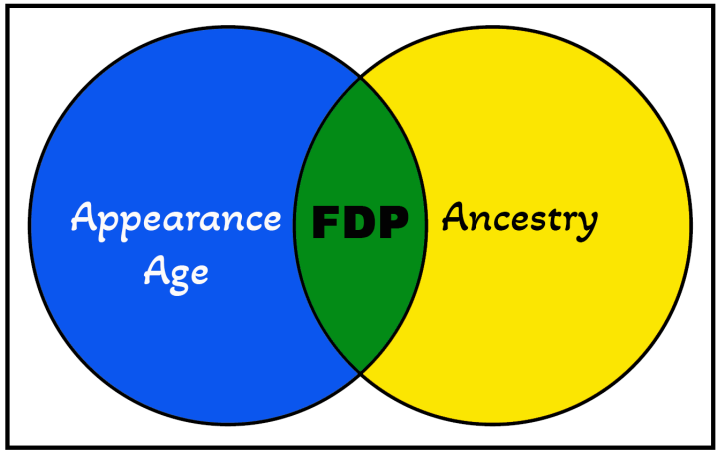
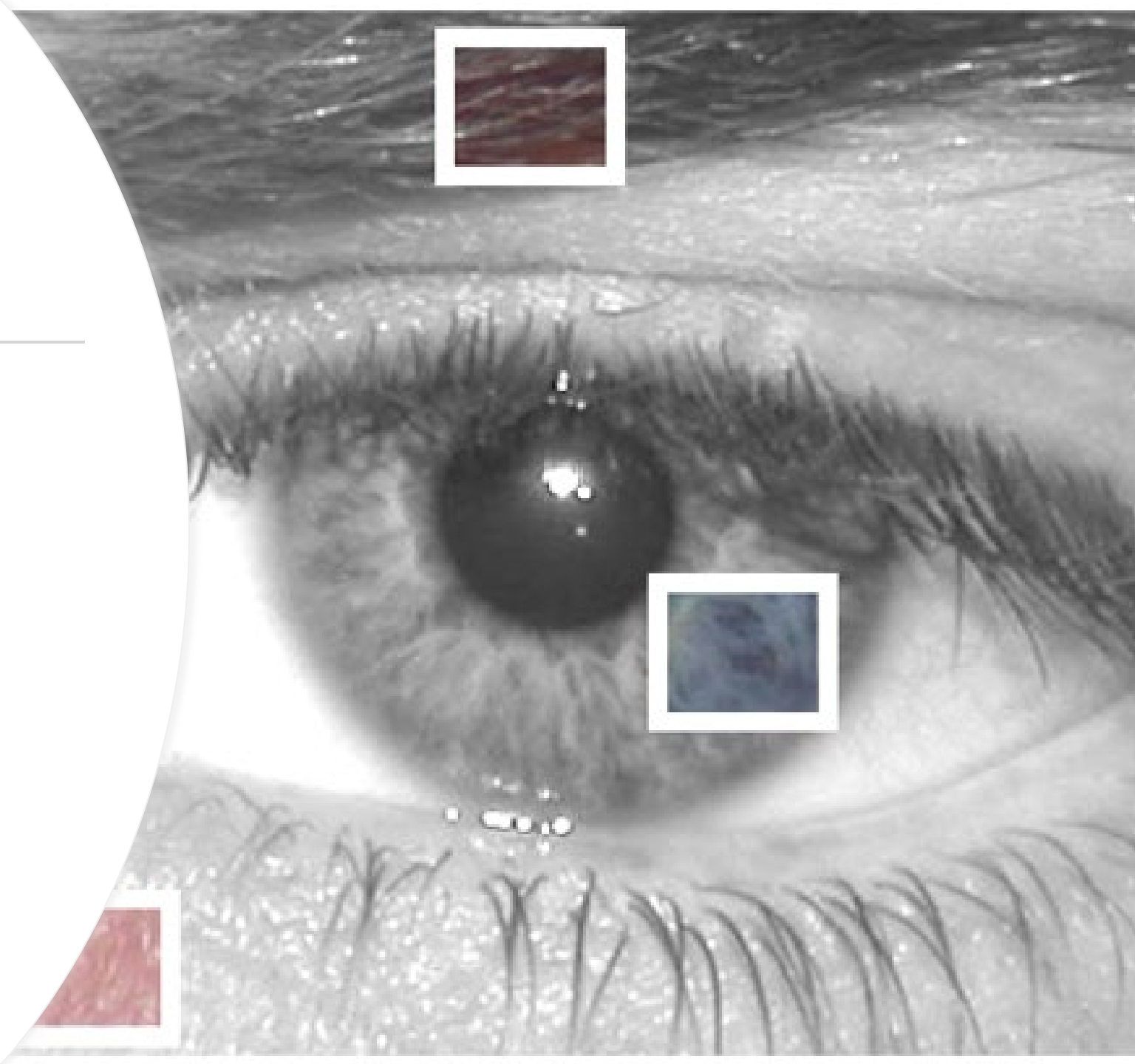
Dr. Susan Walsh
Indiana University Indianapolis

