see original article on page 201

The genetics of albuminuria: from haplotype association mapping in mice to genetic causation in humans

Roel Sterken[1] and Krzysztof Kiryluk[1]

Genome-wide haplotype association mapping (HAM) in inbred mouse strains emerged as an efficient method for identifying novel quantitative trait loci for disease-related phenotypes. In this issue of *Kidney International*, Tsaih *et al.* present the results of the first HAM for age-related kidney damage in mice and examine the detected loci in the context of the human genome-wide association study for diabetic nephropathy.

Kidney International (2010) **77,** 173–175. doi:10.1038/ki.2009.457

Murine genetic models of human disorders have contributed numerous insights into disease pathogenesis. Unlike the genetic variation in humans, there is relatively little genetic diversity among the commonly used strains of laboratory mice. This is primarily because most laboratory strains originated from a small number of ancestral animals.[1,2] Purposeful inbreeding of mice further limits the genetic diversity. After 20 generations of inbreeding, on average at least 98.6% of loci become homozygous.[1] Many strains have been bred for more than 100 generations and thus are completely homozygous at all loci. Such intra-strain isogenicity is particularly useful in studies that seek to eliminate genetic variation as one of the experimental variables. On the other hand, the inter-strain genetic and phenotypic differences have been the backbone of traditional quantitative trait locus (QTL) linkage mapping in mouse genetics, where the segregation of known genetic markers is tested against the segregation of quantitative phenotypes in F_2 or backcross offspring populations. While powerful, these classic QTL analyses depend on many years of expensive scientific labor to breed, genotype, and phenotype the appropriate intercrosses for analyses, as each F_2 mouse is genetically unique. In addition, because of the limited amount of recombination in these populations, the identified QTL intervals are usually very large, on the order of 40–50 Mb, making it difficult to pinpoint the causal genes.

The more recent haplotype-based methods demonstrate great promise for tackling the issues of cost, time, and resolution.[3] These methods take full advantage of the fact that inbred strains are homozygous (that is, they have no intra-strain individual genetic variation), thus their haplotype structure is fixed. This makes them quasi-immortal, presenting an opportunity to comprehensively genotype each strain once, while phenotyping can be repeated indefinitely. Additionally, replication of phenotype measurements in genetically identical organisms reduces the effect of environmental factors. Since inbred laboratory mice share common ancestors, historic recombinations have scrambled the inter-strain genetic variation, resulting in genomes that are mosaics of the ancestral haplotypes (Figure 1). The method of haplotype association mapping (HAM) treats inbred mouse lines as if they were individuals from a larger outbred population. The HAM approach tests each of the haplotype blocks across the genome for association with a phenotype of interest. Haplotype blocks are potentially powerful in association studies because each block can be distinguished by genotyping a relatively small fraction of polymorphisms. As is evident from Figure 1, the size of haplotype blocks decreases with the number of carefully selected strains, increasing the final resolution of the method.[4]

The most important progress in this type of approach can be attributed to the development of high-throughput sequencing technologies that enabled accurate discovery of single-nucleotide polymorphisms (SNPs) across the mouse genome. Dense genotypic information for an increasing number of inbred strains is becoming publicly available.[5,6] For example, the Mouse Phenome Database (http://www.jax.org/phenome) currently contains the genotype data for 16 strains assayed at over 8 million SNPs and 72 strains assayed at more than 500,000 locations, with a mixture of actual and imputed data being available for 74 strains at 7.8 million locations. Moreover, 97 different phenotypes have been mustered across inbred strains and are publicly available for download. In addition, the ongoing Collaborative Cross project aims at establishing a large panel of recombinant inbred mouse strains derived from a set of eight well-characterized inbred founder strains.[7] This project is designed specifically to provide adequate power and resolution for QTL detection in complex trait analyses. Public availability of these data holds a great potential to accelerate gene discovery for many disease-related quantitative traits.

Tsaih and colleagues[8] (this issue) applied the HAM approach to identify potential QTLs for age-related proteinuria in mice. It is well established that the aging process elicits a number of structural and functional changes in kidneys.[9] In humans, the most characteristic is the reduction of functioning kidney mass

[1]*Division of Nephrology, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York, USA*
Correspondence: *Krzysztof Kiryluk, Division of Nephrology, College of Physicians and Surgeons, Columbia University, 622 West 168th Street, PH4 Stem, Room 124, New York, New York 10032, USA. E-mail: kk473@columbia.edu*

