



Published in final edited form as:

*Cancer Res.* 2017 November 01; 77(21): e67–e70. doi:10.1158/0008-5472.CAN-17-0584.

## The Mouse Tumor Biology Database (MTB): A Comprehensive Resource for Mouse Models of Human Cancer

Debra M. Krupke<sup>1</sup>, Dale A. Begley<sup>1</sup>, John P. Sundberg<sup>1</sup>, Joel E. Richardson<sup>1</sup>, Steven B. Neuhauser<sup>1</sup>, and Carol J. Bult<sup>1,\*</sup>

<sup>1</sup>The Jackson Laboratory, Bar Harbor, ME 04609

### Abstract

Research using laboratory mice has led to fundamental insights into the molecular genetic processes that govern cancer initiation, progression, and treatment response. Although thousands of scientific articles have been published about mouse models of human cancer, collating information and data for a specific model is hampered by the fact that many authors do not adhere to existing annotation standards when describing models. The interpretation of experimental results in mouse models can also be confounded when researchers do not factor in the effect of genetic background on tumor biology. The Mouse Tumor Biology database (MTB) is an expertly curated, comprehensive compendium of mouse models of human cancer. Through the enforcement of nomenclature and related annotation standards, MTB supports aggregation of data about a cancer model from diverse sources and assessment of how genetic background of a mouse strain influences the biological properties of a specific tumor type and model utility.

### Keywords

mouse models; cancer; databases; informatics; curated

### Introduction

Laboratory mice have been used as models for investigation into the genetic basis of human cancers for over a century (1). The laboratory mouse's physiological and genetic similarity to humans combined with the amenability of the mouse genome to precise manipulation (e.g. transgenes, targeted mutations, recombineering, *CRISPR/Cas9*, etc.) contributes to the position of the mouse as the premier animal model for investigating human biology and disease (2, 3). Although the limits of animal models for recapitulation of all aspects of human disease must be appreciated (4, 5), the laboratory mouse remains a powerful genetic model for analyzing a myriad of clinically relevant aspects of cancer biology (6, 7). Genetic mapping studies in mice have been instrumental in identifying genes associated with cancer susceptibility and resistance (8). Key insights into the molecular genetics of cancer initiation

\*To whom correspondence should be addressed: Carol J. Bult, 600 Main St., Bar Harbor, ME 04609, Phone: (207) 288-6324, Fax: (207) 288-6830, carol.bult@jax.org.

Conflict of interest disclosure statement: The authors declare no potential conflicts of interest.

and progression have been discovered using Genetically Engineered Mouse Models (GEMMs) (9).

The Mouse Tumor Biology database (MTB; <http://tumor.informatics.jax.org>) is a comprehensive information resource of mouse models of human cancer. The central mission of the resource is to facilitate the use of the laboratory mouse as a model system for investigating the genetic and genomic factors that underlie human cancers. Evaluating the influence of genetic background on phenotypic variation has long been acknowledged to be critical for the generation and proper utilization of valid mouse models of human cancer (10). To achieve its mission MTB provides the community with expertly curated, semantically consistent data needed to select appropriate strains and mutant mice for experimentation. Expert curation to enforce semantic consistency is central to MTB's operating principles. The rigorous application of gene and strain nomenclature standards is essential for aggregation of data about the same model from diverse sources and for comparing models one to another.

MTB was first released online in 1998 (1) and continues to be a freely available community informatics resource. The database currently contains information on over 46,000 mouse models (representing >82,000 individual tumor frequency records) from nearly 7,000 different cohorts of mice. It also includes >6,500 images, derived from both histopathological studies and cytogenetic assays. MTB's reference collection holds data on nearly 15,000 scientific articles of mouse models of cancer, each indexed to their associated tumor types. Historically the scope of MTB covered GEMM and conventional inbred and hybrid model systems. Renewed interest in xenograft models using immunodeficient and "humanized" mice has led to an expansion of scope to include over 400 Patient-Derived Xenograft (PDX) models. These models present a unique opportunity for studying human cancers *in vivo* and to address clinically relevant questions about therapy response, cancer heterogeneity, and acquired and *de novo* resistance to therapy (4).

