A Cautiously Optimistic View of the Future of Data-driven Genomically Informed Medicine

Isaac S. Kohane, MD, PhD
Overview

- Where we were
- Where we are heading now
- Where we could be.
Fundamental Medical Challenge: True Names of Disease

Ordo 1.

PRIMATES.

Dentes primores superiores IV paralleli. Mammæ pectorales, binae.

1. HOMO nosee Te ipsum.

1. H. diurnus. (*) vagans cultura, loca.
   a. H. rufus, cholericus, rectus.
   f. H. albus, sanguineus torōlus.
   γ. H. luridus, melancholicus rigidus.
   d. H. niger, phlegmaticus, laxus.
   t. H. montrosus folo (a), vel arte (b. c.)
      a. Alpini parvi, agiles, timidi: Patagonici magni, segnes.
      b. Monorchides ut minus ferriles: Hottentotti.
      Juncet puella abdomine attenuato: Europæae.
      Plagiocephali capite antice compresso. Canadenses.

2. Homo nocturnus. Oorang Oorang Bont. iuv. 84. r. 84.

Genus Trogloditæ seu Oorang Ooutang ab Homine vero difficilium, additum quovis omnis attionen, obtinere non potui, nisi admerem nostrum lubrican, in aliis generibus non constantem. Ne Dentestianii minus a reliquis remoti, nec Nympheæ caffar, quibus enant Simias, kume ad Simias reducere adimitabant. Inquirant atque in vivo, qua ratione, modo nostræ aliquæ existant, ab Huinis genere separari queat, nam inter Simias veritantem oportet esse Simian. Apollodor.
Narrative data (NLP text extractions)

Codified data (ICD9 codes, etc)
Cohorts
Subset 1: (\Public Studies\EGP0001\Biomarker Data\Non Omics\Mutation Detection\KRAS Mutation\NA\) 

**Analysis:** Survival Analysis

**Variable Selection**

**Time**
Select time variable from the Data Set Explorer Tree and drag it into the box. For example, "Survival Time". This variable is required.

**Category**
Select a variable on which you would like to sort the cohort and drag it into the box. For example, "Cancer Stage". If this variable is continuous (e.g., Age), then it should be "binned" using the option below. This variable is not required.
• R module in tranSMART

• Published figure in JCO
Common-Rare:Weak-Strong Spectrum

10^6 patients

Tay-Sachs

ASD

DM2

10^8 patients

Huntington’s

CAD

MODY

10^4 diseases

Deterministic “highly penetrant”

10^2 diseases

Weak effect, not predictive, dominated by environment
Needs

Rare &
COMMON

BIG

DATA
SHRINE: Governance over Technology

- Search routine clinical records from 5 major hospitals for:
  - Demographics
  - Diagnosis
  - Medications
  - Lab Results
- Reach N
  - Rare Dx
  - Small Effects
- 10 billion FACTS
- 6 million patients
Finding rare events of interest.

Cardiac angiogenic imbalance leads to peripartum cardiomyopathy

Ian S. Patten1,2,3, Sarosh Rana4,5, Sajid Shahul6, Glenn C. Rowe1, Chooson Jang1, Laura Liu1, Michele R. Hacker1, Julie S. Khee1, John Mitchell1, Feroze Mahmood7, Philip Hess4, Caitlin Farrell1, Nicola Koukis2, Eliyahu V. Khankin7, Suzanne D. Burke1, Igor Tudorache1, Johann Battezachs2, Federica del Monte3, Denise Hiltifer-Kleiner2, S. Ananth Karumanchi3,8 & Zoltan Arany1

17 MAY 2012 | VOL 485 | NATURE | 333

samples from subjects with PPCM have been previously described3. Patients in both studies were predominantly Caucasian. Retrospective analyses of PPCM and pre-eclampsia in the Harvard teaching hospitals were performed using the Harvard [Shared Health Research Information Network (SHRINE)]3, a de-identified repository of aggregate patient information.
Where is the missing information?

\[ V_P = V_G + V_E + V_{GE} \]
“Big Data” Map of Patient Health

Courtesy
Griffin Weber
La Jornada

Veracruz: reporta agente municipal extraño brote epidémico que ha cobrado dos vidas

La funcionaria de La Gloria informó que el raro padecimiento ha afectado a 60 por ciento de sus tres mil habitantes con infecciones respiratorias.
Wednesday, Sep. 09, 2009

Is a Swine Flu Outbreak Coming? Ask Your iPhone

By Bryan Walsh
iPhone Submissions vs CDC sentinel surveillance

R^2=0.74
CLARITY competition
(Crowdsourced discovery)
http://genes.childrenshospital.org
• Three families each with one affected child
• Nine exomes (Life) and genomes (CompleteGenomics)
• Three family clinical summaries
• $25K prize
• 30 international (commercial and academic) contest entrants!
Unexpected contest outcomes
Fast forward 10 years...
2023: Every patient encounter is learning opportunity (for provider and patient)
2023
Yearly check up includes exposure scan

WGS data interpreted in light of interact with environmental factors and current clinical status.

<table>
<thead>
<tr>
<th>rs13266634(SLC30A8)</th>
<th>OR (95% CI)</th>
<th>N(cases):</th>
</tr>
</thead>
<tbody>
<tr>
<td>trans-β-carotene (low(-1SD))</td>
<td>1.8 [1.3, 2.6]</td>
<td>1699(167)</td>
</tr>
<tr>
<td>trans-β-carotene (mean)</td>
<td>1.1 [0.8, 1.5]</td>
<td></td>
</tr>
<tr>
<td>trans-β-carotene (high(+1SD))</td>
<td>0.67 [0.41, 1.1]</td>
<td></td>
</tr>
<tr>
<td>p-value (FDR):5e-05 (0.015)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rs2237895(KCNQ1)</th>
<th>OR (95% CI)</th>
<th>N(cases):</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCB170(low(-1SD))</td>
<td>0.44 [0.21, 0.93]</td>
<td>1009(88)</td>
</tr>
<tr>
<td>PCB170(mean)</td>
<td>0.61 [0.34, 1.1]</td>
<td></td>
</tr>
<tr>
<td>PCB170(high(+1SD))</td>
<td>0.85 [0.5, 1.5]</td>
<td></td>
</tr>
<tr>
<td>p-value (FDR):0.023 (0.24)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chirag Patel
Healthcare system no longer major source of trial patients.

FlashTrials.gov/.com (cf. FlashMob) composed on the fly match volunteers to appropriate therapeutic trials. New study designs enable randomized partial trial participation.
In era of big data, well-characterized patients are the only non-commodity.

- WGS is no more expensive than electrolyte panel.
- BLUE button liquifies all data in any healthcare institution on behalf of the patient.
- Lifetime record is maintained by third party bonded data agency.
  - Includes everything we currently think of as health related but only the patient or guardian can provide access.
2023, more health data generated in 1 home than an ICU bed in 2013
Pathophys. Precia
- Nystagmus
- Diaphoresis
- Tropic Thunder
- Less than Zero

Medications
- Blue pill
- Red pill
- Bitter pill

DDx
- ADHD
- Narcissism
- Terminal technophilia

Social History
- Etoh+
- Nicotine+++ 
- WD40++++
What are the questions/challenges?

• Who is going to curate these data?
• What is the vector of FHx and genomic data?
• Is there a role for the healthcare system?
• How do we create a larger market for the information commons?
• What are we going to do to uniquely identify patients.