Using R for the Analysis of BeadArray Microarray Experiments

Mark Dunning, Natalie P Thorne, Michael Smith, Isabelle Camilier, Simon Tavaré
Department of Oncology, University of Cambridge, England

The development of diseases such as cancer is caused by fundamental changes in the function and morphology of cells in an organism. These changes are governed by the regulation of proteins produced in the cell, which are in turn regulated by the amount of expression of particular genes. Therefore it is of great interest to medical researchers to be able to compare the expression levels of genes between different conditions or samples. In recent years, the technology of microarrays has made it feasible to measure the gene expression levels of many thousands of genes in cells taken from different samples. The sheer volume of data produced even by a simple microarray experiment has led to a new inter-disciplinary subject within Bioinformatics which uses the expertise of biologists, computer scientists and mathematicians to be able to manipulate, analyse and draw meaningful biological conclusions from microarray experiments.

Illumina have created an alternative microarray technology (BeadArray) based on randomly arranged beads, each of which carries copies of a gene-specific probe. Random sampling from an initial pool of beads produces an array containing, on average, 30 randomly positioned replicates of each probe type. This degree of replication makes the gene expression levels obtained using BeadArrays more robust whilst spatial effects do not have such a detrimental effect as they do with conventional arrays, where there is often little or no replication of probes over an array. BeadArrays are already being used in a number of high-throughput experiments (eg www.hapmap.org).

Until now, analysis of BeadArray data was carried out by using Illumina's own software package and therefore did not utilise the wide range of Bioinformatic tools already available via the Bioconductor website (www.bioconductor.org). Also, the data output from this software only gives a single measurement for each bead type on an array, thus losing information about the replicates. The intention of our project was to create an R package (*beadarray*) for the analysis of BeadArray data incorporating ideas from existing microarray analysis R packages. We aimed to provide a flexible and extendable means of analysing BeadArray data both for our own research purposes and for the benefit of other users of BeadArray technology. The *beadarray* package also gives users access to the full dataset for each array.

We will describe the methods available in our R library for the reconstruction and analysis of BeadArray data. We will demonstrate the low variability and high reproducibility of data generated by BeadArray experiments along with methods for quality control. An important step in the analysis of microarray data is normalisation in which data from separate experiments are made comparable by removing any systematic variation between arrays. We will present our results on investigations in comparing and assessing the performance of various normalisation approaches (including the different preprocessing and background correction steps) for BeadArray data.

The latest version of the *beadarray* package is available now at
http://www.bioconductor.org/packages/bioc/1.8/html/beadarray.html