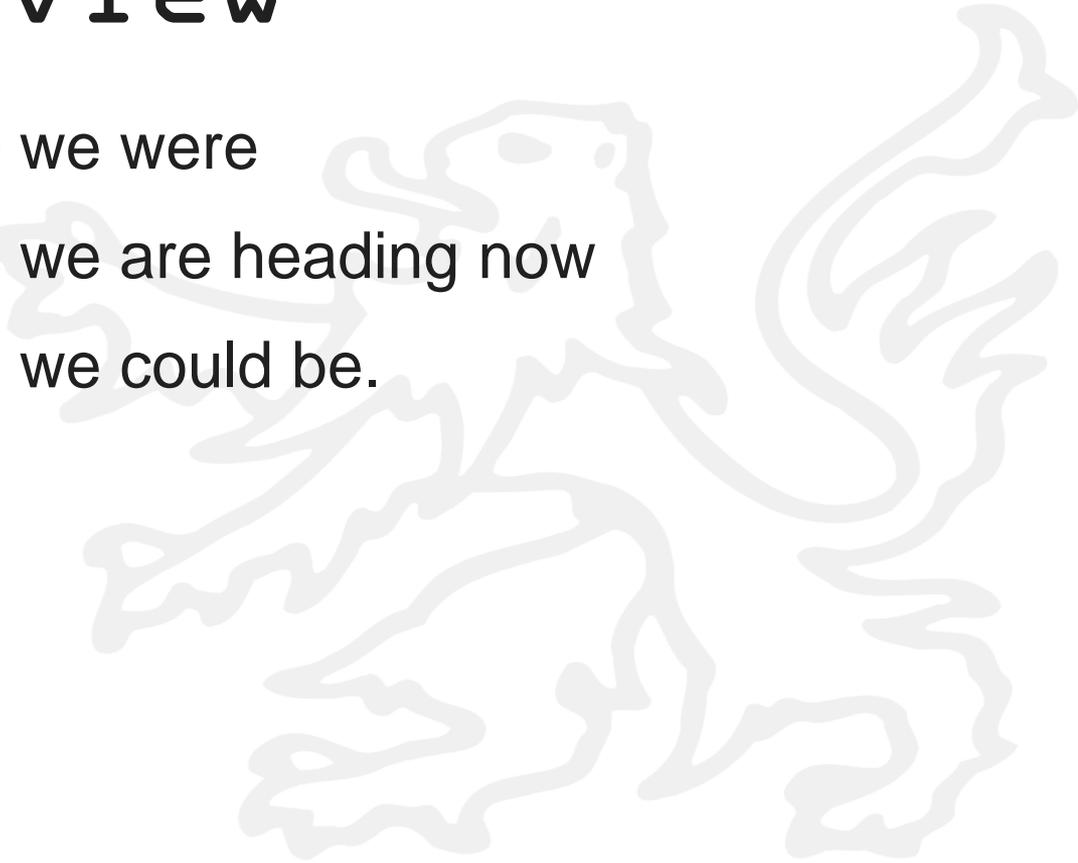


# A Cautiously Optimistic View of the Future of Data-driven Genomically Informed Medicine

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 I.

## PRIMATES.

Dentes primores superiores IV paralleli.  
Mammæ pectorales, binæ.

### I. HOMO nosce Te ipsum.

1. H. diurnus. (\*) *vagans cultura, loco.*

a. H. rufus, cholericus, rectus.

β. H. albus, sanguineus, torosus.

γ. H. luridus, melancholicus, rigidus.

δ. H. niger, phlegmaticus, laxus.

ε. H. monstruosus solo (a), vel arte (b. c.)

a. *Alpini parvi, agiles, timidi: Patagonici magni, segnes.*

b. *Monorchides ut minus ferriles: Hottentotti.*

*Juncæ puellæ abdomine attenuato: Europæ.*

c. *Macrocephali capite conico. Chineses.*

*Plagioccephali capite antice compresso. Canadenses.*

2. Homo nocturnus. *Ourang Outang Bont. jav. 84. t. 84.*

*Genus Trogloditæ seu Ourang Outang ab Homine vero distinctum, adhibita quamvis omni attentione, obtinere non potui, nisi assererem notam lubricam, in aliis generibus non constare. Nec Dentes lanarii minime a reliquis remoti; nec Nymphae castræ, quibus carent Stivæ, hunc ad Simias reducere admittebant. Inquirant exacte in vivo, qua ratione, modo notæ aliquæ existant, ab Hominis genere separari queat, nam inter Simias versantem oportet esse Simiam. Apollodor.*

Americanus.  
Europæus.  
Asiaticus.  
Afer.



# Narrative data (NLP text extractions)

The screenshot displays the i2b2 Workbench interface for Rheumatoid Arthritis. The top navigation bar includes 'Navigate', 'Find Terms', 'Query Tool', and 'Correlation Analysis Cell'. The user is identified as Robert Plenge, MD, PhD, with a green status indicator and a 'Wiki' link.

