

# Integrating Clinical and Biological Information in a Shanghai Biobank: An Introduction to the Sample Repository and Information Sharing Platform Project

Wenbin Cui,<sup>1</sup> Peiyong Zheng,<sup>2</sup> Jiahong Yang,<sup>2</sup> Rong Zhao,<sup>2</sup> Jiechun Gao,<sup>2</sup> and Guangjun Yu<sup>1</sup>

Biobanks are important resources and central tools for translational medicine, which brings scientific research outcomes to clinical practice. The key purpose of biobanking in translational medicine and other medical research is to provide biological samples that are integrated with clinical information. In 2008, the Shanghai Municipal Government launched the “Shanghai Tissue Bank” in an effort to promote research in translational medicine. Now a sharing service platform has been constructed to integrate clinical practice and biological information that can be used in diverse medical and pharmaceutical research studies. The platform collects two kinds of data: sample data and clinical data. The sample data are obtained from the hospital biobank management system, and mainly include the donors’ age, gender, marital status, sample source, sample type, collection time, deposit time, and storage method. The clinical data are collected from the “Hospital-Link” system (a medical information sharing system that connects 23 tertiary hospitals in Shanghai). The main contents include donors’ corresponding medication information, test reports, inspection reports, and hospital information. As of the end of September 2014, the project has a collection of 16,020 donors and 148,282 samples, which were obtained from 12 medical institutions, and automatically acquired donors’ corresponding clinical data from the “Hospital-Link” system for 6830 occurrences. This project will contribute to scientific research at medical institutions in Shanghai, and will also support the development of the biopharmaceutical industry. In this article, we will describe the significance, the construction phases, the application prospects, and benefits of the sample repository and information sharing service platform.

## Introduction

IN RECENT YEARS, THE SHANGHAI Municipal Government has increased its investment in the field of translational medicine by encouraging universities, research institutions, and hospitals to perform research in the biobanking field.<sup>1</sup> The goal is to improve the level of medical care and resolve real-life medical issues, namely reducing morbidity and mortality, improving recovery rates and quality of life,<sup>2</sup> and decreasing healthcare costs. Many places in China have recently established biobanks, but biobank construction is scattered and redundant. In particular, many current biobanks have the same or similar contents, leading to wasted manpower, materials, and financial resources. There are currently no unified procedures for operation or quality control, so different collections in China employ different standards for storage, nomenclature, coding, and quality control. The biobank information management systems are underdeveloped, making integration of information on

samples and clinical data difficult, and also making resource sharing difficult. In brief, there is limited exchange between different biobanks, the available samples are not used efficiently, and the relevant laws and regulations are imperfect. And there are still many social, ethical, and legal issues that need to be considered.

The samples in a biobank need to be connected to the diverse types of clinical information available for hospital patients, and a platform should be established which connects different hospitals and allows sharing of relevant information. In this way, the data from diverse clinical studies can be integrated and utilized for various medical applications. In China, there is no biobank network that has been established and led by a public institution. Therefore, the Shanghai Hospital Development Center, a state-run and nonprofit public institution, which was approved to be established by the Shanghai Municipal People’s Government in 2005, and is in charge of municipal public hospital investment, construction, operation, management and

<sup>1</sup>Children’s Hospital of Shanghai, Shanghai Children’s Hospital, Shanghai Jiao Tong University, Shanghai, China.

<sup>2</sup>Department of Medical Affairs, Shanghai Hospital Development Center, Shanghai, China.

assessment, is initiating a network biosample bank model. This model project combines decentralized and centralized biosample storage, management, and sharing of information.

To a great extent, this project has achieved a certain breakthrough in the field of biobanking. The construction, management, and maintenance of a biobank require significant resources. In order to facilitate the sharing of medical resources and encourage basic and clinical studies of important diseases, this project will standardize the biobank data which are distributed to different hospitals. It will also develop a link between the biobank and the clinical data sharing platform of the “Hospital-Link” project and establish a platform that integrates clinical practice and biological information, to provide an information-sharing service that can be used in diverse medical and pharmaceutical research projects.

