

Research and Education for Biomedical Informatics at Tokyo Medical and Dental University

Hiroshi Tanaka, Jun Nakaya

School of Biomedical Sciences, Information Center for Medical Sciences, Tokyo Medical and Dental University

Summary

Objectives: Based on a basic concept of “Systems Life Science: understanding life and disease as a unified system”, we move forward in research, empirical implementation, and making contributions to healthcare policy.

Methods: We integrate bioinformatics and medical informatics for identifying critical issues in biological science and solving medical challenges with a concept of “Systems Life Science” which consists of “Systems Evolutionary Biology” for basic science, “Systems Pathology” for clinical sciences, and an empirical medical informatics for future medicine.

Results: Our laboratory is an integrated laboratory consisting of a computational biology group in the School of Biomedical Sciences (SBS), a bioinformatics group in the Medical Research Institute (MRI), and a medical informatics group in the Information Center for Medical Sciences (ICMS) with a philosophy of “Empirical Systems Life Science”.

Conclusions: Based on the philosophy of “Empirical Systems Life Science”, we continue to forward our research, education, systems implementations, and international standardization efforts. We believe that this approach will become a fundamental and effective way to uncover many of the secrets of life processes, and to help/solve complex issues for future medicine in this post genomic era with exceedingly rapidly growing amounts of -omics data and knowledge.

Keywords

Empirical systems life science, systems evolutionary biology, systems pathology, systems medical informatics, translational research informatics

Geissbuhler A, Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2007. *Methods Inf Med* 2007; 46 Suppl 1: 157-62

1. Basic Concept of our Research

The recent explosive expansion of data and knowledge in the biomedical fields, especially relating to comprehensive molecular information such as genomic, proteomic and other kinds of “omics-data”, makes informatics approaches crucial in exploring and clarifying the secrets of life as well as their contributions in resolving wide-ranged medical issues arising from clinical practice and healthcare management. We have been engaged both in bioinformatics and medical informatics for identifying critical issues in biological science and solving medical challenges, from the standpoint of a basic concept of “**Systems Life Science: understanding life and disease as a unified system**”. From this integrated concept, we have moved forward in research, empirical implementation, and making contributions to healthcare policy.

1.1 Basic Sciences: Systems Evolutionary Biology

Considering life as an “evolving network composed of biological molecules”, we explore the secrets of life and in an effort to move towards an overarching theory for the biological sciences, which we call “Systems Evolutionary Biology”.

1.2 Clinical Sciences: Systems Pathology

Based on the concept of “Systems Pathology” based on much -omics background, we try to resolve issues arising in medical science and clinical practice. This concept is derived from the fact that each disease can be seen as a kind of system with a hierarchical aggregation of networks from the molecular level to that of the individual organism. We believe that our concept of “Systems Pathology: understanding a disease as a system” can be a key theory in helping develop molecular medicine or personalized medicine.

1.3 Medical Informatics for Future Medicine

We gather and combine all of our research to make use of them practical and effective as Systems Medical Informatics. From the standpoint of policy making and construction of the social-infrastructure for better healthcare into the 22nd century, we research and implement socially-required medical informatics technologies such as the Electronic Health Record (EHR), Ubiquitous IT for Medicine, global interoperable technologies, database integration, in cooperation with international standardization efforts at SDOs like ISO/CEN/HL7. Prof. Hiroshi Tanaka as president of Japan Healthcare IT Initiative is responsible for policy

making, the introduction, and direction of Medical IT in Japan.

Our laboratory is an integrated laboratory consisting of a computational biology group in the School of Biomedical Sciences (SBS), a bioinformatics group in the Medical Research Institute (MRI), and a medical informatics group in the Information Center for Medical Sciences (ICMS) with a philosophy of “Empirical Systems Life Science”. Prof. Hiroshi Tanaka conducts research and educations as Dean of Biomedical Science for the PhD program in SBS in conjunction with MRI, and as Director of ICMS.

SBS was established in 2003 and is the only postgraduate school engaged in the education and research of post genome medicine in Japan. Students taking courses of Biomedical Science PhD Program attend lectures and experimental practice during the first 3 months, and in the remainder of the program conduct practical research in the interdisciplinary area of medicine and bioscience. Research at SBS is assisted and advised not only by the staff of SBS but also by the researchers in MRI and Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University (TMDU), as well as by the staff of allied research institutions outside TMDU. The number of the staff in SBS is far larger than that of students. The curriculum is wide and balanced, including the practice of genetic engineering. The lectures and practice cover both basic and advanced topics so that students not majoring in modern bioscience can also conduct cutting edge research. We accept students having an undergraduate education in medicine, biology, bioscience, chemistry, pharmacology, information and systems engineering.

