The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data

Cynthia L. Smith · Janan T. Eppig

Abstract The Mammalian Phenotype Ontology (MP) is a structured vocabulary for describing mammalian phenotypes and serves as a critical tool for efficient annotation and comprehensive retrieval of phenotype data. Importantly, the ontology contains broad and specific terms, facilitating annotation of data from initial observations or screens and detailed data from subsequent experimental research. Using the ontology structure, data are retrieved inclusively, i.e., data annotated to chosen terms and to terms subordinate in the hierarchy. Thus, searching for “abnormal craniofacial morphology” also returns annotations to “megacephaly” and “microcephaly,” more specific terms in the hierarchy path. The development and refinement of the MP is ongoing, with new terms and modifications to its organization undergoing continuous assessment as users and expert reviewers propose expansions and revisions. A wealth of phenotype data on mouse mutations and variants annotated to the MP already exists in the Mouse Genome Informatics database. These data, along with data curated to the MP by many mouse mutagenesis programs and mouse repositories, provide a platform for comparative analyses and correlative discoveries. The MP provides a standard underpinning to mouse phenotype descriptions for existing and future experimental and large-scale phenotyping projects. In this review we describe the MP as it presently exists, its application to phenotype annotations, the relationship of the MP to other ontologies, and the integration of the MP within large-scale phenotyping projects. Finally we discuss future application of the MP in providing standard descriptors of the phenotype pipeline test results from the International Mouse Phenotype Consortium projects.

Introduction

Systematic collection and curation of phenotypic descriptions began in the 1940s as textual synopses for mouse mutants (Snell 1941). Then, maintaining and updating text manually was easy (very few mutants known) and there were no electronic records requiring search mechanisms or computer-parsable formats. Now there are over 24,300 mutant alleles that have been identified in mice with established phenotypes, representing mutant alleles in over 9,600 genes (Table 1). Also, there are nearly 4,700 QTL (quantitative trait loci) that represent genomic regions associated with particular phenotypic traits. Phenotypic characterization data continue to expand rapidly. The Mouse Genome Informatics (MGI http://www.informatics.jax.org) (Blake et al. 2011; Eppig et al. 2012) database, the primary international database for mouse, adds several hundred new mutant alleles with reported phenotypes to the database each month.

Researchers increasingly develop sophisticated new mouse models of human disease and analyze phenotypes in mice carrying complex engineered and mutant allele combinations on multiple genetic backgrounds. The unit of annotation for a phenotype, therefore, must be the animal(s) or “whole” genotype assessed, annotated to MP terms and accompanied by key conditional variables (e.g.,

---

1 This number refers only to mutant alleles that are “in mice.” The number of alleles known exceeds 730,000 when also counting those alleles that exist only as targeted or gene-trapped mutations in ES cell lines.
Auwerx et al. 2004; Nord et al. 2006). These new data sets (Clark et al. 2004; Cook et al. 2006; Goldowitz et al. 2004) as the corner-stone for phenotype annotation; (c) application of the MP ontology to ongoing curation of phenotypes in MGI and the retirement of text-based descriptions; (d) development of new, robust access to phenotypes via redesigned web interfaces, tracks on the MGI Mouse Genome Browser (http://gbrowse.informatics.jax.org/cgi-bin/gbrowse/mouse_current), and contribution of data to other genome browser resources such as University of California Santa Cruz (UCSC, http://genome.ucsc.edu), Ensembl (http://www.ensembl.org) and NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov); and (e) development of a human disease view of mouse phenotypes utilizing MGI’s annotations to OMIM (Online Mendelian Inheritance in Man, http://www.omim.org) disease terms.

The significance of the mouse as a model organism, the availability of its fully sequenced genome, and the accessibility of mouse tissues for experimentation at all life-stages invite new applications and exquisite experimental manipulation to address key scientific questions. Integration of experimental data using standard data descriptions and nomenclatures is of paramount importance in maximizing the value of the mouse model system. MGI has loaded large-scale data from ENU phenotyping centers and laboratories and is poised to load new phenotyping data from the developing International Mouse Phenotyping Consortium (IMPC, www.mousephenotype.org), allowing integration of these data with data derived from individual laboratories and the biomedical literature. A common interface to phenotypes in MGI exists that allows critical phenotype and genotype comparisons. Further, alignment of mouse mutant phenotypes with human disease symptoms will aid in identifying mouse genetic models with phenotypic matches as well as the currently captured experimentally demonstrated mouse models for human disease.

### The Mammalian Phenotype Ontology (MP)

The MP is the workhorse for standardizing phenotypic descriptions in mouse, rat, and other mammals. The MP is a “precomposed” ontology, structured as a DAG (directed acyclic graph) and using phenotype terms recognized by research biologists and clinicians that include simple compound concepts (e.g., liver hyperplasia, MP:0005141) and aggregate concepts (e.g., glomerular crescent, MP:0011506) (Fig. 1).

The MP is a flexible, expandable tool that can grow to accommodate the anticipated rapid increase in phenotyping data, can be applied to maximize precision and breadth of user phenotype searches, and can facilitate an efficient curation stream of incoming phenotype data. By annotating phenotypes from these data sets using MP, the standardization and concurrent retrieval of terms is achieved. This stands in contrast to natural language text, where there is

| Data type                                                        | Count  |
|-----------------------------------------------------------------|--------|
| Total mutant alleles (in ES cell lines and mice)                | 738,364|
| Mutant alleles in mice                                          | 24,339 |
| Genes with mutant alleles                                       | 14,743 |
| Genes with mutant alleles in mice                               | 9,636  |
| Mammalian phenotype ontology (MP) terms                         | 8,744  |
| Genes with phenotype annotations                                | 8,903  |
| Genotypes with phenotype annotations                            | 43,335 |
| Total MP annotations to genotypes                               | 227,169|
| Human diseases with one or more genotypic mouse models          | 1,148  |
| Mouse genotypes modeling human diseases                         | 3,668  |
| Quantitative trait loci (QTL)                                   | 4,696  |
| Total recombinase (Cre)-expressing transgenes and alleles       | 1,739  |

\( ^a \) Data as of May 5, 2012, [www.informatics.jax.org](http://www.informatics.jax.org). New data are added to the MGI database daily; thus, actual counts will be higher than those shown here.

\( ^b \) Mutant allele counts include spontaneous, induced (e.g., by ENU), and genetically engineered alleles. Transgenes, which are not part of the normal mouse genome, are not included.

\( ^c \) Mutants present only in ES cell lines versus those created in mice or made into mice from ES cells are distinguished in several table counts. All phenotype-related data refer to mutations present in mice.

Table 1 Mutant allele and phenotype data in the Mouse Genome Informatics database, MGI\(^{a,b,c}\)

---

© Springer
As of May 2012, the MP contains 8,744 terms describing morphological, physiological, and behavior anomalies. The top nodes are organized into 27 categories representing biological systems, mortality terms, and behavior, with abnormal morphological and physiological system terms at the next node level. Phenotype data can be annotated at any point along the structure, depending on the detail available from information sources. Each term is distinct and defined, aiding both curators and users in selecting the appropriate term for their needs. In addition, attributes and relationships among the terms are described in the form of a DAG (Fig. 1). This allows more flexibility than that of a simple tree, since each term can have multiple relationships to broader parent terms and more specific child terms. The more specific terms are subsumed by parent terms as one moves up the graph, which allows for more complete grouping, searching, and analysis of annotated data.

Multiple resources provide browser formats for viewing the MP, including the Ontology Lookup Service (OLS, http://www.ebi.ac.uk/ontology-lookup/ontologyList.do), Bioportal (http://bioportal.bioontology.org/ontologies), and MGI’s MP browser (http://www.informatics.jax.org/searches/MP_form.shtml). Figure 1 shows a sample page from MGI’s MP Browser for the phenotype term ventricular septal defect (MP:0010402). Each term in the MP browser is followed by a link to MGI genotypes associated with that term or any children of that term. Right A graphical representation of the DAG structure for the term “ventricular septal defect.” The two alternate paths from the node “cardiovascular system phenotype” are shown.
hierarchy, thus effectively flattening the DAG structure for web viewing. The MP file in OBO format is available for download from the MGI ftp site (ftp://ftp.informatics.jax.org/pub/reports/index.html#pheno); it is also available in OBO and OWL formats from the Open Biomedical Ontologies (OBO, http://www.obofoundry.org) foundry site, OLS, and Bioportal.

