Genetic discovery and risk prediction for type 1 diabetes in individuals without high-risk HLA-DR3/DR4 haplotypes

Carolyn McGrail[1,#], Joshua Chiou[1,*], Ruth Elgamal[1], Amber M Luckett[2], Richard A Oram[2,3], Paola Benaglio[4], Kyle J Gaulton[4,#]

1. Biomedical Sciences Graduate Program, UC San Diego, La Jolla, CA
2. University of Exeter College of Medicine and Health, Exeter, UK.
3. Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK
4. Department of Pediatrics, UC San Diego, La Jolla, CA
Corresponding authors

* Present affiliation: Internal Medicine Research Unit, Pfizer Research and Development, Pfizer, Cambridge, MA, USA

Contact: Carolyn McGrail, cmcgrail@health.ucsd.edu; Kyle J Gaulton, kgaulton@ucsd.edu

ABSTRACT

Over 10% of type 1 diabetes (T1D) cases do not have high-risk HLA-DR3 or DR4 haplotypes with distinct clinical features such as later onset and reduced insulin dependence. To identify genetic drivers of T1D in the absence of DR3/DR4, we performed association and fine-mapping analyses in 12,316 non-DR3/DR4 samples. Risk variants at the MHC and other loci genome-wide had heterogeneity in effects on T1D dependent on DR3/DR4, and non-DR3/DR4 T1D had evidence for a greater polygenic burden. T1D-assocated variants in non-DR3/DR4 were more enriched for loci, regulatory elements, and pathways for antigen presentation, innate immunity, and beta cells, and depleted in T cells, compared to DR3/DR4. Non-DR3/DR4 T1D cases were poorly classified based on an existing genetic risk score GRS2, and we created a new GRS which highly discriminated non-DR3/DR4 T1D from both non-diabetes and T2D. In total we identified heterogeneity in T1D genetic risk and disease mechanisms dependent on high-risk HLA haplotype and which enabled accurate classification of T1D across HLA background.

Type 1 Diabetes (T1D) is an autoimmune disease characterized by the destruction of insulin producing beta cells[1,2] and which has complex etiology[2]. T1D is highly polygenic with over 90 known risk loci[3–8]12/2/23 9:34:00 PM and the largest genetic risk factors map to the Major Histocompatibility Complex (MHC) locus[7–9] which encodes cell surface receptors that present antigenic peptides to $CD4^+$ and $CD8^+$ T cells[10]. Haplotypes of the class II HLA-DR and -DQ genes *DRB1*0301-DQA1*0501-DQB1*0201* (DR3) and *DRB1*04:01/02/04/05/08-DQA1*03:01-DQB1*03:02/04* (DR4)[7,11] confer substantial risk of T1D[11] and are detected in 90% of European T1D cases[12,13]. These high-risk DR3/DR4 haplotypes are thought to increase T1D risk by altering peptide binding to class II MHC and presentation of autoantigens to T cells[14].

The remaining 10% of individuals of European ancestry that develop T1D without high-risk non-DR3/DR4 haplotypes have lower rates of T1D autoantibody positivity[15] and later onset of disease with lower insulin dependence, which can be misdiagnosed as type 2 diabetes (T2D) or latent autoimmune diabetes in adults (LADA)[16–22] leading to mismanagement and complications[16]. In addition, the preventative therapy Teplizumab is most effective in at-risk individuals who are DR4-positive and DR3-negative, and it is less established if it effectively delays onset in non-DR3/DR4 individuals[23]. Furthermore, existing genetic risk scores (GRS) for T1D heavily weigh DR3/DR4 haplotypes due to their large effect, and thus T1D in non-DR3/DR4 individuals may be poorly predicted by these scores. There is increasing evidence that adult onset T1D, which has a lower rate of DR3/DR4 haplotypes[24], is more prevalent than previously estimated[25,26], further underscoring the need to understand T1D in the absence of DR3 and DR4.

To address these gaps, we performed T1D association analyses in individuals without high-risk DR3/DR4 haplotypes (**Figure 1**). We obtained genotype data for 29,723 T1D and control individuals from five European ancestry cohorts (**Supplementary Table 1**), and imputed genotypes into 308M variants in TOPMed v2 and 56,310 variants in Michigan HLA reference panels[27,28]. We partitioned individuals by HLA-DR and -DQ haplotype resulting in 8,808 T1D and 8,199 controls with DR3/DR4 and 1,292 T1D and 11,424 controls without DR3/DR4 (**Supplementary Table 1**). We then tested for T1D

association in the non-DR3/DR4 group, as well as the DR3/DR4 group for comparison. Six non-MHC loci *PTPN22, INS, IFIH1, PTPN2, IKZF4,* and *OSTN,* as well as the MHC locus, reached genome-wide significance ($p<5x10^{-8}$) for T1D in non-DR3/DR4 (**Figure 1b, Supplementary Table 2**). Through conditional analysis (**see Methods**), we identified 12 additional signals at locus-wide significance ($p<1x10^{-5}$), all of which were at the MHC locus (**Figure 1d, Supplementary Table 3**). For all signals we then generated 99% 'credible sets' of variants likely causal for T1D (**Figure 1c,e, Supplementary Table 4**).

We compared the effects of risk variants at these loci on T1D between non-DR3/DR4 and DR3/DR4 using a Breslow-Day (BD) test (**see Methods**). Of the six significant non-HLA loci, five are known while one locus near *OSTN* is previously unreported **(Fig. 2a)**. At the *OSTN* locus, the lead variant had moderate effect in non-DR3/DR4 and limited effect in DR3/DR4 (b=0.4871, b= 0.0281, BD test $p=1.29x10^{-4}$) **(Fig. 2b)**. We also observed a stronger effect in non-DR3/DR4 for lead variants at the *IFIHI* (b=-0.35, b=-0.091, $p=3.88x10^{-5}$) and *PTPN2* (b=-0.37, b=-0.21, p=0.018) loci **(Fig. 2c),** both of which are implicated in beta cell function[29,30], and this heterogeneity remained after conditioning on other known signals at these loci (**Supplementary Table 5**). We further examined lead variants of primary signals at 88 known T1D loci[3], and observed heterogeneity ($p<0.05$) at 9 additional loci. At 5 loci there was evidence for larger effects in non-DR3/DR4 including *PRR15L*, *RAD51B, PRF1, PRKD2,* and 6q27, several of which have been implicated in both immune cells and beta cells. By comparison, 4 loci had evidence for smaller effect in non-DR3/DR4 14q32*, IL2RA*, *IL2*, and *CD69* **(Supplementary Fig. 1, Supplementary Table 6),** which are loci that largely affect T cell function.

We next determined whether individuals with T1D in the absence of HLA-DR3/DR4 may carry greater polygenic risk. We first determined whether known risk variants for T1D broadly had stronger effects in non-DR3/DR4 compared to DR3/DR4. An increased proportion (60.2%) of lead variants at 88 known T1D loci had stronger effect in non-DR3/DR4, although this was not significant (binomial p=0.069). We next determined whether non-HLA heritability differed between groups, and there was a small increase in estimated heritability for T1D in non-DR3/DR4 compared to DR3/DR4 (non-DR3/DR4

h^2=0.2846, se=0.0795; DR3DR4 h^2=0.2693, se=0.0428). Finally, there was a significant increase in the ability to distinguish T1D from controls in non-DR3/DR4 compared to DR3/DR4 (AUC=0.752, AUC=0.733, p=0.0238) using a risk score derived from 87 non-MHC lead variants in a recent T1D GWAS[3]. The ability to distinguish T1D from control in non-DR3/DR4 was further improved when using variant effects derived from the non-DR3/DR4 T1D GWAS (AUC=0.768, p=1.66×10^{-5}). Together these results provide evidence for a larger polygenic contribution to T1D risk in the absence of high-risk DR3/DR4 haplotypes.

