# Knowledge Platform to use BioMedical Semantic WEBs individually: Bridging Gene Ontologies and Clinical Ontologies at individual site

*Jun Nakaya[†], Tetsuo Shimizu[††], and Hiroshi Tanaka[†††]*

*[†]Clinical Genome Informatics Center, School of Medicine, Kobe University, 650-0017 Japan*
*[††]Institute of Medical Science, University of Tokyo, 108-8639 Japan*
*[†††]Tokyo Medical and Dental University, 113-8510 Japan*

**Summary**
Objective: In genomic research with the background of increasing Semantic Networks on Internet, the virtual merging or alignment of gene ontologies (GO) and clinical ontologies (CO) with single scope at each user's location is inevitable consequence to meet individual needs as piecing ontological fragments from Internet or refilling ontological gaps. Methodology: We advocate the knowledge platform that can bridge GO and CO virtually at a client site with three sided basic concepts as (1) the logically extended anatomical index from micro to macro, (2) the knowledge representation based on feature described logical conceptual unit, and (3) the EBM (Evidence Based Medicine) based quality evaluation of knowledge. Results: Scheme and Prototype of the knowledge platform was constructed. Conclusion: The logically extended anatomically hierarchical structure offered the seamless and logical continuity from genome to human. Logical conceptual unit offered unification. EBM based quality assessment enhanced the reliability of knowledge.

***Key words:***
*Semantic WEB, Clinical Ontology, Gene Ontology, Evidence Based Medicine*

## 1. Introduction

The findings based on genomic science and the knowledge derived from genomic research are expected to contribute effectively to clinical medicine in post genomic era. However, the knowledge between genome-scientific domain and clinical domain still has a wide gap. In the real Internet, this knowledge is fragmented and scattered on semantic network as the ontologies like Gene Ontology (GO), Medical Ontology (MO), or Clinical Ontology (CO). These ontologies are expanding and are still scattered on Internet. To refill this wide gap and to piece fragments, a sharable knowledge platform that enables logical integration of GO, MO, and CO is desired. In the context of Internet expansion and its semantic network explosion, the virtual merging or alignment of these ontologies at the personal location with the personal intension is inevitable consequence. On the other hand, the quality of knowledge has many different shades and textures among ontological contents. This situation needs the acceptable and sharable qualification methodology of knowledge. The EBM based knowledge qualification is the most prospective one. This paper proposes the core concept of the sharable knowledge platform that aims to bridge GO, MO, and CO in the context of increasing semantic network on Internet. We also discuss how we should treat ontologies in the Semantic Field. Thus, we brought in prototype that focuses hematology and did its evaluation.

## 2. Background

As indicated in NIH Roadmap [1], the development of the novel tools is important to improve the effectiveness of biomedical research. Here the one of the most important technology mentioned in the roadmap is the technology for the prediction as noted in FDA stagnation/innovation [2]. To establish the prediction technology, the reliable and certified knowledge is needed and the knowledge should be arranged as ontological contents. The knowledge itself and its handling tools will have got to gain an important role to enhance post-genomic research. In the context of the informational explosion over Internet and the

Manuscript revised January 2006.

