Concept Clearance: ML/AI Tools to Advance Genomic Translational Research

Sandhya Xirasagar, Ph.D., Office of Genomic Data Science
Ajay Pillai Ph.D., Division of Genomic Science
Robb Rowley M.D., Division of Genomic Medicine

National Advisory Council for Human Genome Research
September 18, 2023

Outline

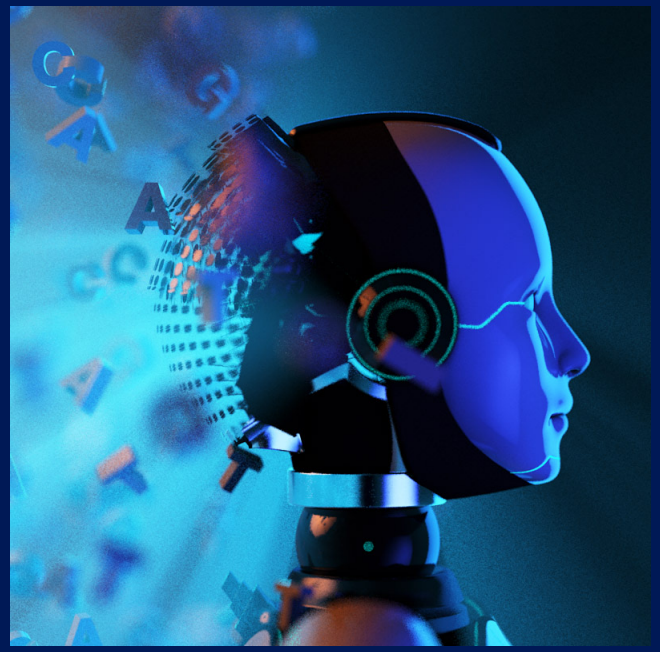
- Objective
- Background and Rationale
- Scope
- Budget
- Acknowledgements
- Feedback

Objective

Spur the development of novel Machine Learning/Artificial Intelligence (ML/AI) tools to explore their potential to advance genomic translational research

Specifically, the aim is to model pleiotropy and variable penetrance through the learning and classification capabilities of ML/AI to uncover novel relationships between genotypes and phenotypes. Tools will be developed in a shared, agreed upon Ethical, Legal, and Social Implications (ELSI) framework .

Machine Learning/Artificial Intelligence



Artificial intelligence (AI) is the capability of a computer system to mimic human cognitive functions such as learning and problem-solving.

Machine learning (ML) is an application of AI where mathematical models of data are used to help a computer learn without direct instruction.

https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/artificial-intelligence-vs-machine-learning
https://www.genome.gov/about-genomics/educational-resources/fact-sheets/artificial-intelligence-machine-learning-and-genomics

ML/AI in Research and Medicine

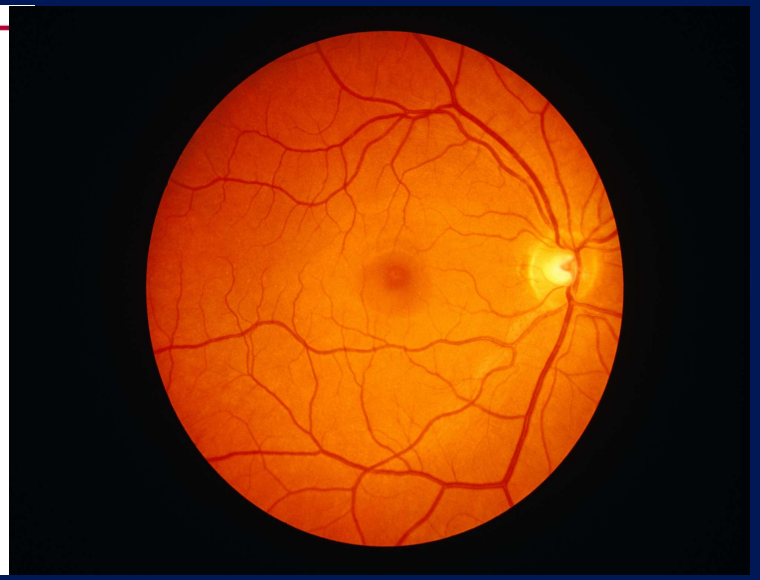
A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study

Carol Y Cheung, An Ran Ran*, Shujun Wang*, Victor T T Chan, Kaiser Sham, Saima Hilal, Narayanaswamy Venketasubramanian, Ching-Yu Cheng, Charumathi Sabanayagam, Yih Chung Tham, Leopold Schmetterer, Gareth J McKay, Michael A Williams, Adrian Wong, Lisa W C Au, Zhihui Lu, Jason C Yam, Clement C Tham, John J Chen, Oana M Dumitrascu, Pheng-Ann Heng, Timothy C Y Kwok, Vincent C T Mok†, Dan Milea†, Christopher Li-Hsian Chen†, Tien Yin Wong†*

