CHAPTER 3.

# Biological sample collection, processing, storage and information management

*Jimmie B. Vaught and Marianne K. Henderson*



## Summary

The collection, processing and storage of biological samples occur in the larger context of organizations known as biological resource centres or biospecimen resources. Biological resource centres are (1,2) service providers and repositories of living cells, as well as genomes of organisms, archived cells and tissues, and information relating to these materials. The US National Cancer Institute (3) defines a biospecimen resource as a "… collection of human specimens and associated data for research purposes, the physical entity where the collection is stored, and all relevant processes and policies." The complexities involved in proper sample management policies and procedures are often underestimated. Prior to initiating a study that will involve the collection of biological samples, many decisions need to be made that will affect the quality of the samples and the outcome of the study. The appropriate sample type(s) needs to be chosen. The processing protocol that will result in samples of suitable quality for the intended laboratory analyses must be selected from among various possible protocols. Consideration must be given to the proper storage conditions to maintain sample quality until analyses are completed. All of these activities must be monitored and controlled by appropriate sample tracking and laboratory informatics systems. A comprehensive quality management system, with standard operating procedures and other appropriate controls, is necessary to assure that biological samples are of consistent quality and right for the intended analyses and study goals.

## Introduction

Although biological specimens have been collected for use in a variety of molecular epidemiology, clinical trial and basic research studies for many years, it has only recently been recognized that the protocols and practices involved in collecting, processing and storing specimens actually comprise "biospecimen science." As a result, many organizations (Appendix 3.1) have engaged in producing guidelines and best practices for these endeavours, now known as biological resource centres or biospecimen resources.

## **Appendix 3.1.** Existing guidelines and best practices for biorepositories

| Title | Authors/Origin | Reference/Link |
|---|---|---|
| Tissue Banking for Biomedical Research | National Cancer Centre/Singapore | http://www.bioethics-singapore.org/uploadfile/52533%20PMHT%20AppendixB-Dr%20Kon.pdf |
| Biorepository Protocols | Australian Biospecimen Network, ABN/ Australia | http://www.abrn.net/ |
| Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology | OECD/International | http://www.oecd.org/dataoecd/55/48/2487422.pdf |
| European Human Frozen Tumor Tissue Bank – TUBAFROST | TUBAFROST/The Netherlands | http://www.tubafrost.org/ |
| Human Tissue and Biological Samples for use in Research: Operational and Ethical Guidelines | MRC/UK | http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002420 |
| Best Practices for Repositories I: Collection, Storage, and Retrieval of Human Biological Materials for Research | International Society for Biological & Environmental Repositories/USA | Cell Preserv Technol 2008;6:3-58 |
| First-Generation Guidelines for NCI-Supported Biospecimen Resources | NCI/USA | http://biospecimens.cancer.gov/bestpractices |
| UN Recommendations on the Transport of Dangerous Goods. Model Regulations | UN Economic Commission for Europe, UNECE/International | http://www.unece.org/trans/danger/publi/unrec/rev13/13files_e.html |
| Specimen Collection, Preparation, and Handling | LabCorp/International | http://www.labcorp.com/datasets/labcorp/html/frontm_group/frontm/section/speccol.htm |

Several organizations have published guidelines and best practices relevant to the discussion in this chapter (54). This table is adapted from the IARC publication *International Network of Biological Resource Centres for Cancer Research: Recommendations on Common Minimal Technical Standards (2).*

These terms reflect the fact that specimen management takes place in an environment that includes a wide range of policies concerning the specimens and data, as well as the physical structure, the biorepository. Biological resource centres are engaged in many activities beyond storage, such as acquiring, processing (e.g. aliquoting, DNA extraction) and distributing biological materials. The practices and policies that have been organized into formal documents testify to the importance of following proper steps that will result in the highest quality specimens for research purposes. The use of proper procedures to produce biological specimens of the appropriate quality, as well as the collection of relevant clinical, epidemiologic and quality control data, gives the biospecimens their value in research.

## Context and public health significance

Biological specimens (or biospecimens), such as blood, urine, saliva, and many other types, are collected for a variety of reasons, for normal patient monitoring and care as well as for basic, clinical and epidemiologic research studies. Many medical advances, including studies of heart disease, AIDS and cancer, have resulted from preliminary developmental studies that have relied on access to and proper use of the appropriate biospecimens. The sources of biospecimens for these studies have been varied, as has their quality (1–4).

For molecular epidemiology studies, the ultimate success of a study depends on reliable laboratory analyses of these specimens. In order for laboratory analyses to be reliable, the collection, processing and storage of specimens must be performed under strictly controlled procedures. As the sensitivity and specificity of analytic techniques have increased to an extraordinary degree in recent years (see Chapter 4), it has become even more important to assure that biospecimens are of the highest quality. In addition, from the point in time that the specimens are collected until laboratory results are analysed and reported, all of the relevant information concerning the specimen, as well as data concerning the study participant and laboratory analyses, must be properly stored in interoperable information management systems. This could mean multiple systems or multiple databases interconnected in a single system. All of these steps must be performed under a well-planned quality assurance programme, and according to relevant legal and ethical standards (discussed in Chapter 2).

### *Examples/case studies*

Prior to initiating a study that involves specimen collection, several key points must be considered. The answers to these questions will be important in determining whether the appropriate materials, equipment and procedures are in place:

• What are the goals of the study?

• What laboratory analyses will be needed to accomplish the study goals?

• What type of biospecimens will be necessary for the intended laboratory analyses?

• How many specimens will be collected? If necessary, a biostatistician should be consulted to assist in determining the number required to achieve statistical significance.

• What volume or size will be required for each specimen to assure that it is adequate for the intended analyses? Will it be necessary to store smaller volumes in aliquots for future unplanned use to avoid thawing a larger aliquot? For example, it is important to consider that new technologies have resulted in more sensitive analytical techniques to apply to older samples (see also Chapters 4 and 7), or older samples may become sources of information to study the natural history of a seemingly 'new' disease.

• What quality standards do the specimens need to meet for valid laboratory analyses? Have such quality measures been validated?

• Have specimen collection, processing and storage protocols been standardized and validated in pilot studies?

• If the specimens will be stored for some period of time before analysis, has the stability of the intended biomarker, or other analyte, been determined for the planned storage conditions?

• Will specimens need to be shipped to distant locations for analysis? If so, have packaging and shipping protocols been validated to assure the stability and safety of the specimens and personnel who will handle them?

• Have all other logistical issues been resolved, such as proper coding, labelling and identifying the types of storage vessels?

• What data will be collected with the sample and the study, and is an appropriate informatics system available to collect and process this information?

• Have all appropriate informed consent, privacy and other ethical and legal rules and regulations been reviewed and adhered to in the study planning?

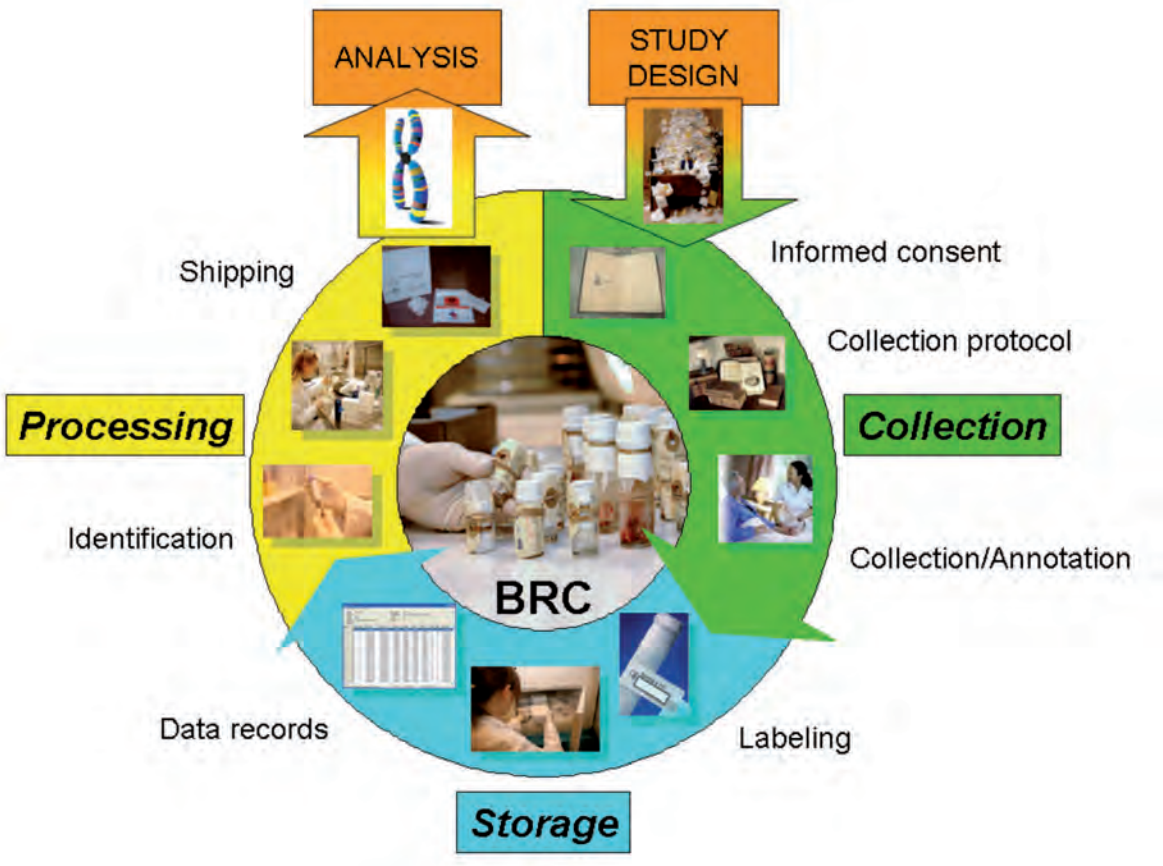
• Are funding and other resources for the proposed study's specimen collection adequate? Will it be necessary to consider lower cost alternate methodologies?

