

H HEURISTICS · RESEARCH REPORT

The Human Variant Hypothesis

*DNA, Creation Structures, and the
Recurrence of Human Form*

Why unrelated people sometimes look alike, and what that reveals about how a shared genetic and developmental architecture generates human form.

Hunter Hughes

October 2026

ABOUT THIS REPORT

The Human Variant Hypothesis

DNA, Creation Structures, and the Recurrence of Human Form

Author	Hunter Hughes
Publication	H Heuristics — Research Report
Date	October 2026
Report type	Conceptual synthesis and research proposal

Abstract

Human appearance shows both substantial diversity and striking recurrence. Across the global population, individuals repeatedly develop similar facial proportions, anatomical features and combinations of visible traits, sometimes producing exceptionally close resemblance between people with no known close family relationship. This report proposes that such recurrence can be investigated through the concept of *creation structures*: the generative rules, biological constraints and developmental processes that channel variation into recognizable forms.

The **Human Variant Hypothesis** argues that human doppelgängers are not anomalous departures from biological individuality but an especially visible expression of structured phenotypic variation. A shared species-specific developmental architecture, polygenic inheritance, gene regulation, evolutionary history and population structure together constrain the space of possible human appearances, allowing similar configurations of traits to arise in different individuals.

Building on this foundation, the report develops a proposed framework of **Generative Phenotypic Recurrence** to classify, measure and explain recurring patterns of human resemblance. It distinguishes morphological similarity from genetic identity, common ancestry and social prominence, treating each as a separate empirical question. The broader aim is to establish whether the recurrence of human form can be understood as a specific biological expression of the general principles through which complex systems generate constrained but variable outcomes.

How to read this report

Each section begins on a new page and can be read on its own. Sources are hyperlinked in the text, mostly through persistent DOI links, and are listed in full in Appendix B. Figures labeled *illustrative* are conceptual simulations created for this report to make an argument visible; they are not measured data. The term *human variant* is a proposed analytical term, not an established biological classification.

Suggested citation

Hughes, H. (2026). *The Human Variant Hypothesis: DNA, Creation Structures, and the Recurrence of Human Form*. H Heuristics Research Report. hheuristics.substack.com

© 2026 Hunter Hughes. All rights reserved.

CONTENTS

Table of Contents

Executive Summary	5
1. Introduction: The Problem of Recurring Faces	8
1.1 The puzzle.....	8
1.2 Why the question matters	8
1.3 What this report is, and is not	9
1.4 Structure of the report.....	9
2. Creation Structures: Generative Rules and Constrained Form	10
2.1 The idea of a creation structure	10
2.2 Four families of creation structures	10
2.3 From creation structures to constrained variation	11
2.4 The conceptual bridge, and its limits.....	12
3. From Genome to Face: The Genetic Architecture of Human Appearance	13
3.1 The scale of human genetic variation.....	13
3.2 Polygenic inheritance and the many-to-one map	13
3.3 Heritability of facial form	14
3.4 Mapping the genetics of the face.....	15
3.5 Gene regulation: where facial variation lives.....	15
3.6 Shared genetic architecture beyond the face	16
3.7 Implications for the hypothesis.....	16
4. Development as a Constraint Engine	17
4.1 From instructions to anatomy	17
4.2 Canalization: the landscape metaphor.....	17
4.3 Modularity and integration	18
4.4 Facilitated variation and evolvability.....	19
4.5 Environment, plasticity and time	20
4.6 Summary: why development produces recurrence	20
5. Population History and the Structure of Human Variation	21
5.1 A young, recently expanded species	21
5.2 Most variation lies within populations.....	21
5.3 Different traits, different patterns	22
5.4 Ancestry-associated resemblance, correctly understood	22
5.5 Implications for the hypothesis.....	23
6. The Human Variant Hypothesis	24
6.1 Formal statement.....	24
6.2 Four components	24
6.3 What makes someone a human variant?.....	25
6.4 The spectrum of resemblance.....	27
6.5 A hierarchy of explanation	27
7. The Doppelgänger Problem: Evidence and Probability.....	29
7.1 Are faces unique?.....	29
7.2 Look-alikes in the laboratory.....	29
7.3 How perception defines a look-alike.....	30

7.4 A probabilistic model of look-alikes	31
7.5 The role of social prominence	31
7.6 Summary	32
8. A Framework of Generative Phenotypic Recurrence	33
8.1 Purpose	33
8.2 The five stages	33
8.3 Core metrics	34
8.4 Decomposing resemblance	35
8.5 Separating four questions	36
8.6 Propositions.....	37
9. Research Design and Methods.....	38
9.1 Design overview	38
9.2 Data sources and sampling.....	38
9.3 Phenotyping protocol.....	39
9.4 Analytical plan	39
9.5 Quality and reproducibility.....	40
10. Falsifiability, Limitations and Alternative Explanations	41
10.1 Falsification criteria	41
10.2 Limitations	41
10.3 Alternative explanations	42
10.4 A note on adaptive storytelling	42
11. Ethics, Governance and the Boundary Against Typology	43
11.1 The historical shadow.....	43
11.2 Five boundaries	43
11.3 Privacy and biometric data.....	44
11.4 Forensic and commercial uses	44
11.5 Language and communication	44
12. Conclusion: The Recurrence of Human Form	45
12.1 Summary of the argument	45
12.2 What has been established, and what is proposed.....	45
12.3 The broader significance	46
Glossary	47
References	49

OVERVIEW

Executive Summary

Why do unrelated people sometimes look alike, and what does that tell us about how human form is generated?

Human appearance has two features that seem to pull against each other. Across more than eight billion people, no two faces are exactly the same. Yet people keep producing similar facial proportions, cranial structures, body types and combinations of visible traits. Sometimes two people with no known family connection look so alike that friends confuse them. This report argues that both observations come from the same source: the structured way human biology turns a limited set of instructions into a very large but bounded range of outcomes.

The report calls these generative rules, biological constraints and developmental processes **creation structures**. The idea draws on a long tradition in biology, from D'Arcy Thompson's [On Growth and Form](#) to Alan Turing's [chemical theory of morphogenesis](#). In that tradition, form is not an arbitrary choice from all imaginable shapes. Physics, mathematics and development channel it into recurring patterns. The report applies that idea to the human face and body. It does not claim that DNA follows the same mechanisms as crystal growth or leaf arrangement. It claims that development is one more area where generative rules produce constrained, recurring forms, and that human resemblance is a case worth studying within that broader frame.

THE HUMAN VARIANT HYPOTHESIS

Human doppelgängers and recurring appearance patterns arise because a shared genetic and developmental architecture structures human physical variation. Individuals carry different combinations of genetic variants and have different life histories, but the processes that build faces and bodies limit the range of viable outcomes. Similar combinations of visible traits can therefore arise in many individuals, producing recurring phenotypic configurations across the human population.

Core argument

The argument has three linked levels. **Level I** covers creation structures in general: symmetry, scaling, self-organization and generative rules that limit which forms physical systems can reach. **Level II** covers the genetic and developmental machinery that builds a human: DNA variation, gene regulation, the cranial neural crest, canalization, modular growth and environmental input. **Level III** is the observable result: recurring configurations of human appearance, from loose family resemblance to striking look-alikes between strangers.

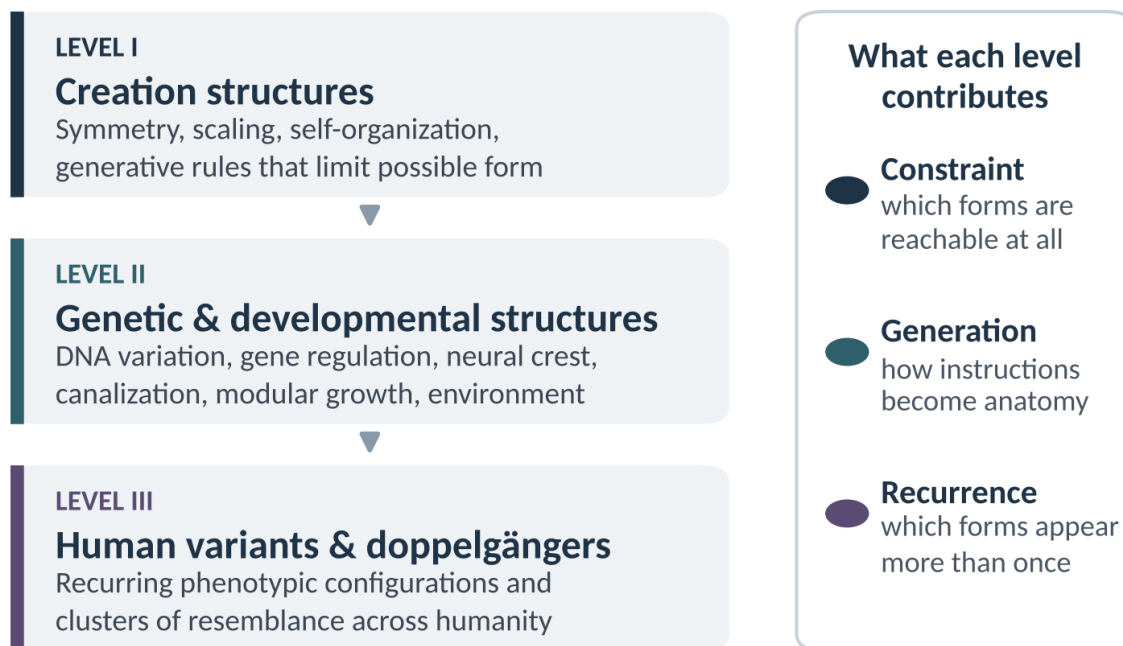


Figure 1. The three levels of explanation used in this report. Level I describes general principles of recurring form, Level II describes the biological machinery that generates human anatomy, and Level III describes the observable recurrence of similar appearances. Arrows show the direction of explanation, not physical causation between levels.

Four observations from the published literature support the argument:

- **Facial form is highly polygenic.** Genome-wide studies have linked roughly two hundred genomic signals to normal-range facial shape ([White et al., 2021](#)). Each signal has a small effect, and they act at both global and local scales ([Claes et al., 2018](#)). Under polygenic inheritance, very different genotypes can produce nearly identical trait values.
- **Development is a constraint engine.** Developmental systems are canalized, modular and integrated ([Waddington, 1942](#); [Klingenberg, 2008](#)). Realized forms therefore occupy a small, structured region of all conceivable combinations of traits.
- **Look-alikes carry a measurable genetic signal.** In a multi-omic study of look-alike pairs verified by facial-recognition software, about half of the most similar pairs clustered together on common genetic variants. Their epigenetic and microbiome profiles were much less alike ([Joshi et al., 2022](#)).
- **Exact facial duplication is still extremely rare.** Measured precisely, the chance that two people match on eight facial dimensions falls below one in a trillion ([Lucas & Henneberg, 2015](#)). Most perceived doppelgängers are matches on a smaller set of perceptually dominant dimensions.

The proposed framework

To move from observation to science, the report proposes a framework of **Generative Phenotypic Recurrence (GPR)**. It treats resemblance as a measurable, decomposable quantity rather than an impression. GPR has five stages: generative inputs, a constrained phenotype manifold, standardized measurement, recurrence mapping and attribution. It introduces a set of working metrics, including a **Recurrence Index** and an **effective dimensionality** measure. It also separates four questions that are usually blurred together:

- **Morphological similarity:** do the two people actually share facial and bodily geometry?
- **Genetic identity:** are they genetically identical, as monozygotic twins are?
- **Common ancestry:** do they share recent or deep ancestry that raises the chance of shared alleles?
- **Social prominence:** is the resemblance noticed mainly because one person is famous or widely photographed?

What the report concludes

The evidence supports a moderate version of the hypothesis. Human resemblance is not spread at random across an unlimited space of appearances. It concentrates where development and inheritance make certain configurations likely. Theory and modeling in this report predict that look-alikes should be most common among people with typical facial configurations, should often be partial (shared within one facial module but not across the whole face), and should depend heavily on how many dimensions the observer, human or algorithm, actually compares.

The report also sets firm boundaries. Human variants are **not** biological types, races or categories of persons. They are recurring regions in a continuous space of variation. Most human genetic variation lies within populations rather than between them ([Lewontin, 1972](#); [Rosenberg et al., 2002](#)). Any study of resemblance must avoid repeating the errors of typological anthropology and physiognomy. Section 11 sets out the governance safeguards these commitments require.

Table 1. Summary of principal findings and confidence levels

Finding	Basis	Confidence
Facial form is highly polygenic and largely regulatory	Large 3D-imaging GWAS; enhancer studies in model organisms	High
Development constrains and channels facial variation	Canalization, modularity and integration research	High
Most human genetic variation lies within populations	Classic and genome-wide apportionment studies	High
Unrelated look-alikes share more common variants than random pairs	One multi-omic study of verified look-alike pairs	Moderate: needs replication
Look-alike frequency depends on dimensionality and typicality	Probabilistic modeling; face-space research	Moderate: model-based
Recurrence can be decomposed into distinct causal channels	Proposed GPR framework	Proposed: untested

SECTION 01

Introduction: The Problem of Recurring Faces

Individuality and recurrence are two sides of one biological process.

1.1 The puzzle

Faces are among the most individual things in nature. Humans recognize thousands of them, often after a glance, and social life depends on telling one person from another. Evolutionary geneticists have argued that human faces evolved partly *to be distinctive*. Facial traits are unusually variable and unusually weakly correlated with one another, which is what a signal of individual identity would need ([Sheehan & Nachman, 2014](#)).

