# GWAS in R

GBIO0009

Kridsadakorn Chaichoompu

University of Liege

# GenABEL package

- R package for:
  - Designed for GWA analysis
  - Graphical results of GWA analysis
- To install:

  install.packages("GenABEL")
- Tutorial

http://www.genabel.org/sites/default/files/html_for_import/GenABEL_tutorial_html/GenABEL-tutorial.html#x1-180004

# GenABEL: Exercise 1

- Exploring IDs in srdta

1. How many people are included in the study?
all = nids(srdta)

2. How many of these are males?
gender = male(srdta)
no.male = length(gender[which(gender == 1)])

3. How many are females?
no.female = length(gender[which(gender == 0)])

4. What is male proportion?
no.male / all

# GenABEL: Exercise 2

1. What are names of markers located after 2,490,000 b.p.?

```
pos = map(srdta)
snp = snpnames(srdta)
snp[which(pos > 2490000)]
```

2. Between 1,100,000 and 1,105,000 b.p.?

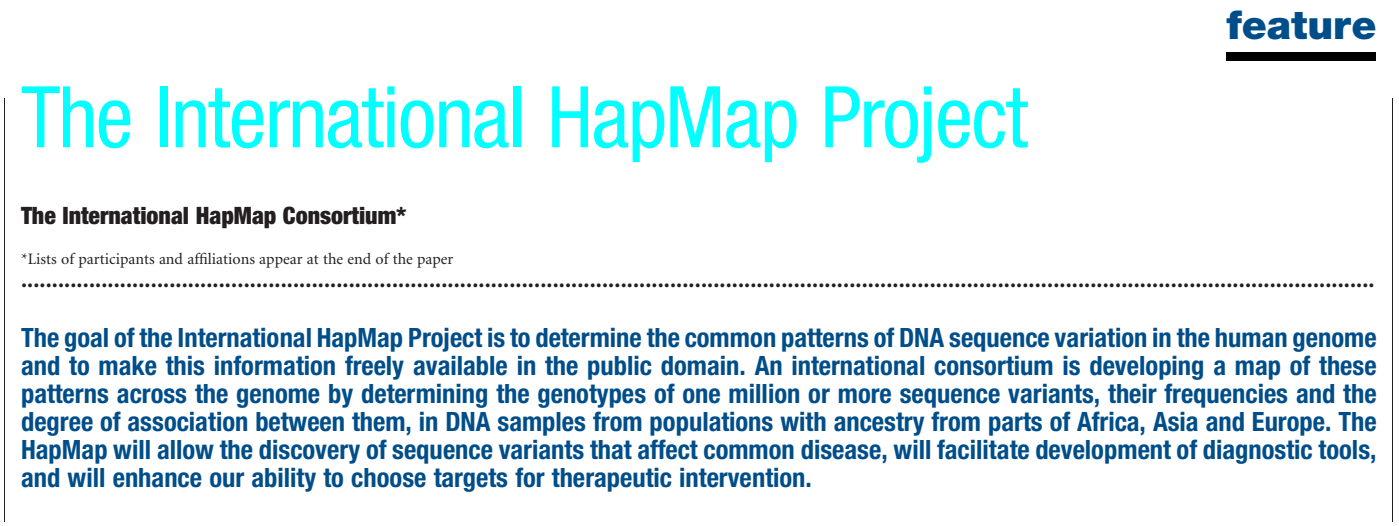
```
snp[which(pos >= 1100000 & pos <= 1105000 )]
```

# Homework 1

- We are going to use the HAPMAP data as case study

http://www.montefiore.ulg.ac.be/~chaichoompu/CK/?Courses___2016_-_GBIO0009_-_Topics_in_Bioinformatics

