

The UK DNA banking network: a “fair access” biobank

Martin Yuille · Katherine Dixon · Andrew Platt ·
Simon Pullum · David Lewis · Alistair Hall ·
William Ollier

Received: 31 October 2008 / Accepted: 28 July 2009 / Published online: 12 August 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The UK DNA Banking Network (UDBN) is a secondary biobank: it aggregates and manages resources (samples and data) originated by others. The network comprises, on the one hand, investigator groups led by clinicians each with a distinct disease specialism and, on the other hand, a research infrastructure to manage samples and data. The infrastructure addresses the problem of providing secure quality-assured accrual, storage, replenishment and distribution capacities for samples and of facilitating access to DNA aliquots and data for new peer-reviewed studies in genetic epidemiology. ‘Fair access’ principles and practices have been pragmatically developed that, unlike open access policies in this area, are not cumbersome but, rather, are fit for the purpose of expediting new study designs

and their implementation. UDBN has so far distributed >60,000 samples for major genotyping studies yielding >10 billion genotypes. It provides a working model that can inform progress in biobanking nationally, across Europe and internationally.

Keywords Research infrastructure · Biobanking · Genetic epidemiology

Introduction

New experimental approaches in genetic epidemiology such as genome wide association (GWA) studies have been developed successfully in recent years. These studies seek to identify the multiple loci which each contribute to genetic risk for a given disease. As these studies progress, they require ever larger numbers of cases and appropriate controls (Davey Smith et al. 2005) with greater accuracy in sample management and deeper and more precise phenotypic information.

Identifying the required numbers is usually only possible by researchers collaborating together. However, after identifying where in the world such numbers exist, a major logistics problem of collection aggregation has to be solved. To address this problem in genetic epidemiology, managing and overseeing access to samples and data has become a specialism undertaken by a dedicated research infrastructure: the biobank. Few research laboratories have the culture,

M. Yuille (✉) · K. Dixon · A. Platt · W. Ollier
Centre for Integrated Genomic Medical Research,
The University of Manchester, Oxford Road,
Manchester M13 9PT, UK
e-mail: martin.yuille@manchester.ac.uk

S. Pullum
Azura Solutions Limited, Martells Barn Offices,
Slough Lane, Ardleigh CO7 7RU, UK

D. Lewis
Health Protection Agency, Porton Down SP4 0JG, UK

A. Hall
Leeds General Infirmary, University of Leeds, G Floor,
Jubilee Building, Great George St., Leeds LS1 3EX, UK

resources, expertise or time to support this specialist work. Biobanking encompasses the activities of an investigation that occur after study design and before data analysis. It involves processing samples and managing data in a secure, quality-assured environment. To optimise collaboration and efficiently drive the discovery process, funders and researchers alike are looking to use secondary biobanks—biobanks that manage resources originated by others.

The UK DNA Banking Network (UDBN www.dna-network.ac.uk) is such a biobank and is presented here as a case study of good practises. The research infrastructure has evolved to provide secure quality-assured accrual of resource, storage, replenishment and distribution of samples, supporting the movement of DNA aliquots and data for the purposes of new peer-reviewed studies in genetic epidemiology.

Here we outline the preparatory phase of the biobanking network; describe policies associated with the network and policies on the receipt and distribution of samples and data; summarise how sample and data management is undertaken and quality is controlled; and draw some conclusions.

Preparatory phase

In its 1999–2003 strategic plans, the UK Medical Research Council (MRC) decided to “develop new

collections of DNA samples with associated phenotypic, health and environmental information, as a unique national resource for research into multifactorial disease” (Plan 1999). £12 million had been earmarked for this unique initiative, which MRC wished to develop in conjunction with the UK charity sector and the National Health Service.

In 2000, MRC issued a Call for Proposals entitled “Establishing a UK network for DNA sample banking and genotyping”. This identified genetic epidemiology as a “strategically important area for translating knowledge of the human genome sequence into real benefits for human health”. Its proposed national biobanking network would for the first time create a research infrastructure—a biobank—to support genetic epidemiology in general. Whereas genetic epidemiology research infrastructure had hitherto been supported for ‘primary biobanking’, in which a researcher designs, accrues, manages and investigates a resource, this was a plan for ‘secondary biobanking’ where resource management is undertaken centrally.

