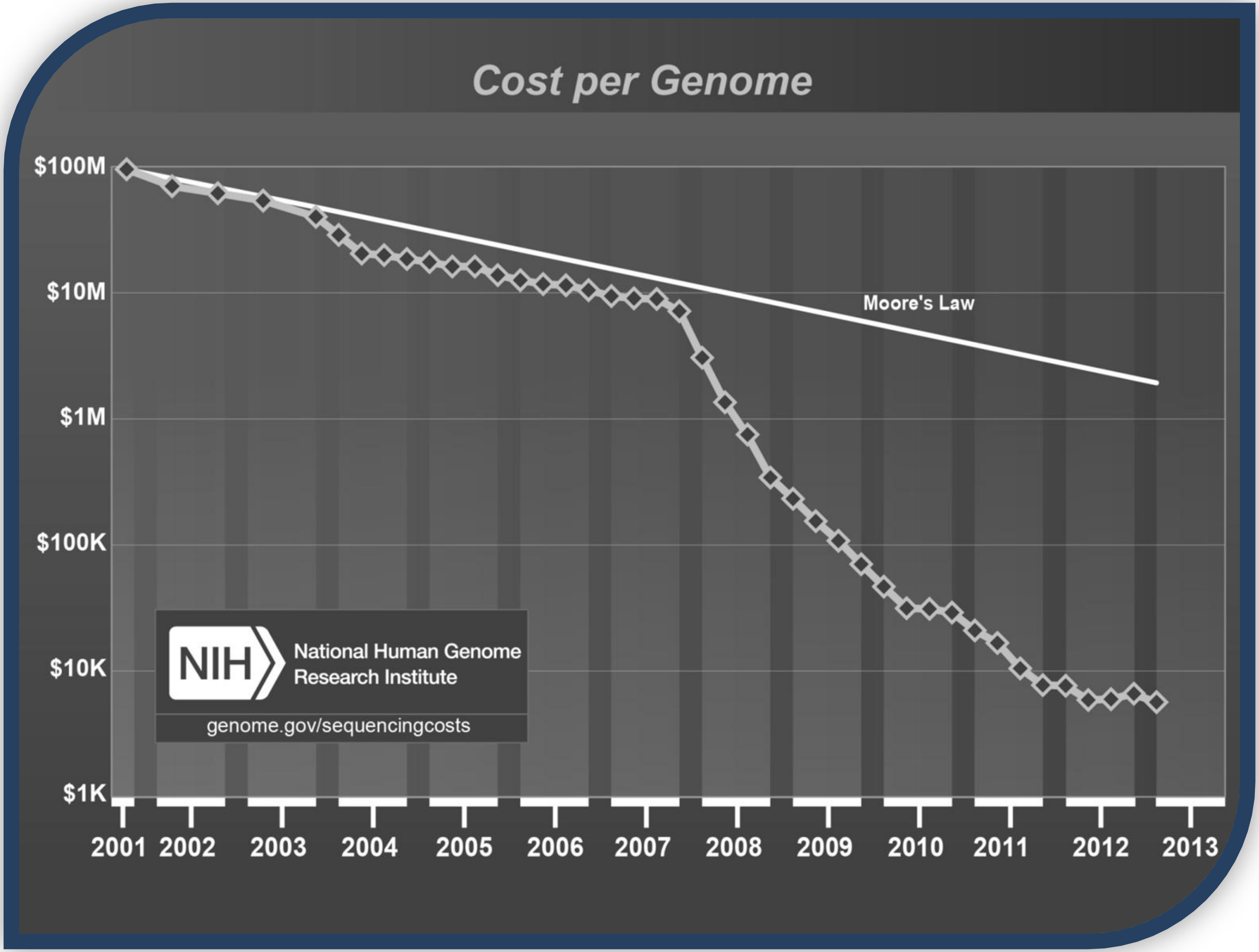
DNA privacy

Vinicius Vielmo Cogo

Cost per Genome



[1]

MinION?

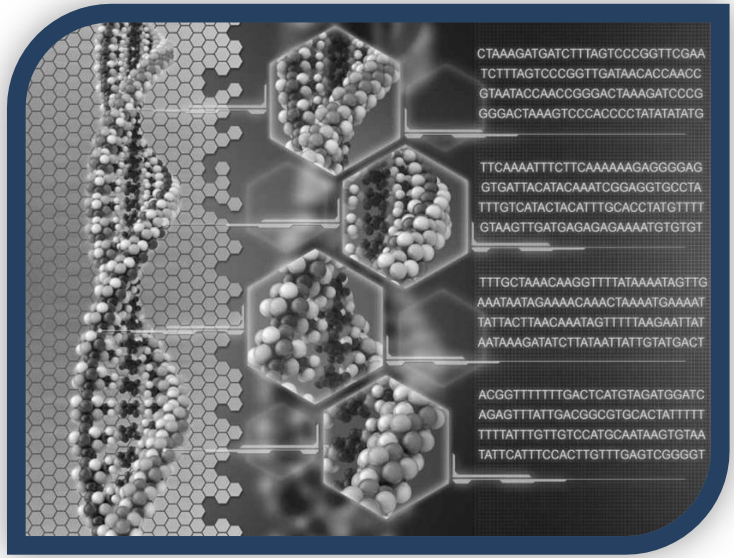
MinION

MinION

Oxford NANOPORE Technologies

[2]

Question 1



Would you like to have your DNA sequenced?

