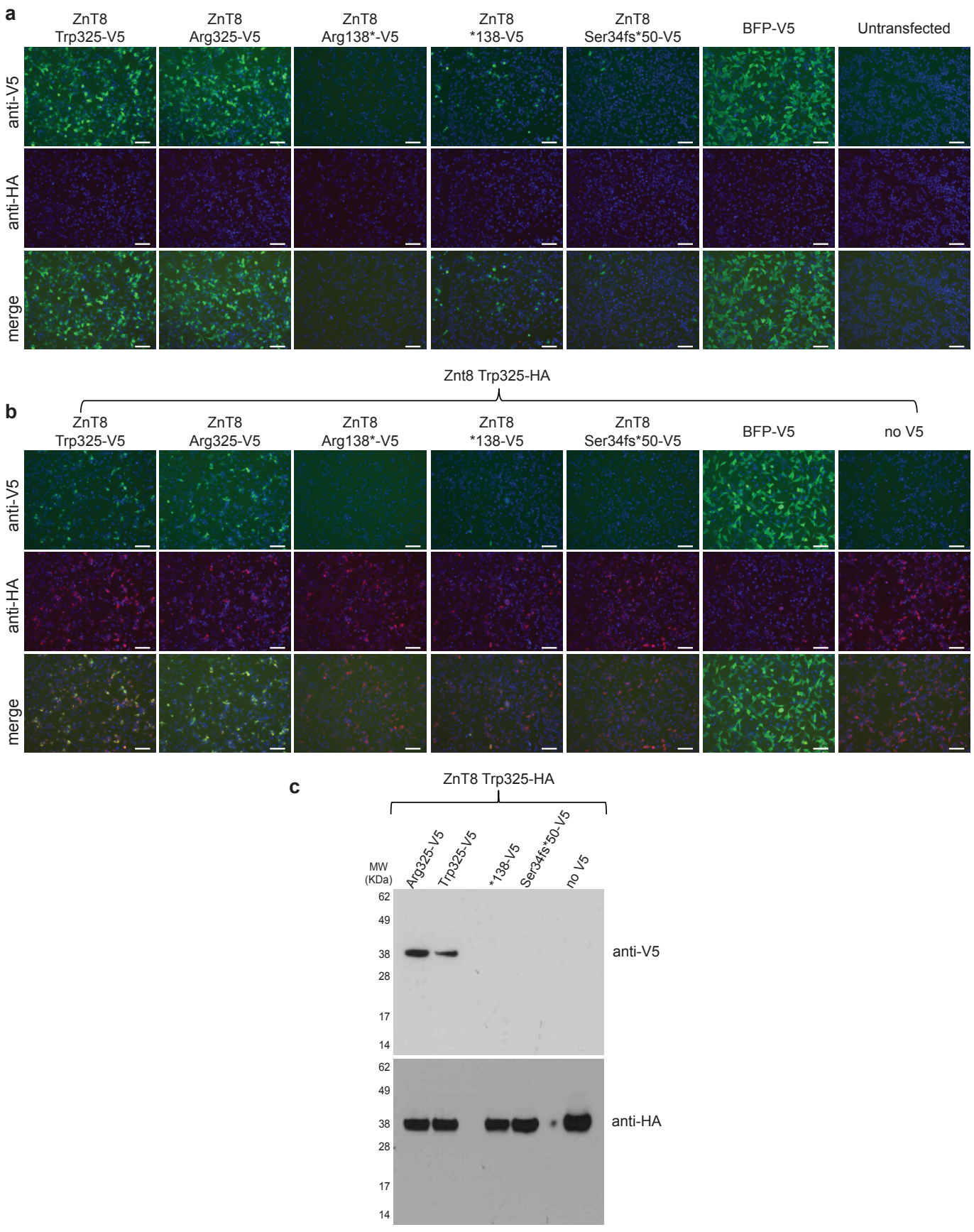
Supplementary Figure 6: **Position of the p.Lys34Serfs*50 frameshift mutation.** Shown is the genomic position of the frameshift mutation p.Lys34Serfs*50, as visualized in the UCSC genome browser [16].



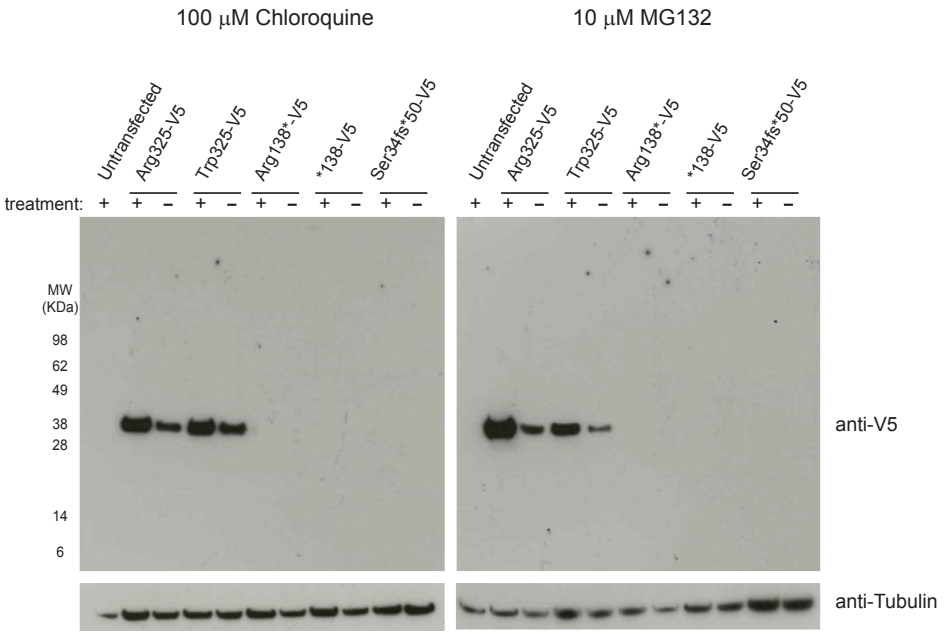
Supplementary Figure 7: **Partial sequence chromatogram for the p.Lys34Serfs*50 frameshift variant.** Sanger sequencing was used to confirm carriers of the frameshift mutation. Shown is a chromatogram for one individual carrier (from Norway) heterozygous for the variant.



