Data Confidentiality and Gene Privacy

Ju Han Kim, M.D., Ph.D.
Professor and Founding Chair, Div. of Biomedical Informatics, Seoul Nat'l Univ. College of Medicine
juhan@snu.ac.kr

No one wants to expose…

- ***credit card numbers***
- ***bank account numbers***
- ***passwords***
- ***sensitive data (patient data)***

Protecting…

- What?
  - Security
  - Privacy
  - Confidentiality
- Against what?
  - Evil hackers
  - Malicious insiders
  - Stupidity

Privacy

- Right to be alone; e.g.:
  - applies mostly to known individuals
- Correlation among pervasive databases
  - sensus
  - marketing
  - health

Confidentiality

- *Use of sharing information by multiple users at many institutions*
- *Should be controlled by coherent policy*
- *Enforced by appropriate technology*
- *E.g., who may use your results of life insurance, for what purposes?*

프라이버시 / 기밀유지 / 보안

- *Privacy: managing your own information to suit your needs*

right to remain unknown

- *Confidentiality: managing someone else's information to protect their privacy*
- *Security: physical security*

Privacy & Intruders

Freedom from

1. **intrusion**
2. **surveillance**
3. right to self control → (**Patient control**)
   - ✓ Giving patients control of the use of their data
   - ✓ To be informed and to control who, when, how, and why their health information is accessed/used
   - ✓ Broader concept than the right to inspect/read

Privacy as the right to life

- Competition
  - ✓ Competition at equality condition
- Autonomy and Freewill
  - ✓ Right to choose religion and good and evil
  - ✓ Right to belief
- Right to forget or not recognize discrimination
  - ✓ Race, gender, regional sentiment

*** Privacy Case Nydia VelÃ¡zquez Â (1982)** Three weeks after Nydia VelÃ¡zquez won the New York Democratic Party's nomination to serve in the U.S. House of Representatives, somebody at St. Claire Hospital in New York faxed VelÃ¡zquez's medical records to the New York Post. The records detailed the care that VelÃ¡zquez had received at the hospital after a suicide attempt--an attempt that had happened several years before the election.

Database Nation: The Death of Privacy in the 21st Century, Simson Garfinkel, Jan 2000, 1-56592-653-6

The intruders

- The Big Brother
- The Little Sisters
- Intrusive Technologies
- Stupidity
- Internal breaches
- Ever increasing stakeholders
- Data integration
- Re-identification of the de-identified

The Intruders – Big Brother

Big brother is watching you!

- **Governmental DBs**
- **National Surveillance**
- **International Collaborations**

The Intruders – Little Sisters

The Sisters are nearer than the Bros

- Flaming
- Flame war
- Cyberbullying
- Internet Trolling
- Smack Talk

The Intruders –Technologies

1. Face recognition, Biometrics (DNA, fingerprints, iris, gait)
2. Video Surveillance, Ubiquitous Networks (Sensors)
3. Semantic Web, Data Mining, Bio-Terrorism Surveillance
4. Professional Assistants (email and scheduling), Lifelog recording
5. E911 Cell Phones, IR Tags, GPS
6. Personal Robots, Intelligent Spaces, CareMedia
7. Peer to peer Sharing, Spam Blockers, Instant Messaging
8. Tutoring Systems, Classroom Recording, Cheating Detectors
9. DNA sequences, Genomic data, Pharmaco-genomics

The Intruders – Stupidity

A14 National Report

FRIDAY, APRIL 4, 1987

The New York Times

Patient Files Turn Up in Used Computer

By JOHN MARKOFF

SAN FRANCISCO, April 3 – The last thing that C. J. Prime expected to find when she hooked up the used I.B.M. computer she had purchased at an Internet auction were 2,000 patient records from Smitty's Supermarkets pharmacy in Tempe, Ariz.

But that is exactly what Ms. Prime found last month when she turned on the computer, which she bought for $159 from the Onsale Corporation, based in Mountain View, Calif.

