

## General structure for GABRIELA project (“gabriela”):

Folder “Data” (*not provided*):

- Contains general non-genetic datasets used in for analyses
  - originally delivered datasets + documentations as pdf files
  - created datasets used for analyses, mostly used here:
    - “5imps.rda”: 5 imputations for the relevant variables for environment, family history, demographics
    - “dat.outcome.rda”: dataset containing the different asthma outcomes
    - “ind.part.rda”: objects indexing which observations to be used for (cross-validation within) training data and test/validation data

“1-Environmental-data.r”:

- builds imputed data sets via MICE
- creates prepared data files for environment (and family history and demographics) (- creates index-vectors in object “fin.ind.rda” for cross-validation and outer cross validation; only used in first analyses; see next paragraph)

“2-Partitioning-data.R”:

- index-vectors saved in object “ind.part.rda”, needed for
  - indexing for cross-validation in training data
  - indexing external validation data
  - indexing farm only/ non-farm only/ all children

“3-Build-outcomes.R”:

- building the outcome variables

Folder “ImputedSNPs”:

- original genome-wide (imputed) SNP data
- codes for creation of prepared data sets
- prepared data sets

Folder “Literature SNPs”:

- previous analyses versions
- Data preparation for literature SNPs
- Literature SNPs

Folder “Results” (folder structure self-explaining):

- all codes for presentations of results
- plots/figures
- tables

Folders “2-validate-innsbruck”, “3-onlyfarm”, “4-onlynonfarm” are the analyses on ALL children, ONLY FARM children, and ONLY NON-FARM children, respectively. They have all identical structure:

- “ImputedSNPs”: analyses where the SNPs are genome-wide (and imputed)
  - R-Codes for the analyses involving genome-wide SNPs
  - “Data”: relevant data produced for genome-wide analyses
  - “Results\_genomewide”:
    - Folders “imp1”, “imp2”, ..., “imp5”: results for analyses on 1.,2.,...,

5. imputation data set, respectively

- Folder "aucs": aggregated results (calculated AUCs) from folders "imp1", "imp2", ..., "imp5"
- "LiteratureSNPs": analyses where the SNPs selected via literature
- Folder "2-LearningMethods":
  - R-codes for different types of analyses involving 19 literature SNPs
  - "Results19":
    - Folders "imp1", "imp2", ..., "imp5": results for analyses on 1.,2.,..., 5. imputation data set, respectively
- Folder "aucs": aggregated results (calculated AUCs) from folders "imp1", "imp2", ..., "imp5"
- Folder "finalmodel": objects created for the final model applied on the whole GABRIELA data set (without Austrian arm)
- Folder "Additional": analyses on asthma and non-asthma children with farm as outcome

Folder "5-validate-pasture":

- non-imputed data set "pasture.new.rda"
- code for preparation of data from PASTURE study
- prepared PASTURE data (5 imputed data sets + outcomes)