Novel insights from genetic and epigenetic studies in understanding the complex uraemic phenotype

Anna Witasp1,2, Tomas J. Ekström2,3, Bengt Lindholm1, Peter Stenvinkel1, Martin Schalling2,4 and Louise Nordfors1,2,4

1Divisions of Renal Medicine and Baxter Novum, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden, 2Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden, 3Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden and 4Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

Correspondence and offprint requests to: Peter Stenvinkel; E-mail: peter.stenvinkel@ki.se

ABSTRACT

Like in many other common complex disorders, studies of chronic kidney disease (CKD) can now make use of the increasing knowledge of the human genome, its variations and impact on disease susceptibility, initiation, progression and complications. Such studies are facilitated by novel readily available high through-put genotyping methods and sophisticated analytical approaches to scan the genome for DNA variations and epigenetic modifications. Here, we review some of the recent discoveries that have emerged from these studies and expanded our knowledge of genetic risk loci and epigenetic markers in CKD pathophysiology. Obstacles and practical issues in this field are discussed.

Keywords: chronic kidney disease, ethnicity, epigenetics, genetics, GWAS

INTRODUCTION

Common complex disorders such as chronic kidney disease (CKD) are influenced by environmental triggers and by multiple common genetic components [1]. Whereas we were limited to the study of rare chromosomal rearrangements and a handful monogenic disease genes only a decade ago, the study of genetic factors behind multifactorial diseases has had an exponential upswing with the launch in 2001 of the sequence of the human genome [2, 3] and, 3 years later, the HapMap database of common human DNA variations [4, 5]. To date, over 30 million single nucleotide polymorphisms (SNPs) have been discovered, most of which can be found in The National Center for Biotechnology Information’s public repository database of SNPs (dbSNP, http://www.ncbi.nlm.nih.gov/snp) together with multiple small-scale variations (e.g. insertions/deletions, microsatellites) and rare variants.

As a consequence, analysis of SNPs covering large parts of the genome has thus been made possible, often in semi-automated chip-based format. Genome-wide association studies (GWASs) now enable screening of the genome for associations with a quantitative trait or disease of interest in an unbiased manner. The first successful GWAS (on age-related macular degeneration) in 2005 included only 96 cases and 50 controls with 116 204 SNPs genotyped [6], while current gene chips may cover up to 5 million SNPs, and sample sizes may reach 2 000 000 individuals. During the past decade, GWASs have become instrumental in the study of complex, polygenic traits [7]. Numerous associations between genetic variants and phenotypes, including renal impairment and CKD, have been established and previously unappreciated disease pathways and mechanisms have been illuminated [8]. This review summarizes some of the findings from GWASs, focusing on the risk and manifestations of CKD. It briefly discusses some methodological restraints in renal populations, as well as the emerging role of epigenetics in the complex uraemic milieu.

NOVEL INSIGHTS FROM GWASs ON MEASURES OF KIDNEY DISEASE

Genetic determinants of renal function and CKD

GWASs have identified and independently replicated more than 50 genetic variants that are associated with indices of renal function, such as estimated glomerular filtration rate (eGFR) or other signs of CKD [8] (Table 1). The first GWAS on eGFR, based on creatinine (eGFR_crea), cystatin C (eGFR_cys) and CKD (eGFR_crea < 60 mL/min/1.73 m²), launched in 2009, was set up...
### Table 1. Summary of key findings from GWAs on measures of CKD

