
Imaging Biobanks, Big Data, and Population-Based Imaging Biomarkers

13

Annalisa Mantarro, Paola Scalise,
and Emanuele Neri

13.1 Introduction

The modern medicine is constantly looking for the link between genetics, etiology, clinical manifestations, and treatment of diseases. In this regard, the anatomical and physiological interindividual differences and similarities, the predisposition to disease development, and its responsiveness to treatment depend widely from detailed information enclosed in the genoma [1].

The advent of biomedical research promoted the creation of an increasing number of facilities for long-term storage and retrieval of human cell and tissue samples. These biorepositories are known with the term of “biobanks,” which represent organized collections of biological samples (usually of human origin), centrally stored for one or more research purposes [2]. Human biobanks include biological material of healthy subjects and/or patients with specific pathologies (disease oriented), of which the most frequent are cancer related. However, there is some confusion

about the meaning of this term: some definitions are general, including all facility types for biological sample collection, while others are specific, comprehending strictly human sample collections [3]. Therefore, a clear definition of the term is an important step toward fostering the collaboration among researchers, allowing easy access to potential sample sources [4].

The history of biobanks starts with the pathology collections on the eighteenth–nineteenth century. During the second half of the twentieth century, the biomedical research was promoted in the United States and then in Europe with numerous collections of human samples for research purposes. In the recent decade, several international initiatives have emerged in order to promote and coordinate all existing and new biobanks and to develop standardized protocols and metrics [5]. The main goal of these initiatives is the implementation of infrastructural projects, aimed to improve the biomedical research by encouraging generic interoperability [6, 7].

The Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) was one of the first European Research Infrastructure projects funded by the European Commission in January 2011.

Actually, BBMRI is the largest organization of biobanks and biospecimen collections worldwide, including a 53-member consortium with over 280 associated organizations from over 30 countries.

A. Mantarro, MD • P. Scalise, MD • E. Neri, MD (✉)
Diagnostic and Interventional Radiology –
Department of Translational Research and New
Technologies in Medicine and Surgery, University
of Pisa, Pisa, Italy
e-mail: annalisamantarro@gmail.com;
scalisepl@gmail.com; emanuele.neri@med.unipi.it;
emanueleneri1@gmail.com; e.neri@ao-pisa.toscana.it

BBMRI is implemented under the European Research Infrastructure Consortium (ERIC) legal entity and its headquarters, located in Graz (Austria), and is responsible for the coordination of national activities in all participating countries [8].

Nevertheless, all data deriving from radiological imaging were not included in such biobanks; only recently, several projects have been started up for creating large repositories of image data, called “imaging biobanks” [9]. In this context, the registration of all imaging biobanks is essential, as well as the definition of structured approach for imaging data storage and retrieval. The latest goal is the research of a connection between the imaging and tissue biobanks, providing a deep association between the phenotype and genotype, by means of possible imaging biomarkers [2].

13.2 Radiomics and Personalized Care

The recent advent of high-throughput techniques for molecular analysis, including genomics, transcriptomics, proteomics, metabolomics, and imaging techniques, has allowed the storage of a large collection of data for identifying biomarkers used in the disease stratification, prediction, and early diagnosis of diseases [10, 11].

In this regard, the “radiomics” can be defined as the science that deals with the high-throughput extraction, storage, and analysis of a large amount of quantitative imaging features (*imaging biomarker*) in order to create accessible databases from radiological images and to reveal quantitative predictive or prognostic associations between images and medical outcomes [12, 13]. The modern and multiparametric imaging, characterized by digital and quantifiable informations, provides a set of biomarkers of the same patient that allow us to quantify the information. These biomarkers may refer to the organ function or neoplastic mass characteristics, and they are expressed by a number.

