ESM Table 1: Analysis of *RBPJL* gene region for association with T2D in American Indians

Genetic Variants			Type 2 Diabetes Association					
			Full-Pima (n = 3,719)		Non-Full-Pima (n = 3,940)		Combined (N = 7,659)	
Variant ID	Alleles[1]	Risk allele/ Freq[2]	OR (95%CI)[3]	*P*-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]
rs76474829	C/G	G, 0.03	1.49(1.04-2.14)	**0.03**	1.69(1.17-2.45)	**0.005**	1.57(1.19-2.07)	**0.001**
chr20:43945378	G/A	A, 0.01	1.08(0.69-1.69)	0.73	2.74(1.46-5.13)	0.002	1.43(0.97-2.12)	0.07
rs985586	G/A	A, 0.09	1.11(0.86-1.44)	0.41	1.18(0.94-1.47)	0.15	1.15(0.96-1.38)	0.12
rs138228405	C/G	C, 0.91	1.14(0.92-1.41)	0.23	1.13(0.88-1.45)	0.34	1.13(0.95-1.34)	0.17
rs143311170	A/G	A, 0.89	1.22(1.01-1.47)	0.04	0.94(0.74-1.18)	0.59	1.11(0.95-1.29)	0.21
rs147593522	A/G	A, 0.87	1.09(0.91-1.31)	0.35	1.06(0.84-1.34)	0.61	1.08(0.93-1.26)	0.29
rs2076026	T/C	C, 0.34	1.04(0.84-1.19)	0.60	1.05(0.90-1.23)	0.53	1.04(0.93-1.16)	0.49
rs2425696	G/A	A, 0.85	1.06(0.88-1.28)	0.53	1.03(0.85-1.24)	0.79	1.05(0.91-1.21)	0.52
rs33961254	AG/-	-, 0.18	1.01(0.86-1.20)	0.89	1.10(0.91-1.33)	0.32	1.04(0.92-1.19)	0.52
rs2072792	T/C	T, 0.24	0.99(0.85-1.16)	0.92	1.09(0.93-1.27)	0.29	1.04(0.92-1.17)	0.55
rs2076027	T/C	T, 0.81	1.03(0.88-1.2)	0.70	1.04(0.87-1.25)	0.65	1.03(0.91-1.17)	0.62
rs148973596	T/-	T, 0.77	1.02(0.88-1.18)	0.84	1.05(0.88-1.25)	0.60	1.03(0.92-1.16)	0.64
rs11698812	A/C	C, 0.29	0.96(0.83-1.11)	0.56	1.10(0.94-1.28)	0.22	1.02(0.92-1.14)	0.73
rs59339622	G/A	A, 0.44	0.98(0.86-1.11)	0.74	1.09(0.94-1.26)	0.26	1.02(0.92-1.13)	0.75
chr20:43932165	G/A	G, 0.95	1.03(0.79-1.33)	0.81	1.01(0.71-1.45)	0.94	1.02(0.82-1.28)	0.86
rs79312216	T/C	C, 0.26	1.03(0.90-1.19)	0.64	0.97(0.83-1.14)	0.70	1.01(0.90-1.12)	0.88
rs720063	T/G	T, 0.77	1.04(0.89-1.21)	0.64	0.97(0.82-1.14)	0.70	1.01(0.89-1.14)	0.90

Whole-genome sequencing in 335 Pima Indians identified 168 non-coding variants with a MAF ≥0.05 and 2 non-synonymous variants with a MAF ≥0.01 in the *RBPJL* gene region. Among these 170 variants, 162 (95.3%) were captured by 17 tag SNPs with an r^2 >0.8. [1]Alleles are reference allele/variant allele. [2]Risk allele and risk allele frequency. [3]OR and *P*-values were adjusted for age, sex, birth-year and first five principal components from a Pima Indian specific GWAS. OR are given per copy of the risk allele. [4]*P*-values shown are after controlling for inflation by genomic control method. All annotations are based on genome-build 37 (hg19). Variant highlighted in bold meets experiment-wide Bonferroni corrected significance ($P \leq 0.001$) and is in linkage disequilibrium (LD) with a missense variant (rs200998587:C>T) in *RBPJL*.

