

Monday November, 13
4.30pm – 6.30pm
Poster Session 2
Bioinformatics Pages
Posters P28 – P49

P28 ANNOTATION OF THE MOUSE GENOME: AUTOMATION MEETS MANUAL EXPERTISE

C Amid, JP Almeida, S Donaldson, A Frankish, RC Gibson, EA Hart, K Kivinen, GK Laird, JE Loveland, JM Mudge, AO Mujica, JA Rajan, HK Sehra, CA Steward, MM Suner, MD Thomas, LG Wilming, JL Harrow
Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

P29 MGI: ONE STOP “SHOPPING” FOR ORTHOLOGOUS INFORMATION GENE SETS IN MOUSE, HUMAN, AND RAT

LJ Maltais, JA Blake, MA Dolan, M McAndrews-Hill, TBK Reddy, CJ Bult, JT Eppig, MGI Staff
The Jackson Laboratory, Bar Harbor, ME, United States

P30 UTILIZATION OF WHOLE GENOME SNP PANELS FOR EFFICIENT GENETIC MAPPING IN THE MOUSE

JL Moran¹, A Bolton¹, J Moore², D Mirel², S Gabriel², DR Beier¹

¹Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, ²The Broad Institute of MIT and Harvard, Cambridge, MA, United States

P31 INTEGRATED ONLINE RESOURCES FOR THE MOUSE GENETIC RESEARCH COMMUNITY

DR Shaw, JA Blake, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig
The Jackson Laboratory, Bar Harbor, ME, United States

P32 LONG CONSERVED NONCODING SEQUENCES

T Kimura¹, Y Sakuraba², H Noguchi³, H Masuya², H Sezutsu⁴, K Takahashi², A Toyoda⁵, Y Sakaki⁵, R Fukumura², T Murata², N Fujimoto², R Motoi², M Yamamura¹, S Wakana², T Noda², T Shiroishi², Y Gondo²

¹Interdisciplinary Graduate School of Science and Engineering, Tokyo, Yokohama, Kanagawa, Japan, ²Functional Genomics Research Group, RIKEN Genomic Science Center (GSC), Yokohama, Kanagawa, Japan, ³Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, ⁴Transgenic Silkworm Research Center, National Institute of Agrobiological Science, Tsukuba, Ibaraki, Japan, ⁵Sequence Technology Team, RIKEN Genomic Science Center (GSC), Yokohama, Kanagawa, Japan

P33 ACCESSING PHENOTYPES: MOUSE MUTANTS AND HUMAN DISEASE MODELS IN MGI

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

P34 MOUSECYC: A CURATED BIOCHEMICAL PATHWAYS DATABASE FOR THE LABORATORY MOUSE

CJ Bult, E Patek
The Jackson Laboratory, Bar Harbor, ME, United States

P35 MOUSE GENOME RESOURCES AT NCBI

MJ Landrum, DM Church, the NCBI Genome Annotation Team, the CCDS Collaboration
National Center for Biotechnology Information, Bethesda, MD, United States

P36 THE MOUSE GENOME INFORMATICS DATABASE – HIGHLIGHTS OF WHAT'S NEW IN 2006

JR Lewis, JA Blake, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig
The Jackson Laboratory, Bar Harbor, ME, United States

P37 FINDING MY MOUSE: THE INTERNATIONAL MOUSE STRAIN RESOURCE (IMSR)

JT Eppig, B Richards-Smith, P Frost, DO Walton
The Jackson Laboratory, Bar Harbor, ME, United States

P38 WHAT'S NEW IN THE MOUSE TUMOR BIOLOGY DATABASE

JT Eppig, DA Begley, DM Krupke, MJ Vincent, JP Sundberg, CJ Bult
The Jackson Laboratory, Bar Harbor, ME, United States