Forensic DNA Phenotyping (FDP)

We must first define the term as used by

- Scientists in the field
- the commercial sector
- the public (due to media)



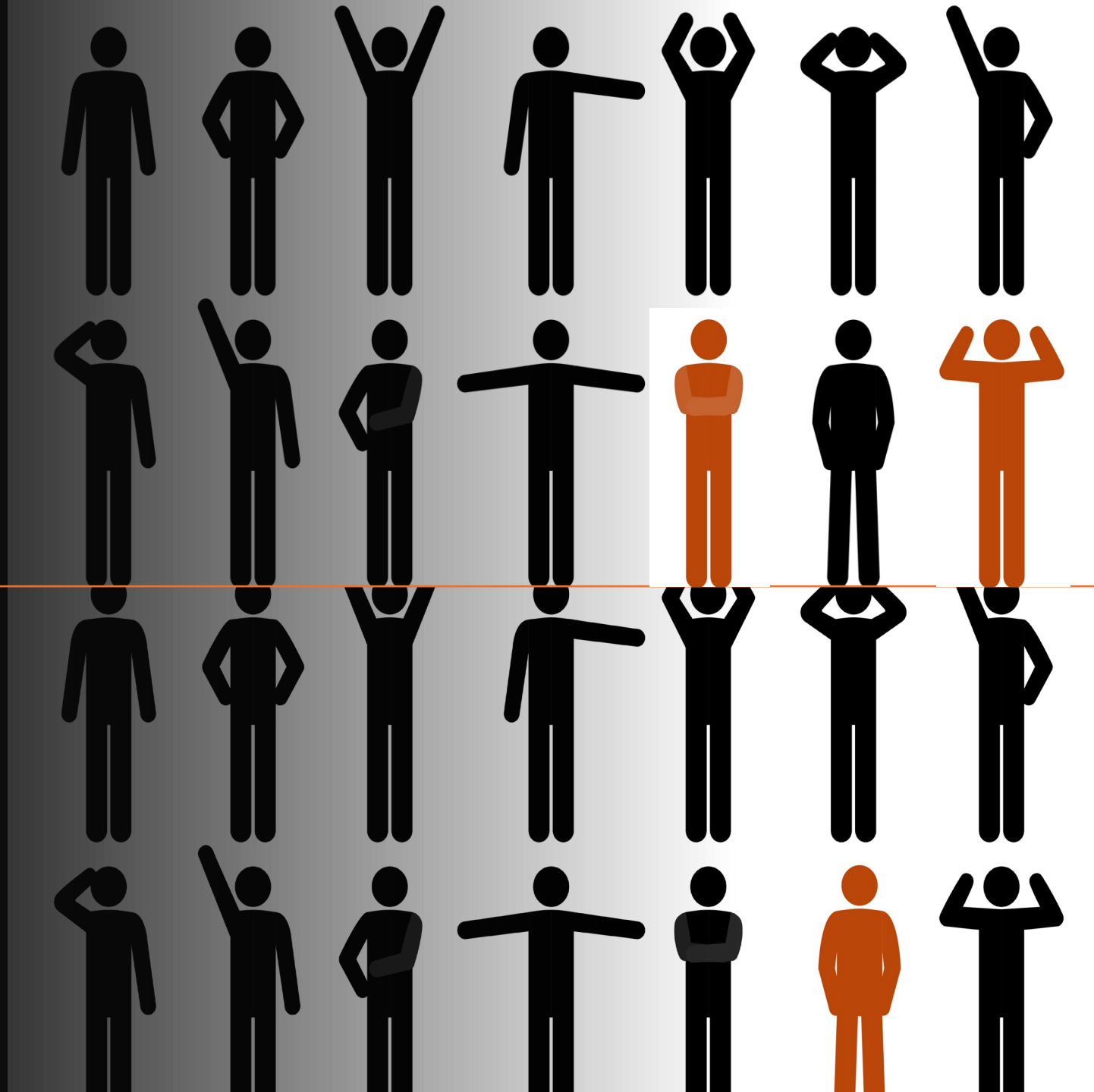


Appearance Traits

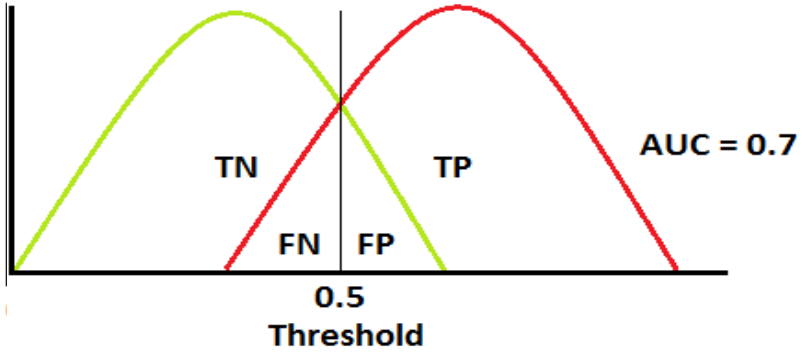
- Each human trait is independently researched
- Traits defined as categorical or continuous
- Variants/Genes are found associated with the trait

- Prediction models are built and tested
- Method, Model and Performance Metrics are peer reviewed & published

IDEA: Accumulation of *intelligence* on the combined traits can narrow down the list of suspects/missing person



Appearance & Age Prediction



- **Pigmentation:** **AUC performance metric**
  Eye – Blue (0.94), Brown (0.95), Intermediate (0.74)
  Hair – Blond (0.8), Brown (0.72), Black (0.83), Red (0.92)
  Skin – Very Pale (0.74), Pale (0.72), Intermediate (0.73), Dark (0.88), Dark-to-Black (0.96)