If there is a significant amount of uncertainty in answering the above questions, then additional thought and planning will be needed before beginning the study. For example, before initiating the collection of blood and urine from 500 000 study participants in 2007, the United Kingdom Biobank conducted a series of sample processing validation studies (5–7). These studies showed the effects of sample processing delays, as well as storage conditions, on the results of the wide variety of assays to be conducted on samples that will be collected over a four-year period, but may be used for studies for 20 years or more. The long-term success of such a large and costly project depends on this careful approach to planning the most efficient specimen collection and processing to maintain the stability of the resulting sample aliquots, which are expected to number approximately 15 000 000 (5).

Among the issues outlined above, cost is a major consideration, especially when designing a study that will include a large collection of biospecimens. Often the costs of collecting, processing and storing biospecimens are not well understood or estimated before starting a study. The design and operation of the physical biorepository also needs to be well thought out. Baird and Frome (8) have outlined the major elements of cost and design for a large biorepository. It is also important to plan biospecimen collections with careful attention to the costs of analyses and storage, especially if long-term storage will be necessary. For example, if a study requires only nanogram quantities of DNA for genotyping purposes, one should consider collecting small amounts of blood or saliva on filter cards, instead of a large volume of blood that will yield hundreds of micrograms of DNA and incur larger processing and storage costs. Other alternate processing and storage approaches that may result in cost savings are considered in the Specimen collection section.

As shown in Figure 3.1, specimen collection, processing and storage are components of a series of steps that are used in any study involving

**Figure 3.1.** The lifecycle of biospecimens in biological resource centres. Used with permission from (2).



the collection of biospecimens. Each of these steps is discussed in turn in the following sections.

## Specimen collection

### *Specimen types*

A wide variety of specimen types may be collected for storage, and in many molecular epidemiology studies more than one of the following (discussed in detail) may be necessary, depending on the study goals (2,3). Additional collection, processing and storage guidance can be found in the International Society for Biological and Environmental Repositories (ISBER), National Cancer Institute (NCI) and International Agency for Research on Cancer (IARC) documents (2–4).

- Blood and blood fractions (plasma, serum, buffy coat, red blood cells)
- Tissue (from surgery, autopsy, transplant)
- Urine
- Saliva/buccal cells

Many other types of specimens may be collected, depending on availability and study goals, for example:

- Placental tissue, meconium, cord blood
- Bone marrow
- Breast milk
- Bronchoalveolar lavage
- Cell lines
- Exhaled air
- Feces
- Fluids from cytology (ascites, pleural fluid, synovial fluid, etc.)
- Hair
- Nail clippings
- Semen

Each of these specimen types should be collected, processed, and stored under conditions that preserve their stability with respect to the intended future analyses. Of particular interest for molecular epidemiology studies are those specimen types that can be collected most conveniently and efficiently, and at the lowest cost for large population-based studies. The most common specimen types collected for these studies are discussed in the following sections: blood, tissue, urine and saliva.

### *Collection procedures*

Collection procedures will vary according to specimen type and the intended analyses, but all procedures should be carefully designed and documented. It is normally a good practice to perform pilot studies to validate new specimen collection methods and protocols (4). The discussion in this section focuses on the specimens most commonly collected for molecular epidemiology studies. Additional information and collection protocols may be found in several references (2–4). Also see Chapter 12, Table 12.2 for additional information about specimen types collected for epidemiologic studies, and their advantages and disadvantages.

#### *Blood collection*

Collection of blood specimens (9) should be carried out by trained phlebotomists to avoid causing study participant discomfort, or compromising the quality or quantity of the sample. Standard protocols recommended by well-established organizations should be used.

An evacuated tube system (e.g. Becton-Dickenson Vacutainer®) with interchangeable glass or plastic tubes is commonly used to collect blood. The tubes, some with additives appropriate to a specific application, are differentiated by their colour-coded stoppers. Blood collection tubes should be drawn in a specific order to avoid cross-contamination of additives (10,11) (also see Chapter 12).

As shown in Table 3.1, blood is often fractionated before being analysed or stored (10,11). Fractionation of blood results in the following components:

- Mononuclear leukocytes (peripheral blood mononuclear cells, PBMCs) are the only cell type in blood that can be maintained in a viable state.
- Neutrophils (the most abundant type of granulocytes) are also nucleated and another source of DNA.
- Erythrocytes can be used to study adducts of haemoglobin.
- Plasma is obtained from an anticoagulated blood sample by separating out the cellular components.

Serum isolation requires no anticoagulants. To reduce contamination, serum should be separated from other blood components as soon as possible. Serum allows for improved analyses of antibodies, nutrients, lipids and lipoproteins. Either serum or plasma may be used for proteomic analyses, although according to recent Human Proteome Organization (HUPO) guidelines there are advantages and disadvantages in the use of either specimen (12). For studies intended to investigate the broadest array of proteins and peptides, plasma is the better choice, as the process of blood coagulation results in the loss of many proteins. Some differences in endogenous hormone analytical results have been found between serum and plasma, but as noted in Chapter 12, both are acceptable as specimens for such analyses.

Depending on the intended laboratory analyses, blood should be collected anticoagulated (consisting of plasma, buffy coat and red blood cells) or coagulated

**Table 3.1.** General guidelines for blood collection and processing

| Blood Fraction | Collection additive | Preferred uses | Limitations/problems |
|---|---|---|---|
| Whole Blood | Anticoagulant (ACD, heparin, EDTA); protease inhibitor for proteomics | Genomics studies; Source of DNA, RNA | Anticoagulant effects need to be considered |
| Buffy Coat | Anticoagulant | DNA extraction; source of lymphocytes, cell lines as unlimited DNA source | Limited yield if blood not properly processed.<br>As a source of DNA, whole blood collection is generally more economical |
| Serum | None | Proteomics; Source of DNA; Multiple analytes | DNA yield low (nanograms) but suitable for genomics applications |
| Plasma | Anticoagulant, possibly protease inhibitor | Proteomics (preferred sample) | DNA yield low (nanograms) but suitable for genomics applications. |
| | | Source of DNA, multiple analytes | Analytical results may differ in serum and plasma. |
| Blood Clot | None | Source of DNA | Extraction difficult, costly |

(consisting of serum and red blood cell clot) (9). There are several types of anticoagulants which need to be chosen carefully to avoid problems with certain laboratory applications (9,11). Other special collection tubes such as Serum Separator Tubes® and Cell Preparation Tubes® (SST, CPT, Becton-Dickenson) allow for more convenient separation of blood fractions, but some problems have been encountered in their use (9). Special collection tubes with protease inhibitors have been developed, which preserve proteins for proteomics analyses (9,12). The analysis of trace metals in blood also requires caution, as they may be present in the evacuated collection tubes. Lot-to-lot variation in the quality of collection tubes is also a potential source of spurious laboratory results.