ESM Table 2: Analysis of *RBPJ* gene region for association with T2D in American Indians.

Genetic Variants			Type 2 Diabetes Association					
			Full-Pima (n = 3,719)		Non-Full-Pima (n = 3,940)		Combined (N = 7,659)	
Variant ID	Alleles[1]	Risk allele/freq[2]	OR (95%CI)[3]	*P*-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]
rs186035024	A/C	C, 0.11	1.20(1-1.45)	0.04	1.23(0.96-1.56)	0.10	1.22(1.04-1.43)	0.01
rs186895314	G/A	A, 0.21	1.13(0.98-1.31)	0.10	1.2(1-1.43)	0.05	1.15(1.02-1.3)	0.02
rs114530054	C/T	C, 0.95	1.18(0.88-1.56)	0.27	1.23(0.87-1.75)	0.23	1.20(0.95-1.52)	0.12
rs13144326	C/T	T, 0.78	1.1(0.94-1.29)	0.25	1.1(0.93-1.3)	0.27	1.1(0.97-1.24)	0.14
rs12639629	G/C	G, 0.40	1.04(0.91-1.18)	0.58	1.12(0.97-1.30)	0.11	1.08(0.97-1.19)	0.17
rs2788865	G/A	A, 0.87	1.17(0.96-1.44)	0.12	1.03(0.84-1.26)	0.81	1.1(0.95-1.28)	0.21
chr4:26455229	G/C	C, 0.08	1.16(0.93-1.44)	0.19	1.06(0.81-1.4)	0.67	1.12(0.94-1.34)	0.21
rs7655165	G/A	A, 0.82	1.14(0.95-1.36)	0.15	1.04(0.87-1.23)	0.70	1.09(0.95-1.24)	0.23
rs73245775	G/A	G, 0.79	1.09(0.93-1.27)	0.29	1.08(0.90-1.28)	0.45	1.08(0.95-1.22)	0.24
rs78445835	T/G	G, 0.04	1.20(0.88-1.64)	0.25	1.08(0.74-1.56)	0.70	1.15(0.89-1.47)	0.29
rs78672655	T/C	T, 0.93	1.15(0.92-1.45)	0.22	1(0.74-1.35)	0.99	1.1(0.91-1.33)	0.34
rs79533922	G/A	G, 0.92	1.15(0.90-1.47)	0.25	1.05(0.83-1.32)	0.69	1.09(0.91-1.30)	0.34
rs12650452	C/T	T, 0.72	1.05(0.9-1.22)	0.52	1.02(0.87-1.19)	0.83	1.03(0.92-1.16)	0.59
rs77056130	-/A	-, 0.90	1.03(0.84-1.27)	0.75	1.02(0.8-1.3)	0.87	1.03(0.88-1.22)	0.69
rs13116206	G/A	A, 0.84	1.05(0.87-1.27)	0.59	1(0.83-1.21)	0.99	1.03(0.89-1.18)	0.70
rs6853254	T/G	G, 0.90	1.09(0.86-1.37)	0.5	0.96(0.76-1.22)	0.73	1.03(0.86-1.22)	0.76
rs185848565	G/A	G, 0.90	0.97(0.79-1.19)	0.79	1.14(0.88-1.47)	0.33	1.02(0.87-1.20)	0.78

Whole-genome sequencing in 335 Pima Indians identified 212 variants with MAF >0.05 in the *RBPJ* gene region. Among these variants, 207 (97.6%) were captured by 17 tag SNPs with an r^2 >0.8. [1]Alleles are reference allele/variant allele. [2]Risk allele and frequency of risk allele. [3]OR and *P*-values were adjusted for age, sex, birth-year and first five principal components from a Pima Indian specific GWAS. OR are given per copy of the risk allele. [4]*P*-values shown are after controlling for inflation by genomic control method. All annotations are based on genome-build 37 (hg19).

