

Population Genomics

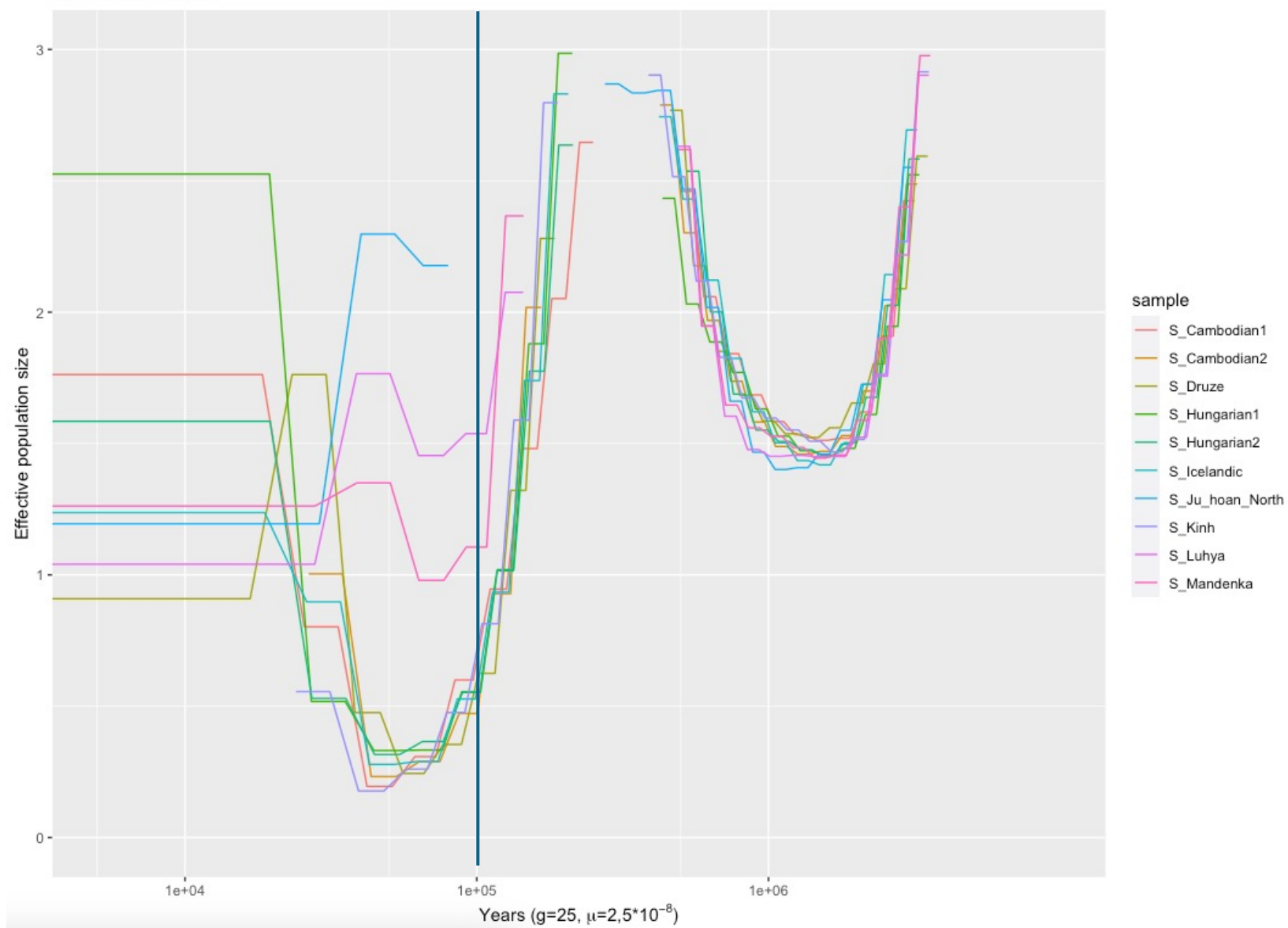
Week 4 Phasing

The subtle art of knowing haplotypes

Overview

- Wrap up from last week (PSMC results)
- Introduction to Phasing
- Exercise
- BREAK
- Exercise
- Questions
- If time, weird summary statistics used for scientific racism

Results of PSMC



Introduction to Phasing

Phasing, also known as haplotype estimation, refers to the process of statistical estimation of haplotypes from genotype data¹. It involves separating maternally and paternally inherited copies of each chromosome into haplotypes². This is a key step in many downstream analyses³.

