Cell Systems

Article

A platform for experimental precision medicine: The extended BXD mouse family

Highlights

- There are now 140 fully inbred BXD strains available, with high-quality genotypes
- More strains, new genotypes, and new models have improved power and precision
- We have high power even for traits with low heritability or small effect sizes
- A phenome of >100 omics datasets and >7,500 classic phenotypes is freely available

Authors

David G. Ashbrook, Danny Arends, Pjotr Prins, ..., Johan Auwerx, Lu Lu, Robert W. Williams

Correspondence

dashbrook@uthsc.edu (D.G.A.), lulu@uthsc.edu (L.L.), rwilliams@uthsc.edu (R.W.W.)

In Brief

Ashbrook et al., have expanded the BXD family to 140 strains, providing a new tool for translational precision and predictive biology, and extended the usefulness of the deep phenome of >100 omics datasets and >7,500 classical phenotypes already available. They show increased precision and power by using new genotypes, updated models, and more strains.

Ashbrook et al., 2021, Cell Systems *12*, 235–247
March 17, 2021
https://doi.org/10.1016/j.cels.2020.12.002

CellPress

Cell Systems

CellPress

Article

A platform for experimental precision medicine: The extended BXD mouse family

David G. Ashbrook,[1,7,8,*] Danny Arends,[2,7] Pjotr Prins,[1] Megan K. Mulligan,[1] Suheeta Roy,[1] Evan G. Williams,[3] Cathleen M. Lutz,[4] Alicia Valenzuela,[4] Casey J. Bohl,[1] Jesse F. Ingels,[1] Melinda S. McCarty,[1] Arthur G. Centeno,[1] Reinmar Hager,[5] Johan Auwerx,[6] Lu Lu,[1,*] and Robert W. Williams[1,*]

[1]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN 38163, USA
[2]Lebenswissenschaftliche Fakultät, Albrecht Daniel Thaer-Institut, Humboldt-Universität zu Berlin, Invalidenstraße 42, 10115 Berlin, Germany
[3]Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, L-4365 Esch-sur-Alzette, Luxembourg
[4]Mouse Repository and the Rare and Orphan Disease Center, the Jackson Laboratory, Bar Harbor, ME 04609, USA
[5]Division of Evolution & Genomic Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
[6]Laboratory of Integrative Systems Physiology, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: dashbrook@uthsc.edu (D.G.A.), lulu@uthsc.edu (L.L.), rwilliams@uthsc.edu (R.W.W.)
https://doi.org/10.1016/j.cels.2020.12.002

SUMMARY

The challenge of precision medicine is to model complex interactions among DNA variants, phenotypes, development, environments, and treatments. We address this challenge by expanding the BXD family of mice to 140 fully isogenic strains, creating a uniquely powerful model for precision medicine. This family segregates for 6 million common DNA variants—a level that exceeds many human populations. Because each member can be replicated, heritable traits can be mapped with high power and precision. Current BXD phenomes are unsurpassed in coverage and include much omics data and thousands of quantitative traits. BXDs can be extended by a single-generation cross to as many as 19,460 isogenic F1 progeny, and this extended BXD family is an effective platform for testing causal modeling and for predictive validation. BXDs are a unique core resource for the field of experimental precision medicine.

INTRODUCTION

The development of predictive, preventive, personalized, and participatory health care requires deep and coherent phenome data as prerequisites. These data must be matched to sophisticated causal models of complex interactions among sequence variants, phenotypes at many levels, and environmental exposures (Auffray et al., 2010; Hood and Friend, 2011; Williams, 2009). For most diseases, it is not yet practical to predict risk and best treatment options for the simple reason that each human is unique (Schüssler-Fiorenza Rose et al., 2019). This *N-of-1* problem inhibits development and refinement of predictions. However, it is now possible to create optimized experimental platforms that fill this gap in the form of large and fully characterized families of isogenic rodents, such as the BXD.

The BXD family was started 50 years ago by crossing fully inbred female C57BL/6J (B6 or B) to male DBA/2J (D2 or D) mice. The first set of BXD recombinant inbred (RI) strains was used to map Mendelian traits (Taylor et al., 1973, 1999), but BXDs are now used mainly to map complex traits, including cancers, metabolic and cardiovascular disease (Grizzle et al., 2002; Koutnikova et al., 2009; Lee et al., 1995; McGinnis et al., 1993), brain structure (Belknap et al., 1992a; Rosen et al., 2009; Seecharan et al., 2003; Zhou and Williams, 1999), behavior, and pharmacology (Ashbrook et al., 2018a; Belknap et al., 1992b, 1993; Grisel et al., 1997; Jones et al., 2006; Knoll et al., 2018; Palmer et al., 2006; Phillips et al., 1998; Rodriguez et al., 1994; Weimar et al., 1982). We started a second wave of BXDs in the late 1990s (Peirce et al., 2004) using advanced intercross (AI) progeny (Darvasi, 1998) (Figures 1 and S1; Table S1). These AI-derived family members incorporate twice as many recombination junctions as do conventional BXDs (Broman, 2005; Crow, 2007; Haldane and Waddington, 1931; Williams et al., 2001) (Figure 1C)—an attribute that improves mapping precision by about 2-fold.

BXDs are well suited for systems biology (Williams and Williams, 2017), thanks to a well-integrated phenome consisting of over 7,500 classical phenotypes and over 100 omics data sets (Williams and Auwerx, 2015; Williams et al., 2001). This phenome extends back to Taylor's (1973) analysis of cadmium toxicity, through to recent studies of metabolism (Roy et al., 2020; Wang et al., 2013, 2016b; Williams et al., 2016, 2020; Wu et al., 2014), addiction (Dickson et al., 2016, 2019; Mulligan et al., 2013), behavior (Carhuatanta et al., 2014; Graybeal et al., 2014; Mulligan and Williams, 2015; Philip et al., 2010), infectious disease (Hayes et al., 2014; McKnite et al., 2012; Wang et al., 2020), epigenetics (Baker et al., 2019; Sandoval-Sierra et al., 2020), and even indirect genetic effects (Ashbrook

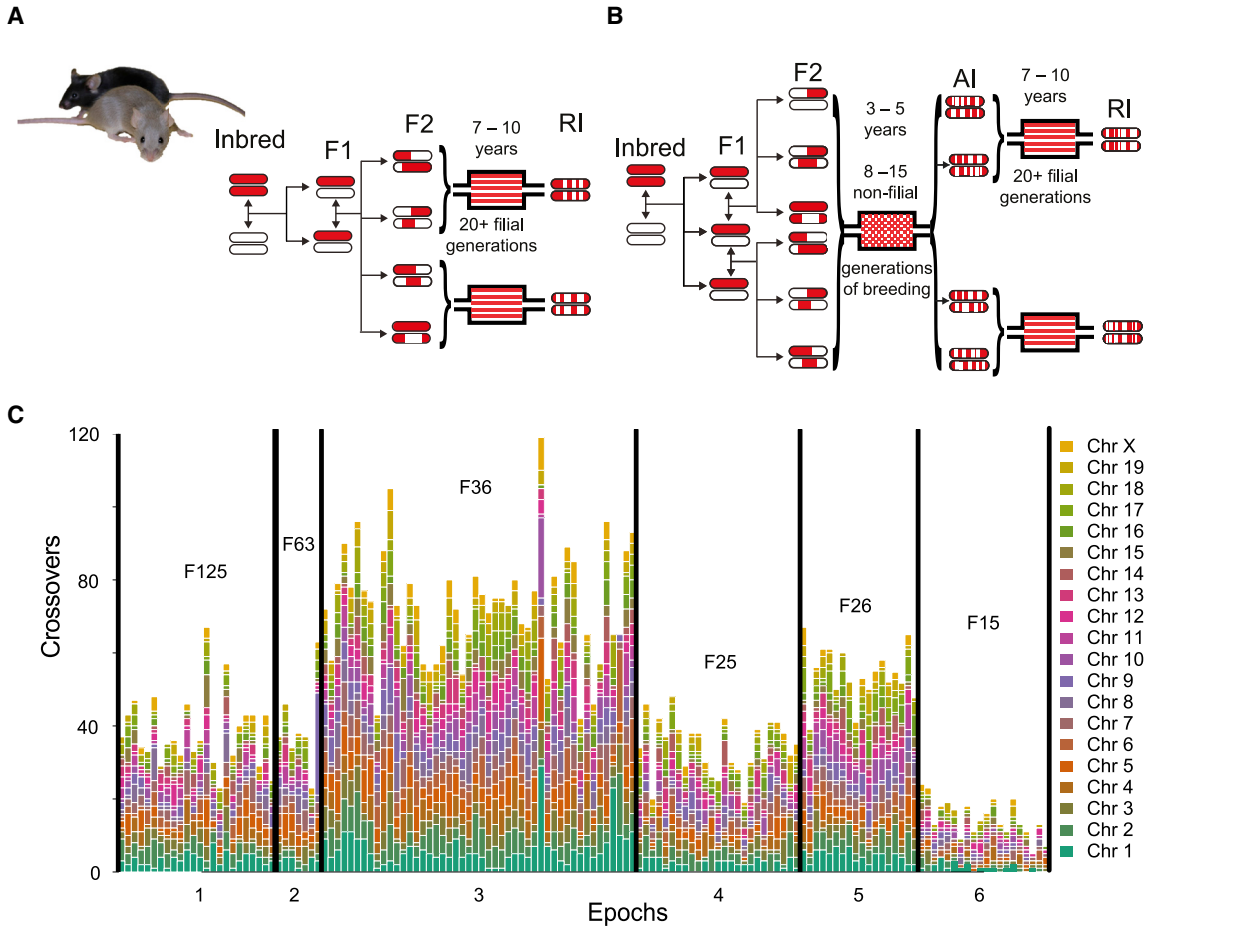

Figure 1. Production of the BXD family by standard F2

(A) or Advanced Intercross (B), and Recombination Differences (C) Four epochs of the BXD were derived from F2 (A) (~75 strains; epochs 1, 2, 4, 6 in C). Two epochs were derived from advanced intercross (B) (~65 strains; epochs 3, 5 in C). Red represents regions of the genome coming from C57BL/6J (B6), white represents regions from the DBA/2J (D2). Solid lines represent a single generation. Adapted from Peirce et al. (2004)(Williams and Auwerx, 2015).

(C) Genome-wide visualization of recombinations in the BXD. Number of recombinations per strain (nRecS), ignoring heterozygous and unknown genotypes plotted using a color gradient. Strain numbers on the x axis, horizontal lines separate the epochs, epochs are annotated on the x axis. Mean number of inbreeding generations is shown for each epoch. Epoch 6 appears to have fewer recombinations due to a large number of heterozygous loci at genotyping. Details in Figure S1; Table S1.

et al., 2015a, 2017; Baud et al., 2017). BXDs have been used to test developmental and evolutionary hypotheses (Hager et al., 2012; Oren et al., 2015; Seecharan et al., 2003), to study GXE and consequences of treatments as a function of age, diet, and sex (Fleet et al., 2016; Philip et al., 2010; Roy et al., 2020; Sandoval-Sierra et al., 2020; Williams et al., 2016, 2020), gene pleiotropy (Wang et al., 2016a), and to test behavioral predictions based on differences in brain architecture (Yang et al., 2008).

Here, we summarize the current status of this resource with a focus on the genetic structure and the power and precision of mapping trait variance to loci and genes. We have almost doubled the size of the BXD family from ~80 (Peirce et al., 2004) to 140 extant strains. Each of the 140 fully inbred and isogenic strains of mice is an immortal genometype that can be resampled at any stage, in both sexes, and under varied but controlled conditions to quantify gene-by-environmental interactions (GXE) and to test the accuracy and errors of genome-to-phenome prediction. Compared with the progeny of conventional intercrosses, AIs, or heterogeneous stock, BXDs are advantageous when heritabilities of traits are low or when measurement errors are high because the genetic signal can be boosted by resampling many isogenic cases (Belknap, 1998; Crow, 2007). A further benefit is that each data set adds quadratically to the number of trait-to-trait associations, and even the oldest data become more powerful as genetic, genomic, and phenomic contexts are improved and enlarged. The BXD, like other replicable reference populations, are therefore well adapted to extensible multisystems analyses, to modeling, and to quantitative prediction (Andreux et al., 2012; Chesler et al., 2003; Jha et al., 2018b, 2018a; Li et al., 2018; Williams et al., 2016). We have therefore assembled deep companion resources, including full sequence for both parents (Baker et al., 2019; Keane et al., 2011; McKnite et al., 2012; Wang

et al., 2016b; Wu et al., 2014). Access to data and statistical tools are available from open-source web services (GeneNetwork.org and Systems-Genetics.org) (Li et al., 2018; Sloan et al., 2016; Williams and Williams, 2017). High-density genetic maps are now combined with linear mixed model mapping to boost power and accuracy.

RESULTS

The BXD family has been expanded to 140 inbred strains

We have approximately redoubled the number of extant BXD strains, from ~35 (Taylor et al., 1999), to ~80 (Peirce et al., 2004), to 140 here. We started producing 108 new BXD strains between 2009 and 2013 (BXD104–BXD220). A total of 123 BXD lines are currently distributed by the Jackson Laboratory (JAX, Table S1). Seventeen additional strains are available at UTHSC but will soon be donated to JAX. All strains are available under a standard material transfer agreement.

We have identified three sets of strains that have almost identical genomes. These are marked by the addition of "a" and "b" suffixes. For example, BXD65, BXD65a, and BXD65b are identical by descent across >90% of their genomes. Without correction, the inclusion of three BXD65 substrains will bias results. There is also more subtle kinship among the BXD family members that results from drift, backcrossing, and unintended selection. Correction of both types of kinship when mapping traits can be accomplished using linear mixed models during analyses, for example, using *Genome-wide Efficient Mixed Model Association* (GEMMA) (Zhou and Stephens, 2012) or R/qtl2 (Broman et al., 2003, 2019) discussed below.

Improved mapping using updated markers

Affymetrix and Illumina arrays have been used to genotype at least more than 100,000 informative markers in 198 BXD strains (Morgan et al., 2015; Peirce et al., 2004; Shifman et al., 2006; Taylor et al., 1999; Zhang et al., 2012). Previously, a subset of 3,830 markers has been used for mapping, but here we have boosted marker numbers nearly by 2-fold to 7,324. This high-density marker set defines proximal and distal limits of chromosomal intervals that are non-recombinant. Markers are spaced at 0.63 Mb ± 1.0 SE, closely matching the expected asymptotic map resolution of ±1 Mb.

