The Mouse Tumor Biology Database: A Public Resource for Cancer Genetics and Pathology of the Mouse

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

Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine 04609

Abstract

Developing genetic mouse models for cancer research has been recognized as an “exceptional opportunity” by the National Cancer Institute. The establishment of bioinformatics resources to facilitate access to published and unpublished data on the genetics and pathology of cancer in different strains of the laboratory mouse is critical to developing and using mouse models of human disease. In this article, we review the Mouse Tumor Biology Database (MTB), a public resource for information on cancer genetics, epidemiology, and pathology in genetically defined mice. We outline current content, data acquisition strategies, and query mechanisms for MTB. MTB is accessible on-line at http://tumor.informatics.jax.org.

Introduction

Mice have been used in cancer research for nearly a century (1, 2). Many of the first lines of genetically homogeneous mice were produced in the 1920s for cancer research specifically. For instance, MacDowell and Taylor (3) used a selective inbreeding strategy to establish a strain of mice, C58, that displayed a 90% incidence of leukemia. Today, there are hundreds of inbred strains that carry distinct combinations of alleles in a fixed state and vary greatly in the types of tumors they develop. For instance, AKR/J mice characteristically present with lymphocytic leukemia (4), whereas a related strain, A/J (5), shows a high frequency of lung tumors (6, 7). Inbred mouse strains represent unique genetic resources for the identification of alleles or quantitative trait loci that predispose to oncogenesis. For example, nine colon cancer susceptibility loci, Sccl through Sccl9, that are distributed throughout the mouse genome have been mapped in crosses between tumor-resistant BALB/c or C3H/HeN mice and the susceptible strain, STS/A (8, 9). Inbred strains are also the raw material for the generation of genetically engineered mice (transgenes and gene-targeted mutants) that have become indispensable tools for cancer research. They have already added much to our knowledge of the effects of specific genetic anomalies on the cellular processes related to cancer (10).

Knowing about the inherent cancer phenotype of a particular mouse strain is of utmost importance for the development of new mouse models for cancer research. However, finding this information can be difficult because of the dispersed nature of the literature and inconsistencies in the nomenclature used for describing tumors in mice and designating mouse strains. In response to an NCI initiative, MTB was developed at The Jackson Laboratory as an electronic resource focusing on the genetics, epidemiology, and pathology of cancer in genetically defined strains of mice. The primary objective of MTB is to collect information regarding the effect of genetic background on tumorigenesis in mice. To accomplish this goal, MTB curators integrate published and unpublished data on tumor frequency or incidence, genetics, and pathology for different strains of laboratory mice.

MTB is designed to aid researchers in determining the intrinsic cancer-related phenotypes of different mouse strains, reviewing patterns of mutations in specific tumors, and identifying genes that are commonly mutated across a spectrum of cancers in the mouse (11–13).

Data Acquisition and Curation. The main data source for MTB is the peer-reviewed scientific literature. The curation staff of the Mouse Genome Database (14) and MTB collaboratively review >350 journals each month. In addition, electronic searches of PubMed at the National Library of Medicine (Bethesda, MD) are performed periodically to identify recent articles describing mouse cancer models.

MTB also contains unpublished information obtained from extramural contributors and routine health surveillance studies of The Jackson Laboratory mouse colonies (15).

MTB focuses on providing the most comprehensive coverage possible of tumor pathobiology data associated with inbred strains of mice commonly used in cancer research and recently developed genetic mouse models of cancer. The latter category includes genetically engineered mice, chemically induced mutants (e.g., strains emerging from N-ethyl-N-nitrosourea mutagenesis projects), and spontaneous mutants. MTB does not contain information on tumor cell lines or preclinical trials using tumor xenograft models, because these types of data are available from other web-based resources (Table 1).

Priority for data entry in MTB is given to the types of cancers that are listed by the NCI as being the most prevalent in the United States population (16). As of December 2001, MTB contained 9429 entries of tumors observed in 2393 strains or cohorts of mice, representing a body of 679 publications (Table 2). Since its initial public release in 1998, use of MTB by the scientific community has increased steadily, reaching a current average of ~450 queries/day. Users of the database include scientists from academic and private research institutions, as well as biotech and pharmaceutical companies.

