



Biobanking and translation of human genetics and genomics for infectious diseases



Ivan Branković^{a,1}, Jelena Malogajski^{a,1}, Servaas A. Morré^{a,b,*}

^a Institute for Public Health Genomics, Department of Genetics and Cell Biology, Research School GROW, Maastricht University, Maastricht, The Netherlands

^b Laboratory of Immunogenetics, Department of Medical Microbiology and Infection Control, VU University Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 28 January 2014

Received in revised form 22 February 2014

Accepted 2 April 2014

Keywords:

Biobank

Infectious disease

Host genomics

HIV

HPV

Chlamydia trachomatis

ABSTRACT

Biobanks are invaluable resources in genomic research of both the infectious diseases and their hosts. This article examines the role of biobanks in basic research of infectious disease genomics, as well as the relevance and applicability of biobanks in the translation of impending knowledge and the clinical uptake of knowledge of infectious diseases. Our research identifies potential fields of interaction between infectious disease genomics and biobanks, in line with global trends in the integration of genome-based knowledge into clinical practice. It also examines various networks and biobanks that specialize in infectious diseases (including HIV, HPV and *Chlamydia trachomatis*), and provides examples of successful research and clinical uptake stemming from these biobanks. Finally, it outlines key issues with respect to data privacy in infectious disease genomics, as well as the utility of adequately designed and maintained electronic health records. We maintain that the public should be able to easily access a clear and detailed outline of regulations and procedures for sample and data utilization by academic or commercial investigators, and also should be able to understand the precise roles of relevant governing bodies. This would ultimately facilitate uptake by researchers and clinics. As a result of the efforts and resources invested by several networks and consortia, there is an increasing awareness of the prospective uses of biobanks in advancing infectious disease genomic research, diagnostics and their clinical management.

© 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Over time, the definition of a “biobank” has moved away from an early view of a biobank as population-based to include a wider typology of biobanks that we find in the literature today. In a survey conducted in 2012, researchers involved in managing sample collections were asked about their definition of biobanks. The results of the survey showed the consensus among respondents that the term biobank may be applied to biological collections of human, animal, plant or microbial samples. Additionally, the term biobank should only be applied to sample collections with associated sample data, and to collections that are managed according to professional standards (Hewitt and Watson, 2013).

In post-Human Genome Project research, the role of biobanking as a component of research infrastructure is broadening, as knowledge

from biobanks contributed to the understanding of the etiology of multifactorial diseases caused by both mutations in a variety of genes and the influence of environmental factors and lifestyle (Brand and Probst-Hensch, 2007; Knoppers et al., 2012). Furthermore, biobanks have paved the way for the evolution of personalized medicine, especially the development of “tailored” drugs (Gottweis and Zatloukal, 2007). In recent years, integration, analysis and interpretation of data originating from biobanks have begun to play a growing role in our understanding of genetic susceptibilities of infectious diseases. The actuality of infectious diseases and the perpetual challenge they pose for researchers and physicians are reflected in both the high prevalence and the high mortality of existing and growing incidence of emerging infectious diseases (Jones et al., 2008). In fact, infectious diseases represent a major health threat worldwide, and are a particularly significant burden to developing countries (Frodsham and Hill, 2004).

The importance of genetic factors in the pathogenesis of infectious diseases has transformed our understanding of such diseases by incorporating host genetic determinants that modulate immune responses as factors of pathogenesis. We now understand that host responses can determine the outcome of an infection as much as – if not more than – the properties of the pathogen itself (Peng et al., 2009). Genomics outlined molecular biomarkers and pathways as targets for diagnosis or

* Corresponding author at: Institute for Public Health Genomics, Department of Genetics and Cell Biology, Research Institute GROW (School for Oncology & Developmental Biology), Faculty of Health, Medicine & Life Sciences, University of Maastricht, Universiteitssingel 50, 6229 ER, Location UNS.50-H1.330, Maastricht, The Netherlands (Thursday & Friday). Tel.: +31 20 44 49375.

E-mail address: samorretravel@yahoo.co.uk (S.A. Morré).

¹ These authors contributed equally to the manuscript

intervention in the field of infectious diseases (Hill, 2006). Relations between genetic factors and susceptibility to the course and the outcome of infectious diseases are predominantly studied through candidate genes, genome wide associations, and twin studies (Hill, 2001). This research means that biobanks (especially large international networks of biobanks driven by the needs of researchers, who require large collections of samples) are an imperative infrastructure for research in host genomics (Meijer et al., 2012).

According to Gottweiss and Zatloukal (Gottweiss and Zatloukal, 2007), there are four main types of biobanks:

- (1) *clinical case/control biobanks*, which contain biological samples taken from patients with specific diseases and from healthy control patients
- (2) *population-based biobanks*, which contain samples from smaller or larger subsets of a population with or without a certain disease;
- (3) *population isolate biobanks*, which contain homogenous genetic material of the population represented; and
- (4) *twin registries*, which contain samples from monozygotic and dizygotic twins.

