



Designing out Racism: Towards better use of race, ancestry and ethnicity in genomics

Sophia Luu

Design Research Lead, Diverse Data Programme



Few years ago...

- BAME
- Asian
- Mixed Race
- Irritable Bowel Syndrome (IBS)



Now...

- ~~• BAME~~
- ~~• Asian~~
- ~~• Mixed Race~~
- ~~• Irritable Bowel Syndrome (IBS)~~



Now...

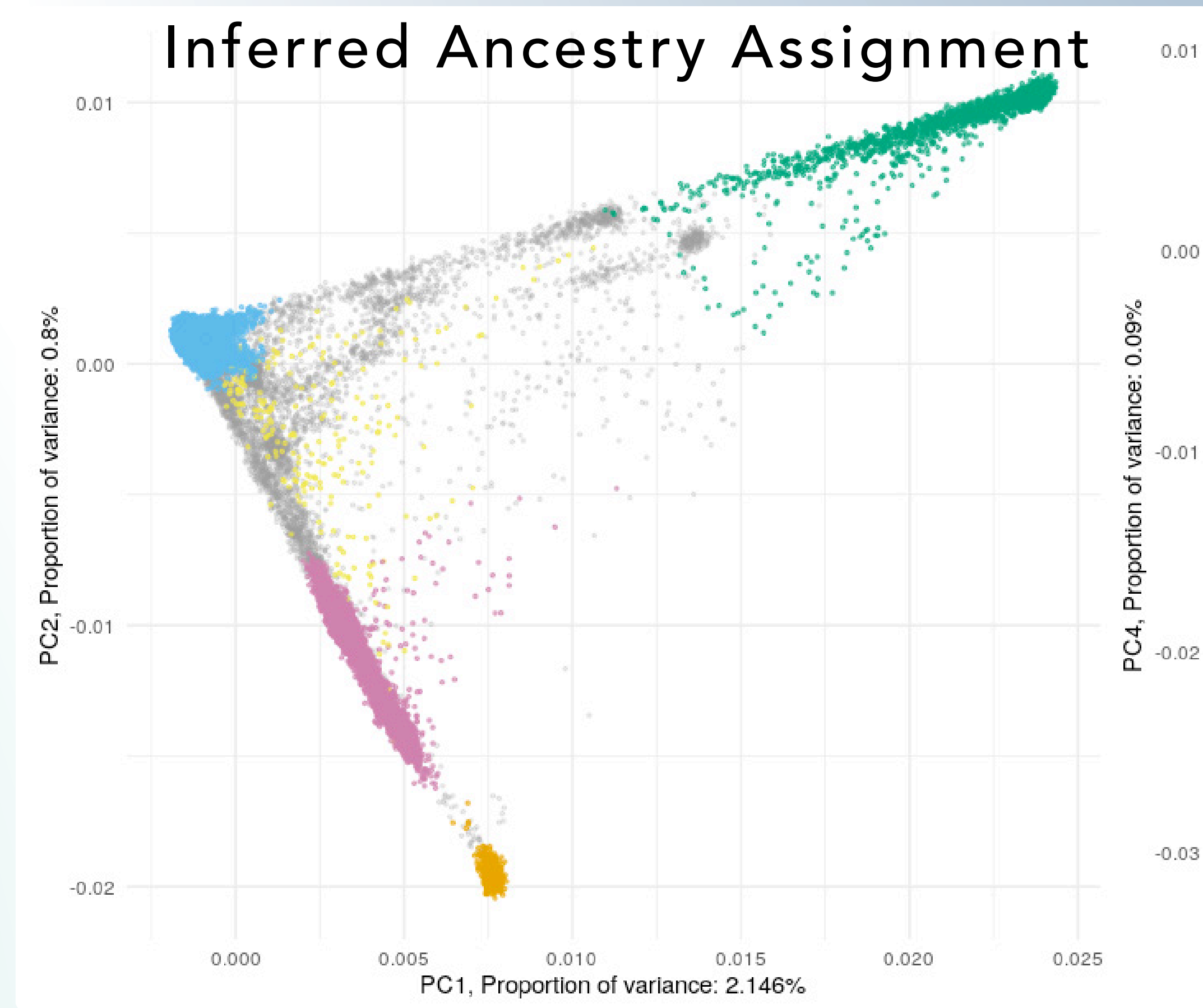
- Person of colour
- British-Vietnamese
- Mixed
- Lactose Intolerant



How does genomics describe me?



- Admixed individual
- Likely European and East Asian
- Likely "Unassigned Ancestry"
- (Variant of LCT gene affects lactose digestion)



And then I dug further...

“

You can tell someone's race from their Genome

”

When we say East Asian superpopulation, we actually mean Han-Chinese. We don't have enough of the others in our dataset.

”

We need more individuals of Black Ancestry

and further...

WORLD VIEW | 24 August 2021

Too many scientists still say Caucasian



Racist ideas of categories for human identity continue to warp research and medicine.

[Alice B. Popejoy](#)



Of the ten clinical genetics labs in the United States that share the most data with the research community, seven include 'Caucasian' as a multiple-choice category for patients' racial or ethnic identity, despite the term having no scientific basis. Nearly 5,000 biomedical papers since 2010 have used 'Caucasian' to describe European populations. This suggests that too many scientists apply the term, either unbothered by or unaware of its roots in racist taxonomies used to justify slavery – or worse, adding to pseudoscientific claims of white biological superiority.

