BIC-seq: a fast algorithm for detection of copy number alterations based on high-throughput sequencing data

Genome **Biology**

Open Access

BIC-seq: a fast algorithm for detection of copy number alterations based on high-throughput sequencing data

Ruibin Xi[1], Joe Luquette[1], Angela Hadjipanayis[2], Tae-Min Kim[1], Peter J Park[1,3,4*]

From Beyond the Genome: The true gene count, human evolution and disease genomics
Boston, MA, USA. 11-13 October 2010

DNA copy number alterations (CNA), which are amplifications and deletions of certain regions in the genome, play an important role in the pathogenesis of cancer and have been shown to be associated with other diseases such as autism, schizophrenia and obesity. Next-generation sequencing technologies provide an opportunity to identify CNA regions with unprecedented accuracy. We developed a CNA detection algorithm based on single-end whole-genome sequencing data for samples with matched controls. This algorithm, called BIC-seq, can accurately and efficiently identify the CNAs via minimizing the Bayesian information criterion (BIC). We applied BIC-seq on a glioblastoma multiforme (GBM) tumor genome from the Cancer Genome Atlas (TCGA) project and identified hundreds of CNVs, some were as small as 10 bp. We compared these CNAs with those detected using the array Comparative Genomic Hybridization (CGH) platforms and found that about one third were 'missed' by the array-CGH platforms, most of which were CNAs less than 10 kb. We selected 17 of the CNAs not detected by the array-based platforms for validation, ranging from 110 bp to 14 kb, and found that 15 of them are true CNAs. We further extended BIC-seq to the multi-sample case to identify recurrent CNAs in across multiple tumor genomes.

Author details
[1]Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, Massachusetts 02115, USA. [2]Harvard Medical School, 77 Louis Pasteur Avenue, Boston, Massachusetts 02115, USA. [3]Department of Medicine, Brigham and Women's Hospital, 77 Louis Pasteur Avenue, Boston, Massachusetts 02115, USA. [4]Harvard-MIT Health Sciences and Technology Informatics Program at Children's Hospital, 300 Longwood Ave., Boston, Massachusetts 02115, USA.

[1]Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, Massachusetts 02115, USA
Full list of author information is available at the end of the article