

**Genetic variants predisposing most strongly to type 1 diabetes diagnosed under age 7 years lie near candidate genes that function in the immune system and in pancreatic beta cells.**

*Jamie R.J. Inshaw MSc, Antony J. Cutler PhD, Daniel J.M. Crouch PhD, Linda S. Wicker PhD and John A. Todd PhD*

JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, University of Oxford, United Kingdom

**Running Title:** Genetics of early-diagnosed type 1 diabetes.

**Corresponding authors:** Mr Jamie Inshaw, Professor John Todd

JDRF/Wellcome Diabetes and Inflammation Laboratory,  
Wellcome Centre for Human Genetics,  
NIHR Oxford Biomedical Research Centre,  
Nuffield Department of Medicine,  
Roosevelt Drive,  
Oxford,  
OX3 7BN

01865 287859

[jinshaw@well.ox.ac.uk](mailto:jinshaw@well.ox.ac.uk)

[jatodd@well.ox.ac.uk](mailto:jatodd@well.ox.ac.uk)

**Word Count:** 4070

**Number of Tables:** 1. **Number of Figures:** 3.

# **Abstract**

## **Objective**

Immunohistological analyses of pancreata from patients with type 1 diabetes suggest distinct islet pathology between those diagnosed at <7 years (<7 group) and those diagnosed at age ≥13 years (≥13 group), with both B and T lymphocyte islet inflammation common in children in the <7 group, whereas B cells are rare in the ≥13 group. Based on these observations, we sought to identify differences in genetic susceptibility between these pre-specified age-at-diagnosis groups, to inform on the aetiology of the most aggressive form of type 1 diabetes that initiates in the first years of life.

## **Research Design and Methods**

Using multinomial logistic regression models, we tested if known type 1 diabetes loci (17 within the HLA region and 55 non-HLA regions) had significantly stronger effect sizes in the <7 group compared to the ≥13 group, using genotype data from 27,071 individuals (18,485 controls, 3,121 cases diagnosed at <7, 3,757 at 7-13 and 1,708 at ≥13).

## **Results**

Six HLA haplotypes/classical alleles and six non-HLA regions, one of which functions specifically in beta cells (*GLIS3*), and the other five likely affecting key T cell (*IL2RA*, *IL10*, *IKZF3*, *THEMIS*), thymus (*THEMIS*) and B cell development/functions (*IKZF3*, *IL10*) or in both immune and beta cells (*CTSH*) showed evidence for stronger effects in the <7 group.

## **Conclusions**

A subset of type 1 diabetes associated variants are more prevalent in children diagnosed under the age of 7 and are near candidate genes that act in both pancreatic beta and immune cells.

Type 1 diabetes is a multifactorial disease in which the insulin-producing beta cells of pancreatic islets are destroyed or rendered dysfunctional by an autoimmune process that often initiates in the first few months of life, causing a pre-diabetic, non-symptomatic state in approximately 0.4% of children <sup>1</sup>. However, the actual diagnosis often happens many years after this prodromal phase, the joint environmental and genetic mechanisms of which remain ill defined, with the median age-at-diagnosis being around age 11 years. Even after diagnosis there often remains sufficient endogenous insulin production to lower the requisite levels of insulin treatment and reduce the probability of developing later in life complications <sup>2</sup>. The exceptions to this are the children diagnosed under the age five years in whom there is little insulin production shortly after diagnosis <sup>2,3</sup>. This subgroup represents the largest unmet clinical challenge, since they suffer the greatest complications of the disease <sup>3</sup>. Yet any intervention of type 1 diabetes autoimmunity in these young children must be as safe and precise as possible, targeting and modulating the causative molecules, cells and pathways. Hence we need to identify the specific mechanisms underlying early-diagnosed type 1 diabetes.

Recent evidence suggests that children diagnosed under 7 years of age may have a different, more aggressive form of islet inflammation (insulitis), characterised by a B lymphocyte infiltrate coincident with a T cell insulitis (CD4<sup>+</sup> and CD8<sup>+</sup> T cells), compared with children aged 13 years and above, who have reduced B cell participation <sup>4</sup>. In cases diagnosed between 7 and 12 years there is a mixture of islet infiltrate phenotypes, some with the “under 7” B cell infiltrate and others with “13 and over” phenotype.

There is already evidence that autoantigen-presenting genes HLA class II and class I are associated with reduced age-at-diagnosis, which provides insight into the biology

of this most beta-cell destructive form of the disease <sup>5-8</sup>. More recently, a genome-wide association analysis of age-at-diagnosis of type 1 diabetes identified a locus on chromosome 6q22.33 that acts almost exclusively in type 1 diabetes cases diagnosed under age 5 years <sup>9</sup>, encoding the protein tyrosine phosphatase receptor kappa (PTPRK) and the thymocyte-expressed molecule involved in selection (THEMIS) genes. However, this approach had to apply a stringent genome-wide multiple testing correction criterion ( $p < 5 \times 10^{-8}$ ) and informative, true signals were likely to have been missed. In the present study, we analysed only the association of specified known type 1 diabetes risk regions, thereby reducing the multiple testing burden. In addition, we use existing knowledge of differences in pancreatic immune cell infiltrates between those diagnosed at  $<7$  years compared to  $\geq 13$  years to stratify patients according to their age-at-diagnosis, and examine genetic differences between these groups of individuals, as opposed to treating age-at-diagnosis as a continuous phenotype.

If type 1 diabetes has a particular pancreatic immunophenotype then it might be expected for it to have distinct genetic features, characterised by susceptibility genes with larger effects in the under 7 group. Moreover, the intermediate group, age-at-diagnosis 7-13 years, would have risk for these age-at-diagnosis-sensitive genes lying between the under 7's and the 13's and over.

## Research Design and Methods

### Study populations

Our dataset consisted of 18,485 controls, 3,121 type 1 diabetes cases diagnosed at <7 years ('<7 group'), 3,757 at  $\geq 7$  to <13 years ('7-13 group') and 1,708 at  $\geq 13$  years (' $\geq 13$  group'). The majority of individuals were from the UK (Table 1), although individuals recruited from countries with different ancestries were also included (Supplementary Figure 1). Related individuals were removed from the analysis (Supplementary approaches).

