 Nucleic Acids Research, 1999, Vol. 27, No. 1 99–105

Electronic access to mouse tumor data: the Mouse Tumor Biology Database (MTB) project

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

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

Received September 1, 1998; Revised October 14, 1998; Accepted October 30, 1998

ABSTRACT

The Mouse Tumor Biology (MTB) Database supports the use of the mouse as a model system of hereditary and induced cancers by providing electronic access to: (i) tumor names and classifications, (ii) tumor incidence and latency data in different strains of mice, (iii) tumor pathology reports and images, (iv) information on genetic factors associated with tumors and tumor development, and (v) references (published and unpublished data). This resource has been designed to aid researchers in such areas as choosing experimental models, reviewing patterns of mutations in specific cancers, and identifying genes that are commonly mutated across a spectrum of cancers. MTB also provides hypertext links to related on-line resources and databases. MTB is accessible via the World Wide Web at http://tumor.informatics.jax.org . User support is available for MTB by Email at mgi-help@informatics.jax.org

INTRODUCTION AND SCOPE OF THE PROJECT

The purpose of the Mouse Tumor Biology (MTB) Database project is to support the use of the mouse as a model system of hereditary and induced cancers. The data in MTB include endogenous tumors that arise out of normal cells in a mouse, typically through multistage pathways, as well as established (often subcutaneous) murine tumors that are serially transplanted or form upon inoculation of cultured tumor cells. MTB does not include data on tumors that arise in humans or other species and subsequently are transplanted in mice. MTB integrates the following primary data types into an electronically accessible public information resource on the World Wide Web:

- curated tumor names and classifications,
- tumor incidence and latency data in different strains of mice,
- tumor pathology reports and images,
- genetic factors associated with tumors and their development, and
- references (published and unpublished data).

MTB is one of several database projects in the Mouse Genome Informatics (MGI) Group at The Jackson Laboratory. Other MGI projects include the Mouse Genome Database (MGD; 1,2) and the Gene Expression Database (GXD; 3,4). MGD and GXD are integrated databases and are available at http://www.informatics.jax.org . MTB is currently a stand-alone, prototype database and is not integrated with either MGD or GXD. This manuscript describes the first release of MTB.

MTB was developed under the general assumption that the genetic background in which a tumor originates and the mode of tumor induction (i.e., whether a tumor arises spontaneously or is experimentally induced) are important factors for studying tumor biology. The organization of the information in MTB directly reflects these underlying assumptions. Each tumor record in MTB is assigned an MTB ID number based on the unique combination of the following information:

- the organ in which the tumor originates,
- the organ affected (for metastatic lesions),
- the general tumor type (e.g., adenocarcinoma, sarcoma, lymphoma, etc.),
- the mode of tumor induction (induced or spontaneous),
- the agent of tumor induction (for induced tumors), and
- the strain and sex of the mouse in which the tumor cells originated (genetic background).

Tumors are classified as ‘induced’ in MTB if the animal is subjected to an experimental protocol designed to test for an association between a treatment and tumorgenicity. Otherwise tumors are classified as ‘spontaneous’. For example, in the absence of any other treatment, tumors that arise in mice that carry an endogenous virus are classified as spontaneous. However, if a viral load is introduced to a mouse that normally does not carry the virus, the resulting tumors are classified as virally induced. Tumors arising against a particular genetic background are classified as spontaneous in MTB, even in the cases of transgenic or targeted mutation (knockout) mice.

Each tumor record can be associated with additional information about one or more of the primary data types. A schematic of the organization of information in MTB is provided in Figure 1.

The MTB query process

Two main search options are available for MTB. A ‘Quick Search’ option is available from the MTB home page that allows users to search the database by an organ or tissue name or by the name of a tumor. Several advanced query forms are also available. The advanced query forms are designed for detailed queries about the primary data types in MTB (see descriptions

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

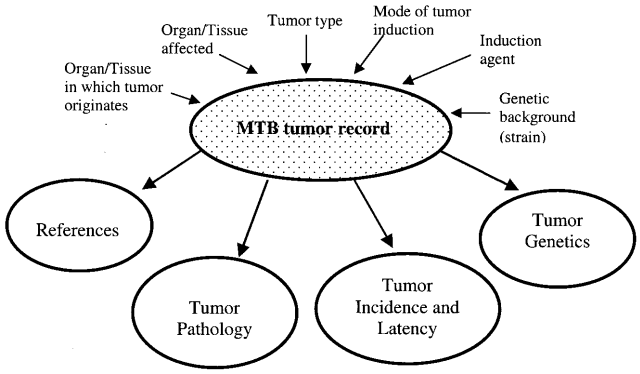


Figure 1. A schematic showing the organization of data in MTB. Each tumor record (stippled oval) is assigned a unique ID in the database based on the unique combination of information shown. Each MTB tumor record can be associated with the information types shown in the open ovals.

below). These forms provide access to multiple fields that users can fill in to refine and restrict their queries.

