Genetics of Diabetic Nephropathy: Are There Clues to the Understanding of Common Kidney Diseases?

B.R. Conway\textsuperscript{a} \ A.P. Maxwell\textsuperscript{b}

\textsuperscript{a}Centre for Inflammation Research, University of Edinburgh, Edinburgh, and \textsuperscript{b}Nephrology Research Group, Queen’s University of Belfast, Belfast, UK

**Key Words**

Genetics \cdot Susceptibility gene \cdot Polymorphism \cdot Diabetic nephropathy \cdot Kidney disease

**Abstract**

Diabetic nephropathy is the most common cause of end-stage renal disease in the Western world. There is evidence for a genetic susceptibility to diabetic kidney disease, but despite intensive research efforts it has proved difficult to identify the causative genes. Improvements in genotyping technologies have made genome-wide association studies (GWAS), employing hundreds of thousands of single nucleotide polymorphisms, affordable. Recently, such scans have advanced understanding of the genetics of common complex diseases, finding more than 100 novel susceptibility variants for diverse disorders including type 1 and 2 diabetes, coronary heart disease, Crohn’s disease and rheumatoid arthritis. In this review, type 2 diabetes is highlighted to illustrate how genome-wide association studies have been used to study the genetics of complex multifactorial conditions; in addition, diabetic nephropathy will be used to demonstrate how similar scans could be employed to detect genetic factors predisposing to kidney disease. The identification of such variants would permit early identification of at-risk patients, enabling targeting of therapy and a move towards primary prevention. In addition, these powerful research methodologies may identify genes that were not previously known to predispose to nephropathy, thereby enhancing our understanding of the pathophysiology of renal disorders and potentially leading to novel therapeutic approaches.

**Introduction**

In recent years there have been major advances in our understanding of the genetic variation that underpins susceptibility to common multifactorial diseases. Genome-wide association studies (GWAS), such as the Wellcome Trust Case Control Consortium (WTCCC) \cite{1}, that simultaneously assess hundreds of thousands of genetic markers have identified novel disease susceptibility variants for common disorders including: type 1 and 2 diabetes \cite{1}, coronary heart disease \cite{1}, inflammatory bowel disease \cite{1}, rheumatoid arthritis \cite{1}, bipolar disorder \cite{1}, cancers of the breast \cite{2} and prostate \cite{3} and traits such as body mass index \cite{4}. There is a notable absence of any renal disorders from this list; however, we will describe how similar methodologies might be utilised to identify genes that predispose to common kidney diseases, using type 1 diabetic nephropathy as an example.
There is evidence supporting the concept of a genetic susceptibility to nephropathy in patients with type 1 diabetes. Firstly, diabetologists have long recognised that a proportion of patients develop nephropathy despite apparently good glycaemic control and conversely some patients with chronic hyperglycaemia, reflecting indifferent control of diabetes, seem to be spared from this complication. Secondly, if the development of nephropathy was directly related to cumulative glycaemic burden, the prevalence would be expected to rise inexorably with increasing duration of diabetes. However, the peak incidence of nephropathy occurs between 10 and 20 years after the onset of diabetes with the incidence falling steeply thereafter; indeed only 25–40% of patients with type 1 diabetes will ever develop nephropathy [5–8]. Thirdly, several studies have demonstrated familial clustering of nephropathy [9, 10]; if two or more siblings have diabetes, the risk of the second sibling developing nephropathy is approximately threefold greater if the first sibling has diabetic kidney disease. Furthermore, diabetic offspring of parents with hypertension or cardiovascular disease are more likely to develop nephropathy, implying that genes that confer risk of hypertension in the general population predispose to kidney disease in diabetic patients [11].

Discovery of genetic variants that underpin susceptibility to nephropathy could yield important insights into this condition. Firstly, it would permit identification of patients at risk of nephropathy shortly after diagnosis of diabetes rather than much later when persistent microalbuminuria develops, by which time there is already histological evidence of renal injury. This would facilitate targeted therapeutic interventions aimed at primary prevention rather than secondary treatment of established nephropathy. Secondly, and perhaps more importantly, if the susceptibility variants are located in genes that have not previously been implicated in diabetic nephropathy, this may lead to improved understanding of its pathophysiology and development of novel therapies. Furthermore, it is possible that the genetic susceptibility variants may not be specific for type 1 diabetic nephropathy, but they may be common to type 2 diabetic nephropathy or indeed other forms of renal disease.

