

# Challenges and Opportunities for Machine Learning in Genomics

Bradley Malin, Ph.D.

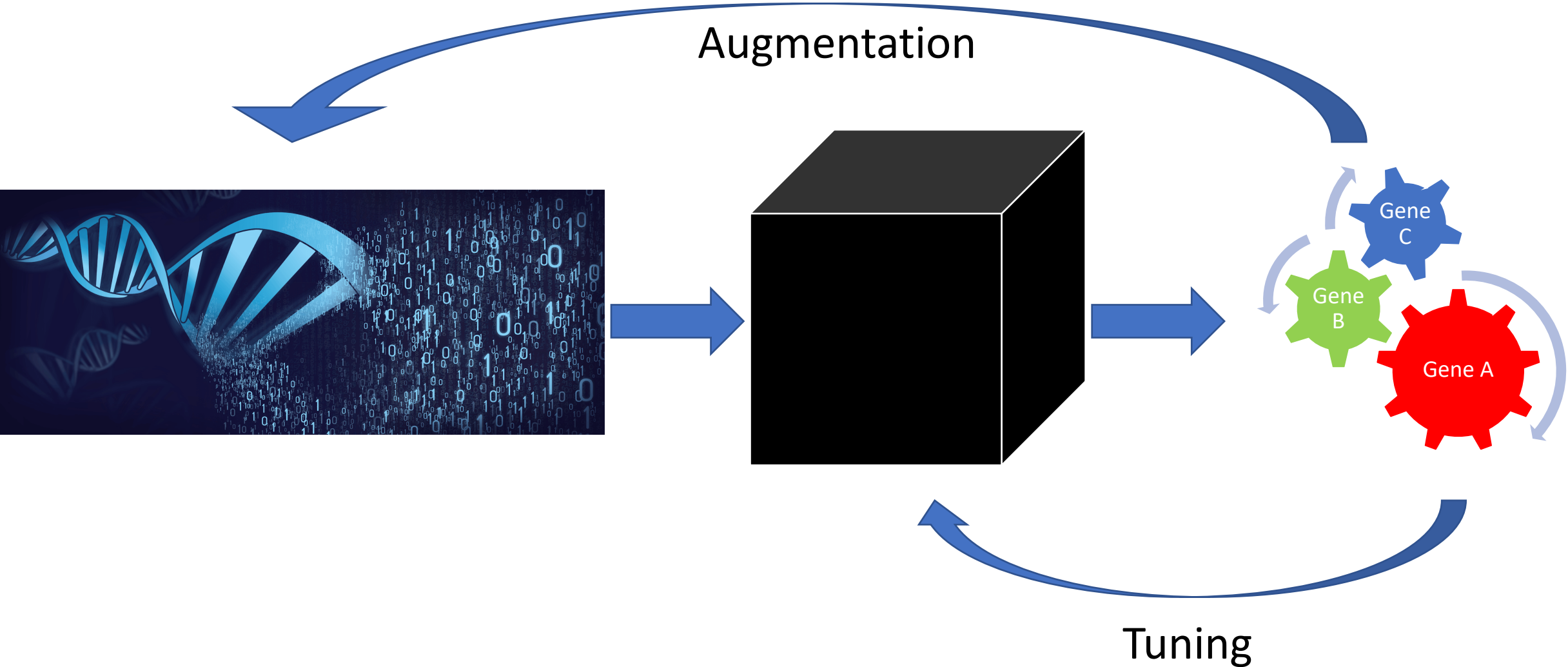
Accenture Professor of Biomedical Informatics, Biostatistics, & Computer Science

Co-Director, Health Data Science Center

Co-Director, Center for Genetic Privacy and Identity in Community Settings

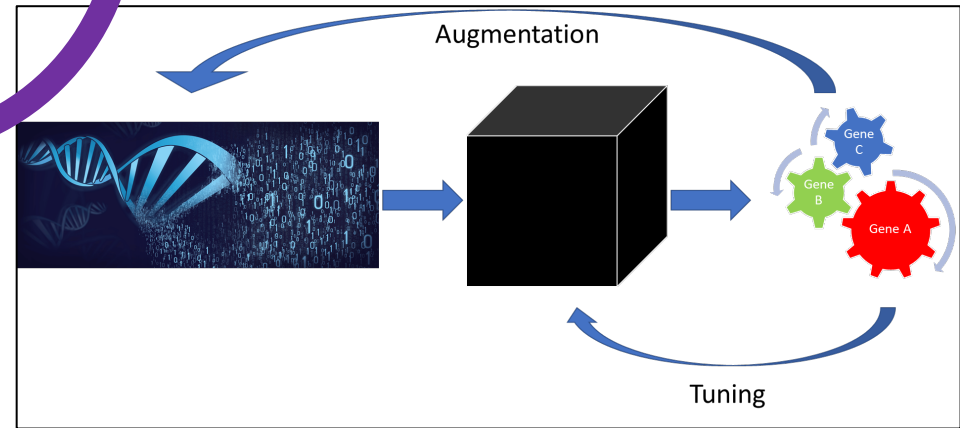
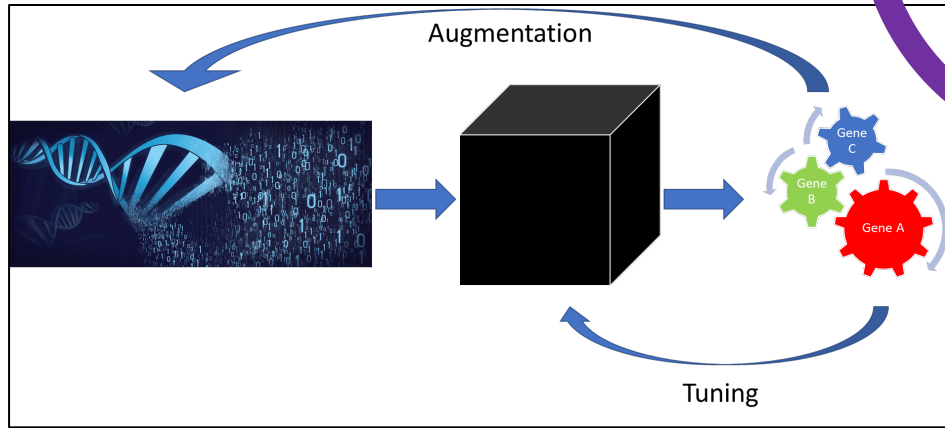
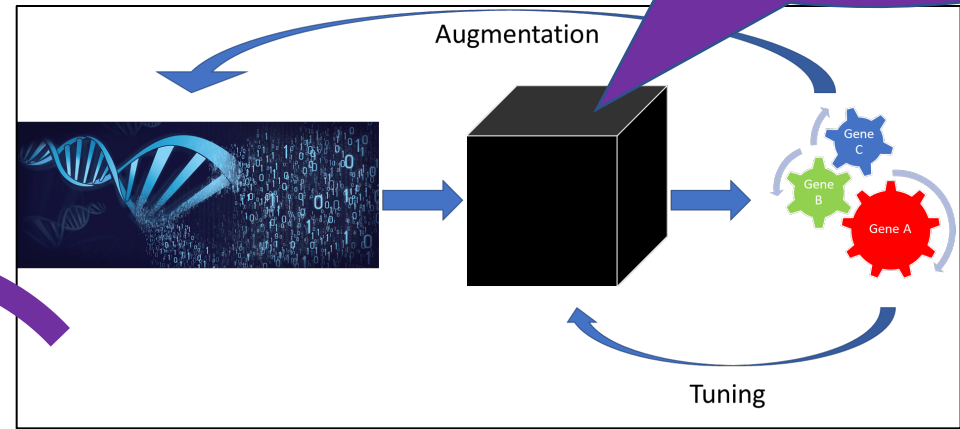
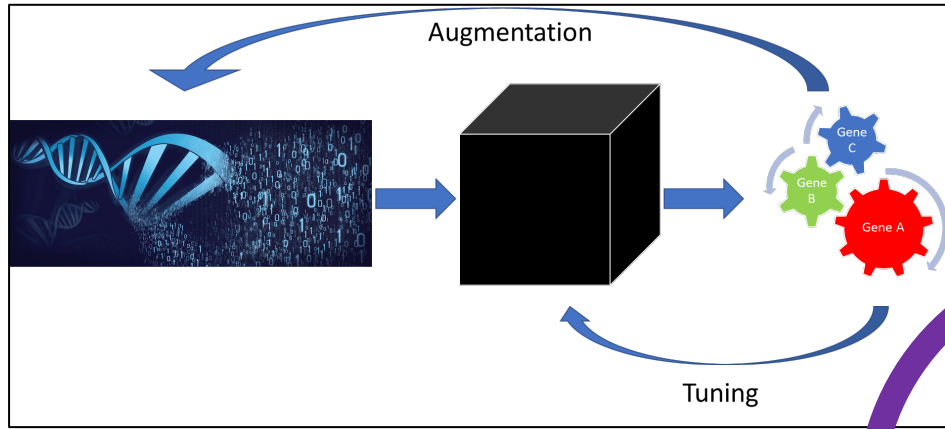
Vanderbilt University Medical Center