The New York Times

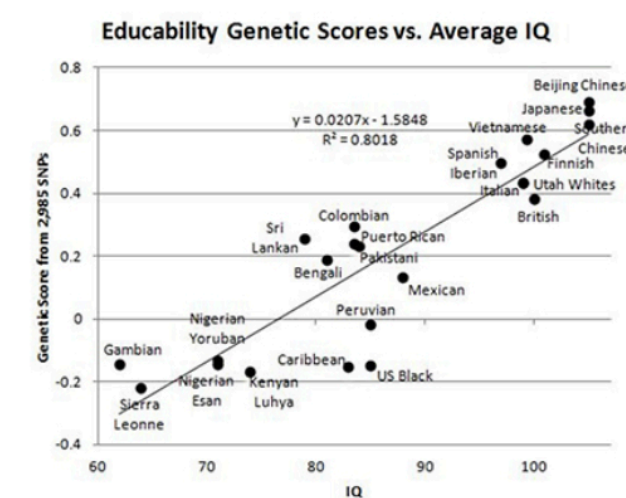
SUBSCRIBE

Why White Supremacists Are Chugging Milk (and Why Geneticists Are Alarmed)

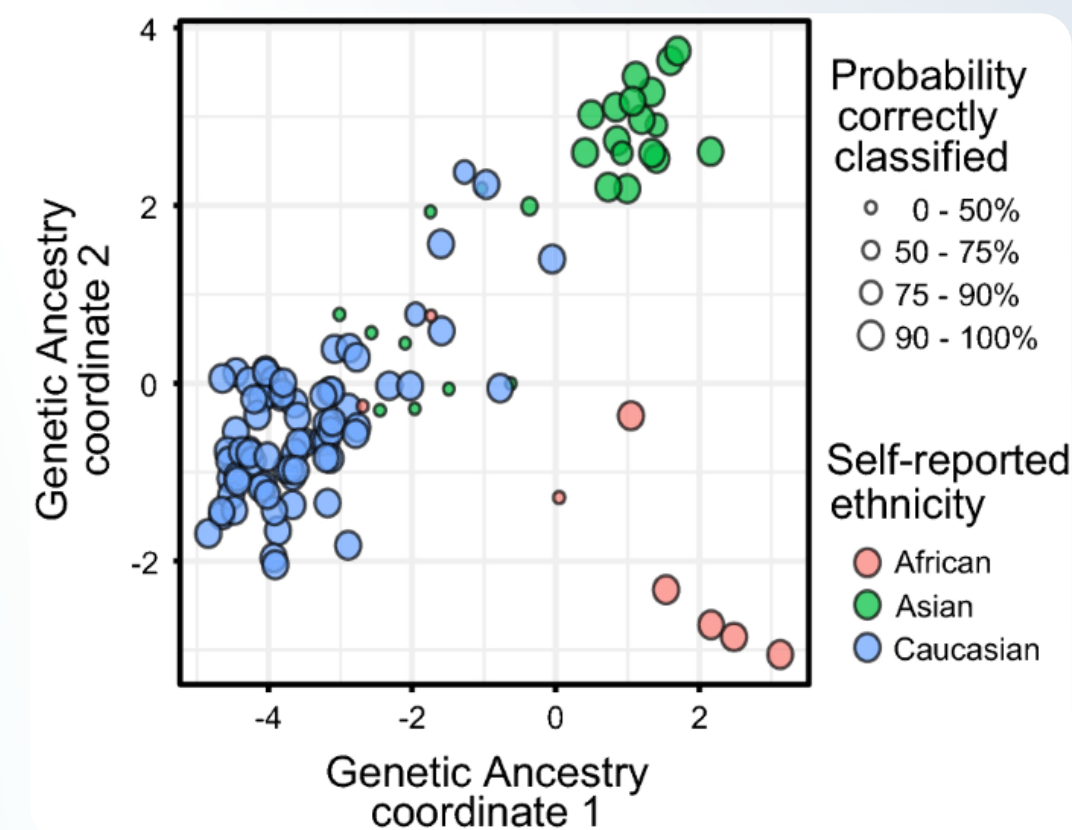
Give this article



400



ancient and modern Egypt. However, the method was only able to detect strong signs of discontinuity between our ancient populations and modern Ethiopians.



**The point is, language is always changing
and context specific.**

The point is, language is always changing and context specific.