Supplementary Figure 8: **Transcript levels of *SLC30A8* variants.** nCounter analysis of *SLC30A8* transcript levels in HeLa cells, following transient over-expression of C-terminal, V5-tagged ZnT8 variants and control ORFs. Data shown are mean, normalized mRNA counts $\pm$ s.e.m. of three independent plasmid transfections from two experiments. Non-specific binding is indicated by the red line. Gene expression detected by probes directed against the *SLC30A8* sequence encoding the N terminus, the sequence encoding the V5 tag, and *TUBB* are shown.



Supplementary Figure 9: **Coexpression of *SLC30A8* variants.** HeLa cells overexpressed either **(a)** singly transfected C-terminal, V5-tagged ZnT8 variants or **(b)** cotransfections that also included C-terminally HA-tagged Trp325-ZnT8 for 24h, after which ZnT8 protein levels were observed. Cells were immunostained for ZnT8 expression using anti-HA and anti-V5 antibodies and costained with Hoechst 33342 to mark nuclei. Scale bars, 100μm **(c)** Protein blot analysis of HeLa cell lysates following cotransfection of Trp325-HA ZnT8 and V5-tagged ZnT8 variants. ZnT8 expression was detected using anti-V5 and anti-HA antibodies.



Supplementary Figure 10: **Inhibition of protein degradation.** HeLa cells transiently overexpressing V5-tagged ZnT8 variants were treated (+) with chloroquine (100 μM) or MG132 (10 μM) or left untreated in medium absent these chemicals, over a 4-h incubation period, to inhibit lysosomal and proteasomal degradation, respectively. Protein blot analysis was performed using anti-V5 and anti-tubulin antibodies.

Supplementary Note

Consortia information

The GoT2D Consortium

Gonçalo Abecasis[1], Vineeta Agarwala[2], Peter Algren[3], David Altshuler[2,4,5,6,7,8,9], Martin Hrabé de Angelis[10], Eric Banks[2], Richard Bergman[11], Thomas Blackwell[1], Michael Boehnke[1], Lori Bonnycastle[12], David Buck[13], Martijn van de Bunt[14], Noël Burtt[2], Peter Chines[12], Francis Collins[12], Mark DePristo[2], Peter Donnelly[13], Timothy Fennell[2], Jason Flannick[2,5], Pierre Fontanillas[2], Timothy Frayling[15], Christian Fuchsberger[1], Stacey Gabriel[2], Kyle Gaulton[13], Christian Gieger[10], Harald Grallert[10], Todd Green[2], Leif Groop[3], Christopher Hartl[2], Andrew Hattersley[15], Bryan Howie[16], Cornelia Huth[10], Jeroen Huyghe[1], Bo Isomaa[17,18], Anne Jackson[19], Goo Jun[1], Hyun Min Kang[1], Jasmina Kravic[3], Jennifer Kriebel[10], Ashish Kumar[13], Phoenix Kwan[1], Claes Ladenvall[3], Cecilia Lindgren[13], Adam Locke[1], Gerton Lunter[13], Clement Ma[1], Anubha Mahajan[13], Alisa Manning[2], Mark McCarthy[13,14,20] Gil McVean[13], Christa Meisinger[10], Thomas Meitinger[21], Karen Mohlke[22], Andrew Morris[13,23], Loukas Moutsianas[13], Martina Müller-Nurasyid[10], Pål Njølstad[24,25], Richard Pearson[13], John Perry[13], Annette Peters[10], Ryan Poplin[2], Inga Prokopenko[13,26], Wolfgang Rathmann[10], William Rayner[13], Janina Ried[10], Manuel Rivas[13], Neil Robertson[13], Laura Scott[1], Khalid Shakir[2], Xueling Sim[1], Kerrin Small[27], Timothy Spector[27], Michael Stitzel[28], Konstantin Strauch[10], Heather Stringham[1], Tim Strom[21], Adrian Tan[1], Tanya Teslovich[1], Tiinamaija Toumi[17,18], Jaakko Tuomilehto[29]

The T2D-GENES Consortium

Gonçalo Abecasis[1], Marcio Almeida[30], David Altshuler[2,4,5,6,7,8,9], Jennifer Asimit[31], Gil Atzmon[32], Mathew Barber[16], Nir Barzilai[32], Nicola Beer[14], Graeme Bell[16,33], Jennifer Below[34], Tom Blackwell[1], John Blangero[30], Michael Boehnke[1], Donald Bowden[35], Noël Burtt[2], John Chambers[36,37,38], Han Chen[39], Peng Chen[40], Peter hines[12], Sungkyoung Choi[41], Claire Churchhouse[2], Pablo Cingolani[42], Belinda Cornes[43], Nancy Cox[16,33], Aaron Day-Williams[31], Ravindranath Duggirala[30], Josée Dupuis[39], Thomas Dyer[30], Shuang Feng[1], Juan Fernandez-Tajes[13], Teresa Ferreira[13], Tasha Fingerlin[44], Jason Flannick[2,5], Jose Florez[2,5,6], Pierre Fontanillas[2], Timothy Frayling[15], Christian Fuchsberger[1], Eric Gamazon[33], Kyle Gaulton[13], Saurabh Ghosh[45], Benjamin Glaser[46], Anna Gloyn[14], Robert Grossman[33,47], Jason Grundstad[47], Craig Hanis[34], Allison Heath[47], Heather Highland[34], Momoko Hirokoshi[13], Ik-Soo Huh[41], Jeroen Huyghe[1], Kamran Ikram[48,43,49,50], Kathleen Jablonski[51], Young Jin Kim[52], Goo Jun25, Norihiro Kato[53], Jayoun Kim[41], Ryan King[54], Jaspal Kooner[37,38,55], Min-Seok Kwon[41], Hae Kyung Im[54], Markku Laakso[56], Kevin Koi-Yau Lam[40], Jaehoon Lee[41], Selyeong Lee[41], Sungyoung Lee[47], Donna Lehman[57], Heng Li[2], Cecilia Lindgren[13], Xuanyao Liu[40,58], Oren Livne[16], Adam Locke[1], Anubha Mahajan[13], Julian Maller[13,59], Alisa Manning[2], Taylor Maxwell[34], Alexander Mazur[60], Mark McCarthy[13,14,20], James Meigs[6,61], Byungju Min[41], Karen Mohlke[22], Andrew Morris[13,23], Solomon Musani[62], Yoshihiko Nagai[60], Maggie Ng[35], Dan Nicolae[16,33,12], Sohee Oh[41], Nicholette Palmer[35], Taesung Park[41], Toni Pollin[63], Inga Prokopenko[13,26], David Reich[2,4], Manuel Rivas[13], Laura Scott[1], Mark Seielstad[64], Yoon Shin Cho[65], E-Shyong Tai[48,40,66], Xueling Sim[1], Robert Sladek[60,67], Philip Smith[68], Ioanna Tachmazidou[31], Tanya Teslovich[1], Jason Torres[16,33], Vasily Trubetskoy[16,33], Sara Willems[69], Amy Williams[2,4], James Wilson[70], Steven Wiltshire[71], Sungho Won[72], Andrew Wood[15], Wang Xu[66], Yik Ying Teo[73,74,75,58,76], Joon Yoon[41], Jong-Young Lee[77], Matthew Zawistowski[1], Eleftheria Zeggini[31], Weihua Zhang[36], Sebastian Zöllner[1,78]

