HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2022 May 20.

Published in final edited form as:

Methods Mol Biol. 2019 ; 2018: 233–247. doi:10.1007/978-1-4939-9581-3_11.

Using Heterogeneous Stocks for Fine-Mapping Genetically Complex Traits

Leah C. Solberg Woods, **Abraham A. Palmer**

Abstract

In this chapter we will review both the rationale and experimental design for using Heterogeneous Stock (HS) populations for fine-mapping of complex traits in mice and rats. We define an HS as an outbred population derived from an intercross between two or more inbred strains. HS have been used to perform genome-wide association studies (GWAS) for multiple behavioral, physiological, and gene expression traits. GWAS using HS require four key steps, which we review: selection of an appropriate HS population, phenotyping, genotyping, and statistical analysis. We provide advice on the selection of an HS, comment on key issues related to phenotyping, discuss genotyping methods relevant to these populations, and describe statistical genetic analyses that are applicable to genetic analyses that use HS.

Keywords

Genetic mapping; Outbred; Advanced intercross; Heterogeneous stock; Diversity outbred

1 Introduction

In the past 10 years, heterogeneous stock (HS) rodents have been used for fine-mapping and identification of underlying causal genes and genetic variants of multiple complex traits. HS are highly recombinant rodent populations that are created from two or more (most commonly eight) inbred founder strains. Examples of HS populations include advanced intercross lines (AIL), HS mice and rats, HS-CC, and the diversity outbred (DO) mouse. The original purpose of the early HS populations was to serve as a source of genetic diversity for selection studies, whereas more recently created HS (HS-CC and DO) were intended primarily for genome-wide association studies (GWAS). In each case, a collection of inbred strains were chosen to capture high levels of genetic diversity. Breeding schemes for these populations require that many families are maintained in each generation and that familial relationships among breeding pairs are minimized. With each generation of outbreeding, the distance between recombination events is reduced, allowing for genetic mapping to increasingly smaller intervals. The individuals within an HS represent a random mosaic of haplotypes of the inbred founder strains, with each animal being genetically and phenotypically distinct. Thus, HS more closely resemble the variation found in natural populations (including humans) when compared to “reference populations” such as recombinant inbred (RI; [1]), collaborative cross (CC; [2]), and other inbred panels. New mutations within HS populations are generally assumed to be negligible, although future work is needed to test this assumption.

There are several advantages to using HS populations for genetic mapping relative to traditional mapping strategies such as an F_2 intercross or backcross. The first is that they allow mapping of complex traits to small regions, addressing a critical weakness of traditional mapping strategies [3-5], and decreasing the number of candidate genes within each interval. A second advantage is that the founders of all extant HS have been fully sequenced and these data are available in public databases, making it possible to identify informative markers, potentially causal coding variants and more complex repetitive or structural differences. Ancestral haplotype information can also be used to inform which founder allele (s) contribute to the phenotypic trait (*see* [6, 7]). Third, the use of HS models for genetic mapping minimizes the chance that mapped alleles will be highly dependent on a specific genetic background/genetic context (*see* [8]), because loci are identified on a mix of genetic backgrounds. Finally, due to the genetic diversity within HS populations, there is often a high degree of phenotypic variability, ranging from physiological to behavioral traits, allowing genetic mapping of many different traits in the same cohort. Advantages and disadvantages of HS rats relative to other mapping strategies have been previously discussed [3-5].

For genetic mapping studies in HS populations, four key steps are necessary: selection of an appropriate HS population, phenotyping, genotyping, and statistical analysis. All four are discussed in more detail below.

2 Selection of an Appropriate HS Population

Below we review available rat and mouse HS populations that can be used for genetic studies. In deciding whether to use HS rats or mice, the main consideration should be the phenotype to be studied. Rats are better suited for certain behavioral phenotypes that do not work well in the mouse [9]. Rats are also larger and thus may be more amenable to phenotypes involving surgery. Mice are less expensive to maintain and benefit from the availability of a greater diversity of genetic tools and resources. Because genetic mapping is dependent on the heritability of the trait, it is important to establish heritability of a phenotype before beginning a large GWAS. This can sometimes be accomplished by measuring differences among the inbred founder strains; however for HS rats, the original founder strains are no longer available; therefore, heritability must be established using the most closely related inbred strains or by pedigree or marker-based relatedness (e.g. [10]).

2.1 HS Mice

The oldest HS mouse population, which is now extinct (personal communication, Dr. Jerry Stitzel 6-30-18), was the Boulder HS which was created in UC Berkeley but thereafter maintained at the Institute for Behavioral Genetics in Boulder, CO (aka HS/IBG; [11]). The HS/IBG was created by intercrossing eight inbred strains: A, AKR, BALB/c, C3H/2, C57BL, DBA/2, Is/Bi, RIII (note that the exact strain/vendor designations are unknown). A second HS population, called the Northport HS (aka HS/NPT; [12]), which is alive in the laboratory of Dr. Robert Hitzemann as of July 1, 2018, was created using six of the same inbred strains: A/J, AKR/NCrl, BALB/cAnNCrl, C3H/HeNCrl, C57BL6/NCrl, DBA/2J, but replaced Is/Bi and RIII with CBA/J, and LP/J. Both the HS/IBG and HS/NPT were

originally created to provide a genetically diverse population for selection studies, but were later used for genetic fine-mapping. The HS/NPT was one of the earlier stocks used for genetic fine-mapping, with the goal of narrowing previously identified loci discovered using an F_2 cross [12, 13]. Using HS for genome-wide analysis, in which discovery of genetic loci and fine-mapping for multiple phenotypic traits were done in the same cohort, occured several years later [14]. HS mice have since been used to map multiple traits including fear behavior [15], ethanol consumption [16], and arthritis [17].

