JIM Original Article

doi: 10.1111/joim.13187

Genome-wide search for genes affecting the age at diagnosis of type 1 diabetes

A. Syreenj[1,2,3], N. Sandholm[1,2,3], C. Sidore[4], F. Cucca[4,5], J. Haukka[1,2,3], V. Harjutsalo[1,2,3,6], P.-H. Groop[1,2,3,7] & the FinnDiane Study Group

From the [1]Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki; [2]Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki; [3]Abdominal Center, Nephrology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; [4]Instituto di Ricerca Genetica e Biomedica, CNR, Monserrato; [5]Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy; [6]National Institute for Health and Welfare, Helsinki, Finland; and [7]Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia

Abstract. Syreeni A, Sandholm N, Sidore C, Cucca F, Haukka J, Harjutsalo V, Groop P-H; the FinnDiane Study Group (University of Helsinki; Folkhälsan Institute of Genetics, Folkhälsan Research Center; University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; Instituto di Ricerca Genetica e Biomedica, CNR, Monserrato; Università degli Studi di Sassari, Sassari, Italy; National Institute for Health and Welfare, Helsinki, Finland; Monash University, Melbourne, Victoria, Australia). Genome-wide search for genes affecting the age at diagnosis of type 1 diabetes. *J Intern Med* 2021; **289**: 662–674.

Background. Type 1 diabetes (T1D) is an autoimmune disease affecting individuals in the early years of life. Although previous studies have identified genetic loci influencing T1D diagnosis age, these studies did not investigate the genome with high resolution.

Objective and methods. We performed a genome-wide meta-analysis for age at diagnosis with cohorts from Finland (Finnish Diabetic Nephropathy Study), the United Kingdom (UK Genetic Resource Investigating Diabetes) and Sardinia. Through SNP associations, transcriptome-wide association analysis linked T1D diagnosis age and gene expression.

Results. We identified two chromosomal regions associated with T1D diagnosis age: multiple independent variants in the HLA region on chromosome 6 and a locus on chromosome 17q12. We performed gene-level association tests with transcriptome prediction models from two whole blood datasets, lymphocyte cell line, spleen, pancreas and small intestine tissues. Of the non-HLA genes, lower *PNMT* expression in whole blood, and higher *IKZF3* and *ZPBP2*, and lower *ORMDL3* and *GSDMB* transcription levels in multiple tissues were associated with lower T1D diagnosis age (FDR = 0.05). These genes lie on chr17q12 which is associated with T1D, other autoimmune diseases, and childhood asthma. Additionally, higher expression of *PHF20L1*, a gene not previously implicated in T1D, was associated with lower diagnosis age in lymphocytes, pancreas, and spleen. Altogether, the non-HLA associations were enriched in open chromatin in various blood cells, blood vessel tissues and foetal thymus tissue.

Conclusion. Multiple genes on chr17q12 and *PHF20L1* on chr8 were associated with T1D diagnosis age and only further studies may elucidate the role of these genes for immunity and T1D onset.

Keywords: age of onset, genome-wide association study, transcriptome-wide association analysis, type 1 diabetes.

Introduction

Over 60 loci in the genome contribute to genetic predisposition to type 1 diabetes (T1D) [1-5] in which insulin deficiency results from an autoimmune attack against insulin-producing beta cells of the pancreatic islets. Heterogeneity in the disease aetiology is recently acknowledged and immunological processes leading to T1D in individuals diagnosed later in life appear different from the processes in individuals having disease onset in early childhood, in which B cells are involved in the pathological process in the pancreas [5]. Different genes and genetic variants may thus affect disease course at varying ages, also suggested by the high diagnosis age correlation (r^2 = 0.95) in Finnish monozygotic twins concordant for T1D [6]. Of the known T1D risk loci,

however, only the HLA locus and a few non-HLA loci, have been associated with age at diagnosis [7-10]. Genetic risk score combines risk-increasing alleles into a single score and the genetic risk score for T1D has already been suggested for clinical use for screening of infants at highest T1D risk [11]. All disease-susceptibility variants are included in the score, but only a few known T1D variants have stronger effects in individuals with early-onset disease [10].

Whilst many T1D-associated SNPs show enrichment in lymphoid cell enhancers [4] or regulate tissue-specific gene expression [12], the true biological effect of many T1D susceptibility loci remains unknown. Nevertheless, enrichment analyses examining whether risk variants preferentially lie in regulatory regions can highlight tissues or cell types that are likely implicated in disease aetiology. Transcriptome-wide association analysis is another method to transfer information from genetic variation to biological effects and gene expression in multiple tissues.

The latest and largest meta-analyses for T1D [4] and T1D diagnosis age [9] have been performed with variants from the ImmunoChip, a large scale but targeted genotyping platform which covers only loci previously associated with immunological diseases. We now took a genome-wide approach by performing a large genome-wide association study (GWAS) meta-analysis in 12,539 individuals with T1D from the Finnish Diabetic Nephropathy (FinnDiane) Study, the UK Genetic Resource Investigating Diabetes (UK GRID), and Sardinia cohorts. Our aim was to identify variants affecting T1D diagnosis age and thereafter, utilizing the genome-wide coverage of our analysis, we aimed to link the variants to open chromatin indicating active gene expression in different cell types and finally, we performed transcriptome-wide association analyses in disease-relevant tissues.

Methods

Genotyping and imputation

Local Ethics committees approved studies in cohorts from Finland, U.K. and Italy participating in this study. In the FinnDiane, we included 5162 individuals with diabetes onset < 40 years and insulin treatment started within one year from diagnosis but also accepted individuals with T1D diagnosis < 18 but missing data on insulin treatment. These individuals had earlier passed quality control after genotyping and imputation performed as previously described [13] and the details provided in Table S1. The UK GRID collection was divided into two strata according to the genotyping platform: 1925 individuals were genotyped using the Affymetrix GeneChip Mapping 500 K and for 3976 using the Illumina 550 K Infinium microarray platform (Illumina, San Diego, CA). Genotyping had been performed as part of earlier genetic studies [1, 3]. Sardinian individuals were recruited from a larger autoimmunity cohort studying multiple sclerosis and T1D, involving 8546 individuals genotyped at 883 557 SNPs with either Affymetrix 500 K (5673 individuals) and Illumina Omni Express (3068 individuals). Amongst these, 1558 individuals had T1D of which 1476 were eligible for this study. Imputation method was Minimac3 with a custom reference panel of Sardinian individuals as in Sidore *et al.* [14].

GWAS meta-analysis

GWAS was performed in each cohort using natural logarithm of age at diagnosis (0.5–40 years, Fig. S1) as the phenotype. Sex and cohort specific covariates (genotyping batch and population stratification variables) were included in a linear regression frequentist-test with SNPTEST v.2 [15], (UK GRID and Sardinia) or score-test with RVTESTS software [16], (FinnDiane). We selected variants with imputation quality score info ≥0.6 and minor allele frequency (MAF) ≥1% for meta-analysis. We combined effect sizes (Beta) and standard errors (SEs) with fixed effect meta-analysis using METAL software [17]. Final results comprised 7 374 092 variants with association results in ≥2 cohorts.

