

September 5, 2012

Comprehensive transcriptome analysis of human ENCODE cells

Transcription starting sites identified using Cap Analysis of Gene Expression (CAGE) technique provide powerful data sets for delineating functional elements across the human genome

ENCODE, an international research project led by the National Human Genome Research Institute (NHGRI), has produced and analyzed 1649 data sets designed to annotate functional elements of the entire human genome. Data on transcription starting sites (TSS) contributed by a research team at the RIKEN Omics Science Center provided key anchor points linking the epigenetic status of genes observed at the 5' end directly to their RNA output.

The ENCODE (Encyclopedia of DNA Elements) project aims to delineate all functional elements encoded in the human genome. Thirty-two institutes from five countries have contributed to the project, each providing their own unique technologies and expertise. The project has developed methods and performed a large number of sequence-based studies mapping functional elements including RNA transcribed regions, protein-coding regions, transcription-factor-binding sites, chromatin structure, and DNA methylation sites.

A team of researchers at the RIKEN Omics Science Center led by Dr. Piero Carninci contributed to the mapping of RNA transcribed regions through their identification of TSSs using RIKEN's original CAGE technique. Subcellular compartments (whole cells, nuclei, and cytosol) from 15 cell lines were fractionated before RNA isolation. For one particular cell line (K562), further fractionation was performed to obtain chromatin, nucleoplasm, and nucleoli.

Isolated RNAs were then divided depending on their length, and long RNAs were further fractionated into polyadenylated and non-polyadenylated long RNA's. Each of the RNA fractions were then characterized for function analysis.

The data set was integrated with data sets provided by other research groups for further analysis, which included modeling transcription levels from histone modification/transcription factor-binding patterns and prediction of transcription activities at distal enhancer regions. Overall, this comprehensive data, together with other data sets, contributed to assigning biochemical functions for 80% of the human genome, particularly in areas outside of well-studied protein-coding regions. Another striking result is the pervasive presence of lowly-expressed RNA transcripts, whose localization is restricted to the cell nucleus.

“Scientists at the RIKEN Omics Science Center are particularly pleased with this work because the CAGE technology, developed earlier, was employed as one of the standard technologies for analyzing the output of the genome,” Dr. Carninci said. “This international collaboration is in line with the OSC mission to understand the function of the genome. OSC has pioneered the field with the Fantom project, which provided a first comprehensive annotation of the mouse and human genome using CAGE, and identified a transcriptional network that controls the cell fate. The current ENCODE dataset provides a comprehensive set of data that strengthens and complements our previous and current work, aimed at understanding the function and regulation of the genome in health and disease states. OSC is committed to further characterize the genome output for much larger number of cells.”

For more information, please contact:

Dr. Piero Carninci
LSA Technology Development Group
RIKEN Omics Science Center
Tel: +81-(0)45-503-9222 / Fax: +81-(0)45-503-9216
Email: carninci@riken.jp

Global Relations Office
RIKEN
Tel: +81-(0)48-462-1225 / Fax: +81-(0)48-463-3687
Email: pr@riken.jp

Reach us on Twitter: [@rikenresearch](https://twitter.com/rikenresearch)

References

1. The ENCODE Project Consortium, "An integrated encyclopedia of DNA elements in the human genome", *Nature*, 2012 doi: 10.1038/nature11233
2. S. Diebali., et. al, "Landscape of transcription in human cells", *Nature*, 2012 doi: 10.1038/nature11247

About RIKEN

RIKEN is Japan's flagship research institute devoted to basic and applied research. Over 2500 papers by RIKEN researchers are published every year in reputable scientific and technical journals, covering topics ranging across a broad spectrum of disciplines including physics, chemistry, biology, medical science and engineering. RIKEN's advanced research environment and strong emphasis on interdisciplinary collaboration has earned itself an unparalleled reputation for scientific excellence in Japan and around the world.

About the Omics Science Center

Omics is the comprehensive study of molecules in living organisms. The complete sequencing of genomes (the complete set of genes in an organism) has enabled rapid developments in the collection and analysis of various types of comprehensive molecular data such as transcriptomes (the complete set of gene expression data) and proteomes (the complete set of intracellular proteins). Fundamental omics research aims to link these omics data to molecular networks and pathways in order to advance the understanding of biological phenomena as systems at the molecular level.

Here at the RIKEN Omics Science Center, we are developing a versatile analysis system, called the "Life Science Accelerator (LSA)", with the objective of advancing omics research. LSA is a multi-purpose, large-scale analysis system that rapidly analyzes molecular networks. It collects various genome-wide data at high throughput from cells and other biological materials, comprehensively analyzes experimental data, and thereby aims to elucidate the molecular networks of the sample. The term "accelerator" was chosen to emphasize the strong supporting role that this system will play in supporting and accelerating life science research worldwide.