We next investigated whether variants within the MHC locus contribute to heterogeneity between non-DR3/DR4 and DR3/DR4 (**see Methods**). Of the 12 MHC signals identified in non-DR3/DR4 T1D, the lead variants of 8 signals exhibited heterogeneity between groups after conditioning on preceding signals from stepwise regression (**Fig. 2a, Supplementary Table 7**). Most signals were in linkage disequilibrium (LD) with known T1D-associated alleles, except for two apparent novel signals linked to HLA-DRB1*09:01-DQA1*03:01-DQB1*03:03 and B*44:02, respectively (**Fig. 2a, Supplementary Table 8**). Among a larger set of 40 known MHC T1D risk alleles, 22 two-digit class I and class II alleles had evidence for heterogeneity (p<0.05) (**see Methods, Fig. 2d,e, Supplementary Table 9**). In class I, HLA-B*39:06 had increased risk in non-DR3/DR4 (p=8.27×10^{-3}), HLA-A*01:01 was more protective in non-DR3/DR4 (p=2.7×10^{-4}), and HLA-B*18:01 had opposed effects in non-DR3/DR4 and DR3/DR4 (p=5.41×10^{-6}) **(Fig. 2d,e)**. In class II, DRB1*15:01-DQB1*06:02 was more protective in DR3/DR4 (DRB1*15:01 p=8.187×10^{-8}, DQB1*6:02 p=8.62×10^{-7}) **(Fig. 2d,e)**, DRB1*13:01-DQB1*06:03 was more protective in non-DR3/DR4 (p=6.3×10^{-3}) **(Fig. 2d,e)**, and DPB1*03:01 had increased risk in DR3/DR4 (p=0.035) **(Fig. 2d,e)**[31]. Together this reveals MHC alleles with heterogeneity in effect on T1D in a non-DR3/DR4 background.

As a higher proportion of individuals without DR3/DR4 haplotypes have later onset of disease, we next examined whether heterogeneity in T1D risk between non-DR3/DR4 and DR3/DR4 was consistent across age of onset. We compared the effects of non-DR3/DR4 signals in each individual cohort included in our meta-analysis, including the

WTCCC and T1DGC cohorts where all cases had age of onset <17 (**Supplementary Table 1**). Among loci significant in non-DR3/DR4 that also exhibited heterogeneity with DR3/DR4, we observed no difference in T1D effects in WTCCC and T1DGC compared to other cohorts (Cochran Q p>0.05). (**Supplementary Table 2**, **Supplementary Fig. 2**). In addition, signals with heterogeneity between non-DR3/DR4 and DR3/DR4 were largely distinct from those with age-dependent association from a separate study[17], and HLA-B*39:06 had larger effect in non-DR3/DR4 and was associated with younger age of onset (**Supplementary Fig. 1**). This supports that heterogeneity in non-DR3/DR4 appears distinct from T1D age of onset-dependent effects.

Given heterogeneity in effects at individual T1D loci in non-DR3/DR4 and DR3/DR4, we next determined if there were broader differences in T1D association between groups. We first tested T1D associated variants in each group for enrichment of candidate *cis*-regulatory elements (cCREs) in immune cell types using stratified LD score regression[32–34]. In the non-DR3/DR4 group, there was enrichment (FDR<0.10) of T1D associated variants in memory B-cells, bulk B cells, mature NK cells, and unstimulated T regulatory cells **(Fig. 2f, Supplementary Table 10)**. Dendritic cells and monocytes were also enriched in non-DR3/DR4 but did not pass multiple test correction. By comparison, the DR3/DR4 group was most enriched for cCREs in T cell populations including T follicular helper, naive T regulatory, central memory $CD8^+$ T, and effector $CD4^+$ T cells (**Supplementary Table 10)**. We next performed gene set enrichment in each group using MAGMA (**Supplementary Table 11)**. The strongest enrichments (p<0.05) in the non-DR3/DR4 included regulation of lipid storage, macrophage stimulating factor, stress-induced apoptotic signaling, and complement activation **(Fig. 2g)**. By comparison, DR3/DR4 was most enriched for T cell-related pathways including IL-2 signaling, and T cell differentiation, activation, and proliferation **(Fig. 2h)**. Collectively, this suggests a greater APC, innate immune and beta cell contribution, and lower T cell contribution, to T1D in non-DR3/DR4 compared to DR3/DR4.

Finally, we evaluated the ability of genetic risk scores (GRS) to predict T1D in non-DR3/DR4 individuals. Over 35% of non-DR3DR4 T1D cases were below the 5th percentile

for T1D in the published GRS2[35], and only 7.5% were above the 50th percentile used to classify T1D with high specificity. In line with this, GRS2 had less ability to distinguish non-DR3/DR4 T1D cases from all controls (AUC=0.764, **Fig. 3a,e**) as compared to all T1D cases (AUC=0.927)[35]. We thus generated a non-DR3/DR4-specific T1D GRS using the 18 risk signals identified in our non-DR3/DR4 GWAS (**see Methods, Supplementary Table 12**). This GRS improved discrimination of non-DR3/DR4 T1D from all controls compared to GRS2 (AUC=0.835; $p=6.30\times10^{-15}$) (**Fig. 3a,e,f**). We next generated a larger 100-variant non-DR3/DR4 T1D GRS by including all 82 non-MHC known T1D risk loci (**Methods, Supplementary Table 12**), which further discriminated T1D from all controls compared to GRS2 (AUC=0.867; $p=7.48\times10^{-32}$) (**Fig. 3b,e,g**). When both T1D cases and controls were subset to non-DR3/DR4 individuals, the 100-variant nonDR3/DR4 GRS again had improved discrimination of T1D over both the 18-variant GRS (AUC=0.894, AUC=0.871; $p=5.70\times10^{-26}$) (**Fig. 3c,f,g**) and GRS2 (AUC=0.878; p=0.0264) (**Fig. 3d,e,g**). We confirmed the ability of the 100-variant non-DR3/DR4 GRS to predict non-DR3/DR4 T1D using an independent test set of 100 T1D and 300 control non-DR3/DR4 samples excluded from the initial association analyses (AUC=0.887) (**Supplementary Fig. 3a,b**).

Since non-DR3/DR4 T1D individuals tend to have later onset and a lower dependence on insulin therapy leading to potential misdiagnosis as T2D[16,17,24,36], we next evaluated the ability of a GRS to differentiate T1D from T2D using the WTCCC1 cohort (**Fig. 4**). Using GRS2 there was limited ability to distinguish non-DR3/DR4 T1D from all T2D (AUC=0.759) (**Fig. 4a,e**). In contrast, our non-DR3/DR4 GRS strongly discriminated non-DR3/DR4 T1D from T2D and was a significant improvement over GRS2 (18-variant AUC=0.890, $p=1.51\times10^{-32}$; 100-variant AUC=0.907, $p=4.94\times10^{-44}$) **(Fig. 4a,b,e,f,g)**. When both T1D and T2D cases were subset to non-DR3/DR4 individuals, the 100-variant non-DR3/DR4 GRS improved on both the 18-variant GRS (AUC=0.920, AUC=0.905, p=0.0775) (**Fig. 4c,f,g**) and GRS2 (AUC=0.876, $p=1.89\times10^{-6}$) (**Fig. 4d,e,g**). The predictions were consistent when utilizing an independent set of 100 T1D samples (AUC=0.906) (**Supplementary Fig. 3c,d**). The ability to distinguish non-DR3/DR4 T1D from T2D appears largely driven by the improved estimation of effects at the MHC locus

(T1D vs T2D HLA AUC=0.885), as the extensive HLA combinations in GRS2 still do not enable differentiating non-DR3DR4 T1D from T2D (T1D vs T2D GRS2 HLA AUC=0.698).

