

# Analysing complex genetic traits with chromosome substitution strains

Joseph H. Nadeau<sup>1,2</sup>, Jonathan B. Singer<sup>3</sup>, Angabin Matin<sup>1</sup> & Eric S. Lander<sup>3,4</sup>

---

Many valuable animal models of human disease are known and new models are continually being generated in existing inbred strains<sup>1,2</sup>. Some disease models are simple mendelian traits, but most have a polygenic basis. The current approach to identifying quantitative trait loci (QTLs) that underlie such traits is to localize them in crosses, construct congenic strains carrying individual QTLs, and finally map and clone the genes. This process is time-consuming and expensive, requiring the genotyping of large crosses and many generations of breeding. Here we describe a different approach in which a panel of chromosome substitution strains (CSSs) is used for QTL mapping. Each of these strains has a single chromosome from the donor strain substituting for the corresponding chromosome in the host strain. We discuss the construction, applications and advantages of CSSs compared with conventional crosses for detecting and analysing QTLs, including those that have weak phenotypic effects.

---

The current approach to identify QTLs that control phenotypic differences between two inbred strains, A and B, involves three steps.

First, initial detection of the QTL involves genetic mapping in a two-generation cross. One conducts either a backcross ((A×B)F1×B) or an intercross ((A×B)F1×(A×B)F1) of the two strains, phenotypes the progeny for the trait of interest, genotypes the progeny using a genetic linkage map and then correlates phenotype with genotype using appropriate analytical methods for QTL mapping<sup>3</sup>.

This first step entails considerable work and yields only modest map resolution. Large crosses are required to provide sufficient power to detect typical QTLs. For example, nearly 300 intercross progeny are required to detect a QTL responsible for at least 10% of the total variance. Genome scans involve determining tens of thousands of genotypes. For example, typing 100 loci in 300 individuals requires determining 30,000 genotypes. Moreover, the QTLs are localized with relatively poor resolution, typically approximately 20 cM, or about one-quarter of a mouse chromosome.

The second step towards identifying QTLs is the construction of congenic strains. In linkage crosses, segregating QTLs contribute 'phenotypic noise' that makes it difficult to be certain whether a given animal has inherited a specific QTL allele. Congenic strains, which were originally developed to dissect the complex genetic basis for histocompatibility<sup>4</sup>, resolve much of the uncertainty that is inherent in linkage crosses. Repeated backcrossing and selection for at least ten generations are used to transfer the chromosomal segment containing the QTL from strain B onto strain A (Fig. 1*a,b*). The resulting congenic strain is denoted A.B-x, where x designates the particular QTL. Once the congenic strain has been constructed, a comparison of the congenic strain A.B-x with parental strain A is used to assess the phenotypic effect of the QTL.

The third step to identify the QTL is fine-structure mapping. Backcrosses between the congenic strain A.B-x and host strain A

enable the precise location of the QTL to be mapped. Progeny that are recombinant within the congenic segment are identified with appropriate genetic markers and phenotyped, either directly or by subsequent progeny testing, to determine whether the recombinant chromosome carries the more or less penetrant allele. The QTL can be identified through standard genetic mapping and positional cloning, a task that will be facilitated by the availability of the mouse genome sequence.

QTL mapping studies have already identified hundreds of polygenic factors affecting medically and physiologically relevant traits. The overall process of QTL identification, however, remains tedious, requiring 10–15 generations of linkage mapping and congenic construction, corresponding to 3–5 years for the mouse and longer for the rat.

## Chromosome substitution strains

QTL mapping can be accelerated using chromosome substitution strains (CSSs), which enable the first two steps above to be omitted. We propose replacing the name 'consomic strains' with 'chromosome substitution strains' to clarify the nature of the strains and the process that is used to produce them. This change has been approved by the Internal Committee on Genetic Nomenclature for Mice.

