Effect of parental autoimmune diseases on type 1 diabetes in offspring can be partially explained by HLA and non-HLA polymorphisms: a nationwide registry and biobank-based study in 7.2M Finns

Feiyi Wang†[1,2], Aoxing Liu†[1,3,4,5], Zhiyu Yang[1], Pekka Vartiainen[1], Sakari Jukarainen[1], Satu Koskela[6], Richard A. Oram[7], Lowri Allen[8], Jarmo Ritari[6], Jukka Partanen[6], FinnGen, Markus Perola[9], Tiinamaija Tuomi*[1,10,11,12], Andrea Ganna*[1,3,4,5]

1. Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland
2. Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, UK
3. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
4. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA
5. Analytic and Translational Genetics Unit, ATGU, Massachusetts General Hospital, Boston, MA, USA
6. Finnish Red Cross Blood Service, Helsinki, Finland
7. University of Exeter, Exeter, UK
8. Cardiff University and University Hospital of Wales, Cardiff, UK
9. Finnish Institute for Health and Welfare (THL), Helsinki, Finland
10. Abdominal Center, Endocrinology, Helsinki University Hospital, Helsinki, Finland
11. Folkhälsan Research Center, Helsinki, Finland
12. Lund University Diabetes Center, Malmo, Sweden

† Joint first authorship
* Joint senior authorship

Corresponding author: Andrea Ganna (andrea.ganna@helsinki.fi) and Tiinamaija Tuomi (tiinamaija.tuomi@hus.fi)

Abstract

Type 1 diabetes (T1D) and other autoimmune diseases (AIDs) co-occur in families. We studied the aggregation of 50 parental AIDs with T1D in offspring and the contribution of a shared genetic background, which was partitioned into HLA and non-HLA variation. Leveraging nationwide registers of 7.2M Finns, including 58,284 family trios, we observed that 15 parental AIDs, such as coeliac disease and rheumatoid arthritis, were associated with an increased risk of T1D in offspring. The identified epidemiological associations were then tested by comprehensive genetic analyses performed on 470K Finns genotyped in the FinnGen study (12,563 trios). The within-family genetic transmission analysis further demonstrated that the aggregation of parental AIDs with offspring T1D could be partially explained by HLA and non-HLA polymorphisms in a disease-dependent manner. For example, the associations with offspring T1D for coeliac disease and psoriasis were mainly driven by HLA while autoimmune hypothyroidism and rheumatoid arthritis also had non-HLA contributors. We, therefore, proposed a novel parental polygenic score (PGS), integrating variations in both HLA and non-HLA genes, to understand the cumulative risk pattern of T1D in offspring. This raises an intriguing possibility of considering parental PGS, in conjunction with clinical diagnoses, to inform individuals about T1D risk in their offspring.

Type 1 diabetes (T1D) is an autoimmune disease (AID) characterized by a severe insulin deficiency resulting from the destruction of insulin-producing pancreatic beta cells. While the pathogenic mechanism remains largely unknown, T1D usually appears in genetically susceptible individuals[1,2] and triggered by environmental exposures[3,4] - mostly before 20 years of age, with incidence increasing from birth and peaking at age 10-14 during puberty[5-7]. Notably, in many populations such as Europeans, Asians, and Latins, genetic polymorphisms in human leukocyte antigen (HLA) genes were reported to account for up to half of the T1D heritability, for which the strongest effects were attributable to two HLA class II haplotypes, HLA-DR3-DQ2 and DR4-DQ8[8-14]. In addition to the well-documented large-effect HLA contribution, non-HLA susceptibility also plays a role[2]. Large-scale genome-wide association studies (GWAS) have identified dozens of independent non-HLA signals across the genome[15-18], that together with HLA polymorphisms can be used to construct polygenic scores (PGS) to evaluate genetic predisposition of T1D prior to the disease onset[19-21] as well as better differentiate T1D from other types of diabetes[22-24].

Interestingly, T1D tends to co-occur with some other AIDs, both in the same individuals and within families (**Supplementary notes 1**). Previous studies mainly explored the co-occurrence from two angles. A popular and straightforward one was to adopt a population-based design and epidemiological analysis for the risk of T1D among people whose relatives were diagnosed with some AIDs. We have provided a detailed review of these studies in **Table S1**. To obtain conclusive findings, a sufficient number of trios or families is needed[25-27]. Another set of studies attempted to leverage genetic information for a better understanding of T1D-AID aggregation among genetically related individuals[28-30]. In one such study, among the one-third of children with T1D who were found to have a relative affected by an AID, having HLA-DR3-DQ2 haplotype was associated with celiac disease (CD) in their extended family[29]. Whereas this type of studies mainly collected family history from questionnaires and focused only on HLA haplotype analysis due to data limitation, other studies systematically researched the disease associations of common variants in either HLA or non-HLA genes. Instead of individual-level data, many studies utilized publicly accessible summary statistics generated from population-based analyses. For example, a recent study captured intercorrelation of AIDs, such as T1D, rheumatoid arthritis (RA), and autoimmune hypothyroidism (HYPO), in two large HLA hotspots (locus 961 and 964)[31]. Further, GWAS revealed shared genetic susceptibility loci beyond the HLA regions for T1D and five other AIDs, including CD, RA, multiple sclerosis (MS), Crohn's diseases, and systemic lupus erythematosus[32].

To date, there has been a lack of comprehensive investigation that combines the two angles and encompasses a wider range of AID spectrum than CD, RA, HYPO and MS. Meanwhile, it remains an enigma why individuals with similar genetic backgrounds can develop such heterogeneous diseases ranging from potentially lethal insulin deficiency to mild hypothyroidism - even among members of the same family who also tend to share the same environment. One way to approach this question is to combine evidence from population-based multi-generational cohorts and large-scale biobanks with high-resolution diagnostic data, preferably of the same population to avoid any bias or confounding caused by differences in across-population socio-cultural contexts and clinical practices. In this study, we aim to comprehensively evaluate the genetic determinants of familial aggregation of T1D and other AIDs using the nationwide multi-generational health registers of the whole Finnish population in FinRegistry[33], as well as the genomic and family trio data available in Finnish biobanks through the FinnGen Study[34]. Considering the world's highest incidence of T1D in Finland[35], these detailed and structured data resources provide us a unique opportunity to quantify the associations between parental AIDs and offspring T1D, and more importantly, to answer three fundamental questions: 1) Which parental diagnoses of AIDs are associated with T1D in offspring? 2) To what extent is the shared genetic background between AIDs and T1D driven by genetic components, separately for HLA and non-HLA variants? And, 3) Can AID-related genetic information in parents complement disease family history be used to estimate the risk of having a child developing T1D?

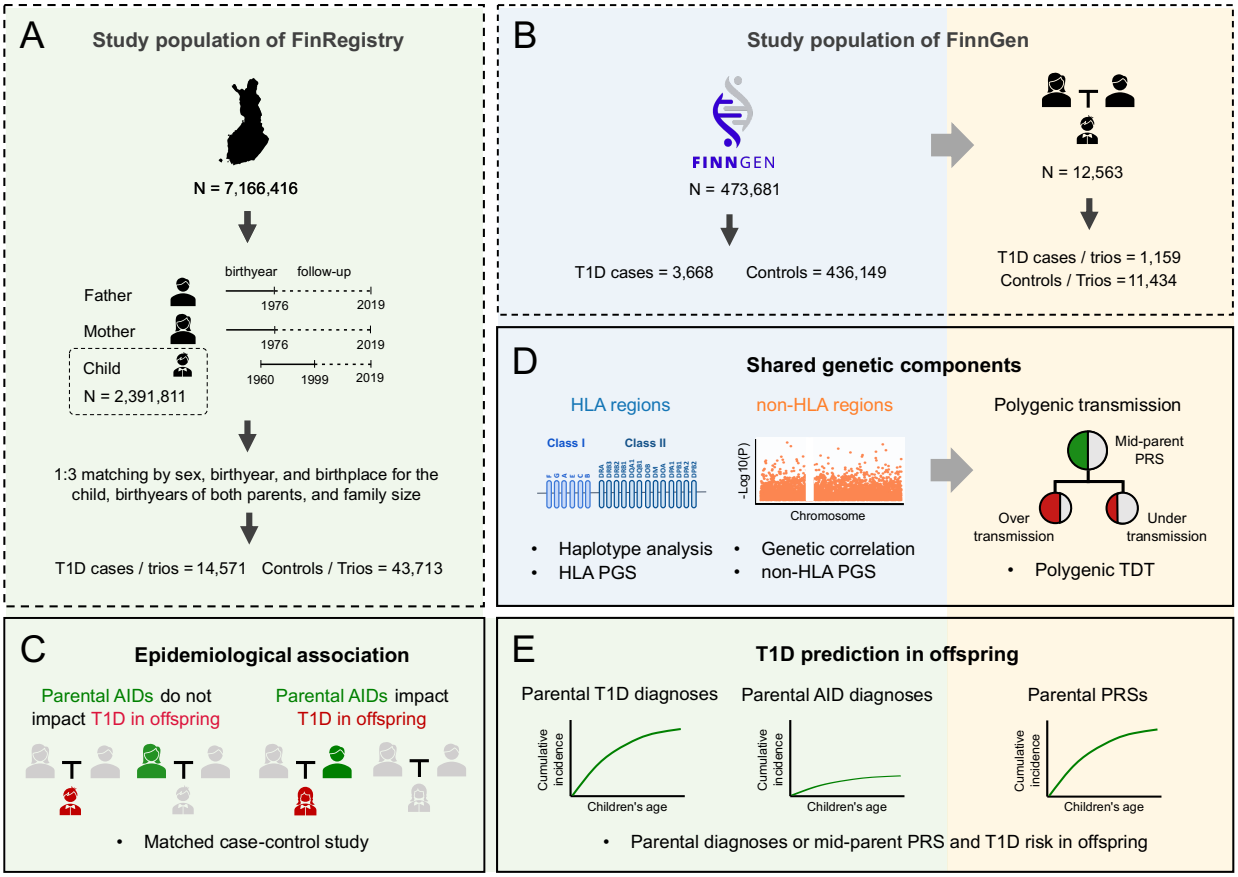

Figure 1: An overview of the study design and study populations. Panel **A**, the study population of FinRegistry (7.2 million individuals) represents every Finn alive in 2010 (5.3 million individuals) and their first-degree relatives. To maximize the coverage of diagnoses of AIDs for parents and T1D for children, we included only family trios with both parents born before 1976 and children between 1960-1999 (the follow-up time in 2019 was at least 45 years for parents and 20 years for children). The solid lines denote the birth year range of the study population and the dashed lines the years of follow-up. Among the 2.4 million family trios, 14,571 had a child ever diagnosed with T1D (T1D trios), and for each of the T1D-trios, we matched three control trios based on sex, birth year, birthplace, and the number of siblings of the child, as well as birth years of both parents. In total, we included 14,571 T1D-trios and their matched 43,713 control trios. Panel **B**, the study population of FinnGen includes 473.7K Finns enrolled in a nationwide network of Finnish biobanks, of which 12,563 family trios can be accurately constructed with available genomic information. FinnGen includes 3,370 T1D cases, 385,786 controls, as well as 1,129 T1D-trios and 11,434 control trios. Panel **C** illustrates a matched case-control study conducted in the 58,284 FinRegistry family trios to examine the association between parents' AID and the T1D status of offspring. The different symbols denote father, mother, and offspring, and colors the disease status (red, offspring with T1D; green, parents with AIDs; grey, individuals unaffected by T1D or other AIDs). Panel **D** depicts the examination of shared genetic components between T1D and other AIDs using population-based analyses (left panel with blue background) or trio-based family analyses (right panel with yellow background). A haplotype and PGS-based analysis of HLA variants conducted in 473.7K FinnGen participants (left); a genetic correlation analysis of the non-HLA variants utilized GWAS summary statistic data (middle); and a polygenic transmission disequilibrium test (pTDT) examined in FinnGen participants whether the AID-associated common variants as a whole were over-transmitted from AID-unaffected parents to their T1D-affected offspring (right). Panel **E**, the average AID PGS of the parents and its predictive performance of

T1D in offspring among 12,563 genotyped family trios in FinnGen. AID, autoimmune disease; T1D, type 1 diabetes; HLA, human leukocyte antigen; PGS, polygenic score; GWAS, genome-wide association study.