All of the software that the pharmacy had used for record keeping was still on the computer's hard disk, including patient names, addresses, Social Security numbers and a chronological list of all the medicine that they had bought at the pharmacy.

"I was alarmed at seeing the patient records," said Ms. Prime, who is a self-employed computer technician in Pahrump, Nev. "I know a lot of people who would be devastated if they knew this kind of information was floating around."

Experts in privacy law say Ms. Prime stumbled upon a growing

C. J. Prime of Pahrump, Nev., displaying the used I.B.M. computer she bought that had 2,000 patient records from a pharmacy in Tempe, Ariz.

- frigidity
- ignorance
- divide

Internal breaches

The dark side

누군가 당신 병력을 본다면 … 대형병원 전자의무기록 접근 허술

http://blog.joins.com/dreronca1/4959210

조회(79) / 추천

등록일 : 2005-06-23 05:04:14

■ 의료법 제21조의 2(전자의무기록)

3. 누구든지 정당한 사유없이 전자의무기록에 저장된 개인정보를 탐지하거나 누출·변조 또는 훼손하여서는 아니된다.

■ 제66조(벌칙)

5년 이하의 징역 또는 2000만원 이하의 벌금에 처한다.

서울대병원·세브란스병원·서울아산병원·삼성서울병원 등 전산망을 통한 전자의무기록체제를 채택하고 있는 국내 10개 대형병원을 조사한 결과 10개 모두 환자의 진료내역이 주치의 등 해당 진료의사 외 병원 직원들에게 고스란히 노출돼 있는 것으로 밝혀졌다. 입원 환자의 성명만 단말기에 기입하면 환자의 주민등록번호와 주소 등 신상명세는 물론 혈액검사 등 주요 검사 결과와 현재 어떤 약을 투여하고 있고, 어떤 수술을 받았는지 모두 알 수 있다. 여기엔 각종 성병과 인공유산수술 과거력, 정신병 여부 등 민감한 프라이버시가 담긴 내용이 담겨 있다.

진료과목이 다른 의사나 병동이 다른 간호사는 물론 원무과 행정직원도 아름만 먹으면 쉽게 알 수 있게 돼 있다. 서울 N병원의 경우 심지어 해당 진료과목이 아닌 의사라도 전산망에 들어가 주치의의 오더(투약과 처치 등 환자치료 지시)까지 바꿀 수 있는 것으로 밝혀졌다.

의료사고전문 전현희 변호사(대외법률사무소)는 "의료진이라 하더라도 진료 목적이 아닌 열람은 위법의 소지가 있다"며 "관련 법률의 정비가 필요하다"고 말했다.

일본 등 전자의무기록이 보편화된 선진국의 경우 환자의 프라이버시가 철저하게 보호된다. 도쿄대학 법학부 히구치 노리오 교수는 최근 서울에서 개최된 대한의사협회 '현대의료정보와 프라이버시'심포지엄에서 "일본은 2003년 5월 개인정보보호법을 통해 환자의 의무기록은 자문의뢰 등 비록 진료 목적이라도 다른 의료인에 통보할 경우 환자의 동의를 거치도록 의무화했다"고 말했다.

출처: 중앙일보 2005년6월23일자

The Intruders - ever increasing stakeholders

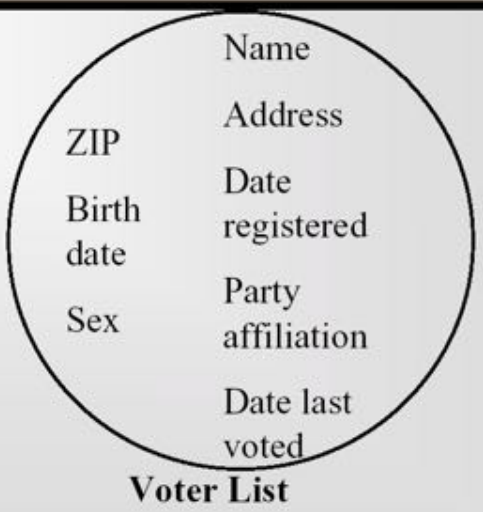
- Dr.
- Nr.
- Therapists
- Laboratory
- Radiology
- Pharmacy
- Admissions
- Administrations