development of the novel High Throughput Screening technologies, the various contents are developed as ontologies.
Nowadays the integration of these ontologies is getting to be the principal subject, but it is difficult because the scopes or the viewpoints are different among ontologies. On the other hand, the expectation to the knowledge management is increasing as the annotation project or the functional genomics. This leads to a trend that is the use or the reuse of this scattered knowledge on Internet. To reuse many differently scoped ontologies, the virtual window that can scope the knowledge on Internet from end user's point of view is important in the current context as being called in question exit research. Here the end users are the medical experts, then this virtual window should be designed in an easy-to understand way for the medical experts. One of the comprehensive structures for the medical experts is the disease-based structure propounded as GALEN [3] or the anatomical based structure propounded as SNOMED CT [4]. Our structure is the mixed structure of both that means the anatomical structure combined with the disease structure. In this paper, we introduce our basic platform for knowledge and its prototype. Then we also discuss about the concept of the virtual window for the Semantic WEBs as the progress type.
Clinically this platform tries to support the genome targeting translational research. Translational research itself has been popular in cancer research; and the genome-science based trial has started recently [5]. Genome targeting translational research is a practical interdisciplinary research aiming to bridge genomic science and clinical trial. Its clinical endpoints are to improve QOL (Quality Of Life), clinical safety and clinical efficiency. Knowledge is being bridged here and the essence of bridging is bi-directional migration of knowledge between genomic science and clinical trial. The integration of this huge knowledge into an electronic representation on a computer is necessary to establish bi-directional migration and bridging at first. Nevertheless, their integration is not easy because they are multi-disciplinary, fragmented, and they have multi-layered multi-dimensional non-linear relation of cause and effect. Too much and vast background information of knowledge in this Internet society needs IT power to assemble knowledge. The simplified electronic representation and its indexing on computer are important for the integration, the bi-directional migration, bridging and handling the vast information. In this way, this work can be placed in the criteria of the Translational Research Ontology in semantic WEBs.
As concerns applying methods of electronic knowledge technology to the medical field, a tool project based on ontology called PROTÉGÉ [6] is under way. Based on PROTÉGÉ, Stanford is developing the EON system [7] for protocol-based care. Recently the genomic ontological dictionary is constructed as Genetic Ontology (GO) by the gene ontology consortium [8]. The global integration of genomic and clinical knowledge has started as some project, but it has not been achieved yet.
We compartmentalized knowledge as the logical conceptual unit. The logical conceptual unit has its philosophical background in the logical atomism that is a kind of logical system theory [9]. Logical system theory fits to both genomic science and clinical medicine, because these two domains are also kinds of the logical system based on the logical unit. These logical systems need the logical consistency and their reasonability. As for describing the features of the logical unit, we try to describe knowledge only by text with focusing on the verbal side of ontology.
Data mining and text mining are the methodologies for the knowledge collection. Some companies like Adriane Genomics Inc. or Cellomics Inc. distribute tools for these purposes [10]. Data mining is a powerful method, but the accuracy of interpretation is contrary to the quantity of the extracted information. Here by presetting structure of knowledge as a skeleton format enables the deep knowledge formation without losing accuracy of the interpretation. This skeleton format also offers the standardized and normalized representation from molecular world to human world in single manner. Our initial target is to establish integrated hematological knowledge as a prototype. We design knowledge platform for this focused area and try to support genomic translational research clinically and experimentally. Through prototyping in this focused area, we try to make problems simple, to cut out problems clearly, and to makes the concept of the platform practical.

## 3. Methods

### 3.1 Knowledge Platform

Our knowledge platform has a logically extended anatomically hierarchical index (LEAHI) from micro (molecule) to macro (human) as a backbone of whole knowledge (Fig. 1). This knowledge platform also has the evidences that can give the quality criteria of knowledge. The LEAHI bridges genome, human, and environment with single scope that is an extended anatomical view. This single scope gives a logically consistent view to the whole combined knowledge that covers the domain area from the micro anatomy including genome and proteome to the macro anatomy. Knowledge is defined as knowledge unit (KU) at any hierarchy with the same manner. According to LEAHI, KUs are located at their defined positions. The positions are defined at data mining phase. KU is defined as a set of a logical unit (LU) and its supportive evidences. A LU is a minimized conceptual unit at each hierarchical

sub domain and is described with its features. Each feature has its supportive evidences. The supportive evidences are practically links to the original contents. The quality of evidences can be estimated with the modified EBM concept. LEAHI is combined with the scale free network index that can represent non-hierarchical relations like hormone network or scale free network.

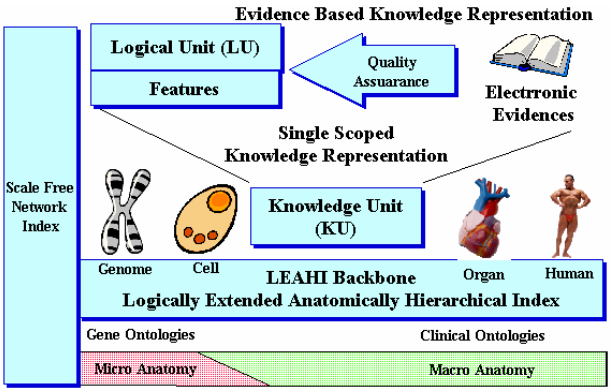


Fig. 1 Knowledge Platform having a Logically Extended Anatomically Hierarchical Index (LEAHI) as a unified backbone of whole knowledge.

## 3.2 Index

This knowledge platform has two indexes internally. The one is logically extended anatomically hierarchical index (LEAHI) of knowledge, and another is the scale free network index. LEAHI is based on the molecular hierarchy, the micro cellular anatomy, and the human anatomical structure. This anatomical view from micro to macro offers single scope and this leads to the continuous, seamless and hierarchical structure from micro to macro level. LEAHI includes DNA, RNA, proteins, micro organella, cell, micro organs as assemblies of cells, organs, individual as assemblies of organs, environment as assemblies of humans. This index can bridge genome, cellome, organome, and phenome as human with single manner. LEAHI locates the KU to the comprehensive position and represents the hierarchical and longitudinal relations of knowledge. We designed LEAHI based on the commonly acceptable knowledge that is almost stable, certain and that is described commonly in many books. Anatomical structure offers the sharable easy understandability between both biologists and clinicians; here they are the domain users. Additionally anatomy based index will not suffer so many changes to its basic structure in the future.

