Review and Perspective

Infectious Diseases Biobanking as a catalyst towards personalized medicine: 
* {\em M. tuberculosis} paradigm

Fotini Betsou\textsuperscript{1}\*, Shreemanta K. Parida\textsuperscript{2}, Martine Guillerm\textsuperscript{3}

\textsuperscript{1}Integrated Biobank of Luxembourg; 6 rue Ernest Barble, L-1210 Luxembourg

\textsuperscript{2}Vaccine Grand Challenges Program, Dept. of Biotechnology, Govt. of India, New Delhi, India

\textsuperscript{3}The UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (WHO/TDR)

Disclaimer: The opinions expressed in this article are those of the authors and are not to be construed as official or as representing the opinion of respective organizations.

Keywords: infectious diseases, tuberculosis, biobank, biomarker

*Corresponding author: F. Betsou, Integrated Biobank of Luxembourg, 6 rue Ernest Barble, L1210 Luxembourg. Tel. 0035227446456. Fax 0035227446464. Email fay.betsou@ibbl.lu
Abstract

Research for biomarkers supporting personalized medicine in infectious diseases is needed, especially for tuberculosis in which the existing toolbox does not yet address the public health priorities. Biobanks are essential infrastructures in this effort by collecting, authenticating and preserving human and/or bacterial specimens. A broad range of specimens should be collected prior to, during and following treatment, with a comprehensive characterisation of the sample donors and the samples themselves to accommodate the most recent technological platforms in biomarker research. This review explains current state-of-the-field biobanking practices in tuberculosis and suggests technical and managerial improvements to ensure long-term preservation and optimal use of the specimens. Open-access and certified biobanks are an essential component of a strategy supporting the development of drugs and diagnostic tests for both public health and personalised medicine. Biobanks have a role to play in the interaction between these two - not always compatible - approaches.
Introduction

Tuberculosis is a major public health problem which infects one third of the world population and kills more than two million people every year\(^1\). The progression of *M. tuberculosis* infection to disease manifestations characterized by chronic granuloma and caseation necrosis, or containment of the disease manifested by granuloma formation with arrest of bacterial multiplication, depends on resistance and immune responses of the host. The granuloma formed as the result of the infection may either resolve or progress to massive necrosis of lung tissue, or develop into a chronic lesion. Though TB primarily affects lungs, it can also affect other organs such as lymph nodes, bones, gastro-intestinal system, genito-urinary system, skin and even eyes. The diagnostic specimens (though predominantly are fresh sputum and blood) may thus include gastric washings, urine, pleural fluid, cerebrospinal fluid, joint fluid or biopsy material.

The differential development of clinical disease probably has a genetic component. This has not been clearly delineated but seems to involve the natural resistance associated macrophage protein 1 (NRAMP1), which regulates macrophage activation\(^2\). The relatively low proportion of individuals who progress to active disease after infection can probably be ascribed to innate resistance in most infected individuals, since vaccination using BCG or a previous episode of TB does not work reliably or effectively to confer protection in high burden parts of the world\(^3\). Environmental factors, such as under-nutrition and various other “personal” factors, including age, cell-mediated immunity, anti-inflammatory treatments, and coexisting diseases (diabetes, other infections) may also influence the disease outcome.

Multi-drug resistance to both isoniazid and rifampin is an increasing worldwide problem and second-line drugs, such as kanamycin or ofloxacin, are more toxic and/or less effective. In addition to the threat of multi-drug resistance (MDR) TB, the last few years have seen an explosion of the newly-emerging threat of extensively-drug resistant (XDR) TB which is resistant to isoniazid and rifampin plus resistant to any fluoroquinolone and at least one of three injectable second-line drugs (i.e. amikacin, kanamycin or capreomycin) \(^4\). Significant research efforts are oriented towards identification of biomarkers with potential application as vaccine antigens (immunogenic virulence factors), as early and specific diagnostic tools of latent or active disease, prognostic tools of severity of disease, and as predictive tools of response to therapeutic regimens\(^5,6,7\). Early response to treatment might be an indication for a shortened course of antibiotic therapy. This response pattern potentially affecting treatment
time is why early classification of patients into risk groups requiring different durations of antibiotic therapy may be important in improving adherence and clinical outcome - which is the goal of personalized medicine (the right treatment to the right patient at the right time, and in the right dosage).

