

Monday November 13
Oral Presenters Abstracts

COMPARATIVE SEQUENCE ANALYSIS

ORAL PRESENTATION

MONDAY NOVEMBER, 13

9.00AM – 9.15AM

O1

MAPPING LONG-RANGE ENHANCERS OF BMP4 AND EXPLORING THE ROLE OF LONG-RANGE EVOLUTIONARILY CONSERVED REGIONS FLANKING BMP4

KJ Chandler, RL Chandler, DP Mortlock
Vanderbilt University, Nashville, TN, United States

Bmp4 is a critical, multi-functional developmental gene that has been suggested to play a role in human disease; however little is known about the transcriptional regulation of *Bmp4*. Studies indicate the *Bmp4* proximal promoter is unable to recapitulate all endogenous expression patterns in vivo. Using comparative analysis, we found three noncoding sequences conserved across 450 million years of evolution that reside 50-100 kilobases from the *Bmp4* promoter and are maintained in a syntenic group across vertebrates. To test our hypothesis that *Bmp4* has maintained long-range ECRs that are required for tissue-specific expression of *Bmp4*, we have utilized several experimental approaches. To map regulatory elements within a 398kb genomic segment containing *Bmp4*, transgenic mice derived from two overlapping BAC reporter transgenes were generated and analyzed for reporter expression. Our findings indicate multiple tissue-specific enhancers such as lung, heart, bone, brain, and vasculature reside greater than 28kb 5' or 3' to the mouse *Bmp4* transcription unit. To test the sufficiency of each ancient ECR to direct tissue-specific expression, reporter constructs were tested in vivo. These experiments suggest 2/3 ECRs reliably function as tissue-specific enhancers in fish. In addition, we are testing the requirement of each ECR to direct tissue-specific expression of *Bmp4* in mouse by deleting each ECR from reporter BAC transgenes and testing deletion BACs for reporter activity in vivo. In sum, these experiments shed light on the poorly understood regulatory landscape of *Bmp4* and elucidate the enhancer potential of ancient, long-range noncoding sequences flanking *Bmp4*.

COMPARATIVE SEQUENCE ANALYSIS**ORAL PRESENTATION****MONDAY NOVEMBER, 13****9.15AM - 9.30AM****O2****ANALYSIS OF LARGE-SCALE VARIATION AND SEGMENTAL DUPLICATION IN THE MOUSE GENOME**

D Church, X She, G Cheng, J Cherry, M DiCuccio, W Hlavna, Y Kapustin, P Meric, E Eichler, D Maglott
University of Washington, Seattle, WA United States

The reference assembly for mouse (based on C57BL.6J) in Build 36 (Feb, 2006 data freeze) is composed of largely BAC based finished sequence with less than 2% of the assembly coming from draft or whole genome shotgun (WGS) sequence. It is clear that the remaining 105 gaps will need considerable effort to reach closure.

It is clear from studies in human that there is a strong correlation between segmental duplication and large-scale genomic variation. Assessment of the reference assembly suggests that at least 4% of the mouse genome is duplicated, twice the amount estimated from the initial draft assembly. The chromosome distribution pattern in mouse shows a preponderance of intra-chromosomal duplication. Duplicated regions in human are gene rich and over-represented for genes involved in environmental response. We are currently assessing the gene content of the duplicated regions in mouse.

As the level of duplication in mouse is approaching that of human, it is likely that large-scale genomic differences are also present. Anecdotal data already exists from the breeding of inbred strains with regions of recombination suppression suggesting inversion polymorphisms. Despite the dearth of finished sequence from other strains, alignment of the sequence from these strains to the reference uncovers regions of genomic polymorphism. Current data suggests that at least 10% of the loci we have identified have some sort of large-scale variation, many of these involving genes. This has important implication for large-scale mapping studies as well as ongoing genome-wide analysis.

COMPARATIVE SEQUENCE ANALYSIS

ORAL PRESENTATION

MONDAY NOVEMBER, 13

9.30AM - 9.45AM

O3

TRANS-ACTING GENES CONTROLLING THE RECOGNITION AND ACTIVITY OF RECOMBINATION HOTSPOTS

K Paigen, K Sawyer, PM Petkov

The Jackson Laboratory, Bar Harbor, ME, United States

Mammalian meiotic recombination occurs at very limited regions of the genome, termed hotspots, which are less than 2 kb in size, vary over several orders of magnitude in their activity, and are separated by long stretches of no recombination. In hotspots, recombination initiates as double-strand breaks which are either processed to generate crossover chromatids or repaired using the homologous chromosome sequence as a template, producing gene conversion.

Although the functions of several key proteins required for all recombination have been relatively well studied, for example Spo11, Msh4 and Mlh1, little is known about the factors determining hotspot recognition and relative activity.

Here we present evidence for the existence of a new class of recombination factors which are coded by trans-acting genes that direct hotspot placement and activity in an allele specific manner. We compared the recombinational properties of the distal 90Mb region of Chr 1 in a congenic cross (B6xB6.CAST-1T) with its interstrain counterpart (B6xCAST). This comparison allowed us to determine how the lack of CAST alleles over the rest of the genome in F1 hybrids of the congenic cross influenced the activity of specific hotspots.

The overall recombination rate over the entire 90-Mb region was similar in the two crosses, and many hotspots were present in both. However, there were also notable differences. Several hotspots highly active in the B6xCAST cross disappeared completely in the congenic cross; recombination appeared in a region that was devoid of recombination in the interstrain cross, and there were regions showing significant differences in recombination rates.

