

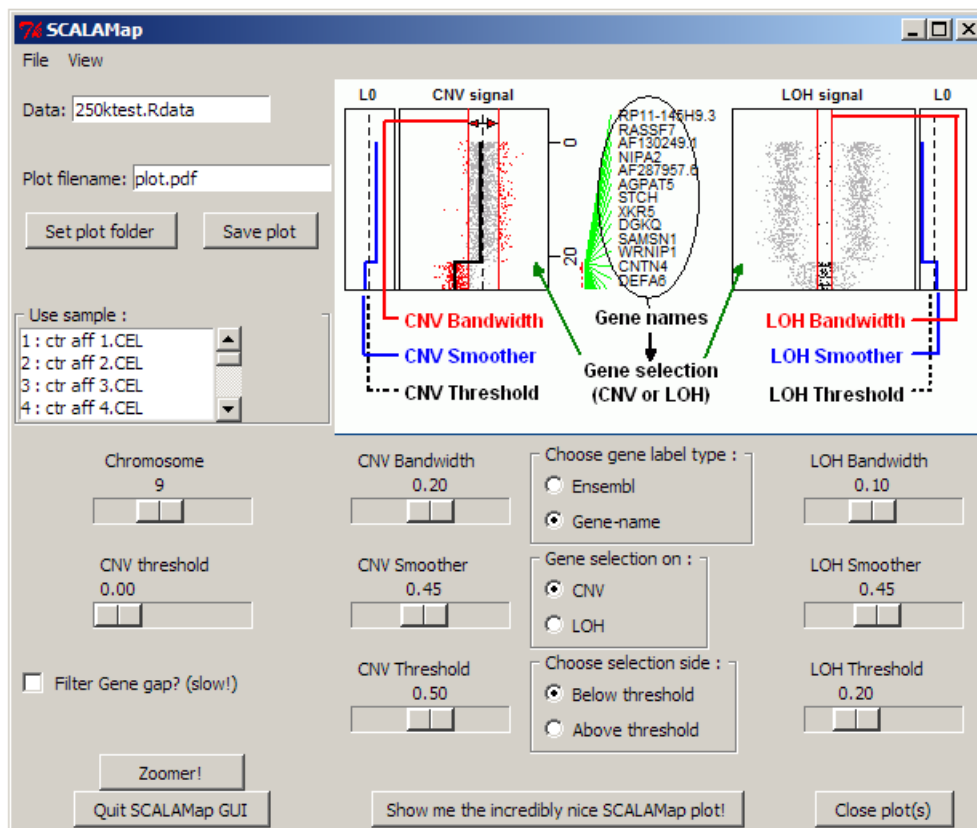
Calibrate and Map SNP arrays with

---

# SCALAMap

---

Version 2.0 for R 2.7.2 +



*Author:*  
Ralph C.A. Rippe

*Maintainer:*  
Ralph C.A. Rippe

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Installation</b>	<b>3</b>
2.1	Getting R . . . . .	3
2.2	Installing on own PCs . . . . .	3
2.3	Installing in the EMC . . . . .	3
2.3.1	Portable R . . . . .	3
2.4	Starting . . . . .	3
<b>3</b>	<b>R Packages in general</b>	<b>4</b>
3.1	Installing a package . . . . .	4
3.2	Loading a package . . . . .	4
3.3	Installing required packages for SCALA File Convertor . . . . .	5
<b>4</b>	<b>Starting the SCALA File Convertor</b>	<b>6</b>
4.1	Starting the interface . . . . .	6
4.2	File locations . . . . .	6
<b>5</b>	<b>Starting the SCALAMap software</b>	<b>7</b>
<b>6</b>	<b>Options and controls</b>	<b>8</b>
6.1	General options . . . . .	8
6.2	File locations . . . . .	9
6.3	Future improvements . . . . .	9

# Chapter 1

## Introduction

This manual accompanies the SCALAMap software package we are currently working on. It can work with (calibrated) SNP array signals and map these signals using their chromosomal location. Genes can be plotted according to anomalies in the CNV or LOH signal, detected by a *smoothed LO-penalty*.

Part of the current functionality includes sample and chromosome selection, detection bandwidth in the signals as well as detection thresholds. The main plot window is automatically generated at package startup and should NOT be closed during use of the GUI. If closed, the GUI has to be restarted. The software was originally written for R version 2.4.1, but it should also work with more recent versions (2.10.0 and above).

R can be used on any PC (e.g. at work or at home). It can even be run from a USB-device, and above all: R is free!

If something doesn't work, please contact the software maintainer at [ripppe /at/ fsw.leidenuniv.nl](mailto:ripppe@fsw.leidenuniv.nl).