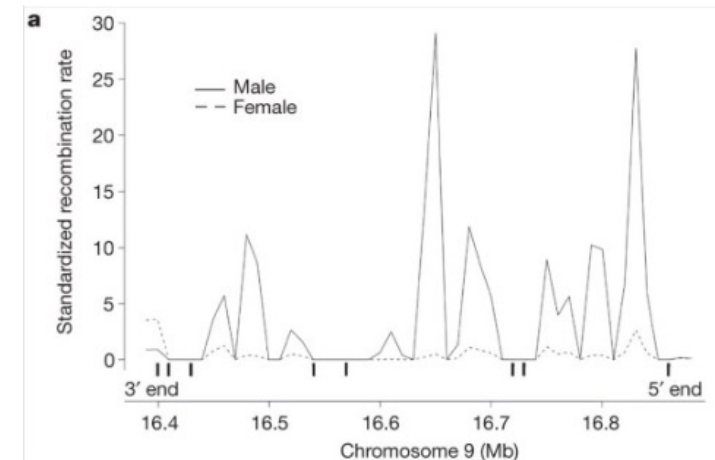
Maternal: ATTGG Paternal: ATCGT	Normal VCFs					
		A	T	T	G	G
				C		T
	Phased	A	T	T	G	G
		A	T	C	G	T

Reasons for Phasing

- More information
- Genetic variation
- Tracking of alleles

Rich Information

- Phased haplotype sequences are a key component in many population genetic analyses. Variation in haplotypes reflects the action of recombination, selection, and changes in population size



Genetic Variation

- Phasing allows us to get a complete picture of genetic variation². It helps in understanding gene expression patterns for genetic disease research

Red Mutations Significantly correlated with disease
But might only be so when they are placed in separate copies of the gene

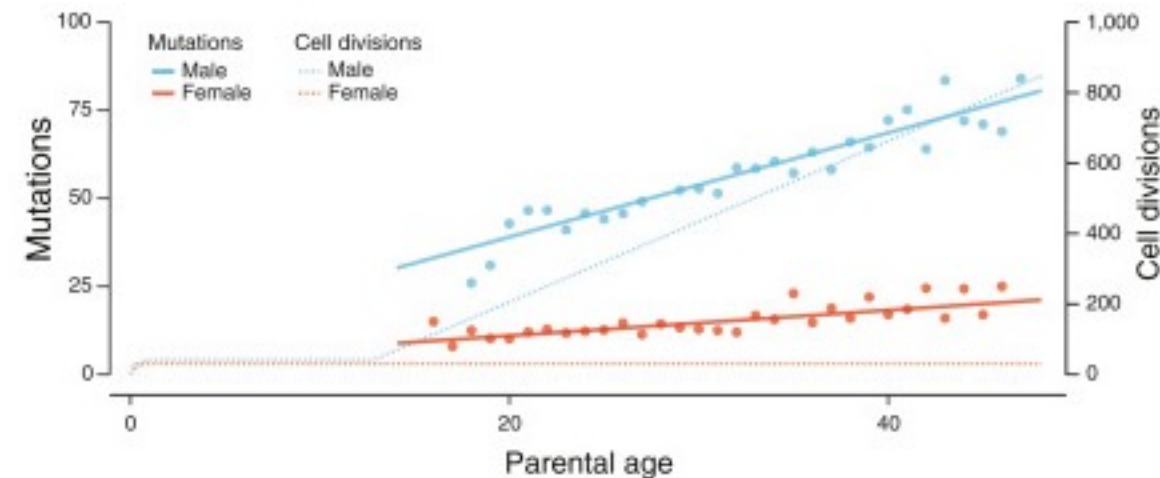
	A	T	T	G	G
			C		T
	A	T	T	G	G
	A	T	C	G	T
	A	T	T	G	T
	A	T	C	G	G

Track Alleles

Haplotype data contain more information than genotype data, as they make it possible to track single alleles or haplotype segments back in the pedigree

Germ Line mutation

Did the the mutation occur in the mother or father.