The majority of cases in the  $\geq 13$  group were diagnosed before the typical age at type 2 diabetes diagnosis (~ 45 years), with 68% diagnosed at  $\leq 18$  years and 99.9% diagnosed at  $\leq 45$  years. To further increase confidence that none of the cases in the oldest age-at-diagnosis group were type 2 diabetics, we took all index variants associated with type 2 diabetes<sup>10</sup> that were also present on the ImmunoChip (n=30) and built a genetic risk score (GRS) to compare the mean score across all groups included in the analysis (Supplementary approaches). The mean type 2 diabetes GRS were similar between all groups (controls=2.47, <7 group=2.49, 7-13 group=2.48,  $\geq 13$  group=2.47) and there were no individuals in the  $\geq 13$  group with a type 2 diabetes GRS above that of the highest value amongst controls (3.47).

### Loci studied

We examined eight HLA class II haplotypes and nine HLA class I classical alleles for their association with type 1 diabetes diagnosed within each age group, which were a subset of the most type 1 diabetes-associated haplotypes identified to date<sup>11</sup> that we also found to be associated with type 1 diabetes in our analysis after conditioning on the other associated HLA haplotypes (logistic regression Wald test  $p < 0.01$ ).

Supplementary Table 1 summarises which haplotypes and classical alleles were

examined, how they were defined and whether they were common enough to include in our analysis, defined as having at least five individuals from each group carrying the classical allele/haplotype.

We also examined 55 loci outside the HLA, which have previously shown association with type 1 diabetes either in univariable analyses or via fine mapping (Supplementary Table 2). Each locus contains an 'index' variant, chosen to be the most strongly disease associated from a set of variants in linkage disequilibrium (LD) that constitute a single genetic signal. We allocated locus names to each of these variants based on a candidate gene(s), but the named genes may not be causal for type 1 diabetes.

### **Imputation**

To impute classical HLA alleles, the HiBag<sup>12</sup> R Bioconductor package was used with the classifiers as calculated from the training dataset in the original publication.

Alleles called with a probability of less than 0.5 were treated as missing. Some individuals were genotyped for a subset of their classical HLA alleles<sup>8</sup> so accuracy of imputation was assessed for a proportion of individuals.

When variants of interest failed genotype quality control (Supplementary approaches), we imputed those variants plus 0.5 Mb surrounding regions using the IMPUTE2<sup>13</sup> software; with the 1000 Genomes Project data as the reference haplotypes. We used this same imputation strategy to impute around index variants for fine mapping non-HLA regions found to be differentially-associated between the <7 and ≥13 groups.

### **Identifying loci with larger effects in the <7 compared to the ≥13 group**

To examine heterogeneity in effect size between age-at-diagnosis groups, we used a method analogous to performing three case-control analyses, comparing the

prevalence of the allele under consideration in controls to the <7 group, 7-13 group and  $\geq 13$  group, respectively, estimating an effect size for each group. The difference in the estimated effect size of the genetic variant between the first and third analyses were used to determine the level of evidence (P-value) for heterogeneity in the effect size of the variant between the youngest and oldest age-at-diagnosis groups, and the effect size in the second analysis (7-13) provided additional supporting evidence. More precisely: we fitted two multinomial logistic regressions per locus, one allowing for different effect sizes for the genetic variant at that locus in the <7 and  $\geq 13$  groups and the other constraining the effect size for the genetic variant in the <7 and  $\geq 13$  groups to equal each other. We compared the likelihoods of these models using a likelihood ratio test <sup>14</sup>, and if the model allowing different effect sizes fitted the data better than the model constraining the effect sizes to equal each other, the locus was considered heterogeneous in effect size between age-at-diagnosis groups. Both models included as covariates sex and the ten largest principal components derived from the set of ImmunoChip variants passing quality control filters (Supplementary approaches). This analysis was performed using the multinomRob R package <sup>15</sup>. For HLA loci, we additionally adjusted for other HLA haplotypes/alleles to account for the high levels of LD in the region. When examining the HLA class II haplotype effects other than DR3-DQ2/DR4-DQ8, we examined only individuals without the DR3-DQ2/DR4-DQ8 diplotype to remove any confounding effects of this diplotype. Supplementary Table 1 shows which individuals were included in each analysis and which classical haplotypes/alleles were adjusted for in each analysis.

### **Sensitivity analyses – non-HLA results**

To exclude the possibility of spurious associations due to population structure in our data, we repeated the analysis using only individuals from the UK and Northern



Ireland and adjusted for sex and the five largest genetic principal components in these individuals only. Additionally, to test sensitivity of our results to age-strata thresholds, we performed the same analysis but instead compared individuals diagnosed at  $<6$  years to the  $\geq 13$  group and also individuals diagnosed at  $<5$  years compared to the  $\geq 13$  group.

We declared a locus differentially-associated if the heterogeneity p-value was associated to a False Discovery Rate (FDR) of  $<0.1$  (Supplementary approaches). To explore whether there were more age-at-diagnosis associated variants which we could not detect in the present analysis due to a lack of statistical power, we examined all loci which did not reach the association threshold ( $\text{FDR} < 0.1$ ) and counted how many loci had the largest effect in the  $<7$  group, the intermediate effect in the 7-13 group and the smallest effect in the  $\geq 13$  group and compared this to the expected frequency of this ordering using a binomial test (Supplementary approaches).

### **Fine mapping and colocalisation analyses**

For the non-HLA loci with the evidence of heterogeneity in effect size between age-at-diagnosis groups ( $\text{FDR} < 0.1$ ), we fine mapped a 0.5 Mb region around the index variant to identify potentially causal variants for type 1 diabetes diagnosed at  $<7$  years. Analysis was limited to individuals from the UK and Northern Ireland, amounting to 2,884 cases diagnosed at  $<7$  years and 11,071 controls, in order to examine a relatively homogeneous population, as fine mapping is sensitive to differences in LD structure between ancestrally divergent groups. We used GUESSFM,<sup>16</sup> which carries out a Bayesian variable selection stochastic search to identify the combinations of variants constituting separate genetic susceptibility to type 1 diabetes.

