Chapter 1.2

Resources for Systems Genetics

Robert W. Williams^, Evan G. Williams*

Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis TN, 38163 ^ and Institute for Molecular Systems Biology, Department of Biology, ETH Zürich, Zürich, Switzerland * Correspondence: rwilliams@uthsc.edu

Summary

A key characteristic of systems genetics is its reliance on populations that vary to a greater or lesser degree in genetic complexity—from highly admixed populations such as the Collaborative Cross and Diversity Outcross to relatively simple crosses such as sets of consomic strains and reduced complexity crosses. This protocol is intended to help investigators make more informed decisions about choices of resources given different types of questions. We consider factors such as costs, availability, and ease of breeding for common scenarios. In general, we recommend using complementary resources and minimizing depth of resampling of any given genome or strain.

Key words; Genetic reference population (GRP), Recombinant inbred (RI), Collaborative Cross (CC), congenic lines, consomic and chromosome substitution lines, recombinant congenic strains, RI intercross (RIX) and RI backcross (RIB) progeny, heterogeneous stock (HS), diversity outcross (DO), hybrid diversity panel (HDP), reduced complexity cross (RCC)

1. Introduction

A large number of innovative resources for systems genetics have been developed over the last 15 years (Williams and Auwerx 2015). There are at least two reasons for this burst of activity. The first catalyst was the introduction of far easier, cheaper, and more comprehensive methods of genotyping (Dietrich et al. 1995, Petkov et al. 2004) that we already take for granted. State-of-the-art genotyping for recombinant inbred (RI) strains consisted of ~1600 microsatellite markers (dinucleotide repeats) in 2001 (Williams et al. 2001). Over the next five years this number increased to more than 10,000 SNPs (Shifman et al. 2006), and we now rely on genotypes at more than 100,000 SNPs using Affymetrix or Illumina platforms (Yang et al. 2009, Morgan et al. 2015) at modest cost—well under $0.01 per marker. The second reason was rapid progress on ways to map quantitative traits with progressively higher precision and power (Lander and Botstein 1989, Darvasi and Soller 1995, Darvasi 1998, Talbot et al., 1999, Williams et al. 2001, Complex Trait Consortium 2003, Churchill et al. 2004, Singer et al. 2004, Flint et al. 2005, Broman 2005), culminating in the establishment of the Complex Trait Consortium in 2002 (Threadgill et al. 2003). A good problem we now face is selecting wisely from among the many options and resources that are available. Any choice is a major commitment. This protocol highlights factors researchers should consider and balance.

Guidance on using this protocol. The goal of this protocol is to step through the decisions associated with selecting resources for both QTL mapping and systems genetics. The first issue is to define classes of questions. Different questions benefit from different types and mixtures of resources—the cliché "different horses for different courses" applies. In **Part 3** we review current murine resource used in QTL mapping and systems genetics. In **Part 4** we consider one multipurpose experimental design that will work reasonably well for a range of questions. Consider this design a starting point for your discussions and decisions. We provide some notes on the pros and cons of the resources, many in a simple question-and-answer format. Since everyone has their own biases, ask others for their opinions.

These are among the main considerations or themes that go into the choice of resources for systems genetics:

1. Cost and availability (strains, hybrids, cases)
2. Phenotype diversity, heritability, and genetic architecture
3. Marker density, mapping precision, and power
4. Sequence diversity and genetic blind spots
5. Selective phenotyping or genotyping
6. Complexity of QTL intervals
7. Population structure, admixture, and analytic methods
8. Depth of genetic, omics, and phenome data resources
9. Robustness, replicability, extensibility (G×E), and translatability

To foreshadow our conclusions: Most researchers currently rely on a single type of resource or cross, and while there are good historical reasons for this focus, this is no longer an optimal or advisable strategy. We now have such a range of powerful genetic resources optimized for different purposes that it makes sense to take advantage of combinations of complementary crosses and even multiple species (Malmanger et al. 2006, Ghazalpour et al. 2012, Houtkooper et al. 2013, Williams et al. 2014, Wang et al. 2016). Analytic methods do get more complex when using combinations of resources, but some of the same methods used to handle admixed human cohorts in genome-wide and phenome-wide association studies (GWAS and PheWAS) have now been adapted to handle combined experimental cohorts (Lippert et al. 2011, Zhou and Stephens 2012, Fulotte et al. 2014).

Our other conclusion is that mapping resolution of about 1 Mb will usually be adequate to transition to validation, including translational analysis of human GWAS and PheWAS data sets (Koutnikova et al. 2009, Wang et al. 2016), analysis of knockout (KO) and knockin (KI) phenotypes, bioinformatic and omics dissection, and pharmacological intervention. This is especially true in an era of super high precision but mechanistically unanchored GWAS. The need for high precision mapping in mouse has been supplanted by an acute need for powerful resources to understand and accurately predict genome-to-phenome (G2P) relations under a wide range of environments and treatments.

2. Types of Questions

We consider four main types of questions.

Type 1 Questions: The classic forward genetic question—what are the polymorphic genes and sequence variants that modulate a phenotype or disease risk? This is by far the most common question our research community has dealt with over the last two decades and will probably remain so for the next several decades. Almost all human GWASs have this same simple reductionist motivation—a simple generalization of the classic Mendelian approach but applied to messier and continuously variable quantitative traits.

The repeated mapping of large numbers of QTLs and their causal QT genes (QTGs) quickly leads to complex systems-level questions—a transition we now are beginning to see in human GWAS. This shift has happened gradually over the past decade, including pioneering work by Wakeland and colleagues on the family of gene variants that contribute to autoimmune disease is a fine example (Subramanian et al. 2006). The work of Hunter and colleagues on metastasis networks (Hunter and Crawford 2008, Hu et al. 2012) and of Morahan and colleagues on type I diabetes (Morahan 2012) provide two other examples of this movement from QTL analysis to complex systems genetics. This shift is leading to the discovery of new biomarkers, diagnostics, mechanisms, and treatments.

Type 1 questions are usually approached in two steps: the first involves mapping QTLs to confidence intervals of 0.5–5 Mb, while the second and more problematic step involves proving to your own satisfaction and that reviewers that a polymorphic candidate gene has been validated as a source of trait variance (Complex Trait Consortium 2003). Almost all of the technical motivation and innovation in the late 1990s and early 2000s in the field of QTL mapping addressed mapping precision, with less explicit consideration given to statistical power. There was, and still is a good reason for this focus of precision: once the right gene has been identified, it becomes possible to switch from genetic causality defined by loci and LOD scores, to actionable molecular mechanisms modulated by differences of protein expression or sequence. Thanks to many human GWAS, we now understand much better how to control the risk of false discovery using populations that incorporate more and more recombinations and complex admixture. One goal of this protocol is to help you get to a sweet spot with a balance of power and precision. A second goal is to help ensure that the results are robust and translatable.

Type 2 Questions: Questions related to gene-by-environment interactions (G×E) and treatment effects on phenotypes. These types of questions will be crucial to those interested in systematic manipulations of diet, environmental stressors, age, pathogens, drug exposure, and differences in social interactions. Mice and other inbred and isogenic model organisms are extremely well suited to evaluate complex experimental effects in the context of QTL mapping. The ability to impose well-controlled perturbations across large cohorts is among the strongest motivations to use model organisms. This kind of design is already the most common and critical in agricultural genetics.

Type 3 Questions: Questions related to the global genetic modulation of single traits or of systems of correlated phenotypes. These types of large-scale questions often fall under the heading of "genetic architecture." This term encompasses the analysis of many components of heritable and non-heritable variation, particularly the number and effect sizes of loci, independence and interactions among loci, and the roles of the environment, epigenetics, parental effects, and developmental noise (Mackay 2001). Oddly enough, before it became easy to map QTLs, these types of hard questions were at the heart of quantitative genetics. In fact, major branches of statistics had their birth in questions of genetic architecture, including

ANOVA and path analysis (Fisher 1918; Wright 1921). The diallel cross—the production of a matrix of F1 hybrids from inbred strains—is one of the mainstays of this type of quantitative genetics (Lenarcic et al. 2012). Recent examples include studies by Airey et al. (2001), Crowley et al. (2014), and Percival et al. (2016) who have used diallel sets of RI strains and the founders of the Collaborative Cross (CC).

Type 4 Questions related explicitly to predicting G2P relations. Given summed effects of gene variants (Type 1 questions), G×E interactions (Type 2), and the architecture of all sources of variance (Type 3), can we assemble predictive models of disease risk as a function of age, environment, diet, and drugs? This is the core question and quandary of precision health delivery. Precision medicine will have a short grace period, but if geneticists, molecular biologists, statisticians, and computational scientists have not delivered something impressive to match the hype, this term and the field risks being dismissed as a misnomer in the same way that *artificial intelligence* (AI) was dismissed and left unfunded for long periods—so-called AI winters. We need great experimental resources to generate and help validate predictions efficiently. The next section provides quick definitions and commentaries on the pros and cons of the important resources.

