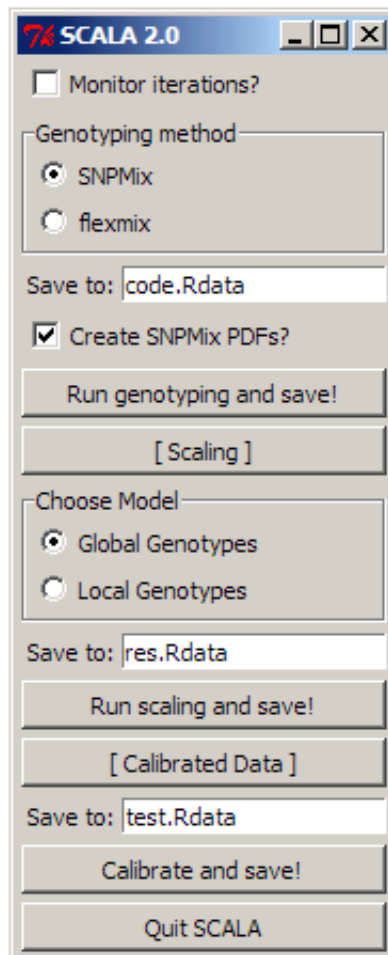


Calibrate SNP bead arrays with

SCALA

Version 2.0 for R 2.7.2 +



Author:
Ralph C.A. Rippe

Maintainer:
Ralph C.A. Rippe

Contents

1	Introduction	2
2	General usage	3
2.1	Dependencies	3
2.2	Getting SCALA to work with the GUI	3
2.2.1	Using just R	3
2.2.2	Using Tinn-R and R	3
2.3	Main program options	4
3	Genotyping with SNPmix	5
4	SCALA	6
4.1	Returned object(s)	6
5	Known bugs and Limitations	7
5.1	Bugs	7
5.2	Limitations	7

Chapter 1

Introduction

This manual accompanies the SCALA software you just downloaded. It can perform genotyping and capture the stability of SNP and sample strengths as well as quantify genotypes. To update all the readers, a very short introduction about what SNP *are*.

Single Nucleotide Polymorphisms (SNPs) are small parts (nucleotides Adenosine, Cytosine, Thymine or Guanine; A, C, T or G) of genetic code (DNA). The major part of our DNA is fixed for all individuals. However, at some position (about 1 in every 1000) one of the nucleotides can change. These are called polymorphisms and this results in a possibly variable phenotype for an individual. The underlying genotype can be homozygotic (AA, BB), heterozygotic (AB, BA) or mutant (B0, AAB, BBB etc), where both A and B can be one of the four nucleotides. For SCALA, the primary interest is in the non-mutant occurrences, since these can already have major implications in disease onset and development, so an accurate understanding is therefore very useful in e.g. pathology and medicine.

From the SCALA FileConvertor we get two vectors with fluorescence intensities (one for the A allele and one for B) per sample. All samples are then merged into matrices X and Y , containing n SNPs in the rows and m samples in the column. These intensities have an underlying genotype, which are initially unknown, but these can be reconstructed from the data in many ways.

In SCALA we incorporate two models. The first model quantifies overall genotypes (global model), whereas the second model quantifies the genotypes per SNP (local model) (Rippe et al, 2010).

The software should work with all R versions from 2.7.2 and up. If it doesn't, please contact the author at rrippe@fsw.leidenuniv.nl.

Chapter 2

General usage

Currently the program is controlled by a Graphical User Interface only.

2.1 Dependencies

The program uses functions from the `rpanel` package for the interface and `flexmix` for the genotyping (optionally, a built-in procedure is provided with `SCALA`). All other functions are included locally in the `SCALA` files (Table 2.1).

<i>file</i>	<i>description</i>
<code>GUI-SCALA.r</code>	defines the GUI for <code>SCALA</code>
<code>SCALA_mix.r</code>	the <code>SNPMix</code> functions for genotyping
<code>SCALA_model1.r</code>	<code>SCALA</code> functions for overall gamma's (global model)
<code>SCALA_model2.r</code>	<code>SCALA</code> functions for per-SNP gamma's (local model)
<code>SCALA_run.r</code>	the file that controls the actual function calls

2.2 Getting `SCALA` to work with the GUI

2.2.1 Using just R

- Put the software from the file "`SCALA.zip`" in a directory of your choice, say "`Work`" (unpack all files).
- Start up your version of R
- Change (in R) the working directory to "`Work`" (or whatever you called it)
- Load "`GUI.r`" into R **as a script**
- Run this code in R

2.2.2 Using `Tinn-R` and R

- Put the software from the file "`SCALA.zip`" in a directory of your choice, say "`Work`" (unpack all files).

- Start up Tinn-R
- Load "GUI.r"
- Start up R from the editor
- Source this code to R (see your editor manual for details).

