

# Algèbre linéaire, une introduction

C. Ambroise

Equipe Statistique et Génome  
UMR CNRS 8071



1 Introduction

2 Objectifs et plan de cours

1 Introduction

Objectifs et plan de cours

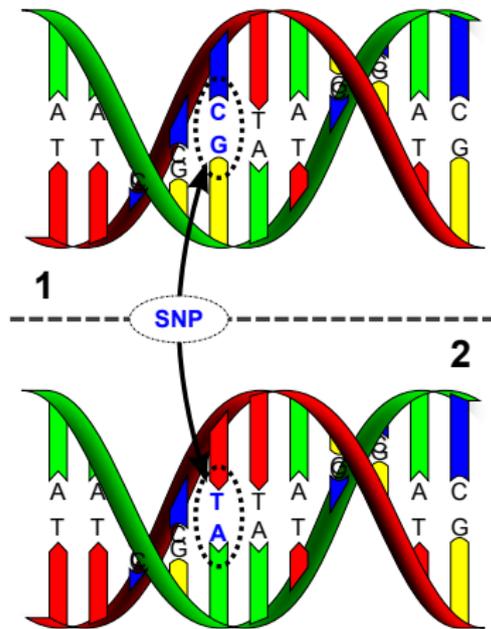
Dans l'article de Nature "Gene Mirror Geography Within Europe", Novembre et al., 2008,  
Les données

- 3000 individus
- décrits par 500, 000 SNP (Single Nucleotide Polymorphism)

peuvent être représenté par un tableau à

- 3000 lignes (disons 1 mètre de haut en écrivant petit)
- 50,000 colonnes (166 mètre de large!!!)

- 90 % de l'ensemble des variations génétiques humaines,
- SNP avec une fréquence allélique supérieure à un pour cent sont présents toutes les cent à trois cent paires de bases en moyenne dans le génome humain,
- où deux SNP sur trois substituent la cytosine avec la thymine.



SNP (wikipedia)

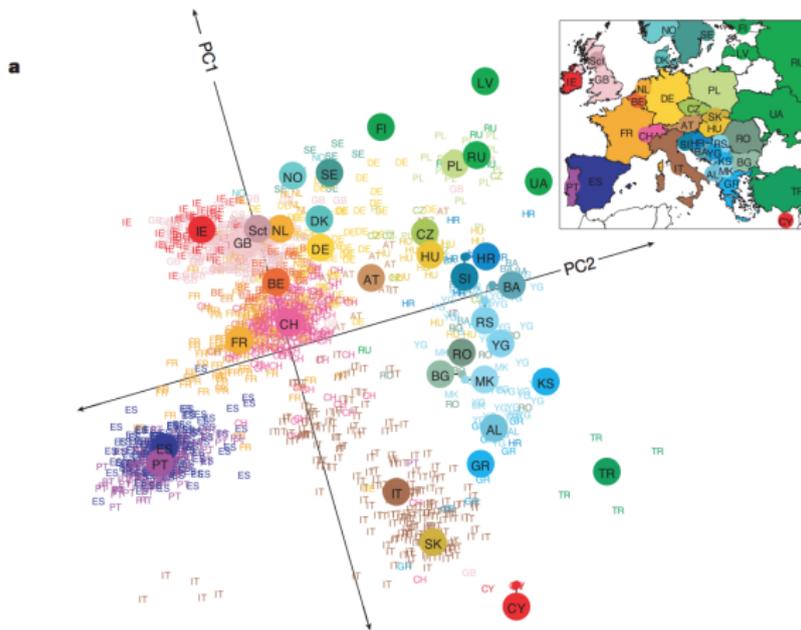
## Quatre configurations C/T

Mère	Père	
	C	T
C	0	1
T	1	2

- Comment résumer ce tableau ?
- Chaque individu appartient à un espace à 500,000 dimension

# Solution : l'analyse en composantes principales

Résumer 500,000 colonnes en 2 colonnes



ACP de la population européenne

● Introduction

② Objectifs et plan de cours

## Objectif

- Connaissances : Comprendre les notions mathématiques utilisées par l'ACP
- Compétences : Réaliser et interpréter une ACP

## Connaissance

Connaitres les concepts

- Matrices,
- Systèmes linéaires,
- Espaces vectoriels

## Compétences

Appliquer

- pivot de Gauss
- Déterminant
- Décomposition en valeurs singulières

Il faut comprendre un peu d'algèbre linéaire pour comprendre l'analyse des données ...

- Partie 1 : Algèbre linéaire appliquée
  - ① Systèmes linéaires, familles linéairement indépendantes, applications linéaires
  - ② Matrices
  - ③ Espaces vectoriels
  - ④ Valeurs propres et valeurs singulières
- Partie 2 : Analyse des données
  - ① Analyse en composantes principales

## Quelques points d'organisation

- Pour la partie “Algèbre” : COURS/TD
  - Il y aura 3 quizz courts en TD (un exercice proche du TD de la fois d'avant). Les 2 meilleurs notes seront gardées et compteront pour 25% de la note finale.
- Pour la partie “Analyse des données” : COURS et TP
  - Il y aura 1 TD à rendre qui comptera pour 25% de la note finale
- L'examen final compte pour 50% de la note.
- Attention : les TD et TP sont obligatoires