The Continuing

Evolution of Biobanking

MARCHING TOWARD STANDARDIZATION

Fluidigm®

Recognizing and Maintaining the Important Combination of **Proper Sample Identification** and High Sample Quality

About this White Paper

A principal investigator's research on biomarkers for cardiovascular disease. A pharmaceutical company screening for responders versus non-responders of a particular drug. A population-wide study looking at causal variants associated with alcoholism and drug addiction. While each of these scientific efforts is distinct and multifaceted, all rely on access to high-quality specimens.

As a result, there is an increasing global reliance on biorepositories, which serve a critical function for many efforts in clinical and basic research. The aim of this white paper is to focus on a key biobank issue—specifically: the importance of proper sample identification and high sample quality.

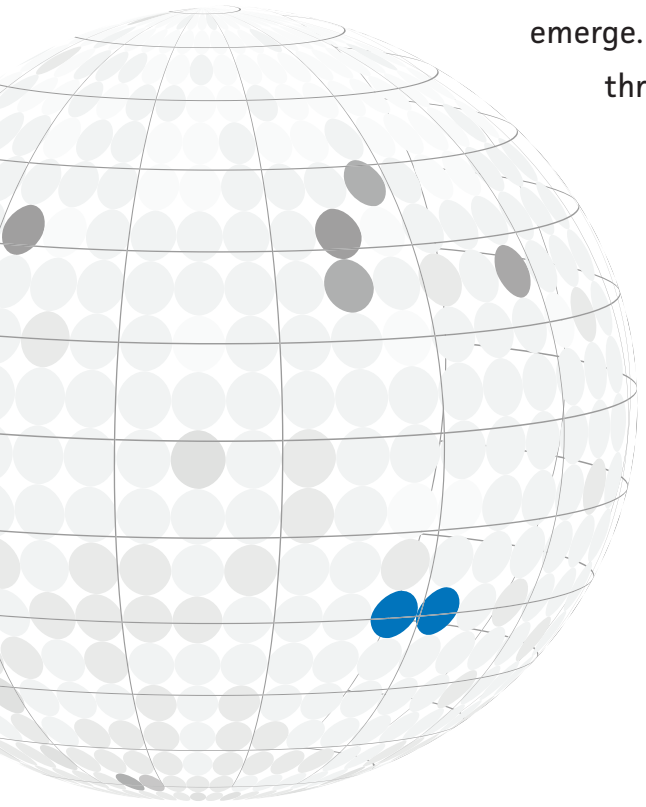
Index

1 The **Challenge**

The last two decades have witnessed the development of, and increasing dependence on, biorepositories for advancing basic, translational, and clinical sciences. This reliance can be explained by the explosion of genomic data, which has lent considerable support to the understanding that many diseases are heterogeneous. For example, a single condition such as breast cancer is in actuality a complex disease with tumors that can be divided into multiple subtypes based on specific molecular signatures.[i]

For the researcher, this understanding means that samples needed for studying breast cancer should attempt to represent each subtype. In addition, scientists now appreciate more than ever that diseases are population-specific, and international collaboration amongst investigators is increasing.

Therefore, leading biobanks should aim to store samples from patients around the globe[ii] in order to have a fuller representation of populations and diseases. For the biobank, the natural consequence of disease and population heterogeneity is that more and more samples are being required, and these samples become highly valued for a very specific set of characteristics; however, as the number of samples deposited into a successful biobank continues to increase, a number of challenges emerge. For example, as it is very common for samples to pass through many hands, mislabeling, and thus identity errors, can readily be introduced.



Combined with additional sources of sample error to be discussed below, the net effect for biobanks is that operational risks increase. How can a leading biobank manage these risks?

Let's take a closer look at the problem in order to understand the potential solution.

2 Root Cause Analysis: **Diverse sample origins** are the main source of quality variance.

From the private sector to academic collaborations, biobank stakeholders at all levels can agree that the quality of the samples and related data should be consistent and standardized. Today, biospecimen quality assurance (QA) and quality control (QC) systems are part of the larger ecosystem of biobank best practices, which also include technical, legal, ethical, and managerial issues.[iii]

39%

Percentage of cancer researchers who reported difficulty obtaining biospecimens of **adequate number**

47%

Percentage of cancer researchers who reported difficulty obtaining biospecimens of **adequate quality**

As we look a level deeper, these bank-wide QA/QC systems and processes are ultimately built on individually deposited samples. However, if individual samples deposited do not meet established standards, then one can only conclude that bank-wide quality, reputation, and ultimately scientific progress will suffer. In support of this assertion, the National Cancer Institute conducted a survey evaluating the biospecimen needs of cancer researchers.[iv]

The authors found that "large proportions reported difficulty obtaining biospecimens of adequate numbers (39%) and quality (47%)." In addition, "low-quality biospecimens resulted in 60% questioning their findings and 81% limiting the scope of their work."

