Session 8: Identification and Privacy

GENE 220

Learning Goals

- **Analyze** the role that government, private and public databases play in the application of forensic genetics in the criminal justice system globally.
- **Examine** risks and benefits of identification and privacy from private and public entities.

Naiomi Hunter
(she/her)

Alanna Pyke
(she/her)

Rachel Ungar
(she/her)

Emily Greenwald
(she/her)

Outline

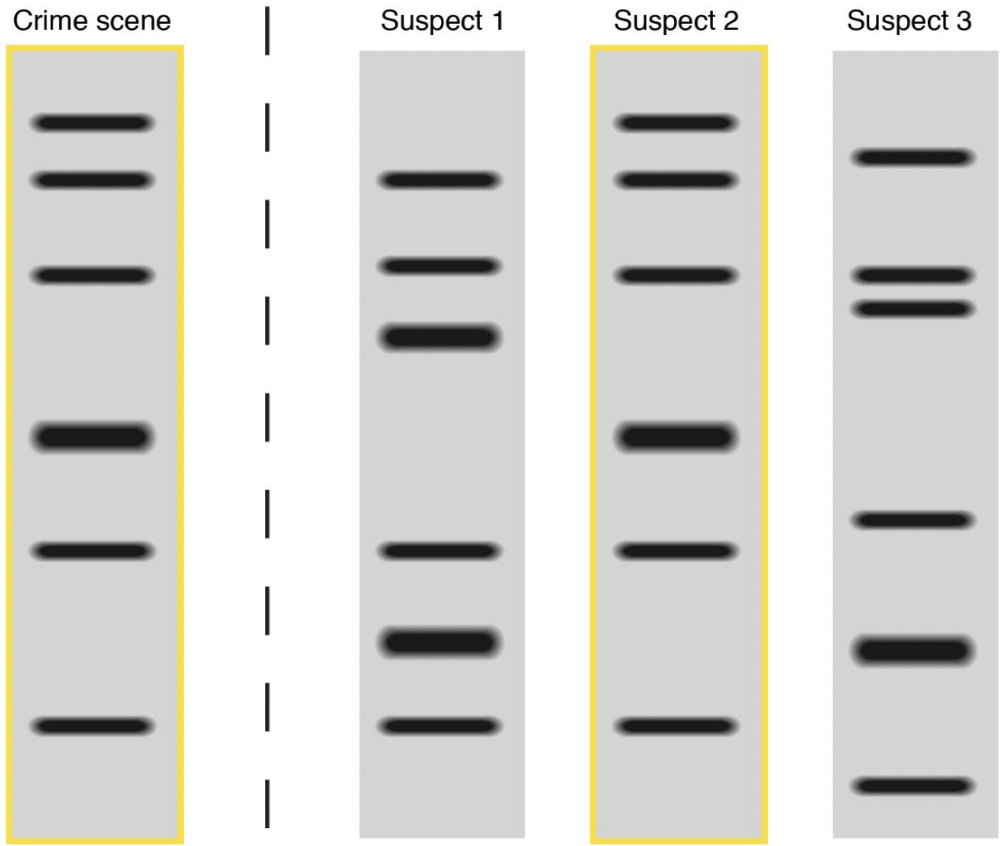
1. Forensic Genetic Databases
2. Research/DTC Genetic Databases
3. Data Policy & Stewardship
4. Databases and Identification
5. Current Protections and Approaches for Genetic Databases

What concerns did family, friends, uber drivers, and other people voice about genetic data?

Forensic Genetic Databases

DNA profiling:

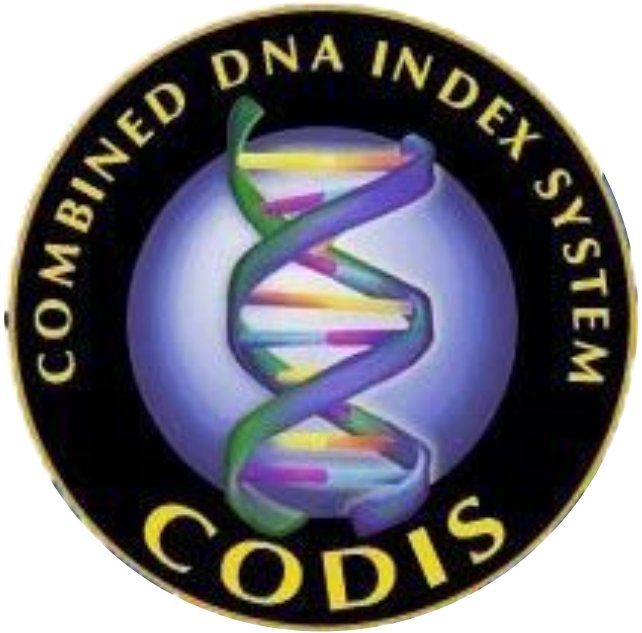
- DNA extraction
- PCR amplification
- Gel electrophoresis
- Profile matching



(Image: Green, E. NHGRI. *genome.gov*)

What is “CODIS”?