- P39 OMICBROWSE DISPLAYS CAGE TAGS' STATISTICS ALONG WITH MULTIDIMENSIONAL OMICS ANNOTATIONS**
N Heida, N Kobayashi, T Endo, T Toyoda
 RIKEN Genomic Sciences Center, Yokohama, Kanagawa, Japan
- P40 EUROPHENOME: ONLINE MOUSE PHENOTYPING RESOURCE**
 S Greenaway¹, A Blake¹, A Shukla², ECJ Green¹, C Duran¹, A-M Mallon¹, JM Hancock¹
¹MRC Mammalian Genetics Unit, Harwell, United Kingdom, ²University of Oxford, Oxford, United Kingdom
- P41 A WEBSITE FOR MOUSE PHENOTYPE REPRESENTATION, PHENOSITE (PHENOTYPE STATISTICAL INFORMATION AND TERMINOLOGY OF EXPERIMENTS)**
H Masuya, N Heida, S Yoshikawa, H Motegi, T Furuse, O Minowa, H Toki, Y Wada, M Inoue, H Kaneda, K Kobayashi, T Toyoda, S Wakana, Y Gondo, T Noda, T Shiroishi
 RIKEN GSC, Tsukuba, Ibaraki, Japan
- P42 COEXPRESSING STRUCTURE REGULATED AT HIGHER LEVEL IN MOUSE GENOME**
T Endo¹, N Mise², K Abe², T Toyoda¹
¹Genomic Science Center, RIKEN, Yokohama, Kanagawa, Japan, ²BioResource Center, RIKEN, Tsukuba, Ibaraki, Japan
- P43 MUSBANKS: AN ARTIFICIAL INTELLIGENCE INTEGRATING WORLD-WIDE MOUSE MUTANT RESOURCES VIA BIOMOLECULAR NETWORKS**
N Kobayashi, N Heida, T Toyoda
 Riken GSC, Yokohama, Kanagawa, Japan
- P44 MODELLING REGULATORY NETWORK RESPONSES TO TRYPANOSOME INFECTION**
 H Hulme¹, G Sanguinetto², HA Noyes³, N Lawrence², L Zeef¹, SJ Kemp³, M Agaba⁴, M Rattray¹, A Brass¹
¹Department of Computer Science, University of Manchester, Manchester, United Kingdom, ²Department of Computer Science, University of Sheffield, Sheffield, United Kingdom, ³School of Biological Sciences, University of Liverpool, Liverpool, United Kingdom, ⁴International Livestock Research Institute (ILRI), Nairobi, Kenya
- P45 THE EUROPEAN MOUSE MUTANT ARCHIVE (EMMA)**
M Hagn¹, G Tocchini Valentini², Y Herault³, S Brown⁴, U Lendahl⁵, J Demengeot⁶, E Birney⁷, M Hrabe de Angelis¹
¹GSF National Research Center for Environment and Health, Munich, Germany, ²CNR, Consiglio Nazionale delle Ricerche, Monterotondo, Italy, ³CNRS, Centre National de la Recherche Scientifique, Centre de Distribution, de Typage et d'Archivage animale, Orleans, France, ⁴MRC, Medical Research Council, Mammalian Genetics Unit, Harwell, United Kingdom, ⁵Karolinska Institutet, Stockholm, Sweden, ⁶Fundacao C. Gulbenkian, Instituto Gulbenkian de Ciencia, Oeiras, Portugal, ⁷European Bioinformatics Institute, Hinxton, United Kingdom
- P46 BATCH QUERIES IN MGI**
KL Forthofer, DO Walton, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig, JA Blake
 The Jackson Laboratory, Bar Harbor, ME, United States
- P47 MOUSE PHENOME PROJECT**
MA Bogue, SC Grubb, TP Maddatu
 The Jackson Laboratory, Bar Harbor, ME, United States
- P48 MUTANT MOUSE MODELS OF HUMAN DISEASE IN THE JACKSON LABORATORY REPOSITORY**
DB Lane, D Bergstrom, B Change, LR Donahue, KR Johnson, CM Lutz, SF Rockwood, M Sasner, MT Davisson, the Repository Team
 The Jackson Laboratory, Bar Harbor, ME, United States

P49 SNP2RFLP: A COMPUTATIONAL TOOL TO FACILITATE GENETIC MAPPING USING BENCHTOP ANALYSIS OF SNPS

WA Beckstead², BC Bjork¹, DR Beier¹, S Sunyaev¹

¹Harvard Medical School, Boston, MA, United States, ²Brigham Young University, Provo, UT, United States

P28**ANNOTATION OF THE MOUSE GENOME: AUTOMATION MEETS MANUAL EXPERTISE**

C Amid, JP Almeida, S Donaldson, A Frankish, RC Gibson, EA Hart, K Kivinen, GK Laird, JE Loveland, JM Mudge, AO Mujica, JA Rajan, HK Sehra, CA Steward, MM Suner, MD Thomas, LG Wilming, JL Harrow
Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

After the initial sequencing and analysis of the mouse genome and with the ever growing interest in other genomes the major task remains the identification and accurate prediction of their entire gene sets. Automatic annotation is necessary as a first-pass analysis of genomes, but ab initio gene prediction algorithms typically predict fewer than 50% of known exons and 20% of complete genes accurately [1-3]. Other approaches, like Ensembl, incorporate homology to known expressed sequences or use sequence conservation between distantly related genomes [4]. However, the expertise of manual curators has been shown to be critical for accurate annotation of splice variants, pseudogenes and gene clusters. For the annotation of the mouse genome Ensembl automatic and Vega manually curated annotation resources at the Sanger Institute exchange and merge their information to provide the best possible gene data set to the community. An increasing number of projects seek curator expertise to validate gene structures and protein families of interest. An overview of current and future genomic annotation projects is presented.

[1] Guigo R and Wiehe T, Gene prediction accuracy in large DNA sequences, In: Frontiers in computational genomics, Caister Academic Press, Norfolk, UK (2003)

[2] Zhang MQ, Nat. Rev. Genet. 3 (2002) 698-709

[3] Rogic S, Mackworth AK, Quellet FB, Genome Research 11 (2001) 817-32

[4] Parra G, et al., Genome Research 13 (2003) 108-117

P29**MGI: ONE STOP "SHOPPING" FOR ORTHOLOGOUS INFORMATION GENE SETS IN MOUSE, HUMAN, AND RAT**

LJ Maltais, JA Blake, MA Dolan, M McAndrews-Hill, TBK Reddy, CJ Bult, JT Eppig, MGI Staff
The Jackson Laboratory, Bar Harbor, ME, United States

For a number of years the Mouse Genome Informatics (MGI) resource, www.informatics.jax.org, has provided orthology information about orthologous genes of mammalian species, especially for mouse, human and rat through the on-going curation efforts of the database and selective downloads of Homologene information. This work complemented the shared efforts of the mouse, human, and rat nomenclature committees.

Recently MGI expanded the knowledge available about mammalian gene orthologs by offering users an opportunity to obtain a wealth of information for the three species. The orthology reports now include links to data sources [HGNC, RGD], Entrez Gene, RefSeq, UniProt, Ensembl, Gene, comparative maps, references for all species or selected species, protein superfamily detail pages, VISTA on UCSC Browser, Homologene, and mouse-human-rat comparative GO graphs. In addition sequences for each species are provided with the ability to download one or more in FASTA format or forward to MouseBLAST, MGI's sequence comparison tool. Supporting documentation for the established orthologous relationships are displayed using evidence codes from the curation effort and published references.