Biobanks contain both samples and data; this twofold nature is the root of much of the legal and ethical controversy surrounding biobanks today. Issues concerning privacy health-related information, informed consent, secondary use of samples, and harmonization of legislation and networking of biobanks are often researched in conjunction with the term “biobanking” (Townend, 2012). The potential impact of biobank-generated knowledge (and its becoming an integral part of public health policies) on our understanding of the etiology of disease, on improving diagnosis and treatment, and ultimately on the health of individuals and populations as a whole has been largely ignored thus far (Knoppers et al., 2012).

Uses of biobanks for public health are (Brand and Probst-Hensch, 2007):

- timely, responsible and effective integration of genome-based health technologies and information into health research, policy and practice;
- supporting the translational process from basic knowledge generated in existing biobanks to the development and implementation of health policies, interventions and programs;
- recognizing the multi-tasking nature of biobanks in the accommodation of different needs by enhancing the ability of biobanks to serve researchers and other relevant stakeholders with particular public health perspectives.

In 1990, Lee identified what he considered to be the ideal properties of a biospecimen bank: a secure, ongoing source of funding; a cryogenic storage facility; selection criteria for obtaining and keeping the best samples in storage; and ensuring the continuation of research to optimize the collection and handling of samples (Lee, 1990). De Paoli (De Paoli, 2005) also identified what he considered to be the main roles of biobanking in microbiology research:

- to enable unfettered epidemiological research: prospective use of biobanking is key to detecting and tracking different strains, comparing new strains with previously stored ones, determining modes of transmission, and ultimately combating infections
- to ensure progress in diagnostics: by comparing samples taken from the same subjects over time or by comparing samples taken from different subjects at the same point in time, or by applying novel diagnostic tools to the analysis of samples that exhibit increased sensitivity and specificity.
- to manage studies with large sample sizes: this may refer to research with sample collections coming from different geographical locations, or research that is conducted in several remote laboratories.
- to establish biorepositories with characterized host cell lines: cell

lines can be used for research, diagnostics and quality control, and other scientific pursuits.

- to assist in building a microbial tree of life: such biobanks provide the basis for mapping out microbial diversity and evolution. Given the increasingly imminent threat that emerging highly virulent or therapy-resistant strains pose for global health, the importance of these types of biobank collections will likely rise in the near future.

In addition to these roles, biobanks can serve as the foundation for conducting research in host genomics and other ‘omic’ sciences, elucidating the role and interactions of the host’s immunogenetic factors in infections (Ballana et al., 2012), as well as driving prospective diagnostic and therapeutic advances (Haralambieva and Poland, 2010).

This article examines the role of biobanks in the basic research in infectious disease genomics, and also observes the relevance and applicability of biobanks in both the translation of impending knowledge and the clinical uptake of biobank-generated knowledge in infectious diseases. Our research identifies potential fields of interaction between infectious disease genomics and biobanks, in line with the global trend of integration of genome-based knowledge into clinical practice.

2. Materials and methods

A literature search was performed on the identification of the existing links between biobanking and infectious diseases; on points of potential collaboration between biobanks, clinics and surveillance agencies; and on the examination of the relevance of electronic health records (EHR) in genomic research of infectious diseases. The study focused on examples of the translation of biobank-generated genome-based knowledge to everyday clinical practice. Databases (PubMed, Cochrane library, Google Scholar), electronic journal collections (Maastricht University EJ collection) and the websites of relevant organizations, networks and consortia (OECD, EAPM, P³G, BBMRI) were searched for appropriate references. The terms used in the searches were [“biobank*” AND (“infectious disease*” OR “genomic*”). Retrieved articles were further selected based on relevance. Additional search terms were “public health”, “data management” and “data privacy”.

This article provides examples of existing biobanks with substantial resources for infectious disease research, such as those for human immunodeficiency virus (HIV), Human papillomavirus (HPV), and *Chlamydia trachomatis* (CT).

3. Results

3.1. Human versus microbial sample biobanks

There is an obvious delineation between samples taken from human individuals suffering from a pathological condition related to an infection (who may or may not be infected with or are carriers of the pathogen) and samples of the infectious agent itself. The former is primarily relevant to host ‘omic’ research, as it provides the material basis for both candidate gene/SNP-approaches and genome-wide association research in seeking (co)morbidity associations, and also investigates the pathogen interactions with host proteomes (Zhang et al., 2010). These samples should be paired with relevant categories of patient phenotypic data. Pathogen samples enable epidemiological studies, investigate genetic strains of that species, and develop better diagnostic tools and novel therapies (De Paoli, 2005). Biomaterial samples may differ in processing and storage, as well as in shelf-life. As biomedical science moves away from the deconstruction of living systems and turns towards a more integrative, all-encompassing approach (through the likes of systems biology) (Khoury et al., 2007), it is reasonable to expect that researchers increasingly begin to assess host ‘omic’ data together with infectious agents’ ‘omic’ profiles. As a prerequisite, adequate samples should always be accompanied by relevant data.

