Integrating Human and Rodent Data to Identify the Genetic Factors Involved in Chronic Kidney Disease

Michael R. Garrett,* Marcus G. Pezzolesi,† and Ron Korstanje‡

*Department of Medicine and Kidney Disease Center, Medical College of Wisconsin, Milwaukee, Wisconsin; †The Research Division, Joslin Diabetes Center, and Harvard Medical School, Boston, Massachusetts; and ‡The Jackson Laboratory, Bar Harbor, Maine

ABSTRACT

The increasing numbers of patients with chronic kidney disease combined with no satisfying interventions for preventing or curing the disease emphasize the need to better understand the genes involved in the initiation and progression of complex renal diseases, their interactions with other host genes, and the environment. Linkage and association studies in human, rat, and mouse have been successful in identifying genetic loci for various disease-related phenotypes but have thus far not been very successful identifying underlying genes. The purpose of this review is to summarize the progress in human, rat, and mouse genetic studies to show the concordance between the loci among the different species. The collective utilization of human and nonhuman mammalian datasets and resources can lead to a more rapid narrowing of disease loci and the subsequent identification of candidate genes. In addition, genes identified through these methods can be further characterized and investigated for interactions using animal models, which is not possible in humans.


The extent of chronic kidney disease (CKD) is a growing national concern. Approximately 25.6 million people in the United States exhibit some degree of kidney injury and/or decline in kidney function.1 Some estimate the number of CKD patients progressing to end-stage renal disease (ESRD) ballooning to more than 2 million in the next 20 years.2 Current treatment options are limited and serve only to slow progression, not to cure or reverse specific conditions. Therefore, understanding the genetic basis of kidney disease is of considerable importance because it could provide early diagnosis or more options for potential targets useful in developing new treatments.

HUMAN LINKAGE AND GENOME-WIDE ASSOCIATION STUDIES

The primary goal of studying the genetics of complex disease in humans is to identify those genetic variants that either predispose or cause the disease. Evidence of familial aggregation and the heritability of biomarkers, such as albuminuria, for the development and progression of CKD, as well as various indices of renal function, provide a foundation for investigating the genetic predisposition of CKD.3–12 Efforts to map loci contributing to CKD susceptibility often rely on family-based linkage approaches, which look to establish statistical associations using genetic markers with a given dichotomous (affected versus unaffected) or quantitative (GFR or degree of albuminuria) phenotypic trait.13 The identification of genomic locations demonstrating linkage provides the first line of evidence suggesting where disease-associated genes are likely to reside. The power of linkage analysis in CKD is best demonstrated by the study of rare familial, Mendelian forms of nephrotic syndrome, culminating in renal failure (Supplementary Tables 1 and 2). In these cases, there has been strong linkage between observed phenotypes and specific chromosomal regions. This approach has yielded several successes, culminating in the identification of mutations in several key genes (NPHS1, NPHS2, ACTN4, TRPC6, and PLCE1) that play structural and functional roles in maintaining the glomerular filtration barrier.14–18 Unraveling the genetics of more common forms of kidney disease has been slower to progress, largely because of phenotypic limitations of these studies, their ability to only detect major genetic effects or their reliance on family-based collections and, perhaps, misconceptions about the underlying genetic model contributing to their susceptibility; that is, the existence of major disease loci and potential gene–gene and gene–environment interactions.

More recently, genome-wide association (GWA) studies, examining as many as 300,000 to 1 million single-nucleotide polymorphisms (SNPs) across the entire genome in thousands of unrelated patient and control subjects, are powerful approaches to detecting disease loci. The recent success of several GWA studies has helped improve our relative under-
standing of the genetic basis of complex human disease. To date, more than 100 loci for diseases such as coronary heart disease, type 1 and type 2 diabetes, bipolar disorder, Crohn disease, and rheumatoid arthritis have been identified using this approach, providing valuable insight to previously unsuspected biologic pathways and improving our understanding of the allelic architecture that underlies these conditions.\textsuperscript{19,20} These studies reveal the effect size of common variants to be more modest than previously suspected, with odds ratios per risk allele generally less than 1.4.\textsuperscript{19} Not surprisingly, GWA studies have proven far superior to linkage-based studies, which are powered to identify major loci—loci with effect sizes greater than 2.0—in identifying genetic variants in common diseases.