Rationale for Gene Discovery in Diabetic Kidney Disease

There is evidence supporting the concept of a genetic susceptibility to nephropathy in patients with type 1 diabetes. Firstly, diabetologists have long recognised that a proportion of patients develop nephropathy despite apparently good glycaemic control and conversely some patients with chronic hyperglycaemia, reflecting indifferent control of diabetes, seem to be spared from this complication. Secondly, if the development of nephropathy was directly related to cumulative glycaemic burden, the prevalence would be expected to rise inexorably with increasing duration of diabetes. However, the peak incidence of nephropathy occurs between 10 and 20 years after the onset of diabetes with the incidence falling steeply thereafter; indeed only 25–40% of patients with type 1 diabetes will ever develop nephropathy [5–8].

Thirdly, several studies have demonstrated familial clustering of nephropathy [9, 10]; if two or more siblings have diabetes, the risk of the second sibling developing nephropathy is approximately threefold greater if the first sibling has diabetic kidney disease. Furthermore, diabetic offspring of parents with hypertension or cardiovascular disease are more likely to develop nephropathy, implying that genes that confer risk of hypertension in the general population predispose to kidney disease in diabetic patients [11].

Discovery of genetic variants that underpin susceptibility to nephropathy could yield important insights into this condition. Firstly, it would permit identification of patients at risk of nephropathy shortly after diagnosis of diabetes rather than much later when persistent microalbuminuria develops, by which time there is already histological evidence of renal injury. This would facilitate targeted therapeutic interventions aimed at primary prevention rather than secondary treatment of established nephropathy. Secondly, and perhaps more importantly, if the susceptibility variants are located in genes that have not previously been implicated in diabetic nephropathy, this may lead to improved understanding of its pathophysiology and development of novel therapies. Furthermore, it is possible that the genetic susceptibility variants may not be specific for type 1 diabetic nephropathy, but they may be common to type 2 diabetic nephropathy or indeed other forms of renal disease.

Allelic Architecture of Complex Common Diseases

Identifying gene variants that predispose to diabetic nephropathy is challenging. Unlike monogenic disorders, in which a relatively rare mutation in a single gene generally confers a very high risk of developing disease, susceptibility to diabetic nephropathy is likely to be determined by a large number of relatively common allelic variants, each individually conferring a modest increase in relative risk (fig. 1). For example, the majority of alleles that have been found to be associated with type 2 diabetes have an odds ratio of less than 1.2 (table 1). The overall risk of developing nephropathy is therefore a summation of the effects of the alleles present at each susceptibility locus interacting with each other and with environmental influences such as long-term glycaemic and blood pressure control.

The commonest forms of variation in the human genome are single nucleotide polymorphisms (SNPs), which are single base substitutions, insertions or deletions that occur at a frequency of at least 1% in a given population. It is estimated than there are more than 10 million SNPs
in the human genome, occurring at an approximate frequency of one in every 300 bp [12]. Additional genomic complexity is generated by inherited duplications or deletions of short segments of chromosomes; each individual being assessed may therefore have fewer or more than the standard two parental alleles for each gene, a feature known as copy number variation [13]. Furthermore, genetic and environmental factors both contribute to variation in epigenetic phenomena such as DNA methylation, which may affect gene expression and the risk of developing diseases [14].

Modest Success in Identifying Nephropathy Susceptibility Genes

Given the complexity involved, it is not surprising that despite investment of significant resources, there has been limited success in identifying variants that modify the risk of developing renal disease. Studies that test for association between gene variants and diabetic nephropathy have tended to produce inconsistent findings (table 2), with only the angiotensin converting enzyme insertion/deletion polymorphism (ACE I/D) demonstrating positive association by meta-analysis [15]. This implies that there are major limitations in the principal research methodologies, linkage analysis and candidate gene association studies that have been employed to date.