For those fine-mapped regions with high ( $>0.8$ ) posterior probability of a single causal variant in the region, we conducted colocalisation analyses with expression quantitative trait loci (eQTL) associations in whole blood from a dataset of over 30,000 individuals<sup>17</sup>. This enabled us to estimate which genes the most likely causal variants are regulating and what direction the effects are on gene transcription and disease risk. The coloc R package was used to carry out this analysis<sup>18</sup> (Supplementary approaches).

### **Heritability estimates by age-at-diagnosis group**

To estimate how the proportion of phenotypic variance could be explained by all variants on the ImmunoChip in each age-at-diagnosis group, we compared chip heritability estimates between type 1 diabetes cases diagnosed at  $<7$  years to those diagnosed at 7-13 years or  $\geq 13$  years. We used the GCTA software (<https://cnsgenomics.com/software/gcta/#Overview>) to fit linear mixed models and estimate heritability for each age-at-diagnosis group with a shared set of controls. We estimated heritability,  $h_g^2$ , on the liability scale, as derived in<sup>19</sup>. We adjusted for sex and the ten largest genetic principal components and included individuals from all ancestry backgrounds in the analysis. We repeated the analysis but excluded the entirety of the HLA region. The prevalence of type 1 diabetes may vary by age; we assumed a prevalence of 0.4% in all age groups in the primary analysis but tested stability of the estimates by performing sensitivity analyses where the assumed prevalence in the  $<7$  group was higher at 0.5% and also, in separate analyses, calculated the heritability with an assumed prevalence of 0.2% and 0.3% in both the 7-13 and  $\geq 13$  groups.

### **Code availability**

The scripts used to carry out these analyses are available at

[https://github.com/jinshaw16/AAD\\_t1d](https://github.com/jinshaw16/AAD_t1d).

## Results

### Multinomial logistic regression: HLA

Six HLA variables were differentially-associated between the <7 and  $\geq 13$  group (with  $\text{FDR} < 0.1$ ). The most strongly differentially-associated locus between the age-at-diagnosis groups was the DR3-DQ2/DR4-DQ8 diplotype, where the diplotype was more prevalent in the <7 group compared to the  $\geq 13$  group, whilst the protective DRB1\*15:01-DQB1\*06:02 and DRB1\*07:01-DQB1\*03:03 haplotypes encoded greater protection from type 1 diabetes in the <7 group compared to the  $\geq 13$  group. Class I alleles A\*24:02 and B39\*06 showed more susceptibility to type 1 diabetes in the <7 compared to and  $\geq 13$  group (Figure 1). Comparison of imputed classical 4 digit HLA alleles with directly genotyped 4 digit HLA alleles showed concordance of over 91% for each gene examined (Supplementary Figure 2).

### Multinomial logistic regression: non-HLA regions

Outside the HLA, seven regions were differentially-associated between the <7 and  $\geq 13$  group ( $\text{FDR} < 0.1$ ), near Cathepsin H (*CTSH*), GLIS family zinc finger 3 (*GLIS3*), Ikaros family zinc finger 3 (*IKZF3*), Chymotrypsinogen B1 (*CTRB1*), the third index variant at interleukin 2 receptor alpha (*IL-2RA*), *THEMIS* and interleukin-10 (*IL-10*), (Figure 2), with *CTSH* surviving Bonferroni correction ( $p < 0.05/55 = 0.00091$ ). At each locus associated with  $\text{FDR} < 0.1$ , the 7-13 group had a larger effect size than the  $\geq 13$  group and smaller than the <7 group. Given the  $\geq 13$  group comprises just 1,708 individuals, it is probable that with increased sample size and hence statistical power, other type 1 diabetes risk loci might reach statistical significance with regards to heterogeneity (Supplementary Figure 3). Of the 48 variants not satisfying an  $\text{FDR} < 0.1$ , 21 had the strongest signal in <7s, weakest in  $\geq 13$ s and intermediate in 7s-13s, compared to eight occurrences in that order expected by chance ( $p = 9.74 \times 10^{-6}$ ,

binomial test), suggesting the presence of substantial additional signal overall despite not showing evidence individually.

### **Stability of non-HLA results**

In the UK-specific sensitivity analysis, performed to ensure no loci were declared differentially-associated due only to differences in ancestry between groups, we found five of the seven loci declared heterogeneous from the primary analysis were heterogeneous between the  $<7$  and  $\geq 13$  group ( $\text{FDR} < 0.1$ ) (Supplementary Figure 4). The locus near *CTRB1* showed no differential association between age-at-diagnosis groups ( $p=0.272$ ) and was thus removed from the set of differentially-associated loci, whilst the locus near *IL10* had a p-value of 0.07, which we considered differentially associated, given the decrease in statistical power in this sensitivity analysis.

In addition, when changing the threshold for the early-diagnosed group to  $<6$  and  $<5$  to ensure loci were not declared differentially-associated due only to a chance association observed due to the choice of age-at-diagnosis group cut-off, all six associated loci from the primary analysis and UK-specific analysis were heterogeneous ( $\text{FDR} < 0.1$ ) (Supplementary Figures 5 and 6).

Minor allele frequency plots by age-at-diagnosis for the six differentially-associated loci are shown in Supplementary Figures 7-12, whilst Supplementary Tables 3 and 4 summarise the most likely causal genes at these loci.

### **Fine mapping**

Fine-mapping analyses showed that five of the six differentially-associated loci had high posterior probability of a single causal variant in the region ( $>0.8$ ), whilst *IL2RA* had a higher posterior probability of having more than one causal variant in the region. Variants prioritised in the GUESSFM analysis for each region examined are

listed in Supplementary Tables 5-10, though it is possible that variants excluded due to low imputation quality that are in LD with these variants could also be causal.