<table>
<thead>
<tr>
<th>SNP (dbSNP number)</th>
<th>Chromosome</th>
<th>In gene/nearest gene</th>
<th>Renal phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12917707</td>
<td>16</td>
<td>UMOD</td>
<td>CKD, ESRD</td>
<td>Köttgen et al. [9], Chambers et al. [12], Böger et al. [14], Liu et al. [25]</td>
</tr>
<tr>
<td>rs17319721</td>
<td>4</td>
<td>SHROOM3</td>
<td>eGFRcrea, UACR</td>
<td>Köttgen et al. [9, 13], Chambers et al. [12], Ellis et al. [15]</td>
</tr>
<tr>
<td>rs2467853</td>
<td>15</td>
<td>GATM-SPTA5L1</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [9, 13], Chambers et al. [12], Ellis et al. [15]</td>
</tr>
<tr>
<td>rs13038305</td>
<td>20</td>
<td>CST3-CST9</td>
<td>eGFRcrea, CKD</td>
<td>Köttgen et al. [9, 13], Chambers et al. [12], Ellis et al. [15]</td>
</tr>
<tr>
<td>rs17131274; rs10109414</td>
<td>8</td>
<td>STC1</td>
<td>eGFRcrea, CKD</td>
<td>Köttgen et al. [9, 13], Chambers et al. [12], Ellis et al. [15]</td>
</tr>
<tr>
<td>rs10206899; rs13538</td>
<td>2</td>
<td>ALMS1/NAT8</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Chambers et al. [12], Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs267734</td>
<td>1</td>
<td>LASS2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs1260326</td>
<td>2</td>
<td>GCKR</td>
<td>eGFRcrea, eGFRcyst, CKD, ESRD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs347685</td>
<td>3</td>
<td>TFPDP2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs11959928</td>
<td>5</td>
<td>DAB2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs6420099</td>
<td>5</td>
<td>SLC34A1</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs881858</td>
<td>6</td>
<td>VEGFA</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs7805747</td>
<td>7</td>
<td>PRKAG2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs4747412</td>
<td>9</td>
<td>PIP5K1B</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs653178</td>
<td>12</td>
<td>ATXN2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs626277</td>
<td>13</td>
<td>DACH1</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs1394125</td>
<td>15</td>
<td>UBE2Q2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs4805834; rs12460876</td>
<td>19</td>
<td>SLC7A9</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Köttgen et al. [13]</td>
</tr>
<tr>
<td>rs7422339</td>
<td>2</td>
<td>CPS1</td>
<td>eGFRcrea</td>
<td>Chambers et al. [12], Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs3127573; rs2279463</td>
<td>6</td>
<td>SLC22A2</td>
<td>eGFRcrea</td>
<td>Chambers et al. [12], Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs6465825</td>
<td>7</td>
<td>TMEM60</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs10794720</td>
<td>10</td>
<td>WDR37</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs10747021</td>
<td>12</td>
<td>SLC6A13</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs491567</td>
<td>15</td>
<td>WDR72</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs9895666</td>
<td>17</td>
<td>BCAS3</td>
<td>eGFRcrea</td>
<td>Köttgen et al. [13], Liu et al. [25]</td>
</tr>
<tr>
<td>rs267734</td>
<td>1</td>
<td>ANXA9</td>
<td>eGFRcrea, CKD</td>
<td>Böger et al. [14], Liu et al. [25]</td>
</tr>
<tr>
<td>rs12278026</td>
<td>11</td>
<td>MPPED2</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
<tr>
<td>rs6431731</td>
<td>2</td>
<td>DDX1</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
<tr>
<td>rs12124078</td>
<td>17</td>
<td>SLC7A41</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
<tr>
<td>rs11078903</td>
<td>17</td>
<td>CDK12</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
<tr>
<td>rs2453580</td>
<td>1</td>
<td>CASP9</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
<tr>
<td>rs2928148</td>
<td>15</td>
<td>INO80</td>
<td>eGFRcrea, eGFRcyst, CKD</td>
<td>Pattaro et al. [16]</td>
</tr>
</tbody>
</table>

**GWASs on DN**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Renal phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1888746; rs13289150</td>
<td>9</td>
<td>T1DM and T2DM DN</td>
<td>Pezzolesi et al. [17]</td>
</tr>
<tr>
<td>rs739401; rs451041</td>
<td>11</td>
<td>T1DM DN</td>
<td>Pezzolesi et al. [17]</td>
</tr>
<tr>
<td>multiple</td>
<td>7</td>
<td>T1DM and T2DM DN</td>
<td>Shimazaki et al. [21], Leak et al. [22], Pezzolesi et al. [23]</td>
</tr>
<tr>
<td>rs7588377</td>
<td>2</td>
<td>T1DM DN</td>
<td>Sandholm et al. [24]</td>
</tr>
<tr>
<td>rs12437854</td>
<td>15</td>
<td>RGMA-MCTP2</td>
<td>Sandholm et al. [24]</td>
</tr>
</tbody>
</table>