### Methods and Results

#### Project goals and construction contents

The general goals in constructing the Shanghai biobank are to resolve key technical problems during biobank construction, develop a core information platform, demonstrate the use of the biobank platform, and provide a technical foundation for biobank construction in Shanghai. The construction goals are to establish: (a) a biobank with decentralized and centralized forms of storage for major diseases in Shanghai; (b) a case-centric biobank information system which integrates information from clinical practice, laboratory analyses, and biological samples; and (c) an international biotechnical center that provides technical support and sharing of services for the development of large-scale clinical trials and basic medical research. Figure 1 shows the

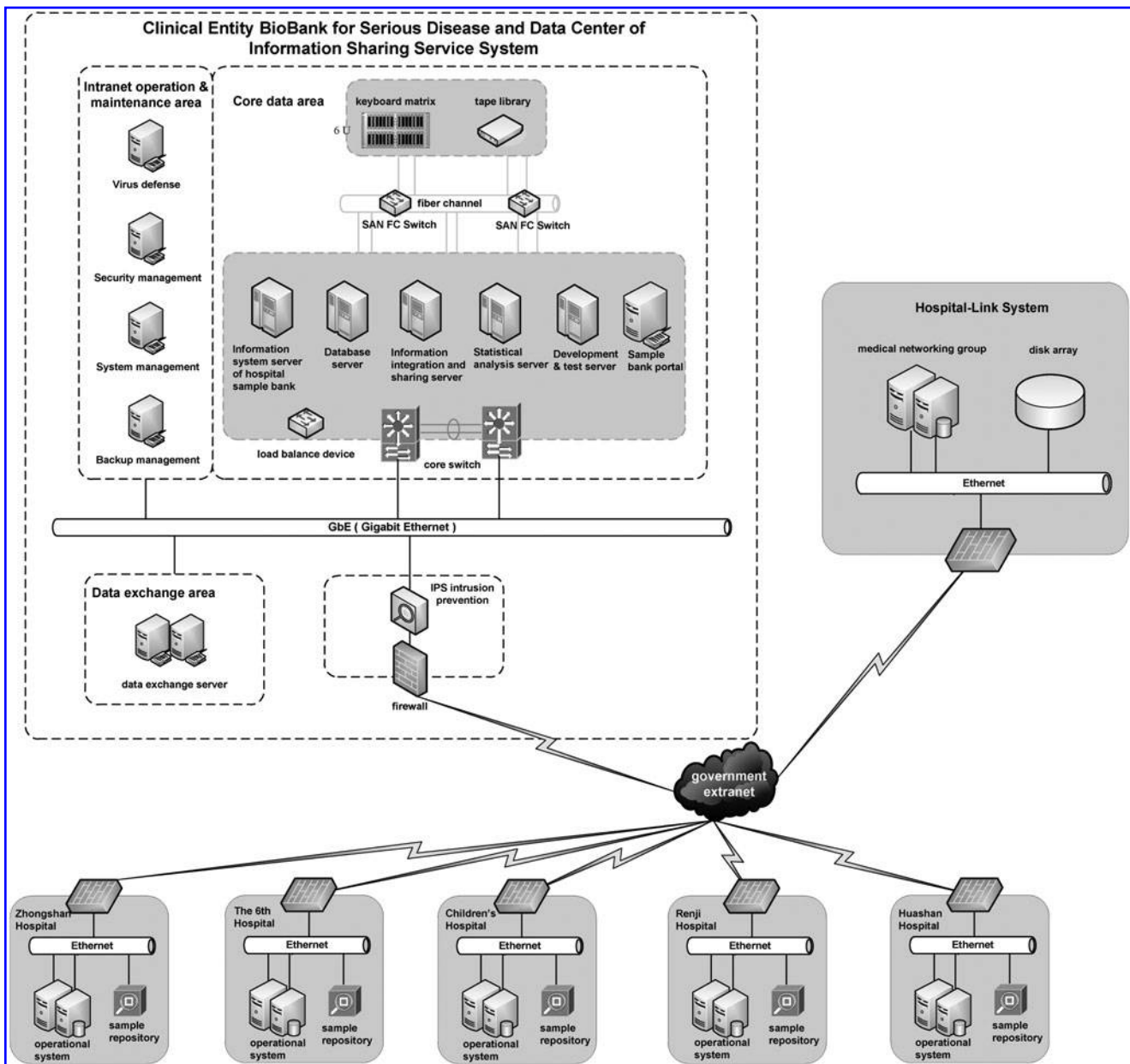
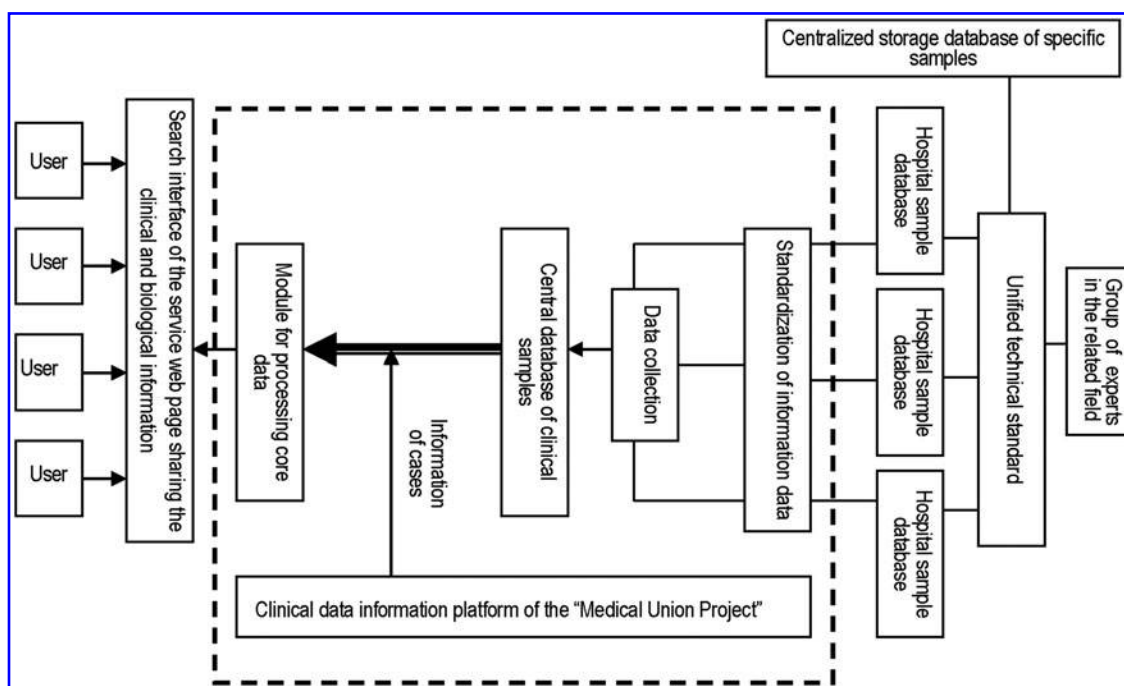