The increase in numbers of markers has significantly improved power and precision of linkage analysis (Figures 2 and S2) across all chromosomes (Figure S2). To quantify these improvements in mapping, we computed linkage for the entire BXD phenome—currently 7,562 phenotypes—using three different genotype files: the original file from 2001 (Williams et al., 2001), the "classic" genotypes used from 2005 to about 2016, and the updated genotypes. Each of these stages roughly doubled the numbers of markers (from 1,578 to 3,811 to 7,324). We counted the number of traits associated with the logarithm of the odds (LOD) linkage scores at three cutoffs: 2.2 (suggestive; a likelihood of one false positive per genome scan), 3.6 (significant; one false positives per 20 scans), and 5.4 (highly significant; one false positives per 100 scans) (Lander and Kruglyak, 1995) (Figure 2C). The largest boost in LOD was achieved for traits with linkage between 3.6 and 5.4 LOD—up by 30% compared with older marker sets (Figures 2C and 2D).

We can further improve power and precision using linear mixed models

Compared with the traditionally used Haley-Knott mapping (Haley and Knott, 1992), more recently developed algorithms can correct for variable kinship among genometypes and further improve mapping (Broman et al., 2003, 2019; Sul et al., 2016; Zhou and Stephens, 2012). We calculated linkage for 3,300 phenotypes using both the Haley-Knott (H-K) method that does not correct for kinship and the linear mixed model (LMM) mapping that does correct for kinship (Broman et al., 2019). Genome-wide significance was estimated based on 5,000 permutations and the improvement is substantial (Figure 3; Table S2). Using H-K, 638 phenotypes reach $p_{gw} < 0.05$ significance, whereas 763 reach the same criterion using the LMM, a 20% increase. To check that this was not due to phenotypes close to the threshold, we counted the number of phenotypes with a peak LOD $\geq$ 1 above threshold: 196 for H-K and 262 for LMM, a 33% increase. Note that there is neither a substantial average increase in LOD scores when using an LMM compared with H-K (a mean increase of 0.06 and a median of 0.07) nor is there a substantial change in the mean p_{gw} threshold (LMM 3.51 versus H-K 3.55; Table S2), demonstrating that LMM does not simply inflate all values.

Empirical mapping precision ranges from 0.25 to 5 Mb

We estimated the empirical precision of linkage maps using ~270,000 *cis*-acting expression QTLs (eQTLs) that have LOD scores greater than 3.0 and within 10 Mb of the mRNA transcription start site (Figure 4A). These criteria are conservative because an acceptance window of ±10 Mb will include a small number of regional trans-acting variants. However, this approach guards against overly optimistic estimates of precision. The mean precision for 17,805 eQTLs with LODs above 5.0 is about 1.0 Mb using only 60–80 BXD strains. The median precision is about 0.50 Mb. In other words, the offset between the location of a gene and the location of the marker with the highest LOD is generally less than 1 Mb using half the available strains. However, regional variation in precision is high (Figures 4A and 4B), and intervals with low precision have either few markers or low recombination rates—a problem that is currently being addressed by generating sequence-based infinite marker maps (Ashbrook et al., 2018b). There may also be regions that contain large structural variants—duplication, inversions, insertions—compared with the reference genome of C57BL/6J.

It is possible to achieve subcentimorgan mapping precision using only half of the full set of BXD strains (Figure 4)—a mean resolution of 500 Kb and a median resolution of 250 Kb for these Mendelian traits. Three factors contribute to this precision: (1) well-balanced distribution of only two haplotype across the genome (minor allele and haplotype frequencies close to 0.5), (2) the ability to boost the effective heritability of traits by resampling (Belknap, 1998), and (3) the high density of recombination junctions captured collectively within the BXD family (Figure 1C). Surprisingly, this level of map precision does not differ appreciably from that typical of the Collaborative Cross, the diversity outbred, or heterogeneous stock (HS), mainly because the minor

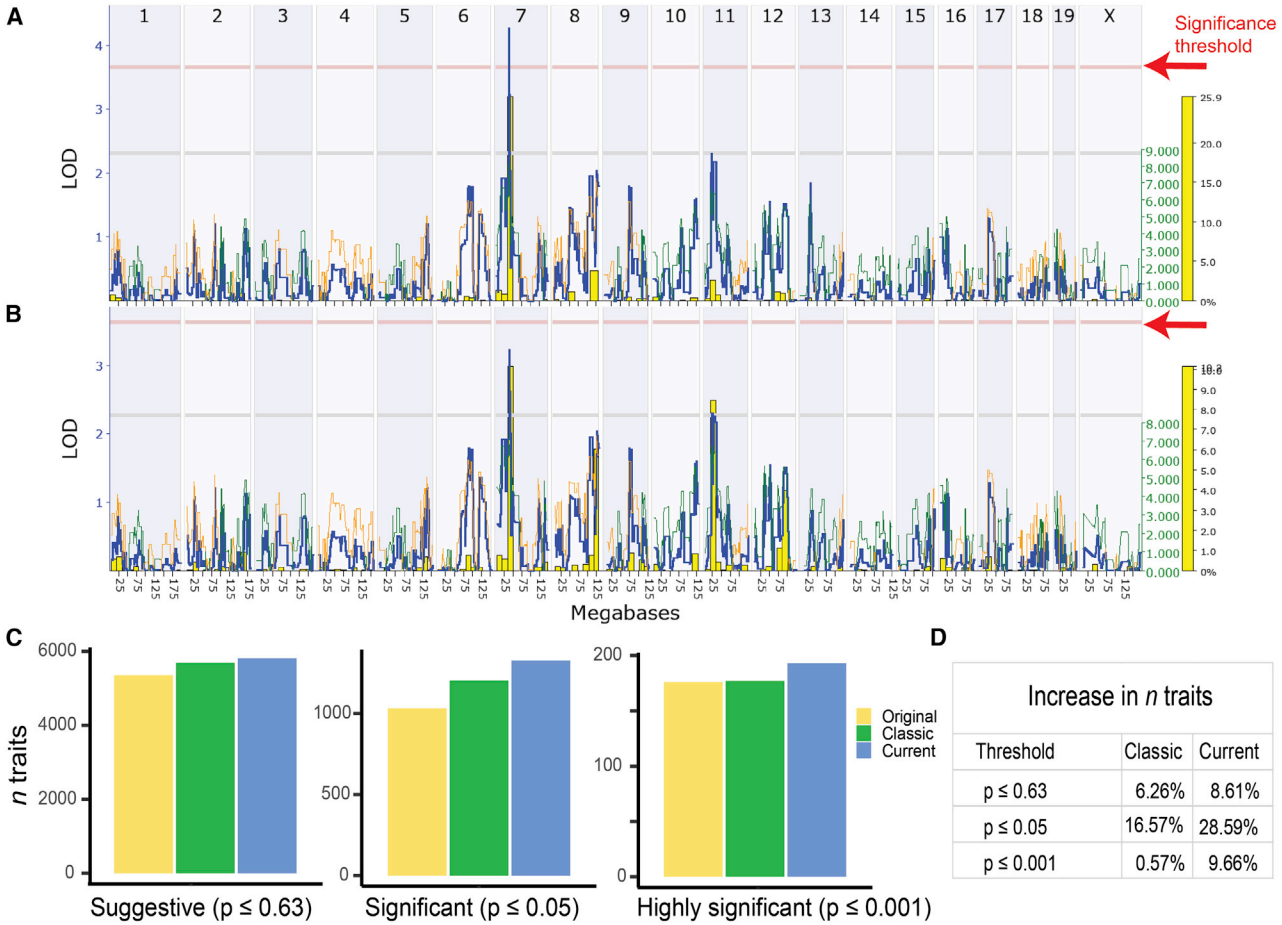

Figure 2. Improved, denser genotypes increase linkage across chromosomes, decades of work, and number of strains

(A and B) Comparison of current (A) and classic (B) genotypes for GeneNetwork: BXD_10666, cytotoxic T cell (CTL) response (5×10^9 PFU AdLacZ iv), measured as interleukin 6 (IL-6) cytokine expression (pg/ml), published by Zhang et al. (2004), measured in 23 strains. A QTL on chromosome 7 is now significant (above the red p = 0.05 significance line) using the current genotypes (A), compared with the classic genotypes (B). Additional examples in Figure S2.

(C) Bar chart showing the number of phenotypes with a peak LOD score over the suggestive (LOD $\geq$ 2.2; p $\leq$ 0.63), significant (LOD $\geq$ 3.6; p $\leq$ 0.05), and highly significant (LOD $\geq$ 5.4; p $\leq$ 0.001) thresholds, using original, classic or current marker maps.

(D) Percentage increase in the number of traits passing the thresholds in the classic or current genotype map, compared with the original map.

haplotype frequencies (MHF) at loci are about 4-fold higher in BXDs than those typical of crosses made using many parental strains (MHFs of 0.05 to 0.10). In any case, precision much finer than this, while welcome, will often not be critical. The fuzzy functional boundaries of genes and the high density of sequence variants in linkage disequilibrium shifts the burden of proof from pure mapping to functional genomics, comparative analysis of human cohorts, complementary animal models, and direct pharmacological and genetic engineering (Smemo et al., 2014).

Mapping with BXDs has high power

How many replicates and strains are needed to detect and resolve QTLs? To start with the conclusion—it is almost always better to study small numbers of as many strains as possible (Andreux et al., 2012; Belknap, 1998). Studying as few as 2–4 replicates of each of 100 or more strains may seem counterintuitive to those used to typing knockouts and their wild-type controls, but a shallow comprehensive coverage is correct even for traits with low heritability. We have developed an R/shiny application (power.genenetwork.org) that provides guidance on power and resampling for designs using the BXD. When heritability is low, the gain in power by increasing replicates is significant up to an *n* of 4–6 cases per genometype but falls off rapidly at heritabilities above 0.5. At heritabilities above 0.5, power is high even using only two replicates (Figure 5). It may still make sense to add replicates for those strains that are in the tails of distributions because these extremes are both potential sampling errors but, if verified, are true research opportunities.

The effect sizes for loci in BXDs are usually high compared with those of outbred populations and humans. Two factors contribute to higher levels of genetic variance. First, nearly all loci in BXDs are homozygous. The lack of heterozygotes increases the genetic variance by 2-fold compared with a matched intercross and four times as much variance as in a backcross (Belknap, 1998). The drawback is obvious: we cannot detect dominance effects without adding at least a subset of

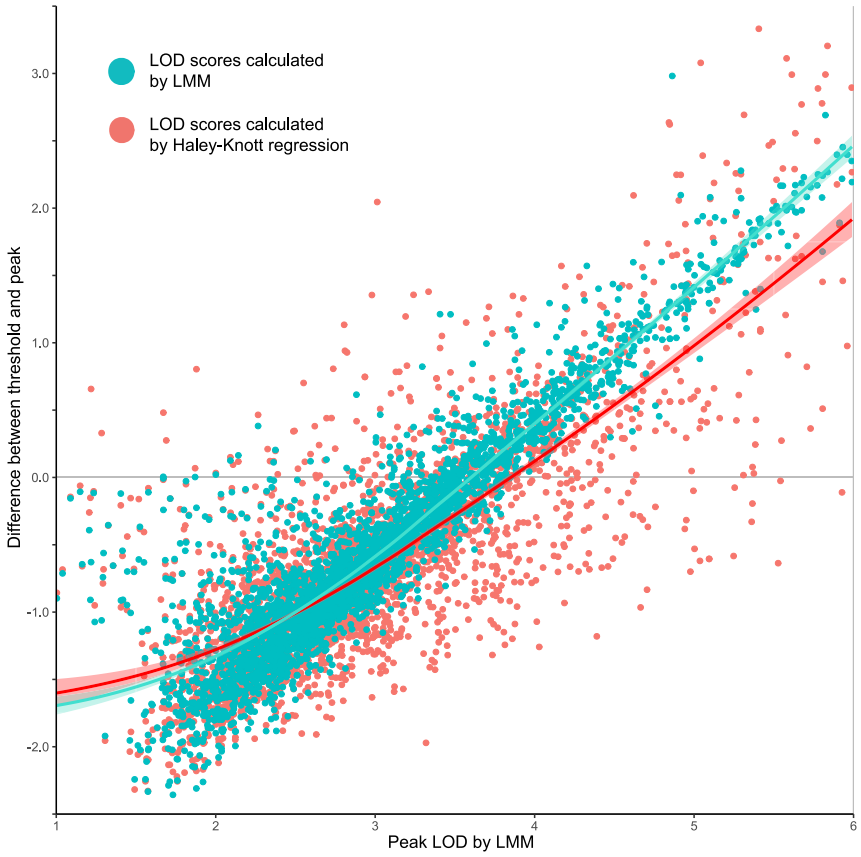

Figure 3. Improvement in mapping by using linear mixed models (LMMs) versus haley-knott regression (H-K)

Linkage measured in 3,300 phenotypes from GeneNetwork, and mapped using Haley-Knott regression or an LMM in R/qtl2 (Broman et al., 2003, 2019). y axis is the difference in LOD from the significance threshold (calculated by 5,000 permutations) and the peak LOD. x axis is the peak LOD, calculated by LMM, so that points can be directly compared. Each point represents a phenotype, LMM results are plotted in turquoise and the Haley-Knott are plotted in red. A smoother was fitted in the same colors for comparison. Underlying data in Table S2.

heterozygous loci of the type we can generate easily with sets of F1 progeny and a diallel cross. The second contributor is replicability (Figure 5). When the effect size is treated as the proportion of total variance explained by genometypes, this increases as the environmental "noise" is suppressed by replication (Belknap, 1998). To demonstrate this effect, we collected all *cis*-eQTLs (n = 6,867) for a brain gene expression data set—55,683 measurements across 129 individuals and 37 BXD strains (GeneNetwork.org: ID GN381). We calculated the proportion of variance in expression explained by *cis*-eQTL markers using either each individual as a unit or each strain mean as a unit. The mean proportion of variance explained was greatly increased by the latter approach (Figure S3), and for variants with smaller effect sizes ($z \leq 0.2$; n = 800) a 2.5-fold improvement was observed. The average proportion of explained variance was about 1.5-fold higher when using strain means than when using individual values. Across all *cis*-eQTLs, the effect size using an individual as unit was 0.49, whereas using strain means was 0.66—a 56% increase.

