



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Novel Linkage Peaks Discovered for Diabetic Nephropathy in Individuals With Type 1 Diabetes

Citation for published version:

FinnDiane Study Group 2021, 'Novel Linkage Peaks Discovered for Diabetic Nephropathy in Individuals With Type 1 Diabetes', *Diabetes* . <https://doi.org/10.2337/db20-0158>

Digital Object Identifier (DOI):

[10.2337/db20-0158](https://doi.org/10.2337/db20-0158)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Diabetes

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Novel linkage peaks discovered for diabetic nephropathy in individuals with type 1 diabetes

**Jani Haukka^{1,2,3}, Niina Sandholm^{1,2,3}, Erkka Valo^{1,2,3}, Carol Forsblom^{1,2,3}, Valma Harjutsalo^{1,2,3,4},
Joanne B. Cole^{5,6,7,8}, Stuart J. McGurnaghan⁹, Helen M. Colhoun⁹, Per-Henrik Groop^{1,2,3,10},
on behalf of the FinnDiane Study Group.**

¹ Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki

² Abdominal Center, Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

³ Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland

⁴ National Institute for Health and Welfare, Chronic Disease Prevention Unit, Helsinki, Finland

⁵ Division of Endocrinology, Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts, USA

⁶ Programs in Metabolism, Broad Institute, Cambridge, Massachusetts, USA

⁷ Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

⁸ Programs in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA

⁹ Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

¹⁰ Department of Diabetes, Central Clinical School, Monash University, Melbourne, Australia

Corresponding author: Per-Henrik Groop, MD, DMSc, FRCPE
Folkhälsan Institute of Genetics
Folkhälsan Research Center
Biomedicum Helsinki (C318b)
PO Box 63, FIN-00014 University of Helsinki
Finland
Email: per-henrik.groop@helsinki.fi

Abstract: 200

Word count: 4241

Figures: 3

Tables: 3

Supplementary Figures: 4

Supplementary Tables: 11

Abstract

Genome-wide association studies (GWAS) and linkage studies have had only limited success in discovering genome-wide significantly linked regions or risk loci for diabetic nephropathy in individuals with type 1 diabetes (T1D). As GWAS cohorts have grown, they have also included more documented and undocumented familial relationships. Here, we computationally inferred and manually curated pedigrees in a study cohort of more than 6,000 individuals with T1D and their non-diabetic relatives. We performed linkage study for 177 pedigrees consisting of 452 individuals with T1D and their relatives using a genome-wide genotyping array with more than 300,000 SNPs and the PSEUDOMARKER software. The analysis resulted in genome-wide significant linkage peaks on eight chromosomal regions from five chromosomes (logarithm of odds [LOD]>3.3). The highest peak was localized at the HLA region on chromosome 6p, but whether the peak originates from T1D or diabetic nephropathy, remains ambiguous. Of the other significant peaks, the chromosome 4p22 region is localized on top of a gene associated with focal segmental glomerulosclerosis, *ARHGAP24*, suggesting that the gene may play a role in diabetic nephropathy as well. Furthermore, rare variants have been associated with diabetic nephropathy and chronic kidney disease near the 4q25 peak, localized on top of *CCSER1*.

Introduction

Diabetic nephropathy (DN, diabetic kidney disease) is a microvascular complication of diabetes that causes progressive decline in kidney function. Approximately one third of individuals with type 1 diabetes (T1D) and half of the individuals with type 2 diabetes develop some degree of kidney function impairment (1). For a significant proportion of these individuals, DN eventually leads to severe impairment of kidney function, end-stage renal disease (ESRD), that can only be treated with dialysis or a kidney transplant.

Both genetic and environmental factors play a role in the development and progression of the disease. DN segregates in families with sibling recurrence risk of 2.3. Overall, genetic factors are thought to explain approximately one third of the DN risk in those with T1D, although after excluding static risk factors, e.g. sex, the genetics was shown to explain half of the disease risk (2; 3).

Although **up to twenty** susceptibility genes and loci have been found for DN by recent Genome-Wide Association Studies (GWASs) with increasingly large sample sizes, much of the predicted genetic risk of the disease remains unexplained (4; 5). However, while GWASs are best powered to detect common variants, many common diseases also have rarer variants that affect the disease risk at the individual level even more than the common variants do (2). Traditional GWASs do not capture rare variants, and whole exome or genome sequencing of unrelated individuals would need considerable sample sizes to detect such rare and low frequency variants. Many of these variants may also be population specific, and using pedigree-based data, linkage methods may serve as a more efficient design for the initial variant discovery (6).

Several previous linkage studies found a genome-wide significant or suggestive linkage peak on chromosome 3q in Finnish and other populations (7). More recently, a significant linkage peak was found on chromosome 22 in Danish, Finnish and French sibling pairs (8). However, these linkage peaks do not occur in overlapping regions with GWAS findings, and the genetic background of the findings remains largely unclear.

Here we present a genetic linkage study performed in small pedigrees consisting of Finnish individuals with T1D and genotyped with a modern genome-wide genotyping chip that included additional exonic variants on top of common genome-wide variants. The related individuals were extracted from the GWAS data, and the

pedigrees were built computationally based on the identity-by-descent matrix (IBD) of their genetic distances; the 2nd and 3rd degree relationships were manually assessed and confirmed to construct the pedigrees.

Research Design and Methods

Genotyping and Imputation

In total, 6,152 participants, including individuals with T1D and their relatives, were genotyped on three batches with Illumina HumanCoreExome Bead arrays 12-1.0, 12-1.1, and 24-1.0 (Illumina, San Diego, CA, USA) at the University of Virginia. Genotyping, variant calling, quality control (QC) and imputation have been described earlier (9). Variants were called with zcall algorithm to optimize calling of rare SNPs. QC included removal of variants with low genotyping call rate, deviation from Hardy-Weinberg equilibrium, allele frequency difference >20% (MAF 5%) or allele frequency difference >5% (MAF>5%) with the 1000 Genomes EUR population. The human reference genome GRCh37 was used for genotype coordinates.

A total of 6,019 individuals passed QC, when samples with genotyping rate <0.95, extreme heterozygosity, sample mix-ups and genetic outliers, were removed. After genotype quality control, phased haplotypes (SHAPEIT v2r837) of 316,899 SNPs were used for imputation with 1000 Genomes EUR phase 3 version 5, using Minimac3/Minimac3-omp v1.0.14 (10-12).

Study Cohort

At the Finnish Diabetic Nephropathy (FinnDiane) Study visit, each patient underwent a thorough clinical investigation as described earlier (13). Micro- or macroalbuminuria was defined based on the urinary albumin excretion rate (AER) in 2 out of 3 consecutive timed overnight or 24h urine collections as previously described (13). ESRD was defined as ongoing dialysis or the patient having received a kidney transplant. ESRD diagnosis year was verified from hospital records.