There is no fixed time period that can be recommended for collecting and processing blood. However, depending on the intended analyses, the stability of blood with respect to various laboratory analyses may be affected or controlled as follows (9,11):

- Anticoagulants used in blood collection, as described above and in Table 3.1.
- Stabilizing agents are necessary to preserve some analytes, and should be included in the collection device or added as soon as possible after collection.
- The time elapsed between blood collection or removal from a storage unit and subsequent processing may be important, depending on the intended analyses. See the United Kingdom Biobank validation study summary for examples of such effects (summarized in reference 7).
- The temperatures at which blood specimens are processed and stored may be important, depending on the intended analyses (13).
- Thaw/refreeze cycles should generally be avoided due to the potential for instability of some analytes. However, thaw/refreeze effects are not well documented for all analytes and may need to be evaluated through pilot tests (13).
- Enzymatic degradation affects many biochemical markers. RNA and proteins are particularly susceptible to this and require special procedures to maintain their integrity during collection and processing. The addition of commercially available RNase inhibitors preserves RNA integrity.
- Special collection systems (ex. PAX DNA® Blood Collection System by PreAnalytiX®) allow for the collection, shipping, and short-term storage of blood at room temperature, and for subsequent extraction of DNA according to a single-tube protocol (14).

### *Tissue collection*

The primary sources of tissues for research are biopsy, surgery and autopsy. As noted in the ISBER Best Practices and IARC Biological Resource Centre Guidelines (2,4), tissues must be collected under strict ethical and legal guidelines, and the collection of samples for research must never compromise the diagnostic integrity of a specimen. Generally it is preferable for a trained pathologist to be involved in the actual procurement of the tissue specimen during a surgical or autopsy procedure.

Other important considerations in collecting tissue are (adapted

from ISBER Best Practices and IARC Biological Resource Centre Guidelines (2,4)):

• Timing. In general, it is important to minimize the time between collection and stabilization and processing of tissue specimens. This time will vary according to the intended use, since different biomolecules degrade at different rates. The effects of collection timing on tissue and macromolecule preservation have not been well studied. The best approach is to collect, stabilize (freezing or fixing) and process tissue specimens as rapidly as possible. It is recommended that surgical or biopsy specimens be preserved within 1 hour (or less if possible) of excision; however, tissue subject to a delay up to two hours should still be collected (15). Detailed records of the timing of events from excision to fixation or freezing should be kept. Tissue banking staff must be present in pathology to freeze or fix the tissue as quickly as possible. Tissues must be snap frozen either directly or enclosed in a container immersed in the freezing medium (e.g. precooled isopentane). Liquid nitrogen is not recommended as a suitable freezing medium for direct snap freezing, due to the potential formation of cryo-artefacts. When dry ice or liquid nitrogen are not readily available, tissue collection into RNAlater® (16) may be a good alternative, provided that the tissue is not required for diagnostic purposes and permission is given by the pathologist.

• Surgical specimens. Remnant samples may be collected from diagnostic procedures or, with proper IRB approval, specimens may be resected specifically for research. Depending on the intended use, specimens may be transported or frozen immediately. Samples requiring snap freezing can be frozen in a Dewar flask of liquid nitrogen or on dry ice at the time of collection. Otherwise, it is recommended that samples be transported in saline on wet ice to the repository or laboratory for additional processing.

• Autopsy specimens. It is important to know the time interval between death and collection and processing of the specimen, as specimens may degrade quickly after death. Autopsy procedures may yield "normal" tissues (i.e. normal lung), or large quantities of a specimen that would not otherwise be available from surgical procedures. Tissue specimens collected at autopsy should be appropriately labelled as to the organ site, tissue type, and time of resection, and then immediately placed in a container of saline on wet ice for transport to the tissue repository for processing.

• Transplant tissue and organs that are inappropriate for transplant may sometimes be made available for research. Often transplant tissue is of a higher quality than either surgical or autopsy specimens, due to the special efforts made to preserve the integrity of the transplant organs.

### *Tissue fixation*

Formalin- or alcohol-fixation and paraffin embedding may be used to preserve tissues at relatively low cost when adequate freezing procedures and storage facilities are not available (2). Formalin-fixation is also the standard practice for preservation of tissues collected during surgery or autopsy. Fixed paraffin blocks may be stored in light- and humidity-controlled facilities at room temperature (18–22°C). Formalin-fixed tissues may be used for DNA extraction. The DNA is usually fragmented but remains suitable for PCR-based analysis of short DNA fragments.

Due to degradation issues, formalin-fixed, paraffin-embedded tissues are of limited use as a source of RNA. However, RNAlater® (16) is a commercial aqueous, non-toxic tissue storage reagent that rapidly permeates tissues to stabilize and protect cellular RNA and eliminates the need to immediately freeze or otherwise stabilize tissue samples. Tissue samples can be harvested and submerged in RNAlater® for storage for specific periods without jeopardizing the quality or quantity of RNA extracted at a later time or date. However, specimens processed in RNAlater® cannot be further used for histomorphopathological analyses.

Alternatives to formalin fixation include ethanol, Optimal Cutting Temperature (OCT) media, methacarn, and Carnoy's solution, among others. To achieve an acceptable balance between the preservation of tissue morphology and nucleic acid integrity, it may be necessary to alter fixation methodology to achieve a study's goals. Several studies have explored the effects of the above standard fixatives, as well as newer ones for special applications (17–20). Although formalin-fixation remains the standard tissue preservation method, these alternatives should be considered for special research applications that require the preservation of particular macromolecules or morphological features.

### *Urine collection (see also Chapter 12)*

Many analytes, such as steroid hormones, pesticides and a wide variety of drugs and their metabolites, can be measured in urine for molecular epidemiology studies (11), making it a convenient

specimen for a variety of studies. Urine collection can performed under several conditions, depending on the study design and analytical goals (4,11):

- First morning. Collected immediately upon rising in the morning, recommended for analytes requiring concentration for detection in laboratory assays.
- Random urine specimens are appropriate for drug monitoring and cytology studies.
- Fractional specimens. The study participant fasts after the last evening meal, and the second morning urine is collected. These specimens are used to compare urine analyte levels with their concentrations in blood.
- Timed urine collections (e.g. 12 and 24 hour) are used to allow comparisons of excretion patterns.

Urine collections should be maintained on ice or refrigerated for the duration of the collection. Collection vessels are generally larger than for other liquid specimens, and may range from 50 to 3000 mL. Depending on the analyte to be measured, a preservative may be needed. The type of preservative may differ according to test methodologies, time delay, and transport conditions. EDTA and sodium metabisulfite are examples of preservatives commonly used in urine collections (11).

### *Saliva/buccal cell collection*

Saliva, with exfoliated buccal cells, is an excellent source of DNA for genetic studies (21). Self-collection of buccal cells is a safe, convenient method that can be used to reduce the cost of specimen collection and is often preferred over blood collection by study participants (discussed in Chapter 12). Several methods have been developed for collecting buccal cells, including swabs, cytobrushes and a mouthwash protocol. The mouthwash protocol has been successfully used in large population-based studies and has been shown to yield DNA of good quality and quantity for genetic analyses (21). However there are limitations to buccal cell DNA, as described below.

New methods are being developed for saliva collection. One such method has been developed by DNAGenotek (22). A proprietary reagent, Oragene, preserves saliva (and DNA) at room temperature. The method has been successfully used in epidemiologic studies (23). The yield and quality of DNA from the Oragene collection is similar to that for the mouthwash method.

### *Collection of blood, saliva on treated cards*

New technologies, such as whole-genome amplification methods to increase genomic DNA yields, and the high cost of collecting and processing blood or mouthwash samples, have led to renewed consideration of treated filter paper cards as a method to collect DNA from blood (24) and buccal swabs (25) (also discussed in Chapter 12). Filter paper cards have been pre-treated to retard bacterial growth, inhibit nuclease activity, and release DNA during processing (26). The cards may be easier to use in paediatric and elderly populations to collect specimens, and can be mailed in an envelope with a desiccant at a nominal cost.

Blood collected on filter cards is well established as a source of DNA for genetic studies, as well as for a variety of other research and clinical applications. The US Centers for Disease Control and Prevention (CDC) uses blood spot cards in its nationwide neonatal screening programme (24). The US Armed Forces collect blood spot cards from all service members and stores them for possible identification purposes, as well as research and clinical purposes. DNA can be easily extracted from blood spots in amounts more than sufficient for genetic studies. This process has been automated, especially for forensic applications (27).

In addition to standard filter cards, new technologies for dry-state specimen collection have been developed. GenVault (28) uses small elements of treated filter paper in 384-well plates for storage of blood, DNA, plasma and serum specimens at room temperature. DNA and protein can be eluted from the elements by relatively straightforward methods.

