Nephrol Dial Transplant (2015) 30: iv26–iv34
doi: 10.1093/ndt/gfv087

ndt
Nephrology Dialysis Transplantation

Full Review

Genome-wide studies to identify risk factors for kidney disease with a focus on patients with diabetes

Florina Regele[1], Kira Jelencsics[1], Dov Shiffman[2], Guillaume Paré[3], Matthew J. McQueen[3], Johannes F.E. Mann[3,4] and Rainer Oberbauer[1]

[1]Division of Internal Medicine 3, Department of Nephrology and Dialysis, Medical University Vienna, Vienna, Austria, [2]Quest Diagnostics, Alameda, CA, USA, [3]Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, ON, Canada and [4]Department of Nephrology and Hypertension, Friedrich Alexander University, Erlangen, Germany

Correspondence and offprint requests to: Rainer Oberbauer; E-mail: rainer.oberbauer@meduniwien.ac.at; www.meduniwien.ac.at/nephrogene

ABSTRACT

Chronic kidney disease (CKD) affects 10–13% of the general population and diabetic nephropathy (DN) is the leading cause of end-stage renal disease (ESRD). In addition to known demographic, biochemical and lifestyle risk factors, genetics is also contributing to CKD risk. In recent years, genome-wide association studies (GWAS) have provided a hypothesis-free approach to identify common genetic variants that could account for the genetic risk component of common diseases such as CKD. The identification of these variants might reveal the biological processes underlying renal impairment and could aid in improving risk estimates for CKD. This review aims to describe the methods as well as strengths and limitations of GWAS in CKD and to summarize the findings of recent GWAS in DN. Several loci and SNPs have been found to be associated with distinct CKD traits such as eGFR and albuminuria. For diabetic kidney disease, several loci were identified in different populations. Subsequent functional studies provided insights into the mechanism of action of some of these variants, such as *UMOD* or *CERS2*. However, overall, the results were ambiguous, and a few of the variants were not consistently replicated. In addition, the slow progression from albuminuria to ESRD could limit the power of longitudinal studies. The typically small effect size associated with genetic variants as well as the small portion of the variability of the phenotype explained by these variants limits the utility of genetic variants in improving risk prediction. Nevertheless, identifying these variants could give a deeper understanding of the molecular pathways underlying CKD, which in turn, could potentially lead to the development of new diagnostic and therapeutic tools.

Keywords: chronic kidney disease, diabetes mellitus, genome-wide association studies

INTRODUCTION

Chronic kidney disease (CKD) affects about 10–13% of the general population, and the incidence and prevalence are increasing in most countries [1, 2]. CKD has multiple aetiologies; the two major aetiologies are hypertensive nephropathy and diabetic nephropathy (DN). DN affects up to 40% of all patients with diabetes and is the leading cause of end-stage renal disease (ESRD) [3]. In patients with diabetes, albuminuria often precedes the decline of kidney function as measured by eGFR and is therefore considered a first sign of DN [4].

While many demographic, biochemical and lifestyle risk factors of CKD have been established, a portion of the risk of CKD remains unexplained by these factors pointing towards a possible genetic contribution [5]. The genetic component of CKD has been demonstrated in familial aggregation studies in families with diabetes and hypertension, which estimated the heritability to range from 36 to 75% for glomerular filtration rate (GFR) and from 16 to 49% for albuminuria [6, 7].

Given the many potential genetic risk factors for common diseases such as CKD, a genome-wide association study is an excellent screening tool to discover genetic risk

loci. In addition to improving risk prediction for CKD, identification of genetic markers associated with CKD could also provide us with important insights into the underlying biological processes of the renal impairment.

Genome-wide association studies (GWAS) focus on the most common kind of genetic variation in the human genome, a *single nucleotide polymorphism* (SNP). SNPs are common substitutions of a single base with another, which occur with high frequency in the human genome (1 every 300–500 base pairs) [8]. Although most of them have no functional outcome, some SNPs might change the amino acid sequence of the resulting protein, the stability of the mRNA transcript or the transcription factor binding activity, splicing or epigenetics regulation. These changes could result in biological changes, which could play a role in disease susceptibility.

For each SNP, there are typically two possible alleles (i.e. two possible base-pairs). The frequency of these alleles in any given population can be assessed, and the term *minor allele frequency* (MAF) refers to the frequency of the less common allele in the population.

In contrast to the rare diseases that are caused by single-locus mutations (Mendelian diseases), the genetic component of common polygenic diseases such as CKD is thought to involve many common genetic variants. The *common disease/common variance* (CD/CV) hypothesis states that common SNPs with their high minor allele frequencies and small effect sizes constitute the genetic risk in CD (Figure 1) [9]. Because a single SNP explains only a small proportion of the trait's variance, multiple genetic variants are required to account for the total genetic risk of a disease.

A GWAS is frequently conducted in a case–control study design. In this design, the allele frequency of each SNP is compared between individuals with a disease or trait (cases), and individuals without the disease or trait (controls). SNPs with different allele frequency in cases and controls are identified as being associated with the specific disease or trait. A precise definition of cases and controls is crucial, because this 'allocation' is done *post hoc*, which makes case–control studies prone to selection bias, which occurs when controls are not representative of the population of cases [10]. A common alternative to the binary case–control study design is the analysis of quantitative traits, such as GFR or albuminuria, which improves the power of a GWAS and does not require a differentiation between cases and controls. Quantitative traits can also be studied in a case–control design by defining quantitative trait thresholds that would discriminate between cases and controls.

The study cohort can be population-based, i.e. a sample of unrelated individuals derived from the general population, or consist of a selected population with a certain trait, e.g. only patients with diabetes. A common alternative design is family-based studies, which study the association of variants with disease among trios of affected and unaffected parents/offspring or among twins or siblings [11].

