



ANNUAL REVIEWS **Further**

Click [here](#) for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

The Future of Mouse QTL Mapping to Diagnose Disease in Mice in the Age of Whole-Genome Association Studies*

Kent W. Hunter and Nigel P.S. Crawford

Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892; email: hunterk@mail.nih.gov

Annu. Rev. Genet. 2008. 42:131–41

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi:
10.1146/annurev.genet.42.110807.091659

Copyright © 2008 by Annual Reviews.
All rights reserved

0066-4197/08/1201-0131\$20.00

*The U.S. Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

Key Words

QTLs, whole-genome association, mouse models

Abstract

Genome-wide association analysis is emerging as a powerful tool to define novel genes and molecular pathways involved in susceptibility to human complex disorders. However, in spite of recent successes, this approach is not without its limitations, the most notable of which is inconsistent phenotype penetrance due to varied environmental exposures. Mouse models do, however, circumvent some of these drawbacks by allowing for a much higher degree of control over genetic variation and environmental exposure, and although their application to human complex genetics is not always straightforward, they do serve as a powerful means of complementing observations in human populations. Mouse quantitative trait locus mapping has proven a successful, yet technically demanding method for defining trait susceptibility. In this review, we focus upon recent advances that are both reducing the technical burden traditionally associated with quantitative trait locus mapping, and enhancing the applicability of these approaches to human disease.

SNP: single nucleotide polymorphism

Quantitative trait locus (QTL): a stretch of genomic DNA closely linked to the genes that underlie the trait in question. These may be considered analogous to “susceptibility loci” in the case of a disease phenotype

Population substructure: a collection of genetically differentiated subpopulations

INTRODUCTION

Advances in genomic technologies have heralded a new era in genetic analysis of common diseases. Most recently this has been demonstrated by the advent of microarray-based platforms that allow researchers to genotype hundreds of thousands of single nucleotide polymorphisms (SNPs) within a single individual. This capability has enabled a number of groups to perform whole-genome association studies that have identified SNPs associated with a variety of common diseases such as breast cancer (13), prostate cancer (63), colorectal cancer (64), type 1 and 2 diabetes (20, 44, 47, 59), Crohn disease (42, 59), celiac disease (51), and restless leg syndrome (60). As a result, our understanding of the underlying genetic variants that play a role in establishing susceptibility for many common clinical disorders has been significantly increased. Furthermore, the identification of novel genes and/or molecular pathways involved in disease susceptibility has opened additional avenues of research that continue to give us greater insight into the pathological basis of complex disorders.

As further advances in genotyping and sequencing reduce the costs of these studies, many more exciting results will undoubtedly follow. Much focus has thus deservedly been placed on these studies. However, like all research strategies and methods, whole-genome association analyses suffer from intrinsic limitations and cannot give us a complete insight into the complex genetic and environmental interactions associated with common disease phenotypes. No single strategy has ever been able to completely encompass all the components and factors necessary to provide a comprehensive understanding of disease etiology, particularly not in the enormously complex arena of human polymorphism, population dynamics, and environmental exposures. The continuing development of mouse models, tools, resources, and methodologies should therefore remain a prominent parallel strategy by which to investigate the genetic and environmental components involved in human disease susceptibility. In this review

we briefly discuss the advances in quantitative trait locus (QTL) analysis over the past decade, primarily by using mouse cancer models as an example. We then attempt to illustrate how mouse QTL and whole-genome association analyses are actually complementary approaches that can serve to enhance our understanding of the genetic basis of human disease.

WHOLE-GENOME ASSOCIATION ANALYSIS

With the current flood of whole-genome association papers demonstrating the power of this approach, a very relevant and pertinent question is, why continue the development of mouse susceptibility genetics? The answer to this question essentially lies in consideration of those inherent limitations of whole-genome association analysis that prevent capture of all relevant biological information. For example, it is known that cancer susceptibility is the result of an interaction between constitutional genetic factors and varying environmental exposures. However, the effects of genetic variants may only be detectable in those individuals with the appropriate environmental exposures. For instance, substantial environmental contributions to breast cancer susceptibility appear to result from dietary (32) and reproductive factors (5), alcohol consumption (45), as well as other exposures. Inability to account for such variables in whole-genome association studies may reduce the power of analyses or obscure the ability to identify true causative susceptibility loci (2).

A further complication is that of population substructure. Differing levels of relatedness among individuals or undetected subpopulations are potent sources of both false-negative and false-positive outcomes in whole-genome association analyses (48). Furthermore, this issue will likely become increasingly important in future studies given that larger sample sizes, assembled from multiple different populations, will be required to detect the substantially more modest effects of those loci not captured in first-generation association studies. In addition, replication of the initial results will

become increasingly more challenging if successively larger sample sizes are needed to detect significant associations.