Yet anyone who has been told "you look just like someone I know" has met the opposite fact. Faces recur. Strangers in different cities, and sometimes on different continents, share a jawline, a set of the eyes or a whole facial gestalt. Photographic projects, social-media "twin stranger" searches and the steady supply of celebrity look-alikes show that close resemblance between unrelated people is not vanishingly rare. It is common enough to be a recognized cultural phenomenon, the *doppelgänger*.

This report starts from the view that these two facts are not in conflict. Both come from the same generative process. An extremely high-dimensional system, the human genome acting through development, produces outcomes in a space that is vast in detail but heavily structured in its broad features. Whether people look unique or recurrent depends on how closely we look and how many features we compare.

1.2 Why the question matters

The question matters for at least four reasons.

- **Basic biology.** Explaining how a genome of about 3.1 billion base pairs, which differs between unrelated people at only a few million positions ([1000 Genomes Project Consortium, 2015](#)), produces a range of faces that is both diverse and constrained is a central problem of the genotype–phenotype map.
- **Evolution and development.** The way variation is distributed shows how developmental systems bias the raw material of evolution, a theme going back to classic work on [developmental constraints](#).

- **Technology.** Facial-recognition systems, forensic DNA phenotyping and biometric identity checks all quietly assume something about how often faces recur. When look-alikes are more common than assumed, false matches follow ([Grother et al., 2019](#)).
- **Human meaning.** Resemblance affects how people perceive kinship, belonging and identity. A careful scientific account can correct folk beliefs, including the persistent idea that appearance reveals character, which has caused real harm.

1.3 What this report is, and is not

This report is a **research proposal and conceptual synthesis**. It reviews the relevant science, states a hypothesis in testable form, offers a framework for measuring and decomposing resemblance, and outlines a research program. It does not present new empirical data. Figures labeled *illustrative* are conceptual simulations meant to make an argument visible, not reports of measured results.

The term **human variant** is a proposed analytical term, not an established biological classification. It refers to a recurring phenotypic configuration: a combination of observable traits that resembles configurations found in other individuals. It does not refer to a kind of person, a subspecies, or any division of humanity. Section 11 covers why that distinction must be guarded carefully.

A NOTE ON TERMINOLOGY

Creation structures is used here in a scientific and philosophical sense: the generative rules, constraints and processes through which physical and biological systems produce form. The term carries no theological claim. It points to a shared logic across domains in which simple rules, applied repeatedly under constraint, generate complex but recurring patterns.

1.4 Structure of the report

The report has three parts.

- **Foundations (Sections 2–5).** Section 2 introduces creation structures as a general idea. Section 3 traces the genetic architecture of human appearance. Section 4 treats development as a constraint engine. Section 5 places human variation in the context of population history.
- **Hypothesis and evidence (Sections 6–7).** Section 6 states the Human Variant Hypothesis formally and proposes a taxonomy of variant types. Section 7 reviews the direct evidence on doppelgängers and develops a probabilistic account of why they occur.
- **Framework and program (Sections 8–12).** Section 8 presents the Generative Phenotypic Recurrence framework. Section 9 outlines a research design. Section 10 sets out falsification criteria and alternative explanations. Section 11 addresses ethics and governance. Section 12 concludes.

A glossary and a full reference list follow the main text. References are hyperlinked in the text, mostly through persistent DOI links.

SECTION 02

Creation Structures: Generative Rules and Constrained Form

Across nature, simple rules applied under constraint produce a limited set of recurring forms.

2.1 The idea of a creation structure

A *creation structure* in this report is any combination of rules, constraints and processes that turns a small amount of specifying information into a much larger amount of form. The defining feature is **compression**: the instructions are far shorter than a full description of the result. A sunflower head holds hundreds of seeds in an intricate double spiral, but the plant does not store each seed's position. It runs a simple local rule repeatedly, and the spiral emerges.

Compression has a consequence that sits at the center of this report. When outcomes are generated from compact rules rather than specified one by one, outcomes are not equally likely. Some forms are easy to generate and appear again and again. Others are hard or impossible to reach. Creation structures therefore produce **recurrence**: similar outcomes in independent instances, because those instances share the same generative logic.

2.2 Four families of creation structures

Four families of generative principle recur across the physical and biological sciences. Each has some bearing on human form.

Symmetry and its breaking

Symmetry is the cheapest kind of form, because one specification serves two or more parts. Bilateral symmetry, organized along a left–right axis, structures the human body and face. Development builds one side and mirrors the instructions to the other. Departures from perfect symmetry, known as fluctuating asymmetry, are small and are themselves informative about developmental stability. For resemblance, symmetry halves the number of independent facial dimensions. A face is not two independent half-faces.

Self-organization and reaction–diffusion

In 1952, Alan Turing showed that two interacting chemicals diffusing at different rates can spontaneously form stable spatial patterns such as spots, stripes and labyrinths from a nearly uniform

start (Turing, 1952). Reaction–diffusion is now an established framework for biological pattern formation (Kondo & Miura, 2010). In vertebrate limbs, a Turing-type network involving BMP, SOX9 and WNT signaling patterns the digits (Raspopovic et al., 2014). Hox genes tune the wavelength of that pattern and so help set digit number (Sheth et al., 2012). The key lesson is that a small number of parameters controls a small number of qualitatively distinct outcomes.

Iteration and phyllotaxis

The spiral arrangement of leaves, florets and seeds, called phyllotaxis, repeatedly converges on the "golden angle" of about 137.5 degrees. Physical experiments have shown that this arrangement emerges from a simple iterative rule: each new element forms where there is most room, given the elements already present (Douady & Couder, 1992). No genetic program for Fibonacci numbers is needed. The recurrence is a property of the generative process.

Allometry and transformation

D'Arcy Thompson's famous transformation grids showed that the forms of related species can often be mapped onto one another by smooth geometric deformations. Many differences in shape are differences in *relative growth* rather than in basic design (Thompson, 1917). Human faces vary in just this way. Much of the difference between two faces can be described as coordinated stretching, widening or projection of a shared template.

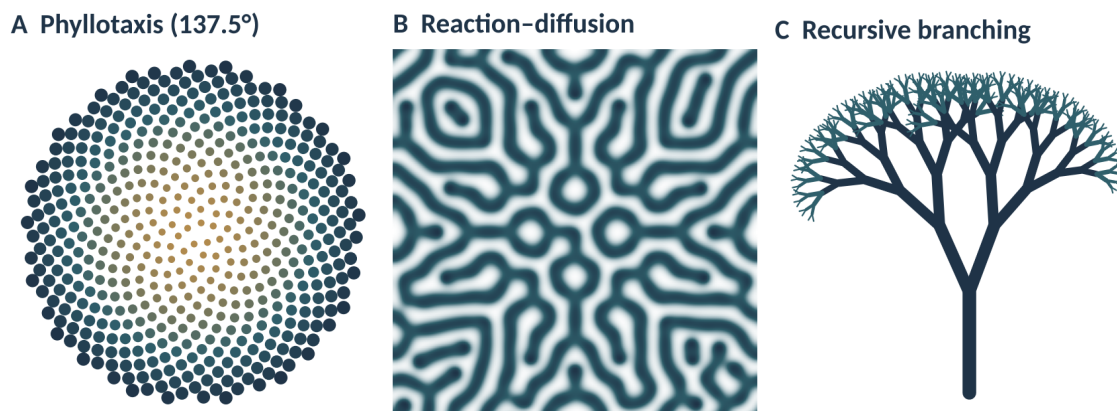


Figure 2. Three classic creation structures, simulated. (A) Phyllotaxis: points placed by repeated rotation through the golden angle self-organize into interlocking spirals. (B) A Gray–Scott reaction–diffusion system: two interacting chemicals settle into a stable labyrinthine pattern from near-uniform starting conditions. (C) Recursive branching: one rule, applied repeatedly with fixed angle and length ratios, produces a complex tree-like form. Each pattern is specified by a handful of parameters.

2.3 From creation structures to constrained variation

The most important property of creation structures for this report is not that they produce beautiful patterns. It is that they **constrain variation**. When generative parameters change slightly, outcomes change in structured ways. Some directions of change are easy, others hard, and some forms cannot be reached at all.

Biologists have long seen this. In a landmark study of developmental abnormalities, Pere Alberch showed that even severe malformations fall into a limited and predictable set of types. That suggests development permits only certain kinds of deviation from the norm ([Alberch, 1989](#)). If even *errors* are constrained, normal variation is likely to be constrained as well.

A widely cited consensus statement defined a developmental constraint as a bias on the production of variant phenotypes, or a limit on phenotypic variability, caused by the structure, character, composition or dynamics of the developmental system ([Maynard Smith et al., 1985](#)). That definition carries the concept this report needs. The space of human appearances is not a uniform cloud. Its shape comes from how humans are built.

2.4 The conceptual bridge, and its limits

The link proposed here is an **analogy of logic, not of mechanism**. Phyllotaxis arises from competition for space at a growing tip. Reaction–diffusion arises from chemical kinetics. Human facial development arises from gene regulatory networks acting on migrating cells, growing cartilage and remodeling bone. These are different mechanisms. What they share is a structure of explanation:

- A **compact generative specification** (a rule, a parameter set, a genome).
- **Iterative or self-organizing execution** (growth, signaling, feedback).
- A **constrained outcome space** in which some forms are much more likely than others.
- **Recurrence** of similar forms in independent instances.

Using this shared structure as a lens leads to sharper questions about human resemblance. How compact is the generative specification of a face? How constrained is the outcome space? Which configurations are favored, and why? The remaining sections take up these questions in turn.

CAUTION AGAINST OVER-REACH

Patterns that look mathematically elegant invite over-interpretation. Not every resemblance reflects a deep generative principle, and not every spiral is Fibonacci. The framework in this report requires every claim of constraint to be tested against an appropriate null model (Section 8). The aim is explanation, not numerology.

SECTION 03

From Genome to Face: The Genetic Architecture of Human Appearance

A face is built by hundreds of small genetic effects acting through shared developmental pathways.

3.1 The scale of human genetic variation

Each human inherits two copies of a genome of about 3.1 billion base pairs. Two unrelated people are identical at the great majority of positions. The international 1000 Genomes Project found that a typical genome differs from the human reference sequence at roughly 4.1 to 5.0 million sites, mostly single-nucleotide polymorphisms (SNPs) and short insertions or deletions ([1000 Genomes Project Consortium, 2015](#)). Each person also carries a few dozen new mutations not present in either parent. In a large Icelandic study the typical count was in the range of roughly 60 to 70, rising with parental age ([Jónsson et al., 2017](#)).

Reproduction mixes this variation on a huge scale. Independent assortment of 23 chromosome pairs alone allows 2^{23} (about 8.4 million) distinct chromosome combinations in each egg or sperm, before recombination shuffles segments within chromosomes. Two parents can therefore produce on the order of 70 trillion chromosome-level combinations, far more than the number of humans who have ever lived. In genetic terms, every person apart from a monozygotic twin is effectively unique.

The question for this report is how this genetic uniqueness produces appearances that are *not* always unique.

3.2 Polygenic inheritance and the many-to-one map

The answer starts with the nature of polygenic traits. More than a century ago, R. A. Fisher showed that the continuous, bell-shaped distributions of traits such as height could be reconciled with Mendelian inheritance if many genes each contribute a small effect ([Fisher, 1918](#)). Later theory and data confirmed that most variation in complex traits is additive ([Hill, Goddard & Visscher, 2008](#)). The "omnigenic" model has argued that, through regulatory networks, a very large share of the genome may affect any given complex trait ([Boyle, Li & Pritchard, 2017](#)).

Polygenic inheritance creates a **many-to-one genotype–phenotype map**. When a trait is the sum of many small contributions, a huge number of different genotypes yield the same or nearly the same value. Middle values are reachable by far more genotype combinations than extreme ones. This is pure combinatorics: there are many more ways to score "average" than "maximum."

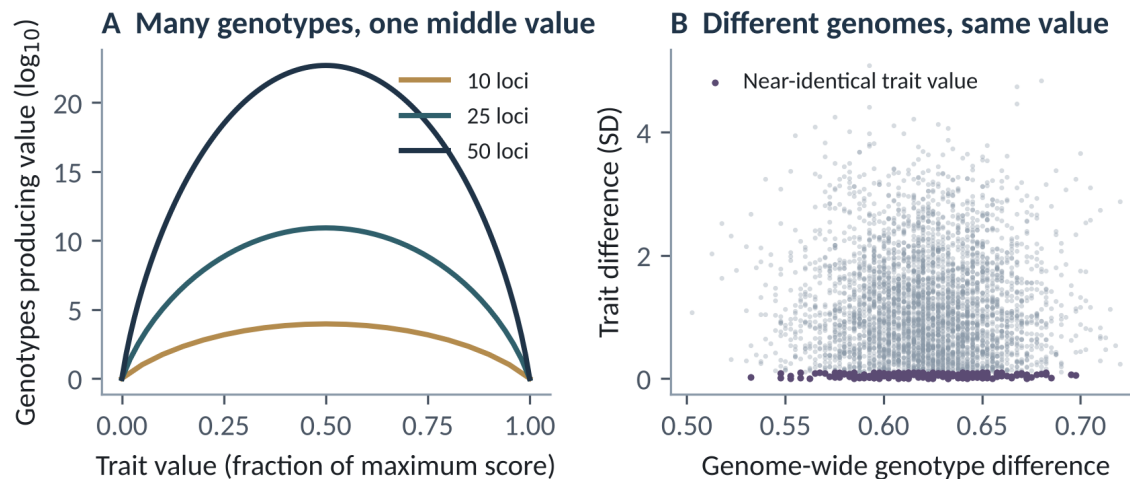


Figure 3. Illustrative simulation of polygenic convergence. (A) For an additive trait controlled by 10, 25 or 50 loci, the number of distinct genotypes producing each trait value (log scale) peaks sharply at intermediate values; with 50 loci, more than 10^{22} genotypes produce the central value. (B) Among 600 simulated individuals with 400 loci, pairs whose trait values are nearly identical (purple) are spread across the whole range of genome-wide genetic difference. Close phenotypic resemblance does not require close genetic similarity.

