Genetic basis of local adaptation

Marine Genomics Week 7

Get started downloading R packages!

The R packages needed this week take FOREVER to download. Let's get them started before we get on with the lecture.

```
install.packages("devtools")
install.packages(c("BiocManager","vcfR","pcadapt"))
# type "yes" when prompted
```

```
BiocManager::install("qvalue")

# When prompted, type "a" — this will take some time
library(devtools)
install_github("whitlock/OutFLANK")
```

From PCA
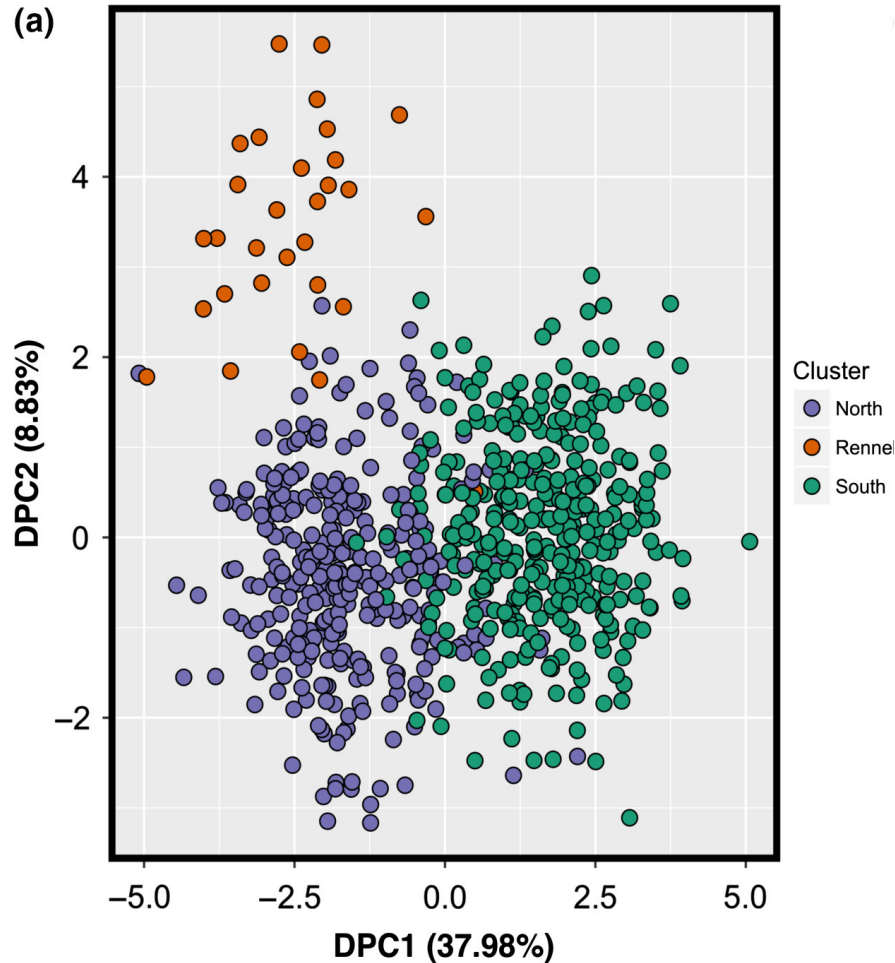
These populations look "genetically different" - but why?

- Neutral differences
- Selection

Learning objectives

- Define local adaptation and how it is inferred from genomic data

- Describe basic measures of genetic diversity and population differentiation

- Use standard analyses to identify 'outlier' SNPs as candidates for local adaptation

What is local adaptation?

