

The background of the entire page is a faded, orange-tinted aerial photograph of a city. A prominent feature is a large, ornate cathedral with a tall spire, situated on a riverbank. The river flows through the city, and other buildings and structures are visible in the distance. The overall color palette is warm, dominated by shades of orange and brown.

**Sunday November 6**

10.00am – 12.30pm

**Poster Session 1**

Bioinformatics Pages 86 - 94

Posters P-1 – P-14

## Bioinformatics

- P-1 EVIDENCE FOR NONRANDOM GENE ORDER IN MOUSE TESTIS SAGE TRANSCRIPTOMES**  
P Divina, R Storchova, P Strnad, C Vlcek, V Paces, J Forejt  
Institute of Molecular Genetics, AS CR, Prague, Czech Republic
- P-2 THE PHENOTYPE RESOURCE OF MGI: FACILITATING THE MOUSE AS A MODEL SYSTEM**  
M Tomczuk, C L Smith, H Dene, D L Burkart, L L Washburn, I Lu, SM Bello, A Anagnostopoulos, M Cassell, H Onda, B Richards-Smith, J T Eppig  
The Jackson Laboratory, Bar Harbor, ME, United States
- P-3 THE MOUSE TUMOR BIOLOGY DATABASE (MTB)**  
J T Eppig, D M Krupke, T C Allio, M J Vincent, J P Sundberg, C J Bult  
The Jackson Laboratory, Bar Harbor, ME, United States
- P-4 BASELINE PHENOTYPE DATA IN THE CONTEXT OF STANDARD PROTOCOLS IN THE EUROPHENOME DATABASE**  
A Shukla<sup>1</sup>, E Green<sup>2</sup>, H Lad<sup>3</sup>, J Weekes<sup>4</sup>, G Gkoutos<sup>5</sup>, S Greenway<sup>6</sup>  
<sup>1</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>2</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>3</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>4</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>5</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>6</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>7</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>8</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom
- P-5 FINDING A MOUSE: THE INTERNATIONAL MOUSE STRAIN RESOURCE (IMSR)**  
B A Richards-Smith, P J Frost, D O Walton, J T Eppig  
The Jackson Laboratory, Bar Harbor ME, United States
- P-6 A NEW METHOD FOR HIGH-THROUGHPUT IDENTIFICATION OF PHENOTYPE-RESPONSIBLE MUTATIONS BY UTILIZING DESIGNABLE DNA-CHIPS AND BIOINFORMATICS**  
T T Toyoda, Y H Hasegawa, Y M Mochizuki, K P Player, N H Heida, E K Kaminuma, N K Kobayashi, Y S Sakuraba, Y G Gondo, J K Kawai, Y H Hayashizaki, K I Ikeda, H M Masuya, S W Wakana, T S Shiroishi  
Genomic Sciences Center, RIKEN, Yokohama, Kanagawa, Japan
- P-7 ARTADE: A NEW BIOINFORMATICS TOOL FOR TILING-ARRAY-BASED ELUCIDATION OF TRANSCRIPTIONAL STRUCTURES AND ALTERNATIVELY SPLICED VARIANTS FOR CODING AND NON-CODING GENES**  
T Toyoda, Y Mochizuki, N Heida, K Ikeda, H Masuya, S Wakana, T Shiroishi, J Kawai, Y Hayashizaki, Y Sakuraba, Y Gondo  
Genomic Sciences Center, RIKEN, Yokohama, Kanagawa, Japan
- P-8 LONGSAGE ANALYSIS SIGNIFICANTLY IMPROVES GENOME ANNOTATION: IDENTIFICATIONS OF NOVEL GENES, ALTERNATIVE TRANSCRIPTS AND ANTISENSE TRANSCRIPTS IN THE MOUSE**  
MB Wahl, U Heinzmann, K Imai  
GSF-National Research Center for Environment and Health, Neuherberg, Germany
- P-9 RECENT ADVANCES IN MOUSE GENOME INFORMATICS**  
S E McClatchy, J A Blake, C J Bult, J A Kadin, J E Richardson, M Ringwald, J T Eppig  
The Jackson Laboratory, Bar Harbor, Maine, United States

