Heterogeneous Stock Populations for Analysis of Complex Traits

Leah Solberg Woods and Richard Mott

Corresponding Author:

Leah Solberg Woods

Medical College of Wisconsin

8701 Watertown Plank Road

Milwaukee, WI 53130

lsolberg@mcw.edu

Running Head: heterogeneous stocks for complex trait analysis

Summary/Abstract

HS populations allow for fine-resolution genetic mapping of a variety of complex traits. HS mice and rats were created from breeding together eight inbred strains, followed by maintaining the colony in a manner that minimizes inbreeding. After 50 or more generations of breeding, the resulting animals' chromosomes represent a genetic mosaics of the founders' haplotypes, with the average distance between recombination events in the centiMorgan range. This allows for genetic mapping to only a few Mb, a much smaller region than what can be identified using traditional F2 intercross or backcross mapping strategies. HS animals have been used to fine-map a variety of complex traits including anxiety and fear behaviors, diabetes, asthma and heart disease, among others. Once a quantitative trait locus (QTL) has been identified, founder sequence and expression analysis can be used to identify underlying causal genes. In the following review, we provide an overview of how HS rats and mice have been used to identify genetic loci, and in some cases the causal genes, underlying complex traits. We discuss the creation and breeding strategies for both HS rats and mice. We then discuss the statistical analyses used to identify genetic loci, as well as strategies to identify causal genes underlying these loci. We end the Chapter by discussing limitations faced when using HS populations, including several statistical challenges that have not been fully resolved.

Key Words

Genetic mapping, outbred mice and rats, expression analysis

Introduction:

HS populations of mice and rats are created from breeding together eight inbred strains, followed by maintaining the colony in a manner that minimizes inbreeding (see figure 1). After 50 or more generations of breeding, the resulting animals' chromosomes are a genetic mosaic of the founders' haplotypes, with the average distance between recombination events in the centiMorgan range allowing genetic mapping to only a few Mb [1, 2]. HS animals exhibit a high degree of both genetic and phenotypic diversity, allowing high-resolution genetic mapping for a wide variety of traits. While both HS rats and mice were originally created to be a resource population for experimental and selection studies [3, 4], Flint and colleagues demonstrated in 1999 that these populations can be used to narrow a previously identified quantitative trait locus (QTL) for anxiety [5] to only 0.8 cM [6], a large improvement over mapping studies using traditional F2 intercross or backcross approaches which generally map to 30 cM or more. Since that time, multiple studies have used HS rats and mice for genetic mapping of complex traits. Whilst this chapter focuses on mouse and rat HS, similar inbred types of populations - in which each individual's genome is a mosaic of the founders - have been made in these and other species [7-10]. Other outbred populations, such as advanced intercross lines [11, 12] the mouse Diversity Outcross (also created from eight inbred founder strains; [13]) and commercially available outbreds [14, 15], are also available and have been reviewed previously [16].

Upon demonstrating success of the HS strategy for fine-mapping a single locus for a behavioral trait, the Flint lab went on to use HS mice to conduct a large multi-phenotype study, including traits involved in fear and anxiety behaviors, diabetes and asthma, and others. The study included the largest cohort of mice at that time (1904 mice) and resulted in

the identification of 843 QTL at 25% false discovery rate (FDR) with an average confidence interval of only 2.8 Mb [17]. In separate studies, HS mice have also been used to fine-map traits such as fear [18, 19], ethanol-induced locomotor activity [20] and arthritis [21]. Two HS mouse colonies have been created: the Boulder HS [4] and the Northport HS [20]. Six of the inbred founder strains are shared between these stocks, namely A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, DBA/2J. The additional founders of the Boulder HS are the strains Is/Bi and R111, while those of the Northport HS are CBA/2J and LP/J. Colonies were created and maintained with 24-40 breeder pairs using either a circular or random mating scheme.

With the success of HS mice, investigators began to also use HS rats for genetic fine-mapping. The HS rat colony (N:NIH-HS) was first established by the NIH in 1984 using the following inbred strains: ACI/N,BN/SsN,BUF/N, F344/N, M520/N, MR/N, WKY/N, WN/N [3]. The colony was maintained using 60 breeder pairs using a random mating strategy. Upon the retirement of the colony's originator, Dr. Carl Hansen, in 2006 the colony was transferred to the laboratory of Dr. Eva Redei at Northwestern University where it was maintained for two years. In 2006, Dr. Redei transferred breeder pairs to the Medical College of Wisconsin in the United States and the Autonomous University of Barcelona in Spain. The Medical College of Wisconsin currently maintains 60 breeder pairs and is using a random breeding strategy. A smaller colony is also currently being maintained at Barcelona (Fernandez Teruel, personal communication). The first genetic mapping study in HS rats fine-mapped a single locus for glucose tolerance from 60 Mb to only 2.4 Mb [22]. Using expression QTL analysis and founder sequence data, *Tpcn2,* was identified as the likely causal gene within this region within only a few years [23]. Since that time, Flint and colleagues again conducted a large multi-phenotype study, including traits involved in

anxiety, heart disease, multiple sclerosis among others [24]. They were able to fine-map 355 QTL for 122 traits at 10% FDR. An example of a scan for platelet aggregation is shown in Figure 2. Using a merge analysis (described below) and protein modeling, they were able to identify 35 probable causative genes within these loci.   These data are described in detail and publicly available at [25]. We note that different analysis methodologies were used in the mouse and rat HS experiments, reflecting methodological improvements (principally the development of mixed models, described below) in the interim. A recent study has also fine-mapped bone structure and strength in HS rats [26] and additional studies have demonstrated phenotypic variability, and thus suitability for future genetic studies, for additional phenotypic traits including kidney traits [27], drug abuse behaviors [28, 29] and behavioral and physiological responses to stress [30-32] and to ethanol [33-35]. Because of the rich history of the rat in behavioral studies, the HS rat will likely prove a useful model for genetic dissection of behaviors that are not easily modeled in the mouse (see [36]. The utility of the HS rat will also be enhanced by the recent availability of gene knock-outs and other genetic manipulations now available in the rat [37, 38].