Secondary analyses of the top variants

HLA imputation

We analysed the meta-analysis top variants in HLA–risk groups in the FinnDiane cohort. We successfully imputed the HLA haplotypes for 5152 out of 5162 individuals in our GWAS cohort with the SNP2HLA software [18] with default parameters and Type 1 Diabetes Genetic Consortium reference panel consisting of 3924 SNPs and 4-digit HLA-genotypes. We used previously genotyped HLA-alleles (4- or 2-digit accuracy) for 4279 FinnDiane study participants [19] for imputation quality validation. The imputed alleles for *HLA-DQA1, HLA-DQB1*, and *HLA-DRB1* matched the genotyped alleles with ≥97.2% accuracy. Table S2 lists all HLA II-locus haplotypes we used to define

HLA–risk groups, but they were in brief (i) T1D high-risk: DR4-DQ8/DR3-DQ2 heterozygotes (n = 1300), (ii) medium-risk: one or two copies of DR4-DQ8 or DR3-DQ2 (n = 3167), and (iii) low-risk: protective HLA II haplotypes or other DR4 and DR3 haplotypes (n = 681).

Stratifying according to age at diagnosis

Wide T1D diagnosis age distribution in the Finn-Diane (0.5–39.9) enabled us to divide the cohort into individuals diagnosed before (n = 3157) or after (n = 2005) the age of 16. We tested the association between meta-analysis top variants and age at diagnosis separately in these two groups.

Enrichment analysis

GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction (GARFIELD) v.2 software package provided annotations, linkage disequilibrium data, location of SNPs relative to transcription start sites (TSS), and scripts for enrichment analysis [20]. Annotation data originated from Encyclopedia of DNA elements (ENCODE, [21]) and NIH Roadmap Epigenomics Consortium [22] and consisted of 424 DNase I hypersensitive-site (DHS) annotations and 41 footprint-annotations from the same tissues, both indicating open and active DNA in multiple tissues, cells and cell lines.

On average, each tissue or cell data included 31 505 genomic DHS loci and 6117 footprint-annotations. Calculated number of effective annotations (223) set the significance threshold to $P = 2.25 \times 10^{-4}$. We performed enrichment analysis of GWAS meta-analysis loci with $P \leq 10^{-5}$ and $P \leq 10^{-4}$ with similar settings used in the original paper by Iotchkova *et al.* [20]: number of bins was 15 for LD and 5 for TSS. Variants were pruned to independent variants ($r^2 < 0.1$) based on UK10K collection with 3621 samples from TwinsUK and ALSPAC population cohorts. The second step was annotation of all independent variants, as well as all variants having $r^2 \geq 0.8$ with the pruned SNPs. Finally, for all different annotations, enrichment of age at diagnosis-associated variants was calculated.

Transcriptome-wide association analysis

We implemented transcriptome-wide association analysis with MetaXcan that uses S-PrediXcan, extension of PrediXcan, and it only requires GWAS SNP-level summary statistics [23, 24]. We downloaded pre-calculated transcriptome prediction models (MASHR-based) and SNP covariance tables from PredictDB (http://predictdb.org/). We selected five relevant tissues and cells (GTEx v.8) for the primary analysis: whole blood, Epstein-Barr virus (EBV)-transformed lymphocytes, and spleen for their relevance in immunity, pancreas as a target organ and small intestine due to the potential role of gut immune system in the pathogenesis of T1D. We harmonized and imputed our SNPs to GTEx variants as described in ref [25]. Significant results had $P \leq 1.3 \times 10^{-5}$ and false discovery rate (FDR) of 0.05 calculated with Benjamini–Hochberg procedure with combined results of the five tissues (a total of 63 947 genes analysed).

We also checked gene-level association results for selected age at diagnosis GWAS meta-analysis loci genes from sigmoid colon, adrenal gland, thyroid, liver and lung tissues. In addition, we selected another whole blood transcriptome prediction model from 922 individuals from Depression Genes and Networks (DGN) study [26] to compare the whole blood transcriptome prediction models. Pividori and colleagues [27] ran MetaXcan (S-PrediXcan) for 4091 traits of which 4049 were from UKBB, with transcription prediction models generated with GTEx v.8 data. We searched transcriptome-wide association results of two diabetes-related UKBB phenotypes: 'Age diabetes diagnosed' and 'E10:Insulin-dependent diabetes mellitus' from http://apps.hakyimlab.org/phenomexcan/. We queried all our significantly associated genes (at FDR = 0.05) from the results generated with UKBB data. In addition, we compared three top genes per tissue found in UKBB Age diabetes diagnosed –phenotype with our results from T1D diagnosis age.

eQTLs and trait-associations

We queried from GTEx v.8 (www.gtexportal.org) and Database of Immune Cell EQTL/Expression/Epigenomics (www.dice-database.org) whether the index SNPs associated with T1D diagnosis age were estimated quantitative trait loci (eQTLs) i.e. associated with gene transcription. We also examined phenotypic associations of chromosome 17 locus index SNP from the UKBB round 2 GWAS summary statistics (released August 2018) from Open Targets (https://genetics.opentargets.org/, accessed October 2019) that also presents publicly available

GWAS results including T1D ImmunoChip meta-analysis [4].

Chromosome-conformation capture

We used Capture HiC Plotter [28] at https://www.chicp.org/chicp/ to look for chr17q12 region genomic interactions. We uploaded our meta-analysis results with $P < 1 \times 10^{-4}$ to be integrated in the plots.

Data resource and availability

The datasets analysed during the current study are not publicly available because we do not have the permission to release patient genotypic data. The scripts used to perform meta-analysis and subsequent analyses are available from corresponding author upon reasonable request.

Results

GWAS meta-analysis

A total of 275 variants were associated with age at diagnosis with genome-wide significance ($P < 5 \times 10^{-8}$, Fig. 1). Most associations were within the HLA region where altogether 21 independent ($r^2 < 0.1$ in 1000G EUR) significant or suggestive ($P < 1 \times 10^{-5}$) age at diagnosis-associations emerged in HLA I, HLA II and HLA III regions, of which eleven were significant at FDR = 0.01 and seven at genome-wide level (Table S3). The strongest association was observed for rs116763857 (G > T, GRCh37:chr6:31,141,482) with major G-allele associated with lower diagnosis age (Beta = −0.295, SE = 0.006, $P = 1.2 \times 10^{-14}$, Fig. S2). Due to the MAF < 0.01 in the Sardinian cohort, the meta-analysis for this SNP included only results from UK GRID (Illumina: MAF = 0.032, Affymetrix: MAF = 0.034) and FinnDiane (MAF = 0.020).

Outside the HLA region, altogether 18 independent variants ($r^2 < 0.1$ in 1000G EUR) showed suggestive association ($P < 5 \times 10^{-5}$) with T1D diagnosis age, of which 12 were significant at FDR = 0.05 (Table 1). Whilst 17 of the suggestive associations were at genomic loci not previously associated with T1D, the only genome-wide significant association outside HLA was at known T1D susceptibility locus chr17q12, where the major C-allele of the locus index SNP rs2941522 (C > T) was associated with lower T1D diagnosis age (Beta = −0.069, SE = 0.012, $P = 7.3 \times 10^{-9}$; Fig. 2a). Direction of effects was consistent in FinnDiane and UK GRID (Fig. 2b) whilst in the Sardinia cohort data, rs2941522 was filtered out from the imputation reference panel. SNP rs9747973 in complete LD with rs2941522 ($r^2 = 1.0$ in 1000G EUR) was not associated with diagnosis age in the Sardinian cohort (Beta = 0.012, SE = 0.026, $P = 0.641$), and neither was the chr17q12 locus meta-analysis 2nd top SNP rs11078921 that is in moderate LD ($r^2 = 0.44$ in 1000G EUR) with rs2941522 (Fig. S3).

