

Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse

Cynthia L. Smith^{*}, Judith A. Blake, James A. Kadin, Joel E. Richardson, Carol J. Bult and the Mouse Genome Database Group

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

Received September 22, 2017; Revised October 11, 2017; Editorial Decision October 12, 2017; Accepted October 19, 2017

ABSTRACT

The Mouse Genome Database (MGD; <http://www.informatics.jax.org>) is the key community mouse database which supports basic, translational and computational research by providing integrated data on the genetics, genomics, and biology of the laboratory mouse. MGD serves as the source for biological reference data sets related to mouse genes, gene functions, phenotypes and disease models with an increasing emphasis on the association of these data to human biology and disease. We report here on recent enhancements to this resource, including improved access to mouse disease model and human phenotype data and enhanced relationships of mouse models to human disease.

INTRODUCTION

The laboratory mouse is used extensively as a model for investigating the etiopathogenesis of human disease. The relevance of laboratory mouse to biomedical research includes its extensive experimental genetics capabilities, fully-sequenced inbred strain genomes, published genotype to phenotype associations, and data for genome wide coverage of induced variation from mouse large-scale mutagenesis programs (1–4) and reviewed in (5). Resources such as the Collaborative Cross (6) and Diversity Outbred projects (7) capture most of the genetic variation present in laboratory mouse strains and serve as unique platforms for studies of human relevant quantitative traits and complex inherited syndromes. Furthermore, the International Mouse Phenotyping Consortium (IMPC) is generating a catalogue of gene function through systematic generation and phenotyping of a genome-wide collection of traditional gene knockout (KO) and CRISPR mutant lines in the mouse (8).

The Mouse Genome Database (MGD) is the primary community knowledgebase for mouse phenotype and gene function and mouse models of human disease (9). MGD's goal is to advance understanding of human biology and disease by facilitating access to integrated genetics and ge-

netic data for the laboratory mouse. To this end, MGD serves as an authoritative resource for the catalog of mouse genes and genome features connecting reference genomic sequence information to mouse biological data, including (i) molecular function, biological process and cellular location information encoded using the Gene Ontology (GO); (ii) a comprehensive listing of mouse mutations, variants and human disease models, with mutant genotypes annotated to Mammalian Phenotype (MP) terms and Disease Ontology (DO) terms and (iii) provide authoritative nomenclature and identifiers for mouse gene names, symbols, alleles and strains as the recognized primary community resource (Table 1). Standardization of mouse nomenclature and annotation of data with commonly used biological ontologies and standardized vocabularies ensure that data are consistently annotated, making precise data mining possible.

MGD is a core component of the Mouse Genome Informatics (MGI) consortium (<http://www.informatics.jax.org>). Other database resources coordinated through the MGI consortium include the Gene Expression Database (GXD) (10), the Mouse Tumor Biology Database (MTB) (11), the Gene Ontology project (GO) (12), MouseMine (13), the International Mouse Strain Resource (IMSR) (14) and the CrePortal database of recombinase expressing mice (www.CrePortal.org, unpublished). Data and information for these resources are obtained through a combination of expert curation of the biomedical literature and by automated or semi-automatic processing of data sets downloaded from more than fifty other data resources. Metrics of current MGD content is shown in Table 2.

In this report, we highlight several improvements to the capture, annotation, integration and presentation of data associated with mouse models of human disease. These include the incorporation of the Disease Ontology (DO), new disease detail pages and ontology browser with listings of associated genes and mouse models, improved ontology vocabulary browsers and incorporation of a Human Phenotype Ontology (HPO) browser. We added human disease to phenotype relationships from Orphanet, which expands our number of human diseases annotated to HPO

^{*}To whom correspondence should be addressed. Tel: +1 207 288 6663; Fax: +1 207 288 6830; Email: cynthia.smith@jax.org