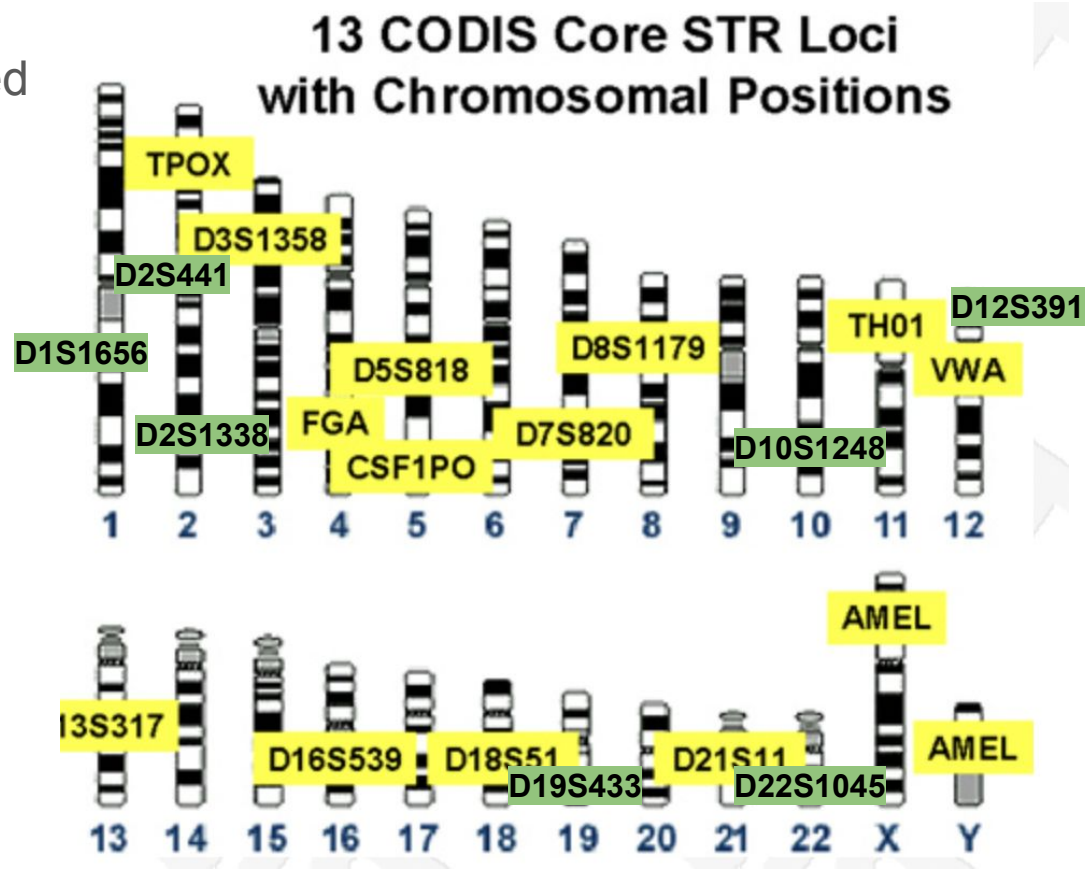
- CODIS=Combined DNA Index System
- FBI databases, the software used to maintain them and perform analyses, and the genetic loci sequenced for these purposes
- DNA Identification Act of 1994 approved the creation of an “index,” or database that can be searched for genetic matches
    - Used for criminal proceedings and identifying human remains



(Image: Biometric Analysis: Combined DNA Index System (CODIS). *FBI.gov*)

DNA fingerprinting: CODIS loci

- CODIS stands for Combined DNA Index System
- Short tandem repeats
- Noncoding
- Updated to 20 loci in 2017



(Image: National Institute of Standards and Technology. 2015. *strbase.nist.gov*)

Databases are searched for matches to crime scene DNA

- There are national, state, and local databases (Office of the Inspector General. 2006. *CODIS Audit.*)

- In 12 states as of 2018, partial-matches to crime scene DNA are used for familial searches (Grabell, Shomron. 2021. *Methods Mol Bio*.)

- After 2013 Supreme Court decision *Maryland v King*, **arrestees**, regardless of conviction status, can have DNA collected and added to CODIS databases in 30 states (*FBI.gov)*

California allows arrestees to be added to CODIS database and, being the most populous state, has the most information in the FBI database

	Total
Arrestee (Not convicted)	909,059
Offender (Convicted)	2,152,679
Investigations aided	102,341

(Data reformatted from : CODIS and NDIS statistics. 2021. *FBI.gov*.)

Racial biases of genetic databases

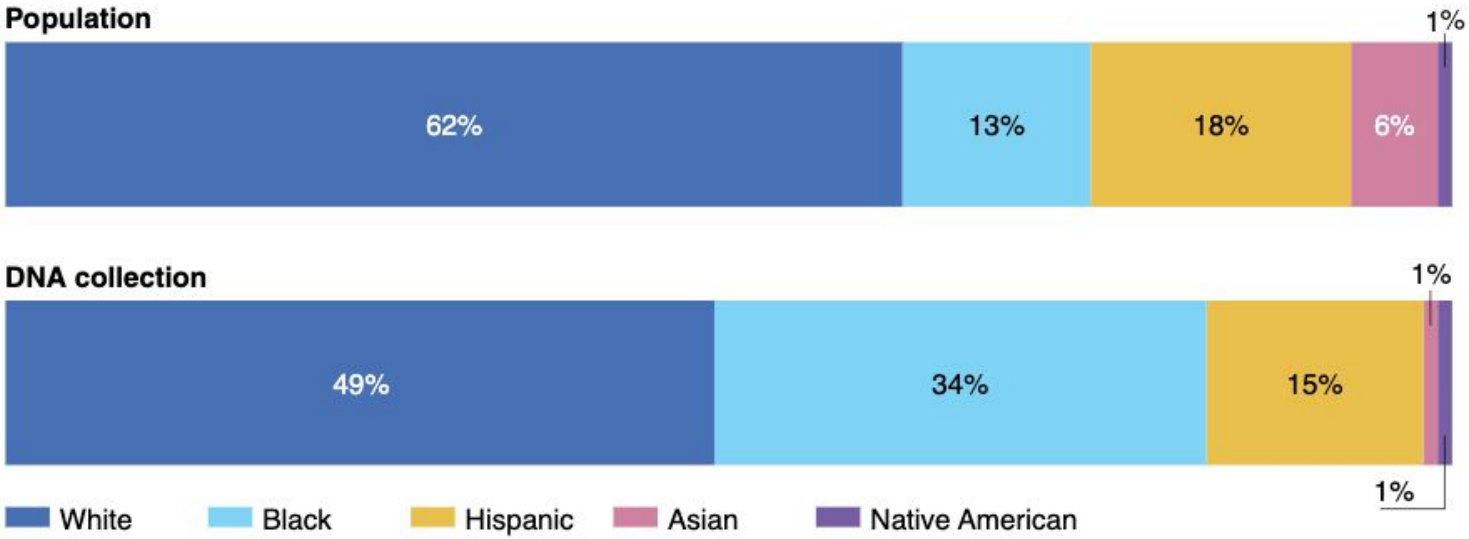


Fig. 1 | Comparison of US population with DNA collection for NDIS deposition, categorized as defined by the US Census. Data from ref. [3].

(Images: Court, Denise. 2021. *Nature Computational Science*.)