- P-10 MOUSE GBROWSE AT MGI: AN INTERACTIVE GENOME SEQUENCE FEATURE MAP FOR THE LABORATORY MOUSE**  
CJ Bult, R Sinclair, D Qi, S Zhu, D Reed, D Bradt, K Frazer, I Witham, M McCrossin  
The Jackson Laboratory, Bar Harbor, Maine, United States
- P-11 PRIORITIZATION OF CANDIDATE GENES FOR EMBRYO-LETHAL MUTATIONS**  
V P Mancuso<sup>1</sup>, O Valladares<sup>2</sup>, H Shao<sup>2</sup>, C Stoeckert<sup>3</sup>, M Bucan<sup>2</sup>  
<sup>1</sup> CAMB Graduate Group, Philadelphia, PA, United States, <sup>2</sup> Department of Genetics, Philadelphia, PA, United States, <sup>3</sup> Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA, United States
- P-12 THE GENE EXPRESSION DATABASE (GXD): INTEGRATED ACCESS TO EXPRESSION INFORMATION FROM THE LABORATORY MOUSE**  
M Ringwald, D A Begley, I J McCright, J H Finger, T F Hayamizu, D P Hill, C M Smith, R M Baldarelli, L E Corbani, J Campbell, S E McClatchy, J A Blake, C J Bult, J T Eppig, J A Kadin, J E Richardson  
The Jackson Laboratory, Bar Harbor, Maine, United States
- P-13 MOUSE NOMENCLATURE: A YEAR IN REVIEW**  
L J Maltais, J A Blake, M McAndrews-Hill, L McLellan  
The Jackson Laboratory, Bar Harbor, ME, United States
- P-14 MOUSE PHENOME PROJECT**  
MA Bogue, S C Grubb  
The Jackson Laboratory, Bar Harbor, ME, United States

**P-1**

**EVIDENCE FOR NONRANDOM GENE ORDER IN MOUSE TESTIS SAGE TRANSCRIPTOMES**

P Divina, R Storchova, P Strnad, C Vlcek, V Paces, J Forejt  
Institute of Molecular Genetics, AS CR, Prague, Czech Republic

Increasing body of evidence suggests that some groups of genes are not distributed randomly along the chromosomes, but tend to form clusters and/or accumulate on sex chromosomes. Here, we present an analysis of gene order in the transcriptomes of cells from total mouse testis and testicular somatic cells created by serial analysis of gene expression (SAGE). An analysis of the testis SAGE transcriptomes in conjunction with the available SAGE transcriptomes of seven non-testis tissues provided the following evidence for non-random gene distribution. (1) Considering all genes expressed in the tissues, we observed an under-representation of X-linked genes in the total testis but not in the testicular somatic cells and other tissues. (2) When tissue-specific genes were examined, a significant 3.2-fold enrichment of the proportion of X-linked genes specific for testicular somatic cells was detected, while the proportions of X-linked genes specific for total testis and for other tissues were comparable. (3) The genes with preferential expression in testis tissues exhibit a certain tendency to form clusters encompassing a significantly higher number of genes than expected by chance. Our results provide new evidence in favor of the theory of male-biased genes accumulation on the X chromosome in testicular somatic cells and indicate the opposite action of the meiotic X-inactivation in testicular germ cells. This corresponds well to the data obtained in mice and humans, but conflicts with the data obtained in *Drosophila*. To elucidate the reason for these discrepancies, more data on other organisms are needed.