When choosing quantitative traits and case definitions in GWAS of CKD, it is important to consider that pathogenesis and the associated genetic variants may differ between aetiologies and in different stages of disease. Different renal traits may be associated with distinct genetic variants [12]. Another essential aspect in GWAS design is calculation of the statistical power of the study. The size of the study, the magnitude of the effect and the allele frequency determine the power to detect an association in a given study. Because of the typically small effect size of individual SNPs, many thousands of participants are needed to achieve sufficient power to detect an association with genetic variants while accounting for testing hundreds of thousands of SNPs [13]. This is often accomplished by conducting a meta-analysis of GWAS

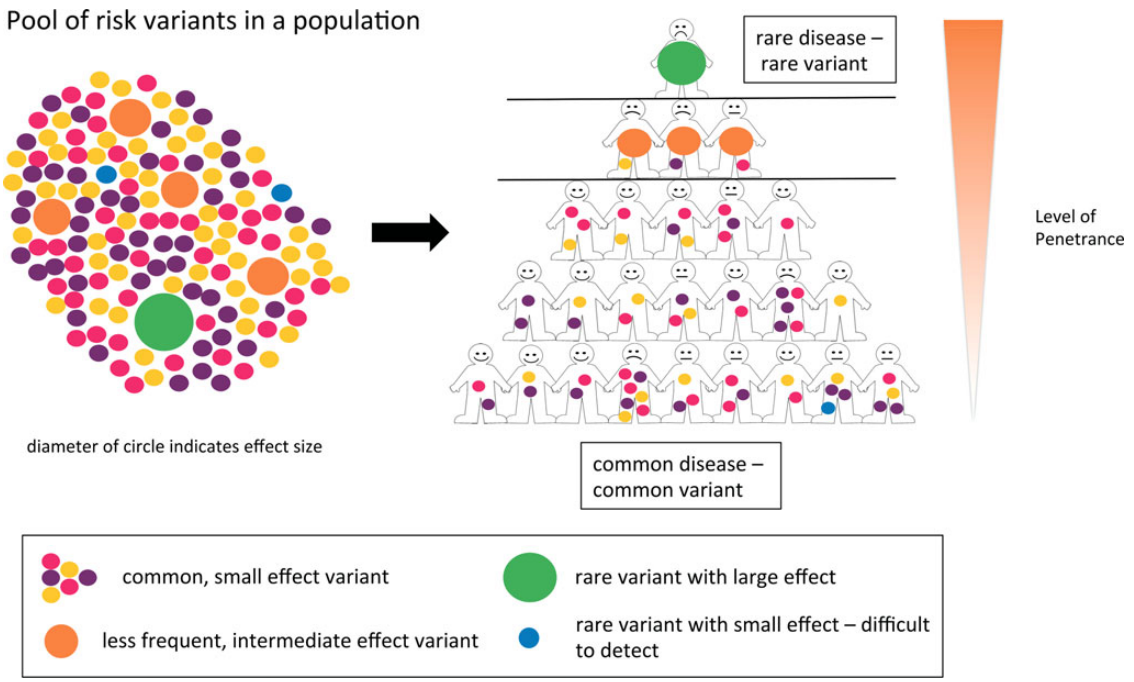

FIGURE 1: Allele frequencies and effect sizes in human diseases.

data from many different studies, which can increase the overall power of the studies considerably.

Designing GWAS has been greatly facilitated by the growing understanding of the structure of genetic variation in the human genome and the development of whole-genome databases, such as the International HapMap Project and the 1000 Genomes Project (http://www.1000genomes.org/) [8]. The International HapMap Project established a publicly available database of SNP frequencies and haplotypes in different populations. A haplotype is a linked set of SNP alleles that are likely to be inherited together through generations and are statistically associated, as can be described by the degree of linkage disequilibrium (LD). Haplotype information in any given genomic region allows the identification of tag SNPs (tSNPs) that can be used to predict the genotypes of other SNPs in the region. Therefore, genotyping a small number of tSNPs could provide information on additional SNPs in the same genomic region. Imputation, the prediction of genotypes based on tSNPs, increases the number of analysed SNPs and improves the genomic coverage of the study.

Other databases provide us with information about the potential functional outcome of SNPs, such as the potential effect of amino acid change on protein function, or potential effect of DNA or RNA regulatory elements, which can affect gene transcription, splicing or mRNA stability. This information can indicate potential functional consequences of SNPs, which further helps to decide which SNPs should be interrogated/further investigated.

QUANTITATIVE EVALUATION OF GWAS STUDIES; STATISTICAL BACKGROUND

For a biallelic gene locus, the three potential genotypes for each individual are major homozygous (i.e. carrying two versions of the more common allele, usually denoted as *AA*), minor homozygous (denoted as *aa*) or heterozygous (i.e. carrying one major and one minor allele, denoted as *Aa*). It is important to note that these genotypes occur with different frequencies in a population, and that these frequencies might differ depending on ethnicity. The frequency of genotypes depends on the frequency of the two alleles *A* and *a* in the population. The probability of carrying allele *A* is usually denoted as p, and the probability of carrying allele *a* as q, with $p + q = 1$ (i.e. 100%). Consequently, the probability of being homozygous for allele *A* (genotype *AA*) is p^2, the probability of being homozygous for allele *a* is q^2, and the probability of being heterozygous is $2pq$.

The relation between allele frequencies and genotype frequencies is illustrated in Table 1 for two alleles (*A* and *a*) that occur in 95 and 5% of the population, respectively.

From Table 1, it can be seen that the proportion of alleles and genotypes can be described by the formula $p^2 + 2pq + q^2 = 1$, known *as* the Hardy–Weinberg equilibrium (HWE). Alleles in the gene pool are expected to be at HWE as long as a population is not subject to any evolutionary influences, such as strong selection pressure, non-random-mating, migration, mutations or genetic drift. Under these conditions, the alleles in the gene pool are merely 'shuffled' over time, which explains why even rare alleles with low MAF do not disappear, and allele and genotype frequencies remain stable over generations. These conditions are never fully met in reality, because most populations are subject to one or many evolutionary influences. However, for most populations, evolutionary changes happen at such a slow rate that they appear to be close to HWE. Testing for deviations from HWE by a Chi-square test can therefore be used for quality control test to identify systematic genotyping errors or population stratification in the study, or to indicate whether a marker is associated with case status if deviation is found only in the case group [14, 15].