Two consequences follow, and both are central to the Human Variant Hypothesis:

- **Convergence.** Two people can reach nearly the same value of a polygenic trait from very different genetic starting points. For a single trait, phenotypic resemblance and genome-wide similarity are only loosely coupled.
- **Central concentration.** Intermediate trait values are produced by many genotypes, so populations cluster around them. This is the statistical root of the "typical" face, and Section 7 shows why typical configurations should recur most often.

3.3 Heritability of facial form

Twin and family studies show that much variation in facial size and shape is heritable, although estimates vary by trait, population and method (Cole et al., 2017). A meta-analysis covering nearly three thousand twin studies and more than fourteen million twin pairs found that, across all human traits examined, the average heritability was close to 49 percent (Polderman et al., 2015).

Heritability has a specific and often misunderstood meaning. It is the share of trait *variance* in a given population and environment that is associated with genetic differences. It does not measure how "genetic" a trait is in any absolute sense, and it can change from one environment to another (Visscher, Hill & Wray, 2008). For resemblance, the key point is that facial similarity among relatives is substantially, but not entirely, genetic. Development, environment and chance account for the rest.

3.4 Mapping the genetics of the face

The genetics of normal facial variation has moved quickly. Early genome-wide association studies (GWAS) of facial landmarks found a handful of loci, including variants near *PAX3*, *PRDM16* and *TP63* (Liu et al., 2012). Advances in 3D surface imaging then allowed much richer descriptions of shape.

A major step came from a **global-to-local** approach. Instead of choosing a few measurements in advance, researchers split the 3D facial surface hierarchically, from the whole face down to small regions, and tested genetic association at every level ([Claes et al., 2018](#)). Later work combining large cohorts found roughly two hundred genomic signals associated with normal-range facial variation. Many lie near genes known from craniofacial development and craniofacial syndromes ([White et al., 2021](#)). Studies across ancestry groups have found loci with consistent effects in European and East Asian samples ([Xiong et al., 2019](#)). That suggests a shared core architecture for facial variation within the human species.

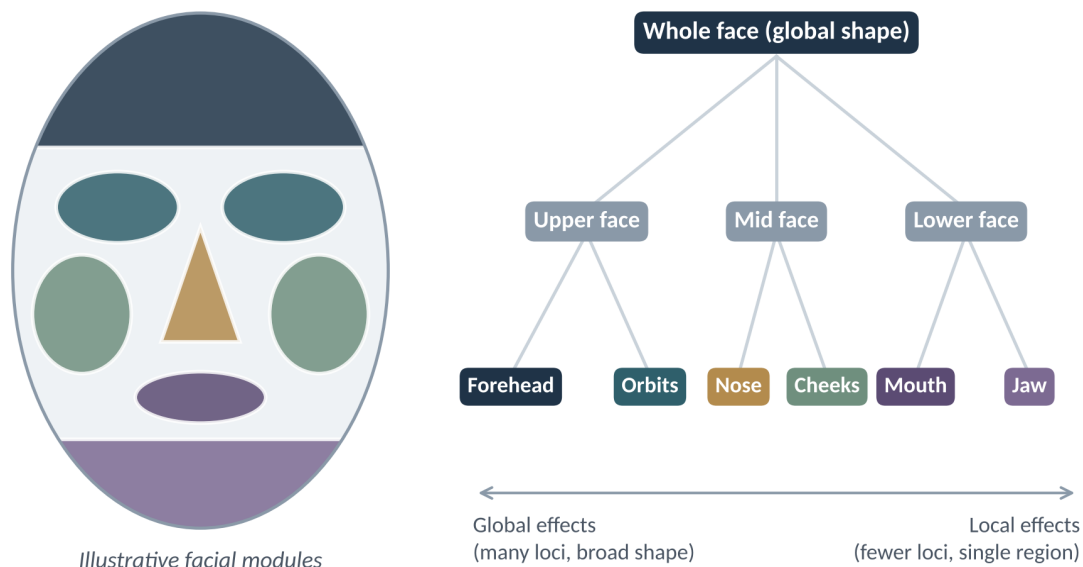


Figure 4. Global-to-local organization of facial genetics (schematic). Genetic effects can be mapped to the whole face, to broad facial zones, or to specific modules such as the nose or jaw. Global segments are influenced by many loci, and local segments often by fewer loci with more focused effects. Module boundaries shown are illustrative.

3.5 Gene regulation: where facial variation lives

Most facial-shape signals lie not in protein-coding sequence but in **regulatory DNA**: enhancers and other elements that control when, where and how strongly genes are expressed. Experimental work in mice identified thousands of candidate craniofacial enhancers. Deleting individual enhancers produced subtle but measurable changes in skull and facial shape without gross malformation ([Attanasio et al., 2013](#)). Comparisons of human and chimpanzee neural crest cells found widespread differences in enhancer activity, consistent with regulatory change as a main driver of species-specific facial form ([Prescott et al., 2015](#)).

This regulatory architecture matters for the hypothesis. Regulatory variants usually adjust the *degree* of a developmental process rather than inventing a new one. They shift the dials of a shared machine. Different people therefore vary along the same set of axes, and that shared set of axes is what makes recurrence possible.

3.6 Shared genetic architecture beyond the face

Facial genetics overlaps with other traits. Many loci influence both face shape and brain shape, reflecting their shared developmental origins in the embryonic head (Naqvi et al., 2021). Other loci affect body proportions, pigmentation and hair together. This **pleiotropy** means that resemblance in one domain can carry partial resemblance in another, which may help explain why some look-alike pairs are similar in more than face shape (Section 7).

Table 2. Key features of the genetic architecture of human appearance

Feature	What the evidence shows	Implication for recurrence
Polygenicity	Roughly two hundred genomic signals linked to normal facial shape, each of small effect	Many genotypes converge on similar faces
Additivity	Most heritable variation in complex traits is additive	Trait values cluster around central configurations
Regulatory basis	Most signals lie in enhancers and other regulatory DNA	Variation shifts shared developmental dials rather than inventing new structures
Hierarchical effects	Loci act on global facial shape and on local modules	Resemblance can be whole-face or modular
Cross-population consistency	Several loci have consistent effects across ancestry groups	Look-alikes are possible across populations
Pleiotropy	Facial loci overlap with brain shape, body and pigmentation loci	Facial resemblance may co-occur with other physical similarity

3.7 Implications for the hypothesis

Section 3 supplies the hypothesis with a genetic engine. Faces are built from many small, mostly additive, mostly regulatory effects acting on a shared developmental program. This architecture predicts that (1) many genotypes converge on similar facial configurations, (2) configurations near the population center are especially reachable, and (3) variation is organized along shared axes rather than scattered in arbitrary directions. All three are preconditions for the recurrence of human form.

SECTION 04

Development as a Constraint Engine

Embryonic development turns genetic variation into form, and filters it along the way.

4.1 From instructions to anatomy

A genome does not contain a picture of a face. It contains instructions that, executed during development, generate one. In the early embryo, the human face forms from a set of tissue swellings called facial prominences, populated largely by **cranial neural crest cells**. These are migratory cells that originate near the developing neural tube and form much of the facial skeleton, cartilage and connective tissue ([Cordero et al., 2011](#)). The prominences grow, meet and fuse in a tightly choreographed sequence under the control of conserved signaling pathways and positional codes, including the Hox system that patterns the vertebrate body axis ([Krumlauf, 1994](#)).

This process is a creation structure in the full sense. Signaling gradients, cell migration, proliferation and mechanical interaction among tissues execute a compact set of instructions iteratively, in feedback with one another. The result is a face that is new in detail but deeply constrained in organization.

4.2 Canalization: the landscape metaphor

In 1942, C. H. Waddington introduced **canalization**: development is buffered so that it reliably reaches the same end-state despite genetic and environmental disturbance ([Waddington, 1942](#)). He pictured development as a ball rolling down a landscape of branching valleys. The shape of the landscape, set by the genome, guides the ball into one of a limited number of channels. Small pushes rarely move the ball out of its valley.

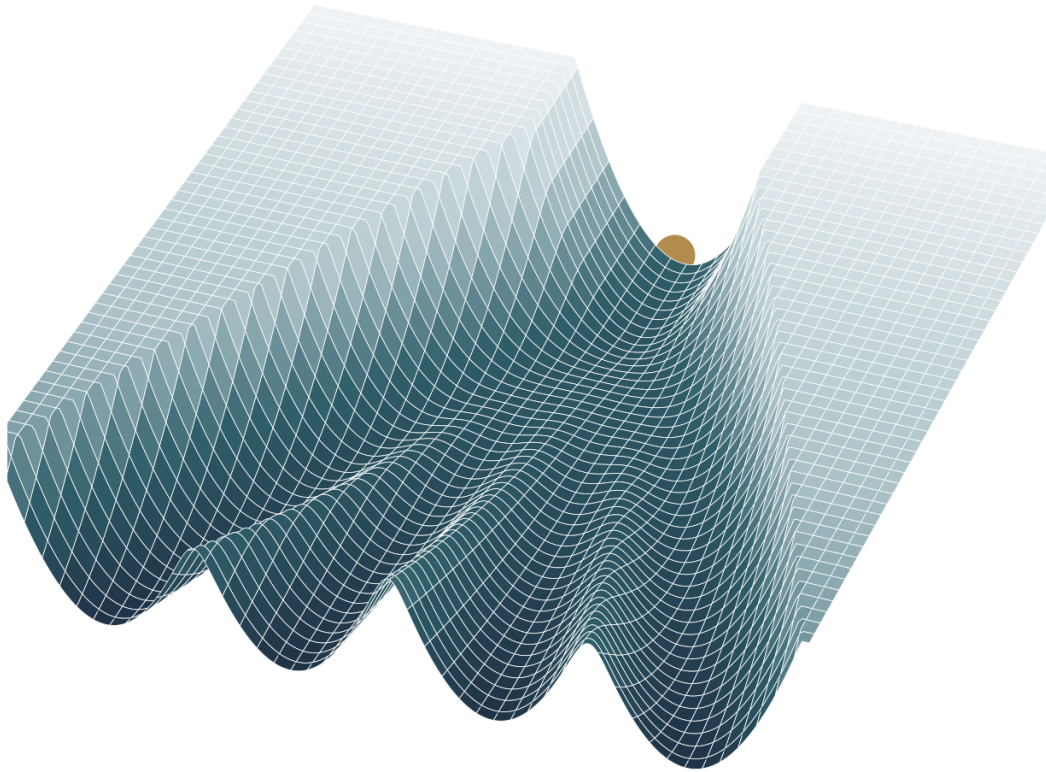


Figure 5. A Waddington-style developmental landscape (illustrative). The developing system (gold sphere) begins at the back, in a single broad valley. As development proceeds toward the front, the landscape branches into a small number of deep valleys, which represent canalized outcomes. Many different starting perturbations end in the same valley, so canalized development produces recurring end-states rather than a continuous spread of possibilities.

Canalization contributes to recurrence in two ways. First, it **suppresses variation**: many small genetic and environmental differences are absorbed, so outcomes are more similar than inputs. Second, where canalization is incomplete, variation is released along **preferred directions**, the valley floors, rather than at random. Both effects concentrate realized forms into a limited part of the possible space.

4.3 Modularity and integration

Development is organized into partly independent **modules**: sets of traits that develop together and covary strongly with one another, but only weakly with traits in other modules ([Wagner & Altenberg, 1996](#); [Klingenberg, 2008](#)). In the skull and face, likely modules include the orbits, the nasal region, the oral region, the mandible and the cranial vault, although the boundaries depend on scale and method. Covariation measured in adults is a layered record of many developmental processes acting at different times, which researchers have called a palimpsest ([Hallgrímsson et al., 2009](#)).