Affiliations

1. Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

2. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA

3. Department of Clinical Sciences, Diabetes and Endocrinology, Clinical Research Centre, Lund University, Malmö, Sweden

4. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

5. Center for Human Genetic Research and Diabetes Research Center (Diabetes Unit), Massachusetts General Hospital, Boston, MA, USA

6. Department of Medicine, Harvard Medical School, Boston, MA, USA

7. Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA

8. Department of Molecular Biology, Harvard Medical School, Boston, MA, USA

9. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

10. Institute of Genetic Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany

11. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

12. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

13. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

14. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

15. Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK

16. Department of Human Genetics, University of Chicago, Chicago, IL, USA

17. Department of Medicine, Helsinki University Hospital, Helsinki, Finland

18. Folkhälsan Research Center, Helsinki, Finland

19. Department of Pathology, University of Michigan, Ann Arbor, MI, USA

20. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Headington, UK

21. Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany

22. Department of Genetics, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA

23. Department of Biostatistics, University of Liverpool, Liverpool, UK

24. Department of Clinical Science, University of Bergen, Bergen, Norway

25. Department of Pediatrics, Haukeland University Hospital, Bergen, Norway

26. Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden

27. Kings College of London, Department of Twin Research and Genetic Epidemiology, London, UK

28. Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA

29. Department of Public Health, University of Helsinki, Helsinki, Finland

30. Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA

31. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

32. Department of Medicine, Department of Genetics, Albert Einstein College of Medicine, NY, USA

33. Department of Medicine, University of Chicago, Chicago, IL, USA

34. Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA

35. Center for Genomics and Personalized Medicine Research, Center for Diabetes Research, Department of Biochemistry, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

36. Department of Epidemiology and Biostatistics, Imperial College London, London, UK

37. Imperial College Healthcare NHS Trust, London, UK

38. Ealing Hospital National Health Service (NHS) Trust, Middlesex, UK

39. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

40. Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

41. Seoul National University, Seoul, South Korea

42. McGill Centre for Bioinformatics, McGill University, MontrÃľal, Quebec, Canada

43. Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore

44. Department of Epidemiology, Colorado School of Public Health, Aurora, CO, USA

45. Department of Textile Technology, Indian Institute of Technology Delhi, New Delhi, India

46. Department of Endocrinology, Hadassah Medical Center, Kiryat Hadassah, Jerusalem, Israel

47. Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL, USA

48. Duke-National University of Singapore Graduate Medical School, Singapore, Singapore

49. Department of Ophthalmology, National University of Singapore and National University Health System, Singapore, Singapore

50. Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands

51. The Biostatistics Center, George Washington University, Rockville, MD, USA

52. Department of Neurology, Konkuk University School of Medicine, Seoul, South Korea

53. Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan

54. Department of Health Studies, University of Chicago, Chicago, IL, USA

55. National Heart and Lung Institute (NHLI), Imperial College London, Hammersmith Hospital, London, UK

56. Department of Medicine, University of Eastern Finland, Kuopio Campus and Kuopio University Hospital, Kuopio, Finland

57. Division of Clinical Epidemiology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

58. Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore

59. Department of Statistics, University of Oxford, Oxford, UK

60. McGill University, Montréal, Québec, Canada

61. General Medicine Division, Massachusetts General Hospital, Boston, MA, USA

62. Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

63. Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

64. University of California San Francisco, San Francisco, CA, USA

65. Department of Biomedical Science, Hallym University, Chuncheon, Gangwon-do, South Korea

66. Department of Medicine, National University of Singapore, National University Health System, Singapore, Singapore

67. Department of Medicine, Royal Victoria Hospital, Montréal, Québec, Canada

68. National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, MD, USA

69. Department of Genetic Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands

70. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA

71. Centre for Medical Research, Western Australian Institute for Medical Research, The University of Western Australia, Nedlands, WA, Australia

72. Chung-Ang University, Seoul, South Korea

73. Department of Epidemiology and Public Health, National University of Singapore, Singapore, Singapore

74. Centre for Molecular Epidemiology, National University of Singapore, Singapore, Singapore

75. Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore

76. Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

77. Center for Genome Science, Korea National Institute of Health, Osong Health Technology Administration Complex, Chungcheongbuk-do, South Korea

78. Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

Selection of individuals and genes for initial sequencing experiment

For targeted sequencing, we drew from 27,500 individuals spanning several previously described prospective cohorts from Finland [Botnia] or Sweden [Malmo] [17–25]. The participants gave their written informed consent and the study protocol was approved by the Ethics Committees of Helsinki University Hospital, Finland, and Lund University. Individuals were ranked according to a liability model that measured risk for type 2 diabetes (T2D) as previously described [1], with cases selected to have low predicted risk for diabetes and controls selected to have high predicted risk for diabetes (Supplementary Table 3). The diagnosis of diabetes was based upon an oral glucose tolerance test (OGTT) applying ADA/WHO criteria for glucose. In the case of diabetic glucose values, the diagnosis was confirmed with a fasting glucose measurement or a new OGTT on a separate day.

