LITERATURE TRIAGE AND INDEXING IN THE MOUSE GENOME INFORMATICS (MGI) GROUP

Randal P. Babiuk, Dale A. Begley, Susan M. Bello, Dirck W. Bradt, Donna L. Burkart, Howard Dene, Alexander D. Diehl, Harold J. Drabkin, Jacqueline H. Finger, Terry F. Hayamizu, David P. Hill, Michelle Knowlton, Debra M. Krupke, Ira Lu, Lois J. Maltais, Monica McAndrews-Hill, Terrence F. Meehan, Li Ni, Hiroaki Onda, Dmitry Sitnikov, Constance M. Smith, Cynthia L. Smith, and Monika Tomczuk.
The Jackson Laboratory, Bar Harbor, Maine, USA 04609

http://www.informatics.jax.org

ABSTRACT

The Mouse Genome Informatics (MGI; http://www.informatics.jax.org) group is comprised of several collaborating projects including the Mouse Genome Database (MGD) Project, the Gene Expression Database (GXD) Project, the Mouse Tumor Biology (MTB) Database Project, and the Gene Ontology (GO) Project. Literature identification and collection is performed cooperatively amongst the groups.

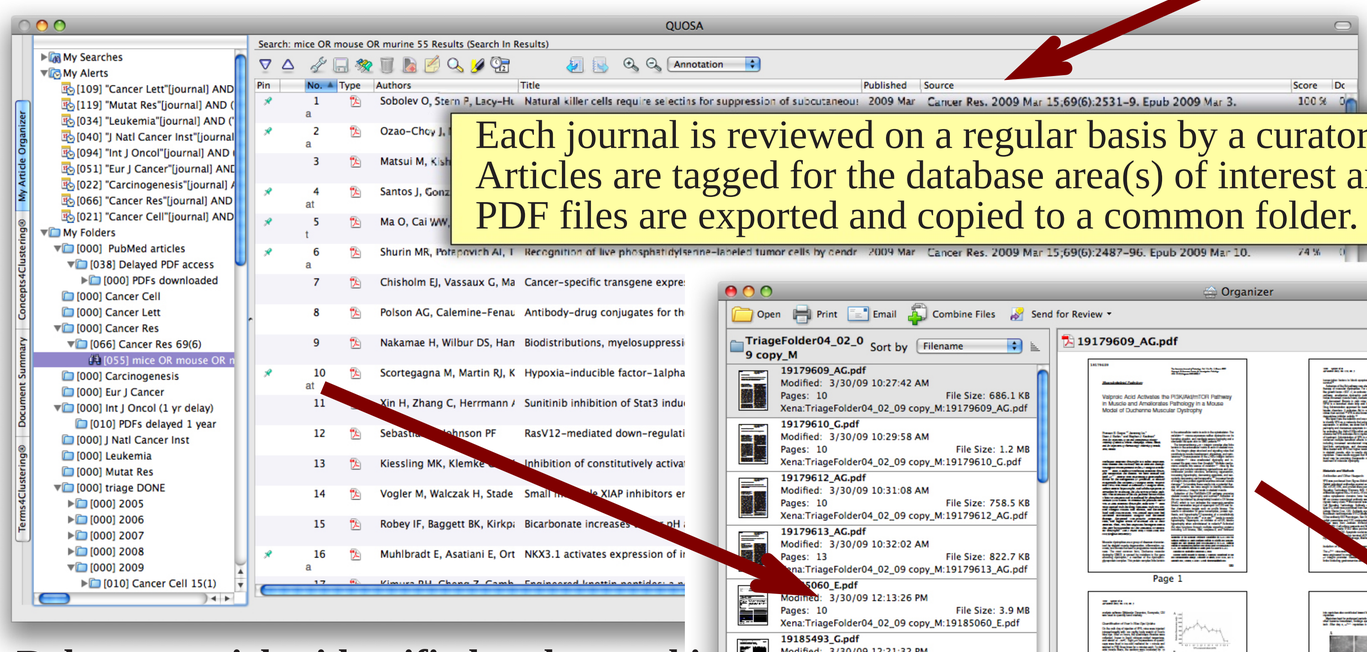
In recent years many institutional libraries have transitioned from a focus largely on print holdings to one of electronic access to journals. This change has necessitated adaptation on the part of the MGI curatorial group. Whereas the majority of journals covered by the group used to be surveyed in paper form, those journals are now surveyed electronically. Approximately 160 journals have been identified as those most relevant to the various database groups. Each curator in the group has the responsibility of scanning several journals for articles relevant to any of the database projects. Articles chosen via this process are marked as to their potential significance for various projects. Each article is catalogued in a Master Bibliography section of the MGI database system and annotated to the database sections for which it has been identified as relevant. A secondary triage process allows curators from each group to scan the chosen articles and mark ones desired for their project if such annotation has been missed on the initial scan.