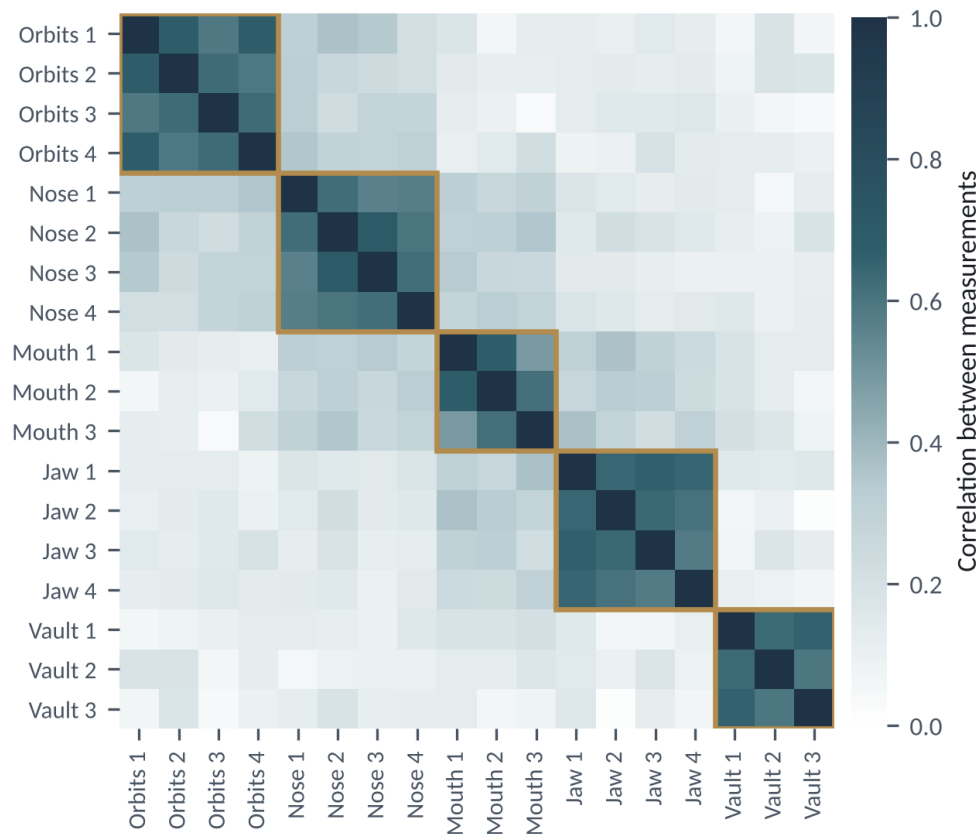


Figure 6. Illustrative correlation structure among 18 facial measurements grouped into five developmental modules (gold outlines). Measurements within a module are strongly correlated; measurements in different modules are weakly correlated, with moderate integration between neighboring modules. Values are simulated to show the concept.

Modularity has a direct consequence for resemblance that has received little attention: **resemblance should often be modular**. Two people may share a mid-face configuration and differ strongly in the jaw, or share orbits and brow while differing in the mouth. This is the familiar experience of someone who "has her eyes" or "has his nose," whether among relatives or strangers. Section 6 treats modular resemblance as a distinct class of human variant.

4.4 Facilitated variation and evolvability

Kirschner and Gerhart argued that conserved core processes, such as signaling pathways, cell behaviors and regulatory linkages, are deeply robust yet *easily regulated* (Kirschner & Gerhart, 1998). Variation is then channeled into changes of degree, timing and location, which tend to yield viable and integrated forms. Development does not merely tolerate variation. It shapes and biases it.

For the face, this means variation is predominantly *coordinated*. A change that lengthens the mid-face also shifts the position of the eyes, nose and mouth together. Realized faces therefore occupy a lower-dimensional space than a count of individual measurements would suggest. Section 7 shows that this **effective dimensionality** is the single most important parameter governing how often look-alikes occur.

4.5 Environment, plasticity and time

Development is not purely genetic. Nutrition, hormones, mechanical loading during chewing, health in early life and many other factors influence craniofacial growth. Even monozygotic twins, who share their entire inherited genome, accumulate epigenetic differences over their lifetimes ([Fraga et al., 2005](#)) and are not perfectly identical in appearance.

The face also changes across the life course. Proportions shift through childhood and adolescence, and soft tissue changes with age, weight and sun exposure. Resemblance is therefore **time-indexed**: two people may look alike at 25 and less so at 60, or the reverse. A rigorous account of human variants must treat age as a coordinate of the phenotype rather than as noise.

4.6 Summary: why development produces recurrence

Development turns genetic diversity into constrained phenotypic diversity through four mechanisms:

- **Buffering (canalization)** reduces the variance of outcomes relative to the variance of inputs.
- **Channeling (facilitated variation)** directs remaining variation along a few preferred axes.
- **Modularity** lets parts of the face vary partly independently, producing partial resemblances.
- **Integration** links traits within modules, so resemblance in one measurement often brings resemblance in others.

Together these mechanisms produce a phenotype space that is structured, anisotropic and much smaller than the space of conceivable forms. That structured space is the stage on which human variants appear.

SECTION 05

Population History and the Structure of Human Variation

Shared ancestry shapes resemblance, but human variation is overwhelmingly continuous and overlapping.

5.1 A young, recently expanded species

All living humans descend from populations in Africa, and genetic diversity is highest among African populations. As humans spread across the world, successive founder events gradually reduced genetic diversity. Heterozygosity falls with geographic distance from East Africa in a pattern consistent with a **serial founder effect** ([Ramachandran et al., 2005](#)). Population history therefore adds another layer of structure to human variation. People who share recent ancestry share more alleles on average, including alleles that influence appearance.

5.2 Most variation lies within populations

The most important single fact about human population structure is that **most human genetic variation is found within populations rather than between them**. In a classic 1972 analysis, Richard Lewontin estimated that about 85 percent of genetic diversity in the protein markers he studied lay within local populations ([Lewontin, 1972](#)). Later genome-wide analyses confirmed the broad pattern: within-population differences among individuals account for roughly 93 to 95 percent of variation at the markers studied ([Rosenberg et al., 2002](#)). The same work also showed that individuals can be assigned to broad ancestry clusters when many markers are combined. Both findings are true at once, and both matter here.

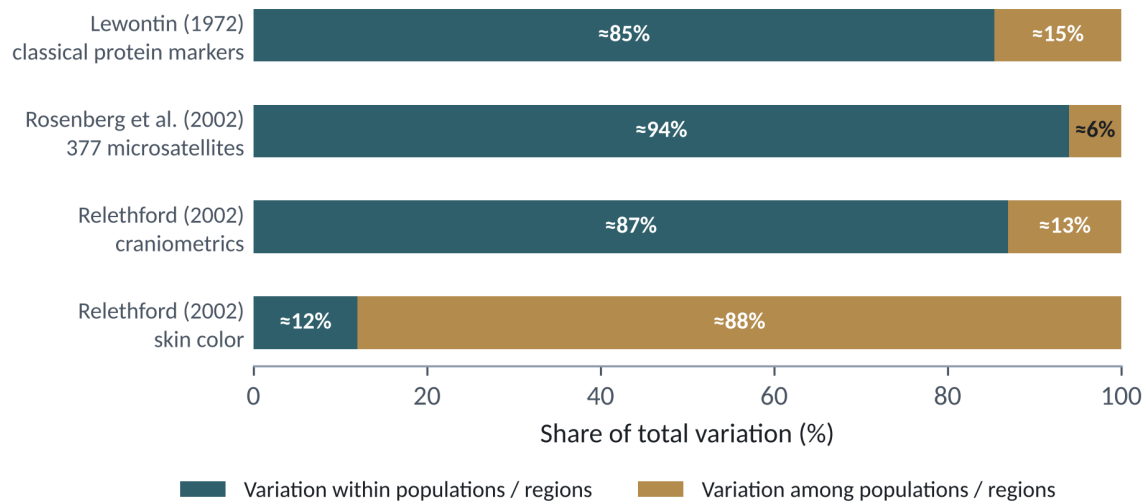


Figure 7. Apportionment of human variation within and among populations or geographic regions, as reported in four classic analyses (values approximate and rounded). Neutral genetic markers and craniometric measurements show most variation within populations. Skin color is the striking exception: it is strongly differentiated among regions, reflecting adaptation to ultraviolet radiation.

5.3 Different traits, different patterns

Morphological traits do not all follow the same geographic pattern. Craniometric variation broadly resembles neutral genetic variation, with most diversity within regions. Skin color is the opposite: most of its variation lies *between* regions ([Relethford, 2002](#)). Skin pigmentation closely tracks ultraviolet radiation levels, which indicates strong natural selection ([Jablonski & Chaplin, 2000](#)). Some aspects of nose shape, such as nostril width, correlate with temperature and humidity, consistent with modest climate adaptation ([Zaidi et al., 2017](#)).

This trait-by-trait mosaic matters for perceived resemblance. Human observers give heavy weight to a few highly visible, regionally differentiated traits such as skin tone, hair form and eye shape. As a result, people may judge two individuals as alike or unlike on the basis of traits that account for only a small share of overall morphological variation. Section 7 returns to this perceptual bias.

5.4 Ancestry-associated resemblance, correctly understood

Two people from the same population are, on average, slightly more likely to share facial-shape alleles than two people from distant populations. All else being equal, they are therefore somewhat more likely to look alike. This report calls this **ancestry-associated resemblance**. It is a real and measurable effect, and any study of look-alikes must control for it. Otherwise shared ancestry will masquerade as something more specific.

Three qualifications are essential:

- **Ancestry is continuous.** Human genetic variation changes gradually across geography, with extensive gene flow. Ancestry clusters are statistical summaries that depend on sampling and on the number of clusters requested. They are not natural boundaries.
- **Overlap dominates.** Because most variation lies within populations, two people from the same population can differ more than two people from different populations. Look-alikes across ancestry groups are expected under the hypothesis, and they are observed.
- **Labels are not biology.** Social categories of race do not correspond to discrete biological groups. Professional bodies have recommended that genetics research move away from racial typology in its population descriptors ([National Academies, 2023](#); [AABA, 2019](#)).

DESIGN PRINCIPLE

In the framework proposed here, ancestry is a **covariate to be measured and controlled**, never a category into which people are sorted. A human variant is defined by a phenotypic configuration that can appear in anyone, not by membership in a population.

5.5 Implications for the hypothesis

Population history adds a third source of structure to the space of human appearance, alongside polygenic combinatorics and developmental constraint. Shared ancestry modestly raises the baseline chance of resemblance. Shared developmental architecture across the whole species makes look-alikes possible *across* ancestry groups. The hypothesis predicts both: look-alikes enriched within populations, and look-alikes present throughout humanity. The balance between the two is an empirical question that the framework in Section 8 is designed to measure.

SECTION 06

The Human Variant Hypothesis

A formal statement, its components, and a working taxonomy of recurring human configurations.

6.1 Formal statement

Building on Sections 2 to 5, the hypothesis can now be stated precisely.

THE HUMAN VARIANT HYPOTHESIS

Human doppelgängers and recurring appearance patterns emerge because a shared genetic and developmental architecture structures human physical variation. Individuals carry different combinations of genetic variants and have different life histories, but the processes that generate human faces and bodies limit the range of viable outcomes. Similar combinations of visible traits can therefore arise in multiple individuals, producing recurring phenotypic configurations across the human population.

The central research question follows: **can the distribution of human appearances be modeled in a way that identifies recurring configurations more systematically than ordinary descriptions of resemblance?**

6.2 Four components

The hypothesis has four components. Each can be examined separately, and each draws on a different body of evidence.

Component 1: a shared biological foundation

Humans develop from a common, species-specific developmental architecture. That architecture explains much of the broad similarity among human bodies and faces: the same facial prominences, the same neural crest origins, the same signaling pathways and the same layout of parts. Evidence: comparative developmental biology, the conserved regulatory architecture of craniofacial development (Section 4), and the consistency of facial-shape loci across ancestry groups ([Xiong et al., 2019](#)).

Component 2: a constrained space of variation

Genetic and developmental processes do not generate every imaginable body plan or facial structure with equal probability. Variation is canalized, channeled along preferred axes, and organized into modules. Evidence: developmental constraint theory ([Maynard Smith et al., 1985](#)), studies of

morphological integration, and the low effective dimensionality of facial shape in morphometric analyses.

Component 3: recurring phenotypic configurations

Multiple people can independently show similar combinations of traits even when they are not closely related. Evidence: the many-to-one genotype–phenotype map of polygenic traits (Section 3), and direct studies of unrelated look-alikes (Section 7).

Component 4: a spectrum of resemblance

Human variants range from partial similarities to exceptionally close look-alikes. They do not form a set of perfectly discrete categories. Resemblance is a matter of degree, measured along a continuous scale, and any thresholds used for analysis are conventions rather than natural boundaries.

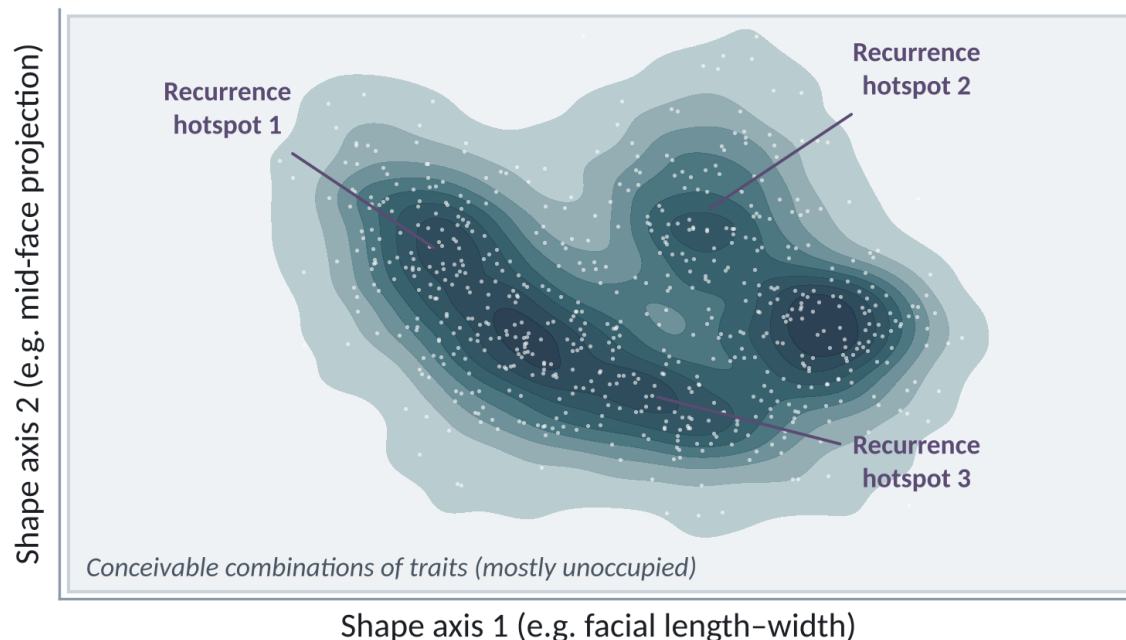


Figure 8. Constrained occupancy of phenotype space (illustrative). The pale rectangle represents all conceivable combinations of two shape features; the shaded contours show where simulated individuals actually fall. Realized forms occupy a limited, structured region with zones of high density, labeled here as recurrence hotspots, where similar configurations are most likely to appear in unrelated individuals.

