HELP to Use the macro FAMT

1 Download

Download the FAMT macro (Excel file) and open it.

![Figure 1: Illustration of the FAMT macro](image)

The macro contains 4 sheets:

- The "Macro" sheet displays the results of the different steps of the Factor Analysis for multiple testing package.

- The "Expression" sheet contains the gene expressions data frame: genes are in rows without row names and arrays in columns (the column names are the identifier of arrays) (see Figure ??).

- The "Covariates" sheet gives information about the experimental conditions: the identifier of each row (arrays), as used in the column names of Expression, is provided, with the value of the main explanatory variable in the testing issue and possibly other covariates (see Figure ??).

- The "Annotations" sheet provides additional information about the response variables of the multiple testing procedure to be used to describe the results. One column must be named ID and gives the variable (gene) identifier (see Figure ??).
The data must be stored in each sheet as above-mentioned. Note that Covariates and Annotations datasets are optional. The number of columns of expression must correspond to the number of rows of covariates and furthermore expression and annotations must have the same number of rows. If covariates dataset is not provided the procedure aims at testing the significance of the mean expression. If annotations dataset is not provided, a basic annotations dataset is created with row indices as variables identifiers.

**Import your data** (copy and paste the data in each sheet or use the excel menu via import extern data). Be careful to the decimal mark (comma or point). The following figures present the different sheets with the data provided in the package. This dataset concerns hepatic transcriptome profiles for 9893 genes of 43 half sib male chickens selected for their variability on abdominal fatness.

![Figure 2: An example of the Expression sheet](image)

Another sheet, called "graph", is automatically created to store all the graphs provided by the functions of the package. The buttons in the "Macro" sheet allow to run the package by step. The CLEAR button clears the "Macro" sheet and the "graph" sheet.
2 The DATA button

The DATA button enables you to define the datasets, to create the FAMT data and to summarize the FAMT data. You have to select the available datasets (expression, covariates and/or...
annotations), then some other dialog box allow to precise the class of variables.

2.1 Statistics of the FAMT data

When you have defined the data frame, you can ask for summary of the FAMT data by clicking on Statistics of data.

The results are displayed in the "macro" sheet. The function provides:

- For Expression: the number of tests which corresponds to the number of rows, the sample size which is the number of columns.

- For Covariates and Annotations: classical summaries.
3 The FAMT MODEL button

The FAMT model button enables to implement the FAMT complete multiple testing procedure. When you click on the FAMT model, a form with 4 text box is opened (see Figure ??):

- The first box determines the experimental condition and the optional covariates.
- The second box corresponds to the experimental condition on which the test is done.
- The third and fourth box are optional and they refer to the number of factors. You can select a number of factors to fit the FA model (in the third box) or this number is estimated. The last box allows to change the default value of the maximum number of factors tested to estimate the optimal number of factors.

If you don’t fill the box, the default values are kept for the fitting of the FA model:

- The experimental condition is the column 1 in the "Covariates" sheet (x=1)
- The test is done on the 1st column of the previous vector (x[1]). If x=1, test = 1 too.
- The function estimates the optimal number of factors
- The maximum number of factors tested to estimate the optimal number of factors is 8.

In our illustrative example (data provided with the FAMT package), we test the significance of the relationship between each gene expression and the abdominal fatness (6th column of covariates), taking into account the effect of the dam (4th column of covariates). So, in the first box, we write 4 and 6, column numbers corresponding to the experimental condition and covariates (column numbers are separated by semicolon), and in the second box, we type 6, column number of the explanatory variable of interest (see Figure ??).
The optimal number of factors used to fit the model is given in the "Macro" sheet. The "graph" sheet, automatically added in the excel file, contains three graphs (see Figure 7):

- The values of the variance inflation criteria for each number of factors are plotted
- The histograms of p-values and adjusted p-values.

If you want to use the FAMT method as a classical multiple testing procedure without any modeling for the dependence structure across the variables, choose 0 for the number of factors to adjust the FA model.

This step builds the "FAMT model", and enables you to analyze results (with the Results button). You can fit different models but the last one is used for the analysis of the results.
4 The RESULTS button

The RESULTS button proposes three functions to analyse and display the results (see Figure ??). The first function provides information about the rejected genes, the second one gives an estima-
tion of the proportion of true null hypotheses, and the last one helps the user to describe and interpret the factors.

4.1 Statistics of the FAMT model

The selection of "Statistics of the FAMT model" gives the number of rejected genes according to raw analysis and FAMT analysis, the annotations characteristics of significant genes, and the estimated proportion of true null hypotheses.

The number of positive tests is provided for each level of False Discovery Rate (FDR) control chosen by the user (the default value is 0.15). If you want to change the level of FDR control, you have to define the range of the FDR control. In our illustrative example, we select a range from 0 to 0.3 with increment of 0.05 (see Figure ??).

The significant genes are listed with the genes identification and array names in the original data frames. You can change the identifiers (add some characteristics for example) by clicking in the check box "identification of the significant genes" (see Figure ??).
A new dialog box is displayed and you can select the identifiers among the annotations variables. Results are shown in the "Macro" sheet (see Figure ??). The list of positive genes is given for the highest level of FDR.

If you don’t fill the box of the "Statistics of the model", the default values are selected: the FDR control is 0.15 by default, the significant genes are characterised by the genes identification and their name in the Annotations file.
4.2 Estimation of the proportion of true null hypotheses

The function estimates the proportion of true null hypotheses (pi0). The histogram of the p-values with the estimate of pi0 null line is plotted in the "graph sheet. An additional graph is displayed
Figure 13: Display of the results of the Statistics of the model showing the spline curve used to estimate pi0 (see Figure ??).

Figure 14: Display of the results of the pi0

The algorithm used to estimate the proportion of true null p-values is the "smoother" method (this method uses the smoothing spline approach proposed by Storey and Tibshirani (2003)).
4.3 FAMT factors description

This function provides diagnostic plots to interpret and describe the factors using external information either on genes or arrays. To use this option, the FAMT data must contain "Annotations" dataset. You have to fill three items:

- the axes: a length 2 vector specifying the factors to plot
- the covariates
- the factors of annotations.

The default value of the axes is the factors 1 and 2.

The function takes all covariates except those used in the model and the array name. The function takes all variables of annotations of factor type.

In our illustrative example (see Figure ??), the axes are the two first factors, the external covariates are the column 3 (Pds9s: the body weight) and the column 5 (Lot: the hatch), the external annotations are the columns 2 (Block) 3 (Column) 4 (Row) which correspond to the location on the microarray and the column 5 (Length: oligonucleotide size).

Figure 15: FAMT factors description

Graphical devices are plotted in the "graph" sheet if the FAMT model has more than one factor (see Figure ??).

The tables of p-values are displayed in the "Macro" sheet: p-values of the test whereas the score of each factor are affected by the selected covariates, and p-values of the test whereas the score of each factor are affected by the selected annotations.
Figure 16: Display of the results of the FAMT factors description