MRC envisaged a network of centres that would house large DNA collections, manage associated databases and make the resources (i.e. samples and data) available to researchers. The initial collections populating the network were in 14 high-impact diseases (Table 1). One disease collection based overseas was unable to join the network because local law prohibited export of human DNA. These awards

Table 1 MRC collections managed by UDBN

ICD block	Collection	Stock DNA	FTA spot	PBL	LCL
C00-D48	Colorectal Ca	5,583	5,570		
	Breast Ca	227	0		
E00-E90	Type 2 diabetes	800	0	2,845	1,286
F00-F99	Unipolar depression	1,063	0	1,324	112
G00-G99	Parkinson’s	462	462		
	Alzheimer’s	375	0	2,730	626
	Multiple sclerosis	1,678	822	811	451
H00-H59	Age-related macular degeneration	1,696	0		
I00-I99	Hypertension	864			3,584
	Coronary heart disease	4,635	4,635	4,637	108
J00-J99	Asthma	863	0		850
L00-L99	Eczema	1,044	0		950
N00-N99	Glomerulonephritis	5,832	3,159	2,985	199
TOTALS	26,559	14,648	15,332	6,366	8,166

ICD, WHO International Classification of Disease (v.10); FTA, DNA spotted on Whatman FTA paper; PBL, peripheral blood lymphocytes; LCL, lymphoblastoid cell lines

had novel conditions of funding: the collections were to be managed “as shared national resources” and “made available to collaborators”. Awardees were “required to transfer a portion of each sample” to the network and “to add any genotype data they obtain to the common database”. These award conditions were critical to the success of UDBN.

Looking to the future, MRC envisaged that further collections in the network might be funded either by the MRC itself or by other funders. It saw the network as “complementary” to and being “closely coordinated” with what was at that time an early stage proposal for a healthy cohort study on 500,000 volunteers [now known as the UK BioBank project (Ollier et al. 2005)].

Competitive bids to build UDBN were reviewed by an international panel and support commenced in 2003 for three activities:

- Management of resources. DNA sample and data management is undertaken at the Centre for Integrated Genomic Medical Research (CIGMR) in the University of Manchester. Cell management is undertaken at the Health Protection Agency’s European Cell and Culture Collection (ECACC). Management of sample annotations and UDBN’s interactive website is out-sourced to a specialist bioinformatics consultancy.
- Management of access. Principles of access to samples and data have been approved by an MRC-appointed international Steering Committee that has provided oversight to the project as a whole. Terms of reference are summarised in Box A.
- Coordination of the biobanking network.

Box A UDBN management committee key terms of reference

To advise the local management committee on protocol development, operational issues, risk management strategies and good practice principles

To monitor applications to deposit and access DNA samples

To alert the Medical Research Council of any ethical or other implications which may put the participants in the research, the public and/or council at risk

Policy and its development

The policies of a secondary biobank need to govern receipt, storage and distribution of samples and data. They should satisfy the interests of all stakeholders: the patient or subject, the collector, the biobank, the recipient of samples or data, the various funders and

institutions. Some of these interests are expressed in current law. Others arise from the practicalities of carrying out a study. Where interests conflict, coordinating activities within the network of collectors and infrastructure have assisted in arriving at compromise solutions. Working through these considerations in a pragmatic fashion and recognising the need for simple, cost effective solutions, UDBN has rejected the unrestricted ‘open access’ model and has developed a ‘fair access’ model. The principle of fair access has been promulgated through the United Nations Educational Scientific and Cultural Organisation (UNESCO) International Declaration on Human Genetic Data (2003) which provides that states should regulate the flow of data and samples “to ensure fair access” (http://portal.unesco.org/en/ev.php-URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html).

