CrossMark

Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database

Molly A. Bogue[1] · Gary A. Churchill[1] · Elissa J. Chesler[1]

Abstract The Mouse Phenome Database was originally conceived as a platform for the integration of phenotype data collected on a defined collection of 40 inbred mouse strains—the "phenome panel." This model provided an impetus for community data sharing, and integration was readily achieved through the reproducible genotypes of the phenome panel strains. Advances in the development of mouse populations lead to an expanded role of the Mouse Phenome Database to encompass new strain panels and inbred strain crosses. The recent introduction of the Collaborative Cross and Diversity Outbred mice, which share an extensive pool of genetic variation from eight founder inbred strains, presents new opportunities and challenges for community data resources. A wide variety of molecular and clinical phenotypes are being collected across genotypes, tissues, ages, environmental exposures, interventions, and treatments. The Mouse Phenome Database provides a framework for retrieval, integration, analysis, and display of these data, enabling them to be evaluated in the context of existing data from standard inbred strains. Primary data in the Mouse Phenome Database are supported by extensive metadata on protocols and procedures. These are centrally curated to ensure accuracy and reproducibility and to provide data in consistent formats. The Mouse Phenome Database represents an established and growing community data resource for mouse phenotype data and encourages submissions from new mouse resources, enabling investigators to integrate existing data
into their studies of the phenotypic consequences of genetic variation.

Introduction

Understanding the causes of variation in complex disease-related phenotypes, how these traits relate to one another, and which phenotypic outcomes most resemble human disease requires detailed characterization and integration of data across phenotyping domains. Deep phenotyping of model organisms is a powerful approach to basic and translational research of human disease. The laboratory mouse is an especially efficient and versatile model system. Mice have a relatively short lifespan, 99 % of mouse genes are shared with humans (Boguski 2002), and a rich repertoire of phenotyping modalities is available to study the physiology, behavior, and genetics of mice in normal development, aging, and disease. Mice present many advantages over direct study of diseases in humans including precise control of experimental conditions, low costs, access to tissues and interventions, and repeatability of experimentation. As a result, the laboratory mouse remains the most widely studied and most well-characterized model organism.

Thousands of inbred and genetically modified strains are currently available as live animals or as cryopreserved stocks, and more are being created and phenotyped (Ringwald et al. 2011; Brown and Moore 2012). Crosses of inbred mouse strains have revealed the genetic basis of numerous complex traits through quantitative trait locus (QTL) mapping (for numerous examples, see QTL Archive at phenome.jax.org). New genetic reference populations have been developed including the Collaborative Cross (CC) inbred strains (Churchill et al. 2004; Chesler et al.

✉ Molly A. Bogue
  molly.bogue@jax.org

[1] The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

2008; Iraqi et al. 2008; Morahan et al. 2008; Welsh et al. 2012; Threadgill and Churchill 2012) and their complementary high-precision mapping population, the Diversity Outbred (DO) mice (Churchill et al. 2012; Svenson et al. 2012a; Chesler 2013). The genomes of many widely used mouse strains, including the founders of the CC and DO, have been fully sequenced (Keane et al. 2011; Yalcin et al. 2012; Wong et al. 2012; Ananda et al. 2014), and high-density genotyping arrays are available (Yang et al. 2009).

Unlike the human population, genetic variation present in the mouse has been stabilized, characterized, and segregated (both randomly and non-randomly) across a variety of different populations. The shared genetic variation in mouse populations provides a basis for data integration and a means to discover the causal genetic variants for disease-related phenotypes. Dense genotyping and sequencing technologies enable characterization of genomic similarity of individual mice and strains. Our ability to relate this complete picture of genetic variation to phenotypic observations of individual mice enables identification and validation of the genetic basis of complex, disease-related traits, increasingly with single nucleotide resolution.

Collecting data on widely used populations provides significant opportunities to extend findings through data reuse but only if the data are harmonized and integrated. Dissemination of primary mouse phenotype data is an imperative complement to research publications; however, additional effort—beyond releasing data in supplemental files—is needed to facilitate coherent integrative analyses across multiple studies.

Primary data access is crucial for three reasons: (1) integrative analysis to find consensus among diverse studies, (2) reanalysis in light of new developments, and (3) reproducibility. Unfortunately, phenotypic data often exist in diverse and sometimes non-computable stores with insufficient documentation and restricted access. With increasing potential for data integration to provide new insights, it is critical that we provide access to carefully curated data in standardized formats that allow researchers to build upon previous studies. Recent advances in meta-analysis techniques have demonstrated the value of combining primary data across multiple mouse studies (Kang et al. 2014; Bubier et al. 2014).

The Mouse Phenome Database

The Mouse Phenome Database (MPD; http://phenome.jax.org) (Grubb et al. 2014) stores harmonized primary data, including per mouse phenotypes acquired over multiple trials, measures, or conditions. Now in its 14th year, MPD collects, annotates, and disseminates quantitative phenotype data and protocols in an integrated relational database

to facilitate faceted search and other capabilities. MPD provides a repository of mouse phenotype data and a suite of tools for comparative and quantitative analysis. Originally developed as a repository for phenotype data collected on a small and defined set of inbred strains (Paigen and Eppig 2000), the scope and role of MPD as a general repository for primary data collected on individual mice has expanded. MPD data are organized around a catalog of phenotype ontology terms; assays and protocols are extensively documented; analysis tools provide summary statistics and data visualization; and, importantly, the common data framework enables easy data access to and integration of data from multiple labs and experiments. Data come from investigators around the world and represent a broad scope of physiological and behavioral traits in naïve mice and those exposed to drugs, environmental agents, or other treatments. Access to phenotype data and protocols from different sources enables researchers to reproduce experiments; reanalyzes data with new algorithms and up-to-date bioinformatics resources; and discovers unexpected relationships among trait data that were collected in different times and places. The high standards of documentation and curation and the stability of the program at The Jackson Laboratory have made MPD a primary resource for investigators to archive and retrieve quantitative mouse phenotypic data and protocols.

