Hunter KW, Williams RW (2002) Quantitative trait locus analysis in cancer research. ILAR J 43: 80–88.

Complexities of Cancer Research: Mouse Genetic Models

Kent W. Hunter and Robert W. Williams

Abstract

Cancer susceptibility is a complex interaction of an individual's genetic composition and environmental exposures. Huge strides have been made in understanding cancer over the past 100 yr, from recognition of cancer as a genetic disease, to identification of specific carcinogens, isolation of oncogenes, and recognition of tumor suppressors. A tremendous amount of knowledge has accumulated about the etiology of cancer. Cancer genetics has played a significant role in these discoveries. Analysis of high-risk familial cancers has led to the discovery of new tumor suppressor genes and important cancer pathways. These families, however, represent only a small fraction of cancer in the general population. Most cancer is instead probably the result of an intricate interaction of polymorphic susceptibility genes with the sea of environmental exposures that humans experience. Although the central cadre of cancer genes is known, little is understood about the peripheral genes that likely comprise the polymorphic susceptibility loci. The challenge for cancer genetics is therefore to move forward from the Mendelian genetics of the rare familial cancer syndromes into the field of quantitative trait loci, susceptibility factors, and modifier genes. By identifying the genes that modulate an individual's susceptibility to cancer after an environmental exposure, researchers will be able to gain important insights into human biology, cancer prevention, and cancer treatment. This article summarizes the current state of quantitative trait genetic analysis and the tools, both proven and theoretical, that may be used to unravel one of the great challenges in cancer genetics.

Key Words: cancer; genetics; modifiers; QTLs; susceptibility

The study of cancer genetics has provided insight into the cause and control of neoplastic processes. Knudson's "two-hit" model of tumor initiation (Knudson et al. 1975), for example, traces back to studies of rare familial malignancies. In addition to the classical oncogenes (genes whose products promote cell proliferation) (Huebner and Todaro 1969; Todaro and Huebner 1969), Knudson advanced the idea of a complementary gene class—the tumor suppressors—whose function is to inhibit or check cell proliferation. Inheritance of an inactivated allele of a suppressor gene (the first hit) increases the risk of cancer. The stochastic somatic inactivation of the second allele (the second hit) then triggers uncontrolled proliferation. This tumor suppressor model gained momentum a decade later with the cloning and molecular dissection of the retinoblastoma gene and protein (Dryja et al. 1986). Other suppressors followed in rapid succession, including WT1 (Haber et al. 1990), NF1 (Wallace et al. 1990), and p53 (Baker et al. 1989; Eliyahy et al. 1989; Finlay et al. 1989; Levine et al. 1989; Nigro et al. 1989; Stanbridge 1989).

The discovery and analysis of these and other genes have catalyzed progress in cancer research. First, identification of cancer gene variants has exposed novel molecular pathways associated with apoptosis (Lowe et al. 1993), DNA damage control (Savitsky et al. 1995), and novel cell cycle checkpoint mechanisms (Riese et al. 1999). Second, our understanding of complex interactions among oncogenes and tumor suppressors has been enriched (Chen and Defendi 1992). Molecular pathway analysis has revealed the convergence and apposition (or possibly, opposition) of the effects of growth-promoting oncogenes and the growth-regulatory action of tumor suppressors in cancer etiology. Tumor suppressor genes often act as guardians at key points in pathways, overseeing and regulating signals transduced by proto-oncogenes and cofactors (Picksley et al. 1996; Sherr 1996).

The focus of this review is on a third contribution of genetics to cancer research, namely, the growing appreciation of the genetic complexity of cancer. Cancer is not a single disease but is instead a family of related diseases. Although the hallmark of cancer—the unregulated proliferation of cells—is common among all tumors, the genes and the genetic and genomic events leading to neoplastic disease are distinct for each tumor type. The war on cancer has been long-drawn-out due to this characteristic. Cancer research is evolving to develop new methods of diagnostics, including molecular or expression array analysis, to define the tumor type better and to develop more effective, tailored therapeutic approaches based on the particular tumor etiology.