## Chapter 2

# Installation

### 2.1 Getting R

The latest version of R (and older versions too) can be downloaded from the CRAN website<sup>1</sup>. Versions for PC, Mac, Linux and Unix operating systems are available. Every few months an updated version of R is released.

### 2.2 Installing on own PCs

If you download the Windows binary (.exe), double-clicking the file will start the installer. Just follow the instructions and all will be fine.

### 2.3 Installing in the EMC

Depending on the department in the EMC, R is already included in one of the startmenus. There should be a button labelled 'R'. Clicking this will ask for installation confirmation. All installation steps are performed automatically. When finished, R can be started from Programs list in windows.

#### 2.3.1 Portable R

The methods above install R on a single pc and links some filetypes (e.g. *.r*, *.Rd* and *.Rdata*) to R. However, you can also install R to a USB drive if you want. This way you can always use R directly from USB (or any other local storage device). If you choose this option, make sure that you do NOT register the filetypes for R (*.R*, *.Rda*, *.Rdata* etc) on the pc you are working on!

### 2.4 Starting

After installation, R can be run via the shortcuts (on the desktop and in Start menu), or directly by double-clicking it in the R/bin installation folder.

---

<sup>1</sup>see <http://www.r-project.org/> and click CRAN

## Chapter 3

# R Packages in general

The installed R distribution comes with some default packages that can do basic math and basic statistical models. If you want more complex models, special functions or groovy plotting procedures, you can install additional libraries (packages). These are also open-source and absolutely free. If you download packages from CRAN or BioConductor, these will usually run fine, since they have to abide by some programming rules. It is however also possible to use non-CRAN or non-BioConductor packages, which are also user-maintained but come without any guarantee.

### 3.1 Installing a package

A CRAN-package can be installed if you click *Install packages...* in R, then select a repository (a database with available packages on that specific location). From the list that comes up, select your package and click install. If the package needs or depends on any other packages, these will also be installed automatically.

If the packages you are searching for is not in the repository, it usually helps to try some other repositories (US, UK and Australia, e.g.).

### 3.2 Loading a package

Loading an R package (it doesn't matter whether it is from CRAN, BioConductor or any other source) can only be done after it has been installed to R. To load a library into R's working memory (you have to do this after every restart of R), you can click *Load package...* and select your packages, or you can type

```
library(package_name)
```

From now on, you can use the functions provided in the package. Please do **not** use

```
load(package_name)
```

since this function is used to *load working data* (like the SPSS .sav or .csv files) into memory.

### **3.3 Installing required packages for SCALA File Convertor**

1. Start R
2. Click File - Source script
3. Open script 'packageinstaller.r'.
4. The packages will be installed and loaded automatically

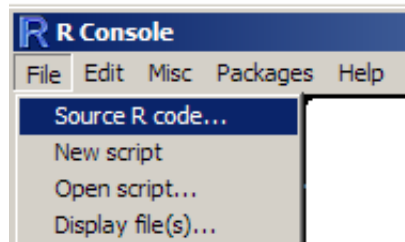
For the packages needed for SCALAMap, run the equivalent script in the SCALAMap folder you created.

## Chapter 4

# Starting the SCALA File Convertor

### 4.1 Starting the interface

- Equivalent to 'packageinstaller.r', but now use 'GUI-fileconversion.r'.
- The SCALA File Convertor interface starts



