

Information Processing

Pessimism of the Intellect, Optimism of the Will [Favorite posts](#) [Twitter: @hsu_steve](#)



Stephen Hsu
 Follow

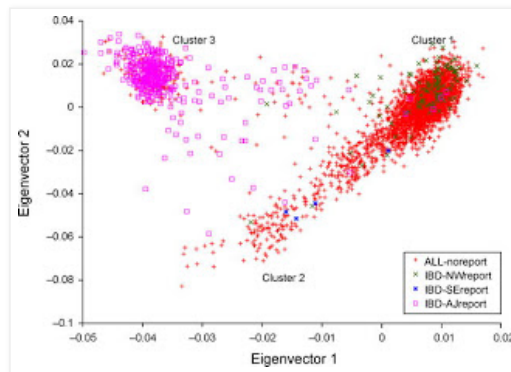
[View my complete profile](#)

Physicist, Startup Founder, Blogger, Dad

Tuesday, January 22, 2008

"No scientific basis for race"

"It's just a social construction" -- a picture is worth a million words...



Caption: Each point is an individual, and the axes are two [principal components](#) in the space of genetic variation. Colors correspond to individuals of different European ancestry. (Via [gnxp](#).)

The figure is from the following paper, reporting on a study of over 4000 individuals. The researchers can group most Europeans into a geographical cline (NW vs SE, that's the red band in the lower right of the figure; there are two clusters but also individuals who are in-between) + Ashkenazim (the pink isolated cluster in the upper left) using a few hundred markers. I'm sure even better resolution can be obtained with more loci.

[Discerning the Ancestry of European Americans in Genetic Association Studies](#)

Abstract: European Americans are often treated as a homogeneous group, but in fact form a structured population due to historical immigration of diverse source populations. Discerning the ancestry of European Americans genotyped in association studies is important in order to prevent false-positive or false-negative associations due to population stratification and to identify genetic variants whose contribution to disease risk differs across European ancestries. Here, we investigate empirical patterns of population structure in European Americans, analyzing 4,198 samples from four genome-wide association studies to show that components roughly corresponding to northwest European, southeast European, and Ashkenazi Jewish ancestry are the main sources of European American population structure. Building on this insight, we constructed a panel of 300 validated markers that are highly informative for distinguishing these ancestries. We demonstrate that this panel of markers can be used to correct for stratification in association studies that do not generate dense genotype data.

The money paragraph: "...Here we mine much larger datasets (more markers and more samples) to identify a panel of 300 highly ancestry-informative markers which accurately distinguish not just northwest and southeast European, but also Ashkenazi Jewish ancestry. This panel of markers is likely to be useful in targeted disease studies involving European Americans."

For previous discussion of genetic clustering of human populations, see [here](#) and [here](#). It has been known for some time that major continental groups ("races") form distinct clusters. [Improved data](#) allow for much finer exploration of clusters within clusters.

This post is getting a lot of traffic from [metafilter](#), and judging from the comments people are confused. I offer the following from the [second link](#) in the paragraph above:

...no matter what genetic markers you choose: SNPs, STRs, no matter how you choose them: randomly or based on their "informativeness", it is relatively easy to classify DNA into the correct continental origin. Depending on the marker types (e.g., indel vs. microsatellite), and their informativeness (roughly the distribution differences between populations), one may require more or less markers to achieve a high degree of accuracy. But, the conclusion is the same: after a certain number of markers, you always succeed in classifying individuals according to continental origin.

Thus, the emergent pattern of variation is not at all subjectively constructed: it does not deal specifically with visible traits (randomly chosen markers could influence any trait, or none at all), nor does it privilege markers exhibiting large population differences. The structuring of humanity into more or less disjoint groups is not a subjective choice: it emerges naturally from the genomic composition of humans, irrespective of how you study this composition. Rather than proving that race is skin-deep, non-existent, or unimportant, modern genetic science is both proving that it is in fact existent, but also sets the foundation for the study of its true importance, which is probably somewhere in between the indifference of the sociologists and the hyperbole of the racists.