3.2. Infectious disease genomics and the inflow of data

Advances in sequencing technologies have resulted in a relentless influx of data that need to be interpreted. Currently, researchers are generating data more rapidly than can be analyzed. In particular, the genetic variability of bacteria accumulated through evolution is enormous and significantly increases the volume of datasets. Public health genomics specifically emphasizes the need to examine all 'omics' (Khoury et al., 2007; Mardis, 2009), which, in the case of infectious disease, involves both the host and the pathogen. This emphasis dramatically increases demand for the effective deciphering of large amounts of data. There are, however, efforts among members of the microbiology community to develop strategies that would make this data manageable. For example, by grouping select loci of the pathogenic bacterial strain into schemes (the so-called *gene-by-gene approach*) and by implementing themes schemes in conjunction with conventional (sequence-based) schemes in adequate database platforms, data that is obtained through different studies and can be used more effectively in combined analyses (Maiden et al., 2013). In this way, "genotype summaries" of selected genes could be linked to phylogenetic relationships and functional characteristics of bacteria, thereby helping researchers navigate vast bacterial genomic diversity.

3.3. Existing infectious disease biobanks and networks

The number of laboratories that are creating their own biobanks is difficult to quantify with full precision. However, the number of organizations that are developing nation-wide or transnational collections and are building large consortia and networks is ever increasing (De Paoli, 2005). Large networks such as the Public Population Project in Genomics and Society (P³G) (<http://p3g.org/>) and Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) (<http://bbmri.eu/>) are changing the landscape of international collaboration in biobanking, harmonizing regulatory legislation and thus facilitating the use of biobanks in research (Wichmann et al., 2011). Promising nation-wide models have also emerged, including the United Kingdom Biobank (<http://www.ukbiobank.ac.uk/>), the Swedish National Biobank programme (now the Swedish arm of BBMRI, <http://www.bbmri.se/>), the Iceland Biobank, and others. Some biobanks, however, are not positioned as public domain entities; in the case of Iceland, for example, the biobank is a private company that has been given a commercial data license (Mitchell, 2010).

This article presents several infectious disease biobanks and their effective usage, which has led to cases of successful clinical uptake (or its near-future prospects) for HIV, CT and HPV (for which there is a more substantial body of literature available). These biobanks are founded and governed by hospitals and academic institutions and provide clear examples of how biobanking can stimulate research in infectious disease genomics.

3.3.1. HIV

The Infectious Diseases Biobank (IDB) at King's College London is an oft-cited example of an infectious disease-oriented biobank (Kozlakidis et al., 2012; Towie, 2007; Williams et al., 2009). The IDB is actively collecting samples from patients infected with HIV, hepatitis B and C viruses, and invasive bacteraemias (such as the methicillin resistant *Staphylococcus aureus* (MRSA)). The IDB is also collecting samples from healthy control subjects. The number of HIV patients with archived materials in the IDB is steadily increasing, resulting in the availability of distinct patient cohorts (in meaningful numbers) to researchers. Data on the IDB's website indicate that by September 2010, HIV sample donations had reached 500 annually. Examples of research stemming from this collection are (as stated on the IDB website): the roles of vpu gene and tetherin in HIV/AIDS pathogenesis; gene expression signatures in *in vivo* and *in vitro* HIV-1 infection; non-infectious HIV co-morbidities; renal function and bone homeostasis in patients

on HAART; the definition of CD161 + CD8 + T cell subset function in HIV infection and their response to therapy; the effect of Maraviroc on microbial translocation in HIV infected individuals receiving antiretroviral therapy; and the metabolic impact of Darunavir/ritonavir maintenance monotherapy after successful viral suppression with standard Atripla in HIV-1-infected patients. Aside from archiving biological samples, the BioBank runs a database in which the following clinical information on HIV donors is archived: histories of CD4 + cell numbers and plasma viral loads; last known HIV negative result and first positive HIV test; birth date; ethnicity; gender; HAART administering; and other complications. The database also features sample processing information (dates and times of venepuncture, processing and freezing); and details on aliquots that have been stored or transferred to researchers.

Another example of an infectious disease-oriented biobank is the Spanish HIV BioBank (Garcia-Merino et al., 2009). The primary objective of this biobank is to further scientific knowledge about HIV infection by providing biological samples from HIV-infected patients that are included in cohorts for the objective of carrying out research. The HIV BioBank receives samples from 28 hospitals, spread across Spain, which are grouped into 6 cohorts of HIV-patients, each with defined characteristics. Any member of the AIDS Network, or any party to a relevant collaboration with a member can apply for samples. Sample release applications are evaluated by members of the Scientific Committee. If the project is approved, the researcher signs a Release Agreement with the director of the BioBank and with the coordinator of the Cohort. The BioBank and the Cohort are responsible for locating the type and number of samples needed to carry out the project. Once a year, after the samples have been released, the principal researcher sends a scientific report to the BioBank containing his or her results, such that the BioBank can maintain up-to-date records on all projects.

