A critical analysis of cancer biobank practices in relation to biospecimen quality

Amanda Rush¹ · Kevin Spring^{2,3,4} · Jennifer A. Byrne^{1,5}

Received: 26 August 2015 / Accepted: 17 September 2015 / Published online: 22 October 2015

© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2015

Abstract There are concerns that a substantial proportion of published research data is not reproducible, which may partially explain the frequent failure to translate pre-clinical results to clinical care. High-quality cancer biospecimens are needed for robust, reproducible research findings, with most researchers obtaining these specimens from cancer biobanks or tumour banks. This review provides an overview of the types of quality control (QC) activities conducted within cancer biobanks that pertain to biospecimen quality and of biospecimen quality reporting tools, including SPREC and BRISQ. We examine how QC assay results and other biospecimen data are communicated from biobanks to researchers, and whether these activities lead to improved biospecimen quality reporting within the literature and/or to improved research outcomes. We also discuss operational factors that limit QC activities within biobanks and evidence gaps requiring further research. In summary, whereas the provision of quality biospecimens is a common aim of cancer biobanks,

QC activities remain underreported and are rarely discussed in the literature, compared with other aspects of biobank operations. Further research is required to determine how biobanks can most efficiently optimise biospecimen quality, and how communication between biobanks and researchers can be improved.

Keywords Biobank · Cancer · Biospecimen quality · Research reproducibility · Quality control · Reporting

Research reproducibility—a problem for all who generate and use research data

Many researchers and commentators state that irreproducible pre-clinical data are largely responsible for the failure to translate research results to the clinic, and the need to improve the current situation has been widely recognised (Buck 2015; Freedman et al. 2015; Ioannidis 2014). Inadequate reproducibility may be at least partially responsible for the mismatch identified between research inputs and outputs, such as improvements in human life expectancy between 1965 and 2014 (Bowen and Casadevall 2015). While the generation of irreproducible data may achieve short-term gains for researchers in the form of publications and further funding, this can lead to years of wasted effort from potentially numerous research groups who attempt to replicate and/or extend these results, as well as wasted research funds (Ioannidis 2014). Furthermore, where irreproducible results contribute to clinical trials or other patient interventions, these can result in patients being administered treatments that do not work (at best) or produce active harm (at worst). In both these cases, potentially effective treatments can also be wrongly withheld and healthcare resources wasted. Even in the highly regulated environment of cancer diagnostics, false-positive and false-

✉ Jennifer A. Byrne
jennifer.byrne@health.nsw.gov.au

¹ Molecular Oncology Laboratory, Children's Cancer Research Unit, Kids Research Institute, The Children's Hospital at Westmead, Locked Bag 4001, Westmead 2145, NSW, Australia

² Ingham Institute, PO Box 3151 (Westfield Liverpool), Liverpool 2170, NSW, Australia

³ Liverpool Clinical School, University of Western Sydney, Locked bag 7103, Liverpool 1871, Australia

⁴ South Western Sydney Clinical School, University of New South Wales, Liverpool, NSW 1871, Australia

⁵ The University of Sydney Discipline of Paediatrics and Child Health, The Children's Hospital at Westmead, Locked Bag 4001, Westmead 2145, NSW, Australia

negative results persist, in part due to variable biospecimen quality (Hicks et al. 2011; Wolff et al. 2013).

The changing landscape of biospecimen use in cancer research

Biomedical research typically consists of pre-clinical or discovery research, which requires validation in independent sample cohorts, followed by testing through clinical trials before these discoveries can be translated into clinical practice (Fig. 1). Pre-clinical cancer research frequently involves experiments using model systems such as cancer cell lines and animal models of disease. Cancer cell lines are accessible, inexpensive and, in some cases, represent adequate disease models (Wilding and Bodmer 2014). Concerns over the impacts of cell line cross-contamination and mis-identification are now being addressed through better cell culture practices (Geraghty et al. 2014) and journal-imposed requirements for researchers to authenticate cell lines using short tandem repeat (STR) profiling (Lichter et al. 2010). However, other limitations, such as those imposed by routine cell culture systems, the consequences of long-term passaging, including loss of tumour heterogeneity, and genetic drift from genomic instability, result in cell lines incompletely representing the original tumour state within the host (LaBaer 2012; Wilding and Bodmer 2014). The growth of cancer cell lines as xenograft models provides a three-dimensional tissue environment (Wilding and Bodmer 2014), but not always one that is orthotopic (Salletta et al. 2014). Xenograft models also present additional barriers, such as high cost and low-throughput capacity (Wilding and Bodmer 2014), in addition to the physiological differences between small animal models and humans (see review of Lal et al., this edition of *Biophysical Reviews*). Because of these limitations, researchers are increasingly seeking to perform pre-clinical studies using human cancer samples and associated tissue controls. As cell lines which reflect all human cancers are not yet available (Boehm and Golub 2015), studying human cancers directly can provide a better representation of disease (Lawrence et al. 2014). Well-annotated biospecimens can enable associations to be tested between measures of interest and provide data on the host, the disease present, the treatment(s) given and patient outcome.

Biospecimens have also gained prominence in cancer research through the rise of transcriptomics, genomics and proteomics technologies, which permit the analysis of hundreds or thousands of analytes per individual sample (Micheel et al. 2012). These high-throughput techniques place increased requirements upon sample numbers, as they require large cohorts to reduce the ever-present problem of data overfitting (Micheel et al. 2012). An analysis of the numbers of biospecimens used in three different cancer research programmes revealed that the numbers of biospecimens

employed increased by sixfold between 2001–2003 and 2010–2012, and average cohort sizes increased from approximately 50 to 200 cases (Braun et al. 2014). The requirement for increased sample numbers is also further driven by the need for independent test and validation cohorts in order to provide pre-clinical data suitable for further translation (Fig. 1).

