Mouse Genome Informatics (MGI) Workshop

Mouse Genome Informatics (MGI) provides free, public access to integrated data on the genetics, genomics and biology of the laboratory mouse. In this self-guided tutorial you will use MGI to:

- learn how to navigate the MGI website efficiently
- locate mouse genotypes associated with specific phenotypes
- find mouse models of human diseases
- find candidate genes in a region of interest
- identify distributors of stocks carrying mutations of interest
- identify strain-specific SNPs and sequences
- retrieve upstream regulatory sequence of a gene
- find gene expression assays and images for specific anatomical structures
- perform batch queries

Phenotypes

Alleles of the human *APOE* gene are implicated in hypercholesterolemia, atherosclerosis and Alzheimer's Disease. How can you find more information on the mouse *Apoe* gene?

1. Go to the Mouse Genome Informatics (MGI) (**www.informatics.jax.org**) web site, enter apoe in the Quick Search field at the top of the web page and click the Quick Search button.
2. Click on the *Apoe* gene symbol to see its Gene Detail page, the page with links to all the information we have on this mouse gene, including genome coordinates and links to sequences, alleles, SNPs, Gene Ontology, expression data and references.

How can you find information on the different mouse *Apoe* alleles?

1. In the Mutations, Alleles and Phenotypes section of the *Apoe* Gene Detail page, click on the All Mutations and Alleles link. This will display a summary of the known alleles of the mouse gene and show official and unofficial (synonyms) nomenclature, and any high-level phenotype terms and human disease model information associated with each allele.

This allele summary provides the best overview of the alleles but what if you were more interested in seeing more specific phenotype data for the *Apoe* alleles?

1. Backup to the *Apoe* Gene Detail page and click on the first link for the Phenotype Summary in the Mutations, Alleles and Phenotypes section, the ### phenotypes from ## alleles in ## genetic backgrounds.

 This will show the lower level phenotype terms for each single mutant allele genotype. (For multiple-allele genotypes, use the second link for the Phenotype Summary.)

Jan Breslow's lab has knocked out the mouse gene. What percent of these mice on a 129P2/OlaHsd * C57BL/6 background develop develop cerebral xanthoma on a high cholesterol diet?

1. Scroll down and click on the *Apoe^{tm1Bres}* allele symbol.
2. In the Phenotypes section of this Allele Detail page, click on hm3 (hm for homozygous) for this strain background.

Genotype/Background:		Genotype	Genetic Background
	hm1	Apoe^{tm1Bres}/Apoe^{tm1Bres}	B6.129P2-Apoe^{tm1Bres}
	hm2	Apoe^{tm1Bres}/Apoe^{tm1Bres}	either: (involves: 129P2/OlaHsd * BALB/c) or (involves: 129P2/OlaHsd * C57BL/6J)
	hm3	Apoe^{tm1Bres}/Apoe^{tm1Bres}	involves: 129P2/OlaHsd * C57BL/6

(Note: *Apoe^{tm1Bres}* means this is the first targeted mutation (tm1) in *Apoe*, from the Jan Breslow lab (Bres).

How can you order these mice?

1. At the top of the page, in the Find Mice section, notice that though there are 10 mouse strains available carrying this mutation and there are 81 strains or lines available carrying any *Apoe* mutation. Click the 10 strains available link to search the IMSR.
2. This will search for stocks with this *Apoe* mutation from several providers. In your IMSR search results, the Strain Name links to the repository's site, the Repository links to an email form so you can ask about the stock and the Allele Symbol links to the MGI Allele Detail page.

Note: The Find Mice (IMSR) section also appears on Allele Detail pages. The IMSR (International Mouse Strain Resource) is a searchable online database of mouse strains and stocks available worldwide. Neither it, nor MGI, house any stocks. You can search the IMSR directly by clicking on the Find Mice (IMSR) link in the dark blue banner near the top of most

MGI web pages. If a desired allele is unavailable, it usually means the authors didn't contribute the stock to a public repository. Your best option then is to try contacting the authors of the original publication, noted at the bottom of the allele page.