… more than 70

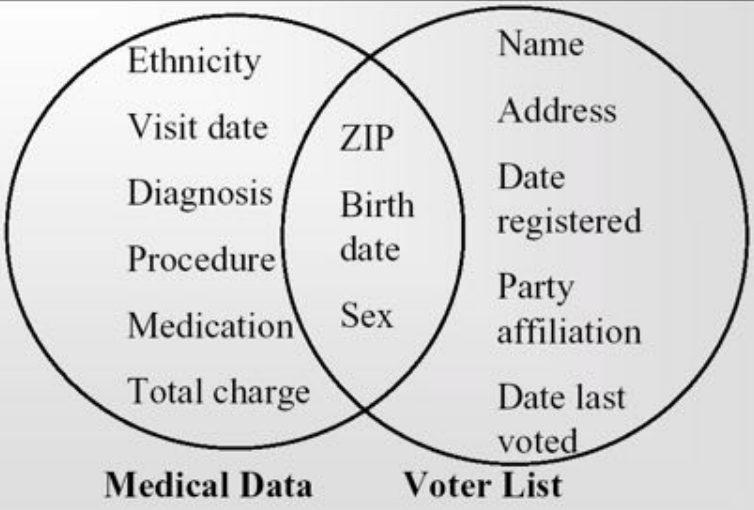
- Managers
- Patients
- Payers
- Reviewers
- Gov. Institutions
- Insurance Company, Pharma
- Hackers
- and more and more people…

The Intruders – Data Integration

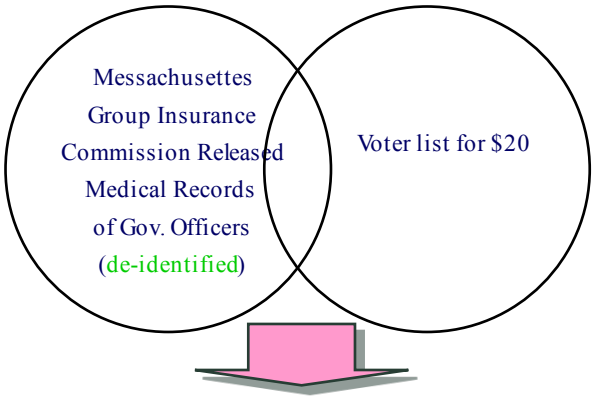
The Intruders – Data Integration



The Intruders – Data Integration



De-identification &Re-identification



Former Governor, William F. Weld

Group Insurance Commission Record Decoded

출처: Sherman E. It doesn't take much to make you stand out. Cambridge, Mass.: Harvard University Extension School Bulletin, Fall 2001

Reidentification of Individuals in Chicago's Homicide Database
A Technical and Legal Study

Salvador Ochoa Jamie Rasmussen Christine Robson Michael Salib

Collective address: reidentify@mit.edu

Names of the 35% of the victims were reidentified (only with public data)

Abstract

Many government agencies, hospitals, and other organizations collect personal data of a sensitive nature. Often, these groups would like to release their data for statistical analysis by the scientific community, but do not want to cause the subjects of the data embarrassment or harassment. To resolve this conflict between privacy and progress, data is often deidentified before publication. In short, personally identifying information such as names, home addresses, and social security numbers are stripped from the data. We analyzed one such deidentified data set containing information about Chicago homicide victims over a span of three decades. By comparing the records in the Chicago data set with records in the Social Security Death Index, we were able to associate names with, or reidentify, 35% of the victims. This study details the reidentification method and results, and includes a legal review of U.S. regulations related to reidentification. Based on the findings of our project, we recommend removal of these databases from their online locations, and the establishment of national deidentification regulations.