Linkage analysis has been integral to the discovery of mutations that cause single-gene disorders; however, it has limited power to detect gene variants that confer a relatively low risk of disease (fig. 1). For effective analysis, large family pedigrees are necessary and these are particularly difficult to collect in the context of diabetic nephropathy as parents of such patients have often died prematurely due to their higher prevalence of cardiovascular disease [11].

The candidate gene strategy utilises our current understanding of the pathophysiology of diabetic kidney disease to prioritise genes to test for association with ne-
Nephropathy usually by employing a case-control study. However, in reality, selection of genes is effectively limited to an educated guess (table 2) and recent genome-wide association scans have indicated major deficiencies in the current understanding of the pathogenesis of many diseases. For instance, few of the genes listed in table 1 would previously have been considered to be putative candidate genes for susceptibility to type 2 diabetes.

Gene Hunting Using Genome-Wide Association Scans

Given the limitations of candidate gene studies it is now appropriate to consider genome-wide association strategies to search for nephropathy susceptibility genes. Whilst such a screening method is more expensive than an individual candidate gene study, in overall terms it is likely to be more cost-effective per susceptibility variant detected. In type 2 diabetes, for example, the majority of susceptibility variants detected to date have been identified from just five GWAS [1, 16–19], with only two variants identified from the much larger number of previously conducted candidate gene studies. Such GWAS have recently become feasible due to the development of several key research tools.

Firstly, the HapMap project has catalogued over three million SNPs within the human genome which can be used as markers to identify nearby gene variants that associate with disease [20]. Importantly, the HapMap has also characterised the patterns of linkage disequilibrium (LD) within the human genome [12]. LD refers to the concept that when two polymorphisms lie close to each other along a chromosome, there is very low risk of recom-

Table 2. Candidate genes in which polymorphisms have demonstrated association with type 1 diabetic nephropathy

<table>
<thead>
<tr>
<th>Class</th>
<th>Gene name</th>
<th>Symbol</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renin-angiotensin system</td>
<td>angiotensin-converting enzyme 1</td>
<td>ACE</td>
<td>17q23</td>
</tr>
<tr>
<td></td>
<td>angiotensinogen</td>
<td>AGT</td>
<td>1q42–43</td>
</tr>
<tr>
<td></td>
<td>angiotensin II receptor, type 1 and 2</td>
<td>AGTR1</td>
<td>3q21–25</td>
</tr>
<tr>
<td>Glucose metabolism</td>
<td>aldose reductase</td>
<td>AKR1B1</td>
<td>7q35</td>
</tr>
<tr>
<td></td>
<td>glucose transporter-1</td>
<td>SCL2A1</td>
<td>1p35</td>
</tr>
<tr>
<td></td>
<td>receptor for advanced glycosylation end-products</td>
<td>RAGE</td>
<td>6p21.3</td>
</tr>
<tr>
<td>Growth factors</td>
<td>transforming growth factor-β1</td>
<td>TGFBI</td>
<td>19q13.1</td>
</tr>
<tr>
<td></td>
<td>transforming growth factor-β, receptors I–III</td>
<td>TGFBR1/2/3</td>
<td>9q22, 3p22, 1p33</td>
</tr>
<tr>
<td></td>
<td>vascular endothelial growth factor</td>
<td>VEGF</td>
<td>6p12</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>superoxide dismutase 1 and 2</td>
<td>SOD1/2</td>
<td>21q22 6q25.3</td>
</tr>
<tr>
<td></td>
<td>haptoglobin</td>
<td>HP</td>
<td>16q22</td>
</tr>
<tr>
<td></td>
<td>paraoxonase 1 and 2</td>
<td>PON1/2</td>
<td>7q21–22</td>
</tr>
<tr>
<td></td>
<td>catalase</td>
<td>CAT</td>
<td>11p13</td>
</tr>
<tr>
<td></td>
<td>glutathione peroxidase 1</td>
<td>GPX1</td>
<td>3p21.3</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>apolipoprotein e</td>
<td>APOE</td>
<td>19q13.2</td>
</tr>
<tr>
<td></td>
<td>adiponectin</td>
<td>ADIPOQ</td>
<td>3q27</td>
</tr>
<tr>
<td></td>
<td>peroxisome proliferator activated receptor gamma</td>
<td>PPARy</td>
<td>3p25</td>
</tr>
<tr>
<td>Cytoskeletal genes</td>
<td>caldesmon</td>
<td>CALD</td>
<td>7q35</td>
</tr>
<tr>
<td></td>
<td>beta-adducin</td>
<td>ADD2</td>
<td>2p14</td>
</tr>
<tr>
<td>Inflammation</td>
<td>intercellular adhesion molecule-1 regulated upon</td>
<td>ICAM1</td>
<td>19p13.3</td>
</tr>
<tr>
<td></td>
<td>activation, normal T cell expressed and secreted</td>
<td>CCR5</td>
<td>3p21</td>
</tr>
<tr>
<td></td>
<td>gene receptor</td>
<td>IL1B</td>
<td>2q14</td>
</tr>
<tr>
<td>Other</td>
<td>unc-13 homolog B</td>
<td>UNC13B</td>
<td>9p11–12</td>
</tr>
<tr>
<td></td>
<td>endothelial nitric oxide synthase</td>
<td>NOS3</td>
<td>7q36</td>
</tr>
<tr>
<td></td>
<td>Protein kinase C, β1</td>
<td>PRKCB1</td>
<td>16p11.2</td>
</tr>
</tbody>
</table>
bination occurring between them during meiosis; therefore, they will tend to be inherited together within a population, even when individuals are separated by many thousands of meioses. Hence in a genome-wide scan many SNPs are redundant as their genotype can be inferred by typing adjacent SNPs, and a subset of approximately 500,000 to 1,000,000 carefully chosen marker SNPs known as haplotype tag SNPs (htSNPs) [21] can be employed to genotype the majority of the common variation within the genome.