One thing commenters seem particularly confused about is the difference between phenotypic and genetic variation. The clustering data show very clearly that, in certain subspaces, the genetic variation within a particular population cluster is less than between clusters. That is, the genetic "distance" between two individuals within a cluster is typically much less than the distance between clusters. (Technical comment: this depends on the number of loci or markers used. As the number gets large the distance between clusters becomes much larger than the individual cluster radius. For continental clusters, if hundreds or thousands of markers are used the intercluster distance dominates the intracluster size. Further technical comment: you may have read the misleading statistic, spread by the intellectually dishonest [Lewontin](#), that 85% percent of all human genetic variation occurs within groups and only 15% between groups. The statistic is true, but what is often falsely claimed is that this breakup of variances (larger within group than between group) prevents any meaningful genetic classification of populations. This false conclusion neglects the correlations in the genetic data that are revealed in a cluster analysis. See [here](#) for a simple example which shows that there can be dramatic group differences in phenotypes even if every version of every gene is found in two groups -- as long as the frequency or probability distributions are distinct. Sadly, understanding this point requires just enough mathematical ability that it has [eluded](#) all but a small number of experts.) **Update:** see [here](#) for an explanation in pictures of Lewontin's fallacy. I also edited the paragraph above for clarity.

On the other hand, for most phenotypes (examples: height or IQ, which are both fairly heritable, except in cases of extreme environmental deprivation), there is significant overlap between different population distributions. That is, Swedes might be taller than Vietnamese on average, but the range of heights within each group is larger than the difference in the averages. Nevertheless, at the tails of the distribution one would find very large discrepancies: for example the percentage of the Swedish population that is over 2 meters tall (6'7") might be 5 or 10 times as large as the percentage of the Vietnamese population. If two groups differed by, say, 10 points in average IQ (2/3 of a standard deviation), the respective distributions would overlap quite a bit (more in-group than between-group variation), but the fraction of people with IQ above some threshold (e.g., >140) would be radically different. It has been [claimed](#) that 20% of all Americans with IQ > 140 are Jewish, even though Jews comprise only 3% of the total population.

...The imbalance continues to increase for still higher IQ's. New York City's public-school system used to administer a pencil-and-paper IQ test to its entire school population. In 1954, a psychologist used those test results to identify all 28 children in the New York public-school system with measured IQ's of 170 or higher. Of those 28, 24 were Jews.

There is no strong evidence for specific gene variants (alleles) that lead to group differences (differences between clusters) in behavior or intelligence, but progress on the genomic side of this question will be rapid in coming years, as the price to sequence a genome is dropping at an exponential rate.

What seems to be true (from preliminary studies) is that the gene variants that were under strong selection (reached fixation) over the last 10k years are different in different clusters. That is, the way that modern people in each cluster differ, due to natural selection, from their own ancestors 10k years ago is not the same in each cluster -- we have been, at least at the genetic level, experiencing divergent evolution. In fact, [recent research](#) suggests that 7% or more of all our genes are mutant versions that replaced earlier variants through natural selection over the last tens of thousands of years.

Posted by [Stephen Hsu](#) at [12:01 AM](#)

[20 Comments](#)



Labels: [ashkenazim](#), [genetics](#), [political correctness](#), [science](#)



Disqus seems to be taking longer than usual. [Reload?](#)

[Newer Post](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)



[Stephen Hsu](#)

[Follow](#) < 0

[View my complete profile](#)

Blog Archive

- ▶ [2019](#) (12)
- ▶ [2018](#) (128)
- ▶ [2017](#) (150)
- ▶ [2016](#) (206)
- ▶ [2015](#) (167)
- ▶ [2014](#) (184)
- ▶ [2013](#) (211)
- ▶ [2012](#) (222)
- ▶ [2011](#) (266)
- ▶ [2010](#) (234)
- ▶ [2009](#) (204)
- ▼ [2008](#) (255)
 - ▶ [12](#) (25)
 - ▶ [11](#) (31)
 - ▶ [10](#) (34)
 - ▶ [09](#) (31)
 - ▶ [08](#) (16)
 - ▶ [07](#) (16)
 - ▶ [06](#) (19)
 - ▶ [05](#) (21)
 - ▶ [04](#) (16)
 - ▶ [03](#) (20)
 - ▶ [02](#) (10)
 - ▼ [01](#) (16)

[L'Affaire Kerviel: the mystery persists](#)

[Societe Generale: a proud giant laid low](#)

[Fake alpha, tail risk and compensation in finance](#)

[L'Affaire Kerviel and a rogues' gallery of rogue t...](#)