Official mouse gene nomenclature located on the Marker Detail Page results from the extensive collaborations of the nomenclature groups, MGI, HGNC, and RGD. Assigning the equivalent gene symbol to orthologous genes reduces interspecies nomenclature confusion and allows researchers to easily retrieve and discuss specific genes.

HUGO Gene Nomenclature Committee

Rat Genome Database

Gene Ontology

Supported by NIH grant HG00330 and HG02273

P30

UTILIZATION OF WHOLE GENOME SNP PANELS FOR EFFICIENT GENETIC MAPPING IN THE MOUSE

JL Moran¹, A Bolton¹, J Moore², D Mirel², S Gabriel², DR Beier¹

¹Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, ²The Broad Institute of MIT and Harvard, Cambridge, MA, United States

We have previously described a fixed whole genome panel of 394 mouse SNPs that we used to successfully map 44 loci, including ENU-induced and spontaneous mutations, modifier loci, quantitative trait loci and loci that demonstrate loss-of-heterozygosity (Moran et al., Genome Research 2006). The major utility of a whole genome SNP panel is that complete genome haplotype characterization is obtained in a single analysis. This facilitates the efficient discrimination between true and false positive association, as well as the discovery of unexpected modifier effects.

We have now developed a new whole genome panel of 768 SNPs that is analyzed using the Illumina genotyping platform. SNP density averages 3 Mb across autosomes and 7 Mb across Chr. X. The 768 panel was designed to maximize the number of markers that would be informative for crosses made using C57BL/6J, such that the average number of informative SNPs between C57BL/6J and common inbred strains is 550. However, to insure that the panel is suitable for many strain combinations, a haplotype binning strategy was used to maintain informativeness across the genome between other strains. The new panel and analysis protocol allows for genetic mapping at higher resolution and higher throughput. Initial results using the panel demonstrate it is extremely robust with respect to both informativeness and allele discrimination, and we have successfully mapped 23 monogenic mutants (74% of mutants genotyped) thus far; some with as few as 3 affecteds. Utilization of this resource is available as a service for the mouse genetics community.

P31

INTEGRATED ONLINE RESOURCES FOR THE MOUSE GENETIC RESEARCH COMMUNITY

DR Shaw, JA Blake, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig

The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Genome Informatics Database (MGI, <http://www.informatics.jax.org/>) provides extensive information on the biology of the laboratory mouse by integrating data ranging from sequences to phenotypes. Core data include gene characterization and functions, phenotype and disease model descriptions, DNA and protein sequence data, SNPs and PCR polymorphisms and genetic and physical mapping data. Controlled vocabularies such as the Gene Ontology (GO), Mouse Phenotype (MP) Ontology and Mouse Anatomical Dictionary facilitate queries against the datasets. Another controlled vocabulary, the Human Disease Vocabulary, uses a catalog of human genetic disorders from Online Mendelian Inheritance in Man (OMIM) to associate mouse models with human diseases. Researchers can also query gene expression data detailing when and where a gene is expressed and what genes are expressed in specific tissues and developmental stages. MGI provides query forms that allow users to compose complex searches that simultaneously address functional information, phenotype, gene expression, sequence and mapping data. For example, users can find all genes mapped to Chromosome 2 between 150 and 168 Mbp whose products have transcription factor activity and where the gene has mutant alleles associated with vision or eye defects. Another search could find all SNPs in that region for selected strains. MGI is updated with new data nightly and continually adds new datasets and enhances the user interface.

MGI's dedicated User Support group is available at

mgi-help@informatics.jax.org.

MGI is supported by NIH grants HG00330, HG02273, HD33745.

P32**LONG CONSERVED NONCODING SEQUENCES**

T Kimura¹, Y Sakuraba², H Noguchi³, H Masuya², H Sezutsu⁴, K Takahashi², A Toyoda⁵, Y Sakaki⁵, R Fukumura², T Murata², N Fujimoto², R Motoi², M Yamamura¹, S Wakana², T Noda², T Shiroishi², Y Gondo²

¹Interdisciplinary Graduate School of Science and Engineering, Tokyo, Yokohama, Kanagawa, Japan, ²Functional Genomics Research Group, RIKEN Genomic Science Center (GSC), Yokohama, Kanagawa, Japan, ³Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan, ⁴Transgenic Silkworm Research Center, National Institute of Agrobiological Science, Tsukuba, Ibaraki, Japan, ⁵Sequence Technology Team, RIKEN Genomic Science Center (GSC), Yokohama, Kanagawa, Japan

Highly conserved genomic DNA sequences have been found between human and the other species by comparative sequence analysis even in noncoding sequences. By comparing such noncoding sequences after masking repetitive elements, we found 628 segments that are longer than 500 bp with more than 95% identity between human and mouse genomes. We call them Long Conserved Noncoding Sequences (LCNS). LCNS are distributed all over the genome in human and mouse. Many of the LCNS are highly conserved not only in other mammals but also in chicken and fish. In particular, we have intensively searched LCNS homologs in the noncoding sequences of dog, chicken, zebrafish and Tetraodon by using BLAST. As a result, we found 561 and 311 LCNS-like sequences that are longer than 500 bp in the dog and chicken genomes, respectively. Sixty-nine and fifty sequences in the zebrafish and Tetraodon genomes, respectively, also exhibited more than 200 bp homologies to the LCNS. In order to study the mechanism(s) of the conservation during evolution, we are conducting transcriptional and gene ontological analyses of the LCNS as well as the nearby genomic sequences to the LCNS. We are also investigating the mutation rate of the LCNS to examine if the LCNS are mutational cold spots or not.