In addition, although a number of associations have been detected for genes with the most potent phenotypic effects (13, 24, 33, 37, 54, 61), most disease susceptibility is probably not due to a collection of genes with additive effects. Rather there are likely a number of genes that modulate susceptibility primarily through nonadditive interactions. These interactions can be multiplicative or conditional interactions, where the main effect of one gene is dependent on the presence of a particular allele of a second locus. Identification of these causative polymorphisms poses a significant problem for a variety of reasons. First, while a single SNP requires the analysis of 3 genotype classes, analysis of 2 SNPs requires the characterization of 9 different 2-locus combinations, 3 SNPs 27 genotype classes, and so on (34). This “curse of dimensionality” (4) may rapidly exceed the number of individuals who can be recruited into epidemiological studies to provide sufficient statistical power to detect associations (34). Furthermore, if SNPs do not have effects independent of interactions, then analyzing hundreds of thousands of SNPs in pairwise or higher-order fashion will undoubtedly exceed currently available computational capacity (34).

MOUSE MODELS OF COMPLEX DISEASE: A COMPLEMENTARY APPROACH

Mouse models, despite their limitations, provide a means of circumventing at least some of the limitations of human whole-genome association studies. There is growing evidence in the literature demonstrating that genetic susceptibility plays a very similar role in complex disease pathogenesis in both humans and in mouse models in spite of interspecies differences. An example of a complex trait mapping to the same genes or orthologous regions in the mouse and human genome includes plasma concentrations of high-density lipoprotein cholesterol, an im-

portant determinant of atherosclerotic vascular disease (57). In this case, the mouse and human experiments have provided complementary clues for each system during the analysis (46), or mouse QTLs have been demonstrated to translate into human disease susceptibility loci (8, 9, 14, 15). These successes suggest that the development of mouse complex trait analysis should be continued in parallel with the further investigation of human complex disease susceptibility. Indeed, utilization of these complementary approaches may well facilitate the development of novel methodologies and resources to help unravel the extremely complex interwoven factors affecting disease etiology.

Technological Advances Aid QTL Identification

What is the current status of modeling genetic susceptibility in the mouse? The difficulties reported to be associated with QTL mapping in the mouse have caused many investigators to explore alternative strategies for candidate gene identification (36). Specifically, the lack of mapping resolution and the difficulty of identifying genes, much less appropriate candidate genes, significantly impaired the ability of complex trait genetics to reach the “end game,” i.e., the identification of specific susceptibility gene(s) for any given trait. Fortunately, technical advances in the interim have significantly alleviated many of these difficulties. The completion of the human (30, 52) and mouse (58) genome sequencing projects, the preeminent technical advances, made it possible for the first time to identify all genes within a given region of interest. This in turn has allowed for the targeted development of reagents to analyze specific candidate genes without the laborious and time-consuming genomic screening previously required to simply identify potential transcripts.

Our ability to identify and characterize candidate regions has continued to expand dramatically as additional species have been sequenced (for examples, see 6, 27, 31). This has permitted identification of evolutionarily conserved sequence domains, and more recently has

Complex trait: a trait that develops through interaction of a number of genes with environmental factors. This type of inheritance is common in human disease and is referred to as polygenic inheritance

Identical by descent:

the finding that two alleles in a locus are identical because of descent from the common ancestor

Recombinant congenics:

an inbred mouse strain used for genetic mapping experiments, produced by mating two genetically distinct inbred strains, backcrossing their descendants to one of the parental strains for two generations or more, and then initiating sister x brother inbreeding without selecting for particular markers or phenotypes

Advanced intercross:

mouse lines produced by repeated intercrossing over ten or more generations. Resulting mice are heterozygous

enabled direct visualization of SNPs among many of the commonly used inbred mouse strains by the use of chip-based sequencing and large-scale polymorphism screening (17). Furthermore, sequence availability across species has permitted further acceleration of candidate gene identification for those traits that have been mapped in multiple species. Identification of orthologous chromosomal segments and their breakpoints in genetically defined candidate regions can help refine QTL localization and candidate gene lists by limiting searches to those loci shared between the two species (12). The most successful application of this approach to date has been the identification of a number of loci associated with cholesterol metabolism (56). A further advantage of the flood of sequence information is the ability to analyze haplotype structure, which was originally developed in human genetic mapping (53). Knowledge of those segments of the genome that are identical by descent (i.e., have the same haplotype structure) between mouse strains is a valuable method of refining regions of interest. Use of partially inbred mouse strains (14) or multiple genetic mapping studies with different inbred strain combinations can rapidly reduce the number of candidate genes that need to be screened for effects (14, 38, 55).

