U1167 - Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement

JOLIBIO

Benjamin Grenier-Boley

09/11/2023

Institut Pasteur de Lille

Unit Inserm U1167 RID-AGE

Director : P. Amouyel / 5 teams / 132 persons

1. Santé publique et épidémiologie moléculaire des maladies liées au vieillissement (A. Meirhaeghe)
2. Déterminants moléculaires du remodelage cardiaque et de l'insuffisance cardiaque (F. Pinet)
3. Déterminants moléculaires de la maladie d'Alzheimer et syndromes apparentés (J-C. Lambert)
4. Biologie structurale intégrative (I. Landrieu)
5. Glycation : de l'inflammation au vieillissement (E. Boulanger)

Localization

- Institut Pasteur de Lille (Teams 1/2/3/4)
- Campus Haute Borne Université de Lille (Team 4)
- Centre Hospitalier Universitaire de Lille (Team 5)

Alzheimer's disease - Introduction

Definition

- Neurodegenerative disease (dementia) cause by neuronal loss leading to death
- Memory impairment, executive functions, orientation, ...

Numbers

- 1,2 million cases in France (35 million worldwide)
- 150 million dementia cases in 2050 (2/3 Alzheimer's disease)

Forms

- Familial : 1% cases , << 65 years old, mutations (APP, PSEN1, PSEN2)
- Sporadic : 99% cases, < 65 years old (early) or > 65 years old (late)
  - 20-50% environmental factors : age, education, hearing loss, social isolation, depression, physical activity, diabetes, obesity, ...
  - 50-80% genetic factors : mutations on genome (APOE ε_4, 1993, ...)

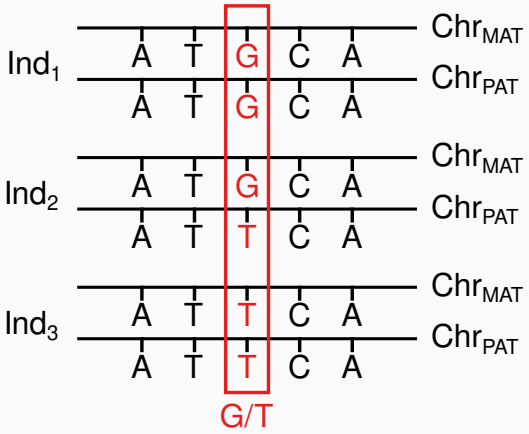
Alzheimer's disease - Analyzing genetic factors

Goal

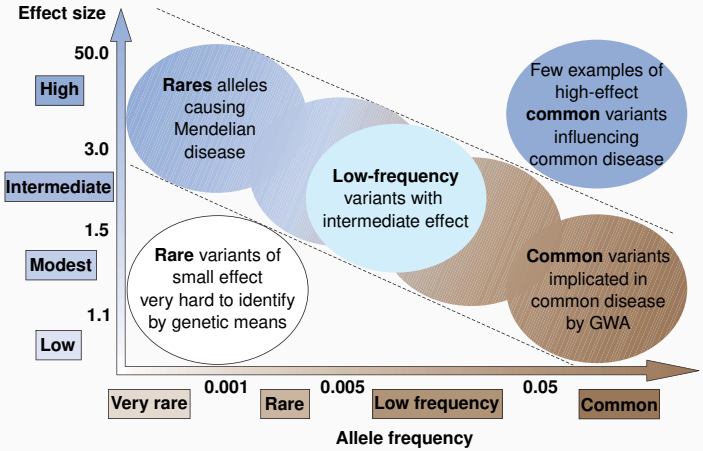
- Discover new genetic factors
- Identify new metabolic pathways
- Understand disease mechanisms
- Discover new biomarkers

Genetic epidemiology

- Analyzing millions of mutations in hundreds of thousands samples
- Comparing frequencies between cases and controls
- Using standardized pipelines or new methods
- Bioanalysis/Biostatistics/Bioinformatics

