Functionality of the Advanced Statistical Module
From the workflow to case studies

C.H.E.M. Adriaans
Index

• Introduction
• Advanced Statistical module
• Aim of the study
• Mock up forms
• Results of two case studies
• Usefulness of the advanced statistical module
• Future developments
• Summary
Introduction

Computer processing programs and tools developed to analyse vast growing amount of Big data

1. Microarray dataset
2. Raw data CEL files
3. Quality control and data pre-processing
4. Normalised data
5. Statistical analysis
6. Statistical results
7. Pathway analysis
8. Pathway analysis results
9. Results interpretation and literature validation

Adapted from Baetke et al
Adapted from Eijsen et al

A - webservice

Statistical analysis

B - Local host

Adapted from Eijsen et al
Advanced Statistical Module

• Integration of paired testing with a pairing variable
• Covariates and an interaction model
Aim of the study

1. Developing a new layouts for the forms

2. Show the functionality and usefulness of a workflow by running two case studies.
Mock up forms

**Paired Dataset**

Is your data paired? (constricted to one variable pairing)

- Yes
- No

Please fill in your pairing variable

**Covariates**

Does your dataset have covariates?

- Yes
- No

Please choose if your covariates are numeric or factorial

- Factorial
- Numeric

Please fill in whether or not you want to have a baseline-effect

- Baseline

Does your model have an interaction?

- Yes
- No

Please fill the interactions in

*(Interaction model is only possible if the covariate has a base-effect)*

Your formula is:

**COMPUTED AUTOMATICALLY**
Adapted from Eijsen et al
Two case studies

• Gene Expression Omnibus

• GSE34111; test the functionality of the paired testing

• GSE36245; test the functionality of the covariates and interaction module
Dataset GSE34111; understanding muscle wasting in cancer patients

- Twelve patients
  - 10 males
    Ages 56 to 82
  - 2 females
    Ages 59 and 63

- Paired biopsy
  - Pre-resection
  - 8 months post resection
Quality control results of GSE34111

Density histogram after GCRMA
Curves should be comparable between arrays

Boxplot after GCRMA
Distributions should be comparable between arrays
Statistical analysis results of GSE34111

BiGCaT
oxidation by cytochrome P450 pathway
Results of GSE34111

Down-regulated:
CYP 2W1
CYP2C19
and CYP3A4

Up-regulated
CYP2E1
Dataset GSE36245; glioblastoma

- 46 patients:
  - 23 males; ages 1 to 48
  - 23 females; ages 6 to 46
- 3 groups:
  - MUT1
    - H3AFA mutation
    - 12 samples
  - MUT2
    - IDH1 mutation
    - 9 samples
  - WT
    - 18 samples
Qual

QC1: M → Sample

QC2: M

QC3: M
Statistical analysis results of GSE36245

Age - numerical

Age - factorial
Statistical analysis results of GSE36245

- Covariate Age
  - Factorial VS Numerical

- Covariate Sex
  - Non-interaction VS interaction model
## Covariate Sex

<table>
<thead>
<tr>
<th>Non-interaction</th>
<th>Interaction model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mut_f</td>
<td>WT_f</td>
</tr>
</tbody>
</table>

### Covariate Sex with interaction model was used for the pathway analysis

- **Non-interaction model:** gene-expression male/female do not need to be the same

- **Interaction model:** gene-expression male/female not the same, but the effect between WT and the mutation may differ
Pathway analysis results of GSE36245

1. Notch Signaling &

2. TGF Beta Pathway
   - Both showed similarities in regulation between males and females
   - Both showed lower of Z-score lower in males

3. DNA Replication
   - Power of Z-score lower in females; while for the males it is the highest Z-score
   - no significance found in females
DNA Replication Pathway in males

1. Assembly of the pre-replicative complex

2. Activation of the pre-replicative complex

3. DNA replication initiation

4. Switching of origins to post-replicative state

5. Leading/lagging strand synthesis
Pathway Replication males vs females

• ORC6L:
  - females: down-regulated
  - males: up-regulated

  **Difference in sex?!**

  - females: down-regulated
  - males: up-regulated

• RNA primer-DNA primer:
  - females: down-regulated
  - males: up-regulated
Usefulness of advanced statistical module

- Ability to test paired datasets
- Ability to test the influences that covariates may have on a dataset
  - Including the interacting that covariates can have with one another on the dataset
Future development

• User interface
  – PHP; to integrate to the webservers
Summary

1. New layouts for the forms of the statistical module were made
2. Functionality and usefulness of a workflow including the advanced statistical module functionality by running two case studies
3. Future development of PHP; to integrate to the webserver
Mock up forms

Statistical analysis of gene expression data

Before running this module, you may visit its referred user guide

Enter your normalized data: a file containing normalized expressions

Browse normalized data file

Choose File

Or discover [Statistical analysis] using an example dataset (Example1)

Email address (recommended)

Disclaimer: your e-mail address will be used for sending your results, and will be deleted off our server in 14 days. It will not be used for other purposes.