An advanced intercross line (AIL) is an HS that is composed of only two inbred strains [18]. At least two AIL have been created using C57BL/6J and DBA/2J [19, 20], but the only currently extant AIL was created by crossing the LG/J and SM/J inbred strains and has been used for numerous mapping studies (e.g., [19, 21-34]).

More recently, several HS mouse populations that were specifically intended for genetic fine-mapping have been developed. The HS-CC was created by intercrossing eight inbred strains: A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HlLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ [35]. Those strains were selected because they were the same eight strains used to make the collaborative cross (CC; for further discussion of the CC *see* [36]). Another HS termed the HS4, which is now extinct, was similarly created by intercrossing only four inbred strains: C57BL/6J, DBA/2J, BALB/cJ, and LP/J [35]. Using the partially inbred founders of the CC, another HS was created at Jackson Labs and is termed the diversity outbred (DO; [37]). The DO are currently being maintained using 175 breeding pairs [7], a much larger number than used for other HS [38], which should significantly reduce inbreeding. The DO have been used for a wide variety of GWAS and related projects (e.g., [7, 39-44]). Both the HS-CC and DO are made up of the same eight inbred strains. That said, the HS-CC were started frm the original inbred strains whereas the DO were started from the partially inbred CC lines. For more information about the similarities and differences between the HS-CC and the DO, *see* [45].

2.2 HS Rats

The N/NIH HS is the only rat HS population that we are aware of; it was first established at the National Institute of Health (NIH) in 1984 [46] and was derived from the following eight inbred progenitor (founder) strains: ACI/N, BN/SsN, BUF/N, F344/N, M520/N, MR/N, WKY/N, and WN/N. The HS rat colony was maintained using a rotational breeding strategy and 60 breeder pairs by Dr. Carl Hansen at NIH until 2003. From 2003 to 2006, the colony was transferred to the laboratory of Dr. Eva Redei at Northwestern University and the number of breeder pairs decreased to 25. In 2006, the HS colony was transferred to two locations: Dr. Solberg Woods at the Medical College of Wisconsin and Dr. Alberto Fernando Teruel at the Autonomous University of Barcelona in Spain. At that time, the colony had been through 55 generations of breeding (50 at the NIH and 5 at Northwestern University). At MCW the number of breeding pairs was increased to 46 and the colony was maintained using a rotational breeding strategy [1, 23]. The HS animals at the Medical College of Wisconsin were named NMcwi:HS (Rat Genome Database identification number: 2314009). In 2013, after 15 generations of breeding at MCW (70 total), the colony was expanded to include 64 breeder pairs per generation and have been maintained in this

way since that time. At that time a new breeding strategy that uses kinship coefficients was employed (https://github.com/pcarbo/breedail). As of early 2018, the colony had been through 81 generations of breeding. This HS rat colony is currently a national resource funded through a NIDA Center of Excellence for genome-wide association studies (GWAS) in outbred rats (P50DA037844) which seeks to identify genetic loci underlying drug abuse behaviors (*see* www.ratgenes.org). Dr. Solberg Woods' laboratory currently ships rats to several investigators throughout the United States, on a cost-recovery basis (enquiries to: lsolberg@wakehealth.edu). As of this writing, the HS colony is being maintained at both the Medical College of Wisconsin (MCW) and Wake Forest School of Medicine (WFSM); however, the MCW colony will be phased out by the end of 2019.

HS rats have also been used for GWAS. They were first used to fine-map a previously identified locus for glucose tolerance [47, 48], which led to identification of *Tpcn2* as the likely underlying causal gene [49]. They were subsequently used to map multiple behavioral and physiological traits [50]. Since then they have been used for genetic mapping of adiposity [6] and studies show that the HS rat will be a promising model for mapping kidney-related traits [51], bone fragility [52], drug abuse behavior [53-55], depression-like behavior [10], as well as behavioral and physiological responses to stress [56-58] and ethanol [59-61].

2.3 Breeding Considerations

In many cases, HS mice or rats can be obtained from an existing colony for much less than the cost of maintaining an independent colony. When maintaining a colony, both the breeding methods (i.e., circular vs. random) and the number of families per generation are critical parameters. Rockman and Kruglyak [62] conducted a series of simulation studies that compare several outbred designs. They concluded that randomized breeding, using equal contributions from each breeder pair for the next generation (one male and one female from each breeder pair, no sib matings), is one of the most effective designs. Random breeding results in an expanded genetic map, decreased bin sizes between recombination breakpoints and controls genetic drift. An added advantage is that it is relatively easy to implement. Both inbreeding and genetic drift are also dependent on population size, with smaller populations more vulnerable to genetic drift [63]. In this regard, it is best to maintain the largest colony that is economically feasible. Rockman and Kruglyak [62] found that 64 breeder pairs increased map expansion, decreased bin size, and controlled genetic drift relative to a population of only 16 breeder pairs. In general, more breeder pairs are always better for achieving these goals. Dr. Palmer's lab has created software that implements a more sophisticated breeding scheme that uses the pedigree to minimize relatedness across all pairs (https://github.com/pcarbo/breedail).

