Genome Wide Haplotype analyses of human complex diseases with the EGEE grid

Tregouet David – david.tregouet@upmc.fr

INSERM UMRS937 – UPMC – Paris - France

www.eu-egee.org

e-infrastructure

Genome Wide Association Studies (GWAS)

- **Principle**

  Testing the association between a large number (~500K) of single nucleotide polymorphisms (SNPs) and a variable of interest (e.g: a disease) in a large cohort of individuals

- **How ?**

  Estimate the SNP allele frequencies in cases and controls and calculate the corresponding statistical test yielding a pvalue

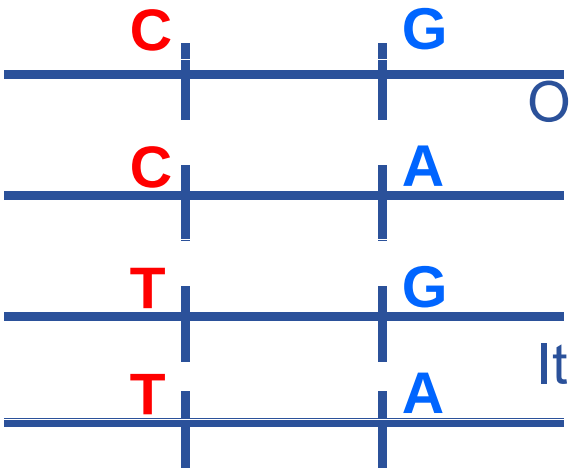
- **SNP definition**

  Genetic variation in a DNA sequence that occurs when a single nucleotide (~ base: A,C,G,T) in a genome is altered. Often considered as a binary 0/1 variable

GWAS' main limits

- **Only single SNP associations are tested**

- **May miss 'haplotypic' interaction between SNPs located in the same gene (or region)**
  - Haplotype: Combination of alleles on a given chromosome
  - For example , with 2 SNPs (C/T & G/A) → 4 haplotypes



One may want to test for difference in haplotype frequencies between cases and controls

It may happen that only one haplotype is at risk

Genome Wide Haplotype Analysis (GWHAS)

- **Is it possible ?**

2 SNPs : up to 4 haplotypes (i.e 00|01|10|11)

3 SNPs : up to 8 haplotypes (i.e 000|001|010|011|100|101|110|111)

In a window (eg a gene or a region) of n SNPs, up to 2^n haplotypes

- **Yes...but**

a large number of tests / comparisons have to be carried out to identify which combination of SNPs is the best predictor for the disease ?

Genome Wide Haplotype Analysis (GWHAS)

- **Is it possible ?**

2 SNPs : up to 4 haplotypes (i.e 00|01|10|11)

3 SNPs : up to 8 haplotypes (i.e 000|001|010|011|100|101|110|111)

In a window (eg a gene or a region) of n SNPs, up to 2^n haplotypes

Example: In a window of 10 adjacent SNPs, restricting the haplotypes of length 4 lead to 375 combinations to be tested:

[SNP1 + SNP2]
[SNP1 + SNP3]
.....
[SNP1 + SNP10]
[SNP2 + SNP3]
.....
[SNP2 + SNP10]
.....
[SNP9 + SNP10]

[SNP1 + SNP2 + SNP3]
[SNP1 + SNP2 + SNP4]
.....
[SNP1 + SNP9 + SNP10]
[SNP2 + SNP3 + SNP4]
.....
[SNP3 + SNP6 +SNP8]
.....
[SNP8 + SNP9 + SNP10]

[SNP1 + SNP2 + SNP3 +SNP4]
.....
[SNP1 + SNP6 + SNP7 +SNP10]
.....
[SNP7 + SNP8 + SNP9 + SNP10]

Genome Wide Haplotype Analysis (GWHAS)

- **GWHAS are possible but are extremely computationnally demanding !!!!**

- **Distribution of the haplotypic calculations on EGEE**
  - Development of an easygLite interface
  - Python & Perl script for results ' visualization

GWHAS on Coronary Artery Disease (CAD)

- **WTCCC data: 1926 CAD patients & 2938 healthy controls**
- **378,000 SNPs**
- **Sliding windows approach on each chromosome**
  Windows of size 10
  Haplotype composed of up to 4 SNPs