Table 1. Data for which MGD serves as an authoritative source. In addition to providing unique IDs and symbols for genes, alleles and strains, MGD expertly curates functional, phenotype and disease model data from literature into MGD

Data type	Description
Unified mouse genome feature catalog	Catalog of integrated predictions from Ensembl, NCBI, Havana/Vega; used by NCBI, IMPC, etc.
Gene Ontology (GO) annotations for mouse	Curated from literature and integrated from others
Mouse phenotype annotations	Curated from literature, integrated with data from large scale projects
Mouse models of human disease	Curated mouse models of human disease annotated with Disease Ontology
Gene to nucleotide sequence association	Co-curation with MGA (Mouse Genome Annotation) group
Gene to protein sequence association	Co-curation with UniProt and Protein Ontology groups
Mammalian Phenotype (MP) Ontology	Developed, distributed and used by MGD; also used by RGD, IMPC, DMDD, etc.
Gene and genome feature symbols, names and IDs	Created using International Nomenclature Guidelines in coordination with human and rat nomenclature groups
Mutation symbols, names and IDs	Nomenclature and IDs for mouse mutations are assigned and provided by MGD
Mouse strain nomenclature and IDs	Created and provided by MGD; nomenclature assistance is also provided to other mouse repositories

Table 2. Summary of MGD content, September 2017

Genes and genome features with nucleotide sequence data	47 693
Genes with protein sequence data	24 317
Genes with human orthologs	17 089
Genes with rat orthologs	18 509
Genes with GO annotations	24 502
Total GO annotations	312 109
Mutant alleles in mice	51 378
Genes with mutant alleles in mice	12 401
Mouse QTL	6257
Genotypes with phenotype annotation (MP)	60 951
Total MP annotations	315 657
Mouse models (genotypes) associated with human diseases	6027
References in the MGD bibliography	237 578

terms. We now include phenotype data from Deciphering the Mechanisms of Developmental Disorders (DMDD) project. Finally, we are participating with the new Alliance of Genome Resources member groups in efforts to create a new data resource for comparative biology (<http://www.alliancegenome.org>).

NEW FEATURES AND IMPROVEMENTS

The three primary areas of improved and new functionality in MGD center on disease and phenotype annotation and the user interfaces used to deliver these annotations to the biomedical research community.

Disease Ontology (DO) is incorporated into MGD

The Disease Ontology (DO) is a community effort to provide standard terms for annotating phenotypic data (15). It is a hierarchical ontology, built on a Directed Acyclic Graph (DAG) structure, that integrates vocabularies from MeSH, ICD, NCI Thesaurus, SNOMED, UMLS, Orphanet, EFO and OMIM. Its hierarchical structure permits a range of detail from high-level, broadly descriptive terms to low-level, very specific terms. This range is useful for annotating mouse model data to the level of detail known and for searching for this information using either whole systems or specific terms as search criteria.

We have adopted the use of Disease Ontology in MGD to annotate mouse models of human disease and participated in updating and making new additions to DO. Previously we relied on disease titles from the Online Mendelian Inheritance in Man (OMIM) database (16) for annotations

of mouse models of human disease. This approach limited the scope of disease terms because of the focus of OMIM on human diseases with familial inheritance. Furthermore, the OMIM term list does not exist in a complete hierarchical structure, limiting search and retrieval efficiency. Existing MGI mouse-human disease model annotations to OMIM terms have been translated to DO terms. The new MGD Disease Ontology Browser (Figure 1) allows users to navigate the ontology and see associated genes and mouse models, either in a tree or graphical view. Links to other disease vocabularies and ontologies, including OMIM, are provided when available. The additional genes tab and models tab for terms show all data for the selected term and also for any of the more specific subclasses in the ontology. The MGI Quick Search field, Human-Mouse Disease Connection, and other advanced query forms will support searches using the DO terms or IDs.