# Machine Learning is a Process



Bigger  
&  
Better

# Safety in Numbers



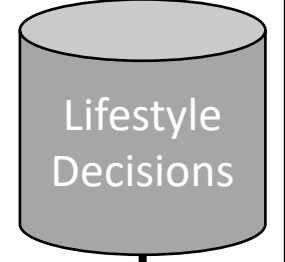
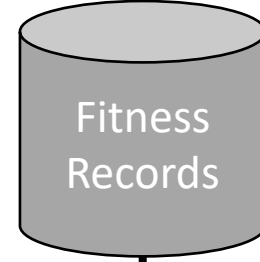
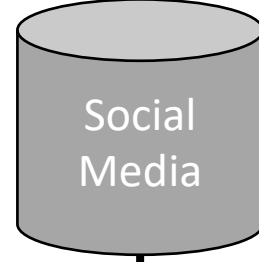
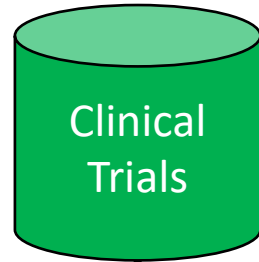
Do you see what I see?



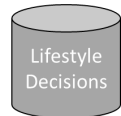
# Broader

## Data

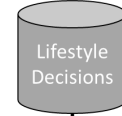
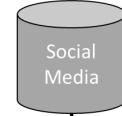
Parent



Child



Child





By  
**Vicky Stein**

21  
comments

# Genetic research has a white bias, and it may be hurting everyone's health

[Science](#) Updated on Mar 22, 2019 10:46 AM EDT – Published on Mar 21, 2019 4:28 PM EDT

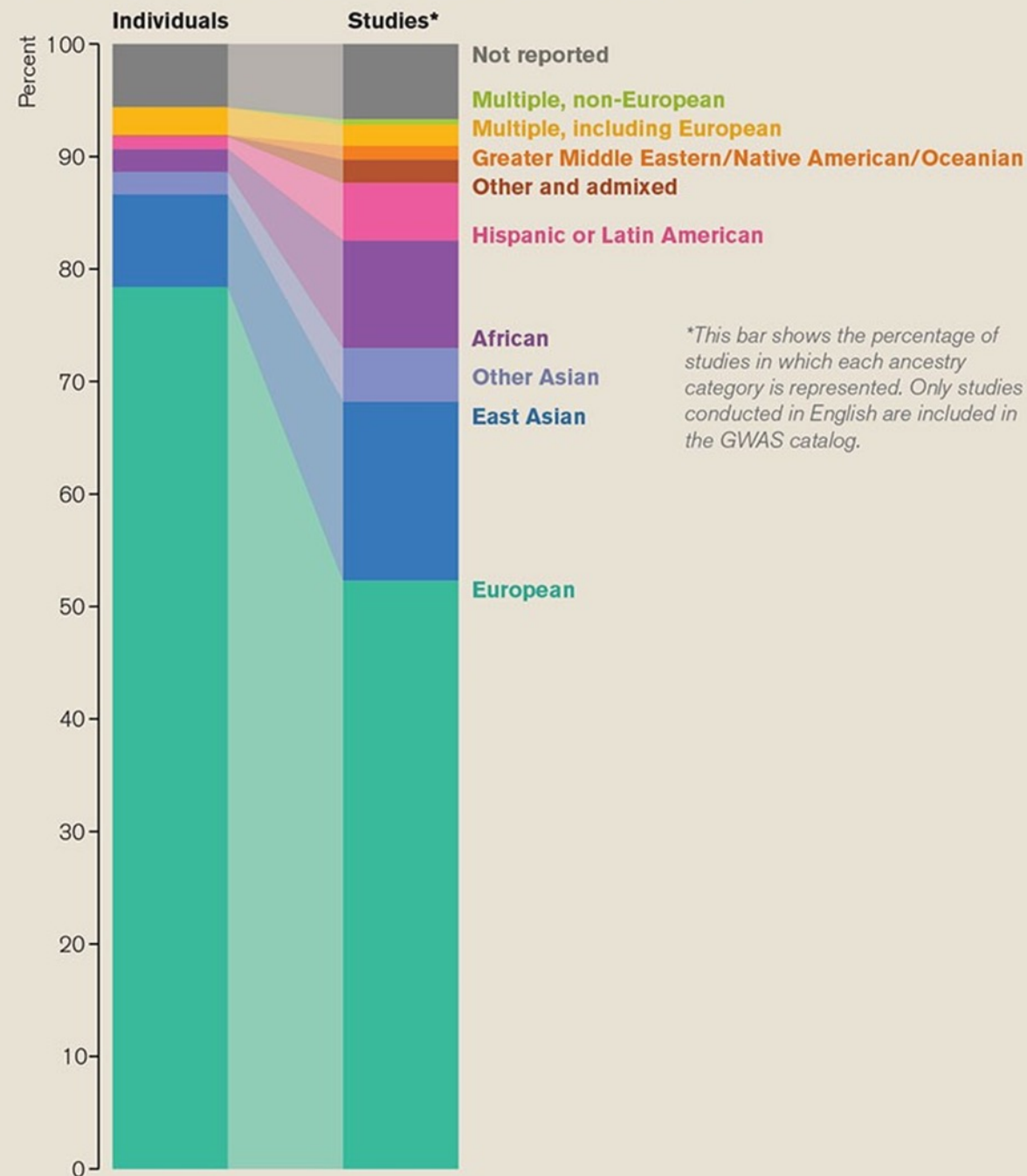
INNOVATIONS IN · 04 DECEMBER 2019

# All of Us

DNA-based medicine needs more diversity to avoid harmful bias. One big research project is fixing that.