A custom hybrid capture array was designed to target the 115 genes that either (i) lie within a genomic region associated with risk of T2D prior to 2008 (within 350kb of the single nucleotide polymorphism [SNP] with strongest reported association), or (ii) contain variants reported to cause monogenic forms of diabetes. A full list is shown in Supplementary Table 1.

Sequencing, variant calling, data QC, and annotation for initial sequencing experiment

DNA libraries were barcoded using the Illumina index read strategy and sequenced with an Illumina HiSeq2000. Reads were mapped to the human genome hg19 with the BWA algorithm [26] and processed with the Genome Analysis Toolkit (GATK) [27] to recalibrate base quality-scores and perform local realignment around known indels. Target coverage or each sample was also computed with the GATK. Single nucleotide variants (SNVs) and small insertions and deletions (indels) were called with the Unified Genotyper module of the GATK and filtered to remove SNVs with annotations indicative of technical artifacts (such as strand-bias, low variant call quality, or homopolymer runs). Samples with fewer than 76% of targeted bases covered to 20x, with an abnormally high number of non-reference alleles or heterozygosity, or with an abnormally low concordance with prior SNP array genotypes (based on the distribution across all samples) were excluded from analysis. Any sample genotype at a site with fewer than 10x coverage in the sample was ignored (e.g. set as missing). Variants were annotated with the Variant Effect Predictor [2].

Genotyping of p.Arg138* variant

From among the predicted protein truncating SNVs identified in the initial sequencing experiment, a subset were further genotyped in additional individuals (Supplementary Table 4) from Finland [Botnia] and Sweden [Malmo] (drawn from the same cohorts used for sequencing). Variants were genotyped on the Illumina HumanExome v1.1 array (Exome Chip; http://genome.sph.umich.edu/wiki/Exome_Chip_Design), which was designed to contain a near comprehensive catalog of non-synonymous variation identified from exome sequencing of thousands of individuals.

For genotyping of these cohorts, DNA samples were sent to the Broad Institute and prepared for genetic analysis with two quality control measures. First, DNA quantity was measured by Picogreen, and then all samples with sufficient total DNA and minimum concentrations for downstream activities were genotyped for a set of 24 SNPs using the Sequenom iPLEX Assay. These 24 validated markers include 1 gender assay and 23 SNPs located across the autosomes. The genotypes for these SNPs were used as a quality filter to advance samples, as well as a technical fingerprint validation (when applicable) for array genotypes.

All genotyping was performed at the Broad Institute Genetic Analysis Platform. DNA samples were placed on 96-well plates and genotyped using the Illumina HumanExome v1.1 SNP array. Genotypes were assigned using GenomeStudio v2010.3 with the calling algorithm/genotyping module version 1.8.4

using the custom cluster file HumanExomev1_1_CEPH_A.egt. Subsequent processing of genotype calling was done by zCall [28]. Samples with 2 or more discordant fingerprint genotypes and/or call rates below 97% were excluded from data analysis.

Individuals with call rates below 95% on Metabochip or Exome chip were excluded from analysis, as were SNVs with call rates below 80% or with extreme deviations from Hardy-Weinberg equilibrium. Principal component analysis was performed on a set of LD-independent SNPs using PLINK [29] and EIGENSTRAT [30].

Further genotyping of the p.Arg138* variant was performed in additional cohorts drawn from five countries of Europe (Supplementary Table 4). In three countries, phenotypic association was performed: Sweden (the PIVUS [31] and ULSAM [32] cohorts [PIVUS/ULSAM]), Denmark (the Inter99, Health2006/2008, Vejle Biobank, ADDITION, and Steno Diabetes Center cohorts [33] [Danish]), and Finland (the FUSION study [34], the DR's EXTRA study [35], the FIN-D2D survey [36], the FINRISK 2007 study [37], and the Metabolic Syndrome in Men (METSIM) study [38] [Finnish]). In two countries, frequency alone was estimated and phenotypic association was not performed: Poland (the Genomics Collaborative Initiative [39, 40]) and Britain (the WTCCC/UKT2D consortia [41, 42], the UK Adult Twin Registry [43], the Oxford Biobank [44], and DARTS [4]). Individuals from Poland were genotyped on a custom Sequenom assay, while the remaining individuals were genotyped on the Exome Array.

Sequencing and genotyping of Icelanders

Variants were genotyped in Icelanders, drawn from population-based projects conducted by deCODE genetics [deCODE], using a combination of whole-genome sequencing and imputation (Supplementary Tables 4,10). The study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland.

The Icelandic set of 3,548 T2D cases was recruited over a number of years. Samples were initially collected through a long-term epidemiological study and through screening for T2D amongst participants in deCODE genetics-funded family-based studies on cardiovascular and metabolic traits. Further collection included patients attending the Diabetes Clinic at Landspitali University Hospital, the main hospital in Iceland. Individuals with a discharge diagnosis of T2D from any inpatient ward at the hospital were also recruited; diagnosis of T2D was validated through careful revision of patient files. GWAS studies using these samples have been previously described [45]. Additional cases with clinical diagnosis of T2D were recruited through collaboration with other major diabetes clinics in Iceland. Clinical diagnosis of T2D in Iceland is based on fasting serum glucose $\geq$7.0 mmol/L on two separate occasions.

Measurements of blood pressure, serum glucose, fasting glucose, insulin, HDL, total cholesterol and triglycerides were performed at the Research Laboratory at Mjodd (RAM) and at the Landspitali University Hospital in the years 1990 to 2010. Fasting and non-fasting serum glucose levels, total cholesterol, triglycerides and HDL were measured with a Vitros analyzers and corresponding multilayer reagents (Ortho Clinicla Diagnostics, Rochester, NY, USA) and serum insulin levels were measured using an electrochemiluminescence immunoassay and Elecsys analyzers (Roche Diagnostics, Mannheim Germany). The LDL values were calculated from the HDL, triglyceride and total cholesterol values using the Friedewald equation, excluding individuals with triglyceride values > 4.52 mmol/L. BMI values, collected for studies on cardiovascular and metabolic diseases and cancer, were available for 74,735 Icelandic individuals, of which 23% were self-reported. Of those, 73,961 were included in the analysis. All measurements were adjusted for age at measurement and standardized to a standard normal distribution using quantile-quantile standardization for each sex separately. The blood pressure measurements were adjusted for year of birth and standardized to a standard normal distribution using quantile-quantile standardization.

