BeadDataPackR: A Compression Tool For Raw Illumina Microarray Data

Mike L. Smith[1,*], Andy G. Lynch[1]

1. Department of Oncology, University of Cambridge, Cancer Research UK Cambridge Research Institute, Robinson Way, Cambridge CB2 0RE, UK

*Contact author: mike.l.smith@cancer.org.uk

Microarray technologies have been an increasingly important tool in biological research in the last decade, and a number of initiatives have sought to stress the importance of the provision and sharing of raw microarray data. Illumina BeadArrays provide a particular problem in this regard, as their random construction and high number of replicate probes simultaneously adds value to analysis of the raw data and obstructs the sharing of those data, with many of the standard repositories having no facilities for storing the raw data. This leaves the burden of distributing such data with the individual researcher. The **BeadDataPackR** package provides an interface to a novel compression scheme for raw Illumina BeadArray data, designed to ease the storage and bandwith concerns that providing access to such data brings. It offers two key advantages over off-the-peg compression tools. First it uses knowledge of the data formats to achieve greater compression than other approaches, and second it does not need to be decompressed for analysis, but rather the values held within can be directly accessed by existing tools, such as the popular **beadarray** package.

References

M.L.Smith and A.G.Lynch (2010). BeadDataPackR: A Tool to Facilitate the Sharing of Raw Data from Illumina BeadArray Studies. *Cancer Informatics 9*, 217–227.