GENETICS

# INHERITANCE OF FACIAL CHARACTERISTICS

20.11.2023

There is a fascinating story from Barcelona. Maybe I missed the [NYT article] last year but here it is: [Cell Reports 40, 111257, August 23, 2022]

Joshi et al. reported that look-alike pairs identified by facial recognition algorithms share genotypes but not DNA methylomes and microbiomes.

CellPress
OPEN ACCESS

Cell Reports
Report

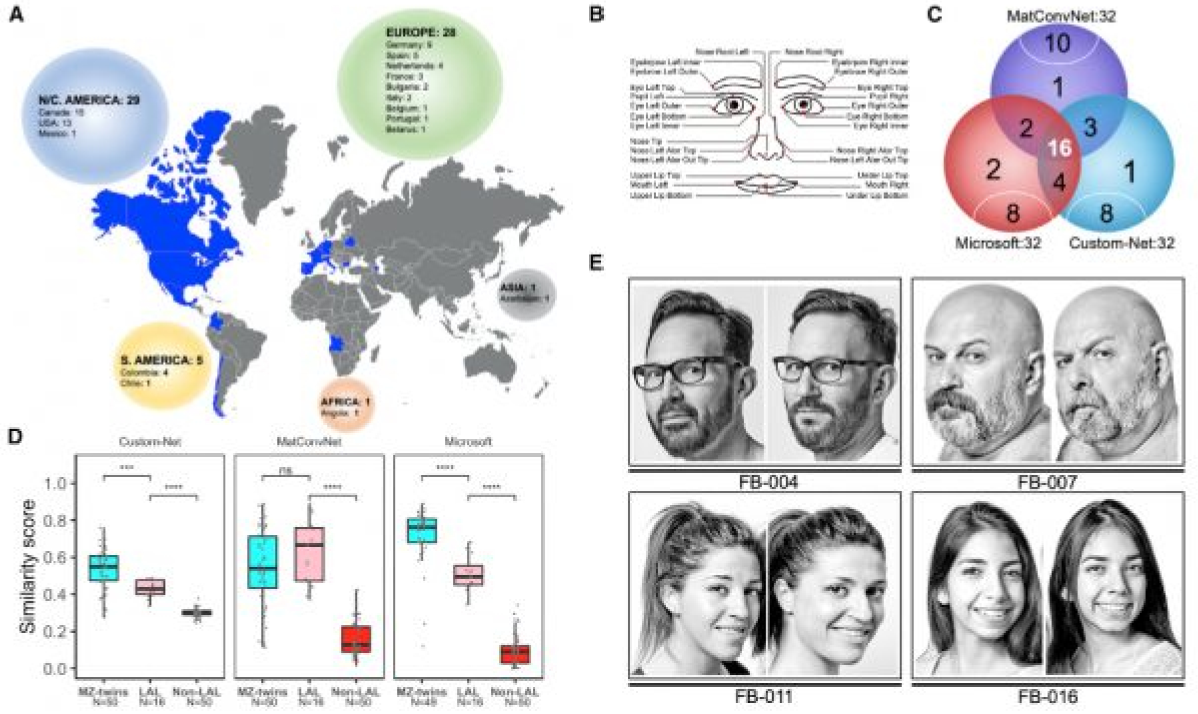


**Figure 1. Recruitment and objective determination of look-alike human pairs**

(A) Representation of the global worldwide distribution of 32 look-alike pairs (n = 64) in this study.

(B) 27 facial parameters by which the Microsoft Oxford Project face API (Microsoft) objectively performs face detection.

(C) Venn diagram showing the number of look-alike pairs discerned and jointly identified in the three facial recognition programs: MatConvNet, Custom-Net, and Microsoft. Numbers within the semi-circle present the pairs that did not cluster in each software.

(D) Boxplots showing unbiased quantitative similarity scores comparing each facial recognition software (MatConvNet, Custom-Net, Microsoft) for monozygotic twins (MZs; blue), look-alike pairs (LALs; rose), and random non-LALs (red). The x axis represents the different cohorts analyzed. The y axis exhibits similarity scores measured between 0 and 1. 1 represents identical facial image, and 0 represents two totally different photographic entities. "N" indicates the number of couples. Differences calculated using two-sided Mann-Whitney-Wilcoxon test: ****$p < 0.0001$; ***$p < 0.001$; ns, non-significant.

