

Genomic Testing and Privacy Issues

Emiliano De Cristofaro
University College London
<https://emilianodc.com>

Who Am I?

- Reader in Security & Privacy Technology, UCL Computer Science (2013–)
- Head of Information Security Research Group (2018–)
- Faculty Fellow at the Alan Turing institute (2018–)
- Research Scientist, Xerox PARC (2011–2013)
- PhD in Computer Science, University of California (2011)
- Cybersecurity
- Privacy Enhancing Technologies
- Privacy and Ethical Issues in AI
- Data-Driven measurements of “interesting” issues on the Web

Privacy in Computer Science

Privacy in Computer Science

Mostly defined as a set of **information security** properties...

Privacy in Computer Science

Mostly defined as a set of **information security** properties...

Confidentiality: keeping a user's data secret

Privacy in Computer Science

Mostly defined as a set of **information security** properties...

Confidentiality: keeping a user's data secret

Control: giving control to the individual about the use of their personal information

Privacy in Computer Science

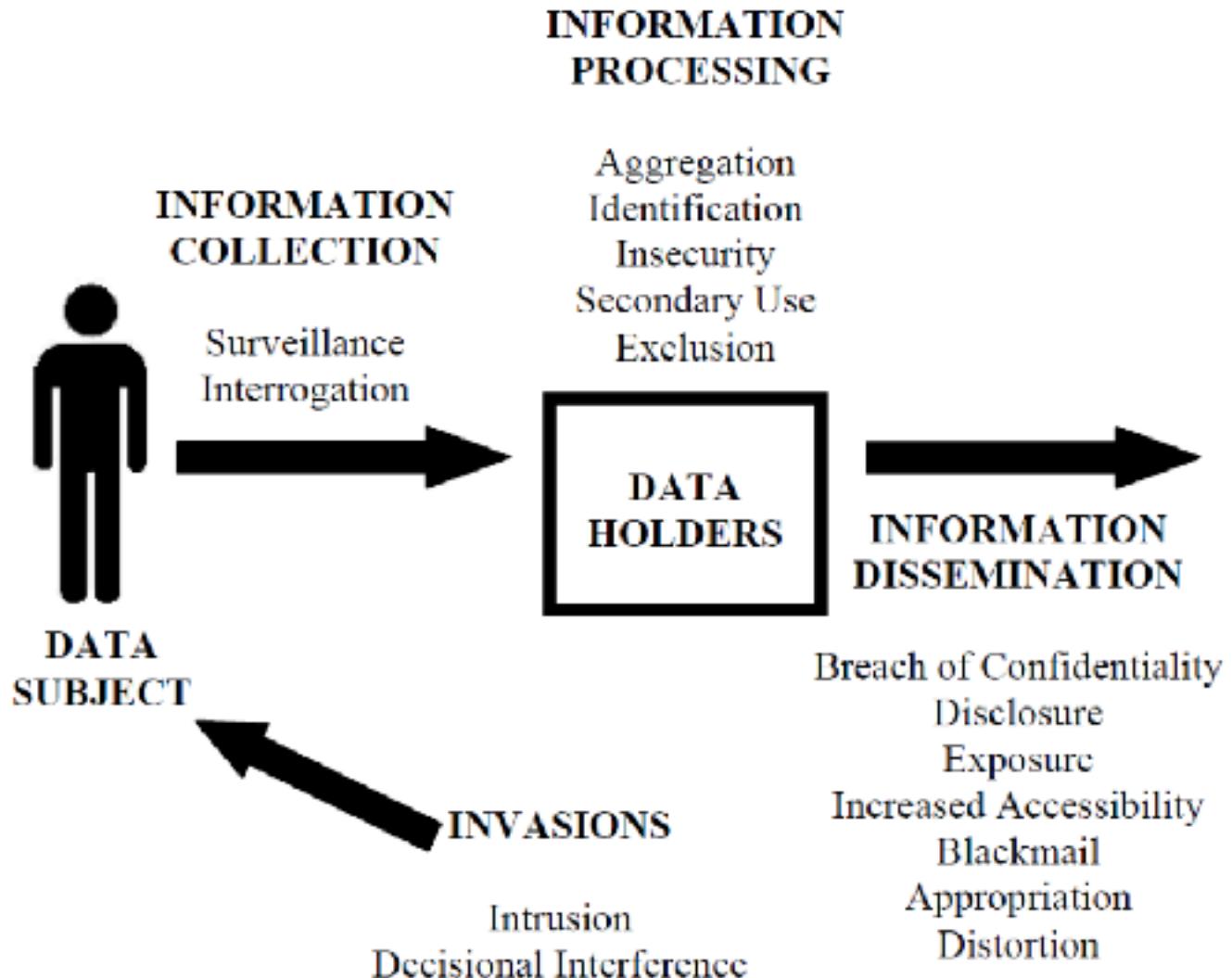
Mostly defined as a set of **information security** properties...

Confidentiality: keeping a user's data secret

Control: giving control to the individual about the use of their personal information

Self-actualization: allowing the individual to use their information environment to further their own aims

Taxonomy of Privacy Harms



Example Privacy Harms (Solove)

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Reporters deceitfully gain entry to a person's home and secretly photograph and record the person

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Reporters deceitfully gain entry to a person's home and secretly photograph and record the person

New X-ray devices can see through people's clothing, amounting to what some call a "virtual strip-search"

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Reporters deceitfully gain entry to a person's home and secretly photograph and record the person

New X-ray devices can see through people's clothing, amounting to what some call a "virtual strip-search"

The government uses a thermal sensor device to detect heat patterns in a person's home

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Reporters deceitfully gain entry to a person's home and secretly photograph and record the person

New X-ray devices can see through people's clothing, amounting to what some call a "virtual strip-search"

The government uses a thermal sensor device to detect heat patterns in a person's home

A company markets a list of five million elderly incontinent women

Example Privacy Harms (Solove)

A newspaper reports the name of a rape victim

Reporters deceitfully gain entry to a person's home and secretly photograph and record the person

New X-ray devices can see through people's clothing, amounting to what some call a "virtual strip-search"

The government uses a thermal sensor device to detect heat patterns in a person's home

A company markets a list of five million elderly incontinent women

Despite promising not to sell its members' personal information to others, a company does so anyway

Ex

GCHQ data collection violated human rights, Strasbourg court rules

A newspaper re

Reporters dece
record the pers

New X-ray devi
call a “virtual st

The governmer
person's home

A company ma

Despite promis
company does

Spies breached right to privacy in programme revealed by Edward Snowden, judges say



▲ The GCHQ building in Cheltenham. Photograph: GCHQ/PA

Two families of privacy technologies

Two families of privacy technologies

Soft Privacy Technologies

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Hard Privacy Technologies

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Hard Privacy Technologies

Stronger focus on data minimization

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Hard Privacy Technologies

Stronger focus on data minimization

Assumption: there exists no single third party that may be trusted with user data

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Hard Privacy Technologies

Stronger focus on data minimization

Assumption: there exists no single third party that may be trusted with user data

Threat model: a service is in the hands of the adversary; may be coerced; may be hacked.