Individuals with micro-, macroalbuminuria or ESRD were set as cases and those with normal AER for at least 15 years were set as controls. We included individuals with T1D with diabetes onset before the age of 40,

and insulin treatment started within one year of the diagnosis of diabetes. All participants gave their written consent for this study.

Pedigrees

Pedigrees were built based on the individuals' pairwise genetic distances in the SNP data. KING software was used to compute the pairwise genetic distances (based on the proportion of alleles with IBD=0, IBD=1, and IBD=2) between the subjects in the genotypic data (14). Pairs of individuals with 3rd degree relationship or closer ($IBD=0 < 1 - \frac{1}{2^{5/2}}$, kinship coefficient $> \frac{1}{2^{9/2}}$), were considered to the automatic inferring of the pedigrees with the PRIMUS software(15).

As the IBD=0 and kinship coefficients may be highly similar between different types of 2nd and 3rd degree relative pairs, we only used the automatic pedigree construction methods for pedigrees with complete parent-offspring and full-sibling familial relations. We used the more distant relation predictions as a basis to search and verify the actual relationship types using multiple sources such as population registries, medical records and questionnaires. The 2nd degree relationships (representing either grandparent-grandchild pairs, avuncular pairs, or half-siblings) were labeled as avuncular, if the age difference between the subjects was more than 30 years (unlikely to be half-siblings) but less than 35 years (unlikely grandparent-grandchild pairs). Unconfirmed 2nd and 3rd degree familial relations were not added to the pedigrees, and the final pedigree structures were checked using the Pedcheck software for the general structure. PLINK was used to check the Mendelian error rate for SNPs (16; 17).

Parametric Two-point Linkage Analysis

The linkage analysis for the autosomal markers was performed with the PSEUDOMARKER software for the directly genotyped SNPs by using default parameters and recessive inheritance model, resembling affected sibling pairs model for concordant sibling pairs. The frequency of phenocopies was 0 and a LOD-score of 3.3 was considered as genome-wide significant threshold (18; 19). We subsequently ran the linkage analysis on imputed markers for ± 1 Mbp from the genome-wide significant linkage peaks, except for the HLA region. The

markers were limited to those with imputation quality r^2 over 0.75. We also performed separate analysis with pedigrees with only discordant and concordant individuals for DN, along with complex trees that included at least one unaffected individual and two or more affected ones. We also performed an additional linkage analysis using the dominant inheritance model to confirm overlapping peaks in both models.

HLA Allele Prediction and Validation of HLA Allele Imputation

We have previously genotyped the HLA alleles for the *DQA1*, *DQB1* and *DRB1* genes for 4,279 FinnDiane study participants (20). As almost half of the individuals in the pedigrees did not have HLA alleles typed, we imputed the HLA alleles using SNP2HLA software based on GWAS data from the HLA-region and the Type 1 Diabetes Genome Consortium's (T1DGC) reference genotype-HLA-allele panel (21).

***In silico* eQTL and mQTL Mapping Using Public Datasets**

We performed an *in silico* eQTL study for markers with genome-wide significant linkage LOD-scores for DN by searching two databases with kidney eQTL data: Human Kidney eQTL Atlas and NephQTL (22). We also queried the GTEx database for eQTL changes on various tissues and methylation quantitative trait loci (mQTL) from the mQTLdb database for changes on CpG-island methylation in blood (23; 24).

***In silico* Replication of Linkage Peaks**

Of the diabetic nephropathy studies thus far, the study by Salem et al. (2019) includes the highest number of individuals (n= 19,327). Therefore, to search for further evidence for replication of our linkage peaks we used the publicly available meta-analysis results of that study (<https://t2d.hugeamp.org/datasets.html>). Based on the phenotype similarity with the current study and the replication dataset, we used the albuminuria-based "all diabetic kidney disease" phenotype (micro- or macroalbuminuria or ESRD vs. normal AER). Of note, the data include also unrelated individuals from the current linkage study.

Results

Individual and Pedigree Characteristics

Of the total of 177 pedigrees, more than half ($n = 95$) consisted of sibling pairs either concordant or discordant for DN. Furthermore, the dataset included 40 pedigrees with sibships and more complex relationship structures, resulting in a median pedigree size of three. A total of 263 subjects were affected DN cases (micro- or macroalbuminuria or ESRD) and 120 individuals with T1D had had normal AER for at least 15 years and were assigned as unaffected individuals. Of note, 57 individuals had either unknown DN status, or short disease duration, or they were non-diabetic parents of the included participants; these individuals were included in the families but with an unknown phenotype (**Table 1**). Nine of the complex pedigrees are described in **Supplementary Figure 1**.

There were more males (59.6%) among the affected individuals, whereas 52.3% of the unaffected individuals were women (**Table 2**). As expected, the affected individuals had higher systolic and diastolic blood pressure and longer diabetes duration. They were also younger and had slightly higher HbA_{1c}.

Genome-Wide Linkage Analysis

We performed the parametric two-point linkage analysis for the 177 pedigrees and exome chip marker data using PSEUDOMARKER software. The analysis resulted in genome-wide significant linkage peaks (LOD-score >3.3) on 2q24.3, 2q37.2, 4q21-22, 4q25, 6p21-22, 20q13.2 and 22q12.1 (**Figure 1, Table 3, Supplementary Figure 2**). The highest linkage peak occurred at the chromosome region 6p21–6p22, located at 22.7–45.2Mb, with LOD-score up to 7.33 and altogether 165 markers on the region exceeding the $\text{LOD} \geq 3.3$ (**Supplementary Figure 3**). The multiple linkage peaks on chromosomes 2 and 4 likely originate from separate genetic signals due to long distance between the regions with significant LOD-scores. The linkage analysis was also run for imputed genotypic data for each significant linkage peak (except HLA-region), starting 1Mb prior to the first genome-wide significant marker and ending 1Mb after the last significant marker. The imputed marker set resulted in an increased number of markers reaching LOD-score >3.3 but did not yield higher linkage LOD-scores for the imputed regions (**Table 3**). Additionally, we identified suggestive linkage peaks with $\text{LOD} \geq 3.0$ on chromosomes 1q42.3, 11q22.1, 12p11.22, 17p13.3 and 18p11.23 (**Table 3**).

Markers with significant linkage LOD score show high genotyping quality and heterozygosity (**Supplementary Table 1**). Additional genome-wide linkage analysis with dominant inheritance showed that peaks on chromosome 4q25, 6p21-22 and 20q13.2 had genome-wide significant LOD scores in both recessive and dominant models (**Supplementary Table 2, Supplementary Figure 4**). We also performed linkage analysis with individuals with T1D and macroalbuminuria and ESRD as affected and individuals with normal AER as controls. The analysis showed that 3 of 8 significant peaks had a LOD-score ≥ 3.3 also with macroalbuminuria and ESRD as affected phenotype (**Supplementary Table 3**).

