EHIP 2009

Electronic Health Information & Privacy Conference

November 19, 2009 - Ottawa, Canada

Gold Sponsors:

Bell

PRIVACY ANALYTICS

GenomeCanada

CHEO Research Institute Institut de recherche

Silver Sponsor:

Microsoft®

www.ehip.ca

PROGRAM

REGISTRATION & WELCOME (8:00 – 8:30)

OPENING REMARKS & PLENARY (8:30 – 9:45)

Richelieu/Frontenac Room

"Reconsidering Privacy in the Genomic Era"

Mark A. Rothstein, Herbert F. Boehl Chair of Law and Medicine & Director of the Institute for Bioethics, Health Policy and Law, University of Louisville School of Medicine

BREAK (9:45 – 10:15)

TRACK 1

Richelieu/Frontenac Room

Panel 1A (10:15 – 12:00)

Public Release of Genomic Data

Session Chair: Patricia Kosseim, Genome Canada

Panelists:

Yann Joly, McGill University

Laura Rodriguez, NIH/NHGRI

Aled Edwards, Structural Genomics Consortium, University of Toronto

TRACK 2

Joliet Room

Session 2A (10:15 – 12:00)

Data Linkage and Privacy

Session Chair: Liam Peyton, University of Ottawa

Speakers:

Andrew Borthwick, Intelius, Inc.

Frederick Bieber, Harvard Medical School

Stanley Trepetin,New York City Department of Health and Mental Hygiene

LUNCH (12:00 – 13:00)

Richelieu/Frontenac Room

Session 1B (13:00 – 14:45)

De-identification of Genomic Data

Session Chair: Bradley Malin, Vanderbilt University

Speakers:

Murat Kantarcioglu, University of Texas at Dallas

Chris Cassa, Harvard Medical School and MIT

Bradley Malin, Vanderbilt University

Session 2B (13:00 – 14:45)

Privacy Considerations in Disease Surveillance

Session Chair: Philip AbdelMalik, PHAC

Speakers:

Anita Fineberg, Anita Fineberg & Associates

Jay Mercer, Canadian Medical Association/Practice Solutions

Khaled El Emam, CHEO Research Institute & University of Ottawa

BREAK (14:45 – 15:15)

Session 1C (15:15 – 17:00) Genomics On-line	**Session 2C** (15:15 – 17:00) Health Privacy in Practice
Session Chair: Patricia Kosseim, Genome Canada	**Session Chair:** Michael Power, Privacy Lawyer
Speakers: Mike Spear, Genome Alberta Rose Geransar, University of Calgary & Farah Mohamed, University of Alberta Brenda Wilson, University of Ottawa	**Speakers:** Peter McLaughlin, Foley & Lardner Mike Gurski (Bell Canada)

Table of Contents

Introduction

2009 Electronic Health and Information Privacy Conference

More and more health information is being collected about us – and much of that data is collected, transmitted and stored electronically. This not only includes clinical information, but increasingly life style and genetic information as well.

There is growing demand to use this personal health information for research, administrative, and policy making purposes. At the same time, there have recently been at least 143 data breaches in Canada and US from medical establishments affecting more than 6.3 million records (see http://www.ehealthinformation.ca/dataloss). This has multiple negative consequences: from reducing the trust of patients in the public and private organizations that manage their personal information, to patients adopting privacy protective behaviors that may be detrimental to their well being. This trust, once lost, is difficult to regain.

The theme for the 2009 conference is the collection and use/disclosure of genetic information. We will address issues concerning the consent and security mechanisms around the construction of biobanks, including linking to other data sources. The conference will also cover a number of very relevant contemporary privacy issues: privacy considerations in the context of syndromic surveillance (for example, when trying to detect influenza like illnesses from various hospital and practice sources), and the expected significant changes to the US Health Insurance Portability and Accountability Act (HIPAA) and HIPAA enforcement. The focus will be on policy as well as technical issues and solutions.

Organizing Committee:
Khaled El Emam, CHEO RI & University of Ottawa
Patricia Kosseim, Genome Canada
Brad Malin, Vanderbilt University

Keynote: Reconsidering Privacy in the Genomic Era

Mark A. Rothstein, University of Louisville

Abstract:

Privacy is a popular concept in the abstract, but one that eludes a consensus definition and quickly becomes contentious in its numerous applications. After attempting to simplify and demystify the concept of privacy, this talk will focus on the challenges to privacy raised by new genomic technologies. The talk will address whether it would be better to address genetic and genomic privacy by enacting special legislation or by having more general protections for informational health privacy. It also will discuss some of the specific privacy issues raised by genomics in research, electronic health records, and other areas.

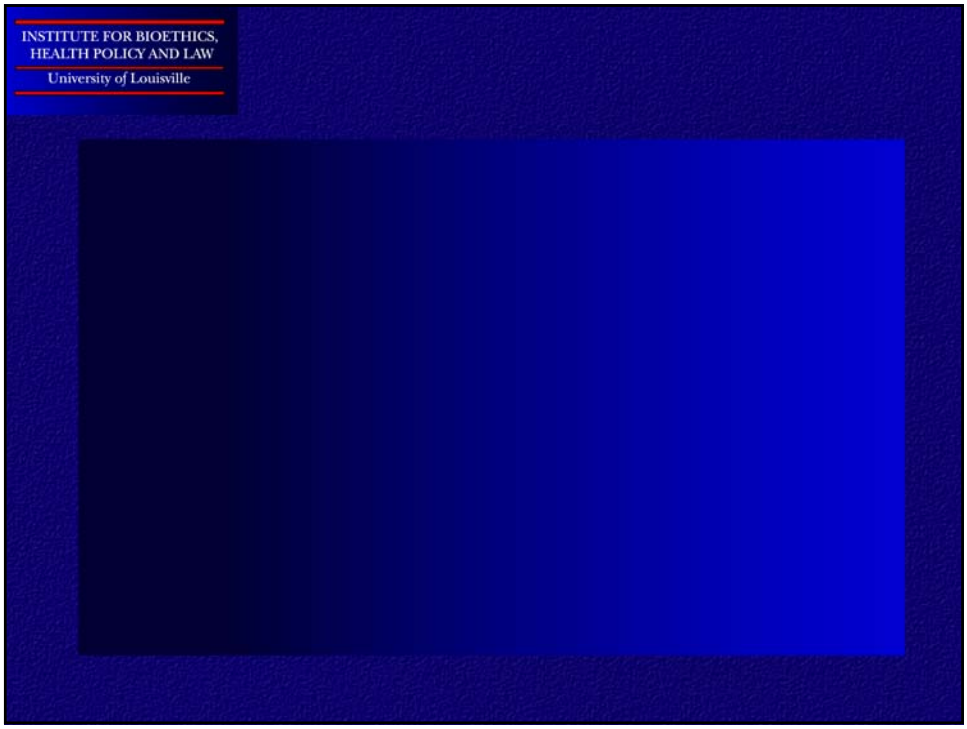
Bio:

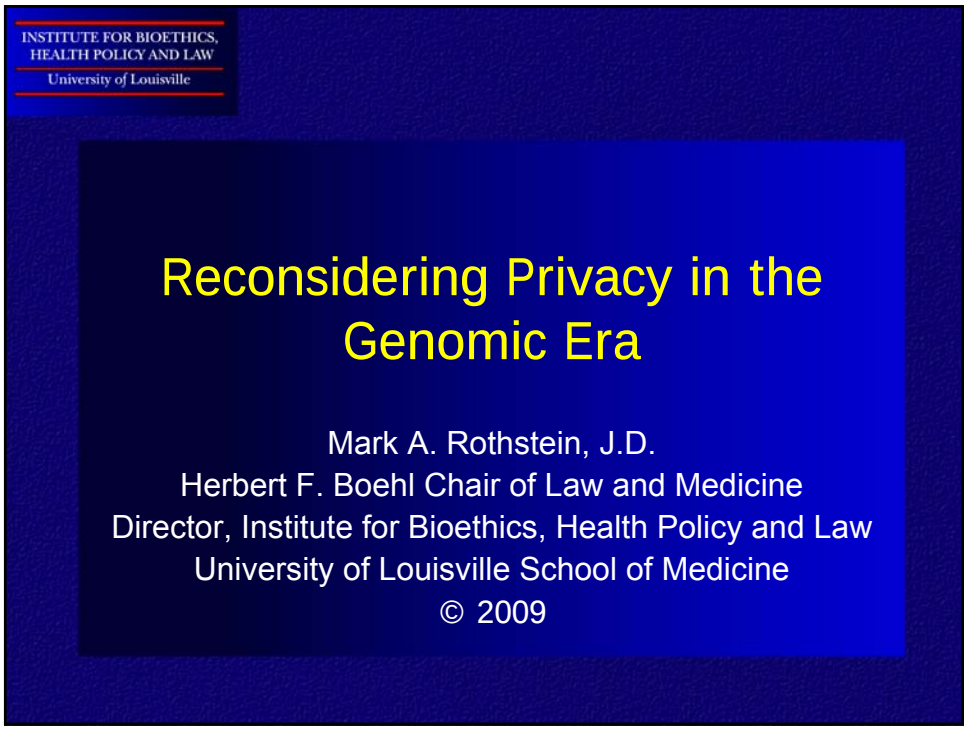
Mark A. Rothstein holds the Herbert F. Boehl Chair of Law and Medicine and is the Founding Director of the Institute for Bioethics, Health Policy and Law at the University of Louisville School of Medicine. He received his B.A. from the University of Pittsburgh and his J.D. from Georgetown University.

Professor Rothstein has concentrated his research on bioethics, genetics, health privacy, public health law, and employment law. From 1999-2008, he served as Chair of the Subcommittee on Privacy and Confidentiality of the National Committee on Vital and Health Statistics, the statutory advisory committee to the Secretary of Health and Human Services on health information policy. He is past president of the American Society of Law, Medicine and Ethics.

He is the author or editor of 19 books and nearly 200 book chapters and articles in leading journals of bioethics, law, medicine, and public health.

Link to video of this presentation.


INSTITUTE FOR BIOETHICS,
HEALTH POLICY AND LAW
University of Louisville


INSTITUTE FOR BIOETHICS,
HEALTH POLICY AND LAW
University of Louisville
Reconsidering Privacy in the Genomic Era
Mark A. Rothstein, J.D.
Herbert F. Boehl Chair of Law and Medicine
Director, Institute for Bioethics, Health Policy and Law
University of Louisville School of Medicine
© 2009

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Focus of this talk

1. Conceptualizing genetic privacy
2. Assessing genetic exceptionalism
3. Exploring the effects on privacy of specific technologies

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

1. Conceptualizing genetic privacy.

Privacy is a befuddling concept.

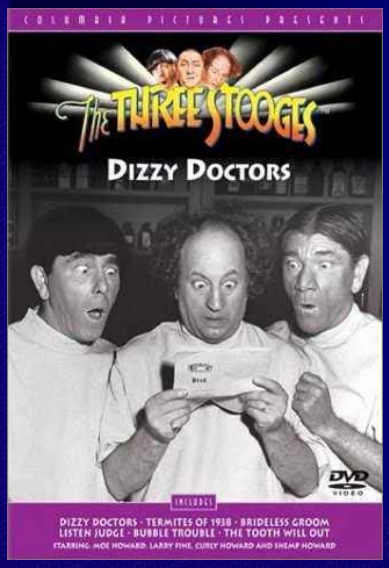
INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy is universally praised . . . in the abstract.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

There is widespread disagreement with regard to specifics, such as the definition of privacy, the relative importance of privacy compared with other values, and the best way to protect privacy.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville



Health privacy is especially befuddling, including for those who work in health care.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Definition:

"Privacy is a condition of limited access to the person or personal information."

In its essence, privacy has the following two aspects.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

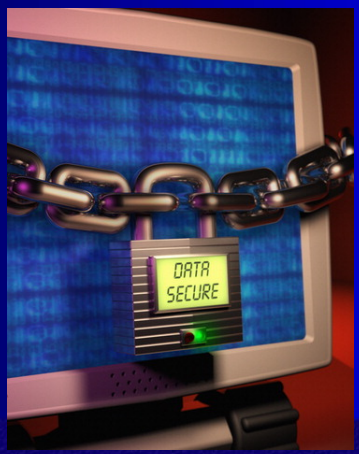
1. Intrusional privacy

* physical presence
* visual or aural intrusion
* electronic surveillance

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

2. Informational privacy

* educational
* financial
* health



INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Informational **HEALTH** privacy includes behavioral, reproductive, and genetic privacy

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy

Informational Intrusional

Educ. Financial **Health**

Behavioral **Genetic** Repro.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

OTHER COMMONLY ASSERTED ASPECTS OF PRIVACY

- Decisional privacy (autonomy)
- Proprietary privacy (appropriation)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

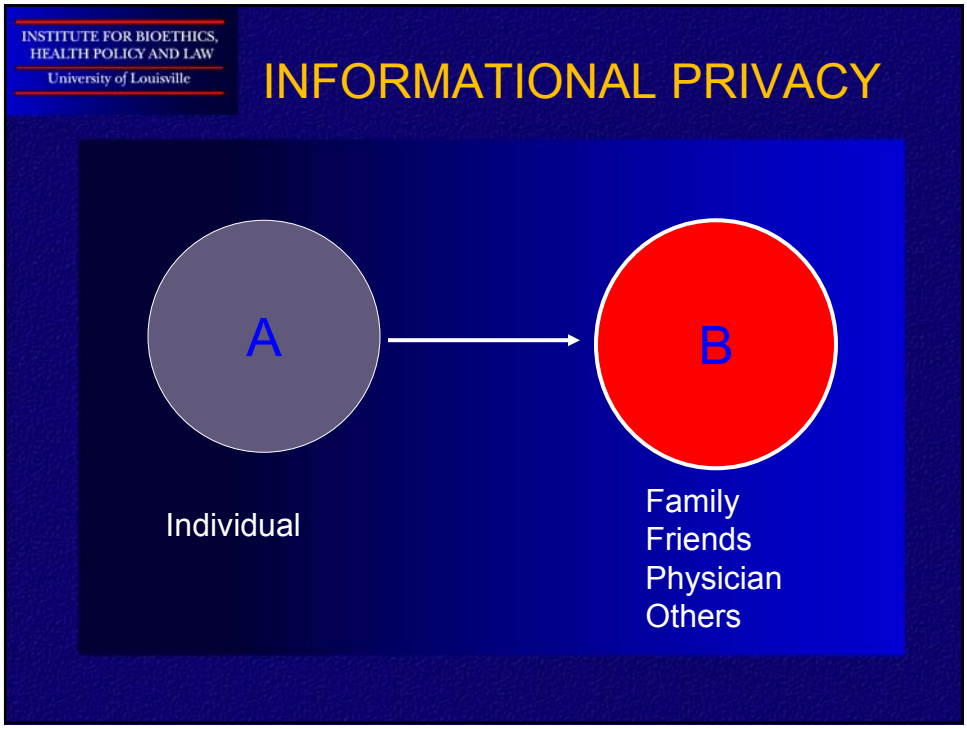
Informational Privacy

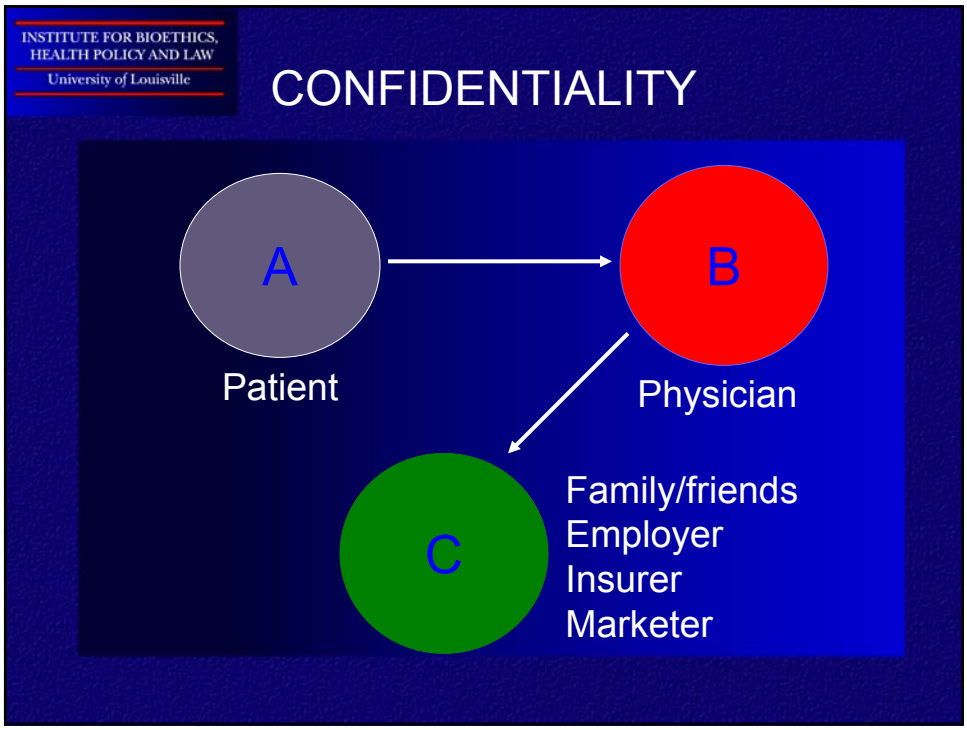
vs.

Confidentiality

vs.

Security


INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville
INFORMATIONAL PRIVACY
A
B
Individual
Family
Friends
Physician
Others


INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville
CONFIDENTIALITY
A
B
Patient
Physician
C
Family/friends
Employer
Insurer
Marketer

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

SECURITY

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy
vs.
Right to Privacy

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

MORE PRECISE NOTIONS OF PRIVACY

1) Privacy simpliciter	Objective sense
2) Right to privacy	Normative sense
3) Right to privacy	Legal sense

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Right to Privacy – normative sense
ethics / morality / philosophy / policy

- Privacy promotes human dignity.
- Privacy permits the development of intimate relationships.
- Privacy enables medical care.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Conclusion :

Privacy ought to be protected to advance important individual and societal interests.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Right to Privacy – Legal sense

- Constitutional law (e.g., unreasonable search and seizure prohibited by Canadian Charter of Rights and Freedoms; U.S. 4th Amendment)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Right to Privacy – Legal sense

- Statutory/regulatory law (e.g., Canada-provincial privacy statutes; U.S.-HIPAA Privacy Rule)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Right to Privacy – Legal sense

- Common law (e.g., tort of invasion of privacy)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

COMMON LAW TORTS FOR INVASION OF PRIVACY (U.S.)

1. Unreasonable intrusion upon seclusion
2. Publicity placing another in a false light
3. Public disclosure of embarrassing private facts
4. Appropriation of another's name or likeness

5. (Breach of confidence)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Conclusion :

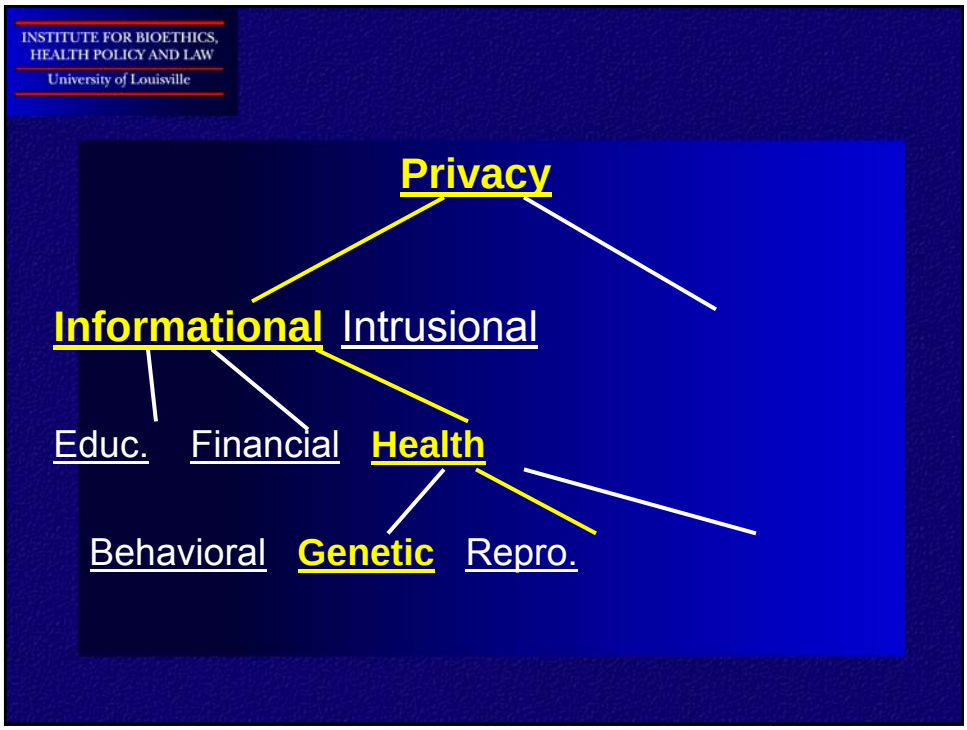
Legal right to privacy concerns discrete categories of protected interests.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

A FEW ADDITIONAL POINTS

- Privacy has an intrinsic and an instrumental value.
- Privacy is not absolute and does not always trump other interests.
- There is a social cost to privacy (e.g., clinical care, public health, research, law enforcement), but there is also a social cost to a lack of privacy.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville



INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

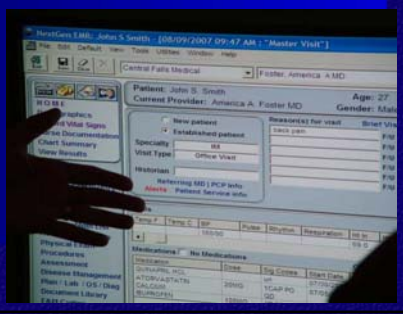
2. Assessing genetic exceptionalism

Much genetic information is sensitive.

- It implicates the health of family members.
- It is transgenerational.
- It often carries stigma and has led to discrimination and eugenics.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Because genetic information includes sensitive health information, it should be protected from wrongful access, use, and disclosure.



INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy, confidentiality, and security measures should be in place.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

But separate, genetic-specific legal measures are <u>not</u> desirable and are often counterproductive.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

1. It is impossible to define "genetic."

- Scientifically
- Legally

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

2. It is extraordinarily difficult to segregate genetic from non-genetic information in health records.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

3. Genetic exceptionalism can be a self-fulfilling prophecy.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Genetic exceptionalism is a politically expedient strategy if the choices are:

- ☐ Do nothing
- ☐ Enact comprehensive measures
- ☑ Enact a genetic-specific law

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

3. Exploring the effects on privacy of specific technologies

A. Biobanks (repositories of human biological specimens)

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

BIOBANKS DIFFER FROM TYPICAL RESEARCH

- Single protocol / many protocols
- Current protocol / future protocols
- Individual risk-benefit / individual and group risk-benefit
- Mostly physical harm / exclusively non-physical harm

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Some Key Ethical Issues

1. Consent to future uses
2. Linkage with medical records
3. Withdrawal of sample
4. Recontact

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

B. Electronic Health Records

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

EHRs and EHR NETWORKS

- Network of networks
- Interoperable
- Longitudinal
- Comprehensive

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

ADVANTAGES OF EHRs

1. Enhance coordination of care
2. Avoid duplication of services
3. Improve effectiveness of care
4. Improve efficiency of care
5. Facilitate outcomes and other research

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

DISADVANTAGE OF EHRs (FOR PRIVACY)

It eliminates the chaos of disaggregated, unconnected, fragmented, and largely paper-based health records.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

SOME KEY PRIVACY ISSUES TO BE RESOLVED

Should patients be allowed to control access to their records?

Should patients be allowed to control the content of their records?

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Security and confidentiality safeguards are not enough.

Patients are concerned about the psychic harm from having their sensitive health information accessible to both health care providers and third parties.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Should any special protections be put in place for certain classes of sensitive health information, such as mental health, substance abuse, STDs, and genetic test results?

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

C. Personal health records

- Allergies
- Diagnoses
- Family history
- Hospitalizations
- Lab values
- Medications
- Surgeries

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Freestanding PHRs

Google health

Microsoft HealthVault

PassportMD

revolutionhealth
Your home for health and balance

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

"Tethered" PHRs

Cleveland Clinic

HUMANA. Guidance when you need it most

KAISER PERMANENTE®

United States Department of Veterans Affairs

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy Issues

Business models of freestanding PHRs

- Charge monthly fee
- Sell data
- Sell ads
- Combination of the above

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Other concerns

- Lost portable devices
- Hacking

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

D. Direct-to-consumer genetic testing

Numerous commercial, often unregulated, web-based labs offer a wide range of genetic testing for the following purposes:

- Paternity
- Genealogy
- Curiosity
- Health risk assessment

These are home collection tests, not home performed tests (e.g., pregnancy).

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Scientific concerns about DTC genetic testing

- Lack of clinical validity
- Lack of inter-lab concordance
- Variance in interpretations
- Lack of genetic counseling
- "Cascade effect" of genetic testing
- Lack of clinical utility
- Psychosocial harms

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Privacy concerns about DTC genetic testing

vacy laws

nformation

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Other ethical concerns about DTC genetic testing

- Consumer protection issues (e.g., nutrigenomics)
- Lack of regulation
- Autonomy/paternalism
- Public education
- Resource consumption

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

Conclusion

"Reconsidering Privacy in the Genomic Era"

But, it's also the . . . Biobank era

- Electronic health record era
- Personal health record era
- Direct-to-consumer genetic testing era

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

- It's also the Patients Like Me era.

- It's also the facebook, myspace, Linked in era.

INSTITUTE FOR BIOETHICS, HEALTH POLICY AND LAW
University of Louisville

- We need to reconsider on an ongoing basis what privacy is and what steps we are prepared to take to protect it.


INSTITUTE FOR BIOETHICS,
HEALTH POLICY AND LAW
UNIVERSITY OF
LOUISVILLE

Panel 1A: Public Release of Genomic Data

Session Chair: Patricia Kosseim, Genome Canada

Bio:

Patricia Kosseim has recently joined Genome Canada on a two-year Executive Interchange arrangement to lead a national strategy for addressing ethical, economic, environmental, legal and social (GE^3LS) issues related to large-scale genomics research. She joins Genome Canada from the Office of the Privacy Commissioner of Canada (OPC), where she has held the position of General Counsel since January 2005, responsible for the activities of the Legal Services, Policy and Parliamentary Affairs Branch.

Before joining OPC, Patricia spent five years building and heading up the Ethics Office of the Canadian Institutes of Health Research. During this period, she was briefly seconded to Canada Health Infoway Inc. to advise on privacy issues related to the development of pan-Canadian electronic health record systems.

Patricia worked in Montreal for over six years with the national law firm of Heenan Blaikie, practicing primarily in the areas of health law, human rights, labor & employment law, privacy law, administrative law, professional liability and civil litigation.

Called to the Québec Bar in 1993, Patricia holds degrees in Business (B.Com '87) and Laws (B.C.L. / LL.B. '92) from McGill University, and a Master's Degree in Medical Law and Ethics (M.A.'94) from King's College, University of London (U.K.).

Public Release of Genomic Data: Ethical and Legal Perspectives

Yann Joly, McGill University

Abstract:

Data creation and release is exponential. The genomic research community understands that data sharing is a necessity for economic, scientific and ethical reasons. However, for all its promise, data sharing raises important ethical, legal and social challenges. This presentation will focus on two of the main issues in data sharing practices: 1) how can we motivate data producers to share their data in a timely manner with the rest of the scientific community? and 2) how can the confidentiality and the autonomy of research participants be respected in open release and access? These issues as well as some potential solutions will be contextualized by the use of a case study based on the model of the International Cancer Genome Consortium.

Bio:

Yann Joly, Ph.D. (DCL), Lawyer, is an Assistant Professor at the Faculty of Medicine, Department of Human Genetics at McGill University, as well as an ethics and legal consultant in the private sector. He is the North American coordinator of the Association de recherche et de formation en droit medical (ARFDM). His research activities lie at the interface of the fields of intellectual property, health law (biotechnology and other emerging health technologies) and bioethics. Yann Joly is the current Data Access Officer of the International Cancer Genome Consortium (ICGC).

Public Release of Genomic Data: Ethical and Legal Perspectives

Bartha M. Knoppers, P.h.D, O.C.
And
Yann Joly, P.h.D.
Centre of Genomics and Policy

Table of contents

- 1) The return of open science
- 2) Hurdles along the way
- 3) Case Study: The ICGC

1) The return of open science

- It is a sad commentary on the success of the control culture that even conversations around freedom rely on the vocabulary and ideologies of those who emphasize protection, and that freedom isn't free unless someone can get sued. But nothing other than the public domain really works from the perspective of data integration. And data integration is coming at us at exponential speed.

John Wilbanks (2008) Journal of Science Communication

1) The return of open science

Theoretical foundations:

- Heritage of Humanity (Grotius)
- Global Public Good (Hume)
- Norms of Science (Merton)
- Open Source (Stallman)

1) The return of open science
Bermuda Principles (1996)

- **Primary Genomic Sequence Should be in the Public Domain**
- "It was agreed that all human genomic sequence information, generated by centres funded for large-scale human sequencing, should be freely available and in the public domain in order to encourage research and development and to maximise its benefit to society."

1) 1) The return of open science
Ft. Lauderdale Meeting (2003)

Introduction

- "The meeting concluded that pre-publication release of sequence data by the International Human Genome Sequencing Consortium, and other sequence producers, has been of tremendous benefit to the scientific research community in general."

2) Hurdles along the way Confidentiality scare?

- "In 2004, Zhen Lin and colleagues illustrated that access to just 30–80 statistically independent single nucleotide polymorphisms (SNPs) was sufficient to uniquely identify an individual (Lin, Owen, and Altman 2004). Recently, Homer and colleagues demonstrated that an individual's SNP profile could potentially be identifiable even when it is aggregated with 1,000 or more other samples (Homer et al. 2008)".

Amy L. McGuire , The American Journal of Bioethics, (2008)

2) Hurdles along the way Confidentiality scare?

- "The ease of identifying people from DNA or genomic data, without breaking laws, should not be overstated; it takes competence, perhaps a laboratory equipped for the purpose, computational power, perhaps linking to other data, and determined effort. But some risks are real. [P]rotection of identifiability is obligatory for maintaining the trust of our most important research partners, the public."

Lowrance WW, Collins FS. *Science* (2007).

2) Hurdles along the way Confidentiality scare?

HapMap Project

- "Researchers will use the genetic variation information in the database to create a genetic map that summarizes the patterns of genetic variation, called haplotype map or "HapMap". The HapMap will be put on the Internet. The HapMap will not include medical information, but researchers will use it as a tool in future studies to find genes related to many diseases."
- "If your sample is used, lots of genetic information from your sample will be put in the database, and lots of people will be able to look at it for any purpose."

2) Hurdles along the way Confidentiality scare?

1000 Genomes Project

- "Although we will not collect any names or medical information, and we will take many measures to protect your privacy, we will generate lots of genetic information about each person whose sample is studied. This information will be put in open access scientific databases, available on the Internet to anyone who wants to look at it."
- "As technology advances, there may be new ways of linking information back to you that we cannot foresee now […]. We believe that the benefits of learning more about human genetic variation and how it relates to health and disease outweigh the current and potential future risks […]".

2) Hurdles along the way
Confidentiality scare?

- **Prepublication Data Sharing/ Toronto Statement (2009)**

  "For aggregated data that cannot be used to identify individuals, databases are open access, but for clinical and genomic data that are associated with a unique, but not directly identifiable individual, access may be restricted."

  Nature 461, 168-170 (2009)

2) Hurdles along the way
Protection instinct/ competitive science

- **Prepublication Data Sharing/ Toronto Statement (2009)**

  "If data producers request a protected time period to allow them to be the first to publish the data set, this should be limited to global analyses of the data and ideally expire within one year. If the citable statement is a 'marker paper' it should be subjected to peer review and published in a scientific journal".

2) Hurdles along the way
Protection instinct/ competitive science

Paper Retracted Following Genome Data Breach

Scooped. Another team broke the database embargo and published a paper using Laura Bierut's data.

Constance Holden, Science, 2009

2) Hurdles along the way
Protection instinct/ competitive science

Paper Retracted Following Genome Data Breach

Actions taken :

- Yale took down a press release it had posted about the study.
- NIH froze the researchers' access to dbGaP.
- PNAS retracted the paper from its print edition and added a retraction notice to the online edition.

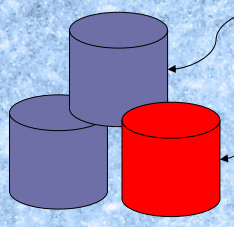
3) International Cancer Genome Consortium (ICGC)

- Moratorium to limit data users from carrying out and publishing global analyses. (users are free to use and publish data that targets specific genes or mutations).
- Duration: until 1) after the initial global analysis is published by the ICGC member project; or, 2) one year after the quantity of data on which the initial global analysis will be carried out has been released via the ICGC database or other public databases; or, 3) two years after its initial release, whichever occurs first.

3) ICGC Prospective consent

- One of the purposes of the ICGC is to support the sharing of coded data with the international research community in order to achieve its goal of facilitating and accelerating research into the causes and control of cancer. The ICGC also respects the individuals who contribute to ICGC projects and will strive to protect their confidentiality. To accomplish these aims, the ICGC has established a policy that data from participants be organized and placed into two databases, **Open and Controlled-Access.**

3) ICGC
Open vs protected data



DCC Data Mart

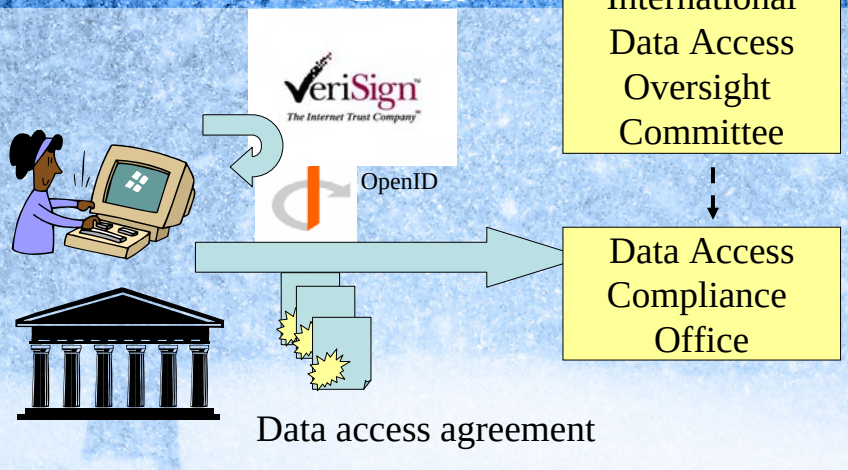
Binary access policy:

Access to protected data is all or none.

3) ICGC
Prospective consent

- **Open-access: Information in this database will be publicly accessible, but will not contain any** information that could be used to identify you specifically. It will include information about your tumour, your age range and your gender.
- **Controlled-access: This database will only contain your coded medical information and** information from the more detailed analyses of your coded samples.

3) ICGC, Obtaining Authorization for Controlled Data



Conclusion

- The rapid prepublication release of sequencing data has served the field of genomics well. The Toronto meeting participants acknowledged that policies for prepublication release of data need to evolve with the changing research landscape, that there is a range of opinion in the scientific community, and that actual community behaviour (as opposed to intentions) need to be reviewed on a regular basis.

Toronto International Data Release (2009) 461 *Nature* 168-170.

Conclusion

- (S)cholars have sought to restructure privatization to better accord with certain ideals. (W)e suggest an analogous move with respect to the public domain. We argue that leaving a resource in the public domain is not enough to satisfy societal ideals. It matters how that public domain is to be structured.

  Anupam Chander, Madhavi Sunder , "The Romance of the Public Domain" (2004) 92 California Law Review 1331

Sharing Genomic Data in the Face of Advancing Technologies and Statistics

Laura Rodriguez, National Human Genome Research Institute, National Institutes of Health

Abstract:

In 2006 the U.S. National Institutes of Health (NIH) announced a draft policy to create a central repository of individual-level genotype and phenotype data that would serve as the foundation for the creation of a community resource database to support the emerging area of genome-wide association studies (GWAS). Under the final version of this policy, which was released in 2007 and effective in early 2008, genotype and phenotype data from any study submitted for funding to conduct GWAS was expected to be deposited to the NIH GWAS Data Repository, known as the database for Genotypes and Phenotypes, for subsequent data sharing for appropriate research purposes. The data access model developed for this new type (and considerably expanded volume) of data was two-tiered. Open Access portions of the database made summary level information about studies available to anyone, and Controlled Access portions of the database displaying the individual-level information contained within the raw genotype and phenotype data were made available only to users approved by NIH Data Access Committees (DACs) to conduct investigator-specified research projects. Included within the summary-level information accessible through the Open Access portions of the database were aggregate genomic statistical tables (allele frequencies, p values for measured SNPs, etc.), as well as aggregate phenotype data tables. Sharing data in these forms was a standard and long-held practice within the research community as the data types were commonly accepted to pose no risk to the privacy or confidentiality of individual participants in the original studies. However, in August 2008, Homer et. al published a seminal paper demonstrating innovative statistical methods to resolve a known DNA genotype from within a complex mixture of DNA samples. Although this new methodology did not enable the definitive identification of data contributors to aggregate genomic data collections (such as those made publicly available through dbGaP and other GWAS resources) unless a full genomic analysis was available from an already identified source, the NIH determined that there had been a sufficiently substantive change to the risks to individual confidentiality to warrant revising GWAS policy to move aggregate genomic data from the Open Access pages within dbGaP to accessibility only through Controlled Access mechanisms. An overview of the data access policy within the NIH GWAS data sharing model and the changes that were made following the publication of the new statistical methods will be provided.

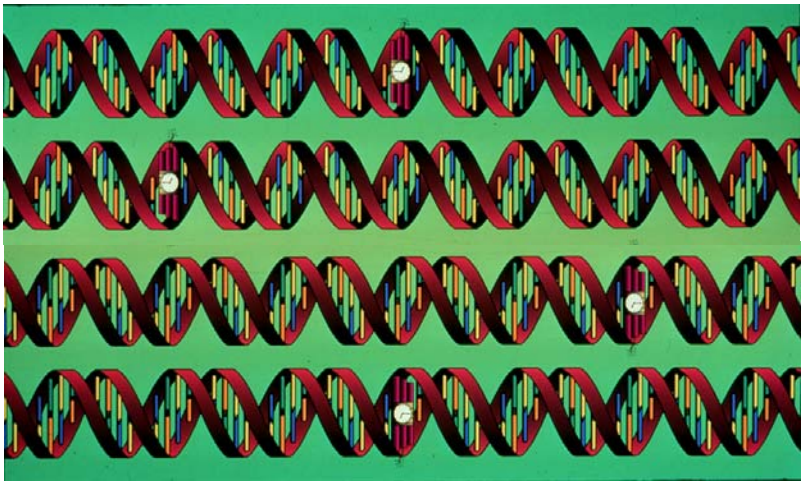
Bio:

Laura Lyman Rodriguez, Ph.D., is the Acting Director for the Office of Policy, Communication, and Education and the Senior Advisor to the Director for Research Policy at the National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH). Dr. Rodriguez works to develop and implement policy for research initiatives at the NHGRI, as well as trans-NIH programs. She is particularly interested in the policy and ethics questions related to the inclusion of human research participants in genomics and genetics research. Dr. Rodriguez is also interested in the policy and organizational issues associated with the development and establishment of strategic partnerships. Among other activities, Dr. Rodriguez provided leadership for many of the policy development activities pertaining to the Genetic Association Information Network (GAIN) as well as the development and implementation of the trans-NIH Policy for Data Sharing in Genome-Wide Association Studies (GWAS).
Dr. Rodriguez received her bachelor of science with honors in biology from Washington and Lee University in Virginia and earned a doctorate in cell biology from Baylor College of Medicine in Texas.