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

Hard Privacy Technologies

Stronger focus on data minimization

Assumption: there exists no single third party that may be trusted with user data

Threat model: a service is in the hands of the adversary; may be coerced; may be hacked.

k-out-of-n honest third parties

Two families of privacy technologies

Soft Privacy Technologies

Focus on compliance and “internal controls”

Assumption: a third party is entrusted with the user data

Threat model: third party is trusted to process user data according to user wishes

E.g.: Tunnel encryption (SSL/TLS)

“Keeping honest services safe from insiders / employees”

Hard Privacy Technologies

Stronger focus on data minimization

Assumption: there exists no single third party that may be trusted with user data

Threat model: a service is in the hands of the adversary; may be coerced; may be hacked.

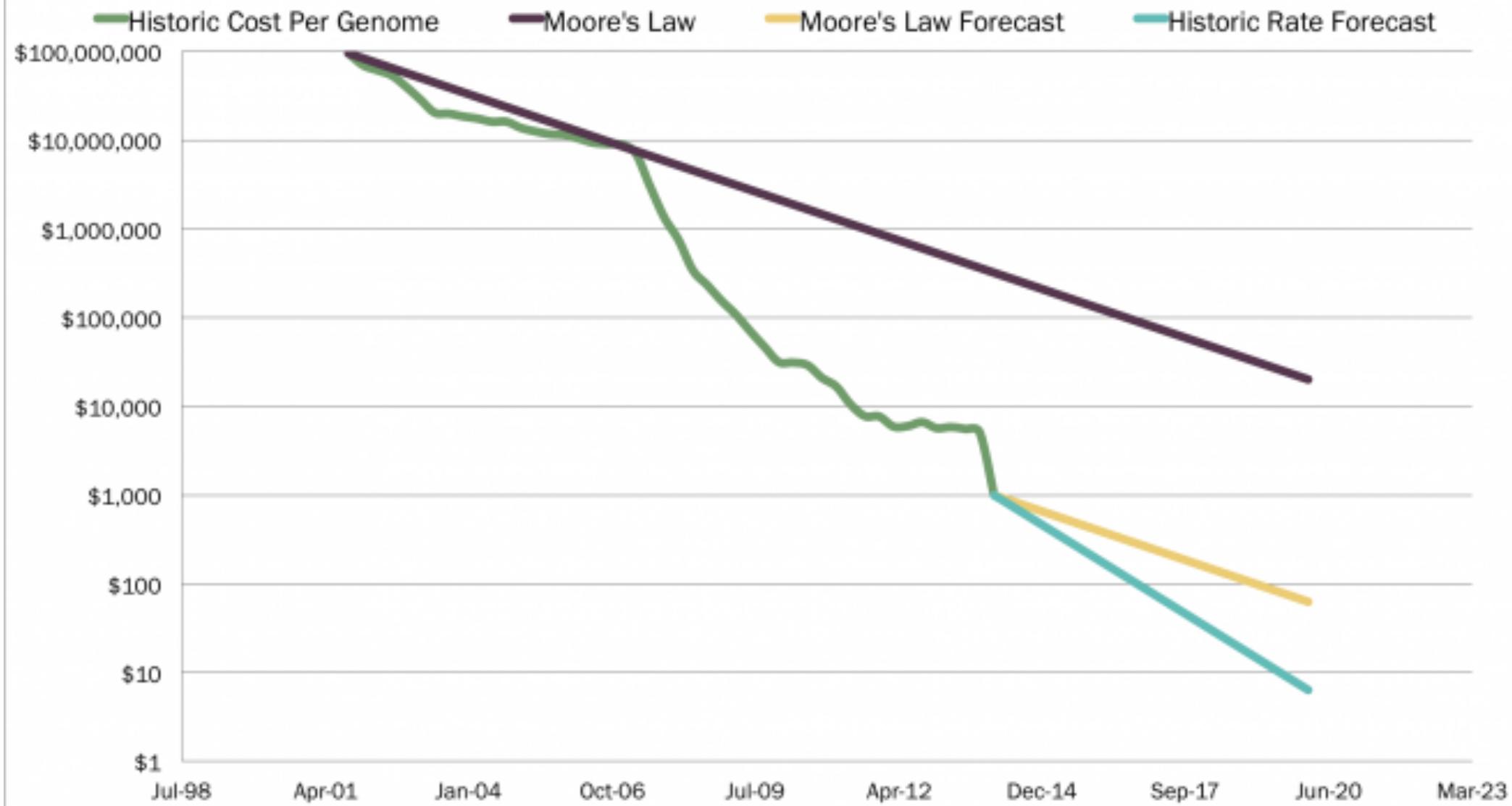
k-out-of-n honest third parties

E.g., Tor

May relay on service integrity if auditing is possible

A Case Study: Genomic Testing

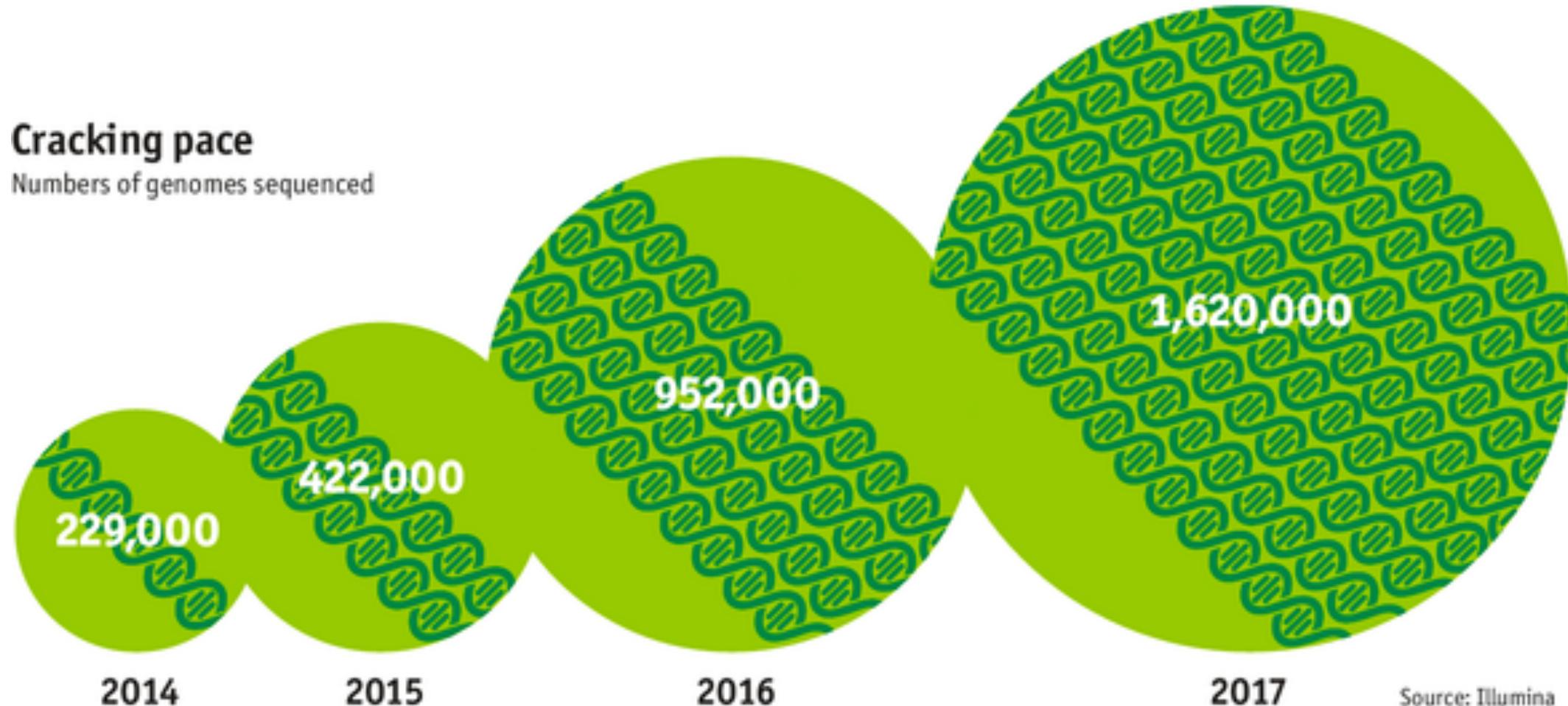
Cost Declines of Genome Sequencing



From: James Bannon, ARK

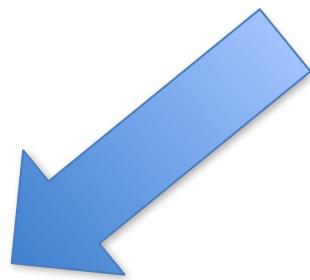
Cracking pace

Numbers of genomes sequenced



From: The Economist

How to read the genome?



Genotyping

Testing for genetic differences using a set of markers



Sequencing

Determining the full nucleotide order of an organism's genome

The First Child Saved By DNA Sequencing

[+ Comment Now](#) [+ Follow Comments](#)



The First Child Saved By DNA Sequencing

[+ Comment Now](#) [+ Follow Comments](#)



Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability

Farah R. Zahir , Jill C. Mwenifumbo , Hye-Jung E. Chun , Emilia L. Lim , Clara D. M. Van Karnebeek , Madeline Couse , Karen L. Mungall , Leora Lee , Nancy Makela , Linlea Armstrong , Cornelius F. Boerkoel , Sylvie L. Langlois , Barbara M. McGillivray , Steven J. M. Jones , Jan M. Friedman [†] and Marco A. Marra [†]

BMC Genomics 2017 18:403

<https://doi.org/10.1186/s12864-017-3671-0> | © The Author(s). 2017

Received: 4 November 2016 | Accepted: 29 March 2017 | Published: 24 May 2017

The First Child Saved By DNA Sequencing

[+ Comment Now](#) [+ Follow Comments](#)



Genomics promises a leap forward for rare disease diagnosis

Faster and cheaper DNA sequencing brings new hope to patients



Jessica suffers from a rare condition that was diagnosed through DNA analysis

Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability

Farah R. Zahir , Jill C. Mwenifumbo , Hye-Jung E. Chun , Emilia L. Lim , Clara D. M. Van Karnebeek , Madeline Couse , Karen L. Mungall , Leora Lee , Nancy Makela , Linlea Armstrong , Cornelius F. Boerkoel , Sylvie L. Langlois , Barbara M. McGillivray , Steven J. M. Jones , Jan M. Friedman [†] and Marco A. Marra [†]

7 18:403

[186/s12864-017-3671-0](#) | © The Author(s). 2017

er 2016 | Accepted: 29 March 2017 | Published: 24 May 2017

The First Child Saved By DNA Sequencing

[+ Comment Now](#) [+ Follow Comments](#)



Genomics promises a leap forward for rare disease diagnosis

Faster and cheaper DNA sequencing brings new hope to patients



Jessica suffers from a rare condition that was diagnosed through DNA analysis

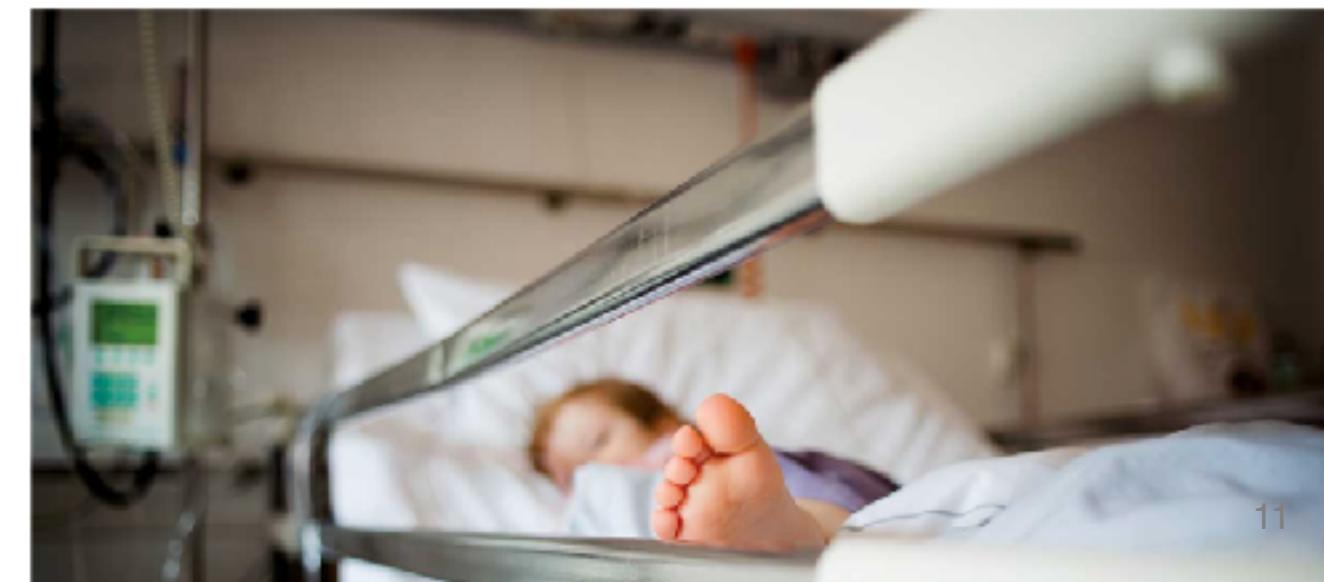
Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability

Farah R. Zahir , Jill C. Mwenifumbo , Hye-Jung E. Chun , Emilia L. Lim , Clara D. M. Van Karnebeek , Madeline Couse , Karen L. Mungall , Leora Lee , Nancy Makela , Linlea Armstrong , Cornelius F. Boerkoel , Sylvie L. Langlois , Barbara M. McGillivray , Steven J. M. Jones , Jan M. Friedman [†] and Marco A. Marra [†]

[Home](#) | [News](#) | [Health](#)

THIS WEEK 26 March 2011

Three critically ill children helped by speedy genome sequencing



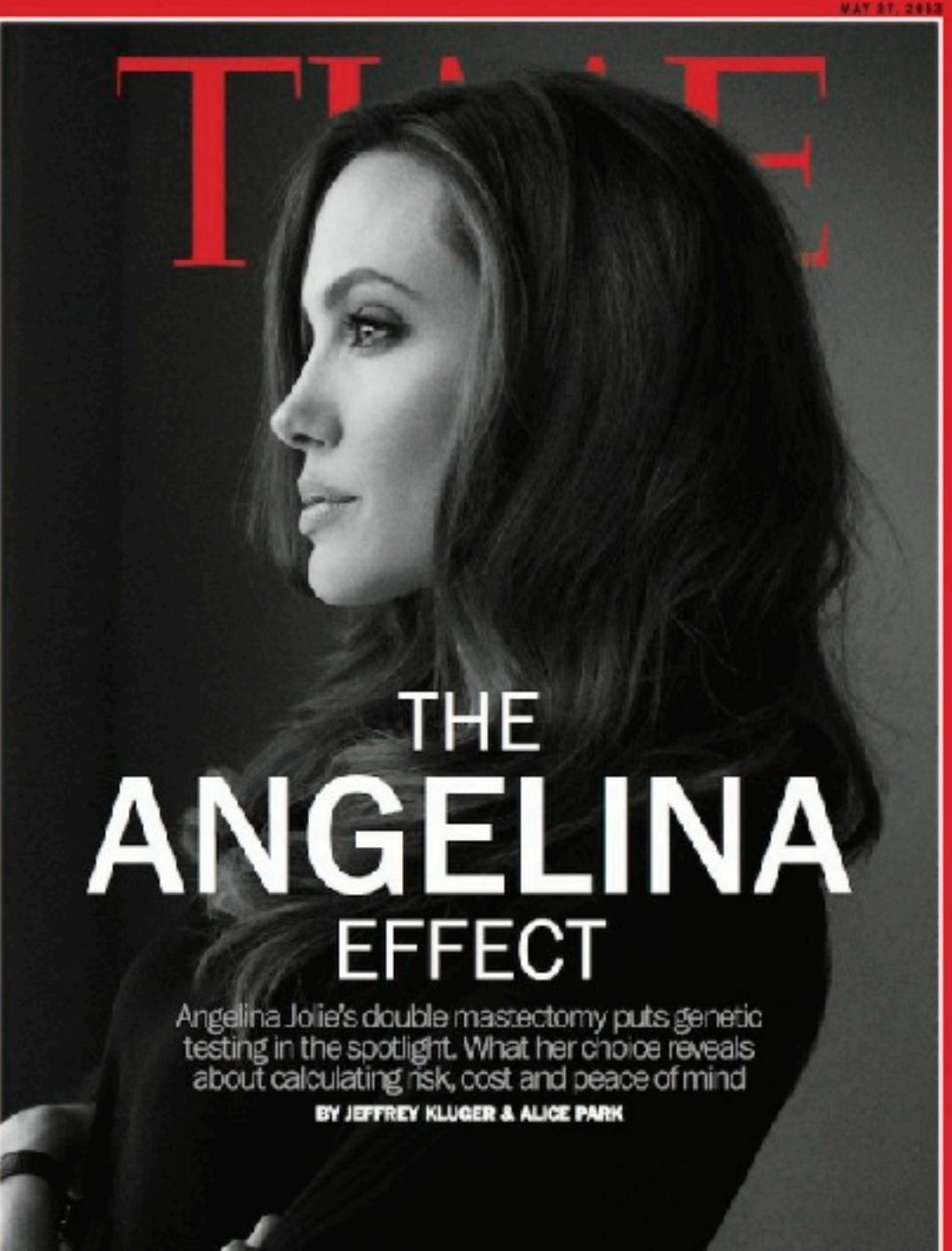
'Angelina Jolie effect' boosted genetic testing rates, study suggests

Actor's call for women to seek testing for breast and ovarian cancer mutations raised screening rates but may not have reached those most at risk



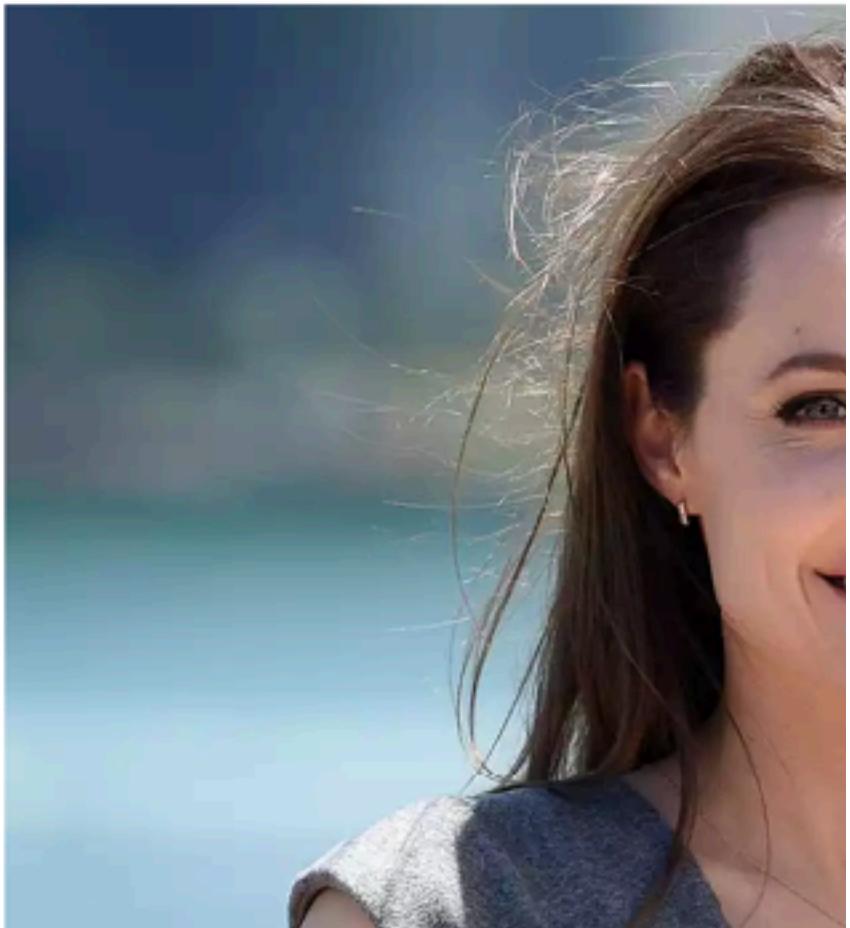
'Angelina Jolie' testing rates, st

Actor's call for women to seek mutations raised screening rates at risk



'Angelina Jolie effect' testing rates, study su

Actor's call for women to seek testing for t mutations raised screening rates but may at risk



Angelina Jolie gene testing for all?