Increased information at the genomic and transcriptomic level has also allowed researchers to understand the degree of inter-tumour heterogeneity that exists in diseases that were previously subject to much simpler classification systems (Vogelstein et al. 2013; Watson et al. 2013). The availability of transcriptomic and genomic data for large numbers of cancer biospecimens has enabled the identification of recurrent gene expression and sequence alterations in common cancer types (Watson et al. 2013). Similarly, many individual gene alterations are present in only small percentages of tumours diagnosed (Lawrence et al. 2014). Depending upon the inherent mutation rate, which varies substantially between cancer types (Lawrence et al. 2013), very large sample cohorts will be required to document all relevant genetic changes in many cancers (Lawrence et al. 2014). While this may appear to be an exercise in diminishing returns, the presence of targetable alterations in even small proportions of different cancers can extend the benefits of personalised medicine to patients beyond those groups for which the targeted therapy was designed. Good examples are the agents/drugs which target ERBB2, which were originally developed for treating ERBB2-positive breast cancer. Recognition of less frequent ERBB2 alterations in other cancers has allowed these drugs to benefit additional patient populations characterised by poor disease outcomes (Yarden and Pines 2012). It is also important to recognise that these fundamental advances in medicine require the use of biospecimens at every stage (Fig. 1).

Overview of cancer biobanks

Cancer biospecimens are increasingly sourced from formal cancer biobanks (Simeon-Dubach and Watson 2014; Vaught and Lockhart 2012). We have previously defined a cancer biobank as a standardised, not-for-profit collection of human biospecimens with associated data records and consent process for the purpose of distribution to cancer researchers for population or disease-specific research (Rush et al. 2015). Cancer biobanks vary in size, the number of staff employed, the types of specimens collected (cancer vs. normal samples), the formats of specimens collected, the types of clinical and sample information collected, their underpinning financial support models and, most importantly, in terms of their biospecimen access policies (Rush et al. 2015). Some cancer biobanks are linked with particular research studies and/or

Fig. 1 Independent biospecimen cohorts (labelled A–E) are required in order to translate research discoveries into clinical practice. A single biobank is shown for illustrative purposes, but in practice, biospecimen cohorts are likely to be sourced from and/or deposited in multiple biobanks. Discovery and validation processes require independent biospecimen cohorts (A–C). Clinical trials both require and generate biospecimens (cohort D), which may also be subsequently used for future research, and clinical practice generates biospecimens (cohort E) for future research and diagnostic use

only make samples available to collaborators, whereas others support unspecified research and as such employ open access policies (Watson and Barnes 2011).

Cancer biobanking is not a new activity, but since approximately 2000, there has been greater investment in and focus upon biobanking as a requirement for robust cancer research (Ahmed 2011; Barnes et al. 2008; Zika et al. 2011). Despite this increased focus on the requirement for biobanks to store and supply cancer biospecimens of sufficient quality, The Cancer Genome Atlas (TCGA) sequencing projects encountered major and unexpected challenges in sourcing the required frozen tumour and matched normal samples that met histological quality control assessments (Blow 2009; Olson and Berger 2011). This reported crisis in biospecimen sourcing was reflected in survey results indicating that many cancer researchers were unable to access sufficient quantities of high-quality biospecimens (Masset et al. 2011). Similarly, a more focused study of academic and industry leaders revealed that biospecimen quality was ranked as the top challenge to obtaining the biospecimens required by both participant groups (Myles et al. 2011).

Access to biospecimens of sufficient quality remains a major challenge for cancer researchers today. Biological reagents and reference materials have been described as the largest single contributor to irreproducibility of pre-clinical results (Freedman et al. 2015). Ongoing difficulties in locating and obtaining appropriate biospecimens of high quality lead many biomarker researchers to continue to employ “samples of convenience”, which are by definition accessible, but of uncertain integrity (Anderson et al. 2013). Such studies produce the lowest level of associated evidence, but are the most common type of biomarker study reported (Simon et al. 2009). The continued support of uncoordinated, fragmented research

using heterogenous biospecimens in non-standardised assays will therefore only delay the availability of candidate markers entering clinical validation (Poste 2012).

Introduction to biospecimen quality

Many individual factors impact upon the quality of a biospecimen, which can also be described as whether it is fit for purpose (Lim et al. 2011; Moore et al. 2011). Cancer biospecimens can be used in a variety of experiments requiring DNA, RNA, protein, lipids and other cellular constituents. These cellular components exhibit different stability and robustness (Shabikhani et al. 2014). For example, it is widely accepted that DNA is relatively stable, being designed to carry genetic information between generations. In contrast, the stability of RNA and protein species varies widely according to their function and/or site of expression or activity. Post-translational modifications to proteins and metabolism intermediates have in many cases evolved to signal transient information and are therefore highly labile (LaBaer 2012). Consequently, biospecimen quality can be considered as a sliding scale, depending upon the end user’s specific experimental requirements. It is possible that a given set of biospecimens may be of sufficient quality for DNA sequence analyses, but that the same cohort may perform variably, or poorly, in a proteomics study. It is also recognised that fitness for purpose will evolve with increasing knowledge, or may yet to be discovered. For this reason, some biobanks advocate minimal sample processing in order to cater for future applications where requirements are as yet unknown (Elliott et al. 2008). The advantage of storing biospecimens intact without processing was illuminated through the discovery of

microRNAs that would have been lost through routine processing for protein, mRNA and DNA (Watson et al. 2014).