The NIH GWAS Policy:
Sharing data & protecting privacy

Laura Lyman Rodriguez, Ph.D.
Office of Policy, Communications,
& Education, NHGRI
National Institutes of Health

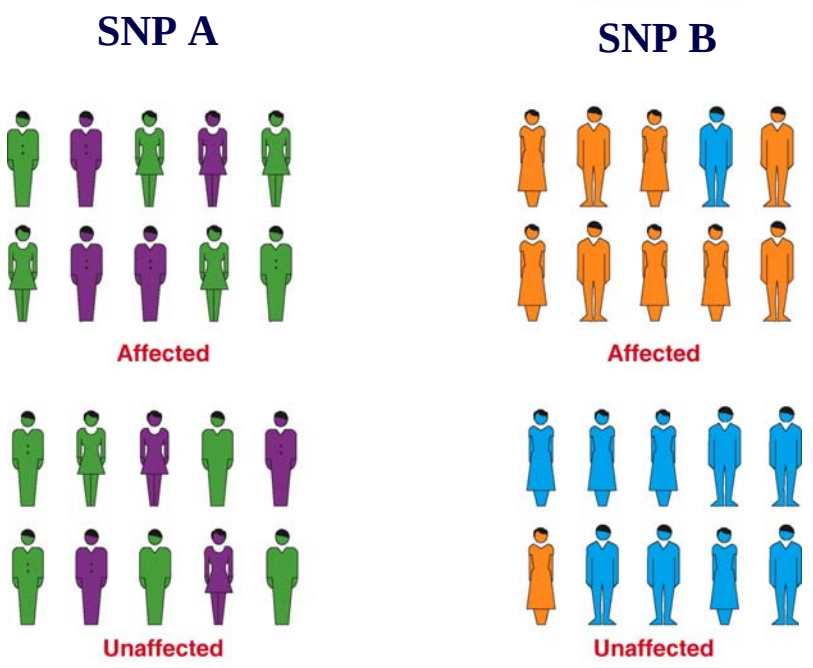
Electronic Health Information and Privacy Meeting
November 19, 2009



Everyone has a number of genetic variants that influence their risk of disease

A Genome-Wide Association Study is …

- Method for interrogating all 10 million variable points across the human genome
- Variation is inherited in “blocks” (haplotypes), so not all 10 million points have to be tested
- Recent technology advances drastically decreased costs for measuring large numbers of variant points in the genome (“SNPs”)
- Now possible to design studies with sufficient power to identify “common” variation with “modest” risk



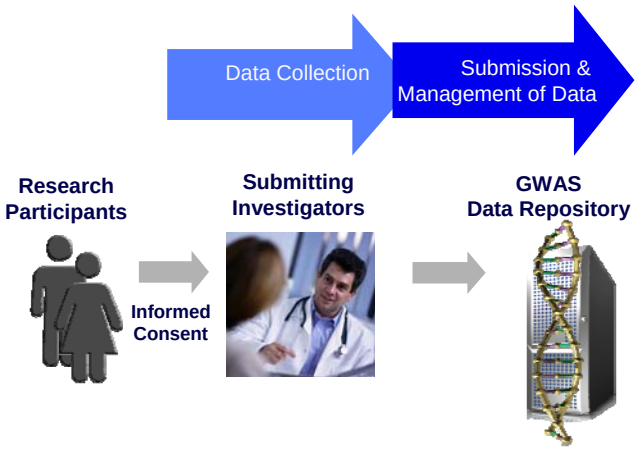
Why is this different?

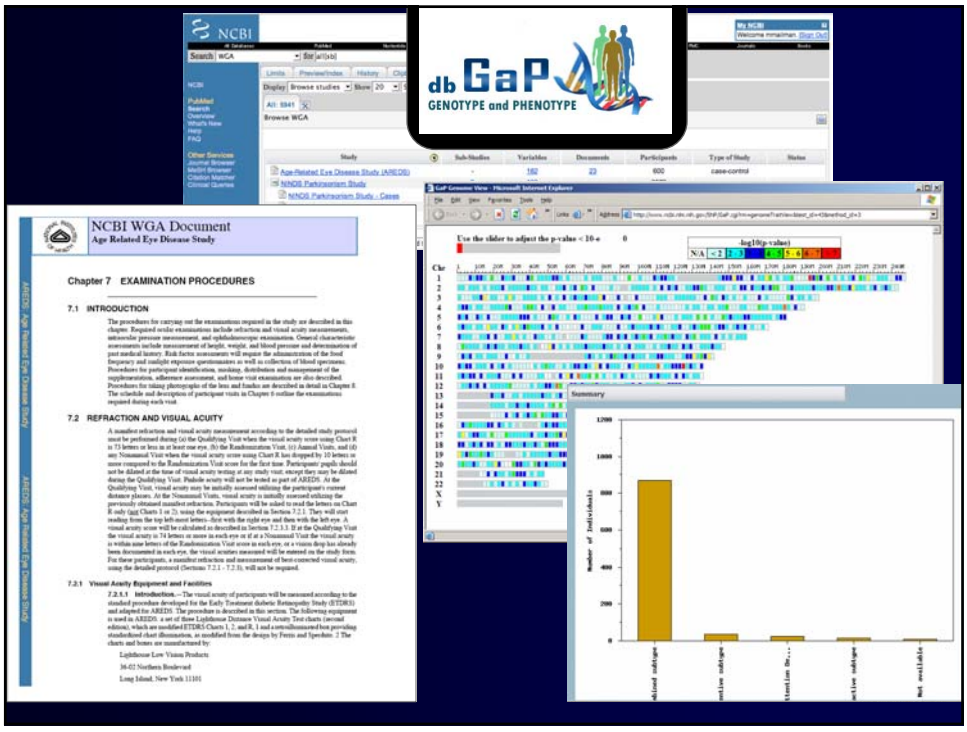
- Unprecedented opportunity to advance understanding of common diseases (e.g., diabetes, cancer, heart disease)
- The data generated is far richer than what a single investigator or a collaborative team can fully explore
  - Many different questions may be asked
  - Cross-study analyses are possible, which increases the capacity to address complex questions

Guiding Principle

The greatest public benefit will be realized if data from GWAS are made available, under terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of investigators.

GWAS Data Management Overview

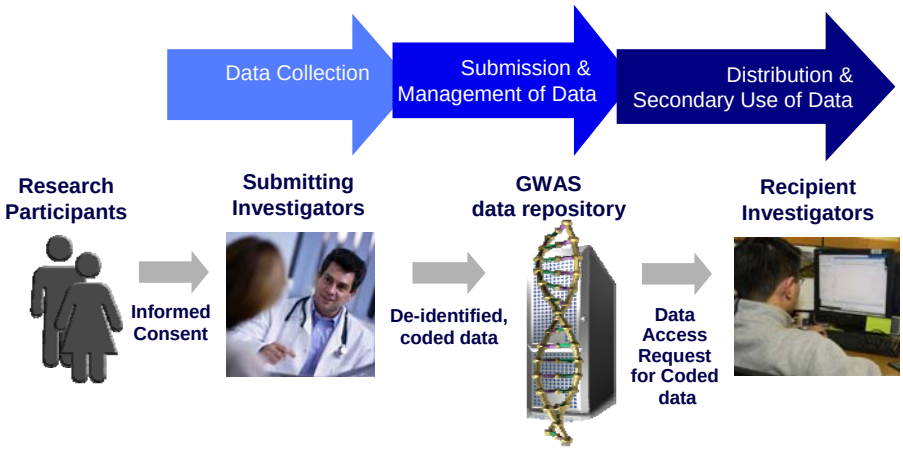




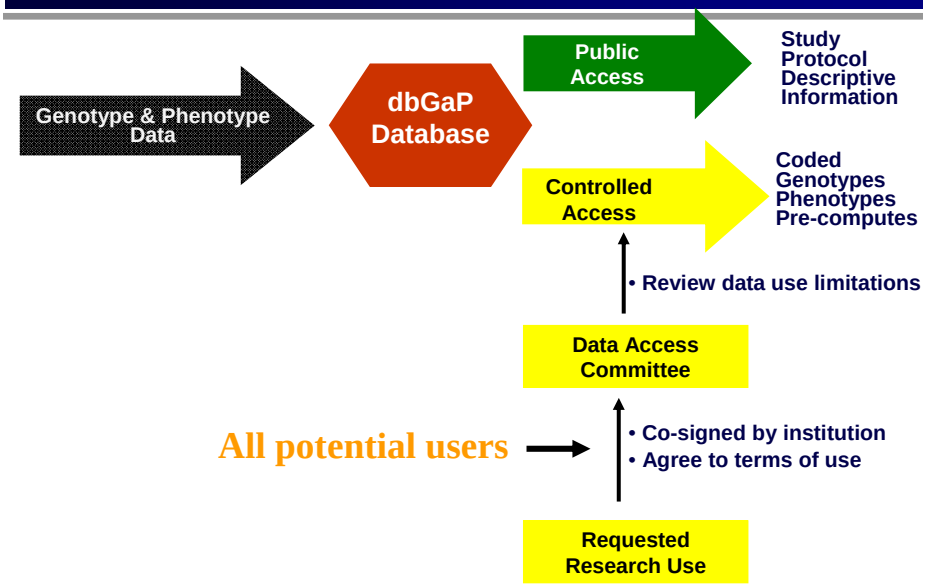
Responsibilities in Data Submission

- The **PI** will remove HIPAA identifiers and retain the keycode to the data
- **Local institution** will certify approval of submission to GWAS data repository, including statements that:
  - data are provided in accord with applicable laws and regulations
  - an **IRB** or Privacy Board has reviewed the submission plans
- Any limitations on data use are requested at time of application (e.g., limitations imposed by existing informed consent).

GWAS Data Management Overview



Data Access is Two-Tiered



Controlled Access

- Individual-level data organized by consent group
- Terms and conditions for data use agreed to by PI and home organization through Data Use Certification
- Data Access Committees (DACs) review requests for consistency with data use limitations
  - Monitor data use through review of annual reports from approved data users

Data Use Certification Agreement

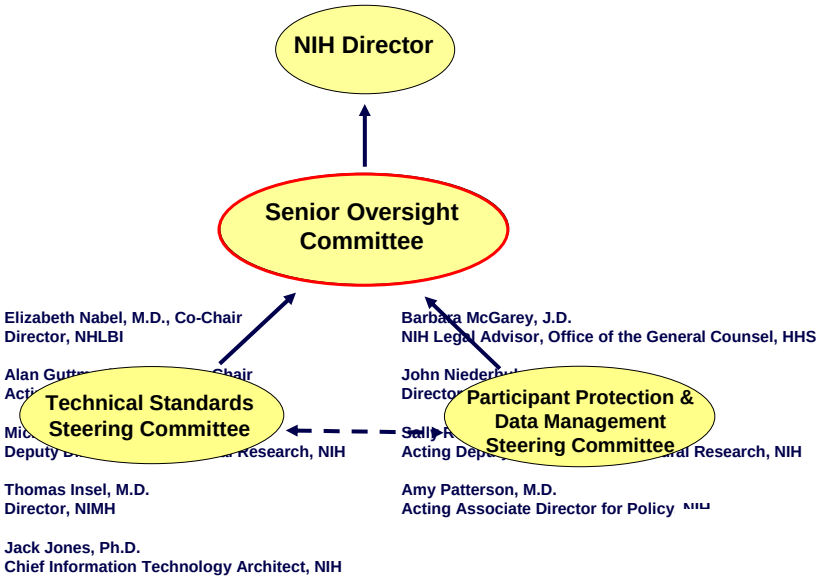
- There is a common framework for all NIH Data Use Certifications (DUCs)
- Terms and conditions include that requesters will:
  - be responsible for compliance with federal, state, and local policies
  - only use the data for the specified research use
  - not identify study participants
  - not transfer data beyond approved users
  - immediately notify the DAC if a security breach occurs
  - submit brief annual updates on research and publications
  - be identified as an Approved User within the dbGaP
  - acknowledge other GWAS policies

dbGaP by the Numbers

As of Fall 2009:

- 39 deposited studies involving 79 institutions
- 57,612 phenotypes measured
- Over 500 approved users with at least 1 project
  - Investigators span research sectors, but primarily reside in academic-based institutions
- Users from 196 institutions in 25 countries
- 48 additional studies in process

NIH GWAS Oversight Structure



OPEN ACCESS Freely available online

PLoS GENETICS

Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays

Nils Homer[1,2], Szabolcs Szelinger[1], Margot Redman[1], David Duggan[1], Waibhav Tembe[1], Jill Muehling[1], John V. Pearson[1], Dietrich A. Stephan[1], Stanley F. Nelson[2], David W. Craig[1]*

1 Translational Genomics Research Institute (TGen), Phoenix, Arizona, United States of America, 2 University of California Los Angeles, Los Angeles, California, United States of America

Abstract

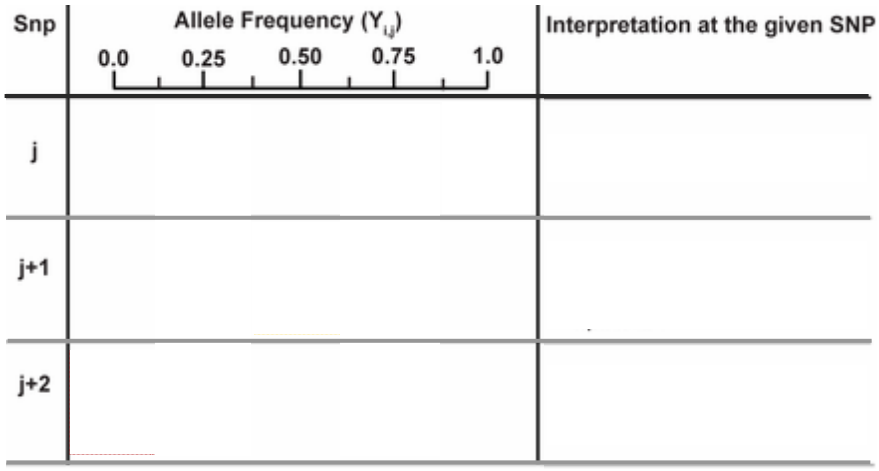
We use high-density single nucleotide polymorphism (SNP) genotyping microarrays to demonstrate the ability to accurately and robustly determine whether individuals are in a complex genomic DNA mixture. We first develop a theoretical framework for detecting an individual's presence within a mixture, then show, through simulations, the limits associated with our method, and finally demonstrate experimentally the identification of the presence of genomic DNA of specific individuals within a series of highly complex genomic mixtures, including mixtures where an individual contributes less than 0.1% of the total genomic DNA. These findings shift the perceived utility of SNPs for identifying individual trace contributors within a forensics mixture, and suggest future research efforts into assessing the viability of previously sub-optimal DNA sources due to sample contamination. These findings also suggest that composite statistics across cohorts, such as allele frequency or genotype counts, do not mask identity within genome-wide association studies. The implications of these findings are discussed.

Homer N et al, *PLoS Genet* 2008 Aug 29;4(8):e1000167.

Initial Response….You can do what?!

- Authors shared manuscript with NHGRI
- Extensive review and discussion with authors
- Designed test, provided masked data
  - Allele frequencies of 547,000 SNPs from mixture of 1,000 persons
  - Individual genotype data from 100 persons
    - 38 in mixture
    - 55 not in mixture
    - 7 parents of mixture members
  - Algorithm re-run at NCBI, results matched exactly

Inferring Placement from Allele Frequencies



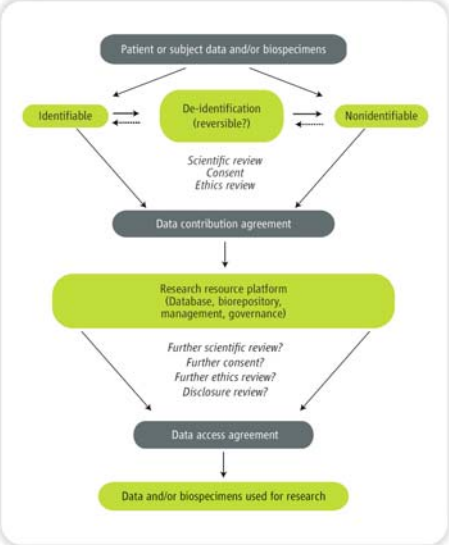
Y = Person of Interest; Pop = Reference Population; M = Mixture

Homer N et al, *PLoS Genet* 2008 Aug 29;4(8):e1000167.

Summary of Policy Actions

- Aggregate data for datasets with individual-level data moved to controlled access requests portions of the database
- Looked at ways to facilitate sharing of summary data for other datasets through additional statistical analyses
- After additional consideration, the GWAS SOC voted to make the interim policy decision final and maintain aggregate data under controlled access

Looking for Balance …

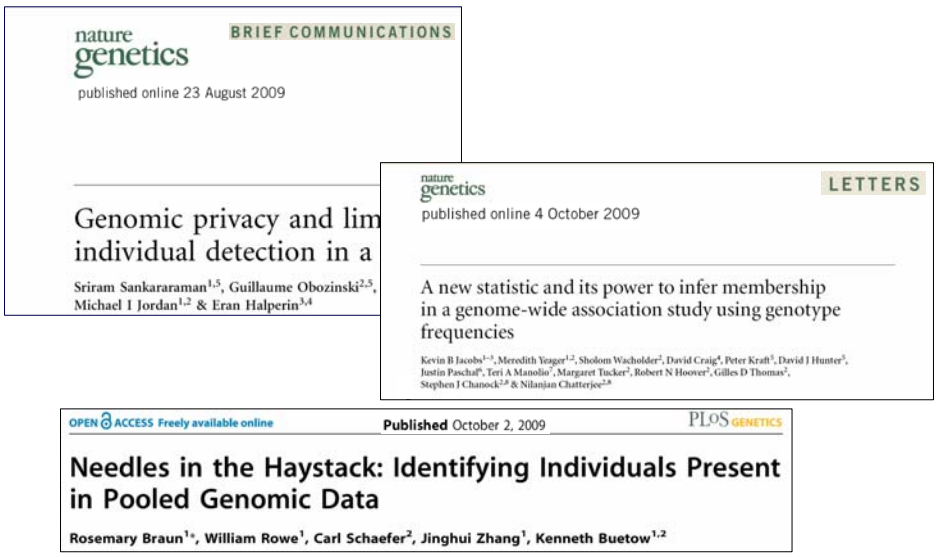


Source: Lowrance and Collins 2007

Balance scientific potential with public trust and participant protection

- Different definitions of "identifiable"
- Variety of means to render data "identifiable"
- Uncertain and debatable risk calculation

Recent Papers Continue the Discussion



nature genetics BRIEF COMMUNICATIONS

published online 23 August 2009

Genomic privacy and lim
individual detection in a

Sriram Sankararaman[1,5], Guillaume Obozinski[2,5], Michael I Jordan[1,2] & Eran Halperin[3,4]

nature genetics LETTERS

published online 4 October 2009

A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies

Kevin B Jacobs[1–3], Meredith Yeager[1,2], Sholom Wacholder[2], David Craig[4], Peter Kraft[5], David J Hunter[5], Justin Paschall[6], Teri A Manolio[7], Margaret Tucker[2], Robert N Hoover[2], Gilles D Thomas[2], Stephen J Chanock[2,8] & Nilanjan Chatterjee[2,8]

OPEN ACCESS Freely available online **Published** October 2, 2009 PLoS GENETICS

Needles in the Haystack: Identifying Individuals Present in Pooled Genomic Data

Rosemary Braun[1]*, William Rowe[1], Carl Schaefer[2], Jinghui Zhang[1], Kenneth Buetow[1,2]

Acknowledgements

GWAS Senior Oversight Committee

and

Marianna Bledsoe, NIH
John Burklow, NIH
Stephanie Burrows, NHLBI
Sarah Carr, NIH
Stephen Chanock, NCI
Mike Feolo, NCBI
Kevin Jacobs, NCI
Raynard Kington, NIH
David Lipman, NCBI
Teri Manolio, NHGRI
Jim Ostell, NCBI
Justin Paschall, NCBI
Larry Thompson, NHGRI
Peggy Tucker, NCI

Headlines…

Los Angeles Times

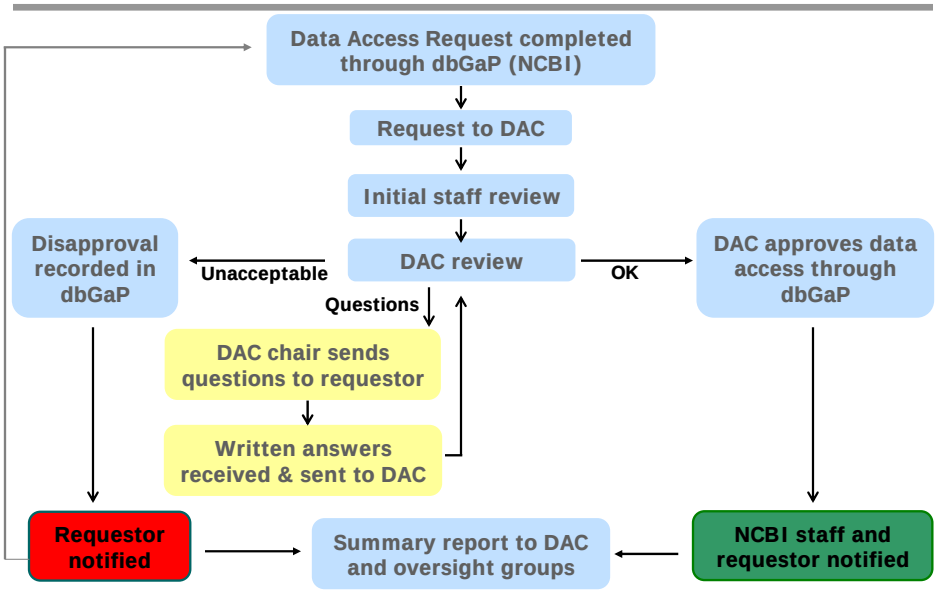
DNA databases blocked from the public
The National Institutes of Health removes patients' genetic profiles from its website after a study reveals that a new type of analysis could confirm identities.
By Jason Felch
Los Angeles Times Staff Writer

Good for Cops, Bad for NIH

By Jennifer Couzin
*Science*NOW Daily News
29 August 2008

Forensic Breakthrough Stirs NIH to Close GWAS Data from Public View
August 29, 2008
By Matt Jones,
a GenomeWeb staff reporter

Data Request Review Procedures



GWAS Policy Elements

- **Data Management**
  - **Data Submission Procedures**
  - **Data Access Principles**
  - **Protection of Research Participants**
- **Scientific Publication**
- **Intellectual Property**

Points to Consider for IRBs

- Provides investigators & IRBs with information on important participant protection considerations related to submission of data
- Not intended to serve as a checklist
- Topics include:
  - Background on the scientific opportunities presented by GWAS
  - Discussion of the ethical issues relevant to the review of submission plans for GWAS datasets
  - Specific points to consider in the evaluation of informed consent documents

Annual Reports Elements

- Summary of research progress
- Proposed plans for further research utilizing currently approved NIH GWAS datasets
- List of all completed or accepted scientific presentations that include (or will include) findings made with the individual-level NIH GWAS data accessed through dbGaP.
- List of manuscripts submitted
- Description of any intellectual property generated as a result of using the NIH GWAS individual-level data
- Summary information on any inappropriate data release incidents or other data security issues
- General comments on process & Suggestions for improving dbGaP, NIH GWAS, study-specific data access, or NIH GWAS policy or procedures in general

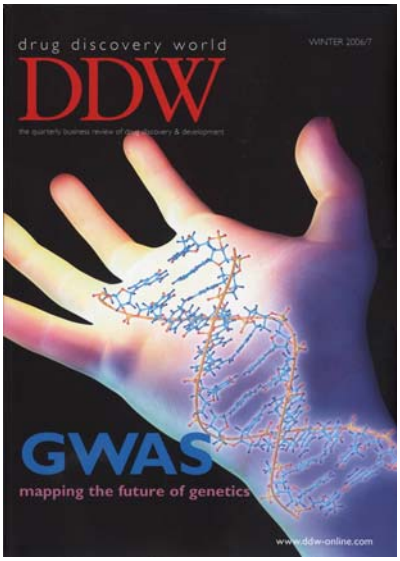
Scientific Publication

- Contributing PIs will have the exclusive right to submit publications for twelve months after a GWAS dataset is made available
  - This includes any form of public dissemination
- All other appropriate uses of the data are permitted during this period

Intellectual Property

- Consensus is that GWAS data should be pre-competitive
  - Automated calculations to identify first round genetic associations will be made available through dbGaP
- NIH urges that associations remain available to all investigators & discourages premature claims
- Users & data submitters must “acknowledge” this position
- NIH encourages broad use of GWAS data consistent with NIH’s Best Practices for Licensing with Genomic Inventions.

What is a GWAS?



"any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition."

Any study what ???

Open Access Clinical Trials to Achieve Clinical Proof-of-Concept

Aled Edwards, University of Toronto

Abstract:

The release of genome sequence information is to some extent “yesterday’s news”; we must prepare for public release of data that link genotype to function. One area of science that could benefit immensely from open access to information is drug discovery. Drug discovery resources in academia and industry are currently not used efficiently. Duplication could be reduced, productivity could be increased, and fewer patients subject to harm by performing clinical proofs of concept within open access industry-academia partnerships. The power of these experiments is dependent on having access to patient metadata and genome data. Efforts are underway to launch such studies, and it would be great if the framework for data release and management could anticipate these studies.

Bio:

Aled Edwards, Ph.D. is Banbury Professor of Medical Research at the University of Toronto, Canada, Visiting Professor of Chemical Biology at the University of Oxford and the Director and CEO of the Structural Genomics Consortium, an Anglo-Canadian-Swedish public-private partnership devoted to open-access drug discovery science.

Dr. Edwards co-founded Affinium Pharmaceuticals, a Toronto-based anti-infectives company and Scate Consultants Inc, a company that commercializes bioremediation intellectual property. He also served as the Scientific Consultant on the Canadian television drama ReGenesis. He has served in management and advisory capacities for several biotechnology companies, international research consortia and funding agencies.

SGC

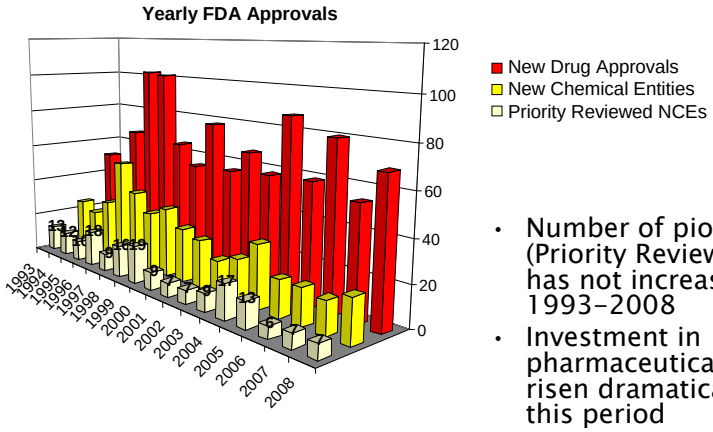
Open Access Clinical Probes

EHIP

November 2009

SGC Toronto | SGC Oxford | SGC Stockholm

The Challenge of Pioneer Drug Discovery



Public Data from Center of Drug Evaluation and Research: www.fda.gov/cder/

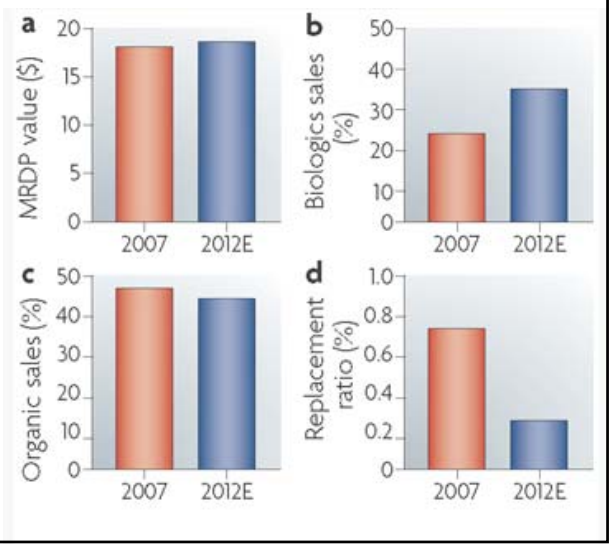
- Number of pioneer drugs (Priority Reviewed NCEs) has not increased from 1993-2008
- Investment in pharmaceutical R&D has risen dramatically over this period
- >90% failure rate in clinical trials for pioneer drugs due to lack of efficacy

Cost of drug discovery

- >>300,000 scientists work in pharma R+D (on top of ~$60B in basic research funded by governments and charities)
- ~ 60 approved drugs are approved
- ~ 40-45 of these are “re-worked” existing drugs
- ~ 20 are “new” chemical entities
- ~ 8 are “pioneer” medicines (fast tracked through FDA)
- 1 approved medicine for every ~5,000 people-years of work
- 1 novel medicine for every 15,000 people-years of work
- 1 pioneer medicine for every 35,000 people-years of work

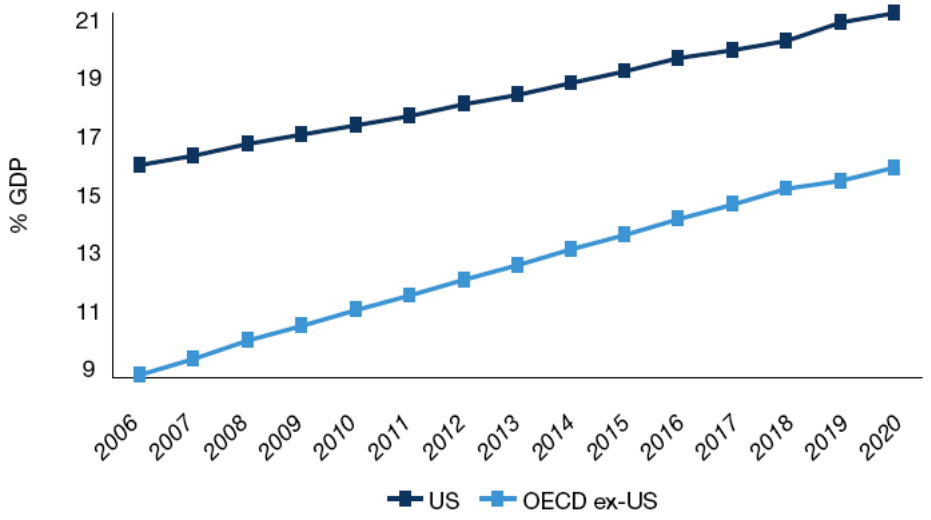
Pharma is losing its luster to investors

1.1% growth predicted over next 5 years

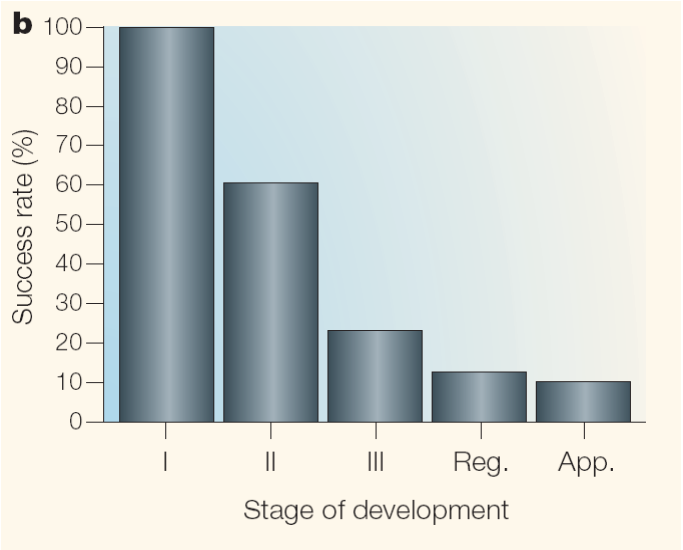


Governments are more price-conscious

Figure 6: Health expenditure as a percentage of GDP is increasing rapidly in the OECD countries



The best thinkers are not useful



Major Sources of Attrition: Efficacy & Toxicity

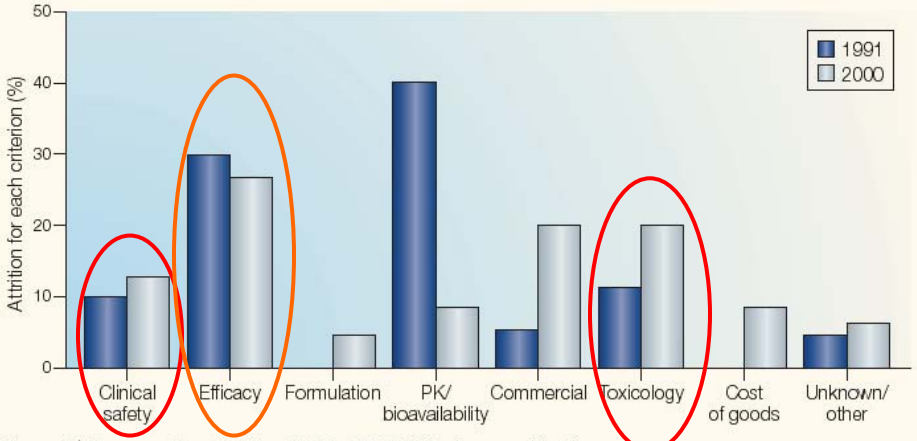


Figure 3 | **Reasons for attrition (1991–2000).** PK, pharmacokinetics.

What can we do?

1. Smarter target selection
   a. Improve understanding of biochemical/signalling networks & pathways
   b. Early access to chemical probes for pathway deconvolution
   c. Combine complementary approaches, eg. RNAi, KO, small molecule, HCS
   d. Use human cells/tissue in early target validation, rely less on orthologues
   e. Early development of biomarkers predicting efficacy

2. Smarter molecule selection
   a. Improve predictive toxicology (avoid dose-limiting toxicity)
   b. Improve PK/PD understanding: is the drug reaching the target?

3. Smarter organization of resource
   a. Stop multiple duplication of high-risk research
   b. Redefine pre-competitive boundaries for novel targets & pathways
   c. Pool resources and carry out high-risk PoC studies in the open

Open access science to promote "pioneer" drug discovery

Chemical probes in epigenetics

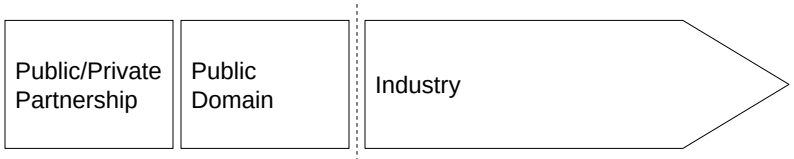
What can we do?

1. Smarter target selection
   a. Improve understanding of biochemical/signalling networks & pathways
   b. Early access to chemical probes for pathway deconvolution
   c. Combine complementary approaches, eg. RNAi, KO, small molecule, HCS
   d. Use human cells/tissue in early target validation, rely less on orthologues
   e. Early development of biomarkers predicting efficacy
2. Smarter molecule selection
   a. Improve predictive toxicology (avoid dose-limiting toxicity)
   b. Improve PK/PD understanding: is the drug reaching the target?
3. Smarter organization of resource
   a. Stop multiple duplication of high-risk research
   b. Redefine pre-competitive boundaries for novel targets & pathways
   c. Pool resources and carry out high-risk PoC studies in the open

Pre-Competitive Chemical Biology

- Situation
  - Drug discovery scientists stay away from many pioneer/potential drug targets because of the high probability of failure in clinical trials
  - Engagement of the academic community in target validation is hampered by lack of proper tools
- Aims
  - Develop interest for these targets by generating chemical inhibitors ("chemical probes") and providing them to academic and industrial scientists, unencumbered
- Scientific Plan
  - Partnership to develop quality chemical probes using biological expertise in academia and medicinal chemistry from industry
  - Fund effort using both public and private resources
  - Place chemical probes into the public domain, unencumbered
  - Scientists will use probes to enhance knowledge about human biology

Model for Pre-Competitive Chemistry



Chemical Probes	**Target Validation**	**Drug Discovery**
Screening	No IP	(re)Screening
Chemistry	No restrictions	Chemistry
Structure	Publication	Lead optimization
Bioavailability		Pharmacology
		DMPK
		Toxicology
		Chemical development
		Clinical development

Creative commons *Proprietary*

The Chemical Biology Consortium

SGC

Objective: make 37 probes and data publicly available (no restriction on use) over 4 years

Participants:

- **SGC – Toronto (HMTs, Royal Family, HATs)**
- **SGC – Oxford (KDMs, Bromo domains)**
- **SGC – Stockholm (PARPs)**
- **GSK Exploratory Chemistry (8 med chemists)**
- **NIH Chemical Genomics Center (20 HTS)**
- ***OICR medicinal chemistry (3 FTE)***
- ***Frye Lab, UNC (2 FTE)***

Funder

Province $4.6M

Wellcome T. $8M

(Pending)

SGC

Why should academia carry out open-access medicinal chemistry? High Quality Chemical Probes have significant "academic" impact

Data compiled from Google Scholar, October 5, 2007 (Tim Willson, GSK)

Compound	*Receptor*	*Papers*	*Citations*	*Years*	*h-index*[a]	*g-index*[b]	*Comments*
GW1929[c]	PPARγ	317	11063	14	47	100	Agonist. Oral activity
GW0742[d]	PPARδ	392	7212	10	41	78	Agonist. Oral activity
GW4064	FXR	250	4482	8	37	61	Agonist. Oral activity
SR12813	PXR	127	4628	8	33	67	Agonist. In vitro probe only
GW9662	PPARγ	528	4513	8	32	50	Irreversible antagonist. In vitro probe o
GW3965	LXR	181	3073	7	29	53	Agonist. Oral activity
GW7647	PPARα	118	2312	7	22	47	Agonist. Oral activity
CITCO	CAR	73	711	5	14	24	Agonist. In vitro probe only

[a] Hirrsch's h-index: a metric of academic impact, combining quality with quantity
[b] Egghe's g index: a modification of the h-index with more weight on highly-cited articles
[c] includes citations for close analog GW7845
[d] includes citations for close analog GW501516

SGC

Clinical proof-of-concept
Opportunity for greater impact?

SGC Toronto | SGC Oxford | SGC Stockholm

What can we do?

1. Smarter target selection
   a. Improve understanding of biochemical/signalling networks & pathways
   b. Early access to chemical probes for pathway deconvolution
   c. Combine complementary approaches, eg. RNAi, KO, small molecule, HCS
   d. Use human cells/tissue in early target validation, rely less on orthologues
   e. Early development of biomarkers predicting efficacy

2. Smarter molecule selection
   a. Improve predictive toxicology (avoid dose-limiting toxicity)
   b. Improve PK/PD understanding: is the drug reaching the target?

3. Smarter organization of resource
   a. Stop multiple duplication of high-risk research
   b. Redefine pre-competitive boundaries for novel targets & pathways
   c. Pool resources and carry out high-risk PoC studies in the open

Pre-Competitive Clinical Probes

Situation	Action	Outcome
Over the past decades • Fewer drugs approved despite increased investment by government and industry • Clinical trials are main expense • 80% of clinical trials fail for pioneer drug targets due to lack of efficacy or target-related side effects • Drug development programs aiming to exploit the same target or pathway often are launched in parallel by multiple companies - and often fail in parallel • Rarely does the first drug molecule for a pioneer target proceed through clinical development to drug registration	• Increase the number and quality of clinically-validated drug discovery targets • Form public-private partnerships to provide clinical validation for multiple pioneer targets, rather than focus in parallel on fewer targets. • Perform clinical trial within an open-access model, to capitalize on the breadth of experience available in academia, industry and regulatory agencies, without impeding knowledge flow • Fund through a charitable or non-profit intermediary organization	• Portfolio of clinically-enhanced validated pioneer targets, with consequent reduction in scientific and economic risk to drug discovery • Freedom to operate for all parties increases opportunity for true collaboration • Systematic approach leads to improved understanding of human physiology, human pharmacology and disease processes • Increased awareness within the lay public of the scientific and organizational challenges of drug discovery science • Increased number of pioneer drug approvals

Drug discovery is secretive and duplicative

Costly examples

- MMPs for cancer
- FTI for cancer
- Substance P receptor for pain

And it continues – Aurora Kinase Inhibitors

>60

- Antimitotic kinase – potential treatment for numerous cancer types
- Will also affect healthy proliferating cells – risk of low TI
- >60 separate organizations have pre-clinical programs with patents
- 11 compounds in Phase I
- Further 4 compounds in Phase II
- Estimated total expenditure >£200M
- No data available on outcomes of clinical studies, apart from rumours

11

AT9283
PF03814735
AS703569
AMG-900
KW-2449
CYC116
AZD-1152
MLN-8054
VX-667
SU-6668
SNS-314

4

MLN-8237
PHA-739358
VX-680
ENMD-981693

Preclinical | Phase I | Phase II

SGC

Conclusion:
The current model to discover medicines needs a change and change will not come from within

SGC Toronto | SGC Oxford | SGC Stockholm

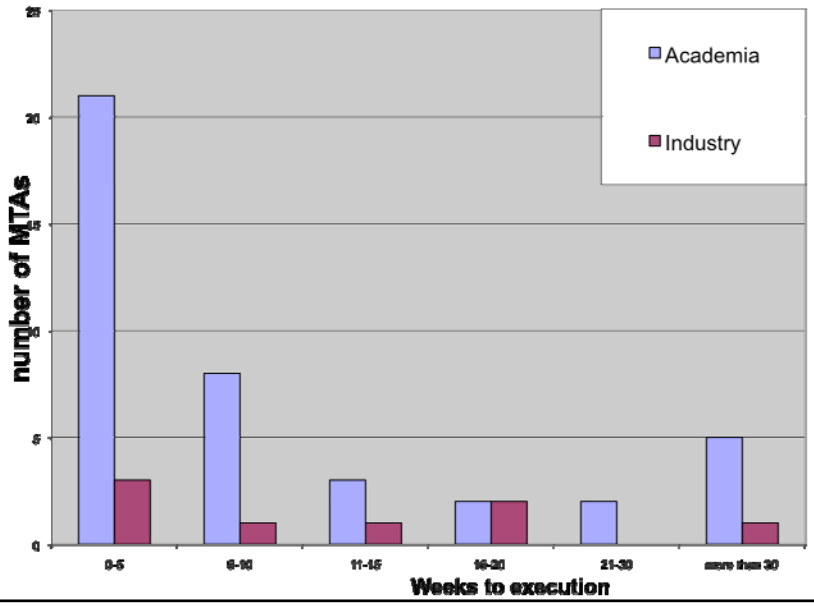
Open access clinical probe model

1. Create PPP to oversee “open” clinical PoC trials
   - Involve regulators and lay public
2. Prioritize target areas
   - Focus on pioneer targets
3. Perform all experiments without patent protection
   - Eliminates negotiations over “who owns” and allows access to best basic and clinical scientists
4. Release all information from basic to clinical studies into public domain (with adequate protection of privacy)

Outcomes

1. High quality science linking target to disease
2. Target invalidation (in 90% of cases)
3. Positive clinical PoC to reduce probability of failure in late clinical trials, and to reduce harm to patients
4. Greater understanding of challenges of drug discovery
5. More clinical PoC’s per $ invested

Why do I focus on the need for open access?



