

COMMENTARY

# Current standards for the storage of human samples in biobanks

Tim Peakman<sup>†1</sup> and Paul Elliott<sup>\*†2</sup>

## Abstract

Biobanks are diverse in their design and purpose; the idea of fully harmonizing historical and future biobanks is unaffordable and unfeasible. Biobanks should focus their efforts instead on developing and maintaining high-quality collections of samples capable of providing a wide range of biological information using processes that minimize introduced variability. A full data audit trail on sample processing, archiving, and quality control procedures should also be provided. This should enable the data derived from biobanks to contribute as part of wider collaborative efforts with other similar resources.

## Biobanks: the need for standardization

Biobanks are heterogeneous in their design and use, and they range in size from, say, 1,000 patients to 500,000 or more volunteers. They may contain data and samples from family studies, or from patients with a specific disease (plus ideally, matched controls), or they may be part of large-scale epidemiologic collections, or collections from clinical trials of new medical interventions. The samples collected will typically include whole blood and its fractions, extracted genomic DNA, whole cell RNA, urine, as well as, variously, saliva, nail clippings, hair and a variety of other tissues and material relevant to the design of specific studies. Inevitably, data and samples are collected under different conditions, to different standards and for different purposes. Some biobanks take a highly centralized approach to the collection, processing and archiving of samples (for example, UK Biobank [1]) where participant samples undergo minimal processing at the collection site, but are shipped to a central

processing and storage facility. While ensuring robust quality control and data integrity and security, this approach inevitably introduces a delay between collection and cryopreservation that may result in the loss of labile species in the samples. Conversely, other large studies will aim to collect and process participant samples as quickly as possible (for example, the American Cancer Society Cancer Prevention Study-3 [2]). Here, samples are collected at fundraising events and in workplace settings and are processed within a few hours by local laboratories before low-temperature archiving. The challenges here are to maintain consistency of collection, shipping and processing. A hybrid approach is taken in other studies where a proportion of the participant samples are processed and stored locally, with a second set stored in a centralized archive. Here the challenges lie in process consistency, inventory control, and management of the use of the depletable aspects of the resource. This method is being considered for the Helmholtz consortium Biobank, which is under development in Germany.

Not surprisingly, given the challenges of data collection and sample storage within particular studies, there has been little standardization across biobanks. However, a number of international initiatives are aiming to provide guidance and protocols to address this issue going forward (for example, the DataSHaPER tools developed by the Public Population Project in Genomics (P<sub>3</sub>G) [3]). The aim is to facilitate data sharing between different resources, thereby increasing effective sample size and statistical power, especially for rare diseases [4]. Rather than striving for uniformity across diverse studies, we believe it is more realistic to focus on developing and testing protocols that produce high-quality data and samples, with full information describing their collection and processing. In this way, studies will be optimized for the specific questions being investigated, while also potentially contributing to collaborative efforts that take advantage of samples from several biobanks.

## Design and implementation of biobanks: what are the basics?

Four key areas should be addressed in designing and implementing biobanks, regardless of their size and use.

<sup>†</sup>Equal contributors

\*Correspondence: p.elliott@imperial.ac.uk

<sup>2</sup>MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's Campus, Imperial College London, Norfolk Place, London W2 1PG, UK

Full list of author information is available at the end of the article

### Design and validate the sample collection protocol before main recruitment starts

An important early decision is whether samples collected from volunteers at multiple locations should be processed as quickly as possible at the collection site or shipped to a central processing facility. The first approach has the advantage that parameters that are rapidly lost within a sample may be captured, as well as avoiding possible degradation of the latent information during shipment; the second allows for a centralized approach to sample handling and processing, which may be cost-effective and result in better quality control. Either way, it is essential to minimize, as far as possible, the impact of the collection, processing, shipping and archiving protocol on the integrity of the samples. This requires properly designed pilot studies followed by robust procedures to ensure that the samples are collected, processed and handled strictly according to protocol [5-7].