For the biobank administrator, the logical question then emerges: What are the fundamental traits necessary in an individual repository sample to maintain an appropriate standard?

The answer is that, at the minimum, each individual sample should be properly identified. Ideally, this extends beyond the label on the outside of the tube, but also includes the identity and quality of the genetic biomaterial (e.g., DNA) inside the tube. As alluded to above, this critical task of controlling sample identification and individual sample quality is being pushed to the very forefront of biobank management as the number of samples being deposited continues to grow.

One estimate, as of 2008, puts the number of deposits at 20 million samples per year, in the US alone.[v] Ironically, the challenge of maintaining the integrity of the data, namely the sheer number of samples being deposited from a wide mix of populations, is the very thing that makes these repositories so attractive to researchers.

It is generally accepted that a major challenge to biobank specimen quality is pre-analytical factors that are outside of their control--including the methodology and consistency of specimen collection, processing, and shipping.[vi, vii]

To help address this challenge, there is an emerging field known as biospecimen science that is critically evaluating these pre-analytical variables. Simeon and Watson recently published a paper reviewing evidence-based, customer-focused biobanking.[viii] In this paper, the authors point out that a specimen may pass through as many as 20 hands prior to biobank deposit, introducing a high likelihood of pre-analytical errors, including labeling errors.[ix] This begs the question—As a biobank, how sure can you be of what you have?

Beyond the "chain of custody" problem, the Simeon review highlighted an example where a combination of data sets from multiple research sites gave rise to the possibility that the same biobank participant was, in reality, present in the same study, yet disguised as different, distinct individuals from different sites. The authors concluded that this would increase type I and type II errors.

As we look further at the consequences of sample misidentification and poor sample quality, the issue of reproducibility comes

into focus. Prinz et al. reports a high percentage (65%) of inconsistent results among researchers, further signifying the complications resulting from the insufficient methodology of biobanks.[x] The consequences are not trivial, with the National Institutes of Health (NIH) and many scientific journals all increasingly sensitive to the impact of results that cannot be reproduced. This sensitivity is rapidly transforming into stringent policies requiring documentation regarding reproducibility for funding and publication consideration. In short, biobank managers have to accept that they have limited control over the sample collection process even though it directly and dramatically impacts their operation.

In spite of this limited control, a bank's reputation and future growth as a trusted resource depends on the integrity of its deposits.

Let's look further at the issue of reputation.

It is likely apparent that a biobank with known sample quality issues will suffer a loss in credibility. And while fear of reputation loss is a valid motivator, biorepository leaders should view quality as an added benefit. In other words, the quality standards a bank uses should be seen entirely as an investment, one which delivers returns in the form of increasing sample deposits, reputation, and ultimately, scientific impact, resulting scientific directions and conclusions can be considerably negatively affected.

With so many unknowns, one can logically ask--How does a biobank manage this specific risk? What's the cost of that risk management? How do I make my biobank stand out as a leader?

3 The **Movement** Toward Standardization

Facing a history of documented issues and the need to ultimately manage biobank risk, key stakeholders in biorepositories have pushed for standard operating procedures, guidelines, oversight, and accreditation as a base for operations. Today, there are a number of national and international resources available to biobank administrators, including the College of American Pathologists (CAP); the European, Middle Eastern, and African Society of Biopreservation & Biobanking (ESBB); and the International Society For Biological And Environmental Repositories (ISBER). These organizations shed valuable light on all operational-related matters, including the need for a robust QA/QC system. In the NCI Biorepositories and Biospecimen Research Branch (BBRB) recommendations, the economic

aspects of bank operation are also being factored into standard operating procedures. To the bank administrator, the overlay of economic and scientific considerations is an important, everyday balance. If we examine where a return on investment makes sense, an investment ensuring individual specimen identity and sample (e.g., DNA) quality is a natural. In short, better quality samples and reproducibility lead to better data and outcomes, translating into greater chances of grant renewals for researchers and re-engagement of the biobank.

Armed with this knowledge, leading biobanks should not wait for researchers to ask for high-quality samples; but, should make this a part of their standard service offering.