We finally determined the diagnostic value of a GRS for T1D in a non-DR3/DR4 background. Non-DR3/DR4 individuals with T1D on the published GRS2 scale do not meet the minimum requirement for diagnostic feasibility at a Youden index of 0.50 (max Youden index=0.427). When T1D GRS2 is subset to only non-DR3/DR4 and re-scaled (**Table 1**), a score of 12.40 at the 50th percentile of T1D has a sensitivity of 50% and specificity of 95.4% and meets diagnostic criteria (max Youden index=0.609) at the 25th T1D percentile. When comparing T1D to T2D using GRS2, there was limited ability to distinguish T1D from T2D (max Youden index=0.414). By comparison, the 18 variant non-DR3/DR4 GRS has a score of 12.50 at the 50th percentile for T1D with a sensitivity of 50.0% and specificity of 95.2% (**Table 1**) and a peak Youden index of 0.607 at a score threshold of 11.05. Similarly, in the 100-variant non-DR3/DR4 GRS, the 50th percentile of T1D has a sensitivity of 50.0% and specificity of 96.9% and a peak Youden index of 0.644 at a score threshold of 23.21 (**Table 1**). There was also diagnostic value in the 100-variant non-DR3/DR4 GRS in differentiating non-DR3/DR4 T1D from T2D (max Youden index=0.687) and all controls (max Youden index=0.585). Overall, our results provide a GRS with clinical value in distinguishing non-DR3/DR4 T1D from non-diabetes and T2D.

Discussion

We observed heterogeneity in T1D genetic risk dependent on high-risk DR3/DR4 haplotypes at class I and class II MHC alleles as well as other loci genome wide. Multiple genes at loci with larger effects on T1D in non-DR3/DR4 such as *PTPN2, IFIH1, PRR15L,* and *RAD51B* are implicated in inflammatory signaling[37,38] and survival[29,39,40] in immune cells and beta cells. We also identified a novel T1D locus in non-DR3/DR4 near *OSTN,* which has been linked to diabetic cardiomyopathy in mouse models[41]. Conversely, many loci with reduced effect in non-DR3/DR4 such as *IL2RA*, *IL2* and *CD69* impact T cell function and activation[42,43]. More broadly, there were distinct patterns of enrichment in non-DR3/DR4 T1D where genomic annotations and molecular pathways related to B cells, NK cells, and beta cells were more enriched, and T cells less enriched, compared

to DR3/DR4. These results collectively suggest that mechanisms of T1D in a non-DR3/DR4 background may depend more on inflammation and beta cell dysfunction than T cell activation[16–20]. Given modest evidence for a higher polygenic burden in T1D in a non-DR3/DR4 background, individuals without this high-risk factor may require additional polygenic risk to lead them to develop T1D.

Based on previous GRS percentile scales, many non-DR3/DR4 individuals with T1D would not have been predicted to develop T1D. The GRS reported here enables accurate discrimination of T1D from non-diabetes and T2D in non-DR3/DR4 individuals. To improve accessibility, the variants in the non-DR3/DR4 GRS are derived from publicly available imputation panels enabling the GRS to be more easily calculated from genotyping array data. While the subset of individuals that develop T1D without DR3/DR4 exhibit later onset and lower insulin dependence, they have similar risk profiles to the majority of T1D cases and not to T2D. The accurate prediction of T1D in the non-DR3/DR4 population in this study will therefore help avoid misdiagnosis of T2D and prevention of ketoacidosis and future complications. While non-DR3/DR4 T1D background has generally later onset, it can occur at any age and our results suggest that the effects of variants on T1D risk in a non-DR3/DR4 background are generally consistent across age of onset. As this study was conducted in individuals of European ancestry, similar studies of T1D in other ancestries may help improve risk prediction of T1D in these populations. Furthermore, the GRS reported here may have value in discriminating T1D in populations where high-risk haplotypes such as DR3/DR4 are uncommon.

The ability to distinguish at-risk individuals is also critical for determining eligibility for clinical trials and therapies. The preventative therapy Teplizumab modulates T cell activity and is most efficacious in individuals with HLA-DR4[23], and several clinical trials aiming to preserve beta cell function at onset preferentially recruit DR3/DR4 individuals[44,45]. Given evidence for differences in disease mechanisms, in particular a less prominent contribution from T cells, alternate therapies may be needed to prevent T1D in non-DR3/DR4 individuals. In addition, as autoantibodies are seen at lower rates in non-DR3/DR4 T1D, additional biomarkers are needed for this group[15].

Overall our findings are in line with a growing body of literature [46–49] which support that T1D is a heterogeneous disease consisting of sub-types with distinct pathophysiological features. More broadly, stratifying by high-risk genetic background may be an effective strategy to discover genetic and mechanistic heterogeneity in other complex diseases.

ONLINE METHODS

Research subjects and genotype imputation

We compiled genotype data from 10,100 T1D and 19,623 control individuals of European ancestry from publicly available cohorts (**Supplementary Table 1**). T1D case cohorts were matched to control cohorts based on country of origin and genotype array where possible as previously described[3]. We applied the HRC imputation preparation program (version 4.2.9, https://www.well.ox.ac.uk/~wrayner/tools/) and used PLINK[50] (version 1.90) to perform quality control prior to imputation to remove variants with MAF <1%, missing genotypes >5%, in violation of Hardy-Weinberg equilibrium (HWE $P < 1\times10^{-5}$ in control cohorts and HWE $P<1\times10^{-10}$ in case cohorts), difference in allele frequency > 0.2 compared to HRC r1.1 reference panel[51], and variants with allele ambiguity.[3,50]. We then imputed genotypes for all samples into the TOPMed v2 and Michigan Multi-ethnic HLA reference panels using the respective imputation servers[52,53]. We additionally used SNP2HLA to impute genotypes into the specialized European HLA TDGC reference panel[54]. In the genome-wide imputation, we removed variants with an imputation accuracy score of $r^2<0.3$. In the HLA imputed imputation, we removed variants with an imputation accuracy of $r^2<0.5$ and a standard deviation in control allele frequency >0.055 across cohorts. Variants that passed QC filters in all cohorts were tested for association.

Association testing and meta-analysis

We defined distinct DR3/DR4 and non-DR3/DR4 groups based on HLA-DR and HLA-DQ haplotype status using four-digit HLA alleles imputed from the T1DGC reference panel with SNP2HLA. DR3 was identified by the presence of HLA-DRB1*03:01-DQB1*02:01, while DR4 was identified by the presence of HLA-DRB1*04:01/02/04/05/08-

DQB1*03:02/04/02:02[47]. We excluded 23 individuals who were identified as non-DR3/DR4 via SNP2HLA alleles, but had DR3 or DR4 tag SNPs[55,56]. From the non-DR3/DR4 risk group, 100 cases and 300 control samples we removed prior to performing association analyses. In both non-DR3/DR4 and DR3/DR4 groups, we tested variants for T1D case and control association using firth bias corrected logistic regression in EPACTS[57]. We tested variants with MAF>1% for association using sex and the first 4 genotype PCs as covariates. Summary statistics were combined across cohorts in a fixed effects inverse variance weighted meta-analysis using METAL[58]. The genomic inflation (lambda GC) was 1.069 for non-DR3/DR4 and 1.098 for DR3/DR4. We used LDSC[32] to test for heritability in the summary statistics outside of the MHC for each group using a population prevalence of 1% for each group, and a sample prevalence of 51% for DR3/DR4 and 10% for non-DR3/DR4.

Fine-mapping of independent signals

We identified 1 Mb regions around the lead variants of all genome-wide significant loci. We performed conditional analysis at each locus using stepwise analysis iteratively including the most significant variant in the regression model and re-performing the meta-analysis until no significant variants remained. In conditional analysis we used a locus wide significance threshold of $p < 1x10^{-5}$. We then performed Bayesian fine mapping to create credible sets of likely causal variants for each signal[59]. From the summary statistics, effect size and standard error were used to calculate the approximate Bayes factor (BF) for each variant in $r^2>0.1$ with the lead variant. We calculated the probability of association (PPA) for each variant by dividing the BF by the total sum of BFs. We then created 99% credible sets by including variants in descending order of PPA until the cumulative posterior probability was at least 99% (**Supplementary Table 12**).

