



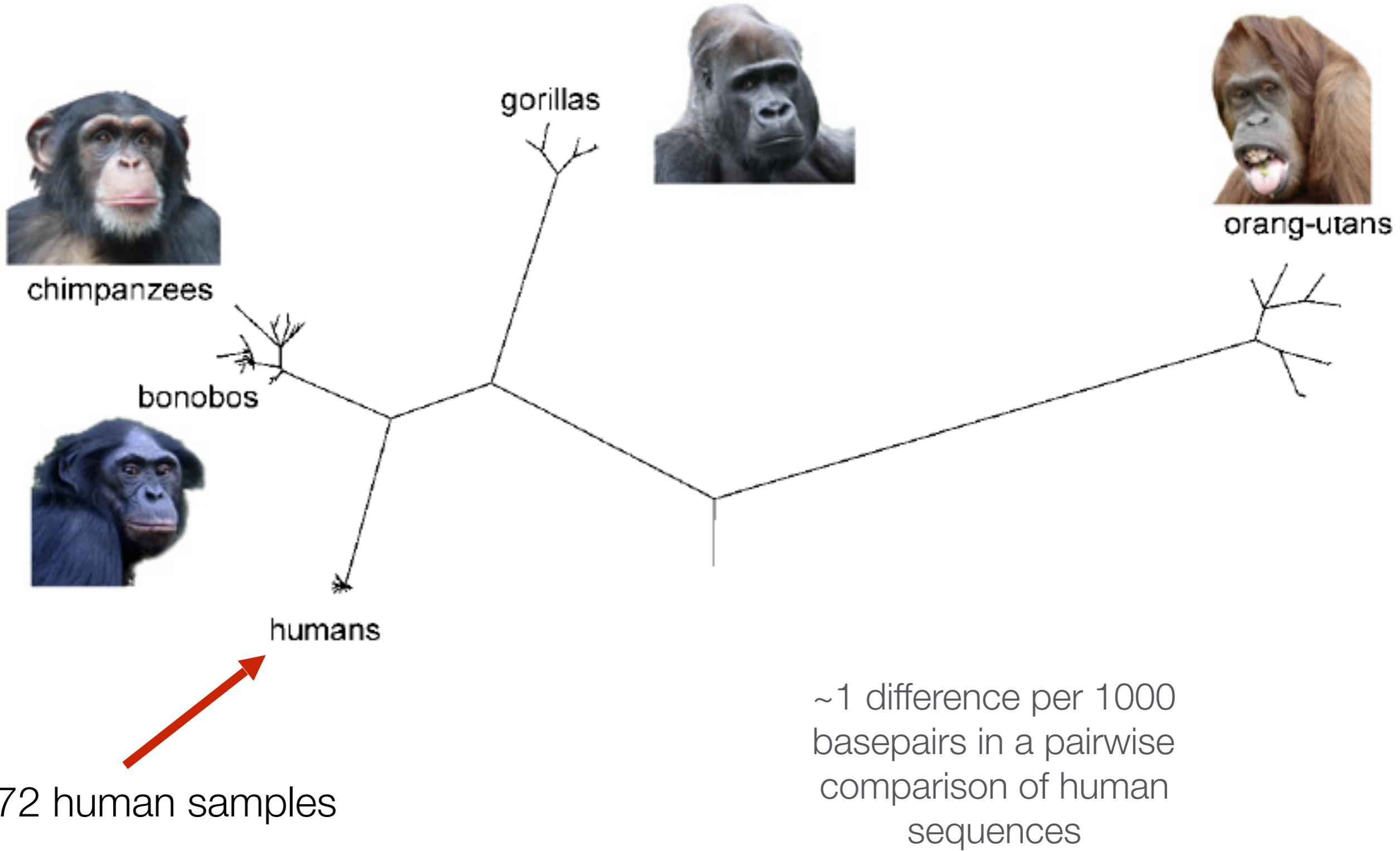
Genes and Geography: Understanding human genetic diversity