In FinnDiane, T1D was diagnosed 1.3 years earlier in individuals with the homozygous major allele CC genotype at rs2941522 (T1D diagnosis age median [IQR] = 13.1 [7.7–20.4]), when compared to individuals homozygous for the minor T allele (14.4 [9.5–22.8]). Interestingly, the secondary analysis of the stratified FinnDiane cohort showed that SNP rs2941522 as well as nearly all other top variants in the meta-analysis were associated with diagnosis age ($P < 0.05$) only in individuals with T1D diagnosis age < 16, but not in the late-onset group (Fig. S4).

Another secondary analysis in FinnDiane showed that rs2941522 association was independent of HLA II haplotypes: Effect sizes for the major C-allele were similar in individuals with high-risk DR4-DQ8/DR3-DQ2 heterozygous diplotype (Beta = −0.095, $P = 0.0023$), in those having medium HLA risk (Beta = −0.053, $P = 0.0044$), and amongst individuals with other/protective HLA II haplotypes (Beta = −0.098, $P = 0.017$, heterogeneity test in METAL, $P = 0.378$).

Focus on chr17q12 locus

Index variant

Our index variant rs2941522 at chr17q12 lie in an established T1D susceptibility region: in the T1D ImmunoChip meta-analysis [4], the rs2941522 risk allele was major C (T1D $P = 1.4 \times 10^{-7}$) and the same allele was associated with lower age at diagnosis in our study. In addition, rs2941522 is strongly correlated with the reported T1D locus index SNPs rs2290400 [3] and rs12453507 [4] ($r^2 > 0.8$; 1000G EUR). The age at diagnosis-lowering C-allele of rs2941522 is also correlated with known functional polymorphisms; splice-site-affecting C-allele of rs11078928 ($r^2 = 0.73$) resulting in lower *GSDMB* (Gasdermin B) expression and rs4065275 A-allele disrupting a CTCF-binding motif ($r^2 = 0.74$) causing lower *ORMDL3* (ORMDL Sphingolipid Biosynthesis Regulator 3) expression

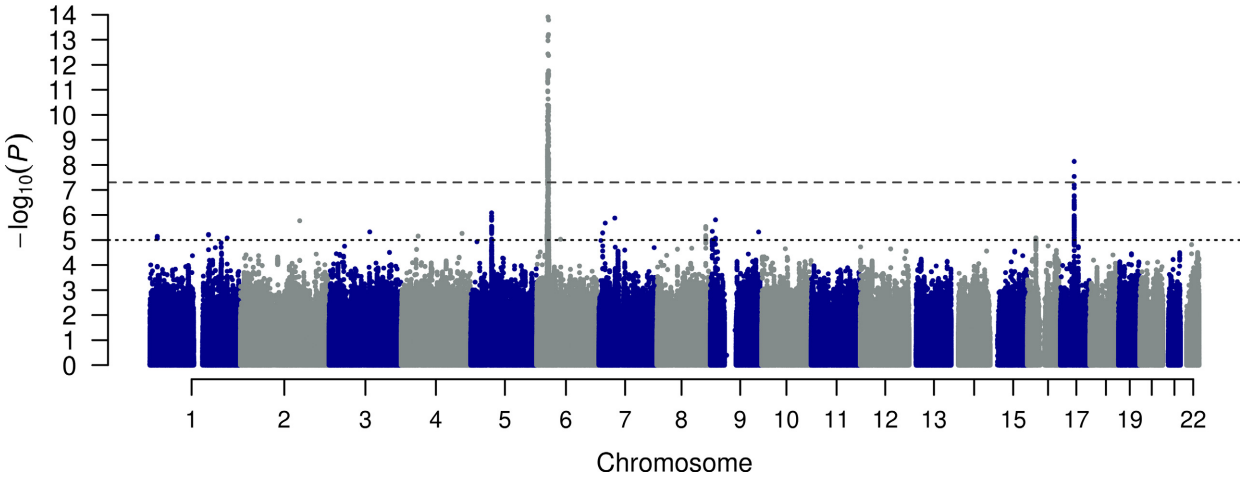

Fig. 1 *Type 1 diabetes diagnosis age GWAS meta-analysis results. Two horizontal lines indicate* P *value thresholds for genome-wide significant (log_{10}(P) > 7.30; P < 5 × 10^{-8}) and suggestive [(log_{10}(P) > 5; P < 10^{-5}] associations. Two peaks emerge in HLA region on chromosome 6 and on chr17q12.*

in whole blood. These correlations are only apparent in 1000G Europeans, whilst in Africans the SNPs are not in LD (Fig. S5).

UKBB

In the UKBB GWAS data, the age at diagnosis-lowering C-allele of rs2941522 was strongly associated with lower white blood cell and neutrophil counts, and lower neutrophil but higher lymphocyte and monocyte percentages ($9.5 \times 10^{-101} \leq P \leq 2.6 \times 10^{-15}$, Table S4). In addition, C-allele was associated with higher age at asthma diagnosis ($P = 3.9 \times 10^{-46}$).

eQTL evidence

Our chr17q12 index variant rs2941522 correlates most strongly with *IKZF3* (IKAROS Family Zinc Finger 3) top eQTL in whole blood (GTEx v.8, $r^2 = 0.972$ with rs907091). SNP rs2941522 is also correlated ($r^2 > 0.80$) with top eQTL variants for *GSDMB* and *ORMDL3* in whole blood, EBV-transformed lymphocytes, spleen and small intestine tissues. The rs2941522 itself is associated with expression of a total of 16 genes in eQTLgen whole blood data [29] ($P \leq 7.6 \times 10^{-6}$) with the strongest association with *IKZF3, GSDMB* and *ORMDL3* expression ($P = 3.3 \times 10^{-310}$). In immune cell subtypes, the C-allele was associated with lower *GSDMB* expression most strongly in naïve B cells (adjusted $P = 7.9 \times 10^{-14}$), but also in many T-cell subtypes, but not in monocytes. eQTL evidence for rs2941522 on *ORMDL3* expression was similar, but less evident in the same cell types (e.g. adjusted $P = 6.3 \times 10^{-8}$ in naïve B cells). Unfortunately, eQTL data for *IKZF3* was modest in the DICE-database (top eQTL had adjusted $P = 0.004$).

Chromatin conformation

Only the B-lymphocyte-derived GM12878 cell line and CD34 haematopoietic stem cells in the Hi-C dataset by Mifsud *et al.* [30] had detectable chromatin conformation interactions for the genomic region containing rs2941522. This region had 3D-interactions (ChiGAGO score ≥ 5) with transcription start sites of multiple genes including *ORMDL3* and *GSDMB* in the GM12878 cell line (Fig. S6) but only one interaction in the haematopoietic stem cells (CD34) with *GSDMA* gene.