ATAC-seq peak calling

ATAC-seq data for 175 samples from 20 different immune cell types (Bulk B, Mem B, Naïve B, Effector CD4pos T, Follicular T Helper, Memory Teffs, Memory Tregs, Naïve Teffs, Th1 precursors, Regulatory T, Th2 precursors, Th17 precursors, Naïve Tregs, Effector memory CD8pos T, Naïve CD8 T, Central memory CD8pos T, CD8pos T,

Gamma delta T, Monocytes, Mature NK) at resting and stimulated conditions were obtained from the NCBI GEO database at the accession GSE118189[34] and processed to generate peak coordinates. Reads were aligned using STAR[60] to hg19 and duplicate reads, reads mapping to blacklisted regions from ENCODE, and read with mapping quality Q<30 were filtered. Peak calling was performed using MACS2[61] on BAM files further filtered for read pairs with insert size no larger than 140 bp (macs2 callpeak --nomodel --nolambda --keep-dup all --call-summits -f BAMPE -g hs –q 0.01), combining individual samples for each cell type and treatment, for a total of 40 distinct peak sets. We then used bedtools multiinter to obtain one consensus set of peaks and featureCounts[62] to obtain the peak x read counts in each cell-type.

GWAS enrichment analysis

For each group, we performed partitioned heritability LD-score regression to estimate genome wide enrichment in stimulated and unstimulated immune cell accessible chromatin generated above[32–34]. We used the summary statistics for each group excluding the MHC locus and formatted it for input to LD score regression using the munge_sumstats.py script. We generated binary annotations from each accessible chromatin bed file and computed cell specific LD enrichment scores for each risk cohort using the version 2.2 1000G baseline model. We then corrected for multiple tests in each group across all cell types using false discovery rate (FDR) and considered FDR<0.10 significant. Additionally, we performed gene set enrichment in GO, KEGG and REACTOME pathways with the summary statistics for each group using MAGMA[63] with default parameters and reported the strongest enrichments using uncorrected p-values.

Testing differences in T1D risk between groups

We tested for differences in T1D effect between non-DR3/DR4 and DR3/DR4 using merged PLINK files of Michigan HLA and TOPMed imputed variants for all 29,723 samples. We tested for heterogeneity in marginal effects on T1D using Breslow-Day (BD) tests with the “-bd” flag in PLINK[50]. We performed BD tests for lead variants for six non-HLA loci identified in non-DR3/DR4 samples, and lead variants for 88 previously known

T1D risk loci[3]. At MHC and other loci with multiple signals, we tested for heterogeneity in effects on T1D conditional or other known variants. We generated regression models with PLINK using the "-glm interaction firth" flag including sex, the first 4 genotype PCs, DR3/DR4 status, and additional variants as covariates and evaluated the interaction with DR3/DR4 status. For the 12 MHC signals identified in non-DR3/DR4 samples, we conditioned on preceding lead variants from the stepwise regression. For the *IFIH1* and *PTPN2* loci, we conditioned on lead variants for all other known signals at the locus[50]. For the HLA locus, we conditioned on HLA-DRB1*03:01, HLA-DQB1*02:01, HLA-DRB1*04:01, HLA-DRB1*04:02, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*04:08, HLA-DQB1*03:02, and HLA-DQB1*03:04 to examine heterogeneity in 40 known two-digit HLA risk alleles independent of these 9 DR3 and DR4 alleles. To then obtain the conditional effect of each variant within non-DR3/DR4 or DR3/DR4, we performed logistic regression separately for each group using EPACTS including sex, 4 genotype PCs, and the same variants from the interaction tests above.

Generation and statistical analysis of the GRSs

We calculated GRS1 by using SNP2HLA imputed genotypes to define DR3 and DR4 tag SNPs and TOPMed imputed genotypes for the remaining 28 signals[55]. We calculated GRS2 by using TOPMED variants where possible (60 variants), Michigan HLA for rs116522341, rs1281934, and the Michigan HLA proxies DQB1*06:02, B*18:01, DPB1*03:01, rs1611547, and rs114170382, for rs17843689, rs371250843, rs559242105, rs144530872, and rs149663102 respectively. We excluded individuals with more than 2 HLA-DR/DQ calls, for GRS2, in line with the published methods[35]. For each of the signals identified in the non-DR3/DR4 group, we generated an 18-varaint GRS by summation of all signals weighted by the beta (B) for each effect allele (X) (**Equation 1**, **Supplementary Table 12**).

$$GRS = \sum_{i=0}^{N} B_i X_i$$

We additionally leveraged 82 non-HLA T1D risk loci from the largest T1D GWAS conducted to date[3] weighted by non-DR3/DR4 marginal effects to generate a 100-variant

combined GRS. We determined the ability of GRS2, the 18-variant GRS, and 100-variant GRS to distinguish non-DR3/DR4 T1D from either all control samples (DR3/DR4 and non-DR3/DR4) or non-DR3/DR4 control samples. We tested the ability of each GRS to discriminate T1D and non-diabetes using the area under the curve (AUC) of the receiver operator characteristic (ROC) statistics[64]. We calculated the difference between AUCs of each GRS using the deLong test and calculated the diagnostic criteria using the Youden Index. We then generated percentiles for how many non-DR3/DR4 T1D or control individuals fall at various GRS score thresholds. We calculated sensitivity at each GRS score as TP/ (TP + FN) and specificity as TN/(TN+FP)[65]. We further tested the ability to differentiate T1D from non-disease in the 100-variant non-DR3/DR4 T1D GRS using an independent test group of 100 cases and 300 control samples removed prior to association analyses. We also compared the ability of each GRS to differentiate T1D from T2D using 1,999 T2D individuals from the WTCCC study[66]. After imputing the T2D genotypes into TOPMed and Michigan HLA reference panels, we calculated the scores for each GRS as described above and derived ROC statistics comparing T1D to T2D. We additionally validated the performance of the 100-variant non-DR3/DR4 GRS to distinguish T1D from T2D using the 100 T1D case test group and all 1,999 T2D samples.