We define CSS A.B-Chr(i) to be a homozygous inbred strain that is identical to strain A except that chromosome 'i' has been replaced by the corresponding chromosome from strain B (Fig. 1*c*). In the mouse, a CSS panel consists of 21 strains, corresponding to the 19 autosomes and two sex chromosomes. A CSS panel thus neatly partitions the genome into a collection of single chromosome substitutions on a defined and uniform genetic background. This contrasts with segregating crosses, recombinant congenic strains (RCSs) and recombinant inbred strains (RISs), in which the genomes of the progenitor strains are fragmented in a random, overlapping fashion among the panel of

---

<sup>1</sup>Department of Genetics, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. <sup>2</sup>Center for Human Genetics, University Hospitals of Cleveland, Cleveland, Ohio, USA. <sup>3</sup>Whitehead Institute for Biomedical Research and <sup>4</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence should be addressed to J.H.N. (e-mail: [jhn4@po.cwru.edu](mailto:jhn4@po.cwru.edu)).

strains (Fig. 1*d,e*). Although RISs and RCSs have little direct use for mapping multigenic traits, they have been used as a source of parental strains for specialized crosses aimed at QTL mapping.

A CSS panel facilitates the mapping of QTLs that control trait differences between its progenitor strains. One assays the strains in the panel for a phenotype of interest: if CSS A.B.Chr(*i*) differs from host strain A, there must be at least one QTL on chromosome *i*. This approach has several advantages over traditional QTL mapping. First, no crosses need be performed. Second, genotyping is not required because the segregation of the genome among the strains is already known in the panel of CSSs. Third, the design offers certain advantages in statistical power compared with segregating crosses, RISs and RCSs.

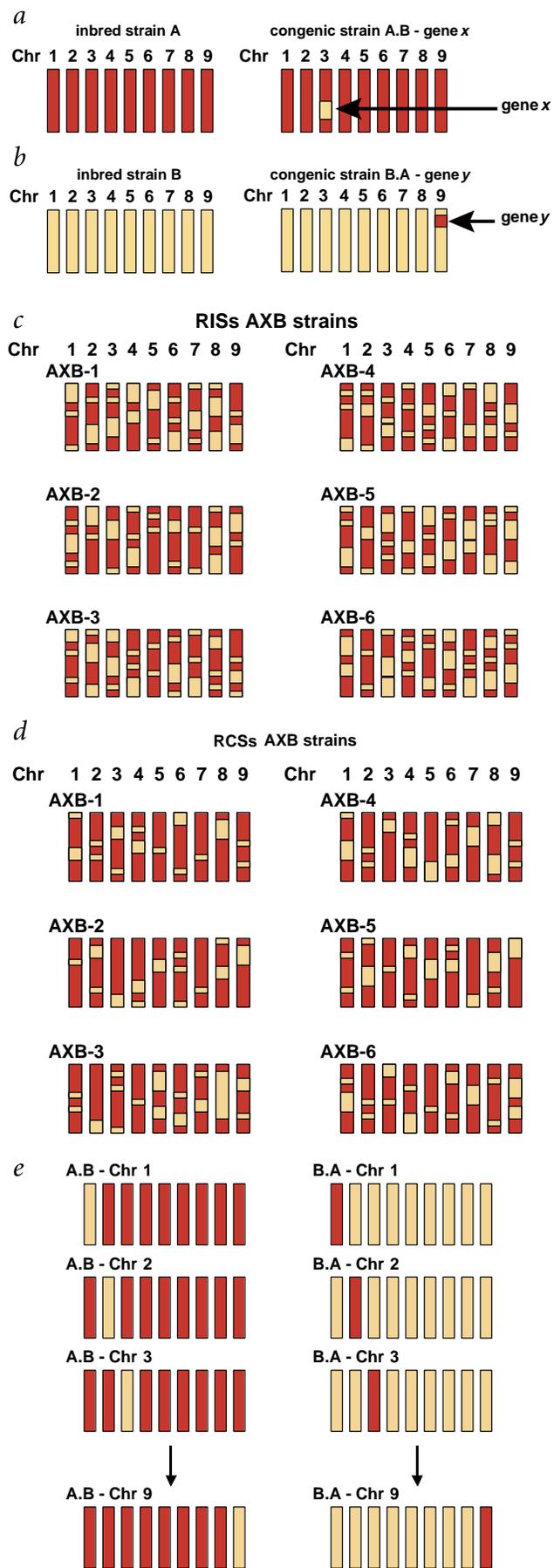
CSSs have two limitations. QTLs are initially assigned to an entire chromosome by CSS mapping, rather than to an approximately 20-cM region as in a cross. In addition, CSSs do not distinguish among multiple QTLs on the substituted chromosome. The ability, however, to proceed directly to fine-structure mapping without investing years in the construction of special congenic strains offsets these limitations. One can immediately backcross the appropriate CSS strain to host strain A, collect recombinant progeny, test whether more than one QTL accounts for the trait difference on the substituted chromosome and localize each QTL with considerable map resolution.

