

P1**A TISSUE-SPECIFIC GENE EXPRESSION ATLAS OF UNIPARENTAL DUPLICATION AND METHYLATION MUTANT MICE**

Reiner Schulz¹, Kathryn Woodfine¹, Trevelyan R. Menheniott¹, Deborah Bourc'his², Timothy Bestor³ and Rebecca J. Oakey¹

¹King's College London School of Medicine at Guy's, King's College & St Thomas' Hospitals, Department of Medical & Molecular Genetics; ²U741 Inserm/Paris 7, Institut Jacques Monod, Paris; ³College of Physicians & Surgeons of Columbia University

Genomic imprinting, the different epigenetic programming of the parental genomes during gametogenesis, leads to significant differential expression between the maternally and paternally inherited alleles of certain genes. Many of these imprinted genes are associated with human genetic diseases. Uniparental duplication (UpDp) refers to the inheritance of the same region on both chromosomes in a homologous pair from either only the mother (matUpDp) or the father (patUpDp). Imprinted genes in this region are highly differentially expressed between matUpDp and patUpDp samples. *Dnmt3l* is required to establish the DNA methylation marks in the oocyte that in the offspring, regulate the parent-of-origin-specific expression of most imprinted genes. Consequently, expression of imprinted genes controlled by maternal methylation differs between maternal imprint-free offspring of *Dnmt3l*^{-/-} dams (*Dnmt3l*^{+/+}) and wildtype mice. We have interrogated gene expression in multiple tissues of UpDp and in *Dnmt3l*^{+/+} mice using microarrays. These experiments have led to the discovery of several novel and often tissue-specifically imprinted genes. We have collated the data into a web-accessible expression atlas (<https://atlas.genetics.kcl.ac.uk/atlas.php>) that specifically addresses the tissue-specificity of imprinted expression and its dependence on maternal germline methylation. The atlas comprises three integrated components: a literature-based tissue- and developmental stage-specific catalog of known imprinted genes, a genome browser displaying the microarray data in genomic context, and a self-organizing map representation of the data that aids their similarity-based exploration.

P2**RETROTRANSPOSITION: A MECHANISM FOR GENERATING NEW IMPRINTED GENES?**

Ruth McCole, Andrew J. Wood, Reiner Schulz and Rebecca J. Oakey
Kings College London

Imprinted genes undergo epigenetic modifications during gametogenesis, which lead to transcriptional silencing of either the maternally or the paternally derived allele in the subsequent generation. In the mouse, a number of imprinted genes have been described that originated by retrotransposition. *Nap115*, *U2af1-rs1*, and *Inpp5f_v2* are all imprinted retrogenes that share three additional genetic properties: their promoters overlap CpG islands, are located within introns of host transcripts and are derived from parental genes on the X chromosome. Using these sequence features alone, we identified *Mcts2*, a novel imprinted retrogene on mouse Chromosome 2. A bioinformatic screen has identified a number of candidate genes whose imprinting status is currently under investigation. Retrogenes lacking just one of the sequence features described above are also being analysed for evidence of imprinting in order to identify which features are necessary for imprinting establishment. We can then interrogate the evolutionary pressures acting upon these retrogenes which have caused them to become imprinted. This has wider importance for the evaluation of both mechanistic and selection pressure-based theories of how imprinting could have evolved.

P3**PROGRESS TOWARDS INTEGRATION OF MOUSE PHENOME RESOURCES**

John Hancock, Mouse Phenotype Database Integration Consortium
MRC Harwell, Harwell, UK

The integration of bioinformatics resources holding information on mouse phenotype data of different types will be a critical step in building an infrastructure that will allow the identification of relationships between genotype and phenotype. The mouse bioinformatics community has identified three main areas that require development to achieve this goal: a standard framework for the description of phenotyping procedures, a standard ontological structure for the description of phenotypes, and a format for the interchange of phenotype information between databases and applications [The Mouse Phenotype Database Integration Consortium (2007) Integration of mouse phenome data resources. *Mamm. Genome* 18: 157-163]. We describe progress towards these goals and interactions between the Consortium and other related projects.

P4**COMPARATIVE ANALYSIS OF THE POLYLEUCINE STRUCTURE IN THE MAMMALIAN TASTE RECEPTOR, T1R1 GENE.**

Mitsuru Ebihara
Dept. Food Science, Ishikawa Prefectural University

Expansions of trinucleotide repeat cause several diseases, such as Huntington disease by polyglutamine expansion in the *Htt* gene, congenital central hypoventilation syndrome by polyalanine expansion in the *phox2b* gene. These mutations could be associated with protein misfolding and dysfunction, but their roles remain to be determined. Recently, it has been reported that contraction mutants of polyalanine in the *phox2b* gene which has twenty alanine tracts as a normal repeat number, had half function, but expansion mutant showed no activity as a transcription factor. Some of HPAA tracts are shorter in prokaryote than in eukaryotes, indicating that HPAA got longer and more active in evolution. HPAA database showed approximately 1% of total proteins in prokaryotes and eukaryotes had HPAA in their sequences, suggesting that HPAA tracts had important roles in the function of proteins. Leucine repeat number in the taste receptor, *T1R1* gene is seven in human and chimpanzee, six in mouse, five in dog and chicken and two or three in fishes. Because the *T1R1* has function for perception of umami such as glutamate, it was suggested that these size difference may be associated with preference of food and appetite. Here, polyleucine region of the *T1R1* gene from various mammals were isolated and sequenced to determine the repeat number. Carnivores such as *Panthera tigris*, *Paguma larvata* and *Panthera pardus* had five leucine repeats in number, but omnivorous animals had longer, suggesting that diversity of food preference could be associated with repeat number of leucine in the *T1R1* gene.

P5 IF WE COULD TALK TO THE ANIMALS

Mathew W. Wright, Michael J. Lush, Elspeth A. Bruford
HUGO Gene Nomenclature Committee (HGNC)

The HUGO Gene Nomenclature Committee (HGNC) endeavours to avoid confusion by creating a common nomenclature, so that each human gene has a unique symbol that everyone can recognise. We also maintain a close collaboration with the MGNC and RGNC, and aim to assign the equivalent symbol to orthologous human, mouse and rat genes.

The HGNC Comparison of Orthology Predictions search tool, HCOP (<http://www.genenames.org/hcop>), enables users to compare orthologs predicted for a specified human gene, or set of human genes. HCOP shows orthology predictions between human and 7 other genomes (mouse, rat, chimp, dog, chicken, zebrafish, fruitfly), and currently includes data from Ensembl, Evola, HGNC, Homologene, Inparanoid, MGI, PhlGS, PhyOP, Treefam and ZFIN.

When determining orthology between human, mouse and rat, orthologous genes in other mammalian species are often identified. However other mammalian genomes do not currently have gene nomenclature committees, and this means that genes in these species are often represented by anonymous identifiers in the public databases. Where there is a consensus for one-to-one orthology mappings between human and other mammalian genomes then such genes could in theory directly adopt the human nomenclature, and this could be done a semi-automated way. However this automated approach would not work well for species-specific genes and paralogs, which need to be manually curated. Ideally appropriate gene names should be assigned to all genes in all sequenced genomes, providing each species with their own genetic language. This issue remains most pressing for species where there is currently no established gene nomenclature committee.

P6 EVOLUTIONARY CONSERVED CDT2-INTS7 INTERGENIC REGION ACTS AS A BIDIRECTIONAL PROMOTER AND IS ASYMMETRICALLY REGULATED

Kenichi Yoshida
Department of Life Sciences, Meiji University School of Agriculture

Recent genome-wide studies have revealed that the genetic order in mammalian genomes is not completely random, but that genes with comparable and/or coordinated expression tend to be clustered together. Many genes are unexpectedly coupled by shared transcribed regions in antisense orientation by bidirectional promoters. The linkage of two genes by bidirectional promoters has been shown to facilitate control of functionally related genes. On the basis of this, the CDT2, a regulator of DNA replication, and an integrator complex subunit 7 (INTS7), an interactor of the largest subunit of RNA polymerase II, provide a unique example considering that the two genes contribute to a different phenotype, namely, DNA replication and transcriptional regulation, respectively, though both phenotypes are characterized as part of the fundamental process of life. We present evidence that CDT2 and INTS7 may well be tightly linked by a bidirectional promoter in an evolutionary conserved manner. Within a short intergenic region, E2Fs could up-regulate gene expression in the direction of the CDT2 gene. The tissue distribution of mRNA for CDT2 and INTS7 was inconsistent with each other. Moreover, the presence of similarities between mouse and rat tissue mRNAs was abundant, but these patterns were quite different from the results obtained from human tissues. These findings add a unique example and help to understand the mechanistic insights into the regulation of gene expression through an evolutionary conserved intergenic region of the mammalian genome.

P7**COMPARATIVE GENOMIC MAPPING OF THE HUMAN GHRELIN GENE REVEALS NOVELS EXONS GIVING RISE TO ALTERNATIVE SPLICE VARIANTS, INCLUDING NATURAL ANTISENSE TRANSCRIPTS**

Inge Seim, Adrian Herington, Lisa Chopin

Institute of Health and Biomedical Innovation and School of Life Sciences, Faculty of Science, Queensland University of Technology, Brisbane, Australia

Ghrelin is a multifunctional peptide hormone expressed in a range of normal tissues and pathologies, including growth hormone release, appetite regulation, gut motility and cancer cell proliferation. Despite its widespread and important physiological actions, its precise regulatory mechanisms remain ambiguous. Compared to other preprohormones, the genomic structure of ghrelin is thought to be simple, consisting of four coding exons and a short 20 bp first exon (exon 0). In this study, comparative sequence analysis of the human and murine ghrelin locus revealed intergenic evolutionary conserved regions (ECRs) in the 10 kb region upstream of the preproghrelin start codon. RT-PCR, 5' RACE and cap-analysis gene expression (CAGE) tag analysis confirmed that one of these regions corresponds to a novel exon (exon -1) 2.6 kb upstream. Moreover, 5' extensions to exon 0 and exon 1 were identified. The expression of extended exon 0 and exon -1 were examined by RT-PCR, revealing, in particular, a complex exon -1 to exon 4 splice pattern that includes natural antisense transcripts to ghrelin (ghrelinOS). Strand-specific RT-PCR, 5' RACE and CAGE tags provided further confirmation that ghrelinOS is present on the opposite strand to ghrelin. Our preliminary work supports the validity of employing comparative genomics to identify putative far upstream exons and (resulting splice variants) that would otherwise be difficult or prohibitively expensive to detect. Current experiments are directed at further characterising the identified splice variants, in particular antisense transcripts that could play a role in regulating the ghrelin gene, especially in appetite.