Whole-genome sequence was performed on 2,230 samples selected for various conditions from the deCODE database, as previously described [46–51]. Briefly, samples were sequenced to at least 10x coverage on Illumina GAIIx and/or HiSeq 2000 instruments, following manufacturer's instructions. Reads

were aligned to NCBI Build 36 of the human reference sequence with the BWA algorithm and processed with the GATK to recalibrate base quality-scores and perform local realignment around known indels. SNVs and indels were identified, genotyped, and filtered with the GATK and annotated with SNP effect predictor (snpEff) [52].

All identified variants were imputed by either 'direct' imputation, via long-range phasing into 95,085 chip-typed Icelanders, or by genealogy based imputation, into an additional 294,082 Icelanders [46–51, 53]. The Icelandic chip-typed samples were assayed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics. Long range phasing of all chip-genotyped individuals was performed with methods described previously [46]; in brief, phasing is achieved using an iterative algorithm which phases a single proband at a time given the available phasing information about everyone else who shares a long haplotype identically by state with the proband. The second genealogy-based imputation step was applied to relatives of chip-typed individuals and has also been previously described [46]; in brief, allele counts are estimated for a proband's paternal and maternal haplotypes based on fully phased imputed and chip genotypes of all individuals within two meioses of the proband.

To validate the imputed genotypes for the p.Lys34Serfs*50 *SLC30A8* frameshift deletion, we genotyped 467 individuals, including 243 predicted carriers of the mutation, using Sanger sequencing. All but seven of the predicted carriers were confirmed as carriers of the mutation, and no new carriers were identified. The genotypes for the 467 individuals were combined with the genotypes of the 2,230 whole-genome sequenced individuals and the combined set of genotypes were used as a training set for re-imputation of the variant. These re-imputed genotypes were then used in all the association tests presented in the paper.

Genotyping of Norwegian individuals

Norwegian individuals (Supplementary Table 4) were drawn from the HUNT2 Study [HUNT2], a subset (aged $>$20 years) of an extensive population-based health survey (The Nord-Trøndelag Health Study) conducted in a Norwegian county with 127,000 inhabitants [54]. Diabetes was self-reported or identified by standard tests if random glucose was >8.0 mmol/l. Informed consent was obtained from all participants and the study was approved by the Regional Committee for Research Ethics and the Norwegian Data Inspectorate.

The *SLC30A8* p.Lys34Serfs*50 variant was genotyped by a fragment-length-based method using differentially fluorescently labeled primers, and genotypes were confirmed by Sanger sequencing in all positive samples.

Sequencing of individuals from additional ethnicities

The exons of *SLC30A8* were sequenced in 12,294 additional individuals as part of the whole-exome sequencing studies performed through the Genetics of Type 2 Diabetes (GoT2D) and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia. Individuals were selected spanning 5 ethnicities: European (the FUSION study [34] [FUSION], the METSIM study [38] [METSIM], KORA-gen [55] [KORA], the WTCCC/UKT2D consortium [41, 42] and the UK Adult Twin Registry [43] [UKT2D], as well as Ashkenazi individuals recruited from the metropolitan New York region [56, 57] [Ashkenazim] and small number of individuals from the Finnish [Botnia] and Swedish [Malmo] prospective cohorts used for the initial sequencing experiment [17–25]), African-American (the Jackson Heart Study (JHS) [58] cohort [JHS] as well as additional individuals recruited from North Carolina, South Carolina, Georgia, Tennessee, or Virginia [59] [WFS]), South Asian (the London Life Sciences Prospective Population Study (LOLIPOP) [60, 61] [LOLIPOP] and Singapore Indian Eye Study (SINDI) [62] [Singapore Indians]), East Asian (the Korean Association REsource (KARE) [63] [KARE] as well as the Singapore Dia-

betes Cohort Study (SDCS) and Singapore Prospective Study Program [64–67] [Singapore Chinese]), and Hispanic (the San Antonio Family Heart Study (FHS) [68], the San Antonio Family Diabetes/Gallbladder Study (SAFDGS) [69], the Veterans Administration Genetic Epidemiology Study (VAGES) [70], the Family Investigation of Nephropathy and Diabetes (FIND) [71], San Antonio component [San Antonio], and individuals from Starr County, TX [72] [Starr County]). Data generation and processing was performed in an identical fashion as for the initial sequencing experment, although target capture was performed with the Agilent SureSelect Human All Exon platform rather than a custom hybrid capture array.

Functional characterization of ZnT8 variants

Plasmids

Plasmids encoding C-terminal, V5-tagged human *SLC30A8* (NM_173851.2) variants were synthesized and subcloned into the pLX304 lentiviral vector by Genscript. The W325 and R325 plasmids express cDNAs encoding alternative alleles of the common missense SNP rs13266634. The p.Arg138* nonsense allele was engineered in two different ways: Arg138* expresses the full length R325 cDNA with the nonsense mutation at amino acid 138, disrupting downstream translation of the V5 tag; *138 encodes the truncated ORF generated by this variant (amino acids 1-137) in-frame with the V5 tag. The p.Lys34Serfs*50 frameshift allele, corresponds to the ORF produced by the 7 bp deletion. Control ORFs (BFP, GFP, HcRed, and Luciferase) in pLX304 lentiviral vector were obtained from the RNAi Consortium at the Broad Institute [73]. C-terminal, HA-tagged W325 was amplified by PCR and sublconed into the pLX301 lentiviral vector.

Cell culture

HeLa cells were cultured in Dulbecco's Modified Eagle's Medium supplemented with 10% heat-inactivated fetal calf serum and penicillin/streptomycin. Cells were maintained at 37° C in the presence of 5% CO_2.

Gene expression analyses

Expression levels of *SLC30A8* variant ORFs were determined using the NanoString nCounter system [74]. Seventy-one genes, including *SLC30A8*, V5-tagged genes, and housekeepers, were measured in RLT lysates from HeLa cells transiently expressing C-terminal, V5-tagged *SLC30A8* variants or V5-tagged control ORFs, as per manufacturer's instructions. Data was normalized in two steps. First, variation in sample processing was normalized using the spiked-in positive control probes provided by the nCounter system. Then, variation in input was normalized by median centering. The background level of non-specific binding was determined by calculating the mean + 2 standard deviations of the spiked-in negative control probes.

