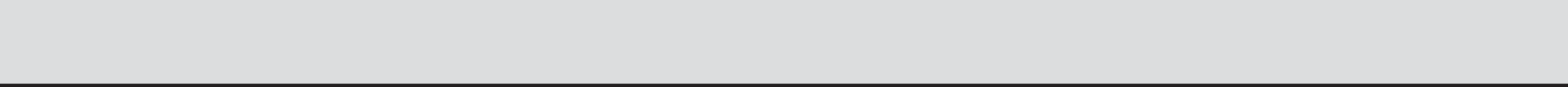


**Monday October 29**  
**Oral Presenters Abstracts**



**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****9.00 – 9.15am**

O1-1

**PITUITARY TRANSCRIPTOME: 5 LIBRARIES REVEAL NOVEL GENES AND PATHWAYS  
ACTIVE IN ORGANOGENESIS**

S. A. Camper<sup>1</sup>, N. M. Solomon<sup>1</sup>, M. L. Brinkmeier<sup>1</sup>, S.W. Davis<sup>1</sup>, P. Carninci<sup>2</sup>, T. Shiraki<sup>2</sup>, T. Arakawa<sup>2</sup>, J. Kawai<sup>2</sup>, A. H. Mortensen<sup>1</sup>, J. W. MacDonald<sup>1</sup>, D. Ghosh<sup>1</sup>, Y. Hayashizaki<sup>2</sup>, R. H. Lyons<sup>1</sup>

<sup>1</sup>University of Michigan, Ann Arbor, MI 48109, USA; <sup>2</sup>Riken Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan

Hayashizaki and colleagues developed a catalog of the mammalian transcriptome by sequencing cDNA clones from 246 mouse tissues (Carninci et al., *Genome Research* 13:1273, 2003). Three volumes arose from pituitary glands dissected when organogenesis is affected by Prop1 mutations: E14.5Prop1df/df, WT E12.5 and E14.5. Each of these had a high novelty value, ranking in the top 20/246. We report comparison with two novel cDNA subtraction libraries: E14.5WT - E14.5Prop1df/df; E14.5WT - E12.5WT. 55,755 cDNA sequences were assembled in a searchable database, representing 12,014 expressed genes. Each subtracted library exhibited the expected content skewing and contributed significantly to the pituitary transcriptome. The Prop1 mutant pituitary transcriptome resembles the e12.5 more than e14.5, as anticipated for developmental arrest. Gene ontology analysis demonstrated that cDNAs common to each library include housekeeping genes, and revealed both expected and novel categories in the mutant, i.e. the characteristic reduced vascularization and increased apoptosis fits the category 'response to oxygen radical'. We validated the coverage of the libraries by detecting all 14 of the known homeobox gene transcription factor cDNAs as full-length or longer. The full-length nature is important for information content and functional studies. 31 homeobox genes not previously known to be part of the pituitary transcriptome were detected and validated by RT-PCR and/or in situ hybridization. A similar approach was taken for discovering novel signaling pathways, which yielded members of the WNT, NOTCH, and BMP pathways. These genes are exciting candidates for regulators of pituitary development and function in mouse and man. (NICHD, UM-CCMB)

**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****9.15 – 9.30am****O1-2****CRACKING THE SECOND GENETIC CODE**

Timothy R. Hughes<sup>1</sup>, Gwenael Badis-Breard<sup>1</sup>, Michael Berger<sup>2,3</sup>, Shaheynoor Talukder<sup>1</sup>, Andrew Gehrke<sup>2</sup>, Anthony A. Philippakis<sup>2,3</sup>, Esther Chan<sup>1</sup>, Savina Jaeger<sup>2</sup>, Quaid D. Morris<sup>1</sup> and Martha L. Bulyk<sup>2,3</sup>

<sup>1</sup>University of Toronto, Toronto ON, CANADA; <sup>2</sup>Brigham and Women's Hospital, Boston, MA, USA;

<sup>3</sup>Harvard Medical School, Boston, MA, USA

The encoded sequence preferences of DNA-binding proteins represent the primary mechanism by which cells interpret the genome, and the relationship between transcription factors and cis-regulatory elements has been termed the “second genetic code” Despite the central importance of DNA-binding activities to cell biology, physiology, development, and evolution, the complete binding activity of only a small minority of DNA-binding proteins has been experimentally established. We have used microarrays containing all possible 10-base sequences to examine the binding specificities of over two hundred mouse DNA-binding proteins representing twenty-two different structural classes. Our results reveal a surprisingly complex landscape of DNA-binding activities, with most proteins possessing unique sequence preferences. Moreover, individual proteins often recognize multiple distinct motifs, such that consensus and position weight matrix representations can fail to identify some of the best binding sequences. These data indicate that a complete vocabulary of protein-DNA affinity will be required to completely understand the function of the genome. We propose that the evolutionary success of DNA-binding protein families, typically attributed to modularity in spatial expression or protein-protein interactions, is also be due to diversity and malleability in DNA sequence recognition, which in turn could facilitate evolution of regulatory programs. We anticipate that the data resulting from this effort will be invaluable for understanding both gene regulation and genome evolution.

