

TECHNOLOGY FEATURE

BUILDING BETTER BIOBANKS

High-quality, data-rich samples are essential for future research. But obtaining and storing these samples is not as straightforward as many researchers think.

L. RUBENSTEIN/BROAD INST.



Larger biobanks have invested in automated storage and retrieval systems to track samples and ensure that they are maintained at a constant temperature.

BY MONYA BAKER

Across the world, freezers and cabinet shelves are full of human samples. Biobanks — collections of biological material set aside for research — vary tremendously in size, scope and focus. Samples can be collected from the general population, from patients who have had surgery or a biopsy and from people who have recently died. Some collections date back decades. The Aboriginal genome, for instance, was sequenced from a lock of hair originally given to British ethnologist Alfred Cort Haddon in the 1920s; he crisscrossed the world gathering samples that are now housed at the University of Cambridge,

UK. Most collections contain dried or frozen blood, but tissues such as eye, brain and nail are also held. Some biobanks address different questions from others: a population-based biobank that collects dried blood and health data may be used to determine the genetic risk factors for breast cancer, whereas a disease biobank that collects tumour samples might be used to reveal different molecular forms of breast cancer.

The number of tissue samples in US banks alone was estimated at more than 300 million at the turn of the century and is increasing by 20 million a year, according to a report¹ from the research organization RAND Corporation in Santa Monica, California. Those numbers are

probably an underestimate, says Allison Hubel, director of the Biopreservation Core Resource at the University of Minnesota in Minneapolis.

But many scientists still say they cannot obtain enough samples. A 2011 survey² of more than 700 cancer researchers found that 47% had trouble finding samples of sufficient quality. Because of this, 81% have reported limiting the scope of their work, and 60% said they question the findings of their studies.

Whereas researchers would once have inspected biological specimens under a microscope or measured only a handful of chemical constituents, or analytes, now they want to profile hundreds of molecules, including DNA, RNA, proteins and metabolites. The ►

INDIVIDUAL

► popularity of genome-wide association studies, in which researchers scan the genome to look for genetic markers, has trained scientists to go on statistical hunts that require both more quantitative measurements and greater numbers of samples. “The manner in which biomedical researchers use biospecimens has changed substantially over the past 20 years,” says Stephen Hewitt, a clinical investigator at the National Cancer Institute in Bethesda, Maryland, and an expert on sample quality. “Our knowledge of the factors that impact a biospecimen has not kept up, nor has the education of the users about how fragile a biospecimen is.”

IN THE COLD

“If you don’t treat the sample properly, it can limit what you can do,” says Kristin Ardlie, director of the Biological Samples Platform at the Broad Institute in Cambridge, Massachusetts. She recalls a project to isolate RNA from placenta samples, which are full of RNA-degrading enzymes. After several tries and no success, a bit of detective work revealed that a collaborator had put the samples in a -20°C freezer to begin with, and only moved them to a -80°C freezer several hours later.

“Researchers think ‘freezing is freezing,’” says Ardlie, but a typical freezer is not cold enough to stop degradative enzymes. Except for DNA, few biomolecules are preserved well at -20°C . Most samples can be stored at -80°C , but certain specimens, such as live cells, need to be kept at temperatures close to -200°C , at which point enzymes are thought not to be able to function at all³.

Worse than having nothing to analyse are analytes that change in unpredictable ways.

One study⁴ showed that the concentration of two cancer biomarkers seemed to increase by around 15% from the time that the serum samples were collected and frozen to when they were thawed and measured again about 10 years later. In another experiment⁵, designed to simulate long-term freezing, researchers examined how several cancer biomarkers

changed in serum samples that were frozen and then thawed. Some protein biomarkers seemed to be stable for decades even with multiple freeze-thaw cycles. However, vascular endothelial growth factor — an extensively studied biomarker implicated in diabetes, arthritis and cancer — was so unstable that the authors recommended that it should never be measured in samples that have been frozen.



Reports have probably underestimated the true number of samples in US biobanks.

Allison Hubel



Liquid-nitrogen freezing stops samples degrading.

Not all biobanks document whether a sample has been thawed for analysis and then restocked, nor do they monitor freezer temperatures, says Daniel Simeon-Dubach, a biobanking consultant based in Switzerland. Even short-term fluctuations in temperature can allow sample-damaging ice crystals to form, but Simeon-Dubach says he has seen researchers hold freezer doors open for minutes at a time to show off their specimens. “I think ‘what are you doing? Show me a picture!’”