FIG. 1. Network topology of the sample repository and information sharing platform.



**FIG. 2.** Approach used to establish a biobank platform that integrates clinical practice and research of biological samples.

network topology of the sample repository and information sharing platform, and Figure 2 summarizes the approach used to establish the Shanghai biobank platform. By the end of September 2014, the project had collected and stored 148,282 samples from 16,020 donors, which were obtained from 12 participating hospitals, and acquired 6830 donors' corresponding clinical data from the "Hospital-Link" system.

#### *Selection of pilot hospitals and diseases, establishment of groups of experts, and unification of technical and management standards*

Some hospitals in Shanghai have constructed their own biobanks.<sup>3</sup> For example, Zhongshan Hospital has constructed a liver cancer biobank with an area 300m<sup>2</sup> with space available for 2000 blood biospecimens and 3000 tissue biospecimens. Ruijin Hospital has constructed a hypertension biobank, which has an area of 300m<sup>2</sup> and space available for 1000 blood biospecimens, 5000 DNA biospecimens, and 1000 bodily fluid specimens. In this project, twelve municipal general hospitals and specialized hospitals that operate under the "Hospital-Link" platform have been selected as pilot sites for the construction of a "decentralized sample storage database." Two sites have been chosen to store specific samples from each hospital to form a "centralized sample storage database," and a sample bank network management model is being implemented by combining decentralization and centralization. Acute and chronic hepatitis, rheumatoid arthritis, malignant tumors, and diabetes are the disease areas to be examined initially, and information of samples from patients with these diseases is included in the pilot units. Groups of experts are engaged in each of these fields. The expert team includes those who specialize in clinical research, basic research, and bio-informatics, who can guide collection, preservation, information management, safety management, and application management of the biospecimens of patients with these diseases.