CSS panels enable the genetic dissection of any quantitative trait by phenotyping progeny from each of the strains without additional genotyping or generating crosses. Indeed, many phenotypes can be measured in parallel. By being able to proceed directly to fine-structure mapping, positional cloning becomes more tractable, especially with the availability of the DNA sequence of the mouse and rat genomes in the near future.

### Constructing CSSs

The catch, of course, is that a CSS panel must first be generated for the strain combination of interest. For the mouse, a total of 21 strains must be constructed through a 'marker-assisted' breeding program of roughly 2–3 years in duration. To create a CSS for chromosome *i*, one starts with (A×B)F1 progeny and performs successive backcrosses to strain A (Fig. 2). At each generation, an offspring that has a non-recombinant chromosome *i* from strain B is identified and used as the parent for the next generation. Genotyping the offspring for a set of genetic markers spanning the length of chromosome *i* reveals those that are heterozygous across the entire chromosome (that is, heterosomic).

The genetic markers can be chosen from the dense map of more than 6,000 markers available for mouse<sup>5</sup> and more than 5,500 available for the rat<sup>6</sup>. At each successive backcross generation, one continues to select from progeny that are heterosomic for chromosome *i*. Homozygosity elsewhere in the genome is rapidly achieved, as the proportion of unlinked loci remaining heterozygous is halved with each backcross. After a sufficient



**Fig. 1** Hypothetical composition of inbred strains and their derivatives. **a**, Two hypothetical genetically defined inbred strains, designated A and B, composed of nine chromosomes in the haploid genome. **b**, Two congenic strains, designated A.B-x and B.A-y, indicating that gene *x* has been transferred from strain B to strain A, and that gene *y* has been transferred from strain A to strain B, respectively. **c**, Recombinant inbred strains (RISs) made by intercrossing the progenitor inbred strains A and B, then inbreeding the progeny of intercrossed hybrid mice. The chromosome maps show the segments that are derived from progenitor strains A and B; note that the segments derived from strain A and strain B occur in an approximately 1:1 ratio. **d**, Recombinant congenic strains (RCSs) made by crossing (A×B)F1 hybrid mice to progenitor inbred strain A, followed by inbreeding the progeny of this backcross. Strain A-derived segments are more frequent and longer than strain B-derived segments, as evidence of the additional backcross to strain A. **e**, The construction of chromosome substitution strains (CSSs), A.B-Chr(*i*) and B.A-Chr(*j*) and B. A-y, whereby each strain has a single chromosome from the donor strain substituting for the corresponding chromosome in the host strain.

number of backcross generations, heterosomic mice are intercrossed and progeny that are homozygous for the substituted chromosome are selected and propagated as a strain with repeated brother-sister matings. The number of backcross generations traditionally used for constructing mouse congenic strains is ten, after which the expected proportion of the donor genome remaining at unlinked loci is about 0.002 ( $=0.5^9$ ). The resulting CSS thus carries an intact chromosome substituted from the donor strain (at least within the bounds of the most distal genetic markers used for the introgression) and few, if any, residual, unlinked chromosome segments from the donor strain.

