Mouse Phenome Database

Stephen C. Grubb, Terry P. Maddatu, Carol J. Bult and Molly A. Bogue*

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

ABSTRACT

The Mouse Phenome Database (MPD; http://www.jax.org/phenome) is an open source, web-based repository of phenotypic and genotypic data on commonly used and genetically diverse inbred strains of mice and their derivatives. MPD is also a facility for query, analysis and *in silico* hypothesis testing. Currently MPD contains about 1400 phenotypic measurements contributed by research teams worldwide, including phenotypes relevant to human health such as cancer susceptibility, aging, obesity, susceptibility to infectious diseases, atherosclerosis, blood disorders and neurosensory disorders. Electronic access to centralized strain data enables investigators to select optimal strains for many systems-based research applications, including physiological studies, drug and toxicology testing, modeling disease processes and complex trait analysis. The ability to select strains for specific research applications by accessing existing phenotype data can bypass the need to (re)characterize strains, precluding major investments of time and resources. This functionality, in turn, accelerates research and leverages existing community resources. Since our last NAR reporting in 2007, MPD has added more community-contributed data covering more phenotypic domains and implemented several new tools and features, including a new interactive Tool Demo available through the MPD homepage (quick link: http://phenome.jax.org/phenome/trytools).

INTRODUCTION

The laboratory mouse is an invaluable model organism for investigating the genetic basis of human disease. Studies have demonstrated the efficacy of comparative mouse–human genomics to identify novel mechanisms of human disease progression, underscoring the need to make mouse strain data widely available for community access. Using inbred strain data for integrative studies leverages their fixed genotypes and expands their utility to determine molecular relationships between disease and associated risk factors.

The Mouse Phenome Project was launched as an international collaboration to complement the mouse genome sequencing effort and provide a research resource and integral tool for complex trait analysis (1). This powerful approach, termed phenomics, captures complexities of entire biological pathways that are not accessible through conventional approaches. A central database was built to support the Project and provide a repository for the large amounts of data collected. The database, called the Mouse Phenome Database (MPD; www.jax.org/phenome), has been publicly available since 2001 (2). MPD is a grant-supported effort with three full-time staff members headquartered at The Jackson Laboratory (JAX), a non-profit biomedical research institute with a focus on the mouse as a model for understanding human biology and disease (http://www.jax.org).

The Mouse Phenome Project promotes and facilitates strain surveys that follow a set of recommendations proposed by members of the research community to standardize testing across laboratories and over time, and ultimately to maximize data reproducibility and value. A set of diverse inbred mouse strains was carefully chosen for systematic phenotyping to generate the building blocks of the phenome of the laboratory mouse. The Project is open to researchers with expertise in any biomedically-relevant field of study. Strain characteristics data are received from members of the scientific community and added to the MPD standardized framework, providing users a platform for data exploration, analysis and hypothesis testing. Project recommendations, priority strains and data submission guidelines are accessible through the MPD homepage.

The ability of investigators to use MPD to find causal genes and biomarkers of human disease will be significantly enhanced by the capacity to integrate human data with comprehensive information on the laboratory mouse. International efforts are underway to address integration

*To whom Correspondence should be addressed. Tel: +1 207 288 6016; Fax: +1 207 288 6079; Email: molly.bogue@jax.org

issues for several public mouse resources holding phenotypic data (3), including Europhenome at Harwell (UK), PhenoSITE at Riken (Japan) and MPD. Discussions are in progress to coordinate data formats and reporting standards that ensure interoperability across databases. We have also been involved in the development of minimum information for mouse phenotyping procedures (MIMPP; www.interphenome.org) as part of the larger community-wide effort for minimum information for biological and biomedical investigations (MIBBI) (14) that fosters coordination of minimum information checklists such as minimum information about a microarray experiment (MIAME). These checklists ensure adequate descriptions about the biological material being tested (or used for testing) and the assays employed for measuring biological or behavioral manifestations (traits). Until community standards are in place for reporting phenotypic data, we will continue using the definitions adopted when MPD was launched in 2001 (Table 1).

DATA IN MPD

Our last NAR update was in 2007 (5). Most of the discussion points, figures and URLs set forth there are still current. Before presenting our recent updates, we will review some fundamental points about MPD. Every MPD project has a dataset and detailed protocols, health status and environmental parameters of the test animals, and any other information essential to understand and evaluate the data (Table 1). MPD is also a repository for protocol information where a library of procedures and assays are maintained so that others in the community may benefit from their use. Most phenotypic datasets in MPD are in strain survey format. For example, an expert in lipid metabolism participating in the Project and following Project recommendations might take readings on 10 females and 10 males of 40 strains and submit the individual animal data in a spreadsheet having one row per mouse and multiple columns for various lipid measurements. We would then annotate and format the data to meet MPD standards. Each measurement is classified and integrated in the MPD phenotype category structure. We compute summary statistics, where our unit of analysis is an MPD measurement with strain (by sex) being our analysis group (we do not combine male and female data nor do we combine data from different MPD measurements). Individual animal data and summary statistics are available for downloading as well as protocols and other metadata. To identify possible biological correlations (related phenotypes may indicate common genes or pathways), we further analyze each measurement by regression analysis with every other measurement currently in the database and store the results to support queries based on measurement correlations (see below) (2). In addition to phenotypic data, strain genotypes are collected and stored in MPD so that phenotypic and genotypic data can be juxtaposed, facilitating the ability to determine how allele-specific variations translate to differences in mouse phenotype.

Current contents

At the present time MPD contains around 1400 phenotypic measurements and ~740 million single nucleotide polymorphism (SNP) allele calls. Over 600 strains of mice are represented in MPD where phenotypic and/or genotypic data are available (most of the data are for MPD priority strains and their derivatives). Around 200 people are currently registered as principal investigators of MPD projects (phenotyping and genotyping), representing ~130 institutions in 12 countries, and supported by ~60 funding agencies and research foundations worldwide. Phenotypic measurements are from 75 investigator-contributed projects (~20 other projects are pending), with coverage in a number of important areas (summarized in Table 2). Several large phenotyping initiatives utilize MPD as the official repository for their strain survey data, including the Jackson Aging Center (Nathan Shock Center of Excellence in the Basic Biology of Aging) and the Heart, Lung, Blood and Sleep Disorders Center (NHLBI Program for Genomic Applications) (6).