## Data Acquisition and Curation

The primary source of data in MTB is the peer-reviewed scientific literature. Data are also obtained from direct researcher submissions and from a variety of cancer research related bioinformatics resources such as PathBase (<http://www.pathbase.net>) and the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>). All data in MTB are reviewed to ensure conformance to official nomenclature for genes, alleles, and mouse strain names (11). Internally developed controlled vocabularies for anatomy, cell type, and tumor type are used to annotate records in the database to ensure that database searches are complete and accurate. Members of the MTB curation team are participating in ongoing community efforts to develop and improve ontologies for these biological concepts. Comments, suggestions, and data contribution from the cancer research community are encouraged. Community feedback can be provided using the User Support feedback form that can be accessed from every web page in MTB.

## Display and Search Features

MTB supports several methods for searching and browsing the database content including faceted search interfaces, interactive tables and graphics, and other web-based search forms. (See Video 1.)

### Model summary table

The model summary table on the MTB home page lists counts of mouse models in the database for the most common types of human cancer (Figure 1). The table displays mouse models where the tumor type is reported at a frequency of 80% in a colony/study size of 20 mice. This allows users of the resource to focus their search on high potential models instead of models where a specific tumor type may appear only sporadically. For example, a lung cancer researcher might begin by choosing the link corresponding to potential lung cancer models in mutant mice. The data summary returned from this link can be narrowed or broadened using the appropriate facet selections. The same lung cancer researcher, for example, could narrow their search by choosing to see only those models with *Kras* mutations. Subsequently they might choose to broaden their search by deselecting the facet for “human lung tissue” thereby opening up the search to display all potential cancer models with a *Kras* mutation.

In addition to inbred and genetically engineered mouse models, MTB also contains data on PDX models distributed by The Jackson Laboratory. Data for these models can also be accessed from the model summary table on the home page. Each PDX model is displayed with its key characteristics including diagnosis, cancer stage and grade, cancer type, summarized genomic data, and therapy response data.

### Interactive graphical data summaries

High-level comprehensive visual summaries are useful to enhance comprehension of trends in large and diverse data sets and for drawing attention to areas of interest or patterns held within those data. The Tumor Frequency Grid presents a high-level summary of all the data in MTB regarding spontaneous tumors in a subset of inbred strains of mice. Patterns of tumor susceptibility can be recognized in the “heat map” pattern of the Grid and the underlying data can be accessed by clicking on a Grid cell. A dynamic iteration of the Grid presents the data in a user-configurable format so that the strains and organs can be refined to reflect the specific interests of the user. MTB’s Cancer QTL Viewer presents a graphical summary of published cancer-related QTL studies in the mouse that are integrated with the rich biological annotations of mouse genes available from the Mouse Genome Informatics (MGI) database (12). Users can upload their own unpublished annotations and display them in the context of published QTL regions.

### Other web-based search forms

Conventional web-based search forms in MTB allow users to craft searches with very specific criteria for tumor characteristics, mouse strain names and types, genes and allele type, and literature references. Researchers can also search the database using human gene symbols, such as a list of genes commonly mutated in a human cancer. The input list is

mapped to corresponding mouse orthologs and a summary table with links to relevant records in MTB is returned. MTB also maintains a search tools to aid researchers with finding relevant genomic data sets for mouse models of human cancer in public repositories such as GEO. To make external genomic data sets about mouse models “findable”, MTB curators associate data sets with semantically consistent metadata for strain name and tumor type.