Independent test set of **194 European individuals**: eye color 80% accurate, hair color 77% accurate, skin color 80% accurate using online webtool – hirisplex.erasmusmc.nl

  Eyebrow color – Blond (0.7), Brown (0.62), Black (0.68)**
  Freckling – Presence/Absence (0.75)**

- **Hair Shape:**
  Straight Hair: Yes/No (0.68)**
  Male Pattern Baldness: Yes/No (0.69)**

- **Body Height:** >195 cm men, >180 cm women
  Tall/Non Tall (0.79) – *no model available*.

- **Age:** **MAE performance metric**
  range due to tissue type MAE 3.2 – 5.1yrs
  webtool available at mathgene.usc.es/snipper/

$$\text{MAE} = \frac{|(y_i - y_p)|}{n}$$

y_i = actual value
y_p = predicted value
n = number of observations/rows

** European VISAGE software tool only

Ancestry Prediction

where genetic data is used to <u>estimate</u> the geographic origins of a persons <u>recent</u> ancestors

- Can be performed using large scale genetic data or select ancestry informative markers (AIMs)

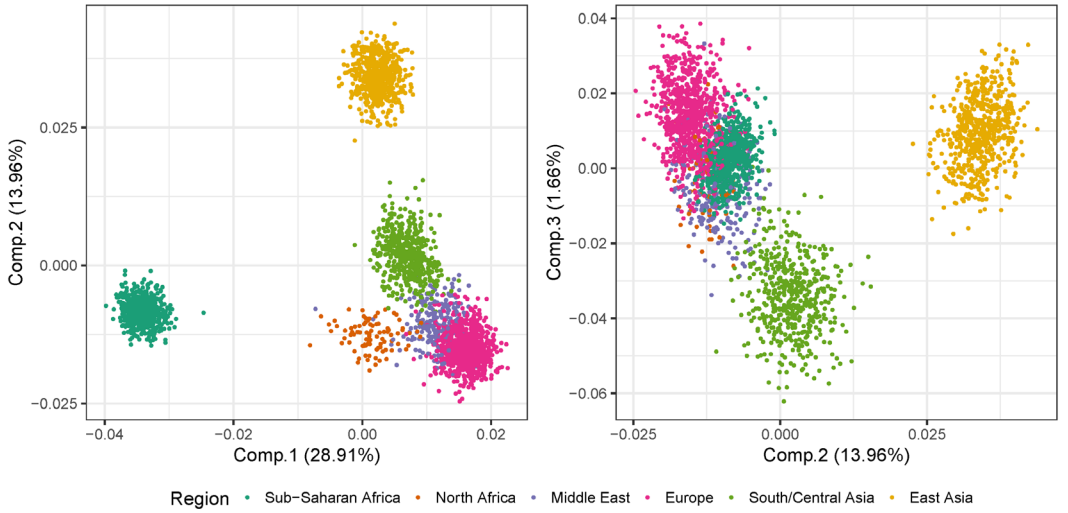
  (webtool available at mathgene.usc.es/snipper/)