Session 2A: Data Linkage and Privacy

Session Chair: Liam Peyton, University of Ottawa

Bio:

Liam Peyton, Ph.D., P.Eng., is a principal investigator for the Intelligent Data Warehouse laboratory and Associate Professor at the University of Ottawa. He is a member of the Hospital Data Warehouse Association and an active research collaborator with The Ottawa Hospital around issues related to the collection, protection and dissemination of healthcare data for improving quality of care. He has degrees from Aalborg University (Ph.D. 1996), Stanford University (M.Sc. 1989), and McGill University (B.Sc. 1984).

Data Linkage and Privacy at a People Search Engine

Andrew Borthwick, Intelius, Inc.

Abstract:

A great deal of public information is available about a large fraction of the population. Examples of such public information sources include the telephone white pages, real estate transactions, criminal convictions, records of civil suits, and information posted on the Internet. While all of this information is theoretically available to anyone, people search engines add value for consumer and business clients by linking this information together into a single profile for each individual. Once this information is linked, it can be tremendously valuable. For instance, a woman can learn before going out on a date whether a man has a criminal background. A prospective employer can avoid unpleasant surprises before extending a job offer.

Achieving this linkage poses many technical and ethical challenges, however. This talk will first give a high level description of how Intelius is able to match this information using modern information extraction and machine learning technologies. We will then turn to some of the ethical and business challenges surrounding data linkage, including:

- Tolerance for error.
- Representing uncertain information
- “Privacy by obscurity” is now impossible
- Benefits of modern linkage technologies
- Privacy hazards of data linkage
- Steps that can be taken to mitigate privacy loss

Bio:

Andrew Borthwick is Principal Scientist at Intelius, Inc., where he works primarily on information extraction (finding information about people from the web), person matching (is the “John Smith” on web page A the same as the “John Smith” on web page B?), and people search. Prior to Intelius, Dr. Borthwick worked for Spock Networks in these areas up until its acquisition by Intelius and founded a company, ChoiceMaker Technologies, which focused on person matching. Dr. Borthwick’s credentials in the field of person matching include two U.S. patents, multiple published papers and invited talks, and serving as the principal investigator of a series of Small Business Innovation Research Grants from the National Science Foundation. He was retained as an expert witness on person matching issues in two voting rights cases by the Brennan Center for Social Justice and by American Express in a major commercial case. Dr. Borthwick received his Ph.D. in Computer Science in 1999 from New York University.

Data Linkage and Privacy at a People Search Engine

Andrew Borthwick, Ph.D.
Principal Scientist
Intelius, Inc.

November 19, 2009

About Intelius

- Primary business is empowering consumers & businesses with services to make intelligent decisions about personal safety & security
- See reports on people covering
  - Property ownership
  - Background check
  - Phone verifications
  - Web presence
  - Vital Records

Intelius Customers

- Business
  - Screening potential hires
  - Screening renters
  - Screening potential clients/business partners
- Consumer
  - Screening people before a date
  - Identity Monitoring & Management
  - Family Protection & Personal Safety

Intelius Data Sources

- Telephone White Pages & Connection Status
- Public records data
  - Real estate transactions
  - Civil & Criminal Records
  - Demographic Records
- Web data
  - Public social network profiles
  - Blogs, news articles
  - Biographical data from general web
    - I.e. Bio's of corporate officers

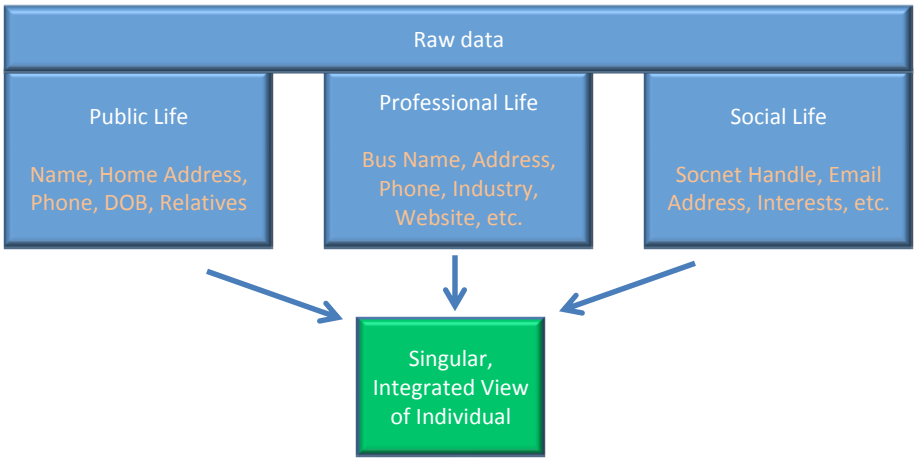
Why data linkage at Intelius?

- Get all information about a given person on one report
- Don't mix information from multiple people on same report
  - This is more serious, from Intelius' point of view

Health applications

- Immunization registries
  - Track every shot administered to every kid
    - Notify parents when child needs a shot
    - Let doctors know what shot a child has received
- Epidemiological information
  - Link death registry and cancer registry
- Contain epidemics
  - Track who has a disease

Task Overview



Record Linkage Stages

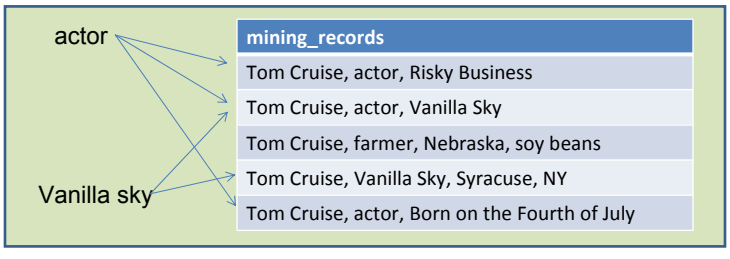
- Normalization
- Blocking
- Pair-wise decision-making
- Clustering

Normalization

- Standardize addresses
  - 42 Chestnut Street East #5 → 42 Chestnut St E Apt 5
- Map nicknames to canonical names
  - Andy, Drew → Andrew

Blocking

- If we test each database record for linkage against every other database record, we'll do almost n-squared pair-wise checks
  - 1 million records → c. 1 trillion comparisons
- Solution: Divide the records into "blocks" that match on name, job, education, location, etc…



• Only test linkage against database records that have at least one matching field

Core Pair-wise Matching

- Determine if pairs of records proposed by blocking are, in fact, the same
- Use indicators of match/no-match decision called “features” or “field-comparators”
- Features are combined into an overall judgment

Example Features

- Names match and have frequency X
- Names are an approximate match
  - Phonetic codes (i.e. Soundex) match
- Locations are Y miles apart
- Birthday’s match/don’t match
- etc.

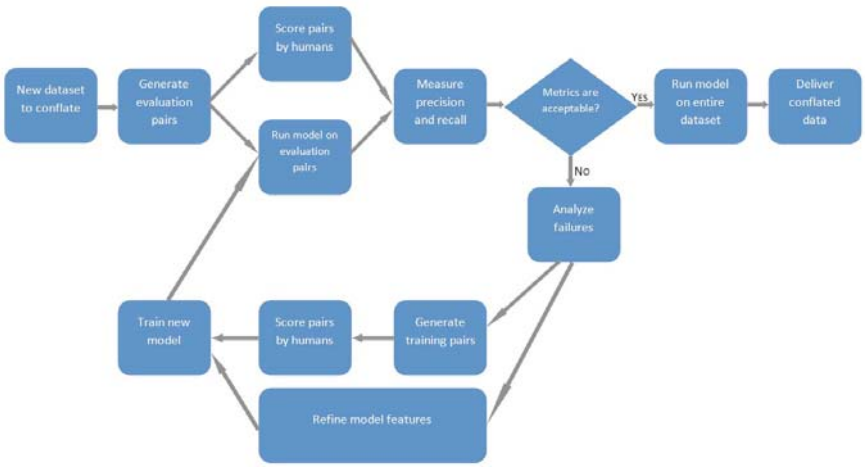
Pairwise Decision Making

- Given the features indicating a match or non-match, is the pair a match?
- Two ways to do this:
  - Rules
    - Match on first name, last name, and zipcode → match
  - Machine learning
    - Automatically determine the feature combinations that indicate a match

Data Requirements

- Training data
  - Machine learning uses to identify useful features, assign weights, and build decision trees
  - Human uses it to develop features
- Evaluation data
  - What is the accuracy of my model?
- Randomly sample pairs of records and human-tag them to generate training and evaluation data

Model Building Cycle



Clustering—Example

- A pair-wise model could say
  - A=B
  - A=C
  - B≠C
- Example
  - A and B are linked by location
  - A and C are linked by middle name
  - B and C are clearly not the same due to differing birthdays

Basic Accuracy Measures

- Precision
  - True Positive/(True Positive + False Positive)
- Recall
  - True Positive/(True Positive + False Negative)

	Machine Match	Machine No-Match
Human Match	True Positive	False Negative
Human No-match	False Positive	True Negative

Societal Implications

- Privacy
- Benefits and Hazards
- Tolerance for error

Privacy by Obscurity

- Some data used to require a private eye and a trip to the court house
  - Marriage
  - Civil Suits
  - Property ownership and property tax
  - Criminal records
- Difficult to track people down who don't want to be found

Benefits and Hazards

- Benefits
  - Know if your date is married/a criminal
  - Women know where abusive ex-husband is living
- Hazards
  - Abusive ex-husband knows where ex-wife is living
  - Dogged for life by youthful indiscretions

Mitigating the bad stuff

- Search engines can work with organizations representing people who don't want to be found
  - Battered women's shelters
  - Police officers unions
- People can correct their data or request that it be removed

Error Tolerance (1)

- Do you care more about precision or recall error?
  - For a search engine, a precision error is worse than a recall error
    - Don't falsely mix two people's profiles
  - For an airline, a recall error is worse
    - Need to identify all possible terrorists who are getting on the plane
- The is always a tradeoff between precision and recall

Error Tolerance (2)

- Decisions are made based on probabilities. Can't be 100% certain
  - Possible that there could be two people named "Andrew Borthwick" in Palo Alto, CA

For further information

- Technical overview
  - Duplicate Record Detection: A Survey. AK Elmagarmid, PG Ipeirotis, VS Verykios - IEEE Transactions on knowledge and data engineering, 2007
- Evaluation
  - Web Person Search 1 and 2
  - http://nlp.uned.es/weps/
- Open Source Toolkit
  - FEBRL: Python based. Good implementations of many string comparison algorithms.
    - http://datamining.anu.edu.au/projects/linkage.html

Questions?

Andrew Borthwick, Ph.D.
Principal Scientist
Intelius, Inc.
aborthwick@intelius.com

Privacy and Policy Considerations in Use of DNA Data Banks

Frederick Bieber, Harvard Medical School

Abstract:

Dr. Bieber will present an overview of the rationale and utility of forensic DNA testing, with an introduction to the issues relevant to DNA profiling of convicted offender registries in the U.S., Canada and other countries. The presentation will highlight the most compelling contemporary policy considerations relating to genetic privacy, disposition of DNA samples and use of new electronic data searching algorithms. He will comment on the evolving and expanding categories for inclusion in offender registries.

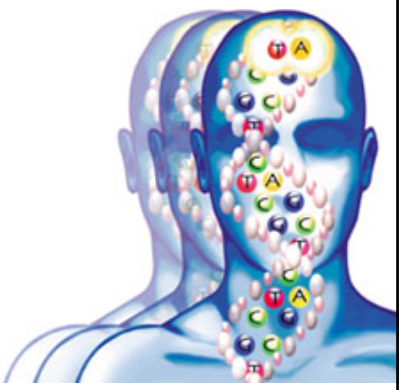
Bio:

Dr. Frederick R. Bieber serves as Medical Geneticist at Brigham and Women's Hospital and as Associate Professor of Pathology at Harvard Medical School in Boston, MA. His work focuses on the forensic aspects of DNA-based human identification, leading to his involvement in hundreds of civil and criminal cases. He has participated in the publication of over 100 articles, chapters, and books in human genetics,pathology and forensic medicine. He has testified as an expert witness in state, federal, and military courts in the U.S. and abroad, providing pro bono service to Innocence Projects representing clients who were convicted before the modern era of DNA testing and now seek exoneration by testing old evidence. He has expertise in firearms and ballistics and firearms injuries and lectures on this subject to physicians and medical students in various courses at Harvard. Dr. Bieber serves on advisory boards of the Federal Bureau of Investigation, the Royal Canadian Mounted Police, and the United States Department of Defense.

Link to video of this presentation.

Privacy and Policy Considerations in Use of DNA Data Banks

Frederick R. Bieber
Brigham and Women's Hospital
Harvard Medical School
Boston, MA



Electronic Health Information and Privacy Conference
Ottawa, Canada
19 Nov 2009

OPEN ACCESS Freely available online PLoS GENETICS

Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays

Nils Homer[1,2], Szabolcs Szelinger[1], Margot Redman[1], David Duggan[1], Waibhav Tembe[1], Jill Muehling[1], John V. Pearson[1], Dietrich A. Stephan[1], Stanley F. Nelson[2], David W. Craig[1]*

1 Translational Genomics Research Institute (TGen), Phoenix, Arizona, United States of America, 2 University of California Los Angeles, Los Angeles, California, United States of America

Finally, it is important to consider these findings in light of GWA studies. Indeed, the push to develop high-density SNP genotyping arrays is largely driven by the desire to identify common variants predisposing to a disease. For many GWA studies, the overall cost of genotyping thousands of individuals is substantial. However since genotype data is transferable and can be combined with data from other studies, there is a considerable effort to make experimental data publicly available. As part of this effort, many studies provide pooled allele frequency data in the form of summary statistics (e.g. allele frequencies or genotype counts), in part to mask individual-level genotype data. Though counter-intuitive, our findings show a clear path for identifying whether specific individuals are within a study based on summary-level statistics. Such approaches may have specific utility for identifying redundant individuals when new individual-level genotype data is combined with previous studies sharing only summary statistics.

Considering privacy issues with genetic data, it is now clear that further research is needed to determine how to best share data while fully masking identity of individual participants. However, since sharing only summary data does not completely mask identity, greater emphasis is needed for providing mechanisms to confidentially share and combine individual genotype data across studies, allowing for more robust meta-analysis such as for gene-environment and gene-gene interactions.

Canada OPC Priorities

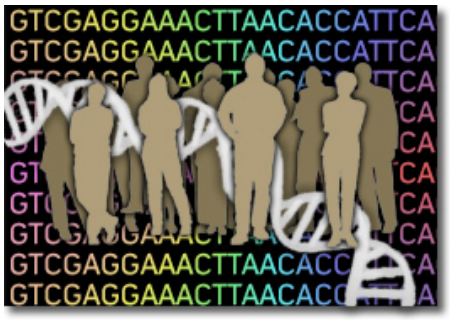
- Information Technology
- National Security
- Identity Theft
- **Genetic Information**

OPC- Canada

- **Genetic Information**
- Advances in genetics have important implications for privacy. Genetic testing for employment, criminal matters, research, medical care, access to insurance and genetic testing to determine biological relationships all raise significant privacy issues.
- **Our aims are to:**
  - Advance research and knowledge to address some of the new challenges posed by genetics in the context of traditional data protection regimes. These challenges include the right *not* to know and the concept of privacy in a world where a genetic sample offers information about not only an individual, but also about his or her family members.
  - Raise public awareness about the many potential uses of genetic information.

Genetic testing

1. Diagnostic
2. Carrier testing
3. Prenatal or preconception testing
4. Newborn screening
5. Predisposition testing
6. Pharmacogenetics
7. Forensic and paternity testing
8. Bio-geographic ancestry prediction



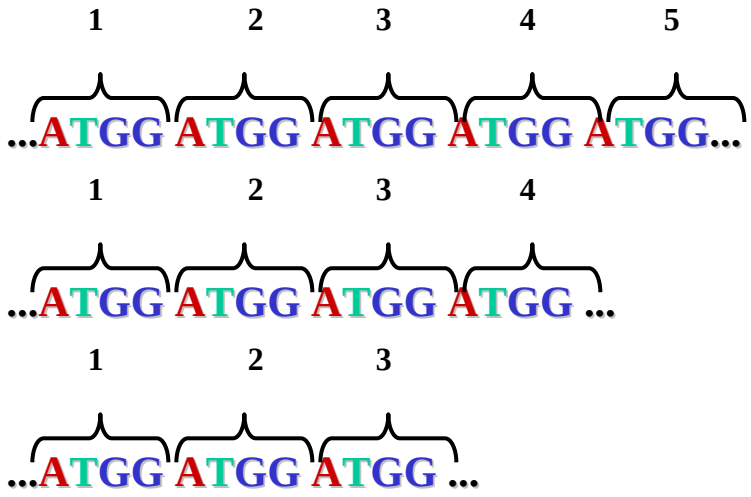
Possible Social Uses of DNA Typing

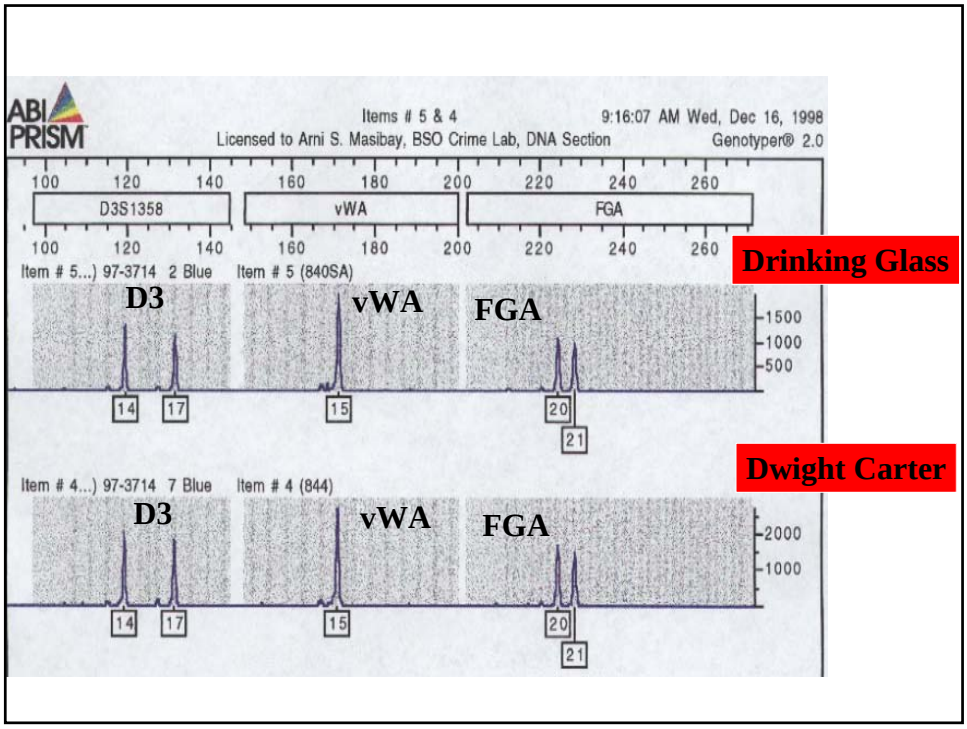
- **Medical**
  - **Diagnostics, treatment, disease prediction**
- **Identity Testing/Forensics**
  - **paternity, child support**
  - **patient sample/nursery mix-ups**
  - **immigration, inheritance**
  - **missing persons, unknown soldiers**
  - **victim ID (war, natural disaster)**
  - **crime investigations**
- **Population Genetics**
  - **human evolution/migration**
  - **genome diversity**

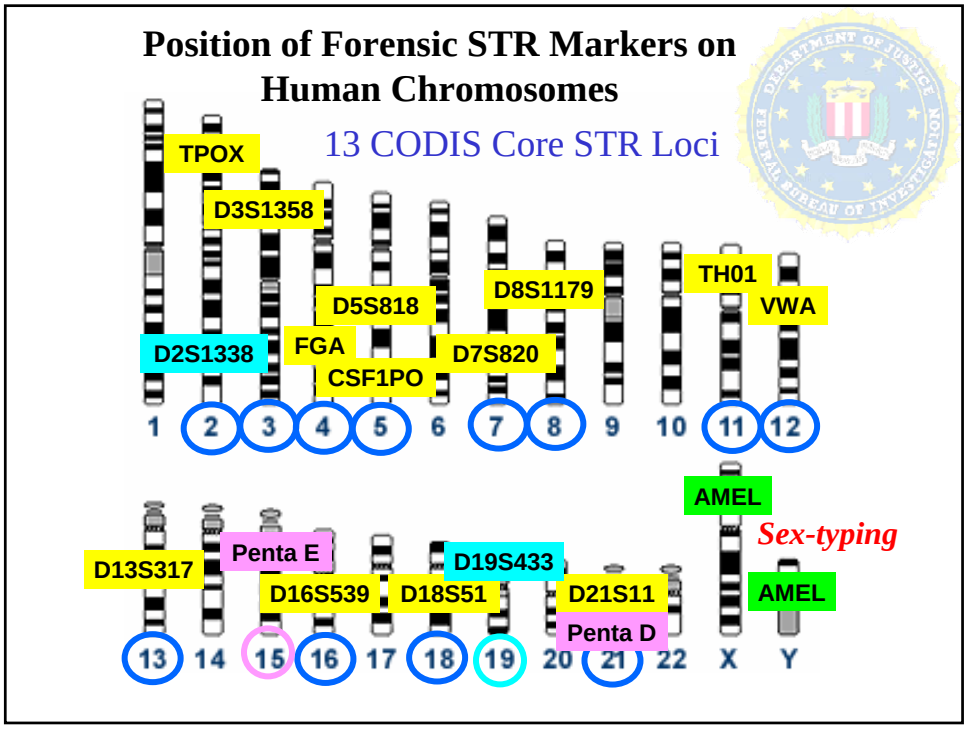
Human Identity Testing

- Forensic cases -- **matching suspect with evidence**
- Paternity testing -- **identifying father**
- Historical investigations
- Missing persons investigations
- Mass disasters -- **putting pieces back together**
- Military DNA “dog tag”
- Convicted felon DNA databases

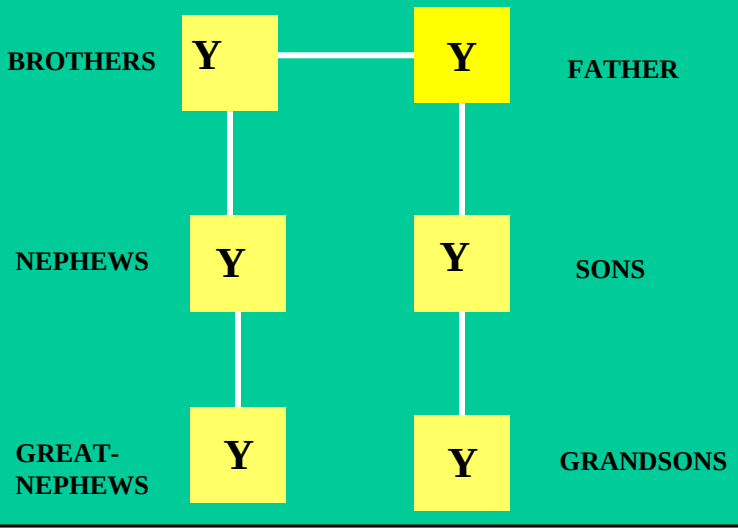
STR: Sequence Length Differences




ABI PRISM
Items # 5 & 4
9:16:07 AM Wed, Dec 16, 1998
Licensed to Arni S. Masibay, BSO Crime Lab, DNA Section
Genotyper® 2.0
D3S1358
vWA
FGA
Item # 5...) 97-3714 2 Blue
Item # 5 (840SA)
D3
vWA
FGA
Drinking Glass
Item # 4...) 97-3714 7 Blue
Item # 4 (844)
D3
vWA
FGA
Dwight Carter

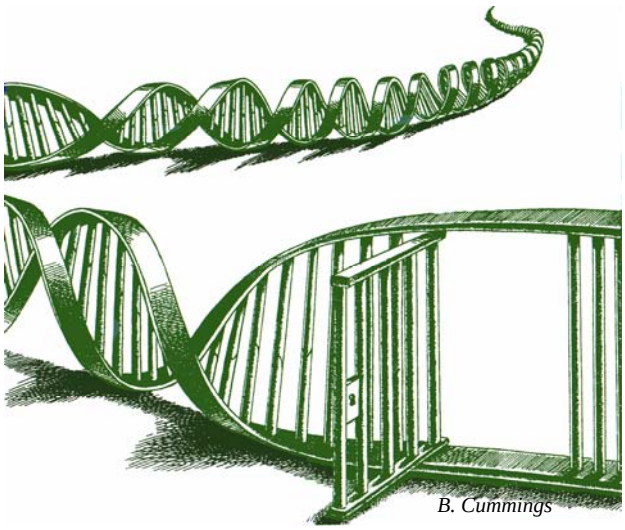

Position of Forensic STR Markers on Human Chromosomes
13 CODIS Core STR Loci
TPOX
D3S1358
D5S818
D8S1179
TH01
VWA
D2S1338
FGA
D7S820
CSF1PO
AMEL
Sex-typing
D13S317
Penta E
D19S433
D16S539
D18S51
D21S11
AMEL
Penta D

Y-chromosome



DNA Typing

Probative Evidence for Convictions and Exonerations



Published online 30 October 2009 | Nature | doi:10.1038/news.2009.1050

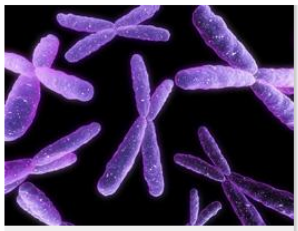
News

Lighter sentence for murderer with 'bad genes'

Italian court reduces jail term after tests identify genes linked to violent behaviour.

Emiliano Feresin

An Italian court has cut the sentence given to a convicted murderer by a year because he has genes linked to violent behaviour — the first time that behavioural genetics has affected a sentence passed by a European court. But researchers contacted by *Nature* have questioned whether the decision was based on sound science.



A court in Italy has cut a prisoner's jail term because he has genes associated with aggressive behaviour.

University Backs Away From New-Hire DNA Testing

University of Akron (Ohio)

Criminal background checks for university employees

(2) If the preferred candidate has lived in the state of Ohio for the past five consecutive years, a state of Ohio criminal background check will be conducted through the Ohio bureau of criminal identification and investigation ("BCI&I"). If the candidate has not lived in the state of Ohio for the past five consecutive years, a federal criminal background check will be conducted through the "BCI&I" website in addition to the Ohio criminal background check. Residence length will be verified during the employment reference checking procedure.

(3) Certain positions at the university of Akron, if required by law or contract, will be subject to both state of Ohio and federal criminal background checks regardless of how long the preferred candidate has resided in Ohio. Further, at discretion of the university of Akron, any applicant may be asked to submit fingerprints or DNA sample for purpose of a federal criminal background check. The

http://www.uakron.edu/ogc/docs/11-22_8-5-09.pdf

DNA Registries

Blood/tissue or DNA collected from:

- Offenders/arrestees
- Crime scene evidence

- Missing persons
- Mass disaster
- Medical
- Research
- Military/government

- Voluntary
  - Exclusion
  - Relatives of missing/lost/victims
- Genealogy

Forensic DNA Data Banks USA and Canada

- National DNA Index (NDIS)
- 7,434,897 convicted offender profiles
- 285,425 crime scene profiles
- 98,700 hits

- National DNA Data Bank of Canada
- 176,628 convicted offender profiles
- 52,285 crime scene profiles
- 13,324 hits

National DNA Data Bank Investigations Assisted

Offence	Total
Murder	897
Sexual assault	1,761
Attempted murder	332
Robbery (Armed)	1,551
Break and entering with intent, committing offence or breaking out	7,444
Assault (+)	940
Other	399
Total:	13,324

DNA-based Indirect Identifications

Genetic Kinship Analysis

- **Paternity Trios** (mother, child, tested man)
  - civil
    - child support, custody, immigration, estate
  - criminal
    - incest, rape
- **Mass Disasters/Missing persons**
  - victim identifications
  - family reunifications
  - military, national security
- **Forensic Investigations**

Tribute to the Unknown Soldier

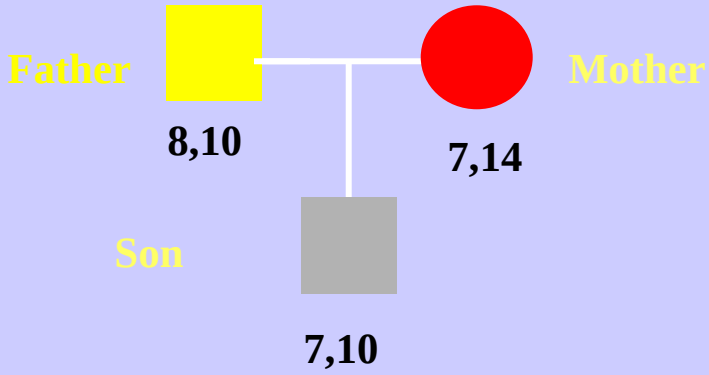
- We do not know his name, or his age, not his unit, or exactly when he died.
- We don't know his religion or what region of the country he came from.
- We DO know that he was Canadian.
- He is everyone's father, brother, husband, and son.
- He is our sense of pride and our sense of loss.
- He is every soldier who has ever fallen.

Missing Persons and Unknown Soldiers

- Military
  - POW/MIA, non-recovered remains
  - war crimes, crimes against humanity
- Civilian
  - lost/missing/runaway children and adults
  - natural disasters (e.g. flood, quakes, hurricane)
  - mass disasters (e.g. air crash, explosion)
  - criminal (e.g. kidnap, exploitation, terrorism)

Paternity

child receives one allele from each parent

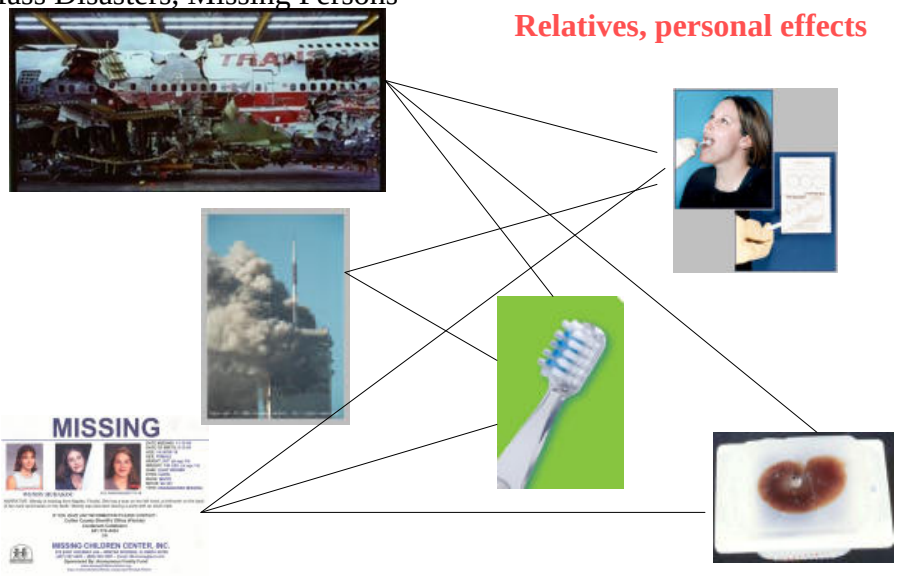


DNA Profile Comparisons

Direct and Indirect Identifications

Mass Disasters, Missing Persons

Relatives, personal effects



DNA Identifications After the 9/11 World Trade Center Attack

- **Victim Identifications** (9/11/05)
  - **1594**

  ~25% of all DNA IDs used **indirect kinship analysis**

  ~**850** IDs based solely on DNA

  - 20% of these solely from "mini-STRs"
  - 10 from SNPs alone
  - 10 from SNPs + STRs

See: Science 310:1122, 2005

Katrina Victim Identifications

Victim ID Center, Carville, LA

- 892 identified
  - Personal effects, pathology, fingerprints, dental records, field case notes, radiographic, photographic, anthropology
- **146 required DNA (almost all by indirect kinship analysis)**
- 152 RM
- 37 remains in mortuary (26 JUN 2006)

Data updated: 1600 hrs, 26 June 2006

Moving from Traditional Kinship Analysis to Possible Uses in Forensic Investigations

DNA Registries

Blood/tissue or DNA collected from:

- Offenders/arrestees
- Crime scene evidence

- Missing persons
- Mass disaster
- Medical
- Research
- Military/government

- Voluntary
  - Exclusion
  - Relatives of missing/lost/victims
  - Genealogy

- **Proxy profiles**
  - **MZ twins of those in registries**
  - **Close relatives**

"BTK" serial killer investigation
Forensic Reverse Parentage

Semen DNA
Evidence from crime scenes

???

DNA from suspect's daughter's biopsy

45 CFR 164.512 sIii A-C

Shaking the Family Tree

Searching for Suspects using Mendel's Laws

Can Genetic Kinship Analysis be applied to DNA Database profiles?

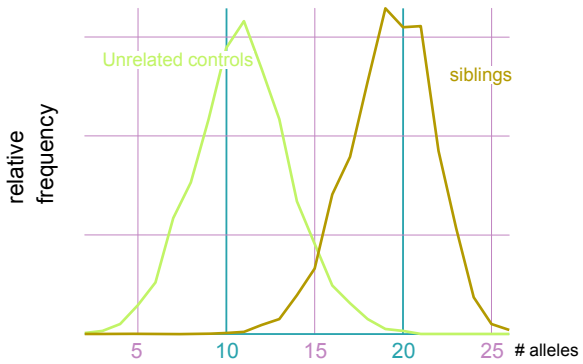
Mass Disasters		Crime Investigations
Human remains	are to	unidentified suspects
	as	
Volunteer relatives	are to	potential relatives in CODIS offender index

STR Allele Sharing in Sibs

Bourke, Ladd, Bieber et al/AAFS, 2/2000

	Full Sibs n=104	Unrelated Pairs n=112
• **Locus Identity**	**4.3**	**1.0**
– range	0-9	0-4
• **Alleles shared**	**16**	**8**
– range	11-23	3-13

Sibling screening – allele sharing



Sibling common allele distribution; 13 CODIS loci

Forensic DNA Analysis by "Data Mining"

- **Simple methods**
  - Partial profile/reduced stringency search
  - Allele sharing
  - Rare alleles
- **Sophisticated**
  - Kinship analysis

Murder Investigation into the murder of Lynette White

- Valentine's Day 1988
- Brutal stabbing of Lynette White
- Conviction of

  Cardiff 3
- Convictions later quashed

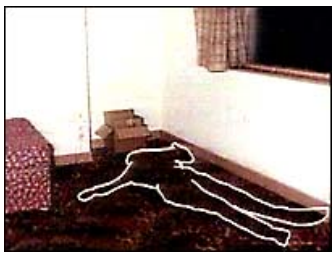
Police investigator Paul Williams

Cardiff, Wales
Renewed Investigation

1988 Murder of Lynette White

- Fresh DNA samples taken from crime scene
- Typed with 10 STR loci
- All 5 original defendants excluded
- Searched DNA database- **no match**
- Screened 300+ people in Cardiff area, **no match**

- **DNA expert Andrew MacDonald noted rare allele**



Rare or Common Allele Search

Find the Fly Yellow Ferrari

Jeffrey Gafoor

The Cellophane Man

- Narrowed search from 600 to 70
- One stood out, **a 14 yo boy**
- Uncle of 14 yo boy
- Excluded boy's father, but found complete profile match to uncle, Jeffrey Gafoor, who confessed to murder

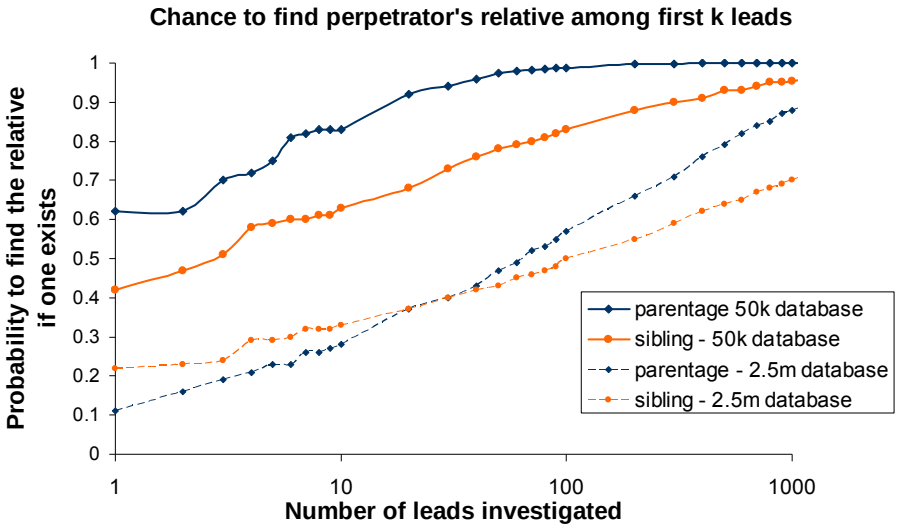
DNA Registries

Blood/tissue or DNA collected from:

- Offenders/arrestees
- Crime scene evidence
- Missing persons
- Mass disaster
- Medical
- Research
- Military/government
- Voluntary
  - Exclusion
  - Relatives of missing/lost/victims
- **Proxy profiles**
  - **MZ twins of those in registries**
  - **Close relatives**

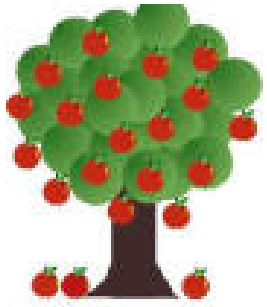
Chance of identifying close relative of offender in simulated database

Bieber, Brenner, Lazer
Science, 2006



Does the Apple Fall Far from the Tree?

- **Crime in Families**
  - Organized crime families
  - Neighborhood gangs
  - Dysfunctional family units



Family Background of Jail Inmates, 1996

Table 4.18. Family background of jail inmates, by sex and race/Hispanic origin, 1996

	Percent of jail inmates						
	Total	Male	Female	White non-Hispanic	Black non-Hispanic	Hispanic	Other
Family member ever incarcerated							
Any*	46.1%	44.7%	58.2%	46.7%	49.0%	37.1%	54.8%
Father	17.1	17.2	16.5	21.1	14.4	13.1	27.0
Mother	4.4	3.9	9.1	4.1	4.9	3.7	7.2
Brother	30.3	29.9	33.8	27.7	34.9	25.4	32.5
Sister	6.2	5.5	12.7	6.0	7.0	4.3	10.2
Spouse	3.3	2.1	13.6	4.9	1.7	2.3	9.5
Child	1.3	1.1	3.2	1.5	1.2	0.9	3.0

Source: Correctional Populations in the United States, 1996
BJS, U.S. DOJ

Home > Denver & the West

Denver uses "familial DNA evidence" to solve car break-ins

BOOKMARK
PRINT EMAIL
4 COMMENTS

By Howard Pankratz
The Denver Post

POSTED: 11/16/2009 02:17:23 PM MST
UPDATED: 11/16/2009 02:19:28 PM MST

A man who broke into two cars in Denver in February 2008 has been identified and convicted using a method of DNA identification called "familial DNA evidence," Denver police and prosecutors said today.

Lynn Kimbrough, spokeswoman for Denver district attorney Mitch Morrissey, said it is the first successful use of familial DNA evidence to identify and convict a person in Denver.

The suspect, Luis Jaimes-Tinajero, 21, was linked to two car break-ins that occurred on Feb. 20, 2008 and Feb. 21, 2008,.

When he broke the car windows, Jaimes-Tinajero cut himself and left blood at the scene, said Kimbrough.

The blood left at the scene was used to develop a DNA profile that was submitted to Denver's local DNA database.

The sample was run through a familial search software program designed by the Denver DA's office and the Denver police department crime lab.

It was a close match to another DNA profile already in the database, which indicated the blood may have been left behind by a close family member, most likely a sibling, said Kimbrough.

