

Methodology article

Open Access

A new set of BXD recombinant inbred lines from advanced intercross populations in mice

Jeremy L Peirce*^{†1,2}, Lu Lu^{†1}, Jing Gu, Lee M Silver² and Robert W Williams¹

Address: ¹Center for Neuroscience Department of Anatomy and Neurobiology University of Tennessee Health Science Center 855 Monroe Avenue Memphis, Tennessee 38163 USA and ²Louis Thomas Laboratory Department of Molecular Biology Washington Road Princeton University Princeton, NJ 08544 USA

Email: Jeremy L Peirce* - jlpeirce@princeton.edu; Lu Lu - lulu@nb.utmem.edu; Jing Gu - jgu@utmem.edu; Lee M Silver - lsilver@princeton.edu; Robert W Williams - rwilliam@nb.utmem.edu

* Corresponding author †Equal contributors

Published: 29 April 2004

Received: 08 January 2004

BMC Genetics 2004, 5:7

Accepted: 29 April 2004

This article is available from: <http://www.biomedcentral.com/1471-2156/5/7>

© 2004 Peirce et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Recombinant inbred (RI) strains are an important resource for mapping complex traits in many species. While large RI panels are available for *Arabidopsis*, maize, *C. elegans*, and *Drosophila*, mouse RI panels typically consist of fewer than 30 lines. This is a severe constraint on the power and precision of mapping efforts and greatly hampers analysis of epistatic interactions.

Results: In order to address these limitations and to provide the community with a more effective collaborative RI mapping panel we generated new BXD RI strains from two independent advanced intercrosses (AI) between C57BL/6J (B6) and DBA/2J (D2) progenitor strains. Progeny were intercrossed for 9 to 14 generations before initiating inbreeding, which is still ongoing for some strains. Since this AI base population is highly recombinant, the 46 advanced recombinant inbred (ARI) strains incorporate approximately twice as many recombinations as standard RI strains, a fraction of which are inevitably shared by descent. When combined with the existing BXD RI strains, the merged BXD strain set triples the number of previously available unique recombinations and quadruples the total number of recombinations in the BXD background.

Conclusion: The combined BXD strain set is the largest mouse RI mapping panel. It is a powerful tool for collaborative analysis of quantitative traits and gene function that will be especially useful to study variation in transcriptome and proteome data sets under multiple environments. Additional strains also extend the value of the extensive phenotypic characterization of the previously available strains. A final advantage of expanding the BXD strain set is that both progenitors have been sequenced, and approximately 1.8 million SNPs have been characterized. This provides unprecedented power in screening candidate genes and can reduce the effective length of QTL intervals. It also makes it possible to reverse standard mapping strategies and to explore downstream effects of known sequence variants.

Background

Recombinant inbred (RI) strains have been an important resource for investigation and genetic mapping of Mendelian and quantitative traits in the mouse over the past sev-

eral decades [1]. Conventional mouse RI strains are developed by crossing two inbred parental strains and repeatedly mating the resulting siblings for 20 generations or more to ensure that they are at least 99% inbred [2].

The resulting strains have a genome with an average 4-fold increase in recombination compared to a single generation genetic map [3,4]. RI strains are especially useful for mapping complex traits, since they create an immortalized mapping population that allows researchers to phenotype as many animals per genome as desired over extended periods of time. Multiple phenotypic data points per genome lower the effect of environmental noise. This facilitates more precise phenotypic estimates invaluable to mapping complex traits with low to moderate heritability, including such traits as CNS architecture [5], alcohol related phenotypes [6,7], basal locomotor activity [8], body weight, growth rate, litter size, and sex ratio [9].

The replicable nature of RI strain data is also useful in examining phenotypes using multiple behavioral, pharmacological, physiological, and biochemical techniques, all of which may yield valuable information about the identity of underlying genetic differences. In addition, RI strains are uniquely valuable in examining the interaction of genes with environments, a property which has greatly encouraged their use in the plant genetics community, but which has not yet been exploited by experimental mammalian geneticists.

Because the inbreeding necessary for RI generation takes 4–5 years, and exploitation of newly generated strains requires a relatively dense linkage map, the development of a novel strain panel requires considerably more effort for less immediate return than the phenotyping and sparse genotyping effort associated with even a reasonably large conventional intercross or backcross. This high initial barrier has kept the number of mouse RI strains low and increased the popularity of mapping strategies utilizing segregating crosses. The resulting lack of power has given RI strains a poor reputation among some mouse geneticists – a reputation that in turn provides an unfortunate disincentive to the creation of adequately powered strain sets. However, when lines can be inbred by selfing, as is the case with *Arabidopsis* [10], sunflower [11,12], beans [13], tomato [14], and maize [15-17], and/or cost of multi-generation crosses is lower as in *Drosophila melanogaster* [18] and *C. elegans* [19], large RI sets are common. In the case of maize, for instance, a set of approximately 1000 strains is currently available, and loci of small effect, in addition to epistatic interactions, can be readily detected and mapped [17].

The number of recombinations archived per strain, while considerably higher than the number available in members of an intercross or backcross, necessarily limits the usefulness of a small RI strain set for fine mapping of Mendelian and quantitative traits. Given a sufficiently dense linkage map, an ideal RI-like mapping resource

would include considerably more recombinations per strain as well as a larger number of strains. A higher density of recombination, especially in combination with an RI-based mapping paradigm like the RI intercross (RIX) mapping proposed by Threadgill and colleagues [20,21], might also reduce the number of strains necessary to achieve a given level of total recombination density, similarly reducing the expense associated with maintenance of a large RI colony. RIX mapping involves generating F1 crosses between RI strains. Since the parental genotypes are known and homozygous, the genotypes of the resulting offspring are fixed and easily ascertained for a given pair of strains.

Development of a high recombination density population useful for fine mapping of quantitative traits generally requires multiple generations during the course of which recombinations accumulate in the population. This is certainly the case for the advanced intercross line (AIL) approach [22,23]. Unfortunately, AILs, while a useful mapping tool, are not a stable resource that can be indefinitely used by the community. Having invested considerable time and energy in the creation of AIL populations generated by crossing B6 and D2, we decided to create derivative inbred lines in the hope that these lines would permanently archive a large number of the unique recombinations present in our AIL populations. The resulting lines, some of which are described here, are similar to conventional RI lines but offer noticeably higher recombination densities. In addition to facilitating new studies, these strains can be used to revise and extend results previously published in the BXD RI strain set.

Results

By inbreeding animals from two separate B6 × D2 AIL populations, we have developed a large set of BXD ARI lines, each with more recombinations than an equivalent BXD RI line. Fig. 1 outlines the breeding protocol, which is described more completely in the methods section. Fig. 2 shows the recombinations present on chromosome 15 in the currently genotyped subset of the ARI population. The complete set of genotypes for this subset [see additional file 1: genotypes.xls], in addition to a later generated, small number of paired genotypes for nearly the entire ARI strain set [see additional file 2: genopairs.xls], and a very recently generated less dense set of genotypes for nearly all strains [see additional file 3: 286newgenotypes.xls] accompany this paper. Additional genotypes will be made available at <http://www.nerwenet.org/papers/ari.html>. For simplicity, we have referred to the lines derived from the Princeton AIL as Group A and those derived from the University of Tennessee Health Science Center (UTHSC) AIL as Group B, regardless of the institution where the strains were actually inbred.