Query results for tumor types/names using either the advanced forms or the Quick Search option are returned as an MTB Tumor Type Summary Screen. The summary table contains hyperlinks to all of the data in MTB that are available for a particular tumor record (Fig. 2). A common query strategy is to start with a very general query, review the results, and then refine the query so that fewer records are returned. For example, a query on the organ/tissue, mammary gland, returns hundreds of records. A user may wish to refine the query by requesting only those tumors that arise spontaneously or by specifying a specific strain name (e.g., BALB/c, C57BL/6) or strain type (e.g., transgenic, knockout, inbred). If no records that match the user-defined criteria are found, an error message and new query form are returned to the user. When a tumor record in MTB represents only negative or control data (i.e., mice that do not have a specific tumor), the cells in the summary table are shaded gray to distinguish them from other records in the database.

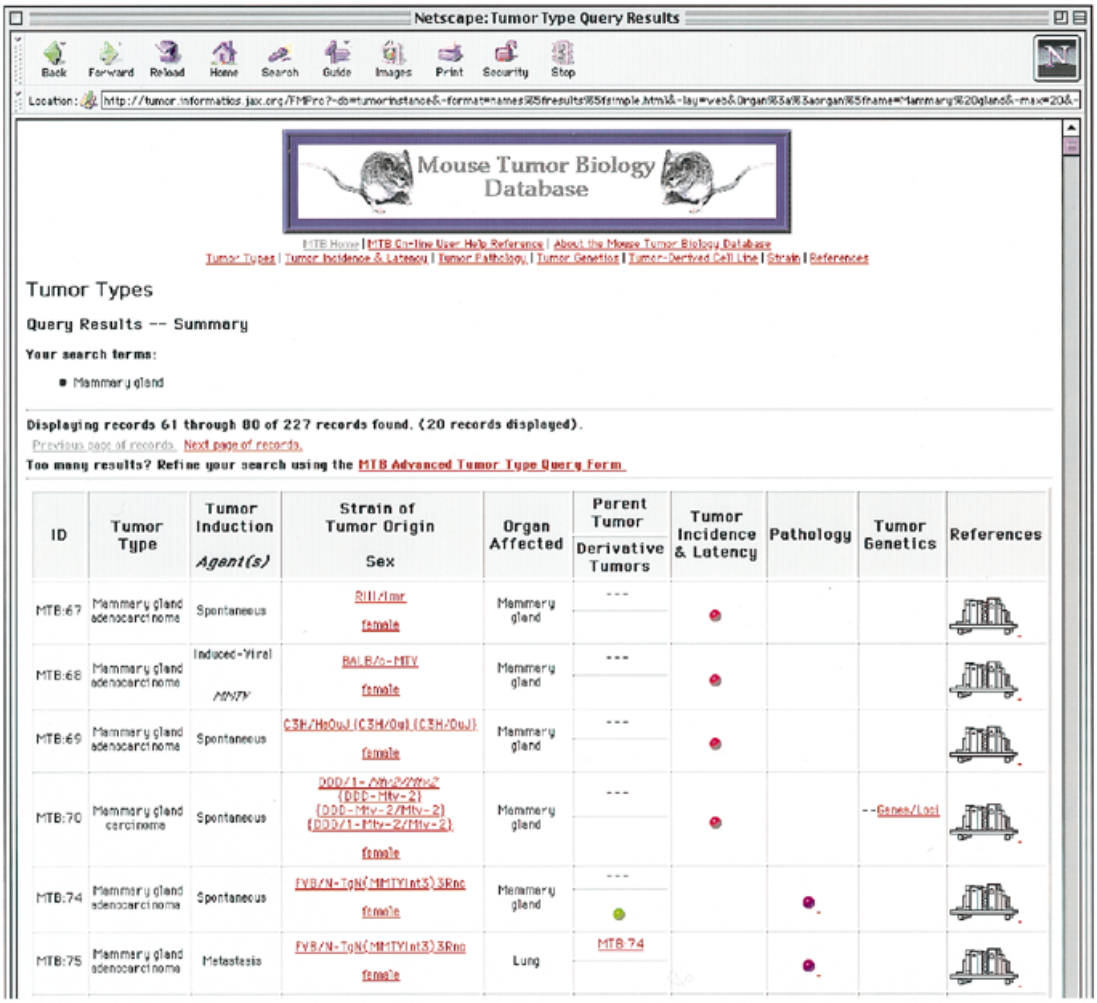


Figure 2. Example of an MTB Tumor Type Summary Screen. If a cell in the summary table contains hyperlinked text or image, the user can mouse-click on the link to view the associated data. In the other database reports within MTB, the MTB ID is hyperlinked to the Tumor Type Summary Screen, facilitating easy navigation between the primary related data types in MTB. Empty cells indicate that no data are available.