Substrains and epochs allow rapid determination of causal variants

Although kinship and epoch structure of BXDs have been framed as a problem that needs to be solved using LMMs, kinship can also provide unique opportunities. In some cases, substrain sets show substantial phenotypic differences and can be used to generate the so-called "reduced complexity crosses" (Bryant et al., 2020). Due to the low genetic variation between substrain sets, causal variants can often be identified rapidly, and these sets can be used as models of disease (Chang et al., 2006; Cook et al., 2006; Rosen et al., 2013).

Each epoch of BXDs has been generated using unique B6 and D2 parents. Although nominally considered identical, these parents have inherited small numbers of spontaneous mutations (Figure 1A). These *de novo* sequence variants that accumulated between epochs lead to expression variation and differences in phenotypes, and due to their very small number, these variants are especially tractable. It becomes straightforward to define the quantitative trait nucleotides causing this higher order variation among epochs using reduced complexity crosses and coisogenic crosses (Heiker et al., 2014; Kirkpatrick and Bryant, 2014; Kumar et al., 2013; Mulligan et al., 2008). Well-defined epoch effect variants include *Gpnmb* in vision (Lu et al., 2011, 2019), *Gabra2* in CNS function (Figure 6; (Hawkins et al., 2020; Mulligan et al., 2019), and *Taar1* in methamphetamine abuse (Reed et al., 2017; Shi et al., 2016; Stafford et al., 2019).

DISCUSSION

Kinship and genetic drift

The expanded family of BXDs is a well powered resource for both forward and reverse genetic analyses of genome-to-phenome linkage. As this family has grown, relations among individual strains have become complex, requiring the use of LMMs (Arends et al., 2010; Sul et al., 2016; Zhou and Stephens, 2014) or nonparametric equivalents such as mixed random forests (Stephan et al., 2015) that account for kinship, epoch, and other cofactors. The family has kinship at several levels. First, there is random fixation and selection of alleles and haplotypes that occur during inbreeding (Williams et al., 2001). Second, there is a strong kinship among AI-derived strains, including BXD43 to BXD102 and BXD160 to BXD186 (Peirce et al., 2004), that have their origins mainly in the 8–14 generations of intercrossing that preceded inbreeding (Darvasi, 1998). Third, there are coisogenic pairs such as BXD29/TyJ and BXD29-Tlr4<lps-2J>/J

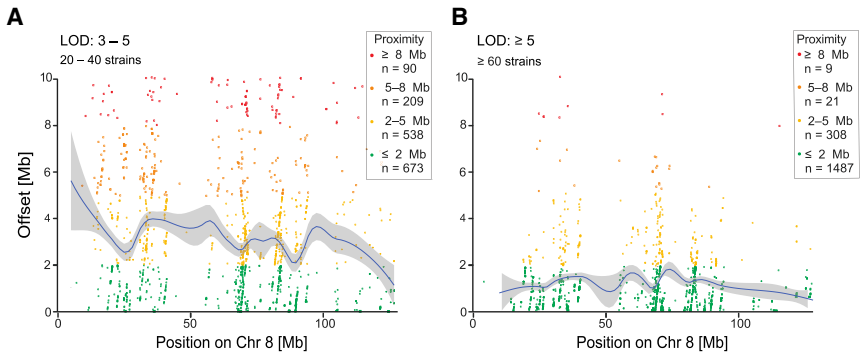

Figure 4. Empirical QTL mapping precision estimated using *Cis*-eQTLs

Each point shows the distance between a probe and an eQTL.

(A) Precision achieved when mapping modest QTLs with LOD scores of 3–5 and using only 20–40 strains and 2–4 samples per strain. Number of replicates within each strain will affect the precision to a greater extent for QTLs with low heritability and modest LOD scores by increasing the effective heritability (Belknap, 1998). eQTL studies generally use only 1–4 replicates, so precision values are conservative. Mean offset across the genome is 4.67 Mb, median 2.72 Mb.

(B) Precision achieved when mapping highly significant QTLs with LOD > 5, using 60–80 strains, with > 2 replicates. Mean offset is 1.38 Mb, median 0.76 Mb.

(Rosen et al., 2013). One member of each pair has at least one highly penetrant spontaneous mutation. Other recombinant congenic strains result from breeding admixture, for example, the two BXD48s and the three BXD73s. These small sibships are interesting and useful and have been used to quickly identify causal mutations for immune disfunction in *Tlr4* (Cook et al., 2006) and for retinal degeneration in *Cep290* (Chang et al., 2006). Finally, there is kinship that results from the epochs of BXD production.

Improved power and precision of mapping BXD phenotypes—A 50-year path of progress

The first set of BXDs was used to map traits to sparse linkage groups, not chromosomes, in the early 1970s (Taylor et al., 1973). While postmillennial mapping resources are obviously far better, there is still much room for improvement. We compared three sequential sets of markers used to map the BXDs—the original genotypes from 2001 (Williams et al., 2001), the set used from 2005 to 2016 (the "classic" set) (Shifman et al., 2006), and the new genotype file released here. This most recent set increased the yield of loci detected at genome-wide significance by about ~25%.

We highlight two key points. First, using the genetic resources we present in this paper, it is possible to map quantitative traits with modest LOD scores with good precision, even when using comparatively small numbers of strains (n = 25 to 50; e.g., Chintalapudi et al., 2017; Houtkooper et al., 2013; Williams et al., 1998). Second, an effective way to transition from QTLs to causal genes and biological processes is to take advantage of complementary resources, including other murine mapping resources, efficient *in vitro* and *in vivo* screens (Houtkooper et al., 2013; Williams and Auwerx, 2015), and human GWAS data (Ashbrook et al., 2014a, 2015b; Jha et al., 2018b, 2018a; Koutnikova et al., 2009). The brute-force way to improve precision and power is to just phenotype larger numbers of BXD strains. With 140 strains and nearly 20,000 easily made F1 hybrids, traits can be mapped with subcentimorgan precision.

Identification of causal variants is becoming easier, thanks to improved genotypes (Rau et al., 2015; Shifman et al., 2006; Simecek et al., 2017; Williams et al., 2001), full genome sequence (Keane et al., 2011; Wang et al., 2016b), powerful omics resources (King et al., 2015; Williams et al., 2016, 2018; Wu et al., 2014), better Bayesian fine-mapping methods (Gonzales and Palmer, 2014; Mulligan et al., 2017; Wakefield, 2008; Wellcome Trust Case Control Consortium et al., 2012; Zhou and Stephens 2014), and efficient molecular validation methods (Li et al., 2018). BXDs have been used to define specific genes and variants corresponding to well over 20 QTLs, including two tightly linked genes, *Iigp2* and *Irgb10*, for *Chlamydia* infectivity (Miyairi et al., 2007, 2012), *Fmn2* as a master controller of tRNA synthetases in neurons (Mozhui et al., 2008), *Ubp1* for blood pressure (Koutnikova et al., 2009), *Hc* for H5N1 influenza resistance (Boon et al., 2014), *Comt* as a master controller of neuropharmacological traits (Li et al., 2010), *Alpl* for hypophosphatasia (Andreux et al., 2012), *Mrps5* for longevity (Houtkooper et al., 2013), *Bckdhb* for maple syrup urine disease, *Dhtkd1* for diabetes (Wu et al., 2014), *Hp1bp3* for cognitive aging (Neuner et al., 2016), *Ahr* for locomotor activity (Williams et al., 2014), and *Gabra2* and *Taar1* for behavioral traits (Mulligan et al., 2019; Reed et al., 2017).

The BXD family Is the largest and the most deeply phenotyped mammalian genetic reference panel

Data sets for the BXD encompass multiple levels—from single molecules to complex behavioral repertoires and to traits measured with environmental perturbations, including exposure to alcohol and drugs of abuse (Dickson et al., 2019; Mulligan et al., 2018; Théberge et al., 2019; Zhou et al., 2018), infectious agents (Boon et al., 2014; Chella Krishnan et al., 2016; Russo et al., 2015), dietary modifications (Fleet et al., 2016; Jha et al., 2018b, 2018a; Jones and Jellen, 2017; Reyes Fernandez et al., 2016; Rodrigues et al., 2017), stress (Diessler et al., 2018; Jung et al., 2017), and even as a function of age (Sandoval-Sierra et al., 2020; Williams et al., 2020). This phenome is linked to over 500 publications, including state-of-the-art proteome, metabolome (Williams et al., 2018, 2020), epigenome (Baker et al., 2019; Sandoval-Sierra et al., 2019), and metagenome (Perez-Munoz et al., 2019) data sets available on GeneNetwork.org.

Genetic research is moving rapidly toward causal modeling of health and disease risk and toward the predicted efficacy of prevention and interventions (Hood and Flores, 2012). The size and depth of the BXD phenome makes it a strong foundation for this experimental precision medicine. The genetic architecture of traits can be dissected and causal relations among networks

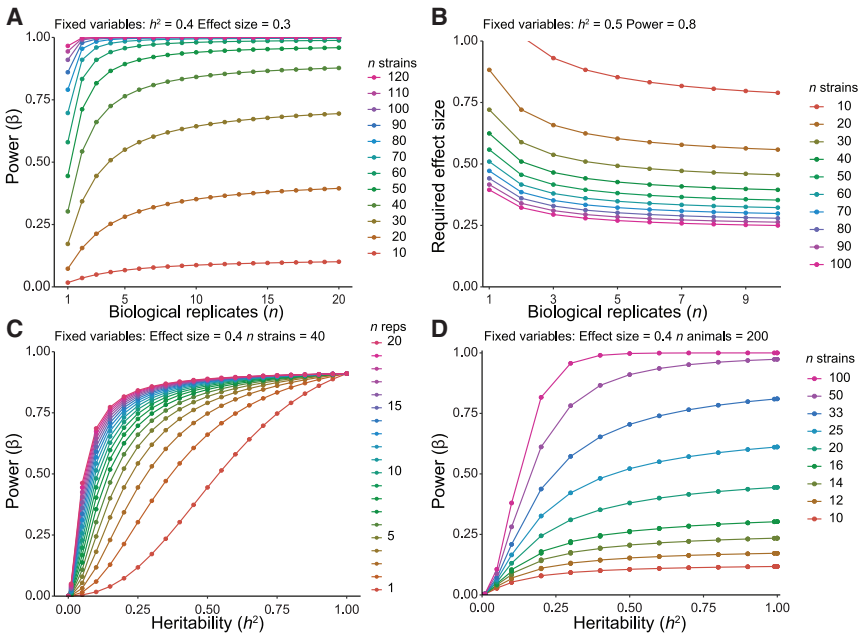

Figure 5. Relations between power, numbers of strains and replicates in BXD

(A) Power to detect a given effect for *n* replicates and *n* strains. Locus effect size and heritability constant at 0.3 and 0.4, respectively. More replicates increase power but beyond four replicates the gain in power is marginal.

(B) Effect size detectable, given *n* replicates and strains. Heritability and power kept at 0.5 and 0.8.

(C) Power to detect a given effect size, dependent upon heritability and *n* biological replicates. Locus effect size and *n* strains kept at 0.4 and 40. With very low h^2 there is no power, and with lower h^2 the gain in power is stronger with increasing number of replicates compared with high h^2. Beyond four replicates, the gain in power is marginal.

(D) Power to detect a given effect size, dependent upon heritability, with a total of 200 animals. These range from 100 lines with 2 replicates to 10 lines with 20 replicates. In contrast to what one might assume, even for low levels of h^2, power always increases with more lines rather than more replicates. Effect size kept at 0.4. Values generated using qtlDesign (Sen et al., 2007). An empirical example of power in the BXD can be found in Figure S3.

can be explored (Roy et al., 2020). The ability to resample genomes across a stable reference family enables the expansion of data in almost any direction. Improved phenotype ontologies and better access to FAIR data (Wilkinson et al., 2016) enables cross-species translation. This also enables more diverse researcher communities to engage in replicable trans-disciplinary studies of genome-phenome prediction.

Future directions: Epistasis, pleiotropy, epigenetics, gene-by-environment interactions

Most work using the BXD family has had the simple goal of defining single gene variants and processes that contribute to heritable differences in disease risk. But we need to tackle a more important problem—the complex interplay among sets of variants, constellations of phenotypes, and different treatments and environments (Ashbrook et al., 2014b; Mulligan and Williams, 2015; Williams and Auwerx, 2015; Williams et al., 2016). BXDs are now sufficiently large to evaluate predictive models of biological processes of these more complex types (Li et al., 2008; Miyairi et al., 2012; Wang et al., 2016b) and to map epistatic interactions. But the power to fit complex models can be greatly amplified by crossing BXD strains to generate a diallel cross of up to 19,460 F1 progeny. All isogenic BXD F1 lines are replicable and have entirely defined genomes. They have the advantage of being non-inbred and heterozygous. A subset of this massive diallel cross (DAX) can be generated efficiently from a central repository for cost-effective cell-based assays, and for matched *in vivo* predictive validation. Any part of this DAX can be used to study the parent-of-origin and sex chromosome effects (Ashbrook and Hager, 2013). It is also practical to cross BXDs to other lines—for example, to the humanized 5XFAD line used in Alzheimer's disease research by Kaczorowski and colleagues (Neuner et al., 2019a, 2019b). By crossing humanized or genetically engineered lines on a single genetic background to a diverse but defined subset of BXDs, variants that modify traits—such as memory loss in the 5XFAD model—can be mapped and can provide valuable information on disease process and outcomes. Kaczorowski and colleagues have also evaluated the efficacy of "reverse translation" from human to mouse (Neuner et al., 2019a). They generated a polygenic genetic risk score using 21 human genes that increase Alzheimer's risk and showed that allele dosage was significantly associated with cognitive outcomes in mice. This demonstrates that naturally occurring variation in these networks may have overlapping effects in mouse and humans.

In humans, tens of thousands of variants have now been mapped to haplotype blocks of under 100 Kb (Huan et al., 2015; Schizophrenia Working Group of the Psychiatric Genomics, 2014; Wood et al., 2014). We need to revamp murine genetic resources to contribute in more effective and complementary ways, in an era flooded with GWAS hits. How are rodent resources best repositioned to help deliver on the still unmet and much more integrative promises of predictive genetics and personalized precision health care? The short answer is that we need large genetically complex resources with matched multiscalar and multisystems phenome data. We need resources that define and test progressively more sophisticated computational models as a function of genometype, stage, and exposure. We need replicable populations with the same intrinsic genetic complexity and admixture as humans but without the challenges of clinical research—high cost, marginal compliance and control, confidentiality, and of most importance, ethical constraints on designs and interventions. Rodent populations are ideally positioned to be replicable and extensible testbeds with which the power, accuracy, and limits of precision health care can be defined.