The promise of GWA approaches has recently begun to reap rewards in CKD, as five GWA studies have been published to date (Supplementary Table 3). The strongest association identified thus far, and one of only two reported associations to reach a level indicative of “genome-wide significance,” was reported in a meta-analysis of 41,000 subjects, including more than 4300 patients with CKD.\textsuperscript{21} In this study, Kottgen et al. identified significant associations at the uromodulin (\textit{UMOD}) locus on chromosome 16p for both CKD and the quantitative analysis of GFR. Three GWA studies have been conducted in patients with diabetic nephropathy.\textsuperscript{22–24} Variants in the engulfment and cell motility 1 (\textit{ELMO1}) gene on chromosome 7p were initially associated with diabetic nephropathy in a GWA study performed in a Japanese cohort.\textsuperscript{22} Variants at this locus have since been reported in two large African-American cohorts with type 2 diabetes and ESRD and in a collection of Caucasian type 1 diabetic patients from the Genetics of Kidneys in Diabetes (GoKinD) study.\textsuperscript{25,26} Interestingly, this same region has also been linked with ESRD and variation in GFR.\textsuperscript{27,28} A second GWA study in Pima Indians with type 2 diabetes identified a strong association in intron 8 of the plasmacytoma variant translocation (\textit{PVT1}) gene.\textsuperscript{23} In the largest GWA study of diabetic nephropathy to date, Pezzolesi et al. recently identified associations on chromosomes 7p, 9q, 11p, and 13q.\textsuperscript{24} Lastly, in a GWA study of participants from the Framingham Heart Study, Hwang et al. reported associations with albuminuria at loci on chromosomes 11q and 21q.\textsuperscript{29}

As expected, the odds ratios detected in the majority of GWA studies of CKD have been rather modest, ranging from 1.25 to 1.45—effect sizes far below the threshold detectable using linkage-based approaches. Despite the early successes of GWA studies in identifying loci associated with CKD, a number of challenges still lie ahead that must be met to advance our understanding of the genetic factors that cause these diseases. At the forefront of these challenges is the need for replication. Although support for several of the loci identified by Kottgen et al. and Pezzolesi et al. has been demonstrated in cohorts independent of the original association signal, additional confirmation is necessary to bolster support for these loci. Currently, a number of efforts are underway to replicate these findings in additional collections, and more GWA studies in CKD are on the horizon, including a study in the Family Investigation of Nephropathy and Diabetes (FIND) collection. Continued efforts to map CKD disease loci, specifically those sharing a common genetic mechanism of modest effect, will rely on meta-analyses of these data. However, such approaches are challenging because of the inherent phenotypic heterogeneity of these collections and the resulting datasets.

Additionally, despite offering far superior resolution in comparison with linkage-based approaches (disease loci can be localized to regions approximately 5 to 100 kbp in length in GWA studies versus several mega-base pair regions in linkage studies), GWA studies do not pinpoint the causal functional variant. A complete inventory of all correlated variants within the associated region, their functional prioritization, and additional genotyping are required to fine-map this signal. Together with the HapMap project (www.hapmap.org), the recent launch of the 1000 Genomes Project (www.1000genomes.org) promises to facilitate the identification of such variants. Pending its completion, however, resequencing relevant individuals is still currently required to establish a comprehensive catalog of potential causal variants for subsequent functional analyses. Finally, current GWA platforms poorly assess the contribution of rare variants to disease susceptibility, and analytical methodologies used to evaluate the effects of gene-gene (epistatic) or gene-environment interactions on disease remain limited.