Re-identification < Rare disease >

Malin and Sweeney at Carnegie Mellon Univ. integrated
(1) Illinois' publicly available de-identified discharge summary data (1990-1997) with (2) Census data and (3) Voter list,
surprisingly re-identifying real names of rare disease patients by using the publicly available data only

Cystic fibrosis: 33%
Huntington disease: 50%
Fanconi Anemia: 70%
Refsum disease: 100%

How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems

Bradley Malin, and Latanya Sweeney

Data Privacy Laboratory, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

Received 23 December 2003. Available online 28 May 2004.

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

Health-Information Altruists — A Potentially Critical Resource

Isaac S. Kohane, M.D., Ph.D., and Russ B. Altman, M.D., Ph.D.

One of the key ideas behind sequencing the human genome was the promise of "personalized medicine." The idea was that genetic information could be used to make health care more precise, efficacious, and safe. The Human Genome Project showed us that among humans, DNA sequences are 99.9 percent similar, but the remaining 0.1 percent, in the context of environmental and epigenetic factors, produces the entirety of genetic variability within the human population. How can we use the information about human genetic variation to achieve these stated goals of the genome-sequencing effort?[1] Investigators are currently collecting phenotypic information about patients (their disease diagnoses, prognoses, and treatments) and comparing it with their DNA sequences. Methods used to obtain these large phenotypically annotated populations may not be adequately productive because of concerns about privacy and disclosure of genotypic and phenotypic data. We think these concerns are real but addressable sociologically, technologically, and legislatively. The basic idea is that giving patients control of the use of their health data will provide a practical mechanism for harnessing the volunteerism of our populations and gathering research data.

the National Human Genome Research Institute, has called for large cohorts (at least 200,000 subjects) to be assembled simply to achieve the necessary sample sizes to overcome the problems of cross-sectional studies.[2]

Public standards and patients' control: how to keep electronic medical records accessible but private

Kenneth D Mandl, Peter Szolovits, Isaac S Kohane

British Medical Journal (2001)

and never used to discriminate
consequence, researchers inve
in removing any information from research data sets that could be used to identify the specific participants.

However, a recent study by Malin and Sweeney concerning database security has shown that apparently de-identified subjects often can be either unambiguously re-identified or partially identified by means of filtering the data to a very small subgroup of potential matches.[3] Malin and Sweeney took publicly availat
data from
combined them with Census data and voter-regis-

New England Journal of Medicine (2005)

Efforts

regional and international

Legislative efforts in Korea

- Constitution
  - ✓ 제17조: "모든 국민은 사생활의 비밀과 자유를 침해받지 아니한다."
- Criminal Laws:
  - ✓ 제316조, 비밀침해 행위 처벌;
  - ✓ 제317조, 의사, 한의사, 치과의사, 약제사, 조산사 등이 업무처리 중 지득한 타인의 비밀을 누설시 처벌
- Privacy Act
- Acts on Information and Communication:
  - ✓ 정보통신망이용촉진및정보보호등에관한법률 제21조(전자문서 등의 공개 제한) 및 제49조(비밀 등의 보호)
  - ✓ 전자서명법 제24조(개인정보의 보호)
  - ✓ 공공기관의개인정보보호에관한법률 제13조(처리정보의 열람제한)
- Medicine-related Acts
  - ✓ 보건의료기본법 제12조(비밀보장)
  - ✓ 의료법 제19조(비밀누설의 금지)
  - ✓ 전염병예방법 제54조의 6
  - ✓ 후천성면역결핍증예방법 제7조
  - ✓ 장기이식등에관한법률 제27조

26/62

HIPAA

Health Insurance Portability and Accountability Act

- Since 1996, U.S. congress
  - data interchange standards
  - data security
  - patient privacy
- HIPAA Security and Electronic Signature Standards, 1998
- HIPAA Standards for Privacy of Individually Identifiable Health Information, 2000
- HIPAA regulation starts in 2003