Family Searching

Science, Law, Ethics

- **Opportunities**
  - Solve crimes
  - Enhance public safety
  - Exonerate innocent
- **Challenges**
  - 4th amendment issues?
    - Probable cause?
  - reaction by media, legislature
    - could have unfortunate implications for missing persons and mass disaster programs

Family Searching

Intersection of Science, Law, Ethics

- **Pros**
  - demand for public safety outweighs privacy interests?
  - makes full use of DNA data (and samples) in ongoing basis
- **Cons**
  - 4th amendment?
  - targeting subset of population (relatives of convicted/arrested) for life?
  - Could intrude on privacy of those adventitiously identified (false +s)

Common Uses of Partial Information in Criminal Justice Efforts

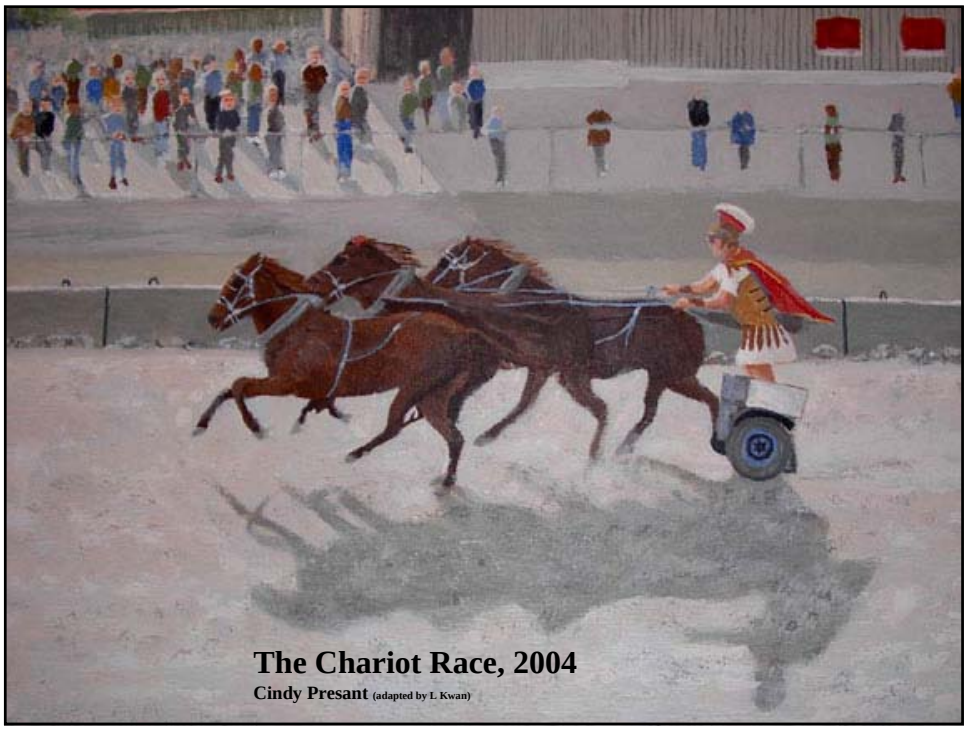
- Partial license plates
- Partial fingerprints
- Credit card traces to family members
- Phone records
- Email, blogs, on-line social networks

Summary

- **Male kin are in CODIS "by proxy"**
  - expands database 20-40% or more
- **Shaking the Family Tree technically feasible now**
  - Rare allele searches
  - Allele sharing comparisons
  - Kinship analyses using likelihood calculations
    - Y-chromosome typing would eliminate most false leads
- CODIS MP database software upgrades
- Database managers
- Public perception

Science, Law, Ethics, Policy, Privacy

- Rights of Victims
- Rights of Accused
- Rights of Citizens
- Safeguards for thoughtful use of DNA technology



The Chariot Race, 2004
Cindy Presant (adapted by L Kwan)

Anonymous Fuzzy String Comparisons in Healthcare Record Linkage Applications

Stanley Trepetin, New York City Department of Health and Mental Hygiene

Abstact:

Medical privacy continues to be a key issue as policy research continues to show people's demand for health organizations to protect patients' personal data. Health organizations need personally identifiable data for unhampered decision making; yet, identifiable data are often the basis of information abuse if such data are improperly transmitted, stored, or disposed. This talk shows how health organizations may use de-identified data for some strategic organizational operations.

Mr. Trepetin will demonstrate a new idea for anonymous record linkage. For a variety of health applications there is a need to perform linkage among data set records to connect data about the same individual or event so that further analysis becomes possible. However, the privacy of the individuals in the records must also be better protected. He will show how linkage can be effectively performed based not on the actual data but on an anonymous form of the data, without diminishing the ability to link records whose identifiers are only "close" to each other, not equal, because of typical recording errors. Mr. Trepetin will show how to embed additional information from a record into the anonymization process to increase the security and error-handling during string comparisons.

Finally, Mr. Trepetin will discuss how the proposed technique was tested on a real record linkage platform, IBM's QualityStage, with real person-level data. The matching results were essentially the same when compared to matching results using personally identifiable data.

Bio:

Stanley Trepetin is the Chief Information Security Officer at the New York City Department of Health and Mental Hygiene (DOHMH). At DOHMH he sets organizational IT security strategy and policy. Stanley completed his PhD at MIT in Health Informatics in 2006. At MIT, he designed new ways to anonymously match data and quantify the benefit of implementing information privacy within health organizations. Prior to MIT he worked for IBM for 10 years where he provided large systems software support to Fortune 500 clients and was a software developer and project manager. He has a Master's Degree from Duke University focusing on patent usage within biotechnology and an undergraduate degree from Cornell in computer science and mathematics.

Anonymous fuzzy string comparisons in healthcare record linkage applications

Stan Trepetin, PhD
Chief Information Security Officer
New York City Department of Health and Mental Hygiene

Outline

- Usage of string comparisons within healthcare record linkage applications
- Need for privacy
- Existing approaches trying to provide secure string comparisons
- A new approach
- Test of new approach within an actual public health environment

Healthcare Record Linkage

- Healthcare organizations increasingly rely on record linkage.
- Record linkage:
  - One or more unique or non-unique identifiers may be used.
  - String similarity computations are useful to catch erroneous insertions, deletions, etc.

Privacy requirement

- Privacy is important, too:
  - Surveys, recent HIPAA changes, etc.
- Question: can string comparison computations be done in a privacy-preserving manner?

Other Approaches

- Access control: data behind an access control, inaccessible to analyst.
- Problem: difficult to control the dynamic nature of access control (changing employee roles, backups, etc).

Other Approaches

- [Song et al., 2000] suggests encrypting all possible errors within an identifier and comparing the resulting lists during record linkage.
- Problem: matching potential declines because one cannot easily identify all identifier errors

Other Approaches

- [Pang et al., 2006] suggests submitting encrypted reference strings and their distances to a third party to determine which distances are below the threshold.
- Problem: Generic distance computations are often inappropriate when assessment of particular character positions is important.

New approach

- Use padding from the same record to encrypt a field to be character analyzed
  - Padding must generally be unique for each individual in linkage file
- Use a deterministic scheme like AES (ECB mode), triple DES, etc. for the encryption

Example of anonymous comparison

Before privacy enhancement

Rec. First_ name	Last_name
1538 Jim	Smith
3294 J***e***m	Smith

After privacy enhancement

Rec. First_ name	Last_name
1538 hd: E(J|Smith)->E(i|Smith)->E(m|Smith)	Smith
3294 hd: E(J|Smith)->***E(e|Smith)***->E(m|Smith)	Smith

Solution security

- Frequency analysis would not work.

Quality Stage

- The DOHMH wanted to use this technology.
- I ran a proof-of-concept.
- Used IBM's QualityStage, version 7.5.
- A live public health data set that DOHMH had access to: half million records; over 45 columns.

Quality Stage

- Protocol already existed to find duplicate records:
  - Data uploaded into QS from MS SQL 2005.
  - Multi-pass protocol (each pass had its own blocking and matching variables)
  - Four identifiers (SSN, last name, birthdate, zip code) were compared using string comparison functions.

Quality Stage

- I converted the four character-analyzed variables into enciphered forms.
- Impact:
  - The protocol had to be changed somewhat to account for the new weights.
  - The field width increased because each character position was now a full enciphered value instead of one byte.

Quality Stage

- Result: the Quality Stage produced results containing 99% of the same number of duplicates as running the protocol over identifiable data.
- Conclusion: this protocol provides security and can be applied in real record linkage settings.

Q & A

Session 1B: De-identification of Genomic Data

Session Chair: Bradley Malin, Vanderbilt University

Bio:

Bradley Malin is an Assistant Professor of Biomedical Informatics in the School of Medicine and an Assistant Professor of Computer Science in the School of Engineering at Vanderbilt University. He is the founder and director of the Vanderbilt Health Information Privacy Laboratory (HIPLab), which focuses on basic and applied research in a number of health-related areas, including primary care and secondary sharing of patient-specific clinical and genomic data. His research has received several awards of distinction from the American and International Medical Informatics Associations. For the past several years, he has directed a data privacy research and consultation team for the Electronic Medical Records and Genomics (eMERGE) project, a consortium sponsored by the U.S. National Human Genome Research Institute. He has served as a program committee member and workshop chair for numerous research conferences and has edited several volumes for Springer Lecture Notes in Computer Science, a special issue for the journal Data and Knowledge Engineering, and is currently on the editorial board of the journal Transactions on Data Privacy. He received a Bachelor's in biology (2000), Master's in knowledge discovery and data mining (2002), Master's in public policy & management (2003), and a Doctorate in computation, organizations & society (2006) from the School of Computer Science at Carnegie Mellon University.

Privacy-Preserving Storage and Querying of Genomic Data

Murat Kantarcioglu, University of Texas at Dallas

Abstract:

In this talk, we present a novel cryptographic framework that enables organizations to support genomic data mining without disclosing the raw genomic sequences. Organizations contribute encrypted genomic sequence records into a centralized repository, where the administrator can perform queries, such as frequency counts, without decrypting the data. We discuss the evaluation results of our framework with existing databases of single nucleotide polymorphism (SNP) sequences and demonstrate that the time needed to complete count queries is feasible for real world applications. We further show that approximation strategies can be applied to significantly speed up query execution times with minimal loss in accuracy. The framework that is presented can be implemented on top of existing information and network technologies in biomedical environments.

Bio:

Dr. Murat Kantarcioglu is currently an assistant professor of computer science at University of Texas at Dallas. He had a Ph.D. degree from Purdue University in 2005. He received his master's in Computer Science from Purdue University in 2002 and his bachelor degree in computer engineering from METU, Ankara, Turkey in 2000. He is also a recipient of NSF CAREER Award. His research interests lie at the intersection of Privacy, Security, Data Mining and Databases: Security and Privacy issues raised by data mining; Distributed Data Mining techniques; Security issues in Databases; Privacy issues in health care. His current research is funded by grants from NIH, NSF, AFOSR, ONR and IARPA.

Link to video of this presentation.

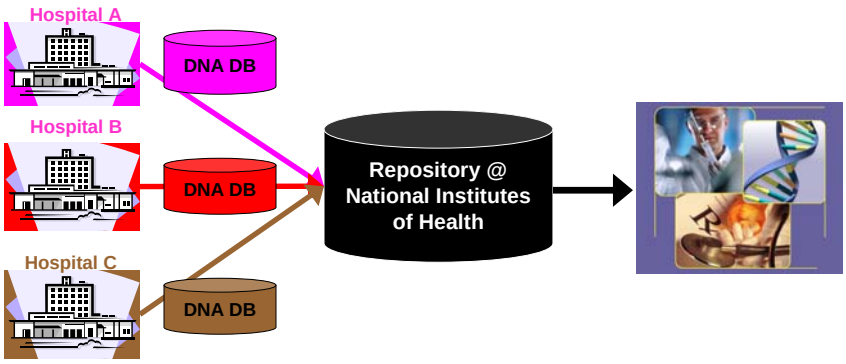
A Privacy Preserving Framework for Integrating, Storing and Querying Biological Data

Murat Kantarcioglu, Ph.D.
University of Texas at Dallas

Joint work with Brad Malin and Wei Jiang

Example

- **Goal:** Construct repositories of person-specific DNA for pharmacogenetic and biomedical research



- **Challenge:** Need to merge, store, query records securely without violating privacy

Privacy Enhancing Technologies

- **Cryptographic techniques**
  - Symmetric key systems
  - Public key systems
  - Homomorphic Encryption
    - Certain operations on the encrypted data sets are possible using Homomorphic encryption
  - Id-based encryption
    - Any string (bob@company.com) could be public key in Id-based encryption
  - Cryptographic Hardware
    - Encryption keys can be stored securely.

Privacy Enhancing Technologies

- **Anonymization techniques**
  - Non-individually identifiable data sets with formal privacy guarantees could be released using techniques such as k-anonymity.
- **Data analysis & Digital Forensics techniques**
  - Audit logs
  - Online auditing tools
  - Digital Forensics techniques

Homomorphic Encryption

- Can compute the encrypted sum of two messages via their ciphertexts
- Three properties of interest for computation
  - Property 1: Efficient Multiplication by Constant
  - Property 2: Probabilistic Encryption
  - Property 3: Additively Homomorphic

Homomorphic Encryption

- **Property 1:** Multiplication by a constant

$$E(km) := k \times_h E(m)$$

which yields

$$D(E(km)) = km$$

Homomorphic Encryption

- **Property 2**: Probabilistic Encryption
- Given m, the following occurs with ↑↑↑ probability

$$c_1 = E(m) \qquad c_2 = E(m)$$

$$D(c_1) = D(c_2)$$

$$c_1 \neq c_2$$

Homomorphic Encryption

- **Property 3:** Additively Homomorphic
  - Given cyphertexts $E(m_1)$ and $E(m_2)$
  - There exist efficient algorithm to compute

$$E(m_1+m_2) := E(m_1) +_h E(m_2)$$

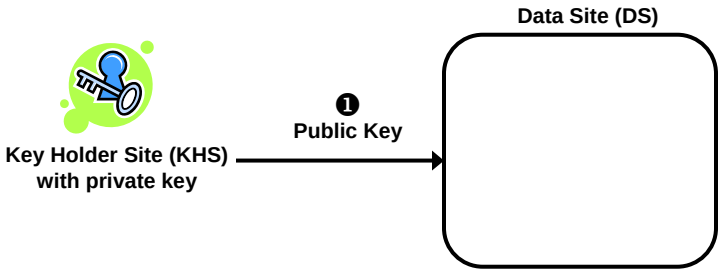
which yields

$$D(E(m_1+m_2)) = m_1+m_2$$

Example: Secure Record Management

(Kantarcioglu, Jiang, Liu, Malin, IEEE TITB 2008)

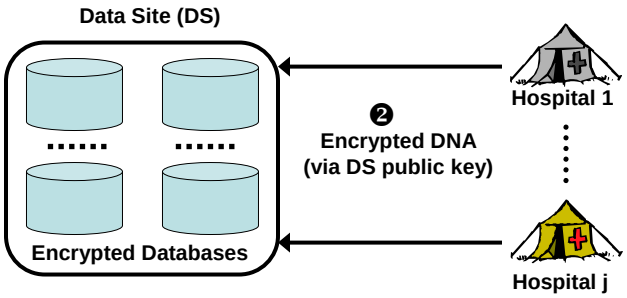
- Data Providers
- Third Party Data Managers ← Required
- Data Users



Key Generation: <public, private>

Architecture

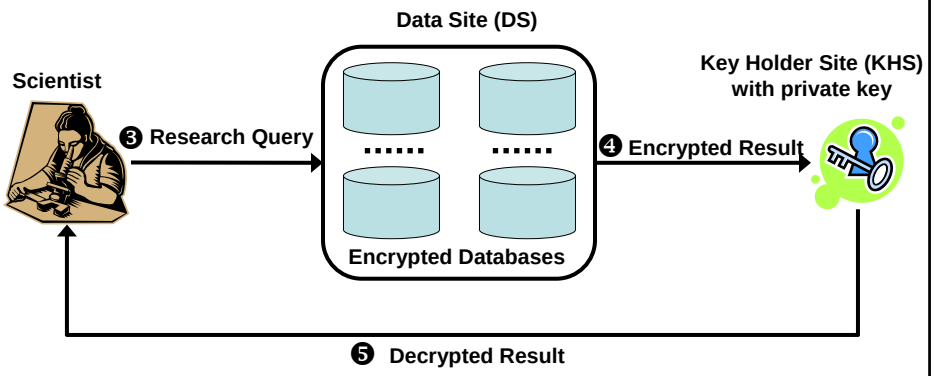
(Kantarcioglu, Jiang, Liu, Malin, IEEE TITB 2008)



Data Encryption

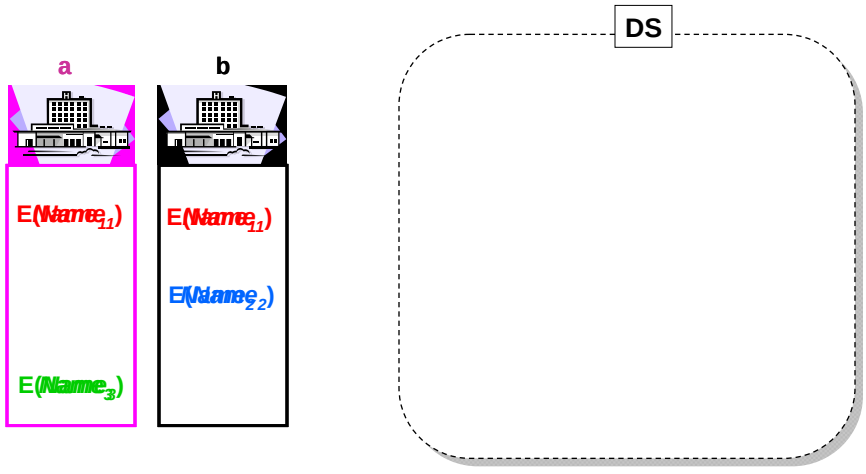
Architecture

(Kantarcioglu, Jiang, Liu, Malin, IEEE TITB 2008)



Query Issuance | Result Decryption | Query Processing

Secure Data Integration

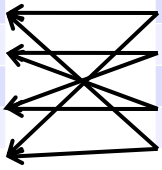


Leveraging *k*-anonymity for secure data integration

- Relax semantic security provided by encryption techniques in a controlled manner
- *k*-anonymize records before sharing them
  - Each record is equivalent to at least *k*-1 other records over the combination of identifying attributes

Example

Age	Sex	Zip
30	M	15213
33	M	15217
33	F	15213
30	M	15213



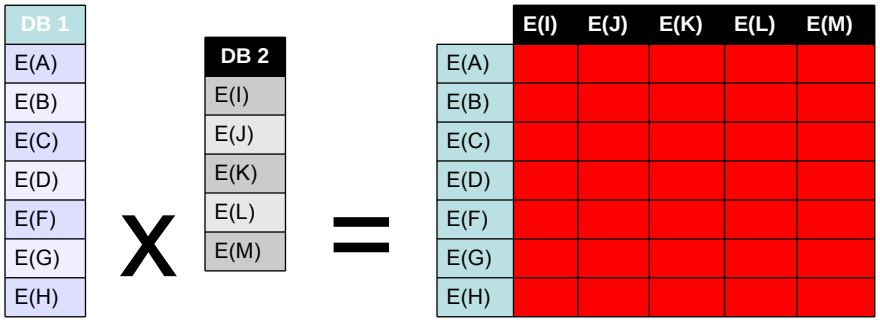
Age	Sex	Zip
30	M	15213
33	*	1521*
33	*	1521*
30	M	15213

Private Records

2-Anonymous

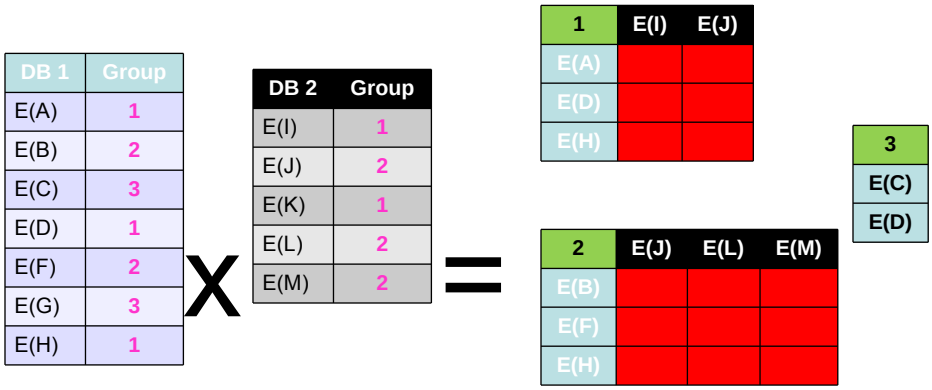
Leveraging *k*-anonymity

- *k*-anonymize records before sharing them
- *k*-anonymized demographics → hash keys



Leveraging *k*-anonymity

- *k*-anonymize records before sharing them
- *k*-anonymized demographics → hash keys



Other Issues

- **Other capabilities of our system**
  - Privacy-preserving audits
- **Current limitations**
  - Slower compared to non-secure versions.
  - More storage is needed.

Conclusions

- Existing technologies enable privacy-preserving biological data integration, storage and querying.
- The trade-off is between cost versus privacy
  - Almost any task could be achieved without violating privacy.
- With more research, all the potential benefits of "biological data" could be unlocked at a reasonable cost without violating individual privacy.

References

- For the details of our proposed system , please see the following references. (**Joint work** with **Wei Jiang** (Purdue University), and **Bradley Malin**, (Vanderbilt University))
  - Murat Kantarcioglu, Wei Jiang, Ying Liu, and Bradley Malin, "**A Cryptographic Approach to Securely Share and Query Genomic Sequences**", IEEE Transactions on Information Technology in Biomedicine, Vol. 12, No. 5, pp 606-617 (2008)
  - Murat Kantarcioglu, Wei Jiang, and Bradley Malin, "**A Privacy-Preserving Framework for Integrating Person-Specific Databases** ", Privacy in Statistical Databases, 2008, LNCS 5262, pp. 298–314, 2008.

Privacy Implications for Rare Mutation Data

Christopher Cassa, Harvard Medical School and MIT

Abstract:

Sequencing of an individual's DNA may reveal single nucleotide variants that have not been documented or previously identified. These variants include nonsense and missense mutations, insertions or deletions, and other lesions. Presence of such mutation data in a shared or published sequence substantially increases the ability to identify the individual whose data are shared. In the case of a de novo germline mutation, I will discuss the privacy implications for carrying a specific mutation.

I will first explore general identifiability issues for mutant loci, and how likely a match would be among 1000 people. If a mutation is not de novo, I will show that it is necessary to adjust estimates using the effective population size and prevalence in the population.

Bio:

Christopher Cassa, Ph.D., a graduate of the Harvard-MIT Division of Health Sciences and Technology, is a research fellow at the Children's Hospital Informatics Program at Harvard Medical School in Boston, MA. He has researched a wide range of medical privacy and identifiability issues. Applying quantitative approaches, he has helped develop two anonymization techniques for geographical data and investigated the re-identification potential of geographical data shared in textual and map form. His most recent work has investigated the ability to infer genotypes from family members of research proband, and how readily research datasets can be used to identify family members and familial phenotypes.

Link to video of this presentation.

Privacy Implications for Rare Mutation Data

Christopher Cassa, PhD

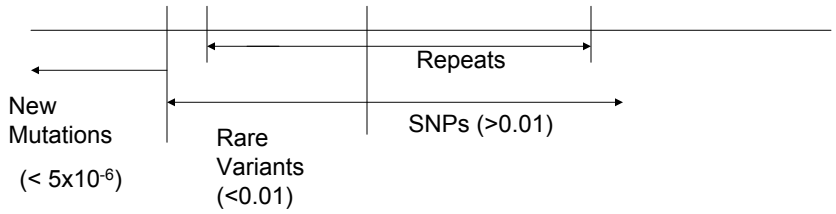
Children's Hospital Informatics Program
Harvard-MIT Division of Health Sciences and Technology

Growth of Mutation Data

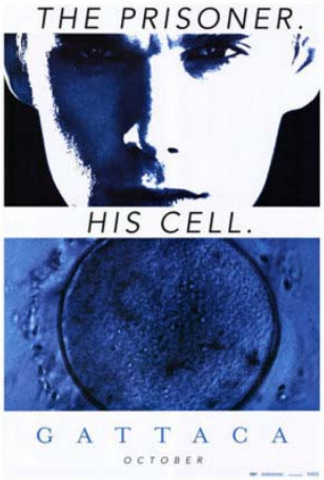
- Research studies sharing sequence and expression data for other investigators
- Public Studies:
  - Growth of GWAS Studies
- Growth of deep sequencing and small resequencing technologies
- Direct-to-consumer genetic screenings

23andMe Navigenics deCODE ME Knome

Population Frequencies of Variants



Broad Fear of DNA Use in Society



What Protections are in Place?

- Genetic Information Non-Discrimination Act
  - Passed in 2008, just came into effect
- State Laws Protecting Similar Items
- Data use agreements protecting against attempts at re-identification

Genomic Data Pose Unique Risks

- Discrimination Concerns
  - Insurance, workplace discrimination
  - Life, disability, and long term care insurance uncovered
- Genetic Knowledge and Personal Decision Making
- Implications for Family Members
- May Carry Surname

Why Risk GATTACA?

- Correlate clinical outcomes with genomic data
- Individual participation necessary – sharing genotypic and clinical data with investigators
- Methods to help individuals with risk assessment and to preserve privacy with such disclosures needed

Balance between Privacy and Data Use

- Pervasive in research, medicine, and public health investigations, posing risk to privacy
- Disclose identity, medical conditions, and hereditary data

Balance between privacy and research and public health

Collaborators

- The work was supported by the National Library of Medicine, National Institutes of Health

Kenneth Mandl MD MPH (CHIP)

Peter Szolovits PhD (MIT)

Topics

1. Estimating the prevalence of a previously unknown mutation
2. How likely is it that a two individuals will match at an observed mutation?
3. How do these estimates change in different populations?

Example:

- You are given a specific single nucleotide variant found in an individual. You are able to quickly verify that this is not a variant previously documented in the human genome, dbSNP or other project.
- Is it possible to estimate how well one could identify that individual?
- How would these results differ if the individual is of European descent or of African descent?

Various Approaches

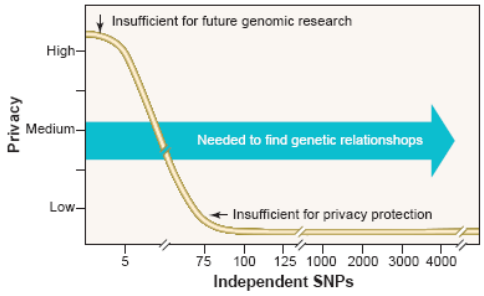
- Is this a de novo germline mutation? In the case of a de novo mutation, we could evaluate how identifiable a person is with that mutation (and how likely a match would be among 1000 people) directly. **[Covered First]**
- If this mutation is not de novo, we will need to adjust our estimate with population genetics adjustments using population size and estimates of prevalence in the population. **[Covered Last]**

General Assumptions:

- This is a de novo germline mutation that is not distributed widely in the population.
- Region-specific mutation frequency in the genome is even enough to use the mean frequency values.
- Considering an autosomal mutation
- We will be gender-neural in our analysis as there is gender mutation bias found in men passing mutations (due to the increased number of mitotic events and risk of transmission of new mutations).

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=10978293

Altman: Privacy Decreases Sharply with a Small Set of Independent SNPs



Trade-offs between SNPs and privacy.

Genomic Research and Human Subject Privacy

- At a low number (35-70) of identified **independent** SNPs, the amount of privacy dramatically decreases.
- Two random people in a population matching

Zhen Lin, *Art B Owen*, *Russ B Altman*. **Science**. Washington: Jul 9, 2004.Vol.**305**, Iss. **5681**; pg. **183**

Example Roadmap

- Step 1: Treat a mutation as a rare allele with frequency ***p***
- Step 2: Calculate ***p***, the allele frequency of that mutant in specific population, and $\pi_{A\alpha}$ from p(Aa)
- Step 3: Calculate the probability of a match of that mutated base pair in a second person, μ_i
- Step 4: Calculate the probability that those two people are the same given a match observed with probability μ_i

What is the Probability of Observing this Specific Mutation?

- Let pop. freq. of mutant variant at locus *i* be p_i:

$$p_i = (r_{region,type})P_{sub-type}$$

- $r_{region,type}$ is the region-specific, type specific mutation rate per base pair, per generation
  - Example: *Transition mutation rate in a CpG locus*
- $P_{sub\text{-}type}$= probability of the specific sub-type of the mutation class (normalized for type)
  - Example: *A→G mutation of all transition mutations*

Scarano, E., Iaccarino, M., Grippo, P., Pirisi, E. 1967. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. Proc. Natl. Acad. Sci. U. S. A. 57, 1394-140

Children's Hospital Boston
Informatics Program
Harvard-MIT Division of
Health Sciences and Technology

Evaluating Data Sources

- We start to evaluate the previous match estimate by obtaining population-specific values for
  - $r_{region,type}$ mutation rate estimates
  - $P_{sub\text{-}type:}$ Using data from the Cardiff Human Gene Mutation Database, we can calculate $p_{sub\text{-}type}$, which must be normalized among all main type mutations

Region-Specific Mutation Rates ($r_{region,type}$)

- $r_{region,\ type}$ estimates of mutation rate for different sites and different classes of mutation

Mutation type	Mutation rate
Transition at CpG	1.6 x 10-7
Transversion at CpG	4.4 x 10-8
Transition at non-CpG	1.2 x 10-8
Transversion at non-CpG	5.5 x 10-9
All nucleotide substitutions	2.3 x 10-8
Length mutations	2.3 x 10-9
All mutations	2.5 x 10-8

- Rates calculated on the basis of a divergence time of 5 mya, ancestral population size of 10^4, generation length of 20 yrs, and rates of molecular evolution.

http://www.genetics.org/cgi/content/full/156/1/297/T4

Children's Hospital Boston Informatics Program
Harvard-MIT Division of Health Sciences and Technology

HGMD Statistics for Missense Mutations ($P_{sub\text{-}type}$)

Wild type	G	T	A	C	Total
Guanine	--	2228	7140	2290	11658
Thymine	1481	--	1045	3609	6135
Adenine	2947	734	--	1048	4839
Cytosine	1619	4785	1376	--	7780

HGMD Statistics for Nonsense Mutations ($P_{sub\text{-}type}$)

Wild type	G	T	A	C	Total
Guanine	--	1009	1028	0	2037
Thymine	224	--	325	0	549
Adenine	0	273	--	0	339
Cytosine	499	3178	727	--	4817

All Transition Missense Mutations ($P_{sub\text{-}type}$)

Wild type	G	T	A	C	Total
Guanine	--	***	7140	***	7140
Thymine	***	--	***	3609	3609
Adenine	2947	***	--	***	2947
Cytosine	***	4785	***	--	4785

http://www.hgmd.cf.ac.uk/ac/hoho1.php

Example Part II:

- What is the likelihood that you could pick out that individual out of a group of 1000 genotyped at the same locus.

Probability of finding a match in rare mutation alleles

- Technically, the probability that two people carry the same allele must include the possibility that either of them is homozygotic minor or heterozygotic at the locus.
  - For de novo mutations, this is not likely
  - Not important for very small frequencies
- The chance of two unrelated people with mutations matching at any locus, μ_i, is the frequency of heterozygotes, π_i, from $2p_i(1\text{-}p_i)$

$$\mu_i = (q_i^2)^2 + (2 p_i q_i)^2$$

$$for \quad small \quad \pi_i, \ \mu_i \approx \pi_i \approx 4(p_i)^2$$

P(same|match at M bases)

- We can subsequently evaluate the posterior probability of a match at *M* mutant loci using Bayes' Theorem.
  - Caveat: all M mutant loci must be independent
- Suppose the adversary assumes a conservative prior model that research subjects are uniformly sampled from a population of *N* people.
- The probability that a person is subject *i*, given that they share a set of *M* mutations can be directly calculated:

$$p(same \mid match) = \frac{p(match \mid same)\, p(same)}{p(match \mid same)\, p(same) + p(match \mid !same)\, p(!same)}$$

Example

- How identifiable is a Caucasian male with a missense A→G mutation in a CpG locus among 1000 others?
- We have region-specific information here, since this is a CpG locus, and we also know it's a transition mutation, since it is a purine/purine mutation.
  - $p_i = (p_{sub\text{-}type})(r_{region,type})$
  - $p_i = (0.386)(1.6 \times 10^{-7})$
  - $p_i = \mathbf{6.2 \times 10^{-8}}$

Example (Continued)

$\pi_i = \mathbf{(6.2 \times 10^{-8})(1\text{-}6.2 \times 10^{-8})}$

$\mu_i = \mathbf{(1.54 \times 10^{-14})}$ + very small terms

$$p(same \mid match)$$
$$= \frac{p(match \mid same)\,p(same)}{p(match \mid same)\,p(same) + p(match \mid !same)\,p(!same)}$$
$$= \frac{(\frac{1}{1E3})}{(\frac{1}{1E3}) + (1.54E-14)(1-\frac{1}{1E3})}$$
$$= 0.99993 \approx 1$$

Example (Continued)

- If the probability of one person matching is μ_i = ***1.54 × 10^{-14}***, the likelihood of a identifying a match in 1000 people is an example of the birthday problem:
  - p(n)=1-p!(n), approximated by $1-e^{-(n(n-1))/2(1.54E-16)}$
  - p(1000)=0.000000325; 14,599,883 people required for 50% chance of a match
- This answer should be close to the answer in an African population, as long as neutral assumptions hold because the localized mutation rate should not be different among populations

At least one match?

- It's possible to determine how likely one match or fewer will occur, or the probability of having more than one match in a set of N individuals using the Binomial Distribution

Binomial Distribution

n=N people, p= μ_i (p(Aa)), k=1
E(X) = (n)(p) = 1000 μ_i = 6.2E-5
Var(X) = np(1-p) = 1000 $\mu_i(1-\mu_i)$ = 6.2E-5

$$f(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$F(x;n,p) = \Pr(X \le x) = \sum_{j=0}^{\mathrm{Floor}(x)} \binom{n}{j} p^j (1-p)^{n-j}$$

$$= \sum_{j=0}^{1} \binom{1000}{j} (\pi_i)^j (1-\pi_i)^{1000-j}$$

p(1 or fewer in 1000) = 0.999999998

Example Part III:

- How would this estimate differ in an African population versus a European population?

Possible Approaches

- Use population-specific, region-specific mutation frequency databases to calculate the probabilities of mutations [+ simpler; - not many thorough databases]
- Alter estimates for different populations taking into consideration their effective population size and local population-specific heterozygosity rate

Population-Specific Mutation DBs

Database	Address	Curators
Arab Disease Mutation Database	http://www.ctga.org.ae	
Cypriot Disease National Database	http://www.goldenhelix.org/cypriot (Encouraged by HGVS)	George P. Patrinos, Erasmus University Medical Center Rotterdam, The Netherlands Marina Kleanthous, Anthi Drousiotou, Roula Christodoulou, Philippos Patsalis and Marios Kariolou Dept Molecular Genetics, Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus Costantinos Deltas Dept of Biological Sciences, University of Cyprus, Nicosia, Cyprus
Finnish Disease Heritage	http://www.findis.org (Encouraged by HUGO-MDI)	Pertti Aula Dept. of Medical Genetics, University of Helsinki, Dept. of Molecular Medicine, National Public Health Institute, Biomedicum Helsinki, Finland
Hellenic Disease National Database	http://www.goldenhelix.org/hellenic (Encouraged by HGVS)	George P. Patrinos, Erasmus University Medical Center Rotterdam, The Netherlands Manos Papadakis, Center for Thalassemia Athens, Greece
Iranian Human Mutation Gene Bank	http://www.IHMGB.com (Encouraged by HUGO-MDI)	Hossein Najmabadi, Maryam Neishabury, Farhad Sahebjam, Kimia Kahrizi, Yousef Shafaghati, Nushin Nikzat, Marya Jalalvand, Farahnaz Aminy, Susan Bany Hashemi, Babak Moghimi, Ali Reza Noorian, Ali Jannati, Mehrdad Mohamm Khalil Javan Genetics Research Centre, Univ. of Welfare Sciences & Rehabilitation, Tehran, Iran
Israeli Populations	http://www.tau.ac.il/medicine/NLGIP/nlgip.htm	Nat. Lab. Genetics of. Israeli Populations Tel Aviv Univ.
The Lebanese National Mutation Frequency Database	http://www.goldenhelix.org/lebanese	Prof. Andre Megarbane, Dr. Eliane Chouery, Saint Joseph University, Faculty of Medicine, Unit of Medical Genetic Beirut, Lebanon George P. Patrinos, Erasmus University Medical Center Rotterdam, The Netherlands
Korean Disease Mutation Database	Under construction	to be added soon
Chinese Disease Mutation Database	Under construction	to be added soon
Singapore Human Mutation and Polymorphism Database	http://shmpd.bii.a-star.edu.sg/	Marie Loh, Bioinformatics Institute, National University of Singapore Ene-choo Tan, Defence Medical and Environ Research Institute, Singapore
Turkish Disease	http://bioserver.bio.boun.edu.tr	A. N. Basak, S. H. Caglayan , M. U. Caglayan

Population Differences

- In an African population with greater local heterozygosity and greater effective population size one might expect:
  - Older (mean) mutation that is more prevalent
  - Easier to find a match with one person if population frequency is higher
  - Harder to randomly match two people at a set of polymorphic SNPs or repeats due to heterozygosity
  - Less susceptible to drift variance, but less likely to be fixed (Kimura), with 1/N probability of fixing any allele

Conclusion

- Rare mutations are highly identifying, while commonly varying markers range over a large set of population frequencies so matching a rare variant is much more informative than matching a common one.
- It is unlikely that two people share a new rare mutant allele in a sample of 1000 in any population
- It may not be possible to preserve privacy in a dataset containing rare mutations

Surveying the Landscape of Privacy in Clinical Genomics Research Databases

Bradley Malin, Vanderbilt University

Abstract:

The increasing adoption of electronic medical record systems into healthcare, combined with decreasing costs of high-throughput and storage systems, has enabled the collection of detailed person-specific clinical and genomic data. Scientists can now data mine for relationships between complex disorders and genomic features, as well as environmental factors, but need to share records across institutional boarders to strengthen statistical power in complex association studies, allow verification of findings, and comply with a host of regulations. To support a data sharing culture and prevent stagnancy in biomedical research, it is crucial that organizations protect the anonymity and confidentiality of the corresponding research participants. In this talk, I will review various real-world policies and technologies that various organizations have developed to protect research participants in such environments. I will further review the extent to which such systems are resistant to emerging adversarial threats in the context of varying amounts of an adversary's background knowledge. This talk will conclude with a discussion of recent research developments and challenges for data protection in emerging clinical genomics research databases.

Bio:

See Session Chair Bio, page 125.

Link to video of this presentation.

Surveying the Landscape of Privacy in Clinical Genomics Research Databases

Bradley Malin, Ph.D.
Assistant Prof. of Biomedical Informatics, School of Medicine
Assistant Prof. of Computer Science, School of Engineering
Vanderbilt University
November 19, 2009

TUESDAY, JUNE 20, 2006

THE TENNESSEAN

171 6 SECTIONS 3

VU to put patient DNA in vast research pool

Blood samples included unless people opt out

By CLAUDIA PINTO
Staff Writer

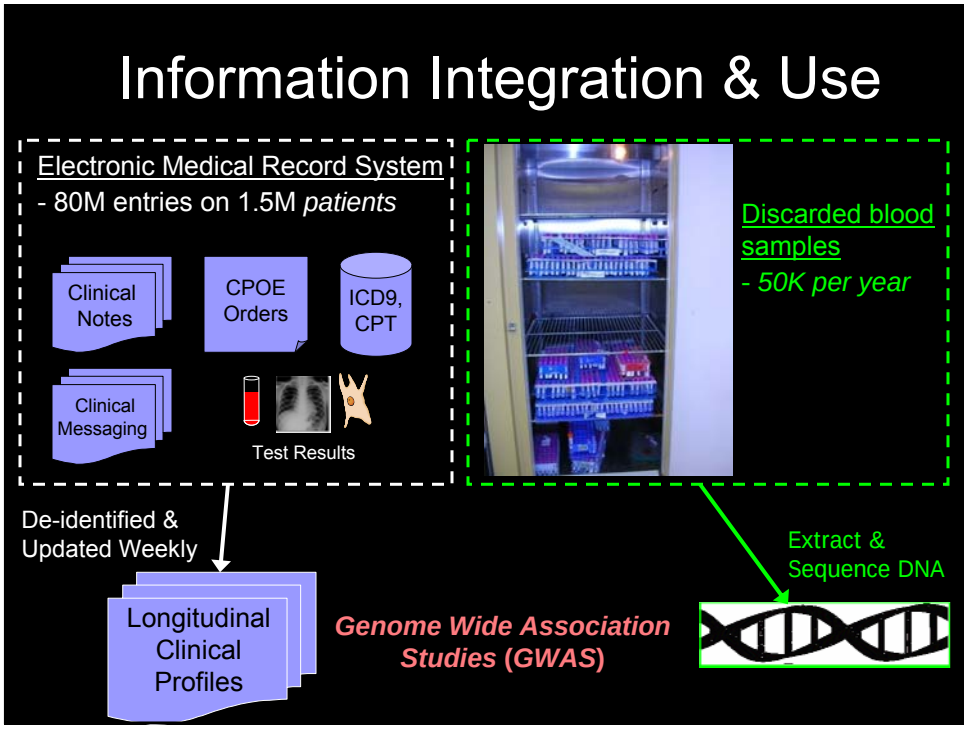
DNA from as many as 400,000 people will be fed over five years into a database at Vanderbilt University Medical Center under a $5 million research program expected to launch in the fall.