On the other hand "scale free network index" can represent the crosswise relations. The scale free network index is prepared to support subsidiary relations as steroidal network or disease complications. It can represent the molecular network including gene network, disease complication, and any other lateral or scale free relations.

Based on these two indexes, knowledge is represented its relations. These indexes are also used to the data mining to define its classification criteria.

## 3.3 Knowledge Unit and Logical Unit

A domain concept is represented as the knowledge unit (KU). A KU is defined as a set of a logical unit (LU) and its supportive evidences. LU is minimized conceptual unit at each hierarchical domain. The LU is represented with its features that are diversified attributes of the LU. The features are normalized within the belonged hierarchical domain group to avoid repetition. The evidences of features are represented as links to electronic publications practically. These evidences can be the criteria of the quality of knowledge that is described in the features. The quality classification of each evidences are evaluated with an assessment table.

Considering its survivability and taking a stance of the verbal side of ontology, we try to describe the features with text as possible. The diverse representations like image or sound can be also converted to the extracted features that are written with text. In this way, we can say that the features of LU are a kind of the ultimate essence of the human understanding.

## 3.4 Skeleton

We prepared two kinds of skeleton formats. The one is for the domain knowledge of diseases (Fig. 2.), and the other is for the domain knowledge of the molecular science (Fig. 3.). We describe knowledge units of almost all diseases according to this skeleton format. This skeleton format is standardized among all diseases. Applying disease knowledge to this standardized skeleton enables us to normalize disease knowledge with the single mannered description way. This disease skeleton has features like disease name, other name, etiology etc. In the most cases, molecular findings are located at etiology-cause-gene or etiology-cause-protein or in etiology-mechanism. The standardization gives logically consistent description among diseases and we can have the same structure among KUs. This means KUs are comparable with single manner. The KU is tied to LEAHI. The LU is described with features and is a concept of a disease in Fig. 2. The evidences represent the reliability of features and have links to their assisting features. As an example, genomic knowledge of entity is inserted in the clinical knowledge of path like disease-etiology-cause-gene and is supported by evidence linking to Internet publication. This skeleton format is generated after processing third level normalization among all clinical fields. This skeleton has hierarchical expression in its feature description and can cover cross-clinical fields with single standard.

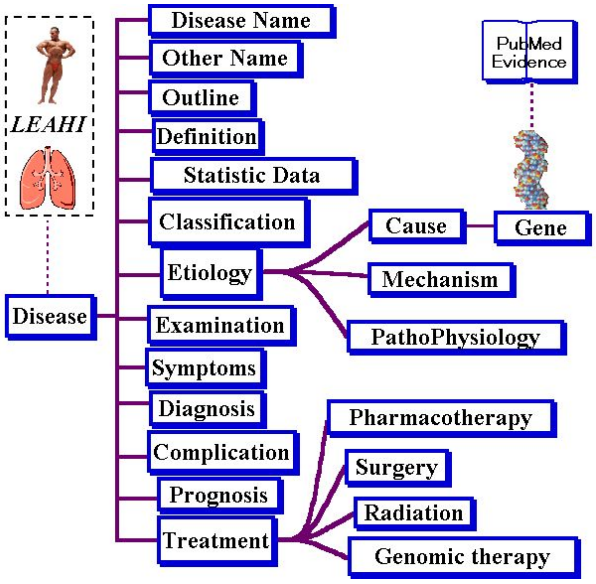


Fig. 2. A part of the skeleton for diseases.

The molecular findings are represented with entity and relation (Fig. 3.). Entity skeleton has features of type, effect, and its character. Relation skeleton has feature of type, degree, direction, and elements. In the molecular world, the KU is each entity or each relation. Each entity is related to the logically extended anatomically hierarchical index (LEAHI) with feature of the "Location" in "Character". Here the molecular entity can be related to phenome disease through LEAHI. The relation will be defined within a hierarchical domain. These categorizations of features are based on theory of elementary process that is described in Feynmann Physics [11].

