

Out of the cold	175
Out on a LIMS	177
Box 1: Treasure chest of transcripts	174
Box 2: Getting inside the bank	175

Biobanking: freezer burn

Nathan Blow

Biobanking is gaining momentum as more and more patient samples and clinical information are being stored in facilities around the globe. New technology is helping everyone—from national efforts to smaller research laboratories—to process and track their biospecimen collections.

Efforts to collect, store and distribute biospecimens for research and diagnostics has led to the development of a new field, biobanking. “Every scientist that has lots of stored samples actually has his own biobank,” suggests Rolf Muller, chief scientific officer at Biomatrix in San Diego. Although larger biobanking efforts tend to be most visible, large collections of DNA, RNA and tissue samples that exist in freezers in individual labs around the world in fact far outnumber the sample collections at national and institutional biobanks. But using samples from different locations can be fraught with trouble and is one of the primary challenges in biobanking today.

“It is my office that has identified the tissue sources for The Cancer Genome Atlas (TCGA) project,” says Carolyn Compton, director of the Office of Biorepositories and Biospecimen Research (OBBR) at the National Cancer Institute (NCI) in Bethesda, Maryland, USA, “and I can tell you that this has been a major learning experience.” TCGA is a \$100 million dollar joint initiative between the NCI and the National Human Genome Research Institute to interrogate the genome of cancer. Three years ago, Compton’s office put out a call to researchers around the world for frozen tissue from different tumor types along with matched normal reference samples for TCGA projects. The response from the community was energetic, leading researchers at OBBR to suspect they would be able to collect all 1,500 initial samples from only four to six institutions. But over time Compton came to realize that they were sorely mistaken.

When researchers searched their freezers for samples to ship to OBBR for



Tim Peakman says they took an industrial approach to the development of the UK Biobank.

testing, the dropout rate approached 99%. “They would pull the samples out of their freezers and recognize they were not big enough; or cut a section, and it would not have enough tumor, or they were missing the matched normal from the patient,” she says; “the winnowing effect was amazing.” For those samples that did make it to the OBBR and passed the histological quality control assessment, nearly 30% failed when it came to molecular testing of sample quality. Now, OBBR is working with 54 institutions, both inside and outside the United States, to support TCGA efforts. “It was an eye-opener that a transformative project of this importance was not going to have the samples that it needed to fuel its pipeline from existing collections.”

Finding samples

According to many researchers, Compton’s experiences obtaining tissue samples are not unique; many investigators have difficulty finding high-quality human biospecimens, along with clinical information about the sample, for their research efforts. For tissues, biobanking can be somewhat of a tradeoff: whereas frozen samples tend

to be easier to work with, they are far less common than formalin-fixed paraffin embedded (FFPE) samples, which are easier to store but often problematic for researchers interested in downstream molecular applications (Box 1).

“The way biobanking is done at most institutions is people just put things into the freezer without looking at them,” says Compton, “so they do not really know what is in the biobank.” These realizations have led to the emergence of several larger biobanking efforts aimed at standardizing the collection and storage of not only tissues, but also blood samples and nucleic acids.

Currently, when it comes to large biobanking initiatives, the \$105 million UK Biobank hosted at the University of Manchester in South Manchester, UK, has few peers. The UK Biobank is attempting to collect, store and eventually distribute 500,000 participant samples along with related health histories and information that will be gathered from 30–35 different clinic sites located throughout Great Britain during three and a half years. “The critical thing is we have to get at least 100 donors to go through each clinic every day, six days a week, to do this within budget,” says Tim Peakman, executive director of the UK Biobank. These numbers become more staggering when sample processing efforts are taken into account: the blood and urine collected from each donor at the clinic is fractionated by the UK Biobank to produce 28 aliquots, which means on a typical day the UK Biobank processes around 20,000 aliquots of human biological samples.

Although they rely on standard robotic systems for most simple liquid-handling

BOX 1 TREASURE CHEST OF TRANSCRIPTS

Millions of formalin-fixed paraffin embedded (FFPE) tissue samples are archived in hospitals and university around the world, representing a potential treasure trove of information about how diseases such as cancer progress. But the challenge for researchers interested in exploring these samples at the molecular level has always been the wide range in quality of the biomolecules extracted from these fixed samples and finding tools that can effectively exploit those biomolecules. This has led many scientists to avoid using FFPE samples and to the suggestion that only samples of the highest quality be banked and used in the future, leaving those millions of samples already collected sitting on the sidelines.