ON THE SHELF

Many of the larger biobanks are buying sophisticated freezers to maintain a constant temperature. Systems from companies such as Hamilton Storage Technologies in Hopkinton, Massachusetts, start at around US\$1 million and can hold between 250,000 and 10 million samples. Rather than opening freezer doors, researchers place sample tubes in a hatch, and a mechanical arm then moves them to interior shelves. Researchers can even use laboratory information-management systems to search for appropriate samples for a particular study, such as those from donors of a particular age or weight, and then transmit their request to be retrieved. The samples are deposited in a delivery hatch and an e-mail is sent when they are ready to be picked up. The -80°C freezer also records how many times each sample is removed from frozen storage and for how long. Other companies, such as Brooks Automation in Chelmsford, Massachusetts, also sell automated freezers, and Freezerworks in Mountlake Terrace, Washington, sells data-management software for storing and tracking samples.

Even without such sophisticated equipment, freezer storage can be expensive. A typical

epidemiological study might have 100,000 samples from 10,000 patients that would fill five freezers, each of which costs \$6,000 a year to maintain properly, says Jim Vaught, deputy director of the National Cancer Institute’s Office of Biorepositories and Biospecimen Research (OBBR). And although freezing is considered the best way to preserve biomolecules and live cells, it can distort the appearance of tissues.

To cut storage costs, most researchers study morphology by relying on a preservation technique that harks back more than a century. Tissue taken from a patient is soaked in the preservative formalin, and pieces of the ‘fixed’ tissue are then embedded in blocks of paraffin. The Joint Pathology Center in Silver Spring, Maryland, has some 28 million of these blocks, dating back to the First World War. The blocks allow a thin slice of tissue to be taken and stained for microscope slides, but biomolecules are not preserved as effectively. “The tissue actually drowns in fixative,” explains Hewitt. Hypoxia in the dying cells degrades RNA and alters proteins; formalin crosslinks protein and DNA into complexes, and causes nicks in RNA and DNA. When researchers go to recover biomolecules, removing the paraffin can cause more damage.

Although DNA and RNA have been extracted from paraffin-embedded samples, the quality varies and analysis is difficult. Mike Hogan is vice-president of research at IntegenX in Pleasanton, California, which sells products for storing DNA and RNA at room temperature. He believes that formalin fixation can be modified to preserve biomolecules. The main causes of biomolecular degradation are not due to the formalin directly but to hydrolysis and oxidation, he says. Freezing works because the chemical reactions driving degradation occur more slowly at low temperatures. Scientists at IntegenX and at the University of North Carolina, Chapel Hill, are working on techniques to slow hydrolysis and oxidation by removing water and reactive oxygen-containing molecules. If this technique works, it would allow researchers to study biomolecules but still maintain the morphology standards and staining protocols developed over decades of formalin fixation.

Other approaches focus on removing formalin from the process. In 2009, Qiagen, based in Hilden, Germany, released a product called PAXgene Tissue, which uses a proprietary, alcohol-based fixative to preserve biomolecules and to allow tissue specimens to be embedded in paraffin. The tissue can be stored for up to seven days at room temperature, four weeks at 4°C and months at -20°C without compromising morphology or biomolecules, says Daniel Groelz, a senior scientist at Qiagen. Researchers there are working on ways in which pathologists can analyse more kinds of PAXgene-preserved histological samples. “There is a huge amount of specialized staining techniques that have been optimized for

N.G. JOHNSON



Researchers are investigating the best way to analyse biomolecules in samples stored in paraffin blocks.

formalin,” says Groelz. The company is working on adapting protocols for PAXgene, particularly antibody-based protocols to stain proteins, he explains.

This preservative method is starting to be used in place of deep freezing. A pilot project for one of the most ambitious tissue-collection studies, which aims to correlate gene expression and common genetic variation within dozens of tissue types, examined more than 20 tissue types preserved using four different methods. The Genotype-Tissue Expression (GTEx) programme, a collaborative effort involving several groups at the US National Institutes of Health, as well as academic institutions ultimately chose PAXgene. Jeffrey Struewing, a programme director for the US National Human Genome Research Institute in Bethesda, Maryland — who works on the project — explains that not only did the technique preserve RNA, the logistics necessary to ship ultracold samples could have hampered collection. Struewing says that it is too early to know how PAXgene will work over many years or for biomolecules such as proteins. “There is no preservation method that is going to work for every analyte in every sample.”