By James Gallagher
Health and science correspondent, BBC News

18 January 2018 | [Comment](#)

[Share](#)



GETTY IMAGES

Testing all women for the "Angelina Jolie gene", even if not considered at risk, would prevent cancers, save lives and is cost effective, say doctors.

Genetic Risk Factors (11)

REPORT	RESULT
Alpha-1 Antitrypsin Deficiency	Variant Absent; Typical Risk
Alzheimer's Disease (APOE Variants)	e4 Variant Absent
Early-Onset Primary Dystonia (DYT1-TOR1A-Related)	Variant Absent; Typical Risk
Factor XI Deficiency	Variant Absent; Typical Risk
Familial Hypercholesterolemia Type B (APOB-Related)	Variant Absent; Typical Risk

[See all 11 genetic risk factors...](#)

Traits (41)

REPORT	RESULT
Alcohol Flush Reaction	Does Not Flush
Bitter Taste Perception	Can Taste
Blond Hair	28% Chance
Earwax Type	Wet
Eye Color	Likely Brown

[See all 41 traits...](#)

Inherited Conditions (43)

REPORT	RESULT
Beta Thalassemia	Variant Present
ARSACS	Variant Absent
Agenesis of the Corpus Callosum with Peripheral Neuropathy (ACCPN)	Variant Absent
Autosomal Recessive Polycystic Kidney Disease	Variant Absent
Bloom's Syndrome	Variant Absent

[See all 43 carrier status...](#)

Drug Response (12)

REPORT	RESULT
Proton Pump Inhibitor (PPI) Metabolism (CYP2C19-related)	Rapid
Warfarin (Coumadin®) Sensitivity	Increased
Phenytoin Sensitivity (Epilepsy Drug)	Increased
Sulfonylurea Metabolism	Greatly reduced
Abacavir Hypersensitivity	Typical

[See all 12 drug response...](#)

Genetic Risk Factors (11) ?

REPORT	RESULT
Alpha-1 Antitrypsin Deficiency	Variant Absent; Typical Risk
Alzheimer's Disease [APOE Variants]	ε4 Variant Absent
Early-Onset Primary Dystonia (DYT1-TOR1A-Related)	Variant Absent; Typical Risk
Factor XI Deficiency	Variant Absent; Typical Risk
Familial Hypercholesterolemia Type B (APOB-Related)	Variant Absent; Typical Risk

[See all 11 genetic risk factors...](#)

Traits (41) ?

REPORT	RESULT
Alcohol Flush Reaction	Does Not Flush
Bitter Taste Perception	Can Taste
Blond Hair	28% Chance
Earwax Type	Wet
Eye Color	Likely Brown

[See all 41 traits...](#)

Inherited Conditions (43) ?

REPORT
Beta Thalassemia
ARSACS
Agenesis of the Corpus Callosum with Peripheral Neuropathy (ACCPN)
Autosomal Recessive Polycystic Kidney Disease
Bloom's Syndrome

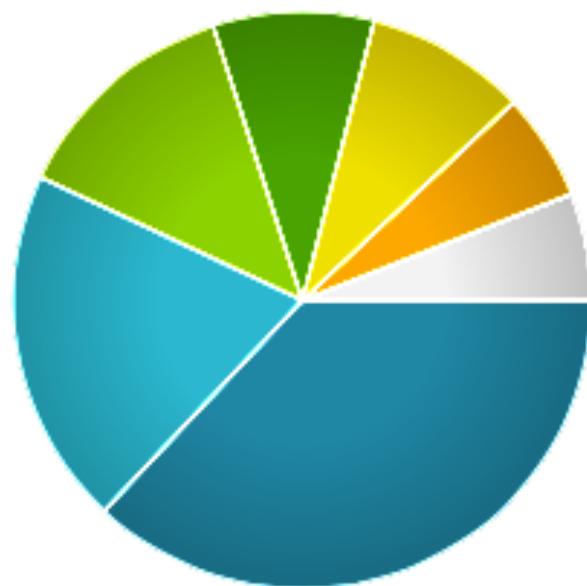
[See all 43 conditions...](#)

Drug Response (12) ?

REPORT
Proton Pump Inhibitor (PPI) Metabolism (CYP2C-related)
Warfarin (Coumadin®) Sensitivity
Phenytoin Sensitivity (Epilepsy Drug)
Sulfonylurea Metabolism
Abacavir Hypersensitivity

[See all 12 drug responses...](#)

Genetic Ethnicity



List View

Map View

Surname View

search matches

Show: both sides

Sort: relationship

25 per page

1 – 25 of 424



Male

You

UPDATE YOUR PROFILE



Female

2nd to 3rd
Cousin
1.66% shared, 5
segments

J2a2

Send an Introduction



Female

3rd to 4th
Cousin
1.30% shared, 3
segmentsUnited States | Alsace-Lorraine (Strasbourg), France | Personal
[REDACTED] | Surname | 5 more | USb2Public Match
Send a Message

Male

3rd to 4th
Cousin
1.03% shared, 2
segments

H10a1a | R1b1c2

Send an Introduction



Female

3rd to 5th
Cousin
0.45% shared, 2
segments

H7

Send an Introduction



Female

3rd to 5th
Cousin
0.42% shared, 2
segments

H1

Send an Introduction



Male

3rd to 5th
Cousin
0.40% shared, 2
segmentsUnited States | Reno, Nevada | San Diego, California
Tucker | Littlefield | Warga | 4 more | H1c | G2aPublic Match
Send a Message