Ontology browser redesign

Improvements to the Mammalian Phenotype (MP), Gene Ontology (GO), and the Adult Mouse Anatomy browsers include implementation of new format and search options including an autocomplete option to assist in finding relevant terms (Figure 2). Tree views include easy navigation with links to superclasses and to annotated mouse phenotype, function and expression data. Toggles in the tree structure allow expansion or collapse of particular sections of interest.

MGD has also implemented a new browser created by MGI that features the Human Phenotype Ontology (17)

A
Disease Ontology Browser

Leber congenital amaurosis (DOID:14791)
Synonyms: LCA; Leber's amaurosis; Leber's congenital amaurosis; Leber's disease
Alt IDs: MESH:D057130, OMIM:PS204000, ORDO:65, UMLS_CUI:C0339527
Definition: A retinal disease that is characterized by nystagmus, sluggish or no pupillary responses, and severe vision loss or blindness.

Term Browser
Genes (27)
Models (21)

Parent term(s)

retinal disease +

Term with siblings

Leber congenital amaurosis +

angioid streaks +
 blindness +
 brachyopsia
 choriorretinal scar +
 cone dystrophy
 hypertensive retinopathy
 macular holes
 nerve fibre bundle defect
 night blindness +
 proliferative vitreoretinopathy

Child term(s)

- Leber congenital amaurosis 1
- Leber congenital amaurosis 10
- Leber congenital amaurosis 11
- Leber congenital amaurosis 12
- Leber congenital amaurosis 13
- Leber congenital amaurosis 14
- Leber congenital amaurosis 15
- Leber congenital amaurosis 16
- Leber congenital amaurosis 17
- Leber congenital amaurosis 2
- Leber congenital amaurosis 3
- Leber congenital amaurosis 4
- Leber congenital amaurosis 5
- Leber congenital amaurosis 6
- Leber congenital amaurosis 7
- Leber congenital amaurosis 8
- Leber congenital amaurosis 9

B

Excel File | Text File | *Disease is associated /modeled with this Gene or a homolog. More...

Disease Term	Human Homologs	Mouse Homologs	Mouse Models	Homology Source
Leber congenital amaurosis 10	CEP290*	Cep290*	3 models	HomoloGene and HGNC
Leber congenital amaurosis 12	RD3*	Rd3*	1 model	HomoloGene and HGNC
Leber congenital amaurosis 14	LRAT*	Lrat*	1 model	HomoloGene and HGNC
Leber congenital amaurosis 2	RPE65*	Rpe65*	4 models	HomoloGene and HGNC
Leber congenital amaurosis 3	SPATA7*	Spat7*	1 model	HomoloGene and HGNC
Leber congenital amaurosis 4	AIPL1*	Aipl1*	3 models	HomoloGene and HGNC
Leber congenital amaurosis 5	LCA5*	Lca5*	1 model	HomoloGene and HGNC
Leber congenital amaurosis 6	RPGRIP1*	Rpgrip1*	1 model	HomoloGene and HGNC
Leber congenital amaurosis 7	CRX*	Crx*	2 models	HomoloGene and HGNC

C

Term Browser | Genes (27) | Models (21)

Excel File | Text File | All mouse models of Leber congenital amaurosis with phenotypic similarity to the human disease