The biospecimen lifecycle—what can and cannot be controlled

The biospecimen lifecycle has been described to explain the steps involved in biospecimen collection and processing (Moore et al. 2011). The lifecycle involves specimens being sampled, collected, processed, stored within the biobank, shipped to recipient(s), stored within the recipient's site and subjected to analysis procedures, with the additional possibility that excess material may also be shipped back to the biobank for further storage and redistribution (Fig. 2). The overall pre-analytical phase includes the lifecycle stages from patient to sample distribution (Fig. 2). Variations at each stage of the biospecimen life cycle can individually and collectively impact sample quality or whether a sample is fit for purpose.

Biospecimens have been described as being subjected to “industrial-strength biologic stresses” during the sample collection process, which take place both before and after sample acquisition (Olson and Berger 2011). These stressors can activate pro-survival signalling pathways within the biospecimen, which could be confused with underlying disease processes, particularly when high-resolution, multi-analyte assays are employed (Lim et al. 2011). Furthermore, in the context of cancer biobanking, if matched tumour and normal specimens are not subjected to the same collection protocols, differences in processing may create additional apparent class differences beyond those caused by disease status (Lim et al. 2011). Differences in biospecimen quality may therefore contribute to or further complicate the interpretation of batch effects within high-throughput data (Leek et al. 2010).

Some elements of the biospecimen lifecycle cannot be controlled by biobanks (Riegman et al. 2015). The disease within the host usually cannot be influenced (Olson and Berger 2011), although it has been proposed that prospective biobanking can lead to earlier disease diagnosis and similarly to better patient outcomes (LaBaer 2012). Anaesthesia (type and duration) and arterial clamp time reflect surgical requirements, as opposed to those of downstream biospecimen researchers (Olson and Berger 2011; Lim et al. 2011). Similarly, sampling and collection procedures may reflect clinical requirements and protocols, as opposed to those preferred by biobanks and researchers. Variables associated with post-sample acquisition include ambient temperature and duration, fixatives used and fixation procedures, rate of freezing, storage temperature and aliquot sizes (Olson and Berger 2011). Once a biospecimen is under the control of a biobank, it is important that collection, processing, storage and shipping occur according to defined, evidence-based protocols which

are in place to maximise sample integrity. Considerations regarding the procurement, optimal storage temperatures (including freeze–thaw cycles) and quality control (QC) tools for tissue and blood samples have recently been reviewed in detail elsewhere (Ahmed 2011; Hubel et al. 2014; Shabikhani et al. 2014). Biophysics research using methods such as confocal Raman spectroscopy (Dong et al. 2010) has been critical for increasing our understanding of nucleation and ice crystal growth within cells and tissues, and has thereby provided an evidence base for biospecimen freezing and low-temperature storage protocols (Hubel et al. 2014). The comparative analysis of other biophysical properties of transformed versus non-transformed cells, such as resistance to fluid shear stress (Barnes et al. 2012), may also be of future relevance to bio-preservation technology.

Biospecimen QC in cancer biobanks

Biobanks undertake a range of possible QC activities to measure or predict biospecimen quality. The results of these activities, which include confirming patient identity, histological diagnosis, and percentage of tumour cells, as well as routine database audits (Shabikhani et al. 2014), may determine whether samples are retained within the biobank and/or provided to researchers (Herpel et al. 2010) (Fig. 3). The net effects of variations in pre-analytical factors are typically assessed using measures of sample integrity, such as hemolysis in the case of blood samples, and the integrity of extracted RNA and DNA. In the case of proteins, quantitative and sample integrity reference proteins are currently lacking, but these would be welcomed by proteomics researchers (LaBaer 2012).

Concomitantly with authentication becoming more frequent among the large community of researchers employing cell lines (Geraghty et al. 2014; Lichter et al. 2010), the potential for sample misidentification in biobanks is being recognised and acted upon, although data regarding the rates of sample misidentification are generally lacking. Sample authentication can be achieved by STR profiling of archived blood spots on Guthrie cards collected prior to sample archiving that can be assayed as required (Cardoso et al. 2010). However, this assay is being superseded by high-throughput technologies that allow precise sample identification by molecular bar coding, with iPlex Pro Sample ID panel of Agena Bioscience (San Diego, CA) and the SNP Trace ID panel of Fluidigm Corp. (South San Francisco, CA) leading in this area. The iPlex Pro Sample ID panel uses MassARRAY technology to annotate 44 single nucleotide polymorphisms (SNPs), three gender markers and five DNA copy number controls. Alternatively, the SNP Trace ID panel uses the BioMark HD real-time PCR system and incorporates 96 SNPs to provide accurate authentication of samples.

Fig. 2 The different elements of the biospecimen lifecycle. *Single-headed arrows* temporal sequence of events involving biospecimens and/or information, *double-headed arrows* biospecimen lifecycle stages reported by the Standard PReanalytical Code (SPREC) and/or Biospecimen Reporting for Improved Study Quality (BRISQ)

Biospecimen QC reporting by cancer biobanks

Most cancer researchers have indicated that they want more information about the biospecimens that they use (Masset et al. 2011). For some biospecimen quality measures, researchers can perform their own assays to derive quality measures, such as A260/A280 ratios to measure nucleic acid purity, and RNA integrity numbers and other measures of nucleic acid integrity. In the case of samples supplied as formalin-fixed, paraffin-embedded specimens, researchers can request local advice to confirm histopathological diagnoses and percentages of tumour cells. This is likely to be more challenging in the case of frozen samples, and such assays can also reduce the amount of tissue that remains available for research. However, some quality measures cannot be simply replicated

by researchers, such as confirmation of sample identity and the integrity of data supplied (Kiehnopf and Krawczak 2011). It is also important to recognise that experimental QC assay results reflect the cumulative effects of potentially many individual pre-analytical variables, and provide little or no information about which variables most directly contributed to biospecimen quality in particular cases. However, biobanks are ideally placed to directly measure or otherwise collate a range of information relevant to the biospecimen lifecycle and to communicate this information to researchers.

