Avoiding Fake Diseases and Biomarkup In Pediatrics

Isaac S. Kohane, MD, PhD
Google Maps: GIS layers
Organized by Geographical Positioning

Information Commons
Organized Around Individual Patients

Transportation
Land Use
Census Tracts
Structures
Postal Codes
Raster Imagery

Exposome
Signs and Symptoms
Genome
Epigenome
Microbiome
Other Types of Patient Data
Individual Patients

Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease
Report from National academy of science, USA, 2011
Some of the points I will make

• Most population genetics and clinical medicine addicted to categorical diagnosis
• Fake/Wrongly taxonomized diseases makes social & commercial construction of disease likely.
• Genomics-first/Genomics only unnecessarily limiting
• Clinical predictive accuracy does not imply shared physiology
• Disease overlaps demonstrates a fundamental problem
• Clinical data can be used to lesson the confusion.
• But multi-modal approaches are the most robust.
Blood Gene Expression Detection

Train with P1 and Predict P2

Train with P1 and Predict P2

Train with P2 and Predict P1

AUC 0.78

AUC 0.80

Male 0.73

Female 0.75
“I think when Ari [Ne’eman] talks about autism and I talk about autism, we’re talking about people with different clusters of autism. I know he doesn’t like the word ‘cure.’ If my daughter could function the way Ari could, I would consider her cured,” says Singer. “I have to believe my daughter doesn’t want to be spending time peeling skin off her arm.”
Extract and pool pituitary glands

CJD prions

Homogenize

Inject

30-year incubation

Aβ seeds

CJD and Aβ deposits

CJD and Aβ deposits

HARVARD MEDICAL SCHOOL

DEPARTMENT OF Biomedical Informatics
GIANT study

- 100’s of genes
- Not like diseases at tails
- So what do you call short stature and is that a diagnosis?
Criteria for Treatment

- “Growth hormone deficiency (GHD)”
- “Idiopathic short stature (ISS), defined by height standard deviation score ≤-2.25” associated with growth rates unlikely to result in normal adult height, in whom other causes of short stature have been excluded

and a little story from 25 years ago
But about once a month, an ambitious parent with cash to burn asks Desrosiers, a pediatric endocrinologist in Florida, if he would be willing to give growth hormones to a short but otherwise healthy child.

Desrosiers turns these patients away, but he says the requests still come.

Just last week, the father of a young baseball player -- a 14-year-old who was already 5 feet 6 inches tall -- expected Desrosiers to prescribe recombinant growth hormone (rGH) to add height to his budding athlete.

"He wanted to make his kid big, and he thinks he's going to walk out with the shots," said Desrosiers, director of pediatric endocrinology at Arnold Palmer Children's Hospital in Orlando. "He was willing to pay more than $45,000 a year, and didn't even bat an eyelash."

Desrosiers said he even gets requests for growth hormone from "Jolly Green Giant" families, where children are likely to be tall.

In 2003, the U.S. Food and Drug Administration approved the use of rGH for children with idiopathic -- or unexplained -- short stature, without a diagnosed metabolic hormone deficiency.
Pathological Interaction Between Clinical Annotation and Genetics.
Hypertrophic Cardiomyopathy (HCM)
- Heart failure
- Arrhythmias
- Obstructed blood flow
- Infective endocarditis
- Sudden cardiac death

Prevalence 1:500
Autosomal Dominant
HCM Prevalence = 1:500
HCM Inheritance = Autosomal Dominant
Genetics-induced Health Disparities

![Bar chart showing minor allele genotype prevalence for European Americans and African Americans for various genes.