P33**ACCESSING PHENOTYPES: MOUSE MUTANTS AND HUMAN DISEASE MODELS IN MGI**

H Onda, A Anagnostopoulos, R Babiuk, SM Bello, DL Burkart, MA Cassell, H Dene, I Lu, B Richards-Smith, CL Smith, M Tomczuk, LL Washburn, JT Eppig

Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine, United States

A primary goal of the Mouse Genome Informatics Database (MGI, <http://www.informatics.jax.org>) is to facilitate the use of mouse as a model system for studying human biology and disease. To this end, MGI integrates a wide range of data from DNA sequence to phenotype/disease. This integration enables analysis and querying of complex disease phenotypes in the context of underlying genetic cause, gene function, gene and protein sequence, expression data and genome location.

To describe phenotypic abnormalities and similarities to human disease, we are developing and utilizing a structured vocabulary of mouse anomalies (the Mammalian Phenotype Ontology) and human disease terms from the Online Mendelian Inheritance in Man (OMIM). These standard terms provide a backbone for robust annotation, allow comprehensive searches, and support web-based and computational access to phenotype and disease data.

Currently, MGI contains over 22,000 mouse phenotypic mutants and over 16,000 genotypes annotated with one or more phenotypic terms. Over 1,700 mouse models are associated with human diseases or syndromes.

We will describe how MGI can be used as a tool for comparing phenotypes, finding appropriate models for human disease, and mined for answers to complex biological questions.

Supported by NIH grant HG00330..

P34

MOUSECYC: A CURATED BIOCHEMICAL PATHWAYS DATABASE FOR THE LABORATORY MOUSE

CJ Bult, E Patek

The Jackson Laboratory, Bar Harbor, ME, United States

The availability of the complete genome sequence for the laboratory mouse provides a powerful platform for predicting genes and other genome features. However, building a catalog of genome annotations is just the beginning for biology in a "post-genome" era. To derive new insights into fundamental biological processes using complete genome sequences will require understanding how genome features interact in pathways and networks in the cell and how perturbations of these interactions contribute to disease processes. Toward this end, we have implemented a new database of curated biochemical pathways for the laboratory mouse called MouseCyc.

The MouseCyc database represents a significant advance for biomedical researchers wanting to access mouse genetic and genomic data in the context of physiological and cellular processes. The initial focus for the development of MouseCyc is on metabolism and includes such cell level processes as biosynthesis, degradation, energy production, and detoxification. MouseCyc differs from existing pathway databases and software tools because of the extent to which the pathway information in MouseCyc is integrated with the wealth of biological knowledge for the laboratory mouse that is available from the Mouse Genome Informatics (MGI) database.

Supported by NIH HG003622.

P35

MOUSE GENOME RESOURCES AT NCBI

MJ Landrum, DM Church, the NCBI Genome Annotation Team, the CCDS Collaboration

National Center for Biotechnology Information, Bethesda, MD, United States

In May 2006, NCBI released Build 36.1 of the mouse genome containing three genome-wide assemblies and seven strain specific partial assemblies. The reference assembly (based on C57BL/6J), produced by the Mouse Genome Sequencing Consortium (MGSC) and NCBI, is comprised of ~95% BAC-based finished sequence, with <2% of sequence based on draft sequence or whole genome shotgun (WGS) sequence from the Mouse Genome Sequencing Consortium version 3 (MGSCv3) assembly. The original MGSCv3 has been re-annotated and provided to allow assessment of the differences between the different assembly methods. The Celera mixed strain assembly is also provided. Over 50 Mb of sequence from other strains (predominantly variants of 129) are available and placed relative to the reference assembly when possible. All of the assemblies are annotated and available as part of the standard NCBI resource set: Map Viewer, BLAST, the NCBI ftp site, Entrez Gene, etc. The NCBI annotation is generated with a suite of software tools available from our website (<http://www.ncbi.nlm.nih.gov>), including Splign, a cDNA-to-genomic alignment program (<http://www.ncbi.nlm.nih.gov/sutils/splign/>), and Gnomon, a Hidden Markov Model-based gene prediction program (<http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.html>), available in Genome Workbench, <http://www.ncbi.nlm.nih.gov/projects/gbench/>). Current annotation provides prediction of gene models, a set of genomic and transcript based Reference Sequences, clone placement (BACs and fosmids based on end-sequence alignment), variation, STSs, Gene Trap clones, MICER clones, human and rat mRNAs, and locations of phenotypes defined by placement of genes or linked STS markers. In addition, annotation from outside sources, such as Ensembl, is available. Assessment of the NCBI and Ensembl gene models is ongoing with the goal of producing a Consensus CDS (CCDS) set for the mouse genome (<http://www.ncbi.nlm.nih.gov/projects/CCDS/>). Preliminary results from this effort as well as other tools allowing access to the genome and related information will be discussed.

P36**THE MOUSE GENOME INFORMATICS DATABASE – HIGHLIGHTS OF WHAT'S NEW IN 2006**

JR Lewis, JA Blake, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Genome Informatics (MGI) database resource integrates mouse genetic, genomic, and biological data to facilitate the use of mouse as a model system for understanding human biology and disease. MGI is continually adding features and integrating new data to improve this resource for its users. Visit <http://www.informatics.jax.org> to see this year's highlights.