HLA Haplotype Imputation and HLA Haplotype-stratified Linkage Analysis

The HLA-region and in particular the MHC-II haplotypes are the most important genetic risk factors for T1D, and a high linkage peak on the HLA-region is well documented for T1D. Therefore, we performed a separate linkage analysis for chromosome 6 linkage region with HLA haplotype stratification, based on imputed HLA-alleles and pedigree types. Based on comparison within a subset of FinnDiane individuals with directly genotyped HLA alleles, the imputation of the MHC-II alleles was highly accurate: the prediction of HLA-DQA1 matched the genotyped allele with 99.1%, HLA-DQB1 with 98.0%, and HLA-DRB1 with 97.2% accuracy (2- or 4-digit depending on the HLA-typing accuracy). The allele predictions were combined as haplotypes, and the MHC-II haplotype risk scores for T1D were based on a chart by Erlich et al (25).

For 89.2% of the subjects at least one of the two HLA-haplotypes was either DRB1*0301-DQA1*0501-DQB1*0201 (DR3) or DRB1*0401-DQA1*0301-DQB1*0302 (DR4) (**Figure 2**). Furthermore, 68.6% of subjects had at least one DR4 haplotype, which indicates that the subjects have highly similar HLA-haplotype-based genetic risk for T1D. In the stratified analysis including only individuals with the DR4 serotype as cases or controls, the linkage remained with LOD-score 7.43 at 6p22 (33.28Mb).

Sub-pedigree-type Based Linkage Analysis

In order to find out which pedigree type was the leading contributor to the linkage signals with $\text{LOD} > 3.3$, we run the linkage analysis separately for different pedigree types. Most of the linkage signals originated from the 57 pedigrees concordant for DN, with less signal originating from the 77 discordant pedigrees or from

the 9 larger pedigrees that included at least 2 concordant individuals and 1 discordant member (**Supplementary Table 4**). As the HLA-region is the main T1D genetic risk component, the trees concordant for DN cannot distinguish between the linkage signal originating from DN and T1D. Outside the HLA-region, the DN linked regions do not overlap with markers or regions previously linked or associated with T1D (**Supplementary Table 5**).

***In silico* Replication of the Linkage Peaks**

We performed an *in silico* replication for the linkage peaks using the newest JDRF Diabetic Nephropathy Collaborative Research Initiative (DNCRI) GWAS for DN with 19,327 individuals (5). Of the regions with genome-wide significant linkage, we found significant association on *CCSER1* intron on chr4q25 (92,5Mb away) with lead SNP rs538044833 (P-value= 2.795×10^{-8}). More detailed examination indicated that the association was only found in one cohort (The Scottish Diabetes Research Network Type 1 Bioresource, N=4,689); thus, this association was not previously reported. The alternative allele of the lead SNP was present in 10 cases and 20 controls, yielding a raw odds ratio of 3.0. Furthermore, suggestive associations were found in DNCRI at chr2:163.9Mb/2q24.3 (P-value= 2.189×10^{-5}), chr2:234.8Mb/2q37.2 (P-value= 2.619×10^{-5}), chr4:88.1Mb/4q21.21 (P-value= 6.403×10^{-6}) and chr22:29.1Mb/22q12.1 (P-value= 5.709×10^{-6}) (**Figure 3, Supplementary Figure 2, Supplementary Table 6**).

***In silico* Associations with Gene Expression and DNA Methylation**

We studied whether the lead variants had expression or methylation quantitative trait loci (eQTL; mQTL) associations within the flanking region. Of the 11 non-chromosome 6 variants with $\text{LOD} \geq 3.3$, there were three nominal eQTL associations for either glomerular or tubular allele specific expression in the NephQTL database: rs10033307-*BMP2K* (Glomerulus, P-value: 0.022), rs10014992-*CCSER1* (Tubule, P-value: 0.011) and rs200655-*TSHZ2* (Tubule, P-value: 0.043). Furthermore, five SNPs showed significant eQTL associations in various tissues in the GTEx database (**Supplementary Table 7**). In the mQTL database, three DN-linked variants at chr4:85.6-91.5Mb showed significant mQTL associations with CpG islands in blood (**Supplementary Table 8**).

Discussion

Here, we performed a linkage analysis for DN in 177 computationally inferred and manually curated pedigrees using dense genome-wide genotyping chips combined with analytical methods that allow analysis across a combination of multiple types of pedigrees. To our knowledge, this is the first genetic linkage study for DN or any diabetic complication performed with a dense marker set. We found evidence of genetic linkage for DN on multiple novel linked regions on 2q24.3, 2q37.2, 4q21-22, 4q25, 6p21-22, 20q13.2 and 22q12.1.

Recent GWASs have yielded only few susceptibility loci for DN and replication in other populations have had limited success. Out of the newly identified linkage peaks, only few regions overlap with these GWAS loci. This is not surprising as the pedigree-based linkage studies are expected to find linked regions within families. For example, the best-known susceptibility variants for breast cancer in *BRAC1* and *BRAC2* genes were originally found with linkage studies, long before GWASs with sample sizes of tens or hundreds of thousands of individuals were possible (26). Regarding diabetic complications, our recent *CACNB2* association with diabetic retinopathy was initially discovered by suggestive linkage signal at chromosome 10p12 (27).

Based on the gene function, expression in kidneys, and associated kidney diseases, we identified plausible DN genes located under, or close to these linkage peaks (**Table 3**). Aside from the linkage peak at the HLA-region on chromosome 6, the chromosome 4q peak is the widest with markers with significant LOD-scores between 79Mb and 91Mb. It remains uncertain whether the peak occurs due to one or multiple signals, but long distances between the markers with significant LOD-scores (i.e. rs10033307 at chr4:79.3Mb, rs4129430, rs11097033 and rs1482085 at chr4:85.9–86.6Mb and rs10014992 chr4:91.6 Mb) point towards multiple separate signals (**Supplementary Figure 2 C-E**). The 91.0-92.5Mb peak had a genome-wide significant LOD score also with a dominant inheritance model (**Supplementary Table 2**). These markers with significant linkage LOD-scores occur on top of *FRAS1* (79.0-79.5Mb), *ARHGAP24* (86.4-86.9Mb) and *CCSER1* (91.0-92.5Mb), respectively. All of these genes have been associated with kidney function or diseases in previous studies (**Supplementary Table 9**).

