

Abstract

***MutationView* : An Integrated Knowledge-Base for Variation in Human Disease Genes - Intelligent Search System and Genome-wide Data Construction -**

Ohtsubo, M.¹⁾, Mitsuyama, S.¹⁾, Kawamura, T.¹⁾, Shimizu, N.¹⁾ and Minoshima, S.^{1),2)}

1) Department of Molecular Biology, Keio University School of Medicine, 2) Photon Medical Research Center, Hamamatsu University School of Medicine, Japan

More than 1000 disease-causing genes have been molecularly cloned. In order to better understand correlation between disease and genetic diversity, we have developed an integrated knowledge-base system with graphical environment, *MutationView* (<http://mutview.dmb.med.keio.ac.jp>). The characteristic features of the *MutationView* are as follows: (1) Various data display: genomic/cDNA structure, functional domain of protein, histogram of case number for mutations, changes in the nucleotide/amino acid sequence and restriction sites, (2) Analysis functions: classification based on mutation type, hereditary pattern and symptom. To date, we have collected 9375 entries of mutations from 1648 literatures dealing with 235 genes involved in 217 distinct diseases focused on nine categories of diseases, such as those related to ophthalmology, brain, muscle, otolaryngology, heart, syndrome, autoimmunity and familial tumor.

Currently, *MutationView* is principally focusing on monogenic disease mutations, and molecular biological data are mainly being collected separately for each gene. For more effective use of *MutationView*, systematic knowledge on diseases including symptom and patient/family information as well as various functions of genome-wide aspects are necessary. Here, we report a new version of *MutationView* which has functions to automatically extract disease-associated knowledge from the literature. It extends the scope of *MutationView* to genome-wide by the following two features: (1) A search system using relational index of words in OMIM: Armed with the categorized dictionary for various fields such as clinical medicine, histology and anatomy, significant relations among words are being extracted by statistical analysis based on coincidental appearance in the same section of each OMIM document. (2) Genome-wide data presentation function: Various information such as chromosome band, contig, gene, transcription, and marker covering all the chromosomes was automatically imported from Ensembl. Moreover, gene structure was compiled from 857 disease causative genes.