



Sunday November 6

10.00am – 12.30pm

Poster Session 1

Genome Sequencing & Comparative Analysis Pages 96 - 110

Posters P-15 – P-37

Genome Sequencing & Comparative Analysis

- P-15 COMPLETE SEQUENCE OF THE MITOCHONDRIAL GENOME OF *MUS MUSCULUS MUSCULUS* STRAIN PWD/PH**
Z Trachtulec, J Forejt
 Institute of Molecular Genetics, Acad.Sci.Czech Rep., Czech Republic
- P-16 OVERLAPPING TRANSCRIPTION AND IMPRINTING STATUS OF THE GENES AROUND THE *PDCD2* LOCUS ON MOUSE CHROMOSOME 17**
O Mihola, Z Trachtulec, J Forejt
 Institute Mol. Genet., Acad.Sci., Prague, Czech Republic
- P-17 CHARACTERISATION OF 1.5 MB OF THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS OF THE 129/SV MOUSE**
I Retter¹, M Scharfe², A Mauhar³, C Chevillard³, M Ludewig², G Nordsiek², H Blöcker², R Riblet³, W Müller¹
¹German Research Centre of Biotechnology, Department of Experimental Immunology, Braunschweig, Germany ²German Research Centre of Biotechnology, Department of Genome Analysis, Braunschweig, Germany ³Torrey Pines Institute for Molecular Studies, San Diego, USA
- P-18 GENOME-WIDE HIGH RESOLUTION EXPRESSION MAPPING IN THE ADULT MOUSE BRAIN**
S Sunkin, C Dang, M Hawrylycz, E Lein, P Wohnoutka, A Jones
 Allen Institute for Brain Science, Seattle, WA, United States
- P-19 EVOLUTIONARY BREAKPOINTS IN MAMMALIAN, BIRD AND FISH CHROMOSOMES**
R W Elliott
 Roswell Park Cancer Inst., Buffalo, NY, United States
- P-20 ORIGINS OF MOUSE INBRED STRAINS DEDUCED FROM WHOLE-GENOME SCANNING BY POLYMORPHIC MICROSATELLITE LOCI**
H Yonekawa¹, T Sakai¹, Y Kikkawa¹, I Miura², K Moriwaki², T Shiroishi³, Y Satta⁴, N Takahata⁴
¹ Tokyo Metro. Inst. Med. Sci., Tokyo, Japan, ² BRC, Riken, Ibaraki, Japan, ³ Nat. Inst. Genetics, Shizuoka, Japan, ⁴ Grad. Univ. Adv. Stu., Kanagawa, Japan
- P-21 THE MAJOR URINARY PROTEINS: GENOMIC ORGANISATION, FUNCTIONALITY AND ONGOING EVOLUTION**
J M Mudge¹, S D Armstrong², D H L Robertson², S A Cheetham², M Thom², R J Beynon², J L Hurst², L G Wilming¹, J L Harrow¹, T Hubbard¹
¹ Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, ² University of Liverpool, Liverpool, United Kingdom
- P-22 GENE-DRIVEN SCREENING OF ENU-INDUCED MUTATIONS IN THE CIS-REGULATORY REGION CONTROLLING SHH EXPRESSION IN DEVELOPING MOUSE LIMB BUDS**
H Masuya¹, H Sezutsu¹, Y Sakuraba¹, T Sagai², M Hosoya², I Miura¹, A Shimizu¹, J Nagano¹, S Kaneko¹, H Yokoyama¹, T Noda¹, Y Gondo¹, S Wakana¹, T Shiroishi¹
¹ RIKEN GSC, Tsukuba, Ibaraki, Japan, ² National Institute of Genetics, Mishima, Shizuoka, Japan
- P-23 IMGT OVERVIEW:**
1. THE MOUSE IMMUNOGLOBULIN HEAVY IGH GENES
 G Folch, V Giudicelli, C Jean, C Ginestoux, G Lefranc, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-24 IMGT OVERVIEW:**
2. THE MOUSE IMMUNOGLOBULIN KAPPA IGK GENES
 G Folch, C Jean, C Ginestoux, V Giudicelli, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France

- P-25 IMGT OVERVIEW:
3. THE MOUSE IMMUNOGLOBULIN LAMBDA IGL GENES**
G Folch, D Scaviner, D Guiraudou, C Ginestoux, V Giudicelli, G Lefranc, [M-P Lefranc](#)
IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-26 IMGT OVERVIEW:
4. THE MOUSE T CELL RECEPTOR ALPHA TRA GENES**
N Bosc, D Scaviner, G Folch, C Ginestoux, V Giudicelli, [M-P Lefranc](#)
IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-27 IMGT OVERVIEW:
5. THE MOUSE T CELL RECEPTOR BETA TRB GENES**
N Bosc, G Folch, C Ginestoux, V Giudicelli, [M-P Lefranc](#)
IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-28 IMGT OVERVIEW:
6. THE MOUSE T CELL RECEPTOR GAMMA TRG GENES**
G Folch, D Scaviner, C Ginestoux, V Giudicelli, [M-P Lefranc](#)
IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-29 IMGT OVERVIEW:
7. THE MOUSE T CELL RECEPTOR DELTA TRD GENES**
N Bosc, G Folch, C Ginestoux, V Giudicelli, [M-P Lefranc](#)
IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France
- P-30 COMPARATIVE GENOME MAPPING OF THE DEER MOUSE (*PEROMYSCUS MANICULATUS*) WITH REFERENCE TO *MUS MUSCULUS***
C M Ramsdell, E L Thames, J L Weston, M J Dewey
Peromyscus Genetic Stock Center, Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA
- P-31 EFFICIENT GENETIC MAPPING IN THE MOUSE USING A WHOLE GENOME SNP PANEL**
[J L Moran](#)¹, A D Bolton¹, A Brown², N Dwyer³, D Nelson¹, P Tran¹, Y Yun¹, A Chesebro¹, B Bjork¹, C Li⁴, D J Kwiatkowski², R Kucherlapati², T Wiltshire⁵, D R Beier¹
¹ Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, ² Harvard Partners Center for Genetics and Genomics, Cambridge, MA, United States, ³ Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, ⁴ Dana Farber Cancer Institute, Harvard School of Public Health, Boston, MA, United States, ⁵ Genomics Institute of the Novartis Research Foundation, San Diego, MA, United States
- P-32 BIOLOGICAL FUNCTIONS UNDERLYING THE LINKAGE DISEQUILIBRIUM DOMAINS OF INBRED MICE**
[K Paigen](#), B L King, P Petkov, J H Graber, G Churchill, N Herz
The Jackson Laboratory, Bar Harbor, ME, United States
- P-33 CLONED SINGLE-MOLECULE ARRAYS FOR GENOME-WIDE RESEQUENCING**
[T Smith](#)
Solexa Inc., Little Chesterford, Cambridge, United Kingdom
- P-34 FORWARD AND REVERSE MUTAGENESIS SCREENS OF A SUBSET OF MOUSE GENES AND REGULATORY REGIONS**
[J S Weber](#)¹, G Boschult¹, G Schnell², J Kovar³
¹ University of Nebraska, Lincoln, NE, United States, ² University of North Carolina, Chapel Hill, NC, United States, ³ LI-COR Biosciences, Lincoln, NE, United States