Important applications of forensic genetics beyond criminal proceedings

- Exonerations of wrongfully convicted people
- Identifying human remains
- Reuniting families that have been separated

Reading + small-group discussion (25 min)

5 min Read the article
5 min Small group discussion
10 min Large group discussion

Readings

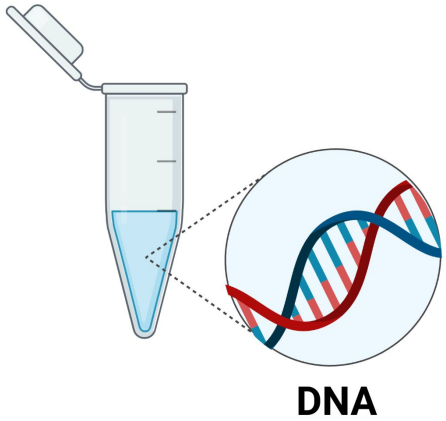
- Congress should support access to post-conviction DNA testing
- Police used a rape survivor's DNA to later arrest her
- Scientists say DNA can reunite separated migrant families
- China is collecting DNA from Tens of Millions of Men and Boys Using US Equipment

Questions

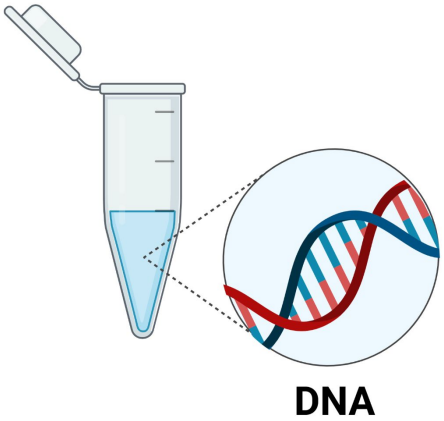
- How was genetics used in this case?
- What, if any, were the benefits of using genetics in this case?
- What, if any, were the harms of using genetics in this case?

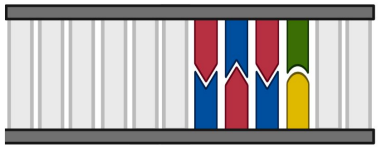
Research / Direct-To-Consumer (DTC) Genetic Databases

What types of data can be generated?



What types of data can be generated?

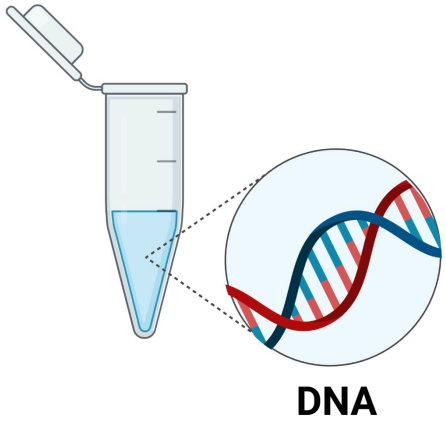


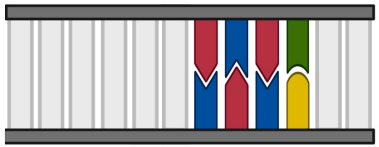


Targeted Sequencing

- Information about specific region
- Hypothesis driven & clinical testing

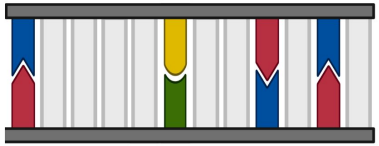
What types of data can be generated?





Targeted Sequencing

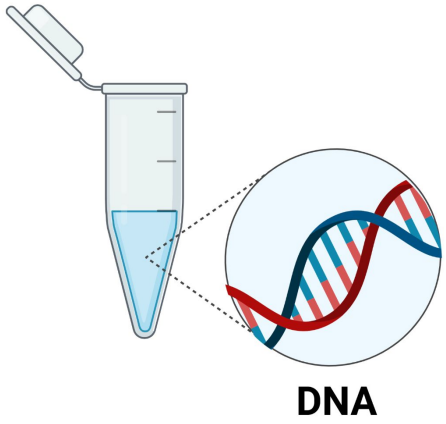
- Information about specific region
- Hypothesis driven & clinical testing

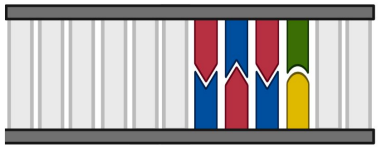


Genotyping Arrays

- genome-wide
- ~1% bases
- Population genetics and genetic epidemiology

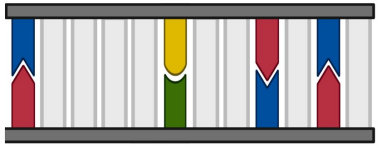
What types of data can be generated?





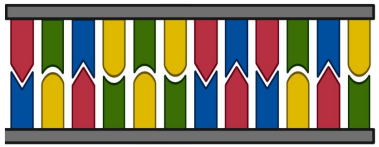
Targeted Sequencing

- Information about specific region
- Hypothesis driven & clinical testing



Genotyping Arrays

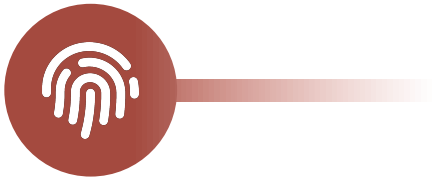
- genome-wide
- ~1% bases
- Population genetics and genetic epidemiology



Whole Ome Sequencing

- genome-wide
- as many bases as possible
- Rare and *de novo* variant detection

Types of Data Access



Private

- data is not shared outside the initial team

Types of Data Access

Private

- data is not shared outside the initial team

Restricted

- requires permission from a data manager
- often phenotype & genotype data access is separate

European Genome-phenome Archive (EGA)

NCBI database of Genotypes and Phenotypes (dbGaP)

Types of Data Access

Private

- data is not shared outside the initial team

Restricted

- requires permission from a data manager
- often phenotype & genotype data access is separate

European Genome-phenome Archive (EGA)

NCBI database of Genotypes and Phenotypes (dbGaP)

Unrestricted

- only hoop is finding the correct download link
- de-identified with no phenotype data

