

## Prospects for complex trait analysis in the mouse

Richard Mott · Jonathan Flint

Received: 1 April 2008 / Accepted: 25 April 2008 / Published online: 21 May 2008  
© The Author(s) 2008

We want to discuss how recent dramatic progress in whole-genome association analysis (WGA) applied to human case-control studies will affect complex trait analysis in the mouse. At first sight it might now seem that one can identify the genetic determinants of human complex disease directly. However, the picture is less straightforward, for although a significant number of genes have been identified and replicated across different human WGA studies, in most diseases the genetic variation segregating at these genes explains only a small fraction of cases that should be accounted for by genetic causes. The question remains whether the missing genetic signal is from common variants in other genes but with very small phenotypic effect, or is caused by rare variants in the same genes, or a combination of both. In either case, to continue making progress directly with the human WGA methodology it will be necessary to increase sample sizes significantly (Zeggini et al. 2008).

Where does this leave mouse complex genetics? Of course, as the readers of *Mammalian Genome* will appreciate, working with mice does have certain advantages in that it is possible to design and control experiments and take detailed measurements in a way that is impossible with humans. Nevertheless, the mouse complex genetics community must address three key issues in order to contribute to our understanding of human biology and disease.

The first of these is mapping resolution. We need to establish suitable mouse populations in which high-resolution mapping is the norm. The small extent of linkage disequilibrium (LD) in most human populations means that a positive signal in a human WGA is localized to a few tens

of kilobases, usually the span a single gene. Assuming that the functional genetic variant acts on the nearest gene (which is not always the case), then human WGA delivers single-gene resolution. In contrast, a detected QTL in an  $F_2$  intercross between two inbred laboratory mouse strains that explains 5% of the phenotypic variance will be mapped into a 95% confidence interval of approximately 30 Mb, containing approximately 300 genes. In some cases, combining information from multiple  $F_2$  crosses with the haplotype map of the mouse genome can refine QTL localization (Hitzemann et al. 2002; Li et al. 2005). Several other strategies have been proposed to solve this problem, by using populations of mice with more recombinants, and consequently steeper LD decay profiles. Heterogeneous stocks (HS) are descended from eight known inbred strain of mice that are outcrossed using a rotational breeding scheme for many generations until the genomes are relatively fine-grained mosaics of the founder haplotypes; mapping resolution of 2–3Mb is obtainable (Mott et al. 2000; Talbot et al. 1999; Valdar et al. 2006). However, despite being much more accurate than an  $F_2$  cross, this is still too crude for single-gene mapping for which we require mapping resolution of about 100 kb.

Commercial mouse breeders maintain very large genetically heterogeneous outbred populations, some of which are suitable for complex trait analysis. As proof of principle, Yalcin et al. (2004) showed how one such population, MF1, could be used to fine-map a QTL for behavior down to the gene *Rgs2*. Our group is now in the process of evaluating the genetic variability and LD profile of several commercially available outbred populations. Preliminary data suggest that there are some populations that have suitable properties but that others are not useful, having been recently rederived from a small number of animals and therefore containing extensive LD. There is

---

R. Mott (✉) · J. Flint  
Wellcome Trust Centre for Human Genetics, Roosevelt Drive,  
Oxford OX3 9JN, UK  
e-mail: Richard.Mott@well.ox.ac.uk

also one tantalizing study of outbred wild mice that suggests they have an LD structure similar to that of humans (Laurie et al. 2007).

The disadvantage of working with outbreds is that each animal is unique so that it is impossible to perform repeat experiments on the same genetic background, which may be necessary; for example, in a study to measure gene expression changes during development. Furthermore, the cost of high-density genotyping required limits the size of experiments. In contrast, inbred strains of mice permit these types of studies, need only be genotyped once, and there is a synergy in accumulating data from different experiments on standardized genetic backgrounds. There has been considerable debate over the direct use of the standard laboratory inbred strains for WGA. The main point of contention is that the number of independent inbred strains available is limited (for example, the mouse phenome database <http://www.phenome.jax.org/pub-cgi/phenome/mpdcgi> uses less than 40 priority strains, and even these strains share haplotypes to a considerable degree). Although the method may work for major QTLs explaining over 50% of the phenotypic variance, results are mixed for complex traits (Payseur and Place 2007) where QTLs of small effect are lost among false-positive signals in a genome scan. On the other hand, the sharp LD decay profile of the inbred strains is very attractive, so that if one can be sure that a QTL is segregating in a region (for example, from an  $F_2$  cross), then an analysis of inbred strains across the region may identify the gene in some cases. In addition, there is a considerable saving in genotyping costs.

