Review

Mammalian Genome
Incorporating Mouse Genome

Mammalian Genome 13, 175–178 (2002).
DOI: 10.1007/s00335-001-4001-Y

Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort

David W. Threadgill,[1] Kent W. Hunter,[2] Robert W. Williams[3]

[1]Department of Genetics, CB#7264, Lineberger Comprehensive Cancer Center, Rm 11-109, University of North Carolina, Chapel Hill, North Carolina 27599, USA
[2]Laboratory of Population Genetics, DCEG/NCI/NIH, Bldg 41, Rm 702, 41 Library Drive, Bethesda, Maryland 20892, USA
[3]Center for Genomics and Bioinformatics, University of Tennessee Health Science Center, 858 Madison Avenue, Rm 101A, Memphis, Tennessee 38163, USA

Received: 29 November 2001 / Accepted: 17 December 2001

During the previous century, isogenic lines of mice became one of the major resources responsible for advancing biomedical and genetic research. From humble beginnings in 1909, marked by derivation of the DBA inbred strain by Clarence Little, founder of The Jackson Laboratory, over 500 unique strains have been developed, many purposely derived as models of common human diseases (Beck et al. 2000). Many fundamental discoveries in the biological sciences have been attributed to the existence of inbred mice. Discoveries like the major histocompatibility complex by George Snell and monoclonal antibodies by George Kohler and Cesar Milstein are examples, both recognized with Nobel Prizes (Festing and Fisher 2000). With the development and refinement of powerful genetic tools including transgenesis (Gordon et al. 1980), gene targeting (Doetschman et al. 1987; Thomas and Capecchi 1987), and ENU mutagenesis (Russell et al. 1979; Bode 1984), the impressive power of inbred strains in functional genomics has been overshadowed, but not forgotten.

Ken Paigen, director of The Jackson Laboratory, reawakened community interest in the extraordinary genetic resources and phenotypic diversity archived in extant inbred strains (Paigen and Eppig 2000). As a result, over 50 strains are currently being screened as part of the Mouse Phenome Project (http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home) to rigorously define the remarkable range of phenotypes already available for genetic dissection. The trick in the next decade will be to partition all of the dramatic strain differences in hundreds of traits to sets of polymorphic genes. To do so in a concerted and effective way will require significantly more resources than are currently available.

A complex trait consortium

During the 15th International Mouse Genome Conference (IMGC) held in Edinburgh, Scotland on October 21–24, 2001, a group of 20 geneticists interested in complex trait analysis gathered to discuss potential community resources that could bring the concept of complex trait dissection and quantitative trait locus (QTL) gene identification much closer to reality. From this meeting, the idea of a Complex Trait Consortium (CTC) was developed. The first formal meeting of the consortium is scheduled for May 15–17, 2002 in Memphis, Tennessee (http://www.complextrait.org).

The goals of the CTC are to galvanize researchers interested in more complex biological questions and to develop an international infrastructure to facilitate rapid advances in the field similar to the recent resurgence witnessed for large-scale mutagenesis projects with the formation of the International Mouse Mutagenesis Consortium (Nadeau et al. 2001). These large international programs have, by any measure, been tremendously successful and will provide years of resources to assign function to individual genes (Hrabe de Angelis et al. 2000; Nolan et al. 2000). The value and sophistication of mutant production has been eloquently described and put into context with the more rudimentary status of classical QTL analysis (Nadeau and Frankel 2000). Although mutagenesis approaches are ideally suited to address many important biological questions, equally important questions exist that are not so easily solved by this approach. The more difficult problems revolve around efficient dissection and identification of genes regulating complex and quantitative traits. In light of the success that a community-wide effort brought mutagenesis programs (Nadeau et al. 2001), the time is right to reassess community efforts and resources available to tackle more complex issues of modern mammalian biology.