We searched for replication of the linkage loci in a recent GWAS meta-analysis on DKD, and found a genome-wide significant association with an *CCSER1* intronic variant rs538044833 and DKD close to the 4q25 (91.5Mb) peak ($P\text{-value}=2.795\times 10^{-8}$) (**Figure 3, Supplementary Figure 2 F, Supplementary Tables 5 and 9**) (5). Another rare *CCSER1* intronic variant rs553908921 was also associated with CKD at genome-wide significant level ($P\text{-value}=2.19\times 10^{-8}$) in the recent HUNT study (28). Both associated variants were rare, and in case of rs538044833, the association was only observed in the Scottish cohort from the meta-analysis. Furthermore, eQTL data suggest that the linkage analysis lead SNP rs10014992 is associated with tubular *CCSER1* expression. While further studies are needed to establish *CCSER1* as DN and CKD associated gene, these studies strongly support rare variation in *CCSER1* as a risk factor for DN.

The *ARHGAP24* gene directly under the linkage peak at 4q22.1 (85.9–86.6Mb) has been associated with familial focal segmental glomerulosclerosis (FSGS), a rare familial kidney disease, but not with DN (**Supplementary Figure 1 D, Supplementary Table 9**) (29). Furthermore, the orthologous Q156R mutation reduces the capability of mouse *Arhgap24* to deactivate Rac1 (29). The Salem et al 2019 meta-analysis also showed a suggestive association 1.6Mb from the peak (**Supplementary Figure 1 D, Supplementary Table 6**).

Of the SNPs with significant linkage, the chromosome 4q variants account for the majority of the identified eQTL and mQTL activity (**Supplementary Table 7, Supplementary Table 8**). For example, the blood based mQTL database shows significant mQTL effect with the *ARHGAP24* intronic variant rs1482085 (LOD-score=3.39) and a CpG-island cg20784207 at birth ($P\text{-value}=3.30\times 10^{-11}$) and childhood ($P\text{-value}=3.87\times 10^{-10}$) (**Supplementary Table 8**). Although the associations were not found in the kidney tissue, possibly due to lack of large enough eQTL and mQTL databases for the kidney, blood-based associations could also be considered interesting due to the microvascular nature of DN.

Other plausible DN genes under the linkage peaks include *AGAP1* on chr2q37.2, *GRB14* at 2q24.3 and *TSHZ2* close to 20q13.2, which are associated with insulin resistance, Rac1 activation pathway and renal development (**Supplementary Table 9**). For these linkage peaks outside the chromosome 4, the Salem et al

2019 meta-analysis showed suggestive associations with DKD for variants on top or close to regions on 2q24.3, 2q37.2 and 22q12.1 (**Supplementary Table 6**) (5).

The HLA region has been identified as a leading genetic factor in some kidney diseases such as the Idiopathic membranous nephropathy (30). For DN, previous studies have yielded contradictory findings whether genes localized on the HLA region, such as *AGER* and *TNF- α* , play a significant role in the development of DN. Of note, no significant associations with DN have been found on the HLA region in GWASs.

In this study, the widest chromosome 6 linkage peak with 165 genome-wide significant markers ranges from 22.8Mb to 45.1Mb. The region with 13 markers with LOD-score >5 occurred between 33-35Mb (highest with LOD 7.3 on 33.5Mb). *TNF- α* , *AGER*, as well as the MHC-II genes, are located at 31.5-33Mb. There were also several signals throughout the region including a peak at 25Mb, which occurs on top of the *LRRC16A* gene, which has been associated with gout (31). Several studies including our previous linkage analysis showed a linkage peak at chromosome 6p21 (8). As this linkage peak occurred on the HLA region and the concordant and discordant sibling pairs showed similar haplotype inheritance pattern, it was presumed that the peak originated from the genetic risk of T1D, and not from DN. We HLA-stratified the pedigrees by assigning the affected/unaffected phenotypes for individuals with high T1D risk HLA DRB1-0401-DQA1-0301-DQB1-0302 haplotype. This analysis showed an increased LOD-score in the HLA region, when compared to the main linkage analysis. We further studied the origin of the signal by running the linkage analyses separately for different pedigree types (concordant individuals only, discordant individuals and complex trees). These analyses showed that the pedigrees including only concordant affected individuals were responsible for the majority of the LOD-score (**Supplementary Table 4**). Due to the superior power of affected concordant individuals only analysis compared to discordant individuals analysis, and relatively small amount of complex pedigrees, this may not be surprising. However, as concordant individuals only analysis cannot distinguish between DN and T1D, the origin of the peak remains ambiguous.

In general, the majority of the LOD-scores for the other DN-linked markers originated from the pedigrees including only concordant individuals. However, outside the HLA-region, most of these peaks did not co-occur

with previously found T1D-associated variants or linked regions (**Supplementary Table 5**). As the genetic studies of T1D have been more powerful compared to DN studies due to larger sample sizes, we do not expect the non-overlapping linked regions to be novel T1D regions.

Of note, the HLA-allele imputations performed for the haplotype stratified linkage analysis were significantly more accurate than the previous HLA-allele predictions using Illumina 610 K genotyping chip data for 161 Finnish individuals (32). With the SNP2HLA and HLA*IMP –software and HapMap2 reference panel they achieved <20% imputation accuracy for HLA-DRB1-gene, compared to 99.1% for *HLA-DQA1*, 98.0% for *HLA-DQB1* and 97.2% for *HLA-DRB1* in our current study (32). We speculate that the main cause for the drastic difference in imputation accuracy between the studies could be due to the different reference panels and different genotyping platform used in the studies; i.e. Illumina CoreExome chip includes more exomic variants compared to most other genotyping chips, which may also improve the imputation quality for the MHC-II alleles.

Altogether five suggestive linkage peaks had linkage LOD-scores between 3 and 3.3 (**Table 3**). The chromosome 1q42.3 peak with LOD-score of 3.2 is especially notable as it is located between *NID1* and *GPR137B*. Previous studies have suggestively associated *NID1* with Goodpasture's syndrome, also called glomerular basement membrane disease. Nidogen-1, encoded by *NID1*, is able to bind to other basement membrane proteins such as type IV collagen (33). Rare variants in the *COL4A3* gene encoding the $\alpha 3$ subunit of type IV collagen have been associated with Alport syndrome, and recently, also a common *COL4A3* variant was associated with DN on large GWAS done with 19,406 individuals with T1D (5; 34). This is one example, in which a serious, usually monogenic disease is caused by a deleterious rare variant, whereas the risk of a complex disease affecting the same tissue type or organ is affected by common variants located in the same gene.

