

Editorial Reviews

Genome-wide association studies in kidney diseases: Quo Vadis?

Francesco Pesce and Francesco Paolo Schena

Department of Emergency and Organ Transplantation, Renal, Dialysis and Transplant Unit, University of Bari, Bari, Italy

Correspondence and offprint requests to: Francesco Paolo Schena; E-mail: fp.schena@nephro.uniba.it

Keywords: genetic; glomerulonephritis; kidney diseases; microarray

Introduction

A genome-wide association (GWA) study is a genetic epidemiology approach designed to scan genetic variation across the entire human genome in order to identify genetic associations with phenotypic traits as well as the presence or absence of a disease. Hundreds of thousands of single-nucleotide polymorphisms (SNPs), the most common form of genetic variant, serve as markers. SNPs are assayed and related to diseases or health-related conditions applying bioinformatics algorithms. This has become feasible thanks to the recent technological improvements in the so-called high-throughput technologies. The analysis identifies regions (loci) with statistically significant differences in allele or genotype frequencies between cases and controls and so the variations are said to be 'associated' with the disease. The completion of the Human Genome Project in 2003 and then the HapMap Project allowed GWA studies to be conducted, cataloguing common human genetic variants and providing the 'tag SNPs' (markers) included in the DNA microarrays used for the genotyping [1]. GWA studies are usually structured in four parts (1) collection of a large number of blood sample from individuals with the disease/trait of interest and from a control group; (2) DNA isolation and genotyping; (3) statistical tests for associations between the SNPs and the disease/trait; (4) replication of identified associations in an independent population sample and examination of functional implications experimentally [2]. Four hundred replicated associations have now been reported for more than 70 common diseases, conditions and biological parameters (see the National Human Genome Research Institute catalogue of published GWA studies, <http://www.genome.gov/gwastudies>) [3]. In this review, we present the results of GWA studies performed for kidney diseases published to date (Table 1) and point out the main features of this intricate methodological approach in order to better understand what opportunities are offered in nephrology. Studies have been identified through 'HuGE Navigator', 'A Catalog of Published Genome-Wide Association Studies' and 'PubMed' search [3,4].

GWA studies in kidney diseases

Chronic kidney disease and glomerular filtration rate

The group of Caroline S. Fox recently indicated susceptibility loci for estimated glomerular filtration rate (eGFR) and chronic kidney disease (CKD) [5]. They conducted meta-analyses of specific GWA studies for indices of renal function and for CKD from four population-based, unselected cohorts of European-ancestry part of the CHARGE Consortium: Atherosclerosis Risk in Communities Study (ARIC), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS) and Rotterdam Study (RS). A total of 19 877 participants with 2388 CKD cases were analysed. They tested for replication in 21 466 participants with 1932 CKD cases. The eGFR was measured by serum creatinine (eGFR_{crea}) and cystatin C (eGFR_{cys}), and CKD was defined as eGFR_{crea} <60 ml/min/1.73 m². The replication samples did not have cystatin C measurement available.

Choosing a broad phenotype definition for CKD without narrowing for the main underlying diseases, such as hypertension or diabetes, is a natural application of GWA studies and may allow the identification of common mechanisms. For CKD, they identified SNPs in the Uromodulin gene (UMOD; OMIM 191845) on chromosome 16p12.3. The same SNP rs12917707 found at the UMOD locus had the strongest association with eGFR_{crea} too. Additional SNPs associated with eGFR_{crea} were the intronic SNP rs17319721 in the shroom family member 3 (SHROOM3; OMIM 604570) on chromosome 4q21.1, the intronic SNP rs6040055 in the Jagged 1 (JAG1; OMIM 601920) on chromosome 20p12 and the intronic SNP rs2467853 in spermatogenesis associated 5-like 1 (SPATA5L1) at the GATM-SPATA5L1 locus on chromosome 15q21.1.

