Building a Biobank for Clinical Research

Roberto Caricchio, MD
Professor of Medicine
Chief, Section of Rheumatology
1. Why Your Job is Your Most Valuable Asset
2. Repurposing the Suburbs
3. Survival-Store Shopping
4. Biobanks; Saving Your Parts
6. The New Calvinism
7. Ecological Intelligence
8. Amortality; Forever Young
9. Africa; Open for Business
10. Reinventing the Highway
What are biobanks?

A collection of biological material and the associated data and information stored in an organized system, for a population or a large subset of a population.

Organization for the Economic Cooperation and Development

Organized collections and storage of human biological samples and associated data of great significance for research and personalized medicine.

BBMRI-European Commission (EC)
The rise of Biobanks

Human studies support the experimentally based notion of oxidative DNA damage as an important mutagenic and apparently carcinogenic factor. However, the proof of a causal relationship in humans is still lacking.....

.....This could possibly be supported by demonstration of the rate of oxidative DNA damage as an independent risk factor for cancer in a prospective study of biobank material using a nested case control design.....

Biobank in the scientific literature

Biobank in the scientific literature

Biobank in the scientific literature

• Understanding of genomic information and genetic mechanisms in diseases,

• Developments in IT sector and bioinformatics.
What are biobanks?

- Biological Human sample (Biospecimen)
- Attached or connected information
- Consent and patient data safety and protection
What are biobanks?

Bio-banking main components

- Biologic material handling
  - Biospecimen collection
  - Biospecimen processing
  - Biospecimen storage
  - Biospecimen distribution

- Database management system
  - Informed consent
  - Medical history and personal information
  - Life style data
  - Confidentiality
  - Security
The ideal Biobank

- Genetic Data
- Genealogical data
- Proteomic and metabolic data
- Lifestyle and environmental data
- Clinical Data
Centralized vs Federated biobanking

**Centralized**
- Central Biobank:
  - collection
  - processing
  - storage
  - access

**Federated**
- Virtual Database
- Researchers
- Biobank: collect, process, store
- Data
The process of biobanking

1. Consent
2. Biological sample
3. Associated information
4. Processing
5. Storage
6. Research/study
7. Results
8. Scientific publication
## Biospecimen

<table>
<thead>
<tr>
<th>Blood</th>
<th>Plasma</th>
<th>Serum</th>
<th>RBC</th>
<th>White cells</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>proteins</td>
<td>Cell lines</td>
<td>urine</td>
<td>CSF</td>
<td>SF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amniotic fluid</td>
<td>Buffy coat</td>
<td>BMSC</td>
<td>BM Tissue</td>
<td>Tissue</td>
<td></td>
</tr>
</tbody>
</table>
Storage in biobanks

Samples are stored in a way appropriate for the sample material and the intended research purpose.

• Blood, plasma, serum, and DNA are stored in -80°C freezers.

• Tissues and cell lines are preserved in liquid nitrogen freezers at -196°C.
Types of Biobanks

Human biobank classification is based on:

- **Tissue type** (tumor tissue, cells, blood, DNA or RNA);
- **Purpose/intended use** (research, forensics, transplantation, source for therapeutics, e.g., umbilical blood, stem cell biobanks for individual or community use, or diagnostics);
- **Ownership** (academic and research institutions, hospitals, biotechnology and pharmaceutical companies or government run).
Types of Biobanks

• Population-based biobanks
  focused on the study of development of common, complex diseases over time

• Disease oriented biobanks
  - biobanks of tissue samples and clinical data also referred to as disease oriented or clinical biobanks.

European Biobanking and BioMolecular resources Research Infrastructure (BBMRI)
Lifecycle of the Biospecimen

Specimen is viable and biologically reactive

Molecular composition subject to further alteration/degradation

Pre-acquisition

Post-acquisition
Biobanks requirements

- Consenting processes
- Collection services
- Storage infrastructure
- Specimen Access and Distribution
- Accounting
- Regulatory compliance
Biospecimen Availability and Quality: Major Challenge for Research

The lack of high quality, appropriately collected, stored and annotated specimen is the limiting factor for translational research or basic research.

High quality specimen enables:

- Drug discovery
- Translational Research and basic research
- Clinical development
- Assay variability
- Specimen variability
- Diagnostic development
- Personalized medicine
- Targeted therapies
Why is it so difficult to obtain high quality specimens and data?

- Collection, processing, storage procedures differ
  - Degree and type of data annotations varies
  - Scope and type of patient consent differs
  - Access policies are lacking or unknown to potential users
  - Materials transfer agreement conditions differ
  - Supporting IT structures differ in capacity and functionality
  - Funding sources varies
Pre-analytical Variables

Pre-acquisition variables:
- Antibiotics
- Other drugs
- Type of Anesthesia
- Duration of Anesthesia
- Arterial Clamp time
- Blood pressure variations
- Intra-op blood loss
- Intra-op blood administration
- Intra-op fluid administration
- Pre-existing medical conditions
- Patient gender
- Ethnicity