Why? (pros)

Personal purposes

- Print and place it in my bookshelf

DNA profiling

- Obtain future hypotheses and statistics

Diagnosis

- Confirm a disease
- Personalized medicine

Participate in a cohort study

- As a case or control

Contribute to general science

- Participating in future studies

Make my DNA public

- Contribute to any research

Why not? (cons)

Personal purposes

- I don't have an empty bookshelf at home
- I'm afraid by my privacy

DNA profiling

- I don't want to know my hypotheses
- I don't believe in fortune tellers

Diagnosis

- We don't know enough about DNA
- It's difficult to predict a disease based only in the DNA sequence

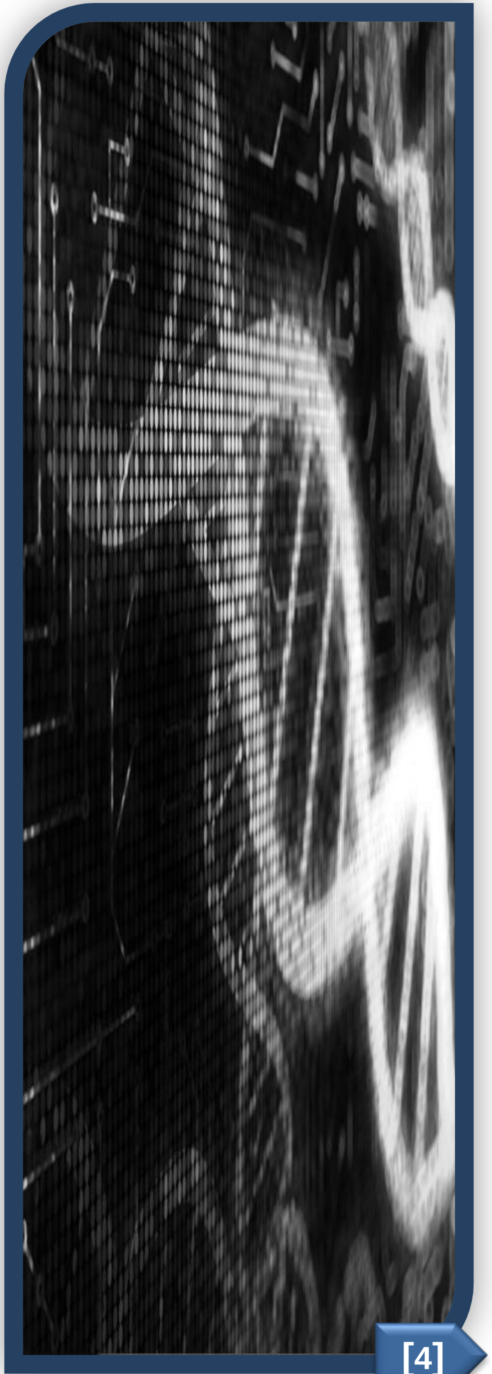
Other reasons

- Current high cost
- There is no way to safely store and analyze it

[3]

Risks

- Data leakage
- Cryptographic problems
- Aggregating information
- Targeted attacks
- Personal reputation
- Genetic discrimination



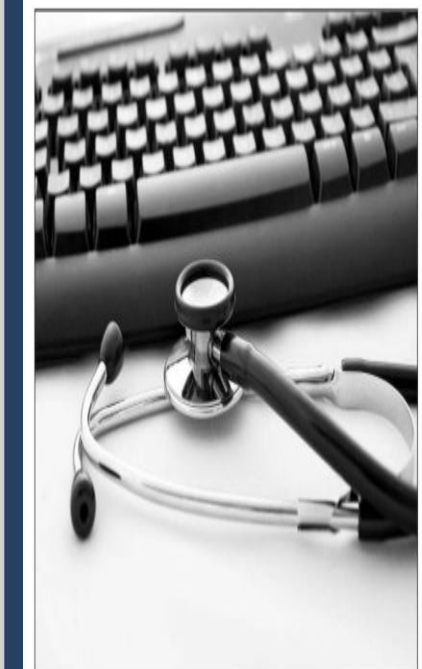
[4]