Three loci were found in association with eGFR_{cys}. The strongest association was for the intergenic SNP rs13038305 between cystatin C (CST3; OMIM 604312) and cystatin 9 (CST9). The other SNPs were the intergenic SNP rs1731274 in the stannocalcin 1 gene (STC1; OMIM 601185) on chromosome 8p11.2-21 and the rs12917707 at the UMOD locus. UMOD encodes the Tamm-Horsfall protein that is the most abundant protein in the urine of healthy individuals. UMOD knockout mice have been shown to have 63% lower creatinine clearance compared to

Table 1. Published genome-wide association studies in kidney diseases

Author	Disease/trait	Locus	Region	SNP
Köttgen <i>et al.</i> [5]	Chronic kidney disease Estimated glomerular filtration rate (creatinine)	UMOD	16p12.3	rs12917707
		SHROOM3	4q21.1	rs17319721
		SPATA5L1, GATM	15q21.1	rs2467853
		UMOD	16p12.3	rs12917707
		JAG1	20p12	rs6040055
	Estimated glomerular filtration rate (cystatine)	STC1	8p11.2-21	rs1731274
		UMOD	16p12.3	rs12917707
		CST3-CST9	20p11.2	rs13038305
		SLC12A3	16q13	rs2289116
		ELMO1	7p14	rs741301
Maeda <i>et al.</i> [6]	Diabetic nephropathy (type 2 diabetes)	FRMD3	9q21-22	rs1888746
				rs13289150
		CARS	11p15.5	rs739401
				rs451041
Pezzolesi <i>et al.</i> [10]	Diabetic nephropathy (type 1 diabetes)	PVT1	8q24	rs2720709
				rs2648875
Hanson <i>et al.</i> [11]	End-stage renal disease in diabetic nephropathy	PVT1	8q24	rs2720709
Obara <i>et al.</i> [12]	IgA nephropathy	PIGR	1q31-41	Not reported

wild-type mice. SHROOM3 encodes a protein expressed in human kidney and involved in epithelial cell shape regulation. JAG1, a ligand of the Notch receptor involved in cell differentiation and morphogenesis, could play a role in kidney diseases but the association found in the discovery group did not replicate. GATM encodes glycine amidinotransferase, an enzyme involved in creatine biosynthesis. The genes in the CST superfamily encode cystatin proteins and STC1 encodes stanniocalcin 1, a hormone regulating calcium homeostasis in fish that is highly expressed in the renal nephron and may influence local calcium and phosphate homeostasis through a paracrine mechanism. In this study, the risk of CKD was 0% in those without any risk alleles and 12.1% in individuals carrying all six risk alleles [5].

Diabetic nephropathy

A GWA study for diabetic nephropathy (DN) in type II diabetes in a Japanese population confirmed the role of transcription factor 7-like 2 (TCF7L2; OMIM 602228) in type II diabetes as previously reported in Icelandic, Danish and American populations [6]. The TCF7L2 gene maps on chromosome 10q25; it is a part of the Wnt signalling pathway and could have pleiotropic effects. Its product is a transcription factor implicated in blood glucose homeostasis. Variants in this gene may play a role in response to certain classes of hypoglycaemic agents and have also been related to colon cancer. What is really interesting is that TCF7L2 might be associated with renal function and CKD development, independent of its effect to increase the risk for overt diabetes as shown by Köttgen *et al.* [7,8]. Maeda *et al.* [6] identified solute carrier family 12 (sodium/chloride transporters) member 3 (SLC12A3; OMIM 600968) and engulfment and cell motility 1 gene (ELMO1; OMIM 606420) as candidates for DN. SLC12A3 encodes a renal thiazide-sensitive sodium-chloride cotransporter that mediates sodium and chloride reabsorption in the distal convoluted tubule. Mutations in this gene cause Gitelman syndrome. While many large-scale genotyping studies performed on Japanese subjects with type 2

