The Translational Research Informatics (TRI)

Jun Nakaya[†]

[†]*Clinical Genome Informatics Center, School of Medicine, Kobe University, 650-0017 Japan*

Summary
This paper introduces the Translational Research Informatics (TRI), its definition, its requirements, its technological background, and the perspectives. The TRI is the essential informatics to support the translational research (TR) phase that is a part of the early clinical trial phase. The TR is believed to be a key pipeline for the success of the omics based medicine. We analyzed the requirements to the TRI and defined the required informational technologies. In concrete terms, the integration of data and knowledge is the fundamental technology, and the establishment of the knowledge based prediction is principal to achieve the safe and efficient TR. The TRI is the critical informatics to meet the practical success in this post genomic era.

Key words:
Translational Research Informatics, Translational Research, clinical trial, pipeline, informatics, knowledge processing

1. Introduction to TRI

Basically the definition of the Translational Research Informatics (TRI) is the informatics to support the translational research (TR) phase. The TR phase is a part of the early clinical trial phase. Here the explicit definitions of the TR phase are slightly different among the nations. In Australia the TR indicates the phase including preclinical, I, and IIa, while the TR in USA indicates the phase including preclinical and I [1]. The difference is the inclusion of the phase IIa. In Japan, Japanese TRI study group defines the TR phase as the preclinical, I, and IIa [2]. Meanly among international countries, the TR indicates the phase including preclinical, I, and IIa. Based on these investigations, we define that the target phase of the TRI is the combined phase of the preclinical phase, the phase I, and the phase IIa.

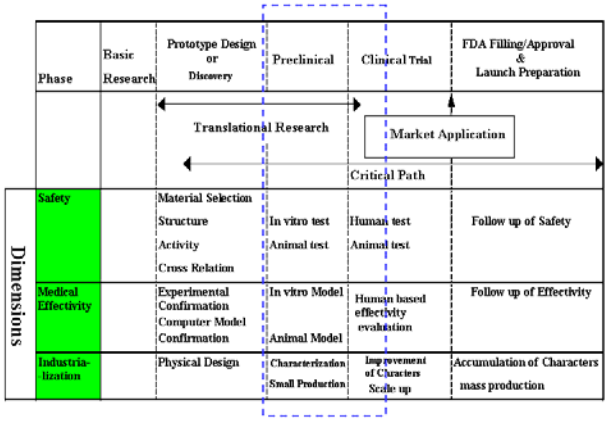
This TR phase is recognized as the principal exit research phase of all biomedical researches in USA, Australia, and EU countries. The TR phase has a role as a social pipeline to bridge the fruits of basic research to the general public. The TR phase is essential to earn the social consensus or the legal background for applying the cutting edge medicine like genome medicine to the general practice.

Viewing the TR phase from the point of informatics side, we can rephrase the TR phase as the translational phase of the information or the knowledge from basic science to clinical practice. That is to say that the TR phase is the transitional phase from basic scientific knowledge to clinical knowledge for the effective use of the knowledge. The object of TRI is to support this knowledge transition process from the IT side of view with keeping the maximized safety of the human based clinical trials and the maximized efficiency.

2. Requirements to the TRI

Recently in US, the biomedical research cost increases abruptly [3], while the number of the new drug application decreases [4]. This means that the developmental cost per a new drug significantly increases. In concrete terms, it increased from 1.1 B$ (2000) to 1.7B$ (2002). The reason of this increase was based on the critical path phase that increased both the time consuming and the economical cost. This critical path phase includes the TR phase. Specifically the TR phase is an initial part of this critical path, and needs for the optimization of the TR phase was recognized. In other aspect, the TR phase is the essential phase to install a novel clinical method based on a novel discovery to the general public. The TR phase is also essential to energize the drug discovery market. The requirement analysis on overall technologies to support this pipeline including the TR phase has done at US Food and Drug Administration (FDA) and National Cancer institute (NCI) [5].