Statistical analysis and systems genetics in HS populations

In order to perform genetic mapping in HS populations, the underlying ancestral haplotype mosaics must first be determined. In this way, one can compute the probability that a particular locus in a given individual descended from which of the eight founder strains, thus providing increased information over simply analyzing genotype data (often based on bi-allelic single nucleotide polymorphisms or SNPs). HAPPY, a program developed by Mott and colleagues [2], uses a hidden Markov Model to determine the ancestral probabilities and has been shown to significantly improve genetic mapping results. Other programs, such as

DOQTL which was more recently developed for use in the Diversity Outbred (DO) mouse population, a population of mice created from eight founder strains, very similar to the HS, [39] can also be used to determine ancestral probabilities in the HS. Once probabilities are determined, regression modeling is conducted on the underlying mosaic structure to identify QTL.

Several programs are available for identifying QTL including Bagpipe [40], QTLRel [41] and DOQTL [39]. As with other highly recombinant populations, it is important to account for the complex family relationships within the HS population when conducting the analysis [24, 40, 42]. Most colonies are maintained using either a random or circular breeding strategy with anywhere from 40-60 breeding pairs, generally sufficient to minimize inbreeding and control genetic drift. Random mating strategies have the advantage of also avoiding reduced map expansion and therefor may be preferable to circular mating strategies [43]. As a result of the closed nature of the breeding strategy, animals within a colony are all related to each other but to differing degrees. If not accounted for, false positive QTL will be identified simply on the basis of relatedness, as opposed to actually pertaining to the phenotype. Moreover, in HS studies, in order to generate a large sample of animals for phenotyping, it is necessary to expand the colony size resulting in a large number of families. Thus the analysis of the phenotyped individuals acquires a mixture of linkage (due to family structure) and association (because all individuals are ultimately descended from eight founders). If the parents of the phenotyped generation are genotyped along with the phenotyped mice, then it is also possible to infer the maternal and paternal origin of the alleles, allowing the study of parent-of-origin effects [44].

There are several ways to account for unequal degrees of relatedness in outbred populations. These include mixed modeling approaches such as EMMA [45] or by resample model averaging [40]. When genome-wide genotype information is unavailable, and/or when the full pedigree is unknown, family can also be accounted for by including this as a random term in the model, as previously demonstrated for a single locus on rat chromosome 1 [22, 46]. When full pedigree information is known, QTLRel can be used to account for family relationships in highly recombinant animal populations such as the HS [41]. Resample model averaging approaches make use of genome-wide genotypes to determine genetic relatedness directly, and may prove advantageous under certain circumstances, particularly when pedigree information is unknown [40]. For mixed models, a kinship matrix is used to determine the genetic relatedness of each pair of individuals. This can be simply computed from SNP data (in the same way that is it computed in human studies) or from the ancestral haplotype mosaics [44]. Baud et al [24] used a mixed model to control for differences in relatedness, whilst Valdar et al 2006 [17] used resample model averaging (developed further in [40]). Each method has advantages and disadvantages. The mixed-model methodology is essentially equivalent to transforming both phenotypes and genotypes by multiplication by the square root of the inverse of the variance-covarinace matrix, to create an uncorrelated dataset that can be analysed by ordinary least squares. The method is well-established and works well on phenotypes that are approximately normally distributed. Resampling methods, on the other hand, work well on phenotypes that are strongly skewed in distribution, but are slower than mixed models.

Once QTL are identified, there are several methods that can be used to identify the underlying causal gene(s) within the locus. These include a statistical merge analysis [47] (a form of genotype imputation), expression QTL mapping and protein modeling. To date,

complete genomes have been sequenced in the founder strains of the HS mice [48] and rats [24]. Relative to the respective reference genomes, more than 4 million SNPs per strain have been identified in the mouse [48] and more than 2 million SNPs per strain have been identified in the rat [24], in addition to structural variants and insertion/deletions. Because a repetitive portion of the genome (~15% in mouse and ~12% in rat) could not be reliably mapped to the reference strains, the number of variants identified is likely much larger than reported [24, 48]. The available sequence information can be used in several ways to identify candidate genes and/or variants within a fine-mapped QTL. By coupling founder sequence with relatively dense genotyping of the outbred population, it is possible to impute HS genotypes at all possible SNPs within a QTL. This can be followed by a merge analysis to narrow the potential causative variants within the QTL [47]. Briefly, a merge analysis uses probabilistically inferred descent to impute genotype dosages at unobserved loci, and then surveys those multiple imputed SNPs for their association with the phenotype. Using this method, two statistical models are compared: the haplotype model and the allelic model. In the haplotype model, the underlying ancestral probabilities at each SNP are used to model the QTL; with each founder haplotype permitted to carry a different phenotypic effect. This haplotype model is compared with one in which only the alleles for that SNP are used (allelic model). In the allelic model, the founder strain alleles are "merged" into two groups for each diallelic SNP: those containing allele "a" at a locus of interest and those containing allele "b" at this locus [47]; with each group having a single phenotypic effect. Potentially causative variants are those in which the allelic model provides a better fit, that is, explaining the same amount but with far fewer parameters, than the haplotype model (see [21, 47]. This method has proven useful in narrowing the number of causative variants within QTL mapped in HS mice [48] and rats [24]. A single causal variant, however, is rarely identified and follow-up studies are often needed to identify the causal variant. In addition, the method works less

well when multiple causal variants underlie a single QTL [24], but can be used to show that a QTL cannot be explained by a single biallelic variant (as was the case for about half the QTLs detected by Baud et al). Merge analysis is also useful for excluding genetic variants that cannot be causal.

