

# Establishment of a Network-Based Intra-Hospital Virtual Cancer Biobank

Lianhai Zhang,<sup>1-3\*</sup> Xiaojiang Wu,<sup>2,4\*</sup> Ying Hu,<sup>1\*</sup> Xiaohong Wang,<sup>1</sup> Zhonghu He,<sup>5</sup> Yuntao Xie,<sup>6</sup> Kaifeng Pan,<sup>7</sup> Ning Wang,<sup>8</sup> Zhihua Dong,<sup>9</sup> Lei Zhang,<sup>9</sup> and Jiafu Ji<sup>1</sup>

There is a growing interest in integrating biomaterial repositories into larger infrastructures in order to meet research demands. However, even for a single hospital or institute, where both population-based and multiple disease-based biobanks have existed for a long time, the integration of existing separate biobanks into a virtual cancer biobank is still challenging. The guidelines and procedures for biobanking are varied and not universally enforced or followed in separate biobanks. Within the last 2 years, we initiated a project to establish a centralized biobank facility in a common storage environment. Analyzing the challenges and interests of stakeholders for the biobanks, a working group comprised of representatives from the central and separate banks, ethic committees, and research administration offices reached an agreement to implement a central facility by following the ISBER best practices for biobanking, and including regular project reviews by the ethical and scientific boards. Furthermore, by implementing a modified minimum information system with biobank data sharing, a network based intra-hospital virtual cancer bank was established to facilitate sharing information of samples held by separate banks. Meanwhile, this virtual biobank network, which has integrated patient information from hospital health care systems, will gradually integrate follow-up information from the cancer registry office and data from epidemiology studies, providing controlled access for sample providers and resource users. In the future, this infrastructure designed for a single hospital may be helpful for building a broader virtual network for data and specimen exchanges.

## Introduction

IT IS WELL KNOWN THAT THE MAIN PURPOSE of biobanks is quick and efficient access to collections of human biological material and associated clinical information.<sup>1</sup> When properly collected and well-preserved, these resources reduce the time for achieving conclusive results and promote scientific progress. In order to satisfy the growing interest in sample and data acquisition from the research community, one approach is to establish and sustain a large clinical data integrated biobank network between many organizations and stakeholders. Therefore, collaboration and cooperation among network members must extend beyond disputes over sample and data ownership and systems to enable the exchange of data and samples and to form strategic alliances.<sup>2,3</sup>

Despite advances in biobanking, significant limitations remain and restrict the progression of research efforts. Building a virtual biobank network among different cancer biobanks, by identifying and balancing the interests of

stakeholders, is still a significant challenge. In a single hospital such as the one described in this report, where multiple biobanks have existed for a long time, separate biobanks significantly increase the difficulty for implementing biobank quality management and standardization.

The purpose of this article is to provide an overview of the current situation of biobanking in our hospital, to describe a harmonious solution reached through stakeholder analysis to establish a centralized virtual biobank, to include sharing data between separate banks and integrating clinical data, in order to facilitate collaboration and networking among the existing banks.

### *Biobank situation overview: Coexistence of separate banks*

Peking University Cancer Hospital (PUCH) is a 790-bed public hospital in North China, incorporating a large institute focusing on translational cancer research. In PUCH, clinical

<sup>1</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Cancer Biobank and Departments of <sup>2</sup>Surgery, <sup>5</sup>Genetics, <sup>6</sup>Breast Cancer, and <sup>7</sup>Epidemiology, <sup>3</sup>Research Administration Office, <sup>4</sup>Ethics Committee, <sup>8</sup>Beijing Office for Cancer Prevention and Control, Peking University Cancer Hospital and Institute, Beijing, China.

<sup>9</sup>Goodwill Information Technology Co. Ltd, Beijing, China.

\*LZ, XW, and YH contributed equally to this work.

investigations into specific cancers and research into basic oncologic processes have proceeded hand in hand for many years.

In 1996 an initiative was started to establish a biobank as a common resource for future cancer research in the hospital and its community, to support clinical and basic investigations into cancer genetics, oncogenesis, pathogenesis, and experimental therapeutics. Initially, it was proposed that the biobank would be operated by the Department of Surgery and Pathology, primarily to collect surgically removed tissue samples and matched peripheral blood from patients with gastric, colorectal, liver, esophagus, lung, and other cancers. In 2006, the central biobank published a book about its standard operating procedures (SOPs) for biobanking. In 2009, PUCH biobank was admitted as a member of the Biobank Branch of the China Medicinal Biotech Association, and later in 2013 as a member of the Beijing Biobank Alliance. Starting in 2011, the biobank began to build its own database using its servers and barcoding its collected samples. And during the past 2 years, the central biobank has engaged commercial IT support to integrate clinical data, and has followed ISBER best practices<sup>4</sup> as an additional step toward quality management.