Conclusions

Systems genetics using rodent models has been revitalized over the last decade, thanks to several resources, including the BXD

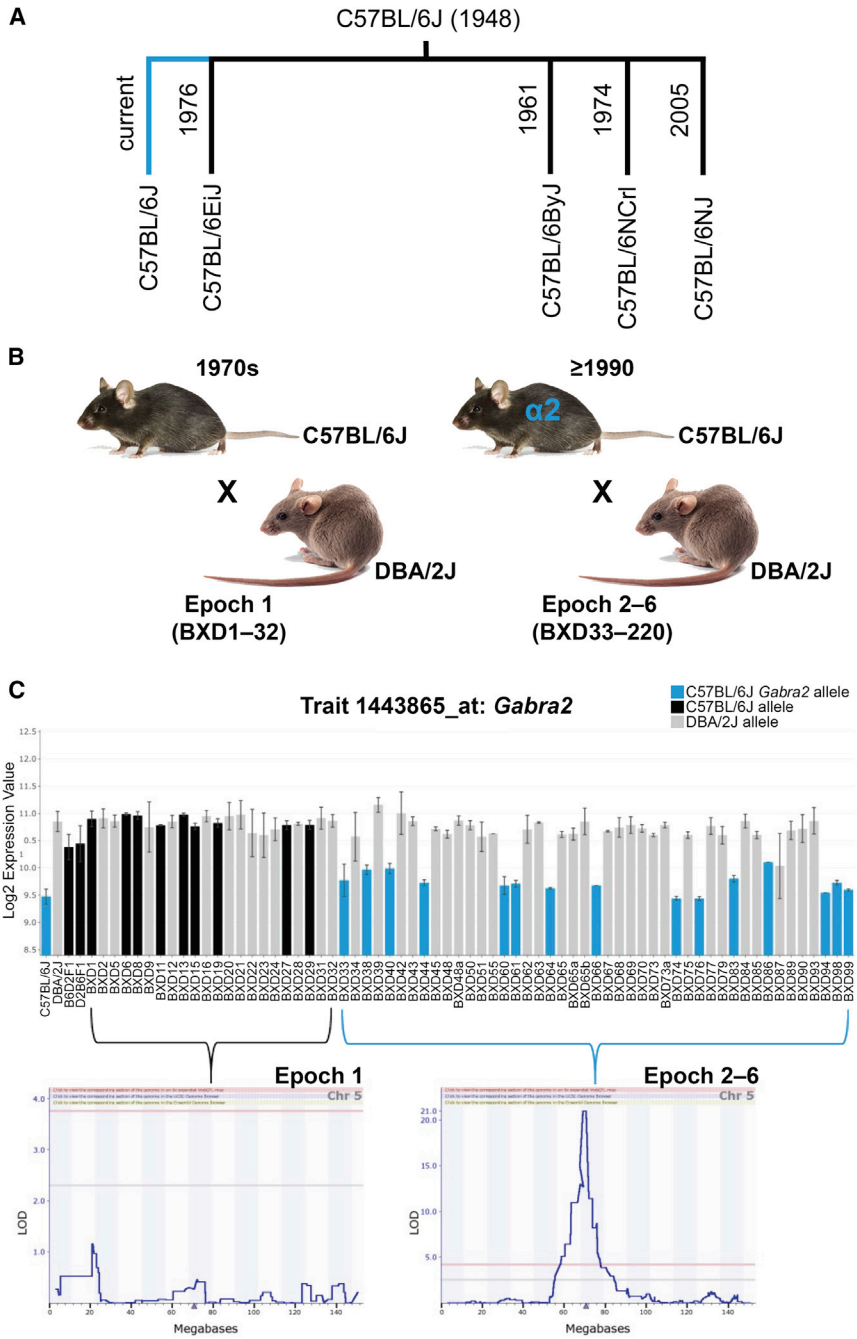

Figure 6. Impact of epoch-specific mutations in the BXD family

(A) A private mutation (single nucleotide intronic deletion) in the Gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 2 *(Gabra2)* gene occurred in the C57BL/6J lineage (turquoise line).

(B) The spontaneous mutation was fixed in Jackson Laboratory C57BL/6J foundation stock after separation of the C57BL/6EiJ substrain (1976), and (B) after the creation of the first BXD epoch (1970s). All BXD epochs (2–6) created after 1990 are derived from separate crosses between C57BL/6J and DBA/2J mice and segregate for the C57BL/6J *Gabra2* mutant allele (shown in turquoise).

(C) Inheritance of the C57BL/6J mutant allele (turquoise) results in a reduction of brain *Gabra2* expression levels relative to the ancestral C57BL/6J (black) or DBA/2J (gray) allele. Allele color coded based on *Gabra2* genotype at rs13478320. *Gabra2* mRNA expression values (Genetwork trait 1443865_at) are shown for the hippocampus but are replicated in other brain regions and at the protein level. QTL mapping in epoch 1 or later epochs demonstrates how the absence or presence of segregating C57BL/6J mutant *Gabra2* alleles controls gene expression.

family (Peirce et al., 2004), the Hybrid Mouse Diversity Panel (Bennett et al., 2015; Ghazalpour et al., 2012), and the Collaborative Cross (Churchill et al., 2004; Morgan and Welsh, 2015; Schughart and Williams, 2017; Valdar et al., 2006). The main limitation has been relatively modest mapping power and precision—a simple problem caused by small numbers of strains. With 140 strains now readily available, and expandable to 19,460 isogenic F1s, the extended BXD family has overcome this problem. All phenome data are multiplicatively useful, and like a fine vintage, old data get better with age.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cels.2020.12.002.

ACKNOWLEDGMENTS

Thanks to Dr. Abraham Palmer for providing B6D2 AI progeny (G8 and G9) to make BXD160 through BXD186. We thank Drs. Gerald McClearn and Lisa Tarantino for 60 eighth-generation (G8) AI progeny from Pennsylvania State University that contributed to epoch 3. We thank Dr. Benjamin A. Taylor for initiating the BXD and for continued support and encouragement. We thank Dr. Saunak Sen for consultation on power calculations and qtlDesign. We thank members of the EPFL Center for Phenogenomics, in particular, Drs. Xavier Warot and Emilie Gesina for sharing data from their BXD colony, and Erik Soehnel from Scionics for assistance in obtaining the full breeding statistics.

The UTHSC Center for Integrative and Translational Genomics (CITG) has supported production of the BXD colony at UTHSC and will continue to support this colony for the duration of the grant. The CITG also provides generous support for computer hardware and programming associated with GeneNetwork, and our Galaxy and UCSC Genome Browser instances.

We thank the support of the UT Center for Integrative and Translational Genomics and funds from the UT-ORNL Governor's Chair, NIDA grant P30DA044223 and NIAAA U01 AA013499, U01 AA016662, and U01 AA014425 for the work at UTHSC. Work in the Auwerx lab on the BXDs is supported by the École Polytechnique Fédérale de Lausanne, the ERC (AdG-787702), the SNSF (310030B-160318), the AgingX program of the Swiss Initiative for Systems Biology (RTD 2013/153), and the NIH (R01AG043930). The BXD Resource at the Jackson Laboratory is supported by NIH P40 OD011102 awarded to Dr. Cathleen M. Lutz.

The data sets generated and/or analyzed during the current study are available in the GeneNetwork repository, https://www.genenetwork.org/.

AUTHOR CONTRIBUTIONS

Conceptualization, R.W.W., L.L., and J.A.; Methodology, D.G.A., D.A., P.P., J.A., L.L., and R.W.W.; Software D.G.A., D.A., P.P., A.G.C., L.L., and R.W.W.; Formal Analysis, D.G.A., D.A., E.G.W., R.H., and S.S.; Investigation; S.R., C.M.L., A.V., C.J.B., J.F.I., and M.S.M.; Resources, C.M.L. and A.V.; Data Curation, D.G.A., D.A., and A.G.C., and R.W.W.; Writing – Original Draft, D.G.A., D.A., R.H., and R.W.W.; Writing – Review & Editing, D.G.A., D.A., P.P., M.K.M., S.R., E.G.W., R.H., J.H., L.L., and R.W.W.; Visualization, D.G.A., D.A., and E.G.W.; Supervision, J.A., L.L., and R.W.W.; Funding Acquisition, R.W.W., J.A., and C.M.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 29, 2020
Revised: August 29, 2020
Accepted: December 21, 2020
Published: January 19, 2021

REFERENCES

Andreux, P.A., Williams, E.G., Koutnikova, H., Houtkooper, R.H., Champy, M.F., Henry, H., Schoonjans, K., Williams, R.W., and Auwerx, J. (2012). Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. Cell *150*, 1287–1299.

Arends, D., Prins, P., Jansen, R.C., and Broman, K.W. (2010). R/qtl: high-throughput multiple QTL mapping. Bioinformatics *26*, 2990–2992.

Ashbrook, D.G., and Hager, R. (2013). Empirical testing of hypotheses about the evolution of genomic imprinting in mammals. Front. Neuroanat. *7*, 6.

Ashbrook, D.G., Williams, R.W., Lu, L., Stein, J.L., Hibar, D.P., Nichols, T.E., Medland, S.E., Thompson, P.M., and Hager, R. (2014a). Joint genetic analysis of hippocampal size in mouse and human identifies a novel gene linked to neurodegenerative disease. BMC Genomics *15*, 850.

Ashbrook, D.G., Delprato, A., Grellmann, C., Klein, M., Wetzel, R., Overall, R.W., and Badea, A. (2014b). Transcript co-variance with nestin in two mouse genetic reference populations identifies Lef1 as a novel candidate regulator of neural precursor cell proliferation in the adult hippocampus. Front. Neurosci. *8*, 418.

Ashbrook, D.G., Gini, B., and Hager, R. (2015a). Genetic variation in offspring indirectly influences the quality of maternal behaviour in mice. eLife *4*, e11814.

Ashbrook, D.G., Williams, R.W., Lu, L., and Hager, R. (2015b). A cross-species genetic analysis identifies candidate genes for mouse anxiety and human bipolar disorder. Front. Behav. Neurosci. *9*, 171.

Ashbrook, D.G., Sharmin, N., and Hager, R. (2017). Offspring genes indirectly influence sibling and maternal behavioural strategies over resource share. Proc. Biol. Sci. *284*, 20171059.

Ashbrook, D.G., Roy, S., Clifford, B.G., Riede, T., Scattoni, M.L., Heck, D.H., Lu, L., and Williams, R.W. (2018a). Born to cry: a genetic dissection of infant vocalization. Front. Behav. Neurosci. *12*, 250.

Ashbrook, D.G., Arends, D., Mulligan, M.K., Williams, E.G., Lutz, C., Valenzuela, A., Bohl, C., Ingels, J., McCarty, M., Centeno, A., et al (2018b). Sequencing the BXD family, a cohort for experimental systems genetics and precision medicine. IBANGS Meeting: The 20TH Annual Genes, Brain & Behavior Meeting, (Rochester, MN USA: International Behavioral and Neural Genetics Society) 2018, pp. 4–5. https://ibangs.memberclicks.net/assets/IBANGS2018ProgramWithAbstracts.pdf

Auffray, C., Charron, D., and Hood, L. (2010). Predictive, preventive, personalized and participatory medicine: back to the future. Genome Med. *2*, 57.

Baker, C.L., Walker, M., Arat, S., Ananda, G., Petkova, P., Powers, N.R., Tian, H., Spruce, C., Ji, B., Rausch, D., et al. (2019). Tissue-specific trans regulation of the mouse epigenome. Genetics *211*, 831–845.

Baud, A., Mulligan, M.K., Casale, F.P., Ingels, J.F., Bohl, C.J., Callebert, J., Launay, J.M., Krohn, J., Legarra, A., Williams, R.W., and Stegle, O. (2017). Genetic variation in the social environment contributes to health and disease. PLoS Genet. *13*, e1006498.

Belknap, J.K. (1998). Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behav. Genet. *28*, 29–38.

Belknap, J.K., Phillips, T.J., and O'Toole, L.A. (1992a). Quantitative trait loci associated with brain weight in the BXD/Ty recombinant inbred mouse strains. Brain Res. Bull. *29*, 337–344.

Belknap, J.K., Crabbe, J.C., Plomin, R., McClearn, G.E., Sampson, K.E., O'Toole, L.A., and Gora-Maslak, G. (1992b). Single-locus control of saccharin intake in BXD/Ty recombinant inbred (RI) mice: some methodological implications for RI strain analysis. Behav. Genet. *22*, 81–100.

Belknap, J.K., Metten, P., Helms, M.L., O'Toole, L.A., Angeli-Gade, S., Crabbe, J.C., and Phillips, T.J. (1993). Quantitative trait loci (QTL) applications to substances of abuse: physical dependence studies with nitrous oxide and ethanol in BXD mice. Behav. Genet. *23*, 213–222.

Bennett, B.J., Davis, R.C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R.R.S., Eskin, E., Yan, M., et al. (2015). Genetic architecture of atherosclerosis in mice: A systems genetics analysis of common inbred strains. PLoS Genet. *11*, e1005711.

Boon, A.C.M., Williams, R.W., Sinasac, D.S., and Webby, R.J. (2014). A novel genetic locus linked to pro-inflammatory cytokines after virulent H5N1 virus infection in mice. BMC Genomics *15*, 1017.

Broman, K.W. (2005). The genomes of recombinant inbred lines. Genetics *169*, 1133–1146.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., and Churchill, G.A. (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics *211*, 495–502.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889–890.

Bryant, C.D., Smith, D.J., Kantak, K.M., Nowak, T.S., Williams, R.W., Damaj, M.I., Redei, E.E., Chen, H., and Mulligan, M.K. (2020). Facilitating complex trait analysis via reduced complexity crosses. Trends Genet. *36*, 549–562.

Carhuatanta, K.A.K., Shea, C.J.A., Herman, J.P., and Jankord, R. (2014). Unique genetic loci identified for emotional behavior in control and chronic stress conditions. Front. Behav. Neurosci. *8*, 341.