Stephanie Devaney

## Racial Backgrounds in Published Gene-Association Studies

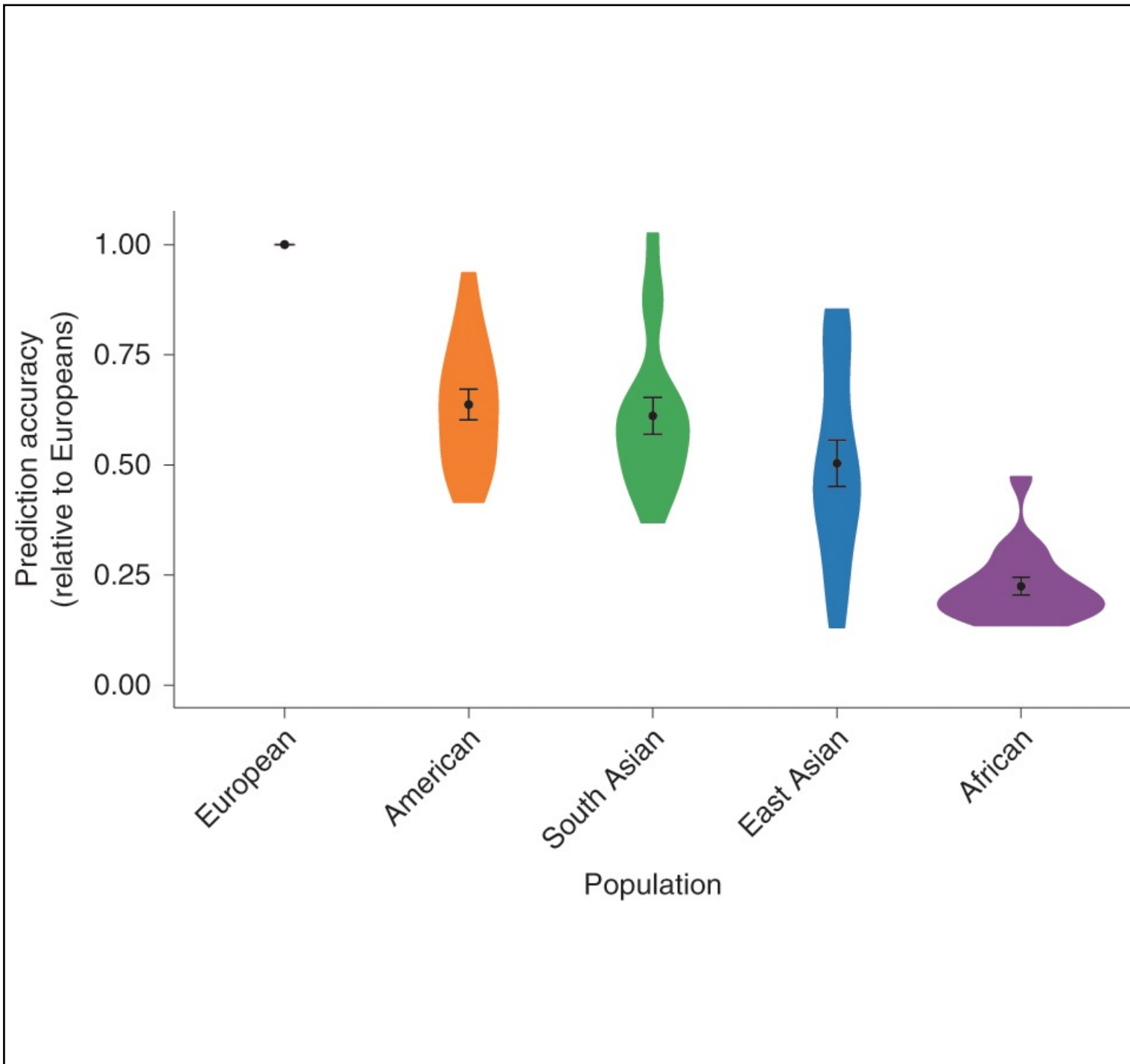


Review Article | [Open Access](#) | Published: 25 February 2020

## Evaluating the promise of inclusion of African ancestry populations in genomics

Amy R. Bentley, Shawneequa L. Callier & Charles N. Rotimi [✉](#)

*npj Genomic Medicine* 5, Article number: 5 (2020) | [Cite this article](#)

