Use of class I and class II HLA loci for predicting age at onset of type 1 diabetes in multiple populations

Ana M Valdes[1], **Henry A Erlich**[2,3], **Joyce Carlson**[4], **Michael Varney**[5], **Priscilla V Moonsamy**[3], and **Janelle A Noble**[2]

[1]Dept of Twin Research and Genetic Epidemiology, King's College London, UK

[2]Children's Hospital Oakland Research Institute, Oakland, California

[3]Roche Molecular Systems, Pleasanton, California

[4]Clinical Chemistry, University Hospital, Malmö, Sweden

[5]Victorian Transplantation and Immunogenetics Service, Australian Red Cross Blood Service, Melbourne, Australia

Abstract

Aims—to assess in multiple populations the role of HLA alleles on early and late age at onset of type 1 diabetes.

Methods—Stepwise linear regression models were used to determine which HLA class I and class II risk alleles to include. High resolution genotyping data for patients from the Type 1 Diabetes Genetics Consortium (T1DGC) collection (n=2278) and four independent cohorts from Denmark, Sardinia and the US (HBDI and Joslin) (n=1324) patients (total n=3602) were used to assess the role of HLA variation on age of onset and predict early onset (age   5) and late onset (age 15) of type 1 diabetes.

Results—In addition to carriage of HLA class I alleles A*24:02, B*39:06, B*44:03 and B*18:01, HLA class II DRB1-DQB1 loci significantly contributed to age at onset explaining 3.4% of its variance in the combined data. HLA genotypes with gender were able to predict late onset in all cohorts studied with area under the curve (AUC) values ranging from 0.58 to 0.63. Similar AUC values (0.59–0.70) were obtained for early onset for most cohorts except in the Sardinian study in which none of the models tested had significant predictive power.

Conclusions—HLA associations with age of onset are consistent across most Caucasian populations and HLA information can predict some of the risk of early and late onset of type 1 diabetes. Considerable heterogeneity was observed between Sardinian and other populations particularly with regards to early age of onset.

Corresponding Author: Ana M Valdes PhD Dept of Twin Research & Genetic Epidemiology, King's College London St Thomas Hospital, London SE1 7EH Tel +44(0)207-188-6765 Fax +44(0)207-188-6718 ana.valdes@kcl.ac.uk.

Introduction

Type 1 diabetes is one of the most widely studied complex genetic disorders, and the genes in the human leukocyte antigen (HLA) region are reported to account for approximately 40–50% of the familial aggregation of type 1 diabetes [1]. Age at onset of type 1 diabetes may modify the metabolic phenotype of the patients and may influence the risk of late complications of diabetes. For example, age at onset of type 1 diabetes significantly modifies the long-term risk of proliferative retinopathy. The highest risk for retinopathy is seen in age-at-onset group 5–14 years, whereas the lowest risk is in age-at-onset group 15–40 years [2]. Similarly patients with onset of diabetes after age 15 have been observed to have a lower risk of diabetic nephropathy and end stage renal disease than do patients diagnosed during adolescence [3]. On the other hand recent studies indicate higher mortality for type 1 diabetes patients diagnosed in late adolescence or adulthood than for patients diagnosed earlier [4].

A significant genetic component for age at onset has been reported, in particular, a contribution by specific human leukocyte antigen (HLA) alleles [5–7]. However, previous studies investigating the role of HLA alleles on age of onset have all come from a single cohort and have analysed sample sizes of only a few hundred patients at a time [5–7]. The aim of this study was to investigate HLA class I and class II classical loci genotyped in a large collection of patients from the T1DGC collection to assess their effect on age at onset of type 1 diabetes. We have studied different populations of European descent focusing on the role of specific DRB1-DQB1 genotypes, DPB1 and HLA class I alleles that have been previously implicated in risk of type 1 diabetes or on age at diagnosis. We have compared genetic prediction risk models for early age at onset (age <5) and for late age at onset (15).

Subjects and Methods

Study subjects

The T1DGC is a large, worldwide, collaborative study aimed at collecting and genotyping new type 1 diabetes families in a highly standardized fashion from multiple populations, to aid in the search for additional type 1 diabetes genes within and outside the HLA region [8]. An individual was designated as affected if he or she had documented type 1 diabetes with onset at 37 years of age, had used insulin within 6 months of diagnosis, and had no concomitant disease or disorder associated with diabetes. High resolution HLA genotyping was performed at eight classical MHC loci by four genotyping centers using standardized typing protocols, reagents, and quality control procedures [9]. In addition to the patient clinical samples collected by the T1DGC, genotyping was also carried out in existing clinical collections. Age at onset and high resolution genotyping data were also available for samples and data collected outside of the T1DGC framework and contributed for inclusion in various T1DGC projects; including the Danish, HBDI, Joslin and Sardinian collections

Proband status—For the T1DGC collection, the proband was identified as the first child diagnosed with type 1 diabetes in the family. The `proband' variable within the data set identifies the first child diagnosed with type 1 diabetes. For the existing cohorts, the criteria for proband assignment was not readily available for all pedigrees.