The principles of fair access on which UDBN sought to develop its policies are listed in Box B. Some comment is appropriate here:

Box B Principles of Fair Access

Fair to the subject (or patient)

Privacy and confidentiality

Ethical use of samples and data

Consent management: national open methods to permit effective withdrawal of consent

Public engagement: understanding and goal-setting

Fair to the collector

Exploration rights: right to sole access restricted to specified experiments

Fair to the recipient

Collaboration management: to ensure transparency

Access to usable published/unpublished data

Long term availability of sample: long-term support and replenishment of sample

Minimum of administration

Fair to collector’s and investigator’s institution

Intellectual property rights management: long term tracking of samples, data, uses

- The central concern of researchers is control: who controls access to samples and data? On this question, UDBN opted simply to rely on an existing type of relationship between researchers that is as old as science itself: the relationship of collaboration. This has the advantages of having

well-understood and accepted rules, of being largely self-monitoring and of having proven effectiveness and value. UDBN avoided the ideal of “open access” wherein non-collaborators could access samples and full data sets. Open Access requires the creation of a cumbersome oversight mechanism to minimise misuse of data. This misuse can arise unintentionally: it could be a consequence of the fact that biomedical data categories have not been standardised. If the biomedical research community were to agree a common ontology, then it would be advisable to consider this matter again. However, most future collectors are likely to continue to introduce some non-standard (i.e. experimental) biomedical data categories into their accrual work. Such innovation is essential and desirable but may allow unintentional misuse of openly accessible biomedical data by a non-collaborating investigator. The counter-argument to the problem of this unintentional misuse is that peer review of the resultant manuscript would discover the misuse. However this counter-argument is flawed unless peer review is carried out by the collector. This would then contradict the principle that peer review should not be undertaken by a person with a competing interest.

- National consent management is an ambition of UDBN primarily because of the benefits it can provide to the subject. It would create a single contact point where members of the public can go if they wish to withdraw consent. For now, a subject must recall precisely who consented them for what. It also benefits collectors by reducing administrative burdens. A consent database needs to be isolated, with linkage to other data sets only permitted to a ‘consent guardian’. To realise this ambition, substantial work needs to be undertaken with all stakeholders.
- It is reasonable to assert the right of a collector to exclusive access to his/her collection for the purposes of the investigational goals stated in the initial collection proposal. A collection proposal with no stated investigational goals is unlikely to attract support. A proposal with too many such goals seems unfocussed and so is unlikely to attract support. A collector who attains his initial goals will be in the strongest possible position to set the next goals—with or without collaborators.

This ‘first mover’ advantage is inevitable and probably desirable. Quashing that advantage with an ‘open access’ policy seems undesirable because of its possible demotivating effect.

- When an investigator seeks to collaborate with a collector, anecdotal evidence suggests that the relationship can come under stress. This can endanger the security of shared confidential data as well as the joint investigation. By tracking communications between the two parties, the biobank encourages them to have a more transparent relationship and is well-placed to mediate if stresses arise.

Coordination of the network

The need for coordination and continuous dialogue between, on the one hand, the 13 groups of investigators involved in making disease-specific collections and, on the other, the research infrastructure was recognised at an early stage by one of the clinical investigator groups. As a result, the MRC made funding available specifically for networking meetings. The case for networking was based on the fact that a number of the investigator groups were already funded for other collection activities by other funders and so already had well-established procedures. This case was strengthened by recognition of the rapidly evolving research environment into which UDBN was entering.

Inclusive and flexible biannual network meetings have therefore been held to facilitate the process of integration of the collections at UDBN. Throughout, each party has chosen freely whether to participate in developments within the network. At times, when a universal consensus view has been slow to emerge, a system of phased integration has been used. This has entailed early adoption, delayed adoption or even late adoption of a proposed common practise. Participants have thus been able to engage at a speed and in a manner suited to their own particular needs. These needs sometimes included the ability to observe the effective engagement of others. This process has seen a change from restrictive and competitive patterns of academic behaviour to a pattern of broad and effective collaboration as reflected, for example, in the success of the Wellcome Case Control Consortium (2007) seven of whose investigators hold collections in UDBN facilities.

Sample receipt at UDBN

Under the terms of their MRC awards, collectors were required to deposit an aliquot of DNA at UDBN. A Material Transfer Agreement (MTA) between the collector's institution and UDBN's institution (University of Manchester) was therefore appropriate. Its main terms are shown—simplified—in Box C. UDBN needs reassurance that all material was collected ethically and needs information so that it can act if advised of a subject's withdrawal of consent. The collector needs reassurance that UDBN will implement a proper duty of care for the resource.