The need for new resources

Tracking down complex trait and QTL genes, gene variants that contribute to graded and usually subtle differences among humans and other organisms, is regarded as an exceptionally difficult, if not quixotic task. However, the solution is technically straightforward and, if appropriately supported, no more expensive than a moderately sized mutagenesis program. These efforts are well justified given the growing appreciation that variations in susceptibility to the most pervasive diseases, like cancer, diabetes, infections, and mental illness, are almost invariably modulated by a substantial number of gene variants that interact with other genetic and non-genetic factors in complex and not easily predictable ways.

The CTC group recognized that several valuable intermediate resources have been or are being developed. Among the more useful resources are various panels of consomic and congenic lines (Nadeau et al. 2000; Iakoubova et al. 2001). These panels will be excellent resources for partitioning individual complex and quantitative trait loci with sufficient strength and treating them as Mendelian loci. Besides incorporating a limited genetic diversity, these and other resources currently being developed are optimal primarily for individual gene function studies or the genetic dissection of simple genetic traits. Consequently, the current resources are but a small component necessary to efficiently pursue the many complex biological questions important to human health and disease prevention.

Correspondence to: D. Threadgill; E-mail: dwt@med.unc.edu

Table 1. Proposed resources to stimulate and enhance complex trait research.

Resource	Use	Cost[a]
Mus SNPs	Mapping, association studies	3–5M
Mus haplotypes	Association studies	0.3M
Second-generation RI panel	QTLs, epistasis, and gene-environment interactions	10–15M
Transcript and proteome maps	Selection of candidates genes	3–5M
BAC libraries	Candidate gene validation	1–2M
ES cell lines	Candidate gene validation	1 M
Mutagenized ES cell libraries	Candidate gene validation	1–2M

[a] Estimated multi-year total in million (M) US$.

Many important biological problems where the mouse could contribute substantially will not be solved with the current resources. For example, additional resources are needed to dissect epistatic interactions, to efficiently identify disease susceptibility genes, and to bring epidemiological findings to the gene level through gene-by-environment interactions. Other less obvious examples are the genetic dissection of traits that have reduced penetrance or low heritability, or traits with high stochastic variation. If the full power of mouse genetics is to be realized, additional reagents will be required. As such, several critically important community resources were identified at the CTC organizational meeting that would propel complex trait and QTL analysis forward (Table 1).

Mus germ-plasm haplotypes

Many inbred mouse strains share ancient haplotypes randomly spread across their genomes. Unlike simple repeats, single nucleotide polymorphisms (SNPs) are stable over many generations. Furthermore, many SNPs in extant inbred mouse strains can be traced back to a common ancestral population. This genealogical relatedness, along with the accumulation of recombinations over hundreds of generations, will facilitate complex trait and QTL mapping through haplotype association studies. Thus, one high-priority resource identified by the CTC is a database containing dense genome-wide haplotype maps of 50–100 extant inbred mouse strains. Participants of the CTC planning meeting agreed, based upon available polymorphism data, that 100 strains would harbor the majority of the extant *Mus* genus germ-plasm and be sufficient to survey the extent of *Mus* haplotype diversity. All agreed that this resource should be linked directly to the recently initiated Mouse Phenome Project. Needless to say, this resource is dependent upon an appropriately supported mouse SNP discovery program.

A major attribute of *Mus* haplotype maps will be the tremendous breadth of characterized germ-plasm available to support other studies. In addition to association studies, *Mus* haplotype maps will promote more informed selections of strains for modifier detection and mapping. Conceivably, the selection of a panel of appropriately chosen strains for test matings might allow localization of modifiers, or better yet, identification of candidates without genome-wide genotyping.

Multi-parental recombinant inbred panel

A second resource identified at the CTC organizational meeting is a second-generation recombinant inbred panel. The original RI lines were not developed with complex traits in mind (Bailey 1971). Rather, they were designed to efficiently localize Mendelian traits. With a 1,000-line advanced, multi-parental RI panel, single-gene resolution would be approached if each line archived 100 recombination sites. Such a purposely designed RI panel could serve as a new community mapping resource by combining diverse germ-plasms derived from a variety of inbred strains with the capture of more than 100,000 precisely defined recombination breakpoints.