Secondly, technological advances permit the genotyping of large numbers of SNPs in thousands of individuals within the scope of a large project grant. For example, high-throughput commercial genotyping platforms such as Affymetrix (http://www.affymetrix.com) and Illumina (http://www.illumina.com) are now able to genotype up to one million SNPs in several thousand individuals at a cost of less than GBP 1,000,000. In addition, such platforms can assess some of the copy number variation and the extent of DNA methylation present in individual genomes.

Thirdly, GWAS generate many millions of data-points and the bioinformatic and statistical tools required for analysing this information have evolved in parallel. Crucially, current statistical methodologies can detect false-positive results due to bias introduced by population stratification and incorrect genotype assignment.

Thus a GWAS is now a practical option for identifying nephropathy susceptibility variants, but given the still significant costs of such a scan, careful consideration must be given to the design of the study in order to maximise the chance of success.

**Phenotyping of Cases and Controls**

Accurate phenotyping of cases and controls is essential as misclassification will introduce statistical noise, thereby reducing the power of the study. In particular a key criterion in selection for cases should be overt proteinuria, as although microalbuminuria is a risk factor for subsequent development of nephropathy, a sizeable proportion of patients may have spontaneous regression of microalbuminuria [22]. It is also important to appreciate that most case-control studies are cross-sectional in nature so there is a risk that patients recruited as controls could develop nephropathy at a later date. However, given the natural history of diabetic nephropathy, recruiting control patients who have no evidence of microalbuminuria after at least 15 years of diabetes is a reasonably pragmatic approach as these individuals are at very low risk of subsequently developing nephropathy [5].

However, employing rigorous inclusion and exclusion criteria renders it difficult to collect a sufficiently large number of cases and controls from a single centre in order to achieve adequate statistical power. For instance, for a study to have 80% power to detect a variant that conferred a 1.5-fold increased risk of nephropathy and where the minor allele frequency is 10% would require more than 1,500 cases and controls if the level of significance was set at $1 \times 10^{-7}$, an appropriate threshold for GWAS (see below). Hence multicentre collaborations have been initiated, such as the GoKinD collection (http://www.jdrf.org/gokind), which has succeeded in recruiting its target of more than 600 cases and 600 controls in both the USA and the UK. However even this large collection remains inadequately powered; therefore, multi-national collaborations between consortia such as GoKinD, the FIND study in the USA (http://darwin.cwru.edu/FIND/), and the Finnish Finn-Diane study (http://www.folkhalsan.fi/default.asp?path=228;273;280;287;285;1170) are being explored. Such co-operation between investigators has underpinned the success in discovering gene variants that predispose to type 1 and 2 diabetes.