QUALITY COLLECTION

Some of the most intractable difficulties occur before preservation begins, says Carolyn Compton, the first director of the OBBR and chief executive of the Critical Path Institute in Tucson, Arizona — a not-for-profit organization for improving drug development. “Biospecimens are parts of people’s bodies that get removed from their setting. They are undergoing biological stresses they would never experience in your body.” When cut off from a blood supply

and exposed to abrupt changes in temperature, the cells’ behaviour becomes hard to predict. Gene expression and protein phosphorylation fluctuate wildly and cellular self-destruct pathways may be activated. Researchers must ask themselves whether analyses of samples reflect the biology of the patients they come from, says Compton. “You can have an absolutely perfect test but still get the wrong answer.”

Even if tissue is preserved well, it may not tell the full biological story. “The problem is not just the post-mortem interval,” says Hewitt. “What’s difficult is the pre-mortem lack of vitality.” Tissues collected from patients who have been on ventilators may not resemble those from healthy patients. “If you took a biopsy of muscles in my arm after I went rowing in the morning, my RNA profile will look a lot different than if I’ve been sleeping for a while,” Hewitt says. “You’ve got to interpret your data within the limit of what they can tell you.”

Blood, urine and saliva samples from non-hospitalized volunteers can be collected during scheduled appointments. But solid-tissue samples are usually collected in hospitals as part of more urgent procedures. Medication, the anaesthesia regime and how blood is shunted from the tissue being removed all affect the sample. So does the length of time the sample stays at room temperature before it is frozen, the time and type of fixative, the rate at which it is frozen and the size and shape of the aliquots.

Medical staff will always be focused on the patient on the operating table, but a greater

awareness of the impact that samples can have on medical research and patients’ diagnoses is having an effect. At a conference organized by the OBBR this February, Gene Herbek, a pathologist at the Nebraska Methodist Hospital in Omaha, described working with surgical teams so that tissues reached pathologists’ laboratories within an hour of excision.

Biotechnology company Indivumed in Hamburg, Germany, collects samples within ten minutes of excision by having designated nurses on surgical teams. These nurses prepare for surgery along with the rest of the team, receiving information about a patient’s treatment and condition. Once the tissue is removed, it is taken into a room next to the surgical suite, where it is sectioned into pieces that are then fixed and frozen. “The solution is not technology; it is process,” says Helge Bastian, managing director of the company.

“The rule of thumb for how long you have from taking a sample to starting to process it is 15 minutes,” says Simeon-Dubach, adding that this is a very ambitious goal. Specifics vary by organ — gastrointestinal organs such as the stomach should be processed much faster.

Speed is also important for tissues collected post-mortem. Staff collecting tissues for GTEx are expected to be ready around the clock so that they can begin work as soon as the team collecting donated organs has finished. More than half of the tissue samples in the programme have been collected and fixed within six hours of death, says Struewing.

Tissues are shipped to the Broad Institute, also part of GTEx, for gene-expression profiling. Before beginning the profiling, Ardlie’s team isolates the RNA and checks it for quantity and quality using a metric called the RNA integrity number; it is an imperfect measure, but excluding samples with low integrity numbers maintains consistency.

ASSESSING QUALITY

Researchers need better biomarkers of sample quality both to prevent expensive experiments on inappropriate material and to reduce artefacts, says Hewitt, who is working with Ardlie to find measures of RNA quality that work in paraffin-embedded tissue. Scott Jewell, director of the programme for biospecimen science at the Van Andel Research Institute in Grand Rapids, Michigan, is evaluating markers of oxygen deprivation and various types of cell death (autophagy, apoptosis and necrosis). Documenting how samples are collected and maintained is important, but may be insufficient, he says. Researchers need specific recommendations. “We want markers that can say, ‘This is a bad sample. This is a good sample.’” Such a process is important not only for choosing samples to include in a particular study, but also for understanding how best to preserve them.

Many biobanking experts find that researchers give little thought to sample quality. An

analysis of 125 biomarker discovery papers published in open-access journals between 2004 and 2009 found that more than half included no information about how specimens had been obtained, stored or processed⁶. Perhaps this is not surprising; biobanking practices have come under scrutiny only recently. The OBBR, established in 2005, released its first official set of best-practice guidelines in 2007, and last year released *Biospecimen Reporting for Improved Study Quality* to guide researchers on documenting how biospecimens are collected, processed and stored.

The International Society of Biological and Environmental Repositories in Bethesda, Maryland, published its first edition of best practice in 2005, and a coding system — Standard PREanalytical Code — for describing what tissue had been collected and how in 2010. The European Union has funded a four-year programme called Standardisation and Improvement of Generic Pre-analytical Tools and Procedures for *In Vitro* Diagnostics, a multi-institution project coordinated by Qiagen with the aim of improving and standardizing sample handling for *in vitro* diagnostics. In

addition, societies are advocating that journals request this information for peer-reviewed articles.