Table 1 Three dimensions and Important Techniques



Manuscript revised July 5, 2006.

For the optimization of the TR phase, the technological requirement analysis had done by US FDA and Institute of Medical Science of the University of Tokyo (IMSUT) [3,6,7]. Table 1 shows the outline of the result. The benchmarks of optimization are the clinical safety, the clinical efficiency, and the industrialization. In TR phase as a kind of a critical toxicological phase for human, the mostly prioritized benchmark should be the clinical safety, the second one is the clinical efficiency, and the third one is the industrialization. The point that the maximum safety is prioritized than the industrialization is different from the other industries. In table 1, based on these three benchmarks, the important technologies are indicated. Many technologies as the animal model, the testing methodology, and the scale-up methodology are indicated for whole pipeline process. As for the TR phase, the documentation technology and the prediction technology that can predict the final clinical result at an earliest timing possible are principal. In other words, we require the prediction technology that can predict the final result of the human based clinical trial at an earliest stage possible. Followings are the example of this technology:

A. The cause and effect model that can be used to predict the results of human based clinical trial or economical effect

The examples are the correlation table between the bio-system (transgenic mouse, cell line, etc) experimental result and the human clinical result, the cause and effect model of "pulse wave - arrhythm - death" at neoplastic patient. The object of this technology is to avoid the critical trial, unfruitful trial, and unproductive investment by predicting the final conclusion of the human based clinical results at an earliest stage wherever possible. At the stage of novel creation of the medical methodologies, the prediction technology is expected to predict the result of the human based clinical effects, the side effects, the clinical risks, the economical risks, and the clinical paths (the clinical conditions that patient will pass) with enough accuracy. This prediction technology does not need to stand on the deterministic model, if the prediction model has enough accuracy. Considering the context that the current deterministic technology does not have enough technological level that can predict the complicated phenomenon as clinical facts with enough exactness, adopting the cause and effect models that have enough accuracy is a reasonable choice to establish the prediction model for the TR optimization. The TRI stands on the standpoint that we need the practically usable prediction model based on present technologies.

B. Data collective fundamentals to establish the cause and effect relational model

On the assumption to establish the cause and effect relational model, we need to collect the clinical case data and knowledge. To collect the clinical case data and knowledge, we need the data collective fundamentals that is ordinary called as the social informational fundamental technology. This social informational fundamental technology is the primary demanded technology and is based on the collection, the integration, and the systematization of information and knowledge. The specific technologies are as follows:

- Ontology as a technology for index/knowledge/terminology
- Markup Language as a technology of data exchanging format
- Integrated database as a technology for data store (Compatibility with Markup Languages must be considered because this technology premises data exchanging)

The data and knowledge derived from these technologies must be sharable internationally from the view of data availability. This implies that these technologies must link up with the international standardization activities as International Organization for Standardization (ISO) [8], Health Level Seven (HL7) [9].

C. Optimized clinical planning and management methodology from informatics aspect

Because the TR phase is a so-called human based experimental phase, the TR protocol must be planned with error-free optimization. The planned TR protocol must be executed exactly and must be managed appropriately. To achieve these, the optimized protocol generation technology and the protocol execution management system must be required. IMSUT and Kobe university CGI center developed these technologies [10]. The developed technologies are based on the rule generation and the management of the protocol execution. The both are based upon the protocol knowledge that is a kind of the summation of the rules. This technology implies the knowledge migration technology with copy based migration from a TR facility to a clinical trial facility.

D. Novel biomarker or statistic marker and measurement technology having causal relation with the human clinical result or the final economical effect of the probes.

New biomarkers and statistical measurement as a marker for psychological effect are required to establish more reliable and profound prediction.

E. Benchmarks to evaluate the predictive accuracy

The benchmark itself is required to evaluate the novel technologies from the point of the predictive accuracy. The benchmarks for causal effect models and the logical background of evaluations are required.