Querying the MTB Database. The major data types in MTB include mouse strains, genes and mutations, tumor types, organs/tissues or cell types of tumor origin, pathology images, and literature frame protein, Cdkn2a, cyclin-dependent kinase inhibitor 2A; MMHCC, Mouse Models for Human Cancers Consortium; Apc, adenomatosis polyposis coli.

Received 8/20/01; accepted 12/21/01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Supported by grant CA89713 and in part by Cancer Core Grant CA34196 from the National Cancer Institute to The Jackson Laboratory.

2 To whom requests for reprints should be addressed, at The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609. Phone: (207) 288-6039; Fax: (207) 288-6132; E-mail: dnat@informatics.jax.org.

3 The abbreviations used are: MTB, Mouse Tumor Biology Database; NCI, National Cancer Institute; Ttp53, transformation related protein 53; p19ARF, alternative-reading frame protein; Cdkn2a, cyclin-dependent kinase inhibitor 2A; MMHCC, Mouse Models for Human Cancers Consortium; Apc, adenomatosis polyposis coli.


5 Internet address: http://www.informatics.jax.org.

6 MTB distinguishes between males and females of a particular strain or cohort of mice to account for genetic and physiological differences between the sexes. The term cohort refers to populations of mice that were not maintained as breeding strains (e.g., F1 siblings of a particular cross), or to individual animals (e.g., chimeras, transgenic founders).
citations. Each of the primary data types is represented on the MTB home page by an icon that is linked to a query form (Fig. 1). MTB supports several different query mechanisms, including browsing using controlled vocabularies and customizable query forms.

Researchers who are primarily interested in tumors of a particular organ or tissue can use the “Quick Organ/Tissue Search” option on the MTB home page (Fig. 1). This tool consists of a scrollable list of tissues or cell types for which database entries are available. A database user simply selects the name of the organ/tissue of interest to retrieve database records that match the selected term. Similarly, the “Genetics Search” option can be used to query MTB for tumors associated with mutations in a particular gene of interest. For instance, a search for Trp53 (p53) will retrieve all of the records of tumors that arose in Trp53-mutant mice and records for other mice in which somatic Trp53 mutations were detected in the tumor tissue.

An overview of the cancer spectrum characteristic of inbred mouse strains can be obtained from a graphical matrix, the Tumor Frequency Grid (Fig. 1). For ease of comparison, strain families are grouped according to their genealogy (5). Color-coded fields indicate the highest reported frequency of spontaneous tumors in a particular organ or tissue. The Tumor Frequency Grid doubles as a query tool; a mouse-click on a colored cell returns a summary table of all of the tumor records pertinent to that particular organ or tissue in the selected strain.

The “Tumor Search” form can be used to perform custom queries. As shown in Fig. 2, the user determines the specificity of the query depending on the number of fields selected from pull-down menus or parameters entered into text fields. This allows database users to cast a broad net (e.g., “show me all records of lymphoblastic leukemias”) or to ask very specific questions (see legend to Fig. 2).

To facilitate database searches, MTB uses standardized nomenclature and controlled vocabularies. The diagnostic terminology comes from standard pathology texts (17–19). Anatomical terms are based on the Mouse Anatomical Dictionary (20) that was developed for the International Committee on Standardized Genetic Nomenclature for Mice. MTB also contains the use of nonstandard genetic nomenclature for database queries. As shown in Fig. 2, a user interested in tumors of p19ARF knockout mice may type the acronym “ARF” into the Strain Name field of the Strain Search form and select “targeted mutation (knockout)” from the Strain Type list. MTB will retrieve all of the records pertinent to mice carrying targeted mutations in the Cdkn2a locus that affect expression of p19ARF.

Data Presentation. In response to a query by a user, MTB initially returns a tabular summary of all records that match the selected query terms and constraints. The number of records displayed depends on the specificity of the query. The user can then select an individual record from the summary table and retrieve a Detail Information page (Fig. 3).