Biobanks, as organized collections and providers of biospecimens, have become strategic and essential tools in research efforts for identification and validation of surrogate biomarkers. Biobankers collect, authenticate, preserve and offer independent access to biological materials, such as specimens and cultures for research purposes. Biospecimens are prepared in multiple aliquots for long-term storage so that future researchers will be able to use them as new technologies and knowledge evolve over time. In the area of infectious diseases, human material of infected sample donors and isolated microbial strains are key elements of such biobank initiatives. In this case, microbial strains isolated from culture of collected human samples correspond to derivatives of these samples. Other derivatives include host and microorganism nucleic acids, proteins, or isolated immune cells. Accurate characterization/authentication of the derivative strains is crucial for correct diagnosis as well as for epidemiological analyses and our understanding in the spectrum of pathogenesis of the disease with different variants. These efforts are critically important, for both prevention of disease transmission and efficient use of the stored samples in research. In this paper, we argue that biobanks need to guarantee the authenticity and reliable characterization of all the supplied biomaterials, in order to fully realize their value as enhancers of research towards personalized medicine and public health amelioration in the infectious diseases area, and we use the M. tuberculosis as example.

**Definitions:**

A ‘bank’ is a place where individuals or groups deposit their own assets for short- or long-term periods. The bank holds it safely for them, uses it, takes care of it (and protects its value) - but doesn’t own it. And a depositor is free to withdraw his/her assets anytime. Similarly, a biobank is a facility for safekeeping and storage of biological specimens for research.

Research participant: A research participant is someone who freely consents to be part of a prospective research project (not usually the case in retrospective research).

Sample donor: A ‘sample donor’ is someone who freely consents to make a gift /’donation’, independent of a specific research project.
The status of *M. tuberculosis* biobanking

Apart from collections of biological material which may be stored in hospitals or clinical laboratories, organized *M. tuberculosis* biobanks are very few. Organized biobanks, having established procedures and technologies for the systematic collection, annotation and very long term preservation of biospecimens, are expected to achieve both the necessary quantity and quality of samples to allow efficient research to be conducted. In terms of numbers of donors, and assuming a 5% relapse rate after initial therapy of drug sensitive TB patients over a 6-9 months period, a biobank should collect samples from at least 2,400 donors, in order to be able to provide 120 recurrence cases. Assuming a 25-35% relapse rate of MDR TB patients over the same period, a biobank should collect samples from at least 350-480 drug resistant patients, in order to be able to provide 120 recurrence cases. The numbers must be adequate to differentiate in the test systems to give statistical power for meaningful conclusions and sufficient to provide reference intervals of newly identified biomarkers. Guidelines promulgated by the Expert Panel on Theory of Reference Values of the International Federation of Clinical Chemistry (IFCC) specify that a minimum sample size of 120 reference subjects should be used for estimation of a reference interval. The ideal sample type for reliable tuberculosis biomarkers is still unknown; therefore, as many different types of samples as possible should be collected (within funding capabilities).

The TB Immunology Group Research Tissue Bank in the UK has announced collection of blood, sputum, urine, pleural fluid, pleural biopsies, pericardial fluid, ascitic fluid, CSF, bronchoalveolar lavage specimens, transbronchial lung biopsies, open lung biopsies, renal, adrenal, liver, gastric/bowel, laryngeal, muscle, bone marrow, lymph node and skin biopsies, to be used in collaborative research studies on the physiopathology of tuberculosis. The Global Alliance for TB Drug Development (TBA) has launched the REMox TB clinical trial incorporating moxifloxacin treatment. Funding has recently been secured, allowing development of a biobank. This biobank will collect whole blood, RNA of *M. tuberculosis* strains, sputum, and urine from study participants with tuberculosis undergoing treatment in the trial. TBA has also recently proposed the establishment of a Consortium for TB Biomarkers (CTB2) with the National Institute of Allergy and Infectious Diseases (NIAID)’s AIDS Clinical Trials Group (ACTG) and the U.S. Centers for Disease Control and Prevention (CDC)’s TB Trials Consortium.
(TBTC), to combine their biorepository efforts in order to 1) agree on standards for collection, processing and storage of a core set of relevant samples, including use of common data standards and elements, where possible; 2) establish an appropriate Biorepository of the agreed upon samples, and 3) establish a peer review system for making the samples available to qualified scientists throughout the world for TB biomarker and diagnostics discovery and development, to enable discovery and ultimately qualification of biomarkers of TB drug effect, including a marker of stable TB cure vs. relapse as the first priority (personal communication, E. Gunter, A. Ginsberg).