Allele selection and genotype coding—The genetic contribution of DRB1-DQB1 genotypes was encoded as DR3/DR4=4, DR3/DR3=3 DR4/DR4=2, DR4/DRx, DR3/DRx=1, DRx/DRx=0. Where DR3=DRB1*03:01-DQB1*02:01, DR4= DRB1*04:01/2/4/5/8/13-DQB1*03:02 or 03:04 or 02:01 and x is any other haplotype including DRB1*04:03 or other DRB1*04 carrying haplotypes with DQB1*03:01. Genotypes that included the highly protective allele DQB1*06:02, the haplotype DRB1*14:01 DQB1*05:03 or DRB1*07:01 DQB1*03:03 were categorized as DRx/DRx. This ranking was based on previous reports of predisposing, protective and neutral DR-DQ haplotypes [15]. HLA alleles at loci other than DRB1-DQB1 that have been convincingly implicated in risk of type 1 diabetes were also included in the model. These included DPB1 alleles 02:02, 03:01 and 04:02 and class I alleles A*24:02, A*02:01, A*11:01, A*30:02, A*32:01, A*66:01, B*18:01, B*35:02, B*57:01, C*03:03, B*39:06, B*44:03, C*07:02 [6,10–11]

Risk factor selection—Stepwise linear regressions were carried out using age at onset as the outcome continuous variable including all the above genetic factors in addition to gender, cohort of origin and proband status.

Outcome variables—Having identified which HLA variables to include, logistic regressions were carried out in each cohort separately, adjusting for proband status and including the HLA alleles that were found to influence age at onset in the step-wise linear regression analyses. Two binary outcome variables were defined: (1) "early age at onset" coded as 1 if age at onset 5 (28.6% of patients) or 0 if age at onset >5 (71.4% of patients), (2) "late age at onset" coded as 1 if age at onset >15 (21.5% of patients) or 0 if age at onset 15 (78.5% of patients). These age cut-offs were defined based on the ages at which differences in rates of complications and mortality have been observed [3–4].

Inter study heterogeneity—Inter-study hetereogeneity was assessed using a DerSimonian Laird random effects meta-analysis and computing the heterogeneity variance τ. The rmeta library in R was used (http://cran.r-project.org/web/packages/rmeta/rmeta.pdf).

Calibration and discrimination—The predictive power of a given diagnostic is usually summarized by a receiver operating characteristic (ROC) curve. In this type of analysis, subjects are ranked in descending order of their predicted risk and the cumulative proportion of subjects who develop disease (cases) is plotted against the corresponding cumulative proportion of the population, i.e., the sensitivity (true positive fraction) is plotted in the *y*-axis vs 1- specificity (the false negative fraction) in the *x*-axis [12]. A perfect diagnostic would be represented by a line that starts at the origin, travels up the y axis to 1 and then across the origin to an *x*-axis value of 1, thus having a total area under the curve (AUC) of 1. A test with AUC=0.5 on the other hand has zero diagnostic value. Whereas discrimination examines the ability to correctly classify subjects into different groups, calibration assesses how closely the predicted probabilities reflect actual risk [12]. Calibration and discrimination abilities of the models were examined in the independent cohorts described above.

A risk score was calculated for each individual using the logit equation:

$$\text{Logit} = \ln(p/1 - p) = \alpha + \beta 1 X_i + \ldots + \beta_i X_i$$

Where p is the probability of the outcome (early or late age at onset), α is the constant and β is the natural logarithm value of the odds ratio for a specific predictor X_i.

The logit operator maintains the linearity of the model and allows the calculation of a probability of the outcome (in this case early or late age at onset) given the different sets of predictors, according to p= exp(logit)/ (1+exp(logit)). Thus the higher the risk score, the greater the risk of the outcome.

The individuals were classified into different sub-groups according to the risk scores. Observed and predicted frequencies of the disease in subgroups were calculated. The Hosmer–Lemeshow $\chi2$ statistics for goodness-of-fit were used for calibration to compare observed and predicted risk [13]. Non significant p-values for this test indicate good calibration. Both discrimination and calibration of risk models were carried out using the PredictABEL package for R (http://cran.r-project.org/web/packages/PredictABEL/index.html).