P8**EXPLORING THE MOLECULAR BASIS FOR ATHLETIC PERFORMANCE TRAITS IN NORTH SWEDISH TROTTERS**

Gabriella Lindgren¹, Ingrid Jacobsson², Aneta Ringholm², Knut Roed³, Sofia Mikko¹, Leif Andersson²

¹Dept of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Sweden. ²Dept of Medical Biochemistry and Microbiology, Uppsala University, Sweden. ³Dept of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, Norway

The horse constitutes an exciting model for physiological genetic research in a functional genomics context. The origin of the North Swedish trotter provides a unique opportunity to identify genes influencing body constitution and racing performance. The North Swedish trotter originates from the North Swedish horse that is a draught horse used in farming and forestry. It is well known that there has been some crossbreeding between Standardbreds and North Swedish trotters before obligatory paternity testing was introduced in Sweden. We hypothesize that we will be able to detect a gene flow that has occurred between Standardbreds and North Swedish trotters due to this cross-breeding, and a subsequent strong selection for racing performance in the North Swedish trotters. A remarkable improvement in racing performance of the North Swedish trotter has occurred during the last fifty years. This process should leave "genetic footprints" in the genome of North Swedish trotters in the form of chromosome segments originating from Standardbreds. We are using a population genomics approach to identify chromosome regions under positive selection to achieve our goals. Chromosome regions under strong positive selection are likely to harbor genes affecting morphological and physiological traits important for physical performance traits in this system. Presently, we compare the genetic makeup of North Swedish trotters (20 individuals), North Swedish horses (10 individuals) and Standardbreds (10 individuals) by performing a high-density genome scan. To date, 167 microsatellite markers have been analyzed within the horse material. Estimation of overall Fst values between the three horse breeds support that crossbreeding has occurred between the Standardbred and the North Swedish Trotter. Detection of genes that regulate racing speed is of particular interest as this could give a model for further studies of the genetic basis for complex traits such as muscle capacity and oxygen uptake.

P9

MAUSDB: THE GERMAN MOUSE CLINIC OPEN SOURCE PHENOTYPE- AND MOUSE MANAGEMENT SYSTEM

Holger Maier, Christoph Lengger, Ralph Steinkamp, Karlheinz F. Schaeble, Helmut Fuchs, Valerie Gailus-Durner, Martin Hrabe de Angelis

GSF-National Research Center for Environment and Health, Institute of Experimental Genetics, Munich/Neuherberg, Germany

Mutant mouse lines are important tools to elucidate gene function as observed phenotypes can mostly be attributed to a known genotype. The German Mouse Clinic (GMC, <http://www.mouseclinic.de>) as an open-access platform offers standardized and comprehensive phenotype analysis of mutant mouse lines and screens for potential new mouse models of human diseases. In the GMC, mouse cohorts pass through 14 different screening modules in a strictly defined workflow in the course of primary screen where up to 240 physiological parameters per mouse are measured. Screening many mouse cohorts in multi-parallel workflows is a logistical challenge and requires appropriate IT support and well-defined data infrastructure. Therefore, we developed a web-based database application - MausDB - for the GMC that serves as a central data platform accessible by all GMC users. MausDB supports scheduling of mouse lines to the phenotyping pipelines by work list management functions. Phenotyping data upload to MausDB and re-export is done via spreadsheet files. MausDB also offers standard mouse management functions (breeding support, cage management, etc.) and integrates phenotype data with line/genotype data and other metadata on the individual mouse level. For the GMC, this is a prerequisite for inter-line data analysis, data mining and data exchange in a cross-European phenotyping effort (EUMODIC). Although primarily developed for the GMC, MausDB also proved to be useful for other mouse facilities due to its general purpose design and intuitive user interface. Hence, we offer MausDB to the mouse community as open source software under the terms of the GNU General Public License.

P10

COMPARATIVE FUNCTIONAL ANNOTATION: THE MOUSE PERSPECTIVE

Judith A. Blake, David P. Hill, Mary E. Dolan, Harold J. Drabkin, Li Ni, and Dmitry Solnikov

Mouse Genome Informatics Group, The Jackson Laboratory

The ability to use functional annotations for data analysis and to perform comparative biology investigations of genomic data depends on comprehensive and high-quality annotations derived from experimental data published in the biomedical literature. The Functional and Comparative Annotation Team of the Mouse Genome Informatics (MGI) resource provides such annotations for mouse genes. The MGI team participates in the Gene Ontology (GO) Reference Genome Project, a shared annotation effort among the GO Consortium members who provide annotations for the primary model organisms such as mouse, fly and yeast. We are collaborating in efforts in providing comprehensive annotations for the human genes and their orthologs. As the curators are simultaneously working on the same set of genes, they are also updating the ontologies, providing an orthology set for these organisms, and improving documentation of the GO annotation processes. The comprehensive annotations of the well-studied model organisms serve as a basis for the annotation of emerging genomes. For all experimentalists and computational biologists using GO functional annotations in their work, the Reference Genome Project will provide deeper and complete annotations for genes and gene products. I will present strategies, goals and project status from the perspective of the MGI and GO project objectives.

The Gene Ontology project is supported by NHGRI grant HG002273, The MGI project is supported by NIH grants HG00330.

P11**REARRANGEMENT EVENTS HAVE SYNCHRONIZED WITH NUCLEOTIDE SUBSTITUTIONS DURING MAMMALIAN EVOLUTIONARY HISTORY**

Tsuyoshi Hachiya, Yasunori Osana, Kris Popendorf and Yasubumi Sakakibara
Department of Biosciences and Informatics, Keio University

Accurate detection of orthologous segments plays a key role in genome rearrangement analysis, where an orthologous segment is a conserved segment descended from the common ancestor without rearrangements. Genomic distances are defined as the minimum number of rearrangement operations needed to transform one order of orthologous segments to another. Phylogenetic trees based on genomic distance can be reconstructed by neighbor-joining methods or MGR algorithm. Previous programs for finding orthologous segments require some prior parameter values. These parameter values, however, are difficult to set manually, because the appropriate values can vary over different set of genomes to be compared. Here, we propose a new program, OSfinder (Orthologous Segment finder), which can automatically determines its optimal parameter values for each data set. Our evaluation demonstrates that OSfinder achieves the better performance compared with existing programs, such as ADHoRe, DAGChainer and AXTCHAIN. We applied OSfinder to six mammalian genomes (human, chimpanzee, macaque, mouse, rat, and dog) and elucidated the unrooted phylogenetic tree of the six mammals by neighbor-joining methods. The comparison between rearrangement-based phylogenetic tree and nucleotide substitution-based phylogenetic tree showed a remarkably high value of Pearson correlation coefficient (0.98) by taking profile of branch lengths as a feature vector of phylogenetic tree. This result means that the relative frequencies of rearrangements and substitutions in each lineage are almost the same, and indicates that the both events have synchronized with each other at least during mammalian evolutionary history.

P12**MURASAKI: LANGUAGE-THEORY BASED HOMOLOGY DETECTION ACROSS MULTIPLE LARGE-SCALE GENOMES**

Kris Popendorf, Yasunori Osana, Hachiya Tsuyoshi, Yasubumi Sakakibara
Keio University

As the number of whole genome sequences available continues to increase rapidly, the raw scale of the sequence data being used in analysis is the first hurdle for comparative genome analysis. When performing whole genome alignments, large-scale rearrangements make it necessary to first find out roughly which short well-conserved segments correspond to what other segments (termed anchors). Successful results have been achieved by adapting tools like BLAT and BLASTZ a problem-to-problem basis, but the work required to perform a single alignment is considerable. Recently, new programs such as Mauve and PatternHunter can handle slightly larger inputs, but the memory/time requirements for sequences like Human and Chimp X chromosomes are prohibitive for most computational environments. Our novel algorithm, which we have implemented in a program called Murasaki (available at <http://murasaki.dna.bio.keio.ac.jp>), makes it possible to identify anchors of multiple large sequences on the scale of several hundred megabases (eg. three mammal chromosomes) in a matter of minutes. We also apply the traditionally natural-language information retrieval technique of term-frequency inverse document-frequency (TF-IDF) term weighting to word frequencies in DNA to achieve a robust filtering system. We also tested and confirmed the performance of Murasaki against a variety of genomes, including Human, Mouse, Rat, Chimp, Rhesus, and a variety of Mycobacteria, Mycoplasma, and fungi. We compared the performance of TF-IDF to hits and length as other possible anchor filters, and found TF-IDF to exhibit reliably better performance.

