

Genome-Wide Analysis Studies and Chronic Kidney Disease

Doloretta Piras^a Magdalena Zoledziewska^b Francesco Cucca^{b, c}
Antonello Pani^a

^aDivisione di Nefrologia e Dialisi, Azienda Ospedaliera G. Brotzu, and ^bIstituto di Ricerca Genetica e Biomedica (IRGB-CNR), Cagliari, and ^cUniversità degli Studi di Sassari, Sassari, Italy

Keywords

Chronic kidney disease · Rare renal diseases · Genome-wide analysis studies · Genetic markers · Kidney traits

Abstract

In recent years, the very high worldwide prevalence of chronic kidney disease (CKD) has led some authors to talk of an “epidemic.” The progression of CKD varies considerably among individuals despite similar aetiologies, optimal blood pressure, and glycaemic control. Over the last decade, through genome-wide association studies (GWAS), more than 50 genetic loci have been identified in association with CKD. Understanding the genetic basis of CKD could provide a better knowledge of the biology of the involved pathways, thus potentially leading to novel tools for the diagnosis, prevention, and therapy of CKD. In this review, we will analyse the role of GWAS in the study of CKD. © 2017 S. Karger AG, Basel

This work was presented at the international conference “The Kidney in Genetic and Rare Diseases,” Naples, Italy, October 27–29, 2016.

Introduction

In recent years, the very high worldwide prevalence of chronic kidney disease (CKD) has led some authors to talk of an “epidemic” [1]. By 2030, in the United States, end-stage renal disease (ESRD) is estimated to involve over 2 million people [2]. The progression of CKD varies considerably among individuals despite similar aetiologies, optimal blood pressure, and glycaemic control. Clinical factors account for less than one half of the observed variability. The genetic basis for progression of renal disease is, in part, distinct with regard to onset and progression.

The rationale for the study of the genetic background in CKD using genome-wide association studies (GWAS) lies in the high prevalence of CKD and in the variability in the GFR slope among healthy individuals. Moreover, several studies have shown high heritability for GFR (from 36 to 75%), and, albeit to a lesser extent, for albuminuria (from 16 to 49%) [3, 4]. One of the main aims of GWAS in nephrology is to find possible common genetic variants which account for the variability of the kidney trait.

Methodology in Genetic Studies

Several methodologies can be used in genetic studies on kidney traits. Linkage analysis is hypothesis-free since we do not need to know which genes may be involved in specific diseases; however, it is used to study monogenic traits, inherited according to Mendel's law, while it is not useful in the study of polygenic diseases (i.e., CKD) or qualitative traits (i.e., eGFR). Other possible approaches include "candidate gene association studies," to look for the association between common genetic variants (minor allele frequencies >5% in the general population) in plausible candidate genes and phenotypes. However, since it is not hypothesis-free, it is not useful when we are searching for new possible variants related to common diseases. Currently, the most often used approach for this purpose is GWAS, a common method used in associating single nucleotide polymorphisms (SNPs) with the disease or trait being studied. It has several advantages: it is hypothesis-free, so we do not need to know where to look for a possible variant, but we can study the whole genome; it may be used as a complementary approach to linkage studies or when robust linkage analysis is not feasible; with just one analysis, we can study more than 1 million SNPs, and this number can be increased through methods such as imputation. Statistical tests, like linear or logistic regression, are usually used to find an association between SNPs and a common disease or a quantitative trait. However, the limitation is that very large samples are needed since the thresholds of significance in GWAS analysis are far lower than those used in clinical studies (i.e., p values $<5 \times 10^{-8}$) [5]. This problem is often solved by resorting to meta-analysis in which several cohorts are lumped together to be studied. GWAS can be used both in cross-sectional and in longitudinal studies, while they are less frequently used in case-control studies.

In cross-sectional studies, the end point can be continuous (eGFR/albuminuria), thus having the highest power; when it is dichotomic (i.e., eGFR $<$ vs. >60 mL/min) the power is lower, but sometimes the interpretation is easier for clinicians. In many of these studies, eGFR is evaluated both by sCr and Cys; the advantage of using 2 different biomarkers to estimate GFR is that genetic factors affecting their production, metabolism, and secretion can be evaluated. Neither CKD onset nor CKD progression can be studied with this design.

Longitudinal studies are useful for analysing CKD onset, kidney function decline, and progression of CKD, while studies on ESRD are limited by low power since

only a small number of individuals progress to ESRD (i.e., in the ARIC study, only 101 [0.9%] of 11,677 initially healthy individuals of European descent progressed to ESRD over 17 years of follow-up). In the setting of ESRD, case-control studies can be useful.

