SUPPLEMENTARY DATA

Recessive mutations in *PCBD1* cause a new type of early-onset diabetes

Methods for Xenopus and mouse experiments

Xenopus embryo manipulation

Microinjections and dissections were performed as described (7). Morpholino antisense oligonucleotide targeting the first intron-exon boundary of the Xenopus pcbd1 pre-mRNA was designed and purchased from GeneTools LLC. The sequence of the splice-blocker morpholino is: 5'-gcgctttccctgccatgtctgctga-3'. For anterior endodermal (AE) explants, dorsal vegetal cells were microdissected at stage 10 and cultured in 0.5x MMR in until desired stage. RNA extraction and RT-qPCR on Xenopus explants was performed as described (7). Standard number of qPCR cycles (40 cycles) was performed using SYBR Green Master Mix (Roche) on a StepOnePlus^{TM} system (Applied Biosystems). Ornithine Decarboxylase (ODC) was used as reference gene in Xenopus RT-qPCR analysis. All the values were normalized to the reference gene and calculated using the delta CT method. Data were determined in triplets. All experiments were repeated at least three times, unless otherwise stated.

Immunofluorescence and In situ Hybridization

For immunofluorescence on mouse cryosections mouse embryos were fixed in 4% paraformaldehyde at 4°C from 2 hours to overnight. Subsequently, samples were equilibrated in 20% sucrose solution and embedded in OCT compound (Sakura). Cryosections (10µm) were incubated with TSA (Perkin Elmer) blocking buffer 1hr at RT and afterwards with primary antibodies at appropriate dilution. Primary antibodies used were: anti-Pdx1 (abcam, ab47308); anti-Insulin (GeneTex, GTX27842); anti-E-cadherin (Invitrogen, 13-1900) and anti-Pcbd1 (Atlas Antibodies, HPA037575). AlexaFluor-conjugated secondary antibodies (Invitrogen) were used at a 1:750 dilution. All confocal images were acquired with an LSM 700 confocal laser-scanning microscope.

Whole-mount in situ hybridization was performed according to (7). Stained embryos were embedded in gelatin and sectioned using a vibratome (Leica VT1000S).

SUPPLEMENTARY DATA

Supplementary Table 1. Statistics of whole-genome sequencing performance of family 1.

Sample	Fully called genomic positions, %	% of fully called genomic positions covered at ≥10x	Total genomic variants	Fully called exonic positions, %	% of fully called exonic positions covered at ≥10x	Total exonic variants
I-I	95.6	96.0	4,091,829	98.2	98.6	23,500
I-II	95.7	96.6	4,141,813	98.1	98.4	23,915
II-III	95.8	96.4	4,111,511	98.4	98.8	23,624
II-II	95.8	96.8	4,138,100	98.2	98.6	23,760
III-II	96.0	97.4	4,049,355	98.4	98.9	23,282

Supplementary Table 2. PCR primers used to amplify *PCBD1* exons for Sanger sequencing.

Primer name	Primer sequence	Exon	Annealing temperature
PCBD1_1bF	gatggtctcacgagggaaaca	1	60
PCBD1_1aR	gcaggggactcgaaaagact		
PCBD1_2F	cccagcctattgctcaaaga	2	58
PCBD1_2R	ctggatgagtgtggtgtctga		
PCBD1_3F	aggatgtcaaggggaaatg	3	58
PCBD1_3R	aggcatgtgcaatctcagtg		
PCBD1_4F	actggccagctgctattctg	4	58
PCBD1_4R	ttggattgacctgtggaaaag		

Supplementary Table 3. *Xenopus* primers used for SyberGreen RT-qPCR.

Gene	Forward primer	Reverse primer
odc	ttcgggtgattccttgccac	gccattgtgaagactctctccaatc
pcbd1	catgacaagggtggctctcc	ctacatcactattggtaatgtgtttcct
pcbd2	ctttaaccaggcatttggatt	gtgtagtcagagttatctgaac
hnf1a	ccatggcaaaacttatggatttaga	ggagatggggtactctgactg
hnf1b	gaagaaagagaagctttagtgg	gactatatctcagcccttgc
pdx1	gttccctcagctgcttatcg	taccaaggggttgctgtagg
ptf1a	atggaaacggtcctggagca	gaggatgagaaggagaagttg
insulin	aggcttcttctactaccctaag	acaatccccctcttcattt
sox9a	caactaattgcgcactgggg	tcttcagcaaaggcacccaa
foxa2	taccaacatcaactccatgagc	gtaacttcgccggtaagttttg
prox1	ctgatatctcaccttattcgg	tgggaggtgatgcatctgttg
hex	cctttccgcttgtgcagagg	aacagcgcatctaatgggac
fibrinogen	aagatgactcagtgggcagc	ttcaatgccgccttctcctt

SUPPLEMENTARY DATA

Supplementary Table 4. Final set of homozygous mutations from whole-genome sequencing dataset of family 1 after selection of novel, protein affecting, conserved and damaging variants.