ESM Table3: Analysis of *PTF1A* gene region for association with T2D in American Indians

Genetic Variants			Type 2 Diabetes Association					
			Full-Pima (n = 3,719)		Non-Full-Pima (n = 3,940)		Combined (N = 7,659)	
Variant ID	Alleles[1]	Risk allele/freq[2]	OR (95%CI)[3]	P-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]	OR (95%CI)[3]	*P*-value[3,4]
rs7916519	G/A	A, 0.19	1.15(0.97-1.38)	0.11	1.23(1.04-1.45)	0.02	1.19(1.05-1.36)	0.008
rs16923098	A/C	C, 0.12	1.18(0.96-1.43)	0.12	1.23(1-1.52)	0.05	1.2(1.03-1.39)	0.02
rs10741021	A/G	G, 0.67	1.1(0.96-1.25)	0.18	1.12(0.96-1.32)	0.15	1.11(0.99-1.23)	0.06
rs7090683	G/A	A, 0.43	1.04(0.92-1.18)	0.53	1.15(1-1.32)	0.06	1.09(0.98-1.2)	0.1
rs116846325	C/A	C, 0.76	1.11(0.96-1.27)	0.17	1.06(0.88-1.27)	0.54	1.09(0.97-1.23)	0.14
rs4748844	T/G	T, 0.61	1.04(0.91-1.18)	0.55	1.1(0.95-1.27)	0.19	1.07(0.96-1.18)	0.21
rs4454613	A/C	C, 0.54	1.05(0.93-1.2)	0.42	1.08(0.93-1.23)	0.32	1.06(0.96-1.18)	0.22
rs7904665	T/C	C, 0.91	1.01(0.78-1.33)	0.92	1.19(0.94-1.52)	0.15	1.11(0.92-1.35)	0.28
rs187391034	G/A	G, 0.94	1.03(0.81-1.32)	0.81	1.15(0.84-1.56)	0.39	1.08(0.88-1.32)	0.49
chr10:23522489	C/T	C, 0.96	1.05(0.78-1.41)	0.75	0.91(0.56-1.47)	0.7	1.02(0.78-1.32)	0.9

Whole-genome sequencing in 335 Pima Indians identified 30 variants with MAF>0.05 in the *PTF1A* gene region. Among these variants, 26 (86.6%) were captured by 10 tag SNPs with an r^2 >0.8. [1]Alleles are reference allele/variant allele. [2]Risk allele and frequency of risk allele. [3]OR and *P*-values were adjusted for age, sex, birth-year and first five principal components from a Pima Indian specific GWAS. OR are given per copy of the risk allele. [4]*P*-values shown are after controlling for inflation by genomic control method. All annotations are based on genome-build 37 (hg19).

ESM Table 4: Missense variants in *RBPJL* and *PTF1A* identified in whole-genome sequence data

Variant Information[1]			Allele Frequency[2]					Functional Prediction[4]		
Variant ID/Gene	Variant	AA change	Africans	Americans	Asians	Caucasians	Pimas[3]	PROVEAN	SIFT	Polyphen-2
rs200998587/*RBPJL*	**c.839C>T**	**p.(Thr280Met)**	**0.00**	**0.01**	**0.00**	**0.00**	**0.03**	**Deleterious**	**Damaging**	**Probably Damaging**
chr20:43945378/*RBPJL*	c.1333G>A	p.(Asp445Asn)	-	-	-	-	0.02	Neutral	Tolerated	Possibly Damaging
chr20:43940497/*RBPJL*	c.347C>T	p.(Thr116Met)	-	-	-	-	0.002	Neutral	Tolerated	Possibly Damaging
chr10:23481878/*PTF1A*	c.419G>A	p.(Gly140Asp)	-	-	-	-	0.002	Neutral	Damaging	Benign
rs7918487/*PTF1A*	c.787T>C	p.(Ser263Pro)	0.60	0.62	0.85	0.47	0.93	Neutral	Damaging	Benign