The MP is a dynamic ontology, actively used and developed by those annotating phenotypes in mouse and other species. Requests for new terms, term revisions, and suggestions for structural organization modifications to the MP are frequently proposed by curators and user groups. Suggestions for improvement and additions from the community are submitted through the Open Biomedical Ontologies Mammalian Phenotype Requests tracker system at SourceForge (https://sourceforge.net/tracker/?atid=1109502&group_id=76834) or by email to pheno@jax.org.

Expansion of the MP ontology and review of its hierarchical structure occurs in collaboration with new phenotype annotation projects when the need for additional granularity of terms is anticipated. In addition, collaborative review of particular systems by expert editors together with subject area specialists helps create terms and structures that are intuitive and useful to those communities. Recent additions and revisions include the respiratory system, renal/urinary system, and cardiovascular system (with significant structural reorganization) that expanded the MP by 714 terms. To accommodate data being generated by large-scale phenotyping efforts at the Wellcome Trust Sanger Institute (hereafter, Sanger Institute) Mouse Genetics Program (http://www.sanger.ac.uk/mouseportal) and from the EUMORPHIA (Brown et al. 2005; Mandillo et al. 2008) and EUMODIC (Beck et al. 2009; Morgan et al. 2010) European large-scale phenotyping efforts, MP added 38 new population-level lethality terms. These lethality terms also will support data forthcoming from the IMPC projects. Furthermore, 196 new MP homeostasis terms now describe the results of phenotype pipeline tests generated by these centers. When new MP terms are added or revised from these annotation projects or from user requests, relevant existing phenotype annotations at MGI are triggered for review and revised to reflect the new terminology as appropriate.

Along with cardiovascular system term revisions, Fyler codes (Keane et al. 2006), a systematic, hierarchical classification of congenital heart disease (see example in Fig. 1), were included as secondary IDs to the primary MP ID. Fyler codes align the MP to current standards of the cardiac disease research community and its representation in the research and clinical literature. These codes are consistent with the International Pediatric and Congenital Cardiac Codes (IPCCC, http://www.ipccc.net) and enable users to search for congenital heart defects using these codes, IDs, or term names, with comprehensive retrieval of information.

### Applying the MP to phenotype annotations

A number of resources use the MP to describe abnormal phenotypes (see Table 2), including MGI, the Rat Genome Database (RGD, http://rgd.mcw.edu), Online Mendelian Inheritance in Animals (OMIA, http://omia.angis.org.au/home), the Sanger Institute Mouse Genetics Program, MRC Harwell’s MouseBook (http://www.mousebook.org), Europhenome (http://www.europhenome.org), and the IMPC, among others. In addition, the MP is used by mouse repositories to annotate phenotype data (or reflect downloaded MGI phenotype data) for describing available mouse strains and stocks. These include the Jackson Laboratory Repository (JAX® Mice, http://jaxmice.jax.org), the European Mouse Mutant Archive (EMMA, http://www.emmanet.org), and the Mutant Mouse Regional Resource Centers (MMRRC, http://www.mmrrc.org), among others.

MGI contains information on published spontaneous, induced, and genetically engineered mouse mutations (Table 1), as well as contributed or downloaded data from large-scale mouse mutagenesis projects, including ENU, gene trap, and knockout mutagenesis projects (see Table 3 for a list of mutagenesis projects with data integrated into MGI). All of these data are integrated with all of the other genomic, expression, function, tumor and pathway data in MGI to facilitate knowledge discovery and hypothesis building. The MGI website presents a mutant allele in all of its studied contexts, which is key for discerning multigenic disease models, genetic background effects, and allelic interactions. Tools also are available at MGI for users who wish to separate genotypes carrying single-gene mutations from more complex genotypes such as conditional genotypes or those carrying transgenes, mutations in multiple genes, or large genomic rearrangements.

Figure 2 shows an example of MP annotations to genotypes involving the Fgfr2<sup>gmi/scl</sup> mutant allele (a mutation in the fibroblast growth factor receptor 2 gene, the first targeted mutation of this gene from the laboratory of Joseph Schlessinger). Phenotype data are viewed in a matrix summary format to facilitate comparison of multiple genotypes and genetic backgrounds or by genotype. Clicking the links in these sections leads to an expanded view, including terms and additional details organized by physiological system, as well as mouse disease model data and images.

Phenotype data associated with specific terms are retrieved in a variety of ways. A MP term or ID entered in the Quick Search box on any MGI page will retrieve a list of genes, alleles, and vocabulary terms. MP terms entered in the Phenotype/Human Disease section of the advanced Genes
| Resource                                      | URL                                      |
|----------------------------------------------|------------------------------------------|
| *AnnotQTL*                                   | http://annotqtl.genouest.org             |
| Biportal                                      | http://bioportal.bioontology.org/ontologies |
| Chemical Entities of Biological Interest (ChEBI) | http://www.ebi.ac.uk/chebi             |
| Collaborative Cross (CC)                      | http://csbio.unc.edu/CCstatus           |
| Disease Ontology (DO)                         | http://disease-ontology.org             |
| Diversity Outcross (DO)                       | http://cgdjax.org/datasets/phenotype/SvensonDO.shtml |
| Drosophila Genome Database (FlyBase)          | http://flybase.org                      |
| Ensembl                                       | http://www.ensembl.org                  |
| *European Mouse Mutant Archive (EMMA)*        | http://www.emmanet.org                  |
| *Europhenome*                                | http://www.europhenome.org              |
| Foundational Model of Anatomy (FMA)           | http://sig.biosstr.washington.edu/projects/fm/AboutFM.html |
| Gene Ontology (GO)                            | http://www.geneontology.org            |
| *Gene Weaver*                                | http://www.GeneWeaver.org               |
| Getting an Understanding of LOGical definitions (GULO) | http://compbio.charite.de/svn/hpo/trunk/src/tools/gulo |
| Human Phenotype Ontology (HP)                 | http://www.human-phenotype-ontology.org |
| International Classification of Diseases (ICD) | http://www.cdc.gov/nchs/icd.htm   |
| International Knockout Mouse Consortium (IKMC) | http://www.knockoutmouse.org          |
| *International Mouse Phenotyping Consortium (IMPC) | http://www.mousephenotype.org        |
| International Pediatric and Congenital Cardiac Code (IPCCC) | http://www.ipccc.net |
| *Jackson Laboratory Mouse Repository (JAX Mice) | http://jaxmice.jax.org                |
| KEGG Pathways Database                        | http://www.genome.jp/kegg/pathway.html    |
| *Mammalian Phenotype Enrichment Analysis (MamPhEA)* | http://evol.nhri.org.tw/phenome/index.jsp?platform=mmus |
| *Mammalian Phenotype Ontology (MP) browser    | http://www.informatics.jax.org/searches/MP_form.shtml |
| *Mammalian Phenotype Ontology download from MGI site | ftp://ftp.informatics.jax.org/pub/reports/index.html#pheno |
| Mammalian Phenotype Requests Tracker          | https://sourceforge.net/tracker/?atid=1109502&group_id=76834 |
| MEDIC disease vocabulary                      | http://ctdbase.org/voc.go?type=disease   |
| Medical Subject Headings (MeSH)               | http://www.nlm.nih.gov/mesh/MBrowser.html |
| *MGI Mouse Genome Browser (Mouse GBrowse)     | http://gbrowse.informatics.jax.org/cgi-bin/gbrowse/mouse_current |
| *MouseFinder*                                 | http://www.mousemodels.org              |
| *MRC Harwell MouseBook                        | http://www.mousebook.org                |
| *Mutant Mouse Regional Resource Centers (MMRRC) | http://www.mmrrc.org          |
| National Center for Biotechnology Information (NCBI) | http://www.ncbi.nlm.nih.gov |
| *Online Mendelian Inheritance in Animals (OMIA) | http://omnia.angis.org.au/home |
| Online Mendelian Inheritance in Man (OMIM)    | http://www.omim.org                     |
| Ontology Lookup Service (OLS)                 | http://www.obi.ac.uk/ontology-lookup/ontologyList.do |
| Open Biomedical Ontologies (OBO)              | http://www.obofoundry.org               |
| Orphanet                                      | http://www.orpha.net/consor/cgi-bin     |
| *PhenoHM*                                     | http://phenome.cchmc.org/PhenoBrowser/Phenome |
| *PhenomeNet*                                  | http://phenomebrowser.net               |
| *PhenomicDB*                                  | http://www.phenomicdb.de                |
| Phenotype and Trait Ontology (PATO)            | http://code.google.com/p/pato           |
| *Rat Genome Database (RGD)                    | http://rgd.mcw.edu                      |
| Reactome                                      | http://www.reactome.org                 |
| *Sanger Institute Mouse Resources Portal      | http://www.sanger.ac.uk/mouseportal    |
| Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) | http://www.ihtsdo.org/snomed-ct     |
| *ToppGene: Candidate gene prioritization      | http://toppgene.cchmc.org/prioritization.jsp |
| Uberon, cross-species anatomy ontology        | http://obofoundry.org/wiki/index.php/UBERON:Main_Page |
| *UCSC genome browser                          | http://genome.ucsc.edu                   |