The WHO/TDR( World Health Organization Special Programme for Research and Training in Tropical Diseases) assembled a collection of well characterized serum, saliva, sputum and urine samples using a standardized protocol for prospective collection, characterization, storage and maintenance. Samples were collected in collaboration with research teams from 13 different countries worldwide from untreated TB suspect patients. Clinical and biological information is recorded from each patient. The WHO/TDR TB Specimen Bank applies an open access policy through which specimens are available for no cost, (except shipping and distribution fees) to any scientist or diagnostic product developer after examination of the request by an independent scientific Review Committee. The TDR Specimen Bank is the only open-access collection of samples from well-characterized patients from different geographic locations representing a wide spectrum of samples from different patient cohorts, which are made available to any researcher or organization interested in developing diagnostic tools for use in poor resource settings. Since its formal launch in 2000, 95 applications have been reviewed and 65 (68%) approved for use by interested investigators across the world. The specimen bank was intended to stimulate commercial R & D for simple point-of-care assay development, set high standards of quality for tools in development, assist in quality control, limit the need for field trials, facilitate the approval process and simplify direct comparison of new and existing diagnostic kits. For 10 years this biobank has been approached by private companies and academic researchers and facilitated the development and the evaluation of a wide range of tuberculosis specific diagnostic tests [http://apps.who.int/tdr/svc/diseases/tuberculosis/specimen-bank](http://apps.who.int/tdr/svc/diseases/tuberculosis/specimen-bank). Table 1 presents the characteristics of the main TB biobanking initiatives.

**From biobanks to personalized medicine**
Most of the discourse around personalized medicine is focused around cancer because of its higher prevalence in affluent countries. However, personalized medicine also applies to infectious diseases treatment, and is not limited simply to avoiding penicillin administration to allergic patients. Surrogate and “personalized” biomarkers are now needed to classify patients in genetic predisposition groups; to administer the most appropriate antibiotics for the most appropriate durations, to accurately and differentially diagnose primary active infections from chronic infections, re-infections, or reactivations; and to adjust vaccine administration to previous immune status and/or HLA background. In this perspective, TB represents a case where effective implementation of surrogate biomarkers could allow stratification of patients before starting chemotherapy, and thus have a significant impact on the global burden of the disease. Therapy could become more efficient, and resistance due to patient non-adherence could be diminished, and evaluation of novel anti-tuberculosis drugs could also be accelerated. Moreover, there is a need for effective personalized prognostic biomarkers in clinical trials, since the only unequivocal clinical end-point to date is bacteriological cure at the completion of therapy and absence of relapse for one to two years, which makes the duration and cost of such clinical trials extremely high.

Table 2 provides a summary of examples of research with potential personalized medicine applications. Predictive, prognostic and diagnostic studies have indicated multiplex signatures, based on either soluble markers or gene expression levels. Multiplex biosignatures are generally more promising in terms of diagnostic performance than single biomarkers. These recent studies have broadened our understanding in deciphering the disease pathogenesis and thereby paving the way for development of better diagnostics, therapeutics and vaccines. However, the introduction of these biomarkers in clinical practice is still pending reproduction of the initial results in validation population sets.

Finally, personalized approaches do not only concern the patient himself, but also the bacterial strains. Closely related clinical isolates of *M. tuberculosis* have been shown to vary as related to a significant number of their structural and functional proteins by iTRAC proteomics analysis. Genotypically, distinct phylogeny of *Mtb* has been described with specific geographic prevalence which has been implicated in the disease diversity. In this respect, TB represents a particular challenge, since different bacillary sub-populations, with different anatomical localisations and metabolic profiles can coexist in a patient and semi-dormant bacillary populations could be more resistant to antibiotics. Most of the currently
used anti-tubercular drugs act on the multiplying \textit{Mtb}, hence less efficient on the metabolically altered \textit{Mtb} sub-populations representing persisting or dormant bacilli.

Despite the important number of studies performed, very few biomarkers have succeeded in entering clinical practice. The example of interferon-gamma release assays (IGRA) from PBMCs, incubated with mycobacterial antigens, is borderline, since IFN-\(\gamma\) can hardly be considered as a “specific” biomarker. All of these studies are based on the analysis of human samples or microbial derivatives. Thus, the accurate characterization and the biomolecular quality of the samples used have a direct impact on the validity and applicability of the results. Yet, not all studies use samples from well-organized biobanks as either test or validation sets\(^26\). Moreover, as it always happens with retrospective studies, establishment of associations is not informative on whether the biomarker is a cause or a result of the corresponding clinical endpoint, and does not guarantee predictive value at the individual level. For instance, IFN-\(\gamma\) in sputum or sIL-2R in serum have both been independently associated to treatment response but their predictive value is unknown\(^27, 28\). Prospective studies, and thus prospective biobanking with high traceability of the follow-up of the sample donors, are required for this purpose.