Collaborative Cross strains, Diversity Outbred mice, and related populations

Genetic reference populations have defined genomes, can be retrieved or reproduced indefinitely, and enable data integration through genetic correlation across genome-matched individuals or strains. The Collaborative Cross, a multi-parent recombinant inbred strain panel was created to provide greater allelic diversity and better randomization of genotypes and genetic recombinations than was previously available in genetic reference populations and thereby to improve the power, precision, and accuracy of genetic mapping and phenotype correlation analysis (Threadgill et al. 2002; Churchill et al. 2004; Chesler et al. 2008; Iraqi et al. 2008; Morahan et al. 2008). The eight founder strains of the CC include three wild-derived strains to maximize diversity (Fig. 1). This population structure reflects the design characteristics for CC strains (CCC 2012). Dense genotyping and haplotype reconstructions of the CC inbred strains are freely available (for review, see Morgan and Welsh 2015, this issue) and, together with the founder genome sequences, represent a comprehensive knowledge of these stable, reproducible genomes. There are currently about 150 extant CC strains. Each strain can provide genetically identical mice to support as many different

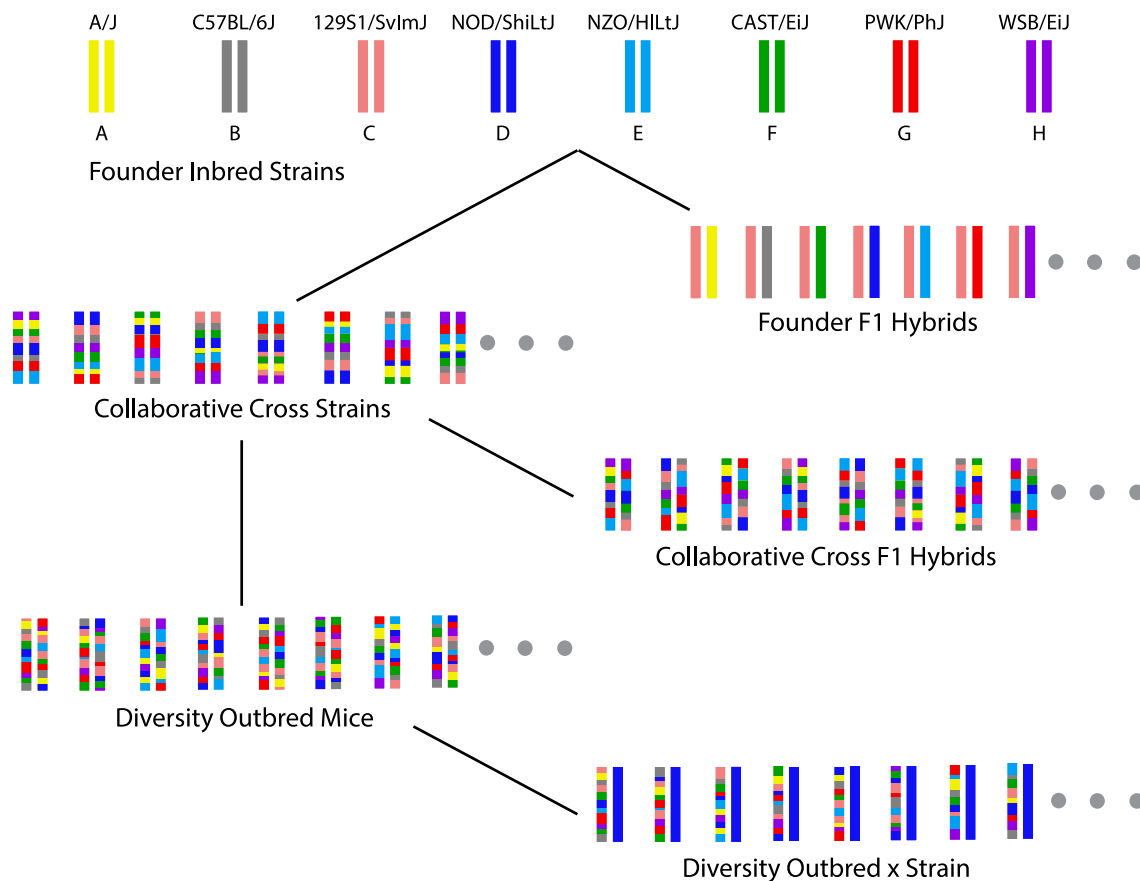## Collaborative Cross and Diversity Outbred Mouse Resources

Fig. 1 Collaborative Cross (CC) and Diversity Outbred (DO) mouse resources. The eight founder strains are depicted at the *top* as color-coded chromosomes. Color-coding for these strains has been agreed upon by the research community and standardized. Founder F1 hybrid offspring are produced from crossing two founder strains and are thus reproducible. Theoretically, there are 102 possible F1 hybrids (including reciprocal crosses). CC strains are inbred (homozygous at most loci) and reproducible. CC F1 hybrids are offspring produced from crossing two CC strains and are thus reproducible. There are many possibilities for CC F1 hybrids. DO mice were derived from CC strains through strict breeding protocols to maintain heterogeneous stocks and are thus not reproducible. Theoretically, there is an unlimited supply of DO mice. They are typically tested as a population of over 100 unique mice. "DO × strain" mice are produced from a DO mouse crossed to an inbred strain. Because DO mice are not reproducible, "DO × strain" mice are likewise not reproducible

studies as resources allow. Phenotypes collected on CC mice are directly cumulative, facilitating the comparison of new and historical data. An impressive range of phenotypic diversity has been reported recently for traits including susceptibility to melanoma (Ferguson et al. 2015), allergic airway responsiveness (Kelada et al. 2014), immune cell counts (Phillippi et al. 2014), and susceptibilities to H1N1 (Ferris et al. 2013), Ebola (Rasmussen et al. 2014), West Nile Virus (Graham et al. 2015), and *Klebsiella* (Vered et al. 2014). The inbred CC strains are limited in mapping resolution and power by the number of extant strains, and do not enable the detection of non-additive effects due to the lack of heterozygosity.