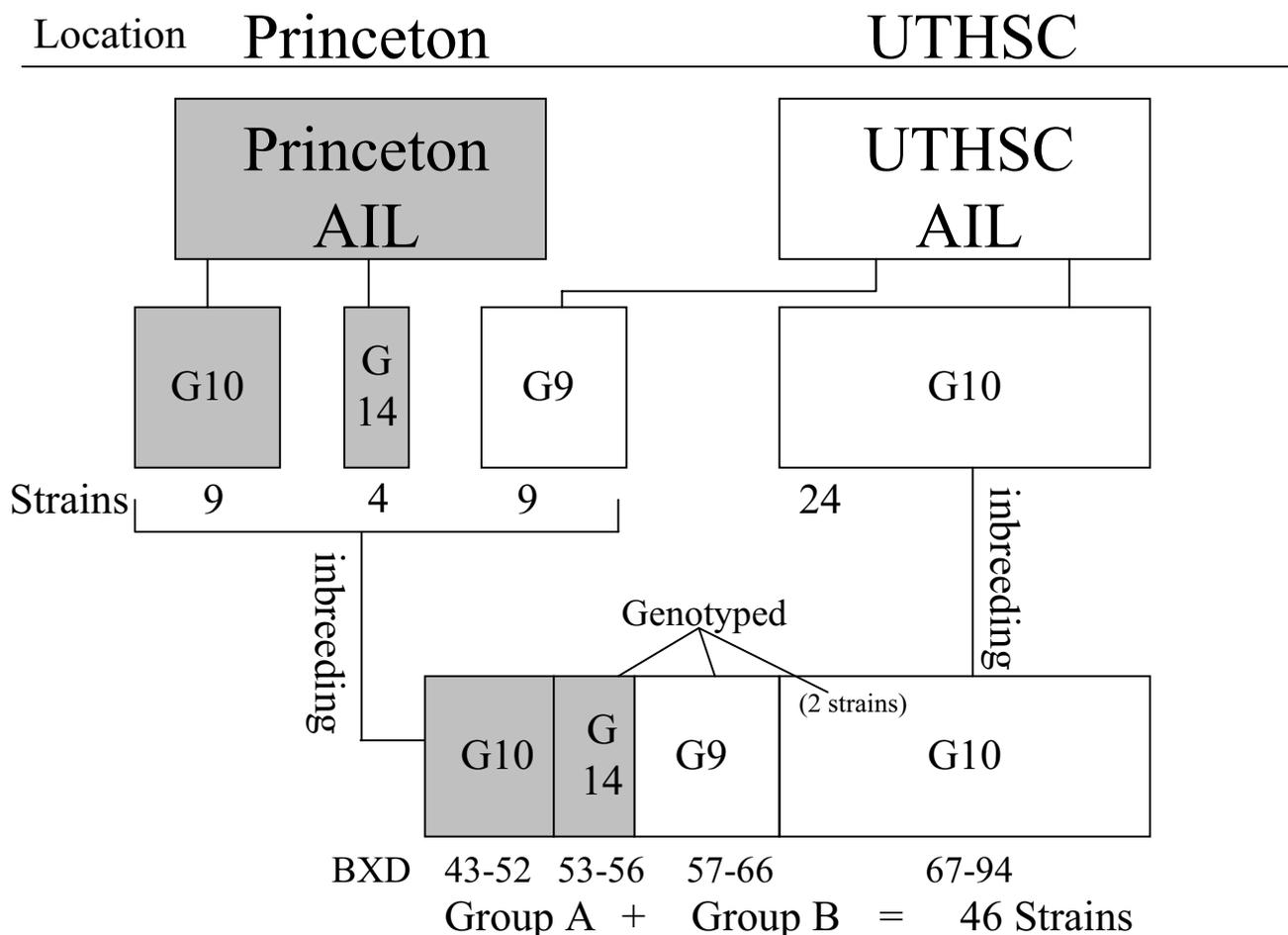


Figure 1
ARI breeding schematic. This figure shows the breeding history, genetic derivations, and genotyping status of the BXD ARI strain set. We commonly refer to animals derived from the BXD AIL generated at Princeton University as "Group A" and animals derived from the BXD AIL generated at the University of Tennessee Health Science Center as "Group B." Group A animals are shown in gray. Genotyping refers to our dense, 588 marker effort. Nearly all strains in both groups are genotyped at 268 loci across the genome.

Current status

The Jackson Laboratory currently has a total of 34 BXD RI strains available from live stock. We have added 46 BXD ARI strains, for a total of 80 BXD RI strains, many of which will be available by the time of publication. These strains more than double the number of BXD-based RI strains commonly accessible. Table 1 describes the strains, their current availability, and their genotyping status. 17 BXD ARI strains are fully genotyped and currently available. As of January 7, 2004, a total of 7 strains (6 available) are inbred at F20 or higher and all but 6 strains are inbred to at least F14 (92% of the genome fixed in both parents [2]). Table 1 and Fig. 1 outline the average status and his-

tory of the full strain set. In addition to the genotype data analyzed here [see additional files 1: genotypes.xls and 2: genopairs.xls], we have very recently completed a less dense (268 markers) set of genotypes for nearly all BXD ARI strains, which is available for the convenience of potential investigators [see additional file 3: 286newgenotypes.xls] but it is not analyzed further here.

Heterozygosity

The number of heterozygous intervals, as well as the total length of heterozygous regions is shown in Table 2. Our initial calculation of heterozygosity simply involved summing the length of the heterozygous regions in each strain

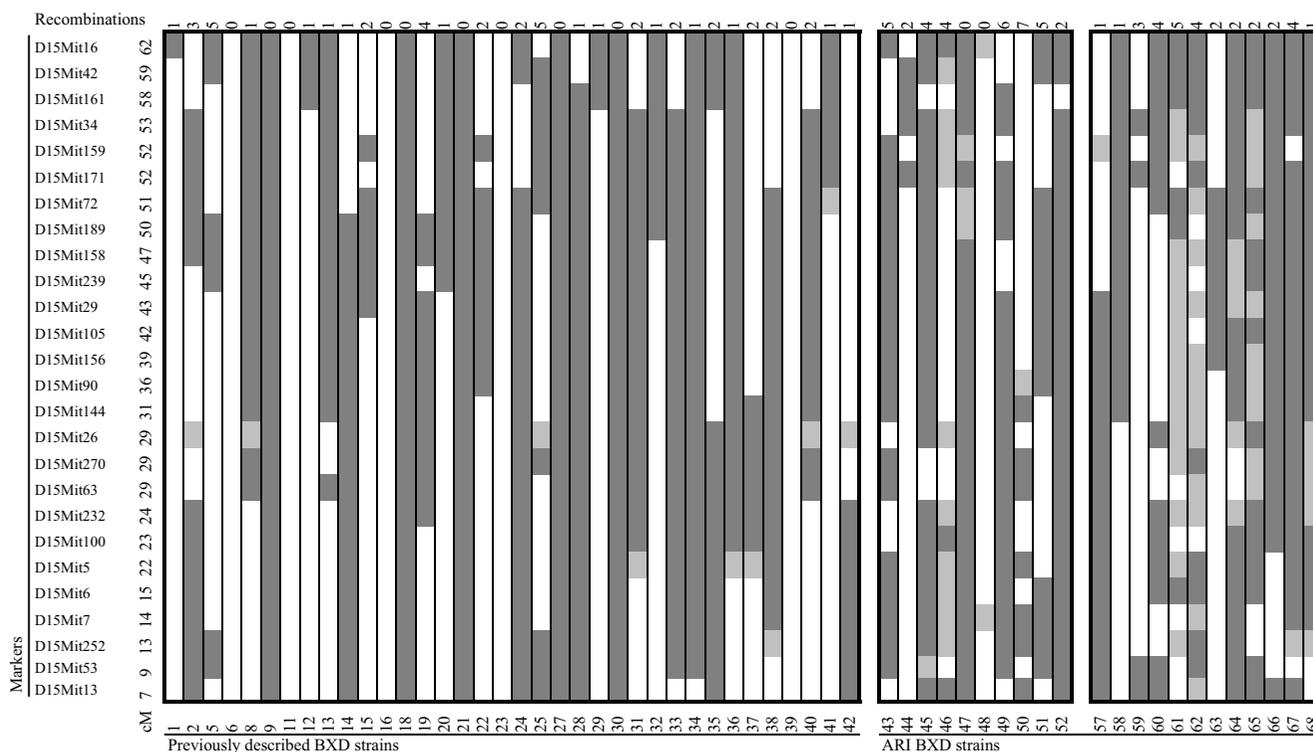


Figure 2
ARI chromosome 15. Chromosome 15 in 22 genotyped ARI strains and the 34 BXD RI strains. Dark grey regions are homozygous B6, white regions are homozygous DBA, and light grey regions are heterozygous. All BXD data are taken from [4] and are publicly accessible via <http://www.nervenet.org/papers/bxn.html>. Number of recombinations per strain is displayed below each strain. Where heterozygous regions were present, the number of recombinations was calculated based on the **smallest** number of recombinations that might ultimately be resolved from that pattern. For example BXD47 has 0 recombinations rather than 2. Even discounting additional recombinations from resolution of heterozygous regions, the ARI animals have a considerably larger number of recombinations, 3.0 per strain compared to 1.3 per strain in BXD animals for Chr 15. BXD37, BXD41, BXD47, and BXD58, are now extinct.

Table 1: ARI Strains, Current Inbreeding and Availability

Strains BXD	Parental AIL	AIL Gen.	Inbred at	Inbreeding Generation	Strains Genotyped	Generation Genotyped	Availability
43–52	Princeton	G10	Princeton	18	Yes	F9-F13	7/9 ^a
53–56	Princeton	G14	Princeton	14	No	N/A	Pending
57–66	UTHSC	G9	Princeton	16	Yes	F8-F10	8/9 ^a
67–68	UTHSC	G10	UTHSC	16	Yes	F8-F9	2/2
69–94	UTHSC	G10	UTHSC	15	Yes ^b	N/A	Pending

^aBXD47, BXD58, and BXD72 are now extinct. BXD71, BXD81, and BXD82 have no offspring and are likely to become extinct, so are not counted towards the total number of strains. ^bGenotyped only at a small set of closely spaced marker pairs and a moderately dense set more recently generated set of 268 markers

and dividing by the entire length of the genome. These fractions are reported as actual heterozygosity in Table 2 [see additional file 1: genotypes.xls], and compared with

expected heterozygosity as calculated by Green's method [24]. It is important to note that both actual and expected

Table 2: Heterozygosity and Genotype mis-assignment

Strain BXD	Inbreeding Generation	Heterozygous ^a Intervals	Heterozygous Region (cM)	Expected % Heterozygotes	Actual % Heterozygotes (Individual)	Actual % Heterozygotes (Population)
43	13	6	37	6%	2%	3%
44	11	12	61	10%	4%	6%
45	12	13	73	8%	5%	8%
46	9	46	232	14%	15%	24%
47	13	43	226	6%	15%	24%
48	13	72	442	6%	29%	45%
49	11	8	26	10%	2%	3%
50	11	20	101	10%	7%	11%
51	12	8	44	8%	3%	5%
52	11	14	103	10%	7%	11%
57	11	10	61	10%	4%	6%
58	9	28	169	14%	11%	17%
59	10	5	31	12%	2%	3%
60	10	30	188	12%	13%	20%
61	9	33	148	14%	10%	16%
62	9	24	119	14%	8%	12%
63	8	22	123	16%	8%	12%
64	8	24	178	16%	12%	18%
65	8	69	370	16%	25%	38%
66	8	18	130	16%	9%	14%
67	9	14	116.5	14%	8%	12%
68	8	21	124	16%	8%	12%

^aHeterozygosity here refers to individual heterozygosity except as specified.

heterozygosity here refer to individual heterozygosity rather than population heterozygosity.