**Critical factors in the catalysis of personalized medicine by biobanking**

**Types and processing of samples**

The type and processing of biospecimens may differ depending on the anticipated end-use. It is often difficult for a biobank to anticipate all the different future uses of the samples; therefore, the most stringent processing requirements should be followed for the biobank to be able to provide samples to as many researchers as possible.

Mycobacterial infection can be localized or systemic; therefore, the types of samples that can be collected during clinical practice include body secretions, exudates, tissue, or fluids. Using left-overs from clinical diagnostic procedures for biobanking purposes is generally not the optimal practice. Instead, specimens dedicated to biobanking and research should be collected at the same time with specimens for clinical diagnosis, but in separate containers and processed through separate standardized workflows. As usual, the devil is in the sample processing details and in each processing method’s critical steps.
Specimens intended for microbiological analyses:

Direct detection of Mycobacteria can be performed by smear microscopy or culture. Mycobacteria are relatively resistant to chemical agents because of the hydrophobic nature of the cell surface with high lipid content and their clumped growth. Therefore, tubercle bacilli may survive for long periods in dried sputum. Furthermore, the intracellular localization of mycobacteria in monocytes, reticuloendothelial cells, or giant cells facilitates microbial persistence. Specimens intended for direct diagnosis by acid-fast stains or culture from non-sterile sites, such as sputum, can be liquefied with 2% N-acetyl-L-cysteine, decontaminated with 1% NaOH, neutralized with buffer, and concentrated by centrifugation. However, best practice is to send unenriched specimens to a reference laboratory to increase the efficiency of strain identification. First morning sputum and urine samples are optimal for recovery of acid-fast bacteria. It is recommended to separately process at least two sputum samples during a two-day period; one on the spot sample and the second on a next morning sample (some countries use three samples; however, two samples identify 95% of smear-positive cases).

In cases of suspected renal tuberculosis, three consecutive first morning urine specimens should be submitted to the reference laboratory. Sputum and urine specimens should be processed within 2 hours from collection. If a delay of more than two hours is anticipated, urine and expectorate sputum should be refrigerated, whereas induced sputum should be kept at room temperature for a maximum of 24 hours. Since children are often unable to produce sputum, tracheal aspirates can be collected in pediatric populations. When a specimen is centrifuged to sediment the mycobacteria, the relative centrifugal force (RCF) applied is also a critical factor and should be at least 3000 x g for 15 minutes. If culture is performed, the selections of optimal media, inoculation, incubation conditions, and decontamination are all critical issues. Most gram-positive antimicrobials cannot be used, since they will also inhibit mycobacteria. Suspensions of positive primary cultures, adjusted to the McFarland No. 1 turbidity standard, can be stored at -70 °C ± 10 °C.

Blood samples to be used for identification of mycobacteria by today’s standard methods should be collected in either sodium polyanethosulfonate (SPS) or heparin-containing tubes, but not in EDTA or ACD tubes. The delay between collection and separation should not exceed 24 hr at room temperature. However, Nucleic Acid Amplification Tests (NAATs) for direct detection of mycobacteria in clinical specimens involve processing steps such as heat,
sonication, alkali and/or nonionic detergent treatment. Contrary to the above specifications, in this case, heparin collection tubes should be avoided, and EDTA tubes are preferred.

Specimens intended for non-microbiological analyses:

If samples are intended to be used for immunological, molecular biology, or proteomic analyses, critical in-vitro preanalytical steps should be accurately recorded for biological fluids or solid tissues collected. For biological fluids, this information includes type of primary collection tube, pre-centrifugation time delay and temperature, centrifugation conditions, post-centrifugation time delay and temperature, and long-term storage duration and temperature. For solid tissues, this descriptive information includes warm and cold ischemia times, type and duration of fixation, and long-term storage duration and conditions.

Finally, if metabolomics applications are anticipated, in-vivo preanalytical data, including the time of the day when the blood or urine samples were collected, medications, and food intake should also be collected and recorded in appropriate capability databases. A review of Standard Operating Procedures (SOPs) for different types of biospecimens being collected and processed in the context of tuberculosis biobanks is currently under preparation by the London School of Hygiene and Tropical Medicine (personal communication, G. Eckhoff, R. Mcnerney).