Examples of biomarkers are the diameter, volume, computed tomography (CT) density measurement, magnetic resonance (MR) signal

intensity, standard uptake value (SUV) in positron emission tomography (PET) imaging, contrast enhancement (valuated in MR, CT, or ultrasound examination), perfusion parameters (i.e., blood flow, blood volume, mean transit time, and permeability), tissue elasticity in elastosonography, tissue pattern (texture analysis), morphological pattern, and much more. Moreover, beyond radiology, other types of images can be collected, for example, from endoscopy, microscopy, and surgery, providing measurable personalized data. Each of these biomarkers is patient specific and will be stored, analyzed, and correlated as part of a cluster of biomarkers of that patient [14].

The main focus of imaging biobanks is the “personalized medicine,” where the treatment is increasingly tailored on the basis of specific characteristics of the patient and their disease [15]. Quantitative medical imaging, with the identification of imaging biomarkers, represents a crucial part of personalized medicine providing selection criteria and follow-up strategies, tailored to the patient’s needs [16]. All these imaging informations should be considered as the phenotypic expression of a patient and can be correlated to the genotype. In this setting, the radiogenomics, which is the extension of radiomics, aims to identify a link between genotype and phenotype imaging [17].

13.3 Imaging Biobanks: Current Status

The imaging biobanks are wide data collection including medical images and their correlated imaging biomarkers. The content of these biobanks, linked to that of biorepositories, should be available in a shared workflow among all researchers. A European network of imaging biobanks could significantly enhance the validation of new imaging biomarkers that could be potentially used as prognostic and predictive descriptors in the clinical practice.

In March 2014 the European Society of Radiology (ESR) instituted a dedicated working group (ESR WG on imaging biobanks) aimed at

monitoring all existing imaging biobanks in Europe, promoting the federation and communication among them in a white paper [2]. Furthermore, the ESR Working Group promoted the realization of imaging biobanks and techniques for the analysis and processing of imaging biomarkers, stimulating the integration of existing image data repositories and also the link between the imaging biobanks and traditional biobanks, as well as encouraging the researcher cooperation for the standardization, validation and benchmarking of all data stored. The development of imaging biobanks is focused on imaging data collection and sharing for clinical research programs (i.e., clinical trials). By the definition and validation of new biomarkers, the imaging biobanks meet the need for storage, diffusion, and comparison of disease-specific data [14]. In this context, the international research collaboration promotes the comparing of imaging tools, protocols, data, and expertise, in order to establish common acquisition protocols and to ensure high image quality. These data collections could be based on the regional/national screening programs or clinical trials (i.e., performed for colorectal cancer, breast cancer, or lung cancer). The oncologic imaging represents the most suitable field for the discovery and validation of new biomarkers from multiple imaging modalities, since the oncologic patients are frequently monitored for staging and follow-up of treatment response. Nowadays, there is a significant need of detailed and accurate biomarkers, in order to reduce cancer morbidity and mortality, promoting the progression of the traditional “one size fits all” strategy toward a new “personalized” cancer therapy [18].

13.4 Imaging Data Standardization

Accordingly to the dissemination and implementation of imaging biobank, the imaging collection and storage standardization are needed. The development of data standards promotes the communication among all the biobanks, using a standardized format, in order to integrate and

share suitable informations for all researchers, as well as to provide legal regulation in the institutions [19]. Nowadays, this cooperation among all researchers about imaging biobanks is very poor. On these bases, the main focus of imaging biobanks is the endorsement of high-quality standard levels, yielding harmonized datasets for biomarker extraction thus reducing the inter-variability [20]. All imaging researchers should cooperate to improve and standardize the image acquisition protocols and archiving, the softwares for data analysis and processing, and further methodologies for imaging biobanks.

The current standardization efforts promote the spreading of new techniques for medical image acquisition, visualization, storage, and sharing. In this setting, the Picture Archiving and Communication System (PACS) is a medical imaging technology, which provides the storage and access of digital imaging datasets deriving from multiple modalities through a network connection. The universal format for PACS image storage and transfer is the Digital Imaging and Communications in Medicine (DICOM) that encourages the interoperability between various systems of the healthcare institutions [21]. Moreover, this collaboration between healthcare enterprise professionals and industries resulted in the Integrating the Healthcare Enterprise (IHE), an international initiative, which defines how existing standards should be used (integration profile) [22]. The Cross-Enterprise Document Sharing (XDS) integration delineates the guidelines for sharing documents among all healthcare institutions, promoting the connection between the imaging centers and imaging biobanks [23].

