Genetics of kidney failure and the evolving story of APOL1

David J. Friedman1,2,3 and Martin R. Pollak1,3,4

1Nephrology Division, Department of Medicine, and 2Center for Vascular Biology Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. 3Harvard Medical School, Boston, Massachusetts, USA. 4Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

Chronic kidney disease (CKD) results from a wide array of processes that impair the kidney’s ability to perform its major functions. As many as 20 million Americans suffer from CKD and nearly a half million from end-stage renal disease, but there are also examples of centenarians with adequate renal function. Family-based and genome-wide studies suggest that genetic differences substantially influence an individual’s lifetime risk for kidney disease. One emerging theme is that evolution of genes related to host defense against pathogens may limit kidney longevity. The identification of these genetic factors will be critical for expanding our understanding of renal development and function as well as for the design of novel therapeutics for kidney disease.

Introduction

Natural selection determines which genes are successfully transmitted to future generations, but the dramatic increase in the human lifespan over the past century has presented an unexpected challenge to our organ systems. The nearly half-million Americans with end-stage renal disease (ESRD), and many millions with chronic kidney disease (CKD), indicate that for many individuals, preservation of kidney function limits longevity (1). Increasing numbers of centenarians with adequate renal function suggests that some kidneys can outlast the upper limits of the current human lifespan. We can then ask why some kidneys fail after less than a decade, while others can span a century without signs of deterioration. The answer to this question has profound implications for understanding how the kidney works, finding treatments for renal diseases, and planning for an aging society.

CKD results from a wide array of processes that impair kidney function. While collapsing a multitude of specific diagnoses into the term “CKD” has clinical utility, it obscures the complexity of the diverse biological processes that cause decreased kidney function. We are beginning to recognize that many types of CKD have strong genetic components. In some cases, identical clinical syndromes may be caused by mutations in different genes. In other cases, seemingly distinct disease entities are driven by the same gene mutation. Identifying such genes can illuminate biological processes and overturn dogmas about disease pathogenesis. Connecting genetic variants with clinical entities will be a major goal in the years ahead.

Genetic diseases can be broadly classified as Mendelian or complex. Conceptually, Mendelian and common complex diseases lie at two ends of a spectrum represented by very rare but very highly penetrant variants on one end, and very common but poorly penetrant variants on the other. Evolutionary pressures may lead to variants and associated phenotypes that deviate from this pattern, with unusual combinations of frequency and effect size (2). Here, we briefly consider Mendelian and complex genetic variants leading to kidney failure and highlight one surprising example in this rarer category of common but powerful disease-causing genetic variants.

Mendelian kidney diseases may be surprisingly common

Polycystic kidney disease (PKD) is the best-known and most easily recognizable inherited kidney disease. It is seen frequently, not because of multiple common variants in the genome (each with small effect), but because the variants that cause PKD in the PKD1 and PKD2 genes are common for such highly penetrant mutations (Table 1). Most PKD patients develop kidney disease after the reproductive years, which explains why these variants are not rapidly eliminated from the population. Although PKD is the most common Mendelian genetic disease leading to kidney failure, penetrant PKD mutations are found on roughly 0.1% or fewer of all chromosomes (3).

Other Mendelian renal diseases are not so easily recognized. A common presentation in the nephrologist’s office is the individual with mildly decreased kidney function and mild to moderate proteinuria, with or without hypertension. The nonspecific nature of this phenotype can obscure the genetic heterogeneity of the underlying diseases. For instance, family-based approaches have shown that point mutations in the ACTN4, TRPC6, and INF2 genes can all lead to similar phenotypes characterized by progressive deterioration in renal filtration function and a pattern of injury known as focal segmental glomerulosclerosis (FSGS) (4–6). Mutations in any of these genes cause highly – but incompletely – penetrant autosomal-dominant disease. Although each of these mutations is individually rare, highly penetrant mutations in the known and unknown kidney disease genes taken together likely account for a nontrivial fraction of disease. Individually, these variants remain rare because they become apparent early in life and likely affect the number of offspring a patient will have, decreasing the probability that the variants will be transmitted.