Data receipt at UDBN

Extensive discussions were held with collectors leading to decisions that:

- Personal data (e.g. name, address, National Health Service number) should not be held by UDBN. The policy re-assures any collectors who might retain concerns over theoretical breaches of

patient confidentiality. For most collections, UDBN has a sample identifier but no personal identifier. This policy limits the closeness of collaboration between a collector and a third party researcher. It also delays development of centralised consent management. Such management is needed to increase the transparency of research and thus maintain public trust. For some newer collections, UDBN undertakes project management during the accrual phase. This can require UDBN to hold some personal information. However only appropriate, registered, named individuals (e.g. Principal Investigators) can access this information.

- Searchable summary data on 10 phenotype data categories specified by each collector should be made accessible to *bona fide* researchers.

UDBN was built before GWA studies had been initiated. However, it was clear that GWA data would become a major resource and that the bioinformatics challenge of managing this data should be addressed by the community as a whole. UDBN is working with

Box C Material transfer agreement between collector and UDBN (simplified)

Depositor institution undertakings

The depositor institution agrees to:

- Provide UDBN with the principal investigator's successful proposal(s) seeking support to undertake the collection and with the corresponding award letter(s) and terms and conditions of support
- Provide UDBN with ethics committee approval(s) of the collection and the form(s) of consent to be signed by the subjects of the collection
- Affirm that the consent form allows retention of samples in a place whose definition embraces UDBN and allows genetic characterisation of the samples
- Provide UDBN with the questionnaire(s) to the subjects of the collection and with a list describing the clinical and other phenotypic tests performed
- Delegate to the principal investigator the responsibility to enable appropriate access to information and data on the collection
- Require the Principal Investigator to inform UDBN promptly of the unique identifier of DNA material when the corresponding subject is deemed by the principal investigator to have withdrawn consent described elsewhere

UDBN undertakings

UDBN agrees to:

- Exercise a proper duty of care for the collection, taking all reasonable measures to ensure
 - Safekeeping and conservation of the DNA sample
 - Replenishment of the DNA sample to the extent that funds allow
 - Consistency between the information on the collection made available via UDBN and the information provided to UDBN by the principal investigator and registered users
 - The confidentiality of all items or types of unpublished information on the Collection specified by the principal investigator
 - Remove from the collection and take reasonable steps to destroy DNA material for a given subject when the principal investigator deems that subject to have withdrawn consent
 - Promote on the UDBN website and in other media the appropriate use of the collection
-

the European Bioinformatics Institute's European Genome-phenome Archive (www.ebi.ac.uk/ega) on this.

Data distribution

Figure 1 illustrates the steps that need to be taken by a website user to access data and DNA aliquots from UDBN.

Data are distributed via the project website (www.dna-network.ac.uk) to third party researchers. Their *bona fides* are established when a website user applies online for registration. If successful, this yields a user name and password to allow access to a restricted area of the website. An applicant must supply their personal workplace email address and the workplace email address of their systems administrator. This allows UDBN to verify the applicant's association with the workplace and then, separately, to establish its nature. Any workplace that is not an academic research body is carefully scrutinised. If the workplace business were in the legal, insurance or law enforcement field, the application would be referred to lawyers at UDBN's host institution. Virtually all applicants have been from European academic institutions. Applicants from drug development

companies had previously established collaborations directly with UDBN collectors.

Information about and summary data on each collection is held in a relational database held on UDBN's secure server at the University of Manchester. The number of phenotype data items varies between collections from four to 35. This is based currently on data snapshots provided by collectors. A search tool allows registered users—researchers—to explore data sets and sub-sets. The researcher is able to communicate online with the group of collectors responsible for a collection to negotiate and pursue collaboration. The communications are logged to allow UDBN to provide an audit trail and to permit UDBN to contribute to negotiations where necessary.

After collaboration is agreed between a collector and a researcher, it is likely that, in designing their collaborative study, they will want to share data in full, not just summary data. This is implemented by UDBN giving the collector the ability to vary permissions so that the named researcher may access fuller data.