- **European Americans**
  - TNNT2 (K247R): 27.14%
  - OBSCN: 15.27%
  - TNNI3 (P82S): 4.07%
  - MYBPC3 (G278E): 3.15%
  - JPH2 (G505S): 2.92%
  - Remaining mutations: 7.18%

- **African Americans**
  - TNNT2 (K247R): 2.88%
  - OBSCN: 0.33%
  - TNNI3 (P82S): 0.03%
  - MYBPC3 (G278E): 0.02%
  - JPH2 (G505S): 0.80%
  - Remaining mutations: 6.67%

The chart illustrates the potential impact of different genetic mutations on health disparities between European Americans and African Americans.]
<table>
<thead>
<tr>
<th>Age</th>
<th>Ethnicity</th>
<th>Report Year</th>
<th>Originally Reported Status</th>
<th>Current Status</th>
<th>Indication for Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>Unavailable</td>
<td>2005</td>
<td>P</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
<tr>
<td>75</td>
<td>Unavailable</td>
<td>2005</td>
<td>P</td>
<td>B</td>
<td>Family History and Clinical Symptoms of HCM</td>
</tr>
<tr>
<td>32</td>
<td>Black or African American</td>
<td>2005</td>
<td>P</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
<tr>
<td>34</td>
<td>Black or African American</td>
<td>2005</td>
<td>U</td>
<td>B</td>
<td>Clinical Diagnosis and Family History of HCM</td>
</tr>
<tr>
<td>12</td>
<td>Black or African American</td>
<td>2006</td>
<td>U</td>
<td>B</td>
<td>Family History of HCM</td>
</tr>
<tr>
<td>40</td>
<td>Black or African American</td>
<td>2007</td>
<td>U</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
<tr>
<td>45</td>
<td>Black or African American</td>
<td>2007</td>
<td>U</td>
<td>B</td>
<td>Clinical Features of HCM</td>
</tr>
<tr>
<td>16</td>
<td>Asian</td>
<td>2008</td>
<td>U</td>
<td>B</td>
<td>Clinical Diagnosis and Family History of HCM</td>
</tr>
<tr>
<td>59</td>
<td>Black or African American</td>
<td>2006</td>
<td>P</td>
<td>B</td>
<td>Clinical Features of HCM</td>
</tr>
<tr>
<td>15</td>
<td>Black or African American</td>
<td>2007</td>
<td>P</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
<tr>
<td>16</td>
<td>Black or African American</td>
<td>2007</td>
<td>P</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
<tr>
<td>22</td>
<td>Black or African American</td>
<td>2007</td>
<td>P</td>
<td>B</td>
<td>Clinical Diagnosis and Family History of HCM</td>
</tr>
<tr>
<td>48</td>
<td>Black or African American</td>
<td>2008</td>
<td>U</td>
<td>B</td>
<td>Clinical Diagnosis of HCM</td>
</tr>
</tbody>
</table>

P = Pathogenic and Presumed Pathogenic  
U = Pathogenicity Debated and Unknown Significance

P82Ser  
Gly278Glu

DEPARTMENT OF  
HARVARD  
Biomedical Informatics  
MEDICAL SCHOOL
Valsartan for Attenuating Disease Evolution In Early Sarcomeric HCM (VANISH)

This study is currently recruiting participants. (see Contacts and Locations)

Verified July 2013 by New England Research Institutes

Sponsor:
New England Research Institutes

Collaborator:
National Heart, Lung, and Blood Institute (NHLBI)

Information provided by (Responsible Party):
New England Research Institutes

ClinicalTrials.gov Identifier:
NCT01912534

First received: June 5, 2013
Last updated: October 5, 2015
Last verified: July 2013

History of Changes

Full Text View Tabular View No Study Results Posted
Group 1 (Overt HCM Cohort)
1. LV wall thickness ≥12 mm and ≤25 mm or z score ≥3 and ≤18 as determined by rapid assessment by the echocardiographic core laboratory
2. NYHA functional class I or II; no perceived or only slight limitations in physical activities
3. No resting or provokable LV obstruction (peak gradient ≤30 mmHg) on clinically-obtained Exercise Tolerance Test (ETT)-echo within the past 24 months or transthoracic echo with Valsalva maneuver within the past 12 months
4. Age 8-45 years
5. Able to attend follow-up appointments, complete all study assessments, and provide written informed consent