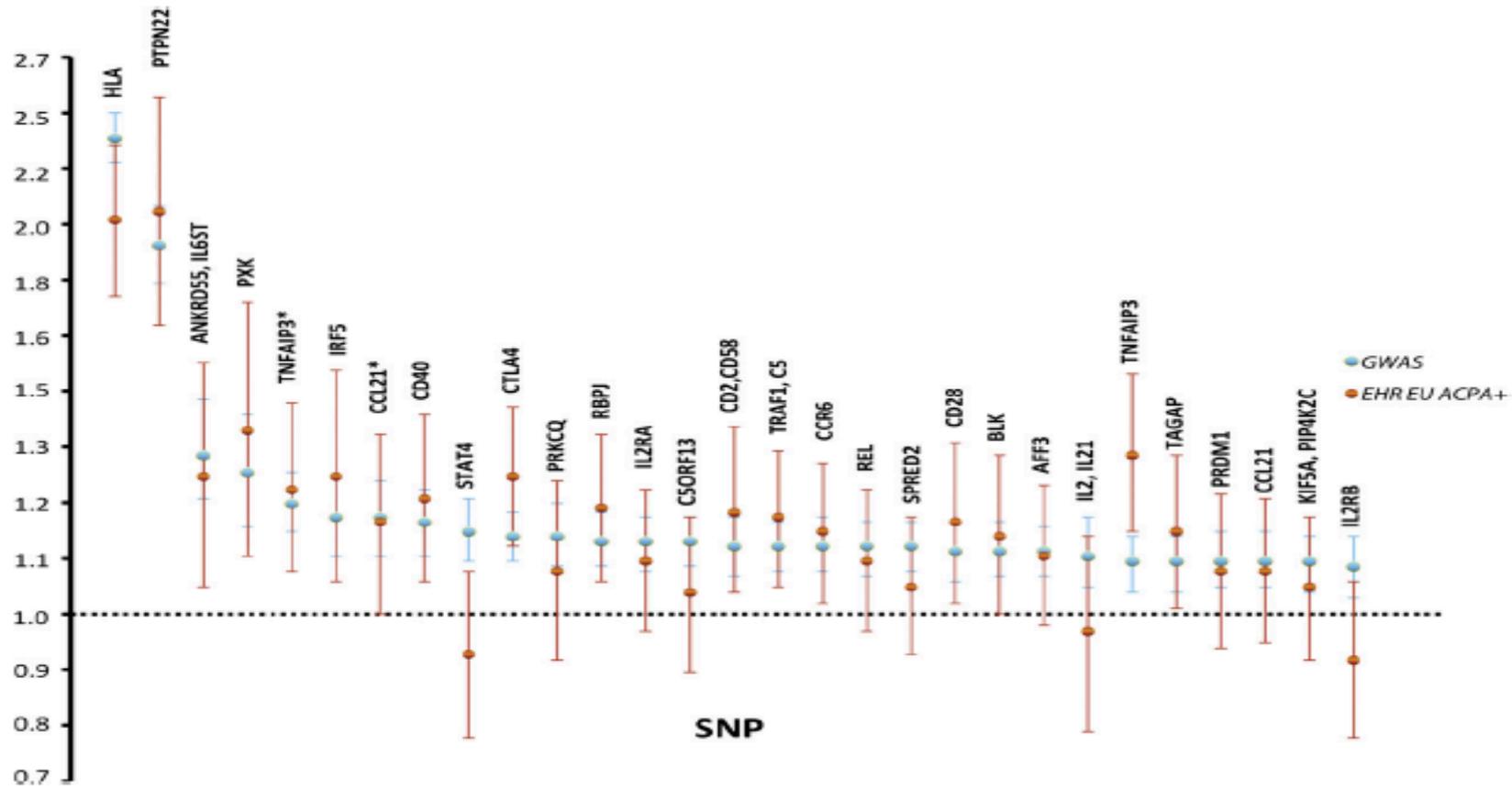
The central 'Query Tool' section shows a query named 'True\_3-ACR Cri@09:32:30'. It features three columns for 'Group 1', 'Group 2', and 'Group 3', each with 'Dates', 'Occurs > 0x', and 'Exclude' options. The 'Group 2' column is populated with categories such as 'ACR Criteria', 'Diseases', 'Erosions', 'Labs', 'Medications', 'LMR Notes', 'BWH Arthritis Center', and 'Rheumatoid arthritis and other i'. A 'Run Query' button is located at the bottom of this section, and a status message indicates 'Patient(s) returned: 2641'.

Below the query tool is the 'Timeline View' section, which includes buttons for 'Create model for Timeline' and 'Render a Timeline'. The main area of the timeline view displays a grid of data points for various categories: 'ACR\_Criteria', 'Diseases', 'Erosions', 'Medications', 'LMR\_Notes', 'BWH\_Arthritis\_C...', and 'Rheumatoid\_arth...'. The x-axis represents time, with a scale from 0 to 90. The y-axis lists the categories. Blue vertical bars indicate the occurrence of events over time.

On the right side of the interface is the 'Import Wizard' panel, which is currently in 'Step 1 - Choose Data'. It offers two options: 'Import' and 'Drag Patient'. Below this is 'Step 2 - Verify Data', which includes 'Patient Mapping' and 'Event Map' tabs. The 'Patient Mapping' tab shows a table with columns for 'Status', 'Patient', and 'Pat'. At the bottom of the wizard is 'Step 3 - Upload Data', which includes a 'Progress' indicator and an 'Upload' button.

Two large black arrows point from the text labels at the top and bottom of the image to the 'BWH Arthritis Center' and 'LMR Notes' categories in the query tool, and to the 'Timeline View' section, respectively.

Odds Ratio





Generate Summary Statistics | 
 Summary | 
 Clear | 
 Save

[Comparison](#)
[Advanced Workflow](#)
[Results/Analysis](#)
[Grid View](#)
[Data Export](#)
[Export Jobs](#)

Analysis ▾

## 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.

X

...\\Delai survie globale\

### 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 (ex. Age), then it should be "binned" using the option below. This variable is not required.