Another clinically important, genetically heterogeneous entity is pediatric nephrotic syndrome, which has long been regarded as an immunologically mediated disease. The identification of defects in the NPHS2 gene and its protein product, podocin, in patients with early-onset nephrotic syndrome demonstrated that abnormalities in the glomerular slit diaphragm can present similarly to immune nephropathies (7). As other slit diaphragm genes have subsequently been identified as the cause of related clinical presentations, we now suspect that some large fraction (20%–30%) of childhood steroid-resistant nephrotic syndromes result from structural and/or

Conflict of interest: The authors are named on a provisional patent for an APOL1 genetic test.

Citation for this article: J Clin Invest. 2011;121(9):3367–3374. doi:10.1172/JCI46263.
signaling failure due to inheritance of two mutant copies of glomerular slit-diaphragm genes (8). Consequently, we have become less likely to treat these genetic diseases in children with immunomodulatory therapy (9).

Sequencing data of the 1000 Genomes Project indicate that each person carries 250–300 loss-of-function variants (10). Gain-of-function variants may also be common. This unexpectedly large number suggests that the fraction of kidney disease explained by individually rare variants may, in aggregate, also be greater than expected.

Identifying genetic factors in common forms of kidney disease

The high prevalence of kidney disease poses the question of whether common phenotypes result from rare genetic variants with large effects or from multiple common (but weak) genetic variants that influence disease in aggregate. Recent genome-wide association studies (GWAs) using estimated glomerular filtration rate (eGFR) as the phenotype of interest have identified several loci that may affect renal function (11, 12). The large number of individuals genotyped has afforded impressive statistical power to these studies, but effect sizes are quite modest, and the fraction of variance explained is very small. However, GWASs have identified some common variants in genes that also cause familial juvenile hyperuricemic nephropathy (13, 14). The observation that the causal SNP is not known with certainty.

Although investigators are making progress toward identifying genetic variants associated with specific types of common kidney disease, few compelling loci have yet been pinpointed. Diabetic nephropathy is among the most common causes of renal disease in essentially every developed country (17). Dozens of SNPs for type 1 diabetes (18) and for type 2 diabetes (19) have been identified, but no rare familial forms of IgA nephropathy and the typical sporadic presentation is not known. The large ethnic and geographical differences in IgA nephropathy, such as its high prevalence in Asia, suggest that environmental factors may place different evolutionary pressures on the frequencies of IgAN-influencing alleles (31).

A recent GWAS for idiopathic membranous nephropathy, a less common disease (1 case per 100,000 individuals per year), revealed two robust associations with larger than typical effect sizes (32).
Common SNPs at HLA-DQA1 and PLA2R1 both demonstrated strong signals, and surprisingly, the allelic effects of the two SNPs appear to be multiplicative, potentially indicating a common mechanism of pathogenesis (33). The association for PLA2R1 is particularly exciting, because it confirms suggestive results from small candidate gene studies (34, 35) as well as substantive biological evidence that PLA2R1 is an autoantigen in many individuals with membranous nephropathy (36).

**Kidney disease in individuals of African descent**

African Americans have 4- to 5-fold more kidney failure than European Americans (1). Although old reports have hinted at this disparity for some time, it was neither widely appreciated nor reliably quantified until relatively recently (39, 40). Furthermore, the phenomenon remains poorly characterized. Do we mean African Americans per se, or any individual with recent African ancestry? Does regional African ancestry matter (i.e., Kenya versus Nigeria)? And how do we define kidney failure, given that so many people with kidney dysfunction die of other related cardiovascular diseases, rather than from ESRD itself?

Large differences in disease prevalence between two ancestral groups (even if highly heterogeneous, such as Africans and Europeans) suggest that relatively common genetic variants must account for these differences. It would be statistically unlikely and biologically implausible for many different rare, independent variants in TRPC6, ACTN4, NPHS2, and other kidney disease genes to all be more common in African Americans than European Americans.