We also assessed heterozygosity in our second, smaller set of genotypes [see additional file 2: genopairs.xls]. We chose a subset of 13 markers previously genotyped in 16 Group A and Group B strains. Overall heterozygosity in this subset of loci decreased by the time the second set of genotypes were generated, from a directly measured individual level of 12% in the original genotypes to 7.5% (averaging individual male and female results).

Increased resolution in ARIs

The relationship between F2 recombination rates (per meiosis rates) and observed recombination rates for recombinant inbred strains derived from advanced intercross-like progenitors at various generations has been described [25]. For densely spaced markers (less than 1 cM), we expect our ARIs to achieve 2.1 (G9-based), 2.3 (G10-based), and 2.8 (G14-based) times the resolution of a standard RI. This agrees well with our results. For the 113 markers with spacing between 1.0 and 1.5 cM (averaging UTHSC and MIT positions), for instance, the average spacing was 1.14 cM and the observed average recombination fraction was 9.3 cM (G9) or 9.0 cM (G10). This agrees quite well with the calculated values of 8.7 cM

(G9) or 9.2 cM (G10). Other intervals involving average spacing below 3 cM behaved similarly (data not shown).

Total recombinations

The average number of recombinations for currently genotyped strains at the current 588 markers/strain genotyping resolution is 77 recombinations/strain (78 in genotyped Group A and 75 in genotyped group B). Number of recombinations resulting from transitions between homozygous genotypes as well as an estimate of recombinations resulting from remaining heterozygous patches and an estimate of the number of recombinations detected at the current BXD RI strain set genotyping resolution is given in Table 3. The latter estimate takes into account the fact that approximately 11% more recombinations are detected at the current 936 marker resolution of the BXD RI set than are detected with an evenly spaced set of 588 markers in the same set. Applying this result to the ARI strain sets results in an estimated average of 85 recombinations/strain once these strains are fully inbred and genotyped at a similarly high marker resolution.

Unique recombinations

¹Determining the number of unique recombinations per strain is rather difficult given the complex breeding history of the AIL members serving as ARI progenitors. ²Table

Table 3: Estimated total recombinations per ARI line

Strain ^a	Homozygous Recombinations ^b	1/2 Heterozygous Recombinations ^b	Total Recombinations	Proportional Estimate at 936 Markers/Strain
43	53	9.5	62.5	69.2
44	42	26.5	68.5	75.8
45	73	6	79.0	87.4
46	42	28	70.0	77.5
47	44	32.5	76.5	84.7
48	56	23.5	79.5	88.0
49	53	20	73.0	80.8
50	39	26	65.0	72.0
51	37	68	105.0	116.2
52	56	18.5	74.5	82.5
57	64	5.5	69.5	76.9
58	53	10	63.0	69.7
59	73	12.5	85.5	94.6
60	53	47	100.0	110.7
61	53	42	95.0	105.2
62	22	65	87.0	96.3
63	67	7.5	74.5	82.5
64	69	18	87.0	96.3
65	44	8	52.0	57.6
66	56	13	69.0	76.4
67	72	13.5	85.5	94.6
68	55	21.5	76.5	84.7

^aBXD47 and BXDA58 are extinct. ^bHomozygous recombinations refers to the number of transitions between homozygous markers of differing genotype. Heterozygous recombinations refer to the number of transitions between a homozygous marker of either genotype and a heterozygous marker. Estimated recombinations at 936 markers/strain estimates recombination density at this higher resolution for comparison with the BXD RI strain set.

Table 4: Estimated total and unique recombinations per genotyped ARI strain

	G9 (12 strains)		G10 (10 strains)	
	Recombinations per Strain	Density ^a	Recombinations per Strain	Density ^a
Conservative estimate, adding missed inbreeding-derived recombinations only	59.7	1.4	61	1.5
Proportional estimate, including missed AIL and inbreeding-derived recombinations	67.3	1.6	69.1	1.7
Experimental estimate using closely spaced pairs of loci	77.6	1.9	86.6	2.1
Estimated total recombinations at 936 markers per strain	84.3	2.0	86.6	2.1

^aFold change in strain recombination density compared to the currently available BXD RI strains. (BXD RI density equals 1.0)

4 summarizes the methods used to estimate the number of additional recombinations. All ARI strains will have recombinations derived from the AIL progenitors and from inbreeding. ³The total number of recombinations

present in the AIL populations used as progenitors is constant, and an increasing number of samples drawn from that constant pool will inevitably result in drawing the same recombination multiple times as the number of

strains increases. ⁴Because of this, the number of unique recombinations archived per ARI line will necessarily depend on the number of strains analyzed, especially as sampling of recombinations present in the AIL is saturated. ⁵In all analyses of unique recombinations we have considered only strains that have been fully genotyped for a common set of markers.

The conservative estimate of the number of unique recombinations was determined as described in the methods section. This estimate assumes that only two recombinations, each representing a transition between genotypes in a given direction (B6 to D2 or the reverse), can be unique, regardless of the number of lines considered. For the current marker density, this method results in an estimate of approximately 48 recombinations/line in genotyped animals. Adjusting for additional inbreeding-derived recombinations missed by the lower density marker set by adding 12 recombinations per line gives an estimate of 60 recombinations per line. This estimate is quite conservative, and can be taken as a minimum estimate of the number of recombinations in each ARI strain in the initial strain sets inbred at Princeton. This estimate can be directly compared to the number of recombinations in the BXD RI strain set genotyped with 936 markers [4] (41.4 recombinations per line) since the method corrects for both resolution and reductions in unique recombinations detected using the minimal estimation method at current ARI genotyping resolution.

The proportional estimate considering both additional inbreeding derived recombinations and additional AIL-derived recombinations is more likely to be a reasonable estimate of the actual number of unique recombinations since it includes an estimate of the additional AIL-derived recombinations missed by the minimal estimate. Since an average of 29.2 recombinations are detected by application of the minimal estimation method to samples of the lower density genotyped BXD RI strain set, compared with the full estimate of 41.4 recombinations per line in this strain set, the ratio of full resolution to low resolution detected recombinations is 1.42:1. Application of this ratio to the minimum estimation method gives the proportional estimate of 67.3 and 69.1 recombinations per line in the Group A and Group B strain sets, respectively. This is relatively close to the estimate of total recombinations, suggesting that only 13% and 11%, respectively, of recombinations are likely to be duplicates in the current strain sets.

The proportional estimate is still somewhat lower than the estimate generated using a set of pairs of closely spaced loci [see additional file 2: genopairs.xls]. Even in these extremely small intervals, which averaged 0.48 ± 0.12 Mb, some intervals contained transitions in more

than one direction, which must be independent. Our first set of analyses using these pairs of loci examined unique recombinations in the subset of currently genotyped strains. We were able to analyze 9 of the 10 currently genotyped Group A animals, which have a total of 13 recombinations in our intervals, 11 of which were unfixed transitions, and no non-unique recombinations. In two cases, a single interval contained unfixed transitions in both directions, but otherwise there was only one recombination per interval. The resulting estimate of 0% redundant recombinations is clearly optimistic since it is based on one fewer strain than the currently genotyped set and since we know that there are common recombinations in the starting population. However, it does serve to indicate that there is probably very little redundancy and suggests that there will also be relatively little redundancy between these strains and the additional 4 Group A strains.

Estimating unique recombinations in the currently genotyped Group B strains was more difficult since we were only able to analyze 9 of the 12 currently genotyped strains directly. In order to estimate the fraction of unique recombinations in the full set of 12 strains we randomly selected 3 strains from the other available Group B strains. The average result from 500 such partially random sets was 92% (77.6) unique recombinations per strain.

In addition to estimating unique recombinations in the currently genotyped strains, we are interested in understanding the saturation of unique recombinations in the larger set of ARIs, most of which belong to Group B. We currently have genotypes for a total of 35 Group B strains. In this strain set, a minimum of 65% of the recombinations will be unique, accounting for the fraction of heterozygotes likely to resolve as recombinations. Fig. 3 shows the relationship between population size and unique recombination frequency for a variety of intermediate numbers of ARI lines.