**P-2**

**THE PHENOTYPE RESOURCE OF MGI: FACILITATING THE MOUSE AS A MODEL SYSTEM**

M Tomczuk, CL Smith, H Dene, DL Burkart, LL Washburn, I Lu, SM Bello, A Anagnostopoulos, M Cassell, H Onda, B Richards-Smith, JT Eppig  
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Genome Informatics Database (MGI, <http://www.informatics.jax.org>) integrates mouse genetic, genomic, and biological data to facilitate the use of mouse as a model system for understanding human biology and disease. Thus the cataloging of mutant alleles, the descriptive annotations of phenotypes and the associations of mouse mutants to human disease is key to MGI's goals.

Phenotypic data are curated using the Mammalian Phenotype (MP) Ontology, a vocabulary developed to provide standardized terms for describing phenotypic characteristics. Additionally, disease terms from Online Mendelian Inheritance in Man (OMIM) are used to associate mouse models and human diseases. These controlled vocabularies assist in searching, grouping, comparing and analyzing data that are ineffectively retrieved through text searches. Users can query the database by phenotypic terms, gene or allele names, type of mutant allele (spontaneous, chemically induced, targeted, etc.) and disease terms. Each allele or QTL annotation contains a molecular description (if known), is described relative to genetic background strain(s), is associated with MP Ontology terms containing specific details, and may be correlated with human disease. Users may also browse the phenotype and human disease vocabularies and retrieve data associated with particular terms. The information in MGI aids in model building and in addressing complex questions by allowing users to carry out sophisticated queries such as those simultaneously addressing phenotypic, biochemical function and process, expression, sequence and mapping data.

Additional MGI posters: Blake (Gene Ontologies), Bult (GBrowse, SNPs), Eppig (Disease models/OMIM, MTB), Maltais, (Nomenclature), McClatchey (MGI), Richards-Smith (IMSR), Ringwald (GXD).

Supported by NIH grant HG00330.

**P-3****THE MOUSE TUMOR BIOLOGY DATABASE (MTB)**

J T Eppig, D M Krupke, T C Allio, M J Vincent, J P Sundberg, C J Bult  
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Tumor Biology Database (MTB, <http://tumor.informatics.jax.org>) supports the use of mouse as a model system of hereditary and induced cancers. The database is designed around the principle that genetic background is a key factor influencing the kinds and onset of cancers observed in different strains of genetically defined mice. MTB provides researchers with access to data on mouse cancer models including tumor names and classifications, tumor incidence and latency data in different strains, tumor pathology reports and images, information on genetic factors associated with tumors and tumor development, and literature references.

The MTB web interface supports complex ad hoc queries – for example, “Which transgenic strains of mice have a high incidence of lung adenocarcinomas?” or “Show me tumor records where a point mutation was detected in *Kras2*”. Such queries are possible in MTB by integrating mouse tumor data from many sources, primarily the published literature, and using standardized nomenclature and controlled vocabularies. Terms for classifying tumors, organ and tissue names, strain type, reproductive status, and mutation type are all standardized. This feature allows consistent searching and enables links from MTB records to other resources such as the Mouse Models of Human Cancer Consortium (<http://emice.nci.nih.gov>). Nomenclature for genes, alleles, and strains is integrated with the Mouse Genome Informatics resources (<http://www.informatics.jax.org>), so users can access a wide range of additional related genetic and phenotypic information. MTB provides the scientific community with a central resource for rapidly finding and evaluating the ever-expanding volume of mouse tumor data.

Supported by NIH grant CA89713

**P-4****BASELINE PHENOTYPE DATA IN THE CONTEXT OF STANDARD PROTOCOLS IN THE EUROPHENOME DATABASE**

A Shukla<sup>1</sup>, E Green<sup>2</sup>, H Lad<sup>3</sup>, J Weekes<sup>4</sup>, G Gkoutos<sup>5</sup>, S Greenway<sup>6</sup>

<sup>1</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>2</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>3</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>4</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>5</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>6</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>7</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom, <sup>8</sup> Medical Research Council, Harwell, Didcot, Oxford, United Kingdom

ENU mutagenesis is widely used to generate mutant mice and identify models of human genetic disease. Characterisation of these mutants requires standard ways of determining phenotypic information. An important aspect of this is describing the phenotypes of inbred mouse strains to allow comparison with mutant lines. Eumorphia (<http://www.eumorphia.org>) is an integrated program for the development of standard methods for analysing mouse phenotypes. Application of these methods on normal mice generates baseline data to which mutant mice can be compared.