3 Phenotyping

We will not discuss any particular phenotype as these are beyond the scope of this chapter. Phenotypic traits studied in HS mice or rats, however, should be amenable to high-throughput phenotyping, as large numbers of animals are needed to achieve sufficient statistical power (*see* Subheading 3.3 below). Because each HS animal is genetically and

phenotypically unique, it is not possible to have biological replicates when working with HS populations. For this reason, it is important to use phenotypes that are robust, stable, and replicable—heritable traits should fulfill these criteria. Due to the cost of genotyping, it is advantageous to phenotype multiple traits in each animal. Examples of studies with multiple high-throughput phenotypes include Valdar et al. [14], Baud et al. [50], and Gonzales et al. [25].

3.1 Tracking Animals and Covariates

Record keeping is critical to all scientific work, and is especially so for a large GWAS with HS. Each animal must be assigned a unique ID. At the time of weaning, all animals should receive some indelible identifier, typically an ear, neck tag or a transponder ID chip, secondary identifiers, like ear punches can also be helpful. Pertinent information that should be recorded includes sex, coat color (convenient for identifying the animal while alive and for checking genetically predicted coat color versus recorded coat color, which can identify sample mix-ups), date of birth, date of death, and the ID of the mother and the father. Phenotypic data that are collected should always refer to the individual by their unique ID; abbreviating this ID is sometimes tempting, but always unwise. Covariates collected at the time of phenotyping such as batch, experimenter, and time of day should also be carefully recorded; a list of possible confounding environmental covariates has been presented by [64]. As discussed in Subheading 5, covariates can then be regressed out in advance or included in the statistical model.

3.2 Data Checking

Once all phenotype data have been collected, it is important to remove erroneous data and to locate missing data; in practice, identifying outliers is a good first step. Once outliers are identified, it can be determined whether the outlier is legitimate or represents some type of error. Often batches of data (e.g., data collected on a given day) can be compared to data collected in all other batches, to determine whether batch-wide errors have occurred.

3.3 Statistical Power

The number of animals to be used depends greatly on the heritability of the trait, which can be estimated in advance. However, power also depends on the effect sizes of the QTLs, which are not knowable in advance. Power also depends on the degree of recombination among the HS founders, with greater recombination increasing mapping precision but also decreasing statistical power, thus increasing the sample size required (*see* [4]). Previous power calculations have been run in multi-founder populations and suggest that 1000–1500 HS animals provide sufficient power for mapping QTL explaining 5% of the variance [65-67]. These simulations do not account for the confounding effects of relatedness (e.g., [22, 68]), or marker ascertainment (e.g., [69]). That said, previous studies using >1000 rats or mice have had high degrees of success in mapping many traits [14, 50]. Indeed, we have had success mapping adiposity traits with only 742 HS rats, although the QTL identified just reached the significance level despite explaining >5% of the variance [6]. Expression QTL (eQTL) mapping generally requires far fewer animals because eQTL explain a greater percentage of the variance.

4 Genotyping

A critical consideration for any quantitative trait locus (QTL) mapping study is the genotyping platform. The number of genetic markers needed is a function of the recombination in the HS population; in general, more generations of breeding will require more markers. Prior studies have used anywhere between a few thousand to a few million markers. If marker density is not sufficient, QTLs may not be detected. For example, in a study conducted a decade ago, we used a single nucleotide polymorphism (SNP) genotyping microarray that contained ~4,500 informative markers to study a 34th generation AIL [22]; subsequent reanalysis with a denser set of markers showed that some apparently true loci were not discovered due to inadequate coverage (https://www.biorxiv.org/content/10.1101/387613v2). Similarly, in one of our recent HS rat studies, we used a 10,000 SNP array, containing 8218 informative markers in the HS rat population [6]. Although successful for genetic mapping, we found that this relatively low number led to uncertainty when we tried to impute founder haplotypes (*see* Subheading 5 below).

Many recent mouse studies have used the Mouse Universal Genotyping Array (MUGA) which contains about 8000 SNPs [7], the MegaMUGA which contains about 78,000 SNPs and the more recent GigaMUGA, which contains about 143,000 SNPs; these arrays are available from Neogen (http://genomics.neogen.com/en/mouse-universal-genotyping-array). Previous studies in the HS rat have used an 800K array [50]; however this array is no longer commercially available.

Microarrays are typically designed with a particular population in mind, which means they may not provide satisfactory coverage for other populations. For example, the above mentioned MUGA arrays, despite their name, were predominantly designed for the DO mouse and may be less informative in other populations (e.g., *see* [70, 71]). Designing a new array incurs a significant cost. As the price of next generation sequencing has decreased, strategies have been developed to obtain genotypes via sequencing rather than from microarrays. The two most prominent sequencing-based strategies are genotyping-by-sequencing and low-coverage whole-genome sequencing.

For low-coverage whole-genome sequencing, the entire genome of each rat or mouse is sequenced at very low coverage (~0.2X). An imputation algorithm is then used to call genotypes. Davies et al. [72] recently developed a genotype imputation algorithm, Sequencing to Imputation Through Constructing Haplotypes (STITCH), that is tailored to low-coverage sequence data for which no reference haplotypes are available. This method has recently been applied to commercially available outbred mice [70].In the case of HS, for which reference haplotypes are available, conventional imputation software such as Beagle [73] may perform equally well [72].