Local adaptation – individuals have the highest fitness at their home site
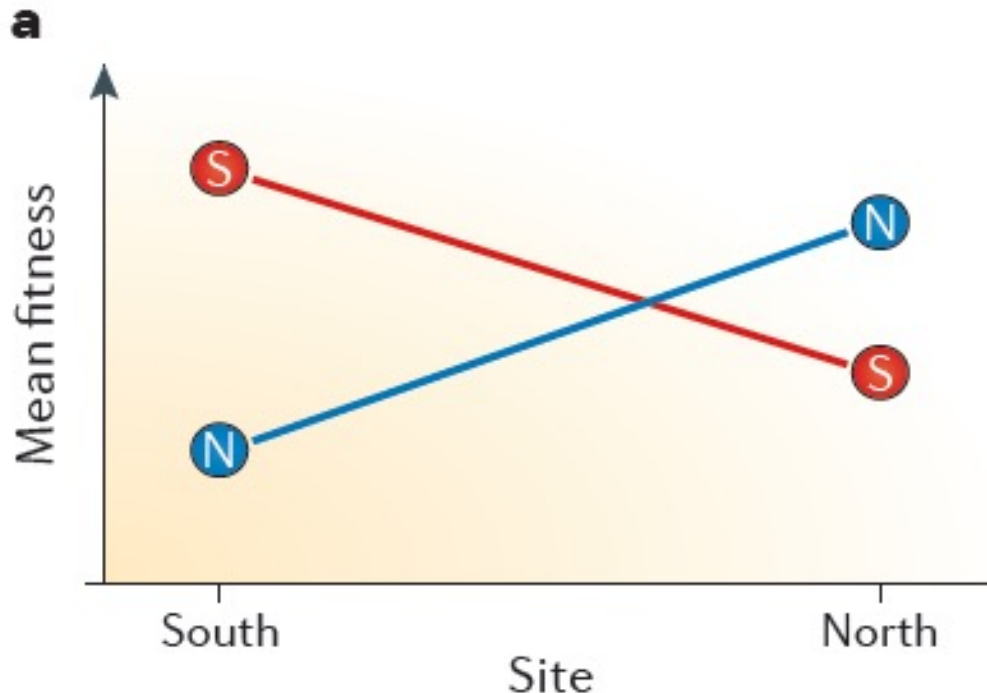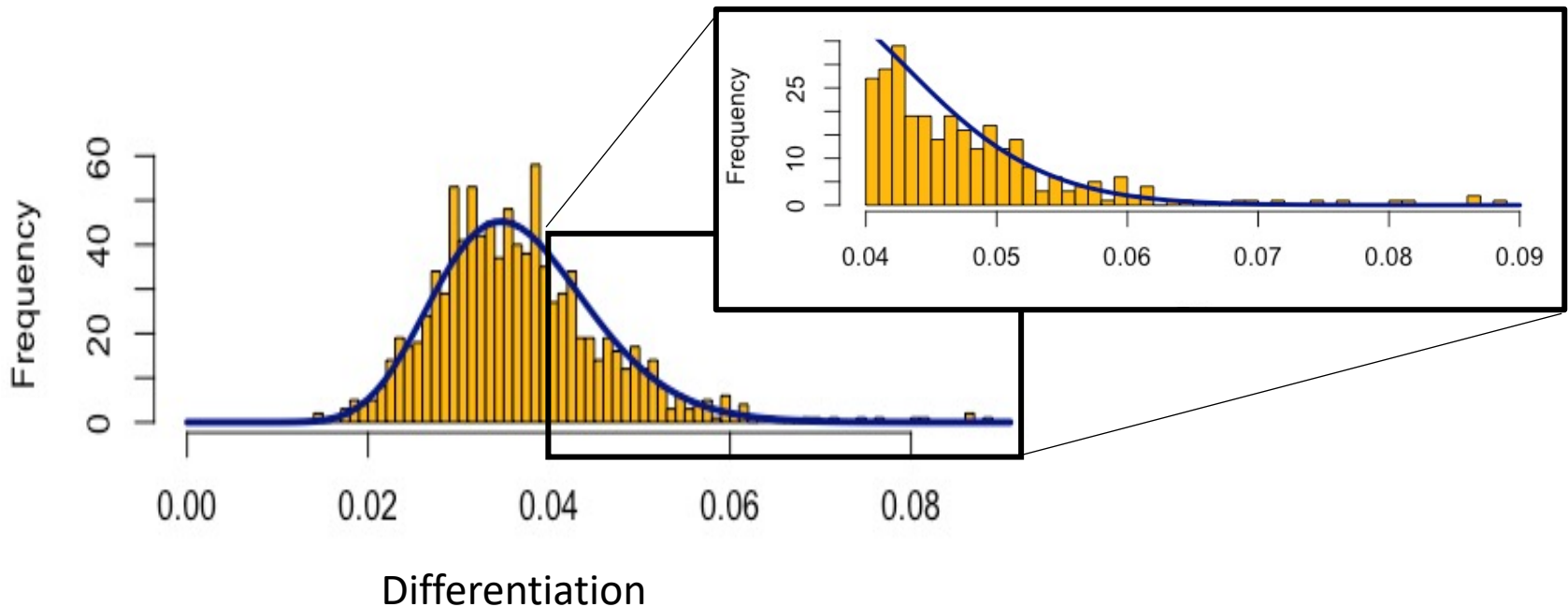