Incoming baseline data relates to multiple categories and sources, resulting in various data formats, ambiguity in data, lack of common vocabulary to describe and compare the data, and absence of important metadata concerning experimental conditions. We are developing a resource called EuroPhenome for the efficient acquisition, storage and query of this baseline data by solving the above data issues. Access to EuroPhenome is web-based using the client-server mechanism with MySQL as a back-end. The EuroPhenome search module allows comparative search and querying of the data by combining search categories, e.g. assay, lab, mouse-age, gender, genotype and strain. Alternatively, the user can browse and download the data for a specific category for further analysis. Importantly in view of recent developments in phenotype ontologies, EuroPhenome relates phenotype data directly to standard SOPs, accessible through the EMPReSS resource (<http://empress.har.mrc.ac.uk>). By referring to the experimental metadata it may be possible to identify sources of variation in the data.

EuroPhenome will act as a phenotype data source and tool-box for the European mouse community.

**P-5**

**FINDING A MOUSE: THE INTERNATIONAL MOUSE STRAIN RESOURCE (IMSR)**

B A Richards-Smith, P J Frost, D O Walton, J T Eppig  
The Jackson Laboratory, Bar Harbor ME, United States

The International Mouse Strain Resource (IMSR, <http://www.imsr.org>) is a searchable online database of mouse strains and stocks available from facilities around the world. The IMSR currently tracks over 5,800 strains and stocks from 19 repositories in the U.S., Canada, Europe, Japan, and Australia. In addition, over 10,300 ES cell lines from BayGenomics are represented. Additional mouse repositories' holdings and the International Gene Trap Consortium (<http://www.igt.org.uk>) lines will soon be included in the IMSR database.

Users can search the IMSR using one or more of the following parameters; strain or stock name or accession ID, strain state (e.g. live, cryopreserved), strain type (e.g. inbred, congenic), gene or allele symbol/name or MGI ID, mutation type (e.g. spontaneous, targeted), chromosome, and geographic location or specific repository.

A search of the IMSR database returns those strains satisfying the user's query. Information provided includes strain/stock designation, holder site, state, strain type, synonyms, chromosome, allele symbol, allele name, gene name, and mutation type. From this table users can link directly to: 1) additional details about the strain as provided on the holder's website; 2) an e-mail form for contacting the holder with questions or to order the mouse; and 3) the Mouse Genome Informatics Database's (MGI's) phenotype descriptions associated with the strain's mutant alleles. Finally, users of MGI (<http://www.informatics.jax.org>) will find links directly to IMSR from Phenotypic Allele Detail pages for mutant alleles, further facilitating "finding my mouse".

Also see MGI posters by: Blake, Bult, Eppig, Maltais, McClatchey, Ringwald, Tomczuk  
Supported by NIH grants MH061915, HG00330

**P-6**

**A NEW METHOD FOR HIGH-THROUGHPUT IDENTIFICATION OF PHENOTYPE-RESPONSIBLE MUTATIONS BY UTILIZING DESIGNABLE DNA-CHIPS AND BIOINFORMATICS**

T T Toyoda, Y H Hasegawa, Y M Mochizuki, K P Player, N H Heida, E K Kaminuma, N K Kobayashi, Y S Sakuraba, Y G Gondo, J K Kawai, Y H Hayashizaki, K I Ikeda, H M Masuya, S W Wakana, T S Shiroishi  
Genomic Sciences Center, Riken, Yokohama, Kanagawa, Japan