**Selection of Makers and Study Design for Genome-Wide Scans**

While careful selection of htSNPs can improve the efficiency of genome-wide scans, the amount of genotyping required may be further reduced by focusing on the relatively small proportion of SNPs that lie adjacent to or within genes, for example non-synonymous SNPs that alter the amino acid sequence [23, 24]. However, the disadvantage of this strategy is that it will not detect intergenic polymorphisms that regulate gene expression. Indeed, the majority of variants that have been found to associate with type 2 diabetes are located either within introns or between genes, often lying more than 100 kb away from the nearest gene (table 1).

Arguably, a more efficient approach to improve genotyping success is the use of a staged strategy (fig. 2). In the first stage, a proportion of all the available patients would be genotyped for the entire panel of SNP markers. Subsequently, a nominal proportion of typed SNPs (e.g. top 0.1% demonstrating the strongest association with nephropathy) would then be genotyped in an additional case-control population; the rationale being that the top 0.1% of SNPs associated with nephropathy will be enriched for the susceptibility alleles. The data from the two stages are then combined, thereby reducing the amount of genotyping required, but without significantly impacting on the power of the study [25, 26].

Strategies to Detect Common Susceptibility Variants for Kidney Disease
Risk of False Positives

One problem with genotyping vast numbers of SNPs in a GWAS is that the majority of associations detected will be false positives (type 1 errors). For example, if 1,000,000 SNPs were genotyped and a conventional p value of 0.05 was employed, approximately 50,000 SNPs would associate with disease by chance alone, swamping the relatively small number of true positive associations. To reduce this risk, the significance threshold is corrected for multiple testing; for example, a threshold of $1 \times 10^{-7}$ for each SNP would result in a whole genome false-positive rate of approximately 1 in 20. In addition, the odds that a positive association is a true positive is proportional to the power of the study; therefore, smaller studies require more stringent significance thresholds in order to have a similarly low risk of false-positive results [1].

However, despite the use of very high significance thresholds, a proportion of the associations detected will still represent false positives, and therefore replication of the results in additional populations is required before the association can be declared valid (fig. 2). For example, the variant in the FTO gene that confers susceptibility to obesity has been validated in 13 separate cohorts, comprising a total of almost 40,000 individuals [4].

Identification of Causal Variants

Even after detection of valid associations between SNP markers and disease, more detailed mapping is required to identify the causal variants. Blocks of LD may extend for hundreds of kilobases; therefore, distinguishing causal genetic variants from those that are simply associated markers can be challenging. Determining whether a SNP is a causal genetic variant or an associated marker may be less important when used in individual risk prediction, but it is much more important when trying to dissect the pathogenesis of a condition such as diabetic nephropathy. For instance, markers in an area on chromosome 12q24 associate with type 1 diabetes; however, there is extensive LD in this region, spanning more than 1.5 million base pairs and implicating in theory up to 15 genes [1]. One potentially helpful strategy is to use a more ancient population, in the context of evolution, such as an African
population. In such a population many more meioses have occurred and therefore the blocks of LD are shorter. Once the region of interest has been defined, fine genetic mapping by extensive DNA sequencing of the region may identify particularly attractive novel variants that alter protein structure or gene expression. The efficiency of fine mapping may be improved by studying persons with extreme phenotypes whose individual genomes are likely to be enriched for susceptibility or protective genetic variants (fig. 2). For instance, such phenotypes might include cases that have developed nephropathy early despite well controlled diabetes or controls who have no evidence of nephropathy despite exceptionally long duration of diabetes and poor glycaemic control.

**Limitations of Genome-Wide Association Scans**

The recent GWAS have, however, only identified a small proportion of all of the genetic variants that predispose to disease. For example, while the sibling relative risk (or $\lambda_s$ value) for type 2 diabetes is 3.0, the combined $\lambda_s$ value provided by all of the susceptibility loci detected to date is only 1.07 [27]. While this limits the utility of such loci to predict an individual’s risk of disease [28, 29], it also offers hope that many more susceptibility variants await discovery.