Group 2 (Preclinical HCM Cohort (G+/LVH-))
1. LV Wall Thickness <12 mm and z score <3, as determined by rapid assessment by the echocardiographic core laboratory
2. Age 10-25 years
3. E' z score ≤-1.5 OR ECG abnormalities other than NSSTW changes (Q waves, T wave inversion, repolarization changes) OR LV wall thickness z-score 1.5-2.9 combined with LV thickness to dimension ratio ≥0.19 (as determined by rapid assessment by the echocardiographic core laboratory)
4. Able to attend follow-up appointments, complete all study assessments, and provide written informed consent
Is Prediction the Acid Test of Diagnosis?
## Survival 3 Years After a WBC Test
(White, Male, 50-69 Years; Using Last WBC Between 7/28/05 and 7/27/06)

<table>
<thead>
<tr>
<th>Repeat Interval</th>
<th>Result Time</th>
<th>WBC Value</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>Normal</td>
</tr>
<tr>
<td>&lt; 1 Day</td>
<td>12a-8a</td>
<td>43.33%</td>
<td>84.68%</td>
</tr>
<tr>
<td></td>
<td>8a-4p</td>
<td>54.55%</td>
<td>86.61%</td>
</tr>
<tr>
<td></td>
<td>4p-12a</td>
<td>77.30%</td>
<td>77.30%</td>
</tr>
<tr>
<td>&lt; 1 Year</td>
<td>12a-8a</td>
<td>47.83%</td>
<td>79.58%</td>
</tr>
<tr>
<td></td>
<td>8a-4p</td>
<td>76.96%</td>
<td>90.73%</td>
</tr>
<tr>
<td></td>
<td>4p-12a</td>
<td>81.65%</td>
<td>92.99%</td>
</tr>
<tr>
<td>&gt; 1 Year</td>
<td>12a-8a</td>
<td>95.65%</td>
<td>95.65%</td>
</tr>
<tr>
<td></td>
<td>8a-4p</td>
<td>97.30%</td>
<td>98.13%</td>
</tr>
<tr>
<td></td>
<td>4p-12a</td>
<td>92.68%</td>
<td>97.35%</td>
</tr>
<tr>
<td>Any</td>
<td>73.17%</td>
<td>91.79%</td>
<td>78.11%</td>
</tr>
<tr>
<td>Patients</td>
<td>1286</td>
<td>18775</td>
<td>3052</td>
</tr>
</tbody>
</table>
Clinical data reflect both patients’ health AND their interactions with the healthcare system.

<table>
<thead>
<tr>
<th>Patient Pathophysiology</th>
<th>Healthcare System Dynamics</th>
<th>Data Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Demographics</td>
<td>Number of Observations</td>
<td>Data Entry Errors</td>
</tr>
<tr>
<td>Diagnoses</td>
<td>Time of Day of Observations</td>
<td>Dictation Mistakes</td>
</tr>
<tr>
<td>Laboratory Test Results</td>
<td>Time Between Observations</td>
<td>Data Compression Loss</td>
</tr>
<tr>
<td>Vital Signs</td>
<td>Cost of a Test or Treatment</td>
<td>Unstructured Data</td>
</tr>
<tr>
<td>Genetic Markers</td>
<td>Clinical Setting / Clinician Type</td>
<td>Missing Data</td>
</tr>
</tbody>
</table>

**Clinical Encounter**

- **Patient** → **Clinician**

**Electronic Health Record (EHR) Data**

**Patient Pathophysiology**

- **Normal**
  - Best Outcomes
  - Moderate Outcomes
- **Abnormal**
  - Moderate Outcomes
  - Worst Outcomes
Predicting Survival from **Ordering** a Lab Test

% Reduction in Unexplained Variance of Survival Compared to Age-Sex-Race-Adjusted Mortality Rate

Odds Ratio of Death
Predicting Survival Using Lab Value & HSD
Clinical Data To Clarify Diagnostic Boundaries?
Mothers told me about bowel problems but pediatricians told me...