Data from a 2012 survey of U.S. biobank managers (Henderson et al. 2013a) showed that 78 % of biobanks surveyed use standardised Material Transfer Agreements (MTAs) to communicate information from biobanks to researchers (Henderson et al. 2013b). MTAs can therefore also

Fig. 3 Biospecimen and information flow between biobanks and researchers. Activities shown in the box delineated by dotted lines take place within biobanks, whereas activities within the box delineated by broken lines take place within research laboratories. Biospecimens and information are supplied to researchers, based upon the results of quality control (QC) assays conducted within the biobank and the researcher's requirements. Biospecimens may then be subjected to further QC analysis by researchers, the results of which may be fed back to the biobank

be used as a tool to communicate biospecimen quality and/or limitations, as well as to describe requirements for the return of data and/or biospecimens from researchers back to the biobank (Simeon-Dubach and Watson 2014). The results of QC assays can also be summarised in reporting tools, such as the Standard PREanalytical Code (SPREC) and Biospecimen Reporting for Improved Study Quality (BRISQ), which complement methods designed to improve the reporting of tumour biomarker research results [REporting recommendations for tumor MARKer prognostic studies (REMARK)] and diagnostic accuracy [Standards for Reporting of Diagnostic Accuracy (STARD)] (Bossuyt et al. 2003; McShane and Hayes 2012).

The SPREC was developed in 2009 by members of the International Society for Biological and Environmental Repositories (ISBER) Working Group on Biospecimen Science, who recognised a requirement for standardised nomenclature to describe biospecimen handling, from specimen collection to analysis. This seven-element code differs slightly for fluids and solid tissues and is designed to integrate into biobank quality management and inventory systems (Betsou et al. 2010). The SPREC was updated in 2012, in response to user feedback and a broadening of its scope (Lehmann et al. 2012), and is now incorporated, in compliance with the aim of the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI), into the Swedish data set MIABIS (Minimum Information About Biobank Information Sharing) (Norlin et al. 2012) and interfaces with protocols submitted to the Molecular Methods Database (Klingström et al. 2013). To facilitate biobank use of this coding system, two support tools have been developed for generating (SPRECalc) and encoding (SPRECware) relevant information (Lehmann et al. 2012).

The BRISQ was developed in 2011 in response to identified variations in the amount and type of preanalytical information reported in the literature (Moore et al. 2011, 2012, 2013). The authors created a tiered coding system to be used by biorepositories in reporting biospecimen quality variables to researchers and by researchers in reporting these variables in publications (Moore et al. 2011, 2012, 2013). The elements in BRISQ were included to enable evaluation and replication of reported studies involving human biospecimens, and include references to each stage of the biospecimen lifecycle (Moore et al. 2011, 2012, 2013). Tier 1 of the coding system includes items considered to be recommended to report, with lesser priority assigned to Tier 2 (items beneficial to report) and Tier 3 (additional items to report) (Moore et al. 2011, 2012, 2013). Reporting variables for the SPREC and BRISQ codes overlap somewhat, however the SPREC places greater focus on the biospecimen collection mechanism (Fig. 2) and includes a standardised method for reporting variables (Lehmann et al. 2012). Conversely, the BRISQ encompasses more stages of the biospecimen timeline (Fig. 2), including those prior to biobank specimen receipt, such as patient

disease status and clinical characteristics, and allows the direct reporting of biospecimen QC results (Moore et al. 2011).

Although the SPREC and BRISQ codes are not intended to replace evaluations of biospecimen quality, they represent an important tool for transparent reporting of the biospecimen quality measures that have, and have not, been undertaken. Both tools acknowledge that biobanks and/or researchers may not be able to report on an entire suite of variables due to procedural or logistical shortfalls (Lehmann et al. 2012; Moore et al. 2011, 2012, 2013; Simeon-Dubach and Moore 2014). However, reporting on preanalytical variables where at all possible is purported to improve translational study outcomes (Simeon-Dubach and Moore 2014). This has been recognised by *The Journal of Pathology*, *Histopathology* (Simeon-Dubach et al. 2012) and, more recently, *Biopreservation and Biobanking* (Simeon-Dubach and Moore 2014), which all endorse the use of BRISQ reporting in manuscript submissions. Indeed, the main driver for reporting complete BRISQ data sets may represent journal requirements enforced through the reviewer community (Braun et al. 2014). In response to the poor reporting of Tier 1 BRISQ data elements in the cancer research literature, an online BRISQ reporting tool has been created (Cheah et al. 2012). The BRISQ Report tool allows biobanks to provide BRISQ data to biospecimen recipients in the form of a structured report, which can be included in submitted/published manuscripts as supplementary data (Cheah et al. 2012).