## ***Preserving specimen stability during collection***

As noted above for tissue biospecimens, the elapsed time for collection, and between collection and stabilization, should be minimized, and the tissue temperature should be reduced as soon as possible after collection. This is especially important if freezing is the stabilization endpoint. If fixation is the stabilization endpoint, control of processing time between maximum and minimum durations may be required. Rapid processing may not be as critical for other types of biospecimens, such as blood. Optimal processing times vary depending on the analysis method for which a biospecimen is used.

Biorepositories should use the processing method that preserves the greatest number of analytes. The best scheme to preserve analytes is to divide specimens into aliquots or fractions of appropriate size or volume and/or preserve them by multiple processing methods.

## Specimen processing

Specimens are processed according to the study design and the methods most appropriate for preserving the analytes of interest. For a particular specimen type and analysis, several processing methods may be appropriate. The IARC standards (2) list some of the more routine processing protocols. The general guidelines in this section outline some of the important considerations when choosing processing methods for specimens most commonly collected for molecular epidemiology studies. Additional issues concerning the processing and analyses of specimens for proteomic, metabolomic, physical, chemical, and immunologic applications are discussed in Chapters 4 and 7.

### *Blood – separation into fractions (e.g. plasma, serum, buffy coat, red blood cells)*

The processing method used for blood specimens depends on the laboratory analyses to be performed. Cryopreservation is a cost-effective way of preserving viable lymphocytes for subsequent recovery of DNA, or for Epstein-Barr Virus (EBV) transformation to create lymphoblastoid cell lines as a source of unlimited amounts of DNA (29). Cryopreservation typically involves the use of a cryoprotectant, such as dimethyl sulfoxide (DMSO). However, commercial cryoprotectants that are less toxic have been developed (30). Whole blood may also be cryopreserved as an efficient and cost-effective approach to centralized processing and storage of viable cells in large-scale epidemiological studies (29).

### *Tissue – processing after surgery, autopsy*

Specimens resected specifically for research may be either processed in the operating room or pathology suite, shortly after the time of collection, or may be transported to the repository for processing, depending upon the requirements of the specific protocol. Additional details are discussed above, and may also be found in the ISBER and IARC Guidelines (2,4).

### *Urine*

Processing of urine before storage is fairly straightforward. The primary decision is the size of the aliquots to be stored and is based on the expected analyses. If the analytes are stable to thaw/refreeze cycles then larger aliquots can be stored.

### *Saliva/buccal cell processing from mouthwash protocol specimens*

Buccal cells collected using the mouthwash protocol (21) are processed by centrifugation of the cell suspension, resuspension in a buffer, and either processed immediately or frozen for future use. Usually, additional processing involves DNA extraction. Note that a special consideration in processing buccal cell DNA is the high percentage of bacterial DNA present in these specimens, which requires special quantitation by real-time PCR.

### *DNA extraction*

DNA extraction methodology is well established for a variety of specimen types, including whole blood, blood fractions, buccal cells, fresh and frozen tissues, and paraffin tissue blocks (31). The gold standard for DNA extraction is generally considered to be phenolchloroform extraction, but other standard methods that are more efficient, less expensive, and that utilize less toxic chemicals provide similar yields and DNA of similar molecular weight. Companies such as Gentra and Qiagen have collected DNA stability data of over 12 years' duration (32).

Techniques for measuring the quality and quantity of DNA range from absorbance at 260nm and 280nm, to fluorescence methods, to real-time PCR for detection of less than 25 picograms DNA. The A260/A280 ratio is a rough measure of DNA purity and protein contamination. Additional methods of measuring DNA quality include gel electrophoresis. The accuracy of DNA quantitation by these methods can vary widely and can affect the quality of downstream genomic analyses. Genomic assays may be very sensitive to the quantity of DNA. A study by the US National Institute of Standards and Technology found a great deal of variability between various methods and among laboratories participating in a DNA quantitation study (33). Great care must be taken to assure that DNA concentration is accurately measured before use in any assay, especially PCR-based genomic applications that require precise quantities of DNA.

RNA is less stable than DNA and is more difficult to extract intact. However, special methods and reagents have been developed that allow for preservation of RNA in blood and other specimens, as noted in the discussion of tissue fixation.

Saliva or blood collected on treated paper cards is available, for example, from Whatman® for laboratory applications. Enough DNA can be obtained from a 2mm punch of a paper card for about 500

single nucleotide polymorphism (SNP) genotypes. The extraction of DNA from blood spot cards can be automated as noted above (27).

Table 3.2 summarizes source material for nucleic acid extraction, and some of the procedural and methodological issues encountered with each specimen type.

### *Aliquoting*

Dividing specimens into smaller sample aliquots is usually necessary to preserve them in volumes useful for routine analyses. The aliquoting protocol should be designed only to store the number of aliquots necessary for the intended analyses, plus additional long-term archival samples that will be available for unforeseen uses. In developing an aliquoting protocol, the consequences of repeated thawing and refreezing cycles should be considered. Although many analytes, such as steroid hormones (discussed in Chapter 12), are stable, other analyses may be affected by one or more thaw-freeze cycles (2,3).

### *Automated systems for specimen processing*

Automated systems have been developed for specimen processing, and several of these systems are useful in processing specimens for molecular epidemiology studies. Generally automation is most applicable to DNA extraction and specimen aliquoting.

For DNA extraction several automated systems are available, depending on the specimen type and volume. For blood specimens, and other blood fractions and suspensions of buccal cells up to 10 mL, the Gentra AutoPure is one of the preferred systems (32). The AutoPure has been validated for use with plasma, serum, buffy coat, buccal cell and other cell suspensions. For smaller samples, in the volume range of 50 uL to 1 mL, the Qiagen EZ-1 and M-48 systems are available (32). Other commercial and custom systems have been developed for specialized automated applications.

The other major biorepository activity that is amenable to automation is aliquoting. DNA in solution, as well as for example serum and plasma, must be stored in volumes suitable for downstream laboratory analyses. If standard collection and storage vessels are used, and a standard aliquoting protocol can be developed, then aliquoting can be automated. An example of a system for automated aliquoting is from TECAN (34).

### Storage

Depending on the intended laboratory analyses, and other considerations, specimens and their aliquots may be stored under

**Table 3.2.** Common DNA sources and extraction issues

| Specimen source | Collection method | Extraction method | DNA yield | Advantages | Challenges |
|---|---|---|---|---|---|
| Whole Blood | Evacuated tube with anticoagulant | Manual or automated | 100s of micrograms | High yield, minimal processing | Refusal to participate |
| Blood -Buffy Coat | Processing of anti-coagulated blood | Manual or automated (with some processing) | 100s of micrograms | High yield, minimal storage volume | Variable yield and quality of buffy coat cellular material |
| Blood - Plasma, Serum | Processing of blood, with or without anticoagulant | Manual or automated | Nanograms | Good use of samples collected for other purposes | Low yield |
| Saliva | Mouthwash, Oragene | Manual or automated | 10-50 micrograms | High compliance rate | Bacterial DNA |
| Blood clot | Evacuated tube, no anticoagulant | Manual (special processing necessary) | Variable | Good use of 'extra' samples | Extractions expensive, |
| DNA fragmented | None | Source of DNA | Extraction difficult, costly | | |
| Tissue – Fresh or Frozen | Surgery, autopsy | Manual | Variable | Most appropriate sample for some studies | DNA fragmented, RNA quality low |
| Paraffin Embedded Tissue | Tissue sections from surgery, autopsy | Manual | Variable | Easily stored | DNA fragmented, RNA quality low |

a variety of conditions as shown in Table 3.3. Most common specimens such as plasma, serum or DNA may be securely stored in mechanical freezers at −80 °C. However, lymphocytes, or other cellular specimens, should be stored in the vapour phase of liquid nitrogen at −150 °C or lower, when long-term cellular viability is necessary. Other storage conditions that are optimal for the preservation of specimen stability should be considered, for example for endogenous hormones, as discussed in Chapter 12. Although generally not necessary in terms of sample and analyte stability, storage in the liquid phase of a liquid nitrogen tank at −196 °C is an excellent option. Although thorough cost analyses have not been performed, it is generally accepted that over the long term, liquid nitrogen freezers are less expensive to maintain than mechanical freezers, due to lower electrical requirements for the equipment and less need to cool the equipment space. In addition, liquid nitrogen freezers are less susceptible to mechanical failure and can withstand power outages for long periods with no temperature deviations.

In situations where freezer systems may not be available, a lower-cost option is collection of saliva or blood spots on filter cards and storage at room temperature. Below are some general storage considerations (1,2,4):

• Adequate back-up storage capacity for low temperature units should be maintained. The power supply must be connected to a back-up generator system that immediately provides power during an electrical outage. Standard operating procedures and techniques for rapidly transferring material to back-up units during such emergencies should be documented.

