

Table S1. Descriptive statistics of race, sex, treatment, stratum, risk level, use of insulin and age among 670 participants in DPT-1. P-values measures associations of individual category, in comparison with the reference with the progression to T1D onset, while p-value for age measures its association with the progression to T1D onset.

Covariates	Category	N	coef	HR	SE	Z	p
Race	White	608	0.00	Reference			
	Asian	3	0.88	2.40	0.71	1.23	2.18E-01
	Black	9	-0.18	0.83	0.58	-0.31	7.54E-01
	Hispanic	29	-0.36	0.70	0.41	-0.86	3.91E-01
	Other	7	0.66	1.93	0.51	1.30	1.92E-01
	Unkown	14	-0.53	0.59	0.58	-0.91	3.62E-01
Sex	Male	378	0.00	Reference			
	Female	292	0.11	1.11	0.13	0.82	4.13E-01
Age (10 years)			-0.23	0.79	0.08	-2.94	3.28E-03
Treatment	Oral Placebo	177	0.00	Reference			
	Oral Insulin	159	-0.23	0.80	0.21	-1.09	2.77E-01
	Observation	157	0.65	1.91	0.18	3.51	4.42E-04
	Parenteral Insulin	177	0.57	1.77	0.18	3.11	1.86E-03
Risk Level	Low	354	0.00	Reference			
	High	316	0.72	2.05	0.13	5.43	5.75E-08

Table S2. Distribution of race/ethnicity, sex, treatment and use of insulin among 354 low risk (oral insulin group) and 316 high risk participants (insulin injection group). P-values measure if row frequencies are the same between low and high-risk groups.

Variable	Categories	Low Risk (n=354)	High Risk (n=316)	p
Race/ethnicity	White	314 (88.7)	294 (93.04)	6.14E-02
	Asian	2 (0.56)	1 (0.32)	1.00E+00
	Black	7 (1.98)	2 (0.63)	1.83E-01
	Hispanic	20 (5.65)	9 (2.85)	8.78E-02
	Other	5 (1.41)	2 (0.63)	4.56E-01
	Unknown	6 (1.69)	8 (2.53)	5.90E-01
Sex	Female	139 (39.27)	153 (48.42)	1.92E-02
Treatment	Oral Placebo	177 (50)		
	Oral Insulin	177 (50)		
	Observation		157 (49.68)	
	Intervention		159 (50.32)	
Use of insulin	Yes	177 (50)	159 (50.32)	9.38E-01

Table S3. Distribution of DQA1-DQB1 haplotypes, and their association results on haplotype-progression: estimated coefficient (coef), hazard ratio (HR), standard error (SE), Z-score, p-value and q-value for false positive error rate for individual haplotypes.

ID	DQA1-DQB1	N	coef	HR	SE	Z	p	q
1	*01:01-*05:01	84	-0.02	0.98	0.18	-0.12	9.04E-01	9.34E-01
2	*01:02-*05:02	15	-0.28	0.76	0.51	-0.55	5.81E-01	7.40E-01
3	*01:02-*06:04	43	-0.19	0.82	0.26	-0.74	4.58E-01	7.40E-01
4	*01:03-*06:03	39	-0.18	0.83	0.29	-0.64	5.20E-01	7.40E-01
5	*02:01-*02:02	55	-0.39	0.68	0.26	-1.51	1.32E-01	3.66E-01
6	*03:01-*03:02	446	0.31	1.36	0.10	3.09	2.01E-03	2.82E-02
7	*03:02-*03:03	10	0.64	1.90	0.45	1.42	1.57E-01	3.66E-01
8	*03:03-*02:02	10	-0.39	0.68	0.70	-0.56	5.78E-01	7.40E-01
9	*03:03-*03:01	79	-0.48	0.62	0.23	-2.09	3.69E-02	2.58E-01
10	*03:03-*03:02	70	0.31	1.36	0.20	1.58	1.15E-01	3.66E-01
11	*04:01-*04:02	26	0.04	1.04	0.36	0.10	9.20E-01	9.34E-01
12	*05:01-*02:01	350	-0.01	0.99	0.11	-0.08	9.34E-01	9.34E-01
13	*05:05-*03:01	56	-0.48	0.62	0.27	-1.76	7.77E-02	3.62E-01
	Rare Hap (<10)	57	-0.21	0.81	0.26	-0.82	4.10E-01	7.40E-01