The College of American Pathologists, a professional society based in Northfield, Illinois, has developed an accreditation programme for biorepositories. It began accepting applications this year and has so far had a good response. Facilities receive a checklist of required practices, such as whether they have tracked if samples have been thawed and refrozen, and whether they have installed freezer alarms. If facilities meet the list of requirements, they

A kingdom's worth of samples and data

In the past few years, more than a half a million people in the United Kingdom have collectively peed into cups, spat into tubes and had needles stuck in their arms. They have spent hours having their weight, blood pressure, memory, lung volume and grip strength measured and recorded, and answering extensive questionnaires about their lifestyles.

Deciding what samples and information to collect was several years in the planning, says Paul Downey, director of operations, who joined the UK Biobank project in 2004, years before the first participants were enrolled. "Tracking 14 million small tubes of biological fluid from the point of creation through to storage for, potentially, 30 years is an ask," he says. Everything had to be completed within the bank's budget and in a way that was acceptable to volunteers. Initially, for example, some researchers hoped for fat biopsies, but that would have discouraged volunteers and increased the cost and complexity of collection.

The decision was made to collect blood, urine and saliva. Samples were collected using seven kinds of blood tubes, containing different mixtures of anticoagulants and preservatives to support anticipated assays. The team decided not to include additives in the saliva and urine samples because this might introduce variability or affect future analyses.

On a typical day, the project processed samples from between 700 and 1,000 volunteers. This took four or five laboratory personnel and more than a dozen robots, which together could do the work of about 50 technicians, says Downey. "It does look like a science-fiction set." One six-armed robot centrifuges blood samples, images the layers that form and then pipettes each layer into its own tube,

which is bar-coded and prepared for storage. The robots are still busy processing samples for disease-specific biobanks and follow-up specimens from participants, says Downey.

Each donor's samples are split between two facilities, which serve as a back-up for

to a delivery port, minimizing the exposure of other samples to warmer temperatures. Nineteen kilometres away, a manual back-up system cooled by liquid nitrogen holds 6 million samples at -196°C — the temperature recommended for storing live cells. This facility contains 50 round, stainless-steel tanks, each the size of a small car. Eventually, says Downey, the set-up will let researchers compare how various biomolecules survive under the two conditions.

Steps have also been taken to preserve samples during analysis, says Downey. For example, research facilities for measuring DNA, RNA and certain proteins are available on-site, minimizing the need to ship precious samples to investigators. Eventually, the biobank itself may measure parameters that will probably be of interest to many researchers, including blood lipid levels. Informatics systems will allow data from the experiments to be shared; they will also combine requests for analyses into batches to minimize the number of times a sample is thawed and refrozen.

The project, which is nearly ten years in the making, has so far cost £85 million (US\$134 million), and is still in its infancy. The UK Biobank began accepting research proposals for ways to study these samples and associated data in March and received some 150 "credible applications" within the first three weeks, says Downey. Scientists from universities and health-care companies in any country are eligible.

Samples will grow more valuable as data accumulate. In ten years, an estimated 9,000 of the original donors will have developed Alzheimer's disease, 10,000 will have breast cancer and 28,000 will have died from heart disease. **M.B.**



The UK Biobank stores millions of samples from UK citizens.

each other. The first, which holds 9 million samples at -80°C , is essentially a freezer the size of six double-decker buses, containing a 1.5-tonne robot. Researchers can request samples by bar-code, and the robot will pluck them from their racks and bring them

can apply for accreditation and schedule an inspection. “I’ve been hearing for ten years that someone should step up and do this,” says Nilsa Ramirez, director of the biopathology centre at Nationwide Children’s Hospital in Columbus, Ohio, and co-chair of the accreditation working group. “I think it will allow investigators to have a sense that what they are dealing with is the highest possible quality.”

One difficulty with these efforts is that published guidelines are generally based on researchers’ impressions and experience, not dedicated experiments that test for the best ways to preserve samples, says Vaught, whose office is now awarding grants for assessing and developing storage technology, and is maintaining a hand-curated database of relevant peer-reviewed literature. Biobank professionals often develop their own practices after a few pilot studies but do not publish them, he says. “There are no international standards based on solid research.”