F. Biomedical knowledge quantification methodology for quantified prediction

To achieve the precision prediction, we need to quantify the prediction to compare the accuracy of the prediction. To quantify the prediction, we need the quantified

knowledge basically to calculate the predictions. The biomedical knowledge that is described with the multi dimensional features [11] may have the potential of the quantification.

Among these technologies, the main technologies are A, B, and C. The others as D, E, and F are the subsidiary technologies for A, B, and C.

3. Social Informational Fundamentals of TRI

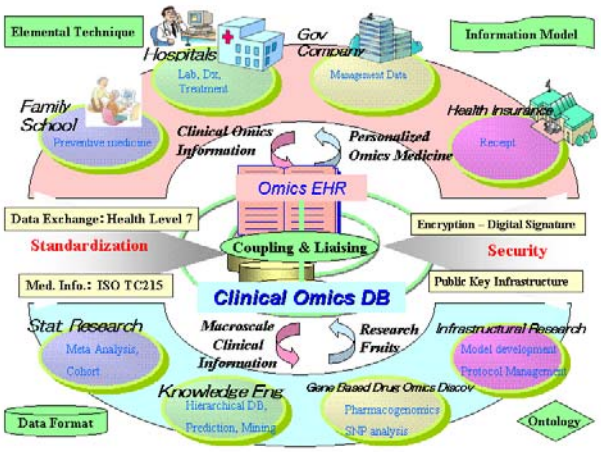


Fig. 1 Social Informational Fundamentals of TRI

In this post genomic world, we must consider the novel clinical practice that is based on the Omics information. The Omics information is the summation of the hierarchical omics data, and these data must be integrated with logical order. According to the above analysis, the required social informational fundamentals in TR phase are shown in Fig. 1. Basically the principal technologies are the omics electronic health record (EHR) and clinical omics databases. Coupling and liaising these two technologies enable the high efficient legal/special documents generation/utilization, the high efficient data collection as research fundamentals, and the safe and efficient offering of novel medical methods to the general public. Three supporting elemental techniques for these two technologies are the ontology, the information model, and the data format.

In order to go on these technologies realistically in general public, we need the standardization of the elemental techniques, the data sharing, and the securing of the data security. The standardization of the elemental techniques is important for data sharing, meta analysis, technological compatibility, and reconciliation of facility's gap. Considering that the TR phase does not have a scaling up capacity of number of patients on the reason that it is a kind of toxicological phase, the utilization of data and knowledge in meta analysis by the standardization has a significant meaning. This characteristic of TR phase is different from the middle/latter part of clinical trials. In this post genomic era, the data that pass this TR phase will have omics data containing genome data fatally. Because these clinical omics/genome data are the ultimate personal information, the securement of the data security is the considerable prerequisite of the data sharing and the standardization of the elemental techniques. The current elemental techniques for the data security are public key infrastructure (PKI), digital signature, and encryption.

The practical pattern to apply the basic scientific findings as the fruits of the genomic science to the clinical practice is summarized in two applications as the novel diagnostic method or the novel treatment technique (Fig. 2). In order to make these novel applications usable at general clinical scene, all novel applications as drugs/devices must pass the social pipeline including the TR phase and the clinical trials to earn the legal background and social consensus. Here the essences of what pass the pipeline are data, information, and knowledge. And the informational fundamentals that can support the transition process and the data passing are significant. This transformed knowledge through these informational fundamentals is used for general public as general practice, health promotion, and disease prevention. In normal course, a novel general practice technique based on the novel methodology will be designed from the standard guidance that is generated from the collected data and the transformed knowledge. The continuous market research and the continuous data collection of the novel product after opening market can be good basis to get the future target and to find the hidden issues/chances. The earned data from the clinical trial process and the continuous evaluations will be analyzed with statistic methods and stored in the integrated knowledge base, this knowledge is used for the targeting process or the creation process of novel methods.