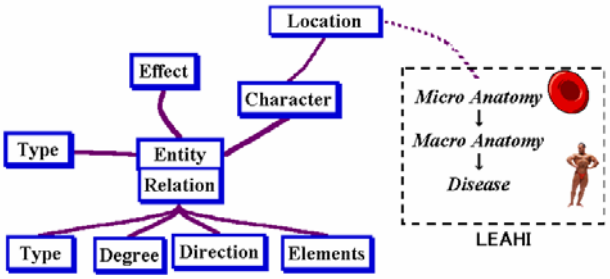


Fig. 3. A part of the skeleton for molecular findings.

## 3.5 Evidences and EBM based qualification table

The evidences for each feature are set with the feature in KU. It represents reliability of each feature of LU. In fact, evidences are links or references to substantiality. Substantial evidences are the publications or the facts in Semantic WEBs on Internet. The criterions of quality evaluation of evidences are summarized in a table (Table. 1.: quoted from [12]). Through this table, we evaluate the quality of evidences. This table is based on the EBM Cochran collaboration design [13]. The criterion for books and sub-minor class are added to make order of evidence numerically clear. The sub-minor class include impact factor and citation index in order to enable continuous numerical evaluation. In this table, the highest qualified evidence has the many sided universality in many books. They are based on the knowledge criteria that are predefined in the client ontological classification table.

Table 1: Quality assessment table. This table is based on EBM cochran collaboration design. We add books and sub-minor class to evaluate seamlessly. In this table, the highest quality evidence is supported by many sided universality in many books (quoted from [12])

| Quality | Major class | Sub-major class | Minor class | Sub minor class |
|---|---|---|---|---|
| High | Book | Many-sided universality | Number of supports | IF, CI |
| | Paper | Randomized Controlled Trial (RCT) | Multiple support | IF, CI |
| | Paper | RCT | Meta analysis over multiple RCT's (conclusions may differ) | IF, CI |
| | Paper | RCT | Single RCT conclusion | IF, CI |
| | Paper | Non Randomized Controlled Trial (non RCT) | Comparative studies, correlation studies etc. | IF, CI |
| | Paper | Quasi experimental studies | Cohort studies, case controlled studies, etc | IF, CI |
| | Paper | Case series, case report, etc | Number of supports | IF, CI |
| Low | WEB site | Opinions from experienced experts | Comments of committee, etc | IF, CI |

## 3.6 Knowledge collection

The knowledge collection in Semantic Networks on Internet is achieved with locating the extracted KU collections to the Logically Extended Anatomically Hierarchical Index (LEAHI) (Fig. 4). Here knowledge is collected with general data mining technology according to the theme derived from LEAHI. The theme is the concrete entity or relation that are described in LEAHI. The figure shows ideal automation process. We use some

data mining tools (Cell Space, Pathway Assist) [14] to collect knowledge parts manually in part as prototype. Each gathered information is diversified into the features according to the skeleton format and the diversified features are stored as the knowledge source with its supportive evidences or reasons in temporary database. The stored features are used to compose the logical unit. Each feature of the logical unit has the links to its evidences or reasons. This means that the knowledge unit is formed as a set of the logical unit and its supportive evidences. According to LEAHI, we arrange and locate the collected knowledge from Semantic WEBs on Internet with data mining technology. Knowledge units are tied to LEAHI according to the preset collection theme. The knowledge resources in Semantic WEBs on Internet are public databases or ontologies as electronic books, PubMed, MEDLINE plus, GenBank, PDB, PIR, Swiss Prot, SCOP, Japana Centra Revuo Medicina Database, etc.

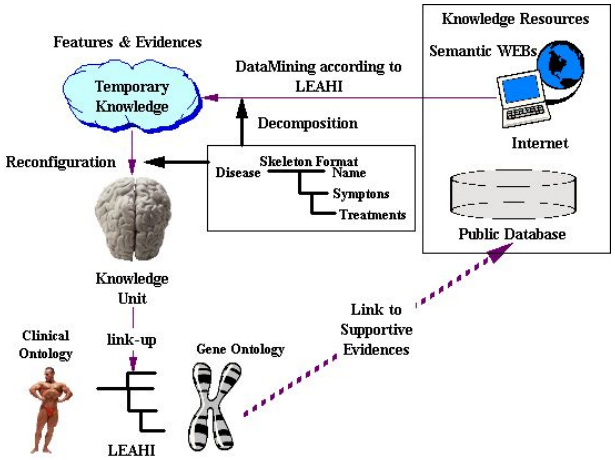


Fig. 4. Knowledge Collection in Semantic WEB on Internet

### 3.7 External Program Relation

From each KU, we can call the external programs as the external computer simulators or the external protocol management tools that lie in Semantic WEBs on Internet. The biomedical conditions for computer simulations should be described in the same LU. Its demanded features are entity, relation, motion, and its time course. On the other hand, the clinical conditions should also be included in the preset conditions of the same LU. This type of LU that describes the computer simulation in Semantic WEBs on Internet is a different type LU that we can call it as a dynamic LU.