MRI was established in 1973 as an integrated research institute having 17 departments for dealing with issues in the medical sciences. The current re-

search focus of MRI is to tackle issues in medical science with a hope to contribute to develop treatments for patients suffering from intractable diseases. These disorders include metabolic, neurological, psychiatric, cardiovascular, loco-motor, immunological, genetic, infectious, and neoplastic ones.

ICMS was established in 1995 and is the only information processing center having a postgraduate school engaged in education and research in medical informatics. Students taking the program of Biomedical Information Science attend lectures and have experimental practice, and in the remainder of the program conduct practical research in the interdisciplinary area of clinical practice, information science, bioinformatics, and medical informatics.

2. Research and Education Activities

2.1 Biological Science - Systems Evolutionary Biology

2.1.1 Evolutionary Dynamics of Olfactory Receptor Genes in Vertebrates

Mammals have over 1000 olfactory receptor (OR) genes, which form the largest known multigene family. To understand the evolutionary dynamics of OR genes in vertebrates, we conducted a phylogenetic analysis of all OR genes identified from the genome sequences of zebrafish, pufferfish, frogs, chicken, humans, and mice. The results suggested that the most recent common ancestor between fishes and tetrapods has at least nine ancestral OR genes, and all OR genes identified were classified into nine groups each of which originated from one ancestral gene. Eight out of the nine group genes are still observed in current fish species, while only two group (α and γ) genes were

found from mammalian or avian genomes with a few exceptions, showing that the OR genes family in fishes is much more diverse than that in mammals or birds. In mammals or birds, group γ genes are nearly absent in fishes, while four major group genes present in fishes are completely absent in mammals or birds. The expansion of group γ genes also occurs in frogs, but frogs retain the group genes that are abundant in fishes, indicating that the frog OR gene family has both mammal-like and fish-like characters. These observations can be explained by environmental change that organisms experienced after the divergence between fishes and tetrapods. This article appeared on the cover of Proceedings of the National Academy of Sciences of the US[1,2], and was introduced in Editor's Choice in Science [3].

2.1.2 The Missing Link in the Evolution of Hox Clusters

The goal of this research is to understand evolution from the viewpoint of computational systems evolution (systems evolutionary biology). We especially targeted developmental systems: e.g., the Hox signaling system. Hox cluster has key roles in regulating the patterning of the antero-posterior axis in a metazoan embryo. It consists of the anterior, central and posterior genes: the central genes have been identified only in bilaterians, but not in cnidarians, and are responsible for archiving morphological complexity in bilaterian development. However, their evolutionary history has not been revealed (missing link) (Figure 1). We showed the evolutionary history of Hox clusters of 18 bilaterians and 2 cnidarians by using a new method, “motif-based reconstruction”, examining the gain/loss processes of evolutionarily conserved sequences, “motifs”. We recon-

