Meeting report

Can we find the genes involved in complex traits?

Mathew Pletcher and Tim Wiltshire

Address: Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA.

Correspondence: Tim Wiltshire. Email: timw@gnf.org

Published: 22 September 2004

Genome ***Biology*** 2004, **5:**347

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2004/5/10/347

A report on the third Complex Trait Consortium meeting, Bar Harbor, USA, 6-9 July 2004.

The Complex Trait Consortium (CTC) [http://www.complex-trait.org] came together for the first time two years ago in Memphis, Tennessee to formalize a loose association of scientists focused on using methods for mapping quantitative trait loci (QTLs) to unlock the genetics of complex phenotypes. The study of QTLs attempts to pinpoint the associations between the natural phenotypic and genotypic variation found in existing populations for the purpose of uncovering the mechanism of complex diseases. Since its first meeting, the CTC has grown, and it met, in July, at The Jackson Laboratory for its third annual meeting to evaluate the tools and resources that are available to advance the study of QTLs and to review the progress towards identification of quantitative trait genes (QTGs). The largest impact on mouse genetics, and not just QTL studies, in the short history of the CTC has come from the public accessibility of a nearly complete mouse-genome assembly. The importance of that achievement carried through the entire meeting from Rick Woychik's (The Jackson Laboratory, Bar Harbor, USA) introductory remarks to spirited concluding discussions about the appropriateness of current strategies to generate genome assemblies of additional strains of mice.

The presence of a mouse genome assembly provides the context with which to analyze the growing resource of single-nucleotide polymorphisms (SNPs) for the purposes of studying and defining the genomic ancestry of the common inbred strains. Using a SNP set of 1,513 markers in 60 strains of mice, Ken Paigen (The Jackson Laboratory) showed that large regions of the mouse genome, including a 40 megabase (Mb) region on the X chromosome, exist in linkage disequilibrium (LD). LD is a measure of the co-inheritance of markers, whereby the expected frequencies of haplotypes are not found within a population. By Paigen's estimate, over 50% of the genome of inbred mouse strains exists in LD blocks of more than 6 Mb, meaning that there are large segments of the genome that are selectively co-inherited in order to maintain complementary allelic combinations in the neighboring genes (that is, recombination in this region is selected against). One of us (T.W.) discussed his extension of the current SNP data to over 10,000 markers in 48 different strains. He also proposed the re-evaluation of the idea of *in silico* mapping, presenting work in which haplotype structure was inferred from SNP data and statistical associations were found between those inferred haplotypes and phenotypic distributions among the inbred strains.

Mark Daly (Whitehead Institute, Cambridge, USA) also demonstrated the utility of defining ancestral relationships on the basis of SNP patterns between strains, but as a means to follow-up initial QTL mapping rather than for defining new QTLs. By combining SNP data with information from multiple crosses that defined the QTL, and hypothesizing that the SNP pattern has to be different in the area of the QTG, the candidate region can be cut down to less than a megabase. This same methodology was used by a number of speakers as a means to narrow down candidate gene lists for their favorite QTLs. Daly also announced his group's plan to begin work on an Affymetrix-style 'SNP chip' that will be used to provide strain genotypes for 200,000 SNP markers. This announcement of the further extension of current SNP data was added to by Richard Mott (Wellcome Trust Centre for Human Genetics, Oxford, UK), who solicited mouse strains to be typed against a 15,000-SNP panel that will be used to genotype a 2,000-strain heterogeneous stock set, which is currently being used to investigate behavioral QTLs. Mott's solicitation led to the inclusion in his genotyping effort of all available recombinant inbred strains and inbred lines that make up the 'Priority List' of the Mouse Phenome Project [http://www.jax.org/phenome].

Combining genomic assembly data, genotyping data and microarray expression data has led to the rapid expansion of the study of expression QTLs, or eQTLs - variations in gene-expression level that are responsible for quantitative variation in phenotype. Ken Manly (University of Tennessee, Memphis, USA) presented his web-based tool, WebQTL [http://webqtl.org] and the QTL Reaper software package, for studying the associations between expression differences in genes and allelic patterns in recombinant inbred lines. Even considering the false-positive rates associated with this methodology, John Belknap (Oregon Health and Science University, Portland, USA) and Eric Schadt (Rosetta Inpharmatics, Kirkland, USA), demonstrated successful integration of eQTL analysis into their respective studies of alcohol preference and obesity phenotypes. Kent Hunter (National Cancer Institute, Bethesda, USA) showed how informative eQTLs can be even without an understanding of the underlying genetics that produced them. He confirmed in mice that the expression levels of particular genes correlate with, and can even predict, the susceptibility of different strains to metastatic tumors. If the susceptible expression pattern can be reset - for example, Hunter was able to accomplish this with administration of caffeine - to resemble resistant expression patterns, the mouse is no longer susceptible to tumorigenesis or metastasis.

Despite the initial success of integrating the mouse genome assembly into the process of QTL mapping, it is certainly not the final answer for identifying QTGs, and many other tools and concepts were discussed that might have just as large an effect. A recent grant solicitation (from the National Institute of Environmental Health Sciences) for the resequencing of 15 additional strains of mice beyond the current C57BL6/J assembly was welcomed as a means of further defining haplotype patterns. It was suggested that an approach focused on the coding regions of the genes could be most beneficial, ensuring greater overlap of sequences between more strains and providing a catalog of the SNPs likely to be most informative. Woychik described preliminary discussions for the creation of a public knockout mouse library. Conversation focused on whether such a library should be purchased from a commercial entity or regenerated publicly, and whether strains would be more useful if they were all conditional knockouts, although this resource is currently only at the contemplative stage with no funding. Plans for the 'collaborative cross', a 1,000 line recombinant inbred resource developed from a cross of eight different strains, were organized by Rob Williams (University of Tennessee Health Science Center, Memphis, USA) and Gary Churchill (The Jackson Laboratory), and some of the first breeding crosses might be initiated this year by Churchill. The goal of the collaborative cross is to be able to define QTLs of only a megabase in size, although the statistical models for analyzing the data from such a cross are still being developed. Mott introduced the idea that outbred mice could be used both to confirm and to refine mapping done in inbred strains, as well as to provide additional diversity for the identification of new QTLs. Work by Karl Jepsen (Mount Sinai School of Medicine, New York, USA) on bone fragility underscored the benefit of incorporating distinct disciplines, in this case engineering, in defining and understanding new phenotypes. He also pointed out the importance of understanding the relationships between closely related phenotypes, because their interdependence could mean that the characteristic assumed to be under investigation is actually secondary to a phenotype that is not of interest. Likewise, David Threadgill (University of North Carolina, Chapel Hill, USA) presented clear examples of how environmental contributions such as bacterial flora can drastically change the phenotypic readout of genetically identical mice.

In the end, the discussion of current and future resources, uses of the genome assemblies, different kinds of crosses and phenotyping all feed back to the idea of moving from QTL to QTG, and for that the most positive news to the field of QTLs and complex disease study came from Bev Paigen (The Jackson Laboratory). She has been able to progress to QTGs for no fewer than five loci by incorporating traditional genetic cross data with haplotype structure, human QTL synteny, and, in the case of the *Abca1* gene, expression data. She provided some of the best evidence that a critical mass of converging information and understanding has now tipped the scale and will, in the near future, allow more rapid progress to the identification of genes from loci of interest.