RReportGenerator : Automatic Reports from Routine Statistical Analysis using R

UseR! 2008, Dortmund, Aug 11-14 2008, Focus Session “Reporting”

http://www-bio3d-igbmc.u-strasbg.fr/~wraff/

wolfgang.raffelsberger (at) igbmc.fr

Laboratoire de **B**io-Informatique et **G**énomique **I**ntégratives (LBGI)

IGBMC, Strasbourg

Automatic Data Analysis

: **Powerful Means for Statistical Analysis**

- ☺ vast collection of additional modules (CRAN, Bioconductor)
- ☹ command-line structure : "inaccessible" for Biologists,…

Our Motivation : Make R available in convenient way for

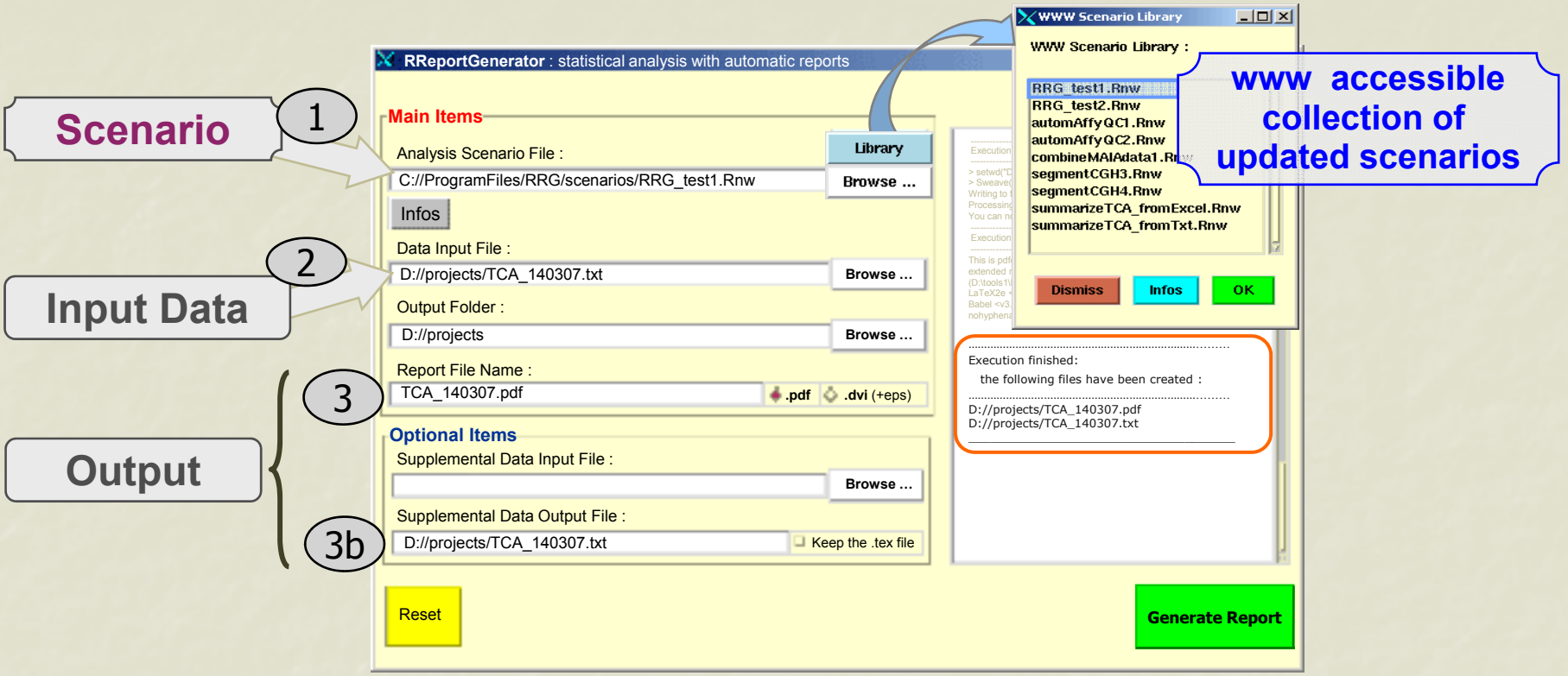
- automatic routine use
- via graphical user interface (GUI)
- allowing to access prewritten R-code
- creating reports in pdf-format

RReportGenerator : Principle

GUI **managing tasks of** :