Where is cancer genetics heading now, and what roles will genetics play in cancer biology in the postsequence era? Most of the rare familial cancer genes that have contributed so much to our understanding of the central processes of tumorigenesis have been identified. Now that the human genome has been sequenced (Lander et al. 2001; Venter et al. 2001), with the mouse and other vertebrates soon to follow, the few remaining inherited cancer syndromes will surely be identified quickly. Simple computer-based"in silico" experiments comparing genomes for conserved sequences will enable researchers to identify genes and evolutionarily important transcription control elements (Dehal et al. 2001; Scherf et al. 2001). Additional genes that play important roles in cell cycle control may be identified by sequence homology or by the genome-scale proteomics projects.

Is cancer genetics therefore entering the declining phase of useful techniques and strategies? No, clearly not. Although cancer genetics has helped define the key molecules and genes that play critical roles in cell proliferation, there are still large areas of comparative ignorance. The vast majority of cancer incidence is not associated with familial cancer syndromes. Cancer in the general population is instead probably the result of an intricate interaction of polymorphic susceptibility genes with the sea of environmental exposures that humans experience.

Although the central cadre of cancer genes is known, little is understood about the peripheral genes that likely comprise the polymorphic susceptibility loci. Some candidates for cancer susceptibility genes (e.g., the xenobiotic-metabolizing pathway members) appear obvious. However, the biology of the cell does not always meet our preconceived notions. For example, it is difficult to imagine that the NF2 gene would have been selected as a tumor suppressor gene based on its functional interaction with the cytoskeleton (Gutmann 1994). Likewise, it is difficult to imagine how investigators will be able to identify low penetrant cancer modifier loci that may have very important roles

in cancer susceptibility in the general population simply by biochemical interactions or sequence comparisons.

The challenge for cancer genetics is therefore to move forward from the mendelian genetics of the rare familial cancer syndromes into the field of quantitative trait loci (QTLs1), susceptibility factors, and modifier genes. By identifying the genes that modulate an individual's susceptibility to cancer after an environmental exposure, researchers will be able to gain important insights into human biology. Among these might be (1) a better understanding of how cellular biochemistry and metabolism interacts with the cell cycle, (2) greater understanding of molecular interactions of the cell's biochemical networks, (3) novel genetic polymorphisms to be used in conjunction with the molecular tools currently under development to yield more accurate diagnosis and subsequent therapies, and (4) an expanded understanding of the effects of the interaction of human inheritance and environment.

Although this review is intended to be a discussion of quantitative trait analysis in cancer research, it should be noted that the concepts and methodologies discussed here are common to all complex trait genetic analysis. As a result, the discussion is presented in a general format, rather than specifically focusing on cancer. Key to the success of any complex trait genetic analysis, however, is the ability to phenotype accurately. Whether investigators are studying spontaneous tumor initiation, chemical or radiation sensitivity, tumor multiplicity, tumor growth or progression, or other non-cancer-related phenotypes, it is critical to be able to define and accurately quantitate the variable in question. Incorrect phenotyping, like miscalled genotypes, can significantly reduce the power to detect genetic linkage in any type of complex trait analysis. Careful delineation of the problem and accurate and systematic measurement of the phenotypes will therefore significantly improve the probability of identifying and mapping genetic modifiers.

What then are the tools available for quantitative trait analysis in cancer research? Because the ultimate goal of cancer research is to address human disease, the most relevant experimental system would obviously be human populations and patient material. The low penetrance of the cancer susceptibility genes in the general population, in combination with the unequal exposures to environmental carcinogens, precludes standard linkage analysis because susceptibility would not appear as a familial trait. Association studies using large populations to find linkage disequilibrium between polymorphisms and cancer susceptibility could be utilized to circumvent this problem. Recent analysis, however, suggests that performing a complete genome scan in human associations studies would require assaying tens to hundreds of thousands of polymorphic loci per individual (Dunning et al. 2000; Kruglyak 1999). With current technologies, performing genome-wide association studies in human populations would be prohibitively expensive and are therefore, at least at present, not a routine practical experimental approach.

For the foreseeable future, therefore, the rodent models will be the workhorses for cancer quantitative trait genetics. Rodents, particularly the mouse, have many advantages for susceptibility genetics. Mice are small, are relatively inexpensive to house, and have a short generation time. Many different inbred strains are available, and there is an increasing body of evidence to suggest that there is a tremendous amount of genetic diversity among the inbred strains to be studied. High-resolution genetic and genomic resources are available, as well as the ability to manipulate the genome at both the

individual nucleotide and the gross chromosomal level. As a result, it is possible to engineer animals to test the effect of specific polymorphisms or mutations on a phenotype of interest directly.

