Innovative Technology and Its Contribution to Biobanking

Kurt Zatloukal M.D.
Institute of Pathology
Medical University of Graz
Austria
Content

• Drivers of innovation in biobanking
• Innovation fields
• Increasing relevance of data
• Image data a novel biobanking component
Content

• Drivers of innovation in biobanking
• Innovation fields
• Increasing relevance of data
• Image data a novel biobanking component
**Biobanking: a rapidly growing market**

Overall world revenue for Biobanking will surpass $53bn in 2027, our work calculates. This will be driven by the use of biobanking in drug development and medical research, as well as the expected rise in number of commercial biobanks over the forecast period.  

[https://www.wallstreet-online.de/nachricht/9849463-biobanking-market-forecasts-2017-2027](https://www.wallstreet-online.de/nachricht/9849463-biobanking-market-forecasts-2017-2027)

**Global Forecast to 2022**", The biobanking market is expected to reach USD 2.69 Billion by 2022 from USD 1.85 Billion in 2017, at a CAGR of 7.8%. Factors driving the growth of this market include the increasing number of genomics research activities for studying diseases; advances in biobanking and the growing trend of conserving cord blood stem cells of newborns; government & private funding to support regenerative medicine research; and the growing need for cost-effective drug discovery and development. MarketsandMarkets™
User Needs: Tackling the great health challenges and impact on biobank content

Environment

Life style and exposure data

Food, phys. exercise, stress, toxins, pathogens

Pathogens

Oral microbiome
Stool

Tissues
Cells
iPS
Blood - CTC
Serum
Plasma - cfDNA
Urine
Saliva
Exhale condensate

Genome

Microbiome

Transcriptome

Proteome

Metabolome

Health Disease

K. Zatloukal, GBW 2017
An Unsolved Issue: Limited reproducibility of scientific results

Too many of the findings that fill the academic ether are the result of shoddy experiments or poor analysis (see pages 21-24). A rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated. Even that may be optimistic. Last year researchers at one biotech firm, Amgen, found they could reproduce just six of 53 “landmark” studies in cancer research. Earlier, a group at Bayer, a drug company, managed to repeat just a quarter of 67 similarly important papers. A leading computer scientist frets that three-quarters of papers in his subfield are bunk. In 2000-10 roughly 80,000 patients took part in clinical trials based on research that was later retracted because of mistakes or improprieties.
A Major Opportunity for Biobanks

The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman, Iain M. Cockburn, Timothy S. Simcoe

Estimated US Annual Preclinical Research Spend

US$56.4B

US$28.2B

50%

50%

Categories of Preclinical Irreproducibility

Biological Reagents and Reference Materials
(36.1% of total)

Study Design
(27.6% of total)

Data Analysis and Reporting
(25.5% of total)

Laboratory Protocols
(10.8% of total)
A New Gold Rush

- Big data
- Data mining
- Machine learning
- Deep learning
- Artificial intelligence
- Computational modelling
- Semantic searches
- Ontologies
- Knowledge discovery
Future of Biobanks

from sample to data storage
from raw material to knowledge
New Legal and Regulatory Requirements

REGULATION (EU) 2017/746 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 5 April 2017


REGULATIONS

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

WORLD VIEW A personal take on events

We must urgently clarify data-sharing rules

Scientists have worked hard to ensure that Europe’s new data laws do not harm science, but one last push is needed, says Jan-Eric Litton.

26 JANUARY 2017 | VOL 541 | NATURE | 437

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

K. Zatloukal, GBW 2017
Content

• Drivers of innovation in biobanking
• Innovation fields
• Increasing relevance of data
• Image data a novel biobanking component
Innovation Fields

- Biosamples and data types
- Sampling technologies
- Sample stabilization
- Sample storage
- Sample analytics
- Data management solutions
- Application fields and business models
Examples of Novel Developments

- Biosamples and data types:
  Microbiome, body fluids, imaging data
- Sampling technologies
  Circulating free DNA
- Sample stabilization
  Room temperature stabilization of biomolecules
- Sample analytics
  Metabolomics, proteomics, NMR
- Data management solutions
  Machine learning, AI
- Application fields and business models
  Public–private–partnerships
Content

- Drivers of innovation in biobanking
- Innovation fields
- Increasing relevance of data
- Image data a novel biobanking component
Inventory of 100 million biosamples in Europe

- Tissue, Blood, DNA
- Blood, DNA

Surface is proportional to number of samples

BB MUG

Tissue, Blood, DNA

Blood, DNA

Surface is proportional to number of samples
Inventory of 60 million biosamples in Europe

How to find the right samples?