6.3 What makes someone a human variant?

The report defines a **human variant** as a *recurring phenotypic configuration*: a combination of observable traits that resembles a configuration found in other individuals. The term describes a *region of phenotype space* that is occupied more than once, not a person and not a type of person. Every human is unique. Every human also lies in some region of phenotype space shared, to some degree, with others.

Four working classes of human variant are proposed, distinguished by *which* aspects of phenotype recur.

Table 3. A working taxonomy of human variant classes

Variant class	What recurs	Primary measurement	Main confounders
Morphological	Broad facial geometry: jaw, cheekbones, facial length, feature placement	Aligned 3D landmarks or dense surface distances	Age, body composition, expression
Proportional	Relative proportions and overall structure, independent of size	Scale-free ratios and angles	Measurement choice; growth stage
Composite	A distinctive combination of facial, hair and bodily traits	Multi-scale geometry plus appearance and perceptual ratings	Hairstyle, grooming, photography, rater bias
Ancestry-associated	Statistical tendencies shaped by shared population history	Genetic ancestry estimates as covariates	Sampling design; social categorization

Morphological variants

Individuals who share broad facial geometry, such as jaw shape, cheekbone prominence, facial length, or the relative positions of facial features. Morphological variants are best measured with 3D landmark or dense-surface methods after aligning shapes to remove differences in position, orientation and scale (Rohlf & Slice, 1990).

Proportional variants

Individuals whose features differ in detail but share similar *relative proportions* and overall facial structure. Proportional resemblance survives differences in absolute size and in fine morphology. This is why a child can resemble an adult relative, and why caricaturists can capture a likeness with a few strokes.

Composite variants

Individuals who share a distinctive combination of facial, hair and bodily traits that makes them look especially alike as a whole. Composite resemblance is what most people mean by a doppelgänger. It is the most perceptually striking class, and also the class most affected by non-biological factors such as hairstyle, grooming, expression and photographic angle.

Ancestry-associated resemblance patterns

Similarities shaped partly by shared population history and inherited variation. This class recognizes that human genetic variation overlaps extensively across populations. It is defined as a *statistical tendency within a continuum*, not a boundary, and it serves mainly as a baseline against which the other classes are measured.

6.4 The spectrum of resemblance

Resemblance between two individuals can be placed on a continuous scale. At one end lies the ordinary partial similarity shared by most members of a population. At the other lies near-indistinguishability, normally reached only by monozygotic twins. Unrelated look-alikes sit in the upper tail of the distribution for unrelated pairs. They are rare individually, but in a population of billions the tail contains many pairs in absolute terms.

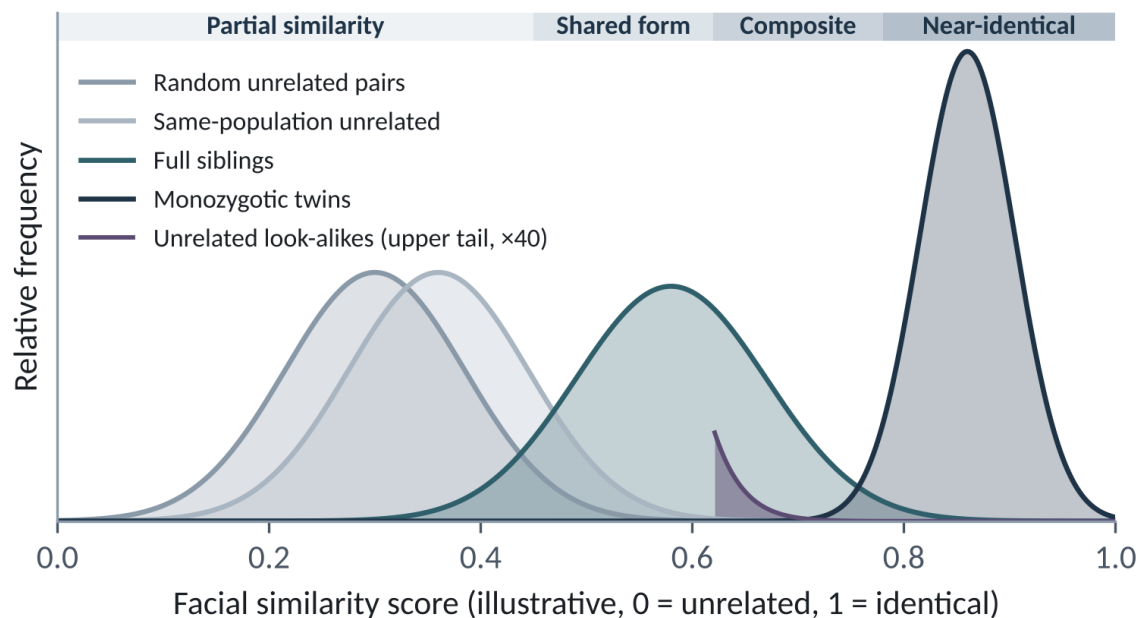


Figure 9. The spectrum of resemblance (illustrative). Curves show the expected distribution of a facial-similarity score for different classes of pairs. Unrelated look-alikes (purple, magnified 40×) are the upper tail of the distribution for same-population unrelated pairs, overlapping the range typical of siblings. Bands across the top mark descriptive zones of the spectrum; their boundaries are conventions, not natural categories.

Two features of this picture matter. First, unrelated look-alikes are **not a separate category**. They are the extreme of a continuous distribution, as Component 4 requires. Second, the extreme of the unrelated distribution **overlaps the sibling distribution**. Some strangers resemble each other as much as typical siblings do. This overlap is exactly what one would expect if resemblance were driven partly by convergent allele sharing rather than by kinship alone.

6.5 A hierarchy of explanation

The hypothesis does not claim that every resemblance has the same cause. It proposes that resemblance between any two people comes from a mixture of distinct channels, whose shares differ from pair to pair:

- **Kinship:** shared alleles inherited by descent from recent common ancestors.
- **Convergent configuration:** independent arrival at similar trait values through the many-to-one genotype–phenotype map.

- **Shared development and environment:** similar developmental histories, nutrition and life-course exposures.
- **Presentation:** hairstyle, grooming, expression, pose, lighting and photography.
- **Perception and sampling:** how observers or algorithms weigh features, and which pairs come to attention at all.

Section 8 turns this list into a measurable decomposition. Section 7 first looks at what is directly known about look-alikes.

SECTION 07

The Doppelgänger Problem: Evidence and Probability

What studies of look-alikes show, and why their frequency depends on how closely we look.

7.1 Are faces unique?

Forensic science has long assumed that faces, like fingerprints, are effectively unique. A direct test measured faces in a large image database to estimate the chance that two people match closely on a set of facial dimensions. For eight measurements taken together, the estimated chance of an exact match between two people was **lower than one in a trillion** ([Lucas & Henneberg, 2015](#)). Matches on only a few measurements are far more common. The rarity of a "perfect" double comes from multiplying many partly independent dimensions together.

This result is not evidence against the hypothesis. It frames it. Exact duplication is effectively impossible. **Perceived** duplication is common. The gap between the two is the core of the doppelgänger problem.

7.2 Look-alikes in the laboratory

The most direct evidence comes from a 2022 study of unrelated look-alikes ([Joshi et al., 2022](#)). The researchers recruited look-alike pairs photographed by the Canadian artist François Brunelle for his long-running project on people who resemble each other. They scored each pair's facial similarity with three independent facial-recognition algorithms. Roughly half the pairs were rated as highly similar by all three, at levels comparable to monozygotic twins on the same measures. The team then collected DNA, DNA-methylation and microbiome data along with questionnaire data on physical and lifestyle traits.

Their main findings, as reported by the authors:

- Among the most similar pairs, a substantial subset (9 of 16) **clustered together on common genetic variants**, sharing many more SNP genotypes than random pairs.
- Many shared variants lay in or near genes previously associated with facial and other physical traits.
- DNA methylation profiles were **less** similar between look-alikes, and microbiome profiles showed little similarity, suggesting the resemblance was mainly linked to inherited sequence variation rather than to shared environment.

- Look-alike pairs also tended to be similar in some physical traits, such as height and weight, and in some self-reported behavioral traits.

READING THE 2022 STUDY CAREFULLY

The study supports Component 3 of the hypothesis, convergent configurations linked to shared common variants. It also has limits that a full research program must address. The sample was small and recruited through a photographic project rather than at random. Participants were mostly of European ancestry, which narrows generalization and complicates control for population structure. Common-variant sharing does not show which variants cause resemblance. Associations with behavior should not be read as evidence that appearance predicts character (see Section 11).

7.3 How perception defines a look-alike

Whether two people "look alike" is judged by a perceiver, human or machine. That perceiver compresses the face into a limited representation.

In humans, faces are processed by specialized neural systems ([Tsao & Livingstone, 2008](#)). Recordings from face-selective neurons in primates indicate that facial identity can be encoded along roughly fifty dimensions of shape and appearance ([Chang & Tsao, 2017](#)). Psychological models describe faces as points in a multidimensional **face space** centered on an average face. Distinctive faces lie far from the center and are recognized more easily. Typical faces lie near the center and are more easily confused ([Valentine, 1991](#)).

Human face matching is also less reliable than people assume. Different photographs of the *same* unfamiliar person can be judged to show different people ([Jenkins et al., 2011](#)). Face-matching ability also varies widely between individuals. So-called super-recognizers perform far above average ([Russell, Duchaine & Nakayama, 2009](#)), and the most accurate identifications come from combining trained human examiners with algorithms ([Phillips et al., 2018](#)). People are generally better at telling apart faces from groups they have more experience with, the so-called own-group or other-race effect ([Meissner & Brigham, 2001](#)). This means perceived resemblance between two strangers depends partly on the observer.

Machine perception does the same kind of compression. Classic "eigenface" methods showed that faces can be approximated by a few dozen principal components ([Sirovich & Kirby, 1987](#); [Turk & Pentland, 1991](#)). Modern deep-learning systems map each face to a compact embedding vector of a few hundred dimensions ([Schroff et al., 2015](#); [Deng et al., 2019](#)). Large-scale government testing has found that many algorithms produce false matches between different people at rates that vary across demographic groups, in some cases by factors of ten to one hundred ([Grother et al., 2019](#)). Algorithmic look-alikes are partly a product of algorithmic compression.

7.4 A probabilistic model of look-alikes

A simple model shows why look-alikes can be both rare and common. Represent each face as a point in a space of d independent standardized dimensions, and call two faces look-alikes if they differ by less than a set tolerance on every dimension, on average. The expected number of look-alikes for a given person in a population of about 8.1 billion can then be computed directly.

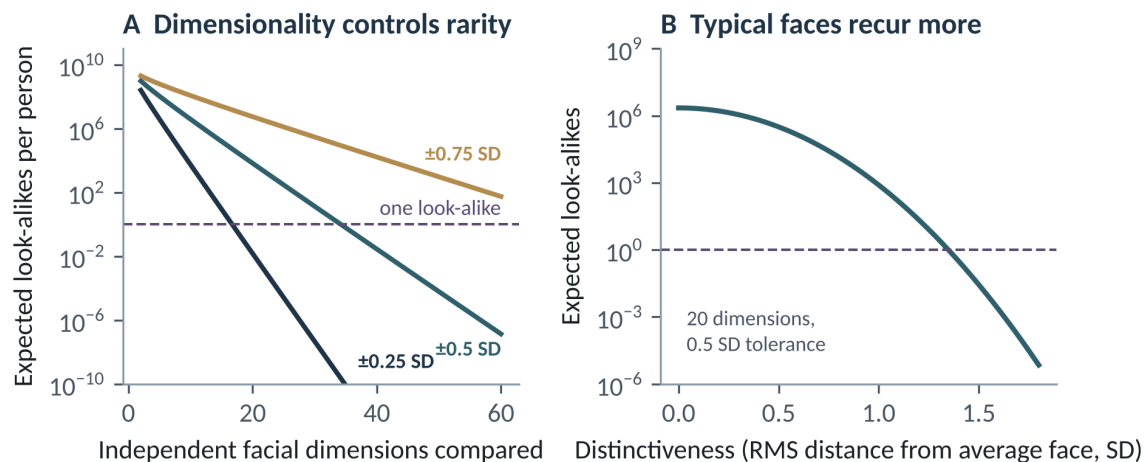


Figure 10. Why look-alikes are both rare and common (illustrative model). (A) Expected number of look-alikes per person in a population of 8.1 billion, as a function of the number of independent facial dimensions compared, for three tolerance levels. With few dimensions or loose tolerance, everyone has many look-alikes; with many dimensions and tight tolerance, essentially no one does. (B) For a fixed comparison of 20 dimensions at 0.5 SD tolerance, people with typical faces (near the population average) have many more expected look-alikes than people with distinctive faces.

