Published in final edited form as:

Vet Pathol. 2010 November ; 47(6): 1016–1020. doi:10.1177/0300985810374845.

Pathbase and the MPATH ontology: Community Resources for Mouse Histopathology

Paul N. Schofield^{1,2}, Michael Gruenberge¹, and John P. Sundberg²

¹Department of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge, UK

²The Jackson Laboratory, Bar Harbor, Maine, USA

Abstract

Pathbase, the database of images of mouse histopathology, was developed as a resource to provide free access to representative images of lesions in background and mutant strains of laboratory mice. When utilized in conjunction with diagnostic workups or phenotyping of mutant mice, it can provide a “virtual second opinion” for those working without access to groups of experienced pathologists. This is a community resource, and facilitates the sharing of expertise and data between members of the pathology community worldwide. The MPATH pathology ontology was originally developed alongside Pathbase for the annotation of images. This now represents an important resource for the coding of diagnoses which permits sophisticated data retrieval and computational analysis of mouse phenotypes. The structure and use of MPATH is discussed together with current and future challenges for the coding of mutant mouse phenotypes.

Keywords

mouse pathology; database; ontology; MPATH; mouse anatomy; phenotyping

The major challenge of the post-genomic era is the attribution of function to genes and pathways. The use of model organisms, such as the mouse, to provide phenotype/genotype relations is now established as a key approach to establishing normal gene function, and traditional hypothesis-driven research has now been augmented by the co-ordinated large scale mutagenesis and phenotyping projects established internationally over the last decade. ¹⁶ Mouse Genome Informatics (<http://www.informatics.jax.org/>), at the time of writing, lists 23,600 mutant alleles and 34,038 genotypes with phenotypic information. The International Knockout Mouse Consortium (<http://www.knockoutmouse.org/>)⁸ aims systematically to generate in the order of 20,000 new null or conditional alleles within the next two years, and through the EUMODIC program initially to phenotype 650 new mutant mouse lines. The scale of the resources now available and promised is huge and the potential impact of using these new tools enormous. However, achieving this potential requires systematic annotation of phenotypes, integration of disparate and complex data types, and resources for sharing the data.

The pathological analysis of mutant mice is a key mode of investigation for both small and large scale approaches to phenotyping, and generates what are often the key data needed to interpret and classify the phenotypic consequences of gene mutation or dysregulation. Pathbase and MPATH represent developments made in response to community demand for a systematic, computable description framework for mouse pathology and access to primary data and reference images for mutant mouse phenotyping. These terms and codes can be

integrated into diagnostic databases which, when online and linked to Pathbase can provide access to images and definitions thereby providing a “virtual second opinion”.^{22,26}

Addressing the expertise gap in mutant mouse pathology

The increasingly urgent need for trained pathologists to phenotype genetically engineered mutant mice highlights a now well-recognized deficiency.^{2,6,11,19} Encouraging young pathologists to enter into this field as a full time endeavor, beyond toxicological pathology/drug safety in industry, has been difficult. One reason is that there are few universities that really specialize in genetics-based mouse research to the level of creating large groups to provide an environment to encourage career development. Alternatively, small pathology phenotyping services develop, often lacking a critical mass of experts, much less with a large and diverse enough case load to provide experience over time or accessibility to consultation.²⁴ A number of specialty courses have become available to help address some of the training issues²³ as well as “telepathology”, building on the various slide scanning technologies and software to enable group discussions and consultations online. This is an extension of consortia of pathologists forming on each continent to network as yet another attempt to help each other to maintain a quality program.