Transcript abundance levels, based on RNAseq or microarray data, can also be used to identify causal genes underlying a QTL, as well as identification of gene networks that play a role in a given trait. Expression QTL (eQTL) analysis in HS populations allows investigators to map both cis and trans-eQTL to within only a few Mb of the transcript itself [49]. Overlap of cis-eQTL with physiological QTL can then be used to identify candidate genes within an interval, as demonstrated in HS rats by Tsaih et al., [23]. Our group identified *Tpcn2* as a cis-eQTL within a physiological QTL for fasting glucose and insulin levels and demonstrated that glucose levels strongly correlated with *Tpcn2* expression levels in the HS rats. We then demonstrated that fasting glucose and insulin in response to a glucose challenge were altered in *Tpcn2* knock-out mice, and *Tpcn2* was nominally correlated with fasting insulin levels in humans, providing evidence that *Tpcn2* is the likely causal gene within this region. Transcript abundance levels can also be used to identify gene networks (groups of correlated transcripts) that may play a role in disease [50-53]. Although this strategy has not been applied specifically to HS populations to-date, it offers a promising avenue of research for the future. Similarly, regions of the genome associated with open chromatin, such as DNAse-1 hypersensitive sites, identified in the mouse reference genome by the mouse ENCODE project [54] and across the HS founder strains [55] may be used to help identify causal variants and genes.

Further Considerations and Limitations

Using outbred HS populations offers several advantages over traditional F2 intercross or backcross strategies. The first is the ability to fine-map to only a few Mb, greatly decreasing the number of potential candidate genes within a given locus. Once loci are identified, full genome sequence is available for founder strains of the HS mouse [48] and HS rat [24] and this information has proven to be invaluable for identifying causative genes and variants within fine-mapped QTL.   Despite these advantages, there are also several disadvantages that should be considered prior to embarking on a study with HS populations. Each rat is genetically and phenotypically distinct, so once a QTL, or even a gene, has been identified, there is no inbred model to go back to for functional testing, although the inbred founders might be used for this purpose. In addition, large numbers of animals are needed to have sufficient power for each mapping study and high density genotyping platforms (generally 10K or greater) are required. Because each animal is unique, all animals need to be genotyped and phenotyped with each new study, as opposed to recombinant inbred lines where genotypes need be collected only once. As a result, it is beneficial to gather as much phenotype information as possible from the same group of animals, so that genotyping only needs to be done once and this information can be used to map multiple traits (e.g., [17, 24]). A further disadvantage is that these populations have been created through a single funnel (i.e., combining founder genomes only once), leading to loss of certain alleles and potentially unbalanced representation of the founder genomes. There are further considerations regarding confirmation of a potentially causal gene, and several statistical challenges that have not been resolved when using highly recombinant populations such as the HS.

Once a candidate gene is identified, follow-up studies are needed to confirm or disprove the role of that candidate in the trait. In addition to replication in a separate cohort, conducting a cross-species comparison can help provide support for the gene of interest. Of particular interest is human genome-wide association data which is often publicly available and can be mined to determine if a gene of interest falls just below the genome-wide significance threshold in humans. Once there is sufficient evidence to suspect a causal role for a specific gene, one of the most popular methods used to verify this gene is to study it in a knock-out model. Such methods have been available in the mouse since 1990 [56] and have recently become available in the rat [57]. Although popular because of their relative ease of constructing a knock-out, it is important to recognize that showing a change in phenotype in a knock-out model neither proves nor disproves a causative role of this gene at the QTL [58], particularly because there is no way to create a knock-out on the same background in which the QTL was identified. Methods such as quantitative complementation offer an alternative approach to test a causal role of the gene or variant [59, 60]. More importantly, new gene targeting approaches which allow for changes in single base pairs are now being used and offer a more realistic approach than a full gene knock-out (see [61]. The revolution in gene editing due to CRISPR/Cas9 technology (recently applied to rats in [38]) suggests that gene confirmation will become more straightforward in the future, at least for isolated coding variants. However, cases where multiple causal variants, carried on a single haplotype, are implicated, will likely remain a challenge to prove causality, particularly if their effect is regulatory. It is therefore important for investigators to assess all available information, including expression, sequence, cross-species comparisons, results from a knock-out, allelic changes using CRISPR/Cas9 modifications, as well as possible *in vitro* studies to assess the potential causative role of a particular gene and/or variant.

In addition to the above considerations, there are several statistical concerns that need to be taken into consideration when analyzing outbred HS populations. One of these is how best to determine significance thresholds. Cheng and Palmer [62] recently compared four different methods used in an advanced intercross population. They found that as long as an appropriate statistical model (i.e., one that takes into account the complex family relationships) is used, all methods worked relatively well, with gene-dropping (a simulation technique used in pedigree analysis) decreasing false QTL even when family is not taken into account. The best way to determine QTL confidence intervals is another challenge. Many studies use the 1.5 LOD drop method, (e.g., [46, 63, 64], which was initially developed for use in F2 intercross populations. An alternative approach is to use non-parametric bootstrapping [65], in which the QTL is re-estimated under alternative datasets based on the original, with each alternative data-set created by resampling the individuals with replacement [66]. Although this method has been shown to be overly conservative [67], it does provide a complementary estimate of how sensitive the localization of the top QTL peak is to resampling, thus providing insight into whether more than one locus may underlie the QTL (see [46].