But Todd Golub, a cancer researcher at the Broad Institute in Cambridge, Massachusetts, USA, has a different idea. In the November 6, 2008 issue of *The New England Journal of Medicine* Golub and his colleagues described how they were able to transcriptionally profile FFPE samples, up to 24 years old in some cases, by adapting current technology¹. "I think our work is a clear illustration of the fact that suitable quality of samples is completely dependent on the technology available at any one time," says Golub.

In their hunt for molecular signatures of hepatocellular carcinoma recurrence, Golub and his colleagues wanted to profile the global transcription patterns of these tumors and surrounding tissues from FFPE samples. The problem was that existing approaches, which allowed limited numbers of transcripts to be examined, could not enable the large-scale transcriptional profiling that the team was seeking. But early on, Golub and his team recognized that the core technology developed by Illumina in San Diego for their cDNA-mediated annealing, selection, extension and ligation (DASL) assay could be thought about in a slightly different way, where

instead of profiling a specific set of candidate transcripts, with some minor modifications the system could be extended into a discovery tool for whole-genome profiling. "I think it was more a mind-set change than a technical advance," suggests Golub.

The minor modifications involved a bioinformatics analysis to demonstrate that high-quality data could be generated with fewer probes per transcript than had been previously reported, saving space on the DASL assay, and that despite scaling up it was possible to use and integrate the information from four arrays spanning the genome instead of just a single one.

Armed with their modified DASL-assay, Golub's team went back to the FFPE samples to interrogate 6,000 genes from 307 patient samples with hepatocellular carcinoma identifying a 186-gene signature of the likelihood of recurrence from surrounding tissue. "Our success rate with these FFPE samples has been around 90%," says Golub, which was essential to finding the hepatocellular carcinoma gene signature. Although he notes that the metric of success can be debated and that some sample collections have a lower success rate, they have now looked at several thousand different tissue samples, collected from different hospital sites, and are still finding a 90% rate of success.

"The 6,000-transcript array was really a proof of concept that you could use this platform for discovery purposes," says Golub, who is also a researcher at the Dana-Farber Cancer Institute. His group is now doing follow-up studies on hepatocellular carcinoma as well as extending the approach to other tumor types. And since their report, Illumina has generated a complete genome DASL assay which is now commercially available to researchers interested in tapping into the treasure chest for themselves. *NB*

and sample-moving tasks, Peakman says one automation issue they had to overcome involved fractionating blood cells from a single tube. Here the problem is the many variables that can affect the separation boundaries between white cells, red cells and plasma in a tube of patient blood. Between any two blood samples, chances are the ratios and volumes will be quite different if, for example, the nurse puts slightly more or less blood in one tube than the other or if one of the participants suffers from a dyslipidaemic condition or even has a particularly fatty meal before blood collection, resulting in more turbid plasma.

To solve this variability issue, the UK Biobank sought the help of RTS Life Science of Manchester, UK, in designing and manufacturing an automated blood fractionation system incorporating machine vision. The RTS system identifies the sample, centrifuges it and then uses

a high-resolution digital camera to take two pictures of the tube at different angles using a cold but very even light source to minimize sample damage. An image-processing algorithm then determines the boundaries between blood cells and plasma, and codes the tube for transport to another robot, which uses pipetting needles to extract the different cell types based on the machine vision results.

But robots were only one ingredient when it came to the design of the UK Biobank. "You come in here and think it is all about the robots, but the background of the people who have put this together is in understanding the industrialization of biological processes," says Peakman. This was critical, he explains, as what they are striving for at the biobank is sample consistency, which will allow for the production of good sample data in a form that researchers can access in a timely manner.

With everything up and running now, Peakman says the bank should collect the samples from half a million participants by the summer of 2010, at which point sample distribution efforts will start at the bank. "We designed the system so that it can issue 1,000 randomly distributed samples in a day," says Peakman, noting a typical case study of 10,000 individuals along with 5,000 controls would be with a researcher in three weeks time. "I think that is reasonable."