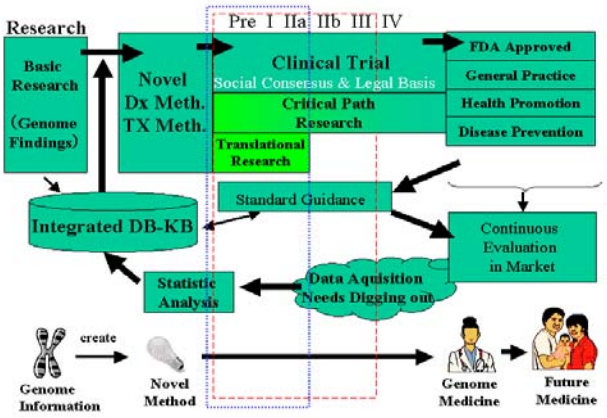


Fig. 2 The Flow of Pipeline and Genome Medicine

The TRI needs above informational infrastructure that should have an information cycle as for the utilization and the collection of information. This infrastructure is required as a social platform to open the door to the new medicine.

4. Inception of TRI

The TRI has just begun and is in its infancy. The current status of the TRI is in the stage of construction of the data collection platform. The elemental techniques of the data collection platform are the information model, the ontology, and the data format. These are important to establish the integration of omics EHR and clinical omics database.

Among the above-mentioned elemental techniques, as the medical information exchange technique based on the information models, ISO and HL7 are constructing the information models for the medical domain. These efforts are adopted in the national EHR project in some countries, and the international data collective infrastructure is being put into place [12]. Recently, HL7 has started the cooperation with the CDISC (Clinical Data Interchange Standards Consortium) that promotes the computerization of Clinical Trial from a pharmaceutical company's view [13].

As the ontological technique, the Gene Ontology (GO) is the biological ontology mainly for the genes [14], while the SNOMED-CT (Systematized Nomenclature of Medicine for Clinical Terms) is the clinical terminology for clinical use [15]. With holding many discussions, these go near to be an international standard. In US and UK, the national projects as the integration of these ontologies and the connection with the messaging models have started. In Japan, there is a project of clinical omics ontology framing to support the TR [16]. The basic platform for the TR system is going ahead.

As for the data exchanging format, now, the Genomic Sequence Variation Markup Language (GSVML) is in international standardization at ISO [17]. The GSVML is the data exchanging format of genomic sequence variation data to use it mainly in human clinical medicine. The object of GSVML is to enhance the data exchanging of genomic sequence variation data with focusing on SNP mainly for human based clinical use internationally. The GSVML receives an expectation to contribute the international collection of clinical genomic data with coupling on EHR information models that is in process of international standardization. The GSVML is originally from Japan and is the world first international standard in genomic medical information domain.

5. Horizon of TRI in Post Genomic Era

Getting into the post genomic era after the completion of the human genome project [18], currently the productions of the practical fruits are expected from the genomic researches. The genome medicine is one of the main targets of the expected outcomes. The genome medicine is still in the stage of experimental medicine, and it needs to pass through the TR phase as the social pipeline to realize the genome medicine as the general practice in general public. The TRI is the informatics to support the social pipeline and is expected to play an important role in future.

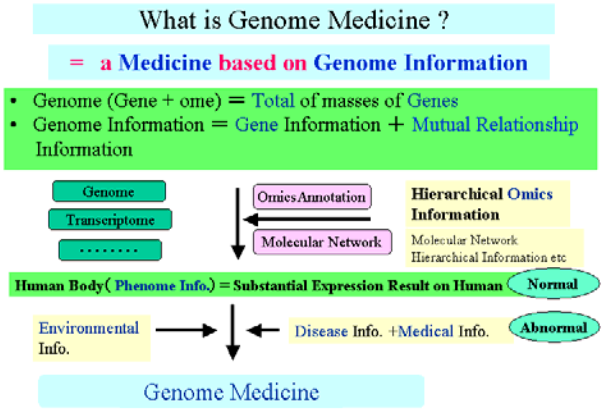


Fig. 3 Definitions and their Relations in the Genome Medicine

The genome medicine is a medicine based on the genome information (Fig. 3). The genome information is the summation of the information of respective genes and the cross relationships of the genes. To connect the genome information to the genome medicine, we need more information as the phenome information that is substantial information of expression in human body, the hierarchical omics information that exist between the genome information and the phenome information, the disease information that represent the abnormal state of the human body, the clinical information that is for the diagnosis or the treatment of the disease, and the environmental information that can be the cause of the disease.