*Denotes resources that incorporate Mammalian Phenotype (MP) terminology for phenotype data.
and Markers Query Form or Phenotypes Query Form return genes or alleles associated with genotypes annotated to that term. Selecting terms in the MP Browser displays links at the term level to genotypes annotated to that term or any child of that term. For example, a search in MGI using either the identifier “MP:0010402” or the term name “ventricular septal defect” returns a list of 300 genotypes with 337 annotations representing 311 matching alleles in 232 genes, transgenes, and markers (Fig. 3). These results include annotations to terms listed below “ventricular septal defect” in the hierarchy such as “inlet ventricular septal defect” and “perimembraneous ventricular septal defect.” Thus, use of an ontology allows the retrieval of all information associated with a term and its children.

Phenotype data also are retrieved from the MGI Batch Query Form and the MGI BioMart. MGI also maintains a suite of public reports containing phenotype data for download. Using MGI’s public reports and web services, MGI data can be exported to a variety of other data providers such as NCBI, EBI, UCSC, OMIM, and mouse mutant repositories, where they are incorporated to enrich those resources.

Table 2 continued

| Resource | URL |
|----------|-----|
| UniProt-GOA (Gene Ontology Annotation) | http://www.ebi.ac.uk/GOA |
| *VeryGene | http://www.verygene.com |
| Zebrafish Database (ZFIN) phenotype example | http://zfin.org/action/phenotype/phenotype-statement?id=53698 |

Relationship of the Mammalian Phenotype Ontology to other ontologies for model organism phenotypes and human disease data

Comparing phenotypes among organisms as well as against human phenotypes (and thereby with human disease) makes it possible to discover commonalities of gene function, pathways, and mechanisms. Because all organisms currently have significant gaps in the experimental knowledge of mutations and phenotypes for all genes and in the understanding of the function and interactions for each gene, comparative analyses can provide clues and direction for new experimental validation and research avenues.

At present, there is no universal phenotype ontology for all species that could easily facilitate comparative phenotyping. For mammals, the MP is widely accepted and applied (see above and Table 2). For human, the Human Phenotype Ontology (HP, http://www.human-phenotype-ontology.org) (Robinson et al. 2008), also a precomposed ontology, is actively being developed. For other model organisms, approaches vary, from species-specific vocabulary lists (e.g., in FlyBase, the *Drosophila* model organism database, http://flybase.org) to the “EQ” (entity + quality) approach (e.g., in ZFIN, the Zebrafish Information Network). In the EQ approach, terms are composed de novo at the time of annotation using an “entity” term found in other existing ontologies [e.g., Gene Ontology (GO, http://www.geneontology.org), Chemical Entities of Biological Interest (ChEBI, http://www.ebi.ac.uk/chebi)] plus a “quality” from the Phenotype and Trait Ontology (PATO, http://code.google.com/p/pato) (Gkoutos et al. 2005) that provides the term modifier (e.g., pale, enlarged, absent). For example, ZFIN combines the anatomy term “pericardium” (ZFA:0000054) with the PATO term “edematous” (PATO:0001450) to create a complete phenotype (EQ) statement “pericardium edematous, abnormal” (http://zfin.org/action/phenotype/phenotype-statement?id=53698). The MP Ontology contains a single precomposed term “pericardial edema” (MP:0001787).

Mappings between terms of phenotype ontologies harmonize these different approaches. For the MP, direct mappings to other precomposed phenotype ontologies such as the HP, or indirect mappings of MP terms to EQ statements (Mungall et al. 2010) are used. The EQ mapping consists of developing a “logical definition” in ontological parlance. Logical definitions for MP and HP can be combined with annotations from other species databases that use EQ statements to describe phenotypes, making multispecies phenotype data integration and comparisons possible (Mungall et al. 2010; Washington et al. 2009). Importantly, logical definitions enhance the MP by establishing relationships of terms to a wider suite of interoperating ontologies. However, aggregate terms such as hydrocephaly or glomerular crescent require representation as multiple EQ statements, diminishing the meaning and recognition of these scientific/clinical terminologies.

Ontologies developed as annotation tools (e.g., MP, HP, and GO) are improved by mapping concepts to a common reference framework based on existing standard ontologies (such as the global anatomy ontology Uberon, http://obofoundry.org/wiki/index.php/UBERON:Main_Page) (Mungall et al. 2012). Maintenance of ontologies with multiple inheritance pathways becomes increasingly difficult with increased size and complexity, and they are particularly difficult to
| Center | Reference | URL |
|--------|-----------|-----|
| **ENU Mutagenesis Projects** | | |
| Australian Phenomics Facility at ANU | Nelms and Goodnow (2001) | http://www.apf.edu.au/about/projects/index.shtml |
| Cardiovascular Development Consortium (CvDC) | Kaltman et al. (2010) | http://www.devbio.pitt.edu/research/mouse_muta.html |
| Heart, Lung, Blood & Sleep Center (HLBS) | Svenson et al. (2003) | http://pga.jax.org/protocol_008.html |
| Helmholtz Zentrum München IDG (formerly GSF) | Hrabe de Angelis et al. (2000) | http://www.helmholtz-muenchen.de/en/idg/mouse-genetics/large-scale-mutagenesis/enu-mutagenesis/ |
| Mouse Mutagenesis for Developmental Defects (BCM) | Kile et al. (2003) | http://www.har.mrc.ac.uk/research/mutagenesis/ |
| MRC Harwell Mutagenesis Project | Nolan et al. (2000) | http://mutagenetix.utsouthwestern.edu/ |
| Mutagenix | Hoebe and Becker (2005) | http://mutagenetix.utsouthwestern.edu/ |
| Neuroscience Mutagenesis Consortium (NMICE) | Goldowitz et al. (2004) | http://www.neuromice.org/About/ParticipatingCenters.jsp |
| Reproductive Genomics (ReproGenomics) | Lessard et al. (2004) | http://reproductivegenomics.jax.org |
| RIKEN Japan | Gondo et al. (2010) | http://www.brc.riken.jp/labs/gsc/mouse/ |
| Sloan-Kettering Mouse Project | Anderson (2000) | https://mouse.mskcc.org/mutant/mutantBase.php |
| Toronto Centre for Modeling Human Disease (CMHD) | Cordes (2005) | http://www.cmhd.ca/en_mutationism |
| **Gene Trap Projects (reported in IGTC, www.genetrap.org)** | | |
| BayGenomics | Stryke et al. (2003) | http://www.mmrrc.org/catalog/overview_BG.php |
| European Mouse Mutagenesis Consortium (EUCOMM) | Friedel et al. (2007) | http://www.helmholtz-muenchen.de/en/idg/mouse-genetics/large-scale-mutagenesis/eucomm/ |
| Exchangeable Gene Trap Clones (EGTC) | Araki et al. (1999) | http://egtc.jp |
| German Gene Trap Consortium | Wiles et al. (2000) | http://www.helmholtz-muenchen.de/en/idg/mouse-genetics/large-scale-mutagenesis/german-gene-trap-consortium |
| Lexicon Genetics OMNIBANK | Zambrowicz et al. (1998) | Resource now available through TIGM |
| RIKEN BioResource Center | Matsuda et al. (2004) | http://www2.brc.riken.jp/lab/mouse_es/index.html.en |
| Sanger Institute Gene Trap Resource | Guo et al. (2004) | http://www.sanger.ac.uk/resources/mouse/sigtr/ |
| Soriano Lab Gene Trap Database | Friedel and Soriano (2010) | http://research.mssm.edu/soriano/lab/gene_trap.html |
| The Institute for Genomic Medicine (TIGM) | Hansen et al. (2008) | http://www.tigm.org/ |
| TIGEM-IRBM Gene Trap | Roma et al. (2008) | http://gene trap.tigem.it/public/index.php |
| Toronto Centre for Modeling Human Disease (CMHD) | To et al. (2004) | http://www.cmhd.ca/gene trap |
| **Knockout Mouse Projects** | | |
| Deltagen | Moore (2005) | http://www.deltagen.com |
| European Mouse Mutagenesis Consortium (EUCOMM) | Auwerx et al. (2004) | www.eucomm.org |
| European Mouse Mutagenesis Consortium Tools (EUCOMMTools) | Skarnes et al. (2011) | www.eucommtools.org |
| Knockout Mouse Project (KOMP) | Asp et al. (2004) | http://www.knockoutmouse.org |
| Lexicon Genetics | Frööde et al. (2003) | http://www.lexicon-genetics.com |
| North American Conditional Mouse Mutagenesis (NorCOMM) | Collins et al. (2007a) | http://www.norcomm.org |
| The Institute for Genomic Medicine (TIGM) | Collins et al. (2007b) | http://www.tigm.org |