Genotype-by-sequencing (GBS), which was originally developed for use in plants, has been adapted for use in both mice and rats [25, 71, 74]. In this method, the DNA is cut using one or more restriction enzymes, ligated to bar-coded adapters, pooled, and sequenced. As the cost of sequencing continues to decrease, these methods may supplant the use of

microarrays; however, it is important to note that the bioinformatic analysis required for these sequencing-based methods is more burdensome than for microarrays.

Genotype data can be used to perform several important quality checks. Markers on the X chromosome should be homozygous (actually hemizygous) in males, the presence of too many heterozygous genotypes in a male, or too few in a female, suggests that the sex was recorded incorrectly or that samples were inadvertently switched. Such errors should be corrected or the affected samples should be excluded. When sequencing data are available it is also possible to compute the fraction of reads that map to the X chromosome; females are expected to have twice as many such reads. As with genotype information, deviations from this expectation indicate some sort of error. Genotype information can also be used to determine whether the genotypes are consistent with the recorded pedigree; this can be done in many ways, including by performing “Mendelization checks” or by comparing identical by descent statistics to kinship as calculated from the pedigree. Finally, the genotype at known coat color alleles can be compared to the recorded coat color; deviations from expectation may indicate genotyping error at that locus, errors in the recording of coat color, or sample mix-ups.

5 Statistical Analysis

Once phenotype and genotype data have been quality checked, the data are statistically analyzed. Association analyses can be conducted on SNP genotypes or on ancestral haplotypes, as described below.

GWAS is widely used in human genetics and is a general term for many similar methods that seek to estimate the association between a genetic marker (typically a SNP) and a phenotype of interest. Rodent populations have relatively small effective population sizes, meaning that there will be variable degrees of relatedness among subjects (siblings, cousins, etc.), which must be accounted for when performing a GWAS [22, 68, 75-78]. The most commonly used approach to account for relatedness is to employ a linear mixed model (LMM *see* [6]). Various software packages are available for this purpose (reviewed in [78]) that can include covariates such as experimenter, time of day, and sex. The only LMM software packages specifically targeted at model organisms are QTLRel [7, 75], DOQTL [79], and R/qtl2 [80], all of which have been used to analyze multi-founder populations like the HS. In addition, GEMMA, while originally designed for use in human GWAS, has been widely used for both human and rodent GWAS (e.g., [71, 81]). Non-normally distributed phenotypes may require transformations as with other statistical tests that assume normality. Web-based tools are also available to allow dynamic analyses with historical or user-supplied datasets (*see* www.genenetwork.org).

An advantage of HS populations is the ability to impute founder haplotypes. Imputation is commonly used in human genetics to obtain genotypes at markers that are not directly genotyped. However because an HS has a finite number of haplotypes and because parental lines have been sequenced, ancestral haplotype probabilities can be used for genetic mapping, potentially providing information beyond that obtained using only SNP genotypes [66]. There are several methods for haplotype imputation, including HAPPY, originally

developed by Mott and colleagues [66], DOQTL, developed by Gatti and Churchill [79] or QTL2Geno, developed by Broman [80, 82]. All methods employ a Hidden Markov model to determine the probability of each possible haplotype across the genome. To date, a comparison of the three methods has not been published. Once genetic loci are identified, founder haplotype effects can be determined using tools within DOQTL [79] or using the Diploeffect model (*see* [6, 83]) or simply by comparing strain distribution patterns for the lead SNP and adjacent SNPs.

5.1 Significance Thresholds

In human GWAS, 5×10^{-8} is an almost universally accepted threshold for significance [84]. In HS rodent populations, where the degree of linkage disequilibrium is highly variable [4], it is more common to use empirically derived thresholds. Genome-wide significance thresholds can be determined using Bonferroni thresholds (which are highly conservative), estimated by parametric bootstrap samples from the fitted null [47, 68], or estimated using permutation, if and only if an LMM was used to account for relatedness [76]. There are also more recently proposed methods (e.g., multiTrans), and our experience suggests that thresholds obtained using multiTrans are very similar to those obtained using permutation (unpublished observation).

5.2 Confidence Intervals

To determine confidence intervals, LD intervals for the detected QTL can be defined by including neighboring markers that meet a set level of LD, measured with the squared correlation coefficient r^2 as previously described [6]. An r^2 threshold of 0.4 is widely used in human genetics.

6 Post-GWAS: How to Find Causal Genes and Variants

Genetic loci identified using HS rats or mice typically span just a few Mb or less (there are several examples of loci that are less than 1 Mb), depending on many factors including the population used, the effect size of the locus, the density of the markers, and the local LD structure. In rare cases these loci contain only one gene, however it is common to identify loci that include many genes. It is therefore important to follow up genetic mapping with additional strategies to identify the causal genes and/or variants. Non-synonymous variants within the QTL that are both highly conserved and predicted to be damaging by SIFT (http://sift.jcvi.org/) or Polyphen (http://genetics.bwh.harvard.edu/pph/) should be considered, if and only if they match the strain distribution pattern of the QTL. As demonstrated in our lab [6] and by others [50], this can be followed up by using protein modeling strategies to demonstrate a functional effect of the amino acid change. Although this strategy has been successful, human GWAS suggests that most causal variants are likely to be regulatory instead of coding. Although there are still challenges to identify the causal regulatory variant, RNA expression, eQTL mapping (e.g., [49, 71]), co-localization [85], and mediation analysis [6] can be used to identify candidate genes that underlie these loci. When combined, we have demonstrated that these strategies allow identification of multiple candidate genes underlying a single locus [6]. Merge analysis [86] has also been

used to narrow the number of potentially causal variants within a QTL, thus further refining candidate genes within a locus.