Phenotypic data currently available can be classified as baseline (72%), longitudinal aging data (14%), or controlled studies of intervention effects (14%) such as administering drugs or high-fat diet, or exposure to toxins or pathogens. Each measurement contains data from multiple strains of mice with as many as 60 strains tested (the average per measurement is 20 strains). Most projects involve both sexes (84%) and use MPD priority strains (82%). The remaining 18% are special strain panels where the progenitors are often MPD priority strains. Analysis tools for phenotypic data are available in the MPD Toolbox depicted in Figure 1. To see how these tools work, see the interactive Tool Demo available through the MPD homepage (quick link: http://phenome.jax.org/phenome/trytools).

Genomic characterization of mouse strains is currently supported in MPD by way of SNP data. Copy number variant (CNV) data will be added in the future. SNP datasets are supplied by investigators (or institutions) either directly or as freely available data downloads. The MPD SNP collection currently includes 8 + million unique genomic locations for 16 strains in our high-density merged dataset (about 3.5 SNP locations per 1 kb) and lesser amounts of SNP data for approximately 125 additional strains plus 7 recombinant inbred (RI) strain panels. Overall, there are 18 SNP data sources represented in MPD, including SNPs from Broad, Celera, Perlegen (NIEHS), Wellcome Trust, Genomics Institute of the Novartis Research Foundation (GNF), and The Jackson Laboratory. To provide maximal utility for different research applications, MPD consolidates SNPs from multiple sources based on SNP density and the complement of strains assayed. Currently there are five datasets with four degrees of SNP density (high, as defined above; to very low ~2000 SNPs per entire genome). SNP and gene annotations from external resources such as Mouse Genome Informatics (MGI; http://www.informatics.jax.org) (7), NCBI dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP), (8) and Ensembl (http://www.ensembl.org) (9) are part of the merge operation. NCBI dbSNP also provides

Table 1. MPD Definitions

Term	Definition	Relationships	Comments
MPD Project	Entity that logically binds a scientific investigation's unique dataset, protocols and other documentation necessary to evaluate and use the data.	An MPD project has only one primary dataset and one protocol.	A project usually represents the characterization of a cohort of animals tested for multiple traits; a project can stand alone in that the source (submitting investigator), all the data, and all the information needed to understand the data are bound.
MPD Protocol	Entity that binds one or more specific procedures, contains information about test animals, and their environment, anesthesia, experimental design, interventions, workflow and other overarching concepts that apply to one or more component procedures of the protocol.	There is one protocol for every project. A protocol may contain more than one procedure.	An intervention is a controlled perturbation (or treatment) that is part of the study, such as high-fat diet, ethanol in drinking water or toxin exposure. For every intervention, there should be a control (baseline).
Procedure	Detailed information about an experimental method, containing descriptions about equipment, reagents, solution preparation, safety issues, special definitions, formulas and data analysis.	A project may involve multiple procedures bound by a single protocol. A procedure may involve one or more assays.	
Assay	Analytical test that determines a range of values (preferably quantitative) for one or more biological or behavioral manifestations (traits).	A procedure may involve multiple assays. An assay quantifies one or more traits.	
Trait	Biological or behavioral manifestation of an individual that can be measured, quantified or scientifically categorized. A trait is a product of an organism's genome, its natural history (e.g. age), its environmental history (e.g. fostered pup) and controlled experimental perturbations (e.g. high-fat diet); when a trait is quantified for an individual or strain, it is called a characteristic (or parameter).	An assay may quantify one or more traits, and a trait may be deconstructed into component traits; a phenotype is determined by one or more traits.	When a trait is measured on a particular set of individuals as part of a specific scientific investigation (MPD project), the resulting set of data points is collectively called an MPD measurement.
MPD Measurement	Collection of data points that measure a trait in individuals of a population; measurement values span the range of biological possibilities for that population, gathered as part of a particular scientific investigation (MPD project) which follows a defined procedure and well-controlled assay.	An MPD measurement quantifies (or otherwise defines) one trait. There may be multiple. MPD measurements per project. MPD measurements are the unit of analysis (by strain and sex). Each measurement is annotated and has a set of attributes or is otherwise linked to essential information, such as:	• Accession ID • Variable name • Project symbol • Protocol [procedure(s)] • Short description • Tag for baseline/control/intervention • Units • Strains and sex tested • Sample sizes • Age of mice • Data type • Classification annotations • Supplemental information

For example, a protocol might describe multi-system testing for a panel of inbred strains, which would require the description of multiple procedures, one of which might be hematology. The procedure 'hematology' involves multiple assays, including hematocrit and complete blood count (CBC). The CBC is an assay that measures multiple traits (WBC, RBC, etc). RBC is a quantifiable trait. The values from an assay (data points) are collectively called an MPD measurement.

It should be noted that a trait may be measured multiple times. For example RBC could be measured in more than one study or measured at multiple ages within a single study. Because one or more conditions of testing are different each time a trait is quantified for a population, the measurement receives a unique name, accession number and other attributes (above). Some redundancy in testing is encouraged for validation purposes because test conditions are rarely completely identical when animal age, environment and protocol nuances are considered.