P-35 ARRAY CGH ON MOUSE USING TOTAL GENOMIC DNA WITH OLIGONUCLEOTIDE ARRAYS

N Sampas, E Lin, A Scheffer, P Tsang, L Bruhn, K Shannon
Agilent Technologies, Palo Alto, United States

P-36 FUNCTIONAL GENOMICS OF ALL MOUSE PROTEIN KINASES AND PHOSPHATASES

A R R Forrest¹, D F Taylor¹, M L Crowe¹, N J Waddell^{1,2}, G Kolle¹, R Kodzius^{5,6}, S Katayama^{5,6}, J L Fink¹, M M Gongora¹, C Flegg¹, H Suzuki³, M Kanamori³, C Kai³, J Kawai³, P Carninci³, Y Hayashizaki³, S M Grimmond¹.

¹ Institute for Molecular Bioscience and ARC Centre in Bioinformatics, Queensland, Australia, ²Queensland Institute for Medical Research, Queensland, Australia, ³Genome Exploration Research Group (Genome Network Project Core Group), ⁴RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Kanagawa, Japan, ⁵Genome Science Laboratory, Discovery Research Institute, ⁶ RIKEN Wako Institute, Saitama, Japan

P-37 MOUSE GENOME RESOURCES AT NCBI

D M Church, M Landrum and the NCBI Genome Annotation Team
DHHS/NIH/NLM/NCBI; Bethesda, MD

P-38 STUDENT ORAL ABSTRACT S-1

SEQUENCE CONSERVATION AT ODZ4, A COMPLEX LOCUS REQUIRED FOR MOUSE DEVELOPMENT, INDICATES THAT MAMMALIAN GENE REGULATION SPANS 450 MILLION YEARS OF EVOLUTION

A C Lossie, T Asgedom and MJ Justice
Baylor College of Medicine, Houston, Texas, United States

P-39 STUDENT ORAL ABSTRACT S-2

THE REGULATORY MUTATION, MVWF1, IS A COMMON M. M. DOMESTICUS ALLELE AND THE MAJOR CAUSE OF PROLONGED A PTT IN MICE

J Johnsen¹, J Baines², D Tautz³, and D Ginsburg^{1,4}
¹University of Michigan, USA. ²University of Munich, Germany. ³University of Cologne, Germany. ⁴Howard Hughes Medical Institute, USA.

P-15**COMPLETE SEQUENCE OF THE MITOCHONDRIAL GENOME OF *MUS MUSCULUS MUSCULUS* STRAIN PWD/PH**Z Trachtulec, J Forejt

Institute of Molecular Genetics, Acad.Sci.Czech Rep., Czech Republic

There is an increasing acknowledgement of the importance of genes encoded by the mitochondrial genome. Here we present the complete mitochondrial sequence of the wild-derived strain PWD/Ph. The analysis of the restriction sites of several endonucleases characteristic for murine subspecies confirmed its *M. m. musculus* origin. Our sequence is therefore the first complete mitochondrial genome sequence of *M. m. musculus*. The PWD/Ph mitochondrial genome sequence is about three times as divergent from C57BL/6J as the nuclear genome sequence. The PWD/Ph mitochondria-encoded proteins are 98.3 to 100% identical to the C57BL/6J proteins. A conplastic strain carrying the PWD/Ph mitochondrial genome along with C57BL/6J nuclear genome was prepared in our laboratory and our mitochondrial sequence will play an important role in the clarification of phenotypes of this strain.

This work was supported by grant No. A5052406 from the Grant Agency of the ASCR.

P-16**OVERLAPPING TRANSCRIPTION AND IMPRINTING STATUS OF THE GENES AROUND THE *PDCD2* LOCUS ON MOUSE CHROMOSOME 17**O Mihola, Z Trachtulec, J Forejt

Institute Mol. Genet., Acad.Sci., Prague, Czech Republic

The Programd cell death 2 (*Pdcd2*) gene on mouse chromosome 17 was evaluated as a member of a highly conserved synteny, a candidate for an imprinted locus, and a candidate for the *Hybrid sterility 1* (*Hst1*) gene. Three new transcripts were identified at this locus: an alternative *Pdcd2* mRNA skipping the last two exons, an antisense RNA overlapping the entire alternative exon and a part of the intron unique to the alternative mRNA and another antisense transcript ending upstream *Pdcd2* (*Updcd2*). The antisense RNA was determined to be an alternative isoform of the neighboring *Tbp* gene. *Updcd2* could be a new noncoding gene sharing the promotor with *Pdcd2* or another alternative *Tbp* isoform overlapping the entire *Pdcd2* gene. Allelic sequencing and transcription studies did not reveal any support for the candidacy of *Pdcd2* for *Hst1*. A relationship between the transcription of alternative or overlapping RNAs was sought by molecular and bioinformatical methods. No correlated expression of *Pdcd2* with other two genes of the highly conserved synteny was observed. Moreover, *Pdcd2* and five other genes from this region were not imprinted in embryonic tissues obtained from crosses of mouse strains PWD/Ph and C57BL/6J.