The statistical analysis of GWAS data generally aims to determine if any of the genotyped markers is associated with case status or a quantitative trait at the predefined level of significance. Such analysis needs to account for multiple testing of the roughly 10 million SNPs in the genome. For categorical outcomes in a case–control study, the main objective is to determine if any genotype, for example being homozygous for a potential risk allele, constitutes a higher risk for the disease compared with another genotype (for example being homozygous for the other allele).

A first analysis can be done using a contingency table, which lists the genotype distribution and allele frequency for cases and controls separately. Table 2 is a contingency table for a variant with a C risk allele and a T non-risk allele.

The odds ratio (OR) is the ratio of the odds of being a case when having a specific genotype to the odds of being a case when having a reference genotype.

From Table 2, the OR for the risk genotype CC in reference to the genotype TT can be calculated as

$$\mathrm{OR}\ (\mathrm{CC\ versus\ TT}) = \frac{\mathrm{CC}_{\mathrm{cases}}/\mathrm{CC}_{\mathrm{controls}}}{\mathrm{TT}_{\mathrm{cases}}/\mathrm{TT}_{\mathrm{controls}}} = \frac{571/548}{39/68} = 1.81$$

Table 1. Example for distribution of allele and genotype frequencies in the population

Alleles	***A*** 95% of alleles in a population (P = 0.95)	***a*** 5% of alleles in a population ($q = 0.05$)
A P = 0.95	***AA*** occurs with a frequency of $p^2 = 0.9025$ 90.25% of population	***Aa*** occurs with a frequency of $pq = 0.0475$ 4.75% of population
a P = 0.05	***aA*** occurs with a frequency of $pq = 0.0475$ 4.75% of population	***aa*** occurs with a frequency of $q^2 = 0.0025$ 0.25% of population

Table 2. Example contingency table showing genotype distribution for cases and controls

	Cases ($n = 1000$)	Controls ($n = 1000$)
Risk allele homozygotes (C/C)	571	548
Heterozygotes (C/T)	390	385
Non-risk homozygotes (T/T)	39	68

Alternatively, to test the association of single alleles (for example, C versus T), a similar equation using the allele counts is used as follows:

$$\text{OR (C versus T)} = \frac{C_{cases}/C_{controls}}{T_{cases}/T_{controls}} = \frac{1532/1481}{468/521} = 1.15$$

A Chi-square test or Fisher's exact test is used to assess the significance of the association.

Multivariable regression models, such as linear regression for continuous outcomes (e.g. GFR), or logistic regression for dichotomous outcomes (e.g. CKD/no CKD), can be used to assess the influence of other factors, such as sex, age, population substructure or study site on the association.

The choice of statistical model depends on the number of variants investigated. When using classical regression models, it has to be noted that the large number of hypothese tested in a GWAS increases the probability of false-positive results unless appropriate multiple testing correction is applied. There are several approaches to correct for this problem of multiple testing, such as the Bonferroni correction, the false discovery rate or a Bayesian approach [10, 16].

Simulation studies aimed to develop standards for genome-wide significance [17, 18]. Based on the distribution of LD and resulting independent genomic regions, the effective number of independent tests in a dense genome-wide scan has been estimated as 1 million for European populations. Subsequently, the unadjusted P-value of 5×10^{-8} is commonly accepted as a significance threshold for GWAS in populations of European descent. For African populations, which have a greater genetic diversity, the threshold is closer to 10^{-8}.

Alternatively to the use of classical regression models and an adjusted P-value, there are more complex approaches such as lasso, ridge regression or elastic net to account for the huge abundance of predicting variables compared with the number of observed subjects [19, 20].

Another important factor to consider is the possibility of spurious associations caused by population stratification, which describes differences in allele frequencies due to ethnicity or population substructure differences between cases and controls. Furthermore, it is important to consider that when an SNP is associated with disease, it is not necessarily causally related to the disease. The associated SNP could merely be a tSNP for the causal variant. To identify potentially causal variant(s), the genomic region identified by the tSNP should be fine mapped by a denser genotyping.

EXTERNAL VALIDATION

A key element in GWAS is the validation of significant SNPs from the discovery phase in an independent study. The discovery study often overestimates the effect size of the discovered variants, a phenomenon known as 'winner's curse'. The overestimation of the effect size results in an underestimation of sample size for a validation study, and an underpowered validation study may, in some cases, explain the failure to replicate the results of the discovery study [21].

GENOME-WIDE STUDIES TO IDENTIFY GENETIC RISK FACTORS FOR CKD

Köttgen *et al.* [22, 23] performed two large meta-analyses of over 20, predominantly population-based, studies, altogether including over 90 000 individuals of European ancestry. Two of the studies in the replication stage consisted entirely of diabetic patients; in the other studies, diabetes prevalence ranged from 2.6 to 15.3%. Both eGFR (estimated by creatinine and cystatin C) and CKD (defined as eGFR < 60 mL/min/1.73 m^2) were investigated in these studies. They identified 16 loci that were associated with renal function and CKD at genome-wide significance level ($P < 5 \times 10^{-8}$). Several of these loci had previously been linked to renal disease, as for example rare, mutations in the *UMOD* locus, which cause rare, autosomal-dominant renal diseases such as familial juvenile hyperuremic nephropathy and medullary cystic kidney disease type 2 [23, 24]. *UMOD* encodes the most abundant protein excreted in urine, uromodulin, also known as Tamm–Horsfall protein; however, its physiological function is not fully understood [25]. Other variants identified by Köttgen *et al.*, were in genes related to nephrogenesis glomerular filtration barrier formation and podocyte function, angiogenesis, solute transport, metabolic functions of the kidney and the function of primary cilia [23]. Together, the 16 loci accounted for only 1.4% of variability of eGFR. These 16 eGFR-associated SNPs were also evaluated for their association with baseline eGFR, baseline albuminuria and time to stage 3B CKD in 3028 patients with type 2 diabetes. The association with eGFR was replicated for 3 SNPs in *UMOD*, *GCKR* and *SHROOM3*, but none of the 16 SNPs were associated with albuminuria or time to stage 3B CKD [26].