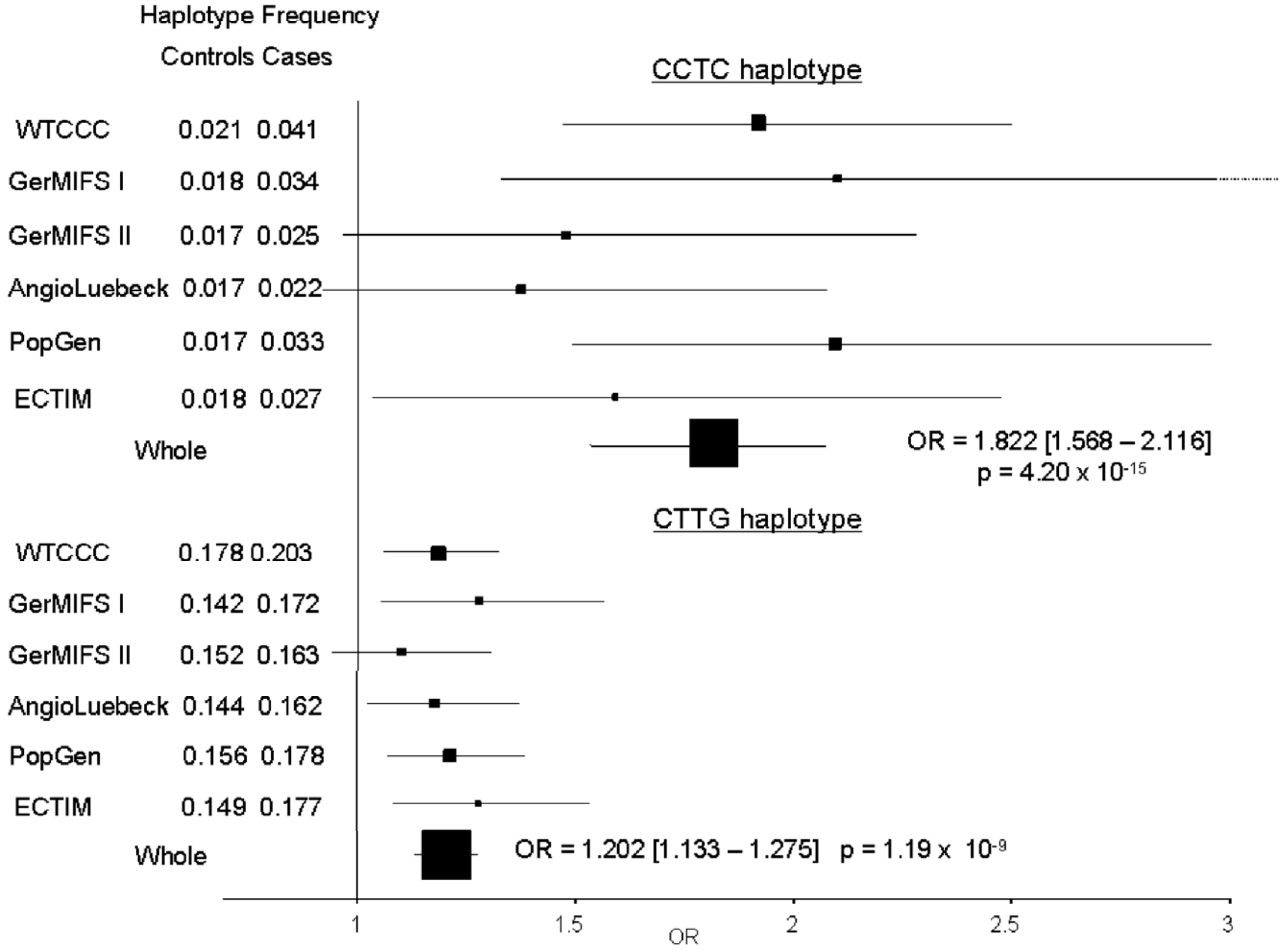
1 to 10	2 to 11	3 to 12	.....	(n-10) to n

- **Search for regions where haplotypes are stronger predictors of CAD risk than SNP alone**

GWHAS on Coronary Artery Disease

- 8.1 millions of combinations tested in less than 45 days (instead of more than 10 years on a single Pentium 4)
- 29 regions where haplotypes could be better predictors than SNPs alone were identified
- To control for false positives , replication was investigated in about 7000 CAD patients and 7000 controls
- One region on chromosome 6 was confirmed

Nature Genetics doi:10.1038/ng.314



Conclusions

- Genome Wide Haplotype Association Studies are now a reality thanks to the use of Grid technology
- Using EGEE, we were able to identify a cluster of 3 genes where haplotypes are strongly associated with CAD risk (Tregouet et al. *Nature Genetics March 2009*)
- Possibility to apply such tool to other human diseases (Diabetes, Cancer....)
- Possibility to use EGEE to investigate interactions between SNPs that are not necesseraly in the same gene/region

Credits

Inserm

UMRS 937

Institut national
de la santé et de la recherche médicale

UNIVERSITE PIERRE & MARIE CURIE
LA SCIENCE A PARIS

François Cambien
Alexandru Munteanu
Laurence Tiret
Claire Perret

cardiogenics

Nilesh Samani
Heribert Schunkert
Inke König
Jeannette Erdmann
Andreas Ziegler
....

cnrs
advancing the frontiers

UMR 8623

LRI

Cécile Germain