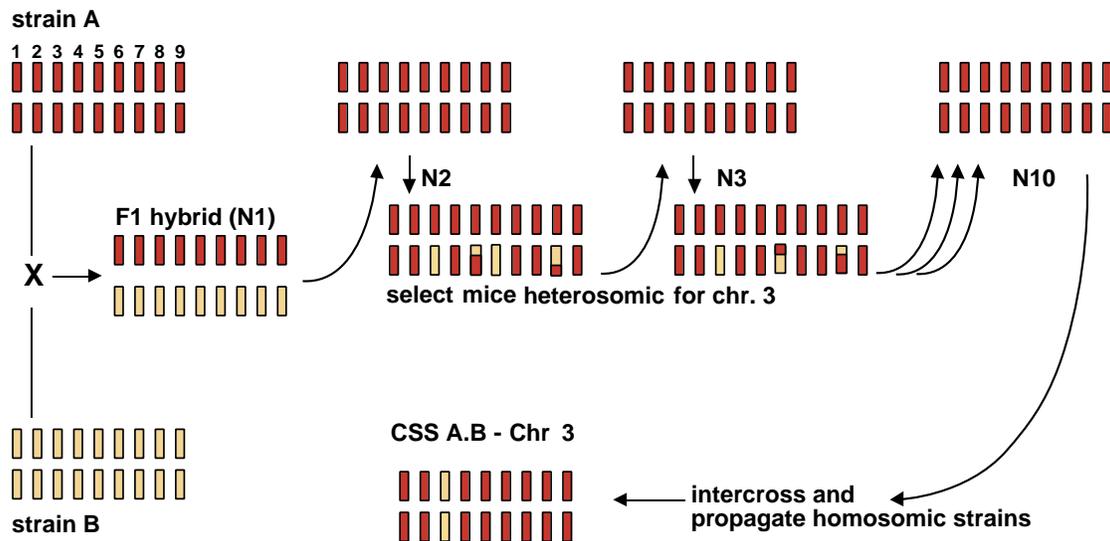
Obtaining a suitable heterosomic animal at each generation is simple. Assuming no crossover interference (that is, crossovers occurring independently) on a chromosome of length  $d$  centimorgans, the expected proportion of heterosomic progeny would be  $1/2e^{-d/100}$ , which ranges from 16% for long chromosomes to 35% for short chromosomes in the mouse. Crossovers are not completely independent due to crossover interference; therefore the actual proportion of progeny with a non-recombinant chromosome is marginally lower than expected. In a breeding program to create a complete B6.A-Chr(i) CSS panel, heterosomic progeny occurred at frequencies ranging from 7% for chromosome 2 to 41% for chromosome 15, with an overall average frequency of 20% (659 of 3,254 backcross progeny; Table 1).

A breeding program to generate a CSS panel can be improved by using genetic markers to follow the segregation of chromosomes from the donor strain. At each generation, progeny that have lost more than 50% of the remaining donor genome can be selected, a process called “speed congenics”<sup>7-9</sup>. Such marker-assisted breeding can potentially decrease the required number of backcross generations to six or seven. In addition, one can show that no donor alleles remain at unlinked loci—in which case the expected proportion of the donor genome is smaller than the 0.002 expected under random mendelian segregation. With a

sufficiently dense map, the presence of donor segments remaining as a result of double crossovers can be excluded. Although the possibility of donor segments being retained owing to crossovers beyond the most distal markers screened or gene conversion events cannot be eliminated, the total contribution of such events should be small.

The substituted chromosome should thus account for any phenotypic differences between the CSS and the host strain in almost all cases. Evidence of this is provided by showing that the trait difference segregates with the chromosome in the first round of fine-structure mapping.