Along with the construction of the central biobank, two longitudinal epidemiology studies in Shandong Province and Henan Province collected a large number of specimens in rural areas for cancer epidemiology and etiology research projects. The frozen biopsy tissues and blood, and some extracted derivatives such as DNA and RNA were continuously transported back to PUCH and stored in  $-80^{\circ}\text{C}$  freezers in their respective departments inside the institute building. In addition, some independent biobanks collecting samples from the dominant cancer types seen in their institutions, such as breast cancer, lung cancer, melanoma, and hepato-pancreato-biliary cancer, were established by respective principal investigators with grant support from a variety of sources. The chronology of the development of the central biobank and separate specimen storage banks and their respective sizes are shown in Figure 1.

*Challenges of separate banks for appropriate governance*

The coexistence of the central biobank and separate banks remained in place for many years. As shown in Table 1,

each bank was operated by a respective department following different (or even no fixed) SOPs for cancer specimen collection, and each submitted its research proposals separately to a central ethical review board, scientific board, and research administration office. There was no strict oversight. The location of these banks was scattered in the hospital and samples were at risk without redundant power supplies, alarm systems, and proper monitoring. Some banks even stored sample information and clinical data in personal computers, and as a result, these data were isolated from clinical information and follow-up information in the hospital health care system. In our opinion, these inconsistencies with best practices were mainly influenced by such factors as inadequate funding to support Information Technology (IT) infrastructure, lack of personnel with biospecimen management experience, and the lack of availability of professional and customized software services.<sup>5</sup>

The long-time coexistence of these separate operations indicates that on one hand the services of the central biobank may not fully accommodate the demands of researchers. However, on the other hand and more importantly, conflicts of interests exist between separate banks with the central biobank or among clinical practitioners over publication and attribution, and opportunity for financial remuneration. As a result, biological materials and related data collected in separate banks could not be considered an “open” resource for the whole research community. Realizing the extensive number of stakeholders involved, a working group, comprised of representatives from central biobanks and separate biobanks (including respective clinical practitioners and biobank staff and related lab staff) and the ethics committee and research administration office, decided to undertake a proper analysis, and to reach a coordinated solution for constructing a centralized biobank.

**Materials and Methods**

A systematic stakeholder analysis with respect to biobank projects is key to the initiation of a central biobank for harmonization of policies and procedures and long-term sustainability. We followed the methodology from a previous publication as a key reference for this analysis.<sup>6</sup> In our setting, which is comprised of a hospital and related research institute, we first established a working group consisting of a number of stakeholders, such as clinical

