RHAPSODY WP3 Pre-Diabetes Analysis Plan

(Version 0.8.1)

Mickaël Canouil

30 May 2018

Files description

At this stage, you were provided with the following files from the (hidden) Gist (don't share outside RHAPSODY):

- `README.Rmd`, the file you are actually reading;
- `RHAPSODY_WP3_PreDiab.Rmd`, a Rmarkdown script which perform:
  - opal node access,
  - phenotype QC,
  - preliminary modelling,
  - VCFs formatting/filtering,
  - variants analysis;
- `RHAPSODY_WP3_PreDiab_DEBUG.R`, a small part of the main Rmarkdown script, which includes only node access and phenotype QC;
- `RunAnalysis.sh`, a shell script to run the whole analysis (parameters have to be filled properly according to your informations);
- `handleVCF.sh`, the shell script used to format/filter VCFs files prior to the variant analysis. This shell script is also included within the main script, but you can use it directly as a standalone, if so (or if your already done formatting and don't want to do it again), parameter `format_vcfs` must be set to `FALSE`).
- `opal_credentials.txt`, the server, login and password informations to access phenotype data on your local node;
- `Install_Rpackages.R`, a R script which install a predefined version of all R packages loaded in the Rmarkdown script, only if needed.

Run the analysis

R Setup

Please use R version 3.4.2 (2017-09-28) (if not possible, newer version “3.4.*" can be used).
If VCFtools is not available on your machine, please install it: VCFtools. Be sure the needed packages are installed (this is also done within the Rmarkdown script) or copy/paste the following command.

```
check_packages <- function(package) {
  if (!package %in% installed.packages()[, "Package"]) {
    install.packages(
            package,
            repos = c(
              "https://rhap-fdb01.vital-it.ch/repo/",
              "https://cran.rstudio.com/",
              "http://cran.obiba.org"),
            dependencies = TRUE
        )
  }
  library(package = package, character.only = TRUE)
}

list_packages <- c(
  "parallel",
  "grid",
  "scales",
  "broom",
  "viridis",
  "readxl",
  "writexl",
  "cowplot",
  "knitr",
  "kableExtra",
  "lme4",
  "lmerTest",
  "Hmisc",
  "data.tree",
  # "RCurl",
  # "rjson",
  "opal",
  # "datashieldclient",
  # "dsCDISC",
  "tidyverse"
)

invisible(sapply(list_packages, check_packages))
```

Run the Rmarkdown script

1. First, create (or change) the `opal_credentials.txt` with your informations.

2. Fill in the proper informations.

3. Set the `analysis_step`:

    1. Node access
    2. Phenotype QC
    3. Run Mixed Model without variants
    4. Check VCFs files
    5. Format VCFs files
    6. Run Mixed Model with (imputed) variants
    7. Check output files

4. Run the R code below within interactive R session.

5. a/ If everything worked, you should have a HTML report.
   Increment `analysis_step` and repeat steps 3 to 5.
   b/ If something went wrong, you should have an error message, for example:

   ```
   label: list_vcf (with options)
   List of 2
    $ eval  : language params_steps$step_5
    $ engine: chr "sh"

   Quitting from lines 1597-1598 (RHAPSODY_WP3_PreDiab.Rmd)
   Error in if (options$eval) { : argument is of length zero
   ```

   Either try to solve the issue yourself and/or send the informations to mickael.canouil@cnrs.fr

```r
# set the output directory or leave as is;
# output will be generated where the Rmarkdown file is
working_directory <- './'

reportName <- 'RHAPSODY_WP3_PreDiab'

# Analysis
rmarkdown::render(
  input = "RHAPSODY_WP3_PreDiab.Rmd",
  output_format = "html_document",
  output_file = paste0(reportName, ".html"),
  output_dir = working_directory,
  params = list(
    cohort_name = 'CohortName',
    author_name = 'Firstname LASTNAME',
    opal_credentials = 'opal_credentials.txt',
    vcf_input_directory = './vcfs',
    imputation_quality_tag = 'INFO', # To be set according to VCF (could also be "R2")
    vcftools_binary_path = './vcftools/vcftools_latest/bin',
    output_directory = working_directory,
    analysis_step = 1,
    format_vcfs = TRUE,
    variants_analysis = TRUE,
    n_cpu = 2,
    echo = FALSE, # Should R code be printed in the report
    warning = TRUE, # Should warnings be printed in the report
    message = FALSE # Should messages be printed in the report
  ),
  encoding = "UTF-8"
)
```