We are constructing a CSS panel with C57BL/6J as the host strain and A/J as the donor strain. The numerous polygenic trait differences between A/J and C57BL/6J (<http://www.informatics.jax.org>) make these an ideal combination of inbred strains to evaluate the usefulness of CSSs. The breeding program uses marker-assisted selection both to retain the desired chromosome and to monitor and select against donor contributions to other regions (speed-congenic breeding). All 21 strains have remained viable and continue to yield heterosomic progeny at the desired frequency. With the most recent generation of breeding, the strains have now produced progeny that have retained the desired chromosome and lost all detectable donor strain contributions in unlinked regions. Several generations are required to produce homozygous lines, after which the strains will be expanded and made available to the research community. We have also begun construction of additional CSS panels, including an A.B6 panel as well as B6.129 and 129.B6 panels. We encourage others to create CSS panels for strain combinations that provide important polygenic trait models. It should be noted that some chromosomes cannot be successfully substituted in a strain combination involving C57BL/6J and an inbred strain derived from *Mus spretus* (J.-L. Guénet, pers. comm.). It remains to be determined whether all chromosomes can be successfully substituted in other strain combinations.



**Fig. 2** Strategy for constructing CSSs. The first step requires making hybrids between strains A and B, which are then backcrossed to host strain A. Progeny with a non-recombinant chromosome derived from donor strain B, in this case chromosome 3, are identified in this and subsequent backcrosses. These mice are backcrossed to host strain A at each generation. At the tenth backcross generation (N11, counting the F1 hybrids as the N1 generation), males and females with the non-recombinant chromosome derived from donor strain B are intercrossed. Progeny of this intercross that are homozygous (homozygous) for chromosome 3 are used to propagate the homozygous strain. Special considerations apply to the X and Y chromosomes. A Y chromosome CSS is made by selecting males for backcrossing to the host strain at each backcross generation<sup>16-18</sup>. To ensure that the Y chromosome is not unintentionally transferred from the donor to the host strain, a heterosomic carrier female is used for at least one backcross. Similarly, to eliminate the X chromosome derived from the donor strain, at least one backcross requires a male that is heterosomic for the desired autosome and has an intact X chromosome from the host strain.

**Table 1 • Frequency of non-recombinant chromosomes in backcrosses**

Host	Donor	Chr	Length (cM) <sup>a</sup>	Total no. of mice	No NR <sup>b</sup>	Per cent NR	Expected
Strain name:	129.MOLF-Chr19						
129/Sv	MOLF/Ei	19	47.0	233	47	20%	31%
Strain name:	B6.A-Chr(i)						
C57BL/6J	A/J	1	112.5	158	26	16%	16%
		2	96.2	215	15	7%	19%
		3	64.5	140	25	18%	26%
		4	82.0	205	38	19%	22%
		5	82.0	212	36	17%	22%
		6	64.5	165	31	19%	26%
		7	67.8	230	34	15%	25%
		8	71.0	142	24	17%	25%
		9	64.4	218	35	16%	26%
		10	75.4	136	31	23%	24%
		11	80.9	148	34	23%	22%
		12	59.0	121	32	26%	28%
		13	57.9	162	41	25%	28%
		14	67.7	147	32	22%	25%
		15	60.1	123	50	41%	27%
		16	51.4	122	28	23%	30%
		17	49.2	178	36	20%	31%
		18	37.1	218	57	26%	35%
		19	57.9	143	32	22%	28%
		X	57.9	71	22	31%	28%

<sup>a</sup>Source of genetic lengths is available ([http://carbon.wi.mit.edu:8000/ftp/distribution/mouse\\_sslp\\_releases/may99/05-99](http://carbon.wi.mit.edu:8000/ftp/distribution/mouse_sslp_releases/may99/05-99)). <sup>b</sup>NR, non-recombinant chromosome.

### Advantages of CSSs

We consider features of CSSs that are useful for dissecting complex traits and offer advantages over other methods.

**Partitioning of the genome.** CSSs offer advantages over RISs and RCSs in that the contribution of the donor genome is partitioned into individual chromosomes. As a result, the pattern of phenotypic differences between CSSs and the host strain reveals the chromosomal location of QTLs. By contrast, RISs and RCSs contain random subsets of the progenitor genomes and, because the subsets are overlapping rather than disjoint, the chromosomal location of a QTL cannot be directly inferred from the phenotypes. Indeed, typical RIS and RCS panels can be used only for direct mapping of monogenic traits<sup>10</sup>. To study polygenic traits, one must perform additional crosses involving individual RISs or RCSs.

