HHS Public Access

Author manuscript

Trends Genet. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

Trends Genet. 2019 July ; 35(7): 501–514. doi:10.1016/j.tig.2019.04.003.

High-diversity mouse populations for complex traits

Michael C. Saul[1], **Vivek M. Philip**[1], **Laura G. Reinholdt**[1], **Center for Systems Neurogenetics of Addiction**[1,2,3,4], and **Elissa J. Chesler**[1,*]

[1]The Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME

[2]UNC Chapel Hill, Chapel Hill, NC

[3]SUNY Binghamton, Binghamton, NY

[4]Pittsburgh University, Pittsburgh, PA

Abstract

Contemporary mouse genetic reference populations are a powerful platform to discover complex disease mechanisms. Advanced high-diversity mouse populations include the Collaborative Cross strains, Diversity Outbred stock, and their isogenic founder strains. When used in systems genetics and integrative genomics analyses, these populations efficiently harnesses known genetic variation for precise and contextualized identification of complex disease mechanisms. Extensive genetic, genomic, and phenotypic data are already available for these high diversity mouse populations, and a growing suite of data analysis tools have been developed to support research on diverse mice. This integrated resource can discover and evaluate disease mechanisms relevant across species.

Keywords

systems genetics; mouse; complex genetics; genetic diversity; complex traits

The Challenge of Complex Disease

Complex diseases present compelling biomedical challenges that can be studied using human and non-human animal genetics. Though advances in human genetics have identified loci for many heritable complex diseases [1–8], there are several well-known limitations of genome-wide association studies (GWAS). 1) For many human disease loci, the biological mechanism of action is unknown. 2) When little is known about a locus, the path from genetic association to clinically actionable targets is unclear. 3) Disease process or developmental trajectory is not always obvious from a causal genetic variant. 4) GWAS results may not generalize across human sub-populations. 5) Medical records and participant phenotyping is often incomplete, imprecise, and retrospective whereas model organism

*correspondence: elissa.chesler@jax.org (Elissa J. Chesler).

phenotyping can include in-depth, prospective and standardized measures. 6) Sample size requirements are formidable in human GWAS. 7) Power is insufficient to study interacting genetic loci. 8) Studies of genetic interaction with development, environment, and sex are largely intractable. Further, heterogeneous diseases like psychiatric disorders manifest with overlapping symptoms, intertwined disease trajectories [9], and complex genetic regulation [9,10]. In summary, the SNP-to-disease association model underlying GWAS cannot readily capture complex disease biology without further biological context. These challenges are often tractable with discovery genetics in model organisms.

Human genetic studies are essential to identify causal loci for human disease. However, there is growing recognition that these studies merely point to etiology and mechanisms rather than therapeutic interventions. Further, these findings are often specific to the population under investigation, so precise causal information is of limited general utility. Model organisms simplify discovery of molecular networks associated with disease phenotypes in temporal and anatomical context. Thus, when a human variant is found, non-human animal studies may connect it to a relevant biological process – and often to a druggable target. When the goal is to elucidate biological mechanisms of disease for diagnosis and intervention, consilience between human and non-human animal genetics provides an efficient and inexpensive approach to biological discovery and clinical translation [11]. While consilience is rare at the variant level, it often manifests at the level of orthologous genes, pathways, and molecular networks. The mouse remains the predominant resource for disease related research [12–15]. Recent advances in sequencing, computation, and gene editing that have advanced human genetics have also greatly augmented the capabilities of mouse genetics, resulting in new tools and resources for mechanistic discovery and translation.

Capturing and using genetic variation in laboratory mice

Much biomedical research has utilized C57BL/6 mice and **outbred stocks** such as Hsd:ICR, and Swiss Webster mice. **Inbred mouse strains** such as C57BL/6 promise experimental rigor and reproducibility [16]. Individuals of the same sex from within 10 generations of properly maintained **isogenic** stocks have essentially identical genomes. Modern colony maintenance further reduces accumulation of spontaneous mutations [17,18]. Standardized genomes allow direct comparison of individuals from isogenic stocks across laboratories separated by years and continents.

As a consequence of genome standardization, inbred mice are widely used in biomedical science. Unfortunately, standardization impedes many important research applications. Generalizability is limited with results gleaned from a single strain. Further, individuals from inbred strains exhibit variability within and across laboratories, attributable to limited diversity in stabilizing mechanisms in response to environmental variation [19].

Naturally occurring genetic diversity of mice provides allelic variation for biological discovery within a well-characterized, easily manipulated experimental system. Findings from genetically diverse populations are more likely to generalize across species or strains. Early experimental crosses and other populations derived from inbred mouse strains

harnessed the diversity of laboratory mice [20] and have been used to detect complex trait mechanisms in behavioral science, immunology, and physiology [21]. Using **quantitative trait locus (QTL)** mapping and **genetic correlation** in mouse populations, researchers have detected variants in genes orthologous to candidates from human GWAS [22]. However, early QTL mapping studies reported broad regions containing many genes and variants that were difficult to prioritize due to limited contemporaneous genetic and genomic resources. One advance was the development of **systems genetics** approaches, which integrate QTL mapping with genome-wide molecular phenotyping to discover, characterize, and contextualize molecular and phenotypic network variation across biological scale. These studies often used **genetic reference populations** such as the BXD recombinant inbred strains to integrate data across experiments.

An integrative resource for complex disease biology

Over the past two decades, community-based efforts to improve mouse resources gave rise to three new advanced mouse populations. The Complex Trait Consortium (CTC), a group dedicated to producing new genetics tools to study complex traits in diverse populations, proposed the first of these resources in the early 2000s [23]. The CTC sought to improve mapping precision and systems genetics power. The resulting mouse populations – the Collaborative Cross (CC), Diversity Outbred (DO), and their founder inbred strains – provide complimentary genetic reagents for the study of complex traits. They comprise an integrated resource harboring a substantial pool of shared genetic variation. That variation is randomized throughout the genome in the CC and DO populations.

The eight founder strains capture approximately 90% of the genetic diversity seen in the *Mus musculus* species [24]. These strains are A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HlLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ (Figure 1A). Five are common laboratory strains descendent of fancy mice of the *Mus musculus domesticus* subspecies. The remaining strains are wild-derived inbred representatives of three *Mus musculus* subspecies: the European *domesticus*, the north Asian *musculus*, and the south Asian *castaneus.* Wild mice display traits that were bred out of domesticated mice [21,25] while providing genetic diversity intrinsic to the mouse species [24].

The DO and CC genetic reference population are recombinant populations systematically derived from the founders. Randomized breeding design greatly reduces population structure effects that limit mapping resolution in collections of extant inbred strains. The CTC intended to breed a panel of 1,000 recombinant inbred strains [26–28] to obtain a reproducible mapping population. The CTC worked to provide the highest diversity resource achievable. However, there were some challenges. When the founders were chosen, only limited genotypic data were available. While the committee that chose the founder strains made efforts to maximize genetic variation, additional consideration was given to disease susceptibility, particularly in cancer and diabetes research. The CTC simulated many designs to introduce high diversity with randomization. In the absence of genotypic information, an eight founder design was attractive, but a simpler four founder design would have had a higher 25% minor allele frequency for strain-specific variants and could have simplified breeding logistics. Furthermore, such a design allows all possible breeding configurations to

be performed, whereas only a fraction of the some 80,000 possible matings could be sampled in the CC, requiring extensive efforts to randomize and balance mitochondrial and sex chromosome composition [29].

Three international breeding sites produced hundreds of incipient CC funnels [29–31] and a wide variety of insights into mammalian genetics [32–34], but inbreeding depression, infertility, and other factors led to a high rate of attrition [35]. Approximately 50 finished CC strains are publicly available today, capturing a representative cross-section of founder haplotypes [36] and providing adequate power for genetic correlation studies. Their genetic equidistance makes them a suitable replacement for the non-uniformly related extant inbred strain collection that comprise the original Mouse Phenome Panel [37] and Hybrid Mouse Diversity Panel [38].