X

...\\M\  
...\\NM\

High Dimensional Data

[Search by Subject](#)
[Navigate Terms](#)
[Across Trials](#)

EGP0001 (173)

Biomarker Data (173)

Non Omics (173)

Immunological (172)

Mutation Detection (173)

BRAF Mutation (173)

abc M (5)

abc NA (2)

abc NM (166)

KRAS Mutation (173)

abc M (68)

abc NA (8)

abc NM (97)

NRAS Mutation (173)

Clinical Data (173)

Demographics (173)

SEX (173)

abc F (78)

abc M (95)

123 AGE (172)

Outcome (173)

Deces (173)

123 Delai survie globale (173)

123 Duree reponse (172)

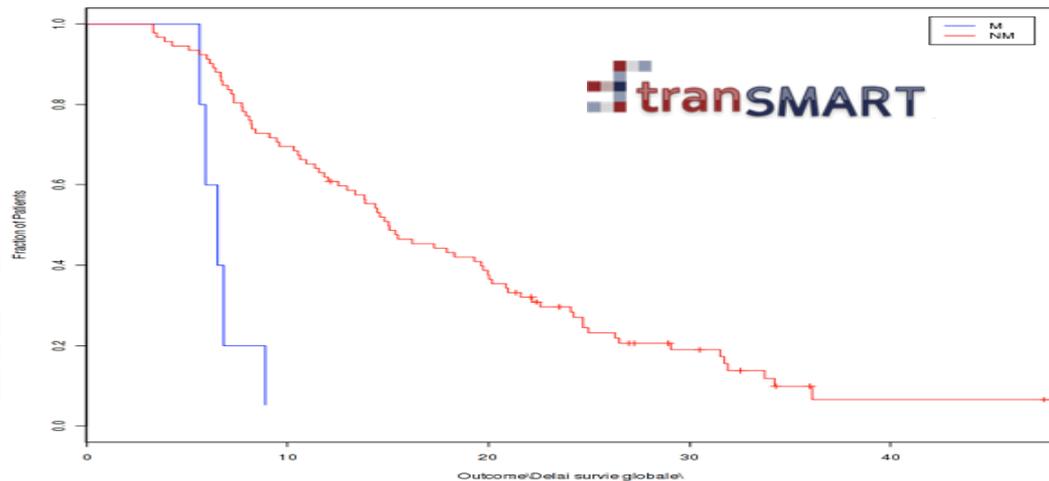
123 OMS Score (159)

123 Progression (172)

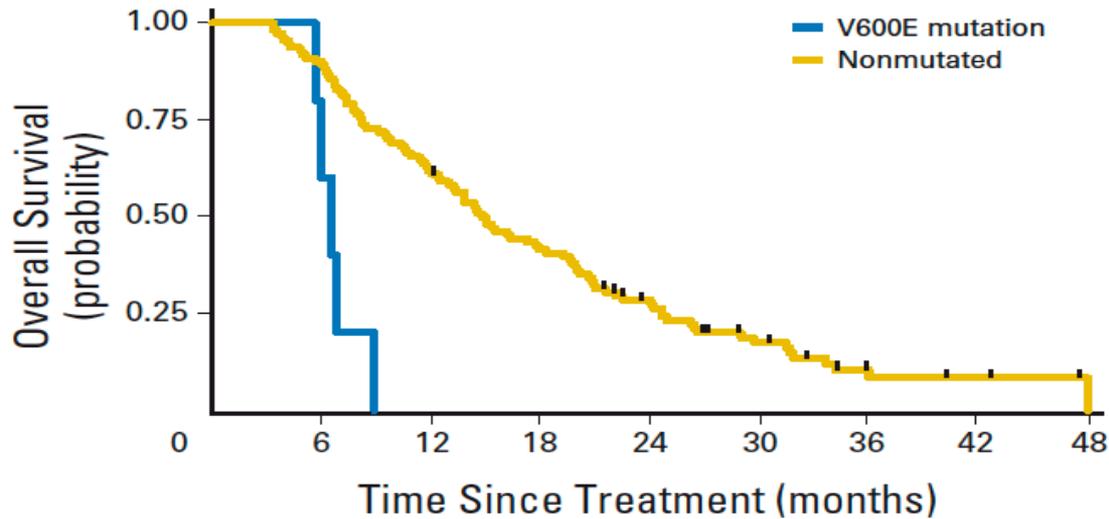
Treatment (173)