**Replication.** Crosses often yield suggestive evidence for QTLs that fall short of being statistically significant. Following up such observations requires generating, phenotyping and genotyping a second, larger cross. In contrast, suggestive evidence of a phenotypic difference in a CSS can be established in a limited number of animals and then immediately confirmed by scoring additional animals in that strain.

**Tests of dominance.** A comparison of the phenotype of a homosomic A.B-Chr(i) CSS with the heterosomic (A.B-Chr(i)×A)F1 hybrid allows a direct test of dominance, without the complicating effects of other segregating QTLs on a heterogeneous background.

**Detecting a QTL with an additive effect.** The power to detect a QTL contributing to an additive trait depends on the proportion of the overall phenotypic variance influenced by the locus<sup>3</sup>. Segregation of unlinked QTLs in linkage crosses causes 'phenotypic noise' that can obscure the phenotypic effect of individual QTLs. In contrast, such noise is eliminated by comparison of a CSS with its host strain. As a result, QTLs with weaker phenotypic effects can be detected in fewer progeny by comparing a given CSS with its host strain than by analysing a segregating cross<sup>11</sup>.

**Studying epistasis.** The analysis of panels of complementary CSSs, such as A.B-Chr(i) and B.A-Chr(i), is a novel and powerful

method for discovering epistasis in polygenic traits. Suppose that a phenotype depends on the presence of the A allele at each of three loci on different chromosomes (i, j and k). Genetically mapping these factors in an ((A×B)F1)×(A×B)F1 intercross would be difficult, because only 1 of 64 progeny would have the phenotype, assuming that the trait is fully recessive. The principal clue that the trait is epistatic rather than additive is absence of the trait in the A.B-Chr(i), A.B-Chr(j) and A.B-Chr(k) CSSs, as well as in the reciprocal CSSs B.A-Chr(i), B.A-Chr(j) and B.A-Chr(k). Absence of the phenotype from the A.B-Chr(i), A.B-Chr(j) and A.B-Chr(k) CSSs indicates that an essential allele is located on each of these three chromosomes; absence of the phenotype in the reciprocal B.A-Chr(i), B.A-Chr(j) and B.A-Chr(k) shows that these chromosomes are required together to make the trait. These contrasting effects distinguish additive and epistatic polygenic traits in reciprocal panels of CSSs.

**Studying individual loci.** CSSs are ideal starting points for studying individual QTLs, such as fine-structure mapping in two-generation crosses or constructing congenic strains carrying a segment of the substituted chromosome. Construction of such congenics is rapid, requiring only two backcross generations to obtain a chromosome with crossovers flanking the critical region followed by intercrossing to homozygose the chromosome segment of the region.

**CSSs as resources for future studies.** The outcome of most studies with linkage crosses, including advanced intercross lines<sup>12</sup> (AILs) and heterogeneous stocks<sup>13</sup> (HSs), is information about the chromosomal location of QTLs. Specialized strains must then be constructed to pursue genetic mapping and functional studies. By contrast, CSSs not only enable the mapping of QTLs but can also be used immediately for subsequent genetic and functional studies<sup>11</sup>.

### Proof of concept

Two of us (A.M. and J.H.N.) recently reported<sup>11</sup> the construction of a CSS to map susceptibility loci in the inbred 129/Sv strain (currently called 129S3/SvImJ), which is a model of spontaneous testicular cancer<sup>11</sup>. A previous genome scan of nearly 300

tumour-bearing and tumour-free progeny of intercrosses and backcrosses of the 129/Sv and MOLF/Ei inbred strains reported only weak evidence for linkage of susceptibility loci, with the best test scores for chromosome 19 (ref. 14). Rather than test hundreds of additional mice with the hope of attaining statistical significance, a CSS was constructed for chromosome 19 (ref. 11). A sample of 28 homosomic mice was sufficient to demonstrate evidence for a susceptibility locus on chromosome 19. Thus, a weak effect in a large, segregating cross was converted to a stronger effect in fewer progeny in a CSS.