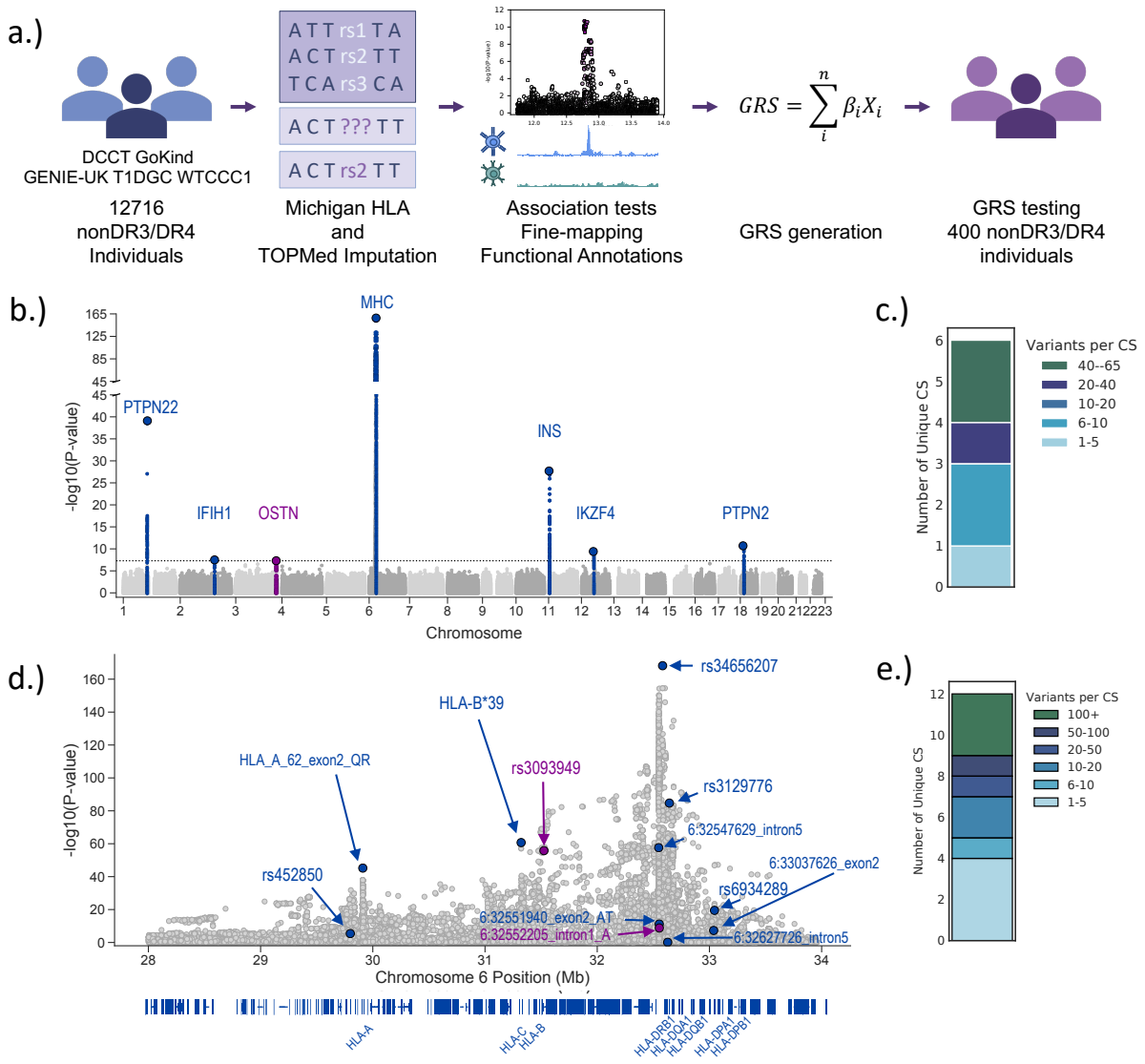
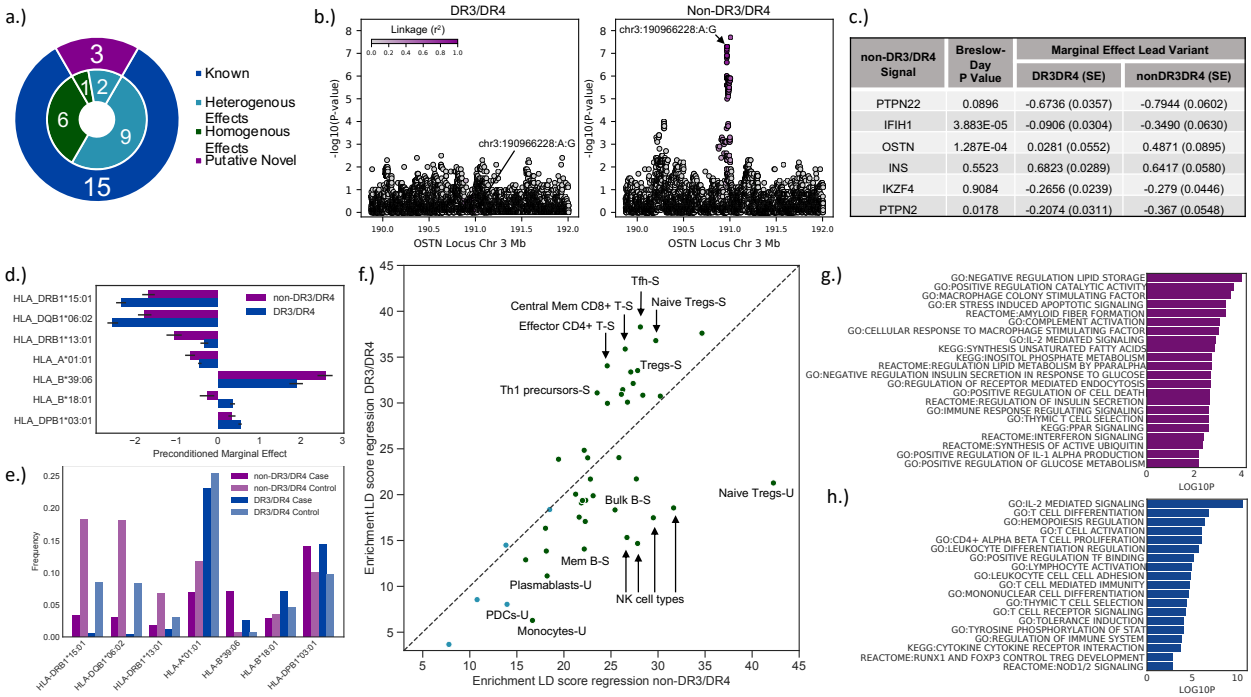


Fig. 1. Genetic Discovery in Non-DR3/DR4 T1D. **a.** Overview of genetic discovery and risk prediction for T1D in non-DR3/DR4 individuals. **b.** Genome-wide T1D association ($\log_{10}$ *P* values from meta-analysis of *n*=12,316 samples) in individuals without DR3/DR4 haplotypes. Known T1D loci are colored blue and novel loci are colored purple. All loci are labeled with the nearest gene. **c.** Number of 99% credible set (CS) variants in fine-mapped non-HLA T1D risk signals. **d.** T1D association at the MHC locus ($\log_{10}$ *P* values from marginal association in meta-analysis of *n* = 12,316 samples). Known signals are colored blue and novel signals are colored purple. The location of class I and II HLA genes are shown on the bottom. **e.** Number of 99% credible set variants in fine-mapped MHC signals.



non-DR3/DR4 Signal	Breslow-Day P Value	Marginal Effect Lead Variant	
		DR3DR4 (SE)	nonDR3DR4 (SE)
PTPN22	0.0896	-0.6736 (0.0357)	-0.7944 (0.0602)
IFIH1	3.883E-05	-0.0906 (0.0304)	-0.3490 (0.0630)
OSTN	1.287E-04	0.0281 (0.0552)	0.4871 (0.0895)
INS	0.5523	0.6823 (0.0289)	0.6417 (0.0580)
IKZF4	0.9084	-0.2656 (0.0239)	-0.279 (0.0446)
PTPN2	0.0178	-0.2074 (0.0311)	-0.367 (0.0548)

Fig. 2: Heterogeneity of genetic and biological mechanisms in non-DR3/DR4 and DR3/DR4 T1D a. Number of known and novel signals with heterogeneity in effect on T1D in non-DR3/DR4 and DR3/DR4 background. **b.** Locus plots of T1D association at the novel *OSTN* locus in DR3/DR4 (left) and non-DR3/DR4 (right). **c.** non-HLA loci with genome wide significant T1D association in non-DR3/DR4 analyses with effect sizes for lead variants in each group and difference in effect using interaction tests. **d.** T1D effect sizes of class I and II HLA alleles in each risk group after conditioning on 9 DR3 and DR4 risk alleles. **e.** T1D case and control frequencies of known HLA alleles in non-DR3/DR4 and DR3/DR4. **f.** Enrichment scores for T1D-associated variants in DR3/DR4 and non-DR3/DR4 in stimulated and unstimulated immune cell accessible chromatin using LD score regression. Dark green points are significant at FDR<0.10 in non-DR3/DR4 and DR3/DR4 while light blue points are FDR significant in only DR3/DR4. **g.** Gene pathway enrichment in non-DR3/DR4 T1D using MAGMA. **h.** Gene pathway enrichment in DR3/DR4 T1D using MAGMA.