This work was supported by grant No. 301/05/0738 from the Czech Science Foundation and No. 1M6837805002 – Center for Applied Genomics - from the Czech Ministry of Education, Youth and Sports.

P-17

CHARACTERISATION OF 1.5 MB OF THE IMMUNOGLOBULIN HEAVY CHAIN LOCUS OF THE 129/SV MOUSE

I Retter¹, M Scharfe², A Mauhar³, C Chevillard³, M Ludewig², G Nordsiek², H Blöcker², R Riblet³, W Müller¹

¹German Research Centre of Biotechnology, Department of Experimental Immunology, Mascheroder Weg 1, 38124 Braunschweig, Germany ²German Research Centre of Biotechnology, Department of Genome Analysis, Mascheroder Weg 1, 38124 Braunschweig, Germany ³Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, CA 92121-1122, USA

The sequence of the proximal part of the immunoglobulin heavy chain locus (IgH) of the 129/Sv mouse strain has been elucidated and analysed. 117 variable genes (V genes) and pseudo genes have been identified, representing 12 of 15 murine IgH-V gene families. 47 V genes are supposed to be functional. Of that, 27 sequences are described in the germline configuration for the first time. Furthermore, 2 new D segments have been identified. The order of D segments in the 129/Sv strain is DFI16.3, DST4.2, DFI16.1, DSP2.9, DSP2.3, DST4.3, DFI16.2, DSP2.5, DSP2.2, DSP2.11, DSP2.2, DSP2.8, DSP2.7, DST4, DQ52.

The availability of 1.5 Mb genomic sequence of the IgH locus allows the comparison between different mouse strains on the level of the coding sequence as well as on the level of the locus' structure. The 129/Sv mouse has been assigned to the IgH^a haplotype. Unexpectedly, the comparison of the D region, J region and constant region genes revealed differences to the BALB/c strain which also has the IgH^a haplotype. The comparison of the two IgH^a strains to the C57BL/6 strain of the IgH^b haplotype will give insight to the evolution of immunoglobulin variable gene segments. The genomic sequences are available from the EMBL database by the accession numbers AJ851868-AJ851885.

P-18

GENOME-WIDE HIGH RESOLUTION EXPRESSION MAPPING IN THE ADULT MOUSE BRAIN

S Sunkin, C Dang, M Hawrylycz, E Lein, P Wohnoutka, A Jones
Allen Institute for Brain Science, Seattle, WA, United States

The Mouse Genome Sequencing Consortium has shown that approximately 99% of mouse genes have a homologue in the human genome, supporting the utility of the mouse model for examining gene expression. Several techniques available to study large-scale gene expression, such as microarrays, real-time polymerase chain reaction, and serial analysis of gene expression, have been applied to expression analysis of large structures in the brain; however, these techniques lack the resolution to distinguish cellular heterogeneity within these regions. The Allen Institute for Brain Science (AIBS) has developed a high-throughput *in situ* hybridization platform to generate gene expression maps of the mouse brain. The goal of our inaugural project, the Allen Brain Atlas (ABA), is to map the expression of ~18,000 genes in C57BL/6J. These data will be searchable through the atlas database, a resource that will assist in more global approaches to the molecular analysis of the nervous system. Currently, images of expression patterns for over 6,000 genes are in the database. Analysis of the results obtained for the first several thousand genes indicates a tremendous diversity of cell types in the brain. Numerous genes show tremendous specificity for functionally discrete brain regions, and often particular cell types within these regions. In some cases these genes subdivide known anatomical regions, suggesting further functional divisions within these regions. This regional gene expression provides the means to correlate molecular function and phenotypic differences between different cell types in the brain. Images of gene expression patterns and analytic tools can be viewed at www.brainatlas.org.

P-19**EVOLUTIONARY BREAKPOINTS IN MAMMALIAN, BIRD AND FISH CHROMOSOMES**

R W Elliott

Roswell Park Cancer Inst., Buffalo, NY, United States

Positions of loci on sequence-based maps of several mouse chromosomes have been compared with the positions of orthologs in rat, dog, chicken, zebrafish and human chromosomes. Conserved regions between rat and mouse are very long, often consisting of whole chromosomes. Mouse Chr 17 shows the largest number of evolutionary breakpoints with the rat genome. These breaks have been analyzed relative to the human map, as well as to another mammal, the dog, and a bird, the chicken. Most of these breaks represent ancestral breaks already present in the mammalian line. A few represent inversions that occurred in the rodent line. Many of the breaks between mouse and chicken are also found between mouse and dog or mouse and human. However, some are unique to the chicken. Conserved regions between fish and mammalian genomes are very short, often containing only two genes.