7 Conclusions

HS populations allow genetic fine-mapping of complex disease traits to relatively small intervals of the genome. Due to genetic and phenotypic diversity, these model systems are useful for mapping multiple behavioral and physiological traits. We have outlined the general methods used for genetic mapping using outbred rodent resources, outlining considerations for phenotyping, genotyping, and statistical analysis.

Acknowledgments

L.S.W. and A.A.P. were both supported by P50DA037844. L.S.W. is also supported by R01 DK106386.

References

1. Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004) A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet 5:7. 10.1186/1471-2156-5-7 [PubMed: 15117419]
2. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J et al. (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36(11):1133–1137. 10.1038/ng1104-1133 [PubMed: 15514660]
3. Mott R, Flint J (2013) Dissecting quantitative traits in mice. Annu Rev Genomics Hum Genet 14:421–439. 10.1146/annurev-genom-091212-153419 [PubMed: 23834320]
4. Parker CC, Palmer AA (2011) Dark matter: are mice the solution to missing heritability? Front Genet 2:32. 10.3389/fgene.2011.00032 [PubMed: 22303328]
5. Solberg Woods LC (2014) QTL mapping in outbred populations: successes and challenges. Physiol Genomics 46(3):81–90. 10.1152/physiolgenomics.00127.2013 [PubMed: 24326347]
6. Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A et al. (2018) Genetic fine-mapping and identification of candidate genes and variants for adiposity traits in outbred rats. Obesity (Silver Spring) 26(1):213–222. 10.1002/oby.22075 [PubMed: 29193816]
7. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ et al. (2012) High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics 1902):437–447. 10.1534/genetics.111.132597 [PubMed: 22345611]
8. Sittig LJ, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA (2016) Genetic background limits generalizability of genotype-phenotype relationships. Neuron 91(6):1253–1259. 10.1016/j.neuron.2016.08.013 [PubMed: 27618673]
9. Parker CC, Chen H, Flagel SB, Geurts AM, Richards JB, Robinson TE et al. (2014) Rats are the smart choice: rationale for a renewed focus on rats in behavioral genetics. Neuropharmacology 76(Pt B):250–258. 10.1016/j.neuropharm.2013.05.047 [PubMed: 23791960]
10. Holl K, He H, Wedemeyer M, Clopton L, Wert S, Meckes JK et al. (2018) Heterogeneous stock rats: a model to study the genetics of despair-like behavior in adolescence. Genes Brain Behav 17(2):139–148. 10.1111/gbb.12410 [PubMed: 28834208]
11. McClearn GE, Wilson JR, Meredith W (1970) The use of isogenic and heterogenic mouse stocks in behavioral research. In: Lindzey G, Thiessen D (eds) Contributions to behavior-genetic analysis: the mouse as a prototype. Appleton Century Crofts, New York, pp 3–22
12. Demarest K, Koyner J, McCaughran J Jr, Cipp L, Hitzemann R (2001) Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. Behav Genet 31(1):79–91 [PubMed: 11529277]
13. Talbot CJ, Nicod A, Cherny SS, Fulker DW, Collins AC, Flint J (1999) High-resolution mapping of quantitative trait loci in outbred mice. Nat Genet 21(3):305–308 [PubMed: 10080185]

14. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet 38(8):879–887 [PubMed: 16832355]
15. Talbot CJ, Radcliffe RA, Fullerton J, Hitzemann R, Wehner JM, Flint J (2003) Fine scale mapping of a genetic locus for conditioned fear. Mamm Genome 14(4):223–230 [PubMed: 12682774]
16. Hitzemann R, Edmunds S, Wu W, Malmanger B, Walter N, Belknap J et al. (2009) Detection of reciprocal quantitative trait loci for acute ethanol withdrawal and ethanol consumption in heterogeneous stock mice. Psychopharmacology 203(4):713–722. 10.1007/s00213-008-1418-y [PubMed: 19052728]
17. Ahlqvist E, Ekman D, Lindvall T, Popovic M, Forster M, Hultqvist M et al. (2011) High-resolution mapping of a complex disease, a model for rheumatoid arthritis, using heterogeneous stock mice. Hum Mol Genet 20(15):3031–3041. 10.1093/hmg/ddr206 [PubMed: 21565963]
18. Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141(3):1199–1207 [PubMed: 8582624]
19. Parker CC, Cheng R, Sokoloff G, Palmer AA (2012) Genome-wide association for methamphetamine sensitivity in an advanced intercross mouse line. Genes Brain Behav 11(1):52–61. 10.1111/j.1601-183X.2011.00747.x [PubMed: 22032291]
20. Peirce JL, Broman KW, Lu L, Chesler EJ, Zhou G, Airey DC et al. (2008) Genome Reshuffling for Advanced Intercross Permutation (GRAIP): simulation and permutation for advanced intercross population analysis. PLoS One 3(4):e1977. 10.1371/journal.pone.0001977 [PubMed: 18431467]
21. Carroll AM, Cheng R, Collie-Duguid ES, Meharg C, Scholz ME, Fiering S et al. (2017) Fine-mapping of genes determining extrafusal fiber properties in murine soleus muscle. Physiol Genomics 49(3):141–150. 10.1152/physiolgenomics.00092.2016 [PubMed: 28087756]
22. Cheng R, Lim JE, Samocha KE, Sokoloff G, Abney M, Skol AD et al. (2010) Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. Genetics 185(3):1033–1044. 10.1534/genetics.110.116863 [PubMed: 20439773]
23. Cheverud JM, Lawson HA, Bouckaert K, Kossenkov AV, Showe LC, Cort L et al. (2014) Fine-mapping quantitative trait loci affecting murine external ear tissue regeneration in the LG/J by SM/J advanced intercross line. Heredity (Edinb) 112(5):508–518. 10.1038/hdy.2013.133 [PubMed: 24569637]
24. Ehrich TH, Hrbek T, Kenney-Hunt JP, Pletscher LS, Wang B, Semenkovich CF et al. (2005) Fine-mapping gene-by-diet interactions on chromosome 13 in a LG/J x SM/J murine model of obesity. Diabetes 54(6):1863–1872 [PubMed: 15919810]
25. Gonzales NM, Seo J, Hernandez-Cordero AI, St. Pierre CL, Gregory JS, Distler MG et al. (2018) Genome wide association study of behavioral, physiological and gene expression traits in a multigenerational mouse intercross. BioRxiv. 10.1101/230920
26. Hernandez Cordero AI, Carbonetto P, Riboni Verri G, Gregory JS, Vandenbergh DJ, Gyekis JP et al. (2018) Replication and discovery of musculoskeletal QTLs in LG/J and SM/J advanced intercross lines. Phys Rep 6(4). 10.14814/phy2.13561
27. Lawson HA, Lee A, Fawcett GL, Wang B, Pletscher LS, Maxwell TJ et al. (2011) The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J x SM/J murine model. Mamm Genome 22(3–4):197–208. 10.1007/s00335-010-9313-3 [PubMed: 21210123]
28. Lawson HA, Zelle KM, Fawcett GL, Wang B, Pletscher LS, Maxwell ,TJ et al. (2010) Genetic, epigenetic, and gene-by-diet interaction effects underlie variation in serum lipids in a LG/JxSM/J murine model. J Lipid Res 51(10):2976–2984. 10.1194/jlr.M006957 [PubMed: 20601649]
29. Norgard EA, Lawson HA, Pletscher LS, Wang B, Brooks VR, Wolf JB et al. (2011) Genetic factors and diet affect long-bone length in the F34 LG,SM advanced intercross. Mamm Genome 22(3–4):178–196. 10.1007/s00335-010-9311-5 [PubMed: 21170743]
30. Parker CC, Carbonetto P, Sokoloff G, Park YJ, Abney M, Palmer AA (2014) High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. Genetics 198(1):103–116. 10.1534/genetics.114.167056 [PubMed: 25236452]

31. Parker CC, Cheng R, Sokoloff G, Lim JE, Skol AD, Abney M et al. (2011) Fine-mapping alleles for body weight in LG/J x SM/J F(2) and F (34) advanced intercross lines. Mamm Genome 22(9–10):563–571. 10.1007/s00335-011-9349-z [PubMed: 21761260]
32. Parker CC, Sokoloff G, Cheng R, Palmer AA (2012) Genome-wide association for fear conditioning in an advanced intercross mouse line. Behav Genet 42(3):437–448. 10.1007/s10519-011-9524-8 [PubMed: 22237917]
33. Rai MF, Schmidt EJ, Hashimoto S, Cheverud JM, Sandell LJ (2015) Genetic loci that regulate ectopic calcification in response to knee trauma in LG/J by SM/J advanced intercross mice. J Orthop Res 33(10):1412–1423. 10.1002/jor.22944 [PubMed: 25989359]
34. Samocha KE, Lim JE, Cheng R, Sokoloff G, Palmer AA (2010) Fine mapping of QTL for prepulse inhibition in LG/J and SM/J mice using F(2) and advanced intercross lines. Genes Brain Behav 9(7):759–767. 10.1111/j.1601-183X.2010.00613.x [PubMed: 20597988]
35. Iancu OD, Darakjian P, Walter NA, Malmanger B, Oberbeck D, Belknap J et al. (2010) Genetic diversity and striatal gene networks: focus on the heterogeneous stock-collaborative cross (HS-CC) mouse. BMC Genomics 11:585. 10.1186/1471-2164-11-585 [PubMed: 20959017]
36. Consortium CC (2012) The genome architecture of the Collaborative Cross mouse genetic reference population. Genetics 190(2):389–401. 10.1534/genetics.111.132639 [PubMed: 22345608]
37. Matsumoto Y, Goto T, Nishino J, Nakaoka H, Tanave A, Takano-Shimizu T et al. (2017) Selective breeding and selection mapping using a novel wild-derived heterogeneous stock of mice revealed two closely-linked loci for tameness. Sci Rep 7(1):4607. 10.1038/s41598-017-04869-1 [PubMed: 28676693]
38. Yalcin B, Flint J (2012) Association studies in outbred mice in a new era of full-genome sequencing. Mamm Genome 23(9–10):719–726. 10.1007/s00335-012-9409-z [PubMed: 22847376]
39. Gatti D, French JE, Schughart K(2017) QTL Mapping and identification of candidate genes in DO mice: a use case model derived from a benzene toxicity experiment. Methods Mol Biol 1488:265–281. 10.1007/978-1-4939-6427-7_12 [PubMed: 27933529]
40. Logan RW, Robledo RF, Recla JM, Philip VM, Bubier JA, Jay JJ et al. (2013) High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. Genes Brain Behav 12(4):424–437. 10.1111/gbb.12029 [PubMed: 23433259]
41. Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ (2014) Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene. Mamm Genome 25(5–6):211–222. 10.1007/s00335-014-9508-0 [PubMed: 24700285]
42. Shorter JR, Huang W, Beak JY, Hua K, Gatti DM, de Villena FP et al. (2018) Quantitative trait mapping in Diversity Outbred mice identifies two genomic regions associated with heart size. Mamm Genome 29(1–2):80–89. 10.1007/s00335-017-9730-7 [PubMed: 29279960]
43. Smallwood TL, Gatti DM, Quizon P, Weinstock GM, Jung KC, Zhao L et al. (2014) High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. G3 4(12):2353–2363. 10.1534/g3.114.014704 [PubMed: 25344410]
44. Winter JM, Gildea DE, Andreas JP, Gatti DM, Williams KA, Lee M et al. (2017) Mapping complex traits in a diversity outbred F1 mouse population identifies germline modifiers of metastasis in human prostate cancer. Cell Syst 4(1):31–45 e36. 10.1016/j.cels.2016.10.018 [PubMed: 27916600]
45. Chesler EJ, Gatti DM, Morgan AP, Strobel M, Trepanier L, Oberbeck D et al. (2016) Diversity Outbred mice at 21: maintaining allelic variation in the face of selection. G3 6(12):3893–3902. 10.1534/g3.116.035527 [PubMed: 27694113]
46. Hansen C, Spuhler K (1984) Development of the National Institutes of Health genetically heterogeneous rat stock. Alcohol Clin Exp Res 8(5):477–479 [PubMed: 6391259]
47. Solberg Woods LC, Holl K, Tschannen M, Valdar W (2010) Fine-mapping a locus for glucose tolerance using heterogeneous stock rats. Physiol Genomics 41(1):102–108. 10.1152/physiolgenomics.00178.2009 [PubMed: 20068026]