  - Likelihood Ratio (LR) of closest member population provided and/or proportions

  - Admixed samples difficult to elucidate

  - Prediction is only as good as the reference populations – anchors

  - Can be complemented by maternal/paternal ancestry inference using Mito and Y markers



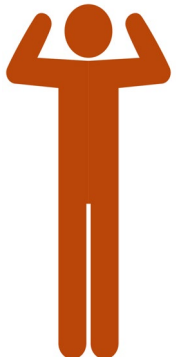
Figures taken from Review of the Forensic Applicability of Biostatistical Methods for Inferring Ancestry from Autosomal Genetic Markers

<u>OF NOTE</u>: It is **<u>not</u>** the role of ancestry prediction to predict a phenotype and vice versa. They are and should be treated as independent tests that complement each other as intelligence, not substitute.

The Scientists and FDP

Intelligence versus Identity – Why?

- Predictions are currently group-based – blue eye color, age range, continental ancestry etc.

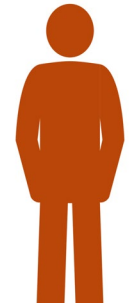


- Appearance Traits are Polygenic – multiple genes needed to produce trait

- Appearance Prediction uses SNPs – single nucleotide polymorphisms
    - Trait expressed goes beyond the genotype i.e. environmental
    - Also influenced by epigenetics, epistasis

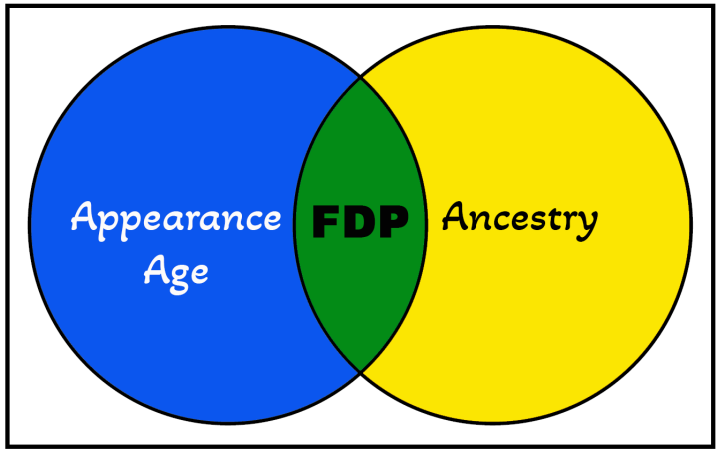
- Unlike the use of STRs – Short Tandem Repeats,
  Using SNPs in the context of FDP is **NOT** comparable profiling, it is **NOT** individual specific

FUTURE: Yes the accumulation of traits, including aspects of facial morphology prediction in the future, in addition to ancestry information may highlight a singular individual, but it more than likely will highlight several, this depends on the case.



Its intelligence role is to shift the priority list for police investigation and questioning

Prediction Result Example using current scientific knowledge



Probability Estimates & Verbal Description – MALE PROFILE (STR KNOWLEDGE)

Blue 0.93 Intermediate 0.02 Brown 0.05
Most probable eye color is Blue

Blond 0.03 Brown 0.31 Black 0.65 Red 0.01; Shade: Dark 0.93
Most probable hair color Dark Brown/Black

Straight Hair Yes 0.8 No 0.2
Straight hair is the most probably prediction

Balding Yes 0.1 No 0.9
Absence of balding is the most probable prediction

Very Pale 0.1 Pale 0.5 Intermediate 0.33 Dark 0.06 Dark-Black 0.01
Most probable skin color Pale to Intermediate