Chang, B., Khanna, H., Hawes, N., Jimeno, D., He, S., Lillo, C., Parapuram, S.K., Cheng, H., Scott, A., Hurd, R.E., et al. (2006). In-frame deletion in a novel centrosomal/ciliary protein CEP290/NPHP6 perturbs its interaction with RPGR and results in early-onset retinal degeneration in the rd16 mouse. Hum. Mol. Genet. *15*, 1847–1857.

Chella Krishnan, K., Mukundan, S., Alagarsamy, J., Hur, J., Nookala, S., Siemens, N., Svensson, M., Hyldegaard, O., Norrby-Teglund, A., and Kotb, M. (2016). Genetic architecture of group a streptococcal necrotizing soft tissue infections in the mouse. PLoS Pathog. *12*, e1005732.

Chesler, E.J., Wang, J., Lu, L., Qu, Y., Manly, K.F., and Williams, R.W. (2003). Genetic correlates of gene expression in recombinant inbred strains: a relational model system to explore neurobehavioral phenotypes. Neuroinformatics *1*, 343–357.

Chintalapudi, S.R., Maria, D., Di Wang, X., Bailey, J.N.C., NEIGHBORHOOD consortium, International Glaucoma Genetics consortium, Hysi, P.G., Wiggs, J.L., Williams, R.W., and Jablonski, M.M. (2017). Systems genetics identifies a role for Cacna2d1 regulation in elevated intraocular pressure and glaucoma susceptibility. Nat. Commun. *8*, 1755.

Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W., et al. (2004). The collaborative cross, a community resource for the genetic analysis of complex traits. Nat. Genet. *36*, 1133–1137.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. *74*, 829–836.

Cook, D.N., Whitehead, G.S., Burch, L.H., Berman, K.G., Kapadia, Z., Wohlford-Lenane, C., and Schwartz, D.A. (2006). Spontaneous mutations in recombinant inbred mice: mutant toll-like receptor 4 (Tlr4) in BXD29 mice. Genetics *172*, 1751–1755.

Crow, J.F. (2007). Haldane, Bailey, Taylor and recombinant-inbred lines. Genetics *176*, 729–732.

Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. Nat. Genet. *18*, 19–24.

Dickson, P.E., Miller, M.M., Calton, M.A., Bubier, J.A., Cook, M.N., Goldowitz, D., Chesler, E.J., and Mittleman, G. (2016). Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. Psychopharmacology (Berl.) *233*, 701–714.

Dickson, P.E., Roy, T.A., McNaughton, K.A., Wilcox, T.D., Kumar, P., and Chesler, E.J. (2019). Systems genetics of sensation seeking. Genes Brain Behav. *18*, e12519.

Diessler, S., Jan, M., Emmenegger, Y., Guex, N., Middleton, B., Skene, D.J., Ibberson, M., Burdet, F., Götz, L., Pagni, M., et al. (2018). A systems genetics resource and analysis of sleep regulation in the mouse. PLoS Biol. *16*, e2005750.

Falconer, D.S., and Mackay, T.F.C. (1996). Introduction to Quantitative Genetics (Longman).

Fleet, J.C., Replogle, R.A., Reyes-Fernandez, P., Wang, L., Zhang, M., Clinkenbeard, E.L., and White, K.E. (2016). Gene-by-Diet interactions affect serum 1,25-dihydroxyvitamin D levels in male BXD recombinant inbred mice. Endocrinology *157*, 470–481.

Ghazalpour, A., Rau, C.D., Farber, C.R., Bennett, B.J., Orozco, L.D., van Nas, A., Pan, C., Allayee, H., Beaven, S.W., Civelek, M., et al. (2012). Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mamm. Genome *23*, 680–692.

Gonzales, N.M., and Palmer, A.A. (2014). Fine-mapping QTLs in advanced intercross lines and other outbred populations. Mamm. Genome *25*, 271–292.

Graybeal, C., Bachu, M., Mozhui, K., Saksida, L.M., Bussey, T.J., Sagalyn, E., Williams, R.W., and Holmes, A. (2014). Strains and stressors: an analysis of touchscreen learning in genetically diverse mouse strains. PLoS One *9*, e87745.

Grisel, J.E., Belknap, J.K., O'Toole, L.A., Helms, M.L., Wenger, C.D., and Crabbe, J.C. (1997). Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains. J. Neurosci. *17*, 745–754.

Grizzle, W.E., Mountz, J.D., Yang, P.A., Xu, X., Sun, S., Van Zant, G.E., Williams, R.W., Hsu, H.C., and Zhang, H.G. (2002). BXD recombinant inbred mice represent a novel T cell-mediated immune response tumor model. Int. J. Cancer *101*, 270–279.

Hager, R., Lu, L., Rosen, G.D., and Williams, R.W. (2012). Genetic architecture supports mosaic brain evolution and independent brain-body size regulation. Nat. Commun. *3*, 1079.

Haldane, J.B., and Waddington, C.H. (1931). Inbreeding and linkage. Genetics *16*, 357–374.

Haley, C.S., and Knott, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity (Edinb) *69*, 315–324.

Hawkins, N.A., Nomura, T., Duarte, S., Williams, R.W., Homanics, G.E., Mulligan, M.K., Contractor, A., and Kearney, J.A. (2020). Gabra2 is a genetic modifier of Dravet syndrome in mice. bioRxiv https://www.biorxiv.org/content/10.1101/2020.04.19.048546v1.

Hayes, K.S., Hager, R., and Grencis, R.K. (2014). Sex-dependent genetic effects on immune responses to a parasitic nematode. BMC Genomics *15*, 193.

Heiker, J.T., Kunath, A., Kosacka, J., Flehmig, G., Knigge, A., Kern, M., Stumvoll, M., Kovacs, P., Blüher, M., and Klöting, N. (2014). Identification of genetic loci associated with different responses to high-fat diet-induced obesity in C57BL/6N and C57BL/6J substrains. Physiol. Genomics *46*, 377–384.

Hood, L., and Flores, M. (2012). A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. N. Biotechnol. *29*, 613–624.

Hood, L., and Friend, S.H. (2011). Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat. Rev. Clin. Oncol. *8*, 184–187.

Houtkooper, R.H., Mouchiroud, L., Ryu, D., Moullan, N., Katsyuba, E., Knott, G., Williams, R.W., and Auwerx, J. (2013). Mitonuclear protein imbalance as a conserved longevity mechanism. Nature *497*, 451–457.

Huan, T., Meng, Q., Saleh, M.A., Norlander, A.E., Joehanes, R., Zhu, J., Chen, B.H., Zhang, B., Johnson, A.D., Ying, S., et al. (2015). Integrative network analysis reveals molecular mechanisms of blood pressure regulation. Mol. Syst. Biol. *11*, 799.

Jha, P., McDevitt, M.T., Gupta, R., Quiros, P.M., Williams, E.G., Gariani, K., Sleiman, M.B., Diserens, L., Jochem, A., Ulbrich, A., et al. (2018a). Systems analyses reveal physiological roles and genetic regulators of liver lipid species. Cell Syst. *6*, 722–733.e6.

Jha, P., McDevitt, M.T., Halilbasic, E., Williams, E.G., Quiros, P.M., Gariani, K., Sleiman, M.B., Gupta, R., Ulbrich, A., Jochem, A., et al. (2018b). Genetic regulation of plasma lipid species and their association with metabolic phenotypes. Cell Syst. *6*, 709–721.e6.

Jones, B.C., and Jellen, L.C. (2017). Systems genetics analysis of iron and its regulation in brain and periphery. Methods Mol. Biol. *1488*, 467–480.

Jones, L.C., McCarthy, K.A., Beard, J.L., Keen, C.L., and Jones, B.C. (2006). Quantitative genetic analysis of brain copper and zinc in BXD recombinant inbred mice. Nutr. Neurosci. *9*, 81–92.

Jung, S.H., Brownlow, M.L., Pellegrini, M., and Jankord, R. (2017). Divergence in Morris water Maze-based cognitive performance under chronic stress Is associated with the hippocampal whole transcriptomic modification in mice. Front. Mol. Neurosci. *10*, 275.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

King, R., Lu, L., Williams, R.W., and Geisert, E.E. (2015). Transcriptome networks in the mouse retina: an exon level BXD RI database. Mol. Vis. *21*, 1235–1251.

Kirkpatrick, S.L., and Bryant, C.D. (2014). Behavioral architecture of opioid reward and aversion in C57BL/6 substrains. Front. Behav. Neurosci. *8*, 450.

Knoll, A.T., Jiang, K., and Levitt, P. (2018). Quantitative trait locus mapping and analysis of heritable variation in affiliative social behavior and co-occurring traits. Genes Brain Behav. *17*, e12431.

Koutnikova, H., Laakso, M., Lu, L., Combe, R., Paananen, J., Kuulasmaa, T., Kuusisto, J., Häring, H.U., Hansen, T., Pedersen, O., et al. (2009). Identification of the UBP1 locus as a critical blood pressure determinant using a combination of mouse and human genetics. PLoS Genet. *5*, e1000591.

Kumar, V., Kim, K., Joseph, C., Kourrich, S., Yoo, S.H., Huang, H.C., Vitaterna, M.H., de Villena, F.P.-M., Churchill, G., Bonci, A., and Takahashi, J.S. (2013). C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. Science *342*, 1508–1512.

Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. *11*, 241–247.

Lee, G.H., Bennett, L.M., Carabeo, R.A., and Drinkwater, N.R. (1995). Identification of hepatocarcinogen-resistance genes in DBA/2 mice. Genetics *139*, 387–395.

Li, H., Wang, X., Rukina, D., Huang, Q., Lin, T., Sorrentino, V., Zhang, H., Bou Sleiman, M., Arends, D., McDaid, A., et al. (2018). An integrated systems genetics and omics toolkit to probe gene function. Cell Syst. *6*, 90–102.e4.

Li, R., Svenson, K.L., Donahue, L.R.B., Peters, L.L., and Churchill, G.A. (2008). Relationships of dietary fat, body composition, and bone mineral density in inbred mouse strain panels. Physiol. Genomics *33*, 26–32.

Li, Z., Mulligan, M.K., Wang, X., Miles, M.F., Lu, L., and Williams, R.W. (2010). A transposon in Comt generates mRNA variants and causes widespread expression and behavioral differences among mice. PLoS One *5*, e12181.

Lu, H., Wang, X., Pullen, M., Guan, H., Chen, H., Sahu, S., Zhang, B., Chen, H., Williams, R.W., Geisert, E.E., et al. (2011). Genetic dissection of the Gpnmb network in the eye. Invest. Ophthalmol. Vis. Sci. *52*, 4132–4142.

Lu, Y., Zhou, D., Lu, H., Xu, F., Yue, J., Tong, J., and Lu, L. (2019). Investigating a downstream gene of Gpnmb using the systems genetics method. Mol. Vis. *25*, 222–236.

McGinnis, J.F., Lerious, V., Pazik, J., and Elliott, R.W. (1993). Chromosomal assignment of the recoverin gene and cancer-associated retinopathy. Mamm. Genome *4*, 43–45.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One *7*, e39191.

Miles, C.M., and Wayne, M. (2008). Quantitative trait locus (QTL) analysis. Nat. Educ. *1*, 208.

Miyairi, I., Tatireddigari, V.R.R.A., Mahdi, O.S., Rose, L.A., Belland, R.J., Lu, L., Williams, R.W., and Byrne, G.I. (2007). The p47 GTPases Iigp2 and Irgb10 regulate innate immunity and inflammation to murine Chlamydia psittaci infection. J. Immunol. *179*, 1814–1824.

Miyairi, I., Ziebarth, J., Laxton, J.D., Wang, X., van Rooijen, N., Williams, R.W., Lu, L., Byrne, G.I., and Cui, Y. (2012). Host genetics and Chlamydia disease: prediction and validation of disease severity mechanisms. PLoS One *7*, e33781.

Morgan, A.P., Fu, C.P., Kao, C.Y., Welsh, C.E., Didion, J.P., Yadgary, L., Hyacinth, L., Ferris, M.T., Bell, T.A., Miller, D.R., et al. (2015). The mouse universal genotyping array: From substrains to subspecies. G3 (Bethesda) *6*, 263–279.

Morgan, A.P., and Welsh, C.E. (2015). Informatics resources for the Collaborative Cross and related mouse populations. Mamm. Genome *26*, 521–539.

Mozhui, K., Ciobanu, D.C., Schikorski, T., Wang, X., Lu, L., and Williams, R.W. (2008). Dissection of a QTL hotspot on mouse distal chromosome 1 that modulates neurobehavioral phenotypes and gene expression. PLoS Genet. *4*, e1000260.

Mulligan, M.K., Abreo, T., Neuner, S.M., Parks, C., Watkins, C.E., Houseal, M.T., Shapaker, T.M., Hook, M., Tan, H., Wang, X., et al. (2019). Identification of a functional non-coding variant in the GABAA Receptor α2 subunit of the C57BL/6J mouse reference genome: major implications for neuroscience research. Front. Genet. *10*, 188.

Mulligan, M.K., Dubose, C., Yue, J., Miles, M.F., Lu, L., and Hamre, K.M. (2013). Expression, covariation, and genetic regulation of miRNA Biogenesis genes in brain supports their role in addiction, psychiatric disorders, and disease. Front. Genet. *4*, 126.

Mulligan, M.K., Mozhui, K., Prins, P., and Williams, R.W. (2017). GeneNetwork: A toolbox for systems genetics. Methods Mol. Biol. *1488*, 75–120.

Mulligan, M.K., Ponomarev, I., Boehm, S.L., Owen, J.A., Levin, P.S., Berman, A.E., Blednov, Y.A., Crabbe, J.C., Williams, R.W., Miles, M.F., and Bergeson, S.E. (2008). Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. Genes Brain Behav. *7*, 677–689.

Mulligan, M.K., and Williams, R.W. (2015). Systems genetics of behavior: a prelude. Curr. Opin. Behav. Sci. *2*, 108–115.

Mulligan, M.K., Zhao, W., Dickerson, M., Arends, D., Prins, P., Cavigelli, S.A., Terenina, E., Mormede, P., Lu, L., and Jones, B.C. (2018). Genetic contribution to initial and progressive alcohol intake among recombinant inbred strains of mice. Front. Genet. *9*, 370.