In a large meta-analysis involving 31 850 individuals of European ancestry and 6981 African Americans, an SNP in the *CUBN* locus was found to be associated with higher levels of albuminuria, but not with eGFR or CKD. The results were replicated in 27 746 individuals and were independent of the hypertension or diabetes status. Only 0.15% of variance in albuminuria levels was explained by the SNP.

In the context of the large SysKid collaborative project (www.syskid.eu), Shiffman *et al.* investigated the association of the 16 eGFR-associated SNPs identified by Köttgen *et al.* with the annual rate of increase of albuminuria among 3723 diabetic patients of European and non-European ancestry [27]. One variant, rs267734 in *CERS2*, was found to be associated with the rate of increase of albuminuria ($P = 0.0015$). The annual rate of increase of albuminuria was 11.3% for homozygote carriers of the risk allele, compared with 5.0% for heterozygotes and 1.7% for non-risk homozygotes. For patients who were normo-albuminuric at baseline, each risk allele was associated with a 50% increased risk of incident albuminuria after adjustment for age, sex, ethnicity, principal component of genetic variability, baseline hypertension, eGFR, uACR, smoking status, study and treatment group. Another SNP, rs267738, which was in high LD with the original SNP and encoded an amino acid change in the *CERS2*-encoded protein, was also associated with progression of albuminuria with $P = 0.0013$. The association of the initially identified SNP (rs267734) with

albuminuria progression was confirmed in an independent large population of 4390 participants of the ORIGIN study (P = 0.02) [28].

In recent years, several other loci associated with DN and ESRD have been discovered. Variants in the engulfment and cell motility 1 (*ELMO1*) locus were associated with renal disease in individuals with type 1 diabetes in a Caucasian population, and type 2 diabetes in a Japanese and African American population [29–31]. Overexpression of *ELMO1* was shown to contribute to the progression of chronic glomerular injury [32].

A meta-analysis of 21 studies showed that the *MTHFR* gene 677TT genotype might confer a moderately increased risk for DN and diabetic retinopathy [33]. *MTHFR* encodes the methylenetetrahydrofolate reductase, which plays an important role in homocysteine metabolism. The TT genotype was also shown to be associated with cardiovascular disease in ESRD [34].

Two studies showed that variants in *MYH9* account for most of the 2- to 4-fold increased risk for non-diabetic ESRD and focal segmental glomerulosclerosis in African Americans compared with Europeans [35, 36]. Two years later, it was found that this risk was actually conferred by variants in the neighbouring *APOL1*, which are in high LD with the *MYH9* SNPs and encode non-synonymous amino acid changes [37]. Variants in *APOL1* were shown to be associated with higher rates of ESRD and a more rapid progression of CKD in black patients, when compared with white patients, regardless of the diabetes status [38]. Other studies investigating the *MYH9/APOL1* region in DN found an association of several variants in *MYH9* with DN in African Americans and European Americans [35, 39, 40], however, this association was not confirmed in a UK population [41].

Several studies investigated genetic risk factors specifically for type 1 or type 2 diabetes. A large meta-analysis of GWAS in DN type 1 identified two variants associated with diabetic ESRD at genome-wide significance, one in the *AFF3* gene and one intergenic SNP between *RGMA* and *MCTP2*. Functional data suggest that *AFF3* influences renal tubule fibrosis via the transforming growth factor-beta (TGF-β1) pathway. The strongest association with DN (defined as macroalbuminuria or ESRD due to DN) was detected for an SNP in the *ERBB4* gene, which however did not reach genome-wide significance ($P = 2.1 \times 10^{-7}$). *ERBB4* encodes a member of the EGF receptor tyrosine kinase family and modulates kidney tubule proliferation and polarity during nephrogenesis. Subsequent pathway analysis of genes co-expressed with *ERBB4* indicated potential involvement of ERBB4 in fibrosis [42].

Several gender-specific associations with CKD have been reported, most convincingly with the rs4972593, which is associated with ESRD in women, but not in men with type 1 diabetes. This SNP is located on chromosome 2q31 between the Sp3 transcription factor (*SP3*) and the cell division cycle associated 7 (*CDCA7*) genes. *SP3* is a transcription factor which shows higher glomerular expression level in women. *CDCA7* is a transcription factor regulating cell proliferation [43].

Several candidate genes for DN are related to the renin–angiotensin–aldosterone–system. A meta-analysis from Wang *et al.* suggested that an insertion/deletion polymorphism in the angiotensin-converting enzyme might contribute to DN development, especially in an Asian population with type 2 diabetes mellitus [44].

Variants in the angiotensinogen (AGT) and angiotensin II receptor type 1 (*AGTR1*) have also been shown to be involved in the development of DN. A meta-analysis suggested that the *AGTR1* A1166C polymorphism may contribute to DN development, particularly in type 2 diabetes mellitus patients [45].

A variant in the gene encoding acetyl-coenzyme A carboxylase beta (*ACACB*) was shown to be associated with proteinuria and ESRD in patients with type 2 diabetes mellitus in several Asian and European American populations. *ACACB* is an important enzyme in fatty acid oxidation and was shown to be expressed in podocytes and tubular epithelial cells in mice [46, 47].

A summary of the recent findings on DN can be found in Table 3.

CONCLUSION AND OUTLOOK

Many GWAS and meta-analyses have aimed at identifying genetic risk factors for kidney disease during the last few years. Many genetic loci have been identified and replicated for eGFR, CKD, albuminuria and distinct kidney diseases, such as non-diabetic ESRD or focal-segmental glomerulosclerosis in African Americans, idiopathic membranous nephropathy or IgA nephropathy [23, 36, 51, 59–61]. However, association data for DN were less conclusive. Although several DN associated loci have been identified, only few of them have been validated and many remained unvalidated. Even variants in one of the more promising candidate genes, ELMO1, were not significant in all the studies. Consistent with the aetiological heterogeneity of CKD, there has been little overlap in genetic markers associated with different kidney diseases—or even different measures of renal function, such as albuminuria and eGFR—which could indicate different underlying disease processes for these traits [62]. Furthermore, there is evidence that different genetic variants are involved in different stages of diseases, and that the functional effect of a genetic variant may differ depending on ethnicity and population.