A number of different approaches have been developed to tackle the problem of identifying genes that match up with QTLs. Most recently it has been suggested that large-scale mutagenesis projects are the most efficient method of identifying modifier genes (Nadeau and Frankel 2000). A number of different centers have been established worldwide and are currently generating and analyzing mutagenized animals for a variety of behavioral, developmental, and biochemical phenotypes (e.g., Hrabe de Angelis et al. 2000; Nolan et al. 2000). Although this strategy is enormously powerful for identifying new genes and novel mutations leading to a phenotype of interest, it is likely to have limited utility for cancer susceptibility genetics for a number of practical reasons. Cancer is usually a disease of the aged, both in humans and in rodents. Except for the retrovirally induced cancer, spontaneous tumors in the mouse generally arise late in life. Therefore, to observe significant changes in tumor incidence in mutagenized animals would require aging large numbers for many months, which would be a significant burden to perform in a systematic way. In addition, except for mammary tumors or some skin cancer models, in many cases early-stage cancers in mice are not obvious. Systematic palpation, whole body imaging, or postmortem pathological analysis is required to identify and characterize the tumor load and spectrum on each individual animal. It would then be necessary to use cryopreserved gametes to repeat the process to confirm heritability. Again, the logistics and expense of a thorough study, testing all of the progeny of a mutagenesis program for effects on a particular cancer type, would likely preclude its general use.

In spite of the expense and difficulties, a number of mutagenesis centers are, and should be, aging mutagenized animals to identify cancer phenotypes. Ethylnitrosourea mutagenesis will clearly affect those genes in cancer pathways as well as those in behavioral or developmental pathways. As a result, many of the mutagenized animals must harbor mutations in important cancer genes. However, unless these animals are maintained for long periods of time (>1 yr), it is likely that the cancer phenotypes detected in the mutagenesis studies will be early-arising tumors. They will likely be due to inactivating mutations in tumor suppressor genes or activating events in proto-oncogenes that result in strong oncogenic potential. Although scientists will undoubtedly identify a number of new cancer-causing point mutations in the known genes and will discover potentially new members of the tumor-promoting and -suppressing gene families, these discoveries will likely represent models of the familial cancer families, with inherited cancer-inducing mutations, rather than tumor susceptibilities that are present in the bulk of the human population. Mutations that result in weakly penetrant cancer phenotypes would be less likely to be observed unless significant numbers of each pedigree were aged for long periods of time. To use the mutagenesis strategy to search for these low penetrance genes or polymorphisms that likely more closely model the differential susceptibility to cancers in the general human population, it would be necessary to reintroduce perhaps the most significant variable eliminated in many spontaneous cancer models---the environmental component.

The vast majority of cancer in the human population is not the result of mendelian inheritance, an activated proto-oncogene, or inactivated tumor suppressor. Rather, it is

due to the intersection of genetic predispositions caused by polymorphisms in metabolic or regulatory genes and exposure to carcinogenic environmental agents. Thus, individuals with similar genetic susceptibility may have significantly different cancer incidences depending on differences in regional or lifestyle-associated carcinogen exposure. Conversely, tumor incidence can vary greatly among individuals with similar degrees of high-level exposure to potent genotoxic agents, presumably due to differences in inherited susceptibilities. Modeling the genotype/phenotype intersection in mutagenized mice would therefore require exposing large numbers of animals to a variety of agents and maintaining them for extended periods of time.

The most practical tools then are the experimental crosses of strains known to differ significantly in tumor susceptibility and gene-environment interactions. A variety of different methods have been developed or adapted to extract gene loci that modify cancer development. The most commonly used strategy is mapping in backcross or intercross mapping. More sophisticated strategies include combinations of recombinant inbred mapping panels and experimental crosses, recombinant congenics, chromosomal substitution strains, recombinant inbred segregation tests, and, most recently, recombinant inbred intercrosses. Each of these strategies has its strengths and weaknesses and is discussed briefly below.