Characterization of samples

Accurate characterization of the samples supplied by a biobank concerns both the authentication and the integrity of the biomaterial. When a biobank supplies a serum from a patient with acute *M. tuberculosis* infection, this should indeed represent *M. tuberculosis*, true acute infection status (authenticity), and the serum sample should not have been compromised by any type of pre-analytical bias. Initial characterization of the bacterial strain, which is a derivative of the sputum sample, should be performed by a gCLP (Good Clinical Laboratory Practice), CLIA (Clinical Laboratory Improvement Amendments) or ISO15189 (Medical Laboratories - Particular Requirements for Quality and Competence)-compliant subcontracted laboratory, unless the biobank itself runs such a facility.

Method validation is an important aspect in all sample or derivative characterization assays. Method validation allows confidence that detected differences are due to different clinical populations, and not to different organism strains, different human genetic backgrounds, or different methodologies. For instance, with each analysis, and whenever new reagents are
introduced, a known positive and a known negative sample should be incorporated. In comparative studies of cytokine levels in culture supernatants, it may be important to verify through lot-testing that blood collection tubes do not contain any endotoxin which may induce cytokine production during subsequent culture of blood cells. This time, the devil is in the assay’s details and in each assay method’s critical steps.

Authentication:

If phenotypic methods are used for strain identification, growth rate, pigment production, microscopic morphology of culture, colony morphology and conventional biochemical tests should be documented. If genotypic methods are used, these may be based either on use of probes (in-solution hybridization, solid format reverse hybridization), on amplification followed by DNA sequencing, or on amplification followed by restriction enzyme analysis. A rapid identification test of *M. tuberculosis* is commercialized, based on immunochromatographic detection of a species-specific secreted antigen. This test can be used on cultures, but not directly on clinical specimens.

Integrity:

Little is known about the kinetics and stability of the different genomics and proteomics biomarkers assessed, but this knowledge is necessary for corresponding method validation. Even in the case of the IGRA commercialized assays, the stability of the positive responses over time is not well known. Once again, prospective studies and thus prospective biobanking with high traceability of the follow-up of the sample donors are required for validation.

There are actually no data available on the value of potential quality control tools allowing assessment of the collection procedures, shipping and storage conditions. However, homogeneity in these steps is key to the quality of multicenter research studies. In order to proceed to effective quality control of retrospective collections, biobank managers can proceed in different ways. Quality control can be performed on every specimen received at the biobank. In some instances, this is highly recommended and cost effective. For example, quality control of the sputum samples to assess how representative of the lower respiratory tract secretions they are, and therefore how suitable for culture applications, can be done by simple direct gram-negative staining and microscopic observation. A properly collected specimen should contain a minimum of gram-negative bacteria, squamous epithelial cells (signaling gross contamination with oropharyngeal flora), and significant numbers of
polymorphonuclear leucocytes. A good quality specimen should have less than 10 squamous cells per 100X field. In other instances, generalized quality control at biobank reception is not cost-effective; for example, if the intended use of the specimens is DNA extraction and analysis. In this case, quality control before distribution of samples to researchers may be performed (eg DNA concentration, purity, Taq amplifiability), provided such quality control is not destructive for the sample. Finally, retrospective quality control can always be performed. Here, two options are available: either apply quality control testing to a defined percentage of the collected specimens, selected on the basis of randomization, or apply quality control testing to the samples which are susceptible to having undergone the most “inconsistent” processing. The first approach allows comparisons between different collection sites, whereas the second approach allows targeted assessment of the “highest risk” samples.