In 2008, two studies took advantage of this disparity in renal disease prevalence to demonstrate that a locus on chromosome 22 conferred most or nearly all of the increased risk for nondiabetic kidney disease in African Americans (41, 42). Mapping by admixture linkage disequilibrium (MALD) uses differences in allele frequencies between populations with differing susceptibility to disease to determine chromosomal regions that are likely to harbor disease-causing genes (reviewed in ref. 43). A large excess of African ancestry among African Americans with either FSGS, hypertension-associated ESRD (H-ESRD), or HIV nephropathy (HIVAN) at this single locus on chromosome 22 (Figure 1) confirmed a genetic basis for the observed ancestry-related disparity in renal disease prevalence. Initial studies suggested that the disease-causing variants might be located in the MYH9 gene, but fine mapping and sequencing did not identify any convincing causal variants.

**Figure 1**

APOL1 risk variants: admixture and MALD. (A) APOL1 risk variants arose approximately 4,000 years ago in Africa and rose quickly to high frequency. In Yoruba (Nigeria), approximately 46% of chromosomes contain either the G1 or G2 allele. The ancestors of modern Europeans left Africa many millennia before the origin of these risk alleles, so the risk alleles are not found in Europeans. Today, approximately 36% of all African Americans (AA) carry the G1 or G2 alleles. (B) MALD. In the US, African Americans have roughly 80% African ancestry and 20% European ancestry. Since African Americans have much more renal disease than European Americans, MALD tests for chromosomal regions in African Americans with renal disease (cases) that are disproportionately African in origin. Genome-wide, there is no excess African ancestry in renal disease cases. However, at the chromosome 22 locus, there is a marked excess of African ancestry in cases. The odds of this distribution by chance alone are several billion to one (41, 42).
In the case of the chromosome 22 region identified by MALD, the size of the nephropathy-associated haplotype containing the risk alleles is considerably longer than the haplotypes of the non-risk alleles (44). The unusual length of this risk haplotype suggests the occurrence of a selective sweep: natural selection may have favored the renal risk allele for some advantage it conferred or favored another allele in close linkage disequilibrium (LD) with the renal risk allele. The implication is that natural selection swept the renal disease risk allele to high frequency despite its potentially deleterious effects. The presence of a selective sweep also led to an initial underestimation of the size of the risk interval (41, 42). Factoring in the long haplotypes that are associated with selective sweeps, the risk interval could then be expanded beyond MYH9 to include the apolipoprotein L (APOL) family of genes.

APOL1
The 1000 Genomes Project aims to catalog genetic variations in diverse human populations (45). Although there are many genetic variants with large differences in allele frequency between native west Africans and Europeans at the chromosome 22 locus, two particular coding sequence variants in APOL1 — not cataloged in HapMap, but identified in 1000 Genomes data — demonstrated by far the strongest association with kidney disease in African Americans (44, 46). Importantly, these S342G and I384M substitutions (in almost complete LD; referred to as G1) and a deletion of two amino acid residues, N388 and Y389 (referred to as G2), alter the amino acid sequence of the encoded ApoL1, which suggests functionality. While identified initially by testing African Americans with FSGS, a similarly strong effect on the risk of H-ESRD was also demonstrated. In these initial studies, individuals with one risk allele (G1 or G2) have no or only minimally increased kidney disease risk. However, individuals with two risk alleles (one on each chromosome) have 7- to 10-fold increased risk of FSGS and H-ESRD. Thus, APOL1-mediated kidney disease risk follows essentially autosomal-recessive inheritance, unlike most common complex disease variants that follow additive or multiplicative patterns.

The G1 and G2 alleles are common; on the order of 50% of African American individuals have either one or two risk alleles, and 10%-15% have two risk alleles (44). The unusually large effect size,
high frequency, and recessive mode of inheritance raises the question of whether APOL1-mediated renal disease in African Americans is more accurately characterized as a Mendelian disease with important modifiers rather than as a complex disease.