Results

Study populations and disease diagnoses

We leveraged nationwide socio-demographic and health registers of the entire Finnish population collected in the FinRegistry study[33] (N = 7.2 million) and a subset of 473,681 genotyped Finns from the FinnGen study[34] (**Figure 1A** and **1B**). To avoid potentially biased estimates caused by incomplete medical records, we excluded individuals who had migrated in or out of Finland by 2019 (572,640 were excluded for FinRegistry and 9,196 for FinnGen). To have comparatively complete coverage of AID diagnoses for parents and T1D diagnoses for children, we considered parents born before 1976 and children born between 1960 and 1999, corresponding to at least 20 years of follow-up for children by the end of 2019 (**Figure S2.2**). We defined T1D cases as those who had their first diagnosis recorded before the age of 40. In total, we included 14,571 T1D cases in FinRegistry (0.61%) and 3,668 in FinnGen (0.83%). To examine the association between AIDs in parents and T1D in offspring, as well as the extent to which genetic factors could contribute to the identified association, we constructed 2.4 million family trios in FinRegistry based on Finnish multi-generation registers and 12,563 genotype-inferred trios in FinnGen (**Methods**). Of these trios, 14,571 in FinRegistry and 1,129 in FinnGen had children with T1D.

To have broad coverage of AIDs affecting different tissues or organ systems, we defined 50 AIDs or autoimmune-related disorders based on Finnish versions of the International Classification of Diseases (ICD) codes 8-10 (**Supplementary notes 2.1**). After excluding 13 AIDs with less than 50 parental cases in FinRegistry and 11 AIDs that were highly correlated with other AIDs (**Table S2.1**), we included 26 AIDs for further analyses: T1D, adrenocortical insufficiency (ADDISON), autoimmune hyperthyroidism (HYPER), autoimmune hypothyroidism (HYPO), autoimmune haemolytic anaemias (AIHA), allergic purpura, vitamin B12 deficiency anemia (B12A), idiopathic thrombocytopenic purpura (ITP), sarcoidosis, primary biliary cholangitis (PBC), celiac disease (CD), inflammatory bowel disease (IBD), IgA nephropathy, ankylosing spondylitis (AS), mixed connective tissue disease (MCTD), rheumatoid arthritis (RA), Sjögren's syndrome (SjS), systemic sclerosis, Wegener granulomatosis (WG), systemic lupus erythematosus (SLE), Guillain-Barré syndrome (GBS), multiple sclerosis (MS), myasthenia gravis (MG), alopecia areata (AA), psoriasis, and vitiligo. The number of parental cases ranged from 55 for AIHA (corresponding to a prevalence of 0.05%) to 11,485 for HYPO (prevalence = 9.85%) (**Table S2.1** and **S3.1.1**).

We designed four sets of complementary analyses (**Figure 1**). We first examined the association between parental AID and offspring T1D risk in FinRegistry. For AIDs showing significant epidemiological associations, we explored the HLA-related associations with haplotype and PGS-based analyses in FinnGen, and the genome-wide non-HLA associations with a summary statistic-based genetic correlation analysis. A polygenic transmission disequilibrium test (pTDT) was then used to examine the within-family transmission of AID genetic signals from AID-unaffected parents to their T1D-affected and unaffected offspring. Lastly, we proposed a novel parental PGS method, integrating both HLA and non-HLA contributors, to assess the risk of developing T1D in offspring.

Epidemiological associations between parental AIDs and T1D in offspring

We first wanted to examine the associations between parental AIDs and T1D in children using the large number of family trios available in the FinRegistry study. For each of the 14,571 trios with T1D-affected children, we employed a 1:3 matched case-control design considering the information of both the child (sex, birth year, birthplace, and number of siblings) and parents (birth year) (**Methods** and **Figure 1C**). For other available socio-demographic factors that were not used in matching, we saw limited differences between T1D trios and their matched controls (**Table S3.1.2**). Overall, children with T1D (42.2% [95% CI 41.4%, 43.0%]) were more likely to have AID-affected parent(s) compared to those without T1D (31.9% [31.5%, 32.3%]).

Of 26 parental AIDs examined in FinRegistry, fifteen were associated with increased risk of T1D in offspring at a nominal P-value <0.05, and nine remained statistically significant after Bonferroni correction for multiple-testing (0.05/26 = 0.002) (**Figure 2** and **Table S3.2.1.1**). The strongest association was seen for T1D (odds ratio (OR) [95% CI], 6.77 [5.44, 8.42], $P=3.8\times10^{-66}$), followed by CD (2.14 [1.90, 2.42], $P=2.5\times10^{-35}$), B12A (1.76 [1.54, 2.02], $P=1.4\times10^{-16}$), MCTD (1.54 [1.24, 1.92], $P=8.5\times10^{-5}$), HYPO (1.53 [1.46, 1.60], $P=4.8\times10^{-73}$), RA (1.48 [1.37, 1.60], $P=1.8\times10^{-23}$), sarcoidosis (1.45 [1.28, 1.65], $P=1.1\times10^{-8}$), HYPER (1.42 [1.20, 1.68], $P=4.3\times10^{-5}$), and SjS (1.33 [1.12, 1.59], $P=1.6\times10^{-3}$). In terms of sex, we didn't observe significant differences either for parents or for offspring (**Figure S3.2.2**). However, as previously reported in the literature[36-38], fathers with T1D had a higher risk of having children with T1D than mothers with T1D (7.49 [5.68, 9.86] vs. 5.71 [4.01, 8.13], $P_{difference}= 0.24$), but this was not significant. We also noted that, a paternal history of T1D could significantly impact the onset age of T1D among boys, resulting in T1D being diagnosed 1.85 [0.29, 3.41] years earlier compared to having a T1D-unaffected father (P=0.02) (**Supplementary notes 3.2.3**). A sensitivity analysis restricted to early-onset T1D (before age 20) yielded stronger associations for parental early-onset T1D and parental SLE from the main analysis (before age 40). For example, the OR increased from 6.77 [5.44, 8.42] to 10.90 [7.90, 15.04] for parental early-onset T1D ($P_{difference}=0.02$) and from 1.29 [0.96, 1.73] to 1.77 [1.24, 2.51] for parental SLE ($P_{difference}=0.18$) (**Table S3.2.1.2**).

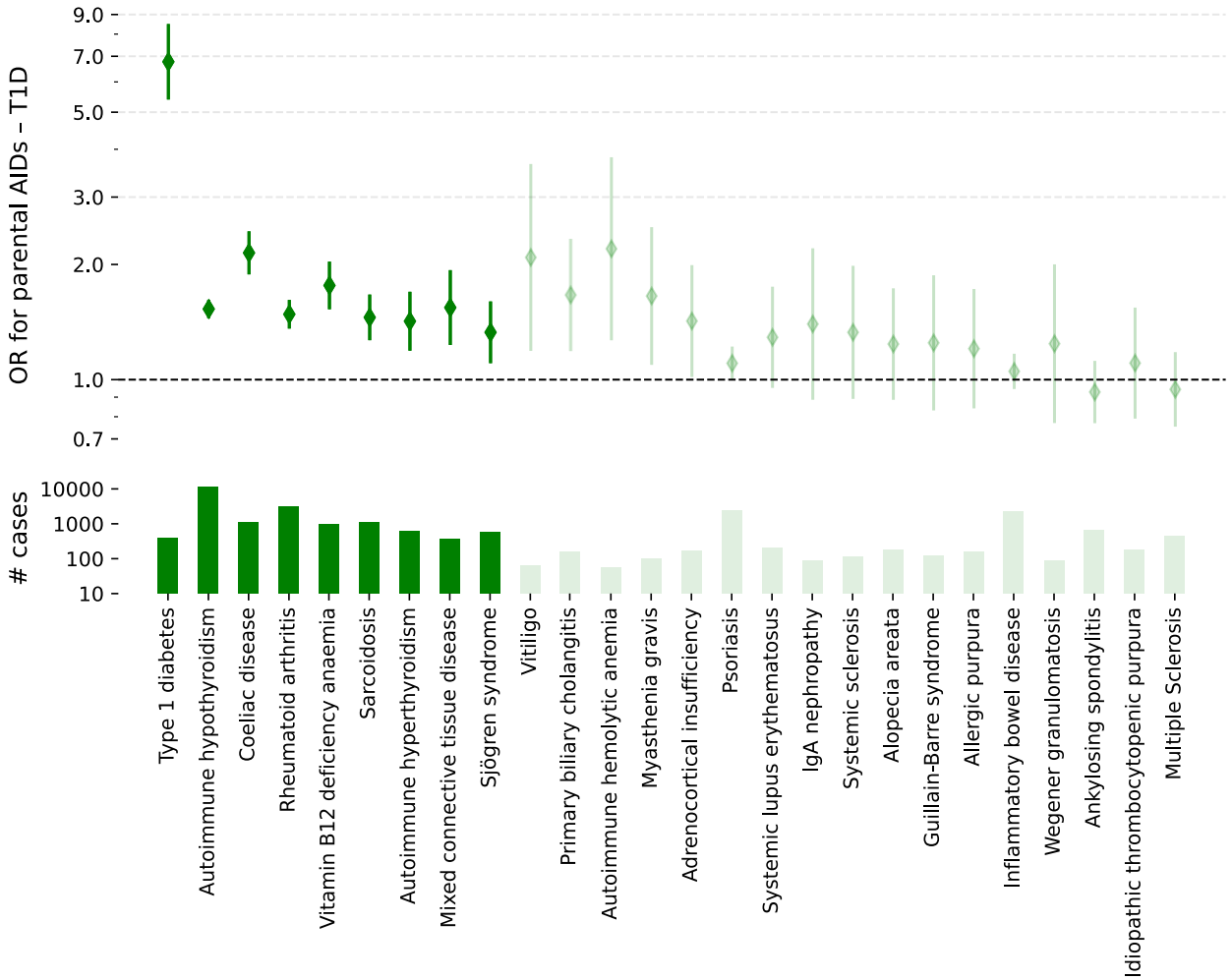

Figure 2: Epidemiological associations between parental AIDs and T1D in offspring ordered by P-value. Upper panel: Association (OR, and 95% CI) between parental AIDs and T1D in offspring using a matched case-control design in Finnish nationwide registry data (FinRegistry). Lower panel: number of T1D cases with parents having a given AID diagnosis. The dark green diamonds and bars indicate AIDs that are significantly associated with T1D after multiple testing corrections. AID, autoimmune disease; T1D, type 1 diabetes; OR, odds ratio; CI, confidence interval.

Shared genetic components from HLA and non-HLA variants at a population level

Parents and offspring share many environmental exposures and genetic components. We next aimed to quantify the extent to which the identified epidemiological associations are attributable to a shared genetic background. Given that T1D is known to be impacted by both HLA and non-HLA variations[2,8-18], we analyzed HLA and non-HLA regions separately at a population level (**Methods**). We first examined the association of the main T1D predisposing / protective HLA haplotypes with other AIDs to understand to what extent a single haplotype can drive the identified epidemiological associations. Next, we combined multiple HLA imputed alleles to create multi-allelic scores for HLA for each AIDs and examine the association with T1D. The analyses were conducted on 439,817 individuals from FinnGen (**Figure 1B**, **Table S3.1.1**). For non-HLA variations, we estimated genetic correlation between T1D and other AIDs using the latest GWAS summary statistics.

Shared effect of HLA haplotypes for T1D and other AIDs

Given that for T1D, the strongest genetic associations are for Class II haplotypes, in particular HLA-DR3 and DR4 haplotypes[8-14,39], we constructed 19 *DRB1-DQA1-DQB1* haplotypes based on 64 imputed alleles of these genes in FinnGen[40] (**Methods**).

