

# Centralized Biorepositories for Genetic and Genomic Research

Geoffrey S. Ginsburg, MD, PhD

Thomas W. Burke, PhD

Phillip Febbo, MD

**M**EDICAL PRACTICE IS BEING TRANSFORMED BY MOLECULAR analyses of biological samples that provide prognostic and diagnostic information for clinical care. The analytical methods and technologies that have accompanied the sequencing of the human genome have been the subject of research on their accuracy, reliability, and precision, but far less attention has been paid to the storage methods and archiving of the specimens required for analysis. Indeed, well-annotated biospecimen collections have enabled the recent identification of genes and genetic loci thought to contribute to susceptibility for several complex diseases.<sup>1</sup> Some of these genome-wide association studies were based on government-sponsored, centralized collections.<sup>2-4</sup> However, biorepositories with standardized procedures, informatics, and embedded regulatory compliance remain rare and the state of storage of human biospecimens is often decentralized and poorly organized at many of the US medical centers.

Human tissue has been collected and stored at medical and research institutions in the United States for more than 100 years.<sup>5</sup> By 2000, more than 300 million human biospecimens had been stored in freezers across the United States with an estimated accrual rate of an additional 20 million specimens annually.<sup>6</sup> In 2004, the National Cancer Institute estimated that it spends more than \$50 million yearly on banking samples from cancer patients as part of 125 funded research programs and projects.<sup>7</sup> Academic medical centers historically have discharged the responsibility of biospecimen sample and data storage to individual investigators who have developed “home remedies” to meet specific research objectives, often with limited resources and inconsistent funding.

The result is a decentralized system of ad hoc solutions that, despite facilitating great advances, often lack standards, are inefficient, and create a liability for the institutions and their researchers. Most institutions today cannot readily access a list of samples stored on institutional premises, let alone ascertain the conditions under which they are stored or who donated them. With a greater reliance on sample-derived data for genetic and genomic re-

search and clinical care, improved standards and informatics for sample procurement, storage, and analysis are required to maximize the value of tissue collection for research participants, investigators, medical centers, and funding agencies.

## Drivers for Centralization

The main drivers of centralization of biospecimen collections are the direction of medical research, efficiencies gained through economies of scale, improved quality and reliability, limitation of liability, and the provision of quality patient care. The scope of today’s genetic and genomic research requires comparisons of data and results across multiple studies for replication and/or validation. A uniform approach at every level of a centralized repository from sample information systems to standards for procurement and storage both within and between academic medical centers will be key to facilitate and optimize the use of resources in genomic research.

Academic medical centers already unwittingly bear a significant amount of the cost for biorepository activities. Cost drivers of sample banking include freezers, space, investigator and technician time, and computer software for tracking and annotation. Summed across a medical center, millions of dollars per year can be spent on biobanking, and much of this cost may be “hidden.” Moreover, these costs may increase several-fold if samples are rendered useless by a suboptimal system for tracking or processing them or if investigators need to reascertain sample collections that unknowingly exist at local or networked institutions. Institutional investment in sample repository infrastructure would create synergies and introduce efficiencies to produce economies of scale that not only decrease the financial burden to the individual investigator and to the institution itself but also reduce the barriers to entry for early stage investigators and improve opportunities for research funding.

Heterogeneity in regulatory oversight of sample collections can also potentially place host institutions and their investigators at significant legal risk and may compromise the value of biospecimen resources for future translational

**Author Affiliation:** Duke University, Institute for Genome Sciences & Policy, Durham, North Carolina.

**Corresponding Author:** Geoffrey S. Ginsburg, MD, PhD, Center for Genomic Medicine, Duke Institute for Genome Sciences & Policy, Box 3382, 101 Science Dr, Durham, NC 27708 (geoffrey.ginsburg@duke.edu).

### Box. Stakeholder Needs Fulfilled by a Centralized, Institutional Biorepository

#### Research participants

- Clarify use of samples for research
- Optimize use of donated specimens
- Ensure protection of confidentiality

#### Investigators

- Biospecimens collected in a defined manner from a diversity of disease types and populations based on a continual review of researcher needs
- Timely and widespread access to biospecimens and associated data through a centralized peer-review process
- Clear guidelines for acceptable sample use

#### Limit liability

#### Medical centers

- Provide valuable resource for institutional researchers
- Optimize use of research space and increase opportunities for research funding
- Enable provision of high-quality care
- Facilitate research
- Limit liability