Age at diagnosis and T1D susceptibility loci

We specifically examined 62 T1D-associated variants outside the HLA for association with diagnosis age. None of the T1D loci index variants was associated with age at diagnosis at genome-wide significance (Table S5). Nineteen variants (31%), however, had nominal age at diagnosis association ($P < 0.05$), and in all instances, previously reported T1D risk alleles were associated with lower diagnosis age (Binomial test $P = 5.5 \times 10^{-16}$). We replicated ($P < 0.05$) previously shown age at diagnosis-associations at five loci with consistent effect directions with the earlier findings (Table S6), and the strongest association being at chr17q12 locus

Table 1. *Independent SNPs[a] with P < 10^{-5} in the age at diagnosis GWAS meta-analysis*

Position (hg19)	SNP	EA/OA	EA freq[b]	Beta	SE	*P*[c]	Significant at FDR	Effect direction[d]	Average imputation info	Closest gene TSS[e]	In gene[f]
1:21,020,405	rs17407280	C/T	0.777	0.058	0.013	7.2×10^{-6}	0.10	+ + + +	0.94	*KIF17*	*KIF17*
1:161,508,617	rs10919543	G/A	0.346	−0.055	0.012	6.0×10^{-6}	0.10	− − ? −	0.95	*FCGR3A*	*RP11-25K21.6*
1:212,643,660	rs7524087	G/A	0.019	0.285	0.064	8.3×10^{-6}	0.10	? + ? +	0.94	*AC092803.1*	*NA*
2:162,504,240	rs77155259	G/A	0.944	0.142	0.030	1.7×10^{-6}	0.05	+ + ? +	0.71	*SLC4A10*	*SLC4A10*
3:111,593,569	rs111821448	C/T	0.972	0.164	0.036	4.8×10^{-6}	0.05	+ + + +	0.82	*ABHD10*	*PHLDB2*
4:46,374,297	rs139881548	C/T	0.984	0.222	0.049	7.0×10^{-6}	0.10	+ + + ?	0.81	*AC095060.1*	*GABRA2*
4:166,837,480	rs34740712	C/T	0.104	−0.088	0.019	5.4×10^{-6}	0.05	− − ? −	0.88	*TLL1*	*TLL1*
5:57,133,145	rs4513644	T/A	0.794	0.067	0.014	8.2×10^{-7}	0.05	+ + + +	0.99	*LINC02225*	*NA*
6:65,458,587	rs78824139	C/T	0.986	0.294	0.066	9.3×10^{-6}	0.10	+ + ? ?	0.65	*EYS*	*EYS*
7:10,096,655	rs10258381	G/T	0.049	−0.119	0.026	5.2×10^{-6}	0.05	− − − +	0.93	*AC004936.1*	*NA*
7:17,186,779	rs6461299	C/T	0.422	−0.078	0.016	2.1×10^{-6}	0.05	? − − −	0.88	*AC019117.3*	*AC003075.4*
7:43,447,790	rs10255565	C/T	0.154	−0.086	0.018	1.3×10^{-6}	0.05	− − − −	0.78	*LUARIS*	*HECW1*
8:133,787,972	rs35766765	G/C	0.092	−0.091	0.019	2.9×10^{-6}	0.05	− − − −	0.95	*PHF20L1*	*PHF20L1*
9:5,421,234	rs74629033	G/A	0.950	0.116	0.025	4.5×10^{-6}	0.05	+ + + +	0.89	*PLGRKT*	*PLGRKT*
9:14,228,471	rs10961433	G/A	0.867	0.079	0.017	1.5×10^{-6}	0.05	+ + + +	0.91	*AL136366.1*	*NFIB*
9:132,505,659	rs4837404	G/A	0.368	0.064	0.014	4.8×10^{-6}	0.05	+ + ? +	0.68	*PTGES*	*PTGES*
16:22,980,465	rs1110458	C/T	0.239	−0.055	0.012	8.2×10^{-6}	0.10	− − − −	0.99	*AC127459.2*	*NA*
17:37,910,368	rs2941522	C/T	0.575	−0.069	0.012	**7.3×10^{-9}**	0.01	− − − ?	0.92	*GRB7*	*NA*

[a]SNPs in the table are not in LD (r^2 < 0.1 in 1000 Genomes European population). 1000G LD data accessed through LDlink (https://ldlink.nci.nih.gov/).
[b]Effect allele frequency in the meta-analysis cohorts.
[c] Genome-wide significant *P* value in bold.
[d]Direction of the effect (Beta – or +) in the meta-analysis cohorts. Order of the cohorts: FinnDiane, UK GRID Illumina, UK GRID Affymetrix, and Sardinia. A question mark denotes missing data in the cohort.
[e]Nearest gene transcription start site (TSS): data from http://genetics.opentargets.org (Accessed September 2019).
[f]Genomic location in Basic Gene Annotation Set from GENCODE Version 28lift37 (Ensembl 92), Accessed throughhttp://genome.ucsc.edu/.

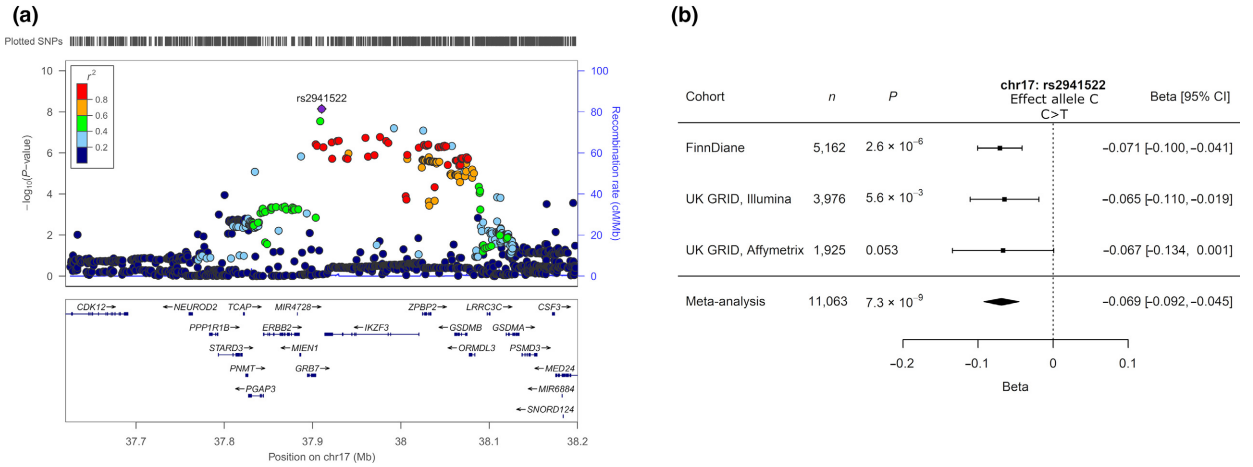

Fig. 2 *(a) Locus zoom plot of GWAS meta-analysis results for T1D diagnosis age on chromosome 17 locus: index SNP rs2941522 ± 290 kilobases. Chromosomal position is hg19 and linkage disequilibrium data (r^2) is from 1000G European population. Blue line denote for recombination rate (b) Forest plot of rs2941522 association with natural logarithm of age at diagnosis in FinnDiane and UK GRID Illumina and Affymetrix cohorts and the corresponding meta-analysis result. In the Sardinian cohort, this variant was absent.*

lead SNP rs2290400 ($P = 3.9 \times 10^{-6}$). The most recent *IL2* locus lead SNP rs75793288 [4] is triallelic and was missing from all our cohorts likely due to filtering out in quality control steps. Therefore, we searched for association results of the previous index variant rs2069763 [31]. T1D risk alleles for rs2069763 (*IL2*), rs12416116 (*RNLS*), rs6476839 (*GLIS3*) and rs34593439 (*CTSH*) were nominally associated with lower diagnosis age in our study (*IL2*; $P = 9.1 \times 10^{-5}$, *RNLS*; $P = 0.014$, *GLIS3*; $P = 0.005$, *CTSH*; $P = 0.001$). SNPs rs72975913 and rs802719 in *PTPRK/THEMIS* locus on chromosome 6q22.33 that showed association with lower disease onset and higher T1D risk in early childhood (disease onset ≤ 5 years) [9], were similarly associated with lower diagnosis age in our study (rs72975913; $P = 4.1 \times 10^{-4}$ and rs802719; $P = 0.037$).