P-20**ORIGINS OF MOUSE INBRED STRAINS DEDUCED FROM WHOLE-GENOME SCANNING BY POLYMORPHIC MICROSATELLITE LOCI**H Yonekawa¹, T Sakai¹, Y Kikkawa¹, I Miura², K Moriwaki², T Shiroishi³, Y Satta⁴, N Takahata⁴¹ Tokyo Metro. Inst. Med. Sci., Tokyo, Japan, ² BRC, Riken, Ibaraki, Japan, ³ Nat. Inst. Genetics, Shizuoka, Japan,⁴ Grad. Univ. Adv. Stu., Kanagawa, Japan

Microsatellite loci are uniformly distributed at approximately 100 kbp intervals on all chromosomes except the chromosome Y, and genetic information about more than 9,000 loci and high-throughput polymorphism analysis are now available. Taking advantage of these properties, we carried out whole genome scanning using 8 common inbred strains of laboratory mice (CIS) including A/J, C57BL/6J, CBA/J, DBA/2J, SM/J, SWR/J, NC/Nga, 129/SvJ and 8 wild-derived inbred strains (WIS), BGL2/Ms, CAST/Ei, JF1/Ms, MSM/Ms, NJL/Ms, PGN2/Ms, SK/CamEi and SWN/Ms. We selected and located 1226 informative loci at 1.2 cM average intervals on all of the chromosomes of the sixteen strains, and compared the polymorphisms of the eight CIS with those from the eight WIS as subspecies representatives. More than 50 % of the loci can be identified as WIS- (therefore, subspecies specific-) alleles in the CIS genomes. We also discovered that the CIS chromosomes form a mosaic structure with an average ratio of *domesticus*- to non-*domesticus* alleles of 3:1. Furthermore, the *domesticus*- alleles were present much more frequently on the CIS chromosome X than on their autosomes, suggesting that successive backcrossing of non-*domesticus*-stocks to *domesticus*-stocks had been undergone at the beginning of CIS history.

P-21**THE MAJOR URINARY PROTEINS: GENOMIC ORGANISATION, FUNCTIONALITY AND ONGOING EVOLUTION**

J M Mudge¹, S D Armstrong², D H L Robertson², S A Cheetham², M Thom², R J Beynon², J L Hurst², L G Wilming¹, J L Harrow¹, T Hubbard¹

¹ Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, ² University of Liverpool, Liverpool, United Kingdom

Male *Mus musculus* mice are highly territorial, marking their range with urine deposits which modulate male and female behaviour. This information transfer is facilitated by the obligate excretion of the Major Urinary Proteins (MUPs). Phenotypic data indicates that MUPs are highly polymorphic within wild populations; only close siblings express the same complement of molecules. Biochemical and behavioural studies have shown that mice respond to these differences, indicating that MUPs confer a consistent signature of recognition. Whilst MUP polymorphism is well documented at the protein level in wild and inbred populations, corresponding genomic data is lacking. Here, we describe the content, architecture and evolution of the MUP gene cluster on chromosome 4 within C57BL6 mice. We describe multiple genes, and structural and phylogenetic analyses allow us to classify three regions within the cluster, differing in terms of age, organisation and potential to exhibit ongoing genomic instability. We analyse MUP expression at the protein level and correlate this to the annotated genome, identifying genes which demonstrate strong sexual dimorphism. Many questions regarding MUP function and evolution remain, and to this end we describe the outlines of a multi-disciplinary comparative project, beginning with resequencing and phenotyping in another inbred strain. The mechanisms by which evolution operates at the level of behaviour remain poorly understood. The discernable chain of causation from MUP genotype to expression profile to behavioural response suggests that a thorough study of this gene family will be of great interest within the field of evolutionary biology.

P-22**GENE-DRIVEN SCREENING OF ENU-INDUCED MUTATIONS IN THE CIS-REGULATORY REGION CONTROLLING SHH EXPRESSION IN DEVELOPING MOUSE LIMB BUDS**

H Masuya¹, H Sezutsu¹, Y Sakuraba¹, T Sagai², M Hosoya², I Miura¹, A Shimizu¹, J Nagano¹, S Kaneko¹, H Yokoyama¹, T Noda¹, Y Gondo¹, S Wakana¹, T Shiroishi¹

¹ RIKEN GSC, Tsukuba, Ibaraki, Japan, ² National Institute of Genetics, Mishima, Shizuoka, Japan

Comparative genomics approach reveals non-coding sequences that are conserved among evolutionally remote species. In many cases, it is difficult to predict functions of those sequences. Large-scale ENU-mutagenesis enables us to obtain point mutations in any given regions in the mouse genome. Here we report gene-driven approach in a large-scale ENU-mutagenesis program for screening mutations in mammals-fish-conserved-sequence 1 (MFCS1), which is highly conserved sequence that acts limb specific cis-regulator of Sonic hedgehog (*Shh*). Mouse preaxial polydactyly (PPD) mutations showed ectopic expression *Shh* in the anterior mesenchyme of the limb buds. MFCS1-KO mice completely lose *Shh* expression in the limb buds. Hence those mouse mutations demonstrated that MFCS1 have a major limb-specific *Shh* enhance and repressor(s) to down-regulate *Shh* expression in anterior mesenchyme of developing limb buds. In this study, we conducted gene-driven screening for this region, and have obtained three new point mutations (single-base substitutions). Phenotype analysis of these three mutations indicated that a mutation *M101116* exhibits PPD phenotype, although the penetrance is lower than those of existing PPD mutations, *M100081* and *Hx*. In contrast, two other mutations *M101117* and *M101192* showed no abnormality of the limbs in hetero- and homozygous mice. The result of this study and previous ones suggested that multiple specific-sites are involved in down-regulation of MFCS1 in anterior mesenchyme of limb buds. Thus, gene-driven screening for ENU-mutagenized mice is a powerful approach to explore function of conserved non-coding sequences that are revealed by comparative genomics.