Disease Term	Allelic Composition	Genetic Background	Reference	Phenotypes
Leber congenital amaurosis 2	Rpe65 ^{rd12} /Rpe65 ^{rd12}	B6(A)-Rpe65 ^{rd12} /J	J:104929	View
Leber congenital amaurosis 2	Rpe65 ^{tm1Tmr} /Rpe65 ^{tm1Tmr}	involves: 129S1/Sv * 129X1/SvJ	J:136882	View
Leber congenital amaurosis 2	Rpe65 ^{tm1Tmr} /Rpe65 ^{tm1Tmr}	involves: 129S1/Sv * 129X1/SvJ * C57BL/6	J:51293, J:61889, J:71529, J:81027, J:85796	View
Leber congenital amaurosis 2	Cnga3 ^{tm1Biel} /Cnga3 ^{tm1Biel} Rpe65 ^{tm1Tmr} /Rpe65 ^{tm1Tmr}	involves: 129S1/Sv * 129X1/SvJ * C57BL/6	J:71529	View
Leber congenital amaurosis 3	Spat7 ^{tm1Mrd} /Spat7 ^{tm1Mrd}	involves: 129S7/SvEvBrd	J:216279	View
Leber congenital amaurosis 4	Aipl1 ^{tm1Visu} /Aipl1 ^{tm1Visu}	involves: 129S1/Sv * 129X1/SvJ * C57BL/6J	J:92617	View
Leber congenital amaurosis 4	Aipl1 ^{tm1Tll} /Aipl1 ^{tm1Tll}	involves: 129S4/SvJae * C57BL/6	J:92601	View
Leber congenital amaurosis 4	Aipl1 ^{tm1Mad} /Aipl1 ^{tm1Mad}	involves: 129S7/SvEvBrd	J:94655	View
Leber congenital amaurosis 5	Lca5 ^{Gt(AG0283)Wtsi} /Lca5 ^{Gt(AG0283)Wtsi}	involves: 129P2/OlaHsd * C57BL/6	J:174018	View
Leber congenital amaurosis 6	Rpgrip1 ^{tm1Tll} /Rpgrip1 ^{tm1Tll}	involves: 129S4/SvJae * C57BL/6	J:82755	View
Leber congenital amaurosis 7	Crx ^{Rip} /Crx ⁺	C57BL/6J-Crx ^{Rip}	J:203337	View

Figure 1. Disease Ontology Browser at Mouse Genome Informatics. Example of a new disease detail page and associated mouse and human data available for Leber congenital amaurosis. (A) Shown are the Disease Ontology (DO) term, definition, synonyms and IDs for this DO term. Additional matching terms at other resources are shown and linked when available at another resource. Graphical views of the ontology are also available. (B) The genes tab of the browser lists all human and mouse orthologs associated with the disease or subclasses of the disease. Mouse models are listed, if available. (C) The models tab lists the disease and genotype of mouse models of the disease along with supporting reference information and links to phenotype data for each model.

The screenshot displays the Gene Ontology Browser interface. On the left, the 'GO Search' section shows a search for 'cell adhesion' with 50+ terms listed, including 'cell adhesion molecule binding', 'cell adhesion involved in single-species biofilm formation in or on host organism', and 'cell adhesion mediated by integrin'. The main 'GO Term Detail' section for 'chorio-allantoic fusion' (GO:0060710) provides a definition: 'The cell-cell adhesion process in which the cells of the chorion fuse to the cells of the allantois.' It lists parent terms: 'is-a cell-cell adhesion', 'part-of labyrinthine layer morphogenesis', and 'is-a reproductive process'. The category is 'Biological Process'. The 'GO Tree View' on the right shows a hierarchical list of terms, with 'chorio-allantoic fusion' highlighted in blue and noted as having 10 genes and 10 annotations. Other terms in the tree include 'cell-cell adhesion involved in sealing an epithelial fold', 'cell-cell adhesion mediated by cadherin', and 'epithelial cell-cell adhesion'.

Figure 2. Gene Ontology Browser. The new MGD ontology browsers have an autocomplete search feature and display linked matching terms. Shown here is an example from the Gene Ontology Browser. GO term names, definitions, IDs, superclasses (parent terms) and term relations are listed and terms are shown in a tree view. The example shown here is a search for 'cell adhesion' showing all terms containing this text string. The subclass 'chorio-allantoic fusion' is selected. Different branches may be viewed by clicking on the different parent terms under the definition. Each biological term is connected to annotated genes via a link in the tree view following the term name; in this example, ten different genes are annotated to this biological process term. Other terms with subclasses are expanded by clicking on the triangle toggle to the left of the term name.

developed by Peter Robinson and colleagues (http://www.informatics.jax.org/vocab/hp_ontology). Users can either browse or search the HPO browser to view terms, definitions and links to disease detail pages that are associated with the phenotypic feature at the Human-Mouse Disease Connection (HMDC) portal.

