A Data Standard for Sourcing Fit-for-Purpose Biological Samples in an Integrated Virtual Network of Biobanks

Philip R. Quinlan,¹ Gita Mistry,² Helen Bullbeck,³ and Anne Carter,⁴
on behalf of the Confederation of Cancer Biobanks (CCB) Working Group 3

Human tissue biobanks are at the epicenter of clinical research, responsible for providing both clinical samples and annotated data. There is a need for large numbers of samples to provide statistical power to research studies, especially since treatment and diagnosis are becoming ever more personalized. A single biobank cannot provide sufficient numbers of samples to capture the full spectrum of any disease. Currently there is no infrastructure in the United Kingdom (UK) to integrate biobanks. Therefore the National Cancer Research Institute (NCRI) Confederation of Cancer Biobanks (CCB) Working Group 3 looked to establish a data standard to enable biobanks to communicate about the samples they hold and so facilitate the formation of an integrated national network of biobanks. The Working Group examined the existing data standards available to biobanks, such as the MIABIS standard, and compared these to the aims of the working group. The CCB-developed data standard has brought many improvements: (1) Where existing data standards have been developed, these have been incorporated, ensuring compatibility with other initiatives; (2) the standard was written with the expectation that it will be extended for specific disease areas, such as the Breast Cancer Campaign Tissue Bank (BCCTB) and the Strategic Tissue Repository Alliances Through Unified Methods (STRATUM) project; and (3) biobanks will be able to communicate about specific samples, as well as aggregated statistics.

The development of this data standard will allow all biobanks to integrate and share information about the samples they hold, facilitating the possibility of a national portal for researchers to find suitable samples for research. In addition, the data standard will allow other clinical services, such as disease registries, to communicate with biobanks in a standardized format allowing for greater cross-discipline data sharing.

Introduction

THE USE OF BIOLOGICAL SAMPLES to facilitate research is not new; equally, the banking of samples for research purposes is not a new development. There remain, however, key challenges in connecting researchers to the samples required for their research. Access to samples of suitable quality is still posed as one of the major challenges facing research.^{2–4} The difficulty facing a researcher trying to find relevant sources of samples is partly driven by the fact that it can be hard for any single biobank to provide all of the samples needed to give statistical rigor to a study. This is especially true for samples from rare diseases, where a single biobank may only collect a handful of cases over a year. The study by Curtis et al.⁵ found that even for a common disease (e.g., breast cancer), there was a need to contact five different biobanks in order to source sufficient numbers of samples.

Combining samples obtained from multiple centers can, however, introduce an extra level of complexity as each center may collect and store samples to different standards. Although guidelines do exist, such as the International Society for Biological and Environmental Repositories (ISBER) Best Practices for Repositories⁶ and those published by the National Cancer Institute (<http://biospecimens.cancer.gov/practices/2011bp.asp>), or data standards that recommend a certain level of information to be made available about samples collected for use in research,^{7–9} these standards are not universally enforced, or even followed, by all biobanks or national/international directories.

The National Cancer Research Institute (NCRI) Confederation of Cancer Biobanks (CCB) (<http://www.ncri.org.uk/ccb/>) is a consortium of organizations within the UK involved in the development, management, and use of biobank resources for cancer research. The CCB recently undertook a

¹University of Dundee, Division of Cancer Research, Ninewells Hospital and Medical School, Dundee, United Kingdom.

²Children's Cancer and Leukaemia Group (CCLG), University of Leicester, Leicester, United Kingdom.

³Brainstrust, Cowes, Isle of Wight, United Kingdom.

⁴National Cancer Research Institute, Angel Building, London, United Kingdom.

project to formulate a set of criteria against which a biobank can be assessed and accredited; these criteria cover the gamut of biobanking activities, which include consenting donors, risk management, and sample quality. From the initiation of the project, the aim was to devise data standards that would be applicable to all research biobanks, independent of any disease focus. The core long-term goal of the CCB is to create a catalogue and portal where all accredited biobanks are listed, and researchers can search for tissue samples across the accredited network safe in the knowledge that all samples meet a certain standard. Therefore, a data standard was created specifically detailing what information biobanks must be able to provide about the samples they hold; other criteria will be reported elsewhere.

