

Genome Sequencing and Comparative Analysis**Oral Presentation****Saturday November 5****2.30pm – 2.45pm****O-1****PROMOTING MAMMALIAN TRANSCRIPTION**

P Carninci², A Sandelin¹, B Lenhard³, S Katayama¹, K Shimokawa¹, J Ponjavic¹, C A M Semple⁴, M S Taylor¹, V B Bajic⁷, D A Hume⁶, Y Hayashizaki¹

¹ Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama, Kanagawa, Japan, ² Genome Science Laboratory, Discovery and Research Institute, RIKEN Wako Institute, 2-1 Hirosawa, Wako, Saitama, Japan, ³ Center for Genomics and Bioinformatics, Karolinska Institutet, Stockholm, Sweden, ⁴ MRC Human Genetics Unit, Western General Hospital, Edinburgh, United Kingdom, ⁵ University of Oxford, Oxford, United Kingdom, ⁶ ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, Brisbane Qld, Australia, ⁷ Knowledge Extraction Laboratory, Institute for Infocomm Research, Singapore,

To better understand transcription regulation, we have systematically isolated and sequenced CAGE (cap-analysis gene expression) tags corresponding to the 5'-ends of mRNAs in order to map transcription start sites (TSSs) and core promoters in the mouse and human genomes, and to analyze their structure, evolution and diversity. We have produced ~8.9 million CAGE tags from mouse and ~5.5 million CAGE tags from human, to identify 236,498 mouse and 190,513 human TSSs.

We have then produced for the first time a comprehensive hierarchical clustering based on TSS usage. Surprisingly, different classes of promoters determine transcription of different type of RNA transcripts. Our analysis shows that mammalian promoters can be separated into two classes, conserved TATA box enriched promoters, which initiate at a well-defined site, and more plastic, broad and evolvable CpG-rich promoters. Different tissues and families of genes make differential use of these types of promoters.

Our tagging methods allow quantitative analysis of promoter usage in different tissues, and reveal that differentially regulated alternative TSSs are a common feature in protein coding genes. We also identify novel motifs responsible for specific transcription of certain non-coding RNAs. The data can be used to identify tissue-specific promoters, and to analyze the cis-acting elements associated with them. As an example of the utility of the data set, we identify the transcription factors associated with promoters during macrophage differentiation and activation by lipopolysaccharide.

Our data provide the most comprehensive promoter map in mammalian, which paves the way for transcriptional network analysis and systematic system biology.

Genome Sequencing and Comparative Analysis**Oral Presentation****Saturday November 5****2.45pm – 3.00pm****O-2****SINGLE MOLECULE ANALYSIS TO RESOLVE UNASSEMBLED REGIONS OF THE B6 GENOME**

J Amos-Landgraf¹, S Goldstein², S Zhou², J Herschleb², B Teague², L N Kwong², A Shedlovsky¹, W F Dove¹, D C Schwartz²

¹ McArdle Cancer Research Laboratories, University of Wisconsin, Madison, WI, United States, ² Laboratories for Molecular and Computational Genomics, University of Wisconsin, Madison, WI, United States

With the latest assembly of the mouse genome sequence, most of the unique regions of the genome have been resolved; however, repetitive elements have remained refractory to traditional assembly methods. Such repetitive sequences have been implicated in evolution, disease breakpoints, and as recombinogenic substrates. The availability of an optical map covering ~95% of the mouse genome now enables analysis of previously inaccessible regions within the genome. The analysis platform utilizes hundreds of thousands of half-megabase-sized genomic DNA molecules that are individually bar-coded by restriction digestion and then assembled into contigs spanning the entire genome. Alignments of these map contigs with Build 34 sequence has led to the identification of regions of misassembly and putative forms of variation. The maps bridge ~14000 sequence gaps <50 kb and ~250 gaps >50kb, providing an initial visualization of many previously uncharacterized regions. Importantly, optical maps extend an average of ~250 kb on 14 chromosomes into unassembled pericentromeric and telomeric regions. The applicability of such direct visualization can be further illustrated by our analysis of an intestinal cancer modifier, which maps to the pericentromeric region of chromosome 18. Unexpectedly, intrastrain variation was found between the current JAX B6 stock and Build 34 genome sequence located within the pericentromeric region of chromosome 18 and other regions of the genome. Optical mapping provides a necessary approach not only to examine repetitive regions of the genome, but also to assess genome wide germline isogenicity and somatic rearrangements in a small number of tumor cells.