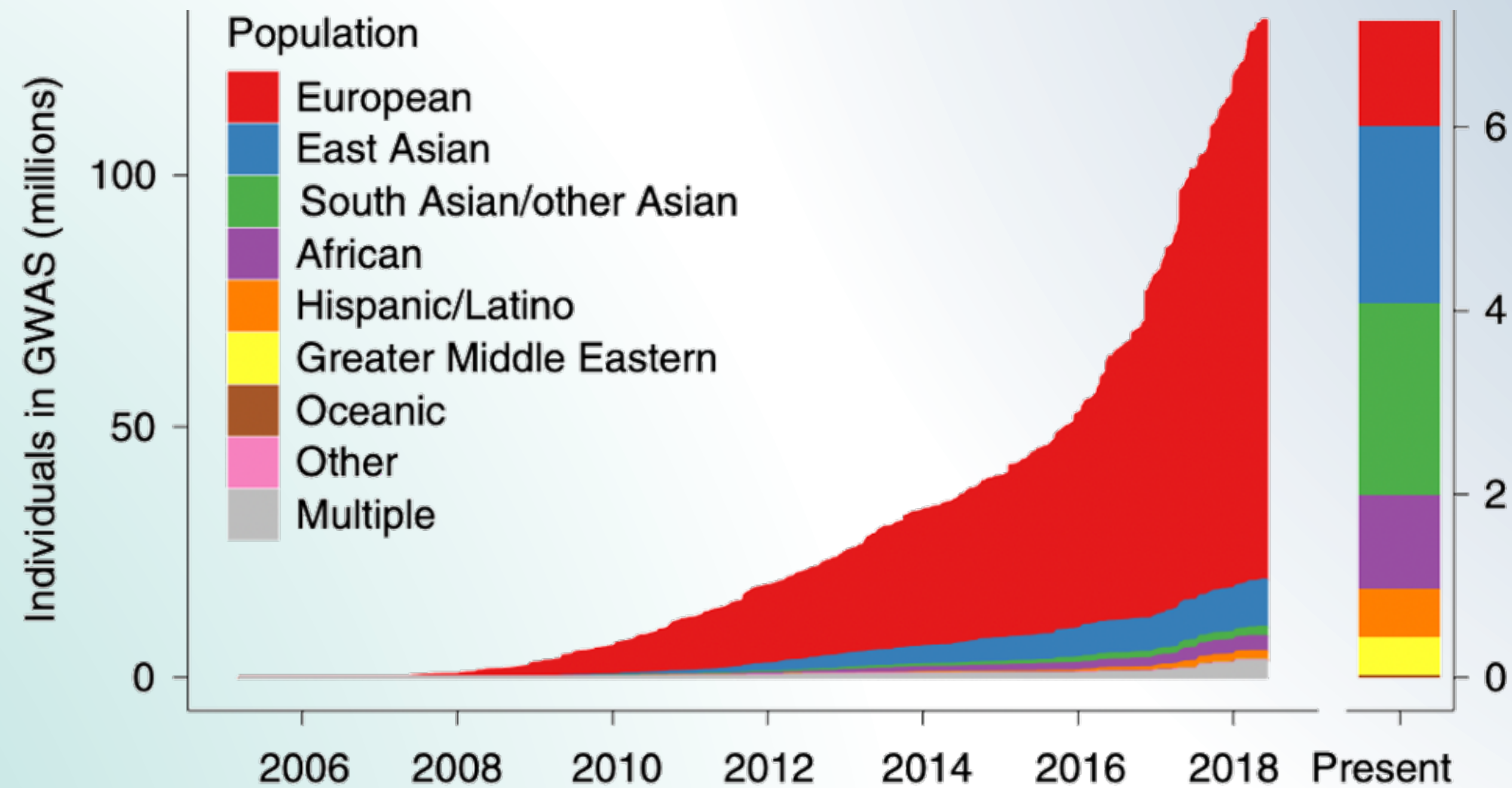
... but when we're dealing with genomics, it's essential to understand when context switching causes bias, unclear science and siloes in problem solving

Working at Genomics England

- Set up and owned by the United Kingdom Department of Health and Social Care
- Whole genome sequencing diagnostics for rare disease and cancer.
- We also equip researchers to find the causes of disease and develop new treatments



Number of individuals by ancestry in Genome Wide Association Studies (GWAS) compared to Global Population



- 78% - European Ancestries in Genome Wide Association Studies (GWAS) compared to 2% of people from African Ancestries
- Polygenic Risk Scores 4.5x more accurate for European ancestries over African ancestries

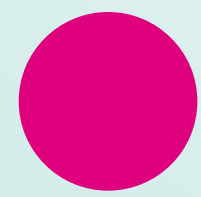
My job as a designer...



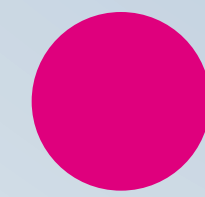
The Problem

A Solution

My job as a designer...