Results

The mean and standard deviation of age at onset by gender and proband status for each of the cohorts are summarized in Table 1. Genotyping and age at onset information for a total of 3602 type 1 diabetes patients corresponding to 1801 affected sib pairs from the T1DGC, including the extant collection, were included. The overall range for age at onset was 0–37 years.

Age at onset was found to be significantly higher in males (p<0.005) than in females. In the T1DGC collection, where the proband was defined as the first child to develop type 1 diabetes, a strong difference in age at onset is seen between probands and non probands (Table 1). In the T1DGC collection and some of the pre-existing collections probands status appears to be a confounding variable for younger onset.

A stepwise logistic regression was carried out, which included DPB1 alleles 02:02, 03:01 and 04:02, class I alleles A*24:02, A*02:01, A*11:01, A*32:01, A*66:01, B*18:01, B*35:02, B*57:01, C*03:03, B*39:06, B*44:03, and C*07:02, adjusting for gender, cohort of origin, and proband status. This analysis revealed that, for HLA, only DR-DQ genotype, A*24:02, and B*39:06 contributed significantly to age of type 1 diabetes onset, with B*44:03 and B*18:01 nearing statistical significance (Table 2).

Individual Population Effects—The allele/genotype frequencies for DR3/4, A*24:02, B*18:01, B*39:06, and B*4403 are shown in Table 3. The frequencies are stratified by early age at onset (age<=5) versus not, by late age at onset (age>=15) versus not and by early onset (age<=5) versus late onset (age>=15). Differences in allele and genotype frequencies are seen among the various populations, as is expected for the genes in the HLA region. The effect of these alleles and genotypes on early and late onset was assessed by multiple logistic regression including all genetic variables in the model in addition to gender and probands status. We observe a striking difference in the frequency of B*18:01 among Sardinian patients compared to the other groups (Table 3). A much higher frequency of certain DR3 haplotypes in this population, compared to other European populations, has already been reported [14].

Using random effects meta-analysis we investigated whether there was evidence of statistically significant heterogeneity between study cohorts, i.e. whether, regardless of the frequency, the effect on age of onset was different for each of the five genetic variables studied. We found no evidence of inter-study heterogeneity for any of the early onset genetic effects nor for any of the early versus late onset associations with the smallest p-value for heterogeneity being p=0.20. From both of these traits (early versus other and early versus late), meta-analyses of the genetic effects of B*18:01 and B*44:03 did not reach statistical

significance; all other associations were statistically significant overall. For late age of onset we observed evidence for inter-study heterogeneity for the effect of B*39:06 on this outcome yielding τ=0.086 p=0.05. By meta-analysis the only genetic effects that were significantly associated with late age of onset were the DR-DQ genotype and A*24:02, indicating that these are the most consistent effects throughout the age of onset distribution. In the absence of significant heterogeneity within the T1DGC sub-cohorts we have merged them for the risk prediction analysis.

We then assessed whether these HLA markers could predict early or late age at onset

Risk Prediction Models—A logistic regression on early onset and late onset outcomes was then fitted using three different models for each outcome: (1) gender as the only risk factor (2) HLA as the only risk factor (3) gender and HLA as risk factors. Early versus late models were not fitted given the small sample sizes involved.

The following models were fitted:

> **A.** Early age at onset
>
>> **1)** gender only
>>
>> Logit= −1.00254+0.07075 gender
>>
>> **2)** HLA only
>>
>> Logit=−1.611245+ 0.12472 B*18:01 +0.855414 B*39:06 −0.3341 B*44:03 + 0.2947 A*24:02 +0.1869 DR-DQ
>>
>> **3)** HLA + gender
>>
>> Logit= −1.91141+ 0.096 gender+ 0.1225 B*18:01 +0.29309 A*24:02 +0.8556 B*39:06 − 0.3394 B*44:03 + 0.18356*DR-DQ
>
> **B.** Late age at onset
>
>> **1)** gender only
>>
>> Logit= −1.3957−0.1790 gender
>>
>> **2)** HLA only
>>
>> Logit= −0.8281− 0.008191 B*18:01 −0.60403 B*39:06 +0.11133 B*44:03 −0.4537 A*24:02 +0.18383 DR-DQ
>>
>> **3)** HLA + gender
>>
>> Logit=−0.6588 −0.207 proband status − 0.207 gender − 0.0478*B*18:01 −0.5087A*24:02 −0.4058B*39:06 + 0.2339B*44:03 − 0.1621*DR-DQ

The risk discrimination and calibration results from these models in all cohorts are shown in Table 4.