Accurate determination of the joint effects of diplotypes (i.e., combinations of founder haplotypes – effectively any departure from the assumption of additivity in the haplotype effects) is also an on-going statistical challenge in outbred populations. Using the DO mouse population, investigators have looked at the effects of just the founder allele effects within the QTL [13, 63]. This has been useful in conducting haplotype analysis and narrowing the region of the QTL. However, within the HS or DO populations, there are in effect 36 possible diplotype combinations, and founder effects account for only eight of these. We

[46] have recently published methods that accounts for all 36 possible diplotype effects and work in this area is on-going (see [68]).

A final statistical concern is that of statistical power. Although previous power calculations have been run in multi-founder populations and suggest that 1,000-1,500 animals provide sufficient power for mapping QTL explaining 5% of the variance [2, 69, 70], these simulations do not account for the confounding effects of relatedness (eg, [40, 42]), or marker ascertainment (eg, [71]), and are therefore likely overstated. Previous studies in the diversity outbred mouse population have used as few as 150 mice, however this study provided sufficient power to map only 11 of 113 traits that were measured [13]. Studies in HS populations have used over 1000 animals, successfully mapping most traits analyzed [17, 24], and demonstrating the increased power of these studies. In order to have more accurate power estimates for future studies in these populations, power calculations will need to account for both family structure and polygenes. That said, increasing sample size has many benefits: there is greater power to detect QTLs of small effect, QTL effect sizes are less likely to be over-estimated due to Winner's Curse, and confidence intervals are likely to be smaller and more accurate.

Conclusion

HS populations have proven useful for fine-mapping complex traits to only a few Mb, rapidly narrowing the number of potential candidate genes within the locus. Several strategies, including use of full founder sequence and expression QTL mapping have been used to identify the underlying causal genes within these loci. It is important, however, to consider the cost and labor intensive challenges of working with HS populations, as well as the many

unanswered statistical challenges that still remain. Despite these challenges, use of outbred

models such as the HS has and will continue to enhance of knowledge of the genetic

architecture of complex traits.

References

1. Mott, R. and J. Flint, *Dissecting quantitative traits in mice.* Annu Rev Genomics Hum Genet, 2013. **14**: p. 421-39.
2. Mott, R., C.J. Talbot, M.G. Turri, A.C. Collins, and J. Flint, *A method for fine mapping quantitative trait loci in outbred animal stocks.* Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12649-54.
3. Hansen, C. and K. Spuhler, *Development of the National Institutes of Health genetically heterogeneous rat stock.* Alcohol Clin Exp Res, 1984. **8**(5): p. 477-9.
4. McClearn, G.E., J.R. Wilson, and W. Meredith, *The use of isogenic and heterogenic mouse stocks in behavioral research*, in *Contributions to behavior-genetic analysis: the mouse as a prototype.*, G. Lindzey and D. Thiessen, Editors. 1970, Appleton Century Crofts: New York. p. 3-22.
5. Flint, J., R. Corley, J.C. DeFries, D.W. Fulker, J.A. Gray, S. Miller, and A.C. Collins, *A simple genetic basis for a complex psychological trait in laboratory mice.* Science, 1995. **269**(5229): p. 1432-5.
6. Talbot, C.J., A. Nicod, S.S. Cherny, D.W. Fulker, A.C. Collins, and J. Flint, *High-resolution mapping of quantitative trait loci in outbred mice.* Nat Genet, 1999. **21**(3): p. 305-8.
7. Aylor, D.L., W. Valdar, W. Foulds-Mathes, R.J. Buus, R.A. Verdugo, R.S. Baric, M.T. Ferris, J.A. Frelinger, M. Heise, M.B. Frieman, L.E. Gralinski, T.A. Bell, J.D. Didion, K. Hua, D.L. Nehrenberg, C.L. Powell, J. Steigerwalt, Y. Xie, S.N. Kelada, F.S. Collins, I.V. Yang, D.A. Schwartz, L.A. Branstetter, E.J. Chesler, D.R. Miller, J. Spence, E.Y. Liu, L. McMillan, A. Sarkar, J. Wang, W. Wang, Q. Zhang, K.W. Broman, R. Korstanje, C. Durrant, R. Mott, F.A. Iraqi, D. Pomp, D. Threadgill, F.P. de Villena, and G.A. Churchill, *Genetic analysis of complex traits in the emerging Collaborative Cross.* Genome Res, 2011. **21**(8): p. 1213-22.
8. Durrant, C., H. Tayem, B. Yalcin, J. Cleak, L. Goodstadt, F.P. de Villena, R. Mott, and F.A. Iraqi, *Collaborative Cross mice and their power to map host susceptibility to Aspergillus fumigatus infection.* Genome Res, 2011. **21**(8): p. 1239-48.
9. Kover, P.X., W. Valdar, J. Trakalo, N. Scarcelli, I.M. Ehrenreich, M.D. Purugganan, C. Durrant, and R. Mott, *A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana.* PLoS Genet, 2009. **5**(7): p. e1000551.
10. Long, A.D., S.J. Macdonald, and E.G. King, *Dissecting complex traits using the Drosophila Synthetic Population Resource.* Trends Genet, 2014. **30**(11): p. 488-95.
11. Darvasi, A. and M. Soller, *Advanced intercross lines, an experimental population for fine genetic mapping.* Genetics, 1995. **141**(3): p. 1199-207.
12. Parker, C.C., R. Cheng, G. Sokoloff, and A.A. Palmer, *Genome-wide association for methamphetamine sensitivity in an advanced intercross mouse line.* Genes Brain Behav, 2012. **11**(1): p. 52-61.
13. Svenson, K.L., D.M. Gatti, W. Valdar, C.E. Welsh, R. Cheng, E.J. Chesler, A.A. Palmer, L. McMillan, and G.A. Churchill, *High-resolution genetic mapping using the Mouse Diversity outbred population.* Genetics, 2012. **190**(2): p. 437-47.
14. Yalcin, B. and J. Flint, *Association studies in outbred mice in a new era of full-genome sequencing.* Mamm Genome, 2012. **23**(9-10): p. 719-26.
15. Zhang, W., R. Korstanje, J. Thaisz, F. Staedtler, N. Harttman, L. Xu, M. Feng, L. Yanas, H. Yang, W. Valdar, G.A. Churchill, and K. Dipetrillo, *Genome-wide association mapping of quantitative traits in outbred mice.* G3 (Bethesda), 2012. **2**(2): p. 167-74.
16. Solberg Woods, L.C., *QTL mapping in outbred populations: successes and challenges.* Physiol Genomics, 2014. **46**(3): p. 81-90.
17. Valdar, W., L.C. Solberg, D. Gauguier, S. Burnett, P. Klenerman, W.O. Cookson, M.S. Taylor, J.N. Rawlins, R. Mott, and J. Flint, *Genome-wide genetic association of complex traits in heterogeneous stock mice.* Nat Genet, 2006. **38**(8): p. 879-87.

