

## REPRESENTATION AND ANALYSIS OF MOLECULAR NETWORKS INVOLVING DISEASES AND DRUGS

MINORU KANEHISA<sup>1,2</sup>

kanehisa@kuicr.kyoto-u.ac.jp

<sup>1</sup>*Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan*

<sup>2</sup>*Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan*

### Abstract

Twenty years ago the Human Genome Project was initiated aiming to uncover the genetic factors of human diseases and to develop new strategies for diagnosis, treatment, and prevention. Despite the successful sequencing of the human genome and the discovery of many disease related genes, our understanding of molecular mechanisms is still largely incomplete for the majority of diseases. In the KEGG database project we have been organizing our knowledge on cellular functions and organism behaviors in computable forms, especially in the forms of molecular networks (KEGG pathway maps) and hierarchical lists (BRITE functional hierarchies). The computerized knowledge has been widely used as a reference for biological interpretation of large-scale datasets generated by sequencing and other high-throughput experimental technologies. Our efforts are now focused on human diseases and drugs. We consider diseases as perturbed states of the molecular system that operates the cell and the organism, and drugs as perturbants to the molecular system. Since the existing disease databases are mostly for humans to read and understand, we develop a more computable disease information resource where our knowledge on diseases is represented as molecular networks or gene/molecule lists. When the detail of the molecular system is relatively well characterized, we use the molecular network representation and draw KEGG pathway maps. The Human Diseases category of the KEGG PATHWAY database contains about 40 pathway maps for cancers, immune disorders, neurodegenerative diseases, etc. When the detail is not known but disease genes are identified, we use the gene/molecule list representation and create a KEGG DISEASE entry. The entry contains a list of known disease genes and other relevant molecules including environmental factors, diagnostic markers, and therapeutic drugs. The list simply defines the membership to the underlying molecular system, but is still useful for computational analysis. In the KEGG DRUG database we capture knowledge on two types of molecular networks. One is the

interaction network of drugs with target molecules, metabolizing enzymes, transporters, other drugs, and the pathways involving all these molecules. The other is the chemical structure transformation network representing the biosynthetic pathways of natural products in various organisms, as well as the history of drug development where drug structures have been continuously modified by medicinal chemists. KEGG DRUG contains chemical structures and/or chemical components of all prescription and OTC drugs in Japan including crude drugs and TCM (Traditional Chinese Medicine) formulas, as well as most prescription drugs in USA and many prescription drugs in Europe. I will report on our strategy to analyze the chemical architecture of natural products derived from enzymatic reactions (and enzyme genes) and the chemical architecture of marketed drugs derived from human made organic reactions in the history of drug development, towards drug discovery from the genomes of plants and microorganisms.