P-23**IMGT OVERVIEW:****1. THE MOUSE IMMUNOGLOBULIN HEAVY IGH GENES**

G Folch, V Giudicelli, C Jean, C Ginestoux, G Lefranc, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH,
 Montpellier, France

The total number of known immunoglobulin heavy IGH genes per haploid genome in *Mus musculus* laboratory mice is 175-179, with a predicted number of 199-203 (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse IGH genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The definitive IMGT nomenclature of the mouse IGHV genes and the correspondence with the provisional nomenclature have been established. The mouse IGH locus is located on chromosome 12 at 58.0 cM. It comprises, in laboratory mice, 146 IGHV (with a predicted number of 170) that belong to 15 subgroups, 17-20 IGHD (17 in BALB/c and 20 in C57BL/6J), 4 IGHJ and 8-9 IGHC genes, depending on the strains or species. The potential IGH repertoire per haploid genome comprises 130-135 functional genes (on the 175-179 genes): 108 IGHV (on the 146), 10-14 IGHD (10 in C57BL/6J and 14 in BALB/c) that belong to four subgroups, 4 IGHJ, and 8-9 IGHC genes. Analysis of the IGHV genes (germline or rearranged) can be performed by IMGT/V-QUEST [3], and analysis of the V-D-J junctions by IMGT/JunctionAnalysis [4]. IMGT Colliers de Perles and three-dimensional structures of 422 VH domains encoded by rearranged IGHV-IGHD-IGHJ genes are available in IMGT/3Dstructure-DB [5].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] Nucl. Acids Res. 32, W435-440 (2004). [4] Bioinformatics 20, I379-I385 (2004). [5] Nucl. Acids Res. 32, D208-D210 (2004).

P-24**IMGT OVERVIEW:****2. THE MOUSE IMMUNOGLOBULIN KAPPA IGK GENES**

G Folch, C Jean, C Ginestoux, V Giudicelli, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH,
 Montpellier, France

The total number of immunoglobulin kappa IGK genes per haploid genome in *Mus musculus* laboratory mice is 180 (190, if the 10 IGKV orphans are included) (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse IGK genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The mouse (*Mus musculus*) IGK locus is located on chromosome 6 at 30.0 cM. It spans 3200 kb and consists of 174 IGKV genes that belong to 19 subgroups and to three clans, 5 IGKJ genes and a single IGKC. The potential IGK repertoire per haploid genome comprises 99-101 functional genes: 94-96 IGKV, 4 IGKJ and 1 IGKC [3]. Analysis of the IGKV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-J junctions by IMGT/JunctionAnalysis [5]. Mouse IGK gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. IMGT Colliers de Perles and three-dimensional structures of 396 V-KAPPA domains encoded by rearranged IGKV-IGKJ genes are available in IMGT/3Dstructure-DB [6].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] Exp. Clin. Immunogenet. 18, 255-279 (2001). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2004). [6] Nucl. Acids Res. 32, D208-D210 (2004).

P-25**IMGT OVERVIEW:****3. THE MOUSE IMMUNOGLOBULIN LAMBDA IGL GENES**

G Folch, D Scaviner, D Guiraudou, C Ginestoux, V Giudicelli, G Lefranc, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH,
 Montpellier, France

The total number of immunoglobulin lambda IGL genes per haploid genome in *Mus musculus* laboratory mice is 17 (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse IGL genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The mouse (*Mus musculus*) IGL locus is located on chromosome 16 at 13 cM. It spans 240 kb and consists, in laboratory mice, of 8 IGLV genes that belong to 2 subgroups (a third IGLV subgroup has been identified in wild mice), 5 IGLJ and 4 IGLC genes [3]. The potential IGL repertoire per haploid genome comprises 13-14 functional genes: 8 IGLV, 3 IGLJ and 2-3 IGLC. Analysis of the IGLV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-J junctions by IMGT/JunctionAnalysis [5]. Mouse IGL gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. IMGT Colliers de Perles and three-dimensional structures of 33 V-LAMBDA domains encoded by rearranged IGLV-IGLJ genes are available in IMGT/3Dstructure-DB [6].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] In: Molecular Biology of B cells (Honjo., Alt and Neuberger, eds), Academic Press, Elsevier Science, pp. 37-59 (2004). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2004). [6] Nucl. Acids Res. 32, D208-D210 (2004).

P-26**IMGT OVERVIEW:****4. THE MOUSE T CELL RECEPTOR ALPHA TRA GENES**

N Bosc, D Scaviner, G Folch, C Ginestoux, V Giudicelli, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH,
 Montpellier, France

The total number of T cell receptor alpha TRA genes per haploid genome in *Mus musculus* laboratory mice is 159 (including 10 TRAV/DV: 8 rearranged either to TRAJ or to TRDD, 2 assigned by homology) (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse TRA genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The mouse TRA locus is located on chromosome 14 at 19.7 cM. It spans 1650 kb and consists of 98 TRAV genes organized in two clusters upstream of 60 TRAJ genes and of a single TRAC gene [2]. The potential TRA repertoire per haploid genome comprises 110-121 functional genes (including 7-8 TRAV/DV): 71-82 TRAV (including 7-8 TRAV/DV), 38 TRAJ and 1 TRAC. Analysis of the TRAV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-J junctions by IMGT/JunctionAnalysis [5]. Mouse TRA gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. IMGT Colliers de Perles and three-dimensional structures of 17 V-ALPHA domains encoded by rearranged TRAV-TRAJ genes are available in IMGT/3Dstructure-DB [6].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] Dev. Comp. Immunol. 27, 465-497 (2003). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2004). [6] Nucl. Acids Res. 32, D208-D210 (2004).