The Diversity Outbred was initiated as an ultra-high precision mapping population [39]. Intercrossing the incipient CC lines results in recombination to reduce **linkage disequilibrium (LD)** blocks to their biological limits. The result is a **heterogeneous stock**, whose genetics derive from an equal contribution of each of the eight founder strains in a random configuration. This population is maintained primarily through pseudo-random matings, though interventions have been performed to maintain variation across the genome [40]. Historically, heterogeneous stock populations including the HS/Ibg and HS/Npt were developed as selection base populations, but work in the late 1990s described their use as a resource for fine mapping of complex traits [41,42], leading to a large mapping study of many complex traits [43]. This extremely fine mapping can also be performed in other **advanced intercross lines** [44].

With an estimated 45 million segregating polymorphisms, the CC and DO populations have more genetic diversity than observed across the human population [45]. This may lead to complex patterns of trait regulation at the population level, manifest in high sample size requirements likely attributable to epistatic interactions that stabilize trait variation, but has the benefit of providing detectable genetic variation in every gene and pathway, facilitating discovery of biological mechanisms for virtually any complex phenotype.

Using genetically diverse mice

The advanced diversity populations comprise an integrated resource to discover and explore biological mechanisms underlying complex disease-related traits. Many experimental applications are possible. Four modes of complex trait discovery are widely used: 1) establishing trait heritability, 2) evaluating relationships among traits to test for shared mechanisms, 3) identifying model strains for focused mechanistic studies, and 4) mapping the genetic basis of trait variation.

Each mouse population is best-suited to particular applications (Figure 2, Key Figure), but their shared set of polymorphisms provide extensive opportunities for data integration Molecular, genetic, genomic, and phenotypic resources are available: fully sequenced genomes from the Sanger Mouse Genomes project can be accessed at the Mouse Genome Database (MGD), and trait data are deposited in the Mouse Phenome Database (MPD) [46].

Analytical tools tailored to common applications of multi-parent populations are available (Box 1). Because of their common genetics, studies performed in these mice can be integrated with existing molecular, genetic, and disease-relevant work (Figure 3).

Establishing trait heritability in the founders

Establishing heritability of a trait and its assays (Figure 3A) demonstrates feasibility of genetic dissection. This information aids experimental design and implementation. **Strain surveys** of the eight founders are well-suited for assay optimization to provide robust, informative, and generalizable parameters. Though heritability can be established in any population by estimating trait variation accounted for by kinship, founder strain surveys provide straightforward, reproducible heritability estimates.

Heritability has been established for a number of behavioral, physiological and molecular traits in founders. Heritabilities are determined by sources of variance and are greatest for morphological traits and lower for behavioral traits [47]. Examples include psychotropic drug response [48,49], infectious disease response [50], and transcript splice variation [51] across the founders. Heritability of reward-related behavioral traits that have proven difficult to observe in conventional mouse strains were established in founder strains [52]. However, it should be noted that certain founders may have characteristics precluding specific procedures, whereas their outcrossed progeny and derivative populations may not. Therefore, trait variation may be studied in complementary CC and DO populations even when measurements are not obtainable in all founder strains.

Some founders, CC strains, and DO individuals possess more wildness behavior than conventional laboratory strains [21,32]. As previously noted, behavioral characteristics were most likely selected out of common laboratory strains, leading to several regions of identity by descent [53]. This expanded behavioral repertoire is advantagous for the study of certain behavioral traits, but can lead to issues in handling, testing, and husbandry [25]. DO males are more aggressive than most laboratory mice and are often singly housed. Certain widely used behavioral assays, including elevated mazes are not suitable. However, the high exploratory behavior also leads to rapid acquisition of drug self-administration [54] and other interesting trait variation.

Characterizing trait correlations in the Collaborative Cross

Understanding how traits covary allows the identification of common underlying mechanisms of closely related traits (Figure 3B). **Trait correlation** is widely used in recombinant inbred strains; nearly 40 years of data are available for BXD strains [55,56]. As a highly diverse, stable, and reproducible reference population, the Collaborative Cross was designed for this application. Phenotypic studies performed in a panel of CC animals can be compared across experiments and laboratories. Such experiments can be extended using **recombinant inbred cross (RIX)** strains, the F1 progeny of systematically intercrossed pairs of inbred strains to deterministically generate testable genomes. CC RIX mice are an interesting subset of reproducible genomes: more iterations are possible (over 1,000 for approximately 50 CC strains) and the progeny are heterozygous at most loci. The

correlations among complex traits observed in the CC, DO and BXD strains recapitulate relations observed in the human population, for example the relationship of novelty and sensation seeking with drug intake and addiction related behaviors [57,58].

With the public availability of finished CC strains, we expect more widespread characterization and utilization in trait correlation studies deposited in the MPD over time [37]. Additional examples of trait correlation in the CC are emerging, and include studies of host-microbiome interactions and complex traits [59]. Further, replicable CC genomes are ideal for studying genome-by-environment interactions. For example, CC strains have been useful for elucidating host genetic effects on pathogen and infectious disease response [60–66]. Multiple studies are presently working to perform large-scale phenotyping within the CC with results available in the MPD.

Identifying complex disease model strains in the Collaborative Cross

Often, Collaborative Cross studies identify strains with disease-related characteristics (Figure 3B). In contrast to much maligned historic “disease models”for pre-clinical testing that feature a single gene perturbation on a single background, these strains possess multiple disease associated variants in many pathways, better reflecting disease heterogeneity in the population. **Extreme strains**, which sometimes express a phenotype more strongly than the founder strains, can be used to test interventions. Trait correlations among the CC can identify **multivariate outlier strains** with traits within the normal range on two disease relevant traits that don not exhibit the expected trait correlation, for example, a paradoxical increase in fat deposition in response to exercise [67]. Complex disease models may also be identified through genotypes at disease relevant loci by extrapolation of human genetic loci to CC mice through expression QTL to identify strains with cumulative high and low risk variants. For example, CC strains with extreme sperm motility phenotypes were identified as male infertility models [35]. An additional benefit of the common genetic background shared by these mouse populations is that CC mice can be used to validate predictive genetic results from the DO.

Finding regulatory mechanisms for trait variation in Diversity Outbred mice

For complex traits exhibiting continuous distributions, QTL mapping is used to identify loci driving phenotypic variation. The Diversity Outbred population is far more precise than historic mapping populations because of its dense recombinations, high genetic diversity, unlimited sample size, and genetic randomization (Figure 2C). DO mice are heterozygous at most loci and the minimum minor allele frequency for an allele private to a single founder strain is expected to be 12.5%, so high mapping resolution can be achieved with reasonable sample sizes. QTL mapped in the DO are precise (Figure 3C) and can sometimes be resolved to individual variants.

DO mice have been successfully used for mapping multiple complex traits including: cardiovascular phenotypes [68,69], metabolic syndrome related traits [70], environmental toxicity [71], cancer modifier traits [72], behavioral traits [73,74], and meiotic drive [40]. Early large-scale studies in incipient CC strains successfully mapped many traits [32], and

CC lines were used to map motor performance and body weight [75], energy balance traits [76], exercise physiology [67], toxicology [77], perinatal nutrition in CC RIX lines [78], kidney phenotypes [79], and hematologtical phenotypes [80]. Although CC mice were intended as a genetic mapping population and were used as such in these early studies, the power of the extant strains is only sufficient to large effects alleles, typically observed in studies of Mendelian traits [81]. They are therefore useful for reproducible mapping of molecular phenotypes such as cis-eQTL and epigenetic regulation, which typically exhibit Mendelian genetic variation.