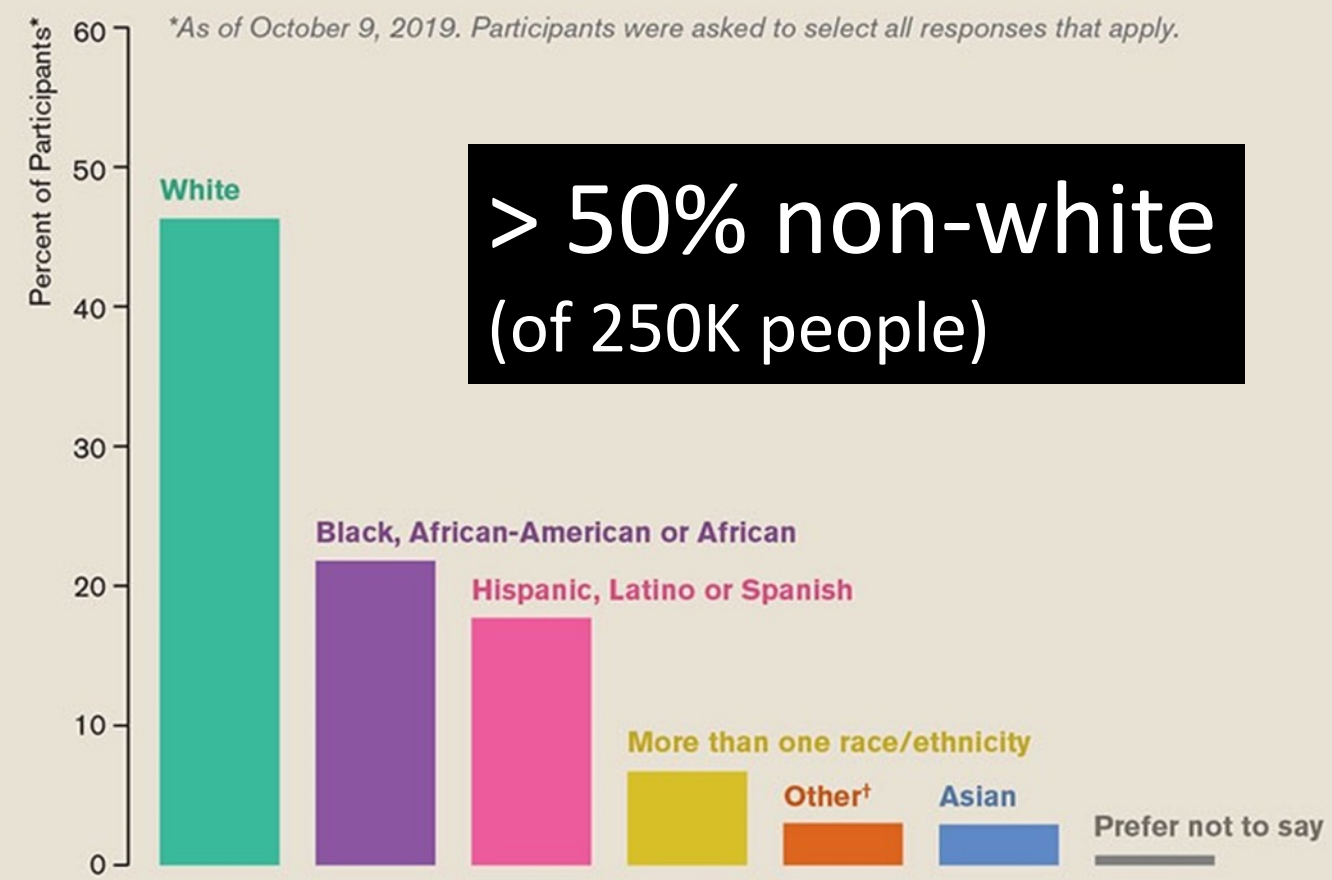


INNOVATIONS IN · 04 DECEMBER 2019

# All of Us

DNA-based medicine needs more diversity to avoid harmful bias. One big research project is fixing that.

Stephanie Devaney



> 50% non-white  
(of 250K people)

†This group includes the following selections: American Indian or Alaska Native; Middle Eastern or North African; Native Hawaiian or other Pacific Islander; None of these describe me.





# Plan for Increasing the Number of Underrepresented Minorities Trained in Machine Learning

## Plan for Increasing the Number of Underrepresented Minorities<sup>1</sup> Trained in Genomics and ELSI<sup>2</sup> Research

### Section I - Action Plan

Goal

Principles

Plan

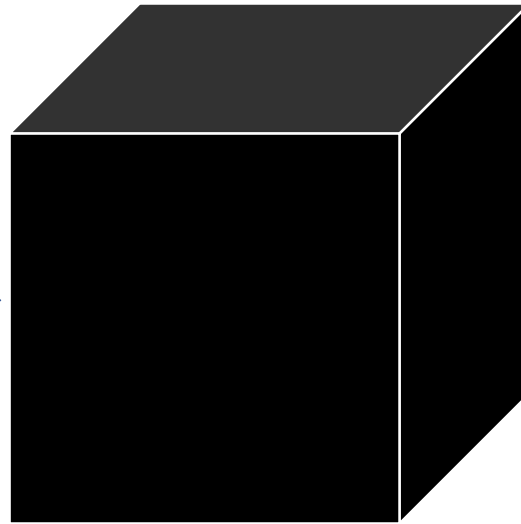
#### A. Office of Policy and Public Affairs (OPPA)

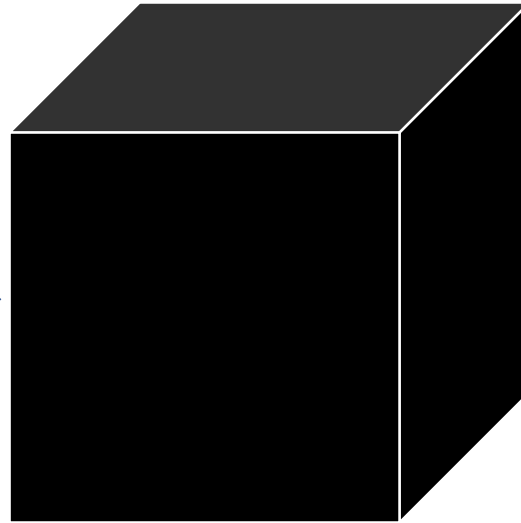
1. Minority Conferences
2. K-12 Education "Exploring our Molecular Selves" Educational Kit
3. Genome Minority Outreach Video for Inclusive Education (Genome Movie)

