The challenges of accessing samples and data for research

Claude Chelala
Bioinformatics Unit, Centre for Molecular Oncology
Post-TCGA era and biobanking

- Value for participants and researchers
- Support the next generation of –omics research questions
- Support longitudinal studies, specimens and data
- Promote data sharing and collaborations
Data integration for patient benefit

• Biobanking ecosystem for personalised medicine
• Interlinking clinical, molecular & *in silico* resources:
  – Bioinformatics: Breast Cancer Now Tissue Bank
  – Bioinformatics and IT: Pancreatic Cancer Research Fund Tissue Bank (PCRFTB)
Pancreatic Cancer Research Fund Tissue Bank (PCRFTB)

• A unique resource of biological materials and clinical data from patients as well as a bioinformatics backing to support cutting-edge research for the benefit of patients.

'New research hope' from pancreatic cancer tissue bank

By Jane Dreaper
Health correspondent, BBC News

14 January 2016  Health
• A research tissue bank to support research for the benefit of pancreatic cancer patients.
  – Collect, store and use specimens under a single HTA license and ethics approval
  – Standardised consent and information sheets for the donors
  – Longitudinal samples available to scientific community
  – SOPs in place from donor recruitment through sample/data collection and processing to sample distribution

• Aims:
  – Accelerates research for early diagnosis and novel therapies.
  – Promote integrative research combining clinical data with genomics data.
PCRF Tissue Bank (PCRFTB)

Sample Repository

Primary sample: Serum, plasma, urine, saliva, fixed & frozen tissue
Derivatives: DNA, RNA, Protein, CTC
Super-derivatives: Tissue microarray

Patient Registry

Auditable clinical annotation database

Bioinformatics Backbone

Support data return and subsequent data mining and analysis
PCRF Tissue Bank (PCRFTB)

Sample Repository
- Governance, regulations, management,
  Sustainability, standardised and harmonised procedures
- Public, patients and clinical engagement
- Retrospective/prospective collections

Patient Registry
- Data completeness, Data security
- Clinical data linkage and harmonisation
- Data integrity
- Tissue Request and allocation system

Bioinformatics Backbone
- Data quality, Data security,
  Confidentiality, heterogeneity, and data sharing
# Patients Registry

## Data Portal

### Patient Details

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Patient (primary)</td>
</tr>
<tr>
<td>Subject Id</td>
<td>Select..</td>
</tr>
<tr>
<td>NHS number</td>
<td></td>
</tr>
<tr>
<td>Initials</td>
<td></td>
</tr>
<tr>
<td>Date of birth</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Unknown</td>
</tr>
<tr>
<td>Height</td>
<td>CM</td>
</tr>
<tr>
<td>Weight</td>
<td>Kilogram</td>
</tr>
<tr>
<td>Date of consent</td>
<td></td>
</tr>
<tr>
<td>Is the consent full?</td>
<td>Yes</td>
</tr>
<tr>
<td>Patient alive?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Update & Continue] [Skip >>] [Exit]
Sample Repository and Patients Registry

- Real-time Web-based data entry and retrieval system for users.
- Tissue collection procedure is complemented by collection of clinical and pathological data.
- A purpose-built clinical data structure is in place to capture patient healthcare history.
Sample Repository and Patients Registry

- Editor: Tissue Collection Officer
  - Donor interview
  - Manual extraction of CRS data (relevant to specific visit only)

- Auditor: Clinical Research Fellow
  - Verification of interview and CRS data
  - Manual extraction of CRS data (relevant to specific visit only)

- Administrator: Data Manager, Coordinator, PI
  - Quality control
Data Content: What we have

Single patient clinical journey

DATA

Specimen

Treatment

Scan

Blood test

Medical History

TIMELINE

2014  2015  2016  2017
Data Content: What we want

- Primary care data
  - Before and after tertiary care
- Uncollected secondary/tertiary care data
- Validation of collected tertiary care data
Barts Pancreas Tissue Bank (BPTB)

- BPTB has been used to model and tests systems for the national collaborative venture
- Samples are mainly collected from patients at the Royal London Hospital, and hosted at Barts Cancer Institute.
BPTB Data Source