**GWASs on Asian and African populations**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Renal phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>multiple</td>
<td>22</td>
<td>MYH9</td>
<td>ESRD, FSG</td>
</tr>
<tr>
<td>multiple</td>
<td>22</td>
<td>APOL1</td>
<td>ESRD</td>
</tr>
<tr>
<td>rs7111394</td>
<td>11</td>
<td>KCNQ1</td>
<td>eGFRcrea</td>
</tr>
<tr>
<td>rs8555246</td>
<td>18</td>
<td>DOK6</td>
<td>UACR</td>
</tr>
<tr>
<td>rs2880072</td>
<td>6</td>
<td>FNDC1</td>
<td>UACR</td>
</tr>
<tr>
<td>rs3828890</td>
<td>6</td>
<td>MHC region</td>
<td>eGFRcrea</td>
</tr>
<tr>
<td>rs10277115</td>
<td>7</td>
<td>UNCX</td>
<td>eGFRcrea</td>
</tr>
<tr>
<td>rs963837</td>
<td>11</td>
<td>MPPED2-DCDCS</td>
<td>eGFRcrea</td>
</tr>
</tbody>
</table>

CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; DN, diabetic nephropathy; FSG, focal segmental glomerulosclerosis; T1DM/T2DM, Type 1/Type 2 diabetes mellitus; UACR, urine albumin-to-creatinine ratio.

as a meta-analysis by the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium and included 19 877 individuals [9]. One key finding was a previously unrecognized association between one SNP (rs12917707; minor allele present in 18% of the study population) within the UMOD gene (encoding Tamm–Horsfall glycoprotein, also known as uromodulin) and CKD (odds ratio 0.76; P-value = 2.8 × 10⁻⁹) and eGFRcrea (β = 0.022; P = 3.0 × 10⁻¹¹). The protein, which is involved in inflammation and infection, is in fact the most abundant protein in normal urine [10] and rare variants in the UMOD gene are known to cause severe monogenic renal disorders; i.e. medullary cystic kidney disease 2 and familial juvenile hyperuricaemic
nephropathy [11]. Moreover, two additional eGFRcys associated loci, located in the SHROOM3 and GATM genes, as well as two eGFRcyst loci in/near the CST and STC1 genes were identified in this pioneering renal GWAS [9].

Later, a European GWAS including 23 812 participants reported associations between 109 SNPs distributed over five loci, including the previously reported SHROOM3, UMOD and GATM, and serum creatinine [12]. Two of the novel top-ranking SNPs, rs10206899 (near NAT8) and rs4805834 (near SLC7A9), were also shown to associate with CKD (odds ratios 0.85 and 0.84, respectively) in a follow-up analysis of an independent sample set. Concurrently, a GWAS meta-analysis conducted by the CKDGen Consortium including 67 093 individuals uncovered about 20 additional genome-wide significant loci associated with eGFR and CKD [13]. A follow-up study on these findings confirmed 13 novel loci, in or near LASS2, GCKR, ALMS1, TFDP2, DAB2, SLC34A1, VEGFA, PRKAG2, PIP5K1B, ATXN2, DACH1, UBE2Q2 and SLC7A9, which were associated with renal function (accounting for 1.4% of the variation in eGFRcrea) and CKD (odds ratios ranging from 0.93 to 1.19), and 7 loci, in or near CPS1, SLC22A2, TMEM60, WDR37, SLC6A13, WDR72 and BCAS3, which were associated with eGFRcrea, but not eGFRcyst. Altogether, these GWASs introduced large-scale genetic studies to the nephrology community and illuminated previously unrecognized metabolic, solute and drug-transport pathways as well as mechanisms for nephrogenesis, podocyte function and angiogenesis that confer susceptibility to CKD.

**Extended and stratified analyses of renal phenotypes**

Based on the results from these previous GWASs, 16 eGFR-related loci were selected and subjected to extended analyses of possible associations with incident CKD, ESRD, urine albumin-to-creatinine ratio (UACR) and albuminuria. These studies included longitudinal and cross-sectional cohort data, which were largely derived from the CKDGen consortium [14, 15]; four European case–control study samples [14]; and data from the Candidate-gene Association Resource (CARe) Renal Consortium (African ancestry) [15]. Six loci (UMOD, PRKAG2, ANXA9, DAB2, DACH1 and STC1) associated with incident CKD (odds ratios ranging from 0.76 to 1.19); two SNPs in UMOD (odds ratio 0.92) and GCKR (odds ratio 0.93) were associated with ESRD [14]; and one eGFR SNP (SHROOM3) was found to associate with UACR (β = −0.034, P = 0.0002), but not with albuminuria [15]. Thus, incident CKD seems to be influenced by eGFR-related loci to a greater extent than ESRD, whereas genetic susceptibility to UACR or albuminuria is most likely involving variants that are yet to be identified and distinct from these investigated eGFR-related loci.