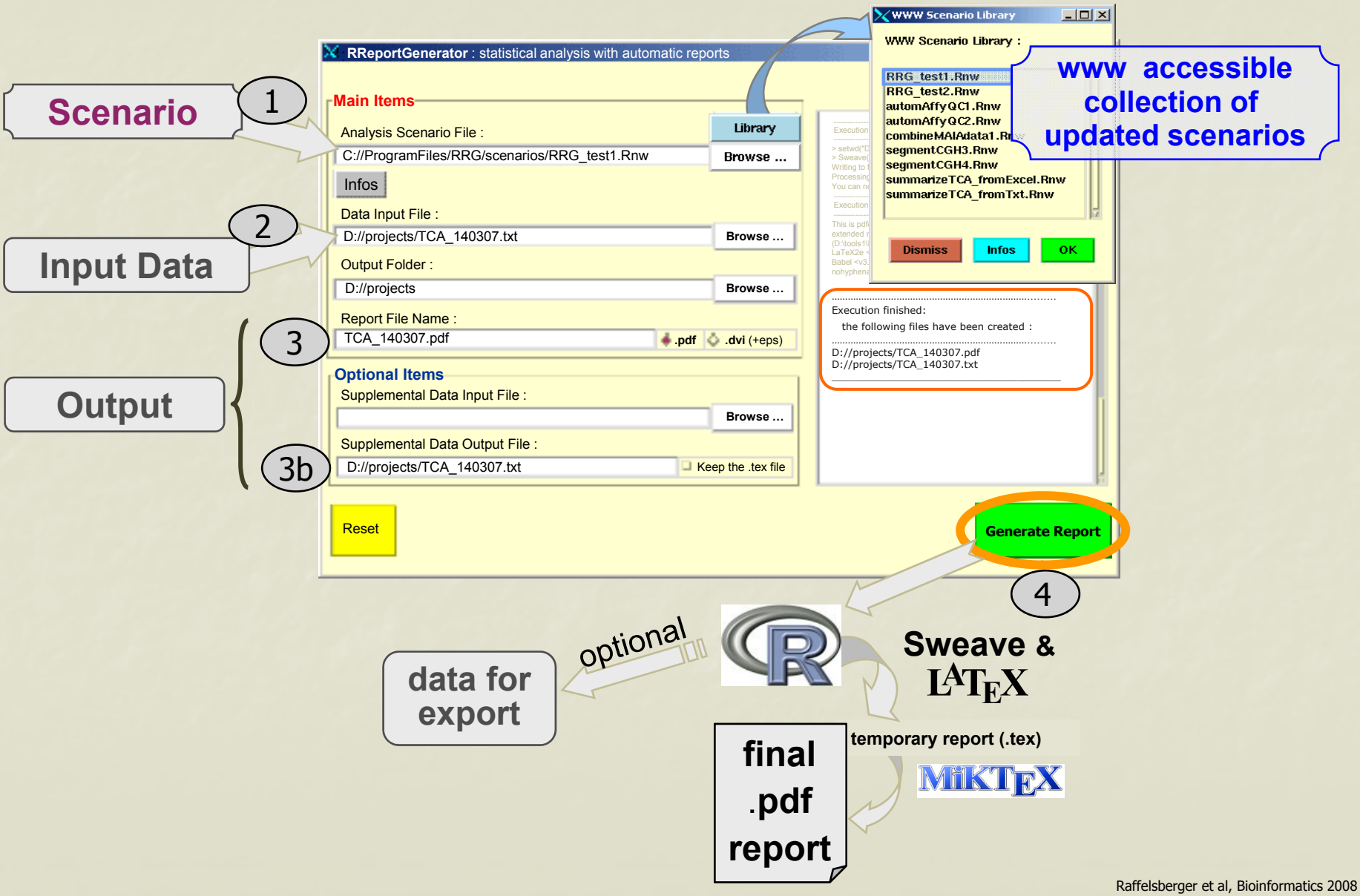
- **selecting the data to be treated**,
- **calling command-line R**,
- **passing** *Sweave* **code**
  (customize by substituting specific terms)
- **file-name & location of report**
- Customizing options : preferred search paths, …

