

# Associating Alzheimer disease with brain MRI and genomic data

Anh Vu Le





# Outline

- Introduction to genetics
- Genetic disorders
  - Alzheimer's disease
- Partial least squares
  - PCA intuition
  - Application on Alzheimer's data



# Gentle introduction to genetics

Each organisms is coded by its **DNA**

DNA is endless string of **letters A, C, T, G**

Human DNA is 3,000,000,000 letters long





# Gentle introduction to genetics

**99.9% shared** among all individuals

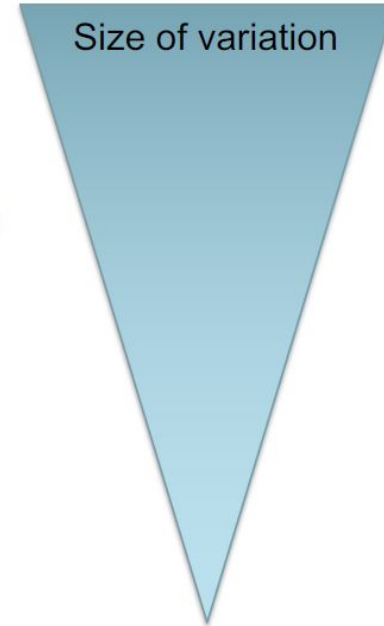
**3,000,000 letters** vary  
(1 floppy disk)





# Genetic disorders

- Duplication or loss of entire chromosomes
  - Trisomy 21, Turner syndrome
- Duplication or loss of parts of chromosomes
  - Deletion syndromes, e.g., 22q11.2
  - Duplication syndromes
  - Translocations (balanced, unbalanced)
  - Inversions
  - ...
- Short insertions or deletions (indels)
- Single Nucleotide Variations (SNVs)