diabetes have implicated polymorphisms in SLC12A3, genetic variation at this locus is unlikely to explain the risk for advanced DN among type 2 diabetic Caucasians [9]. The authors suggest that ELMO1 expression, under high glucose conditions, might contribute to the development and progression of DN influencing the expression of fibronectin. Pezolesi *et al.* [10] performed a GWA study in the genetics of kidneys in diabetes (GoKinD) sample collection, to identify loci associated with the risk of DN in type 1 diabetes. In 820 cases with DN and 885 controls with type 1 diabetes, 360 000 SNPs were genotyped. Among the cases, 284 had proteinuria and 536 end-stage renal disease (ESRD). Samples of the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) study, designed to investigate the development of diabetes-associated complications, were used as replication sample. Thirteen SNPs in four genomic loci were found to be associated with DN. The '4.1 protein ezrin, radixin, moesin (FERM) domain' containing 3 (FERMD3; OMIM 607619) was the strongest associated locus together with cysteinyl-tRNA synthetase (CARS; OMIM 123859) locus. β -chimerin isoform 2/serine carboxypeptidase vitellogenic-like (CHN2/CPVL; OMIM 602857/609780) and an intergenic region on chromosome 13q were also found associated but not in replication sample. FERMD3 maps on chromosome 9q21-22 and is related to a multifunctional protein essential for maintaining erythrocyte shape and membrane mechanical properties in several cells including mouse nephron. It is detectable in adult ovaries, fetal skeletal muscle, brain and thymus. CARS maps on chromosome 11p15.5 and is involved in the attachment of each of the 20 naturally occurring amino acids to their cognate tRNA isoaccepting families. It is expressed in mesangial and proximal tubule cells and has a role in the pathogenesis of cystinosis. CPVL is carboxypeptidase like angiotensin-converting enzyme and bradykinin. It is highly expressed in proximal tubules.

In order to assess the genetic variants contributing to ESRD in type 2 diabetes, Hanson *et al.* [11] performed a GWA study of 115 352 SNPs in pools of 105 unrelated case subjects with ESRD and type 2 diabetes and

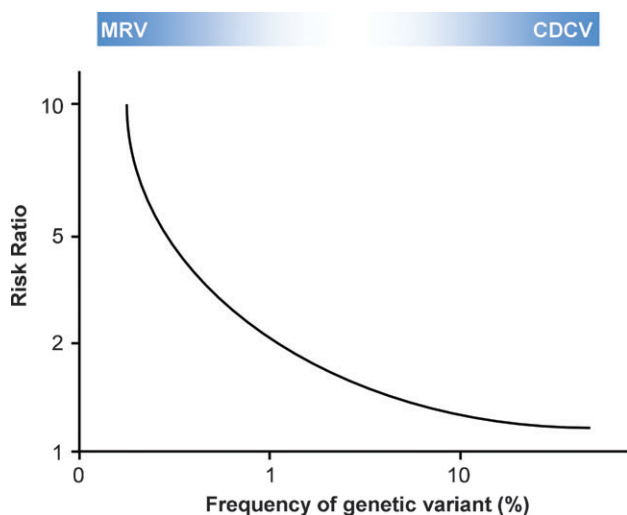


Fig. 1. Relation between the frequency of a genetic variant and risk ratio in respect to Common Disease/Common Variant (CDCV) and Multiple Rare Variants (MRV) hypotheses. GWA studies are based on CDCV hypothesis and lose their detecting power in the case of MRV causing a disease.

102 controls with type 2 diabetes for over 10 years without macroalbuminuria. He found strong associations with rs2720709, located in the plasmacytoma variant translocation gene (PVT1; OMIM 165140) together with rs2648875 that maps to intron 8 of the same gene.

IgA nephropathy

A GWA study was performed by Obara *et al.* in 389 Japanese IgA Nephropathy (IgAN) patients and 465 controls analysing ~80 000 SNPs. He identified a significant association between this glomerulonephritis and six SNPs located in the polymeric immunoglobulin receptor (PIGR; OMIM 173880) gene at chromosome 1q31–41. The PIGR is expressed on several glandular epithelia mediating transcellular transport of polymeric immunoglobulin molecules. Its binding with dimeric IgA on the basolateral surface mediates their transcytosis to the apical surface [12].

Walking through a GWA study

The basic concept of GWA studies is the Common Disease/Common Variant (CDCV) hypothesis, according to which a common complex disease is underlined by relatively few common genetic variants, present in >1% of the population. Association studies lose their detecting power in the case of multiple rare variants (MRV) as the primary cause of a disease. According to MRV hypothesis, complex traits result from many different rare mutations, but with strong effect (high-risk ratios). The relation between the risk ratio and the frequency of a genetic variant in respect to MRV and CDCV hypotheses is shown in Figure 1. These two hypotheses do not exclude each other and with their interconnection generate a continuum [13]. However, the relative contribution of rare versus common variants to disease susceptibility in kidney diseases is not clearly defined. The factors influencing the statistical power of GWA studies are sample size, allele frequency and magnitude of