Newly integrated data

- ⊙ Build 36 NCBI, Ensembl, and Vertebrate Genome Annotation Group (VEGA) gene models.
- ⊙ SNPs – includes calculation of SNPs located near genes.
- ⊙ Protein Information Resource (PIR) SuperFamilies – includes supporting sequences from mouse, human, and rat.
- ⊙ Deltagen and Lexicon data incorporated in concert with NIH acquisition of mice for placement in repositories.

Improved functionality

- ⊙ Query for data by a range of markers or genes – includes MIT markers.
- ⊙ Tab-delimited output for Marker results for download into EXCEL.
- ⊙ Orthology page enhancement to include supporting sequences.

New features

- ⊙ Initial Batch Queries for genes and markers with web or tab-delimited results.
- ⊙ New online book – *“Origins of Inbred Mice”*, editor, Herbert C. Morse III
- ⊙ Gene Ontology (GO) annotation graphs – a new display format for GO annotations.
- ⊙ GO comparative graphs – a new display format for mouse-human-rat GO annotations.

Supported by NIH grants HG00330, HG02273 and HD33745.

P37**FINDING MY MOUSE: THE INTERNATIONAL MOUSE STRAIN RESOURCE (IMSR)**

JT Eppig, B Richards-Smith, P Frost, DO Walton
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 7,600 strains and stocks from 15 repository consortia representing 25 individual repositories in the U.S., Canada, Europe, Japan, and Australia. In addition, over 13,600 ES cell lines from BayGenomics and the Nagy and Soriano laboratories are included. Additional mouse repositories' holdings and the International Gene Trap Consortium (<http://www.genetrap.org>) lines will soon be incorporated into the IMSR database.

Users can search 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.

When searched, IMSR returns a table of results satisfying the query and including the following information for each strain found: 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: additional details about the strain as provided on the holder's website; an email form for contacting the holder with questions or to order the mouse; and the Mouse Genome Informatics Database (MGI) 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”.

Supported by NIH grants MH061915, HG00330

P38

WHAT'S NEW IN THE MOUSE TUMOR BIOLOGY DATABASE

JT Eppig, DA Begley, DM Krupke, MJ Vincent, JP Sundberg, CJ Bult
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Tumor Biology Database (MTB, <http://tumor.informatics.jax.org>) facilitates selection of experimental models for cancer research, evaluation of mouse genetic models of human cancer, review of patterns of mutations in specific cancers, and identification of genes commonly mutated across a spectrum of cancers. Integrated searches, addressing complex ad-hoc questions, are enabled by use of controlled vocabularies and adherence to nomenclature standards.

Recently added MTB features include:

1. A new Mouse Pathology Submission interface allows contributors to create records for mice; specify strain and genetic mutations; generate tumor diagnoses including pathology and treatment descriptions; attach images to diagnoses; and perform edits and reviews of their submitted data.
2. An Advanced Search Form has been added allowing users to simultaneously address questions covering organ and tissue of tumor origin, tumor classification, induction treatment, mouse strain name and type, and specific genes or alleles.
3. The Mouse Tumor Frequency Grid, an interactive grid displaying tumor frequencies in different organs for inbred mice, is now in a new dynamic presentation format. Users can expand the Grid axes for individual strain families and organ systems, thus "zooming-in" on specific sub-strains and tissues of interest. Color-coded frequencies represented in the Grid cells are active, generating a database query for the underlying data. A mouse-over feature provides the highest tumor frequency observed in a strain as well as the number of reports contributing to the cell's value.

Our presentation will highlight examples using these new MTB features.

Supported by NIH grant CA89713

P39

OMICBROWSE DISPLAYS CAGE TAGS' STATISTICS ALONG WITH MULTIDIMENSIONAL OMICS ANNOTATIONS

N Heida, N Kobayashi, T Endo, T Toyoda
RIKEN Genomic Sciences Center, Yokohama, Kanagawa, Japan

OmicBrowse is a browser to explore multiple datasets coordinated in the omic space integrating omics knowledge ranging from genomes to phenomes and connecting evolutionary correspondences among multiple species. Here we present a new function of OmicBrowse, creating statistical analysis reports of CAGE expression-tags' data. OmicBrowse dynamically calculates the statistics of CAGE tags existing within a user-specified chromosomal region, so users can confirm the significance of expression levels of the transcripts in the region. Thus, OmicBrowse is highly appropriate for positional-cloning purposes. It displays both genetic maps and genomic annotations within wide chromosomal intervals and assists a user to select candidate genes by filtering their annotations or associated documents against user-specified keywords or ontology terms. As a genome browser OmicBrowse can color-code gene annotations by text-matching and significances of gene expressions by CAGE expression-tag analysis. OmicBrowse also assists users to share views by e-mail, to print out views, to register frequently-viewed chromosomal intervals, and to define groups of often browsed datasets as menu items. OmicBrowse integrates multiple data servers into a single omic space through secure peer-to-peer server communications, so that a user can easily obtain an integrated view of distributed data servers, e.g., an integrated view of numerous whole-genome tiling-array data retrieved from a user's in-house private-data server, along with various genomic annotations from public internet servers. OmicBrowse is developed by the Genome-Phenome Superbrain Project and is released as free open-source software under the GNU General Public License at <http://omicspace.riken.jp>.