### 3.8 Prototyping and Evaluation

Through constructing the prototype of the integrated hematological knowledge according to the knowledge platform concept, we evaluated our concept. This prototype is concerning about hematology. It has the integrated hematological knowledge and the handling programs. The deepness of the knowledge is almost internship level as the clinical knowledge and master degree level as the biological knowledge. The prototyped computer programs have the functions for the searching, the processing and the inputting knowledge.
The evaluation was done through this prototyping. Through construction of a prototype for the integrated hematological knowledge and its control program, our platform concept is evaluated.

## 4. Results

Fig. 5 is concerning about Fanconi anemia, Umbilical Cord Blood Stem Cell Transplantation (UCBSCT), and ex vivo expansion. This figure shows a sample window for UCBSCT. In this case, you can see the molecular relation model of the ex vivo expansion at right center window with the differentiation-function coordinate order. Clicking UCBSCT in the treatment pops up the features of its detailed explanation as right upper window. Clicking the ex vivo expansion in the UCBSCT opens its Cell relation model. The evidences for the clicked entities are shown in right below.

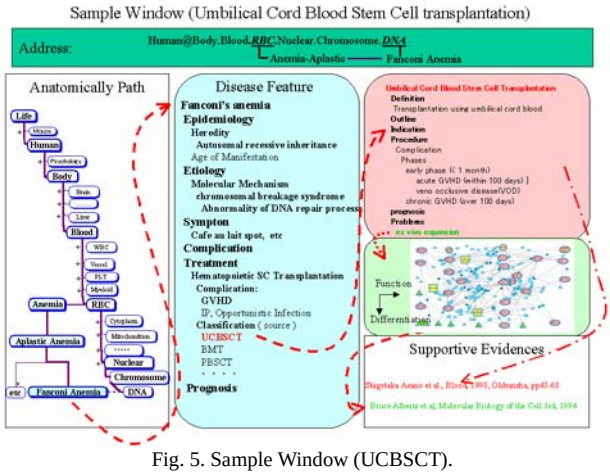


Fig. 5. Sample Window (UCBSCT).

Fig. 6 shows a sample window for the Protocol management. This is concerning about Fanconi anemia, UCBSCT, and the protocol management. In this case, you can see the care protocol of the Umbilical Cord Blood Stem Cell Transplantation at right center window. Clicking the protocol in the UCBSCT treatment pops up the features of its detailed explanation as right center window.

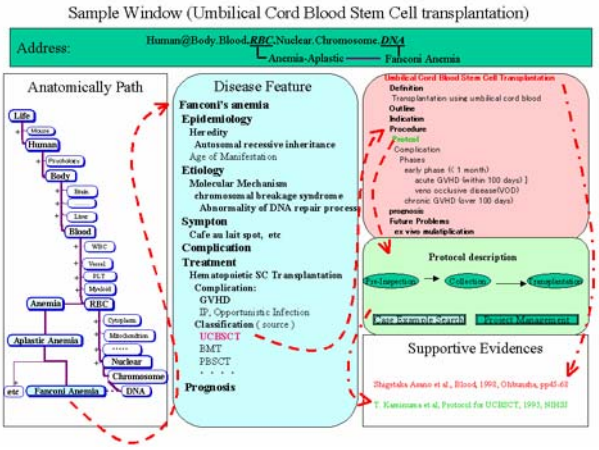


Fig. 6. Sample Window (Protocol Management and UCBSCT).

Fig. 7 shows a sample window for the Iron deficiency anemia. In this IDA case, you can see the erythroblast simulator at right upper window. Clicking the Execution starts the erythroblast simulator. This figure is quoted from previous work [12].