# Cost Effective Learning

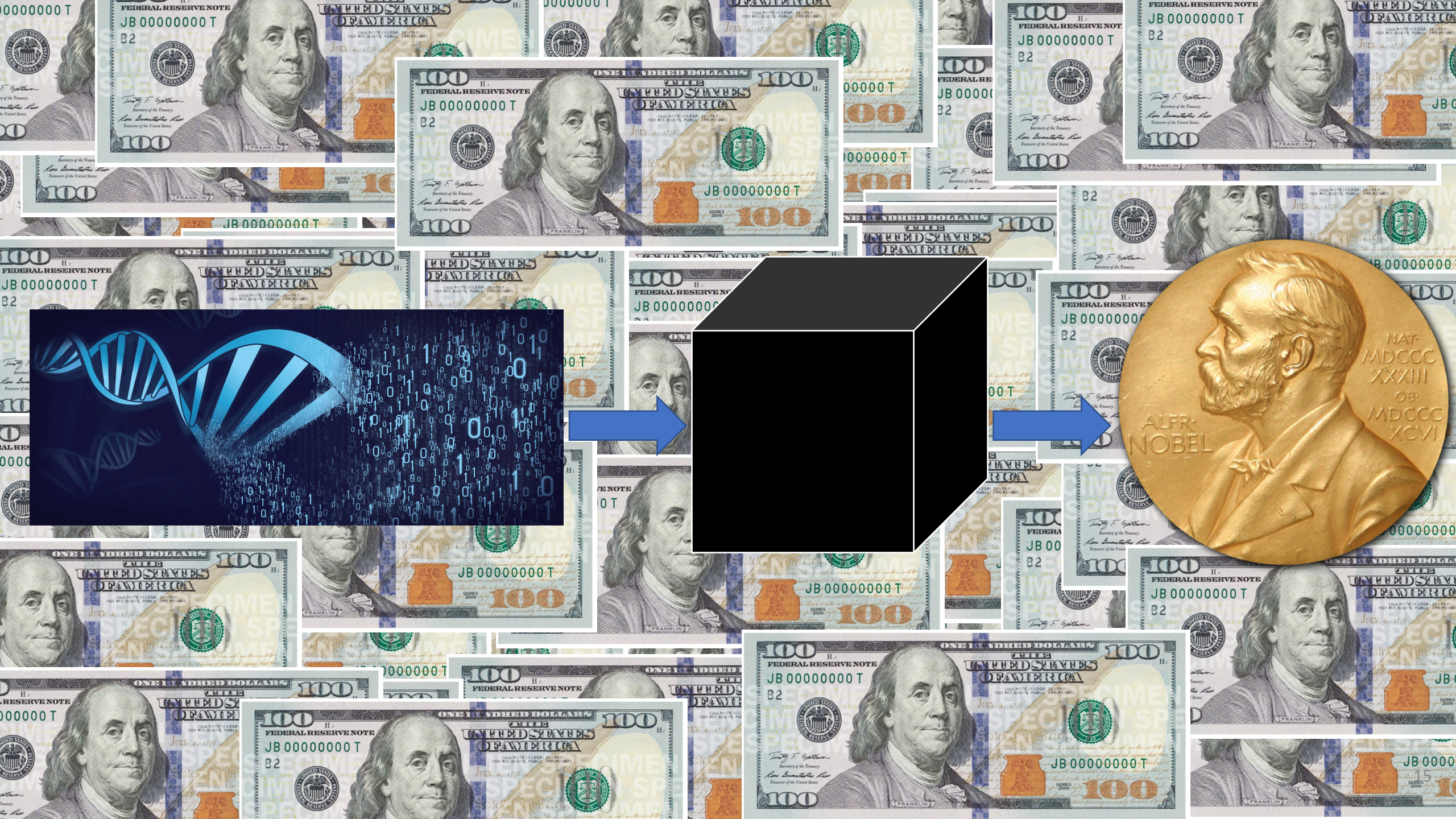














# Data Sharing

SHARE



1



Joseph DeAngelo, the suspected Golden State Killer RANDY PENCH/TNS/NEWSCOM

## We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans

By [Jocelyn Kaiser](#) | Oct. 11, 2018, 2:00 PM

SHARE



1



Joseph DeAngelo, the suspected Golden State Killer RANDY PENCH/TNS/NEWSCOM

## We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans

By Jocelyn Kaiser | Oct. 11, 2018, 2:00 PM

## Most White Americans' DNA Can Be Identified Through Genealogy Databases



Only two percent of the population needs to have done a DNA test to identify nearly everyone else, researchers found. Leonard Lessin/Science Source

By Heather Murphy

Oct. 11, 2018

# Secure Multiparty Computation?

- Machine learn models over encrypted genomic data
- But if we go in this direction, we need
  - Software tools to rapidly design new functions to support evolving statistical techniques
  - Automatically blend software and hardware to optimize speed
  - Accountability for one's actions (blockchain can help)
  - To get comfortable with not seeing individual-level data

# Synthetic Data?

## Ensuring electronic medical record simulation through better training, modeling, and evaluation

Ziqi Zhang, Chao Yan , Diego A Mesa, Jimeng Sun, Bradley A Malin

[Author Notes](#)

*Journal of the American Medical Informatics Association*, Volume 27, Issue 1, January 2020, Pages 99–108, <https://doi.org/10.1093/jamia/ocz161>

**Published:** 08 October 2019 **Article history** ▼



# Synthetic Data?

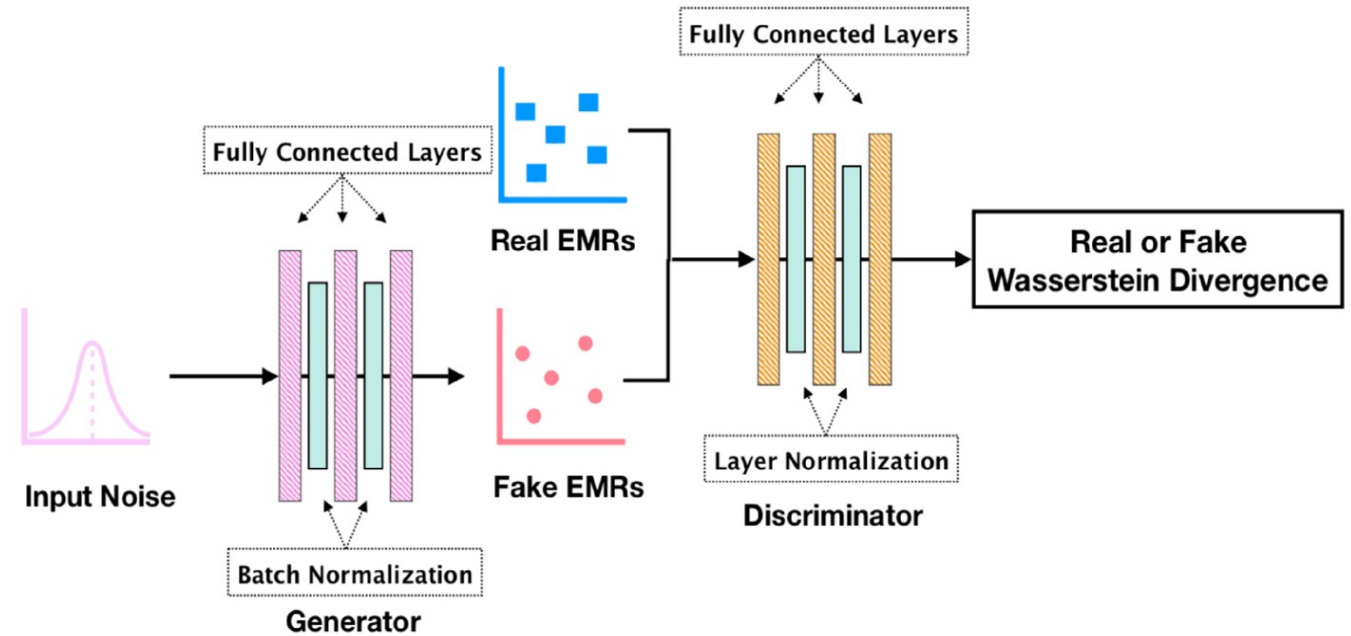
## Ensuring electronic medical record simulation through better training, modeling, and evaluation FREE

Ziqi Zhang, Chao Yan ✉, Diego A Mesa, Jimeng Sun, Bradley A Malin

[Author Notes](#)

*Journal of the American Medical Informatics Association*, Volume 27, Issue 1, January 2020, Pages 99–108, <https://doi.org/10.1093/jamia/ocz161>

**Published:** 08 October 2019 **Article history** ▼



# Synthetic Data?

## Ensuring electronic medical record simulation through better training, modeling, and evaluation FREE

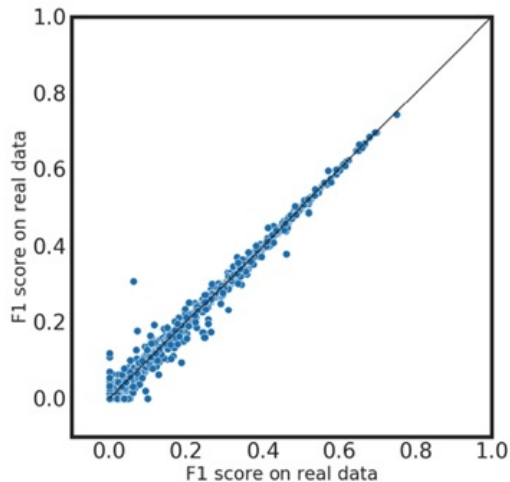
Ziqi Zhang, Chao Yan ✉, Diego A Mesa, Jimeng Sun, Bradley A Malin

[Author Notes](#)

