MA2C Manual

Harvard School of Public Health
Dana Farber Cancer Institute

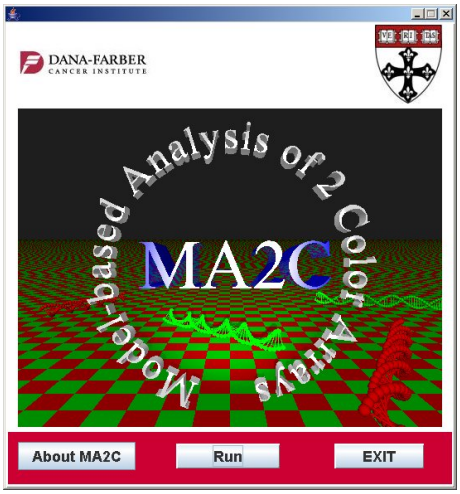
1 Installation

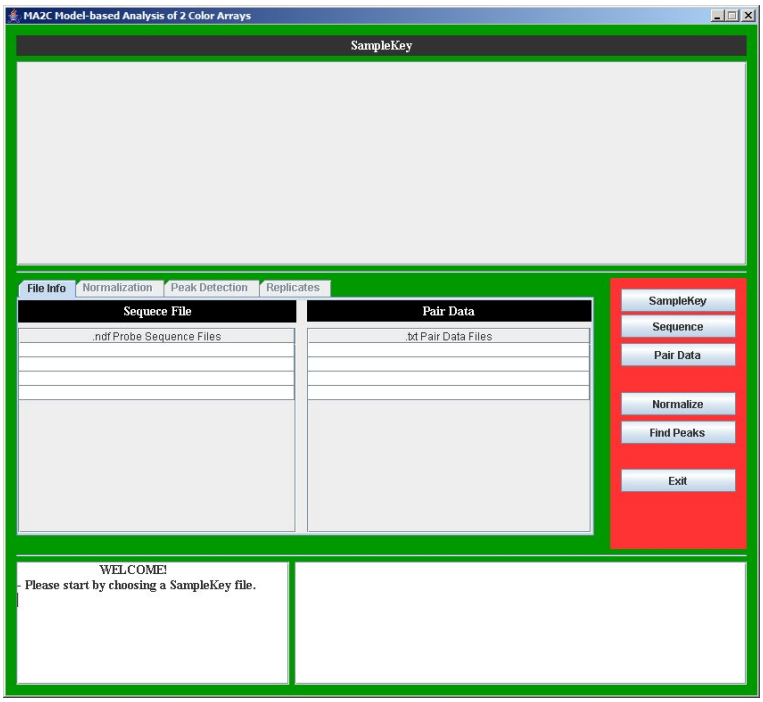
- **REQUIREMENT: You need Java Runtime Environment (JRE) 5.0 or higher.** You can download it from `http://java.sun.com/`.
- Download the package `MA2C.zip` and uncompress it using your favorite method.
- Windows: Double-click on `MA2C\dist\MA2C.bat` to launch the program. By default, the script allocates 600 MB of memory; you can change this amount by editing the script with a text editor.
- Linux, Unix or Mac: change directory to `MA2C/dist/` and execute the command

  java -Xmx600m -jar MA2C.jar

  On Mac, double-clicking on `MA2C.jar` should also start the program automatically; but if you want to pre-allocate memory, then use the command above.

2 Loading Files





- Upon launching the program, click on the "Run" button to display the main interaction panel.

- Click on the "SampleKey" button to select a `SampleKey.txt` file that came with your NimbleGen CD. Ideally, the data structure should be

  `SampleKey.txt, DesignFiles/*.ndf,*.pos, PairData/*.txt`

  If your data structure follows this pattern, then MA2C will automatically look for sequence (`.ndf, .pos`) files and pair data (`.txt`) files and display the information on the GUI.

  Some CD's arrive at your door with a quite rebellious disposition, so if some files or directories don't load properly, then check their names!

- For some reason, if the sequence and pair data folder names do not follow the above convention, then you can select the folders manually via the "Sequence" and "Pair Data" buttons. (But, we recommend that you instead change the folder names to conform to the above convention.)