**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****9.30 – 9.45am****O1-3****RESCUE OF MULTI-MAP CAGE TAGS: IMPLICATIONS FOR STUDIES OF PROMOTER ARCHITECTURE AND REPEAT ELEMENT EXPRESSION**

Geoffrey J Faulkner<sup>1</sup>, Alistair RR Forrest<sup>2,3</sup>, Alistair M Chalk<sup>3</sup>, Kate Schroder<sup>1</sup>, Yoshihide Hayashizaki<sup>3</sup>, Piero Carninci<sup>2,4</sup>, David A Hume<sup>5</sup> and Sean M Grimmond<sup>1</sup>

<sup>1</sup>The Institute for Molecular Bioscience, University of Queensland, Queensland, Australia. <sup>2</sup>Genome Exploration Research Group (Genome Network Project Core Group), RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045, Japan. <sup>3</sup>The Eskitis Institute for Cell and Molecular Therapies, Griffith University, QLD 4111, Australia. <sup>4</sup>Genome Science Laboratory, Discovery Research Institute, RIKEN Wako Institute, Wako, Saitama 351-0198, Japan. <sup>5</sup>Roslin Institute, University of Edinburgh, Roslin EH25 9PS, Scotland, United Kingdom

A major drawback of the CAGE approach thus far has been the removal of as many as 40% of CAGE sequence tags due to their mapping to multiple genomic locations. In this work based on FANTOM3 mouse CAGE data we addressed the origins of multi-map tags and presented a novel strategy for their inclusion into CAGE-based studies of the mammalian transcriptome, resulting in a dramatic improvement in terms of the number of mouse loci detected by CAGE (5516 more TKs). This also led to the majority of multi-map tags being re-incorporated, or “rescued”, back into the CAGE data set, whilst maintaining an acceptable correlation with array data. Finally, we investigated the characteristics of promoters primarily supported by rescued multi-map tags and discovered that these promoters presented a narrow single-peak (SP), TATA-box associated architecture at a substantially higher rate than promoters mostly affiliated with single-map tags. We concluded that the additional promoters identified by rescued multi-map tags represent highly conserved SP-class subfamilies preserved throughout gene duplication events.

**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****9.45 – 10.00am****O1-4****WHOLE GENOME SHOTGUN SEQUENCING OF M. M. MOLOSSINUS-DERIVED MSM/MS AND  
DETECTION OF VAST AMOUNT OF SNPS AGAINST C57BL/6**

Toyoyuki Takada<sup>1,2</sup>, Toshinobu Ebata<sup>1</sup>, Takanori Narita<sup>1</sup>, Tadasu Shin-I<sup>1</sup>, Kuniya Abe<sup>3</sup>, Yoshiyuki Sakaki<sup>4</sup>,  
Atushi Toyoda<sup>4</sup>, Tomoko Sagai<sup>1</sup>, Akihiko Mita<sup>1</sup>, Kazuo Moriwaki<sup>3</sup>, Yuji Kohara<sup>1</sup>, Toshihiko Shiroishi<sup>1,4</sup>

<sup>1</sup>National Institute of Genetics, Shizuoka, Japan, <sup>2</sup>ROIS TRIC, Tokyo, Japan, <sup>3</sup>RIKEN BRC, Ibaraki, Japan,

<sup>4</sup>RIKEN GSC, Kanagawa, Japan

An inbred strain MSM/Ms was established from Japanese wild mice, *Mus musuculs molossinus*. Large extent of variations in phenotype and genome are observed between MSM/Ms and common laboratory strains, which are predominantly derived from west European *M. m. domesticus*. To unveil the genetic bases of inter-subspecific difference in various traits, we are conducting systematic collection of SNP data between MSM/Ms and common laboratory strains. We undertook whole genome shotgun sequencing of MSM/Ms. Nearly 10 million high-quality sequence reads were generated from genome DNA of MSM/Ms females. Over 6 Gbp of nucleotide sequence data were accumulated in total, which corresponds to approximately 2.4-fold coverage for the mouse genome. Excluding long repetitive elements and GAP regions, at least 77% of the euchromatic region of C57BL/6 genome is now assigned by the MSM/Ms genome sequence. Direct comparison of the MSM/Ms genome sequence with that of C57BL/6 successfully detected over 10 million SNPs. Average SNP frequency is 0.82% between these two strains. Moreover, it is notable that approximately 6.7% of the C57BL/6 genome sequence is almost identical (SNP freq. is less than 0.01%) to that of MSM/Ms. We carried out PCR-based resequencing of several inbred strains with various genetic origins for more than 250 genome regions that have sequence identity between C57BL/6 and MSM/Ms. As a result, no other inbred strains showed such a high identity to the C57BL/6 sequences. It demonstrated that at least 6.7% of the C57BL/6 genome has introgression from *M. m. molossinus*.