The first step to produce such an RI panel would involve selecting four to eight strains to generate advanced multi-parental intercross progeny. The selection of the progenitor strains will require further community discussion. Once inbreeding of the advanced intercross progenitors begins, molecular genotyping, which did not exist when the first generation RIs were developed, could be used at each generation to assure maintenance of Hardy-Weinberg equilibrium of progenitor strains and to identify and preserve recombination sites in gene-rich regions. Likewise, selected genotyping could be used to speed the final stages of inbreeding. It is anticipated that the RI inbreeding efforts would be an international effort composed of five to ten core institutions, each being responsible for the production of 100 or more lines.

The majority of a large RI set could be cryopreserved after initial genotyping. A core set of 50–100 lines could be kept available at The Jackson Laboratory or other mouse resource center at any one time. This core set could be used for first-stage, coarse mapping much like the two-tiered approach developed to create the high-density C57BL/6 X *m. spretus* backcross genetic map by The European Collaborative Interspecific Backcross Mapping Consortium (Rhodes et al. 1998). These core RI lines could be transformed by intercrossing into almost 5,000 unique recombinant inbred intercross (RIX) genomes (Fig. 1); RIX are F_1 hybrids between pairs of parental RI lines and have the genetic complexity of advanced intercrosses, but can be reproduced at will from the parental RI lines. RIXs provide more reliable trait means (lower coefficients of error) than RIs because of their heterogeneous genome structure (Threadgill et al. submitted). This also makes them a better model for complex human genomes. The frozen bank of resource RIs, containing tens of thousands of archived recombination sites, could be tapped for second-stage mapping. Since the resource RIs would already be genotyped at high resolution, additional RIs containing recombination breakpoints in the target QTL region could be identified and used to fine-map, at subcentiMorgan resolution, putative loci detected in the core strains. This approach is ideally suited to the inevitable shift from simple "one-gene-at-a-time" functional genomics to "complex network" functional genomics.

Outlook and additional resources

With dense genotypes of 100 diverse inbred strains and a well-structured 1,000-member RI panel, complex and quantitative trait analysis would be transformed profoundly. Using either association studies or linkage analysis, investigators would be able to fine-map loci controlling most heritable complex traits and to test multi-locus epistatic interactions. Uncovering gene/QTL pleiotropy would be feasible for the many traits acquired with sophisticated methods. Furthermore, it will finally become practical to study gene–environment interactions, bringing epidemiological observations to the gene level. With such a resource, it is even conceivable to perform combinatorial genetics by combining genetic mapping, environmental exposures, and transcript and proteome profiling methods, treating each element as a quantitative variable for genetical genomic studies (Jansen and Nap 2001).

Additional resources that would aid both mutagenesis programs and complex and quantitative trait approaches can be envisioned (Table 1). Strain- and tissue-specific transcript and proteome profiles linked to the genetic map would be a valuable start. Others include community access to validated embryonic stem (ES) cell lines from a wide range of strains to engineer background-specific mutations for candidate gene validation; BAC libraries from an equally diverse set of inbred strains for efficient allele transfer via BAC transgenesis; and libraries of in vitro mutagenized ES cells from a variety of genetic backgrounds for ef-