The best prediction for early onset was seen in the JOS cohort for an HLA only model yielding an area under the curve (AUC) = 0.700 (Table 4). For late onset, the best AUC was seen in the DAN cohort for a model including both gender and HLA (AUC=0.644). For all other cohorts and outcomes except one, at least one of the models had an AUC value significantly higher than a value of 0.5 at which the test would have no predictive value. None of the risk prediction models is significantly different from 0.5 in the Sardinian cohorts.

For late onset the models that included gender and HLA had calibration problems in two of the cohorts because of the heterogenous relationship between gender and age of onset among cohorts. We also note that the large confidence intervals for AUC in some of the cohorts is likely due to the smaller sample size available in those studies.

Discussion

In the current study we have investigated the role of HLA high resolution genotypes on age at T1D onset in various populations of European descent. To our knowledge this is the first study to compare the role of HLA on age of onset in different populations.

We further investigated whether such genotyping information could have any predictive value for assessing the risk of very young age at onset in contrast to late onset for type 1 diabetes. Using the largest data set to date to address this question we found that the strongest genetic contribution to age of onset appears to come from the DRB1-DQB1 genotypes, which also have the strongest influence on disease risk [15]. In addition a few select class I alleles, notably A*24:02, B*18:01, B*39:06, and B*44:03 also influence age at onset. Of these the most consistent effect is that of A*24:02, whereas the other class I alleles either do not influence specific cut-offs of age of onset (early versus late) or show evidence of strong heterogeneity across populations (e.g. B*39:06 for late age of onset).

In both the T1DGC collection and other independent extant collections, we find that genotypes for classical class I and class II HLA can have some modest predictive power for these two outcomes. On the one hand, this confirms the role of HLA polymorphism in influencing age at onset. On the other hand, it highlights that other risk factors not included in our models must also be influencing age at type 1 diabetes onset.

Current approaches for the prediction of type 1 diabetes in screening studies take advantage of the major genetic risk factors, genotyping for HLA-DR and HLA-DQ loci and screening for autoantibodies directed against islet-cell antigens [16]. For example, children who carry both of the highest-risk HLA haplotypes (DR3/DR4–DQB1*03:02) have a risk of approximately 1 in 20 for a diagnosis of T1D by the age of 15 years [16]. The results presented here may help improve such models by taking into account also the role of genetic risk factors on age at onset.

We found that gender had little or no predictive value and that because the relationship with age of onset was not consistent across cohorts in most instances it did not improve the AUC. For the DAN and HBDI where the difference in age of onset between genders was strongest, the inclusion of gender did show a slight improvement but not in an additive way. This consistent with what has been reported for the combination of genetic and non genetic factors for other disease areas [17].

We note several study limitations. Our analyses have used data derived from affected sib pair cohorts of European descent, selecting for patients with a strong genetic contribution to type 1 diabetes and therefore possibly also to its age at diagnosis. The current data are thus reflective of the prediction of HLA in a group of patients enriched for genetic risk. On the other hand, these data are relevant to clinical research, as studies of first degree relatives (follow-up, prevention trials) involve those who have a family member [16, 18] already diagnosed with T1D and genetic factors combined with other factors could be applied to the analysis of data from cohorts of relatives. In addition, these results highlight the differences between European descent populations and illustrate the limits and the extent to which HLA may be helpful in predicting age at disease onset.

We have developed and calibrated three risk prediction models for age at early and late onset of type 1 diabetes, based on five independent patient collections. We hope that these models may be used as pilots to lead further research in defining risk prediction for age at onset using other risk factors (e.g., environmental exposures, autoantibodies). The models may be applied at the individual level to predict the most likely category of age at onset (early or late), but also at the population level, with reference to other relative risks from published studies, to estimate the potential population risk reduction that may be gained by primary prevention of any modifiable risk factors that influence type 1 diabetes and the ensuing complications.

Acknowledgments

REFERENCES

1. Noble JA, Valdes AM. Genetics of the HLA Region in the Prediction of Type 1 Diabetes. Curr Diab Rep. Dec; 2011 11(6):533–42. [PubMed: 21912932]

2. Hietala K, Harjutsalo V, Forsblom C, Summanen P, Groop PH. FinnDiane Study Group. Age at onset and the risk of proliferative retinopathy in type 1 diabetes. Diabetes Care. Jun; 2010 33(6): 1315–9. [PubMed: 20185730]

3. Finne P, Reunanen A, Stenman S, Groop PH, Grönhagen-Riska C. Incidence of end-stage renal disease in patients with type 1 diabetes. JAMA. 2005; 294:1782. [PubMed: 16219881]