Finally, the radiology report communicates all informations to the patient and referring physicians. For these purposes, it should be uniform, comprehensive, and easily understood. In 2008, the Radiological Society of North America (RSNA) promoted the dissemination of “structured” report templates, consisting of clinical data, coded terminology (e.g., SNOMED, RadLex), technical parameters, measurements, annotations, and key images [24]. Furthermore, the IHE profile (Management of Radiology Report Templates) defines the appropriate use of

templates, resulting in a wider diffusion of best practices and an improvement of radiology communication quality [25].

13.5 Ethical Issues

The recent biobank development gives rise to some issues about the big data management, in order to guarantee the respect of donors' privacy. In this context, the ethical aspects as well as legal and security issues represent crucial steps in biobank building and diffusion [2]. The ethical framework comprehends various steps, such as informed consent, donors' privacy, data protection, and sharing across the biobanks, thus sustaining of public trust [26–28].

- *Informed consent*

In the first instance, each participant must be informed about how to store, manage, and share own data and samples. On this ground, after the understanding of research purposes, the donor will be asked to sign a very clear consent, legally proving his voluntary authorization to data treatment in an anonymous form. No participant should be exposed to any research risk without their consent. However, there is still no consensus about the type of consent: some authors promoted a standardized consent form, ensuring comparability among all biobanks; while, according to others, the consent should respect geographical, social, and religious diversity [29, 30].

- *Data protection*

The data protection remains a very challenging legal task in biobanking. Among the European Union (EU) member countries, cross border data sharing and protection has been coordinated by the EU Data Protection Directive, leaving some margins for management by the EU member states [5]. The identity protection of research participants must be respected in the biobank framework [31]. The best way to ensure donors' privacy is the data collection in anonymous form; however, this process destroyed the correlation between the genomic and phenotypic informations. For

these reasons, many authors refused permanent anonymization and supported the coding of information as the most appropriate way for data protection. In this setting, the Ethics Review Boards ensure the identity protection of all participants, according to the ethical and legal frameworks and national legislation [29]. The international collaboration in data exchanging should be promoted, in order to minimize the risks for donors [32, 33].

Conclusion

The imaging biobanks represent virtual and accessible databases, focused on the discovery and validation of imaging quantitative biomarkers in order to guarantee a “personalized medicine,” which is increasingly tailored on the specific characteristics of the patients and their diseases.

These imaging biomarkers are employed in the early disease diagnosis, follow-up, and response to treatment, creating a link between the phenotype and genotype data (traditional biobanks).

The worldwide dissemination of the imaging biobanks is already ongoing; particularly in Europe, the ESR Working Group established the definition, management, legal and ethical issues, and federation among them.

References

1. Venter JC, et al. The sequence of the human genome. *Science*. 2001;292(5523):1838.
2. European Society of Radiology (ESR). ESR position paper on imaging biobanks. *Insights Imaging*. 2015;6(4):403–10.
3. Hewitt R, Watson P. Defining biobank. *Biopreserv Biobank*. 2013;11(5):309–15.
4. Shaw DM, et al. What is a biobank? differing definitions among biobank stakeholders. *Clin Genet*. 2014;85(3):223–7.
5. Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) Biobanks and the Public. *Governing Biomedical Research Resources in Europe*. 2013.
6. van Ommen GJ, et al. BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based expert centres. *Eur J Hum Genet*. 2015;23(7):893–900.