In a GWAS published in 2012, Pattaro et al. [16] included more than 130 000 participants of European descent (CKDGen Consortium) and performed stratification for key risk factors such as age, sex, hypertension and diabetes. This study uncovered six new loci, in or near MMPED2, DDX1, SLC47A1, CDK12, CASP9 and INO80, for kidney function (assessed by both eGFRcrea and eGFRcyst) and CKD (odds ratios ranging from 0.98 to 1.05). The authors also found that whereas younger individuals showed a stronger association with the CDK12 SNP compared with older individuals, most of the identified SNPs showed no cross-strata differences. Although it is largely unknown whether the association between genetic risk loci and renal phenotypes may be differentially influenced by the presence of various CKD risk factors, this study implies that genetic impact on kidney function is weakly, or not at all, modified by traditional CKD risk factors. Accordingly, it appears likely that many eGFR-associated loci are generalizable beyond existing risk factor strata. Clearly, such information is of specific interest in studies of CKD populations with significant pathophysiological heterogeneity.

**GWASs on diabetic nephropathy**

Extensive GWAS data have emerged also from diabetic nephropathy (DN) populations with both Type 1 and 2 diabetes mellitus (T1DM and T2DM), including the Genetics of Kidneys in Diabetes (GoKinD) collection of 820 cases (284 with proteinuria and 536 with ESRD) and 885 control subjects, all with T1DM [17]. Following replication in 1304 Type 1 diabetic participants of the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) study, a total of 13 SNPs located in four genomic loci were found to associate with DN. The novel candidate loci were located near the FMRD3 (4.1 protein ezrin, radixin, moesin [FERM] domain containing 3) and CARS (cytosine-tRNA synthetase) genes, with odds ratios of 1.45 and 1.36, respectively, and implicated previously unsuspected pathogenic pathways. Interestingly, both FMRD3 and CARS are expressed in human kidney and, further, FMRD3 variants appear to impact susceptibility to Type 1 and 2 DN in populations with African ancestry [18], but not in Chinese [19] and Japanese [20] populations. In contrast, both a Japanese [21] and an African American [22] GWAS demonstrated associations between T2DM DN and SNPs in the ELMO1 gene. Although these findings were not replicated in the GoKinD collection of European T1DM individuals (including normoalbuminuric controls and advanced DN cases), novel ELMO1 SNPs (rs11769038 and rs1882080; odds ratio of 1.24 and 1.23, respectively), at different positions from the ones previously reported, were discovered [23]. Additionally, the Genetics of Nephropathy: an International Effort consortium recently merged three collections on T1DM DN individuals (All Ireland, Warren 3, Genetics of Kidneys in Diabetes UK; Finnish DN Study and the GoKinD US study) into a large GWAS meta-analysis, in which two new susceptibility loci associated with ESRD and one locus suggestively associated with DN in T1DM were discovered [24]. The ESRD-associated SNPs were located in the AFF3 gene and between RGMa and MCTP2 (on chromosome 15q26), respectively, and functional follow-up analyses of the associated AFF3 variant implied a role in renal tubule fibrosis, thus highlighting novel mechanisms in the pathogenesis of DN.

**GWASs in Asian and African populations**

To uncover further loci for kidney traits, large-scale studies have been performed in African Americans as well as in Asian populations. More than 8000 participants in the CARe Consortium with African ancestry have been included in a recently performed meta-analysis on eGFR, CKD (eGFR < 60 mL/min/1.73 m²), UACR and microalbuminuria [25]. Both GWAS and candidate gene-based array data were
examined, focusing on the 250 kb flanking region around 24 SNPs previously identified in the European Ancestry renal GWAS analyses. Remarkably, 23 of the 24 SNPs were associated with eGFR in African Americans following replication studies. Furthermore, discovery analyses yielded three novel loci, including KCNQ1 in association with eGFR and DOK6 and FNDC1 in association with UACR.

In Asian populations, the HUGO Pan-Asian SNP project [26] has further broadened the current knowledge on risk variant load and ethnicity. In 2012 [27], a GWAS meta-analysis, including 71 149 East Asian individuals derived from the Asian Genetic Epidemiology Network, identified associations between 17 loci and kidney function-related traits. These findings included variants associated with the concentrations of blood urea nitrogen (MTX1–GBA, PAX8, MECOM, UNCX, MPPED2–DCDC5, C12orf51, WDR72, BCAS3 and GNAS), uric acid (MAF at 16q23), serum creatinine (the major histocompatibility (MHC) region, UNCX, MPPED2–DCDC5 and ALDH2) and eGFRcrea (the MHC region, UNCX and MPPED2–DCDC5). Together, these 17 loci explained 1.3, 0.54, 0.55 and 2.3% of variation in blood urea nitrogen, serum creatinine, eGFRcrea and uric acid, respectively. In silico replication of these loci in individuals of European ancestry from the KidneyGen, CKDGen and GUCG consortia, identified multiple overlap: 9 of the 15 loci that reached P < 5.0 × 10^-8 for serum creatinine, eGFRcrea and uric acid measures also showed significant associations in the European samples [27]. Hence, despite a high degree of resemblance, these studies highlight fundamental ethnic differences in genetic susceptibility in renal disease, which may have important consequences for an individual’s disease progress and prognosis.