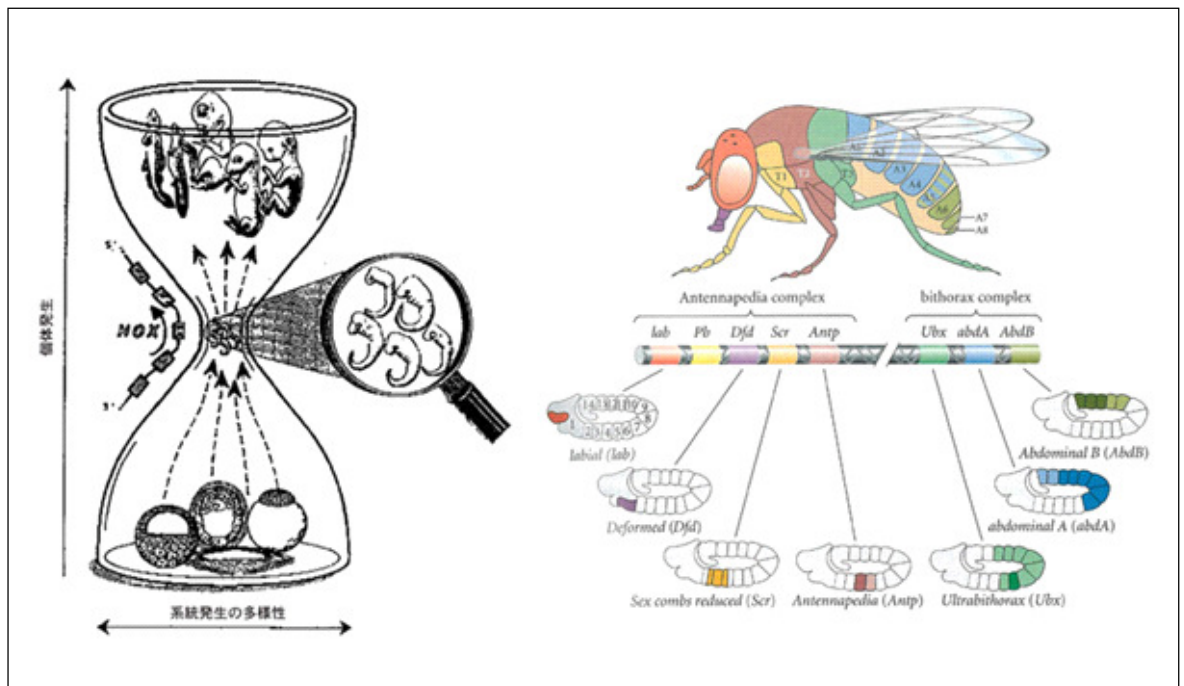


Fig. 1 Systems evolutionary biology to elucidate missing link in Hox clusters' evolution

structed the evolutionary process of central genes of Hox clusters which leads to the explosive diversity of bilaterian body plan. We identified the acoela Hox cluster corresponds with the missing link in the evolution of the Hox cluster between the cnidarian-bilaterian ancestor and the bilaterians [4].

2.1.3 Evolutionary Mechanism of Protein-protein Interaction Networks (PIN)

The PINs in an organism contain many fully connected sub-networks that are evolutionary conserved (motifs). Conventionally, PINs are thought to emerge by the following evolutionary mechanism: A randomly selected node duplicates, a link connecting to the replica is removed with a (approximately) constant probability, and a new link is generated between the replica and a randomly selected node. However, this model cannot explain the presence of motifs correctly, because it produces sparse sub-network structures. By assuming that a new link is generated be-

tween a replica and a target (heterodimerization model), the presence of a half of the motifs can be explained. This implies that another mechanism must exist. To uncover the mechanism, we examined the functional similarities among proteins and found that the proteins showing high functional similarities tend to be located two steps apart in the PINs. We therefore assumed that a new link is generated between a replica and its second neighbor and incorporated this mechanism into the conventional model [5]. Numerical analyses showed that our new model could indeed reproduce the motif structures of PINs nearly completely. Therefore, our second-neighbor model could reflect the actual evolutionary mechanism for the formation of PINs.

2.1.4 An Empirical Examination of the Utility of Codon Substitution Models in Phylogeny Reconstruction

Models of codon substitutions have been commonly used to compare pro-

tein-coding DNA sequences and are particularly effective for detecting signals of natural selection acting on the protein. However, their utility in reconstructing molecular phylogenies and in dating species divergences has not been explored. We applied codon models to 106 genes from eight yeast species to reconstruct phylogenies using the maximum likelihood method, in comparison with nucleotide and amino acid-based analyses [6]. The results show that nucleotide-based analysis was efficient in recovering recent divergences whereas amino acid-based analysis performed better at recovering deep divergences. Codon models appeared to combine the advantages of amino acid and nucleotide data and had good performance in recovering both recent and deep divergences. Although the computational burden makes codon models unfeasible for tree searches in large data sets, we suggest that they may be useful for comparing candidate trees.

2.2 Medical Science - "Systems Pathology" and "Omics Medicine"

2.2.1 Clinical Omics Database for Cancer

As a Japanese national project for critical issues [7], we are moving forward to construct a "Clinical Omics Database for cancer" having omics data such as genomic data, transcriptomic data, proteomic data, other omics data, and clinical data in one integrated set. This database is annotated with network relations resulting from bioinformatics technologies like Data-Mining methods and statistical correlation analyses.

Significance of relating and combining omics data with conventional clinical information is in developing a "Personalized Medicine" as part of the new post-genomic medicine age..

We believe that this database can be a basis to move forward "Personalized Medicine" and its transitional process: "Translational Research" from the standpoint of omics based "Systems Pathology" (Fig. 2).

2.2.2 Education of "Bio-Omics-Medical Informatics"

As a Japanese national project for human resource nurturing [8], we are promoting the training of those who can contribute to a "Bio-Omics-Medical Informatics" that is a multidisciplinary academic field consisting of bioinformatics, medical informatics, biomedical statistics, and clinical omics medicine.