We examined 19 haplotypes (frequencies > 0.5%, **Table S3.3.1.1**), of which six were strongly associated with T1D ($P<1.0 \times 10^{-50}$), with *DRB1*04:01-DQA1*03:01-DQB1*03:02* (OR=7.54 [7.00, 8.13], $P<1.0 \times 10^{-602}$) conferring the strongest susceptibility and *DRB1*15:01-DQA1*01:02-DQB1*06:02* (0.06 [0.04, 0.07], $P=2.0\times10^{-106}$) the strongest protection (**Table S3.3.1.2**). We then tested the association of these six haplotypes with the 24 other included AIDs using logistic regression and considering the age, sex, and first ten principal components (PCs) as covariates (**Figure 3A**). Many of these T1D-associated haplotypes were also associated with other AIDs ($P<1.0 \times 10^{-4}$) (**Table S3.3.1.3**). For example, the T1D risk haplotype *DRB1*03:01-DQA1*05:01-DQB1*02:01* increased the risk of both CD (15.83 [14.56, 17.20], $P<1.0 \times 10^{-918}$) and autoimmune hyperthyroidism (2.42 [2.23, 2.63], $P=6.3\times10^{-100}$), and *DRB1*04:01-DQA1*03:01-DQB1*03:02* increased the risk of RA (1.93 [1.84, 2.02], $P=5.1 \times 10^{-174}$). Of note, opposite effects were seen for some AIDs: the strongest haplotype protecting against T1D (*DRB1*15:01-DQA1*01:02-DQB1*06:02*) was the lead risk haplotype for MS[41] (2.85 [2.60, 3.12], $P=7.9\times10^{-112}$) and the strongest T1D susceptibility haplotype (*DRB1*04:01-DQA1*03:01-DQB1*03:02*) protected against IBD (0.85 [0.80, 0.90], $P=4.7\times10^{-8}$).

Multi-allelic scores (PGS) of HLA alleles for AIDs and T1D

Having observed that multiple HLA haplotypes could contribute to the susceptibility to T1D and other AIDs independently, we constructed polygenic HLA scores (HLA PGSs) to summarize the overall multi-allelic HLA effects for each AIDs. Considering the widespread variation in HLA allele frequencies across populations[42], we opted to construct HLA PGSs within the FinnGen samples 187 imputed Class I and Class II alleles using weighted ridge regression (**Methods**). We were able to construct reliable HLA PGSs ($P\leq0.05/26$ and partial correlation $|\rho|\geq2\%$ for predicting the susceptibility of the disease itself) in 23,336 individuals for 14 AIDs (including T1D). Of these, HLA PGSs for 13 AIDs exhibited significant associations with T1D susceptibility after multiple-test correction (0.05/14 = 0.003), with exceptions for psoriasis ($P=3.6\times10^{-3}$) and AS ($P=6.1\times10^{-2}$) (**Figure 3B** and **Table S3.3.2.1**). Overall, the effects of AID HLA PGS on individuals' own T1D risk were highly correlated with the effects of parental AIDs on offspring T1D obtained from our FinRegistry epidemiological analyses (ρ from a Spearman's rank correlation = 0.63, $P=6.1\times10^{-4}$). Among all AIDs other than T1D (OR=5.33 [4.79, 5.93], $P =3.74\times10^{-207}$), strongest associations were seen for ADDISON (OR=2.30 [2.13, 2.49], $P =4.2\times10^{-100}$), RA (OR=2.10 [1.95, 2.27], $P=1.8\times10^{-78}$), and CD (OR=2.07 [1.94, 2.22], $P =2.4\times10^{-96}$). Consistent with the HLA haplotype analysis, we observed negative associations with T1D for HLA PGS of IBD (OR=0.64 [0.60, 0.69], $P=1.2\times10^{-30}$) and MS (OR=0.81 [0.75, 0.88], $P=2.4\times10^{-7}$). These associations were further replicated with a distinct analytic approach - leave-one-group-out (LOGO) (**Supplementary notes 2.3.2** and **Table S3.3.2.2**).

Genome-wide non-HLA correlation between T1D and AIDs

247 In addition to the major HLA contributors, non-HLA variants identified in the GWAS studies are known
248 to impact AID susceptibility as well. We, therefore, explored the shared correlation between T1D and other
249 AIDs using a summary statistics-based method - LD score regression (LDSC)[43,44], excluding the HLA
250 regions (**Methods**). Significant positive genetic correlations (r_g) were observed between T1D and ten AIDs
251 ($P<0.05/25$ after Bonferroni correction) (**Figure 3C**), including HYPO, CD, RA, B12A, sarcoidosis,
252 HYPER, vitiligo, PBC, MG, and SLE. The first six AIDs were also among the parental AIDs that exerted
253 significant effects on offspring T1D in our FinRegistry epidemiological analyses (ρ from a Spearman's
254 rank correlation=0.30, P=0.15). The highest r_g were seen for B12A (r_g=0.48 [0.34, 0.63], $P=6.4\times10^{-11}$),
255 followed by HYPO (0.43 [0.34, 0.51], $P=2.9\times10^{-23}$) and RA (0.43 [0.25, 0.61], $P=3.9\times10^{-6}$) (**Table**
256 **S3.3.3.1**). Contrary to the negative HLA association, we did not observe a significant non-HLA genetic
257 correlation between T1D and MS or between T1D and IBD. A sensitivity analysis excluding all
258 chromosome 6 variants yielded similar estimates, suggesting that the observed non-HLA results were not
259 driven by variants in highly linkage disequilibrium with HLA alleles (**Table S3.3.3.2**).

260 In summary, both the HLA and non-HLA analyses recapitulated the multi-generational epidemiological
261 effects observed in nationwide registers. While the results conferred that the risk of T1D was positively
262 associated with the risk of many other AIDs regarding both HLA and non-HLA genetic variants (e.g.,
263 HYPO, CD, RA), the opposite held for MS and IBD concerning the HLA susceptibility whereas no genetic
264 correlation was observed in non-HLA regions.

265

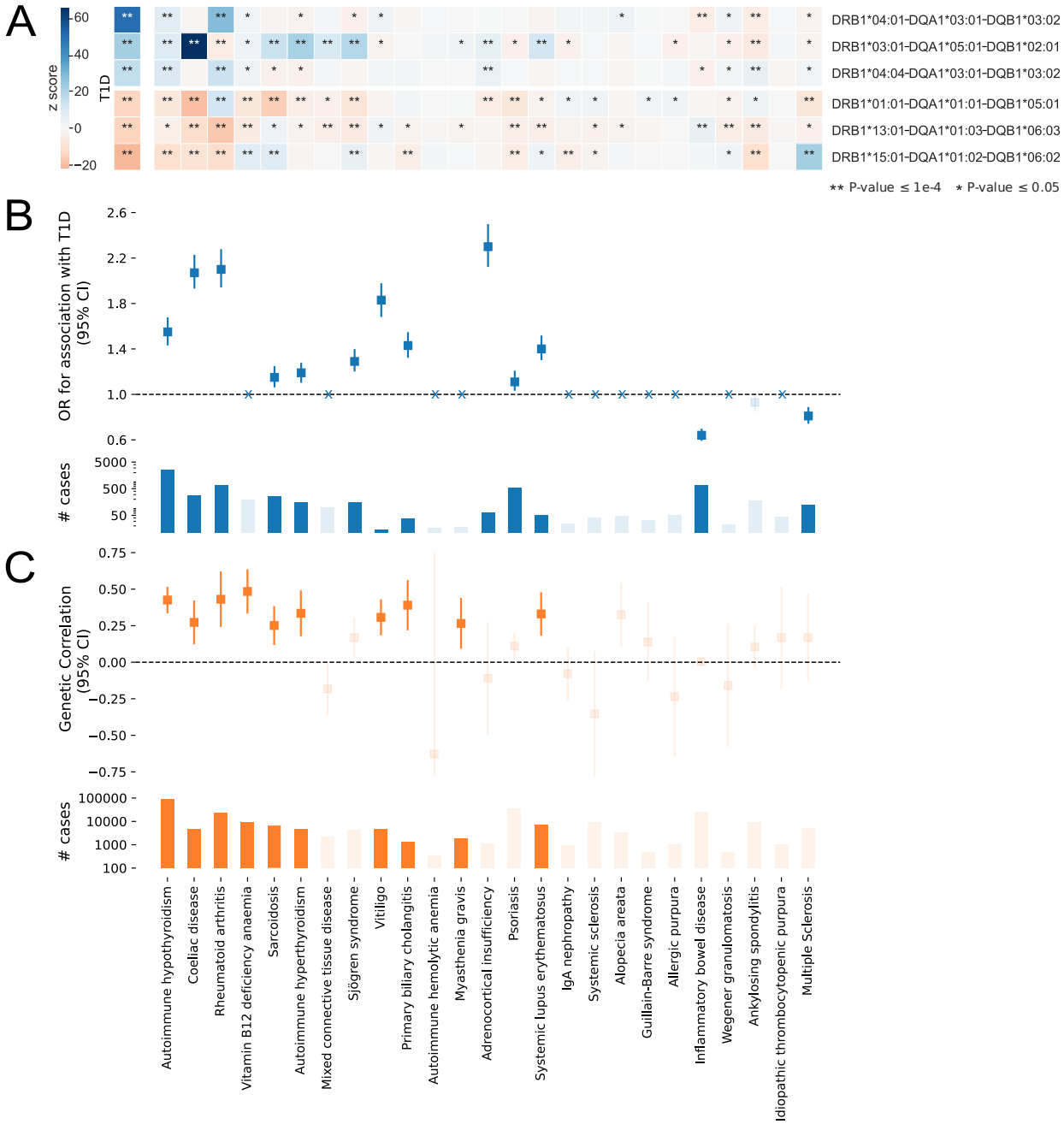

Figure 3: Shared genetic background between T1D and other AIDs, stratified by HLA and non-HLA variations at a population level. **Panel A**, HLA haplotypes strongly associated with T1D and their associations with other AIDs in FinnGen (3,668 T1D cases and 436, 149 T1D controls). The strength of the association is shown as a heatmap with blue depicting a susceptible haplotype associated with the AID in question. The haplotypes are sorted by the P-value for T1D as shown on the left. **Panel B**, upper panel: association (OR, and 95% CI) between HLA PGSs for different AIDs and T1D in FinnGen. The dark blue squares indicate that HLA PGS for a given disease has a significant association with T1D after multiple testing corrections. "×" denotes that an HLA PGS could not be robustly constructed for that AID (**Methods**). Bottom panel: blue bars represent the number of individuals with the given disease. **Panel C**, upper panel: non-HLA based genetic correlations (r_g and 95% CI) between AIDs and T1D using GWAS summary

statistics from European populations. Bottom panel: the number of cases for the given disease. The dark orange squares or bars indicate that the AIDs have a significant r_g with T1D after multiple testing corrections. T1D, type 1 diabetes; AID, autoimmune disease; HLA, human leukocyte antigen; OR, odds ratio; CI, confidence interval; PGS, polygenic score; r_g, genetic correlation; GWAS, genome-wide association study.

Transmitted and non-transmitted genetic liability within families

Polygenic transmission disequilibrium for HLA and non-HLA genetic factors

Following the population-based epidemiological evidence for clustering of other AIDs and T1D in the trios, and the genetic evidence for a shared genetic origin of the AIDs and T1D at a population level, we wanted to examine the transmission of the HLA and non-HLA within families using a polygenic transmission disequilibrium test (pTDT)[45]. The pTDT examines whether the AID risk variants are observed to be over-transmitted (or under-transmitted) to the offspring with T1D (compared to the expected transmission rate) and thus it is immune to many of the potential confounders arising from population studies on unrelated individuals. We considered 12,563 trios (genotyped in FinnGen), of which 1,159 had offspring with T1D (9.2%). The prevalence of T1D in these trios was higher than expected in the general population because of the inclusion of several studies have specifically targeted individuals with T1D. We analyzed 10 AIDs with reliable PGSs ($\rho \geq 2\%$) for both HLA and non-HLA genes and additional six diseases with only reliable PGSs for the HLA regions (**Methods**; **Supplementary notes 3.3.2** and **Table S3.4.1**). For offspring with T1D, both HLA and non-HLA PGSs for T1D deviated significantly from their mid-parent value (1.23 [1.16, 1.30], $P=2.6\times10^{-168}$ and 0.69 [0.62, 0.75], $P=9.6\times10^{-74}$) while no such deviation was observed in unaffected siblings (-0.06 [-0.13, 0.01], P=0.08 and 0.00 [-0.06, 0.05], P=0.88) (**Figure 4** and **Table S3.4.2**). Overall, two major patterns were seen across the analyzed AIDs. The first group encompassed T1D, HYPO, RA, and SLE, for which significant over-transmission in offspring with T1D was seen for both HLA and non-HLA PGSs. The second group exhibited significant transmission only for HLA but not non-HLA PGSs: while psoriasis exhibited significant over-transmission in offspring with T1D compared to unaffected siblings, IBD and MS had a significant under-transmission. Taking IBD as an example, among 2,040 genotyped trios, 756 offspring with T1D presented under-transmitted HLA PGS (-0.38 [-0.45, -0.31], $P=3.1\times10^{-22}$) while their 1,269 unaffected siblings had comparable PGS to their parents (-0.01 [-0.08, 0.06], P=0.17). No diseases showed significant pTDT only regarding non-HLA PGS without a significant association for HLA-PGS.