P13**DEVELOPMENT OF INFORMATIONAL INFRASTRUCTURES TO SHARE MOUSE PHENOTYPE AND EXPERIMENTAL PROCEDURES OF PHENOTYPING PLATFORM WITH MULTIPLE INSTITUTIONS**

Hiroshi Masuya¹, Sumi Yoshikawa¹, Naohiko Heid¹, Norio Kobayashi¹, Takeshi Arikuma³, Tetsuro Toyoda¹., Riichiro Mizoguchi², Tetsuo Noda¹, Yoichi Gondo¹, Shigeharu Wakana¹ and Toshihiko Shiroishi¹

¹RIKEN GSC, ²Department of Knowledge Systems, ISIR, Osaka University, ³Tokyo Institute of Technology

The life science is now on the way of transition from functional analysis of individual genes to understanding a whole life system by referring “complete set” of “Omic data” In the field of mouse functional genomics, a number of large-scale knock out mouse programs have launched, and will lead to exponential increase of mouse mutant resources in this few years. The “phenotype” is core information representing properties of these mutant resources. In the international efforts to develop infrastructure for objective (or systematic) evaluation of phenotype information of these mutant resources, it is desirable to integrate highly reliable information obtained from different phenotypic platforms. Since definition of a specific phenotype often depends on tacit understanding of concept that tends to vary among different facilities, it is necessary to define phenotypes based on the explicit evidence of assay results. We have developed a website termed PhenoSITE (Phenome Semantics Information with Terminology of Experiments: <http://www.gsc.riken.jp/Mouse/>), in which we are trying to integrate phenotype-related information using an experimental-evidence-based approach. In addition, we are developing a format to compare parameters of experimental procedures used among different institutions and projects, and an ontology for describing the structure of experimental genetics.

P14**COMMON AND DIFFERENTIAL RESPONSES TO STEROID HORMONES VIA THEIR COGNATE RECEPTOR-BINDING SITES IN HUMAN GENOME**

Kuniko Horie-Inoue¹, Ken-ichi Takayama^{1,2}, Satoshi Inoue^{1,2}

¹Res. Ctr. Genomic Med., Saitama Med. Univ., Hidaka-shi, Saitama, Japan, 350-1241; ²Dept. Geriatric Med., Grad. Sch. Med., Univ. of Tokyo, Bunkyo-ku, Tokyo, Japan, 113-8655

The activity of NR3C subfamily steroid hormone receptors, including androgen receptor (AR), glucocorticoid receptor (GR), mineralocorticoid receptor (MR), and progesterone receptor (PR), is mainly mediated by the identical palindromic DNA sequences termed steroid hormone receptor elements (SHREs). Nevertheless, each steroid hormone receptor modulates the transcription of distinct sets of target genes. The question remains how these receptors recognize and distinguish SHREs in a common or specific manner. We recently identified androgen receptor binding sites (ARBSs) and their neighboring androgen response genes in prostate cancer LNCaP cells, by the combination of chromatin immunoprecipitation (ChIP) and tiling array (ChIP-chip). Here we investigate whether these ARBSs can recruit other hormone receptors ligand-dependently and their adjacent genes can respond to other steroid hormones. In terms of the ten bona-fide ARBSs for the ENCODE regions, we perform ChIP assays for GR/MR using 293 cells stably expressing either receptor or empty vector treated with vehicle or ligands (10 nM dexamethazone for GR and 100 nM aldosterone for MR, respectively). Several ARBSs exhibit ligand-dependent GR/MR recruitment in the receptor-transfected cells (>2-fold over vehicle treatment by 1h). The ARBSs in the PGC proximal promoter and CDH2 intron 1 recruit both GR and MR significantly upon ligand treatment. Regarding the expression of genes close to the ARBSs, UGT1A1 and PGC can be induced by dexamethazone or aldosterone (>2-fold over vehicle by 24 h). The present data suggest that the NR3C steroid hormone receptors may have potentials to share several common DNA binding sites in human genome.

P15**ANALYSES OF NON-CODING RNAs IN BOVINE GENOME**

Evgeny Glazov, Sean McWilliam, Wes Barris, David Tang, and Brian Dalrymple
CSIRO Livestock Industries

The bovine genome (*Bos taurus*) is the first fully-sequenced genome of a ruminant mammal. It also represents a group of relatively large organisms with a rapid postnatal development. These and other morpho-physiological differences of this mammal make bovine genome an attractive model for comparative genomics studies. As part of bovine genome annotation consortium we focused on identification and analyzes of non-coding structural and regulatory RNAs in the bovine genome. This study overviews common trends in evolution of the main classes of non-coding RNAs (tRNAs, snoRNAs, miRNAs, mRNA-like ncRNAs) in mammalian genomes. Using a combination of comparative genomics approaches and EST data from cow, sheep and other mammalian species, we identified a subset of several hundreds on putative ruminant-specific mRNA-like non-coding RNAs and analyzed their expression pattern. Comparative genomics analyzes of other classes on non-coding RNAs also identified several examples of the bovine-specific expansion and losses in gene copy numbers in several groups of ncRNAs.

Specifics roles of these ncRNAs in the adaptive evolution of ruminants are discussed.

P16**EXPRESSION OF HUMAN CIS NATURAL ANTISENSE TRANSCRIPTS AFFECTED BY THEIR OVERLAPPING ARRANGEMENTS IN THE GENOME**

Naoki Osato, Yoshiyuki Suzuki, Kazuho Ikeo, Takashi Gojobori
DNA Data Analysis Laboratory, Center for Information Biology and DNA Data Bank of Japan

For a significant fraction of mRNAs, their expression is regulated by other RNAs, including *cis* natural antisense transcripts (*cis*-NATs) that are complementary mRNAs transcribed from opposite strands of DNA at the same genomic locus. The regulatory mechanism of mRNA expression by *cis*-NATs is unknown, although several possible explanations have been proposed. To find *cis*-NATs that affect their expression, we conducted a large-scale analysis of the currently, available data and examined how the overlapping arrangements of *cis*-NATs affect their expression level. As a result, we found that (1) the expression level of *cis*-NATs decreases as the length of the overlapping region increases. (2) The distribution of the expression level of *cis*-NATs changes according to different types of the overlapping pattern of *cis*-NATs in the genome. (3) A positive correlation of the expression level between sense and antisense mRNAs was observed much more in *cis*-NATs than randomly selected pairs of transcripts. Moreover, the degree of a positive correlation varies with different types of the overlapping pattern of *cis*-NATs in the genome. (4) In spite of the expectation from a possible regulatory mechanism of gene expression by *cis*-NATs, a negative correlation was not significantly observed. From the results of (1), (2) and (3), it suggests that overlapping arrangements of *cis*-NATs significantly affect their expression.

P17

MOLECULAR EVOLUTION OF GP25L GENE FAMILY

Shoichi Matsukuma, Mituyo Yoshihara

Kanagawa Cancer Center Research Institute, Division of molecular pathology and genetics

gp25L, an ER-Golgi protein gene, has been considered to have a role in protein transportation and one of the family gene, Tmp24 (p23, Tmed10), was shown to be indispensable in embryonic development in mouse. We present that gp25L showed pancreas specific expression in mouse and human. In human, there is a nonsense mutation in exon 3 but a splice variant isoform is formed by exon 3 skipping. This nonsense mutation and splice variant of exon 3 was not present in chimpanzee but, in orangutan, the same splice variant was used as well as the entire form without nonsense mutation in exon 3. Recent advance of mammalian genome analysis allowed us to examine gp25L in various mammals, marsupials and chick and to compare with the related family genes, namely, gp25L2 (Tmed9), p24alpha (Tmed4) and Tmp24. Evolutionary relations among these family members are: gp25L diverged from Tmed4 and Tmed9 before the separation of the latter two and these three are derived from an ancestral gene common to Tmed10. Each of Tmed4, Tmed9 and Tmed10 showed lower variation in mammalian species than gp25L did in mammals. The former showed homologs in marsupials, platypus, fish, fly and nematoda, whereas the latter showed homologs in marsupial and chick but not in frog, fish, fly or nematoda. From expression of gp25L specific to pancreas, evolutionary development of pancreas in vertebrates could be related to evolutionary divergence of gp25L as an independent gene, although the mouse deficient in gp25L showed no obvious morphological defect and survived normally.

P18

SEQUENCE COMPARISON OF FSH-RECEPTORS AMONG 12 MOUSE STRAINS

Osamu Suzuki

National Institute of Biomedical Innovation

Superovulation induction by gonadotropin treatment is widely used in reproductive experiments. However, differential response among various mouse strains to gonadotropins is still a serious problem for practical applications (Suzuki et al., *Reprod. Fertil. Dev.*, 8:975-80, 1996). To elucidate the mechanism underlying the strain difference, we compared FSH-receptor sequences among strains to clarify whether the structural difference in FSH-receptors is involved in the strain difference in superovulation efficiency. Amino acid sequences of FSH-Receptors were deduced from cDNA sequences determined using RT-PCR with testis RNA from 12 strains of mice (129x1/SvJ, A/J, AKR/N, BALB/cCr, C3H/HeJ, C57BL/6J, CBA/N, DBA/1J, DBA/2Cr, NZB/N, NZW/N, and SJL/J) in combination with direct sequencing of the PCR products. Four sets of primers were designed to cover the entire coding sequence of the FSH-receptor on the basis of C57BL/6J sequence (accession number: NM-013523). The deduced FSH-receptor amino acid sequences in all strains examined were essentially the same as that of C57BL/6J except that the 436th amino acid was lysine in C57BL/6J and glutamine in the other strains. Thus, it is unlikely that the strain difference in the response to gonadotropin treatment is due to structural difference in the FSH-receptor. We should focus other mechanisms, such as the strain difference in level, location, and timing of FSH-receptor expression, for the strain difference in superovulation induction. This work was supported by a grant from the Ministry of Health, Labor, and Welfare of Japan.