A key problem is the dissemination of experts around the world and lack of up to date resources for training and reference. Much of this can now be addressed through use of the World Wide Web. There are a number of websites available that are highly specialized and some provide reference images of mouse histopathology.^{5,27} There are also, gratifyingly, numerous books on the subject with more being published annually.²⁷

Pathbase (<http://www.pathbase.net/>) was developed in 1999 by a group of European and North American Pathologists as a community response to the above problems and was initially funded by the European Commission. It contains photomicrographs of representative images^{17,18} annotated to a set of defined controlled vocabularies and ontologies which provide a public resource for the sharing of images of normal and abnormal tissues from mutant and background strains of laboratory mice (Fig. 1).¹⁷ It currently holds over 2000 Tiff and Jpg images covering a wide range of normal, neoplastic, and non-neoplastic lesions from more than a hundred mutant and inbred background strains. Recently Pathbase has included zoomable whole-slide images (“virtual slides”) in its collection with plans to expand the proportion of these valuable images in the future. The images are annotated with details of strain, genotype, anatomical location, and diagnosis with key annotations derived from controlled vocabularies or OBO-compliant bio-ontologies²⁰ (MPATH, MA, the mouse anatomy ontology, the EMAP developmental anatomy ontology, CL the cell ontology and GO, the gene ontology; see below). Nomenclature for mouse strains and mutant gene symbols are included, when provided, and follow the International Mouse Genetic Nomenclature Committee formats.²⁵ This allows for comparison between studies addressing modifier genes or strain specific diseases that might confuse interpretation.

Using Pathbase

Pathbase is completely open access and is used for reference, training, and data-sharing. Users range from individual investigators carrying out hypothesis-driven academic research using the mouse as a model organism, to large-scale mouse mutagenesis and phenotyping programs. The images are contributed by major research centers, by individual investigators, and from legacy sources where rare and valuable material which is at risk of being lost is digitized and made available to the community. Pathbase is open to anyone to submit images directly through the web page. These are checked by curators and reviewed by members of the pathology panel before being made available. Thereafter, a feedback facility, available

though each record page, enables users to comment on the images and to share their expertise.

The use of Pathbase as a data sharing resource is becoming increasingly important. Histopathology images on which the conclusions of publications are based are frequently either placed on the “supplementary information” sites of journals or are not available at all in a useable size or resolution. The increasing occurrence of egregious diagnostic errors in papers published in major journals has caused great concern. For example neoplasms which investigators claim to have found in new mutant mouse lines are often hyperplasias, dysplasias, cystic lesions, or other degenerative conditions, while too many publications base the entire evidence for phenotypes on erroneous pathological diagnosis.⁷

Submitting Images to Pathbase

Individual users may sign up to submit their own images to Pathbase using the “submit images” link at the bottom of the index page. The user is given a series of menus and input boxes to add the details of the image, strain, diagnosis, etc. and the clicking “upload” allows the user to browse their own machine, select the image and upload directly to Pathbase. The images are then subject to curation and checking before being made publicly available. At the moment the majority of images being submitted are Tiffs and Jpgs, but it is possible to upload WSI images in Aperio or Hamamatsu format using the same system. Because of image sizes and users’ available bandwidth, this may be inefficient. Users should contact Pathbase to arrange for a batch upload facility using FTP.

Integration with other projects and databases

Pathbase has established links with the Jackson Aging Center,³ the Mouse Phenome Database (MPD),⁴ and the Mouse Tumor Biology Database (MTB)¹³ to curate and host images from ongoing studies of age-related lesions and normal tissue variation from 31 inbred strains of mice. Quantitative data for the other systems-based parameters measured are loaded in MPD and the two datasets are integrated between MPD and Pathbase. Similar integration has been recently established with the ERA database²⁸ and with the Northwestern University Janus radiobiology database (<http://janus.northwestern.edu/janus2/>)¹⁴ who have coded 50,000 individual mouse records to MPATH to link the two datasets.

