Sharing cancer data using the blockchain
The promise of precision medicine in oncology

Zhang et al., Precision Clinical Medicine, 2018
Cancer involves multiple layers of the “phenome”

Topol EJ, Cell. 2014
Electronic Health Records (EHR) capture many such layers

**Data type examples:**

**Clinical**
- Diagnoses
- Procedures
- Lab test results
- Imaging
- Medications
- Notes

**Non-clinical**
- Demographics
- Insurance
- Location
- Lifestyle
We can gain insight from these data: **towards a learning health system**

---

*Fig. 1| A deep-learning healthcare system.* A schematic representation of a deep-learning healthcare system is shown.
Valuable data are locked in silos

Pathology
Radiology
Surgery
Ob/Gyn
Patient Reported Outcomes
ER

From Data Silos to Standardized, Linked, and FAIR Data for Pharmacovigilance: Current Advances and Challenges with Observational Healthcare Data

Vassilis Koutkias

© Springer Nature Switzerland AG 2019
...and these data are not easy to work with

• EHRs are challenging to represent health state
  o heterogeneous
  o noisy
  o incomplete
  o structured / unstructured
  o redundant
  o subject to random errors
  o subject to systematic errors
  o ...and so and so forth
For cancer, these data are needed to fill in gaps!

Genomics

PERSISTENT BIAS

Over the past seven years, the proportion of participants in genome-wide association studies (GWAS) that are of Asian ancestry has increased. Groups of other ancestries continue to be very poorly represented.

2009
373 studies
1.7 million samples

2016
2,511 studies
35 million samples

96% European ancestry
81% European ancestry
4% Non-European ancestry
19% Non-European ancestry


Clinical trials

Participation in Cancer Clinical Trials
Race-, Sex-, and Age-Based Disparities

Vivek H. Murthy, MD, MBA; Harlan M. Krumholz, MD, SM; Cary P. Gross, MD

Author Affiliations

Table 1. Participants in National Cancer Institute Cooperative Group Breast, Colorectal, Lung, or Prostate Cancer Therapeutic Trials, 1996-2002 (N = 75,215)*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Trial Participants, No. (%)</th>
<th>Proportion of Incident Cancer Patients, %†</th>
<th>Proportion of US Population, %†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White non-Hispanic</td>
<td>64,355 (65.6)</td>
<td>83.1</td>
<td>76.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2,392 (2.1)</td>
<td>3.8</td>
<td>9.1</td>
</tr>
<tr>
<td>Black</td>
<td>6,882 (6.2)</td>
<td>10.9</td>
<td>10.8</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1,448 (1.5)</td>
<td>2.6</td>
<td>3.8</td>
</tr>
</tbody>
</table>

These data exist! But are silo’d
The blockchain can help solve this need

- De-centralized
- Flexible
- Immutable
- Free

Implementing Blockchains for Efficient Health Care: Systematic Review

Anuraag A Vazirani1, Odhran O’Donoghue1, David Brindley1, MEng, MSc, DPhil, FRSA; Edward Meinert2, MA, MSc, MBA, MPA, CEng FCIBS.

1Healthcare Translation Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom
2Global Digital Health Unit, Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

https://www.cbinsights.com/research/what-is-blockchain-technology/
Introducing the Cancer Gene Trust (CGT)

- Sharing **de-identified** clinical, imaging, and genomic data for **consented** patients
- Files are shared via an off-blockchain store (IPFS) by **stewards**
Rigorous IRB process

IRB Approval for CGT
10-month approval process

Initial Comments
“Innovative and potentially important”
1. Need objective measures to assess the success of sharing (as listed in primary endpoint).
2. Details necessary on the gathering and de-identification of data

Resubmission & 2nd comments
1. Further clarify the reason and aims of this study
2. Clearly identify data uploaded and information not kept
3. Re-write for patient-friendly language (8th grade level)

Resubmission & 3rd comments
1. Clearly state risks for sharing confidential medical information
2. Detail safeguards for de-identifying data, importing data, and maintaining database security

Final IRB Approval
Who is part of this pilot trial?

- The PI of the study (Eric Collisson, MD) has recruited **18** cancer patients
- All patients consented to have their data shared via CGT
What data are shared?

- Genetic (somatic) variants from biopsy (e.g., *BRAF v600e mutation*)
- Relevant image data (e.g., CT-scan)
- Relevant Clinical EHR data (e.g., demographics, diagnoses, drugs)

### SEER Registry

- Standardized form
- Cancer-specific
- Manual and prospective data collection

https://seer.cancer.gov/

### OMOP EHR data

- EHR data transformed into *common data model*
- Standardized vocabulary and structure
- Disease-agnostic
- Retrospective data

https://www.ohdsi.org/
What format of EHR is most robust to share?

