Design File

Lori Shepherd and Jonathan Dare

July 10, 2007

Contents

Every sample must have a design specified in the inventory. This design may be a character that simply represents a particular design of a chip or it may reference a design file. The design file must be in a certain format and contain specific data. The purpose of this example is to build a design file in the correct format.

The design file will give information on how the chip was printed. This includes data such as the BAC/Probe printed at each spot, the plate, platerow and platecol the spot occured, spot repetition, and which pin of the pin group performed the spotting.

This file may be created with any text editor such as emacs, excel, etc. The columns MUST only be seperated by a comma (therefore when completing the file in programs such as excel the user must save as a comma delimited text file). The first line of data should be names of columns. It is vital to have this header line. The file may contain any number of columns, however, there are two columns that must be in the file as the first two columns of data. These columns are MasterRow and MasterCol.

MasterRow and MasterCol refer to the spot location relative to the entire chip. Microarray chips are spotted using an arrayer. This arrayer has a group of pins that spot the chip. This creates a grid within a grid environment. The inner grid represents the pins; it will be the dimension of the pin group. The outter grid represents the group of pins as a whole, and where the arrayer spots the chip. If we say the pin group as a whole is a 'Block', and that Block spotted the chip 24 times, it could be in a 12 x 2 pattern, 4 x 6 pattern, etc. We will use a 4 x 6 grid. The chip would have a pattern like the following:

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	7	8	9	10	11	12
3	13	14	15	16	17	18
4	19	20	21	22	23	24

Each one of these 'Block' locations is a grid. The inner grid is the dimensions of the pin group used to spot. If the arrayer used a 18 x 12 pin group, each

block (1 thru 24) has 216 spots arranged in an 18 X 12 pattern. Pick a block in the above table, 1,9, 22, they all have the following pattern as the inner grid:

	1	2	3	...	11	12
1	1	2	3	...	11	12
2	13	14	15	...	23	24
3	25	26		...	35	36
.						
.						
.				...		
17	193	194	195	...	203	204
18	205	206	207	...	215	216

If these were the dimensions of a chip, 6 x 4 and 18 x 12, there would be a total of 5184 spots per chip. The dimensions of all spots over the entire chip would be a 72 x 72 [(18 x 4) x (12 x 6)].

MasterRow and MasterCol, therefore would each have numeric values one to 72 representing the location over the entire chip. If we kept the pin group as an 18 x 12 set but the arrayer spotted a 12 x 4 matrix grid, the entire chip would be a 216 x 48 [(18x12) x (12x4)] grid. MasterRow in this case would consist of numbers from one to 216 and MasterCol would be from one to 48. Again, MasterRow and MasterCol must be the first two columns of the design file and the names of these columns must match exactly including capitalization ("MasterRow","MasterCol",).

The design file may contain any number of additional columns. It should include columns that the user wishes to correct for; some examples may be plate, pin, platerow, platecol, and/or repetition. This allows the user to correct for plate bias, pin bias, etc. Data in these columns should be equivalent to what was used at each particular grid spot.

The design file may be complete or incomplete. The aCGHplus package does not require what we call a complete design file. Some chip locations, such as corners of grids etc., may not be used for quality control purposes. Some spots also may be used as controls and be left blank or have water in them. In these cases the user may opt to not include these spots in the design file. The package does contain a few wrappers to other bioconductor packages, such as marray, which do however operate on a complete design file. By complete, we mean that every possible location on the chip; in the example above all 5184 different spot locations should be listed in the design file including those that were not used or left blank, etc. The user may use the wrappers with an incomplete file, however functionality for the packages will be limited or may not work at all depending on the package.