**FIG. 1.** The development of a central biobank and separate specimen storage banks.

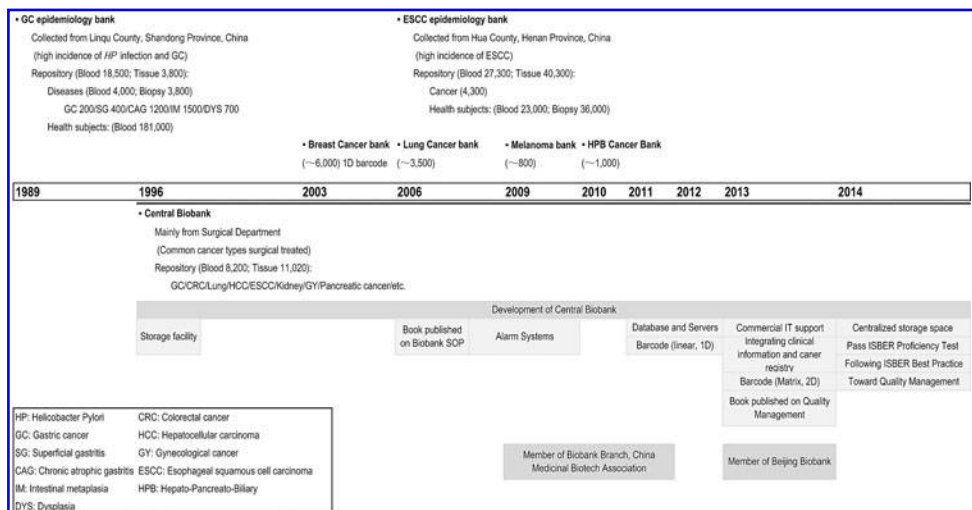


TABLE 1. CHALLENGES FOR SEPARATE BANKS AND POTENTIAL SOLUTIONS TO ESTABLISH CENTRALIZED BIOBANK

Item	Challenges	Proposal for addressing
1	Lack of supervision from the central boards	Centralization of administration: practices of each bank should be reviewed by an independent ethical review board, scientific board, and research administration office
2	Lack of SOPs or guidelines followed	Quality Management: to implement guidelines and SOPs based on Best Practices
3	Conflict of interests between separate banks and centralized hospital management	Publication acknowledgement and grant compensation for sample providers or banks
4	Isolated databases maintained in personal computers; Data is subject to physical loss or damage; No integration to clinical information	IT infrastructure support: Integration of clinical information and data from cancer registry; Providing controlled access to authorized users
5	Lack of monitoring for storage equipment	Centralized storage facility: to provide power supply, alarm system, bar coding equipment, etc.

practitioners, researchers, research institutions, regulatory bodies, and funders, who potentially have a strong influence upon operation of a biobank, or are affected by biobanking policy. The entire stakeholder analysis was then conducted by the working group in a series of roundtable discussions, which resulted in an agreement among the participants.

In order to develop a standard for data sets, allowing all biobanks to share information about the samples they hold, the working group decided on the minimum information for biobank data sharing (MIABIS) after discussing the existing data standards<sup>7,8</sup> available to biobanks.

**Results**

*Stakeholder analysis*

The stakeholder analysis consisted of five steps: identification, attribution, prioritizing, engagement, and monitoring.<sup>6</sup> At the initiation phase, we completed the first four steps. In the identification step, based on the environment and regulations faced by hospitals in China, we formed a modified generic model (from Reference 6) with the following nine stakeholder groups: (i) political bodies; (ii) hospital owners; (iii) funding sources; (iv) our own organization (the hospital and institute); (v) public administrators; (vi) patients and their relatives; (vii) biobank facilities;

(viii) resource users; (ix) and others, such as media, charities, and public organizations (Fig. 2).

After attributing values to stakeholders by using a “power-interest matrix,” the working group decided to target some stakeholders deemed to have the combination of high power and high interest (i.e., patients and their relatives, hospitals and institutes, central biobanks, and clinical practitioners engaged in activities from the separate banks). Following discussion with the major separate storage banks, the working group welcomed collaboration of all stakeholders, and was especially open to discussion of concerns about ownership of the samples held by separate banks. Eventually, a high-level decision was reached on the establishment of a centralized biobank and a formal working group.

The consensus that resulted from roundtable discussion includes:

- (i) Emphasis on the central administration: practices for each biobank should be reviewed by the ethical review board, scientific board, and research administration office.
- (ii) Maintain the investment by the hospital for the centralized biobank facility, to provide sufficient infrastructure to support existing separate banks.
- (iii) Establish a virtual centralized biobank. Separate biobanks will voluntarily join the central biobank, follow the best practices for biobanking, agree on minimum

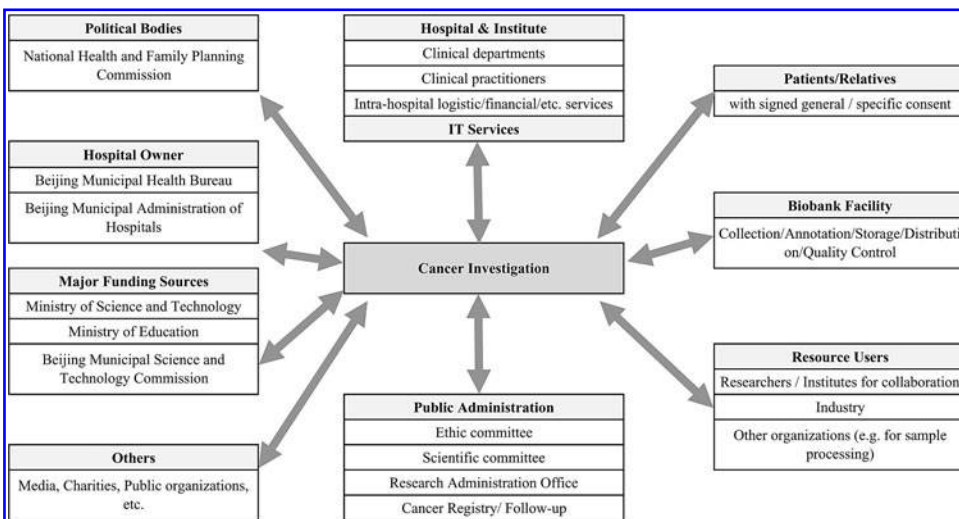


FIG. 2. Identification of stakeholders and their interests for cancer investigation.

