Mouse Tumor Biology Database (MTB): enhancements and current status

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

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

Received October 1, 1999; Accepted October 5, 1999

ABSTRACT

The Mouse Tumor Biology Database (MTB) is a Web-based resource that provides access to information on tumor frequency and latency, genetics and pathology in genetically defined mice (transgenics, targeted mutations and inbred strains). MTB is designed to serve as an information resource for cancer genetics researchers who use the laboratory mouse as a model system for understanding human disease processes. Data in MTB are obtained from the primary scientific literature and direct submissions by the research community. MTB is accessible from the Mouse Genome Informatics Web site (http://www. informatics.jax.org). User support is available for MTB via Email at mgi-help@informatics.jax.org

INTRODUCTION

The central organizing principle for the Mouse Tumor Biology Database (MTB) is that patterns of tumorigenesis are often conditioned by genetic background. The types of spontaneous tumors observed in laboratory mice are characteristics of each strain. To reflect this fact in MTB, each unique combination of tumor type and strain of mouse (i.e., genetic background) constitutes a single database record. Diverse information related to tumor biology can be linked to a database record, including information on tumor frequency and latency, metastatic potential, genotype (of the mouse strain and of the tumor), pathology and literature citations. Researchers seeking appropriate mouse models for studying cancer in humans or for the identification of genetic elements that condition cancer susceptibility in mice will find the ability to view biological data according to the genetic background of the mouse strain particularly useful.

Access to the data in MTB is provided using multiple Web-based query forms. Each form corresponds to one of six key information areas in MTB (tumor type, tumor frequency and latency, tumor genetics, tumor pathology, mouse strain and reference). Using one or more of these forms, database users can customize their search criteria. Queries can be broad (e.g., Show me all of the records in MTB for tumors of the mammary gland) or very specific (e.g., Show me all of the records in MTB for FVB strains carrying a human *HRAS* transgene driven by a mouse *Wap* promoter). Regardless of the query form used, the results are displayed so that it is easy to link to the other primary

information areas represented in MTB (when relevant data are available). Certain data in MTB are also linked to other databases. For example, gene names and symbols are linked to the Mouse Genome Database (MGD; 2) and Gene Expression Database (GXD; 3); published references are linked to MEDLINE (via MGD); strain names for mice distributed by The Jackson Laboratory are linked to the JAX Mice database; and some mouse strains used to model breast cancer in humans are linked to the Biology of the Mammary Gland Web site at NIH.

ENHANCEMENTS TO MTB

MTB has been available as a Web-accessible database since October 1998 (1). Since the first release of MTB, we have focused on several areas of database development including data acquisition, tumor pathology image enhancement, data visualization tools, and database-derived information summaries for the types of data in MTB. In this article we describe the enhancements listed above and the current status of the database.

Data acquisition

The published scientific literature serves as the primary source for data in MTB. Journal articles identified as containing information appropriate for the database by the scientific curation staff are subsequently categorized according to the organ, tissue or cell type affected by the primary tumor. The principal focus for MTB currently is on spontaneous solid tumors in genetically defined mice (transgenics, targeted mutations and inbred strains). Data entry is prioritized based on the most common types of tumors observed in humans (1). Although some data related to commonly used tumor cell lines are included in the database, a comprehensive representation of all tumor cell lines is not a priority for the project. Tumor transplantation studies, toxicological research reports and studies on tumor induction are also not a priority for data acquisition for MTB.

Pathology image resource enhancement

In the first release of MTB <50 tumor pathology images were available and all were of mammary gland tumors in different strains of transgenic mice. A major focus for database development during the past year was to increase the number and diversity of tumor histopathologic images and to enhance their Web presentation. With the current release described here, the database contains >300 images for >20 different types of tumors. The image data in MTB are obtained from The Jackson Laboratory Pathology Program. Many of the images in the database are of

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

tumors observed during routine disease surveillance studies of mouse colonies at The Jackson Laboratory. These data often represent unpublished observations and would not normally be widely available to the scientific community.

One of the significant enhancements to the Web presentation of the histopathology information in MTB is the addition of a detailed description of each image [provided by The Jackson Laboratory's veterinary pathologist (JPS)]. Many of these images now have labels to highlight the diagnostic cellular features for a particular type of tumor. The descriptive material and labeling facilitates more accurate interpretation of the pathology images by non-pathologists.

Data visualization

The complexity and level of detail available for tumor biology data in MTB can make it difficult for users to discern general trends and patterns in the information. We are developing various ways to summarize information in the database to facilitate data comparisons. For example, the current release of MTB includes a Tumor Frequency Grid that graphically summarizes tumor frequency data for several inbred strains of mice. The grid is organized with the organ or tissue affected across the top of the grid and inbred strain name along the side of the grid. When a user clicks on a particular cell within the grid, the details of tumor frequency data for the corresponding strain and organ combination are displayed. Tumor frequency data for 33 organs (or tissues) from 23 inbred strains are currently accessible from the grid. The tumor frequency grid as of October 1999 contains only inbred strains of mice. However, it will be expanded in the next release to include transgenic and targeted mutation mice.