A third route of entry into the MTB database is through the listing of database record counts organized by anatomical system, organ/tissue of tumor origin, tumor class, general tumor type, and gene/locus name. Links to the record count summaries are provided on the MTB home page. Clicking on the hyperlinked data in a record count table returns a summary screen for the selected category displayed as shown in Figure 2.

Data sources

Data in MTB are obtained from two sources, the primary literature and locally generated information at The Jackson Laboratory. Journals are scanned for articles relevant to MTB using the same literature triage process that is used for populating the MGI databases (1,2). For MTB, a data curator familiar with cancer biology captures the appropriate information from relevant articles and enters the data into the database. A large body of data on tumor incidence, latency and pathology is generated from the routine animal health screenings of mouse colonies reared at The Jackson Laboratory. These data are valuable for selecting the appropriate mouse models for specific cancers but are often unpublished and are usually not accessible in an electronic, queryable format. Much of the unpublished data in the current version of MTB comes from The Jackson Laboratory Animal Pathology Department and from the laboratories of individual investigators at The Jackson Laboratory. Regardless of the source, data in MTB are reviewed by a data curator to ensure consistency and accuracy of the information.

An electronic data submission process for MTB is planned for the future but is not yet in place. Individual researchers interested in submitting incidence and pathology data to MTB are encouraged to contact support staff by Email at: submissions@informatics.jax.org

Because of the large number of articles that are published describing the use of mice as model systems of cancer, we have prioritized which studies are entered into the database based on the organ/tissue in which a tumor originates. The priority structure is based largely on the rates of cancer in humans as reported by the National Cancer Institute (NCI; http://www.nci.nih.gov) as well as the types of cancers being studied as part of the Cancer Gene Anatomy Project (http://www.ncbi.nlm.nih.gov/ncicgap/). In approximate descending order of cancer rates in humans, the priority for entry into MTB is: lung, breast, prostate, colon, ovary, pancreas, blood, brain, uterus, kidney, stomach, cervix, esophagus, bladder and skin.

PRIMARY DATA TYPES IN MTB

There are six primary data types in MTB that are described in detail below: tumor type (including tumor-derived cell lines), tumor incidence and latency data, tumor pathology, tumor genetics, strains, and references.

Tumor types

Achieving consistency for tumor names is problematic because multiple naming conventions are currently in use in the cancer biology community. As a result, the same general type of tumor can be referred to by multiple names in the literature. In MTB, tumor names are a combination of the name of the organ/tissue in which the tumor cells originate and a general tumor type (e.g., mammary gland adenocarcinoma). The organ/tissue names used in MTB are, in large part, concurrent with the controlled vocabulary for organ and tissue names used to describe gene expression in adult mice in the Gene Expression Database (GXD; see this issue of *Nucleic Acids Research*). The nomenclature for general tumor types in MTB follows the practice and standards of the animal pathology community. To some extent, assigning tumor types in this manner represents a diagnosis based on knowledge of the anatomical structure(s) in which the tumor cells arose. Our convention is that when the anatomical origin of a tumor is not provided or known, a general tumor name will be used. For example, a description of tumors in the mammary gland could be described very specifically as mammary gland adenocarcinomas or generally as mammary gland tumors when the cellular origin of the tumor is unknown or uncertain.

Cell lines are referenced in MTB both by the primary tumor from which the cell line was derived as well as the cell line name. A user can, therefore, query the database for cell lines derived from a specific primary tumor (e.g., mammary gland adenocarcinoma) or by a cell line name (e.g., EMT6, D2, etc.). For those cell lines that have been transfected, the user can also query by the transfected gene and/or promoter carried in the construct.

Tumor name and tumor cell line names in the literature that do not conform to naming standards are recorded as a synonym of the standard name. Users can search the MTB database by non-standard tumor names, but the results returned use the curated name for a specific tumor type. For example, a search using the tumor name, hepatocellular carcinoma, will return records with the curated name, liver-hepatocyte carcinoma.

Relationships between tumors and subcultured cell lines are maintained in MTB and are represented as part of the Tumor Type Summary Screen (Fig. 2). If a record in MTB represents a tumor that is derived from another tumor (or subcultured from another cell line), the parent tumor MTB ID is displayed as hyperlinked text in the 'Parent Tumor' column on the Summary Screen (Fig. 3). If a record in MTB represents a tumor that gives rise to other tumors through such mechanisms as subculturing of cell lines or metastasis of primary tumors, an image appears in the column titled 'Derivative Tumors'. Clicking on the image in the Derivative Tumors column will return a summary table of all of the tumors (or cell lines) derived from a specific parent tumor or cell line (Fig. 3).