In our ARI population we estimate that we will have, on completion of inbreeding, archived a total of approximately 1100 recombinations in the 13 Group A strains and 2800 (at least 1700 unique) recombinations in the 33 Group B strains. Combined with existing RI strains, these strains provide a total of approximately 5300 recombinations – a 3.8-fold increase in the total number of recombinations available in this strain set. We estimated unique contribution for the Group A animals from the simulation of 13 member Group B strain sets (86%), which should be quite conservative given the larger number of animals at each generation and the fact that 4 strains accumulated 4–5 additional generations worth of recombinations in the AIL before inbreeding. Therefore the Group A strains should contribute at least 1000 unique recombinations, for a total of 2700 unique recombinations archived in 46

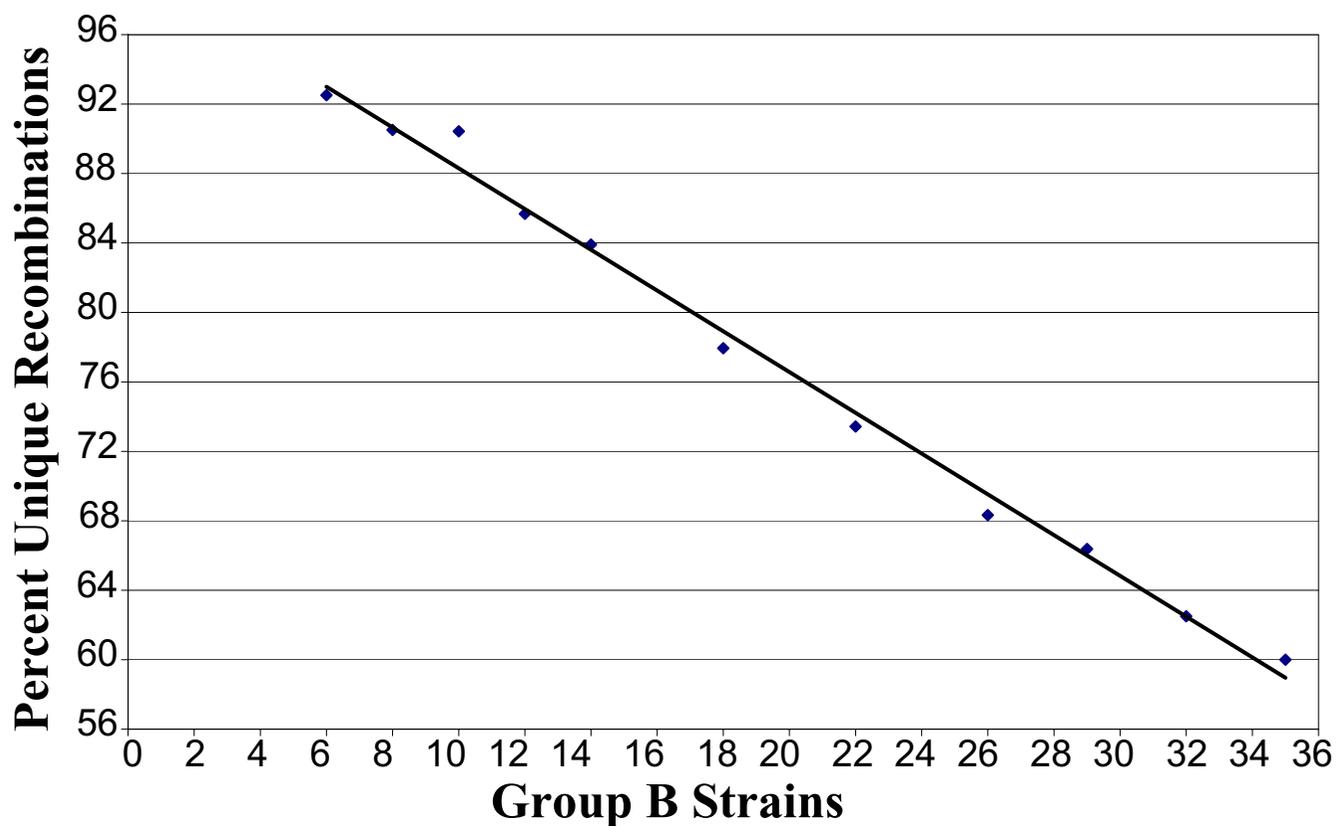


Figure 3
Population size and unique recombinations in the UTHSC ARI lines. Fraction of unique recombinations as a function of number of strains sampled in the set of UTHSC-derived strains. We derived each data point in this highly conservative estimate by randomly sampling the indicated number of strains from the UTHSC-derived strain set, using 100 samples per data point. Up to one directional transition from B6 to DBA and the reverse was scored for each marker interval and the total was divided by the number of intervals with more than 0 recombinations. Heterozygous regions were handled as described for Fig. 2. Fraction of unique recombinations depends on a number of variables including AIL population size, so other crosses may differ.

strains. When combined with the approximately 1400 recombinations available in the 34 BXD RI strains we have a total of approximately 4100 available unique recombinations, approximately 3-fold the previously available number.

Discussion

We have generated a novel set of 46 RI lines based on progenitors from two B6 × D2 advanced intercrosses. These lines have considerably more recombinations than the BXD RI set of lines, archiving an estimated 2.1-fold more recombinations per line. Over the subset of 22 genotyped ARI lines, the ARI strains archive a minimum of 1.4-fold and an estimated 1.6-fold to 1.9-fold increase in unique recombinations per line. Even using the more conservative estimate and considering only the 20 well-genotyped

ARI lines, we have at least doubled the number of available recombinations in the BXD RI background. The additional 26 strains will approximately triple the number of unique and nearly quadruple the number of total recombinations available for analysis in this background.

BXD ARI advantages

The advantages and issues involved in utilizing RI strains for mapping have previously been extensively discussed [26-28]. The ARI strains retain many of these characteristics. The chief advantages of ARI strains versus conventional RI strains, however, are greater potential mapping precision and lower cost per archived total and unique recombination, both of which stem from the higher recombination density. We estimate that the set of 46 ARI strains will ultimately provide a number of unique,

characterized recombinations equivalent to at least 90 conventional F2-derived BXD RI strains – a considerable saving in facility space and costs per archived recombination. The increased recombination density in these strains is also ideal for mapping techniques such as RIX mapping, providing better mapping resolution with a greater fraction of useful strains. We strongly recommend this method as a means of developing a long-term resource from any currently existing advanced intercross lines.

In addition, extending the number of available BXD RI strains allows researchers to take advantage of the extensive work that has already been done using these strains and their parental lines. Since the BXD RI strains were also the largest previously available mouse RI population, they have been extensively phenotyped. At least 626 phenotypes, including a large number of alcohol-related phenotypes [6,7,29-32] and a wide variety of observations from methamphetamine response [33] to stem cell number [34], have been studied in the currently available BXD RI strains. These data are easily accessible via the published phenotypes database and QTL analysis tools that are part of the WebQTL project [35-37]. The availability, also via WebQTL, of a large set of forebrain gene expression phenotypes derived from Affymetrix expression studies of the previously available BXD RI strains further increases the value of this extended RI set. Previously, the 34 existing BXD RI strains had the power to reliably detect (power = 0.80, $p < 0.0001$) only QTLs accounting for 47% of between strain variance [38]! These additional strains make it possible to reliably detect QTLs accounting for only 24% of genetic variance. With a second, independent population for statistical confirmation (power = 0.80, $p < 0.05$), the additional strains allow reliable detection of QTLs accounting for as little as 9% of genetic variance. Having more strains available will also give us sufficient power to characterize some simple epistatic interactions for loci with relatively large effects.

Additionally, sequence data is available for both parental strains and, by imputation, for all well characterized BXD RI strains. B6 sequence data [39] is publicly available and D2 sequence data is available via Celera Discovery System [40] subscription. Since a QTL should typically relate to a DNA polymorphism between the parental strains, a list of all such polymorphisms in a QTL region is a valuable tool.

ARI disadvantages

The chief disadvantage of ARI strains compared to RI strains is the relatively complicated relationship between ARI strains and the inability to assume that ARI recombinations are unique, which is particularly important in fine mapping efforts. Another major disadvantage is that ARI strains are more difficult and time consuming to create than conventional RI strains, since they require a well

developed AIL cross. Also, a given AIL can only be profitably used to create a limited number of ARI strains.

In addition, the AIL population from which the ARI progenitors were drawn is not in Hardy-Weinberg equilibrium. While the overall frequency of B6 and DBA alleles in the AILs is similar (55% and 53% B6 alleles for Group B and Group A, respectively), the frequency of alleles at a given locus varies widely, affecting the likely composition of the resulting ARI population on a per locus basis. For instance, the genotypes for the Group A proximal chromosome 4 and the majority of the Group B chromosome 10 are almost entirely B6 derived.

In our analysis of heterozygosity and evaluation of unique recombinations (using the set of pairs of very closely spaced markers), there was a larger than expected number of cases where one parent had a B, H or D, H genotype while the other was B, B or D, D. These cases represent a considerable fraction of recombinations in the ARI population, and are somewhat surprising given the inbreeding of most of these strains, suggesting that either heterozygotes have some selective advantage or that a small number of genotyping errors have occurred. Unfortunately in most cases there are not flanking markers close enough to meaningfully check these data and distinguish between these possibilities. These cases increase heterozygosity of the population, decrease unique recombinations, and generally provide a conservative bias to these measures.

Early genotyping and heterozygosity

Full inbreeding (20 generations) of a mouse inbred line takes an average of four to five years, though the great majority of inbreeding is accomplished in the first half of that time. In order to gain several potential years of useful analysis and to make our strains available to the community more quickly, we genotyped a total of 22 strains relatively early in the inbreeding process. Naturally, there are a significant number of heterozygous regions still present, and, in fact, there was considerable variation in the number and size of these regions between lines, suggesting that some lines, BXD48 and BXD65 for example, may actually have experienced several generations of cousin-cousin, rather than brother-sister mating. Ultimately this will only serve to increase the number of recombinations in these strains, but proximally they have fewer defined recombinations.

Ultimately, we will re-genotype all strains after full inbreeding is achieved. Early genotypes greatly facilitate the current usefulness of the strains, but must be treated with caution. For instance, heterozygous regions in these strains should be treated as unknown regions, and researchers should be aware of potential mis-assignment

of homozygotes in a small number of cases. Likewise, caution should be exercised in comparing phenotypes between animals at intermediate stages of inbreeding and animals comprising the resulting fully inbred lines, though for highly polygenic traits this will be less important. An easy precaution is to take DNA samples from phenotyped animals and confirm genotypes at loci of interest via pooled genotyping if needed. For applications where a somewhat higher noise level is tolerable, early genotyping is a valuable means of accelerating the usefulness of RI-like lines by several years. We have, for instance, successfully used 20 ARI lines in a small QTL mapping study of alcohol preference (manuscript in preparation). There was some indication that a QTL was present for 4 of 8 previously observed QTLs [41] in this small set, a reasonable result given that some of the previously observed QTLs may be false and that we do not expect to reliably detect real QTLs of modest effect size with this limited number of strains.

