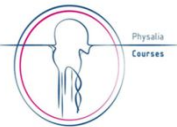


# Towards the GWAS pipeline: revising the steps



Christian Werner

*(Quantitative geneticist and biostatistician)* **EiB, CIMMYT**, Texcoco (Mexico)

Filippo Biscarini

*(Biostatistician, bioinformatician and quantitative geneticist)* **CNR-IBBA**, Milan (Italy)



HerrFalloppio

Oscar González-Recio

*(Computational biologist and quantitative geneticist)* **INIA-UPM**, Madrid (Spain)



OscarGenomics



# Recap: demystifying the practicals

## Illustrations

- local (your computer)
- RStudio
- R Markdown: interactive scripts that mix code and explanations
- basic workflow in R

(more free to experiment/change as you please; beware of local settings and paths to data/scripts)

## GWAS full workflow

- on the server
- bash/R/stand-alone software (Plink, Beagle)
- self-contained scripts
- Pipeline

(common server settings, more structured workflow; hurdle of working on the server with tools like the Linux shell, vim and git)



# The GWAS **workflow**

1. **Get the data:** phenotypes and genotypes
2. **Filter** the data
3. **Impute missing** SNP genotype data
4. **Build** and **run** the **GWAS model**

[GWAS workflow in one slide](#)



# step 1 - getting the data

- **1.get\_data.sh**
  - Download the data
  - Prepare the data



## step 2 - filter the data

- **2.steps\_filtering.sh**
  - Filter genotype data:
    - MAF
    - Missing rate



## step 3 - **imputing** missing genotypes

- **3.step\_imputation.sh**
  - Impute missing genotype data:
    - LHCI



# the stand-alone **GWAS** scripts

- Run GWAS programmatically
  - gwas\_rrblup.R
  - gwas\_statgengwas.R
  - gwas\_sommer.R



## step 4 - run the GWAS

- **4.gwas.sh**
  - Run GWAS through the stand-alone script
  - (try multiple scripts: gwas\_rrblup.R, gwas\_statgengwas.R, gwas\_sommer.R)





## NEXT LECTURE

Collaborative **exercise**