John Novembre

Department of Human Genetics

Department of Ecology and Evolutionary Biology

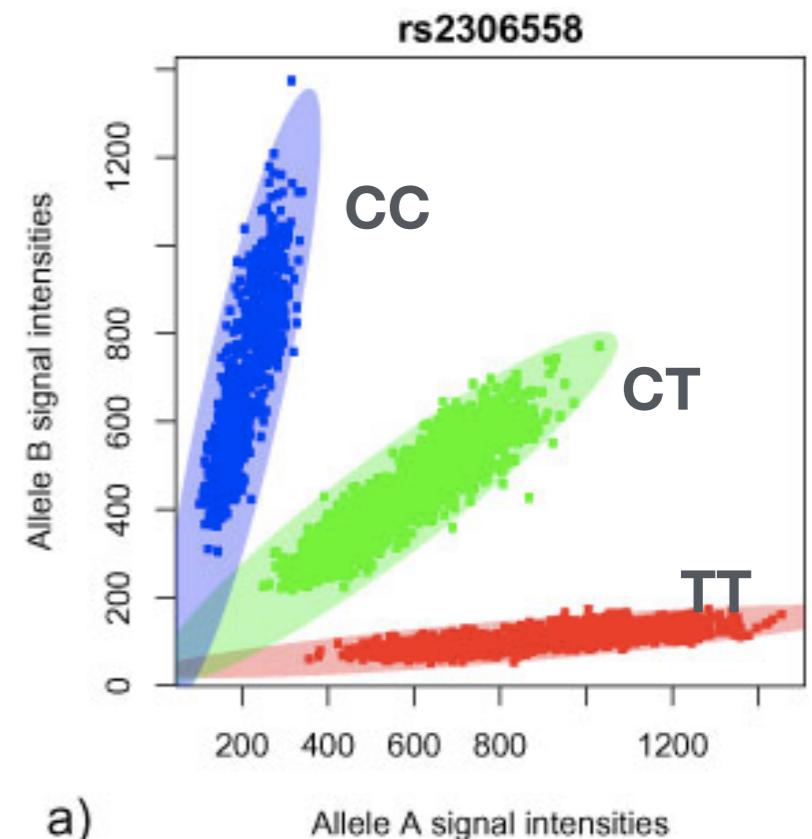
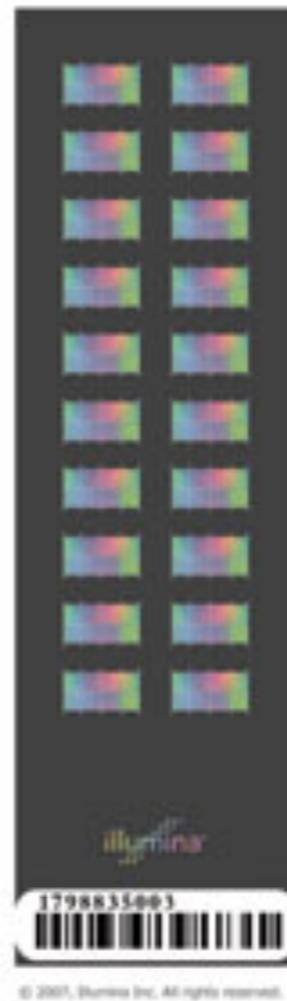
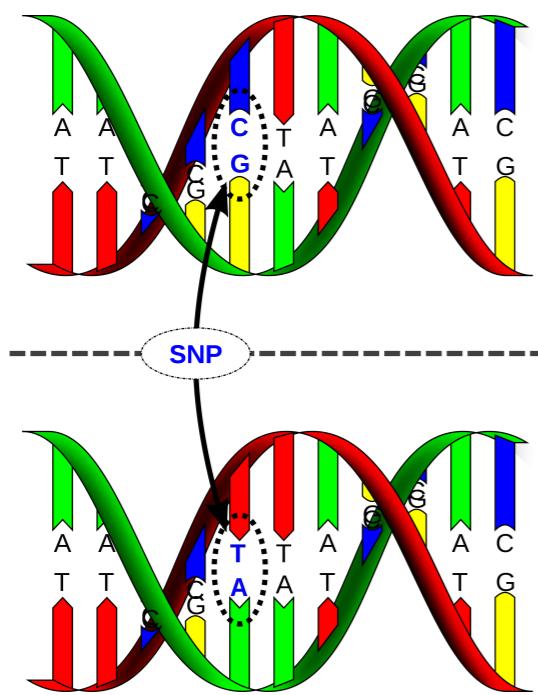
University of Chicago