Three of the five regions fine-mapped with high posterior probability of one causal variant in the region showed evidence of colocalisation with whole-blood eQTLs. The *CTSH* locus showed evidence of colocalisation with a *CTSH* eQTL (posterior probability of colocalisation=0.998); the minor protective allele for type 1 diabetes is associated with decreased expression of *CTSH*, or equivalently, the major allele is the type 1 diabetes risk allele and is associated with increased *CTSH* expression. The *IKZF3* locus fine mapping results prioritised an LD block containing 103 variants, all of which could be causal. This same block of variants is also associated with altered expression of at least three genes ( $p < 5 \times 10^{-150}$ ), with evidence of colocalisation of disease signal and whole-blood eQTL for *IKZF3*, *GSDMB* and *ORMDL3* (posterior probability of colocalisation for type 1 diabetes and eQTL with *IKZF3*=0.982, *GSDMB*=0.981 and *ORMDL3*=0.981); the minor alleles at the block of variants associated with a decrease in type 1 diabetes risk are associated with decreased *IKZF3* expression and, in contrast, with increased expression of *GSDMB* and *ORMDL3*.

Finally, the most likely causal variants in the *THEMIS* region colocalise with a *THEMIS* whole-blood eQTL (posterior probability of colocalisation=0.953); the minor allele is protective for type 1 diabetes and is associated with decreased *THEMIS* expression (Figure 3). There was minimal evidence of colocalisation between the disease risk variants and *PTPRK* expression in whole blood (posterior probability of colocalisation=0.023).

There was no evidence of colocalisation between disease risk variants and whole blood eQTL for any genes in the region of *GLIS3* or *IL10* (posterior probability of colocalisation<0.02, Supplementary Figures 13-16).

### **Heritability estimates by age-at-diagnosis group**

We found the chip heritability estimate on the liability scale to be highest in the <7 group, intermediate in the 7-13 group and lowest in the  $\geq 13$  group (<7  $h_g^2=0.366$ , 7-13  $h_g^2=0.301$ ,  $\geq 13$   $h_g^2=0.233$ ). This trend remained when altering the assumed disease prevalence by age-at-diagnosis group and when excluding the HLA region (Supplementary Table 11).

## Conclusions

The stratification of patients by age-at-diagnosis according to islet phenotypes has provided a rich source of candidate genes and corresponding pathways with greater effects in children diagnosed with type 1 diabetes under age 7 years. We expected to see differential associations with the HLA class II haplotypes, in particular the heterozygous diplotype DR3-DQ2/DR4-DQ8, as well as HLA class I alleles, A\*24:02 and B\*39:06<sup>5-8</sup>. Here, we show for the first time that the protective HLA class II haplotypes DRB1\*15:01-DQB1\*06:02 and DRB1\*07:01-DQB1\*03:03 are less prevalent amongst individuals diagnosed at <7 years compared with those diagnosed at ≥13 years. Therefore, the earliest and most aggressive phenotypic subtype of type 1 diabetes results primarily from carriage of high risk alleles of the HLA class II and I genes. These likely act at one or more of four levels: (i) altering the T cell receptor repertoire in favour of anti-islet antigen reactivity and/or reducing the protective repertoire of T regulatory cells; (ii) providing a strong autoantigen presentation environment in the islets, enabling the infiltration and cytolytic activity of CD8<sup>+</sup> T cells but also by disrupting B cell anergy,<sup>20</sup> permitting binding and presentation of autoantigen to provide potent help to T cells; (iii) affecting the immune response to the viral infections that are involved in the disease; (iv) affecting how the gut microbiome develops in early life, a system that is known to affect type 1 diabetes susceptibility<sup>21</sup>.

In addition to the HLA heterogeneity, we obtained robust evidence of differences in effect size between the age-at-diagnosis groups at six non-HLA loci. Of these loci, one plausible candidate gene, *GLIS3*, most likely acts in the islet beta cells, given the expression levels in the pancreas, lack of expression in immune cells, colocalisation with type 2 diabetes risk variants<sup>22</sup> and lack of association with other autoimmune



diseases (<https://genetics.opentargets.org>). Genes in two of the loci, *CTSH* and *IKZF3*, could act in the islets or elsewhere, whilst all of the other candidate causal genes (*IL2RA*, *IL10*, *THEMIS*, *IKZF3/ORMDL3/GSDMB* and *CTSH*) have known functions in T and/or B cell biology (Supplementary Table 4). This implies that in addition to HLA-susceptibility, risk of type 1 diabetes in the very young is also impacted by particular malfunctions in the infiltrating T and B cells, resulting in a perfect storm of immune infiltration, antigen recognition and a rapid destruction of beta cells.

Of the non-HLA risk regions with the strongest evidence of heterogeneity between age-at-diagnosis groups, we focus on *CTSH*, *IKZF3* and *THEMIS*, which colocalise with whole blood eQTLs. The minor T allele at one of the candidate causal variants at the *CTSH* locus, rs2289702 (C>T) is associated with protection from type 1 diabetes and decreased expression of CTSH mRNA in multiple cell types and tissues, or equivalently, the major C allele is the type 1 diabetes risk allele and is associated with increased expression of CTSH (Supplementary Table 4). The locus has previously been implicated in type 1 diabetes aetiology by altering sensitivity of beta cells to apoptosis<sup>23</sup>, where rs3825932 (C>T) was investigated, which is in low LD ( $r^2=0.26$ ) with the disease-associated variant reported here. However, the type 1 diabetes risk allele counter-intuitively resulted in protection from beta-cell apoptosis, thus, beta-cell apoptosis may not be the primary mechanism underlying disease aetiology in this region. *CTSH* functions as an endopeptidase and can cleave the N-terminus of the Toll-like receptor 3 (TLR3) protein, increasing its functionality<sup>24</sup>. Given TLR3 is expressed in islets<sup>25</sup>, it is possible that the increase in CTSH expression associated with the type 1 diabetes susceptibility allele results in increased TLR3 N-terminus cleavage, heightened responses to viral infections and increased release of type 1 interferon. This may increase baseline risk of type 1 diabetes and specifically the risk

of early-diagnosed type 1 diabetes in individuals carrying this allele, since viral infections are more frequent in childhood.