In an attempt to improve the “discoverability” of sample collections, central directories of samples are proliferating. Ensuring that each biobank collects appropriate data has led to the development of data standards such as the MIABIS standard,¹ which determines the minimum amount of data that should be made available to a central directory in order to allow researchers to find the samples they require. The data standard presented here was drafted with some central core principles. The first was to ensure that this standard was compatible with other standards; where it overlapped with other standards, such as MIABIS,¹ the data terms present in those standards would be used rather than reinvented. The standard would not dictate the data terms to be used; instead it would ensure that the biobank supplies the current data definitions in use. While the data standard was devised by the CCB, it is intended to cover all biobanks and therefore has no disease focus; thus, it must facilitate extensions to be developed that ensure key disease-specific information can be added. Finally, the data standard must allow the description of every patient and individual sample aliquot to ensure that core information relating to the provenance of the sample can be found by the researcher, while maintaining patient confidentiality.

Materials and Methods

A working group (WG 3) consisting of CCB members from several biobanks was assembled and complemented with lay representation, researchers, and noncancer biobankers for the purpose of determining the requirements for a data standard. The existing data standards^{1,7,9} were first examined to ensure that should a standard that meets the requirements of the CCB already exist, it would be recommended rather than implementing a new standard. As no single data standard was found that met all the needs of the CCB, a new data standard was developed that would, where appropriate, adopt the existing standards wherever overlap was present. The focus of WG 3 was to develop a data standard that can be used in the development of a national catalogue of samples, irrespective of disease focus. In order to achieve this aim, the four key goals were to: 1) not repeat or conflict with similar work; 2) not focus on the definition of data terms; 3) offer a mechanism for extension; and 4) include patient and sample level data.

The draft data standard was also reviewed by other biobanks with active online catalogues, as well as by national projects examining the development of such catalogues to ensure the proposed standard had compatibility with these initiatives, and work was undertaken to provide a proof of concept catalogue system.

Results

As the membership of CCB is broad and open to any institution or researcher that collects cancer samples, the confederation includes a wide array of ‘biobank’ types. As such, it may not be possible for all custodians of samples to provide the same level of information as the more established biobanks; even within one biobank there may be some collections that have available varying levels of data. The data elements that must be supplied by all biobanks are provided in Table 1. The data required to meet the minimum standard is described in Table 2, and the data required to meet the best practice standard are described in Table 3. The main difference between the two standards is that the minimum standard includes brief information about the biobank and aggregated information about the sample, and the best practice standard includes information at the individual patient and individual sample levels.

Meeting the key goals

Not repeat or conflict with similar work. The minimum data standard is conceptually very similar to that documented in the MIABIS standard,¹ in particular in describing the biobank (Table 1a) and a collection of samples (Table 2a). Similarly, the best practice standard (Table 3) is more akin to the data standard used by the Breast Cancer Campaign Tissue Bank (BCCTB) (<https://breastcancertissuebank.org/about-tissue-bank.php>). Therefore, to achieve the first goal, the CCB data standard has, where appropriate, adopted the relevant terms from these other standards as indicated in the ‘Source’ column of Tables 1–3. The main difference from the MIABIS standard¹ is the inclusion of Table 2b (Sample Data). This addition provides information on different types of aggregated samples. For example, it is conceivable that one collection under the custody of a biobank may contain both cancerous and noncancerous materials or tissues, and fluid samples. In this scenario, there would be one entry in the Collection table (Table 2a) and two in the Sample Data table (Table 2b), one for the cancerous and one for the noncancerous samples.

Not focus on the definition of data terms. Focusing on the structure and format of the data terms within the standard rather than defining the specific terms to be used ensured this goal was achieved. The key focus of the standard was to seek agreement on the data that should be collected and how this should be structured. Therefore, the terms to be used by a biobank for the diagnoses as well as the terms to represent the organs from which the samples originate must be supplied by the biobank itself, represented in Tables 1b and 1c, respectively. This means that the biobanks included in the catalogue can continue to use their existing databases and ontologies. Some data may not be available from all biobanks; for example, it may not be possible for every biobank in the network to supply the time that the blood flow to the sample was cut-off prior to collection. It is inconceivable, within a network of biobanks, that all biobanks will always be able to provide the same level of information. Therefore, for every term within the data set, it is possible for the bank to mark that data as Unknown/Unavailable/Inappropriate.

