Biobanks and EMR:s – a perfect marriage?

Olli Carpén
Helsinki biobank
• Why combine biobanks and EMRs
  • Personalized medicine

• Infrastructure requirements
  • Societal, organizational

• Experiences from hospital integrated large scale biobanks

• Digital phenotyping
Implementation of PM will not happen without tools to translate novel information into health care

CURRENT
“reactive medicine”

• Evidence based = slow changes

FUTURE
“P4 medicine”

Methods for earlier

Population based high quality biological specimens and electronic medical records play a key role in the transition = deeper phenotyping, “real life data”

Better trial design by cohort stratification based on clinical and molecular features

Population based high quality biological specimens and electronic medical records play a key role in the transition = deeper phenotyping, “real life data”
YOU ONLY WIN WITH A COMPREHENSIVE PACKAGE!
Integrated pathway to personalized medicine

Biology (OMICS)

Health (Phenotype)

Contribution by Jaana Sinipuro (Sitra)
Fully hospital integrated - streamlined and affordable consenting and sample collection

- Biobank consent as part of hospital routine
- Sample collection by accredited procedures
- DNA purification by accredited procedures
- Sequencing and reporting by clinical diagnostic standards

- Medical research, New health care modalities, R&D

EVERY PATIENT HAS THE RIGHT OF BEING A RESEARCH PATIENT
The details of disease

Precision medicine demands precise matching of deep genomic and phenotypic models — and the deeper you go, the more you know.
Disease trajectories provide understanding of individual variation within disease categories

A step towards digital phenotypes

Samu Kurki
Combining biological and EHR information in ovarian cancer

- Ovarian cancer causes more fatalities in than any other gynecological malignancy (in Western world)
- Current treatment is surgery combined with platinum-taxane chemotherapy
- Not all women respond to treatment and most relapse becoming resistant to therapy
- **No diagnostic tests available for outcome or response prediction**

**Need to identify responsive patient cohorts:**
- Save patients from unnecessary treatments
- Provide the best treatment option
- Save money and unnecessary suffering
- Better stratification for clinical trials
"Ovarian cancer is a genomic mess" – how can we identify clinically meaningful biomarkers by combining biology with EHR information?

Conventional approach:
1. Identify disease subtypes on the basis of genetic or other biomarkers
2. Correlate genetic groups (GG) with outcome (survival etc.)

Alternative approach:
1. Identify clinical feature combinations that group individuals based on disease trajectories
2. Use these biologically meaningful groups (BMG) to stratify patients for biomarker search.

Conventional approach:

![Conventional approach diagram]

Alternative approach:

![Alternative approach diagram]
### Feature Importance

<table>
<thead>
<tr>
<th>Feature</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum CA125 value</td>
<td>0.291</td>
</tr>
<tr>
<td>The longest remission interval</td>
<td>0.251</td>
</tr>
<tr>
<td>Cumulative CA125 (integral)</td>
<td>0.180</td>
</tr>
<tr>
<td>Number of responses</td>
<td>0.122</td>
</tr>
<tr>
<td>Comorbidity, based on diagnosis</td>
<td>0.079</td>
</tr>
<tr>
<td>Speed of response</td>
<td>0.077</td>
</tr>
</tbody>
</table>

### All features prediction by kmeans

<table>
<thead>
<tr>
<th>known classification</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died 1y</td>
<td>7</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>Died in 5y</td>
<td>167</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td>Survived over 5y</td>
<td>72</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Jani Salmi
Prognostic subtypes in HGS-OvCa patients treated with platinum-taxane

- **Left**: 67 transcripts that divide patients into three prognostic subtypes
- **Middle**: When combined with **BRCA mutation** status, the transcript profile identifies patient groups with very different outcome expectations: *e.g.* extreme responders (Good-BRCA mut, green) and primarily chemorefractive patients (Poor-BRCA WT, purple)
- **Right**: Poor I primary cells (OC002) respond to demethylating agents

Summary

• Large scale hospital biobanking, by combining biological and longitudinal comprehensive phenotype information, provides a perfect testbed for discovery and development.

• For optimal biobank performance, a platform which includes longitudinal (both retrospective and prospective) EHR information, and capabilities to procure and utilize it, is required.

• The real world evidence, combining large scale biological and phenotypic information, paves the path towards stratified medicine.