Large-scale production of ENU-induced mutant mice has been underway in RIKEN GSC, and hundreds of interesting phenotypes have been obtained so far. However, identification of the mutated genes responsible for the phenotypes requires the conventional forward-genetics approach, which involves a lot of time and effort. Here we introduce a bioinformatics-assisted method named "*in silico* positional cloning" designed to contribute to high-throughput identification of the responsible genes. Our method consists of four steps as follows. (1) Marker-based genetic rough mapping, by which the approximate location of the responsible mutation point is narrowed down to a short interval (5 to 15 Mbp) on a chromosome. (2) Bioinformatics-based ranking of the candidate genes in the interval by using integrated databases ranging from genome to phenome (Omic Space). (3) Design of DNA-chips with probes densely covering the sequences of the candidate genes, followed by identification of candidate SNPs by using the DNA chip. As an initial screening of candidate SNP points, each mRNA extract from the wild-type and the mutant is hybridized with the oligomer probes of the DNA-chip. Then, computationally we try to find those probes showing both high intensities in the wild-type sample and diminished ones in the mutant sample, due to the affinity loss caused by a point mutation. (4) Verification of the SNPs by direct sequencing, followed by functional analyses. We have started applying the above-mentioned method to several ENU-mutant mouse cases, and currently, we have one successful result.

**P-7****ARTADE: A NEW BIOINFORMATICS TOOL FOR TILING-ARRAY-BASED ELUCIDATION OF TRANSCRIPTIONAL STRUCTURES AND ALTERNATIVELY SPLICED VARIANTS FOR CODING AND NON-CODING GENES**

T Toyoda, Y Mochizuki, N Heida, K Ikeda, H Masuya, S Wakana, T Shiroishi, J Kawai, Y Hayashizaki, Y Sakuraba, Y Gondo

Genomic Sciences Center, RIKEN, Yokohama, Kanagawa, Japan

Tiling arrays of high-density oligonucleotide probes spanning the entire genome are powerful tools for the discovery of new genes and are expected to elucidate new forms of alternatively spliced variants. However, it is difficult to determine the structure of the spliced product of a structurally unknown or ambiguous gene from noisy array signals only. Here we introduce a statistical method that estimates the precise splicing points and the exon/intron structure of a structurally unknown gene by maximizing the odds or the ratio of the likelihood of observed array-signal intensities, nucleic-acid sequences based on a bi-directional Markov model and length likelihood of exons and introns. Our method predicted more accurately the gene structures than the simple threshold-based method, and more correctly estimated the expression values of structurally unknown genes than the window-based method. It was observed that the Markov model contributed to the precision of splice-points, and that the statistical significance of the expression (P value) represented the reliability of the estimated gene structure and expression value well. We have implemented the method as a program ARTADE and have developed some useful tools for the analysis of alternatively spliced variants and novel gene structures based on tiling array intensities. The ARTADE program suite is freely available for academic and non-for-profit users at <http://omicspace.riken.jp/ARTADE/>.

**P-8****LONGSAGE ANALYSIS SIGNIFICANTLY IMPROVES GENOME ANNOTATION: IDENTIFICATIONS OF NOVEL GENES, ALTERNATIVE TRANSCRIPTS AND ANTISENSE TRANSCRIPTS IN THE MOUSE**