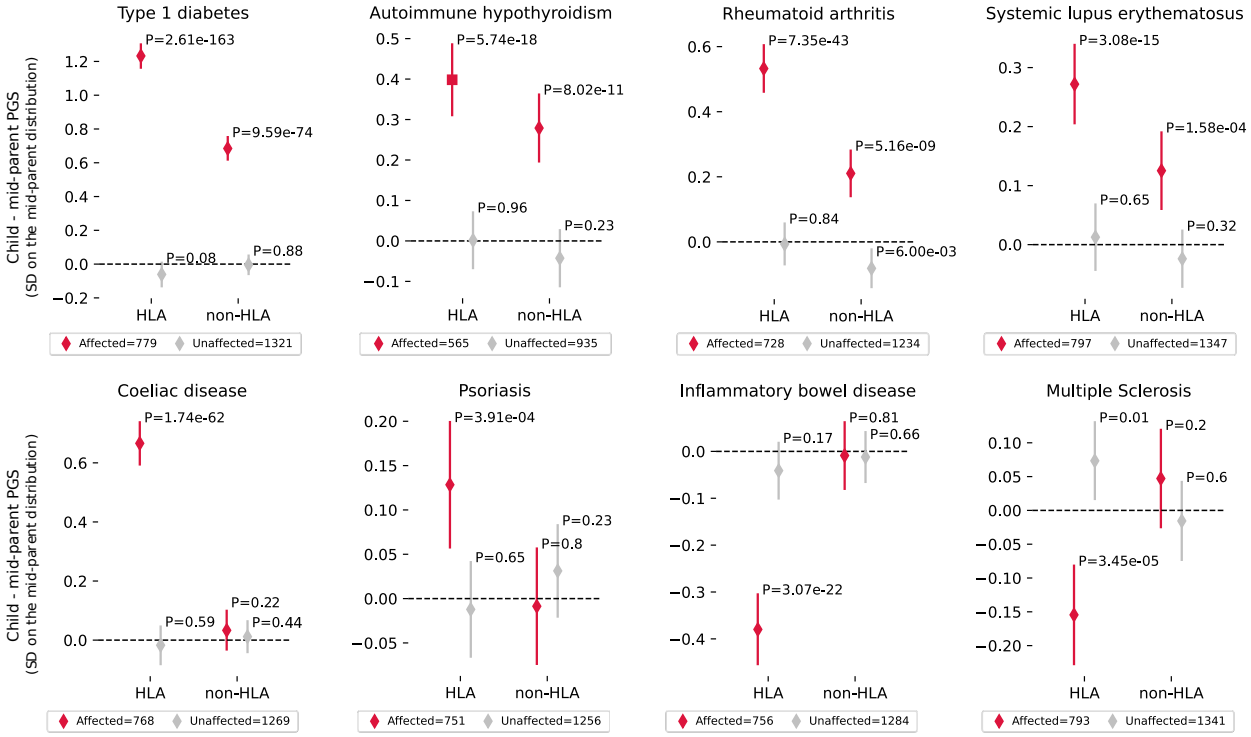

Figure 4: Polygenic transmission disequilibrium tests (pTDT) assess whether AID PGSs transmitted to offspring significantly deviated from the parental PGS in T1D-affected offspring and their unaffected siblings. Deviations from parental PGS (mean and 95% CI) of 1 SD change in offspring PGS are shown separately for HLA and non-HLA PGSs, with the P value obtained from a two-sided t-test. Red color denotes offspring with T1D and gray unaffected siblings. AID, autoimmune disease; PGS, polygenic score; T1D, type 1 diabetes; CI, confidence interval; SD, standard deviation; HLA, human leukocyte antigen.

Offspring T1D risk based on parental AIDs clinical history and genetic information

The PGS for T1D has shown potential to predict disease susceptibility before its onset[19-21]. We asked whether T1D risk in offspring can be assessed using a couple's genetic information before they plan to have a child, and how this compares to estimating the risk based on family history of T1D and other AIDs, as is currently being done in clinical practice. To address this question, we adopted a novel strategy to construct a full PGS (Full-PGS) by considering the contribution ratio of HLA PGS relative to non-HLA PGS (**Methods**; **Supplementary notes 3.5**). We tested eight AIDs (T1D, HYPO, RA, SLE, CD, psoriasis, IBD, and MS) with at least one reliable PGS in HLA regions or non-HLA regions for T1D prediction ($|\rho|\geq 1\%$) (**Table S3.5.1**) using five-fold cross validation. By modelling the HLA and non-HLA variations separately and using imputed HLA alleles (187 alleles in 10 classical HLA genes rather than common genotyped SNPs), the Full-PGSs we proposed outperformed the standard SNP-based PGS methods such as PRS-CS[46] at an individual level (**Table S3.6.1**; **Table S3.6.2**) and when predicting T1D in offspring using AID-specific parental PGS among the 12,563 FinnGen genotyped trios (**Table S3.6.3**; **Figure S3.6.4**).

Having established that a parental Full-PGS for T1D was strongly associated with T1D risk in offspring (AUC_{mean}=0.817), we tested whether the parental Full-PGSs for all the eight AIDs could add any additional information to the parental Full-PGS for T1D. However, limited difference was seen by integrating the PGS of eight AIDs (AUC_{mean}: 0.817 vs 0.820, $P_{difference}$=0.77). This suggested that, instead of a direct effect, the

impact of parental AID PGS on offspring T1D was most likely to be mediated through the genetic correlation between T1D and AID, and therefore provided limited additional information to the offspring T1D prediction when parental T1D PGS was already included and sufficiently accurate.

We thus focused on parental Full-PGS for T1D and estimated the cumulative risk of T1D in offspring using a Cox proportional hazards model adjusted for the first 10 PCs and birthyear of the child among genotyped trios (**Methods**). When stratified PGS into four groups by percentiles (0-50th, 50-90th, 90-99th, 99-100th), we observed distinct trajectories in terms of T1D cumulative incidence rates (0.24%, 0.74%, 2.29%, 6.96% by age 20) (**Figure 5A**). That is, children whose parents were in the top percentile of T1D Full-PGS had a 29-fold higher risk of developing T1D than children whose parents were within the bottom 50% of T1D Full-PGS. When further stratified by the sex of children (**Figure 5B** and **5C**), the adjusted survival curves suggested that sons overall had a higher cumulative incidence than daughters across all PGS groups. For example, in the top T1D parental Full-PGS group, sons (11.48%) could have a 2-fold higher cumulative risk than daughters (5.29%) by age 20. Overall, the cumulative incidence curves of T1D started to level off around the age of 14 to 16.

We then wanted to understand how informative the Full-PGS was compared to parental AID clinical history. In general, the offspring of T1D couples had a higher cumulative incidence of T1D risk (**Figure 5D**) than T1D-unaffected couples and such differences were larger for boys (**Figure 5E**) than girls (**Figure 5F**). Having either father or mother with T1D ($prevalence_{father}$ = 0.22%, $prevalence_{mother}$ = 0.15%) resulted in a lower cumulative risk than having both parents in the top percentile of Full-PGS for T1D (1.49% and 4.16% vs 6.95% by age 20 for children). Having both parents affected by T1D (a very rare event with only a prevalence of 8.5 per million) resulted in a higher cumulative risk (11.30%) than having parents in the top percentile of Full-PGS for T1D (6.95%). Among couples only affected by AIDs other than T1D, offspring also had a higher cumulative incidence of T1D than the general population (**Fig. 5 G**, **H**, and **I**) (0.96 % vs 0.48%), although the cumulative incidence would be smaller than that for the offspring from couples with physician-diagnosed T1D or couples in the top decile of the Full-PGS for T1D.

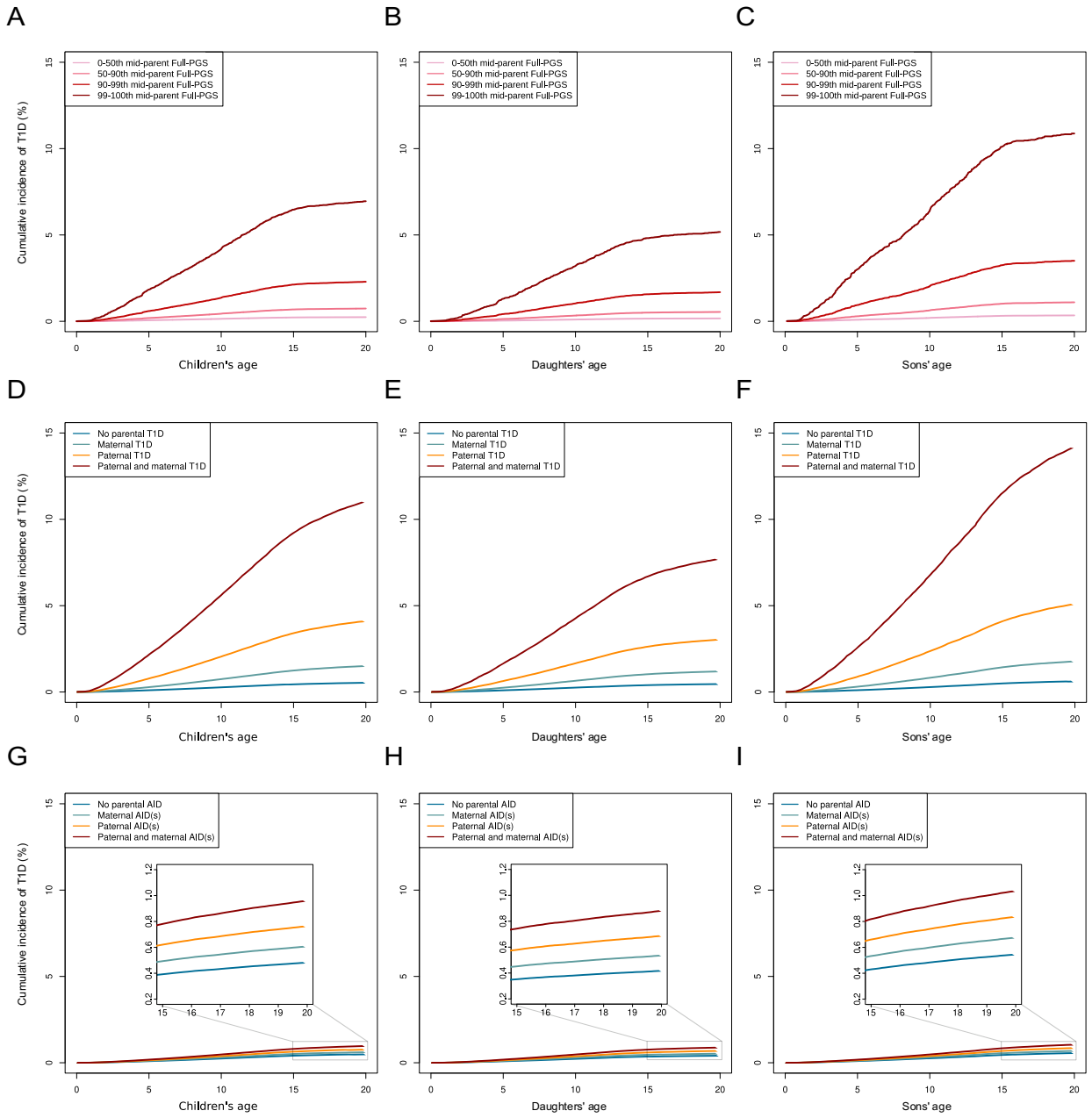

Figure 5: Adjusted survival curves from Cox proportional hazards models for cumulative incidence of T1D in offspring by age 20, stratified by parental full T1D PGS percentiles comprising both HLA and non-HLA variants (**Panels A-C**; 8,872 FinnGen genotyped trios), parents' T1D status (**Panels D-F;** 3,048,812 FinRegistry trios), parents' other AIDs (**Panels G-I**; 3,037,723 FinRegistry trios after removing trios having parent(s) diagnosed with T1D), including T1D, autoimmune hypothyroidism, rheumatoid arthritis, systemic lupus erythematosus, coeliac disease, psoriasis, inflammatory bowel disease, and multiple sclerosis; The panels show pooled data for both sexes (left), daughters (middle), and sons (right). T1D, type 1 diabetes; AID, autoimmune disease; HLA, human leukocyte antigen; PGS: polygenic score.