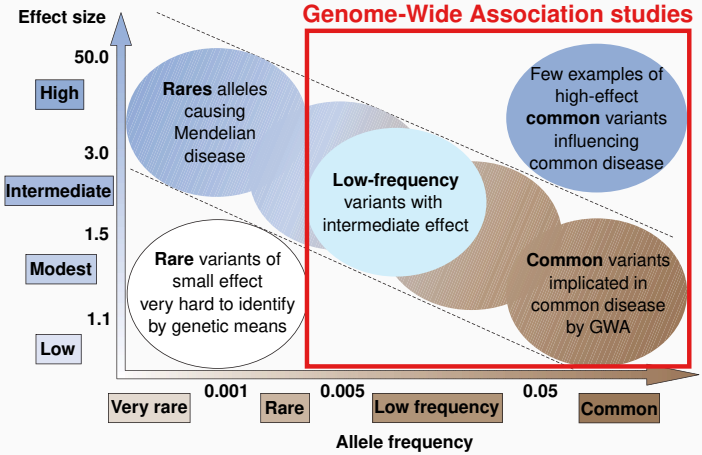


Genetic epidemiology - Principles



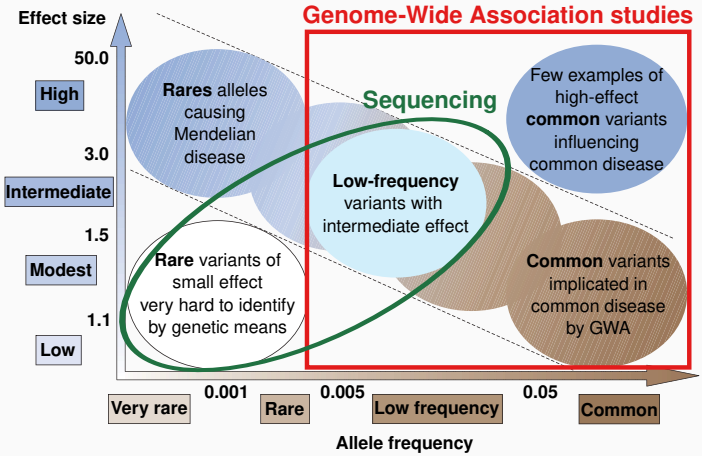
Adapted from Manolio *et al.*, Nature, 2009

Genetic epidemiology - Principles



Adapted from Manolio *et al.*, Nature, 2009

Genetic epidemiology - Principles

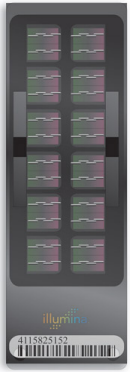


Adapted from Manolio *et al.*, Nature, 2009

Genome-Wide Association studies (GWAS)

Genotyped

- **Mutations :** 200,000 to 4,500,000
- **Cost per sample:** ~ 25€
- **Data Size:** 1.5Gb for 10,000 samples and 600,000 mutations
- **Run time:** Couple of days on a single CPU



Imputed

- **Idea: :** Use of available reference panels to infer missing variations (LD[1])
- **Mutations :** Up to 300,000,000 mutations (Trans-Omics for Precision Medicine[2])
- **Cost per sample:** Cost free (GWAS data)
- **Data Size:** 450Gb for 10,000 samples and 230,000,000
- **Run time:** Thousands of hours on a single CPU

https://www.illumina.com; 1.Linkage disequilibrium; 2.TOPMed: https://imputation.biodatacatalyst.nhlbi.nih.gov

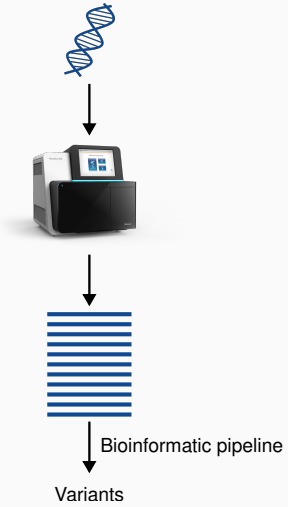
Sequencing

Whole-Exome Sequencing (WES)

- **Data:** All genes sequences (2% of the genome)
- **Cost per sample:** ~ 120€
- **Data size:** 5-10Gb per sample
- **Run time:** 8 years for 1000 samples on a single CPU

Whole-Genome Sequencing (WGS)

- **Data:** All genome sequence
- **Cost per sample:** ~ 350€
- **Data size:** 25-30Gb per sample
- **Run time:** centuries for 1000 samples on a single CPU



Unit data - GWAS and Sequencing data

GWAS

- 110,000 samples
- 230,000,000 imputed variants (TOPMed)
- 4.2Tb initial data
- 43Tb processed data

Sequencing

- 1,400 WES
- 1,500 WGS
- 76Tb initial data
- 135Tb processed data

Require a dedicated storage and computing facility

Unit resources - Mesocentre of University of Lille[1]

Storage

- 1.6Pb of dedicated storage

Computing

- High Performance Computing cluster (HPC)
  - 340Tflops, 6,264 cores and 38Tb of RAM
  - GPU nodes (2 Nvidia A100/80Gb)
  - Pooled resources
- Cloud OpenStack
  - 78.5Tflops, 1,812 cores and 21.6Tb of RAM
- Free of cost for research community

1.https://hpc.univ-lille.fr

Unit main publications

GWAS[1]

- European Alzheimer & Dementia Biobank (EADB)
- 111,326 cases/proxy cases and 677,663 controls
- 75 risk loci found

Sequencing[2]

- Alzheimer Disease European Sequencing (ADES)
- Whole-Exome sequencing
- 16,036 cases and 16,522 controls
- Rare mutations found in 5 genes

1.Bellenguez *et al.*, Nature Genetics, 2022

2.Holstege *et al.*, Nature Genetics, 2022

Unit perspectives

Perspectives

- Continue European effort in EADB : Data/resources opened for collaborators
- Sequence more samples : Focus on WGS
- Other types of analyses : Gene-based, Pathway, Machine learning, ...
- Other types of variations : Structural variants, ...
- Integrate multi-omics data : RNA-seq, Metabolomics, ...
- Multi-ancestry analyses

Acknowledgments

- U1167 RID-AGE
- European/US collaborators
- UK/Finngen Biobanks
- All clinicians contributing to data
- All patients part of these studies
- Fundings

Thank you for your attention