#### *Adjustment of existing information management systems for establishment of a sample bank in pilot hospitals to assure unified format of collected data*

Based on the key characteristics of the above-mentioned diseases, the core frameworks of the biobank information management system have been unified in the pilot hospitals, based on the sample coding norms established by the specialists. This includes standards for sample preservation and data collection. The information management system of the sample bank in the pilot units meet the following basic requirements: (a) supplement and adjust the entry fields of original case information, and unify the field names according to the standards of data collection for each disease; (b) identify the sample codes of pilot hospitals by data standardization in a unified manner according to the standards for sample coding and storage; and (c) provide an open data interface of the original sample information management system in the pilot hospitals, to assure that the biobank information of pilot hospitals can be collected in a unified manner. So far, there are 12 hospitals in Shanghai participating in the project: Zhongshan Hospital, Huashan Hospital, Renji Hospital, the 6<sup>th</sup> People's Hospital, Shanghai Cancer Center, Changhai Hospital, Changzheng Hospital, the 9<sup>th</sup> People's Hospital, Obstetrics and Gynecology Hospital, Children's Hospital, Shanghai Chest Hospital, and Shanghai Pulmonary Hospital, are all now adopting the unified standards of sample preservation and data collection.

#### *Establishment of a central clinical biological sample database*

A central clinical biological sample database connects the various pilot hospitals, regularly collects biological sample data from the pilot hospitals, and updates data in the central database. Patient information is the key field for identification of patients and establishing the connection between the

central database and the clinical information data of the original “Hospital-Link” project (e.g., medical record information, pathological examination information, laboratory test information, and imaging examination information). Establishment of an information sharing system that includes clinical data, biological sample data, data searching, and statistical analysis will enable the combined use of information on clinical diagnosis, medication, laboratory tests, and sample collection, and preservation information. The system will be based on the case characteristics as the search parameters to explore an integrated analysis model and support research in translational medicine.

There have been some problems encountered in the process of the project construction. The first problem is that data labeling and coding are not standardized, because the biobank information management systems within the above 12 hospitals were developed by five different companies, and data items such as sample type, disease code, also have different reference specifications. As a result, the collected data in the data center cannot be classified. In order to solve this problem, members from the participating units, representatives of manufacturers, and experts in biobanking were specially convened to discuss the formulation of a sample data set standard for reference.

Another problem is missing sample data. The information management systems of most participating hospitals have not been connected with other information systems. The biobank information management system is an information island and most sample information is input manually by management personnel. This approach results in low efficiency as well as a high error rate, resulting in loss of data items, such as patient identification and disease diagnosis. To solve this problem, we coordinated with the hospital information center, to assure that the biobank information management system and hospital information system (HIS) or clinical information system (CIS) were interconnected, in order for key information corresponding to the sample to be automatically obtained from the clinical system.

#### *Establishment of an operation and management system for the platform*

The framework, which was established for organizing and managing platform construction and operation, identifies the responsibility of running each aspect of the platform to ensure smooth construction and operation. The established information management system includes the ability to search for information, ensure the safety of database information, the protection of patient privacy, and regular data backup and server maintenance. The sharing service management will include specialists in the related field and pilot hospitals which will establish biological sample sharing management procedures. These include application processes, academic and ethical review, protection of the rights of sample providers, and compensation method, for the decentralized storage of samples in each hospital and for the centralized storage of samples. Through utilizing the established operation and management system, one decentralized storage unit, Zhongshan Hospital, is collecting, processing, and storing donors’ related tissue and fluid specimens from four tumors (lung cancer, gastric cancer, colorectal cancer, and renal cell carcinoma tissues and fluids), and to date has collected 500 samples of each tumor.

Another unit, Hua Shan Hospital, is collecting, processing, and storing donors’ related tissue and fluid specimens from nervous system tumors, and has collected 1500 samples. Children’s Hospital is collecting, processing, and storing donors’ related tissue and fluid specimens for three disease categories (childhood cancer, hereditary metabolic diseases, and birth defects), and has collected 2500 samples of each disease type. To encourage the cross-unit sharing of biological samples, specialists in the related fields and the pilot hospitals have established an evaluation and incentive system for sharing. Practices of cross-unit sharing of biological samples will be increased by the incentive program. For instance, for one hospital, the more biological samples it shares, the more development funds and policy support it will receive from the government.