3. Pros and Cons of Resources and Crosses

We list of some of the major types of resources, from most simple to most complex in terms of level of genetic variation and complexity. The types of crosses and how they are generated are shown schematically in **Figure 1** with numbers that correspond to subsections.

1. Single fully inbred strains, such as DBA/2 and C57BL/6, are often the starting point for *in vivo* studies. We usually do not think of inbred strains in isolation as a resource for systems genetics, but a family of knockouts can be bred into a single isogenic strain (Dowell et al., 2010) or a single KO can be crossed into a hundred different inbred strains (e.g. Bennett et al., 2015) to generate interesting cohorts.

Large sets of distinct inbred strains incorporate a great deal of genetic variation (three are shown in Figure 1), and collectively may also be used as a core resource for systems genetics (Bogue et al. 2014). Genome sequence data are available for more than 36 inbred strains (Keane et al. 2011, www.sanger.ac.uk/science/data/mouse-genomes-project) most of which are also part of the Mouse Phenome Project (Bogue and Grubb 2004). Such collections of inbred strains—often termed diversity panels—provide a quick and ready resource for profiling how traits vary across a wide range of genomes, but there are not enough easily available strains to map QTLs effectively. Power is low and FDRs are high. However, sets of common inbred strains combined with sets of RI strains (more below) are an excellent joint resource for systems genetics—a combination called a hybrid diversity panel to which we return below.

The most commonly used inbred strains have often been split into sets of substrains. These will carry different sets of a few spontaneous mutations that have been picked up over decades of maintenance in different colonies. In mice, C57BL/6J and C57BL/6N are the genetic backgrounds strains used for almost all KO, KI, and transgenic modifications (www.mousephenotype.org).

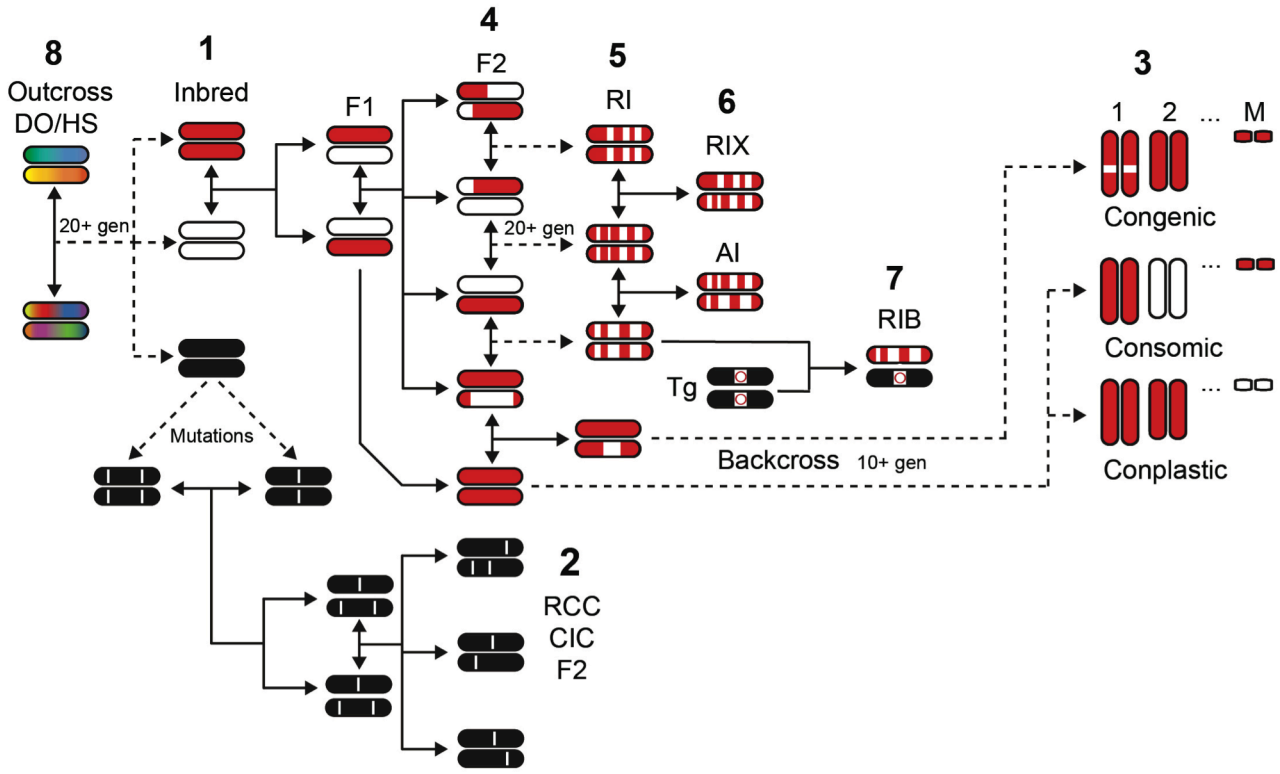

Fig. 1: Breeding schemes used to generate the resources. Short bars symbolize pairs of chromosomes. The colors (usually red, white, or black) denote the haplotypes/genotypes of the chromosomes. The large numbers correspond to Section 3 subheadings. Tg - a transgenic line. For all other abbreviations see text of Section 3.

2. Reduced complexity cross (RCC) or coisogenic cross (CIC): Both are novel types of "postgenomic" intercrosses between very closely related substrains (Kumar et al. 2014, Heiker et al. 2014) or even coisogenic pairs. For example, genomes of the C57BL/6J and C57BL/6N substrains differ at a total of about 36 known coding variants (Keane et al. 2011) but these substrains also differ for a surprisingly large numbers of phenotypes, including responses to several drugs and treatments (Khisiti et al. 2006, Mulligan et al. 2008, Simon et al. 2013, Kirkpatrick and Bryant 2015). BXD29/TyJ and BXD29-Tlr4<lps-2J>/J are a coisogenic pair that differ at two or three loci (Rosen et al. 2013). How is it possible to map an F2 that has almost no sequence variants? Once two substrains have been sequenced deeply (>30-fold coverage), there will almost always be a large enough number of spontaneous non-coding mutations to assemble a sparse genome-wide panel of SNPs and indels for mapping sources of phenotypic differences.

While the mapping precision of an F2 RCC or CIC will be poor (intervals of 20 Mb or more), the small number of segregating variants within any interval means that it can be practical to identify candidate QTGs and even QT nucleotides (QTNs) efficiently (Cardin et al. 2104). Kumar used this approach to define a mutation in *Cyfip2* that controls response to cocaine and methamphetamine (Kumar et al. 2014). The utility of an RCC in mapping and even in systems genetics points out that the key variable in "cloning" QTLs is not mapping precision per se but the number of polymorphic genes and sequence variants within a QTL's confidence interval. A 5–10 Mb interval containing only a single sequence variant will be far more easily reduced to cause and mechanisms than a highly polymorphic 0.1 Mb QTL containing five genes and hundreds of sequence variants (Li et al. 2010, Kumar et al. 2014).

3. Consomic and congenic whole genome panels: By backcrossing two inbred strains to each other while tracking genotypes of progeny over several generations, it is possible to effectively transplant whole chromosomes from donor strain *A* into recipient strain *B*. A full set of consomic strains will consist of 22 lines, each with one swapped chromosome plus the recipient control strains. There are now two sets of consomic strains—crosses of A/J or PWD/Ph into C57BL/6J (Singer et al. 2004, Gregorova et al. 2008). Buchner and Nadeau (2015) have considered the pros and cons of consomic sets (2015) and their efficiency relative to other resources.

A whole genome congenic panel is basically a finer-grained version of a consomic set, but now each strain contains only a piece of a single donor chromosome (Davis et al. 2005). The main utility of consomic and congenic sets is their high power to map phenotypes to single chromosomes. They have been used more recently to study epistasis and epigenetic effects (Buchner and Nadeau 2015). Their main disadvantage is that mapping QTLs requires the production of a secondary F2 intercross or a set of interval specific congenic strains. Whole chromosome effect sizes will almost inevitably decrease during this process (Bryant et al. 2012).

Off-target mutations and isogenic strains. One important factor to consider before using congenic and consomic strain sets is their sensitivity to spontaneous mutations that will accumulate gradually and progressively on the recipient (non-transplanted) background chromosomes. Spontaneous mutations or allele conversion events that arise on these other 20 chromosomes can cause variant phenotypes, and these new phenotypes risk being misattributed to putative variants on the donor chromosome, effect—essentially off-target effect (Williams 1999). It is therefore useful—sometimes even essential—to verify that traits map to the introgressed chromosome by making a small F2 from the consomic or congenic stock. Tracking down off-target variants is difficult because there are no known polymorphisms with which to map the other chromosomes. Sequencing consomic strains and using RCC methods is the obvious, but costly solution.