**ETHNIC DIFFERENCES IN GENETIC RISK FOR RENAL DISEASE AND COMPLICATIONS**

For a number of kidney diseases, the prevalence, progression and outcome vary substantially between individuals with different ethnic descents. Such ancestry disparities are evident in the North American population, where African Americans are at a much greater risk of developing progressive kidney disease and to progress to ESRD than European Americans [28]. The findings from admixture-mapping genome scans (i.e. mapping by admixture linkage disequilibrium, MALD) in African American individuals suggested that variants positioned in the myosin heavy chain type II isoform A (MYH9) gene, expressed in kidney podocytes, may play a role. The reported genetic variants at the Chromosome 22 MYH9 locus correlated with a dramatically increased burden of non-diabetic ESRD [29], including focal segmental glomerulosclerosis and hypertensive ESRD [30] in African Americans, with extremely high odds ratios ranging up to 8. The correlation between MYH9 locus variants and T2DM-associated ESRD was less pronounced (odds ratio 1.2–1.4) [31]. Compellingly, it was shown that each copy of the European ancestral allele conferred a lowered relative risk (0.5) for non-diabetic ESRD, but not diabetic ESRD, compared with the African ancestry allele [29]. Associations were also found between another MYH9 variant, rs4821480, and CKD in a European study sample [32]. Later, even more powerful ESRD candidate risk loci were unravelled in the apolipoprotein L1 (APOL1) gene, which is located close to MYH9 on Chromosome 22 [33–35]. These adjacent APOL1 variants, designated the G1 and G2 alleles, were found to generate a substantially stronger statistical association than those in MYH9 in African Americans [33], but did not explain disease susceptibility in individuals with European descent, where they are virtually absent [32]. It is fascinating that these kidney disease risk alleles may have provided an important evolutionary survival advantage against trypanosomiasis (sleeping sickness) and are enriched in populations originating from regions of sub-Saharan Africa [33]; about 50% of African Americans carry at least one risk allele [34].

This set of studies has been imperative for the current knowledge about ethnic differences in genetic risk loci for renal disease susceptibility and complications and is particularly important to consider when making therapeutic choices. Even when the clinical diagnosis is the same there may be important clinical and histological differences between patients with different ethnicities, which may lead to considerable variations in the efficiency of therapeutic interventions and their ability to halt progression. Because differences in geographic origin correlate with differences in genetic variation, and these genetic factors contribute to variations in susceptibility to disease and the rate of renal function decline observed between patient subgroups, even when given the same treatment, genetic information may be implemented to develop personalized treatment regimes in renal disease (as will be discussed in a sequential review from our group).

**CONCERNS IN RENAL GWASs**

The key ingredients for a successful genetic association study are sufficiently large human study samples and complete, robust molecular measures and clinical observations. The transition to GWASs has posed an even greater size challenge due to the large amount of tested SNPs with presumably low effects. Insufficient sample sizes result in underpowered studies, in which the risk of making both Type I and II errors is substantial. The solution to this problem should be the formation of international research consortia, and within the Nephrology community several successful multi-center initiatives have resulted in the identification of important genetic associations. However, the gene variants identified by current GWASs contribute to a disappointingly low fraction of the estimated disease heritability. In the CKD population, only 2% of the estimated eGFR heritability is explained by the findings made so far [13]. This is being referred to as the missing heritability of genetic association studies and multiple issues have been raised to explain this. Nevertheless, identification of loci with very modest relative risk has repeatedly pointed towards previously undiscovered aetiopathological pathways, effectively opening up completely novel lines of research and discovery in renal research. As is evident from the APOL1 MYH9 results discussed above [32, 33], failure to define the population according to the geographical origin may in some cases dramatically reduce the strength of the finding.
**Phenotypic heterogeneity**