Biobank networks as a means of improving standardisation and harmonisation

Stand-alone biobanks have been reported to experience increased difficulty in quality management and standardisation implementation, with challenges relating to a lack of harmonised standard operating procedures (SOPs) and biobanking guidelines (Zhang et al. 2015). Developing separate SOPs based on the lack of (or variable) best practice documentation can lead to inter-biobank variations in pre-analytical handling of biospecimens (Kiehnopf and Krawczak 2011; Morente et al. 2006; Watson et al. 2011; Yuille et al. 2010). Biobank networking links can encourage a greater focus on biospecimen quality (Meir et al. 2014) via the use of best practice protocols (Catchpoole et al. 2007). For example, funding for the Canadian Tumour Repository Network (CTRNet) is restricted to the provision of network activities, such as certification programmes, SOPs and biobanking policies (Watson et al. 2014). Biobank networking confers advantages to researchers by providing capacity to access larger sample numbers and increasing experimental statistical power (Asslaber and Zatloukal 2007), and to funding bodies through more effective use

of infrastructure and resources (Kiehnopf and Krawczak 2011; Morente et al. 2006; Zika and Defrasne 2010). For example, a networked biobank hub model may allow hiring and cross-training of dedicated staff, as well as sharing of high-cost equipment (Simeon-Dubach and Watson 2014).

Biobank networks are located across North America, Europe and the Asia-Pacific region (Ahmed 2011; Catchpoole et al. 2007; Meir et al. 2014), and are usually based on common geography, sample type/s collected, and/or formats of supplied samples. However, there is still further potential to develop additional, improved biobank networks (Meir et al. 2014; Zika et al. 2011). A survey of European biobanks reported increasing biobanking activity in recent years (Zika et al. 2011). This increases the likelihood of biobank networking and provides strong motivation to pursue further linkages between biobanks, both nationally and internationally (Zika et al. 2011). A major barrier to biobank networking has indeed been the lack of biobank inventories (Ahmed 2011), which may in turn reflect challenges in identifying individual biobanks (Boyer et al. 2012). This was reflected in a European survey, where variable numbers of biobanks per country reflected the existence/absence of national biobank networks (Zika et al. 2011). The need for better biobank inventories is being recognised through, for example, an on-line catalogue of European biobanks established through the European Union BBMRI (Wichmann et al. 2011).

Biobank accreditation/certification

It has been argued that achieving the strict standards required for genome-based medicine will require a level of biobank oversight achievable through either voluntary accreditation/certification or an official regulatory framework (Browman 2012). Standards could also be applied by institutional review boards to research proposals involving biobanks (Browman 2012). Accreditation of biobanks has been proposed to promote the adherence to regulations by biobank staff members, as regulations are controlled through audits and reviews (Herpel et al. 2010).

The relative merits of biobank certification (which addresses adherence to designated standards) and accreditation (which addresses result-oriented performance and competence to carry out specific tasks) have been debated (Herpel et al. 2010; Matzke et al. 2012). CTRNet established a certification programme in response to strong support from Canadian biobank stakeholders (Matzke et al. 2012). This programme includes requesting a set of biospecimens from each biobank, although the results of this assessment are not employed as a factor in determining certification (Matzke et al. 2012). In contrast, Herpel et al. (2010) advocated an

accreditation model, as they concluded that certification did not permit the assessment of core biobanking practices, such as entry and exit tissue controls. Regardless of whether individual biobanks choose accreditation or certification, barriers exist in terms of the personnel and resources needed to prepare for and complete the accreditation process, and to maintain these standards going forward (Herpel et al. 2010).

Implementation and effectiveness of biobank QC activities

Despite increasing recognition of the importance of biospecimen quality, there is limited evidence in the literature that QC reporting systems are being taken up at the biobank coalface, and that information about biospecimen quality is routinely communicated to researchers. A 2003 review of human biobanks in six European Union countries concluded that while good practice guidelines were generally followed and quality controls performed, quality procedures were not always clearly explained (Hirtzlin et al. 2003). An analysis of a cross-section of publications in two cancer journals in 1998, 2004 and 2010 revealed poor reporting of most BRISQ biospecimen-related data elements, and this trend did not significantly improve over the study period (Cheah et al. 2012). More focussed literature analyses also revealed significant differences in the reporting of clinical chart versus biobank BRISQ data field groups (Braun et al. 2014). Clinical chart fields were reported in at least 50 % of the manuscripts examined, whereas some biobank chart fields were reported in less than 25 % of publications (Braun et al. 2014). While it is unclear whether rates of BRISQ chart field reporting reflect the priorities of biobanks, researchers or both groups, this result suggests an underestimation of the influence of pre-analytical factors upon research results. Any such underestimation by researchers could partly derive from biospecimen science publications appearing in a range of scientific journals, leading to researchers being unaware of these collective results (Moore 2012).

The biobank literature frequently cites the reasonable assumption that networking and the top-down imposition of standards through certification and/or accreditation will improve adherence to standards, and increase inter-bank operational harmonisation (Ahmed 2011; Matzke et al. 2012). However, there is little demonstrable evidence that this occurs in practice. Perhaps this is partly because biobank accreditation/ certification are relatively recent activities, and there has been insufficient time to reliably measure their impact on downstream processes. There are similarly few reports describing which QC activities are undertaken by biobanks that do not belong to formal networks, or have not submitted to accreditation or certification. In a large US biobank cohort, it was unclear to

which extent best practices were followed in the absence of systems of accreditation (Henderson et al. 2013b). However, in a smaller Australian biobank cohort, we noted significantly improved reported awareness of biobanking best practices and use of QC measures by poly-user versus mono/oligo-user biobanks (Rush et al. 2015). Open-access (or poly-user) biobanks also employed more staff, were more likely to provide formal staff training, and reported better access to governance support (Rush et al. 2015). There is also little published information regarding how biobanks communicate measures of sample quality to researchers, despite the availability of appropriate tools (Cheah et al. 2012). Further research is clearly required to better describe understand barriers to biobank QC activities and reporting in a range of biobank settings.