# HEGP tranSMART

- R module in tranSMART



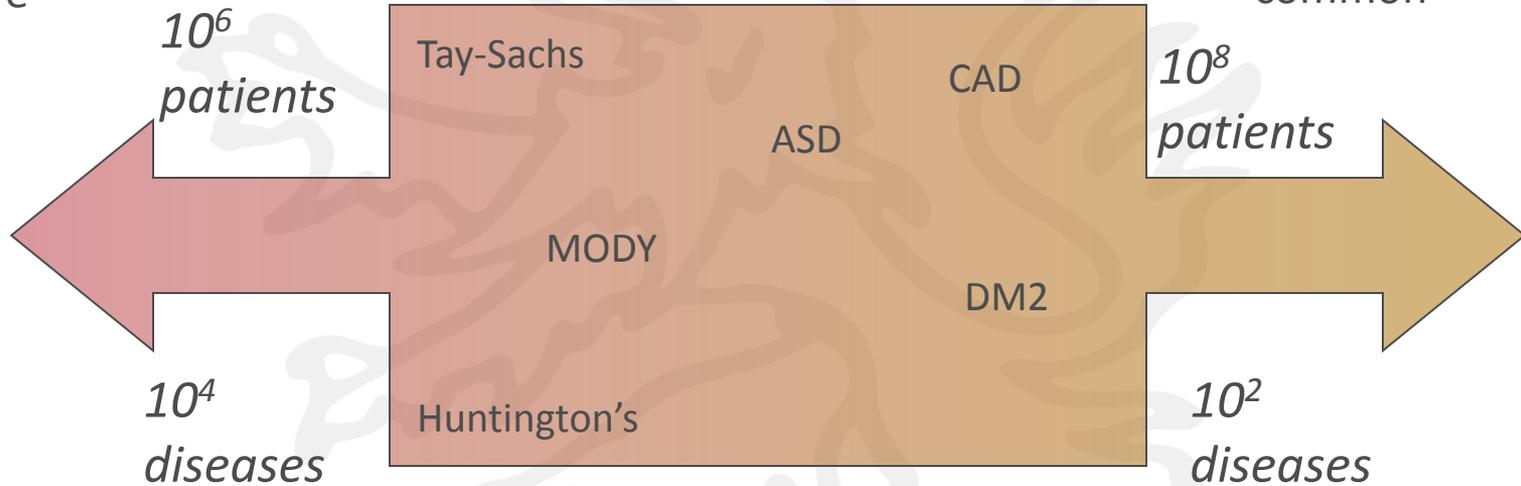
**B**



- Published figure in JCO

# Common-Rare: Weak-Strong Spectrum

rare



common

Deterministic  
“highly penetrant”

Weak effect, not predictive,  
dominated by environment

Rare &  
**COMMON**

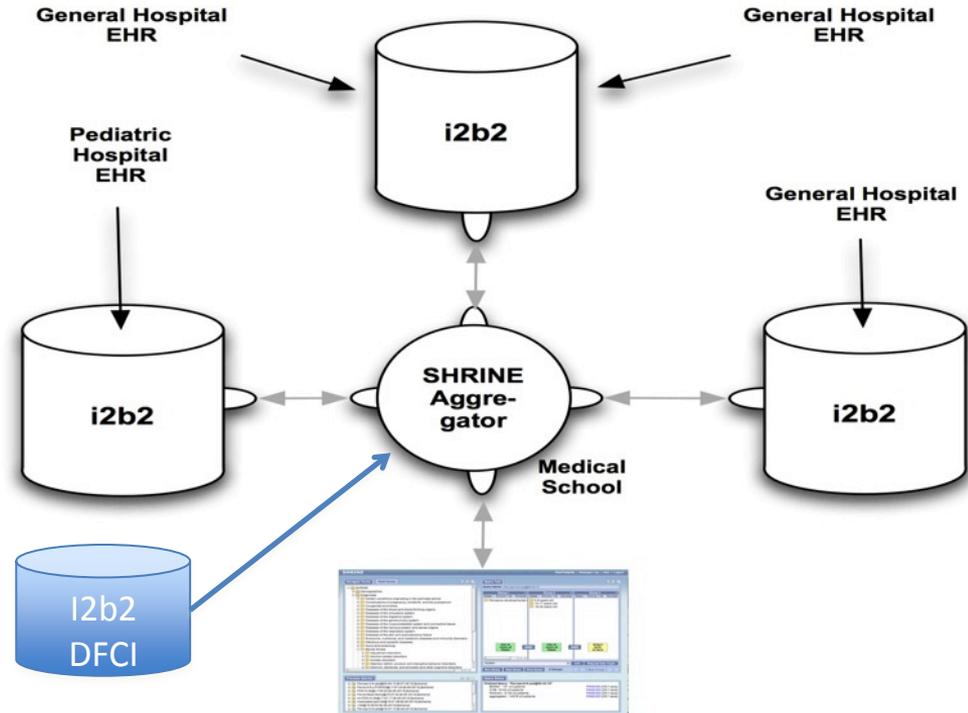
Needs

**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



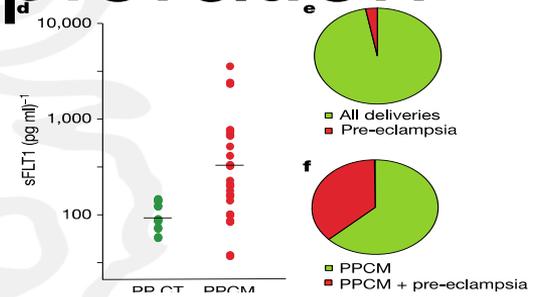
# Finding rare events of interest.

## Cardiac angiogenic imbalance leads to peripartum cardiomyopathy

Ian S. Patten<sup>1,2\*</sup>, Sarosh Rana<sup>3\*</sup>, Sajid Shahul<sup>4</sup>, Glenn C. Rowe<sup>1</sup>, Cholsoon Jang<sup>1</sup>, Laura Liu<sup>1</sup>, Michele R. Hacker<sup>3</sup>, Julie S. Rhee<sup>3</sup>, John Mitchell<sup>4</sup>, Feroze Mahmood<sup>4</sup>, Philip Hess<sup>4</sup>, Caitlin Farrell<sup>1</sup>, Nicole Koullis<sup>1</sup>, Elyahu V. Khankin<sup>5</sup>, Suzanne D. Burke<sup>5,8</sup>, Igor Tudorache<sup>6</sup>, Johann Bauersachs<sup>7</sup>, Federica del Monte<sup>1</sup>, Denise Hilfiker-Kleiner<sup>7</sup>, S. Ananth Karumanchi<sup>5,8</sup> & Zoltan Arany<sup>1</sup>  
17 MAY 2012 | VOL 485 | NATURE | 333

samples from subjects with PPCM have been previously described<sup>3</sup>. 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 \(SHRINE\)](#)<sup>5,3</sup>, a de-identified repository of aggregate patient information.