Immunocytochemistry

HeLa cells plated on collagen-coated 96-well plates were transiently transfected with expression plasmids encoding C-terminus, V5-tagged proteins and fixed with 4% paraformaldehyde 24 h post-transfection. Ins1e cells plated on 804G matrix-coated 96-well plates were transduced with lentivirus expressing C-terminus, V5-tagged proteins. Media was replaced 24 h post-infection and cells were fixed with 4% paraformaldehyde 48 h post-infection. Immunostaining was performed using antibodies against the V5 epitope (1:1500, Life Technologies) or N-terminus of ZnT8 (1:50, Abcam, ab105353). Alexa Fluor® 488 goat anti-mouse and Alexa Fluor® 594 goat anti-rabbit secondary antibodies were used. Nuclei were stained with Hoechst 33342. Images were captured using a Zeiss Cell Observer.

Inhibition of Protein Degradation

For studies involving inhibition of protein degradation, HeLa cells transiently expressing ZnT8 protein variants were changed into media containing inhibitors 24 h after transfection. Cells were treated for 4 h and then fixed with 4% paraformaldehyde for immunofluorescent staining or collected for Western Blot analysis. MG132 (10uM, Calbiochem) was used to inhibit proteasomal degradation; while chloroquine (100uM, Sigma Aldrich) was used as an inhibitor of lysosomal degradatory enzymes.

Western blots

HeLa cells plated on 6-well plates were transiently transfected with expression plasmids using Lipofectamine 2000 (Life Technologies). Cells were harvested 24 hr after transfection, pelleted via centrifugation, and snap frozen. Cell pellets were stored at -80 C for a minimum of 4 h prior to protein extraction. Thawed cell pellets were resuspended in RIPA buffer (50mM Tris pH 7.5, 150mM NaCl, 0.1% sodium dodecyl sulphate, 1% sodium deoxycholate, 0.2% Triton-X; all Sigma Aldrich) containing protease inhibitor (Complete Protease Inhibitor Cocktail, Roche). Protein concentration was determined using the BCA assay, according to manufacturer's guidelines (Pierce Thermo Fisher Scientific). Protein samples were denatured via heating at 42 C for 20 minutes in NuPAGE LDS Sample Buffer (Life Technologies) supplemented with 1% SDS and DTT. Following SDS-PAGE and transfer to nitrocellulose, membranes were blocked with 5% nonfat milk prior to overnight incubation with antibodies. The following antibodies were used: HRP-conjugated, mouse anti-V5 (46-0708 [R96125], Life Technologies), rabbit anti-HA (Abcam, ab9110), or rabbit anti-tubulin (sc-9104, Santa Cruz Biotechnology). HRP-conjugated secondary antibodies were purchased from Jackson ImmunoResearch (West Grove, PA). Membranes were developed using ECL Western Blotting Substrate (Pierce Thermo Fisher Scientific).

Statistical analysis

Statistical association with T2D was performed in a several staged procedure. Statistics were computed separately for p.Arg138*, p.Trp152*, p.Lys34Serfs*50, and the remaining variants, using methods to correct for sample structure (genetic relatedness and population stratification). Variant counts and frequencies were obtained through analysis of all individuals, including those genetically related to other individuals in the study.

Association analysis of p.Arg138* was computed separately for the Botnia, Malmo, PIVUS/ULSAM, Danish, and Finnish cohorts using a mixed linear model, as implemented by FaST-LMM [75] (for Botnia and Malmo) or EMMAX [76] (for the remaining cohorts). Further analysis of p.Arg138* in the additional sequenced individuals (from the Singapore Indians and UKT2D cohorts) was computed using the same procedure, although all sequenced individuals were included in the analysis (rather than being split by cohort). A mixed linear model accounts for different layers of sample structure, including population stratification and sample relatedness. A kinship matrix was first computed for each cohort separately using independent SNPs (MAF >1%) present on the Exome Array using either FaST-LMM or EMMAX; association p-values were then computed for p.Arg138*. As neither FaST-LMM nor EMMAX produce effect size estimates for dichotomous traits, point estimates for odds ratios were computed via a standard logistic regression as implemented in PLINK [29] with 10 principal components as covariates (computed via the EIGENSTRAT [77] software package from the same SNPs as for the kinship matrix). These were then transformed into 95% Wald confidence intervals using standard error estimates back-calculated from the p-values produced by the linear mixed model (alternative approaches of accounting for sample relatedness, such as exclusion of closely related individuals, produced similar estimates of statistical significance, albeit with high variability in effect size estimates depending on which specific individuals were excluded). The resulting six association statistics were combined via an inverse variance based fixed-effects meta-

analysis (as implemented in the METAL software package [78]) to obtain an estimated odds ratio and p-value for association of p.Arg138* with type 2 diabetes.

An association statistic for p.Trp152* was computed via the same methodology as for p.Arg138*, albeit restricted to the individuals from the Botnia cohort who were originally sequenced (the variant was not observed in other cohorts).

Association statistics were computed for p.Lys34Serfs*50 separately for Icelandic and Norwegian individuals. Icelandic individuals were analyzed using previously described methods [46–51]. For T2D, logistic regression was used, treating T2D as the response and expected genotype counts from imputation or allele counts from direct genotyping as covariates. Testing was performed using the likelihood ratio statistic. When testing for association, controls were matched to cases based on the informativeness of imputed genotypes, such that for each case C controls of matching informativeness were chosen, where C is chosen as the ratio of cases to controls in groups of individuals clustered based on mean informativeness values. For quantitative traits, a linear mixed model was used to test for association, with kinship matrix estimated from the Icelandic genealogical database. As it is not computationally feasible to use the full model to analyze all individuals, individuals were split into smaller clusters for the calculation. To account for relatedness and stratification within the case and control sample sets, genomic control [79] was applied. The inflation λ_g in the χ^2 statistic was estimated based on a subset of about 300,000 common variants and the p-values adjusted by dividing the corresponding χ^2 values by this factor. Association statistics for Norwegian samples were computed via a simple logistic regression, with statistical significance computed via the Score statistic [80] as implemented in the R programming language. The resulting two association statistics were combined via an inverse variance based fixed-effects meta-analysis (as implemented in the METAL software package [78]) to obtain an estimated odds ratio and p-value for association of p.Lys34Serfs*50 with type 2 diabetes.