(E) Photographic examples of LALs used in this study.

*https://doi.org/10.1016/j.celrep.2022.111257*

Based on an earlier study, the authors collected a set of 32

> individuals, characterized by their high likeliness, are what are called, in lay-language, look-alike humans, unknown twins, twin strangers, doubles, or Doppelgänger in German [I corrected the German language].

collected by a Canadian artist.

My first  expectation is of course that there will be no major effect neither by the methylome nor the microbiome . This seems to be true. So is this a chance event in the genetic lottery  or are there cryptic relatedness effects as region of origin is usually shared?

We have about 4,3 million variants on the Omni5 chip while there are 19.277 shared SNPs (in 3.730 genes, the whole data are at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142304 and the analysis at https://github.com/mesteller-bioinfolab/lookalike). 19.277/4.300.000=0,45% which can be looked up in an earlier PLOS paper.

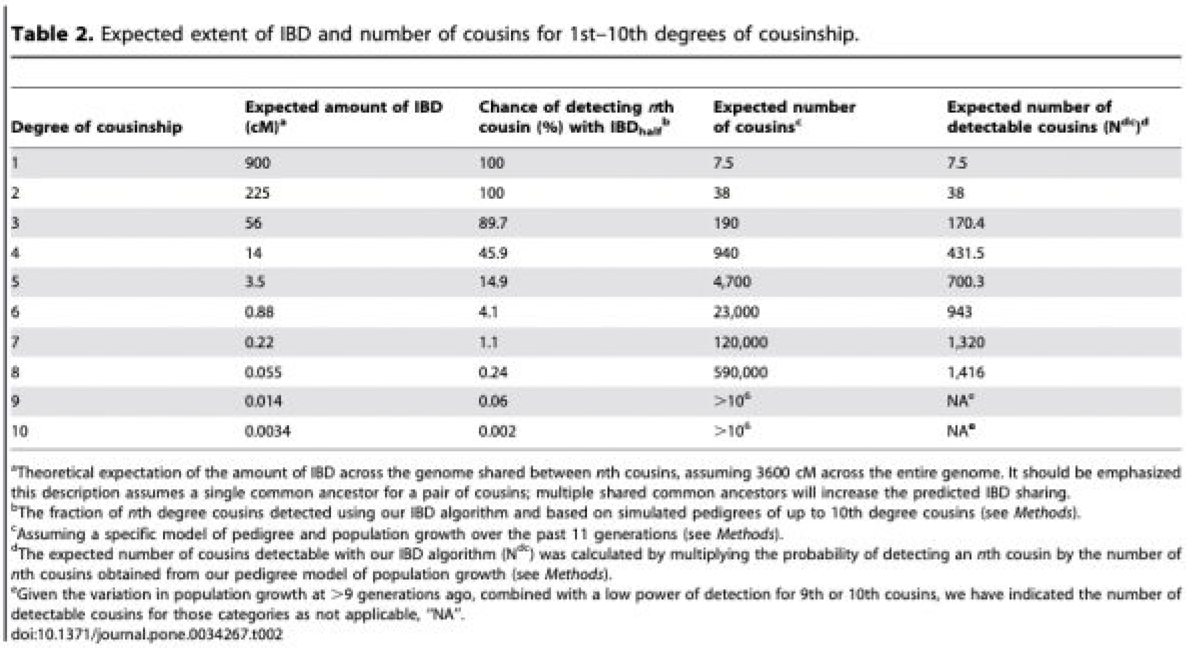


**Table 2.** Expected extent of IBD and number of cousins for 1st–10th degrees of cousinship.

| Degree of cousinship | Expected amount of IBD (cM)[a] | Chance of detecting nth cousin (%) with $IBD_{half}$[b] | Expected number of cousins[c] | Expected number of detectable cousins ($N^{dc}$)[d] |
|---|---|---|---|---|
| 1 | 900 | 100 | 7.5 | 7.5 |
| 2 | 225 | 100 | 38 | 38 |
| 3 | 56 | 89.7 | 190 | 170.4 |
| 4 | 14 | 45.9 | 940 | 431.5 |
| 5 | 3.5 | 14.9 | 4,700 | 700.3 |
| 6 | 0.88 | 4.1 | 23,000 | 943 |
| 7 | 0.22 | 1.1 | 120,000 | 1,320 |
| 8 | 0.055 | 0.24 | 590,000 | 1,416 |
| 9 | 0.014 | 0.06 | $>10^6$ | NA[e] |
| 10 | 0.0034 | 0.002 | $>10^6$ | NA[e] |