• Where liquid nitrogen freezers are used, an adequate supply of liquid nitrogen must be maintained. Vapour phase liquid nitrogen storage is preferred over liquid phase storage, where cross-contamination of specimens may occur. Cryovials must be capable of withstanding liquid nitrogen temperatures. Screw cap vials that will not leak are necessary. A good storage container in liquid phase nitrogen is the CryoBio Systems plastic straw (35).

• Alarm systems should be in place to monitor the temperature of mechanical freezers, or in the case of liquid nitrogen freezers, the liquid nitrogen level and temperature.

• Dry ice is frequently used as a refrigerant for shipping and emergency back-up for mechanical freezers.

• A system for maintenance and repair of storage equipment, support systems and facilities should be in place.

• All equipment should be validated before use, or following



**Table 3.3.** General specimen storage guidelines

| Temperature in °C | Preservation method | Recommended for |
|---|---|---|
| +18 to +20 | Room temperature | Slides, tissue blocks |
| 0 to +4 | Refrigerator | Processing fresh specimens |
| −0.5 to −27 | Freezer | Short-term DNA stability |
| −27 to −40 | Freezer | DNA stability |
| −40 to −80 | Freezer | DNA/RNA stability |
| −80 to −130 | Freezer | Recommended for urine, blood, blood fractions (plasma, serum etc) |
| −130 to −150 | Liquid nitrogen vapour | Recommended for storage of tissues, preservation of cellular viability |
| −196 | Liquid nitrogen liquid phase | Storage of living cells |

Adapted from (2).

repairs that affect the instrument's accuracy or other capabilities.

• Labels for storage vessels must be capable of withstanding the required storage conditions, i.e. the label material must not deteriorate and printing must be readable or scanable after long-term storage.

### *Automated freezer systems*

Automated freezer systems are available for convenient storage and retrieval of samples. Commercial automated freezer systems include a custom system built for ARUP Laboratories (36) and systems developed by REMP (37). Generally automated systems are developed for storage at −80 °C, although some liquid nitrogen systems are available.

Automation is most useful for studies and facilities that are focused on one or a few specimen types that will be collected in large numbers and processed and stored in a systematic way. If samples can be stored in microplates (for example, 96- or 384-well), then automated storage and retrieval systems should be considered. However, due to the wide variety of specimen types and processing methods used in molecular epidemiology studies, it is often difficult to justify expensive automated storage and retrieval systems.

### *Storage system maintenance*

Freezers and other storage equipment should be validated and maintained according to the manufacturer's recommendations. In addition, the biorepository should develop additional protocols to assure that equipment functions properly (3,4). A preventive maintenance programme should be in place, with maintenance performed at regularly established intervals.

Special procedures should be developed to assure that freezers are properly validated, in terms of maintaining their optimal temperatures, during initial installation and at regular intervals. As noted in the ISBER best practices: "…any device that provides a readout, data, or has a meter movement, is considered an instrument, and requires calibration." (4).

### *Freezer temperature monitoring*

Freezer temperatures must be continuously monitored to assure proper storage conditions for samples. For mechanical freezers (−20° to −80 °C), temperatures are displayed on each freezer. For small biorepositories, regular (twice daily) manual logging of temperatures may be adequate. However, larger biorepositories should have additional automated systems for remote monitoring of temperatures to efficiently respond to malfunctions (4).

Liquid nitrogen freezers require monitoring of both temperature and liquid nitrogen levels. Temperature monitoring is performed as for mechanical freezers. Liquid nitrogen levels should be recorded manually, on a regular basis, with a stick to assure that normal levels (usually 8–10 cm) are maintained. It is possible for liquid nitrogen freezers to overfill, which is detrimental to samples. Automated systems should be used that can detect and sound alarms for levels of liquid nitrogen that are either too low or too high.

## Information management

Driven by advances in molecular technologies, including genomics and proteomics, information management is critical to the molecular epidemiology research enterprise (38). Collation and analysis of the data associated with the collected specimens that support biomedical research require robust interoperability to allow maximum usage of the collections (3,4). Information management and analysis tools across the spectrum of biomedical research are challenged to provide high performance, scalability and user-friendly interfaces. Also, as data sharing and collaboration between global investigators increases, secure interfaces for data transfer among institutions is paramount.

To manage the vast amounts of data in a variety of formats and environments, robust, flexible and extensible informatics systems are required (38). Too often, initial research plans do not include a well-thought-out approach to handle the results of an investigation. Deliberate planning for data management is far less costly and time consuming compared with ad hoc efforts that occur post-collection. A plan for the various disparate data types and formats should be included with special considerations for multisite collection protocols. A major part of the integrated informatics system for molecular epidemiology is support for biospecimen collection, shipping, processing, storage, inventory and retrieval processes.

### *Specimen tracking*

Today, biospecimen collections are documented and tracked by many forms of data management tools, spanning from laboratory notebooks for a few hundred sample vials to real-time, multiuser software implementations, which support collections with millions of vials. Clearly, there is a need for automated information systems, but the level of

informatics sophistication needed for a collection is limited by the availability of funding. In addition, it is incumbent upon the custodian of human biospecimens to adhere to ethical standards to protect and use the samples (3,4). Documentation of the study protocol number and the informed consent for the study subject should be easily linked back to the biospecimen to guarantee that the specific use of the specimens has been verified before distribution. Information technology software for specimen tracking features secure, validated environments that adhere to ethical practices. As more and more collections are shared among investigators all over the world, information on patient/subject consent, sample collection techniques and processes, and annotation of the sample must be easily retrievable, exportable, and traceable through time.

Biorepository information systems should support inventory functions by tracking all phases of sample acquisition, processing, handling, quality control and distribution from collection site (patient/subject) to utilization (researcher) (3,4). The inventory tracking should include significant events, such as thaws, loss, depletion and destruction of specimens, whether intentional or accidental. Restocking of returned, unused samples from the researcher, if allowed per protocol, must also be documented. Current guidelines for biorepository information systems recommend the use of electronic (linear or two-dimensional) labels or barcodes to document and associate a unique identification number to the samples. No identifying information about the specimen should be encoded as part of the identifier (3,4). The system should also be able to track any pre-existing, external biospecimen identifiers, such as vial type, and notations from hand-written vial labels. Standard operating procedures for the development of identifiers should be maintained with the system and updated to include all labelling paradigms used in the repository.

Bar code scanning technologies have become faster and more accurate in recent years. There are several varieties of software solutions to generate bar codes, from stand-alone programs to those embedded within other applications. Bar code printing options are recommended based on the volume of labels being printed. For high-volume label printing, thermal transfer or direct thermal bar code printers are the instruments of choice (39). When choosing a device, the conditions under which the scanner will be used, the frequency of use, the type of bar code (linear or 2-D), and the distance from which the scanning will be performed should be considered (39). Cost considerations may influence the selection of the bar code scanning technology employed by the biospecimen resource (4).

Biorepository information systems can report available space in the repository and assign and reserve space for incoming specimens. The location of a specimen should be tracked, but should not be used as part of the identifier naming convention, as locations of specimens may change in time.

The user interface of the system must provide tools to search the inventory based on various specimen characteristics, as well as support the requisition of samples to use in research studies. Query interfaces should be easy to navigate by experienced and inexperienced users. Standard and customizable queries are available in all commercially available systems, although ease of use varies. Many of the currently available biorepository inventory systems include web-based access portals to make the systems easier to deploy and navigate.

### *Informatics system security*

The size and scale of the informatics needs of the molecular epidemiology group will determine if the biorepository information system should include the subjects' demographic and study annotation, or whether these data can be held within another database. Robust biorepository management systems provide controlled user access for system security (39). The system should include role-based security for all repository staff, study coordinators and scientists with a need to access the biospecimens inventory. If the study annotation is held within the same data system, security measures should be enacted to protect the subjects' personal health information (PHI) from disclosure to unauthorized users of the data. Regulations governing the protection of individual identifying information vary from country to country, so it is important to reference the guidelines for the specific locations of study and analysis in the study planning process (4).

If the biospecimen inventory is physically separated from the study annotation, these systems should be designed to interoperate and easily link the full study data, to maximize the ability to mine and analyse the data. If the links between systems are unstructured, the result can be an extraordinarily challenging, expensive and time-consuming effort to produce scientific findings from the study.

The system security architecture for information systems can be two- or three-tiered, depending on

the separation of the user interface client (tier one) from the application server (tier two), and then optionally from the data storage (tier three). Three-tiered systems are more flexible and scalable for groups that have large concurrent user needs with heavy data load requirements (39).