Genetic variation in molecular mechanisms

Transcriptomic analysis contextualizes and resolves mapping results. Expression QTL mapping (eQTL) studies treat transcript abundance as a phenotype. Significant cis-eQTL – genes whose expression is significantly associated with a proximal genotype – often coincide with trait QTL because genetic variants that influence expression frequently influence trait variation. Significant trans-eQTL – genes encoded by loci distal to the QTL – may also regulate complex traits. Variants affecting the structure and function of gene products are also sources of trait variation. Network analysis can be used to map a coexpression module to a QTL (mQTL), allowing interrogation of the genetic regulatory landscape for entire networks of genes. eQTL can connect non-coding regulatory variants to genes, providing a bridge between distal regulatory elements and their targets. Other mapping applications employing quantitative molecular phenotypes include protein QTL (pQTL), chromatin accessibility QTL (caQTL or dsQTL), histone modification QTL (hQTL), and many others [82].

Other applications of diversity mice

For general biomedical applications that typically use a single isogenic strain or outbred stock, DO mice may have significant advantages. Results from studies of genetically diverse mice are likely robust and generalizable to a diverse population rather than idiosyncratic to a single inbred genome [19]. Unlike isogenic mice, each Diversity Outbred animal's genome is unique. Therefore, reproducibility of Diversity Outbred studies occurs at the level of replicate samples. With addition of mice to these studies, one may identify the genetic basis of individual differences.

Selective breeding projects produce novel, polygenic, and reproducible models of disease through artificial selection. The CC and related CC-HS population been used as a selection stock in past experiments [83–86]. The DO population is recommended for new selection experiments. Its great capacity for selection arises from high heterozygosity and high diversity, lowering the expense of acquiring profound selectable variation needed for a selection stock.

Experimental design considerations

Trait and environmental variability, heritability, effect size, and allele frequencies of the causal variant can affect power to detect QTL. Power simulations over a range of

experimental parameters [87], demonstrate that QTL explaining >20% of variance can be detected with 90% power with as few as 200 DO mice. With 1,000 DO mice, QTL can be detected explaining 5% of phenotypic variance with 90% power. Behavioral trait mapping in populations of 300 DO have detected some significant QTL [73,74]. Though noise increases sample size requirements for mapping, gene expression QTL are typically more robust and can be mapped with 400 DO mice or less [69]. Unlike less complex mouse populations, which may segregate fewer trait regulatory variants, an increased sample size will yield increased saturation of genetic effects.

For the CC, mapping power is best obtained by preferentially sampling as many strains as possible before subsampling within strains. Only large effect QTL accounting for >50% variance can be mapped with single samples from 50 strains at 80% estimated power. Increased subsampling within strains may enable mapping of moderate effects explaining >20% variance [81]. These effect sizes are typically only seen in highly penetrant Mendelian traits or molecular traits.

The population structure of the DO necessitates specialized mapping models on microarray-derived genotyping data (Box 1). The underlying regression model can accommodate covariates such as sex, cohort, treatment and additional systematic sources of variance such as experimenter, particularly important for behavioral studies. When coupled to information about the biological effects of SNP variation, this approach can lead to causal variant identification [88,89].

Resources for discovery and validation

Once QTL and trait-associated molecular mechanisms are found, validation resources are used to confirm the molecular and trait-level effects. Companion resources including tissue and cell line biobanks are being generated for these populations. One emerging resource, a set of stem cell lines including mouse embryonic stem cells, (mESCs) and induced pluripotent stem cells (iPSCs) derived from these populations [90,91], enables *in vitro* systems genetics and molecular phenotyping QTL validation.

DO mESCs have been used to genetically dissect pluripotent ground state metastability eQTL and caQTL. Patient-derived pluripotent stem cell lines exhibit phenotypic variability, which affects differentiation and impedes universal protocol development to produce clinically relevant cell types. Patient donor genetic variation is a primary driver of inter-line variability [92–94], and caQTL and eQTL have been mapped in large genetically diverse panels of these differentiated human iPSCs [95]. However, low genetic resolution in these small sample size human studies limits the functional validation of variants underlying QTL [94,96]. In contrast, modestly-sized DO mESC panels, combined with robust analytical approaches, offer the genetic resolution and importantly, validation capabilities needed to demonstrate causality.

Validation of QTL discovered in genetically diverse mice is powered by the ease of gene editing technologies like CRISPR/Cas9. Genetic polymorphisms within guide/donor sequences potentially confounds editing work in diverse mice; sequences are typically

designed using the mouse reference genome (C57BL/6J). Fortunately, whole genome assemblies available through the MGP [97] facilitate nucleotide design.

In some cases, validation may be achieved *in vitro*. For example, variants driving molecular QTL like eQTL and caQTL, can be validated using CRISPR/Cas9 engineering followed by molecular readouts easily measured in cultured cells. The stem cell lines available from the founder strains [90] are particularly useful for validating molecular QTL in both undifferentiated cells and *in vitro* differentiated cultures, which is relatively affordable and fast compared to *in vivo* work. Promising *in vitro* lines can be used to create engineered mice through traditional ES cell microinjection approaches, allowing for further *in vivo* validation.

Validation makes use of existing and emerging tools for highly-controlled manipulation of engineered loci. These include constitutive, tissue-specific, or inducible Cre driver systems as well as an expanded CRISPR repertoire exploiting nuclease-deficient dCas9 and various types of chromatin effector proteins that control gene expression without editing [98]. Finally, CC and CC-RIX mice themselves can validate QTL discovered in DO populations against a range of background variability. In this application, CC or CC-RIX lines are selected for their genotype within a QTL region. Phenotype predictions based on these genotypes can be easily tested *in vivo*. For follow-on studies of extreme phenotypes in genetically distinct DO mice, specialized approaches can be employed. Derivation of iPSCs using non-integrating reprogramming approaches offers a method for propagating unique DO genomes *in vitro* for validation studies [99,100]. Further production of 100% iPSC derived chimeras from DO iPSCs is also possible, albiet with low throughput and requirements for sophisticated embryo manipulation.

Concluding remarks

Arguments against the use of mice in discovery genetics often reflect an incomplete understanding of progress in mouse genetic resources, mouse-human orthology relationships, and comparative genomics experimental work [101,102]. Many mouse genetic studies have anticipated genetic findings from human genetics studies using GWAS or low-throughput rare variant sequencing studies on candidate genes, e.g. *MPDZ* in alcohol related traits [103,104] and *OPRM1* in addiction [105,106]. Global matching of human and mouse regulatory variants demonstrates that mouse genetics identifies orthologs comparable to the genes identified by human GWAS with greater power and efficiency [22]. Integrating phenotypes over a genetic reference population with human genetics has been highly successful [107]. By integrating regulatory variation – including non-coding variants through orthologous targets across species – genetic variants detected in the mouse may feasibly be translated to human genetic implications. Tools such as GeneWeaver [108], the Monarch Initiative [109], and novel non-coding variant prediction algorithms [110] facilitate multi-species translation.

Mouse systems genetics links molecular measures with genetic loci, providing mechanistic context for human genetic variation. Integration of human GWAS data with mouse complex trait genetics is critical to contextualize and interpret the implications of trait variation (see

Outstanding Questions). This approach was successfully used for, e.g., cocaine related traits *FAM53B* in mice and humans [57]. However, early populations generated low- precision results that often required years for resolution. The Collaborative Cross, Diversity Outbred, and founder populations provide an efficient, integrated platform for systems genetics of disease-relevant complex traits. These diversity mice can dissect complex traits at multiple levels by establishing heritability, identifying co-regulation with other traits, finding interesting model strains, and making mechanistic genetic insights. These mouse populations and data analysis tools advance the mouse as a versatile platform for discovery of the biological mechanisms of disease.

Acknowledgements

This work was supported by NIH grants P50 DA039841 and R01 DA037927 to Elissa J. Chesler and the Center for Systems Neurogenetics of Addiction.