The following Quality Control assays may be performed by either the biobank or the end-user who finally receives the samples or a sub-contracting laboratory. Quality control tests allowing more accurate characterization and ensuring more efficient downstream analyses include, but are not limited to, the following: (1) C-Reactive Protein (CRP) measurement in serum allows assessment of the inflammation degree and corresponding normalization of downstream proteomic analyses. (2) Albumin measurement in urine allows evaluation of the risk of adsorption of low concentration antigens in downstream immunoenzymatic assays. (3) *M. tuberculosis* mRNA measurement in sputum of culture positive patients allows evaluation of usability of the samples in RT PCR applications, but also assessment of bacterial viability. Quality control tests allowing assessment of shipping, processing and storage conditions include, but are not limited to, the following: (1). Serum IgM detection in culture positive patients allows assessment of the possible use of the serum samples in immunologic assays targeting specific IgM. (2) Serum sCD40L measurement allows assessment of the time the serum samples were exposed to ambient temperatures. (3) Haemoglobin measurement in serum or plasma allows assessment of haemolysis which may have taken place during collection or prolonged pre-centrifugation delays of blood samples. (4) Serum fingerprinting can be performed for assessment of the identity of different serum samples. (5) Microparticles can be counted in serum or plasma to assess centrifugation conditions and efficiency. (6) Selected platelet components can be measured in order to assess platelet activation during sample processing.
These considerations for method validation and quality control highlight how biobanking joins biospecimen research. For both targeted and whole genome, transcriptome, or proteome-derived biomarkers, biospecimen research allows assessment of the robustness of the biomarker relative to the pre-analytical variations, which are anticipated to take place during sample collection and processing in the real life. For instance, cytokines such as G-CSF, CXCL10, MIF, Serpin E1, CXCL12 have been shown to decrease with increasing freeze-thaw cycles, and careful attention should be paid in studies targeting cytokines to avoid any preanalytical bias.

One suggestion in biobanking for research on infectious diseases would be to systematically keep record of the most important preanalytical steps, as part of the biobank database. One possible way of doing this is using the Standard PReanalytical Code (SPREC) developed by the International Society for Biological and Environmental Repositories (ISBER).

**Biobanking, personalized medicine and public health objectives**

The physiology and transmissibility of *M. tuberculosis* have led to a tuberculous infection risk in the form of an iceberg (Figure 1). A visible level includes sputum-positive, contagious subjects, but an invisible high risk level includes both previously-infected subjects and uninfected subjects coming in contact with the bacterium. The donors of samples to biobanks represent the larger community, and generously consent to the use of their samples for research in the hope and the expectation that this research will finally result in their community’s health improvement. The protection and improvement of community health requires effective preventive medicine. In the case of tuberculosis, this corresponds to two levels of action: decrease of infection rate through association of diagnosis and treatment, and protection of non-contaminated people either by vaccination or by chemoprophylaxis.

The most effective public health means of prevention of future TB cases is rapid diagnosis followed by immediate and effective treatment. There may be a fundamental incompatibility between these public health prerogatives and personalized medicine. A diagnostic biomarker with no specificity for latent versus acute *M. tuberculosis* infection and with 70% positive predictive value in highly endemic countries may be judged as inappropriate in a personalized medicine approach in developed countries. Research in the *M. tuberculosis* area should primarily consider the emerging countries’ public health concerns and adapt the personalized medicine approach to the former.
Biobanking still has a crucial role to play in this respect. Research teams and diagnostic industry should not lose time in conducting evaluation studies on unreliable biomarkers. Unreliable biomarkers may have been identified by previous comparisons between cases and controls in which samples were collected, processed and stored in different conditions. Here, the difference between cases and controls may simply reflect pre-analytical differences instead of biological-clinical differences\textsuperscript{26}. High-level biobanking organization and traceability are required to avoid this loss of time and financial resources. Whereas such pre-analytical sensitivity is to be avoided, high-quality biobanking should not drive us into the trap of excessive standardization which could be detrimental to the public health requirements and induce incompatibility between personalized medicine and public health needs. Indeed, a diagnostic, prognostic, or predictive biomarker whose detection depends on very strict pre-analytical requirements (e.g., pre-centrifugation time of less than 30 min, and temperature of less than 10 °C), although potentially useful in clinical trials, would prove useless in everyday clinical practice, especially in emerging countries. Therefore, useful biomarkers in clinical practice will be those which have sufficient pre-analytical robustness allowing their largest possible implementation and use. High-level biobanking organization and traceability are again required because the variability in the sample processing and storage conditions should be recorded and effectively managed.

Thus, the biobanking discipline can drive personalized medicine to converge with public health considerations, considerations which are highly prevalent to the infectious diseases arena.

**Management of samples**

If the previously described technical and scientific considerations are critical to the fulfillment of the promise of biobanks to the public, certain organizational and logistical considerations are also absolutely essential\textsuperscript{36}.