Unique recombinations and saturation of the AIL-derived recombination pool

The ARI strains archive a large number of recombinations per line. However, making ARI strains based on AIL progenitors is not a fully extensible strategy for making strains with high recombination densities. Since there is a limited, constant pool of AIL-derived recombinations, only inbreeding-derived recombinations will be novel once sampling of the AIL-derived recombinations is saturated.

In the currently genotyped lines the saturation level is quite low – between 0% (Group A) and 8% (Group B). This is reasonable considering that the initial pool from which the ARI progenitors were drawn consisted of 90–100 animals in the case of Group A and 40–60 animals in the case of Group B. The degree of saturation is an important issue for other investigators considering creation of similar strains from pre-existing AILs, since eventually the process will yield returns of unique recombinations approaching F2-based RI strains. Group B includes a total of 33 strains, and can serve as a partial model for this decision, albeit an imperfect one because different AILs will be based on different family sizes, breeding schemes, and generations.

Since it is difficult, in the absence of extremely precise genotyping information, to determine which recombinations are unique, we developed several approaches to this problem. Ultimately, the true average number of unique recombinations present in the genotyped ARI lines will fall between the average determined using our conservative estimate (59) and the estimated total average number of recombinations (85). It is more likely, however, that the number of unique recombinations will resemble our proportionate or experimental estimates given the known

number of total recombinations. That is, the conservatively corrected estimate serves as a reasonable minimum number of recombinations/strain that can be expected for our current strain set, while our best guess at the actual number of recombinations is considerably higher.

In either case, it is clear we have not yet reached a point of seriously diminished return on the creation of new ARI lines. Investigators considering this approach can expect to generate at least 30–40 valuable strains from a single AIL population.

A similar heterogeneous stock based approach

A similar approach to the problem of archiving large numbers of recombinations per strain would be to use a heterogeneous stock (HS) [42] as a progenitor. This approach would have a recombination density likely to be superior to an AIL-based approach, especially for longstanding HS populations, and has the additional advantage and complication of incorporating chromosomal segments from multiple strains. Because of the incorporation of input from many strains, mapping with these strains is likely to be both more versatile and more complex. The limitations of this approach with respect to expense, time of initial establishment, determination and treatment of unique recombinations and eventual diminishing returns, are quite similar to ARI lines. Additionally, however, detection of rare alleles could be problematic.

A seemingly similar approach using a HS population has been taken by Bennett and colleagues (Bennett, personal communication), who created a large 76 strain RI population (LXS) from a pair of inbred strains (ILS, ISS) derived from a randomly mated HS population based on 8 progenitor strains. This HS population was used to select populations that differed with respect to long and short sleep time in response to a hypnotic dose of ethanol [42] and members of these selected populations were subsequently inbred. Because this effort started with two fully inbred strains and immediately commenced inbreeding, it is actually much more similar to an F2-based standard RI approach than to an AIL-based approach or the theoretical HS-based approach above. These RI strains will be extremely useful, especially in research on alcohol-related phenotypes.

Improving on the ARI model

Ideally, an RI-like mapping population should maintain a high density of archived, fully independent recombinations. One approach to generating such a population would be to start with 2^n F2 × F2 breeding cages, where $n/2$ is the desired genome expansion prior to inbreeding. Breeding would then proceed as illustrated in Fig. 4. Briefly, each of the F2 animals carries an independent set of recombinations. The 2^n initial crosses generate an

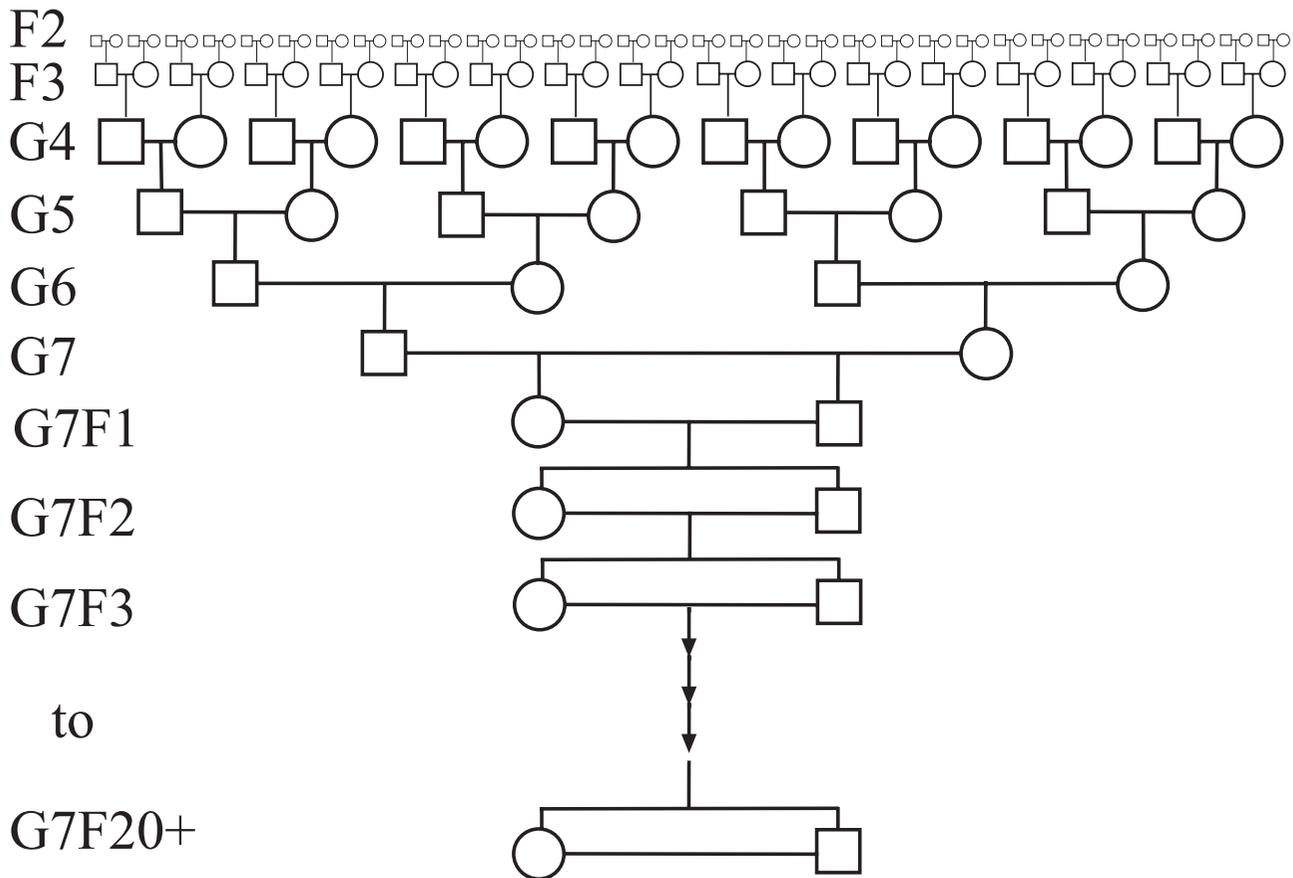


Figure 4
New RI-like breeding scheme. Novel proposed method for maximizing unique recombinations archived in a 2-way RI-like cross. The breeding scheme shown above and discussed more generally in the text results in a single strain with 75% more unique recombinations than a standard RI strain. This example begins with 7 generations of crosses (parental intercross and F1 × F1 cross not shown) prior to inbreeding. All recombinations generated using this method are independent.

independent set of F3 animals that will carry half of the recombinations present in the F2 population in addition to recombinations from the current generation cross. In each subsequent generation there will be a total of $2^{n/2^g}$ breeding cages per line where g is the number of generations following the initial F2 cross. Each F3 animal can then be crossed with another F3, and so on. Since these animals share no common ancestors, all accumulated recombinations will be independent, and since at each generation half of the novel recombinations will be passed to the following generation, the genome expansion should proceed at a predictable rate of $n/2$. As an example, with an initial set of 32 crosses per resulting strain it should be possible to achieve a 2.5-fold expansion from pre-inbreeding breeding in addition to the usual inbreeding expansion. This is not as large an expansion

as that of the current ARI lines (approximately a 75% improvement on the usual 3.3–3.4-fold expansion from inbreeding), but all recombinations will be independent and unique, so the technique is extensible to any desired number of strains. While the initial number of breeding cages and animals per line may seem excessive, this number decreases rapidly, and investigators can set up lines sequentially to minimize needed space and funding.

Another, potentially even more valuable approach is the creation of RI-like lines based on a number of initial progenitors larger than two. Such lines will include a more dense set of recombinations than the typical two progenitor approach and will allow analysis of a wider array of traits, especially given the rather limited diversity of a cross incorporating two parental strains that may often

already share common ancestry. A large number of such strains would be a suitable community-wide resource for efficient fine mapping of complex traits, analysis of epistasis, and a wide variety of other interesting approaches and questions that currently await an appropriate tool. Of course, the proposal above is compatible with multiple strains by generating a population as described with each pair of animals to be included. The outputs of the population would replace the F1 animals in the multi-way cross for any cross design. The 1K Collaborative Cross proposed by the Complex Trait Consortium (CTC) uses such a design [43].