Open chromatin - enrichment analysis

To study genomic location of age at diagnosis-associated SNPs, we performed enrichment analysis with GARFIELD. The number of independent loci from our GWAS meta-analysis with $P < 1 \times 10^{-5}$ were 40 or 17 when excluding HLA-region variants, and 171 or 138 with $P < 1 \times 10^{-4}$ threshold. We chose the $P < 1 \times 10^{-4}$ threshold with higher number of loci as the more relevant for enrichment analysis. These genomic loci were enriched in open chromatin (DHSs) in many tissues, although they were most often located in blood cells (Fig. 3). Performing the same analysis without HLA-region SNPs showed their drastic impact on the results and dropped the number of enriched annotations from 193 to 43, out of which 16 (37%) were different blood cells, 7 (16%) were blood vessel tissues, and 4 (9%) were foetal thymus-tissue-annotations (Binomial test $P < 0.05$ showing enrichment in these three tissues/cells, Table S7). Analysis of another open chromatin signature, genomic footprints, showed four enriched tissues ($P < 2.25 \times 10^{-4}$), of which only skin AG10803–tissue appeared also amongst the significant DHS-annotations (Table S8).

Transcriptome-wide association analysis

To utilize SNPs affecting gene expression (eQTL SNPs) in a genome-wide setting, we studied the association of age at diagnosis and gene expression with transcriptome-wide association analysis. Primary analysis with five GTEx tissues showed 17 significant associations between gene expression and T1D diagnosis age at FDR = 0.05 ($P \leq 1.2 \times 10^{-5}$, Table 2). In whole blood, only higher *IKZF3* (chr17q12) expression was associated earlier T1D diagnosis ($P = 1.8 \times 10^{-6}$). Higher *IKZF3* expression was associated with lower age at diagnosis also in EBV-transformed lymphocytes,

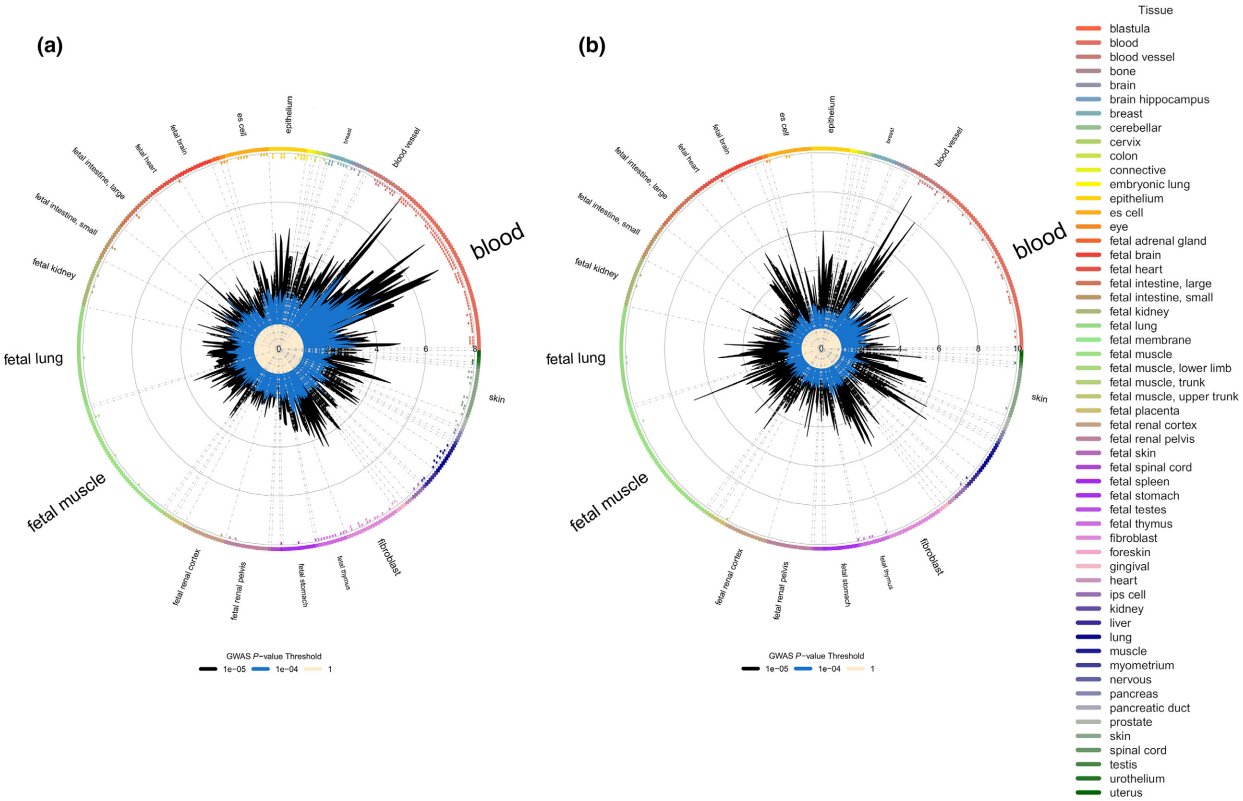

Fig. 3 *DNAse I hypersensitive-site annotations enrichment analysis results of age at diagnosis GWAS meta-analysis results with* P *value thresholds* $< 10^{-5}$ *(inner dots) or* $< 10^{-4}$ *(a) and the same analysis omitting the HLA-region variants (b). The blue and black colours in the centre show the enrichment odds ratios of different annotations for different GWAS meta-analysis result* P *value thresholds as defined in the figure.*

whilst several other genes at the same chr17q12 locus were associated with lower diagnosis age also in other tissues: lower expression of *GSDMB* and *ORMDL3* in EBV-transformed lymphocytes, small intestine and spleen, and higher *ZPBP2* expression in EBV-transformed lymphocytes and small intestine. Of the HLA-region genes, higher *HLA-G* expression in EBV-transformed lymphocytes, higher *LST1* in spleen, and lower *MICA* expression in EBV-transformed lymphocytes and pancreas were associated with lower age at diagnosis.

The other significant ($P \leq 1.3 \times 10^{-5}$) gene in pancreas was *PHF20L1* (PHD Finger Protein 20 Like 1) on chromosome 8, with increased expression associated with lower diagnosis age. Other tissues with higher *PHF20L1* expression associating with lower diagnosis age were EBV-transformed lymphocytes and spleen. Of note, this gene is located under the suggestive age at diagnosis meta-GWAS peak on chromosome 8 ($P < 1 \times 10^{-5}$) and the locus index SNP rs35766765 is in the first intron of the gene and 368 bp from the TSS (Fig. S7).

We analysed five additional tissues (GTEx) and retrieved chr17q12 region and *PHF20L1* (chr8) gene association results. In most of the analysed tissues, *GSDMB*, *ORMDL3*, *PNMT*, and *PGAP3* expression were associated with diagnosis age, whereas *ZPBP2* expression results were missing from other tissues, and additional *IKZF3* association with diagnosis age was seen only in lung tissue ($P = 1.8 \times 10^{-6}$, Table S9). Predicted *PHF20L1* expression was associated with diagnosis age also in thyroid gland, liver, and lung ($P < 1.5 \times 10^{-5}$).