Multi-center studies - The challenges

- Registries and Large databases
  - ✓ Cancer
  - ✓ Childhood immunizations
  - ✓ Cardiovascular surgery
  - ✓ Mammography screening
- Quality improvement and assurance
- Technologic advancement, large-scale data sharing
- Federal, state laws & institutional policies
- Collection, storage, utilization and sharing

Categories of Information

TABLE 1. Categories and definitions of confidential information

Category of information	Definition
Public	Information or data collected, compiled, utilized, or generated that is intended for public distribution and use or that may be obtained under freedom of information legislation. Generally, this includes aggregated data in published form, such as articles in medical journals about patterns of care, accuracy, and other related topics. This does not include confidential information.
Internal	Information or data collected, compiled, utilized, or generated by the organization that may be shared with employees and authorized consultants and contractors only. Authorization for external distribution or access must be obtained from the Principal Investigator. Examples of internal information include site lists, technical reports, and research proposals in stages of preparation.
Restricted	Confidential information collected, compiled, utilized, and/or stored by the organization that contains identifying links with specific individuals or medical practices, such as name, address, and Social Security number. Confidential registry data and reports fall within this category, as do any personal identifiers collected as part of a registry (including diagnoses that, when linked with geographic location, could identify an individual or number of patients served by a facility that could identify provider participants).

Member sites

- Research endeavor vs. confidentiality protection
- Protect from unauthorized access
- Usage only in sanctioned and approved ways
- Prompt report and corrective measures against breaches of the policy
- Prompt response to inquires from concerned participants

Categories of Information

TABLE 2. Types of protection offered by federal or state governments and individual institutions

Type of protection
Federal
Public health service certificate of confidentiality
IRB* requirements for protection of subjects from the risk of loss of confidentiality
State
Laws protecting the confidentiality of records used in medical research
Laws protecting cancer or mammography registries
Quality assurance or peer-review statutes
Laws regulating physician-patient privilege
Laws on Patient's Bill of Rights
Laws governing confidentiality of patient's medical records
Institutional
Data security
Limiting data access with key or password protection
Outlining the specifics of all data handling using a standardized protocol
Shredding unneeded paper data
Formalizing all data requests and establishing a review process for release of research data
Developing a firewall for all computer systems
Maintaining off-site backups of computerized databases
Using a specially designed encryption program to convert data before sending it over the Internet

* IRB, institutional review board.

UK Association of Cancer Registries

- Regulation 2 of the Statutory Instrument (SI) on confidentiality – No. 1438, The Health Service (Control of Patient Information) Regulations 2002 – permits cancer registries to receive patient identifiable data without the need for informed consent.
- However, there remains uncertainty about the circumstances when cancer registries are allowed to disclose patient identifiable data held by them to third parts.
- PIAG has requested UKACR to develop explicit guidance for cancer registries advising them that they must comply with requests from patients to delete identifiable data about themselves from their databases.

UK Association of Cancer Registries

- The basic idea for protecting patient privacy has been de-identification.
- However, the dichotomy of identifiable vs. non-identifiable distinction cannot be made.
- In reality, most of health data are 'Potentially Identifiable'.
  - ✓ Individual records
  - ✓ Tabular data, based on small geographic areas, with cell counts of fewer than five cases/events (or where counts of less than five can be inferred by simple arithmetic)
  - ✓ Tabular data containing cells that have underlying population denominators of less than approximately 1000

34/62

Potentially identifiable data

- the intended use(s) of the data should be stated clearly
- the use(s) of the data should be justified and the data should not be used for any other purposes
- the registry should not release data that are more detailed than necessary to fulfill the stated purpose
- the data should not be passed on to other third parties or released into the public domain
- the data should be kept securely for the period of time that can be justified by the stated purpose, and then destroyed
- no attempt should be made to identify information pertaining to particular individuals or to contact individuals
- no attempt should be made to link the data to other data sets, unless agreed with the data providers
- any public domain reports or papers resulting from analyses of the provided data should be shared prior to publication with the cancer registry (or registries) supplying the information.