1000 Genomes

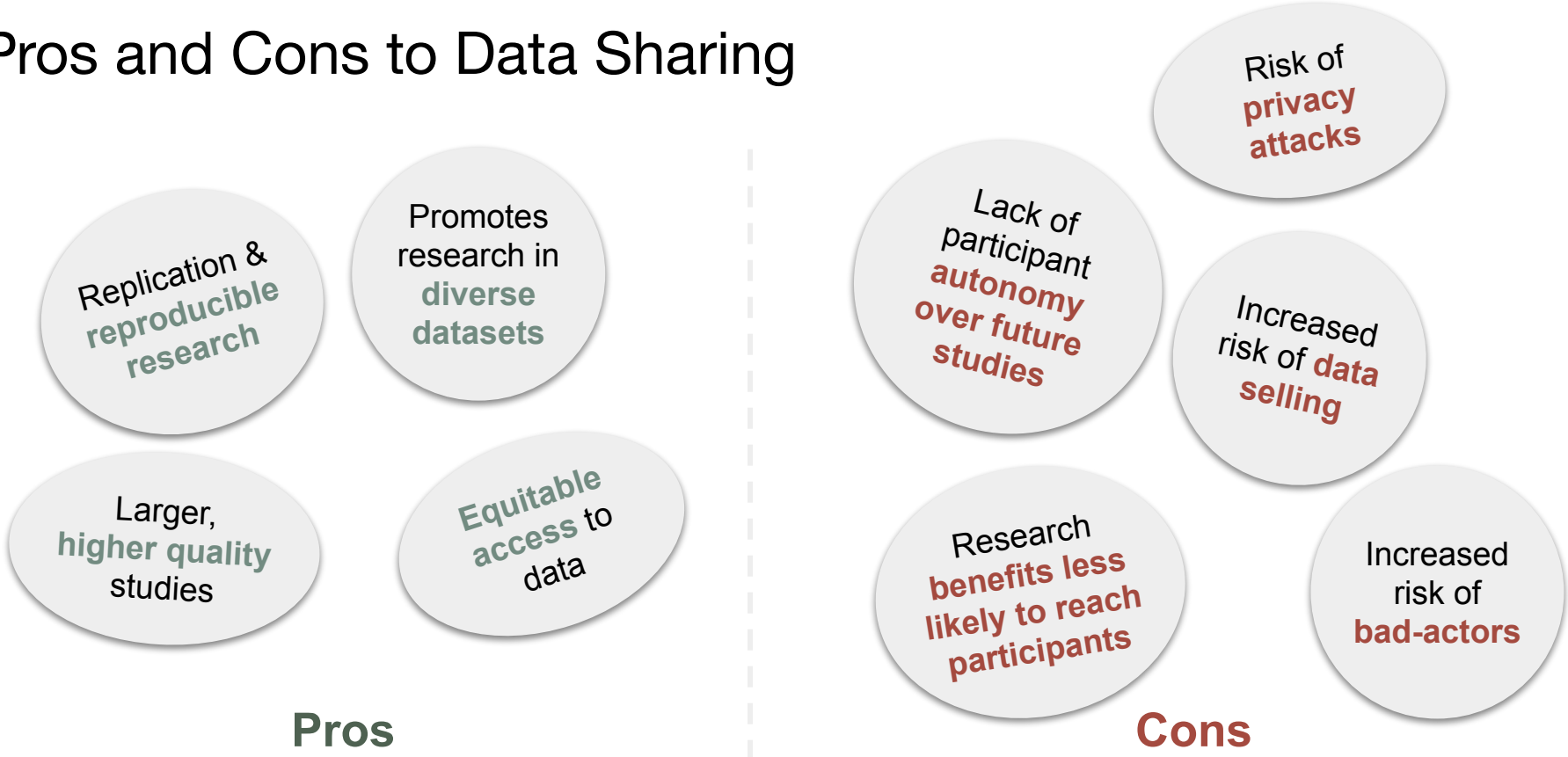
The Genome Aggregation Database (gnomAD)

Pros and Cons to Data Sharing

Pros

Cons

Pros and Cons to Data Sharing



Data Policy & Stewardship

Data policy

- Collection of data necessitates **data policy**, which outlines:
  - Who **owns** the data
  - Where the data will be **stored**
  - Who the data can be **shared with**
  - What can be **done** with the data (e.g., what **questions** can be asked)
  - Etc.

Data policy and **stewardship**

- **Data stewardship** - actively managing data and ensuring adherence to data policy
- Often, data policy needs to be outlined in a **protocol** that needs to be approved by an **Institutional Review Board**
- Consumers / research participants should be told about data policy **during the consent process**

Challenges of assessing stakeholdership in study design (Review)

1. Identify **who the stakeholders are**
2. **Solicit input** from stakeholders
3. **Balance conflicting perspectives**
4. Appropriate **communication** with each group

Entities that deal with human genetic data: four examples

Consumer-facing companies

23&Me: “DNA genetic testing for health, ancestry, and more”

Ancestry: “Family tree, genealogy and family history records”

Research-focused entities

H3Africa: “Human heredity and health in Africa… aims to… study of **genomics and environmental determinants of common diseases** with the goal of improving the health of **African populations**”

UK Biobank: “large-scale biomedical database… **genetic and health information** from half a million **UK** participants… [enables] vital research into the most **common and life-threatening diseases**”

Each entity has participant-facing **policy documents** that explain how data will be managed