Genetic association studies on complex diseases, such as CKD, with a broad phenotypic spectrum with many clinical sub-phenotypes and large inter-patient variation run a significant risk of generating inconsistent or even uninformative results. A number of strategies to reduce confounding by phenotypic heterogeneity are currently available. These include the use of more narrowly defined intermediate phenotypes (also called endophenotypes) instead of the clinical disease phenotype. For example, blood lipids or proinsulin levels have been exploited to find susceptibility loci for cardiovascular disease and T2DM [36, 37]. Other means to circumvent this issue is to apply a case–case approach to identify correlations between a genetic risk factor and specific disease subtype, such as non-diabetic ESRD (see above). In the setting of CKD, a case–case design could be useful in comparing sites of genetic variations between persistently inflamed patients versus non-inflamed patients or in patients with history of CVD versus patients without history of CVD. Alternatively, analysing phenotypic extremes by data stratification, using different cut-offs from the tails of quantitative traits, may offer superior power when marked heterogeneity is present [38].

This has been observed, e.g. in a study of T2DM patients with a great inter-individual variation in body mass index (BMI) [39]. When the authors stratified the T2DM cases by BMI, they found not only additional risk variants, but also that lean patients may have a stronger genetic predisposition to T2DM. Further, GWAS results suggest that loci that are identified via selection of phenotypic extremes are generalizable to the general population [40]. In this way, a case-only design is able to identify genetic differences between different disease subtypes/categories within the patient group and may generate a more comprehensive risk factor profile. As the case–case design increases the statistical power over more classic case–control designs, it is particularly suitable for discovery GWASs aiming at identifying the strongest genetic factors for the tested phenotype and, consequently, could prove to be a cost-effective alternative to classical cohort or case–control GWASs. Finally, the study by Freedman et al. [18] on APOL1 G1/G2 risk variants and the MYH9 E1 risk haplotype in T2DM-ESRD African Americans proposes an elegant application of a case-only design to stratify for known genetic risk factors.

**Selection of well-matched control groups**

Another essential argument for implementing a case-only design is that the problem of recruiting adequate control individuals may be circumvented. In a typical case–control design, the crude inclusion criteria of the control subjects are based upon the collection of an unselected cohort with a very basic matching of age, gender and ethnic origin from the same geographical uptake area as the patient recruitment. The criteria may also include the absence of the studied phenotype (or a family history) effectively making the controls ‘super controls’. Chronically ill patients, such as CKD patients, are however burdened not only with the kidney disease *per se*, but also by a variety of stress factors that are inherent to the disease, including co-morbid complications, medications as well as disease-associated effects on everyday life, economic situation and other psychosocial factors. Thus, it is plausible to reason that these variables create a background noise that may mask the core problem/biology, either depleting or reducing the magnitude of true genotype–phenotype signals or creating spurious associations. Conversely, it is possible that uraemia may serve to reveal clinically relevant genetic differences that would otherwise be hidden behind the ‘environmental noise’. In this context, uraemia may be looked at as an environmental stressor that unmasks pertinent genetic risk factors for, i.e. inflammation or CVD that would otherwise not have been observable (i.e. phenotypic plasticity). Thus, using a case-only approach instead of the classical case–control approach may eliminate unknown confounding factors and in fact enhance our ability to detect otherwise hidden genetic risk factors.

**Late-stage CKD and dialysis patients in genetic analyses**

As a majority of North-American CKD patients die during the course of the progression of the disease and therefore never reach ESRD [41], dialysis patients could be considered as a group of survivors, who potentially harbour a unique set of survival enhancing genetic variants or other characteristics compared with other CKD populations [42]. It may very well be that genetic variants present at a higher frequency in late-stage compared with early-stage CKD, or in healthy controls, are protective rather than conferring increased risk. The plausibility of erroneously proposing that variants detected at a higher frequency in ESRD are risk factors may be reduced by the case–case approach discussed above as well as by putting the genetic variant in its biological context and making comparisons with its role in other, less lethal disorders. Studies on the functional implications of genetic variants during different stages of CKD are therefore warranted.

