Genetic basis of diabetic kidney disease and other diabetic complications

Current Opinion in Genetics and Development, Issue 50, Molecular and genetic basis of disease

Niina Sandholm\textsuperscript{a,b,c} and Per-Henrik Groop\textsuperscript{a,b,c,d}

a. Folkhälsan Institute of Genetics, Folkhälsan Research Center, 00290 Helsinki, Finland
b. Abdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, 00290 Helsinki, Finland
c. Research Programs Unit, Diabetes and Obesity, University of Helsinki, 00290 Helsinki, Finland
d. Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia

Corresponding author:
Niina Sandholm, DSc (Tech)
FinnDiane Study Group
Folkhälsan Institute of Genetics, Folkhälsan Research Center
Postal Address:
Biomedicum C329b
Haartmaninkatu 8
00290 Helsinki
Finland
Tel. +358 9191 25695
Fax +358 9191 25382
niina.sandholm@helsinki.fi

Word count: 2476 words

Figures: 1

Boxes: 2

Abstract (169 words)
Diabetic kidney disease and other long-term complications are common in diabetes, and comprise the main cause of co-morbidity and premature mortality in individuals with diabetes. While familial clustering and heritability have been reported for all diabetic complications, the genetic background and the molecular mechanisms remain poorly understood. In recent years, genome-wide association studies have identified a few susceptibility loci for the renal complications as well as for diabetic retinopathy, diabetic cardiovascular disease and mortality. As for many complex diseases, the genetic factors increase the risk of complications in concert with the environment, and certain associations seem specific for particular conditions, e.g. \textit{SP3-CDCA7} associated with end-stage renal disease only in women, or \textit{MGMT} and variants on chromosome 5q13 associated with cardiovascular mortality only under tight glycaemic control. The characterization of the phenotypes is one of the main challenges for genetic research on diabetic complications, in addition to an urgent need to increase the number of individuals with diabetes with high quality phenotypic data to be included in future genetic studies.

Key words: Diabetic kidney disease; diabetic nephropathy; genome-wide association study; GWAS; end stage renal disease; whole exome sequencing; diabetic retinopathy; cardiovascular disease
Diabetes is a true epidemic with 415 million adults with diabetes worldwide\[1\]. One third of these patients develop severe microvascular complications such as diabetic kidney disease (DKD), sight threatening diabetic retinopathy (DR), and diabetic neuropathy. Furthermore, individuals with diabetes carry an increased risk of cardiovascular disease (CVD), a risk that is particularly high in those with DKD. Consequently, those that develop end stage renal disease (ESRD) requiring dialysis or kidney transplant for survival have 18 times higher premature mortality compared with the general population [2].

Although there is strong evidence for a genetic influence on the development of diabetic complications, environmental factors such as exposure to high blood glucose also contribute, and the disease outcomes are likely a complex interplay between genetics, epigenetic gene regulation, and environment. In this review, we will highlight recent large-scale genetic studies on various diabetic complications. Apart from improving our understanding of the molecular mechanisms behind the diabetic complications, the genetic findings also hold promise of identifying novel biomarkers for earlier identification of patients at risk, and novel therapeutic target molecules.

**Diabetic kidney disease**

Diabetic kidney disease (DKD) clusters in families [3,4]. While the 25-year cumulative incidence of DKD was 25% in diabetic siblings of probands without DKD, the risk was 43% and 58% in siblings of probands with DKD or ESRD, respectively, resulting in a more than twofold sibling risk ratio for DKD [5]. Recently, by using genome-wide genotyping data of unrelated individuals, the narrow sense heritability of DKD was estimated to be 35% [6].

**Genome-wide association studies on DKD**

Multiple candidate genes have been studied for DKD, but robust replication has been challenging [7,8]. One of the first genome-wide association studies (GWAS) on DKD suggested association at rs10868025 in the FRMD3 gene (p=5.0×10^{-7})[9], and the finding was supported by many [9-11], although not all subsequent studies [6,8]. The first genome-wide significant findings for ESRD were reported from a GWAS meta-analysis from the GEnetics of Nephropathy – an International Effort (GENIE) consortium including 6691 individuals with T1D from three discovery cohorts, and up to 11,847 individuals in the joint meta-analysis including the replication cohorts. Variants in the AFF3 (rs7583877, p=1.2×10^{-8}) and on the RGMA – MCTP2 gene region (rs12437854, p=2.0×10^{-9}) were associated with ESRD (Figure 1), while suggestive evidence of association with DKD was found for variants in the ERBB4 gene (rs7588550 p=2.1×10^{-7}) [12].