Patients at the hospital and its clinics will have the option to call a hot line and opt out.

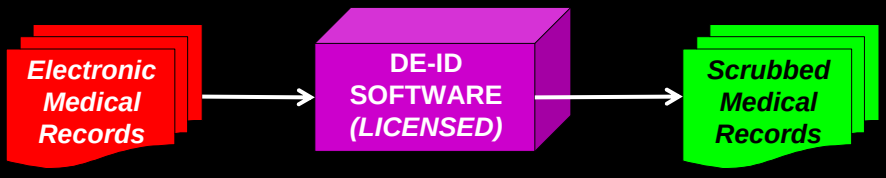
The data will be extracted from blood that would otherwise be thrown out, from lab tests or other uses. Researchers primarily will use the data to conduct research on how to eliminate adverse drug reactions, which kill 100,000 people nationally each year.

"We find that a one-size-fits-all approach to therapy or diagnosis is often inadequate," said Dr. Jeff Balser, Vanderbilt's associate vice chancellor for research. "A good example might be can-

Information Integration & Use



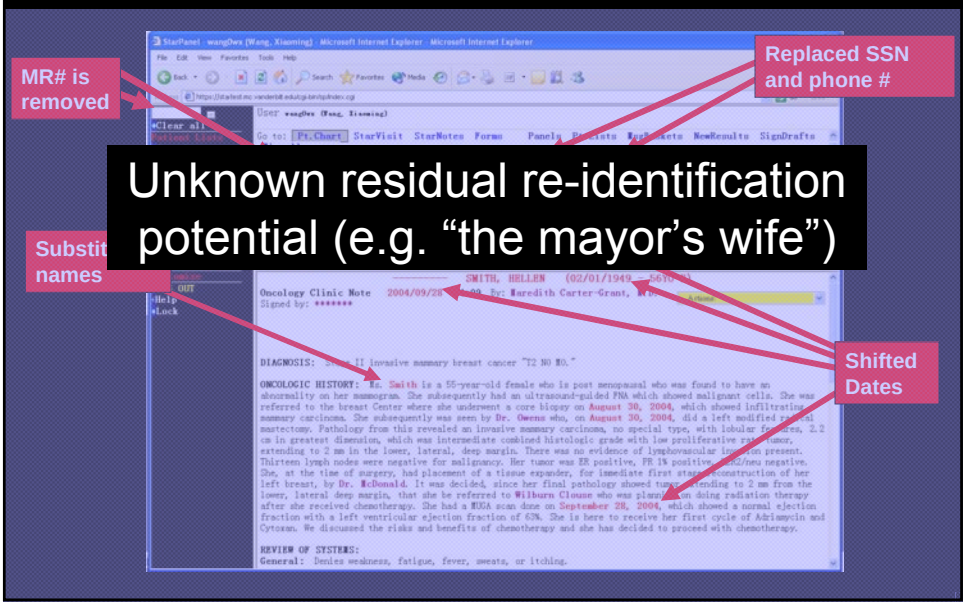
Text De-Identification Process



Scrubbed HIPAA identifiers

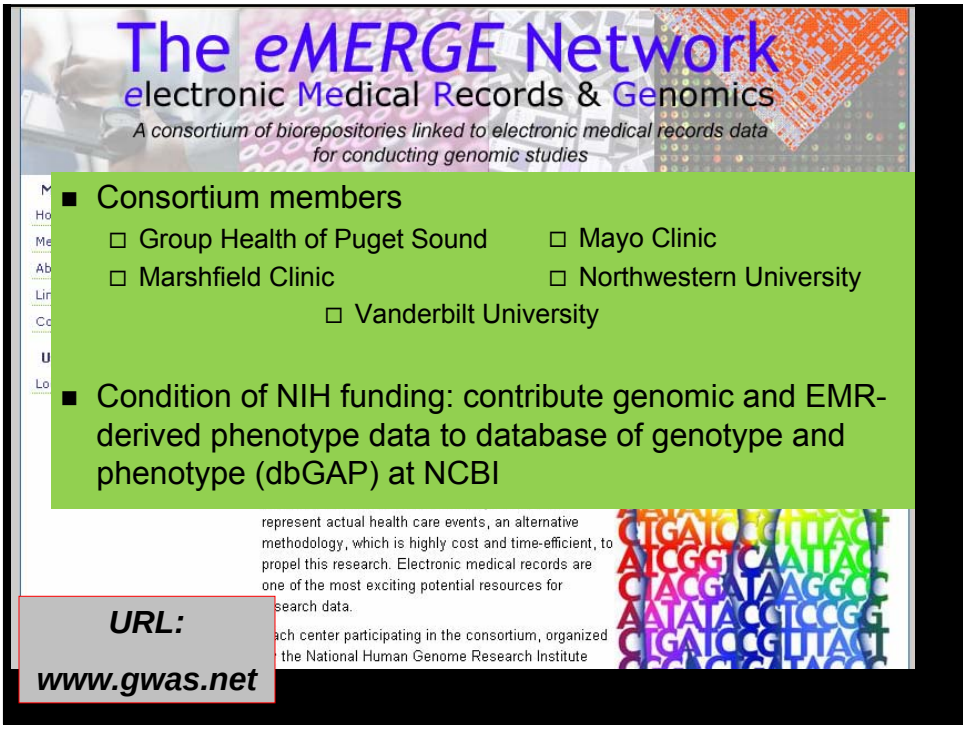
Names → **NAME[XXX, YYY]
Geographical → **PLACE , **INSTITUTION, **STREET-ADDRESS, **ZIP-CODE
Dates → **DATE, **AGE
Phone & Fax Numbers → **PHONE
Email Addresses → **EMAIL
SSN/MRN//Other IDs → **ID-NUM
Device → **DEVICE-ID
URLs / IP Addresses → **WEB-LOC
Pathology Specimen # → **PATH-NUMBER

Example De-identified Medical Record



Technology + Policy

- Databank access restricted to Vanderbilt employees
  - it is NOT a public resource
- Databank users sign Data Use Agreement prohibiting “re-identification”
- Access approved on project-specific basis by Operations Advisory Board and Institutional Review Board
- Project-specific user ID and password; all data access logged and audited



NIH GWAS Policy

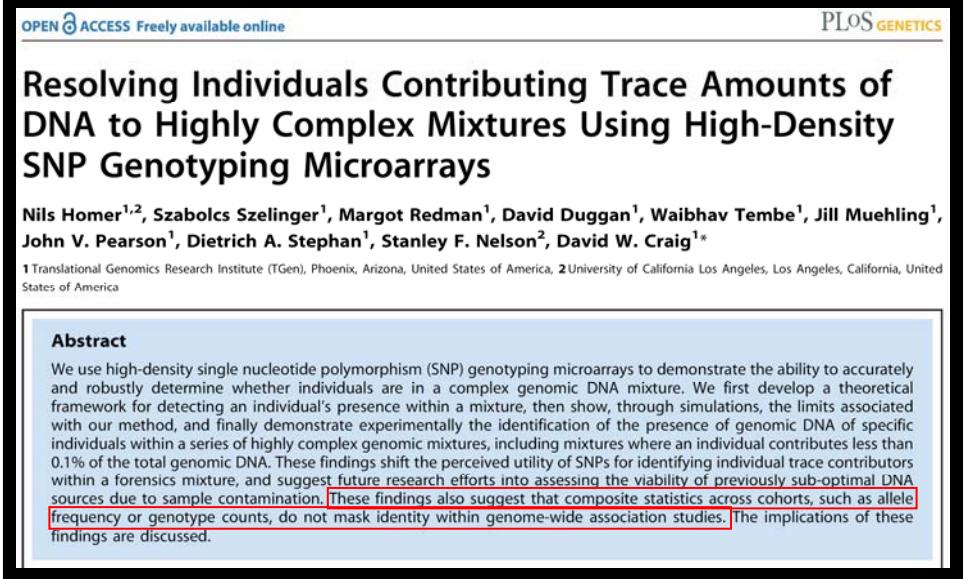
- Data de-identified & accompanied by written certification:
  - Remove the 18 HIPAA identifiers
  - Identities cannot be readily ascertained
  - Submitter has no knowledge data could be used to identify the subject [does not cover coded data]

NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) . NOT-OD-07-088. Aug 28, 2007.

"HIPAA" & Identifiers

- Safe Harbor – must remove
  - Biometric identifiers, including finger and voice prints
  - Any other unique identifying number, characteristic, or code
    - A code is an identifier if the person holding the coded data can re-identify the individual
  - DNA is highly unique (~75 SNPs can distinguish an individual*) ← Is this a "key"?

**Lin Z, Owen A, Altman R. Science. 2004.*



OPEN ACCESS Freely available online PLoS GENETICS

Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays

Nils Homer[1,2], Szabolcs Szelinger[1], Margot Redman[1], David Duggan[1], Waibhav Tembe[1], Jill Muehling[1], John V. Pearson[1], Dietrich A. Stephan[1], Stanley F. Nelson[2], David W. Craig[1]*

1 Translational Genomics Research Institute (TGen), Phoenix, Arizona, United States of America, 2 University of California Los Angeles, Los Angeles, California, United States of America

Abstract

We use high-density single nucleotide polymorphism (SNP) genotyping microarrays to demonstrate the ability to accurately and robustly determine whether individuals are in a complex genomic DNA mixture. We first develop a theoretical framework for detecting an individual's presence within a mixture, then show, through simulations, the limits associated with our method, and finally demonstrate experimentally the identification of the presence of genomic DNA of specific individuals within a series of highly complex genomic mixtures, including mixtures where an individual contributes less than 0.1% of the total genomic DNA. These findings shift the perceived utility of SNPs for identifying individual trace contributors within a forensics mixture, and suggest future research efforts into assessing the viability of previously sub-optimal DNA sources due to sample contamination. These findings also suggest that composite statistics across cohorts, such as allele frequency or genotype counts, do not mask identity within genome-wide association studies. The implications of these findings are discussed.

Homer N et al, PLoS Genet 2008 Aug 29;4(8):e1000167.

Headlines…

Los Angeles Times

DNA databases blocked from the public

The National Institutes of Health removes patients' genetic profiles from its website after a study reveals that a new type of analysis could confirm identities.

By Jason Felch
Los Angeles Times Staff Writer

Good for Cops, Bad for NIH

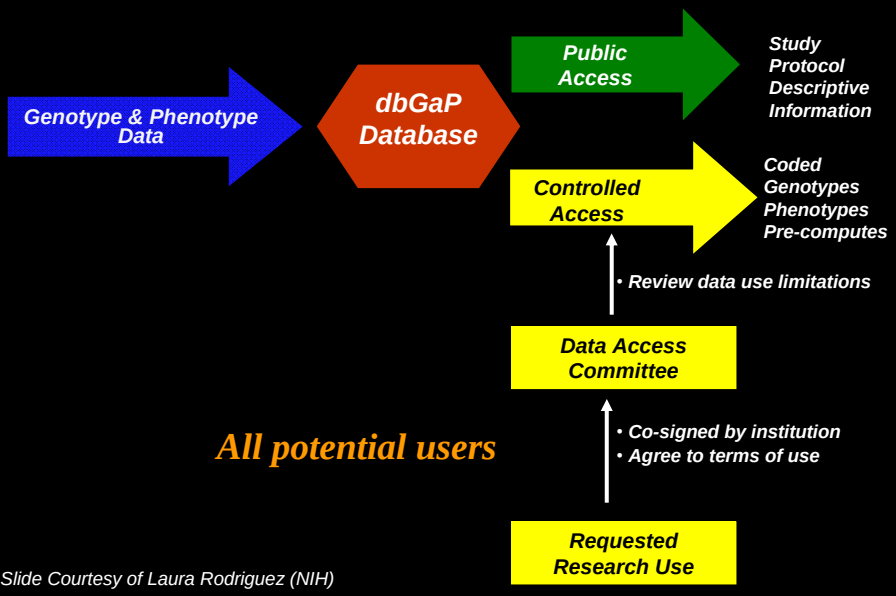
By Jennifer Couzin
*Science*NOW Daily News
29 August 2008

Forensic Breakthrough Stirs NIH to Close GWAS Data from Public View

August 29, 2008
By Matt Jones,
a GenomeWeb staff reporter

Slide Courtesy of Laura Rodriguez (NIH)

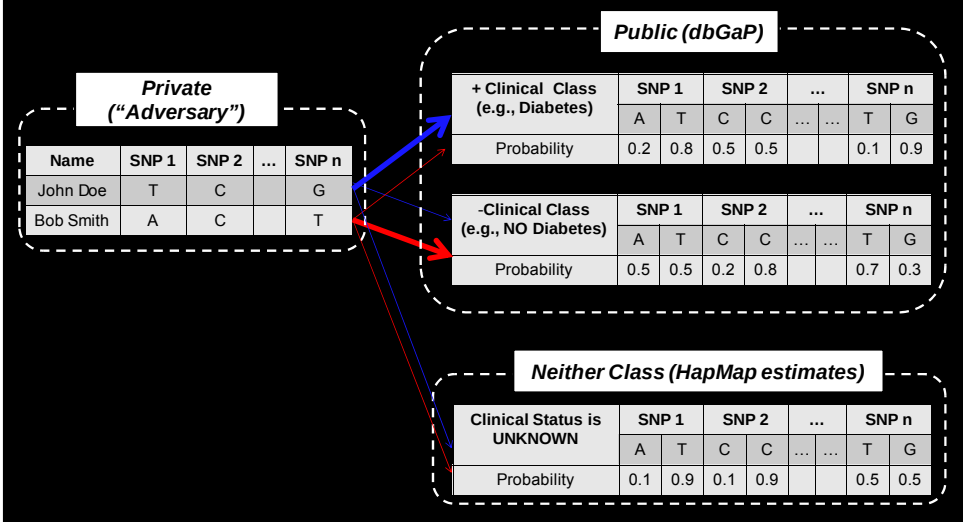
Data Access is Two-Tiered



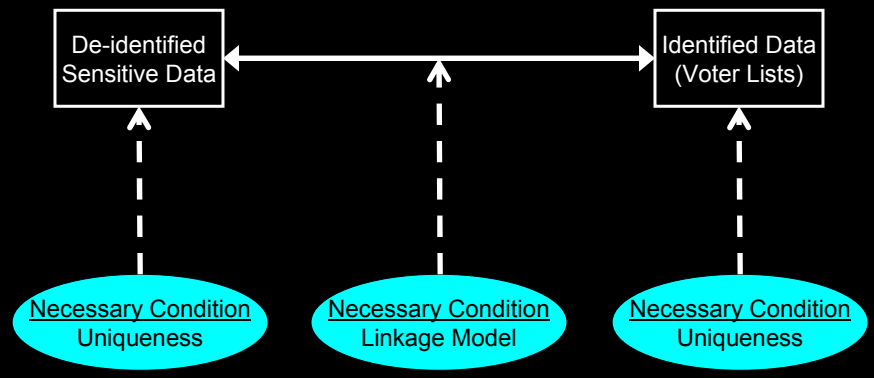
Slide Courtesy of Laura Rodriguez (NIH)

We Fear What We Don't Understand

The "Homer" Attack
(Attribute Inference)

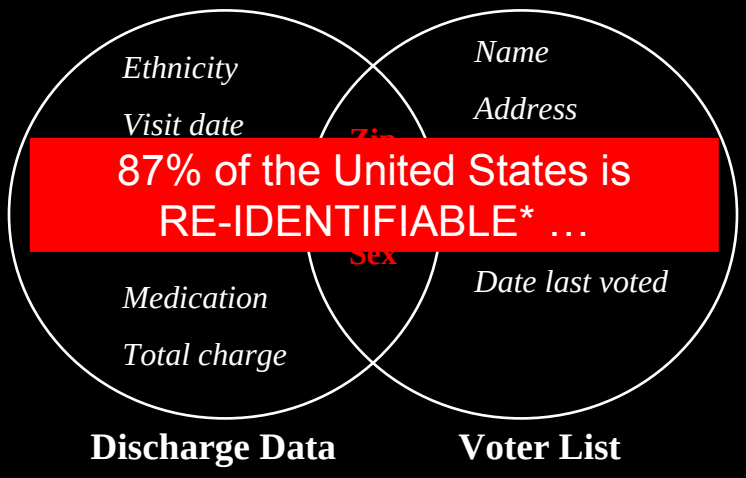


Central Dogma of Re-identification



*Malin B, Kantarcioglu M, Cassa. Book chapter forthcoming in 2010.

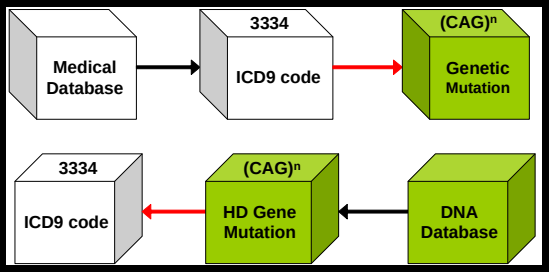
Outsiders Link to Re-identify Data



*Sweeney L. Journal of Law, Medicine, and Ethics. 1997.

DNA Re-identification

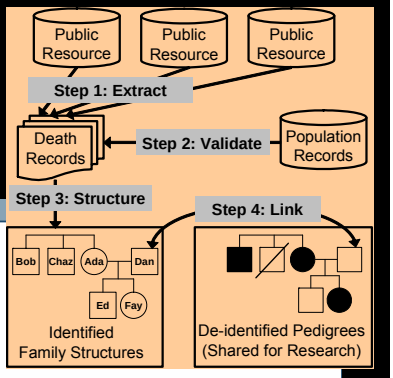
- Many deployed genomic privacy technologies leave DNA susceptible to re-identification*
- DNA is re-identified by automated methods, such as:
  - Genotype – Phenotype Inference **



*Malin B. Journal of the American Medical Informatics Association. 2005; 12(1): 29-34.
**Malin B & L Sweeney. AMIA Symposium 2000; PSB 2002.

Familial Re-identification

- *IdentiFamily*: software that links de-identified families to named individuals*
- Uses publicly available information, such as death records, to build genealogies



OBITUARIES

Richard R. Mann

1924-2007

Richard R. Mann, 82, of Cheyenne died Jan. 12 at Cheyenne Regional Medical Center.

He was born June 29, 1924, in Allentown, Pa., and had lived here since 1956.

Mr. Mann served in the Army Air Corp during World War II in South Africa and Italy.

He retired as a flight engineer for the Wyoming Air National Guard.

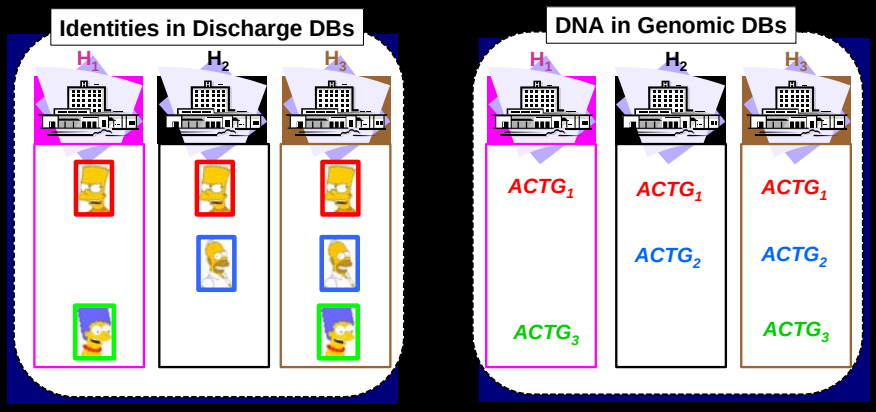
Mr. Mann was a member of St. Mary's Catholic Church, Elks, Moose and the Knights of Columbus, where he had been a past grand knight and state deputy.

He is survived by two sons, Gerald Mann and Thomas Mann, both of Cheyenne; seven daughters, Teresa Johnson, Kathryn Schroll, Judith Oldenburg, Cheryl Thibault, and Jon Cameron, all of Cheyenne, Lou Ann Golden of Sidney, Neb., and Kimberly Byron of Littleton, Colo.; his companion, Katie Heaton of Cheyenne; 25 grandchildren and two great-grandchildren.

He was preceded in death by his wife of more than 50 years, Patricia A. Mann; two daughters, Mary Constance Grant and Jeanane Rhodes; his parents, Russell and Viola Mann; two brothers, Roland Mann and Robert Mann; and a sister, Rochelle Behrandt.

*Malin B. AMIA Symposium. 2006: 524-528.

Trails!



*Malin B, Sweeney L. AMIA 2001; Journal of Biomedical Informatics 2004; AMIA 2005.
*Malin B, Airoldi E. Privacy Enhancing Technologies Conference 2006.

Inside Attacks

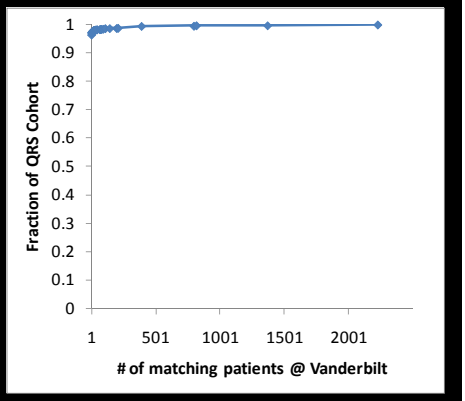
De-identified EMR + DNA ⟷ Identified EMR only

Diagnosis Codes

- ~50% of Vanderbilt patients with at least 1 diagnosis code are unique!
- ~75% " " " " " " " 2

Leveraging Diagnosis Codes

- Cohort: 2500 Vanderbilt patients in a GWAS
- Each patient: set of ICD-9 codes
- "distinctiveness" with respect to entire Vanderbilt population (1.5 million)
- ~97% unique



**Loukides G, Denny J, & Malin B. AMIA Symposium. 2009 (presented two days ago!)*

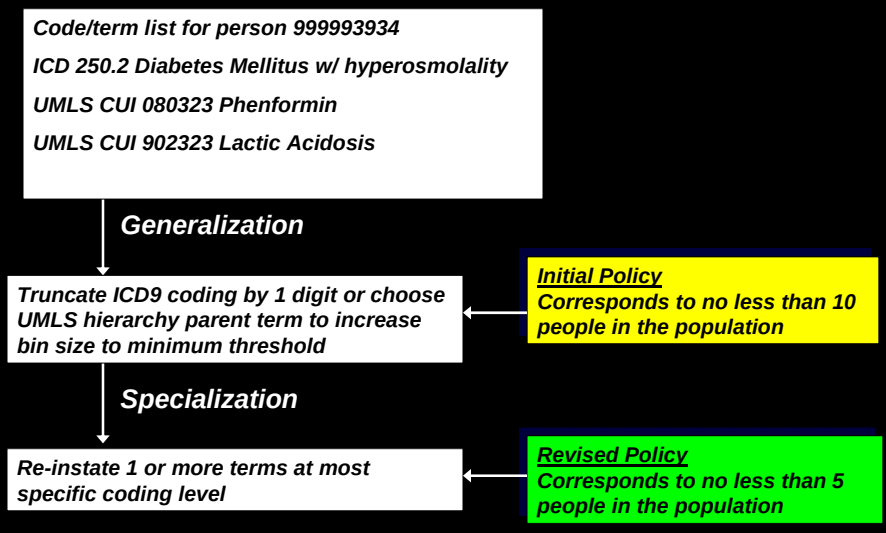
Is All Hope Lost?

Should we Give Up?

No! We can Protect Data

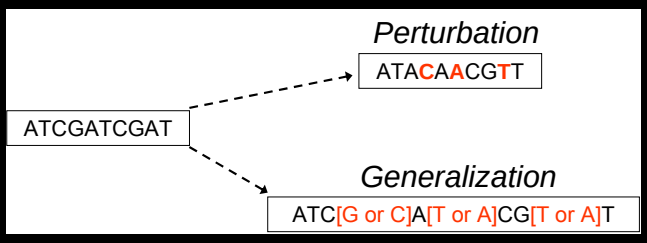
- Many ways to prevent these problems
  - Threat Modeling (the How)
  - Access Control (the Who)
  - Disclosure Control (the What)

Generalization / Specialization of EMR Coded Data



Protection of DNA?

- Alternatives: Perturbation*
- Alternative: Generalization of data**
  - Retains semantics
  - Given enough data – can reconstruct aggregate distributions and associations



*Lin Z, Owen A, Altman R. Science. 2004.
**Malin B. Methods of Information in Medicine. 2005.

K-Protection

K-Map: Every Record maps to K people in the population

Name	Year of Birth	Zip	Warfarin Metabolism	SNPs
*	1963	3720*	High	{A,C}
*	1963	3720*	Low	{A,T}
*	1961	3720*	High	{A,C}
*	1964	3720*	Medium	{A, T}

Sample

Name	Year of Birth	Zip	Warfarin Metabolism	SNPs
*	1963	3720*	High	{A,C}
*	1963	3720*	Low	{A,T}
*	1961	3720*	High	{A,C}
*	1964	3720*	Medium	{A, T}
*	1963	3720*	?	{A,C}
*	1963	3720*	?	{A,T}
*	1961	3720*	?	{A,C}
*	1964	3720*	?	{A, T}

Population

*Sweeney L. International Journal of Uncertainty, Fuzziness, & Knowledge-based Systems. 2002.

K-Protection

Population Unknown?

No Problem: Enforce Protection on the sample

k-Anonymity: Every record maps to K people in the sample

Name	Year of Birth	Zip	Warfarin Metabolism	SNPs
*	1963	3720*	High	{A,C}
*	1963	3720*	Low	{A,T}
*	1961	3720*	High	{A,C}
*	1964	3720*	Medium	{A, T}

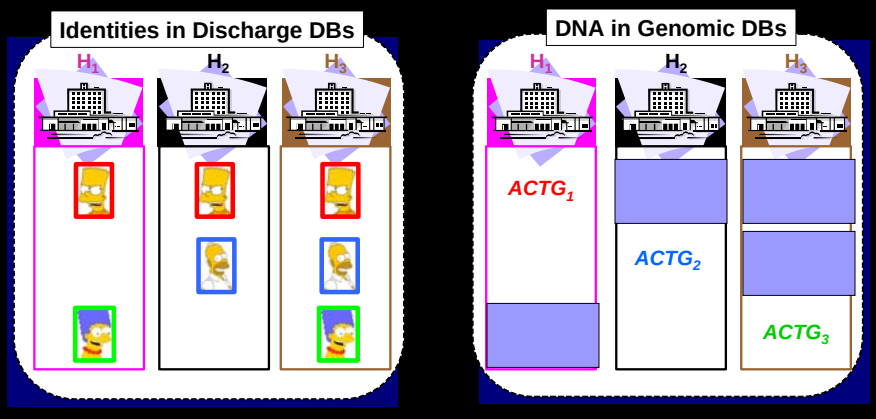
2-Map

Name	Year of Birth	Zip	Warfarin Metabolism	SNPs
*	196[1 OR 3]	3720*	High	{A,C}
*	196[3 OR 4]	3720*	{Low or Medium}	{A,T}
*	196[1 OR 3]	3720*	High	{A,C}
*	196[3 OR 4]	3720*	{Low or Medium}	{A, T}

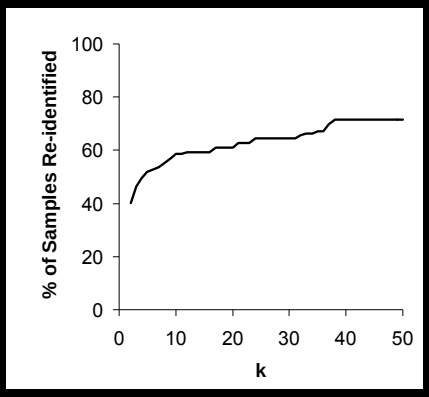
2-Anonymous Sample

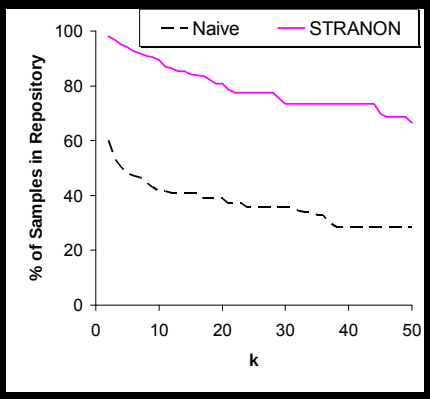
*Sweeney L. International Journal of Uncertainty, Fuzziness, & Knowledge-based Systems. 2002.

Trail Anonymization



Trail Anonymization





BEFORE Protection
100% Samples In Repository

AFTER Protection!
0% Samples Re-identified

*Malin B. Artificial Intelligence in Medicine. 2007.
*Malin B. Artificial Intelligence in Medicine. 2009.

Addressing Homer Directly

Private ("Adversary")

Name	SNP 1	SNP 2	…	SNP n
John Doe	T	C		G
Bob Smith	A	C		T

Public (dbGaP)

+ Clinical Class (e.g., Diabetes)	SNP 1		…		SNP n	
	A	T	…	…	T	G
Probability	0.2	0.8			0.1	0.9

-Clinical Class (e.g., NO Diabetes)	SNP 1		…		SNP n	
	A	T	…	…	T	G
Probability	0.5	0.5			0.7	0.3

Neither Class (HapMap estimates)

Clinical Status is UNKNOWN	SNP 1		SNP 2		…		SNP n	
	A	T	C	C	…	…	T	G
Probability	0.1	0.9	0.1	0.9			0.5	0.5

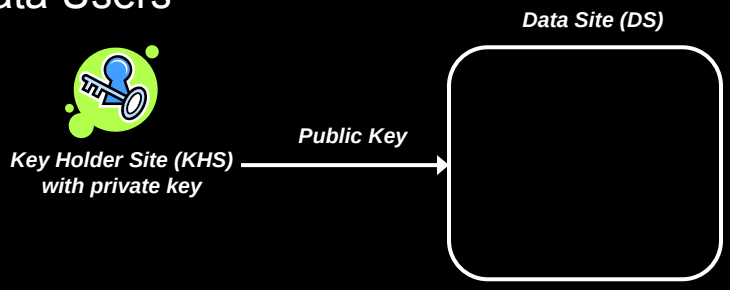
- Publish a subset of SNPs!
- Which SNPs?
  - In combination, the probability of classifying the known person is low.

*Sankararam S et al. Nature Genetics. 2009; 41: 965-967.

Secure Record Management*

(Murat Kantarcioglu will talk about this)

- Data Providers
- Third Party Data Managers ← Required
- Data Users



Key Generation: <public, private>

*Kantarcioglu M, Liu Y, Jiang W, Malin B. IEEE Transactions on Information Technology in Biomedicine. 2008.

The Landscape

- There **exists** a potential for privacy compromise
- There **exists** a means to thwart the problems
- Recommendations
  - Model the **threats** (Who? Cost? When?)
  - Consider the **replicability** of features
  - Apply Formal **risk mitigation** methods

Acknowledgements

- Vanderbilt Health Information Privacy Laboratory (www.hiplab.org)
  - Kathleen Benitez
  - Elizabeth Ashley Durham
  - Aris Gkoulalas-Divanis, Ph.D.
  - Grigorios Loukides, Ph.D.
  - **Bradley Malin, Ph.D.**
  - Acar Tamersoy
- Collaboroators
  - Edoardo Airoldi, Ph.D. (Harvard)
  - Wei Jiang, Ph.D. (MUST)
  - Murat Kantarcioglu, Ph.D. (UTD)
  - Ying Liu, Ph.D. (UTD)
  - Latanya Sweeney, Ph.D. (CMU)
- Funding
  - NHGRI/NIH U01HG004603 (Roden)
    - "Vanderbilt Genome Electronic Records Project"
  - NLM/NIH R01LM009989 (Malin)
    - "Technologies to Enable Privacy in Biomedical Databanks"

Session 2B: Privacy Considerations in Disease Surveillance

Session Chair: Philip AbdelMalik, PHAC

Bio:

Philip AbdelMalik is currently an Epidemiologist and Senior GIS Analyst at the Public Health Agency of Canada's Office of Public Health Practice.

Prior to joining the Agency, Philip was a research coordinator at the Clinical Genetics Research Program, at the University of Toronto / Centre of Addiction and Mental Health, where his work focused on the epidemiology and genetics of schizophrenia, particularly in relation to head trauma.

Since joining the Agency in early 2004, Philip's primary research focus has been the use and promotion of Geomatics in epidemiology and public health, with particular emphasis on issues of location-privacy. Philip completed his M.H.Sc. in Community Health and Epidemiology at the University of Toronto, and is currently a Ph.D. candidate in Public Health Informatics at the Peninsula Postgraduate Health Institute in the UK.

Sharing Personal Health Information for Syndromic Surveillance: Lessons learned in Ontario

Anita Fineberg, Anita Fineberg & Associates

Abstract:

This presentation will initially describe the privacy issues that needed to be addressed in the development of a data sharing agreement between hospitals and a public health unit for the purposes of a Syndromic Surveillance project. While such an agreement is not legally required, the hospitals were reluctant to disclose personal health information in the absence of written assurances with respect to how the public health unit would subsequently use and manage the hospital data.

During the development of the agreement, it became clear that several misunderstandings and misperceptions exist within the healthcare community with respect to the sharing of personal health information for "public health purposes". Questions were raised relating to the circumstances in which an agreement was needed, as well as the timing and scope of mandatory reporting requirements of personal health information. Healthcare professionals were also uncertain as to whom discretionary disclosures could be made for these purposes.

Changes were made to both Ontario's public health and personal health information privacy legislation – respectively, the Health Protection and Promotion Act and the Personal Health Information Protection Act, 2004 – as a result of the report issued by Mr. Justice Campbell, the Commissioner investigating The Introduction and Spread of SARS in Ontario. These changes were made in order to facilitate necessary sharing of personal health information. However, the experience of drafting the Syndromic Surveillance agreement revealed that these changes have not been communicated effectively to those within the healthcare community who need to know. Using the agreement prepared for the Syndromic Surveillance project as a baseline, the presentation will review the "lessons learned in Ontario" with respect to the spectrum of disclosures for "public health purposes" which may and must be made by health information custodians in the province.

Bio:

Anita Fineberg, LL.B., CIPP/C is the President of Anita Fineberg & Associates Inc., a recently incorporated consulting company with a mandate to provide superior, cost-effective and practical privacy solutions for the private sector, government and other public sector entities. She is both a lawyer and a CIPP/C (Certified Information Privacy Professional/Canada). Anita has:

Close to 20 years of experience providing advice on complex access to information and privacy issues with a specialization in health information privacy

Expertise in the interpretation and application of all Canadian privacy legislation, including the Personal Information Protection and Electronic Documents Act (PIPEDA), provincial private and public sector laws and health information privacy legislation

As Corporate Counsel & Chief Privacy Officer at IMS Health Canada and Latin America, successfully managed all internal compliance privacy matters and government advocacy initiatives

Advised the Ontario Ministry of Health and Long-Term Care on government privacy compliance and the privacy implications of new legislation and technologies: the former Smart Systems for Health Agency, Smart Cards, Public Key Infrastructure and the development of health information privacy legislation

Acted as Counsel to the Ontario Information and Privacy Commissioner

Anita is a frequent speaker and course leader at privacy conferences and workshops both domestically, in the U.S. and around the world. She holds a B.A. (Hons.) degree in psychology from Queen's University and an LL.B. from the University of Toronto.

Sharing Personal Health Information for Syndromic Surveillance: Lessons Learned in Ontario

EHIP 2009
Privacy Considerations in Disease Surveillance
November 19, 2009

Anita Fineberg, LL.B., CIPP/C
Barrister & Solicitor
President
Anita Fineberg & Associates Inc.

Agenda

- The ASSET Project
- The Privacy Issues
- The Data Sharing/Research Agreement
- Concerns
- The Framework for Disclosure of PHI for "Public Health Purposes" in Ontario
- SARS déjà vu?
- Lessons Learned
- Contact information

The ASSET Project

- Background
  - routine monitoring of specific indicator variables; e.g. school absenteeism, OTC drug sales ; emergency room patients with symptoms typical of specific diseases -> early indicator of disease outbreak
  - application of information technology to convert free-text emergency room records to standardized format for statistical analysis
  - goals of this phase: (i) deploy the system in 4 emergency departments in Ottawa; (ii) develop, deploy and evaluate an improved ASSET system; (iii) develop response protocols specific for Ottawa Public Health (OPH)
  - this phase hypothesized that a syndromic surveillance system could be successfully deployed in Ottawa

Provide objective data regarding the feasibility of using information technology to mine medical records for population-based health applications

The ASSET Project

- The Information
  - real-time data collection from the emergency departments of four Ottawa hospitals
  - supplied to OPH
  - no patient identifiers (name, dob, address and OHIP#) collected but data assumed to be identifiable personal health information (PHI) because of the scope of the elements
  - patient consent impractical because of volume

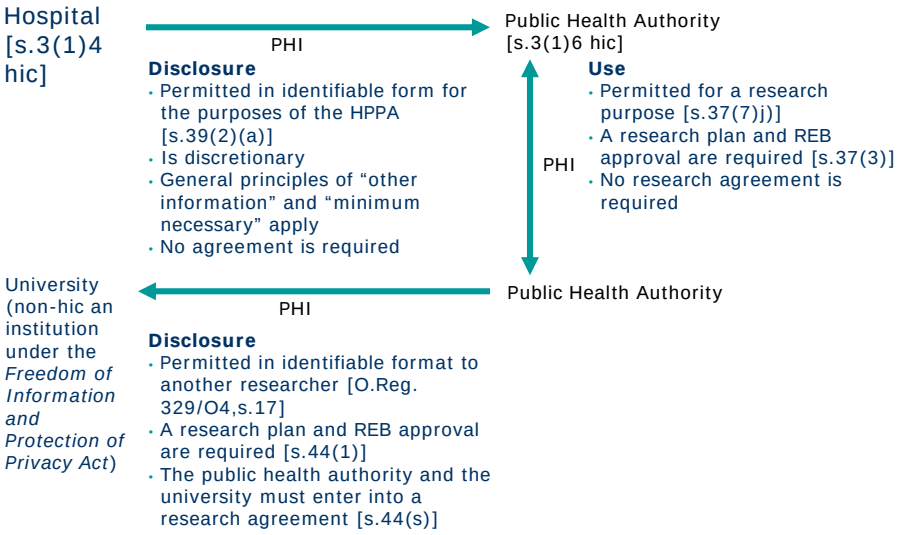
The ASSET Project

- Information Uses and Disclosures
  - used by OPH for the purposes of disease surveillance and research
  - disclosed by OPH to external entities for research purposes
- Data Confidentiality and Security
  - data committee including privacy officers of the Ottawa hospitals and OPH
  - signing of confidentiality agreements
  - controlled data access
  - physical and technical security safeguards

Privacy Issues

- Participating Ottawa hospitals are health information custodians (HICs) under the Ontario *Personal Health Information Protection Act, 2004* (PHIPA)
- Disclosures of PHI by HICs to OPH:
  - for public health purposes
    - no patient consent required
    - no agreement required
  - for internal research purposes
    - no patient consent required if REB approval of research plan
    - research agreement between hospitals and OPH
- Further disclosures of PHI by OPH to researchers:
  - research agreement required by OPH and the external researchers

Privacy Issues



The Data Sharing/Research Agreement

- Template drafted to address non-specific information for "situational awareness"
- Combined data sharing and research agreement
  - determined that the disclosure of Hospital PHI to meet the objective of the ASSET project was a "public health purposes" within the meaning of the HPPA
  - while not legally necessary, the Agreement was drafted to provide hospitals with a "comfort level" re: disclosure of PHI to OPH for "public health purposes"

The Data Sharing/Research Agreement

- Combined data sharing and research agreement (cont'd)
  - Lead to challenges related to inconsistent provisions of PHIPA depending upon whether OPH was a "researcher" or a HIC receiving PHI from another HIC(the Hospitals):
    - notification of individuals in the event of a data breach – communication had to be made with the Hospitals to ensure accurate data identification for potentially numerous individuals
    - requests by individuals for access to their own PHI had to be referred back to the Hospitals
    - discretionary disclosures by the OPH to be discussed with the Hospitals but OPH had the decision-making authority
    - OPH to advise the Hospitals in the event that it received a court order, subpoena; i.e. mandatory disclosure demands
    - OPH to notify the Hospitals if it received any privacy complaints from individuals or notification of a self-initiated review by the Ontario Privacy Commissioner
    - timing by which and nature of the information which the OPH was to advise the Hospitals of any of these events

The Data Sharing/Research Agreement

- Research issues
  - the combined "data sharing and research agreement" was deemed by the parties to constitute a "research agreement" for the purposes of PHIPA
  - identification of which REB could be used to approve a research plan – left to the mutual agreement of the parties
- General matters
  - security safeguards that were required of any entities, including OPH and researchers down the line, who would have access to the PHI
  - Hospitals had the authority to request and the OPH to provide a copy of any and all of is policies related to privacy, confidentiality and security of data
  - detailed provisions for retention and destruction/return of the data

Concerns

- Hospital concerns re: disclosure of PHI to OPH without some form of an agreement
- No ability to share PHI outside of Ontario with provincial counterparts or provision to federal agencies such as the Public Health Agency of Canada or those outside of Canada
- What's the difference between the PHI being used for the ASSET project and reporting requirements under the *Health Protection and Promotion Act* (the "HPPA)?
- ASSET uses a software program to categorize cases; mandatory reporting under the HPPA requires the opinion of a healthcare professional

The framework for disclosure of PHI for "public health purposes" in Ontario

- HICs <u>may</u> disclose PHI without consent:
  - to the CMO or medical officer of health within the HPPA for the purposes of the HPPA
  - to the Ontario Agency for Health Protection and Promotion for a purpose of the *Ontario Agency for Health Protection and Promotion Act, 2007* (OAHPPA); or
  - to a public health authority similar to (i) and that is:

    established under laws of **Canada**, **another** Canadian **province or territory**, or **another jurisdiction**

    If

    the disclosure is for a purpose substantially similar to a purpose of the HPPA
- No agreement is required

This is a <u>discretionary</u> disclosure: HICs can choose to do so or not

The framework for disclosure of PHI for "public health purposes" in Ontario

- Identified healthcare practitioners and enumerated other groups; e.g. school principals, must disclose (report) PHI without consent in situations set out in the HPPA
- Influenza example
  - influenza is a listed communicable and reportable disease
  - ealthcare practitioners who form the opinion that a person has or may have a reportable disease, or is or may be infected with an agent of a communicable disease must report to the Medical Officer of Health
  - hospital administrators have a duty to report if an entry in the records of a hospital in- or out-patient indicates that the person has or may have a reportable disease or is or may be infected with an agent of a communicable disease
  - reports must contain the patient's name, dob, sex etc.