The current incomplete phenotype data sets on the website are adequate for the purpose of third parties who wish to initiate a collaboration. However, there is in fact no risk to collectors associated with presenting summary data on all their standard phenotypes. In the special case of novel phenotypes devised by a collector, release of summary data can be delayed until those phenotypes have been described in a peer-reviewed journal. Such publication provides an adequate measure of quality assurance for those novel phenotypes. If publication is to trigger website data release, then good communications between the infrastructure and the collector need to be maintained over an extended period. It is important that collectors see secondary biobanking as improving their access to new collaborations since, as this is increasingly recognised, their motivation to release all summary data sets increases. Therefore networking events are necessary not only to establish but also to maintain the scientific value of the biobank.

The website registration system is of value because it [a] acts to reassure collectors; [b] will facilitate agreement on access to genotyping data and; [c] will aid use of new tools that UDBN is developing to streamline sample accrual.

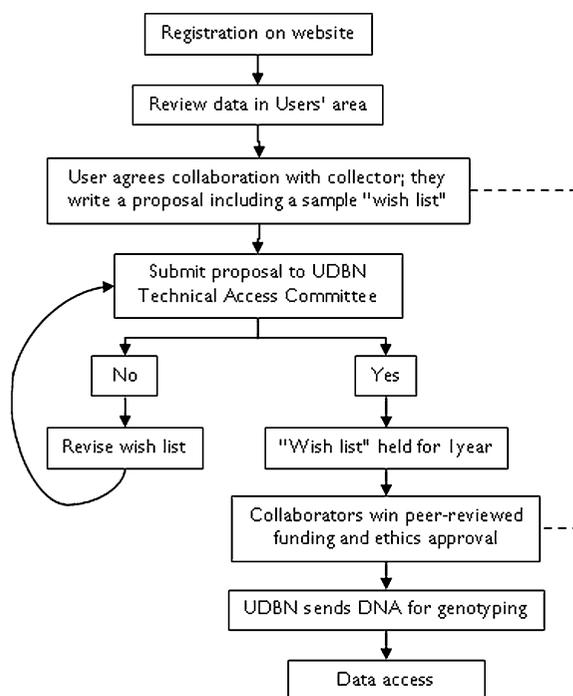


Fig. 1 Approval process for sample distribution

Sample distribution from UDBN

Registered researchers can generate online a “wish list” of samples that they want to test in collaboration with a collector. Since the study they are planning requires peer review and may require ethics approval, the researcher and the collector need to prepare a proposal (off line). This proposal will contain the wish list and will identify the planned tests on those samples.

UDBN requires a copy of the proposal so that its Technical Access Committee (TAC) (a sub-committee of the MRC Steering Committee) can determine whether the proposal is acceptable from a technical point of view. TAC members consider nine questions (Box D) aimed at ensuring that efficient and high quality work is planned; that renewable DNA sources will be used where possible; that the resulting data will in due course enrich the resource for other investigations. After TAC approval, the wish list becomes a “pick list” for the biobank’s Laboratory Information Management System (see below). This will then permit samples to be picked out of the storage system. However no picking starts until the biobank is informed of successful peer review and ethics approval. If TAC finds the proposal unacceptable, the researcher and collector are informed and biobank staff will help them modify their proposal to address TAC’s concerns.

Before DNA sample picking starts, the recipients are required to accept terms and conditions listed in Box E. This ensures that DNA is used for the stated

purposes only and not for commercial gain; and that UDBN will receive access to all genotype data at the time of its publication.

UDBN is minded to modify this procedure to bring it into line with UK BioBank. This infrastructure plans directly to commission genotyping or other tests specified by a researcher for an approved investigation. In this way, data return to the infrastructure is guaranteed and consistency of experimental data quality can be assured.

Sample and data management

Sample management

Samples can be accepted at UDBN’s DNA laboratory either as 7–10 ml EDTA blood samples or as DNA aliquots in wells or tubes. On arrival, samples are logged-in to UDBN’s Laboratory Information Management System (LIMS) using the collector-supplied unique donor identifier. The identifier is typed in twice and is compared with an external list of identifiers from the collector. Blood samples are extracted robotically using proprietary M-PVA Magnetic Beads (chemagen Biopolymer-Technologie) and DNA is dispensed into 1.0 ml tubes (Griener). DNA concentration estimation is performed using emission spectrophotometry (FluoroSkan) run on a Tecan Genesis Freedom 2000 automated liquid handling workstation.