Discussion

In this study, we comprehensively explored the genetic determinants of the familial aggregation of T1D and other AIDs with data from high-dimensional nationwide registers and rich genetic information of the Finnish population. Our long follow-up (≥20 years) in offspring and the early age of T1D onset (median age = 12.7) allows us to maximize the number of T1D cases identified in the study. With systematically designed epidemiological and genetic analyses, we seek to answer three key but underexplored questions: which parental AIDs are related to T1D in offspring, to what extent the identified parental AID-offspring T1D association is attributable to genetic polymorphisms of HLA and non-HLA genes, and last, for any couples that are planning to have children, how well we could use their genetic information to evaluate the T1D risk of their offspring.

The nationwide registers of 7.2 Finns collected in the FinRegistry study cover longitudinal health and sociodemographic information of 58,284 family trios which provides us a unique opportunity to perform a comprehensive exploration of 50 autoimmune diseases for their impacts on offspring T1D, on a scale and detailed level that could be difficult to achieve in questionnaire- or survey-based studies. This rich familial and health information enabled us to perform an efficient matched case-control design to better control for unmeasurable confounding factors than the standard unmatched design used in previous large population-based studies[25-27]. In total, we detected ten parental AID-offspring T1D associations, including T1D, HYPO, CD, RA, HYPER, and B12A, that have previously been linked to T1D risk in offspring in large population-based studies (**Table S4**), and more importantly, six novel associations encompassing AIHA, MCTD, MG, AA, psoriasis, and vitiligo.

Following the epidemiological analyses, we leveraged the genetic information of 470K genotyped Finns enrolled in the FinnGen study to better understand the genetic causes of the observed family aggregations. Given that T1D is known to be impacted by both HLA and non-HLA variations[2,8-18], we designed a set of genetic analyses in a hypothesis-free manner to separate the contributions from HLA and non-HLA regions for each analyzed AID. These genetic analyses were conducted both at a population level, studying the shared effects of HLA and non-HLA variations among unrelated individuals, and, within families using pTDT in 12,563 FinnGen family trios, which to the best of our knowledge, is the largest family-based genetic analysis performed for T1D and AIDs. Similar to the epidemiological family-based study that is widely accepted to yield less biased estimates than population-based studies, pTDT can condition out the impact of shared familial factors through within-family comparisons, allowing, for the first time, to examine under- or over-transmission of genetic risk factors for AIDs in individuals with T1D and their unaffected siblings.

Generally, two major patterns were seen for the transmission of AID-associated variants to offspring with T1D, with one group encompassing T1D, HYPO, RA, and SLE showing significant over-transmission for both HLA and non-HLA variants, and another group comprising IBD, MS, CD, and psoriasis that exerted significant deviated transmission only for HLA. We noted that even among the disease group presenting over-transmission for both HLA and non-HLA, more significant transmission was seen regarding HLA compared to non-HLA. This is consistent with the previous studies which reported strong HLA associations for many AIDs[2,47,48].

Previous works over decades have shown evidence for variation in genes in both HLA and non-HLA regions being associated with and between AIDs, especially after the advances in the development of GWAS and multi-omics data in recent years[47,49-51]. Specifically, genes modulating adaptive immunity through T cell activation and signaling (e.g., *PTPN22*[52,53], *TAGAP*[54-57], *CD226*[58-60]), B cell activation (e.g., *KIF5A/PIP4K2C*[61-63], *IL10*[64-66]), T helper cells (e.g., *STAT1-STAT4*[67-72], *PRKCQ*[73-75]), and regulatory T cells[57,76-79], are involved. Variation has also been shown in genes impacting innate immunity, which may play an important role in viral or microbial infections through the pathogen recognition receptor pathways,

427 transcription factors, apoptosis, autophagy, and immune-complex clearance. A cross-disease genomic
428 analysis of nine AIDs, such as T1D, RA, SLE, suggested that the top prioritized genes of each analyzed
429 AID converged on the same common pathways relevant to T cell activation and signaling, although distinct
430 genes were prioritized across diseases[80]. Many of these works indicated that T1D shared biological
431 pathways with other AIDs, especially RA, CD, and SLE, which to some extent explained why they were
432 tightly associated with T1D in our analyses.

433 The interplay of HLA and non-HLA variants on disease susceptibility turned out to be complex and disease-
434 dependent. Especially, our HLA haplotype analysis showed that multiple T1D lead haplotypes exerted
435 strong effects on other AIDs, but largely in a disease-dependent manner. The HLA effects regarding other
436 AIDs and T1D might also be in opposite directions as illustrated by some HLA haplotypes. For example,
437 being the most protective haplotype of T1D, *DRB1*15:01-DQA1*01:02-DQB1*06:02* meanwhile strongly
438 increased the risk of MS. This opposite pleiotropic HLA was consistantly observed by multi-allelic HLA
439 PGS. However, in non-HLA regions, we observed no link between T1D and MS from either genetic
440 correlation or pTDT. IBD had a similar pattern in these analyses.These findings, taken together, can point
441 to potential explanations for the discrepancies observed between the epidemiological evidence of familial
442 aggregation and the genetic results. In particular, our results suggest that HLA risk factors for IBD and MS
443 are protective of T1D, while their non-HLA risk factors are not associated with T1D. The overall lack of
444 familial aggregation suggested that this HLA protective effect could be counterbalanced by other non-
445 captured genetic effects or by environmental effects shared within the family.

446 Most AIDs are more prevalent in women than in men and disease progression and severity could also differ
447 between the two sexes[37,81]. T1D is an exception in that a slight majority of at least pediatric cases are boys
448 and twice as many fathers than mothers of the patients with T1D also have T1D[37,38]. In this study, we also
449 observed that among individuals with T1D, men had a higher risk of having offspring with T1D than women,
450 while significant sex-specific associations affecting transmission were overall not observed among parental
451 AIDs. Further, when at least one parent had T1D, especially if it was the father, the male offspring would
452 have a higher risk of developing T1D than the female offspring. To date, the precise cause of a higher rate
453 of T1D transmission from fathers than mothers has not been identified, but potential hypotheses largely
454 revolve around in utero exposure to different aspects of maternal T1D and or its consequences e.g. exposure
455 to maternal hyperglycemia, exogenous insulin, maternal islet antibodies, and maternal enterovirus
456 antibodies[38]. It has also been proposed that a protective effect of maternal T1D could be mediated by an
457 *E.coli* dominant maternal microbiome, which fosters the development of neonatal T cells in the infant[82-85].

458 Previous studies also showed that people with AIDs, are less likely to have children[86]. This is likely to
459 reflect multiple factors, including a higher rate of miscarriage amongst women with AIDs including T1D,
460 amongst whom the rate of miscarriage is 15-30% higher than in the background population [87-90]. In some
461 cases, individuals with AIDs may choose not have children due to a fear of transmitting disease to offspring.
462 Being able to estimate early T1D risk trajectories for a couple's offspring, utilizing their PGS and AID
463 disease status, will equip them with more accurate information to make such decisions around family
464 planning, and provide reassurance to those whose offspring have a low predicted risk of T1D. For
465 individuals with offspring predicted to be at high risk of disease, recent advances in T1D research may
466 mean that targeted screening and/or the option of early preventative therapies may be a realistic option.

467 So far, we are not aware of any previous study using parental PGS in estimating an AID. We introduced a
468 new approach to calculate PGSs for diseases that have a strong HLA signal, by training different models
469 for HLA and non-HLA regions and combining them based on their expected contribution to the disease.
470 We showed that parental PGS for T1D is strongly predictive of T1D risk before age of 20 and that parental
471 PGS for other AIDs does not help in better predicting T1D. We showed that a T1D PGS that represents the
472 genetic risk factors from parents is positively associated with developing T1D in offspring, especially sons.
473 The cumulative risk trajectories start to plateau around the age of 14-16, especially for children with high

474 parental PGS (top 10%), which is consistent with the incidence peak at 10-14 years reported in previous
475 studies[5-7]. The effect of high parental PGS on T1D risk in children was in many instances higher than a
476 diagnosis of T1D in their mother or father. For example, having a mother with T1D resulted in a cumulative
477 T1D risk of 1.49% by age 20, while the top 10% of parents with the highest PGS had a cumulative risk of
478 2.88%. Considering that only 0.15% of mothers with children had T1D in Finland, the PGS provides
479 additional value on top of family history. These results indicate the intriguing possibility of considering
480 parental PGS, in conjunction with clinical diagnoses, to inform parents about T1D risk in their offspring.

481 This study also has several limitations. One limitation was that AID diagnoses for some patients might not
482 be well captured in the healthcare register. For example, the specific ICD-9 code for T1D was not
483 implemented until 1987 although the Care Register for Finnish Health Care began as early as in 1969.
484 Another limitation was that rather than being nationally representative, the amalgamated Finnish biobanks
485 represent a collection of biobanks with diverse methods of data collection. Some of the diseases might have
486 a higher prevalence in the Finnish biobanks compared to the nationally representative FinRegistry.
487 Although these aspects might limit the generalizability of the results, we noted that our findings in the
488 Finnish biobanks were consistent with those we observed from FinRegistry and matched well to previous
489 studies[26,29,31,80,81]. Thirdly, rather than extending the analysis to siblings or other relatives, in this study, we
490 focused on parents and offspring. We note the HLA and non-HLA association patterns with T1D across
491 AIDs would be similar to what we observed here although the effect sizes would differ. Finally, while many
492 risk factors (e.g., socioeconomic factors, genetic nurture) might have an impact on the development of T1D,
493 our study mainly focused on family history of diseases and genetic factors by controlling these factors with
494 delicate analytical designs.

495 In conclusion, our results, while confirming the existence of general familial aggregation of AIDs and T1D,
496 highlight a substantial heterogeneity in the impact of different AIDs on T1D. This heterogeneity is partially
497 explained by different genetic effects within and outside the HLA regions and demonstrated two different
498 transmission patterns regarding shared genetic liabilities in HLA and non-HLA. Overall, genetic effects
499 inside and outside HLA regions are consistent with observational analyses, but in the case of IBD and MS
500 we revealed unexpected divergence between the genetic effects and epidemiological observations.
501 Investigating the mechanisms behind these findings may provide valuable insights into the origins of T1D
502 and the etiology of the familial aggregations of AIDs. Moreover, the relative contribution of HLA and non-
503 HLA genetic risk can be leveraged to create powerful PGSs that can complement a family history of AIDs
504 as a tool to inform parents about expected T1D risk in their offspring.

505

506 **References**

1 Kavvoura, Fotini K., and John PA Ioannidis. "CTLA-4 gene polymorphisms and susceptibility to type 1 diabetes mellitus: a HuGE Review and meta-analysis." American journal of epidemiology 162.1 (2005): 3-16.

2 Pociot, F., and M. F. McDermott. "Genetics of type 1 diabetes mellitus." Genes & Immunity 3.5 (2002): 235-249.

3 Åkerblom, Hans K., et al. "Environmental factors in the etiology of type 1 diabetes." American journal of medical genetics 115.1 (2002): 18-29.

4 Knip, Mikael, et al. "Environmental triggers and determinants of type 1 diabetes." Diabetes 54.suppl_2 (2005): S125-S136.

5 Maahs, David M., et al. "Epidemiology of type 1 diabetes." *Endocrinology and Metabolism Clinics* 39.3 (2010): 481-497.

6 Gregory, Gabriel A., et al. "Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study." *The Lancet Diabetes & Endocrinology* 10.10 (2022): 741-760.

