

The Genome Austria Tissue Bank (GATiB)

M. Asslaber^a P.M. Abuja^a K. Stark^{a,f} J. Eder^f H. Gottweis^g M. Trauner^b
H. Samonigg^c H.J. Mischinger^d W. Schippinger^c A. Berghold^e H. Denk^a
K. Zatloukal^a

^aInstitute of Pathology, ^bDepartment of Medicine, Laboratory of Experimental and Molecular Hepatology, Division of Gastroenterology and Hepatology, ^cDivision of Oncology, Department of Internal Medicine, ^dDivision of General Surgery, Department of Surgery, and ^eInstitute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, and ^fFaculty of Computer Sciences, and ^gDepartment of Political Science, University of Vienna, Vienna, Austria

Key Words

Biobank · Biobanks, ethical, legal and social issues (ELSI) · Personalized medicine · Biomarkers · Drug development · Standardization · ELSI

Abstract

In the context of the Austrian Genome Program, a tissue bank is being established (Genome Austria Tissue Bank, GATiB) which is based on a collection of diseased and corresponding normal tissues representing a great variety of diseases at their natural frequency of occurrence from a non-selected Central European population of more than 700,000 patients. Major emphasis is put on annotation of archival tissue with comprehensive clinical data, including follow-up data. A specific IT infrastructure supports sample annotation, tracking of sample usage as well as sample and data storage. Innovative data protection tools were developed which prevent sample donor re-identification, particularly if detailed medical and genetic data are combined. For quality control of old archival tissues, new techniques were established to check RNA quality and antigen stability. Since 2003, GATiB has changed from a population-based tissue bank to a disease-focused biobank comprising major cancers such as colon, breast, liver, as well as metabolic liver diseases and organs affected by the metabolic syndrome. Prospectively collected tissues are associated with blood samples and detailed data on the sample donor's disease, lifestyle and envi-

ronmental exposure, following standard operating procedures. Major emphasis is also placed on ethical, legal and social issues (ELSI) related to biobanks. A specific research project and an international advisory board ensure the proper embedding of GATiB in society and facilitate international networking.

Copyright © 2007 S. Karger AG, Basel

Introduction

Biobanks containing human biological samples, such as tissues, blood or body fluids, in combination with donor-related clinical data, are essential resources for the understanding of the function and medical relevance of human genes as well as for exploring the biological networks in which specific genes are operating [1–3]. Particularly valuable information can be obtained from biobanks comprising normal and diseased human tissues of high quality [4]. Such material provides information on genetic and epigenetic alterations and on modifications of gene products which cause diseases or influence their outcome. Moreover, large tissue collections provide insight into the great variability of human disease manifestations and the individual's response to medical treatment and are, therefore, an essential basis for the advancement of personalized medicine [5–8].

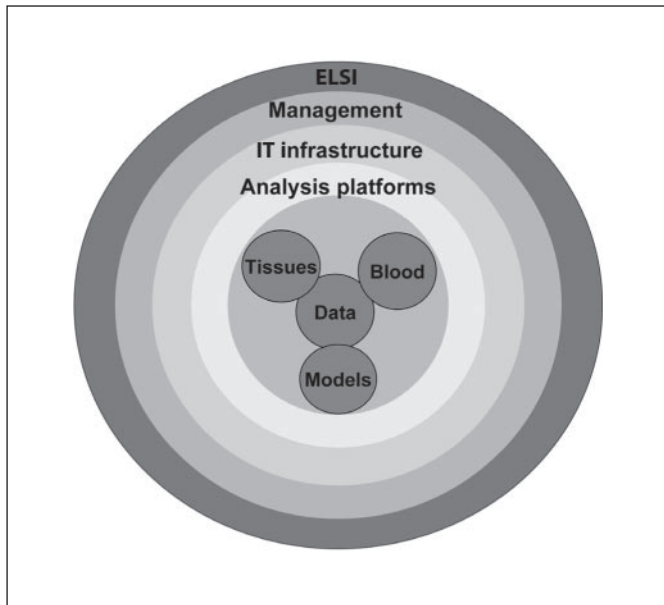


Fig. 1. Structure of GATiB: GATiB contains tissues, blood samples as well as clinical data and data generated by analysis of samples. Human samples are complemented by animal models for major diseases. Standardized analysis of samples is enabled by a broad spectrum of associated ‘-omics’ platforms. The IT infrastructure supports sample and data storage, sample tracking, data analysis, and guarantees patient confidentiality. A management team is responsible for the operation of GATiB as a core facility of the MedUG. A specific research project and an international advisory board provide guidance for ethical, legal and social issues (ELSI).