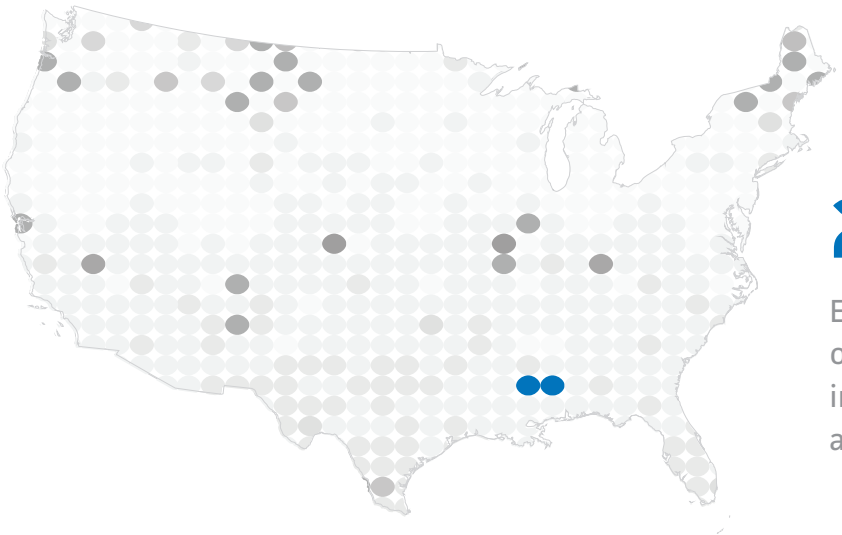
4 Investment in **Risk Management**

How does one justify the cost of sample testing for identity and quality?

Funds are wasted when poor-performing samples are carried through expensive downstream genomic analysis such as genotyping by microarray or by sequencing, so samples must be discarded and data quality suffers. However, if these poor-performing samples were identified prior to analysis, this could result in significant cost savings, as demonstrated by B. Marosy [xi] from the Center for Inherited Disease Research at Johns Hopkins University.

In this study, on average, 4% of samples within a project were found to be problematic, with a range of 0.3%–10.7%. The cost of downstream genomic analysis of this 4% average was more than 2X the cost of testing all samples within the project. For the worst-case scenario of 10.7%, the downstream analysis had more than 6X the cost of testing all samples.

From this data, it is clear that investment in a solution for sample testing is justifiable and will pay for itself in a short period. Besides yielding financial benefits, sample testing helps maintain the integrity of the scientific study.



20,000,000

Estimated number of **deposits per year**, in the United States alone, as of 2008.

5 Standing Out

A significant advantage in being able to assess DNA quality is to offer valuable insight to clients through real-time feedback during the collection phase, both on sample mix-ups, which can be immediately corrected, and on collection site quality.

The conditions that a sample is exposed to, such as temperature and handling before extraction, can greatly affect the quality of the nucleic acid extracted later on. Being able to adjust collection SOPs or identify low-quality collection sites in order to provide the highest-quality samples can be a valuable service offering to biobank customers, giving the biobank a competitive edge for client business.

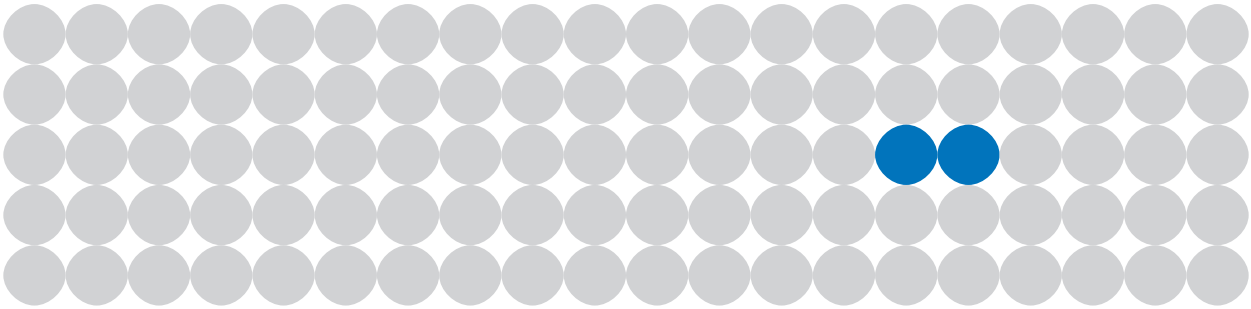
If sample mix-up errors are caught within weeks of collection, there is greater opportunity to reconcile the error and re-collect. On the other hand, if samples are stored and analyzed only after the collection phase is completed, it may be years until mix-ups are discovered. It is highly unlikely that these errors will be resolved before the samples are prepared for lab use.

Advance assurance that many of the available samples can be utilized will provide value to the biobank and the added value of peace of mind to the customer.

6 An Ideal Solution: Cost-Effective **Identity & Sample Quality Testing**

We have already established the costs—both direct and indirect—of poor sample integrity to the biobank. We have also confirmed that the causational factor of poor samples is pre-analysis. The question then becomes, how can biobanks reconcile differences and inconsistencies that happen before a sample even reaches the biobank?

98% of errors happen **before** a sample even reaches a biobank.