Table 2. SNAPSHOT of Selected MPD Content

Aging	blood chemistry • hematology • survival curves • urinalysis
Appearance	coat color
Behavior	activity • alcohol • anxiety • exploratory • learning and memory • stress reactivity • wildness
Blood chemistry	electrolytes • glucose • proteins (enzymes, hormones)
Blood hematology	CBC • coagulation • red cell parameters
Blood lipids	cholesterol • fatty acids • phospholipid • triglycerides
Body composition	fat • fat pads • lean
Body weight & size	length • weight • growth curves
Bone	geometry • bone mineral content • strength • physiology
Brain	morphology • physiology and function
Cancer	metastatic progression • tumor growth • tumor histopathology
Cardiovascular	blood pressure • ECG • heart rate • organ weight • atherosclerosis (aorta fatty-streak lesions)
Drinking preference	alcohol • salt solutions
Ear	hearing • tympanometry • acoustic startle response • morphology (length)
Endocrine	adrenal • hormones
Eye	morphology • vision • degeneration
Immunity	H2 haplotype • thymus • spleen • peripheral blood lymphocytes
Infectious disease	Bacillus anthracis • pathogen-accelerated atherosclerosis
Kidney	urinalysis • metrics • pathology
Liver and gallbladder	function • morphology • pathology (gallstones)
Metabolism	activity • energy (intake, production) • food intake • water intake
Muscle	skeletal (weight, area) • function (grip strength)
Nervous system	autonomic function • neuromuscular function • sensorimotor function
Neurosensory	hearing • nociception • prepulse inhibition • vision
Reproduction	assisted reproduction technologies • colony reproductive performance
Respiratory	lung capacity • allergen-induced inflammation
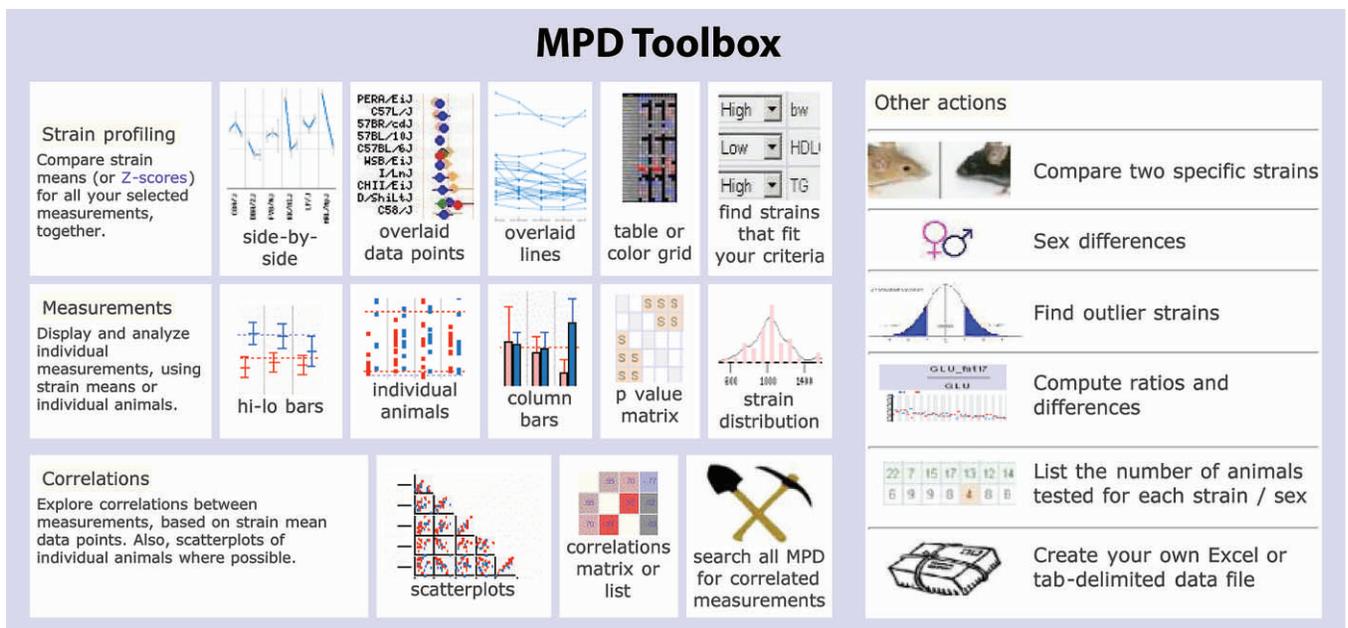
Figure 1. MPD Toolbox. Screenshot of MPD analysis tools, grouped by function: strain profiling (identifying mouse models with specific characteristics), measurement displays, correlations, and other actions. Some of our new tools are featured elsewhere: side-by-side plot and color-grid are shown in Figure 5, the overlaid-data-points plot in Figures 4 and 6. Try the interactive Tool Demo from the MPD homepage or go to http://phenome.jax.org/phenome/trytools.

the service of updating SNP locations when the mouse genome reference assembly is updated, and MPD mirrors these updates when they become available. MPD does not store flanking sequences or other lower-level trace information, but we maintain links to NCBI dbSNP and other resources holding this data. MPD SNP tools for retrieval and analysis are illustrated in Figure 2.

New phenotype strain survey data and functionality

Coat color has been a classic model for many studies in mouse genetics. Since our last NAR update, photographs of 60 strains have been made publicly available (Figure 3), with many strains having a composite of up to four different photos. In addition to coat color, there are new

Figure 2. MPD SNP interface tools for retrieval and filtering SNPs. SNPs may be retrieved by gene symbol or genomic location (left panel), or by more complex criteria. A SNP wizard (top right) has been added to assist users, showing possible options for each retrieval method. Users must select the optimal SNP dataset for their particular research application (see text for details). MPD provides information about each dataset to facilitate the process. To narrow SNP results as much as possible, options for additional criteria are offered (right panel), such as various filtering modes or by annotations (Ensembl, NCBI, MGI). An option to set the confidence interval for imputed SNPs is provided for the CGD SNP dataset (see text for more details and Figure 7).

postings of quantitative measurements that can be classified as baseline strain surveys, longitudinal aging data and controlled intervention studies. New data highlights include studies of bone density, chemically-induced tumorigenesis, assisted reproduction, anxiety and exploratory behavior, vision and eye morphology (for example, see Figure 4). In addition to inbred strains, data have been added for chromosome substitution panels and an eight-way F1 cross panel (see a list of selected projects and participants released since our last NAR update in Table 3). Several phenotype analysis tools have been improved or developed for better visualization and pattern recognition (see Figure 5 for examples and details). Of particular note is a new tool that helps users link phenotype and genotype (see 'Find Genomic Regions' below).