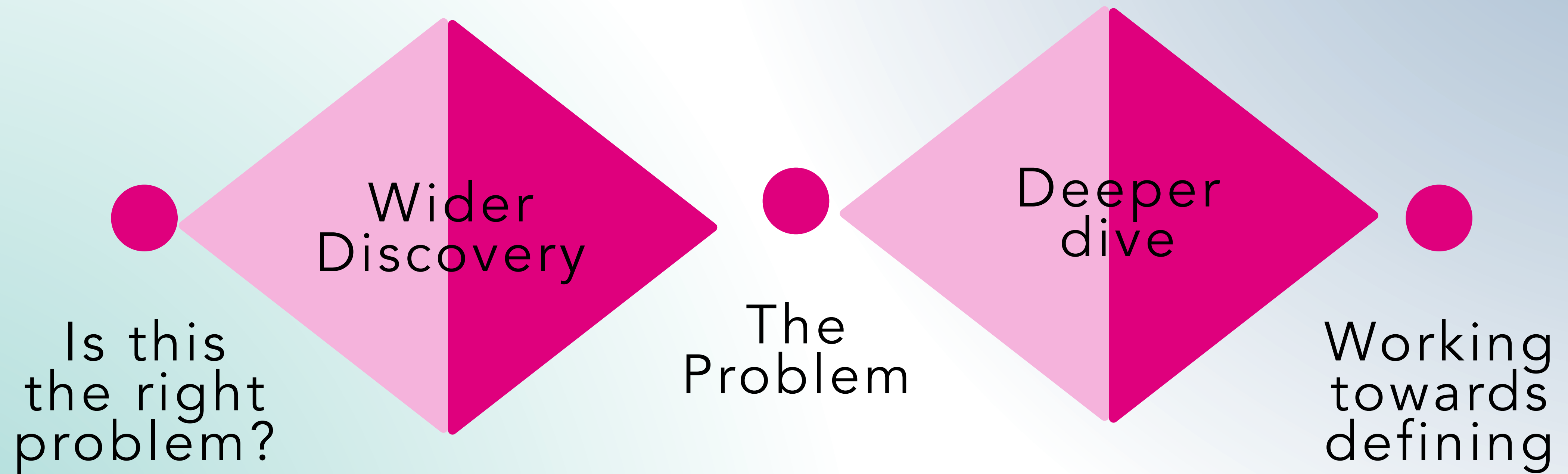


The Problem



A Solution

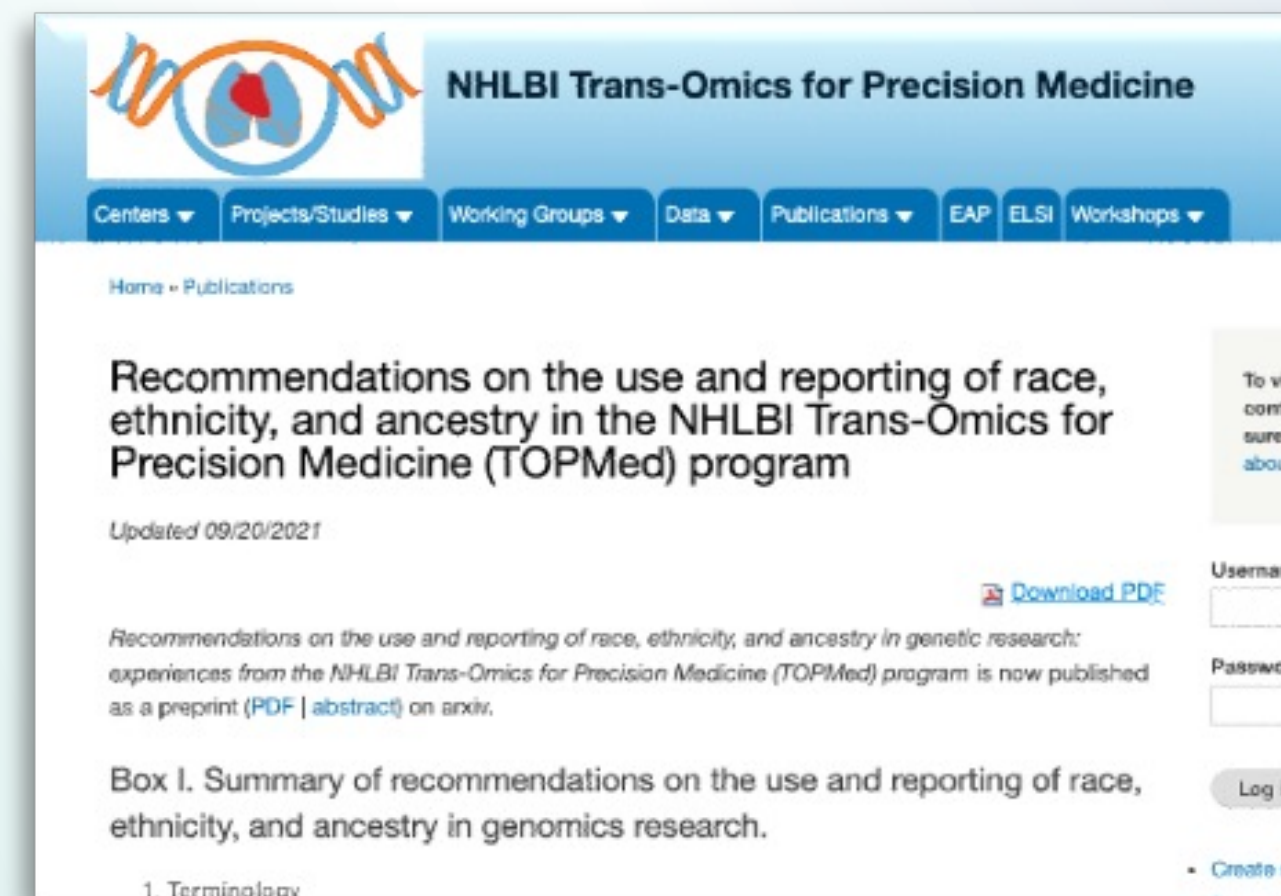
My job as a designer...



How do we reinvent the way language is reviewed and used in genomics, especially with terms that also have a socio-political meaning?

*we used Race, Ethnicity and Ancestry as a starting point

Past Efforts:



> *Hum Mutat.* 2018 Nov;39(11):1713-1720. doi: 10.1002/humu.23644.

The clinical imperative for inclusivity: Race, ethnicity, and ancestry (REA) in genomics

Alice B Popejoy¹, Deborah I Ritter², Kristy Crooks^{3,4}, Erin Currey⁵, Stephanie M Fullerton⁶, Lucia A Hindorf⁶, Barbara Koenig⁷, Erin M Ramos⁶, Elena P Sorokin¹, Hannah Wand¹, Mathew W Wright¹, James Zou¹, Christopher R Gignoux⁴, Vence L Bonham⁸, Sharon E Plon², Carlos D Bustamante¹,
Clinical Genome Resource (ClinGen) Ancestry and Diversity Working Group (ADWG)

Affiliations + expand

PMID: 30311373 PMCID: PMC6188707 DOI: 10.1002/humu.23644

[Free PMC article](#)

Abstract

The Clinical Genome Resource (ClinGen) Ancestry and Diversity Working Group highlights the need to develop guidance on race, ethnicity, and ancestry (REA) data collection and use in clinical genomics. We present quantitative and qualitative evidence to characterize: (1) acquisition of REA data via clinical laboratory requisition forms, and (2) information disparity across populations in the Genome Aggregation Database (gnomAD) at clinically relevant sites ascertained from annotations in ClinVar. Our requisition form analysis showed substantial heterogeneity in clinical laboratory ascertainment of REA, as well as marked incongruity among terms used to define REA categories. There was also striking disparity across REA populations in the amount of information available about clinically relevant variants in gnomAD. European ancestral populations constituted the

But there is a gap...