Previous studies have found linked regions on multiple chromosomes, including the most recent findings on chromosomes 3q21-25 and 22q11 (7; 8). Our linkage analysis did not result in significant linkage peaks with these previously reported ones (**Supplementary Table 10, Supplementary Figure 2 C, H**). Nevertheless, the

former 3q21-25 region contained three markers with linkage LOD-score >2.5 . Likewise, a suggestive linkage signal with LOD-score of 2.80 was found at 22q11, which could indicate lack of power to detect a genome-wide signal in this region rather than lack of replication. Of the 440 individuals in the pedigrees in this study, 175 were also included in our previous study, but most of the previous linkage studies were conducted using sparse microsatellite platforms, and the lower linkage LOD-scores could be due to platform differences between the SNP microarray and microsatellites (35).

The study included only a limited amount of pedigrees, and a limited number of individuals within the pedigrees. While this has been a common limitation of pedigree-based studies, this study has benefited from the development of modern analysis methods that enable studying larger variety of pedigree types together. The supporting association evidence for chromosome 4q peak is also based on rare variants, and they could be invalidated as more data becomes available. However, finding regions with low frequency and rare variants with large effect sizes are main advantage of the linkage studies compared to GWASs. We also note that the validity of eQTL associations have met increasing criticism (36). As genotyping and sequencing becomes more affordable, and with the rise of large nationwide genotyping and sequencing initiatives, studies with larger families, including multiple siblings, cousins etc. will become possible.

As the Finns have well-documented differences in low-frequency and rare variant patterns compared to other European populations, linkage peaks in these Finnish pedigrees do not necessarily exist in pedigrees in other countries. Therefore, replication in other Finnish families would be ideal. However, the FinnDiane study already covers a substantial proportion of the Finnish individuals with T1D and there is no other T1D study cohort with pedigrees and GWAS data available in Finland without significant overlap with the FinnDiane study cohort. As only a few individuals included in this study have participated in our WES and WGS studies, it is currently not statistically feasible to study variants under the linkage peaks using our sequencing data. However, as more individuals will be sequenced, we expect to be able to pinpoint the variants behind these signals in our future studies.

Conclusions

Our study aimed to identify genetic loci linked to DN in Finnish families including individuals with T1D. The results include eight regions with a genome-wide significant linkage signal. The analysis also resulted in peaks on top or close to genes, such as *ARHGAP24*, that have previously been associated with kidney diseases or kidney function, but not yet with DN, and a peak on top of *CCSER1*, where rare variants in two studies have been associated with DN and CKD. Although the most conspicuous peak occurred at the HLA-region of chromosome 6, it is likely to originate at least in part from the T1D risk. To pinpoint and confirm linked genes, more studies are needed.

Acknowledgements

The skilled technical assistance of M. Parkkonen, M. Korolainen, A.-R. Salonen, A. Sandelin and J. Tuomikangas (Folkhälsan Research Center, Finland) are gratefully acknowledged. The authors also acknowledge all the physicians and nurses at each center taking part in the enrolment and clinical characterization of the participants (see **Supplementary Table 11** for a list of study centres and investigators involved in the FinnDiane study).

Guarantors

The first author and Jani Haukka and the Principal Investigator Professor Per-Henrik Groop are the guarantors of this work. They full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

This study was supported by funding from the JDRF (Ref. 1SRA-2016-333-M-R), Folkhälsan Research Foundation, Wilhelm and Else Stockmann Foundation, Liv och Hälsa Society, Helsinki University Hospital Research Funds (EVO TYH2018207), Academy of Finland (275614, 299200, and 316664), Novo Nordisk Foundation (NNF OC0013659), and European Foundation for the Study of Diabetes (EFSD) Young Investigator Research Award funds.

Author contributions

J.H. performed the analyses and wrote the manuscript, N.S. supervised analyses and edited the manuscript, C.F. and V.H. helped in constructing pedigrees with distant relatives, E.V. performed quality controls for the genotyping data, J.C. helped with the meta-analysis's replication data, P-H.G. guided the analyses and edited the manuscript.

Data availability statement

The datasets generated and/or analyzed during the current study are not publicly available as the patients' written consent does not allow data sharing. Data are locally available from the corresponding author on reasonable request.

Disclosures

P-HG has served on advisory boards for AbbVie, Astellas, AstraZeneca, Bayer, Boehringer Ingelheim, Cebix, Eli Lilly, Janssen, Medscape, MSD, Mundipharma, Novartis, Novo Nordisk, Sanofi, and has received lecture honoraria from Astellas, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Elo Water, Genzyme, Medscape, MSD, Mundipharma, Novartis, Novo Nordisk and Sanofi. P-H G has also received investigator-initiated grants from Eli Lilly and Roche.

References

1. Forbes JM, Cooper ME: Mechanisms of diabetic complications. 2013;93:137-188
2. Freund MK, Burch K, Shi H, Mancuso N, Kichaev G, Garske KM, Pan DZ, Pajukanta P, Pasaniuc B, Arboleda VA: Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. 2018:324558
3. Sandholm N, Van Zuydam N, Ahlqvist E, Juliusdottir T, Deshmukh HA, Rayner NW, Di Camillo B, Forsblom C, Fadista J, Ziemek DJ: The genetic landscape of renal complications in type 1 diabetes. 2017;28:557-574
4. Dahlström E, Sandholm N: Progress in defining the genetic basis of diabetic complications. *Current diabetes reports* 2017;17:80
5. Salem RM, Todd JN, Sandholm N, Cole JB, Chen W-M, Andrews D, Pezzolesi MG, McKeigue PM, Hiraki LT, Qiu C: Genome-wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. *Journal of the American Society of Nephrology* 2019;30:2000-2016
6. Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, Esko T, Mägi R, Inouye M, Lappalainen T: Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS genetics* 2014;10:e1004494
7. Rogus JJ, Poznik GD, Pezzolesi MG, Smiles AM, Dunn J, Walker W, Wanic K, Moczulski D, Canani L, Araki S: High-density single nucleotide polymorphism genome-wide linkage scan for susceptibility genes for diabetic nephropathy in type 1 diabetes: discordant sibpair approach. *Diabetes* 2008;57:2519-2526
8. Wessman M, Forsblom C, Kaunisto MA, Söderlund J, Ilonen J, Sallinen R, Hiekkalinna T, Parkkonen M, Maxwell AP, Tarnow L: Novel susceptibility locus at 22q11 for diabetic nephropathy in type 1 diabetes. 2011;6:e24053
9. Syreeni A, Sandholm N, Cao J, Toppila I, Maahs DM, Rewers MJ, Snell-Bergeon JK, Costacou T, Orchard TJ, Caramori ML: Genetic determinants of glycated hemoglobin in type 1 diabetes. *Diabetes* 2019;68:858-867
10. Consortium GP: A global reference for human genetic variation. *Nature* 2015;526:68
11. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue MJ: Next-generation genotype imputation service and methods. 2016;48:1284
12. Delaneau O, Marchini J, Zagury J-F: A linear complexity phasing method for thousands of genomes. *Nature methods* 2012;9:179
13. Syreeni A, El-Osta A, Forsblom C, Sandholm N, Parkkonen M, Tarnow L, Parving H-H, McKnight AJ, Maxwell AP, Cooper ME: Genetic examination of SETD7 and SUV39H1/H2 methyltransferases and the risk of diabetes complications in patients with type 1 diabetes. *Diabetes* 2011;60:3073-3080
14. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M: Robust relationship inference in genome-wide association studies. 2010;26:2867-2873
15. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE, Genetics UoW: PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. 2014;95:553-564
16. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. 1998;63:259-266
17. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ: PLINK: a tool set for whole-genome association and population-based linkage analyses. 2007;81:559-575
18. Gertz EM, Hiekkalinna T, Le Digabel S, Audet C, Terwilliger JD, Schäffer AA: PSEUDOMARKER 2.0: efficient computation of likelihoods using NOMAD. 2014;15:47
19. Hiekkalinna T, Schäffer AA, Lambert B, Norrgrann P, Göring HH, Terwilliger JD: PSEUDOMARKER: a powerful program for joint linkage and/or linkage disequilibrium analysis on mixtures of singletons and related individuals. 2011;71:256-266
20. Söderlund J, Forsblom C, Ilonen J, Thorn L, Wadén J, Parkkonen M, Groop P-H, Diabetologia FSGJ: HLA class II is a factor in cardiovascular morbidity and mortality rates in patients with type 1 diabetes. 2012;55:2963-2969
21. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, Raychaudhuri S, de Bakker PI: Imputing amino acid polymorphisms in human leukocyte antigens. 2013;8:e64683