Association analysis of the remaining nine variants were computed via an aggregate gene based test. Individuals were scored according to the presence of any of the nine variants (e.g., carriers assigned a 1 and non-carriers a 0). Analysis was then performed using a linear mixed-model, as for analysis of p.Arg138*, across all 12,294 sequenced individuals. A kinship matrix for the analysis was computed using independent SNPs (MAF >1%) observed across the entire exome sequencing experiment. Effect size and standard error estimates were obtained via the same procedure as for p.Arg138*.

The four association statistics (for p.Arg138*, p.Trp152*, p.Lys34Serfs*50, and the remaining nine variants), were then combined with a random-effects meta-analysis (via the METASOFT [81] software package). A fixed-effects meta-analysis produced similar (if slightly more significant) effect size and p-value estimates.

Quantitative traits were analyzed identically as for T2D. Traits were scaled into standardized units prior to analysis.

For association tests in the initial sequencing experiment (Supplementary Figure 2), as well as tests of the other variants genotyped in additional individuals (Supplementary Table 6), the analytical procedure was analogous: tests of individual variants were performed with a linear mixed model using the EMMAX software package, and tests of variants aggregated within genes were performed by first collapsing variants into a single indicator variable and then applying the the same mixed model approach. For tests in the initial sequencing experiment, all individuals were analyzed jointly. For tests in the additional genotyped individuals, analysis was performed separately for the Botnia and Malmo cohorts and then combined with a fixed-effects meta-analysis (in the manner described above for p.Arg138*).

Sequence read data for variants called from additional sequencing

Available as additional supplementary data are graphical representations (produced via the Integrative Genomics Viewer [82]) of the protein-truncating variants in *SLC30A8* called from the additional sequencing experiment. Each file corresponds to a variant (named according to Table 1) and shows read traces for

each individual genotyped as a carrier, with a non-carrier individual shown for reference.

A detailed legend for these files is available at http://www.broadinstitute.org/igv/AlignmentData. Briefly, for each file, the top panel shows the position of the variant on the chromosome, at two scales of resolution. The bottom panel shows the human reference sequence (colored based on nucleotide) and a cartoon of the *SLC30A8* transcript.

Each other panel corresponds to a sequenced individual; the top individual is genotyped as a non-carrier (homozygous reference) of the variant, while the remaining are genotyped as (heterozygous) carriers. Within each panel, the top row is a series of histograms representing depth of coverage and the balance of observed alleles for each genomic position; these are colored according to the observed alleles (using the same scheme as for the human reference sequence), with gray representing reference alleles. Below this row is a representation of the read data: each gray bar is an individual read, with strand indicated by the arrow at the end of the bar. Positions in the read that differ from the reference sequence are colored; deletions are shown as horizontal black lines and insertions as vertical purple lines. Reads are colored other than gray if they do not map to the genome as expected. Two vertical dashed lines indicate the position at which the relevant variant was called. Not all reads are shown due to limited vertical space; thus the coverage information at the top of the panel reflects the full distribution of reads, which may not all be visible.

References

[1] Guey, L. T. *et al.* Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genetic epidemiology* (2011).

[2] McLaren, W. *et al.* Deriving the consequences of genomic variants with the ensembl api and snp effect predictor. *Bioinformatics* **26**, 2069–70 (2010).

[3] Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

[4] Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).

[5] Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).

[6] Strawbridge, R. J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–34 (2011).

[7] Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).

[8] Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).

[9] Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

[10] Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).

[11] Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).

[12] Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat. Genet.* **42**, 376–384 (2010).

[13] Boger, C. A. *et al.* CUBN is a gene locus for albuminuria. *J. Am. Soc. Nephrol.* **22**, 555–570 (2011).

[14] Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

[15] Ehret, G. B. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).

[16] Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

[17] Groop, L. *et al.* Metabolic consequences of a family history of niddm (the botnia study): evidence for sex-specific parental effects. *Diabetes* **45**, 1585–93 (1996).

[18] Forsen, T. *et al.* The diabetes registry in vaasa hospital district - direva. *The Journal of General Practitioner* **6**, 17–21 (2011).

[19] Lindholm, E., Agardh, E., Tuomi, T., Groop, L. & Agardh, C. D. Classifying diabetes according to the new who clinical stages. *European journal of epidemiology* **17**, 983–9 (2001).

[20] Parker, A. *et al.* A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. *Diabetes* **50**, 675–80 (2001).

[21] Berglund, G., Elmstahl, S., Janzon, L. & Larsson, S. A. The malmo diet and cancer study. design and feasibility. *Journal of internal medicine* **233**, 45–51 (1993).

[22] Berglund, G. *et al.* Long-term outcome of the malmo preventive project: mortality and cardiovascular morbidity. *Journal of internal medicine* **247**, 19–29 (2000).

[23] Lyssenko, V. *et al.* Clinical risk factors, dna variants, and the development of type 2 diabetes. *The New England journal of medicine* **359**, 2220–32 (2008).

[24] Isomaa, B. *et al.* A family history of diabetes is associated with reduced physical fitness in the prevalence, prediction and prevention of diabetes (ppp)-botnia study. *Diabetologia* **53**, 1709–13 (2010).

[25] Bog-Hansen, E. *et al.* Risk factor clustering in patients with hypertension and non-insulin-dependent diabetes mellitus. the skaraborg hypertension project. *Journal of internal medicine* **243**, 223–32 (1998).

[26] Li, H. & Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).

[27] DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

[28] Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: Genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).

[29] Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–75 (2007).

[30] Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–9 (2006).

[31] Lind, L., Fors, N., Hall, J., Marttala, K. & Stenborg, A. A comparison of three different methods to evaluate endothelium-dependent vasodilation in the elderly: the Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS) study. *Arterioscler. Thromb. Vasc. Biol.* **25**, 2368–2375 (2005).

[32] Hedstrand, H. A study of middle-aged men with particular reference to risk factors for cardiovascular disease. *Ups J Med Sci Suppl* **19**, 1–61 (1975).

[33] Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298–310 (2013).

[34] Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).