In general, the GWAS have several strengths and limitations. The hypothesis-free approach of GWAS enables the identification of new genes and new genomic regions and might improve the understanding of underlying mechanisms of disease. After identifying genomic regions of interest by GWAS, it is essential to conduct follow-up studies to determine the consequences and potential clinical value of GWAS findings. Fine-mapping, gene expression data and further *in vitro* and *in vivo* experiments are necessary to identify the actual causal variant and to further investigate its molecular mechanism and biological effect. This has been done to some degree for the aforementioned *UMOD* locus, which had been found to be strongly associated with eGFR and CKD [22]. In a follow-up analysis, the presence of the *UMOD* SNP rs4293393 was found to be associated with uromodulin levels, and elevated uromodulin levels preceded the development of CKD

Table 3. GWAS and meta-analyses in diabetic kidney disease

Implicated gene	SNP	Ethnicity	Lowest P-value	Significance	Phenotype	Sample size	Reference	Year	Study type
ACACB	rs2268388	Japanese	$P = 5.35 \times 10^{-8}$	GW	T2DM proteinuria	3919	Maeda *et al.* [47]	2010	Meta-analysis
ACE I/D		White Asian	P = 0.01	Significance <0.05, only one SNP tested	T1DKD and T2DKD	26 580	Wang *et al.* [44]	2012	Meta-analysis for ACE
AFF3	rs7583877	White	$P = 1.2 \times 10^{-8}$	GW	T1DKD	12 564	Sandholm *et al.* [42]	2012	GWAS
AGTR1	rs5186	Caucasian Asian	Odds ratio = 2.11	Sign. corr.	T1DKD and T2DKD	9282	Ding *et al.* [45]	2012	Meta-analysis for renin–angiotensin–aldosterone genes
	rs12695897	African American	P = 0.032		T2DM ESRD	1984	Palmer *et al.* [48]	2014	GWAS
APOL	E2 allele E4 allele	Chinese Han	P = 0.01	Significance <0.05, only three haplotypes tested	T2DKD	7482	Yin *et al.* [49]	2014	Meta-analysis for apolipoprotein E
CDCA7-Sp3	rs4972593	White	$P = 5 \times 10^{-8}$	GW	T1DKD ESRD	7761	Sandholm *et al.* [43]	2013	GWAS; sex specific
CERS2	rs267734 rs267738	European and non-European	P = 0.0015 P = 0.0013	Sign. corr. sign. corr.	T1DM and T2DM albuminuria	3723	Shiffman *et al.* [27]	2014	GWAS
Chr 2 AC14777.4	rs7560163	African American	$P = 7 \times 10^{-9}$	GW	T2DKD	6449	Palmer *et al* [50]	2012	GWAS
CNDP1	rs4892249	African	P = 0.043	Sign. corr.	T2DM	1984	Palmer	2014	GWAS
	rs6566815	American	P = 0.0076	Sign. corr.	ESRD		*et al.* [48]		
CUBN	rs1801239	European African American	$P = 4 \times 10^{-8}$	GW	T1DM and albuminuria	31 580 6981	Boger *et al.* [51]	2011	Meta-analysis
ELMO1	rs7785934	Caucasian	$P = 3.3 \times 10^{-4}$	Sign. corr.	T1 DKD ESRD	1705	Pezzolesi *et al.* [52]	2009	Meta-analysis
	Intron 18 + 9170	Japanese	$P = 8 \times 10^{-6}$	Sign. corr.	T2 DKD	880	Shimazaki *et al.* [29]	2005	GWAS
	intron 1, 5 and 13	African American	P = 0.001–0.004	Significance: 0.0001–0.0002		1261	Leak *et al.* [31]	2009	GWAS
	rs741301	Chinese	P = 0.004	Significance <0.05, only six SNP tested		200	Wu *et al.* [53]	2013	GWAS
EPO	rs161740	White	$P = 2 \times 10^{-9}$	GW	T1DM and ESRD	7007	Williams *et al.* [54]	2012	Meta-analysis
FRMD	rs10868025		$P = 5 \times 10^{-7}$	Did not meet their calculated genome-wide sign. of 1.4×10^{-7}	T1DM proteinuria, ESRD	820	Pezzolesi *et al.* [52]	2009	Meta-analysis
	rs13288659		$P = 9.7 \times 10^{-5}$	Sign. corr.	T1DKD	6366	Williams *et al.* [54]	2012	Meta-analysis
GCKR	rs1260326		$P = 3.23 \times 10^{-3}$	Significance <0.003	T2DM eGFR	3028	Deshmukh *et al.* [26]	2013	GWAS
MTHFR	rs1801133	Caucasian Asian African Latin-American	P = 0.042	Significance <0.05, only one SNP tested	T1DKD and T2DKD	7807	Niu *et al.* [33]	2012	Meta-analysis for methylenetetrahydrofolate reductase
MYH9	rs4821480	European Americans	P = 0.053	Significance <0.05	T2DM ESRD	1963	Cooke *et al.* [40]	2012	GWAS
		African American	P = 0.0381			1903	Freedman *et al.* [39]	2009	
	rs2032487		P = 0.054			1963	Cooke *et al.* [40]	2012	
			P = 0.0449			1903	Freedman *et al.* [39]	2009	
	rs4281481		P = 0.055			1963	Cooke *et al.* [40]	2012	
			P = 0.0477			1903	Freedman *et al.* [39]	2009	