Strength in numbers

The cross proposed by the CTC has another important aspect – it would consist of at least 1000 independent lines. As has been amply demonstrated in plants and other organisms, RI strains have many advantages as a mapping resource when a sufficient number exist to adequately power the investigations in question. This strain set, 80 lines in combination with the original BXD RI strains, will be the largest and most recombinant strain set available in mice but will still be much smaller than the strain set available in maize. If this strain set and the 77 member LXS strain set show promise at all, relative to the much smaller strain sets currently available, they should be considered proof-of-principle for a much larger enterprise. The sketchy reputation of RI-based complex trait mapping in the mouse genetics community will evaporate rapidly if we borrow a leaf from our colleagues in the plant genetics community and create a tool adequate to the statistical requirements of our desired results.

Availability of strains

We intend to make the BXD ARI lines widely available to the academic community. The first set of lines available will be those inbred at Princeton, as these are already extensively genotyped. Since the Princeton facility has a number of pathogens, which prevent export to most other animal facilities, we have rederived all but four of these strains and are establishing breeding colonies in the SPF facility at UTHSC. Once these colonies are established, strains will be made available to the academic research community both prior to and after complete inbreeding. We expect that at least 17 genotyped strains will be available by publication, with the remaining densely genotyped strains available within a few months. Subject to breeding constraints, we intend to make additional strains available as rapidly as possible.

Conclusions

Over the past decade, complex trait mapping has become considerably more sophisticated. Part of that sophistication has involved a shift from early RI-based approaches that were primarily useful for detecting loci of large effect

to segregating populations and multi-stage approaches to mapping. These approaches, and others including consomic or chromosome substitution strains (reviewed in [44]) using existing knockouts as congenics [45,46] multiple cross mapping [47] and others are extremely valuable tools for complex trait geneticists, but ultimately a more powerful community-wide resource is needed. The currently existing RI lines incorporate many of the characteristics of such a resource, but do not offer the necessary variability, power, or resolution to be a general-purpose mapping tool.

The ARI approach outlined in this article is a means of creating a remarkably powerful mapping resource. Making inbred strains from AILs is not sufficiently extensible to serve as direct model for the creation of an extremely large strain set, but in the process of creating the ARI lines we have learned many valuable lessons that will hopefully facilitate the creation of such a resource. Perhaps the most important of these lessons is that the development and design of tools and resources for the community should be considered a high priority. Over the long periods of time that popular mouse models tend to be used by the research community, a popular strain set will be used by many investigators, and small improvements in resource design and maintenance cost can impact the quality and price of science for many years.

In addition, of course, the current and near-term availability of the ARI strains greatly extends the utility of the popular BXD RI strain set and we expect and hope that these strains will be useful to a wide variety of investigators in the coming years.

Methods

The ARI lines described here originated from two separate B6 × D2 advanced intercross lines, one generated at the University of Tennessee Health Science Center (UTHSC) and the other at Princeton University (Princeton). Some animals from the UTHSC AIL were transferred to Princeton at the G9 AIL generation. These animals, in addition to the 13 lines derived from the Princeton AIL, were inbred at Princeton to generate advanced recombinant inbred (ARI) lines. The remaining ARI lines were inbred at UTHSC (see Fig. 1). Differences between protocols at the two institutions have been noted.

Mice

B6, D2, and B6D2F1 male and female animals were ordered from The Jackson Laboratory (Bar Harbor, ME) and bred at Princeton or UTHSC as described below. The Princeton facility harbors several murine pathogens, including EDIM and MHV, while the UTHSC facility is specific pathogen free (SPF). All but 4 lines have been rederived and are now housed at UTHSC.

Advanced intercross generation

At Princeton, B6D2F1 animals procured from The Jackson Laboratory were bred to generate B6D2F2 animals. F2 animals were randomly chosen and mated to create a G3 population consisting of approximately 45 breeding cages with two male and two female animals per breeding cage. Following the G3 generation, breeding followed a version of the advanced intercross technique described by Darvasi and Soller [23]. Briefly, at each generation, matings were chosen to minimize the number of common parents. The Princeton AIL maintained a rotating breeding schedule to ensure that animals shared no more than one ancestor in the previous three generations, and usually no more than one ancestor in the previous four generations.

The Tennessee AIL was generated in a similar manner, minimizing common ancestors in each cross, with approximately 30 breeding cages per generation. One important difference, however, is that the F1 generation of the UTHSC AIL was generated from reciprocal crosses of the parental strains rather than from commercially acquired B6D2F1 animals. Because of this difference, the Y chromosome and mitochondrial genome of strains in the Tennessee AIL may come from B6 or D2 parents, while the Y chromosome of strains in the Princeton AIL comes exclusively from the D2 parent and the mitochondrial genome from the B6 parent. Another important difference is that, because of the smaller number of breeding cages per generation, the UTHSC AIL population will have experienced a greater amount of random fixation per generation (in the AIL breeding phase) than the Princeton AIL.

Inbreeding

The 46 BXD ARI strains are currently sufficiently inbred for most mapping purposes, and have already been successfully used for mapping alcohol preference loci with good agreement between ARI and RI results (unpublished result). The strains are an average of 16 generations inbred, and will require only 4 more generations before they are formally considered to be fully inbred strains [2]. It should be noted that to conserve cage space and insure breeding in the both the Princeton and Tennessee inbreeding programs, two animals of each sex were selected for each cage at each generation, if available. Care was taken to ensure that animals selected for breeding were siblings rather than cousins, but it is possible that cousins were occasionally selected in cases where litters were born within 1–2 days of each other. Cousin-cousin matings slow the increase of homozygosity, so the effective inbreeding generation may be slightly lower than the reported generation.

Genotyping

A total of 588 microsatellite loci polymorphic between B6 and D2 strains, distributed across all autosomes and the X

chromosome (average interval between markers 2.5 cM), were amplified for 10 Group A and 10 Group B strains inbred at Princeton [see additional file 1: genotypes.xls]. We also generated a set of 30 genotypes to analyze independence of recombinations, using DNA separately taken from each parent contributing to the subsequent generations. This set [see additional file 2: genopairs.xls] was genotyped in 9 Group A strains and 35 Group B strains, consisted largely of 16 pairs of unlinked loci spaced an average of 0.49 Mb apart (range 0.24 to 0.77). These same strains were also genotyped across Chr 1 at relatively high resolution, which may be useful for researchers with QTLs in this region.

A third set of 286 loci [see additional file 3: 286newgenotypes.xls] was generated for nearly all ARI strains. This set of genotypes was generated using DNA pooled from multiple animals from each strain. Since this set of genotypes was completed after initial submission of this manuscript, it was not used for analysis but is included for researchers interested in a broader and more up to date, though less dense, set of genotypes for these animals.

Genotyping was performed using a modified version of the PCR protocol of Love and colleagues [48] and Dietrich and colleagues [49] described in detail at <http://www.nerinet.org/papers/PCR.html>. DNA for the initial genotyping pass was purified from Princeton tail samples using standard phenol-chloroform extractions from single animals that contributed to the subsequent generations. Briefly, primer pairs purchased from Research Genetics (Huntsville, AL) were amplified using a high-stringency touchdown protocol in which the annealing temperature was lowered progressively from 60°C to 50°C in 2°C steps over the first 6 cycles [50]. After 30 cycles, PCR products were run on cooled 2.5% Metaphor agarose gels (FMC Inc., Rockland ME), stained with ethidium bromide, and photographed. Gel photographs were scored and directly entered into relational database files.

It is worth noting that since the initial genotype data [see additional file 1: genotypes.xls] for each strain was based on a single animal, some of the genotypes scored as homozygous are fixed in the genotyped animal but not in the population of two animals actually contributing to the following generation. A small number of such loci may ultimately be fixed as the opposite allele (eg. genotyped as B6, ultimately fixed as D2). Since this will not be an issue in genotypes of the inbred population, is a small source of error in any case, and is easily avoided by genotyping DNA pooled from both parents, it will be of minimal interest to investigators currently beginning projects in this strain set and is only relevant to investigators who have already completed pilot projects using these strains.

Estimating heterozygosity

For calculation of actual heterozygosity, we assumed that all points between typed heterozygous loci were heterozygous. Where a single locus was heterozygous, we assumed that half the distance between this locus and the surrounding homozygous loci was also heterozygous. For determination of percent heterozygosity we used the sum of the distances between the centromere and the terminal genotyped marker, in all cases using marker positions from Williams and colleagues [4]. The length of the genome from centromere to terminal marker of all chromosomes was 1501 cM, which agrees well with an average of several other estimates of the length of the mouse genome at 1453 cM [2].

We used a small second set of genotypes, generated several generations later, to assess the success of our estimate. This set of genotypes consisted of one marker from each of the closely spaced pairs mentioned above that had also been assayed in the earlier set of genotypes. Homozygous genotypes in this set were compared with homozygous genotypes in the earlier set, and markers with heterozygous genotypes in either set were discarded.

Determining number of total recombinations

With a sufficiently dense marker map, estimating the number of recombinations in a fully inbred RI strain is trivial. In a still partly heterozygous strain, however, the eventual fixation of heterozygous intervals to either homozygous state will add a number of recombinations during the remaining process of inbreeding. A heterozygous region flanked on both sides by the same genotype (eg BBBHHBBB) will either contribute 0 or 2 recombinations when inbred, depending on which haplotype (BBBBBBBB or BBDDDBBB) is ultimately inherited, whereas a heterozygous region flanked by regions of opposite genotype (eg BBBHDDDD) will yield one recombination regardless of the inherited haplotype (BBBBBDDDD or BBDDDDDD). In both cases, each heterozygous region contributes an average of one recombination, so each transition from a homozygous to heterozygous genotype or the reverse will contribute an average of 0.5 recombinations.