The data deluge

SNP genotyping arrays: A revolution

SNP= single nucleotide polymorphism



Typically 250,000 - 1 million SNPs per array

An example of “SNP” data

expand by 50,000 to 8 million times

by 100-1000
times

1021	HGDP00001	0	0	1	-9	A	A	A	A	C	C	C	G	T	T	T	G	G	A	A	A	A	G	A
1022	HGDP00003	0	0	1	-9	A	A	A	A	C	C	C	G	G	T	T	G	G	A	A	A	A	G	A
1023	HGDP00005	0	0	1	-9	G	A	A	A	C	C	T	C	T	T	C	T	A	G	G	A	A	A	A
1024	HGDP00007	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1025	HGDP00009	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1026	HGDP00011	0	0	1	-9	A	A	A	A	C	C	C	G	T	T	T	G	G	A	A	A	A	G	A
1027	HGDP00013	0	0	1	-9	A	A	A	A	C	C	C	G	G	T	T	G	G	G	A	G	A	G	A
1028	HGDP00015	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1029	HGDP00017	0	0	1	-9	A	A	A	A	C	C	C	G	T	T	T	G	G	A	A	A	A	G	A
1030	HGDP00019	0	0	1	-9	A	A	A	A	C	C	C	G	T	T	T	G	G	A	A	A	A	A	A
1031	HGDP00021	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1032	HGDP00023	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1033	HGDP00025	0	0	1	-9	G	A	A	A	C	C	T	C	T	T	C	T	A	G	G	A	A	A	A
1034	HGDP00027	0	0	1	-9	A	A	A	A	C	C	C	G	T	T	T	G	G	A	A	A	A	G	A
1035	HGDP00029	0	0	1	-9	G	A	A	A	C	C	T	C	T	T	C	T	A	G	G	A	A	A	A
1036	HGDP00031	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1037	HGDP00033	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1038	HGDP00035	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1039	HGDP00037	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1040	HGDP00039	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1041	HGDP00041	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1042	HGDP00043	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1056	HGDP00070	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A
1057	HGDP00072	0	0	1	-9	A	A	A	A	C	C	C	T	T	T	T	G	G	A	A	A	A	A	A

3,000 individuals, 500,000 SNPs is ~100,000x this:

Genetic data can be overwhelming



"We finished the genome map, now we can't figure out how to fold it."

