

Introducción a la Genómica

UNAL nov 2017

Alejandro Caceres
ISGlobal, Barcelona

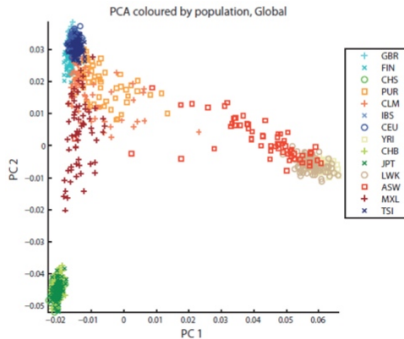
November 9, 2017

Metodos Multidimensionales

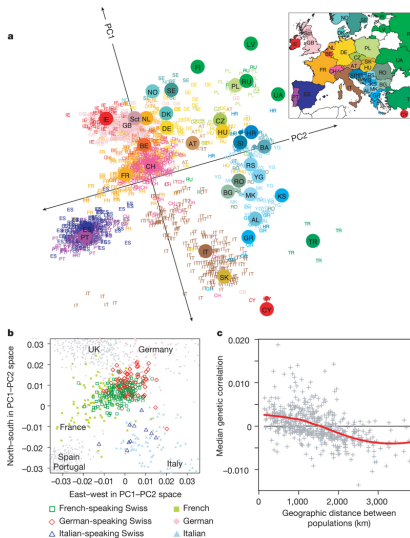
los que mas destacan son el análisis de componentes principales (PCA) y el escalamiento multidimensional (MDS)

- ▶ Miden la variabilidad genética en una muestra de individuos
- ▶ Determinan la distancia genética entre poblaciones
- ▶ Se pueden usar para determinar ancestría

Analisis de Componentes Principales



Analisis de Componentes Principales



Metodos Multidimensionales

Producen variables que son combinaciones lineales de los SNPs de tal forma que

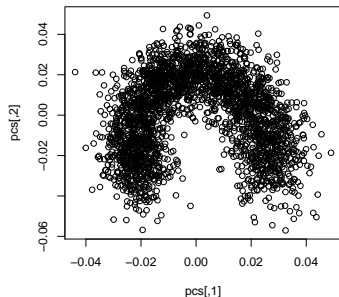
- ▶ explican la mayor variabilidad posible (PCA). SNPstats and PLINK
- ▶ explican la mayor distancia posible entre individuos (MDS). PLINK

El número de variables que se calculan depende de que tanto queramos explicar los datos.

snpStats

Veamos como se calcula PCA en SNPstats

```
library(snpStats)  
  
load("datos/NewsnpSNPstats.RData")  
xxmat <- xxt(NewsnpSNPstats)  
evv <- eigen(xxmat)  
pcs <- evv$vectors[,1:2]  
plot(pcs)
```



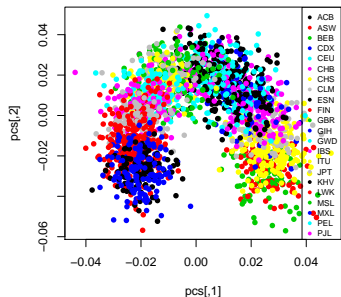
snpStats

Podemos ver que para esta región el PCA discrimina poblaciones

```
ids<-read.table("datos/20130606_g1k.ped", sep="\t", header=TRUE)

rownames(ids)<-ids$Individual.ID
pops<-ids[rownames(NewsnpSNPstats),]$Population

plot(pcs, col=as.numeric(pops), pch=16)
legend("topright", legend=levels(pops), pch=16, col=1:length(levels(pops))
```



PLINK

Debido a que los analisis multivariantes son constosos computacionalmente es mejor usar codigo compilado PLINK.

```
plink --bfile mydata --pca --out mydata  
plink --bfile mydata --mds-plot --out mydata --neighbour 1
```

Los resultados deben ser muy similares. `--neighbour 1 15` calcula las distancias del primer al 15 vecinos mas cercanos para determinar si un individuo es un outlier en terminos de ancestría

Inversiones

Los análisis multivariantes sirven para detectar inversiones polimórficas.

Ejercicio:

- ▶ Descargar los datos de los 1000 Genomas del cromosoma 17 (página de PLINK)
- ▶ cargarlos con snpStats
- ▶ seleccionar los genotipos de la región chr17:43670925-44329730
- ▶ hacer un PCA con los genotipos seleccionados