Another result of the HapMap project is the identification of markers that differ between ancestral populations. A novel strategy for identifying genes underlying ancestry-driven diseases, like some forms of CKD, is mapping by admixture linkage disequilibrium. Using mapping by admixture linkage disequilibrium, both Kopp et al. and Koa et al. identified strong associations with focal segmental glomerulosclerosis and nondiabetic ESRD in African-Americans on chromosome 22q, and subsequently localized these associations to the myosin heavy-chain type II isoform A (\textit{MYH9}) gene.\textsuperscript{30,31} Although this is a powerful and relatively inexpensive method to identify genes that can explain the increased burden of these diseases among some ethnic groups, these genes are most likely only a small subset of the genes involved in the disease and can only be applied to very specific cohorts.

**STUDIES IN RATS AND MICE IDENTIFY CONCORDANT LOCI**

The rat has been extensively used as a model system to study the pathophysiology and genetics of kidney injury or disease,\textsuperscript{32} including models that also exhibit hypertension. A number of well-characterized models, including Dahl S, FHH, MWF, and SBH, have been used for linkage analysis of kidney-related traits, including proteinuria, albuminuria, serum creatinine, and creatinine clearance. To date, complete genome scans have been performed on 13 experimental crosses, identifying more than 40 quantitative
trait loci (QTLs) linked to proteinuria alone (Supplemental Table 4). Many of the QTLs have been confirmed and the chromosomal locations further refined using congenic strain analysis,33–40 a method more often used in the rat. Congenic strains are developed by transferring an entire chromosome (or chromosomal segment) from one strain onto the genetic background of another through an iterative breeding paradigm that continues until the QTL interval is reduced to contain a relatively small set of genes, allowing for a detailed analysis of each individual gene contained in the linked region.41 To date, the majority of congenic rat strains have been developed to study cardiovascular and renal traits and, importantly, such tools have proven vital in further refining the causative disease locus in these models.42

Although not as extensively used as the rat, several groups have reported linkage studies in mice over the past few years, mostly using standard inbred strains (C57BL/6J, BALB/c, KK/TaJcl, 129S1/SvImJ, 129S6/SvEvTac, A/J, and DBA/2J), but also using a hyperlipidemia-prone Apoe knockout strain43 and the FGS/Kist strain,44 which was selectively bred for proteinuria over several generations. Including two reports using strains that are considered models for lupus nephritis,45,46 a total of 18 loci have been mapped for renal damage phenotypes (Supplemental Table 5). Contrary to studies in rat, most of the loci identified in mouse are single observations, with the exception of the distal locus on chromosome 2, which was found in six independent crosses. The multiple linkage analyses performed in rat and mouse reveal nearly all chromosomes harbor at least one renal-related QTL. Interestingly, several of these loci are concordant between the two species and, even more importantly, many of these loci overlap with homologous regions in humans (Figure 1). Some QTLs are repeatedly detected in multiple genetic linkage analyses using different rodent strains, likely indicating these loci play more prominent roles in contributing to renal injury relative to other chromosomal regions. Moreover, the striking concordance between human, rat, and mouse loci, even when different renal phenotypes are considered, suggests common disease mechanisms link these “subphenotypes” of renal damage.

COMBINING HUMAN, RAT, AND MOUSE DATA TO NARROW GENETIC LOCI

There are several approaches one can use to combining data:

**Comparative Mapping**

Interdisciplinary approaches that incorporate data from animal models, including both mouse and rat, with those from human studies provide valuable insight to understanding the causative genes that underpin CKD, and they offer a complementary approach to identifying the causative disease genes in humans. Comparative mapping using combined data across species is also a useful tool to narrow loci that are concordant between species. Because of chromosomal rearrangement between the species, the homologous regions will not completely overlap, and the regions that do not overlap can be excluded. This approach was recently used to narrow a locus influ-