P40**EUROPHENOME: ONLINE MOUSE PHENOTYPING RESOURCE**

S Greenaway¹, A Blake¹, A Shukla², ECJ Green¹, C Duran¹, A-M Mallon¹, JM Hancock¹

¹MRC Mammalian Genetics Unit, Harwell, United Kingdom, ²University of Oxford, Oxford, United Kingdom

The broad aim of biomedical science in the postgenomic era is to link genomic and phenotype information systematically to allow deeper understanding of the processes leading from genomic changes to altered phenotype and disease. Essential to developing such a linkage are databases which contain information on both normal phenotypes of inbred mouse strains and mutant phenotypes.

EuroPhenome (<http://www.europhenome.org>) is an online mouse phenotyping resource which was initially developed to store baseline data generated across a collection of standardised protocols (SOPs) called EMPReSS (<http://empres.har.mrc.ac.uk>). EMPReSS is the European Mouse Phenotyping Resource of Standardised Screens and was developed by groups of expert scientists to enable rapid, easy and reproducible assessment of phenotype in all major body systems. The baseline data were produced to validate the reproducibility of these protocols between laboratories.

Currently the database contains baseline data on 31 distinct phenotyping SOPs representing 6 phenotyping domains. In the future EuroPhenome will also be the repository for phenotype data generated by the EUMODIC project, which starts in 2007 and will phenotype up to 500 knock-out lines generated by the EUCOMM project. The EuroPhenome database is a MySQL relational database and has been designed to allow data from new SOPs or new projects to be added easily.

The EuroPhenome interface allows the user to access the data via the SOP, the phenotyping domain or from a Mammalian Phenotype ontology term. It also allows the user to visualise the data in a variety of ways, allowing assessment of inter-strain, -gender and -laboratory variation for the various protocols. The future aims for EuroPhenome are to provide other routes to the data such as via the genome and to extend the statistical analysis of the data.

P41**A WEBSITE FOR MOUSE PHENOTYPE REPRESENTATION, PHENOSITE (PHENOTYPE STATISTICAL INFORMATION AND TERMINOLOGY OF EXPERIMENTS)**

H Masuya, N Heida, S Yoshikawa, H Motegi, T Furuse, O Minowa, H Toki, Y Wada, M Inoue, H Kaneda, K Kobayashi, T Toyoda, S Wakana, Y Gondo, T Noda, T Shiroishi

RIKEN GSC, Tsukuba, Ibaraki, Japan

Construction of high throughput platforms for various phenotype assays in large-scale mutagenesis programs may bring potentially great progress on phenomics study in mouse. We have developed a database system "MUSBASE" for storing phenotype data as an infrastructure of the comprehensive phenotyping platform of the mouse ENU mutagenesis program in RIKEN GSC. Based on this system, we recently released new webpage named "PhenoSITE" (<http://www.gsc.riken.jp/Mouse/>). Through this website, we implemented some reusable methodologies for sharing phenotype information among many different laboratories. For example, we published our baseline data obtained with multiple inbred strains. We also propose ontology styled representation of the contents of phenotyping platform in RIKEN GSC. These systems would contribute to common method for interpretation of the phenotypes based on the measuring methods or experimental procedures. We plan to develop detailed SOP information page in PhenoSITE, which would provide fundamental structure for direct comparison of phenotype data generated from different laboratories. In addition, we are developing some "viewers" of outlier data of mutants generated from our mutagenesis program, which are simultaneously displayed on the baseline data. The "PhenoSITE" would contribute to integration of phenomics information as an international effort.

P42

COEXPRESSING STRUCTURE REGULATED AT HIGHER LEVEL IN MOUSE GENOME

T Endo¹, N Mise², K Abe², T Toyoda¹

¹Genomic Science Center, RIKEN, Yokohama, Kanagawa, Japan, ²BioResource Center, RIKEN, Tsukuba, Ibaraki, Japan

Gene transcription is not only regulated at the gene level but also at the polygenic level. There are some observations that genes (without apparently related functions) showing similar expression patterns are often clustered along the genome.

It is plausible that such coexpressing events correspond to some epigenetic phenomena such as DNA methylation and/or histone modification.

Coexpression was detected by calculating the correlation between the expression of two genes.

We hypothesized that a genomic region is a coexpressing section when the number of highly correlated pairs is greater than expected, even if the sources of cDNA are different. In order to display how frequent coexpressing regions are in the mouse genome, expression data from 24 samples including ES cells and adult somatic tissues were collected using microarray.

Statistical tests with window analysis revealed that there were many significant coexpressing regions in mouse chromosomes. Moreover, the coexpressing regions varied through cell types.

Somatic tissues have very few coexpressing regions shared with stem cells, and even in stem cells groups different coexpressing regions exist for cells derived from different developmental stages.

Our result implies that higher level transcriptional regulation may be achieved by epigenetic mechanisms operating at regional level.