[35] Kouki, R. *et al.* Diet, fitness and metabolic syndrome–the DR's EXTRA study. *Nutr Metab Cardiovasc Dis* **22**, 553–560 (2012).

[36] Kotronen, A. *et al.* Non-alcoholic and alcoholic fatty liver disease - two diseases of affluence associated with the metabolic syndrome and type 2 diabetes: the FIN-D2D survey. *BMC Public Health* **10**, 237 (2010).

[37] Coletta, D. K. *et al.* Thirty-five-year trends in cardiovascular risk factors in Finland. *Int J Epidemiol* **39**, 504–518 (2009).

[38] Stancakova, A. *et al.* Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* **58**, 1212–1221 (2009).

[39] Winckler, W. *et al.* Association of common variation in the hnf1alpha gene region with risk of type 2 diabetes. *Diabetes* **54**, 2336–42 (2005).

[40] Florez, J. C. *et al.* Association testing in 9,000 people fails to confirm the association of the insulin receptor substrate-1 g972r polymorphism with type 2 diabetes. *Diabetes* **53**, 3313–8 (2004).

[41] Consortium, W. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–78 (2007).

[42] Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).

[43] Spector, T. D. & Williams, F. M. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* **9**, 899–906 (2006).

[44] Tan, G. D. *et al.* The in vivo effects of the Pro12Ala PPARgamma2 polymorphism on adipose tissue NEFA metabolism: the first use of the Oxford Biobank. *Diabetologia* **49**, 158–168 (2006).

[45] Steinthorsdottir, V. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).

[46] Gudmundsson, J. *et al.* Discovery of common variants associated with low tsh levels and thyroid cancer risk. *Nature Genetics* **44**, 319–22 (2012).

[47] Jonsson, T. *et al.* A mutation in APP protects against alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96–9 (2012).

[48] Sulem, P. *et al.* Sequence variants at cyp1a1-cyp1a2 and ahr associate with coffee consumption. *Human molecular genetics* **20**, 2071–7 (2011).

[49] Stacey, S. N. *et al.* A germline variant in the tp53 polyadenylation signal confers cancer susceptibility. *Nature Genetics* **43**, 1098–103 (2011).

[50] Rafnar, T. *et al.* Mutations in brip1 confer high risk of ovarian cancer. *Nature Genetics* **43**, 1104–7 (2011).

[51] Holm, H. *et al.* A rare variant in myh6 is associated with high risk of sick sinus syndrome. *Nature Genetics* **43**, 316–20 (2011).

[52] Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).

[53] Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).

[54] Johansson, S. *et al.* Studies in 3,523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 alpha (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. *Diabetes* **56**, 3112–3117 (2007).

[55] Wichmann, H. E., Gieger, C. & Illig, T. KORA-gen–resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, 26–30 (2005).

[56] Permutt, M. A. *et al.* A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* **50**, 681–685 (2001).

[57] Atzmon, G. *et al.* Evolution in health and medicine Sackler colloquium: Genetic variation in human telomerase is associated with telomere length in Ashkenazi centenarians. *Proc. Natl. Acad. Sci. U.S.A.* **107 Suppl 1**, 1710–1717 (2010).

[58] Taylor, H. A. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis* **15**, S6–4 (2005).

[59] Yu, H., Bowden, D. W., Spray, B. J., Rich, S. S. & Freedman, B. I. Linkage analysis between loci in the renin-angiotensin axis and end-stage renal disease in African Americans. *J. Am. Soc. Nephrol.* **7**, 2559–2564 (1996).

[60] Chahal, N. S. *et al.* Ethnicity-related differences in left ventricular function, structure and geometry: a population study of UK Indian Asian and European white subjects. *Heart* **96**, 466–471 (2010).

[61] Chahal, N. S. *et al.* Does subclinical atherosclerosis burden identify the increased risk of cardiovascular disease mortality among United Kingdom Indian Asians? A population study. *Am. Heart J.* **162**, 460–466 (2011).

[62] Lavanya, R. *et al.* Methodology of the Singapore Indian Chinese Cohort (SICC) eye study: quantifying ethnic variations in the epidemiology of eye diseases in Asians. *Ophthalmic Epidemiol* **16**, 325–336 (2009).

[63] Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527–534 (2009).

[64] Hughes, K. *et al.* Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor levels. *J Epidemiol Community Health* **44**, 29–35 (1990).

[65] Tan, C. E., Emmanuel, S. C., Tan, B. Y. & Jacob, E. Prevalence of diabetes and ethnic differences in cardiovascular risk factors. The 1992 Singapore National Health Survey. *Diabetes Care* **22**, 241–247 (1999).

[66] Hughes, K., Aw, T. C., Kuperan, P. & Choo, M. Central obesity, insulin resistance, syndrome X, lipoprotein(a), and cardiovascular risk in Indians, Malays, and Chinese in Singapore. *J Epidemiol Community Health* **51**, 394–399 (1997).

[67] Cutter, J., Tan, B. Y. & Chew, S. K. Levels of cardiovascular disease risk factors in Singapore following a national intervention programme. *Bull. World Health Organ.* **79**, 908–915 (2001).

[68] Mitchell, B. D. *et al.* Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. *Circulation* **94**, 2159–2170 (1996).

[69] Hunt, K. J. *et al.* Genome-wide linkage analyses of type 2 diabetes in Mexican Americans: the San Antonio Family Diabetes/Gallbladder Study. *Diabetes* **54**, 2655–2662 (2005).

[70] Coletta, D. K. *et al.* Genome-wide linkage scan for genes influencing plasma triglyceride levels in the Veterans Administration Genetic Epidemiology Study. *Diabetes* **58**, 279–284 (2009).

[71] Knowler, W. C. *et al.* The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. *J. Diabetes Complicat.* **19**, 1–9 (2005).

[72] Hanis, C. L. *et al.* Diabetes among Mexican Americans in Starr County, Texas. *Am. J. Epidemiol.* **118**, 659–672 (1983).

[73] Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **8**, 659–661 (2011).

[74] Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).

[75] Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).

[76] Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).

[77] Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–9 (2006).

[78] Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

[79] Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155–166 (2001).

[80] Lin, D. Y. & Tang, Z. Z. A general framework for detecting disease associations with rare variants in sequencing studies. *Am. J. Hum. Genet.* **89**, 354–367 (2011).

[81] Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).

[82] Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).