- **Editor: Tissue Collection Officer**
  - Donor interview
  - Manual extraction of CRS data (relevant to specific visit only)

- **Auditor: Clinical Research Fellow**
  - Verification of interview and CRS data
  - Manual extraction of CRS data (relevant to specific visit only)

- **Administrator: Data Manager, Coordinator, PI**
  - Quality control
  - **Automated extraction of longitudinal EHR**
    - Add verified so-far-complete healthcare trajectory data associated with the samples.
    - Unlock the potential of longitudinal EHRs in a consented Tissue Bank
    - Allow applicant researchers to design innovative studies and build new insights on pancreatic cancer
Bioinformatics Challenges

• Quality and format of the data: Technical challenges.

• Tackling the resistance to data sharing in the biobanking community: challenge due to concerns that a secondary user would misinterpret the data, and political and social factors, such as the structure of incentives and the inherent competitiveness in biomedical research.
PCRFTB: Data Generation

**Project “A”**

- Blood
- Urine
- Tissue

Samples → -omics data

<table>
<thead>
<tr>
<th>Patient 1</th>
<th>Patient 2</th>
<th>Patient 3</th>
<th>Patient n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteomics</td>
<td>Mutation</td>
<td>Copy</td>
<td>number</td>
</tr>
<tr>
<td>Gene</td>
<td>expression</td>
<td>Proteomics</td>
<td>Mutation</td>
</tr>
<tr>
<td>Gene</td>
<td>expression</td>
<td>Gene</td>
<td>expression</td>
</tr>
</tbody>
</table>

Project “A” involves collecting samples from patients and generating -omics data. Patients 1, 2, 3, and n are involved in the project.
Project “A”

Project “B”

Project “C”

Project “X”

Bioinformatics platform

Reduce duplication
Enhance sample use
Promote integrative research
Bioinformatics support
PED: Bioinformatics platform for PCRFTB

• URL: www.pancreasexpression.org
• Four releases so far in BMC Genomics 2007, Nucleic Acids Research: 2011, 2014 and 2018
## Available Analyses

<table>
<thead>
<tr>
<th>Exploratory</th>
<th>Investigative</th>
<th>Interpretative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Principal component analysis</td>
<td>• Molecular profiling</td>
<td>• Survival</td>
</tr>
<tr>
<td>• Tumour purity</td>
<td>• Expression plots</td>
<td>• Gene network</td>
</tr>
<tr>
<td></td>
<td>• Correlation</td>
<td>• Data integration</td>
</tr>
<tr>
<td></td>
<td>• Copy number alterations</td>
<td>– expression</td>
</tr>
<tr>
<td></td>
<td>• Mutational profiles</td>
<td>– copy number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– mutation</td>
</tr>
</tbody>
</table>
Available Analyses

- Exploratory
- Investigative
- Interpretative
PCRFTB Query: Bioinformatics connectivity
.... And allow researchers to ask useful questions

• How many PDAC donors with at least 3 longitudinal plasma samples are available from the BPTB?
• Show me their clinical data summary
• Have any of the donors I selected been subject to whole genome sequencing?
• Show me the landscape of mutations in the KRAS signalling pathway
• Show me the expression level of my genes of interest
• Explore their correlation, related networks or prognostic potential
PCRFTB IT/Bioinformatics

• A unified research platform
  – Reduce duplication of effort and costs
  – Maximise data sharing
  – Enrich clinical data with new attributes (molecular data, mutations, new biomarkers etc.)
  – Accelerate research
  – Maximise research output on samples for patient benefit
Acknowledgements

Bioinformatics Unit
Dayem Ullah
Ope Banowa
Stefano Pirro
Emanuela Gadaleta
Helen Ross Adams
Lavanya Sivapalan
Pauline Fourgoux
Jorge Oscanoa
Henry Li
Graeme Thorn
Jacek Marzec
Syed Haider
Ajanthah Sangaralingam
Ros Cutts

Barts Cancer Institute
Nick Lemoine
Hemant Kocher

Centre for Computational Biology, Life Sciences Initiative, QMUL