How you can you find recombinase (cre) alleles detected in muscle?

1. Return to an MGI web page, then select Recombinase (cre) from the Search Menu or click on the blue Recombinases tab on most MGI web pages.
2. In the Recombinase activity assayed in: field, type: *skel mus* and select skeletal muscle from the list. Note: items are grayed out when we have no cre activity data for that tissue.
3. Click the Search button.
4. Scroll down to the Tg(Camk2a-cre/ERT2)1Aibs allele, where *Camk2a* is the driver (promoter) and click on the Matrix View icon in the second column. This allows you to compare endogenous wild-type *Camk2a* gene expression with reported recombinase activity data for this allele (highlighted) and other *Camk2a* driven alleles.
5. Return to the cre search results page and click on the allele symbol, Tg(Camk2a-cre/ERT2)1Aibs, to access the Transgene Details.
6. On the Transgene Detail page, in the Recombinase Activity section, click the blue triangle after muscle to expand the section.
7. Click on the check mark in the Images column for muscle.
8. The pattern of expression is described and the images can be resized and moved.
9. Note that in the top section of the page, 1 strain is available carrying this transgene.

The Gene Ontology (GO) project uses defined vocabularies to describe the biological processes, molecular functions and cellular components associated with a gene's products. How can you find the GO terms associated with the *Lrp4* gene?

1. Go to the MGI *Lrp4* Gene Detail page and scroll down to the Gene Ontology (GO) Classifications section.
2. Click on the All GO Annotations link.
3. Jump down the page by clicking on the Tabular View link.
4. Click on any Classification Term to view its place in the vocabulary hierarchy, see its definition, and to find a link to all the genes annotated to the term.

You are interested in a QTL defined by MIT markers, *D1Mit11* and *D1Mit445*. How can you see all the genes mapped between these markers?

1. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select Genes —> Genes & Markers Query.
2. In the Feature type section, check protein coding gene. (Click <u>Show</u> to see all the types and the counts for each.)
3. In the Genome location section of the page, enter d1mit11 and d1mit445 in the Marker range fields.

    Marker range: *use current symbols*

    between `d1mit11` and `d1mit445`

    e.g., between "D19Mit32" and "Tbx10"
4. Click the Search button

Now that you have all the candidate genes in this region, you can search for stronger candidates by adding phenotype or GO terms to your search parameters. Use the <u>Click to modify search</u> button above the results. Also notice that you have four Export options.

You are interested in regions of two human chromosomes. How can you see phenotypes and diseases associated with genes in these regions?

1. Click on the Human Disease tab at the top of an MGI web page.
2. Toggle "Please select a field" to "Genome Location."
3. Select the species and build: Human (GRCh38).
4. Enter the region: 19:43341000-44910000
    (This format is also supported: Chr19:43341000-44910000)
5. Click the Add button.
6. Toggle the second "Please select a field" to "Genome Location."
7. Select the species and build: Human (GRCh38).
8. Enter the region: 21:25001000-26971000
9. Toggle the default Boolean from AND to OR.
10. Click the Search button.

The search finds human genes mapped to those regions and returns them, their mouse homologs, and phenotypes and diseases associated with all the returned genes.

11. Scroll out to the right to see diseases associated with these genes. Blue squares indicate mouse data and tan is for human data.
12. Click on the blue square at the intersection of *PLAUR* and immune system to see the mouse *Plaur* genotypes with immune system abnormalities.

Because free text searches can produce spurious results, MGI creates and annotates to defined vocabularies such as the Mammalian Phenotype (MP) Ontology. The Ontology can be searched

or browsed. For example, some *Lrp4* alleles display abnormal motor neuron morphology. How can you find other alleles that show this phenotype?