["No scientific basis for race"](#)

[A higher intelligence at Goldman](#)

[So long, Bobby](#)

[Honors college survey](#)

[And the winner is...](#)

[LHC short films](#)

[What is the entropy of the universe?](#)

[Exodus to the virtual world](#)

[CES, Las Vegas](#)

[Harvard Magazine](#)

[Understanding the subprime crisis](#)

[How far to fall?](#)

- ▶ [2007](#) (196)
- ▶ [2006](#) (154)
- ▶ [2005](#) (253)
- ▶ [2004](#) (100)

Labels

- [physics](#) (347)
- [genetics](#) (281)
- [finance](#) (261)
- [globalization](#) (258)
- [brainpower](#) (248)
- [technology](#) (205)
- [genomics](#) (199)
- [american society](#) (194)
- [China](#) (180)
- [economics](#) (168)
- [psychometrics](#) (166)
- [innovation](#) (161)
- [photos](#) (160)
- [science](#) (159)
- [psychology](#) (147)
- [ai](#) (141)
- [travel](#) (136)
- [universities](#) (119)
- [machine learning](#) (116)
- [credit crisis](#) (115)
- [higher education](#) (114)
- [biology](#) (111)
- [human capital](#) (110)
- [iq](#) (103)
- [genetic engineering](#) (102)
- [startups](#) (100)
- [cognitive science](#) (97)
- [careers](#) (80)
- [autobiographical](#) (78)
- [credit crunch](#) (78)
- [statistics](#) (75)
- [elitism](#) (72)
- [evolution](#) (71)
- [gilded age](#) (71)
- [income inequality](#) (69)
- [quantum mechanics](#) (67)
- [political correctness](#) (65)
- [genius](#) (62)
- [politics](#) (62)
- [talks](#) (62)
- [books](#) (61)
- [caltech](#) (60)
- [social science](#) (60)
- [geopolitics](#) (59)
- [mma](#) (55)
- [history of science](#) (53)
- [sci fi](#) (51)
- [academia](#) (50)
- [bounded rationality](#) (50)
- [harvard](#) (50)
- [kids](#) (50)
- [silicon valley](#) (50)
- [podcasts](#) (49)
- [bgi](#) (48)
- [education](#) (48)
- [mathematics](#) (47)

- [MSU](#) (46)
- [cdo](#) (45)
- [realpolitik](#) (44)
- [derivatives](#) (42)
- [behavioral economics](#) (39)
- [jiujitsu](#) (39)
- [history](#) (38)
- [intellectual history](#) (38)
- [literature](#) (38)
- [neuroscience](#) (37)
- [physical training](#) (37)
- [bubbles](#) (36)
- [mortgages](#) (36)
- [ufc](#) (36)
- [computing](#) (35)
- [film](#) (35)
- [bjj](#) (34)
- [expert prediction](#) (32)
- [google](#) (32)
- [hedge funds](#) (32)
- [foo camp](#) (30)
- [many worlds](#) (30)
- [race relations](#) (30)
- [affirmative action](#) (29)
- [black holes](#) (29)
- [economic history](#) (29)
- [quants](#) (29)
- [video](#) (29)
- [efficient markets](#) (28)
- [movies](#) (28)
- [feynman](#) (27)
- [nuclear weapons](#) (27)
- [sports](#) (27)
- [von Neumann](#) (27)
- [biotech](#) (26)
- [music](#) (26)
- [entrepreneurs](#) (25)
- [housing](#) (25)
- [obama](#) (25)
- [security](#) (25)
- [singularity](#) (25)
- [subprime](#) (25)
- [taiwan](#) (24)
- [berkeley](#) (23)
- [conferences](#) (22)
- [meritocracy](#) (22)
- [ultimate fighting](#) (22)
- [athletics](#) (21)
- [wall street](#) (21)
- [cds](#) (20)
- [internet](#) (20)
- [quantum field theory](#) (20)
- [scifoo](#) (20)
- [blogging](#) (19)
- [venture capital](#) (19)
- [gender](#) (18)
- [goldman sachs](#) (18)
- [treasury bailout](#) (17)
- [university of oregon](#) (17)
- [algorithms](#) (16)
- [cryptography](#) (16)
- [freeman dyson](#) (16)
- [new yorker](#) (16)