The number of MPD measurements has grown substantially, and we do not expect this trend to wane. To improve browsing and search capabilities, we have refined our measurement classification scheme to present measurement listings in a more compact and readable way by grouping measurements with common metadata, for example measurements in a time or dose series are grouped together conserving space and eliminating the redundancy of repeated text (see example in Figure 6). In addition, we have split out 'intervention' and 'age' from the category hierarchy which simplifies the classification scheme further and makes it easier on the eye to browse lists of measurements. In some situations, listing measurements without groupings is helpful, so we have retained this option for users (see Figure 6 comparing these options).

### New genotype (SNP) data and functionality

We have made various incremental improvements to the MPD SNP interface such as adding a SNP wizard
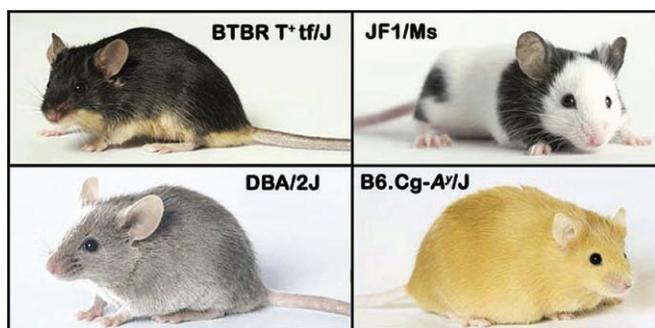
Figure 3. Mouse strain coat color and appearance. Sixty strains have been professionally photographed under standardized conditions (lighting, background, etc.). Four strains are shown here to illustrate the wide range of phenotypes found in laboratory strains for coat color and appearance. DBA/2J is one of the oldest inbred strains in existence. BTBR T^+ tf/J, an inbred strain developed more recently, has a severe defect in corpus collosum development and exhibits extreme behavioral phenotypes. JF1/Ms, a wild-derived inbred strain from Japan (10), has congenital eye abnormalities (Figure 4) and has remarkably high percent body fat although its total body weight is relatively low compared to other strains; and B6.Cg-A^y/J is a congenic strain that exhibits severe obesity-related phenotypes. MPD contains data for inbred strains and their derivatives, such as congenic, consomic and recombinant inbred strains. Photographs by Stanton Short, The Jackson Laboratory.

interface and offering more flexible polymorphism filtering options. New SNP data from several sources have been added recently, including a 12 000-location set for 43 strains (Merck-Rosetta) (11) and mitochondrial characterization of 22 strains (University of Porto, Portugal) (12). The largest new addition is a dataset from the Center for Genome Dynamics (CGD; http://cgd.jax.org) containing a mixture of actual and mathematically imputed allele calls, covering 7.8 + million genomic locations for 74 strains, built by merging data from a number of public data sources and then applying a hidden Markov model algorithm to impute calls that are missing, and attaching a confidence level probability value to each imputed call (13). After importing and processing this dataset, we found that 78% of the SNPs are imputed, and of those, 72% have a confidence level of 0.9 or higher, while 86% have a confidence level of 0.6 or higher. MPD supports queries on this imputed dataset where data are listed based on a specified minimum confidence level threshold, for example 'show only <u>actual</u> calls' or 'show only <u>imputed</u> calls with confidence <u>level</u> of 0.9 or higher' (the right panel of Figure 2 shows this option).

A new exploratory SNP-based tool (called 'Find Genomic Regions') has been developed based on the concept of identity by decent (IBD) whereby two strains or strain sets can be compared across the entire mouse genome, to find regions where the two strain sets differ the most. We make the assumption that phenotypic differences reflect genotypic differences and that differences in a causative element (gene or regulatory region) are present in ancestral variation and are not due to recent mutations. This tool is based on SNP data from several large datasets (Perlegen, Broad, Celera) which together cover 8 + million genomic locations for 16 strains (14–16). This tool can be
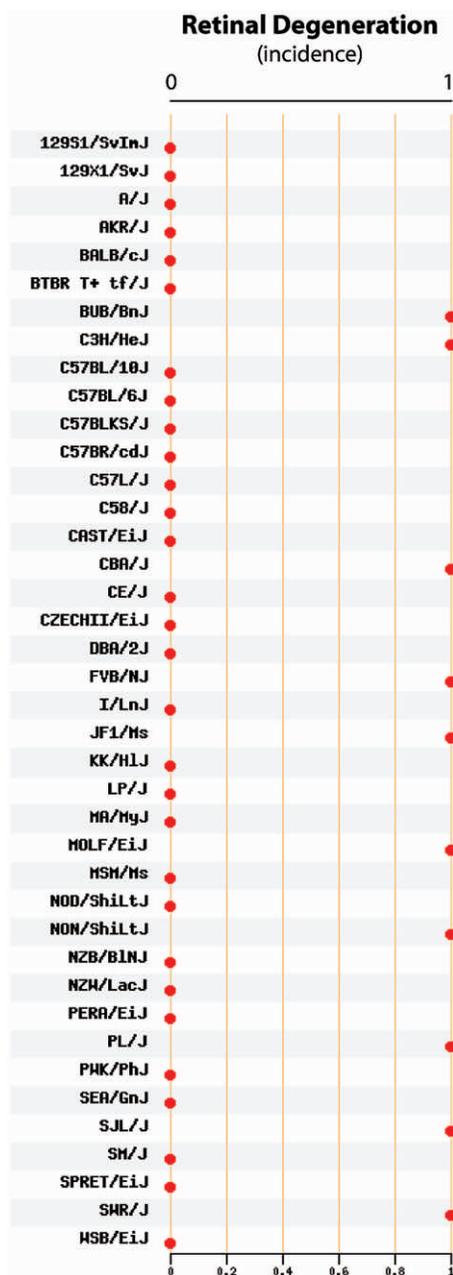