Comparative genetic analysis of diseased and corresponding non-diseased tissue allows identification of germ-line mutations, polymorphisms and somatic mutations relevant to disease. Furthermore, within the diseased tissue an imprint of epigenetic alterations, changes in gene expression levels, post-translational modifications of gene products, and metabolites are preserved [9]. Using novel high-throughput (‘-omics’) technologies, these alterations can be identified providing insight not only into the genetic but also into the acquired non-genetic, particularly environmental (infectious, toxic, allergic, etc.) causes of systemic and localized diseases [10, 11].

Several pathology institutes have established large archives of diseased tissues, which have been obtained for diagnostic purposes. However, in order to fulfill the needs of modern research, the requirements of a biobank are much more far-reaching than those of medical archives for diagnostic service [12].

In medical tissue archives, samples are usually not directly associated with clinical data, which substantially limits their usability for research. The annotation of archival samples with data from medical records which greatly enhances their value for scientific use is a major challenge for biobanks [13, 14]. For instance, the impact of alterations in gene expression on disease outcome or drug response can only be verified in context of a variety of clinical background information. However, such data, as far as they are electronically accessible for older archived samples are often stored in different departments and hospitals in different databases, with the additional obstacle that data formats have changed over the years. Moreover, samples have been collected over long periods of time, where standards for preservation have not been considered a priori [15] and criteria of disease classification may have changed.

Another aspect is that since ethical standards have considerably changed during the last two decades, modern standards cannot be applied to old archived material. While for presently collected samples informed consent (IC) by the donor is compulsory, for older archived material IC was not required [4]. In accordance with emerging international standards, such samples can only be used after approval of specific research projects by the local ethical committee (see for instance, Opinion on Biobanks for Research of the German National Ethics Council, www.ethikrat.org).

A major task of the Genome Austria Tissue Bank (GATiB) project is the transformation of the existing tissue archive at the Medical University of Graz (MedUG), currently consisting of more than 3 million samples from over 800,000 patients, into a biobank by addressing the problems outlined above. Furthermore, the collection is expanded by adding new tissue samples, complemented with corresponding blood samples and detailed patient-related information, particularly considering the different requirements for archive material and prospectively collected samples and data (fig. 1) (for further information, see www.bioresource-med.at).

GATiB: Its Concept and Content

Historical Background

The Institute of Pathology of the MedUG was originally a center of pathology service for the southern part of the Austrian-Hungarian monarchy and because of this tradition it is still the pathology service provider for the University Hospital in Graz and for a further 33 public