7 Hormazábal-Aguayo, Ignacio, et al. "Incidence of type 1 diabetes mellitus in children and adolescents under 20 years of age across 55 countries from 2000 to 2022: A systematic review with meta-analysis." *Diabetes/Metabolism Research and Reviews* (2023): e3749.
8 Noble, Janelle A., et al. "The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families." *American journal of human genetics* 59.5 (1996): 1134.
9 Thomson G, Robinson WP, Kuhner MK *et al*. Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus *Am J Hum Genet* 1988 43: 799–816
10 Lambert, A. Paul, et al. "Absolute risk of childhood-onset type 1 diabetes defined by human leukocyte antigen class II genotype: a population-based study in the United Kingdom." The Journal of Clinical Endocrinology & Metabolism 89.8 (2004): 4037-4043.
11 Luo, Shuoming, et al. "HLA genetic discrepancy between latent autoimmune diabetes in adults and type 1 diabetes: LADA China Study No. 6." *The Journal of Clinical Endocrinology & Metabolism* 101.4 (2016): 1693-1700.
12 Ikegami, Hiroshi, et al. "Genetics of type 1 diabetes: similarities and differences between Asian and Caucasian populations." *Annals of the New York Academy of Sciences*1079.1 (2006): 51-59.
13 Kawabata, Yumiko, et al. "Asian-specific HLA haplotypes reveal heterogeneity of the contribution of HLA-DR and-DQ haplotypes to susceptibility to type 1 diabetes." *Diabetes* 51.2 (2002): 545-551.
14 Rojas-Villarraga, Adriana, Diana Botello-Corzo, and Juan-Manuel Anaya. "HLA-Class II in Latin American patients with type 1 diabetes." *Autoimmunity reviews* 9.10 (2010): 666-673.
15 Todd, John A., et al. "Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes." *Nature genetics* 39.7 (2007): 857-864.
16 Barrett, Jeffrey C., et al. "Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes." *Nature genetics* 41.6 (2009): 703-707.
17 Evangelou, Marina, et al. "A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations." *Genetic epidemiology* 38.8 (2014): 661-670.
18 Robertson, Catherine C., et al. "Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes." *Nature genetics* 53.7 (2021): 962-971.
19 Winkler, Christiane, et al. "Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes." Diabetologia 57 (2014): 2521-2529.
20 Sharp, Seth A., et al. "Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis." Diabetes care 42.2 (2019): 200-207.
21 Padilla-Martínez, Felipe, et al. "Systematic review of polygenic risk scores for type 1 and type 2 diabetes." International journal of molecular sciences 21.5 (2020): 1703.
22 Patel, Kashyap A., et al. "Type 1 diabetes genetic risk score: a novel tool to discriminate monogenic and type 1 diabetes." *Diabetes* 65.7 (2016): 2094-2099.
23 Yaghootkar, H., et al. "Type 1 diabetes genetic risk score discriminates between monogenic and Type 1 diabetes in children diagnosed at the age of< 5 years in the Iranian population." *Diabetic Medicine* 36.12 (2019): 1694-1702.
24 Billings, Liana K., et al. "Utility of polygenic scores for differentiating diabetes diagnosis among patients with atypical phenotypes of diabetes." *The Journal of Clinical Endocrinology & Metabolism* 109.1 (2024): 107-113.
25 Nielsen, Nete M., et al. "Type 1 diabetes and multiple sclerosis: a Danish population-based cohort study." Archives of neurology 63.7 (2006): 1001-1004.
26 Hemminki, K., et al. "Familial association between type 1 diabetes and other autoimmune and related diseases." Diabetologia 52.9 (2009): 1820-1828.
27 Yen, Fu-Shun, et al. "Maternal autoimmune disease associated with a higher risk of offspring with type 1 diabetes: A nationwide mother-child cohort study in Taiwan." Diabetes & Metabolism (2023): 101443.
28 Vyse, Timothy J., and John A. Todd. "Genetic analysis of autoimmune disease." Cell 85.3 (1996): 311-318.
29 Parkkola, Anna, et al. "Extended family history of autoimmune diseases and phenotype and genotype of children with newly diagnosed type 1 diabetes." European Journal of Endocrinology 169.2 (2013): 171-178.
30 Wägner, Ana M., et al. "Predictors of associated autoimmune diseases in families with type 1 diabetes: results from the Type 1 Diabetes Genetics Consortium." Diabetes/metabolism research and reviews 27.5 (2011): 493-498.

31 Werme, Josefin, et al. "An integrated framework for local genetic correlation analysis." *Nature genetics* 54.3 (2022): 274-282.
32 Lettre, Guillaume, and John D. Rioux. "Autoimmune diseases: insights from genome-wide association studies." Human molecular genetics 17.R2 (2008): R116-R121.
33 Viippola, Essi, et al. "Data Resource Profile: Nationwide registry data for high-throughput epidemiology and machine learning (FinRegistry)." *International Journal of Epidemiology*(2023): dyad091.
34 Kurki, Mitja I., et al. "FinnGen provides genetic insights from a well-phenotyped isolated population." Nature 613.7944 (2023): 508-518.
35 Patterson, Christopher C., et al. "Incidence trends for childhood type 1 diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study." *The Lancet* 373.9680 (2009): 2027-2033.
36 Parkkola, Anna, et al. "Extended family history of type 1 diabetes and phenotype and genotype of newly diagnosed children." Diabetes care 36.2 (2013): 348-354.
37 Harjutsalo, Valma, Antti Reunanen, and Jaakko Tuomilehto. "Differential transmission of type 1 diabetes from diabetic fathers and mothers to their offspring." *Diabetes* 55.5 (2006): 1517-1524.
38 Allen, Lowri A., et al. "Maternal type 1 diabetes and relative protection against offspring transmission." *The Lancet Diabetes & Endocrinology* (2023).
39 Erlich, Henry, et al. "HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families." *Diabetes* 57.4 (2008): 1084-1092.
40 Ritari, Jarmo, et al. "Increasing accuracy of HLA imputation by a population-specific reference panel in a FinnGen biobank cohort." *NAR genomics and bioinformatics* 2.2 (2020): lqaa030.
41 Ramagopalan, Sreeram V., and George C. Ebers. "Multiple sclerosis: major histocompatibility complexity and antigen presentation." *Genome medicine* 1.11 (2009): 1-5.
42 Arrieta-Bolaños, Esteban, Diana Iraíz Hernández-Zaragoza, and Rodrigo Barquera. "An HLA map of the world: A comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II." *Frontiers in Genetics* 14 (2023): 866407.
43 Bulik-Sullivan, Brendan K., et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies." *Nature genetics* 47.3 (2015): 291-295.
44 Bulik-Sullivan, Brendan, et al. "An atlas of genetic correlations across human diseases and traits." *Nature genetics* 47.11 (2015): 1236-1241.
45 Weiner, Daniel J., et al. "Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders." *Nature genetics*49.7 (2017): 978-985.
46 Ge, Tian, et al. "Polygenic prediction via Bayesian regression and continuous shrinkage priors." *Nature communications*10.1 (2019): 1776.
47 Anaya, Juan-Manuel, et al. "Autoimmunity: from bench to bedside [Internet]." (2013).
48 Ritari, J., Koskela, S., Hyvärinen, K. and Partanen, J., 2022. HLA-disease association and pleiotropy landscape in over 235,000 Finns. Human Immunology, 83(5), pp.391-398.
49 Serrano, Norma C., Paula Millan, and María-Carolina Páez. "Non-HLA associations with autoimmune diseases." *Autoimmunity reviews* 5.3 (2006): 209-214.
50 Simmonds, M. J., and S. C. L. Gough. "The HLA region and autoimmune disease: associations and mechanisms of action." *Current genomics* 8.7 (2007): 453-465.
51 Shiina, Takashi, et al. "The HLA genomic loci map: expression, interaction, diversity and disease." *Journal of human genetics*54.1 (2009): 15-39.
52 Bottini, Nunzio, et al. "A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes." *Nature genetics* 36.4 (2004): 337-338.
53 Begovich, Ann B., et al. "A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis." *The American Journal of Human Genetics* 75.2 (2004): 330-337.
54 Hunt, Karen A., et al. "Newly identified genetic risk variants for celiac disease related to the immune response." *Nature genetics* 40.4 (2008): 395-402.
55 Smyth, Deborah J., et al. "Shared and distinct genetic variants in type 1 diabetes and celiac disease." *New England Journal of Medicine* 359.26 (2008): 2767-2777.
56 Raychaudhuri, Soumya, et al. "Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk." *Nature genetics* 41.12 (2009): 1313-1318.

57 Todd, John A., et al. "Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes." *Nature genetics* 39.7 (2007): 857-864.
58 Hafler, Jason P., et al. "CD226 Gly307Ser association with multiple autoimmune diseases." *Genes & Immunity* 10.1 (2009): 5-10.
59 Löfgren, Sara E., et al. "A 3′-untranslated region variant is associated with impaired expression of CD226 in T and natural killer T cells and is associated with susceptibility to systemic lupus erythematosus." *Arthritis & Rheumatism* 62.11 (2010): 3404-3414.
60 Qiu, Zhi-Xin, et al. "CD226 Gly307Ser association with multiple autoimmune diseases: a meta-analysis." *Human immunology* 74.2 (2013): 249-255.
61 Raychaudhuri, Soumya, et al. "Common variants at CD40 and other loci confer risk of rheumatoid arthritis." *Nature genetics* 40.10 (2008): 1216-1223.
62 Fung, E. Y. M. G., et al. "Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus." *Genes & Immunity* 10.2 (2009): 188-191.
63 Cohen, Jose E., Svetlana Boitsova, and Eyal Itshayek. "Cerebral venous sinus thrombosis." *Isr Med Assoc J* 11.11 (2009): 685-688.
64 Franke, Andre, et al. "Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility." *Nature genetics* 40.11 (2008): 1319-1323.
65 Barrett, Jeffrey C., et al. "Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes." *Nature genetics* 41.6 (2009): 703-707.
66 Gateva, Vesela, et al. "A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus." *Nature genetics* 41.11 (2009): 1228-1233.
67 International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), et al. "Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci." *Nature genetics* 40.2 (2008): 204-210.
68 Martinez, A., et al. "Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases." *Arthritis & Rheumatism* 58.9 (2008): 2598-2602.
69 Tong, Guanghui, et al. "Association between polymorphism in STAT4 gene and risk of rheumatoid arthritis: a meta-analysis." *Human immunology* 74.5 (2013): 586-592.
70 Liang, Ya-ling, et al. "Association of STAT4 rs7574865 polymorphism with autoimmune diseases: a meta-analysis." *Molecular biology reports* 39 (2012): 8873-8882.
71 Barrett, Jeffrey C., et al. "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease." *Nature genetics* 40.8 (2008): 955-962.
72 Jakkula, Eveliina, et al. "Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene." *The American Journal of Human Genetics* 86.2 (2010): 285-291.
73 Cooper, Jason D., et al. "Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci." *Nature genetics* 40.12 (2008): 1399-1401.
74 Raychaudhuri, Soumya, et al. "Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk." *Nature genetics* 41.12 (2009): 1313-1318.
75 Hinks, Anne, et al. "Overlap of disease susceptibility loci for rheumatoid arthritis and juvenile idiopathic arthritis." *Annals of the rheumatic diseases* 69.6 (2010): 1049-1053.
76 Adamovic, Svetlana, et al. "Association study of IL2/IL21 and FcgRIIa: significant association with the IL2/IL21 region in Scandinavian coeliac disease families." *Genes & Immunity* 9.4 (2008): 364-367.
77 Maiti, Amit K., et al. "Confirmation of an association between rs6822844 at the Il2–Il21 region and multiple autoimmune diseases: evidence of a general susceptibility locus." *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 62.2 (2010): 323-329.
78 Daha, Nina A., et al. "Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis." *Arthritis & Rheumatism* 60.5 (2009): 1255-1260.
79 Hinks, A., et al. "Association of the AFF3 gene and IL2/IL21 gene region with juvenile idiopathic arthritis." *Genes & Immunity* 11.2 (2010): 194-198.
80 Demela, Pietro, Nicola Pirastu, and Blagoje Soskic. "Cross-disorder genetic analysis of immune diseases reveals distinct gene associations that converge on common pathways." *Nature Communications* 14.1 (2023): 2743.
81 Conrad, Nathalie, et al. "Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK." *The Lancet*401.10391 (2023): 1878-1890.