Figure 4. Retinal degeneration. Forty inbred strains were examined for eye abnormalities (retina, cornea, lens, iris). Twenty-five percent of the strains exhibit retinal degeneration by 6–7 weeks of age. This study underscores the importance of using strain characteristics data to choose optimal strains for testing. An investigator using a behavioral apparatus that uses visual cues for scoring would not choose JF1/Ms for the study. Without knowing that JF1 has severe vision problems, the investigator might incorrectly conclude that JF1 is unintelligent, anxious or lethargic. Data from Hawes1 MPD:267 (2008).

used in concert with strain survey data to locate genomic regions that may have an effect on a given phenotype (see example in Figure 7). This tool operates not by tabulating individual SNP locations (which would take much too long for a web-based tool) but rather by scanning an intermediate file that has been produced in advance,

Table 3. SNAPSHOT of selected MPD projects added since last NAR reporting

Author	Institution	Description
Crabbe JC	Oregon Health & Science University	Testing the effects of alcohol by quantitating motor incoordination using the parallel rod floor apparatus
Metten P, Crabbe JC	Oregon Health & Science University	Ethanol-induced intoxication and withdrawal severity
Finn DA, Murillo A, Yoneyama N, Crabbe JC	Oregon Health & Science University	Voluntary ethanol consumption in 22 inbred strains
Richfield EK, Mhyre TR, Cory-Slechta DA, Thiruchelvam M, Chesler EJ	EOHSI; University of Rochester Medical Center	Behavioral, neurochemical, neuroanatomical, and neurotoxicological characterization of the midbrain dopamine system
Brown RE, Schellinck HM, Gunn RK, Wong AA, O'Leary TP	Dalhousie University (CANADA)	Anxiety, exploratory behavior and motor activity; Visual ability and spatial, motor and olfactory learning and memory
Gershenfeld HK	University of Texas – Southwestern	Imipramine response and tail suspension test
Graubert TA, Watters JW, McLeod H	Washington University School of Medicine	ENU-induced tumorigenesis
Churchill GA, Baldwin C	JAX; Boston University	Bone characteristics and body composition of an 8-way diallele cross
Donahue L, Beamer WG, Bogue MA, Churchill GA	JAX	Models of skeletal geometry and bone strength
Donahue L	JAX	Bone mineral density, body composition, and craniofacial characterization
Hawes NL, Chang B, Davisson MT	JAX	Morphological examination of the eye in 41 inbred strains of mice
Chang B, Hawes NL, Davisson MT	JAX	Electroretinogram (ERG) examination in 17 inbred strains of mice
Sugiyama F, Tsukahara C, Paigen B	University of Tsukuba (JAPAN); JAX	Blood pressure for 25 strains
Tomasini-Johansson BR, Mosher DF	University of Wisconsin	Concentration of fibronectin in mouse plasma
Peters LL	JAX	Aging study: Blood hematology
Yuan R	JAX	Aging study: Blood chemistry
Yuan R, Rosen CJ, Beamer WG	JAX	Aging study: IGF-1 and body weight
Korstanje R	JAX	Aging study: Urine albumin and creatinine
Seburn KL, Xing S, Burgess RW	JAX	Aging study: Grip strength and gait analysis
Taft RA, Byers SL	JAX	Assisted reproductive technologies (ARTs)
JAX Phenotyping Services	JAX	Comprehensive survey of 11 inbred strains
Svenson KL, Forejt J, Donahue L, Paigen B	JAX	Multi-system analysis of mouse physiology, C57BL/6J-Chr#PWD chromosome substitution strain panel
Nadeau JH, Hill AE	Case Western Reserve University School of Medicine	C57BL/6J-Chr#A/J chromosome substitution strain panel: Diet-Induced Obesity
Lake J, Donahue L, Davisson MT	JAX	Comprehensive phenotype survey, C57BL/6J-Chr#A/NaJ chromosome substitution strain panel
Palmer A, Ponder CA, Munoz M, Gilliam C	University of Chicago	Innate anxiety-like behavior and fear conditioning in C57BL/6J-Chr#A/J chromosome substitution strain panel
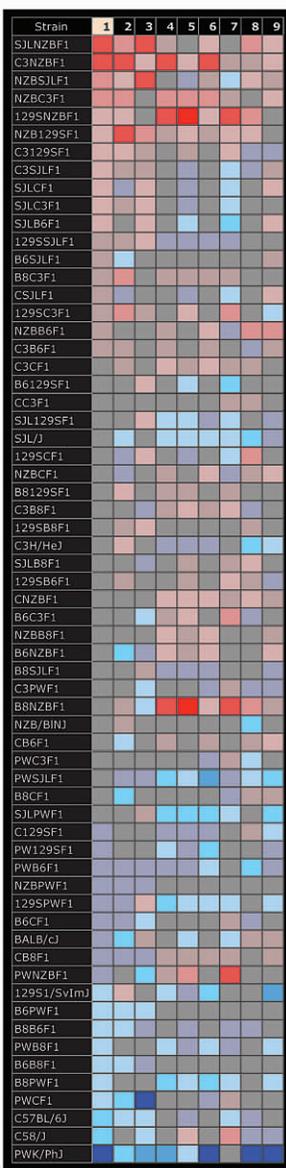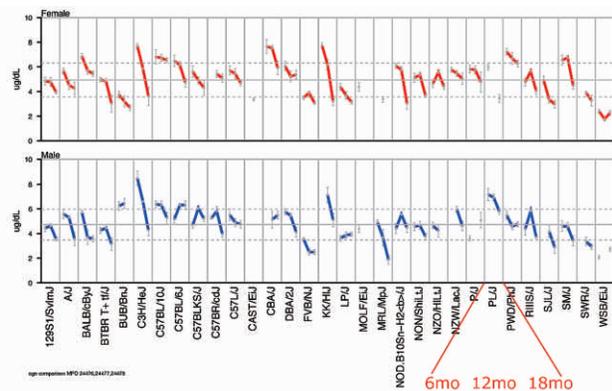
Figure 5. New phenotype tools for strain profiling and identifying important new mouse models for research. The Jackson Aging Center is in the process of testing 32 inbred strains for a wide variety of phenotypic traits at 6, 12, 18 and 24 months of age. A new tool has been developed to visualize aging trends graphically (above). In this example, three time points for thyroxine (T4) are shown for each strain. Such a tool is critical for understanding aging processes which

containing tabulations of strain differences for successive 50 kb windows.