Other large biobanking efforts are starting to emerge as well, with Vanderbilt University School of Medicine in Nashville, Tennessee, USA, beginning a 250,000 person DNA study; a 20,000 person DNA biobanking effort under way at the Marshfield Clinic in Marshfield, Wisconsin, USA; and a recently announced \$8.6 million effort by Oakland, California, USA-based health care company Kaiser Permanente for the creation of a

BOX 2 GETTING INSIDE THE BANK

Chances are, if you are between 40 and 69 years of age and living in one of the major towns or cities in Great Britain, you will be contacted by the UK Biobank. “If you live within 10 to 15 miles of a center, we will get your name from an arm of the National Health Service,” says Tim Peakman, executive director of the UK Biobank. That initial information, which includes name, address, date of birth, gender, National Health Service number and the address of your general physician, is being provided to the UK Biobank under strict scrutiny, says Peakman.

As biobanking efforts expand and more banks spring up around the world, many researchers and biobanking leaders are grappling with figuring out the most ethical ways to obtain and eventually distribute participant samples. But initiatives such as UK Biobank are being viewed by some researchers as test cases for how to do large-scale biobanking. “We take issues of data security incredibly seriously,” says Peakman.

Participants identified by the UK Biobank receive a letter asking them to come into an assessment center near their home for a two-hour appointment. When they arrive at the clinic, they start by signing an electronic signature pad for consent, which gives added security by recording not only signature but also the direction and pressure of the pen stroke. The participants then answer a health questionnaire, have physical measures such as height, weight and bone

density taken, and finally give six tubes of blood and one of urine. “At that point, you receive a piece of paper with your signed consent on one side and your physical measures along with measures of a healthy person of your age and sex on the other,” explains Peakman, adding that that is the only feedback given to the participant.

Once these samples and health data have been collected, it is the responsibility of the UK Biobank to secure the information. When it comes to how the blood samples are stored and maintained, the UK Biobank is licensed through the UK Human Tissue Authority, a governmental organization set up to maintain standards in this area. They also test integrity of their computer databases by having the UK National Computing Center perform ‘penetration testing’ or attempts to break into the database to test the overall security of the system.

Shortly, the UK Biobank will have to face the ethical question sure to trouble many biobanks when it comes to their collections: access. “We have to reassure the participants that there are some foundations on which the access policy is built,” says Peakman. For the UK Biobank, that means that there will be no preferential treatment for those investigators closely involved with the project, and Peakman says they will not allow access to insurance or marketing companies and contest requests from the police unless under court order.

NB

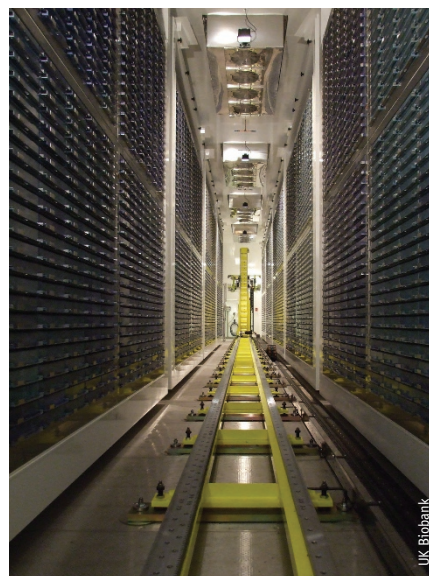
500,000 person DNA biobank in California. And Compton says that in December 2008 the NCI funded a one-year, \$1.2 million pilot project to look into the establishment of a national human cancer biobank to store tissue samples. Still, ethical and security questions associated with storing large numbers of patient samples along with their health histories continues to be an issue at leading biobanks. “The hurdles that you have to go over to get these data are quite high, and you have to demonstrate that you are a legitimate organization that can handle and store these data without leaks and compromise,” explains Peakman (Box 2).

Out of the cold

Although all the blood samples at the UK Biobank will be archived in either fully automated $-80\text{ }^{\circ}\text{C}$ freezers or in back-up manual liquid nitrogen tanks, several developers are looking to move away from freezers when it comes to storage and banking of other biomolecules.

Rolf Muller, who started San Diego-based Biomatrix in 2004, found the inspiration for his company in the most unlikely of places: a toy store. “It was just

unbelievable, here I was spending all this money to store my samples in $-80\text{ }^{\circ}\text{C}$ freezers, and I see Sea-Monkeys just hanging on the shelf all dried up,” recalls Muller. Inspired by these packets of