Indeed, resources for research and facilities for preserving biospecimen quality are in short supply. “People want to allocate funds for the research project and analysis, not the infrastructure that supports it,” says Compton. First-year start-up costs for a mere 50,000 samples will probably be between \$3 million and \$5 million, not including information systems. Ten-year operating costs⁷ could be more than \$10 million. Obtaining funds for ongoing expenses is also a challenge: academics are used to getting samples from their colleagues, rather than paying a repository for high-quality samples.

PRESERVING PATIENT DATA

Several organizations exist to help researchers access the samples needed for their studies. The Cooperative Human Tissue Network — a network of divisions across the United States initiated by the US National Cancer Institute to improve access to human tissue — asks researchers to collect tissue and fluids from routine surgery and autopsies, as does the National Disease Research Interchange in Philadelphia, Pennsylvania, which specializes in rarer specimens, such as eyes. Several governments have initiated large biorepository projects (see ‘A kingdom’s worth of samples and data’). The Biobanking and Biomolecular Resources Research Infrastructure (BBMRI), a network of European biobanks, is creating policies to allow researchers to share specimens and data. The Public Population Project in Genomics (P³G) based in Quebec, Canada, offers not only open-source software for documenting some aspects of informed consent, sample collection and processing, but also a database of biorepositories and their collections. Future research questions will require larger numbers of samples for rigorous statistical analysis, says Isabel Fortier, director of research and development at P³G and a researcher at McGill University Health Centre



WELLCOME LIBRARY, LONDON

Bar-coded samples can be scanned and tracked from collection to analysis.

in Quebec. “There is no way to think that just one study, even a big study, will have enough samples,” she says. “We need to give a second life to the data that we have already collected.”

Ongoing health information about a donor is increasingly desired by researchers, along with information on the preservation of any particular class of biomolecule. This has already prompted considerable reanalysis of appropriate informed consent and data policies⁸, as well as innovations in how data can be stored and mined. “It’s really going to have to be the wave of the future for biobanking,” says Jewell. “Without knowing how to manage the continuous flow of data, you’ll be a static biobank. We want to be able to constantly update the clinical record.”

The more information that is available about a specimen, the more valuable it becomes to other researchers. Scientists studying the effects of a particular gene on a cancer pathway could save years of effort and thousands of dollars if they have ready access to a collection of tumour samples with mutations of interest. David Cox, a senior vice-president at drug-company Pfizer and a member of the BBMRI’s scientific advisory board, believes that the way to get the most out of biological specimens is not prospectively banking samples but finding ways to reuse samples that researchers have already collected for their own questions. “You can’t store everything. This concept that you’re going to get all the samples and store them and then decide what to do is too expensive and it’s hard to maintain the quality.” At the same time, he says, individual negotiations with every group that collects specimens is also

inefficient. He envisions loosely coordinated ‘centres of excellence’, in which researchers store samples and track clinical information for their own research questions, but also agree to a common structure for sharing samples and maintaining their quality.

One problem is who pays for what, says Cox. “People are trying to make money off of these individual pieces instead of trying to get them all together.” Government funding is tight; pharmaceutical companies are willing to fund studies that can lead to new products, and individuals are generally willing to donate specimens and data for the public good, but not for corporate profit. One idea is that research would be conducted for pharmaceutical companies within the biobanking infrastructure, but that companies would not retain the exclusive rights to the data; however, it is too early to say whether this would be viable. There needs to be a way to link infrastructure and information “in a precompetitive fashion, so we can understand the biology better, and we can make better medicines,” says Cox. Perhaps the hardest problem of all will be establishing — and maintaining — investment. ■

Monya Baker is technology editor for *Nature* and *Nature Methods*.

1. Eiseman, E. & Haga, S. *Handbook of Human Tissue Sources* (RAND, 1999).
2. Massett, H. A. et al. *J. Natl Cancer Inst. Monogr.* **2011**, 8–15 (2011).
3. Hubel, A., Aksan, A., Skubitz, A. P. N., Wendt, C. & Zhong, X. *Biopreserv. Biobank.* **9**, 237–244 (2011).
4. Kugler, K. G. et al. *J. Clin. Bioinforma.* **1**, 9 (2011).
5. Kisand, K., Kerna, I., Kumm, J., Jonsson, H. & Tamm, A. *Clin. Chem. Lab. Med.* **49**, 229–235 (2011).
6. Simeon-Dubach, D. & Perren, A. *Nature* **475**, 454–455 (2011).
7. Vaught, J., Rogers, J., Carolin, T. & Compton, C. *J. Natl Cancer Inst. Monogr.* **2011**, 24–31 (2011).
8. Scott, C. T., Caulfield, T., Borgelt, E. & Illes, J. *Nature Biotechnol.* **30**, 141–147 (2012).