7. Riegman PH, et al. The organization of European Cancer Institute Pathobiology Working Group and its support of european biobanking infrastructures for translational cancer research. *Cancer Epidemiol Biomarkers Prev.* 2010;19(4):923–6.
8. Biobanking and BioMolecular Resources Infrastructure. Available at <http://bbmri-eric.eu/>, Lastly Accessed 6 Dec 2015.
9. Woodbridge M, et al. MRIdb: medical image management for biobank research. *J Digit Imaging.* 2013;26(5):886–90.
10. Galli J, et al. The biobanking analysis resource catalogue (BARCdb): a new research tool for the analysis of biobank samples. *Nucleic Acids Res.* 2015;43(Database issue):D1158–62.
11. Lambin P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441–6.
12. Kumar V, et al. QIN “radiomics: the process and the challenges”. *Magn Reson Imaging.* 2012;30(9):1234–48.
13. Bourcier C, et al. Radiomics: definition and clinical development. *Cancer Radiother.* 2015;19(6–7):532–7.
14. Hsu W, et al. Biomedical imaging informatics in the era of precision medicine: progress, challenges, and opportunities. *J Am Med Inform Assoc.* 2013;20(6):1010–3.
15. Aerts HJ, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5:4006.
16. Hewitt RE. Biobanking: the foundation of personalized medicine. *Curr Opin Oncol.* 2011;23(1):112–9.
17. Goyen M. Radiogenomic imaging-linking diagnostic imaging and molecular diagnostics. *World J Radiol.* 2014;6(8):519–22.
18. Mordente A, et al. Cancer biomarkers discovery and validation: state of the art, problems and future perspectives. *Adv Exp Med Biol.* 2015;867:9–26.
19. Quinlan PR, et al. A data standard for sourcing fit-for-purpose biological samples in an integrated virtual network of biobanks. *Biopreserv Biobank.* 2014;12(3):184–91.
20. Peakman T, Elliott P. Current standards for the storage of human samples in biobanks. *Genome Med.* 2010;2(10):72.
21. Huang HK. Medical imaging, PACS, and imaging informatics: retrospective. *Radiol Phys Technol.* 2014;7(1):5–24.
22. Viana-Ferreira C, et al. A framework for integration of heterogeneous medical imaging networks. *Open Med Inform J.* 2014;8:20–32.
23. Fatehi M, et al. Data standards in tele-radiology. *Acta Inform Med.* 2015;23(3):165–8.
24. Bosmans JM, et al. Structured reporting: a fusion reactor hungry for fuel. *Insights Imaging.* 2015;6:129–32.
25. IHE Radiology Technical Committee. IHE Radiology Technical Framework V13.0. Supplement. Management of Radiology Report Templates (MRRT). IHE International; 2014. Available at http://www.ihe.net/uploadedFiles/Documents/Radiology/IHE_RAD_Suppl_MRRT.pdf.
26. Hansson MG. The need to downregulate: a minimal ethical framework for biobank research. *Methods Mol Biol.* 2011;675:39–59.
27. D’Abramo F, et al. Research participants’ perceptions and views on consent for biobank research: a review of empirical data and ethical analysis. *BMC Med Ethics.* 2015;16:60.
28. Pullman D, et al. Personal privacy, public benefits, and biobanks: a conjoint analysis of policy priorities and public perceptions. *Genet Med.* 2012;14(2):229–35.
29. Budimir D, et al. Ethical aspects of human biobanks: a systematic review. *Croat Med J.* 2011;52(3):262–79.
30. Artizzu F. The informed consent aftermath of the genetic revolution. An Italian example of implementation. *Med Health Care Philos.* 2008;11(2):181–90.
31. Schulte in den Baumen T, et al. Data protection in biobanks – a european challenge for the long-term sustainability of biobanking. *Rev Derecho Genoma Hum.* 2009;(31):13–25.
32. Zika E, et al. Sample, data use and protection in biobanking in Europe: legal issues. *Pharmacogenomics.* 2008;9(6):773–81.
33. Mee B, et al. Development and progress of Ireland’s biobank network: ethical, legal, and social implications (ELSI), standardized documentation, sample and data release, and international perspective. *Biopreserv Biobank.* 2013;11(1):3–11.