P43

MUSBANKS: AN ARTIFICIAL INTELLIGENCE INTEGRATING WORLD-WIDE MOUSE MUTANT RESOURCES VIA BIOMOLECULAR NETWORKS

N Kobayashi, N Heida, T Toyoda

Riken GSC, Yokohama, Kanagawa, Japan

Numerous numbers of mouse mutants are being generated worldwide and databases featuring a catalog list of available mice resources have been constructed. Most of the catalog records only address a mutated gene and its phenotypic effects; however, they are not well linked to pathways or physiological networks, which are described in other databases, e.g. MEDLINE. The separation between mutant information and network information prevents biologists from the full use of the bio-resources. An intelligent database system is desired for biologists to search mutants through network information connecting a user-specified keyword and the gene altered in the mutants. MusBanks associates user's arbitrary keywords with the bio-resources via network information. First, MusBanks performs a full-text search of the MEDLINE abstracts, and finds significant gene names or symbols that exist within the matched abstracts with statistical tests, and makes a ranked list of the genes, and finally displays a list of the mutant mice whose altered genes are associated to the keywords. (i.e., MusBanks can search for connections of Keyword -> Medline -> Gene1 -> Mutant.) Simultaneously, MusBanks searches for related genes via estimated gene-gene relationships, and shows connections of Keyword -> Medline -> Gene1 -> Gene2 -> Mutant. MusBanks is a powerful tool which enables us to connect phenotypic functions and bio-resources via not only gene-gene interactions but also drug-protein and orthologs' data. MusBanks is a web application and available at <http://omicspace.riken.jp/>.

P44**MODELLING REGULATORY NETWORK RESPONSES TO TRYPANOSOME INFECTION**

H Hulme¹, G Sanguinetto², HA Noyes³, N Lawrence², L Zeef¹, SJ Kemp³, M Agaba⁴, M Rattray¹, A Brass¹

¹Department of Computer Science, University of Manchester, Manchester, United Kingdom, ²Department of Computer Science, University of Sheffield, Sheffield, United Kingdom, ³School of Biological Sciences, University of Liverpool, Liverpool, United Kingdom, ⁴International Livestock Research Institute (ILRI), Nairobi, Kenya

Identifying transcription factors that are controlling gene expression is a key step towards discovering regulatory networks. We have applied a Bayesian algorithm that was developed to identify regulatory networks using chromatin immuno-precipitation data to the identification of regulatory networks from gene expression data. The algorithm takes a list of transcription factor binding sites for each gene and a global expression data set and identifies the transcription factors that are controlling changes in gene expression. Transcription factor binding sites have been predicted for all human genes by the identification of conserved sequences in the upstream regions of human, dog, rat and mouse genes (*Nature*, **434**, 338-345). We have translated those predictions onto the orthologous murine genes and a set of gene expression data from the liver of two strains of mice infected with *Trypanosoma congolense* and sampled pre-infection and at four time points post infection. The algorithm predicted that transcription factors that have a plausible role in the response to infection were controlling this response. These predictions are currently being tested.

P45**THE EUROPEAN MOUSE MUTANT ARCHIVE (EMMA)**

M Hagn¹, G Tocchini Valentini², Y Herault³, S Brown⁴, U Lendahl⁵, J Demengeot⁶, E Birney⁷, M Hrabe de Angelis¹

¹GSF National Research Center for Environment and Health, Munich, Germany, ²CNR, Consiglio Nazionale delle Ricerche, Monterotondo, Italy, ³CNRS, Centre National de la Recherche Scientifique, Centre de Distribution, de Typage et d'Archivage animale, Orleans, France, ⁴MRC, Medical Research Council, Mammalian Genetics Unit, Harwell, United Kingdom, ⁵Karolinska Institutet, Stockholm, Sweden, ⁶Fondacao C. Gulbenkian, Instituto Gulbenkian de Ciencia, Oeiras, Portugal, ⁷European Bioinformatics Institute, Hinxton, United Kingdom

The EMMA (European Mouse Mutant Archive) was established as part of a worldwide network of repositories for retaining mutant strains essential for biomedical and medical research. Besides cryopreservation of mouse mutant lines in form of embryos or sperm, the distribution of these lines to qualified investigators is the main focus. EMMA maintained lines are supplied as a service to the research community at large and solely for research purposes. EMMA strains are available to academic institutions from all around the world. Applications for depositing and requesting mutant strains can be submitted through the EMMA website at www.emmanet.org.

In order to make the transfer of biological information more effective, appropriate databases will be further developed in the framework of the EMMA and in collaboration with existing databases e.g. the IMSR (International Mouse Strain Resource). An essential role of the EMMA project is to foster the virtual coupling of the EMMA stock centers and related information systems through the establishment and maintenance of a dedicated resource database (EMMA-RDB). This is implemented as a fully integrated web-based database and serves as an interface to the scientific community.

EMMA as part of the global resource network for mutant mouse lines (FIMRE, the International Federation of Mouse Resources) actively cooperates with other leading repositories like the MMRRRC in the US and the Riken Labs from Japan. EMMA is supported by primary public research institutions of the participating countries and by the European Commission FP 6 Research Infrastructures Programme.

P46

BATCH QUERIES IN MGI

KL Forthofer, DO Walton, CJ Bult, JA Kadin, JE Richardson, M Ringwald, JT Eppig, JA Blake
The Jackson Laboratory, Bar Harbor, ME, United States

The Mouse Genome Informatics (MGI) Database provides integrated access to genetic, genomic and biology data for the laboratory mouse. Prototypes for new functionalities can be found at <http://proto.informatics.jax.org/>, linked via "Additional MGI Tools and Links" on the MGI home page, <http://www.informatics.jax.org/>. A new prototype is the Gene/Marker Batch Query.

Using the Batch Query, researchers submit a list of sequence or gene accession identifiers from MGI or other sources and retrieve selected data from MGI for all of the identifiers in the list. Allowable IDs include GenBank/RefSeq, Entrez Gene, UniProt, Ensembl, Gene Ontology (GO), UniGene, or RefSNP. Users also may upload a file listing IDs in tab-delimited or comma-separated format. Duplicate identifiers are removed; identifiers that have no known association with data in MGI are identified as such.