Biospecimen QC reporting from researchers to biobanks

In addition to biobanks communicating pre-analytical data and measures of biospecimen quality to researchers, there is also scope for researchers to return unused samples, research results and independent measures of biospecimen quality to biobanks (Fig. 3). In a European Union biobank survey, the majority of biobanks required researcher feedback regarding experimental outcomes, as well as sample return, although the degree of compliance with these requirements was not reported (Zika et al. 2011). In a U.S. cohort, biobanks that required researchers to pay for biospecimens (in addition to handling and shipping costs) were significantly less likely to require biospecimen return, suggesting that payment for biospecimens may result in less reach-through control over specimens and the data generated (Henderson et al. 2013b). A reduced requirement for researchers to communicate with biobanks may therefore be an unintended consequence of increased pressures upon biobanks to achieve cost recovery. There is currently heightened interest in the role of biobanks in managing returned research findings and incidental results (Bledsoe et al. 2013; Wolf et al. 2012). At present, these discussions pertain to the return of clinically significant information (Bledsoe et al. 2013; Wolf et al. 2012) and do not yet extend to measures of biospecimen quality and suitability for purpose. Nonetheless, any focus upon researcher–biobank communication is welcome and should highlight how little attention this topic has previously received.

Summary and concluding remarks

Biobanks and biospecimen researchers have a naturally symbiotic relationship that could be further enhanced for the

benefit of many. Cancer researchers need access to large cohorts of high-quality biospecimens in order to generate robust data. This increases the demand upon biobanks, which is a welcome pressure, as biobanks increasingly need to demonstrate that their samples are used for research (Simeon-Dubach and Watson 2014). Supplying samples of sufficient quantity and quality also allows biobanks to receive contributions from researchers, such as manuscript co-authorship and cost reimbursement (Riegman et al. 2015).

Biobanks and their research clients are indeed ideal partners in the study and promotion of the use of high-quality biospecimens in research. Biobank staff have unique access to and an unparalleled capacity to report on a range of pre-analytical variables. Although researchers can perform some degree of independent validation of biospecimen quality, they are heavily reliant upon biobanks for detailed information on how biospecimens were sourced and processed. In turn, researchers are expert in their fields of investigation, and in the analytes that they are assessing, and have access to further expertise through collaborative networks (Lim et al. 2011). The routine supply of detailed biospecimen data and subsequent integration of these variables into research analyses could lead to the identification of unexpected associations between pre-analytical variables and research measures, just as collaborations between laboratory scientists and data analysts can facilitate the identification of batch effects (Leek et al. 2010). This would extend biospecimen science from a discrete field to a routine activity carried out by all biospecimen researchers. Better communication between biobanks and researchers could have other benefits. Researchers can educate biobank staff about their biospecimen format and quality requirements, which could assist the biobank in tailoring its current activities and improving its long-term sustainability (Cole et al. 2012; Simeon-Dubach and Watson 2014). In turn, biobank staff can help researchers to understand how biobank operations affect biospecimen quality, allowing researchers to design better, more feasible experiments and overall studies. There needs to be a much greater emphasis placed on the importance of ongoing communication between biobanks and researchers, not least because this will be a requirement for the return of research results. As one way to progress this, cancer research conferences could aim to include more presentations about biospecimen quality and research reproducibility in shared conference sessions involving both cancer researchers and biobankers.

There are many gaps in our current knowledge. We lack a detailed understanding of how pre-analytical variables impact upon individual research analytes. For example, there is a lack of information on the effects of physical forces on solutes during liquid–solid transitions that could contribute to the development of better evidence-based freeze–thaw processes (Lim et al. 2011). We need more information on the QC

activities that are actually performed by biobanks, as opposed to the activities that are recommended or expected by top-down networking models. We need to understand the level of QC result reporting back to researchers, which has been poorly investigated, so that the uptake of QC reporting in the research literature can be understood and improved. While improved biospecimen data reporting should increase the rate of research result validation and translation, research is also required to examine whether this occurs in practice (Cheah et al. 2012). Similarly, we need baseline descriptions of biobank–researcher communication, in order to design and test strategies for this communication to improve.

There has been a great deal of discussion around the ethics of biobanking, which tends to focus upon the custodianship of samples, privacy protection and data return to participants (Wheelock et al. 2015). However, the use of inadequate biospecimens in research is also unethical, leading to wasted research resources (Vaught and Lockhart 2012) and downstream consequences for patients and healthcare systems. Improving our understanding and assessment of biospecimen quality, and of how communication between biobanks and researchers can improve biospecimen use, could have a substantial impact upon the validity of research obtained from human biospecimens, and thereby accelerate the translation of research results to patients.

Acknowledgements AR and JB gratefully acknowledge funding from the Cancer Institute NSW (14/BIO/1-05) to the Kids Cancer Alliance, which supported the preparation of this review. KS gratefully acknowledges funding support from CONCERT, a Cancer Institute NSW-funded Translational Cancer Research Centre. We are also grateful to members of the Cancer Institute NSW Biobanking Stakeholder Network and the Children’s Hospital at Westmead Tumour Bank for discussions.

Compliance with Ethical Standards

Conflict of interest None declared.