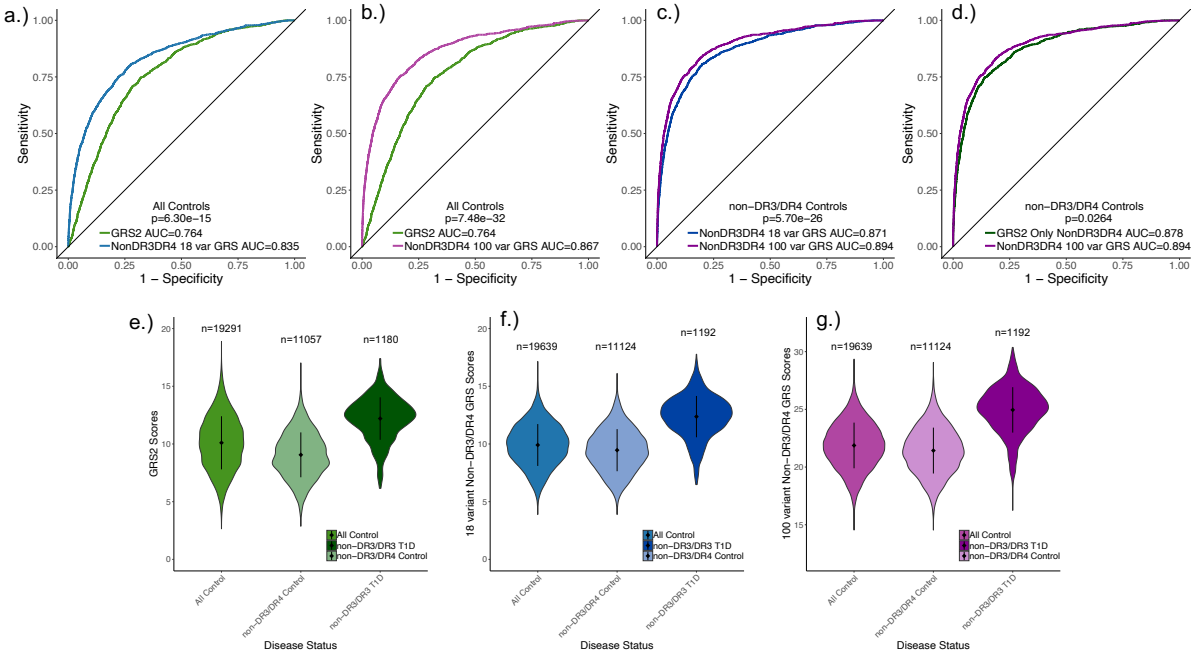


Fig. 3. Genetic Risk Prediction in non-DR3/DR4 T1D. Receiver operating characteristic (ROC) curves showing the ability of genetic risk scores (GRS) to differentiate T1D from non-diabetes in the non-DR3/DR4 population and corresponding violin plots. The AUCs are shown for each GRS and the p-values comparing predictive ability of GRS are calculated using the deLong test. **a.** GRS2 compared to the 18 variant non-DR3/DR4 GRS in non-DR3/DR4 T1D vs. all controls. **b.** GRS2 compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D vs. all controls. **c.** The 18-variant non-DR3/DR4 T1D GRS compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D and controls. **d.** GRS2 subset to non-DR3/DR4 T1D and controls compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D and controls. Violin plots for risk scores in all controls, non-DR3/DR4 controls, and T1D cases depicted for **e.** GRS2 **f.** 18-variant non-DR3/DR4 GRS and **g.** 100-variant non-DR3/DR4 GRS.

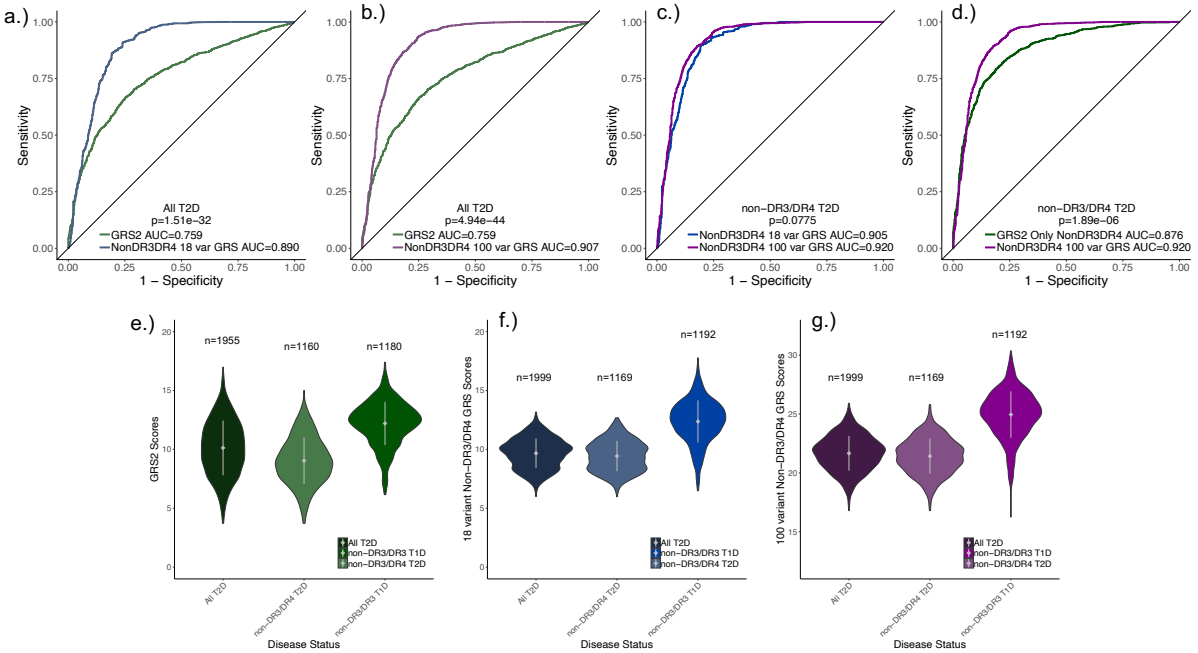


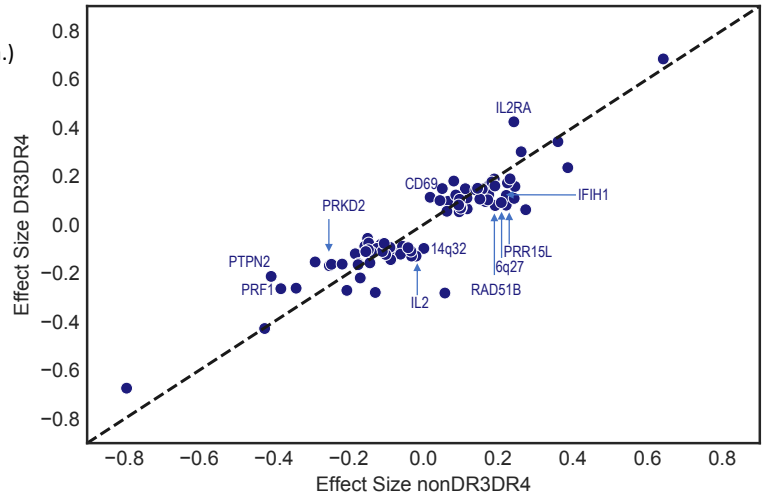
Fig. 4. Genetic Risk Prediction of non-DR3/DR4 T1D and T2D. ROC curves showing the ability of GRS to differentiate T1D from T2D in the non-DR3/DR4 population and corresponding violin plots. The AUCs are shown for each GRS and p-value for difference in prediction of T1D from T2D is calculated using the deLong test. **a.** GRS2 compared to the 18 variant non-DR3/DR4 GRS in non-DR3/DR4 T1D vs. all T2D. **b.** GRS2 compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D vs. all T2D. **c.** The 18-variant non-DR3/DR4 T1D GRS compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D and T2D. **d.** GRS2 subset to non-DR3/DR4 T1D and T2D compared to the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D and T2D. Violin plots for risk scores in all T2D, non-DR3/DR4 T2D, and T1D cases depicted for **e.** GRS2 **f.** 18-variant non-DR3/DR4 GRS and **g.** 100-variant non-DR3/DR4 GRS.