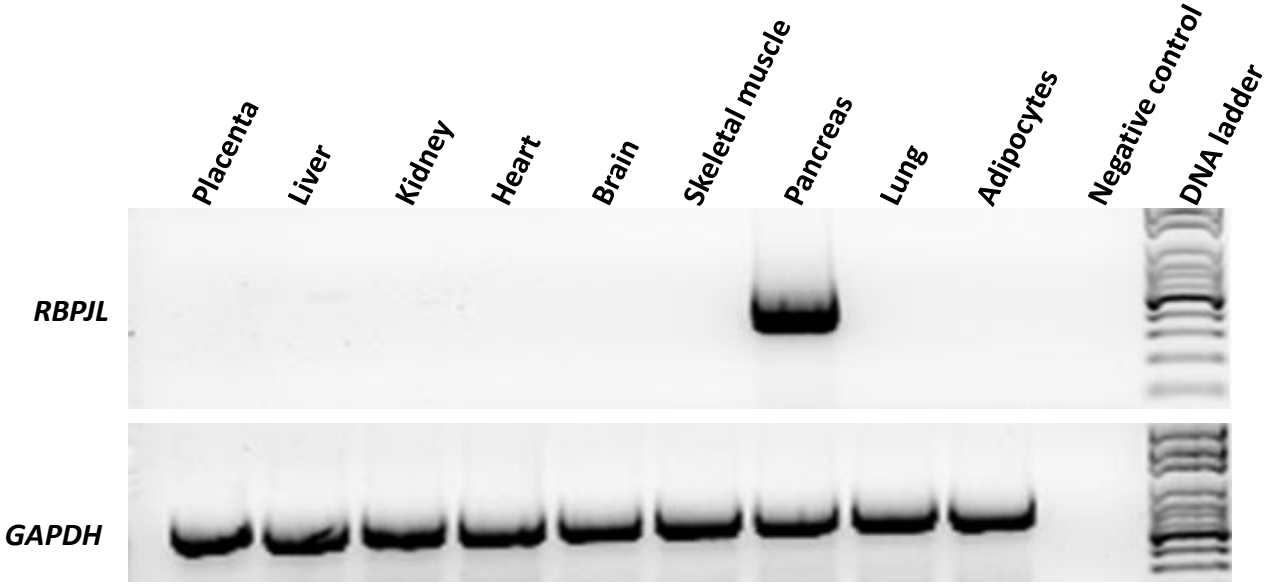
[1]Variant information given is based on genome build 37 (hg 19), refSeq: NM_014276.3 (*RBPJL*) and NM_178161.2 (*PTF1A*). AA is predicted amino acid change. [2]Frequency of the variant allele. [3]Frequency of the variant allele in 335 American Indian samples. [4]The online prediction tools PROVEAN, SIFT, and Polyphen-2 were used to predict whether the AA changes may impair the biological function of the proteins [ESM ref 1-3]. No missense variants were identified in *RBPJ*.

ESM Table 5: Association of c.839C>T (p.(Thr280Met)) in *RBPJL* with type 2 diabetes in cross-sectionally studied urban-dwelling American Indians

Tribal affliation[1]	Sample size	T2D Prevalence (%)	RAF[2]	OR (95% CI)[3]	*P*-Value[3]
I	**843**	**26.9**	**0.03**	**2.77(1.45-5.31)**	**0.002**
II	393	41.7	0.02	2.07(0.66-6.51)	0.21
III	702	30.7	0.02	0.71(0.31-1.61)	0.42
IV	1,055	22.2	0.08	0.96(0.64-1.45)	0.86
Combined[d]	**2,993**	**28.1**	**0.04**	**1.19(0.87-1.62)**	**0.27**