The answer is not taking a passive stance and hoping people follow the rules of collection and transport, but active sample verification. These errors can be corrected immediately by the biobank to further prevent the propagation of these errors in downstream studies.[xii]

Furthermore, the ideal solution to this challenge is applicable to a variety of specimen types including blood, saliva, and difficult, degraded specimens, including formalin-fixed, paraffin-embedded (FFPE) samples. It should be quick and useful across global populations, have the ability to identify unique samples in a large cohort across hundreds of thousands of samples, and be easily comparable across different labs to allow for verification at different time-points. Finally, the tool should be flexible and scalable, so as to adjust to the changing number of samples on a day-to-day basis.

7 Conclusions and **Next Steps**

In short, the ideal solution to the challenge outlined above is to verify the genetic identity, as well as the quality, of an individual sample before the sample is ever deposited.

The successful operation of a biobank is an ever-evolving process, with technical, legal, ethical, and financial challenges around every corner.

The aim of this technical note was to focus on quality assurance and to elucidate the following:

1. Sample quality is the cornerstone of a biobank's reputation.
2. Reliance on high-fidelity sample collection is fraught with error and thus a high-risk proposition.
3. Biobank leadership can actively and cost-effectively promote quality and add to perceived sample value with SNPtrace™ ID Solution.

To learn more about the SNPtrace™ ID Solution, watch Dr. Andrew Brooks, Chief Operating Office at RUCDR Infinite Biologics[c] present a general overview, or speak to a Fluidigm sales representative.

8 Sources

i, page 3 D. Vuong, P.T. Simpson, B. Green, M.C. Cummings, and S.R. Lakhani, "Molecular classification of breast cancer," Virchows Arch. 465(1) (2014): 1–14.

ii, page 3 J. Vaught, A. Kelly, and R. Hewitt, "A review of international biobanks and networks: success factors and key benchmarks," Biopreserv. Biobanking 7(3) (2010): 143–150.

iii, page 4 www.isber.org

iv, page 4 H.A. Massett, N.L. Atkinson, D. Weber, R. Myles, C. Ryan et al., "Assessing the need for a standardized cancer Human Biobank (caHUB): findings from a national survey with cancer researchers," J. Natl. Cancer Inst. Monogr. (42) (2011): 8–15.

v, page 5 AS.B. Haga and L.M. Beskow, "Ethical, legal, and social implications of biobanks for genetics research," Adv. Genet. 60 (2008): 505–44.

vi, page 5 D.F. Ransohoff and M.L. Gourlay, "Sources of bias in specimens for research about molecular markers for Cancer," J. Clin. Oncol. 28(4) (2010): 698–704.

vii, page 5 G. Lippi, J.J. Chance, S. Church, P. Dazzi, R. Fontana et al., "Preanalytical quality improvement: from dream to reality," Clin. Chem. Lab. Med. 49(7) (2011): 1,113–26.

viii, page 5 D. Simeon-Dubach and P. Watson, "Biobanking 3.0: evidence based and customer focused biobanking," Clin. Biochem. 47(4-5) (2014): 300–8.

ix, page 5 W. Weyers, "Confusion-specimen mix-up in dermatopathology and measures to prevent and detect it," Dermatol. Pract. Concept. 4(1) (2014): 27–42.

x, page 6 F. Prinz , T. Schlange and K. Asadullah, "Believe it or not: how much can we rely on published data on potential drug targets?" Nat. Rev. Drug. Discov. 10(9) (2011): 712.

xi, page 8 B. Marosy, "Impact of sample pre-testing in a high throughput genotyping facility," poster session presented at American Society of Human Genetics (20-24 October 2009); Honolulu, HI, USA.

xii, page 10 A. Brooks, "Optimizing samples for future use: innovative technology to improve the functional quality control of DNA samples," webinar (2014).

About **Fluidigm**

Fluidigm provides a rapid, simple, and high-throughput SNP genotyping panel that biobanks and biorepositories can use to identify DNA samples and ensure traceability throughout the banking and downstream analytical processes. SNPtrace™Panel is an expertly designed panel using Fluidigm's SNPtype™ Assays. The BioMark HD™ System and the JunoTM System, combined with the SNPtrace™ Panel, deliver a cost-effective, easy to deploy, and easy to use workflow to determine or verify uniqueness, gender, and quality of a DNA sample prior to storage or for use in downstream genomic analysis.

We sell to leading academic institutions, clinical laboratories, and pharmaceutical, biotechnology, and agricultural biotechnology companies worldwide. Our systems are based on proprietary microfluidics and multi-parameter mass cytometry technology, and are designed to significantly simplify experimental workflow, increase throughput, and reduce costs, while providing excellent data quality. Fluidigm products are provided for Research Use Only. Not for use in diagnostic procedures.

Corporate Headquarters
7000 Shoreline Court, Suite 100 | South San Francisco, CA 94080
Toll-free: 1.866.359.4354 | Tel: 650.266.6000 | Fax: 650.871.7152