Data leakage

From

- A computer, laptop, tablet, smartphone
- A pen drive, hard disk, SD card
- A sequencing company that transcribed my genome
- A doctor or hospital that analyzed my profile
- A biobank that stores my genome
- An external hosting service for the hospital or biobank
- A researcher that was investigating the genome
- A bookshelf

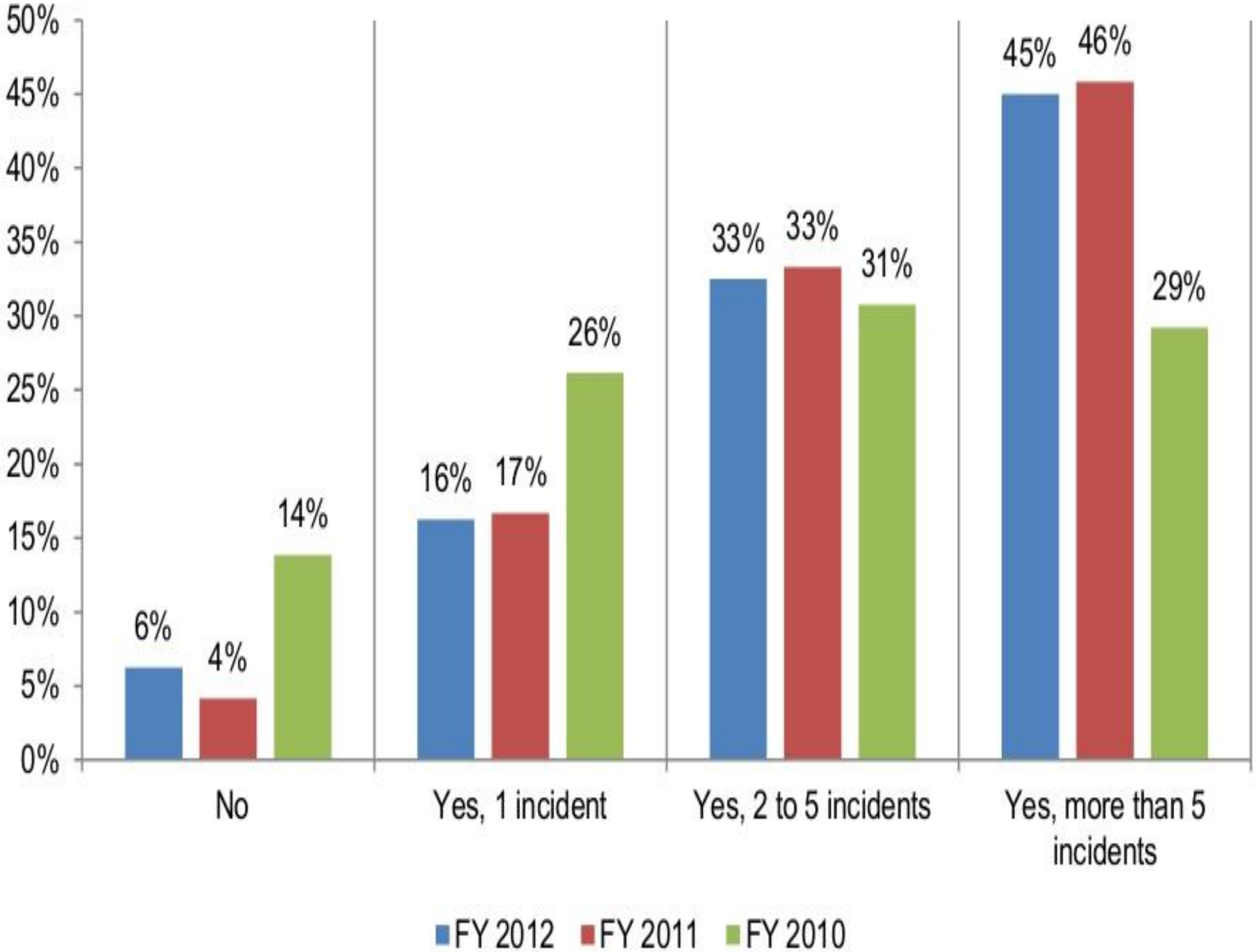
Ponemon INSTITUTE



Third Annual Benchmark Study on Patient Privacy & Data Security

Data leakage

Figure 1. Experienced a data breach involving the loss of patient data in the past two years