GWAS in Nephrology

The finding of common genetic variants thanks to GWAS may lead to a better understanding of the variability of GFR and albuminuria in the general population. It could increase our knowledge of the biology of the involved pathways, thus potentially leading to novel tools for the diagnosis, prevention, and therapy of CKD.

According to McCarthy et al. [6], identifying new susceptibility variants which result in novel biological insights can lead to clinical advances such as the discovery of new therapeutic targets, biomarkers, and tools for prevention. Moreover, the understanding of aetiological processes could lead to tailored medicine in diagnosis, prognosis, and treatment [7].

Defining the phenotype is fundamental when performing genetic analysis. In nephrological studies, several phenotypes can be used, and they can be dichotomic (i.e., presence or absence of CKD or of albuminuria), qualitative (i.e., pathological data), or continuous (eGFR, eGFR slope, renal volumes). However, sometimes it is difficult to collect some kinds of data for a large number of individuals. For example, in order to calculate eGFR slope, we need a longitudinally followed-up cohort; pathological data are not easy to use since some features are very specific for some diseases (i.e., membranous nephropathy [MN]), but not for CKD. To date, renal volumes have not been used as phenotypes for genetic studies.

GWAS Analysis in CKD

The era of GWAS analysis in CKD started in 2009 with a study by Köttgen et al. [8]. It was a meta-analysis and involved more than 20,000 individuals, 2,400 of whom had CKD. The phenotypes that were used as dependent variables were CKD (defined as eGFR <60 mL/min/1.73 m²), eGFR based on serum creatinine, and eGFR based on cystatin C. The genes that were eventually identified as being related to renal phenotypes were UMOD, SHROOM3, and STC1.

Subsequently, Chambers et al. [9] independently demonstrated the association between serum creatinine levels and the loci reported by Köttgen et al. [8]. Moreover, other loci were described; the genes that were identified close to these loci encoded for proteins with different functions. For example, SLC7A9 and SLC34A1 encoded for solute transporters expressed in renal proximal tubular cells; NAT8 for N-acetyltransferase; ALMS1, the causal gene for Alstrom syndrome, a disease characterised by progressive liver and kidney failure; VEGFA for vascular endothelial growth factor A, which is produced by podocytes and is needed for the barrier function of the glomerulus. Stratifying the cohort for hypertension or diabetes did not affect the findings, suggesting that these associations were independent of the most common underlying aetiologies of CKD.

The genetic markers described in these 2 studies were not associated with many of the more common causes of CKD (i.e., diabetes and hypertension) and they were not involved in the RAS pathway. These genes were highly expressed in the tubular compartment: the stress caused by hypertension, diabetes, and possibly xenobiotics may have an effect on a common pathway centred in the renal epithelium [10].

In 2010, another GWAS meta-analysis was carried out by Köttgen et al. [11]. The role of the first 3 initial loci was confirmed, and 13 more new loci were identified as being associated with renal function. The important finding was that 16 loci accounted for only 1.4% of eGFR variance in the face of an estimated heritability of 36–75%.

“Missing heritability” has been described, and it is the discrepancy between the high levels of observed heritability of common diseases and traits and the small effect size attributable to the identified variants [12].

In 2012, Pattaro et al. [13] described 6 new loci associated with CKD and confirmed the role of the 23 known ones. This GWAS is interesting also because the results were stratified for age. For example, UMOD showed a stronger association in older individuals. The same finding was observed in a founder population in Iceland: the effect of UMOD on serum creatinine increased with age. According to the authors, the UMOD variant may influence the adaptation of the kidney to age-related risk factors of kidney disease.

Pattaro et al. [14] demonstrated that many SNPs are in regulatory regions (in particular in enhancer regions) and that many SNPs are expressed in renal epithelial cells but not in endothelial cells, and furthermore that some genes play a role in embryonic development.

The CRIC study group added new SNPs to this scenario, but only 4 reached the threshold for genome-wide significance and only in black individuals, while none of these SNPs were significant in Caucasians [15].

Data from GWAS to Epidemiological Studies

Using longitudinal cohorts, Böger et al. [16] analysed the association between the SNPs described in the literature and found them to be associated with renal function, ESRD, and CKD incidence: 11 of 16 SNPs were associated with kidney traits. UMOD was found to be associated both with incident CKD (adjusted for basal eGFR) and with ESRD.

Moreover, an association between 5 loci and CKD (DAB2, PRKAG2, ANXA9, DACH1, STC1) was described, and one locus was associated with ESRD (GCKR). However, the other 4 loci were not associated with either CKD or ESRD after correction for basal eGFR.