The framework for disclosure of PHI for "public health purposes" in Ontario

- PHIPA addresses such mandated reporting/disclosures under the HPPA
- It permits HICs to disclose PHI without consent if "… permitted or required by law …
- Because HPPA mandates the reporting, HICS may rely on the section of PHIPA above as authorization for the PHI disclosure

SARS déjà vu?

- "Whatever the precise path of legislative reform, privacy, while vital, should not impede the necessary sharing between agencies and governments of information required to protect the public against an outbreak of infectious disease."

The SARS Commission Interim Report – SARS and Public Health in Ontario
by The Honourable Mr. Justice Archie Campbell Commissioner

SARS déjà vu?

- Perception that the law inhibited and prohibited disclosures of information needed to contain and deal with disease spread
- SARS was not listed in the HPPA specification as a reportable, communicable or virulent disease
- Even once listed, unclear how much information and to whom healthcare providers could share

SARS déjà vu?

- Implementation of recommendations in Mr. Justice Campbell's report resulted in:
  - Amendments to both PHIPA and the HPPA
  - Disclosures for the purposes of the HPPA: to provide for the organization and delivery of public health programs, the prevention of the spread of disease and the promotion and protection of the health of the people of Ontario
  - HICS may disclose PHI to the CMO or a medical officer of health even in the absence of a duty to report under the HPPA or a request for information

Lessons Learned

- Numerous challenges in drafting data sharing/research agreements
- Concerns and misunderstandings still exist re: ability of healthcare practitioners to share information in an environment of disease epidemic

The New [*Health Information*] *Protection Act, 2003* allows the health information custodian to disclose – it says "may" and not "shall" – about information of an individual to the Chief Medical Officer of Health or Medical Officer of Health and is very broad. It says for the purpose of that Act. I understand that … there has been a lot of opposition to that particular section. I think that section is great because it will help public health move quickly and collect information that it needs when faced with a situation such as SARS or another influenza pandemic; I am concerned that section is going to be wiped out in the future iteration of the Bill.

Anonymous public health official
quoted in the *Interim Report*

Contact Information

Anita Fineberg, LL.B., CIPP/C
Barrister & Solicitor
President
Anita Fineberg & Associates Inc.

afineberg@sympatico.ca
Bus: 416.762.4583
Cell: 416.565.5007
Fax: 877.475.7096

Real Time Privacy Assessment in H1N1 Reporting: Protecting privacy in an evolving clinical scenario

Jay Mercer, Canadian Medical Association & Practice Solutions

Abstract:

In response to the H1N1 pandemic, a team composed of clinicians, industry, researchers, public health officials and a privacy specialist came together to carry out a project that would permit physicians with electronic medical records to report cases of influenza on a near real time basis, while protecting patient privacy. As the project progressed, the reporting requirements changed numerous times in response to developing understanding of the illness. Real time evaluation of privacy implications became a critical part of keeping the project on track. This presentation will describe the process that was used to develop the project, and also explain the role and importance of including a privacy specialist on the development team when the project is being conducted in a highly dynamic environment.

Bio:

Jay Mercer divides his time between family medicine in a fully automated office in Ottawa, as Medical Director of Practice Solutions Inc., the Canadian Medical Association's group of technology companies, and as a Senior Physician Advisor to the CMA in the area of practice technology. Previously, he spent several years in a combined family medicine and emergency practice in Midland, Ontario. Jay speaks and writes frequently about practice automation for physician groups across Canada. Previously, he was the project leader for the CMA's Physician Website and Patient Portal initiatives, as well as several other activities which integrate technology into patient care. Trained initially as a strategist, Dr. Mercer completed an MD degree in 1993, followed by training in Rural Family Medicine and Emergency Medicine. Prior to working with the CMA, Dr. Mercer's was a consultant involved in assisting governments, large organizations and private companies with healthcare strategic planning and program development in the areas of information management, security and privacy.

Real time privacy assessment in H1N1 reporting: Protecting privacy in an evolving clinical scenario

Dr. Jay Mercer MD, CCFP, FCFP
Dr. Tom Wong MD, FRCPC
Dr. Khaled El Emam, PhD

Project Objective

Carry out a project that would permit physicians with electronic medical records to report cases of influenza on a near real time basis, while protecting patient privacy

Who is on the team?

- Public Health
- Industry
- ITAC
- Clinicians

What went on?

- Project launched
- Understanding of the illness changed
- Public health needs changed
- Reporting requirements changed
- Functional specification changed

Theoretical Underpinning

"Privacy by design"

How do you protect privacy in this scenario?

- Get a privacy specialist on your development team
- Get them on early, and
- Listen to them

Who is on the team?

- Public Health
- Industry
- ITAC
- Clinicians
- Privacy Specialist

Issues Reviewed

- Risk of collecting data on vaccination, pneumonia & 4 key co-morbidities
- Risk of collecting data on pregnancy
- Risk of collecting data on FSA of the practice
- Role of PHAC in data protection
- Consent for data collection
- Consent for data sharing
- Provincial legislation and data sharing
- Press release on project activity

Timeline / Workload

- Start 27 April, 2009
- Privacy specialist engaged 3 weeks later
- 578 messages generated
- 125 retained
- 9 from privacy specialist
- 14 versions of the functional specification
- First test reporting site 29 Sept, 2009

But then the unexpected happened

- We decided that we needed Health Canada REB approval

Who is on the team?

- Public Health
- Industry
- ITAC
- Clinicians
- Privacy Specialist

Theoretical Underpinning

"Privacy by design"

REB Outcome

Public Health Agency of Canada / Agence de la santé publique du Canada

OCT 2 2 2009

Your file / Votre référence

Our file / Notre référence

Dr. Tom Wong
Director, CAID, CCDIC
Infectious Disease and Emergency Preparedness Branch
Public Health Agency of Canada
100 Eglantine Driveway, Room 2391
Ottawa, Ontario
K1A 0K9

Dear Dr. Wong:

Protocol Number: **REB-2009-0035**
Protocol Title: **Rapid Real Time Surveillance and Monitoring of Pandemic Influenza Severity & Associated Risk Factors Using Primary Care Electronic Medical Records**

Following Health Canada's Research Ethics Board's (HC REB) review of the above-listed research protocol and its recommendation for approval which took place on October 8, 2009, I am pleased to inform you that your project may proceed.

The responsibilities of Principal Investigators are set out in the enclosed Certificate of Ethics Approval. Please retain one copy for your records and return one signed copy to the:

Research Ethics Board (REB) Secretariat
Strategic Policy Branch, Health Canada
Holland Cross, Tower B
1600 Scott Street, Room 410
Address Locator 3104A
Ottawa, Ontario K1A 0K9

.../2

Canada

Issues Reviewed

- Risk of collecting data on vaccination, pneumonia & 4 key co-morbidities
- Risk of collecting data on pregnancy
- Risk of collecting data on FSA of the practice
- Role of PHAC in data protection
- Consent for data collection
- Consent for data sharing
- Provincial legislation and data sharing
- Press release on project activity
- Preparation for and presentation to HC REB

Theoretical Underpinning

"Privacy by design"

Can Patients be Re-identified from Emergency Department Data?

Khaled El Emam, CHEO RI & University of Ottawa

Abstract:

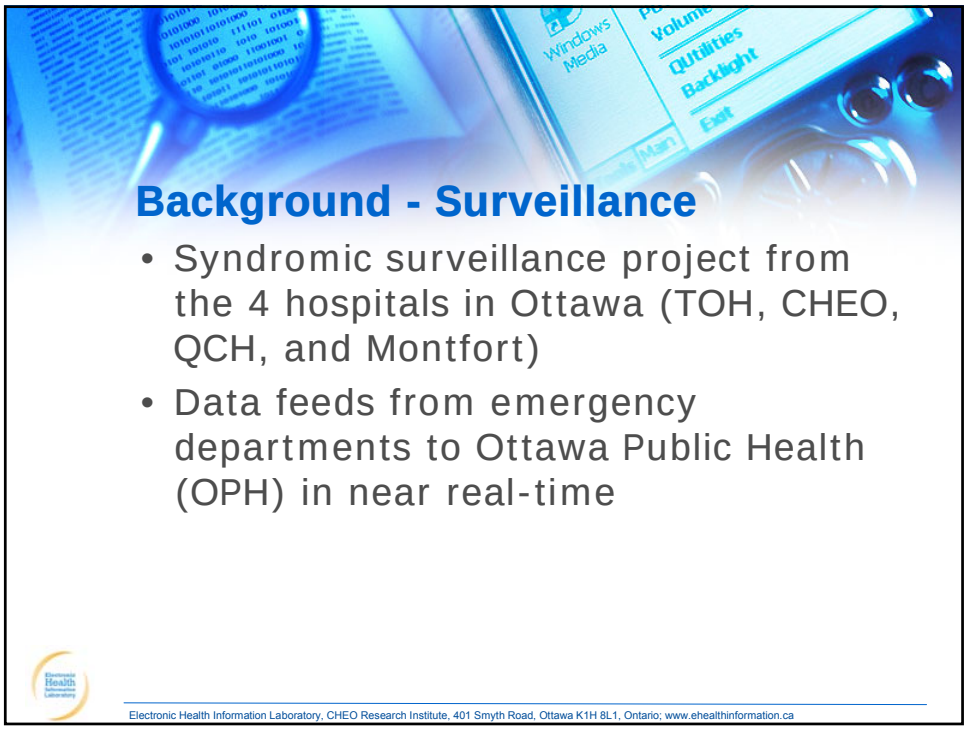
There is continuing reluctance to disclose health information for public health purposes unless it is de-identified. In this presentation we describe a re-identification risk assessment for emergency department data in the context of syndromic surveillance. We also provide methods for de-identifying location information so that it can be shared.

Bio:

Khaled El Emam, PhD, is an Associate Professor at the University of Ottawa, Faculty of Medicine and the School of Information Technology and Engineering. He is a Canada Research Chair in Electronic Health Information at the University of Ottawa. Previously Khaled was a Senior Research Officer at the National Research Council of Canada, and prior to that he was head of the Quantitative Methods Group at the Fraunhofer Institute in Kaiserslautern, Germany. In 2003 and 2004, he was ranked as the top systems and software engineering scholar worldwide by the Journal of Systems and Software based on his research on measurement and quality evaluation and improvement, and ranked second in 2002 and 2005. He holds a Ph.D. from the Department of Electrical and Electronics, King's College, at the University of London (UK). His lab's web site is: http://www.ehealthinformation.ca/.


Can Patients be Re-identified from Emergency Department Data ?
Khaled El Emam
CHEO RI
uOttawa


www.ehealthinformation.ca
Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario; www.ehealthinformation.ca



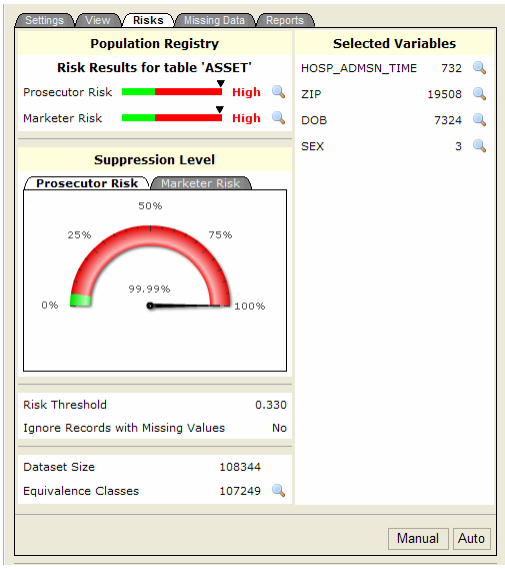
Background - Surveillance

- Syndromic surveillance project from the 4 hospitals in Ottawa (TOH, CHEO, QCH, and Montfort)
- Data feeds from emergency departments to Ottawa Public Health (OPH) in near real-time

Key Data Fields

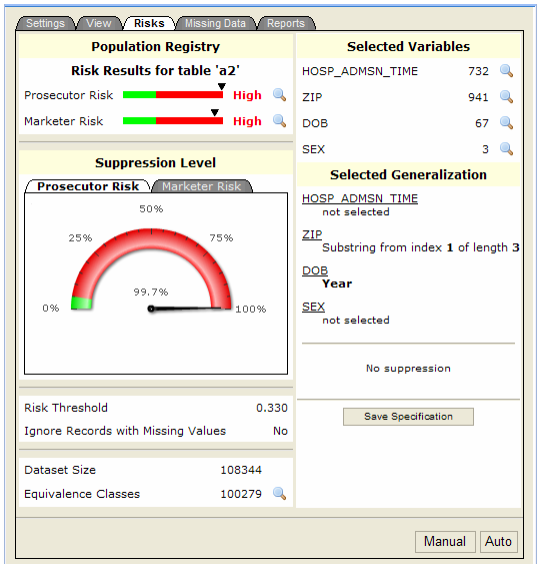
- Hospital identifier
- Date and time of presentation
- Chief complaint
- Patient postal code
- Patient date of birth
- Gender
- Case number (can be encrypted MRN)

Risk Assessment – Original Data



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario; www.ehealthinformation.ca

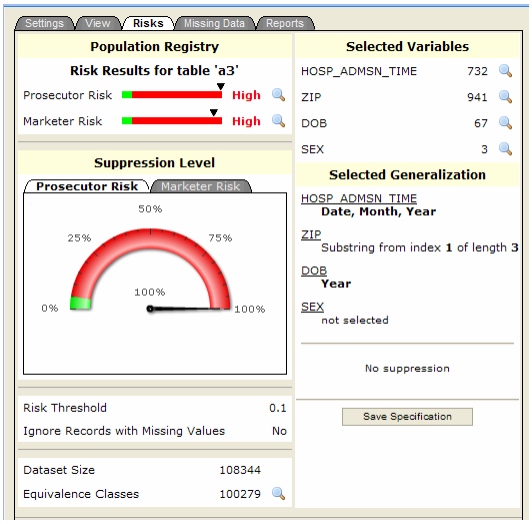
Risk Assessment – OPH Feed



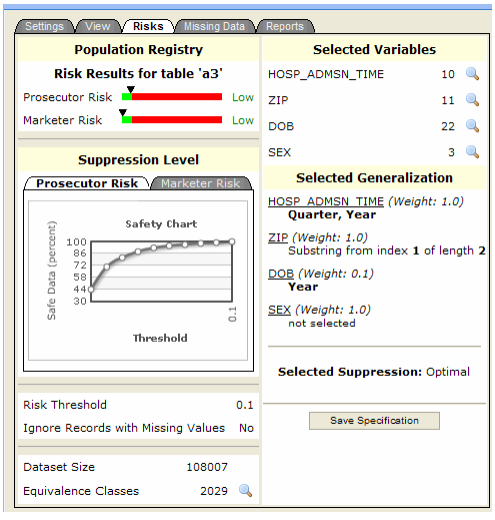
Is it De-identified ?

- The risk of re-identification is above the 0.33 threshold, but there is no legal risk from providing this data
- Sets the bar a little higher than providing the original data
- Even at a 0.5 threshold, ~86% of the records were high risk

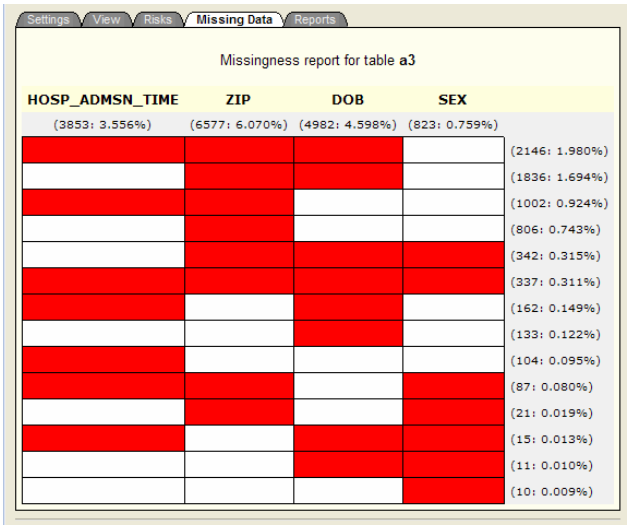
Researcher Request



Optimal Solution for Researchers



Missingness (max=10%)



De-identification for Researchers

	Solution 1 0.2 (9%)	Solution 2 0.33 (6.4%)	Solution 3 0.33 (10%)
Presentation Date	mm/yyyy	mm/yyyy	mm/yyyy
Gender	M/F	M/F	M/F
DoB	Quarter/Year	10 year interval	5 year interval
Location	Region	FSA	FSA


www.ehealthinformation.ca
Pan-Canadian De-Identification Guidelines for Personal Health Information
Heuristics for De-identifying Health Data
De-identification
Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario; www.ehealthinformation.ca

Session 1C: Genomics On-line

Session Chair: Patricia Kosseim, Genome Canada

Bio:

Patricia Kosseim has recently joined Genome Canada on a two-year Executive Interchange arrangement to lead a national strategy for addressing ethical, economic, environmental, legal and social (GE^3LS) issues related to large-scale genomics research. She joins Genome Canada from the Office of the Privacy Commissioner of Canada (OPC), where she has held the position of General Counsel since January 2005, responsible for the activities of the Legal Services, Policy and Parliamentary Affairs Branch.

Before joining OPC, Patricia spent five years building and heading up the Ethics Office of the Canadian Institutes of Health Research. During this period, she was briefly seconded to Canada Health Infoway Inc. to advise on privacy issues related to the development of pan-Canadian electronic health record systems.

Patricia worked in Montreal for over six years with the national law firm of Heenan Blaikie, practicing primarily in the areas of health law, human rights, labor & employment law, privacy law, administrative law, professional liability and civil litigation.

Called to the Québec Bar in 1993, Patricia holds degrees in Business (B.Com '87) and Laws (B.C.L. / LL.B. '92) from McGill University, and a Master's Degree in Medical Law and Ethics (M.A.'94) from King's College, University of London (U.K.).

Meome, Myome, Let's Share Our Genome

Mike Spear, Genome Alberta

Abstract:

In 1953 colour TV was just making it into our homes, cell phones were a dream, and Watson and Crick were letting the world know about the Double Helix structure of DNA. It is now 2010 and you can have a personal genome sequencing done for under a thousand dollars, store it on your phone while watching colour TV on the same phone, use the iPhone Merck Manual app to learn about some of the conditions you may have, and Tweet your followers about the results.

Once back at your computer you can run the raw data through a 3rd party SNP database to get even more information, find a group of people online who share a common interest, trait or disease, and do a more detailed comparison with people you have never met. With the help of Google, SNPedia, and Facebook you dig into the information in more detail, figure out a diet and exercise regime, and make an appointment with your doctor for some tests.

Is this useful or is it even accurate? You'll get a lot of different answers depending on your role and alliances in this online game of Risk. In this presentation you'll be prompted to think seriously about where research and policy should position themselves in the game as an incredible amount of health information is swirling around a growing number of people.

Bio:

Mike Spear is Director of Corporate Communications, Genome Alberta. He cut his teeth in the media business as a journalist, Producer, Executive Producer, and Program Manager with the CBC. His background includes the prestigious CBC President's Award, a media training mission to Croatia with the Washington D.C. based National Democratic Institute, lead on CBC Olympic coverage, and founder of CBC Radio's "Business Network". Mike moved over to other side of the journalist's microphone in 2006 and is currently Director of Corporate Communications with the not-for-profit research funding organization, Genome Alberta.

As part of his efforts to better understand genomics and to find a novel way of raising the profile of the science Mike has had his own personal genome sequencing done by 23andMe, deCODE, Navigenics, and the DNA Ancestry Project. He blogs about the experience at www.genomealberta.ca/blogs ,uses Twitter extensively as @mikesgene to talk about many aspects of genomics and has developed a Facebook news application called GenOmics (http://facebook.genomealberta.ca) to collect and distribute news, video, and blogs to people interested in many of the 'omics' sciences. Drawing from online experience that goes back to his early involvement with CompuServe and the Electronic Frontier Foundation, Mike speaks extensively at social media conferences and workshops.

Link to video of this presentation.

Meome, Myome, Let's Share Our Genome

Mike Spear
Genome Alberta
mspear@genomealberta.ca
Twitter: @mikesgene

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have had the DTC tests offered by:

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have had the DTC tests offered by:

- 23andMe www.23andme.com

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have had the DTC tests offered by:

- 23andMe www.23andme.com
- deCODE www.decode.com

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have had the DTC tests offered by:

- 23andMe www.23andme.com
- deCODE www.decode.com
- Navigenics www.navigenics.com

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have had the DTC tests offered by:

- 23andMe www.23andme.com
- deCODE www.decode.com
- Navigenics www.navigenics.com
- DNA Ancestry Project www.dnaancestryproject.com

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why ?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why ?

Because it was a good way to learn the science

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why ?

Because it was a good way to learn the science

Because it gave me an outreach tool to talk to the public

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why ?

Because it was a good way to learn the science

Because it gave me an outreach tool to talk to the public

Because I could

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have no kids

Why ?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have no kids

My sister is adopted

Why ?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why ?

I have no kids

My sister is adopted

My parents are well into their 80s

GenomeAlberta

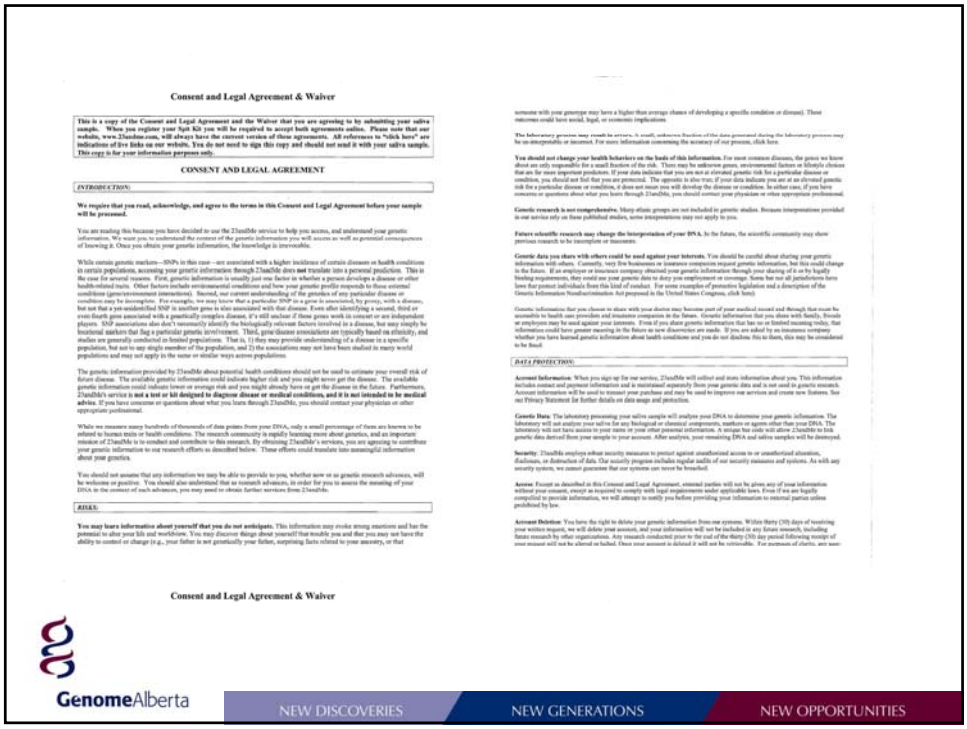
NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

There were detailed consent forms to sign and lots of information on the respective websites.

There were no surprises !

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES



Consent and Legal Agreement & Waiver

This is a copy of the Consent and Legal Agreement and the Waiver that you are agreeing to by submitting your saliva sample. When you register your Spit Kit you will be required to accept both agreements online. Please note that our website, www.23andme.com, will always have the current version of these agreements. All references to "click here" are indications of live links on our website. You do not need to sign this copy and should not send it with your saliva sample. This copy is for your information purposes only.

CONSENT AND LEGAL AGREEMENT

INTRODUCTION:

We require that you read, acknowledge, and agree to the terms in this Consent and Legal Agreement before your sample will be processed.

You are reading this because you have decided to use the 23andMe service to help you access, and understand your genetic information. We want you to understand the context of the genetic information you will access as well as potential consequences of knowing it. Once you obtain your genetic information, the knowledge is irrevocable.

While certain genetic markers—SNPs in this case—are associated with a higher incidence of certain diseases or health conditions in certain populations, accessing your genetic information through 23andMe does **not** translate into a personal prediction. This is the case for several reasons. First, genetic information is usually just one factor in whether a person develops a disease or other health-related traits. Other factors include environmental conditions and how your genetic profile responds to these external conditions (gene/environment interactions). Second, our current understanding of the genetics of any particular disease or condition may be incomplete. For example, we may know that a particular SNP in a gene is associated, by proxy, with a disease, but not that a yet-unidentified SNP is another gene is also associated with that disease. Even after identifying a second, third or even fourth gene associated with a genetically complex disease, it's still unclear if these genes work in concert or are independent players. SNP associations also don't necessarily identify the biologically relevant factors involved in a disease, but may simply be bacterial markers that flag a particular genetic involvement. Third, gene/disease associations are typically based on ethnicity, and studies are generally conducted in limited populations. That is, 1) they may provide understanding of a disease in a specific population, but not to any single member of the population, and 2) the associations may not have been studied in many world populations and may not apply in the same or similar ways across populations.

The genetic information provided by 23andMe about potential health conditions should not be used to estimate your overall risk of future disease. The available genetic information could indicate higher risk and you might never get the disease. The available genetic information could indicate lower or average risk and you might already have or get the disease in the future. Furthermore, 23andMe's service is **not a test or kit designed to diagnose disease or medical conditions, and it is not intended to be medical advice.** If you have concerns or questions about what you learn through 23andMe, you should contact your physician or other appropriate professional.

While we measure many hundreds of thousands of data points from your DNA, only a small percentage of them are known to be related to human traits or health conditions. The research community is rapidly learning more about genetics, and an important mission of 23andMe is to conduct and contribute to this research. By obtaining 23andMe's services, you are agreeing to contribute your genetic information to our research efforts as described below. These efforts could translate into meaningful information about your genetics.

You should not assume that any information we may be able to provide to you, whether now or as genetic research advances, will be welcome or positive. You should also understand that as research advances, in order for you to assess the meaning of your DNA in the context of such advances, you may need to obtain further services from 23andMe.

RISKS:

You may learn informative about yourself that you do not anticipate. This information may evoke strong emotions and has the potential to alter your life and worldview. You may discover things about yourself that trouble you and that you may not have the ability to control or change (e.g., your father is not genetically your father, surprising facts related to your ancestry, or that someone with your genotype may have a higher than average chance of developing a specific condition or disease). Those outcomes could have social, legal, or economic implications.

The laboratory process may result in errors. A small, unknown fraction of the data generated during the laboratory process may be un-interpretable or incorrect. For more information concerning the accuracy of our process, click here.

You should not change your health behaviors on the basis of this information. For most common diseases, the genes we know about are only responsible for a small fraction of the risk. There may be unknown genes, environmental factors or lifestyle choices that are far more important predictors. If your data indicate that you are not at elevated genetic risk for a particular disease or condition, you should not feel that you are protected. The opposite is also true; if your data indicate you are at an elevated genetic risk for a particular disease or condition, it does not mean you will develop the disease or condition. In either case, if you have concerns or questions about what you learn through 23andMe, you should contact your physician or other appropriate professional.

Genetic research is not comprehensive. Many ethnic groups are not included in genetic studies. Because interpretations provided in our service rely on these published studies, some interpretations may not apply to you.

Future scientific research may change the interpretation of your DNA. In the future, the scientific community may show previous research to be incomplete or inaccurate.

Genetic data you share with others could be used against your interests. You should be careful about sharing your genetic information with others. Currently, very few businesses or insurance companies request genetic information, but this could change in the future. If an employer or insurance company obtained your genetic information through your sharing of it or by legally binding requirements, they could use your genetic data to deny you employment or coverage. Some but not all jurisdictions have laws that protect individuals from this kind of conduct. For some examples of protective legislation and a description of the Genetic Information Nondiscrimination Act proposed in the United States Congress, click here).

Genetic information that you choose to share with your doctor may become part of your medical record and through that route be accessible to health care providers and insurance companies in the future. Genetic information that you share with family, friends or employers may be used against your interests. Even if you share genetic information that has no or limited meaning today, that information could have greater meaning in the future as new discoveries are made. If you are asked by an insurance company whether you have learned genetic information about health conditions and you do not disclose this to them, this may be considered to be fraud.

DATA PROTECTION:

Account Information: When you sign up for our service, 23andMe will collect and store information about you. This information includes contact and payment information and is maintained separately from your genetic data and is not used in genetic research. Account information will be used to transact your purchase and may be used to improve our services and create new features. See our Privacy Statement for further details on data usage and protection.

Genetic Data: The laboratory processing your saliva sample will analyze your DNA to determine your genetic information. The laboratory will not analyze your saliva for any biological or chemical components, markers or agents other than your DNA. The laboratory will not have access to your name or your other personal information. A unique bar code will allow 23andMe to link genetic data derived from your sample to your account. After analysis, your remaining DNA and saliva samples will be destroyed.

Security: 23andMe employs robust security measures to protect against unauthorized access to or unauthorized alteration, disclosure, or destruction of data. Our security program includes regular audits of our security measures and systems. As with any security system, we cannot guarantee that our systems can never be breached.

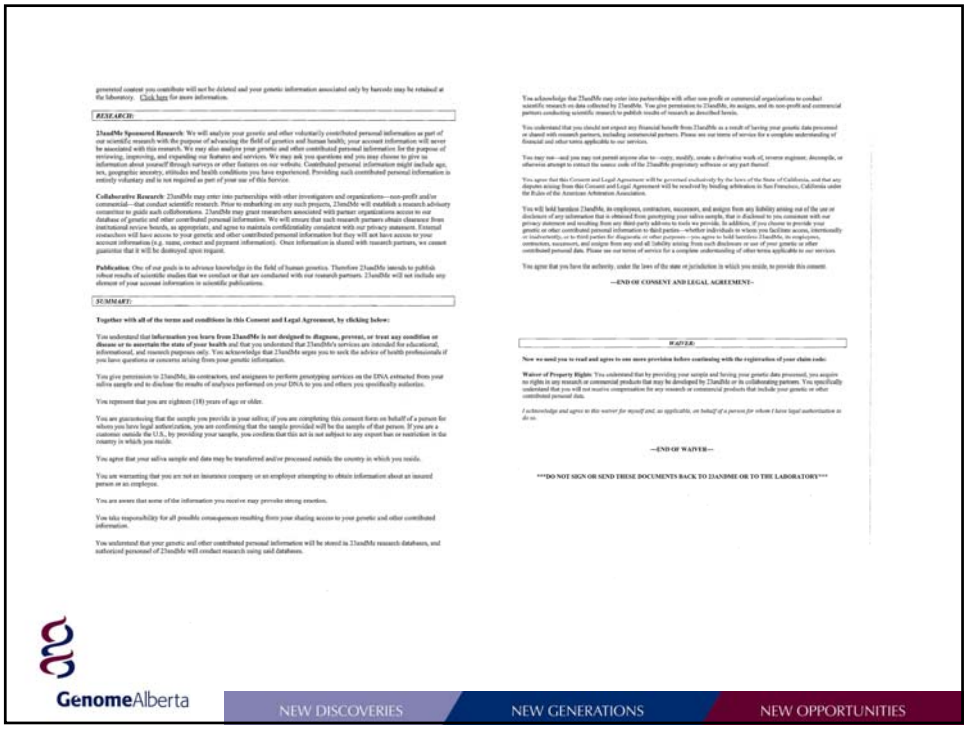
Access: Except as described in this Consent and Legal Agreement, external parties will not be given any of your information without your consent, except as required to comply with legal requirements under applicable laws. Even if we are legally compelled to provide information, we will attempt to notify you before providing your information to external parties unless prohibited by law.

Account Deletion: You have the right to delete your genetic information from our systems. Within thirty (30) days of receiving your written request, we will delete your account, and your information will not be included in any future research, including future research by other organizations. Any research conducted prior to the end of the thirty (30) day period following receipt of your request will not be altered or halted. Once your account is deleted it will not be retrievable. For purposes of clarity, any user-

Consent and Legal Agreement & Waiver

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES



generated content you contribute will not be deleted and your genetic information associated only by barcode may be retained at the laboratory. Click here for more information.

RESEARCH:

23andMe Sponsored Research: We will analyze your genetic and other voluntarily contributed personal information as part of our scientific research with the purpose of advancing the field of genetics and human health; your account information will never be associated with this research. We may also analyze your genetic and other contributed personal information for the purpose of reviewing, improving, and expanding our features and services. We may ask you questions and you may choose to give us information about yourself through surveys or other features on our website. Contributed personal information might include age, sex, geographic ancestry, attitudes and health conditions you have experienced. Providing such contributed personal information is entirely voluntary and is not required as part of your use of this Service.

Collaborative Research: 23andMe may enter into partnerships with other investigators and organizations—non-profit and/or commercial—that conduct scientific research. Prior to embarking on any such projects, 23andMe will establish a research advisory committee to guide such collaborations. 23andMe may grant researchers associated with partner organizations access to our database of genetic and other contributed personal information. We will ensure that such research partners obtain clearance from institutional review boards, as appropriate, and agree to maintain confidentiality consistent with our privacy statement. External researchers will have access to your genetic and other contributed personal information but they will not have access to your account information (e.g. name, contact and payment information). Once information is shared with research partners, we cannot guarantee that it will be destroyed upon request.

Publication: One of our goals is to advance knowledge in the field of human genetics. Therefore 23andMe intends to publish robust results of scientific studies that we conduct or that are conducted with our research partners. 23andMe will not include any element of your account information in scientific publications.

SUMMARY:

Together with all of the terms and conditions in this Consent and Legal Agreement, by clicking below:

You understand that **information you learn from 23andMe is not designed to diagnose, prevent, or treat any condition or disease or to ascertain the state of your health** and that you understand that 23andMe's services are intended for educational, informational, and research purposes only. You acknowledge that 23andMe urges you to seek the advice of health professionals if you have questions or concerns arising from your genetic information.

You give permission to 23andMe, its contractors, and assignees to perform genotyping services on the DNA extracted from your saliva sample and to disclose the results of analyses performed on your DNA to you and others you specifically authorize.

You represent that you are eighteen (18) years of age or older.

You are guaranteeing that the sample you provide is your saliva; if you are completing this consent form on behalf of a person for whom you have legal authorization, you are confirming that the sample provided will be the sample of that person. If you are a customer outside the U.S., by providing your sample, you confirm that this act is not subject to any export ban or restriction in the country in which you reside.

You agree that your saliva sample and data may be transferred and/or processed outside the country in which you reside.

You are warranting that you are not an insurance company or an employer attempting to obtain information about an insured person or an employee.

You are aware that some of the information you receive may provoke strong emotion.

You take responsibility for all possible consequences resulting from your sharing access to your genetic and other contributed information.

You understand that your genetic and other contributed personal information will be stored in 23andMe research databases, and authorized personnel of 23andMe will conduct research using said databases.

You acknowledge that 23andMe may enter into partnerships with other non-profit or commercial organizations to conduct scientific research on data collected by 23andMe. You give permission to 23andMe, its assigns, and its non-profit and commercial partners conducting scientific research to publish results of research as described herein.

You understand that you should not expect any financial benefit from 23andMe as a result of having your genetic data processed or shared with research partners, including commercial partners. Please see our terms of service for a complete understanding of financial and other terms applicable to our services.

You may not—and you may not permit anyone else to—copy, modify, create a derivative work of, reverse engineer, decompile, or otherwise attempt to extract the source code of the 23andMe proprietary software or any part thereof.

You agree that this Consent and Legal Agreement will be governed exclusively by the laws of the State of California, and that any disputes arising from this Consent and Legal Agreement will be resolved by binding arbitration in San Francisco, California under the Rules of the American Arbitration Association.

You will hold harmless 23andMe, its employees, contractors, successors, and assigns from any liability arising out of the use or disclosure of any information that is obtained from genotyping your saliva sample, that is disclosed to you consistent with our privacy statement and resulting from any third-party address to tools we provide. In addition, if you choose to provide your genetic or other contributed personal information to third parties—whether individuals to whom you facilitate access, intentionally or inadvertently, or to third parties for diagnostic or other purposes—you agree to hold harmless 23andMe, its employees, contractors, successors, and assigns from any and all liability arising from such disclosure or use of your genetic or other contributed personal data. Please see our terms of service for a complete understanding of other terms applicable to our services.

You agree that you have the authority, under the laws of the state or jurisdiction in which you reside, to provide this consent.

---END OF CONSENT AND LEGAL AGREEMENT---

WAIVER:

Now we need you to read and agree to one more provision before continuing with the registration of your chain code:

Waiver of Property Rights: You understand that by providing your sample and having your genetic data processed, you acquire no rights in any research or commercial products that may be developed by 23andMe or its collaborating partners. You specifically understand that you will not receive compensation for any research or commercial products that include your genetic or other contributed personal data.

I acknowledge and agree to this waiver for myself and, as applicable, on behalf of a person for whom I have legal authorization to do so.

---END OF WAIVER---

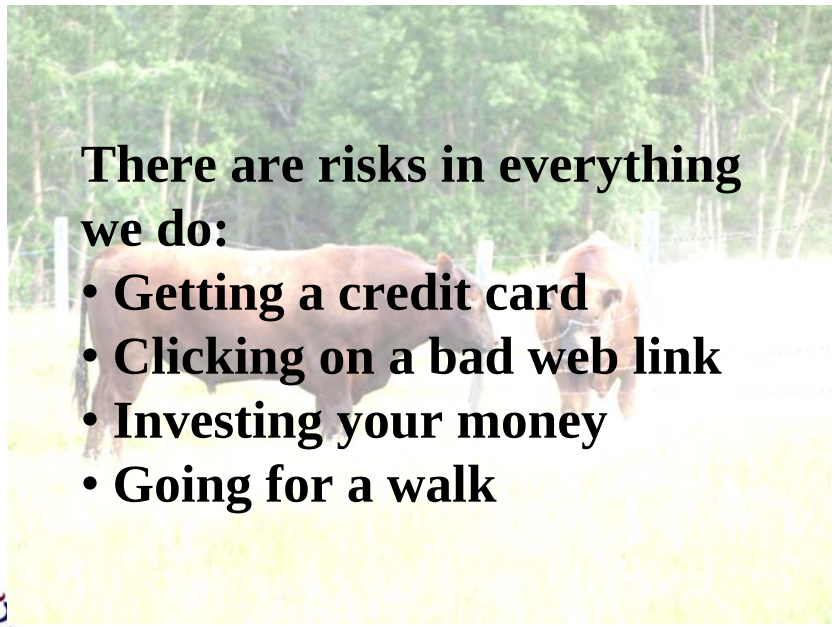
*****DO NOT SIGN OR SEND THESE DOCUMENTS BACK TO 23ANDME OR TO THE LABORATORY*****

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES



GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why take the risk ?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Why take the risk ?

It is my genome.

My meome as Jen McCabe dubbed it on Twitter.

Here is a sample of what the online community has to say:

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

The only problem I could foresee would be being denied health insurance, but since I already have trouble with that due to pre-existing conditions I don't feel it will make a difference. (not holding my breath for Insurance reform here in the US) I don't really worry that anyone here on 23andME would have a nefarious use for my genetic information. If we are sharing, then I also have access to their information. I don't believe anyone would want to clone me or steal my organs.

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have no conserns as I enter a controled sphear by the internet and a huge aray of electronic defenders. Any breaches, I can simply change my e-mail. I think 23andme does a good job. The young lady, Chica (sp) is a great hall monitor and stimulates blog ideas. My wife and I were thrilled at the medical info.

I have no fear at all about sharing.
My information does not need to be protected. I have nothing to hide and I don't care who has access to it.

