Getting Personal with the Healthcare Enterprise for Discovery and Better Care

Harvard Medical School
Center for Biomedical Informatics
i2b2 National Center for Biomedical Computing
Isaac S. Kohane, MD, PhD

Example: PPARγ Pro12Ala and diabetes

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Sample selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oh et al.</td>
<td>Deeb et al.</td>
</tr>
<tr>
<td>Mancini et al.</td>
<td>Clement et al.</td>
</tr>
<tr>
<td>Hegele et al.</td>
<td>Hasstedt et al.</td>
</tr>
<tr>
<td>Lei et al.</td>
<td>Ringel et al.</td>
</tr>
<tr>
<td>Hara et al.</td>
<td>Meirhaeghe et al.</td>
</tr>
<tr>
<td>Douglas et al.</td>
<td>Altschuler et al.</td>
</tr>
<tr>
<td>Mori et al.</td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td></td>
</tr>
</tbody>
</table>

Overall \( P \) value = 2 \times 10^{-7}

Odds ratio = 0.79 (0.72-0.86)

Ala is protective

Courtesy J. Hirschhorn

And here comes commercialization (MD’s not required)

Knome has launched the first commercial whole-genome sequencing and analysis service for individuals for $350,000 per genome. The sequence data will undergo comprehensive analysis from a team of...
Evidence of Disintermediation of Healthcare System

Use of cancer susceptibility testing among primary care physicians

Physician Use of Genetic Testing for Cancer Susceptibility: Results of a National Survey

US Physicians' Attitudes Toward Genetic Testing for Cancer Susceptibility

Clinical Informatics Ambitions c. 1970

Abstract Rapid advances in the information sciences, coupled with the political commitment to broad extensions of health care, promise to bring about basic changes in the structure of medical practice. Computing science will probably exert its major influence by augmenting and, in some cases, largely replacing the intellectual functions of the physician. As the "intelligent" use of the computer influences the fundamental tasks of both physician manpower and quality of medical care, it will also inevitably exact important social costs — psychologic, organizational, legal, economic and technical. Only through consideration of such potential costs will it be possible to introduce the new technology in an effective and acceptable manner.

- William Schwartz, 1970
Problems

<table>
<thead>
<tr>
<th>LEADING CAUSES OF DEATH¹</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diseases of the Heart</td>
<td>726,974</td>
</tr>
<tr>
<td>Cancer (malignant neoplasms)</td>
<td>539,577</td>
</tr>
<tr>
<td>Cerebrovascular Disease</td>
<td>159,791</td>
</tr>
<tr>
<td>Chronic Obstructive Pulmonary Disease</td>
<td>109,029</td>
</tr>
<tr>
<td>Medical Errors²</td>
<td>44,000–98,000</td>
</tr>
<tr>
<td>Accidents and Adverse Effects</td>
<td>95,644</td>
</tr>
<tr>
<td>(motor vehicle accidents = 43,458; all others = 52,186)</td>
<td></td>
</tr>
<tr>
<td>Pneumonia and Influenza</td>
<td>86,449</td>
</tr>
<tr>
<td>Diabetes</td>
<td>62,636</td>
</tr>
<tr>
<td>Suicide</td>
<td>30,535</td>
</tr>
<tr>
<td>Kidney Disease</td>
<td>25,331</td>
</tr>
<tr>
<td>Liver Disease</td>
<td>25,175</td>
</tr>
</tbody>
</table>


A Healthcare System Not Meeting the Challenge

- A lot of promise (genomics)
- Unresponsive (expensive) healthcare systems
- Willingness of the public and the private sector to consider alternatives.

Three Prongs to Efficiently Reach Large $N$

- High throughput genotyping
- High throughput phenotyping
- High throughput sample acquisition

DHHS Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS) argues for the health value of a 500,000 to 1M subject study. Estimated cost: $3,000,000,000

Cost of the pediatric 100,000 study recently launched $>>$ 1B + decades.
NLP (and comedy) is not pretty

But it works

- 96,000 asthma patients identified out of 2.5M PHS patients
  - Stratified by severity, pharmaco-responsiveness and exposures
  - Now with cases and controls (from extrema) reconcensed and biomaterials obtained for genome-wide scans ++
  - 3 methods of tissue acquisition
The three prongs of High Throughput Instrumentation

- $250-$500 for 500,000 SNP’s
- $50-100K for good quality phenotyping of 100K++ individuals
- What about the samples (consented)
  - $650/patient
    - Dozens a week
  - Wait in clinic: $450+/patient
- Crimson
  - Lynn Bry, MD

Crimson: Core Functions

- Mined Phenotypes
- Matched Anonymous ID
- Clinical discard
- Richly annotated biospecimens

Meeting Expectations
i2b2 Hive: A Translational Toolkit

- Adoption 18+AHC’s
- Commercial and academic development efforts
- Free and open source

https://www.i2b2.org/software/
Free-will, autonomy, volunteerism, fairness

**The New England Journal of Medicine**

**Sounding Board**

**Health-Information Altruists — A Potentially Critical Resource**
Isaac S. Kohane, M.D., Ph.D., and Russ B. Altman, M.D., Ph.D.

One of the key ideas behind sequencing the human genome was the promise of "personalized medicine." The idea was that genetic information could be used to make health care more precise, efficacious, and safe. The Human Genome Project showed us that among humans, DNA sequences are 99.9 percent similar, but the remaining 0.1 percent, in the context of environmental and epigenetic factors, produces the entirety of genetic variability the National Human Genome Research Institute, has called for large cohorts (at least 200,000 subjects) to be assembled simply to achieve the necessary sample sizes to overcome the problems of cross-sectional studies.²

**Cause for Concern: No Perfect Anonymity**

One point on the spectrum of volunteerism