2.3 Main program options

The program provides several options and switches, which are discussed below. At GUI startup a folder containing the input files is required. The result files are saved to this same folder.

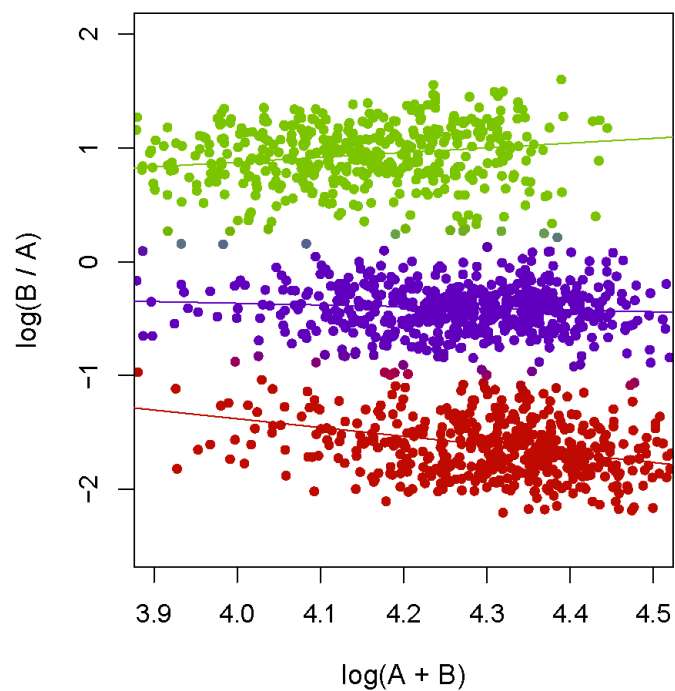
<i>file</i>	<i>description</i>
Monitor iterations	shows the global or local model iterations program in the R window
Genotyping method	use the built-in procedure SNPMix or the external library flexmix; if flexmix is not available, the program automatically switches to SNPMix.
Save to [1]	file to save the genotyping results to; hit [Enter] after changing, otherwise input is lost.
Create PDFs	save the genotyping mixture plots (per sample); only if SNPMix is used, otherwise ignored.
Run button [1]	Runs the genotyping procedure; saves the results <i>and</i> keeps them in R working memory.
Choose model	choose between Global or Local scaling model
Save to [2]	file to save the scaling results to (hit [Enter])
Run button [2]	Run the scaling model as requested and saves results; keeps results in memory for optional calibration.
Save to [3]	file to save the calibrated data to.
Calibrate and save	performs calibration (based on model chosen for scaling) and saves new data. If Global model was selected, genotype-free calibration is performed, for the local model, genotype-based calibration is performed.
Quit SCALA	ends the program and closes GUI. Keeps created data objects in memory!

Chapter 3

Genotyping with SNPmix

This is a stripped down version of the full SNPmix package ¹, which also includes output and sample plotting procedures. The version currently incorporated in SCALA only plots individual samples, but doesn't create any additional output.

The resulting genotypes are just formatted as [1, 2, 3], where 1 and 3 indicate homozygous genotypes AA and BB and 2 is the heterozygous genotype. These calls are based on the highest of their respective cluster probabilities from the mixture model.



¹as described in Rippe et al, 2010

Chapter 4

SCALA

This chapter describes the workings of the SCALA software. The input data can be created with the SCALA FileConvertor.

4.1 Returned object(s)

The available objects after an analysis depend on the chosen model. For the global model, two vectors are returned and can be used for genotype-free calibration:

- α_x (length n) contains the overall SNP levels based on matrix X
- α_y (length n) contains the overall SNP levels based on matrix Y .

Using the local model, two vectors and two matrices are returned:

- α_x (length n) contains the selected SNP calibration values based on matrix X . The selected values are based on the estimated genotypes.
- α_y (length n) contains the selected SNP calibration values based on matrix Y . The selected values are based on the estimated genotypes.
- Γ_x (dim $n \times m$) contains all of the genotype parameters per individual SNP, based on matrix X
- Γ_y (dim $n \times m$) contains all of the genotype parameters per individual SNP, based on matrix Y

Of course, you have full control over all of the objects, so you can specify any evaluation analysis you like.

Chapter 5

Known bugs and Limitations

5.1 Bugs

- SCALA 2.0 crashes if the required .Rdata file (containing the objects created by the SCALA File Converter) is not loaded.

5.2 Limitations

- This version is optimized for R version 2.7.2 and higher. It should however be downwards compatible, but this was not tested.
- The software does not allow for interactive plotting or (easy) plot modifications by the user.
- The program takes external genotypes, but only in a specific file format that is used in all SCALA-related packages. The genotypes are contained in a matrix H with n rows for the SNPs and m columns for the samples. Own calls in this format can be used by simply loading the .Rdata object into once the GUI started.