### Future proof the sample collection

While some studies involving biobanks are designed to address specific questions, they may find broader use in the future (particularly as new or lower-cost analytical technologies become available). Collecting and processing samples from large numbers of volunteers is expensive and time consuming. During the design stage, it is therefore important to consider whether collection of additional samples will have the potential to produce useful data in the future, either as an adjunct to the study in hand or as part of a broader biobanking initiative. If possible, samples should be collected in a way that will allow as wide a range of assay types as can be predicted. As an example, UK Biobank collects a range of biological samples (blood, urine, saliva) that were tested in pilot studies using different analytical techniques, including standard biochemistry, proteomics and metabonomics [5,6]. In order to future proof the samples as far as possible, both plasma and serum were collected in a range of tubes with different additives (Figure 1). A similar set of samples is being collected in the Ontario Health Study [8].

### Implement quality programs from the start of the study

The sample collection and processing protocol should be underpinned by a study-wide quality program with the aim of producing samples and data that are fit for research purposes. This should include quality assurance (preventing errors and variability from occurring) and quality control procedures (detecting errors and variability if they occur) that should be built into the study design from the outset. Many studies are implementing quality schemes, such as ISO9001:2008; these are suited to biobanks because they focus specifically on the quality of the samples and data. ISO accreditation also requires

Vacutainer tube	Fraction	Number of aliquots	
		-80°C	Liquid N <sub>2</sub>
EDTA (9ml) x 2	Plasma	6	2
	Buffy coat	1	1
	Red cells	0	2
Lithium heparin (PST)	Plasma	3	1
Silica clot activator (SST)	Serum	3	1
Acid citrate dextrose	DMSO blood	-	2
Tempus tube	Whole blood (RNA)	-	6
Saliva	Mixed saliva sample	-	2
EDTA (4 ml)	Hematology (Immediate)	-	-
Urine	Urine	4	2
Total aliquots		17	19

**Figure 1. Sample collection, processing and archiving in the UK Biobank baseline assessment visit.** A variety of samples are collected in different collection vessels appropriate to their anticipated end use. Samples are fractionated and stored as aliquots in one of two low-temperature archives to protect them from degradation caused by freeze-thawing, or loss due to breakdown of a single archive site. Footnote to Figure 1: DMSO, dimethyl sulfoxide; EDTA, ethylenediaminetetraacetic acid; PST, plasma separator tube; SST, serum separator tube.

measurement of critical processes (for example, time from sample collection to ultra-low-temperature archiving) and continuous improvement efforts to optimize the performance of the organization. In UK Biobank, there has been the successful transfer of much from Japanese manufacturing quality approaches to optimize technology, processes and systems involved in sample processing [7]. By paying careful attention to the critical points in the pathway, it has been possible to reduce the time from sample collection to ultra-low-temperature archiving from an average 25.6 h (standard deviation = 3.5) to 24.6 h (standard deviation = 2.6), close to the target of 24 h based on pilot studies [9].

### Centralize and standardize as much as possible and limit the impact of variability

As noted, the degree to which sample collection and processing can be centralized will vary between studies. However, standardization and centralization of processing at a dedicated single site bring benefits in robustness of the data trail, reduced cost and increased achievable throughput and accuracy of sample handling and picking; for example, through the use of automation (Figure 2). It also limits the impact of analytical variability and thereby improves the power of subsequent analyses in which data derived from the samples are used. What should be avoided at all costs is non-detectable systematic error introduced by variable (typically manual) processing at multiple sites. Given that these resources are established to explore the etiology of complex diseases where the impact of exposure to specific risk factors will often be low (odds ratio typically 1.5 or below), this type of error



**Figure 2. Sample storage and aisle robotics used to archive and retrieve samples in UK Biobank.** Samples identified by individual barcodes are held in automation compatible racks at  $-80^{\circ}\text{C}$  in independent storage towers maintained at temperature by liquid nitrogen circulating in a closed evaporator system. All sample transfer and retrieval processes are automated to ensure accuracy and speed.