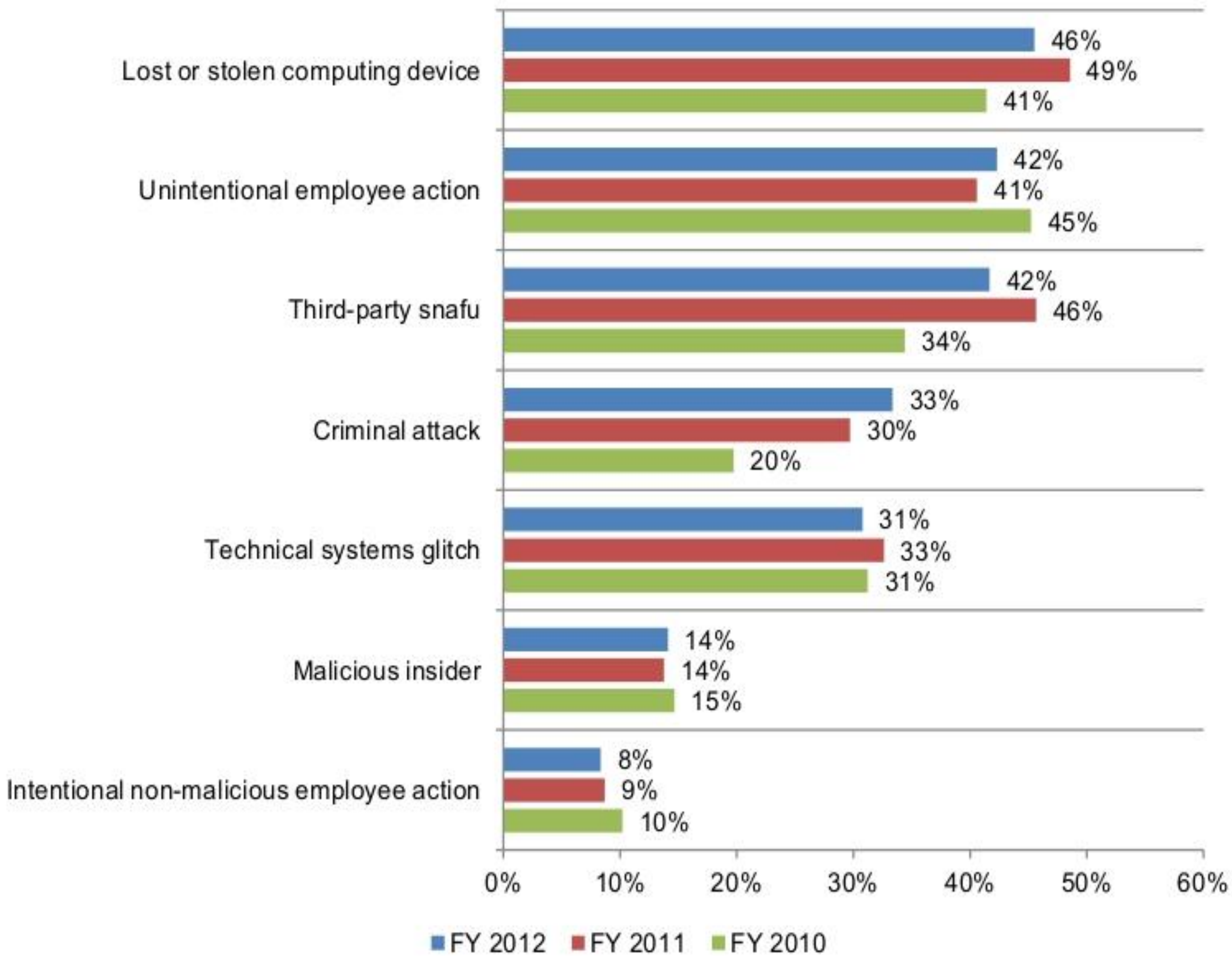
** 80 health care organizations participated in this study*

Ponemon INSTITUTE

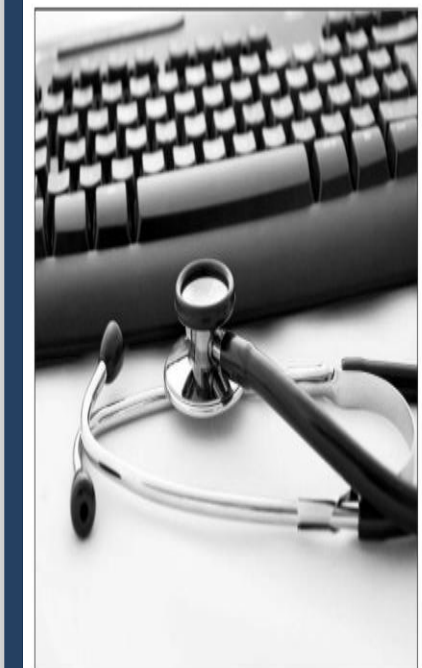
Third Annual Benchmark Study on Patient Privacy & Data Security