These data clearly reveal the influence of trans-acting factors in determining the locations and activity of recombinational hotspots, and open the possibility of mapping and cloning the relevant genes using sperm assays of hotspot activity as the critical phenotype.

COMPARATIVE SEQUENCE ANALYSIS**ORAL PRESENTATION****MONDAY NOVEMBER, 13****9.45AM - 10.00AM****O4****MOUSE CHROMOSOME X: HOW IT COMPARES TO HUMAN**

L G Wilming, J P Almeida, C Amid, S Donaldson, A Frankish, R C Gibson, E A Hart, K Kivinen, G K Laird, J E Loveland, J M Mudge, A O Mujica, J Rajan, H K Sehra, C A Steward, M-M Suner, M Thomas, J L Harrow
Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

As one of the mammalian sex chromosomes, X has always been a chromosome attracting a great deal of interest. Chromosome X contains various tumor and testis expressed genes and is linked to a disproportionately large number of diseases. The chromosome is also interesting for its evolutionary history, *i.e.* its somatic origin and recombination with chromosome Y.

At the Wellcome Trust Sanger Institute we have now manually annotated the complete chromosome X sequence of both mouse and human. Here I will present a selection of annotated features from mouse X and the comparison of mouse X with its human counterpart. Highlights include the Xlr and Xmr gene families in mouse, which are completely absent from human, and the RhoX cluster of reproductive homeobox genes which has expanded considerably in mouse.

COMPARATIVE SEQUENCE ANALYSIS

ORAL PRESENTATION

MONDAY NOVEMBER, 13

10.00AM - 10.15AM

O5

A GENOMIC APPROACH TO IDENTIFY CIS-ACTING TRANSCRIPTIONAL REGULATORY ELEMENTS

SK Loftus¹, A Antonellis², L Umayam², NISC Comparative Sequencing Program³, TG Wolfsberg², ED Green², WJ Pavan¹

¹Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, MD, United States, ²Genome Technology Branch, National Human Genome Research Institute, NIH, Bethesda, MD, United States, ³NIH Intramural Sequencing Center, Rockville, MD, United States

The neural crest is a population of stem cells that give rise to diverse cell lineages including melanocytes, endocrine cells, craniofacial cartilage, glia and schwann cells. The mechanisms controlling crest development are complex and require precise transcriptional regulation. We are combining gene expression pattern studies, comparative genome sequencing and informatics analysis to identify cis-acting transcriptional regulatory elements required for neural crest development. We have selected 14 neural crest-expressed genes and are generating a detailed analysis of gene expression, using whole mount *in situ* hybridization in wild type and neural crest transcription factor mutant mouse embryos. The image-based gene expression patterns will be converted into a qualitative "heatmap" of the mouse embryo using 9 independent regional assessments defining each expression pattern. These data will allow for sub-grouping of genes based upon: 1) variations in spatial expression patterns in normal embryos; 2) variations in stage of neural crest development; and 3) variations in response to mutations in neural crest transcription factors. In order to identify key transcriptional regulatory elements contributing to the expression patterns defined above, non-coding sequence from the 14 neural crest-expressed genes will be searched for elements that are conserved throughout evolution. For all 14 genes, we have isolated BACs from seven different species and generated high-quality finished sequence. We are using multiple algorithms including ExactPlus and WebMCS to identify evolutionarily conserved sequence elements and are developing informatic approaches to assess if correlations exist between conserved sequence elements and subclasses of genes defined by expression indices.

**COMPARATIVE SEQUENCE ANALYSIS
ORAL PRESENTATION****MONDAY NOVEMBER, 13 10.15AM - 10.30AM****O6****GENOME WIDE PHYLOGENETIC ANALYSIS AND PATTERNS OF DIVERSITY IN SEQUENCED MOUSE STRAINS**H Yang¹, T A Bell², G Churchill¹, F Pardo-Manuel de Villena²¹Jackson Laboratory, Bar Harbor, ME, United States, ²UNC-CH, Chapel Hill, NC, United States

Three years ago the NIEHS entered into a contract with Perlegen Sciences for the genome-wide resequencing of the following 15 inbred strains DBA/2, A, 129/S1, AKR, BALB/C, C3H, FVB, NOD, BTBR, KK, NZW, CAST, MOLF, PWD and WSB. The last four strains were selected to maximize genetic diversity by including a wild-derived strain from each of four subspecies of the *Mus musculus* genus. This project is coming to fruition offering the first opportunity to analyze the patterns of genetic variation found in laboratory mice within a solid evolutionary context. The most recent release spans 1,346 Mb and identifies 5,556,401 SNPs. The false positive rate is very low (1.4%). However, the false negative rate is high (66%) and it is strongly biased against SNPs with low minor allele frequency. This situation leads to the under-representation of SNPs from highly divergent lineages that are represented by a single sample in the NIEHS set and should be taken into account in phylogenetic and evolutionary analyses. We have used the WSB, PWD and CAST strains to determine the subspecific origin (i. e., domesticus, musculus and castaneus, respectively) of each genomic region in each the remaining 12 strains. This analysis confirms directly that the majority of the genome of common laboratory strains is of domesticus origin. The subspecific origin can be visualized at the Center for Genome Dynamics website. Surprisingly, our analysis also demonstrates that MOLF, though to be a pure molossinus strain, has SNPs defining extended haplotypes of domesticus origin spanning 25% of its genome. Finally, we are extending these analyses to detect positively selected genes in the genomes of the three major *Mus musculus* subspecies.