Web-based access to mouse models of human cancers: the Mouse Tumor Biology (MTB) Database

Carol J. Bult*, Debra M. Krupke, Dieter Näf, John P. Sundberg and Janan T. Eppig

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

ABSTRACT

The Mouse Tumor Biology (MTB) Database serves as a curated, integrated resource for information about tumor genetics and pathology in genetically defined strains of mice (i.e., inbred, transgenic and targeted mutation strains). Sources of information for the database include the published scientific literature and direct data submissions by the scientific community. Researchers access MTB using Web-based query forms and can use the database to answer such questions as 'What tumors have been reported in transgenic mice created on a C57BL/6J background?', 'What tumors in mice are associated with mutations in the *Trp53* gene?' and 'What pathology images are available for tumors of the mammary gland regardless of genetic background?'. MTB has been available on the Web since 1998 from the Mouse Genome Informatics web site (http://www.informatics.jax.org). We have recently implemented a number of enhancements to MTB including new query options, redesigned query forms and results pages for pathology and genetic data, and the addition of an electronic data submission and annotation tool for pathology data.

INTRODUCTION

The laboratory mouse has long served as an important animal model for human disease because it is known to resemble humans physiologically, is highly similar to humans in both genome content and organization, is well characterized genetically and is easily manipulated experimentally (1). Developing mouse models that accurately reflect the genetics and histopathology of human cancers was recognized in 1998 as an 'exceptional opportunity' by the National Cancer Institute (http://www.nci.nih.gov) (2). Mouse models provide the means to explore genetic and cellular aspects of disease progression and to test therapeutic strategies that might ultimately be used clinically in humans (2,3).

Different inbred strains of mice vary in their intrinsic tumor susceptibility. Standard inbred mice are not usually appropriate models for human cancers because of the relatively low frequency and late onset of sporadic cancers in mice. However, knowing the characteristic cancer 'profile' of a particular genetic background is critical to the process of selecting the appropriate mouse strain for developing transgenic or targeted mutation mice whose disease progression patterns may be more useful for modeling genetic and molecular aspects of a specific human disease. Much of the data about tumor susceptibility and resistance in genetically defined strains of mice (i.e., inbred lines, transgenics, targeted mutation strains) are not available in a format that allows researchers to compare different strains of mice to one another or to compare the cancer profile of a standard inbred strain to that of a transgenic or targeted mutation line created on the same inbred background. Integrating diverse data about genetics and pathobiology for genetically defined strains of mice in a queryable database system is the primary mission of the Mouse Tumor Biology (MTB) Database (4,5).

In a recent survey of Web-based resources for cancer genetics research, we identified over 70 databases and information resources related to basic cancer genetics research (6). The majority of existing cancer-related resources and databases focus on single genes or specific cancer syndromes. Only a handful of the sites we surveyed provided information about mouse models of human cancers; even fewer sites provided detailed information about the pathobiology of laboratory mice. The MTB Database is unique among existing resources in both its scope and degree of integration of data about cancer genetics and pathology in laboratory mice.

The MTB Database has been accessible via the World Wide Web since 1998 (5). The primary data types represented in MTB are tumor types, mouse strain, genetics, pathology and references (both published and unpublished references are included in the database). These areas, in turn, represent the main Web-based forms that are used to query the database. MTB is an extension of the informatics infrastructure developed for representing genetic and biological information about the laboratory mouse established by the Mouse Genome Informatics (MGI) Group at The Jackson Laboratory (http://www.informatics.jax.org). The nomenclature used in MTB for genes and strains of mice comes from the official mouse nomenclature represented in the Mouse Genome Database (http://www.informatics.jax.org/mgihome/nomen) (7). Anatomical terms used in the database come from a controlled vocabulary of mouse anatomy supported by the Gene Expression Database (GXD) (8). Much of the pathology and diagnostic terminology used in MTB comes from the *Pathobiology of the Aging Mouse* (9), a standard mouse pathology text.

*To whom correspondence should be addressed. Tel +1 207 288 6000; Fax +1 207 288 6131; Email: cjb@informatics.jax.org

ENHANCEMENTS TO MTB IN 2000

Details concerning the design and implementation of MTB have been described elsewhere (4,5). The purpose of this report is to describe new features and recent enhancements to MTB. The most common input we received from our database users during the past year was to provide additional query options and reports for pathology and genetic data. Users also requested that we redesign some of the data summary pages so that they did not have to follow as many hypertext links to retrieve the information they were seeking. The details of the changes to the system in response to user feedback are described below. Screen shots and web links illustrating these changes can be viewed in the online version of this article (Supplementary Material).