In order to compare the total number of recombination events detected in the ARI lines to the total number detected in the BXD RI lines, it is necessary to estimate the likely recombination density detected with a given number of markers. In order to accomplish this, we first reduced the genotyping resolution of the BXD RI set – 936 markers with an average of 41.4 recombinations per line as reported [4] – to 588 approximately evenly spaced markers, a number equal to the current ARI genotyping resolution and the best case scenario for detecting recombinations. At this marker density, there are an average of

37.4 recombinations per conventional BXD RI line. We then used the ratio of recombinations at the higher and lower BXD resolutions to estimate the likely number of recombinations in the ARI strains at a higher resolution using a simple ratio:

$$(\text{BXD high resolution} / \text{BXD low resolution}) \times (\text{ARI low resolution})$$

In simulations comparing BXD Chr 1 with two additional recombinations per strain to unmodified BXD Chr 1, computing the high density map for the high-recombination genome was always conservative (data not shown) over a 3-fold change in marker density. Applying the equivalent ratio to the low density ARI map should likewise be conservative. Intuitively, the set with the higher density of recombinations is more likely to lose a detected recombination when any given marker is removed since a higher density implies a lower mean distance between recombinations.

Estimated minimum number of unique recombinations

Estimating the number of unique recombinations is more difficult than estimating the number of total recombinations because of the complex, shared ancestry of the ARI strains. While recombinations between any two markers from the two ARI sets (Group A and Group B) are certain to be independent, shared lineage from the AIL will account for some of the recombinations within A or B sets of lines. As more lines are added, the effects of shared lineage become more pronounced. While all lines accumulated independent recombinations during the process of inbreeding, it is difficult to determine which recombinations are independent and which are shared. We have taken a three-part approach to the problem; (1) estimating the minimum number of unique recombinations per line and adjusting using a conservative set of assumptions, (2) providing a more likely theoretical estimate of the expected number of recombinations, and (3) directly estimating the fraction of unique recombinations using a set of very closely spaced markers.

The minimum number of unique recombinations per line was initially estimated by counting up to 1 B→D and 1 D→B recombination per marker pair and adding, for those marker pairs where there were no B→D transitions, 0.5 recombinations for the first (B or D)→H or H→(B or D) transition involving a possible inheritance of an equivalent transition on resolution of the heterozygous region. Since each heterozygous region is characterized by two transitions, each heterozygous region contributes one total recombination on average. The maximum number of recombinations per marker pair is 2, regardless of the number of lines sampled, and so this estimate is sensitive to number of lines considered and grows increasingly

conservative with the addition of more lines. In order to conservatively estimate total recombinations and facilitate comparison with the more densely genotyped BXD RI lines, we used existing genotyping data from the lower resolution BXD RI set described above. From the total low resolution BXD RI data set, (588 markers) we then tested 200 sets of 10 randomly chosen strains (the genotyped strains from the A and B strain sets generated at Princeton each include at most 10 strains). The difference between the actual number of recombinations detected per BXD line at maximum current resolution and the number detected using the minimal method above at a lower resolution represents the minimum number of independent recombinations undetected in the ARI set due to unidentified unique recombinations and lower marker resolution. This estimate makes the extremely conservative assumptions that the only uncounted unique recombinations using the minimal method are those derived from the inbreeding process and that the number of such recombinations likely to be missed in the ARI lines is the same as the number in the RI lines. The latter assumption is also quite conservative given the higher recombination density and resulting likely higher false negative rate for unique recombinations in the ARI lines. It is necessary to consider marker resolution and missed unique recombinations due to application of the minimal estimation method because these factors are related. For instance, at an infinite marker density, the number of unique BXD RI recombinations missed by the minimal estimation method is 0.

A conservative estimate of the number of unique recombinations

A more likely treatment of unique recombinations starts with the minimal method initially described above. However, instead of adding the estimated number of recombinations missed by the lower resolution ARI genotyping effort, we determined the fraction of recombinations missed by application of the minimal method and reduction of resolution in the low resolution BXD RI set, as compared to the average number of recombinations detected in the full resolution BXD RI set. We then applied this ratio to the minimal number of recombinations in the ARI set. This analysis assumes that the fraction of unique recombinations undetected by the minimal method at a given resolution is the same between the BXD RI and ARI sets. While this method is less strict than the previous method, the assumption is not unreasonable. Adding non-unique recombinations increases the total number of detected recombinations, but not the minimal number. Since the ratio of total to minimal recombinations in the BXD RI line is the ratio when all recombinations are actually unique, it is a reasonable ratio to apply to the minimal number of detected recombinations, which are also unique.

Experimental estimate of the number of unique recombinations

We experimentally tested these calculations by genotyping 16 pairs of tightly linked loci using DNA samples from both the male and female that contributed to the next generation of sibling inbreeding. In pairs of loci with at least one recombination we totaled the number of directional transitions, counting (B or D) \rightarrow H and H \rightarrow (B or D) transitions as 0.5 recombinations. Where only one of the two parents showed a recombination between markers, that transition was counted as 0.5 transition, representing the 50% likelihood that it will be inherited in the final strain set. Ideally all intervals would have been small enough to contain at most one unique recombination and its identical-by-descent counterparts. However, several intervals contained transitions that differed in direction and were therefore independently derived. We treated these intervals as two separate intervals, one for each directional set of transitions, for purposes of determining independent recombinations. Where the only contributions were from transitions to or from heterozygotes or situations where only one parent showed a transition (in other words, any situation where inheritance of a transition in the interval was not assured), the contribution of the interval was calculated as the likelihood that there would be at least one recombination in the interval. The ratio of total recombinations to intervals is a measure of the fraction of independent recombinations. This approach allowed us to estimate the shared recombinations for the entire set of ARI lines in addition to the more limited set of densely genotyped lines. We also estimated the shared recombinations for subpopulations of varying sizes within the Group B population in an effort to define the population size/unique recombination relationship for a population based on an AIL of this size by evaluating 1000 randomly generated populations and inspecting each increment in number of ARI lines.

Authors' contributions

JLP was responsible for inbreeding of BXD43-66, initially suggesting generating inbred lines from AIL progenitors, and authorship of this paper. LL was responsible for inbreeding of BXD67-BXD94 as well as coordinating rederivation of BXD43-BXD66 with JLP, transfer of these lines to UTHSC, and coordinating genotyping efforts with JG. JG performed all genotyping discussed in this paper. LMS provided advice and support to JLP during the generation of the lines described. RWW initiated the inbreeding program at UTHSC and provided advice and support to LL and JG during the inbreeding and genotyping process. RWW and LMS provided guidance to JP during preparation of this manuscript.

Additional material

Additional File 1

Excel spreadsheet of ARI genotypes with strain and marker headings for a subset of ARI strains.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2156-5-7-S1.xls>]

Additional File 2

Excel spreadsheet of pairs of closely spaced genotypes used for estimation of unique recombinations.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2156-5-7-S2.xls>]

Additional File 3

Excel spreadsheet of 286 new genotypes for nearly all ARI strains.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2156-5-7-S3.xls>]

Acknowledgements

Our thanks for financial support from (1) The Informatics Center for Mouse Neurogenetics; (2) P20-MH62009 from NIMH, NIDA, and NSF to RWW. (2) INIA grants U01AA13499 and U24AA135B from NIAAA to RWW, and (3) R37 HD20275 from NICHD to LMS. We thank Shuhua Qi, Zhiping Jia, for animal care and genotyping; Irina Agulnik, Olga Chertkov, and Edward Gomez for animal care; Arthur Centeno for computer support; Pamela Franklin and Barbara Smith for administrative assistance; and John LeVorse, Daniel Goldowitz, and Kristin Hamre for rederivations.