Diversity Outbred (DO) mice are derived from incipient Collaborative Cross strains and represent a complementary resource with the same allelic diversity as the CC strains, maintained in a heterogeneous stock (Churchill et al. 2012; Chesler 2013). This population is an ideal resource for genetic mapping and selective breeding studies. The high fecundity and large stock population enables sampling of extremely large numbers of unique genomes derived from the same eight founders. Over successive generations, the recombination frequency increases, resulting in ultra-high-precision mapping studies (Svenson et al. 2012a; Logan et al. 2013b; Recla et al. 2014; Kelada et al. 2014; Smallwood et al. 2014; Gatti et al. 2014; Church et al.
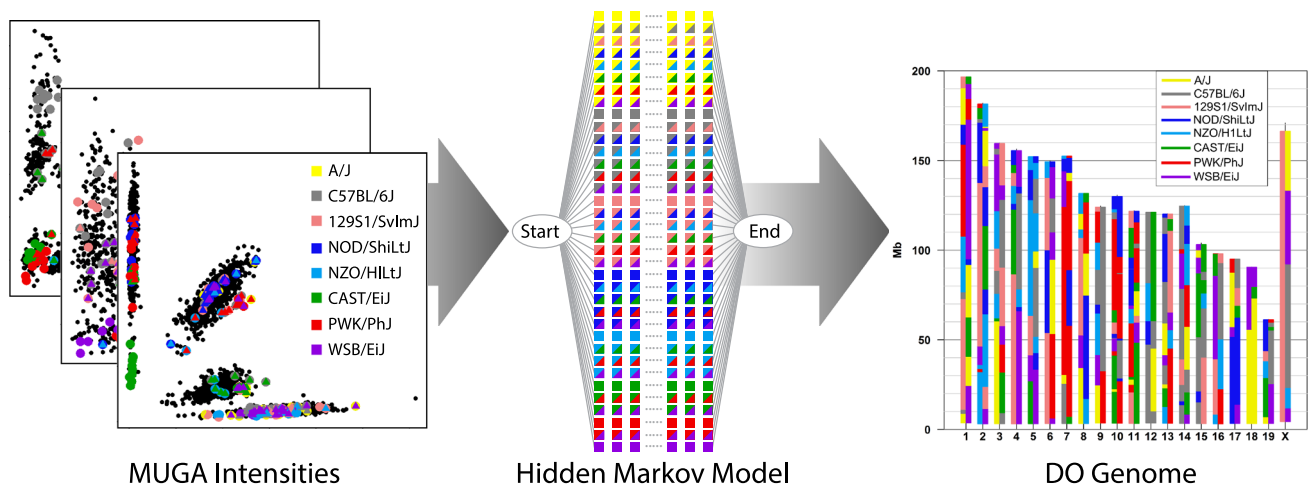
Fig. 2 DO genotyping. High-density genotyping platform—mouse universal genotyping arrays (MUGA)—have been developed to assay the genetic diversity of CC and DO mice. The raw intensity data (*left*) from genotyping arrays provide the additional information that is not available from the binary SNP genotyping calls. Specialized Hidden Markov Model software (*center*) has been developed to recover the parental haplotype block structure of individual mice (*right*). This example is for a DO mouse

2014; French et al. 2015; Church et al. 2015). Each animal is genetically unique with high levels of heterozygosity, enabling precise estimation of QTL location and allelic effect. Each DO mouse requires genotyping (see below) and haplotype reconstruction for QTL analysis (Gatti et al. 2014). CC strains and DO populations have been used to map QTLs to intervals that are less than 5 and 2 Mb, respectively (Philip et al. 2011; Logan et al. 2013b).

Integration of data from DO phenotyping studies is not as direct as with the CC strains. Aggregation and integration of data across DO studies can be achieved through common phenotypes and through shared genotypes at specific genetic loci. DO mice provide a unique opportunity to evaluate the effects of a causal genetic variant in the context of a variety of genetic backgrounds.

In addition to CC inbred strains and individual DO mice, there are numerous possibilities for study of related genotypes (Fig. 1) including the eight founder inbred strains, F1 hybrids of the founder strains, F1 hybrids of CC strains (CC-RIX), and DO mice backcrossed to inbred strains. Recombinant inbred crosses of CC-RIX provide genetically retrievable and defined F1 hybrids of CC strains, which carry all the benefits of inbred genetic reference populations with the added benefits of heterozygosity. The mapping resolution of CC-RIX is theoretically identical to that of a population of CC strains due to the fixed location of recombinations in the population. Together, these advanced mouse populations make a rich and powerful genetic resource with the potential to refine trait correlations and improve trait mapping. Founder strains and DO mice are available through The Jackson Laboratory (http://www.jax.org/); CC strains are available through the UNC Systems Genetics website (http://csbio.unc.edu/CCstatus/).

Data resources for the CC and DO

Phenotype data from early CC and DO studies have been disseminated through a variety of platforms including the MPD. In order to reap the benefits of integrated access, MPD will collect and harmonize these data in a common repository. There, the new data can be directly integrated with the QTL Archive (phenome.jax.org) and inbred strain phenotypes that make up the original MPD data resource. Integration of data from the CC and DO together with data from inbred strains, crosses, and other sets of genetically defined mice provides an opportunity for new discoveries of the function of the vast numbers of genetic variants available in laboratory mice. Access to primary data from these populations will enable and encourage reanalysis of historical data.