American College of Epidemiology Policy Statements

- Routine anonymization of archived medical data :
  - ✓ difficulty in tracing back to individuals
  - ✓ Unable to predict what linkage might be useful in the future investigations
- Individual informed consent
  - ✓ Untenable administrative, financial, and logistical burdens
  - ✓ Non-participation and selection bias

ACE with bigger challenges

Table 1. Summary of Recommendations

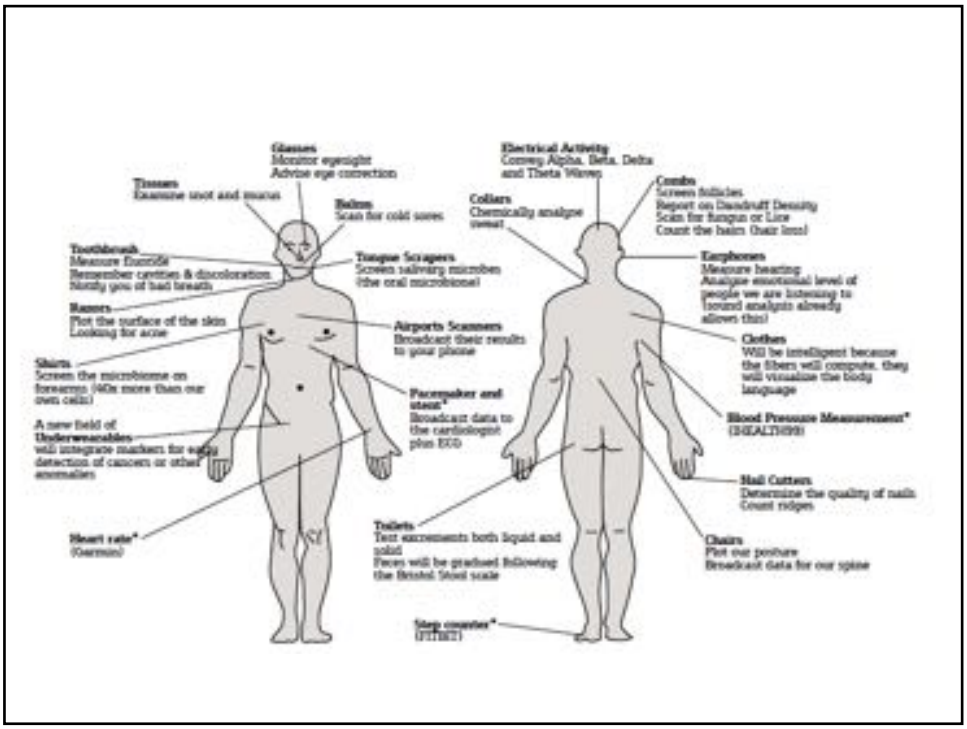
Recommendation	Description
Create a scientific forum on population sciences	The NHLBI should convene a scientific forum to anticipate the major scientific questions and methodological needs in epidemiology and population science over the next 10–20 years.
Launch electronic epidemiology, particularly in collaboration with other organizations and agencies	The NHLBI should actively engage in studies to establish the validity, reliability, and scalability of electronic tools for primary data collection. In doing so, the NHLBI should partner with other organizations and agencies.
Build the data-science workforce	The NHLBI should help establish an adequate workforce to conduct population sciences "of the future," and one approach is to create multifaceted and complementary career development grants.
Develop a dynamic compendium of epidemiologic resources	Resources should be dedicated to creating a dynamic compendium of large epidemiologic resources, including cohort studies, clinical trials data sets, registries, biorepositories, and other relevant epidemiologic resources, to assist the research community in identifying and accessing key existing resources and to improve the return on the investment from these studies.
Integrate epidemiology and clinical trials	Where genuine efficiencies can be created, the NHLBI should encourage the integration of clinical trials and epidemiologic studies.
Create a cohort consortium	The NHLBI should create a cohort consortium to support large-scale collaborations and provide a coordinated, interdisciplinary approach to addressing scientific questions, achieving economies of scale, creating opportunities for collaboration, and accelerating the pace of research and the implementation of new methods.
Implement competitive external evaluation of cohorts	The NHLBI should implement a competitive peer review–based model for its portfolio of large epidemiologic and population studies.