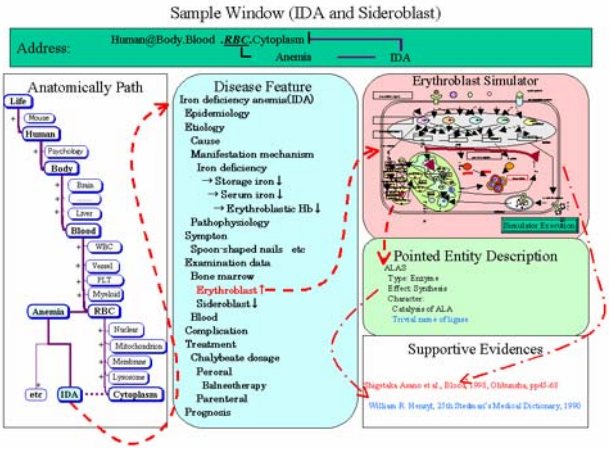


Fig. 7. Sample Window (Erythroblast Simulator and IDA).
(quoted from [12])

## 5. Discussion

### 5.1 Significance

Integrating the multi-disciplinary discrepant knowledge on computer enables to fill the gap between the genome-scientific domain knowledge and the clinically scientific domain knowledge. The basic idea of the integration with keeping logical consistency is to describe both kinds of knowledge with mono-scoped view. Defining knowledge as the logical unit according to LEAHI is the single scoped logical simplification of knowledge. This logically simplified knowledge has the easy understandability and the fast accessibility to the location of the necessary knowledge. This easy understandability and the fast accessibility are the advantage points to perform the knowledge simulation and the data mining on Semantic WEBs. Through performing the knowledge simulation along with LEAHI, we can see the way clear at the simulation because of anatomical understanding. This perspective enables us to cut out the relevant range of the knowledge for each simulation. Through the knowledge collection on Internet with the single scoped index, we can deepen and increase the knowledge without losing accuracy. Performing the data mining from the scattered information on Internet according to LEAHI, the deep and fast knowledge collection can be achieved without losing accuracy. A view on the anatomically hierarchical structure from the micro to macro scope is a good choice as a unique scale that can integrate the genome-scientific domain knowledge and the clinical domain knowledge.

Storing evidences in the knowledge unit makes EBM practical in both genome scientific domain and clinical domain. The evidence based knowledge makes the quality assessment and the judgment of the reliability possible. In big buzz for EBM [15] as scientific reasoning for the diagnosis or the treatment, the evidence based knowledge qualification can be an answer to a question how we can establish EBM practically in the genomic medicine and its translational research.

Our initial clinical target field is the hematology. Here hematology is one of the highly researched genomic fields. In clinical practice of UCBSCT, the problem lies in the low volume of collected stem cell from umbilical cord blood. It makes transplantation difficult to be applied to heavy weight patients. To solve this problem, the ex vivo expansion is important. Though the ex vivo expansion with the single culture of the stem cell is too difficult, the co-culture of the stromal cells from the placenta and the stem cells from the umbilical cord blood is expected to be a successful ex-vivo-expansion methodology in laboratory research. In order to establish this in clinical practice safely, we need more knowledge about reciprocal reaction between the stem cell and the stromal cell including immunology. This knowledge will also be effective for understanding and preventing the GVHD (Graft Versus Host Disease) and the Graft Failure. In the clinical practice, generally, we avoid this critical situation by the HLA donor matching, but practically the unexpected sideration occurs sometimes. The methods for the prediction and for the avoiding these critical situations are seriously needed. If we can train the stem cells immunologically properly during the co-culture by the forward control with the in silico knowledge, we can prevent the GVHD and the graft failure.

### 5.2 Limitations: Validation through example

EBM classification table needs more modification to classify the common-sense knowledge. In our prototype, the knowledge written in books was referred to construct the common-sense knowledge. The common-sense knowledge is classified into single class of "book" that we added to the EBM Cochran classification table. However, we need more classification grades to classify the common-sense knowledge.

### 5.3 Future work.

Based on the collected knowledge and our platform, we can achieve the prediction of the situation through performing the knowledge simulation. Predicting immunological reaction and avoiding the critical clinical situations can improve the clinical safety and the efficiency. Especially the human experimental research phase like translational research needs avoiding the critical clinical situation. This integrated knowledge will be the basis of this kind of the knowledge simulation. Our evidence based knowledge representation makes the knowledge accurate. Increased accuracy for prediction of the clinical events leads to reduce the side effects and the biological hazards. The optimization of the research path minimizes the wasted resources of every research.

The prediction based on the knowledge features is a kind of analogy that can propose the possible situational hypothesis from the existing knowledge. The way of the prediction is similar to the thinking way of human. It means understandable, checkable, and reliable. The expert system, the fuzzy inference, and the neural network model are kinds of method to process knowledge and to infer the situation, but they are more machinelike [16]. The molecular modeling simulation, the pathway model, GON [17], and the virtual reality technology are kinds of method to estimate from the point of molecular modeling, but this kinds of technology still have a long distance to be practical. The clinical biostatistics is a method to analyze clinical data and to estimate situation with its interpolation, but this kind of formulaic analysis does not match to individual medicine like genome-based medicine [18].