1. You can search for this phrase in MGI's "Genes and Markers," "Phenotypes and Alleles," and "Mammalian Phenotype Browser" forms but today, just browse the Phenotype Browser. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select Phenotypes —> Mammalian Phenotype (MP) Browser. (Note: you can always return to the MGI Home Page by clicking on the MGI logo.)

2. You can search or browse the ontology. Today, browse by clicking on <u>nervous system phenotype</u>.

3. Click on <u>abnormal nervous system morphology</u> (a black triangle preceding a term means the term has subterms).

4. Click on <u>abnormal neuron morphology</u> and then <u>abnormal motor neuron morphology</u>.

5. Click on the number of genotypes.
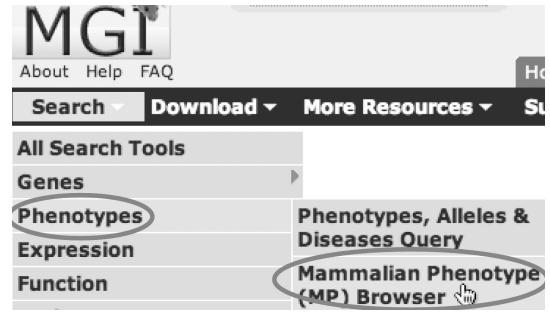
The MGI GO Browser works in a similar manner.
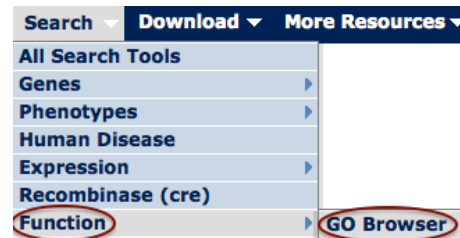
MGI also provides a database (MTB) of mouse tumor biology information. To access MTB, click on the light blue <u>Tumors</u> tab below the Quick Search button on the upper right hand side of most MGI web pages.
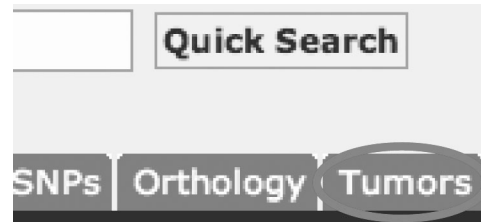
Or, on the MGI Home Page, click on <u>Mouse Models of Human Cancer</u>.

For now, just take a peek at the <u>Tumor Frequency Grid</u> under Additional Resources in the left hand column of the web page.

1. The grid displays published tumor incidence for inbred strains. Click the blue triangles to view substrain and sub-tissue information.

Patient Derived Xenograft (PDX) models for cancer research are created by the implantation of human cells and tumor tissue into immune compromised NOD-SCID-Gamma2 mouse hosts. How can you find SNPs/Variants for a given tumor?

1. In the Additional Resources section on the left hand side of the page, click on the new PDX Finder.

You can think of the PDX Finder as an IMSR for PDX Models, and model providers worldwide contribute to the web resource.

2. In the Search field, start typing breast and select Breast Cancer from the list that appears.
3. Currently more data are available for JAX mice, so in the left hand side of the page, for PDX Model, click on DATASOURCE and check JAX.
4. In your results, click on the Model: TM00091.
5. Though some Genomic Data are available through this site, more information is available at MTB, so click on the button to VIEW DATA AT JAX.
6. Click to expand the Variant Summary section.
7. You can also expand the Gene Expression section or see Copy Number Variants by expanding the Gene CNV section. Dosing studies are available at the bottom of the page.

Polymorphisms

MGI currently gets its SNP data from NCBI's dbSNP. How can you find SNPs for a region of mouse Chr 2?

1. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select Strains / SNPs —> SNP Query.
2. Click the Search by Region tab.
3. In the Genome Region section, select Chromosome 2 and enter these coordinates: 3102200-3292000.
4. You can limit your search to particular strains but today we will search all strains.
5. We will set a Reference Strain for Comparison. In that section, select C57BL/6J. This will limit your search to SNPs where there are data for B6.
6. Click the Search button.