- code counts
  - 0-6 months
  - 6-12 months
  - 12-18 months

patients

<table>
<thead>
<tr>
<th>code counts</th>
<th>code counts</th>
<th>code counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6 months</td>
<td>6-12 months</td>
<td>12-18 months</td>
</tr>
</tbody>
</table>

patient clustering
Autism or Autisms?
What about increasing overlaps across diagnoses?
What genes are shared across co-morbid conditions?

Classification accuracy reveals shared biology
What is the disease?

- Schizophrenia
- Bipolar Disorder
- Major Depressive Disorder

Shared Genes?
Shared Symptoms?
Autism(s)

• Implications for study and treatment
• Currently not obvious from a genetics-first/only approach
• Why clinicians might miss it
• Unsupervised vs self-referential & circular supervised
• Multi-modal-first exploration.
What about when there is NO diagnosis?
UDN Data Process Overview

Data Mgt. Step #1:
Patient Application & initial response

UDN Gateway

UDN Record Archive and Electronic Trial Mgt.

Data Mgt. Step #2:
referral and data transfer to a UDN clinical center

Data Mgt. Step #3:
UDN Data Warehouse

UDN Gateway

Metrics outcomes

Public summaries

UDN Data Warehouse

UDN Record Archive and Electronic Trial Mgt.

Data Mgt. Step #4:
post visit reporting and review

Digital files, workflow, admin reports, phenotype, genotype

Data Mgt. Step #3:
if accepted

If accepted

Data Mgt. Step #2:
referral and data transfer to a UDN clinical center
Seven clinical sites
1 Baylor College of Medicine and Texas Children's Hospital
2 Duke Medicine with Columbia University Medical Center
3 Harvard Teaching Hospitals (BCH, BWH, MGH)
4 National Institutes of Health
5 Stanford Medicine
6 UCLA School of Medicine
7 Vanderbilt University Medical Center

Six additional research sites
8 Central Biorepository
9 Coordinating Center
10 DNA Sequencing Core Facilities
11 Metabolomics Core Facility
12 Model Organisms Screening Center
13 University of Oregon

Applications Received: 822
Applications Under Review: 311
Participants Accepted: 298

Model for the hard cases?
When there is no diagnosis

• If there is a high probability causal variant:
  – Diagnostic label links clinical findings in that patient to that variant

• Do all individuals with that variant have disease?
  – How does genetic background/environment contribute in the general population
  – $p(D|V) \neq p(V|D)$ (cf. HFE & Hemochromatosis)
Why is medicine so dependent on categorical diagnoses?
Medicine’s Uncomfortable Relationship With Math: Calculating Positive Predictive Value

Survey Responses

\( n = 61 \)

Most common answer – 95%

\( n = 27 \)

Correct answer – 2%

\( n = 14 \)
Summary

• Addiction to categorical diagnoses is cognitively useful
  – But our patients have beaten us to the Google-reflex
• Categorical diagnosis can result in spurious biological & clinical inference
  – Often manipulated for $$$, 2ndry agenda
  – Single measurement modality (incl. genomics) easier to manipulate
• Prediction of diagnostic class not necessarily evidence of biological etiology
• Diagnoses are more robust and useful when
  – Data-driven
  – Formally model health systems dynamics
  – Statistically-informed
  – Multi-modal
  – Unsupervised or lightly supervised.
Acknowledgments

Nathan Palmer
Susanne Churchill
Arjun “Raj” Manrai
Andrew Beam
Denis Agniel
Matt Might
Griffin Weber
Rachel Ramoni
Sumaiya Nazeen
Bonnie Berger
Alexa McCray