Continued

Table 3. Continued

Implicated gene	SNP	Ethnicity	Lowest P-value	Significance	Phenotype	Sample size	Reference	Year	Study type
RGMA-MCTP2	rs12437854	White	$P = 2.0 \times 10^{-9}$	GW	T1DKD	12 564	Sandholm *et al.* [42]	2012	GWAS
SHROOM3	rs17319721		$P = 3.18 \times 10^{-3}$	Significance <0.003	T2DM eGFR	3028	Deshmukh *et al.* [26]	2013	GWAS
SLC12A3	rs11643718	Japanese	P = 0.0002	Sign. corr	T2DKD	870	Tanaka *et al.* [55]	2003	GWAS
			P = 0.021	Significance <0.05, only one SNP tested	T2DM, albuminuria	264	Nishiyama *et al.* [56]	2005	Retrospective study
		Malaysian Chinese, Malay, Malaysian Indians	P = 0.038	Significance <0.05, eight SNPs tested	T2DKD	383	Seman *et al.* [57]	2014	
TGF-β1	rs1800470 (T869C)	Asian Caucasian African	P = 0.005 P = 0.04	Significance <0.05, only one SNP tested	T1DKD and T2DKD	4863	Zhou *et al.* [58]	2014	Meta-analysis for TGF-β1
UMOD	rs12917707		$P = 8.84 \times 10^{-4}$	Sign. corr.	T2DM eGFR	3028	Deshmukh *et al.* [26]	2013	GWAS

T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; T1DKD, type 1 diabetic kidney disease; T2DKD, type 2 diabetic kidney disease; sign. corr., significant after correction for multiple testing; GW, genome-wide significance; GWAS, genome-wide association studies.

[63]. Whether the findings of this rather small study can be replicated in larger study populations and eventually put into clinical application depends on the outcomes of further studies.

The biological role of CERS2 has also been investigated by two studies in CERS2-deficient mice. Both studies reported severe liver pathologies and hepatocellular carcinoma. However, only one study described discrete loss of renal parenchyma [64], whereas the other reported normal kidney morphology and function [65]. It remains to be determined whether the identified SNPs influence CERS2 activity or stability, and how CERS2 activity affects progression of albuminuria.

While the potential benefit of an increased understanding of disease is obvious, the prognostic value of the identified markers is questionable, as it is compromised by the discrepancy between the observed high heritability of traits and the small effect sizes of currently identified variants, which explain only a small proportion of the genetic risk or the variance of a trait. For example, the 16 SNPs identified by Köttgen *et al.* accounted for only 1.4% of the variability of eGFR. It is unclear why there is such a big difference between the observed heritability and the fraction of the variance explained by known genetic variants ('missing heritability'). Potentially, there are many more common variants with small effect that await discovery. Other explanations include the existence of undetected rare variants with larger effects, genetic variants other than SNPs [66]. It is also possible that the heritability has been overestimated by neglecting the contribution from gene–gene or gene–environmental interactions to a trait's variability [67].

Regardless, the small impact of known variants on disease risk does not currently allow the identification of individuals at higher risk of disease or disease progression. Whether or not the future identification of more risk variants, each of small effect, would enable the determination of individual genetic risk profile is a subject to debate [68]. However, the detection of rare or low-frequency variants, with potential large effect sizes could be achieved by sequencing genomic regions identified by GWAS. Whole-genome or, the less expensive whole-exome sequencing has been made technically feasible through the emergence of rapid next-generation sequencing and will allow the detection of novel, rare variants not captured by GWAS.

Another challenge of GWAS in kidney research is the definition of phenotypes. GFR is usually estimated by filtration markers in serum, which may also be influenced by factors other than renal function. To differentiate between genetic markers associated with renal function and those associated with biomarker metabolism, it has proven useful to estimate GFR by two different markers (creatinine and cystatin) [22].

While in many kidney diseases, the diagnosis is based on histology and is therefore quite clearly defined, the phenotype of DN is in most studies based on clinical criteria and their progression. There is no consistent definition of cases and controls across all cross-sectional studies on DN. In addition, the degree of albuminuria is known to vary and might regress to normoalbuminuria in early stages, which increases the risk of case/control misclassification [69, 70]. Longitudinal studies could reduce the risk of misclassification and are suited to capture CKD initiation, the slope of decline of renal function and progression to ESRD. They are, however, limited in their power by the moderate rate of progression of albuminuria and the small proportion of individuals progressing to ESRD.

The advantages and disadvantages of GWAS and their systematic approach are shared by other 'omics' areas. Omics refers to fields of research that deal with the collective pools of biomolecules in a cell or tissue, such as genes (*genomics*), mRNA (*transcriptomics*), protein (*proteomics*) or metabolites (*metabolomics*). An interesting future prospect of GWAS could be found in integrating their results with data of other 'omics' areas to create and analyse larger models.

In conclusion, GWAS in kidney disease have been successful in reproducibly identifying genetic variants and loci for some distinct kidney diseases and renal function. For diabetic

kidney disease, several loci have been identified and validated, but the results were quite heterogenic across different populations and depended on the type of diabetes and stage of disease.