genetic effect. Since we do not exactly know neither allele frequency nor the magnitude of genetic effect until the genetic variant is well characterized, we can work on the sample size as the major controllable factor. It has been calculated that ~500 cases and 500 controls are required to achieve a power of 80% to detect an allelic odds ratio (as a measure of risk) of 2.0 at a minor allele frequency of 0.1 at an appropriate level of significance. A key point here is that many nephropathy susceptibility genetic variants may confer a lower magnitude of risk and as consequence we would need a larger number of patients. To detect an allelic odds ratio of 1.5, the number of patients and control raises to 2000 versus 2000. Moreover, the number of patients required increases exponentially as the minor allele frequency decreases. So a GWA study is not fitting if rare variants are responsible for the majority of genetic predisposition to nephropathy and this approach should be attempted only if a sufficient power to detect variants that confer at least a moderate risk of nephropathy can be achieved [14].

The number of tagSNPs included in the microarray, as markers for genome coverage, also affects power [15]. Then after genotyping several hundred thousands of tag SNPs throughout the entire genome, the main analysis problem is the multiple hypothesis testing that can lead to Type I error. Thus, the Bonferroni correction is applied to minimize the false positive results, and very stringent *P*-values are needed, usually 1.0×10^{-7} or 1.0×10^{-8} . A stringent significance threshold is applicable when the sample size is large and the power is adequate [15]. A major obstacle to overcome is the difficulty to move beyond statistical association and to describe the functional biological basis of the link between a genomic interval and the phenotypic trait under investigation. It is not easy to translate a statistical significance of a stand-alone GWA study into results that are meaningful clinically. A support may be provided by another high-throughput technology simultaneously assaying global gene expression. The data sets provided by GWA studies, integrated with those from gene expression should better address the finding for causative variants as proposed by Cookson *et al.* [16]. What is interesting about GWA studies is that such studies permit a quite comprehensive scan of the whole genome and are hypothesis-free. There is no bias or presumptive list of candidate genes thus having the potential to identify novel susceptibility genes. In fact, most of the loci identified through GWA studies had not previously been related to the disease under investigation. The hint is that molecular 'subphenotypes' may exist. Although different pathways might potentially be involved in the development of a particular disease when all cases are considered, in every single patient only one, or a subset, of these pathways may be involved. In the same way, the concept of 'diseaseome' underlines how the relation found between one genomic interval and two, or more, apparently different diseases traces an overlapped network interconnecting several pathways and diseases [17]. However, before moving to 'subphenotypes' we should be sure that we selected a precise clinical/histological phenotype because for case-control studies, as GWA studies, rigorous criteria to correctly assign phenotype are required.

While this major strength of GWA studies is suitable for general conditions such as ESRD, the suggestions of

Goldstein may better fit the genetic of kidney glomerular disorders. He says that the apparently modest effects of common variations probably reflect the efficiency of natural selection. He believes that we should shift attention from genome scans of ever larger samples to studies of rarer variants of larger effect. Searching the full human genome for rare variants requires sequencing capacity and thoughtful selection of the most appropriate groups of individual genomes to resequence, and thoughtful evaluation and prioritization of the many rare variants identified [18]. Replication of a GWA study has to face towards the issue of population stratification. Even in a relatively homogenous population, with well-matched cases and controls recruited from the same geographical area, the effect of population stratification remains. This effect, as a major confounding factor, is amplified in the GWA studies where a very large sample size is needed. Investigators from different countries, as partner of a consortium, usually contribute to complete this step of the project. However, the associations found in a population are not always applicable to another one [15].