P-27**IMGT OVERVIEW:****5. THE MOUSE T CELL RECEPTOR BETA TRB GENES**

N Bosc, G Folch, C Ginestoux, V Giudicelli, M-P Lefranc

IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France

The total number of T cell receptor mouse beta TRB genes per haploid genome in *Mus musculus* laboratory mice is 53 (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse TRB genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The mouse TRB locus on chromosome 6 (20.5 cM) spans 700 kb and consists of 35 TRBV belonging to 31 subgroups. Except for TRBV31, localized downstream of the TRBC2 gene, in inverted orientation of transcription, all the other TRBV genes are located upstream of a duplicated “1 TRBD, 7 TRBJ, 1 TRBC” cluster. The potential TRB repertoire per haploid genome consists of 36-37 functional genes: 21-22 TRBV, 2 TRBD, 11 TRBJ and 2 TRBC. Analysis of the TRBV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-D-J junctions by IMGT/JunctionAnalysis [5]. Mouse TRB gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. IMGT Colliers de Perles and three-dimensional structures of 18 V-BETA domains encoded by rearranged TRBV-TRBD-TRBJ genes are available in IMGT/3Dstructure-DB [6].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. D256-D261 (2005). [3] Exp. Clin. Immunogenet. 17, 216-228 (2000). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2004). [6] Nucl. Acids Res. 32, D208-D210 (2004).

P-28**IMGT OVERVIEW:****6. THE MOUSE T CELL RECEPTOR GAMMA TRG GENES**

G Folch, D Scaviner, C Ginestoux, V Giudicelli, M-P Lefranc

IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH, Montpellier, France

The total number of T cell receptor gamma TRG genes per haploid genome in *Mus musculus* laboratory mice is 15 (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse TRG genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The mouse TRG locus is located on chromosome 13 at 10.0 cM. It spans 200 kb and consists of 7 TRGV genes belonging to 5 subgroups organized in four clusters which comprise, for the first one, “4 TRGV, 1 TRGJ, 1 TRGC”, for the others “1 TRGV, 1 TRGJ, 1 TRGC” (the third cluster being in inverted orientation) [3]. The potential TRG repertoire per haploid genome consists of 14 functional genes: 7 TRGV, 4 TRGJ and 3 TRGC. Analysis of the TRGV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-J junctions by IMGT/JunctionAnalysis [5]. Mouse TRG gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. There is no known three-dimensional structures of mouse V-GAMMA but IMGT Colliers de Perles of the germline TRGV and TRGC genes are available in IMGT Repertoire.

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] Res. in Immunol. 141, 692-695 (1990). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2005).

P-29**IMGT OVERVIEW:****7. THE MOUSE T CELL RECEPTOR DELTA TRD GENES**

N Bosc, G Folch, C Ginestoux, V Giudicelli, M-P Lefranc
 IMGT, the international ImMunoGeneTics information system®, LIGM, UM2, CNRS UPR1142, IGH,
 Montpellier, France

The total number of T cell receptor delta TRD genes per haploid genome in *Mus musculus* laboratory mice is 11 (19, with 8 TRAV/DV rearranged either to TRDD or to TRAJ) (IMGT Repertoire, IMGT, the international ImMunoGeneTics information system® <http://imgt.cines.fr> [1]). The mouse TRD genes and alleles and the corresponding IMGT reference sequences are available in IMGT/GENE-DB [2] and were provided to Mouse Genome Informatics MGI in July 2002. The TRD locus is located on chromosome 14, at 19.7 cM and spans 160 kb, nestled in the TRA locus. It comprises: 5 TRDV, 2 TRDD, 2 TRDJ, 1 TRDC, and in 3' of TRDC, 1 TRDV in inverted orientation. The potential TRD repertoire per haploid genome comprises 10 functional genes (17-18 with the TRAV/DV): 5 TRDV (12-13 with the TRAV/DV), 2 TRDD, 2 TRDJ and 1 TRDC. Analysis of the TRDV genes (germline or rearranged) can be performed by IMGT/V-QUEST [4], and analysis of the V-D-J junctions by IMGT/JunctionAnalysis [5]. Mouse TRD gene and allele identification, classification, description, numerotation are based on the IMGT-ONTOLOGY concepts. Codon and amino acid numbering are according to the IMGT unique numbering for V-DOMAIN and C-DOMAIN. There is no known three-dimensional structures of mouse domains but IMGT Colliers de Perles of the germline TRDV and TRDC genes are available in IMGT Repertoire.

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Nucl. Acids Res. 33, D256-D261 (2005). [3] Dev. Comp. Immunol. 27, 465-497 (2003). [4] Nucl. Acids Res. 32, W435-440 (2004). [5] Bioinformatics 20, I379-I385 (2004).

P-30**COMPARATIVE GENOME MAPPING OF THE DEER MOUSE (*PEROMYSCUS MANICULATUS*) WITH REFERENCE TO *MUS MUSCULUS***

C M Ramsdell, E L Thames, J L Weston, M J Dewey
 Peromyscus Genetic Stock Center, Department of Biological Sciences, University of South Carolina, Columbia,
 South Carolina 29208, USA

The ability to make significant insights via genome comparisons is now possible with the creation of genomic maps for several species. To add to this growing and potent source of information, we present the initial results for a comparative genomic map of the deer mouse (*Peromyscus maniculatus*) with reference to *Mus musculus*. This map will also be instrumental in conducting genetic analyses of unique deer mouse phenomena by utilizing the power of the detailed *Mus* map. The map is being constructed with both Type I (protein coding) and Type II (microsatellite) PCR-based markers. These markers were typed on either one or both of a 116 animal interspecific (*P. maniculatus* x *P. polionotus*) backcross panel and a 5000 rad whole-genome radiation hybrid cell panel, both developed by us. The radiation hybrid cell panel consists of 103 cell lines and has an average marker retention frequency of 59.5%. The radiation hybrids have effectively detected linkage in the deer mouse genome between markers as far apart as 6.7 cM and resolved markers that are, in the *Mus* genome, as close as 0.2 Mb. Thus far, combined results from both panels have indicated two linkage groups with a high degree of both linkage conservation and gene order conservation with reference to *Mus* Chromosome 11.