P19**MOUSE GENE NOMENCLATURE: WHERE ARE WE IN GENE NAMING?**

Lois J. Maltais, Monica McAndrews-Hill, Judith A. Blake, Carol J. Bult, Janan T. Eppig, Mouse Genome Informatics (MGI) Staff
The Jackson Laboratory

The Mouse Genome Database (MGD), www.informatics.jax.org, is the authoritative source of official names for mouse genes, alleles, and strains. Nomenclature follows the rules and guidelines established by the International Committee on Standardized Genetic Nomenclature for Mice and is implemented through the Mouse Genomic Nomenclature Committee (MGNC). MGNC collaborates extensively with the HUGO Gene Nomenclature Committee (HGNC) and the Rat Genome Database (RGD) in establishing standardized nomenclature for orthologous genes for mouse, human, and rat. In addition all three groups offer advice on nomenclature issues to other species, such as zebrafish, dog, and chicken, and provide official nomenclature to various databases, including Entrez gene, Ensembl, Vega, UCSC Browser, Uniprot, Treefam CYP450, Rfam, ADAMTS, and Wnt. Some of the current challenges facing the nomenclature committees revolve around sequence-based resolution of genes and gene families leading to genes being merged, split, or added. Often times interactions with sequence data and genome annotation providers (e.g., NCBI, Ensembl, Vega) and gene family experts are needed to resolve questionable issues. Paralogous genes pose particular dilemmas when determining orthology across species. Build 37 in mouse is helping to sort out some of the issues of orthology and paralogy, which in turn will lead to official symbols and names for the newly identified genes. Here examples demonstrating resolutions to common nomenclature issues are discussed. This work is supported by HG000330

P20**APPLICATION OF SNP GENOTYPING ASSAY FOR TYPING OF HUMAN MITOCHONDRIAL DNA CONTROL REGION IN TAIWAN**

Shiao-ping Huang, and Jing-Jing Lai
Department of Medical Technology, Fooyin University

Analysis of human Mitochondrial DNA (mtDNA) sequence profiles using comparisons to standard sequence, either Cambridge Reference Sequence (CRS) or revised Cambridge Reference Sequence (rCRS), provides the valuable information for characterizing the variation among human populations. Three hypervariable regions (HVR) including HVR-I, HVR-II, and HVR-III appear to be polymorphic in the control regions of mtDNA. To characterize the hypervariable regions of mtDNA for developing a rapid typing method, the SNP genotyping assays were designed with 10 chosen highly polymorphic sites in Taiwan. The designed primers and probes were synthesized and then subjected to perform the SNP genotyping assay. The polymorphic positions with high frequency of polymorphism and applied in this study were T16362C, C16223T, G16129A, T16172C, T16304C, and G16319A in HVR-I, C150T, T152C, and T199C in HVR-II and T489C in HVR-III. The blood samples were collected from 250 random and unrelated individuals. After DNA extraction, the SNP genotyping assays were performed to analyze the genotype of each polymorphic site in HVR of mtDNA. By using the 10 SNP assays, the 250 samples were divided into various haplotypes successfully. Furthermore, the power of discrimination (D) for SNP genotyping assay was estimated and compared with PCR-DNA sequencing method. Therefore, the results allow us to apply the SNP genotyping assay for high throughput and rapid typing of mtDNA in Taiwan.

P21

FROM PHENOTYPE TO GENE USING THE MOUSE GENOME INFORMATICS DATABASE

Carol Bult, Judith Blake, Janan Eppig, James Kadin, Joel Richardson, Martin Ringwald, and the Mouse Genome Informatics Group
The Jackson Laboratory

The nearly complete and annotated genomes for human and mouse provide a powerful framework for *in silico* methods of identifying candidate genes for genetically mapped disease phenotypes. Using publicly available genome browsers a researcher can, in a matter of minutes, generate a list of all known and predicted genes annotated to specific regions of a genome. A list of genes, however, is not sufficient to effectively identify potential candidate genes. The Mouse Genome Informatics (MGI) contains a comprehensive catalog of mouse QTL and genetically mapped mutants. The MGI database facilitates the identification of attractive candidate genes for genetically-defined phenotypes by providing access to well-curated, integrated biological knowledge about mouse genes including gene expression, sequence variation (e.g., SNPs), gene function, phenotype associations, and associations with orthologous human disease genes. As knowledge of genome annotations and gene function change over time, using MGI to re-evaluate previously published QTL loci can often result in new candidate genes or reinforce the quality of previously identified candidate genes. We will illustrate the use of MGI to evaluate candidate genes for previously identified QTL for lung cancer susceptibility (pulmonary adenoma susceptibility, Pas) in mice.

Supported by NIH HG00330.

P22

EVOLUTION OF THE SCHLAFEN GENES, A GENE FAMILY POSSIBLY INVOLVED IN EMBRYONIC LETHALITY, MEIOTIC DRIVE, IMMUNE SYSTEM DEVELOPMENT AND POXVIRUS VIRULENCE

Saijal Naik, Olivia Bustos, Gayle Ayers, Aaly Meheraly, Ellen Pritham and Elena de la Casa-Esperon
Department of Biology. University of Texas at Arlington. USA

To date, little is known about the Schlafen genes in spite of the interesting phenotypes they appear to be associated with. Initial studies found they are expressed in several lymphoid tissues and seem to be involved in T cell development. In mouse, they cluster in the Om region in mouse chromosome 11. Interestingly, we also mapped in this region two interesting phenotypes: embryonic lethality and unequal segregation of chromatids during meiosis (meiotic drive). In order to better understand their role in these diverse traits, we are doing a comparative sequence analysis of this family of genes. We have found that the Schlafen genes have increased in copy number since the divergence of mammals and multiple duplications are present in humans and other species. Moreover, they have been horizontally transferred to orthopoxviruses, where they may play a role in adapting the virus to survive host immune defense mechanisms. Our studies show that some of the Schlafen members are rapidly evolving and, thus, are acquiring novel functions.

P23

GENETIC CHARACTERIZATION OF THE WILD-DERIVED HOUSE MICE IN EAST ASIA

Jun J. Sato¹, Yasunori Yamaguchi¹, Junpei Ueta¹, Hitoshi Suzuki², Wang Chunyan³, Alexei P. Kryukov⁴, Kazuyuki Mekada⁵, Naoyuki Takahata⁶, Kazuo Moriwaki⁵

¹Fukuyama University, ²Hokkaido University, ³Lanzhou Institute of Biological Products, ⁴Russian Academy of Sciences, ⁵RIKEN BRC, ⁶Graduate School of Advanced Studies

Our previous studies showed that four Hemoglobin beta chain (*Hbb*) haplotypes, *d*, *p*, *w1*, and *w2* in the wild-derived house mouse, *Mus musculus*, have been established with the complex history among different subspecies groups, *castaneus* and *musculus*. In the intergenic spacer region between *Hbb* b1 and b2 genes, recombination points were specified among these haplotypes, suggesting the existence of multiple recombination-inducing sequence factors. This implies that the *Hbb* locus might have possibility to reflect the past hybridization among different types of mice along with the detection of the recombination, and therefore possess useful property to characterize the genetic origins of the laboratory mice. In this study, the most polymorphic second intron sequences of both b1 and b2 genes were examined for 56 mice individuals collected from China, Japan, and Russia to trace the origin of the widely used laboratory strains and, in particular, to evaluate the origin of the Japanese mouse strain which was recently suggested to genetically contribute to the common laboratory strains. The Neighbor-joining analyses showed different patterns of similarity between b1 and b2 gene trees, in which the mice collected from Japan represented close affinity to the *p* and *d* haplotypes in the b1 gene tree and to the *p* and *w1* haplotypes in the b2 gene tree. This suggests that the Japanese mice possess the recombinant *p* haplotype and it is thus implied that the genome components of the widely used laboratory strains have also been associated with the complicated genealogies of the East Asian mice.

P24

A MATHEMATICAL MODEL FOR AFFYMETRIX GENECHIP PROBE LEVEL DATA BASED ON FUNCTIONAL STATES OF GENE - ON/OFF

Keiko Otani¹, Megu Ohtaki², Keiko Hiyama², Eiso Hiyama², Kenichi Satoh², Takeshi Shimamoto², Dokki Mohamad², Masataka Andoh², Tetsuji Tonda², Masahiko Nishiyama²

¹Hiroshima Cancer Therapy Development Organization, ²Hiroshima University

Motivation: The Affymetrix GeneChip (PM, MM) probe pair is designed with the intention of measuring non-specific binding. The removal of non-specific binding or cross-hybridization signal from a PM measurement is a key issue in the analysis of microarray data. Many researchers have pointed out that direct subtraction of MM from PM is unlikely to be useful, because MM contains mostly target-specific signal as well as PM. As for methods for summarizing probe intensities in a single index, numerous model-based algorithms have been proposed, however the most algorithms are too complicated with many parameters having poor biological implication. Biologists who want to analyze GeneChip microarray data might be bewildered with the availability of so many procedures with varying results.

Results: We provided a new model-based algorithm for summarizing probe level data using the probability of a gene being ON or OFF, in which both PM and MM values are assumed to have a same distribution if a gene is in the OFF state. This means that the probability that MM > PM is equal to that of MM < PM for OFF genes. Then the probability of a gene being ON or OFF is able to be estimated using the number of times MM > PM per probe set. This is useful to discriminate specific bindings from non-specific bindings. The performance of a new summarization method was examined using U95A dataset of spike-in controls.

P25

A COMPARATIVE STUDY OF THE ABCG2 GENE IN MILK PRODUCTION

Jerry Wei, Julie A. Cavanagh, Mehar S. Khatkar, Lisa G. Riley, Paul A. Sheehy, Herman W. Raadsma and Peter Williamson

Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, and CRC for Innovative Dairy Products, University of Sydney, Australia.

Comparative studies provide a powerful approach to analysis of complex traits. In dairy cattle, quantitative trait locus (QTL) analyses for milk production traits have identified a region on BTA6. Recently a nucleotide variant (SNP) in the bovine ABCG2 gene was associated with an effect on milk protein, and milk fat. As a member of the ABC family, ABCG2 is a transmembrane pump for effluxing lipophilic compounds. ABCG2 has also been used as a stem/progenitor cell marker for side population (SP) phenotype selection. However, role of ABCG2 in mammogenesis or lactogenesis, and in the control of milk composition, remains unclear.