Improvements to the Human-Mouse Disease Connection (HMDC)

The Human-Mouse Disease Connection (HMDC, <http://www.disease-model.org>) is a translational tool designed for exploring and comparing human and mouse phenotypes and their associations with known human diseases. It also provides rapid access to mouse model resources and supporting references. Searches can be initiated based on human or mouse data using one or more parameters, including genes, genomic locations, phenotypes and diseases. New features and data include searches by Disease Ontology terms to group and display disease classes, and the incorporation of human disease and phenotype relationships from

Orphanet (<http://www.orpha.net>). These new data sets add to the existing OMIM disease-to-phenotype data from the HPO project (17) implemented previously.

Inclusion of Deciphering the Mechanisms of Developmental Disorders (DMDD) data

MGD now integrates mouse embryonic mutant phenotype data generated by the Deciphering the Mechanisms of Developmental Disorders (DMDD) project (18). MGD currently has information for 63 mouse lines, and more phenotype data for ~200 additional lines will be added as they become available. All of the DMDD mouse lines are derived from International Knockout Mouse Consortium (IKMC)'s Knockout Mouse Project (KOMP) or European Conditional Mouse Mutagenesis Program (EUCOMM) ES cells (19) or CRISPR-induced lines (20). Mouse mutations are annotated to MP terms by DMDD and are shown in parallel with other published and submitted data on these mutations. Expression data for these lines are also available from the Gene Expression Database (GXD) at MGI (10).

Affected Systems <small>show or hide all annotated terms</small>	hm1		hm2	hm3			
	Sex: ♀		♂	♀			
	Source: ?	IMPC - WTSI	DMDD	IMPC - WTSI	DMDD	MGI	IMPC - WTSI
adipose tissue		✓		✓			
behavior/neurological		✓					✓
cardiovascular system			✓		✓		
craniofacial		✓					
embryo		✓		✓			
endocrine/exocrine glands		✓		✓			
growth/size/body		✓		✓			
hematopoietic system							✓
homeostasis/metabolism		✓	✓	✓	✓	✓	✓
decreased fasted circulating glucose level		✓		✓			
increased circulating alkaline phosphatase level		✓		✓			
retropleural edema			✓		✓		
subcutaneous edema			✓		✓		
abnormal mineral level							✓
increased circulating magnesium level		✓		✓			
decreased circulating phosphate level							✓
immune system							✓
liver/biliary system		✓		✓			
neoplasm		✓		✓			
nervous system		✓		✓			✓
skeleton		✓	✓	✓	✓		
vision/eye		✓		✓			

Figure 3. Phenotype systems grid for *Slc20a2^{tm1a(EUCOMM)Wtsi}*. Shown is Phenotype systems grid from the allele detail page for the *Slc20a2^{tm1a(EUCOMM)Wtsi}* allele featuring DMDD data integrated with IMPC data and published data curated by MGI. Blue triangular toggles to the right of the system name may be opened for comparison of data details across genotypes. Further phenotype details including detailed phenotype terms and references may be accessed by clicking genotype links (e.g. 'hm1') or checkmarks on the phenotypes grid. Other included information on the allele detail page for this mutation (not shown) are links to the gene detail page, mutation origin and description, human disease information, links to repositories carrying the mouse mutation and references.

Shown in Figure 3 is an example of DMDD data submitted for the *Slc20a2^{tm1a(EUCOMM)Wtsi}* mutation together with data submitted by IMPC and from published literature curated at MGI.