A gender-stratified GWAS identified variants between the SP3 and CDCA7 genes to be associated with ESRD in women with T1D, and the finding replicated in other GENIE cohorts (rs4972593 p=3.9×10^{-6})[13]. Intriguingly, the nearby SP3 shows higher expression in female glomeruli (top 0.3%), and Sp3 transcription factor directly binds to estrogen receptor α, providing a plausible explanation for the gender-specific association.

Urinary albumin excretion rate (AER) is one of the early signs of renal complications in diabetes, and commonly used to diagnose and classify DKD. A GWAS on AER in 1925 individuals with T1D from the Finnish Diabetic Nephropathy Study (FinnDiane) found variants in the GLRA3 gene associated with AER (rs1564939 p=8.4×10^{-9}), however, replication in non-Finnish Europeans showed a trend in the opposite direction (p=0.03), suggesting a population specific effect[14]. Replication in further Finnish individuals is ongoing to confirm the finding.

A trans-ethnic GWAS meta-analysis with 6197 individuals with T2D identified rs12523822 on the SCAF/CNKSR3 locus associated with DKD with a particularly strong effect seen in American Indians (p=5.7×10^{-9}, odds ratio (OR) = 0.57). Furthermore, variants in the MYH9 locus were near genome-wide significant in the African American group, likely due to a large proportion of the individuals with T2D with
co-occurring non-diabetic renal disease, as MYH9 is one of the main susceptibility loci for ESRD in the general population with African ancestry [15].

GWAS on estimated glomerular filtration rate (eGFR) in 133,413 individuals from the general population identified 53 loci for eGFR. The rs12917707 in the UMOD gene was significantly associated with eGFR also in the subset of 16,477 individuals with diabetes (mainly T2D). Nominal association (p<0.05) was found in the subset of individuals with diabetes for 19 of the lead loci [16]. Whereas nearly all individuals with T1D and DKD show histological findings characteristic of diabetic nephropathy, in T2D only 30-50% of individuals with albuminuria have true DKD, while the rest may have kidney disease related to hypertension, overweight or ageing [17]. Thus, the genetic factors affecting eGFR in the general population may also contribute to the renal complications in individuals with T2D, while little overlap is found in T1D.

A recent GWAS in 5156 individuals with T1D from the SUrrogate markers for Micro- and Macro-vascular hard endpoints for Innovative diabetes Tools (SUMMIT) consortium utilised multiple phenotypic definitions based on various thresholds of either AER, eGFR, or both [6]. No locus reached genome-wide significance after joint analysis including 12,540 individuals, but variants in the previously reported AFF3 were among the lead SNPs; of note, there was a substantial overlap between this and the previous studies. Variants associated with T2D and body mass index (BMI) were associated with DKD, suggesting that metabolic changes leading to T2D, and BMI, are causal risk factors for DKD. Furthermore, there was an inverse genome-wide correlation between DKD and smoking cessation.

For many complex diseases, the number of loci has increased after including some tens of thousands of individuals. Larger GWAS meta-analyses on DKD are underway e.g. in the JDRF – DNCRI (JN Todd, abstract in Diabetes 2016, 65: A100), and in the SUMMIT consortium (NW Rayner, abstract in Diabetologia 2015, 58: S200).

Taking advantage of genetic data, a causal link was reported between BMI and DKD in a Mendelian randomization study, which utilises genetic information to infer causality between a biomarker and an outcome [18]. Similar studies reported kidney injury molecule 1 (KIM-1) as a likely causal risk factor for reduced eGFR [19], while no evidence of causal link was found for serum uric acid [20].

Whole exome sequencing for DKD

While the GWASs are best suited for detecting common genetic variants, many of the low frequency and rare variants can only be captured with sequencing based methods such as whole exome and whole genome sequencing (WES and WGS). While WES only covers the protein coding regions of the genome, it is more affordable than WGS; furthermore, as the non-synonymous exon variants may directly affect the protein function, their effect sizes, and thus the statistical power, may be markedly higher than for intronic or intergenic variants affecting regulation of gene expression, making WES an attractive research strategy. Nevertheless, findings from WES and WGS on DKD remain sparse. In WES including 997 T1D individuals from the SUMMIT consortium, a rare intronic variant rs188427269 in the NVL gene reached exome-wide significance (p<5×10^-7) for association with ESRD, but no coding variant reached this threshold and further replication is required to validate this finding [6]. Nevertheless, preliminary results are also emerging from WGS in siblings with and without DKD (Öhman MK, abstract in J Am Soc Nephrol 2016, 27:429A).
Diabetic Retinopathy (DR)

Also, DR clusters in families [21], and heritability estimates range from 25 to 52% [22,23]. In a meta-analysis of candidate genes, only variants in the aldose reductase AKR1B1 gene were associated with DR ($p=1\times10^{-4}$) [24]. Recently, in a thiamine (vitamin B1) transporter candidate gene study, variants in SLC19A3 were associated with reduced risk of DR, and the association reached genome-wide significance for the combined microvascular endpoint (DR and ESRD) in meta-analysis with replication samples ($rs=7.1\times10^{-9}$) [25].