Freckles Yes 0.3 No 0.7
Absence of freckling is the most probable prediction

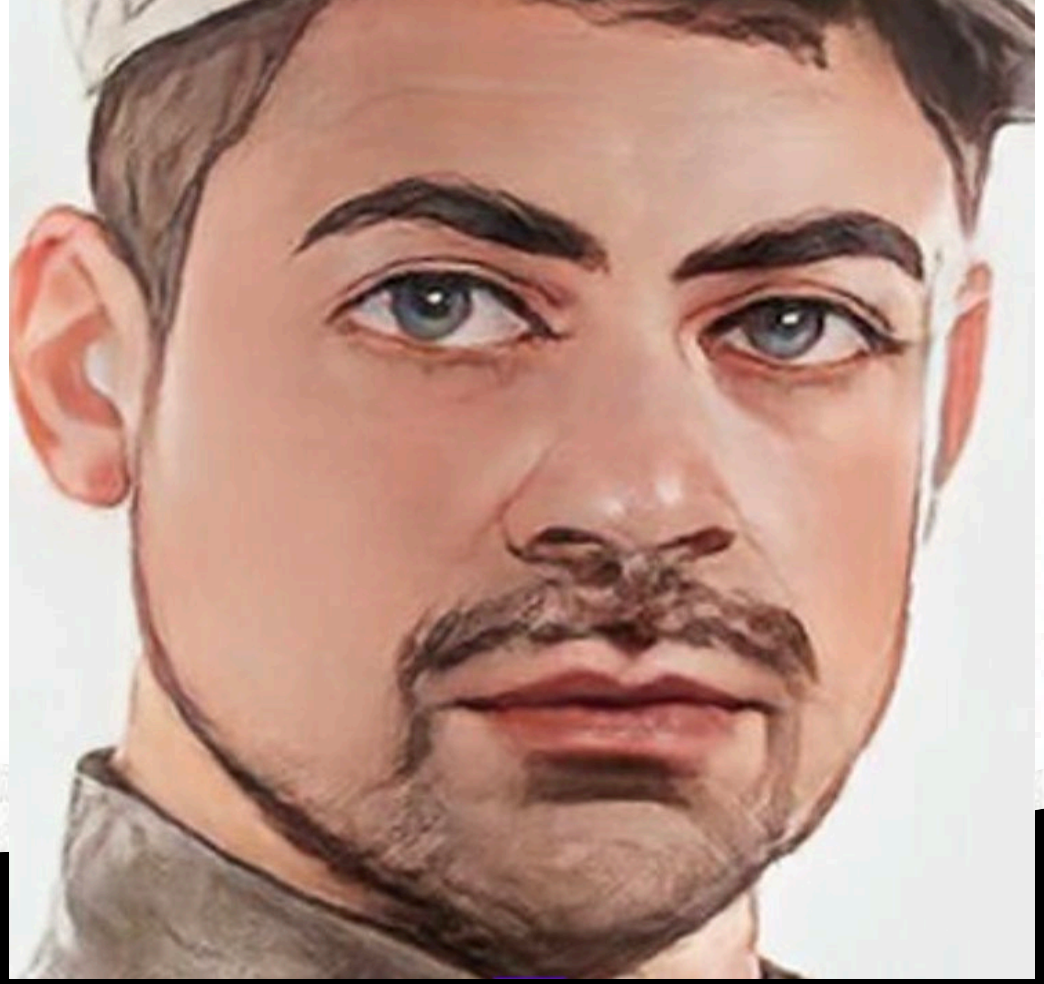
Blond 0.05 Brown 0.4 Black 0.55
Most probable eyebrow color Black to Dark Brown

Age Prediction from blood sample 26±3.2 years
Most probable age range predicted 20-30 years old.

VISUAL INTERPRETATION OF VERBAL
SUBJECTIVE

<u>Ancestry proportions*</u>
European 0.6
North African 0.2
Sub Saharan African 0.01
East Asian 0.12
South/Central Asia 0.05
Middle East 0.02

**Should not reflect the phenotype prediction*



At present, relaying the result is based on probability – interpreting the result is subjective

How can we do this better?

More research needed

provide several visual outputs that are supported by published science and can aid interpretation

NEVER provide a single image based solely on DNA
FDP is not identifying, it is intelligence

The Future of FDP

To allow the **inclusion of predictable traits into the field of FDP**
they need to pass certain *criteria* for true transparency

1. Peer-reviewed publications on

a) the science (genetics, informatics, statistics),

b) design and use of the prediction models (method, training data, provide tool to test)

c) the performance of the predictions on global independent test sets.

2. Publish guidelines on reporting the result, including caveats, to

users, law enforcement, and the public

with training guides available online

3. A standardized set of samples should be made available

to examine both the science of the method, and the impact of result interpretation by law enforcement

This could aid in the transition from benchtop into the legal framework whilst incorporating public perception of both aspects