3.3.2. Chlamydia trachomatis

Partners of the EpiGenChlamydia Consortium (urogenital and ocular CT infections), coordinated by the London School of Hygiene and Tropical Medicine (by David Mabey and Robin Bailey and their Gambian partners) who are researching ocular Chlamydia-related conditions, have already defined and secured 1500 case-control pairs ($n = 3000$). More than 4000 specimens that are currently in use have been collected by Dutch partners, and 10,000 specimens are available for further studies (Morre et al., 2009). One goal was to build a biobank and data warehouse—a biomedical, ethically-developed and run central sample collection and data management system. The Consortium is investigating possibilities for conducting genetic and epidemiologic Chlamydia research with samples from existing biobanks in Northern European countries.

The Consortium also aims to structure trans-national research to such a degree that comparative genomics and genetic epidemiology can be performed on large numbers of unrelated individuals. The most pivotal deliverables of this project were biobanking and data-warehouse building. These deliverables will allow for continuous generation of scientific knowledge on CT–host interaction genetic predisposition to CT infection, and the development of tools for early detection of this predisposition. The study of sequence variation (mainly SNPs) is a technique employed by different consortium partners to gain insight into the differences in clinical courses of infection, in order to identify genetic markers for susceptibility.

A review by Malogajski et al. (Malogajski et al., 2013) gives an overview of immunogenetic factors that have a demonstrable effect on human susceptibility to and the severity of CT. These immunogenetic factors are alleles (determined by specific SNPs) of pathogen recognition receptor genes. Women carrying one or a combination of these alleles are at higher risk of contracting CT, or are at significantly higher risk of developing subfertility-related complications, such as tubal pathology. The review proposes the development of novel diagnostic tools for assessing individual risk faced by CT-positive women. Currently, clinicians employ CT IgG serology when assessing these risks (Broeze

et al., 2011). Due to limited sensitivity and specificity of CT serology, the predictive value of this serology is weak, and as a result, many physicians recommend that women undergo additional invasive, stressful, and costly diagnostic procedures. It is estimated that 40–45% of women undergoing laparoscopy do not have tubal pathology. Additionally, false negative serology results account for about 20% of women whose tubal pathology will not be properly and timely diagnosed (Lal et al., 2013). The proposed tool would introduce a diagnostic approach based on a combination of two factors: a predictive SNP load; and serological markers for CT infection. On-going research aims to validate this set of SNPs and a subsequent cut-off score for diagnostic purpose. This tool would be the first to use a genetic trait in the diagnosis of infectious diseases severity in the triage of women.

3.3.3. HPV

Similar to the translation of CT biobank-derived data into diagnostic applications for subfertility, the translation of HPV research results should contribute to better diagnostics of cervical cancer and its pre-neoplastic stages, cervical intraepithelial neoplasia (Malogajski et al., 2013). Large biobanks and patient cohorts are used to achieve this result. Despite the anticipated outcomes of HPV vaccination (which should lead to a drop in cervical cancer incidence in a matter of decades), the needs of generations of women who were above the age of expected exposure to HPV virus (and were therefore left out of vaccination programs) ought to be addressed. The cervical scraping cytology assessment, better known as the PAP test, is routinely used throughout the world as a screening tool for cervical lesions; however, the PAP has low sensitivity. The introduction of high-risk HPV (hrHPV) assessment will increase sensitivity and ease (especially in the case of self-collected vaginal swabs) to determine a woman's risk of developing cervical cancer. Referral to a gynecologist is only needed where a woman is found to be infected by a hrHPV type. Some authors propose the development of a triage tool for high-risk HPV positive women based on methylation markers; this means that a woman should only be referred for further examination if she tested positive for one or more hrHPV types and at the same time carries a combination of methylation markers indicative of a progression of pre-neoplastic stages. This approach builds upon a number of studies that show how epigenetic alterations become increasingly present with each successive stage of cervical lesion and cervical cancer, mainly in genes that are important to the progression of cancer (e.g. tumor suppressors and cell adhesion molecules), or genes coding microRNAs, whose role is to bind to the viral nucleic sequences, thereby making them inaccessible to enzymes that are replicating or transcribing. Based on the available studies at that time, the review (Malogajski et al., 2013) highlighted methylation patterns in MAL and CADM1 genes as optimal markers for the development of a potential triage test (Overmeer et al., 2011). A more recent review by Litjens et al. (2013) reached the same conclusion, but added p16^{INK-4a}/Ki-67 dual immunostaining and viral integration to the proposed set of markers.

3.4. Relevance of electronic health records

The rise in usage – and usability – of electronic health records (EHRs) is a demonstrably promising catalyst in the efforts to better utilize and standardize biobanks (Kohane, 2011). This pertains to handling the information on the biomaterial from the large cohort studies on a plethora of diseases, including infectious diseases. 'EHR-driven genomic research' can ideally be achieved using two distinct workflows. Firstly, patients with a particular infectious disease or related sequelae could be selected from the EHR by using language processing tools, such as natural language processing (NLP) tools. In this case, the selected population could thereupon be recruited, either for the purpose of providing samples for genomic research, or in order to verify whether residual samples taken on previous occasions could be utilized. Secondly, EHRs can be used to broaden and advance clinical characterization by adding

new relevant data to the files of those individuals whose samples are already stored in another biobank or have been used in the context of a cohort study (Kohane, 2011). Electronic systems for the automated detection of notifiable diseases have, in fact, been tested using EHRs. In past decades, the term of preference was electronic medical records (EMRs). This term has gradually been overtaken by the previously mentioned EHRs, as focus slowly expanded from the inclusion of basic clinical patient data to the provision of a more complete insight into their health background and care.