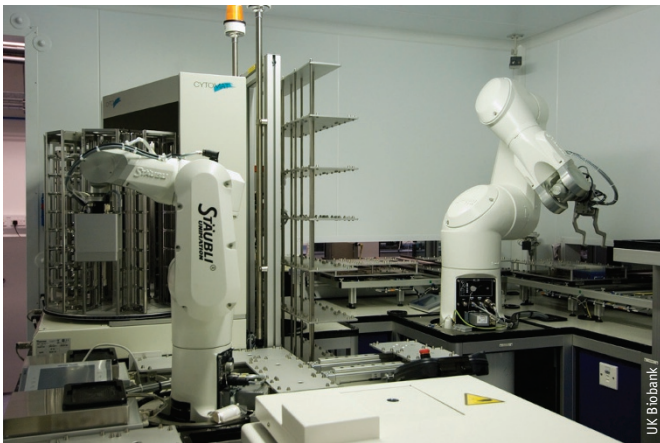


At large-scale biobanks, robotics are used to store biospecimens in $-80\text{ }^{\circ}\text{C}$ freezers.

dried brine shrimp, along with accounts of other extremophiles surviving in dry states for up to 120 years, Muller, along with his wife Judy Muller-Cohn, founded Biomatrix to look for the molecular mechanisms underlying this process of anhydrobiosis, or the tolerance of some organisms to desiccation.

Through research efforts at Biomatrix, Muller and his group not only were able to understand how brine shrimp are stable and can be revived with water by eager young children after purchase, but by using their insights along with a synthetic chemistry approach, they developed new reagents to stabilize DNA, RNA and complex samples such as blood for long-term storage at room temperature.

“What we create is a thermostable, dissolvable glass that packages every single molecule in a sample,” explains Muller. And he claims the DNA is stable for extended time periods. In accelerated aging experiments in which the encased DNA samples are placed at $65\text{ }^{\circ}\text{C}$ for 26 months—an equivalent of 30 years at room temperature—Muller says they have found no signs of degradation or sample loss.



The UK Biobank implemented an automated fractionation system for processing blood samples.

Other companies have also developed room-temperature methods for storing DNA. “DNA is a chemical polymer and doesn’t need to be handled in the same way as cells or tissues,” says Michael Hogan, chief scientific officer at GenVault, located in Carlsbad, California, USA. GenVault has

developed a mineral storage matrix, called GenTegra, for storing purified nucleic acids, which is in many ways akin to DNA stored in bone. But unlike extracting DNA from bone, which can very often prove difficult to remove, Hogan says the GenVault team learned how to store purified DNA in their

mineral matrix in a way where it could be extracted with the addition of only water.

Room-temperature sample storage is beginning to invade the world of biobanking. Hogan says GenVault recently worked with Genome Quebec in Montreal to set up sample storage for their new biobank. Blood samples entering the bank are fractionated, with the plasma fraction going into cryogenic storage for proteomics and the remaining cellular fraction stored in GenVault’s GenPlate system that uses Flinders technology associates (FTA) paper, a macroporous cellulose matrix for long-term storage of crude DNA samples.

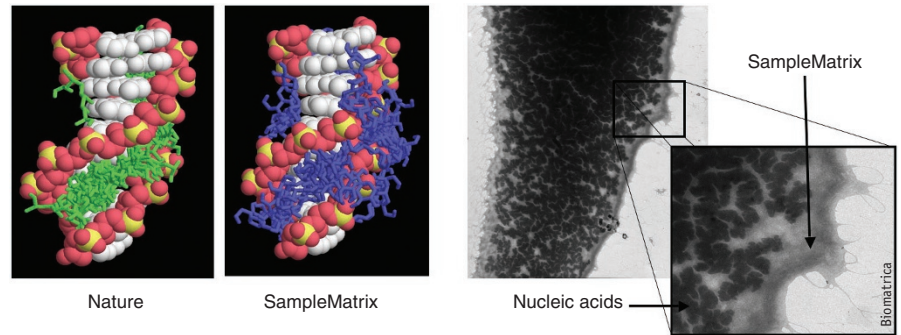
Muller also sees a bright future for room-temperature storage stabilization reagents when it comes to biobanking. “Most biobanks are storing in plates, not a single tube system so when they cherry pick a sample out of a plate they have to defrost everything,” he notes, which can damage some of the sample with each freeze-thaw cycle. With growing interest in their technologies from researchers

around the world, both Biomatrica and GenVault are not stopping with DNA and RNA: scientists at both companies are now working on room-temperature storage reagents for other biomolecules including plasma proteins and even cells.

Out on a LIMS

With sample numbers growing and new technology making storage potentially even easier and less expensive for all, researchers are looking more closely at how best to track and organize the samples and patient information flowing into and out of biobanks. This interest has resulted in a rise in development of laboratory information management systems (LIMS) specifically designed to meet the specialized needs of biobanking.