Summary

Background There is no simple model to screen for Alzheimer's disease, partly because the diagnosis of Alzheimer's disease itself is complex—typically involving expensive and sometimes invasive tests not commonly available outside

Lancet Digit Health 2022; 4: e806–15



ML/AI in Research and Medicine

Original Investigation

FREE

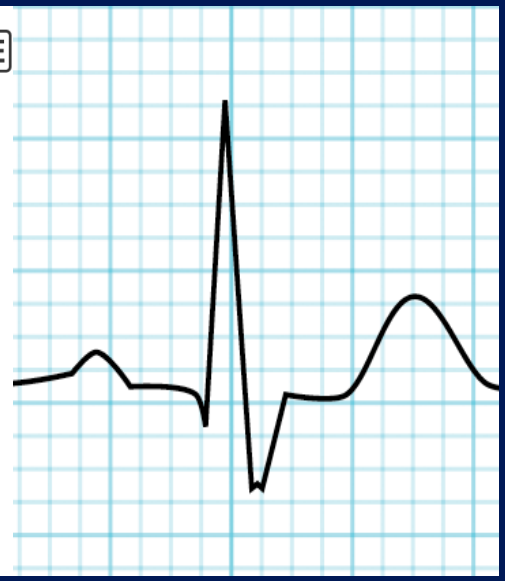
April 3, 2019

Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram

Conner D. Galloway, MSc[1]; Alexander V. Valys, BS[1]; Jacqueline B. Shreibati, MD[1]; et al

» Author Affiliations | Article Information

JAMA Cardiol. 2019;4(5):428-436. doi:10.1001/jamacardio.2019.0640



Use Case - Pathogenic Variants

Penetrance and Pleiotropy → Phenotypic Differences

CFTR - Cystic Fibrosis

Table 1

Hierarchy of associations with mutations in the cystic fibrosis transmembrane regulator gene

		Genetic/other influences	
Phenotypes associated with *CFTR* mutations	*CFTR*	Non-*CFTR* gene modifiers	Environment
'Atypical' CF*			
CBAVD	+++	+	+
Mild pulmonary disease	+++	+	+
ICP†	+++	+	+
Associated with mutations in CFTR‡			
Sinusitis	+	++	+
ABPA	+	++	+++
Asthma	+/-§	+++	++

Genes 2021, 12, 562.

Use Case - Pathogenic Variants

Penetrance and Pleiotropy → Phenotypic Differences

CFTR - Cystic Fibrosis

Environmental factor

Mycobacteroides abscessus

	Odds ratio **95% CI) *P*-value**
Manganese (1-log unit)	***0.74* (0.60, 0.90) *0.002***
Mercury (1-log unit)	***1.45* (1.09, 1.93) *0.010***
Molybdenum (1-log unit)	***1.36* (1.17, 1.59) *0.0001***
Phosphorus (1-log unit)	***1.25* (1.05, 1.49) *0.012***

Environ. Epi. 7(5):p e266, October 2023.

Use Case - Pathogenic Variants

Penetrance and Pleiotropy → Phenotypic Differences

CFTR - Cystic Fibrosis

Environmental factor + Genetic modifier

NHGRI Rationale

- Feedback from NHGRI workshops
  - Genomic Medicine XIII, 2021
  - Machine Learning In Genomics, 2021
- Aligns with NHGRI 2020 Strategic Vision