#### Funding agencies

- Minimize inefficiency in monies spent on biobanking
- Ensure development of functional and enduring resources
- Improve quality and access to samples

research.<sup>8</sup> Chain of custody and adherence to the research permitted under informed consent must be a priority; in addition, privacy and safety must be ensured for research participants. Currently, individual investigators and institutional review boards are charged to ensure that sample-based research is being conducted in concert with informed consent language but the compartmentalized nature of current biorepository practices challenges the successful discharge of regulatory responsibilities. In addition, when investigators leave institutions there are often no clear procedures for transferring this responsibility.<sup>9</sup> Centralized informatics systems coupled with an oversight system will help uphold legal responsibilities and high ethical standards while enabling efficient access to samples.

### Development and Implementation of Standards

The International Society for Biological and Environmental Repositories and the National Cancer Institute's Office of Biorepositories and Biospecimen Research have released first generation guidelines and best practices.<sup>10</sup> In addition, the National Cancer Institute has begun shaping a new generation of standard operating procedures for all aspects of activities involving human biospecimens along with data-

driven quality indicators for biorepositories and legacy specimen collections.

The development of commonly accepted and readily implemented standards is essential for future research and patient care and to benefit all stakeholders (BOX). However, the adoption of such standards may present significant practical and cultural challenges to institutions and investigators and will require both trust and engagement of senior leadership and stakeholders. Investigators need to be educated about and have time to adopt new methods. It may take several years to integrate these prior to becoming institutional policy. How an institution enforces procedures will require discussion, but tying adherence to guidelines for the approval process for human subject research is one possibility.

Three priority areas for investment to bring existing repositories into compliance are informed consent, informatics systems, and governance.

**Informed Consent.** A significant challenge of informed consent in the context of banking specimens is the need to use samples for future research purposes. The value of a specimen is increased by its potential use in future research. While it remains untenable to fully anticipate the specifics of future use, it is critical to inform prospective participants about the potential for future use, track consent information to ensure appropriate use, and adopt standards that address when and how results should be communicated back to the participant. Developing and providing standardized informed consent templates with modular provisions for specific human specimens and data allows investigators to adapt the template and establishes common approaches to issues such as future use.

**Informatics Systems.** Robust informatics systems are fundamental to optimizing the value of biospecimen collections.<sup>11</sup> The upfront investment in centralized systems that facilitate the tracking of specimens and annotation of data collected under specific provisions of consent will maximize the future potential of samples. In addition, the architecture of data storage will need to address chain of custody tracking, database federation, and standardized nomenclature.

Repository informatics systems ideally will provide chain-of-custody tracking for each specimen and derived material through all stages of acquisition, processing, handling, banking, quality control, distribution, and ultimate disposition. By combining this capability with appropriate consent processes, samples can be further annotated through a secure intermediary to clinical data including patient demographic information, symptoms, physical findings, images, treatments, historical clinical events, and future outcomes. This requires an informatics infrastructure that is amenable to federation or integration with disparate and dynamic types of data and databases while ensuring privacy and data security to the individual donors.

For researchers to identify specimens that meet their research needs, the systems for sample tracking and annotation must have a shared standard nomenclature. The development of common data elements is a major goal of both

caBIG (cancer Biomedical Informatics Grid) of the National Cancer Institute and the National Library of Medicine's Unified Medical Language System Metathesaurus, which includes SNOMED CT and the National Cancer Institute's Thesaurus. These efforts are essential toward supporting molecular research as well as health services research and routine clinical care, and necessary for the development of integrated biorepositories.<sup>12-14</sup>

A major goal of centralization and standardization of biobanking activities is to facilitate collaboration and to enable efficient biospecimen research. When implemented properly with the appropriate restrictions and safeguards, the centralized informatics system for a biorepository facilitates efficiency in research by functioning as a "transparent safe-deposit box" in which samples can be "viewed," but at the same time remain private and secure. Such a system would be effective and powerful as investigators need not enroll research participants and collect specimens when appropriate specimens already exist and are available for the planned research. A query system would allow viewing all samples in the inventory to all investigators, but access to specific samples would be dictated by the principle investigator responsible for them or a specimen review board in the case of specimen collections supported by a group of investigators. The availability of specific specimens for research is dependent on restrictions in the informed consent and study protocol under which a specimen was collected and the specimen custodian(s).