---

**Figure 1.** Ideograms of human chromosomes with concordant human, mouse, and rat CKD loci are shown. Human CKD loci identified using linkage (blue) and GWA studies (O−) are indicated on the left side of the chromosome. Concordant loci found in rat (green) and mouse (purple) are shown on the right side of the chromosome.
The proteinuria QTL was successfully narrowed to a small genomic segment through congenic strain analysis. Another example is the previously mentioned locus on mouse chromosome 2 located on chromosome 1,47–49 whereas no linkage has been observed on the homologous region on chromosome 4. Together, these data provided the means to narrow the potential list of candidate genes to focus on in the rat.50 Another example of comparative mapping centers on the neophropathy locus identified on human chromosome 9q in the GoKinD population24 containing 10 potential disease genes (Figure 3A). The proximal portion of this locus is homologous to mouse chromosome 4 and falls within the 95% confidence intervals of the QTL found in the (C57BL/6J × DBA/2J)F2 cross51 and the (C57BL/6J × NZM)F1 × NZM cross.45 The distal portion of this region is homologous to a region on mouse chromosome 13 in which no QTL are found. Assuming the same underlying gene contributes to the association and linkage observed in human and mouse, respectively, all genes in the interval for which the orthologous gene is found on mouse chromosome 13 can be excluded as likely candidates, leaving only Frmd3 and Rasef as potential disease genes.

**Region-Specific Haplotype Mapping**

Most of the inbred strains of mouse and rat are closely related, and parts of their genomes are identical between strains. Because a QTL can occur only when there is variation in the QTL-causing gene, we can use this characteristic to narrow QTL regions by excluding regions of the genome where shared haplotypes exist between the two strains used for a particular cross.

Considering the human chromosome 9q locus described above, this approach allows further narrowing of the region already reduced using comparative mapping. The mouse chromosome 4 region containing Frmd3 and Rasef as candidate disease genes was found in two crosses: (C57BL/6J×DBA/2J)F2 and (C57BL/6J×NZM)F1 × NZM. When comparing haplotypes between B6 and D2, we conclude that B6 and D2 share a haplotype in the region where Rasef is located, and because there is no genetic variation between the two strains along this haplotype, this region could not cause the QTL identified in this cross, and thus we can discount this region as one likely to contain the disease genes (Figure 3B). The same is true when comparing C57BL/6J with NZM. Together, this suggests that Frmd3 is the most likely candidate gene at this locus.

Another example is the previously mentioned locus on mouse chromosome 2 found in five different mouse crosses and on the homologous rat chromosome 3 in two crosses (Supplemental Table 4). If we assume it is the same gene underlying these QTLs, we can combine the mouse data for haplotype analysis. However, for the (C57BL/6J × A/J)F2 and the (C57BL/6J × 129S1) × 129S1 crosses, we observe the B6 allele as the low allele (low albuminuria, low Col1 deposits, respectively), whereas in the (C57BL/6J × 129S6)F2 cross, we observe the B6 allele as the high allele (increased glomerulosclerosis; Thu Le, personal communication). This suggests in the latter cross the QTL is caused by a different polymorphism. Although 129S1 and 129S6 are related, recent high-resolution genotyping using the Mouse Diversity Array52 shows regions with large variation between the two strains. Com-

