

BIOS data: What, Where and How

Joost Verlouw

BIOS symposium

21-09-2016



BIOS data: What

- Several datasets are available
 - RNA sequencing – n=3824
 - DNA methylation – n=4805
 - Corresponding Geno- and Phenotypes
 - Imputed Genotypes to both GoNL and Human Reference Consortium (HRC) panels
 - Phenotypes such as:
 - Additional Phenotype data can be requested

| | |
|-------------|----------------------|
| Sex | Smoking status |
| Age | CRP |
| Weight | Cholesterol |
| Height | Tryglicerides |
| Cell counts | Technical covariates |



BIOS data: Where

- Three storage/compute mechanisms:
- BIOS Virtual Machine (VM)
 - Pre-processed datasets
 - Compute available
- GRID SRM storage
 - Raw data
- EGA
 - <https://www.ebi.ac.uk/ega/studies/EGAS00001001077>



BIOS data: How

- All information to be found at the BBMRI wiki
- http://www.bbmriwiki.nl/wiki/BIOS_start-
- Contains instructions on (among other things):
 - What datasets and where
 - Description of BIOS data infrastructure
 - How to connect to the VM
 - How to use the BIOSRutils package



BIOS data: How

File data request

Using data request forms on wiki



Code of Conduct

Sign Code of Conduct and mail to management team



Request access to VM

Mail to VM system admins with public key



Perform analysis on dataset

Use VM to run analysis on requested dataset