Data leakage

Figure 5. Nature of the incident
More than one choice permitted



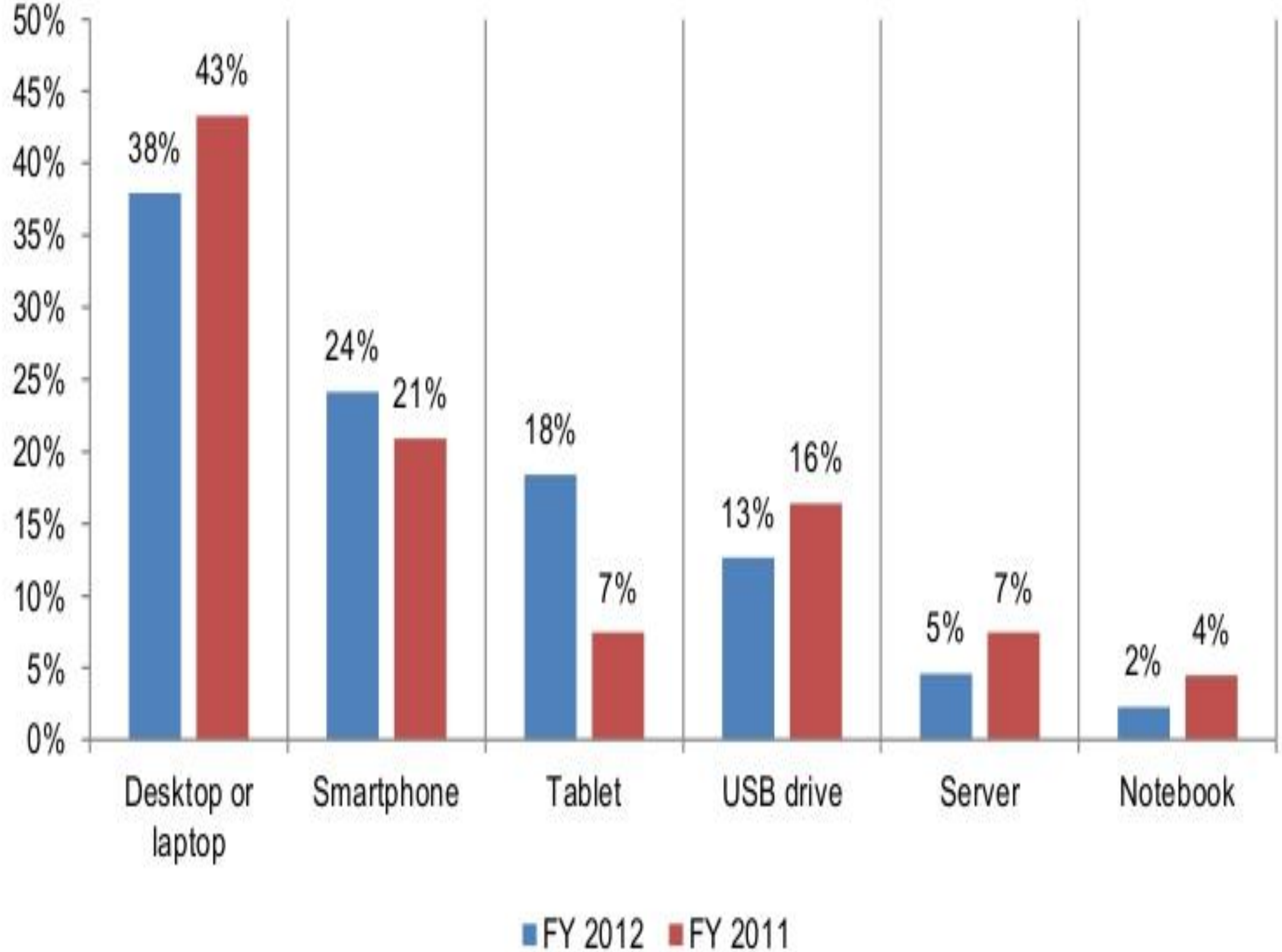
Ponemon INSTITUTE



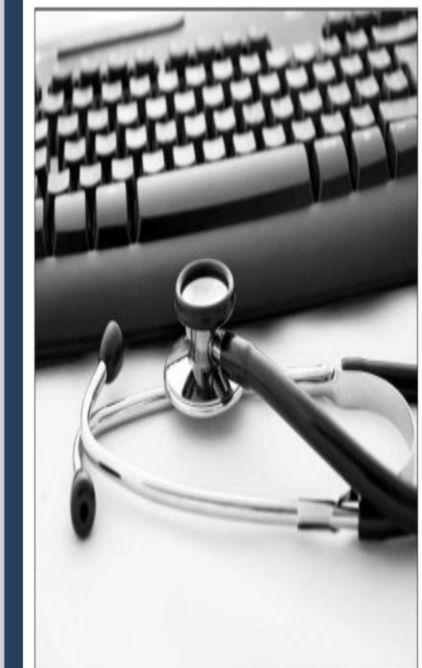
Third Annual Benchmark Study on Patient Privacy & Data Security