The large LD block in the *IKZF3* region is associated with multiple diseases, including asthma and paediatric asthma<sup>26, 27</sup>. The direction of effect of the risk variant is opposite in asthma to autoimmune diseases, where the C allele at a variant within the haplotype, rs921649 (C>T), increases susceptibility to autoimmunity, but protects against asthma<sup>26</sup>. The three genes most strongly associated with the haplotype are expressed in lymphocytes (<https://dice-database.org/>), with good biological candidacy for altering disease risk. *IKZF3* is a transcriptional repressor with a key role in B-cell activation and differentiation<sup>28</sup> and T cell differentiation<sup>29</sup>. *ORMDL3* is a central regulator of sphingolipid biosynthesis<sup>30</sup> and has also been proposed to negatively regulate store-operated calcium, lymphocyte activation and cytokine production<sup>26,31</sup>, while *GSDMB* can act as a pyroptotic protein<sup>32</sup>. Therefore one or more of these genes may be causal for type 1 diabetes risk.

The colocalisation between the type 1 diabetes disease association and a *THEMIS* whole blood eQTL points towards *THEMIS* being the likely causal gene in this region. The minor T allele at one of the candidate causal variants, rs13204742 (G>T) is protective for early-diagnosed type 1 diabetes and coeliac disease<sup>33</sup>, but susceptible for irritable bowel disease<sup>34</sup>. *THEMIS*, expressed in thymocytes and circulating T cells (Supplementary Table 4), is a key signalling molecule for T cell development and survival<sup>35</sup>. In the thymus *THEMIS* sets signalling thresholds at the double positive stage of thymocyte development and influences selection of T cells. Deletion of *THEMIS* reduces transition of double positive to single positive thymocytes<sup>36,37</sup>. Our results suggest that an increase in *THEMIS* expression leads to increased risk of early-diagnosed type 1 diabetes. If we assume the whole blood eQTL with the disease

risk variants is mirrored in the thymus, we hypothesise that the increased *THEMIS* expression would alter the threshold for positive selection and increase the probability of autoreactive T cells entering circulation.

The increased heritability of early-diagnosed compared to later-diagnosed type 1 diabetes should be interpreted with caution given the ImmunoChip design is primarily focussed on immune regions and designed to capture regions of interest in autoimmune diseases. It is possible that those individuals diagnosed with type 1 diabetes later in life have a different profile of susceptibility regions that are not captured on the ImmunoChip due to insufficient statistical power to detect such loci in previous GWAS studies. Nevertheless, it is interesting to note that amongst the susceptibility regions for type 1 diabetes discovered to date, there appears to be a higher heritability in those diagnosed at a young age compared to at  $\geq 13$  years, consistent with increasing age accompanying greater risk and length of exposure to environmental type 1 diabetes causal factors.

One weakness of our analysis is the relatively small sample size in the  $\geq 13$  group compared to the  $< 7$  and 7-13 groups. Increasing the sample size of this group would lead to more accurate estimates of effect sizes of the variants on type 1 diabetes risk at  $\geq 13$  years and thus more robust heterogeneity test results. However, the observed trend, whereby the largest effect tends to be in the  $< 7$  group, the intermediate effect in the 7-13 group and the smallest effect in the  $\geq 13$  group increases confidence in the results, as the 7-13 group is well powered and alone shows several differences in effect sizes with the youngest group.

We have highlighted a number of genetic regions most strongly associated with early-diagnosed type 1 diabetes. The candidate causal genes within these regions suggest early diagnosis could be driven by a fully integrated pathogenic collaboration between

the immune system, the beta cells and viral infection in the initiation and rapid development of extreme insulin-deficiency.

## Acknowledgments

We gratefully acknowledge all participants for allowing the analysis of their genotypic and phenotypic data.

Author contributions: JRJI carried out the analyses and drafted the main body of the manuscript. AJC provided assistance with the manuscript writing and carried out a thorough literature search for age-at-diagnosis associated regions. DJMC provided statistical support and guidance for JRJI as well as critically reviewing the manuscript. LSW provided critical feedback for the manuscript and immunological support and context. JAT supervised the work, proposed the research hypothesis and critically evaluated the manuscript.

JAT is the guarantor for this manuscript.

No authors have any conflict of interests to declare.

This work was funded by the Juvenile Diabetes Research Foundation (JDRF) (9-2011-253, 5-SRA-2015-130-A-N) and Wellcome (091157, 107212) to the Diabetes and Inflammation Laboratory, University of Oxford.

We use data generated by the Wellcome Trust Case Control Consortium (076113).

The Northern Irish Genetic Resource Investigating Diabetes (GRID), Tyypin 1

Diabetekseen Sairastuneita Perheenjäsenineen<sup>[1]</sup><sub>SEP</sub>(IDDMGEN), Tyypin 1

Diabetekseen Genetiikka (T1DGEN) and Warren cohorts were genotyped using the Type 1 Diabetes Genetics Consortium (T1DGC) grants from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the National Institute of Allergy and Infectious Diseases (NIAID), the National Human Genome Research Institute (NHGRI), the National Institute of Child Health and Human Development (NICHD) and the JDRF (U01 DK062418, JDRF 9-2011-530). The research was

supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC.

## References

1. Krischer, J. P. *et al.* Predicting Islet Cell Autoimmunity and Type 1 Diabetes: An 8-Year TEDDY Study Progress Report. *Diabetes Care* **42**, 1051–1060 (2019).
2. Kuhtreiber, W. M. *et al.* Low levels of C-peptide have clinical significance for established Type 1 diabetes. *Diabet. Med.* **32**, 1346–1353 (2015).
3. Rawshani, A. *et al.* Excess mortality and cardiovascular disease in young adults with type 1 diabetes in relation to age at onset : a nationwide , register-based cohort study. *Lancet* **392**, 477–486 (2018).
4. Leete, P. *et al.* Differential insulinitic profiles determine the extent of  $\beta$ -cell destruction and the age at onset of type 1 diabetes. *Diabetes* **65**, 1362–1369 (2016).
5. Howson, J. M. M. *et al.* Evidence of gene-gene interaction and age-at-diagnosis effects in type 1 diabetes. *Diabetes* **61**, 3012–3017 (2012).
6. Valdes, A. M. *et al.* Use of class I and class II HLA loci for predicting age at onset of type 1 diabetes in multiple populations. *Diabetologia* **55**, 2394–2401 (2012).
7. Nejentsev, S. *et al.* Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887–892 (2007).
8. Howson, J. M. M., Walker, N. M., Clayton, D. & Todd, J. A. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. *Diabetes, Obes. Metab.* **11**, 31–45 (2009).
9. Inshaw, J. R. J., Walker, N. M., Wallace, C., Bottolo, L. & Todd, J. A. The chromosome 6q22.33 region is associated with age at diagnosis of type 1