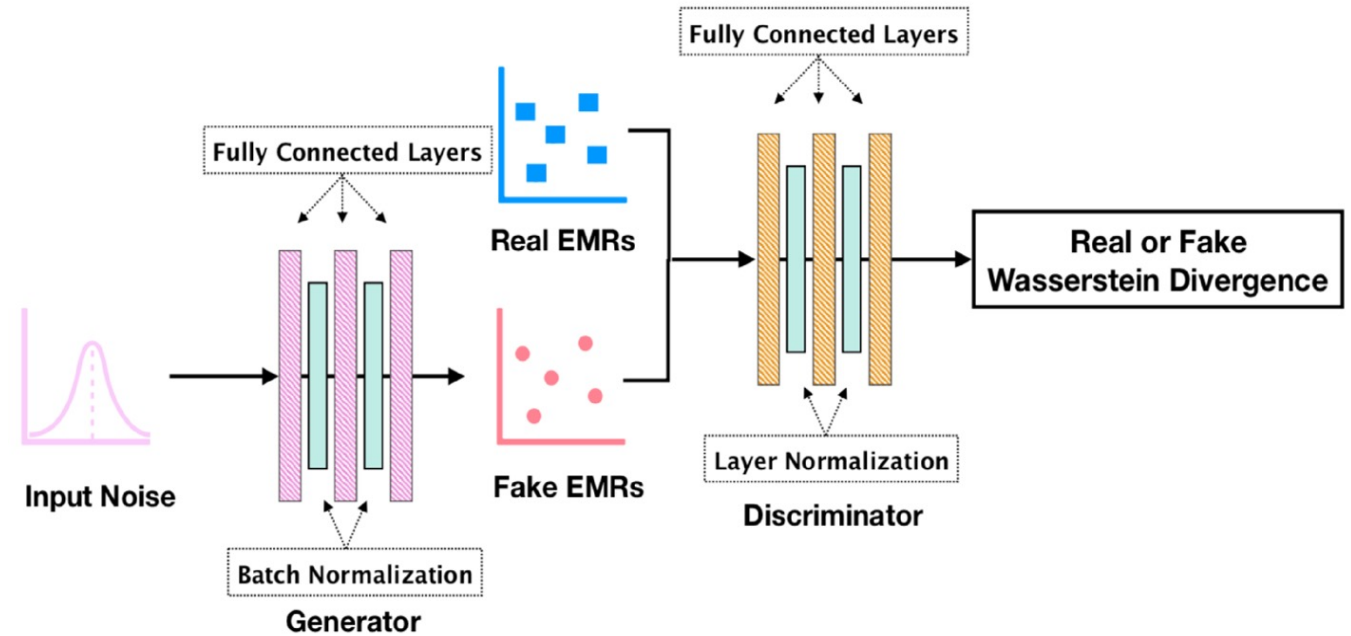
*Journal of the American Medical Informatics Association*, Volume 27, Issue

1, January 2020, Pages 99–108, <https://doi.org/10.1093/jamia/ocz161>

**Published:** 08 October 2019 **Article history** ▾



(a) Real.



# Synthetic Data?

## Ensuring electronic medical record simulation through better training, modeling, and evaluation FREE

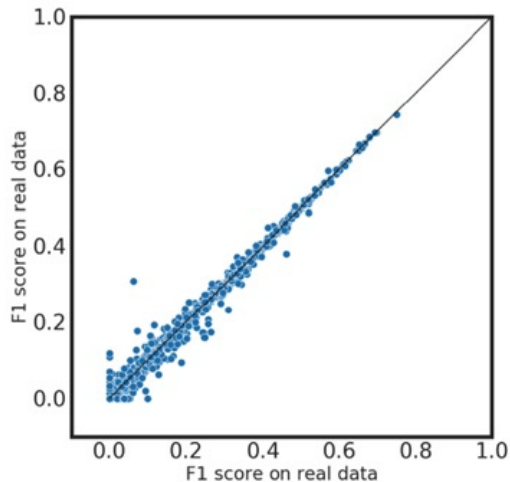
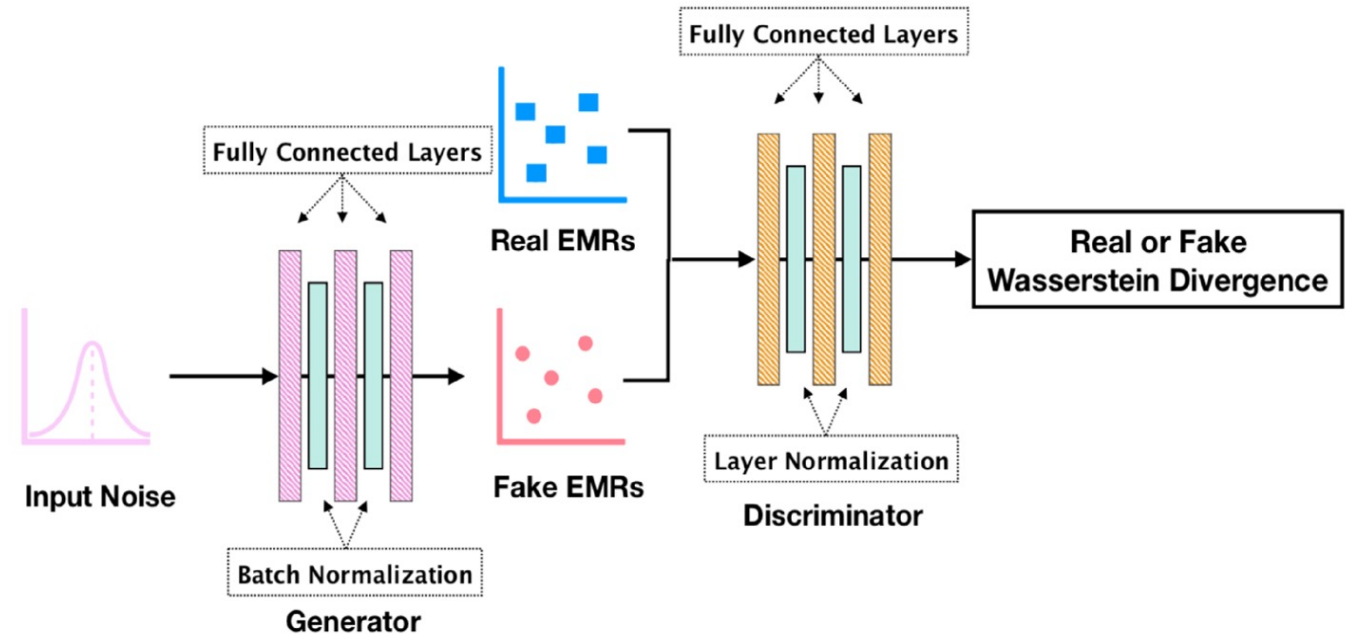
Ziqi Zhang, Chao Yan ✉, Diego A Mesa, Jimeng Sun, Bradley A Malin

Author Notes

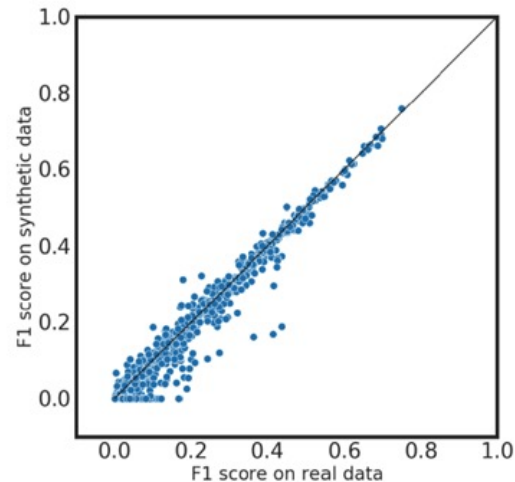
*Journal of the American Medical Informatics Association*, Volume 27, Issue

1, January 2020, Pages 99–108, <https://doi.org/10.1093/jamia/ocz161>

Published: 08 October 2019 [Article history](#) ▼



(a) Real.



(i) EMR-WGAN.