The model gives three results that structure the rest of the report.

- **Dimensionality governs rarity.** The expected number of look-alikes falls exponentially as the number of independent dimensions rises. "Are there doppelgängers?" has no single answer. It depends on how many dimensions are compared and how strictly.
- **Constraint raises recurrence.** Developmental integration and modularity (Section 4) reduce the number of *effectively independent* dimensions. The more constrained the phenotype space, the more look-alikes it produces. This is the formal link between creation structures and doppelgängers.
- **Typicality raises recurrence.** People near the center of the distribution have exponentially more look-alikes than people at the periphery, matching the face-space finding that typical faces are more confusable. It also fits the long-standing observation that mathematically averaged faces look familiar and are rated as attractive ([Langlois & Roggman, 1990](#)).

7.5 The role of social prominence

Most documented look-alikes involve at least one person who is widely photographed: a celebrity, a historical figure, or someone who went viral. This reflects **sampling**, not biology. The more people who

see a face, the more likely someone is to notice a match. Famous faces are effectively compared against millions of observers' mental catalogs, and ordinary faces are not. The framework in Section 8 therefore treats social prominence as a separate, measurable factor that inflates *discovered* resemblance without changing *actual* resemblance.

Table 4. Principal lines of evidence bearing on the doppelgänger problem

Line of evidence	Key finding	Bearing on the hypothesis
Metric uniqueness (Lucas & Henneberg, 2015)	Exact match on eight facial measurements has a probability below one in a trillion	Exact doubles are effectively impossible; perceived doubles must reflect partial matching
Multi-omic look-alike study (Joshi et al., 2022)	Many verified look-alikes share common genetic variants, with less epigenetic and microbiome similarity	Supports convergent genetic configuration (P4)
Identity signaling (Sheehan & Nachman, 2014)	Facial traits are unusually variable and weakly correlated with each other	Selection favors distinctiveness, working against recurrence
Face space (Valentine, 1991)	Typical faces are more confusable than distinctive ones	Supports the typicality proposition (P3)
Neural face code (Chang & Tsao, 2017)	Identity encoded along roughly fifty dimensions	Perception compresses faces into a limited space
Algorithm testing (Grother et al., 2019)	False-match rates vary widely across demographic groups	Algorithmic look-alikes partly reflect compression and bias (P6)

7.6 Summary

The evidence supports a moderate form of the hypothesis. Look-alikes are real, occur across populations and carry a measurable genetic signal, consistent with convergent configuration. Their frequency depends strongly on the dimensionality and precision of comparison, on how typical the faces are, and on who is looking. A science of human variants therefore needs to control all of these factors, which is the purpose of the framework that follows.

SECTION 08

A Framework of Generative Phenotypic Recurrence

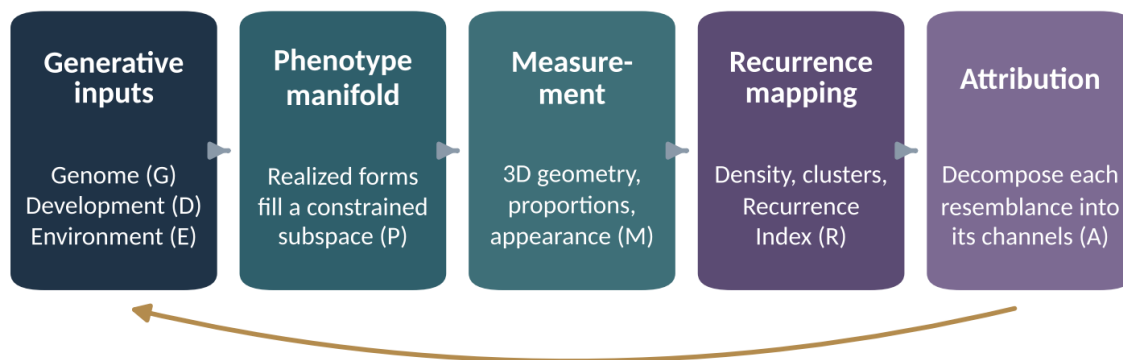
Turning resemblance from an impression into a measurable, decomposable quantity.

8.1 Purpose

The **Generative Phenotypic Recurrence (GPR)** framework is proposed to classify, measure and explain recurring patterns of human resemblance. It is designed to do three things that ordinary descriptions of resemblance cannot:

- **Measure** resemblance on standardized, reproducible scales.
- **Locate** each person and each resemblance within the constrained phenotype space described in Sections 3 and 4.
- **Attribute** each resemblance to its contributing channels: kinship, convergent configuration, shared development and environment, presentation, and perception or sampling.

8.2 The five stages



Feedback: attribution results test and refine the model of generative constraint

Figure 11. The Generative Phenotypic Recurrence framework. Generative inputs (genome, development, environment) produce a constrained phenotype manifold. Standardized measurement places individuals in that space. Recurrence mapping identifies where configurations repeat. Attribution decomposes each resemblance into its contributing channels, and the results feed back to refine the model of generative constraint.

Stage 1: Generative inputs (G, D, E)

Each individual is described by a genome (G), a developmental history (D) and an environmental and life-course history (E). In practice, G is measured by genotyping or sequencing, while D and E are approximated by available covariates such as age, sex, growth history, nutrition and health.

Stage 2: Phenotype manifold (P)

Realized phenotypes occupy a constrained subspace, or manifold, within the space of conceivable forms. The framework estimates this manifold from data, for example with principal component analysis of aligned 3D facial shape, statistical shape models of the kind pioneered in [morphable face models](#), or non-linear embeddings.

Stage 3: Measurement (M)

Resemblance is measured on several complementary scales, kept separate rather than collapsed into one score:

- **Geometric:** distances between aligned 3D landmark configurations or dense surfaces.
- **Proportional:** distances between scale-free ratios and angles.
- **Modular:** geometric distances computed separately for each facial module.
- **Appearance:** color, texture, hair and other surface properties.
- **Perceptual:** similarity judged by human raters and by recognition algorithms, recorded with rater characteristics.

Stage 4: Recurrence mapping (R)

Recurrence mapping asks where in phenotype space configurations repeat, and whether they repeat *more than expected*. The key quantities are defined in Section 8.3.

Stage 5: Attribution (A)

For each pair of interest, the framework estimates how much of their measured resemblance is associated with each channel, using the decomposition in Section 8.4.

8.3 Core metrics

The metrics below are proposed as working definitions. They are deliberately simple so that they can be computed, criticized and refined.

Table 5. Core metrics of the Generative Phenotypic Recurrence framework (working definitions)

Metric	Working definition	What it detects
Effective dimensionality (D_{eff})	Participation ratio of the eigenvalues of the phenotype covariance matrix: $(\sum \lambda)^2 / \sum \lambda^2$	Degree of developmental integration and constraint
Recurrence density (ρ)	Number of individuals within tolerance ϵ of a given configuration, per million sampled	Local crowding of phenotype space
Recurrence Index (R)	Observed neighbors within ϵ divided by neighbors expected under a stated null model	Recurrence beyond chance, given the chosen null
Distinctiveness (δ)	Distance of an individual's configuration from the population mean, in standardized units	Typicality, a predictor of look-alike frequency
Modular resemblance profile	Vector of pairwise distances computed separately for each facial module	Whether resemblance is whole-face or partial
Perceptual amplification (α)	Ratio of perceptual similarity to geometric similarity for a pair	Inflation of resemblance by perception and presentation

Effective dimensionality is the central quantity linking creation structures to doppelgängers. If facial shape were made of many independent traits, D_{eff} would be large and look-alikes vanishingly rare (Section 7.4). If development integrates traits strongly, D_{eff} is small and recurrence common. Estimating D_{eff} for different facial modules, ages and populations is therefore a priority.

The Recurrence Index compares how many neighbors an individual actually has within a set tolerance with how many would be expected under a **null model**. The choice of null model determines what the index detects:

- A null that keeps each trait's distribution but **breaks correlations between traits** tests whether integration (constraint) raises recurrence.
- A null that keeps the full multivariate distribution of the population tests for **local clustering**, meaning hotspots of recurrence beyond what smooth variation would produce.
- A null built **within ancestry strata** tests whether recurrence goes beyond what shared ancestry predicts.

8.4 Decomposing resemblance

For any pair of people, their measured resemblance can be modeled as the sum of contributions from distinct channels plus residual error. With suitable data, each channel can be linked to measurable predictors:

- **Kinship**: genomic relatedness estimated from genome-wide data.
- **Convergent configuration**: similarity in face-associated genetic variants *beyond* that expected from relatedness, such as similarity in genetic scores for specific facial modules after adjusting for kinship and ancestry.

- **Shared development and environment:** similarity in age, growth history, nutrition, body composition and other life-course covariates.
- **Presentation:** the drop in resemblance when hair, grooming, expression and pose are standardized or computationally removed.
- **Perception and sampling:** differences between perceptual and geometric resemblance, rater effects, and the social prominence of the individuals.

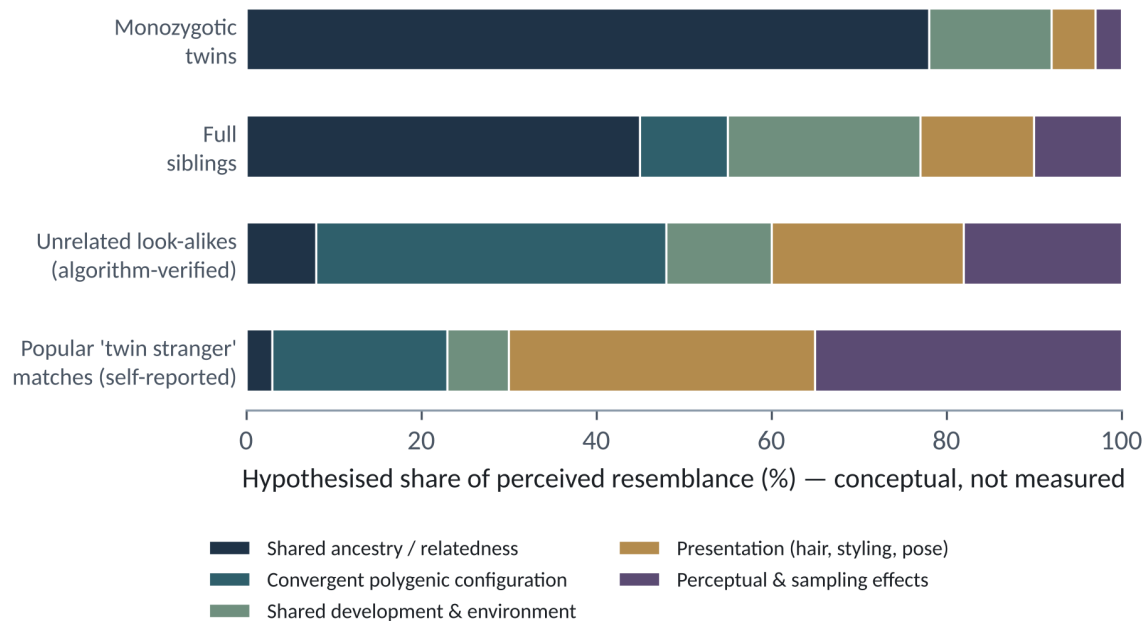


Figure 12. A conceptual decomposition of resemblance for four kinds of pairs. Shares are hypothesized for illustration, not measured. The framework predicts that monozygotic twins resemble each other mainly through shared genotype, siblings through a mix of kinship and shared environment, algorithm-verified unrelated look-alikes largely through convergent configuration, and popularly reported "twin stranger" matches substantially through presentation and perceptual effects.

8.5 Separating four questions

The framework treats four questions as distinct empirical matters that should never be merged:

- **Morphological similarity**, answered by Stage 3 measurement, especially geometric and proportional measures.
- **Genetic identity**, answered by genotyping. Only monozygotic twins (and rare other cases) are genetically identical.
- **Common ancestry**, answered by kinship and ancestry estimation, treated as covariates.
- **Social prominence**, answered by recording how each pair was discovered and how widely each person is photographed or known.

A pair can score high on one dimension and low on another. Unrelated look-alikes are, by definition, high on morphological similarity and low on genetic identity and recent kinship. Celebrity look-alikes are often high on social prominence and only moderate on morphological similarity once presentation is controlled. Keeping these questions apart prevents the most common errors in popular accounts of doppelgängers.

8.6 Propositions

The framework makes six propositions that link creation structures to human variants. Each is testable (Section 9).

- **P1 – Low dimensionality.** The effective dimensionality of human facial shape is far smaller than the number of measured traits, and it is lower within developmental modules than across the whole face.
- **P2 – Structured occupancy.** Realized facial configurations occupy a structured region of phenotype space. Whether that structure is smooth or contains local hotspots beyond a single multivariate distribution is an open empirical question.
- **P3 – Typicality.** The frequency of look-alikes declines with facial distinctiveness, as the model in Section 7.4 predicts.
- **P4 – Genetic convergence.** Unrelated look-alike pairs share more face-associated genetic variation than matched control pairs, beyond what genome-wide relatedness and ancestry predict.
- **P5 – Modular resemblance.** Partial look-alikes resemble each other more within developmental modules than across them.
- **P6 – Perceptual amplification.** Perceived look-alikes outnumber geometry-verified look-alikes, and the excess depends on observer experience, algorithm and social prominence.