Glossary

Advanced Intercross Line
A biparental population of mice where inbreeding is disallowed, large proportions of recombinations have occurred, and most alleles are heterozygous

Extreme Strain
A strain on the extreme ends of a phenotype measure. These strains are often useful for in-depth characterization using low-throughput methods

Genetic Reference Population
A panel of inbred animals showing strong and predictable genetic variability between strains and an unlimited capacity for reproducing the same genomes in different animals

Heterogeneous Stock
A population derived from randomized matings of more than two founder strains, resulting in mosaic genomes with high recombination

Inbred Mouse Strain
A mouse strain derived from at least 20 generations of sibling-sibling matings, which effectively eliminates heterozygosity and results in isogenic offspring

Isogenic
Having the same genotype across the genome; genetically identical

Linkage Disequilibrium
Association of alleles at multiple loci with respect to one another usually caused by close physical proximity on the same chromosome, or coinheritance of alleles across chromosomes in admixed populations

Outbred Mouse Stock
a mouse stock derived from a genetically diverse source that is maintained through matings between unrelated individuals

Multivariate Outlier Strain
Strains outside of an expected range given a trait correlation. These strains are useful for dissociating the biological basis for interrelated traits

Quantitative Trait Locus Mapping
The statistical technique used to associate a complex quantitative trait with the genetic factors governing the trait

Recombinant Inbred Cross (RIX)
The F1 generation resulting from an outcrossing with a recombinant inbred strain

Recombinant Inbred Strain
A strain of mice that underwent inbreeding such that its autosome haplotypes are a mosaic of two or more founder strains

Strain Survey
A study of trait variation across isogenic strains. This technique takes advantage of isogenic strains by assuming a perfect kinship between members of the same strain. Heritability may be simply calculated from ANOVA statistics

Systems Genetics
A complex trait discovery genetics technique coupling genome-scale molecular measures with genetics to identify and contextualize the networked biological mechanisms leading from genotype to trait

Trait Correlation
Quantitative identification of similarities between phenotypes measured in the same panel of mice. Because inbred panels replicate genomes, such studies can be done with data from different animals from the same strains and from different laboratories

References

1. Geschwind DH and Flint J (2015) Genetics and genomics of psychiatric disease. Science 349, 1489–1494. [PubMed: 26404826]
2. Okada Y et al. (2012) Meta-analysis identifies multiple loci associated with kidney function–related trait25friceast asian populations. Nature Genetics 44, 904–909. [PubMed: 22797727]
3. Ehret GB et al. (2016) The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. Nature Genetics 48, 1171–1184. [PubMed: 27618452]
4. Baranov VS et al. (2015) Systems genetics view of endometriosis: A common complex disorder. European Journal of Obstetrics & Gynecology and Reproductive Biology 185, 59–65. [PubMed: 25528731]
5. Locke AE et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197–206. [PubMed: 25673413]
6. Raj T et al. (2017) Genetic architecture of age-related cognitive d25fricanin african americans. Neurology Genetics 3, e125. [PubMed: 28078323]
7. Marioni RE et al. (2018) GWAS on family history of Alzheimer's disease. Translational Psychiatry 8, 99. [PubMed: 29777097]
8. Karaderi T et al. (2015) Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of obesity-related traits. Current diabetes reports 15, 83. [PubMed: 26363598]

9. Arnedo J et al. (2015) Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. American Journal of Psychiatry 172, 139–153. [PubMed: 25219520]
10. Zuk O et al. (2014) Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences 111, E455–E464.
11. Nadeau JH and Auwerx J (2019) The virtuous cycle of human genetics and mouse models in drug discovery. Nature Reviews Drug Discovery In Press.
12. Perrin S (2014) Preclinical research: Make mouse studies work. Nature 507, 423–425. [PubMed: 24678540]
13. Gould SE et al. (2015) Translational value of mouse models in oncology drug development. Nature Medicine 21, 431–439.
14. Justice MJ and Dhillon P (2016) Using the mouse to model human disease: Increasing validity and reproducibility. Disease Models & Mechanisms 9, 101–103. [PubMed: 26839397]
15. Sukoff Rizzo SJ and Crawley JN (2017) Behavioral phenotyping assays for genetic mouse models of neurodevelopmental, neurodegenerative, and psychiatric disorders. Annual Review of Animal Biosciences 5, 371–389. [PubMed: 28199172]
16. Festing MF (2014) Evidence should trump intuition by preferring inbred strains to outbred stocks in preclinical research. ILAR Journal 55, 399–404. [PubMed: 25541542]
17. Taft RA et al. (2006) Know thy mouse. Trends in Genetics 22, 649–653. [PubMed: 17007958]
18. Sarsani VK et al. (2019) The genome of C57BL/6J “Eve”, the mother of the laboratory mouse genome reference strain. bioRxiv
19. Tuttle AH et al. (2018) Comparing phenotypic variation between inbred and outbred mice. Nature Methods 15, 994–996. [PubMed: 30504873]
20. Biggers JD and Claringbold P (1954) Why use inbred lines? Nature 174, 596–597. [PubMed: 13203582]
21. Chesler EJ (2014) Out of the bottleneck: The Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. Mammalian Genome 25, 3–11. [PubMed: 24272351]
22. Attie AD et al. (2017) How mice are indispensable for understanding obesity and diabetes genetics. Current Opinion in Endocrinology, Diabetes, and Obesity 24, 83–91.
23. Threadgill DW et al. (2002) Genetic dissection of complex and quantitative traits: From fantasy to reality via a community effort. Mammalian Genome 13, 175–178. [PubMed: 11956758]
24. Roberts A et al. (2007) The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: Implications for QTL discovery and systems genetics. Mammalian Genome 18, 473–481. [PubMed: 17674098]
25. Wahlsten D et al. (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.
26. Williams RW and Churchill GA The Collaborative Cross: Rationale, implementation, and costs. March-(2004), A presentation at the NIH
27. Vogel G (2003) Scientists dream of 1001 complex mice. Science 301, 456–457. [PubMed: 12881545]
28. Churchill GA et al. (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genetics 36, 1133–1137. [PubMed: 15514660]
29. Chesler EJ et al. (2008) The Collaborative Cross at Oak Ridge National Laboratory: Developing a powerful resource for systems genetics. Mammalian Genome 19, 382–389. [PubMed: 18716833]
30. Morahan G et al. (2008) Establishment of “The Gene Mine”: A resource for rapid identification of complex trait genes. Mammalian Genome 19, 390–393. [PubMed: 18716834]
31. Iraqi FA et al. (2008) The Collaborative Cross, developing a resource for mammalian systems genetics: A status report of the Wellcome Trust cohort. Mammalian Genome 19, 379–381. [PubMed: 18521666]
32. Philip VM et al. (2011) Genetic analysis in the Collaborative Cross breeding population. Genome Research 21, 1223–1238. [PubMed: 21734011]
33. Aylor DL et al. (2011) Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Research 21, 1213–1222. [PubMed: 21406540]