The major benefit of GWAS results is to be found in the increased understanding of disease mechanism and identification of novel pathways and possibly new therapeutic targets. Follow-up studies are important in order to identify variants with specific biological effect and may provide important insight for some identified variants. Given the small effect size of known variants on disease risk, the potential for personalized risk prediction is currently low.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Coresh J, Selvin E, Stevens LA *et al.* Prevalence of chronic kidney disease in the United States. JAMA 2007; 298: 2038–2047
2. Eckardt KU, Coresh J, Devuyst O *et al.* Evolving importance of kidney disease: from subspecialty to global health burden. Lancet 2013; 382: 158–169
3. Gross JL, de Azevedo MJ, Silveiro SP *et al.* Diabetic nephropathy: diagnosis, prevention, and treatment. Diabetes Care 2005; 28: 164–176
4. Adler AI, Stevens RJ, Manley SE *et al.* Development and progression of nephropathy in type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS 64). Kidney Int 2003; 63: 225–232
5. Macisaac RJ, Ekinci EI, Jerums G. Markers of and risk factors for the development and progression of diabetic kidney disease. Am J Kidney Dis 2014; 63(2 Suppl 2): S39–S62
6. Satko SG, Sedor JR, Iyengar SK *et al.* Familial clustering of chronic kidney disease. Semin Dial 2007; 20: 229–236
7. O'Seaghdha CM, Fox CS. Genetics of chronic kidney disease. Nephron Clin Pract 2011; 118: c55–c63
8. International HapMap C. The International HapMap Project. Nature 2003; 426: 789–796
9. Igarashi P, Somlo S. Genetics and pathogenesis of polycystic kidney disease. J Am Soc Nephrol 2002; 13: 2384–2398
10. Bush WS, Moore JH. Chapter 11: genome-wide association studies. PLoS Comput Biol 2012; 8: e1002822
11. Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 2006; 7: 385–394
12. Boger CA, Heid IM. Chronic kidney disease: novel insights from genome-wide association studies. Kidney Blood Press Res 2011; 34: 225–234
13. Li Y, Shiffman D, Oberbauer R. Analysis of single nucleotide polymorphisms in case–control studies. Methods Mol Biol 2011; 719: 219–234
14. Turner S, Armstrong LL, Bradford Y *et al.* Quality control procedures for genome-wide association studies. Curr Protoc Hum Genet 2011; Supplement 68, Unit 1.19, p1.19.1–1.19.8
15. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy–Weinberg equilibrium. Am J Hum Genet 2005; 76: 967–986
16. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. Nat Rev Genet 2009; 10: 681–690
17. Pe'er I, Yelensky R, Altshuler D *et al.* Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008; 32: 381–385
18. Hoggart CJ, Clark TG, De Iorio M *et al.* Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 2008; 32: 179–185
19. Mayer G, Heinze G, Mischak H *et al.* Omics-bioinformatics in the context of clinical data. Methods Mol Biol 2011; 719: 479–497
20. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet 2014; 15: 335–346
21. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case–control data. Am J Hum Genet 2007; 80: 605–615
22. Köttgen A, Glazer NL, Dehghan A *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet 2009; 41: 712–717
23. Köttgen A, Pattaro C, Boger CA *et al.* New loci associated with kidney function and chronic kidney disease. Nat Genet 2010; 42: 376–384
24. Hart TC, Gorry MC, Hart PS *et al.* Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. J Med Genet 2002; 39: 882–892
25. Serafini-Cessi F, Malagolini N, Cavallone D. Tamm–Horsfall glycoprotein: biology and clinical relevance. Am J Kidney Dis 2003; 42: 658–676
26. Deshmukh HA, Palmer CN, Morris AD *et al.* Investigation of known estimated glomerular filtration rate loci in patients with type 2 diabetes. Diabet Med 2013; 30: 1230–1235
27. Shiffman D, Pare G, Oberbauer R *et al.* A gene variant in CERS2 Is associated with rate of increase in albuminuria in patients with diabetes from ONTARGET and TRANSCEND. PLoS ONE 2014; 9: e106631
28. Investigators OtGilbert RE, Mann JF, Hanefeld M *et al.* Basal insulin glargine and microvascular outcomes in dysglycaemic individuals: results of the Outcome Reduction with an Initial Glargine Intervention (ORIGIN) trial. Diabetologia 2014; 57: 1325–1331
29. Shimazaki A, Kawamura Y, Kanazawa A *et al.* Genetic variations in the gene encoding ELMO1 are associated with susceptibility to diabetic nephropathy. Diabetes 2005; 54: 1171–1178
30. Pezzolesi MG, Katavetin P, Kure M *et al.* Confirmation of genetic associations at ELMO1 in the GoKinD collection supports its role as a susceptibility gene in diabetic nephropathy. Diabetes 2009; 58: 2698–2702
31. Leak TS, Perlegas PS, Smith SG *et al.* Variants in intron 13 of the ELMO1 gene are associated with diabetic nephropathy in African Americans. Ann Hum Genet 2009; 73: 152–159
32. Shimazaki A, Tanaka Y, Shinosaki T *et al.* ELMO1 increases expression of extracellular matrix proteins and inhibits cell adhesion to ECMs. Kidney Int 2006; 70: 1769–1776
33. Niu W, Qi Y. An updated meta-analysis of methylenetetrahydrofolate reductase gene 677C/T polymorphism with diabetic nephropathy and diabetic retinopathy. Diabetes Res Clin Pract 2012; 95: 110–118
34. Wrone EM, Zehnder JL, Hornberger JM *et al.* An MTHFR variant, homocysteine, and cardiovascular comorbidity in renal disease. Kidney Int 2001; 60: 1106–1113
35. Kopp JB, Smith MW, Nelson GW *et al.* MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. Nat Genet 2008; 40: 1175–1184
36. Kao WH, Klag MJ, Meoni LA *et al.* MYH9 is associated with nondiabetic end-stage renal disease in African Americans. Nat Genet 2008; 40: 1185–1192
37. Genovese G, Friedman DJ, Ross MD *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 2010; 329: 841–845
38. Parsa A, Kao WH, Xie D *et al.* APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med 2013; 369: 2183–2196
39. Freedman BI, Hicks PJ, Bostrom MA *et al.* Non-muscle myosin heavy chain 9 gene MYH9 associations in African Americans with clinically diagnosed type 2 diabetes mellitus-associated ESRD. Nephrol Dial Transplant 2009; 24: 3366–3371
40. Cooke JN, Bostrom MA, Hicks PJ *et al.* Polymorphisms in MYH9 are associated with diabetic nephropathy in European Americans. Nephrol Dial Transplant 2012; 27: 1505–1511
41. McKnight AJ, Duffy S, Fogarty DG *et al.* Association of MYH9/APOL1 with chronic kidney disease in a UK population. Nephrol Dial Transplant 2012; 27: 3660; author reply 660–661
42. Sandholm N, Salem RM, McKnight AJ *et al.* New susceptibility loci associated with kidney disease in type 1 diabetes. PLoS Genet 2012; 8: e1002921
43. Sandholm N, McKnight AJ, Salem RM *et al.* Chromosome 2q31.1 associates with ESRD in women with type 1 diabetes. J Am Soc Nephrol 2013; 24: 1537–1543
44. Wang F, Fang Q, Yu N *et al.* Association between genetic polymorphism of the angiotensin-converting enzyme and diabetic nephropathy: a