- Co-creation and authorship across disciplines and experiences
- People who are most affected by language use being involved
- Assumption-free terminology choices
- Timely and frequently updated resource
- Practical guidance for use in multiple contexts
- Agreed shared values for prioritisation during conflict

The Language Toolkit

Traffic Light system



Green Listed	Amber Listed	Red Listed	Pending
Generally accepted for use in the context defined	Be aware that there are multiple definitions based on different contexts	This term should be avoided and there are better alternatives	Words suggested by others which are awaiting definition

Green Listed - Generally ac 21	Amber Listed - Be aware th 21	Red Listed - this term shoul 11
Ancestry	Admixed	BAME <small>1</small>
Bias	Assigned Race and Ethnicity	BIPOC
Consanguinity	Background and Descent	Caucasian
Cultural Sensitivity and Competency	Black	Global Majority
Culture	Decolonisation	Global North and Global South

The Language Toolkit

Individual entries

- Definition
- Importance for Genomics England
- Use in other contexts
- Alternatives
- Debate around this word
- Words associated with the debate
- Learn more - including 'golden' examples



Add cover

Admixed

Traffic Light Rating **Amber Listed - Be aware that there are multiple definitions based on different context**

Last updated October 10, 2022

Main Context **Genomics**

+ Add a property

The Language Toolkit

Navigation

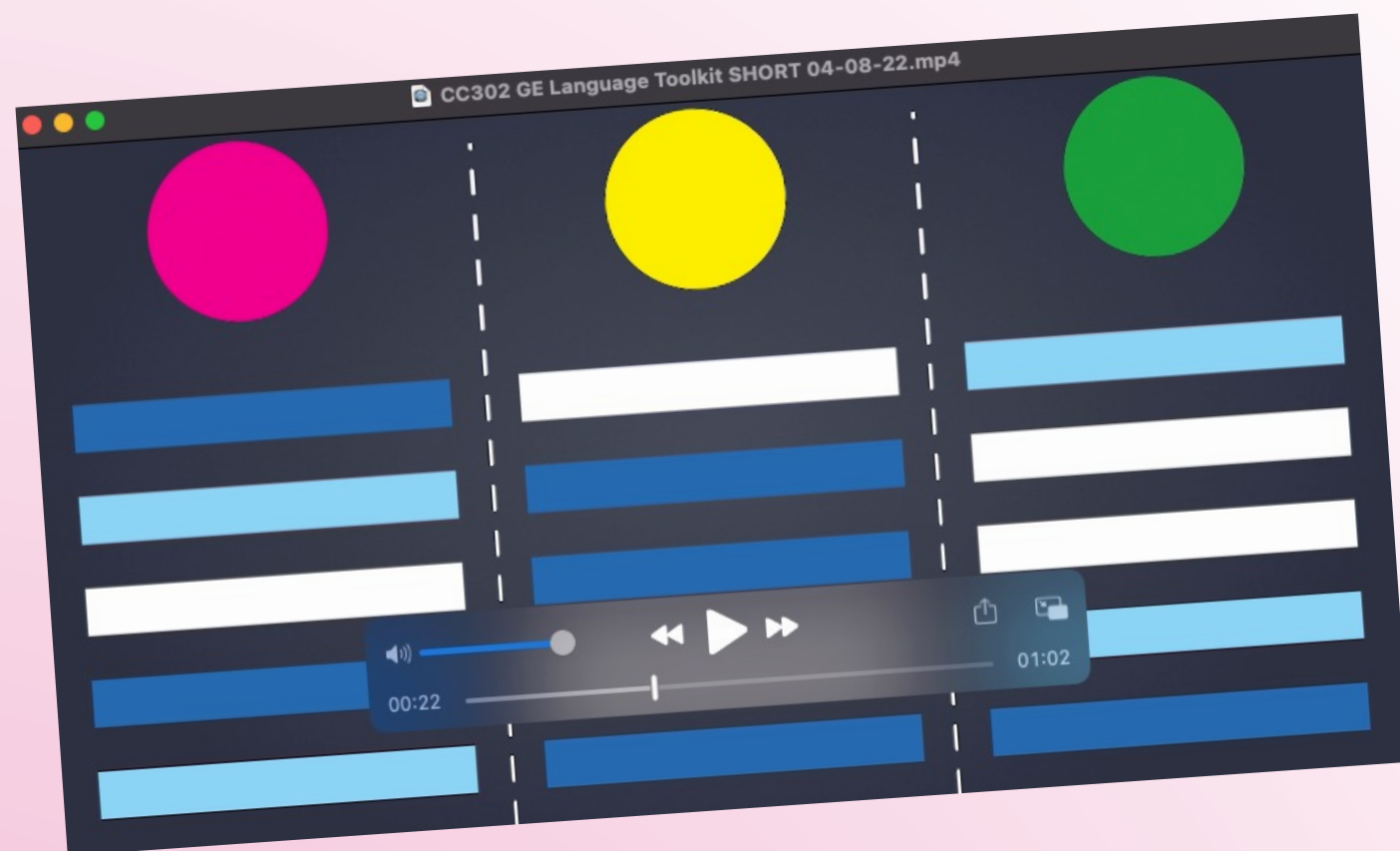
- Alphabetical
- By context
- By rating

The screenshot shows the 'Terminology' interface with a filter menu open. The main context is set to 'Genomics'. The filter menu includes options for 'To-do', 'Multiple Contexts', 'Societal', 'Genomics' (selected), 'In Progress', 'In progress', 'Complete', and 'Done'. The main list shows terms like 'Admixed', 'East Asian', 'Innateness', 'Poor Response', 'Self-Reported Race and Ethnicity', 'South Asian', and 'Super Populations'. A 'Red Listed' section on the right contains terms like 'Caucasian', 'Inbreeding', 'Non-Pure and Pure', 'Self-Reported Ancestry', and 'Trans-Ancestry and Trans-Ethnicity'.