MB Wahl, U Heinzmann, K Imai

GSF-National Research Center for Environment and Health, Neuherberg, Germany

We evaluated our LongSAGE dataset of 202,015 tags (consisting of 41,718 kinds of tags), generated from mouse embryonic tails, and systematically analyzed it against transcript and genome sequence databases. Our results from the experimentally obtained dataset suggested that with the use of a single anchoring enzyme a fraction of LongSAGE tags could not be unambiguously assigned to its gene, due to the presence of widely conserved sequences downstream of particular CATG anchor sites. We observed that 45% of all detected genes with corresponding entries in the Mouse Genome Database and 27% of the detected Ensembl genes gave rise to alternative transcripts, resulting in multiple LongSAGE tags for single genes. Surprisingly, a large fraction of LongSAGE tags with a hit to the genome (66%) could not be assigned to any gene annotated in Ensembl. Thus we tested whether these LongSAGE tags could be assigned directly to a gene using additional evidence. In 2,098 cases, a LongSAGE tag fell into a region containing a putative gene predicted by GenScan, suggesting that such a predicted gene was really transcribed. Furthermore, we identified 9,112 genes that were left out or were wrongly annotated by the Ensembl pipeline. Finally, we found evidence for the presence of 1,260 potential antisense genes, of which 1,001 are not annotated in Ensembl, thereby being regarded as novel. Interestingly their sense counterpart was co-expressed in the majority of the cases.

**P-9**

**RECENT ADVANCES IN MOUSE GENOME INFORMATICS**

S E McClatchy, J A Blake, C J Bult, J A Kadin, J E Richardson, M Ringwald, J T Eppig  
The Jackson Laboratory, Bar Harbor, Maine, United States

The Mouse Genome Informatics (MGI) database, a free online resource, supplies curated genomic, genetic, and biological data about the laboratory mouse. Recent developments additions include:

Sequence assembly coordinates for genes

Mouse GBrowse, a genome browser

An electronic version of The Anatomy of the Laboratory Mouse by Margaret J. Cook

Mammalian phenotype annotation of genotypes and strains and images of mutant phenotypes

Online Mendelian Inheritance in Man (OMIM) disease terms annotated to mouse genotypes

Addition of mouse SNP and other polymorphism data

MGI integrates data gleaned from manual curation of primary literature, submissions of data from investigator laboratories and centers, with and that supplied from databases other database resources such as GenBank, Swiss-Prot and Ensembl. Further insight into mouse gene expression, function and cancer biology is supplied provided by three projects: the Gene Expression Database (GXD), Gene Ontology (GO), and the Mouse Tumor Biology (MTB) database. Integration of rich data from many sources makes MGI a powerful tool for research.

Please visit the MGI website at [www.informatics.jax.org](http://www.informatics.jax.org).

Also see these MGI posters: Bult (GBrowse, SNPs), Eppig (Disease annotations/OMIM, MTB), Maltais (Nomenclature), Ringwald (GXD), Tomczuk (Phenotypes), Richards-Smith (IMSR)

Supported by NIH grants HG00330, HG02273, CA89713, HD37745

**P-10**

**MOUSE GBROWSE AT MGI: AN INTERACTIVE GENOME SEQUENCE FEATURE MAP FOR THE LABORATORY MOUSE**

C J Bult, R Sinclair, D Qi, S Zhu, D Reed, D Bradt, K Frazer, I Witham, M McCrossin  
The Jackson Laboratory, Bar Harbor, Maine, United States

Mouse GBrowse is a web-based interactive map for the graphical display of mouse genome annotations. Like other popular genome browsers, Mouse GBrowse displays annotations as “tracks” of data that can be turned on and off according to the interests of the viewer. The display supports several different levels of resolution as well as pan and zoom functionality. Standard data tracks currently available for Mouse GBrowse include gene predictions from NCBI and Ensembl, mouse SNPs, mouse QTL regions, and mouse genes that are associated with specific phenotypes. The features displayed in Mouse GBrowse are hypertext linked to several informatics resources including dbSNP, Ensembl, EntrezGene, and MGI. Mouse GBrowse supports simple keyword and genome coordinate based searches. In addition to the standard data tracks available on Mouse GBrowse, researchers can upload their own private annotations for display in the context of other genome features.

Mouse GBrowse is accessible from <http://gbrowse.informatics.jax.org>; it was implemented using the Generic Genome Browser software development toolkit that is available from the Generic Model Organism Database (GMOD) web site (<http://www.gmod.org>).