### *Inventory control*

"Inventory control starts with an understanding of the conditions under which errors occur and ends with error-resistant processes, intelligent use of technology, a well-trained and highly motivated workforce, and an ongoing process of continuous improvement" (40).

Inventory controls for biorepository management systems include the creation and storage of audit trails to track data history, data verification routines to assure data quality, and process tracking to assure the integrity of the sample data (39). The audit trails will include any changes/additions/deletions of data identifying the user that made the modifications. The system should have the ability to generate configurable reports and data files to provide the most complete information on the specimen. Inventory controls should include complete documentation of the information management system, updated standard operating procedures for the biorepository processes, security measures, and on-going training for those who access the data system (4).

### *Specimen annotation*

The recognized value of molecular epidemiology studies is the collection of appropriate amounts of data, that when combined with the study subject's specimens and laboratory analyses, can be used to study the environmental and genetic causes of disease. It is important to be able to maintain tight integration of the demographic and clinical annotation of biospecimens, whether the data resides within the same data system or in physically distinct systems. Some study collections may include data-use agreements that require specimens to be de-identified before release from the biorepository for analysis. During the study planning process, the rules that govern specimen access are key factors when considering the use of pre-collected biospecimens in a study (41).

The goals of each molecular epidemiology study will determine the specific clinical annotation that should be maintained. Discussions are ongoing across the international biomedical community to provide guidelines for minimal clinical annotation for various study types (2–4,42), to facilitate data pooling of studies across common research areas. The cohort, case–control, and family-based consortia will benefit from the comparison and harmonization of their study data elements and definitions, and this will allow faster mining to detect underlying patterns across their combined data sets.

### *System interoperability*

Epidemiologists are employing newer genomic technologies within studies, which have resulted in exponentially larger data sets. Legacy databases, however, that were functional with smaller data sets and do not communicate with other systems, may need to be replaced or modified. Large data management challenges require the integration of heterogeneous data and tools in a scalable, high-performance system. These systems can manage vast quantities of data, and provide tools for query and analysis in a secure collaborative environment. Efforts to provide interoperability across many institutions and tools based on grid computing are ongoing. Grid technology can be viewed as an extension or application of the internet framework to create a more generic resource-sharing context (43). Cloud computing is a newer delivery model for large, hosted datacentres which offers various computational and data access on an as-needed, "utility company" model over the internet. It typically involves the provision of dynamically scalable and often virtualized resources, thus avoiding the capital expenditure for purchase and maintenance of infrastructure at each bioresource centre location (44).

Whether the study data is housed within one central data system or in a federated, grid or cloud framework, interoperability is essential for the analysis of the data and the publishing of results. Efficient electronic data exchange or sharing between interoperable systems is based on shared common data element (CDE) definitions (45). When combining data from systems that do not share CDEs, mapping of the data to a shared set of elements is required. Often, these mapping efforts are labour-intensive and can result in a loss of information, as local CDEs are fit into exchangeable definitions. It is possible that small differences in the way questions and responses are worded or presented in epidemiology survey instruments can lead to significant (potentially unrecognized) differences in interpretation. The goal of developing CDEs is to enable semantic interoperability—the ability to represent information precisely enough that it may pass between humans and electronic representations precisely without requiring absolute central control

of data systems or external human expertise (38). Semantic interoperability is a key component to speed data pooling efforts across epidemiologic studies to replicate and validate study findings.

### *Informatics at the US National Cancer Institute*

Biomedical informatics systems are evolving as the technology becomes available to "personalize medicine" for each patient. Towards this end, the NCI Center for Bioinformatics has begun the development of the cancer Biomedical Informatics Grid or caBIGTM (45). This is a voluntary network or grid connecting individuals and institutions to enable the sharing of biomedical data and tools, with a goal of creating a World Wide Web of cancer research. The focus is to speed the delivery of innovative approaches for the prevention and treatment of cancer. The infrastructure and tools created by caBIGTM should have broad utility outside the cancer community. An integral part of the caBIGTM plan is the cancer data standards repository (caDSR) that will be used to build and maintain a repository of CDEs for standardization of terms and data storage practices. Tools for many aspects of biomedical research are becoming available on the caGrid.

### *Information management systems from the US National Cancer Institute and Centers for Disease Control and Prevention*

Several organizations and companies around the US and the world are creating solutions to address the information management challenges presented by molecular epidemiology studies. Informatics activities at the National Cancer Institute, Office of Biorepository and Biospecimen Research (NCI, OBBR) have focused on creating recommendations for best practices associated with biorepository data systems, and the minimal clinical data set that should accompany all NCI-funded specimen collections (3). ISBER is focusing on the creation of best practices for biorepository management data systems. This will foster the development of worldwide standardized methods for collection, long-term storage, retrieval and distribution of specimens that will enable their future use (4).

There is a large variety of highly sophisticated, off-the-shelf, open source, and/or custom software applications for biorepository information management (e.g. http://www.isber.org/ims-products.html). Specific needs of the biorepository and the available funding will help guide the selection of the system employed. One highly-focused custom system serves the CASPIRTM (US Centers for Disease Control and Prevention-ATSDR (Agency for Toxic Substances and Disease Registry) Specimen Packaging, Inventory, and Repository) biorepository (46). CASPIR is a central facility to store biological and environmental biospecimens that the CDC-ATSDR began to develop in 1995. The mission of this biorepository is "… to provide storage for valuable, mostly human, biological samples that have been collected from CDC and ATSDR diagnostic studies, epidemiologic outbreaks, and research studies for possible future use." It has a storage capacity of more than six million biospecimens and is managed through customized data management software called the Archival Specimen Tracking and Retrieval Operations (ASTRO™) system.

The custom BioSpecimen Inventory System-II (BSI-II) was initially developed on contract for the NCI's Division of Cancer Epidemiology and Genetics to support their large biospecimen inventory from hundreds of molecular epidemiology studies (39). The BSI-II is flexible, extensible, and is currently storing data associated with more than 10 million specimens in storage across several contract repositories. The NCI's caBIGTM project has developed an open-source, modular caTissue Suite tool set for biospecimen inventory management, tracking, and annotation. This software permits users to enter and retrieve data concerning the collection, storage, quality assurance, and distribution of biospecimens (47).

## Additional issues

Although the issues discussed in the previous sections are critical to the successful collection and preservation of biospecimens, there are other important considerations, concerning the control of specimen quality, as well as the safety and security of personnel and facilities, that are equally important.

### *Quality assurance and quality control*

A Quality Management System (QMS) is an essential element of biospecimen management (3,4). The key to an effective QMS is the development and adherence to Standard Operating Procedures (SOPs). SOPs should guide the collection, processing, storage and equipment maintenance processes described in this chapter. Biorepository staff should be trained to adhere to all relevant quality systems and SOPs. Additional elements that are important for a

QMS include: appropriate security systems, computerized inventory and specimen quality tracking systems, and a facility disaster plan (4).

Several formal quality programs are appropriate for a specimen QMS, including current Good Manufacturing Practices (cGMP) and International Organization for Standardization (ISO) (48) certification. cGMP certification is used in the USA to maintain quality standards that are appropriate for Food and Drug Administration inspection of laboratories and biorepositories that process and store specimens for clinical applications. For research biorepositories, ISO certification, in general, is more appropriate for organizations that will be collaborating with international partners, and wish to assure that they are operating under a common set of recognized international standards. Both cGMP and ISO require extensive documentation of the sources, quality and performance of materials, equipment, and procedures.

### *Safety in the laboratory and biorepository*

Laboratories and biorepositories should assume that all human biospecimens are potentially infective and biohazardous. A predictable, small percentage of biospecimens will pose a risk to the biorepository workers who process them. All biospecimens should be treated as biohazards (49). In addition to taking biosafety precautions, biorepositories should adhere to key principles of general laboratory safety.

In the United States, the Occupational Safety and Health Administration (OSHA) regulations (50) require that appropriate vaccinations be offered to all personnel who may be potentially exposed to human blood, body fluids and tissues, or other potentially infectious materials. Biorepository work practices should be based on universal precautions similar to those used in laboratories and clinical settings. Good general laboratory work practices are outlined by Grizzle and Fredenburgh (49). The CDC/NIH booklet *Biosafety in Microbiological and Biomedical Laboratories* (51) outlines general biosafety guidelines. All biorepositories that handle human biospecimens should operate under the OSHA (or similar) blood-borne pathogens standards and develop an exposure control plan.

In addition to biosafety, biorepositories should follow strict general safety regulations and procedures regarding chemical, electrical, fire, physical and radiological safety (3,4,50).