This is a not an exhaustive list of major projects.
view for missing terms when additions are largely dictated by the need of curation or projects to define new terms. Missing terms can be logically assumed (i.e., if there is a term “increased X,” the converse term “decreased X” should exist), although such terms might not be biologically relevant phenotypes. Automatic reasoners, software tools that infer the positions of terms in a subsumption hierarchy based on logical definitions, have been exploited to identify missing or erroneous relationships and detect omissions in ontologies (Mungall et al. 2011). To this end, the tool GULO (Getting an Understanding of LOgical definitions, http://compbio.charite.de/svn/hpo/trunk/src/tools/gulo) (Köhler et al. 2011) was applied to refine the MP. Based on the results of reasoner analysis, MP added over 300 new child–parent relationships. This work also uncovered discrepancies in reference ontologies used to construct logical definitions and errors in assignment of EQ statements. Therefore, the use of logical definitions, coupled with software tool reasoners, automates some aspects of ontology review for completeness and term placement in the DAG, although manual verification is needed to ensure accuracy and to place aggregate terms.

A combination of EQ statements and precomposed MP terms can reciprocally meet the needs of automated phenotype annotation pipelines and biomedical researchers interested in exploring data. For example, Europhenome describes parameters tested through the phenotyping pipeline using EQ statements. An eye dysmorphology parameter defined by “eye” (MA:0000261) and “size” (PATO:0000117) defines the test itself, and a phenotypic observation may then be “eye” and “decreased size” (PATO:0000587). This EQ combination corresponds to the MP term “microphthalmia" (MP:0001297), or small eye, a term familiar to scientists and clinicians and used by Europhenome in its web interface and BioMart (Beck et al. 2009; Morgan et al. 2010).

Integration with ongoing and new systematic phenotyping efforts

Systematic phenotyping, where many centers apply common phenotyping protocols under the same conditions (e.g., age, sex), can provide high-quality data sets for analysis and comparison. This interinstitutional standardization of phenotype testing was experimented with and analyzed extensively in the last decade in the European projects of EUROMORPHIA (Brown et al. 2005; Mandillo et al. 2008) and EUMODIC (Beck et al. 2009; Morgan et al. 2010). The role of EUMODIC was to generate phenotype data from the first 500 mutant mouse knockout lines derived from the International Knockout Mouse Project (IKMC, http://www.knockoutmouse.org) (Ringwald et al. 2011; Skarnes et al. 2011) using standardized protocols and pipelines. The results produced by these centers are reported through the Europhenome database (Morgan et al. 2010) and are available through a web interface or through the IKMC BioMart (Oakley et al. 2011). In addition, data generated by the Sanger Institute are available from their mouse resources portal. Both of these resources perform statistical comparisons of experimentally generated mutant mouse phenotype data with control phenotype data to identify significant phenodeviants. Significant results are stored and the relevant MP term for the test result is automatically applied to the mutant line. In addition to automated pipelines, data also are analyzed manually at the Sanger Institute and phenodeviants calls and MP terms are assigned to these results. Both the Europhenome and Sanger Institute’s Mouse Resource Portal sites allow searching and browsing for phenodeviants data using MP terms [see review of both of these efforts in Ayadi et al. (2012)].

MGI is undertaking the importation of these data to integrate them with all the resources that MGI offers. Because these phenotype data are already associated with MP terms, as well as official gene, allele, and strain nomenclature and other standardized parameters, importation is automatable and thus reduces the need for further curation. Similarly, MGI can effectively work with data sets of increasing size, such as those expected from the IMPC, which has a stated goal to carry out high-throughput phenotyping for over 20,000 mutant mouse lines to determine the function of every gene in the mouse genome (Brown and Moore 2012).

Use of phenotype ontologies and mouse phenotype and disease annotations in research and online tools

Mouse genotypes in MGI are annotated to human disease terms from OMIM (Amberger et al. 2011) when an author demonstrates that the phenotype mimics the human disease state. For example, the Fgfr2tm1Schl allele is a model of the human disease Crouzon Syndrome (OMIM ID:123500) (Fig. 2).
### Phenotype Summary

**Key:**
- Phenotype feature: R	R
- Genotype feature: R	R

#### Affected Systems
- craniofacial
- digestive/vascular system
- mortality/aging
- respiratory system
- excision
- vision/eye

#### Disease Models

| Human Disease | Genotype | Allallelic Composition | Genetic Background |
|---------------|----------|------------------------|--------------------|
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |

### Phenotype Data by Genotype

Models with phenotypic similarity to human diseases associated with Fgfr2.

| Human Disease | Genotype | Allallelic Composition | Genetic Background |
|---------------|----------|------------------------|--------------------|
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |

### Mouse Models of Human Disease

| Creutzfeldt-Jakob Disease | Genotype | Allallelic Composition | Genetic Background |
|--------------------------|----------|------------------------|--------------------|
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |
| Creutzfeldt-Jakob Disease | Fgfr2<sup>cre+/-</sup>, Fgfr2<sup>-/-</sup> | inherited: 12991/Sv | 1188299 |

---

1. Human genes are associated with this disease. Mouse models genotypes include mutations in the ortholog of these human genes.
2. Human genes are associated with this disease. Mouse models genotypes do not include mutations in the orthology of these human genes.
Links from MGI mutant allele details to both the OMIM record describing this condition in humans and to MGI’s Human Disease and Mouse Model web pages are provided. These models are searched using the OMIM term or ID from the Quick Search Box on any MGI page or the advanced Genes and Markers Query Form or Phenotypes Query Form, or they may be browsed on the Human Disease Vocabulary browser. OMIM is used as MGI’s source of human genetic disease terms because it provides associated detailed descriptions of human disease and clinical synopses, associates OMIM disease records to human genes, and is recognized and frequently used by clinicians and biomedical researchers as an authoritative information source.