The infective dose of *M. tuberculosis* is very low (ie, ID50<10 bacilli), thus **biosafety** aspects are particularly pertinent to mycobacteriology biobanking. Aerosol-producing manipulations of specimens must be performed under a Biological Safety Cabinet (BSC) at a Biosafety Level 3 (BSL-3) laboratory. Alternatively, mycobacteria can be rendered nonviable by addition of equal volume of 5% sodium hypochlorite to an aliquot of the specimen for 15 min, or by heating if the specimens are meant for DNA analysis.
Shipping of samples must follow international regulations, with biological substances being shipped in triple containers with a UN3373 label (“Biological Substance Category B”) placed on the outside of the outer container\textsuperscript{37}.

Geographical settings and different ethnic origins: In genetic case/control studies, it is important to compare groups of similar ethnic origins to avoid bias linked to different genetic backgrounds. Indeed, several important candidate genes like human leucocyte antigen/alleles and non-human leucocyte antigen genes, such as cytokines and their receptors, chemokines and their receptors, pattern recognition receptors (including toll-like receptors, mannose-binding lectin, and the dendritic cell-specific intercellular adhesion molecule-3 grabbing non-integrin), solute carrier family 11A member 1 (SLC11A), and purinergic P2X7 receptor gene polymorphisms, have been associated with differential susceptibility to TB in various ethnic populations. This heterogeneity has been explained by host–pathogen and gene–environment interactions and evolutionary selection pressures\textsuperscript{38}.

Data standardization and validation is also a critical issue. All clinical and laboratory data for the donor from whom the specimens are collected for the biobank need to be meticulously recorded and associated with the specimens. There are standard methods for reporting the average number of Acid Fast Bacilli (AFB) observed in clinical specimens, proposed by WHO and the CDC, based on the number of AFB found at 1000X per optical field and on the number of colonies observed in culture on solid media. Clinical data of the subjects in standardized and exchangeable formats are essential for meaningful interpretation of the laboratory/research results. Therefore, use of standards, such as the ones proposed by the Clinical Data Interchange Standards Consortium (CDISC, \url{http://www.cdisc.org/standards}) is highly recommended\textsuperscript{39}.

Apart from the biobank databases, which include all relevant data on sample types, numbers and availability, research databases including research results are being developed. The Tuberculosis Database (TBDB) is an integrated database providing access to TB comparative genomics and gene expression data from both \textit{in vitro} experiments and host infected tissues (\url{http://www.tbdb.org}). \textit{Mtb} bacterial genome sequences can also be found on BioHealth Base (\url{http://patricbrc.vbi.vti.edu/portal/portal/patric/IncumbentBRCs?page=bhb}), GenoMycDB (\url{http://157.86.176.108/~catanhogenomycdb/}), Tbrowse (\url{http://tbbrowse.osdd.net}), TB drug resistance associated mutations can be found on the Tuberculosis Drug Resistance Mutation database (\url{http://www.tbdreamdb.com}), while DNA and protein data can be found on
WebTB.org and importantly on the TubercuList (http://genolist.pasteur.fr/TubercuList/help/about.html). Finally, candidate therapeutic targets can be found on the TDR targets database (http://tdrtargets.org) and small molecule libraries of compounds already tested against *Mtb* can be found on the Collaborative Drug Discovery Tuberculosis Database (http://www.collaborativedrug.com)\(^40\). It can be argued that linking these public research databases to local databases with patients’ genomic or transcriptomic information, obtained from biobanked samples, will be the ultimate catalyst towards personalized medicine.

**Open access to biomaterials** is probably the most critical managerial aspect for biobanks to meet the public health requirements. If collected samples remain the “property” of a research scientist or team, with no possible distribution to third parties, this practice definitely minimizes their potential usefulness and the collection’s global return on investment. Providing access to biomaterials should be a *sine qua non* condition for a biobank to be considered and certified as such. Management of this activity can take into account possible benefit sharing considerations, by avoiding exclusive patent applications leading to monopoply situations, or by requiring non-exclusive licenses for the benefit of the local populations who have contributed to the research by donating their samples. Implementation of open access is not always the case in privately funded biobanks (Table 1).

**Legal and ethical issues** cannot be neglected. Informed consent is necessary and should not be restricted to a specific project but rather be open to all kinds of future testing of the samples, including genetic research in the domain of tuberculosis. The benefit-sharing issues need to be addressed upfront with clear operating procedures spelt out and including fair return of the goodwill gesture of the community involved in the voluntary donation of their precious body fluids, in terms of privileged access to the new tools or interventions at preferential terms and prices. TDR prioritized the collaboration with local teams engaged in the field of TB detection by contracting research teams for the specimen collection and characterization. The study protocols were cleared by the WHO and the national legal authorities and all the patients were asked to read and sign a translated inform consent prior collection. In addition to standardizing the collection of specimens, the collaboration resulted in the strengthening of the laboratory capacity for detection of tuberculosis through training and purchase of equipment(s). Conducted outside of clinical trials, the collection of specimens was meant to be a component of a strategy leading to mutual benefits between the collaborating centers, WHO and the potential users of the specimens. Concomitant capacity
building has empowered the community with adequate tools which enhance case detection rate for successful national TB control programs.