Data leakage

Figure 6. Type of device compromised or stolen



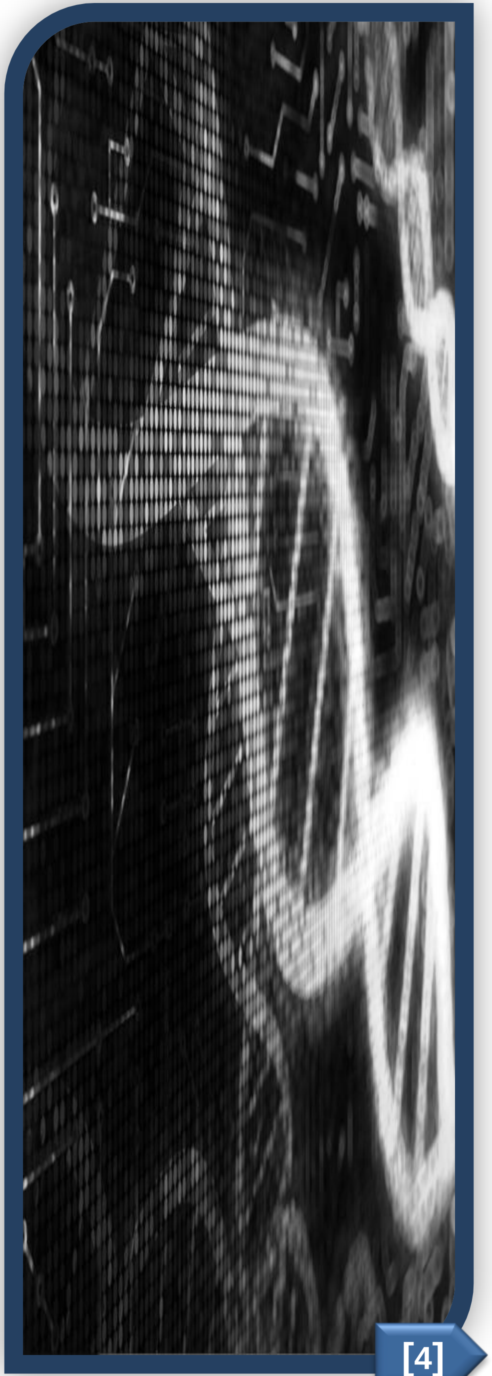
Ponemon INSTITUTE



Third Annual Benchmark Study on Patient Privacy & Data Security

[5]

Cryptographic problems



Is it a solution to data leakage?

- Yes, because encrypted data is meaningless without encryption keys

But, there are problems

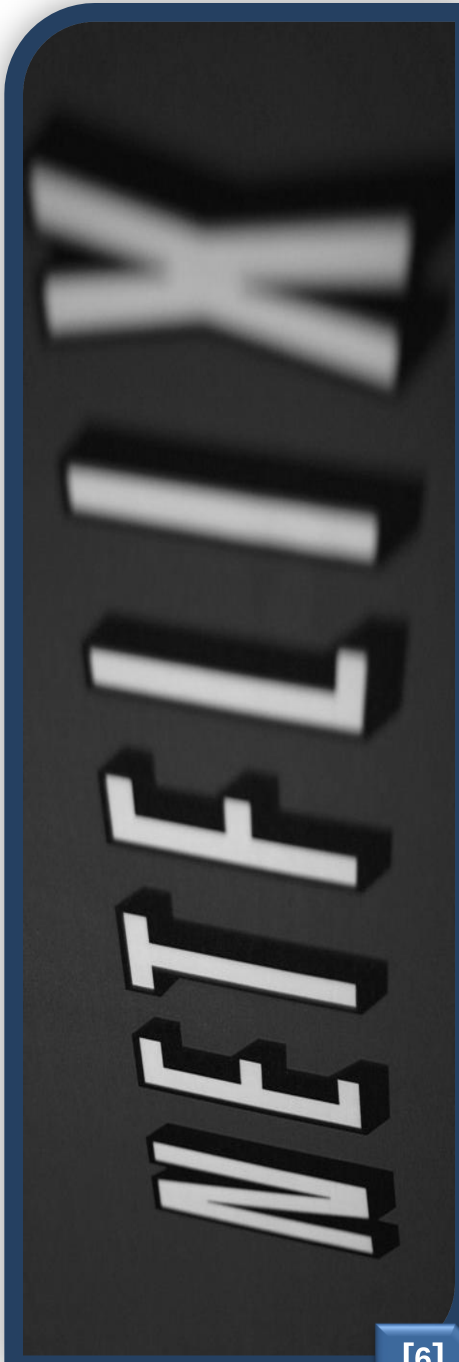
- Computational cost is higher than plain text
- Leakage of cryptographic keys still is possible
- Algorithms vulnerabilities or limitations
- A cryptographic solution may expire with time, but the sensitivity of DNA doesn't