This raises a broad issue that applies to all crosses that are carried for many generations, including standard inbred strains, RI strains, AI progeny, and HS stock: what is the relative impact of inevitable *de novo* mutations on the measured phenotypes and results of different types? The good news is that for most of these resource types, new mutations will be unique to one strain or one case and do not segregate across the whole cross. Provided that the analysis and results are statistical collectives based on a large sample of strains or cases, then rare mutations, even those that are fixed in single strains, will simply be lumped as another source of error variance. In contrast, in situations in which mapping and other results depend on a single case and control—as when using congenic and consomic lines—there is a risk of misattribution of effects.

4. F2 intercrosses and backcrosses: The F2 intercross has been used widely in systems genetics, starting with the work of Damerval, Schadt, Lusis, and colleagues (Damerval et al. 1994, Schadt et al. 2003, 2005). Their main advantage is the ability to make large numbers of progeny quickly from almost any stock (usually inbred strains). F2 and N2 backcrosses have a structure that makes mapping and the analysis of covariance among traits simple. There is no need to correct for population substructure (see **Note 1**)—a problem that arises in almost any multi-generation cross (e.g. heterogeneous stock (HS), AIs, and RI strains, see Williams et al. 2001). It is practical to enhance the complexity and utility of an F2 intercross for systems genetics and for standard QTL mapping by making a four-way F2—for example by crossing A×B F1s to C×D F1s to produce AB×CD F2 progeny. This type of F2 is being used in a long experimental study of life span in mice (Miller et al. 2007).

5. Advanced intercrosses (AI) are simple extensions of F2s in which all subsequent generations are randomly bred, but with careful avoidance of sib matings (Darvasi 1998). The number of recombination events per AI case climbs steadily as the depth of the pedigree increases. At the 8th generation (about two years of breeding), 100 AI progeny, if made correctly, will provide about the same mapping precision as 500 F2 for Mendelian traits (Darvasi and Soller 1995). The countervailing problems with AIs are (1) the more complex logistics of using more than 100 breeders for up to 10 generations has a high cost, (2) the variable kinship among AI progeny needs to be factored into any kind of mapping or other statistical analysis, (3) the need for a significantly higher density of markers, and, perhaps most seriously (4) the loss of power associated with the increased number of recombinations per animal. A solution to some of these issues, first pointed out by Darvasi and Soller (1995), is to generate RI strains from AI stock—so-called Advanced RI (ARI) strains—and both the CC and many of the new BXD strains are actually ARIs.

Trade-Offs. There are important trade-offs between mapping precision and mapping power—the ability to detect QTLs with effects that account for a defined percent of the trait variance assuming a given sample size. As pointed out by Lander and Botstein (1989), the longer the genetic map, the higher the thresholds for statistical significance. The relation is complex, but **Table 1** provides a rough guide of tradeoffs. One column is marked **Recs/case** or recombinations per case, and a second column is marked **LOD Threshold**, or the linkage score that will often be needed to achieve genome-wide significance. **Recs/case** is an index of the potential precision of a resource, whereas the LOD score in this context is an inverse index of statistical power. High **Recs/case** are good for precision, but high LOD score requirements are bad for power.

The goal of course is precision with power. The simplest way to get both is to type larger and larger numbers of cases. (We are all familiar with the exceedingly high numbers of cases that must be used to achieve the genome-wide significance thresholds in human GWAS—typically LODs of around 8.) A better solution is to combine complementary resources—one optimized for power such as a conventional F2 or conventional RI strains, and one optimized for precision—such as the Collaborative Cross (CC), a Hybrid Diversity Panel (HDP), AI, HS or DO stock. The reason why joint resources are not used widely yet is because (1) many of the resources are new, and (2) the computational aspects of the analysis are more involved. But we now have powerful algorithms (e.g., Lippert et al. 2011, Zhou et al. 2012) that can handle dense genotypes and complex cohorts and covariates. Some of these are available online in the new version of GeneNetwork.

5a. RI strains: RI strains were originally made for mapping highly penetrant Mendelian traits (Bailey et al., 1971, Taylor et al., 1973), but they were eventually adopted for the analysis of complex traits (Gora-Maslak et al. 1991). RIs are now a key resource in systems genetics. Their main advantage relative to F2s and HS is that each unique genometype (genetic individual) is represented by a stable inbred strain that can be replicated in large numbers—essentially a sexually reproducing clone. RIs are therefore an excellent resource for studies that benefit from replication across individuals (e.g. dosing and toxicity studies of genotypes) or across environments (i.e. studies on G×E), and for the gradual assembly of deep phenome data that can be used in G2P analysis. In mice, there are now sufficient numbers of RI strains to allow for comparatively precise and well-powered QTL mapping. There are currently two major types of RI strains in mice:

(1) Classic two-parent RI strains. There are a total of about 340 of these types of mouse RI strains, including ~150 BXD available as live stock and many other small RI families: AXB/BXA (29 live), AKXD (20 cryopreserved), BXH (12 live), BRX58N (7

cryopreserved), CXB (12 cryopreserved), ILSXISS (60 cryopreserved), LGXSM (~18), NXSM (15 cryopreserved), SWXJ (13 cryopreserved).

(2) The Collaborative Cross (CC). This is a unique eight-way RI set of about 100 strains that is now in widespread use for QTL analysis and systems genetics (Threadgill et al. 2003, Churchill et al 2004). These strains are available both from UNC Chapel Hill and the Jackson Laboratory.

Classic RI strains that are derived from standard F2 intercrosses harbor more recombinations per genome—about 40 to 50—than do backcrosses (10 to 15), or F2 intercrosses (20 to 30) and therefore deliver better QTL precision than one might expect even with modest samples size (**Fig. 1**, note the alternating red and white haplotype blocks that make up the chromosomes of the RI strains). The ability to resample individuals also reduces non-genetic trait variance—effectively boosting heritability (Belknap 1998). Pandey and Williams (2014) computed the empirical precision of *cis*-acting expression QTLs (*cis*-eQTLs) in the BXD family across the whole genome at different mean LOD scores and at different marker densities (their Figure 8.6). With a cohort of 67 strains and using only two samples per strain, eQTLs with LOD scores of between 3 and 5 were located within ±2 Mb of the cognate or parent gene. Those with LOD scores above 8 were typically within ±1 Mb. Corresponding empirical mapping precision based on cis-eQTLs can now be easily computed for many resource types across the whole genome using data sets and queries built into GeneNetwork (Mulligan et al. 2016, this volume). Examples of doing this for a large AI (n = 811) and a well matched AI-derived RI set (n = 40) are given in **Note 2**.

The CC RI strains are capable of even better mapping precision than standard RIs for two reasons. First, the recombination load (the crossover probability) of CC strains is 1.75 times higher than that of typical two-parent RI strains due in part to the rounds of intercrossing required to merge all eight genomes (Table 5 of Broman 2005). Second, the inclusion of multiple parental genomes within the CC means that it is possible to carry out a fine-grained haplotype contrast analysis that can effectively reduce QTL intervals and numbers of QTG candidates (Yalcin et al. 2005). Haplotype contrasts of the same general type can also be exploited using combinations of conventional RI families, inbred strains, and F2 crosses (e.g., Taylor et al. 1973; Williams et al., 1998, Malmanger et all 2006, Furlotte et al. 2012).

The most important disadvantage of conventional RI strains and other standard two-parent crosses is that they segregate for only a fraction of all known polymorphisms. For example, the BXD family segregates for a total of ~5.2 million sequence variants—about 44% of common variants among standard inbred strains (Roberts et al. 2007). Some stretches of the genome will be almost completely identical by descent (e.g., Yang et al. 2007, Li et al., 2010) and these regions will not normally contribute much to trait variance. This disadvantage however may also be viewed as an advantage when trying to dissect a QTL, since the load of polymorphisms within an interval may be an about 6-fold lower than that of the corresponding interval in the CC or DO stock, and thus the number of viable candidate genes may be much reduced. As shown by Li and colleagues, phenotypes that map into these genetic blindspots can be particularly easy to map to QTNs (Li et al. 2010).