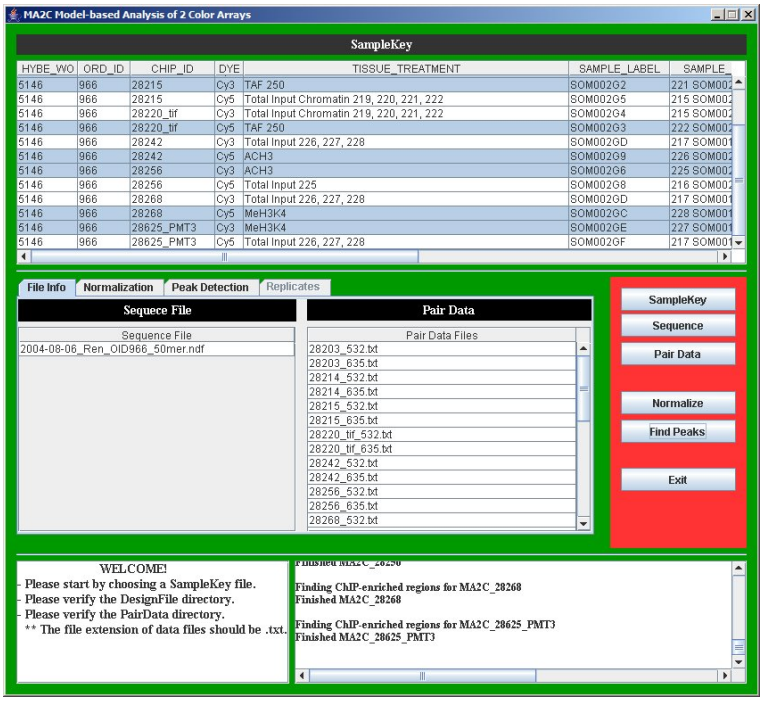
- **NOTE:** The entries in `SampleKey.txt` are tab-delimited and MUST contain the CHIP_ID, DESIGN_ID and DYE information corresponding to each experiment. If necessary, you can create or modify this file using a text editor.

- **NOTE:** It is important that all data files in `PairData/` have the `.txt` extension.

- An example of `SampleKey.txt`:

CHIP_ID	DESIGN_ID	DYE	SAMPLE_DESCRIPTION	...
1111	9876	Cy3	ER Input1	...
1111	9876	Cy5	ER IP1	...
2222	9876	Cy3	ER Input2	...
2222	9876	Cy5	ER IP2	...

Table 1: Example `SampleKey.txt`



3 Analysis

3.1 Normalization

- In the SampleKey table, Ctrl-left click to select the IP DYE for each experiment you want to analyze. In the example given in Table 1, one should thus select only the lines written in blue. It is important that you select only the IP channel. MA2C will automatically find and use the correct Input channel based on your selection of IP channel.

- Choose a normalization method in the "Normalization" tab and click on the "Normalize" button.

- MA2C will generate the following files:

    1. A `.tpmap` file will be generated in the sequence directory, and for each experiment, `MA2C_${CHIP_ID}_raw.txt` and `MA2C_${CHIP_ID}_normalized.txt` will be created in the pair data directory.

    2. A new directory called `MA2C_Output` will be created in the same parent directory as `SampleKey.txt`.

    3. A text log file and a `.pdf` probe histogram file for the run will be written to `MA2C_Output`.

3.2 Peak Detection

- After normalization, check the "Peak Detection" tab for available options, and click on the "Find Peaks" button. You don't need to renormalize the data each time you want to find peaks using different options.
- MA2C will generate the following files in `MA2C_Output`:
  1. `MA2C_${CHIP_ID}.bed`, `MA2C_${CHIP_ID}.gff`, `MA2C_${CHIP_ID}.xls`, `MA2C_${CHIP_ID}.pdf`, and `MA2C_${CHIP_ID}_FDRtable.txt` for peak locations and quality control.

     In `MA2C_${CHIP_ID}.xls`, the column "Peak_Pos" contains the location of peak with respect to the peak starting position; the peak location is computed to be the average between the geometric center and the peak summit. In some cases, two enriched regions may overlap and cause problems with some genome browsers; in `MA2C_${CHIP_ID}.bed` and `MA2C_${CHIP_ID}.gff`, we resolve this issue by truncating the overlapping left-edge, if necessary. `MA2C_${CHIP_ID}.xls` contains the untruncated coordinates for further processing.
  2. `MA2C_${CHIP_ID}.ratio.bar` containing normalized probe ratios.
  3. `MA2C_${CHIP_ID}.MA2Cscore.bar` containing MA2C scores.
- You can load the `.bed` and `.bar` files into Affymetrix's Integrated Genome Browser (IGB).

3.3 Histogram Panels

- Left-click and drag to zoom in.
- Right-click and choose "Save as" to save the image as a PNG file.

4 Agilent Data

To convert Agilent data to a format compatible with MA2C:

- Create a `.ndf` file in `DesignFiles/` containing: DESIGN_ID, SEQ_ID, PROBE_ID, PROBE_SEQUENCE, POSITION, X, Y, where X and Y are array coordinates and POSITION genomic coordinate.
- Create a `.pos` file in `DesignFiles/` containing: CHROMOSOME, PROBE_ID, SEQ_ID, POSITION
- Create `IMAGE_ID.txt` files in `PairData/` containing: IMAGE_ID, X, Y, PM. PM is the background subtracted hybridization intensity, such as F635 Median − B635 Median for Cy5. IMAGE_ID should be of the form 12345_532 and 12345_635, for Cy3 and Cy5, respectively

5 Contact Information

Jun S. Song (jssong@jimmy.harvard.edu)
W. Evan Johnson (wjohnson@hsph.harvard.edu)
Xiaopeng Zhu (zhuxp@jimmy.harvard.edu)
X. Shirley Liu (xsliu@jimmy.harvard.edu)