To assist in the comparison of tumor frequencies among different strains of mice, each cell in the Tumor Frequency Grid is color coded according to five categories of tumor frequency (zero, observed, low, moderate and high). 'Zero' frequency indicates that tumors of a particular organ or tissue have not been reported in any of the literature currently represented in MTB. A frequency denoted as 'observed' indicates that tumors for a particular organ or tissue have been observed but that the exact tumor incidence or frequency values were not reported. 'Low' frequency includes reported tumor frequencies greater than zero and <20% as well as tumors described as having frequencies described as 'sporadic', 'low' and 'very low'. 'Moderate' frequency includes tumors with reported tumor frequencies between 20 and 50% as well as tumors reported in the literature as occurring at a 'moderate' frequency. 'High' frequency includes tumors with reported frequencies >50% as well as those reported as occurring at 'high' or 'very high' frequencies.

Data summaries

The data curation methods employed by the MTB staff include the comparison of names of tumors, strains and genes used in the published literature to appropriate nomenclature standards. This indexing and data integration process makes it more likely that the interrogation of the scientific literature represented in database will return links to all of the relevant studies. The following are examples of information summaries that can be generated from MTB:

• a list of the tumors that have been reported for a particular strain of mouse,

• a list of the tumors reported in a particular reference (published or unpublished),
• a list of the genes that have been analyzed for a particular type of tumor,
• a list of tumors correlated with a particular gene or mutation type (including chromosome aberrations), and
• a list of tumors that have a user-specified value for incidence or frequency.

Standards for mouse strain and gene names used in MTB are from the International Committee on Standardized Genetic Nomenclature for Mice (4). Standards for tumor names are derived from a published reference for mouse pathobiology (5). Tumor and strain names used in the original publications are retained in MTB as synonyms. Users can search the database using outdated or incorrect nomenclature, but the correct nomenclature will be displayed in the query results. Gene name synonyms are not retained in MTB because this information is curated by the MGD staff and is available from that database (2).

The curation of tumor name nomenclature is particularly challenging because the name of a tumor implies a specific diagnosis based, in part, on the cell type from which a tumor originates. In some cases the synonymy of tumor names is obvious. For example, 'mammary gland adenocarcinoma' is equivalent to 'mammary adenocarcinoma'. In other cases, however, translating a published tumor name to a standardized nomenclature is not possible. For example a tumor name published as 'mammary tumor' cannot be made synonymous with a more informative name without an analysis of the original tumor tissue. As a result, tumor names in MTB are a mixture of specific terms (e.g., 'mammary gland adenoacanthoma') and general descriptive terms (e.g., 'mammary gland tumor').

Another information summary resource available from the MTB home page is a compilation of electronically accessible resources for basic cancer genetics researchers (6). The Web sites listed in this resource are sorted into several categories: Animal Models, Cancer Genetics and Genomics, Pathology, Reagents and Protocols, and Cancer Biology. Over 70 resources are currently listed with hypertext links to their respective Web sites. The list of Web sites is updated on a regular basis.

CURRENT STATUS OF MTB

The public version of MTB is updated weekly. As of October 1999 over 370 literature citations have been curated. From these citations >6500 tumor records have been entered, encompassing >80 types of tumors from >140 different organs, tissues or cell types. These tumors arose on >1800 different genetic backgrounds. Data involving 200 different mouse genes have also been curated. Additionally, the database has >9600 tumor frequency records and >300 pathology images.

FUTURE DIRECTIONS

Data acquisition in the focus areas described in this article will continue to be a major goal of MTB to ensure that the database is as comprehensive as possible. For the next release of the database we will focus on increasing the number of studies describing genes of relevance to mouse tumor susceptibility and those showing allelic changes or expression changes in tumor versus normal tissue.

ADDRESSES AND USER SUPPORT

MTB can be accessed at the Mouse Genome Informatics (MGI) Web site (http://www.informatics.jax.org) but is not currently available at MGI mirror sites.

User support for MTB is available via on-line documentation, Email, fax and phone at:

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

The Mouse Genome Informatics group also maintains a community electronic bulletin board that serves as a discussion and announcement forum for mouse researchers. The list currently has >1300 subscribers. Members of the scientific research community can register for this free service at http://www.informatics.jax.org/support/lists.shtml

CITATION OF MTB

Users of MTB are encouraged to cite this article and to the previous description (1) of the database when referring to MTB. 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. WWW (http://www.informatics.jax.org). [Include the date (month/year) when the data were retrieved and the version of the database.]

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.

REFERENCES

1. Bult,C.J., Krupke,D.M. and Eppig,J.T. (1999) *Nucleic Acids Res.*, **27**, 99–105.
2. Blake,J.A. *et al.* (2000) *Nucleic Acids Res.*, **28**, 108–111 (this issue).
3. Ringwald,M. *et al.* (2000) *Nucleic Acids Res.*, **28**, 115–119 (this issue).
4. Maltais,L.J., Blake,J.A., Eppig,J.T. and Davisson,M.T. (1997) *Genomics*, **15**, 471–476.
5. Mohr,U., Dungworth,D.K., Ward,J., Capen,C.C., Carlton,W. and Sundberg,J. (eds) (1996) *Pathobiology of the Aging Mouse*, Volumes 1 and 2. ILSI Press, Washington, DC.
6. Bult,C.J., Krupke,D.M., Tennent,B.J. and Eppig,J.T. (1999) *Genome Res.*, **9**, 397–408.