This genome medicine has three characters as the personalized medicine, the pre-emptive medicine, and the scientific medicine. The personalized medicine is the medicine per nation per region per personal. The genome medicine has characteristics of the tailored medicine and the customized medicine. A large amount of information is essential to achieve the customized medicine, and the genome data especially genomic sequence variation data can offer this great deal of data.

The pre-emptive medicine is a kind of the preventive medicine that is based on the precise prediction of the state enabled by the tons of genome information.

We can say that the preventive medicine is a kind of economical medicine achieved by the disease prevention or the condition increment prevention.

The scientific medicine is the medicine based on the clinical case integrated with the omics background. The substance of the evidences at EBM (Evidence Based Medicine) that is currently accepted in international clinical domain is the clinical cases, and the scientific medicine will support the scientific background with the vast amount of omics data. To generate the omics background, we need the integrated omics database and the analysis technique of omics network.

In the genome medicine, the informatics undertakes a role to support three clinical objectives as the clinical safety, the clinical efficiency, (clinical effect, working efficiency), and the economical efficiency from the informatics side of view. The TRI undertakes a role to support the process of the actualization of the genome medicine and the bridging to the general public from the IT side of view.

From the standpoint of informatics, we can say that the TR phase is a kind of the transitional phase of knowledge from the basic knowledge that is derived from the genome based findings to the medical knowledge. In this way, the TRI is the essential supporting field to establish the genome medicine truly.

6. Global Provisions of TRI

Many efforts of the TRI have started globally.

In US, following the term of "new pathway to discovery" in the NIH roadmap [19], the knowledge informational infrastructure project had started as the toolbox for the researchers on a budget of $139M from 2004. The main focus is in the integration and the utilization of the knowledge for the effective use of the scattered data on this Internet society. This infrastructure is principal to promote the molecular targeting development.

In Australia where is also the advanced country about the TR like US, the IT system is recognized as the infrastructure to enhance the availability of data. They started with the integration of the distributed patient databases [20]. They think this issue is urgent and principal to utilize the accumulated patient data. The principal role of this infrastructure is to minimize the risk of the TR by predicting the result of the clinical trials at early phase.

In Japan, a NTRSC (National Translational Research Support Center) concept is in under contemplation. Fig. 4 shows the concept of the NTRSC. The concept is discussed at the TRI study group. The NTRSC is a kind of social fundamental IT infrastructure and is based on the knowledge processing. Currently the project is in the developmental phase of the prediction techniques based on the knowledge processing and the documentation techniques based on the coupled technology of EHR and DB. Basically this project is prototyping the informational platform infrastructure to collect the nation-wide information.

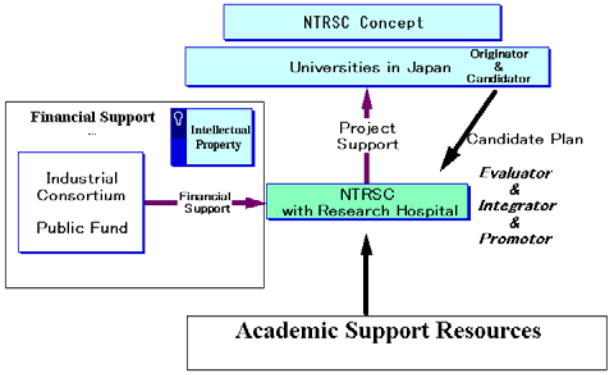


Fig. 4 A Concept of the NTRSC in Japan

7. Conclusion

In days ahead, the TR must be more principal to secure the human health from an aspect of installing the new technologies to the general public. The TRI must work as a social informational fundamental that can support the TR phase where is a part of the social pipeline for the life science. To achieve the strategic utilization of the information in the TRI, we must promote the continued effort to integrate the international information and to develop the practical technologies. The TRI is the critical informatics to fruit the omics researches practically in this post genomic era.