**Rare variants**

Based on the assumptions of the *common disease–common variant hypothesis*, most of the SNPs probed on GWAS chips are common variants (>1% frequency), derived from the HapMap database [43]. However, the impact on disease by rare variants has been highlighted and several recent studies on complex disorders have re-sequenced candidate genes, previously identified through GWASs, in order to identify rare genetic variants with potentially larger effect sizes [44, 45]. In the area of kidney disease, the exons and 4 kb upstream region of *UMOD* were recently sequenced, which uncovered multiple novel rare variants in the *UMOD* region [46]. Disappointingly though, they were not significantly enriched and could not account for the observed GWAS signal in the study material. In addition, fine-mapping efforts in T2DM and coronary artery disease [47] and T1DM [48] have not yet generated any novel insights and recent data from common autoimmune diseases (autoimmune thyroid disease, coeliac disease, Crohn’s disease, psoriasis, multiple sclerosis and T1DM) suggest only a minor impact of rare coding-region variants on disease susceptibility [49].
Other genetic variants

Structural variations and copy number variations (CNVs) have not been studied as extensively as SNPs. CNVs associate with various common diseases, including glomerulonephritis [50] and are over-represented in genes associated with complex, rather than Mendelian, diseases [51]. This can now be studied using new refined chip designs that are able to retrieve information on some rare variants and CNVs, and new exon and whole-genome sequencing tools [52, 53]. However, some genetic variants such as variation in repeat dense areas of the genome, illustrated by variations in telomere repeat length as well as expanding DNA repeats of which there are several examples in the literature, may remain undetected despite being critical for disease or disease progression [54, 55]. Such repeats may not be detectable by currently available GWAS, CNV or sequencing platforms.

Polygenic and pleiotropic effects

The polygenic architecture of complex diseases infers that multiple variants, each with a small effect size, contribute to disease susceptibility. Since the magnitude of each variant may be too small to be captured by currently available sample sizes [56], and because recruitment and genotyping of sufficiently large samples are limiting factors, researchers are striving to develop new, improved, analytical methods for the GWAS. For example, hypothesis-driven GWAS (HD-GWAS) approaches may incorporate knowledge on disease aetiology and pathobiology [57] or information on enrichment patterns of functionally annotated SNPs [58] a priori into statistical methods to systematically prioritize SNPs, to increase the power of detecting effects that are below the traditional genome-wide significance level in GWAS. Indeed, by using a multistep approach, involving systematic application of prior biological knowledge on 24 genes previously identified and replicated in eGFR GWASs, six novel eGFR-associated loci (in or near the FBXL20, INHBC, LR2P, PLEKHAI, SLC3A2 and SLC7A6 genes) were uncovered by Chasman and coworkers [59], highlighting a feasible strategy to infer novel SNPs and disease pathways without the need for using increasingly larger sample sizes. Moreover, tools to handle genetic pleiotropy, i.e. the phenomenon of a single gene or variant being associated with more than one distinct phenotype, are under way, using pleiotropy-informed approaches for GWAS methods to capture more of the phenotypic complexity in common diseases [60].

Epistasis

Another issue to be solved is the possible influence of gene-gene interaction or epistasis. Freedman et al. [18] hypothesized that extremely common variants with large odds ratios, such as the MYH9 E1 haplotype and APOL1 G1 and G2 alleles, may mask the effects from other loci. Indeed, by using a case–case study strategy, they could demonstrate that FRMD3 variants contribute to the risk of DN in African Americans with T2DM. However, this association only became evident when stratifying for T2DM-associated DN and adjusting for MYH9/ APOL1 gene variants. Thus, this implies a potential interaction between FRMD3 and the MYH9 E1 risk haplotype in DN susceptibility, and adds to the genetic risk profile in nephropathy, primarily constituted by APOL1 G1 and G2 variants. So far, epistatic genes have not been sufficiently investigated but the development and introduction of statistical methods such as multilocus models, interaction analyses and stratification analyses are vital to account for their influence in future studies. Again, the analysis of multiple genes in one data set will increase the demand on sample size.

The emerging role of epigenetics in CKD

The epigenetic state, i.e. potentially reversible changes in genetic information other than the primary DNA sequence, determines whether a particular gene, or other genetic elements, will be functioning or not in terms of transcriptional activity or repression. Epigenetic states modifying the genome and direct gene programmes, include DNA-methylation of CpG dinucleotides, nucleosomal histone modifications and other mechanisms [61]. With the rapidly increasing insights into basic mechanisms of epigenetics, it is now becoming possible to better understand the mechanisms whereby environment and life style influence disease development and how the various cell-type specific functional genomes are differently geared to respond to various cues from a dynamic environment. A loss of this plasticity is likely one of the factors behind many diseases. An important aspect not easily captured in GWAS is the concept introduced by Feinberg et al. [62], suggesting that a genetically determined degree of variability in the epigenetic state may be important for disease development. Thus, the ability or propensity to have an epigenetic alteration as a result of environmental challenges may be genetically determined. Furthermore, it is possible that many types of epigenetic modifications are still unaccounted for due to the lack of analytical tools. There are in fact many chemical modifications of histones and DNA that occur but are not routinely measured yet.