MGI is supported, in part, by NIH/NHGRI HG00330.

**P-11****PRIORITIZATION OF CANDIDATE GENES FOR EMBRYO-LETHAL MUTATIONS**

V P Mancuso<sup>1</sup>, O Valladares<sup>2</sup>, H Shao<sup>2</sup>, C Stoeckert<sup>3</sup>, M Bucan<sup>2</sup>

<sup>1</sup> CAMB Graduate Group, Philadelphia, PA, United States, <sup>2</sup> Department of Genetics, Philadelphia, PA, United States,

<sup>3</sup> Penn Center for Bioinformatics, University of Pennsylvania, Philadelphia, PA, United States

To prioritize candidates for embryo-lethal mutations in the mouse, we performed bioinformatics analysis of 615 genes with embryo-lethal phenotypes in the mouse and zebrafish. A set of 300 mouse genes, shown to cause pre- and perinatal lethality in the mouse, were identified by searching the Mouse Genome Database (GO term = "survival: embryonic lethality"), and 315 genes essential for embryonic and early larval development in zebrafish were identified in an insertional mutagenesis screen (Amsterdam, et al., 2004). We asked if essential genes share a common expression pattern and if these genes have a higher rate of conservation than other genes.

We demonstrate that essential genes are highly expressed in the embryo. To exploit this trend, a ratio (mean expression in adult tissues/mean expression in embryo) has been generated to describe the level of embryonic expression of each gene. We were able to obtain mouse expression data for 241 essential mouse genes and 307 orthologs of essential zebrafish genes, and found that 56% of genes known to be essential in the zebrafish and mouse have a higher level of relative expression in the embryo.

Previous studies have also found that essential genes, and genes producing phenotypes when mutated are conserved. (Amsterdam, et al., 2004, Ashburner, et al., 1999) We identified 1:1 orthologs of mouse genes in human, *D. melanogaster*, *C. elegans*, and *S. cerevisiae*, and have asked if orthologs conserved throughout evolution are more common in essential genes. 165/294 zebrafish lethal genes and 24/183 mouse lethal genes for which conservation data were available were conserved through yeast.

We also examined the rate of protein evolution by measuring the KA/KS ratio for mouse and human orthologs, and compared essential genes to a set of 353 genes on mouse Chr5. The mean KA/KS is 0.13 in the Chr5 genes, whereas both the essential genes in the mouse and the orthologs to zebrafish essential genes have a lower rate of protein evolution. (0.084; p-value < 0.0001, and 0.071; p-value < 0.0001). These data indicate that expression and level of conservation may be traits common to essential genes. Using this knowledge we have prioritized candidate genes on mouse chromosome 5 to support positional cloning efforts.

**P-12****THE GENE EXPRESSION DATABASE (GXD): INTEGRATED ACCESS TO EXPRESSION INFORMATION FROM THE LABORATORY MOUSE**

M Ringwald, D A Begley, I J McCright, J H Finger, T F Hayamizu, D P Hill, C M Smith, R M Baldarelli, L E Corbani, J Campbell, S E McClatchy, J A Blake, C J Bult, J T Eppig, J A Kadin, J E Richardson  
The Jackson Laboratory, Bar Harbor, Maine, United States

The Gene Expression Database (GXD), one component of the Mouse Genome Informatics (MGI) system, collects and integrates different types of gene expression data from wild-type and mutant mice. Data are acquired from the literature and via electronic data submissions, brought into standardized formats and integrated with genotypic, phenotypic and other biological information about the laboratory mouse. This integration enables powerful database queries and fosters insights into the molecular networks underlying development and disease. The data content and querying capabilities, and thus the utility of GXD, have increased tremendously during recent years. Highlights of the last year include: the steady addition of expression data from the literature; the integration of large amounts of RNA in situ hybridization data, including image data, from large-scale screens; the incorporation of all publicly available mouse cDNA and EST data combined with our standardized annotation of pertinent cDNA source data (such as strain and tissue); the development of a new version of the Gene Expression Notebook, a tool to organize and store expression data in the laboratory and submit selected data to GXD; and the development of a prototype Web-based tool that will allow users to explore microarray data and to exploit its integration with MGI by creating, viewing, sorting, and correlating customized subsets of data through iterative query refinement.