NOTE: We can't know this from genetic data!

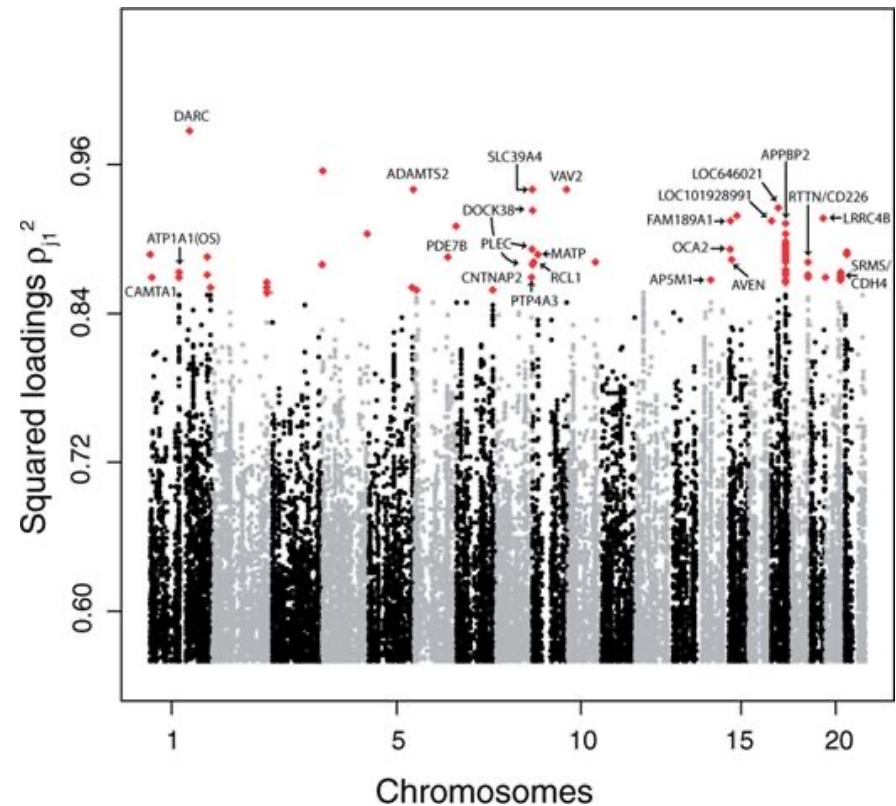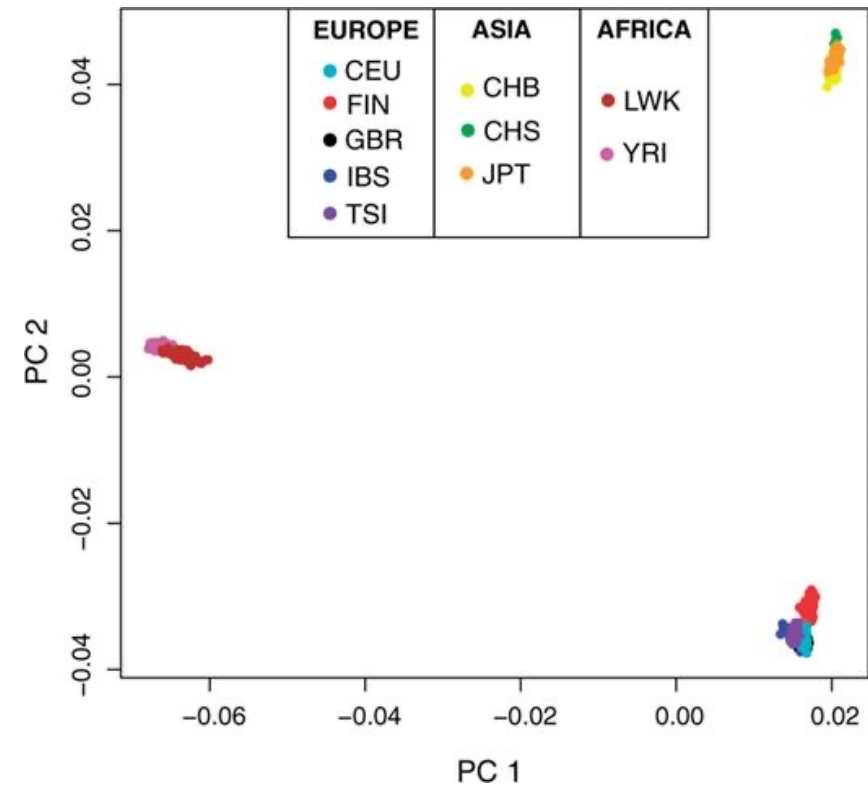# How do we identify the genomic basis of adaptation?