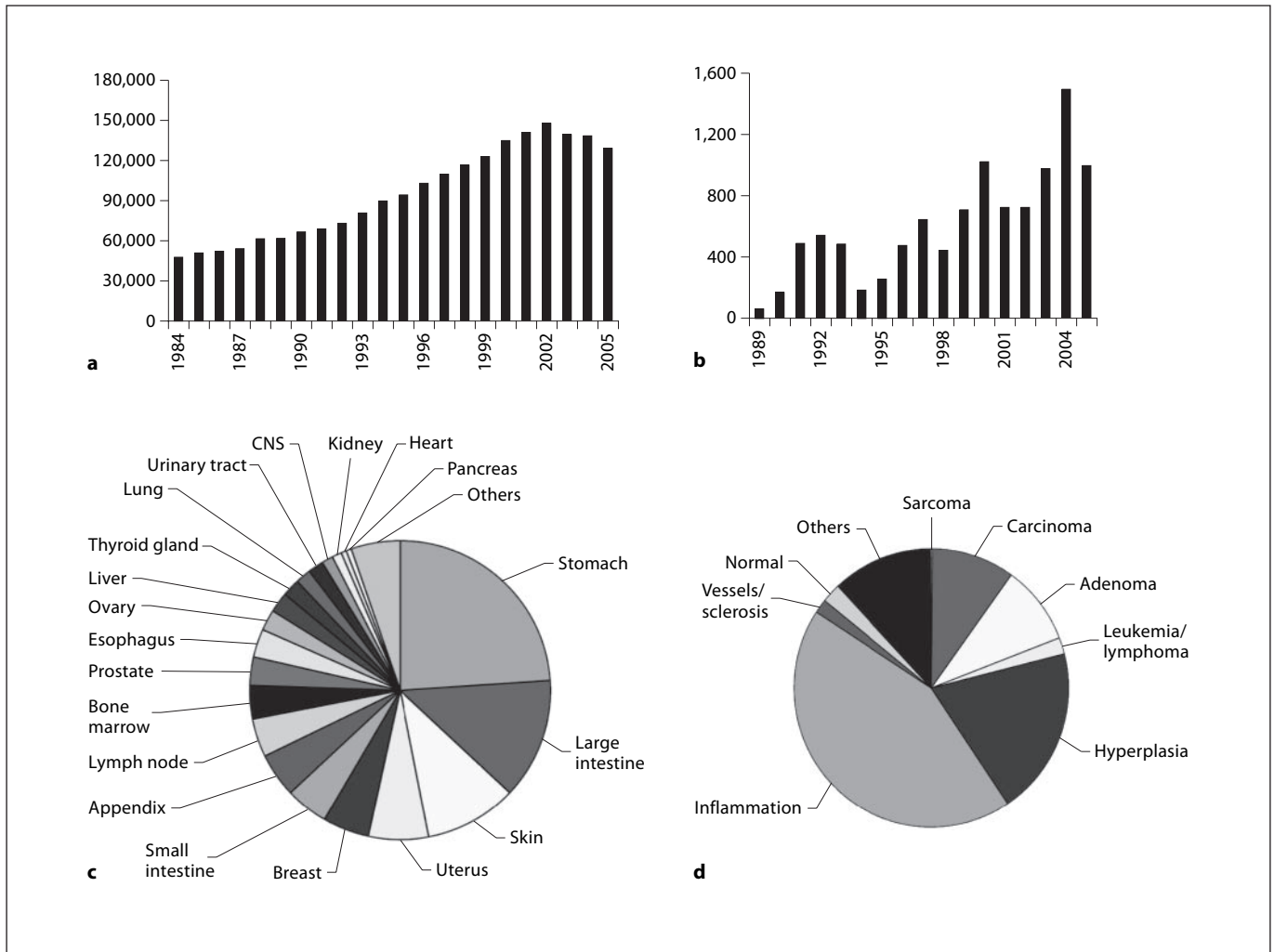


Fig. 2. Annual number of collected formalin-fixed paraffin-embedded (FFPE) (a) and cryopreserved tissues (b). Total content of FFPE tissue samples is shown for major organs (c) and diseases (d).

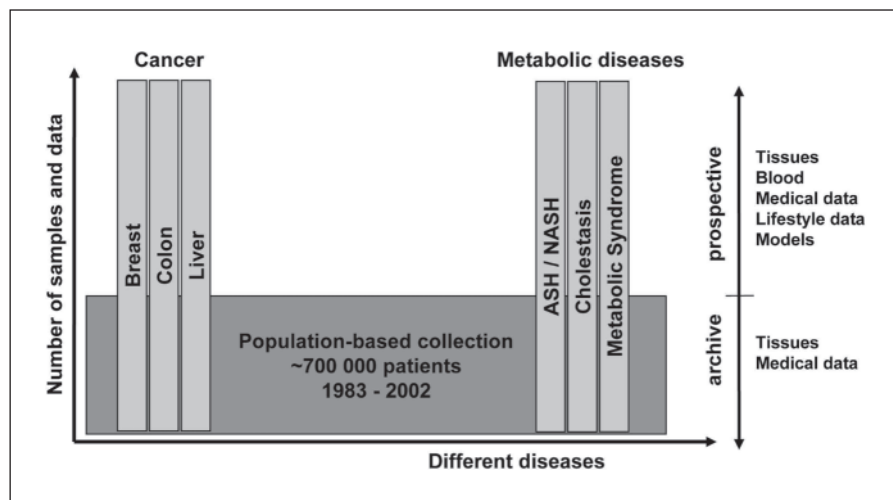
and 10 private hospitals. As a consequence, in south-eastern Austria a single institute is responsible for the preservation, diagnosis and storage of tissue samples, resulting in a large population-based collection of samples. This archive formed the basis for GATiB, contributing samples from 700,000 patients, totaling 2.7 million diseased and normal tissue specimens collected from 1983 to 2002. Over this time period the collection reflects all diseases where tissues have been obtained for diagnostic or therapeutic purposes. Hence, diseases are represented almost at their natural frequency of occurrence characteristic for a mixed Central European population. An important aspect is that all samples have been collected and processed in one institute, so that the same tissue-processing proto-

col was applied over the years, guaranteeing a high level of standardization. Since 2002, the population-based character has not been maintained any longer due to decentralization of the pathology service. Nevertheless, by the end of 2005, about 3 million formaldehyde-fixed, paraffin-embedded (FFPE) and more than 30,000 cryopreserved tissue samples from 800,000 patients were available within GATiB (fig. 2).

