Biospecimen Complexity—the Next Challenge for Cancer Research Biobanks?

Peter H. Watson1,2

Abstract

Purpose: Biospecimens (e.g., tissues, bloods, fluids) are critical for translational cancer research to generate the necessary knowledge to guide implementation of precision medicine. Rising demand and the need for higher quality biospecimens are already evident.

Experimental Design: The recent increase in requirement for biospecimen complexity in terms of linked biospecimen types, multiple preservation formats, and longitudinal data was explored by assessing trends in cancer research publications from 2000 to 2014.

Results: A PubMed search shows that there has been an increase in both raw numbers and the relative proportion (adjusted for total numbers of articles in each period) of the subgroups of articles typically associated with the use of biospecimens and both dense treatment and/or outcomes data and multiple biospecimen formats.

Introduction

It is clear that there is an increasing appetite in health research for access to human biospecimens. We have documented this increased demand for biospecimens through analysis of published research (1), others have highlighted this same issue through researcher surveys (2), and biobankers have generated advice for researchers to enhance access (3, 4) and debated the challenges for biobanks to respond to this demand with sufficient numbers of high-quality biospecimens (5, 6).

One challenge is the relatively low priority given to the activity of biobanking, despite its importance to health research. Human biospecimens contribute to the data presented in up to 40% of cancer research publications (1, 7). However, a recent news article on underuse of biobanks has gained only 20% of the online attention (based on relative altimetric scores) as a companion news feature about the relative merits of the terms personalized and precision in medicine (8, 9).

Conclusions: Increasing biospecimen complexity is a large-ly unrecognized and new pressure on cancer research biobanks. New approaches to cancer biospecimen resources are needed such as the implementation of more efficient and dynamic consent mechanisms, stronger participant involvement in biobank governance, development of requirements for registration of collections, and models to establish stock targets for biobanks. In particular, the latter two approaches would enable funders to establish a better balance between biospecimen supply and research demand, reduce expenditure on duplicate collections, and encourage increased efficiency of biobanks to respond to the research need for more complex cases. This in turn would also enable biobanks to focus more on quality and standardization that are surely factors in the even more important arena of research reproducibility.

Another challenge is that biospecimens are hard to find. This is because biobanking that supports academic research is conducted across a broad landscape and in the context of individual basic research projects, translational studies, and clinical trials, not just "professional" biobanks. This diffuse nature of the activity leads to many research collections but too often their existence and level of quality is unknown and they are under used. Other potential collections lie in hospital archives, but the process of locating and gaining access can be complicated. There is no requirement for biobanks to be listed in searchable registers and no generally accepted biobank classification schema to help researchers appreciate those that may contain appropriate biospecimens (10). Furthermore, when specimens are found, the conditions whereby they were collected and stored is often unknown (11), and their quality can be hard to determine before embarking on costly experimental analysis. Then just one collection of biospecimens is often insufficient to enable selection on the basis of several clinical and pathologic and biomarker criteria, but combining samples from multiple biobanks can compound experimental variation (12).

Solutions are emerging to address these problems. However, an important new pressure that has been largely missed in this discussion is increasing complexity of biospecimens used in research. Complexity can be considered an aspect of quality that describes levels beyond the simple state. When considered in the context of a tumor biospecimen in a tumor biobank, the majority of cases can be categorized as existing in a "simple state" comprising a biospecimen sample and its annotating data. Both sample and data can be limited (e.g., a single aliquot associated
Discussion

"Simple state" cases currently comprise the majority of the stock held by tumor biobanks but demand for complex cases is clearly rising. In cancer research, this demand is currently driven by new research sectors such as genomic exploration of inter and intra tumor heterogeneity (16) and ctDNA/CTCs (17, 18). But the maturation of many large scale and recently initiated healthy population-based cohorts is another significant impending driver. Over time much of the future research value of such cohorts will be dependent on linkage between biospecimens collected from healthy individuals and biospecimens associated with their future medical diagnoses and events among participants to create complex cases.

Compiling complex cases consumes more biobank resources at a time when pressures to increase quality are already challenging with immediate limited annotating data such as patient diagnosis and specimen composition) or more comprehensive (several sample aliquots and data fields concerning the patient, biospecimen collection and composition, the pathology represented in different areas of the biospecimen, or extensive with the clinical features, treatments, and outcomes data) in terms of scope and density, and many quality metrics can pertain to these. Alternatively, a case in a biobank can be categorized as being in a "complex state."