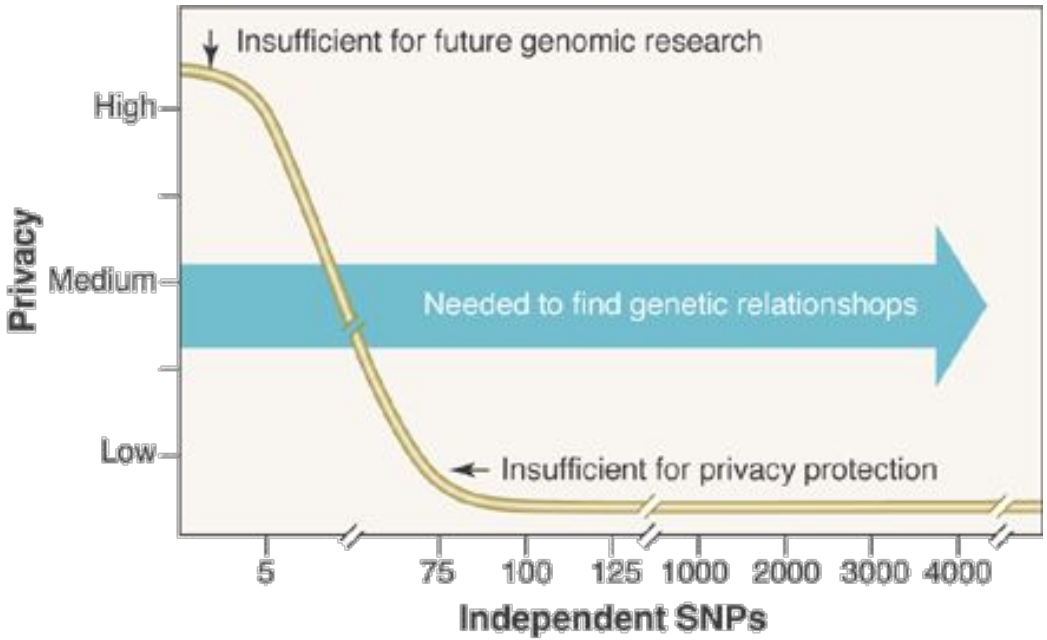
Activity: Partner with someone who read different policies than you and discuss the guiding questions (10 minutes)

Consider both the **research focus** of the database as well as its **privacy and consent policy**.

1. *Who are the stakeholders?*
2. *What did this data policy convey as the benefits of individual participation in this study?*
3. *What did this data policy convey as the harms, if at all?*
4. *What do **you** think are the benefits and harms to individuals and society?*
5. *How well do you think this information was communicated to the participants?*

Databases and Identification

Theoretically, an individual can be uniquely identified with just 30-80 statistically independent SNPs



Genomic Research and Human Subject Privacy
ZHEN LIN, ART B. OWEN, AND RUSS B. ALTMAN

[GED match]

Comprehensive solutions for genetic genealogy and family tree search

GEDmatch provides more applications for comparing your DNA test results with the most people worldwide.

VIEW OUR APPLICATIONS

1.4M Members Globally

45+ DNA Tools

12 Years of Experience

350M People in GEDCOMs

ACCEPTS RESULTS FROM

ancestry · Living DNA · tellmeGen · FamilyTreeDNA · MyHeritage · 23andMe

Applications you can't get anywhere else.

We have tools for

- ESTIMATING ETHNICITY AND HERITAGE
- TRACING SHARED ANCESTORS
- CREATING SUPERKITS AND MORE!

Reidentification in forensic cases: the Golden State Killer

- Serial rapist and killer between 1974-1986
- Upload DNA from a rape kit to GEDmatch
- found the suspect's third cousin
- Narrowed down to six possible suspects using age, crime locations, and genealogical records
- Eye-color prediction tool said this person had blue eyes, only one man did
- Arrest April 2018
- More than 30% of individuals in a forensic database have a sibling, parent, or child in a consumer database

https://slate.com/culture/2023/02/dna-testing-gsk-barbara-rae-venter-book-review.html

By May 2020, most (81%) of GEDmatch's 1.4 million users have not opted in

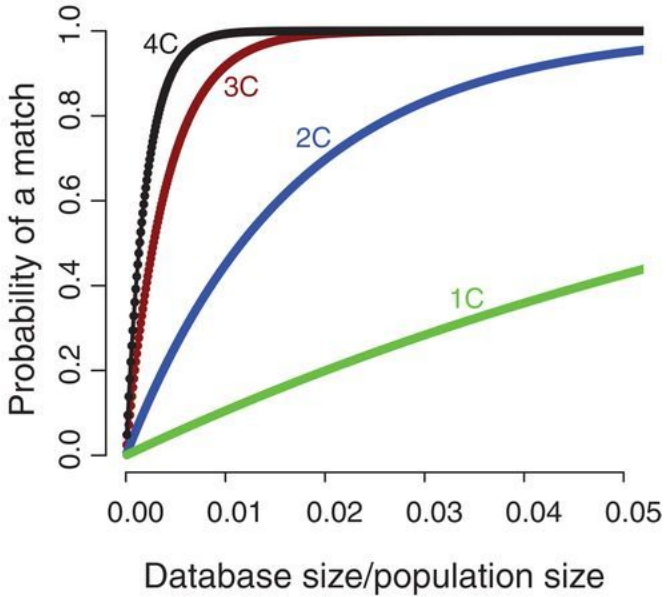
Public Opt-in

We will compare your DNA kit to all other kits in the GEDmatch database to find your matching genetic relatives. Kits in the database include those submitted by users undertaking personal genetic genealogy research, adoptee searches, users (including law enforcement) attempting to identify unidentified human remains, and law enforcement attempting to identify perpetrators of violent crimes. Your kit WILL be compared with kits submitted by law enforcement to identify perpetrators of violent crimes. The operators of GEDmatch encourage everybody to select this option.

Wan, Zhiyu, et al. "Sociotechnical safeguards for genomic data privacy." Nature Reviews Genetics (2022)

Probability of a match of a third-cousin or closer

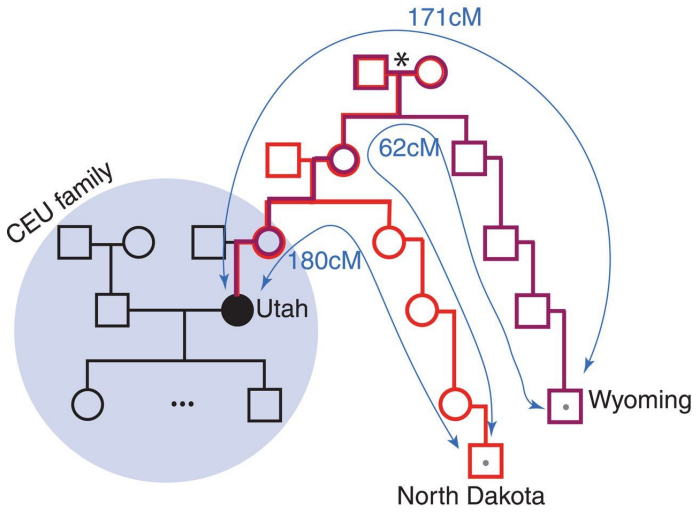
- In a database of 1.2 million individuals, 60% of searches of European individuals will result in finding a third cousin or closer match
- Northern European ancestry 30% more likely to have at least a third cousin match than sub-Saharan African ancestry
- Database needs to cover only 2% of the target population to provide a third-cousin match to nearly any person
- For >99% of europeans to be have at least one third cousin, database size should be 3 million