A tissue sample is best characterized by its image

Surface is proportional to number of samples
(only) Images Capture the Specific Features of Tissue Samples
Histological Slides: New relevance of an old resource
Imaging Biomarkers

ESR statement on the stepwise development of imaging biomarkers

European Society of Radiology (ESR)

---

ESR Position Paper on Imaging Biobanks

European Society of Radiology (ESR)
Supervised Learning by Using Labeled Data

Detecting Cancer Metastases on Gigapixel Pathology Images

Yun Liu¹*, Krishna Gadepalli¹, Mohammad Norouzi¹, George E. Dahl¹, Timo Kohlberger¹, Aleksey Boyko¹, Subhashini Venugopalan²**, Aleksei Timofeev², Philip Q. Nelson², Greg S. Corrado¹, Jason D. Hipp³, Lily Peng¹, and Martin C. Stumpe¹

{liuyun,mnorouzi,gdahl,lhpeng,mstumpe}@google.com

¹Google Brain, ²Google Inc, ³Verily Life Sciences, Mountain View, CA, USA

270 slides pixel-level annotation (Camelyon16 data set)

➢ Few data sets required
➢ Annotation process very laborious, expensive, error prone
Unsupervised Learning by Using Unlabeled Data

Building High-level Features Using Large Scale Unsupervised Learning

- 10 million randomly selected YouTube video thumbnails (200 x 200 pixels)
- Trained to recognizing 20,000 object categories using 16,000 cores

Need for very large data sets and large computing power
Linking Digital Slides and Biosamples

Region of interest

FFPE block  Digital slide  Disease outcome
Finding new
• molecular alterations
• biomarkers
• therapy targets
• redefinition of morphologic criteria

Region of interest

FFPE block  Digital slide  Disease outcome
Challenges

• Availability of digital images and medical data
• High-throughput scanning capacities
• Data storage capacities
• ELSI compliance
  Anonymization
  GDPR compliance
**High-Throughput Scanning**

1600 slides per day
- 1000 slide loading capacity
- Parallelized scanning
- Fast scanning
- Open and flexible design

Test: Scanning time for 40x (0.25\(\mu\)m/pixel): 51 seconds/slide
Decrease of Data Storage Costs

![Graph showing 10-year technology cost/terabyte projections from 2014 to 2023. The graph indicates a decrease in cost for NAND Flash, Disk, and Tape, with CGR values provided for each.]

Source: © Wikibon 2014, from Numerous Sources including Analysts, Consultants, IBM & Oracle.
Existing funding under Horizon 2020 will allow to support the European Open Science Cloud and to kick-start the European Data Infrastructure. Initial estimation of the required additional public and private investment is €4.7 billion in the period of 5 years. This includes €3.5 billion for data infrastructure, €1 billion for a large-scale EU-wide Quantum Technologies flagship and €0.2 billion for actions on widening access and building trust. Additional
Open Data, Open Access, Open Innovation
An Underestimated Issue: Specific Features of Health Data

- Sensitive data (identifyable personal data)
- Highly regulated (data protection, ethics, health care)
- Potential high value, complex ownership
- Heterogeneous data (plain text in national language, imaging, laboratory, lifestyle, -omics etc.)
- Heterogeneous quality
- Data provenance, series (disease course, versioning, evolution)
- Complex access procedures
Analytical Technologies and Changing Relevance of Biobank Resources

FFPE Blocks: RT-PCR, NGS

Blood: Metabolomics, free circulating DNA

Issues: scanners, machine learning

Mind specific quality requirements!
What Is Next?

“The best way to predict the future is to invent it.”
- Alan Kay

from 100YSS
Thank You for Your Attention