Backcross and Intercross Analysis

The most common forms of quantitative trait analysis, backcross and intercross mapping, are robust methods for generating low-resolution localization of genes that modulates phenotypes. The choice between backcross and intercross methods depends on several factors (Darvasi 1998). To generate an idea of the approximate number of modifier loci present and to obtain estimates of their additive or dominance effects, an F2 intercross generally requires fewer animals than a backcross (Darvasi 1998). For loci that act additively, it has been estimated that F2 crosses require approximately 30% fewer animals than a backcross. Conversely, backcross strategies are much more efficient in situations in which a few loci are known to have dominant effects. Furthermore, backcrosses, because there are only two genotype classes at each locus compared with three for intercross analysis, are more efficient in detecting epistatic interactions between unlinked modifiers.

A major strength of both the backcross and intercross methods, unlike many of the other strategies to be discussed, is that any two genetically distinct strains of mice can be used to generate a mapping cross. Thus, the genetic components of any phenotype that differ between two of the multitude of inbred mouse strains can be subjected to genetic dissection. This capability includes both inbred strains, including common inbred strains (Mock et al. 1993), ferally derived mice (Nagase et al. 2001), transgenics (Le Voyer et al. 2000), and knockout mice (Bolivar et al. 2001). A number of laboratories are taking advantage of the strain background affects on transgenic or targeted mutations to map genes that modify a variety of cancer phenotypes, including latency, tumor growth, and metastatic progression (Dragnani et al. 2000; Hunter et al. 2001; Le Voyer et al. 2000, 2001; Moser et al. 2001; Rowse et al. 1998).

A significant disadvantage of using traditional mouse genetic methods is the lack of resolution. Genetic localization obtained by these methods is usually measured in the

tens of centimorgans rather than the subcentimorgan resolution required for efficient positional cloning. As a result, preliminary genome localization of QTLs by these methods must be supplemented with fine mapping analysis, requiring the construction of interval-specific congenic animals and subsequent analysis of subcongenic intervals (Darvasi 1998; Potter et al. 1994). This process therefore is a long, laborious path that requires large numbers of animals to achieve the desired goal.

Recombinant Inbred Mapping

Recombinant inbred (RI1) strains (Figure 1) have been used to map a wide range of mendelian loci and quantitative traits (Grisel et al. 1997; Klein et al. 1998; Tafti et al. 1999). They offer compelling advantages for mapping complex genetic traits, particularly those that have low heritabilities. Each recombinant genome is replicated in the form of an entire isogenic line (Bailey 1981). A major advantage of recombinant inbred panels is that the variance associated with uncontrolled error can be suppressed to very low levels (Belknap 1998). This effectively elevates heritability and greatly improves prospects for mapping QTLs. Gene effects can also be tested under a spectrum of environments and using numerous experimental perturbations. RI strains can therefore be exploited to expose gene-environment interactions. In contrast, interactions between genes and the environment cannot usually be studied using conventional mapping populations in which each animal is unique. A third advantage of RI strains is that genotypes generated by different groups using a variety of methods can be pooled to generate high-density linkage maps of both markers and QTLs (Williams et al., 2001). As a result, loci that segregate in RI sets can usually be mapped without genotyping. This attribute was a significant advantage before the advent of efficient and easy PCR genotyping methods (Bailey 1981).

Although RI strains can significantly reduce the effort required to obtain QTL linkage information, there are a number of limitations. First, RI sets exist for relatively few of the key inbred strains. As a result, only the subset of polymorphic loci that differ among these few groups of inbred mouse strains can be mapped using this strategy. Fortunately, these strains are among the most widely used and well characterized (C57BL/6J, A/J, BALB/cJ, DBA/2J) (Taylor 1996). Second, despite the increased power associated with the ability to assay identical genotypes repeatedly, the limited number of lines in each RI panel (at the time of this writing, still fewer than 40 per set) precludes mapping QTLs, which do not account for a substantial fraction of the genetic variance (Belknap et al. 1996; Pataer et al. 1997). Using the original 26-strain BXD RI set, it was estimated that a QTL would have to account for >26% of the variance to be detected consistently. Mapping QTLs that account for less variance requires a two-tiered approach—identifying putative loci using the RI analysis followed by a second experiment to confirm the QTL assignment an independent cross. Third, like the backcross and intercross strategies, the precision with which QTLs can be mapped is generally poor, with candidate regions often spanning 10 to 20 centimorgans. Resolution and power in RI analysis would be improved by increasing the number of strains available in each set. Each RI strain has roughly the same utility for mapping complex traits as 5-20 F2 progeny; the number varies as a function of heritability and number of QTLs influencing a given trait (Williams et al. 2001). Although this increase will be a