- Population genetic (genotype only)
  - Genetic differentiation outlier tests
- Phenotype mapping (genotype & phenotype)
  - Candidate genes
  - QTL mapping
  - Genome wide association studies (GWAS)
  - Bulk Segregant Analysis
  - Polygenic Scores
- Environment mapping
  - Genotype-Environment Association (GEA)

Outlier analysis

- Which loci are more differentiated than 'expected'
- Null distribution can be simulated for fit to 'neutral' data

Whitlock & Lotterhos 2015. *The American Naturalist*

PCA-associated SNPs

Duforet-Frebourg et al. *Molecular Biology and Evolution* 2015

Quantifying genetic diversity: Allele frequencies

- **Alleles**: B and b
  - *p*=frequency of B
    - 13/30=0.43
  - *q=frequency of b*
    - *17/30=0.57*
  - *p+q=1*

BB   bb   **Bb**

Bb   Bb   Bb   bb

bb   **Bb   BB   Bb**

Bb

Bb   bb

Bb

Quantifying genetic diversity
Genotype frequencies and HWE

- Two alleles with frequencies *p* and *q:*
  - *What is the probability that a randomly chosen individual will have genotype BB? Bb? bb?*

Genotype	BB	Bb	bb
Frequency	p^2	$2pq$	q^2

- $p^2+2pq+q^2=1$

Quantifying genetic diversity Genotype frequencies and HWE

p=0.43, q=0.57

Why are predicted and observed values different?