The so-called ESP (Electronic Medical Record Support for Public Health) algorithmic system, which has been tested on Chlamydia records and others, assists not only in the identification and reporting of cases of notifiable disease, but also in the advancement of public health. Prospective applications of EHRs include syndromic surveillance; clinical decision support; the construction of vaccine registries; and the assessment of areas with higher prevalence of disease (Klompas et al., 2007). The incorporation of patient genome-based information into EHRs would act as a major driving force for genomic medicine. It would enable the investigation of potential comorbidities of genomic associations (Kohane, 2011), and would elucidate the ways in which such associations can individually or synergistically result in increased susceptibility to or severity of infectious disease. That said, the incorporation of patient genome information into EHRs has thus far been a daunting task, since most EHR systems are not designed to include genomic data (Kullo et al., 2013). Although the linear DNA sequence is simple by nature, the sheer volume of data and the complexity of relations among the 'functional components' of DNA are significant hurdles in attempting to devise an EHR system using this information (Kullo et al., 2013; Masys et al., 2012).

3.5. Data privacy and infectious disease genomics

While the issues of data privacy and consent exceed the parameters of this article, we will briefly lay out the state of the art in this field, as well as how legal frameworks, governance of infectious disease biobanks and the handling of sensitive data can affect not only patient rights, but also biomedical research generally. In 2013, during the Irish presidency of the EU Council, the European Alliance for Personalised Medicine (EAPM) hosted a conference on innovation and patient access to personalized medicine, in which experts discussed recent advances in healthcare and formulated conclusions relating to these advances (EAPM, 2013). In order to optimize data security and to facilitate access and consent (which would allow for re-use and secondary use of data), it was concluded that robust legal regulation of personal data in scientific research should be implemented. Moreover, cross-border transfers of data for the purposes of scientific research should be stimulated in cases where such privacy instruments have been deployed. It is important to note that, in harmonizing different systems of governance, a balance must always be struck between the stimulation of cross-border transfers of data and individual rights to privacy.

The importance of data protection cannot be underestimated, especially in the handling of samples taken from individuals who are afflicted with serious infections. Genomic discoveries concerning such infections potentially create various forms of discrimination in the context of future discovery. For example, research has shown that African-American carriers of a polymorphism conferring a Duffy antigen-negative phenotype, DARC -46C/C, are resistant to malaria (*Plasmodium vivax* infection). However, subsequent research into this polymorphism also revealed that carriers have a 40% increased likelihood of becoming infected with the HIV-1 virus (He et al., 2008). In this case, contrary to the protective character of CCR5-Δ32 deletion as witnessed in European populations, the disruption of the expression of a functional receptor is a major disadvantage to the carrier. Evidently, the risks of stigmatization and discrimination arising from genome-based information (the disclosure of a patient's illness or infection status being a potential infringement of patient rights) cannot be ignored.

As stated in the EAPM report conclusions (EAPM, 2013), progress must be achieved by developing trust between researchers and the public, and by promoting the equal treatment of health research data (including genome-based information) and the removal of silos for single-use data. Since this information is theoretically unlimited in terms of longevity, robust data protection mechanisms must be in place for periods longer than the samples' shelf-life (Heeney et al., 2011).

At the same time, different sets of mechanisms are needed in parallel with vigilant data protection. Genome-based research necessitates large sample sizes in order to arrive at more reliable results; as a result, overly-restrictive data protection policies can impede research and innovation (Masys et al., 2012). A large number of samples is necessary to identify patient subgroups of interest. There are also calls for the provision of research data to the general public, particularly in cases where the research itself was funded using taxpayer dollars (Church et al., 2009). Entire human genome sequences, including those of several prominent researchers, have already been made available to the general public online (open access model). In spite of this open access, the debate over balancing the right to consent versus the right to privacy is far from resolved. The fact that fewer than 13–15 genomic locations with variable repeats (or 30–80 statistically independent SNPs) can be used to identify any one individual (Lin et al., 2004) puts requests for the deregulation of data sharing into a new perspective. Samples that contain a 'genomic fingerprint' in combination with data relating to the presence of serious infections pose a new threat to those safeguards that ensure participant anonymity and prevent partial treatment. Due to lack of funding, many academic institutions allow private organizations to handle their genomic databases; as a result, the protection of the rights of human participants may be at risk in any future commercial uses of data (De Paoli, 2005).