We used another whole blood transcriptome prediction model (DGN data) in MetaXcan, and this analysis showed significant associations ($P < 3.4 \times 10^{-5}$, FDR = 0.05) for two genes on chr17q12 (*IKZF3,*

Table 2. *Transcriptome-wide association analysis results*

Tissue	Genome region	Significant associations with age at diagnosis of T1D, FDR 5%[a]			
		Ensemble id	Gene	z-score	*P*
GTEx v.8 tissues					
Whole blood: 12136 genes					
	chr17q12	ENSG00000161405.16	*IKZF3*	−4.77	1.8×10^{-6}
EBV-transformed lymphocytes: 11 798 genes					
	chr6; MHC I	*ENSG00000204632.11*	*HLA-G*	−4.60	4.2×10^{-6}
	chr6; MHC I	*ENSG00000204520.12*	*MICA*	4.50	6.9×10^{-6}
	chr8q24.22	*ENSG00000129292.20*	*PHF20L1*	−4.68	2.9×10^{-6}
	chr17q12	*ENSG00000186075.12*	*ZPBP2*	−4.74	2.2×10^{-6}
	chr17q12	*ENSG00000073605.18*	*GSDMB*	4.74	2.2×10^{-6}
	chr17q12	*ENSG00000172057.9*	*ORMDL3*	4.62	3.8×10^{-6}
	chr17q12	*ENSG00000161405.16*	*IKZF3*	−4.56	5.2×10^{-6}
Spleen: 13 486 genes					
	chr6: MHC III	*ENSG00000204482.10*	*LST1*	−4.49	7.2×10^{-6}
	chr8q24.22	*ENSG00000129292.20*	*PHF20L1*	−4.66	3.1×10^{-6}
	chr17q12	*ENSG00000073605.18*	*GSDMB*	5.02	5.0×10^{-7}
	chr17q12	*ENSG00000172057.9*	*ORMDL3*	4.62	3.80×10^{-6}
Pancreas: 13 132 genes					
	chr6; MHC I	*ENSG00000204520.12*	*MICA*	4.38	1.2×10^{-5}
	chr8q24.22	*ENSG00000129292.20*	*PHF20L1*	−4.79	1.7×10^{-6}
Small intestine, terminal ileum: 13 395 genes					
	chr17q12	*ENSG00000186075.12*	*ZPBP2*	−4.74	2.2×10^{-6}
	chr17q12	*ENSG00000073605.18*	*GSDMB*	4.62	3.8×10^{-6}
	chr17q12	*ENSG00000172057.9*	*ORMDL3*	4.39	1.1×10^{-5}
Whole blood from Depression Genes and Networks (DGN) study: 11 530 genes					
	chr6; MHC I	*ENSG00000204520.8*	*MICA*	5.89	4.0×10^{-9}
	chr6; MHC III	*ENSG00000204304.6*	*PBX2*	5.44	5.4×10^{-8}
	chr6; MCH III	*ENSG00000204438.6*	*GPANK1*	5.01	5.5×10^{-7}
	chr17q12	*ENSG00000141744.3*	*PNMT*	4.24	2.3×10^{-5}
	chr17q12	*ENSG00000161405.12*	*IKZF3*	−4.19	2.8×10^{-5}
	chr20q13.12	*ENSG00000158296.9*	*SLC13A3*	4.23	2.3×10^{-5}

[a]Benjamin-Hochberg FDR-calculation separately for combined GTEx tissues (whole blood, EBV-transformed lymphocytes, spleen, pancreas, small intestine) and DGN whole blood.

PNMT), three genes in the HLA-region (*MICA, GPANK1, PBX2*) and *SLC13A3* on chr20 (Table 2). Furthermore, lower *GSDMB* and *ORMDL3* and higher *PHF20L1* expression were also associated with lower diagnosis age although non-significantly at FDR = 0.05 ($6.5 \times 10^{-4} < P < 1.3 \times 10^{-3}$).

We sought for replication of our transcriptome-wide associations in whole blood, EBV-lymphocytes, spleen, pancreas and small intestine significant at FDR = 0.05 in the UKBB Age diabetes diagnosed transcriptome-wide association study. In addition, we queried top three genes found in the UKBB in the same tissues, from our results. The gene associations at chr17q12 or *PHF20L1* on chromosome 8 did not replicate in the UKBB diabetes diagnosis age transcriptome-wide analysis, which is probably explained by the lack of

association of the region SNPs with the UKBB T1D phenotype (Table S10). Of our age at diagnosis-associated genes within HLA, the *MICA* association was to the other direction: higher *MICA* expression in EBV-transformed lymphocytes and pancreas were associated with lower diabetes diagnosis age in the UKBB data. Although diagnosis age distributions were very different in the UKBB Age diabetes diagnosis- phenotype (1–80) and our study (0.5–40), UKBB top association results *CYP21A2* and *HLA-DQB2* in whole blood, *HLA-DQA2* in EBV-transformed lymphocytes, *CLIC1* in the spleen, *POM* and *MSH5* in the pancreas, and *HLA-DQB1* in the small intestine tissue were nominally associated ($P < 0.05$) with age at diagnosis in our data (Table S11).

Discussion

In our GWAS meta-analysis, as expected, variants in the HLA region were strongly associated with T1D diagnosis age. We identified 21 independent significant or suggestive associations within the HLA, a genomic region, which has extremely high allelic diversity and differences in LD structures between populations. Therefore, our many independent top loci are not necessarily the most important in other cohorts/populations. Our results showed again, however, the importance of the HLA region on T1D onset, and suggested that not only classical HLA genes, but also other genes in class III region too, might affect diagnosis age. Nevertheless, our main interest was the genome beyond HLA, where SNPs on chromosome 17q12 showed an association with age at diagnosis and transcriptome-wide association analyses in multiple tissues pointed out many genes in the same region.

Chromosome 17q12 locus is associated with autoimmune diseases e.g. Crohn disease, ulcerative colitis and rheumatoid arthritis [32]. This locus seems to co-ordinate lymphocyte development, and the locus is also associated with childhood acute lymphoblastic leukaemia [33], in which cancerous cell type is lymphocyte precursor cell, lymphoblast. Interestingly, our chr17q12 locus index SNP rs2941522 was the locus lead SNP also in a recent asthma GWAS [34], and chr17q12 association is strongest with childhood-onset asthma [35]. In a transcriptome-wide analysis similar to ours, multiple genes in the chr17q12 were associated with childhood-onset asthma only [36]. The effects of the risk alleles at this locus are opposite for the asthma and T1D: rs2941522 C-allele associated with protection from asthma [34], in our study, was associated with earlier T1D diagnosis. Asthma-associated variants at the locus were in open chromatin and regulated *GSDMB* and *ORMDL3* expression in multiple immune cells types excluding monocytes and dendritic cells [37]. Our study showed the same: lead variants at chr17q12 locus were present at all enriched DHS-annotations of different blood cells and predicted expression levels of *GSDMB* and *ORMDL3* were associated with T1D diagnosis age.

Earlier studies on chr17q12 locus mostly concern asthma and they are thoroughly reviewed by Stein and colleagues [32]. In the diabetes context, *ORMDL3* might play a role in pancreatic beta cell endoplasmic reticulum stress response during newly-onset diabetes [38]. The study found low *ORMDL3* mRNA levels in leucocytes in children with T1D, and the expression decreased during their disease progression. To our knowledge, no studies for *GSDMB* in diabetes have been reported, but eQTL SNPs for *GSDMB* and *ORMDL3* co-regulate the expression of these genes in Europeans [32]. Our results showed the same; both genes associated with age at diagnosis in transcriptome-wide association analyses in multiple tissues. In addition, within the same chr17q12 locus, expression of *IKZF3*, *ZPBP2, PNMT* were associated with T1D diagnosis age in our study. Of these proteins, transcription factor *IKZF3* involved in T-cell and B-cell differentiation and proliferation has probably the highest relevance for immunity.