BB bb **Bb**

Bb **Bb** **Bb** bb

bb **Bb** **BB** **Bb**

Bb

Bb bb

Bb

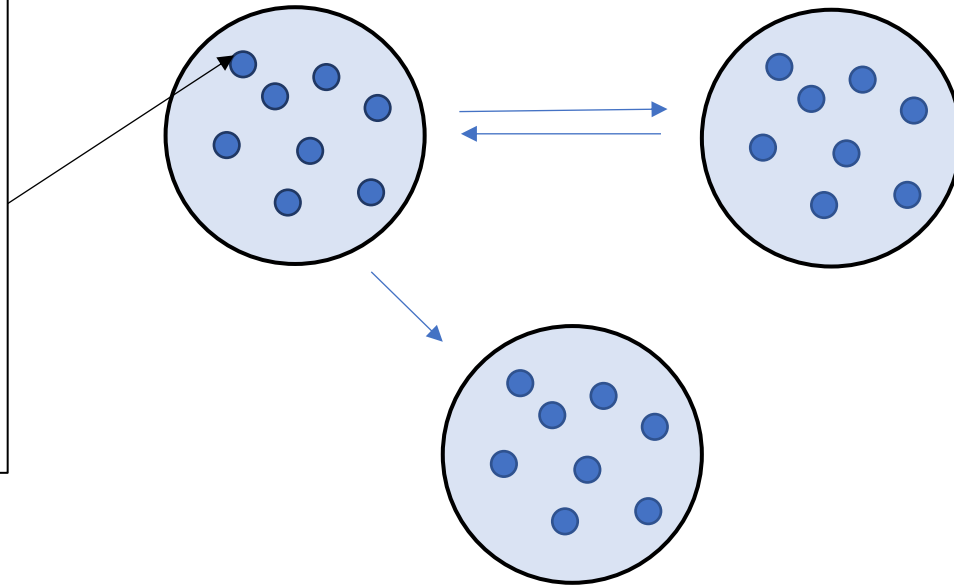| Genotype	BB	Bb	bb
Frequency	p^2	$2pq$	q^2

Quantifying population differentiation

- F-statistics are based on inbreeding: $F = \dfrac{H_{exp} - H_{obs}}{H_{exp}}$

Differentiation between subpopulations (relative to total)

$$F_{ST} = \dfrac{H_T - H_S}{H_T}$$

H_T=expected heterozygosity of total population
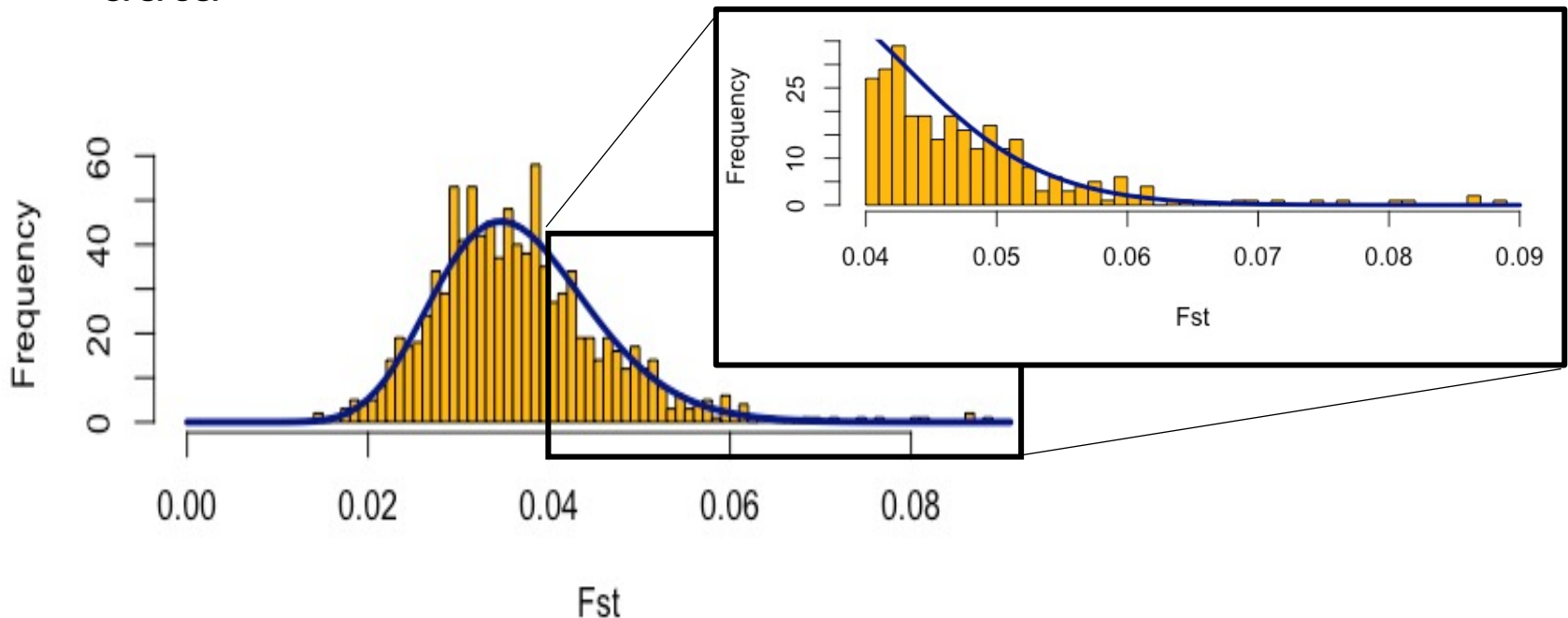H_S=expected subpopulation heterozygosity under HWE