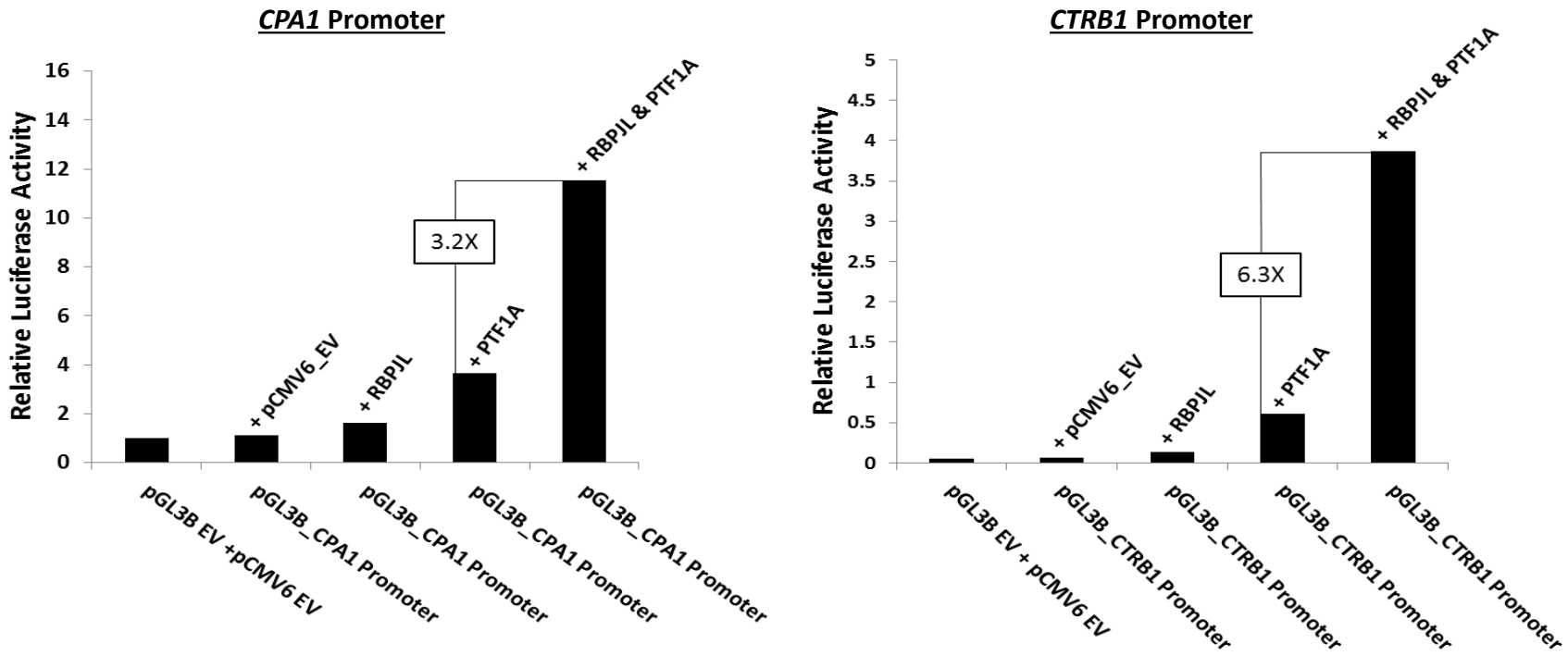
[1]All individuals had ≥ 50% American Indian ancestry. Specific tribal affiliation cannot be reported under the protocol for which these individuals were studied; Groups II-IV represent individuals who affiliate with one of the three specific tribes while Group I represent all others. [2]RAF-risk allele frequency; frequency of the methionine coding allele (T). [3]OR and *P*-values were adjusted for age, sex and first five principal components derived from GWAS. OR are given per copy of the risk allele (Met allele). [4]The *P*-value in the combined analysis was further controlled for inflation using the genomic control approach.