Given the exponentially increasing amount of complex mouse phenotype and human disease model data in MGI and elsewhere and that these data are stored in model organism databases using different methods, computational tools are required that will lead to better data mining and comparison of phenotypic data across different species. There are a number of approaches using lexical matching or ontology mapping of phenotype or clinical terms that map phenotypic similarity between mouse and human genes and variants and suggest candidate genes for human diseases [e.g., PhenomicDB, http://www.phenomicdb.de (Groth et al. 2010); PhenoHM, http://phenome.cchmc.org/phenome (Sardana et al. 2010); MouseFinder, http://www.mousemodels.org (Chen et al. 2012); PhenomeNet, http://phenomebrowser.net (Hoehndorf et al. 2011; Gkoutos et al. 2012; and reviewed in Schofield et al. 2012)].

In addition to comparing data across species, MP annotated phenotype data are used as a parameter by a number of web tools that integrate published and high-throughput data

![Fig. 3 Example of a comprehensive MGI data search result using the term “ventricular septal defect.” Shown are samples of the results listed from this search. All 300 genotypes in MGI that are annotated to this term, or terms below this node in the MP, are returned. For example, genotypes annotated to the term “inlet ventricular septal defect,” a child term to “ventricular septal defect,” are seen in the results list](image-url)
to facilitate gene discovery via enrichment analysis of gene sets or to identify candidate genes for QTL. Among the tools for enrichment analysis are MamPhea (http://evol.nhri.org.tw/phenome/index.jsp?platform=mmus) (Weng and Liao 2010), which enables gene enrichment analysis of genes from multiple species based exclusively on MP annotations from mouse, and ToppGene (http://toppgene.cchmc.org/prioritization.jsp) (Chen et al. 2009), a gene enrichment tool that uses MP as one of many parameters for sorting gene sets.

Other resources include Gene Weaver (http://www.GeneWeaver.org) (Baker et al. 2012), which integrates sets of biological functions (GO), their relations to mutant phenotypes through the MP, KEGG pathways (http://www.genome.jp/kegg/pathway.html), QTL data, and more. VeryGene (http://www.verygene.com) (Yang et al. 2011) links tissue-specific gene expression data to data on gene function (GO), Reactome (http://www.reactome.org), KEGG pathways, MP annotations, disease associations, and targeting drugs. Among the web tools for candidate gene identification is the AnnotQTL tool (http://annotqtl.genouest.org) (Lecerf et al. 2011), which adds mouse MP annotations, as well as mouse and human gene function (GO) annotations, to genes in an identified QTL interval region to assist in predicting candidate genes.

Using the tools described above and elsewhere, a number of recent studies highlight the use of mouse MP annotations in the identification or validation of candidate gene sets in human disease and mouse studies:

- Dickerson et al. (2011) identified 1,965 human disease genes from OMIM’s morbid map and separated them according to whether the knockout phenotype of the mouse ortholog was lethal (essential) or viable from phenotype data coded to the MP in MGI. Human genes in this set with mouse orthologs having a lethal phenotype are over-represented among disease genes associated with cancer and highly connected in protein–protein interaction networks.

- Russell et al. (2012) discovered novel candidate genes for congenital diaphragmatic hernia by expression profiling of mouse embryonic diaphragm, then applying gene enrichment analysis on this identified set with MGI annotated data of muscle development and metabolism terms in the GO and abnormal muscle and cardiovascular phenotype terms from the MP ontology. Twenty-seven new candidate genes were identified. One candidate gene, pre-B-cell leukemia transcription factor one (Pbx1), when mutated, results in a range of previously undetected diaphragmatic defects in mice.

- Meehan et al. (2011) used MGI phenotype data associated with MP terms and mouse model data to create a set of similarly annotated genes/genotypes likely to have previously uncharacterized autistic-like phenotypes. The implicated genes considerably overlapped with a set of over 300 human genes associated with human autism spectrum disorder due to small, rare copy number variants (CNVs, Pinto et al. 2010). Similarly, Gai et al. (2012) identified 12 MP ontology term annotations that are significantly enriched in genes overlapping inherited rare autism CNVs and are consistent with observable phenotypes associated with human autism spectrum disorder behaviors.

- Shaikh et al. (2011) identified a group of genes enriched in human developmental delay-associated CNVs, which when disrupted in mice, result in specific nervous system phenotypes. The most significant term annotated to these genes was an abnormal nervous system white matter tract phenotype, which was used to narrow the candidate gene set for further analysis.

- Bayés et al. (2011) identified 1,461 proteins present in human neocortex postsynaptic density. Mutations in 199 of these genes were associated with human neurological diseases in OMIM. Enrichment analysis revealed 77 MP terms, including cognitive and motor phenotypes associated with mutations in the mouse orthologs that revealed new candidate genes. A similar gene enrichment result was shown using the Human Phenotype Ontology annotations derived from OMIM.

- Hageman et al. (2011) used MP annotations to kidney phenotypes to narrow the genomic intervals and find candidate genes for QTL affecting the urinary albumin-to-creatinine ratio in mice.

Thus, the predictive value of mouse mutant phenotypes in identifying new candidate genes assists researchers in revealing the complex nature of human diseases.

Summary and future prospects

The MP ontology continues to evolve and expand to robustly describe phenotypes. New terms and structural refinements are incorporated as required by phenotype annotation efforts at MGI and other databases, phenotyping centers, mutagenesis projects, investigator research, and review by biological domain experts. MGI curates information on published mouse mutations and electronically imports phenotype and disease model information from other sources.

MGI continues to adapt as new data drive database infrastructure and as public data presentation changes. For example, changes are already underway to accommodate Euphenome and the Sanger Institute’s large-scale phenotype data derived from targeted knockout mutations, as well as future IMPC phenotype data. Additional new sources of mouse allele, variant, and phenotype data will arise from the
Collaborative Cross (CC, http://csbio.unc.edu/CCstatus) (Churchill et al. 2004; Threadgill and Churchill 2012) and the Diversity Outcross (DO, http://cdg.jax.org/datasets/phenotype/SvensonDO.shtml) (Svenson et al. 2012), as well as mutations induced by engineered zinc finger nucleases (Osiak et al. 2011). Other mutation-generation techniques, including transposon-induced mutations (cf. Largaespada 2009; Liang et al. 2009; Takeda et al. 2007; Wang et al. 2008), and the detection by NexGen and whole-exome sequencing of significant numbers of previously undetectable ENU mutations (cf. Arnold et al. 2011; Boles et al. 2009; Guryev and Cuppen 2009; Sun et al. 2012) will further increase the genomic mutations and phenotypic data that require MP and nomenclature standards for integration with existing data. The promise of integrating these many streams of phenotype data with a robust MP ontology will enable a growing reservoir of standardized data for data mining, gene set enrichment studies, candidate disease model identification, and validation of computational predictions.

Many challenges remain in the ability to use computational tools to analyze and compare data from human clinical and mouse phenotype resources. Human GO data are freely available via the Universal Protein Resource GO Annotations (UniProt-GOA, MP Ontology http://www.ebi.ac.uk/GOA) (Dimmer et al. 2012), but genetic, disease, and clinical data are scattered in many databases with differing formats and accessibility, and many resources are not maintained in a computational-friendly format (Küntzer et al. 2010). The HP, now being adopted by resources such as NCBI, is available for standardization of human clinical symptoms (Robinson et al. 2008) and is mapped to OMIM disease records. Logical definitions derived for the HP are mapped to similar ontologies such as the MP (Mungall et al. 2011).

The HP is only one part of the infrastructure needed for human disease data management, however. A comprehensive disease ontology with descriptions and definitions of disease terms in the context of observable clinical features, including a mapping to other phenotype ontologies such as HP and/or MP, is required to maximize the interoperability and computational access to the wide range of human disease data. Current vocabularies for human disease have a number of drawbacks that prevent their wide adoption as a robust source for human disease annotation. OMIM, while an excellent source of text descriptions of disease, lacks a hierarchical structure and is limited to Mendelian disease. The international classification of disease (ICD, http://www.nlm.nih.gov/mesh/MBrowser.html) is designed for physician billing codes and thus is confounded by many nondisease terms such as those for injury and infection. The Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT, http://www.ihtsdo.org/snomed-ct) must be licensed for use by country or affiliation and is thus not a publicly available resource.