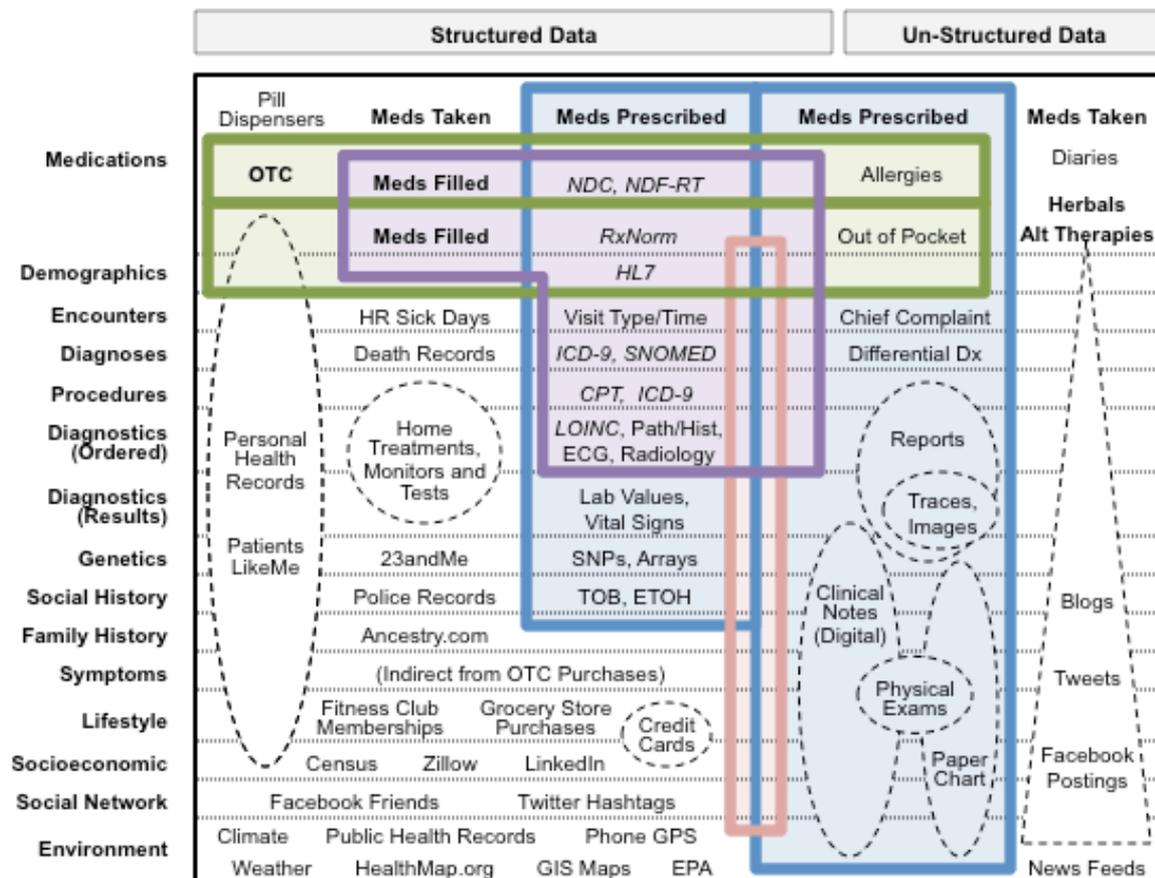
## Importance of real-time exploration



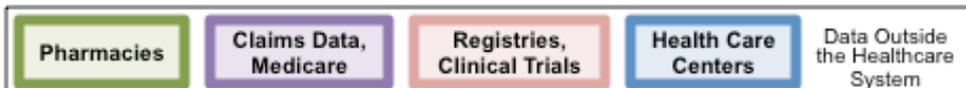
Where is the missing  
information?

$$V_P = V_G + V_E + V_{GE}$$

# "Big Data" Map of Patient Health



Courtesy  
Griffin Weber



Search   [Advanced Search](#) Display Options:



**Outbreak Missing?**  
Add it to the map

POWERED BY Google

**Quick Views** [\[manage\]](#)

- Vaccine preventable diseases
- User Submitted Reports
- H1N1 Swine Flu

**440 Alerts Now Showing**

Date: 24 Sep - 28 Sep | Diseases: [all](#) | Places: [all](#) | Sources: [all](#) | Categories: [all](#)

Source	Date	Summary	Disease	Location	Cases	Deaths	Significance	Stats
	27 Sep	<a href="#">Klimovich: "The investigation proved the intentional poisoning ..."</a>	Poisoning	<a href="#">Togo</a>			☆☆☆☆☆ 0 votes	
	27 Sep	<a href="#">Epidemic outbreak of Lassa fever in Sierra Leone: Two ... - ...</a>	Lassa Fever	<a href="#">Sierra Leone</a>		2	☆☆☆☆☆ 1 vote	
	27 Sep	<a href="#">Plague outbreak in China controlled - UPI.com</a>	Plague	<a href="#">Nyingchi, Tibet, China</a>			☆☆☆☆☆ 1 vote	



Influenza A (H1N1) Reports



**Source**

Informal Sources (Media)  
 Official Sources (ie: CDC, WHO)

[All HealthMap Sources >>](#)

---

**Category**

Ruled Out  
 Suspected Cases  
 Suspected Deaths  
 Confirmed Cases  
 Confirmed Deaths

---

**Zoom to country**

[Mexico](#)



Influenza A (H1N1) Virus, 2009 — Online Monitoring

John S. Brownstein, Ph.D., Clark C. Freifeld, B.S., and Lawrence C. Madoff, M.D.