This discussion suggests there is a strong case for designing and constructing a large population of inbred lines that contains a high density of recombinants and where the lines are independent, in the sense that they do not share any recombination events. This was the motivation behind the Collaborative Cross (CC) (Churchill et al. 2004; Threadgill et al. 2002). Currently, over 400 CC recombinant inbred lines are being bred at Oak Ridge National Laboratory, USA, and Tel Aviv University, Israel, in a collaboration funded by the U.S. Department of Energy (DOE), The Ellison Foundation, National Institutes of Health (NIH), and The Wellcome Trust. The lines are descended from eight genetically diverse founder strains (including three wild-derived strains) and are currently between generation 6 and 10 of inbreeding. They will be fully inbred in about 4 years but are already sufficiently advanced that a pilot project is planned to assess their use for QTL mapping. The CC is almost ideal, except that the expected QTL mapping resolution is about 1 Mb (Valdar et al. 2005)—not quite single gene.

The second issue is that the haplotype structure of the classical laboratory strains of mice is not ideal for complex

trait analysis. To make best use of the mouse, we need complete genome sequences of the common laboratory strains. Already we know from partial resequencing of 16 strains that the so-called classical strains share only a fraction of the genetic variation segregating in wild-derived strains (Frazer et al. 2007; Yang et al. 2007) so there should be fewer QTLs in a study solely using classical strains. Moreover, their genomes are not independent; the pattern of haplotype sharing is not random across the genome, causing the problem of false-positive QTLs alluded to above. By contrast, the haplotype structure of the CC is necessarily random, very rare variants should not exist, and the founding strains contain more genetic variation than is present in some human populations (Roberts et al. 2007). We do not yet know much about the haplotype structure or origins of commercial outbreds.

The third issue is how we should relate discoveries made in the mouse to human disease. It should be reiterated that many studies in mice are not feasible in humans, such as the elucidation of gene networks from most tissues and developmental stages. As it is extremely unlikely that identical causative polymorphisms will be segregating in both species, we should not necessarily expect the same genes to be identified in WGA, although there are examples where this is the case, such as cancer modifiers common to mice and humans (Ruivenkamp et al. 2002). Nevertheless, we should expect the same pathways to be implicated (Emilsson et al. 2008). However, at present the functional annotation of both species is incomplete. Therefore, alongside the development of suitable mapping populations, we need comprehensive annotation of the gene networks in the mouse, and how they vary during development, between tissues and between genetic backgrounds. There is not space here to say more except that it will require a concerted international effort, integrating and extending existing online resources.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD et al (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 36:1133–1137
- Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F et al (2008) Genetics of gene expression and its effect on disease. *Nature* 452:423–428
- Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA et al (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448:1050–1053
- Hitzemann R, Malmanger B, Cooper S, Coulombe S, Reed C et al (2002) Multiple cross mapping (MCM) markedly improves the

- localization of a QTL for ethanol-induced activation. *Genes Brain Behav* 1:214–222
- Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ et al (2007) Linkage disequilibrium in wild mice. *PLoS Genet* 3:e144
- Li R, Lyons MA, Wittenburg H, Paigen B, Churchill GA (2005) Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics* 169:1699–1709
- Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci USA* 97:12649–12654
- Payseur BA, Place M (2007) Prospects for association mapping in classical inbred mouse strains. *Genetics* 175:1999–2008
- Roberts A, Pardo-Manuel de Villena F, Wang W, McMillan L, Threadgill DW (2007) The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm Genome* 18:473–481
- Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C et al (2002) *Ptprj* is a candidate for the mouse colon-cancer susceptibility locus *Sccl* and is frequently deleted in human cancers. *Nat Genet* 31:295–300
- Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC et al (1999) High-resolution mapping of quantitative trait loci in outbred mice. *Nat Genet* 21:305–308
- Threadgill DW, Hunter KW, Williams RW (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome* 13:175–178
- Valdar W, Flint J, Mott R (2005) Simulating the collaborative cross: power of QTL detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172:1783–1797
- Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P et al (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 38:879–887
- Yalcin B, Willis-Owen SA, Fullerton J, Meesaq A, Deacon RM et al (2004) Genetic dissection of a behavioral quantitative trait locus shows that *Rgs2* modulates anxiety in mice. *Nat Genet* 36:1197–1202
- Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F (2007) On the subspecific origin of the laboratory mouse. *Nat Genet* 39:1100–1107
- Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL et al (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645