

— Keynote Address —

## The New Phase of Transcriptome Analysis

**Yoshihide Hayashizaki**

yoshihide@gsc.riken.jp

Genome Exploration Research Group, Riken Genomic Sciences Centre, Riken Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan

### Abstract

We have established a large-scale system named CAGE (CAP-based analysis of gene expression), for identifying the 5' Transcription Start Sites (TSS) and promoter regions. With this system we have obtained over 10,000,000 CAGE tags from human and mouse. We have also determined the sequences of more than 100,000 full-length cDNAs from mouse, which were subsequently used to study the transcriptional landscape in mammals. From this large data set, the 5' and 3' boundaries of 181,047 transcripts with extensive variations arising from alternative promoter usage, splicing and polyadenylation, were identified. Subsequent genomic mapping of the transcriptome reveals transcriptional forests, with overlapping transcription on both strands, separated by deserts in which few transcripts are observed. Additional complex transcriptional genomic regions were observed, named "chains", possessing alternative forms and overlapping transcripts. As a summation, there are 16,247 new mouse protein-coding transcripts, including 5,154 encoding novel proteins. Also including new sense-antisense transcripts, 36,372 cis- and trans-antisense events in full-length cDNAs, 1,457 chains, 1,499 "gene fusions" and non-coding RNA.

Our CAGE tag method allows us to quantitatively analyze promoter usage in different tissues, revealing that differentially regulated alternative TSSs are a common feature in genes. The data permits genome-scale identification of tissue-specific promoters and analysis of associated cis-acting elements.

These data provide a comprehensive platform for comparative analyses of mammalian transcriptional regulation in differentiation and development.