4. Harjutsalo V, Maric C, Forsblom C, Thorn L, Wadén J, Groop PH. Sex-related differences in the long-term risk of microvascular complications by of type 1 diabetes. Diabetologia. 2011; 54:1992–9. [PubMed: 21509443]

5. Valdes AM, Thomson G, Erlich HA, Noble JA. Association between type 1 diabetes age of onset and HLA among sibling pairs. Diabetes. 1999; 48(8):1658–61. [PubMed: 10426387]

6. Valdes AM, Erlich HA, Noble JA. Human leukocyte antigen class I B and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. Hum Immunol. Mar; 2005 66(3):301–13. [PubMed: 15784469]

7. Tait BD, Harrison LC, Drummond BP, Stewart V, Varney MD, Honeyman MC. HLA antigens and age at diagnosis of insulin-dependent diabetes mellitus. Hum Immunol. Feb; 1995 42(2):116–22. [PubMed: 7744614]

8. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–7. [PubMed: 19430480]

9. Mychaleckyj JC, Noble JA, Moonsamy PV, et al. HLA genotyping in the international Type 1 Diabetes Genetics Consortium. Clinical trials (London, England). 2010; 7:S75–87.

10. Varney MD, Valdes AM, Carlson JA, et al. HLA DPA1, DPB1 alleles and haplotypes contribute to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2010; 59(8):2055–62. [PubMed: 20424227]

11. Noble JA, Valdes AM, Varney MD, et al. HLA class I and genetic susceptibility to type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium. Diabetes. 2010; 59(11):2972–9. [PubMed: 20798335]

12. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994; 309:188. [PubMed: 8044101]

13. Hosmer, DW.; Lemeshow, S. Applied Logistic Regression. Wiley; New York: 2000.

14. Koeleman BP, Herr MH, Zavattari P, et al. Conditional ETDT analysis of the human leukocyte antigen region in type 1 diabetes. Ann Hum Genet. May; 2000 64(Pt 3):215–21. [PubMed: 11246473]

15. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes. Apr; 2008 57(4):1084–92. [PubMed: 18252895]

16. Aly TA, Ide A, Humphrey K, et al. Genetic prediction of autoimmunity: initial oligogenic prediction of anti-islet autoimmunity amongst DR3/DR4-DQ8 relatives of patients with type 1A diabetes. J Autoimmun. 2005; 25(Suppl):40–45. [PubMed: 16242305]

17. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ, GRIPS Group. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. Eur J Clin Invest. 2011; 41(9): 1004–9. [PubMed: 21434891]

18. Aly TA, Ide A, Jahromi MM, et al. Extreme genetic risk for type 1A diabetes. Proc Natl Acad Sci U S A. 2006; 103:14074–14079. [PubMed: 16966600]

Table 1

Age at onset and gender distribution by cohort of the patients included in the study.

Collection	cohort	N	Male N	Male mean	Male SD	Female n	Female mean	Female SD	Female p-value	num sib pairs	Probands mean	Probands SD	Not Probands mean	Not Probands SD	Not Probands p-value
T1DGC collection	Asia Pac	300	152	9.86	7.52	148	10.09	7.27	0.79	150	7.82	6.01	12.13	8.00	2.6e-07
	Eur	948	513	11.18	7.29	435	10.23	7.20	0.045	474	8.32	6.00	13.16	7.60	<1.0e-200
	N Am	816	436	8.99	6.60	380	8.36	6.29	0.14	408	7.17	5.25	10.23	7.16	5.8e-012
	UK	214	96	7.82	4.57	118	7.64	4.66	0.61	107	6.59	4.65	8.86	4.29	6.6e-05
Extant collectoins	DAN	196	106	13.33	7.98	90	9.66	7.09	0.0011	98	12.83	8.00	10.46	7.42	0.16
	HBDI	898	484	12.58	8.18	414	11.12	7.82	0.00121	449	12.92	8.46	10.89	7.48	0.00065
	JOS	130	66	11.18	6.66	64	11.86	7.80	0.56	65	11.08	6.86	11.95	7.60	0.24
	SAR	100	46	11.75	8.55	54	13.06	7.25	0.33	50	11.43	7.34	13.48	8.29	0.15
	all	3602	1899	10.89	7.49	1703	9.96	7.15	3.3e-05	1139	9.49	7.10	11.41	7.47	<1.0e-200

Differences in mean age of onset between males and females and between probands and non-rpobands were compared by t- test.