18.   Johannesson, M., R. Lopez-Aumatell, P. Stridh, M. Diez, J. Tuncel, G. Blazquez, E. Martinez-Membrives, T. Canete, E. Vicens-Costa, D. Graham, R.R. Copley, P. Hernandez-Pliego, A.D. Beyeen, J. Ockinger, C. Fernandez-Santamaria, P.S. Gulko, M. Brenner, A. Tobena, M. Guitart-Masip, L. Gimenez-Llort, A. Dominiczak, R. Holmdahl, D. Gauguier, T. Olsson, R. Mott, W. Valdar, E.E. Redei, A. Fernandez-Teruel, and J. Flint, *A resource for the simultaneous high-resolution mapping of multiple quantitative trait loci in rats: the NIH heterogeneous stock.* Genome Res, 2009. **19**(1): p. 150-8.
19.   Talbot, C.J., R.A. Radcliffe, J. Fullerton, R. Hitzemann, J.M. Wehner, and J. Flint, *Fine scale mapping of a genetic locus for conditioned fear.* Mamm Genome, 2003. **14**(4): p. 223-30.
20.   Demarest, K., J. Koyner, J. McCaughran, Jr., L. Cipp, and R. Hitzemann, *Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity.* Behav Genet, 2001. **31**(1): p. 79-91.
21.   Johnsen, A.K., W. Valdar, L. Golden, A. Ortiz-Lopez, R. Hitzemann, J. Flint, D. Mathis, and C. Benoist, *Genome-wide and species-wide dissection of the genetics of arthritis severity in heterogeneous stock mice.* Arthritis Rheum, 2011. **63**(9): p. 2630-40.
22.   Solberg Woods, L.C., K. Holl, M. Tschannen, and W. Valdar, *Fine-mapping a locus for glucose tolerance using heterogeneous stock rats.* Physiol Genomics, 2010. **41**(1): p. 102-8.
23.   Tsaih, S.W., K. Holl, S. Jia, M. Kaldunski, M. Tschannen, H. He, J.W. Andrae, S.H. Li, A. Stoddard, A. Wiederhold, J. Parrington, M. Ruas da Silva, A. Galione, J. Meigs, R.G. Hoffmann, P. Simpson, H. Jacob, M. Hessner, and L.C. Solberg Woods, *Identification of a novel gene for diabetic traits in rats, mice, and humans.* Genetics, 2014. **198**(1): p. 17-29.
24.   Baud, A., R. Hermsen, V. Guryev, P. Stridh, D. Graham, M.W. McBride, T. Foroud, S. Calderari, M. Diez, J. Ockinger, A.D. Beyeen, A. Gillett, N. Abdelmagid, A.O. Guerreiro-Cacais, M. Jagodic, J. Tuncel, U. Norin, E. Beattie, N. Huynh, W.H. Miller, D.L. Koller, I. Alam, S. Falak, M. Osborne-Pellegrin, E. Martinez-Membrives, T. Canete, G. Blazquez, E. Vicens-Costa, C. Mont-Cardona, S. Diaz-Moran, A. Tobena, O. Hummel, D. Zelenika, K. Saar, G. Patone, A. Bauerfeind, M.T. Bihoreau, M. Heinig, Y.A. Lee, C. Rintisch, H. Schulz, D.A. Wheeler, K.C. Worley, D.M. Muzny, R.A. Gibbs, M. Lathrop, N. Lansu, P. Toonen, F.P. Ruzius, E. de Bruijn, H. Hauser, D.J. Adams, T. Keane, S.S. Atanur, T.J. Aitman, P. Flicek, T. Malinauskas, E.Y. Jones, D. Ekman, R. Lopez-Aumatell, A.F. Dominiczak, M. Johannesson, R. Holmdahl, T. Olsson, D. Gauguier, N. Hubner, A. Fernandez-Teruel, E. Cuppen, R. Mott, and J. Flint, *Combined sequence-based and genetic mapping analysis of complex traits in outbred rats.* Nat Genet, 2013. **45**(7): p. 767-775.
25.   Baud, A., V. Guryev, O. Hummel, M. Johannesson, and J. Flint, *Genomes and phenomes of a population of outbred rats and its progenitors.* Sci Data, 2014. **1**: p. 140011.
26.   Alam, I., D.L. Koller, T. Canete, G. Blazquez, C. Mont-Cardona, R. Lopez-Aumatell, E. Martinez-Membrives, S. Diaz-Moran, A. Tobena, A. Fernandez-Teruel, P. Stridh, M. Diez, T. Olsson, M. Johannesson, A. Baud, M.J. Econs, and T. Foroud, *Fine mapping of bone structure and strength QTLs in heterogeneous stock rat.* Bone, 2015. **81**: p. 417-26.
27.   Solberg Woods, L.C., C. Stelloh, K.R. Regner, T. Schwabe, J. Eisenhauer, and M.R. Garrett, *Heterogeneous stock rats: a new model to study the genetics of renal phenotypes.* Am J Physiol Renal Physiol, 2010. **298**(6): p. F1484-91.
28.   Richards, J.B., D.R. Lloyd, B. Kuehlewind, L. Militello, M. Paredez, L. Solberg Woods, and A.A. Palmer, *Strong genetic influences on measures of behavioral-regulation among inbred rat strains.* Genes Brain Behav, 2013. **12**(5): p. 490-502.
29.   Wang, T., W. Han, B. Wang, Q. Jiang, L.C. Solberg-Woods, A.A. Palmer, and H. Chen, *Propensity for social interaction predicts nicotine-reinforced behaviors in outbred rats.* Genes Brain Behav, 2014. **13**(2): p. 202-12.
30.   Diaz-Moran, S., M. Palencia, C. Mont-Cardona, T. Canete, G. Blazquez, E. Martinez-Membrives, R. Lopez-Aumatell, A. Tobena, and A. Fernandez-Teruel, *Coping style and stress*