Mouse Phenome Database (MPD) enables users to share and reanalyze data from the CC and DO and to compare results with those obtained on other mouse strains and strain panels. MPD currently houses primary phenotype data from CC founder strains, F1 hybrids of founder strains, incipient CC strains, and DO mice covering a range of phenotypic domains, including behavior, hematology, renal function, disease susceptibility, and exercise and endurance. Additional datasets will be added to MPD as they become available. Users of these resources are encouraged to contact us about potential data submissions (see below).
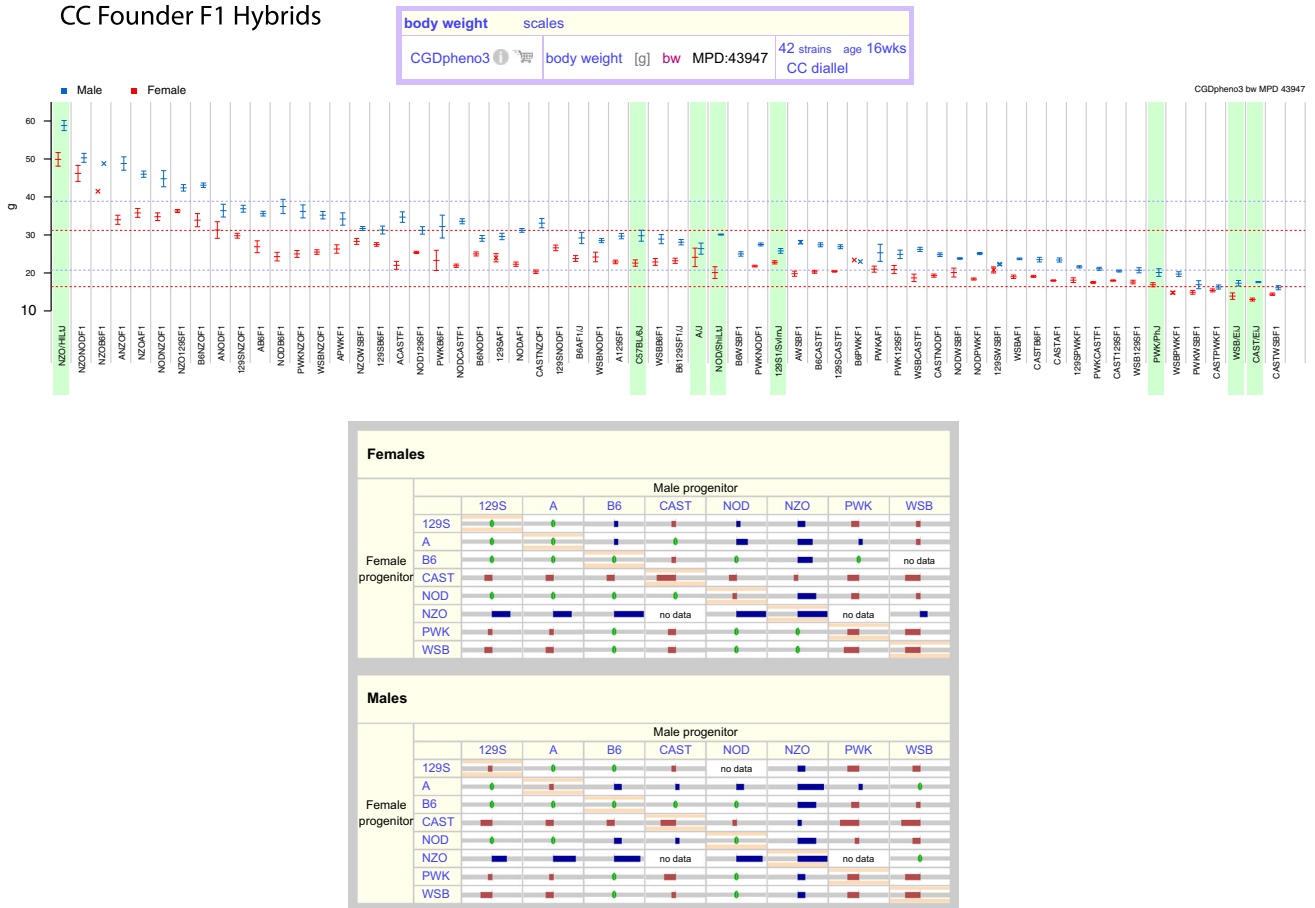
Fig. 3 CC Founder F1 Hybrids. Data for the eight CC founder strains (*green highlight*) and F1 hybrids can be viewed in multiple ways. The *upper panel* shows a measurement plot with strain on the x-axis and body weight on the y-axis. Males are in blue and females in red (mean ± SEM). The overall mean (by sex) and ± one standard deviation are indicated by *dotted lines*. There are plotting options available for this tool, including output format (here we used eps). The *lower panel* illustrates a z-score matrix of the outcome of pairwise matings for body weight. Female offspring are shown in the top matrix and male offspring in the lower matrix. Within each matrix, the female progenitor strain is shown in the first column and the male progenitor strain is shown in the first row. Matings between individual inbred strains (on the diagonal) were conducted at the same time F1 hybrids were generated. In this case, *blue* indicates high-end outliers ($Z > 1.0$), *red* indicates low-end outliers ($Z < -1.0$), and *green* indicates average (within ± 1 standard deviation). "No data" indicate instances where offspring could not be generated. This body weight data is from MPD project CGDpheno3 (Svenson et al. 2012b)

In addition to phenotype data, MPD maintains genotype data for the 8 CC founder strains, including 68 + million genomic locations (Yalcin et al. 2012; Keane et al. 2011; Wong et al. 2012) and 470,000 + locations for the mouse diversity array (Yang et al. 2009). We are expanding the database to accommodate new CC and DO genotype data collected on the "MUGA" Mouse Universal Genotyping Arrays. DO genotypes are derived from primary array intensity data using a Hidden Markov Model to reconstruct a founder haplotype mosaic (Fig. 2). Genotype data are stored as genotype calls at the probe level and as reconstructed genotype probabilities. The latter, together with the founder genomes, can be used to impute the genomes of individual DO mice with high accuracy. In addition to MPD resources, primary and derived genotype data and

information on the CC inbred strains can be obtained from a number of resources described in Morgan and Welsh, 2015, including the UNC resource at http://csbio.unc.edu/CCstatus/, and founder genome sequences can be accessed at http://www.sanger.ac.uk/resources/mouse/genomes/.

MPD visualization and analysis tools

From the MPD homepage, users can search or navigate data by strain, strain panel, investigator, phenotype, or ontology terms. There are several options for downloading data—from a single measurement to bulk download of the entire database. Video tutorials and tool demos are available to walk users through the most common tasks.
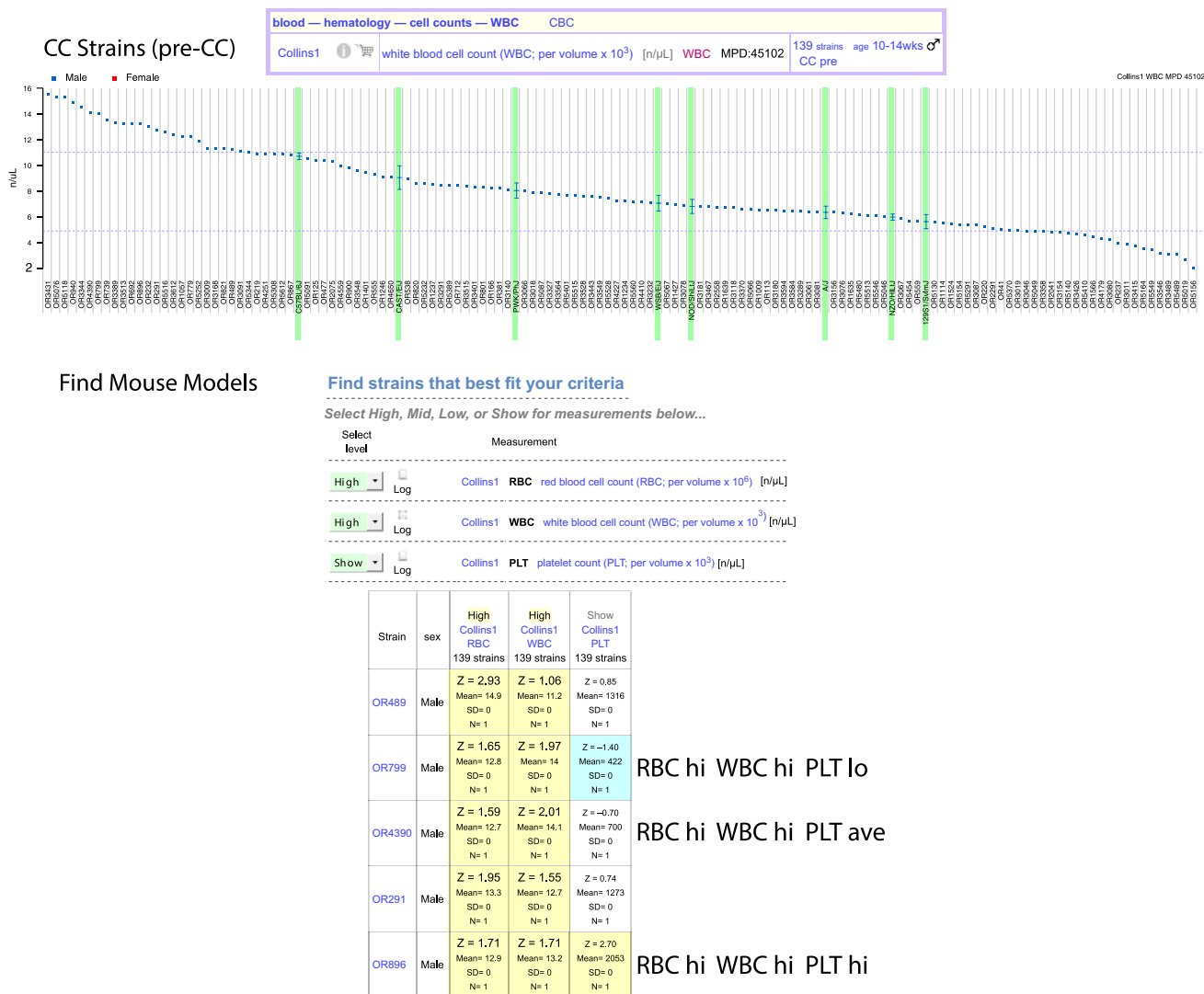
CC Strains (pre-CC)