- We compared data contained with patients’ SEER and OMOP records to their “true” results in EPIC based on gold-standard variables (Conley et al.):

<table>
<thead>
<tr>
<th>Core Clinical Data Elements for Cancer Genomic Repositories: A Multi-stakeholder Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robert B. Conley,1 Dane Dickson,2,4 Jean Claude Zenklusen,3 Jennifer Al Naber,1 Donna A. Messner,1 Aljan Atasoy,4 Lena Chalikovsky,1 Deborah Collyar,5 Carolyn Compton,7 Martin Ferguson,4 Sean Khoozin,8 Roger D. Klein,9 Sri Kotte,10 Razelle Kurzrock,11 C. Jimmy Lin,10 Frank Liu,12 Ingrid Marino,13 Robert McDonough,14 Amy McNeal,15 Vincent Miller,13 Richard L. Schilsky,16 and Lisa I. Wang17</td>
</tr>
</tbody>
</table>
SEER vs. OMOP: how accurate are they?

- While OMOP had slightly higher scores overall:
  • SEER (mean 29.5±11.5) vs. OMOP (mean 32.1±5.4)
  • We found no significant difference between them (p = 0.38)

- We did, however, find OMOP was slightly more accurate for reporting:
  • Ethnicity (p = 0.01)
  • Cancer site and histology (p = 0.01)

- Generally more variability in SEER than OMOP

<table>
<thead>
<tr>
<th>Patient</th>
<th>SEER</th>
<th>OMOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>43</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>15</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>17</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>total</td>
<td>501</td>
<td>545</td>
</tr>
</tbody>
</table>
Standardized EHR formats allow for seamless cross-institution integration

Model/Results

Data

OMOP

Model/Results

Data

Cancer Gene Trust

EPIC

Cerner
The current CGT framework (alpha)

https://www.cancergenetrust.org/

- Clinical data
- Genomics
- Images
Data within pilot CGT release

• **n = 18** patients consented and enrolled
• 17 with **genomic data** (13 with Foundation One; 4 with UCSF 500)
• 2 with **image data**
• 17 with available **OMOP data:**

  • **Conditions:** **11,457 total events** (mean: 674; max: 2,389)  
    (e.g., “Malignant neoplasm of gallbladder”)
  • **Procedures:** **6,935 total events** (mean: 408; max: 1,394)  
    (e.g., “Biopsy, lung or mediastinum, percutaneous needle”)
  • **Drugs:** **15,058 total events** (mean: 886; max: 3,661)  
    (e.g., “trametinib 2 MG Oral Tablet”)
Example use case: patient 17

Gender: Female
Age: 41
Race: Asian

BRAF v600e mutation

PET-CT: FGD avid Hepatic mass, upper abdominal adenopathy, supraclavicular adenopathy

FOLFOX started
XRT: supraclavical mass

PET-CT:
FGD avid Hepatic mass, upper abdominal adenopathy, supraclavicular adenopathy
PatientExploreR: an app to explore, visualize, and download CGT data

GitHub (general, non-CGT version):
https://github.com/BenGlicksberg/PatientExploreR
Limitations

• All patients run the risk of possible re-identification

• No way to automatically identify primary cancer & therapeutic efficacy

• Current implementation not complete: missing valuable EHR data

• Submission is *steward-based*

• Need a system in place to link *same* patient from different stewards

• Data quality is limited by those who enter it
Future Steps

• Better data: Refine OMOP extracts (including measurement data and notes)

• More (diverse) data: Data from more sites

• Enhance PatientExploreR functionality

• Automated image analysis via deep learning
Rate today’s session

Cyberconflict: A new era of war, sabotage, and fear

9:55am - 10:10am Wednesday, March 27, 2019
Location: Ballroom
Secondary topics: Security and Privacy

We're living in a new era of constant sabotage, misinformation, and fear, in which everyone is a target, and you're often the collateral damage in a growing conflict among states. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. Moving from the White House Situation Room to the dens of Chinese, Russian, North Korean, and Iranian hackers to the boardrooms of Silicon Valley, David reveals a world coming face-to-face with the perils of technological revolution—a conflict that the United States helped start when it began using cyberweapons against Iranian nuclear plants and North Korean missile launches. But now we find ourselves in a conflict we’re uncertain how to control, as our adversaries exploit vulnerabilities in our hyperconnected nation and we struggle to figure out how to deter these complex, short-of-war attacks.

David Sanger
The New York Times
Acknowledgements

• Shohei Burns

• Rob Currie

• Ted Goldstein, PhD

• Eric Collisson, MD (PI)

Funding: The Marcus Family (Marcus Program is Precision Medicine Innovation)