Once articles have been identified for each database project a variety of processes are implemented to further categorized and index data from those articles. For example, the Alleles and Phenotype section of the MGD database indexes each article marked for MGD and in this indexing process they identify each mouse gene and allele examined in the article. The GXD database indexing process has a different focus. In this case articles are indexed with regard to the stage of development used in the study as well as the assay technique used. In each case the indexing gives an overview of the data held in the article and assists in the more extensive curation performed in the following step of the curation process. Indexing also provides each group with valuable information used to prioritize and streamline the overall curation process.

The MGI projects are supported by NHGRI grants HG000330, HG002273, HG003622, NICHD grant HD033745, and NCI grant CA089713.

~160 Journals Reviewed

Each journal is reviewed on a regular basis by a curator. Articles are tagged for the database area(s) of interest and PDF files are exported and copied to a common folder.

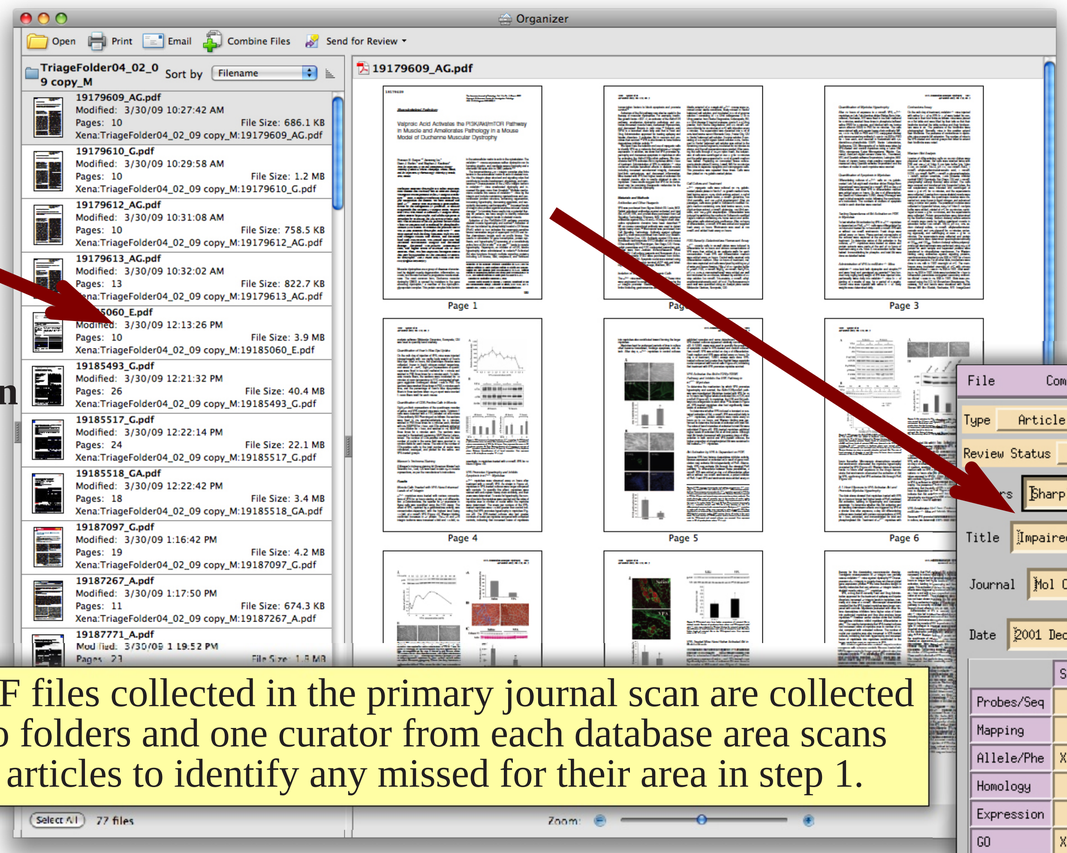


Relevant articles identified and tagged in QUOSA for various database areas

