INTRODUCTION

The laboratory mouse has long served as a model system for investigations into human biology and disease because its physiology is similar to that of humans, because there is a high degree of conservation in genes and genome organization, and because the mouse genome is amenable to precise manipulation (e.g. transgenesis, targeted mutation, recombinant engineering, CRISPR/Cas9, etc.) (1-4). The genetic uniformity of inbred strains contributes to the experimental power of mice to drive discovery of the contributions of specific genes and modifiers for cancer susceptibility and resistance (5). Inbred lines of mice are ideal for identifying and studying low penetrance cancer genes (5,6) which are particularly important for uncovering the genetic basis of susceptibility to cancers in human populations when there is no evidence of familial segregation (7). Studies of mice with induced and engineered changes in specific genes or transgenes have provided fundamental insights into the underlying molecular genetics of cancer initiation and progression (8). Although the limits of model systems for the faithful recapitulation of all aspects of human disease must be appreciated (9,10), the laboratory mouse is widely recognized as the premier animal model for investigating clinically relevant aspects of tumor biology (11-13).

Understanding the influence of genetic background on phenotype variation is critical for the creation of valid mouse models of human cancer and for the appropriate use of these models in cancer research (14). The failure of researchers to appreciate the impact of genetic background can lead to confounding and misleading interpretation of mouse model data. For example, transgenic mice expressing human HRAS on a mixed genetic background of the C57BL/6 and SJL strains were reported to have a mammary gland carcinoma frequency of 45-50% (15). The same transgene allele on an FVB/N strain congenic background resulted in a mammary gland carcinoma frequency of 100% (16). Along similar lines, a recent study reported that both incidence and latency of thymic lymphoma in mouse models varied depending on which genetic background the Atm<sup>imp1Awb</sup> allele was introgressed into (17). The impact of drug treatments for specific tumor types also needs to be evaluated in the context of what is ‘normal’ for the genetic background of the mice in the experiment. For example, a study by Roberts et al. (18) demonstrated the importance of genetic background in using animal models to investigate responses to pharmacological intervention.

One of the major trends in translational cancer research is the renewed interest in xenograft models using immunodeficient mice. Although xenografts have been used in cancer research for decades, progressive improvements in the transplant compliant host animal have greatly enhanced the utility of these models in basic and translational cancer research (19). ‘Human-in-mouse’ xenografts (i.e. Patient-Derived Xenograft models of human cancers)
Derived Xenografts, (PDX) are created by implanting primary human tumor material directly into an immunodeficient host mouse. Several routes for implantation are possible including subcutaneous, under the renal capsule, tail vein injection or orthotopic. PDX models allow researchers to directly study human cells and tissues in vivo (20,21). These models have an advantage over cell lines and cell line xenografts because the tumors retain a more natural architecture and are more reflective of the heterogeneity and histology seen in primary tumors (10). They have an advantage over tumors that arise in genetically engineered mice in that the xenografts retain (at least for a period of time following the initial implantation) a human-specific microenvironment. Several murine hosts are used routinely in xenograft studies, including nude (Foxn1\(^\text{nu}\)) and SCID (Prkdc\(^{scid}\)). However, much of the current state-of-the-art xenograft research in cancer relies on NOD.Cg-Prkdc\(^{scid}\)Il2rg\(^{tm1Wjl}\)/SzJ (NOD scid gamma or NSG) mice (20,22). NSG mice lack mature T cells, B cells and functional Natural killer (NK) cells; they are deficient in cytokine signaling. As with genetic models of human cancer, the genetics of the host strain in xenograft models is important for selecting an appropriate model system. Engraftment success of different host strains is variable and an important factor in experimental design of PDX studies (23). Curating this information in the published cancer xenograft literature is one of the future content enhancements planned for Mouse Tumor Biology (MTB) (see Future Directions, below).

The MTB (http://tumor.informatics.jax.org) database is a freely available community informatics resource designed to support the effective use of the laboratory mouse as a model system for investigating the genetic and genomic basis of human cancer. MTB provides expertly curated, semantically consistent data to guide researchers in the selection of appropriate strains and mutant mice for experimentation. A fundamental biological principle at the heart of MTB’s design and functionality is that the genetic background of mice greatly influences the choice, use, interpretation and translational utility of these models in cancer research. MTB is extending curated content and search functionality beyond the past focus solely on genetic models of human cancer to include mouse models and related data that are important for translational and pre-clinical cancer research. In this paper we describe recent changes to user interface and data acquisition strategies for MTB.