Erlich, Yaniv, et al. "Identity inference of genomic data using long-range familial searches." Science (2018)

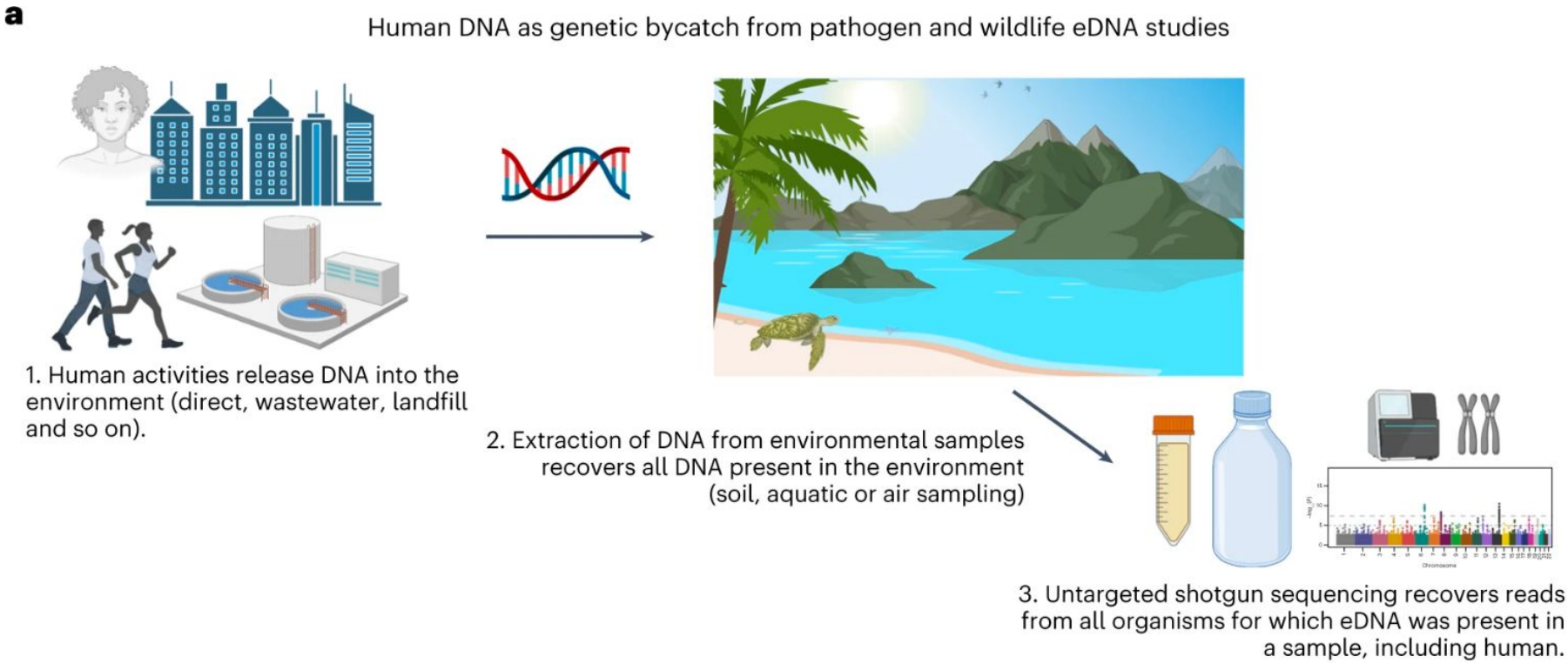
Re-identification of a person from 1000 Genomes database within a day

1. Uploaded publicly available data to GEDmatch
2. GEDmatch returned two relatives
3. Identified shared ancestral couple of all three individuals in public genealogical records
4. Searched for descendants matching the publicly available demographic data of the 1000Genomes sample (expected year of birth and pedigree structure)



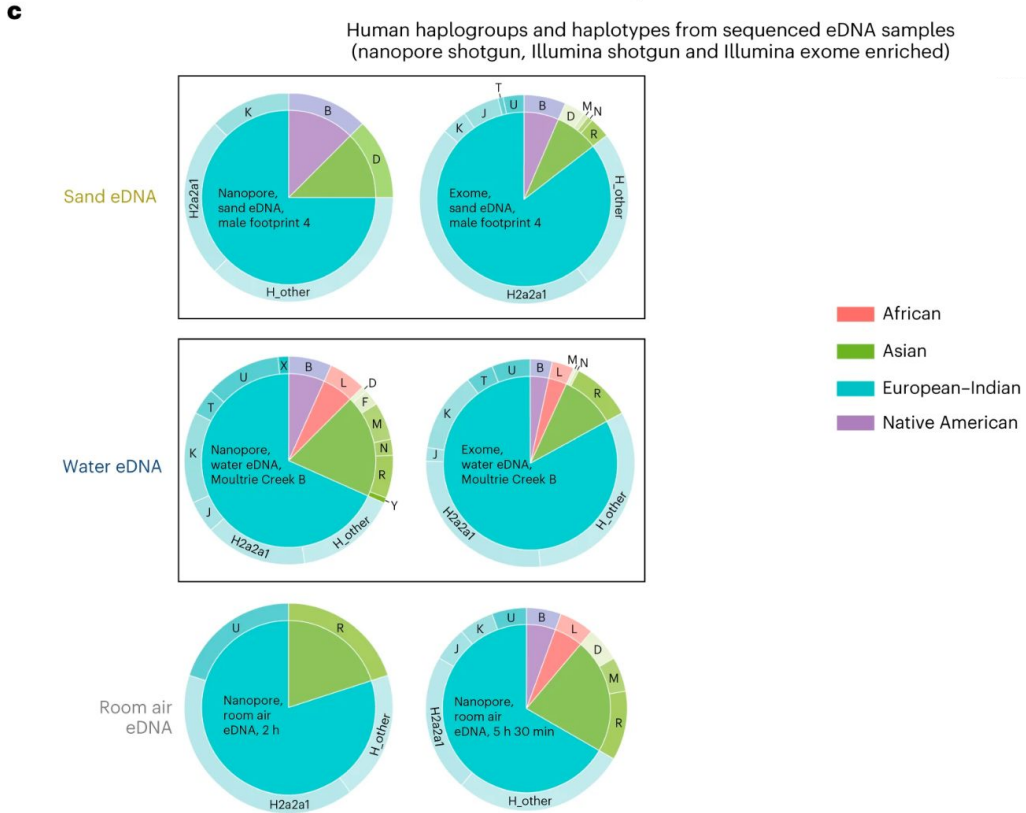
Erlich, Yaniv, et al. "Identity inference of genomic data using long-range familial searches." Science (2018)