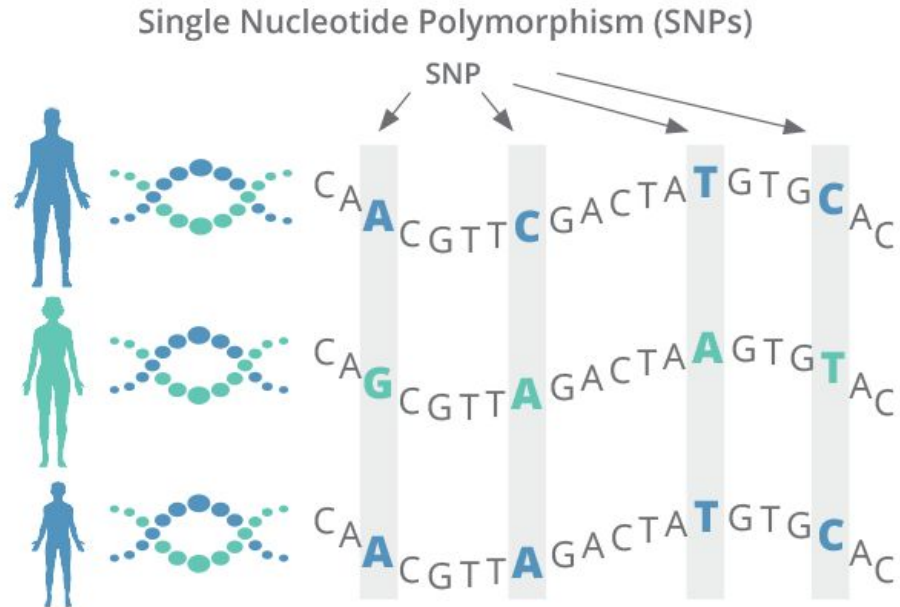


# Single nucleotide polymorphism

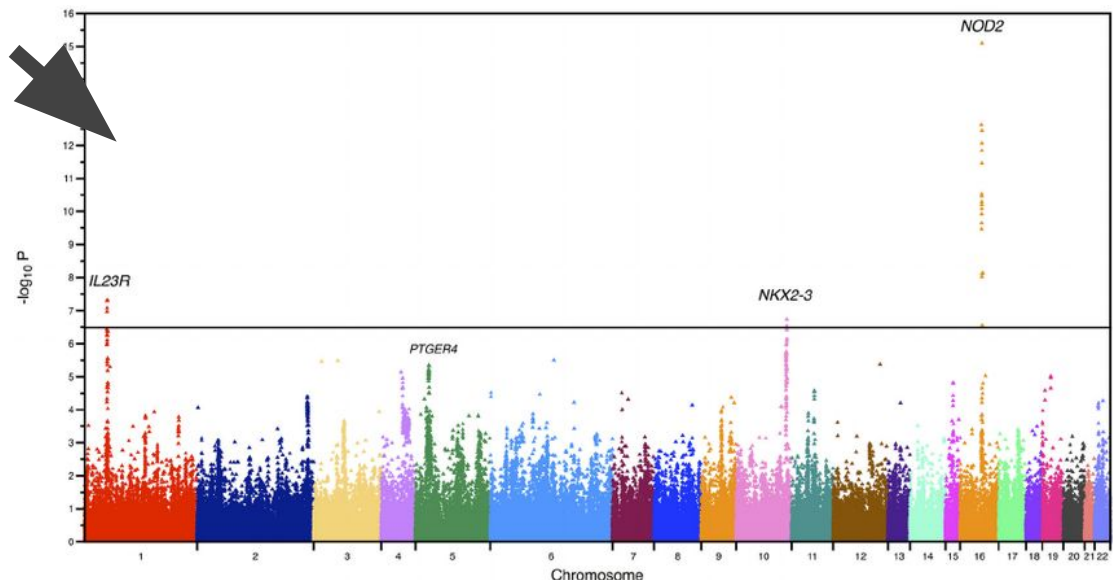
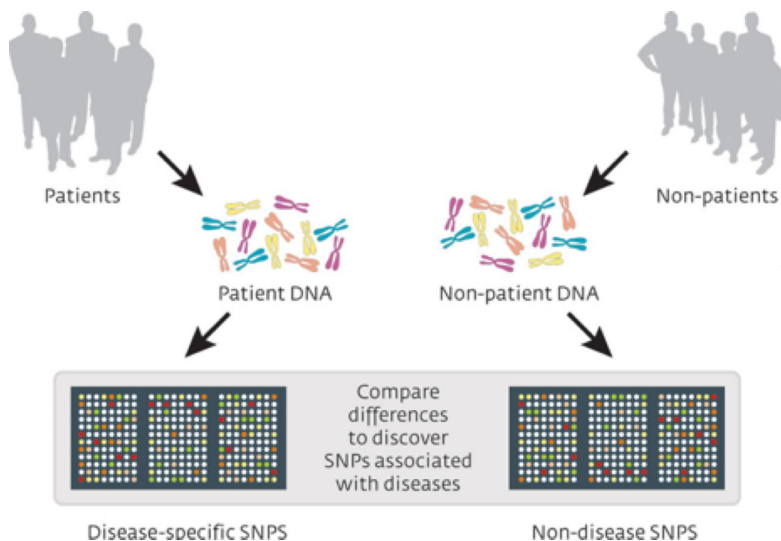
Mutation at a particular location occurring in >1% of population

Millions of location

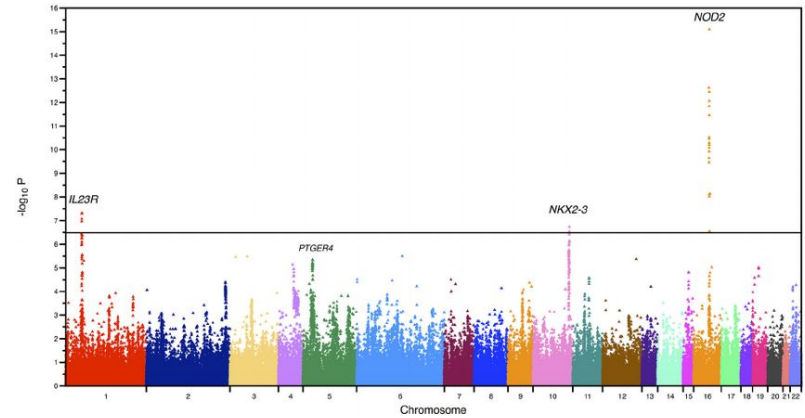
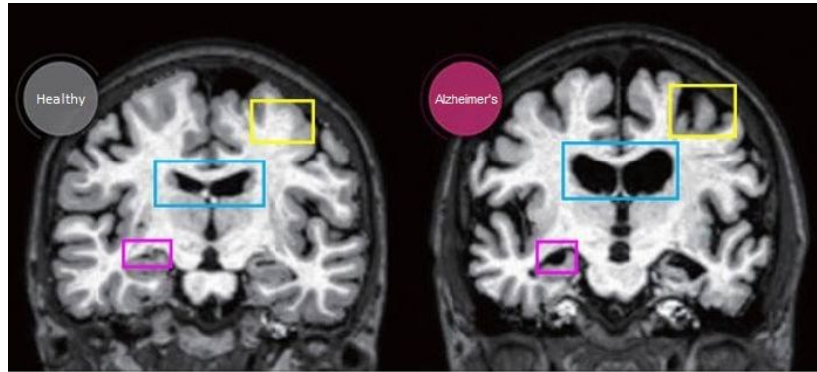
- Millions of locations
- + Locations known