There are several reasons why the recent GWAS have had limited success in identifying disease susceptibility variants. Firstly, given the high genome-wide significance threshold required, true positives may be prematurely disregarded (type 2 errors). For example in the WTCCC study in type 2 diabetes, the association signal from the PPARγ gene did not reach a genome-wide significance level (table 1). It was only when markers in this region were tested in additional diabetic populations that the strength of the association signal exceeded the stringent $1 \times 10^{-7}$ threshold [30]. Hence, in order to reduce the risk of prematurely dismissing true positives, a large number of markers showing the strongest association in stage 1 of a GWAS should be taken forward to stage 2 (fig. 2).

Secondly, the number of cases and controls required to provide adequate power to detect association with disease increases exponentially with decreasing minor allele frequency [31]. GWAS will therefore tend to favour detection of common susceptibility variants which have relatively high minor allele frequencies (table 1) and are unable to identify rare variants, even if they confer modest risk of disease (fig. 1). The ability of GWAS to detect the majority of the variation that underpins susceptibility to complex disease is therefore dependent on whether the genetic predisposition is largely accounted for by common, low impact SNPs or by rare variants which confer a much greater individual risk of disease (fig. 1).

Additional caveats are specifically relevant to GWAS in type 1 diabetic nephropathy. Firstly, many susceptibility variants for complex disease have only been detected following meta-analyses employing large numbers of patients (for example, a total of almost 90,000 cases and controls in type 2 diabetes) [30]. Given the low prevalence of type 1 diabetic nephropathy as compared with type 2 diabetes, cohorts of this magnitude will be very difficult to assemble and replication studies may need to include patients with nephropathy due to type 2 diabetes. Secondly the extent of heritability of diabetic nephropathy is less than that observed in type 1 and type 2 diabetes. This implies that it is likely to be much harder to detect nephropathy susceptibility variants as they may be fewer in number, confer on average a lower risk of disease or their signal may be masked by strong environmental influences.

**Conclusion**

In summary, recent advances in study design and genotyping technologies have permitted the identification of more than 100 loci harbouring gene variants that predispose to various common diseases [32]. Similar strategies can now be utilised to dissect the genetic susceptibility to many common renal diseases such as diabetic nephropathy. Significant pooling of resources by international collaboration will be needed to generate sufficiently sized case-control studies that provide adequate statistical power for genome-wide association. These approaches promise to deliver many novel and surprising insights into the genetic susceptibility to common kidney diseases and lead to the discovery of additional molecular pathways suitable for therapeutic interventions.
References

1. Wellcome TCCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.


This review of genetic approaches for the understanding of the susceptibility and complications of diabetes mellitus (DM) and its associated nephropathy by Conway and Maxwell draws attention to advances in the field of genetics and their applicability to complex chronic diseases including DM and CKD. It highlights the value and improvements made in the field of genomic-wide association studies (GWAS). It points to a number of single nucleotide polymorphisms (SNPs) underlying diabetes and diabetic nephropathy. They often relate to the coding of proteins, cytokines, growth factors and signal transduction mediators thought to be involved in the pathogenesis of diabetes and diabetic nephropathy. In particular, growing attention is focusing on the role of transcription factor 7-like 2 (TCF7L2) polymorphism in the susceptibility, development and complications of type 2 DM including nephropathy. TCF7L2 is a key element of the Wnt signaling pathway. In Europeans, GWAS showed TCF7L2 to be the most important locus predisposing to T2D so far. For the first time, a gene is consistently involved in T2D susceptibility in all major ethnic groups. The presence of the TCF7L2 rs7903146 risk allele increases TCF7L2 gene expression in islets beta cells and potentially directly affects insulin secretion. The association of TCF7L2 polymorphisms with T2DM provides new insights into future genetic predisposition tests but remains the tip of the T2DM genetic iceberg. Other potential associations are highlighted by Conway and Maxwell in their review. Furthermore, this research has also opened the door to the exploration through GWAS of genetic susceptibility to the progression of nondiabetic CKD where TCF7L2 was also found to be of significance in a number of cohorts [Kottgen A, et al: TCF7L2 variants associate with renal function and CKD progression in population-based cohorts. J Am Soc Nephrol 2008;19:1989–1999].