Dual Focus on Old Archived and Prospectively Collected Samples

The archival material is of outstanding value, since it represents tissue specimens of a large non-selected cohort, for most of which long-term follow-up data are

Fig. 3. Dual focus of GATiB: archived material comprises an exceptionally large population-based collection of a non-selected cohort of 700,000 patients. Prospectively collected samples have the main focus on major human cancers as well as metabolic liver diseases and organs affected by the metabolic syndrome. Tissues and associated blood samples as well as patient-related data are collected in a standardized fashion. Furthermore, for specific diseases, animal models are being established and validated for their human disease relevance.



available. Furthermore, certain disease conditions can only be studied in such older archived samples which cannot be obtained anymore, since therapy regimens have changed (e.g. increase of neoadjuvant therapy) and advanced tumor stages are observed less frequently in certain organs due to improved early diagnosis and success of cancer prevention programs.

Most archived material is available as FFPE tissue that was collected in the context of histopathological diagnosis. The relevance of this material was underestimated for a long time. However, FFPE material gains more importance since improved methods for RNA and protein extraction have been developed [16–19]. Although the same preservation technique was used for all specimens in the archive, variations in the process (e.g. longer fixation times over the weekend or variations in ischemia time) can never be excluded. Therefore, particular attention must be paid to sample quality control in order to check whether certain specimens meet the requirements of modern ‘-omics’ methodologies [20]. As an example, the widely used ratio of 18/28s rRNA cannot be applied as a quality marker for RNA conservation in FFPE, since after formaldehyde fixation, ribosomal RNA is tightly cross-linked to ribosomal protein complexes and, therefore, not properly represented in RNA preparations. Therefore, an assay for RNA quality control in FFPE has been developed, which is based on the quantification of the fragment size of preserved mRNA.

Several of the above-mentioned limitations of the use of archived samples can be avoided with prospectively collected material since harvesting, processing, and storage of samples including associated data can be standard-

ized and documented. Although a variety of methods has been developed to extract nucleic acids and proteins from FFPE tissues, several analytical technologies still rely on snap-frozen samples. Hence, systematic collection of material for cryopreservation in liquid nitrogen was started in the early 1990s. Most cryopreserved material was collected in the context of frozen section diagnosis of surgical specimens within 10–20 min after surgical removal. Recently, cryofixation was also extended to other surgical specimens for which frozen section diagnosis is not performed. A research nurse and a 24-hour standby pathology service are in charge of obtaining and cryopreserving surgical material. In addition to tissues, corresponding blood samples are collected before and after surgery as well as during follow-up since matched pairs of diseased tissues and blood samples are particularly suited to rapid identification of novel biomarkers [21]. Patient-related information, family history, lifestyle data, etc. are collected in a standardized fashion using questionnaires. Such processes of standardized patient sample and data collection can only be established in the context of interdisciplinary cooperations with highly committed project partners. Therefore, prospective material collection is focused on the research priorities of the MedUG in the fields of breast, colon and liver cancer as well as on metabolic diseases with special emphasis on metabolic liver diseases and tissues affected in the course of the metabolic syndrome (fig. 3).

In certain disease areas, analysis of human tissue and blood samples is complemented by studies using animal models for these diseases, such as tumor xenografts in nude mice, pharmacological and transgenic models for

the functional characterization of genes [22]. The establishment of animal models in the context of a human biobank also allows detailed comparative studies of models and human diseases which is a prerequisite for model validation. Furthermore, in selected cases, primary cell lines are established from tumor samples. Importantly, the cell lines and xenografts can be directly compared with the original human diseased tissue, which is generally not the case for conventional tumor cell lines.

Preservation Techniques

The development and validation of new fixation and preservation techniques are crucial to guarantee optimal future usability of biological material. A series of novel techniques are being developed for the collection of fresh material and tested in research projects performed in collaboration with GATiB. These improved techniques markedly enhance the value of a tissue bank, allowing better analysis of diseased tissue. Currently, new fixation procedures for improved preservation of DNA and RNA in paraffin-embedded tissues are tested. Moreover, a preservation protocol is being optimized for proteomic analysis with special emphasis on the detection of post-translational modifications.

