Implementing qDNAseq in Galaxy: a whole genome sequencing copy number analysis tool

Stef van Lieshout, Ilari Scheinin, Daoud Sie, Remond J.A. Fijneman, Bauke Ylstra
Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands

DNA copy number aberrations are a hallmark of cancer and can be quantified by shallow whole-genome sequencing (WGS). A robust method has been developed [(1)] that detects copy number aberrations by binning and counting sequence reads in non-overlapping windows (usually of 15kb). Then a combined LOESS correction for mappability and GC content is applied followed by excluding genomic regions from both ENCODE project blacklists and a novel blacklist based on sequence depth of 38 individuals from the 1000 Genomes project.

The procedure is available as a Bioconductor package, QDNAseq [(2)]. The accompanying Galaxy tool uses the popular BAM format as input and reports results in a clear and concise HTML based view within Galaxy itself. Various output formats can be downloaded, including an R data structure file for downstream analysis and a Zipped archive with all the output together.

Due to precalculated bin annotations, current limitations include the support for one genome build (GRCh37/hg19) and one sequencing type (50bp single read). Additional dedicated tools will handle these challenges and future plans include the addition of different strategies for segmenting and calling the copy number data.

Funding was supported by the Center for Translational Molecular Medicine, Translational research IT project (CTMM TraIT).

(1) Ilari Scheinin, Daoud Sie et al. DNA copy number analysis of fresh and formalin-fixed specimens by whole-genome sequencing: Improved correction of systematic biases and exclusion of problematic regions, (submitted).
(2) http://www.bioconductor.org/packages/release/bioc/html/QDNAseq.html