New query options for tumor type searches

We have implemented two enhancements to the query options for tumor types. First, we added the capacity to search the database by anatomical system instead of just by organ name. Users can now submit queries such as 'Retrieve all information from MTB for tumors of the Digestive System'. Second, we added support for constraints on queries based on the status of metastases of a tumor. It is now possible, for example, to search for tumors of the mammary gland that are known to metastasize to the lung or the liver.

Enhancements to pathology queries

In the previous versions of MTB, users could only query for and view photomicrographs and diagnostic descriptions for specific strain/tumor combinations (e.g., 'Show me all mammary gland adenocarcinomas for FVBN-TgN(MMTVPyVT)634Mul female transgenic mice'). In the October 2000 release of the database, we added new query forms to allow users to search for pathology data by more general criteria, including organ system (e.g., 'Retrieve all pathology images for tumors of the liver regardless of strain'), strain name (e.g., 'Retrieve all available pathology images for tumors in BALB/c mice regardless of organ system or type of tumor') and tumor type (e.g., 'Retrieve all available pathology data for mammary gland adenocarcinomas regardless of strain'). With this enhancement, database users can now generate results for broad queries about the pathology data in MTB with a single query instead of having to retrieve tumor/strain combinations one at a time.

A second enhancement to the representation of pathology data in MTB relates to the display of the histology images themselves. In previous versions of the database, users would mouse click on a thumbnail version of the histology images in a pathology summary page to view a version of the image with higher resolution of the cellular features. The higher resolution image replaced the current window and the user could no longer compare the image with the diagnostic text provided for the image. We have implemented a simple solution to this dilemma in which the user can view the higher resolution image and diagnostic text in a separate window by mouse clicking on a button labeled 'View Large Image' that is below the thumbnail version of the photomicrograph. The separate window displaying the higher resolution image can be resized and closed independently of the main web page.

Enhancements to gene queries

In both the conceptual and logical design of the MTB database we separated the concepts of genetic changes in tumor cells from the genetics associated with the background of a particular strain of mouse. As a result, users interested in querying both strains and tumors represented in MTB by gene name or symbol needed to search the database using two different query forms. We have implemented a new query mechanism that allows users to search strain and tumor information by gene symbol simultaneously. Now, for example, a search of MTB using the gene symbol for the retinoblastoma 1 gene (*Rb1*) will return information both on the strains that carry a targeted, induced or spontaneous allele of the *Rb1* gene, as well as on the tumors that have reported genetic alterations (e.g., point mutations, deletions, etc.) in the *Rb1* gene. The query results for gene symbol searches are returned in two parts. First, a list of the alleles for genes represented in MTB is returned. Second, the associations of the alleles with either tumor and/or strains are indicated along with hypertext links to the appropriate detail pages.

Enhancements to the tumor frequency grid

The MTB tumor frequency grid was introduced in 1999 as a graphical means of querying and displaying complex cancer profile information for families of inbred strains of mice (5). The tumor frequency grid includes most of the inbred strains of mice that are being systematically characterized as part of an international collaboration (also known as the mouse 'phenome' project) to establish broad baseline phenotypic data on commonly used and genetically diverse inbred strains of the laboratory mouse (10).

We have made two enhancements to the tumor frequency grid to make it more informative for our users. First, we changed the grid from a three-color coding system to reflect tumor frequency to a five-color system. The display of five colors allows more precise information about tumor frequency to be communicated graphically and has the additional benefit that relative frequencies can be discerned even if the grid is printed out or displayed in black and white. Second, we re-structured the listing of the strains in the grid from an alphabetical order to groupings that are associated with the genealogical relationships of the inbred strains (11). This organization makes it easier to find strains that might be expected to be more similar in their patterns of cancer susceptibility.

Electronic data submission of pathology information: JaxPath

The primary mechanism for data acquisition for MTB is through regular review of the published scientific literature by a staff of biologists with expertise in cancer biology and mouse genetics. To support electronic submission and community curation of pathology information we have implemented a Web-based resource called JaxPath that is accessible from the MTB home page. JaxPath allows users to submit unpublished pathology images and data to MTB or to add supplementary image data that could not be included in an original publication because of space or cost limitations. Users that submit data to JaxPath are assigned a password that allows them to edit the descriptions and annotations of their contributed images via the Web. Attributions citing the contributor(s) of images and

annotations are displayed on the pathology data summary reports on the Web.

Queries by MTB accession ID

As described in a previous report on MTB (5), each instance of a tumor in the database is represented as a combination of tumor name, strain, sex and organ of tumor origin. This way of organizing information in MTB reflects our underlying assumption that genetic background plays an important role in patterns of tumorigenesis. Each tumor instance in MTB is automatically assigned a permanent, unique accession identifier (these accession IDs are displayed on many of the query results pages) that allows us to unambiguously reference that tumor instance and make stable links to other databases. Users who have identified specific records in the database that they are interested in retrieving on a regular basis can now query MTB directly using the appropriate MTB accession ID (http://tumor.informatics.jax.org).