Analysis Platforms

Tissue samples are a valuable and limited resource. This necessitates state-of-the-art technologies for analysis to obtain high-quality data. Therefore, it is indispensable to integrate analysis platforms within the framework of a biobank – not only to guarantee efficient and economical use of tissue, but also to ‘transform tissue into data’. The transition from a sample-focused to a data-focused resource may be regarded as characteristic for the maturation process of a biobank.

In the context of the broad spectrum of analysis platforms, which can be applied to tissues, tissue microarray technology is of central importance in association with a biobank. Hundreds of specimens can be analyzed (e.g. by immunohistochemistry, and in-situ hybridization) in a single procedure, guaranteeing stable conditions with minimal expenditure of time, samples and reagents [23, 24]. A proprietary robotic system for automated and highly standardized tissue microarray production allowing the analysis of thousands of tissues, which is also suitable for industrial applications, has been developed in cooperation with a biotech company (ORIDIS Biomed GmbH). Its key features are precise core selection (digital images of HE-stained tissue sections with marks for core selection are overlaid with images of the paraffin blocks),

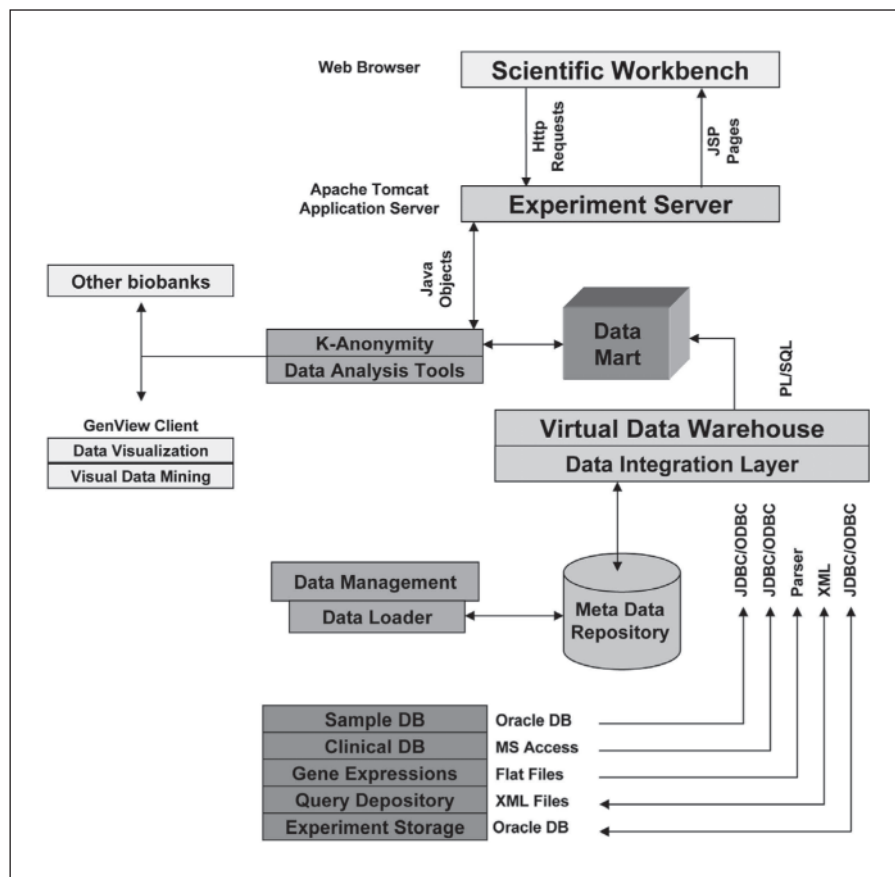
simultaneous generation of multiple array blocks and automated documentation of the manufacturing process which allows to trace back each tissue core to the original sample from which it has been derived.

IT Infrastructure and Data Security

The entire information about the sample collection and the associated data is electronically accessible and searchable via an Oracle®-based sample database. Sensitive data, such as name, date of birth, which directly identify the individual, are stored separately and protected by high-security measures (firewall, authentication system, etc.). Each dataset of different hierarchies is coded with individually generated identity numbers, serving as keys for project partners. Storing of fully anonymized data is not feasible since patients’ identities have to be kept for annotation of samples with clinical data, in particular follow-up data, or for proper assignment of specimens or serum collected at different time points from the same patient [4, 13, 25]. Primary processing of tissue, such as extraction of DNA, RNA or proteins, is also documented and linked to the data generated from the samples. The sample database was designed open, so that any further analysis or documentation of specimens can be recorded in a growing part of the database.