Complexity is manifested in terms of added features such as sample rarity, format, location, and temporal occurrence that all confer added value but also in many instances added cost. Sample rarity is determined by a combination of actual frequency of the pathology in clinical medicine, modified by availability of biospecimens as a result of standards of practice for treatment. Rarity was once defined largely by uncommon diagnoses defined by standard clinical criteria and as such did not confer added costs to the typical biobank. But rarity is increasingly refined by criteria based on specific molecular features that more precisely define subgroups that are now emerging from subclassification of what were previously considered common tumor types (13). To enable selection of these subgroups biobanks often need to collect additional data or conduct additional processing (e.g., creation of tissue microarrays) to enable annotating assays to be performed, both of which add costs. Sample format refers to whether the sample is available in or matched with samples preserved in different ways (e.g., fresh, frozen, FFPE). Location and temporal occurrence refer to samples linked to additional samples representing different sites from the same individual at the same point in time (e.g., a primary tumor and adjacent nodal metastasis) or representing other time points from the same individual relevant to clinical stages or evolution of a disease (e.g., blood samples taken before and after surgery, or tissue samples obtained from a primary tumor and then again from a recurrence many years later). The impact of research customer demand for these forms of complex cases has not been discussed.

Materials and Methods

An indication of this demand can be obtained by performing a search of PubMed. It is not possible to directly query PubMed for inclusion of biospecimens and related variables because of the lack of standardization in terminology related to biospecimens within articles and specific search terms. However, an indirect indication of published cancer research articles that are likely to have used human biospecimens can be obtained using the search terms: "cancer" and "gene" and "expression." The increasing demand for complex cases can then be illustrated by addition to this baseline query of empirically selected search terms relevant to data or biospecimen use. Higher density annotation and/or data complexity within this set of articles is highlighted by addition of the search terms: "survival," "recurrence," metastasis," either alone or in combination. These terms were also chosen empirically on the assumption that analysis of relationships between gene expression and these different outcome events often require data from the primary and secondary events.

Results

A PubMed query was performed on May 10, 2016, to illustrate trends in use of complex biospecimens at 4 intervals over a 20-year period. These intervals were chosen to match those used in several of our previous publications on biospecimen use, but starting with the most recent year possible. However, it became clear that the indexing of articles in 2015 was incomplete within PubMed so the next most recent complete year (2014) was substituted in the final analysis. As shown in Fig. 1 there is an increase in both raw numbers and the relative proportion (adjusted for total numbers of articles in each period) of the subgroups of articles typically associated with the use of biospecimens and dense treatment and/or outcomes data. The increasing demand for complex cases with combinations of biospecimen formats is also highlighted by addition to this baseline query of the search terms: "FFPE or Formalin-fixed paraffin-embedded," "fresh," "frozen," either alone or in combination. These terms were chosen as they emerged as terms that are commonly used and relatively accurate indicators of biospecimen use in articles in our previous literature review studies (1, 7, 14, 15). As shown in Fig. 1, although the relative proportion of articles that are selected by using a single biospecimen format term has not changed significantly over time, there is an increase in the relative proportions of articles associated with multiple biospecimen formats.

Translational Relevance

Biospecimens are critical for translational cancer research to generate the necessary knowledge to guide implementation of precision medicine. Rising demand and the need for higher quality biospecimens are already evident. However, the recent increase in requirements for biospecimen complexity in terms of linked biospecimen types, multiple preservation formats, and longitudinal data, is a largely unrecognized and new pressure on cancer research biobanks. Complex biospecimens are especially important to interrogate biomarkers of tumor evolution over time. Given the already widening gap between the needs of research and the quality that biobanks can deliver, new approaches to cancer biospecimen resources are needed to enable biobanks to focus their efforts on the collection of complex cases. We need more efficient enrollment mechanisms, stronger participant involvement in governance, and coordination of biobanking enabled by registration of collections.

www.aacrjournals.org Clin Cancer Res; 23(4) February 15, 2017

895

Published OnlineFirst August 22, 2016; DOI: 10.1158/1078-0432.CCR-16-1406

Downloaded from clincancerres.aacrjournals.org on January 10, 2022. © 2017 American Association for Cancer Research.
the sustainability of many individual biobanks (6, 19). There are some general solutions to all these pressures that have already begun to emerge in the face of increased demand for quality and numbers of biospecimens. At the system level these include catalogues of biobanks (20) and quality management (21–23), and best practice (11) and reporting standards (24). Here, the focus is principally on making larger biospecimen numbers available that are collected under known and standardized conditions.