Dilution of stock to 100 ng/ul in 0.1 M Tris/0.01 M EDTA pH 7.5 is performed using a Hamilton Star liquid handling work station, after mixing using a Microtec vortexer. 1D bar coded tube racks, fitted with lugs that prevent misleading, hold 96 2D bar coded Matrix tubes and are stored in customised freezers. Location data is captured using a hand-held PDA device (Falcon; Unitech PT930S). Microsoft Active Sync software automatically uploads location data from the PDA to the host PC. –80°C freezers are located in the University of Manchester. Second site storage of an aliquot at the UK BioBank facility near Manchester guards against catastrophic loss.

To withdraw an aliquot, racks are retrieved manually, tubes are picked using the BioMicroLab XL20 Tube Handler, thawed and mixed. Aliquots are removed and dispensed using a liquid handling robot to a specified tray format and layout. Volumes

Box D Issues addressed by technical access committee

-
- The platform(s) is appropriate for the tests
 - There is no other platform available that uses substantially less DNA
 - The out-sourced genotyper is reputable
 - The amount of DNA/sample requested matches the amount of DNA/sample required
 - The requested concentration of DNA is reasonable for the platform
 - The test(s) can be performed with comparable reliability on cell line DNA or on WGA-DNA as on blood-derived DNA
 - Less than 20% of the tests on the samples requested have not been performed or submitted for approval previously
 - The plan for data access is reasonable
 - The peer review processes of the funder are acceptable
-

Box E Terms and conditions of release of DNA material

The Recipient may use the DNA material only in the common good

The Recipient shall not sell, transfer, release or distribute the DNA material to a third party or use the DNA material for profit or any other commercial gain

The Recipient agrees to only use the DNA material in the previously specified tests

The Recipient agrees to give UDBN access to all the data on all tests performed on the DNA Material within 10 days of the date of publication of a paper reporting on any of the tests on any of the samples. Unjustified delay will lead to UDBN advising the funders and the depositor institution of non-compliance

The Recipient agrees that UDBN may publish (at a time not before the date of publication of a paper that describes the results of the specified tests) via the UDBN website

1. General information about the tests to the public
2. Summary data about the results to registered users of the UDBN website
3. De-identified subject-specific data about the results to registered users named by the recipient

The Recipient shall acknowledge UDBN as supplier of DNA from the relevant collection in all written or electronic reports and publications

At the request of UDBN the recipient agrees to return or destroy any remaining DNA material at the end of the project

remaining are measured using the BioMicroLab VolumeCheck. Trays are despatched on dry ice by courier.

The LIMS used to track samples within the laboratory is a fully customised installation of Nautilus (version 8.1) (Thermo Fisher Scientific). It is based on an Oracle 10 g database. The most basic in-house customisation involved the creation of a workflow for each type of sample to be received by the laboratory; the workflow models the processes which may be performed on the particular sample type. Additional in-house customisation provides validation and extra functionality using Infomaker reports (Sybase), Oracle PL/SQL and Microsoft Visual Basic 6 extensions.

The LIMS server makes use of a redundant array of five independent disks. The redundancy ensures that if a disk crashes, LIMS operation is not disrupted. Data is safeguarded using a twenty tape rotation tape backup strategy.

Phenotypes, genotypes and the website

The design of the database was the key issue when developing the website. It needed to be flexible (because new phenotypes are being used) as well as scalable for the large volumes of data expected. The desire for the system to be “open source” for academic institutions helped decide the technologies that would be used.

The database was created using MySQL and when each collection is added to the database, meta data is

created for each phenotype to aid processing and comparison of the data items, both in terms of filtering within the collection and comparison with other collections.

Java web services are used to process the data and manage the workflow for collectors using the system, which is all brought together with a content management system written in PHP that then presents the data and collaboration management features to users.

Using these technologies provides a highly flexible platform allowing easy integration to other systems so that services that might exist elsewhere—as at genotype databases—need not be “re-invented”.