In addition to sequence information, other technologies have also significantly contributed to the ability to sort through QTL candidate lists. One of the most commonly used is genome-wide expression arrays. The observation that most SNPs lie outside of coding regions has led many to hypothesize that most QTLs are likely due to subtle changes in gene expression rather than to missense or nonsense mutations. This suggestion has subsequently led to the screening of genetically defined candidate intervals for genes that display differential expression between the strains of interest (3). This approach, originally designated genetical-genomics (28), provides an unbiased method for rapidly screening through hundreds of potential candidate genes simultaneously to reduce the list for further analysis to a manageable number (12, 22).

Advances in computing technology have enhanced our understanding of haplotype structure both in humans and model organisms, and the combination of human epidemiological association studies and haplotype analysis can also be a powerful tool for the screening, identification, and characterization of candidate QTLs. Reduction of candidate regions to relatively small intervals, followed by screening of potential genes using haplotype-tagged SNPs in appropriate human cohorts, can provide direct evidence of potential pathological roles in human disease and also give important clues into the identity of candidates in mouse systems (8, 9, 14). When appropriate human cohorts exist, this strategy can be a very efficient and rapid method for screening genomic regions, requiring significantly less time and effort than development of new in vitro systems or the generation of specialized mouse models.

Novel QTL Mapping Strategies Enhance Our Understanding of Human Disease

Ultimately, the efficiency of QTL modeling in the mouse is dependent upon good genetic mapping resolution. This has been, and continues to be, the major roadblock to efficient identification of novel modifier genes. Like other forms of complex trait analysis, recent innovations are beginning to alleviate some of the more problematic aspects of this area. Standard backcross and intercross genetic mapping suffers from the inability to refine candidate intervals much below half-chromosome resolution. To address this shortcoming, new genetic mapping strategies have been developed. Recombinant congenics (11) and advanced intercross panels (10), for example, reduce the amount of donor genome interrogated in each member of these specialized mapping panels, thus increasing resolution (see **Figure 1**). The trade-off is an increased amount of breeding over standard mapping strategies. An alternative strategy is to use multiple mapping crosses with different combination of strains to identify common haplotype blocks that segregate within