Neuner, S.M., Garfinkel, B.P., Wilmott, L.A., Ignatowska-Jankowska, B.M., Citri, A., Orly, J., Lu, L., Overall, R.W., Mulligan, M.K., Kempermann, G., et al. (2016). Systems genetics identifies Hp1bp3 as a novel modulator of cognitive aging. Neurobiol. Aging *46*, 58–67.

Neuner, S.M., Heuer, S.E., Huentelman, M.J., O'Connell, K.M.S., and Kaczorowski, C.C. (2019a). Harnessing genetic complexity to enhance translatability of Alzheimer's disease mouse models: a path toward precision medicine. Neuron *101*, 399–411.e5.

Neuner, S.M., Heuer, S.E., Zhang, J.G., Philip, V.M., and Kaczorowski, C.C. (2019b). Identification of pre-symptomatic gene signatures that predict resilience to cognitive decline in the genetically diverse AD-BXD model. Front. Genet. *10*, 35.

Oren, Y., Nachshon, A., Frishberg, A., Wilentzik, R., and Gat-Viks, I. (2015). Linking traits based on their shared molecular mechanisms. eLife *4*, e04346.

Palmer, A.A., Lessov-Schlaggar, C.N., Ponder, C.A., McKinnon, C.S., and Phillips, T.J. (2006). Sensitivity to the locomotor-stimulant effects of ethanol and allopregnanolone: a quantitative trait locus study of common genetic influence. Genes Brain Behav. *5*, 506–517.

Peirce, J.L., Lu, L., Gu, J., Silver, L.M., and Williams, R.W. (2004). A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet. *5*, 7.

Perez-Munoz, M.E., McKnite, A.M., Williams, E.G., Auwerx, J., Williams, R.W., Peterson, D.A., and Ciobanu, D.C. (2019). Diet modulates cecum bacterial diversity and physiological phenotypes across the BXD mouse genetic reference population. PLoS One *14*, e0224100.

Philip, V.M., Duvvuru, S., Gomero, B., Ansah, T.A., Blaha, C.D., Cook, M.N., Hamre, K.M., Lariviere, W.R., Matthews, D.B., Mittleman, G., et al. (2010). High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. Genes Brain Behav. *9*, 129–159.

Phillips, T.J., Belknap, J.K., Buck, K.J., and Cunningham, C.L. (1998). Genes on mouse chromosomes 2 and 9 determine variation in ethanol consumption. Mamm. Genome *9*, 936–941.

Rau, C.D., Parks, B., Wang, Y., Eskin, E., Simecek, P., Churchill, G.A., and Lusis, A.J. (2015). High-density genotypes of inbred mouse strains: improved power and precision of association mapping. G3 (Bethesda) *5*, 2021–2026.

Rédei, G.P. (2008a). Carter-Falconer mapping function. In Encyclopedia of Genetics, Genomics, Proteomics and Informatics (Springer), p. 276.

Rédei, G.P. (2008b). Haldane's mapping function. In Encyclopedia of Genetics, Genomics, Proteomics and Informatics (Springer), p. 836.

Rédei, G.P. (2008c). Kosambi's function. In Encyclopedia of Genetics, Genomics, Proteomics and Informatics (Springer), p. 1066.

Rédei, G.P. (2008d). Morgan. In Encyclopedia of Genetics, Genomics, Proteomics and Informatics (Springer), p. 1255.

Reed, C., Baba, H., Zhu, Z., Erk, J., Mootz, J.R., Varra, N.M., Williams, R.W., and Phillips, T.J. (2017). A spontaneous mutation in Taar1 impacts methamphetamine-related traits exclusively in DBA/2 mice from a single vendor. Front. Pharmacol. *8*, 993.

Reyes Fernandez, P.C., Replogle, R.A., Wang, L., Zhang, M., and Fleet, J.C. (2016). Novel genetic loci control calcium absorption and femur bone mass as well as their response to low calcium intake in male BXD recombinant inbred mice. J. Bone Miner. Res. *31*, 994–1002.

Rodrigues, B.A., Muñoz, V.R., Kuga, G.K., Gaspar, R.C., Nakandakari, S.C.B.R., Crisol, B.M., Botezelli, J.D., Pauli, L.S.S., da Silva, A.S.R., de Moura, L.P., et al. (2017). Obesity increases mitogen-activated protein kinase phosphatase-3 levels in the hypothalamus of mice. Front. Cell. Neurosci. *11*, 313.

Rodriguez, L.A., Plomin, R., Blizard, D.A., Jones, B.C., and McClearn, G.E. (1994). Alcohol acceptance, preference, and sensitivity in mice. I. Quantitative genetic analysis using BXD recombinant inbred strains. Alcohol. Clin. Exp. Res. *18*, 1416–1422.

Rosen, G.D., Azoulay, N.G., Griffin, E.G., Newbury, A., Koganti, L., Fujisaki, N., Takahashi, E., Grant, P.E., Truong, D.T., Fitch, R.H., et al. (2013). Bilateral subcortical heterotopia with partial callosal agenesis in a mouse mutant. Cereb. Cortex *23*, 859–872.

Rosen, G.D., Pung, C.J., Owens, C.B., Caplow, J., Kim, H., Mozhui, K., Lu, L., and Williams, R.W. (2009). Genetic modulation of striatal volume by loci on Chrs 6 and 17 in BXD recombinant inbred mice. Genes Brain Behav. *8*, 296–308.

Roy, S., Sleiman, M.B., Jha, P., Williams, E.G., Ingels, J.F., Chapman, C.J., McCarty, M.S., Hook, M., Sun, A., Zhao, W., et al. (2020). Gene-by-environmental modulation of longevity and weight gain in the murine BXD family. bioRxiv https://www.biorxiv.org/content/10.1101/776559v3.

Russo, L.M., Abdeltawab, N.F., O'Brien, A.D., Kotb, M., and Melton-Celsa, A.R. (2015). Mapping of genetic loci that modulate differential colonization by Escherichia coli O157:H7 TUV86-2 in advanced recombinant inbred BXD mice. BMC Genomics *16*, 947.

Sandoval-Sierra, J.V., Helbing, A.H.B., Williams, E.G., Ashbrook, D.G., Roy, S., Williams, R.W., and Mozhui, K. (2019). Influence of body weight at young adulthood on the epigenetic clock and lifespan in the BXD murine family. bioRxiv https://www.biorxiv.org/content/10.1101/791582v1.

Sandoval-Sierra, J.V., Helbing, A.H.B., Williams, E.G., Ashbrook, D.G., Roy, S., Williams, R.W., and Mozhui, K. (2020). Body weight and high-fat diet are associated with epigenetic aging in female members of the BXD murine family. Aging Cell *19*, e13207.

Schizophrenia Working Group of the Psychiatric Genomics (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Schughart, K., and Williams, R.W. (2017). Systems Genetics (Springer).

Schüssler-Fiorenza Rose, S.M., Contrepois, K., Moneghetti, K.J., Zhou, W., Mishra, T., Mataraso, S., Dagan-Rosenfeld, O., Ganz, A.B., Dunn, J., Hornburg, D., et al. (2019). A longitudinal big data approach for precision health. Nat. Med. *25*, 792–804.

Seecharan, D.J., Kulkarni, A.L., Lu, L., Rosen, G.D., and Williams, R.W. (2003). Genetic control of interconnected neuronal populations in the mouse primary visual system. J. Neurosci. *23*, 11178–11188.

Sen, S., Satagopan, J.M., Broman, K.W., and Churchill, G.A. (2007). R/qtlDesign: inbred line cross experimental design. Mamm. Genome *18*, 87–93.

Shi, X., Walter, N.A.R., Harkness, J.H., Neve, K.A., Williams, R.W., Lu, L., Belknap, J.K., Eshleman, A.J., Phillips, T.J., and Janowsky, A. (2016). Genetic polymorphisms affect mouse and human trace amine-associated Receptor 1 function. PLoS One *11*, e0152581.

Shifman, S., Bell, J.T., Copley, R.R., Taylor, M.S., Williams, R.W., Mott, R., and Flint, J. (2006). A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol. *4*, e395.

Simecek, P., Forejt, J., Williams, R.W., Shiroishi, T., Takada, T., Lu, L., Johnson, T.E., Bennett, B., Deschepper, C.F., Scott-Boyer, M.P., et al. (2017). High-resolution maps of mouse reference populations. G3 (Bethesda) *7*, 3427–3434.

Sloan, Z., Arends, D., W. Broman, K., Centeno, A., Furlotte, N., Nijveen, H., Yan, L., Zhou, X., W. Williams, R., and Prins, P. (2016). GeneNetwork: framework for web-based genetics. J. Open Source Softw. *1*, 25.

Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature *507*, 371–375.

Stafford, A.M., Reed, C., Baba, H., Walter, N.A., Mootz, J.R., Williams, R.W., Neve, K.A., Fedorov, L.M., Janowsky, A.J., and Phillips, T.J. (2019). Taar1 gene variants have a causal role in methamphetamine intake and response and interact with Oprm1. eLife *8*, e46472.

Stephan, J., Stegle, O., and Beyer, A. (2015). A random forest approach to capture genetic effects in the presence of population structure. Nat. Commun. *6*, 7432.

Sul, J.H., Bilow, M., Yang, W.Y., Kostem, E., Furlotte, N., He, D., and Eskin, E. (2016). Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. PLoS Genet. *12*, e1005849.

Taylor, B.A., Heiniger, H.J., and Meier, H. (1973). Genetic analysis of resistance to cadmium-induced testicular damage in mice. Proc. Soc. Exp. Biol. Med. *143*, 629–633.

Taylor, B.A., Wnek, C., Kotlus, B.S., Roemer, N., MacTaggart, T., and Phillips, S.J. (1999). Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. Mamm. Genome *10*, 335–348.

Théberge, E.T., Baker, J.A., Dubose, C., Boyle, J.K., Balce, K., Goldowitz, D., and Hamre, K.M. (2019). Genetic influences on the amount of cell death in the neural tube of BXD mice exposed to acute ethanol at midgestation. Alcohol. Clin. Exp. Res. *43*, 439–452.

Tsaih, S.W., Lu, L., Airey, D.C., Williams, R.W., and Churchill, G.A. (2005). Quantitative trait mapping in a diallel cross of recombinant inbred lines. Mamm. Genome *16*, 344–355.

Valdar, W., Flint, J., and Mott, R. (2006). Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics *172*, 1783–1797.

Wakefield, J. (2008). Reporting and interpretation in genome-wide association studies. Int. J. Epidemiol. *37*, 641–653.

Wang, J., Williams, R.W., and Manly, K.F. (2003). WebQTL: web-based complex trait analysis. Neuroinformatics *1*, 299–308.

Wang, L.S., Jiao, Y., Huang, Y., Liu, X.Y., Gibson, G., Bennett, B., Hamre, K.M., Li, D.W., Zhao, H.Y., Gelernter, J., et al. (2013). Critical evaluation of transcription factor Atf2 as a candidate modulator of alcohol preference in mouse and human populations. Genet. Mol. Res. *12*, 5992–6005.

Wang, L., Jiao, Y., Wang, Y., Zhang, M., and Gu, W. (2016a). Self-confirmation and ascertainment of the candidate genomic regions of complex trait loci - A none-experimental solution. PLoS One *11*, e0153676.

Wang, X., Pandey, A.K., Mulligan, M.K., Williams, E.G., Mozhui, K., Li, Z., Jovaisaite, V., Quarles, L.D., Xiao, Z., Huang, J., et al. (2016b). Joint mouse-human phenome-wide association to test gene function and disease risk. Nat. Commun. *7*, 10464.

Wang, J., Yoon, T.W., Read, R., Yi, A.K., Williams, R.W., and Fitzpatrick, E.A. (2020). Genetic variability of T cell responses in hypersensitivity pneumonitis identified using the BXD genetic reference panel. Am. J. Physiol. Lung Cell. Mol. Physiol. *318*, L631–L643.

Weimar, W.R., Lane, P.W., and Sidman, R.L. (1982). Vibrator (vb): a spinocerebellar system degeneration with autosomal recessive inheritance in mice. Brain Res. *251*, 357–364.

Wellcome Trust Case Control Consortium, Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M., Auton, A., Myers, S., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. *44*, 1294–1301.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. Sci. Data *3*, 160018.

Williams, E.G., and Auwerx, J. (2015). The convergence of systems and reductionist approaches in complex trait analysis. Cell *162*, 23–32.

Williams, E.G., Mouchiroud, L., Frochaux, M., Pandey, A., Andreux, P.A., Deplancke, B., and Auwerx, J. (2014). An evolutionarily conserved role for the aryl hydrocarbon receptor in the regulation of movement. PLoS Genet. *10*, e1004673.

Williams, E.G., Roy, S., Statzer, C., Ingels, J., Bohl, C., Hasan, M., Cuklina, J., Lu, L., Ewald, C.Y., Williams, R.W., et al. (2020). The molecular landscape of the aging mouse liver. bioRxiv https://www.biorxiv.org/content/10.1101/2020.08.20.222968v1.

Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science *352*, aad0189.

Williams, E.G., Wu, Y., Wolski, W., Kim, J.Y., Lan, J., Hasan, M., Halter, C., Jha, P., Ryu, D., Auwerx, J., and Aebersold, R. (2018). Quantifying and localizing the mitochondrial proteome across five tissues in a mouse population. Mol. Cell. Proteomics *17*, 1766–1777.

Williams, R.W. (2009). Herding cats: the sociology of data integration. Front. Neurosci. *3*, 154–156.

Williams, R.W., Gu, J., Qi, S., and Lu, L. (2001). The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. Genome Biol. *2*, RESEARCH0046.

Williams, R.W., Strom, R.C., and Goldowitz, D. (1998). Natural variation in neuron number in mice is linked to a major quantitative trait locus on Chr 11. J. Neurosci. *18*, 138–146.

Williams, R.W., and Williams, E.G. (2017). Resources for systems genetics. In System Genetics: Methods and Protocols, K. Schughart and R.W. Williams, eds. (Springer), pp. 3–29.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.

Wu, Y., Williams, E.G., Dubuis, S., Mottis, A., Jovaisaite, V., Houten, S.M., Argmann, C.A., Faridi, P., Wolski, W., Kutalik, Z., et al. (2014). Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. Cell *158*, 1415–1430.

Yang, R.J., Mozhui, K., Karlsson, R.M., Cameron, H.A., Williams, R.W., and Holmes, A. (2008). Variation in mouse basolateral amygdala volume is associated with differences in stress reactivity and fear learning. Neuropsychopharmacology *33*, 2595–2604.