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I fear that my genetic information will be used against me or my family. More expensive insurance, refusal of insurance, selective service, job discrimination, social exclusion..... Plus I am a private person and I don't think it is anyone else's business.
Strong legal protection for the person and data.
Deidentification.

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

What fears?... you assume, wrongly, that we are afraid.

What protections?... simply ensure that people have the individual control and freedom to share, or not share, as much or as little of their own personal genetic data as they are individually comfortable with. Rephrased: give people the liberty to share whatever they want, but safeguard everything that they choose to keep private

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

From a practical perspective, it's highly unlikely anyone will use my data for malicious purposes. I'm not worried about insurance, but that's because I have faith in our great socialist healthcare system. Plus, I am still young. If 20 years from now, the majority of my peers have been genotyped, do I really have anything to fear from my data being available? We still won't know all there is to know about the human genome, and if any insurance company tried to discriminate against anyone genetically I imagine there would be a huge public outcry. Nonetheless, I hope Canada passes a genetic non-discrimination law sometime (If we have one please correct me). I feel my data is already protected well enough.

Besides, what did it tell me?

Besides, what did it tell me?

- **I 'm going to be bald at an early age**

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- **I 'm going to be bald at an early age**
- **I have a low risk of developing asthma**

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- I 'm going to be bald at an early age
- I have a low risk of developing asthma
- I'm at increased risk for osteoarhritis

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- I 'm going to be bald at an early age
- I have a low risk of developing asthma
- I'm at increased risk for osteoarhritis
- I have sticky ear wax

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- I 'm going to be bald at an early age
- I have a low risk of developing asthma
- I'm at increased risk for osteoarhritis
- I have sticky ear wax
- I have brown hair

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- I 'm going to be bald at an early age
- I have a low risk of developing asthma
- I'm at increased risk for osteoarhritis
- I have sticky ear wax
- I have brown hair
- There is a good chance I'll live to be 100

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- ❑ I 'm going to be bald at an early age
- ❑ I have a low risk of developing asthma
- ❑ I'm at increased risk for osteoarhritis
- ❑ I have sticky ear wax
- ❑ I have brown hair
- ❑ There is a good chance I'll live to be 100
- ❑ I'm at increased risk for ARMD

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Besides, what did it tell me?

- ❑ I 'm going to be bald at an early age
- ❑ I have a low risk of developing asthma
- ❑ I'm at increased risk for osteoarhritis
- ❑ I have sticky ear wax
- ❑ I have brown hair
- ❑ There is a good chance I'll live to be 100
- ❑ I'm at increased risk for ARMD
- ❑ I have an increased risk for Early Onset Alzheimer's

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

If information alone could change people's lives, everybody would be skinny, rich & happy.

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I have a background as a journalist and I work for a genomics organization.

Courtesy Ronald Reagan Library

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

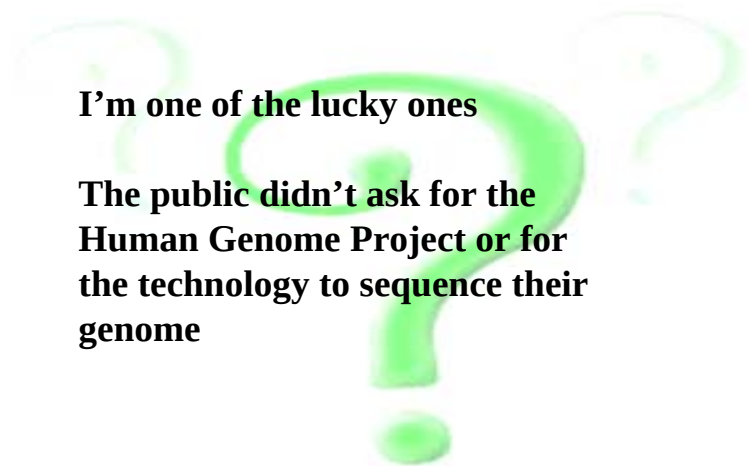
I'm one of the lucky ones

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I'm one of the lucky ones

The public didn't ask for the Human Genome Project or for the technology to sequence their genome



GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I’m one of the lucky ones

The public didn’t ask for the Human Genome Project or for the technology to sequence their genome

Science did that

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I’m one of the lucky ones

The public didn’t ask for the Human Genome Project or for the technology to sequence their genome

Science did that

Science let the Gene-e out of the bottle

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I'm one of the lucky ones

The public didn't ask for the Human Genome Project or for the technology to sequence their genome

Science did that

Science let the Gene-e out of the bottle

GenomeAlberta

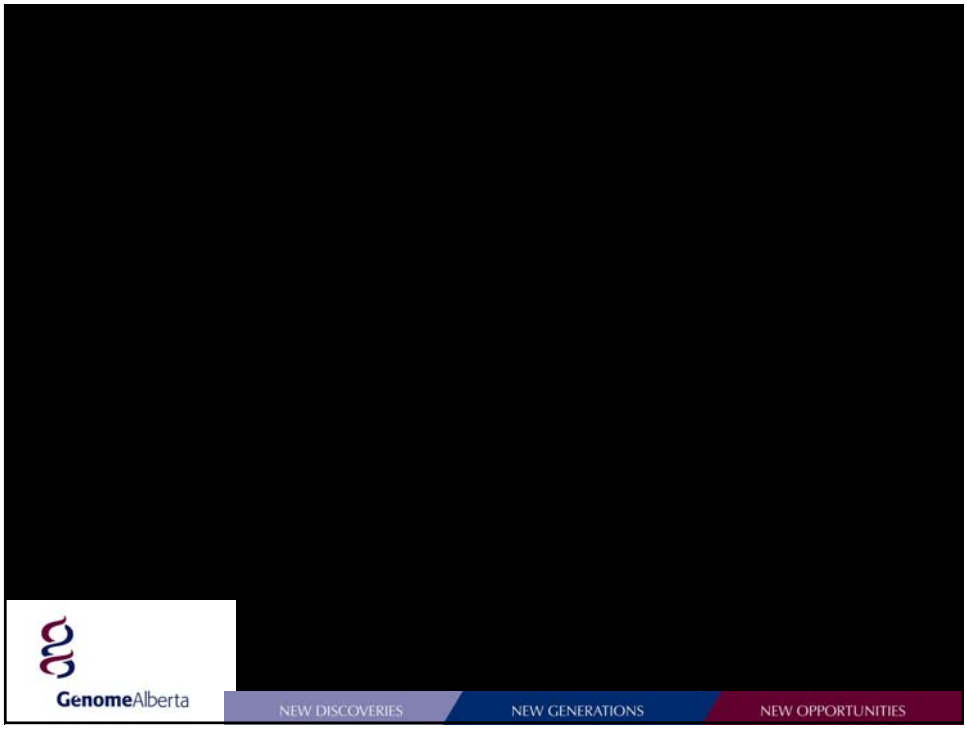
NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

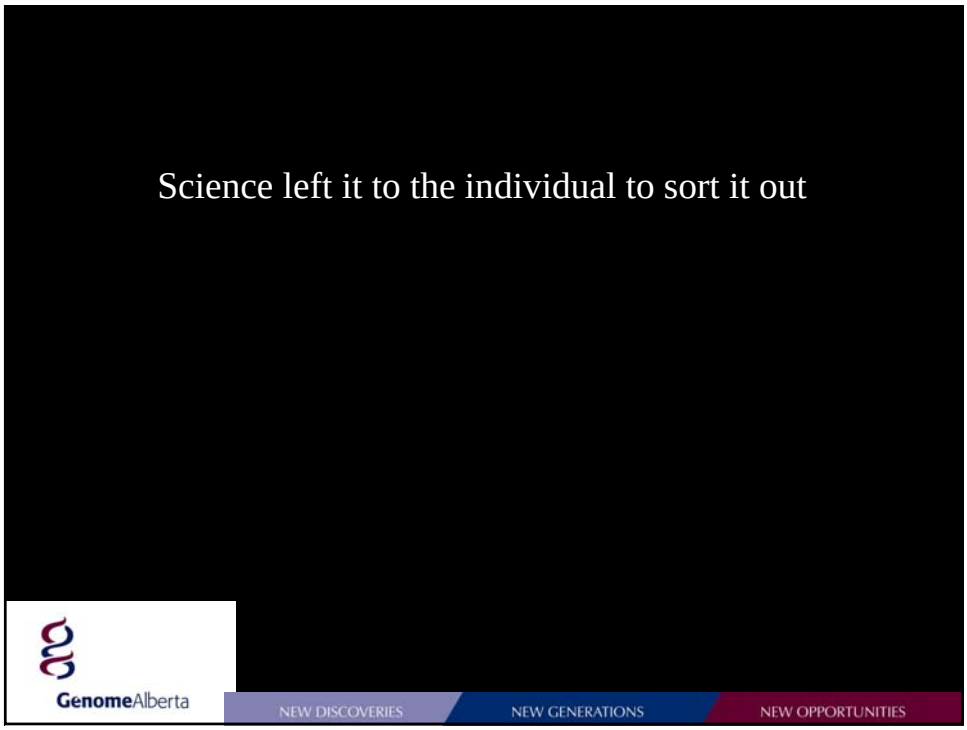
Stem Cell Charter Clip.avi

Alternated clip: FB Stem Cell Charter.mp4

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES


GenomeAlberta
NEW DISCOVERIES
NEW GENERATIONS
NEW OPPORTUNITIES


Science left it to the individual to sort it out
GenomeAlberta
NEW DISCOVERIES
NEW GENERATIONS
NEW OPPORTUNITIES

Science left it to the individual to sort it out

To interpret the ‘meome’

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Science left it to the individual to sort it out

To interpret the ‘meome’

People turned to each other online to get information

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

My data has been downloaded hundreds of times from our website

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

My data has been downloaded hundreds of times from our website

From SNPedia.com

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

My data has been downloaded hundreds of times from our website

From SNPedia.com

With 25 people on 23andME

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

My data has been downloaded hundreds of times from our website

From SNPedia.com

With 25 people on 23andME

And with someone who had my sequence on their laptop in Washington, D.C.

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

What does the online community do with the information other than sharing it?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I shared some of my results with my Gyno, expecting a negative reaction, but she seemed impressed. Because I have high risk for heart disease, high blood pressure, and PAD she did a CRP test on me and seemed hesitant about me taking BC pills. In addition, I don't carry the Breast Cancer variants (which also contribute to Ovarian Cancer) which was my main reason for wanting to go on the pill.
Low risk ovarian cancer+high risk cardiovascular= no bc pills

---

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

I shared some of my results with my Gyno, expecting a negative reaction, but she seemed impressed. Because I have high risk for heart disease, high blood pressure, and PAD she did a CRP test on me and seemed hesitant about me taking BC pills. In addition, I don't carry the Breast Cancer variants (which also contribute to Ovarian Cancer) which was my main reason for wanting to go on the pill.
Low risk ovarian cancer+high risk cardiovascular= no bc pills

---

No behaviour change.
I have discussed hemachromatosis with my doctor.

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

It's one copy of H63D...I'm a carrier. 23andMe says "...no increased risk for iron overload." Still, I don't take vitamins which contain an iron supplement and, as soon as I can, I will again donate blood.

EDIT: Looks like H63d works in concert with ApoE ε4. Does anyone know if 23andMe tests for it?

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

It's one copy of H63D...I'm a carrier. 23andMe says "...no increased risk for iron overload." Still, I don't take vitamins which contain an iron supplement and, as soon as I can, I will again donate blood.

EDIT: Looks like H63d works in concert with ApoE ε4. Does anyone know if 23andMe tests for it?

Please check which variant of the HFE gene you have. The high risk variant for hemochromatosis is two copies of C2982Y. The H63D variant carries a lower risk of hemochromatosis but probably an elevated risk of Alzheimer disease (probably because of increased iron uptake into neurons even with normal peripheral iron levels). 23andMe makes no mention of this anywhere.

I can't say that I've changed any of my habits. We knew that my grandfather suffered from hemochromatosis and that his descendants would also be susceptible. Due to that, I gave blood as often as I could and also let my doctors know. Unfortunately, I developed autoimmune hepatitis and the drug I take to control the symptoms precludes me from giving at this time. Testing through 23andMe indicate that my mother and I are indeed carriers.

Testing also indicated elevated risk to the autoimmune disease rheumatoid arthritis. Related to the hepatitis? I don't know but there's not much I could change as far as habits that would change anything on that front. Just need to make the doctor aware if I ever start having arthritic symptoms.

Other than the RA, 23andMe reports decreased risks for me for Crohn's, type 1 diabetes, celiac disease and age-related macular degeneration. Even the disease risks they categorize as 'typical risk' show slightly lower risks for me and I'm not a carrier for anything but the hemochromatosis.

I brought my 23andme results to my physician. He suggested I start taking a supplement against macular degeneration and as I am a carrier for hemacromatosis that I come in for blood tests every year or so, to check my blood iron levels.

I am also far more aware of my risk of developing type two diabetes and as a consequence I now eat far fewer sweets than I used to.

So yes, it has changed my behavior.

I got my results a month ago, but I haven't showed them to my doctor yet (mostly because my family physician is not easy to get to from university). I expect when I do she will have some recommendations for me.

I am at pretty high risk of macular degeneration, but I have't done anything about it yet. I would prefer to talk to said doctor about it. I am also heterzygous for factor 5 leiden, which isn't something I can change. I don't smoke or take birth control, so I don't think it's that big of a deal.

I am also at increased risk for heart attack and other heart-related conditons, as well as lower HDL and higher LDL; it's not surprising given my dad has high cholesterol and had a heart-attack. While I have tried to change my diet and increased my exercise in the past, I tend to get bored easily. I think it will take awhile but I will eventually improve my habits. I certainly don't want to end up like my dad.

The 23andMe scan has given me tons of information, and I am now incorporating that in many ways.

- I had known for a long time that I am a slow metabolizer for caffeine. Now I have it confirmed genetically. I watch my caffeine intake even more carefully than before. My sleep quality has improved.

- I am a carrier for the HFE variant H63D. While I am not at risk of hemochromatosis, I am at risk for AD (this is currently not reflected in 23andMe's output but I suggested it to them). I will be monitoring my ferritin levels closely. Note: I am a neuroscientist and the area of neurodegeneration is actually my field.

- I am heterozygous for the lactose intolerance variant (as are both of my parents). We have long suspected that lactose intake could be related to some of our problems, but thought this was unlikely because of our northern European heritage. Now with this result, it's more plausible. I have switched from milk/cheese to soy. It's more climate friendly anyway.

Yay .. im negative =) other than common polymorphisms of no known clinical significance.
The waiting has been hard , really hard ..it's such a relief :)
good luck to everyone in this group.. gentle ((hugs))

GenomeAlberta
NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

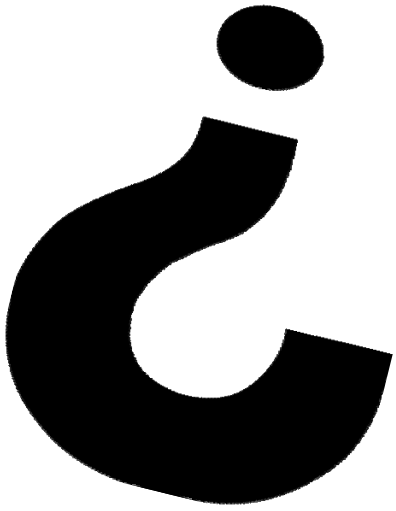
Yay .. im negative =) other than common polymorphisms of no known clinical significance.
The waiting has been hard , really hard ..it's such a relief :)
good luck to everyone in this group.. gentle ((hugs))

I tested negative for both, but had the surgeries anyway. My mother, grandmother and numerous aunts have been diagnosed with the disease and died from it. I decided that it was too risky even with a negative result. In November 2008, I had a double mastectomy with tram flap reconstruction, and hysterectomy with oophrec...tomy. I feel great and know that this is not a decision for everyone. The fear was overpowering me and by doing this I feel that I made the best decision for myself.

How do people choose who to trust ?

Lots of options but from a Communications perspective this tends to hold pretty solid over time:

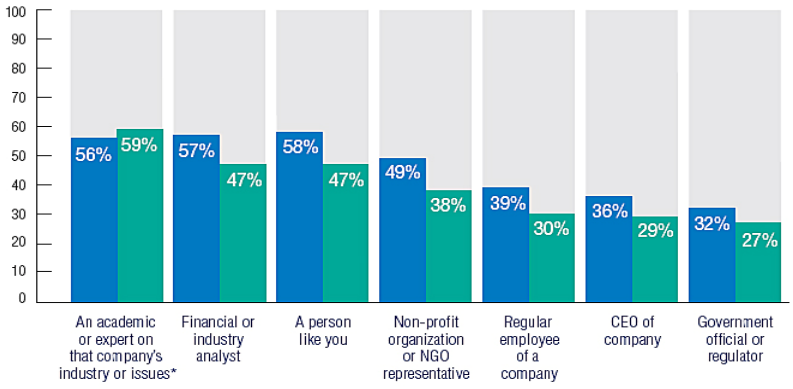
QUESTIONS



GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

If you heard information about a company from each of these people, how credible would it be?



*Asked as "an academic" in 2008
Informed publics ages 35 to 64 in 18 countries
"Extremely credible" and "very credible" responses only

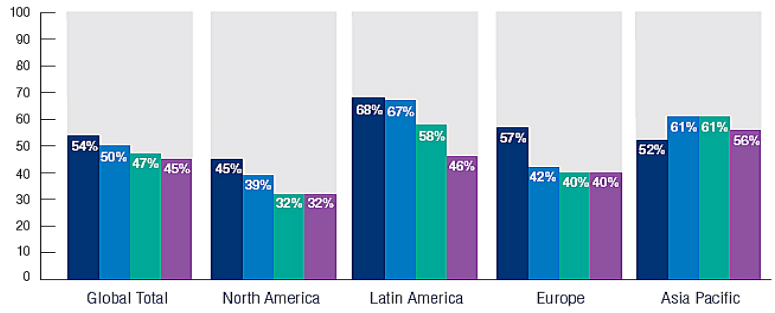
from 2009 Edelman Trust Barometer

GenomeAlberta

NEW DISCOVERIES NEW GENERATIONS NEW OPPORTUNITIES

Figure 8: NGOs most trusted institution in every region except Asia Pacific

How much do you trust each institution to do what is right?



Informed publics ages 35 to 64 in 20 countries
Responses 6-9 only on 1-9 scale; 9 = highest

NGOs Business Media Government

from 2009 Edelman Trust Barometer

Online Direct to Consumer Advertising for Genetic Testing: An examination of credibility markers, consent and privacy provisions

Rose Geransar, University of Calgary & Farah Mohamed, University of Alberta

Abstract:

Rose Geransar and Farah Mohamed will present the following study conducted in conjunction with Edna Einsiedel of the University of Calgary.
Findings. Two strategies were most frequently used by companies to frame risk: underlining the basis of the condition, often with genetic determinist and essentialist undertones, and stressing the commonality of the conditions. Major credibility and trust markers employed were indications of organizational professional accreditation/ recognition and credentials of company executives and staff. The company websites provided limited, vague or misleading information about disease etiology and promoted tests for use in broader at-risk populations than is normally indicated in clinical practice. Available consent forms were varied in the elements of consent that they covered, and were available on only one third of the websites examined. Privacy policies were more widely available, but varied tremendously in both the scope and depth of their content. Implications of these trends for Canadian consumers and clinicians are discussed.
Companies engaging in online direct-to-consumer advertising (DTCA) for genetic testing are continuing to expand and specialize in the types of tests they offer, and are developing more sophisticated websites for communicating with consumers. Because of the long-distance nature of the communicative transactions involved in the provision of services, the communication and handling of issues pertaining to establishing trust and credibility, protecting consumer privacy and obtaining consent are of particular interest. This presentation will summarize the findings of two key studies using samples of companies engaged in internet direct-to-consumer (DTC) advertising for genetic testing. The studies pertain to: 1) the way in which genetic risk information is framed to consumers, including strategies to establish trust and credibility in this context, 2) the information content in the companies' online privacy statements and consent forms.
Methods. Key words specific to genetic test DTC advertising were entered into popular internet search engines to arrive at the respective samples of companies. Representations of benefits and risks on company websites were coded and themes were developed across advertisements. Available consent forms and privacy policies were coded and analyzed for themes.

Bios:

Rose Geransar has a B.Sc. in Biochemistry and is currently a Ph.D. Candidate in the Department of Community Health Sciences at the University of Calgary and Office of Medical Bioethics. Her dissertation is in the area of consent as part of the broader framework of governance in public umbilical cord blood banking, funded by CIHR. She is an active member of the Canadian Bioethics Society and a part of the Genome Canada community of researchers in the area of genomics-related ethical, economic, environmental, economic and social issues (GE3LS). She was the recipient of the 2008 Douglas Kinsella Award for Research in Bioethics.

Farah Mohamed recently completed a Bachelor of Health Sciences (B.HSc. Hons) at the University of Calgary, and is currently studying law at the University of Alberta. She has been involved in research in rehabilitation medicine, mental health, and direct-to-consumer genetic testing.

Link to video of this presentation.

Online Direct-to-Consumer Marketing for Genetic Testing

Credibility Markers, Consent and Privacy Provisions

Edna E. Einsiedel[1], Ph.D., **Rose M. Geransar**[2], B.Sc., Ph.D. Candidate, **Farah Mohamed**[3], B.HSc., LLB. Student

1 Faculty of Communication & Culture; 2 Faculty of Medicine, Department of Community Health Sciences, University of Calgary;
3 Faculty of Law, University of Alberta

November 19, 2009

Direct-to-Consumer Advertising

» "If you have a family history of Alzheimer's, this test will determine your susceptibility."

"Discover how your genes can hold the secret to your well-being."

» "It's hard to find a woman who doesn't worry that one day she will get breast cancer."

2

Direct-to-consumer marketing of genetic tests: risk-benefit frames and credibility markers

GeneLink

Sciona

DNAdirect

GeneDx
DNA DIAGNOSTIC EXPERTS

Put Your Genes to the Test!
mycell
the science of you™

GENOVATIONS™
Predictive Genomics for Personalized Medicine

3

Research Questions (2007)

- How are the clinical value and limitations of genetic tests being framed?
- What risk frames are employed?
- What strategies are used to establish trust and credibility in this context?

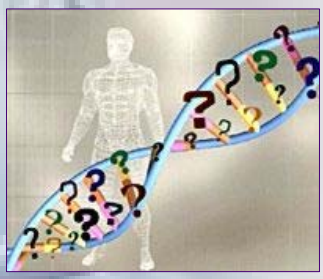
4

Research Questions (2009)

- How do companies engaging in DTC advertising for genetic testing protect the privacy of the health information they collect and the genetic information they generate?
- How are the various elements of consent addressed by these companies?

Methods

- **Sampling**
  - Non-probability sampling
  - Popular search engines (key terms), health websites
  - Simulation of a consumer search process
- **Data collection/ Analysis**
  - Text: Coding, themes
  - Categorical data (nominal)

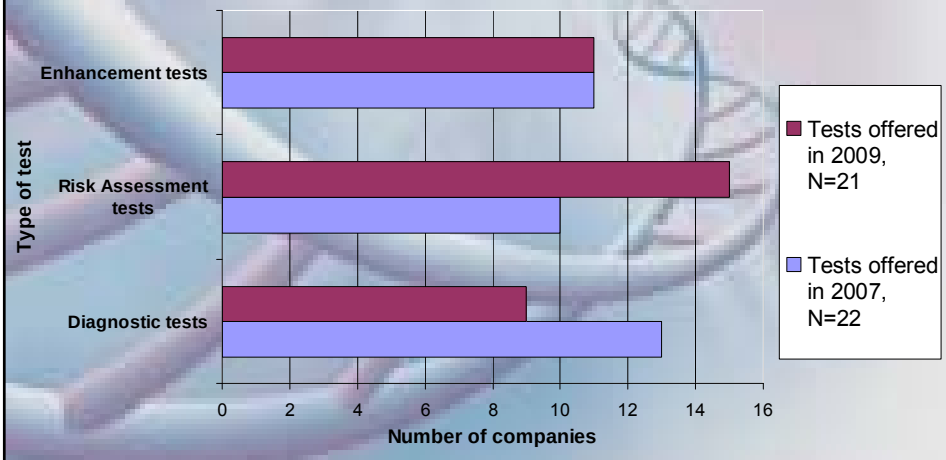


6

Categories of Genetic Tests

Genetic Test Category	Type of gene(s)	Disease causality	Examples of diseases/ genes for which genetic testing is offered
Diagnostic	Single high penetrance gene (one or more alleles)	Single- gene (one or more alleles)	Tay Sachs disease (HEXA) Cystic Fibrosis (CFTR) Huntington's Disease (HD)
Risk Assessment	Single moderate-to-low penetrance gene (one or more alleles)	Multi-gene Multi-factorial	Hereditary breast cancer (BRCA) Early onset Alzheimer's disease (APOE)
Enhancement	Many low penetrance genes	Multi-gene Multi-factorial	Cardiac health profile, Nutrigenetic tests (many genes)

2007-2009: Tests Offered by Sample of Companies



8

MAPPING RISK

- **Risk frames**
  1. Who's at risk
  2. Why know your risk?
  3. How is risk assessment made?

9

1. Who's at risk?

Strategy	Example Claims
Emphasis on genetic basis of condition	*"All 19 genes analyzed influence these five areas of health"*
Emphasis on condition being common	*"Hearing loss is one of the most frequent hereditary defects"*

10

2. Why is it important to know your risk?

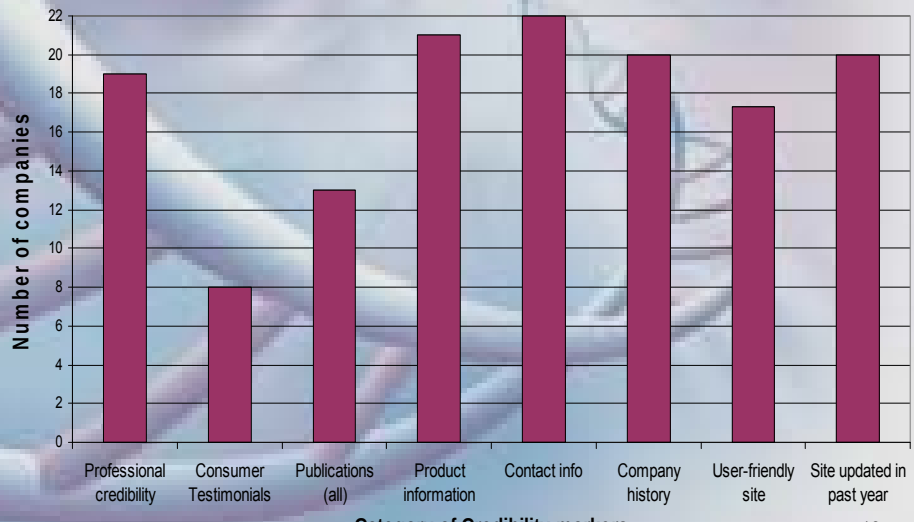
Strategy	Example Claims
Need to resolve problem; consequences of delay	*"You can use prevention strategies earlier in life"*
Backed by scientific research	*"Test X looks for a unique genetic marker that has been validated by independent researchers in studies of tens of thousands of people all around the world"*
Importance of taking control	*"Choose to know, take control!"* *"Learning more may reduce your anxiety"*

11

3. How is the assessment made?

Strategy	Example Claims
Emphasis on simplicity of approach	*"Safe, simple, convenient"* *"Fast and painless"*
Specialized risk assessment tool	*"Enables accurate and definitive diagnosis of many conditions"*
Individualized, tailored outcomes	*"Once your diet, lifestyle and genes have been analyzed, we'll send you a confidential personalized Action Plan of up to 100 pages, telling you how to match your diet and lifestyle to your genes. "*
Pros and cons of test	"Pro: a positive test would motivate you to take preventive measures Con: You are not interested in learning if you have a genetic risk for X"

Trust & Credibility Markers



13

Consent and Privacy Aspects (2009)

Questions:

- How do these companies protect the privacy of the health information they collect and the genetic information they generate?
  - Examination of privacy policies
- How are the various elements of consent addressed by these companies?
  - Examination of consent forms

Examined a modified sample of 21 companies

- 6 with **consent forms**
- 19 with **privacy policies**

(2 had neither, 5 had both)

Elements of consent (Herbst and Merz, 2005)	Number of companies that addressed element (Out of 6)		
	In consent form	In privacy policy	Other part of website
How is my genetic material collected?	4	0	3
How is my blood/ skin/ tissue tested?	2	2	4
What will this test tell me?	5	0	2
What are the risks of this test?	5	0	0
What alternatives are there to taking this test?	1	0	0
How will I possibly benefit from this test?	3	0	2
How will I learn the results of this test?	4	0	1
How reliable are these results?	2	0	0
How will this information about me be kept private?	2	2	1
Who can have access to my results?	5	2	2
To whom will my results be disclosed?	3	4	2
What will happen to my sample after the test is completed?	3	2	0
Will my sample be used for anything other than this diagnostic genetic test (e.g. research)?	1	1	0
Who do I contact if I have questions or concerns about this test?	1	1	4

Herbst, J.L., and Merz, J.F. (2005). Ethical, social, and legal issues related to molecular genetic testing. In W.B. Coleman and G.J. Tsongalis (Eds), "Molecular Diagnostics: For the Clinical Laboratorian. Second Edition. Humana Press. Pp. 545-554.

Consent practices (2009)

What was consent form for?	# of elements of consent addressed (out of 14)	Physician requisition
• Autism diagnostic genetic test	7	Yes
• Nutrition analysis- oxidative stress susceptibility*	10	No
• Cancer drug pharmacogenetic test	10	Yes
• Bipolor disorder and depression (combined genetic test)	9	Yes
• General consent form for molecular genetic test (1)	10	No
• General consent form for molecular genetic test (2)	12	Yes

Commonly mentioned basics

Elements of Consent	Example Quotes from Consent Forms
How the testing is done	*"The test configuration is based on a detailed analysis of the scientific literature. The test looks for DNA variations called single nucleotide polymorphisms (SNPs) in five genes [..]" (C11)*
Purpose of genetic test	*"The purpose of this molecular genetic test is to ascertain if I am, my child is, or my unborn child is [please circle appropriate] carrying mutation(s) predisposing to or causing the specific disease or condition [X]". (C1)*
How you will learn the results	*"The test results will be reported to you by mailing them to an address that you designate or by providing you with access to a secure website (Web results not available in French in Canada.)."*

Rarely mentioned..

Elements of Consent	Example Quotes from Consent Forms
Alternative options	*"This test is not the only way to look for genetic changes, and my physician may recommend this test before or after doing other genetic tests." (C5)*
Reliability of test results	*"I understand that the molecular genetic test may not generate accurate results for the following reasons: sample mix-up, samples unavailable from critical family members, maternal contamination of prenatal samples, inaccurate reporting of family relationships, or technical problems, but not limited to these." (C1)*
Secondary uses of sample	*"After testing is completed, I understand that my blood, body fluid or tissue specimens may be disposed of or retained indefinitely for research, test validation, and/or education by [Company], as long as my privacy is maintained. I understand that no compensation will be given nor will funds be forthcoming due to any invention(s) resulting from research and development using the specimens submitted." (C1)*

Disclosures & Disclaimers that were not made in ads..

Risks	Example Quotes from Consent Forms
Emotional risks	*"You may learn information about yourself that you do not anticipate. This information may evoke strong emotions and has the potential to alter your life and worldview. You may discover things about yourself that trouble you and that you may not have the ability to control or change [..] These outcomes could have social, legal, or economic implications." (C20)*
Possibility of discrimination	*"Genetic testing may expose you to risk of discrimination by health insurance companies, making it more difficult for you to be insured."*

Disclosures & Disclaimers that were not made in ads..

Limitations	Example Quotes from Consent Forms
Communication of uncertainty	*"This is NOT a DIAGNOSTIC TEST. It is a RISK ASSESSMENT TEST. Persons who learn they are positive for variations in one or more of these genes may never experience any discernible harm to their health because of that variation.* *The [condition being tested for] is influenced by many other genetic and environmental factors. It is not possible to estimate the effect of variants in these six genes on overall health." [C11]*
	"As of this date, these gene tests have been validated only in Caucasians of European ancestry. Their meaning and interpretation in other racial and ethnic groups is unclear." [C21]
"Not a medical service"	*"In deciding to take this test, you understand that neither [Company C11] nor any of its staff are agreeing to provide a medical service or offering to render medical care or advice of any kind. You may wish to obtain professional genetic counseling or consult with your health care professional before signing this consent form."*

Companies with Privacy Policies (19)

Theme	Sub-theme	# of Companies
Mention of Legislation	**Specific Legislation**	5
	Reference to General Guidelines	3
Information Collected	Personally Identifiable (Name and Contact Information)	10
	Personal Health Information	3
	Genetic Information	**3**
	Updates to or Deleting Information	9
Use of Information (Client-related purposes)	To perform client-requested services	10
	Communicate with client regarding services	9
	Communicate to client services that may be of interest	8
Disclosures to Other Parties	Parties involved in providing care	6
	Agents (or contractors) providing services to company	9
	Third-parties (with consent)	13
	When required by law	9
Business	**Transfer of Assets or Merger**	4
	Internal Business Operations	4
Legal	Underage policy	7
	Parents role on behalf of under-aged children	3
Security	Network and data protection (e.g. firewall, encryption, backup systems etc.)	9
	E-mail/Internet Caveat	8
	Third-party ("hacking") Caveat	4
	Separating PHI and Personally Identifiable	2
	Personnel	9

What personal information do companies collect?

Health-specific information in the form of ...

- personal health information (3)

"we may collect Phenotypic Information (disease conditions and personal traits) if you choose to participate"

- genetic information (3)

"When you sign up for our service, [our company] 2 collects and stores personal information about you, including ... Genetic Information (the As, Ts, Cs, and Gs at particular locations in your genome)."

How is this information used with respect to business purposes?

- Health-related information is considered by companies to be a type of 'asset'
- How will information be handled in the event that they merge with another company? (4)

"In the event that [the company] goes through a business transition such as a merger ... your personal information and non-personal information will likely be among the assets transferred"

Trends/ Conclusions

- Stable-to-increasing demand for web-based access to genetic services
- Tests with lower clinical validity are more likely:
  1. To be advertised through the use of emotive techniques
  2. Not to require physician mediation
  3. To be offered with long-distance or no genetic counseling.
- Strategies used to build trust focus on genetic testing as a credence good: scientific publications, professional accreditation

24

Trends/ Conclusions

- While ads often (over)emphasized the benefits of the tests, consent form focused more heavily on risks and limitations that were not mentioned on other parts of websites
- Many elements of consent not covered
- Very few companies provided online consent forms on their sites; actual consent practices need to explored
- Most privacy policies and consent forms often did not indicate secondary uses of sample, mechanisms by which confidentiality of health information is protected, reliability of test results, and provisions made in case of company mergers/ transfer of assets.

Acknowledgements

This study was carried out as part of a Genomics, Ethics, Environment, Economics, Law and Society (GE3LS) project on Knowledge Translation in Health Systems funded by Genome Canada (E.F. Einsiedel, PI).

26

Relevant publications

- Geransar, R.M. and E.F. Einsiedel. (2008). Evaluating online direct-to-consumer marketing of genetic tests: informed choice or buyer beware? *Genetic testing*, 12:1, 13-23.

- Einsiedel, E. and R.M. Geransar. (2009). Framing genetic risk: Trust and credibility markers in online direct-to-consumer advertising for genetic testing. *New Genetics and Society*; 28(4): 339-362.

27

Psychological Impacts of Pre-dispositional Genetic Testing: Possible lessons for direct to consumer advertising

Brenda Wilson, University of Ottawa

Abstract:

There are a range of potential benefits from knowing one's genetic predisposition for common disorders, most notably the possibility for improving health outcomes by reducing disease risk and detecting disease early enough for effective intervention. Achieving these outcomes often requires changes in health behaviour. This talk will focus on the current evidence of the effect of genetic testing for predisposition to common adult onset conditions on emotional state, personal risk perception, and health behaviour. It will examine the implications of these findings for DTC marketing of genetic tests, including the issue of offering tests without the requirement for preliminary genetic counselling.

Bio:

Brenda Wilson, M.B., Ch.,B., F.F.P.H., trained as a physician at the University of Edinburgh, and as a public health physician at the University of Newcastle-Upon-Tyne. She is an Associate Professor in Epidemiology & Community Medicine at the University of Ottawa, and conducts public health and health services research relating to genetics. Her research has spanned a range of issues, including genetics education and knowledge tools for non-genetics professionals, the impacts of genetic testing for late onset disorders, outcome measures for genetics health services, family communication and disclosure of genetic information, and the integration of ethical, legal and social issues into genetics technology assessment processes. Her most recent work investigates the empirical value of family health history in complex disease risk prediction, and lay and professional reactions to the (hypothetical) extension of genomic profiling into public health screening programs.

Link to video of this presentation.

Psychological impacts of pre-dispositional genetic testing: possible lessons for direct to consumer advertising

Electronic Health Information and Privacy Conference
November 19, 2009

Brenda J Wilson, MB ChB MRCP(UK), FFPH

uOttawa
L'Université canadienne
Canada's university

Acknowledgements

Crystal Palleschi (nee Dunn), Heather Howley,
Jodi Heshka, Phil Wells

Research Funding Support:
CIHR Interdisciplinary Capacity Enhancement Team Grant
Heart & Stroke Foundation of Ontario Program Grant
Canada Research Chairs Program

Genetic testing has changed radically in 20 years

Type of genetic variation	Examples	Practice model
Single gene disorders, high penetrance, no effective interventions	**Huntington Disease**	**Genetic services, non-directive counselling**
Single gene disorders, high penetrance, effective interventions	**Phenyl-ketonuria**	**Population screening**
Single gene disorders, low or variable penetrance, interventions variables	Hereditary breast/ovarian cancer	Genetic services, counselling may or may not be directive
Genetic variation at one locus or multiple loci	Factor V Leiden; pharmaco-genetic traits	Communication of genetic information regarding future risk of disease and interventions, counselling may be directive

Khoury MJ. Genet Med 2003; 5: 261-8.

Type of genetic variation	Examples	Practice model
Single gene disorders, high penetrance, no effective interventions	Huntington Disease	Genetic services, non-directive counselling
Single gene disorders, high penetrance, effective interventions	Phenyl-ketonuria	Population screening
Single gene disorders, low or variable penetrance, interventions variables	Hereditary breast/ovarian cancer	Genetic services, counselling may or may not be directive
Genetic variation at one locus or multiple loci	Factor V Leiden; pharmaco-genetic traits	Communication of genetic information regarding future risk of disease and interventions, counselling may be directive

Khoury MJ. Genet Med 2003; 5: 261-8.

Classical genetic testing:

- Rare disorders
- High certainty, predictive information
- Usually no effective interventions

Predispositional genetic testing:

- Commoner disorders
- Lower certainty, less predictive information
- (Sometimes) possibility to reduce risk

How does disease predisposition information affect

- Perception of personal risk?
- Emotional well-being?
- Health-related behaviour?
- Use of health services?

Available evidence

1. Systematic reviews of genetic testing
2. Analysis of pilot data on example condition
3. REVEAL Study

Six reviews to January 2008

Study	Intervention reviewed	Condition
Broadstock 2000 11 studies (1990-98)	Predictive testing	Mainly Huntington's Disease
Meiser 2002 12 studies (1980-2000)	Genetic counselling	Breast cancer
Butow 2003 19 studies (1980-2001)	Genetic counselling and testing	Breast cancer
Braithwaite 2004 21 studies (1980-2001)	Genetic counselling	Familial cancer
Wainberg 2004 7 studies (1996-2003)	Surveillance & surgery	BRCA mutation carriers
Heshka 2008 30 studies (2000-2006)	Predispositional testing	Familial cancer Alzheimer disease

Reasons for Seeking Genetic Susceptibility Testing Among First-Degree Relatives of People With Alzheimer Disease

*J. Scott Roberts, ‡Susan A. LaRusse, ‡Heather Katzen, §‖Peter J. Whitehouse, §Melissa Barber, ‖Stephen G. Post, ‡Norman Relkin, ¶Kimberly Quaid, *Robert H. Pietrzak, **L. Adrienne Cupples, *†**†††Lindsay A. Farrer, *†Tamsen Brown, and *†Robert C. Green

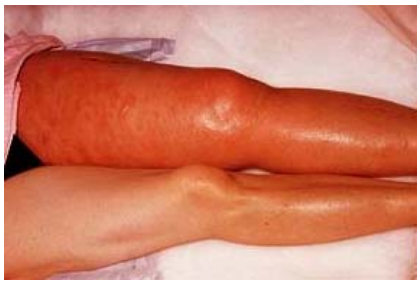
Alzheimer Disease and Associated Disorders, 2003; 17: 86-93

REVEAL Study: randomized controlled trial of APOE genetic testing for Alzheimer's Disease (AD)

- Boston, NYC, Cleveland
- Adult children of person with confirmed AD
- Self-referred and systematically contacted

An exploratory study of the psychological and behavioural impacts of genetic testing for thrombophilia among asymptomatic first-degree relatives of patients with venous thrombosis

Dunn, C. MSc Thesis, University of Ottawa, 2006



Thromboembolic disease

- Relatively frequent
- Significant long term morbidity
- Significant risk of death
- Prevention depends on avoidance of risk factors and acting on early symptoms

An exploratory study of the psychological and behavioural impacts of genetic testing for thrombophilia among asymptomatic first-degree relatives of patients with venous thrombosis

Dunn, C. MSc Thesis, University of Ottawa, 2006

Not a classical genetic disease,
but some gene variants confer increased risk

- 111 people tested – 57 carriers, 54 non-carriers
- Aged 21-78 years
- Psychological measures pre-test, one week and 12 months after test result
- Self-report health behaviour and contact with health services assessed at 12 months

How does predispositional testing affect perceptions of disease risk?