The Language Toolkit

Tailored Learning Aids



Ancestry, race and ethnicity

A simplified quick guide

Genomics England

Genetic ancestry

✓ **Visible in a genome**

Shows where genomic variants have come from, which can be helpful for determining someone's risk of disease.

You need to take a genetic test to know your ancestry. At Genomics England, we follow a standard for grouping ancestry by five larger reference groups, sometimes called "super populations"

- European
- African
- South Asian
- East Asian
- American

My genome has substantial West African ancestry which links to Sickle Cell Disease.

Race

✗ **Not visible in a genome** (though inferences can sometimes be made)

A context-specific identity term, now widely agreed to be a socio-political construct.

Commonly used racial categories in the UK include:

- White
- Brown
- Latino/Latino and Arab
- Black
- South East Asian

I would call myself mixed race (Black and White), but I am often assumed to be White by others because of my lighter skin tone.

Ethnicity

✗ **Not visible in a genome**

Refers to a person's specific cultural, religious and social background.

Commonly used ethnic categories in the UK include:

- White British*
- Indian*
- Syrian
- Nigerian Yoruba
- Taiwanese

*note that only two of these are options on the current UK census

My ethnicity is Jamaican British, because I feel like this best captures the cultural activities and groups that I align with.

A Kitchen Conversation: Creating awareness with East Asians about genomics

THE G WORD - 2021-11-24

0:00 / 17:51

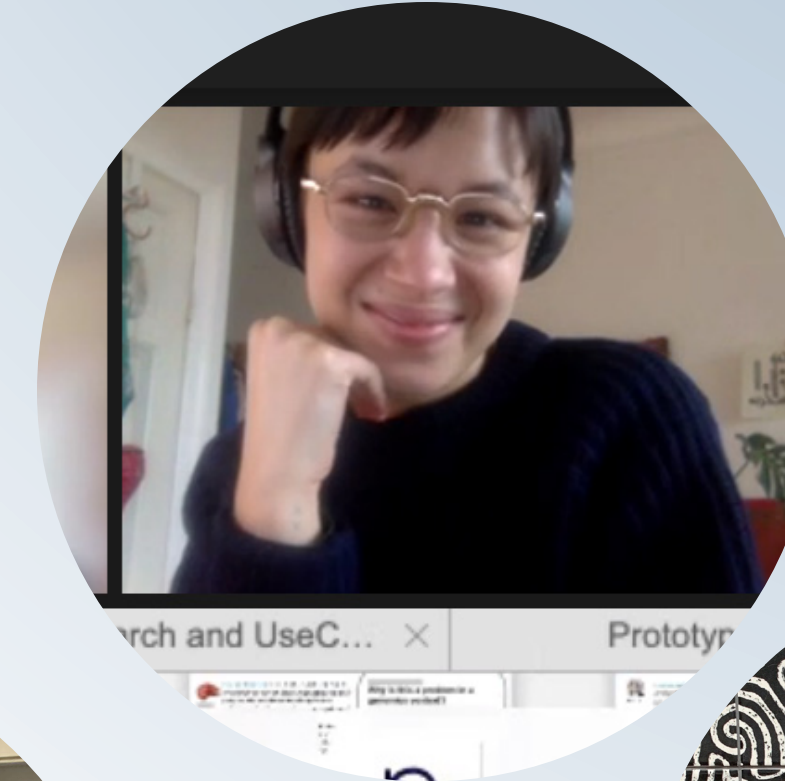
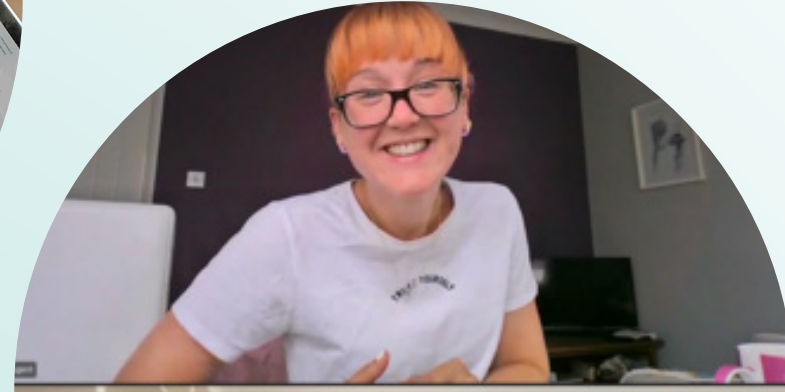
Genomics England

THE G WORD

*[Sophia at Genomics England] is trying to design new ways to get people of different backgrounds to understand what the database is, donate, and ensure that the data gets used. Joining us on today's episode of The G Word is Sophia Luu's family, interviewed by Sophia Luu herself, our Design Researcher on the Diverse Data Programme. She is joined by her grandfather, her grandmother, her father and her mother to discuss why not many people of East Asian descent are in our database.



Co-Creation Sessions ...



