

From gene variants to network variants: a new database for understanding diseases and drugs

Minoru Kanehisa, Mao Tanabe

Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan

Linking the diversity of genomes to the diversity of organisms in the tree of life has been one of the main objectives of the KEGG database project. We developed the KEGG PATHWAY database as a reference knowledge base for understanding conserved and diverse functions of organisms, which is accomplished by the process of KEGG pathway mapping where genes in the genome are mapped to nodes of molecular networks (KEGG pathway maps). Unfortunately, however, KEGG PATHWAY is not sufficient for understanding the diversity among human genomes, especially in relation to diseases and drugs. KEGG pathway maps for cancers, for example, contain oncogenes and tumor suppressor genes with genetic alterations, which are marked in red but are linked to normal genes. Similarly, gene-disease associations are accumulated in the KEGG DISEASE database, but the details of genetic alterations are not given. Thus, we have started developing a new database named KEGG NETWORK for "perturbed" molecular networks involving human diseases and drugs. KEGG NETWORK is a collection of network elements, which are defined in a way somewhat similar to KEGG modules. Based on published literature we accumulate knowledge on how, for example, signaling pathways in cancer are perturbed by gene variants, viruses and environmental factors. Gene variants are linked to ClinVar, dbSNP, and other databases, so that KEGG NETWORK can be used for interpretation of personal genome sequences.