To investigate ABCG2 as a candidate gene we used a comparative approach based on RNA expression profiles and functional assays in mammary tissue. Expression profiles were examined in mouse models, a marsupial (wallaby) model and from public datasets. These were compared to lactation cycle profiles in dairy cattle. The analysis provided clear evidence for differential expression of ABCG2 during mammary gland development. Functional analyses of ABCG2 in mammary epithelial cells suggested that ABCG2 may affect lactation by influencing the proliferation phase in mammogenesis. To identify the potential functional SNPs in the bovine promoter region, we used a pooled DNA sequencing strategy, and developed genotyping assays for each SNP using high resolution melt analysis (HRM). Validation of the haplotype and allele frequency in dairy cattle, and association with dairy traits are underway.

P26

PREDICTION OF NUCLEAR LOCALIZATION PROTEINS WITH CHARGE PERIODICITY OF 28 RESIDUES

Noriyuki Sakiyama, Runcong Ke and Shigeki Mitaku

Department of Applied Physics, Graduate School of Engineering, Nagoya University

Recently, we calculated the autocorrelation function of the electric charges in amino acid sequences, which showed significant positive correlation in eukaryotic genomes and very sharp charge periodicity of 28 residues in vertebrates. In human genome, 2.4% of the total proteins were found to have charge periodicity of 28 residues. The detailed analysis of proteins with charge periodicity of 28 residues (PCP28) suggested that many of them from human genome are localized to nucleus. In this study, we developed a novel method to discriminate PCP28 localized to the nucleus from other proteins, by using PCP28 extracted from the SWISS-PROT database release 48.7. We analyzed these proteins by two steps. The first step was the calculation of the average values of eight physicochemical properties of whole proteins sequences, by which a score for predicting the nuclear localization proteins was calculated. The second step was the analysis of local sequences with highly positive charge for the same purpose. Finally, the total system was developed, using two types of scores. The accuracy of the discrimination of PCP28 of the nuclear localization was better than 90%. By using this system, we compared genomes based on PCP28 in 248 prokaryotic genomes, 7 invertebrate genomes and 10 vertebrate genomes. The result indicated that nuclear PCP28 form a new branch in vertebrate and number of nuclear PCP28 increased as evolution. On the other hand, the correlation of PCP28 except nuclear proteins from vertebrate fitted the correlation of all PCP28 from prokaryote.

P27**A GENOME-WIDE MAP OF 8.27 MILLION SNPs IN THE LABORATORY MOUSE GENOME**

B. Karlak¹, K. Frazer¹, C. Wade^{2,3}, M. Bogue⁴, D. Hinds¹, E. Beilharz¹, R. Gupta¹, J. Montgomery¹, M. Morenzoni¹, G. Neilsen¹, S. Osborn¹, C. Pethiyagoda¹, L. Stuve¹, F. Johnson⁵, M. Daly^{2,3}, D. Cox¹

¹Perlegen Sciences, Mountain View, CA; ²Broad Institute, Cambridge, MA; ³Center for Human Genetic Research, Boston, MA; ⁴The Jackson Laboratory, Bar Harbor, ME; ⁵Toxicology Operations Branch, NIEHS, Research Triangle Park, NC

A dense map of genetic variation in the laboratory mouse genome will provide insights into the evolutionary history of the species and lead to an improved understanding of the relationship between inter-strain genotypic and phenotypic differences. Using our high-density oligonucleotide array-based technology, we re-sequenced the genomes of four wild-derived and eleven classical inbred strains and identified 8.27 million high quality SNPs. The wild-derived inbred strains are separated from one another by over 3 million discordant SNPs uniformly distributed across the genome, while pairs of classical inbred strains are separated on average by 936,000 discordant SNPs. Analysis of the genetic contributions of the four *Mus* subspecies to the classical strains surprisingly revealed that collectively the contributions of the three Asian subspecies - *molossinus* (27%), *musculus* (14%), and *castaneus* (12%) - is larger than that of *domesticus* (46%). The re-sequencing data, mapped to build 36 of the NCBI Mouse Genome, is available at our web site (<http://mouse.perlegen.com/>), and includes information on SNP location, genotypes, gene assignment, PCR primer composition, trace files and quality information. The website allows the data to be viewed interactively or downloaded in batch format.

P28**cSNP MINING FROM 3,210 FULL-LENGTH ENRICHED cDNA SEQUENCES OF KOREAN NATIVE PIGS**

Vijaya Dirisala, Juhyun Kim, Kwangha Park, Nameun Kim, Younshin Oh, Hojun Choi and Chankyu Park
Department of Animal Biotechnology; College of Animal Bioscience and Technology, Konkuk University, Seoul 143-701, Republic of Korea

Sequences from clones of full-length enriched cDNA libraries serve as valuable resources for functional genomic studies. We analyzed 3,210 chromatograms obtained from the 5'-end sequencing of the clones from the brainstem, liver, neocortex and spleen full-length enriched cDNA libraries of Korean native pigs. In addition, 50,000 pig EST sequence trace files were obtained from GenBank and combined with our sequencing information for SNP identification *in silico*. The process generated 8,118 contigs containing 800 putative cSNPs. Of these 33 putative cSNPs were randomly selected and validated using 20 different pigs of four different breeds (Duroc, Landrace, Yorkshire, Korean native pig). 20 out of 33 cSNPs were confirmed (validation rate, 61%). We also identified 15 new cSNPs from the validation process, which were not detected by our *in silico* analysis. Therefore we obtained results on a total of 35 confirmed cSNPs with allele frequency among 4 breeds. Our study shows that large-scale EST sequencing from the Korean native pig can be effectively employed for natural polymorphism-based pig genome analysis.

P29

IDENTIFICATION OF EVOLUTIONALLY CONSERVED FAT REGULATORY GENES USING COMPARATIVE GENOMIC APPROACH.

Ken Yagi, Itoshi Nikaido, Yosuke Mizuno, Yasushi Okazaki

Division of Functional Genomics and Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University

The excess fat storage in visceral adipose tissue is one of the major causes of metabolic syndrome that is being recognized as a critical problem of public health in human.

The genes that control energy homeostasis and fat metabolisms might be conserved from lower organisms to mammals. In *C. elegans*, genome-wide RNAi screening analysis was carried out to discover fat regulatory genes in the body. To discover novel genes that are involved in fat storage in adipocytes in mammals, we searched for the mouse orthologues of *C.elegans* fat regulatory genes. To validate if these genes affect on fat storage in mammals, we performed knock-down assays with siRNAs using a preadipocyte cell line DFAT-D1. Molecular analysis revealed that about half of the genes showed consistent effects in terms of triglyceride accumulation, but others showed reverse or no effects compared with those observed in *C. elegans*.

In conclusion, we searched for mouse orthologues of *C. elegans* genes that affected fat storage. We tested the function of these genes and found nine genes stimulated adipogenesis whereas one gene suppressed adipogenesis.

P30

THE STATE OF ANNOTATION IN HAVANA

Laurens Wilming, J Almeida, C Amid, I Barnes, D Carvalho Silva, S Donaldson, A Frankish, R Gibson, E Hart, R Kinsella, G Laird, J Loveland, J Mudge, A Mujica, J Rajan, H Sehra, E Sheridan, C Steward, M-M Suner, M Thomas, Jennifer Harrow and Tim Hubbard

Informatics Department, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1HH, UK

The Havana (Human And Vertebrate Analysis and Annotation) group at the Wellcome Trust Sanger Institute is tasked with manual annotation of, amongst others, mouse and human genomic sequence. Apart from annotation of entire in-house sequenced chromosomes (mouse 2, 4, 11 and X; human 1, 6, 9, 10, 13, 20, 22 and X), we are involved in a number of large collaborative efforts targeting individual genes rather than chromosomes. One of these projects is the CCDS (Consensus Coding DNA Sequence) project, a collaboration between the Hinxton campus (Havana (Vega) and Ensembl), the RefSeq group at the NCBI and the genome informatics group at UCSC. The aim of this project is to define a gold standard reference set of coding transcripts. Initially limited to human, this has now been extended to mouse. Any CCDS candidate transcripts with disagreement between Hinxton and NCBI annotation are manually inspected and where possible agreement is found on the proposed structure. The end result is effectively a combined non-redundant gene set between the various genome centres. We are involved in another combined gene set effort with Ensembl, producing a Havana-Ensembl merged gene set of Havana annotated transcripts that are identical in Vega and Ensembl. These transcripts are coloured gold in Ensembl. Havana is also collaborating with the EUCOMM (European Conditional Mouse Mutagenesis) and KOMP (Knock-Out Mouse Project) efforts producing the accurate and comprehensive, detailed annotation of coding regions and all splice variants that is needed to optimize design of primers to produce conditional knockout mutants. I will present an update on our current projects and show the utility of manual annotation for these.

P31**POSMED: A WEB DATABASE EXPLORING THE OMIC SPACE BASED ON LITERATURE MINING**

Norio Kobayashi, Yoshikazu Hasegawa, Naohiko Heida, Satomi Matsushita, Keith Player, Yoshiki Mochizuki, Manabu Ishii and Tetsuro Toyoda
RIKEN Genomic Sciences Center

Knowledge discovery from a vast number of biomedical documents is playing an important role in understanding biomedical processes. By applying a text-mining technique, successful results have been reported e.g., extracting relationship between omic items such as diseases and genes from biomedical documents. However, for synthetic understanding of a biological system, we would like a network over various omic items rather than a relationship between specialized items. To realize a network extraction tool, a novel methodology which integrates various omic items and documents into a single framework is required. PosMed has been implemented based on the idea described above. PosMed searches omic items statistically related to a user's keyword by crawling a pre-constructed network. The network consists of 1) relations between documents and items extracted by a text-mining technique and 2) relations between items co-cited in documents utilizing relationships found in step 1. PosMed searches targeted omic items by a full-text search against documents with the user's keyword and applying relations 1 and 2, *i.e.*, via paths keyword -> documents-> item and keyword -> documents -> item -> item. The resultant items are shown as a network as well as a list ranked by statistical significance. PosMed now supports document sets generated from ontologies such as BioPAX, in addition to document sets MEDLINE, OMIM and PPI. It also includes the omic items rat genes and RIKEN researchers in addition to mouse genes and mouse mutants. In this presentation, we will illustrate how PosMed works with examples utilizing newly added data.