significant boon for those studying traits that vary between C57BL/6 and DBA/2J, many of the cancer traits or modifier analysis will not benefit due to lack of availability of large RI sets between many other inbred strains of interest.

Recombinant Congenic Analysis

The complex and often nonlinear interactions among many polymorphic genes and their products are difficult in QTL mapping. Standard backcross, intercross, and recombinant inbred analysis often lack the power required to dissect multiple unlinked genes, particularly if they interact epistatically. To overcome this limitation, a variation of recombinant inbred panels, the recombinant congenics (RC1), has been developed (Groot et al. 1996). Like the RI panels, RC panels are composites of the genomes of two inbred progenitor parent strains. Unlike the RIs, which are generated by serial brother-sister matings after an initial outcross, two rounds of backcrossing to one of the progenitor strains precedes the brother-sister mating to achieve genome homozygosity (Figure 1). The resulting RC strains contain a reduced amount of the donor strain genome (12.5%) compared with backcross, intercross, or RI analysis. The greater partitioning of the donor genome can result in isolation of individual modifier genes into different RC strains where their effects may be studied in isolation. Although RCs cannot resolve very large numbers of genes each with small effects, it has been calculated that effects due to 5 to 7 genes with major effects can be successfully addressed (Demant and Hart 1986). Gene interactions and epistasis can also be addressed effectively in recombinant congenics (Fijneman et al. 1996, 1998a,b). The reduction in the amount of donor genome present in the RC compared with the RI lines increases the probability of separating interacting genes into individual strains where their effects can be studied in isolation. However, like the RIs, RCs are also time consuming to make; and only traits that differ between the progenitor strains can be assayed (Demant and Hart 1986; Fortin et al. 2001).

Congenics, Chromosomal Substitutions

One of the most common strategies used to validate QTLs and then to refine their positions is to generate sets of congenic lines (Graff et al. 1966) in which high and low alleles are swapped between the low and high parental strains. If the QTL is real and has been mapped to the correct interval, then phenotypes of the reciprocal congenic lines should each deviate from the recipient strains by a predictable amount (2a, where a is the additive QTL effect). One exception, however, occurs when the initial localization was due to two or more linked genes each contributing to the phenotypic variance. Separation of the loci by breeding (e.g., during the generation of a high-resolution subcongenic panel) can result in a multiple linked intervals contributing unequally to the phenotypic variance (Morel et al. 2001).

The strength of the congenic method is that it will usually convert a polygenic trait to a monogenic (pseudomendelian) quantitative trait. In other words, any differences in the mean phenotypes of the congenic and the background strain should be generated by a single polymorphic locus in the introgressed interval. When the congenic lines have the expected phenotype, it is not difficult to shave down the size of the interval progressively while at the same time fine-mapping the QTL position. One of the drawbacks with a

congenic approach is that each QTL needs its own mouse resources. There is a substantial genotyping load and many generations of backcrossing (5 generations for speed congenics [Markel et al. 1997] and about 8 to 10 for standard congenics) for the derivation of this mapping resource.

A variation on this theme is the chromosomal substitution strains (Nadeau et al. 2000). Like the interval-specific congenic animals, chromosomal substitution strains carry a genomic segment from a donor strain, in this instance a whole chromosome, in the background of a second inbred recipient strain. If a full panel of 21 strains is constructed (chromosomes 1-19, X, and Y), the chromosomal assignment of a modifier can be rapidly determined for any trait that differs between the parental strains. Like the RIs, RCs, and congenics, chromosomal substitution strains (CSSs1) have defined and fixed genotypes. The CSS increase the power to detect QTLs because the variance due to environmental or experimental fluctuation can be averaged out by repeated analysis of a genotype. CSSs can then be used to develop a series of overlapping interval-specific congenic animals for high-resolution mapping. A disadvantage of this system is that the initial QTL assignment is to a whole chromosome, rather than a chromosomal region. In addition, in the initial analysis, CSSs cannot distinguish between single and multiple QTLs on the donor chromosome. Although the use of reciprocal CSS panels can identify the presence of interactions between host and donor loci, no linkage information on the host modifiers would be obtained. Furthermore, CSS panels, by isolating individual donor chromosomes, preclude the possibility of detecting epistatic interactions between unlinked donor loci.

