MutationView : An Integrated Knowledge Base for Mutations and Polymorphisms in Human Disease Genes

- Automatical Extraction of Disease-Associated Knowledge -

Masafumi Ohtsubo[1] ohtsubo@dmb.med.keio.ac.jp
Susumu Mitsuyama[1] mituyama@dmb.med.keio.ac.jp
Takashi Kawamura[1] kawamura@dmb.med.keio.ac.jp
Nobuyoshi Shimizu[1] shimizu@dmb.med.keio.ac.jp
Shinsei Minoshima[2] mino@hama-med.ac.jp

[1] Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
[2] Photon Medical Research Center, Hamamatsu University School of Medicine, 1-20-1 Handayama, Hamamatsu-shi, Shizuoka 431-3192, Japan

Keywords: mutation, disease gene, distributed database, GUI, language analysis

1 Introduction

One of the important purposes of genome research is to elucidate the nature of genes responsible for monogenic as well as multigenic diseases. In order to better understand correlation between disease and genetic diversity, we developed an integrated database system *MutationView*, which at present mainly deals with monogenic diseases (`http://mutview.dmb.med.keio.ac.jp`). *MutationView* is the system which can search, display and analyze the mutation data with graphical environment.

2 Charactercteristics and Data Contents of *MutationView*

The characteristic features of the *MutationView* are as follows (Fig. 1: right): (i) Several ways are available to access to the gene of interest through the chromosomal map of the gene or disease, anatomical chart of disease-associated organ or tissue and diagram of causative gene product. (ii) Various functions for data analysis and display are available such as genomic/cDNA structure of gene, functional domain of protein, zooming in/out of the nucleotide/amino acid sequences, histogram of mutations, case number, changes in the nucleotide sequence and restriction sites, classification based on mode of inheritance and clinical symptom, practical information such as PCR primers and reaction conditions. (iii) *MutationView* system can be linked to global LSDBs while maintaining their independency. (iv) 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. Numbers of genes and typical diseases in current *MutationView* are shown in Table 1.

3 Prototype of New Version of *MutationView* with Novel Functions

For further effective use of *MutationView*, more systematic knowledge on diseases including symptom and patient/family information is necessary in addition to molecular biological data. Here, we report a new version of *MutationView* with novel functions to automatically extract disease-associated knowledge from the literature and to collect/search/display genome-wide data as follows;

Table 1: Contents of *MutationView*, 2003.

KMDB	#Gene	Typical Disease
Eye	71	Retinitis pigmentosa, Glaucoma, Corneal dystrophy
Heart	24	Cardiomyopathy, Heart dysmorphism
Ear	37	Deafness
Brain/Neuron	52	Familial Parkinsonism, Alzheimer disease
Cancer-related	22	Breast, Retinoblastoma, Neurofibromatosis
Syndrome	38	Waardenburg, LongQT
Autoimmune	2	APECED, APT1
Muscle	31	DMD, MD Fukuyama type
Blood	40	CML, Citrullinemia

(i) A search system using relational index of words appearing in OMIM: Armed with the categorized dictionary for various fields such as clinical medicine, histology and anatomy, significant relations of words were extracted by statistical analysis based on coincidental appearance in the same section of each OMIM document. (Fig. 1: left).

(ii) Genome-wide data import from Ensembl: Various information such as chromosome band, contig, gene, transcription and marker, covering all the chromosomes was automatically imported from Ensembl. Moreover, gene structure data was made about 857 disease causative genes.

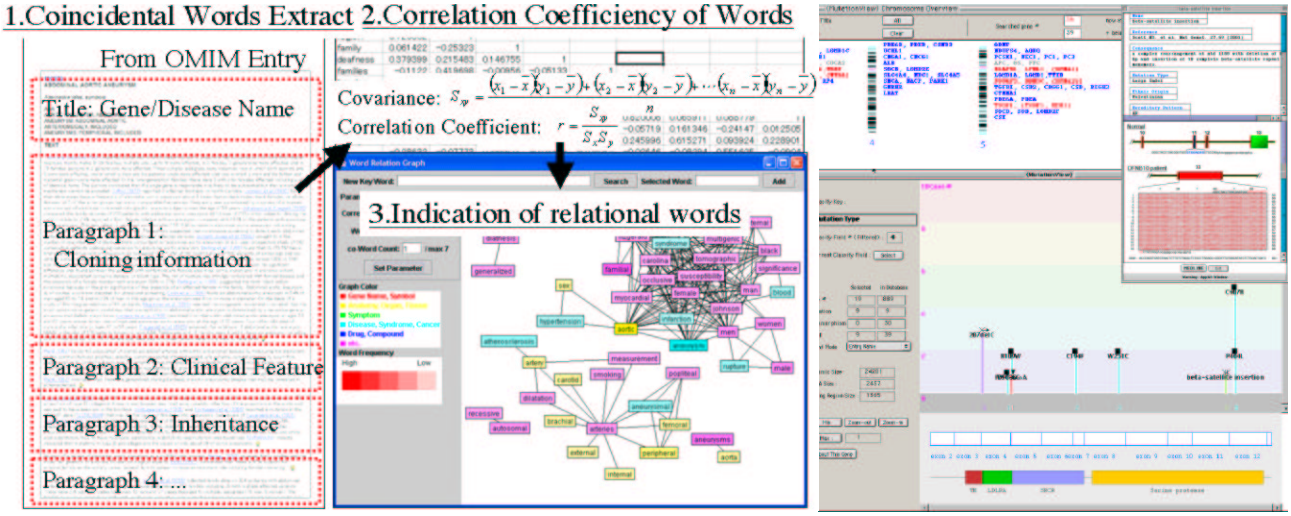

Figure 1: A Concept for Extraction of Disease-associated Knowledge (left) and a typical display of *MutationView* (right).

Acknowledgments

We thank Professor Fumiaki Ito and his colleague of Setsunan University to construct data of cancer-related genes. We also thank Chi Co., Ltd. for their extensive collaboration for coding JAVA applets and servlets. This work was supported in part by a fund for the "Research for the Future" Program from the Japan Society for the Promotion of Science (JSPS) and Ministry of Education, Culture, Sports, Science and Technology (MEXT), and a Grant-in-Aid for Scientific Research on Priority Areas from the MEXT.

References

[1] Minoshima, S., Mitsuyama, S., Ohtsubo, M., Kawamura, T., Ito, S., Ito, F., and Shimizu, N., A new Version of the KMDB/*MutationView*: a Keio Mutation Database for human disease genes, *Genome Informatics*, 11:224–226, 2000.

[2] Minoshima, S., Mitsuyama, S., Ohtsubo, M., Kawamura, T., Ito, S., Ito, F., and Shimizu, N., The *KMDB/MutationView*: a mutation database for human disease genes, *Nucleic Acids Res.*, 29:327-328, 2001.