may give misleading results or mask the presence of real causative associations. This effect may be exacerbated in prospective cohorts where case-control studies are nested within the sample, especially if cases and controls are drawn differentially from different sites. If processing occurs at local sites, substantial effort should be directed into training of staff to agreed and validated operating procedures and in monitoring their performance to ensure quality standards are maintained. Cross-validation between sites will also be required. The problem of locally introduced variability through processing may be exacerbated if disease-specific studies use case and control samples from different collections. It is only by ensuring rigorous consistency and quality within individual studies that biobanks can collaborate effectively and start to exploit the potential of the very large 'virtual' sample size being created across biobanks internationally.

## Conclusions

Rather than attempting to standardize biobanks to a uniform design, effort should be focused on designing and testing the sample collection protocol in a way that produces high-quality data and samples for research use. A full data audit trail should be generated on the sample collection process to allow collaborative use of samples and data across different biobanks. It is vital that quality programs are implemented to minimize the effect of introduced variability on the integrity of the samples and, where possible, consideration should be given to future proofing the collection. In this way sample biobanks should continue to provide valuable information well into the future and provide a long-term return on the initial investment in establishing the resource.

## Competing interests

Tim Peakman is Executive Director of UK Biobank and Paul Elliott is a member of the UK Biobank Steering Committee.

## Authors' contributions

The authors contributed equally to the preparation of this article.

## Author details

<sup>1</sup>UK Biobank, Units 1-2 Spectrum Way, Adswold, Cheshire SK3 0SA, UK.  
<sup>2</sup>MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, St Mary's Campus, Imperial College London, Norfolk Place, London W2 1PG, UK.

Published: 5 October 2010

## References

1. UK Biobank [<http://www.ukbiobank.ac.uk>]
2. American Cancer Society: Cancer Prevention Study-3 [<http://www.cancer.org/Research/ResearchProgramsFunding/Epidemiology-CancerPreventionStudies/CancerPreventionStudy-3/index>]
3. Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'heureux F, Deschênes M, Knoppers BM, Doiron D, Keers JC, Linksted P, Harris JR, Lachance G, Boileau C, Pedersen NL, Hamilton CM, Hveem K, Borugian MJ, Gallagher RP, McLaughlin J, Parker L, Potter JD, Gallacher J, Kaaks R, Liu B, Sprosen T, Vilain A, Atkinson SA, Rengifo A, Morton R, et al: Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *Int J Epidemiol* 2010, doi: 10.1093/ije/dyq139.
4. Burton PR, Hansell AL, Fortier I, Manolio TA, Khoury MJ, Little J, Elliott P: Size matters: just how big is BIG? Quantifying realistic sample size requirements for human genome epidemiology. *Int J Epidemiol* 2009, **38**:263-273.
5. Elliott P, Peakman T: The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008, **37**:234-244.
6. Peakman TC, Elliott P: The UK Biobank sample handling and storage validation studies. *Int J Epidemiol* 2008, **37** Suppl 1:i2-i6.
7. Downey P, Peakman T: Design and implementation of a high throughput biological sample processing facility using modern manufacturing principles. *Int J Epidemiol* 2008, **37**:46-50.
8. Ontario Health Study [<http://www.p3gobservatory.org/catalogue.htm?jssessionid=50373D569771511A84835184B76A6468?studyid=859>]
9. Barton RH, Nicholson JK, Elliott P, Holmes E: High throughput  $^1\text{H}$  NMR-based metabolic analysis of human serum and urine for large scale epidemiological studies: validation study. *Int J Epidemiol* 2008, **37**:i31-i40.

doi:10.1186/gm193

Cite this article as: Peakman T, Elliott P: Current standards for the storage of human samples in biobanks. *Genome Medicine* 2010, **2**:72.