Quality control and assurance

All three infrastructure providers in UDBN implement ISO9001-2000. All procedures are specified in standard operating procedures; all use of procedures is monitored and all versions of procedures are tracked. This ensures reproducibility. A condition of biannual accreditation by the International Standards Organisation is a programme of quality improvement. This condition is, in its effects, a driver for undertaking methods research. The decision to seek ISO accreditation for UDBN was based on the recognition that an infrastructure should add value in resource management in ways that conventional research laboratories will not or cannot. The culture of a conventional research laboratory encourages innovation at all

times, whereas the culture of a research infrastructure seeks to manage innovation within a strict quality control and assurance framework.

Methods research has been undertaken on a number of processes, for example on DNA concentration estimation. UDBN, in conjunction with Public Population Project in Genomics (P3G: www.p3gconsortium.org), has performed an observational study on 14 academic and commercial laboratories for DNA concentration estimation. This has quantitated the errors in estimation that arise both within and between labs that have hitherto been the subject of anecdotal reports (ms in prep). UDBN, jointly with P3G and with European partners in the pan-European Biobanking and BioMolecular Resources Infrastructure initiative (BBMRI: www.biobanks.eu; Yuille et al. 2007) is now planning to identify and collaboratively reduce common major sources of this error. This will reduce costs and the long delays that are seen in large studies after the required samples have been identified and before genotyping has actually begun.

Discussion

UDBN has provided over 60,000 aliquots of the 40,000 samples that it manages for candidate gene and GWA studies, including those conducted through the Wellcome Trust Case Control Consortium (2007). New collector groups are now working with and seeking to work with UDBN on accrual management as well as sample and data management for both cross-sectional and longitudinal studies. Thus this secondary biobank is resolving major bottlenecks in genetic epidemiological investigation and accelerating progress in this field. It is also reducing the need for repeated investment in new infrastructure and resource management, helping clinicians focus on

their core competencies in study design and paving the way for future investigations based on shared phenotypes, genotypes or other exposure.

The debates that have taken place within the network and with colleagues in BBMRI and P3G have led UDBN to formulate some principles of ‘fair access’. They are pragmatic principles based on existing practise. We suggest they meet the basic needs of the genetic epidemiology community for improved access not just to data but also to samples. This is the key advantage of secondary biobanking over data-sharing schemes such as the Genetic Association Information Network (2007).

A summary of the advantages of secondary biobanking is shown in Box F and below are identified some of the main general lessons we have learnt so far from building UDBN.

- Investment by funders in research infrastructure, as distinct from buildings infrastructure, has not been a priority for biomedical funders (Note that in the physical sciences the distinction between the two can be blurred—in astronomy one cannot construct a large optical telescope without a building to house it). UDBN’s resource management requires small flexible robots and data networks. A good primary biobank may attract support for this as part of a specific investigation. However it was the vision of the benefits of secondary biobanking that persuaded MRC to support UDBN. A critical aspect of this support was the condition of support given to collectors requiring them to provide samples to UDBN and to lodge data in a common database. Without this condition, UDBN would have had great difficulty in overcoming fears about confidentiality and ownership that impede resource aggregation in biomedical research.

Box F Advantages of secondary biobanking

Acts as an ‘honest broker’: its policies must be fair to investigators as depositors or recipients of samples or data

Ensures that a collection can outlive its originators

Facilitates extending studies (e.g. from cross-sectional to longitudinal)

Extends the useful life of a non-renewable resource. A primary biobanker has a powerful motivation to perform “just one more experiment” on that resource

Offers and creates economies of scale

Offers advantages in resolving the ethical, legal and societal issues connected with biobanking

Provides an appropriate environment in which to conduct methods research to improve resource management

