Genotype Calls and Copy Number Analysis with the oligo Package

Benilton S. Carvalho[1], Rafael A. Irizarry[1]

[1] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, USA

The fast development of high-throughput technologies enabled us to access to more than 1.5 million markers for genetic variation on a human subject at a time. Under such circumstance, the need for a unified framework to manage the data provided by experiments based on microarrays and methods that would help the user analyse such big datasets in the most efficient manner has increased significantly. We developed an R/BioConductor package, `oligo`, structured to handle all array designs by Affymetrix Inc. and NimbleGen Systems Inc. The package implements preprocessing methodologies and the genotyping algorithm CRLMM for Affymetrix SNP arrays (50K, 250K, SNP 5.0 and SNP 6.0). In this work we demonstrate the main features of the package, focusing on applications using SNP arrays. We present results that attest how CRLMM outperforms standard genotyping tools and strategies for copy number analyses using our software and Affymetrix arrays, allowing the user to identify chromosomal indels.

References

[1] Carvalho B, Bengtsson H Speed TP and Irizarry RA (2007) Exploration, normalization, and genotype calls of high density oligonucleotide SNP array data. Biostatistics, 2007 Apr;8(2):485-99. Epub 2006 Dec 22