The combination of experimental, especially genetic data, with patient data requires further measures for data protection. In this context, a specific challenge is that the patient’s identity can, at least theoretically, be restored from coded data through the combination of large datasets [4, 26]. To address this problem, the different data sources are integrated in a virtual data warehouse which is accessed by a data mart (fig. 4). In this data mart a more far-reaching anonymization layer is implemented, which is based on the concept of K-anonymity. This allows the combination of data only in such a way that a sufficient number of data twins is available, which prevents the indirect identification of individuals by seemingly ‘innocent’ parameters [27]. To eliminate identifying parameters, they can be aggregated by different strategies, such as transformation (e.g., body weight and size into body mass index) or regrouping (e.g., individual age into age groups), by the investigator. Parameters can thus be flexibly tuned to the requirements of a particular study with a minimal loss of information. Researchers have access to coded data records only, the K-anonymity being an additional tool for prevention of re-identification of linked complex datasets. Mainly for security reasons, an open-access infrastructure is unacceptable, as it might conflict with patients’ rights because of the sensitive nature of the

Fig. 4. IT infrastructure: a scientific workbench supports the access to and combination of data stored in different databases (DB). After defining the parameters for a specific experiment the required data are materialized in a specific data mart. Only coded data are released and an additional anonymization layer ensures K-anonymity. The second interface allows visual data-mining for ad-hoc grouping, filtering and analysis of large and different types of datasets. Furthermore, this interface enables data exchange with other biobanks.



information that can be extracted from tissues linked to genetic and clinical data. The wealth of data resulting from different domains enables purely data-based experiments without using laboratory facilities (in-silico experiments) and supports goal-directed planning of further experiments.

Ethical, Legal and Social Guidance

Biobanks operate at the intersection of the public perception of research and the research community. The high operation costs of a biobank require high transparency and the many contributors need to benefit, and so do different project partners and sample donors [4]. To address general demands and concerns of the public as well as the perception of GATiB and its international networking activities, a subproject within GATiB has been set up, which specifically deals with the ethical, legal and social issue (ELSI) requirements. This subproject combines empirical research with reflection on the ethical and social implications of GATiB. In order to obtain a better understanding of the sociopolitical dynamics of

biobanks, different ones are studied in detail. The project will give a global comparative view of major biobanks based on empirical on-site visits and evaluations. This, on the one hand, provides the basis for close cooperation with other similar initiatives worldwide. On the other hand, much can be learnt from other major biobank projects, and these empirical observations and analyses can be useful for creating a positive society-biobank interaction. Additionally, ELSI guidance of GATiB is provided by an international advisory board consisting of distinguished experts in the fields of ethics, social sciences, law and data protection.

Discussion

Because of the enormous efforts required to establish a population-based biobank, many biobank initiatives focus their sample collection on a certain disease group (e.g. LURIC for cardiovascular diseases [28], EPIC for cancer [29], specific tumor registers, etc.). This has the

advantage that sample and data collection procedures are specifically designed for the demands related to the disease of interest. Moreover, research projects can be performed readily after the establishment of a biobank of this type. On the other hand, population-based biobanks have the advantage of serving as valuable resources for genetic epidemiological studies. Furthermore, biomarker development requires access to a cohort representing most human diseases and corresponding normal tissues. The specificity of a biomarker found in a disease-focused study can only be properly validated in a population-based – preferably non-selected – collection, which contains diseased and normal tissues from a broad spectrum of different organs (e.g. the healthy organ panel requested by the FDA) as well as corresponding blood samples [30, 31]. GATiB constitutes a combination of both types in that it contains a population-based collection and is now expanded by a disease-focused approach.

The number of samples stored in a biobank is another relevant aspect if required to support the needs of personalized medicine. For example, breast cancer is a common disease and consequently GATiB contains more than 10,000 breast cancer samples. Nevertheless, molecular profiling of specific tumor entities in combination with specific treatment regimens in different patient subgroups markedly reduces the number of samples fulfilling the inclusion criteria for analysis, so that statistical significance is hardly achieved for some rarer constellations. This underlines that only a large number of samples can provide a sufficient size of specific patient subgroups to be useful for evaluating individual variations of diseases. This becomes even more critical in highly diverse or rare diseases. Therefore, to exploit the full potential of biobanks, networking between individual biobanks is indispensable [2, 32]. As a prerequisite for international cooperation it is not only necessary to define common standards for sample quality and data formats

[33], but also to consider the differences in ethical, legal and social environments in the different countries of partner biobanks [34–36]. Besides public acceptance and networking, biobank management and governance needs to cover a variety of aspects such as compliance with biosafety and biosecurity regulations as well as maintaining a balance between sample use and accrual.