## Software

The MTB software is designed to be stable, extensible, and cost-effective. The system has four major software components:

1. The public web interface runs on Apache HTTP server and Apache Tomcat 7. The web interface utilizes Java, Struts, JSP, JavaScript, and HTML. Solr is used for faceted searches. The EXT JS framework and Google Charts are used for interactive tables and graphics.
2. The curatorial interface for data entry is implemented as a Java Swing desktop application.
3. The backend is a highly normalized relational database running on Postgres.
4. The MTB Application Programming Interface (API) is implemented as SOAP-based web services. The API allows access to MTB data in a platform and language independent manner that is sufficiently flexible to serve the diverse needs of bioinformaticians.

## Discussion and Future Directions

In an era of rapidly evolving technologies that are generating a grand abundance of cancer models and associated data, MTB has a unique role in ensuring that the investment in the development and characterization of cancer models is realized through the implementation of robust data integration practices and development of tools for model comparison. MTB will continue to provide high-quality, value-added information on different kinds of cancer models in a freely accessible community database.


A major focus for MTB in the future will be to expand from supporting researchers focused primarily on the genetics of cancer to supporting translational and pre-clinical research communities by representing data and information needed for cancer model validation and credentialing (13). Examples of new data sources relevant to genetic models of cancer that will be integrated into MTB include cancer phenotypes identified through the international Knockout Mouse Project (14) and complex trait mapping studies using state of the art genetic resources such as Diversity Outbred and Collaborative Cross mice (15). To enhance the relevance of MTB for translational cancer research, we have partnered with other community informatics resources and with cancer researchers to draft minimal information data standards to describe PDX models. These standards, in turn, will be used in a collaborative effort to develop a web portal to support searches of PDX models across multiple, distributed repositories.

## Acknowledgments

Financial support: National Cancer Institute (CA089713) to C. J. Bult


## References

1. Näf D, Krupke DM, Sundberg JP, Eppig JT, Bult CJ. The Mouse Tumor Biology Database: a public resource for cancer genetics and pathology of the mouse. *Cancer Res.* 2002; 62:1235–40. [PubMed: 11888882]
2. Justice MJ, Siracusa LD, Stewart AF. Technical approaches for mouse models of human disease. *Dis Models Mech.* 2011; 4:305–10.
3. Liu ET, Bolcun-Filas E, Grass DS, Lutz C, Murray S, Shultz L, Rosenthal N. Of mice and CRISPR: the post-CRISPR future of the mouse as a model system for the human condition. *EMBO Rep.* 2017; 18:187–93. [PubMed: 28119373]
4. Kopetz S, Lemos R, Powis G. The promise of patient-derived xenografts: the best laid plans of mice and men. *Clin Cancer Res.* 2012; 18:5160–2. [PubMed: 22912394]
5. McGonigle P, Ruggeri B. Animal models of human disease: challenges in enabling translation. *Biochem Pharmacol.* 2014; 87:162–71. [PubMed: 23954708]
6. Khaled WT, Liu P. Cancer mouse models: past, present and future. *Semin Cell Dev Biol.* 2014; 27:54–60. [PubMed: 24718321]
7. Malaney P, Nicosia SV, Dave V. One mouse, one patient paradigm: new avatars of personalized cancer therapy. *Cancer Lett.* 2014; 344:1–12. [PubMed: 24157811]
8. Drinkwater NR, Gould MN. The long path from QTL to gene. *PLoS Genet.* 2012; 8:e1002975. [PubMed: 23049490]
9. Politi K, Pao W. How genetically engineered mouse tumor models provide insights into human cancers. *J Clin Oncol.* 2011; 29:2273–81. [PubMed: 21263096]
10. Sundberg JP, Roopenian DC, Liu ET, Schofield PN. The Cinderella effect: searching for the best fit between mouse models and human diseases. *J Invest Dermatol.* 2013; 133:2509–13. [PubMed: 23812235]
11. Sundberg JP, Schofield PN. A mouse by any other name .... *J Invest Dermatol.* 2009; 129:1599–1601. [PubMed: 19521404]
12. Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE, Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nuclei Acids Res.* 2014; 42:D810–7.
13. Onken MD, Winkler AE, Kanchi KL, Chalivendra V, Law JH, Rickert CG, et al. A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identifies a transcriptional signature predicting metastatic disease. *Clin Cancer Res.* 2014; 20:2873–84. [PubMed: 24668645]
14. Bradley A, Anastassiadis K, Ayadi A, Battery JF, Bell C, Birling MC, et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm Genome.* 2012; 23:580–6. [PubMed: 22968824]
15. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-resolution genetic mapping using the Mouse Diversity Outbred population. *Genetics.* 2012; 190:437–47. [PubMed: 22345611]



[MTB Home](#) [Help](#)

Your input matters.  
Take our short online survey.