# Synthetic Data?

## Ensuring electronic medical record simulation through better training, modeling, and evaluation FREE

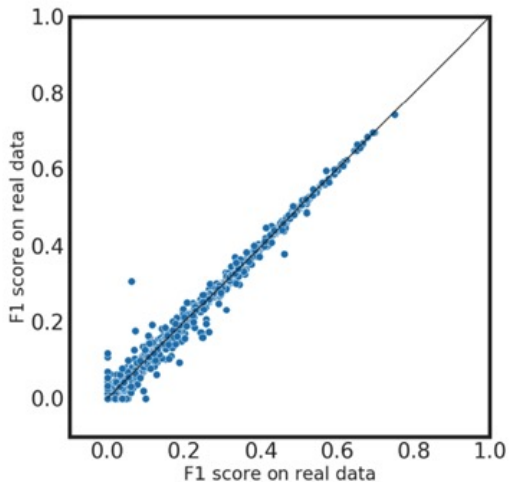
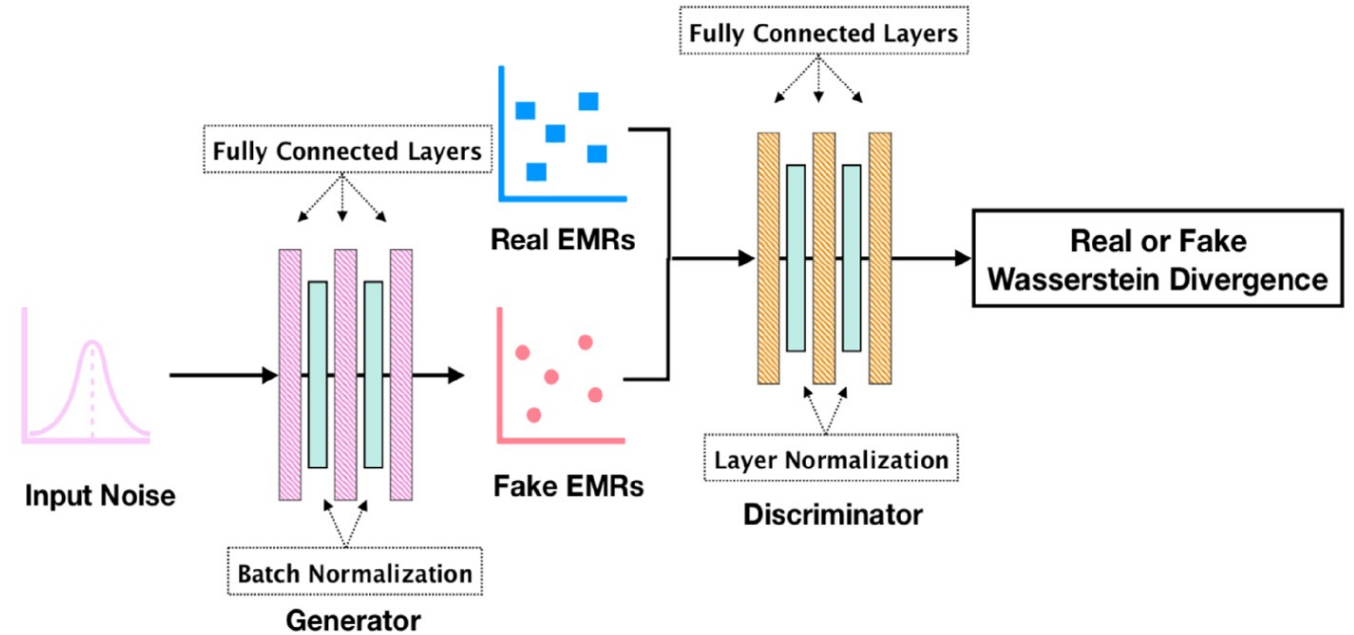
Ziqi Zhang, Chao Yan ✉, Diego A Mesa, Jimeng Sun, Bradley A Malin

Author Notes

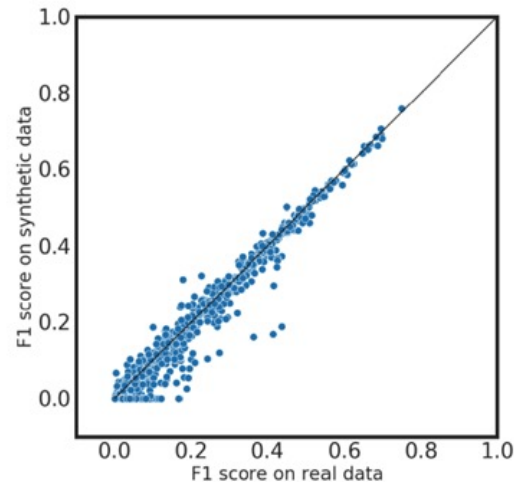
*Journal of the American Medical Informatics Association*, Volume 27, Issue

1, January 2020, Pages 99–108, <https://doi.org/10.1093/jamia/ocz161>

Published: 08 October 2019 [Article history](#) ▼



(a) Real.



(i) EMR-WGAN.

## PLOS GENETICS

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

### Creating artificial human genomes using generative neural networks

Burak Yelmen ✉, Aurélien Decelle, Linda Ongaro, Davide Marnetto, Corentin Tallec, Francesco Montinaro, Cyril Furtlehner, Luca Pagani, Flora Jay ✉

Published: February 4, 2021 • <https://doi.org/10.1371/journal.pgen.1009303>

# Decision Support

# Moving ML Models into Practice





# Moving ML Models into Practice



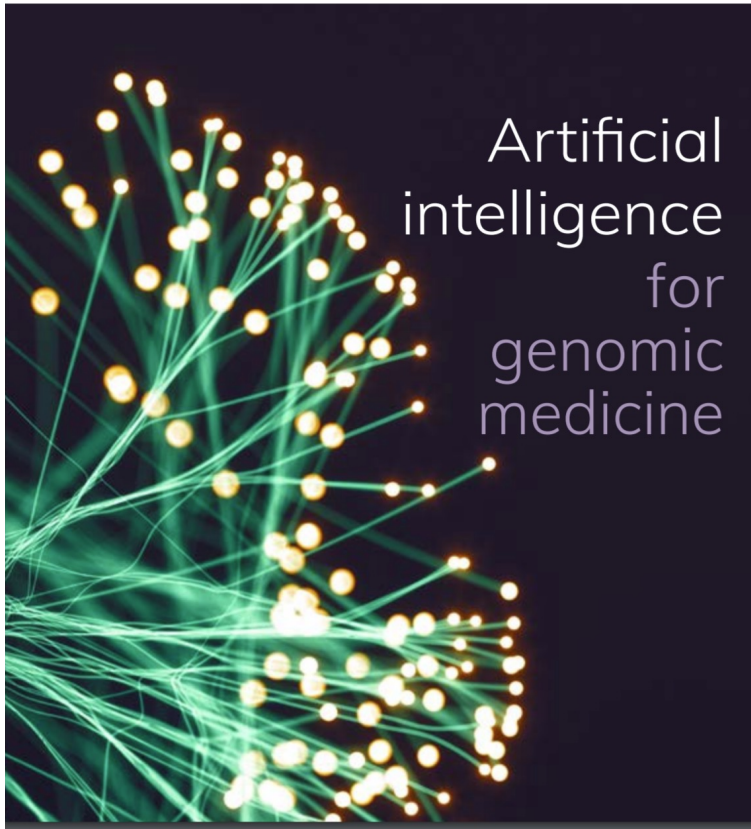
## FDA Cleared AI Algorithms

Our list of FDA cleared AI algorithms provides valuable details on each model, bringing all of the relevant information together for easy access. Convenience area, modality, predicate devices, product testing and evaluation related to product performance, and clinical validation. Our Define-AI use cases match clicking on the model will take you directly to the FDA summary.