**Figure 2.** Comparative map shows overlap of renal susceptibility loci between rat and human. The physical map of the rat QTL on chromosome 2 is shown on the left. The proteinuria QTL was successfully narrowed to a small genomic segment through congenic strain analysis.50 The region in human that is homologous to the rat QTL lies on both human chromosomes 1 and 4. No linkage for any renal-related traits has been observed on human chromosome 4 (151 to 155 Mb), whereas linkage with human chromosome 1 (152 to 157 Mb) has been observed in several studies.47,48,60 Taken together, the region of concordance between the rat and human (1q21) allows the rat QTL to be narrowed by approximately 50%, along with the number of likely candidate genes. Map distances are in base pairs (www.ensembl.org, Ensembl v38; April 2006). Reproduced with permission from APS.50
The 2-Mb interval on chromosome 9 flanking associations identified in the GWA study of the GoKinD collection contains 10 annotated genes. The proximal 1-Mb region is homologous with a region on mouse chromosome 4 that is linked to albuminuria QTL that have been mapped in two crosses ((C57BL/6J × DBA/2J)F2 and (C57BL/6J × NZM)F1 × NZM), whereas the distal 1-Mb region is homologous to a region on mouse chromosome 13 for which no linkage with renal phenotypes has been found. On the basis of the overlapping signal, concordance mapping at this locus implicates only RASEF and FRMD3 as potential candidate CKD genes. (B) QTL mapping detects chromosomal regions that contain genetic variance between the strains used in a particular cross. Regions that are genetically identical between the two strains, conversely, cannot be linked to the trait of interest. The black bars indicate the regions that are genetically different between the parental strains used for the two crosses. Because haplotype information was not available for NZM, we used data from NZB and NZW, which are the progenitors of NZM. Because RASEF is in a region with no genetic variance in both the C57BL/6×DBA/2J and C57BL/6×SM crosses, FRMD3 is the most likely candidate gene for this concordant locus in human and mouse.

Figure 4. Interval-specific haplotyping and human concordance narrow a locus for albuminuria. A QTL for albuminuria was found on the distal part of mouse chromosome 2 in multiple crosses. Interval-specific haplotyping eliminates the regions in which the two parental strains for each cross were genetically identical and resulted in three small intervals with 14, 60, and 13 genes, respectively. The proximal region is homologous to human chromosome 20p, for which linkage was found in human, whereas the distal region is homologous to human chromosome 20q, for which no linkage has been found. On the basis of the concordance at both the mouse and human loci, the candidate gene is most likely among the 14 genes within the proximal region.

comparison of the chromosome 2 region does not show differences between the two substrains (Fernando Pardo-Manuel de Villena, personal communication), but it leaves open the possibility of a modifier gene at another locus, which could explain the differences between these two strains.

When performing haplotype analysis with the strains in the five crosses (after excluding the B6 by 129S6 strain combination), we look for a haplotype that is shared by B6, PL, and BALB (the strains providing the resistant allele) and a haplotype that is shared by NZW, A, KK, DBA/2, and 129S1 (the strains providing the susceptible allele). In addition, the haplotypes must be different between the two groups. Only three small regions on chromosome 2 fit all of these criteria. The first region is approx. 0.7 Mb large and contains 14 genes, the second region is 1 Mb and contains 25 genes, and the third region is only 50 kb and contains two genes (Figure 4). Here, we can also use comparative mapping: the first region is concordant to a human QTL for diabetic nephropathy on chromosome 20p, whereas the other two regions are homologous to human chromosome 20q, with no known linkage or association with renal phenotypes. Among the 14 genes, only one has a nonsynonymous SNP that matches the strain distribution described above. That gene is the P102S change in Rhrbp1, encoding the ribosome-binding protein. Gene expression studies are needed to test the 14 genes for expression differences among the eight strains. At this point, the above methodology can only be employed in the mouse because the necessary resources in the rat have not been developed. Recently, a large-scale SNP discovery project for the rat (STAR Consortium) identified almost 3 million new SNPs, of which 20,238 were evaluated across 167 distinct inbred rat strains and two recombinant inbred panels. These new genetics resources will provide the tools to perform similar analysis as described for the mouse. Because the QTL in the example described above is also found in the rat, it could confirm the mouse analysis and maybe even narrow the region further.