In this post-genomic era, omics information derived from many experiments concerning the transcriptome, proteome, and metabolome are accumulating rapidly and expected to reveal mechanisms of disease, giving rise to much novelty in medicine.

On the other hand, based on the concept of systems biology, thenovel methodologies assume diseases usually result from multiple processes involving hierarchically related networks at multiple levels, ranging from those inside of the cell all the way up to the entire individual organism.

The next generation of medical science will be based on these contexts, and omics information introduces a paradigm shift from conventional pathology to whole-omics systems pathology. In this context, for drug discovery and novel medicine, it is essential to nurture the talents of those who have ability to analyze/integrate/develop/understand disease from the standpoint of "Systems Pathology" based on clinical omics information.

2.2.3 Empirical Systems Evolutionary Biology for HIV

By analyzing the evolutionary process of HIV in the patient body that received a HAART (highly active anti-retroviral therapy) with systems pathological approach, we try to predict prognosis of AIDS therapy and to reveal drug-resistance mechanism of it [9]. HIV evolution in HAART patient is more complex than that in ordinary therapy. We have collected protease sample of HIV in HAART patient for 3 years and analyzed them. We found that AIDS patients have plural virus subgroups regularly in the body, and their composition ratio varies according to variety of anti-HIV drugs.

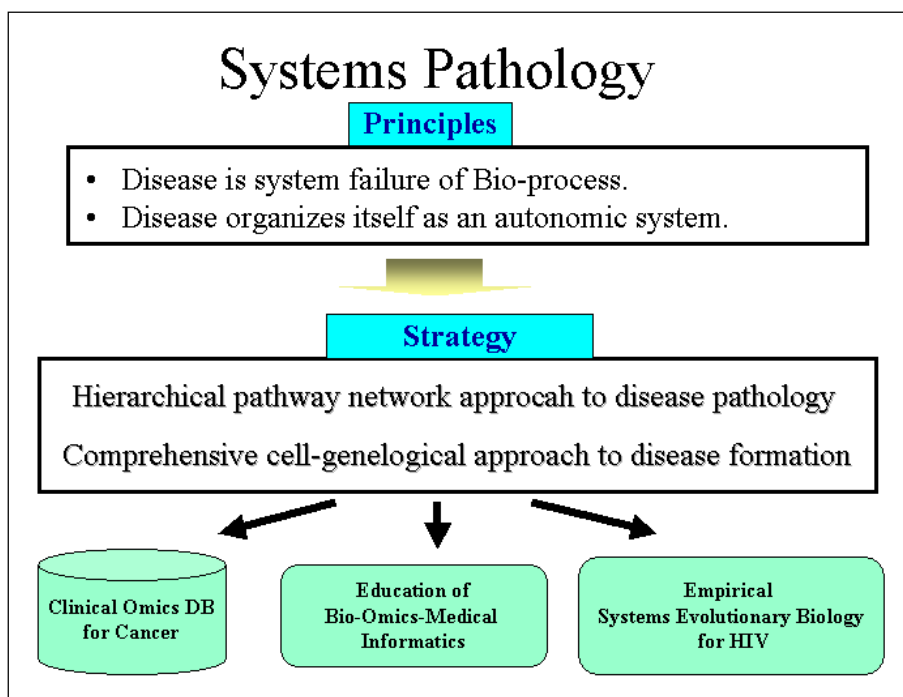


Fig. 2 Basic concept of systems pathology

2.3 Medical Informatics for Future Medicine

2.3.1 Ubiquitous Medical Informatics

To establish a totally computerized clinical practice, not only are information systems with network environments needed, but also hospital spaces themselves need to be computerized. Focusing on compact distributed sensor and electronic tag systems with mobile communication technology, we promote research to make operating rooms, ICUs, hospital wards, outpatient departments, and home care environments computerized as medically intelligent

spaces. Our ubiquitous project is a part of the Japanese national project [10].

2.3.2 International Standardization of Biomedical Informatics Technologies

We are attempting to contribute internationally through promoting international standardization at SDOs such as ISO/HL7/CEN/IHE. Based on outcomes from our research as part of Japanese national projects, we have proposed and have promoted international standardization at SDOs. Currently we are promoting the Genomic Sequence Variation Markup Language (GSVML) with eight countries (US, UK, Canada, Australia, Israel, Italy, Korea, Japan) as an international leader [11,12]. GSVML is a data exchanging format focusing on genomic sequence variation data and human health application domains such as the gene-based medicine and pharmacogenomics (Fig. 3). GSVML is the first international standard in the clinical genomics domain.