Multiple GWASs have been performed on DR in various ethnicities [26-31], but the studies are limited in size (mainly few hundreds of individuals) and only one GWAS meta-analysis exists with 2829 individuals from two studies [27]. Since many of the studies have only reported suggestive signals with no replication, subsequent studies have attempted to replicate the loci but with limited success. Nominally significant replication ($p<0.05$) was obtained e.g. for variants in ARHGAP22 and PLXDC2 that originally showed significant genome-wide association with DR in Taiwanese individuals with T2D [28], but no replication was achieved in other studies [31,34].

GWAS on DR in 844 Australian individuals with T2D found rs9896052 near GBR2 gene to be suggestively associated with DR ($p=6.6\times10^{-5}$), and this finding was replicated in people with T1D, T2D, and in an Indian cohort of individuals with T2D, resulting in genome-wide significance in meta-analysis ($p=4.2\times10^{-8}$). In human histological samples, the GBR2 protein was expressed throughout the retina, and GBR2 expression was increased in a mouse model of DR [31].

WES including 107 Saudi-Arabian individuals with T2D found enrichment of rare variants in those without DR in the NME3, LOC728699, and FASTK genes ($p<5\times10^{-6}$) [35]. Another WES including 70 individuals of mixed ethnicity and mainly T2D, suggested six genes with rare variants in DR cases and reduced gene expression in human retinal endothelial cells, when cultured under high glucose [36]. However, findings from both studies require replication for validation of the association.

Box 1: Research gaps for genetics of DKD

- Need of larger patient cohorts for genetic studies in order to increase power and likelihood of finding susceptibility loci
- Better phenotypic characterization of the renal complications, preferably by promoting collection of renal biopsies, since currently used biomarkers eGFR and AER can serve at the most as intermediate phenotypes and do not provide a fully accurate description of the renal pathology
- Gene expression and eQTL data for cell-specific kidney tissue are limited, hampering functional annotation of identified variants

Box 2: Research gaps in genetics of DR

- Published GWASs are small and only one GWAS meta-analysis exists with 2829 individuals [27]
- Scoring of fundus photographs is relatively straightforward but demanding, and access to carefully scored high quality fundus photographs is limiting the study population.
Neuropathy
Diabetic neuropathies cover a range of symptoms and can be divided into diabetic peripheral neuropathies including painful neuropathy, and autonomic neuropathy [37]. However, diagnosis and phenotyping for research purposes is challenging, and the genetic background remains largely unexplored soil. Literature based meta-analysis of ACE and MTHFR candidate genes in nine studies found nominal evidence of association with peripheral neuropathy for both genes [38]. Using genome-wide genotyping data, the heritability of painful neuropathy was estimated to be 11% [39]. The same GWAS study, including 3063 individuals with T2D, suggested variants near GFRA2 to confer susceptibility to painful neuropathy (rs17428041 p=1.8×10^{-7})[39]. Further stratification by gender in the same study suggested rs71647933 at the ZSCAN20-TLR12P to be associated with painful neuropathy in women (p=2.7×10^{-7}) and rs6986153 near HMGB1P46 in men (p=8.0×10^{-7})[40]. Another GWAS on painful neuropathy in individuals with T1D is underway (E Valo, abstract in Diabetologia 2017, 60: S72-S73).

Cardiovascular complications in diabetes
The risk of CVD is particularly elevated in those with DKD with 40% developing CVD by the age of 40, compared to 7% among people with T1D without DKD [41]. Nevertheless, the genetic background of CVD in diabetes has been studied surprisingly little, despite evidence of familial clustering [42] and high heritability (41%) for intima media thickness, a marker of atherosclerosis [43].

There is evidence that the susceptibility loci for CVD in the general population also affect the risk of CVD in individuals with diabetes[44,45]. On the contrary, a GWAS including 4188 individuals with T2D identified rs10911021 near the glutamate-ammonia ligase (GLUL) gene encoding a glutamine synthetase, to be associated with CVD especially in individuals with diabetes (p=2×10^{-3})[46]. Of note, the lead SNP was associated with GLUL expression in endothelial cells, and with the ratio between the glutamic acid precursor (plasma pyroglutamic acid) and glutamic acid, pointing to a potential mechanism[46].