The reference strain is listed first in your results, though you can drag strain columns to reposition them. Click on an MGI SNP Detail link in the first column to see SNP flanking sequence.

Sequences

How can you get a representative mouse transcript sequence for *Lrp4*?

1. Go to the MGI *Lrp4* Gene Detail page, scroll down to the Sequences & Gene Models section of the page and select the transcript box.
2. Click the Go button to download the FASTA file.

How can you get the *Lrp4* genomic sequence plus 5 Kb of flanking sequence?

1. Use the same section of the Gene Detail page that you used to get the transcript sequence and add 5 to the Flank box to the right of the genomic sequence.

How can you find exon information for the *Lrp4* gene?

1. Return to the *Lrp4* Gene Detail page, scroll down and in the Sequences & Gene Models section of the page, click on the ID for the Ensembl Gene Model.
2. Then click on <u>Sequence</u> in the left hand column.
3. Click on Configure this page in the left hand column to add line numbering.

MGI has begun curating gene interaction data. The first data set describes predicted and verified gene targets of known mouse microRNAs. How can you find microRNAs that interact with or are predicted to interact with *App*, the amyloid beta (A4) precursor protein?

1. Go to the MGI *App* Gene Detail page.
2. Scroll down to the Interactions section and click on the View All button.
3. By default, the table lists the validated interactions first, though you can change the sort by clicking on column heading arrowheads.
4. You can filter out the predicted interactions by clicking on the Validation filter button and checking validated.

    Filter interactions by: Interaction ⏷ Validation ⏷ Data Source ⏷ Score ⏷

5. You can use the Batch Query and MouseMine buttons to get additional data for these microRNAs.

Additional types of interactions and enhancements to the graph are planned for future database updates.

Alliance of Genome Resources

The Alliance of Genome Resources was created to provide a consistent user interface with an integrated view of data from 6 Model Organism Databases and the Gene Ontology (GO) Consortium. The founding database members of the Alliance are Flybase, Mouse Genome Database (MGI), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), Wormbase, and the Zebrafish Information Network (ZFIN). In their 1.0 release, AGR allowed users to search for, and view, gene, functional (GO) data, and disease associations from the above seven sources. A recent release now includes alleles. Orthology data for human and Alliance organisms are built on the DRSC Integrative Ortholog Prediction Tool (DIOPT), which integrates several prediction approaches.

How can you find functional information for orthologs of a binding protein whose mouse gene symbol is *Fkbp5*?

1. Go to The Alliance home page at **www.alliancegenome.org**, or click on their logo at the top right of an MGI web page.
2. Enter Fkbp5 in the search field and toggle All to Gene.

Matching results are highlighted in yellow. For example, the FKBP59 *Drosophila* gene is matched through orthology to Fkbp5 genes.

3. Click on the fkbp5 gene symbol for *Danio rerio*.
4. Scroll down to the Function – GO Annotations section and click on the light green cell for neurological.
5. Notice that this annotation was made in *Drosophila*. This site combines the GO annotations for the 6 species.

The Alliance does not yet have phenotype information. For that, go to the mouse gene:

6. Back up to your search results and click on the mouse *Fkbp5* gene symbol to access the mouse gene page.
7. A phenotype summary is provided by MGI in the Description, though more detail can be found by following the Source link (MGI:104670) to MGI.

Expression Data

Mutations in the human fibulin 5 gene (FBLN5) cause the disease, Cutis Laxa, which is characterized by pulmonary emphysema. Mice with *Fbln5* mutations also exhibit pulmonary emphysema. *Fbln5* has been assigned the Gene Ontology (GO) Biological Process term, cell adhesion. How can you find cell adhesion genes that are expressed in the lung?

1. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select Expression —> Gene Expression Data Query.
2. In the Genes section of the page, use the right hand field and type: *cell adhesion* and select the term from the list. Note that the number of genes annotated to the term is shown.
3. In the Anatomical Structure or Stage section, select "detected in."
4. In the Anatomical Structures field, type: lung and select the term from the list.
5. To reduce the number of results for this example, we are going to limit the search to Assay type: RNA in situ. An easy way to do this is to click the check mark for "Find expression data in any assay type" to uncheck all assay types and then select RNA in situ.
6. Click the Search button.
7. Click the Assays tab and then click on the data link for the *Adamts18* Assay, MGI: 4822986.
8. Click on the Figure link for the Structure, TS23: left lung.
9. Beneath the image, in the Other Database Links section, click on the Eurexpress link. At the Eurexpress site you can zoom in on the image or view a different serial section.

Batch Queries

How can you find information on several genes at once?

1. On an MGI web page, click on Download in the dark blue banner and select Batch Query.

2. In the text field, enter a column of IDs and/or gene symbols, for example:

    P08226
    Shank3
    App

3. In the Output section of the page, check Nomenclature, Genome Location and Gene Ontology (GO). (GO describes the molecular functions, biological processes and cellular locations of the gene product.)
4. You can export your results in spreadsheet or Tab-delimited formats or forward it to MouseMine (see the next example).

We have seen that MGI makes it easy to find genotypes associated with particular phenotypes but the web forms currently do not allow you to export genotype details in tab-delimited formats. If you used the Mammalian Phenotype (MP) Browser to find genotypes associated with emphysema (MP:0001958) and abnormal lung elastance (MP:0011043), how can you get a text file of those genotypes?
You can do this (and much more) with our MouseMine, **www.mousemine.org**.

1. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select MouseMine.
2. Click on the Templates tab near the top of the page.
3. Click on Mammalian phenotypes (MP terms) --> Mouse genes and models.
4. Enter MP:0001958, MP:0011043 in the LOOKUP field and click the Show Results button.
5. Toggle the 25 results so you can see all the results.

This returns a list of genes and transgenes and genotypes associated with these phenotypes terms and their subterms. Now we can modify the query:

6. Click on the Manage Columns button and then click on Add a Column.
7. Click on MGI Type the first time it is listed (under Subject).
8. Click Add 1 new column.
9. Click Apply Changes.
10. This adds the gene feature types to the table. Now you can remove columns you don't want to see by clicking the X in the column heading and you can filter by clicking the last icon in a column heading, the graph icon.
11. Click the Export button to select your formatting options.

References

An overlooked feature of MGI is the ability to search publications we have curated.
1. Go to Search beneath the MGI logo on the upper left hand side of an MGI web page and select References —> Reference Query.
2. Try an author and/or text search. (The Author and Journal fields feature autofills.)
3. Click on in the Title column to view an abstract.
4. The first column provides links to PubMed and Journals and the Curated Data column links to data curated from the reference and is also a way to find official nomenclature for genes and alleles researched in a publication.

Mouse Genome Informatics (MGI) has made electronic versions of several great out-of-print books available through our web site, including: *Biology of the Laboratory Mouse* by the Staff of The Jackson Laboratory, edited by Earl L. Green, *The Anatomy of the Laboratory Mouse* by Margaret J. Cook, *Mouse Genetics* by Lee Silver, *The Coat Colors of Mice* by Willys K. Silvers,

Color Atlas of Neoplastic and Non-neoplastic Lesions in Aging Mice by C. H. Frith and J. M. Ward and *Origins of Inbred Mice*, edited by Herbert C. Morse III.

1. Go to the MGI web site and click on <u>More Resources</u> in the dark blue banner and select <u>Online Books</u>.
2. Click on *Biology of the Laboratory Mouse*.
3. Click on the cover to open the book.
4. Scroll down to the bottom of the left hand frame of the page and click on <u>Image Index</u>.
5. Click on one of the Plates or other Figures.

If you have any questions on the Mouse Genome Informatics databases, contact **mgi-help@jax.org**, or click the <u>User Support</u> link at the very bottom of our web pages.