P32**GMV: A COMPARATIVE GENOME BROWSER FOR MURASAKI**

Yasunori Osana, Kris Popendorf, Yasubumi Sakakibara
Keio University

GMV is a multi-purpose comparative genome browser tailored for a comparative genome system, called Murasaki, which is a scalable and fast language-theory based homology detection tool across multiple genome sequences. GMV provides visualization function and genome analysis function for the results output by Murasaki for multiple genome comparisons. GMV can be executed on multi-platforms including Unix, MacOS X and Windows in both of batch and interactive operation modes. Batch mode is suitable for automated plots to check Murasaki's output at runtime, and in interactive mode is to use as a comparative genome browser. The key features of GMV are: (a) draw homologous segments (called anchors) detected by Murasaki across multiple genomes, (b) display annotation/expression data from GenBank/GFF formatted files attached to anchors, and also from on-line DAS servers such as UCSC, Ensembl or KEGG have, (c) explore the relations among homologous anchors and gene coding regions interactively by using annotation information together, (d) interactive access to databases on the web, (e) filter noisy anchors by employing language-theoretic analysis such as TF-IDF scores from Murasaki output and annotation data, and (f) generate text report on comparative genome analysis using several functions of GMV. GMV can also import Mauve and BlastZ outputs, as a general purpose comparative genome browser. This poster demonstrates the effectiveness of GMV on several mammalian genome comparison analyses. GMV is publicly available at: <http://murasaki.dna.bio.keio.ac.jp/gmv.html>.

P33

SHOE: PROMOTER ANALYSIS TOOL

Natalia Poluliakh^{1,2}, Harada Hajime¹, Wataru Fujibuchi¹, Martin Frith¹, Paul Horton¹

¹AIST, Computational Biology Research Center, ²Sony CSL

SHOE: Promoter Analysis Tool

SHOE (Sequence HOmology & Expression, also mnemonic with phylogenetic footprinting) is a program for promoter analysis which combines information about transfection factor binding affinities from Position Weight Matrix (PWM) libraries, co-expression data from large scale microarray database mining, and conservation information from alignments of human, mouse, and rat promoters.

DATA AND DATA PREPARATION

Promoter Sequences

Promoter region sequences are taken from Refseq and DBTSS, using transcription start sites from DBTSS when possible. Orthologous promoters were obtained from the Mammalian orthology database

Orthologous promoters were first roughly aligned pairwise with SSEARCH and then Clustal W was used to construct a multiple alignment for regions in which the orthologs for all three species (human, mouse, and rat) were available and alignable.

Expression Data

For gene expression data, SHOE uses microarray data from the Gene Expression Omnibus Database.

Co-expressed gene pairs

SHOE uses Pearson's correlation coefficient to measure the degree of co-expression between two genes. We implemented a multipass algorithm to perform this calculation in a memory efficient manner.

PWM libraries

SHOE can use Position Weight Matrix (PWM) libraries from either Transfac or Jaspar to represent the binding site preferences of known transcription binding factors.

Gene Name Unification

Gene names were translated between Transfac names, UniGene, and RefSeq ids as necessary to allow comparison between the various sources of data.

SCORE CALCULATION

To identify sites which are likely to be regulated by a particular transcription factor, we combine the information from PWM libraries with sequence conservation and correlated expression levels. For the PWM score we use the likelihood ratio of the sites being generated by the particular transcription factors PWM to a high order Markov background model. The scores for conservation and correlated expression are currently defined in a more ad hoc manner, based on the empirically observed distribution of "conserved" columns in alignments of unrelated promoters and the expression correlation of all possible gene pairs.

PRELIMINARY RESULTS AND CONCLUSION

A preliminary version of SHOE has been applied to the analysis of the human, G-protein coupling MAP kinase pathway (Polouliakh et al 2006). Several transcription factors implicated in the literature (STAT5A, STAT1, NF-kappaB, IRF-1, MAX, CREB, VDR) were identified in promoters of genes from that pathway but not in promoters of a ribosomal control group. Since SHOE proved useful in that study, (after refining the implementation), we plan to release it to the community in 2008.

P34**THE MOUSE GENOME ASSEMBLY AND MOUSE RESOURCES AT NCBI**

Valerie A. Schneider, Deanna M. Church and the NCBI annotation team
National Center for Biotechnology Information, NIH, Bethesda, MD, U.S.A.

The current reference assembly for the mouse genome (Build 37) consists of a high quality, largely clone-based assembly. The assembly is produced from a single strain (C57BL/6J). Coordination between the sequencing centers and NCBI has made it possible to generate a very high quality assembly. For Build 37, all joins between finished clones that did not meet defined quality standards were manually inspected, as were all gaps between scaffolds, of which there are 105. Over 96% of the reference assembly consists of HTGS_Phase3 (finished) clone based sequence, 3.5% of the assembly consists of gaps, largely representing the centromere and less than 1% of the assembly is either draft or WGS sequence. The vast majority of the remaining WGS sequence is unplaced and typically represents regions of the genome that have been difficult to map in large insert clones.

In addition to the reference assembly, Build 37 contains the Celera assembly as well as several partial assemblies that have been assembled in a strain specific fashion from finished sequence in GenBank. All of these assemblies have been annotated using the NCBI genome annotation process (gpipe, <http://www.ncbi.nlm.nih.gov/projects/genome/guide/build.html>). Features annotated include genes, RefSeq transcripts, genomic sequence not used in the assembly, clones, repeats, gene trap clones and MICER clones. All annotation on all assemblies is available through the NCBI Map Viewer. The Map Viewer allows for the display and integration of multiple coordinate systems such that different assemblies, as well as non-sequence based maps, can be shown together. Gene and transcript, which continue to be updated in collaboration with Mouse Genome Informatics and the scientific community, are also available in Entrez Gene and RefSeq. In addition, we are developing new genome guide pages to assist users in navigating available mouse resources.

P35**ASSESSING EXPRESSION TECHNOLOGIES AND THEIR APPLICATION TO COMPLEX LOCI.**

Alistair M. Chalk, Matthew Anderson, Ryan Wolski, Nick Matigian, Alan Mackay-Sim and Christine Wells
National Center for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University

The ability of high throughput sequencing technologies to quantify gene expression is adding resolution to how we view expression of a gene. Particularly for complex loci it is important to be able to visualise the different signals from the various sequencing and expression technologies and how they relate to the expression of alternative transcripts. We have developed a viewer to display sequence tags (CAGE) as well as platform array expression data (affymetrix/illumina) on a locus by locus manner. In addition we are able to visualise features of the alternative transcript products (i.e. domains, transmembrane topology, and protein localisation).

This approach provides the means combine multiple lines of evidence of transcription across a locus, further providing information on promoter usage, exon usage and the relationships between these and summary (3' UTR) expression data. We aim to integrate genetic data derived from the Hapmap projects as well as variants identified from deep sequencing approaches.

Contact: a.chalk@griffith.edu.au

P36**CASIMIR; TOWARDS A STRATEGY FOR COORDINATION AND SUSTAINABILITY OF MOUSE INFORMATICS RESOURCES**

Paul Schofield¹, Vassilis Aidinis², Ewan Birney³, Steve Brown⁴, Duncan Davidson⁵, Michael Gruenberger¹, Anne-Marie Mallon⁴, Glenn Proctor³, Eli Reuveni⁶, Nadia Rosenthal⁶, Klaus Schughart⁷, Glauco Tocchini-Valentini⁸, Tom Weaver⁹, John Hancock⁴

¹Dept of Physiology Development and Neuroscience, University of Cambridge, UK, ²Institute of Immunology, BSRC Fleming, Athens, Greece, ³European Bioinformatics Institute, Hinxton, UK., ⁴MRC Mammalian Genetics Unit, Harwell, UK, ⁵MRC Human Genetics Unit, Edinburgh, UK., ⁶EMBL-Monterotondo Outstation, EMBL, Monterotondo, Italy. ⁷Dept. Experimental Mouse Genetics, Helmholtz Centre for Infection Research, Braunschweig, Germany. ⁸Istituto di Biologia Cellulare, CNR, Monterotondo, Italy., ⁹Geneservice Limited, Cambridge, UK.

The deluge of mouse genetic and phenotypic data currently being generated from hypothesis driven and high throughput mutagenesis studies requires a substantial informatics infrastructure in order for it to be retrieved and used for computation. Much of the current effort in functional genomics, and our understanding of the biology of human disease, is therefore underpinned by informatics infrastructures currently served by a diverse collection of relevant databases throughout the world. Standardisation of data representation, structure, data transfer and database querying protocols to provide complete database interoperability is essential in the post genomic age. Failure to sustain these databases and other resources in the future will have seriously deleterious effects on existing research in that data resources will remain fragmented and the potential synergy lost. An issue secondary to interoperability therefore, is that of database sustainability, both scientific and financial in terms of funding and support for curation and development. In response to these needs the European Commission has funded CASIMIR, a Coordination Action of the 6th Framework Programme (<http://www.casimir.org.uk>). The aims of CASIMIR are to provide recommendations on data standards, database organisation and funding of mouse informatics in Europe and to facilitate discussion between mouse centred groups on potential areas where integration of information is most vital. The project brings together a core of major European informatics groups and stakeholders, primarily to develop a strategy for the integration and interoperability of European mouse databases concerned with the genetics and biology of the mouse.