P-31**EFFICIENT GENETIC MAPPING IN THE MOUSE USING A WHOLE GENOME SNP PANEL**

J L Moran¹, A D Bolton¹, A Brown², N Dwyer³, D Nelson¹, P Tran¹, Y Yun¹, A Chesebro¹, B Bjork¹, C Li⁴,
D J Kwiatkowski², R Kucherlapati², T Wiltshire⁵, D R Beier¹

¹ Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States, ² Harvard Partners Center for Genetics and Genomics, Cambridge, MA, United States, ³ Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, ⁴ Dana Farber Cancer Institute, Harvard School of Public Health, Boston, MA, United States, ⁵ Genomics Institute of the Novartis Research Foundation, San Diego, MA, United States

We developed a genome-wide panel of 400 multiplexed SNPs that are genotyped by Sequenom mass spectrometry. SNPs were selected based on chromosomal position and polymorphism primarily between three strains: C57BL/6J, DBA/2J, and 129X1Sv/J. Thirty inbred and wild-derived strains have been genotyped, and there is adequate genome coverage of informative SNPs. Analysis of SNP data is performed with dChip software, which generates a color-coded image of SNP alleles and appropriate LOD score calculations.

Map locations for over 30 loci, including ENU mutations, spontaneous mutations and modifier loci, have been identified. The average number of affected mice genotyped for 19 mapped monogenic mutations was 9, and map locations have been obtained by genotyping as few as 4 affected mice. The average recombinant interval size was 43 Mb. Thus, our SNP panel allows for identification of moderate resolution map position with small numbers of mice in a high-throughput manner.

In addition to mapping monogenic mutations, LOH, modifier and potential QTL loci have been identified. Also, the panel is suitable for marker-selected breeding strategies. Our experience has demonstrated that the panel is effective for mapping crosses from many inbred and wild-derived strain combinations. The major utility of this 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 and strain contamination.

This panel was developed with the support of NIH UO1-HD43430 and utilization is available to any investigator.

P-32**BIOLOGICAL FUNCTIONS UNDERLYING THE LINKAGE DISEQUILIBRIUM DOMAINS OF INBRED MICE**

K Paigen, B L King, P Petkov, J H Graber, G Churchill, N Herz
The Jackson Laboratory, Bar Harbor, ME, United States

We have previously reported that among genetically diverse inbred strains of mice, domains of linkage disequilibrium, averaging several megabases in size, comprise one-fourth to one-third of the genome, along with genetic evidence indicating that these domains reflect selection for co-adapted sets of alleles at closely linked genes that share biological functions. We have now been able to extend these results by identifying some of these functions.

The ten domains with strongest LD were tested using the Ingenuity pathways database; four domains contained a total of 13 biological pathways, each with eight or more tightly clustered genes (135 of 435 genes total). In two connected pathways concerned with cell to cell signalling and immune functions, 10 of 35 and 21 of 35 genes (31 of 69 total) are located within a seven megabase domain (167-174 Mb) on mouse Chr 1. Remarkably, the organization of the network correlates with the arrangement of genes on the chromosome; interacting proteins are often chromosomal neighbors. Gene duplications account for only a small fraction of this result.

We have also reported strong LD between domains on separate chromosomes, creating scale-free networks. Using the gene ontology (GO) database we have been able to identify a number of biological functions and processes underlying these networks; one of which, eye morphogenesis, was predicted in advance.

The finding that functionally related genes are often tightly linked genetically has implications for understanding both the evolutionary origins of mammalian chromosomes and for experimental studies identifying the genes underlying complex traits.

P-33

CLONED SINGLE-MOLECULE ARRAYS FOR GENOME-WIDE RESEQUENCING

T Smith

Solexa Inc., Little Chesterford, Cambridge, United Kingdom

Solexa has been developing cloned single-molecule array technology for whole-genome sequencing. This technology promises minimal sample preparation, higher throughput and greatly reduced costs (based on much denser features) than current methods. To enable these methods, we have developed a very robust 4-colour DNA sequencing-by-synthesis technology that employs reversible terminators with removable fluorescence. This sequencing biochemistry has now been shown to support in excess of 25 cycles with high fidelity, both in solution and on surfaces. This read length represents a key milestone towards accurate mammalian genomic re-sequencing in that 25-mers are long enough to routinely align back to the human reference sequence. Allele calls are then made from the aligned sequences. A recent Solexa *in-silico* study has shown that 82% of the human reference genome can be re-sequenced with 25-mer reads.

These cloned arrays or clusters are formed using a surface-amplification method. We have made significant progress in the preparation of such clusters from randomly cut genomic DNA samples and in applying developments in surface chemistry to improving the methodology. Solexa's surface provides both improved stability and decreased non-specific binding of nucleotides. The clusters have been sequenced on-chip using our sequencing-by-synthesis technology and a novel sequencing instrument incorporating highly sensitive fluorescence detection and automated fluidics. Our current capabilities in this area will be described, highlighting the number of features analyzed, the attainable read length and the measured sequencing accuracy.

Progress in sequencing small genomes using this cluster technology will be shown.

P-34

FORWARD AND REVERSE MUTAGENESIS SCREENS OF A SUBSET OF MOUSE GENES AND REGULATORY REGIONS

J S Weber¹, G Boschult¹, G Schnell², J Kovar³

¹ University of Nebraska, Lincoln, NE, United States, ² University of North Carolina, Chapel Hill, NC, United States,

³ LI-COR Biosciences, Lincoln, NE, United States

Although multiple mammalian genomes have been sequenced, the functions of the majority of genes and conserved non-coding sequences remain to be determined. Various genetics-based *in vivo* methods are being used to identify the functions of genes and regulatory regions. We are using ENU mutagenesis to generate point mutations in the mouse genome. We have taken phenotype and genotype-based approaches to screen for functionally important mutations. Phenotype-based mutagenesis screens have resulted in the isolation and characterization of multiple mutants affecting development and maintenance of different organ systems. A few of these mutants will be described, including details of the mutation responsible for the abnormal phenotype. In the genotype-based mutagenesis screen TILLING (Targeting Induced Local Lesions In Genomes) is being used to detect point mutations in genes and conserved non-coding sequences that may be important in immunity and other systems. A detailed description and update of the screen will be presented.