**APOL1 influences the risk of different forms of kidney disease**

The G1 and G2 APOL1 variants exert similar magnitudes of effect on the risk of both FSGS and H-ESRD (44). This observation suggests that a common mechanism may be a major factor in these two disease entities, as some clinicians have long suspected. That the effect size is so large for both diseases may mean that the overlap in etiology is quite substantial. This difference between H-ESRD and FSGS likely reflects, at least in part, the timing of clinical evaluation, the presence of modifier genes (or epistatic effects), and perhaps the nature of environmental exposures.

The APOL1 genotype may prove to be an even stronger factor in HIVAN than FSGS or H-ESRD, with odds ratios estimated to be as high as 30. HIVAN is characterized by marked proteinuria and, often, an aggressive histologic lesion referred to as collapsing nephropathy, in which FSGS is accompanied by podocyte proliferation and the appearance of collapsed glomerular capillaries (47). Widely observed to be much more common in black individuals (48–50), the overwhelming association with the chromosome 22 locus provides a genetic explanation for this difference. Rates of HIVAN in Africa correlate with frequency of chromosome 22 risk phenotypes (46, 51, 52). If the assumption is correct that H-ESRD, FSGS, and HIVAN share a common etiology, a logical implication is that hypertension may not in fact lead to ESRD in the vast majority of cases diagnosed as hypertensive nephrosclerosis. Rather, hypertension-associated kidney disease (or hypertensive nephropathy, hypertensive nephrosclerosis, or whatever synonym one chooses) may be primarily a renal syndrome that precedes and potentially causes hypertension. This hypothesis will be difficult to prove, but has major implications for treatment of both renal disease and hypertension.

**The origins of APOL1 renal risk variants**

It is interesting to consider how APOL1 renal risk variants rose to such high frequency in individuals of African descent. Hemoglobinopathies, paradigmatic examples of an evolutionary compromise, may afford some insight. In malaria-endemic parts of the globe, numerous variants in the hemoglobin gene are observed at high frequency (53, 54). Whereas two copies of variant hemoglobin lead to several well-defined human disorders (e.g., sickle cell disease and the thalassemias), individuals with one copy develop less severe malarial disease. The frequency of sickle cell allele in hemoglobin (HbS) varies greatly, globally and within Africa, in relation to the historical endemicity of malaria. Recent evidence suggests that the allele arose independently multiple times in history, as HbS is found on several different haplotype backgrounds (55, 56). The survival advantage conferred by the HbS allele (a dominant trait) caused it to increase in frequency until the recessive disease became common.

The biology and genetics of APOL1 tell a similar evolutionary story. Circulating ApoL1 can lyse *Trypanosoma brucei* and various other trypanosoma, and this innate immune function protects humans against the trypanosomes that infect cattle and other animals. While humans are resistant to *T. brucei* because of the presence of ApoL1, a subspecies of this parasite emerged (*T. brucei rhodesiense*) that is entirely resistant to Apol1-mediated lysis (57). A T. b. rhodesiense virulence factor called SRA confers this resistance, inactivating ApoL1 by binding its C terminus and preventing parasite lysis (58). ApoL1 protein harboring G1 and G2 variants, located at the site of the Apol1/SRA interaction, can escape SRA-mediated inactivation, providing a (dominant) selective advantage for these variants (ref: 44 and Figure 2). Two copies of variant Apol1 greatly increase risk of renal disease in a recessive fashion, although the penetrance is clearly lower than that for sickle cell disease. Presumably, in areas with high sleeping sickness burdens, variant APOL1 frequency tends to rise until homozygotes become relatively common. In the US and other non-African locales, where there is no apparent selective advantage to variant Apol1, we see only the deleterious consequences.