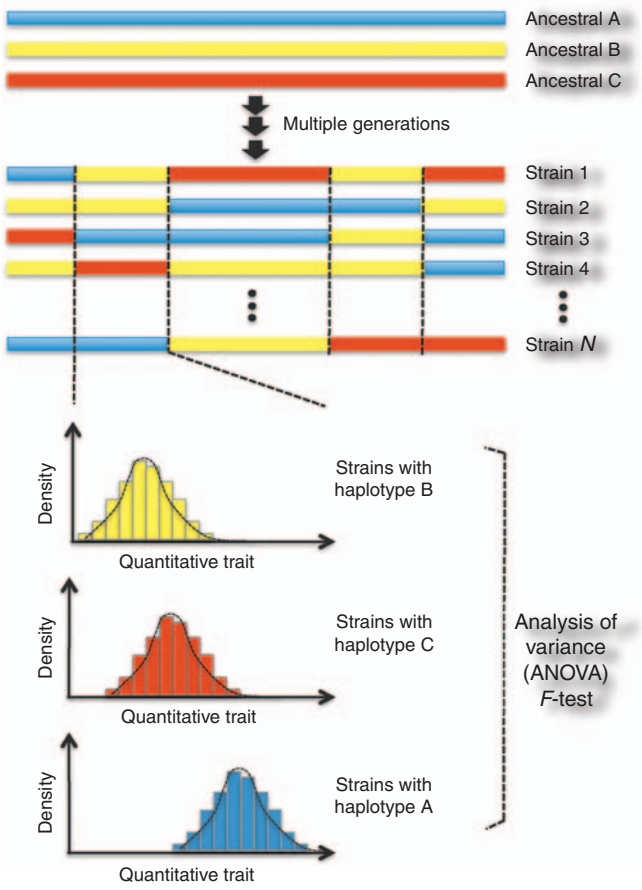

Figure 1 | Principles of haplotype association mapping in inbred strains. A limited number of ancestral haplotypes (A, B, and C) are subjected to many recombination events as generations descend. As a result, the genomic organization of an inbred strain can be viewed as a mosaic of ancestral haplotype blocks delineated by historical recombination break points. In haplotype association mapping, a pattern of haplotype blocks is inferred from single-nucleotide polymorphism genotypes with the use of computer algorithms. For each of the haplotype blocks, the distributions of quantitative phenotypes are compared between the inbred strains grouped by a shared haplotype segment. A statistical test is performed for all inferred haplotype blocks across the genome.

through a decrease in the number of viable glomeruli, leading to a progressive decline of the glomerular filtration capacity. The authors hypothesize that susceptibility to age-related kidney damage is, in part, genetically determined. Since aging in mice is marked by an increase in albuminuria, the authors determined urinary albumin-to-creatinine ratios (ACRs) in groups of 10 male and 10 female mice from 30 inbred strains of different ages (12, 18, and 24 months). In support of the role of genetic background, the ACR levels and progression rates showed striking variation across the 30 strains. In addition, the level of albuminuria in mice was profoundly affected by their sex. Next, the authors selected over 60,000 SNPs from the available public databases that tag the mouse genome at approximately 40-kb intervals. The haplotype blocks were inferred by a hidden Markov model[10] and tested for association with the ACR phenotype. The top nine suggestive QTLs were identified, each containing one to seven positional candidate genes. Tsaih and colleagues[8] attempted to confirm these observations using human genetic data for loci syntenic to the nine mouse QTLs. The genotypes for 820 cases and 885 controls from a previously reported genome-wide association study for diabetic nephropathy were used for this purpose.[11] Two significant association signals were found in these regions: one in the non-coding area between the *MYO16* and *IRS2* genes (chromosome 13q), and one in the intronic region of *NEGR1* (chromosome 1p).

This study exemplifies the utility of *in silico* analysis of mouse genetic data to accelerate the discovery of variants that are potentially involved in human disease. The availability of dense genotype data for many inbred strains allows testing of almost any phenotype that can be potentially associated with disease pathogenesis. The homozygous nature of the inbred strains eliminates most of the genetic complexity of epigenetic, recessive, or dominant effects frequently observed in heterozygotes. Moreover, the online availability of comprehensive phenotype and genotype data allows researchers to reduce genome-wide QTL searches to minutes.

Nevertheless, several limitations of this approach need to be noted. First, the genotyping data for many of the inbred strains are still incomplete, which necessitates the use of imperfect imputation methods to fill the gaps in the analyzed data sets. Second, the information on haplotype block structure continues to evolve, and it is not clear how to best define the borders between the blocks. The genotyping density and the number and choice of strains have an effect on the final haplotype block structure that is used for association testing. Therefore, defining haplotypes on the basis of SNPs spaced at 40-kb intervals might not be ideal, but it is unclear to what degree this compromises the effectiveness of a study design. Third, an underlying pattern of complex relatedness among inbred strains may inflate false-positive rates in association studies. This situation is analogous to the effect of cryptic relatedness or population structure on the results of genome-wide association studies in human populations. Although efficient methods for dealing with this problem have been developed for human studies, different statistical approaches are needed for model organisms.[12,13] Additional problems arise from the fact that ACR is a notoriously difficult quantitative phenotype to model, mainly because of its highly skewed distribution in which many individuals have a score of zero. Because parametric association