Please don’t make changes or click any button while data is uploading

Next
[Statistical Analysis] Describe your dataset

For each “SourceName” (headers of the normalized data file), enter a proper “Experimental Group” (experiments/arrays or the @ sign for annotations. At least two different Experimental Groups are required. You may also load a description file.

Please note that experimental group names may not contain special characters other than a dot(.) or an underscore(,) and will be changed to continue the statistical analysis. Also please note that we assume the Experimentalgroups are factorial, if not please choose the Advanced Statistical Analysis!

SourceName

Experimental Group

Group names will change if a special character other than dot or underscore is given.

Do you want to do advanced statistical analysis? □ Yes □ No
(paired data, covariates)

Please don’t make changes or click any button while data is uploading Next

Back to the previous step
Only one paired function at a time using a single pairing variable.
Not possible: layered pairing variables
[Statistical Analysis] Describe your dataset: Advanced statistics

You have entered 4 experimental groups

Paired Dataset

Is your data paired? *(constricted to one variable pairing)*

- Yes
- No

Please fill in your pairing variable

<table>
<thead>
<tr>
<th>Pairing variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>variable 1</td>
</tr>
</tbody>
</table>

Covariates

Does your dataset have covariates?

- Yes
- No

Please choose if your covariates are numeric or factorial

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>factorial</td>
<td>factorial</td>
<td>factorial</td>
</tr>
<tr>
<td>numeric</td>
<td>numeric</td>
<td>numeric</td>
</tr>
</tbody>
</table>
Covariates

Does your dataset have covariates?
- Yes
- No

Please choose if your covariates are numeric or factorial:

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>factoriel</td>
<td>factoriel</td>
<td>factoriel</td>
</tr>
<tr>
<td>numeric</td>
<td>numeric</td>
<td>numeric</td>
</tr>
</tbody>
</table>

Please fill in whether or not you want to have a baseline-effect:

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>baseline</td>
<td>baseline</td>
</tr>
</tbody>
</table>

Does your model have an interaction?
- Yes
- No

Please fill the interactions in:
(Interaction-model is only possible if the covariate has a base-effect)

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>baseline</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>baseline</td>
</tr>
</tbody>
</table>

Your formula is:
COMPUTED AUTOMATICALLY

Please don't make changes or click any button while data is uploading.

Back to the previous step

Next
Covariates

Does your dataset have covariates?

- Yes
- No

Please choose if your covariates are numeric or factorial:

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>factorial</td>
<td>factorial</td>
<td>factorial</td>
</tr>
<tr>
<td>numeric</td>
<td>numeric</td>
<td>numeric</td>
</tr>
</tbody>
</table>

Please fill in whether or not you want to have a baseline-effect:

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Does your model have an interaction?

- Yes
- No

Please fill the interactions in:

Interaction-model is only possible if the covariate has a base-effect:

<table>
<thead>
<tr>
<th>Covariate 1</th>
<th>Covariate 2</th>
<th>Covariate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>baseline</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓</td>
<td>baseline</td>
</tr>
</tbody>
</table>

Your formula is:

COMPUTED AUTOMATICALLY

*Please don't make changes or click any button while data is uploading*
[Statistical Analysis] Define your dataset

Modify the settings below to customise your analysis

You have entered 4 experimental groups

Default Comparisons

Compute the default comparisons

<table>
<thead>
<tr>
<th>Group 1 compared with</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 compared with</td>
<td>Group 3</td>
</tr>
<tr>
<td>Group 4 compared with</td>
<td>Group 2</td>
</tr>
</tbody>
</table>

add a custom comparison

Custom Comparisons

Compute the custom comparisons

Example: \((-0.5) \times \text{control12} + (0.5) \times \text{control2} + (1) \times \text{treated12}\).

View the example Comparison.

<table>
<thead>
<tr>
<th>Example Comparison</th>
<th>Comparison 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>-0.5</td>
</tr>
<tr>
<td>Group 2</td>
<td>0.5</td>
</tr>
<tr>
<td>Group 3</td>
<td>1</td>
</tr>
<tr>
<td>Group 4</td>
<td></td>
</tr>
</tbody>
</table>

+ add
- remove
Biomedical Sciences

Significant genes list
Compute the significant genes list using the following thresholds:
P-values < 0.05
Log Fold-Change > 2.0
Average Expression > 5.0

Draw histograms and graphs
Visualise the data:
- Plot P-value histograms for each comparison
- Plot adapted Fold Change histograms for each comparison
- Plot a Venn Diagram
- Plot a Volcano plot

Multiple Selections testing
Run a Multiple Selection testing using the following:
- Holm
- Hochberg
- Hommel
- Bonferroni
- Benjamini & Yekutieli (SY)
- Benjamini & Hochberg (BH)
- None
Biomedical Sciences

Summary tables

Summarize the number of genes meeting each value on the following threshold lists:

- **P-values:**
  - 0.1
  - 0.05
  - 0.01
  - 0.001
  - 0.75

- **Adjusted P. values list:**
  - 0.1
  - 0.05
  - 0.01
  - 0.001
  - 0.75

- **Fold Change list:**
  - 1.1
  - 1.2
  - 1.5
  - 2.0
  - 1.75

Please don't make changes or click any button while data is uploading

Back to the previous step

RUN