Several nascent efforts that are developing human disease ontologies/vocabularies are underway. These include the Disease Ontology (DO, http://disease-ontology.org) (Schriml et al. 2012), MEDIC (http://ctdbase.org/voc.go?type=disease) (Davis et al. 2012), and Orphanet (http://www.orpha.net/consor/cgi-bin) (Rath et al. 2012). In addition, the Medical Subject Headings (MeSH)-disease branch at the U.S. National Library of Medicine (Nelson et al. 2004) is increasingly incorporating OMIM disease terms. Ultimately, the successful growth and maturation of one or more of these or other proposed disease ontologies and vocabularies should lead to greater interoperability of human genetic, disease, and clinical data among the scattered resources, as well as integration with model organism data. Adoption of semantic and syntactic standards by the human clinical community will facilitate integration of data from a multitude of resources and allow the ability to compute over many data sets, as has been demonstrated for mouse genetic and phenotype data via the MP.

Acknowledgments We thank S. Bello and R. Babiuk for helpful comments on the manuscript. This work was funded by the National Institutes of Health, National Human Genome Research Institute (NHGRI) Grant HG000330.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Acevedo-Arozena A, Wells S, Potter P, Kelly M, Cox RD, Brown SD (2008) ENU mutagenesis, a way forward to understand gene function. Annu Rev Genomics Hum Genet 9:49–69

Amberger J, Bocchini C, Hamosh A (2011) A new face and new challenges for online Mendelian inheritance in man (OMIM®). Hum Mutat 32:564–567

Anderson KV (2000) Finding the genes that direct mammalian development: ENU mutagenesis in the mouse. Trends Genet 16:99–102

Araki K, Imaizumi T, Sekimoto T, Yoshinobu K, Yoshimuta J, Akizuki M, Miura K, Araki M, Yamamura K (1999) Exchangeable gene trap using the Cre/mutated lox system. Cell Mol Biol (Noisy-le-grand) 45:737–750

Araki M, Araki K, Yamamura K (2009) International gene trap project: towards gene-driven saturation mutagenesis in mice. Curr Pharm Biotechnol 10:221–229

Arnold CN, Xia Y, Lin P, Ross C, Schwander M, Smart NG, Müller U, Beutler B (2011) Rapid identification of a disease allele in mouse through whole genome sequencing and bulk segregation analysis. Genetics 187:633–641

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, Grieder FB, Heintz N,
Hicks G, Insel TR, Joyner A, Koller BH, Lloyd KC, Magnunson T, Moore MW, Nagy A, Pollock JD, Roses AD, Sands AT, Seed B, Skarnes WC, Snoddy J, Soriano P, Stewart DJ, Stewart F, Stillman B, Varumis H, Varticovski L, Verma IM, Vogt TF, von Melchert H, Witkowsky JP, Wurst W, Yancopoulos GD, Young SG, Zambrowicz B (2004) The knockout mouse project. Nat Genet 36:921–924

Auwerx J, Avner P, Baldock R, Ballabio A, Balling R, Barbacid M, Berns A, Bradley A, Brown S, Carmeliet P, Chambon P, Cox R, Davidson D, Davies K, Duboule D, Forejt J, Granucci F, Hastie N, de Angelis MH, Jackson I, Kiousis D, Kaillas G, Lathrop M, Lendahl U, Malumbres M, Melchior H, Muller W, Partanen J, Ricciardi-Castagnoli P, Rigby P, Rosenthal N, Skarnes B, Stewart AF, Thornton J, Tocchini-Valentini G, Wagner E, Wahl W, Wurst W (2004) The European dimension for the mouse genome mutagenesis program. Nat Genet 36: 925–927

Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, Gailus-Durner V, Greenaway S, Houghton R, Karp N, Leblanc S, Lengger C, Maier H, Mallon A, Marshall S, Melvin D, Morgan H, Pavlovic G, Ryder E, Skarnes B, Selloum M, Ramirez Solis R, Sorg T, Teboul L, Vasseur L, Walling A, Weaver T, Wells S, White J, Bradley A, Adams D, Steel K, Hrabé de Angelis M, Brown S, Hervauly (2012) Mouse large-scale phenotyping initiatives: overview of the European mouse disease clinic (EUMODIC) and of the wellcome trust sanger institute mouse genetics project. Mamm Genome 23:000–000. doi: 10.1007/s00335-012-9418-y

Baker EJ, Jay JJ, Bubier JH, Langston MA, Chesler EJ (2012) GeneWebber: a web-based system for integrative functional genomics. Nucl Acids Res 40:D1067–D1076

Bayés A, van de Lagemata LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, Grant SG (2011) Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 14:19–21

Beck T, Morgan H, Blake A, Wells S, Hancock JM, Mallon AM (2009) Practical application of ontologies to annotate and analyse large scale raw mouse phenotype data. BMC Bioinform 10(Suppl 5):S2

Blake JA, Bult CJ, Kadon JA, Richardson JE, Eppig JT; Mouse Genome Database Group (2011) The mouse genome database (MGD); premier model organism resource for mammalian genomics and genetics. Nucl Acids Res 39:D842–D848

Boles JK, Wilkinson BM, Wilming LG, Liu B, Probst FJ, Harrow J, Grafton D, Hentges KE, Woodward LP, Maxwell A, Mitchell K, Risley MD, Johnson R, Hirschi K, Lupski JR, Funato Y, Miki D, Hrabe de Angelis MHH, Fuchs H, Rathkolb B, Soewarto D, Jacob HJ, Duek-Roggli P, Xenarios I, Apweiler R (2012) The UniProt-GO annotation database in 2011. Nucl Acids Res 40:D565–D570

Brown SDM, Moore MW (2012) The international mouse phenotyping consortium: past and future perspectives on mouse phenotyping. Mamm Genome 23:000–000. doi: 10.1007/s00335-012-9427-x

Brown SD, Chambon P, de Angelis MH, Eumorphia Consortium (2005) EMPReSS: standardized phenotype screens for functional annotation of the mouse genome. Nat Genet 37:1155

Chen J, Barde EE, Aronow BJ, Jegga AG (2009) TopGene suite for gene list enrichment analysis and candidate gene prioritization. Nucl Acids Res 37:W305–W311

Chen CK, Mungall CJ, Ghodous GV, Doelken SC, Köhler S, Ruff BJ, Smith C, Westerfield M, Robinson PN, Lewis SE, Schofield PN, Smedley D (2012) MouseFinder: candidate disease genes from mouse phenotype data. Hum Mutat 33:858–866

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Bromw KA, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, Cox RD, Crabbie JC, Crusio WE, Darvasi A, Deschepper CF, Doerge RW, Faber CR, Forejt J, Gailde D, Garlov SJ, Geiger H, Gershenson H, Gordon T, Gu J, Gu, de Haan G, Hayes NL, Heller C, Himmelhauer B, Hitzemann R, Hunter K, Hsu HC, Iraqi FA, Ivandic B, Jacob HJ, Jansen RC, Jepsen KJ, Johnson DK, Johnson TE, Kempermann G, Kindzierski C, Kotb M, Kooy RF, Llamas B, Lammert F, Lassalle JM, Lowenstein PR, Lu L, Lusis A, Manly KF, Marcucio R, Matthews DF, Medrano JF, Miller DR, Mittleman M, 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 JH, Petlonalen-Patolite 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, Sillanpää MJ, Siracusa LD, Snoeck HW, Spearow JL, Svenson K, Tarantino LM, Threadgill D, Toth LA, Valdar W, de Villena FP, Warden C, Whately S, Williams RW, Wiltshire T, Yi N, Zhang D, Zhang M, Zou F, Complex Trait Consortium (2004) The collaborative cross, a community resource for the genetic analysis of complex traits. Nat Genet 36:1133–1137

Clark AT, Goldowitz D, Takahashi JS, Vitaterra MH, Siepka SM, Peters LL, Franklin WN, Carlson GA, Rossant J, Nadeau JH, Justice MJ (2004) Implementing large-scale ENU mutagenesis screens in North America. Genetica 122:51–64

Collins FS, Rossant J, Wurtz W (2007a) The international knockout mouse consortium: a mouse for all reasons. Cell 128:9–13