4. Discussion

The aim of this review was to explore empirical evidence on the role of biobanking in infectious disease genomics and to outline the pertinent issues in setting up and utilizing biobank materials. We note that published material that provides a detailed overview of existing infectious disease biobanks and their uses to date is lacking. In order to facilitate extensive collaboration with researchers and to ensure the continuation of research on infection, infectious disease biobanks must become more visible, and must emphasize their societal impact. Thus far, the authors have encountered underrepresentation of infectious disease biobanks in publications and insufficient information on official websites. The public should be able to easily access a clear, detailed outline of regulations and procedures for sample and data utilization by academic or commercial investigators, as well as an account of the precise roles of governing bodies. Examples of procedural transparency and extensive online visibility include the King's College Infectious Disease BioBank and the Spanish HIV BioBank (Garcia-Merino et al., 2009; Williams et al., 2009). Appropriate regulation should precede the effective translation of biobank-based research to clinical settings; such regulation necessitates intensive efforts, so as to ensure rapid clinical uptake.

In recent years, several biobanking consortia and extensive networks have been formed, and there has been a visible increase in efforts, stakeholders' involvement and resources allocated (Wichmann et al., 2011). Nevertheless, infectious disease biobanks have yet to achieve their full potential. This review does, however, provide several examples of biobanks that have successfully contributed to the translation of data to clinics and patients. In order to successfully utilize biobank information in research on infectious disease, and in order to develop 'tailored' therapies based on pharmacogenomics research, adequate representation of ethnic minorities and neglected populations in biobanking is of paramount importance. Biobanks must be constructed to account for ethnic differences in susceptibility to certain infectious diseases, which themselves have been extensively documented (Dolo et al., 2005; Velez

et al., 2010). In HIV-AIDS therapy research, for example, extrapolations of potential clinical implications of allele frequency differences between different ethnicities could significantly assist doctors when prescribing therapies. Consortium for the BioBank and Pharmacogenetics database of African populations is an example of efforts paving the way for individualized treatments for HIV-AIDS (Matimba et al., 2008). The Consortium's biobank of anonymous samples was used to determine baseline frequency distribution of SNPs of genes affecting drug metabolism; this usage enabled the establishment of a pharmacogenetics database. Certain information can be essential for optimizing therapeutic approaches and reducing ethnic-specific adverse reactions, such as the different drug-metabolizing capacities of particular allelic versions of enzymes (for instance CYP2B6*6 allele) (Matimba et al., 2008). There is an argument to be made, however, that these differences are neither inherent nor applicable to all infections. Some authors argue, in terms of decreased precision of data analysis, against 'blind social inclusivity' in biobank sampling at the *potential* expense of 'analytical acuity.' (Smart et al., 2008). In countries such as the UK or the US, acting more fervently upon the two aforementioned views could lead to a reevaluation of the manner in which biobanks are governed.

EHR system designers need to be encouraged to configure these systems so as to enable the incorporation of genome-based (or 'omic'-based) information. The addition of pathogen 'omic' data to an accompanying registry should be made feasible in order to promote research in infectious diseases. Other forms of research, the clinical uptake of genome-based knowledge, and the advancement of personalized medicine can all invaluablely benefit from a shifting approach to health record management. Several approaches have been proposed to this effect, and each acknowledges the unique nature of genomic data (Jing et al., 2012; Masys et al., 2012).

The unique challenges associated with biobank-based research indicate that it is in some aspects more complex than other types of health or biomedical research. One of the main obstacles to translating biobank data into the clinical setting is confidentiality and privacy, which stem from a necessary pairing of biobank information with personal and unrelated types of health information. The protection of data obtained from samples of patients who are afflicted with serious infections is of particular importance due to the potential in such cases for discrimination. Discrimination can result both from current interpretations of data, and from future research and upcoming innovations in genomic technologies.

4.1. Conclusions

A clear overview of the usage of existing infectious disease biobanks is lacking in present literature, and we maintain that this information should be readily accessible to the public, along with clear regulatory and procedural guidelines for utilization of samples and data. This would ultimately facilitate the currently insufficient uptake by researchers and clinics. Several biobanks have, however, already set high standards in terms of instating appropriate regulation as well as enabling successful translation into clinical setting and can therefore serve as a model to other biobanks. In recent years, efforts and resources that have been invested in biobanking networks and consortia have surged. As a result, there is a higher awareness of the multitude of ways in which biobanking can advance basic research, diagnostics and – most importantly – the clinical management of infectious disease. These advances will ensure that research in biobank-based infectious disease continues to progress.

References

- Ballana, E., Gonzalo, E., Grau, E., Iribarren, J.A., Clotet, B., Este, J.A., 2012. Rare LEDGF/p75 genetic variants in white long-term nonprogressor HIV+ individuals. *AIDS* 26, 527–528.
- Brand, A.M., Probst-Hensch, N.M., 2007. Biobanking for epidemiological research and public health. *Pathobiology* 74, 227–238.