Whilst *GSDMB* and *ORMDL3* expression is ubiquitous, *IKZF3* expression level is highest in EBV-transformed lymphocytes, spleen and small intestine tissue (GTEx v.8). Interestingly, Ram and colleagues showed that T1D risk SNP at chr17q12 regulated *IKZF3* expression (amongst other genes in the locus) most strongly in EBV-transformed B cells [39]. It is intriguing that amongst the T1D susceptibility loci there are genomic regions near three Ikaros-family zing finger proteins: *IKZF1* (7p12.2*), IKZF3* (17q12.2), and *IKZF4* (12q.13.2) that form homo- and heterodimers and have function in lymphocyte differentiation and development [40] as well as in haematological cancers. Given the recent evidence on B-cell involvement in the pathogenesis of T1D only in individuals with early disease onset, our findings suggests that *IKZF3* could play a role in T1D onset in childhood.

Ubiquitously expressed *PHF20L1* was associated with T1D diagnosis age in transcriptome-wide

analysis of multiple tissues and locus index SNP is close to TSS. This gene encodes a transcriptional regulator protein that also recognizes methylated lysine and stabilizes DNA methyltransferase 1 (DNMT1) and prevents its degradation [41]. We found no clear link between diabetes and *PHF20L1* in the literature but in early differentiation of T-cell subtypes, *PHF20L1* expression differ in Th1 and Th2 [42]. Another gene with no previous T1D indications was *SLC13A3* (chr20) that was associated with T1D diagnosis in DGN whole blood in our study. We could not, however, confirm this association in GTEx whole blood (*SLC13A3* missing), or in other studied tissues ($P > 0.05$, data not shown).

Known T1D risk loci show substantial pleiotropy with other autoimmune diseases [4, 43]. The ImmunoChip is useful for detailed investigation of these autoimmune disease-associated loci. Even though our study targeted the whole genome, the top association results located at previously confirmed T1D susceptibility regions at chr6 and chr17, and the genome beyond the confirmed T1D loci carried only suggestive associations with disease onset. Of note, only two of our eighteen suggestive/significant independent non-HLA index variants, however, are in the ImmunoChip genotyping SNP panel. Our study was well powered with 12,539 individuals, but included only three cohorts all of European origin. Our study approach lacked non-diabetic controls, and therefore we can only speculate whether our top variants were associated with T1D *per se*. Search from UKBB PheWAS results showed, however, that most of them were nominally associated ($P < 0.05$) with at least some diabetes-related traits (e.g. T1D, insulin use, age diabetes diagnosed; search from Open Targets, data not shown). In addition, *PHF20L1,* gene not previously implicated in diabetes, was associated with T1D diagnosis age in our transcriptome-wide association study of multiple tissues. This all implies that the genome beyond the ImmunoChip may still carry T1D-risk-increasing variants, urging for further genome-wide search for novel T1D susceptibility loci.

The FinnDiane cohort spans a whole spectrum of T1D diagnosis ages, here defined as 0.5–40 years, whereas the T1D diagnosis in the UK GRID and Sardinian cohorts was made mostly in childhood (<16 years). Because we treated disease-onset age as a continuous variable, the UK GRID and Sardinian cohort had less phenotypic variation. Stratifying the FinnDiane cohort according to the T1D diagnosis age we noticed that even though the effect sizes of the meta-analysis top variants were often the largest in the whole FinnDiane, most of the variants showed significant effects only in the early-onset group, i.e. the genetic differences were evident already amongst the < 16 year-diagnosis group (Fig. S7). This is in line with the previous knowledge how a pathogenic process leading to T1D starts in early childhood and therefore also disease-susceptibility variants might play the largest role in early years of life.

On the other hand, many T1D risk variants seem to confer disease risk similarly in paediatric and adult T1D cases [44]. Still, some variants do show age-effects and we nominally replicated most of the earlier findings. We are not the first to report age-related effects at chr17q12. One study showed rs11078927 to have varying effect sizes according to age at onset [45], and in a Japanese cohort, the T1D susceptibility SNP rs2290400 was associated with T1D risk in those with early disease onset [46]. In the recent study of 8,586 Caucasian individuals with T1D and 18,485 controls and with the UK GRID cohort included, T1D SNP in the locus was associated with higher T1D risk in those diagnosed < 7 years [10]. Our study proved the same; our locus index SNP allele correlated with T1D risk allele, was associated with lower T1D diagnosis age.

Conclusion

To conclude, our GWAS meta-analysis for T1D diagnosis age highlighted multiple independent SNPs in the HLA region, gave suggestive evidence for a variant near *PHF20L1* and confirmed the known T1D susceptibility region on chromosome 17. This region contains many co-regulated candidate causal genes of which our transcriptome-wide association analyses pointed out *IKZF3, GSDMB*, *ORMDL3, PNMT*, and *ZPBP2*. Nevertheless, gene expression is both cell type and developmental time-dependent and to find the most important gene(s) we need to catch right tissues and cells at relevant developmental stages. In T1D, this would mean cells and tissues of children aged < 5 years, when the first signs of autoimmunity might already appear. Only more detailed studies may show, what are biological roles of the genes in the chr17q12 locus, which seem to carry many immune disease-associated variants with effects for age at disease onset.

Acknowledgements

We authors acknowledge Prof. John A. Todd and MD Jamie R. J. Inshaw for giving access to the UK GRID GWAS for age at diagnosis data and providing original idea to perform this kind of study. We kindly thank all individuals taking part in these cohorts and research staff for gathering invaluable data. We also thank doctors and nurses at various FinnDiane centres for their efforts to recruit and characterize the individuals with type 1 diabetes included in the FinnDiane Study (Table S12).

Conflict of Interest statement

P.-H.G. has received research grants from Eli Lilly and Roche; is an advisory board member for AbbVie, Astellas, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Medscape, Merck Sharp & Dohme (MSD), Mundipharma, Novartis, and Sanofi; and has also received lecture fees from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water, Genzyme, Medscape, MSD, Mundipharma, Novartis, Novo Nordisk, and Sanofi. Other authors (A.S., N.S., C.S., F.C., J.H., and V.H.) declare no conflicts of interests relevant to this article.

Sources of Funding

The FinnDiane Study was supported by grants from the Folkhälsan Research Foundation, Wilhelm and Else Stockmann Foundation, Novo Nordisk Foundation (NNF OC0013659), Liv och Hälsa Society, Helsinki University Hospital Research Founds (TYH2018207), Academy of Finland (275614, 299200 and 316664) and European Foundation for the Study of Diabetes (EFSD, N.S.). JDRF supported the genotyping of the FinnDiane subjects (grant 17-2013-7).

Author Contribution

Anna Syreeni: Formal analysis (lead); Investigation (lead); Project administration (lead); Software (lead); Visualization (lead); Writing-original draft (lead); Writing-review & editing (lead). **Niina Sandholm:** Data curation (equal); Formal analysis (supporting); Funding acquisition (equal); Project administration (equal); Supervision (lead); Writing-review & editing (lead). **Carlo Sidore:** Data curation (equal); Investigation (equal); Resources (equal); Writing-review & editing (supporting). **Francesco Cucca:** Data curation (equal); Resources (equal). **Jani Haukka:** Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Writing-original draft (supporting); Writing-review & editing (supporting). **Valma Harjutsalo:** Data curation (equal); Funding acquisition (lead); Investigation (supporting); Resources (equal); Writing-review & editing (supporting). **Per-Henrik Groop:** Funding acquisition (lead); Project administration (lead); Resources (lead); Supervision (equal); Writing-review & editing (equal).