Table 1: Comparisons of resource types

Type of Cross	Recs/case	LOD Threshold	$/Genotyping	$ Case*	Isogenic	Inbred	Phenome	GXE	Breeding	Ref
Consomic and Congenic Sets	1	1 to 2	0	140	Yes	Yes	Yes	Easy	Variable	14, 55
Reduced Complexity Cross	25	1 to 2	25	20	Almost	Almost	Hard	Hard	Easy	44, 45
F2 intercross, 2-way or 4-way	25	2.5 to 3	25	15	No	No	Hard	Hard	Easy	8, 16
Advanced intercross	100	4 to 5	100	100	No	No	Hard	Hard	Hard	9, 10
RI strains and Advanced RI Strains	50 to 80	3 to 4	0	140	Yes	Yes	Yes	Easy	Variable	4, 8, 22
Advanced Intercross RI strains	80	4 to 5	0	140	Yes	Yes	Yes	Easy	Variable	4, 8
RI Intercross F1s (RIX, RIB)	100 to 200	4 to 6	0	50	Yes	No	Hard	Easy	Easy	36, 38, 40
Hybrid Diversity Panel (HDP)	1000	6+	0	20–150	Yes	Yes	Yes	Yes	Easy	18, 19
Collaborative Cross (8-way RI)	135	4 to 6	0	195	Yes	Yes	Yes	Easy	Variable	13, 17
Diversity Outcross (DO HS)	400+	5 to 7	100	55	No	No	Hard	Hard	Easy	84, 85
Outbred Stock (e.g. CD-1, CF-1)	1000	6 +	100	7	No	No	Hard	Hard	Easy	68, 79

A practical disadvantage of RI strains is that they often have poor breeding performance compared to many F2s and outbred stock. While BXD strains average 4–5 pups per litter, some are hard to maintain and can be sensitive to housing conditions. Many CC lines have even lower fecundity. This is one reason why many inbred strains are so much more expensive than outcross or HS animals (**Table 1**) and why they are often cryopreserved rather than kept as live stock. This issue was also a factor motivating the creation of the DO: it provides a way to stabilize recombinations events that were at risk of extinction (Gary Churchill, personal communication). Speaking of the obvious, a final disadvantage of RI strains that is that they are inbred—an anomalous genetic architecture that will not only decrease fitness but will often increase trait variance relative to isogenic F1 hybrids due to the loss of heterosis and allele buffering.

5b. Advanced RI lines. There are also several interesting variants of RI strains. The first of these are highly recombinant RI strains generated from AI progeny (Darvasi and Soller 1995). Many of the new BXD strains (BXD43 and higher) are AI-derived (Williams et al. 2001, Peirce et al. 2004), as are all of the LGXSM strains (Hrbek et al. 2006). Instead of directly inbreeding siblings of an F2, progeny are crossed to avoid sib matings for as many as 30 generations, prior to the inbreeding phase (another 20 generations). The main benefit of using AI stock for making RI strains is a significant increase in potential QTL mapping precision (**Note 2**), but as usual, with loss of power.

5c. RI Backcrosses. The second variant involves making a set of F1 intercrosses between RI strains and a single inbred strain—usually one that carries interesting modifier alleles with a dominant or additive effect. For example Hunter and colleagues crossed 18 AKXD RIs to an FVB strain carrying a dominant cancer gene variant to map modifiers of metastasis (Yang et al. 2005). They refer to this cross as an RI backcross (RIB) because the 18 sets of F1s are similar to a backcross—those chromosomes inherited from the RI parent are recombinant, whereas those inherited from the other strain are not. This idea can also be generalized across multiple RI sets and inbred strains. For example, Bennett and colleagues crossed an APOE transgenic strain to more than 31 common inbred strains and 66 BXD, AXB/BXA, BXH, CXB RI strains (Bennett et al. 2015) to study the genetic architecture of atherosclerosis.

6. RIX panels: RIX panels are a clever new extension of RI strains that have some interesting advantages over RI strains and HS. Given a set of 10 RI strains, it is simple to cross all of them to each other: 1×1, 2×1, 3×1, 3×2 and reciprocal crosses 1×2, 1×3, 2×3, and so on. From

only 10 starting strains one can produce a full diallel set made up of 100 isogenic sets of F1. In a full diallel we do not gain much precision by resampling the same parental haploid genome in different combinations (1×2, 1×3, 1×4 etc.). While no new recombination event occur in making these F1s, one does expose an interesting range of phenotypes, such as those exploited by Rasmussen and colleagues (2014) to develop mouse models of Ebola infection.

What makes RIX particularly attractive now for both mapping and systems genetics is that we have several large sets of RI strains—more than 100 BXDs and close to 100 CC lines. While it is not practical to generate or study a full 200 × 200 matrix of 40,000 RIX progeny and founders, it is practical to sample all 200 of these RI genomes by making 100 non-overlapping sets of RIX litters: 1×2, 3×4, ... 198×199, and 199×200. And two different RI sets can be crossed (e.g., BXD1 to CC001). A set of 100 disjoint (non-overlapping) RIX progeny solves a number of problems—(1) efficient sampling of large RI families that exploits all recombination events in the parental RIs; (2) much lower inbreeding coefficients than inbred parents; (3) genetic complexity much more like that of human populations; (4) ability to study parent-of-origin and dominance effects; (5) fully defined genomes, and (6) deep replication of any particular RIX to increase phenotype precision, (7) more direct analysis of G×E using precisely the same genometypes under two or more conditions; and as a (8) powerful resource to test predictive models of G2P relations.

Disadvantages of RIX sets include the following: (1) they can be costly to generate compared to HS or DO stock; (2) there will be a loss of genetic variance associated with the heterozygozity of RIX progeny compared to homozygous parents (Hegmann and Possidente 1981); (3) breeding and cohort logistics are somewhat more complicated and expensive; and (4) it will be difficult for a community of researchers to define a single reference set of RIXs to use for collaborative phenotyping because there are such huge numbers of potential RIX that can be made.

7. Hybrid Diversity Panels (HDP): A hybrid diversity panel is an aggregate of RI strains and common inbred strains that are usually phenotyped together and used as a single joint mapping resource (Williams et al. 1998, Overall et al. 2009, Ghazalpour et al. 2012). They are used for at least two reasons: (1) to achieve comparatively high mapping precision (intervals of 1–5 Mb) that can match those of HS and DO stock using as few as 100 inbred strains; (2) to make it possible to assemble large phenomes that can be used for G×E analysis. A HDP does not have a rigid definition, and a mouse HDP could and should include CCs, BXDs, and even RIX. Depending on its membership of isogenic genometypes, an HDP will share some of the same problems of any one RI family, but to a lesser degree. For example, the issue of genetic blind spots will be less serious except for a few regions of the genome that tend to be identical-by-descent even in the CC. The main problem of an HDP is the generally low to moderate fecundity of members and their high acquisition costs.

8. Outbred stock (OS), Heterogeneous stock (HS), and Diversity Outcross stock (DO): Outbred stock (OS)—often referred to as Swiss Webster stock (Lynch 1969, Chia et al. 2005)—are the progeny of a nine albinos (two males and seven females) imported from a colony in Lausanne to New York in 1926. They were subsequently distributed to researchers and commercial vendors worldwide as "standard laboratory" mice. As expected given this history, OS do not incorporate much genetic variation. Genomes of 66 OS colonies studied by Yalcin and colleagues (2010) were heterozygous at no more than 34% of polymorphic loci, and a significant number of colonies were almost fully inbred. The theoretical attraction of some OS colonies is their potential high mapping precision with LD blocks that are only a few hundred kilobases.

HS and DO could be considered variants of OS, but here we use a modern definition of HS and DO as special stock generated from well-structured intercrosses and outcrosses

among diverse sets of inbred progenitor strains. HS are almost always maintained using larger colonies—50 or more breeding cages—and breeding schemes that minimize mating of closely related individuals. One original motivation to make HS was to produce new models by intercrossing diverse strains and then selectively breeding progeny for high and low phenotypes in responses to drugs, alcohol, and other treatments (Kakihana et al. 1966, Holmes et al. 1986). The Northport HS (HS-Npt) made by intercrossing A/J, AKR/J, BALBc/J, CBA/J, C3H/HeJ, C57BL/6J, DBA/2J, and LP/J is a good example (Hitzemann et al. 1994). HS have also been used for high precision QTL mapping (Valdar et al. 2006).

The DO is an example of a modern HS made by intercrossing early generations of the CC (Churchill 2012, Svenson 2012). DO mice are significantly more diverse even than HS-Npt or outbred stock for the simple reason that three of the progenitors of the DO and CC—PWK/Ph, CAST/Ei, and WSB/Ei—are inbred strains derived from highly diverse wild *Mus* species and subspecies. DO cohorts are now at the 22nd generation (G22) of outcrossing. The DO segregate for well over 40 million common sequence variants with minor allele frequencies above 10%. These animals breed well and incorporate 4 to 6-fold more genetic variants than the number of common variants in human populations.

There are two key advantages of DO and HS: (1) they have a genetic complexity that equals or exceeds that of most human populations. They are excellent models for precision medicine; (2) like AI cohorts they gradually accumulate large numbers of recombinations and therefore can resolve QTLs with high precision; (3) the high genetic diversity among parental strains ensures that phenotypes will be highly variable and that most regions of the genome will be polymorphic, and finally; (4) they usually have excellent breeding performance, a feature that reduces costs.

The main disadvantages of HS and DO stock is the inevitable flip side: the high recombination load will reduce statistical power and the high genetic complexity and numbers of haplotype can make it difficult to resolve single linked QTGs and QTNs. The last and most obvious experimental disadvantage is that HS and DO animals are genetically unique. This means that it is more difficult to acquire phenomes for these types of resources or to use them as effectively in G×E studies.