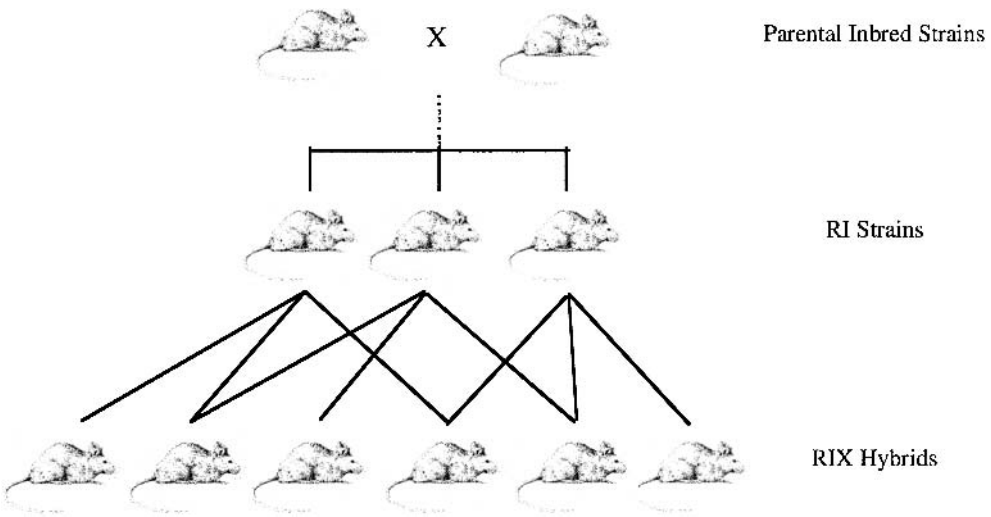

Fig. 1. Diagrammatic representation of RIX lines derived by serially intercrossing parental RI's.

ficient generation of allelic series (Chen et al. 2000). A community discussion on BAC libraries from additional strains is being organized by David Beier (Brigham and Women's Hospital, Boston) and Peter de Jong (Children's Hospital Oakland Research Institute). These and additional needs will be discussed at the planned Memphis CTC meeting in May, 2002.

To establish these public complex and quantitative trait mapping resources will require a number of core infrastructures including genotyping, mouse husbandry, bioinformatics, and statistics cores. Although the development of these core infrastructures will require significant effort and expense, it may be possible to share much of the infrastructure developed for the large mutagenesis centers. Regardless, the CTC will profit enormously from the skills and experience developed by international mutagenesis programs.

Infrastructure cost

What would such an effort cost? Surprisingly, the cost is quite modest when compared with current expenditures for mutagenesis programs. Systematic community efforts would initially require a relatively modest level of support (Table 1). Excluding long overdue resources to identify and catalog large numbers of SNPs in mice, a panel of 100 diverse inbred strains could be genotyped at densities sufficient to permit association studies within a year for a few hundred thousand US dollars. The estimated cost for production of 1,000 new advanced intercross lines would be around US$1.5–2.0 million per year during the initial development stages, sufficient to support 5,000 breeding cages of mice. These estimates include costs associated with required genotyping, bioinformatic, and statistical support cores. These costs are dwarfed by the size of human population studies that have far lower chances of actually identifying genes controlling complex and quantitative traits without strong candidates—candidates that could be provided by expanded studies in mice.

As a comparison, the cost of international mutagenesis programs is now well above US$20 million a year. National Institutes of Health support alone is in excess of US$15 million per year. Although results of the various mutagenesis programs are increasingly impressive, they will eventually reach a level of diminishing returns. An important distinction is that mutagenesis programs are additive in their effect. Each new mutation or phenotype requires independent characterization and mapping. In contrast, a program to provide optimal infrastructure for complex and quantitative trait analyses would be multiplicative in return. Since shared, community-wide complex trait resources will use the same diverse inbred strain set or advanced RI panel, each new analysis can immediately be compared with and contrasted to all previous analyses, opening the door for meta- and multifactorial analyses.

Similar to the less well adapted guest investigator programs at the large mutagenesis centers, well planned community resources consisting of a panel of diverse inbred strains and multi-parental RIs for two-stage mapping would be ideally suited for guest investigators. A guest investigator program would dramatically reduce overhead costs of maintaining large mouse populations at many institutional sites. Furthermore, unlike the reality at a mutagenesis center where guest investigators screen a temporally small window of putative mutants arising while they are in residence, guest investigators phenotyping one of the complex trait resources will have also cloned the underlying genes by virtue of the pre-mapped status of the resources. Likewise, results from guest researcher studies could immediately be compared across all previous studies for epistatic or multi-factorial analyses.