- diabetes and disease risk in those diagnosed under 5 years of age. *Diabetologia* **61**, 147–157 (2018).
10. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
  11. Noble, J. A. & Valdes, A. M. Genetics of the HLA Region in the Prediction of Type 1 Diabetes. *Curr. Diab. Rep.* **11**, 533–542 (2011).
  12. Zheng, X. *et al.* HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* **14**, 192–200 (2014).
  13. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, (2009).
  14. Morris, A. P. *et al.* A Powerful Approach to Sub-Phenotype Analysis in Population-Based Genetic Association Studies. *Genet. Epidemiol.* **34**, 335–343 (2010).
  15. Mebane, W. & Sekhon, J. *Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data.* (2004).
  16. Wallace, C. *et al.* Dissection of a Complex Disease Susceptibility Region Using a Bayesian Stochastic Search Approach to Fine Mapping. *PLOS Genet.* **11**, e1005272 (2015).
  17. Vösa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL meta- analysis. *bioRxiv* 1–57 (2018).
  18. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).



19. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
20. Smith, M. J. *et al.* Loss of B-Cell Anergy in Type 1 Diabetes Is Associated With High-Risk HLA and Non-HLA Disease Susceptibility Alleles. *Diabetes* **67**, 697–703 (2018).
21. Paun, A. *et al.* Association of HLA-dependent islet autoimmunity with systemic antibody responses to intestinal commensal bacteria in children. *Sci. Immunol.* **4**, eaau8125 (2019).
22. Aylward, A., Chiou, J., Okino, M.-L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum. Mol. Genet.* (2018). doi:10.1093/hmg/ddy314
23. Fløyel, T., Brorsson, C., Nielsen, L. B., Miani, M. & Bang-berthelsen, C. H. CTSH regulates  $\beta$ -cell function and disease progression in newly diagnosed type 1 diabetes patients. *PNAS* **111**, 10305–10310 (2014).
24. Qi, R., Singh, D. & Kao, C. C. Proteolytic Processing Regulates Toll-like Receptor 3 Stability and Endosomal Localization. *J. Biol. Chem.* **287**, 32617–32629 (2012).
25. Rasschaert, J. *et al.* Toll-like Receptor 3 and STAT-1 Contribute to Double-stranded RNA+ Interferon-gamma-induced Apoptosis in Primary Pancreatic beta-Cells. *J. Biol. Chem.* **280**, 33984–33991 (2005).
26. Schmiedel, B. J. *et al.* 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat. Commun.* **7**, (2016).
27. Bouzigon, E. *et al.* Effect of 17q21 Variants and Smoking Exposure in Early-Onset Asthma. *N. Engl. J. Med.* **359**, 1985–1994 (2019).

28. Wang, J. *et al.* Aiolos Regulates B Cell Activation and Maturation to Effector State. *Immunity* **9**, 543–553 (1998).
29. Quintana, F. J. *et al.* Aiolos promotes TH17 differentiation by directly silencing Il2 expression. *Nat. Immunol.* **13**, 770–777 (2012).
30. Davis, D., Kannan, M. & Wattenberg, B. Orm / ORMDL proteins: Gate guardians and master regulators. *Adv. Biol. Regul.* **70**, 3–18 (2018).
31. Carreras-Sureda, A. *et al.* ORMDL3 modulates store-operated calcium entry and lymphocyte activation. *Hum. Mol. Genet.* **22**, 519–530 (2013).
32. Panganiban, R. A. *et al.* A functional splice variant associated with decreased asthma risk abolishes the ability of gasdermin B to induce epithelial cell pyroptosis. *J. Allergy Clin. Immunol.* **142**, 1469–1478 (2018).
33. Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
34. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
35. Mehta, M. *et al.* Themis-associated phosphatase activity controls signaling in T cell development. *Proc. Natl. Acad. Sci.* **115**, E11331–E11340 (2018).
36. Johnson, A. L. *et al.* Themis is a member of a new metazoan gene family and is required for the completion of thymocyte positive selection. *Nat. Immunol.* **10**, 831–839 (2009).
37. Fu, G. *et al.* Themis controls thymocyte selection through regulation of T cell antigen receptor-mediated signaling. *Nat. Immunol.* **10**, 848–856 (2009).

## Tables

	<b>Controls</b>	<b>T1D individuals diagnosed at &lt;7 years</b>	<b>T1D individuals diagnosed at 7-13 years</b>	<b>T1D individuals diagnosed at ≥13 years</b>
N	18485	3121	3757	1708
Age-at-diagnosis (years): Median (IQR) [range]	NA	4 (2, 5) [0, 6.9]	10 (8, 11) [7, 12.9]	15 (13, 21) [13, 50.3]
Sex: Female	9771 (52.9%)	1528 (49.0%)	1968 (52.4%)	743 (43.5%)
Asia-Pacific	925 (5%)	28 (0.9%)	26 (0.7%)	52 (3.0%)
Central Europe	1681 (9.1%)	52 (1.7%)	63 (1.7%)	156 (9.1%)
Finland	2824 (15.3%)	100 (3.2%)	152 (4.0%)	434 (25.4%)
Northern Ireland	478 (2.6%)	222 (7.1%)	248 (6.6%)	35 (2.0%)
UK	10593 (57.3%)	2662 (85.3%)	3206 (85.3%)	923 (54%)
USA	1984 (10.7%)	57 (1.8%)	62 (1.7%)	108 (6.3%)

**Table 1:** Characteristics of individuals included in the analysis.

## Figure Legends

**Figure 1:** Classical HLA haplotype/alleles association with type 1 diabetes diagnosed at <7 years old (red circle; mean log-odds ratio age-at-diagnosis  $\pm$  95%CI), 7-13 years old (green circle; mean log-odds ratio age-at-diagnosis 7-13 $\pm$  95%CI) and  $\geq$ 13 years old (blue circle mean log-odds ratio age-at-diagnosis  $\geq$ 13 $\pm$  95%CI), from a multinomial logistic regression. Left panel shows the log-odds ratios with a dashed red line showing a log-odds ratio of 0. The right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic regression constraining the log-odds ratios from the <7 and  $\geq$ 13 groups to be equal compared to an unconstrained model. Red dotted line shows nominal significance in heterogeneity ( $p < 0.05$ ), red dashed line show Bonferroni-corrected significance in heterogeneity.