4. A Multipurpose Design for Systems Genetics

In this section we consider some of the designs that can now be used to address the four types of questions in Section 1. In the first section below (4.1) we consider Type 1 questions with a focus on mapping precision. In the second section, we start to wrap everything together by considering a single adaptable design for systems genetics that will be good for most discussion purposes. We comment on ways to modify or extend this multipurpose design using a Question and Answer format. Much of the text is summarized in Tables 1 and 2.

4.1. Genotypes and genetic maps: What mapping resolution is needed?

The goal is usually to get down to about 1 Mb precision as efficiently as possible. Assume we are completely naive—we only know what traits interest us and that it is somewhat variable among individual mice belonging to a few strains or stocks. We do not have estimates of heritability and we do not yet know what strains or crosses would be most useful.

One of the best resources in this situation is to study phenotypes in a small number of strains and F1 hybrids between these strains. This made sense several decades years ago (e.g., Taylor 1973, Williams et al., 1996) and it makes even more sense today (e.g., Graham et al.

2015) because these initial "survey" data can eventually be wrapped into a mapping study with all other resources—whether HS, DO, CC, or RIX. For example, a study of 6 individuals each of 18 isogenic groups, such as sets of fully inbred strains some of their F1s, will answer questions related to trait heritability, trait dominance, and if you are lucky, even give you hints about genetic complexity and architecture. It may be possible to evaluate if the trait or disease phenotype is controlled by a small number of QTLs (the oligogenic model) or by hundreds of QTLs (the polygenic or "infinitesimal" model) (Taylor 1973, Williams 1996). This 120-case study will also enable you to perfect phenotyping and learn much more about sources of technical error, sex differences, and selecting better resources for the next stages.

The main risk in this type of pilot study is batch effect and phenotype rift. Systematically phenotyping strains A through R at a systematic pace of one genotype per week over 4 months is a poor experimental design, since temporal variance and drift will masquerade as a heritable difference among lines. Interleave the phenotyping to study 10 different genotypes with 1 or 2 individuals each for the first phase of the experiment and then repeat cycles as needed. An interleaved design may not be feasible in all situations, in which case consider re-phenotyping well-known strains throughout a study to check for drift.

Mapping precision: While more mapping precision is always a good thing, there is not much justification to refine maps down to much less than confidence intervals of 1–2 Mb. Intervals of this size can now be efficiently dissected using an impressive and diverse array of data resources—including of course, full genome sequence for all genes in all strains. A small number of candidate genes and variants can now often be tested efficiently now using genetically engineered mice, fish, flies, worms, or human GWAS data sets, *in vitro* analyses, or even phenome-wide association (Wang 2016).

Another reason not to obsess about precision much below 1 Mb is the fuzzy functional definitions of genes. This is highlighted by a recent analysis of one of the strongest loci that modulates obesity in humans—SNPs within intron 1 of the *FTO* gene. While the position of linkage is not in question, these SNPs apparently tag variants in a long-range enhancer of *IRX3*—a small transcription factor 0.5 Mb distal (Smemo et al. 2014). This emphasizes that functional validation is critical, and that the law of diminishing can kicks in with some force under 1 Mb. We consider a 0.5 to 1 Mb as a reasonable goal that can usually be achieved efficiently using a combination of resources described below. This is not quite as precise as what can be achieved with large GWAS, but unlikely human studies we can efficiently transition to molecular mechanisms.

4.1. Assumptions

To develop this multipurpose design we assume almost nothing other than that the traits of interest are heritable and genetically complex, and that the initial focus is not on G×E or treatment effects, developmental stages and ages. We will come back to extensions that these types of questions toward the end of this section.

Sample size and costs of stock. As our starting parameter, we budget for 240 individuals per year over a four-year period—960 cases total at a pace of 20 per month and 1 per day. This is a modest throughput that should be adaptable to almost any type of study, even electrophysiology, advanced imaging and behavioral methods. The cost of mice may range from as little as $20 per case to as much as $200. Standard inbred strains such as those used to generate the CC cost between $20 (C57BL/6J) and $200 per animal (WSB/EiJ) with an average of $102. The average price for most of the resources discussed in this chapter is currently about $150 per case. An experiment using 240 cases/yr will typically require a budget of ~$40,000/yr. Housing costs are variable, but it is safe to assume 25 to 50 cages will

incur a cost of $10,000 to $20,000/yr. If cases must be genotyped (e.g., F2, HS, and DO stock), then factor in a charge of as much as $100 per case (Table 1).

Sex balance. Whenever possible males and females should be used in roughly equal numbers and concurrently. Not only is the use of both sexes becoming a mandate, but results will also be more interesting and robust in terms of their translational relevance. Finally, sex differences can provide mechanistic insight. The inclusion of both sexes in a design does not double the required sample size, even when using isogenic cohorts of RIs, RIXs, or HDPs. A balanced sample of just one or two males and females across multiple genometypes can be a powerful design to detect sex differences. Of course, sparse sampling does not address sex differences within any single strain, but this is a topic that may be worth revisiting in a second phase of work.

While it may look tidy in a Methods section, it is not necessary to get numbers of cases balanced precisely either by sex or genotype. Do not obsess about filling every cell in a design uniformly. If you must obsess about anything, make it (1) batch confounds, (2) drift in phenotyping standards, and (3) quality control for electronic records and case identifiers. When possible consider whether litter effects are a confounding factor in phenotype variation. This is a particular risk for RIX designs in which one single litter may be used for each genotype.

Table 2: Multipurpose design for systems genetics

Group	Types	Type Notes	N	Reps	M	F	Months
Group 1A	8	Fully inbred strains	48	6	3	3	2.4
Group 1B	4	4 F1s using 8 genometypes	24	6	3	3	1.2
Group 1C	8	Your choice	48	6	3	3	2.4
Group 2A	40	CC or BXD, AXB (exploratory)	160	4	2	2	8
Group 2B	40	CC or BXD, AXB (selective)	160	4	2	2	8
Group 3A	40	RIX (semi-exploratory)	160	4	2	2	8
Group 3B	40	RIX (selective)	160	4	2	2	8
Group 4A	100	DO or HS (predictive)	100	NA	50	50	5
Group 4B	100	DO or Your Choice	100	NA	50	50	5
Sums	380		960				48

4.2. Experimental Design for Systems Genetics (see Table 2)

Stage 1: Heritability, technical robustness of assays, effects of sex, and genetic architecture. The main purpose of phase 1 is make sure we understand more about the main sources of variance of phenotypes. It is well worth a 3–6 month pilot to make sure the phenotyping methods and assays work well. The data from this initial work will eventually be useful for mapping.

Group 1A: 6 individuals each of 8 inbred strains. It would make great sense to start with the parents of the CC. Depending on your field of study you could add or substitute AKR, BALB, DBA/2J, FVB or other common strains,
n = 48

Error-checking. Since assignment errors can destroy your results, keep track of coat color, and even better, save tails of animals for post-hoc genetic verification. This is important for all stages of the work.

Group 1B: 6 individuals from each of 4 F1 hybrids made using strains A through H (AB, CD, EF, GH, or the reciprocals AB, BA, CD, DC). The parental strains for the F1s can be selected based either on greater genetic differences or on contrasting phenotypes.
n = 32

Group 1C: 6 individuals from each of 8 additional types based on the initial results above, or to encompass other interesting strains selected from the Mouse Phenome Project (phenome/jax.org) or based on any interest you have in RCC methods. You could also use this set of 48 cases to resolve problems or seize opportunities. This set could include F1 hybrids.
n = 48

Question 1: Is 6 samples per type really enough? ANSWER: If you are not examining different environmental factors, then yes. In fact, you probably should not do 6 per type at any one time or from only 1 or 2 litters, but break work into analysis of 2–4 cases for each of 12 types, and generate data over several batches. You may want to run pairs of male and female (litter mates even) in single batches, since you are likely to be used paired *t* tests. If you find that the batch effects are large, then you have learned something important and may need to rethink the design of the larger study. If you find that there is variation as a function of age, you have also learned something important. Furthermore, after phenotyping 6 per type, you will have a good idea if any particular type needs to be resampled to higher Ns. **Note 3** discusses some of the factors that should be considered when selecting number of biological replicates.

Question 2: Should I use wild strains such as PWD/PhJ, CAST/EiJ or WSB/EiJ? ANSWER: Yes, unless there is some specific contraindication, such as cost, availability, or wildness (Wahlsten 2003). There is no reason to not expose yourself to the remarkably wide range of phenotypes at this stage. (Make sure you unbox wild strains carefully or you will have stories to tell.)

Question 3: Should I use HS or DO stock initially? ANSWER: No, not unless you have already used these types of resources or need them to address a specific hypothesis. You cannot estimate heritability from a single cohort of HS animals

Question 4: Should I phenotype pairs of closely related substrains? ANSWER: Probably not at this stage unless you already know that there are significant differences in related phenotypes among substrains. If you are interested in exploiting RCC methods then include pairs or trios of substrains in Group 1c. Genetic variance will be lower in substrain contrasts, so you will need to increase sample size to 8–12 per type.