RIX Mapping

A novel method recently proposed to increase the power of recombinant inbred mapping is a straightforward extension of RI analysis called RIX mapping, developed by Threadgill, Williams, and colleagues (Williams et al. 2001). This new method relies on a potentially large set of diverse F1 intercrosses (X) generated from many different pairs of RI strains. RIX mapping exploits the well-characterized RI strains and has comparatively high statistical power for detecting QTLs. This method can also improve positional precision somewhat. RIX mapping shares all of the advantages of RI mapping (Bailey 1981; Belknap et al. 1996) and has the added benefit of providing a very large sample of unique (but predictable) genotypes. Two additional advantages are hybrid vigor of the F1 progeny used to assess phenotypes and the ability to make each F1 by reciprocal crosses, a feature that can partly control for parental effects. The RIX method can be used to confirm a QTL and refine its position. This confirmation set will usually involve making a set of RIXs that have known genotypes in the relevant interval hypothesized to contain a QTL. The prediction is that the mean phenotype of each RIX line should conform to expectation given its known genotype in a test interval (e.g., among the BXD RIXs, the genotypes must be BB, BD, or DD).

Similarly, an RI can be backcrossed to one or many more inbred strains to generate recombinant backcross (RIB1) progeny. Unlike RI lines, the RIB crosses share segregation ratios similar to those of standard backcross, with a 1:1 ratio of homozygous to heterozygous genotypes. Each particular litter of RIB mice again have a predictable recombinant genotype defined by the particular combination of parental genomes. RIB

progeny can be used to solve one obvious limitation of RI lines—their lack of heterozygous genotypes. The RIB cross is particularly useful in that modifiers of dominant and semidominant mutations, null alleles, or transgenes can be easily mapped in a single generation cross without genotyping.

The Next Hurdle: Finding the Gene

Mapping cancer QTLs is obviously only a first step. Identification and characterization of the genes and alleles that control cancer susceptibility and progression comprise the ultimate goal, not simply assigning them to genomic "bins." The challenge up to now has been to take those final steps, going from low-resolution mapping to genetic polymorphism or mutation. To date, relatively few cancer QTLs have been identified, by any method (MacPhee et al. 1995; Zhang et al. 1998). Specialized high-resolution mapping experiments, based on either interval-specific or RIX- and RIB-style methodologies, will continue to play a major role in our efforts to identify the genetic components that influence tumor initiation, progression, drug response, and the myriad other cancer phenotypes of interest. The recent proliferation of genomic resources and techniques, however, will greatly change how the strategies used in cancer genetics. First and foremost among these resources will be the complete sequence of the mouse genome and the identification from the sequence and the expressed sequence tag (EST1) databases of the vast majority of the genes. High-resolution mapping will therefore narrow the list of potential candidate genes from the ~30,000 in the genome to a small handful. Comparative genomics will also aid in reducing the number of candidate genes that need to be analyzed. Comparing the results of QTL studies in different species and combining the mapping data with the genomic sequence data to identify transcripts in the region have the possibility of significantly limiting the number of genes that need to be analyzed (e.g., Stoll et al. 2000). The high level of conservation of local gene order between mammalian species suggests that only those genes that are shared between the species on orthologous chromosomal segments are likely candidates for those modifier genes affecting the same trait that map to orthologous regions. Furthermore, knowledge of the tissue expression pattern of these genes and the biochemical pathways in which they operate may reduce the number of potential candidates further. The ability to generate "designer" chip microarrays, as well as the ability to microdissect and assay specific cells from normal or tumor tissue, can further prioritize all of the genes in a candidate interval for analysis by identifying genes in QTL candidate regions expressed in the appropriate tissues or displaying differential expression patterns (Sandberg et al. 2000).