Table 2

Stepwise linear regression of factors associated with type 1 diabetes age at onset 1801 sib pairs

Risk Factor	Coeff.	SE	p Value
Proband (yes=1, no=0)	−1.9909	0.2372	6.6E-17
DR-DQ *(DR3/4=4, DR3/3=3, DR4/4=2, DR4/x or DR3/x=1, DRx/x=0)*	−0.7125	0.0801	9.0E-19
A*24:02 (0, 1 or 2 copies carried)	−1.5917	0.2804	1.5E-08
B*39:06 (0, 1 or 2 copies carried)	−2.6639	0.5084	1.7E-07
B*44:03 (0, 1 or 2 copies carried)	1.0109	0.5996	0.0919
B*18:01 (0, 1 or 2 copies carried)	−0.4958	0.2904	0.0878
Gender (female=2 male=1)	−0.5685	0.1189	1.8E-06
R^2 accounted for by HLA genotype	0.03843		

[#] results are adjusted for cohort of origin. Coefficients for the categorical variable "cohort" are not shown.

[##] DR3= DRB1*03:01 DQB1*02:01 DR4= DRB1*04:01/02/04/05/08 DQB1*02:01 or 03:02 or 03:04 x = any other DRB1-DQB1 haplotypes, DRx/x is also any genotype that carried DQB1*0602, DRB1*14:O1 DQB1*05:03 or DRB1*07:01 DQB1*03:03.

Table 3

Comparison of HLA alleles associated with age of onset and their association with early and late age of onset in different populations. Odds ratios were derived from a logistic regression model including all genetic variables as well as gender and proband status. Outcomes tested are age at onset 5 and late age at onset 15 versus all other patients or versus each other.

HLA allele/genotype	cohort	onset 5# n=1021	onset >5 n=2581	OR (95% CI)	p	onset <15 n=2827	onset 15# n=775	OR (95% CI)	p	onset 5# n=1021	onset 15# n=775	OR (95% CI)	p
A*24:02	T1DGC												
	Asia Pac	19.7%	13.4%	1.34(0.79–2.26)	0.28	16.3%	11.2%	0.62(0.31–1.26)	0.19	19.7%	11.2%	0.52(0.22–1.22)	0.13
	Eur	15.6%	11.8%	1.38(1.0–1.9)	0.05	13.6%	10.3%	0.76(0.53–1.09)	0.13	15.6%	10.3%	0.61(0.38–0.95)	0.0307
	N Am	11.5%	8.5%	1.35(0.95–1.92)	0.089	10.0%	7.1%	0.72(0.42–1.24)	0.24	11.5%	7.1%	0.73(0.39–1.34)	0.30
	UK	10.1%	8.1%	1.58(0.73–3.4)	0.24	8.6%	13.6%	1.29(0.31–5.35)	0.77	10.1%	13.6%	0.7(0.13–3.78)	0.68
	DAN	20.0%	24.3%	0.74(0.29–1.87)	0.53	12.0%	10.8%	0.88(0.37–2.1)	0.78	20.0%	10.8%	1.39(0.39–5.01)	0.61
	HBDI	13.7%	10.0%	1.38(0.94–2.02)	0.10	12.6%	6.2%	0.45(0.29–0.7)	0.0003	13.7%	6.2%	0.44(0.26–0.74)	0.0019
	JOS	33.3%	18.9%	2.27(0.76–6.73)	0.14	13.1%	5.9%	0.21(0.05–0.97)	0.046	33.3%	5.9%	0.11(0.02–0.72)	0.0206
	SAR	15.8%	13.4%	0.26(0.02–2.77)	0.26	7.0%	6.7%	0.41(0.08–2.08)	0.28	15.8%	6.7%	3.71(0.15–90.9)	0.42
B*18:01	T1DGC												
	Asia Pac	14.9%	12.4%	1.23(0.57–2.62)	0.60	6.7%	6.0%	0.97(0.36–2.56)	0.95	14.9%	6.0%	1.16(0.35–3.87)	0.81
	Eur	12.4%	11.3%	1.06(0.76–1.49)	0.73	11.5%	11.8%	1.06(0.75–1.49)	0.75	12.4%	11.8%	1.06(0.67–1.66)	0.81
	N Am	9.1%	6.0%	1.5(1–2.24)	0.051	7.2%	6.3%	0.99(0.55–1.83)	0.99	9.1%	6.3%	0.74(0.38–1.46)	0.39
	UK	19.5%	11.0%	0.98(0.56–1.73)	0.95	12.1%	0.0%	0.00(0.0–3e+14)	0.72	19.5%	0.0%	0(0–inf)	0.70
	DAN	8.9%	16.2%	0.55(0.17–1.78)	0.32	8.5%	3.9%	0.40(0.12–1.28)	0.12	8.9%	3.9%	0.85(0.16–4.6)	0.85
	HBDI	8.3%	8.9%	0.97(0.65–1.45)	0.87	9.0%	7.9%	0.86(0.58–1.28)	0.463	8.3%	7.9%	0.89(0.54–1.47)	0.65
	JOS	22.2%	13.7%	1.61(0.49–5.26)	0.43	8.5%	5.9%	0.62(0.18–2.15)	0.45	22.2%	5.9%	0.33(0.07–1.6)	0.17
	SAR	60.5%	40.2%	3.07(1.18–8.02)	0.022	45.1%	41.7%	1.12(0.53–2.37)	0.76	60.5%	41.7%	0.53(0.18–1.56)	0.25
B*39:06	T1DGC												
	Asia Pac	10.6%	1.9%	5.65(1.56–20.5)	0.009	2.7%	0.9%	0.41(0.05–3.43)	0.41	10.6%	0.9%	0.3(0.03–2.83)	0.29
	Eur	8.0%	4.7%	1.9(1–3.61)	0.049	3.3%	1.5%	0.44(0.19–1.03)	0.06	8.0%	1.5%	0.23(0.08–0.64)	0.0048
	N Am	4.0%	2.2%	1.98(1.07–3.67)	0.030	3.0%	1.8%	0.57(0.2–1.64)	0.30	4.0%	1.8%	0.38(0.12–1.2)	0.10
	UK	1.3%	3.0%	0.41(0.04–4.06)	0.45	1.2%	0.0%	0.00(0.0–2e+58)	0.91	1.3%	0.0%	(0–inf)	0.93
	DAN	17.8%	8.8%	3.21(1.07–9.66)	0.038	5.6%	4.9%	0.68(0.21–2.19)	0.51	17.8%	4.9%	0.18(0.04–0.88)	0.0340