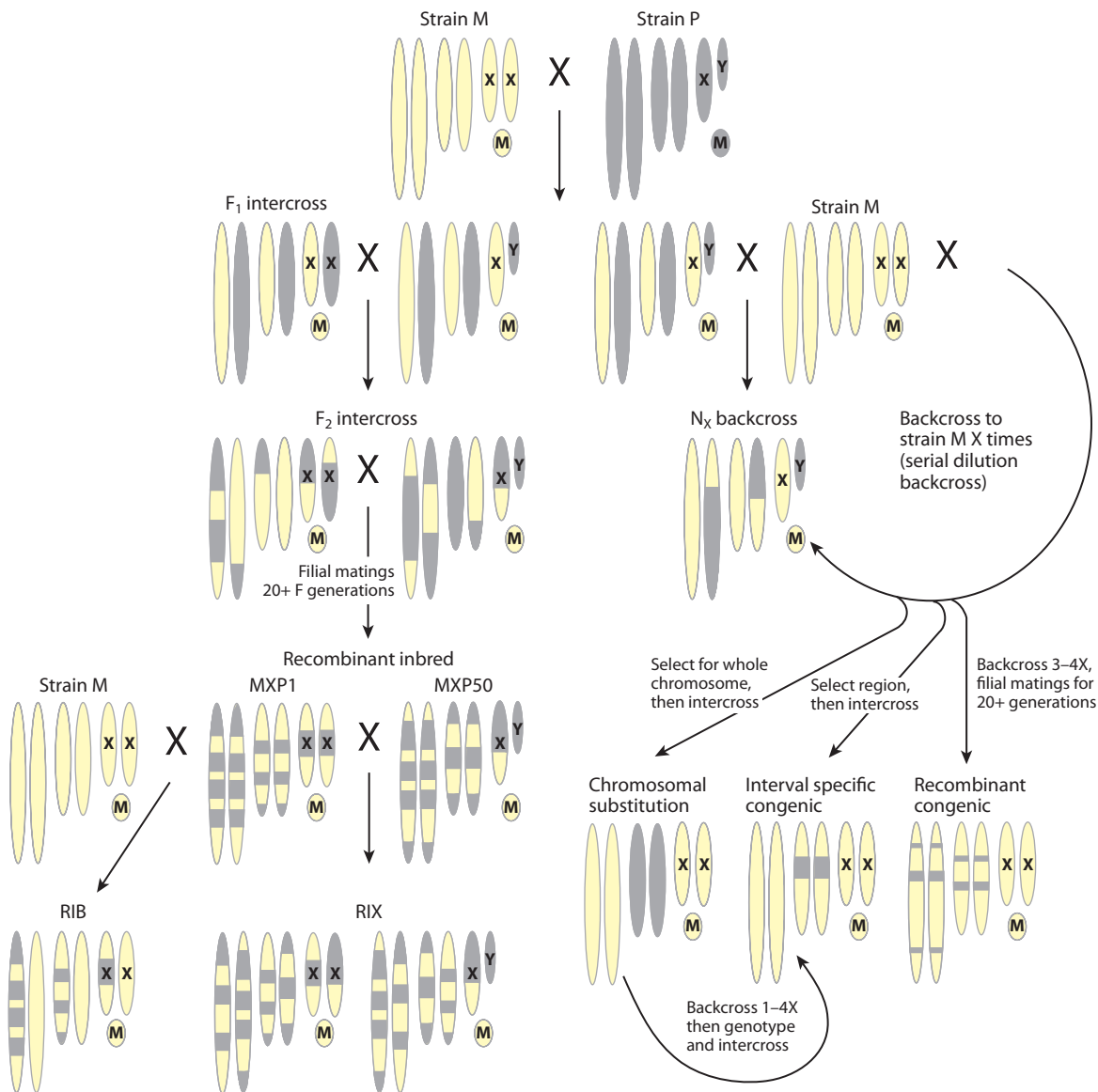


Figure 1

Schematic representation of experimental mapping cross strategies. Chromosomes are represented by ovals. The left and middle pairs represent autosomes, the right-hand pair the sex chromosomes. Mitochondrial DNA and cytoplasmic factors are represented by the small circle labeled M. Chromosomes and chromosomal segments originating from the parental donor strain (P) are gray, while the maternal genomic elements are light yellow. Figure adapted from Hunter & Williams (26).

the genetically defined candidate regions to narrow down the search for genes of interest (22, 23, 38, 55).

Initial higher mapping resolutions can be achieved by use of heterogeneous stock mice

(35). These mice are generated from the random breeding of progeny originating from typically four to eight inbred strains. Random segregation over many generations of breeding results in large numbers of meiotic crossovers,

which is essentially the genomic equivalent of what is observed in outbred human populations. The greater frequency of meiotic crossovers therefore necessitates high-resolution genotyping to capture the information of interest, which in turn allows traits of interest to be mapped at a high resolution. The downside of this strategy is again increased breeding requirements, as well as much more extensive genotyping.

A major shortcoming of most genetic mapping strategies to date is the fact that each animal produced during the meiotic mapping process is genetically unique. However, biological replicates provide an important source of statistical power for identification of candidate genes in most experimental strategies in systems biology. In addition, to date, unique mapping populations have had to be developed for each investigator's phenotype of interest. This has precluded the ability to investigate, at least with any degree of ease, how similar genetic variation influences what on the face may seem to be unrelated phenotypes of interest. However, as our computational abilities expand to enable better modeling of the complex interwoven networks that underlie all biological processes, it will be a significant advantage to directly assay more and more complex traits on the same genetic backgrounds to determine how genetic variation affects the entire organism, as well as to provide a virtually unlimited amount of tissues for future experimental investigation.

The Collaborative Cross—A Powerful New Tool for Complex Trait Analysis

To combine the high-resolution mapping capacity of heterogeneous stock animals with the “immortal” and easily replicable properties of inbred strains, an effort has recently been undertaken to develop a new mapping resource that can be utilized by many different laboratories as a genetic reference panel. Known as the Collaborative Cross, the project aims to develop a large panel of new recombinant inbred lines derived from eight different inbred mouse strains (7, 49, 50). Using a defined breed-