Offer a mechanism for extension. A core aim of the CCB data standard was to ensure that it provides all core

TABLE 1. DATA ELEMENTS REQUIRED OF ALL BIOBANKS

<i>Field Name</i>	<i>Format</i>	<i>Description</i>	<i>Source</i>
<i>(a) Biobank</i>			
Biobank ID	Free text	Free text string of letters starting with the country code (according to standard ISO1366 alpha2) followed by the underscore “_” and post-fixed by a biobank ID or name	MIABIS-01
Name of biobank	Free text	Text string of letters denoting the name of the biobank.	MIABIS-02
Governing body	Free text	Text string of letters denoting the governing body (e.g., a university, hospital trust, commercial company etc.) for the biobank	MIABIS-03
URL	Free text	Text string of letters with the complete http-address for the biobank URL	MIABIS-04
Country code	ISO-standard (3166 alpha2), two letter code	Text string of letters of the two letter code for the country of the biobank according to ISO-standard 3166 alpha2	MIABIS-05
Contact person	Free text	Text string of letters denoting the name of the contact person for the biobank	MIABIS-07
Contact phone	Free text	Phone number for the “Contact person”, including international call prefix	MIABIS-08
Contact email	Free text	Email address of the “Contact person”	MIABIS-09
Contact department	Free text	Department, or corresponding (e.g., division), of affiliation of the “Contact person”	MIABIS-10
Contact address	Free text	Street name and street number or PO Box of the “Contact person”	MIABIS-11
Contact post code	Free text	Post Code of the “Contact person”	MIABIS-12
Contact city	Free text	City of the “Contact person”	MIABIS-13
Contact country	ISO-standard (3166 alpha2), two letter code	Country of the “Contact person”	MIABIS-14
Last updated	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when data about the biobank was last updated in a database	MIABIS-17
<i>(b) Diagnoses</i>			
Diagnosis code	Free text	Text representation of the diagnosis	
Diagnosis description	Free text	Text description of the diagnosis	
Source and version of diagnosis code	Free text	Text representation of the source and version of the diagnosis code (e.g., ICD-10)	
<i>(c) Organs</i>			
Organ code	Free text	Text representation of the organ	
Organ description	Free text	Text description of the organ	
Source and version of organ code	Free text	Text representation of the source and version of the organ code (e.g., SNOMED)	
<i>(d) Available data</i>			
Data type	Free text	Text string of letters detailing the type of data available (e.g., Comorbidity, Clinical, Pathology, Omics)	
Data availability	Free text	Text String of one of the following: Available to all/Available only through collaboration/Other	

terminology used to allow any biobank within the network to communicate about the patients and samples under its custodianship, while providing sufficient areas where the standard could be extended. To facilitate this, the core data standard should not be changed by any disease specific area. However, there are fields named ‘Disease specific data’ (e.g., Table 3a) where the standard can be extended to allow either additional tables or fields to be placed at those levels.

The BCCTB was a leading partner in the development of the data standard and the core structure has been utilized within this breast cancer specific biobank. As well as demonstrating the ability of the standard to be extended, it also demonstrates the appropriate use of the patient and sample information. The CCB data standard was created with the belief that it could be used beyond cancer to describe samples. To test the feasibility of this belief, the CCB