34. Durrant C et al. (2011) Collaborative Cross mice and their power to map host susceptibility to Aspergillus fumigatus infection. Genome Research 21, 1239–1248. [PubMed: 21493779]
35. Shorter JR et al. (2017) Male infertility is responsible for nearly half of the extinction observed in the mouse Collaborative Cross. Genetics 206, 557–572. [PubMed: 28592496]
36. Srivastava A et al. (2017) Genomes of the mouse Collaborative Cross. Genetics 206, 537–556. [PubMed: 28592495]
37. Paigen K and Eppig JT (2000) A mouse phenome project. Mammalian Genome 11, 715–717. [PubMed: 10967127]
38. Ghazalpour A et al. (2012) Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mammalian Genome 23, 680–692. [PubMed: 22892838]
39. Svenson KL et al. (2012) High-resolution genetic mapping using the mouse Diversity Outbred population. Genetics 190, 437–447. [PubMed: 22345611]
40. Chesler EJ et al. (2016) Diversity Outbred mice at 21: Maintaining allelic variation in the face of selection. G3: Genes, Genomes, Genetics 6, 3893–3902. [PubMed: 27694113]
41. Mott RM et al. (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences 97, 12649–12654.
42. Talbott CJ et al. (1999) High-resolution mapping of quantitative trait loci in outbred mice. Nature Genetics 21, 305–308. [PubMed: 10080185]
43. Valdar W et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nature Genetics 38, 879–887, [PubMed: 16832355]
44. Darvasi A and Soller M (1995) Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping. Genetics 141, 1199–1207. [PubMed: 8582624]
45. Keane TM et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294. [PubMed: 21921910]
46. Grubb SC et al. (2013) Mouse phenome database. Nucleic Acids Research 42, D825–D834. [PubMed: 24243846]
47. Visscher PM et al. (2008) Heritability in the genomics era — concepts and misconceptions. Nature Reviews Genetics 9, 255–266.
48. Crowley JJ et al. (2012) Antipsychotic-induced vacuous chewing movements and extrapyramidal side effects are highly heritable in mice. The Pharmacogenomics Journal 12, 147–155. [PubMed: 21079646]
49. Morgan AP et al. (2014) The antipsychotic olanzapine interacts with the gut microbiome to cause weight gain in mouse. PLoS One 9, e115225. [PubMed: 25506936]
50. Leist SR et al. (2016) Influenza H3N2 infection of the Collaborative Cross founder strains reveals highly divergent host responses and identifies a unique phenotype in CAST/EiJ mice. BMC Genomics 17, 143. [PubMed: 26921172]
51. Zheng CL et al. (2015) Splicing landscape of the eight Collaborative Cross founder strains. BMC Genomics 16, 52. [PubMed: 25652416]
52. Dickson PE et al. (2015) Sex and strain influence attribution of incentive salience to reward cues in mice. Behavioural Brain Research 292, 305–315. [PubMed: 26102561]
53. Yang H et al. (2011) Subspecific origin and haplotype diversity in the laboratory mouse. Nature Genetics 43, 648–655. [PubMed: 21623374]
54. Dickson PE et al. (2015) Association of novelty- related behaviors and intravenous cocaine self-administration in Diversity Outbred mice. Psychopharmacology 232: 1011–1024. [PubMed: 25238945]
55. Bubier JA and Chesler EJ (2012) Accelerating discovery for complex neurological and behavioral disorders through systems genetics and integrative genomics in the laboratory mouse. Neurotherapeutics 9, 338–348. [PubMed: 22422471]
56. Taylor B et al. (1973) Genetic analysis of resistance to cadmium-induced testicular damage in mice. Proceedings of the Society for Experimental Biology and Medicine 143, 629–633. [PubMed: 4719448]
57. Dickson PE et al. (2016) Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. Psychopharmacology 233, 701–714. [PubMed: 26581503]

58. Sanchez-Roige S et al. (2019) Genome-wide association studies of impulsive personality traits (BIS-11 and UPPS-P) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. Journal of Neuroscience 39, 2562–2572. [PubMed: 30718321]
59. Bubier J et al. (2018) Systems genetic discovery of host-microbiome interactions reveals mechanisms of microbial involvement in disease. bioRxiv
60. Ferris MT et al. (2013) Modeling host genetic regulation of influenza pathogenesis in the Collaborative Cross. PLoS Pathogens 9, e1003196. [PubMed: 23468633]
61. Lorè NI et al. (2015) Host genetic diversity influences the severity of Pseudomonas aeruginosa pneumonia in the Collaborative Cross mice. BMC Genetics 16, 106. [PubMed: 26310945]
62. Vered K et al. (2014) Susceptibility to Klebsiella pneumonaie infection in Collaborative Cross mice is a complex trait controlled by at least three loci acting at different time points. BMC Genomics 15, 865. [PubMed: 25283706]
63. Graham JB et al. (2015) Genetic diversity in the Collaborative Cross model recapitulates human West Nile virus disease outcomes. mBio 6, e00493–15. [PubMed: 25944860]
64. Gralinski LE et al. (2015) Genome wide identification of SARS-CoV susceptibility loci using the Collaborative Cross. PLoS Genetics 11, e1005504. [PubMed: 26452100]
65. Green R et al. (2017) Oas1b-dependent immune transcriptional profiles of West Nile virus infection in the Collaborative Cross. G3: Genes, Genomes, Genetics 7, 1665–1682. [PubMed: 28592649]
66. Green R et al. (2016) Transcriptional profiles of wnv neurovirulence in a genetically diverse Collaborative Cross population. Genomics data 10, 137–140. [PubMed: 27872814]
67. McMullan RC et al. (2018) CC002/Unc females are mouse models of exercise-induced paradoxical fat response. Physiological Reports 6, e13716. [PubMed: 29924460]
68. Shorter JR et al. (2018) Quantitative trait mapping in Diversity Outbred mice identifies two genomic regions associated with heart size. Mammalian Genome 29, 80–89. [PubMed: 29279960]
69. Smallwood TL et al. (2014) High-resolution genetic mapping in the Diversity Outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. G3: Genes, Genomes, Genetics 4, 2353–2363. [PubMed: 25344410]
70. Tyler AL et al. (2017) Epistatic networks jointly influence phenotypes related to metabolic disease and gene expression in Diversity Outbred mice. Genetics 206, 621–639. [PubMed: 28592500]
71. French JE et al. (2015) Diversity Outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. Environmental Health Perspectives 123, 237–245. [PubMed: 25376053]
72. Winter JM et al. (2017) Mapping complex traits in a Diversity Outbred F1 mouse population identifies germline modifiers of metastasis in human prostate cancer. Cell Systems 4, 31–45. [PubMed: 27916600]
73. Logan RW et al. (2013) High-precision genetic mapping of behavioral traits in the Diversity Outbred mouse population. Genes, Brain and Behavior 12, 424–437.
74. Recla JM et al. (2014) Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene. Mammalian Genome 25, 211–222. [PubMed: 24700285]
75. Mao J-H et al. (2015) Identification of genetic factors that modify motor performance and body weight using Collaborative Cross mice. Scientific Reports 5, 16247. [PubMed: 26548763]
76. Mathes WF et al. (2011) Architecture of energy balance traits in emerging lines of the Collaborative Cross. American Journal of Physiology-Endocrinology and Metabolism 300, E1124–E1134. [PubMed: 21427413]
77. Venkatratnam A et al. (2017) Collaborative Cross mouse population enables refinements to characterization of the variability in toxicokinetics of trichloroethylene and provides genetic evidence for the role of PPAR pathway in its oxidative metabolism. Toxicological Sciences 158, 48–62. [PubMed: 28369613]
78. Schoenrock S et al. (2017) Perinatal nutrition interacts with genetic background to alter behavior in a parent-of-origin-dependent manner in adult Collaborative Cross mice. Genes, Brain and Behavior