hormone responses in genetically heterogeneous rats: comparison with the Roman rat strains. Behav Brain Res, 2012. **228**(1): p. 203-10.

31.   Lopez-Aumatell, R., M. Guitart-Masip, E. Vicens-Costa, L. Gimenez-Llort, W. Valdar, M. Johannesson, J. Flint, A. Tobena, and A. Fernandez-Teruel, *Fearfulness in a large N/Nih genetically heterogeneous rat stock: differential profiles of timidity and defensive flight in males and females.* Behav Brain Res, 2008. **188**(1): p. 41-55.

32.   Lopez-Aumatell, R., E. Vicens-Costa, M. Guitart-Masip, E. Martinez-Membrives, W. Valdar, M. Johannesson, T. Canete, G. Blazquez, P. Driscoll, J. Flint, A. Tobena, and A. Fernandez-Teruel, *Unlearned anxiety predicts learned fear: a comparison among heterogeneous rats and the Roman rat strains.* Behav Brain Res, 2009. **202**(1): p. 92-101.

33.   Bice, P.J., T. Liang, L. Zhang, T.J. Graves, L.G. Carr, D. Lai, M.W. Kimpel, and T. Foroud, *Fine mapping and expression of candidate genes within the chromosome 10 QTL region of the high and low alcohol-drinking rats.* Alcohol, 2010. **44**(6): p. 477-85.

34.   Foroud, T., P. Bice, P. Castelluccio, R. Bo, L. Miller, A. Ritchotte, L. Lumeng, T.K. Li, and L.G. Carr, *Identification of quantitative trait loci influencing alcohol consumption in the high alcohol drinking and low alcohol drinking rat lines.* Behav Genet, 2000. **30**(2): p. 131-40.

35.   Spuhler, K. and R.A. Deitrich, *Correlative analysis of ethanol-related phenotypes in rat inbred strains.* Alcohol Clin Exp Res, 1984. **8**(5): p. 480-4.

36.   Parker, C.C., H. Chen, S.B. Flagel, A.M. Geurts, J.B. Richards, T.E. Robinson, L.C. Solberg Woods, and A.A. Palmer, *Rats are the smart choice: Rationale for a renewed focus on rats in behavioral genetics.* Neuropharmacology, 2013.

37.   Katter, K., A.M. Geurts, O. Hoffmann, L. Mates, V. Landa, L. Hiripi, C. Moreno, J. Lazar, S. Bashir, V. Zidek, E. Popova, B. Jerchow, K. Becker, A. Devaraj, I. Walter, M. Grzybowksi, M. Corbett, A.R. Filho, M.R. Hodges, M. Bader, Z. Ivics, H.J. Jacob, M. Pravenec, Z. Bosze, T. Rulicke, and Z. Izsvak, *Transposon-mediated transgenesis, transgenic rescue, and tissue-specific gene expression in rodents and rabbits.* FASEB J, 2013. **27**(3): p. 930-41.

38.   Zhao, L., E. Oliver, K. Maratou, S.S. Atanur, O.D. Dubois, E. Cotroneo, C.N. Chen, L. Wang, C. Arce, P.L. Chabosseau, J. Ponsa-Cobas, M.G. Frid, B. Moyon, Z. Webster, A. Aldashev, J. Ferrer, G.A. Rutter, K.R. Stenmark, T.J. Aitman, and M.R. Wilkins, *The zinc transporter ZIP12 regulates the pulmonary vascular response to chronic hypoxia.* Nature, 2015. **524**(7565): p. 356-60.

39.   Gatti, D.M., K.L. Svenson, A. Shabalin, L.Y. Wu, W. Valdar, P. Simecek, N. Goodwin, R. Cheng, D. Pomp, A. Palmer, E.J. Chesler, K.W. Broman, and G.A. Churchill, *Quantitative trait locus mapping methods for diversity outbred mice.* G3 (Bethesda), 2014. **4**(9): p. 1623-33.

40.   Valdar, W., C.C. Holmes, R. Mott, and J. Flint, *Mapping in structured populations by resample model averaging.* Genetics, 2009. **182**(4): p. 1263-77.

41.   Cheng, R., M. Abney, A.A. Palmer, and A.D. Skol, *QTLRel: an R package for genome-wide association studies in which relatedness is a concern.* BMC Genet, 2011. **12**: p. 66.

42.   Cheng, R., J.E. Lim, K.E. Samocha, G. Sokoloff, M. Abney, A.D. Skol, and A.A. Palmer, *Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations.* Genetics, 2010. **185**(3): p. 1033-44.