GXD and pertinent tools are accessible through the Mouse Genome Informatics web site at <http://www.informatics.jax.org/>. GXD is supported by NIH grant HD33745.

**P-13**

**MOUSE NOMENCLATURE: A YEAR IN REVIEW**

L J Maltais, J A Blake, M McAndrews-Hill, L McLellan  
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Genome Database (MGD), under the auspices of the International Committee on Standardized Nomenclature for Mice, is the authoritative source for mouse genes, genetic markers, and strain nomenclature. Official nomenclature is employed throughout the Mouse Genome Informatics (MGI) resource ([www.informatics.jax.org](http://www.informatics.jax.org)) and disseminated to databases such as EntrezGene, UniProt, Ensembl, miRNA, and KinBase. The MGD Nomenclature group interacts on a daily basis with the HUGO Gene Nomenclature Committee (HGNC), the Rat Genome Database (RGD), and members of the scientific community for reserving and assigning official nomenclature. In addition, the MGD nomenclature group assists with the assignment and resolution of sequence-to-gene associations in collaboration with other MGI staff and the NCBI (EntrezGene and RefSeq). Journal editors, including those from Nature Genetics, Genomics, and Biology of Reproduction, consult the nomenclature group to ensure published articles will use official nomenclature. Key projects for this year included 1) work with gene family experts to revise the nomenclature for gene families, such as ACOT, ACSL, ARID, UGT, 2) updates to nomenclature for mouse/human orthologs involving RIKEN symbols as part of the Human and Mouse Orthology project, a collaboration with the HGNC, 3) review of outstanding nomenclature issues such as interim symbols, 4) sharing a booth on nomenclature with the HGNC and RGD at the Human Genome Meeting (HGM2005), and 5) hosting a workshop at The Jackson Laboratory to discuss nomenclature issues among the mouse, human, and rat nomenclature groups. Supported by NIH grant HG00330

**P-14**

**MOUSE PHENOME PROJECT**

M A Bogue, S C Grubb  
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Phenome Project is an international collaborative effort to promote the systematic characterization of a set of inbred strains and their derivatives, and to deposit the data in a central database. Community SNPs are collected also. The web interface for the Mouse Phenome Database (MPD) enables data retrieval/downloads, provides tools for online analysis and links to other databases. Selected MPD updates follow: 1) At the request of the community, password-protected web carrels are available for confidential work sessions with no requirement to publicly post private data. MPD carrels are designed for users to mine their own private data using MPD tools; moreover, it is possible to compare/correlate private data across the entire MPD. 2) A new feature has been added to merge multiple SNP records into a single row. This option collapses multiple SNP records sharing the same genomic location into a single row while resolving discrepancies and maintaining links to source data. 3) The identification of mouse models of complex human phenotypes is imperative for many research areas. Strain surveys and careful quantitation of parameters relevant to human health are useful in finding models of best approximation. The MPD helps identify models using a criteria fit tool based on user-selected MPD measurements and user-supplied criteria. 4) Software is under development for studying genotype:phenotype associations. The utility/potential of such a tool is ultimately dependent on the quantity and quality of phenotypic data. We encourage investigators and funding agencies to invest in efforts to quantify those parameters relevant to their disease areas. The prospective benefits—tapping the potential of inbred strains—promise to outweigh the costs.

MPD applications will be demonstrated.

[www.jax.org/phenome](http://www.jax.org/phenome)

The Mouse Phenome Project is funded in part by NHLBI, NHGRI, NIMH (USA).