References

1 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447:** 661–78.
2 Todd JA, Walker NM, Cooper JD *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; **39:** 857–64.
3 Barrett JC, Clayton DG, Concannon P *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41:** 703–7.
4 Onengut-Gumuscu S, Chen W, Burren O *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet* 2015; **47:** 381–86.
5 Leete P, Willcox A, Krogvold L *et al.* Differential insulitic profiles determine the extent of beta-cell destruction and the age at onset of type 1 diabetes. *Diabetes* 2016; **65:** 1362–69.
6 Hyttinen V, Kaprio J, Kinnunen L *et al.* Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 2003; **52:** 1052–55.
7 Valdes AM, Erlich HA, Carlson J *et al.* Use of class I and class II HLA loci for predicting age at onset of type 1 diabetes in multiple populations. *Diabetologia* 2012; **55:** 2394–401.
8 Howson JM, Cooper JD, Smyth DJ *et al.* Evidence of gene-gene interaction and age-at-diagnosis effects in type 1 diabetes. *Diabetes* 2012; **61:** 3012–17.
9 Inshaw JR, Walker NM, Wallace C *et al.* The chromosome 6q22.33 region is associated with age at diagnosis of type 1 diabetes and disease risk in those diagnosed under 5 years of age. *Diabetologia* 2018; **61:** 147–57.
10 Inshaw JRJ, Cutler AJ, Crouch DJM *et al.* Genetic variants predisposing most strongly to type 1 diabetes diagnosed under age 7 years lie near candidate genes that function in the immune system and in pancreatic beta-cells. *Diabetes Care* 2020; **43:** 169–77.
11 Sharp SA, Rich SS, Wood AR *et al.* Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. *Diabetes Care* 2019; **42:** 200–7.
12 Nyaga DM, Vickers MH, Jefferies CA *et al.* Type 1 diabetes mellitus-associated genetic variants contribute to overlapping immune regulatory networks. *Front Genet* 2018; **9:** 535.
13 Syreeni A, Sandholm N, Cao J *et al.* Genetic determinants of glycated hemoglobin in type 1 diabetes. *Diabetes* 2019; **68:** 858–67.
14 Sidore C, Busonero F, Maschio A *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments

association analyses for lipid and blood inflammatory markers. *Nat Genet* 2015; **47:** 1272–81.
15 Marchini J, Howie B, Myers S *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39:** 906–13.
16 Zhan X, Hu Y, Li B *et al.* RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016; **32:** 1423–26.
17 Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26:** 2190–91.
18 Jia X, Han B, Onengut-Gumuscu S *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013; **8:** e64683.
19 Söderlund J, Forsblom C, Ilonen J *et al.* HLA class II is a factor in cardiovascular morbidity and mortality rates in patients with type 1 diabetes. *Diabetologia* 2012; **55:** 2963–69.
20 Iotchkova V, Ritchie GRS, Geihs M *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* 2019; **51:** 343–53.
21 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489:** 57–74.
22 Bernstein BE, Stamatoyannopoulos JA, Costello JF *et al.* The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 2010; **28:** 1045–48.
23 Gamazon ER, Shah KP. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet.* 2015;**47:** 1091–8.
24 Barbeira AN, Dickinson SP, Bonazzola R *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018; **9:** 1825.
25 Barbeira AN, Bonazzola ER, Gamazon Y *et al.* Widespread dose-dependent effects of RNA expression and splicing on complex diseases and traits. *BioRxiv* 2019. https://doi.org/10.1101/814350.
26 Battle A, Mostafavi S, Zhu X *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* 2014; **24:** 14–24.
27 Pividori M, Rajagopal PS, Barbeira A *et al.* PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *BioRxiv* 2019;. https://doi.org/10.1101/833210.
28 Schofield EC, Carver T, Achuthan P *et al.* CHiCP: A web-based tool for the integrative and interactive visualization of promoter capture hi-C datasets. *Bioinformatics* 2016; **32:** 2511–13.
29 Võsa U, Claringbould A, Westra H *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *BioRxiv* 2018;. https://doi.org/10.1101/447367.
30 Mifsud B, Tavares-Cadete F, Young AN *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* 2015; **47:** 598–606.
31 Bradfield JP, Qu H, Wang K *et al.* A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet* 2011; **7:** e1002293.
32 Stein MM, Thompson EE, Schoettler N *et al.* A decade of research on the 17q12-21 asthma locus: Piecing together the puzzle. *J Allergy Clin Immunol* 2018; **142:** 749–64.
33 Wiemels JL, Walsh KM, de Smith AJ *et al.* GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. *Nat Commun* 2018; **9:** 286.
34 Shrine N, Portelli MA, John C *et al.* Moderate-to-severe asthma in individuals of european ancestry: A genome-wide association study. *Lancet Respir Med* 2019; **7:** 20–34.
35 Bouzigon E, Corda E, Aschard H *et al.* Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008; **359:** 1985–94.
36 Pividori M, Schoettler N, Nicolae DL *et al.* Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: Genome-wide and transcriptome-wide studies. *Lancet Respir Med* 2019; **7:** 509–22.
37 Schmiedel BJ, Seumois G, Samaniego-Castruita D *et al.* 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat Commun* 2016; **7:** 13426.
38 Yang W, Sheng F, Sun B *et al.* The role of ORMDL3/ATF6 in compensated beta cell proliferation during early diabetes. *Aging (Albany NY)* 2019; **11:** 2787–96.
39 Ram R, Morahan G. Effects of type 1 diabetes risk alleles on immune cell gene expression. *Genes (Basel)* 2017; **8:** 167.
40 Heizmann B, Kastner P, Chan S. The ikaros family in lymphocyte development. *Curr Opin Immunol* 2018; **51:** 14–23.
41 Estève PO, Terragni J, Deepti K *et al.* Methyllysine reader plant homeodomain (PHD) finger protein 20-like 1 (PHF20L1) antagonizes DNA (cytosine-5) methyltransferase 1 (DNMT1) proteasomal degradation. *J Biol Chem* 2014; **12:** 8277–87.
42 Lund RJ, Löytömäki M, Naumanen T *et al.* Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation. *J Immunol* 2007; **178:** 3648–60.
43 Pociot F. Type 1 diabetes genome-wide association studies: Not to be lost in translation. *Clin Transl Immunol* 2017; **6:** e162.
44 Howson JM, Rosinger S, Smyth DJ *et al.* Genetic analysis of adult-onset autoimmune diabetes. *Diabetes* 2011; **60:** 2645–53.
45 Liley J, Todd JA, Wallace C. A method for identifying genetic heterogeneity within phenotypically defined disease subgroups. *Nat Genet* 2017; **49:** 310–16.
46 Ayabe T, Fukami M, Ogata T *et al.* Variants associated with autoimmune type 1 diabetes in Japanese children: Implications for age-specific effects of cis-regulatory haplotypes at 17q12-q21. *Diabet Med* 2016; **33:** 1717–22.

Correspondence: Per-Henrik Groop, Abdominal Center, Nephrology, University of Helsinki and Helsinki University Hospital, Biomedicum Helsinki, C318b, Haartmaninkatu 8, Helsinki FIN-00290, Finland.
(e-mail: per-henrik.groop@helsinki.fi)

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Genome-wide search for genes affecting the age at diagnosis of type 1 diabetes.■