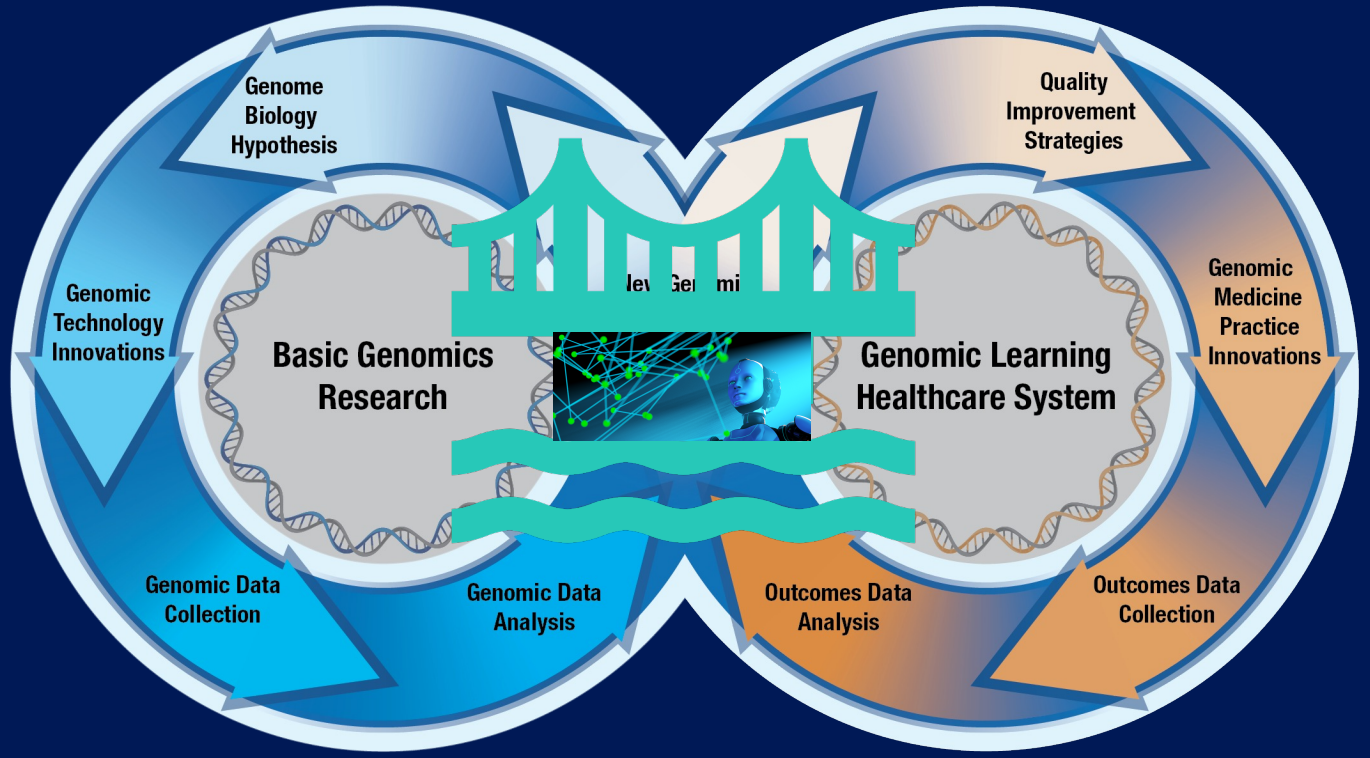
> "Provide a conceptual framing that consistently conveys the role of both genomic and non-genomic contributors to health and disease – routinely considering the importance of social and environmental contributions to human health and the interactions…"

Objective

Spur the development of novel Machine Learning/Artificial Intelligence (**ML/AI**) tools to explore their potential to advance genomic translational research

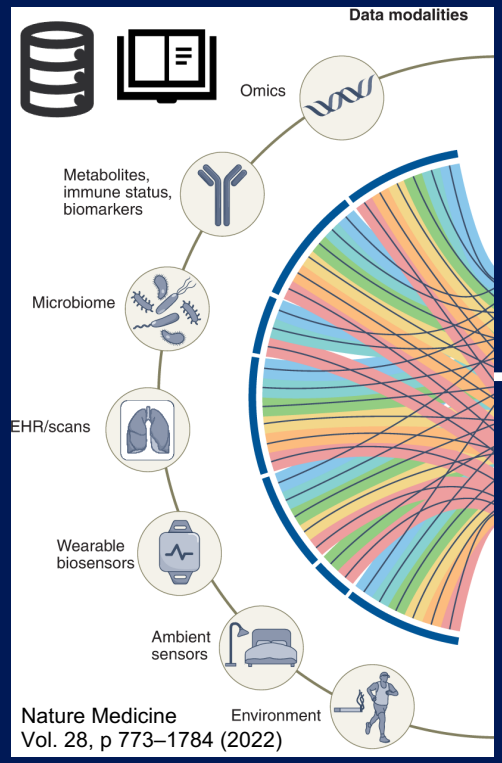
Specifically, the aim is to model pleiotropy and variable penetrance through the learning and classification capabilities of ML/AI to uncover ***novel relationships between genotypes and phenotypes****. Tools will be developed in a shared, agreed upon* ***Ethical, Legal, and Social Implications (ELSI) framework*** *.*

ML/AI Tools to Advance Genomic Translational Research

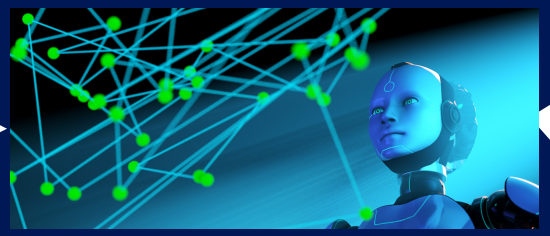


Nature. 2020 Oct;586(7831):683-692.