Check back regularly to see which new algorithms are available and have been added to the list. Send information on AI algorithms that are not listed and

Product	Company	Subspecialty	Body Area
Koios DS	Koios Medical, Inc	Women's Imaging	Breast
SubtlePET	Subtle Medical, Inc.	Neuroradiology	Brain
Vitrea CT Brain Perfusion	Vital Images, Inc.	Neuroradiology	Brain
BriefCase- ICH	Aidoc Medical, Ltd.	Neuroradiology	Brain
HealthCCS	Zebra Medical Vision Ltd.	Cardiac Imaging	Coronary Arteries
OsteoDetect	Imagen Technologies	Musculoskeletal Imaging	Wrist
Viz CTP	Viz. ai, inc.	Neuroradiology	Brain
Quantib Brain	Quantib BV	Neuroradiology	Brain
DM-Density	Densitas, Inc.	Women's Imaging	Breast
ContaCT	Viz. ai, inc.	Neuroradiology	Brain
QuantX	Quantitative Insights, Inc.	Women's Imaging	Breast
Arterys Cardio DL	Arterys Inc.	Cardiac Imaging	Heart
ClearView cCAD	ClearView Diagnostics Inc.	Women's Imaging	Breast
CT CoPilot	ZepMed, LLC.	Neuroradiology	Brain
ClearRead CT	Riverain Technologies	Chest Imaging	Lung
Vitrea CT Lung Density Analysis Software	Vital Images, Inc.	Chest Imaging	Lung
Lung Density Analysis	Imbio LLC	Chest Imaging	Lung
ClearRead +Confirm	Riverain Technologies	Chest Imaging	Chest
DeltaView Model 2.1	Riverain Technologies	Chest Imaging	Chest
Critical Care Suite	GE Medical Systems	Chest Imaging	Lung
HealthPNX	Zebra Medical Vision Ltd.	Chest Imaging	Lung
cmTriage	CureMetrix, Inc.	Women's Imaging	Breast
Aidoc BriefCase- CSF triage	Aidoc Medical, Ltd.	Neuroradiology	Cervical Spine
Aidoc Briefcase- ICH and PE triage	Aidoc Medical, Ltd.	Chest Imaging	Pulmonary Arteries
HealthCXR	Zebra Medical Vision Ltd.	Chest Imaging	Lung
HealthICH	Zebra Medical Vision Ltd.	Neuroradiology	Brain

# Moving ML Models into Practice



## FDA Cleared AI Algorithms

Our list of FDA cleared AI algorithms provides valuable details on each model, bringing all of the relevant information together for easy access. Convenience area, modality, predicate devices, product testing and evaluation related to product performance, and clinical validation. Our Define-AI use cases match clicking on the model will take you directly to the FDA summary.

Check back regularly to see which new algorithms are available and have been added to the list. Send information on AI algorithms that are not listed and

Product	Company	Subspecialty	Body Area
<a href="#">Koios DS</a>	Koios Medical, Inc.	Women's Imaging	Breast
<a href="#">SubtlePET</a>	Subtle Medical, Inc.	Neuroradiology	Brain
<a href="#">Vitrea CT Brain Perfusion</a>	Vital Images, Inc.	Neuroradiology	Brain
<a href="#">BriefCase- ICH</a>	Aidoc Medical, Ltd.	Neuroradiology	Brain
<a href="#">HealthCCS</a>	Zebra Medical Vision Ltd.	Cardiac Imaging	Coronary Arteries
<a href="#">OsteoDetect</a>	Imagen Technologies	Musculoskeletal Imaging	Wrist
<a href="#">Viz CTP</a>	Viz. ai. inc.	Neuroradiology	Brain

Over 100 approvals

<a href="#">ClearView cCAD</a>	ClearView Diagnostics Inc.	Women's Imaging	Breast
<a href="#">CT CoPilot</a>	ZepMed, LLC.	Neuroradiology	Brain
<a href="#">ClearRead CT</a>	Riverain Technologies	Chest Imaging	Lung
<a href="#">Vitrea CT Lung Density Analysis Software</a>	Vital Images, Inc.	Chest Imaging	Lung
<a href="#">Lung Density Analysis</a>	Imbio LLC	Chest Imaging	Lung
<a href="#">ClearRead +Confirm</a>	Riverain Technologies	Chest Imaging	Chest
<a href="#">DeltaView Model 2.1</a>	Riverain Technologies	Chest Imaging	Chest
<a href="#">Critical Care Suite</a>	GE Medical Systems	Chest Imaging	Lung
<a href="#">HealthPNX</a>	Zebra Medical Vision Ltd.	Chest Imaging	Lung
<a href="#">cmTriage</a>	CureMetrix, Inc.	Women's Imaging	Breast
<a href="#">Aidoc BriefCase- CSF triage</a>	Aidoc Medical, Ltd.	Neuroradiology	Cervical Spine
<a href="#">Aidoc Briefcase- ICH and PE triage</a>	Aidoc Medical, Ltd.	Chest Imaging	Pulmonary Arteries
<a href="#">HealthCXR</a>	Zebra Medical Vision Ltd.	Chest Imaging	Lung
<a href="#">HealthICH</a>	Zebra Medical Vision Ltd.	Neuroradiology	Brain



# AI / ML Algorithms Need Support

# AI / ML Algorithms Need Support

- Models based on genomics have potential to include 1000s of variables

# AI / ML Algorithms Need Support

- Models based on genomics have potential to include 1000s of variables
- Decision support needs to be **verifiable**

# AI / ML Algorithms Need Support

- Models based on genomics have potential to include 1000s of variables
- Decision support needs to be **verifiable**
- Decisions need to be **auditable** (and tied to algorithm versions)

# AI / ML Algorithms Need Support

- Models based on genomics have potential to include 1000s of variables
- Decision support needs to be **verifiable**
- Decisions need to be **auditable** (and tied to algorithm versions)
- Guidance must be **explainable**

# AI / ML Algorithms Need Support

- Models based on genomics have potential to include 1000s of variables
- Decision support needs to be **verifiable**
- Decisions need to be **auditable** (and tied to algorithm versions)
- Guidance must be **explainable**
- Accessibility and applicability must be **equitable**

# Parting Thoughts

- We need BIG data... but it must be DIVERSE
- We need COST EFFECTIVE storage and compute
- We need TRUST, which may require bringing the LEARN to the DATA
- IF machine learning models are put into services (e.g., clinical care) - we need EXPLAINABLE findings

# Thank you! Questions?

[b.malin@vumc.org](mailto:b.malin@vumc.org)



Vanderbilt Health Data Science Center

<https://www.vumc.org/heads/>

Center for Genetic Privacy and Identity in  
Community Settings

An NHGRI CEER (RM1 HG009034)

<https://www.vumc.org/getprecise>