- Data visualization : “Data compression for people”
– Kris Sankaran

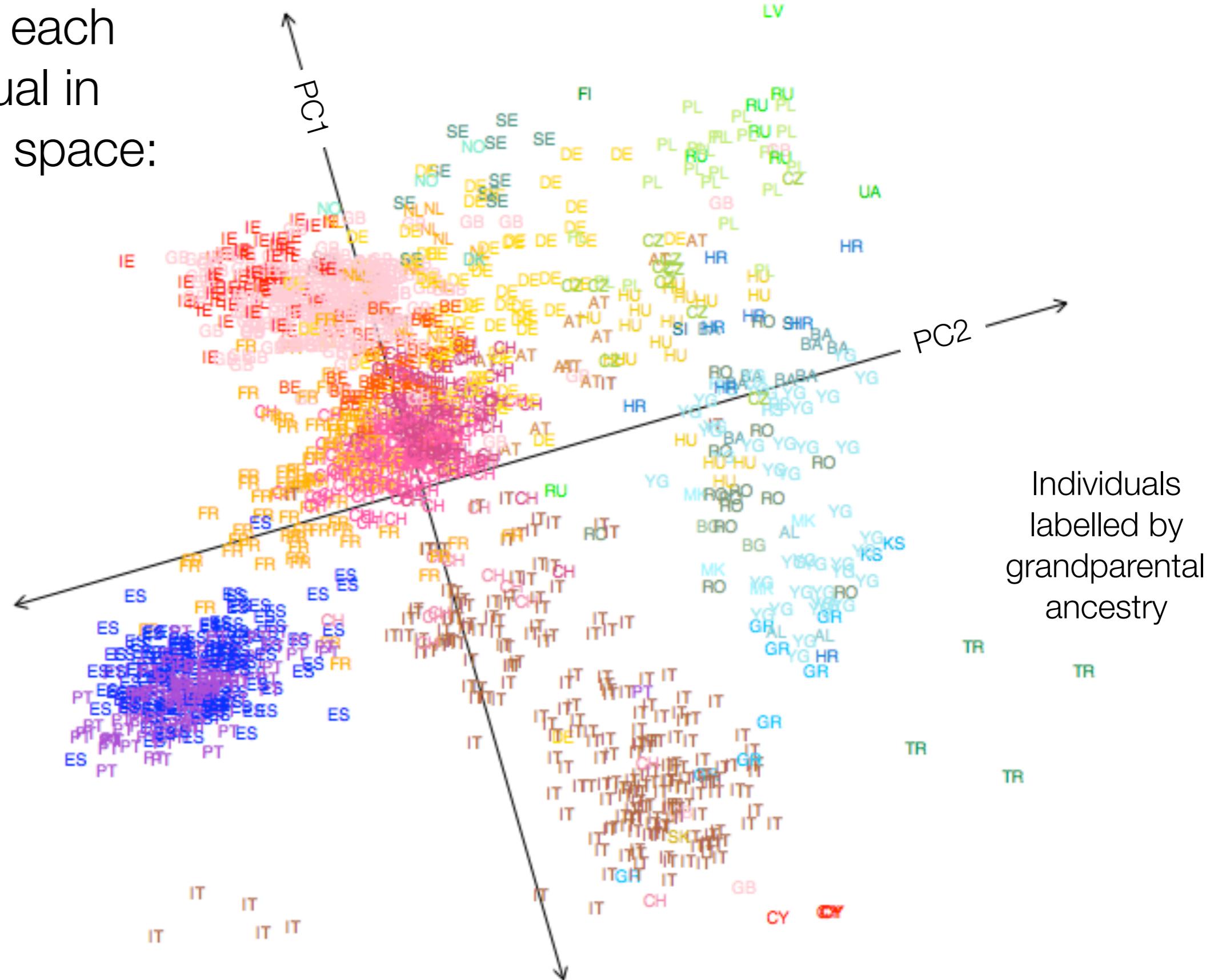


Image from EU Panorama “27 Europeans”

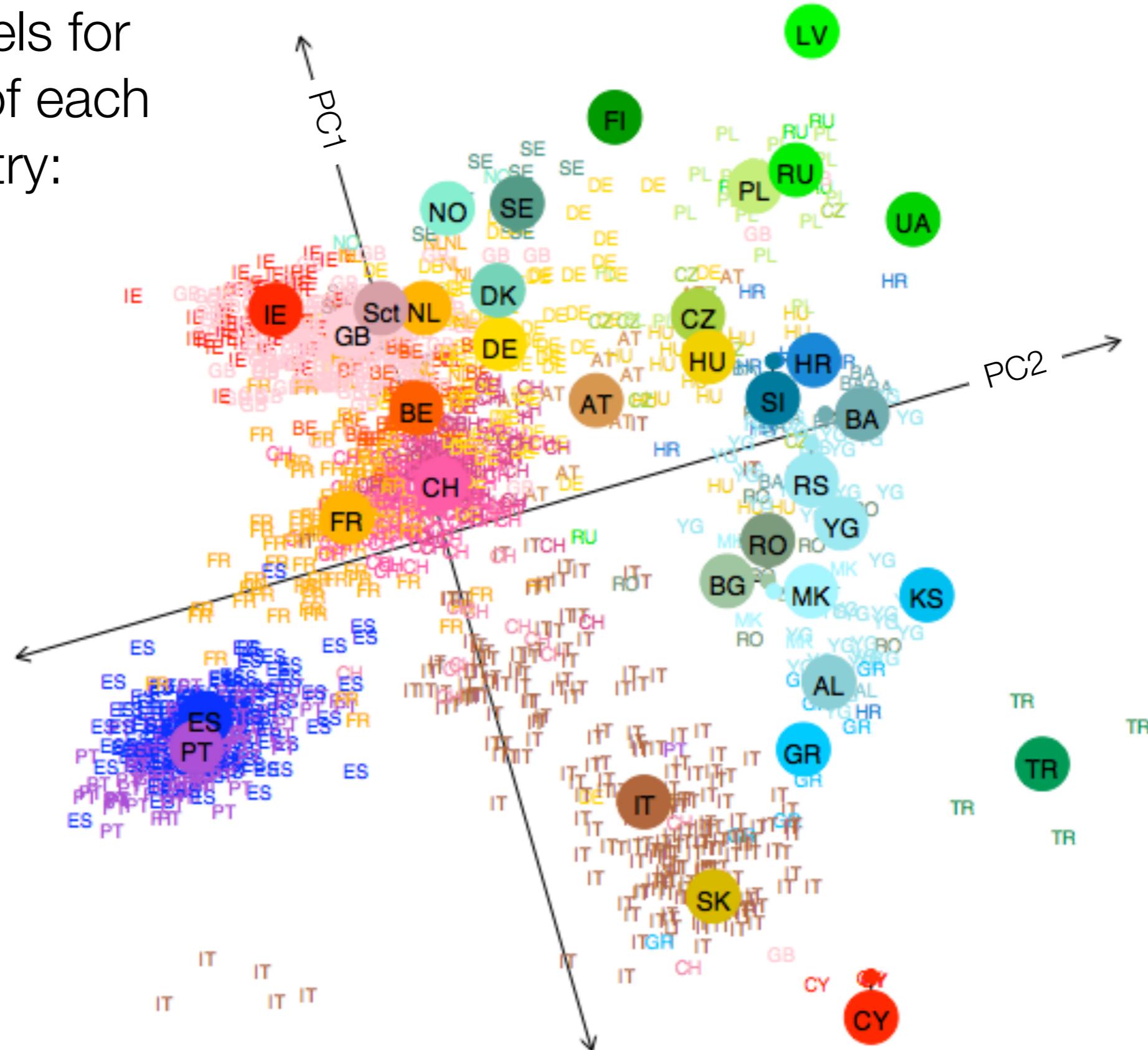
1400 individuals from 37 unique populations in Europe, mainly sampled in London and Lausanne