Mouse Anatomy and Pathology Ontologies

While pathologists are trained to use language carefully, over the centuries since Rudolf Virchow³⁰ and others formalized pathology as a medical specialty, language has evolved both to encompass new discoveries and technologies but also to deal with synonyms, spelling variations, etc. These variables are exacerbated by different training sites, training methods, and institutional specialties. Databases depend on standardized terminologies or cross-references for synonyms as well as spelling. In order to address these generic problems the bioinformatics community has, in recent years, produced complex term hierarchies - ontologies - describing various areas of knowledge (gene properties, anatomies etc) where the terms are linked by relationships (e.g. part of, is a, derived from etc). These ontologies have provided computational tools to capture knowledge within a domain and to express it within a relational framework which can be used by a broad range of clinicians and scientists.¹

The Mouse Anatomy Ontology (MA) was developed to standardize anatomical terms.¹⁰ MA has formal ontological structure built on the kind of framework contained for example in an anatomy reference book¹⁵ but is open and under constant development and

refinement. This is a dynamic process and, as such, MA is updated regularly as users provide input to curators. A textbook on comparative microscopic anatomy of the mouse and human is currently being written (P. Treuting, corresponding editor) and similar ones on embryology are now available.¹² These will provide further extension of comparative anatomical nomenclature.

Pathology nomenclature for the mouse has been captured in the form of an ontology called the Mouse Pathology Ontology (MPATH). This was built using the expertise of the Mouse Pathology Ontology Consortium, a group of 20 D.V.M. and M.D. pathologists and biologists who meet regularly to review and update the ontology. It can be browsed on-line using the click/expand hierarchy browser, searched on-line using free text expressions, or downloaded as an OBO format human readable flat file from the Pathbase site (http://eulep.pdn.cam.ac.uk/Pathology_Ontology/mpath1/mpath_obo_2002_07_26.txt) or from the OBO Foundry repository (<http://www.obofoundry.org/>). The downloaded OBO flat file version may be imported into ontology-aware software or other databases.

MPATH is the only ontology currently available to describe the full range of mouse histopathology in a formally-structured way, with subsumption hierarchies allowing accurate inference. It is segmented into aspects of pathology that are familiar to traditionally trained pathologists. The most current release is fully-defined and contains terms for all the major classes (594 to date) of pathological lesions and processes, relevant to the mouse. Many tissue responses are common to multiple anatomical sites and, as far as possible, the redundancy of specifying a particular response in multiple tissues is avoided by the curatorial creation of cross-products with an appropriate anatomy ontology such as MA, the mouse adult anatomy. The use of cross-products prevents the combinatorial explosion which causes “ontology bloat” in poorly structured ontologies – the inclusion in the ontology of all possible pre-composed variations of instances of an entity. For details of the semantic issues surrounding the description of mammalian phenotypes see reference⁹. Other controlled vocabularies for rodent pathology (eg. that being derived from the recent INHAND initiative of the international societies of toxicopathology),²⁹ have inappropriate content and structure for studies of mutant mice – particularly the kind of phenotyping studies used in functional genomics – and, whilst it is vital to ensure that MPATH can cover the terms in such glossaries, they do not have the computational reasoning power of ontologies needed for informatics.

The definitions used in MPATH range from the simple format found in Stedman’s Medical Dictionary,²¹ to literature references, for example the papers of the NCI Mouse Models of Human Cancer Consortium, and where required, the MPATH consortium created or modified definitions itself, where appropriate or necessary. Synonyms and obsolete terms are included to allow cross-referencing as the ontology evolves. Further development is currently being undertaken to include terms from human pathology to allow MPATH to be used to facilitate integration of human and mouse disease descriptions.

Conclusions

Biology has become Big Science and with the advent of the large scale systematic mutagenesis and phenotyping programs worldwide the place hitherto occupied by high energy physics has been filled by the new biology. As part of this, pathology, as a discipline, has a crucial role in advancing our understanding of gene function in health and disease and pathologists are now key players in this global endeavor. This responsibility presents challenges, those of training, modernization, and standardization, but most of all it requires coordination to make the most of efforts distributed worldwide. We need new resources and we need a new world view in veterinary pathology if we are to play the important role we

have been assigned and learning to use the new resources and informatics is a challenge to which the profession must be sure to rise.