HLA allele/genotype	cohort	onset 5[#]	onset >5	OR (95% CI)	p	onset <15	onset 15[#]	OR (95% CI)	p	onset 5[#]	onset 15[#]	OR (95% CI)	p
		n=1021	n=2581			n=2827	n=775			n=1021	n=775		
	HBDI	7.4%	3.0%	3.02 (1.77–5.16)	5.3E-05	4.8%	2.1%	0.46 (0.23–0.93)	0.031	7.4%	2.1%	0.3 (0.14–0.65)	0.0021
	JOS	3.7%	3.2%	0.77 (0.06–10.1)	0.84	1.1%	2.9%	1.26 (0.98–1.61)	0.07	3.7%	2.9%	11.0 (0.46–261.)	0.14
	SAR	10.5%	1.2%	36.7 (1.3–1040)	0.034	1.4%	1.7%	1.08 (0.12–39.4)	0.61	10.5%	1.7%	0.07 (0–3.51)	0.18
B*44:03	T1DGC												
	Asia Pac	4.3%	4.3%	1.13 (0.93–1.36)	0.22	4.5%	3.4%	0.41 (0.08–2.03)	0.28	4.3%	3.4%	0.47 (0.06–3.49)	0.46
	Eur	3.2%	5.3%	0.43 (0.14–1.33)	0.14	4.8%	4.6%	0.85 (0.41–1.76)	0.66	3.2%	4.6%	1.58 (0.54–4.62)	0.40
	N Am	2.4%	4.7%	0.58 (0.19–1.71)	0.32	3.4%	7.1%	2.27 (0.95–5.43)	0.06	2.4%	7.1%	3.36 (1.06–10.6)	0.0394
	UK	2.5%	5.9%	0.9 (0.71–1.14)	0.39	4.9%	0.0%	0.00 (0.0–5e+37)	0.86	2.5%	0.0%	0 (0–inf)	0.91
	DAN	2.2%	0.7%	0.65 (0.05–9.17)	0.75	0.7%	2.0%	1.92 (0.09–40.8)	0.67	2.2%	2.0%	0.32 (0.01–9.78)	0.51
	HBDI	2.3%	4.5%	1.05 (0.29–3.84)	0.94	3.4%	5.5%	1.19 (0.58–2.44)	0.64	2.3%	5.5%	1.26 (0.42–3.78)	0.67
	JOS	3.7%	3.2%	1.23 (0.89–1.69)	0.20	2.3%	5.9%	2.57 (0.32–20.4)	0.37	3.7%	5.9%	1.29 (0.04–38.5)	0.88
	SAR	5.3%	6.1%	0.95 (0.58–1.55)	0.83	5.6%	6.7%	1.13 (0.15–8.53)	0.91	5.3%	6.7%	0.3 (0.01–7.45)	0.47
DR3/DR4 *	T1DGC												
	Asia Pac	53.2%	40.2%	1.16 (0.97–1.38)	0.11	49.4%	22.4%	0.69 (0.55–0.85)	0.0006	53.2%	22.4%	0.7 (0.53–0.91)	0.0086
	Eur	43.6%	33.8%	1.15 (1.04–1.27)	0.0085	38.3%	30.7%	0.89 (0.8–0.99)	0.0287	43.6%	30.7%	0.83 (0.73–0.95)	0.0087
	N Am	46.7%	33.5%	1.18 (1.07–1.31)	0.0010	40.5%	24.1%	0.83 (0.72–0.96)	0.0121	46.7%	24.1%	0.79 (0.67–0.93)	0.0042
	UK	57.0%	41.5%	1.20 (0.96–1.49)	0.10	49.3%	9.1%	0.72 (0.46–1.12)	0.15	57.0%	9.1%	0.66 (0.4–1.1)	0.11
	DAN	55.6%	40.5%	1.37 (1.07–1.75)	0.0137	47.2%	35.3%	0.80 (0.64–1.01)	0.0628	55.6%	35.3%	0.63 (0.45–0.88)	0.0061
	HBDI	48.1%	35.1%	1.25 (1.12–1.40)	4.75E-05	40.0%	33.6%	0.84 (0.75–0.93)	0.0006	48.1%	33.6%	0.75 (0.65–0.85)	1.66E-05
	JOS	74.1%	42.1%	1.66 (1.19–2.31)	0.0028	52.3%	41.2%	0.87 (0.67–1.12)	0.29	74.1%	41.2%	0.53 (0.35–0.82)	0.0042
	SAR	26.3%	57.3%	0.82 (0.52–1.30)	0.41	57.7%	36.7%	0.60 (0.41–0.87)	0.0067	26.3%	36.7%	0.72 (0.41–1.25)	0.24