22. Gillies CE, Putler R, Menon R, Otto E, Yasutake K, Nair V, Hoover P, Lieb D, Li S, Eddy S: An eQTL landscape of kidney tissue in human nephrotic syndrome. *The American Journal of Human Genetics* 2018;103:232-244
23. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K: Systematic identification of genetic influences on methylation across the human life course. *Genome biology* 2016;17:61
24. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N: The genotype-tissue expression (GTEx) project. *Nature genetics* 2013;45:580
25. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear ALJD: HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. 2008;
26. Ford D, Easton D, Stratton M, Narod S, Goldgar D, Devilee P, Bishop D, Weber B, Lenoir G, Chang-Claude J: Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics* 1998;62:676-689
27. Vuori N, Sandholm N, Kumar A, Hietala K, Syreeni A, Forsblom C, Juuti-Uusitalo K, Skottman H, Imamura M, Maeda S: CACNB2 is a novel susceptibility gene for diabetic retinopathy in type 1 diabetes. *Diabetes* 2019;68:2165-2174
28. Graham SE, Nielsen JB, Zawistowski M, Zhou W, Fritsche LG, Gabrielsen ME, Skogholt AH, Surakka I, Hornsby WE, Fermin D: Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nature communications* 2019;10:1-9
29. Akilesh S, Suleiman H, Yu H, Stander MC, Lavin P, Gbadegesin R, Antignac C, Pollak M, Kopp JB, Winn MP: Arhgap24 inactivates Rac1 in mouse podocytes, and a mutant form is associated with familial focal segmental glomerulosclerosis. *The Journal of clinical investigation* 2011;121:4127-4137
30. Stanescu HC, Arcos-Burgos M, Medlar A, Bockenhauer D, Kottgen A, Dragomirescu L, Voinescu C, Patel N, Pearce K, Hubank MJNEJoM: Risk HLA-DQA1 and PLA2R1 alleles in idiopathic membranous nephropathy. 2011;364:616-626
31. Sakiyama M, Matsuo H, Shimizu S, Chiba T, Nakayama A, Takada Y, Nakamura T, Takada T, Morita E, Naito M: A common variant of leucine-rich repeat-containing 16A (LRRC16A) gene is associated with gout susceptibility. *Human cell* 2014;27:1-4
32. Vlachopoulou E, Lahtela E, Wennerström A, Havulinna A, Salo P, Perola M, Salomaa V, Nieminen M, Sinisalo J, Lokki ML: Evaluation of HLA-DRB1 imputation using a Finnish dataset. *Tissue Antigens* 2014;83:350-355
33. Chew C, Lennon R: Basement membrane defects in genetic kidney diseases. *Frontiers in pediatrics* 2018;6:11
34. Kashtan CE, Ding J, Garosi G, Heidet L, Massella L, Nakanishi K, Nozu K, Renieri A, Rheault M, Wang F: Alport syndrome: a unified classification of genetic disorders of collagen IV α 345: a position paper of the Alport Syndrome Classification Working Group. *Kidney international* 2018;93:1045-1051
35. Osterholm A-M, He B, Pitkaniemi J, Albinsson L, Berg T, Sarti C, Tuomilehto J, Tryggvason K: Genome-wide scan for type 1 diabetic nephropathy in the Finnish population reveals suggestive linkage to a single locus on chromosome 3q. *Kidney international* 2007;71:140-145
36. Huang QQ, Ritchie SC, Brozynska M, Inouye M: Power, false discovery rate and Winner's Curse in eQTL studies. *Nucleic acids research* 2018;46:e133-e133
37. Gondoin A, Hampe C, Eudes R, Fayolle C, Pierre-Eugène C, Miteva M, Villoutreix BO, Charnay-Pouget F, Aitken DJ, Issad T: Identification of insulin-sensitizing molecules acting by disrupting the interaction between the Insulin Receptor and Grb14. *Scientific reports* 2017;7:16901
38. McGregor L, Makela V, Darling SM, Vrontou S, Chalepakis G, Roberts C, Smart N, Rutland P, Prescott N, Hopkins J: Fraser syndrome and mouse blebbed phenotype caused by mutations in FRAS1/Fras1 encoding a putative extracellular matrix protein. *Nature genetics* 2003;34:203
39. Pitera JE, Scambler PJ, Woolf AS: Fras1, a basement membrane-associated protein mutated in Fraser syndrome, mediates both the initiation of the mammalian kidney and the integrity of renal glomeruli. *Human molecular genetics* 2008;17:3953-3964

40. Jenkins D, Caubit X, Dimovski A, Matevska N, Lye CM, Cabuk F, Gucev Z, Tasic V, Fasano L, Woolf AS: Analysis of TSHZ2 and TSHZ3 genes in congenital pelvi-ureteric junction obstruction. *Nephrology Dialysis Transplantation* 2009;25:54-60
41. Funk SD, Lin M-H, Miner JH: Alport syndrome and Pierson syndrome: diseases of the glomerular basement membrane. *Matrix Biology* 2018;71:250-261

Table 1 – Number of trees and individuals in the pedigrees included in the linkage analysis

Pedigrees	N (trees)	Affected	Unaffected	Unknown
Concordant sibling pairs	40	80	0	0
Discordant sibling pairs	55	55	55	0
Sibships	9	14	13	5
Parent-offspring pairs	15	20	9	0
Cousin pairs	7	10	4	0
Avuncular pairs	8	12	3	0
Larger trees	43	64	44	52

Affected: Number of affected individuals with DN; Unaffected: number of individuals with normal AER;
Unknown: Number of individuals with unknown kidney disease status.