- Broeze, K.A., Opmeer, B.C., Coppus, S.F., Van Geloven, N., Alves, M.F., Anestad, G., Bhattacharya, S., Allan, J., Guerra-Infante, M.F., Den Hartog, J.E., Land, J.A., Idahl, A., Van der Linden, P.J., Mouton, J.W., Ng, E.H., Van der Steeg, J.W., Steures, P., Svenstrup, H.F., Tiitinen, A., Toye, B., Van der Veen, F., Mol, B.W., 2011. Chlamydia antibody testing and diagnosing tubal pathology in subfertile women: an individual patient data meta-analysis. *Hum. Reprod. Update* 17, 301–310.
- Church, G., Heeney, C., Hawkins, N., de Vries, J., Boddington, P., Kaye, J., Bobrow, M., Weir, B., 2009. Public access to genome-wide data: five views on balancing research with privacy and protection. *PLoS Genet.* 5, e1000665.
- De Paoli, P., 2005. Bio-banking in microbiology: from sample collection to epidemiology, diagnosis and research. *FEMS Microbiol. Rev.* 29, 897–910.
- Dolo, A., Modiano, D., Maiga, B., Daou, M., Dolo, G., Guindo, H., BA, M., Maiga, H., Coulibaly, D., Perlman, H., Blomberg, M.T., Toure, Y.T., Coluzzi, M., Doumbo, O., 2005. Difference in susceptibility to malaria between two sympatric ethnic groups in Mali. *Am.J.Trop. Med. Hyg.* 72, 243–248.
- European Alliance for Personalised Medicine (EAPM), 2013. Innovation and Patient Access to Personalised Medicine: Report from Irish Presidency Conference March 20th/21st 2013. Brussels <http://euapm.eu/wp-content/uploads/2012/07/EAPM-REPORT-on-Innovation-and-Patient-Access-to-Personalised-Medicine.pdf>.
- Frodsham, A.J., Hill, A.V., 2004. Genetics of infectious diseases. *Hum. Mol. Genet.* 13 (Spec No 2), R187–R194.
- Garcia-Merino, I., de Las Cuevas, N., Jimenez, J.L., Gallego, J., Gomez, C., Prieto, C., Serramia, M.J., Lorente, R., Munoz-Fernandez, M.A., 2009. The Spanish HIV Biobank: a model of cooperative HIV research. *Retrovirology* 6, 27.
- Gottweis, H., Zatloukal, K., 2007. Biobank governance: trends and perspectives. *Pathobiology* 74, 206–211.
- Haralambieva, I.H., Poland, G.A., 2010. Vaccinomics, predictive vaccinology and the future of vaccine development. *Future Microbiol.* 5, 1757–1760.
- He, W., Neil, S., Kulkarni, H., Wright, E., Agan, B.K., Marconi, V.C., Dolan, M.J., Weiss, R.A., Ahuja, S.K., 2008. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. *Cell Host Microbe* 4, 52–62.
- Heeney, C., Hawkins, N., de Vries, J., Boddington, P., Kaye, J., 2011. Assessing the privacy risks of data sharing in genomics. *Public Health Genomics* 14, 17–25.
- Hewitt, R., Watson, P., 2013. Defining biobank. *Biopreserv. Biobanking* 11, 309–315.
- Hill, A.V., 2001. Immunogenetics and genomics. *Lancet* 357, 2037–2041.
- Hill, A.V., 2006. Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet.* 40, 469–486.
- Jing, X., Kay, S., Marley, T., Hardiker, N.R., Cimino, J.J., 2012. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an ehr prototype based on the continuity of care record standard. *J. Biomed. Inform.* 45, 82–92.
- Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. *Nature* 451, 990–993.
- Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A., Bradley, L., 2007. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* 9, 665–674.
- Klompas, M., Lazarus, R., Daniel, J., Haney, G., Campion, F.X., Kruskal, B., Platt, R., 2007. Electronic medical record support for public health (ESP): automated detection and reporting of statutory notifiable diseases to public health authorities. *Adv. Dev. Surveill.* 3, 1–5.
- Knoppers, B.M., Zawati, M.H., Kirby, E.S., 2012. Sampling populations of humans across the world: ELSI issues. *Annu. Rev. Genomics Hum. Genet.* 13, 395–413.
- Kohane, I.S., 2011. Using electronic health records to drive discovery in disease genomics. *Nat. Rev. Genet.* 12, 417–428.
- Kozlakidis, Z., Cason, R.J., Mant, C., Cason, J., 2012. Ethical and legal considerations in human biobanking: experience of the infectious diseases biobank at King's College London, UK. *Biomedical Science, Engineering and Technology* pp. 761–778.
- Kullo, I.J., Jarvik, G.P., Manolio, T.A., Williams, M.S., Roden, D.M., 2013. Leveraging the electronic health record to implement genomic medicine. *Genet. Med.* 15, 270–271.
- Lal, J.A., Malogajski, J., Verweij, S.P., de Boer, P., Ambrosino, E., Brand, A., Ouburg, S., Morre, S.A., 2013. Chlamydia trachomatis infections and subfertility: opportunities to translate host pathogen genomic data into public health. *Public Health Genomics* 16, 50–61.