[4]

Aggregating information

The Netflix prize case

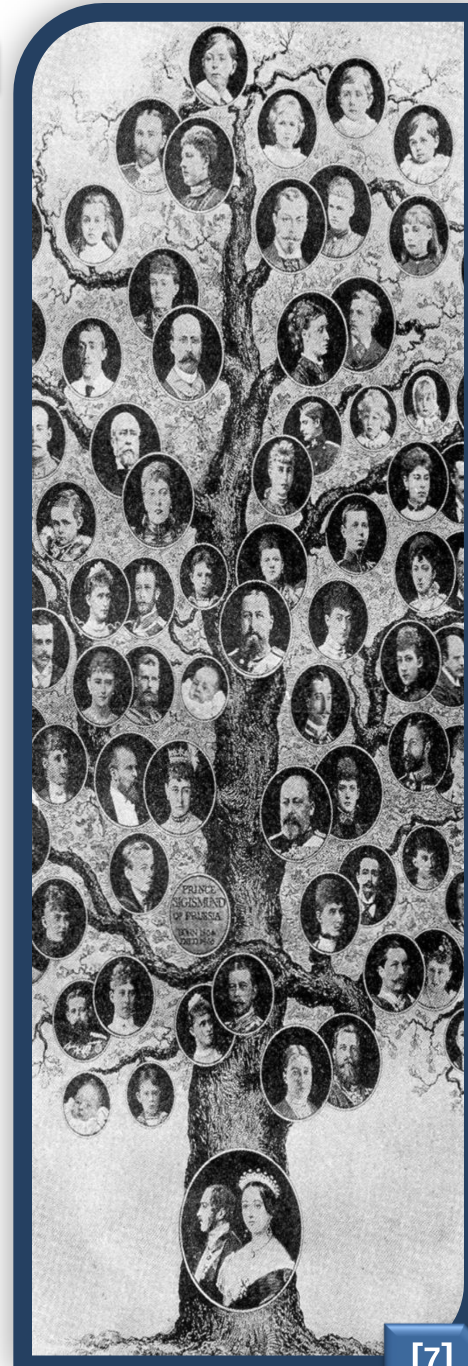
- Data mining contest
- US$ 1-million prize for improvements on movie suggestion algorithms
- Netflix released anonymous preferences of 500 000 subscribers
- Researchers tried to identify subscribers based on complementary public information from IMDB
- They identified 99% of cases with 6-8 movie ratings on database
- Removing the identifying information is not enough for anonymity.



Aggregating information

Identifying personal genomes by surname inference

- Based in short tandem repeats on the Y chromosome (microsatellites)
- Aggregated information from recreational genetic genealogy databases
- Identified the surname of 12% of publicly available genomes from 1000 Genome Project
- Retrieving other metadata (as age and state) they were able to triangulate the identity of the target



[7]

Targeted attacks

Scenarios

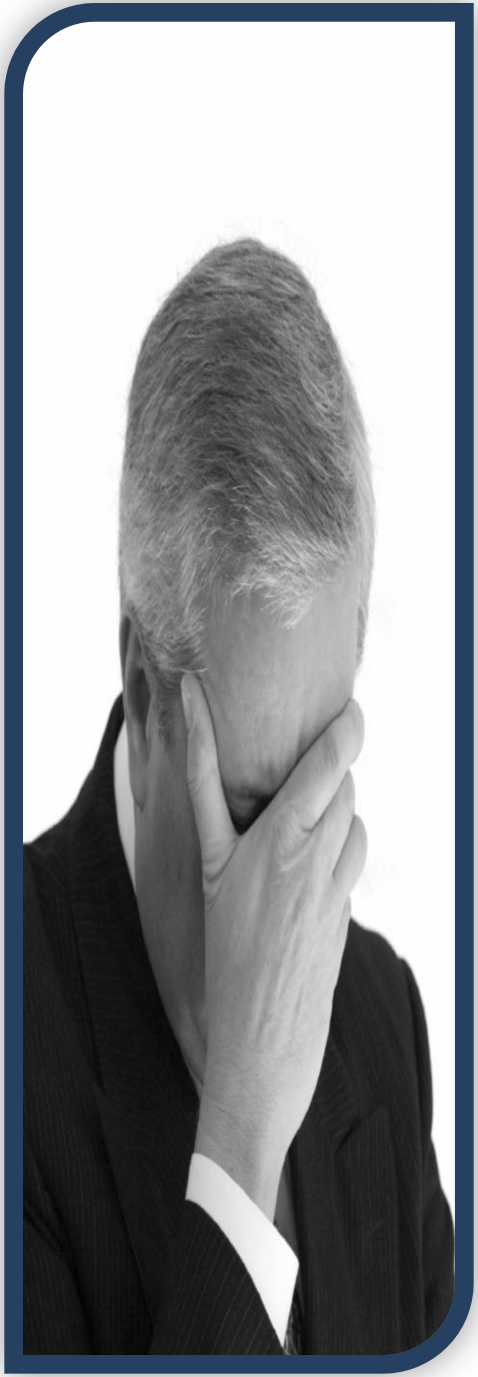
- Learn something about a specific individual or group
- Insert some microorganism or disease in a population
- Bioweapons (e.g. chimera viruses)