**Traces of natural selection in the human genome**

The hypothesis that APOL1 renal risk variants were selected for their advantageous properties in the recent past (~10,000 years) is supported by patterns of genetic variation among modern humans. In general, a new mutation occurs once at a specific locus and remains rare or disappears. The vagaries of population bottlenecks, human migration, mating preferences, and other variables may cause new mutations to reach substantial frequencies in some population groups (referred to as genetic drift) independent of any specific consequences of that mutation (2, 59). This increase in allele frequency takes time; therefore, if an allele becomes common as a result of drift, it is likely very old. Variant alleles that are common in Europeans, Africans, and Asians originated at least 50,000–100,000 years ago, before modern humans migrated out of Africa. If, however, a new mutation confers a selective advantage, it can reach high frequency in a specific population much more rapidly (known as a selective sweep) than expected by drift alone. A recently selected allele that rises rapidly in frequency will tend to
be on a longer haplotype than the nonselected allele at that locus (60). Long haplotype tests for positive selection compare the length of the selected allele’s haplotype with the haplotype length of the alternate allele (60). In the case of the G1 and G2 APOL1 haplotypes, these calculations give strong statistical support for positive selection acting on these alleles over the last 3,000–10,000 years in Africa. Because these DNA changes developed in sub-Saharan Africa after the out-of-Africa migrations of 20,000–100,000 years ago, they are exceedingly rare or nonexistent among people without recent African ancestors (e.g., last few millennia) (44).

Some disease-causing mutations in kidney disease genes occur at very highly conserved loci. For example, the glutamic acid that constitutes residue 255 in human α-actinin-4 (ACTN4) is a glutamic acid in the mouse, yeast, and Drosophila homologs (5). It is not surprising that a nonconservative amino acid substitution at this residue alters this protein’s function, leading to human disease (5). In contrast to ACTN4 and a number of other Mendelian kidney disease genes, the APOL family of genes is relatively new and rapidly changing in response to environmental pressure. Humans have six APOL genes, arranged in a cluster on chromosome 22 (61). APOL1 is a recently evolved member with a limited distribution in the animal kingdom, existing only in humans and a few other primate species (62). For new or nonconserved genes like APOL1, disease-causing variants might be thought of quite differently, with changes in these genes influencing specifically primate—and perhaps uniquely human—phenotypes.

Many puzzles

Although we believe that these two APOL1 variants explain the high rate of kidney disease in African Americans and have risen to high frequency because of a selective advantage that resistance to pathogens has conferred, several puzzles remain to be solved. The renal risk alleles appear to be most common in western sub-Saharan Africa, whereas T. b. rhodesiense is currently endemic in eastern Africa. It is conceivable that distributions of trypanosomes have changed over time, or even that the very high frequency of trypanolytic APOL1 variants in West Africa may have altered trypanosomal distributions through herd immunity. T. b. rhodesiense may not be the only factor driving natural selection; a second trypanosomatida, Leishmania, is also targeted by APOL1 (63). It is certainly possible that ApoL1 has broad innate immunity properties, shaping the relative frequencies of the APOL1 alleles. Just as mysterious are the biological mechanisms by which these variants lead to human kidney disease. Many reasonable hypotheses can be generated regarding the effect of either circulating or intracellular APOL1 on kidney function. Based on what little is known about ApoL1 biology beyond its antitrypanosomal effect, roles in apoptosis, autophagy, ion channel function, and lipid biology could all be important for kidney function in health and disease. Both extensive deep sequencing and kidney-specific functional evidence will be critical in validating and understanding the role of ApoL1 in kidney pathophysiology.

Clinical implications

The presence of two APOL1 high-risk variants is not sufficient to cause disease. It is highly likely that other genetic and environmental factors with highly variable frequencies modify the penetrance and expressivity of the APOL1 genotype. However, the unusually large relative risk associated with this genotype suggests that there may eventually be value in regarding APOL1-mediated renal disease as a distinct clinical entity. CKD presenting in a middle-aged African American individual is more likely than not associated with this high-risk (two-allele) APOL1 genotype. On the other hand, the likelihood that this genotype is present depends heavily on the specifics of an individual’s ancestry. The frequency of the APOL1 risk alleles is essentially zero in Ethiopia, which likely explains the rarity of HIV-associated nephropathy in this population (52). Americans of Ethiopian descent, therefore, are not expected to be at high risk of kidney disease from H-ESRD or FSGS, either. At present, the specifics of how APOL1 genotype should alter clinical care are unknown and deserving of study. It will be important for clinical studies to determine whether any treatment aimed at slowing either kidney disease progression or the underlying disease process will have different efficacy in individuals with or without APOL1 risk alleles.