Male

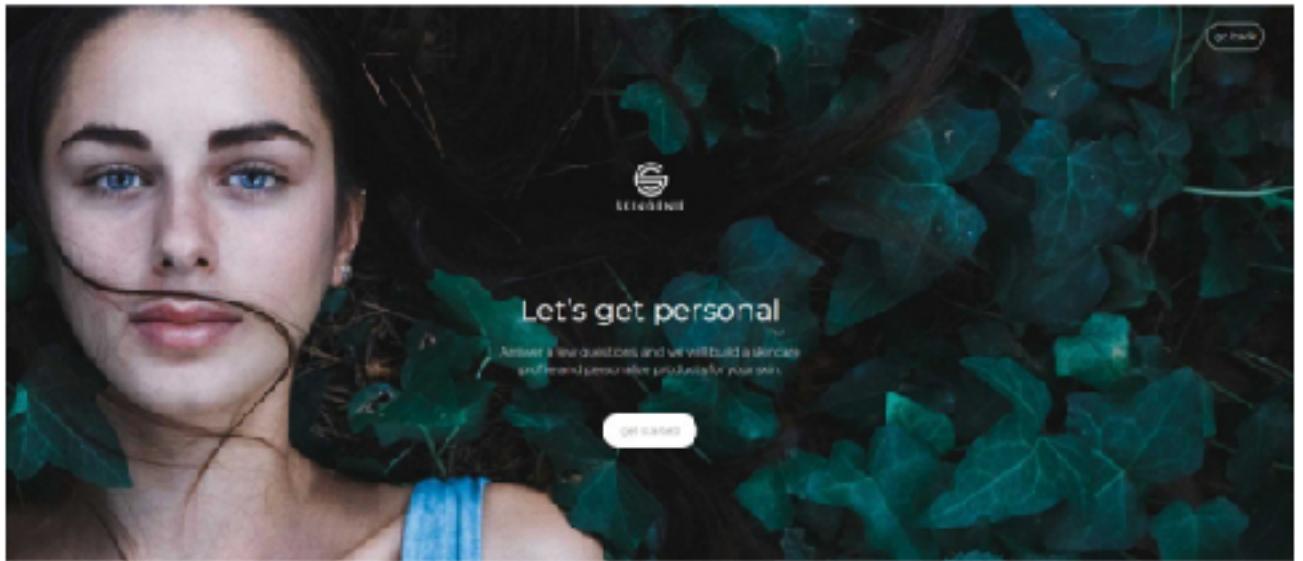
3rd to 5th
Cousin
0.37% shared, 2
segmentsUnited States | Father: father: grandfather: Edward [REDACTED]
[REDACTED] | [REDACTED] | [REDACTED] | K1atb
R1b1b2a1aPublic Match
Send a Message

Male, b. 1978

3rd to 6th
Cousin
0.40% shared, 1
segmentUnited States | New Jersey | Utah | California
Northern Europe | U9b1 | T

Send an Introduction

SkinGenie



SKINGENIE

HOME | SKIN CARE | WINE ACCESSORIES

WINE ACCESSORIES



VINOME

SpareRoom



SPAREROOM



SkinGenie

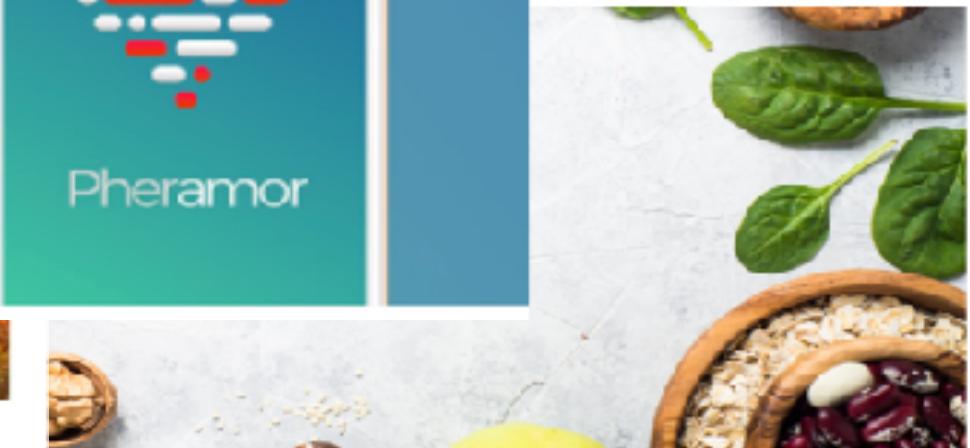


Nice genes, wanna date? This new dating app uses your DNA to find matches

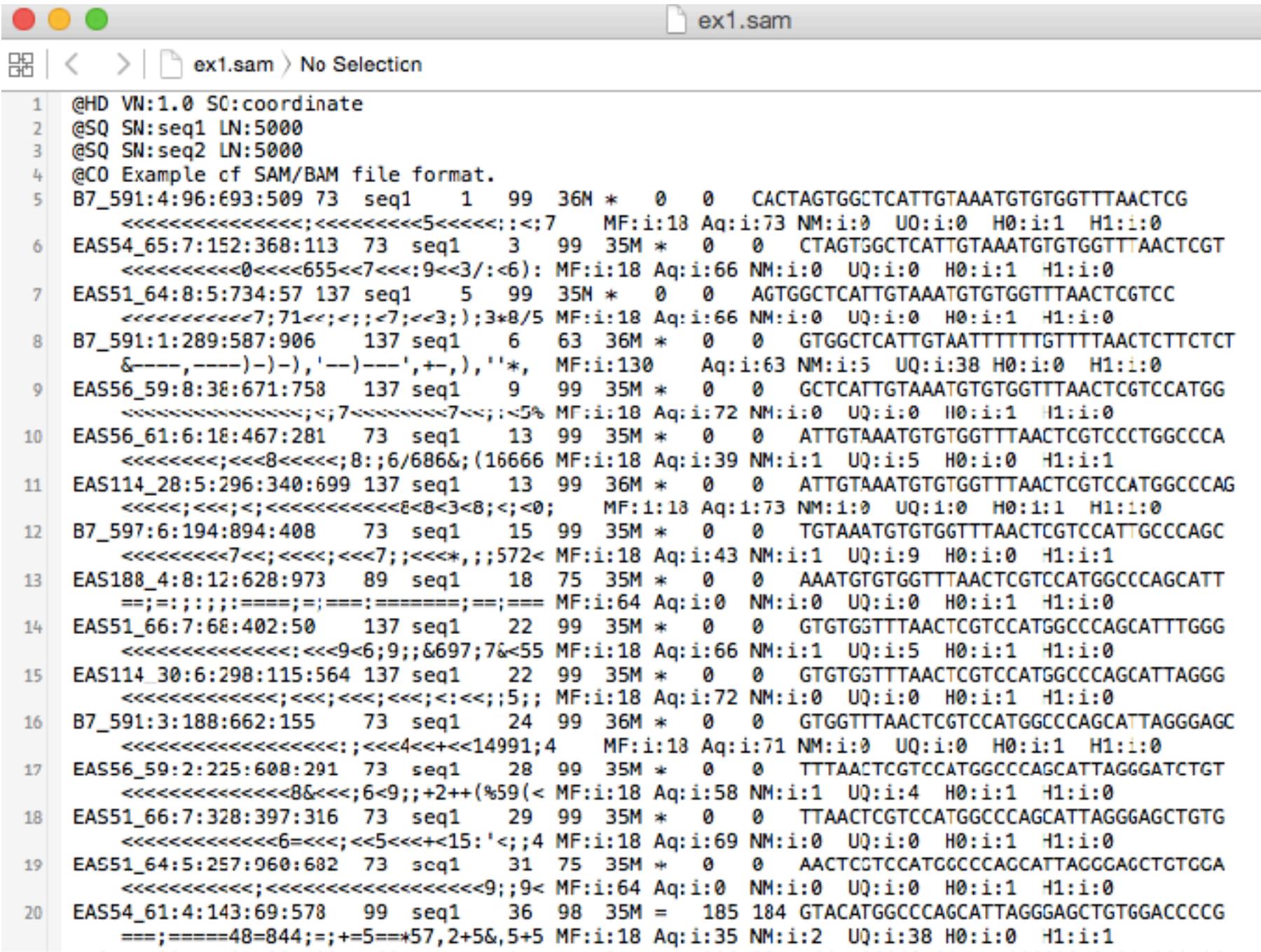


VINOME

SpareRoom







**But... not all data is
created equal!**

Health Data Hacking

Health Data Hacking

Anthem Hacking Points to Security Vulnerability of Health Care Industry

By REED ABELSON and MATTHEW GOLDSTEIN FEB. 5, 2015



The New York Times



An Anthem Health Insurance facility in Indianapolis. Hackers gained access to about 80 million company records, and somehow the stolen data will be used for identity theft. Austin P. Harrington/Getty Images

