# Illustrate usage of child documents in knitr

### Sarah Gerster (sarah.gerster@isb-sib.ch) and Frédéric Schütz (frederic.schutz@isb-sib.ch)

### Bioinformatics Core Facility, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

### March 20, 2014

# What are child documents?

Child documents are sub-files written in the same format (in our case markdown) as the main document, which should be included in the main document when compiling the report. Refer to http://yihui.name/knitr/demo/child/ for additional information.

In this example, we want to keep this main document brief. It should only contain this introduction, the skeleton of the rest of the file (including the child documents), the information about the R session and an appendix. Please explore all input files to see how calls to child documents can be nested.

# Generating some data

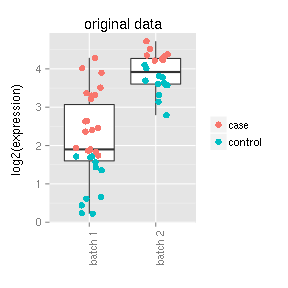
We will simulate some meaningless data to have some values to plot and represent in tables. The code to simulate the data is shown below:

set.seed(42) # for reproducibility  
## simulated values of two genes for samples from different batches  
gene1\_b1 <- sort(abs(rnorm(30, mean = 2, sd = 1)))  
gene1\_b1\_groups <- rep(c("control", "case"), times = c(11, 19))  
gene1\_b2 <- sort(abs(rnorm(20, mean = 4, sd = 0.5)))  
gene1\_b2\_groups <- rep(c("control", "case"), times = c(12, 8))  
  
## put the simulated data in a data frame  
orig\_df <- data.frame(by = as.factor(rep(c("batch 1", "batch 2"), times = c(30,   
 20))), vals = c(gene1\_b1, gene1\_b2), cat = as.factor(c(gene1\_b1\_groups,   
 gene1\_b2\_groups)))

# Plotting the data

Assume the data actually corresponds to measured intensities for a specific gene extracted from a larger gene expression data set. The samples were processed in two batches. We call a customized plotting function to get a first idea of what the data look like (code not shown in the report).

my\_R\_function(df = orig\_df, y\_lab = "log2(expression)", title = "original data")



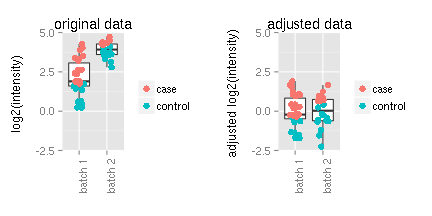
There seems to be a batch effect which we would like to remove.

## Batch effect removal with a simple z-transform:

norm\_df <- data.frame(by = as.factor(rep(c("batch 1", "batch 2"), times = c(30,   
 20))), vals = c(scale(gene1\_b1), scale(gene1\_b2)), cat = as.factor(c(gene1\_b1\_groups,   
 gene1\_b2\_groups)))

To see the effect of our batch effect removal step, we want to plot both, the original and the corrected data next to each other (code not shown in the report):

## plot original data  
plot1 <- my\_R\_function(df = orig\_df, y\_lab = "log2(intensity)", y\_lim = c(-2.5,   
 5), title = "original data")  
## plot modified data  
plot2 <- my\_R\_function(df = norm\_df, y\_lab = "adjusted log2(intensity)", y\_lim = c(-2.5,   
 5), title = "adjusted data")  
grid.arrange(plot1, plot2, ncol = 2)



# Information about the R session

Don't forget to include information about the R session (version of R and used packages) in which the current report was generated:

## R version 3.0.2 (2013-09-25)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
##   
## attached base packages:  
## [1] grid stats graphics grDevices utils datasets methods   
## [8] base   
##   
## other attached packages:  
## [1] xtable\_1.7-1 gridExtra\_0.9.1 ggplot2\_0.9.3.1 knitr\_1.5   
##   
## loaded via a namespace (and not attached):  
## [1] MASS\_7.3-29 RColorBrewer\_1.0-5 colorspace\_1.2-4   
## [4] dichromat\_2.0-0 digest\_0.6.3 evaluate\_0.5.1   
## [7] formatR\_0.10 gtable\_0.1.2 labeling\_0.2   
## [10] munsell\_0.4.2 plyr\_1.8 proto\_0.3-10   
## [13] reshape2\_1.2.2 scales\_0.2.3 stringr\_0.6.2   
## [16] tools\_3.0.2

# APPENDIX: How to generate a .docx file from this R markdown document?

Knit the .Rmd file to produce an .md file. Then, in a shell, run

$ pandoc markdown\_template\_child-docs.md -o report-via-pandoc.docx

By the way, you could also use pandoc to generate a standalone HTML file with numbered sections:

$ pandoc -Ns markdown\_template\_child-docs.md -o report-via-pandoc.html

If you want some sections to be unnumbered, for example the title, authors, affilitation and date in this document, just add {-} at the end of the corresponding lines.

If you are interested in working with markdown, we strongly encourage you to convert the files "manually" from .md to your final format via pandoc. It is much more flexible than the in-built tools in R.