Question 5: Why are F1 hybrids useful? ANSWER: For at least these three reasons: (1) F1 hybrids are used to evaluate effects of gene variants on phenotypes in organisms with a more typical heterozygous genome. F1 hybrids are isogenic so they have many of the advantages of inbred strains. (2) F1 hybrids also enable us to evaluate whether phenotypes are dominant or recessive. (3) Reciprocal F1s can be used to study parent-of-origin effects on phenotypes. Note that some of these advantages do not apply to F1s between closely related substrains.

Stage 2: Low resolution mapping and systems genetics. The purpose is to understand the genetic complexity of phenotypes by low-resolution mapping but with good power. If there are a few QTLs with large effects then even a cross with 40 genometypes will highlight one or two loci. Since we rely on RI strains for this first analysis, it should be possible to compare all new data with all previously generated phenotypes and QTLs. We can be confident to find some interesting leads, generate new hypotheses, and perhaps even gain mechanistic insight.

Group 2A: 4 each of 40 RI strains. Use 4 each if heritability is <0.4, otherwise consider using 2 each of 80 strains, particularly if you suspect that trait variance is controlled by a major effect locus. You can always return to the RI strains to boost your samples size.
n = 160

Group 2B: Same as above, but using a new set of 40 RI strains. You will now already know if you have detected suggestive or significant QTLs. If the answer is yes, then you can selectively phenotype those RI strains that have recombinations between the right haplotypes in the right regions. You might also want to replicate any outlier strains detected in Group 2A. If the results from Group 2A do not yet provide compelling candidates, then just forge ahead with more or different RI strains.
n = 160

Question 6: Could I not use RIX in Group 2B? ANSWER: Yes, since you will have RI strains available, this is an option. However, the RIX will not provide you much more genetic signal unless you use different RI parents to make the RIXs. RI and other fully homozygous strains have twice the genetic variance of F1 hybrids. This gives them a power advantage at early stages of mapping.

Question 7: Should I use BXDs, AXBs, or the CC strains? ANSWER: The CC will almost always be a good choice, as they are likely to exhibit the highest phenotypic variance in any target phenotype. BXDs and AXBs will provide better mapping power *per case* due to their lower genetic complexity, but this benefit can be neutralized by less phenotypic variance. If the parents of the RI panels differ markedly and your focus is more on systems genetics than mapping precision (e.g. C57BL/6J vs DBA/2J), then the BXD may be the best first choice for the simple reason that so much data has been accumulated for these strains. Availability of RI strains can sometimes be the main constraint.

Question 8: Can I mix CC stains with other RI panels? ANSWER: Yes, and this is precisely the motivation for resources such as the HDP. It is probably a good idea to sample at least 16 strains in any one RI set so that you can evaluate whether or not a locus is segregating and so that you can estimate trait covariance to some degree among phenotypes within single RI families.

Question 9: Should I use consomic or congenic panels for this work? ANSWER: No; not unless your screen in part 1 included PWD/Ph and A/J and suggested that these strains differed markedly from C57BL/6J. These are the strains that have been used to make consomic sets. Consomic strains can have good power if you sample each of 20 strains with 6 or more cases, but to achieve mapping precision (±5 Mb), you will have to generate your own derivative crosses, and effect sizes of loci can evaporate during the production of congenics (Bryant 2012).

Question 10: How do I handle outlier strains in the initial QTL analysis? ANSWER: Transform data so that outliers do not have an overwhelming effect on maps and other

statistical results. You can winsorise high and low outliers or use a logarithm transform. Replicate outliers if you suspect technical error.

Stage 3: High resolution mapping and more systems genetics

Group 3A: 4 each of 40 sets of RIX progeny that are produced by crossing within or even across sets of RI strains. You will need 80 RI strains to make 40 non-overlapping RIXs. Vendors may be willing to do this for you if the strains are not available to you. At this point you will almost surely have a small set of reasonably well mapped loci. You will also have enough data to decide if you want to reevaluate your questions. Are you really after QTGs, do you want to test a specific intervention, or do you want to try your luck at G2P prediction using a set of molecular and genetic biomarkers? This first set of 40 RIX progeny should enable you to do all three.
n = 160

Group 3B: Same as above but this set could be generated to test an intervention or age (using Group 3A as a control). Or this RIX group could be created selectively to test multilocus interactions or parent-of-origin effects.
n = 160

Stage 4: High resolution mapping, predictive validation, and systems genetics. The combined results of the three stages should have left you with a set of loci mapped to less than 2 Mb. If that is not the case, then this final stage should help achieve that goal. Ideally, you might want to select DO stock on the basis of genotype, and that may be a service that will soon be available. This would be most useful if only one specific haplotype contrast is generating trait variance (e.g., a 1 vs 7 split of haplotype effects).

Group 4A: DO or HS. DO stock will probably be most accessible and also generally most suitable.
n = 100

Group 4B: Your wildcard. You could continue with a second set of 100 DO mice if the first results strengthened results. Or you could use the DO mice you still have to selectively cross animals with specific combinations of alleles. This would require selective genotyping of specific SNPs. DO mice are a wonderful source of genetic variance, but you may want to select or trim back some of those variants. This will position you well to predict phenotypes based on combinations of haplotypes at two or three loci.

Alternatively, use this group of cases for further studies on the effects of treatment, age or stage (see Group 3B)
n = 100

Question 11: How do I genotype DO or HS? **ANSWER:** Even in the most demanding situation of mapping DO, HS, and wild caught populations markers need only be about 100 Kb apart (Yalcin et al. 2010), and since the mouse genome is about 2.5 Gb, 100,000 well chosen markers will be more than adequate. Virtually any population, no matter how complex its genetic architecture can now be typed using the latest version of the mouse universal genotyping array (the GigaMUGA) or by sparse sequencing for about $100/case (Rat Genome Sequencing and Mapping Consortium et al. 2013).

For selective genotyping of a handful of markers in DO or RCC F2 intercrosses you can use standard protocols that will probably require acquiring sets of PCR primers. Costs may be as high as $1/genotype/case. If you require a few hundred markers per case then a good ballpark cost for custom genotyping is $0.10–0.20 per genotype per case—or

$20–40 for 200 markers for an F2 progeny. Finally almost all inbred, RI strains, an RIX progeny are already well typed and there is no cost at all.

Question 12: Is there a strong justification to use all of these types of resource—RIs, RIX and HS/DO? ANSWER: These resource types perform many of the same functions. However, G×E will be easier to study using RI and RIX. RIX progeny made using CC RIs are genetically similar to DO animals, but incorporate fewer recombinations per animal. Data from RIX cases can also be used to build up a phenome database and are potentially more useful for large collaborative teams, but this advantage may remain theoretical for the next several years. DO/HS animals are logistically far easier to obtain and provide you with access to the ultimate breadth of genetic and phenotypic diversity. They are the closest you can get to a wild-type mouse population short of capturing your own. If you results from Stages 1 to 3 are supported in DO populations, then you can be sure that results will have the maximum replicability and perhaps even translatability to human populations. You may also be able to computationally and genetically "extract" specific disease models from RI, CC, and DO stock.

5. Future Directions and Conclusions

Thanks to the massively reduced cost and increased scope of omics technologies, it is now feasible for small collaborative groups—and even single research groups—to execute large studies in systems genetics. We can anticipate this paradigm to continue, and even accelerate, in the coming years with the advent of new and improved methods of quantifying an individual's proteomes, metabolomes, metagenomes, and epigenomes as a function of cell type, tissue, age, and state. It is great to have the core resources that are needed to take advantage of this rapidly expanding set of omics technologies.

What we have not considered in this chapter is the analytic and synthetic tools needed for high-content systems genetics? How do we actually map aggregated data from 1000 cases with complex substructure? How do we build predictive models and test their fit to empirical data? These questions are taken up in many of the chapters in this volume.

6. Notes

1: What is population substructure and how does it make statistical analysis and mapping trickier? We all have learned that observations used in many statistical tests should be independent. In genetic crosses all F2 progeny are usually treated as independent observations. But what if there are strong litter effects, or batch confounds due to technical errors. These effects can introduce variance into a cross that can obscure the detection of the genuine effects and produce spurious linkage. Similarly, in an AI cross, one mating pair may produce 50 siblings whereas another mating pair produces only 5. In this case we have known and unbalanced pedigree substructure that needs to be corrected even when doing something as simple as computing a correlation coefficient. Large GWASs sometimes combine data from different ethnicities and it is also essential to correct statistically for the kinship relations among members. In some cases we can use the genotypes of cases to compute a matrix of kinship similarity, and use this matrix to correct for the population substructure. If we know

the litter and batch identifiers we can also adjust for these nuisance variables in a statistical model.