Fig. 8 shows the future image of the Umbilical cord blood stem cell transplantation. Clinically our final goal is to improve the Quality Of Life, the clinical efficiency, and the clinical safety. Ultimately it means avoiding or preventing the critical clinical situation like GVHD, or Graft Failure in UCBSCT. For the prevention of the critical situation, we need to predict the future event. In the laboratory scene, the prediction enables the optimized control of the WET devices. It leads to optimize the ex vivo expansion of the stem cells. This makes application of the umbilical cord blood stem cell transplantation to the heavy weight patients possible. Predicting the situation needs the scientifically certified and accurate knowledge. That is the evidence-based knowledge in the current clinical scene. This certification is based on the scientifically certified publications like papers. The bio-statistical confirmation of data is important to form knowledge or judge input data. The integrated Genomic technology like SNPs, DNA chips, Protein chips will be important for the personalized medicine as the integrated scientific input. Here the charts should be computerized to use them as the computer input. All of these future therapies are based on the integrated multi-disciplinary knowledge of the molecular science and the clinical medicine. In addition, the integration must be done logically.

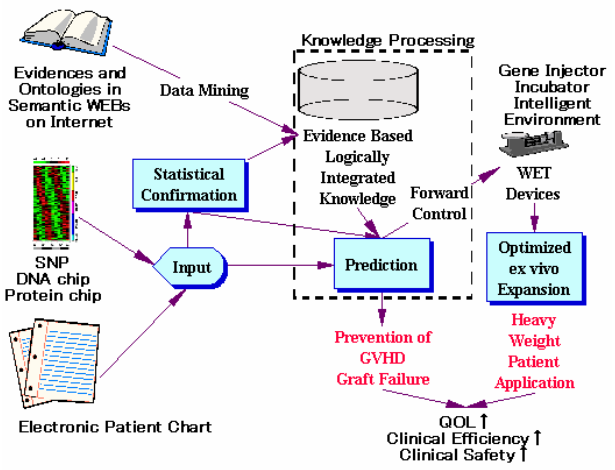


Fig. 8. Future image of Umbilical Cord Blood Stem Cell Transplantation.

In the prototype, we could establish almost all of the clinical knowledge. As for genomic knowledge, we could establish only for ironic metabolism and umbilical cord blood stem cell actions in prototype. Our next challenge is to enrich the genomic knowledge.

## 6. Conclusion

1. Integration of the genome-scientific domain knowledge and the clinical domain knowledge with keeping the logical consistency in semantic WEBs on Internet is the principal demand of both genome-scientific domain and clinical domain.
2. The knowledge representation based on the knowledge unit enables the integration of GO, MO, and CO with single standard. The knowledge representation based on the logical conceptual unit could offer the continuous and seamless logicality from micro to macro. Logical atomism as the system theory fitted to both the genome-scientific domain and the clinical domain.

3. Including the evidences in the knowledge unit (KU) makes the quality of knowledge clear. Storing the Internet links of the evidences in set with the logical unit is the reasonable way to enhance the reliability of the collected knowledge.
4. Evaluating evidences based on the modified Cochran EBM classification table is the commonly acceptable way to qualify the knowledge. Introduction of the evidence and its assessment method make EBM effective in the genome medicine.
5. The logically extended anatomically hierarchical index (LEAHI) makes knowledge continuous and seamless from genome to human. The anatomically hierarchical structure could keep the knowledge logical from micro to macro level in the prototype. The scale free network index was pertinent to represent the molecular network and the disease cross relation.
6. Through construction of a prototype for the integrated hematological knowledge and its handling program, we elucidated the issues that remain to be solved and evaluated our concept.

**Acknowledgment**

We wish to thank Prof. Edward Feigenbaum and Prof. Setsuo Ohsuga for helpful comments concerning about knowledge engineering.