MTB also contains “negative” data. These include records of tumor-free mice that served as negative controls in a particular study and cases where animals carrying a mutation that is associated with human cancer remained disease-free. For example, loss-of-function mutations in the human FANCC gene result in the familial cancer susceptibility disorder, Fanconi anemia, whereas homozygous Fancc knockout mice do not develop malignancies (22, 23).

The high degree of data integration allows the user to retrieve a maximum amount of information from one query. Each tumor record is linked to frequency data, information about the genetics of the mouse strain, age of the animals at diagnosis, and a literature citation. Where applicable, MTB reports the experimental regimen (referred to as “treatment”) used to promote or inhibit tumor formation. Tumor pathology data are represented by annotated histology photomicrographs that are associated with detailed captions authored by veterinary pathologists (Fig. 3). MTB currently contains ~100 images of mouse tumors, some of which represent previously unpublished data obtained from The Jackson Laboratory mouse colonies.

Hypertext links connect MTB to other databases providing supplemental information. For instance, gene and allele symbols are hyperlinked to the Mouse Genome Database, which contains gene mapping and expression data, genomic linkage maps, phenotype

---

**Table 1  Web-based resources relevant to mouse models of cancer**

<table>
<thead>
<tr>
<th>Database Name, Provider</th>
<th>Data scope</th>
<th>Reference</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMHCC Phenotype Database, Mouse Models for Human Cancers Consortium, National Cancer Institute</td>
<td>Phenotypic and preclinical data for mouse strains developed by MMHCC, including availability from the MMHCC Repository. Information regarding tumor cell lines, xenograft models, and experimental protocols.</td>
<td></td>
<td><a href="http://emice.nci.nih.gov">http://emice.nci.nih.gov</a></td>
</tr>
<tr>
<td>Cancer Genome Anatomy Project (CGAP), National Cancer Institute</td>
<td>Data and analysis tools related to genes, gene expression, single nucleotide polymorphisms (SNP), and chromosomal aberrations in mouse and human cancers.</td>
<td>25</td>
<td><a href="http://cgap.nci.nih.gov">http://cgap.nci.nih.gov</a></td>
</tr>
<tr>
<td>HistoBank, NIH</td>
<td>Information on cDNA libraries and BAC clones.</td>
<td>26</td>
<td><a href="http://histology.nih.gov/">http://histology.nih.gov/</a></td>
</tr>
<tr>
<td>Mouse Tumor Biology Database, The Jackson Laboratory</td>
<td>Genetics, epidemiology (tumor frequency), and pathology of tumors in genetically defined mice (inbred strains, spontaneous or induced mutants, transgenics).</td>
<td>11–13</td>
<td><a href="http://tumor.informatics.jax.org">http://tumor.informatics.jax.org</a></td>
</tr>
</tbody>
</table>

---

**Table 2  Number of records in the first release of MTB (October 1998) and the current version**

<table>
<thead>
<tr>
<th>Organ or tissue of tumor origin</th>
<th>Number of MTB records</th>
<th>October 1998</th>
<th>December 2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammary gland</td>
<td>271</td>
<td>1076</td>
<td></td>
</tr>
<tr>
<td>Intestine</td>
<td>10</td>
<td>596</td>
<td></td>
</tr>
<tr>
<td>Prostate gland</td>
<td>0</td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>Uterus</td>
<td>4</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Lymphohematopoietic system</td>
<td>45</td>
<td>1407</td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td>0</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>24</td>
<td>437</td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>14</td>
<td>1431</td>
<td></td>
</tr>
<tr>
<td>Thyroid gland</td>
<td>1</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>1</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>196</td>
<td>3998</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>566</td>
<td>9429</td>
<td></td>
</tr>
</tbody>
</table>

---

7 Tumor frequency data in MTB represent the percentage of affected animals as reported in a particular study. Tumor incidence is a population-level phenomenon that reflects the inherent tendency of a particular mouse strain to develop a specific type of tumor.