48. Solberg Woods LC, Holl KL, Oreper D, Xie Y, Tsaih SW, Valdar W (2012) Fine-mapping diabetes-related traits, including insulin resistance, in heterogeneous stock rats. Physiol Genomics 44(21):1013–1026. 10.1152/physiolgenomics.00040.2012 [PubMed: 22947656]
49. Tsaih SW, Holl K, Jia S, Kaldunski M, Tschannen M, He H et al. (2014) Identification of a novel gene for diabetic traits in rats, mice, and humans. Genetics 198(1):17–29. 10.1534/genetics.114.162982 [PubMed: 25236446]
50. Baud A, Hermsen R, Guryev V, Stridh P, Graham D, McBride MW et al. (2013) Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. Nat Genet 45(7):767–775. 10.1038/ng.2644 [PubMed: 23708188]
51. Solberg Woods LC, Stelloh C, Regner KR, Schwabe T, Eisenhauer J, Garrett MR (2010) Heterogeneous stock rats: a new model to study the genetics of renal phenotypes. Am J Physiol Ren Physiol 298(6):F1484–F1491. 10.1152/ajprenal.00002.2010
52. Alam I, Koller DL, Sun Q, Roeder RK, Canete T, Blazquez G et al. (2011) Heterogeneous stock rat: a unique animal model for mapping genes influencing bone fragility. Bone 48(5):1169–1177. 10.1016/j.bone.2011.02.009 [PubMed: 21334473]
53. King CP, Palmer AA, Woods LC, Hawk LW, Richards JB, Meyer PJ (2016) Premature responding is associated with approach to a food cue in male and female heterogeneous stock rats. Psychopharmacology 233(13):2593–2605. 10.1007/s00213-016-4306-x [PubMed: 27146401]
54. Richards JB, Lloyd DR, Kuehlewind B, Militello L, Paredez M, Solberg Woods L et al. (2013) Strong genetic influences on measures of behavioral-regulation among inbred rat strains. Genes Brain Behav 12(5):490–502. 10.1111/gbb.12050 [PubMed: 23710681]
55. Wang T, Han W, Wang B, Jiang Q, Solberg-Woods Lc, Palmer AA et al. (2014) Propensity for social interaction predicts nicotine-reinforced behaviors in outbred rats. Genes Brain Behav 13(2):202–212. 10.1111/gbb.12112 [PubMed: 24289793]
56. Diaz-Moran S, Palencia M, Mont-Cardona C, Canete T, Blazquez G, Martinez-Membrives E et al. (2012) Coping style and stress hormone responses in genetically heterogeneous rats: comparison with the Roman rat strains. Behav Brain Res 228(1):203–210. 10.1016/j.bbr.2011.12.002 [PubMed: 22178313]
57. Lopez-Aumatell R, Guitart-Masip M, Vicens-Costa E, Gimenez-Llort L, Valdar W, Johannesson M et al. (2008) Fearfulness in a large N/Nih genetically heterogeneous rat stock: differential profiles of timidity and defensive flight in males and females. Behav Brain Res 188(1):41–55. 10.1016/j.bbr.2007.10.015 [PubMed: 18079010]
58. Lopez-Aumatell R, Vicens-Costa E, Guitart-Masip M, Martinez-Membrives E, Valdar W, Johannesson M et al. (2009) Unlearned anxiety predicts learned fear: a comparison among heterogeneous rats and the Roman rat strains. Behav Brain Res 202(1):92–101. 10.1016/j.bbr.2009.03.024 [PubMed: 19447285]
59. Bice PJ, Liang T, Zhang L, Graves TJ, Carr LG, Lai D et al. (2010) Fine mapping and expression of candidate genes within the chromosome 10 QTL region of the high and low alcohol-drinking rats. Alcohol 44(6):477–485. 10.1016/j.alcohol.2010.06.004 [PubMed: 20705418]
60. Foroud T, Bice P, Castelluccio P, Bo R, Miller L, Ritchotte A et al. (2000) Identification of quantitative trait loci influencing alcohol consumption in the high alcohol drinking and low alcohol drinking rat lines. Behav Genet 30(2):131–140 [PubMed: 10979603]
61. Spuhler K, Deitrich RA (1984) Correlative analysis of ethanol-related phenotypes in rat inbred strains. Alcohol Clin Exp Res 8(5):480–484 [PubMed: 6391260]
62. Rockman MV, Kruglyak L (2008) Breeding designs for recombinant inbred advanced intercross lines. Genetics 179(2):1069–1078. 10.1534/genetics.107.083873 [PubMed: 18505881]
63. Falconer DS (1960) Introduction to quantitative genetics. Ronald Press, New York
64. Wurbel H (2002) Behavioral phenotyping enhanced—beyond (environmental) standardization. Genes Brain Behav 1(1):3–8 [PubMed: 12886944]
65. Mott R, Flint J (2002) Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics 160(4):1609–1618 [PubMed: 11973314]
66. Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proc Natl Acad Sci U S A 97(23):12649–12654 [PubMed: 11050180]