Systematic reviews:

- Accuracy of risk perception improves compared with pre-testing perceptions, but there is a persistent tendency to over-estimate risk

How does predispositional testing affect perceptions of disease risk?

Predictive Genetic Testing for Alzheimer's Disease: Impact upon Risk Perception

Theresa M. Marteau,[1]* Scott Roberts,[2] Susan LaRusse,[3,5] and Robert C. Green[4]

Risk Analysis, 2005; 25:397-404.

REVEAL Study

Control group

counselling, individual risk calculated according to gender and family history only

Intervention group

as above with addition of actual genetic test result

- *Test positive* should have **higher** risk perception than control group
- *Test negative* should have **similar** risk perception to control group

Table II. Cumulative Lifetime Risk (Mean (*SD*, Range)), Perceived Personal Risk (Mean (*SD*)) (a) Unadjusted for Lifetime Risk and (b) Adjusted for Lifetime Risk, Age, and Number of Affected Relatives, by APOE ε4 Status

	Cumulative Lifetime Risk	Perceived Personal Risk	
		(a) Unadjusted	(b) Adjusted
APOE ε4+	47.8 (9.6, 25–57)	3.9 (0.7)	3.4 (0.7)
APOE ε4–	24.1 (5.5, 18–29)	2.9 (0.8)	3.1 (0.8)
Controls	26.8 (4.4, 18–29)	3.2 (0.7)	3.4 (0.7)

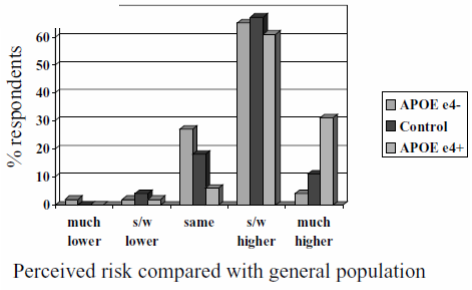


Fig. 2. Perceived personal risk compared with those without a family history of AD.

How does predispositional testing affect emotional well-being?

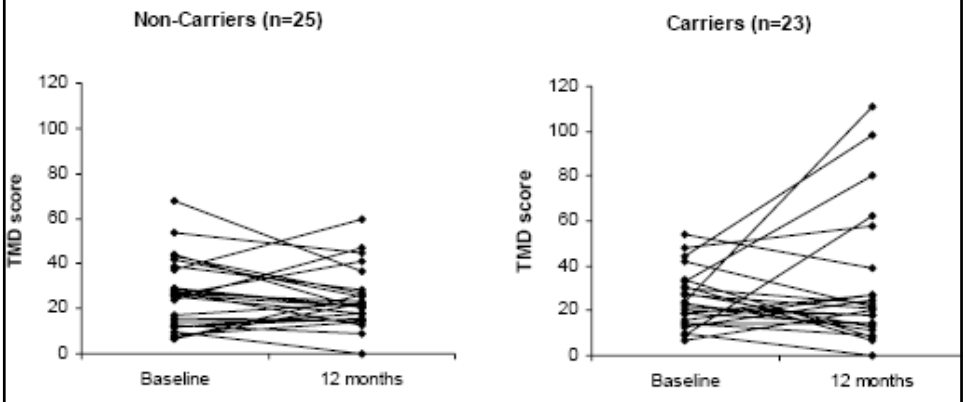
Systematic reviews:

- Generally low levels of psychological morbidity
- Negative test results (no mutation) → relief, positive test results → short term increase in distress, returns to baseline within weeks-months
- ***Patient characteristics***, not test results, generally predict long term psychological outcomes for individual patients

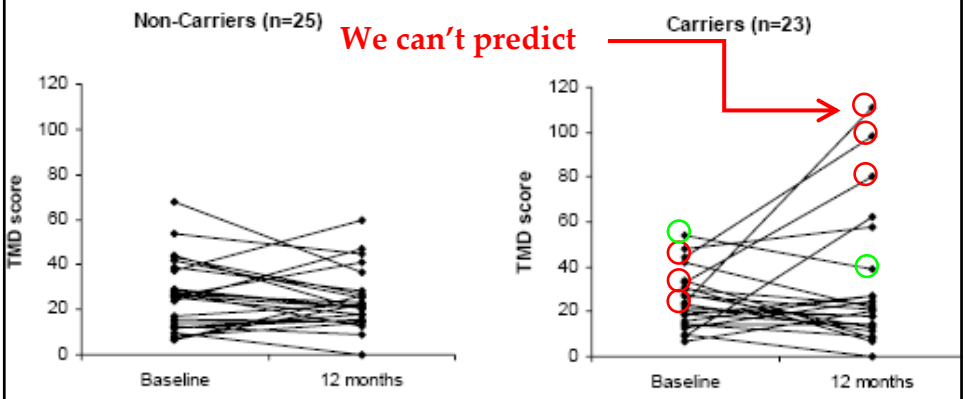
How does predispositional testing affect emotional well-being?

Thrombophilia study:

Total Mood Disturbance



How does predispositional testing affect emotional well-being?



How does predispositional testing affect health related behaviour?

Systematic reviews:

- Uptake of recommended surveillance - unclear effects
- Uptake of chemoprophylaxis – generally low
- Uptake of risk-reduction surgery – highly variable

All dependent on local protocols

Health Behavior Changes After Genetic Risk Assessment for Alzheimer Disease: The REVEAL Study

Serena Chao, MD, MSc, J. Scott Roberts, PhD,† Theresa M. Marteau, PhD,‡ Rebecca Silliman, MD, PhD,*§ L. Adrienne Cupples, PhD,‖ and Robert C. Green, MD, MPH¶#§*

Alzheimer Disease and Associated Disorders, 2008; 22: 94-7.

Participants who learned they were e4 positive were significantly more likely than e4 negative participants to report AD-specific health behavior change 1 year after disclosure (adjusted odds ratio: 2.73; 95% confidence interval: 1.14, 6.54; P=0.02). Post hoc analyses revealed similar significant associations between numerical lifetime risk estimates and self-report of AD-specific health behavior change. Despite lack of preventive measures for AD, knowledge of APOE genotype, numerical lifetime risk, or both, influences health behavior.

How does predispositional testing affect use/expectations of health services?

Thrombophilia study:

		Yes	No	Don't know	p
Visited doctor more often since test	*Carriers*	2 (4.9)	39 (95.1)	0	0.60
	Non-carriers	1 (4.3)	22 (95.7)	0	
Discussed test result with doctor	*Carriers*	32 (78.0)	9 (22.0)	0	0.71
	Non-carriers	11 (47.8)	12 (52.2)	0	
Doctor understands test result	*Carriers*	26 (81.3)	3 (9.4)	3 (9.4)	0.013
	Non-carriers	10 (83.3)	0	2 (16.7)	
Doctor gave advice about risk	*Carriers*	18 (56.3)	12 (37.5)	2 (6.3)	0.157
	Non-carriers	3 (25.0)	7 (58.3)	2 (16.7)	

Overview of knowledge to date

- For most people, genetic tests appear not to have long term negative psychological impact
- However, some people have persistent distress
- Risk perception may not be accurate (over, under)
- Unclear how testing affects subsequent health-related behaviour
- Unclear impacts of testing on health services
- Some evidence that patients expect primary care doctors to know more about genetic tests

Application to DTC genetic tests?

- Is the generally reassuring lack of lasting emotional impact dependent on pre-test counselling?
- What is the likely harm of failing to identify those at high emotional risk beforehand?
- Is there potential for harm through simply being aware that tests are available?

Who seeks genetic testing?

Who seeks genetic susceptibility testing for Alzheimer's disease? Findings from a multisite, randomized clinical trial

J. Scott Roberts, PhD[1], Melissa Barber, ScM[3], Tamsen M. Brown, MS[1,2], L. Adrienne Cupples, PhD[4], Lindsay A. Farrer, PhD[1,2,4,5], Susan A. LaRusse, MS[6], Stephen G. Post, PhD[7], Kimberly A. Quaid, PhD[8], Lisa D. Ravdin, PhD[9], Norman R. Relkin, MD, PhD[9], A. Dessa Sadovnick, PhD[10], Peter J. Whitehouse, MD, PhD[3], John L. Woodard, PhD[11], and Robert C. Green, MD, MPH[1,2,5]

Genetics in Medicine, 2004; 6: 197-203

Of 196 systematically contacted participants, 47, or 24%, progressed from initial contact to RCT enrollment. These participants were more likely to be below age 60 (adjusted OR 3.83, P 0.01) and college educated (adjusted OR 3.48, P 0.01).

Why do they seek testing?

Reasons for Seeking Genetic Susceptibility Testing Among First-Degree Relatives of People With Alzheimer Disease

*J. Scott Roberts, ‡Susan A. LaRusse, ‡Heather Katzen, §‖Peter J. Whitehouse, §Melissa Barber, ‖Stephen G. Post, ‡Norman Relkin, ¶Kimberly Quaid, *Robert H. Pietrzak, **L. Adrienne Cupples, *†***†††Lindsay A. Farrer, *†Tamsen Brown, and *†Robert C. Green

Alzheimer Disease and Associated Disorders, 2003; 17: 86-93

Why do they seek testing?

Reasons for testing	Progressed to disclosure	Did not progress to disclosure
Prepare my spouse or children for my illness	88	71
Contribute to AD research	85	66
Information for family planning	87	73
Arrange long term care	85	71
Arrange personal affairs	84	71

People who decline testing are different from those who seek it

Table 1
Mean (*M*) psychological outcome scores by testing group (raw scores)

Measure	Baseline		7–10 Days		4 Months		12 Months	
	N	M (S.D.)	N	M (S.D.)	N	M (S.D.)	N	M (S.D.)
Breast cancer distress (total score)								
Carriers	30	13.1 (13.1)	25	21.2 (14.4)	25	17.7 (18.6)	20	16.1 (14.9)
Non-carriers	59	13.4 (14.6)	43	13.9 (16.1)	47	8.1 (13.5)	42	8.2 (14.2)
Not tested	51	16.0 (14.8)	45	14.9 (12.3)	50	13.1 (13.5)	43	12.3 (14.8)
State anxiety								
Carriers	25	36.1 (11.2)	24	38.5 (13.8)	26	36.8 (15.3)	22	31.7 (10.5)
Non-carriers	53	33.6 (12.1)	43	31.6 (11.1)	48	32.2 (10.8)	46	36.2 (12.9)
Not tested	47	33.6 (10.7)	46	36.8 (12.1)	48	36.3 (14.2)	46	39.0 (12.2)
Depression								
Carriers	25	5.5 (5.7)	24	5.3 (6.2)	26	6.2 (8.7)	22	4.0 (5.1)
Non-carriers	50	6.3 (6.7)	44	5.7 (7.0)	50	3.6 (5.4)	46	5.4 (6.4)
Not tested	47	5.9 (5.6)	47	7.2 (6.8)	48	6.4 (6.3)	46	6.9 (7.00)

S.D. standard deviation.

Meiser B et al. Psychological impact of genetic testing in women from high-risk breast cancer families. Eur J Cancer 2002; 38: 2025-31.

People who decline testing are different from those who seek it

Table 1
Mean (*M*) psychological outcome scores by testing group (raw scores)

Measure	Baseline		7–10 Days		4 Months		12 Months	
	N	M (S.D.)	N	M (S.D.)	N	M (S.D.)	N	M (S.D.)
Breast cancer distress (total score)								
Carriers	30	13.1 (13.1)	25	21.2 (14.4)	25	17.7 (18.6)	20	16.1 (14.9)
Non-carriers	59	13.4 (14.6)	43	13.9 (16.1)	47	8.1 (13.5)	42	8.2 (14.2)
Not tested	51	16.0 (14.8)	45	14.9 (12.3)	50	13.1 (13.5)	43	12.3 (14.8)
State anxiety								
Carriers	25	36.1 (11.2)	24	38.5 (13.8)	26	36.8 (15.3)	22	31.7 (10.5)
Non-carriers	53	33.6 (12.1)	43	31.6 (11.1)	48	32.2 (10.8)	46	36.2 (12.9)
Not tested	47	33.6 (10.7)	46	36.8 (12.1)	48	36.3 (14.2)	46	39.0 (12.2)
Depression								
Carriers	25	5.5 (5.7)	24	5.3 (6.2)	26	6.2 (8.7)	22	4.0 (5.1)
Non-carriers	50	6.3 (6.7)	44	5.7 (7.0)	50	3.6 (5.4)	46	5.4 (6.4)
Not tested	47	5.9 (5.6)	47	7.2 (6.8)	48	6.4 (6.3)	46	6.9 (7.00)

S.D. standard deviation.

Meiser B et al. Psychological impact of genetic testing in women from high-risk breast cancer families. Eur J Cancer 2002; 38: 2025-31.

Observations

- People who choose to pursue predispositional genetic testing may be better able to cope with test results than those who do not
- All evidence so far is based on studies performed in a clinical context (i.e. with in-depth counselling)

- These are both different from most or all DTC scenarios

Concerns

- As DTC diffuses more widely, will some people undergo testing who otherwise would have declined and protected themselves?
- Greater risk of harm to more vulnerable individuals without automatic provision of support services?
- Testing out of feeling of obligation to others rather than for personal health benefit?

Bibliography

Braithwaite D, Emery J, Walter F, Prevost AT, Sutton S. Psychological impact of genetic counseling for familial cancer: a systematic review and meta-analysis. *J Natl Cancer Inst Monogr* 2004; **96**: 122-33.

Broadstock M, Michie S, Marteau T. Psychological consequences of predictive genetic testing: a systematic review. *Eur J Hum Genet* 2000; **8**: 731-8.

Butow PN, Lobb EA, Meiser B, Barratt A, Tucker KM. Psychological outcomes and risk perception after genetic testing and counselling in breast cancer: a systematic review. *Med J Aust* 2003; **178**: 77-81.

Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. A systematic review of perceived risks, psychological and behavioural impacts of genetic testing. *Genet Med* 2008; **10**: 19-32.

Meiser B, Butow P, Friedlander M, Barratt A, Schnieden V, Watson M, Brown J, Tucker K. Psychological impact of genetic testing in women from high-risk breast cancer families. *Eur J Cancer* 2002; **38**: 2025-31.

Meiser B, Halliday JL. What is the impact of genetic counselling in women at increased risk of developing hereditary breast cancer? A meta-analytic review. *Soc Sci Med* 2002; **54**: 1463-70.

Wainberg S, Husted J. Utilization of screening and preventive surgery among unaffected carriers of a BRCA1 or BRCA2 gene mutation. *Cancer Epidemol, Biomarkers Prev* 2004; **13**: 1989-95.

Session 2C: Health Privacy in Practice

Session Chair: Michael Power, Privacy Consultant

Bio:

Michael Power is a Toronto-based legal advisor/consultant on privacy and information risk management, serving both public and private sector clients. He previously served as Vice-President, Privacy and Security, at eHealth Ontario. Prior to that, Michael was a partner at Gowling Lafleur Henderson LLP (Gowlings), advising on privacy and other information risk issues, where he also acted as Chief Privacy Officer.

Mr. Power writes and speaks extensively on privacy and information security issues and is the author of the Access and Privacy Title of Halsbury's Laws of Canada, co-author of the American Bar Association best-seller Sailing in Dangerous Waters: A Director's Guide to Data Governance. He is a member of the Nova Scotia Barristers Society and the Law Society of Upper Canada; is active in the Cyberspace Committee of the ABA's Business Law Section and is a member of the senior advisory board of the IEEE magazine, Security & Privacy. Michael Power received his LLB and MBA degrees from Dalhousie University.

US Data Security Requirements in EHRs

Peter McLaughlin, Foley & Lardner

Abstract:

Identity theft and fraud have been increasing in the US. The health sector must confront a two-fold challenge of reducing the financial impact as well as protecting the integrity of patient health records. In the US, recent legislation and rules specify the administrative, technical and physical requirements to protect the security and integrity of patient health and financial records. US attorney Peter McLaughlin, former Assistant General Counsel – Privacy and Security, for Cardinal Health, Inc., will discuss current technical and compliance requirements applicable to providers of electronic health records.

Bio:

Peter McLaughlin is senior counsel with Foley & Lardner LLP and a member of the firm's Privacy, Security & Information Management Practice. His experience as a corporate lawyer and business advisor includes international, health & financial privacy compliance, as well as data security and IT transactions. Prior to joining Foley, Mr. McLaughlin was in-house counsel for over eight years, including two years as Assistant General Counsel (Privacy and Security) and the first global privacy leader for Cardinal Health, Inc., a Fortune 20 company. Mr. McLaughlin received his J.D. from Georgetown Law Center and his bachelor's degree from Columbia University.

FOLEY

FOLEY & LARDNER LLP

EHIP 2009: US Data Security Requirements in EHRs

Peter F. McLaughlin

Senior Counsel

Agenda

- Evolving specifications for EHRs
- Enforcement looming
- Data security considerations generally
- EHR implementation recommendations

FOLEY

FOLEY & LARDNER LLP

Certified EHR

- Certified to meet standards adopted by the National Coordinator
- Patient demographic and clinical health information (e.g., medical history and problem lists)
- Ability to provide clinical decision support
- Support physician order entry
- Capture and query information relevant to health care quality
- Exchange electronic health information and integrate with other systems

Certified EHR: Providers

- Meets standards adopted by the National Coordinator
- Patient demographic and clinical health information (e.g., medical history and problem lists)
- Ability to provide clinical decision support for physician order entry
- Capture and query information relevant to health care quality
- Exchange electronic health information and integrate with other systems

Implementation of ARRA Passed Health Care Provisions

- Health IT Provisions
  - HIT Policy Committee: Makes policy recommendations to the National Coordinator for Health IT relating to the implementation of a nationwide health IT infrastructure, including:
    - Where standards, implementation specifications, and certification criteria are needed
    - The collection of quality data and public reporting
    - Biosurveillance and public health
    - Drug safety
    - Technologies to improve quality and safety

Implementation of ARRA Passed Health Care Provisions (cont'd)

- Requires the Office of the National Coordinator for Health IT (established in 2004) to appoint a chief privacy officer

- HIT Standards Committee: Recommends to the National Coordinator standards and certification criteria. Any current standards adopted before stimulus bill may be brought forward and applied. Must develop a schedule for assessing HIT Policy Committee recommendations and shall update the schedule annually

EHR Certification

- 12-31-09: ONC to adopt initial certification standards
  - To be recommended by HIT Policy Committee and tested by NSIT
  - CCHIT to provide guidance and appropriate certification standards

CCHIT and Certification

- September 8th: CCHIT announced plans to launch 2 new certification programs
  - "CCHIT Certified 2011" and "Preliminary ARRA 2011"
    - If final standards include new requirements, CCHIT will offer incremental inspections to vendors (at no charge) to bring preliminary certifications into alignment with final rules.

Practical Impact

- Standards not final, but need to start implementing/acquiring EHR functionality now
  - Vendor discussions and due diligence
  - Contractual protections will be key
    - Don't forget about security features to protect electronic PHI

FOLEY
FOLEY & LARDNER LLP

HITECH Act

- Health Information Technology for Economic and Clinical Health Act (HITECH Act) within the American Recovery and Reinvestment Act of 2009
  - Subtitle D - Privacy
    - Expansion of privacy and security requirements to forward adoption of EHRs
    - Impacts covered entities, business associates, and vendors not currently subject to HIPAA

FOLEY
FOLEY & LARDNER LLP

Heightened Enforcement

- Mandatory formal investigation and penalties for "willful neglect"
- Increased CMP amounts based on level of intent
  - Starts at $100; can go as high as $1.5 million

Heightened Enforcement (cont'd)

- State Attorneys General
  - Provided enforcement authority to bring actions on behalf of individuals
  - Courts can award damages, costs and attorney fees

Heightened Enforcement (cont'd)

- Penalties will be used to fund OCR enforcement activities
  - Portion of penalties to ultimately go to patients
- Business associates will be subject to criminal and civil penalties
- Employees of covered entities now clearly subject to criminal liability

Heightened Enforcement (cont'd)

- Audits
  - Covered entities **and** business associates will be subject to periodic audits

HITECH Act Breach Notification Guidance (cont'd)

- PHI is rendered unusable, unreadable, or indecipherable to unauthorized individuals if one or more of the following "safe harbors" apply:
  - Electronic PHI has been encrypted as specified in the HIPAA Security Rule by "the use of an algorithmic process to transform data into a form in which there is a low probability of assigning meaning without use of a confidential process or key" and such confidential process or key that might enable decryption has not been breached. Encryption processes identified below have been tested by the National Institute of Standards and Technology (NIST) and judged to meet this standard

FOLEY
FOLEY & LARDNER LLP

HITECH Act Breach Notification Guidance (cont'd)

- Valid encryption processes for data at rest are consistent with NIST Special Publication 800-111, *Guide to Storage Encryption Technologies for End User Devices*
- Valid encryption processes for data in motion are those that comply with the requirements of Federal Information Processing Standards (FIPS) 140-2. These included, as appropriate, standards described in NIST Special Publications 800-52, *Guidelines for the Selection and Use of Transport Layer Security (TLS) Implementations*, *800-77, Guide to IPsec VPNs*; or 800-113, *Guide to SSL VPNs*, and my include others which are FIPS 140-2 validated

FOLEY
FOLEY & LARDNER LLP

HITECH Act Breach Notification Guidance (cont'd)

17

- Data is vulnerable in the following states:
  - Data in motion (e.g., network, wireless transmission)
  - Data at rest (e.g., databases, file systems, other storage)
  - Data in use (e.g., being created, retrieved, updated)
    - Most problematic to secure
  - Data disposed (e.g., discarded paper records and electronic media)
- With the possible exception of "data in use", PHI in each of these states may be secured using one or more methods

HITECH Act Breach Notification Guidance (cont'd)

18

- PHI is rendered unusable, unreadable, or indecipherable to unauthorized individuals and thus is not "unsecured PHI" if one or more of the following "safe harbors" apply:
  - Encryption. All encryption can be "cracked," but sometimes don't need to crack (e.g., key logger brings down the mafia). Computationally infeasible. Security depends on:
    - Strength of the encryption algorithm; and
    - Security of decryption key/process
  - Destruction
    - Paper records
    - Electronic media

HITECH Act Breach Notification Guidance (cont'd)

- Encryption
  - Electronic PHI has been encrypted as specified in the HIPAA Security Rule by "the use of an algorithmic process to transform data into a form in which there is a low probability of assigning meaning without use of a confidential process or key" and such confidential process or key that might enable decryption has not been breached
  - List of technologies and methodologies identified in guidance for "safe harbor" is meant to be exhaustive not illustrative

FOLEY & LARDNER LLP

HITECH Act Breach Notification Guidance (cont'd)

- Encryption processes that have been tested by the National Institute of Standards and Technology (NIST) and judged to meet the HIPAA encryption standard
  - "Data at rest" means data that resides in databases, file systems, and other structured storage methods
  - Valid encryption processes for data at rest are consistent with NIST Special Publication 800-111, *Guide to Storage Encryption Technologies for End User Devices. (www.csrc.nist.gov)*

FOLEY & LARDNER LLP

HITECH Act Breach Notification Guidance (cont'd)

- "Data in Motion" means data that is moving through a network, including wireless transmission
- Valid encryption processes for data in motion are those that comply with the requirements of Federal Information Processing Standards (FIPS) 140-2, including:
  - Standards described in NIST Special Publications 800-52, *Guidelines for the Selection and Use of Transport Layer Security (TLS) Implementations,*
  - *800-77, Guide to IPsec VPNs,*
  - 800-113, *Guide to SSL VPNs,* and
  - May include others which are FIPS 140-2 validated

HITECH Act Breach Notification Guidance (cont'd)

- Data disposed means discarded paper records or recycled electronic media
  - The media on which the PHI is stored or recorded must have been destroyed in one of the following ways:
  - Paper, film, or other hard copy media have been shredded or destroyed such that the PHI cannot be read or otherwise cannot be reconstructed
  - Electronic media have been cleared, purged, or destroyed consistent with NIST Special Publication 800-88, *Guidelines for Media Sanitization*, such that the PHI cannot be retrieved
  - Beware hardware support vendors, removable media, and e-bay

Breach Detection

- Detection of Breaches of Unsecured PHI
  - Monitoring use of PHI
    - Review of audit trails – All applications and operating systems generate audit/log files. Can be voluminous. All relevant information may not be captured.
      - Sometimes improperly configured, turned off, or records overwritten
    - Systems that help monitor access, use, and disclosure of PHI

Breach Investigation

- Investigation of Potential Breaches of Unsecured PHI
  - Determining whether a breach of and Electronic Health Records occurred
    - Audit trails
    - Computer Forensics
    - Other

HITECH Act Breach Notification Guidance (cont'd)

- HHS is seeking comment on:
  - Electronic media configurations such as a fingerprint protected Universal Serial Bus (USB) drive, which guidance should specifically address
  - Other methods that should be considered for rendering paper and electronic PHI unusable, unreadable, or indecipherable to unauthorized individuals
  - Circumstances under which the methods discussed above would fail to render information unusable, unreadable, or indecipherable to unauthorized individuals
  - Whether future guidance should specify which off-the-shelf products, if any, meet the encryption standards identified in this guidance

Accounting

- System capabilities
  - Information that can be logged
  - How logs can be created
  - When information can be tracked
  - Whether audit trails can be used for accounting

Accounting

- Accounting Rules Change for EHRs
  - Must account for disclosures related to treatment, payment and health care operations (as well as all other accountable disclosures)
  - Three year period
  - Business Associates may be impacted
  - Regulations to be issued regarding the information that must be collected for an accounting

Accounting (cont'd)

- Effective Date
  - Current users of EHRs: 1-1-14
  - Future users of EHRs (after 1-1-09): 1-1-11 or date EHR is acquired (whichever is later)

HITECH Act Audit Trail Capabilities

- Audit Trail Capabilities:
  - Determine what is tracked and by which application
  - Confirm an audit trail is created by all relevant systems for all relevant data and actions. Look for gaps
  - Audit trail data must be backed up and retained for a period consistent with relevant document retention requirements
  - Audit trails may be intentionally or unintentionally modified, corrupted, or destroyed
  - Consider use of WORM drives or other technology to ensure the integrity of audit trail data
  - Make audit trail functionality part of all relevant IT contracts, including outsourcing engagements

What the HITECH Act Means for Business Associates

- HITECH Act
  - Requires Business Associates to Comply with HIPAA Security Rule
    - Security rule applies to business associate of a covered entity in the same manner that such sections apply to the covered entity (45 C.F.R. §§ 164.308, 164.310, 164.312, and 164.316)
    - Allows Secretary of HHS to conduct periodic audits of covered entities *and* business associates for compliance with security rules
  - Security requirements must be incorporated into Business Associate Agreements
  - Effective 12 months after enactment of legislation

What the HITECH Act Means for Business Associates (cont'd)

- HITECH Act –
  - Requires Business Associates to report breaches of unsecured PHI to Covered Entities
    - Effective 30 days after regulations are issued on breach notification
  - Requires Business Associates to Comply with HIPAA Privacy Rule
    - Subjects Business Associates to enforcement provisions, *e.g.* civil and criminal penalties for HIPAA violations
    - Effective 12 months after enactment of legislation

Business Associate Compliance with the HIPAA Security Rule

- Written Security Plan and Documentation is Required
  - Maintain the policies and procedures in written form
    - Must address standards and implementation specifications
      - Some implementations are required; others are addressable
      - If addressable, entity can assess reasonableness and appropriateness of safeguard to its environment and either implement or document and implement an equivalent alternative measure
  - Maintain a written record of the security assessment
    - Time Limit (Required) – 6 years
    - Availability (Required)
    - Updates (Required)

Overview: Information Security

- Security requires a unified approach
  - Security policies
  - Employee education
  - Use of technology (e.g., firewalls, encryption, intrusion detection systems)
  - Security audits
  - Addressing security in contracts with business partners and other vendors

Industry Practice Sufficient?

- May represent minimum requirements
  - Company must "implement standard practices… where such standards have gained sufficient industry acceptance and adoption such that… adherence to the standards would not unreasonably place [company] at a competitive disadvantage."
    Ziff-Davis FTC Consent Decree
- But not necessarily a guarantee of compliance…
  - "[An industry] never may set its own tests, however persuasive be its usages. Courts must in the end say what is required; there are precautions so imperative that even their universal disregard will not excuse their omission."
    T.J. Hooper case, 60 F.2d 737 (2d Cir. 1932)

Three Step Approach

- ✓ Vendor due diligence
- ✓ Contractual protections
- ✓ Information handling procedures and requirements, generally in the form of contract exhibits

Step One: Due Diligence

- ❖ From the outset, Vendors must be on notice that the information they provide as part of the provider's information security due diligence will be (i) relied upon in making a vendor selection; and (ii) part of the ultimate contract.
- ❖ To ensure proper documentation and uniformity in the due diligence process, providers should develop a "Vendor Due Diligence Questionnaire."

Step Two: Contractual Protections

- Personnel due diligence - Background checks and screening
- Use of subcontractors
  - Strictly limit
  - Approval required
  - Joint and several liability
  - Due diligence
  - Consider use of NDAs to achieve direct contractual privity

Step Two: Contractual Protections

- Control of Personnel
  - Compliance with facility access and security policies
  - Vendor identification card
  - Access scheduling
  - Escorts required
- General Audit Provision
  - Permit audit of vendor compliance with contract terms, including confidentiality, security, personnel, etc.
- No Removal of Data

Step Two: Contractual Protections

- General Security Obligations
  - Take all reasonable measures to secure and defend its systems and facilities from unauthorized access or intrusion
  - Periodically test systems and facilities for vulnerabilities
  - Immediate reporting of breaches
  - Joint security audits
  - Regulatory access and compliance
  - Firewalls, antivirus, use of VPNs, on-demand access

Step Two: Contractual Protections

- Security Breach Notification For PII - - Associated Costs
  - Control of notice
  - Allocate responsibility for costs vendor

Step Three: Information Handling Requirements

- Where appropriate, attach specific information handling requirements in an exhibit
  - Securing PII
  - Encryption
  - Secure destruction of data
  - Securing of removable media
  - Communication and coordination

FOLEY
FOLEY & LARDNER LLP

Questions?

Peter F. McLaughlin
Senior Counsel
pmclaughlin@foley.com
617.502.3265

The Economics of Privacy in Health Care

Mike Gurski, Bell Canada

Abstract:

Based on a Whitepaper, now published by the IEEE and a Think Tank for MOHLTC this session will explore a way to invest strategically in privacy in health care environments. Thus, cost reductions can accrue at the same time as the efficacy of privacy management can grow. Examples of organizations who have begun to adopt components of this privacy investment framework will be cited. The costs of not following this model will also be discussed. At the end privacy professionals should have a privacy story that any senior executive wants to know and can endorse.

Bio:

Mike Gurski is the Director of the Bell Privacy Centre of Excellence and the Privacy Strategist for Bell Information & Communications Technology Solutions. In his responsibilities at the Centre he leads a comprehensive privacy professional services arm for enterprise customers. Mike also heads a research arm focused on developing privacy technologies in areas that include: wireless health care environments, identity theft solutions, and Internet censorship circumvention software. As well Mike is a founding member of the 'The Privacy Network (www.theprivacynetwork.org): a knowledge exchange network that links various privacy communities in Canada.

Mike is also on the Board of Directors for the International Security Trust and Privacy Alliance which is developing a privacy framework to assist organizations in implementing privacy from a systems and technology perspective. Prior to joining BSSI Mike chaired the international Privacy Enhancing Technology Testing and Evaluation Project, to develop privacy technology evaluation standards and was the founding Chair of the Wroclaw Foundation: an international data protection commissioners' vehicle to facilitate international privacy technology standards. He also served as the Senior Technology Advisor to Ontario's Office of the Information and Privacy Commissioner for five years. Currently, he is on the Board of the Privacy Enhancing Technology (PET) Research Workshop; an international research symposium, and chairs both the international PETs Executive Briefing and the University of Waterloo's annual Centre for Applied Cryptographic Research, Privacy and Security Conference (www.cacr.math.uwaterloo.ca).

Mike has written published articles on e-mail encryption, misconceptions of privacy and security, wireless, and P3P (Platform for Privacy Preferences), a privacy specification for the Web: this latter work while a member of W3C team developing P3P. As well he has written papers on Privacy Design Principles and Privacy Impact Assessments for Integrated Justice Technology Systems. This was done in partnership with the United States Justice Department's Office of Justice Programs.

Mike is a frequent speaker on privacy issues and a guest lecturer at number of MBA schools and universities in Canada and abroad. Mike holds degrees from the University of Waterloo's School of Architecture and the Faculty of Arts, St. Jerome's University.

In his spare time Mike pursues research on megalithic architecture and organizes bike trips in Europe.

The Value of a Privacy Culture in Healthcare

EHIP: Ottawa

2009

Bell

Agenda

- Introduction
- Framing the Value Proposition
- Costs and benefits: striking a balance
- The Elements of Culture: An enterprise privacy framework
- Tales from the early adopters

Bell

Background

- Belief that privacy is important
- Move towards eHealth delivery
- The challenge:
  1. Construct an argument to demonstrate to health care providers that privacy is of value in health care
  2. Create a framework for promoting a culture of privacy

Bell

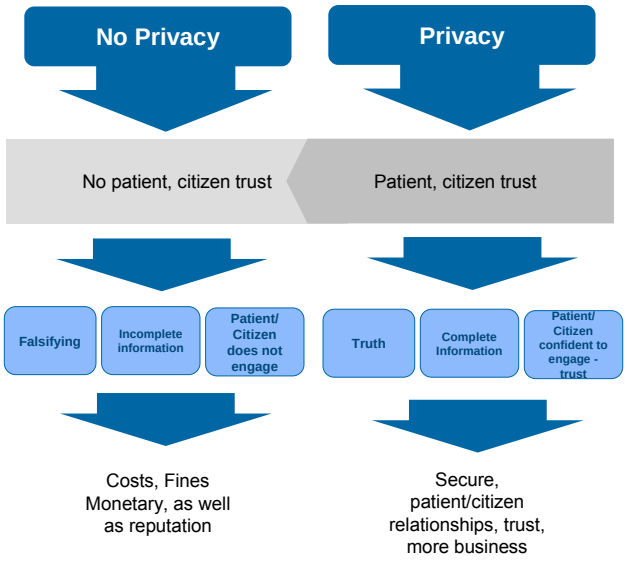
Privacy? Forget about It!

Privacy tends to fall by the wayside for two reasons:

1. the value of privacy is not readily accessible and
2. the return on investment is unclear

Bell

Forget about It? Forget you!



Bell

Lack of Privacy Costs

- Inaccuracies
- Delayed care
- Misdiagnosis
- Errors in care
- Loss of reputation
- Fines / Penalties
- Identity Theft
- Breaches

Bell

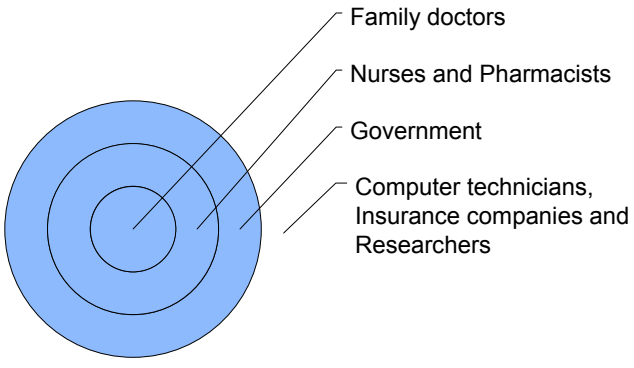
Privacy, a Matter of Trust

- Individuals tend to have the most confidence in those delivering the care directly
- Confidence highest for family doctors and high in both nurses and pharmacists

Source: EKOS Research Associates, Revisiting the Privacy Landscape a Year Later, submitted to the Privacy Commissioner of Canada, March 2006

Bell

Ripple Effect



Bell

People Care

- Media reports on breaches
- Shredder sales are up
- Support for the do-not-call list
- Legislative initiatives (e.g. Identity theft and SPAM)

Bell

Ask 10 people…
get 11 different answers

- Privacy is a multifaceted construct
- Heavily dependent on context

Bell

Different Definitions

"Dad, get out of my room!"
(the right to be left alone)

"Dad, it's none of your business!"
(informational self-determination)

Bell

And Different Approaches

fundamentalists — **about 25% of the public are privacy zealots**

pragmatists — **about 57% of public believe that business should "earn" the public's trust**

don't cares — **about 18% of public have nothing to hide**

-Alan Westin's testimony, The House Committee on Energy and Commerce (May 8, 2001)

Bell

And different contexts

- Environment has an impact
- Let's compare:

A doctor's office

A subway station

- Reasonable expectation of privacy

Bell

Add Technology to the Mix

- Countless benefits in health care
- Countless problems in health care
- One thing is certain: privacy properties are amplified

Bell

More or Less Privacy? Ask the Individual

- Most Canadians claim they're comfortable with EHRs
- And it's ok for researchers to link personal health information

But

- Ask first!
- Consent choices have value

Source: EKOS Research Associates, "Electronic Health Information and Privacy Survey: What Canadians Think – 2007" (August 2007)

Bell

More or Less Privacy? Ask the Institution

- Study showed Research Ethics Boards vary in their approach to consent
- Lack of recognition, particularly among the sites not requiring consent, that linkability could lead to re-identification of individuals
- Cited high level of trust that researchers would not attempt to re-identify individuals

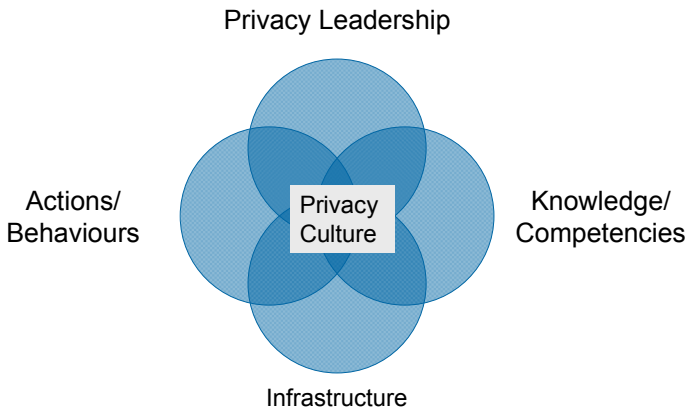
Bell

Oscillating Control

- Over control vs. under control
- Relevant at both individual and institutional levels
- Points to need for privacy framework

Bell

Bell Enterprise Privacy Framework



Bell

Bell Privacy framework/best practices

Leadership	Knowledge	Action	Infrastructure
Selection of a leader (one person) or a core team to facilitate a Privacy framework, strategy and plan Must be positioned to provide strategic insight into business requirements	Institutional knowledge Identification of privacy issues within the context Familiarity with legislative and regulatory requirements Understanding privacy costs and benefits and how these impact individual and institutional decision making Understand privacy issues, challenges and how to address them	Privacy affects all departments and people Must be a holistic (organization-wide) Manifest as individuals function within the system Needs to be proactive – leadership sets out key steps Privacy management as part of daily procedures Pursue public privacy leadership course, through policies, annual reports, etc	Facilities and systems serving the enterprise which support privacy Relies heavily on IT **Examples**: •Communications systems •Database roles and permissions •Compliance tools •Training and Operational Controls

Bell

Privacy Leadership

- Understand privacy issues and how to address them
- Could be an individual or a core team
- Must be positioned to provide strategic insight into business requirements and take a holistic approach to their work

Bell

Knowledge/Competencies

- Institutional knowledge
- Identification of privacy issues within the context
- Familiarity with legislative and regulatory requirements
- Understanding privacy costs and benefits and how these impact individual and institutional decision making

Bell

Actions

- Manifest as individuals function within the system
- Being proactive – leadership sets out key steps
- Privacy management is part of daily procedures
- Pursue public privacy leadership course, through policies, annual reports, etc

Bell

Infrastructure

- Facilities and systems serving the enterprise which support privacy
- Relies heavily on IT
- Examples:
  - Communications systems
  - Database roles and permissions
  - Compliance tools
  - Training and Operational Controls

Bell

The Tools

- Privacy policies and procedures
- Privacy training
- Privacy communications
- Privacy impact assessments

Bell

Tales from the Early Adopters

- Newfoundland & Labrador
- Champlain LHIN
- Hospital for Sick Children: Canadian Pedeatric Wait Times Strategy
- Public Health Agency for Canada
- Regional Municipality of York (the 905)

Bell

Conclusion

- Privacy has value in healthcare
- Trust is key
- Understand privacy issues and recognize the role of context
- Balance of costs and benefits
- Implement a privacy framework which addresses that balance

Bell

Contact

Howard Simkevitz
Senior Privacy and
Information Technology Counsel
Bell Privacy Centre of Excellence
416-229-6698
howard.simkevitz@bell.ca