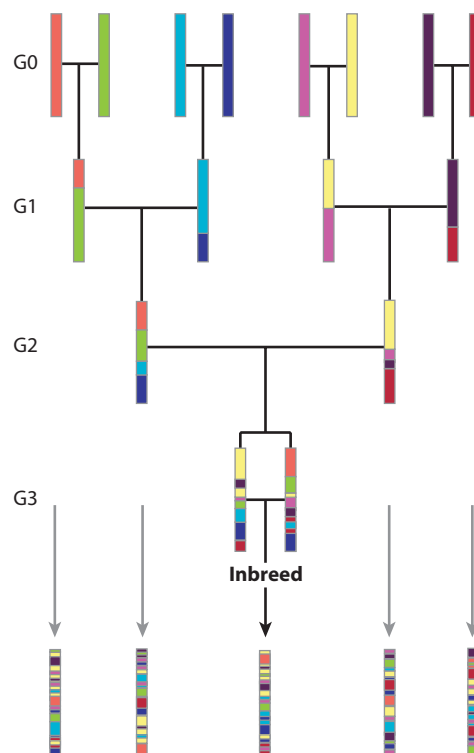


Figure 2

Simplified schematic representation of the Collaborative Cross breeding strategy. An example of a breeding funnel is shown, with each of the eight progenitor inbred strains represented by a different color. The genomes of all eight progenitors are mixed by controlled breeding in the G0–G3 generations, followed by inbreeding. Multiple independent funnels are breeding simultaneously, with different independent permutations of the starting positions of the inbred progenitors, resulting in production of a series of genetically complex recombinant inbred strains (other funnels represented by light gray arrows).

ing scheme and high-density genotyping, the phase and origin of each genomic segment will be known (**Figure 2**). Sufficient meiotic recombination will be incorporated across the panel to enable much higher resolution mapping of quantitative traits. Since the RI panel will be completely inbred, it will be possible to generate virtually unlimited amounts of material for current, and future, in vitro, in vivo, and genomic analytical methods. Furthermore, as

investigators analyze new phenotypes on the panel, it will be possible to start understanding the complex networks of interacting phenotypes by directly comparing the data accumulated on the same set of genetically segregating genotypes.

The Collaborative Cross will also help address another significant shortcoming of using the mouse to model human populations, that of genetic diversity. Due to the relatively low number of progenitors for the common laboratory inbred strains, large regions of the genome are identical by descent (IBD). For pairwise comparisons between any two common laboratory strains, IBD has been estimated to vary between 36% to 62% percent (62). As a result, large genetic “holes” exist that preclude the ability to screen significant portions of the genome. In addition, because of the number of strains included in the Collaborative Cross, including members of the three major mouse subspecies, almost 90% of the genetic variation present in the mouse genome will be available for screening, with similar minor allele frequency as observed in the human population (43).

The Practical Application of Complex Trait Analysis in Mouse Models to Human Disease

The evolution of complex trait strategies is enabling greater and faster analysis and identification of complex traits with higher intricacy. Mouse models permit the inclusion of genetically sensitized strains to address specific, focused but previously unexplored questions. For example, by using the highly metastatic polyoma middle T (PyMT) antigen transgene model, it was demonstrated that mammary tumor metastasis, which was thought to be the result of somatic evolution (16), actually has a significant inherited susceptibility component (25, 39). Further studies have demonstrated that these results translate into the human population. Pilot molecular epidemiological association studies of the human orthologs of the metastasis modifiers *Sipa1* (39) and *Rrp1b* (8) have shown statistically significant associations

(9). In addition, these results have prompted re-examination of clinical data in a descriptive epidemiology study that discovered an increased risk of metastatic disease in the first-degree relatives of probands who had died within five years of diagnosis (21).

These relatively straightforward complex trait analyses have recently been further extended by the investigation of additional factors thought to significantly contribute to human cancer progression, diet and obesity. Introduction of an obesity-sensitive strain, M16i (1), as the mapping partner for the PyMT animal, and a high-fat diet, has enabled direct experimental evaluation of these known factors for human breast cancer progression (29). These experiments demonstrated that, as expected, high-fat diets lead to a more malignant tumor phenotype (18, 19). In addition, a number of associated genetic loci were identified, providing the initial dissection of the underlying genetic contributions. Furthermore, current complex trait strategies are enabling investigation of not only gene-environment interactions, but also of epigenetic contributions to phenotypes. For example, utilization of the compound *Trp53-Nf1* knockout mouse model demonstrated that not only did both glioblastoma (40) and astrocytoma (41) have inherited components, but significant epigenetic events might also play an important role in disease pathogenesis (41).