Table 2: Clinical characteristics of the affected and unaffected individuals in pedigrees

	Affected (n=255)	Unaffected (n=128)	<i>p</i>
Women (%)	40.4	52.3	<0.0001
Age (years)	59.5 ± 11.9	57.3 ± 11.0	ns
BMI (kg·m ⁻²)	26.4 ± 4.6	25.8 ± 3.8	ns
Age of diabetes onset (years)	12.2 ± 7.5	16.0 ± 9.7	<0.004
Diabetes duration (years)	36.9 ± 10.8	30.5 ± 11.4	<0.0001
HbA _{1c} (%)	8.74 ± 1.67	8.23 ± 1.23	<0.05
Systolic BP (mmHg)	142 ± 21	131 ± 17	<0.0001
Diastolic BP (mmHg)	81 ± 12	77 ± 8	<0.05
Albumin excretion rate (mg)	7.99 [4.56-11.955]	157.29 [39.2-955]	<0.0001
eGFR (mL·min ⁻¹ ·1.73m ⁻²)	51 ± 19	97 ± 20	<0.0001
Micro/Macro/ESRD (n)	78/85/92	-	-

Albumin excretion rate (AER) is presented as median and interquartile range

Table 3: Genome-wide significant and suggestive linkage peaks for DN

Region	Linkage LOD-score (imputed*)	N markers with LOD-score ≥ 3.3 (Imputed*)	Noteworthy nearby genes
Genome-wide significant linkage peaks			
2q24.3	3.31 (3.4)	1 (5)	<i>GRB14</i> - Insulin resistance, fat distribution, BMI (37)
2q37.2	3.63 (3.7)	1 (20)	<i>AGAP1</i>
4q21.21	3.62	1	<i>FRAS1</i> - familial Fraser syndrome, kidney function and development (38; 39)
4q22	3.80 (3.9)	2 (93)	<i>ARHGAP24</i> – FSGS (29), <i>PTPN13</i> – Diabetic nephropathy (3)
4q25	3.41 (3.6)	1 (41)	<i>CCSER1</i>
6p21-22	7.42	165	HLA-region
20q13.2	3.58 (3.4)	2 (2)	<i>TSHZ2</i> - renal pelvis development in human and mice (40)
22q12.1	3.63 (5.1)	2 (8)	
Suggestive linkage peaks			
1q42.3	3.25	1	<i>NID1</i> – encodes GBM protein nidogen, plausible associations to GBM diseases (41)
11q22.1	3.1	1	
12p11.22	3.28	3	
17p13.3	3.29	1	
18p11.23	3.16	1	

*Linkage LOD-scores and number of markers with LOD-score ≥ 3.3 are given for the directly genotyped data, and in parentheses, for imputed data.

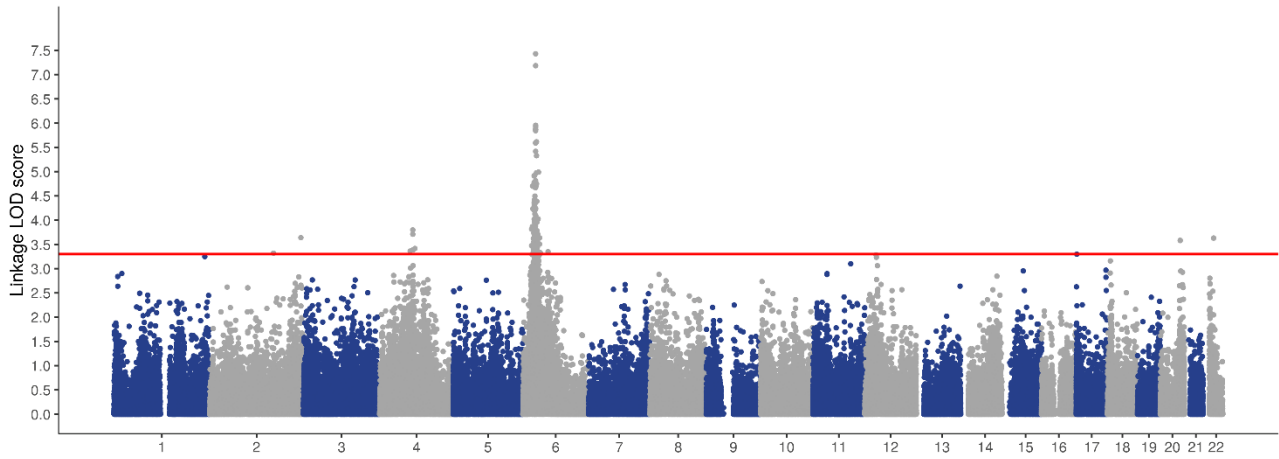


Figure 1 - Manhattan plot of genome-wide linkage analysis. The plot shows results of genome-wide linkage analysis LOD scores. The analysis showed linkage peaks with genome-wide significant LOD scores > 3.3 on chromosomes 2q24.3, 2q37.2, 4q21 - 22, 4q25, 6p21-22, 20q13.2 and 22q12.1. The highest and widest peak occurs on chromosome 6p HLA region.