Analysis using 197,000 SNP loci

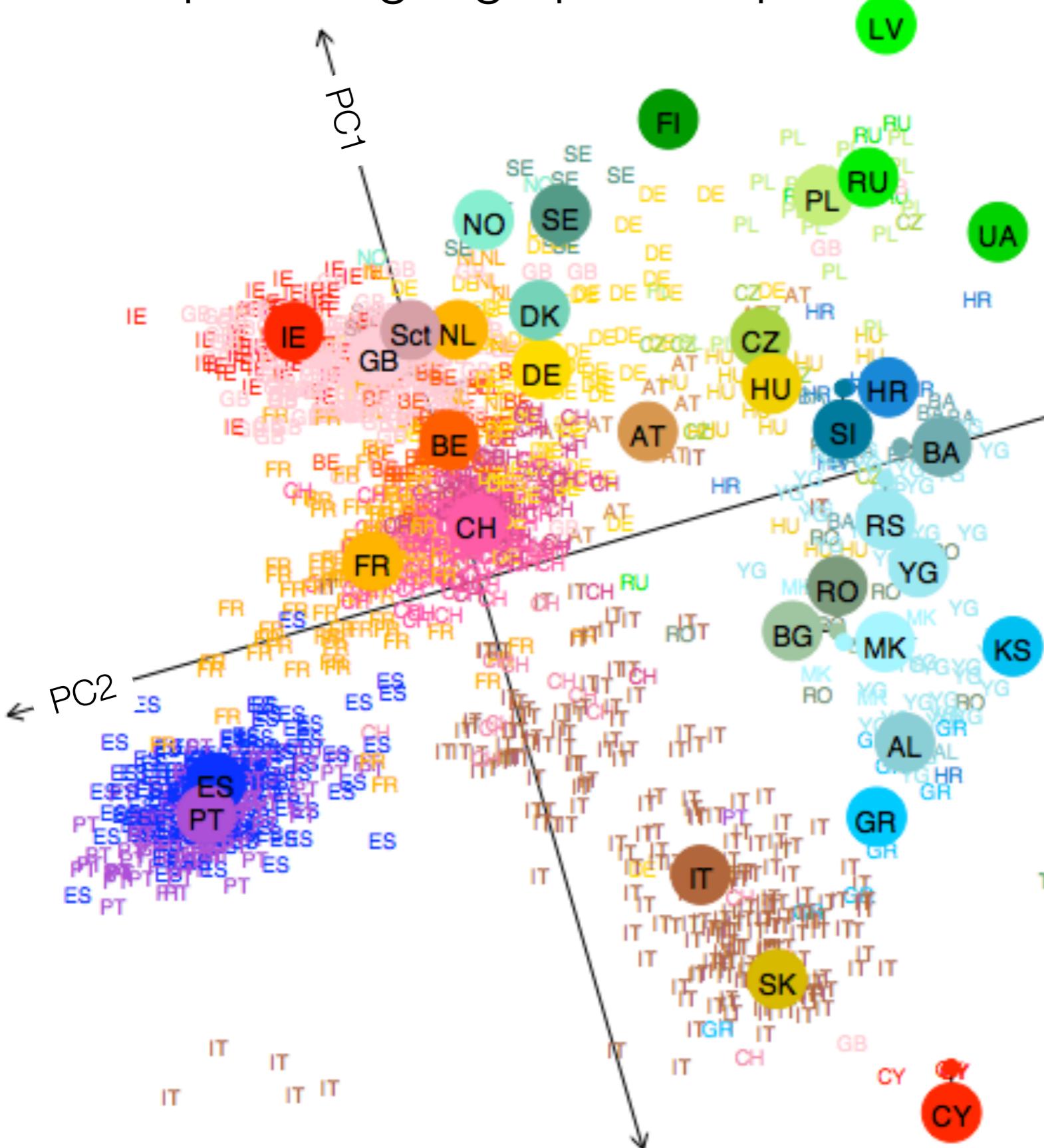
Plotting each individual in PC1-PC2 space:



Add labels for median of each country:



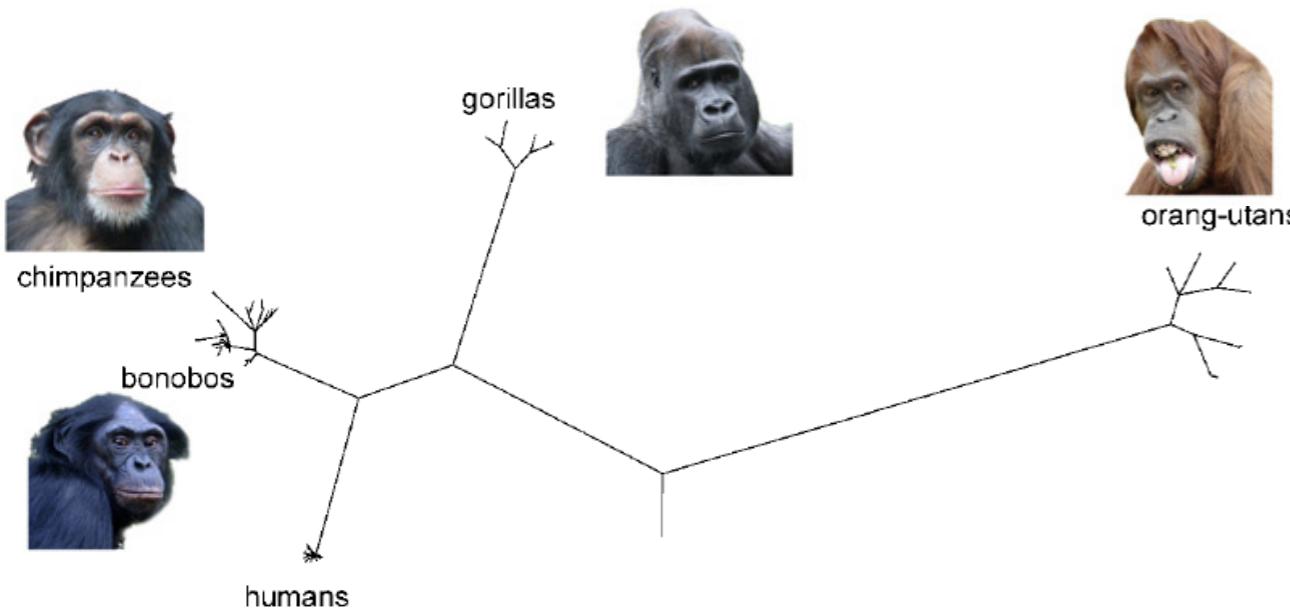
Compare to geographic map:



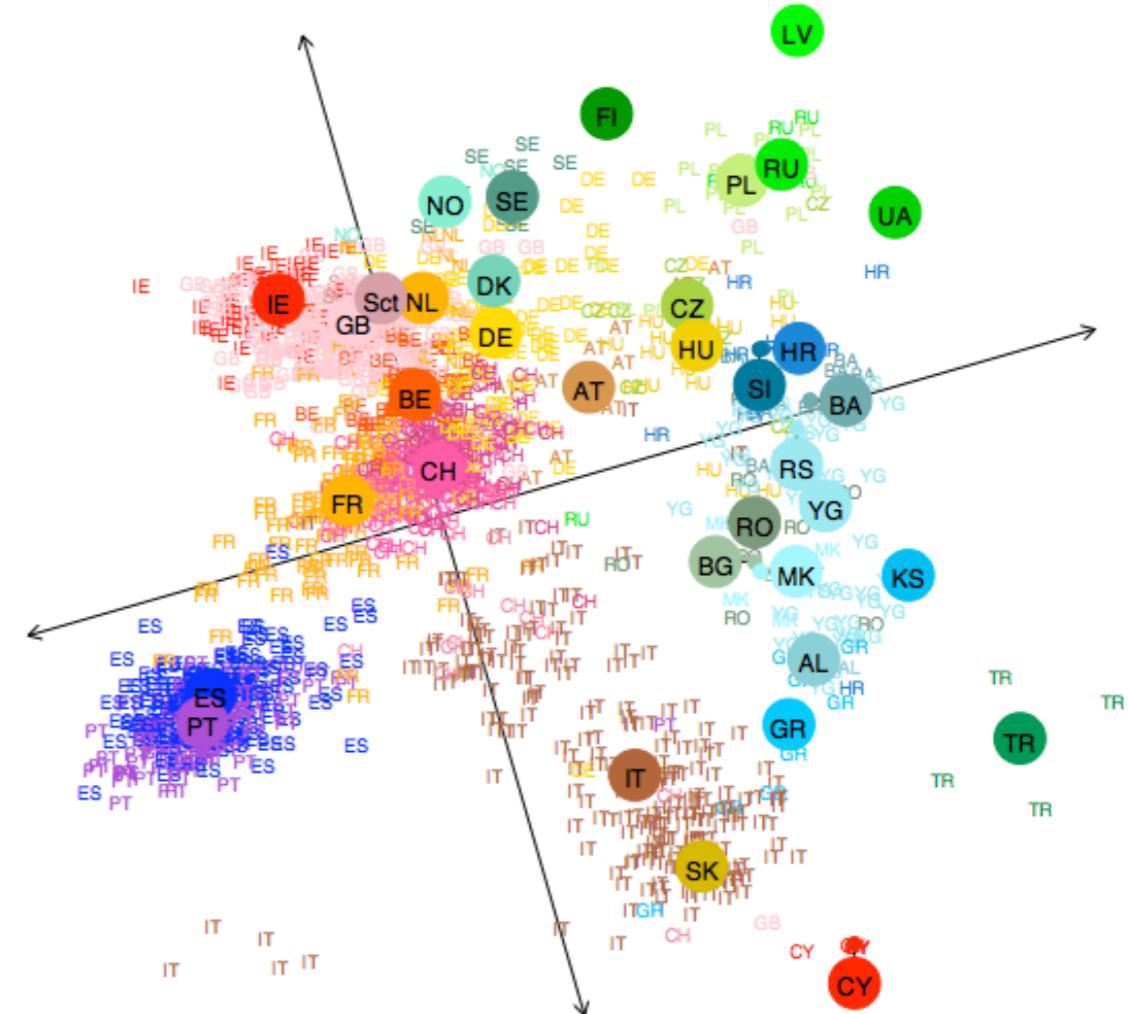
- PC1 prop variation = 0.3 %
- PC2 prop variation = 0.15%
- $F_{ST} = 0.004$

Novembre et al (Nature, 2008)

Both of these are true



Humans comparatively closely related
Sequence divergence is low: 1 per 1000 bp



Sub-continental origins of individuals reflected in genetics

The Kennewick Man: a pre-historic European in the Americas?