RReportGenerator



Raffelsberger et al, Bioinformatics 2008

RReportGenerator



Raffelsberger et al, Bioinformatics 2008

Sample Code : Sweave

```
%@RRG_INFO
% Read a tab-delimited file & draw a scatter plot.
%@RRG_INFO_END

\documentclass[a4paper]{article}
\title{Minimal test scenario for RReportGenerator}
\usepackage{a4wide,Sweave}
\begin{document}
\maketitle

The file selected as 'Data Input File' is read :
<<chunk_read, echo=FALSE, print=TRUE>>=
  mydata <- read.table("<DATA_IN_FILE>")
  t.test(mydata[,1],mydata[,2],paired=TRUE)
@

<<chunk_plot1, echo=FALSE, fig=TRUE>>=
  plot(mydata[,2]~mydata[,1])
@

% If a name for 'Supplemental Data Output File' was given …
<<chunk_save, echo=F, print=F, results=hide>>=
  write.table(summary(mydata), "<DATA_OUT_FILE>", row.names=F)
@
\end{document}
```

Raffelsberger et al, Bioinformatics 2008

RReportGenerator Applications

- **Automatic analyses** for ***routine applications*** in

  - **Transcriptomics** : Quality control, data normalization

  - **Comparative Genomics Hybridization (CGH)**

    Counseling to amplified/lost regions

  - **Transfected cell array (TCA)** :

    Graphical and statistical summary

RReportGenerator

IGBMC

Examples : Transcription Profiling (Affymetrix CQ)

Affymetrix Batch Quality Control using RReportGenerator and R

August 1, 2008

This document was generated by Analysis Type File *'automAffyQC1.Rnw'*, Version: 1.2.1, a protocol for automated QC analysis for Affymetrix expression array data (using R version 2.7.0 on a x86-64-unknown-linux-gnu system).

The aim of this report is to provide information on multiple QC aspects for a set of Affymetrix arrays. The interpretation of QC parameters and QC plots should be done with care since this may lead to delicate decisions. For further information and details about the various plots shown in this report please look at the references section. This analysis protocol was written by wolfgang.raffelsberger (at) igbmc.fr, LGBI, IGBMC (Strasbourg, France).

QC analysis of 16 cel-files of type 'HG-U133A' with 22283 probesets each, found in path :

```
[1] "/genomics/g6/CELs/cancer/calvano2005/A_t0_02"
```

page 3
Both plots give a resume of the original PM values, either as boxplot or as (kernel) density estimate for the signal distribution (similar to Bioconductor-packages simpleaffy and affyQCReport).

page 4
This figure shows the QC plot from the simpleaffy package. For this purpose the 3' to 5' ratio for spiked-in and control genes (triangles for b-Actin and squares for GAPDH) are considered. The filled (blue) dot with the vertical line (heading to 0) represents the scaling factor. Finally, the percentage of 'present calls' and value of average background are shown in red on the left side. Note that the order of arrays is inversed (compared to other plots in this report).

page 5
This pair of graphs shows the RLE (top) and NUSE (bottom) plots from the affyPLM package. The RLE compares for each probeset (across all arrays) its expression value against its median across all arrays. The NUSE plot shows the standard errors for each probe-set (variability intra), standardized across all arrays.

page 6
This page shows false color images for the RLE residuals of all probes (from the affyPLM package, with red intensities corresponding to positive residuals and blue to negative residuals). Note, that smaller deviating zones have little impact on the overal results. For further information, examples and details see e.g. chapter by Bolstad in book by Gentleman *et al.*

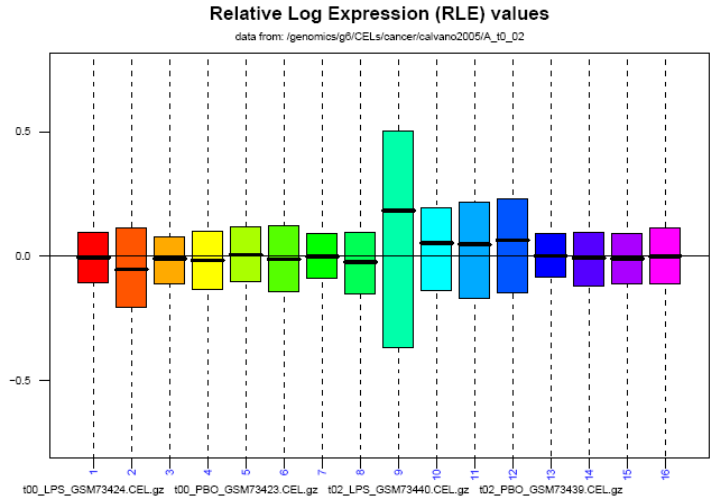
page 7
MA plots (from affyPLM package) of each array against a synthetic median array (constructed from probe-wise medians of all arrays in the current project). The red line represents a lowess fit to the scatter plot and could be helpful in indicating non-linear relationships.