Ethical approval This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Ahmed FE (2011) Biobanking perspective on challenges in sample handling, collection, processing, storage, analysis and retrieval for genomics, transcriptomics and proteomics data. *Anal Methods* 3: 1029–1038
- Anderson NL, Ptolemy AS, Rifai N (2013) The riddle of protein diagnostics: future bleak or bright? *Clin Chem* 59:194–197
- Asslaber M, Zatloukal K (2007) Biobanks: transnational, European and global networks. *Brief Funct Genomics* 6:193–201
- Barnes RO, Parisien M, Murphy LC et al (2008) Influence of evolution in tumor biobanking on the interpretation of translational research. *Cancer Epidemiol Biomarkers Prev* 17:3344–3350
- Barnes JM, Nauseef JT, Henry MD (2012) Resistance to fluid shear stress is a conserved biophysical property of malignant cells. *PLoS One* 7: e50973
- Betsou F, Lehmann S, Ashton G et al (2010) Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. *Cancer Epidemiol Biomarkers Prev* 19:1004–1011
- Bledsoe MJ, Clayton EW, McGuire AL et al (2013) Return of research results from genomic biobanks: cost matters. *Genet Med* 15:103–105
- Blow N (2009) Biobanking: freezer burn. *Nat Methods* 6:173–178
- Boehm JS, Golub TR (2015) An ecosystem of cancer cell line factories to support a cancer dependency map. *Nat Rev Genet* 16:373–374
- Bossuyt PM, Reitsma JB, Bruns DE et al (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Clin Biochem* 40:357–363
- Bowen A, Casadevall A (2015) Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. *Proc Natl Acad Sci USA* 112:11335–11340
- Boyer GJ, Whipple W, Cadigan RJ et al (2012) Biobanks in the United States: how to identify an undefined and rapidly evolving population. *Biopreserv Biobank* 10:511–517
- Braun L, Lesperance M, Mes-Massons AM et al (2014) Individual investigator profiles of biospecimen use in cancer research. *Biopreserv Biobank* 12:192–198
- Browman GP (2012) Special series on comparative effectiveness research: challenges to real-world solutions to quality improvement in personalized medicine. *J Clin Oncol* 30:4188–4191
- Buck S (2015) Solving reproducibility. *Science* 348:1403
- Cardoso S, Valverde L, Odriozola A et al (2010) Quality standards in Biobanking: authentication by genetic profiling of blood spots from donor’s original sample. *Eur J Hum Genet* 18:848–851
- Catchpoole D, de Fazio A, Devereux L et al (2007) The importance of biorepository networks: the Australasian biospecimen network- oncology. *Aust J Med Sci* 28:16–20
- Cheah S, Dee S, Cole A et al (2012) An online tool for improving biospecimen data element reporting. *Biopreserv Biobank* 10:501–510
- Cole A, Cheah S, Dee S et al (2012) Biospecimen use correlates with emerging techniques in cancer research: impact on planning future biobanks. *Biopreserv Biobank* 10:518–525
- Dong J, Malsam J, Bischof JC et al (2010) Spatial distribution of the state of water in frozen mammalian cells. *Biophys J* 99:2453–2459
- Elliott P, Peakman TC, Biobank UK (2008) The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 37:234–244
- Freedman LP, Cockburn IM, Simcoe TS (2015) The economics of reproducibility in preclinical research. *PLoS Biol* 13:e1002165
- Geraghty RJ, Capes-Davis A, Davis JM et al (2014) Guidelines for the use of cell lines in biomedical research. *Br J Cancer* 111:1021–1046
- Henderson GE, Cadigan RJ, Edwards TP et al (2013a) Characterizing biobank organizations in the U.S.: results from a national survey. *Genome Med* 5:3
- Henderson GE, Edwards TP, Cadigan RJ et al (2013b) Stewardship practices of U.S. biobanks. *Sci Transl Med* 5:215cm7
- Herpel E, Röcken C, Manke H et al (2010) Quality management and accreditation of research tissue banks: experience of the National Center for Tumor Diseases (NCT) Heidelberg. *Virchows Arch* 457:741–747
- Hicks DG, Kushner L, McCarthy K (2011) Breast cancer predictive factor testing: the challenges and importance of standardizing tissue handling. *J Natl Cancer Inst Monogr* 2011:43–45
- Hirtzlin I, Dubreuil C, Préaubert N et al (2003) An empirical survey on biobanking of human genetic material and data in six EU countries. *Eur J Hum Genet* 11:475–488
- Hubel A, Spindler R, Skubitz AP (2014) Storage of human biospecimens: selection of the optimal storage temperature. *Biopreserv Biobank* 12:165–175
- Ioannidis JP (2014) How to make more published research true. *PLoS Med* 11:e1001747