Acknowledgment

I wish to thank Tetsuo Shimizu and Hiroshi Tanaka who are the core members of the TRI study group.

References

[1] Lawrie W Powell : Bridging the gap between basic science and clinical medicine: mentors and memories, MJA 177 (11/12): 657-660, 2002

[2] Nakaya J, Shimizu T, Tanaka H, Asano S : Current Translational Research in Australia and Translational Research Supporting Center (TRSC) in Japan, Chem-Bio Informatics Journal, CBIJ1595- 5(2), 27-38, 2005

[3] NEDO in Washington DC : Investigated Report of 2002. ＜http://www.nedodcweb.org/report/2002-10-7.htm＞（URL）[accessed July 11, 2006]

[4] NHMRC : Performance Measurement Report 2000–2003. <http://www.health.gov.au/nhmrc/aboutus/pmf2003.pdf> (URL) [accessed June 10, 2004]
[5] 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, Washington, Mar. 2004
[6] Shigetaka Asano, Jun Nakaya : Key Note "Hurdles to establish Genome Medicine and the NTRSC", Proceedings of the 1st Symposium on Genome Medical Informatics, Tokyo, 2004, pp15-18
[7] Tetsuo Shimizu, Jun Nakaya, et al : Report on RR2002 Project, Tetsuo Shimizu (eds.), Ministry of ECSST（Tokyo), 2004, pp4-45
[8] International Organization for Standardization (ISO) : TC215 Health Informatics . ＜ http://www.iso.org/iso/en/ISOOnline.frontpage＞（URL） [accessed July 11, 2006]
[9] Health Level Seven : What is HL7 ＜http://www.hl7.org/＞（URL）[accessed July 11, 2006]
[10] Jun Nakaya, Tetsuo Shimizu, Hiroshi Tanaka : e-Protocol - A Translational Research (TR) Protocol Generator and Monitor -, International Journal of Computer Science and Network Security, 6 (7), ISSN: 1738-7906, in Printing, 2006
[11] Jun Nakaya, Tetsuo Shimizu: Knowledge Architecture based on Evidence Based Logical Atomism for Translational Research, International Journal of Computer Science and Network Security, vol. 6-2, ISSN: 1738-7906, 175-179, 2006
[12] International Affiliate Committee of Health Level Seven : HL7 IAC. ＜http://www.hl7.org/＞（URL）[accessed July 11, 2006]
[13] Clinical Data Interchange Standards Consortium: CDISC, <http://www.cdisc.org/> (URL) [accessed July 11. 2006]
[14] 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, 25-29, 2000)
[15] Stearns, MQ., Price, C., Spackman, KA., Wang, AY.: SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp. , 662-666, 2001
[16] Jun Nakaya, Tetsuo Shimizu : Knowledge Architecture based on Evidence Based Logical Atomism for Translational Research, International Journal of Computer Science and Network Security, 6 (2), ISSN: 1738-7906, 175-179, 2006
[17] Jun Nakaya : Genomic Sequence Variation Markup Language (GSVML), ISO/#25720 , International Standard Organization TC 215 WG2, N422 Committee Draft, 1-132, 2005
[18] Barbara R. Jasny and Leslie Roberts : Introduction to special issue; Building on the DNA Revolution, Science, 11 April Vol. 300. no. 5617, 277, 2003
[19] Elias, Z. The NIH Roadmap. Science. Oct Vol. 302. 3, 63-72, 2003
[20] BIO21 Australia ltd.: BIO21 member institutions <http://www.bio21.com.au/members.asp> (URL) [accessed July 11, 2006]

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.