# Genome wide association studies



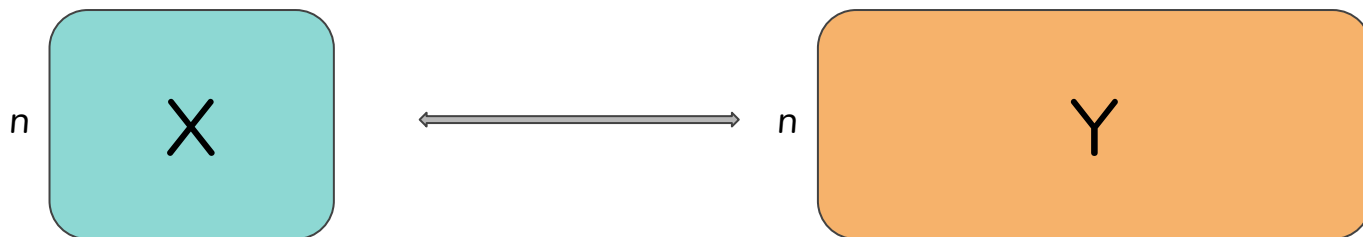
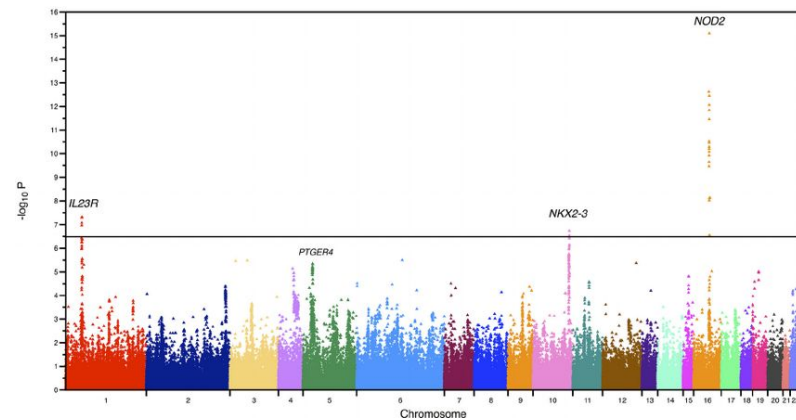
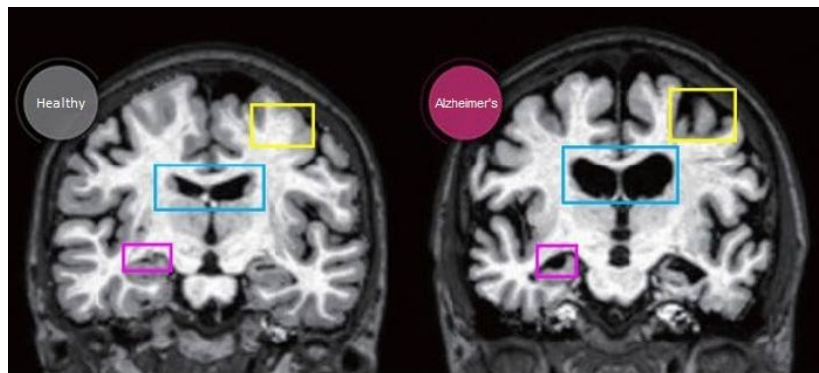
# Alzheimer's disease