P37**A GENOME-WIDE ANALYSIS OF TRANSCRIPTION REGULATION IN MACROPHAGE, THE ACTIVITY IN THE GNP-ECW/FANTOM4**

Harukazu Suzuki and The GNP-ECW/FANTOM4 consortium

Laboratory for Genome Exploration Research Group, RIKEN GSC, Yokohama, Japan

One of the major challenges in current genome biology is to understand how 2000 of transcription factors cooperatively regulate transcription in mammals. Here we introduce our genome-wide approach for the transcription regulation, which is going on in the Expression Cluster Workshop (ECW) of the Genome Network Project (GNP)/FANTOM4. Although our aim is to construct a pipeline to analyze transcription regulation of any biological events in mammals, we chose differentiation and LPS response in macrophage cells as a model system. We took series of experimental data of PMA- and LPS-stimulated THP-1 cells and LPS-stimulated HMDM cells in the time-dependent manner; those are three unique expression profiles (deep CAGE, microarray and TF-qRT-PCR), HTS small RNA sequences, several CHIP on chip profiles (Ac-H3K9, PolII and a few key TFs), TF protein-protein interaction data and etc. Based on those experimental datasets, computational integration analysis has been achieved. We present our current results in our activity.

P38**PREPARATION OF DEEP-CAGE LIBRARIES FOR 454 AND SOLEXA SEQUENCERS**

Mitsuyoshi Murata, Sachiko Ishikawa, Yoshiko Hayashida, Kengo Hayashida, Hiromi Nishiyori, Hiromi Sano, Akira Hasegawa, Michihira Tagami, Shiro Fukuda, Shinji Kondo, Lassmann Timo, Arner Erik, Daub Carsten, Piero Carninci, Yoshihide Hayashizaki

RIKEN

We have proadly applied CAGE technology to identify 5'-end of mRNA effectively and provide expression profiling data and promoter identification. CAGE allows identification of transcription starting sites (TSS), search of promoter region on upstream of TSS and information on transcription factor binding sites. We have improved the method for this kind of libraries preparation and achieved time and cost saving.

The earliest CAGE method takes 12 days. But bland new method takes only 5 days to 7 days it depends on sequencer we use. We don't use any toxic reagents like Phenol or Chloroform. And we also built up new purification method for this protocol.

Here we introduce we have completed the method of CAGE library preparation, it can deal with 454 sequencer (454 LIFE SCIENCES) spend 1/10 cost and analyze approximately 100 fold data and also with Solexa sequencer (Illumina), which costs 1/250 of classic sequencing and allow analyzing 140 fold data compare to RISA sequencer (SHIMADZU BIOTECH), the RIKEN Sanger-technology based instruments.

Such developments makes the cost of sequencing CAGE libraries comparable to standard microarrays and paves the way for a broader diffusion of CAGE.

P39**NEXT GENERATION SEQUENCER DATA PROCESSING AND MANAGEMENT SYSTEM FOR DEEP CAGE AND SHORT RNA SEQUENCES**

Jessica Severin, Timo Lassmann, Ryoko Ishihara, Shintaro Katayama, Hideya Kawaji, Yoshihide Hayashizaki, Carsten Daub

Riken

The next generation sequencers (Selexa, 200 base 454, ABI SOLID, Helicos) are poised to open new windows in the field of genomics. These new instruments can generate millions of sequence reads per experiment, allowing us to not only look deeper into the inner workings of cells, but also can be used for obtaining functional data deriving from RNA (like CAGE, short RNA libraries) or DNA (chromatin immunoprecipitation). But this increased data also imposed new demands on data management and processing systems. Presented here is a new database and processing system to deal with the demands of these new instruments. The system is modular allowing for different algorithms and processes to be added as research demands. It is based on an 'incremental build' design and thus allows each unique sequence to be processed only once. The system uses advanced AI computer techniques (blackboard systems, autonomous agents, emergent behaviour) to allow for large scale parallel computations on potentially heterogeneous clusters. Currently the system does annotation screening of sequence against known RNA databases via BLAST, and does mapping of short sequences to chromosomes via BLAST, Exonerate, and SSAHA2. We plan to add additional mapping algorithms (vmatch, custom heuristic searches) and cluster analysis of mappings in the near future.

P40

MOUSE PHENOME DATABASE: SELECTED UPDATES (OCT 2007)

Terry Maddatu, Stephen Grubb, Carol Bult, Molly Bogue
The Jackson Laboratory

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 sequence, SNP and haplotype data. 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 information. The web interface for the Mouse Phenome Database (MPD) provides tools for online analysis and enables data retrieval/downloads for offline custom analyses. Bibliography Bogue MA, Grubb SC, Maddatu TP, Bult CJ. Mouse Phenome Database (MPD). Nucleic Acids Res. 2007 Jan; 35 (Database issue): D643-9. PMID: 17151079 www.jax.org/phenome

P41

THE CENTER FOR GENOME DYNAMICS RESEARCH OVERVIEW

Randy Smith, The Center for Genome Dynamics*, Gary Churchill
The Jackson Laboratory

The Center for Genome Dynamics mission is to evaluate the role of genome-wide organization in mammalian biology by developing detailed maps of interactions that encompass allelic diversity, functional categories, gene expression, recombination hotspots, and phenotype associations. We believe that understanding these principles will provide new genetic approaches to understanding human health and disease at the systems level. This presentation will highlight recent progress at the Center in the areas of gene expression response to high fat diet, structural model analysis for adiposity in mice, analysis of the genetic structure of mouse populations by the combinatorial analysis of long-range linkage disequilibrium, the use of linear model genome scans for expression QTL analysis, distribution and control of recombination activities in the mouse, and the subspecific origin of the laboratory mouse.

P42**Evola: ORTHOLOG DATABASE OF ALL HUMAN GENES IN H-InvDB WITH MANUAL CURATION OF PHYLOGENETIC TREES**

Ryuichi Sakate¹, Akihiro Matsuya^{1,2}, Yoshihiro Kawahara¹, Yoshiharu Sato¹, Takashi Gojobori^{3,4}, Takeshi Itoh^{3,5}, Tadashi Imanishi³

¹JBIRC, JBIC, Tokyo, Japan, ²Hitachi, Co., Ltd., Tokyo, Japan, ³BIRC, National Institute of AIST, Tokyo, Japan,

⁴Center for Information Biology and DDBJ, NIG, Shizuoka, Japan, ⁵Division of Genome and Biodiversity Research, NIAS, Ibaraki, Japan

Evola is a sub-database of H-InvDB, an integrated database of annotated human genes (<http://h-invitational.jp/>), and contains human orthologs for 11 vertebrates (chimpanzee, macaque, mouse, rat, dog, cow, opossum, chicken, zebrafish, tetraodon and fugu). Orthologs are a pair of genes in different species that evolved from a common ancestral gene by speciation. With the rapid growth of transcriptome data of various species, more reliable orthology information is prerequisite. However, detection of orthologs could be erroneous if pairwise distance-based methods, such as reciprocal BLAST searches, were utilized. Thus, Evola was constructed as a fully curated database of evolutionary features of human genes. In the process of ortholog detection, computational analysis based on genome synteny and transcript sequence similarity was followed by manual curation by researchers examining phylogenetic trees. In total, 18,968 human genes have orthologs among the 11 vertebrates; either computationally detected or manually curated orthologs. Evola provides amino acid sequence alignments and phylogenetic trees of orthologs and homologs. In “N/dS view”, natural selection on genes can be analyzed between human and other species. In “ocus maps” all transcript variants and their exon/intron structures can be compared among orthologous gene loci. We expect Evola to serve a comprehensive dataset to be utilized in comparative analyses for obtaining new knowledge about human genes. Evola is available at <http://jbirc.jbic.or.jp/hinv/evola/>.

P43**ONTOLOGICALDISCOVERY.ORG: A WEB RESOURCE FOR THE EMPIRICAL DISCOVERY OF PHENOTYPIC RELATIONS ACROSS SPECIES AND EXPERIMENTAL SYSTEMS**

Erich J. Baker¹, Zuopan Li², Jeremy Jay², Vivek Philip³, Yun Zhang², Michael A. Langston^{2,3}, Elissa J. Chesler^{3,4}

¹Computer Science Department, Baylor University, Waco, TX, USA, ²Computer Science Department, University of Tennessee, Knoxville TN, USA, ³Genome Science and Technology Program, University of Tennessee, Knoxville TN, USA, ⁴Oak Ridge National Laboratory, Oak Ridge, TN, USA

The Ontological Discovery Environment (<http://ontologicaldiscovery.org>) is a free, public Internet resource for the storage, retrieval and analysis of phenotype-centered genomic data sets. The purpose of the tool is to allow users to define phenotype categories based on naturally occurring biological networks, pathways and systems, rather than on externally manifested constructs. The central organizing metaphor of a gene set may be created, stored and curated by individual users, shared among virtual working groups or made available for access by all users. Gene sets are submitted using a variety of accession numbers, microarray feature IDs, and gene symbols from several model organisms. The use of diverse ID types allows integration across experimental platforms, literature reviews, and other genomic analyses. In addition, each gene is associated with a score, which can be a statistic, binary value, or effect size, indicating the strength of gene-phenotype association. The sets are annotated with several levels of meta-data including through unstructured description, publication information, and structured community ontologies for anatomy, process and function. Gene sets are translated to a user-chosen reference species through gene homology, allowing translational comparison of models regardless of face validity of the experimental systems. Analyses of set similarity, distance, and hierarchical relations among sets are performed through analysis of the bi-partite network of gene-phenotype relations. Submitted and computationally derived gene sets can be integrated into a “Phenome Map” which constructs hierarchical trees of phenotypes based on the genes to which they are associated, allowing for an empirical discovery of the natural phenotype ontology.