Outlook

The amount of diseased tissue which can be collected from individual patients will always be very limited, and some researchers express concern that tissue banks cannot meet the growing demand for material [37]. GATiB's mission is to contribute to the advancement of medicine and excellence in research. Although it is established as a core facility of the MedUG, it is accessible to the international scientific community as well as to industry. However, GATiB does not act as a tissue distributor but supports specific research projects of high quality. Analysis of the samples has to be performed by using standardized and state-of-the-art analysis platforms in a well-coordinated manner to ensure that a broad scientific community can build on the data generated. By such a process, 'tissue' is converted to 'high-quality data' leading to an expanding knowledge base which can be shared with the scientific community.

Acknowledgements

This work has been supported by grants from the Austrian Genome Program (GEN-AU), the European Union, and the Zukunftsfonds of the Province of Styria. The authors wish to acknowledge the efforts, contributions and discussion of all the GATiB coworkers and the colleagues at the MedUG. Also, the valuable IT support of G. Casari at ORIDIS Biomed GmbH is gratefully acknowledged.

References

- 1 Hirtzlin I, Dubreuil C, Preaubert N, Duchier J, Jansen B, Simon J, Lobato De Faria P, Perez-Lezaun A, Visser B, Williams GD, Cambon-Thomsen A: An empirical survey on biobanking of human genetic material and data in six EU countries. *Eur J Hum Genet* 2003;11:475–488.
- 2 Hagen HE, Carlstedt-Duke J: Building global networks for human diseases: genes and populations. *Nat Med* 2004;10:665–667.
- 3 Cambon-Thomsen A: The social and ethical issues of post-genomic human biobanks. *Nat Rev Genet* 2004;5:866–873.
- 4 Godard B, Schmidtke J, Cassiman JJ, Ayme S: Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *Eur J Hum Genet* 2003;11(suppl 2):S88–S122.
- 5 Collins CD, Purohit S, Podolsky RH, Zhao HS, Schatz D, Eckenrode SE, Yang P, Hopkins D, Muir A, Hoffman M, McIndoe RA, Rewers M, She JX: The application of genomic and proteomic technologies in predictive, preventive and personalized medicine. *Vascul Pharmacol* 2006;45:258–267.
- 6 Nicholson JK: Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2006;2:52.