Environmental DNA captures human DNA



Whitmore, Liam, et al. Nature Ecology & Evolution (2023)

Ancestry testing from environmental DNA



Whitmore, Liam, et al. Nature Ecology & Evolution (2023)

Large group discussion (8 minutes)

- What concerns does this raise?
- Can DNA information ever be truly private?
- Has this impacted how you might treat your own personal genetic information?

Current Protections and Approaches for Genetic Databases

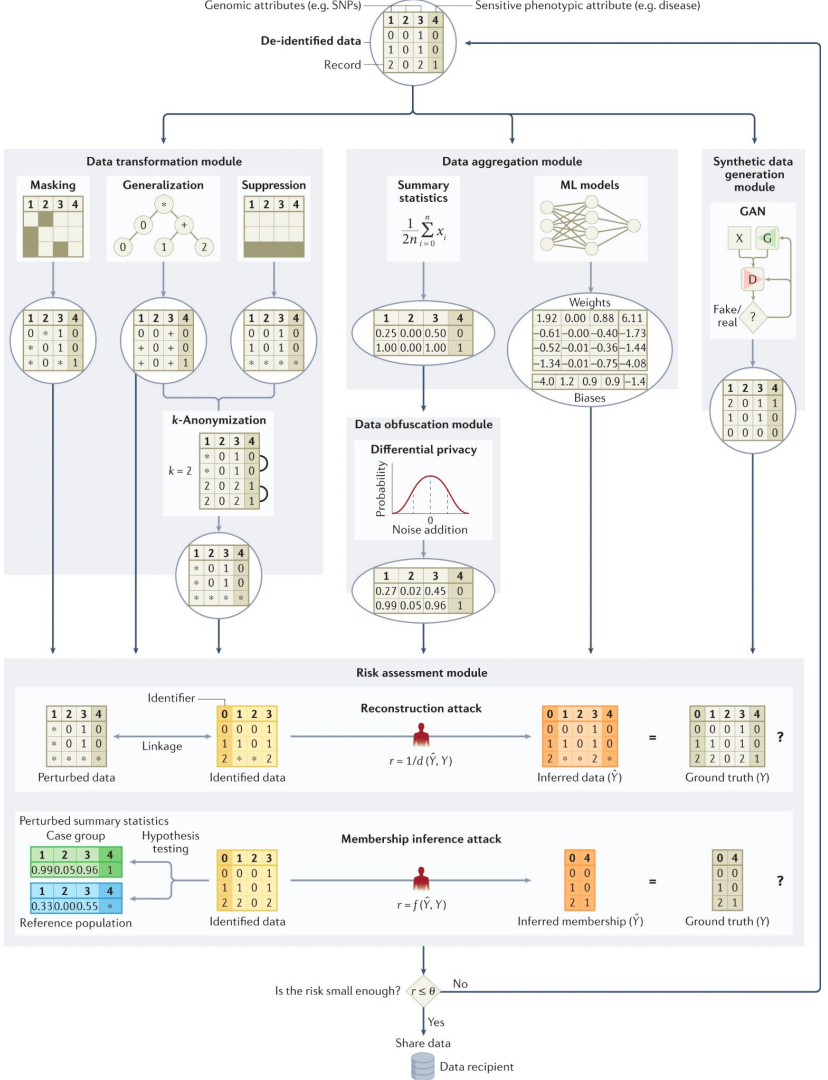
Legal Protections

Genetic Information Nondiscrimination Act of 2008 (GINA)

- prohibits **health insurance discrimination** based on genetic information (exceptions: life insurance, long-term care insurance, disability insurance)
- prohibits **employers from requiring/requesting genetic information or testing** of employees
  (exceptions: U.S. military, employers with fewer than 15 employees)

https://www.genome.gov/about-genomics/policy-issues/Genetic-Discrimination

Scientific Protection: Computer security approaches for privacy protection in genomic data sharing.



Wan, Zhiyu, et al. Nature Reviews Genetics (2022)

Community Protections: Example of community owned: Native BioData Consortium

ABOUT US

We are composed of the nation's leading Indigenous geneticists.

Our Board and Community Advisory Group includes tribal experts in precision health, technology, law, policy, business, ethics, and cultural matters.

The NBDC was established in 2018.

We already have an international reputation in tribal research and policy.

We are leading example for improving Indigenous health equity.

KEEP SAMPLES AND DATA LOCAL

BUILD TRIBAL CAPACITY

FOSTER TRUST IN RESEARCH

RESEARCH THAT BENEFITS TRIBES

What protections should genetic databases have?

4 / 5 groups
10 minutes small white board brainstorm
12 minutes large share out

Potential societal harms	**Relevant institutions or stakeholders**	**Social responsibilities of stakeholders**	**How scientists can mitigate harm**	**What I can do**
Research links racial differences in IQ to genetics	UK Biobank, other large-scale research databases	The UK government has a responsibility to only give data access to those who will ask responsible research questions	Create a list of community standards topics that the data cannot be used for	Criticize publicly any bad-faith papers trying to create this link/have a conversation with my lab to create a list of topics we will not look at