Black

The social implications and legacy of race in genomics



Black

The social implications and legacy of race in genomics



Consanguinity

Biased cultural associations

The screenshot shows a Google search for "consanguinity" with the following results:

- Top Left:** A video thumbnail titled "Cousin Marriage - Consanguinity - YouTube" with the text "First cousin parents trebled their risk of having a child with an autosomal recessive disorder." Source: youtube.com.
- Top Middle:** A table titled "Consanguinity and Relationship, Jewish ..." from bakerblumfamilytree.com.
- Top Right:** A book cover titled "SYSTEMS OF CONSANGUINITY AND AFFINITY OF THE HUMAN FAMILY" by Lewis H. Mumm. Source: biodiversitylibrary.org.
- Bottom Left:** A diagram titled "NEPOTISM: How to Count the Level of Degree of Consanguinity/Affinity?" showing a hierarchy of family members. Source: gabotaf.com.
- Bottom Middle-Left:** A grid of 24 small portraits of diverse people. Source: genassistabcs.com.
- Bottom Middle-Right:** Three pie charts for Kuwait, Oman, and Pakistan, showing the distribution of consanguinity types. Source: medcraveonline.com.
- Bottom Right:** A flowchart titled "Consanguinity and the coefficient of inbreeding" showing the breakdown of consanguinity into various types and their percentages. Source: wordpress.mrreid.org.

Consanguinity

Biased cultural associations

The screenshot shows a Google search for 'consanguinity' with the following results:

- Top Result:** A video thumbnail titled 'Cousin Marriage - Consanguinity - YouTube' with the text 'First cousin parents trebled their risk of having a child with an autosomal recessive disorder.'
- Second Result:** A table titled 'Consanguinity and Relationship, Jewish ...' from bakerblumfamilytree.com.
- Third Result:** A book cover titled 'SYSTEMS OF CONSANGUINITY AND AFFINITY IN THE HUMAN FAMILY' by Lewis H. Mumm.
- Fourth Result:** A diagram titled 'NEPOTISM: How to Count the Level of Degree of Consanguinity/Affinity?' from gabotaf.com, showing a hierarchy of family members.
- Fifth Result:** A grid of diverse human faces titled 'Consanguinity | GENASSIST' from genassistabcs.com.
- Sixth Result:** Pie charts for 'Kuwait', 'Oman', and 'Pakistan' from MedCrave online, showing the distribution of consanguinity types.
- Seventh Result:** A family tree diagram titled 'Consanguinity and the coeffi...' from wordpress.mrreid.org, showing percentages for various relationships like Great-Grandparents (12.5%), Consanguinity (25%), etc.

Super Populations

When genomics is taken out of context

What do the population codes for human allele frequencies mean?

In views like [Population genetics](#) in the variation tab, you may find three letter codes for populations.

These come from the [HapMap project](#), and/or the [1000 Genomes project](#).

The following table describes the population codes, and shows which populations are grouped into super populations.

Population Code	Description	Super Population Code
CHB	Han Chinese in Beijing, China	EAS
JPT	Japanese in Tokyo, Japan	EAS
CHS	Southern Han Chinese	EAS
CDX	Chinese Dai in Xishuanagbanna, China	EAS
KHV	Kinh in Ho Chi Minh City, Vietnam	EAS
CEU	Utah Residents (CEPH) with Northern and Western European ancestry	EUR

Super Populations

When genomics is taken out of context

What do the population codes for human allele frequencies mean?

In views like [Population genetics](#) in the variation tab, you may find three letter codes for populations.

These come from the [HapMap project](#), and/or the [1000 Genomes project](#).

The following table describes the population codes, and shows which populations are grouped into super populations.

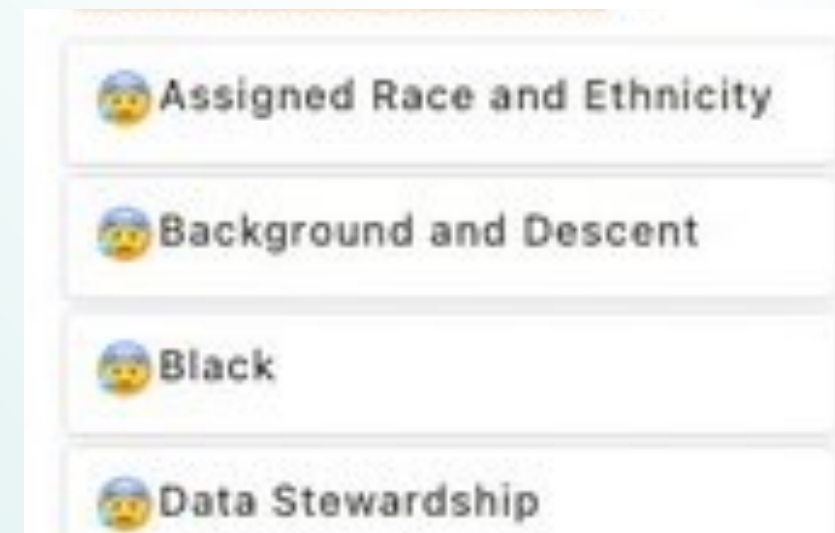
Population Code	Description	Super Population Code
CHB	Han Chinese in Beijing, China	EAS
JPT	Japanese in Tokyo, Japan	EAS
CHS	Southern Han Chinese	EAS
CDX	Chinese Dai in Xishuanagbanna, China	EAS
KHV	Kinh in Ho Chi Minh City, Vietnam	EAS
CEU	Utah Residents (CEPH) with Northern and Western European ancestry	EUR

It's no good if you can't use it!