- 7 Jain KK: A critical review of the Royal Society's report on personalized medicine. *Drug Discov Today* 2006;11:573–575.
- 8 Dielert M, Sers C: Personalized medicine and development of targeted therapies: the upcoming challenge for diagnostic molecular pathology. A review. *Virchows Arch* 2006; 448:744–755.
- 9 Fischer HP: Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery. *Biotechnol Annu Rev* 2005;11: 168.
- 10 Corella D, Ordovas JM: Integration of environment and disease into 'omics' analysis. *Curr Opin Mol Ther* 2005;7:569–576.
- 11 Vemuri GN, Aristidou AA: Metabolic engineering in the-omics era: elucidating and modulating regulatory networks. *Microbiol Mol Biol Rev* 2005;69:197–216.
- 12 Kaiser J: Biobanks. Population databases boom, from Iceland to the USA. *Science* 2002;298:1158–1161.
- 13 Patel AA, Gilbertson JR, Parwani AV, Dhir R, Datta MW, Gupta R, Berman JJ, Melamed J, Kajdacsy-Balla A, Orenstein J, Becich MJ: An informatics model for tissue banks – lessons learned from the Cooperative Prostate Cancer Tissue Resource. *BMC Cancer* 2006; 6:120.
- 14 Gilbertson JR, Gupta R, Nie Y, Patel AA, Becich MJ: Automated clinical annotation of tissue bank specimens. *Medinfo* 2004;11: 607–610.
- 15 Grizzle W, Grody WW, Noll WW, Sobel ME, Stass SA, Trainer T, Travers H, Weedn V, Woodruff K: Recommended policies for uses of human tissue in research, education, and quality control. Ad Hoc Committee on Stored Tissue, College of American Pathologists. *Arch Pathol Lab Med* 1999;123:296–300.
- 16 Little SE, Vuononvirta R, Reis-Filho JS, Natrajan R, Irvani M, Fenwick K, Mackay A, Ashworth A, Pritchard-Jones K, Jones C: Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. *Genomics* 2006; 87:298–306.
- 17 Devries S, Nyante S, Korkola J, Seagraves R, Nakao K, Moore D, Bae H, Wilhelm M, Hwang S, Waldman F: Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J Mol Diagn* 2005;7:65–71.
- 18 Shi SR, Liu C, Balgley BM, Lee C, Taylor CR: Protein extraction from formalin-fixed, paraffin-embedded tissue sections: quality evaluation by mass spectrometry. *J Histochem Cytochem* 2006;54:739–743.
- 19 Ikeda K, Monden T, Kanoh T, Tsujie M, Izawa H, Haba A, Ohnishi T, Sekimoto M, Tomita N, Shiozaki H, Monden M: Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J Histochem Cytochem* 1998;46:397–403.
- 20 Jewell SD, Srinivasan M, McCart LM, Williams N, Grizzle WH, LiVolsi V, MacLennan G, Sedmak DD: Analysis of the molecular quality of human tissues: an experience from the Cooperative Human Tissue Network. *Am J Clin Pathol* 2002;118:733–741.
- 21 Anderson NL, Anderson NG: The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845–867.
- 22 Sausville EA, Burger AM: Contributions of human tumor xenografts to anticancer drug development. *Cancer Res* 2006;66:3351–3354.
- 23 Haedicke W, Popper HH, Buck CR, Zatloukal K: Automated evaluation and normalization of immunohistochemistry on tissue microarrays with a DNA microarray scanner. *Biotechniques* 2003;35:164–168.
- 24 Simon R, Mirlacher M, Sauter G: Tissue microarrays. *Methods Mol Med* 2005;114:257–268.
- 25 Auray-Blais C, Patenaude J: A biobank management model applicable to biomedical research. *BMC Med Ethics* 2006;7:E4.
- 26 Van Veen EB, Riegman PH, Dinjens WN, Lam KH, Oomen MH, Spatz A, Mager R, Ratcliffe C, Knox K, Kerr D, van Damme B, van de Vijver M, van Boven H, Morente MM, Alonso S, Kerjaschki D, Pammer J, Lopez-Guerrero JA, Llombart Bosch A, Carbone A, Ghoghini A, Teodorovic I, Isabelle M, Passiouvov A, Lejeune S, Therasse P, Oosterhuis JW: TuBaFrost 3: regulatory and ethical issues on the exchange of residual tissue for research across Europe. *Eur J Cancer* 2006; 42:2914–2923.
- 27 Stark K, Eder H, Zatloukal K: Priority-based K-anonymity accomplished by weighted generalisation structures. *Comput Sci* 2006; 4081:394–404.
- 28 Winkelmann BR, Marz W, Boehm BO, Zotz R, Hager J, Hellstern P, Senges J: Rationale and design of the LURIC study – a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics* 2001;2:S1–S73.
- 29 Riboli E, Kaaks R: The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;26 Suppl 1:S6–S14.
- 30 Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, Dominiczak AF, Fitzpatrick B, Ford I, Jackson C, Haddow G, Kerr S, Lindsay R, McGilchrist M, Morton R, Murray G, Palmer CN, Pell JP, Ralston SH, St Clair D, Sullivan F, Watt G, Wolf R, Wright A, Porteous D, Morris AD: Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 2006;7:74.
- 31 Ollier W, Sprosen T, Peakman T: UK Biobank: from concept to reality. *Pharmacogenomics* 2005;6:639–646.
- 32 Bouchie A: Coming soon: a global grid for cancer research. *Nat Biotechnol* 2004;22: 1071–1073.
- 33 Pearson H: Summit calls for clear view of deposits in all biobanks. *Nature* 2004;432:426.
- 34 De Montgolfier S, Moutel G, Duchange N, Callies I, Sharara L, Beaumont C, Feingold J, Herve C: Evaluation of biobank constitution and use: multicentre analysis in France and propositions for formalising the activities of research ethics committees. *Eur J Med Genet* 2006;49:159–167.
- 35 Joly Y, Knoppers BM: Pharmacogenomic data sample collection and storage: ethical issues and policy approaches. *Pharmacogenomics* 2006;7:219–226.
- 36 Hansson MG, Dillner J, Bartram CR, Carlsson JA, Helgesson G: Should donors be allowed to give broad consent to future biobank research? *Lancet Oncol* 2006;7:266–269.
- 37 Adam D: Online tumour bank aims to offer ready route to tissues. *Nature* 2002;416:464.