In a GWAS of 2667 individuals with T2D from the intensive treatment arm of the ACCORD study targeting glycated hemoglobin HbA_{1c} levels of <6.0%, minor alleles at rs9299870 in MGMT and rs57922 on chromosome 5q13 were associated with three-fold risk of cardiovascular mortality [47]. Despite the specific design of the discovery study, supporting evidence for the joint effect of the two variants on cardiovascular mortality under tight glycemic control was also found in two other studies. When 65 biomarkers were evaluated in a subset of 351 individuals from the ACCORD study, rs57922 was associated with the 12-month change in the active glucagon-like peptide 1 (GLP-1) in the intensive treatment arm, with increased GLP-1 levels in the C/C homozygotes that derived benefit from intensive glycemic control [48].

While preliminary results have been presented for further GWAS meta-analyses on coronary artery disease (N Van Zuydam, abstract in Diabetologia 2013, 56:S76-77) and peripheral vascular disease in individuals with diabetes (N Van Zuydam, abstract in Diabetes 2015, 64:A15-A15), no GWASs are published for cerebrovascular disease (strokes) in diabetes.

Conclusions
GWASs have thus far identified six susceptibility loci for renal complications in diabetes (p<5×10^{-8} for AFF3, RGMA-MCTP2, and CDCA7-SP3 for ESRD; SCAF8-CNKR3 for DKD; UMOD for eGFR; and GLRA3 for albuminuria in Finnish individuals). Furthermore, variants in the GBR2 locus are associated with DR, variants near GLUL with CVD in diabetes and MGMT and 5q13 with cardiovascular mortality under tight glycemic control. However, compared with many other complex diseases, the number of identified loci remains limited.
Larger number of individuals is required to robustly identify additional loci with moderate effect sizes. Therefore, international collaboration is indispensable. On the other hand, studying well characterized homogenous populations such as Finland or Iceland may be particularly advantageous especially when searching for rare variants; while the population isolates have less variant loci in total, the existing loss-of-function and missense mutations may have markedly higher frequencies, facilitating the identification of novel loci [49].

One of the main challenges in the genetic studies for all microvascular complications is the definition of the phenotype. While multiple markers are used for characterization of the complications, e.g. eGFR and albuminuria for DKD, confirmation of diabetic nephropathy as the underlying pathology would require renal biopsies, which are rarely taken from individuals with diabetes.

As for all complex diseases, the genes do not alone cause disease, but increase the risk in concert with environmental factors. It can be challenging to disentangle the genetic factors under such complex environmental influence, and there is evidence that certain genetic effects are only observed under particular conditions, e.g. \textit{SP3-CDCA7} associated with ESRD only in women[13], or \textit{MGMT} and chromosome 5q13 associated with cardiovascular mortality only under tight glycemic control [47]. Given an unlimited number of patients, these associations might also be observed in non-stratified patient sets, but smaller number of patients is required when only the affected homogenous subpopulation is studied.

Epigenetic changes have been suggested to be the mechanism behind metabolic memory, i.e. sustained effect of hyperglycemia on development of complications over time. Genome-wide DNA methylation studies have shown differently methylated genes for both DKD [50] and DR [51], but further studies are needed before any methylation marks can be used as biomarkers for prediction of complications.

Whilst waiting for whole genome sequencing data for all individuals with diabetes, the genetic research should focus on large GWASs, exome chips, and WES studies in carefully characterized patient sets, aiming to understand the molecular mechanisms and pathophysiology of diabetic complications.

**Funding**

This work was supported by Folkhälsan Research Foundation, the Wilhelm and Else Stockmann Foundation, the Liv och Hälsa Foundation, Helsinki University Central Hospital Research Funds (EVO), the Signe and Ane Gyllenberg Foundation, Finska Läkaresällskapet, the Novo Nordisk Foundation (NNF14SA0003), and the Academy of Finland (134379 and 275614).
References

Recent publications with special (●) and outstanding (●●) interest are annotated


found association with diabetic retinopathy, and when combined with replication studies, the variants were genome-wide significantly associated with the joint microvascular endpoint of retinopathy and ESRD.


Figure legends

Figure 1: Loci reaching genome-wide significant association \((p<5\times10^{-8})\) with renal complications in diabetes. Effect size estimates are given as odds ratios for diabetic kidney disease (DKD) and end stage renal disease (ESRD), but as \(\beta\) for continuous traits \((\beta \text{[log(AER)]}, 10\times\beta \text{[log(eGFR)]})\). References for the loci: rs7583877 \((\text{AFF3})\) and rs12437854 \((\text{RGMA/MCTP2})\) [12]; rs4972593 \((\text{CDCA7/SP3})\) [13]; rs12523833 \((\text{SCAF8/CNKSR3})\) [15]; rs1564939 \((\text{GLRA3})\) [14]; rs12917707 \((\text{UMOD})\) [16].