blood — hematology — cell counts — WBC   CBC

Collins1   white blood cell count (WBC; per volume x 10³) [n/µL]   WBC   MPD:45102   139 strains age 10-14wks ♂
CC pre

Collins1 WBC MPD 45102

■ Male ■ Female

n/µL

Find Mouse Models

Find strains that best fit your criteria

Select High, Mid, Low, or Show for measurements below...

Select level   Measurement

High ▼   Log   Collins1 **RBC** red blood cell count (RBC; per volume x 10⁶) [n/µL]

High ▼   Log   Collins1 **WBC** white blood cell count (WBC; per volume x 10³) [n/µL]

Show ▼   Log   Collins1 **PLT** platelet count (PLT; per volume x 10³) [n/µL]

Strain	sex	High Collins1 RBC 139 strains	High Collins1 WBC 139 strains	Show Collins1 PLT 139 strains	
OR489	Male	Z = 2.93 Mean= 14.9 SD= 0 N= 1	Z = 1.06 Mean= 11.2 SD= 0 N= 1	Z = 0.85 Mean= 1316 SD= 0 N= 1	
OR799	Male	Z = 1.65 Mean= 12.8 SD= 0 N= 1	Z = 1.97 Mean= 14 SD= 0 N= 1	Z = −1.40 Mean= 422 SD= 0 N= 1	RBC hi WBC hi PLT lo
OR4390	Male	Z = 1.59 Mean= 12.7 SD= 0 N= 1	Z = 2.01 Mean= 14.1 SD= 0 N= 1	Z = −0.70 Mean= 700 SD= 0 N= 1	RBC hi WBC hi PLT ave
OR291	Male	Z = 1.95 Mean= 13.3 SD= 0 N= 1	Z = 1.55 Mean= 12.7 SD= 0 N= 1	Z = 0.74 Mean= 1273 SD= 0 N= 1	
OR896	Male	Z = 1.71 Mean= 12.9 SD= 0 N= 1	Z = 1.71 Mean= 13.2 SD= 0 N= 1	Z = 2.70 Mean= 2053 SD= 0 N= 1	RBC hi WBC hi PLT hi

Fig. 4 CC Strains. Data for CC strains can be viewed in a measurement plot as shown in the *upper panel* for white blood cell count (WBC), where strains are along the x-axis and WBC is on the y-axis. In this example, only males were tested (*blue*). There are options for plotting these results as well as output options (here we chose eps). The *lower panel* illustrates the criteria-fit tool. In this example, we wanted to identify CC strains with high red blood cell count (RBC) and high WBC, but with a wide range of values for platelet count (PLT). We first selected our measurements of interest (this step not shown) and applied the tool. Then we selected "High," "Mid," "Low," or "Show" for each measurement according to our criteria. Results (truncated) are shown in a screenshot where high-end outliers are in *yellow* and low-end outliers are in *blue*. These blood cell parameters are from MPD project Collins1 (Collins 2012; Kelada et al. 2012)

CC founder F1 hybrids

Summary data from CC founder F1 hybrids can be viewed graphically as illustrated in Fig. 3 (upper panel). The eight founder strains, highlighted in green, exhibit a wide range of genetic diversity for body weight with strain means ranging from 13 g (CAST/EiJ females) to 58.8 g (NZO/HlLtJ males). We see that the body weights of F1 hybrids are intermediate between the extremes of founder strains.

The same phenotype data can be viewed as a diallel structure using a visual "z-score matrix" to show the relative body weights of offspring from each pairwise cross, (Fig. 3 lower panel). This unique visualization reveals trends in the data; for example, F1 hybrids from male or female NZO/HlLtJ tend to weigh considerably more than other F1 hybrids while matings involving wild-derived inbred strains such as CAST/EiJ, PWK/PhJ, and WSB/EiJ tend to produce offspring with lower body weights.

CC strains

CC data can be summarized visually as shown in the upper panel of Fig. 4. In this example, variation in white blood cell count (WBC) across 139 incipient CC strains from different lineages exceeds the range of values observed in the founder inbred strains (highlighted in green). To identify strains of interest, users can employ a criteria-fit tool to quickly find strains exhibiting specific phenotype profiles, (lower panel of Fig. 4). Here, we wanted to find strains with high red blood cell count (RBC) and high WBC, but with one strain having elevated amounts of platelet counts (PLT) and another strain having lower than normal PLT. This tool allows users to specify characteristics in up to 15 measurements that can come from multiple studies.

DO mice

Individual measurements from DO mice are summarized visually as shown in Fig. 5 (left panel) in a histogram for the tail suspension test (time immobile). Approximately 300 DO mice were tested in this study. Mean and standard error are shown for each of the eight founder inbred strains above the plot (in green). As observed with CC strains in Fig. 4, the eight founder strains exhibit less phenotypic diversity when compared to the DO population. Another view shows correlations for all pairwise measurements in a study (right panel of Fig. 5). This view helps users elucidate relationships among traits. The inset shows a scatterplot, which can be viewed by clicking on any cell of the correlation matrix.