67. Valdar WS, Flint J, Mott R (2003) QTL fine-mapping with recombinant-inbred heterogeneous stocks and in vitro heterogeneous stocks. Mamm Genome 14(12):830–838. 10.1007/s00335-003-3021-1 [PubMed: 14724734]
68. Valdar W, Holmes CC, Mott R, Flint J (2009) Mapping in structured populations by resample model averaging. Genetics 182(4):1263–1277. 10.1534/genetics.109.100727 [PubMed: 19474203]
69. Valdar W, Flint J, Mott R (2006) Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics 172(3):1783–1797. 10.1534/genetics.104.039313 [PubMed: 16361245]
70. Nicod J, Davies RW, Cai N, Hassett C, Goodstadt L, Cosgrove C et al. (2016) Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. Nat Genet 48(8):912–918. 10.1038/ng.3595 [PubMed: 27376238]
71. Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ et al. (2016) Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. Nat Genet 48(8):919–926. 10.1038/ng.3609 [PubMed: 27376237]
72. Davies RW, Flint J, Myers S, Mott R (2016) Rapid genotype imputation from sequence without reference panels. Nat Genet 48(8):965–969. 10.1038/ng.3594 [PubMed: 27376236]
73. Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98(1):116–126. 10.1016/j.ajhg.2015.11.020 [PubMed: 26748515]
74. Fitzpatrick CJ, Gopalakrishnan S, Cogan ES, Yager LM, Meyer PJ, Lovic V et al. (2013) Variation in the form of Pavlovian conditioned approach behavior among outbred male Sprague-Dawley rats from different vendors and colonies: sign-tracking vs. goal-tracking. PLoS One 8(10):e75042. 10.1371/journal.pone.0075042 [PubMed: 24098363]
75. Cheng R, Abney M, Palmer AA, Skol AD (2011) QTLRel: an R package for genome-wide association studies in which relatedness is a concern. BMC Genet 12:66. 10.1186/1471-2156-12-66 [PubMed: 21794153]
76. Cheng R, Palmer AA (2013) A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness. Genetics 193(3):1015–1018. 10.1534/genetics.112.146332 [PubMed: 23267053]
77. Cheng R, Parker CC, Abney M, Palmer AA (2013) Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. G3 3(10):1861–1867. 10.1534/g3.113.007948 [PubMed: 23979941]
78. Gonzales NM, Palmer AA (2014) Fine-mapping QTLs in advanced intercross lines and other outbred populations. Mamm Genome 25(7–8):271–292. 10.1007/s00335-014-9523-1 [PubMed: 24906874]
79. Gatti DM, Svenson KL, Shabalin A, Wu LY, Valdar W, Simecek P et al. (2014) Quantitative trait locus mapping methods for diversity outbred mice. G3 4(9):1623–1633. 10.1534/g3.114.013748 [PubMed: 25237114]
80. Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA (2018) R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multi-parent populations. Genetics 211:495–502. 10.1534/genetics.118.301595. [PubMed: 30591514]
81. Pallares LF, Carbonetto P, Gopalakrishnan S, Parker CC, Ackert-Bicknell CL, Palmer AA et al. (2015) Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. PLoS Genet 11(11):e1005607. 10.1371/journal.pgen.1005607 [PubMed: 26523602]
82. Broman KW (2016) qtl2geno: treatment of marker genotypes for QTL experiments. R package version 0.4-21. http://kbroman.org/qtl2
83. Zhang Z, Wang W, Valdar W (2014) Bayesian modeling of haplotype effects in multiparent populations. Genetics 198(1):139–156. 10.1534/genetics.114.166249 [PubMed: 25236455]
84. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273(5281):1516–1517 [PubMed: 8801636]
85. Hormozdiari F, van de Bunt M, Segre AV, Li X, Joo JWJ, Bilow M et al. (2016) Colocalization of GWAS and eQTL signals detects target genes. Am J Hum Genet 99(6):1245–1260. 10.1016/j.ajhg.2016.10.003 [PubMed: 27866706]

86. Yalcin B, Flint J, Mott R (2005) Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics 171(2):673–681 [PubMed: 16085706]