New QTL analysis archive

At the request of members of the research community, MPD has developed an archive of quantitative trait loci (QTL) analysis datasets. At this writing there are 23 datasets available in a variety of subject areas, many associated with projects that have also contributed inbred strain survey data to MPD. These QTL studies typically involve intercross (F2) or backcross (N2) progeny of strains in the MPD priority list. Currently these data are available in Excel spreadsheets (R/qtl format), where a spreadsheet contains phenotypic measurements for each individual in the population (usually several hundred mice) and their genotypes (typically based on Mit markers). Linkages to MPD phenotype categories are maintained to optimize search capabilities, and links to MGI are maintained for connectivity to other databases. The primary purpose of the QTL archive is to provide a public repository for these datasets so that investigators can easily find and download them for custom analyses, e.g. combined cross analysis to reduce QTL intervals to a more manageable size for subsequent gene testing and validation. We plan to add QTL analysis tools in the near future, including interactive QTL maps.

HIGH LEVEL OVERVIEW OF IMPLEMENTATION

All public access to MPD is via our web site. MPD runs on a Solaris (Unix) computer system and is implemented using an open source software platform that includes relational database, web presentation scripting, and integrated graphical data plotting components. Apache web server software serves our web pages using a CGI method, and web 'cookies' are utilized to manage user preferences and item collections. Some custom-written programs in the 'C' language are invoked for compute-intensive tasks such as computation of statistics and correlations, and for SNP data display. We have a URL interface that web site developers can use to build links to specific MPD data views (visit our web site and search on 'URL').

The database has 70 data tables, including 6 containing mouse biometric data, 30 for SNP data, 17 catalogs and dictionaries, and 8 of various internal and external mappings (our detailed data model can be viewed by visiting our web site and searching on 'schema'). Data are typically contributed using Excel spreadsheets transmitted as

are not always linear over time. This tool helped identify several complex phenotypes which would not have been discovered if examining only one time point. Another new tool useful for identifying mouse models in shown in the lower panel. The color grid tool is based on the heat map concept using Z-scores. Strain names are listed on the left, measurement numbers are shown along the top (1–6) which are fully defined below the grid when viewing online. Shades of red indicate those measurements that are above the overall mean and blue indicates those that are below. Intensity of color tracks with severity, where the more intense colors are the most extreme. More new tools are featured in Figure 6. Data (upper) from Yuan3 MPD:244 (2008); (lower) Churchill1 MPD:171 (2004).
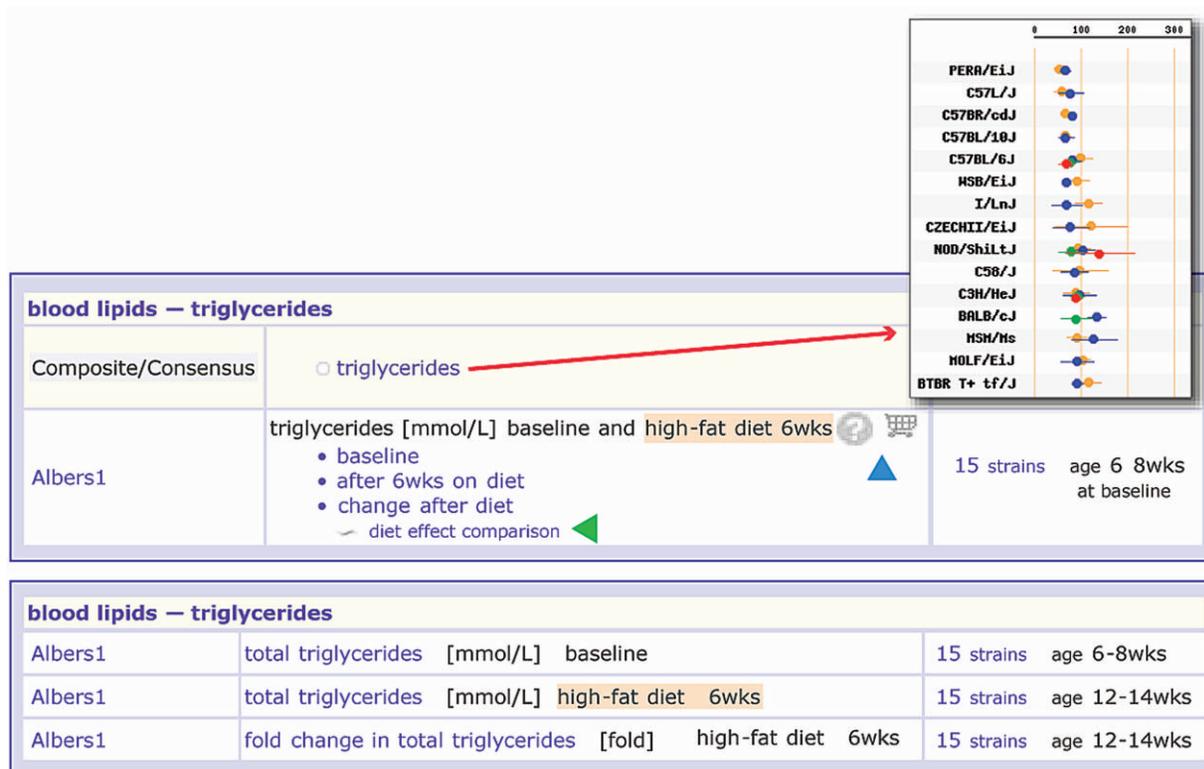
Figure 6. MPD measurement categories and using metadata to organize displays. When new MPD measurements are accessioned, they are classified based on the trait measured and experimental context. In this example, when a set of data containing three triglyceride measurements was submitted to MPD (projects are given a symbol, e.g. Albers1), each measurement was annotated (metadata) to reflect the population tested, the experimental methods (baseline vs. intervention of high fat diet for 6 weeks), and biological parameters (age). The lower panel shows the older MPD display where the metadata is included in every row. The middle panel shows the same measurements and illustrates our new method of displaying measurements based on common metadata. Although redundancy is diminished, each measurement still retains all its originally annotated metadata which is visible in other website views. The grouping display is now the default when browsing by category, but users may toggle between viewing options. The new classification scheme is amenable to adding comparison views. In this case, a plot is generated that shows a diet-effect comparison (click on link at green arrow) showing all three measurements in a single plot. Blue arrow: '?' is a quick link to the protocol and the shopping cart icon is for flagging measurements to create customized datasets, an advanced MPD feature not discussed here. The upper panel illustrates a new feature to show consensus views of related measurements across multiple projects (red arrow). The thumbnail view shows baseline triglyceride levels from four different MPD projects. Strain sets may not overlap 100% as shown here where some strains were tested by only two projects and other strains were tested by all four projects. Albers1 MPD:8 (1999).