CSSs offer a powerful complement to existing approaches for studying polygenic traits—such as crosses, RISs and RCSs. The main drawback is that one must construct a CSS panel for each strain combination of interest. Once this is achieved, however, the use of CSSs can accelerate the genetic analysis of phenotypic traits

that differ between inbred strains. With improvements in genotyping technologies<sup>15</sup>, the construction of additional CSSs to study specific phenotypes is likely to become increasingly feasible.

#### Acknowledgements

J.B.S. was supported by NHGRI grant F32 HG00195. This work was supported by NIH grant RR12305 and CA75056 to J.H.N., a grant from the Keck Foundation to the Department of Genetics, Case Western Reserve University School of Medicine, a grant from the Howard Hughes Medical Institute to Case Western Reserve University School of Medicine and grants from the NIH to E.S.L.

Received 1 December 1999; accepted 27 January 2000.

1. Bedell, M.A., Jenkins, N.A. & Copeland, N.G. Mouse models of human disease. Part I. Techniques and resources for genetic analysis in mice. *Genes Dev.* **11**, 1–10 (1997).
2. Bedell, M.A., Jenkins, N.A. & Copeland, N.G. Mouse models of human disease. Part II. Recent progress and future directions. *Genes Dev.* **11**, 11–43 (1997).
3. Lander, E.S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
4. Snell, G.D. Methods for the study of histocompatibility genes. *J. Genet.* **49**, 87–108 (1948).
5. Dietrich, W.F. *et al.* A comprehensive genetic map of the mouse genome. *Nature* **380**, 149–152 (1996).
6. Steen, R.G. *et al.* A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res.* **9**, 1–8 (1999).
7. Gould, K.A., Dietrich, W.F., Borenstein, N., Lander, E.S. & Dove, W.F. Mom1 is a semi-dominant modifier of intestinal adenoma size and multiplicity in Min/+ mice. *Genetics* **144**, 1769–1776 (1996).
8. Markel, P. *et al.* Theoretical and empirical issues for marker-assisted breeding of congenic strains. *Nature Genet.* **17**, 280–284 (1997).
9. Wakeland, E.K., Morel, L., Achey, K., Yui, M. & Longmate, J. Speed congenics: a classic technique in the fast lane (relatively speaking). *Immunol. Today* **18**, 472–477 (1997).
10. Darvasi, A. Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genet.* **18**, 19–24 (1998).
11. Matin, A., Collin, G.B., Asada, Y., Varnum, D. & Nadeau, J.H. Susceptibility to testicular germ cell tumors in 129.MOLF-Chr19 mice. *Nature Genet.* **23**, 237–240 (1999).
12. Darvasi, A. & Soller, M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207 (1995).
13. Talbot, C.J. *et al.* High-resolution mapping of quantitative trait loci in outbred mice. *Nature Genet.* **21**, 305–308 (1999).
14. Collin, G.B., Asada, Y., Varnum, D.S. & Nadeau, J.H. DNA pooling as a quick method for finding candidate linkages in multigenic trait analysis: an example involving susceptibility to germ cell tumors. *Mamm. Genome* **7**, 68–70 (1996).
15. Wang, D.G. *et al.* Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
16. Hudgins, C.C., Steinberg, R.T., Klinman, D.M., Reeves, M.J. & Steinberg, A.D. Studies of consomic mice bearing the Y chromosome of the BXS mouse. *J. Immunol.* **134**, 3849–3854 (1985).
17. Scribner, C.L. & Steinberg, A.D. The role of splenic colony-forming units in autoimmune disease. *Clin. Immunol. Immunopathol.* **49**, 133–142 (1988).
18. Waters, N.S. & Deneberg, V.H. Analysis of two measures of paw preference in a large population of inbred mice. *Behav. Brain Res.* **63**, 195–204 (1994).