82 Warram, James H., et al. "Differences in risk of insulin-dependent diabetes in offspring of diabetic mothers and diabetic fathers." *New England Journal of Medicine* 311.3 (1984): 149-152.
83 Yang, Siyoung, et al. "Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance." *Science*348.6234 (2015): 589-594.
84 Roth-Schulze, Alexandra J., et al. "Type 1 diabetes in pregnancy is associated with distinct changes in the composition and function of the gut microbiome." *Microbiome*9 (2021): 1-21.
85 Gao, Yuan, et al. "Maternal gut microbiota during pregnancy and the composition of immune cells in infancy." *Frontiers in Immunology* 13 (2022): 986340.
86 Liu, Aoxing, et al. "Evidence from Finland and Sweden on the relationship between early-life diseases and lifetime childlessness in men and women." *Nature Human Behaviour*(2023): 1-12.
87 Miodovnik, Menachem, et al. "Spontaneous abortion among insulin-dependent diabetic women." *American journal of obstetrics and gynecology* 150.4 (1984): 372-376.
88 Dorman, Janice S., et al. "Temporal trends in spontaneous abortion associated with type 1 diabetes." *Diabetes research and clinical practice* 43.1 (1999): 41-47.
89 Lorenzen, T., et al. "A population-based survey of frequencies of self-reported spontaneous and induced abortions in Danish women with Type 1 diabetes mellitus." *Diabetic medicine* 16.6 (1999): 472-476.
90 Penney, Gillian C., et al. "Outcomes of pregnancies in women with type 1 diabetes in Scotland: a national population-based study." *BJOG: an International Journal of Obstetrics & Gynaecology* 110.3 (2003): 315-318.

Acknowledgments

We thank the entire FinRegistry and FinnGen team for making the data available for this study. We also acknowledge Richard Oram (Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK) for his discussion on this study, Samuel Jones (Institute for Molecular Medicine Finland (FIMM), Finland) for discussion on HLA PGS analysis, Bradley Jermy (FIMM, Finland) for discussion on genetic correlation analysis, Mary P Reeve (FIMM, Finland) for discussion on disease definitions, Om Dwivedi (FIMM, Finland) for discussion on T1D PGS, and Alessio Gerussi (University of Milano-Bicocca, Italy) for discussion on results from a clinical standpoint. The We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis Pharma AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB.

Author contributions

FW, AL, TT, and AG designed the study and wrote the manuscript, with comments from all authors. FW processed the registry data, conducted the statistical analyses, and generated all figures and tables; AL and FW preprocessed the FinRegistry pedigree data and the FinRegistry team preprocessed all other registry data of FinRegistry; the FinnGen data team preprocessed the FinnGen data. TT, PV, and the FinnGen clinical team defined disease endpoints; TT

oversaw the project and interpreted the findings from a clinical standpoint. AL, ZY, SJ, and AG advised the registry-based epidemiological analyses; AL, ZY, and AG advised genetic analyses; AL, SK, JR, JP, and TT advised HLA haplotype analyses. AG and TT supervised the study. MP is the principal investigator of the FinRegistry project. All authors discussed the results, revised the manuscript, and had final responsibility for the decision to submit for publication.

Funding

FW was funded by the University of Helsinki and the University of Edinburgh joint PhD program in Human Genomics. AG was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme [grant number 945733], starting grant AI-Prevent. AG was supported by the Academy of Finland (grant no. 323116). This project also received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 101016775. TT was supported by grants from Folkhälsan Research Foundation, The Academy of Finland (336822, 312072 and 336826), and the University of Helsinki.

Competing interests

AG is the founder of Real World Genetics.

RO holds a UK Medical Research Council Institutional Confidence in Concept Grant to develop a 10 SNP biochip T1D genetic test in collaboration with Randox.

Ethics statement

FinRegistry is a collaboration project of the Finnish Institute for Health and Welfare (THL) and the Data Science Genetic Epidemiology research group at the Institute for Molecular Medicine Finland (FIMM), University of Helsinki. The FinRegistry project has received approvals for data access from THL (THL/1776/6.02.00/2019 and subsequent amendments), Digital and Population Data Services Agency (VRK/5722/2019–2), Finnish Center for Pension (ETK/SUTI 22003) and Statistics Finland (TK-53–1451-19). The FinRegistry project has received IRB approval from THL (Kokous 7/2019).

The use of the biobank donor samples is in accordance with the biobank consent and meets the requirements of the Finnish Biobank Act 688/2012. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier

TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 11 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, BB2021_65, Finnish Red Cross Blood Service Biobank 7.12.2017 (000-2018), Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/430/2021 §28, §29, HUS/150/2022 §12, §13, §14, §15, §16, §17, §18, §23, §58 and §59, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and amendments BB_2021-0140, BB_2021-0156 (August 26 2021, Feb 2 2022), BB_2021-0169, BB_2021-0179, BB_2021-0161, AB20-5926 and amendment #1 (April 23 2020) and it´s modification (Sep 22 2021), BB_2022-0262, BB_2022-0256, Biobank Borealis of Northern Finland_2017_1013, 2021_5010, 2021_5018, 2021_5015, 2021_5015 Amendment, 2021_5023, 2021_5023 Amendment, 2021_5017, 2022_6001, 2022_6006 Amendment, BB22-0067, 2022_0262, Biobank of Eastern Finland 1186/2018 and amendment 22§/2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, 27§/2022, 28§/2022, 29§/2022, 33§/2022, 35§/2022, 36§/2022, 37§/2022, 39§/2022, 7§/2023, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), 8§/2021, 9§/2021, §9/2022, §10/2022, §12/2022, 13§/2022, §20/2022, §21/2022, §22/2022, §23/2022, 28§/2022, 29§/2022, 30§/2022, 31§/2022, 32§/2022, 38§/2022, 40§/2022, 42§/2022, 1§/2023, Central Finland Biobank 1-2017, BB_2021-0161, BB_2021-0169, BB_2021-0179, BB_2021-0170, BB_2022-0256, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844: ARC_2021_1001.

Methods

Data sources

Most of our analyses were conducted by leveraging two large datasets in Finland – FinRegistry[33] and FinnGen[34] (**Table S3.1.1**).

FinRegistry is a national registry database in Finland combining disease registries with comprehensive records on individuals' demographics, socioeconomic status, death, drug purchases, prescriptions, and administrative information. It includes all the residents in Finland who were alive on 1st January 2010, and their first-degree relatives (in total approximately 7.2 million individuals). 572,640 individuals were excluded as they migrated in or out of Finland by 2019, remeining 3,412,326 individuals with both father and mother information.

FinnGen (R11) is a national research project using samples and data from Finnish biobanks, compromising genome-wide genotype data and healthcare registry data for approximately 0.47 million Finnish biobank donors. After excluding 9,196 individuals who had migrated in or out of Finland by 2019, removing 10,588 individuals without imputed HLA data and 8,627 individuals who have other missing values in the cohort, we included 3,668 T1D cases and 441,602 controls without T1D included in FinnGen and observed lifetime prevalence and number of cases in FinnGen for each selected AID. The study population features slightly more women (N=251,294, 56.4%).

By using the datasets from the same country, we alleviate concerns about potential bias caused by differences in genetic background, healthcare system, or socio-cultural context.

Finnish population-based case-control study

The epidemiology association between AIDs in parents and T1D in their children was investigated by a case-control study within FinRegistry. To include most of the individuals and their parents in this analysis, we selected all individuals born in Finland between 1960 and 1999 who had both father's and mother's information available. The follow-up period started either from the date they were born or from the year the codes for our selected diseases were included in national patient register and ended on 31 December 2019, which is the latest follow-up date in FinRegistry. We then restricted our analysis to offspring whose father was born between 1917 and 1976 and whose mother was born between 1922 and 1976. Finally, we removed those individuals who died during the follow-up period.

T1D cases were individuals diagnosed with T1D before the age of 40 years. For each trio with T1D-affected children, three controls without T1D diagnosis were matched by following variables: sex, birth year (five-year bin, 8 levels), and birthplace for the child (19 levels), birth year of father (five-year bin, 12 levels), birth year of mother (five-year bin, 11 levels), and family size (represented by the number of siblings: 0, 1, 2, 3, >=4). The differences in socio-demographic characteristics between cases and controls, including socio-economic status, marital status, number of offspring, occupation, education, and mother tongue, were further examined to ensure that the two groups were comparable. Cohen's D[91] ($\geq$0.2) for ordinal characteristics such as number of siblings and number of children and Chi-squared test ($P<0.05$) for categorical variables such as education and marital status were used to test whether the differences are minor.

Exposures of interest were all the parental endpoints related to AIDs mainly defined by Finnish versions 8-10 of International Classification of Diseases (ICD) (**Supplementary notes 2.1**). AIDs with fewer than 50 cases among parents in the final study population in the case-control study were filtered out from the list. We then excluded AIDs with an unclear definition or AID codes that were subtypes of other codes. The final AID list was used across all analyses in this study.

A conditional logistic regression, adjusted by birth year, birth year of father, birth year of mother, and the exact number of siblings, was applied to estimate the effect of having a given AID in parents on the risk of developing T1D in offspring. Bonferroni multiple testing correction was applied, and only the diseases reached significant associations ($P<=0.05/26$) were included in the following analyses. To examine whether the identified associations are shared between father and mother, we used conditional logistic regression stratified by maternal AIDs status and paternal AIDs status. We also investigated whether parental AID status can influence the onset age of disease in offspring by regressing the age at onset of T1D in offspring on the AID status in their parents, with the same covariates described above adjusted. Statsmodels v.0.13.5[92] and SciPy v.1.7.3 [93] were used for the analyses in this section in the context of Python v.3.7.11.

Shared genetic components at a population level

To explore, to what extent, the familial aggregations between AIDs and T1D identified in the population-based case-control study are contributed by genetic factors, we conducted three analyses in both HLA regions and non-HLA regions at individual level. Given the strong associations between HLA and AIDs[4], we first performed an HLA haplotype analysis and an HLA-based PGS analysis using imputed HLA alleles. Then, we performed a genetic correlation analysis using variants in non-HLA regions.

HLA haplotype analysis

The HLA alleles were imputed by a population-specific reference panel using FinnGen data at the individual level[37]. We considered all the 187 unique available alleles on 10 classical genes in class I and II regions for analysis, including 27 alleles for *HLA-A*, 40 for *HLA-B*, 23 for *HLA-C*, 24 for *HLA-DPB1*, 15

671 for *HLA-DQA1*, 16 for *HLA-DQB1*, 33 for *HLA-DRB1*, and 3 each for *HLA-DRB3*, *HLA-DRB4*, and *DRB5*
672 separately. According to previous studies[8-10,36], several HLA haplotypes in Class II have been significantly
673 associated with an increased risk of developing T1D and other AIDs. To understand whether HLA genetic
674 liability to T1D is associated with other AIDs, we conducted a haplotype analysis (**Figure S2.4.1**) and only
675 considered possible haplotypes from gene combination *HLA-DRB1-DQA1-DQB1*. We first removed
676 individuals with any ambiguous genotypes in the given haplotype. We define an ambiguous genotype as a
677 genotype with imputed allele dosage in [0.4, 0.6] or [1.4,1.6]. Next, we extracted the potential haplotypes
678 using the expectation–maximization algorithm and removed those haplotypes whose frequency is less than
679 0.005. We identified the most prevalent HLA haplotypes that were either positively or negatively associated
680 with T1D using a haplotype score test ($P<1.0 \times 10^{-10}$), a statistical test to calculate the contribution of a
681 specific haplotype on a certain phenotype when the linkage phase is unknown[95]. Multiple testing corrections
682 were applied for haplotype selection. We then only selected the three haplotypes with the highest P-value
683 and the three with the lowest. Afterward, a multivariate logistic regression was built to estimate the
684 associations between the remaining haplotypes and AID status, adjusting for age, sex and the first ten
685 principal components (PCs). Haplotype analysis was done in Python v.3.7.13 using statsmodels v.0.13.2[92]
686 and in R v.4.2.2 using gap v.1.3-1[96].

687

688 *HLA multi-allelic score*

689 To tackle the complex linkage disequilibrium (LD) structure amongst HLA alleles, we introduced a PGS
690 method to construct PGS for T1D and other AIDs specific to HLA regions using FinnGen R11. To handle
691 the downstream analyses in later Panels, we considered all individuals of genotyped family trios as our
692 target population and the rest as a training set for weight calculation.