CONCLUSION: FUTURE PROBLEMS

In spite of the clinical importance of the examples discussed above, it is difficult to envision an easy and efficient analysis of such complex interactions in human populations. First, it is virtually impossible to control important covariates such as environmental exposure in human populations. Second, to address the types of questions posed by these types of complex interactions, it becomes necessary to stratify patients into increasingly smaller and more homogenous subgroups. Association studies performed in such undersized subcohorts subsequently suffer from a substantial reduction in

PyMT: polyoma middle T

Metastasis: dissemination of tumor cells from a primary cancer to a distant site and the establishment of a distinct secondary tumor. Spread of cancerous cells typically occurs via the blood stream, lymphatic vessels, or through a body cavity

statistical power to the point where significant numbers of false negatives might rob investigators of a more comprehensive understanding of the disease phenotypes and molecular pathways that underpin them. Mouse models, however, provide a parallel platform for direct experimentation to confirm the results of human epidemiology and association studies. Furthermore, we will probably see an increasing sophistication of mouse models in the very near future. This will likely lead not only to the development of clinically relevant disease models but also to increasingly informative genetic anal-

yses that augment observations from human-based studies. In addition, and more important, it provides a vehicle in which to identify key genes and molecular pathways that may evade capture by current capabilities in human whole-genome association studies. The continued integration of these approaches, with each species informing the biology of the other, will result in a richer comprehension of the diverse and immense complexity of cancer and other disease phenotypes, which in turn will likely lead to significant improvements in clinical management of human disease.

SUMMARY POINTS

1. Whole-genome association analysis is a powerful means of elucidating novel susceptibility loci for complex human disorders.
2. Whole-genome association studies are not, however, without their limitations. These inherent constraints reduce the power of this type of analysis to detect novel susceptibility loci.
3. Mouse models of human disease provide a particularly useful complement to whole-genome association analysis. Quantitative trait locus analysis is an increasingly powerful tool in this respect, and recent technological innovations in this area have improved their applicability to human disease.
4. Novel QTL mapping strategies like the Collaborative Cross are enhancing our understanding of human disease.
5. A number of recent studies have illustrated the relevance of complex trait analysis in mouse models to human disease.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Allan MF, Eisen EJ, Pomp D. 2004. The M16 mouse: an outbred animal model of early onset polygenic obesity and diabetes. *Obes. Res.* 12:1397–407
2. Ambrosone CB. 2007. The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer. *Breast Cancer Res.* 9:114
3. Arbilly M, Pisante A, Devor M, Darvasi A. 2006. An integrative approach for the identification of quantitative trait loci. *Anim. Genet.* 37 (Suppl.) 1:7–9
4. Belman RE. 1961. *Adaptive Control Processes*. Princeton, NJ: Princeton Univ. Press
5. Britt K, Ashworth A, Smalley M. 2007. Pregnancy and the risk of breast cancer. *Endocr. Relat. Cancer* 14:907–33

6. Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87
7. 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.* 36:1133–37
8. Crawford NP, Qian X, Ziogas A, Papageorge AG, Boersma BJ, et al. 2007. Rrp1b, a new candidate susceptibility gene for breast cancer progression and metastasis. *PLoS Genet.* 3:e214
9. Crawford NP, Ziogas A, Peel DJ, Hess J, Anton-Culver H, Hunter KW. 2006. Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. *Breast Cancer Res.* 8:R16
10. Darvasi A, Solter M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141:1199–207
11. Demant P, Hart AA. 1986. Recombinant congenic strains—a new tool for analyzing genetic traits determined by more than one gene. *Immunogenetics* 24:416–22
12. DiPetrillo K, Wang X, Stylianou IM, Paigen B. 2005. Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet.* 21:683–92
13. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–93
14. Ewart-Toland A, Briassoulis P, de Koning JP, Mao JH, Yuan J, et al. 2003. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat. Genet.* 34:403–12
15. Ewart-Toland A, Dai Q, Gao YT, Nagase H, Dunlop MG, et al. 2005. Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* 26:1368–73
16. Fidler IJ, Kripke ML. 1977. Metastasis results from preexisting variant cells within a malignant tumor. *Science* 197:893–95
17. Frazer KA, Esquin 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* 448:1050–53
18. Gordon R, Hunter KW, Sorensen P, Pomp D. 2008. Genotype x diet interactions in mice predisposed to mammary cancer: I. body weight and fat. *Mamm. Genome*. In press
19. Gordon RR, Hunter KW, La Merrill M, Sorensen P, Threadgill DW, Pomp D. 2008. Genotype x diet interactions in mice predisposed to mammary cancer: II. tumors and metastasis. *Mamm. Genome*. In press
20. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, et al. 2007. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 448:591–94
21. Hartman M, Lindstrom L, Dickman PW, Adami HO, Hall P, Czene K. 2007. Is breast cancer prognosis inherited? *Breast Cancer Res.* 9:R39
22. 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.* 1:214–22
23. Hitzemann R, Malmanger B, Reed C, Lawler M, Hitzemann B, et al. 2003. A strategy for the integration of QTL, gene expression, and sequence analyses. *Mamm. Genome* 14:733–47
24. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, et al. 2007. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* 39:870–74
25. Hunter KW, Broman KW, Voyer TL, Lukes L, Cozma D, et al. 2001. Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene Brms1. *Cancer Res.* 61:8866–72
26. Hunter KW, Williams RW. 2002. Complexities of cancer research: mouse genetic models. *ILAR J.* 43:80–88
27. International CGSC. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716
28. Jansen RC, Nap JP. 2001. Genetical genomics: the added value from segregation. *Trends Genet.* 17:388–91
29. Key TJ, Schatzkin A, Willett WC, Allen NE, Spencer EA, Travis RC. 2004. Diet, nutrition and the prevention of cancer. *Public Health Nutr.* 7:187–200
30. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921

31. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803–19
32. Linos E, Willett WC. 2007. Diet and breast cancer risk reduction. *J. Natl. Compr. Cancer Netw.* 5:711–8
33. Miki Y, Swenson J, Shattuck-Eidens D, Futreal P, Harshman K, et al. 1994. Isolation of BRCA1, the 17q-linked breast and ovarian cancer susceptibility gene. *Science* 266:66–71
34. Moore JH, Ritchie MD. 2004. The challenges of whole-genome approaches to common diseases. *JAMA* 291:1642–43
35. Mott R, Flint J. 2002. Simultaneous detection and fine mapping of quantitative trait Loci in mice using heterogeneous stocks. *Genetics* 160:1609–18
36. Nadeau JH, Frankel WN. 2000. The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nat. Genet.* 25:381–84
37. Oldenburg RA, Kroeze-Jansema K, Kraan J, Morreau H, Klijn JG, et al. 2003. The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. *Cancer Res.* 63:8153–57
38. Park YG, Clifford R, Buetow KH, Hunter KW. 2003. Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res.* 13:118–21
39. Park YG, Zhao X, Lesueur F, Lowy DR, Lancaster M, et al. 2005. Sip1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1. *Nat. Genet.* 37:1055–62
40. Reilly KM, Loisel DA, Bronson RT, McLaughlin ME, Jacks T. 2000. Nf1;Trp53 mutant mice develop glioblastoma with evidence of strain-specific effects. *Nat. Genet.* 26:109–13
41. Reilly KM, Tuskan RG, Christy E, Loisel DA, Ledger J, et al. 2004. Susceptibility to astrocytoma in mice mutant for Nf1 and Trp53 is linked to chromosome 11 and subject to epigenetic effects. *Proc. Natl. Acad. Sci. USA* 101:13008–13
42. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, et al. 2007. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* 39:596–604
43. 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* 18:473–81
44. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PI, et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316:1331–36
45. Schatzkin A, Longnecker MP. 1994. Alcohol and breast cancer. Where are we now and where do we go from here? *Cancer* 74:1101–10
46. Schultz JM, Yang Y, Caride AJ, Filoteo AG, Penheiter AR, et al. 2005. Modification of human hearing loss by plasma-membrane calcium pump PMCA2. *N. Engl. J. Med.* 352:1557–64
47. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–45
48. Serre D, Montpetit A, Pare G, Engert JC, Yusuf S, et al. 2008. Correction of population stratification in large multi-ethnic association studies. *PLoS ONE* 3:e1382
49. Threadgill DW, Hunter KW, Williams RW. 2002. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm. Genome* 13:175–78
50. Valdar W, Flint J, Mott R. 2006. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172:1783–97
51. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, et al. 2007. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat. Genet.* 39:827–29
52. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. *Science* 291:1304–51
53. Wade CM, Kulbokas EJ 3rd, Kirby AW, Zody MC, Mullikin JC, et al. 2002. The mosaic structure of variation in the laboratory mouse genome. *Nature* 420:574–78
54. Wang-Gohrke S, Rebbeck TR, Besenfelder W, Kreienberg R, Runnebaum IB. 1998. p53 germline polymorphisms are associated with an increased risk for breast cancer in German women. *Anticancer Res.* 18:2095–99