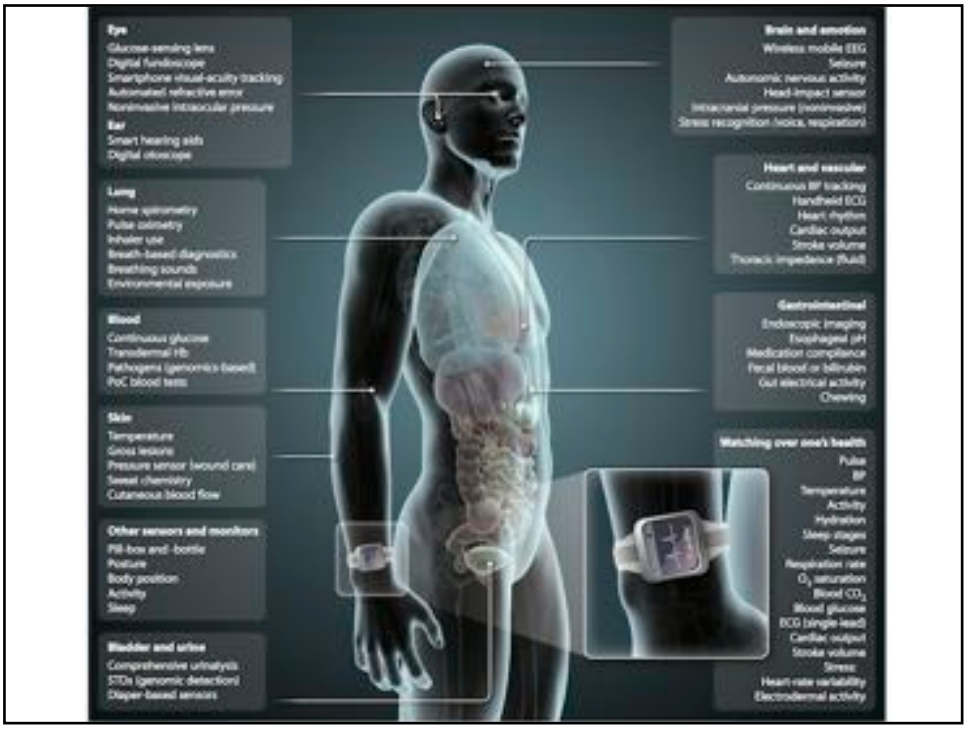
Abbreviation: NHLBI, National Heart, Lung, Blood Institute.

New Challenges

New Challenges

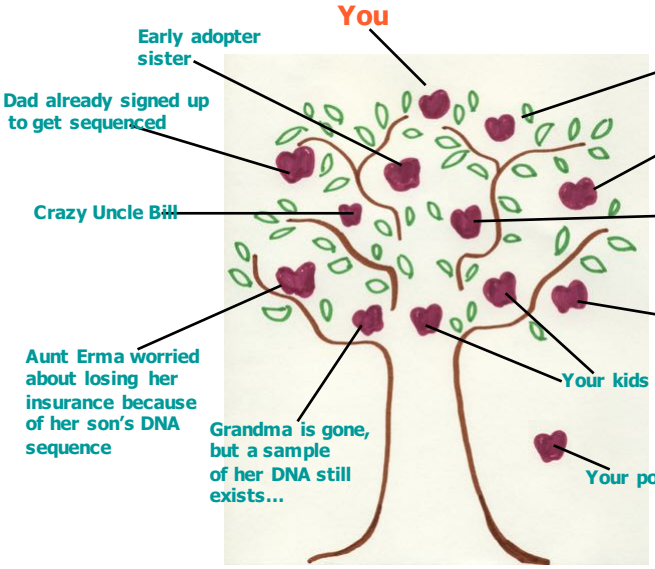
- Personal Genomes
  - ✓ Fundamentally identifiable in itself
  - ✓ Non-editability
  - ✓ Beyond person, shared by family members
- Life logging
- Bio-Banks and biomedical research
- Taxonomy for Secondary Uses