FUTURE DIRECTIONS

The National Cancer Institute recently formed the Mouse Models of Human Cancer Consortium (MMHCC) as a mechanism to speed up the validation of mouse models for human cancer and to achieve consensus nomenclature for tumor names and diagnoses (http://www.nci.nih.gov/dcb/odhome.htm). Although it is not currently publicly accessible, the MMHCC is developing a database of mouse models for human cancer that will include information about the testing of therapeutic agents and experimental protocols that is outside the scope of MTB. Because many of the strains of mice that will be described in the MMHCC database will also be represented in MTB, a major goal for our database group over the next year will be to link MTB to the MMHCC database. The integration of these two resources will allow researchers to move seamlessly from information about baseline cancer phenotype and genetic information (in MTB) to detailed descriptions of pre-clinical and clinical mouse models, experimental protocols and results of therapeutic trials (from the MMHCC).

ADDRESSES AND USER SUPPORT

The MTB Database can be accessed at the MGI web site (http://www.informatics.jax.org). The MTB database is also available at MGI mirror sites. User support for MTB is available in the form of online documentation, email, fax and phone:

WWW: http://www.informatics.jax.org/mgihome/support/support.shtml
Email: mgi-help@informatics.jax.org
Tel: +1 207 288 6445
Fax: +1 207 288 6132

The MGI Group also maintains a community electronic bulletin board that serves as a discussion and announcement forum for issues relating to the genetics or biology of mouse and rat. The list is archived and can be searched using keywords (http://www.informatics.jax.org/mgihome/lists/lists.shtml).

Anyone may search the archive, although only registered users may post messages to the list. Individuals may subscribe to this service on the Web at the following URL: http://www.informatics.jax.org/mgihome/lists/lists.shtml.

CITATION OF MTB

Users of MTB are encouraged to cite this paper when referring to MTB in a publication. The following format is suggested when referring to specific data obtained from MTB:
Mouse Tumor Biology Database (MTB), Mouse Genome Informatics Group, The Jackson Laboratory, Bar Harbor, Maine, USA. World Wide Web (http://www.informatics.jax.org). [Include the date (month/year) when the data were retrieved.]

SUPPLEMENTARY MATERIAL

Additional information for this article is available via NAR Online. This information includes links to web pages and to screen shots showing the results of specific MTB queries (Table S1).

ACKNOWLEDGEMENTS

REFERENCES

1. Paigen,K. (1995) A miracle enough: the power of mice. *Nature Med.*, **1**, 215–220.
2. DePinho,R.A. and Jacks,T. (1999) Mouse models of cancer: Introductory comments. *Oncogene*, **18**, 5248.
3. Klausner,R. (1999) Studying cancer in the mouse. *Oncogene*, **18**, 5249–5252.
4. Bult,C.J., Krupke,D.M. and Eppig,J.T. (1999) Electronic access to mouse tumor data: the Mouse Tumor Biology Database (MTB) project. *Nucleic Acids Res.*, **27**, 99–105.
5. Bult,C.J., Krupke,D.M., Sundberg,J.P. and Eppig,J.T. (2000) Mouse Tumor Biology Database (MTB): enhancements and current status. *Nucleic Acids Res.*, **28**, 112–114.
6. Bult,C.J., Krupke,D.M., Tennent,B.J. and Eppig,J.T. (1999) A survey of web resources for basic cancer genetics research. *Genome Res.*, **9**, 397–408.
7. Blake,J.A., Eppig,J.T., Richardson,J.E., Kadin,J.A., Bult,C.J. and the Mouse Genome Database Group. (2001) The Mouse Genome Database (MGD): Integration Nexus for the Laboratory Mouse. *Nucleic Acids Res.*, **29**, 91–94.
8. Ringwald,M., Eppig,J.T., Kadin,J.A., Richardson,J.E. and the Gene Expression Database Group. (2000) GXD: a gene expression database for the laboratory mouse: current status and recent enhancements. *Nucleic Acids Res.*, **28**, 115–119. Updated article in this issue: *Nucleic Acids Res.* (2001), **29**, 98–101.
9. Mohr,U., Dungworth,D.L., Capen,C.C., Carlton,W.W., Sundberg,J.P. and Ward,J.M. (1996) *Pathobiology of the Aging Mouse*. International Life Science Institute, Washington, DC, USA, vols 1 and 2.
10. Paigen,K. and Eppig,J.T. (2000) A mouse phenome project. *Mamm. Genome*, **11**, 715–717.
11. Beck,J.A., Lloyd,S., Hafezparast,M., Lennon-Pierce,M., Eppig,J.T., Festing,M.F.W. and Fisher,M.C. (2000) Genealogies of mouse inbred strains. *Nature Genet.*, **24**, 23–25.