TABLE 2. DATA REQUIRED TO MEET THE MINIMUM STANDARD

<i>Field Name</i>	<i>Format</i>	<i>Description</i>	<i>Source</i>
<i>(a) Collection</i>			
Sample collection/ Study ID	Free text	Text string depicting the unique ID or acronym for the sample collection or study	MIABIS-18
Study name	Free text	Text string of letters denoting the name of the study	MIABIS-19
Description of collection	Free text	Text string of letters describing the sample collection or study aim (max 200 characters)	MIABIS-20
Type of collection	Free text	Case-control, Cohort, Cross-sectional, Longitudinal, Twin-study, Quality-control, Population-based, Disease- specific, Other	MIABIS-30
Collection start	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format specifying when the sample collection starts	MIABIS-31
Collection end	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format specifying when the sample collection ends, if applicable	MIABIS-32
Sex of donors	Free text	Text string of letters denoting the sex of the sample donors. Can be several values	MIABIS-35
Number of current sampled individuals	Integer	Number of individuals with biological samples in the sample collection/study at the date of Last updated	MIABIS-48
Last updated	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when data about the sample collection was last updated in a database	MIABIS-51
Access conditions	Free text	Text string describing the conditions under which access to the collection is granted	
Collection status	Free text	Text string describing the collection status of the collection	
Biobank ID	Free text	Text string of letters starting with the country code (according to standard ISO1366 alpha2) followed by the underscore “_” and post-fixed by a biobank ID or name	MIABIS-01
<i>(b) Sample data</i>			
Material type	Free text	Most commonly abundant biological samples in biobanks; can be several values	MIABIS-41
Macroscopic assessment	Free text	Where applicable, a text string of letters detailing the macroscopic assessment of the sample (e.g., Tumor, Normal)	
Pathological assessment	Free text	Where applicable, a text string of letters detailing the pathological assessment of the sample (e.g., invasive, ductal)	
Current sampled individuals	Integer	Number of individuals with biological samples in the sample collection/study at the date of Last updated (also see Planned sampled individuals)	MIABIS-48
Organ code	Free text	The code of the Organ; it must match an entry in the Organ table	
Diagnosis code	Free text	The code of the Diagnosis (if applicable); it must match an entry in the Diagnosis table	
Sample collection/ Study ID	Free text	The identifier used in the Collection table	
<i>(c) Collection to available data</i>			
Sample collection/ Study ID	Free text	The identifier used in the Collection table	
Data type	Free text	The data type; must match an entry in the available data table	

TABLE 3. DATA REQUIRED TO MEET THE BEST PRACTICE STANDARD

<i>Field Name</i>	<i>Format</i>	<i>Description</i>	<i>Source</i>
<i>(a) Patient</i>			
Patient's anonymized ID	Free text	An identifier, made of either letters, numbers or both, as an anonymized identifier for the patient	MIABIS-35 NA
Patient's sex	Free text	Text string of letters denoting the sex of the sample donors	
Disease specific data	To be defined	There is an expectation that some diseases will extend the patient table	
<i>(b) Patient diagnosis</i>			
Patient diagnosis ID	Free text	An identifier, made of either letters, numbers or both, as an anonymized identifier for the diagnosis	
Diagnosis date	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when the diagnosis was made	
Patient's anonymized ID	Free text	The identifier used in the Patient table	
Patient's diagnosis code	Free text	The code of the patient's diagnosis; it must match an entry in the Diagnosis table	
<i>(c) Sample group</i>			
Sample group Identifier	Free text	Text string of letters and/or numbers denoting the identifier of the sample group	STRATUM
Date sample group collected	ISO-standard (8601) time format (YYYY-MM-DD)	Date in ISO-standard (8601) time format when the samples were banked	
Consent details	Free text	Text string of letters denoting the consent details of the collection	
Patient diagnosis ID	Free text	The ID of the diagnosis that is most relevant to the samples; it must match one in the Patient Diagnosis table	
Age	Free text	The age of the patient at the time of the sample, stratified in the following age groups: 0–5, 6–12, 13–17, 18–20, 21–30, 31–40, 41–50, 51–60, 61–70, 71–80, 81–90, 91–100, 101–110, 111–120	
Disease specific data	To be defined	There is an expectation that some diseases will extend the sample group table	NA
Patient anonymized ID	Free text	The anonymized ID of the patient that the Sample Group comes from	
<i>(d) Solid specimen</i>			
Solid sample identifier	Free text	Text string of letters and/or numbers denoting the identifier of the solid sample	EUROC-30
Organ code	Free text	Text representation of the Organ Code as shown in the Organ table	
Organ location	Free text	Where applicable, the text string of letters denoting the location of the organ (e.g., Left/Right)	
Disease specific data	NA	There is an expectation that some diseases will extend the Solid Specimen table; see other data standards	Multiple
Sample group identifier	Free text	The identifier of the Sample Group to which the Solid Specimen belongs; this must match an ID in the Sample Group table	
<i>(e) Tissue sample</i>			
Barcode	Free text	An identifier, made of either letters, numbers or both, as an anonymized identifier for the sample	MIABIS41
Material type	Free text	Text string of letters detailing the type of the sample (e.g., solid tissue, biopsy)	
Macroscopic assessment	Free text	Text string of letters detailing the macroscopic assessment of the sample where applicable (e.g., normal, tumor)	

(continued)

TABLE 3. (CONTINUED)