Table 1: Sensitivity, Specificity, Youden Index, and scales for GRS cutoff values in non-DR3/DR4 only GRS2, non-DR3/DR4 18 variant GRS and non-DR3/DR4 100 variant GRS

non-DR3/DR4 only GRS2	T1D Centile	Non-Disease Centile	Sensitivity (%)	Specificity (%)	Youden Index
8.86	5	47.6	95.00	47.63	0.4263
10.33	15	73.7	85.00	73.72	0.5872
11.20	25	85.3	75.00	85.35	0.6035
11.77	35	91.3	65.00	91.27	0.5627
12.40	50	95.4	50.00	95.41	0.4541
13.39	75	98.9	25.00	98.85	0.2385
14.85	95	99.9	5.000	99.95	0.0978
18-variant non-DR3/DR4 GRS	**T1D Centile**	**Non-Disease Centile**	**Sensitivity (%)**	**Specificity (%)**	**Youden Index**
9.066980	5	42.0	94.97	41.95	0.3691
10.585440	15	72.5	84.98	72.60	0.5758
11.358825	25	84.5	75.00	84.53	0.5953
11.850500	35	90.3	65.02	90.26	0.5527
12.497350	50	95.2	50.00	95.20	0.4519
13.553850	75	98.8	25.00	98.78	0.2377
15.080665	95	99.9	5.034	99.87	0.04908
100-variant non-DR3/DR4 GRS	**T1D Centile**	**Non-Disease Centile**	**Sensitivity (%)**	**Specificity (%)**	**Youden Index**
21.254775	5	46.9	94.97	46.90	0.4187
23.057930	15	78.7	84.98	78.68	0.6366
23.843050	25	88.4	75.00	88.41	0.6341
24.434680	35	93.6	65.02	93.56	0.5858
25.108450	50	96.9	50.00	96.90	0.4689
26.283800	75	99.3	25.00	99.33	0.2433
27.820655	95	99.9	5.034	99.96	0.0450

Supplement

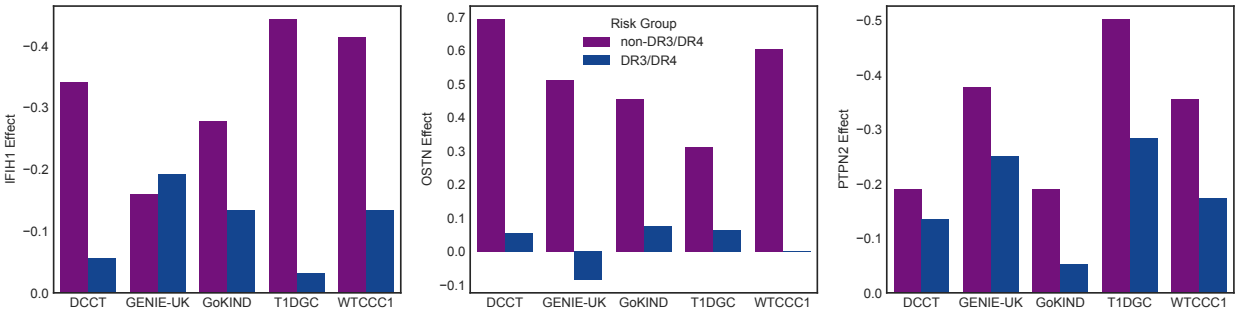
a.)



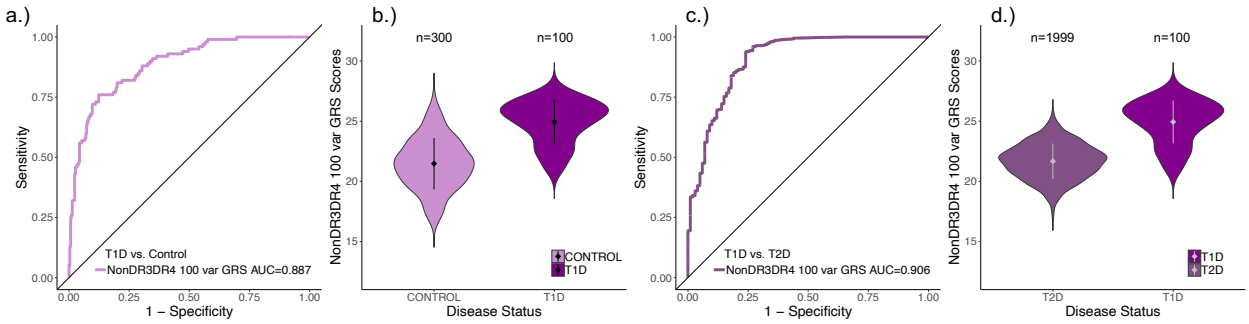
b.)

Locus	Marginal Effect		Breslow-Day P Value
	DR3DR4	nonDR3DR4	
PRR15L	0.08	0.2214	0.004279
PTPN2	-0.2133	-0.4076	0.01389
IFIH1	0.1204	0.2216	0.02353
RAD51B	0.0785	0.1919	0.03476
PRKD2	-0.1692	-0.2519	0.04268
PRF1	-0.264	-0.3816	0.0444
6q27	0.091	0.2081	0.04829
14q32	-0.0981	0.0011	0.01759
IL2RA	0.4233	0.2419	0.02244
IL2	-0.1297	-0.0198	0.04386
CD69	0.1126	0.0178	0.04801

Supplementary Fig. 1. a. The T1D effect size of lead variants in non-DR3/DR4 and DR3/DR4 at primary signals for 88 known T1D loci. Selected loci with variants that have the largest differences in T1D effect are annotated. **b.** Table containing the locus and T1D effect size for non-DR3/DR4 and DR3/DR4 and the p-value for heterogeneity in effect using a Breslow Day test for 11 lead variants with $p < 0.05$.



Supplementary Fig. 2. Effects for discovery non-MHC non-DR3/DR4 signals. The effects of lead variants at the *IFIH1, OSTN* and *PTPN2* loci on T1D risk in non-DR3/DR4 and DR3/DR4 stratified by individual cohort included in this study.



Supplementary Fig. 3. ROC curves showing the ability of GRSs to differentiate T1D from non-disease and T1D from T2D in an independent test population of non-DR3/DR4 T1D and corresponding violin plots. **a.** The AUC is shown for the 100-variant non-DR3/DR4 T1D GRS in non-DR3/DR4 T1D and non-disease and **b.** violin plot for scores in 100 T1D cases and 300 controls from the test group. **c.** The AUC is shown for the combined 100-variant non-DR3/DR4 T1D GRS in non-Dr3/DR4 T1D and all T2D. **d.** violin plot for scores in 100 T1D cases from the test group and 1999 T2D individuals.

1. Gepts, W. Pathologic Anatomy of the Pancreas in Juvenile Diabetes Mellitus. *Diabetes* **14**, 619–633 (1965).
2. Katsarou, A. *et al.* Type 1 diabetes mellitus. *Nat. Rev. Dis. Primer* **3**, 1–17 (2017).
3. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nat. 2021 5947863* **594**, 398–402 (2021).
4. Barrett, J. C. *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
5. Bradfield, J. P. *et al.* A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* **7**, e1002293 (2011).
6. Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
7. Nerup, J. *et al.* HL-A ANTIGENS AND DIABETES MELLITUS. *The Lancet* **304**, 864–866 (1974).
8. Thomsen, M. *et al.* MLC Typing in Juvenile Diabetes Mellitus and Idiopathic Addison's Disease. *Immunol. Rev.* **22**, 125–147 (1975).
9. Noble, J. A. *et al.* The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am. J. Hum. Genet.* **59**, 1134–1148 (1996).
10. Horton, R. *et al.* Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
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. Rønningen, K. S., Spurkland, A., Iwe, T., Vartdal, F. & Thorsby, E. Distribution of HLA-DRB1, -DQA1 and -DQB1 alleles and DQA1-DQB1 genotypes among Norwegian patients with insulin-dependent diabetes mellitus. *Tissue Antigens* **37**, 105–111 (1991).