Other enhancements

In addition to the major enhancements described above, several minor enhancements were implemented. Researchers can now search for genome features that are still on unlocalized and unplaced contigs for the reference genome assembly (GRCm38). MGD data loads from IMPC were modified to load and integrate phenotype data generated from CRISPR mutations in addition to ES cell line knockout mutations, increasing the number of annotated phenotype-genotype data in MGD. MGD also updated the load of GO annotations from the Go consortium site to use the new Gene Product Association Data (GPAD, [http://www.genontology.org/page/gene-](http://www.genontology.org/page/gene-product-association-data-gpad-format)

[product-association-data-gpad-format](http://www.genontology.org/page/gene-product-association-data-gpad-format)) annotation file format which supports inclusion of additional metadata in contrast to the previously used Gene Association File (GAF).

MGD and the Alliance of Genome Resources

MGD is one of the founding members of the Alliance of Genome Resources (AGR), a new data resource integration effort among the major model organism database (MOD) groups and the Gene Ontology Consortium (GOC). The founding members of the AGR are: the Gene Ontology Consortium, Mouse Genome Database (MGD), FlyBase, WormBase, Saccharomyces Genome Database (SGD), Rat Genome Database (RGD), Zebrafish Information Network (ZFIN). The AGR will work to standardize access to and display of common data types from different model organisms to better support comparative biology for biomedical researchers. The formation of the AGR builds on the collaborative activities between the MODs and GO over several years seeking to enhance data integration, exchange, and the use of common data standards. Now these groups will merge key activities and data representations, coordinating data retrieval and analysis within the comparative perspective. The initial release of the public web portal for the AGR (<http://www.alliancegenome.org>) is scheduled for October 2017. Among the data types to be included in the initial release are gene details such as gene name and symbol, genomic location, orthology, function annotations, and disease associations. Longer-term goals of the Alliance include adding other model organisms, data types, and analysis tools within a common shared infrastructure.

IMPLEMENTATION AND PUBLIC ACCESS

The primary MGI database ('production') is a highly normalized relational database designed and optimized for data integration and incremental updating. This database is the locus of data loads and ongoing expert data curation. It resides in a PostgreSQL server behind a firewall and is not accessible by the public. In contrast, our public web interface is backed by a combination of a highly unnormalized databases (also in PostgreSQL) and Solr/Lucene indexes, designed for high performance query and display in a read-only environment. The front end data stores are refreshed from the production master on a weekly basis. The separation between the public and production (private) architectures provides a large measure of flexibility in project planning, as either side can (and often does) change without affecting the other.

MGD broadcasts data in a variety of ways to support basic research communities, clinical researchers and advanced users interested in programmatic or bulk access. MGD provides free public web access to data from <http://www.informatics.jax.org>. The web interface provides a simple 'Quick Search', available from all web pages in the system and is the most used entry point for users. Various advanced query forms are provided to support precise parameter searching. Data may be retrieved from most results pages by downloading text or Excel files, or forwarding results to Batch Query or MouseMine analysis tools (see below).

MGD offers batch querying interfaces for data retrieval for users wishing to retrieve data in bulk. The Batch Query tool (<http://www.informatics.jax.org/batch>) (21) is used for retrieving bulk data about lists of genome features. Feature identifiers can be typed in or uploaded from a file. Gene IDs from MGI, NCBI GENE, Ensembl, VEGA, UniProt and other resources can be used. Users can choose the information set they wish to retrieve, such as genome location, GO annotations, list of mutant alleles, MP annotations, Reference SNP IDs or Disease Ontology (DO) terms. Results are returned as a web display, or in tab delimited text or Excel format. Results may also be forwarded to MouseMine (see below).

MGD data access is available through MouseMine (<http://www.mousemine.org>) (13), an instance of InterMine that offers flexible querying, templates, iterative querying of results and linking to other model organism InterMine instances. MouseMine access is also available via a RESTful API, with client libraries in Perl, Python, Ruby, Java and JavaScript.