Acknowledgments

This work was supported by grants from the European Commission (Contract numbers QLRI-1999-00320; PATHBASE and LSHG-CT-2006-037188; EUMODIC), North American Hair Research Society, the National Institutes of Health (CA089713, RR17436, AR49288), and the Ellison Medical Foundation.

References

1. Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. *Nat Rev Genet* 2004;5:213–222. [PubMed: 14970823]
2. Barthold SW, Borowsky AD, Brayton C, Bronson R, Cardiff RD, Griffey SM, Ince TA, Nikitin AY, Sundberg JP, Valli VE, Ward JM. From whence will they come? - A perspective on the acute shortage of pathologists in biomedical research. *J Vet Diagn Invest* 2007;19:455–456. [PubMed: 17609364]
3. Bell SM, Kelly SA, Hoyle JA, Lewis FA, Taylor GR, Thompson H, Dixon MF, Quirke P. c-Ki-ras gene mutations in dysplasia and carcinomas complicating ulcerative colitis. *BrJCancer* 1991;64:174–178.
4. Bogue MA, Grubb SC, Maddatu TP, Bult CJ. Mouse Phenome Database (MPD). *Nucleic Acids Res* 2007;35:D643–D649. [PubMed: 17151079]
5. Bolon B. Internet resources for phenotyping engineered rodents. *ILAR J* 2006;47:163–171. [PubMed: 16547373]
6. Cardiff RD. Pathologists needed to cope with mutant mice. *Nature* 2007;447 528–528.
7. Cardiff RD, Ward JM, Barthold SW. 'One medicine---one pathology': are veterinary and human pathology prepared? *Lab Invest* 2008;88:18–26. [PubMed: 18040269]
8. Collins FS, Rossant J, Wurst W. A mouse for all reasons. *Cell* 2007;128:9–13. [PubMed: 17218247]
9. Hancock JM, Mallon AM, Beck T, Gkoutos GV, Mungall C, Schofield PN. Mouse, man, and meaning: bridging the semantics of mouse phenotype and human disease. *Mamm Genome* 2009;20:457–461. [PubMed: 19649761]
10. Hayamizu TF, Mangan M, Corradi JP, Kadin JA, Ringwald M. The adult mouse anatomical dictionary: a tool for annotating and integrating data. *Genome Biol* 2005;6:R29. [PubMed: 15774030]
11. Ince TA, Ward JM, Valli VE, Sgroi D, Nikitin AY, Loda M, Griffey SM, Crum CP, Crawford JM, Bronson RT, Cardiff RD. Do-it-yourself (DIY) pathology. *Nat Biotechnol* 2008;26:978–979. [PubMed: 18779800]
12. Kaufman, M.; Nikitin, A.; Sundberg, JP. Comparative embryology of the endocrine system. Boca Raton: CRC Press; 2010.
13. Krupke DM, Begley DA, Sundberg JP, Bult CJ, Eppig JT. The Mouse Tumor Biology database. *Nat Rev Cancer* 2008;8:459–465. [PubMed: 18432250]
14. Paunesku D, Paunesku T, Wahl A, Kataoka Y, Murley J, Grdina DJ, Woloschak GE. Incidence of tissue toxicities in gamma ray and fission neutron-exposed mice treated with Amifostine. *Int J Radiat Biol* 2008;84:623–634. [PubMed: 18661379]
15. Popesko, P.; Rajitov, V.; Hork, J. A Color Atlas of Anatomy of Small Laboratory Animals. Vol. vol. 2. London: Wolfe Publishing, Ltd.; 1992.
16. Rosenthal N, Brown S. The mouse ascending: perspectives for human-disease models. *Nat Cell Biol* 2007;9:993–999. [PubMed: 17762889]
17. Schofield PN, Bard JB, Booth C, Boniver J, Covelli V, Delvenne P, Ellender M, Engstrom W, Goessner W, Gruenberger M, Hoefler H, Hopewell J, Mancuso M, Mothersill C, Potten CS, Quintanilla-Fend L, Rozell B, Sariola H, Sundberg JP, Ward A. Pathbase: a database of mutant mouse pathology. *Nucleic Acids Res* 2004;32:D512–D515. [PubMed: 14681470]