Input IDs are cross-referenced against internal and external accession IDs in the MGI database for markers, their sequences, GO annotations, alleles, Mammalian Phenotype (MP) annotations, and nearby SNPs. Standard output data includes marker nomenclature, genome location, Ensembl ID, and Entrez Gene ID. Additional information may be obtained including GO annotation terms and IDs, MP annotation IDs, MGI allele IDs, GenBank IDs, RefSNP IDs, or UniProt IDs. Output formats include HTML or tab-delimited text allowing for easy import to spreadsheets such as Excel. We seek input regarding the usefulness of the batch query interface and what additional input and output combinations the community would like to see provided by MGI.

Supported by NIH grants: HG00330, HG02273.

P47

MOUSE PHENOME PROJECT

MA Bogue, SC Grubb, TP Maddatu
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 collect and provide phenotypic data in a central, public database. The project is an ongoing effort and continued expansion will assist the research community in maximizing the use of quantitative phenotypic data together with emerging genomic sequence, SNP, and haplotype data. The web interface for the Mouse Phenome Database (MPD) provides tools for online analysis and enables data retrieval/downloads for offline custom analyses. Data for a wide range of parameters are annotated and stored in the MPD along with submitter's contact information, detailed protocols, environmental parameters, and other relevant information. In addition, community SNPs are collected, consolidated, and made available through a SNP interface. The MPD SNP collection contains strain alleles for 7.5+ million genome-wide locations from multiple sources. All SNPs are mapped to the current NCBI mouse genome build, and dbSNP, NCBI, and Ensembl annotations are provided along with links to MGI Mouse GBrowse. With recently developed MPD tools, users are able to view a set of phenotypic data from available strains alongside SNPs within specified genes or regions. These tools and other site updates will be demonstrated.

Mouse Phenome Database (MPD)

www.jax.org/phenome

P48**MUTANT MOUSE MODELS OF HUMAN DISEASE IN THE JACKSON LABORATORY REPOSITORY**

DB Lane, D Bergstrom, B Change, LR Donahue, KR Johnson, CM Lutz, SF Rockwood, M Sasner, MT Davisson, The Repository Team

The Jackson Laboratory, Bar Harbor, ME, United States

The Jackson Laboratory Repository, established as a centralized resource for the international scientific community, currently contains nearly 3000 mouse strains and distributes over 33,000 mice to researchers annually. The Repository adds 200 to 300 new biomedically significant mouse strains per year by importation from external scientists and by development of new spontaneous mutation models within our own program. Mutant mouse strain models for cancer, Alzheimer's disease, cystic fibrosis, X-linked Alport syndrome and inflammatory bowel disease were recently added to the collection of mice held by the Repository. Mutant mouse sets from Deltagen (NIH-sponsored) and the Consortium for Functional Glycomics were also recently accepted. The Repository includes conditional mutation mouse strains, which allow investigators to control temporal and/or tissue-specific expression, and track expression with reporter genes.

The Repository, charged with importing, developing, and preserving these strains, maintains an on-line resource (<http://jaxmice.jax.org/query/>) that includes strain phenotype and genetic descriptions, information about strain development and maintenance, licensing requirements, a list of related references, and other relevant information.

Researchers can submit candidate strains using the on line form available at The Jackson Laboratory website at: <http://www.jax.org/grc/index.html>. Strain donation to the Repository fulfills the requirements for the sharing of mice in accordance with the NIH Model Organism Sharing Policy.

The Jackson Laboratory Repository is supported by the NCR (RR09781, RR11083, RR16049), NIA, The Howard Hughes Medical Institute, The Ellison Medical Foundation and donations from several private charitable foundations.

P49**SNP2RFLP: A COMPUTATIONAL TOOL TO FACILITATE GENETIC MAPPING USING BENCHTOP ANALYSIS OF SNPS**

WA Beckstead², BC Bjork¹, DR Beier¹, S Sunyaev¹

¹Harvard Medical School, Boston, MA, United States, ²Brigham Young University, Provo, UT, United States

Many mouse gene mutations cause phenotypes that serve as models of human genetic disorders. Mapping and positional cloning of these mutations accelerates our understanding of the mouse gene, its human ortholog and the underlying etiology of the disorder. SNP markers have been shown to give high resolution in genetic mapping because they are abundant throughout the genome and can be analyzed in a high-throughput manner using automated technology. However, initial localization via a genome-wide SNP panel often defines a large chromosomal interval, and "bench-top" technologies for fine-mapping using SNPs are often inefficient. We have developed a web-based tool, SNP2RFLP, that can extract region-specific SNPs from dbSNP and identify those SNPs that create restriction fragment length polymorphisms (RFLPs) that can be easily assayed by restriction enzyme digestion of SNP-containing PCR products. The input to SNP2RFLP is the two mouse strains used in the cross, the chromosomal region, and a set of restriction endonucleases. SNP2RFLP extracts the SNPs from dbSNP that are polymorphic between the two strains in the region in question. The program simulates a restriction digest of the SNP-containing sequences with each enzyme to determine whether the SNP creates an RFLP. Informative markers are then analyzed using Primer3, which finds suitable PCR primers surrounding the SNP. The output of SNP2RFLP is the informative SNPs that create RFLPs and the left and right PCR primers. This information can then be used to readily perform the RFLP assays and further refine the region containing the mutation of interest. SNP2RFLP should prove valuable to the general mouse genetics community and is publicly available at <http://genetics.bwh.harvard.edu/snp2rflp>.