DNA methylation in CKD

Epigenetic disturbances are important in a number of human diseases [63, 64]. In CKD, DNA methylation is of particular interest as uraemic derangements, including hyperhomocysteinaemia [65, 66], inflammation [67], dyslipidemia [68, 69] and oxidative stress [70], have been shown to associate with altered DNA methylation homeostasis. Initial analyses on global DNA methylation changes in CKD patients have demonstrated both DNA hypomethylation [65, 66] and hypermethylation [67] and later genome-wide analyses have identified several different loci associated with DNA methylation changes [71, 72]. However, this research area is still in the early discovery phase and further site-specific methylation analyses are needed to gain more detailed knowledge of epigenetic regulators in CKD. In the future, longitudinal studies to monitor epigenetic fluctuations in disease progression may help to define early markers for CKD progression and complications. In addition, hydroxymethyl cytosine in DNA which is more common in our genomes than previously appreciated may have biological significance [73, 74] and may link the epigenome to oxidative stress and active gene transcription, possibly through a demethylation process involving inflammation.
FUTURE PERSPECTIVES AND CONCLUSIONS

While still a young and somewhat rambunctious field, modern biology is evolving rapidly with a promise of major transformations in the way we see complex disorders, such as CKD. The lack of genes with large effects (such as in polycystic kidney disease) should come as no surprise, given the development in genetics of other complex disorders. The recent large-scale genetic studies have unravelled a plethora of novel genes, exposed previously unrecognized relevant pathways in renal function biology, improved our understanding of renal disease pathogenesis and disease aetiology of different renal phenotypes, and the role of ethnic variability for genetic susceptibility. New insights into epigenetics/epigenomics such as genome-wide histone modifications, DNA cytosine methylation and cytosine hydroxymethylation, combined with information of lifestyle and other environmental factors influencing prognosis, may improve our understanding of the links between genotype and epigenotype with risk for CKD and its hazardous complications. Refined genotyping efforts and integrated clinical epidemiological-genetic approaches may reduce translational gaps between genetic/epigenetic information, biological mechanisms and renal pathophysiology, allowing us to make use of the vast novel genetic and epigenetic information in clinical practice.

FUNDING

This work was supported by The Swedish Research Council (P.S. 2010–2847 and L.N. 2012–2721), ALF grants from Stockholm County Council (M.S. and P.S.), Karolinska University Hospital including a genomics Core grant (M.S.), Karolinska Institutet Research Foundations, Fredrik and Ingrid Thuring Foundation and The Strategic Research Programme in Diabetes at Karolinska Institutet (Swedish Research Council 2009–1068).

CONFLICT OF INTEREST STATEMENT

A.W., T.J.E., P.S., M.S. and L.N. have nothing to disclose. B. L. is employed by Baxter Healthcare Corporation. Baxter Novum is the result of a grant from Baxter Healthcare Corporation to Karolinska Institutet.

(See related article by Witasp et al. How can genetics and epigenetics help the nephrologist improve the diagnosis and treatment of chronic kidney disease patients? Nephrol Dial Transplant 2014; 29: 972–980.)

REFERENCES

34. Rosset S, Tzur S, Shemer R et al. Missense mutations in the APO1 gene are highly associated with end stage kidney disease risk previously attributed to the MTH9 gene. Hum Genet 2010; 128: 345–350
36. Strawbridge RJ, Dupuis J, Prokopenko I et al. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes 2011; 60: 2624–2634
37. Perry JR, Voight BF, Yengo L et al. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variant in lean compared to obese cases. PLoS Genet 2012; 8: e1002741
53. 1000 Genomes A Deep Catalog of Human Genetic Variation. http://www.1000genomes.org/
57. Sun L, Rommens JM, Corvol H et al. Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat Genet 2012; 44: 562–569
62. Feinberg AP. Genome-scale approaches to the epigenetics of common human disease. Virochows Arch 2010; 456: 13–21
66. Feinberg AP. Genome-scale approaches to the epigenetics of common human disease. Virochows Arch 2010; 456: 13–21
74. Valinluck V, Tsai HH, Rogstad DK et al. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res 2004; 32: 4100–4108

Received for publication: 9.7.2013; Accepted in revised form: 5.9.2013