ESM Figure 1. RBPJL is predominantly expressed in the pancreas



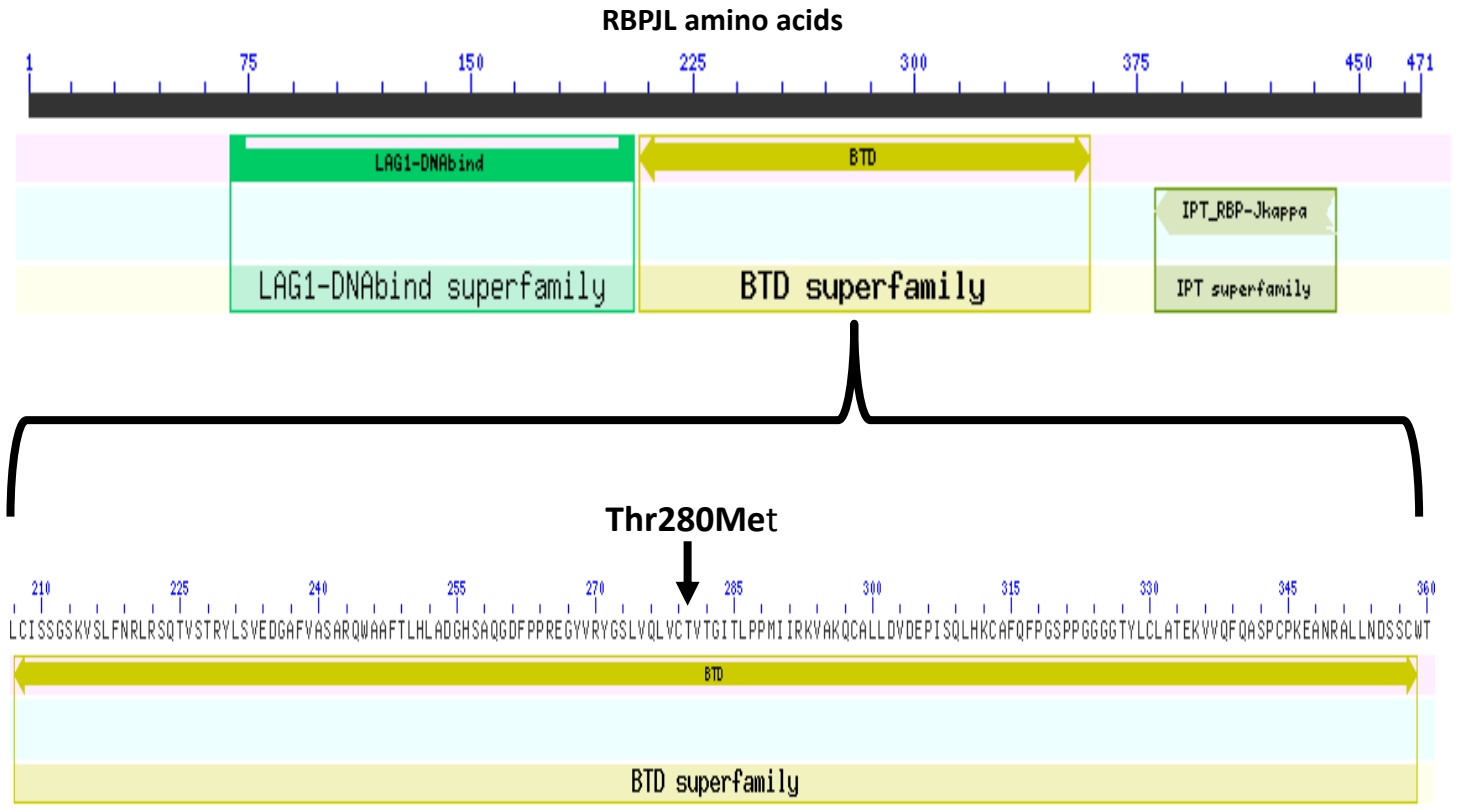
Gene expression levels were analyzed using ≈1ng of human cDNA with primers specific for RBPJL and GAPDH.