- Lee, R.E., 1990. Environmental specimen banking: a complement to environmental monitoring. *Biol. Trace Elem. Res.* 26–27, 321–327.
- Lin, Z., Owen, A.B., Altman, R.B., 2004. Genetics. Genomic research and human subject privacy. *Science* 305, 183.
- Litjens, R.J., Theelen, W., van de Pas, Y., Ossel, J., Reijans, M., Simons, G., Speel, E.J., Slangen, B.F., Ramaekers, F.C., Kruitwagen, R.F., Hopman, A.H., 2013. Use of the HPV MLPA assay in cervical cytology for the prediction of high grade lesions. *J. Med. Virol.* 85, 1386–1393.
- Maiden, M.C., van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat. Rev. Microbiol.* 11, 728–736.
- Malogajski, J., Brankovic, I., Verweij, S.P., Ambrosino, E., van Agtmael, M.A., Brand, A., Ouburg, S., Morre, S.A., 2013. Translational potential in health care of basic genomic and genetic findings for human immunodeficiency virus, *Chlamydia trachomatis*, and human papilloma virus. *Biomed. Res. Int.* 2013, 892106.
- Mardis, E.R., 2009. New strategies and emerging technologies for massively parallel sequencing: applications in medical research. *Genome Med* 1, 40.
- Masys, D.R., Jarvik, G.P., Abernethy, N.F., Anderson, N.R., Papanicolaou, G.J., Paltoo, D.N., Hoffman, M.A., Kohane, I.S., Levy, H.P., 2012. Technical desiderata for the integration of genomic data into electronic health records. *J. Biomed. Inform.* 45, 419–422.
- Matimba, A., Oluka, M.N., Ebeshi, B.U., Sayi, J., Bolaji, O.O., Guantai, A.N., Masimirembwa, C. M., 2008. Establishment of a biobank and pharmacogenetics database of African populations. *Eur. J. Hum. Genet.* 16, 780–783.
- Meijer, I., Molas-Gallart, J., Mattsson, P., 2012. Networked research infrastructures and their governance: the case of biobanking. *Sci. Public Policy* 1–9.
- Mitchell, R., 2010. National biobanks: clinical labor, risk production, and the creation of biovalue. *Sci. Technol. Hum. Values* 35, 330–355.
- Morre, S.A., Ouburg, S., Pena, A.S., Brand, A., 2009. The EU FP6 Epigenchlamydia Consortium: contribution of molecular epidemiology and host-pathogen genomics to understanding Chlamydia trachomatis-related disease. *Drugs Today* 45 (Suppl. B), 7–13.
- Overmeer, R.M., Louwers, J.A., Meijer, C.J., van Kemenade, F.J., Hesselink, A.T., Daalmeijer, N.F., Wiltink, S.M., Heideman, D.A., Verheijen, R.H., Zaal, A., van Baal, W.M., Berkhof, J., Snijders, P.J., Steenbergen, R.D., 2011. Combined CADM1 and MAL promoter methylation analysis to detect (pre-)malignant cervical lesions in high-risk HPV-positive women. *Int. J. Cancer* 129, 2218–2225.
- Peng, X., Chan, E.Y., Li, Y., Diamond, D.L., Korth, M.J., Katze, M.G., 2009. Virus-host interactions: from systems biology to translational research. *Curr. Opin. Microbiol.* 12, 432–438.
- Smart, A., Tutton, R., Ashcroft, R., Martin, P., Balmer, A., Elliot, R., Ellison, G.T.H., 2008. Social inclusivity vs analytical acuity? A qualitative study of UK researchers regarding the inclusion of minority ethnic groups in biobanks. *Med. Law Int.* 9, 169–190.
- Towie, N., 2007. London hospital launches infectious disease 'biobank'. *Nat. Med.* 13, 653.
- Townend, D., 2012. The politeness of data protection: exploring a legal instrument to regulate medical research using genetic information and biobanking (PhD thesis). *Universitaire Pers Maastricht, Maastricht* pp. 19–29.
- Velez, D.R., Wejse, C., Stryjewski, M.E., Abbate, E., Hulme, W.F., Myers, J.L., Estevan, R., Patillo, S.G., Olesen, R., Tacconelli, A., Sirugo, G., Gilbert, J.R., Hamilton, C.D., Scott, W.K., 2010. Variants in toll-like receptors 2 and 9 influence susceptibility to pulmonary tuberculosis in Caucasians, African-Americans, and West Africans. *Hum. Genet.* 127, 65–73.
- Wichmann, H.E., Kuhn, K.A., Waldenberger, M., Schmelcher, D., Schuffenhauer, S., Meitinger, T., Wurst, S.H., Lamla, G., Fortier, I., Burton, P.R., Peltonen, L., Perola, M., Metspalu, A., Riegman, P., Landegren, U., Taussig, M.J., Litton, J.E., Fransson, M.N., Eder, J., Cambon-Thomsen, A., Bovenberg, J., Dagher, G., van Ommen, G.J., Griffith, M., Yuille, M., Zatloukal, K., 2011. Comprehensive catalog of European biobanks. *Nat. Biotechnol.* 29, 795–797.
- Williams, R., Mant, C., Cason, J., 2009. The infectious diseases biobank at King's College London: archiving samples from patients infected with HIV to facilitate translational research. *Retrovirology* 6, 98.
- Zhang, L., Zhang, X., Ma, Q., Zhou, H., 2010. Host proteome research in HIV infection. *Genomics Proteomics Bioinformatics* 8, 1–9.