79. Thaisz J et al. (2012) Genetic analysis of albuminuria in Collaborative Cross and multiple mouse intercross populations. Renal Physiology 303, F972–F981. [PubMed: 22859403]
80. Kelada SN et al. (2012) Genetic analysis of hematological parameters in incipient lines of the Collaborative Cross. G3: Genes, Genomes, Genetics 2, 157–165. [PubMed: 22384394]
81. Keele GR et al. (2018) Determinants of QTL mapping power in the realized Collaborative Cross. bioRxiv 459966.
82. Vandiedonck C (2018) Genetic association of molecular traits: A help to identify causative variants in complex diseases. Clinical Genetics 93, 520–532. [PubMed: 29194587]
83. Iancu OD et al. (2010) Genetic diversity and striatal gene networks: Focus on the heterogeneous stock-collaborative cross (HS-CC) mouse. BMC Genomics 11, 585. [PubMed: 20959017]
84. Colville AM et al. (2018) Regional Differences and Similarities in the Brain Transcriptome for Mice Selected for Ethanol Preference From HS-CC Founders. Frontiers in Genetics 28, 300.
85. Colville AM et al. (2017) Effects of selection for ethanol preference on gene expression in the nucleus accumbens of HS-CC mice. Genes, Brain and Behavior 16, 462–471.
86. Zombeck JA et al. (2011) Selective breeding for increased home cage physical activity in Collaborative Cross and Hsd:ICR mice. Behavior Genetics 41, 571–582. [PubMed: 21184167]
87. Gatti DM et al. (2014) Quantitative trait locus mapping methods for Diversity Outbred mice. G3: Genes, Genomes, Genetics 4, 1623–1633. [PubMed: 25237114]
88. French JE et al. (2015) Diversity Outbred Mice Identify Population-Based Exposure Thresholds and Genetic Factors that Influence Benzene-Induced Genotoxicity. Environmental Health Perspectives 123, 237–245. [PubMed: 25376053]
89. Recla JM et al. (2019) Genetic mapping in Diversity Outbred mice identifies a Trpa1 variant influencing late-phase formalin response. Pain In press.
90. Czechanski A et al. (2014) Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. Nature Protocols 9, 559–574. [PubMed: 24504480]
91. Garbutt TA et al. (2018) Permissiveness to form pluripotent stem cells may be an evolutionarily derived characteristic in Mus musculus. Scientific Reports 8, 14706. [PubMed: 30279419]
92. DeBoever C et al. (2017) Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells. Cell Stem Cell 20, 533–546. [PubMed: 28388430]
93. Féraud O et al. (2016) Donor dependent variations in hematopoietic differentiation among embryonic and induced pluripotent stem cell lines. PloS One 11, e0149291. [PubMed: 26938212]
94. Schwartzentruber J et al. (2018) Molecular and functional variation in iPSC-derived sensory neurons. Nature Genetics 50, 54–61. [PubMed: 29229984]
95. Skelly DA et al. (2019) Genetic variation influences pluripotent ground state stability in mouse embryonic stem cells through a hierarchy of molecular phenotypes. bioRxiv 552059.
96. Alasoo K et al. (2018) Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nature Genetics 50, 424–431. [PubMed: 29379200]
97. Lilue J et al. (2018) Sixteen diverse laboratory mouse reference genomes define strain- specific haplotypes and novel functional loci. Nature Genetics 50, 1574–1583. [PubMed: 30275530]
98. Lu X and Li Q A CRISPR-dCas Toolbox for Genetic Engineering and Synthetic Biology. Journal of Molecular Biology 431, 34–47. [PubMed: 29958882]
99. Yoshioka N et al. (2013) Efficient generation of human iPSCs by a synthetic self- replicative RNA. Cell Stem Cell 13, 246–254. [PubMed: 23910086]
100. Lau S et al. (2014) Direct neural conversion from human fibroblasts using self- regulating and nonintegrating viral vectors. Cell Reports 9, 1673–1680. [PubMed: 25482564]
101. Shay T et al. (2015) Genomic responses to inflammation in mouse models mimic humans: We concur, apples to oranges comparisons won't do. Proceedings of the National Academy of Sciences 112, E346–E346.
102. Takao K and Miyakawa T (2015) Genomic responses in mouse models greatly mimic human inflammatory diseases. Proceedings of the National Academy of Sciences 112, 1167–1172.

103. Karpyak VM et al. (2009) Sequence variations of the human MPDZ gene and association with alcoholism in subjects with European ancestry. Alcoholism Clinical & Experimental Research 33, 712–721.
104. Fehr C et al. (2002) Congenic mapping of alcohol and pentobarbital withdrawal liability loci to a <1 centimorgan interval of murine chromosome 4: identification of Mpdz as a candidate gene. Journal of Neuroscience 22, 3730–3738. [PubMed: 11978849]
105. Smith AH et al. (2017) Genome-wide association study of therapeutic opioid dosing identifies a novel locus upstream of OPRM1. Molecular Psychiatry 22, 346–352. [PubMed: 28115739]
106. Berratini WH et al. (1994) Quantitative trait loci mapping of three loci controlling morphine preference using inbred mouse strains. Nature Genetics 7, 54–58. [PubMed: 8075641]
107. Wang X et al. (2016) Joint mouse–human phenome-wide association to test gene function and disease risk. Nature Communications 7, 10464.
108. Bubier JA et al. (2015) GeneWeaver: Finding consilience in heterogeneous cross- species functional genomics data. Mammalian Genome 26, 556–566. [PubMed: 26092690]
109. Mungall CJ et al. (2017) The Monarch Initiative: An integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Research 45, D712–D722. [PubMed: 27899636]
110. Zhou J and Troyanskaya OG (2015) Predicting effects of noncoding variants with deep learning–based sequence model. Nature Methods 12, 931–934. [PubMed: 26301843]
111. Morgan AP and Welsh CE (2015) Informatics resources for the Collaborative Cross and related mouse populations. Mammalian Genome 26, 521–539. [PubMed: 26135136]
112. Broman KW et al. (2019) R/qtl2: Software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics 211, 495–502. [PubMed: 30591514]
113. Darlington RB and Smulders TV (2001) Problems with residual analysis. Animal Behaviour 62, 599–602.
114. Churchill G and Doerge R (2008) Naive application of permutation testing leads to inflated type I error rates. Genetics 178, 609–610. [PubMed: 18202402]
115. Cheng R and Palmer AA (2013) A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness. Genetics 193, 1015–1018. [PubMed: 23267053]
116. Sen Ś and Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159, 371–387. [PubMed: 11560912]

Box 1: Software and analysis tools

A growing analytical toolkit facilitates work with advanced mouse populations. A recent review discussed these resources in great depth [111]. Most of the software packages available are free and open-source.

An advantage of isogenic founder strains is their known and reproducible underlying genotypes. The sequenced reference mouse strain is one founder, the C57BL/6J mouse strain. The Mouse Genomes Project led by the Sanger Institute has sequenced and assembled the genomes of the other seven founder strains. Their website includes useful tools to query genetic variation among the founder strains, allowing researchers to quickly identify variants likely to drive biological mechanisms (available at https://www.sanger.ac.uk/science/data/mouse-genomes-project). Visualization facilities for genome features are available using the MGI Multiple Genome Viewer (available at http://www.informatics.jax.org/mgv). The genomes of many extant CC mice have been both genotyped and sequenced; their genotypes were released on the UNC Systems Genetics website (available at http://csbio.unc.edu/CCstatus/CCGenomes).

For DO animals, individual genotyping is required for any genetic mapping studies. At present, the most recent Mouse Universal Genotyping Array (GigaMUGA) platform is the predominant whole genome genotyping resources used for DO mice. Genotyping using these arrays can be purchased as a service from the Neogen Corporation (available at http://www.neogen.com/genomics). HaploQA is a web-based software for interpretation of MUGA-derived microarrays (available at http://haploqa.jax.org). Genotyping by RNA sequencing (GBRS) is a potentially promising avenue to high-resolution genotyping that also produces relevant molecular phenotyping data for systems genetics work. The software package GBRS is presently under development for use with advanced mouse populations (available at https://gbrs.readthedocs.io/en/latest) and can be used for alignment and transcript abundance estimation as well as genotyping.

Sequencing experiments with these populations typically involve reconstruction of each founder genomes and annotation, lifting over length variations introduced by short insertions and deletions. The g2g tools software package performs these tasks (available at https://github.com/churchill-lab/g2gtools). Allele-specific expression patterns for RNA-seq can be elucidated using EMASE (available at https://emase.readthedocs.io/en/latest).

Typically, a statistical model used for QTL mapping in Diversity Outbred will model phenotype as a function of genotype and a number of other factors [39]. DO mice display 36 genotype states (8 homozygous and 28 heterozygous diplotypes) compared with 2 or 3 in traditional mapping crosses. To infer these states, a hidden markov model (HMM) on genotype array data represents predicted diplotypes as probabilistic estimates of each diplotype state [112]. A typical QTL mapping model regresses the trait on estimated diplotype probabilities for each of the 8 founder alleles, a reduced form of the more complex 36 state diplotype model that assumes heterozygotes are intermediates of homozygous states for any given trait. In standard modeling notation, this model is:

$$y_i = \beta_s s_i + \beta_g g_i + \sum_{j=1}^{8} \beta_j g_{ij} + \lambda_i + \epsilon_i$$

where y_i is the ith animal's phenotype, $\beta_s s_i$ is the effect of sex for the ith animal, $\beta_g g_i$ is the effect of an additive grouping covariate for the ith animal, $\beta_j g_{ij}$ is the effect of founder allele probability for allele j in the ith animal, λ_i is the polygenic random effect of the ith animal, and ϵ_i is the error term. Statistical software was developed to operationalize this model for R including: includes R/DOQTL (https://bioconductor.org/packages/release/bioc/html/DOQTL.html) and qtl2 (https://kbroman.org/qtl2).