**Figure 2:** Non-HLA type 1 diabetes associated variants, showing on the left hand panel the log-odds ratios for the minor allele for those diagnosed at <7 years old (red circle; log-odds ratio age-at-diagnosis  $\pm$  95%CI), 7-13 years old (green circle; log-odds ratio age-at-diagnosis 7-13 $\pm$  95%CI) and  $\geq$ 13 years old (blue circle log-odds ratio age-at-diagnosis  $\geq$ 13 $\pm$  95%CI), from a multinomial logistic regression; the dashed red line shows a log-odds ratio of 0. The right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic regression constraining the log-odds ratios from the <7 and  $\geq$ 13 groups to be equal compared to an unconstrained model. Red dotted line shows threshold for false discovery rate of  $< 0.1$ , red dashed line shows threshold for Bonferroni-corrected heterogeneity. Showing only loci with a false discovery rate of less than 0.1.

**Figure 3:** Results from colocalisation and fine mapping in the *IKZF3* region (left) *CTSH* region (centre) and THEMIS region (right). Analyses include individuals from the UK and Northern Ireland and only controls and cases diagnosed at <7 years. First panel: association absolute z scores from whole blood eQTL study examining variant effects on *IKZF3* (left), *CTSH* (centre) and *THEMIS* (right) mRNA levels, coloured by LD  $r^2$  to the most strongly associated variant with the respective mRNA expression. Second panel: association absolute z scores from logistic regression examining variant associations with type 1 diabetes risk at < 7 years, coloured by LD  $r^2$  to the most strongly associated variant with *IKZF3* (left), *CTSH* (centre) and *THEMIS* (right) mRNA. Third panel: Gene positions (genome build 37), with arrows indicating direction of transcription. Fourth panel, univariable early-diagnosed type 1 diabetes log-odds ratios and 95% confidence intervals for the minor allele for each of the most likely causally associated variants as prioritised by GUESSFM. Fifth panel:  $\log_e(\text{absolute eQTL z score})$  if z score > 0 and  $-\log_e(\text{absolute eQTL z score})$  if z score < 0 for the effect of the minor allele, so direction of effect can be compared, including only eQTLs with a p value of  $< 5 \times 10^{-150}$  (left),  $< 5 \times 10^{-50}$  (centre) and  $< 5 \times 10^{-25}$  (right). The symbols are coloured red if contained in the set of most likely causal variants, as produced by GUESSFM and the shape corresponds to the gene that the variant is effecting transcription of, with the genes shown in the centre of the figure.

**Supplementary Figure 1:** Largest genetic principal component against second-largest genetic principal component generated using all individuals included in the primary analysis, coloured by country of enrolment.

**Supplementary Figure 2:** Concordance of HiBag imputed versus directly genotyped classical HLA alleles. Concordance is defined as identical 4 digit HLA classical allele at both chromosomes.

**Supplementary Figure 3:** All 55 non-HLA type 1 diabetes associated variants, showing on the left hand panel the log-odds ratios for the minor allele for those diagnosed at <7 years old (red circle; log-odds ratio age-at-diagnosis  $\pm$  95%CI), 7-13 years old (green circle; log-odds ratio age-at-diagnosis 7-13  $\pm$  95%CI) and  $\geq$ 13 years old (blue circle log-odds ratio age-at-diagnosis  $\geq$ 13  $\pm$  95%CI), from a multinomial logistic regression; the dashed red line shows a log-odds ratio of 0. The right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic regression constraining the log-odds ratios from the <7 and  $\geq$ 13 groups to be equal compared to an unconstrained model. Red dotted line shows threshold for nominally significant heterogeneity between groups ( $p < 0.05$ ), red solid line shows threshold for false discovery rate of  $< 0.1$ , red dashed line shows threshold for Bonferroni-corrected significant heterogeneity.

**Supplementary Figure 4:** All 55 non-HLA type 1 diabetes associated variants examined using only individuals from the UK or Northern Ireland, showing on the left hand panel the log-odds ratios for the minor allele for those diagnosed at <7 years old (red circle; log-odds ratio age-at-diagnosis  $\pm$  95%CI), 7-13 years old (green circle; log-odds ratio age-at-diagnosis 7-13  $\pm$  95%CI) and  $\geq$ 13 years old (blue circle log-odds ratio age-at-diagnosis  $> 13 \pm$  95%CI), from a multinomial logistic regression; the dashed red line shows a log-odds ratio of 0. The right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic

regression constraining the log-odds ratios from the  $<7$  and  $\geq 13$  groups to be equal compared to an unconstrained model. Red dotted line shows threshold for nominally significant heterogeneity between groups ( $p < 0.05$ ), red solid line shows threshold for false discovery rate of  $< 0.1$ , red dashed line shows threshold for Bonferroni-corrected significant heterogeneity.

**Supplementary Figure 5:** All 55 non-HLA type 1 diabetes associated variants, showing on the left hand panel the log-odds ratios for the minor allele for those diagnosed at  $<6$  years old (red circle; log-odds ratio age-at-diagnosis  $\pm$  95%CI), 6-13 years old (green circle; log-odds ratio age-at-diagnosis 6-13  $\pm$  95%CI) and  $\geq 13$  years old (blue circle log-odds ratio age-at-diagnosis  $\geq 13 \pm$  95%CI) from a multinomial logistic regression; the dashed red line shows a log-odds ratio of 0. The right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic regression constraining the log-odds ratios from the  $<6$  and  $\geq 13$  groups to be equal compared to an unconstrained model. Red dotted line shows threshold for nominally significant heterogeneity between groups ( $p < 0.05$ ), red solid line shows threshold for false discovery rate of  $< 0.1$ , red dashed line shows threshold for Bonferroni-corrected significant heterogeneity.