Tumor incidence and latency

Strains of mice differ in their predisposition to specific cancers. This predisposition is quantified as tumor incidence (the percentage of a population of mice that are observed to develop a particular type of tumor) and tumor latency (the age at which a strain tends to develop a particular type of tumor). Tumor incidence and latency can, therefore, be considered characteristic of mice having a particular genetic background or mutant gene. Incidence and latency characteristics are important factors to consider in selecting a strain of mouse to use as a model for cancer research. Because resistance to specific cancers is equally important as cancer susceptibility, MTB includes negative incidence and latency information as well as positive data. MTB users have the option of suppressing the display of zero tumor incidence data when they use the advance query form for tumor incidence and latency.

The incidence and latency data in MTB can be queried by a user-specified tumor incidence value (in percent), the age (of a

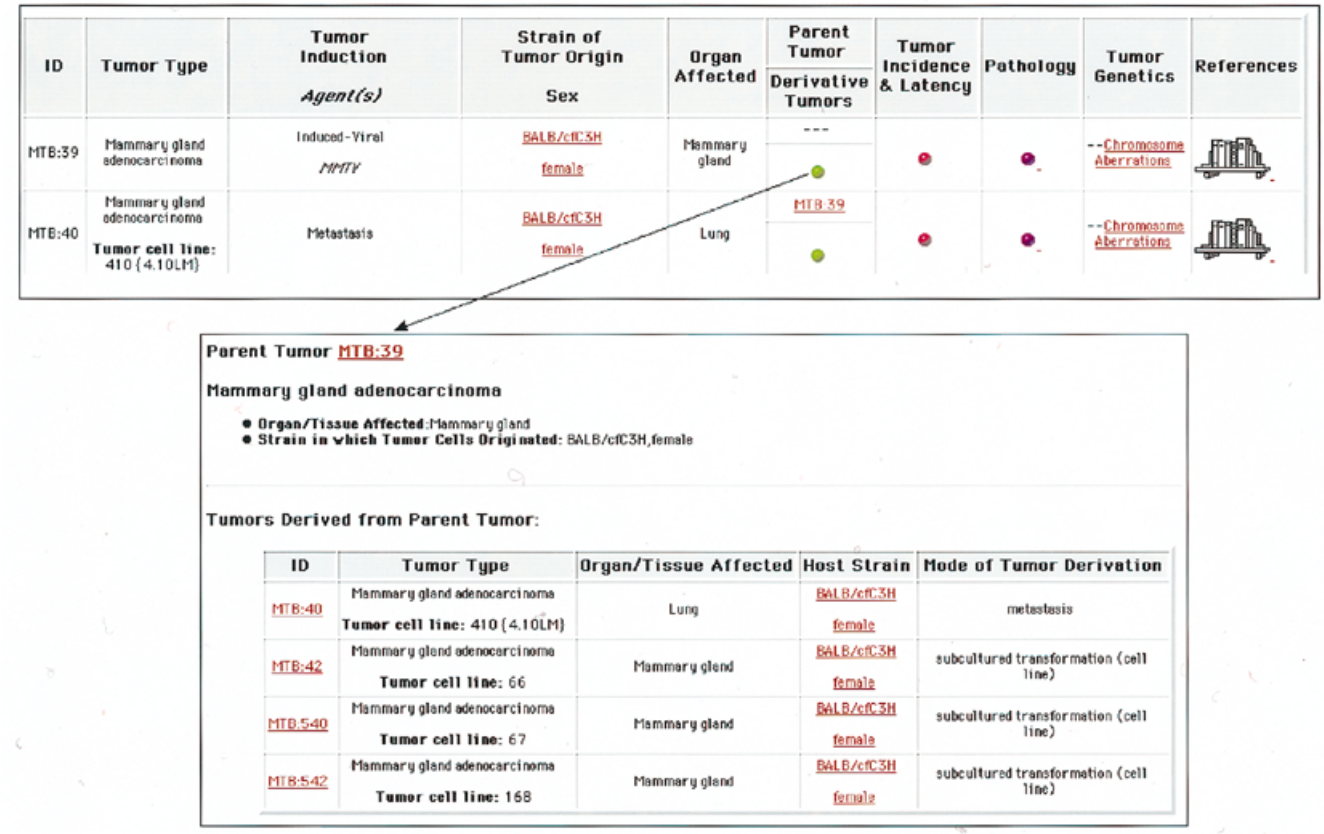


Figure 3. MTB Tumor Type Summary Screen showing links to Parent Tumor and Derivative Tumor data. If the user clicks on the image in the Derivative Tumor column, the list of derived tumors for a particular tumor record is displayed. The arrow points to the results that would be returned if the user mouse-clicked on the green ball graphic in the Derivative Tumors cell indicated.

mouse or a mouse population) at tumor onset, or the age (of a mouse or a mouse population) at tumor detection. Age of onset implies that a researcher systematically surveyed a mouse population to determine the first occurrence of a tumor type. Age of detection is more commonly reported in the literature and implies that tumors were observed when mice were examined or sacrificed at a particular age (Fig. 4).

