Linking genomic knowledge to diseases and drugs

Minoru Kanehisa

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan and Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan

The large-scale datasets generated by genome sequencing and other high-throughput experimental technologies are the basis for understanding life as a molecular system and for developing medical, industrial, and other practical applications. The key to linking such large-scale datasets to practical values lies in advanced bioinformatics technologies, not only in terms of computational methods, but also in terms of knowledge bases. It is essential to organize our knowledge on higher-level systemic functions in a computable form, so that it can be used as a reference for bioinformatics analysis to link from genomic and molecular-level information. We have been developing the KEGG database resource (http://www.genome.jp/kegg/), an integrated resource of genomic, chemical, and systemic functional information. The main component is the KEGG PATHWAY database consisting of manually drawn graphical diagrams of molecular networks, called pathway maps, and representing metabolism, signaling, various other cellular processes, and selected human diseases including cancers, immune disorders, and neurodegenerative diseases. Disease pathway maps present detailed pictures of molecular networks involving causative genes, but it is often the case that our knowledge is too fragmentary. Molecular lists are less detailed but more general and still computable representation, and KEGG DISEASE is being developed for all human diseases with known genetic factors. KEGG DRUG is a chemical structure based information resource for all approved drugs in Japan and the U.S.A. and some herbal medicines in Japan, together with drug structure maps representing our knowledge on chemical structure transformations in drug development. Here I will first describe the genome annotation of over 1000 complete genomes in KEGG and the process of pathway mapping and BRITE mapping for inferring higher-level systemic functions. I will then focus on our attempts to uncover disease mechanisms and druggable natural products by utilizing the KEGG resource.