#### *Establishment of a biobank research core institution and a third-party public platform*

The Shanghai Clinical Research Center and the Shanghai Biochip Company, Ltd. together built the third-party service platform at a cost of 20,000,000 RMB (approximately 3.3 million US dollars). The building covers an area of about 3500 square meters. A council and director have responsibilities for management of the platform. The main tasks include management of the central sample database, coordination of standardized systems, development of ethical guidelines, training of personnel, construction of a quality management and control system, provision of services using the technical service platform, promotion of scientific cooperation, and coordination of sample sharing. The third-party public platform will play an important role in quality management, the timely detection of problems, and developing solutions for biobank construction, operation, and management.

## **Discussion**

The emergence of translational medicine has led to the rapid and efficient translation of basic research into clinical practice<sup>4</sup> and a biobank is a comprehensive repository of biological specimens that can be used for basic research and clinical trials.<sup>5,6</sup> These developments have produced new information that can be used for prevention, diagnosis,<sup>7</sup> examination, treatment,<sup>8</sup> and the development of new drugs<sup>9,10</sup> and health prevention measures that can benefit all patients. The rapid development of translational medicine is closely connected with the development of biobanks,<sup>11</sup> and the establishment of biobanks stimulates the rapid industrialization and clinical applications of scientific research,<sup>12</sup> the ultimate realization of translational medicine. It would be to their advantage for all countries to become engaged in research in translational medicine<sup>13</sup> and provide political, economic, and technical support for such research.<sup>14</sup>

The construction and development of biobanks cannot happen without an informatics platform,<sup>15</sup> which integrates clinical practice with data from biological samples. For the overall construction plan of the Shanghai biobank, and to solve key technical problems involved in the construction process of the biobank project, we researched and developed a core information platform, consisting of a standardized hospital client service system based on the Software-as-a-Service (SaaS)<sup>16</sup> model, a clinical information and biological

information integration and sharing platform, and a convenient and practical one-stop sharing query service system. We implemented a standardized application of the information platform in 12 pilot hospitals, achieving high quality, timeliness and reliability of biospecimen data transmission, providing an application demonstration for the realization of a one-stop search service for all of Shanghai as a sharing information resource. Through 2 years implementation of the project, 12 participating units are interconnected, and through the end of September 2014 the project has collected 148,282 samples from 16,020 donors, and automatically obtained donors' corresponding clinical data from medical system for 6830 occurrences.

There are five operational innovations involved in this project. The first is realizing biological specimens distributed storage in different medical institutions and centralized information management. The second innovation is setting up unified rules for biological specimens collection and quality management; the third is realizing the integration of clinical information and biological information; the fourth is improving and optimizing research on safeguard mechanisms for ethics, sharing service, technology management and other aspects; and the fifth is implementing an application demonstration of translational medicine research based on biobanking. There are also some technology innovations, such as optimizing sample processing technology to build high a quality biospecimen repository, storing sample information based on cloud storage technology, providing analysis tools for translational medicine, and employing informatics for Integrating Biology and the Bedside (i2b2). In addition, the approach provides intelligent search capabilities for sample information and provides a standardized management service for hospital biospecimens using cloud computing technology, and for implementing an application demonstration on the hospital end, based on the Internet of Things (IOT)<sup>17</sup>. Several new breakthroughs have been achieved in the project: setting standards and specifications for sample labeling and data acquisition; realizing the interconnection and integration of heterogeneous sample information systems; realizing integration and sharing of biological information and clinical information; and providing customized analysis tools for scientific research which is based on biobanking.

The established sharing platform integrates clinical practice data and information of biological samples and provides significant social and economic benefits. First, it provides a technical framework and a platform for further integration of biobank information in Shanghai hospitals and for the production of a larger information platform. Second, the implementation of a large-scale medical biobank model will initially combine the decentralized and centralized storage of samples and unify information management and sharing in Shanghai, or even in China at large, as an initial stage for the ultimate construction of the Shanghai biobank. This will serve as an important foundation for subsequent developments. In addition, the information sharing services of the platform, which integrates data from clinical practice with biological samples, revitalizes work with biological samples in numerous hospitals, improves the efficacy of such work, promotes the production and application of scientific research, and enhances translational medicine in general. The construction of a platform

that integrates clinical practice and data on biological samples in Shanghai, and that shares this resource, will allow multicenter cooperation in major medical centers and will provide a platform for international exchange. Although this platform will not produce revenue directly, the outcomes of research using this platform will provide huge economic benefits.