**Governance.** Any institutional core resource that will be made available to all researchers requires a governing board with key stakeholder representation including clinicians, researchers, patients, technologists, and health system leadership. This represents a shift from the view that individual investigators or research consortia govern the samples and reflects the reality that such repositories require cooperation among research participants, their health care system, and the many individuals with special scientific or technical expertise. Academic medical centers that partner with the institution's health system or with other health systems will create numerous opportunities to develop diverse populations of biospecimens for study and may serve to engage the local communities in the organization's banking efforts. Governance bodies will need to reconcile biorepository practices with state and local laws and potentially conflicting local institutional policies.

### An Imperative for Genomic Research

Biospecimen banking is critical for the attainment of a state of predictive, preventive, and personalized medicine. It is an intensive and growing activity, and investments made today in biorepositories are projected to be used for decades and will have significant consequences for determining both the opportunities and the limits of future research, and ultimately medical diagnosis and therapies. The govern-

ments of several countries, including the United Kingdom, Iceland, Estonia, and Singapore, have made a commitment to centralized tissue banking on a national scale. Academic medical centers in the United States need to consider mechanisms with which to provide the appropriate infrastructure to facilitate collaborative genomic research even in the absence of a formal government initiative to do so.

The historically loose organization and lack of standardization of biorepositories, when coupled with the scale demanded by modern genomic methods, presents serious challenges to our ability to leverage this important resource to the fullest extent. Most individual investigators cannot harness the disparate resources required for standardized biobanking activities, which is becoming vital to genomic research and the transformation of health care from intervention to prevention. Institutional investments in comprehensive systems for archiving clinical sample inventories will likely define tomorrow's leading academic medical centers and enable their investigators to provide optimal and molecularly guided care to their patients.

**Financial Disclosures:** None reported.

**Funding/Support:** This work was supported by the Duke Institute for Genome Sciences & Policy and the Duke Translational Medicine Institute.

**Role of the Sponsor:** The sponsors were not involved in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

**Additional Contributions:** We are grateful to Isaac Chan, BA, and Rajesh Dash, MD, and Robert Califf, MD (all from Duke University) for their contributions to this Commentary. Each was compensated as part of his salary.

### REFERENCES

1. Topol EJ, Murray SS, Frazer KA. The genomics gold rush. *JAMA*. 2007;298(2):218-221.
2. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320-323.
3. Helgadottir A, Manolescu A, Thorleifsson G, et al. The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. *Nat Genet*. 2004;36(3):233-239.
4. Stefansson H, Sigurdsson E, Steinthorsdottir V, et al. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*. 2002;71(4):877-892.
5. Eisman E, Bloom G, Brower J, Clancy N, Olmsted S. *Case Studies of Existing Human Tissue Repositories: "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era*. Santa Monica, CA: RAND Corporation; 2003.
6. Eisman E, Haga S. *A Handbook of Human Tissue Sources: a National Resource of Human Tissue Samples*. Santa Monica, CA: RAND Corporation; 2000.
7. Department of Health and Human Services, Public Health Service, National Cancer Institute, 132nd National Cancer Advisory Board. Summary of Meeting, November 30–December 1, 2004. [http://deainfo.nci.nih.gov/Advisory/ncab/132\\_1104/30nov04mins.pdf](http://deainfo.nci.nih.gov/Advisory/ncab/132_1104/30nov04mins.pdf). Accessibility verified February 19, 2008.
8. Evans BJ, Meslin EM. Encouraging translational research through harmonization of FDA and common rule informed consent requirements for research with banked specimens. *J Leg Med*. 2006;27(2):119-166.
9. Charo RA. Body of research-ownership and use of human tissue. *N Engl J Med*. 2006;355(15):1517-1519.
10. ISBER. Best practices for repositories I: collection, storage and retrieval of human biological materials for research. *Cell Preserv Technol*. 2005;3(1):5-48.
11. Patel AA, Gilbertson JR, Parwani AV, et al. An informatics model for tissue banks—lessons learned from the Cooperative Prostate Cancer Tissue Resource. *BMC Cancer*. 2006;6:120.
12. National Cancer Institute. Terminology browser. <http://nciterns.nci.nih.gov/NCIBrowser/Dictionary.do>. Accessed February 15, 2008.
13. National Cancer Institute. Welcome to the caBIG™ community Web site. <https://cabig.nci.nih.gov>. Accessed February 15, 2008.
14. National Cancer Institute. Unified Medical Language System Metathesaurus fact sheet. <http://www.nlm.nih.gov/pubs/factsheets/u/mlsmeta.html>. Accessed February 15, 2008.