43.   Rockman, M.V. and L. Kruglyak, *Breeding designs for recombinant inbred advanced intercross lines.* Genetics, 2008. **179**(2): p. 1069-78.

44.   Mott, R., W. Yuan, P. Kaisaki, X. Gan, J. Cleak, A. Edwards, A. Baud, and J. Flint, *The architecture of parent-of-origin effects in mice.* Cell, 2014. **156**(1-2): p. 332-42.

45.   Kang, H.M., N.A. Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin, *Efficient control of population structure in model organism association mapping.* Genetics, 2008. **178**(3): p. 1709-23.

46.   Solberg Woods, L.C., K.L. Holl, D. Oreper, Y. Xie, S.W. Tsaih, and W. Valdar, *Fine-mapping diabetes-related traits, including insulin resistance, in heterogeneous stock rats.* Physiol Genomics, 2012. **44**(21): p. 1013-26.

47.   Yalcin, B., J. Flint, and R. Mott, *Using progenitor strain information to identify quantitative trait nucleotides in outbred mice.* Genetics, 2005. **171**(2): p. 673 - 81.

48.   Keane, T.M., L. Goodstadt, P. Danecek, M.A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N.A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T.G. Belgard, P.L. Oliver, R.E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C.A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I.J. Jackson, A. Czechanski, J.A. Guerra-Assuncao, L.R. Donahue, L.G. Reinholdt, B.A. Payseur, C.P. Ponting, E. Birney, J. Flint, and D.J. Adams, *Mouse genomic variation and its effect on phenotypes and gene regulation.* Nature, 2011. **477**(7364): p. 289-94.

49.   Huang, G.J., S. Shifman, W. Valdar, M. Johannesson, B. Yalcin, M.S. Taylor, J.M. Taylor, R. Mott, and J. Flint, *High resolution mapping of expression QTLs in heterogeneous stock mice in multiple tissues.* Genome Res, 2009. **19**(6): p. 1133-40.

50.   Chen, Y., J. Zhu, P.Y. Lum, X. Yang, S. Pinto, D.J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S.K. Sieberts, A. Leonardson, L.W. Castellini, S. Wang, M.F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T.A. Drake, A.J. Lusis, and E.E. Schadt, *Variations in DNA elucidate molecular networks that cause disease.* Nature, 2008. **452**(7186): p. 429-35.

51.   Ghazalpour, A., S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E.E. Schadt, T.A. Drake, A.J. Lusis, and S. Horvath, *Integrating genetic and network analysis to characterize genes related to mouse weight.* PLoS Genet, 2006. **2**(8): p. e130.

52.   Keller, M.P., Y. Choi, P. Wang, D. Belt Davis, M.E. Rabaglia, A.T. Oler, D.S. Stapleton, C. Argmann, K.L. Schueler, S. Edwards, H.A. Steinberg, E. Chaibub Neto, R. Kleinhanz, S. Turner, M.K. Hellerstein, E.E. Schadt, B.S. Yandell, C. Kendziorski, and A.D. Attie, *A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility.* Genome Res, 2008. **18**(5): p. 706-16.

53.   Schadt, E.E., J. Lamb, X. Yang, J. Zhu, S. Edwards, D. Guhathakurta, S.K. Sieberts, S. Monks, M. Reitman, C. Zhang, P.Y. Lum, A. Leonardson, R. Thieringer, J.M. Metzger, L. Yang, J. Castle, H. Zhu, S.F. Kash, T.A. Drake, A. Sachs, and A.J. Lusis, *An integrative genomics approach to infer causal associations between gene expression and disease.* Nat Genet, 2005. **37**(7): p. 710-7.

54.   Stamatoyannopoulos, J.A., M. Snyder, R. Hardison, B. Ren, T. Gingeras, D.M. Gilbert, M. Groudine, M. Bender, R. Kaul, T. Canfield, E. Giste, A. Johnson, M. Zhang, G. Balasundaram, R. Byron, V. Roach, P.J. Sabo, R. Sandstrom, A.S. Stehling, R.E. Thurman, S.M. Weissman, P. Cayting, M. Hariharan, J. Lian, Y. Cheng, S.G. Landt, Z. Ma, B.J. Wold, J. Dekker, G.E. Crawford, C.A. Keller, W. Wu, C. Morrissey, S.A. Kumar, T. Mishra, D. Jain, M. Byrska-Bishop, D. Blankenberg, B.R. Lajoie, G. Jain, A. Sanyal, K.B. Chen, O. Denas, J. Taylor, G.A. Blobel, M.J. Weiss, M. Pimkin, W. Deng, G.K. Marinov, B.A. Williams, K.I. Fisher-Aylor, G. Desalvo, A. Kiralusha, D. Trout, H. Amrhein, A. Mortazavi, L. Edsall, D. McCleary, S. Kuan, Y. Shen, F. Yue, Z. Ye, C.A. Davis, C. Zaleski, S. Jha, C. Xue, A. Dobin, W. Lin, M. Fastuca, H. Wang, R. Guigo, S. Djebali, J. Lagarde, T. Ryba, T. Sasaki, V.S. Malladi, M.S. Cline, V.M. Kirkup, K. Learned, K.R. Rosenbloom, W.J. Kent, E.A. Feingold, P.J. Good, M. Pazin, R.F. Lowdon, and L.B. Adams, *An encyclopedia of mouse DNA elements (Mouse ENCODE).* Genome Biol, 2012. **13**(8): p. 418.

55.   Hosseini, M., L. Goodstadt, J.R. Hughes, M.S. Kowalczyk, M. de Gobbi, G.W. Otto, R.R. Copley, R. Mott, D.R. Higgs, and J. Flint, *Causes and consequences of chromatin variation between inbred mice.* PLoS Genet, 2013. **9**(6): p. e1003570.