### MTB CONTENT

The MTB database was first released on the World Wide Web in 1998 (24) with a primary focus on genetically engineered mouse models of human cancer. Data in MTB related to these models includes frequency of mouse tumors in the context of specific genetic backgrounds, histopathology images of mouse tumors and information about specific mutations/allelic variants in mouse tumors. As genome sequencing became a common approach for characterizing human tumors (25), we developed new interfaces for MTB to connect lists of human genes commonly mutated in various cancers to mouse models carrying mutations in those genes. As PDX have become increasingly important as a platform for basic cancer research and translational cancer research (26), we have initiated curation efforts to represent this literature and these models in MTB. A summary of the trends in MTB’s data content since the inception of the resource is provided in Table 1.

Nomenclature and annotation standards are essential for scientific communication and for accurate and complete data retrieval and aggregation (27). The nomenclature and semantic standards implemented in MTB benefit researchers by greatly simplifying what can be a frustrating and time consuming task of finding and comparing data about mouse models. For all of the gene, allele, mutation and strain data in MTB, curators enforce the nomenclature standards set by the International Committee on Standardized Nomenclature for Mice (http://www.informatics.jax.org/nomen/) and the Human Gene Nomenclature Committee (28). An example of the importance of standardized genetic nomenclature is illustrated by alleles of the Fgfr3 gene (fibroblast growth factor receptor 3, MGI:955254). Mouse models carrying mutations in Fgfr3 have been observed to promote skin and lung cancer (29), but finding the relevant model using non-standard nomenclature (e.g. Fgfr3\(^{+/−}\) or Fgfr3\(^{+/-}\)) is complicated by the fact that there are over 25 targeted alleles of Fgfr3 published, with 11 of the alleles originating from the same laboratory that published the observation of increased tumorigenesis in lung and skin. Only one of the 11 alleles, Fgfr3\(^{tm1Cxd}\) (MGI:2135675); in combination with a targeted mutation in Ctnnb1 (Ctnnb1\(^{tm1Mmt}\)) and on a mixed genetic background, is associated with increased lung tumorigenesis.

For terminologies associated with tumor classification annotations, MTB adopts standards set by pathologist working groups that are convened on a regular basis to evaluate neoplasias of specific anatomical systems and to de-

---

**Table 1. Annotation trends in MTB since the project’s inception**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated references</td>
<td>142</td>
<td>1164</td>
<td>2594</td>
<td>3730</td>
</tr>
<tr>
<td>Tumor frequency records</td>
<td>3699</td>
<td>23 778</td>
<td>35 832</td>
<td>61 655</td>
</tr>
<tr>
<td>Genetically defined strains</td>
<td>623(a)</td>
<td>3559(a)</td>
<td>3764</td>
<td>6057</td>
</tr>
<tr>
<td>Tumors associated with specific genes</td>
<td>105</td>
<td>800</td>
<td>1330</td>
<td>2288</td>
</tr>
<tr>
<td>Images (histopathology, SKY, FISH, etc.)</td>
<td>—</td>
<td>1318</td>
<td>3882</td>
<td>5886</td>
</tr>
<tr>
<td>PDX models</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>357</td>
</tr>
</tbody>
</table>

\(a\) Male and female mice were stored as separate strain entries until 2004.
Figure 1. (A) Screenshot of MTB’s Cancer QTL Viewer showing the display of all annotated cancer QTL regions for the laboratory mouse displayed in genome context. (B) Links to QTL Detail pages in the MGI database.

develop standard diagnoses and terminologies. Examples of standard classifications for mouse tumors to emerge from these workshops include the lymphohematopoietic system (30), lung (31), mammary gland (32), gastrointestinal system (33), nervous system (34), pancreas (35) and prostate (36). We also work with related bioinformatics resources such as PathBase (37) on mapping terms to ensure broad dissemination of standard tumor classification and diagnostic terminologies for mouse models.

DATABASE ENHANCEMENTS

MTB can be searched using web-based search forms and interactive graphical summaries of mouse strain characteristics related to cancer phenotypes (24,38). Four recent enhancements to data content and user interfaces have been implemented in MTB to further advance the mouse for understanding the genetic and genomic basis of cancer. These enhancements include (i) customizable genome maps of Quantitative Trait Loci (QTL) associated with cancer phenotypes mapped in mice, (ii) search tools for finding data sets associated with cancer genomics studies in laboratory mice in public archives, (iii) search tools for finding mouse models using human gene symbols and (iv) access to information and data associated with PDX models of human cancer.