- Secondary biobanking has a structuring effect on research. It manages all samples in a standard way regardless of the investigation for which they have been collected. The data that it aggregates, while it may be deposited using different vocabularies, is readily amenable for the development of flexible ontologies that embrace multiple collections in one or more diseases. It therefore facilitates the future aggregation of new or larger resources. Such resources are precisely what genetic epidemiology requires as it seeks to unravel smaller effects. For example, a recent genetic study on breast cancer was based on 50,000 patients and its authors (Easton et al. 2007) concluded: “The detection of further susceptibility loci will require genome-wide studies with more complete coverage and using larger numbers of cases and controls, together with the combination of results across multiple studies.”
- Secondary biobanking, through its focus on resource management, is well placed to implement the highest standards. This promotes research in a neglected area. Standard methods research gets low scores from funders compared with novel methods research.
- When UDBN provides resource management for new collections, the infrastructure cost is marginal. Hence, secondary biobanking is a more efficient use of financial resources since it eliminates the requirement for separate primary biobanking infrastructure. These savings free up funds allocated to genetic epidemiology for more or larger investigations.
- The benefits of a secondary biobank grow over time through cost savings and synergies. Therefore support for such an infrastructure should not be based on short or medium term grants. This has been recognised, for example, in the planned development of the pan-European BBMRI biobanking system (Yuille et al. 2007) where the intergovernmental European Strategy Forum on Research Infrastructures recommends initial support for 10–20 years (European roadmap for research infrastructures report 2006).
- A secondary biobank provides a useful way of achieving public policy goals of research. Biomedical research must engage with the public to

maintain and build trust. It is unreasonable to expect a primary biobank to work on this area. However a secondary biobank is well placed to undertake consent management, outreach work, educational initiatives and public engagement because, if it has a fair access policy, then it is obliged to be proactive in being fair to patients. UDBN has undertaken some educational initiatives (e.g. its Practical Biobanking courses held at the University of Manchester) that start this process.

- Could the UDBN model of secondary biobanking be replicated in other jurisdictions? Could such biobanks form a network? The identification of roadblocks in other jurisdictions to creating a national secondary biobank is beyond the scope of this review. However, the main practical issue is whether national funders decide to adopt this approach and then to impose the necessary grant conditions. It should be noted that the primary distinction between a secondary biobank and a conventional primary biobank is policy-related: an existing primary biobank can adopt the policies of a secondary biobank and thus function as both simultaneously. An international network of secondary biobanks is practical, even if national laws prohibit sample and data export. A network would only be curtailed by one jurisdiction outlawing collaboration with scientists in another jurisdiction. Few, if any, states do this.

Acknowledgments The UDBN is supported by the Medical Research Council. The award was made to the following network members: S. Bhattacharya (Institute of Ophthalmology, University College London); H Campbell (University of Edinburgh); M. Caulfield (William Harvey Research Institute); C. Clarke (University of Birmingham); A. Compston (University of Cambridge); W. Cookson (Imperial College London); M. Dunlop (University of Edinburgh); J. Goodship (University of Newcastle upon Tyne); A. Hall (University of Leeds); A. Hattersley (Peninsula Medical School, Exeter University); D. Lewis (Health Protection Agency); M. McCarthy (University of Oxford); P. McGuffin (Institute of Psychiatry, London); M. Moffatt (Imperial College, London); K. Morrison (University of Birmingham); P. Munroe (Barts and The London Medical School); W. Ollier (University of Manchester); M. Owen (Wales College of Medicine, Cardiff University); A. Rees (University of Aberdeen); N. J. Samani (University of Leicester); E. Solomon (King’s College London School of Medicine); A. Webster (Moorfields Eye Hospital, London); J. Williams (Cardiff University); M. Yuille (University of Manchester).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR (2005) Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 366:1484–1498
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–1093
- European roadmap for research infrastructures report (2006) European Strategy forum on research infrastructures (ESFRI). Office for official publications of the European communities, Luxembourg
- Ollier W, Sprosen T, Peakman T (2005) UK biobank: from concept to reality. *Pharmacogenomics* 6:639–646
- MRC Strategic Plan 1999–2003 (1999) Medical Research Council, London, p 13
- The GAIN Collaborative Research Group (2007) New models of collaboration in genome wide association studies: the genetic association information network. *Nat Genet* 39:1045–1051
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14, 000 cases of seven common diseases and 3, 000 shared controls. *Nature* 447:661–678
- Yuille M, van Ommen GJ, Bréchet C, Cambon-Thomsen A, Dagher G, Landegren U, Litton JE, Pasterk M, Peltonen L, Taussig M, Wichmann HE, Zatloukal K (2007) Bio-banking for Europe. *Brief Bioinformatics* 9:14–24