P-35**ARRAY CGH ON MOUSE USING TOTAL GENOMIC DNA WITH OLIGONUCLEOTIDE ARRAYS**

N Sampas, E Lin, A Scheffer, P Tsang, L Bruhn, K Shannon
Agilent Technologies, Palo Alto, United States

Array-based comparative genomic hybridization (aCGH) is an important tool for studying genomic alterations associated with cancer and developmental disorders in mouse model systems. Previously, we have demonstrated that arrays containing *in situ* synthesized 60-mer oligonucleotide probes can detect genomic lesions including single copy, as well as copy number gains using whole genomic DNA samples¹. Recently, we have designed a 60mer oligonucleotide array containing 42,404 unique probe sequences specific to the mouse genome. Over 70% of these probes are located within the boundaries of known genes, while the remaining probes are biased toward expressed genomic sequences. To validate the performance of this platform, we used several model systems, including a previously characterized cell line that is nominally diploid but contains a known partial amplification on chromosome 16, which yielded probe by probe error rates of less than 12% in the separation of their log-ratio distributions in the triploid region past the breakpoint, where the ratio is nominally 3:2. The mean slope of experimental versus theoretical log-ratios for chromosome X probes on this genome-wide mouse CGH array in XY versus XX hybridizations typically exceeds 0.8, with probe by probe error rates of less than 1% in the separation of their log-ratio distributions where the copy number ratios are 2:1.

¹Barrett, MT, *et al.*, *PNAS* **101**, 17765-17770 (2004).

P-36**FUNCTIONAL GENOMICS OF ALL MOUSE PROTEIN KINASES AND PHOSPHATASES**

A R R Forrest¹, D F Taylor¹, M L Crowe¹, N J Waddell^{1,2}, G Kolle¹, R Kodzius^{5,6}, S Katayama^{5,6}, J L Fink¹, M M Gongora¹, C Flegg¹, H Suzuki³, M Kanamori³, C Kai³, J Kawai³, P Carninci³, Y Hayashizaki³, S M Grimmond¹.

¹ Institute for Molecular Bioscience and ARC Centre in Bioinformatics, Queensland, Australia, ²Queensland Institute for Medical Research, Queensland, Australia, ³Genome Exploration Research Group (Genome Network Project Core Group), ⁴RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Kanagawa, Japan, ⁵ Genome Science Laboratory, Discovery Research Institute, ⁶ RIKEN Wako Institute, Saitama, Japan

Phosphorylation state of a protein regulates their bioactivity, stability, and the ability to interact with other partners. This state is controlled by the opposing interplay of protein kinases and phosphatases which are under tight spatial and context specific control. In mammals, cascades of these phosphoregulators mediate signalling within a cell or between cells, and play crucial roles in differentiation and development. In an effort to define the phosphoregulator network, we have used computational methods to identify all phosphoregulators in mouse. Representative proteins for each component were screened for domain and subcellular localisation motifs. Furthermore, tissue gene expression data, molecular interaction between the p-regs and their catalytic targets and links to the literature have been mined from other sources and integrated into a relational database (www.phoshoreg.imb.uq.edu.au). Due to a paucity of published subcellular localisations for many p-regs, localization of 160 novel members were determined experimentally (IF) and incorporated into the DB. Taken together these analyses provide the foundation for modelling the activity and interplay of these important signalling pathways at a tissue, subcellular and molecular level.

Finally, as part of the FANTOM3 project, the transcriptional complexity of the phosphoregulators in the mouse was also reviewed. These data provided transcriptional evidence for a large number of novel gene products; many with potential dominant negative or dis-regulating functions. The relevance of novel variants has been reviewed by a combination of RT-PCR, RACE and iso-form specific microarray expression profiling.

P-37

MOUSE GENOME RESOURCES AT NCBI

D M Church, M Landrum and the NCBI Genome Annotation Team
DHHS/NIH/NLM/NCBI; Bethesda, MD

Genomic resources for the mouse genome have increased greatly. A Whole Genome Shotgun (WGS) assembly generated by the Mouse Genome Sequencing Consortium (MGSC) was released and published. During the past year, sequencing resources for mouse have shifted towards clone-based (HTGS) sequence. As of July 29, 2005, 2.28 Gb of non-redundant finished sequence and 1.2 Gb of redundant draft sequence were available. Greater than 95% of the HTGS sequence is from the reference strain (C57BL/6J). To leverage all available sequence data, NCBI has been performing composite assemblies that integrate HTGS sequence from C57BL/6J into the MGSCv3. NCBI Build 33 (based on data from May 30, 2004) integrated 1.1 Gb of HTGS phase 3 (finished) sequence.

In addition to producing the reference assembly, NCBI has been producing alternate, strain-specific assemblies. In Mouse Build 35, 6 alternate assemblies were produced. In addition, the Celera mouse assembly (Mmu16) is annotated and available as part of the standard NCBI resource set: Map Viewer, the NCBI ftp site, Entrez Gene, etc.

In addition to producing these assemblies, NCBI provides annotation for all assemblies via a suite of software tools available from our website (<http://www.ncbi.nlm.nih.gov>). Current annotation provides prediction of gene models (based on alignment and *ab initio* prediction), clone placement (BACs and fosmids based on end sequence alignment), variation, STSs, Gene Trap clones, MICER clones, human and rat transcripts and phenotypes (via STS connections). Annotation information will be provided for the current mouse assemblies. In addition, improvements in the NCBI MapViewer and associated resources with respect to clone identification and comparative mapping will be discussed.

See Page 22 for P-38 and P-39