Genome Sequencing and Comparative Analysis

Oral Presentation

Saturday November 5

3.00pm – 3.15pm

O-3

HIGH-RESOLUTION MAPPING OF RECOMBINATION ON MOUSE CHROMOSOME 1

P M Petkov, K Sawyer, L Guccione, K Paigen
The Jackson Laboratory, Bar Harbor, ME, United States

Although it has been known for some years that mammalian meiotic recombination occurs at special sites termed hotspots, rather than occurring randomly along the chromosomes, we know little about the genomic organization of this important biological process. To address this lack, we have constructed the first large-scale, detailed recombination map of a mammalian chromosome using 6000 meioses from crosses of C57BL/6 x Cast/EiJ. 100 Mb of Chr 1 were mapped at 5 Mb resolution, within this 25 Mb at approx 100 Kb resolution and 5 Kb at even greater resolution. We can draw several conclusions from the resulting maps.

Hot spots vary greatly in activity, and all activity classes are nearly equally represented, with only a slight bias against the most active hotspots. As a result, most recombination occurs at very few sites.

Hot spots are not randomly distributed along the chromosome. There are large regions, one Mb or more, with very little or no recombination and “torrid zones” where high activity hotspots are clustered. The result of this is a very “granular” genetic map.

The 1.3X increased recombination on autosomes in females v. males is largely not due to the presence of additional female specific hotspots; the two sexes share the same hotspots which differ in activity in male and female meiosis, although sex-specific recombination sites cannot be ruled out. Surprisingly, the female/male ratio is regionally determined, rather than being an individual property of each hotspot. This last fact requires that we introduce new parameters into our models of mammalian recombination.

Genome Sequencing and Comparative Analysis**Oral Presentation****Saturday November 5****3.15pm – 3.30pm****O-4****IMGT OVERVIEW: 8. MOUSE IMMUNOGLOBULIN AND T CELL RECEPTOR GENES AND IMGT-ONTOLOGY FOR IGSF AND MHCSF**V Giudicelli, M-P Lefranc

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

Except for 24 IGHV genes not yet identified, all the immunoglobulin (IG) and T cell receptor (TR) genes in *Mus musculus* laboratory mice are known: there are 610-614 genes (372-376 IG and 238 TR) in the seven main loci, and 10 IGKV orphans. The number of functional genes is 412-432, comprising 242-250 IG genes and 170-182 TR genes (detailed in the accompanying abstracts). These data result from an exhaustive analysis in IMGT, the international ImMunoGeneTics information system®, <http://imgt.cines.fr> [1], based on the IMGT-ONTOLOGY concepts [2]. The assignment of a sequence to a 'locus', 'group', 'subgroup', 'gene' and 'allele' in IMGT/GENE-DB relies on the 'CLASSIFICATION' concept. The sequence and domain motifs in IMGT/LIGM-DB (22,219 mouse IG and TR sequences) and in IMGT/3Dstructure-DB (544 mouse IG, TR and MHC three-dimensional structures) are described with standardized labels according to the 'DESCRIPTION' concept. Codon and amino acid numbering is based on the 'NUMEROTATION' concept. Interestingly these concepts, and particularly the IMGT unique numbering, initially set up for the domains of the IG, TR and MHC have been extended to the V-LIKE-DOMAIN [3] and C-LIKE-DOMAIN [4] of the IgSF other than IG and TR, and to the G-LIKE-DOMAIN [5] of the MhcSF other than MHC, as illustrated by the IMGT Collier de Perles [6].