Health Data Hacking

Anthem Hacking Points to Security Vulnerability of Health Care Industry

By REED ABELSON and MATTHEW GOLDSTEIN FEB. 5, 2015



The New York Times

Anthem: one of US largest health insurers

60 to 80 million *unencrypted* records stolen in the hack (revealed in February 2015)

Social security numbers, birthdays, addresses, email and employment information and income data for customers and employees, including its own chief executive



An Anthem Health Insurance facility in Indianapolis. Hackers gained access to about 80 million company records, and somehow the stolen data will be used for identity theft. AP Photo/AJ Mast/Globe Staff

US Healthcare “Wall of Shame”

US Healthcare “Wall of Shame”

Around 2 declared breaches per week, each affecting 500+ people

US Healthcare “Wall of Shame”

Around 2 declared breaches per week, each affecting 500+ people

https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

The screenshot shows the 'Breach Report Results' page of the OCR portal. At the top, there's a green header bar with the HHS logo, the text 'U.S. Department of Health and Human Services Office for Civil Rights', and a sub-header 'Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information'. Below the header is a large image of a person's hands typing on a keyboard. The main content area has a title 'Breaches Affecting 500 or More Individuals' and a descriptive paragraph about the reporting requirements under the HITECH Act. A 'Show Advanced Options' link is visible. The main table has columns for Name of Covered Entity, State, Covered Entity Type, Individuals Affected, Breach Submission Date, Type of Breach, and Location of Breached Information. Five rows of data are listed:

Breach Report Results							
	Name of Covered Entity	State	Covered Entity Type	Individuals Affected	Breach Submission Date	Type of Breach	Location of Breached Information
1	Alliance Health Networks, LLC	UT	Healthcare Provider	42372	02/15/2016	Hacking/IT Incident	Network Server
2	Radiology Regional Center, PA	FL	Healthcare Provider	483083	02/12/2016	Loss	Paper/Films
3	DataStat, Inc.	MI	Business Associate	552	02/12/2016	Unauthorized Access/Disclosure	Paper/Films
4	Washington State Health Care Authority (HCA)	WA	Health Plan	91187	02/09/2016	Unauthorized Access/Disclosure	Email
5	SELM JOHNSON, LLP	NE	Business Associate	30972	02/08/2016	Theft	Laptop

De-Anonymization

TECH 4/25/2013 @ 8:47PM | 17,111 views

Harvard Professor Re-Identifies Anonymous Volunteers In DNA Study

[+ Comment Now](#) [+ Follow Comments](#)

A Harvard professor has re-identified the names of more than 40% of a sample of anonymous participants in a high-profile DNA study, highlighting the dangers that ever greater amounts of personal data available in the Internet era could unravel personal secrets.

From the onset, the Personal Genome Project,



Harvard Professor Latanya Sweeney

Melissa Gymrek et al. “*Identifying Personal Genomes by Surname Inference.*” Science Vol. 339, No. 6117, 2013

Aggregation

Re-identification of aggregated data

Statistics from allele frequencies can be used to identify genetic trial participants [1]

Membership Inference

Presence of an individual in a group can be determined by using allele frequencies and his DNA profile [2]

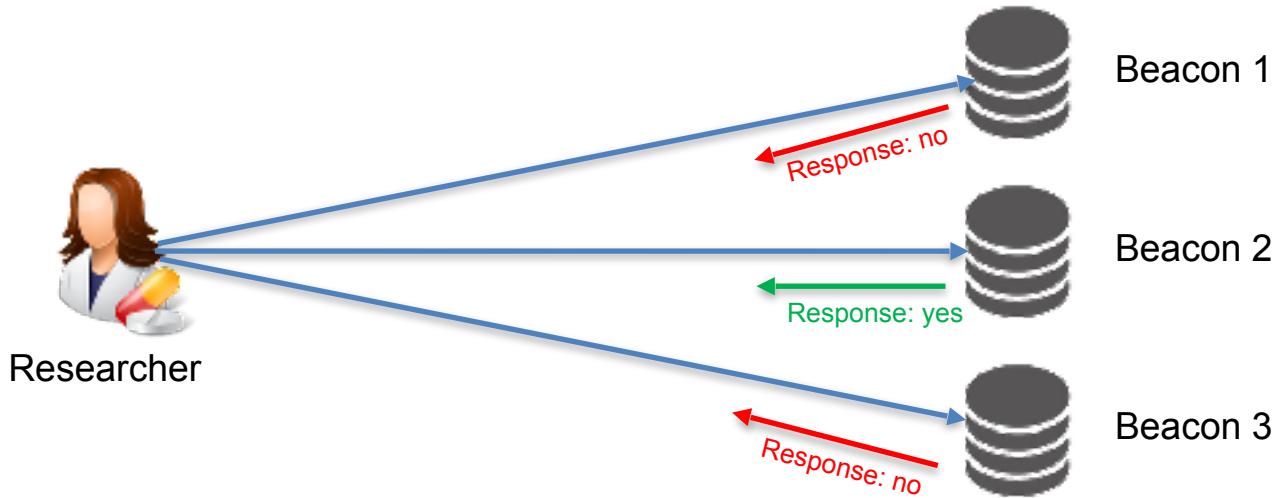
Against individuals contributing their microRNA expressions to scientific studies [3]

[1] R. Wang et al. "Learning Your Identity and Disease from Research Papers: Information Leaks in Genome Wide Association Study." CCS, 2009

[2] N. Homer et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genetics, 2008

[3] M. Backes et al. Membership privacy in MicroRNA-based studies. CCS, 2016.