- Portada
- Edición impresa
- Especiales
- Multimedia
- Servicios
- Escribanos

Usted está aquí: Portada → 2009 → 04 → 01 → Veracruz: reporta agente municipal extraño brote epidémico que ha cobrado dos vidas

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

Enviar esta página a alguien  
 Imprimir esta página

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



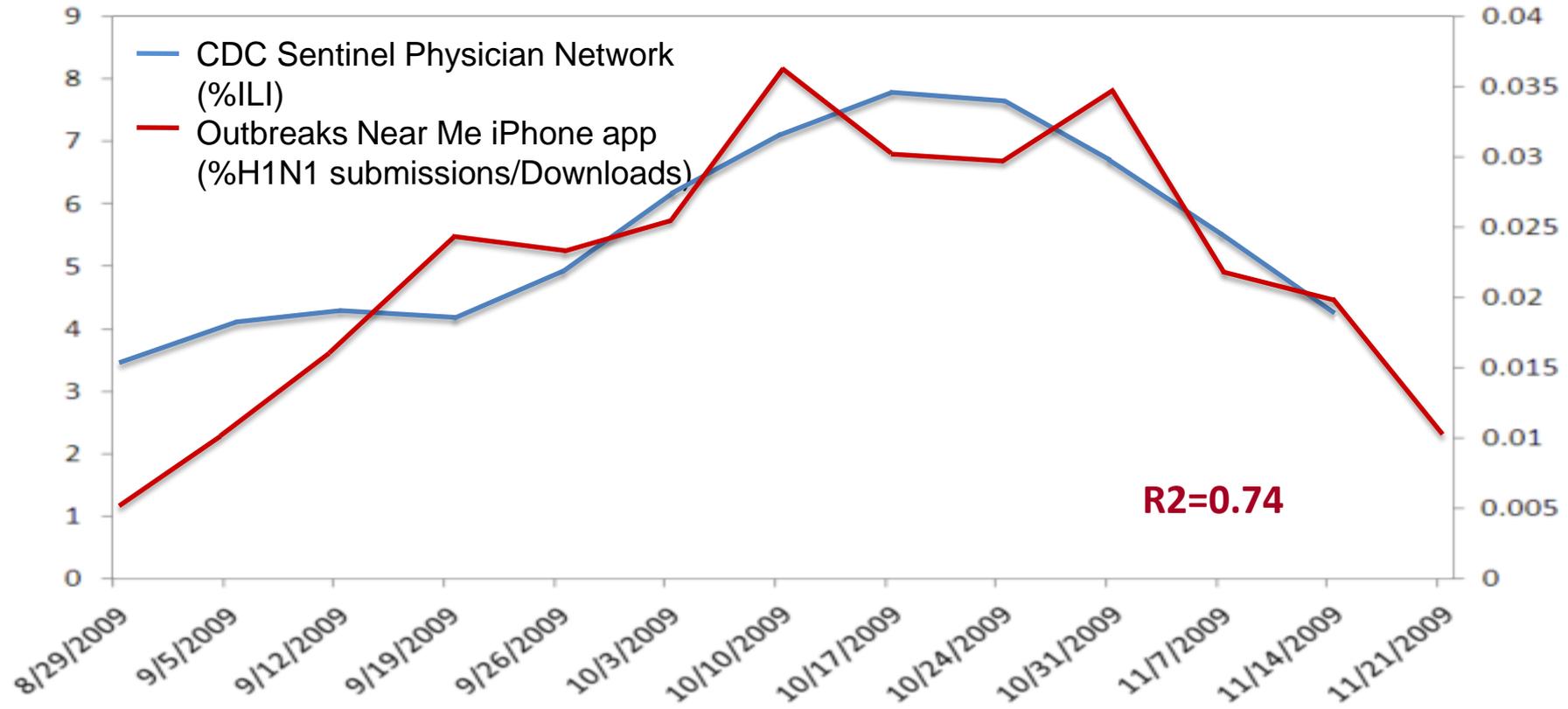
Wednesday, Sep. 09, 2009

## Is a Swine Flu Outbreak Coming? Ask Your iPhone

By Bryan Walsh

**TIME**  
IN PARTNERSHIP WITH **CNN**

# iPhone Submissions vs CDC sentinel surveillance



# 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)

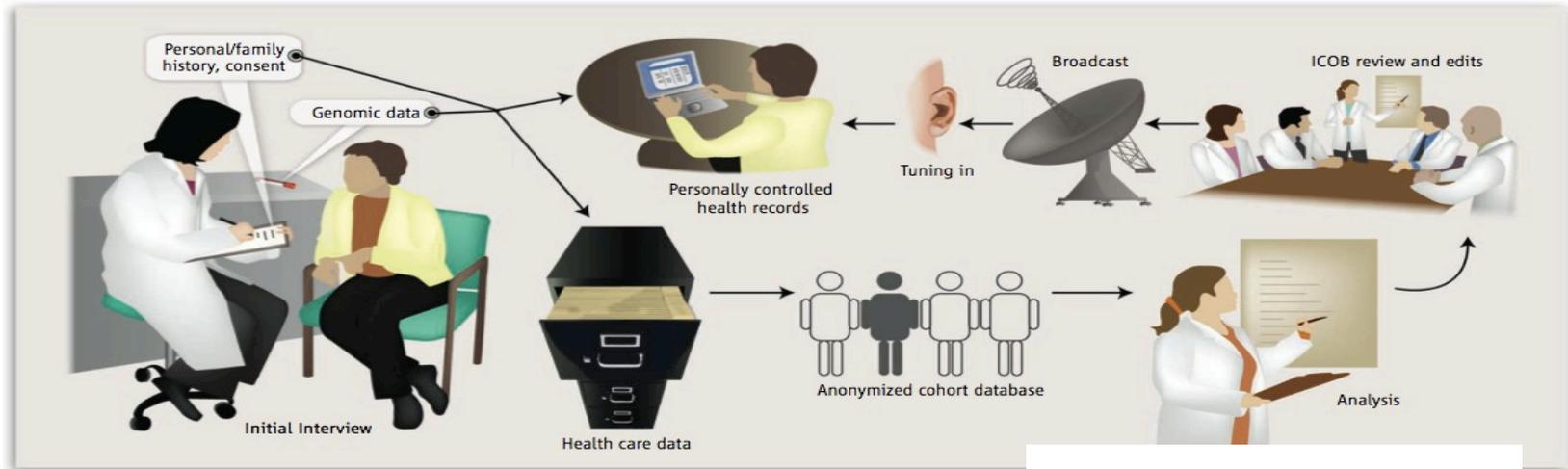
POLICYFORUM

MEDICINE

## Reestablishing the Researcher-Patient Compact

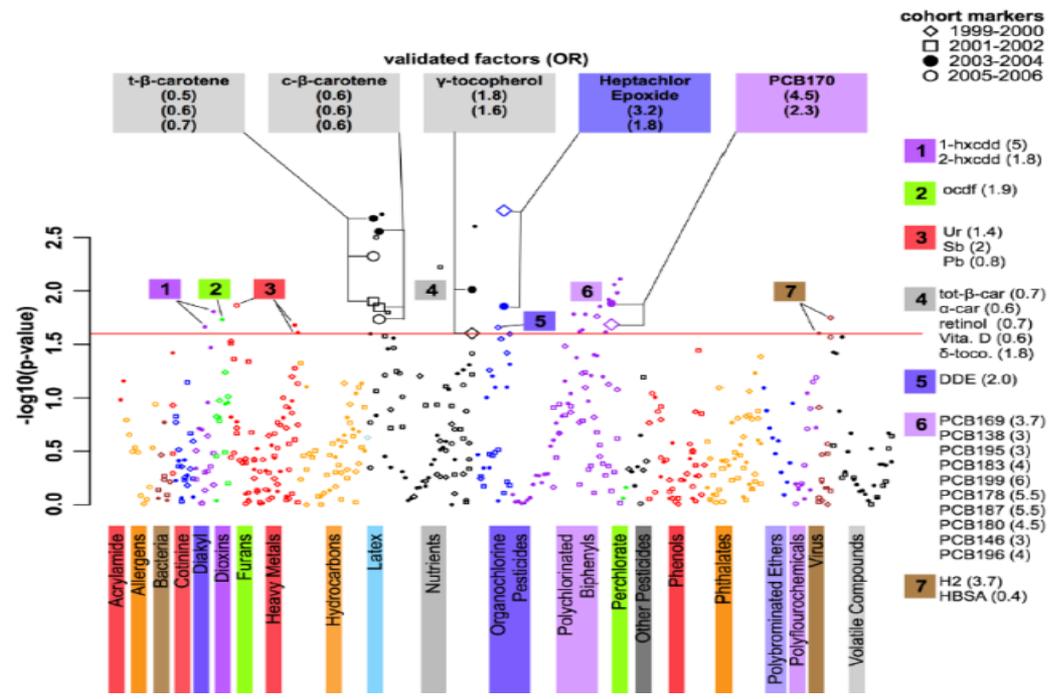