Cancer QTL in mouse

The laboratory mouse is a powerful genetic model for identifying genes associated with cancer susceptibility and resistance (39). MTB’s Cancer QTL Viewer (Figure 1) provides 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 (40). One common starting point for the interface is to display all mapped cancer-related QTL on a genome-wide map. The QTL regions are color coded according to the organ system/cell type associated with the mapping study. In cases where authors provide the symbol of the genetic marker associated with the peak marker-phenotype association score in a mapping study, the QTL is represented by the location of that marker. When authors provide information on the genetic markers that define the boundaries of the QTL region, the entire range of the QTL region is displayed on the map. The cancer QTL graphic in MTB includes a table with links to QTL details in the MGI database. The Cancer QTL Viewer allows users to upload their own unpublished
A. Search results using pre-compiled human cancer genes

<table>
<thead>
<tr>
<th>Human Gene Symbol</th>
<th>Mouse Gene Symbol</th>
<th>Name</th>
<th>Alleles/Transgenes in Strains</th>
<th>Classes of Tumor Specific Alterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Apc</td>
<td>adenomatosis polyposis coli</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>ATM</td>
<td>Atm</td>
<td>ataxia telangiectasia mutated</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>Cdkn2a</td>
<td>cyclin-dependent kinase inhibitor 2A</td>
<td>22</td>
<td>7</td>
</tr>
<tr>
<td>EGFR</td>
<td>Egfr</td>
<td>epidermal growth factor receptor</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>EPHA3</td>
<td>Epha3</td>
<td>Eph receptor A3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EPHA5</td>
<td>Epha5</td>
<td>Eph receptor A5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ERBB4</td>
<td>Erbb4</td>
<td>v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

B. Specific alleles and transgenes associated with the gene

<table>
<thead>
<tr>
<th>#</th>
<th>Gene</th>
<th>Description</th>
<th>Transgene</th>
<th>Spontaneous mutation</th>
<th>Chemically induced mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Egfr, epidermal growth factor receptor (Human) tetO, tetracycline operator (Regulatory - Promoter)</td>
<td>Tg(tetO-EGFR*delta19)9Hev (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Egfr, epidermal growth factor receptor</td>
<td>Egfrwa2/Egfrwa2 (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Egfr, epidermal growth factor receptor</td>
<td>Egfrwa5/Egfr+ (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. Summary of tumor-specific alterations per gene

<table>
<thead>
<tr>
<th>#</th>
<th>Gene</th>
<th>Description</th>
<th>Transgene</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Egfr, epidermal growth factor receptor</td>
<td>Transgene (16)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>EGFR, epidermal growth factor receptor</td>
<td>Point mutation (5)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Egfr, epidermal growth factor receptor</td>
<td>Insertion (3)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Screenshot of a subset of search results using a list of 26 human genes reported as frequently mutated in human lung adenocarcinoma in Ding et al. (43). Search results provide rapid access to (A) alleles/genotypes in different strains of mice and (B) to the types of genetic variants/alterations that have been observed for the gene in mouse tumors.

Cancer genomics

Genomics technologies are now used routinely to characterize the genomes of primary human tumors and of tumors in genetically defined mouse models. Although many of the genome-scale data from these studies are available from public data archives it can be difficult for researchers to find data sets that include specific mouse strain information or are associated with specific tumor types due to the lack of adherence to standardized gene and strain nomenclature. To support rapid access to mouse cancer genomics data, MTB indexes public genomic data archives such as Gene Expression Omnibus (GEO) (41) and ArrayExpress (42) with standardized nomenclature. The Gene Expression Data Set Search Form in MTB allows users to search for data sets in these archives by organ, tumor classification, strain name and assay platform. The search results allow users to rapidly access study information and the data for each sample in the relevant public archive and are also linked to relevant mouse model data in MTB.

To link the results of large-scale human cancer genomics studies to mouse models, MTB curators have compiled lists of human cancer genes identified from published cancer genome surveys. These lists are provided on the Human Gene Search form in MTB and can be downloaded or used as input to search the MTB for mouse models associated with each of the genes in the list. The search form also allows ad hoc searches of mouse models in MTB using human gene symbols. An example of a search of MTB using the precompiled list of 26 human genes identified as being frequently mutated in a survey of 188 human lung adenocarcinomas (43) is shown in Figure 2.
High concordance of morphologic features of the patient tumor and PO tumor.
Dr. Regina Gandour-Edwards, University of California, Davis


PO tumor #434 for PDX model TM00188 (aka, LG0520P). Nests of pleomorphic squamous cells in desmoplastic stroma.

Figure 3. Examples of PDX characterization data accessed from MTB’s PDX model web portal. (A) Description of the concordance of morphological features of primary patient tumor and engrafted tumors. (B) Circos plot showing integration of genomic data from PDX tumor. (C) Tumor growth characteristics of similar tumors in different PDX models. (D) Tumor growth curves from a drug response study comparing four different treatments in tumor-bearing mice.