<i>Field Name</i>	<i>Format</i>	<i>Description</i>	<i>Source</i>
<i>(e) Tissue sample</i>			
Pathological assessment	Free text	Where applicable, a text string of letters detailing the pathological assessment of the sample (e.g., invasive, ductal)	MIABIS-47
Availability	Free text	Text string to represent the availability of the sample	
Time to freeze from excision	Time (in minutes)	The number of minutes from excision to freezing	
Time to freeze from cease blood flow	Time (in minutes)	The number of minutes from the cut-off of blood flow to the sample to freezing	
Storage temperature	Temperature (in Celsius)	The storage temperature of the samples in Celsius.	
Freeze method	Free text	Text description describing how the samples were frozen (e.g., snap frozen, controlled-rate freezing)	Multiple
Storage medium	Free text	Text description of any storage medium added to the sample (e.g., RNALater)	
Disease specific data	NA	There is an expectation that some diseases will extend the Tissue Sample table; see other data standards	
Solid specimen identifier	Free text	The identifier of the Solid Specimen to which the Tissue Sample belongs; this must match an ID in the Solid Specimen table	
<i>(f) Fluid sample</i>			
Barcode	Free text	An identifier, made of either letters, numbers or both, as an anonymized identifier for the sample	MIABIS41
Material type	Free text	Text string of letters detailing the type of the sample (e.g., Whole Blood, Serum, Plasma, Urine)	
Availability	Free text	Text string to represent the availability of the sample	MIABIS-47
Time to freeze	Time (in minutes)	The number of minutes from the blood being taken to frozen	
Storage temperature	Temperature (in Celsius)	The storage temperature of the samples in Celsius	
Freeze method	Free text	Text description describing how the samples were frozen (e.g., snap frozen, controlled-rate freezing)	Multiple
Volume	Volume in ml	The volume of the stored sample or aliquot	
Collection method	Free text	Text string of letters detailing how the Fluid Sample was taken (e.g., EDTA, heparin, no preservative, mid-stream sample)	
Processing method	Free text	Text description describing any processing performed (e.g., single or double centrifugation, gravitational force applied)	
Storage medium	Free text	Text description of any storage medium added to the sample	
Disease specific data	NA	There is an expectation that some diseases will extend the Fluid table; see other data standards	Multiple
Sample group identifier	Free text	The identifier of the Sample Group to which the Fluid Sample belongs; this must match an ID in the Sample Group table	
<i>(g) Sample to available data</i>			
Sample barcode	Free text	The identifier used in the Tissue Sample or Fluid Sample table	
Data type	Free text	The data type; must match an entry in the available data table	

data standard was used within the STRATUM project (<http://www.stratumbiobanking.org/data.html>), which created a data standard for cataloguing respiratory disease samples. The core dataset of the CCB project was retained and extended where appropriate. While the STRATUM project is yet to be implemented, it demonstrates that conceptually the CCB data standard could be applied to a wider biobanking setting.

The core focus of this work was to facilitate the creation of a national catalogue of available samples. Although this has not yet been completed, the Edinburgh and Dundee Experimental Cancer Medicine Centres (ECMC) (<http://www.ecmcnetwork.org.uk/network-centres/edinburgh/>) have adopted the data standard when developing a system to allow researchers to discover what samples are available across these two independent biobanks, one based in Dundee and the other

in Edinburgh. Therefore, even though both biobanks use different terminology for their samples and two independent database systems, the data standard was able to provide a mechanism for the databases of both biobanks to communicate about the samples available at each site.

Include patient and sample level data. The ECMC trial of the data standard showed that each ECMC site could upload an anonymized version of their data pertaining to the patient (Table 3a). Some data relating to the patient, such as the patient diagnosis (Table 3b), may change over time so this is separated from the main Patient table to allow multiple entries to be attributed to the Patient. In a similar fashion, there are some properties linked to the sample that are time sensitive, such as the age of the patient at the time the sample was taken, any diagnosis at the time the sample was taken, or the consent conditions for the sample. The Sample Group (Table 3c) is used to provide information relevant to samples that were all collected on the same day, which allows the time sensitive information (diagnosis, age, consent conditions) to be attributed to all the grouped samples rather than having to link them individually. The Tissue Sample (Table 3e) and Fluid Sample (Table 3f) tables represent the individual aliquots and their properties, including some key quality information. The two were separated as there are clearly additional fields for tissue samples, such as the type of organ and the location of that organ (contained within the Solid Specimen Table 3d) that are not relevant to a fluid sample such as blood. Conversely, there are properties that a blood sample will require, such as volume, that are not applicable to tissue samples. Again, where the field is found in another data standard, such as the Storage Temperature, the definition from that standard has been used. The reason for requesting this level of detail is to ensure that researchers can find combinations of samples that may not have been predefined within a collection of samples using the Collection Table. In addition, the researcher can search for samples based on key quality control parameters.