Supported NIH U01AA13499, NIH U24AA13513.

P44

DIFFERENCES IN microRNA EXPRESSION ACROSS COMMON MOUSE INBRED STRAINS USING TAQMAN RT-PCR

Michael Parsons¹, Christina Grimm², Wilfried Nietfeld², Hans Lehrach², Leonard C Schalkwyk¹

¹Social, Genetic, Developmental Psychiatry Research Centre, Institute of Psychiatry, KCL, London, UK. ²Max-Planck-Institute for Molecular Genetics (MPIMG) in Berlin, Germany.

MicroRNAs (miRNAs) are single-stranded RNAs that are involved in the regulation of protein-coding genes at the level of messenger RNA (mRNA), which likely underlie their involvement in the regulation of numerous traits including developmental timing, apoptosis, immune function and neuronal development. In order to better understand the regulation of miRNA expression, we investigated miRNA expression differences between four different common inbred strains A/J, Balb/c, C57BL/6J and DBA/2J (n=8 for all strains) to see if the genetic variation across these common inbred strains differentially affect miRNA expression. We have initially conducted a total of 42 miRNA Taqman RT-PCR reactions using pooled RNA samples within strain. The six miRNAs that had markedly different miRNA expression levels, were repeated in the individual samples, with four reactions showing significant differences across strain ($p < 0.01$). This is the first observation of miRNA expression differences across inbred mice strains. These candidate miRNAs are currently being investigated in further studies using more complex genetics populations, including recombinant inbred mice (BXD RI) and consomics (A/J on a C57BL/6J background), which will allow us to better characterise the regulation both miRNA expression (eQTL mapping) as well as miRNAs regulation of mRNA expression (www.genenetwork.org).

P45

TRANSCRIPTS AND THEIR EXPRESSION PATTERNS IN THE REGION CONTAINING THE HYBRID STERILITY 1 GENE

Ondrej Mihola^{1,2}, Zdenek Trachtulec^{1,2}, John Schimenti³, and Jiri Forejt^{1,2}

¹Dept Mouse Mol. Genet., ²Center for Applied Genomics; Inst. Mol. Genet., AS CR, Prague, Czech Republic;

³Center for Vertebrate Genomics, Dept Biomed. Sci., Cornell University, Ithaca NY

The alleles of the Hybrid sterility 1 (*Hst1*) gene affect fertility of F1 males in crosses between laboratory strains and *Mus musculus musculus* mice. In certain types of crosses, such as (PWD x C57BL/10) F1, they cause a breakdown of spermatogenesis at the stage of primary spermatocytes. The *Hst1* gene has been mapped to a 360-kb region of mouse chromosome 17. The region contains six known genes. Here we describe alternative mRNA isoforms for five of the six candidate genes and the evaluation for *Hst1* candidacy by real-time RT-PCR on testicular mRNAs of sterile and fertile mice. The isoforms change the open reading frame or 5'-untranslated region by alternative exon inclusion or the 3'-untranslated region due by alternative polyadenylation.

P46**IDENTIFICATION OF NOVEL TRANSCRIPTOME SIGNATURES CHARACTERISTIC OF ERYTHROPOIETIC RESPONSE BY HIGH-RESOLUTION PROFILING OF MOUSE BLOOD USING SOLEXA 1G SEQUENCING TECHNOLOGY**

Matthew N. Fedoruk, Jim L. Rupert
University of British Columbia

Deep sequencing of the blood transcriptome is an important step in elucidating the transcriptional profile of blood cells and how these profiles are influenced in response to physiological or environmental alterations. Blood plays a crucial role in immunity, metabolism and homeostasis in humans and animals, and is routinely used in the differential diagnosis of both hematological and nonhematological disorders, as well as a marker of a number of non-pathological changes, such as erythropoietic response to hypoxia.

DESIGN: Comparative blood transcriptome analysis is being investigated as a potential testing protocol for erythropoietin (EPO) abuse in athletes. Massively parallel signature sequencing (MPSS) has been applied to the transcriptome of pooled C57BL6J mouse whole blood samples exposed to the following conditions: normoxia (control), hypoxia (13.5% O₂), and subcutaneous recombinant EPO injection (10mg/kg), with the ultimate goal of identifying transcripts, or patterns of transcripts, that are diagnostic of exogenous EPO administration.

RESULTS: Using Solexa's unique high-throughput sequencing technology, this study has yielded > 7 million tags representing > 700,000 unique RNA sequences per experimental library. Using the SAGE data analysis software DiscoverySpace 4, we have identified both known and novel tag species that are significantly differentially-expressed between the experimental conditions, and performed tag-to-gene mapping. To assess inter-individual variation in these candidate transcripts, differential expression patterns observed in the pooled samples are being confirmed using qPCR in individual mice. Once robust and consistent candidate transcripts have been identified in mice, their human homologues will be identified and evaluated for their potential as markers for exogenous erythropoietin activity in a blood-based anti-doping test.

CONCLUSION: To our knowledge this study represents the most comprehensive expressed sequence tag analysis of the mouse whole blood transcriptome.

S1-7/P47**AN ENU-INDUCED MUTATION OF A MIRNA ASSOCIATED WITH PROGRESSIVE HEARING LOSS**

Morag Lewis¹, Elizabeth Quint², Agnieszka Rzadzinska¹, Anne Kent-Taylor¹, Helmut Fuchs³, Martin Hrabé De Angelis³, Cordelia Langford¹, Stijn Van Dongen¹, Anton Enright¹, Nick Redshaw⁴, Tamas Dalmay⁴, Karen P Steel^{1,2}
¹Wellcome Trust Sanger Institute, Hinxton, UK; ²MRC Institute of Hearing Research, Nottingham, UK; ³GSF National Research Center for Environment and Health, Munich, Germany; ⁴University of East Anglia, Norwich, UK.

Progressive hearing loss is very common in the human population, but relatively little is known about the molecular or genetic basis. We report here a new ENU-induced mouse mutant called *diminuendo* (*Dmdo*), inherited in a semi-dominant manner. Heterozygotes show progressive loss of auditory responses and hair cell degeneration. Homozygotes show no cochlear responses, have extensive hair cell loss by 4 weeks and show clear head-bobbing. Homozygotes also show anomalies in hair bundle maturation from as early as 5 days after birth, with reduced apical surfaces of hair cells and progressive loss of the staircase organization; however, interstereocilia links appear unaffected at early stages. We mapped the mutation to a 4.5 Mb region of chromosome 6 containing 40 genes, and resequenced the majority of exons within this region. Eventually, we found a single base change in the seed region of a microRNA, *mmu-mir-96*. This miRNA is expressed in sensory hair cells from an early stage, and we are currently investigating candidates for its target sites to elucidate the mechanism of action.

P48**ANALYSIS OF MICRO RNAS INVOLVED IN OSTEOLASTIC DIFFERENTIATION**

Yosuke Mizuno¹, Ken Yagi¹, Tatsuo Suda¹, Takenobu Katagiri², Toru Fukuda², Yasushi Okazaki¹

¹Div. Funct. Genomics & Syst. Med. and ²Div. Pathophysiol., Res. Ctr. for Genomic Med., Saitama Med. Univ.

Micro RNAs (miRNAs) are small RNA molecules of about 22 bases, which don't code protein, and approximately 400 distinct miRNAs have been experimentally discovered. In cytoplasm, miRNAs form the RISC (RNA -induced silencing complex) and inhibit translation of genes which have complementary sequence to the miRNAs. Because a few mismatches are tolerated in the sequence complementarity between miRNAs and their target genes, each miRNA has more than 200 different predicted target genes. Recently miRNAs have been reported to play significant roles during development, proliferation and differentiation in many types of cells and tissues, including granulocytes, myocytes and osteoclasts. However, the roles of miRNA in osteoblast differentiation have not been determined yet. We induced osteoblastic differentiation of mouse bone marrow-derived ST2 cells by treatment with bone morphogenic protein 4 (BMP4). The expression level of each miRNA during the osteoblastic differentiation was analyzed systematically using a miRNA microarray developed in house. The expression levels of several miRNAs have significantly changed during the osteoblastic differentiation of ST2 cells induced by BMP4. To examine functions of these miRNAs, we transfected each miRNA or its antisense strand into ST2 cells and induced by BMP4. We found that several miRNAs significantly changed the levels of differentiation markers for osteoblasts. Taken together, these results suggest that several miRNAs play important roles in the regulation of osteoblastic differentiation.

P49**GLOBAL ANALYSIS OF MICRORNA AND GENE EXPRESSION IN HUMAN CELL LINES**

Yoshinao Ruike¹, Atsuhiko Ichimura¹, Soken Tsuchiya², Kazuharu Shimizu², Ryo Kunimoto³, Yasushi Okuno³ and Gozoh Tsujimoto¹

¹Department of Genomic Drug Discovery Science, ²Department of Nanobio Drug Discovery, ³Department of Pharmacoinformatics, Graduate School of Pharmaceutical Sciences, Kyoto University.

MicroRNAs (miRNAs) are small non-coding RNAs that negatively regulate gene expression at posttranscriptional level. Although considerable progress has been made in studying the function of miRNAs, it still remains largely unclear mainly because of the difficulty in identifying target genes for miRNA. As an initial effort to analyze the correlation between miRNA and target genes, we examined 148 human miRNA and gene expression across 16 human cell lines. Applying computational target prediction algorithm on the data obtained, we could extract physically relevant miRNA-target gene pairs. Furthermore, GO-based annotation and functional enrichment analysis of the extracted miRNA-target gene pairs made it possible to functional interpretation of miRNAs. The approach would be of value for further studying the function of miRNA.