Gene	Chr	Start	End	Reference	Observed	SIFT*	PP2†	MT‡	Linkage	T1DB§	HI‖	Vpa8.5¶	Dpa8.5#	Rank
RECK	9	36091226	36091226	G	A	T	P	D	Yes	Enriched	1.5	9.4	14.2	4
GDF10	10	48426745	48426745	T	C	D	D	D	Yes	Low	0.8	2.1	0.4	5
OGDHL	10	50953532	50953532	C	T	D	D	D	Yes	Moderate	0.5, 1.2, 4.5	16.9	15.7	3
MYPN	10	69955261	69955261	C	T	D	D	D	Yes	Low	0.6	NE**	NE**	6
PCBD1	10	72645644	72645644	G	-	D	NA	D	Yes	Enriched	47.2	72.0	40.6	1
CTSZ	20	57571717	57571717	G	A	D	D	D	Yes	Moderate	152.9	64.7	96.9	2

*SIFT – SIFT prediction (T – tolerant, D - deleterious)
†PP2 – Polyphen-2 prediction (P – possibly damaging, D – probably damaging, NA – not available)
‡MT – MutationTaster prediction (D – disease-causing)
§T1DB – expression in human islets present T1DBase (array data)
‖HI – RNAseq RPKM (Reads Per Kilobase of transcript per Million mapped reads) values in human islets (5)
¶Vpa8.5 – RNAseq FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values in mouse ventral pancreatic bud at E8.5 (6)
#Dpa – 8.5 RNAseq FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values in mouse dorsal pancreatic bud at E8.5 (6)
**NE – not expressed

SUPPLEMENTARY DATA

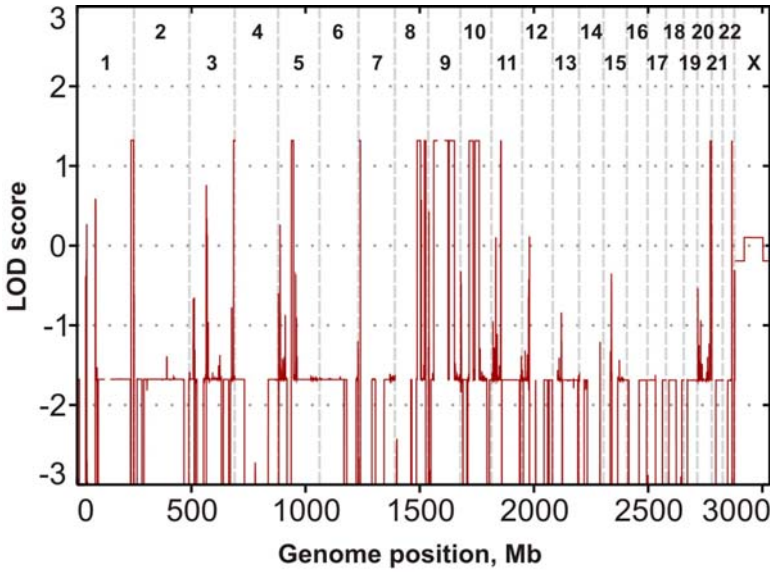
Supplementary Table 5. Characteristics of family members with heterozygous *PCBD1* mutations.

Family	Relationship	Protein alteration	Present age, years	Ethnicity	Diabetes				Weight range*(BMI, kg/m²)
					Diabetes, type	Onset, years	Recent HbA1c, % (mmol/mol)	Treatment	
1	Father	p.[(Leu16Cysfs*5)];[(wt)]	49	Turkish	Yes, T2D	45	6.2 (44.3)	Lifestyle	Obese (31.7)
	Mother	p.[(Leu16Cysfs*5)];[(wt)]	48	Turkish	Yes, T2D	40	6.9 (51.9)	Metformin	Overweight (29.6)
	Maternal grandmother	p.[(Leu16Cysfs*5)];[(wt)]	69	Turkish	Yes, T2D	48	6.9 (51.9)	Metformin, then SU	Overweight (28.5)
2	Father	p.[(Glu97Lys)];[(wt)]	47	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	48	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
3	Father	p.[(Glu87*)];[(wt)]	Died at 50	Ashkenazi Jewish	Yes, T2D	35	N.A.	Metformin, then SU	Obese (>30)
	Mother	p.[(Glu87*)];[(wt)]	55	Ashkenazi Jewish	No	N.A.	N.A.	N.A.	Normal (< 25)
4	Father	p.[(Gln98*)];[(wt)]	36	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
5	Father	p.[(Glu27*;Asp88Gln)];[(wt;wt)]	45	Turkish	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Glu27*;Asp88Gln)];[(wt;wt)]	40	Turkish	No	N.A.	N.A.	N.A.	Normal (< 25)
6	Father	p.[(Gln98*)];[(wt)]	36	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
	Mother	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Normal (< 25)
7	Father	p.[(Gln98*)];[(wt)]	34	Caucasian	No	N.A.	N.A.	N.A.	Overweight (29.1)
	Mother	p.[(Asn71del)];[(wt)]	32	Caucasian	No	N.A.	N.A.	N.A.	Obese (31.2)