Each of the examples above represents a means to review and interpret the results of phenotypic characterization within populations. Several compelling studies have identified genetic variants through the integration of experiments across multiple mapping populations. These include multi-population studies of HDL levels (Burgess-Herbert et al. 2008), efforts to integrate recombinant inbred and standard inbred populations in a single analysis (Bennett et al. 2010; Ghazalpour et al. 2012), and evaluations of SNP segregation patterns across overlapping QTLs (Bubier et al. 2014). Exciting developments in genetic analysis and computation have greatly increased the rigor with which investigators can combine studies across populations through the integration of population structure into mapping analysis (Devlin and Roeder 1999; Yu et al. 2006). Further, the availability of large numbers of densely genotyped strains and deep sequencing for many other strains renders feasible the reconstruction, albeit coarse in
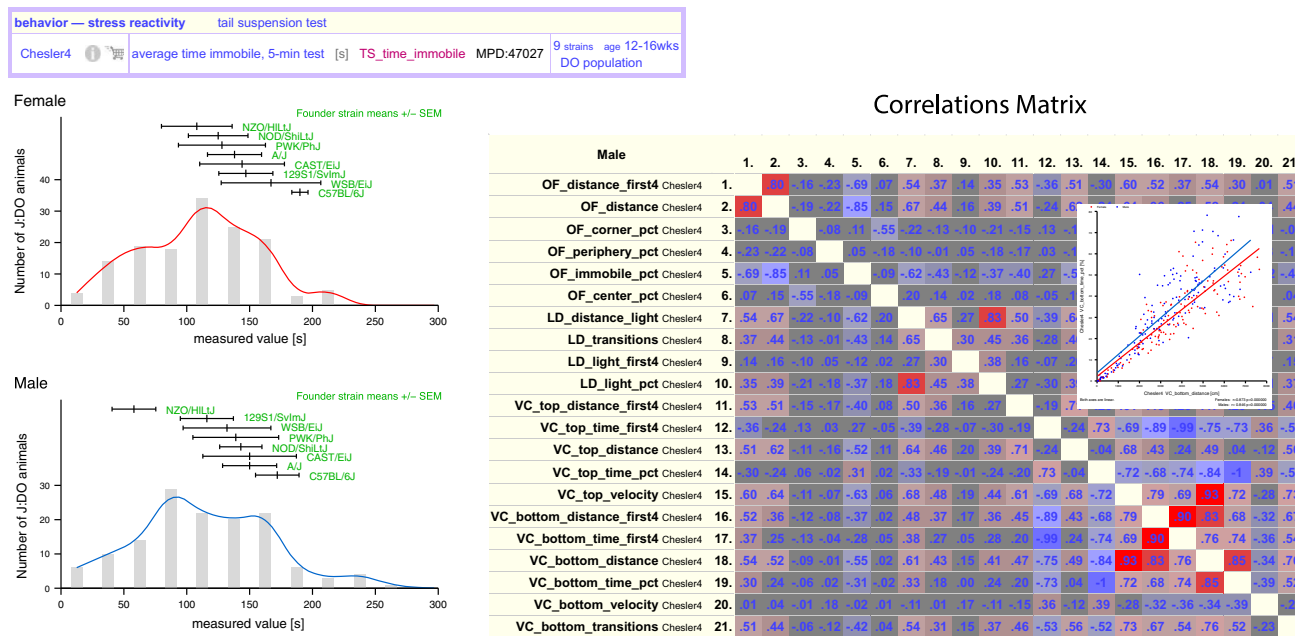

Fig. 5 Diversity Outbred (DO). Individual measurements from DO populations can be viewed as a histogram as shown in the *left panel*. Females (*red*) are above and males (*blue*) below. The founder strains (*green*) are indicated above the histogram (mean ± SEM). There are plotting options available for this tool as well as output options (here we used eps). The *right panel* shows a screenshot of a correlations matrix for all pairwise measurements. *Red* values indicate positive correlations, *blue* indicated negative. The more intense the color, the stronger the correlation. Users can view detailed scatterplots (*inset*) by clicking on cells of the matrix. Correlation coefficients and p values are provided just below the scatterplot. Plot options are available for this tool. Data are from MPD project Chesler4 (Logan et al. 2013a, b)

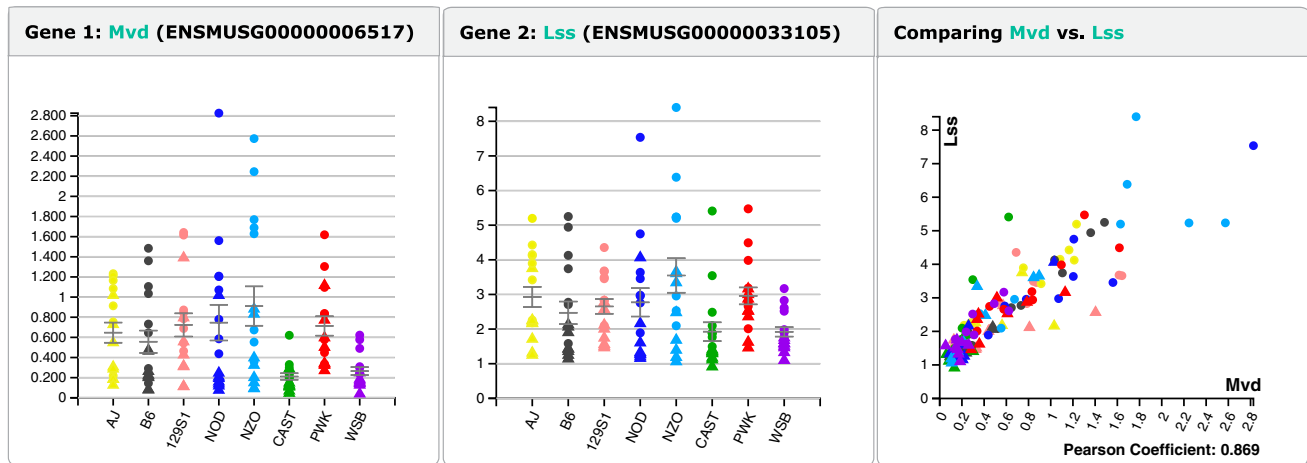## CC Founder Strain – Gene Expression Viewer

Fig. 6 CC founder strain—gene expression viewer at do.jax.org. Users are able to input their gene of interest (in this case, *Mvd*) and generate a plot of the results across all founder strains (*left panel*) through programmatic access to the underlying data. Users can also query the database to find the top 100 correlations between their gene of interest and other gene expression profiles. In this case, the top gene found was *Lss* (*middle panel*). A scatterplot is available to view the gene expression relationship, with Pearson correlation coefficient provided (*right panel*)

some cases, of virtually any mouse genome. With a common framework of sequence variation, it is now possible to integrate a much broader range of mouse phenomic studies, enabling discovery of the role of millions of genetic variants in the development of phenotypic variation and disease. The Collaborative Cross and Diversity Outbred population provide mouse resources with the power, precision and diversity to characterize the roles of these variants in molecular regulation, biological processes, and whole-organism disease-related phenotypes.