In large RI sets such as the BXDs and CC, there is cryptic substructure that may not show up easily in genotypes but that that may still be important. The BXDs for example, were generated in multiple cohorts between 1970 and 2013 using the same parental strains—C57BL/6J and DBA/2J, but of course, 43 years of breeding history will add many new variants to both parents and some of these are already well known to have important effects (Anderson et al. 2002, Wang et al. 2016).

2: To estimate empirical precision for QTLs across a population in GeneNetwork (www.genenetwork.org) you first need to select an expression data set from the pull-down menu. In this example, select **Species** = *Mouse*, **Group** = *B6D2 AI PSU*, **Type** = *Muscle mRNA*, and **Data Set** = *PSU B6D2 AI Muscle...*
Enter this query into the **Get Any** box:

cisLRS=(23 46 50)

where cisLRS is the linkage statistic specifically for the cis-acting eQTLs. The first two values in parentheses are the minimum and maximum LRS values to return (LRS = LOD x 4.61), and the final number is the size in megabases of the acceptance window used to define how close a gene must be to the QTL peak to be considered cis-acting. In this case the acceptance window is very broad, and the peak LRS can be anywhere 50 Mb on either side of the gene.

This search will generate 2086 hits. You can resort and download the results as an Excel table using the **Download Table** button. In this large AI intercross with more than 800 cases generated by Carbonetto and colleagues (2014) between C57B/6J x DBA/2J, the mean offset between 2000 cis-eQTLs with LOD between 5 and 10 and their genes is 7.0 ± 0.21 Mb.

If you try precisely the same set of operations with a matched BXD Advanced RI data set (*EPFL/LISP BXD CD+HFD...Exon Level*) you will find that the mean offset between 4400 cis-eQTL in this data set is 2.0 ± 0.06 Mb. This latter ARI data set is based on ~80 cases (1 replicate of 40 strains under two conditions—high fat and standard chow diet (Williams et al., 2014).

3. Genetic studies usually benefit more by increasing the *n* of genometypes that are phenotyped than by increasing the *n* of replicates per type (e.g., figure 1B in Andreux et al, 2012 and Belknap 1998). All else being equal, a studying of 160 types without replication should be superior in terms of QTL results to one of 40 strains and 4 replicates of each. This is obvious for Mendelian traits such as coat color, but it also holds true for quantitative traits—even those with low heritability. However, at an early stage of a study it is vital to understand heritability and technical confounds and in some cases, replication is easy and cheap. For this reason, it is a good idea to begin work with 6 to 8 replicates of a few "reference" genomes. When using isogenic cases we recommend two replicates minimum, one per sex. Bumping this up to two per sex per strain will improve the comfort level of many reviewers, although to keep them happy you will probably need 6–8 per group. There are also some good reason to study 6 or more cases per genometype even after heritability is known: such as studies of genetics control of variation itself (Rönnegård and Valdar 2011) or pharmacological effect thresholds.

One way to think about the diminishing returns of high replication rates is to compare *t* scores and *z* scores required to achieve statistical significance for simple two-sample comparisons using different sample sizes. The *z* score assumes variance of the population is known and the critical value to reject the null at alpha 0.05 is $z = 1.96$. In contrast, the *t* score

estimates variance from the sample itself, and the critical values start at a woefully high 12.71 for n = 2, but drops toward the asymptote of 1.96 very quickly: 3.182 for n = 4, 2.757 for n = 6, and 2.201 for n = 12.

Acknowledgment

We thank the support of the UT Center for Integrative and Translational Genomics, and funds from the UT-ORNL Governor's Chair.

References

Airey DW, Lu L, Shou S, Williams RW (2002) Genetic sources of individual differences in cerebellum. The Cerebellum 1: 233–240.

Anderson MG, Smith RS, Hawes NL, Zabaleta A, Chang B, Wiggs JL, John SW (2002) Mutations in genes encoding melanosomal proteins cause pigmentary glaucoma in DBA/2J mice. Nat Genet 30: 81-85.

Andreux PA, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, et al. (2012) Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. Cell 150: 1287-1299.

Bailey DW (1971) Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation 11: 325-327.

Belknap JK (1998) Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behav Genet 28: 29-38.

Bennett BJ, Davis RC, Civelek M, Orozco L, Wu J, et al. (2015) Genetic architecture of atherosclerosis in mice: A systems genetics analysis of common inbred strains. PLoS Genet 11: e1005711.

Bogue MA, Grubb SC (2004) The mouse phenome project. Genetica 122: 71-74.

Bogue MA, Peters LL, Paigen B, Korstanje R, Yuan R, et al. (2014) Accessing data resources in the mouse phenome database for genetic analysis of murine life span and health span. J Gerontol A Biol Sci Med Sci. 71: 170-177.

Broman KW (2005) The genomes of recombinant inbred lines. Genetics 169:1133-1146.

Bryant CD, Kole LA, Guido MA, Sokoloff G, Palmer AA (2012) Congenic dissection of a major QTL for methamphetamine sensitivity implicates epistasis. Genes Brain Behav. 11:623-632.

Buchner DA, Nadeau JH (2015) Contrasting genetic architectures in different mouse reference populations used for studying complex traits. Genome Res 25: 775-791.

Cardin S, Scott-Boyer MP, Praktiknjo S, Jeidane S, Picard S, et al. (2014) Differences in cell-type-specific responses to angiotensin II explain cardiac remodeling differences in C57BL/6 mouse substrains. Hypertension 64: 1040-1046.

Chia R, Achilli F, Festing MF, Fisher EM (2005) The origins and uses of mouse outbred stocks. Nat Genet 37: 1181–1186.

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD et al. (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet. 36: 1133-1137.

Churchill GA, Gatti DM, Munger SC, Svenson KL (2012) The diversity outbred mouse population. Mamm Genome 23: 713–718.

Complex Trait Consortium (2003) The nature and identification of quantitative trait loci: a community's view. Nature Genetics Reviews 4: 911–916.

Crowley JJ, Kim Y, Lenarcic AB, Quackenbush CR, Barrick CJ, et al. (2014) Genetics of adverse reactions to haloperidol in a mouse diallel: a drug-placebo experiment and Bayesian causal analysis. Genetics

196:321-347.
Damerval C, Maurice A, Josse JM, de Vienne D (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. Genetics 137: 289-301.
Darvasi A (1998) Experimental strategies for the genetic dissection of complex traits in animal models. Nat Genet 18: 19-24.
Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199-1207.
Davis RC, Schadt EE, Smith DJ, Hsieh EW, Cervino AC, et al. (2005) A genome-wide set of congenic mouse strains derived from DBA/2J on a C57BL/6J background. Genomics 86: 259-270.
Dietrich WF, Copeland NG, Gilbert DJ, Miller JC, Jenkins NA, et al. (1995) Mapping the mouse genome: current status and future prospects. Proc Natl Acad Sci USA 92: 10849-10853.
Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, et al. (2010) Genotype to phenotype: a complex problem. Science 328:469.
Fisher R (1918) The correlation between relatives on the supposition of Mendelian inheritance. Phil Trans R Soc Edinburgh 52: 399-433.
Flint J, Valdar W, Shifman S, Mott R (2005) Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet 6: 271-286.
Furlotte NA, Kang EY, Van Nas A, Farber CR, Lusis AJ, et al. (2012) Increasing association mapping power and resolution in mouse genetic studies through the use of meta-analysis for structured populations. Genetics. 191: 959-967.
Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, 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.
Gora-Maslak G. McClearn GE, Crabbe JC. Phillips TJ, Belknp JK, et al. (1992) Use of recombinant inbred strains to identify quantitative trait loci in psychopharmacology. Psychopharmacol. 104: 413-424.
Graham JB, Thomas S, Swarts J, McMillan AA, Ferris MT, et al. (2015) Genetic diversity in the collaborative cross model recapitulates human West Nile virus disease outcomes. MBio 6: e00493-15.
Gregorova S, Divina P, Storchova R, Trachtulec Z, Fotopulosova V, et al. (2008) Mouse consomic strains: Exploiting genetic divergence between Mus m. musculus and Mus m. domesticus subspecies. Genome Res 18: 509-515.
Hegmann JP, Possidente B (1981) Estimating genetic correlations from inbred strains. Behav Genet 11:103-114.
Heiker JT, Kunath A, Kosacka J, Flehmig G, Knigge A, et al. (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.
Hitzemann B, Dains K, Kanes S, Hitzemann R (1994) Further studies on the relationship between dopamine cell density and haloperidol-induced catalepsy. J Pharmacol Exp Ther 271: 969-976.
Hitzemann R, Malmanger B, Reed C, Lawler M, Hitzemann B, Coulombe S, Buck K, Rademacher B, Walter N, Polyakov Y, Sikela J, Gensler B, Burgers S, Williams RW, Manly K, Flint J, Talbot C (2003) A strategy for the integration of QTL, gene expression, and sequence analyses. Mamm Genome 14: 733-747.
Holmes RS, Petersen DR, Deitrich RA (1986) Biochemical genetic variants in mice selectively bred for sensitivity or resistance to ethanol-induced sedation. Anim Genet 17: 235-44.
Houtkooper RH, Mouchiroud L, Ryu D, Moullan N, Katsyuba E, et al. (2013) Mitonuclear protein imbalance as a conserved longevity mechanism. Nature 497: 451-457.
Hrbek T, de Brito RA, Wang B, Pletscher LS, Cheverud JM (2006) Genetic characterization of a new set of recombinant inbred lines (LGXSM) formed from the inter-cross of SM/J and LG/J inbred mouse strains. Mamm Genome 17: 417-429.
Hsieh SM, Look MP, Sieuwerts AM, Foekens JA, Hunter KW (2009) Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. Breast Cancer Res. 11: R75.