SECTION 09

Research Design and Methods

A phased program to test the hypothesis with open, replicable and ethically governed methods.

9.1 Design overview

Testing the hypothesis requires linking three kinds of data on the same individuals: high-quality **phenotype** data (ideally 3D facial and body imaging), **genetic** data, and **covariate** data on development, environment and life course. It also requires a **perceptual** layer recording how human raters and algorithms judge similarity. The program below is indicative. Timelines and sample sizes would be refined in a full proposal.

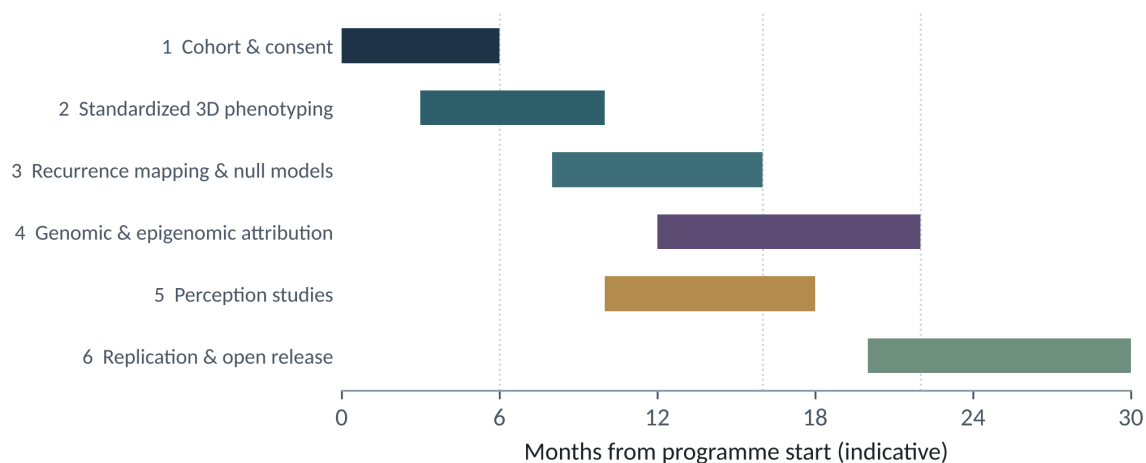


Figure 13. Indicative phasing of a research program to test the Human Variant Hypothesis. Phases overlap so that early phenotyping results can inform recurrence mapping and so that perception studies can run alongside genetic analysis.

9.2 Data sources and sampling

Three complementary sources are proposed:

- **Existing research cohorts with 3D facial imaging and genotypes.** Research consortia have assembled large 3D facial datasets linked to genetic data for studies of normal facial variation and craniofacial disorders. Resources such as [FaceBase](#) show the value of shared, well-governed craniofacial data. Reanalysis of such data, within their consent terms, allows P1, P2, P3 and P5 to be tested without new recruitment.

- **A purpose-recruited look-alike cohort.** Pairs nominated as look-alikes through several channels: public appeals, artist and photography projects, and pairs flagged by algorithms in consented datasets. Recording the discovery channel lets social prominence be measured rather than assumed.
- **Matched control pairs.** For each look-alike pair, control pairs matched for age, sex, and genetically estimated ancestry, drawn at random from the same source population. Controls are essential for P4.

Sampling must be **ancestrally diverse**. Most facial-genetics research to date has been concentrated in European-ancestry samples. A recurrence framework built on one population cannot describe the human species.

9.3 Phenotyping protocol

- **3D surface imaging** of face and head under standardized conditions: neutral expression, hair restrained, no makeup or eyewear, and controlled pose and lighting.
- **Landmark and dense-surface registration** followed by Procrustes alignment, producing geometric, proportional and modular measures.
- **Body measures:** height, weight, body proportions and, where possible, 3D body scans.
- **Appearance measures:** calibrated skin reflectance, eye color and hair characteristics.
- **Repeated measurement** for a subsample, to estimate measurement error and short-term variability.
- **Standard 2D photographs** under both controlled and naturalistic conditions, for perceptual and algorithmic comparison.

9.4 Analytical plan

Estimating the phenotype manifold

Principal component analysis and related methods estimate the dimensions of shape variation for the whole face and for each module. Effective dimensionality is calculated with confidence intervals by bootstrap resampling, and compared across modules, ages, sexes and ancestry strata (P1).

Recurrence mapping

For every individual, neighbor counts are computed at several tolerance levels and compared with the null models defined in Section 8.3, giving a Recurrence Index for each person. Recurrence is then modeled as a function of distinctiveness (P3). Density-estimation and clustering methods test for structured occupancy and local hotspots (P2). Results are reported across a range of thresholds rather than at one arbitrary cut-off.

Genetic attribution

For look-alike and control pairs, genome-wide relatedness, ancestry similarity and similarity in module-specific facial genetic scores are estimated. Mixed models then test whether look-alikes share more face-associated variation than controls after adjusting for kinship and ancestry (P4). Analyses are pre-registered, and multiple-testing corrections apply.

Perception studies

Human raters drawn from diverse backgrounds, and multiple recognition algorithms, judge similarity for the same pairs under controlled and naturalistic image conditions. Comparing perceptual with geometric similarity estimates perceptual amplification and its dependence on rater background and image presentation (P6).

Table 6. Testable propositions, tests and disconfirming results

Proposition	Prediction	Test	Would disconfirm
P1 Low dimensionality	D_{eff} far below the number of measured traits; lower within modules	PCA of aligned 3D shape with bootstrap intervals	D_{eff} close to the trait count
P2 Structured occupancy	Occupied region is compact and structured	Density estimation against shuffled-covariance nulls	Occupancy matches the independent-trait null
P3 Typicality	Look-alike count falls with distinctiveness	Regression of neighbor counts on δ across tolerances	No relation, or the reverse
P4 Genetic convergence	Look-alikes share more face-associated variation than matched controls	Mixed models adjusting for kinship and ancestry	No excess sharing after adjustment
P5 Modular resemblance	Partial look-alikes are more similar within modules than across them	Comparison of modular resemblance profiles with random partitions	Resemblance unrelated to module boundaries
P6 Perceptual amplification	Perceived look-alikes outnumber geometric ones; effect varies by rater and algorithm	Paired perceptual and geometric ratings	Perfect agreement with no observer dependence

9.5 Quality and reproducibility

- **Pre-registration** of hypotheses, metrics, thresholds and analysis plans before data access.
- **Independent replication** in at least one held-out cohort for every primary finding.
- **Sensitivity analyses** across tolerance levels, registration methods and null models.
- **Open methods:** release of analysis code and summary statistics. Individual-level data stay in controlled-access environments consistent with participant consent.

SECTION 10

Falsifiability, Limitations and Alternative Explanations

What would show the hypothesis to be wrong, and what else might explain the evidence.

10.1 Falsification criteria

A hypothesis that explains every possible outcome explains nothing. The Human Variant Hypothesis in its GPR form makes claims that could fail. The strongest tests are:

- **If effective dimensionality is high** and close to the number of measured traits, the claim that developmental constraint drives recurrence is seriously weakened.
- **If unrelated look-alikes share no more face-associated genetic variation than matched controls** after adjustment for kinship and ancestry, the convergent-configuration channel is not supported, and look-alikes would be better explained by presentation and perception.
- **If look-alike frequency is unrelated to distinctiveness**, the face-space model of recurrence is incomplete.
- **If partial look-alikes do not show modular structure**, the link between developmental modularity and resemblance fails.
- **If perceived look-alikes are fully confirmed by geometry** with no observer dependence, the perceptual-amplification proposition fails, which would simplify the science.

10.2 Limitations

- **Measurement dependence.** Every resemblance metric reflects choices about landmarks, alignment, scale and weighting. The framework addresses this by reporting several metrics, but no measure is neutral.
- **Discovery bias.** Look-alikes reach researchers through channels, such as fame, virality and photographic projects, that are not random samples of humanity.
- **Population coverage.** Facial-genetic knowledge is concentrated in a few ancestry groups. Conclusions may not generalize until diversity improves.
- **Environmental confounding.** Shared environment and culture can produce similarities in body composition, grooming and expression that resemble genetic convergence.

- **Causal inference.** Even strong associations between shared variants and resemblance do not identify causal variants or mechanisms. Experimental and functional studies would be needed.
- **The analogy to creation structures.** The conceptual bridge from general generative principles to human development is heuristic. It organizes questions but does not supply mechanisms.

10.3 Alternative explanations

Table 7. Alternative explanations and how the framework distinguishes them

Alternative	Claim	Distinguishing test
Pure chance	Look-alikes are the expected tail of a large population with no special structure	Recurrence Index relative to an independent-trait null; GPR predicts $R > 1$ for constrained traits
Hidden kinship	Look-alikes are undocumented relatives	Genome-wide relatedness estimation; restrict analysis to pairs with no detectable recent kinship
Population stratification	Look-alikes reflect shared ancestry only	Ancestry-matched controls and within-stratum null models
Presentation	Resemblance comes from hair, grooming, pose and photography	Standardized 3D imaging; measure the drop in similarity when presentation is removed
Perceptual artifact	Resemblance is in the eye of the beholder	Compare geometric and perceptual similarity across diverse raters and algorithms
Selective discovery	Only striking or famous pairs are reported	Record discovery channel; recruit through algorithmic screening of consented cohorts

10.4 A note on adaptive storytelling

Biology has a long history of explaining every trait as an adaptation for something. Gould and Lewontin famously criticized this tendency ([Gould & Lewontin, 1979](#)). The hypothesis does not claim that recurring configurations are adaptive, or that look-alikes exist "for" any purpose. Recurrence is expected as a *by-product* of how faces are built. Selection for identity signaling ([Sheehan & Nachman, 2014](#)) would, if anything, work *against* recurrence by favoring diversity. The observed balance between distinctiveness and recurrence may reflect a tension between selection for individuality and developmental constraint, and that tension deserves study in its own right.

SECTION 11

Ethics, Governance and the Boundary Against Typology

Studying human resemblance safely requires clear limits on what the science claims and how its data are used.

11.1 The historical shadow

Research on human appearance has a troubled history. Francis Galton, who pioneered composite photography to find "typical" faces ([Galton, 1879](#)), was also a founder of eugenics. Nineteenth-century physiognomy and criminal anthropology claimed to read character, criminality and worth from facial form. Twentieth-century racial typology sorted people into supposed biological types. These programs were scientifically wrong and caused profound harm.

Modern machine learning has revived some of these claims, including attempts to infer criminality or other personal characteristics from facial images. Researchers have criticized such work as a return of physiognomy in algorithmic form ([Agüera y Arcas, Mitchell & Todorov, 2017](#)). Any research program on human variants must be built to make these errors impossible, not merely unlikely.

11.2 Five boundaries

BOUNDARIES OF THE HUMAN VARIANT HYPOTHESIS

- 1. Variants are not types.** A human variant is a region of phenotype space, never a category of person.
- 2. Variants are not races.** Recurring configurations cut across ancestry, and ancestry is treated as a continuous covariate.
- 3. Appearance is not character.** The framework makes no claims, and permits no inferences, about personality, ability, behavior or worth from appearance.
- 4. Resemblance is not identity.** Looking alike does not make two people interchangeable, related, or the same in any social or legal sense.
- 5. Description is not value.** Typical and distinctive configurations are statistical descriptions, not judgments of beauty, normality or desirability.

11.3 Privacy and biometric data

Facial images and genetic data are among the most sensitive personal data. Under the European Union's General Data Protection Regulation, biometric data used for unique identification and genetic data are special categories that require heightened protection ([GDPR, Art. 9](#)). Several US states also regulate the collection of biometric identifiers. Research on look-alikes adds a specific risk: data showing that two people look alike could be used to impersonate them or to defeat biometric security.

Required safeguards include:

- **Explicit, specific consent** covering facial imaging, genetic analysis, perceptual studies and any data sharing, with the right to withdraw.
- **Consent from both members** of any look-alike pair before pair-level data are analyzed or published.
- **Controlled-access storage** with no release of identifiable images or individual-level genomes in open repositories.
- **Prohibition of re-identification**, and no use of research data for surveillance, policing or commercial face matching.

11.4 Forensic and commercial uses

Research linking DNA to facial shape ([Claes et al., 2014](#); [Sero et al., 2019](#)) has raised the prospect of predicting appearance from DNA. Claims about identifying individuals from genomes by predicting their traits have been strongly debated ([Lippert et al., 2017](#)). The human variant framework strengthens a key caution here. Because many genotypes converge on similar faces, and because perceived resemblance is inflated by compression, **appearance predicted from DNA is inherently non-unique**. In a large population, a predicted face will match many people. Forensic use of facial predictions risks wrongly implicating look-alikes.

The same logic applies to facial recognition. Look-alikes are a structural source of false matches, and false-match rates differ across demographic groups ([Grother et al., 2019](#)). Research on human variants should be used to **reduce** these harms, for example by quantifying look-alike risk for high-stakes identification and by making the case for human review and corroborating evidence.

11.5 Language and communication

How results are described matters. The program should avoid language that implies types, ranks or essences (for example, "kinds of faces" or "facial races"). It should use population descriptors consistent with current guidance ([National Academies, 2023](#)). Every public communication should state the limits in Section 11.2, and community and participant representatives should review communications before release.

SECTION 12

Conclusion: The Recurrence of Human Form

Uniqueness and recurrence come from the same generative process.