email attachments, and all database updates are made by staff via interactive web tools or direct table updates to our development node. MPD's production node is then refreshed from the development node as needed. There is no situation where the database is directly updated by non-staff users.

AN INVITATION TO INVESTIGATORS AND FUTURE MPD DIRECTIONS

Data in all subject areas with potential relevance to translational research towards improvement of human health are of considerable importance. Although many phenotypic domains are currently represented in MPD, the acquisition of new data is open-ended with the goal of collecting data on a broader scope (and in some cases to a deeper level for phenotypes needing more granularity) as well as collecting data generated from new, more sophisticated phenotyping technologies. To expand the scope and maximize the utility of MPD, members of the global scientific community are invited to contribute their strain survey data or join us in a coordinated effort to seek funding that will support systematic strain surveys. It is this spirit of collaboration that has shaped MPD and made it an important community resource and that will continue to guide the future growth and development of MPD.

Researchers interested in contributing data to MPD or in collaborating on new phenotyping projects should contact us at phenome@jax.org. Data submission guidelines are accessible through the MPD homepage 'How to contribute data'.

MPD provides user support through online documentation and via email (phenome@jax.org). PHENOME-LIST is a moderated electronic bulletin board http://phenome.jax.org/phenome/list.html. We welcome user input and suggestions. Our Suggestion Box is accessible from most every MPD page (footer). Suggestions or comments can be submitted anonymously.
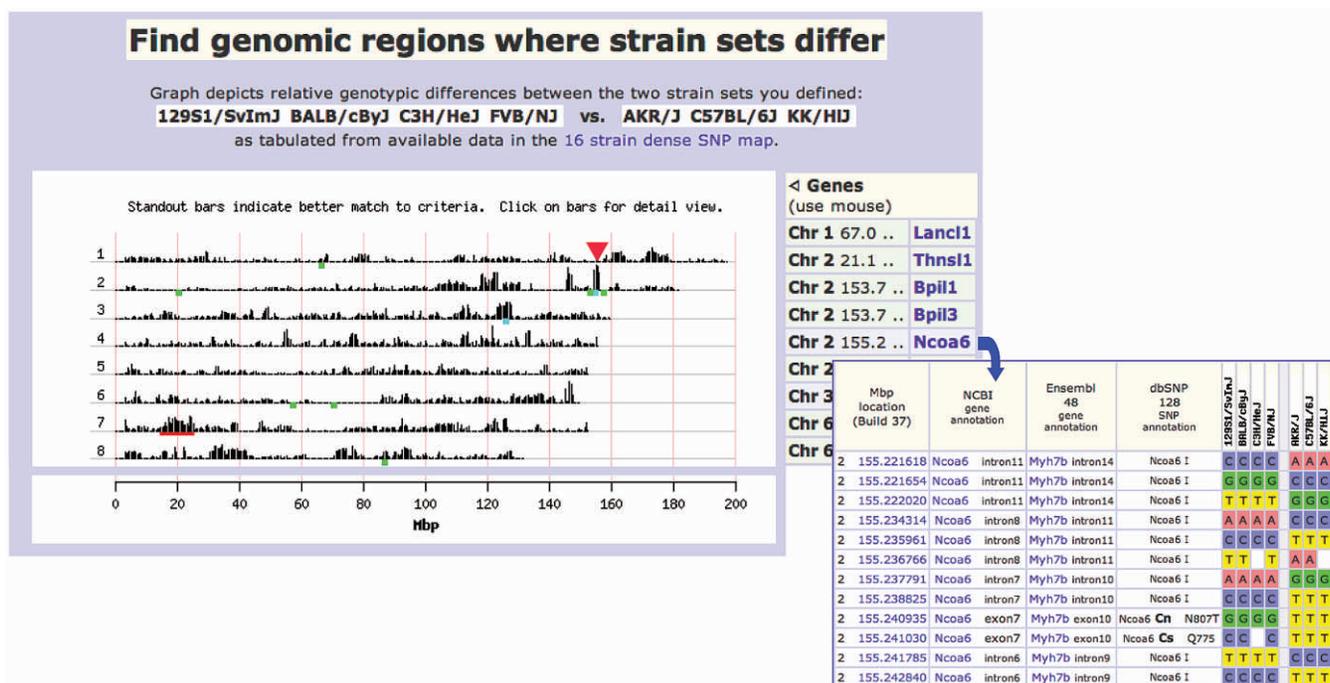
Figure 7. Find Genomic Regions. This new tool is based on the concept of identity by decent (IBD) regarding ancestral inheritance in inbred strains of mice, and on the assumption that phenotypic differences reflect genotypic differences. Therefore, finding regions of the genome that are different between strains with differing phenotypes of interest may help identify causal genes or regulatory regions contributing to the differences in phenotype. Here is an example: a measurement reveals polar phenotypes among strains so that high- and low-end outliers can be grouped (Low: 129S1/SvImJ, BALB/cByJ, C3H/HeJ, FVB/NJ; High: AKR/J, C57BL/6J, KK/HlLtJ) and entered as such in the set up window. The tool is deployed to scan the mouse genome and plot regions where the Low group is most different from the High group (top panel, truncated to show Chr 1–8 only). Genes and other regions of interest can be superimposed on the plot, including user-specified genes (blue), genomic coordinates (red), and locations where genes have been annotated (MGI) with keywords that the user enters (green). Genes and coordinates are listed to the right of the plot. The user can progressively zoom in on particular regions, all the way to listings of individual SNPs. In this example, we drilled down on the 5 Mbp interval on Chr 2 (152–157 Mbp; red arrow), and found this region contains >16K SNP locations and 139 annotated genes. Filtering our SNP retrieval by limiting it to polymorphic locations between our High and Low strain sets, we reduced the region to <3K SNPs. Several genes including Ncoa6 (lower panel) meet our criteria and might be considered good candidate genes for our phenotype. The SNP retrieval shows merged-in annotation from NCBI, Ensembl, and dbSNP. I = intron, Cs = coding synonymous (amino acid (aa) and aa position in the peptide); Cn = coding nonsynonymous (aa encoded, position, aa change).