information about biobank data sharing for communication between the central biobank and separate biobanks, and for later exchanges of data with external organizations. Meanwhile, separate biobanks will maintain priority for their “own” sample use.

- (iv) Establish a network to facilitate availability of controlled-access to integrated clinical information from the hospital health care system, and follow-up information from the cancer registry office; and gradually integrate epidemiology databases.

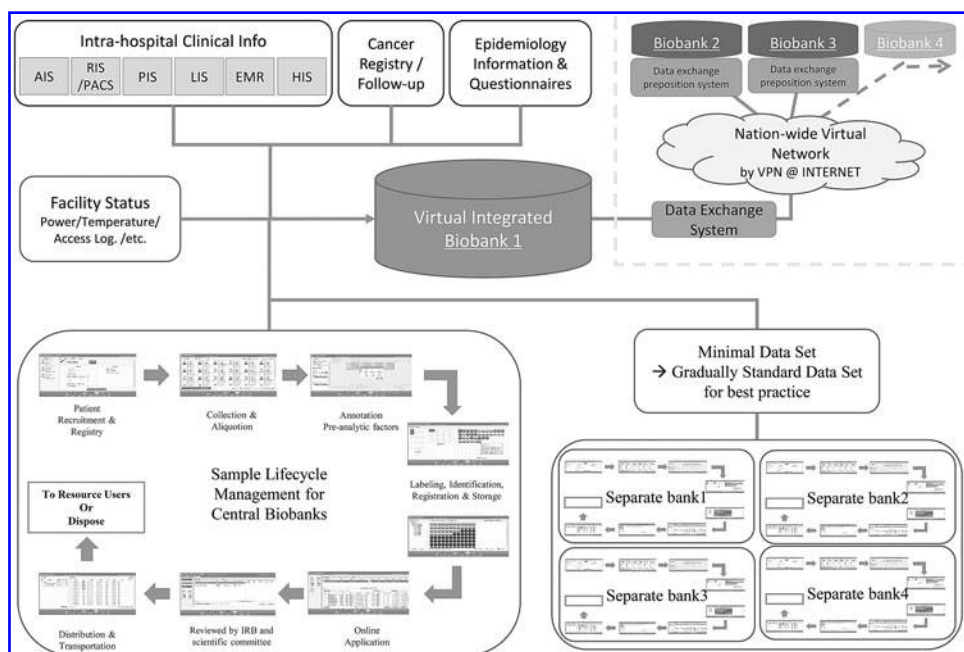
### *Establishment of a network-based virtual biobank*

To establish a network-based intra-hospital virtual biobank, first a standard data set should be developed to allow all biobanks to integrate and share information about the samples they hold, facilitating the possibility of a common portal for researchers or other resource users to find suitable samples.<sup>8</sup> To facilitate this communication, the working group agreed on the minimum information about biobank data sharing from existing data standards available to