In 2014, we tested the SNPs which were described in the literature as being associated with renal function on the SardiNIA cohort (i.e., a founder population in Ogliastra, an isolated region in Sardinia, Italy), taking into consideration the genetic data together with the clinical parameters [17]. The end points we considered were the presence of CKD, that we defined according to the KDIGO guidelines as either eGFR <60 mL/min or the presence of albuminuria, the eGFR slope (continuous data), and fast decline versus slow eGFR decline (dichotomous data). In order to use the genetic data, we created a genetic risk score. We used 13 SNPs previously described in the literature as being associated with renal function, and we analysed the clinical parameters together with the genetic risk score in a multivariate analysis. Since every individual has 2 alleles for each locus, there were 26 alleles altogether; in the best case scenario, they could all have been low-risk alleles, so the genetic risk score would have been 0, while in the worst case scenario (all high-risk alleles) the score would have been 26. Remarkably, in the final model, the genetic risk score was associated with all the end points: (1) odds of CKD; (2) additional change in eGFR; (3) odds of fast eGFR decline. In particular, each allele was found to carry an odds ratio of 1.07 for CKD. Therefore, considering that each risk allele adds a 7% risk of CKD, and that in our population the genetic risk score ranged from a minimum of 6 to a maximum of 24, the person with the highest risk score had a three-fold increased risk of CKD compared to the person with the lowest score.

How to Use Data from GWAS

About 90% of the SNPs associated with complex traits are localised in non-coding regions (enhancer more often than promoter) [18]. Once a new SNP has been located, we have to search the nearby regions to find candidate genes [10]. The search is based on the strength of the association with CKD, i.e. higher renal expression as compared to other organs, and on the strength of the biological rationale for the hypothesised role in kidney disease.

Comments on GWAS

Since most of the SNPs that were found to be associated with CKD are located in non-coding regions, they do not change the protein sequence, but they may influence the protein expression, thus possibly complicating the statistical association. A large “missing heritability” is possible since 16 loci account for only 1.4% of the variability, while the heritability of eGFR is estimated to be about 36–75%. Most studies used several cohorts; therefore, uniformity of the definition of phenotype was not guaranteed, for example with regard to differences in the methods of serum creatinine calibration and the different performance of formulas for estimating GFR in different populations. Moreover, individuals with severe kidney function impairment were under-represented [19].

Main Genes Identified by GWAS and Their Roles

The UMOD gene encodes for uromodulin or Tamm-Horsfall protein. It is a kidney-specific protein that is exclusively synthesised by epithelial cells lining the thick ascending limb of the loop of Henle. It is the most abundant urinary protein under physiological conditions, and it defends against urinary tract infections caused by uropathogenic bacteria, and furthermore, it is protective against kidney stones. Defects in UMOD are associated with several renal disorders such as medullary cystic kidney disease-2, glomerulocystic kidney disease with hyperuricaemia and isostenuria, and familial juvenile hyperuricaemic nephropathy. Moreover, common variants in the UMOD gene are associated with hypertension, eGFR, ESRD, CKD, and kidney stones [20, 21].

SHROOM3 is an actin-associated protein that regulates epithelial cell shape and tissue morphogenesis by binding F actin and regulating its subcellular organiza-

tion. It is needed for the development and maintenance of podocyte cytoarchitecture. In the absence of SHROOM3, podocyte morphology is altered during development leading to podocyte loss and glomerular degeneration [22].

GWAS and Rare Renal Diseases: The Example of Glomerulonephritis

Membranous Nephropathy

GWAS can also be used for the study of rare diseases, such as primary MN. A European group found that genetic variants in an HLA-DQA1 and phospholipase A2 receptor allele were associated most significantly with MN [23].

We studied 94 Sardinian cases of MN versus 1,668 controls, and we analysed 8 million SNPs (SNPs with a major allelic frequency >1%). Six of these variants were associated with MN. We found a trend toward an association for the known PLA2R1 locus (p value = 5.7×10^{-3}), and a strong association with the HLA region (p value = 1.1×10^{-12}).

IgA Nephropathy

A complex genetic background underlies the pathogenesis of IgA nephropathy. An association with the major histocompatibility complex has been found [24, 25, 26], as with genes involved in inflammation [25] and with the complement factor H-related genes CFHR1 and CFHR3 [24]. Moreover, an association between risk alleles and age at onset was described, as was an increased risk of inflammatory bowel disease and an overlap with loci associated with the maintenance of the intestinal epithelial barrier and response to mucosal pathogens [27].

Conclusion

GWAS is a useful tool for studying common genetic variants associated with both CKD and eGFR, but also some specific diseases such as MN and IgA nephropathy. GWAS can identify common risk variants in <100 individuals if they have a large effect (i.e., MN). However, for common, complex diseases such as CKD, more than 20,000 individuals are needed since most of the common risk variants have a small effect.

GWAS is like a roadmap: it can drive scientists towards the knowledge of new pathophysiological path-

ways. Finding a new SNP associated with a clinical trait results in greater knowledge of the loci near that specific SNP, and may lead to the discovery of a new gene and the function of its encoded protein.