“There are two kinds of biobanks,” says Jian Wang, president and chief executive officer at BioFortis in Columbia, Maryland, USA, “one is relatively easy and the other is much more difficult.” Wang sees the centralized large biobank as being ‘easy’ because all the information can be stored in a single, centralized LIMS. The ‘difficult’ ones are for the smaller individual labs, which get their samples from a larger biobank, but the data that they generate with those samples might not go back into a centralized LIMS for other researchers to use. For this reason, BioFortis created a management system called Labmatrix, which according to Wang allows for “the management of the whole biobanking life cycle regardless of whether a central or individual lab is using it.”



Molecular modeling predictions of trehalose disaccharides (left) and SampleMatrix (center) interactions with nucleic acid molecules. Electron microscopy (right) shows the thermostable barrier that forms around the nucleic acid molecules, helping to stabilize and prevent degradation.

Labmatrix is a web-based tool that has an integration engine, allowing for the merging of research, clinical and biobanking data from different sources in different locations, along with a query interface so researchers can locate samples and data of interest. Labmatrix is not the only web-based biobanking LIMS available now: StarLIMS in Hollywood, Florida, USA, also offers a web-based LIMS package, and the NCI, in collaboration with the RAND Corporation, has developed the cancer biomedical information grid (caBIG), a web-based informatics tool for enrolling donors, tracking samples, and extracting and annotating clinical data related to cancer.

Other non-web-based LIMS packages are available as well for centralized biobanks, including Nautilus LIMS from Thermo Fisher Scientific in Waltham, Massachusetts, USA, which is being used at the HUNT Biobank in Norway as part of a population-based health study, and

Sapphire from LabVantage Solutions in Bridgewater, New Jersey, USA.

But according to Wang, the biggest hurdle in biobanking when it comes to LIMS packages is on the horizon. “I think it is less of a challenge in the physical management of the samples now and more of a challenge in the next step, adding the science onto the operational.” And according to Wang, adding in the science is the key to moving from just biobanking samples in large collections to using biobanks for discovery. “You need to look at biobanking not as an end to itself, but as a means to allow for translational research.”

1. Hoshida, Y. *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).

Nathan Blow is the technology editor for *Nature* and *Nature Methods* (n.blow@boston.nature.com).

TECHNOLOGY FEATURE

SUPPLIERS GUIDE: COMPANIES SUPPLYING BIOBANKING EQUIPMENT AND SERVICES

Company	Web address
Aperio	http://www.aperio.com/
Agencourt Biosciences	http://www.agencourt.com/
Agilent Technologies	http://www.agilent.com/
Asterand	http://www.asterand.com/
Asuragen	http://www.asuragen.com/
Bacus Laboratories	http://www.bacuslabs.com/
Beecher Instruments	http://www.beecherinstruments.com/
Biochain	http://www.biochain.com/
BioFortis	http://www.biofortis.com/
Biomatrix	http://www.biomatrix.com/
Biostorage Technologies	http://www.biostorage.com/
Clontech	http://www.clontech.com/
Cybrdi	http://www.cybrdi.com/
Cytomyx	http://www.cytomyx.com/
deCODE Genetics	http://www.decode.com/
Epicentre	http://www.epicentre.com/
Expression Pathology	http://www.expressionpathology.com/
GenVault	http://www.genvault.com/
Hamilton Robotics	http://www.hamiltonrobotics.com/
HistoRx	http://www.historx.com/
Illumina	http://www.illumina.com/
Imgenex	http://www.imgenex.com/
ILSbio LLC	http://www.ilsbio.com/
ISU ABXIS	http://www.abxis.com/
Kreatech Biotechnology	http://www.kreatech.com/
LabVantage Solutions	http://www.labvantage.com/
LabWare LIMS Solutions	http://www.labware.com/
Life Technologies	http://www.lifetech.com/
NuGen	http://www.nugeninc.com/
Pathology Devices	http://www.pathologydevices.com/
Qiagen	http://www.qiagen.com/
Rigaku	http://www.rigaku.com/
RTS Life Science	http://www.rtslifescience.com/
Thermo Fisher Scientific	http://www.thermofisher.com/
StarLIMS	http://www.starlims.com/
US Biomax	http://www.usbiomax.com/
Zyagen	http://www.zyagen.com/