55. Wang X, Korstanje R, Higgins D, Paigen B. 2004. Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res.* 14:1767–72
56. Wang X, Paigen B. 2002. Quantitative trait loci and candidate genes regulating HDL cholesterol: a murine chromosome map. *Arterioscler. Thromb. Vasc. Biol.* 22:1390–401
57. Wang X, Paigen B. 2005. Genetics of variation in HDL cholesterol in humans and mice. *Circ. Res.* 96:27–42
58. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–62
59. Wellcome TCCC. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78
60. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, et al. 2007. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet.* 39:1000–6
61. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, et al. 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378:789–92
62. Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. 2007. On the subspecific origin of the laboratory mouse. *Nat. Genet.* 39:1100–7
63. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, et al. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* 39:645–49
64. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, et al. 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39:989–94



Contents

Mid-Century Controversies in Population Genetics <i>James F. Crow</i>	1
Joshua Lederberg: The Stanford Years (1958–1978) <i>Leonore Herzenberg, Thomas Rindfleisch, and Leonard Herzenberg</i>	19
How <i>Saccharomyces</i> Responds to Nutrients <i>Shadia Zaman, Soyeon Im Lippman, Xin Zhao, and James R. Broach</i>	27
Diatoms—From Cell Wall Biogenesis to Nanotechnology <i>Nils Kroeger and Nicole Poulsen</i>	83
Myxococcus—From Single-Cell Polarity to Complex Multicellular Patterns <i>Dale Kaiser</i>	109
The Future of QTL Mapping to Diagnose Disease in Mice in the Age of Whole-Genome Association Studies <i>Kent W. Hunter and Nigel P.S. Crawford</i>	131
Host Restriction Factors Blocking Retroviral Replication <i>Daniel Wolf and Stephen P. Goff</i>	143
Genomics and Evolution of Heritable Bacterial Symbionts <i>Nancy A. Moran, John P. McCutcheon, and Atsushi Nakabachi</i>	165
Rhomboid Proteases and Their Biological Functions <i>Matthew Freeman</i>	191
The Organization of the Bacterial Genome <i>Eduardo P.C. Rocha</i>	211
The Origins of Multicellularity and the Early History of the Genetic Toolkit for Animal Development <i>Antonis Rokas</i>	235
Individuality in Bacteria <i>Carla J. Davidson and Michael G. Surette</i>	253

Transposon Tn5 <i>William S. Reznikoff</i>	269
Selection on Codon Bias <i>Ruth Hershberg and Dmitri A. Petrov</i>	287
How Shelterin Protects Mammalian Telomeres <i>Wilhelm Palm and Titia de Lange</i>	301
Design Features of a Mitotic Spindle: Balancing Tension and Compression at a Single Microtubule Kinetochore Interface in Budding Yeast <i>David C. Bouck, Ajit P. Joglekar, and Kerry S. Bloom</i>	335
Genetics of Sleep <i>Rozi Andretic, Paul Franken, and Mehdi Tafti</i>	361
Determination of the Cleavage Plane in Early <i>C. elegans</i> Embryos <i>Matilde Galli and Sander van den Heuvel</i>	389
Molecular Determinants of a Symbiotic Chronic Infection <i>Kattherine E. Gibson, Hajime Kobayashi, and Graham C. Walker</i>	413
Evolutionary Genetics of Genome Merger and Doubling in Plants <i>Jeff J. Doyle, Lex E. Flagel, Andrew H. Paterson, Ryan A. Rapp, Douglas E. Soltis, Pamela S. Soltis, and Jonathan F. Wendel</i>	443
The Dynamics of Photosynthesis <i>Stephan Eberhard, Giovanni Finazzi, and Francis-André Wollman</i>	463
Planar Cell Polarity Signaling: From Fly Development to Human Disease <i>Matias Simons and Marek Mlodzik</i>	517
Quorum Sensing in Staphylococci <i>Richard P. Novick and Edward Geisinger</i>	541
Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes <i>Jennifer A. Marshall Graves</i>	565
The Take and Give Between Retrotransposable Elements and Their Hosts <i>Arthur Beauregard, M. Joan Curcio, and Marlene Belfort</i>	587
Genomic Insights into Marine Microalgae <i>Micaela S. Parker, Thomas Mock, and E. Virginia Armbrust</i>	619
The Bacteriophage DNA Packaging Motor <i>Venigalla B. Rao and Michael Feiss</i>	647

The Genetic and Cell Biology of Wolbachia-Host Interactions <i>Laura R. Serbus, Catharina Casper-Lindley, Frédéric Landmann, and William Sullivan</i>	683
Effects of Retroviruses on Host Genome Function <i>Patric Jern and John M. Coffin</i>	709
X Chromosome Dosage Compensation: How Mammals Keep the Balance <i>Bernhard Payer and Jeannie T. Lee</i>	733

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at <http://genet.annualreviews.org/errata.shtml>