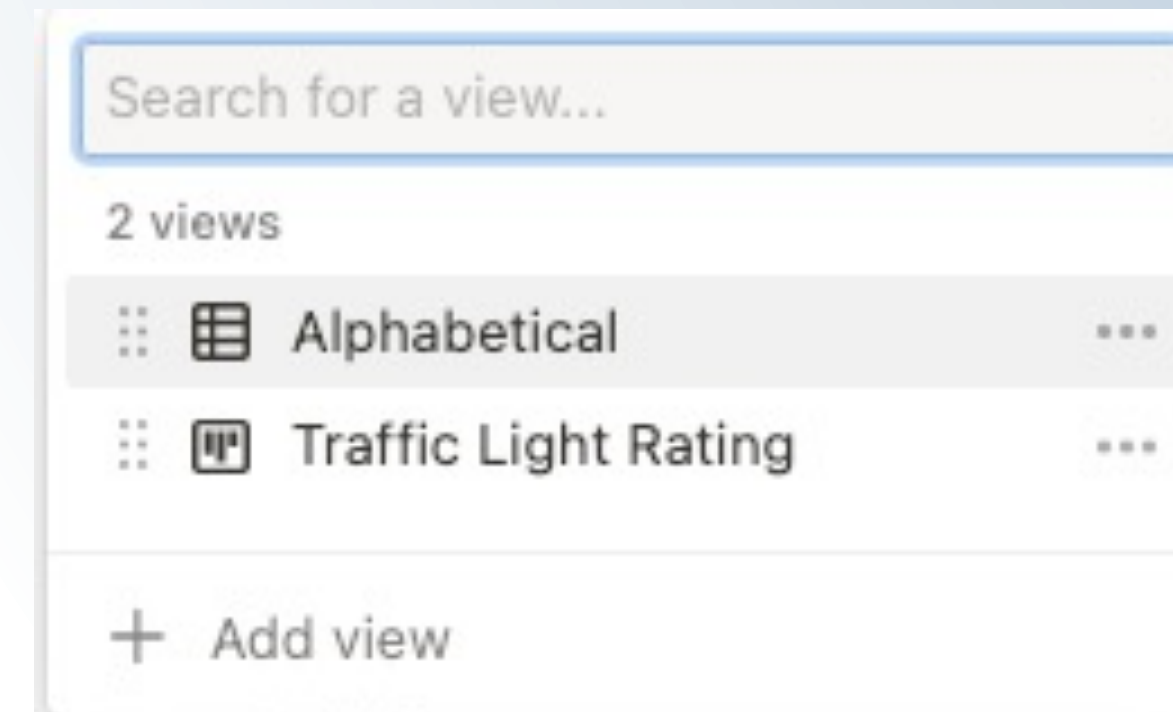
Accessibility

- ▶ Use in other contexts
- ▶ What is the debate around this word?
- ▶ Related words
- ▶ Learn more

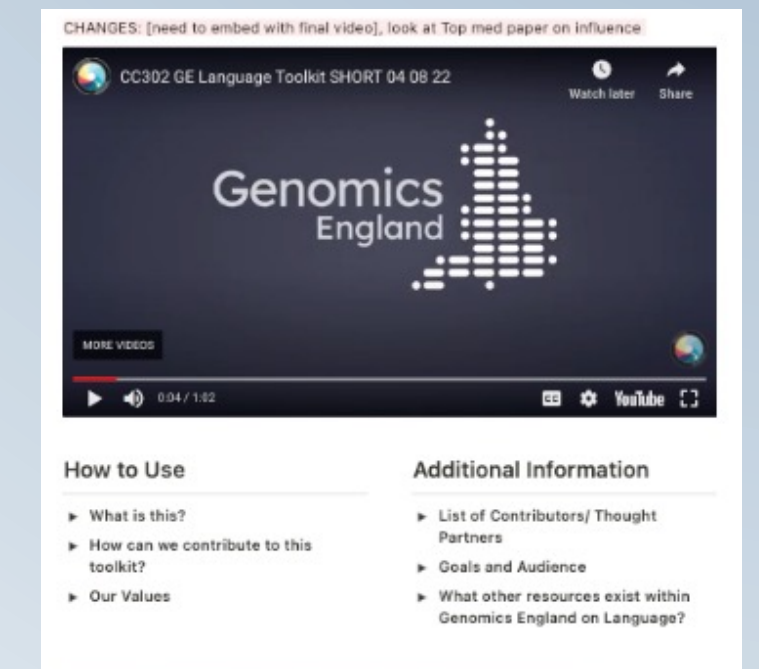
Removing Toggles



Changing visual language - less 'policing'



Adding new filters and views



Removing Homepage Clutter

We are FAR from finished

Inciting long term change



Thank you to our collaborators

Co-Authors

- Semine Long-Callesen, History of Science Researcher
- Dr Sam Tallman - Genomic Data Scientist
- Dr Ed Hollox - Genomic Data Scientist
- Mahantesh Biradar, M.Tech - Genetic Epidemiology and Research Fellow
- Dr. Anna Need - Human Geneticist
- Amanda Pichini - Genetic Counsellor
- Lyra Nightingale - Ethics Lead
- Aman Ali - Community Organiser, Muslim Engagement and Development
- Dr Furaha Asani - Public Scholar and Writer
- Dr Sasha Henriques - Principal Genetic Counsellor
- Salim and Hasha Shaikh, Summaiya - Research & Projects Directors, Smartlyte/ Get Families Talking

Copy, Reviewers and proof-readers

- Dr Alice Popejoy - Reseacher
- Ewan Birney - Director General
- Dr Matteo Fumagalli, Clinician
- Melanie Martin, Copywrytyer
- Dr Daniel Rhodes - Senior Bioinformatician
- Dr Maxine Mackintosh - Diverse Data programme lead
- Thuy Nguyen - Senior Bioinformatician
- Diksha Srivastava - Policy lead

18 in-depth sessions with:

5 patient advocacy groups

5 researchers (science and social)

5 population geneticists/ bioinformaticians

3 clinicians and genetics counsellors



Thank you

If you would like to learn more and contribute to the toolkit, contact

sophia.luu@genomicsengland.co.uk

For more information about the Diverse Data programme:

diversedata@genomicsengland.co.uk