**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****10.00 – 10.15am**

O1-5

**COMPREHENSIVE GENOMIC AND FUNCTIONAL ANALYSIS OF THE SRY-BOX 10 (SOX10) LOCUS:  
IMPLICATIONS FOR GENOME-WIDE PREDICTIONS OF SOX10 TARGET GENES**

Anthony Antonellis<sup>1</sup>, Jimmy Huynh<sup>3</sup>, Gabriel Renaud<sup>1</sup>, Tyra Wolfsberg<sup>1</sup>, Gene Elliot<sup>2</sup>, Andy McCallion<sup>3</sup>,  
Eric D. Green<sup>1</sup>, and William J. Pavan<sup>2</sup>

<sup>1</sup>Genome Technology Branch and <sup>2</sup> Genetics Disease Research Branch, National Human Genome Research Institute, Bethesda, MD 20892. <sup>3</sup>Johns Hopkins University School of Medicine, Baltimore, MD 21205

The transcription factor SOX10 is a key regulator of genes involved in neural crest development, and mutations in the SOX10 gene are associated with human diseases that affect neural crest derivatives. While SOX10 targets have been identified, the mechanisms of SOX10 regulation remain elusive. Recently, we used comparative sequence analysis to identify 9 highly-conserved, non-coding segments upstream of SOX10. We demonstrated that a transgene induced hypomorphic mutant allele of the mouse Sox10 gene has deleted three of these sequences, compromising its expression in some but not all tissues. Thus, Sox10 expression is likely dependent on these and other cis-acting regulatory sequences. Our goal is to characterize cis-acting transcriptional regulatory sequences at SOX10. We have thus pursued three research avenues: (1) Test conserved segments for enhancer potential in cell lines; (2) Determine if segments drive reporter gene expression in developing zebrafish and mice; and (3) Delete segments in a BAC spanning Sox10 and test for altered reporter-gene expression in transgenic mice. These studies have revealed that 4 highly-conserved genomic segments drive reporter gene expression in neural crest derivatives in vitro and in vivo. Two of these segments harbor head-to-head SOX binding sites that are conserved down to chicken. These data are consistent with direct regulation of Sox10 by SOX family members, and/or that Sox10 directly regulates its own transcription. Our work provides a basis for identifying similar DNA sequences in a genome-wide fashion toward identifying novel targets of SOX10, and an outline for this approach will be presented.

**COMPARATIVE GENOMICS / COMPUTATIONAL BIOLOGY  
ORAL PRESENTATION****Monday October, 29****10.15 – 10.30am****O1-6****RNA IN SPERM: TOWARDS AN IDENTIFICATION OF RNA TRANSMITTED FROM SPERM TO ZYGOTE THROUGH FERTILIZATION**

Mitsuoki Mick Kawano<sup>1</sup>, Hideya Kawaji<sup>1</sup>, Chika Kawazu<sup>1</sup>, Toshiyuki Hata<sup>2</sup>, Hiromi Sano<sup>2</sup>, Akira Hasegawa<sup>2</sup>, Shiro Fukuda<sup>2</sup> and Yoshihide Hayashizaki<sup>1,2</sup>

<sup>1</sup>RIKEN, Frontier Research System, Functional RNA Research Program, <sup>2</sup>RIKEN, Genomic Sciences Center

It is widely recognized that all that fathers essentially contribute to the next generation is half of their genome, and mature sperm cells are transcriptionally and translationally inactive. However, an increasing number of reports have shown that RNAs were present in human sperm detected by using RT-PCR, differential display, in situ hybridization and microarrays. One study showed delivery of human RNAs from the sperm to the oocyte on fertilization using the hamster sperm penetration assay. Intriguingly, recent evidence suggests that sperm RNA has the potential to influence phenotype through an epigenetic alteration to gene expression without affecting genotype. Therefore, it is very tempting to identify transcripts in the sperm, which might have an important role in spermatogenesis and/or embryogenesis.

We corrected motile mouse and chimpanzee sperms and isolated total RNA. Then those RNAs were tagged at both ends to make cDNA libraries (15-250 nt) which were sequenced by a 454 sequencer. We could find RNAs including a part of mRNA, rRNA, tRNA, and antisense RNAs and small regulatory RNAs in the sperm. Candidates of transmissible RNAs must be RNAs identified in the sperm but be absent in the unfertilized oocyte and be detected in the fertilized oocyte. And we will try to detect Y chromosome-derived sperm RNAs in the early "female" embryos that do not have the Y chromosome.

We hope that sperm transcriptome analysis will lead to discover new paternal contribution to the progeny.