Zhang, H.-G., Hsu, H.-C., Yang, P.-A., Yang, X., Wu, Q., Liu, Z., Yi, N., and Mountz, J.D. (2004). Identification of multiple genetic loci that regulate adenovirus gene therapy. Gene Ther *11*, 4–14.

Zhang, W., Korstanje, R., Thaisz, J., Staedtler, F., Harttman, N., Xu, L., Feng, M., Yanas, L., Yang, H., Valdar, W., et al. (2012). Genome-wide association mapping of quantitative traits in outbred mice. G3 (Bethesda) *2*, 167–174.

Zhou, D., Zhao, Y., Hook, M., Zhao, W., Starlard-Davenport, A., Cook, M.N., Jones, B.C., Hamre, K.M., and Lu, L. (2018). Ethanol's effect on Coq7 expression in the hippocampus of mice. Front. Genet. *9*, 602.

Zhou, G., and Williams, R.W. (1999). Eye1 and Eye2: gene loci that modulate eye size, lens weight, and retinal area in the mouse. Invest. Ophthalmol. Vis. Sci. *40*, 817–825.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. *44*, 821–824.

Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods *11*, 407–409.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
GeneNetwork	Genenetwork.org	RRID:SCR_002388
Experimental Models: Organisms/Strains		
Mouse: BXD1/TyJ	JAX	RRID:000036
Mouse: BXD2/TyJ	JAX	RRID:000075
Mouse: BXD3	JAX	N/A
Mouse: BXD4	JAX	N/A
Mouse: BXD5/TyJ	JAX	RRID:000037
Mouse: BXD6/TyJ	JAX	RRID:000007
Mouse: BXD7	JAX	N/A
Mouse: BXD8/TyJ	JAX	RRID:000084
Mouse: BXD9/TyJ	JAX	RRID:000105
Mouse: BXD10	JAX	N/A
Mouse: BXD11/TyJ	JAX	RRID:000012
Mouse: BXD12/TyJ	JAX	RRID:000045
Mouse: BXD13/TyJ	JAX	RRID:000040
Mouse: BXD14/TyJ	JAX	RRID:000329
Mouse: BXD15/TyJ	JAX	RRID:000095
Mouse: BXD16/TyJ	JAX	RRID:000013
Mouse: BXD17	JAX	N/A
Mouse: BXD18/TyJ	JAX	RRID:000015
Mouse: BXD19/TyJ	JAX	RRID:000010
Mouse: BXD20/TyJ	JAX	RRID:000330
Mouse: BXD21/TyJ	JAX	RRID:000077
Mouse: BXD22/TyJ	JAX	RRID:000043
Mouse: BXD23/TyJ	JAX	RRID:000098
Mouse: BXD24/TyJ-Cep290rd16/J	JAX	RRID:000031 | rd16
Mouse: BXD24/TyJ	JAX	RRID:005243
Mouse: BXD25/TyJ	JAX	RRID:000081
Mouse: BXD26	JAX	
Mouse: BXD27/TyJ	JAX	RRID:000041
Mouse: BXD28/TyJ	JAX	RRID:000047
Mouse: BXD29/Ty	JAX	RRID:010981
Mouse: BXD29-Tlr4lps-2J/J	JAX	RRID:000029 | defective lipopolysaccharide response, 2J
Mouse: BXD30/TyJ	JAX	N/A
Mouse: BXD31/TyJ	JAX	RRID:000083
Mouse: BXD32/TyJ	JAX	RRID:000078
Mouse: BXD32/TyJ-Galctwi-5J/J	JAX	RRID:003613 | twitcher 5 Jackson
Mouse: BXD33/TyJ	JAX	RRID:003222
Mouse: BXD34/TyJ	JAX	RRID:003223
Mouse: BXD35/Ty	JAX	N/A
Mouse: BXD36/TyJ	JAX	RRID:003225
Mouse: BXD37/Ty	JAX	N/A
Mouse: BXD38/TyJ	JAX	RRID:003227
Mouse: BXD39/TyJ	JAX	RRID:003228

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: BXD40/TyJ	JAX	RRID:003229
Mouse: BXD41/Ty	JAX	N/A
Mouse: BXD42/TyJ	JAX	RRID:003230
Mouse: BXD43/RwwJ	JAX	RRID:007093
Mouse: BXD44/RwwJ	JAX	RRID:007094
Mouse: BXD45/RwwJ	JAX	RRID:007096
Mouse: BXD46	JAX	N/A
Mouse: BXD47	JAX	N/A
Mouse: BXD48/RwwJ	JAX	RRID:007097
Mouse: BXD48a/RwwJ	JAX	RRID:007139
Mouse: BXD49/RwwJ	JAX	RRID:007098
Mouse: BXD50/RwwJ	JAX	RRID:007099
Mouse: BXD51/RwwJ	JAX	RRID:007100
Mouse: BXD52	UTHSC	N/A
Mouse: BXD53/2RwwJ	JAX	RRID:017749
Mouse: BXD54	UTHSC	N/A
Mouse: BXD55/RwwJ	JAX	RRID:007103
Mouse: BXD56/RwwJ	JAX	RRID:007104
Mouse: BXD57	UTHSC	N/A
Mouse: BXD58	UTHSC	N/A
Mouse: BXD59	UTHSC	N/A
Mouse: BXD60/RwwJ	JAX	RRID:007105
Mouse: BXD61/RwwJ	JAX	RRID:007106
Mouse: BXD62/RwwJ	JAX	RRID:007107
Mouse: BXD63/RwwJ	JAX	RRID:007108
Mouse: BXD64/RwwJ	JAX	RRID:007109
Mouse: BXD65/RwwJ	JAX	RRID:007110
Mouse: BXD65a/RwwJ	JAX	RRID:007140
Mouse: BXD65b/RwwJ	JAX	RRID:009677
Mouse: BXD66/RwwJ	JAX	RRID:007111
Mouse: BXD67/RwwJ	JAX	RRID:007112
Mouse: BXD68/RwwJ	JAX	RRID:007113
Mouse: BXD69/RwwJ	JAX	RRID:007114
Mouse: BXD70/RwwJ	JAX	RRID:007115
Mouse: BXD71/RwwJ	JAX	RRID:007116
Mouse: BXD72	UTHSC	N/A
Mouse: BXD73/RwwJ	JAX	RRID:007117
Mouse: BXD73a/RwwJ	JAX	RRID:007124
Mouse: BXD73b/RwwJ	JAX	RRID:007146
Mouse: BXD74/RwwJ	JAX	RRID:007118
Mouse: BXD75/RwwJ	JAX	RRID:007119
Mouse: BXD76	UTHSC	N/A
Mouse: BXD77/RwwJ	JAX	RRID:007121
Mouse: BXD78/2RwwJ	JAX	RRID:024029
Mouse: BXD79/RwwJ	JAX	RRID:007123
Mouse: BXD80	UTHSC	N/A
Mouse: BXD81/RwwJ	JAX	RRID:007125
Mouse: BXD82	UTHSC	N/A
Mouse: BXD83/RwwJ	JAX	RRID:007126

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: BXD84/RwwJ	JAX	RRID:007127
Mouse: BXD85/RwwJ	JAX	RRID:007128
Mouse: BXD86/RwwJ	JAX	RRID:007129
Mouse: BXD87/RwwJ	JAX	RRID:007130
Mouse: BXD88/2RwwJ	JAX	RRID:017750
Mouse: BXD89/RwwJ	UTHSC	N/A
Mouse: BXD90/RwwJ	JAX	RRID:007133
Mouse: BXD91/2RwwJ	JAX	RRID:017751
Mouse: BXD92	UTHSC	N/A
Mouse: BXD93	UTHSC	N/A
Mouse: BXD94	UTHSC	N/A
Mouse: BXD95/RwwJ	JAX	RRID:007138
Mouse: BXD96	UTHSC	N/A
Mouse: BXD97	UTHSC	N/A
Mouse: BXD98/RwwJ	JAX	RRID:007141
Mouse: BXD99/RwwJ	JAX	RRID:007142
Mouse: BXD100/RwwJ	JAX	RRID:007143
Mouse: BXD101/RwwJ	JAX	RRID:007144
Mouse: BXD102/RwwJ	JAX	RRID:007145
Mouse: BXD103	UTHSC	N/A
Mouse: BXD104	UTHSC	N/A
Mouse: BXD105	UTHSC	N/A
Mouse: BXD106	UTHSC	N/A
Mouse: BXD107	UTHSC	N/A
Mouse: BXD108	UTHSC	N/A
Mouse: BXD109	UTHSC	N/A
Mouse: BXD110	UTHSC	N/A
Mouse: BXD111/RwwJ	JAX	RRID:030872
Mouse: BXD112	UTHSC	N/A
Mouse: BXD113/RwwJ	JAX	RRID:024030
Mouse: BXD114		N/A
Mouse: BXD115	UTHSC	N/A
Mouse: BXD116	UTHSC	N/A
Mouse: BXD117	UTHSC	N/A
Mouse: BXD118	UTHSC	N/A
Mouse: BXD119	UTHSC	N/A
Mouse: BXD120	UTHSC	N/A
Mouse: BXD121	UTHSC	N/A
Mouse: BXD122/RwwJ	JAX	RRID:030873
Mouse: BXD122a	UTHSC	N/A
Mouse: BXD123/RwwJ	JAX	RRID:025974
Mouse: BXD124/RwwJ	JAX	RRID:025975
Mouse: BXD125/RwwJ	JAX	RRID:024031
Mouse: BXD126	UTHSC	N/A
Mouse: BXD127	UTHSC	N/A
Mouse: BXD128/RwwJ	JAX	RRID:030874
Mouse: BXD128a/RwwJ	JAX	RRID:030875
Mouse: BXD129	UTHSC	N/A
Mouse: BXD130	UTHSC	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: BXD131	UTHSC	N/A
Mouse: BXD132	UTHSC	N/A
Mouse: BXD133	UTHSC	N/A
Mouse: BXD134	UTHSC	N/A
Mouse: BXD135	UTHSC	N/A
Mouse: BXD136	UTHSC	N/A
Mouse: BXD137	UTHSC	N/A
Mouse: BXD138	UTHSC	N/A
Mouse: BXD139	UTHSC	N/A
Mouse: BXD140	UTHSC	N/A
Mouse: BXD141/RwwJ	JAX	RRID:029865
Mouse: BXD142	UTHSC	N/A
Mouse: BXD143	UTHSC	N/A
Mouse: BXD144/RwwJ	JAX	RRID:025984
Mouse: BXD145	UTHSC	N/A
Mouse: BXD146	UTHSC	N/A
Mouse: BXD147	UTHSC	N/A
Mouse: BXD148	UTHSC	N/A
Mouse: BXD149	UTHSC	N/A
Mouse: BXD150/RwwJ	JAX	RRID:029867
Mouse: BXD151/RwwJ	JAX	RRID:029868
Mouse: BXD152/RwwJ	JAX	RRID:029869
Mouse: BXD153	UTHSC	N/A
Mouse: BXD154/RwwJ	JAX	RRID:025976
Mouse: BXD155	UTHSC	N/A
Mouse: BXD156/RwwJ	JAX	RRID:029871
Mouse: BXD156a	UTHSC	N/A
Mouse: BXD157/RwwJ	JAX	RRID:029872
Mouse: BXD158	UTHSC	N/A
Mouse: BXD159	UTHSC	N/A
Mouse: BXD160/RwwJ	JAX	RRID:029873
Mouse: BXD161/RwwJ	JAX	RRID:025985
Mouse: BXD162	UTHSC	N/A
Mouse: BXD163	UTHSC	N/A
Mouse: BXD164	UTHSC	N/A
Mouse: BXD165	UTHSC	N/A
Mouse: BXD166	UTHSC	N/A
Mouse: BXD167	UTHSC	N/A
Mouse: BXD168/RwwJ	JAX	RRID:029874
Mouse: BXD169/RwwJ	JAX	RRID:029876
Mouse: BXD170/RwwJ	JAX	RRID:029878
Mouse: BXD171/RwwJ	JAX	RRID:029879
Mouse: BXD172/RwwJ	JAX	RRID:029880
Mouse: BXD173	UTHSC	N/A
Mouse: BXD174	UTHSC	N/A
Mouse: BXD175	UTHSC	N/A
Mouse: BXD176	UTHSC	N/A
Mouse: BXD177/RwwJ	JAX	RRID:029881
Mouse: BXD178/RwwJ	JAX	RRID:029882

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: BXD179	UTHSC	N/A
Mouse: BXD180/RwwJ	JAX	RRID:029883
Mouse: BXD181	UTHSC	N/A
Mouse: BXD182	UTHSC	N/A
Mouse: BXD183	UTHSC	N/A
Mouse: BXD184/RwwJ	JAX	RRID:029884
Mouse: BXD185	UTHSC	N/A
Mouse: BXD186/RwwJ	JAX	RRID:029885
Mouse: BXD187/RwwJ	JAX	RRID:031974
Mouse: BXD188	UTHSC	N/A
Mouse: BXD189	UTHSC	N/A
Mouse: BXD190/RwwJ	JAX	RRID:030876
Mouse: BXD191/RwwJ	JAX	RRID:030877
Mouse: BXD192	UTHSC	N/A
Mouse: BXD193	UTHSC	N/A
Mouse: BXD194/RwwJ	JAX	RRID:030878
Mouse: BXD195/RwwJ	JAX	RRID:030879
Mouse: BXD196	UTHSC	N/A
Mouse: BXD197/RwwJ	JAX	RRID:030880
Mouse: BXD198	UTHSC	N/A
Mouse: BXD199/RwwJ	JAX	RRID:030881
Mouse: BXD200	UTHSC	N/A
Mouse: BXD201	UTHSC	N/A
Mouse: BXD202/RwwJ	JAX	RRID:031975
Mouse: BXD203	UTHSC	N/A
Mouse: BXD204	UTHSC	N/A
Mouse: BXD205	UTHSC	N/A
Mouse: BXD206	UTHSC	N/A
Mouse: BXD207	UTHSC	N/A
Mouse: BXD208	UTHSC	N/A
Mouse: BXD209	UTHSC	N/A
Mouse: BXD210/RwwJ	JAX	RRID:031976
Mouse: BXD211	UTHSC	N/A
Mouse: BXD212	UTHSC	N/A
Mouse: BXD213	UTHSC	N/A
Mouse: BXD214/RwwJ	JAX	RRID:030882
Mouse: BXD215/RwwJ	JAX	RRID:031977
Mouse: BXD216/RwwJ	JAX	RRID:031978
Mouse: BXD217	UTHSC	N/A
Mouse: BXD218	UTHSC	N/A
Mouse: BXD219	UTHSC	N/A
Mouse: BXD220	UTHSC	N/A
Mouse: BXD221	UTHSC	N/A
Mouse: BXD222	UTHSC	N/A
Mouse: BXD223	UTHSC	N/A
Mouse: BXD224	UTHSC	N/A
Mouse: BXD225	UTHSC	N/A
Mouse: BXD226	UTHSC	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: BXD227	UTHSC	N/A
Mouse: BXD228	UTHSC	N/A
Software and Algorithms		
GeneNetwork	Genenetwork.org	RRID:SCR_002388
BXDtools		https://github.com/DannyArends/BXDtools
BXD power calculator app	http://power.genenetwork.org/; this paper	https://github.com/Dashbrook/BXD_power_calculator_app
R/qtl2	Broman et al., 2019	RRID:SCR_018181; https://kbroman.org/qtl2/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents may be directed to, and will be fulfilled by, the corresponding author David Ashbrook (dashbrook@UTHSC.edu).