Session 9: Genetics in the Archives Preview

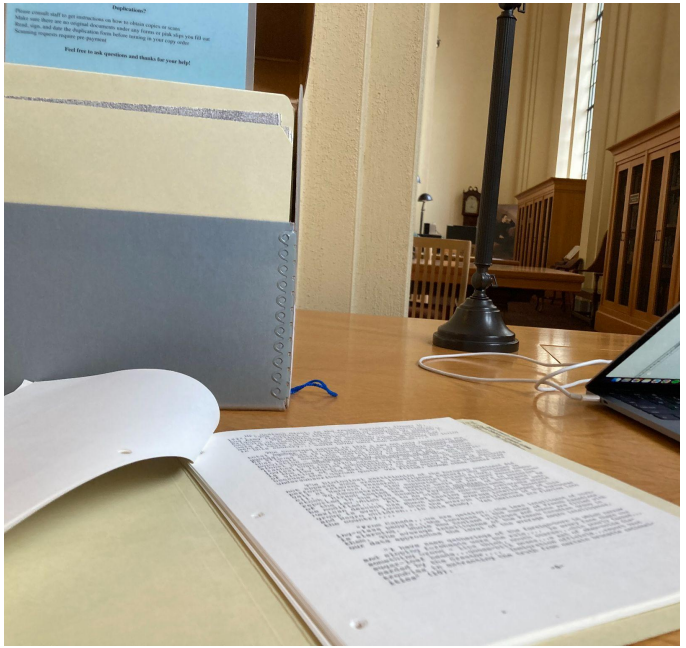
Session 9: Genetics in the Archives

Learning objectives

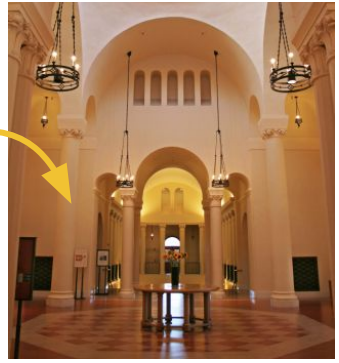
- Understand how individuals acted to counter the misuse of genetics research to further racist and sexist ideologies
- Connect historical examples of ethical problems to the modern-day case studies presented in this course
- Analyze how scientists' ideologies impacted their interpretation of science

Session 9: Genetics in the Archives – Format

- Five tables on different topics
- You will have three long discussion topics (20 minutes each) and two short exposure topics (7 minutes each)



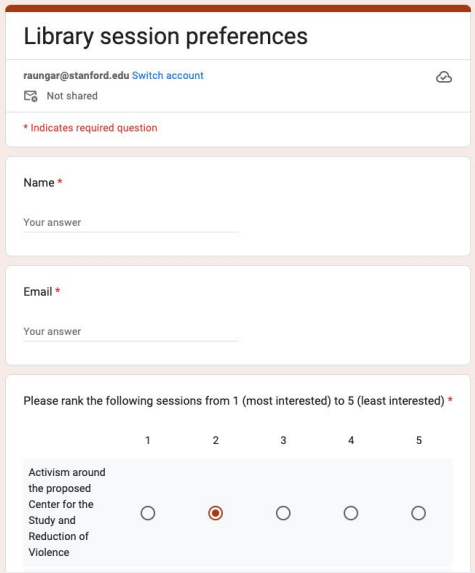
Session 9: Genetics in the Archives – Location



- Will be in Green library!!!!
- **YOU MUST BRING YOUR STANFORD STUDENT ID** (not med school badge)
- Enter through side closest to Main Quad and go up the stairs
- We will meet in Green Library, Bing Wing at the big round oak table at the top of the grand staircase (2nd floor).
- We will be checking our email, please email Rachel/Roshni if you get lost!!

Session 9: Genetics in the Archives – Pre-Work

1. By **Thursday at 5PM** please rank your session preferences



Library session preferences

raungar@stanford.edu Switch account

Not shared

* Indicates required question

Name *

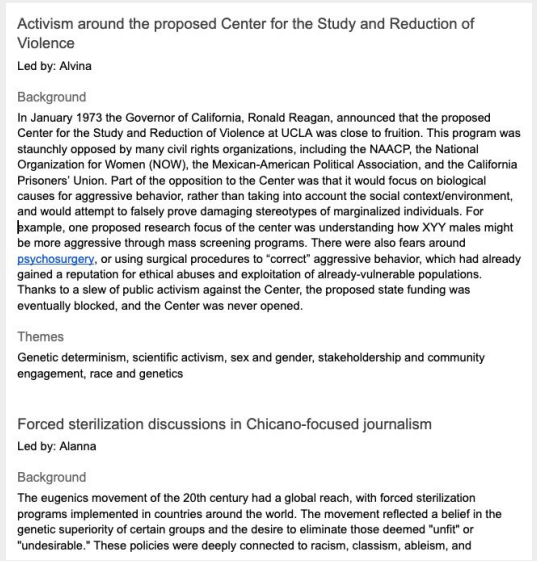
Your answer

Email *

Your answer

Please rank the following sessions from 1 (most interested) to 5 (least interested) *

	1	2	3	4	5
Activism around the proposed Center for the Study and Reduction of Violence	○	◉	○	○	○



Activism around the proposed Center for the Study and Reduction of Violence

Led by: Alvina

Background

In January 1973 the Governor of California, Ronald Reagan, announced that the proposed Center for the Study and Reduction of Violence at UCLA was close to fruition. This program was staunchly opposed by many civil rights organizations, including the NAACP, the National Organization for Women (NOW), the Mexican-American Political Association, and the California Prisoners' Union. Part of the opposition to the Center was that it would focus on biological causes for aggressive behavior, rather than taking into account the social context/environment, and would attempt to falsely prove damaging stereotypes of marginalized individuals. For example, one proposed research focus of the center was understanding how XYY males might be more aggressive through mass screening programs. There were also fears around psychosurgery, or using surgical procedures to "correct" aggressive behavior, which had already gained a reputation for ethical abuses and exploitation of already-vulnerable populations. Thanks to a slew of public activism against the Center, the proposed state funding was eventually blocked, and the Center was never opened.

Themes

Genetic determinism, scientific activism, sex and gender, stakeholdership and community engagement, race and genetics

Forced sterilization discussions in Chicano-focused journalism

Led by: Alanna

Background

The eugenics movement of the 20th century had a global reach, with forced sterilization programs implemented in countries around the world. The movement reflected a belief in the genetic superiority of certain groups and the desire to eliminate those deemed "unfit" or "undesirable." These policies were deeply connected to racism, classism, ableism, and

Sections assignments released by Friday at 5PM

2. Complete short readings for three assigned section prior to class

3. Please show up to Green Library 5 minutes early if possible, and allow yourself time to get lost, so we can get started ASAP! And bring your ID!!