13. Sanjeevi, C. B. *et al.* Polymorphic Amino Acid Variations in HLA-DQ Are Associated With Systematic Physical Property Changes and Occurrence of IDDM. *Diabetes* **44**, 125–131 (1995).
14. Nepom, G. T. A Unified Hypothesis for the Complex Genetics of HLA Associations With IDDM. **39**, (1990).
15. Singh, S. *et al.* Prevalence of Autoantibodies and HLA DR, DQ in Type 1 Diabetes Mellitus. *J. Clin. Diagn. Res. JCDR* **10**, EC09-EC13 (2016).
16. Leslie, R. D. *et al.* Adult-Onset Type 1 Diabetes: Current Understanding and Challenges. *Diabetes Care* **44**, 2449–2456 (2021).
17. Inshaw, J. R. J., Cutler, A. J., Crouch, D. J. M., Wicker, L. S. & Todd, J. A. 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 B-cells. *Diabetes Care* **43**, 169–177 (2020).
18. Barker, A. *et al.* Age-dependent decline of β-cell function in type 1 diabetes after diagnosis: a multi-centre longitudinal study. *Diabetes Obes. Metab.* **16**, 262–267 (2014).
19. Leete, P. *et al.* The Effect of Age on the Progression and Severity of Type 1 Diabetes: Potential Effects on Disease Mechanisms. *Curr. Diab. Rep.* **18**, 115 (2018).
20. Karjalainen, J., Salmela, P., Ilonen, J., Surcel, H. M. & Knip, M. A comparison of childhood and adult type I diabetes mellitus. *N. Engl. J. Med.* **320**, 881–886 (1989).
21. Jones, A. G., McDonald, T. J., Shields, B. M., Hagopian, W. & Hattersley, A. T. Latent Autoimmune Diabetes of Adults (LADA) Is Likely to Represent a Mixed Population of Autoimmune (Type 1) and Nonautoimmune (Type 2) Diabetes. *Diabetes Care* **44**, 1243–1251 (2021).
22. Jones, A. G., McDonald, T. J., Shields, B. M., Hagopian, W. & Hattersley, A. T. Latent Autoimmune Diabetes of Adults (LADA) Is Likely to Represent a Mixed Population of

Autoimmune (Type 1) and Nonautoimmune (Type 2) Diabetes. *Diabetes Care* **44**, 1243–1251 (2021).

23. Herold, K. C. *et al.* An Anti-CD3 Antibody, Teplizumab, in Relatives at Risk for Type 1 Diabetes. *N. Engl. J. Med.* **381**, 603–613 (2019).
24. Howson, J. M. M., Rosinger, S., Smyth, D. J., Boehm, B. O. & Todd, J. A. Genetic Analysis of Adult-Onset Autoimmune Diabetes. *Diabetes* **60**, 2645–2653 (2011).
25. Thomas, N. J. *et al.* Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank. *Lancet Diabetes Endocrinol.* **6**, 122–129 (2018).
26. Harding, J. L. *et al.* The Incidence of Adult-Onset Type 1 Diabetes: A Systematic Review From 32 Countries and Regions. *Diabetes Care* **45**, 994–1006 (2022).
27. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nat. 2021 5907845* **590**, 290–299 (2021).
28. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat. Genet.* **53**, 1504–1516 (2021).
29. Santin, I. *et al.* PTPN2, a Candidate Gene for Type 1 Diabetes, Modulates Pancreatic β-Cell Apoptosis via Regulation of the BH3-Only Protein Bim. *Diabetes* **60**, 3279–3288 (2011).
30. Colli, M. L., Moore, F., Gurzov, E. N., Ortis, F. & Eizirik, D. L. MDA5 and PTPN2, two candidate genes for type 1 diabetes, modify pancreatic beta-cell responses to the viral by-product double-stranded RNA. *Hum. Mol. Genet.* **19**, 135–146 (2010).
31. Varney, M. D. *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* **59**, 2055–2062 (2010).
32. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

33. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

34. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019).

35. Sharp, S. A. *et al.* Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. in *Diabetes Care* vol. 42 200–207 (American Diabetes Association Inc., 2019).

36. Mishra, R. *et al.* Genetic Discrimination Between LADA and Childhood-Onset Type 1 Diabetes Within the MHC. *Diabetes Care* **43**, 418–425 (2019).

37. Looney, B. M., Xia, C.-Q., Concannon, P., Ostrov, D. A. & Clare-Salzler, M. J. Effects of type 1 diabetes-associated IFIH1 polymorphisms on MDA5 function and expression. *Curr. Diab. Rep.* **15**, 96 (2015).

38. Trivedi, P. *et al.* Perforin facilitates beta cell killing and regulates autoreactive CD8+ T-cell responses to antigen in mouse models of type 1 diabetes. *Immunol. Cell Biol.* **94**, 334–341 (2016).

39. Aylward, A. *et al.* Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLoS Genet.* **17**, e1009531 (2021).

40. Benaglio, P. *et al.* Type 1 diabetes risk genes mediate pancreatic beta cell survival in response to proinflammatory cytokines. *Cell Genomics* **2**, 100214 (2022).

41. Zhang, X. *et al.* Osteocrin, a novel myokine, prevents diabetic cardiomyopathy via restoring proteasomal activity. *Cell Death Dis.* **12**, 1–12 (2021).

42. Garg, G. *et al.* Type 1 Diabetes-associated IL2RAvariationlowers IL-2 signaling and contributes to diminished CD4+CD25+ regulatory T-cell function. *J. Immunol. Baltim. Md 1950* **188**, 4644–4653 (2012).

43. Gorabi, A. M. *et al.* The pivotal role of CD69 in autoimmunity. *J. Autoimmun.* **111**, 102453 (2020).

44. Van Rampelbergh, J. *et al.* First-in-human, double-blind, randomized phase 1b study of peptide immunotherapy IMCY-0098 in new-onset type 1 diabetes. *BMC Med.* **21**, 190 (2023).

45. Nowak, C. *et al.* Intralymphatic GAD-Alum (Diamyd®) Improves Glycemic Control in Type 1 Diabetes With HLA DR3-DQ2. *J. Clin. Endocrinol. Metab.* **107**, 2644–2651 (2022).

46. Caillat-Zucman, S. *et al.* Age-dependent HLA genetic heterogeneity of type 1 insulin-dependent diabetes mellitus. *J. Clin. Invest.* **90**, 2242–2250 (1992).

47. Inshaw, J. R. J., Cutler, A. J., Crouch, D. J. M., Wicker, L. S. & Todd, J. A. 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 β-Cells. *Diabetes Care* **43**, 169–177 (2020).

48. Qu, H.-Q. *et al.* Genetic architecture of type 1 diabetes with low genetic risk score informed by 41 unreported loci. *Commun. Biol.* **4**, 1–10 (2021).

49. Leete, P. *et al.* Studies of insulin and proinsulin in pancreas and serum support the existence of aetiopathological endotypes of type 1 diabetes associated with age at diagnosis. *Diabetologia* **63**, 1258–1267 (2020).

50. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

51. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

52. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

53. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).

54. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.-M. & Concannon, P. J. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS ONE* **8**, 64683 (2013).

55. Oram, R. A. *et al.* A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. *Diabetes Care* **39**, 337–344 (2016).

56. Nguyen, C., Varney, M. D., Harrison, L. C. & Morahan, G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. *Diabetes* **62**, 2135–2140 (2013).

57. EPACTS - Genome Analysis Wiki. https://genome.sph.umich.edu/wiki/EPACTS.

58. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190 (2010).

59. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009).

60. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).

61. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

62. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

63. Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Comput. Biol.* **11**, e1004219 (2015).

64. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).

65. Florkowski, C. M. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *Clin. Biochem. Rev.* **29**, S83–S87 (2008).

66. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).