Collins FS, Finnell RH, Rossant J, Wurtz W (2007b) The international knockout mouse consortium: a new partner for the international knockout mouse consortium. Cell 129:235

Cook MC, Vinuesa CG, Goodnow CC (2006) ENU-mutagenesis: insight into immune function and pathology. Curr Opin Immunol 18:627–633

Cordes SP (2005) N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. Microbiol Mol Biol Rev 69:426–439

Davis AP, Wiegers TC, Rosenstein MC, Mattingly CJ (2012) MEDIC: a practical disease vocabulary used at the comparative toxicogenomics database. Database (Oxford) 20:bar065

de Angelis MHH, Flaswinkel H, Fuchs H, Kathkolb B, Soewarto D, Marshall S, Heffner S, Purgent W, Wensch K, Jung M, Reis A, Richter T, Alessandri F, Jakob T, Fuhs E, Kohl H, Kremmer E, Schaeble K, Rollinski B, Roscher A, Peters C, Meitinger T, Strom T, Steckler T, Holsboer F, Klopstock T, Gekeler F, Schindewolf C, Jung T, Avraham K, Behrendt H, Ring J, Zimmer A, Schughart K, Pfeffer K, Wolf E, Balling R (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25:444–447

Dickerson JE, Zhu A, Robertson DL, Hentske KE (2011) Defining the role of essential genes in human disease. PLoS One 6:e27368

Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O’Donovan M, Martin MJ, Bely B, Browne P, Mun Chan W, Eberhardt R, Gardner M, Laiho K, Legge D, Magrane M, Pichler K, Poggioli D, Sehra H, Aucinchloss A, Axelsen K, Blatter MC, Boutet E, Bracon–Quintaje S, Breuza L, Bridge A, Courdet A, Coste E, Crichton T, Alessandri F, Jakob T, Fuchs E, Kohl H, Kremmer E, Schaeble K, Rollinski B, Roscher A, Peters C, Meitinger T, Strom T, Steckler T, Holsboer F, Klopstock T, Gekefer F, Schindewolf C, Jung T, Avraham K, Behrendt H, Ring J, Zimmer A, Schughart K, Pfeffer K, Wolf E, Balling R (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. Nat Genet 25:444–447

Cusin I, Duek-Roggli P, Xenarios I, Apweiler R (2012) The UniProt-GO annotation database in 2011. Nucl Acids Res 40: D565–D570

Gumowski N, Hinz U, Hulo C, James J, Jimenez S, Jungo F, Keller G, Lemercier P, Lieberherr D, Masson P, Moinat M, Pedruzzi I, Poux S, Rivoire C, Roechert B, Schneider M, Stutz A, Sundaram S, Tognolli M, Bouguerelot L, Argoud-Puy G, Cusin I, Duek-Roggli P, Xanarios I, Apweiler R (2012) The UniProt-GO annotation database in 2011. Nucl Acids Res 40: D565–D570
Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database Group (2012) The mouse genome database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucl Acids Res 40:D881–D886

Eswarakumar VP, Ozcan F, Lew ED, Bae JH, Tome´ F, Booth CJ, Adams DJ, Lax I, Schlessinger J (2006) Attenuation of signaling pathways stimulated by pathologically activated FGFR-receptor 2 mutants prevents craniosynostosis. Proc Natl Acad Sci USA 103:18603–18608

Friddle CJ, Abuin A, Ramirez-Solís R, Richter LJ, Buxton EC, Edwards J, Finch RA, Gupta A, Hansen G, Holt KH, Hu Y, Huang W, Jaing C, Key BW Jr, Kipp P, Kohlhauff B, Ma QZ, Markesich D, Newhouse M, Perry T, Platt KA, Potter DG, Qian N, Shaw J, Schrick J, Shi ZZ, Sparks MJ, Tran D, Wann ER, Walke W, Wallace JD, Xu N, Zhu Q, Person C, Sands AT, Zambrowicz BP (2003) High-throughput mouse knockouts provide a functional analysis of the genome. Cold Spring Harb Symp Quant Biol 68:311–315

Friedel RH, Soriano P (2010) Gene trap mutagenesis in the mouse. Methods Enzymol 477:243–269

Friedel RH, Seisenberger C, Kaloff C, Wurst W (2007) EUCOMM: the European conditional mouse mutagenesis program. Brief Funct Genomic Proteomic 6:180–185

Gai X, Xie HM, Perin JC, Takahashi N, Murphy K, Wenocur AS, D’arcy M, O’Hara RJ, Goldmuntz E, Grice DE, Shaikh TH, Hakonarson H, Buxbaum JD, Elia J, White PS (2012) Rare structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry 17:402–411

Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D (2005) Using ontologies to describe mouse phenotypes. Genome Biol 6:R8

Gkoutos GV, Schofield PN, Hoehndorf R (2012) Computational tools for comparative phenomics; the role and promise of ontologies. Mamm Genome 23:000–000. doi:10.1007/s00292-012-0404-4

Goldowitz D, Franken WN, Takahashi JS, Holtz-Vitaterna M, Bult C, Kibbe WA, Snoddy J, Li Y, Pretel S, Yates J, Swanson DJ (2004) Large-scale mutagenesis of the mouse to understand the genetic bases of nervous system structure and function. Brain Res Mol Brain Res 132:105–115

Gondo Y, Fukumura R, Murata T, Makino S (2010) ENU-based gene-driven mutagenesis in the mouse: a next-generation gene-targeting system. Exp Anim 59:537–548

Grotch P, Kaliev I, Iliﬀe T, Raibov K, Leser U, Weiss B (2010) Phenoclustering: online mining of cross-species phenotypes. Bioinformatics 26:1924–1925

Guo G, Wang W, Bradley A (2004) Mismatch repair genes identiﬁed related to urinary albumin excretion. J Am Soc Nephrol 22:73–81

Hansen GM, Markesich DC, Burnett MB, Zhu Q, Dionne KM, Richter LJ, Finnhell RH, Sands AT, Zambrowicz BP, Abuin A (2008) Large-scale gene trapping in C57BL/6N mouse embryonic stem cells. Genome Res 18:1670–1679

Hoebe K, Beutler B (2005) Unraveling innate immunity using large scale N-ethyl-N-nitrosourea mutagenesis. Tissue Antigens 65:395–401

Hoehndorf R, Schoﬁeld PN, Gkoutos GV (2011) PhenomeNET: a whole-phenome approach to disease gene discovery. Nucl Acids Res 39:e119

Kaltman JR, Schramm C, Pearson GD (2010) The national heart, lung, and blood institute bench to bassinet program: a new paradigm for translational research. J Am Coll Cardiol 55: 1262–1265

Keane JF, Fyler DC, Lock JE (2006) Nadas’ pediatric cardiology, 2nd edn. WB Saunders, St Louis, ISBN-13: 978-14160-2390-6

Kile BT, Hentges KE, Clark AT, Nakamura H, Salinger AP, Liu B, Box N, Stockton DW, Johnson RL, Behringer RR, Bradley A, Justice MJ (2003) Functional genetic analysis of mouse chromosome 11. Nature 425:81–86

Köhler S, Bauer S, Mungall CJ, Carletti G, Smith CL, Schoﬁeld P, Gkoutos GV, Robinson PN (2011) Improving ontologies by automatic reasoning and evaluation of logical deﬁnitions. BMC Bioinform 12:418

Küntzer J, Eggle D, Klostermann S, Burtscher H (2010) Human variation databases. Database (Oxford) 2010:baq015. doi: 10.1093/database/baq015

Largaespada DA (2009) Transposon mutagenesis in mice. Methods Mol Biol 530:379–390

Lecerf F, Breuadoue A, Sallou O, Desert C, Blum Y, Lagarrigue S, Demeure O (2011) AnnotQTL: a new tool to gather functional and comparative information on a genomic region. Nucl Acids Res 39:W28–W333

Lessard C, Pedonka JK, Hartford SA, Schimmenti JC, Handel MA, Eppig JJ (2004) New mouse genetic models for human contraceptive development. Cytogenet Genome Res 105:222–227