~10,000 articles added to MGI per year

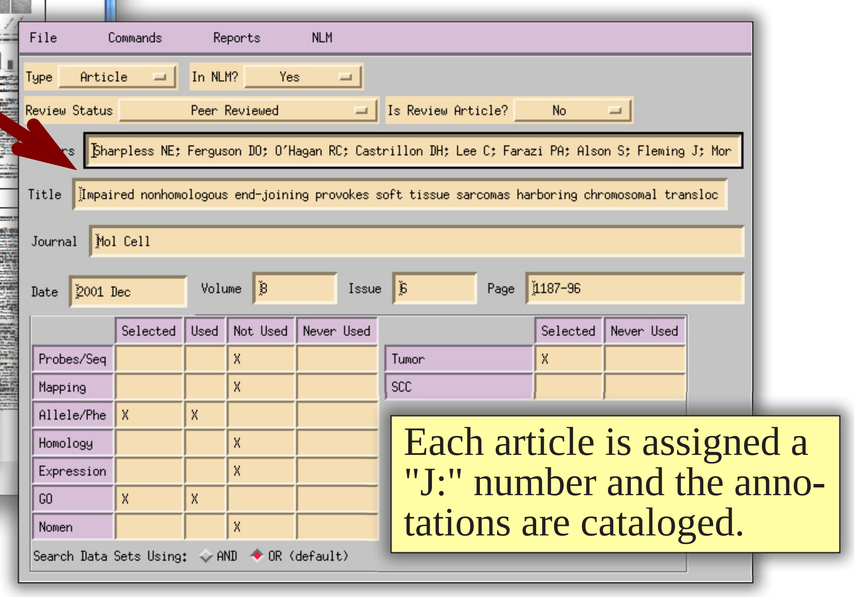
Database Areas:

Alleles & Phenotype
Gene Expression
Gene Ontology (GO)
Mapping
Nomenclature
Tumor



PDF files collected in the primary journal scan are collected into folders and one curator from each database area scans the articles to identify any missed for their area in step 1.

Acrobat used to further identify database areas



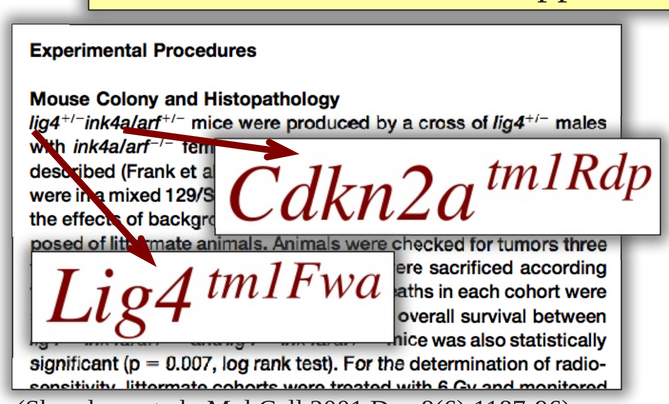
Each article is assigned a "J:" number and the annotations are cataloged.

Master Bibliography catalogs all articles and their chosen area associations

Gene & Allele Indexing

(Abe Y, et al., FEBS J 2008 Jan;275(2):318-31)

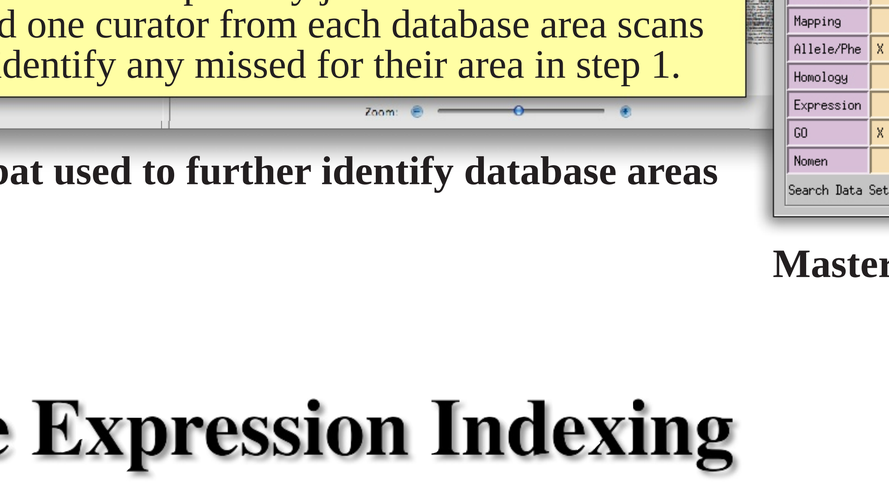
Each relevant reference is indexed for the genes and alleles used and standard nomenclature is applied.



(Sharpless, et al., Mol Cell 2001 Dec;8(6):1187-96)

Gene Expression Indexing

Each relevant reference is indexed for gene expression data, cataloged by gene, age and assay used.



Clicking on the number of results for a gene brings back a tabular summary of those results.

Tumor Indexing

Each relevant reference is indexed for the organs studied for tumorigenesis and is also prioritized for data entry.