F_{ST} = 0 means not differentiation
F_{ST} = 1 means no shared alleles (complete differentiation)
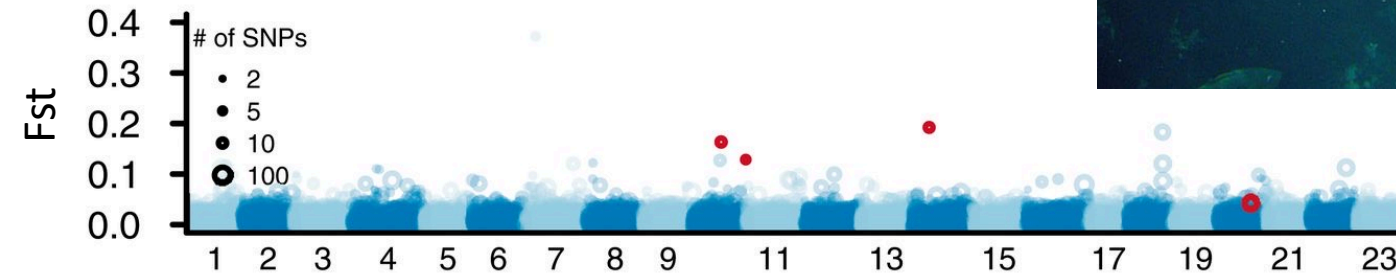
Outlier analysis