GSVML can provide answers specific to the demands for making good use of internationally accumulated genomic sequence variation data resulting from recent genomic research worldwide. We believe that GSVML has the ability to enhance genomic sequence variation data utilization internationally by providing a standardized platform for both clinical and research applications. Developing interfaces to HL7 and CEN, we enhanced the ability of GSVML [13]. Currently we are planning to expand an entry window for GSVML to omics networks.

2.3.3 Biomedical Knowledge Processing, Semantic Web Applications

Based on the concept of “Systems Pathology”, we try to integrate clinical ontologies and biological ontologies. Clinical ontologies such as SNOMED-CT and biological ontologies such as Gene Ontology (GO) are not enough to describe the multi-hierarchical net-

work knowledge existing in the whole omics space [14]. Establishing practical omics ontologies and integrating knowledge with data are urgent demands, and they are also prerequisite to establish -omics related EHRs. Based on an interface analysis between GSVML and SNOMED-CT/en13606 according to an information exchange approach, we have analyzed the fundamental requirements for an integrated ontology. This research is a basis of planning a national integrated -omics database project in Japan.

2.3.4 Translational Research Informatics

Translational Research Informatics (TRI) is the essential informatics to support the translational research (TR) phase that is a part of the early clinical trial phase. TR is believed to be a key pipeline for the success of -omics based medicine. We have analyzed requirements of TRI and defined the required informational technologies. In concrete

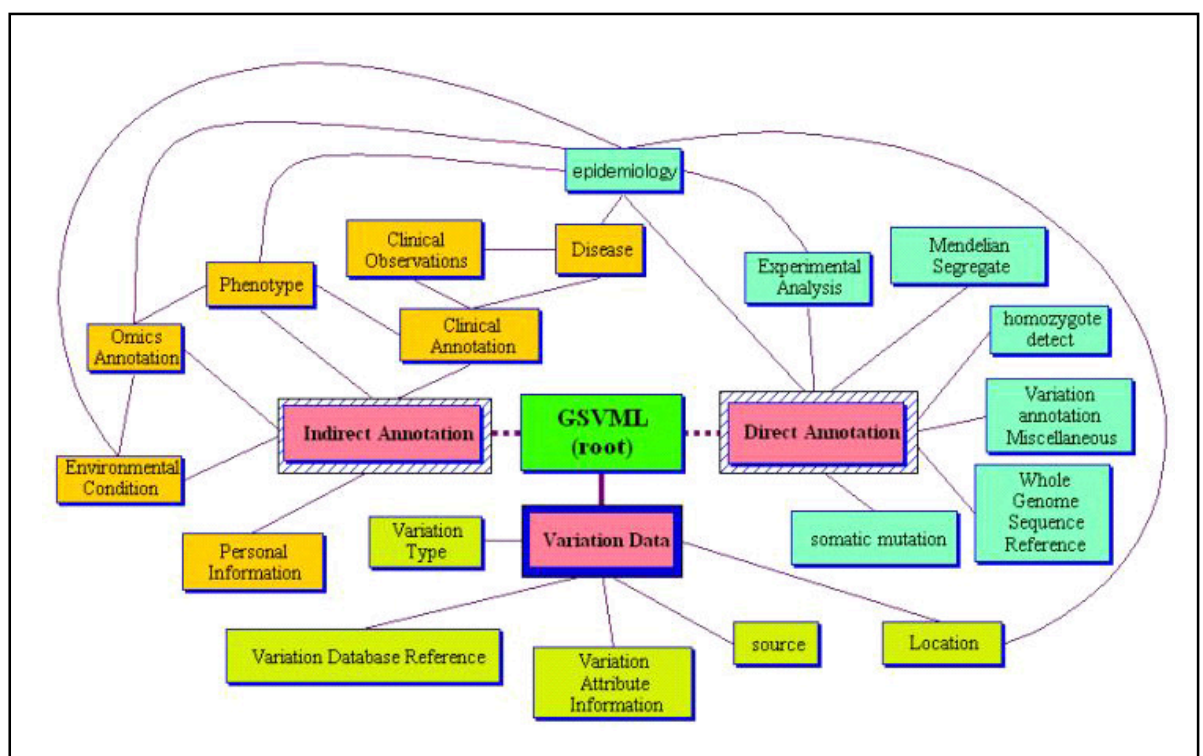


Fig. 3 GSVML as international standard of systems medical informatics in clinical genomics domain

terms, the integration of data and knowledge is the fundamental technology, and the development of knowledge based prediction methods is the principal approach to achieving safe and efficient TR [15]. Currently we are engaged in Japanese national project to establish TRI. We believe that TRI is the critical informatics required for omics-based medicine to become practically successful of in this post genomic era, and we continue to work on it.