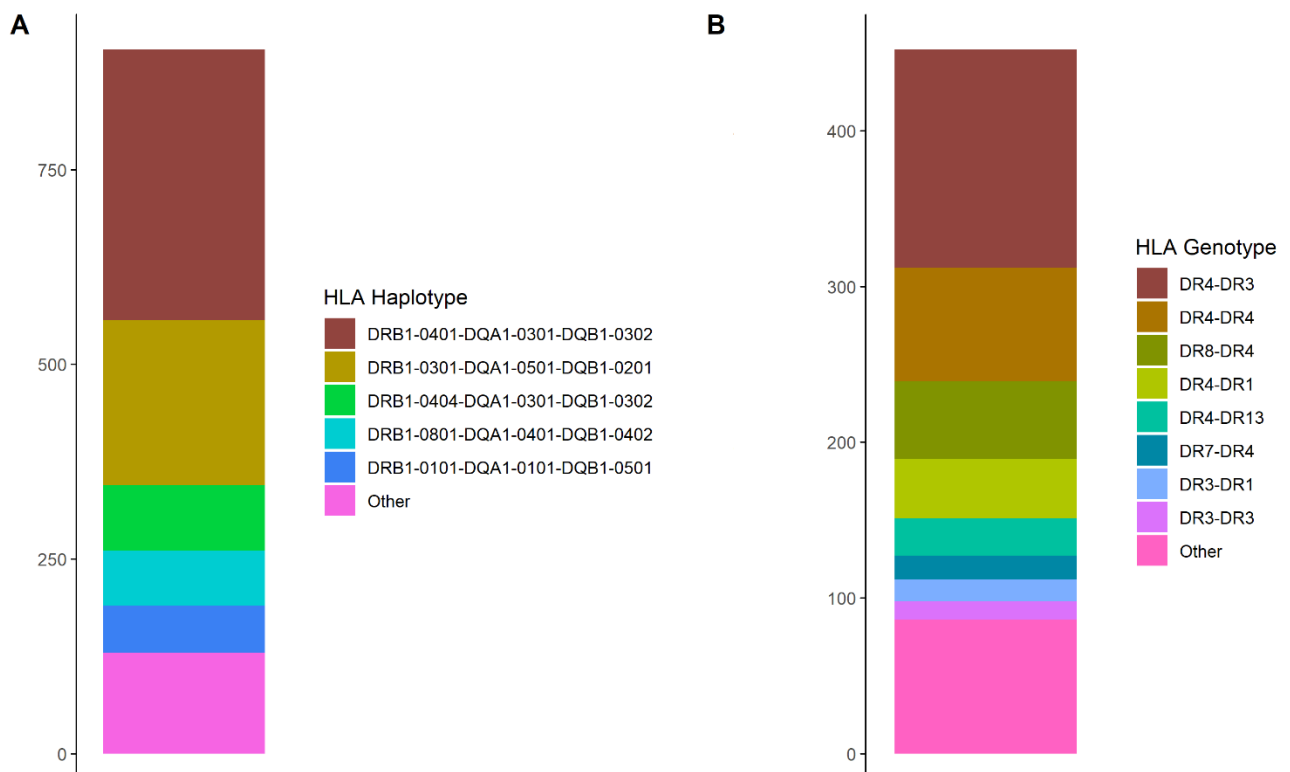


Figure 2 - Imputed HLA haplotype and genotype frequencies for patients in pedigrees. The HLA alleles were imputed using SNP2HLA software and T1DGC reference panel, and then combined into HLA-haplotypes and genotypes. **A** - shows that majority of haplotypes were high T1D risk DR3 or DR4 haplotypes. **B** - Moreover, more than 80% of individuals has either one or two DR4, which further highlights the homogenic genetic risk for T1D on this cohort.

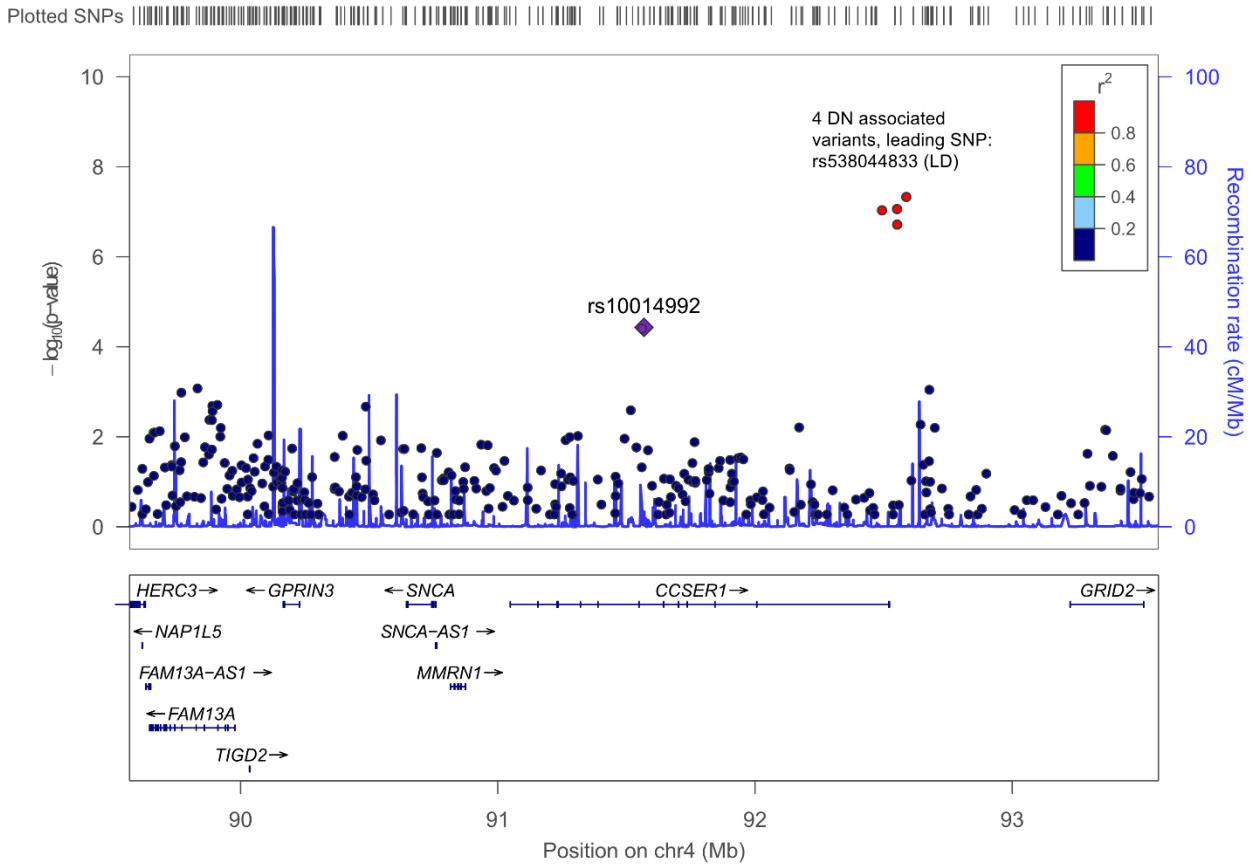


Figure 3 - LocusZoom regional association plot of 4q25, showing the significant linkage and association signals. The y-axis indicates the $-\log_{10}(p\text{-value})$, instead of LOD scores, for both linkage and association results. The purple diamonds show DN linkage P-value for the lead SNP in our current linkage study, whereas the red circles with rs-numbers display association P-values from Salem et al. 2019. The remaining SNPs are coloured according to their r^2 correlation with the lead SNP.