References

- Taylor BA: **Recombinant inbred strains**. *Genetic Variation in the Laboratory Mouse*, 2nd edn Edited by: Lyon M L and Searle A G. Oxford, Oxford University Press; 1989:773-796.
- Silver LM: **Mouse Genetics: Concepts and Applications**. New York, Oxford University Press; 1995.
- Haldane JBS, Waddington CH: **Inbreeding and linkage**. *Genetics* 1931, **16**:357-374.
- Williams RW, Gu J, Qi S, Lu L: **The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis**. *Genome Biol* 2001, **2**:RESEARCH0046.
- Peirce JL, Chesler EJ, Williams RW, Lu L: **Genetic architecture of the mouse hippocampus: identification of gene loci with selective regional effects**. *Genes Brain Behav* 2003, **2**:238-252.
- Grisel JE, Metten P, Wenger CD, Merrill CM, Crabbe JC: **Mapping of quantitative trait loci underlying ethanol metabolism in BXD recombinant inbred mouse strains**. *Alcohol Clin Exp Res* 2002, **26**:610-616.
- Demarest K, McCaughran J., Jr., Mahjubi E, Cipp L, Hitzemann R: **Identification of an acute ethanol response quantitative trait locus on mouse chromosome 2**. *J Neurosci* 1999, **19**:549-561.
- Koyner J, Demarest K, McCaughran J., Jr., Cipp L, Hitzemann R: **Identification and time dependence of quantitative trait loci for basal locomotor activity in the BXD recombinant inbred series and a B6D2 F2 intercross**. *Behav Genet* 2000, **30**:159-170.
- Bennett B, Carosone-Link P, Johnson TE: **Genetics of body weight, growth size, and sex ratio in recombinant inbred strains**. *Physiological Genomics* 2003, **in press**.
- Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F: **Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in Arabidopsis**. *Theor Appl Genet* 2002, **104**:1173-1184.
- Tang S, Heesacker A, Kishore VK, Fernandez A, Sadik ES, Cole G, Knapp SJ: **Genetic Mapping of the Or5 Gene for Resistance to Orobanch Race E in Sunflower**. *Crop Sci* 2003, **43**:1021-1028.
- Flores Berrios E, Gentzbittel L, Kayyal H, Ailbert G, Sarrafi A: **AFLP mapping of in vitro organogenesis traits using recombinant inbred lines in sunflower (Helianthus annuus L.)**. *Theor Appl Genet* 2000, **101**:1299-1306.
- Johnson W, Gepts P: **Segregation for performance in recombinant inbred populations resulting from inter-gene pool crosses of common bean (Phaseolus vulgaris L.)**. *Euphytica* 1999, **106**:45-56.
- Saliba-Colombani V, Causse M, Gervais L, Philouze J: **Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome**. *Genome* 2000, **43**:29-40.
- Limami AM, Rouillon C, Glevarec G, Gallais A, Hirel B: **Genetic and physiological analysis of germination efficiency in maize in relation to nitrogen metabolism reveals the importance of cytosolic glutamine synthetase**. *Plant Physiol* 2002, **130**:1860-1870.
- Fracheboud Y, Ribaut JM, Vargas M, Messmer R, Stamp P: **Identification of quantitative trait loci for cold-tolerance of photosynthesis in maize (Zea mays L.)**. *J Exp Bot* 2002, **53**:1967-1977.
- Stenehjem S, Bruggemann E: **Glumb bar phenotypes in a B73 x Mo17 recombinant inbred population reveal the epistatic interaction between p11 and b1**. *Maize Genetics Conference Abstracts* 2001:43:P80.
- Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF: **Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster**. *Proc Natl Acad Sci U S A* 1997, **94**:9734-9739.
- van Swinderen B, Shook DR, Ebert RH, Cherkasova VA, Johnson TE, Shmookler Reis RJ, Crowder CM: **Quantitative trait loci controlling halothane sensitivity in Caenorhabditis elegans**. *Proc Natl Acad Sci U S A* 1997, **94**:8232-8237.
- Williams RW, Threadgill DW, Manly KF, Rosen GD, Shou S, Airey DC, Gu J, Lu L: **Integrated complex trait analysis of brain and behavior in mouse: near-term prospects for high-resolution QTL mapping using large RI and RIX sets**. *The 4th Annual Meeting of the International Behavioural and Neural Genetics Society San Diego, USA*; 2001.
- Threadgill DW, Hunter KW, Williams RW: **Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort**. *Mamm Genome* 2002, **13**:175-178.
- Darvasi A: **Experimental strategies for the genetic dissection of complex traits in animal models**. *Nat Genet* 1998, **18**:19-24.
- Darvasi A, Soller M: **Advanced intercross lines, an experimental population for fine genetic mapping**. *Genetics* 1995, **141**:1199-1207.
- Green EL: **Genetics and Probability in Animal Breeding Experiments**. New York, Oxford University Press; 1981.
- Winkler CR, Jensen NM, Cooper M, Podlich DW, Smith OS: **On the determination of recombination rates in intermated recombinant inbred populations**. *Genetics* 2003, **164**:741-745.
- Bailey DW: **Strategic uses of recombinant inbred and congenic strains in behavior genetics research**. *Genetic Research Strategies for Psychobiology and Psychiatry* Edited by: Gershon ES, Matthysse S, Breakefield XO and Ciaranello ED. New York, Plenum; 1981:189-198.
- Plomin R, McClearn GE, Gora-Maslak G, Neiderhiser JM: **Use of recombinant inbred strains to detect quantitative trait loci associated with behavior**. *Behav Genet* 1991, **21**:99-116.
- Johnson TE, DeFries JC, Markel PD: **Mapping quantitative trait loci for behavioral traits in the mouse**. *Behav Genet* 1992, **22**:635-653.
- Belknap JK, Metten P, Helms ML, O'Toole LA, Angeli-Gade S, Crabbe JC, Phillips TJ: **Quantitative trait loci (QTL) applications to substances of abuse: physical dependence studies with nitrous oxide and ethanol in BXD mice**. *Behav Genet* 1993, **23**:213-222.
- Browman KE, Crabbe JC: **Quantitative trait loci affecting ethanol sensitivity in BXD recombinant inbred mice**. *Alcohol Clin Exp Res* 2000, **24**:17-23.
- Phillips TJ, Lessov CN, Harland RD, Mitchell SR: **Evaluation of potential genetic associations between ethanol tolerance and sensitization in BXD/Ty recombinant inbred mice**. *J Pharmacol Exp Ther* 1996, **277**:613-623.

32. Risinger FO, Cunningham CL: **Ethanol-induced conditioned taste aversion in BXD recombinant inbred mice.** *Alcohol Clin Exp Res* 1998, **22**:1234-1244.
33. Grisel JE, Belknap JK, O'Toole LA, Helms ML, Wenger CD, Crabbe JC: **Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains.** *J Neurosci* 1997, **17**:745-754.
34. de Haan G, Van Zant G: **Dynamic changes in mouse hematopoietic stem cell numbers during aging.** *Blood* 1999, **93**:3294-3301.
35. Chesler EJ, Wang J, Lu L, Qu Y, Manly KF, Williams RW: **Genetic correlates of gene expression in recombinant inbred strains: a relational model to explore for neurobehavioral phenotypes.** *Neuroinformatics* 2003, **1**:343-357.
36. Wang J, Williams RW, Manly KF: **WebQTL: Web-based complex trait analysis.** *Neuroinformatics* 2003, **1**:299-308.
37. **WebQTL** [<http://www.webqtl.org>]
38. Belknap JK, Mitchell SR, O'Toole LA, Helms ML, Crabbe JC: **Type I and type II error rates for quantitative trait loci (QTL) mapping studies using recombinant inbred mouse strains.** *Behav Genet* 1996, **26**:149-160.
39. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flieck P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grahham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
40. **Celera Discovery System** [<http://www.celeradiscoverysystem.com>]
41. Belknap JK, Atkins AL: **The replicability of QTLs for murine alcohol preference drinking behavior across eight independent studies.** *Mamm Genome* 2001, **12**:893-899.
42. McClearn GE, Kakhiana R: **Selective breeding for ethanol sensitivity: Short-sleep and long-sleep mice.** *Development of Animal Models as Pharmacogenetic Tools* Edited by: McClearn G E, Dietrich RA and Erwin VG. Washington, DC, US Government Printing Office; 1981:81-113.
43. Flaherty L, Abiola O, Angel JM, Avner P, Bachmanov AA, Belknap JK, Bennett B, Blankenhorn EP, A. Blizard D., Bolivar VJ, Brockmann GA, Buck KJ, Bureau J-F, Casley WL, J. Chesler E., Cheverud JM, Churchill GA, Cook MN, Crabbe JC, Crusio WE, Darvasi A, de Haan G, Demant P, Doerge RW, W. Elliot R., R. Farber C., Flint J, Gershenfeld H, Gibson JP, Gu W, Himmelbauer H, Hitzemann R, Hsu H-C, Hunter K, Iraqi F, Jansen RC, Johnson TE, Jones BC, Kempermann G, Lammerert F, Lu L, Manly KF, Matthews DB, Medrano JF, Mehrabian M, Mitelman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Mountz JD, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Paigen B, Palmer AA, Peirce JL, Pomp D, Rosemann M, Rosen GD, Schalkwyk LC, Seltzer Z, Settle S, Shimomura K, Shou S, Sikely JM, Siracusa LD, Spearow JL, Teuscher C, Threadgill DW, Toth LA, Toyee AA, Vadasz C, Van Zant G, Wakeland E, Williams RW, Zhang H-G, Zou F: **The nature and identification of quantitative trait loci: a community's view.** *Nat Genet Reviews* 2004.
44. Nadeau JH, Singer JB, Martin A, Lander ES: **Analysing complex genetic traits with chromosome substitution strains.** *Nat Genet* 2000, **24**:221-225.
45. Peirce J: **Looking at old tools in new ways: using knockouts as congenics to study QTLs.** *Genome Res* 2001, **11**:1469-1471.
46. Bolivar VJ, Cook MN, Flaherty L: **Mapping of quantitative trait loci with knockout/congenic strains.** *Genome Res* 2001, **11**:1549-1552.
47. Hitzemann R, Malmanger B, Cooper S, Coulombe S, Reed C, Demarest K, Koyner J, Cipp L, Flint J, Talbot C, Rademacher B, Buck K, McCaughran J., Jr.: **Multiple cross mapping (MCM) markedly improves the localization of a QTL for ethanol-induced activation.** *Genes Brain Behav* 2002, **1**:214-222.
48. Love JM, Knight AM, McAleer MA, Todd JA: **Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites.** *Nucleic Acids Res* 1990, **18**:4123-4130.
49. Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli NC, Lander ES: **A genetic map of the mouse suitable for typing intraspecific crosses.** *Genetics* 1992, **131**:423-447.
50. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS: **'Touchdown' PCR to circumvent spurious priming during gene amplification.** *Nucleic Acids Res* 1991, **19**:4008.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

