



The RIKEN FANTOM4 Project Paves the Way to Human Gene Network Models

-New Applications for DNAFORM DeepCAGE and short RNA Services-

Our approaches to life sciences are founded in hypothesis-driven studies. The suggested explanation of observable phenomena formulated in any hypothesis, however, requires extensive experimental evaluation to prove or disprove the hypothesis. Although elementary to our intellectual conceptions, life sciences are moving in new directions from “hypothesis-driven studies” to an age of “predictive science” based on systems biology. It is envisioned that predictive sciences could one day lead to engineered designs of new therapies dramatically reducing time and cost for drug development.

A thorough understanding of the regulatory mechanisms underlying biological processes should allow us to accurately predict how a cell, a tissue, or even an organism will respond to environmental signals. For future medical research, data-intensive computing will be required to interpret the massive data sets needed to build reliable models suitable of describing biological processes. At the forefront of such developments, scientists have started to decipher regulatory networks controlling gene expression in lower organisms to handle the complexity of data collection and interpretation.

During the FANTOM4 Project¹, an international consortium under the leadership of the RIKEN Omics Center has for the first time taken on the task of establishing a gene network model for a human cell line, monocytic THP-1 cells. Before attempting this ambitious goal, they had to develop new methodologies for unsupervised genome-wide expression profiling, and directly linking such expression profiles to promoter activities driving gene expression. Accurately measuring promoter activities at defined locations in the genome was the starting point to building a computational model around regulatory elements found in the promoter regions. All results from the experiments and computer model are freely available at RIKEN’s new FANTOM4 web resources (<http://fantom.gsc.riken.jp/4/>)^{2,3}.

The combination of the Cap Analysis of Gene Expression (CAGE) method⁴ for genome-wide digital expression profiling and promoter identification with new next-generation sequencing approaches (also denoted as the DeepCAGE method⁵) provided the in-depth data sets on which the new FANTOM4 gene network is based. DeepCAGE is a powerful technology that is uniquely positioned to elucidate regulatory mechanisms in biological studies. As RIKEN has shown,

DeepCAGE data and computational models derived thereof can be correlated with other experimental data, namely, that obtained by shortRNA analysis, knockdown experiments on transcription factors, chromatin IP, microarray, and protein-protein interactions. In particular, the analysis of shortRNAs during the FANTOM4 Project led to the identification of new RNAs, denoted as “transcription initiation RNAs” or “tiRNAs”, which are associated with RNA polymerase II binding at transcription start sites⁶.

In addition to providing the essential dataset for describing the regulatory gene network, DeepCAGE experiments from the FANTOM projects revealed for the first time transcriptional activities at repeat elements on a genome-wide scale⁷. These observations demonstrate the great power of DeepCAGE in the analysis of rare transcriptional events potentially of great biological importance, here, leading to the identification of 23,000 candidate regulatory regions derived from retrotransposons. Since there are more than 4.3 million repetitive elements in the human genome, their analysis offers enormous promises for the discovery of new markers for medical research.

DNAFORM is pleased to offer DeepCAGE and shortRNA library services giving our customers access to the latest RIKEN technologies for dissecting biological processes. Methods developed for the RIKEN FANTOM4 project provide a powerful way to analyze regulatory pathways relevant to any study. Contact us for more information on DeepCAGE and shortRNA services and on how DNAFORM can take your research to the next level on the way to predictive science.

References:

1. FANTOM Consortium et al., The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet.* 2009 May;41(5):553-62. Epub 2009 Apr 19.
2. Kawaji H et al., The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. *Genome Biol.* 2009 Apr 19;10(4):R40.
3. Severin J et al., FANTOM4 EdgeExpressDB: an integrated database of promoters, genes, microRNAs, expression dynamics and regulatory interactions. *Genome Biol.* 2009 Apr 19;10(4):R39.
4. Kodzius R et al., CAGE: cap analysis of gene expression. *Nat Methods.* 2006 Mar;3(3):211-22. No abstract available.
5. Valen E et al., Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.* 2009 Feb;19(2):255-65. Epub 2008 Dec 11.
6. Taft RJ et al., Tiny RNAs associated with transcription start sites in animals. *Nat Genet.* 2009 May;41(5):572-8. Epub 2009 Apr 19.
7. Faulkner GJ et al., The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet.* 2009 May;41(5):563-71. Epub 2009 Apr 19.