TABLE 2. MINIMUM INFORMATION ABOUT BIOBANK FOR DATA SHARING

<i>Field name</i>	<i>Format</i>	<i>Description</i>
(a) Collection		
Sample collection/Study ID	Free text	Unique study ID or acronym for the sample collection
Study name	Free text	Name of the study
Description of collection	Free text	Description of study aim (max. 200 characters)
Type of collection	Free text	Case-control/Cohort/Cross-section/Longitudinal/Quality Control/Population-based/Disease-specific/Other
Collection start	yyyy-mm-dd	Date when the sample collection starts
Collection end	yyyy-mm-dd	Date when the sample collection ends, if applicable
Number of current samples individuals	Integer	Number of individual biological samples in the collection at the date of last updated
Last update	yyyy-mm-dd	Date when the sample collection was last updated
Biobank ID	Free text	86_PUCH (country code_acronym of hospital)
(b) Patient		
Patient's anonymized ID	Free text	An anonymized identifier for the patient
Patient's sex	Free text	Sex of the sample donor
Year of Birth	yyyy	Year of Birth
(c) Patient Diagnosis		
Patient diagnosis	Free text	Diagnosis (table) for sample donors (ICD-10)
Diagnosis date	yyyy-mm-dd	Date when the diagnosis was made
(d) Sample		
Date sample collected	yyyy-mm-dd	Date when samples were collected
Consent details	Free text	Consent details of the collection
Patient diagnosis	Free text	Diagnosis that is most relevant to the samples (ICD-10)
Age	Free text	The age of the patient at the time of the sample collection
(e) Specimen		
Organ Code	Free text	ICD-10 plus other information
(f) Tissue sample		
Barcoded ID	Free text	An anonymized identifier for the sample
Material type	Free text	Type of the sample (e.g., solid tissue, biopsy)
Pathological assessment	Free text	Pathological assessment of the samples
Storage temperature	Temperature	The storage temperature in Celsius
Aliquot size	Free text	The average size of aliquots (e.g., in microgram)
Aliquot amount	Integer	Amount of aliquot in storage
(g) Liquid sample		
Barcoded ID	Free text	An anonymized identifier for the sample
Material type	Free text	Type of the sample (e.g., whole blood, serum, plasma, urine)
Storage temperature	Temperature	The storage temperature in Celsius
Aliquot size	Free text	The average size of aliquots (e.g., in microliter)
Aliquot amount	Integer	Amount of aliquot in storage

\*For samples of the central biobank from Oct. 2013, the details for pre-analytic factors are recorded when registering sample information with software and a SPREC<sup>14,15</sup> code is automatically generated by system.



**FIG. 3.** The design of a network-based intra-hospital virtual cancer biobank.

biobanks (modified from references 7, 8, and Table 2), based on opinions from all participants in the working group. This minimal data set, rather than a standard one, would be easily applicable to all biobanks, independent of any cancer types. Some items of this data set, such as dates of birth, sample collection, and diagnosis will be partially de-identified to prevent tracing back to the patients.

Since the project was just in the initiation phase, the working group decided not to establish a data standard for other information providers, such as the cancer registry, follow-up, and epidemiology database. In consideration of potential integration into a nation-wide or even global network, we will further negotiate with external experts for extension of our data set to those other stakeholders.

Therefore, even though each separate bank uses a different database for sample registration and annotation, the virtual biobank collecting and maintaining a minimal data set will provide a mechanism for biobanks to communicate with each other about the sample availability. Thus the virtual biobank will be able to provide a common public portal for researchers to search for the most suitable samples available for their research.

The integration of sample lifecycle management data with clinical information has been solved by connecting the central biobank database to the hospital health care system, in order to retrieve data from the laboratory information system (LIS), radiology information system/picture archiving and communications systems (RIS/PACS), electronic health records (EHR), and other data. However, due to the historic lack of structured data and standardization in the healthcare system, and even noticeable variations of existing structured data among different cancer types and treatment algorithms, the “real” integration of electronic health content into sample information is still problematic. Also, as the project is still in the initiation phase, the integration of cancer registry, follow-up data and epidemiology data is ongoing.

In the long run, as shown in Figure 3, this network-based intra-hospital virtual cancer biobank will try to coordinate

planning of a nation-wide virtual network for data and sample exchanges.

**Discussion**

This situation and challenges for our hospital to integrate existing separate biobanks into a virtual cancer biobank may also be faced by other hospitals or institutes in China. Establishing a centralized biobank facility by a high-level decision-making process can at least consolidate all the biobanking equipment in a common storage environment. However, for long-term sustainable development and harmonization of the whole community, analyzing the interests of biobank stakeholders and discussion among their representatives can produce a consensus on appropriate biobanking and following best practices. Several well-known biobank initiatives and projects have engaged with stakeholders during the planning phase, for example, the U.S. National Cancer Institute, the pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI), and the UK Biobank, in order to build harmonious environments.<sup>3,9,10</sup>

The agreement on minimum information about biobank data sharing by all the participating banks helps establish a communication standard to build a network-based intra-hospital virtual biobank among separate banks.<sup>7</sup> Together with the integrated clinical information, this virtual biobank will help researchers reduce the time for investigation and promote scientific progress. And in this new environment, the centralized biobank facility personnel will assist new biobanks by providing advice on planning, setup, protocols, database services, and other services to enhance operations, such as management advice, ethics and regulatory application preparation, and even budget development.