## Conclusion

Construction of a platform that integrates clinical practice and data on biological samples requires a large collection of samples and efficient management procedures, including the ability to query the sample database, as well as robust statistical analysis, biomolecular information, and clinical information. Meeting these criteria will satisfy the requirements for research in translational medicine, and greatly improve the efficacy of scientific research.

We are establishing pilot projects before promoting a comprehensive strategy, and are planning to establish a comprehensive information-sharing service platform and a third-party service platform within the next three years. At that time, we will have connected 6 to 8 biobanks, with biological samples from 350,000 cases, and a central bank with 100,000 cases. This platform integrates clinical practice with data from biological samples, and will be an important resource for the study of major diseases, a critical tool for research in translational medicine, and an important resource for the biopharmaceutical industry. There will be an easy-to-use application process for diverse research projects that seek to improve the efficient use of biological samples. This approach will significantly improve the development of translational medicine in Shanghai and help to establish Shanghai as an international center for medical research and clinical care.

## Acknowledgments

We are thankful to Dr. Yutong Song for the valuable comments.

## Author Disclosure Statement

There are no financial conflicts of interest to disclose. This work was supported by Scientific Research Program of Science and Technology Commission of Shanghai Municipality (program numbers 12DZ2295000 and 13511504904).

## References

1. Ruan L, Song Y, Fan J, et al. The Shanghai Biobanking DNA Quality Control Program. *Biopreserv Biobank* 2014;12:246–254.
2. Riegman PH, Morente MM, Betsou F, de Blasio P, Geary P. Biobanking or better healthcare. *Mol Oncol* 2008;2:213–222.
3. Fan JL, Zhang J. Shanghai Biobank Network (SBN). *Biopreserv Biobank* 2011;9:133–133.
4. Dou XL, Cao ZL, Ying HY, et al. Translational medicine in China: Awareness is increasing and research is advancing. *Sci China Life Sci* 2013;56:861–862.
5. Campbell L, Betsou F, Garcia DL, et al. Development of the ISBER best practices for repositories: Collection, storage, retrieval and distribution of biological materials for research. *Biopreserv Biobank* 2012;10:232–233.

6. International Society for Biological and Environmental Repositories (ISBER). Best practices for repositories I: Collection, storage, and retrieval of human biological materials for research. *Cell Preserv Technol* 2005;3:5–48.
7. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nature Rev Genetics* 2007;8:151–157.
8. De Souza YG, Greenspan J. Biobanking past, present and future: Responsibilities and benefits. *AIDS* 2013;27:303–312.
9. Khleif SN, Doroshow JH, Hait WN. AACR-FDA-NCI Cancer Biomarkers collaborative consensus report: Advancing the use of biomarkers in cancer drug development. *Clin Cancer Res* 2010;16:3299–3318.
10. Zatloukal K, Hainaut P. Human tissue biobanks as instruments for drug discovery and development: impact on personalized medicine. *Biomarkers Med* 2010;4:895–903.
11. Meir K, Gaffney E, Simeon-Dubach D, et al., and Marble Arch International Working Group on Biobanking. The human face of biobank networks for translational research. *Biopreserv Biobank* 2011;9:279–285.
12. Palmirota R, Barbanti P, Ludovici G, et al. Establishment of a biorepository for migraine research: The experience of Interinstitutional Multidisciplinary BioBank (BioBIM). *Neurolog Sci* 2013;34:1659–1663.
13. Wang X, Wang E, Marincola FM. Translational medicine is developing in China: A new venue for collaboration. *J Translat Med* 2011;9:3.
14. Botti G, Franco R, Cantile M. Tumor biobanks in translational medicine. *J Translat Med* 2012;10:204.
15. Vaught J, Kelly A, Hewitt R. A review of international biobanks and networks: Success factors and key benchmarks. *Biopreserv Biobanking* 2009;7:143.
16. Sun W, Zhang K, Chen S-K, et al. Software as a service: An integration perspective. *Lecture Notes Computer Sci* 2007;4749:558–569.
17. Carretero J, García JD. The Internet of things: Connecting the world. *Personal Ubiquit Computing* 2014;18:445–447.

Address correspondence to:

*Dr. Guangjun Yu*  
*Shanghai Children's Hospital*  
*Shanghai Jiao Tong University*  
*No.24, Lane 1400, Beijing West Road*  
*Shanghai*  
*China*

*E-mail: yuguangjun@shdc.org.cn*