ParallelCBS Documentation

Description: Runs the CBS algorithm on multiple samples in parallel.

Author: Marc-Danie Nazaire, gp-help@broadinstitute.org

Summary

Segments DNA copy number data into regions of estimated equal copy number using circular binary segmentation (CBS). This module works by splitting each sample into a separate file and then launching parallel runs of the CBS GenePattern module on each of these files. Afterwards, the results of each CBS run are merged into one file.

References

1. Olshen, A. B., Venkatraman, E. S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. Biostatistics, 5:557–72. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. (1998). High resolution analysis of dna copy number variation using comparative genomic hybridization to microar- rays. Nat. Genet., 20:207–211.
2. Lucito, R., Healey, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., West, J., Rostan, S., Nguyen, K., Powers, S., Ye, K., Olshen, A., Venkatraman, E., Norton, L., and Wigler, M. (2003). Representational oligonucleotide microarray analysis: a high resolution method to detect genome copy number variation. Nat. Genet., 13:2291–305.
3. Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., S, S. L., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). Assembly of microarrays for genome-wide measurement of dna copy number. Nat. Genet., 29:263–4. Venkatraman, E. S. and Olshen, A. B. (2007). A faster circular binary segmentation algorithm for the analysis of array cgh data. Bioinformatics, 23:657–63.

Parameters

Name	Description
input file	A copy number file.
output file	The name of the output file.

Output Files

1. <output file>
   A segmentation file. This is tab delimited text file with six columns: the sample id, the chromosome number, the map position of the start of the segment, the map position of the end of the segment, the number of markers in the segment, and the average value in the segment.

Platform Dependencies

Module type: SNP Analysis

CPU type: any

OS: any

Language: R 2.14

GenePattern Module Version Notes

Date	Version	Description
9/14/2012	1	Initial version.