Summary and Conclusions

Over 20 staff and 40 graduate students (about 20 master students and 20 PhD candidates) are in our Laboratory, cutting across SBS, MRI, and ICMS in TMDU, pursuing many national projects in multidisciplinary academic fields. Based on the philosophy of "Empirical Systems Life Science", we continue to pursue forward our research, education, systems implementations, and international standardization efforts. We believe that this approach will become a fundamental and effective way to uncover many of the secrets of life processes, and to help solve complex issues for future medicine in this post genomic era with exceedingly rapidly growing amounts of -omics data and knowledge.

References

1. Niimura Y, Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci USA* 2005;102:6039-44.
2. Niimura Y, Nei M. Comparative evolutionary analysis of olfactory receptor gene clusters between



Fig. 4 Staffs and graduate students of our laboratory. The man in the center below is Prof. Hiroshi Tanaka and both sides of him are professors. This photograph shows a part of Tanaka lab.

- humans and mice. *Gene* 2005;346:13-21.
3. International Rice Genome Sequencing Project. The map based sequence of the rice genome. *Nature* 2005;436:793-800.
4. Ogishima S, Tanaka H. The missing link in the evolution of Hox clusters. *Gene* 2006; accepted.
5. Ren F, Tsubota A, Hirokawa T, Kumada H, Yang Z, Tanaka H. A unique amino acid substitution, T126L, in human genotype C of hepatitis B virus S gene and its possible influence on antigenic structural change. *Gene* 2006;383:43-51.
6. Ren F, Tanaka H, Yang Z. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst Biol* 2005, 54:808-18.
7. Construction of exhaustive molecular pathological database, outlays for promoting Japanese Science and Technology, Japan Science and Technology Agency. <http://www.jst.go.jp/EN/index.html>; 2007 [accessed 05.02.06].
8. Bio-Medical-Omics Informatics Education Program, outlays for promoting Japanese Science and Technology, Japan Science and Technology Agency. <http://www.jst.go.jp/EN/index.html>; 2007 [accessed 05.02.06].
9. Hasegawa N, Sugiura W, Matsuda M, Mogushi K, Tanaka H and Ren F. Inference of evolutionary forces driving HIV-1 drug-resistance acquisition under HAART using longitudinal HIV-1 protease gene samples. *Antivir Ther* 2005;S10:114.
10. Experimental use of electronic tags in medical field, outlays for promoting Japanese Science and Technology, Japan Science and Technology Agency. <http://www.jst.go.jp/EN/index.html>; 2007 [accessed 05.02.06].
11. International Organization for Standardization, TC215. Genomic Sequence Variation Markup Language, <http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=43182&scopelist=PROGRAMME>; 2006 [accessed 11.07.06].
12. Nakaya J, Ido K, Hiroi K, Yang W, Kimura M. (Best Paper Award) Genomic Sequence Variation Markup Language (GSVML) for Global Interoperability of Clinical Genomics Data. Asia Pacific Association for Medical Informatics 2006 Proceedings; 2006. p.1-8.
13. Hiroi K, Ido K, Yang W, Nakaya J. Interface Analysis between GSVML and HL7 version3. *J Biomed Inform.* In press 2007.
14. Nakaya J, Shimizu T, Tanaka H. Knowledge Platform to use BioMedical Semantic WEBS individually: Bridging Gene Ontologies and Clinical Ontologies at individual site (Leadoff Article). *International Journal of Computer Science and Network Security.* 2006;6 (2A):1-9.
15. Nakaya J, Shimizu T, Tanaka H. e-protocol - A Translational Research (TR) protocol generator and monitor - . *Communication Network, International Journal of Computer Science and Network Security.* 2006;6 (7B): 18-23.

Correspondence to:

Hiroshi Tanaka
Tokyo Medical and Dental University
Information Center for Medical Sciences
1-5-45 Yushima, Bunkyo-ku
Tokyo, 113-8510 Japan
Tel: +81-3-5803-5839
Fax: +81-3-5803-0247
E-mail: tanaka@cim.tmd.ac.jp