Scope



CFTR+



Lung infections

Asthma

Sensitive to Smoke

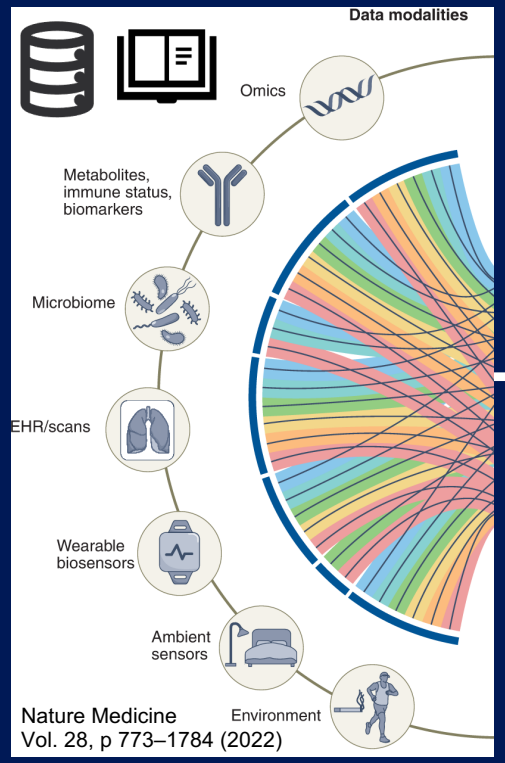
Others…

Existing Datasets
(Genomics, Omics, Phenotypic,
Social Determinants of Health..)

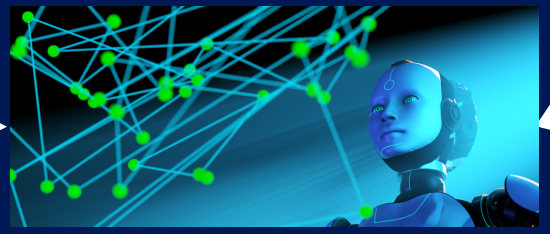
+

Reference Knowledge

Scope



Train



Address ELSI concerns

Develop and validate ML/AI tools within a consortium-established ELSI Framework.

Existing Datasets
(Genomics, Omics, Phenotypic, Social Determinants of Health..)

+

Reference Knowledge

Program Structure

Biphasic, Collaborative, 2RFAs

Development Sites (UG3)
Design Phase

Consortium will jointly:
- Select diseases and pathogenic variants
- Design tool end-points and outputs and validation plan
- Prepare datasets for cross-validation
- Formulate draft best practices and ELSI framework

Have objectives been achieved?

Development Sites (UH3)
ML/AI Tool Development/Validation Phase

Each site will:
- Develop tools according to framework
- Validate according to plan

Consortium will jointly:
- Refine best practices & framework
- Disseminate resources FAIR-ly
  - Tools, datasets
  - ELSI framework
  - Lessons learned etc.,

Y1 Y2 Y3 Y4 Y5

Coordination Center (U01)

Coordinate logistics for all collaborative activities

Years 3-5 Contingent on Continuation of Development Sites

Relationship to Ongoing Activities

Program	Leverage
eMERGE, PRIMED, GREGoR AllofUS, TOPMed, UKBB, Bridge2AI etc.	Datasets
AIM-AHEAD, Coalition of Health AI…	Best practices, ELSI-relevant resources

Budget

- Total annual costs
  - 3-4* sites for a total cost of $4.8M
  - $1.2M for CC
- Total costs for 5 years: $30M

Multidisciplinary team with expertise in ML/AI, data wrangling software development, clinical research, ELSI, coordination

*Co-funding to be sought from other ICOs for additional sites

Acknowledgements

- ***Valentina Di Francesco***
- ***Rene Sterling***

- Larry Brody
- Lisa Chadwick
- Carolyn Hutter
- Dave Kaufman
- Nicolas Keller
- Nicole Lockhart
- Teri Manolio
- Iman Martin
- Joannella Morales
- Jahnavi Narula
- Mike Pazin
- Colette Pollard
- Erin Ramos
- Shurjo Sen
- Helen Thompson
- Simona Volpi
- Nephi Walton
- Chris Wellington
- Ken Wiley (NCATS)

Feedback