Liang Q, Kong J, Stalker J, Bradley A (2009) Chromosomal mobilization and reintegration of sleeping beauty and PiggyBac transposons. Genesis 47:404–408

Mandillo S, Tucci V, Hölter SM, Meziane H, Banchaabouchi MA, Kallnik M, Lad HV, Nolan PM, Ogurcak M, Coghill EL, Gale K, Golini E, Jacquot S, Krezel W, Parker A, Riet F, Schneider I, Marazziti D, Aujexer N, Brown SD, Chambon P, Rosenthal N, Tocchinetti-Valentini G, Wurst W (2008) Reliability, robustness, and reproducibility in mouse behavioral phenotyping: a cross-laboratory study. Physiol Genomics 34:243–255

Matsuda E, Shigeoka T, Iida R, Yamanaka S, Kawaichi M, Ishida Y (2004) Expression proﬁling with arrays of randomly disrupted genes in mouse embryonic stem cells leads to in vivo functional analysis. Proc Natl Acad Sci USA 101:4170–4174

Meehan TF, Carr CJ, Jay JJ, Bult CJ, Chesler EJ, Blake JA (2011) Autism candidate genes via mouse phenomics. J Biomed Inform 44(Suppl 1):S5–S11

Moore MW (2005) High-throughput gene knockouts and phenotyping in mice. Ernst Schering Res Found Workshop 50:27–44

Moran H, Beck T, Blake A, Gates H, Adams N, Dehouzey G, Leblanc S, Lenger C, Maier H, Melvin D, Meziane H, Richardson D, Wells S, White J, Wood J, de Angelis MH, Brown SD, Hancock JM, Mallon AM, EUMODIC Consortium (2010) EuroPhenome: a repository for high-throughput mouse phenotype data. Nucl Acids Res 38:D577–D585

Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M (2010) Integrating phenotype ontologies across multiple species. Genome Biol 11:R2

Mungall CJ, Bada M, Berardini TZ, Deegan J, Ireland A, Harris MA, Hill DP, Lomax J (2011) Cross-product extensions of the gene ontology. J Biomed Inform 44:80–86

Mungall CJ, Tornai C, Gkoutos GV, Lewis SE, Haendel MA (2012) Uberon, an integrative multi-species anatomy ontology. Genome Biol 13:R5

Nelms KA, Goodnow CC (2001) Genome-wide ENU mutagenesis to reveal immune regulators. Immunity 15:409–418

Nelson SJ, Schopen M, Savage AG, Schultman JL, Arluk N (2004) The MeSH translation maintenance system: structure, interface design, and implementation. Stud Health Technol Inform 107(Pt 1):67–69

Nolan PM, Peters J, Strivens M, Rogers D, Hagan J, Spurr N, Gray IC, Vizor L, Brooker D, Whitehill E, Washbourne R, Hough T,
Greenaway S, Hewitt M, Liu X, McCormack S, Pickford K, Selly R, Wells C, Tymowska-Lalanee Z, Roby P, Glenister P, Thornton C, Thaug C, Stevenson JA, Arkell R, Mburu P, Hardisty R, Kierman A, Erven A, Steel KP, Voegeling S, Guenet JL, Nickels C, Sadri M, Nasse M, Isaacs A, Davies K, Browne M, Fisher EM, Martin J, Rastan S, Brown SD, Hunter J (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat Genet 25:440–443

Nord AS, Chang PJ, Conklin BR, Cox AV, Harper CA, Hicks GG, Huang CC, Johns SJ, Kawamoto M, Liu S, Meng EC, Morris JH, Rossant J, Ruiz P, Skarnes WC, Soriano P, Stanford WL, Strung D, von Melchner H, Wurst W, Yamamura K, Young SG, Babbitt PC, Ferrin TE (2006) The international gene trap consortium website: portal to all publicly available gene trap cell lines in mouse. Nucl Acids Res 34:D642–D648

Oakley DJ, Iyer V, Skarnes WC, Smedley D (2011) BioMart as an integration solution for the International Knockout Mouse Consortium. Database (Oxford) 2011:bar028

Osiak A, Radecke F, Guhl E, Radecke S, Dannemann N, Lüttge F, Nord AS, Chang PJ, Conklin BR, Cox AV, Harper CA, Hicks GG, Rath A, Olry A, Dhombres F, Brandt MM, Urber B, Ayme S (2012) The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 83:610–615

Roma G, Sardiello M, Cobellis G, Cruz P, Lago G, Sanches R, Stupka E (2008) The UniTrap resource: tools for the biologist enabling optimized use of gene trap clones. Nucl Acids Res 36:D741–D746

Russell MK, Longoni M, Wells J, Maalouf FI, Tracy AA, Lostcertales M, Ackerman KG, Pober BR, Lage K, Bult CJ, Donaboe PK (2012) Congenital diaphragmatic hernia candidate genes derived from embryonic transcripts. Proc Natl Acad Sci USA 109:2978–2983

Sardana D, Vasa S, Vepachedu N, Chen J, Gudivada AR, Aronow BJ, Jegga AG (2010) PhenoHM: human-mouse comparative phenotype-genome server. Nucl Acids Res 38:W165–W174

Schofield PN, Hoehndorf R, Gkoutos GV (2012) Mouse genetic and phenotypic resources for human genetics. Hum Mutat 33:826–836

Schirim LM, Arze C, Nadendla S, Chang YW, Mazaïtis M, Felix V, Feng G, Kibbe WA (2012) Disease ontology: a backbone for disease semantic integration. Nucl Acids Res 40:D940–D946

Shaikh TH, Haldeman-Englert C, Geiger EA, Ponting CP, Webber C (2011) Genes and biological processes commonly disrupted in rare and heterogeneous developmental delay syndromes. Hum Mol Genet 20:880–883

Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A (2011) A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474:337–342

Smith CL, Eppig JT (2009) The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip Rev Syst Biol Med 1:390–399

Smith CL, Goldsmith C-AW, Eppig JT (2004) The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biol 6:R7

Snell GD (ed) (1941) Biology of the laboratory mouse, 1st edn. Blakiston, New York

Stryke D, Kawamoto M, Huang CC, Johns SJ, King LA, Harper CA, Meng EC, Lee RE, Yee A, L’Italien L, Chuang PT, Young SG, Skarnes WC, Babbitt PC, Ferrin TE (2003) BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. Nucl Acids Res 31:278–281

Sun M, Mondal K, Patel V, Horner VL, Long AB, Cutler DJ, Caspary T, Zwick ME (2012) Multiplex chromosomal exome sequencing accelerates identification of enu-induced mutations in the mouse. G3 (Bethesda) 2:143–150

Svenson KL, Bogue MA, Peters LL (2003) Invited review: identifying new mouse models of cardiovascular disease: a review of high-throughput screens of mutagenized and inbred strains. J Appl Physiol 94:1650–1659

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

Takeda J, Keng VW, Horie K (2007) Germline mutagenesis mediated by sleeping beauty transposon system in mice. Genome Biol 8(Suppl 1):S14

Threadgill DW, Churchill GA (2012) Ten years of the collaborative cross. G3 (Bethesda) 2:153–156
To C, Epp T, Reid T, Lan Q, Yu M, Li CY, Ohishi M, Hant P, Tsao N, Casallo G, Rossant J, Osborne LR, Stanford WL (2004) The centre for modeling human disease gene trap resource. Nucl Acids Res 32:D557–D559
Wang W, Bradley A, Huang Y (2008) A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. Genome Res 19:667–673
Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE (2009) Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol 7:e1000247
Weng MP, Liao BY (2010) MamPhEA: a web tool for mammalian phenotype enrichment analysis. Bioinformatics 26:2212–2213
Wiles MV, Vauti F, Füchtbauer EM, Ruiz P, Füchtbauer A, Arnold HH, Lehrach H, Metz T, von Melchner H, Wurst W (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat Genet 24:13–14
Yang X, Ye Y, Wang G, Huang H, Yu D, Liang S (2011) VeryGene: linking tissue-specific genes to diseases, drugs, and beyond for knowledge discovery. Physiol Genomics 43:457–460
Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392:608–611