meta-analysis comprising 26,580 subjects. J Renin Angiotensin Aldosterone Syst 2012; 13: 161–174
45. Ding W, Wang F, Fang Q *et al.* Association between two genetic polymorphisms of the renin-angiotensin–aldosterone system and diabetic nephropathy: a meta-analysis. Mol Biol Rep 2012; 39: 1293–1303
46. Tang SC, Leung VT, Chan LY *et al.* The acetyl-coenzyme A carboxylase beta (ACACB) gene is associated with nephropathy in Chinese patients with type 2 diabetes. Nephrol Dial Transplant 2010; 25: 3931–3934
47. Maeda S, Kobayashi MA, Araki S *et al.* A single nucleotide polymorphism within the acetyl-coenzyme A carboxylase beta gene is associated with proteinuria in patients with type 2 diabetes. PLoS Genet 2010; 6: e1000842
48. Palmer ND, Ng MC, Hicks PJ *et al.* Evaluation of candidate nephropathy susceptibility genes in a genome-wide association study of African American diabetic kidney disease. PLoS ONE 2014; 9: e88273
49. Yin YW, Qiao L, Sun QQ *et al.* Influence of apolipoprotein E gene polymorphism on development of type 2 diabetes mellitus in Chinese Han population: a meta-analysis of 29 studies. Metabolism 2014; 63: 532–541
50. Palmer ND, McDonough CW, Hicks PJ *et al.* A genome-wide association search for type 2 diabetes genes in African Americans. PLoS ONE 2012; 7: e29202
51. Boger CA, Chen MH, Tin A *et al.* CUBN is a gene locus for albuminuria. J Am Soc Nephrol 2011; 22: 555–570
52. Pezzolesi MG, Poznik GD, Mychaleckyj JC *et al.* Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. Diabetes 2009; 58: 1403–1410
53. Wu HY, Wang Y, Chen M *et al.* Association of ELMO1 gene polymorphisms with diabetic nephropathy in Chinese population. J Endocrinol Invest 2013; 36: 298–302
54. Williams WW, Salem RM, McKnight AJ *et al.* Association testing of previously reported variants in a large case–control meta-analysis of diabetic nephropathy. Diabetes 2012; 61: 2187–2194
55. Tanaka N, Babazono T, Saito S *et al.* Association of solute carrier family 12 (sodium/chloride) member 3 with diabetic nephropathy, identified by genome-wide analyses of single nucleotide polymorphisms. Diabetes 2003; 52: 2848–2853
56. Nishiyama K, Tanaka Y, Nakajima K *et al.* Polymorphism of the solute carrier family 12 (sodium/chloride transporters) member 3, SLC12A3, gene at exon 23 (+78G/A: Arg913Gln) is associated with elevation of urinary albumin excretion in Japanese patients with type 2 diabetes: a 10-year longitudinal study. Diabetologia 2005; 48: 1335–1338
57. Abu Seman N, He B, Ojala JR *et al.* Genetic and biological effects of sodium-chloride cotransporter (SLC12A3) in diabetic nephropathy. Am J Nephrol 2014; 40: 408–416
58. Zhou TB, Jiang ZP, Qin YH *et al.* Association of transforming growth factor-beta1 T869C gene polymorphism with diabetic nephropathy risk. *Nephrology* (*Carlton)* 2014; 19: 107–115
59. Stanescu HC, Arcos-Burgos M, Medlar A *et al.* Risk HLA-DQA1 and PLA (2)R1 alleles in idiopathic membranous nephropathy. N Engl J Med 2011; 364: 616–626
60. Genovese G, Tonna SJ, Knob AU *et al.* A risk allele for focal segmental glomerulosclerosis in African Americans is located within a region containing APOL1 and MYH9. Kidney Int 2010; 78: 698–704
61. Gharavi AG, Kiryluk K, Choi M *et al.* Genome-wide association study identifies susceptibility loci for IgA nephropathy. Nat Genet 2011; 43: 321–327
62. Placha G, Canani LH, Warram JH *et al.* Evidence for different susceptibility genes for proteinuria and ESRD in type 2 diabetes. Adv Chronic Kidney Dis 2005; 12: 155–169
63. Köttgen A, Hwang SJ, Larson MG *et al.* Uromodulin levels associate with a common UMOD variant and risk for incident CKD. J Am Soc Nephrol 2010; 21: 337–344
64. Imgrund S, Hartmann D, Farwanah H *et al.* Adult ceramide synthase 2 (CERS2)-deficient mice exhibit myelin sheath defects, cerebellar degeneration, and hepatocarcinomas. J Biol Chem 2009; 284: 33549–33560
65. Pewzner-Jung Y, Brenner O, Braun S *et al.* A critical role for ceramide synthase 2 in liver homeostasis: II. Insights into molecular changes leading to hepatopathy. J Biol Chem 2010; 285: 10911–10923
66. Manolio TA, Collins FS, Cox NJ *et al.* Finding the missing heritability of complex diseases. Nature 2009; 461: 747–753
67. Zuk O, Hechter E, Sunyaev SR *et al.* The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci USA 2012; 109: 1193–1198
68. Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. PLoS Genet 2009; 5: e1000540
69. Macisaac RJ, Jerums G. Diabetic kidney disease with and without albuminuria. Curr Opin Nephrol Hypertens 2011; 20: 246–257
70. Steinke JM, Sinaiko AR, Kramer MS *et al.* The early natural history of nephropathy in Type 1 diabetes: III. Predictors of 5-year urinary albumin excretion rate patterns in initially normoalbuminuric patients. Diabetes 2005; 54: 2164–2171

Received for publication: 28.1.2015; Accepted in revised form: 9.3.2015