ESM Figure 2. Linked to Figure 2-Effect of RBPJL on the activity of *CPA1* and *CTRB1* promoters in HEK 293 cells



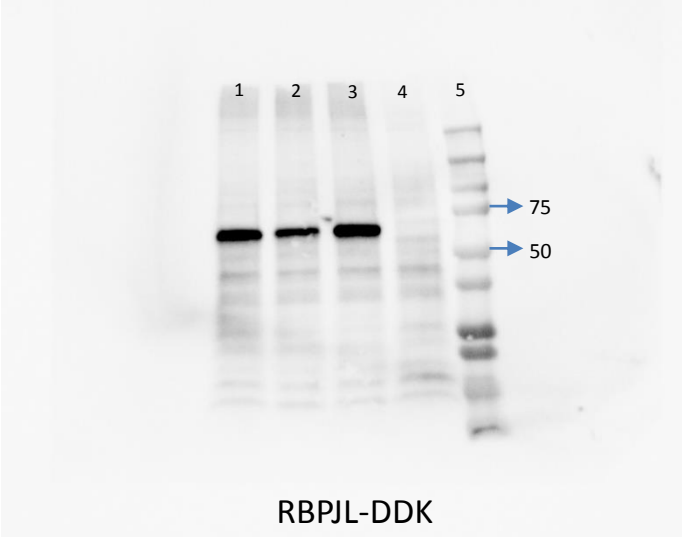
Fold increase in luciferase activity compared to pGL3-basic vector (pGL3B) and pCMV6 empty vector (EV) is shown. Equal amount (200ng) of RBPJL and/or PTF1A pCMV6 expression vectors were co-transfected with 600ng of pGL3B_CPA1 or pGL3B_CTRB1 promoter-reporter constructs along with 50ng of pGL4.74 Renilla luciferase vector as a control for transfection efficiency. Equal amount of total DNA was used for transfection in all conditions by using pCMV6 EV wherever necessary.

ESM Figure 3. Conserved protein domains for RBPJL

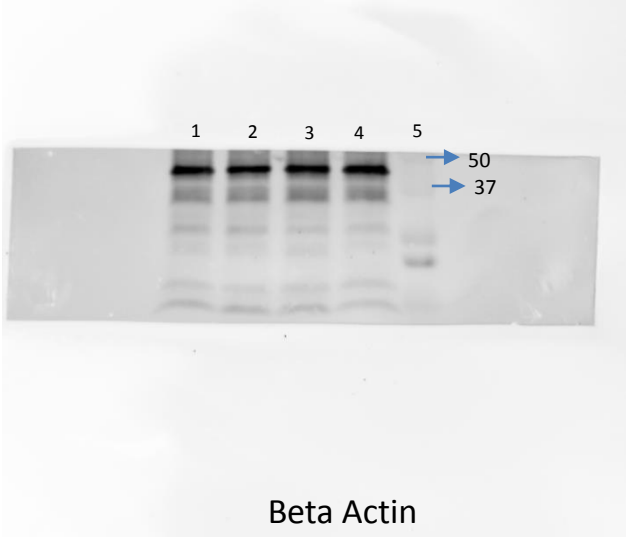


The p.(Thr280Met) missense variant is located in the beta-trefoil DNA binding domain (BTD). Adapted from NCBI Conserved Domain database [ESM ref 4].

ESM Figure 4. Linked to Figure 3A (raw image – Exp1)



RBPJL-DDK



Beta Actin

Lane1 – 445Asn RBPJL (not in Figure 3A)
Lane 2 – 280Met RBPJL (shown in Figure 3A)
Lane 3 – Wild type RBPJL (445Asp, 280Thr) (shown in figure 3A)
Lane 4 – pCMV6 empty vector (shown in figure 3A)
Lane 5 – Dual color precision plus protein standard (Biorad) (not in figure 3A)
*RBPJL predicted molecular weight – 56.6 kDa
*Beta actin predicted molecular weight – 42 kDa

Electronic supplementary material references

ESM ref 1. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015; **31**: 2745–2747.

ESM ref 2. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.

ESM ref 3. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; **Chapter 7**: Unit7.20.

ESM ref 4. Marchler-Bauer A, Bo Y, Han L *et al*. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 2017; **45**; 200–203.