8 Internet address: http://www.informatics.jax.org/mgihome/GXD/aboutGXD.shtml.

9 Internet address: http://www.informatics.jax.org/mgihome/nomen.
descriptions, and links to orthologous genes in other mammals. If a mouse strain is available from The Jackson Laboratory or from the MMHCC, hyperlinks lead to strain and availability information from the JAX Mice10 database or the MMHCC Repository,12 respectively. Where appropriate, MTB provides links to tumor type-specific or strain-specific databases, such as the Biology of the Mammary Gland13 web site at the NIH, and the Transgenic Adenocarcinoma of the Mouse Prostate14 web site at Baylor College of Medicine (Houston, TX).

**Pathology Data Submission.** Although expert curation of the literature is the primary data acquisition strategy, MTB also invites direct submissions of tumor pathology photomicrographs from the research community. Because most print journals impose restrictions on the acceptable number of figures per article, researchers are often forced to select only a few representative images. Consequently, much valuable pathology data remains unpublished. By providing scientists with the opportunity to contribute supplementary photomicrographs, MTB seeks to fill this gap. To facilitate the data submission process, we have implemented JaxPath, a web-based editorial tool that allows researchers to review their images and edit annotations and captions on-line. JaxPath records are password-protected and are merged into the publicly accessible database only with explicit permission from the submitting inves-

---

10 Internet address: http://emice.nci.nih.gov.
11 Internet address: http://jaxmice.jax.org/index.shtml.
12 Internet address: http://web.ncifcrf.gov/researchresources/mmhcc/default.asp.
14 Internet address: http://www-tramp-model.cellb.bcm.tmc.edu.
Fig. 2. Interactive query forms allow users to perform broad or highly specific database searches. A, the Tumor Search Form. In this example, the user asks to retrieve records on "prostate gland adenocarcinomas that arose spontaneously in male T-antigen-transgenic mice, with a frequency >40% and that metastasized to the lung." B, the Strain Search Form. Depicted is an example of a database query using the nonstandard term, "ARF." Because the current official gene symbol for the locus encoding p19ARF is Cdkn2a, MTB will return a list of Cdkn2a knockout mice.
Tumor Detail Page

Fig. 3. Example of a Tumor Detail Page reporting colonic polyps observed in Apc<sup>Min</sup> mutant mice under different feeding conditions. Tumor Detail pages contain information on the host strain, genetic aberrations found in the tumor tissue (in this case, loss of heterozygosity at the Apc locus), pathology, and literature citations. Where appropriate, a camera icon links to annotated photomicrographs.

Future Directions. MTB is the most extensive resource of mouse tumor pathobiology data currently available on the web (24). Maintaining and expanding the database through literature data curation is an integral part of The Jackson Laboratory’s Mouse Genome Informatics program. The MTB information content is updated weekly. New tools are being designed for MTB to enhance the representation of quantitative trait loci mapping data, to accommodate large-scale tumor genotype and gene expression profiling data, and to include images produced with in vivo tumor imaging technologies that are being adapted to mice. Hyperlinks will connect MTB with the mouse cancer phenotype database being developed by the MMHCC. To ensure database interoperability, MTB will adopt the controlled vocabularies of diagnostic disease terms that are being evaluated by the MMHCC. Finally, MTB welcomes critical comments and suggestions from the cancer research community to improve data accuracy and user-friendliness.

Acknowledgments

We thank Dr. Jerrold M. Ward (NCI, Bethesda, MD) for helpful advice. For critical reading of the manuscript we thank Dr. Vivienne I. Rebel (Dana-Farber Cancer Institute, Boston, MA), and Drs. Louise M. McKenzie and Antonio J. Planchart (The Jackson Laboratory, Bar Harbor, ME).

References


The Mouse Tumor Biology Database: A Public Resource for Cancer Genetics and Pathology of the Mouse

Dieter Näf, Debra M. Krupke, John P. Sundberg, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/5/1235

Cited articles
This article cites 18 articles, 10 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/5/1235.full.html#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
/content/62/5/1235.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.