Table S4. Motifs of (-18 β , β 57) and their equivalent DQ haplotypes together with motif frequencies among carriers and non-carriers of DQ8.1 who may carry 0, 1 or 2 copies of respective motifs

Motif	DQ8.1	Non-DQ8.1	Equivalent Haplotypes
AA		427	*02:01-*02:02 *03:03-*02:02 *03:03-*03:04 *05:01-*02:01
AD	80	73	*01:03-*06:01 *03:03-*03:01 *05:03-*03:01 *05:05-*03:01 *05:05-*03:19 *05:05-*03:29 *06:01-*03:01
VA	519	3	*03:01-*03:02 *03:03-*03:02 *05:05-*03:02
VD	10	86	*01:03-*06:02 *01:03-*06:03 *01:04-*05:03 *02:01-*03:03 *03:02-*03:03 *03:03-*04:01 *03:03-*04:02 *04:01-*04:02 *04:02-*04:02 *04:04-*04:02
VS		15	*01:02-*05:02
VV		139	*01:01-*05:01 *01:02-*06:04 *01:02-*06:09 *01:05-*05:01
Total	609	743	Numbers of motif copies among DQ8.1 carriers and others

Table S5. Results from genotypic association analysis of motif (-18 β , β 57) among high risk subjects: genotype frequency, estimated coefficient (coef), hazard ratio (HR), standard error (SE), Z-score and p-value. Rare genotypes with fewer than 10 copies are merged into a single group.

Genotype	freq	coef	HR	SE	Z	p
VA/VA	35	0.68	1.98	0.23	2.99	2.80E-03
AA/AD	17	-2.47	0.08	1.00	-2.46	1.39E-02
AD/VV	10	-0.70	0.49	0.58	-1.21	2.28E-01
VA/VD	21	0.33	1.39	0.30	1.08	2.78E-01
AD/VA	21	0.29	1.34	0.30	0.97	3.32E-01
AA/VA	100	0.14	1.15	0.17	0.83	4.08E-01
AA/VV	16	-0.28	0.75	0.42	-0.68	4.95E-01
VA/VV	29	0.17	1.18	0.29	0.57	5.65E-01
AA/VD	14	-0.13	0.88	0.42	-0.31	7.57E-01
AA/AA	28	-0.02	0.98	0.30	-0.06	9.52E-01
Rares (<10)	25	-0.83	0.43	0.42	-2.00	4.53E-02

Table S6 Functions of differing residues between HLA-DQA1*03:01-B1*03:02 and -DQA1*03:03-B1*03:01

HLA-DQ Molecule/ residue	A1*03:01- B1*03:02	A1*03:03- B1*03:01	Function of residue
a157	Asp	Ala	Participates in the formation of putative homodimer of heterodimers
-b18	Val	Ala	SPa, additional function unknown
b13	Gly	Ala	P4b
b26	Leu	Phe	P4 b
b45	Gly	Glu	Shapes surface electrostatic potential
b57	Ala	Asp	P9 b
b167	Arg	His	If Arg, participates in RGD loop; function and shape of HGD sequence unknown
b185	Ile	Thr	unknown

a) SP = signal peptide; b) P = pocket

Table S7. Result from the association analysis of identified motif of (-) with elevated IAA level at the baseline (430 and 910 participants with IAA level less than or greater than 80): motif frequencies among negative and positive IAA level, estimated odds ratio, Z score and p-value.

Motif	IAA <80	≥ 80	OR	Z	p
AA	149 (34.65)	272 (29.89)	0.86	-1.86	0.06
AD	61 (14.19)	90 (9.89)	0.70	-2.25	0.02
VA	149 (34.65)	370 (40.66)	1.17	2.31	0.02
VD	33 (7.67)	63 (6.92)	0.90	-0.51	0.61
VS	4 (0.93)	11 (1.21)	1.30	0.46	0.65
VV	34 (7.91)	104 (11.43)	1.45	1.92	0.05