Slightly Controlled Information
Exhibitionism in the Genomic Era

A Dangerous (but good?) Idea
Isaac S. Kohane, MD, PhD
One Motivation

- Key promise of the genome sequencing projects was an era of “personalized medicine”
- 99.9% of all human genomes are identical, but the remaining 0.1% produces the entire variability within the human population
- Research projects are collecting phenotypes about patients and correlating them with the precise variations present in the patients’ individual genomes
Points to be Made

- Conventional methodologies to obtain large phenotypically annotated populations are not adequately productive.

- Concerns about privacy and disclosure of genotype and phenotype data are real but addressable sociologically, technologically and legislatively.

- Patient/consumer populations are ahead of the healthcare and research establishment in thinking about these challenges.

- Methods for giving patient control of health data disclosure provide a mechanism for gathering research data, and

- Harnessing the volunteerism of our populations is an important and promising solution today.
Clinical genomic research depends on large, well-phenotyped populations but…

- There are several challenges
  - In size of populations required
  - In obtaining the phenotypes
Example: PPAR\textsubscript{γ} Pro12Ala and diabetes

Estimated risk (Ala allele)  
0.1 0.3 0.5 0.7 0.9 1.1 1.3 1.5 1.7 1.9  
0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2.0  

Ala is protective
Series of screenshots removed due to copyright reasons. Please see: http://spin.nci.nih.gov/
Series of screenshots removed due to copyright reasons. Please see: http://spin.nci.nih.gov/
Series of screenshots removed due to copyright reasons. Please see: http://spin.nci.nih.gov/
Risks

- Real Risks
- Cultural-specific perceptions
Risks: There is no perfect anonymity

- Sweeney
- Altman

Figure removed due to copyright reasons.
Please see:

● The decision to share genetic data affects other people. ‘
   DNA is shared with immediate family members, extended family,
   and ethnic groups
● Carelessness.

Figure removed due to copyright reasons.
The cat is already out of the bag

- Most people are not fully aware of the degree to which their blood samples can be and are used by the pharmaceutical industry
  - When they learn, some wonder if they should participate in the profits that result from their samples.
- Insurance companies, hospitals routinely share data for reimbursement and research.
Perceptions of Risk Vary

- Culturally and personally determined.
- Not surprisingly, the level of concern for health privacy shows a distribution (just as many other phenotypes!) in the population
Figure removed due to copyright reasons.
Implanted RFID

- ISO standard representation of data
- FDA approved
- 16 bytes of healthcare identifiers
- Implanted in the right triceps
- MRI safe
- Airport security safe
- John’s RFID identifier is linked at CareGroup to his medical history
Figure removed due to copyright reasons.
Please see:

Figure removed due to copyright reasons.
Please see:

Annas, G. J. "Rules for research on human genetic variation--lessons from Iceland."
An Alpha-1 patient has one of the most obvious genetic discrimination cases in the country ... Patient, who lost a 30-year-old brother to Alpha-1, was identified as having the disorder ... As a preventative treatment her doctor put her on a plasma augmentation therapy used to slow the progression of lung damage. Shortly after receiving the bill for her first treatment, her employer, who self-insured, abruptly fired her, in spite of the fact that she continued to perform her job, in a manner, which her employer had always referred to, as exemplary.
• The problem:
  • Highly mobile patients
  • No formalized data exchange
  • Mobile patients
  • Disparate providers
  • Incompatible medical data storage
  • Lack of unified view of patient record
Solving this in a centralized fashion is another major risk.
Solution: PING

- Personal Internetworked Notary and Guardian
- Personally controlled records
- Ubiquitous
- Open standards
- Open source
- Web interface
- Public key infrastructure
- XML-based data
Genomic Data for a PHR
Example SNP-Test

Sax U, Cassa C, Kohane IS.
Representation of Genomic Data in
Personal Health Records.
(in preparation) I2B2
Overall Proposed solution

- Seek out volunteers who exist on the end of the spectrum near the exhibitionists.
- Potential damages can be outlined, but the subjects will elect to take the risk in the hopes of helping to save lives.
- Volunteers would donate DNA sequence data as well as molecular and cellular phenotypic data.
Modest (?) Necessary Desiderata

- **Actions by policy makers and legislators** could clear the field for these studies.
  - **Illegal** to act upon the health information contained in research databases would send a simple message to those considering mining these data:
    - you can do it, but you are not allowed
  - Researchers who curate genetic databases should have some protection of activities,
    - as long as they follow an **agreed set of operating guidelines**.
    - Guidelines **would not guarantee privacy** (a standard that is too high),
    - guarantee reasonable physical security and data quality.
- **Patients** should be granted explicit control over the disclosure process.
Real test

- Create pilot studies to test the feasibility of asking patients to accept lower levels of privacy guarantees.
- Require the development of new consents.
- Studies would also limit the disclosed data to include only those subsets of information that the patient is willing to share.
- At the same time, policy makers would provide some protections for the patients and the researchers.
In Summary

- Why should you fear this idea?
- Why should you rejoice that it's beginning to be implemented?
- What should you tell your mom about it?
- What's your most interesting discovery?
  - Oncogenesis recapitulates ontogenesis
- What's your most recent discovery?
  - It’s all over after 40.