tests rely on the assumption of normal distribution, alternative methods such as non-parametric or permutation procedures are needed to derive more accurate *P* values. The trait can also be dichotomized on the basis of a predefined ACR threshold, but this approach generally leads to a significant loss of power. Future studies that examine kidney pathology scores, such as degree of glomerulosclerosis or fibrosis instead of the ACR measurements, may also circumvent this problem by providing more powerful injury-specific phenotypes. Finally, for the purpose of testing for signal concordance in humans, a careful selection of adequately powered human cohorts with disease phenotypes that are closely related to those studied in mice is necessary. Because the susceptibilities of mice to age-related albuminuria and to diabetic nephropathy may have distinctly different genetic determination, one cannot discard the mouse signals that are not observed in the human cohort. Hopefully, with the increasing number of human genome-wide association study data sets, more powerful cohorts will soon become available to make these types of approaches more feasible. Lastly, we must remain aware that significant associations are not a proof of causality but merely represent a screening method for identifying genetic variants that warrant further research.

DISCLOSURE
The authors declared no competing interests.

REFERENCES

1. Beck JA, Lloyd S, Hafezparast M *et al.* Genealogies of mouse inbred strains. *Nat Genet* 2000; **24**: 23–25.
2. Yang H, Bell TA, Churchill GA *et al.* On the subspecific origin of the laboratory mouse. *Nat Genet* 2007; **39**: 1100–1107.
3. Pletcher MT, McClurg P, Batalov S *et al.* Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* 2004; **2**: e393.
4. Cuppen E. Haplotype-based genetics in mice and rats. *Trends Genet* 2005; **21**: 318–322.
5. Frazer KA, Eskin E, Kang HM *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 2007; **448**: 1050–1053.
6. Mott R. A haplotype map for the laboratory mouse. *Nat Genet* 2007; **39**: 1054–1056.
7. Churchill GA, Airey DC, Allayee H *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 2004; **36**: 1133–1137.
8. Tsaih S-W, Pezzolesi MG, Yuan R *et al.* Genetic analysis of albuminuria in aging mice and concordance with loci for human diabetic nephropathy found in a genome-wide association scan. *Kidney Int* 2010; **77**: 201–210.
9. Martin JE, Sheaff MT. Renal ageing. *J Pathol* 2007; **211**: 198–205.
10. Szatkiewicz JP, Beane GL, Ding Y *et al.* An imputed genotype resource for the laboratory mouse. *Mamm Genome* 2008; **19**: 199–208.
11. Pezzolesi MG, Poznik GD, Mychaleckyj JC *et al.* Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 2009; **58**: 1403–1410.
12. Kang HM, Zaitlen NA, Wade CM *et al.* Efficient control of population structure in model organism association mapping. *Genetics* 2008; **178**: 1709–1723.
13. Webb BT, McClay JL, Vargas-Irwin C *et al.* In silico whole genome association scan for murine prepulse inhibition. *PLoS One* 2009; **4**: e5246.

see original article on page 239

Anemia treatment in chronic kidney disease accompanied by diabetes mellitus or congestive heart failure

Steven Fishbane[1] and Nobuyuki Miyawaki[1]

Anemia is common in chronic kidney disease (CKD). The CHOIR study found increased risk of a composite cardiovascular outcome when anemia was treated with epoetin-alfa to a target hemoglobin level of 13.5 as compared with 11.3 g/dl. Whether this increase applies to all patient subgroups equally is unclear. We discuss an analysis by Szczech and colleagues of the effects of the higher hemoglobin target in CKD patients with diabetes mellitus or congestive heart failure.

Kidney International (2010) **77,** 175–177. doi:10.1038/ki.2009.455

In this issue of *Kidney International*, Szczech and colleagues[1] report on a *post hoc* analysis of the Correction of Hemoglobin and Outcomes in Renal Insufficiency (CHOIR) study.[2] The original study was notable for finding increased risk of a composite cardiovascular end point with epoetin-alfa targeted to a hemoglobin (Hgb) level of 13.5 as compared with 11.3 g/dl in patients with chronic kidney disease (CKD). When this is considered together with the CREATE (Cardiovascular Risk Reduction by Early Anemia Treatment with Epoetin Beta) study (CKD patients)[3] and the Normal Hematocrit Cardiac Trial (hemodialysis patients),[4] there was a strong trend in each study for increased risk for mortality with higher Hgb targets during erythropoiesis-stimulating agent (ESA) treatment. As a result, the National Kidney Foundation's Kidney Disease Outcomes Quality Initiative (KDOQI) anemia guidelines were revised in 2007 to include an evidence-based warning to avoid Hgb targets above 13 g/dl in CKD patients treated with ESAs. Little, however, has been learned about whether higher Hgb targets affect certain sub-populations differently. The current elegant analysis by Szczech *et al.*[1] examines the effect of Hgb target in the CHOIR study on two highly relevant sub-populations, CKD patients with diabetes mellitus and those with congestive heart failure (CHF).

[1]*Winthrop-University Hospital, Mineola, New York, USA*.
Correspondence: *Steven Fishbane, Winthrop-University Hospital, 200 Old Country Road, Suite 135, Mineola, New York 11501, USA.*
E-mail: sfishbane@metrorenal.com