Search for

in these sections



- All Sections
- Tumor
- Organ
- Strain
- Genetics

**Search Forms**

- Tumor
- Strain
- Genetics
- Pathology Images
- Reference
- Advanced
- Search MTB Using Human Genes
- Gene Expression Data Sets

**Additional Resources**

- NEW Faceted Tumor Search
- PDX Model Search
- PDX Comparison
- PDX Login
- Cancer QTL Viewer
- Tumor Frequency Grid
- Dynamic Tumor Frequency Grid
- Other Cancer Websites
- Immunohistochemistry
- Lymphoma Pathology

[Citing These Resources](#)  
[Warranty Disclaimer](#)  
& [Copyright Notice](#)  
Send Questions and  
Comments to [User Support](#).

Last Database Update

### Welcome to the Mouse Tumor Biology (MTB) Database

The Mouse Tumor Biology (MTB) Database supports the use of the mouse as a model system of human cancers by providing access to information on and data for:

- spontaneous and induced tumors in mice,
- genetically defined mice (inbred, hybrid, mutant, and genetically engineered strains of mice) in which tumors arise,
- genetic factors associated with tumor susceptibility in mice,
- somatic genetic-mutations observed in tumors,
- Patient Derived Xenograft (PDX) models

Examples of the data accessible from MTB include:

- tumor frequency & latency data,
- tumor genomic data,
- tumor pathology reports and images,
- associations of models to the scientific literature,
- links additional on-line cancer resources

**A**

Cancer Site	2016 ACS Est. Human Mortality Rank	No. of est. deaths USA 2016	Mouse Models of Human Cancer (restricted to reports where n≥20 mice and tumor frequency≥80%)			PDX Models
			Mutant Strains	Other Strains	All Strains	
Lung and other respiratory	[1]	162,510	42	108	150	74
Lymphohematopoietic	[2]	58,320	125	36	161	6
Colon and other intestine	[3]	51,600	60	4	64	90
Pancreas	[4]	41,780	14	0	14	22
Breast	[5]	40,890	108	28	136	30
Liver and bile duct	[6]	27,170	18	11	29	6
Prostate	[7]	26,120	38	0	38	4
Urinary bladder	[8]	16,390	7	1	8	29
Brain and other nervous system	[9]	16,050	39	0	39	34
Esophagus	[10]	15,690	0	5	5	1
Uterus and cervix	[11]	14,590	4	0	4	3
Kidney and renal pelvis	[12]	14,240	5	0	5	11
Ovary	[12]	14,240	9	0	9	13
Skin	[14]	13,650	19	11	30	19
Stomach	[15]	10,730	11	5	16	2
Oral cavity and pharynx	[16]	9,570	4	1	5	4
Soft tissue including heart	[17]	4,990	22	19	41	30
Gallbladder	[18]	3,710	2	0	2	0
Endocrine system	[19]	2,940	55	2	57	1
Bone and joint	[20]	1,490	3	0	3	12

#### What's new in MTB?

##### 2016

- December
  - MTB now holds over **80,000** tumor frequency records.

**Figure 1.** Screen shot of the MTB Home Page. (A) The model summary table is linked to a faceted search interface (to allow further refinement of the search parameters) and to additional PDX data. Other search options on the home page include (B) traditional, web-based search forms, (C) the faceted search interface, (D) an interactive Cancer QTL Viewer, and (E) the graphical Tumor Frequency Grid.