Though this model does not capture dominance, it has been effective in mapping complex traits. A critical component of the regression model, a random effects kinship matrix derived from pairwise diplotype probabilities, captures relatedness among subjects to increase power. Traits can also be adjusted by regression against nuisance factors, using the residuals for mapping, although this approach may suppress genetic signal [113]. These models produce LOD scores for genotypes across the genome, and genome-wide statistical significance is tested via permutation excluding the kinship matrix (kinship violates the assumption of exchangeability) [114,115]. A 95% Bayesian credible interval describes positional confidence intervals [116] often as narrow as two megabases, underscoring the precision of mapping in the DO. Within a QTL, a two-state SNP association model improves precision.

An advantage to using diversity mice is the reusability of data resources when the researcher deposits them in a database. For DO, CC, and founders datasets, the Mouse Phenome Database (MPD, https://phenome.jax.org/) includes measures readily useable for initial exploration, trait correlation, and discovery genetics. A repository of DO QTL studies is maintained at the Jackson Laboratory (https://dodb.jax.org). For various two-parent mouse crosses such as the BXD, GeneNetwork has produced an impressive suite of resources to deposit, analyze, and integrate systems genetics datasets (http://genenetwork.org).

To integrate, compare, and contrast mouse and human data, GeneWeaver facilitates integration of heterogenous genome-scale datasets collected across multiple species (https://geneweaver.org). Knowledge-guided exploration of discoveries from gene sets generated in multiple species can be performed in KnowEnG (https://knoweng.org). Cross- species integration applications such as the Monarch Initiative [109] aim to bridge genomic and phenotypic annotations between commonly used species using orthology information.

Lastly, technologies change quickly. The Jackson Laboratory Genetic Diversity Initiative provides up-to-date information on these mouse populations (available at https://www.jax.org/research-and-faculty/genetic-diversity-initiative).

Outstanding Questions

- What are the most productive paths for utilization of mouse and human integrative genetic analyses to speed the translation of mechanistic insights?
- How do we make use of the mechanistic context provided by systems genetics to discover and validate clinically relevant targets?
- What model organism data will facilitate the interpretation of the contribution of non- coding variants to human disease?
- How do we best use genetic and genomic data to establish phenotype consilience across species?

Highlights

- High-diversity mouse populations with known and reproducible genetic variation make complex trait genetics tractable in a mammalian system.
- Together, these populations are a valuable integrated and scalable tool for discovery genetics in complex trait studies.
- The Collaborative Cross (CC), its founders, and the heterozygous CC-RIX derived from crosses of the CC strains are a fully-reproducible population for exact genome-matched correlational and controlled studies.
- The Diversity Outbred (DO) population displays high genetic and phenotypic variability and enables precise genetic mapping.
- Cross-species genomic analysis of mouse derived results allows comparative and translational applications.

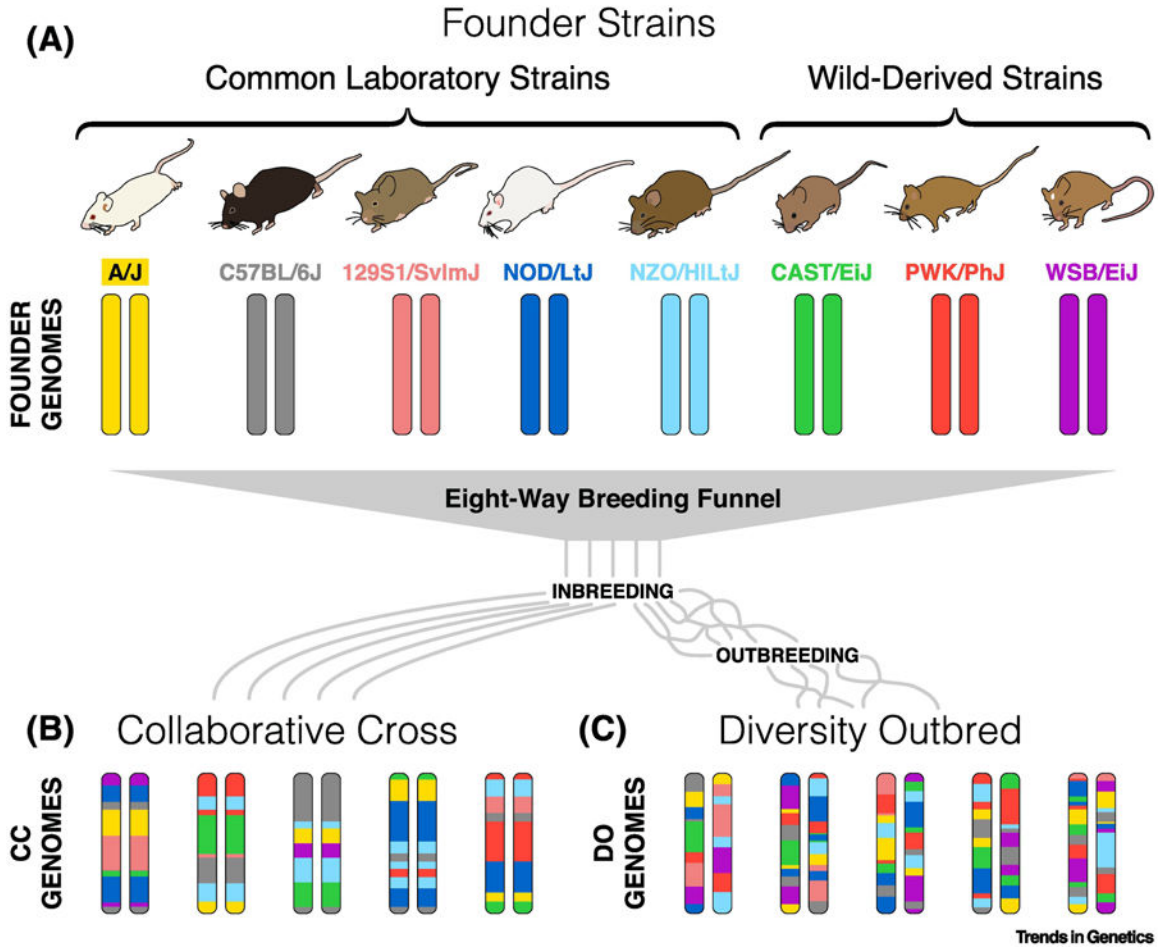

Figure 1.

Summary of the genome structures of the primary advanced mouse strains. The common origin of the genomic variation contained within these mice allows for their use as an integrated set of tools to investigate the genetic basis of complex traits. **A)** Founder strains include five common laboratory inbred strains and three wild-derived inbred strains. Together, these strains recapitulate about 90% of the genetic variation observed in Mus musculus and represent genotypic variation comparable to human populations. **B)** Collaborative Cross strains are a panel of eight-way recombinant inbred strains derived from the founders. Approximately 50 Collaborative Cross strains are presently under distribution. **C)** Diversity Outbred mice were derived from sustained outbreeding of the founder genotypes, resulting in continuous variation in genome structure and high heterozygosity while retaining variants useful for mapping.