Post-acquisition variables:
- Time at room temperature
- Temperature of room
- Type of fixative
- Time of fixative
- Rate of freezing
- Processing method
- Size of aliquots
- Type of collection container
- Biomolecule extraction method
- Storage temperature
- Storage duration
- Storage in vacuum
The Ideal Work-flow

a. Sample collection
   Process Automation
   Aliquoting

b. Mechanical Freezers
   BIOBANK STORAGE
   Liquid nitrogen storage
   - 80 °C
   - 140 °C

c. Management Software
   BIOBANK SAMPLES SHARING
   MTA

   USERS:
   - Internal researchers
   - External researchers
All of US
**Table 1. Data Available to Researchers from the All of Us Cohort.**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current sources</strong></td>
<td></td>
</tr>
<tr>
<td>Health surveys</td>
<td>Initial surveys include information on sociodemographic characteristics, overall health, lifestyle, and substance use, with subsequent modules covering personal and family medical history and access to health care.</td>
</tr>
<tr>
<td>Physical measurements</td>
<td>Per-protocol measurements include blood pressure, heart rate, weight, height, body mass index, and hip and waist circumferences.</td>
</tr>
<tr>
<td>Biospecimens†</td>
<td>Blood and urine samples are tested for DNA, RNA, cell-free DNA, serum, and plasma. If blood specimens cannot be obtained, saliva specimens are obtained.</td>
</tr>
<tr>
<td>Electronic health records</td>
<td>Initial capture of structured data includes billing codes, medication history, laboratory results, vital signs, and encounter records from health care provider organizations. Records will be expanded to include narrative documents. Pilot studies are testing data collection through Sync for Science and other health data aggregators.</td>
</tr>
<tr>
<td>Digital health information</td>
<td>Data can be captured from compatible participant-owned devices such as Fitbit. Pilot studies of other devices and linkage to health apps are being explored.</td>
</tr>
<tr>
<td><strong>Future sources</strong></td>
<td></td>
</tr>
<tr>
<td>Health surveys</td>
<td>Additional modules, including surveys regarding social behavioral determinants of health, are under development.</td>
</tr>
<tr>
<td>Bioassays</td>
<td>Pilot studies for genotyping and whole-genome sequencing are expected to begin by early 2020. Additional pilot studies of bioassays are planned.</td>
</tr>
<tr>
<td>Health care claims data</td>
<td>Systems for the use of claims data, including billing codes and medication data, are under development.</td>
</tr>
<tr>
<td>Geospatial and environmental data</td>
<td>These data include geospatial linkage to measures such as weather, air quality, pollutant levels, and census data. Assays and sensor-based measurements of exposure are under consideration.</td>
</tr>
<tr>
<td>Other sources</td>
<td>Voluntary contributions of data from social networks (e.g., Twitter feeds) and additional biospecimen collections are under consideration.</td>
</tr>
</tbody>
</table>

* Additional information is available at https://www.researchallofus.org/data.
† Types of biospecimens are listed in Table S3 in the Supplementary Appendix.
### Table 2. Scientific Goals of the All of Us Program and Expected Timelines.*

<table>
<thead>
<tr>
<th>Goal</th>
<th>End of 2018 (N = 94,000)</th>
<th>End of 2019 (N = &gt;200,000)</th>
<th>2020–2022 (N = &lt;650,000)</th>
<th>2023–2027 (N = &gt;1 million)</th>
<th>After 2027 (N = &gt;1 million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return data to participants</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
</tr>
<tr>
<td>Establish demonstration projects†</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Discover genetic and environmental correlates with disease</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
</tr>
<tr>
<td>Improve predictions of therapeutic safety and efficacy</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
</tr>
<tr>
<td>Discover disease biomarkers</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Connect mobile health, digital health, and sensor data with clinical outcomes</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Develop new disease classifications</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>++++</td>
</tr>
<tr>
<td>Support clinical trials</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Enable machine-learning applications</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
</tr>
<tr>
<td>Improve understanding of health disparities</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++++</td>
</tr>
<tr>
<td>Develop and test new therapeutic agents</td>
<td>+</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
</tbody>
</table>

* The expected number of participants in the cohort is shown for each time period. The number of plus signs in each cell indicates the anticipated relative degree to which each goal may be accomplished during the estimated timeline for focused research.

† Demonstration projects are scientific studies implemented by the All of Us program to show the quality, usefulness, validity, and diversity of the All of Us research data set and platform. In these projects, the population and data are further characterized, and the data are evaluated with a view to determining whether known associations can be replicated.
The Temple Lupus Cohort

- Born from the need of a systematic database of Lupus patients to improve Clinical Trials and Quality of Care.

- Implemented along with the electronic medical record (EMR) which allows to collect systematic and consistent information on lupus patients at each visit.

- Integration of SLE disease activity (SLEDAI) and SLE damage index (SDI) measure outcomes with electronic charts during clinic encounters since November 2010.

- Systematic collection of blood, urine samples and renal tissue to foster translational research and clinical trials.
Disease scores, lab results, imaging and disease-specific medications are tracked over time in the Rheumatology Synopsis.

Medications courses with start and stop dates.

Provider and patients can see graphic representation of disease activity, and correlate with treatments.
SLE patient and healthy individuals have anti-Curli antibodies
Frequency of bacteriuria correlate with anti-curli/eDNA antibodies
<table>
<thead>
<tr>
<th>Collaborators</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>New York University</td>
<td>Janssen</td>
</tr>
<tr>
<td>Penn State</td>
<td>LuCIN</td>
</tr>
<tr>
<td>LKSoM</td>
<td>NIH</td>
</tr>
</tbody>
</table>
Conclusion

• Biobanks are complex systems of systematically programmed storage of human material and associated data.

• In the past 20 years the science of biobanks has became an integral part of personalized medicine

• A great number of biobanks have been established all over the world to support the dramatic development in diseases prevention, prediction, diagnosis and treatment.
Acknowledgements

• Caricchio Laboratory
  • Ryan Pachuki.
  • Xinyan Zhang
  • Lynne Kohler

• Rheumatology Fellows

• Lupus Patients