Future directions

MPD is being redeployed in a modern software engineering framework to facilitate the incorporation of additional data analysis modules and displays, including those specialized for access and analysis of variation in the Collaborative Cross and Diversity Outbred population. Keeping abreast of new developments in database and computer networking technology, the upcoming release of MPD will feature an applications programming interface (API) to allow programmatic access to the underlying data for analysis and display in the context of other web-based resources. In place of the current duplication of data that exists to support multiple websites, e.g., do.jax.org, access to a common source of reference data through the API will help ensure the integrity and consistency of data while giving developers the freedom to display and disseminate data in different formats and contexts. For example, see Fig. 6.

Contributing data

To submit data from mapping and reference populations for curation, integration, and community access through MPD, please contact us at phenome@jax.org. We acquire new data generated by members of the community; incorporate evolving technologies for archiving, integrating, and analyzing new and existing data; and expand activities that promote research reproducibility within and across resources.

References

Ananda G, Takemon Y, Hinerfeld D, Korstanje R (2014) Whole-genome sequence of the C57L/J mouse inbred strain. G3 (Bethesda) 4:1689–1692

Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang WP, He A, Kayne P, Gargalovic P, Kirchgessner T, Pan C, Castellani LW, Kostem E, Furlotte N, Drake TA, Eskin E, Lusis

AJ (2010) A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res 20:281–290

Boguski MS (2002) Comparative genomics: the mouse that roared. Nature 420:515–516

Brown SD, Moore MW (2012) The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping. Mamm Genome 23:632–640

Bubier JA, Jay JJ, Baker CL, Bergeson SE, Ohno H, Metten P, Crabbe JC, Chesler EJ (2014) Identification of a QTL in *Mus musculus* for alcohol preference, withdrawal, and Ap3m2 expression using integrative functional genomics and precision genetics. Genetics 197:1377–1393

Burgess-Herbert SL, Cox A, Tsaih SW, Paigen B (2008) Practical applications of the bioinformatics toolbox for narrowing quantitative trait loci. Genetics 180:2227–2235

CCC et al (2012) The genome architecture of the collaborative cross mouse genetic reference population. Genetics 190:389–401

Chesler EJ (2013) Out of the bottleneck: the diversity outcross and collaborative cross mouse populations in behavioral genetics research. Mamm Genome 25:3–11

Chesler EJ, Miller DR, Branstetter LR, Galloway LD, Jackson BL, Philip VM, Voy BH, Culiat CT, Threadgill DW, Williams RW, Churchill GA, Johnson DK, Manly KF (2008) The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm Genome 19:382–389

Church RJ, Wu H, Mosedale M, Sumner SJ, Pathmasiri W, Kurtz CL, Pletcher MT, Eaddy JS, Pandher K, Singer M, Batheja A, Watkins PB, Adkins K, Harrill AH (2014) A systems biology approach utilizing a mouse diversity panel identifies genetic differences influencing isoniazid-induced microvesicular steatosis. Toxicol Sci

Church RJ, Gatti DM, Urban TJ, Long N, Yang X, Shi Q, Eaddy JS, Mosedale M, Ballard S, Churchill GA, Navarro V, Watkins PB, Threadgill DW, Harrill AH (2015) Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice. Food Chem Toxicol 76:19–26

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbe JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Farber CR, Forejt J, Gaile D, Garlow SJ, Geiger H, Gershenfeld H, Gordon T, Gu J, Gu W, de Haan G, Hayes NL, Heller C, Himmelbauer H, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kendziorski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews D, Medrano JF, Miller DR, Mittleman G, Mock BA, Mogil JS, Montagutelli X, Morahan G, Morris DG, Mott R, Nadeau JH, Nagase H, Nowakowski RS, O'Hara BF, Osadchuk AV, Page GP, Paigen B, Paigen K, Palmer AA, Pan HJ, Peltonen-Palotie L, Peirce J, Pomp D, Pravenec M, Prows DR, Qi Z, Reeves RH, Roder J, Rosen GD, Schadt EE, Schalkwyk LC, Seltzer Z, Shimomura K, Shou S, Sillanpaa MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, Tarantino LM, Threadgill D, Toth LA, Valdar W, de Villena FP, Warden C, Whatley S, Williams RW, Wiltshire T, Yi N, Zhang D, Zhang M, Zou F (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

Collins FS (2012) Hematological parameters in 8 inbred founder strains and 130 + emerging lines (pre-CC) of the Collaborative Cross. MPD: Collins1, Mouse Phenome Datbase (phenome.jax.org), Bar Harbor, Maine USA

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Ferguson B, Ram R, Handoko HY, Mukhopadhyay P, Muller HK, Soyer HP, Morahan G, Walker GJ (2015) Melanoma susceptibility as a complex trait: genetic variation controls all stages of tumor progression. Oncogene 34:2879–2886

Ferris MT, Aylor DL, Bottomly D, Whitmore AC, Aicher LD, Bell TA, Bradel-Tretheway B, Bryan JT, Buus RJ, Gralinski LE, Haagmans BL, McMillan L, Miller DR, Rosenzweig E, Valdar W, Wang J, Churchill GA, Threadgill DW, McWeeney SK, Katze MG, Pardo-Manuel de Villena F, Baric RS, Heise MT (2013) Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. PLoS Pathog 9:e1003196

French JE, Gatti DM, Morgan DL, Kissling GE, Shockley KR, Knudsen GA, Shepard KG, Price HC, King D, Witt KL, Pedersen LC, Munger SC, Svenson KL, Churchill GA (2015) Diversity outbred mice identify population-based exposure thresholds and genetic factors that influence benzene-induced genotoxicity. Environ Health Perspect 123:237–245

Gatti DM, Svenson KL, Shabalin A, Wu LY, Valdar W, Simecek P, Goodwin N, Cheng R, Pomp D, Palmer A, Chesler EJ, Broman KW, Churchill GA (2014) Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda) 4:1623–1633