There are still some unresolved challenges for establishing this centralized biobank. As clinical practitioners will have various levels of interests and opinions, some violation of best practices and disputes may only be solved through collaboration and partnership with major stakeholders. And

the challenges related to public perceptions of biobanks, bioethical considerations, personal data protection, and authorization for clinical and genetic data access will continue for a long time in China.<sup>11–13</sup> We do hope our experiences in infrastructure construction of this integrated virtual biobank may be helpful for establishing a plan for a nationwide virtual network for data and specimen exchanges.

## Conclusion

In this article we describe how to establish a network based intra-hospital virtual cancer biobank in a single hospital or institute, where multiple disease-based biobanks have existed for many years, after implementing a centralized biobank in a common storage environment. In the process of building this network, the whole community reached an agreement to follow the best practices of biobanking and accept minimum standards for data sharing. This infrastructure design for a virtual biobank network, which integrates information from the patient, will result in a broader virtual network for data and even specimen exchanges, to expedite clinical and basic investigations in the future.

## Author Disclosure Statement

The authors declare the following competing financial interests: LZ and ZD are employees of Goodwill Information Technology Co. Ltd. This study is partially supported by National Key Technology R&D Program (2012AA02A203-B01, 2012AA02A504-B01, 2012AA020101, 2015ZX09307-001) and Beijing Municipal Science & Technology Commission (Z121100007512010, D131100005313010).

## References

1. De Souza YG, Greenspan JS. Biobanking past, present and future: responsibilities and benefits. *AIDS* 2013;27:303–312.
2. Gottweis H, Lauss G. Biobank governance: Heterogeneous modes of ordering and democratization. *J Community Genet* 2012;3:61–72.
3. Harris JR, Burton P, Knoppers BM, et al. Toward a roadmap in global biobanking for health. *Eur J Hum Genet* 2012;20:1105–1111.
4. 2012 best practices for repositories collection, storage, retrieval, and distribution of biological materials for research international society for biological and environmental repositories. *Biopreserv Biobank* 2012;10:79–161.
5. Nussbeck SY, Skrowny D, O'Donoghue S, et al. How to design biospecimen identifiers and integrate relevant functionalities into your biospecimen management system. *Biopreserv Biobank* 2014;12:199–205.
6. Bjugn R, Casati B. Stakeholder analysis: A useful tool for biobank planning. *Biopreserv Biobank* 2012;10:239–244.
7. Norlin L, Fransson MN, Eriksson M, et al. A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank* 2012;10:343–348.
8. Quinlan PR, Mistry G, Bullbeck H, et al. A data standard for sourcing fit-for-purpose biological samples in an integrated virtual network of biobanks. *Biopreserv Biobank* 2014;12:184–191.
9. Myles R, Massett HA, Comey G, et al. Stakeholder research on biospecimen needs and reactions to the development of a national cancer human biobank by the National Cancer Institute. *J Natl Cancer Inst Monogr* 2011;2011:16–23.
10. Mitchell R. National biobanks: Clinical labor, risk production, and the creation of biovalue. *Sci Technol Human Values* 2010;35:330–355.
11. Liu M, Hu Q. A proposed approach to informed consent for biobanks in China. *Bioethics* 2014;28:181–186.
12. Ji X, Zhao XM, Jiang JJ, et al. Clinical biobanks, from the world to china. *Biomed Environ Sci* 2014;27:481–483.
13. Chen H, Gottweis H, Starkbaum J. Public perceptions of biobanks in China: A focus group study. *Biopreserv Biobank* 2013;11:267–271.
14. Lehmann S, Guadagni F, Moore H, et al. Standard pre-analytical coding for biospecimens: Review and implementation of the Sample PREanalytical Code (SPREC). *Biopreserv Biobank* 2012;10:366–374.
15. Betsou F, Lehmann S, Ashton G, et al. Standard pre-analytical coding for biospecimens: Defining the sample PREanalytical code. *Cancer Epidemiol Biomarkers Prev* 2010;19:1004–1011.

Address correspondence to:

*Dr. Jiafu Ji  
Department of Surgery  
Peking University Cancer Hospital and Institute  
52 Fucheng Road  
Haidian District  
Beijing 100142  
China*

*E-mail: jiafuj@gmail.com*