**Supplementary Figure 6:** All 55 non-HLA type 1 diabetes associated variants, showing on the left hand panel the log-odds ratios for the minor allele for those diagnosed at  $<5$  years old (red circle; log-odds ratio age-at-diagnosis  $\pm$  95%CI), 5-13 years old (green circle; log-odds ratio age-at-diagnosis 5-13  $\pm$  95%CI) and  $\geq 13$  years old (blue circle log-odds ratio age-at-diagnosis  $\geq 13 \pm$  95%CI) from a multinomial logistic regression; the dashed red line shows a log-odds ratio of 0. The

right panel shows the association statistics from a likelihood ratio test comparing a multinomial logistic regression constraining the log-odds ratios from the  $<5$  and  $\geq 13$  groups to be equal compared to an unconstrained model. Red dotted line shows threshold for nominally significant heterogeneity between groups ( $p < 0.05$ ), red solid line shows threshold for false discovery rate of  $< 0.1$ , red dashed line shows threshold for Bonferroni-corrected significant heterogeneity.

**Supplementary Figure 7:** Minor allele frequency of the index variant near the *CTSH* gene for controls and individuals diagnosed at various ages.

**Supplementary Figure 8:** Minor allele frequency of the index variant near the *GLIS3* gene for controls and individuals diagnosed at various ages.

**Supplementary Figure 9:** Minor allele frequency of the index variant near the *IKZF3* gene for controls and individuals diagnosed at various ages.

**Supplementary Figure 10:** Minor allele frequency of the index variant near the *IL2RA* gene (third index variant) for controls and individuals diagnosed at various ages.

**Supplementary Figure 11:** Minor allele frequency of the index variant near the *THEMIS* gene for controls and individuals diagnosed at various ages.

**Supplementary Figure 12:** Minor allele frequency of the index variant near the *IL10* gene for controls and individuals diagnosed at various ages.



**Supplementary Figure 13:** Top panel: association absolute z scores from whole blood eQTL study examining variant effects on GLIS3 mRNA levels, coloured by LD  $r^2$  to the most strongly associated variant with GLIS3 mRNA expression. Bottom panel: association absolute z scores from logistic regression examining variant associations with type 1 diabetes risk diagnosed at <7 years, using individuals from the UK and Northern Ireland only and adjusting for the five largest principal components as derived from genotype data, coloured by LD  $r^2$  to the most strongly associated variant with GLIS3 mRNA. Shows most associated disease risk variants are not in high LD with the most associated GLIS3 whole blood eQTLs.

**Supplementary Figure 14:** Top panel: association absolute z-scores from whole blood eQTL study examining variant effects on IL10 mRNA levels, coloured by LD  $r^2$  to the most strongly associated variant with IL10 mRNA expression. Bottom panel: association absolute z scores from logistic regression examining variant associations with type 1 diabetes risk diagnosed at <7 years, using individuals from the UK and Northern Ireland only and adjusting for the five largest principal components as derived from genotype data, coloured by LD  $r^2$  to the most strongly associated variant with IL10 mRNA. Shows most associated disease risk variants are not in high LD with the most associated IL10 whole blood eQTLs.

**Supplementary Figure 15:** Top panel: association absolute z-scores from whole blood eQTL study examining variant effects on IL24 mRNA levels, coloured by LD  $r^2$  to the most strongly associated variant with IL24 mRNA expression. Bottom panel: association absolute z scores from logistic regression examining variant associations

with type 1 diabetes risk diagnosed at <7 years, using individuals from the UK and Northern Ireland only and adjusting for the five largest principal components as derived from genotype data, coloured by LD  $r^2$  to the most strongly associated variant with IL24 mRNA. Shows most associated disease risk variants are not in high LD with the most associated IL24 whole blood eQTLs.

**Supplementary Figure 16:** Top panel: association absolute z-scores from whole blood eQTL study examining variant effects on FAIM3 mRNA levels, coloured by LD  $r^2$  to the most strongly associated variant with FAIM3 mRNA expression. Bottom panel: association absolute z scores from logistic regression examining variant associations with type 1 diabetes risk diagnosed at <7 years, using individuals from the UK and Northern Ireland only and adjusting for the five largest principal components as derived from genotype data, coloured by LD  $r^2$  to the most strongly associated variant with FAIM3 mRNA. Shows most associated disease risk variants are not in high LD with the most associated FAIM3 whole blood eQTLs.

**Supplementary Table 1:** Classical HLA alleles/haplotypes examined in analysis.

**Supplementary Table 2:** Non-HLA variants examined in analysis.

**Supplementary Table 3:** Non-HLA region variants with evidence of heterogeneity in effect size between the <7 and  $\geq 13$  groups: Promoter Capture Hi-C (PCHi-C) candidate genes.

**Supplementary Table 4:** Details of non-HLA variants with evidence of heterogeneity in effect size between the  $<7$  and  $\geq 13$  groups.

**Supplementary Table 5:** Most likely variants causally associated with type 1 diabetes at the *CTSH* locus from GUESSFM fine mapping analysis.

**Supplementary Table 6:** Most likely variants causally associated with type 1 diabetes at the *GLIS3* locus from GUESSFM fine mapping analysis.

**Supplementary Table 7:** Most likely variants causally associated with type 1 diabetes at the *IKZF3* locus from GUESSFM fine mapping analysis.

**Supplementary Table 8:** Most likely variants causally associated with type 1 diabetes at the *IL2RA* locus from GUESSFM fine mapping analysis.

**Supplementary Table 9:** Most likely variants causally associated with type 1 diabetes at the *THEMIS* locus from GUESSFM fine mapping analysis.

**Supplementary Table 10:** Most likely variants causally associated with type 1 diabetes at the *IL10* locus from GUESSFM fine mapping analysis.

**Supplementary Table 11:** Chip heritability estimates by age-at-diagnosis group under various disease prevalence assumptions.