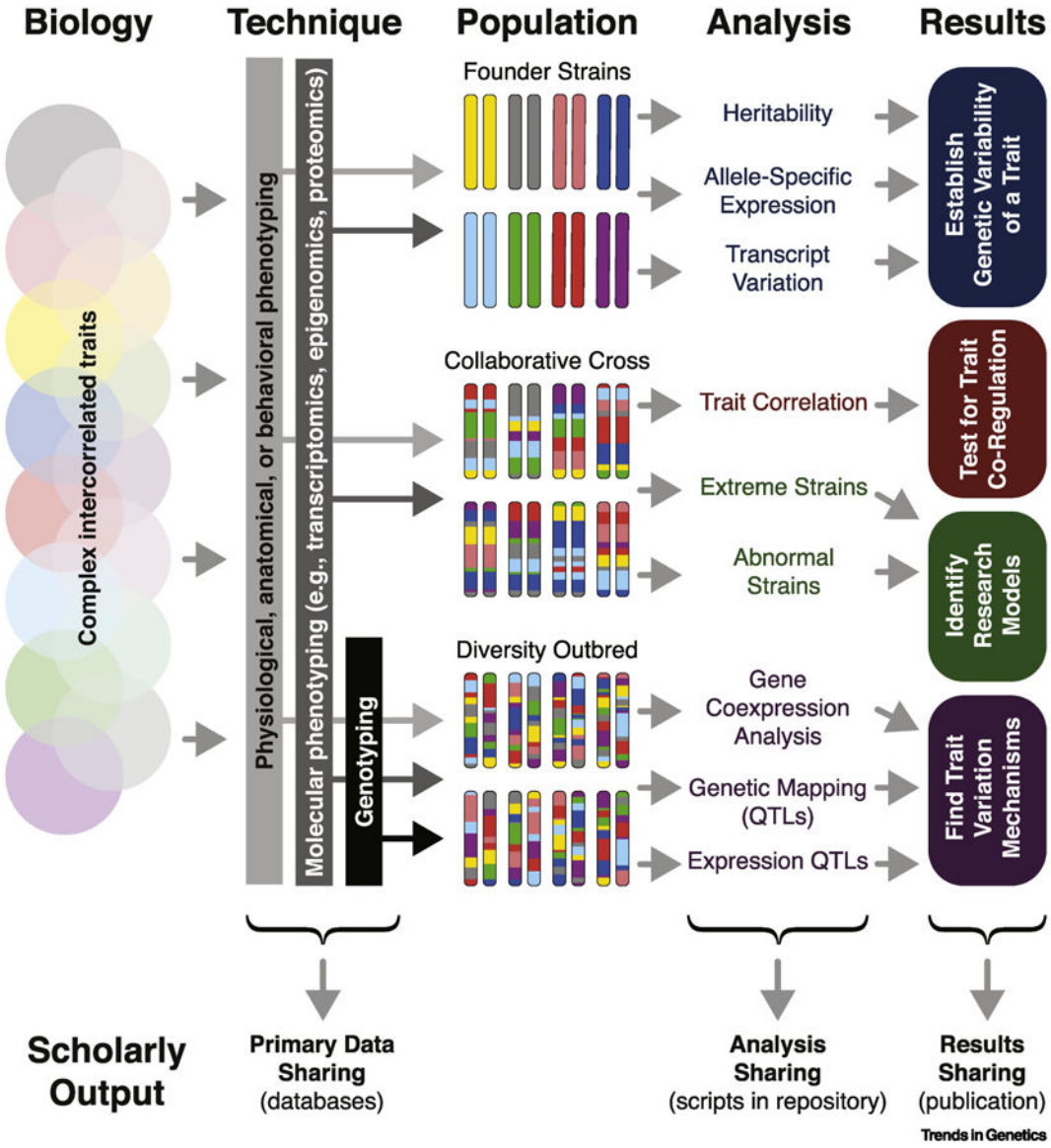

Figure 2, Key Figure.

Typical pipelines for discovery using diversity mice from biological question to results. Diversity mice can contribute through multiple integrated applications to research on complex traits. The selection of the ideal mouse population is dependent upon the research question being asked. Complex traits can be established as heritable, then dissected into multiple phenotypic and genotypic outputs. Furthermore, extreme and multivariate outlier strains allow for establishment of research models that can correlate and dissociate important aspects of biology.

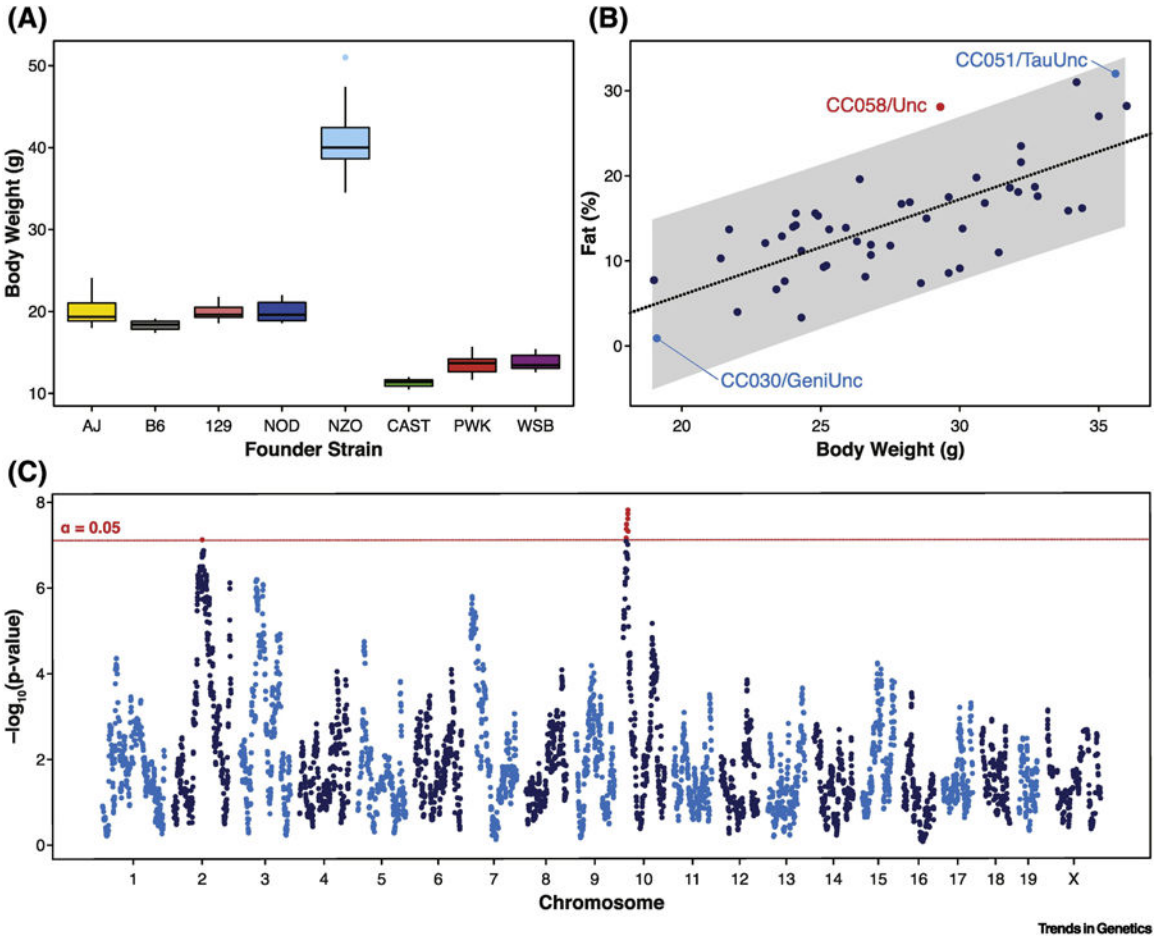

Figure 3.

Multiple diversity mouse resources can be used in separate experiments to dissect a single biological question at multiple levels. In this case, multiple published experiments deposited in the Mouse Phenome Database include information about body weight as a complex trait. **A)** Strain surveys on the founder strains demonstrate heritable variation of, e.g., body weight (MPD: Morgan1). **B)** Trait correlation in the Collaborative Cross strains demonstrate a biologically significant link between traits (MPD: McMullan1). In this example, there is a high correlation between body weight and percentage body fat. This method identifies extreme strains (blue) and a multivariate outlier strain (red) that may be models for future study (gray: 95% prediction interval). **C)** QTL mapping in the Diversity Outbred population (MPD: Recla1). For body weight, significant QTL were identified on chromosomes 2 and 10. Combined with expression data, significant findings can be further resolved to the gene level and contextualized as elements of gene coexpression networks.