*
odds ratios refer to the DRB1-DQB1 genotype variable as encoded in Table 2, not only to the DR3/DR4 genotype.

#
outcomes for logistic regression adjusted for gender and probands status and including all five genetic variables defined in Table 2.

Table 4

Validation of the risk prediction models for type 1 diabetes age at onset using HLA information.

	Early age at onset	Discrimination	Calibration
Cohort	model	AUC [95% CI]	Hosmer–Lemeshow p value
T1DGC	gender	0.514 [0.49 – 0.54]	0.63
	HLA	0.595 [0.57 – 0.62][a]	0.78
	gender + HLA	0.598 [0.57– 0.62][a]	0.99
DAN	gender	0.604 [0.52 – 0.68][a]	1.00
	HLA	0.604 [0.51 – 0.69][a]	1.00
	gender + HLA	0.625 [0.54 – 0.71][a]	0.99
SAR	gender	0.398 [0.28 – 0.52]	1.00
	HLA	0.502 [0.36 – 0.64]	0.98
	gender + HLA	0.500 [0.36 – 0.64]	0.81
HBDI	gender	0.535 [0.49 – 0.57]	1.00
	HLA	0.467 [0.48 – 0.66]	1.00
	gender + HLA	0.565 [0.59 – 0.67][a]	0.99
JOS	gender	0.435 [0.33 – 0.54]	1.00
	HLA	0.700 [0.60– 0.80][a]	0.99
	gender + HLA	0.687 [0.58 – 0.79][a]	0.99

	Late age at onset	ROC [95% CI]	Hosmer–Lemeshow p value
T1DGC	gender	0.538 [0.51 – 0.56][a]	0.21
	HLA	0.584 [0.55 – 0.62][a]	0.52
	gender + HLA	0.594 [0.66 – 0.61][a]	0.99
DAN	gender	0.623 [0.55 – 0.69][a]	0.99
	HLA	0.583 [0.50 – 0.67][a]	0.99
	gender + HLA	0.644 [0.56 – 0.72][a]	0.87
SAR	gender	0.430 [0.32 – 0.53]	0.99
	HLA	0.627 [0.51 – 0.74][a]	0.999
	gender + HLA	0.535 [0.40 – 0.67]	0.001
HBDI	gender	0.564 [0.53 – 0.60]	0.001
	HLA	0.615 [0.58 – 0.66][a]	0.29
	gender + HLA	0.618 [0.58 – 0.66]	0.001
JOS	gender	0.506[0.41 – 0.61]	1.00
	HLA	0.600 [0.50 – 0.71][a]	0.99
	gender + HLA	0.602 [0.50 – 0.72][a]	0.28

Models with area under the curve (AUC) for ROC curves significantly   0.50 and non-significant goodness of fit (Hosmer Lemeshow p-value >0.05) are indicated with an superscript[a]