18. Schofield, PN.; Bard, JBL.; Rozell, B.; Sundberg, JP. Computational pathology: challenges in the informatics of phenotype description in mutant mice. In: Sundberg, JP.; Ichiki, T., editors. Handbook on genetically engineered mice. Boca Raton: CRC Press; 2005. p. 61-81.
19. Schofield PN, Brown SDM, Sundberg JP, Warren M, Dubus P, Ellender M, Fiette L, Rozell B, Quintanilla-Martinez L, Raspa M, Song J-Y, vanderValk M, Arends M, McKerlie C. PRIME importance of pathology expertise. *Nat Biotechnol* 2009;27:24–25. [PubMed: 19131991]
20. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007;25:1251–1255. [PubMed: 17989687]
21. Stedman. Stedman's medical dictionary for the health professions and nursing. 5th ed.. Lippincott, Williams and Wilkins, London: 2005.
22. Sundberg BA, Schofield PN, Gruenberger M, Sundberg JP. A data capture tool for mouse pathology phenotyping. *Vet Pathol* 2009;46:1230–1240. [PubMed: 19605915]
23. Sundberg JP, Hackman RC, HogenEsch H, Nikitin AY, Ward JM. Training Mouse Pathologists: Five years of Pathology of Mouse Models of Human Disease Workshops. *Toxicol Pathol* 2007;35:447–448. [PubMed: 17474065]
24. Sundberg JP, Haschek WM, Hackman RC, HogenEsch H. Developing a comprehensive mouse pathology program. *Comp Med* 2004;54:615–619.
25. Sundberg JP, Schofield PN. A mouse by any other name. *J Invest Dermatol* 2009;129:1599–1601. [PubMed: 19521404]
26. Sundberg JP, Sundberg BA, Schofield PN. Integrating mouse anatomy and pathology ontologies into a diagnostic/phenotyping database: tools for record keeping and teaching. *Mammalian Genome* 2008;19:413–419. [PubMed: 18797968]
27. Sundberg JP, Ward JM, Schofield PN. Where's the mouse info? *Vet Pathol* 2009;46:1241–1244. [PubMed: 19605899]
28. Tapio S, Schofield PN, Adelman C, Atkinson MJ, Bard JL, Bijwaard H, Birschwilks M, Dubus P, Fiette L, Gerber G, Gruenberger M, Quintanilla-Martinez L, Rozell B, Saigusa S, Warren M, Watson CR, Grosche B. Progress in updating the European Radiobiology Archives. *Int J Radiat Biol* 2008;84:930–936. [PubMed: 19016141]
29. Vahle J, Bradley A, Harada T, Herbert R, Kaufmann W, Kellner R, Mann P, Pyrah I, Rittinghausen S, Tanaka T. The international nomenclature project: an update. *Toxicol Pathol* 2009;37:694–697. [PubMed: 19638442]
30. Virchow, R. Cellular pathology as based upon physiological and pathological histology. 2nd ed.. New York: R M DeWitt; 1860. p. 554
31. Yuan R, Tsaih SW, Petkova SB, de Evsikova CM, Xing S, Marion MA, Bogue MA, Mills KD, Peters LL, Bult CJ, Rosen CJ, Sundberg JP, Harrison DE, Churchill GA, Paigen B. Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels. *Aging Cell* 2009;8:277–287. [PubMed: 19627267]

Figure 1.

Pathbase image annotations. Each image is annotated to provide information on the animal in which the lesion originated and the lesion itself. The gene, allele name, and strain name are constrained free text – i.e. using standard nomenclature. Genotype status and type of mutation are drawn from short controlled vocabularies (CV) and the remainder use specified ontologies. There is also additional provision for a free text annotation to provide other non-structured information.