Isaac S. Kohane,<sup>1,2,3\*</sup> Kenneth D. Mandl,<sup>1,2,3</sup> Patrick L. Taylor,<sup>2,4</sup> Ingrid A. Holm,<sup>2,5</sup> Daniel J. Nigrin,<sup>1,2,3</sup> Louis M. Kunkel<sup>2,5,6</sup>

Well-intentioned regulations protecting privacy are denying important information to patient subjects. Advances in information technology mean that a better approach to clinical research is possible.



*Science 2007*

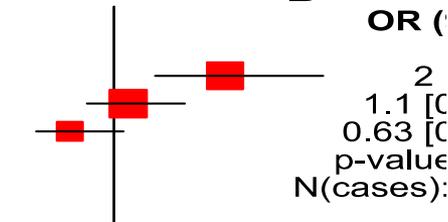
# 2023 Yearly check up includes exposure scan



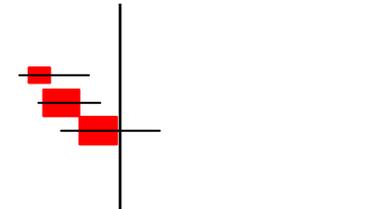
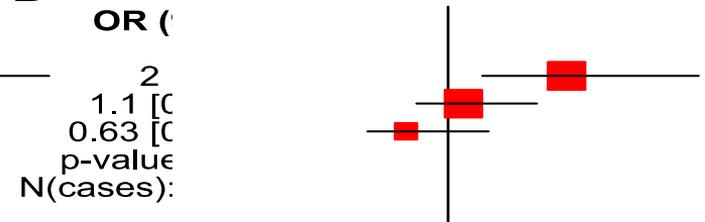
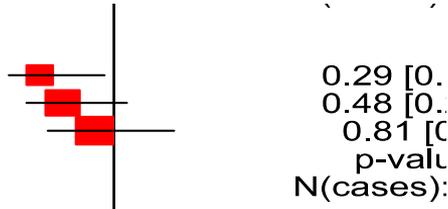
# WGS data interpreted in light of interact with environmental factors and current clinical status.