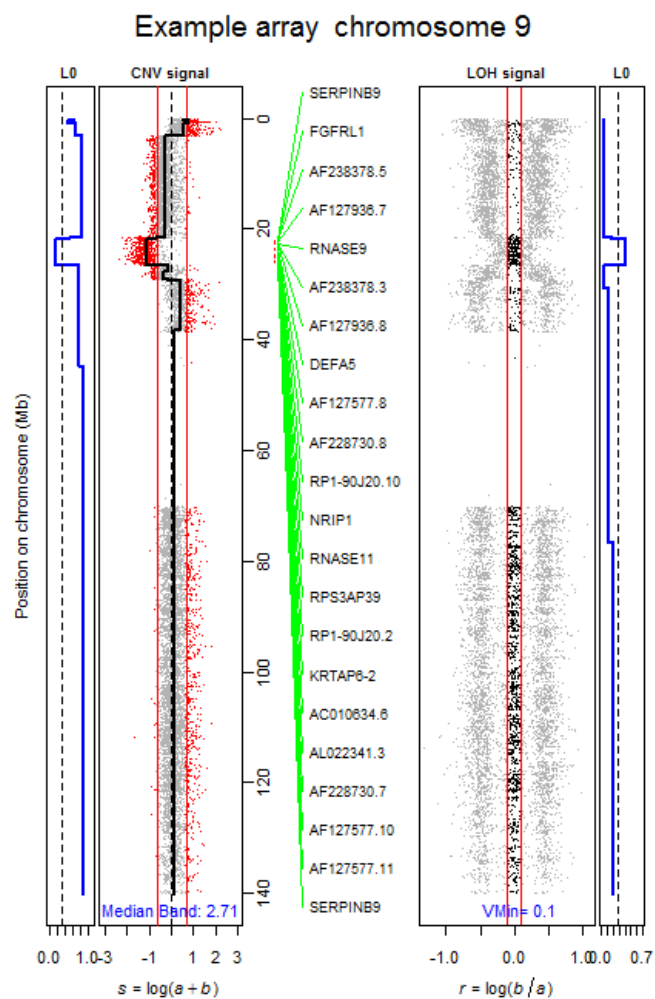
### 4.2 File locations

- The folder in which the files to be processed are located, has to be chosen at file convertor start-up.
- The folder to which the converted (and calibrated) files are written also has to be chosen at file convertor start-up.
- Choose the filename for *uncorrected* data (defaults to 'data.Rdata')
- Choose the filename for corrected data (defaults to 'datacalib.Rdata')

## Chapter 5

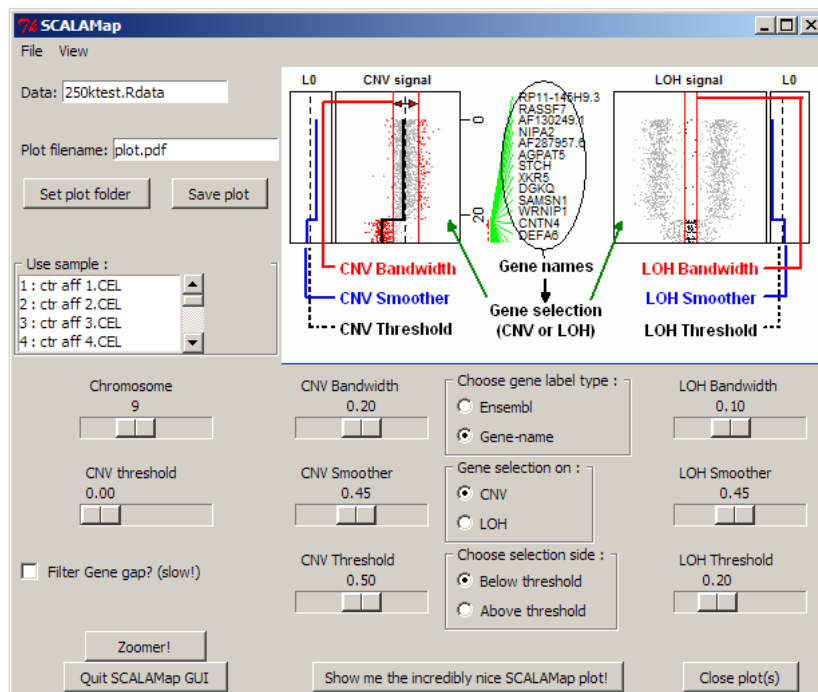
# Starting the SCALAMap software

- Start R, click File - Source R code and open 'GUI-SCALAMap.r' .
- The software now reads the appropriate data and starts the software.
- The plot is created only when the according button is clicked.



# Chapter 6

## Options and controls



### 6.1 General options

With this software you currently can:

- Select your samples
- Select the chromosome for analysis
- Set a signal quality threshold in the current sample. If set to 0, 100% of the data will be shown, if set to the maximum of 75%, the best 25% of data will be used and shown. The threshold is based on the empirical distribution function and therefore fits the active sample.

- Choose of genes of interest should be plotted based on either the CNV or LOH signal
- Set CNV and LOH thresholds (dotted line) and bandwidth (in red) for signal selection
- Choose whether you want the selection to be below or above signal threshold.
- Zoom in at particular regions of the selected chromosome in the active sample.
- Set smoothness of L0 (blue lines), which is used for gene selection. This can be done separately for the CNV and LOH signal.
- Perform selection gap filtering (takes a long time!). This should be used if two areas are selected by the L0. By default, the plot will also show all genes in between the two areas.
- Save current plot from window to PDF (to own filename and folder).

**Left** click on a slider makes the value change one 'point' in clicked direction.

**Rightclick** jumps the slider to the clicked location.

If you change a filename in a textfield, always hit 'Enter' after you are done; otherwise the filename will not be saved and used!

## 6.2 File locations

A saved plot goes into the specified folder with a user-specified filename. The same holds for the exported data go into the 'Results' folder, again with a user-specified name.

The 'Support Files' folder should remain untouched at any time; it contains the gene database as well as some dependency scripts that are used by the main program.

## 6.3 Future improvements

Functionality that is still to come:

- Improved evaluation methods for genotype calls (in the current procedures, no active genotyping is performed)
- User-specified median positioning
- Change the LOH-axis representation to a more beneficial scale.