**ANIMAL MODELS TO TEST CANDIDATE GENES**

As more candidate genes are identified by GWA studies in humans, genetic
studies in animal models, or a combination of both, there will be a need to validate the candidacy of these genes and establish a model in which we can explore the mechanism and possible intervention. There are several tools available both in mouse and rat, and the choice will depend on the nature of the gene and the likely effect of found polymorphisms. When it is believed that a dysfunctional gene is involved in renal damage, a (conditional) knockout model would probably be the best choice. However, when it is overexpression of a gene or a change in the protein structure, one might want to establish a transgenic animal in which the gene is overexpressed or the gene with the altered coding sequence is introduced.

**Knockouts**

Until recently, making a knockout animal with null alleles was restricted to mice because of the need for embryonic stem cells, which are not available for rats. The more classical approach, in which an exon is interrupted by a selective marker, is the easiest but often problematic because the gene is either involved in development or is crucial to processes in other organs, leading to embryonic lethality. The solution to this is making a conditional knockout. Critical coding exons are flanked by target “loxP” sites susceptible to Cre recombinase. The reporter-tagged allele can be converted to the lacZ reporter-tagged null allele by exposure to Cre recombinase, causing a frame-shift mutation and likely triggering nonsense-mediated decay of the mutant transcript, thereby ensuring a null allele. Once the loxP mouse with the desired floxed exon is created, it needs to be crossed to a cre mouse, which contains a Cre recombinase transgene under the direction of a tissue-specific promoter. Many different kidney-specific cre mice have been developed, making it possible to knock out the candidate gene in specific parts of the kidney. Recently, a new approach formed on the basis of zinc finger nuclease methodology has successfully accomplished the first targeted gene knockout in the rat. This emerging technology provides the means to test candidate genes for renal injury in the rat, and first experiments are on the way.

**Transgenics**

The insertion of additional copies of a gene either with its own promoter or driven by a strong exogenous promoter is possible both in mouse and rat. To date, there have been no candidate genes for kidney injury identified via linkage analysis and subsequently followed up using a transgenic approach. However, there are a number of transgenic models, such as Tg human renin and/or angiotensinogen genes that exhibit increased hypertension and renal injury—among many others. Recently, one group developed a transgenic rat with podocyte-specific expression using the human podocin promoter (NPHS2). This study examined the effect of podocyte depletion (through Tg diptheria toxin receptor) and glomerulosclerosis.

**CONCLUSION**

Human GWA studies are a powerful new approach for the identification of loci associated with CKD but require large cohorts and subsequent replication. Although these studies have been successful in identifying numerous disease loci, this approach is limited in its ability to identify causative variants and disease genes that underlie these associations. Data from both rat and mouse studies can complement such human studies and oftentimes can reduce the region of interest while offering additional insight that is currently difficult to ascertain in human studies alone, including analyzing epistatic and gene-environment interactions. Animal models also provide complementary, inexpensive approaches that are effective in pinpointing the causative genes in human. Importantly, the genes identified in animal studies must be verified in human to demonstrate their relevance in human health.

An interdisciplinary approach that incorporates data from animal models with that from human studies can provide valuable insight into our understanding of causative genes that underpin CKD. In this review we have shown that there is concordance in CKD loci between species and the possibilities of narrowing loci by combining human and animal data and resources. Combining data will speed up gene discovery that allows us to better understand the disease mechanisms and identify potential targets for intervention. To achieve this we need careful annotation of each of the loci in the different species and close collaborations between nephrologists and human and mammalian geneticists.

**ACKNOWLEDGMENTS**

M.R.G. is supported by National Institutes of Health (NIH)/National Heart Lung and Blood Institute grant HL094446 and funds from Advancing Healthier Wisconsin; M.G.P. is supported by NIH grants DK77532 and T32 DK007260-32; and R.K. is supported by NIH grant DK069381. We thank Joanne Currer for writing assistance and Jesse Hammer for preparation of the figures.

**REFERENCES**

5. Freedman BI, Spray BJ, Tuttle AB, Buckelew VM Jr: The familial risk of end-stage renal