Materials availability

The BXD strains reported in this study are available from the JAX. JAX Stock No (RRIDs) can be found in the Key Resources Table and in Table S1.

Data and code availability

This paper analyzes existing, publicly available data. These data sets' accession numbers are provided in the Key Resource Table, and throughout the manuscript. Genotype files can be found at http://www.genenetwork.org/webqtl/main.py?FormID=sharinginfo&GN_AccessionId=600.

GeneNetwork.org original code is publicly available at https://github.com/genenetwork/genenetwork2 and https://github.com/genenetwork/genenetwork1.

The scripts used to generate the figures reported in this paper are available in the BXDtools package (https://github.com/DannyArends/BXDtools) or using ggplot2 (version 3.3.2; https://cran.r-project.org/web/packages/ggplot2/index.html) and their use is described in the STAR Methods. Code for computing marker positions and visualization of recombinations is available in the BXDtools package (https://github.com/DannyArends/BXDtools).

Power estimates can be made using http://power.genenetwork.org/, and code is at https://github.com/Dashbrook/BXD_power_calculator_app.

The R/qtl2 software can be found here https://kbroman.org/qtl2/, and its use is described in the STAR Methods.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Between 2009 and 2010 we initiated 74 BXD strains (BXD104 to BXD186). We initiated another 34 strains in 2013 (BXD187 to BXD220). We used both conventional F2 intercrosses (n = 88) and AI progeny (n = 20) to make the 108 new lines. BXD160 through BXD186 were derived from unique matings of AI progeny gifted us by Dr. Abraham Palmer at G8 and G9 late in 2010 (Figures 1 and S1; Table S1).

In cases of low reproductive fitness, we often attempted to rescue lines by outcrossing young male BXDs to C57BL/6J females followed by three or more sequential backcrosses (N3) to the RI male to produce progeny enriched for the BXD genome. In more recent cases, pairs of at-risk strains were crossed to produce RIX progeny (Tsaih et al., 2005; Williams et al., 2001), which were then inbred by sibling mating. BXD221 to BXD228 are RIX-derived strains of this type (Figure S1; Table S1).

Unless otherwise specified, all animals used in this study were raised in a closed-barrier pathogen-free vivarium at the University of Tennessee Health Science Center. All BXD strains are available under a standard material transfer agreement; the most important limitation being that they cannot be sold or distributed without approval of the Jackson Laboratory or UTHSC. Availability information on all BXD lines can be found in Table S1.

METHOD DETAILS

Genetic map construction and genotype error correction

Individuals from each of the 198 BXD lines were genotyped in late 2011 and in late 2015 using the Neogen/GeneSeek MUGA or GigaMUGA arrays. These were combined with previous genotypes generated using Affymetrix and MUGA platforms. Unknown

genotypes were imputed as B (C57BL/6J-like) or D (DBA/2J-like) or were called as H (heterozygous) if the genotype was uncertain.

The full data set contains 37,000 markers and was optimized for mapping efficiency by excluding markers showing identical strain distribution patterns. Markers that flank sites of recombination were retained. As a result, strain distribution patterns are often defined by proximal-distal marker pairs. Whenever possible, we verified and updated genotypes of the original BXD strains (BXD1 - BXD102) to reflect those of stock available from the Jackson Laboratory (Rau et al., 2015) and from Petr Simecek and Gary Churchill; http://cgd.jax.org/data sets/diversityarray.shtml).

Genotypes for all strains were smoothed and curated to remove highly implausible (double) recombination events, e.g. unsupported genotypes (singletons) that introduced two recombination events over less than 100 kb (Williams et al., 2001). In general, we imputed unknown or heterozygous genotypes on the basis of flanking markers. Undefined genotypes between recombinations were coded as heterozygous, and telomere genotypes were imputed using the closest flanking marker.

In the newer BXD strains many regions are still heterozygous. Generated genotypes using standard platforms still show regions of low marker density and a high frequency of recombinations. These regions of low marker density were filled with imputed genotypes using *cis*-eQTLs of genes in the problematic intervals. Microsatellites and *cis*-eQTL genotypes were generated by the Williams/Lu laboratory.

The assembled and error-checked genotype file includes 7,324 markers for 191 independent strains, and 7 substrains, has been available since January 2017 at http://www.genenetwork.org/genotypes/BXD.geno and is the default genetic map used when QTL mapping on GeneNetwork (http://www.genenetwork.org).

Genome-wide visualization of recombinations

For each chromosome we compute the number of observed recombinations per strain from the start of the chromosome (nRecS), ignoring the heterozygous and unknown genotypes. We then visualize this by plotting the number of recombinations for each BXD strain per chromosome using a color gradient. Strain numbers are used for the ordering on the X-axis, and horizontal lines are added to visually separate the different epochs and annotate the epochs on the X-axis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Computing centimorgan positions

For each chromosome we compute the number of observed recombinations from one position to the next, ignoring the heterozygous and unknown genotypes (nRecP). We then compute the recombination fraction (R) by dividing the number of recombinations observed (nRecP) by the population size used to generate the genetic map. The recombination fractions (R) are then deflated using the formula:

$$r = R/(4 - 6R),$$

which is the inverse of the formula $R = 4r/(1+6r)$ for a recombinant inbred line by sibling mating. We then use the function imf.cf available in the R package R/qtl (Arends et al., 2010; Broman et al., 2003) to convert from centiMorgan using the Carter-Falconer map function (Rédei, 2008a). Positions obtained start at 0 on each chromosome, requiring all positions to be shifted so that the first marker on the map matches the known centiMorgan position of the first marker on the genetic map. The old genetic map and the mouse map converter (http://cgd.jax.org/mousemapconverter/) were used to obtain centiMorgan positions for the first marker on a chromosome, all position on a chromosome are shifted to align the genetic map to the known starting positions of the markers. Code for computing marker positions is available in the BXDtools package (https://github.com/DannyArends/BXDtools), and positions can be recomputed using different map functions when required (e.g. Haldane (Rédei, 2008b), Kosambi (Rédei, 2008c), Morgan (Rédei, 2008d)).

Comparing genotype files

Over the past four decades many genotype files have been used for QTL mapping of the BXDs. The first carefully vetted genome-wide map (which we call the 'original') was generated in 2001, and mainly made up of 1,578 microsatellite markers (Wang et al., 2003; Williams et al., 2001). The second 'classic' map was used between 2005–2016 and contained 3,811 SNPs from various array platforms (Shifman et al., 2006). The current third-generation file is based on 7,321 informative SNPs described above.

For the 2001 markers, Mouse Genome Informatics was used to update 1419 microsatellites to the current assembly. This left 159 orphaned markers without mm10 positions. We used the NCBI Probe database (www.ncbi.nlm.nih.gov/probe/) to search for these 159 markers, and extracted primer sequence for 107. We ran Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi) to determine where in the mm10 genome they would amplify. For the remaining 52 markers, we used BLAST to align the target sequence to identify the mm10 position. This left 11 markers which Primer-BLAST either predicted a different chromosome would be amplified, or which did not have any mm10 genomic targets.

These 'original' genotypes only exist for the first 36 BXD strains (BXD1-42), and so the remaining strains were imputed (BXD24a and BXD 43+), using the genotype at the microsatellite's position in the 'current' post-2017 genotyping file. To determine our power to detect QTLs, we ran mapping using QTL Reaper for the 7562 phenotypes collected for the BXD, and the number of phenotypes which

had a LOD greater than three threshold values: 2.2, 3.6 and 5.4. These correspond to suggestive, significant, and highly significant LOD values. The larger the number of phenotypes above these thresholds, the greater our power to detect QTLs.

Precision and resolution

To estimate the empirical precision of mapping across the genome we have extracted data from 40 large-scale genetic studies of gene expression of the BXD family. These studies used between 21 and 79 strains with variable numbers of within-strain replicates over 27 tissues. We defined eQTLs that map within ±20 Mb of the gene associated with the transcript's measurement as *cis*-acting expression QTLs (*cis*-eQTLs). Unlike a standard F2 intercross, this ±20 Mb window is roughly equivalent to a recombination distance of ±40 cM in the highly recombinant BXD family, and the statistical association between markers this far apart is generally quite low (r^2 less than 0.3).

Precision was estimated using the offset between the location of the gene (using the 5' end of the probe or probe set as a reference point), and the location of the SNP with the highest LOD score. Again, this is conservative, as the causal variant is unlikely to be exactly at the 5' end of the gene. There are often two or more neighboring SNPs with equally high LOD scores and we simply take the most proximal marker to compute offset. The light blue local regression (LOESS) smoother curve (Cleveland, 1979) was computed using a window with a size of 0.333% of the genome. Smoothing was carried out and data was plotted using the *ggplot2* R package.

QTL mapping

QTL mapping using GeneNetwork has been described in detail elsewhere (Mulligan et al., 2017). However, in brief, quantitative trait loci (QTLs) are segments of the genome affecting a particular phenotype (Falconer and Mackay, 1996). QTL mapping, identifying QTLs to explain the genetic basis of complex traits, relies on being able to make correlations between genetic markers and phenotypic traits in a population. Individuals are scored for their phenotype for a particular trait, and their genotype at a marker. If there is a difference in mean phenotype between those individuals with one genotype s the other than we can infer that there is a QTL linked to that marker. If there is no difference between the means we can conclude that the locus does not influence the phenotype in that population (Falconer and Mackay, 1996; Miles and Wayne, 2008).

Due to the very high density of markers, the mapping algorithm used to map BXD data sets has been modified and is a mixture of simple marker regression, linear interpolation, and standard Haley-Knott interval mapping (Haley and Knott, 1992). When two adjacent markers have identical strain distribution patterns, they will have identical linkage statistics, as will the entire interval between these two markers (assuming complete and error-free haplotype data for all strains). On a physical map the LOD and the additive effect values will, therefore, be constant over this physical interval. Between neighboring markers that have different strain distribution patterns and that are separated by 1 cM or more, we use a conventional interval mapping method (Haley-Knott) combined with a Haldane estimate of genetic distance. When the interval is less than 1 cM, we simply interpolate linearly between markers based on a physical scale between those markers. The result of this mixture mapping algorithm is a linkage map of a trait that has an unusual profile that is particularly striking on a physical (Mb) scale, with many plateaus, abrupt linear transitions between plateaus, and a few regions with the standard graceful curves typical of interval maps.

Prior to the 2017 release of the genotypes described in this manuscript interval mapping in GeneNetwork relied on 3,795 informative SNP markers across all autosomes and the X chromosome (Wang et al., 2003). These markers were generated using the MUGA array in 2011, along with earlier generated genotypes on the Affymetrix and Illumina platforms (Shifman et al., 2006). As described above, loci are identified in GeneNetwork by the computation of a likelihood ratio statistic score (LRS), and significance was determined using at least 5,000 permutations of the phenotype data.

Updated QTL mapping methods, such as R/qtl (Broman et al., 2003, 2019), Multiple QTL mapping (Arends et al., 2010), GEMMA (Zhou and Stephens, 2012) and pyLMM (Sul et al., 2016), have been implemented on the GeneNetwork site (Sloan et al., 2016). The LMM based methods for kinship between BXD strains by computing a genetic kinship matrix given specific strains used in an analysis, allowing for precise kinship correction and improved QTL mapping performance when mapping any of the BXD phenotypes in the GeneNetwork database.

R/qtl2 (Broman et al., 2003, 2019) was also used to carry out QTL mapping, independently, in the R environment. The functions *scan1* and *scan1perm* were used to calculate linkage for all phenotypes, and permutation thresholds. For Haley-Knott mapping no kinship correction was used, and for LMM the *calc_kinship* function was used, with leave one chromosome out (LOCO).

Power calculation

Power calculations were carried out based on the method of Sen and colleagues (Sen et al., 2007) and implemented in the R program *qtlDesign* based on the h^2_{RIX} defined by Belknap (Belknap, 1998), which we renamed h^2_{ix} as it is applicable to any isogenic strain, not just recombinant inbreds. The *Detectable* function was used, to determine the power available with an RI design to detect a locus of a given effect size.

For calculation of the h^2_{ix}, four variables need to be known: heritability (h^2), the number of within-strain replicates (biological replicates), the number of strains used, and the locus effect size (the proportion of total genetic variance explained by the locus). To calculate h^2, the genomic variance (gen.var) was kept constant at 1, and the environmental variance (env.var) was varied to produce h^2 values ranging between 0 and 1. The number of biological replicates and the number of strains are self-explanatory, and were

capped at 10 and 150 respectively, since >10 replicates produced a marginal increase in power, and because there are a maximum of 150 BXD lines.

The locus effect size is the amount of the variance which is due to genetics (gen.var) explained by a single QTL. That is, a value of 0.2 would mean that a QTL explains 20% of the genetic variance, and a value of 1.0 indicates a Mendelian locus (one QTL which explains all of the genetic variance).

The power given is the ability of the experiment to correctly detect a true positive QTL, given the other values above. The BXD power app can be found at http://power.genenetwork.org/ and uses the *Shiny* R package (https://CRAN.R-project.org/package=shiny).