Certification of biobanks and Reference Materials

Few countries have elaborated regulatory schemas requiring national government-issued authorizations for biobanks. The objective of these systems of national authorizations is to ensure observance of biosafety rules, collection of informed consent and to enable funding allocation through authorization conditions. However, pharmaceutical and diagnostic regulatory agencies acting at either national or international levels have not promoted any kind of requirement in terms of biobank certification yet. Certification of biobanks to international standards such as ISO9001 (Quality Management Systems – Requirements) by an independent certification body is a proof of effective organization and management of the bank. Furthermore, subcontracting to testing laboratories, which are themselves accredited to international standards such as ISO17025 (General Requirements for the Competence of Testing and Calibration Laboratories) or ISO15189 (Medical Laboratories – Particular Requirements for Quality and Competence) by a national accreditation body, is a proof of reliable processes for sample characterization. For the moment, compliance to these standards, however important, essentially remains a voluntary approach of each individual biobank.

Biobanks are expected to provide the means for scientific replication of experiments. Therefore, they may not only be certified for the activities of collection, processing, characterization, preservation and procurement of biological materials, but may also certify these research materials. The American Type Culture Collection (ATCC) provides the first example of a biobank accredited to ISO Guide 34 as reference material producer. A reference material producer is able to provide materials to which he has assigned a quantitative or qualitative value, with a certain degree of confidence. Biobanks in the infectious diseases area can probably take the lead in this aspect, in relation to tumour banks, since characterization methods for the infectious agents is more amenable to accreditation than pathology.

Until now, regulatory agencies delivering authorizations for commercial use of in vitro diagnostic assays, have not expressed clear and explicit requirements relative to the nature and quality of the samples that have been used for assay validation. We suggest that, based on our previously described considerations on management, method validation and quality control, some infectious diseases biobanks could achieve the level of organization allowing
them to produce certified reference panels, such as reference panels, characterized as “*M. tuberculosis* acute” or “latent infection”, or “relapse of disease”. Ideally, evaluation of diagnostic tests should be performed by accredited independent laboratories, receiving the *In Vitro* Diagnostic (IVD) kits from the manufacturer and reference samples from a biobank. The Office of In Vitro Diagnostic Device Evaluation and Safety of the U.S. Food and Drug Administration as well as the CE marking accreditation agencies in Europe could engage into that direction.

**Conclusion**

The discovery of clinically relevant biomarkers can be applied to the development of *M. tuberculosis*-specific diagnostic tools. Both targeted and –omics analyses have identified different genotypes, gene expression signatures and proteins, linked to either bacterial persistence or to host immune responses. If such surrogate biomarkers could be brought into clinical practice, this effort would allow more efficient and timely diagnosis and efficient categorization of patients into specific therapeutic regimens. Such tools are necessary to objectively monitor in a personalized way both the risk of developing disease in exposed individuals and the progression of disease and the response to therapy in infected patients, as well as to develop effective vaccine-based interventions. There are likely important biomarkers of vaccine efficacy or drug effect that would be useful in the context of TB vaccine or drug clinical trials but not at all useful for clinical practice, due to differences in available infrastructure or cost considerations. However, public health concerns are of primary importance in infectious diseases. Therefore, identification and validation of these biomarkers should be conducted in a way ensuring their robustness and their applicability in the real everyday clinical practice. We show that well organized biobanks have a critical role to play in ensuring that both types of research can occur and the health improvement promises to the public will be kept. Because standardization and Quality Assurance are expensive undertakings, research funders should include biobanking in project designing, including that of clinical trials on drug effect or vaccine efficacy.

**Acknowledgements**

We are grateful to Elaine Gunter for critical reading of the manuscript.
References

11. http://www.genomeweb.com/print/947016


Figure 1

The TB “iceberg” including different categories of sample donors (ranging from the most apparent “infected, from clinical trials” to the less apparent, “household contacts”), with the corresponding types of specimens which can be collected by a biobank, and the corresponding types of biomarkers which can be discovered in the research laboratory.