Figure 1.
Increasing demand for cases with higher density annotation and/or complexity is highlighted by the increase in numbers (left) and relative proportions (right) of articles typically associated with the use of biospecimens and dense treatment and/or outcomes data. The PubMed biomedical library was queried for cancer research articles using human biospecimens. The baseline search terms used were: “cancer” and “gene” and “expression,” with filters added for: Human, English, Full Text. The results were then refined by adding search terms: “survival,” “recurrence,” “metastasis” (top), or “formalin-fixed paraffin-embedded or FFPE” (indicated on graphs as FFPE), “fresh” or “frozen” (middle), or combinations of these terms.
operating protocols. The emphasis is also on denser annotation including data on preanalytic variables that may influence interpretation of research results such as collection time (24). But quality comes at a cost, and not all assays and types of research require the same quality measures. Biobankers also worry about the impact of these costs on sustainability of biobanks and have begun to frame this issue in terms of business planning (6, 25, 26), mechanisms to promote appropriate user fees (27, 28), and some other solutions to make biobank operations more efficient (6).

Many biobanks would certainly benefit from even simple strategic or business plans, many have room to implement operational efficiencies and more realistic user fees, and many continue to collect and then hold on too long to cases in excess of known or projected demand (26, 29).

One solution that could more directly address the pressures caused by increasing demand for complex biospecimens is the implementation of new approaches for patient enrollment and consent. Complexity is defined here as an aspect of biospecimen quality, in addition to the inherent integrity of the biological material and the nature of the annotating data. But other important aspects of biospecimen quality are the determinants of research use (i.e., the decision, scope and nature of the patient consent or waiver of consent provided by an ethics board). More efficient approaches to facilitate enrollment and consent could mitigate the influence of factors such as rarity and temporal occurrence that contribute to complexity and so reduce the costs of compiling complex cases. For example, implementing an institution wide program to just obtain permission to contact and access to limited health data from patients can reduce costs of consenting for biobanks and enable the biobank to be more selective in deciding which cases to collect and obtain specific patient consent (30, 31). This would allow biobanks to specifically concentrate on collecting only complex cases such as those where a biospecimen is available from a primary and a recurrence and to approach these patients for consent. Others have proposed capitalizing on the potential of electronic communication interfaces to establish more efficient and ongoing connections between biobanks and participants (32). This ‘dynamic consent’ would also allow specific participants to be approached for different kinds of consent while also allowing these individuals to manage their own consent preferences over time. Another approach is to strengthen the direct involvement of participants in the governance of biobanks (33). This could raise the confidence of an ethics board to approve post-procedure consent protocols or to grant a waiver of consent for collection, both of which would improve the ability of a biobank to selectively focus on collecting complex cases.

Less visible and less debated is the fact that most biobanking is already supported indirectly by research funders within the context of research grants. But funders do not track the creation of biobank resources or demand basic quality requirements, and at least some of the activity supported is a wasted duplication of effort. So another important solution would be the implementation of requirements for registration of collections (21) and development of models to establish stock targets for biobanks (29). Several online catalogues of biobanks exist but these are populated mostly by large professional “poly-user” biobanks (10). Many other forms of biobank (e.g. biobanks arising after collection for research projects and studies), some of high importance and quality, are very difficult to locate. Registers of all collections would establish a better balance between biospecimen supply and research demand, encourage increased efficiency and responsiveness of biobanks to research need for more complex cases within a known funding envelope. It would also enable biobanks to focus on quality and standardization that are surely factors in the even more important arena of research reproducibility (34).

**Conclusion**

Rising demand and the need for higher quality biospecimens in cancer research are already evident. However, the recent increase in biospecimen complexity is a largely unrecognized and additional pressure. New approaches to funding and management of cancer biospecimen resources are needed to address all these pressures and improve the quality, scale, and efficiency of biobanking.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Grant Support**

This work was supported by grants from the Terry Fox Research Institute and the Canadian Institutes of Health Research.

Received June 19, 2016; revised August 8, 2016; accepted August 9, 2016; published OnlineFirst August 22, 2016.

---

**References**

Clinical Cancer Research

Biospecimen Complexity—the Next Challenge for Cancer Research

Biobanks?

Peter H. Watson


Updated version

Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-16-1406

Cited articles

This article cites 34 articles, 4 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/23/4/894.full#ref-list-1

Citing articles

This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/23/4/894.full#related-urls

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, use this link
http://clincancerres.aacrjournals.org/content/23/4/894.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.