56.   Thomas, K.R. and M.R. Capecchi, *Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development.* Nature, 1990. **346**(6287): p. 847-50.

57.	Geurts, A.M., G.J. Cost, Y. Freyvert, B. Zeitler, J.C. Miller, V.M. Choi, S.S. Jenkins, A. Wood, X. Cui, X. Meng, A. Vincent, S. Lam, M. Michalkiewicz, R. Schilling, J. Foeckler, S. Kalloway, H. Weiler, S. Menoret, I. Anegon, G.D. Davis, L. Zhang, E.J. Rebar, P.D. Gregory, F.D. Urnov, H.J. Jacob, and R. Buelow, *Knockout rats via embryo microinjection of zinc-finger nucleases.* Science, 2009. **325**(5939): p. 433.

58.	Flint, J. and E. Eskin, *Genome-wide association studies in mice.* Nat Rev Genet, 2012. **13**(11): p. 807-17.

59.	Long, A.D., S.L. Mullaney, T.F. Mackay, and C.H. Langley, *Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in Drosophila melanogaster.* Genetics, 1996. **144**(4): p. 1497-510.

60.	Yalcin, B., S.A. Willis-Owen, J. Fullerton, A. Meesaq, R.M. Deacon, J.N. Rawlins, R.R. Copley, A.P. Morris, J. Flint, and R. Mott, *Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice.* Nat Genet, 2004. **36**(11): p. 1197-202.

61.	Adams, D.J. and L. van der Weyden, *Contemporary approaches for modifying the mouse genome.* Physiol Genomics, 2008. **34**(3): p. 225-38.

62.	Cheng, R. and A.A. Palmer, *A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness.* Genetics, 2013. **193**(3): p. 1015-8.

63.	Logan, R.W., R.F. Robledo, J.M. Recla, V.M. Philip, J.A. Bubier, J.J. Jay, C. Harwood, T. Wilcox, D.M. Gatti, C.J. Bult, G.A. Churchill, and E.J. Chesler, *High-precision genetic mapping of behavioral traits in the diversity outbred mouse population.* Genes Brain Behav, 2013. **12**(4): p. 424-37.

64.	Samocha, K.E., J.E. Lim, R. Cheng, G. Sokoloff, and A.A. Palmer, *Fine mapping of QTL for prepulse inhibition in LG/J and SM/J mice using F(2) and advanced intercross lines.* Genes Brain Behav, 2010. **9**(7): p. 759-67.

65.	Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition*2009, New York: Springer.

66.	Visscher, P.M., R. Thompson, and C.S. Haley, *Confidence intervals in QTL mapping by bootstrapping.* Genetics, 1996. **143**(2): p. 1013-20.

67.	Manichaikul, A., J. Dupuis, S. Sen, and K.W. Broman, *Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus.* Genetics, 2006. **174**(1): p. 481-9.

68.	Durrant, C. and R. Mott, *Bayesian quantitative trait locus mapping using inferred haplotypes.* Genetics, 2010. **184**(3): p. 839-52.

69.	Mott, R. and J. Flint, *Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks.* Genetics, 2002. **160**(4): p. 1609-18.

70.	Valdar, W.S., J. Flint, and R. Mott, *QTL fine-mapping with recombinant-inbred heterogeneous stocks and in vitro heterogeneous stocks.* Mamm Genome, 2003. **14**(12): p. 830-8.

71.	Valdar, W., J. Flint, and R. Mott, *Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice.* Genetics, 2006. **172**(3): p. 1783-97.

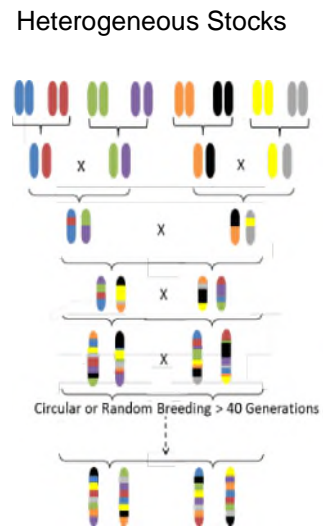**Figures**

Heterogeneous Stocks

Figure 1. Breeding scheme for heterogeneous stock (HS) populations. HS are created by breeding together eight inbred strains. Once all eight genomes are combined, the animals are bred using either a circular strategy or random breeding. Existing HS mouse and rat colonies have been bred for over 50 generations. Figure adapted from Solberg Woods, 2014.
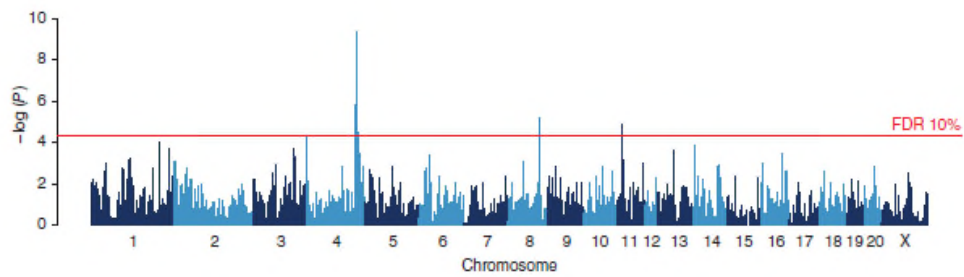
Figure 2. Genome scan for platelet aggregation as shown in Baud et al., 2013. The scan shows the results of a haplotype-based mixed model. The y-axis shows the negative log P values for association with variation in platelet aggregation. The association peak on chromosome 4 harbors the von Willebrand factor gene that was identified through sequence analysis as the causative gene.