# HAPMAP



feature

## The International HapMap Project

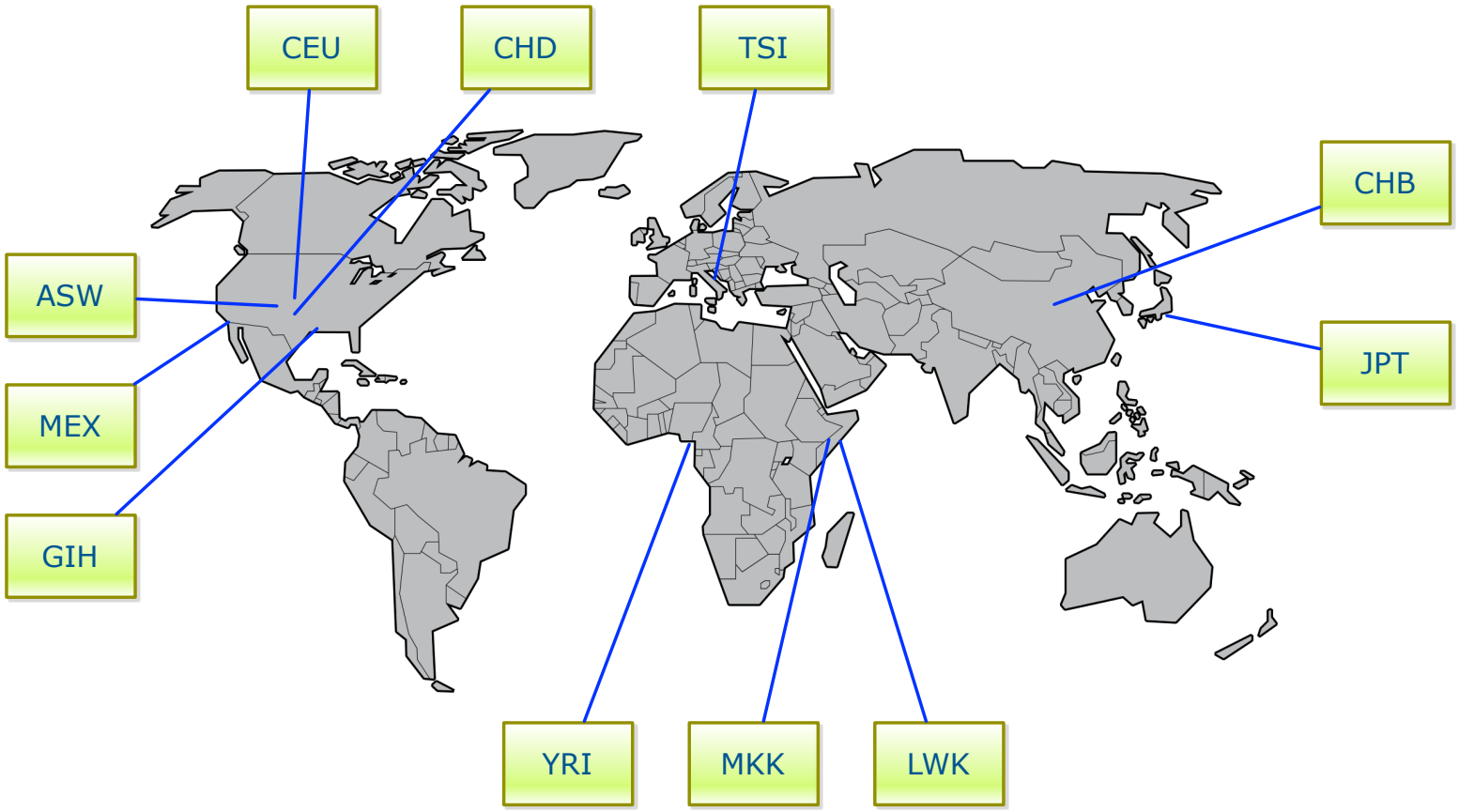
**The International HapMap Consortium***

*Lists of participants and affiliations appear at the end of the paper

**The goal of the International HapMap Project is to determine the common patterns of DNA sequence variation in the human genome and to make this information freely available in the public domain. An international consortium is developing a map of these patterns across the genome by determining the genotypes of one million or more sequence variants, their frequencies and the degree of association between them, in DNA samples from populations with ancestry from parts of Africa, Asia and Europe. The HapMap will allow the discovery of sequence variants that affect common disease, will facilitate development of diagnostic tools, and will enhance our ability to choose targets for therapeutic intervention.**

https://www.genome.gov/10001688/international-hapmap-project/

# HAPMAP samples




ftp://ftp.ncbi.nlm.nih.gov/hapmap

# PLINK formats

PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner.

File formats in PLINK 1.07:

- Text file
  - PED
  - MAP
- Transposed text file
  - TPED
  - TFAM
- Binary file
  - BED
  - BIM
  - FAM
- Data conversion

http://pngu.mgh.harvard.edu/~purcell/plink/data.shtml

# How to load PLINK files with GenABEL?

We can load dataset using the transposed format.

http://www.genabel.org/sites/default/files/html_for_import/GenABEL_tutorial_html/GenABEL-tutorial.html#x1-95000A.2

# HAPMAP: Exercise

- Download the example data from the course website

http://www.montefiore.ulg.ac.be/~chaichoompu/CK/?Courses___2016_-_GBIO0009_-_Topics_in_Bioinformatics

- How many individuals are in the dataset?

  all = nids(df)
- How many males and females are there?

  gender = male(df)

  no.male = length(gender[which(gender == 1)])

  no.female = length(gender[which(gender == 0)])
- How many SNPs?

  pos = map(df)

  length(pos)

# GenABEL: Quality Control Processes

- Hardy–Weinberg equilibrium test
- Minor Allele Frequency filtering

http://www.genabel.org/sites/default/files/html_for_import/GenABEL_tutorial_html/GenABEL-tutorial.html#x1-260005.2

- Linkage disequilibrium pruning

http://www.genabel.org/sites/default/files/html_for_import/GenABEL_tutorial_html/GenABEL-tutorial.html#x1-120003.2

# GenABEL: Association test

- Qtscore - Fast score test for association between a trait and genetic polymorphism

http://www.genabel.org/sites/default/files/html_for_import/GenABEL_tutorial_html/GenABEL-tutorial.html#x1-280005.4