Population Genetics & Ancestry
Genetic Diversity

• Two processes increase genetic diversity in a population
  – **Mutation**: introduces novel variants into the population
  – **Recombination**: re-shuffles the existing patterns of variation (haplotypes)

• The fate of new mutations is affected by drift, selection, and population history
Linkage Disequilibrium (LD)

• Genetic variants are correlated because they occur on a particular haplotype background.

• In the absence of recombination this correlation (LD) would never be broken down and would extend a great distance along chromosomes.

• Recombination breaks down this correlation over many successive generations, leaving a narrower and narrower window of correlation.
Linkage Disequilibrium (LD)

de novo mutation

haplotypes
Linkage Disequilibrium (LD)
Under certain assumptions (neutral evolution, random mating, homogenous recombination), we can model exactly how far this correlation should extend.
Assortative Mating

• **Primary AM**: mates choose each other based on similarity

• **Social homogamy**: mates choose each other due to selected-environment proximity

• **Convergence**: mates become more similar to each other

• Traits that show AM
  – Education
  – Religious participation
  – Political attitudes
  – Height
  – Smoking & drinking
  – NOT personality or mental health
  – Time & place
LD in HapMap populations

<table>
<thead>
<tr>
<th>Panel</th>
<th>% $r^2 &gt; 0.8$</th>
<th>mean max $r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YRI</td>
<td>81</td>
<td>0.90</td>
</tr>
<tr>
<td>CEU</td>
<td>94</td>
<td>0.97</td>
</tr>
<tr>
<td>CHB+JPT</td>
<td>94</td>
<td>0.97</td>
</tr>
</tbody>
</table>
A not-too-terrible illustration, but population flow (movement of people) is NOT unidirectional
Hair Color

Clinal Distribution

Discontinuous Distribution

http://anthro.palomar.edu/vary/vary_1.htm
Principal Components Analysis

• PCA is applied to genotype data to describe continuous axes of genetic variation
• Each axis “explains” as much of the genetic variation in the data as possible, after accounting for the preceding components
• PCs used as covariates to statistically control for non-random differences in LD between people
http://ibg.colorado.edu/cdrom2017/abdellaoui/PCA_practical/PCA_practical.pdf

N = 4,441 Dutch
Multi-dimensional Scaling

- WTCCC
- Excluded samples
- YRI
- CEU
- CHB+JPT
QQ plots

No correction

PCA correction
(top 100 PCs)

Sawcer et al, Nature (2011)