Implications for understanding health disparities

If we are to reduce major disparities in the rates of serious disease between different ethnic or racial groups (however poorly defined these categories may be), we need to understand the genetic and environmental bases of these disparities. In the case of kidney disease, this means rethinking several widely held beliefs. It is too imprecise, and in some cases misleading, to say that African Americans as a group have 4-fold more kidney disease than European Americans. It is somewhat more accurate to say that a subset of black individuals has a 7- to 30-fold increase in kidney disease, depending on the specific kidney phenotype in question. It is better yet to point to the two common coding sequence variants in APOL1 that arose recently in Africa as the cause of these renal diseases, an approach that may benefit the specific individual. As we come to better understand how these variants have spread through populations throughout the globe, we should see a wider appreciation of the fact that neither skin color nor generalizations about ancestry are good predictors of kidney disease for an individual patient.

Implications for new kidney gene discovery

Recently selected kidney disease variants will have different properties than do rare, Mendelian disease genes or common, complex disease genes (Figure 3). Common variants are generally weak, and Mendelian variants are rare, but genes under strong positive selection provide a unique instance in which a variant can have an unusual combination of frequency and penetrance. Many of these variants are unlikely to be widely shared by diverse populations. Any variant common among Europeans, Africans, and Asians is greater than 50,000 years old, occurring before the out-of-Africa migration (64). Mutations with powerful deleterious consequences are eliminated over time if positive selection pressure is relaxed, so it is unlikely that many variants with large effect sizes could persist at high frequencies in environments as different as Europe, Africa, and Asia over a thousand generations. Thus, we should not dismiss the existence of other common, highly penetrant variants, such as APOL1 risk alleles, but rather expect them to be found mostly in specific populations. It is also thought provoking that the genes with the best evidence for association with kidney disease — APOL1, UMOD, PLA2R1/sPLA2 (complement factors, and the MHC loci — all appear to be important in defense against pathogens.

A second issue raised by APOL1 genetics is the role of animal models of kidney disease for understanding human kidney disease. APOL1 is a gene found only in humans and a few primate species, such as baboons and gorillas; not even chimpanzees have functional
APOL1 (62). The human glomerulus, especially the podocyte, is a highly differentiated structure that must endure many physical stresses and toxic insults over many decades, with little apparent capacity for replacement. It is tempting to speculate that ApoL1 may play some role in maintaining the integrity of the glomerulus over a lifetime, especially given its putative role in autophagy and the recently appreciated importance of autophagy for podocyte health (65, 66). The major role of APOL1 variants in human kidney disease does highlight a case in which rodent models are better suited to disease validation and mechanistic understanding than to kidney disease gene discovery.

Future challenges

New technologies and approaches will revolutionize how we connect genotypes and phenotypes for kidney disease. Exome sequencing has already led to identification of genes for rare diseases in families and has the potential to do the same for common diseases as costs decrease. Detailed and thoughtful phenotyping by clinicians will be more important than ever in refining cohorts for analysis. Epigenetic changes and gene-environment interactions represent additional exciting avenues for investigation.

Genetics, as a research discipline, provides insight into biologic mechanisms underlying phenotypic diversity as well as basic biology, physiology, and development. Some challenges, such as understanding mechanisms by which variants in a noncoding sequence of unclear function influence phenotype, will be exceedingly difficult. The discovery of surprising new mechanisms by which DNA sequence regulates cellular function continues to point to much greater complexity in the genome than was entertained just a few years ago. Identification of disease-associated variants will always represent a beginning, not an end, for genetic studies. Discovering how these gene variants cause disease is at least as important as their identification on the path to better methods of treatment and prevention. It will be exciting to watch our understanding of kidney disease evolve.

Address correspondence to: Martin Pollak or David Friedman, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, Massachusetts 02115, USA. Phone: 617.667.0461; Fax: 617.667.0495; E-mail: mpollak@bidmc.harvard.edu (M. Pollak). Phone: 617.667.0283; Fax: 617.667.0495; E-mail: dfriedma@bidmc.harvard.edu (D. Friedman).