- [smpy](#) (16)
- [autism](#) (15)
- [cosmology](#) (15)
- [japan](#) (15)
- [oppenheimer](#) (15)
- [personality](#) (15)
- [happiness](#) (14)
- [height](#) (14)
- [wwii](#) (14)
- [chess](#) (13)
- [christmas](#) (13)
- [fitness](#) (13)
- [hedonic treadmill](#) (13)
- [probability](#) (13)
- [aspergers](#) (12)
- [dna](#) (12)
- [india](#) (12)
- [les grandes ecoles](#) (12)
- [malcolm gladwell](#) (12)
- [neanderthals](#) (12)
- [net worth](#) (12)
- [nobel prize](#) (12)
- [privacy](#) (12)
- [social networks](#) (12)
- [television](#) (12)
- [blade runner](#) (11)
- [entropy](#) (11)
- [geeks](#) (11)
- [nsa](#) (11)
- [philosophy of mind](#) (11)
- [string theory](#) (11)
- [ability](#) (10)
- [football](#) (10)
- [government](#) (10)
- [harvard society of fellows](#) (10)
- [italy](#) (10)
- [mutants](#) (10)
- [nerds](#) (10)
- [olympics](#) (10)
- [pseudoscience](#) (10)
- [Einstein](#) (9)
- [Go](#) (9)
- [complexity](#) (9)
- [crossfit](#) (9)
- [dating](#) (9)
- [encryption](#) (9)
- [eugene](#) (9)
- [flynn effect](#) (9)
- [research](#) (9)
- [turing test](#) (9)
- [alan turing](#) (8)
- [art](#) (8)
- [data mining](#) (8)
- [determinism](#) (8)
- [james salter](#) (8)
- [keynes](#) (8)
- [manhattan](#) (8)
- [pca](#) (8)
- [philip k. dick](#) (8)
- [pop culture](#) (8)
- [real estate](#) (8)
- [robot genius](#) (8)
- [usain bolt](#) (8)

- war (8)
- aig (7)
- alpha (7)
- ashkenazim (7)
- basketball (7)
- energy (7)
- france (7)
- free will (7)
- fx (7)
- game theory (7)
- games (7)
- hugh everett (7)
- poker (7)
- qcd (7)
- russia (7)
- success (7)
- Fermi problems (6)
- anthropic principle (6)
- bayes (6)
- class (6)
- climate change (6)
- democracy (6)
- environmentalism (6)
- nassim taleb (6)
- noam chomsky (6)
- paris (6)
- patents (6)
- prostitution (6)
- quantum computers (6)
- tail risk (6)
- teaching (6)
- volatility (6)
- academia sinica (5)
- bobby fischer (5)
- cold war (5)
- econ talk (5)
- godel (5)
- information theory (5)
- iraq war (5)
- kasparov (5)
- luck (5)
- new york times (5)
- nonlinearity (5)
- perimeter institute (5)
- rationality (5)
- renaissance technologies (5)
- sad but true (5)
- software development (5)
- vietnam war (5)
- warren buffet (5)
- 100m (4)
- iran (4)
- Poincare (4)
- bill gates (4)
- borges (4)
- cambridge uk (4)
- charles darwin (4)
- creativity (4)
- drones (4)
- fake alpha (4)
- feminism (4)
- global warming (4)
- hormones (4)

- humor (4)
- inequality (4)
- intellectual property (4)
- judo (4)
- kerviel (4)
- markets (4)
- microsoft (4)
- mixed martial arts (4)
- monsters (4)
- moore's law (4)
- solar energy (4)
- soros (4)
- trento (4)
- 200m (3)
- babies (3)
- brain drain (3)
- censorship (3)
- charlie munger (3)
- chet baker (3)
- correlation (3)
- demographics (3)
- ecosystems (3)
- equity risk premium (3)
- facebook (3)
- fannie (3)
- fst (3)
- intellectual ventures (3)
- jim simons (3)
- language (3)
- lee kwan yew (3)
- lewontin fallacy (3)
- lhc (3)
- magic (3)
- michael lewis (3)
- nathan myhrvold (3)
- neal stephenson (3)
- olympiads (3)
- path integrals (3)
- risk preference (3)
- search (3)
- sec (3)
- simulation (3)
- sivs (3)
- society generale (3)
- supercomputers (3)
- thailand (3)
- alibaba (2)
- assortative mating (2)
- bear stearns (2)
- bruce springsteen (2)
- charles babbage (2)
- cheng ting hsu (2)
- cloning (2)
- david mamet (2)
- digital books (2)
- donald mackenzie (2)
- eliot spitzer (2)
- empire (2)
- exchange rates (2)
- frauds (2)
- freddie (2)
- gaussian copula (2)
- heinlein (2)

