Gene Level Meta-analysis of Quantitative Traits by Functional Linear Models

Ruzong Fan, NICHD/NIH, 2015

1 Overview

This document describes a R package to implement the functional linear models for gene level meta-analysis of quantitative traits (Fan et al. 2015). Section 2 briefly describes the installation of the program. Section 3 explains how to run the program using one example. Section 4 offers explanation of the results and warnings to use the programs. Section 5 provides some suggestions and parameter choices for real data analysis.

The theoretical basis for this program is given in our research papers in **References**. Please refer to the reference if you use the program in any published work. In case of suggestions and questions and/or problems, you can contact us via e-mail (fanr@mail.nih.gov).

2 Download and Installation

The package is written in R. Download "MetaFLM_fixed_model.R", "MetaFLM_beta_smooth_only.R", "MetaFLM_FPCA.R", "MetaFLM_FPCA_no_position.R", and "MetaFLM_additive_effect_model.R", and an example file of "MetaSKAT_Example_for_MetaFLM.R" from FLM_meta.zip. Put the files in a directory you may access.

3 How to Run the Program

The analysis needs libraries fda, MASS, Matrix, and MetaSKAT in R package. The datasets are from the package MetaSKAT. Make sure to install them before running our codes. Open the "MetaFLM_Example.R" file on an R Console in a PC window. Please change the paths leading to the directories of the package "MetaFLM_fixed_model.R", "MetaFLM_beta_smooth_only.R",

"MetaFLM_FPCA.R", and "MetaFLM_FPCA_no_position.R" on your computer. Then you may run "MetaSKAT_Example_for_MetaFLM.R". Please note that the following codes

```
y1 = rnorm( length(y.list[[1]]) )
y2 = rnorm( length(y.list[[2]]) )
y3 = rnorm( length(y.list[[3]]) )
y = list(length = 3)
y[[1]] = y1
y[[2]] = y2
y[[3]] = y3
pheno = y
```

will generate 3 random samples of quantitative trait values. Therefore, the results will be different from time to time. On April 2, 2015, I got the following results

```
> MetaFlm_beta_smooth_only(L, is.homo = TRUE, y, mode = "Additive", geno, pos, order,
                           bbasis, covariate, base = "bspline", interaction = FALSE)
$LRT
[1] 0.5235781
$Chisq
[1] 0.5235781
$F
[1] 0.5237607
......
> MetaFlm_add_effect(L, is.homo = FALSE, y, mode = "Additive", geno, covariate)
$LRT
[1] 0.8539397
$Chisq
```

```
[1] 0.8539397
$F
[1] 0.8512034
>
> ### The following function is not working since geno[[k]] have different number
      of columns ###
>
> MetaFlm_add_effect(L, is.homo = TRUE, y, mode = "Additive", geno, covariate)
Error in rbind(U, geno[[k]]) :
  number of columns of matrices must match (see arg 2)
```

To make "MetaFlm_add_effect(L, is.homo = **TRUE**, y, mode = "Additive", geno, covariate)" to run, one needs that each individual of the L studies is sequenced at the same variants. However, the function "MetaFlm_add_effect(L, is.homo = **FALSE**, y, mode = "Additive", geno, covariate)" can analyze different genotype data among multiple studies, i.e., individuals of different studies may be genotyped at different genetic markers. The details are provided in Fan et al. (2015).

4 Explanation of the Results and Warnings

As shown in the Section 3, our program can output 3 p-values based on likelihood ratio test (LRT), χ^2, and F-distributed test. The LRT is the same as χ^2, which have correct type I error rates when sample size is large than or equal to 1,500 for single study (Fan et al. 2013, p733, top of the left column), and have correct type I error rates for large sample of multiple meta-analysis studies (Fan et al. 2015). The F-distributed test has conservative and accurate type I error rates (Fan et al. 2013; 2015). If you use the R codes to analyze your data, we recommend to report the p-values of F-distributed test. If you analyze large sample data, both LRT and F-distributed tests can be used.

5 Suggestions and Parameters for Real Data Analysis

In this documentation, we present four R functions to perform gene-based association analysis of quantitative traits. In practice, one may use one of them for data analysis. We suggest to use one of MetaFlm_beta_smooth_only and Metaflm_fixed_model by either B-spline or Fourier spline basis functions. We also suggest the following parameters for a data analysis:

```
order  =  4
bbasis = 15
gbasis = 15
fbasis = 25
```

The two functions, "MetaFLM_FPCA.R" and "MetaFLM_FPCA_no_position.R", are based on functional principal component analysis. We do find that they provide correct type I error rates and similar power levels as MetaFlm_beta_smooth_only and Metaflm_fixed_model, although the results are not presented in Fan et al. (2015).

6 References

Fan RZ, Wang YF, Mills JL, Wilson AF, Bailey-Wilson JE, and Xiong MM (2013) Functional linear models for association analysis of quantitative traits. *Genetic Epidemiology*, 37:726-742.

Fan RZ, Wang YF, Boehnke M, Chen W, Li Y, Ren HB, Lobach I, and Xiong MM (2015) Gene level meta-analysis of quantitative traits by functional linear models. *Genetics*, 200:1089-1104.