[8]

Loss of reputation

Scenarios

- Individual reputation
- Family reputation
- Reputation of sequencing company
- Reputation of a country



Genetic discrimination

Scenarios

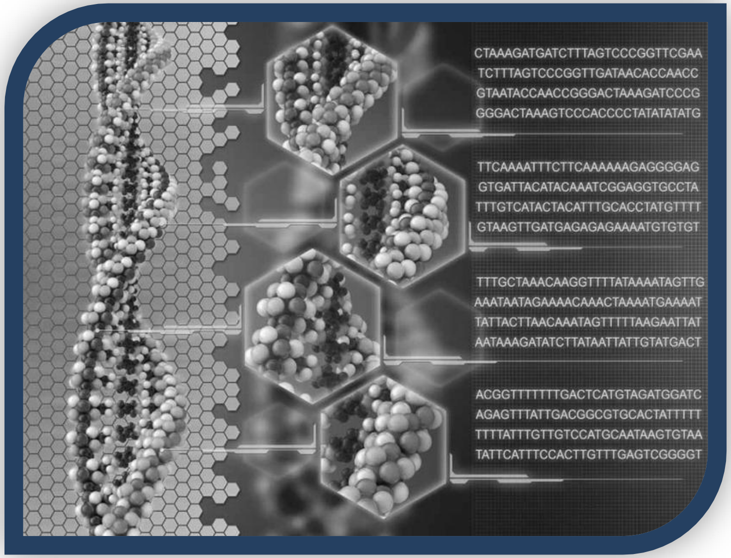
- Health insurance
- Employment
- Education
- Travel and visa

HIPAA

- US act from 1996
- Protecting individuals against genetic discrimination
- Applied to health insurance and employment scenarios
- Individual safety justifies genetic discrimination?

[9]
[10]

Question 2



Would you like to have your DNA sequenced?

What is the message?

- DNA is our ultimate identifier
- There are always privacy risks in sequencing your DNA
- Science needs DNA sequences
- We must define new ways to improve the security on DNA storing and processing

Further information

[1] [DNA sequencing costs]

[2] [Oxford Nanopore Technologies]

[3] [Why I don't want to know my genome sequence]

[4] [Whole Genome Sequencing:
Innovation dream or privacy nightmare?]

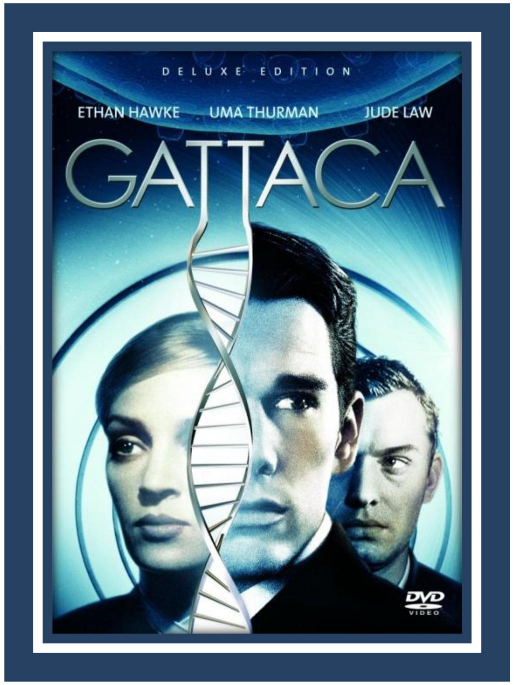
[5] [Study on patient privacy and data security]

[6] [How to Break Anonymity of the Netflix Prize Dataset]

[7] [Identifying Personal Genomes by Surname Inference]

[8] [Hacking the President's DNA (The Atlantic Mag.)]

[9]



[10]

BIOBANKCLOUD
Your PaaS for Biobanking

http://www.biobankcloud.eu/