Tumor pathology

Tumors described in the literature with routine pathology screenings of mouse colonies may have histology images and pathologists' comments or diagnoses associated with them. Most of the images in MTB currently are from health screenings of mouse colonies reared at The Jackson Laboratory. Histological images of tumor cells from the articles in the primary literature are not scanned unless prior approval of the journal is obtained. All images and tumor descriptions are linked to their source to ensure proper attribution of these data. MTB users are encouraged to submit tumor pathology images and descriptions directly to MTB. Information on how to do this can be obtained by sending Email to submissions@informatics.jax.org

Tumor genetics

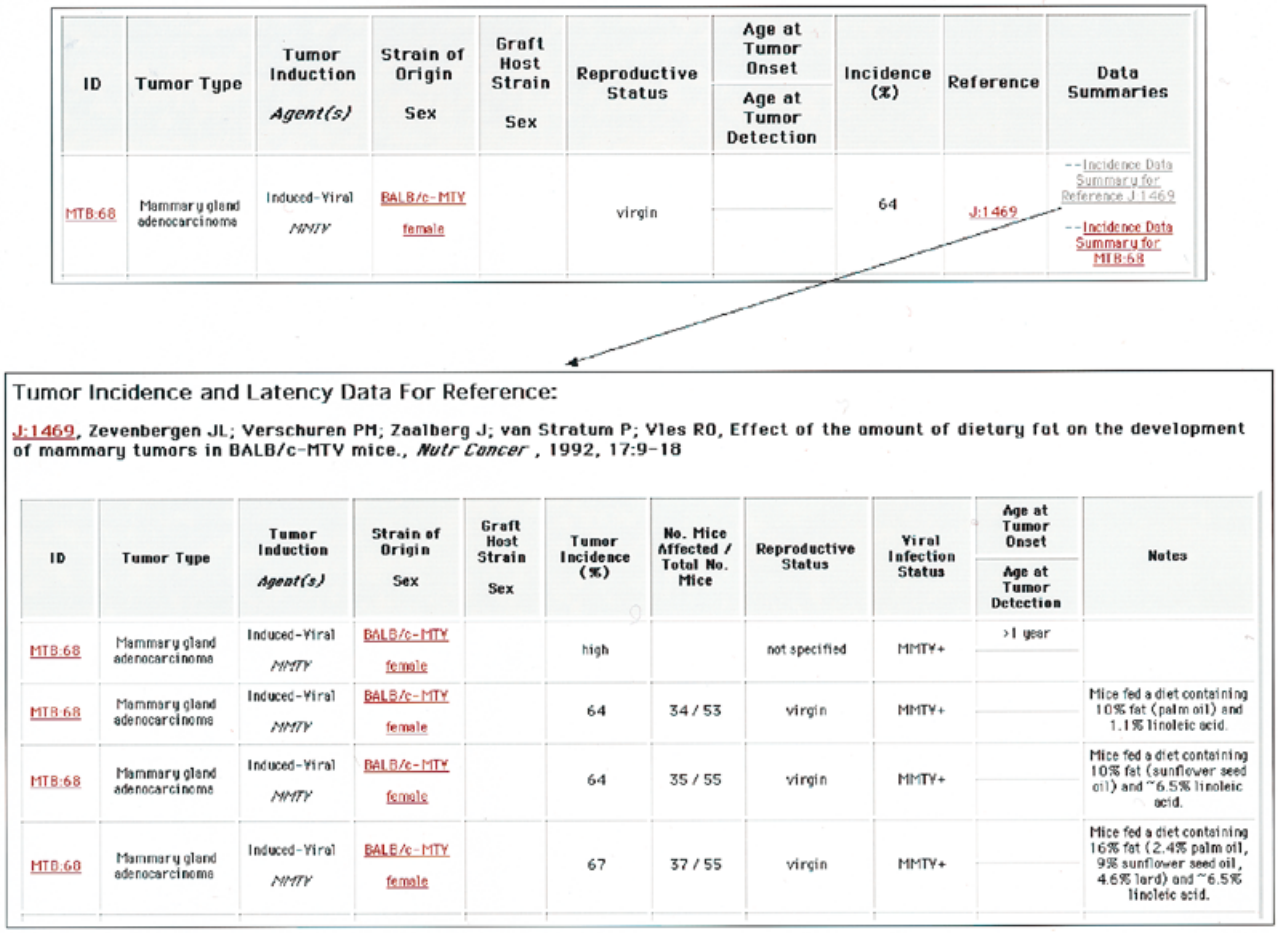
Genetic factors associated with tumor biology that are represented in MTB include the following: (i) alleles and somatic mutations of specific genes or loci, (ii) chromosome aberrations, and (iii) level of gene expression.