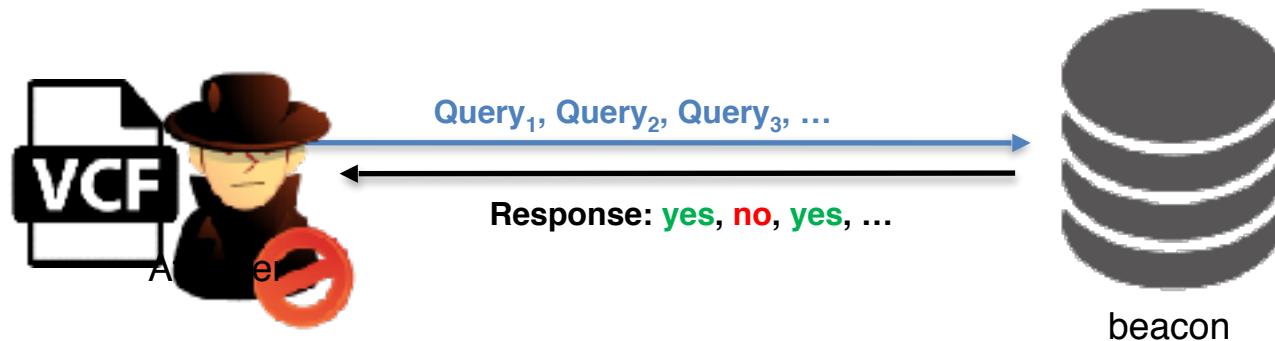
GA4GH Beacon Project



Main features:

Allows researchers to quickly query multiple database to find the sample they need; encourages cross-borders collaboration among researchers
Only minimal responses back in order to mitigate privacy concerns

Shringarpure-Bustamante's Attack



Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics. 2015 Nov 5;97(5):631-46.

Shringarpure-Bustamante's Attack



Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics. 2015 Nov 5;97(5):631-46.

Shringarpure-Bustamante's Attack



Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics. 2015 Nov 5;97(5):631-46.

The attack relies on the assumption that the adversary knows the set of variants (VCF file) of the target individual & the size of the beacon

Shringarpure-Bustamante's Attack



Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics. 2015 Nov 5;97(5):631-46.

The attack relies on the assumption that the adversary knows the set of variants (VCF file) of the target individual & the size of the beacon

The attack is based on a likelihood ratio test where the adversary repeatedly queries the beacon in order to re-identify the individual

Shringarpure-Bustamante's Attack



Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. The American Journal of Human Genetics. 2015 Nov 5;97(5):631-46.

The attack relies on the assumption that the adversary knows the set of variants (VCF file) of the target individual & the size of the beacon

The attack is based on a likelihood ratio test where the adversary repeatedly queries the beacon in order to re-identify the individual

Can be extremely dangerous if the beacon is associated with a sensitive phenotype (e.g., cancer)

Family Relationship Disclosure

Family Relationship Disclosure

Vox

THE GOODS EXPLAINERS POLITICS & POLICY WORLD CULTURE SCIENCE & HEALTH MORE

With genetic testing, I gave my parents the gift of divorce

By George Doe | Sep 9, 2014, 7:50am EDT

f t  SHARE



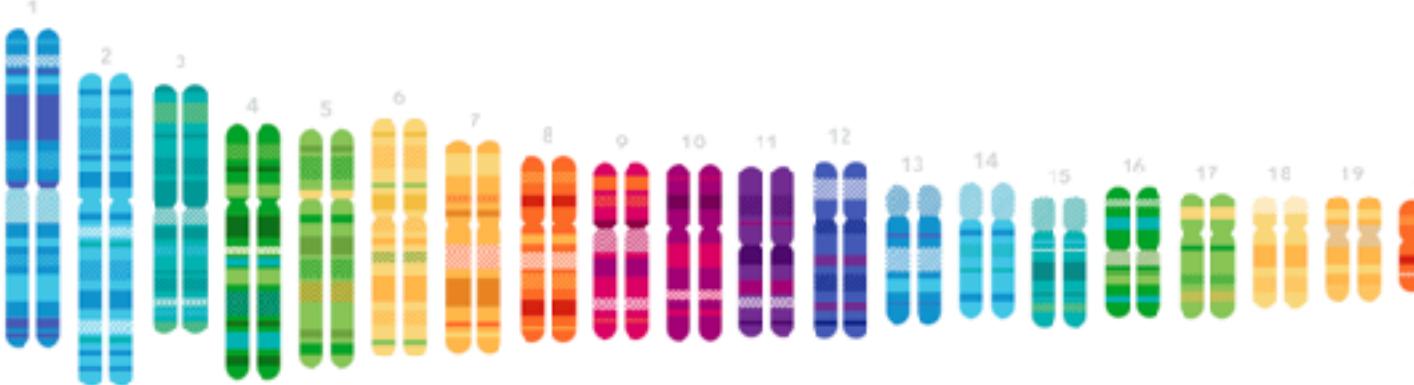
MON

Facebo
because

Family Relationship Disclosure

reddit r/23andme Search r/23andme LOG IN

23andMe
Ancestry 5 reports Genetic Health Risks* 54 reports
Traits 254 reports Carrier Status* 464 reports Wellness 54 reports



Posts

Join the discussion BECOME A REDDITOR

Posted by u/23andmethrowaway22 7 months ago

132 How 23 and Me helped me find my birth father.

(copied from a Facebook post I wrote about this, after taking out identifying data)

Background: I'm 30, female, just took my 23 and Me test in November.

Donor-conceived people are tracking down their biological fathers, even if they want to hide

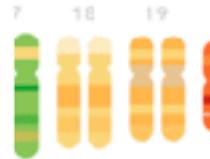


LOG IN

Thanks to DNA tests and the internet, 'anonymous' sperm donation is a thing of the past

Emily Chung, Melanie Glanz, Vik Adhopia · CBC News ·

Posted: Jan 25, 2018 5:00 AM ET | Last Updated: August 20



Posts



132



~~Donna Zuckerman wants you to take down this~~

Deportations on the back of consumer genetic tests worry scientists

BY BRIAN OWENS | 31 AUGUST 2018



Canada's border enforcement agency appears to be using genetic tests and DNA ancestry sites to determine country of origin for would-be deportees

Some issues specific to genomes

Some issues specific to genomes

Genome is treasure trove of sensitive information

Some issues specific to genomes

Genome is treasure trove of sensitive information

Genome data cannot be revoked

Some issues specific to genomes

Genome is treasure trove of sensitive information

Genome data cannot be revoked

Genome is the ultimate identifier

Some issues specific to genomes

Genome is treasure trove of sensitive information

Genome data cannot be revoked

Genome is the ultimate identifier

Access to one's genome ≈ **access** to **relatives'** genome

Are these “new” threats?

Are these “new” threats?

We all leave cells behind after all...

Hair/saliva can be collected and sequenced?

Are these “new” threats?

We all leave cells behind after all...

Hair/saliva can be collected and sequenced?

Are these “new” threats?

We all leave cells behind after all...

Hair/saliva can be collected and sequenced?

...scalability...

Looking Ahead...

How do we overcome the “greater good” vs “privacy” dichotomy?

How do we empower users with informed decisions about their health and genomic data?

How do reason about ethical issues around relatives?



Thank you!