- [industrial revolution](#) (2)
- [james watson](#) (2)
- [ltcm](#) (2)
- [mating](#) (2)
- [mba](#) (2)
- [mccain](#) (2)
- [mit](#) (2)
- [monkeys](#) (2)
- [national character](#) (2)
- [nicholas metropolis](#) (2)
- [no holds barred](#) (2)
- [offices](#) (2)
- [oligarchs](#) (2)
- [palin](#) (2)
- [population structure](#) (2)
- [prisoner's dilemma](#) (2)
- [skidelsky](#) (2)
- [socgen](#) (2)
- [sprints](#) (2)
- [systemic risk](#) (2)
- [twitter](#) (2)
- [ussr](#) (2)
- [variance](#) (2)
- [virtual reality](#) (2)
- [abx](#) (1)
- [anathem](#) (1)
- [andrew lo](#) (1)
- [antikythera mechanism](#) (1)
- [athens](#) (1)
- [atlas shrugged](#) (1)
- [ayn rand](#) (1)
- [bay area](#) (1)
- [beats](#) (1)
- [book search](#) (1)
- [bunnie huang](#) (1)
- [car dealers](#) (1)
- [carlos slim](#) (1)
- [catastrophe bonds](#) (1)
- [cdos](#) (1)
- [ces 2008](#) (1)
- [chance](#) (1)
- [children](#) (1)
- [cochran-harpending](#) (1)
- [computers](#) (1)
- [cpi](#) (1)
- [david x. li](#) (1)
- [dick cavett](#) (1)
- [dolomites](#) (1)
- [drugs](#) (1)
- [dune](#) (1)
- [eharmony](#) (1)
- [epidemics](#) (1)
- [escorts](#) (1)
- [faces](#) (1)
- [fads](#) (1)
- [favorite posts](#) (1)
- [fiber optic cable](#) (1)
- [francis crick](#) (1)
- [gary brecher](#) (1)
- [gizmos](#) (1)
- [greece](#) (1)
- [greenspan](#) (1)
- [hypocrisy](#) (1)

- [igon value](#) (1)
- [iit](#) (1)
- [inflation](#) (1)
- [information asymmetry](#) (1)
- [iphone](#) (1)
- [jack kerouac](#) (1)
- [jaynes](#) (1)
- [jfk](#) (1)
- [john dolan](#) (1)
- [john kerry](#) (1)
- [john paulson](#) (1)
- [john searle](#) (1)
- [john tierney](#) (1)
- [jonathan littell](#) (1)
- [las vegas](#) (1)
- [lawyers](#) (1)
- [lehman auction](#) (1)
- [les bienveillantes](#) (1)
- [lowell wood](#) (1)
- [lse](#) (1)
- [mcgeorge bundy](#) (1)
- [mexico](#) (1)
- [michael jackson](#) (1)
- [mickey rourke](#) (1)
- [migration](#) (1)
- [money:tech](#) (1)
- [myron scholes](#) (1)
- [netwon institute](#) (1)
- [networks](#) (1)
- [newton institute](#) (1)
- [nfl](#) (1)
- [oliver stone](#) (1)
- [phil gramm](#) (1)
- [philanthropy](#) (1)
- [philip greenspun](#) (1)
- [portfolio theory](#) (1)
- [power laws](#) (1)
- [psychology](#) (1)
- [randomness](#) (1)
- [recession](#) (1)
- [sales](#) (1)
- [singapore](#) (1)
- [skype](#) (1)
- [standard deviation](#) (1)
- [star wars](#) (1)
- [starship troopers](#) (1)
- [students today](#) (1)
- [teleportation](#) (1)
- [tierney lab blog](#) (1)
- [tomonaga](#) (1)
- [tyler cowen](#) (1)
- [venice](#) (1)
- [violence](#) (1)
- [virtual meetings](#) (1)
- [war nerd](#) (1)
- [wealth effect](#) (1)