The subsequent search for the polymorphisms or mutations underlying trait variation can and will become faster and simpler. The public and private mouse genome sequencing projects will be a source of polymorphism "in silico." Four different inbred strains have been or are in the process of being sequenced to date (129X1/SvJ, DBA/2J, C57BL/6J and A/J; see <www.celera.com> and <ww.ncbi.nlm.nih.gov/genome/seq/MmHome.html>), providing candidate single nucleotide polymorphisms (SNPs1) in the progenitor strains for the largest RI sets and the reciprocal CSS panels. Investigators using these strains will be able to identify rapidly and prioritize potential SNPs in their candidate regions for validation and further

analysis by simple computer searches. Bioinformatics (<http://cgap.nci.nih.gov/Catalog>) tools are also available to search ESTs, which have been generated from a variety of different inbred strains, for potential additional gene-based SNPs. For other inbred strains not represented in these databases, the availability of the genomic sequence permits the design of primers to sequence exons and promoters in the particular strains of interest, using high-throughput sequencing technologies.

Finding polymorphisms in candidate genes in pathways of interest is always intriguing, but the final proof that any particular genetic alteration is the causative agent of a trait modification is likely to be much more elusive. Polymorphisms or mutations that result in translation chain terminations, like the Mom1 QTL (MacPhee et al. 1995), although relatively easy to test are likely to be rare. A more likely scenario is one of amino acid substitution or promoter mutation. A more subtle possibility would be silent mutations in "splicing enhancers," which might alter the efficiency of proper mRNA processing (Liu et al. 2001). How can one convincingly test the effect of a particular polymorphism on a trait? The ultimate test of course would be to substitute the gene from one of the two strains differing in a trait with the polymorphic allele from the other strain. Unfortunately, due to the paucity of germline-transmissible embryonic stem cells, this is possible for only a few strains (129X1/SvTac, C57BL/6, DBA/1). For other strains, investigators would have to generate "knock-in" animals in one of these strains and then generate congenics to test the effect of a particular polymorphism or mutation. The danger with this strategy is the potential introduction of modifiers from the embryonic stem cell background, either in the congenic interval, linked to the knock-in construction, or potentially elsewhere in genomic regions that were fixed undetected during breeding. As a result, the design and interpretation of these experiments must be made with care.

Does that mean that definitive proof is unlikely for most QTL projects? If it is not possible to perform the cleanest experiments, the knockins, then truly definitive proof may be difficult to obtain. However, as has been suggested in another review (Belknap et al. 2001), it should still be possible to obtain a large enough body of evidence through genetic experiments to provide compelling support for any particular polymorphism or mutation. Combining these experiments with biochemical and in vitro studies, somatic cell genetics and comparative genetics, and genomics should provide a number of avenues to explore the causal role of any particular polymorphism. In addition, the ability to screen thousands to hundreds of thousands of compounds synthesized in combinatorial chemistry libraries offers the opportunity for investigators to pursue "chemical genetics" experiments to validate a particular gene or polymorphism (Stockwell 2000). By identifying compounds that perturb specific pathways, it may be possible for investigators to "replicate" knock-out, knock-in, or transgenic experiments by inhibiting, up-regulating, or modulating a particular pathway or molecule using chemical probes. This strategy has the tremendous advantage of being applicable to any inbred strain, unlike the gene-targeting strategies. Chemical genetics may therefore provide an efficient and effective method to probe modifier genes that are present in strains other than the mainstays of the genetically modified mouse models, 129, C57BL/6, and FVB/N.

Kent W. Hunter, Ph.D., is an Investigator in the Laboratory of Population Genetics, NCI/NIH, Bethesda, Maryland. Robert W. Williams, Ph.D., is Dunavant Chair of Developmental Genetics at the Center of Genomics and Bioinformatics, Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, Tennessee.

1Abbreviations used in this article: CSS, chromosomal substitution strain; EST, expressed sequence tag; QTL, quantitative trait locus; RC, recombinant congenic; RI, recombinant inbred; RIB, recombinant backcross; SNP, single nucleotide polymorphism.