Joint model



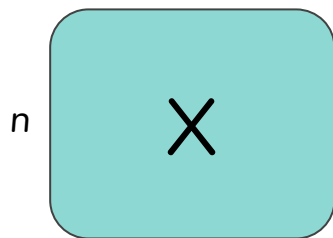
# Joining the data



# Partial least squares

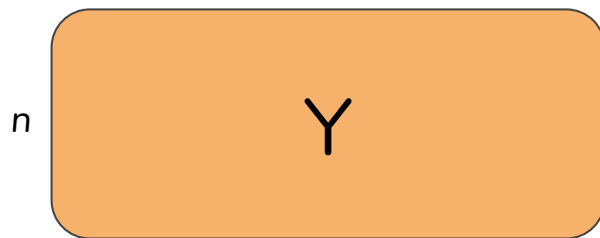
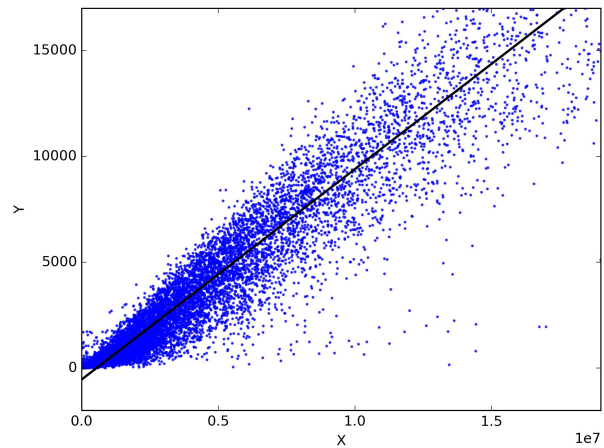
As if *PCA* and *linear regression* had a baby together...

# PCA and regression



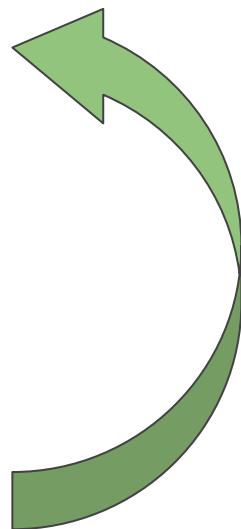
PCA

$n$



PCA

$n$



# From PCA to Partial Least Squares

## PCA iteratively

(1) take a vector  $\mathbf{x}_j$  from  $\mathbf{X}$  and call it  $\mathbf{t}_h$ :  $\mathbf{t}_h = \mathbf{x}_j$  (11)

(2) calculate  $\mathbf{p}'_h$ :  $\mathbf{p}'_h = \mathbf{t}'_h \mathbf{X} / \mathbf{t}'_h \mathbf{t}_h$  (12)

(3) normalize  $\mathbf{p}'_h$  to length 1:  $\mathbf{p}'_{h\text{new}} = \mathbf{p}'_{h\text{old}} / \|\mathbf{p}'_{h\text{old}}\|$  (13)

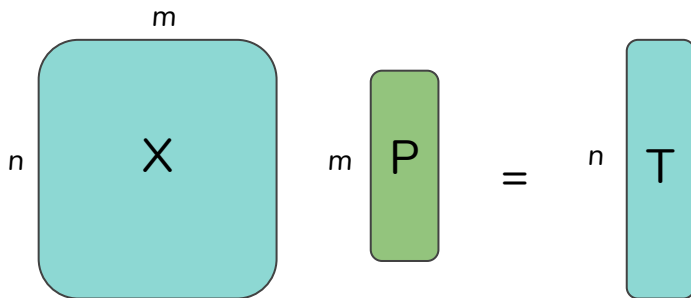
(4) calculate  $\mathbf{t}_h$ :  $\mathbf{t}_h = \mathbf{X} \mathbf{p}_h / \mathbf{p}'_h \mathbf{p}_h$  (14)

(5) compare the  $\mathbf{t}_h$  used in step 2 with that obtained in step 4. If they are the same, stop (the iteration has converged). If they still differ, go to step 2.