Glasses
Monitor eyesight
Advise eye correction
Tissues
Examine snot and mucus
Balms
Scan for cold sores
Toothbrush
Measure fluoride
Remember cavities & discoloration
Nicely you of bad breath
Razors
Plot the surface of the skin
Looking for acne
Shirts
Screen the microbiome on
forearms (We more than our
own cells)
A new field of
Underwearables
will integrate markers for early
detection of cancers or other
anomalies
Heart rate*
(Garmin)
Tongue Scrapers
Screen salivary microbes
(the oral microbiome)
Airports Scanners
Broadcast their results
to your phone
Pacemaker and
stent*
Broadcast data to
the cardiologist
plus ECG
Toilets
Test excrements both liquid and
solid
Feces will be graduad following
the Bristol Stool scale
Step counter*
(FITBIT)
Electrical Activity
Convey Alpha, Beta, Delta
and Theta Waves
Collars
Chemically analyze
sweat
Combs
Screen follicles
Report on Dandruff Density
Scan for fungus or Lice
Count the hairs (hair loss)
Earphones
Measure hearing
Analyze emotional level of
people we are listening to
(sound analysis already
allows this)
Clothes
Will be intelligent because
the fibers will compute, they
will visualize the body
language
Blood Pressure Measurement*
(HEALTHY)
Nail Cutters
Determine the quality of nails
Count ridges
Chairs
Plot our posture
Broadcast data for our spine


Eye
Glucose-sensing lens
Digital fundoscope
Smartphone visual-acuity tracking
Automated refractive error
Noninvasive intraocular pressure
Ear
Smart hearing aids
Digital otoscope
Lung
Home spirometry
Pulse oximetry
Inhaler use
Breath-based diagnostics
Breathing sounds
Environmental exposure
Blood
Continuous glucose
Transdermal Hb
Pathogens (genomics-based)
PoC blood tests
Skin
Temperature
Gross lesions
Pressure sensor (wound care)
Sweat chemistry
Cutaneous blood flow
Other sensors and monitors
Pill-box and -bottle
Posture
Body position
Activity
Sleep
Bladder and urine
Comprehensive urinalysis
STDs (genomic detection)
Diaper-based sensors
Brain and emotion
Wireless mobile EEG
Seizure
Autonomic nervous activity
Head-impact sensor
Intracranial pressure (noninvasive)
Stress recognition (voice, respiration)
Heart and vascular
Continuous BP tracking
Handheld ECG
Heart rhythm
Cardiac output
Stroke volume
Thoracic impedance (fluid)
Gastrointestinal
Endoscopic imaging
Esophageal pH
Medication compliance
Fecal blood or bilirubin
Gut electrical activity
Chewing
Watching over one's health
Pulse
BP
Temperature
Activity
Hydration
Sleep stages
Seizure
Respiration rate
O_2 saturation
Blood CO_2
Blood glucose
ECG (single-lead)
Cardiac output
Stroke volume
Stress:
Heart-rate variability
Electrodermal activity

Impact on Family

personal genetics education project (link)



HSLS, U.Pitt

Ethical and Technological


Retreat1
Who am we?
SNUBILife
Event
Album
B4GM
Schedule
SnuBBS
Research
Projects
Services
Resources
Google
Thank you!
http://www.snubi.org/