Conclusions

GWA studies promise to be the milestone amongst the high-throughput technologies that would eventually lead us to identify novel targets for therapeutic intervention in virtually every field of medicine. The final aim seems to be achievable getting back to the root cause of a disease: gene mutations. The pathophysiology of many kidney diseases is not clearly defined and the chance to search for causative factors throughout the genome in an unbiased way is useful to provide new hypotheses about disease mechanisms. However, GWA studies are pointed out as expensive 'factory science' as we need a large number of case and controls and the cost of genotyping for each of them is high. Resequencing studies are needed in the case of risk alleles with low genotypic relative risks or if genetic risk is conferred by multiple rare SNPs too small to be detected by GWA studies. The next step will combine large GWA studies analyses with resequencing studies for rare variants. After that, the integration with data from gene expression and then biological validation will highlight new significant pathways. Moreover, unknown genetic mechanisms have to be taken into account. Recently, Ptak and Petronis [19] highlighted the importance of epigenetic factors (DNA methylation patterns having genomic effects other than sequence), and we are starting to know copy number variants (CNVs), micro RNAs and long-range promoters. The role of CNV has been described in autoimmune diseases, autism, bipolar disorder and have been reported to affect gene expression. Arrays for detecting simultaneously nearly 1 million SNPs and 1 million CNV have been developed [15]. Correlations between genetic variants and phenotypes are limited by the accuracy in measuring each of them. Now, the ability to measure genotype may exceed our ability to measure phenotype. Continuous ambulatory monitoring, imaging methods and comprehensive ('-omic') approaches to biological samples promise to improve the

accuracy of phenotype measurement [20]. While it is not likely to be clinical utility of GWA studies results in the immediate future, they may hold promise as a valid tool for identifying new mechanisms in the pathogenesis of some kidney diseases.

Conflict of interest statement. None declared.

References

- McCarthy MI, Abecasis GR, Cardon LR *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; 9: 356–369
- Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008; 299: 1335–1344
- Hindorf LA, Junkins HA, Mehta JP *et al.* A catalog of published genome-wide association studies. Available at: <http://www.genome.gov/26525384> (1 June 2009, date last accessed)
- Yu W, Gwinn M, Clyne M *et al.* HuGE Literature Finder. HuGE Navigator. Available at: <http://www.hugenavigator.net/HuGENavigator/startPagePubLit.do/>. (1 June 2009, date last accessed)
- 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
- Maeda S, Osawa N, Hayashi T *et al.* Genetic variations associated with diabetic nephropathy and type II diabetes in a Japanese population. *Kidney Int Suppl* 2007; 106: S43–S48
- George AL Jr. Appraising the value of genomic association studies. *J Am Soc Nephrol* 2008; 19: 1840–1842
- Köttgen A, Hwang SJ, Rampersaud E *et al.* TCF7L2 variants associate with CKD progression and renal function in population-based cohorts. *J Am Soc Nephrol* 2008; 19: 1989–1999
- Ng DP, Nurbaya S, Choo S *et al.* Genetic variation at the SLC12A3 locus is unlikely to explain risk for advanced diabetic nephropathy in Caucasians with type 2 diabetes. *Nephrol Dial Transplant* 2008; 23: 2260–2264
- Pezzolesi MG, Poznik GD, Mychaleckyj JC *et al.* Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes mellitus. *Diabetes* 2009; 58: 1403–1410
- Hanson RL, Craig DW, Millis MP *et al.* Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. *Diabetes* 2007; 56: 975–983
- Obara W, Iida A, Suzuki Y *et al.* Association of single-nucleotide polymorphisms in the polymeric immunoglobulin receptor gene with immunoglobulin A nephropathy (IgAN) in Japanese patients. *J Hum Genet* 2003; 48: 293–299
- Iles MM. What can genome-wide association studies tell us about the genetics of common disease? *PLoS Genet* 2008; 4: e33
- Wang WY, Barratt BJ, Clayton DG *et al.* Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; 6: 109–118
- Seng KC, Seng CK. The success of the genome-wide association approach: a brief story of a long struggle. *Eur J Hum Genet* 2008; 16: 554–564
- Cookson W, Liang L, Abecasis G *et al.* Mapping complex disease traits with global gene expression. *Nat Rev Genet* 2009; 10: 184–194
- Frazer KA, Murray SS, Schork NJ *et al.* Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009; 10: 241–251
- Goldstein DB. Common genetic variation and human traits. *N Engl J Med* 2009; 360: 1696–1698
- Ptak C, Petronis A. Epigenetics and complex disease: from etiology to new therapeutics. *Annu Rev Pharmacol Toxicol* 2008; 48: 257–276
- Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; 322: 881–888

Received for publication: 4.8.09; Accepted in revised form: 6.8.09