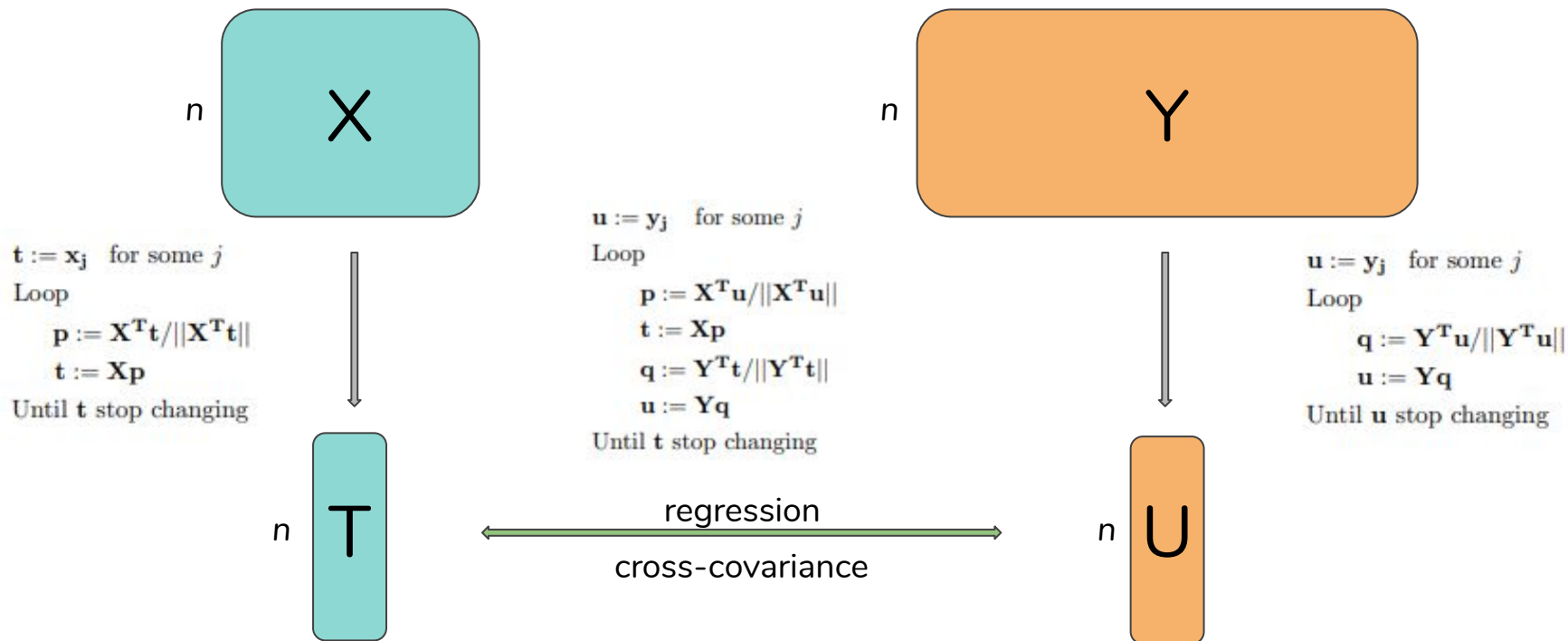
Set a projection of  $\mathbf{X}$  to be one of its columns

Calculate the weights of the principal component

Using the calculated projection weights, project  $\mathbf{X}$  to latent space



# Partial Least Squares





# Alzheimer disease data

409 healthy individuals, 248 patients affected by AD, 460 patients affected by mild cognitive impairment (MCI), and 228 MCI patients subsequently converted to AD during the observational time of the study

**X data:** regional volumes reported in ADNI1/2 for whole brain, ventricles, and average bilateral hippocampi, enthorinal cortex and mid-temporal lobes.

**Y data:** Indicators of 1,167,126 single-nucleotide polymorphisms (SNPs) for each individual



# Results

- Projection of SNPs to lower dimension **explains brain structure**
- Projection of SNPs is the subset of **the most informative variations**
  - Their analysis reveal ones discriminating AD and healthy

SNP	PLS ranking (comp. 1,2) out of 1,167,126 SNPs	Chromosome	Gene	healthy vs AD	MCIs vs MCIs	ENIGMA
rs157580	36th, -	19	TOMM40	1.2e-7	0.034	0.046
rs2075650	1st, 74th	19	TOMM40	1.3e-14	2.17e-6	0.007
rs157582	9th, -	19	TOMM40	5.4e-13	1.3e-8	0.01

**Table 2.** Identified SNPs with largest statistical evidence. Column 2: SNPs ranking relative to the absolute weights of the PLS eigen-components. Columns 5-7: p-value for the group-wise comparison (Mann Whitney non-parametric U-test), or for the correlation with respect to the hippocampal volume in the ENIGMA dataset.



# Resources

Geladi, P., & Kowalski, B. R. (1986). *Partial least-squares regression: a tutorial. Analytica Chimica Acta*, 185, 1–17. doi:10.1016/0003-2670(86)80028-9

Lorenzi, Marco, et al. "Partial least squares modelling for imaging-genetics in Alzheimer's disease: Plausibility and generalization." *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016.

<https://www.diagnosticsolutionslab.com/sites/default/files/u16/snp-similarities.png>

<https://d3i71xaburhd42.cloudfront.net/042f988a290b579554a78dfe4315081b4b6318bc/2-Figure1-1.png>

<https://medium.com/@amarbudhiraja/ml-101-linear-regression-tutorial-1e40e29f1934>