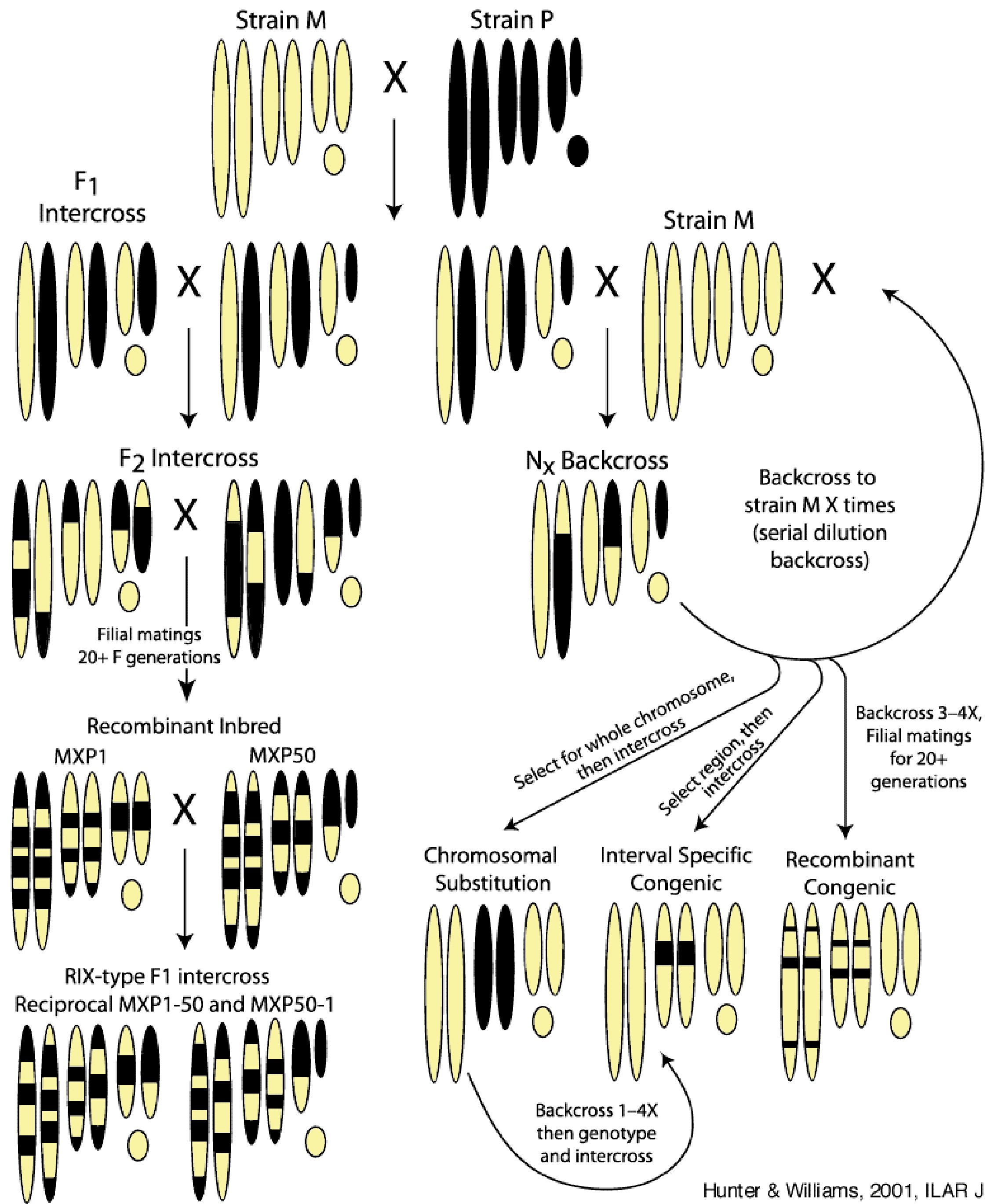

Figure 1 Schematic representation of the generation of experimental crosses and specialized mapping strains. Chromosomes are depicted by the oblong ovals. The middle and left pairs of chromosomes in each set are autosomes, the right hand pair represent the sex chromosomes (X, Y chromosomes). Mitochondrial DNA and cytoplasmic factors are depicted as the small circle beneath the sex chromosomes labeled "M." Chromosomes or chromosomal regions originating from the paternal (strain P)

donor strain are in black; the maternally derived (strain M) chromosomes are light colored.