693 In the training set, to account for the complex LD and high polymorphism among HLA alleles, we applied
694 to calculate the weight for each HLA allele using weighted ridge classifiers with optimal regularization
695 strength, which can address this problem by imposing a penalty on the size of the coefficients. For
696 covariates, we adjusted for age, sex, and the first ten PCs. For the genotyped trios, we calculated PGS by
697 summing up all the allele dosage multiplied by its weight from the corresponding training set. After PGS
698 normalization, we first measured the partial correlation between each AID and its PGS. We dropped all the
699 AIDs whose partial correlation with PGS for the AIDs themselves was lower than 2%, to filter out diseases
700 with low sample sizes or with weak HLA signals. Next, we conducted logistic linear regression to examine
701 the association between PGS and T1D. The Bonferroni correction was applied to account for false positives
702 related to multiple comparisons. Weighted ridge regression and logistic regression were applied using
703 sklearn v.1.0.2 and statsmodels v.0.13.2[92] in Python v.3.7.13 and partial correlation were calculated using
704 ppcor v1.1[97] in R v.4.3.2.

705

706 *non-HLA genetic correlation*

707 To evaluate the genetic overlap between T1D and other AIDs due to shared non-HLA effects, we utilized
708 largeset GWAS summary statistics from European ancestry individuals (**Table S2.3.1**) and estimated
709 genetic correlation in non-HLA regions using linkage disequilibrium score regressions (LDSC), which
710 capitalized the patterns of LD among common genetic variants of the whole genome except for the HLA
711 regions[43,44]. To separate the contribution of HLA from non-HLA genetic signals, we excluded single-
712 nucleotide polymorphisms (SNPs) in the extended HLA regions on chromosome 6 (25–34 Mb). In a
713 sensitivity analysis, we further examined whether the genetic correlations changed significantly while
714 excluding all the SNPs on chromosome 6. We followed the suggested protocol for LDSC analyses from

715 https://github.com/bulik/ldsc and used the package LDSC v.1.0.1[43] in Python v.3.7.11. We considered
716 HapMap3[98] SNPs and constructed LD structures with European ancestry samples in 1000 Genomes
717 project[99].

718

719 **Analyses of inter-generation cross-trait transmission via polygenic transmission disequilibrium test**

720 To understand the transmission of parental AIDs to T1D in offspring, we further estimated the contribution
721 of parental transmitted and non-transmitted genetic liability to T1D in offspring in both non-HLA regions
722 and HLA regions using data in FinnGen. While offspring are expected to receive, on average, half of their
723 parents' risk alleles for a disease, some will over or under-inherit alleles associated with the disease, which
724 will impact their disease risk. To understand how much of the shared genetic liability is on average
725 transmitted to the next generation, we applied a polygenic transmission disequilibrium test (pTDT) to assess
726 whether the mean of the offspring PGS distribution is consistent with its parentally derived expected value[42].
727
728 To construct non-HLA PGS, we used the same GWAS summary statistics used in genetic correlation
729 analysis and adopted PRS-CS[46] (**Table S2.3.2**). We included 12,563 genotyped trios where both parents
730 and at least one offspring were directly genotyped. We considered HLA PGSs and non-HLA PGSs that
731 were robustly associated with the corresponding AID ($P<0.05/19$ and $|\rho|>=2\%$) (**Table S2.3.2**). This
732 resulted in ten diseases with both robust HLA PGS and non-HLA PGS: T1D, HYPO, CD, RA, SLE,
733 sarcoidosis, psoriasis, IBD, MS, and AS, as well as additional six diseases for HLA PGS (B12A, HYPER,
734 PBC, MCTD, SjS, and ADDISON) (**Supplementary notes 3.3.2** and **Table S3.3.4**).

735 For each remaining parental AD, we first removed genotyped families with affected parents and then
736 subtracted mid-parent PGS for that AID from offspring PGS for the same disease. We normalized the
737 obtained difference by dividing it by the standard deviation of the mid-parent distribution. For each child,
738 we grouped his or her family trios according to whether this child was affected, so that we could compare
739 how the mean PGS of children deviate from their mid-parent PGS in the affected sibling group and
740 unaffected sibling group separately, as well as the difference between the two groups. We conducted pTDT
741 using SciPy v.1.7.3[93] in Python v.3.7.13.

742
743
744 **Predicting T1D risk in offspring using parental PGS**

745 To assess how well parental PGSs could be linked to T1D in offspring, we designed several analyses to test
746 the associations and predicting abilities. Considering the prediction performance of PGS for parental AIDs,
747 we only focused on those AIDs that showed significance from previous PGS analyses in HLA regions and
748 non-HLA regions. We introduced a new approach, Full-PGS, combining separately generated HLA and
749 non-HLA PGSs.

750

751 *Full-PGS construction*

752 Theoretically, with pTDT, four patterns could be defined for parental AID - offspring T1D transmissions:
753 1) both HLA PGS and non-HLA PGS over-/under-transmitted from unaffected parents to affected children;
754 2) only HLA PGS over-/under-transmitted; 3) only non-HLA PGS over-/under-transmitted; 4) neither was
755 transmitted.
756

We wondered whether we could improve PGS prediction across diseases by better considering genetic architecture of a target disease. For our target disease, T1D, we constructed HLA PGS and non-HLA PGS separately, and further considered the contribution ratio of HLA PGS relative to non-HLA PGS in a full PGS context.

We proposed an equation to quantify the exact contribution of HLA PGS and non-HLA PGS ($\boldsymbol{ratio_{new}}$) for diseases following pattern 1 ($effect_{HLA} \neq 0$; $effect_{non-HLA} \neq 0$), by taking HLA vs non-HLA contributions for both T1D and another AID into account.

$$\boldsymbol{ratio_{new}} = \frac{2 \times w_{HLA}}{2 \times w_{HLA} + 1 \times w_{non-HLA}} : \frac{1 \times w_{non-HLA}}{2 \times w_{HLA} + 1 \times w_{non-HLA}}$$

Where the contribution ratio between HLA PGS and non-HLA PGS for T1D is 2:1, w_{HLA} and $w_{non-HLA}$ are the weight of HLA PGS and the weight of non-HLA PGS for the AID when predicting the AID itself, $\boldsymbol{ratio_{new}}$ denotes the contribution ratio between HLA PGS and non-HLA PGS when predicting T1D.

For the diseases following pattern 2 ($effect_{HLA} \neq 0$; $effect_{non-HLA} = 0$), since only HLA PGS is assumed to be over-/under-transmitted, we only used HLA PGS for these diseases to predict T1D.

$$\boldsymbol{ratio_{new}} = 1:0$$

For the diseases following pattern 3 ($effect_{HLA} = 0$; $effect_{non-HLA} \neq 0$), similarly, we only considered non-HLA PGS for these diseases to predict T1D.

$$\boldsymbol{ratio_{new}} = 0:1$$

For the diseases following pattern 4 ($effect_{HLA} \neq 0$; $effect_{non-HLA} \neq 0$), neither HLA PGS nor non-HLA PGS was used for T1D prediction.

$$\boldsymbol{ratio_{new}} = 0:0$$

In a scenario of cross-disease prediction, e.g., using PGS for RA to predict T1D, we conducted the analyses with the following steps: 1) assigned RA to one of the four groups according to its pTDT results, 2) quantified the contribution ratio between HLA PGS and non-HLA PGS using $\rho_{\text{HLA PGS}}$, $\rho_{\text{non-HLA PGS}}$ and the proposed equation, and 3) built a Full-PGS using HLA PGS, non-HLA PGS and the ratio of their contributions.

We proposed that this equation could be extended to broader conditions for cross-disease prediction to fully utilize PGS information of a correlated disease to boost the prediction power of a target disease.

$$\boldsymbol{ratio_{new}} = \begin{cases} 0:0, & effect_1 = 0;\ effect_2 = 0 \\ 1:0, & effect_1 \neq 0;\ effect_2 = 0 \\ 0:1, & effect_1 = 0;\ effect_2 \neq 0 \\ \dfrac{w_{T1} \times w_{E1}}{w_{T1} \times w_{E1} + w_{T2} \times w_{E2}} : \dfrac{w_{T2} \times w_{E2}}{w_{T1} \times w_{E1} + w_{T2} \times w_{E2}}, & effect_1 \neq 0; effect_2 \neq 0 \end{cases}$$

where w_{T1} and w_{T2} are the weights of two seperate components of the target disease under a presumed genetic architecture (e.g., partition into HLA variants vs non-HLA variants, or common variants vs rare variants), w_{E1} and w_{E2} are the weights of two separate components of the correlated disease, $\boldsymbol{ratio_{new}}$ is

the contribution ratio between the two genetic components for the correlated disease when predicting the target disease.

Impact of parental AID information and cumulative risk of T1D in children

For each parental AID, we constructed a mid-parent Full-PGS using HLA PGS and non-HLA PGS considering their ratio for cross-disease prediction and examined the association with T1D in offspring. For comparisons, we ran additional models including only using parental diagnosis, only mid-parent PGS calculated by a traditional PGS approach that takes genome-wide variants altogether for calculation, e.g., PRS-CS[46], or both HLA PGS and non-HLA PGS from our previous analyses, assumed to be at least as good as Full-PGS. We used AUC as a primary metric to evaluate model performances. To understand how much additive value parental PGS for other AIDs can be added to a T1D prediction model, we then compare a model with all mid-parent PGSs and one with only mid-parent T1D PGS. We evaluated the two models by comparing the AUCs among 12,563 genotyped trios in FinnGen. To avoid overfitting, the analyses were done using five-fold cross validation.

We then conducted a Cox proportional hazards model to assess the T1D cumulative incidence risk in offspring before the age of 20 years, adjusting for birthyear and the first 10 PCs of offspring. To maximize the statistical power, we considered all the genotyped trios with children born between 1960-2010. Among the 8,827 trios we analyzed, 1,035 had a child with T1D (11.7%). We then grouped the trios by dividing the parental Full-PGS percentile into 0-50th, 50-90th, 90-99th, and 99-100th and estimated cumulative risk within each group given the mean of children's birthyear and PCs. We also stratified the model by the sex of offspring. To make the result generalizable, we calibrated the FinnGen results based on the FinRegistry nationwide T1D prevalence among all children born between 1960-2010 (mean birthyear=1984; N=3,048,812; prevalence=0.6%).

For the analyses in FinRegistry, we used the same cohort as we used for FinnGen calibration. We considered four groups: 1) parents without T1D, 2) maternal T1D only, 3) paternal T1D only, and 4) both parents with T1D among the whole study population. For the analyses excluding children with parental T1D, we considered four groups: 1) parents without AID, 2) only the mother had AID(s), 3) only the father had AID(s), and 4) both parents had AID(s). The parental AIDs in this analysis covered all the significant associations identified from the previous analyses. All the analyses in this session were done using statsmodels v.0.13.2[92] and sklearn v.1.0.2 in Python v.3.7.13 and survival v.3.5.7 in R v.4.3.2.

Code availability

Code for the complete analyses is available at https://github.com/dsgelab/parentalAIDs_T1D.

Reference

91 Cohen, Jacob. Statistical power analysis for the behavioral sciences. Academic press, 2013.

92 Seabold, Skipper, and Josef Perktold. "Statsmodels: econometric and statistical modeling with python." *SciPy* 7 (2010): 1.

93 Virtanen, Pauli, et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature methods* 17.3 (2020): 261-272.

94 Douek, I. F., et al. "Three generations of autoimmune diabetes: an extended family study." Diabetologia 46 (2003): 1313-1318.
95 Schaid, Daniel J., et al. "Score tests for association between traits and haplotypes when linkage phase is ambiguous." *The American Journal of Human Genetics* 70.2 (2002): 425-434.
96 Zhao, Jing Hua. "gap: Genetic analysis package." *Journal of Statistical Software* 23 (2008): 1-18.
97 Kim, Seongho. "ppcor: an R package for a fast calculation to semi-partial correlation coefficients." *Communications for statistical applications and methods* 22.6 (2015): 665.
98 International HapMap 3 Consortium. "Integrating common and rare genetic variation in diverse human populations." *Nature* 467.7311 (2010): 52.
99 1000 Genomes Project Consortium. "A global reference for human genetic variation." *Nature* 526.7571 (2015): 68.

835