The use of liquid nitrogen poses unique safety problems that are not usually noted in laboratory safety documentation. With a liquid temperature of −196 °C, flesh freezes almost instantly if it comes in direct contact with the liquid. Both face and eye protections are required. Oxygen level sensors should always be employed, since oxygen deprivation is a serious hazard in the event of a liquid nitrogen leak.

### *Proper packaging and shipping*

Depending on whether they are known to contain infectious agents, and the intended analyses, specimen shipments may be regulated as infectious substances or as diagnostic specimens. To properly classify the specimens to be included in a shipment, consult references provided in the ISBER Best Practices (4) and by the International Air Transport Association (52).

Specimens are often exposed to temperature fluctuations during transit. The required shipping temperature depends on the intended analyses (3,4). Packaging materials and equipment are available to preserve specimens under ambient, refrigerated and frozen conditions, including liquid nitrogen dry shippers that can preserve specimens frozen at or below –150 °C for up to several weeks (3,4). Devices are available to monitor temperature trends during shipment, either by recording temperatures precisely at certain time intervals, or by changing colour if a certain temperature is exceeded during shipment.

### *Security systems for biospecimen facilities*

Due to the irreplaceable nature of many specimens collected for molecular epidemiology studies, it is critical to protect them from destruction due to electrical outages, equipment failures, and similar problems. The most important systems to have in place are electrical back-up generators and equipment alarms (4).

Generators should be available to provide electrical service to all freezers and any other critical equipment immediately upon the loss of general electric service to the facility. They should be maintained in good working order and started on a regular basis to assure that they are functioning properly (4). The appropriate fuel should be in adequate supply for up to three days of electrical outage during an emergency situation.

Alarm systems should be provided in specimen storage areas to alert the staff when a freezer or other equipment is malfunctioning.

They should be designed to automatically (for example, by cell phone or paging device) notify biorepository staff and other appropriate facilities maintenance personnel during non-working hours. Procedures should be in place to immediately respond to such equipment emergencies, and to either move the specimens to a functioning back-up freezer, or take other appropriate action to preserve their integrity.

In general, these measures should be part of a broader disaster response plan that is designed to protect personnel as well as specimens (4).

## Future directions and challenges

### *Specimen management under adverse or low-resource conditions*

In general, the methods, equipment and supplies described in this chapter are practices that should be adopted under the conditions found in developed countries. However, it is not always possible in some developing countries with fewer resources to have access to liquid nitrogen or mechanical ultra-low freezers, for example, or even electricity in some situations. These special circumstances need to be carefully considered before specimen collection is initiated. Some of the materials described in other sections of this chapter may be useful. For example, if extreme temperatures with little or no local refrigeration is an issue, then blood or saliva can be collected on filter cards and shipped and stored at ambient temperature. Blood can also be collected and shipped at ambient temperature using the PaxGene® collection tubes. Tissues can be fixed in formalin and embedded in paraffin blocks for low-cost storage and transport. If possible, given local conditions, "cool packs" and other supplies can be provided from a central coordinating centre and used to transport specimens at refrigerated temperatures. Note that any such procedures that deviate from documented best practices must be validated in a preliminary pilot study before full-scale adoption.

A specific example of working under such conditions is the Costa Rica HPV Vaccine Trial conducted by the US NCI in collaboration with the Fundacion Inciensa (53). Given the conditions under which specimens had to be collected in Costa Rica, the following factors were considered and accounted for:

- Bad road conditions increase shipment time and specimen shaking. Road conditions change from the dry to rainy season every year, and affect access to some communities.
- Liquid nitrogen may be hard to find in some countries, but not impossible. For example, Nicaragua does not produce any gases, but has hospitals and factories that require oxygen and liquid nitrogen, so oxygen is imported from Costa Rica.
- The cost of liquid nitrogen, equipment and reagents are generally higher in Central America than in developed countries, and in some cases, dealers for a particular country are regional. For example, a particular product produced in the USA may have to be acquired from a Mexican dealer that represents that product for Mexico and Central America.
- In some countries the power supply may be regulated and/or in poor condition. If possible, a back-up power supply should be provided or alternate storage methods should be considered.
- High temperature and humidity during the day are common conditions that may require special shipping containers, such as coolers with cold packs.
- Permits for importation and exportation of human-derived substances and repository operation permits must be obtained before starting operations. Policies and procedures will vary according to the country of origin and the destination.
- Laboratory equipment and reagents may have to be imported, which will require a variable time for customs and regulatory issues or the delivery time policy of the local or international dealer. Because of this, inventory management must be highly coordinated to account for potential delays.

### *Alternate collection technologies*

In addition to dry-state collection and storage on treated cards, other special collection and storage systems have been developed that are beginning to be used in population-based studies. These approaches, mentioned briefly in other sections of this chapter, may gain more widespread use in studies that require the collection of large numbers of specimens that will need purified DNA as the analytical derivative. Some examples are:

- Oragene, developed by DNAGenotek (22). Oragene is a reagent used for saliva collection. The reagent saliva mixture is stable at room temperature. DNA can be readily extracted either by using the company's manual procedure or an automated procedure, such as the Gentra AutoPure. At least one large epidemiology study, performed by the Karolinska Institute, has had success with this protocol (23).

• GenVault (28) has developed a small cellulose element, based on the Whatman treated card, that can be used in a 384-well microplate format to store DNA and other samples in the dry-state. DNA can be eluted from the elements using a simple protocol, and adequate amounts (up to 200 nanograms) of DNA can be extracted from each element, making this a convenient system for long-term economical archiving of DNA.

## *Biospecimen ethical, legal and policy issues*

The ethical, legal and policy aspects of biospecimen collection are as complex, if not more so, than the technical matters outlined in this chapter. The following are some of the issues that have not been fully resolved in the international community:

• *Informed consent.* Formats and details vary greatly among institutions. Policies for handling of biospecimens after withdrawal of consent are not well defined.

• *Ownership*. It is often unclear who 'owns' biospecimens once they have been donated for research. Court cases in the USA have ruled that the study participant does not have any ownership rights after donating a specimen for research. The NCI Best Practices (3) uses the term "custodianship" to reflect the need for a biospecimen resource to develop a plan for long-term care of biospecimens.

• *Specimen and data access*. Biospecimen resources should have clear rules for outside access to specimens and collected data (3).

• *Privacy protection*. Study participants need to be assured that their identity will be protected, with respect to use of specimens they have donated and any resulting data. Privacy regulations are in place for this purpose (3). Due to advances in genomic technologies, it is becoming increasingly difficult to guarantee the protection of an individual's identity.

• *Intellectual property*. Inventions and data arising from research using annotated biospecimens may have commercial value. Institutions should have clear intellectual property guidelines, and use material transfer agreements to assure that the sharing of specimens and data are well controlled. The final disposition of specimens and data should be understood before initiating a transfer.

In summary, the issues surrounding the use of biospecimens in research are complex and must be approached with attention to the many technical factors that may affect the quality of the specimens. In addition, it is important to recognize that the quality of biospecimens is enhanced by the collection and proper control of various types of data. Finally, many issues discussed in this chapter are subject to strict local and national policies and regulations concerning privacy and informed consent.

# References


1. OECD Best Practice Guidelines for Biological Resource Centres. Available from URL: http://www.oecd.org/dataoecd/7/13/38777417.pdf.

2. World Health Organization, International Agency for Research on Cancer. Common minimal technical standards and protocols for biological resource centers dedicated to cancer research. Available from URL: http://www.iarc.fr/en/publications/pdfs-online/wrk/wrk2/Standards_ProtocolsBRC.pdf.

3. National Cancer Institute, Office of Biorepositories and Biospecimen Research. NCI best practices for biospecimen resources. Available from URL: http://biospecimens.cancer.gov/bestpractices.

4. Campbell JD, Skubitz APN, Somiari SB *et al.* (2008). International Society for Biological and Environmental Repositories (ISBER). 2008 Best practices for repositories: collection, storage, retrieval and distribution of biological materials for research. *Cell Preserv Technol*, 6:3–58.

5. Manolio TA (2008). Biorepositories–at the bleeding edge. *Int J Epidemiol,* 37:231–233. doi:10.1093/ije/dym282 PMID:18381397

6. Elliott P, Peakman TC; UK Biobank (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.doi:10.1093/ije/dym276 PMID:18381398

7. Peakman TC, Elliott P (2008). The UK Biobank sample handling and storage validation studies. *Int J Epidemiol,* 37 Suppl 1;i2–i6.doi:10.1093/ije/dyn019 PMID:18381389

8. Baird PM, Frome RJ (2005). Large-scale repository design. *Cell Preserv Technol,* 3:256–266 doi:10.1089/cpt.2005.3.256.