[a] Theoretical expectation of the amount of IBD across the genome shared between *n*th cousins, assuming 3600 cM across the entire genome. It should be emphasized this description assumes a single common ancestor for a pair of cousins; multiple shared common ancestors will increase the predicted IBD sharing.

[b] The fraction of *n*th degree cousins detected using our IBD algorithm and based on simulated pedigrees of up to 10th degree cousins (see *Methods*).

[c] Assuming a specific model of pedigree and population growth over the past 11 generations (see *Methods*).

[d] The expected number of cousins detectable with our IBD algorithm ($N^{dc}$) was calculated by multiplying the probability of detecting an *n*th cousin by the number of *n*th cousins obtained from our pedigree model of population growth (see *Methods*).

[e] Given the variation in population growth at >9 generations ago, combined with a low power of detection for 9th or 10th cousins, we have indicated the number of detectable cousins for those categories as not applicable, "NA".

doi:10.1371/journal.pone.0034267.t002

*http://www.doi.org/10.1371/journal.pone.0034267*

14cM/3600cM=0,39% in this table is next to 0,45% so they are cousins of 5th degree – pointing towards cryptic relatedness. The kinship score in Fig 2C does not provide any generation count as  the whole paper is more a bioinformatics approach than a genetic epidemiology study (the supplement says 3rd degree which is unrelaible). I would prefer the construction of some pedigrees (or at least using a typical sign like the prognatism) although that might be a lot of work.

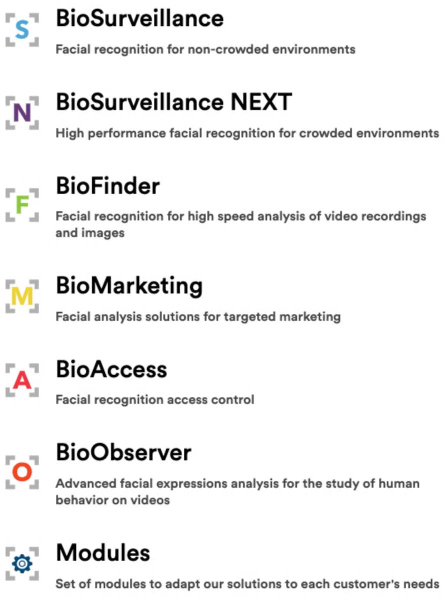
> We observed that almost all the look-alike pairs cluster close to each other according to their countries of origin (or self-attributed ethnic background) (Figure S3). However, they are not more closely related than other pairs of individuals from the same populations taken at random.

This contradicts not only my estimate above but also Fig 2C. It would be interested to full sequence a predigree and go for shared IBD segments as the crude SNP or gene count ends somewhere in the dust of 158 genes in table S5.

Unfortunately he conclusion of the study is clearly off topic and not justified by any data.

> We suggest that these same determinants correlate with both physical and behavioral attributes that constitute human beings.

Behavior cannot be predicted from some facial characteristics. Sorry but we do not need any new eugenics nor any racial genetics as this sentence destroys an otherwise nice paper. Reading it agains, I found one coauthor from a company selling “facial analysis solutions for targeted marketing”…



**BioSurveillance**
Facial recognition for non-crowded environments

**BioSurveillance NEXT**
High performance facial recognition for crowded environments

**BioFinder**
Facial recognition for high speed analysis of video recordings and images

**BioMarketing**
Facial analysis solutions for targeted marketing

**BioAccess**
Facial recognition access control

**BioObserver**
Advanced facial expressions analysis for the study of human behavior on videos

**Modules**
Set of modules to adapt our solutions to each customer's needs

*https://hertasecurity.com/*

CC-BY-NC Science Surf accessed 10.01.2026