Ghazalpour A, Rau CD, Farber CR, Bennett BJ, Orozco LD, van Nas A, Pan C, Allayee H, Beaven SW, Civelek M, Davis RC, Drake TA, Friedman RA, Furlotte N, Hui ST, Jentsch JD, Kostem E, Kang HM, Kang EY, Joo JW, Korshunov VA, Laughlin RE, Martin LJ, Ohmen JD, Parks BW, Pellegrini M, Reue K, Smith DJ, Tetradis S, Wang J, Wang Y, Weiss JN, Kirchgessner T, Gargalovic PS, Eskin E, Lusis AJ, LeBoeuf RC (2012) Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mamm Genome 23:680–692

Graham JB, Thomas S, Swarts J, McMillan AA, Ferris MT, Suthar MS, Treuting PM, Ireton R, Gale M Jr, Lund JM (2015) Genetic diversity in the collaborative cross model recapitulates human west nile virus disease outcomes. MBio 6:e00493

Grubb SC, Bult CJ, Bogue MA (2014) Mouse phenome database. Nucleic Acids Res 42:D825–D834

Iraqi FA, Churchill G, Mott R (2008) The collaborative cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. Mamm Genome 19:379–381

Kang EY, Han B, Furlotte N, Joo JW, Shih D, Davis RC, Lusis AJ, Eskin E (2014) Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4965 mice. PLoS Genet 10:e1004022

Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellaker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, Durbin R, Jackson IJ, Czechanski A, Guerra-Assuncao JA, Donahue LR, Reinholdt LG, Payseur BA, Ponting CP, Birney E, Flint J, Adams DJ (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477:289–294

Kelada SN, Aylor DL, Peck BC, Ryan JF, Tavarez U, Buus RJ, Miller DR, Chesler EJ, Threadgill DW, Churchill GA, Pardo-Manuel de Villena F, Collins FS (2012) Genetic analysis of hematological parameters in incipient lines of the collaborative cross. G3 (Bethesda) 2:157–165

Kelada SN, Carpenter DE, Aylor DL, Chines P, Rutledge H, Chesler EJ, Churchill GA, Pardo-Manuel de Villena F, Schwartz DA, Collins FS (2014) Integrative genetic analysis of allergic inflammation in the murine lung. Am J Respir Cell Mol Biol 51:436–445

Logan RW, Robledo RF, Recla JM, Chesler EJ (2013a) Behavioral and nociception traits in a diversity outbred (DO) mouse population and 8 founder inbred strains. MPD:Chesler4, Mouse Phenome Database (phenome.jax.org), Bar Harbor, Maine USA

Logan RW, Robledo RF, Recla JM, Philip VM, Bubier JA, Jay JJ, Harwood C, Wilcox T, Gatti DM, Bult CJ, Churchill GA, Chesler EJ (2013b) High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. Genes Brain Behav 12:424–437

Morgan AP, Welsh CE (2015) Informatics resources for the collaborative cross and related mouse populations. Mamm Genome

Morahan G, Balmer L, Monley D (2008) Establishment of "The Gene Mine": a resource for rapid identification of complex trait genes. Mamm Genome 19:390–393

Paigen K, Eppig JT (2000) A mouse phenome project. Mamm Genome 11:715–717

Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, Beckmann MA, Spence JS, Jackson BL, Galloway LD, Barker P, Wymore AM, Hunsicker PR, Durtschi DC, Shaw GS, Shinpock S, Manly KF, Miller DR, Donohue KD, Culiat CT, Churchill GA, Lariviere WR, Palmer AA, O'Hara BF, Voy BH, Chesler EJ (2011) Genetic analysis in the collaborative cross breeding population. Genome Res 21:1223–1238

Phillippi J, Xie Y, Miller DR, Bell TA, Zhang Z, Lenarcic AB, Aylor DL, Krovi SH, Threadgill DW, de Villena FP, Wang W, Valdar W, Frelinger JA (2014) Using the emerging Collaborative Cross to probe the immune system. Genes Immun 15:38–46

Rasmussen AL, Okumura A, Ferris MT, Green R, Feldmann F, Kelly SM, Scott DP, Safronetz D, Haddock E, LaCasse R, Thomas MJ, Sova P, Carter VS, Weiss JM, Miller DR, Shaw GD, Korth MJ, Heise MT, Baric RS, de Villena FP, Feldmann H, Katze MG (2014) Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. Science 346:987–991

Recla JM, Robledo RF, Gatti DM, Bult CJ, Churchill GA, Chesler EJ (2014) Precise genetic mapping and integrative bioinformatics in diversity outbred mice reveals Hydin as a novel pain gene. Mamm Genome 25:211–222

Ringwald M, Iyer V, Mason JC, Stone KR, Tadepally HD, Kadin JA, Bult CJ, Eppig JT, Oakley DJ, Briois S, Stupka E, Maselli V, Smedley D, Liu S, Hansen J, Baldock R, Hicks GG, Skarnes WC (2011) The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. Nucleic Acids Res 39:D849–D855

Smallwood TL, Gatti DM, Quizon P, Weinstock GM, Jung KC, Zhao L, Hua K, Pomp D, Bennett BJ (2014) High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. J Bethesda 4:2353–2363

Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L, Churchill GA (2012a) High-resolution genetic mapping using the mouse diversity outbred population. Genetics 190:437–447

Svenson KL, Lenarcic AB, Churchill GA, Valdar W (2012b) Multi-system survey of mouse physiology in 8 inbred founder strains and 54 F1 hybrids of the Collaborative Cross. MPD:CGDpheno3, Mouse Phenome Database (phenome.jax.org) Bar Harbor, Maine USA

Threadgill DW, Churchill GA (2012) Ten years of the collaborative cross. Genetics 190:291–294

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

Vered K, Durrant C, Mott R, Iraqi FA (2014) Susceptibility to klebsiella pneumoniae infection in collaborative cross mice is a complex trait controlled by at least three loci acting at different time points. BMC Genomics 15:865

Welsh CE, Miller DR, Manly KF, Wang J, McMillan L, Morahan G, Mott R, Iraqi FA, Threadgill DW, de Villena FP (2012) Status and access to the collaborative cross population. Mamm Genome 23:706–712

Wong K, Bumpstead S, Van Der Weyden L, Reinholdt LG, Wilming LG, Adams DJ, Keane TM (2012) Sequencing and characterization of the FVB/NJ mouse genome. Genome Biol 13:R72

Yalcin B, Adams DJ, Flint J, Keane TM (2012) Next-generation sequencing of experimental mouse strains. Mamm Genome 23:490–498

Yang H, Ding Y, Hutchins LN, Szatkiewicz J, Bell TA, Paigen BJ, Graber JH, de Villena FP, Churchill GA (2009) A customized and versatile high-density genotyping array for the mouse. Nat Methods 6:663–666

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208