This year, MGD retired its FTP server and other back end upgrades. The files that were accessible via FTP are now downloadable from <http://www.informatics.jax.org/downloads/>. MGD provides a large set of regularly updated database reports from this site. Direct SQL access to a read-only copy of the database is also offered (contact MGI user support for an account). MGI User Support is also available to assist users in generating customized reports on request.

OUTREACH

MGD User Support staff are available for on-site help and training on the use of MGD and other MGI data resources as well as providing off-site workshop/tutorial programs (roadshows) that include lectures, demos and hands-on tutorials. To inquire about hosting an MGD roadshow, email mgi-help@jax.org.

Members of the MGD User Support team can be contacted via email, web requests, phone or fax.

- World wide web: http://www.informatics.jax.org/mgihome/support/mgi_inbox.shtml
- Facebook: <https://www.facebook.com/mgi.informatics>
- Twitter: https://twitter.com/mgi_mouse and https://twitter.com/hmdc_mgi
- Email access: mgi-help@jax.org
- Telephone access: +1 207 288 6445.
- Fax access: +1 207 288 6830.

CITING MGD

For a general citation of the MGI resource, researchers should cite this article. In addition, the following citation format is suggested when referring to datasets specific to the MGD component of MGI: mouse genome database (MGD), MGI, The Jackson Laboratory, Bar Harbor, Maine (URL: <http://www.informatics.jax.org>). Type in date (month, year) when you retrieved the data cited.

MOUSE GENOME DATABASE GROUP

A. Anagnostopoulos, A. Andrews, R.M. Baldarelli, J.S. Beal, S.M. Bello, O. Blodgett, N.E. Butler, K. Christie, L.E. Corbani, H.J. Drabkin, R. Espinoza, J. Franco, S.L. Gianatto, P. Hale, D.P. Hill, L. Hutchins, M. Law, J.R. Lewis, M. McAndrews, N. Mez, D. Miers, H. Motenko, L. Ni, H. Onda, M. Perry, J.M. Recla, D.J. Reed, B. Richards-Smith, D. Sitnikov, M. Tomczuk, L. Wilming and Y. Zhu.

ACKNOWLEDGEMENTS

We are grateful to our collaborators at Disease Ontology (DO) and Rat Genome Database (RGD) for providing updated disease ontology files and help with revisions and additions to the ontology. The use of Human Phenotype Ontology and Human Phenotype Annotation files from Dr Peter Robinson (The Jackson Laboratory) is gratefully acknowledged.

FUNDING

National Institutes of Health (NIH)/National Human Genome Research Institute (NHGRI) [U41 HG000330, R25 HG007053 to M.G.D., U41 HG002223 to Alliance for Genome Resources]. Funding for open access charge: NIH/NHGRI [U41 HG000330].

Conflict of interest statement. None declared.