- Which loci are more differentiated than 'expected'
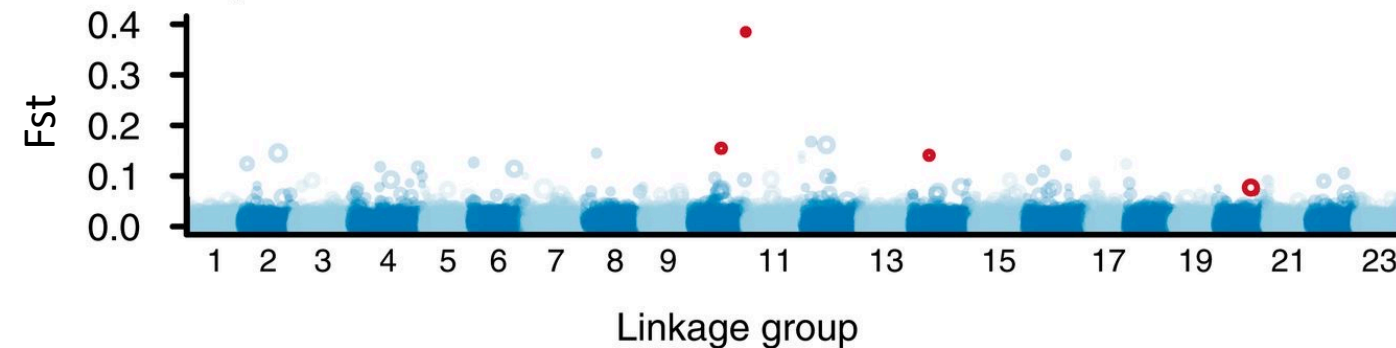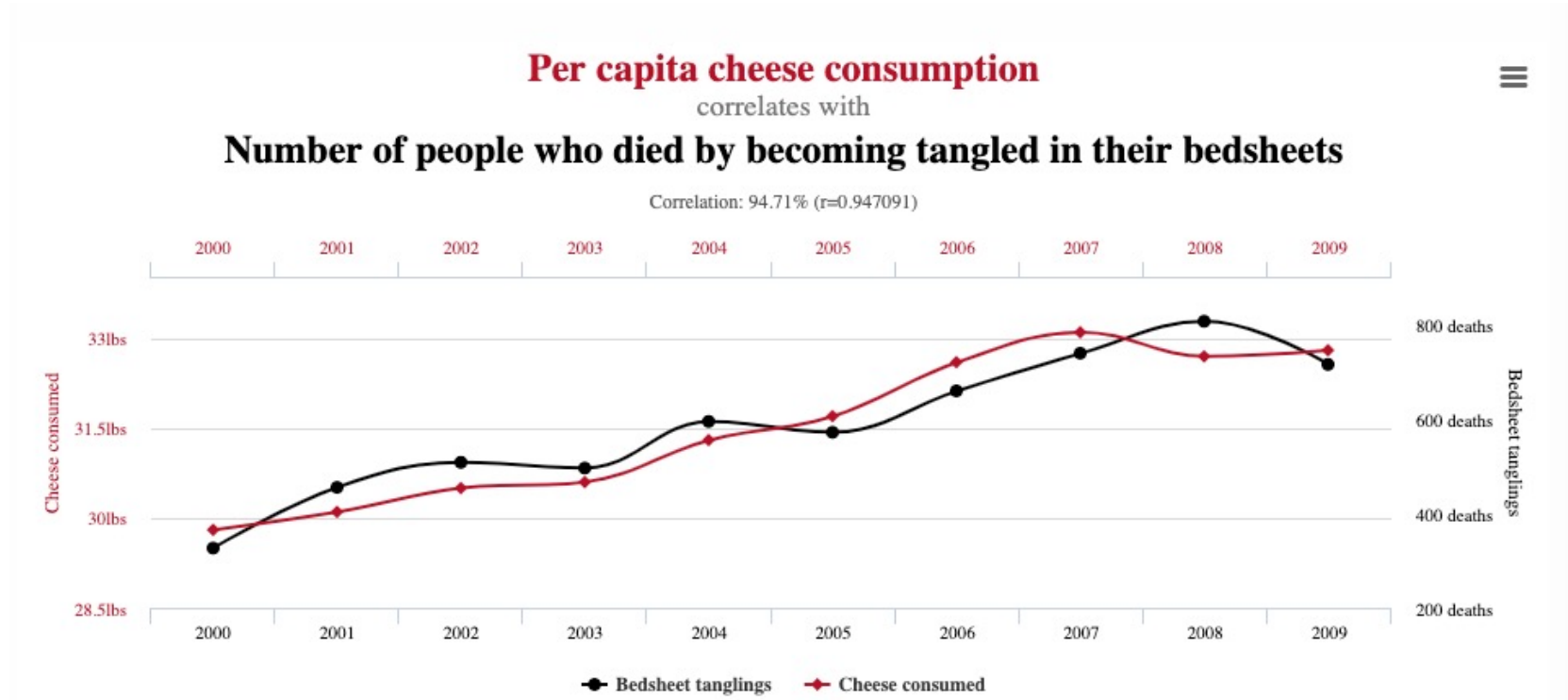- Null distribution can be simulated for fit to 'neutral' data

Whitlock & Lotterhos 2015. *The American Naturalist*

F$_{ST}$ outlier analysis

A Canada 1940–2013

B Norway 1907 –2014

Linkage group

Pinsky et al. *PNAS* 2021

Caution: The problem with tons of data

Per capita cheese consumption correlates with Number of people who died by becoming tangled in their bedsheets. Correlation: 94.71% (r=0.947091)

- When conducting >1000 tests, a
- Some results are likely to be *false positives*
- We use p-value corrections to attempt to minimize this problem

Today's exercise:

- What SNPs drive differentiation along PC axes?

- Which SNPs are most divergent between northern and southernmost sampling sites?