In some instances, a specific mutation, allele, chromosome aberration, or change in the level of gene expression is associated directly with a change in tumor biology. In many cases, however, researchers report an observation of one or more of these genetic factors even when the observation cannot be linked directly to a change in tumor behavior or development. Genotype information about the strain of mouse in which a tumor or tumor cell line originated is provided when available and can be accessed by mouse clicking on hyperlinked strain names or from the advanced query form for strains.

Two data summary formats are provided for tumor genetic data in MTB. Users of the database can view all of the genes/loci or chromosome aberrations associated with a specific tumor record or they can view all of the tumor records associated with a particular gene or chromosome aberration (Fig. 5).

Strains

Users of MTB can search for tumor records associated with a particular strain or strain type (e.g., inbred, congenic, transgenic, knockout). The query results from MTB include information on the genotype of the mouse strain and a summary of the tumors reported to be associated with the strain. Transgenic and knockout mice are being used increasingly in cancer research to study the effects of gene expression and mutations in specific genes. These



ID	Tumor Type	Tumor Induction *Agent(s)*	Strain of Origin Sex	Graft Host Strain Sex	Reproductive Status	Age at Tumor Onset / Age at Tumor Detection	Incidence (%)	Reference	Data Summaries
MTB:68	Mammary gland adenocarcinoma	Induced-Viral *MMTV*	BALB/c-MTV female		virgin		64	J:1469	--Incidence Data Summary for Reference J:1469 --Incidence Data Summary for MTB:68

Tumor Incidence and Latency Data For Reference:

J:1469, Zevenbergen JL; Verschuren PM; Zaalberg J; van Stratum P; Vles RO, Effect of the amount of dietary fat on the development of mammary tumors in BALB/c-MTV mice., *Nutr Cancer*, 1992, 17:9-18

ID	Tumor Type	Tumor Induction *Agent(s)*	Strain of Origin Sex	Graft Host Strain Sex	Tumor Incidence (%)	No. Mice Affected / Total No. Mice	Reproductive Status	Viral Infection Status	Age at Tumor Onset / Age at Tumor Detection	Notes
MTB:68	Mammary gland adenocarcinoma	Induced-Viral *MMTV*	BALB/c-MTV female		high		not specified	MMTV+	>1 year	
MTB:68	Mammary gland adenocarcinoma	Induced-Viral *MMTV*	BALB/c-MTV female		64	34 / 53	virgin	MMTV+		Mice fed a diet containing 10% fat (palm oil) and 1.1% linoleic acid.
MTB:68	Mammary gland adenocarcinoma	Induced-Viral *MMTV*	BALB/c-MTV female		64	35 / 55	virgin	MMTV+		Mice fed a diet containing 10% fat (sunflower seed oil) and ~6.5% linoleic acid.
MTB:60	Mammary gland adenocarcinoma	Induced-Viral *MMTV*	BALB/c-MTV female		67	37 / 55	virgin	MMTV+		Mice fed a diet containing 16% fat (2.4% palm oil, 9% sunflower seed oil, 4.6% lard) and ~6.5% linoleic acid.

Figure 4. Results of a query to display all records in MTB for tumor incidence greater than 50% for mammary gland adenocarcinomas in BALB/c virgin female mice. The arrow points to the results that would be returned if a user clicked on the text indicated on the Incidence Summary Screen.

animal models are represented in MTB as types of strains. MTB users can query the database for tumor studies involving transgenic and knockout mice either by the name of the transgenic/knockout strain or by the transgene construct.

References

MTB references derived from the published literature are stored electronically with the citation information for MGD and GXD. On the Web, MTB references are hyperlinked to the MGI Web page that displays reference information. From here, the user can navigate to the Medline record for a citation, when available. For unpublished data, MTB displays the data provider's name, institution, and contact information.

LINKS TO RELATED SITES

MTB is currently linked directly to two related databases at The Jackson Laboratory, the Mouse Genome Database (http://www.informatics.jax.org) and the JAX® Mice Database (http://jaxmice.jax.org). All of the gene and locus names and symbols used in MTB are hyperlinked to the data in the MGD using MGI accession numbers. Clicking on a hyperlinked gene/locus name or symbol will take the user to MGI to obtain additional mapping and phenotype data for a particular gene or locus. If a publication describes tumor biology research using experimental mice distributed through The Jackson Laboratory, a link to the JAX® Mice Database using the strain stock number is provided in MTB.

Links from MTB records to additional related databases, including the Gene Expression Database (GXD), the Transgenic/Targeted Mutation Database (TBASE; 5), and The Jackson Laboratory Breast Cancer Resource are planned for the future.

CONTROLLED VOCABULARIES

Controlled vocabularies are used extensively in MTB to ensure the accuracy and completeness of responses to user queries and to promote links with other databases. A list of controlled vocabularies currently in use for MTB is provided in Table 1. The lists of terms currently in use for each of the vocabularies listed below are available for viewing from the MTB home page.