PDX

MTB supports online access to all of the available PDX models from The Jackson Laboratory’s PDX Resource. The models in this resource have been comprehensively annotated for clinical information (de-identified) about the patients from whom the primary tumor material was obtained, histopathology and diagnostic marker labeling, whole genome copy number variation (CNV), genome-wide transcriptional profiling and targeted exome sequencing. Subsets of the models have tumor growth data and standard of care drug response data from dosing studies in tumor-bearing mice. When histopathology images of both the primary patient tumor and the engrafted tumors are available, a board certified pathologist reviews the images and provides a summary of the degree to which there is concordance of the morphological features between the primary and engrafted tumors. The user interfaces to the PDX associated data support searches for models by cancer type, patient diagnosis and genomic properties of the engrafted tumor (expression, CNV and/or mutation) (Figure 3). Tabular summaries of variants identified in the engrafted tumors and graphical summaries of gene expression, amplifications and deletions are also provided for each model. Users can request information on the availability of tumor fragments and tumor-bearing mice using a web-based form that is forwarded to the customer services group at The Jackson Laboratory.

FUTURE DIRECTIONS

MTB initially was designed to serve as a centralized resource of information regarding mouse models of human cancer, with a focus on emphasizing how the genetic background of different strains of laboratory mice can influence cancer phenotypes in genetically engineered mice. Mice as genetic models of human cancer will continue to be a focus for the resource and we will integrate two emerging sources of data in this area in the near future. First we will curate cancer-related phenotype information in data emerging from large-scale phenotyping initiatives associated with the international Knockout Mouse Project (44,45). A second source of new data in MTB will come from complex trait mapping studies using new high-precision mapping populations such as Diversity Outbred mice (46).

While mouse model representation will continue to be a primary and unique focus for the MTB database project, many of the planned enhancements to the resource will em-
phasize promoting the use of mouse models in translational and pre-clinical cancer research. We will accomplish this goal by highlighting data used to validate the human cancer relevance of mouse models (i.e. data associated with model ‘credentialing’) and by adding PDX models and data. Specific planned enhancements to MTB include interfaces that highlight the genomic and histopathology concordance of mouse models and human cancers. We also are working with other informatics groups to develop improved interfaces for navigating conserved syntenic regions of the mouse and human genome, enabling better integration of complex trait mapping studies in the mouse with genome-wide association study data in humans. The representation of PDX models in MTB is currently centered on The Jackson Laboratory’s PDX Resource, however we will incorporate information and data from other PDX resources and repositories when possible.

INFRASTRUCTURE

The MTB database system has three major software components:

1) The public web interface runs from version 5.x of the Apache Jakarta Tomcat Web Server and utilizes the Apache Struts Framework. The web interface utilizes Java technology conforming to the 2.4/2.0 Java Servlet/Java Server Pages specifications. We leverage ExtJS and Google Visualizations Javascript libraries for data visualization on the public web interface to MTB.

2) The curatorial interface for MTB is a Java-based desktop application used by the Scientific Curators to enter and curate data. Curator interfaces are built using a rich set of custom Java libraries designed around the Core J2EE Design Patterns.

3) The database itself is a highly normalized relational database currently housed in MySQL. JDBC is used to access the data repository.

The MTB Application Programming Interface (API) facilitates the exchange of data with users and with other database systems. In keeping with our software design goals of ‘write once, run anywhere’, Web Services via SOAP is the primary access to the MTB API suite. This design supports access to MTB data in a platform and language independent manner that suits the differing needs, standards and tools of the community.

USER SUPPORT

The MTB Database can be accessed at the MTB Home Page (http://tumor.informatics.jax.org), which is part of the MGI group web pages (http://www.informatics.jax.org). Announcements about new MTB features are released via the MTB web site as well as the MGI FaceBook page. User support for MTB is available in the form of online documentation, email, fax and phone:

Web: http://www.informatics.jax.org/mgihome/support/support.shtml
Email: mgi-help@jax.org
Tel: +1 207.288.6445
Fax: +1 207.288.6830
FaceBook: https://www.facebook.com/mgi.informatics

DATA USE AND RESOURCE CITATION

MTB data and software are provided freely to the scientific community for promoting education and research activities. Any reproduction or use for commercial purpose is prohibited without the prior express written permission of The Jackson Laboratory.

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://tumor.informatics.jax.org). [Include the date (month/year) when the data were retrieved.]

ACKNOWLEDGEMENTS

We thank Drs Judith Blake and Joel Wagner for helpful comments on earlier versions of this manuscript.

FUNDING

The MTB database is supported by the National Institutes of Health [CA089713]. Research reported in this publication was partially supported by the National Cancer Institute [P30CA034196]. Funding for open access charge: NCI grant CA089713.

Conflict of interest statement. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES


Nucleic Acids Research, 2015, Vol. 43, Database issue D823