BMI – body mass index (kg/m²), HbAc1 – glycated hemoglobin, T2D – type 2 diabetes mellitus, SU – sulfonylurea, N.A. – not available
*Weight range (Normal – BMI < 25 kg/m², Overweight – BMI 25-30 kg/m², Obese – BMI > 30 kg/m²)

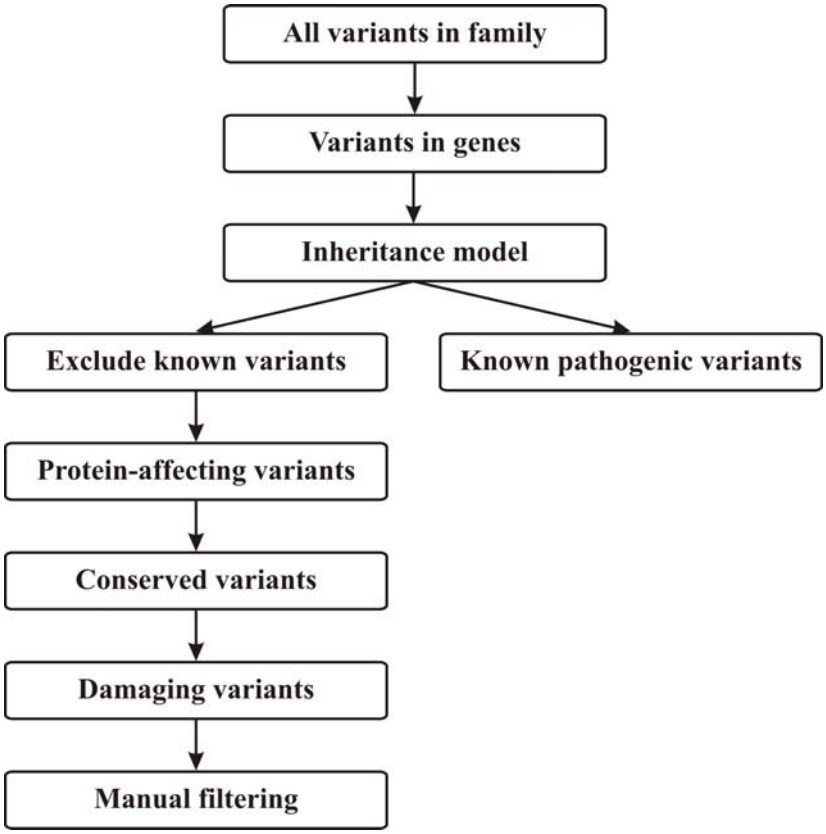
SUPPLEMENTARY DATA

Supplementary Figure 1. Linkage analysis of family 1. 24 genomic regions have positive LOD score, including 13 regions reaching maximal expected LOD score of 1.3, in case of perfect recombination.



SUPPLEMENTARY DATA

Supplementary Figure 2. Filtering strategy employed to reduce all variants of family 1 down to a single candidate gene.



SUPPLEMENTARY DATA

Web resources

MERLIN http://www.sph.umich.edu/csg/abecasis/merlin/index.html
ANNOVAR http://www.openbioinformatics.org/annovar/
Human Genome Reference http://www.ncbi.nlm.nih.gov/refseq/
CGA Tools http://cgatools.sourceforge.net
dbSNP http://www.ncbi.nlm.nih.gov/SNP
NHLBI Exome Sequencing Project (ESP) Exome Variant Server https://esp.gs.washington.edu/EVS
1000 Genomes Project www.1000genomes.org
69 sequenced individuals by Complete Genomics www.completegenomics.com/public-data/69-Genomes
PhastCons scores http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/phastCons46way
SIFT http://sift.jcvi.org
PolyPhen-2 http://genetics.bwh.harvard.edu/pph2
MutationTaster www.mutationtaster.org
T1DBase http://www.t1dbase.org
BIODEF http://www.biopku.org/biodef
The Human Gene Mutation Database (HGMD) http://www.hgmd.org