- Kiehntopf M, Krawczak M (2011) Biobanking and international interoperability: samples. *Hum Genet* 130:369–376
- Klingström T, Soldatova L, Stevens R et al (2013) Workshop on laboratory protocol standards for the molecular methods database. *Nat Biotechnol* 30:109–113
- LaBaer J (2012) Improving international research with clinical specimens: 5 achievable objectives. *J Proteome Res* 11:5592–5601
- Lawrence MS, Stojanov P, Polak P et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214–218
- Lawrence MS, Stojanov P, Mermel CH et al (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505:495–501
- Leek JT, Scharpf RB, Bravo HC et al (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11:733–739
- Lehmann S, Guadagni F, Moore H et al (2012) Standard preanalytical coding for biospecimens: review and implementation of the sample PREanalytical code (SPREC). *Biopreserv Biobank* 10:366–374
- Lichter P, Allgayer H, Bartsch H et al (2010) Obligation for cell line authentication: appeal for concerted action. *Int J Cancer* 126:1
- Lim MD, Dickherber A, Compton CC (2011) Before you analyze a human specimen, think quality, variability, and bias. *Anal Chem* 83:8–13
- Masset HA, Atkinson NL, Weber D et al (2011) Assessing the need for a standardized cancer HUMAN Biobank (caHUB): findings from a national survey with cancer researchers. *J Natl Cancer Inst Monogr* 2011:8–15
- Matzke EM, O'Donoghue S, Barnes RO et al (2012) Certification for biobanks: the program developed by the Canadian tumour repository network (CTRNet). *Biopreserv Biobank* 10:426–432
- McShane LM, Hayes DF (2012) Publication of tumor marker research results: the necessity for complete and transparent reporting. *J Clin Oncol* 30:4223–4232
- Meir K, Cohen Y, Mee B et al (2014) Biobank networking for dissemination of data and resources: an overview. *J Biorepos Sci Appl Med* 2:29–42
- Micheel CM, Nass SJ, Omenn GS (2012) Evolution of translational omics: Lessons learned and the path forward. National Academies Press, Washington DC
- Moore H (2012) The NCI biospecimen research network. *Biotech Histochem* 87:18–23
- Moore HM, Kelly AB, Jewell SD et al (2011) Biospecimen reporting for improved study quality (BRISQ). *J Proteome Res* 10:3429–3438. *Cancer Cytopathol* 119:92–101
- Moore HM, Kelly A, McShane LM et al (2012) Biospecimen reporting for improved study quality (BRISQ). *Clin Chim Acta* 413:1305
- Moore HM, Kelly A, McShane LM et al (2013) Biospecimen reporting for improved study quality (BRISQ). *Transfusion* 53:e1. doi:10.1111/trf.12281
- Morente MM, Mager R, Alonso S et al (2006) TuBaFrost 2: Standardising tissue collection and quality control procedures for a European virtual frozen tissue bank network. *Eur J Cancer* 42:2684–2691
- Myles R, Massett HA, Comey G et al (2011) 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* 2011:16–23
- Norlin L, Fransson MN, Eriksson M et al (2012) A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank* 10:343–348
- Olson S, Berger AC (2011) Establishing precompetitive collaborations to stimulate genomics-driven product development: workshop summary. National Academies Press, Washington DC
- Poste G (2012) Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. *Trends Mol Med* 18:717–722
- Riegman PH, de Jong B, Daidone MG et al (2015) Optimizing sharing of hospital biobank samples. *Sci Transl Med* 7:297fs231
- Rush A, Christiansen JH, Farrell JP et al (2015) Biobank classification in an Australian setting. *Biopreserv Biobank* 13:212–218
- Saletta F, Wadham C, Ziegler DS et al (2014) Molecular profiling of childhood cancer: biomarkers and novel therapies. *BBA Clin* 1:59–77
- Shabikhani M, Lucey GM, Wei B et al (2014) The procurement, storage, and quality assurance of frozen blood and tissue biospecimens in pathology, biorepository, and biobank settings. *Clin Biochem* 47:258–266
- Simeon-Dubach D, Moore HM (2014) BIO comes into the cold to adopt BRISQ. *Biopreserv Biobank* 12:223–224
- Simeon-Dubach D, Watson P (2014) Biobanking 3.0: evidence based and customer focused biobanking. *Clin Biochem* 47:300–308
- Simeon-Dubach D, Burt AD, Hall PA (2012) Quality really matters: the need to improve specimen quality in biomedical research. *J Pathol* 228:431–433. *Histopathology* 61:1003–1005
- Simon RM, Paik S, Hayes DF (2009) Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 101:1446–1452
- Vaught J, Lockhart NC (2012) The evolution of biobanking best practices. *Clin Chim Acta* 413:1569–1575
- Vogelstein B, Papadopoulos N, Velculescu VE et al (2013) Cancer genome landscapes. *Science* 339:1546–1558
- Watson IR, Takahashi K, Futreal PA et al (2013) Emerging patterns of somatic mutations in cancer. *Nat Rev Genet* 14:703–718
- Watson PH, Barnes RO (2011) A proposed schema for classifying human research biobanks. *Biopreserv Biobank* 9:327–333
- Watson PH, Ravid R, Eng CB et al (2011) What are the main roadblocks to transnational biobank collaboration, and how can we overcome them? *Biopreserv Biobank* 9:213–216
- Watson PH, Nussbeck SY, Carter C et al (2014) A framework for biobank sustainability. *Biopreserv Biobank* 12:60–68
- Wheelock AM, Paulson L, Litton JE et al (2015) The EuPA Biobank Initiative: meeting the future challenges of biobanking in proteomics & systems medicine. *J Proteomics*. doi:10.1016/j.jprot.2015.07.014
- Wichmann HE, Kuhn KA, Waldenberger M et al (2011) Comprehensive catalog of European biobanks. *Nat Biotechnol* 29:795–797
- Wilding JL, Bodmer WF (2014) Cancer cell lines for drug discovery and development. *Cancer Res* 74:2377–2384
- Wolf SM, Crock BN, Van Ness B et al (2012) Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med* 14:361–384
- Wolff AC, Hammond ME, Hicks DG et al (2013) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 31:3997–4013
- Yarden Y, Pines G (2012) The ERBB network: at last, cancer therapy meets systems biology. *Nat Rev Cancer* 12:553–563
- Yuille M, Illig T, Hveem K et al (2010) Laboratory management of samples in biobanks: European consensus expert group report. *Biopreserv Biobank* 8:65–69
- Zhang L, Wu X, Hu Y et al (2015) Establishment of a network-based intra-hospital virtual cancer biobank. *Biopreserv Biobank* 13:43–48
- Zika E, Paci D, Braun A et al (2011) A European survey on biobanks: trends and issues. *Public Health Genomics* 14:96–103