Kernel Convolution - Statistical Method for Aberrant Region deTection (KC-SMART) - Manual

C.N. Klijn

November 9, 2007

Contents

1 Introduction

This document contains a short introduction to using the stand-alone Matlab scripts that allow KC-SMART analysis. The stand-alone version is currently in a very early stage of developement and is still on commandline base only. We are currently working on a more user-friendly GUI.

2 Installation

You have to start by installing the Matlab Compiler Runtime files. These files allow you to run our method without a full install of the Matlab software. Please run `MCRInstaller.exe` prior to running the KCSMART files. After you've installed the MCR-files you can run `KCSMART.exe` to start the analysis. It is recommended to run KCSMART in a dedicated DOS-commandline window.

3 Pre-processing

This document assumes that all normalization has already been performed on your data. The KC-SMART importer handles tab separated text files. Please make sure your data-text file is formatted according to the example given in `KCS-exampledata.txt`. All columns shown in this document are essential. The midposition of a probe is essential information. If you don't have the exact mapping (start and end positions) information for the probes in your dataset we recommend to use the following adjustments:

```
start = mid - 1bp
end = mid + 1bp
```

To make sure KC-SMART is able to handle your data, please use only numeric values in the chromosome column, and give X and Y chromosomes an appropriate number. Please enter NaN for any missing value, and make sure there are no missing values in the information columns (such as midpos or chromosome), as this will result in a failed analysis.

4 Running the program

To start the program run `KCSMART.exe` in the directory containing the datafile. Please follow the on-screen prompts. All inputs are case-sensitive and require full input. For example: when required to enter the file to be analysed, also the extension needs to be entered. If any problems arise please dont hesitate to contact the authors (`c.klijn@nki.nl`).

Output of the program consist of a 4 column, tab deliniated file. The first column is the start position of the significant region (in basepairs), the second column is the end position of the significant region. The third column is the chromosome of the significant region and the fourth column is the scale of the analysis.