[1] Nucl. Acids Res. 33, D593-D597 (2005). [2] Bioinformatics 15, 1047-1054 (1999). [3] Dev. Comp. Immunol. 27, 55-77 (2003). [4] Dev. Comp. Immunol. 29, 185-203 (2005). [5] Dev. Comp. Immunol. 29, 917-938 (2005). [6] In Silico Biology 5, 45-60 (2005).

Genome Sequencing and Comparative Analysis

Oral Presentation

Saturday November 5

4.00pm – 4.15pm

O-5

THE CHARCOT-MARIE-TOOTH DISEASE AND SMITH-MAGENIS SYNDROME REGION: COMPARING MOUSE CHROMOSOME 11 AND HUMAN CHROMOSOME 17

L G Wilming, D Adams, R C Gibson, J G R Gilbert, E A Hart, G K Laird, J E Loveland, J Mudge, S Searle, C A Steward, D Swarbreck, J L Harrow, T Hubbard
Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

Mouse chromosome 11 is the first mouse chromosome to be finished with clone based sequence. This allows for a more detailed and accurate annotation, especially of repetitive, re-arranged and polymorphic regions, than whole genome shotgun sequence. Most of the chromosome is syntenic to human chromosome 17, with smaller parts syntenic to 2 and 22. Human chromosome 17 is associated with the congenital neuropathies Charcot-Marie-Tooth (CMT) disease and Smith-Magenis Syndrome (SMS). I will present data on the analysis of chromosome 11, with the emphasis on a detailed comparison of the CMT and SMS regions in mouse and human. Most likely a reflection of the instability of the region in human (both CMT and SMS are caused by chromosomal deletions), the region is highly re-arranged between the two species. At least nine blocks of contiguous orthology totaling around 8 Mb have been shuffled into different order and orientation in mouse. Also, the mouse region does not feature the various duplications of groups of genes seen in human.

The annotation of both the mouse and the human chromosome will be viewable in Vega (vega.sanger.ac.uk).

Genome Sequencing and Comparative Analysis**Oral Presentation****Saturday November 5****4.15pm – 4.30pm****O-6****ESSENTIAL GENES IN THE MOUSE DEDUCED FROM GENETIC SCREENS**K Hentges¹, B Liu², H Nakamura², M Justice²¹ University of Manchester, Manchester, United Kingdom, ² Baylor College of Medicine, Houston, Texas, United States

We have performed two mutagenesis screens for lethal phenotypes using balancer chromosomes. One screen was localized to mouse chromosome 4, between the STS markers D4Mit281 and D4Mit51. The second screen covered the region between *Trp53* and *Wnt3* on mouse chromosome 11. These screens identified all lethal mutations in the balancer regions, without bias towards any phenotype or stage of death. We have isolated 19 lethal lines on mouse chromosome 4, and 59 lethal lines on chromosome 11, many of which are distinct from previous mutants that map to these regions of the genome. We have characterized the mutant lines to determine the time of death and mutant phenotype. Many of the mutants isolated in both screens have cardiovascular defects. We also performed a pair-wise complementation cross to determine if the mutations are allelic. Assuming a Poisson distribution, we calculated the number of essential genes in each region based on allele frequencies. The percentage of essential genes differs between the balancer regions in our study, and is as high as 44% for the chromosome 11 balancer region, indicating that gene functions are not evenly distributed along the genome. An analysis of genetic conservation across species indicates that this high number of essential genes can be correlated with a high degree of linkage conservation throughout evolution. This initial group of mutants provides a functional annotation of mouse chromosomes 4 and 11, and indicates that many novel developmental phenotypes can be quickly isolated in defined genomic intervals through balancer chromosome mutagenesis screens.