9. Vaught JB (2006). Blood collection, shipment, processing, and storage. *Cancer Epidemiol Biomarkers Prev,* 15:1582–1584. doi:10.1158/1055-9965.EPI-06-0630 PMID:16985016

10. BD Diagnostics. Available from URL: http://www.bd.com/vacutainer/pdfs/plus_plastic_tubes_wallchart_orderofdraw_VS5729.pdf.

11. Landi MT, Caporaso NE. Sample collection, processing, and storage. In: Toniolo P, Boffeta P, Shuker DEG *et al.*, editors. Applications of biomarkers in cancer epidemiology. Lyon: IARC Scientific Publication; 1997. p. 223–236.

12. Rai AJ, Gelfand CA, Haywood BC *et al.* (2005). HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics*, 5:3262–3277. doi:10.1002/pmic.200401245 PMID:16052621

13. Jackson C, Best N, Elliott P (2008). UK Biobank Pilot Study: stability of haematological and clinical chemistry analytes. *Int J Epidemiol,* 37 Suppl 1;i16–i22. doi:10.1093/ije/dym280 PMID:18381388

14. PreAnalytix Blood DNA system. Available from URL: http://www.preanalytix.com/DNA.asp.

15. Eiseman E, Bloom G, Brower J *et al.* Case studies of existing human tissue repositories. Santa Monica (CA): RAND Science and Technology; 2003.

16. Applied Biosystems RNAlater tissue collection: RNA stabilization solution. Available from URL: https://products.appliedbiosystems.com/ab/en/US/adirect/ab?cmd=catNavigate2&catID=603386.

17. Stanta G, Mucelli SP, Petrera F *et al.* (2006). A novel fixative improves opportunities of nucleic acids and proteomic analysis in human archive's tissues. *Diagn Mol Pathol,* 15:115–123.doi:10.1097/00019606-200606000-00009 PMID:16778593

18. Vincek V, Nassiri M, Nadji M, Morales AR (2003). A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. *Lab Invest,* 83:1427–1435.doi:10.1097/01.LAB.0000090154.55436.D1 PMID:14563944

19. Cox ML, Schray CL, Luster CN *et al.* (2006). Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. *Exp Mol Pathol,* 80:183–191.doi:10.1016/j.yexmp.2005.10.002 PMID:16332367

20. Olert J, Wiedorn KH, Goldmann T *et al.* (2001). HOPE fixation: a novel fixing method and paraffin-embedding technique for human soft tissues. *Pathol Res Pract,* 197:823–826. doi:10.1078/0344-0338-00166 PMID:11795830

21. García-Closas M, Egan KM, Abruzzo J *et al.* (2001). Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev,* 10:687–696. PMID:11401920

22. DNAGenotek. Oragene DNA stabilization system. Available from URL: http://www.dnagenotek.com.

23. Rylander-Rudqvist T, Håkansson N, Tybring G, Wolk A (2006). Quality and quantity of saliva DNA obtained from the self-administrated oragene method–a pilot study on the cohort of Swedish men. *Cancer Epidemiol Biomarkers Prev,* 15:1742–1745. doi:10.1158/1055-9965.EPI-05-0706 PMID:16985039

24. Mei JV, Alexander JR, Adam BW, Hannon WH (2001). Use of filter paper for the collection and analysis of human whole blood specimens. *J Nutr,* 131 Suppl;1631S–1636S. PMID:11340130

25. Sigurdson AJ, Ha M, Cosentino M *et al.* (2006). Long-term storage and recovery of buccal cell DNA from treated cards. *Cancer Epidemiol Biomarkers Prev,* 15:385–388. doi:10.1158/1055-9965.EPI-05-0662 PMID:16492933

26. Whatman. Filter paper DNA isolation. Available from URL: http://www.whatman.com/NucleicAcidandProteinSamplePreparation.aspx.

27. Tack LC, Thomas M, Reich K *et al.* (2005). Automated forensic DNA purification optimized for FTA card punches and identifier STR-based PCR analysis. *J Assoc Lab Autom,* 10:231–236 doi:10.1016/j.jala.2005.04.004.

28. GenVault. DNA isolation. Available from URL: http://www.genvault.com.

29. Hayes RB, Smith CO, Huang WY *et al.* (2002). Whole blood cryopreservation in epidemiological studies. *Cancer Epidemiol Biomarkers Prev,* 11:1496–1498. PMID:12433734

30. Biolife Solutions. Available from URL: http://www.biolifesolutions.com/.

31. Santella RM (2006). Approaches to DNA/RNA Extraction and whole genome amplification. *Cancer Epidemiol Biomarkers Prev,* 15:1585–1587.doi:10.1158/1055-9965.EPI-06-0631 PMID:16985017

32. Qiagen. DNA extraction. Available from URL: http://www1.qiagen.com/Products/DNA.aspx.

33. Kline MC, Duewer DL, Redman JW, Butler JM (2005). Results from the NIST 2004 DNA quantitation study. *J Forensic Sci,* 50:570–578. doi:10.1520/JFS2004357 PMID:15932088

34. TECAN. Liquid handling and robotics. Available from URL: http://www.tecan.com/page/content/index.asp?MenuID=1&ID=2&Menu=1 &Item=21.1.

35. Cryo Biol System. Available from URL: http://www.cryobiosystem-imv.com/.

36. ARUP. Automated storage and retrieval system. Available from URL: http://www.aruplab.com/LaboratoryExpertise/AutomationInitiative/as_rs.jsp.

37. REMP. Automated freezer systems. Available from URL: http://www.remp.com/.

38. Henderson MK, Mohla C, Jacobs KB, Vaught J (2005). Challenges of scientific data management for large epidemiologic studies. *Cell Preserv Technol,* 3:49–53 doi:10.1089/cpt.2005.3.49.

39. Biological Specimen Inventory System BSI-II. Available from URL: http://bsi-ii.com/.

40. Piasecki DJ. Inventory accuracy: people, processes and technology. Pleasant Prairie (WI): Ops Publishing; 2003.

41. Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). Available from URL: http://grants.nih.gov/grants/guide/notice-files/not-od-07-088.html.

42. Lopez AD, Mathers CD, Ezzati M *et al.* (editors). Global burden of disease and risk factors. Oxford University Press and the World Bank; 2006.

43. Grid Computing. Definition from English Wikipedia. Available from URL: http://en.wikipedia.org/wiki/Open_Grid_Forum.

44. Bernstein D, Ludvigson E, Sankar K *et al.* Blueprint for the intercloud: protocols and formats for cloud computing interoperability. Fourth International Conference on Internet and Web Applications and Services. *IEEE Computer Soc* 2009:328–336.

45. Information about the NCI Cancer Bioinformatics Grid (caBIG). Available from URL: https://cabig.nci.nih.gov/.

46. Gunter EW (1997). Biological and environmental specimen banking at the Centers for Disease Control and Prevention. *Chemosphere,* 34:1945–1953.doi:10.1016/S0045-6535(97)00056-8 PMID:9159897

47. NCI. caBIG's caTissue Suite tools. Available from URL: https://cabig.nci.nih.gov/tools/catissuesuite.

48. International Organization for Standardization. Available from URL: http://www.iso.org/iso/home.htm.

49. Grizzle WE, Fredenburgh J (2001). Avoiding biohazards in medical, veterinary and research laboratories. *Biotech Histochem,* 76:183–206. PMID:11549131

50. Occupational Safety and Health Administration. Hazardous and toxic substances. Available from URL: http://www.osha.gov/SLTC/hazardoustoxicsubstances/standards.html.

51. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. National Institutes of Health. Biosafety in microbiological and biomedical laboratories. 4th ed. Washington (DC): U.S. Government Printing Office; 1999. Available from URL: http://www.cdc.gov/od/ohs/biosfty/bmbl4/bmbl4toc.htm.

52. International Air Transport Association. Infectious substances shipping guidelines. 7th ed. Montreal, Canada: International Air Transport Association; 2010. Available from URL: http://iatabooks.com/.

53. Cortés B, Schiffman M, Herrero R *et al.* (2010). Establishment and operation of a biorepository for molecular epidemiologic studies in Costa Rica. *Cancer Epidemiol Biomarkers Prev,* 19:916–922 doi:10.1158/1055-9965.EPI-10-0066 PMID: 20332271

54. Vaught JB, Caboux E, Hainaut P (2010). International efforts to develop biospecimen best practices. *Cancer Epidemiol Biomarkers Prev,* 19:912–915.doi:10.1158/1055-9965.EPI-10-0058 PMID:20233852