## References

[1] Elias, Z. The NIH Roadmap. Science. Oct Vol. 302. 3 (2003) 63-72
[2] U.S.A., DHHS, FDA (eds.): Innovation or Stagnation; Challenge and Opportunity on the critical path to new medical products. FDA Report, US Department of Health and Human Services Food and Drug Administration, Mar. (2004)
[3] Trombert, P.-B., Rodrigues, JM., Rogers, JE., Baud, R., van der Haring, E., Rassinoux, AM., Abrial, V., Clavel, L., Idir, H.: GALEN: a third generation terminology tool to support a multipurpose national coding system for surgical procedures. Int J Med Inform. Sep. (2000) 58-59:71-85
[4] Stearns, MQ., Price, C., Spackman, KA., Wang, AY.: SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp. (2001) 662-666
[5] William, E.-E., Mary, V.-R.: Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. Science Oct. 286 (1999) 487-491
[6] Musen, MA, Eriksson, H, Gennari, JH, Tu, SW, Puerta, AR.: PROTEGE-II: a suite of tools for development of intelligent systems from reusable components. Proc Annu Symp Comput Appl Med Care. (1994) 1065
[7] Musen, MA.: Domain ontologies in software engineering: use of Protege with the EON architecture. Methods Inf Med. Nov. 37 (4-5) (1998) 540-550
[8] Ashburner, M, Ball, CA, Blake, JA, Botstein, D, Butler, H, Cherry, JM, Davis, AP, Dolinski, K, Dwight, SS, Eppig, JT, Harris, MA, Hill, DP, Issel-Tarver, L, Kasarskis, A, Lewis, S, Matese, JC, Richardson, JE, Ringwald, M, Rubin, GM, Sherlock, G.: The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nat Genet 25 (2000) 25-29
[9] Russell, B.: Philosophy of Logical atomism. Monist 28 (1918) 495-527
[10] KA, Giuliano, Kapur, R, Wang, J, Woo, ES.: Cellular screening of lead compounds in the post-genomic era: An integrated approach to knowledge building from the cellome. Integrated drug discovery technologies Marcel Dekker, New York (2001)
[11] Richard, P., F.: The Character of Physical Law: MIT CambMass, Boston (1967)
[12] Nakaya, J., Ohtsuka, Y., Sakamoto, N., Agishi, Y., "Knowledge Framework aiming at Disease Prevention through Logical Translation of the Balneology to the Modern Medicine," J. J.A. Phys. M. Baln. Clim., vol. 68(4), 209-222, 2005.
[13] Pamela, J., W.: Evidence-based medicine for consumers: a role for the Cochrane Collaboration. J Med Libr Assoc. 90(2) (2002) 218–222
[14] Svetlana, N., Sergei, E., Nikolai, D.: MedScan, a natural language processing engine for MEDLINE abstracts. Bioinformatics. 19(13) (2003) 1699-1706
[15] Friedland, Daniel J. Evidence-based Medicine: A Framework for Clinical Practice. Appleton & Lange, Stamford, CT (1998)
[16] Fukumura, A, Shimura, M, Nagao, S, Tsuji, S.: AI handbook. Japan AI Society: Ohm, (1990) 21-66
[17] Matsuno, H, Doi, A, Hirata, Y, Miyano, S.: XML documentation of biopathways and their simulations in Genomic Object Net. Genome Inform Ser Workshop Genome Inform 12 (2001) 54-62
[18] Spiessens, B, Lesaffre, E, Verbeke, G, Kim, K, DeMets, DL.: An overview of group sequential methods in longitudinal clinical trials. Stat Methods Med Res. 9(5) (2000) 497-515

**Jun Nakaya** received the B.S. and M.S. degrees in Mech. Eng. from Hokkaido Univ. in 1985 and 1987, respectively. After staying in IBM, he received M.D. and Ph.D. degrees from Hokkaido Univ. in 1995 and 1999, respectively. He stayed in M.I.T., Institute of Medical Science, Univ. of Tokyo, and Tokyo Medical and Dental Univ.. Now he is an Associate Professor of graduate school of medicine, Kobe Univ.. He is a member of ISO, HL7, AMIA, ISMH, MIT-J, SSJ, and MSJ.

**Tetsuo Shimizu** received the B.S. degree from Faculty of Science, Univ. of Tokyo in 1971. After working as a researcher in the National Institute of Radiological Sciences, he worked as a scientific manager in the Fujitsu Limited. His research interest is in translational research and ontology. Now he is a Professor of Institute of Medical Science, Univ. of Tokyo.

**Hiroshi Tanaka** received the B.E. and M.E. degrees from Tokyo Univ. in 1974 and 1976, respectively. After staying as lecturer at Tokyo Univ., he received Dr.Med. and Ph.D. degrees from Tokyo Univ. in 1981 and 1983, respectively. He stayed in M.I.T., Uppsala University and Linkoping University, Hamamatsu Univ. School of Med.. Now he is a Director of Center for Information Medicine and a Professor of school of biomedical science, Tokyo Medical and Dental Univ.. He is a President of JAMI and is a member of ISO, MEJ, ICOT, CBIJ, IEEE, and VRJ.