12.1 Summary of the argument

This report began with an apparent paradox. Every human face is unique, yet human faces recur, sometimes so closely that strangers are mistaken for one another. The report argued that the paradox dissolves once human appearance is understood as the output of a **creation structure**: a compact generative specification (the genome) executed through a constrained, iterative process (development) within a structured population history.

That generative logic predicts both observations. Combinatorial genetics and the high dimensionality of fine facial detail make exact duplication effectively impossible. The many-to-one genotype–phenotype map, developmental canalization and integration, modular organization and shared ancestry concentrate realized faces into a structured, lower-dimensional space. In that space, similar configurations recur.

12.2 What has been established, and what is proposed

The report distinguishes three levels of confidence:

- **Well established:** facial form is heritable and highly polygenic, mostly through regulatory variation. Development is canalized, modular and integrated. Most human genetic variation lies within populations. Perceived resemblance depends on how faces are compressed by human and machine perceivers.
- **Supported but preliminary:** unrelated look-alikes share more common genetic variation than random pairs, with less similarity in epigenetic and microbiome profiles. This rests on a small, non-random sample that needs replication.
- **Proposed in this report:** the Human Variant Hypothesis as a unifying account, the taxonomy of variant classes, the Generative Phenotypic Recurrence framework and its metrics, and propositions P1 to P6.

12.3 The broader significance

If the hypothesis holds, the recurrence of human form is one specific biological expression of a general principle: complex systems that generate outcomes from compact rules under constraint produce variable but bounded results, and within those bounds certain forms recur. The same principle governs the arrangement of seeds in a flower head, the stripes on a fish and the branching of a tree. In humans, it governs something more personal, the faces we recognize as our own and as others'.

The study of human variants is therefore not about sorting people. It is about understanding the architecture that makes each person both unique and familiar. A doppelgänger is a visible trace of how deeply shared the human generative process is. However individual each of us looks, we are all built by the same structures.

NEXT STEPS

1. Reanalyze existing 3D facial–genetic datasets to estimate effective dimensionality and Recurrence Index distributions (P1–P3, P5).
2. Develop a pre-registered protocol and governance framework for a diverse look-alike cohort with matched controls (P4, P6).
3. Convene an interdisciplinary advisory group, spanning genetics, developmental biology, perception science, ethics and community representation, to review the framework before data collection.

APPENDIX A

Glossary

Working definitions of key terms as used in this report.

Allometry — The relationship between the size of an organism or part and the shape or proportions of other parts.

Canalization — The buffering of development so that it reliably reaches similar outcomes despite genetic or environmental perturbation.

Composite variant — A proposed class of human variant in which a distinctive combination of facial, hair and bodily traits recurs.

Creation structure — In this report, a set of generative rules, constraints and processes that turns compact specifying information into complex form.

Developmental constraint — A bias on the production of variant phenotypes, or a limit on phenotypic variability, caused by the structure or dynamics of the developmental system.

Distinctiveness (δ) — The standardized distance of a person's facial configuration from the population average.

Doppelgänger — A person who closely resembles another, unrelated person.

Effective dimensionality (D_{eff}) — A summary of how many independent directions of variation a set of traits really has, after accounting for correlations among them.

Enhancer — A regulatory DNA element that increases the transcription of one or more genes, often in specific tissues and developmental stages.

Face space — A psychological model in which faces are represented as points in a multidimensional space centered on an average face.

Generative Phenotypic Recurrence (GPR) — The framework proposed in this report for measuring, mapping and attributing recurring human configurations.

Genome-wide association study (GWAS) — A study that tests many genetic variants across the genome for association with a trait.

Heritability — The proportion of trait variance in a particular population and environment that is associated with genetic differences.

Human variant — A proposed analytical term for a recurring phenotypic configuration. It is not a type, race or category of person.

Modularity — The organization of traits into sets that covary strongly internally and weakly with other sets.

Morphological integration — Covariation among traits that arises from shared developmental, genetic or functional factors.

Neural crest — A population of migratory embryonic cells that forms much of the facial skeleton and connective tissue.

Phenotype manifold — The constrained subspace of all conceivable forms that is actually occupied by realized individuals.

Phyllotaxis — The arrangement of leaves, florets or seeds on a plant, often following spiral patterns linked to the golden angle.

Pleiotropy — The influence of a single genetic variant on more than one trait.

Polygenic trait — A trait influenced by many genetic variants, each usually of small effect.

Procrustes alignment — A method that superimposes landmark configurations by removing differences in position, orientation and scale.

Reaction–diffusion — A mechanism in which interacting, diffusing chemicals self-organize into spatial patterns.

Recurrence Index (R) — The ratio of the observed number of near neighbors in phenotype space to the number expected under a stated null model.

Single-nucleotide polymorphism (SNP) — A position in the genome at which individuals differ by a single DNA base.

APPENDIX B

References

All sources cited in the text, with persistent links.

-
- 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526, 68–74. doi.org/10.1038/nature15393
- American Association of Biological Anthropologists. (2019). AABA statement on race and racism. bioanth.org/about/aaba-statement-on-race-racism/
- Agüera y Arcas, B., Mitchell, M., & Todorov, A. (2017). Physiognomy's new clothes. *Medium*. medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a
- Alberch, P. (1989). The logic of monsters: Evidence for internal constraint in development and evolution. *Geobios*, 22, 21–57. [doi.org/10.1016/S0016-6995\(89\)80006-3](https://doi.org/10.1016/S0016-6995(89)80006-3)
- Attanasio, C., et al. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. *Science*, 342, 1241006. doi.org/10.1126/science.1241006
- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *Proceedings of SIGGRAPH '99*, 187–194. doi.org/10.1145/311535.311556
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169, 1177–1186. doi.org/10.1016/j.cell.2017.05.038
- Chang, L., & Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell*, 169, 1013–1028. doi.org/10.1016/j.cell.2017.05.011
- Claes, P., et al. (2014). Modeling 3D facial shape from DNA. *PLoS Genetics*, 10, e1004224. doi.org/10.1371/journal.pgen.1004224
- Claes, P., et al. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nature Genetics*, 50, 414–423. doi.org/10.1038/s41588-018-0057-4
- Cole, J. B., et al. (2017). Human facial shape and size heritability and genetic correlations. *Genetics*, 205, 967–978. doi.org/10.1534/genetics.116.193185
- Cordero, D. R., et al. (2011). Cranial neural crest cells on the move: Their roles in craniofacial development. *American Journal of Medical Genetics Part A*, 155, 270–279. doi.org/10.1002/ajmg.a.33702
- Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *arXiv:1801.07698*. arxiv.org/abs/1801.07698
- Douady, S., & Couder, Y. (1992). Phyllotaxis as a physical self-organized growth process. *Physical Review Letters*, 68, 2098–2101. doi.org/10.1103/PhysRevLett.68.2098
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399–433. doi.org/10.1017/S0080456800012163
- Fraga, M. F., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *PNAS*, 102, 10604–10609. doi.org/10.1073/pnas.0500398102
- Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *Journal of the Anthropological Institute*, 8, 132–144. doi.org/10.2307/2841021
- GDPR. (2016). Article 9: Processing of special categories of personal data. Regulation (EU) 2016/679. gdpr-info.eu/art-9-gdpr/
- Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm. *Proceedings of the Royal Society B*, 205, 581–598. doi.org/10.1098/rspb.1979.0086

- Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face Recognition Vendor Test (FRVT) Part 3: Demographic effects* (NISTIR 8280). National Institute of Standards and Technology. doi.org/10.6028/NIST.IR.8280
- Hallgrímsson, B., et al. (2009). Deciphering the palimpsest: Studying the relationship between morphological integration and phenotypic covariation. *Evolutionary Biology*, 36, 355–376. doi.org/10.1007/s11692-009-9076-5
- Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, 4, e1000008. doi.org/10.1371/journal.pgen.1000008
- Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of Human Evolution*, 39, 57–106. doi.org/10.1006/jhev.2000.0403
- Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, 121, 313–323. doi.org/10.1016/j.cognition.2011.08.001
- Jónsson, H., et al. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, 549, 519–522. doi.org/10.1038/nature24018
- Joshi, R. S., Rigau, M., García-Prieto, C. A., et al. (2022). Look-alike humans identified by facial recognition algorithms show genetic similarities. *Cell Reports*, 40, 111257. doi.org/10.1016/j.celrep.2022.111257
- Kirschner, M., & Gerhart, J. (1998). Evolvability. *PNAS*, 95, 8420–8427. doi.org/10.1073/pnas.95.15.8420
- Klingenberg, C. P. (2008). Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics*, 39, 115–132. doi.org/10.1146/annurev.ecolsys.37.091305.110054
- Kondo, S., & Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *Science*, 329, 1616–1620. doi.org/10.1126/science.1179047
- Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell*, 78, 191–201. [doi.org/10.1016/0092-8674\(94\)90290-9](https://doi.org/10.1016/0092-8674(94)90290-9)
- Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, 1, 115–121. doi.org/10.1111/j.1467-9280.1990.tb00079.x
- Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology*, 6, 381–398. doi.org/10.1007/978-1-4684-9063-3_14
- Lippert, C., et al. (2017). Identification of individuals by trait prediction using whole-genome sequencing data. *PNAS*, 114, 10166–10171. doi.org/10.1073/pnas.1711125114
- Liu, F., et al. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genetics*, 8, e1002932. doi.org/10.1371/journal.pgen.1002932
- Lucas, T., & Henneberg, M. (2015). Are human faces unique? A metric approach to finding single individuals without duplicates in large samples. *Forensic Science International*, 257, 514.e1–514.e6. doi.org/10.1016/j.forsciint.2015.09.003
- Maynard Smith, J., et al. (1985). Developmental constraints and evolution. *Quarterly Review of Biology*, 60, 265–287. doi.org/10.1086/414425
- Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, 7, 3–35. doi.org/10.1037/1076-8971.7.1.3
- Naqvi, S., et al. (2021). Shared heritability of human face and brain shape. *Nature Genetics*, 53, 830–839. doi.org/10.1038/s41588-021-00827-w
- National Academies of Sciences, Engineering, and Medicine. (2023). *Using Population Descriptors in Genetics and Genomics Research*. The National Academies Press. doi.org/10.17226/26902
- Phillips, P. J., et al. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *PNAS*, 115, 6171–6176. doi.org/10.1073/pnas.1721355115
- Polderman, T. J. C., et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47, 702–709. doi.org/10.1038/ng.3285
- Prescott, S. L., et al. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell*, 163, 68–83. doi.org/10.1016/j.cell.2015.08.036
- Ramachandran, S., et al. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *PNAS*, 102, 15942–15947. doi.org/10.1073/pnas.0507611102

- Raspopovic, J., et al. (2014). Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science*, 345, 566–570. doi.org/10.1126/science.1252960
- Relethford, J. H. (2002). Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, 118, 393–398. doi.org/10.1002/ajpa.10079
- Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39, 40–59. doi.org/10.2307/2992207
- Rosenberg, N. A., et al. (2002). Genetic structure of human populations. *Science*, 298, 2381–2385. doi.org/10.1126/science.1078311
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16, 252–257. doi.org/10.3758/PBR.16.2.252
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *arXiv:1503.03832*. arxiv.org/abs/1503.03832
- Sero, D., et al. (2019). Facial recognition from DNA using face-to-DNA classifiers. *Nature Communications*, 10, 2557. doi.org/10.1038/s41467-019-10617-y
- Sheehan, M. J., & Nachman, M. W. (2014). Morphological and population genomic evidence that human faces have evolved to signal individual identity. *Nature Communications*, 5, 4800. doi.org/10.1038/ncomms5800
- Sheth, R., et al. (2012). Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science*, 338, 1476–1480. doi.org/10.1126/science.1226804
- Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4, 519–524. doi.org/10.1364/JOSAA.4.000519
- Thompson, D. W. (1917). *On Growth and Form*. Cambridge University Press. (Project Gutenberg edition.) www.gutenberg.org/ebooks/55264
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, 31, 411–437. doi.org/10.1146/annurev.neuro.30.051606.094238
- Turing, A. M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B*, 237, 37–72. doi.org/10.1098/rstb.1952.0012
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3, 71–86. doi.org/10.1162/jocn.1991.3.1.71
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A*, 43, 161–204. doi.org/10.1080/14640749108400966
- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era: Concepts and misconceptions. *Nature Reviews Genetics*, 9, 255–266. doi.org/10.1038/nrg2322
- Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150, 563–565. doi.org/10.1038/150563a0
- Wagner, G. P., & Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution*, 50, 967–976. doi.org/10.1111/j.1558-5646.1996.tb02339.x
- White, J. D., et al. (2021). Insights into the genetic architecture of the human face. *Nature Genetics*, 53, 45–53. doi.org/10.1038/s41588-020-00741-7
- Xiong, Z., et al. (2019). Novel genetic loci affecting facial shape variation in humans. *eLife*, 8, e49898. doi.org/10.7554/eLife.49898
- Zaidi, A. A., et al. (2017). Investigating the case of human nose shape and climate adaptation. *PLoS Genetics*, 13, e1006616. doi.org/10.1371/journal.pgen.1006616

H H E U R I S T I C S

Navigating a changing world

H Heuristics is an independent publication by Hunter Hughes exploring the ideas, structures and patterns that shape a changing world, from science and technology to society and human experience.

This report is part of an ongoing series of long-form research essays. It is intended to stimulate rigorous inquiry and discussion. It is not medical, genetic, forensic or legal advice.

Read more and subscribe at hheuristics.substack.com.