April, 1999



The results

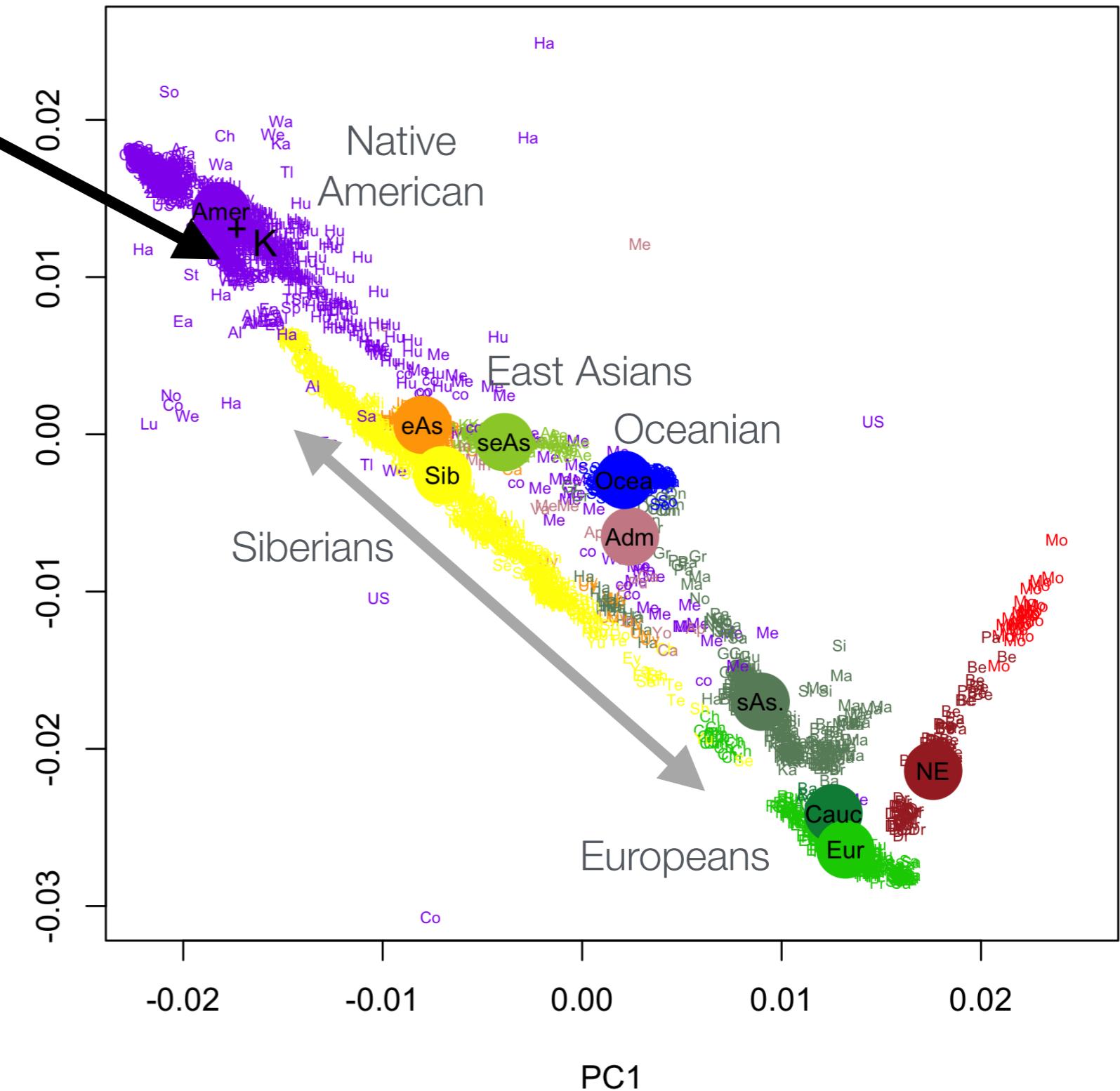
The Kennewick Man / The Ancient One

From Seattle Times, "It's Official: Kennewick Man is Native American":

Five tribes claiming Kennewick Man as a relative will work together to rebury him after the U.S. Army Corps of Engineers said Wednesday it has validated the skeleton is Native American.

Scientists at the University of Chicago this month documented they were able to independently validate last summer's scientific findings as to the skeleton's ancestry by at least three lines of evidence, said John Novembre, associate professor of human genetics at the University of Chicago, who led the review.

The validation was part of a federal process to allow repatriation of the skeleton. The team's finding clears the way for the next steps, in which potential claimants of the remains must document their cultural connection to the Ancient One, as tribes refer to the skeleton.



Reanalysis of Rasmussen et al (2015) data

Closing words: Writing code and writing prose



Thanks to the mentor team!

Power of big data: Methods pool weak signals across many variants

PC1: 1.63% explained