Future of mouse genetics

Mouse geneticists like to keep things simple, but the reality that Allan Balmain (UCSF) presented so well during an IMGC plenary talk is that simple genetics is an oxymoron that will not be sustainable much longer. We really need the right kinds of complex but genetically defined mice to work with in the very near future. The current resources are an acceptable community stop-gap, but they were never intended for efficient complex and quantitative trait analysis and are obviously far from ideal in their present incarnations.

Rather than adapting existing resources, well-planned and specifically designed genetic resources are needed for complex and quantitative trait analyses. These will undoubtedly have a profound impact upon our ability to close the gap that has arisen with Mendelian trait analysis by permitting efficient identification and ultimately validation of genes underlying complex traits. Because the experimental reproducibility afforded by inbred strains permits measurements of even small variations in quantitative traits, the inbred strain, and derivatives thereof, will provide the foundation for future complex and quantitative trait dissection, fulfilling the promise originally espoused by luminaries like Clarence Little almost a century ago.

Francis Collins (NIH) wrote a review for *Nature Genetics* in 1995 entitled "Positional cloning moves from perditional to traditional" (Collins 1995). A witty title once you get beyond thinking perditional is a typographic error. If complex and quantitative trait

mappers can just get their forepaws on the right types of mice, then by 2007 that title could easily read "QTL gene discovery moves from perditional to traditional."

References

Bailey DW (1971) Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation 11, 325–327

Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT et al. (2000) Genealogies of mouse inbred strains. Nat Genet 24, 23–25

Bode VC (1984) Ethylnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the T region of mouse chromosome 17. Genetics 108, 457–470

Chen Y, Yee D, Dains K, Chatterjee A, Cavalcoli J et al. (2000) Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. Nat Genet 24, 314–317

Collins FS (1995) Positional cloning moves from perditional to traditional. Nat Genet 9, 347–350

Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW et al. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature 330, 576–578

Festing MFW, Fisher EMC (2000) Mighty mice. Nature 404, 815

Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci USA 77, 7380–7384

Hrabe de Angelis MH, Flaswinkel H, Fuchs H, Rathkolb B, Soewarto D et al. (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25, 444–447

Iakoubova OA, Olsson CL, Dains KM, Ross DA, Andalibi A et al. (2001) Genome-tagged mice (GTM): two sets of genome-wide congenic strains. Genomics 74, 89–104

Jansen RC, Nap J-P (2001) Genetical genomics: the added value from segregation. Trends Genet 17, 388–391

Nadeau JH, Frankel WN (2000) The roads from phenotypic variation to gene discovery: mutagenesis versus QTLS. Nat Genet 25, 381–384

Nadeau JH, Singer JB, Matin A, Lander ES (2000) Analysing complex genetic traits with chromosome substitution strains. Nat Genet 24, 221–225

Nadeau JH, Balling R, Barsh G, Beier D, Brown SDM et al. (2001) Functional annotation of mouse genome sequences. Science 291, 1251–1255

Nolan PM, Peters J, Strivens M, Rogers D, Hagan J et al. (2000) A systematics, genome-wide, phenotypic-driven mutagenesis programme for gene function studies in the mouse. Nat Genet 25, 440–443

Paigen K, Eppig JT (2000) A mouse phenome project. Mamm Genome 11, 715–717

Rhodes M, Straw R, Fernando S, Evans A, Lacey T et al. (1998) A high-resolution microsatellite map of the mouse genome. Genome Res 8, 531–542

Russell LB, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC et al. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Nat Acad Sci USA 76, 5818–5819

Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mosue embryo-derived stem cells. Cell 51, 503–512