Sometimes, there is an overlap between loci associated with CKD traits and monogenic kidney disease genes (such as for UMOD). An explanation may be that different variants of the same gene can lead to different phenotypes with differing disease severity. The common vari-

ants usually have low penetrance, whilst, in Mendelian diseases, very rare variants have high penetrance.

Translational science is essential for a better understanding of the results of GWAS.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

- 1 Zoccali C, Kramer A, Jager KJ: Epidemiology of CKD in Europe: an uncertain scenario. *Nephrol Dial Transplant* 2010;25:1731–1733.
- 2 Wali RK, Henrich WL: Chronic kidney disease: a risk factor for cardiovascular disease. *Cardiol Clin* 2005;23:343–362.
- 3 O’Seaghdha CM, Fox CS: Genetics of chronic kidney disease. *Nephron Clin Pract* 2011; 118:c55–c63.
- 4 Placha G, Canani LH, Warram JH, Krolewski AS: Evidence for different susceptibility genes for proteinuria and ESRD in type 2 diabetes. *Adv Chronic Kidney Dis* 2005;12:155–169.
- 5 Pe’er I, Yelensky R, Altshuler D, Daly MJ: Estimation of the multiple testing burden for genome-wide association studies of nearly all common variants. *Genet Epidemiol* 2008;32: 381–385.
- 6 McCarthy MI, Hirschhorn JN: Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008;17: R156–R165.
- 7 O’Seaghdha C, Fox C: Genome-wide association studies of chronic kidney disease: what have we learned? *Nat Rev Nephrol* 2011;8: 89–99.
- 8 Köttgen A, Glazer N, Dehghan A, et al: Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* 2009;41:712–717.
- 9 Chambers J, Zhang W, Lord G, et al: Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* 2010;42:373–375.
- 10 Price PM, Hirschhorn K, Safirstein RL: Chronic kidney disease and GWAS: “the proper study of mankind is man.” *Cell Metab* 2010;11:451–452.
- 11 Köttgen A, Pattaro C, Böger C, et al: New loci associated with kidney function and chronic kidney disease. *Nat Genet* 2010;42:376–384.
- 12 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, et al: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9: 356–369.
- 13 Pattaro C, Köttgen A, Teumer A, et al: Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* 2012;8:e1002584.
- 14 Pattaro C, Teumer A, Gorski M, et al: Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* 2016;7:10023.
- 15 Parsa A, Kanetsky PA, Xiao R, et al: Genome-wide association of CKD progression: the chronic renal insufficiency cohort study. *J Am Soc Nephrol* 2017;28:923–934.
- 16 Böger CA, Gorski M, Li M, Hoffmann MM, et al: Association of eGFR-related loci identified by GWAS with incident CKD and ESRD. *PLoS Genet* 2011;7:e1002292.
- 17 Pani A, Bragg-Gresham J, Masala M: Prevalence of CKD and its relationship to eGFR-related genetic loci and clinical risk factors in the SardiNIA study cohort. *J Am Soc Nephrol* 2014;25:1533–1544.
- 18 Susztak K: Understanding the epigenetic syntax for the genetic alphabet in the kidney. *J Am Soc Nephrol* 2014;25:10–17.
- 19 Drawz P, Sedor J: The genetics of common kidney disease: a pathway toward clinical relevance. *Nat Rev Nephrol* 2011;7:458–468.
- 20 Scolari F, Izzi C, Ghiggeri G: Uromodulin: from monogenic to multifactorial diseases. *Nephrol Dial Transplant* 2015;30:1250–1256.
- 21 Gudbjartsson D, Holm H, Indridason O, et al: Association of variants at UMOD with chronic kidney disease and kidney stones – role of age and comorbid diseases. *PLoS Genet* 2010; 6:e1001039.
- 22 Khalili H, Sull A, Sarin S, et al: Developmental origins for kidney disease due to Shroom3 deficiency. *J Am Soc Nephrol* 2016;27:2965–2973.
- 23 Stanescu HC, Arcos-Burgos M, Medlar A, Bockenbauer D, et al: Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *N Engl J Med* 2011;364:616–626.
- 24 Gharavi AG, Kiryluk K, Choi M, Li Y, et al: Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* 2011;43:321–327.
- 25 Yu XQ, Li M, Zhang H, Low HQ, et al: A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet* 2011;44:178–182.
- 26 Feehally J, Farrall M, Boland A, Gale DP, et al: HLA has strongest association with IgA nephropathy in genome-wide analysis. *J Am Soc Nephrol* 2010;21:1791–1797.
- 27 Kiryluk K, Li Y, Scolari F, Sanna-Cherchi S, et al: Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* 2014; 46:1187–1196.