Hu Y, Wu G, Rusch M, Lukes L, Buetow KH, et al. (2012) Integrated cross-species transcriptional network analysis of metastatic susceptibility. Proc Natl Acad Sci USA 109: 3184-3189.

Hunter KW, Crawford NP (2008) The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. Annu Rev Genet 42: 131-141.

Kakihana R, Brown DR, McClearn GE, Tabershaw IR (1966) Brain sensitivity to alcohol in inbred mouse strains. Science. 154: 1574-1575.

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477: 289-294.

Khisti RT, Wolstenholme J, Shelton KL, Miles MF (2006) Characterization of the ethanol-deprivation effect in substrains of C57BL/6 mice. Alcohol 40: 119-126.

Kirkpatrick SL, Bryant CD (2015) Behavioral architecture of opioid reward and aversion in C57BL/6 substrains. Front Behav Neurosci 8: 450.

Koutnikova H, Markku L, Lu L, Combe R, Paananen J, et al. (2009) Identification of UBP1 as a critical blood pressure determinant. PLoS Genetics 5: e1000591.

Kumar V, Kim K, Joseph C, Kourrich S, Yoo SH, et al. (2013) C57BL/6N mutation in Cytoplasmic FMRP interacting protein 2 (*Cyfip2*) regulates cocaine response. Science 342: 1508-1512.

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Lenarcic AB, Svenson KL, Churchill GA, Valdar W (2012) A general Bayesian approach to analyzing diallel crosses of inbred strains. Genetics 190: 413-435.

Li Z, Mulligan MK, Wang X, Miles MF, Lu L, et al. (2010) A transposon in *Comt* generates mRNA variants and causes widespread expression and behavioral differences among mice. PLoS One 5: e12181.

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear mixed models for genome-wide association studies. Nat Methods 8: 833-835.

Lynch CJ (1969) The so-called Swiss mouse. Lab Anim Care 19: 214-220.

Mackay TF (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35: 303–339.

Malmanger B, Lawler M, Coulombe S, Murray R, Cooper S, Polyakov Y, Belknap J, Hitzemann R (2006) Further studies on using multiple-cross mapping (MCM) to map quantitative trait loci. Mamm Genome 17: 1193-1204.

Miller RA, Harrison DE, Astle CM, Floyd RA, Flurkey K, et al. (2007) An Aging Interventions Testing Program: Study design and interim report. Aging Cell 6: 565-575.

Morahan G (2012) Insights into type 1 diabetes provided by genetic analyses. Curr Opin Endocrinol Diabetes Obes 19: 263-270.

Morgan AP, Fu CP, Kao CY, Welsh CE, Didion JP, et al. (2015) The Mouse Universal Genotyping array: From substrains to subspecies. G3 (Bethesda) 6: 263-279.

Mozhui K, Wang X, Chen J, Mulligan MK, Li Z, et al. (2011) Genetic regulation of *Nrxn1* expression: An integrative cross-species analysis of schizophrenia candidate genes. Transl Psychiatry 1:e25.

Mulligan MK, Ponomarev I, Boehm SL 2nd, Owen JA, Levin PS, et al. (2008) Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. Genes Brain Behav 7: 677-689.

Mulligan MK, Mozhui K, Prins P, Williams RW (2016) GeneNetwork – A toolbox for systems genetics. Meth Molecular Biology, in press (this volume).

Overall RW, Kempermann G, Peirce J, Lu L, Goldowitz D, et al. (2009) Genetics of the hippocampal transcriptome in mouse: a systematic survey and online neurogenomics resource. Front Neurosci 3: 55.

Pandey A, Williams RW (2014) Genetics of gene expression in the CNS. Int Rev Neurobiol 116: 195-231.

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.

Percival CJ, Liberton DK, Pardo-Manuel de Villena F, Spritz R, Marcucio R, et al. (2016) Genetics of murine craniofacial morphology: diallel analysis of the eight founders of the Collaborative Cross. J Anat. 228: 96-112.

Petkov PM, Cassell MA, Sargent EE, Donnelly CJ, Robinson P, et al. (2004) Development of a SNP

genotyping panel for genetic monitoring of the laboratory mouse. Genomics 83: 902-911.
Rasmussen AL, Okumura A, Ferris MT, Green R, Feldmann F, et al. (2014) Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. Science 346: 987–991.
Rat Genome Sequencing and Mapping Consortium, Baud A, Hermsen R, Guryev V, Stridh P et al. (2013) Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. Nat Genet 45: 767-775.
Roberts A, Pardo-Manuel de Villena F, Wang W, McMillan L, Threadgill DW (2007) The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. Mamm Genome 18: 473–481.
Rönnegård L, Valdar W (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. Genetics 188: 435-447.
Rosen GD, Azoulay NG, Griffin EG, Newbury A, Koganti L, et al. (2013) Bilateral subcortical heterotopia with partial callosal agenesis in a mouse mutant. Cereb Cortex 23: 859-872.
Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297-302.
Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, et al. (2005) An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 37:710-717.
Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J (2006) A high-resolution single nucleotide polymorphism genetic map of the mouse genome. PLoS Biol 4: e395
Simon MM, Greenaway S, White JK, Fuchs H, Gailus-Durner V, et al. (2013) A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. Genome Biol 14: R82.
Singer JB, Hill AE, Burrage LC, Olszens KR, Song J, et al. (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. Science 304: 445-448.
Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, et al. (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507: 371-375.
Subramanian S, Tus K, Li QZ, Wang A, Tian XH, et al. (2006) A Tlr7 translocation accelerates systemic autoimmunity in murine lupus. Proc Natl Acad Sci USA 103: 9970-9975.
Svenson KL Gatti DM, Valdar W, Welsh CE, Cheng R, et al (2012) High-resolution genetic mapping using the mouse diversity outbred population. Genetics 190: 437-447.
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 21:305-318.
Taylor BA, Heiniger HJ, Meier H (1973) Genetic analysis of resistance to cadmium-induced testicular damage in mice. Proc Soc Exp Biol Med 143: 629-633.
Threadgill DW, Hunter KW, Williams RW (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. Mamm Genome 13: 175-178.
Valdar W, Soberg LC, Gaugiuier D. Burnett S., Klenerman P et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet 38: 879-887.
Wahlsten D, Metten P, Crabbe JC (2003) A rating scale for wildness and ease of handling laboratory mice: Results for 21 inbred strains tested in two laboratories. Genes, Brain, and Behavior 2: 71-79.
Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, et al. (2016) Joint mouse-human phenome-wide association to test gene function and disease risk. Nature Commun 7: 10464.
Williams EG, Auwerx J (2015) The convergence of systems and reductionist approaches in complex trait analysis. Cell 162: 23-32.
Williams EG, Mouchiroud L, Frochaux M, Pandey A, Andreux PA, et al (2014) PLoS Genet 10: e1004673.
Williams RW (1999) A targeted screen to detect recessive mutations that have quantitative effects. Mamm Genome 10: 734-738.
Williams RW, Gu J, Qi S, Lu L (2001) The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. Genome Biol 2(11):RESEARCH0046.
Williams RW, Strom RC, 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 RW, Strom RC, Rice DS, Goldowitz D (1996) Genetic and environmental control of variation in retinal ganglion cell number in mice. J Neurosci 16: 7193-7205.

Wright S (1921) Correlation and causation. J Agri Res 20: 557–585.

Wu Y, Williams EG, Dubuis S, Mottis A, Jovaisaite V, et al. (2014) Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. Cell 158: 1415-1430.

Yalcin B, Flint J, Mott R (2005) Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics 171: 673-681.

Yalcin B, Nicod J, Bhomra A, Davidson S, Cleak J, et al. (2010) Commercially available outbred mice for genome-wide association studies. PLoS Genet 6: e1001085.

Yang H, Bell TA, Churchill GAs, Pardo-Manuel de Villena F (2007) On the subspecific origin of the laboratory mouse. Nat Genet 39: 1100-1107.

Yang H, Crawford N, Lukes L, Finney R, Lancaster M, et al. (2005) Metastasis predictive signature profiles pre-exist in normal tissues. Clin Exp Metastasis 22: 593–603.

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