Chirag Patel

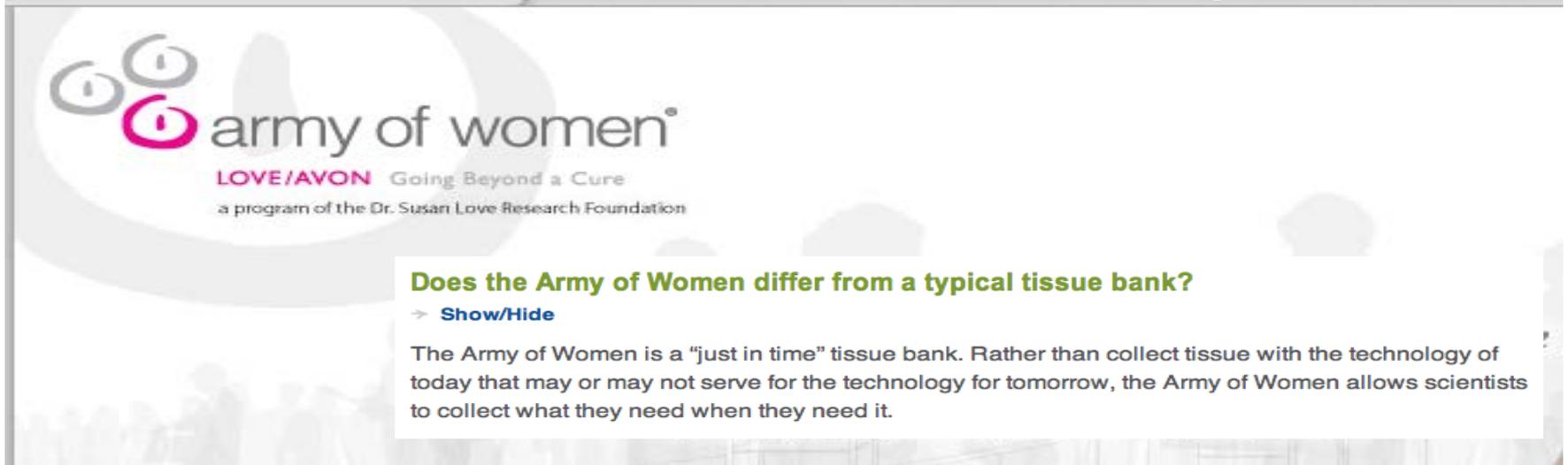
**rs13266634(SLC30A8)**  
 trans- $\beta$ -carotene (low(-1SD)) 1.8 [1.3,2.6]  
 trans- $\beta$ -carotene (mean) 1.1 [0.8,1.5]  
 trans- $\beta$ -carotene (high(+1SD)) 0.67 [0.41,1.1]  
 p-value (FDR):5e-05 (0.015)  
 N(cases):1699(167)



**rs2237895(KCNQ1)**  
 PCB170(low(-1SD)) 0.44 [0.21,0.93]  
 PCB170(mean) 0.61 [0.34,1.1]  
 PCB170(high(+1SD)) 0.85 [0.5,1.5]  
 p-value (FDR):0.023 (0.24)  
 N(cases):1009(88)



# Healthcare system no longer major source of trial patients.



The screenshot shows the Army of Women logo, which consists of three stylized female symbols (a circle with a vertical line) in grey and pink. To the right of the logo, the text reads "army of women" in a grey sans-serif font. Below this, "LOVE/AVON" is written in pink, followed by "Going Beyond a Cure" in grey. At the bottom, it says "a program of the Dr. Susan Love Research Foundation" in a smaller grey font. A white text box with a blue border is overlaid on the right side of the screenshot. It contains the question "Does the Army of Women differ from a typical tissue bank?" in green, followed by a blue link "> Show/Hide". Below the link, the text reads: "The Army of Women is a 'just in time' tissue bank. Rather than collect tissue with the technology of today that may or may not serve for the technology for tomorrow, the Army of Women allows scientists to collect what they need when they need it."

Rollover the areas above to learn more about each part of our model.

FlashTrials.gov/.com (cf. FlashMob) composed on the fly match volunteers to appropriate therapeutic trials. New study designs enable randomized partial trial participation.

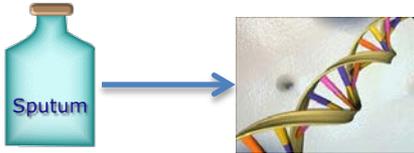
In era of big data, well-characterized patients are the only non-commodity.

- 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 healthdata generated in 1 home than an ICU bed in 2013



KINECT  
for XBOX360





## Pathophys. Precip

- 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.

i2b2.org

smartplatforms.org

**Developmental Medicine**

Lenny Rappaport  
Chuck Nelson  
Stephanie Brewster

**Biotechnology**

Daria Prilutsky  
David Margulies

**Genomics**

Lou Kunkel  
Alal Eran

**Bioinformatics**

Nathan Palmer  
Patrick Schmid  
Sek Won Kong  
Luke Hutchinson  
Bonnie Berger

**Population Science**

Ken Mandl  
Ben Reis  
John Brownstein

**Systems Informatics**

Susanne Churchill  
Rachel Ramoni  
Shawn Murphy  
Griffin Weber  
Doug MacFadden  
Bill Simons  
Mike Mendis  
Nich Wattanasin  
Stan Shaw  
Peter Szolovits  
Kat Liao  
Robert Plenge  
Tianxi Cai  
Finale Doshi  
Roy Perlis  
Jordan Smoller  
David Kreda  
Josh Mandel