FUTURE DIRECTIONS

The first version of the Mouse Tumor Biology (MTB) Database prototype was released in the fall of 1998. In addition to our top priority of populating the database, we also are working to incorporate data about (i) experimental protocols used in cancer biology and (ii) Quantitative Trait Loci (QTL) that contribute to tumor susceptibility and latency. Refinements to both the types of

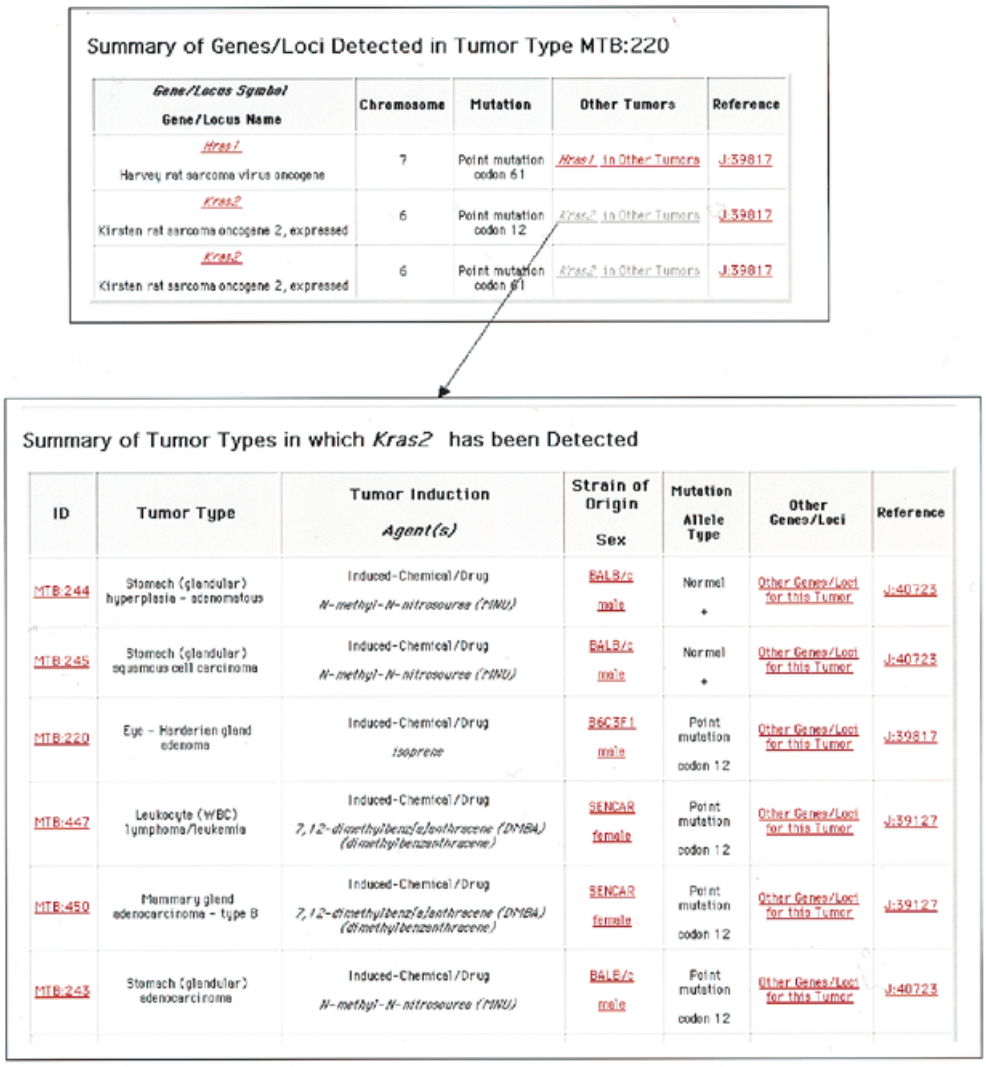


Figure 5. Summary of the genes/loci associated with a specific tumor record in MTB. The arrow indicates the results that would be returned if a user clicked the hyperlinked text indicated.

queries supported and the presentation of data in response to user queries are expected as the database grows and as the experimental approaches to mouse tumor biology change over time.

In the longer term, we will incorporate a mechanism for direct, Web-based, electronic data submission from the mouse cancer biology community. In the meantime, individuals who wish to submit published or unpublished data to MTB are encouraged to contact the authors of this paper or to send Email to: submissions@informatics.jax.org

ADDRESSES AND USER SUPPORT

MTB can be accessed at The Jackson Laboratory at http://tumor.informatics.jax.org

On line documentation and a user help reference are available as Web documents. Users also may contact the User Support staff by Email at: mgi-help@informatics.jax.org , by phone at +1 207 288 6445, or by FAX at +1 207 288 6132.