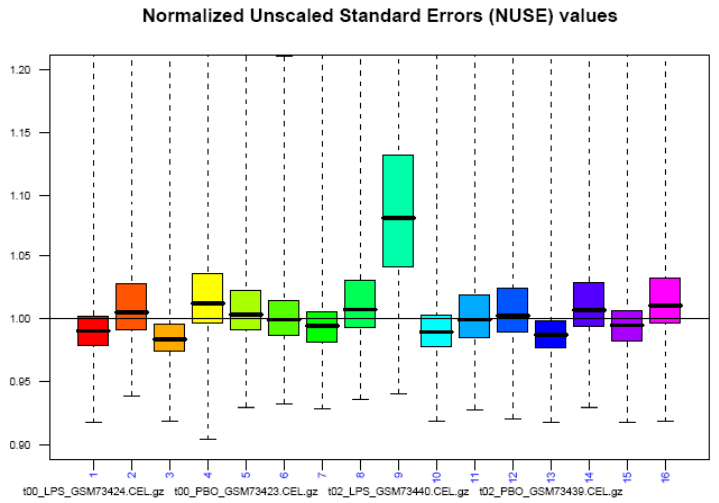
page 8
In the top part the RNA degradation plot shows the average intensity with respect to the sorted 5' to 3' position of the probes in the target-sequence. Depending on the type of microarray specific patterns can be observed (see also references Bolstad, Gentleman et al). The lower figure shows a density estimate for the signal distribution of data resulting from RMA.

page 9
Similarity between RMA summarized samples measured as Euclidean distance : Heatmap of distance values.

page 10
Principal components of RMA summarized samples. The main graphic shows the values of the first and second principal component for each of the samples. The smaller plots show the 2nd vs. 3rd as well as the 1st vs. 3rd component, the bargraph at the bottom indicates the amount of total variance captured by the individual principal components.