CITING MPD

For general citation of MPD, this article may be used. In addition, the following citation format may be used when MPD projects are referred to or MPD datasets used: Investigator(s) name (year project posted) Project title. MPD accession number (MPD:XXX). Mouse Phenome Database Web Site, The Jackson Laboratory, Bar Harbor, Maine USA. World Wide Web (URL: http://www.jax.org/phenome, date of download or access). For more information visit our web site and search on 'citing'.

ACKNOWLEDGEMENTS

We thank participating investigators for contributing their data for worldwide access (http://phenome.jax.org/pub-cgi/phenome/mpdcgi?rtn=projects/list). And we thank Dale Begley, Mary Dolan and Debbie Krupke for reviewing this manuscript.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Bogue,M. (2003) Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *J. Appl. Physiol.*, **95**, 1335–1337.
2. Grubb,S.C., Churchill,G.A. and Bogue,M.A. (2004) A collaborative database of inbred mouse strain characteristics. *Bioinformatics*, **20**, 2857–2859.
3. Mouse Phenotype Database Integration Consortium. Hancock,J.M., Adams,N.C., Aidinis,V., Blake,A., Bogue,M., Brown,S.D., Chesler,E.J., Davidson,D., Duran,C., Eppig,JT. *et al.*, (2007) Mouse Phenotype Database Integration Consortium: integration [corrected] of mouse phenome data resources. *Mamm. Genome*, **18**, 157–163. (Errata in: *Mamm. Genome*. 2007; **18**, 815. *Mamm. Genome*. 2008; **19**, 219–220).
4. Taylor,C.F., Field,D., Sansone,S.A., Aerts,J., Apweiler,R., Ashburner,M., Ball,C.A., Binz,P.A., Bogue,M., Booth,T. *et al.* (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.*, **26**, 889–896.

5. Bogue,M.A., Grubb,S.C., Maddatu,T.P. and Bult,C.J. (2007) Mouse Phenome Database (MPD). *Nucleic Acids Res.*, **35**, D643–D649.
6. Svenson,K.L., Von Smith,R., Magnani,P.A., Suetin,H.R., Paigen,B., Naggert,J.K., Li,R., Churchill,G.A. and Peters,L.L. (2007) Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations. *J. Appl. Physiol.*, **102**, 2369–2378.
7. Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E., Blake,J.A. and the members of the Mouse Genome Database Group. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
8. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Edgar,R., Federhen,S. *et al.* (2007) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36**, D13–D21.
9. Hubbard,T.J., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F., Cutts,T. *et al.* (2007) *Nucleic Acids Res.*, **35**, D610–D617.
10. Kikkawa,Y., Miura,I., Takahama,S., Wakana,S., Yamazaki,Y., Moriwaki,K., Shiroishi,T. and Yonekawa,H. (2001) Microsatellite database for MSM/Ms and JF1/Ms, molossinus-derived inbred strains. *Mamm. Genome*, **12**, 750–752.
11. Cervino,A.C., Li,G., Edwards,S., Zhu,J., Laurie,C., Tokiwa,G., Lum,P.Y., Wang,S., Castellini,L.W., Lusis,A.J. *et al.* (2005) Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. *Genomics*, **86**, 505–517.
12. Goios,A., Pereira,L., Bogue,M., Macaulay,V. and Amorim,A. (2007) mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res.*, **17**, 293–298.
13. Szatkiewicz,J.P., Beane,G.L., Ding,Y., Hutchins,L., Pardo-Manuel de Villena,F. and Churchill,G.A. (2008) An imputed genotype resource for the laboratory mouse. *Mamm. Genome.*, **19**, 199–208.
14. Frazer,K.A., Eskin,E., Kang,H.M., Bogue,M.A., Hinds,D.A., Beilharz,E.J., Gupta,R.V., Montgomery,J., Morenzoni,M.M., Nilsen,G.B. *et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*, **448**, 1050–1053.
15. Mural,R.J., Adams,M.D., Myers,E.W., Smith,H.O., Miklos,G.L., Wides,R., Halpern,A., Li,P.W., Sutton,G.G., Nadeau,J. *et al.* (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science*, **296**, 1661–1671.
16. Wade,C.M. and Daly,M.J. (2005) Genetic variation in laboratory mice. *Nat. Genet.*, **37**, 1175–1180.

APPENDIX