How does Phasing algorithms work

A	T	T	G	T
		C		
A	T	T	G	T
		C		

A	T	T	G	G
				T

A	T	T	G	G
				T

A	T	T	G	G
				T

A	T	C	C	G	T
A	T	T	C	G	G

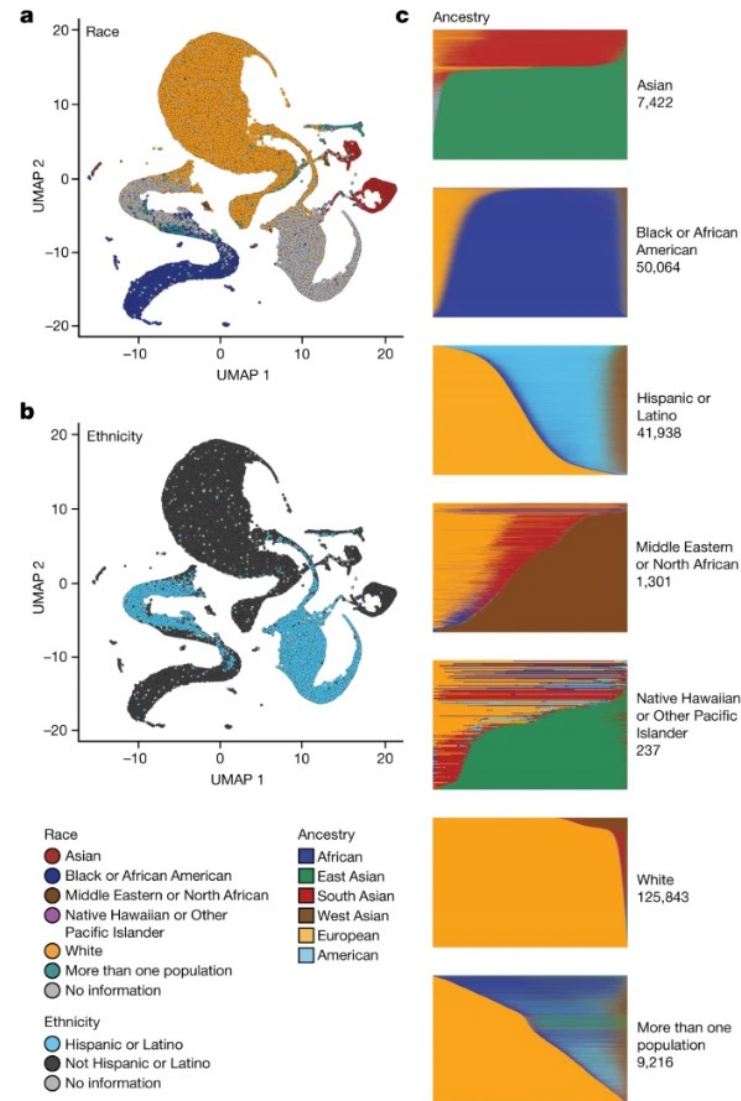
Exercise

A quick lesson in bad communication of genetics

When we do not communicate the results well, they can be used to push political agendas

What is the problem with this figure?

Fig. 2: Genetic ancestry in All of Us.



a, b, Uniform manifold approximation and projection (UMAP) representations of All of Us WGS PCA data with self-described race (**a**) and ethnicity (**b**) labels. **c,** Proportion of genetic ancestry per individual in six distinct and coherent ancestry groups defined by Human Genome Diversity Project and 1000 Genomes samples.



Graham Coop @Graham_Coop · Feb 20

...

The admixture plots go some way to showing that people & self-id race do not map discretely onto clusters. But here again the choice to equate ancestry & self-identified race is misleading (e.g. choice of clustering level & colour matching in legend) 3/n



Graham Coop @Graham_Coop · Feb 20

...

The authors state that these ancestry fractions match “six distinct and coherent ancestry groups” and that this “correspond[s] to participant self-identified race and ethnicity groups”. 4/n

PC data from All of Us samples and the HGDP and 1000 Genomes samples were used to compute individual participant genetic ancestry fractions for All of Us samples using the Rye program. Rye uses PC data to carry out rapid and accurate genetic ancestry inference on biobank-scale datasets⁴⁷. HGDP and 1000 Genomes reference samples were used to define a set of six distinct and coherent ancestry groups—African, East Asian, European, Middle Eastern, Latino/admixed American and South Asian—corresponding to participant self-identified race and ethnicity groups. Rye was run on the first 16 PCs, using the defined reference ancestry groups to assign ancestry group fractions to individual All of Us participant samples.

Race

- Asian
- Black or African American
- Middle Eastern or North African
- Native Hawaiian or Other Pacific Islander
- White
- More than one population
- No information

Ancestry

- African
- East Asian
- South Asian
- West Asian
- European
- American

Ethnicity

- Hispanic or Latino
- Not Hispanic or Latino
- No information



Graham Coop @Graham_Coop · Feb 20



There's a whole set of additional issues raised here. Take for example the "American genetic ancestry." which the authors say is "Latino/admixed american"... 5/n



1



1



9



5K



Graham Coop @Graham_Coop · Feb 20



Yet Latino/Hispanic are an ethnicities that usually includes people with a mixture of recent ancestors from Europe, Africa, and Indigenous American groups, and is not a distinct or separate grouping. 6/n



2



1



14



4.8K