REFERENCES

1. Brown, F.C., Scott, N., Rank, G., Collinge, J.E., Vadolas, J., Vickaryous, N., Whitelaw, N., Whitelaw, E., Kile, B.T., Jane, S.M. *et al.* (2013) ENU mutagenesis identifies the first mouse mutants reproducing human β -thalassemia at the genomic level. *Blood Cells Mol. Dis.*, **50**, 86–92.
2. Li, Y., Klena, N.T., Gabriel, G.C., Liu, X., Kim, A.J., Lemke, K., Chen, Y., Chatterjee, B., Devine, W., Damerla, R.R. *et al.* (2015) Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*, **521**, 520–524.
3. Ha, S., Stottmann, R.W., Furley, A.J. and Beier, D.R. (2015) A forward genetic screen in mice identifies mutants with abnormal cortical patterning. *Cereb. Cortex*, **25**, 167–179.
4. Daxinger, L., Harten, S.K., Oey, H., Epp, T., Isbel, L., Huang, E., Whitelaw, N., Apedaile, A., Sorolla, A., Yong, J. *et al.* (2013) An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biol.*, **14**, R96.
5. Smith, C.L. and Eppig, J.T. (2012) The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm. Genome*, **23**, 653–668.
6. Threadgill, D.W. and Churchill, G.A. (2012) Ten years of the collaborative cross. *Genetics*, **190**, 291–294.
7. Churchill, G.A., Gatti, D.M., Munger, S.C. and Svenson, K.L. (2012) The Diversity Outbred mouse population. *Mamm. Genome*, **23**, 713–718.
8. Ring, N., Meehan, T.F., Blake, A., Brown, J., Chen, C.K., Conte, N., Di Fenza, A., Fiegel, T., Horner, N., Jacobsen, J.O.B. *et al.* (2015) A mouse informatics platform for phenotypic and translational discovery. *Mamm. Genome*, **26**, 413–421.
9. Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J., Anagnostopoulos, A., Baldarelli, R.M., Beal, J.S., Bello, S.M. *et al.* (2017) Mouse Genome Database (MGD)-2017: Community knowledge resource for the laboratory mouse. *Nucleic Acids Res.*, **45**, D723–D729.
10. Finger, J.H., Smith, C.M., Hayamizu, T.F., McCright, I.J., Xu, J., Law, M., Shaw, D.R., Baldarelli, R.M., Beal, J.S., Blodgett, O. *et al.* (2017) The mouse Gene Expression Database (GXD): 2017 update. *Nucleic Acids Res.*, **45**, D730–D736.

11. Bult,C.J., Krupke,D.M., Begley,D.A., Richardson,J.E., Neuhauser,S.B., Sundberg,J.P. and Eppig,J.T. (2015) Mouse Tumor Biology (MTB): a database of mouse models for human cancer. *Nucleic Acids Res.*, **43**, D818–D824.
12. Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
13. Motenko,H., Neuhauser,S.B., O’Keefe,M. and Richardson,J.E. (2015) MouseMine: a new data warehouse for MGI. *Mamm. Genome*, **26**, 325–330.
14. Eppig,J.T., Motenko,H., Richardson,J.E., Richards-Smith,B. and Smith,C.L. (2015) The International Mouse Strain Resource (IMSR): cataloging worldwide mouse and ES cell line resources. *Mamm. Genome*, **26**, 448–455.
15. Schriml,L.M. and Mitraka,E. (2015) The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mamm. Genome*, **26**, 584–589.
16. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
17. Köhler,S., Vasilevsky,N.A., Engelstad,M., Foster,E., McMurry,J., Aymé,S., Baynam,G., Bello,S.M., Boerkoel,C.F., Boycott,K.M. *et al.* (2017) The human phenotype ontology in 2017. *Nucleic Acids Res.*, **45**, D865–D876.
18. Mohun,T., Adams,D.J., Baldock,R., Bhattacharya,S., Copp,A.J., Hemberger,M., Houart,C., Hurles,M.E., Robertson,E., Smith,J.C. *et al.* (2013) Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Dis. Model Mech.*, **6**, 562–566.
19. Bradley,A., Anastassiadis,K., Ayadi,A., Battey,J.F., Bell,C., Birling,M.C., Bottomley,J., Brown,S.D., Bürger,A., Bult,C.J. *et al.* (2012) The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome*, **23**, 580–586.
20. Rosen,B., Schick,J. and Wurst,W. (2015) Beyond knockouts: the International Knockout Mouse Consortium delivers modular and evolving tools for investigating mammalian genes. *Mamm. Genome*, **26**, 456–466.
21. Eppig,J.T., Blake,J.A., Bult,C.J., Kadin,J.A., Richardson,J.E., Anagnostopoulos,A., Babiuk,R.P., Baldarelli,R.M., Beal,J.S., Bello,S.M. *et al.* (2015) The Mouse Genome Database (MGD): Facilitating mouse as a model for human biology and disease. *Nucleic Acids Res.*, **43**, D726–D736.