CITING THE MOUSE TUMOR BIOLOGY DATABASE

Users of MTB are encouraged to cite this paper 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, The Jackson Laboratory, Bar Harbor, Maine, USA. World Wide Web (http://tumor.informatics.jax.org/). [Include the date (month/year) when the data were retrieved and the version of the database.]

IMPLEMENTATION AND DATABASE UPDATES

MTB is implemented as a relational database using FileMaker Pro 4 running on a dedicated server (Power Computing Power Center Pro 210). The database consists of 32 tables that are related to one another using unique identifiers. The Web interface was developed using FileMaker Pro's Web publishing tools (i.e., Web Companion and proprietary Claris Dynamic Markup Language tags). Following a period of evaluating the prototype database, we

Table 1. Controlled vocabularies used in the Mouse Tumor Biology Database

Vocabulary	Description	Comments
Anatomical System	Terms to describe adult organ systems	Based on the GXD adult mouse anatomical dictionary. Examples: Digestive system, Reproductive system, Nervous system, etc.
Organ/Tissue	Specific components of anatomical systems	Based on the GXD adult mouse anatomical dictionary. Examples: mammary gland, liver, bone, lymphocyte, etc.
Tumor Class	A general classification term for a tumor based on the type of tissue in which a tumor originates.	Examples: sarcoma, adenoma, carcinoma, preneoplastic lesion etc.
Tumor Type	A curated name for a tumor or pre-cancerous lesion based on a specific pathology diagnosis	Examples: adenocarcinoma, fibrosarcoma, dysplasia, etc.
Mode of Tumor Induction	Categories for describing how tumors arise	Examples: Spontaneous, Induced-chemical/drug, Induced-viral, etc.
Tumor Agent	A list of the specific agents used to induce tumors	Examples: isoprene, 1,3-butadiene, X-radiation, etc.
Mutation Type	A list of specific types of mutations	Examples: point mutation, deletion, insertion, etc.
Chromosome Aberration	A list of basic types of chromosome aberrations	Examples: trisomy, deletion, chromosome fragmentation, etc.
Gene Expression Level	A general description of the state or rate of transcript production	Examples: up-regulated, down-regulated, expressed, not detected, etc.
Reproductive Status	A description of the hormonal or breeding status of a mouse or mouse population	Examples: virgin, parous, castrated, etc.
Strain Type	A description of the genetic background type of a mouse or mouse population	Examples: inbred, hybrid, congenic, etc

The current versions of these controlled vocabularies can be viewed through http://tumor.informatics.jax.org

plan to migrate MTB to Sybase and to integrate it with other databases in the MGI Group.

At the time this manuscript was written, the data from approximately 100 studies were represented in MTB. Populating the database is top priority in the coming year and users should be aware that MTB presently is not complete with respect to the current literature. The information content of MTB is updated weekly as new data are entered in the curation database. Software changes are released periodically as new database features and modifications to the user interface are implemented. We anticipate one or two software releases each year that will be announced on the MTB home page and via the MGI electronic bulletin board (to subscribe, use the Web form available at http://www.informatics.jax.org/doc/lists.html). Users are encouraged to contact User Support staff to make suggestions or requests about the content of the database and the user interface.

ACKNOWLEDGEMENTS

Pathology images and reports for mammary gland tumors of FVB/N transgenic mice were collected under the auspices of DAMD17-94-J-4016 and contributed by Drs John P. Sundberg and Barbara Tennent. The authors thank the members of the Mouse Genome Informatics Group and Drs Barbara Knowles, Barbara Tennent, Wes Beamer and Carol Linder of The Jackson Laboratory for their input on the design of the MTB user interface and data reports. Drs Barbara Tennent, Carol Linder and Mary Mangan provided valuable comments on earlier versions of this manuscript. The Mouse Tumor Biology Database is supported by a National Cancer Institute contract (97CSX022A) to J.T.E.

REFERENCES

1 Blake,J.A., Richardson,J.E., Davisson,M.T., Eppig,J.T. and the Mouse Genome Informatics Group (1997) *Nucleic Acids Res.*, **25**, 85–91.
2 Blake,J.A., Eppig,J.T., Richardson,J.E., Davisson,M.T. and the Mouse Genome Informatics Group. (1998) *Nucleic Acids Res.*, **26**, 30–137.
3 Ringwald,M., Baldock,R., Bard,J., Eppig,J.T., Kaufmann,M., Nadeau,J.H., Richardson,J.E. and Davidson,D. (1994) *Science*, **265**, 2033–2034.
4 Ringwald,M., Davis,G.L., Smith,A.G., Trepanier,L.E., Begley,D.A., Richardson,J.E. and Eppig,J.T. (1997) *Semin. Cell Dev. Biol.*, **8**, 489–497.
5 Jacobson,D. and Anagnostopoulos,A.V. (1996) *Trends Genet.*, **12**, 117–118.