Genomic Med. (2008) 2:135–136
DOI 10.1007/s11568-009-9121-6

ABSTRACTS

# HGM2008 plenary abstracts: genome informatics to genome biology



## 006: Sequencing revolution and human genome network

**Takashi Gojobori**

DNA Data Bank of Japan, National Institute of Genetics, Mishima 411-8540, Japan; BIRC, National Institute of Advanced Industrial Science and Technology (AIST), Aomi 2-42, Koto-ku Tokyo, 135-0064, Japan

In the Human Genome Network Project, which has been conducted under the Ministry of education, science, culture, and sports (MEXT) of Japan, a huge amount of data for transcription start sites (TSS), protein–protein interaction (PPI), and DNA–protein interaction (DPI) have accumulated so far partly because of the so-called "next generation sequencing machines". All these data are stored in our group at DNA Data Bank of Japan (DDBJ) in the National Institute of Genetics (NIG) in order to construct the database and to provide the users with them. In particular, we successfully mapped to the entire human genome about thirty millions of CAGE segments, which are 20-nucleotide segments from the 5′ of mRNAs that are produced by RIKEN. This mapping revealed the distribution of TSS over the human genome. Moreover, the PPI data newly produced in this project showed that the protein network for each of transcription factors examined can be constituted with high connectivity when we incorporated the publicly available PPI data into our data. The DPI data are also very useful for identifying the locations in the human genome where transcription factors are bound to. Along with the information contained in the integrated human gene database such as the H-Invitational database, which was constructed by our group under supervision of the Ministry of economy, trade, and industry (METI), all these information can be utilized for elucidation of the transcription regulation network in the human genome. In particular, the second and third generation sequencing machines must play a crucial role in not only the efforts of sequencing the complete genomes but also the efforts of elucidating the transcription regulation network. A few years later, I predict that the current time may be remembered as the time of sequencing revolution.

## 007: From protein networks to headache and environment

**Peer Bork**

EMBL Structural and Computational Biology Unit, EMBL, Heidelberg, Germany

Genomes and Metagenomes are sequenced at an ever increasing pace and there are many attempts to connect the derived, one dimensional parts lists, the proteins, to networks in order to provide two dimensional context. Here I want to demonstrate how those networks can be integrated by comparative analysis with a variety of datasets to improve the functional understanding of the biological systems being studied. For example, (1) using STITCH, a resources for protein–chemical interactions, we employ phenotypic side effect similarity between drugs to predict novel drug–targets relations, (2) using STRING, a resources for protein–protein interactions, we integrate cell cycle arrays to discover principles of temporal regulation, and (3) we finally integrate, using iPath, a pathway mapping and visualization tool, metabolic networks derived from different environmental samples to predict spatial metabolic variation of microbial communities at millimetre scale.

## 008: Deciphering biological function from genome variation

**Samir K. Brahmachari**

Council of Scientific and Industrial Research, Anusandhan Bhavan, 2 Rafi Marg, New Delhi 11001, India

Human Genome Project started with little emphasis on non-coding regions and repeats, instead, focusing almost exclusively on finding gene sequences. The current excitement of thousand genome sequencing promises a deluge of data with growing disparity between data and information. Proteins, being the effector molecules of biological function are key to understanding of functional genomics from sequence data. Efforts to establish a correlation between genotype and phenotype has met with limited success specifically for complex disorders due to lack of understanding of the complex regulatory networks that govern phenotypic expression. Concentration of proteins and affinity of their interactions with each other and DNA, modulated by structure, determine the logic of gene regulation at every stage. While models of transcriptional regulation is well-established, subsequent layers of regulation adds a new dimension in understanding biological functions in the post-genome sequencing era. Non-coding RNA molecules offer an elegant mechanism to rapidly and reversibly regulate effector proteins in the cell. We have done genome-wide analysis using computational approaches to identify microRNA-mediated regulation in

biological networks. MicroRNAs can serve as nodal points for post-transcriptional regulation. Feedback inhibition by microRNA in conjunction with transcriptional regulation can provide layers of reinforcing regulatory modules in biological networks. Sequence variation at the genome level translates to structural variation in proteins, alteration of transcription factor binding sites and binding sites of regulatory RNAs, ultimately altering network architecture. The talk will cover how genome variation can affect regulation at each of these levels.