Discussion

The CCB data standard provides a mechanism for building on the work of the previous MIABIS standard,¹ enabling researchers to find suitable available samples based on the individual characteristics of patients. The data standard is designed to empower the researcher with all the information available to help them decide which samples are appropriate for their research. An alternative option would have been to ask every biobank in the UK to provide only the data that was known to be available for all biobanks. This approach of going to the lowest common denominator actively undermines the potential benefit of any national registry, as key quality metrics are not made available to the researcher, especially in a climate where many standards and journal guidance are asking for such information.⁷⁻⁹ In the same light, the data standard should not exclude collections by placing too high a burden on entry, yet should include collections that do not have the same level of detail. Although these collections may not be appropriate for use in all scenarios, they should still be visible to the researcher as they may still be of some use. Therefore, the CCB data standard provides a mechanism to ensure that all sample

collections can be accounted for while introducing a level of quality metrics at the individual sample level.

The CCB data standard avoids detailing the exact terms that each site must use to comply with the standard. Instead, the data standard focuses on agreement on the data to be collected and how it should be structured. This approach does introduce a concern for implementation of the data standard as differences in the meaning of terms will have to be mapped within the central registry. However, the adoption of the data standard by both the ECMC and BCCTB demonstrates that this challenge is technically possible to overcome.

The CCB provides a network of biobanks that are seeking to implement the data standard as part of a harmonization project in which any biobank seeking to be accredited must meet certain pre-defined standards. As such, and in combination with this data standard, the CCB will be able to provide a national registry of samples in the UK, and so provide a one-stop portal for researchers to source the most suitable samples available for their research within a framework of accredited standards.

Acknowledgments

We wish to thank members of Working Group 3 from the NCRI CCB Harmonization Project (Joint Chairs: Ian Forgie and Philip Quinlan; WG Members: Anne Carter, Bill Greehalf, Elwyn Shing, Gita Mistry, Helen Bulbeck, James Flanagan, John Brinsley, Kwok Pang, Mairead MacKenzie, Martin Groves, and Stuart Griffiths); the STRATUM Working Group, Carol Dawson and Paul Mitchell from the ECMC Edinburgh Centre and Breast Cancer Campaign.

Author Disclosure Statement

No competing financial interests exist.

References

1. Norlin L, Eriksson M, Merino-Martinez R, Anderberg M, Kurtovic S, and Litton J-E. Biopreservation and biobanking. A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobanking* 2012; 10:343–348.
2. Thompson A, Brennan K, Cox A, et al. Evaluation of the current knowledge limitations in breast cancer research: A gap analysis. *Breast Cancer Res* 2008;0:R26.
3. Mabile L, Dagleish R, Thorisson GA, et al. Quantifying the use of bioresources for promoting their sharing in scientific research. *GigaScience* 2013;2:7.
4. Eccles SA, Aboagye EO, Ali S, et al., Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. *Breast Cancer Res* 2013'15:R92.
5. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;486:346–352.
6. International Society for Biological and Environmental Repositories, Repositories, I.S.f.B.a.E., 2012 Best Practices for Repositories, Collection, Storage, Retrieval, and Distribution of Biological Materials for Research. *Biopreserv Biobanking* 2012;10:79–161.
7. Moore HM, Kelly A, McShane LM, Vaught J. Biospecimen Reporting for Improved Study Quality (BRISQ). *Transfusion* 2013;53.

8. McShane LM, Altman DG, Sauerbrei W et al., and Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics, REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93: 387–391.
9. Lehmann S, Moore H, Ashton G, et al. [International Society for Biological and Environmental Repositories (ISBER) Working Group on Biospecimen Science], Standard preanalytical coding for biospecimens: Review and implementation of the sample PREanalytical Code (SPREC). *Biopreserv Biobanking* 2012;10:366–374.

Address correspondence to:

*Dr. Philip Quinlan
Division of Cancer Research
Ninewells Hospital and Medical School
University of Dundee
Dundee DD1 9SY
United Kingdom*

E-mail: prquinlan@me.com