page 11 : References

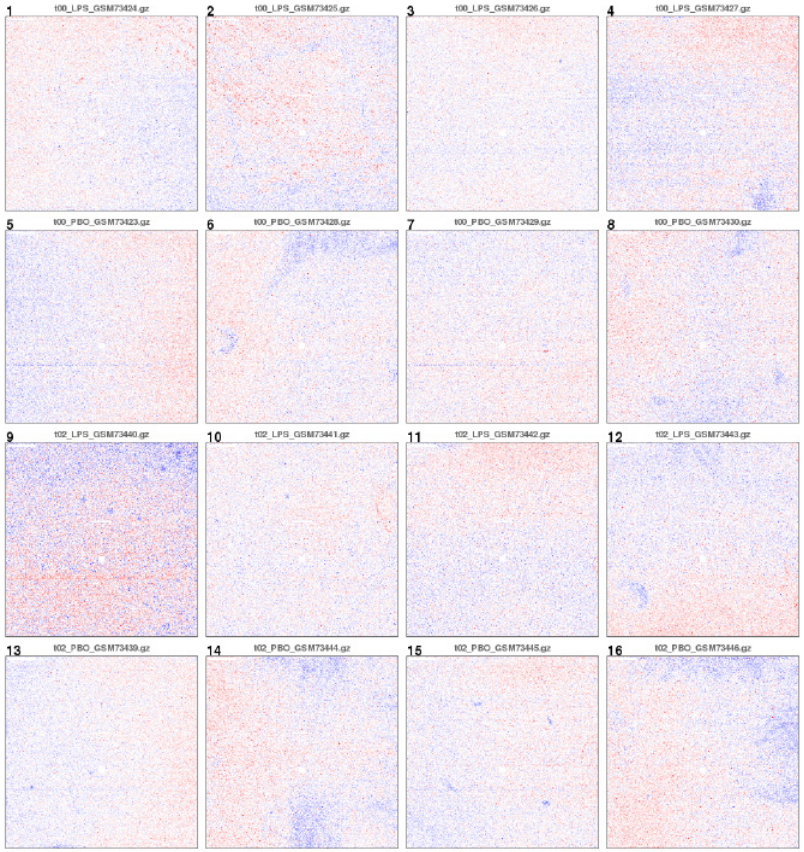




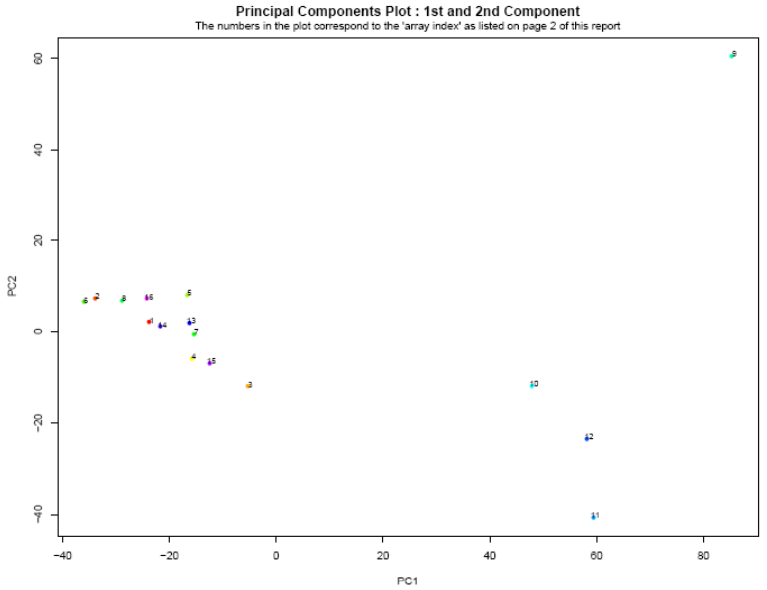
RReportGenerator

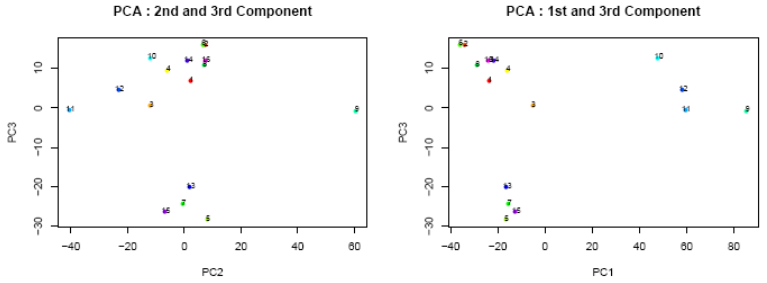
IGBMC

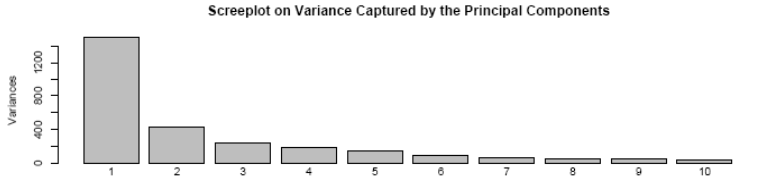
Examples : Transcription Profiling (Affymetrix QC)



6







10

RReportGenerator Availability

Windows & Linux versions

coming with installer, user-manual and basic set of analysis scenarios

under free GNU public license at :

http://www-bio3d-igbmc.u-strasbg.fr/~wraff

RReportGenerator Version 1.3.3

- **Paper Format** : Automatic switch between US and A4 paper format
- Windows : **Customize memory allocation of R**
- **Updated Analysis Scenarios**

More information :

RReportGenerator : Automatic reports from routine statistical analysis using R.
Raffelsberger W, Krause Y, Mouliner L, Kieffer D, Morand AL, Brino L, Poch O
Bioinformatics 2008, 24(2):276-278

Or : http://alnitak.u-strasbg.fr/wikili/index.php/RReportGenerator_English

Acknowledgments

LBGI

Yannick Krause Luc Moulinier
David Kieffer Guillaume Berthommier
Odile Rousselet
all LBGI members
Olivier Poch

CEBGS - IGBMC TCA-Platform

AL Morand, L Brino

Plate-forme Bioinformatique de Strasbourg

IGBMC Microarray Platform

Département de Pathologie Moléculaire

all users of RReportGenerator for feed-back !

Bela Tiwari

INSERM U682

Hôpital Hautepierre, CGE

- Collaborations :

Institut de la Vision (INSERM U592)
CHU Bordeaux
University of Zurich
University College London
CRP Santé Luxembourg
Trinity College Dublin
Lund University