aCGH Segmentation: analysis of a male breast cancer dataset

Iannelli G, Mangia A, Chiarappa P, Paradiso A, Tommasi S

Motivation

In the last years DNA Copy Number has achieved a new role as diagnostic and therapy determinant for cancer. Comparative genomic hybridization (CGH) is a technique by which it is possible to detect and map genetic changes that involve gain or loss of segments of genomic DNA. Microarray formats of CGH provide copy number information at thousands of locations distributed throughout the genome. The aim of this study was to find the most performing algorithm to systematically identify deleted or amplified genomic regions in a set of tumors. This algorithm should have an accurate methodology for detecting the breakpoints delimiting altered regions in genomic patterns.

Methods

The pathological tissues of 25 male breast cancer patients enrolled at the NCC of Bari were hybridized on high-density oligonucleotide aCGH arrays. aCGH was performed using the Agilent Human Genome CGH Microarray Kit (Agilent Technologies, Santa Clara, California, USA) with a resolution of about 100 kb. Data analysis was performed with Nexus Copy Number 5.0 software (Biodiscovery, Inc., El Segundo, CA, USA). This software uses the Rank Segmentation algorithm, a proprietary variation much faster at processing, on Circular Binary Segmentation (CBS) together with the statistical Significance Testing for Aberrant Copy number (STAC) method, to identify non random genomic amplifications and deletions across multiple experiments. We used the modified CBS algorithm to improve the processing speed. It uses a normal distribution function to test for changing points as opposed to the original algorithm based on non-parametric permutation. It is a recursive algorithm that keeps dividing the genome into smaller and smaller segments until no region can be further segmented. The result is to segment the genome into clusters of uniform ratios. The algorithm has a single parameter called Significance Threshold that controls if a region is to be segmented out or not. At the completion of the segmentation process, the entire genome can be represented as a series of segments and each segment having a cluster value which is the median log-ratio value of all the probes in that region. The calling algorithm then uses the cluster values and the user defined threshold to establish regions of copy number variations. According to this algorithm, two regions are considered different when p values are lower than 1.0E-6. Genomic regions of gains were defined as averaged log2 CGH fluorescence ratio & 0.2 and losses as averaged log2 CGH

National Cancer Centre “Giovanni Paolo II” - Bari (I)

ratio ! % 0.2. Frequency significance testing, instead, helps to identify the areas of the genome where there is a statistically significant high frequency of aberrations over the baseline level of aberrations. The algorithm identifies a set of aberrations that are stacked on top of each other such that it would not occur randomly. To integrate our analysis, we compare our dataset with a female breast cancer dataset deposited with the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/projects/geo//query/acc.cqi?acc5GSE12659), applying the same algorithms.

Results

All the 25 males displayed chromosomal instability: 760 gains, 711 losses, 223 high copy gains, 8 homozygous copy losses. Average of 68 aberrations were found in each patient on this study. Amplifications were more frequent in chromosome 7 (50%), 11 (50%), 16 (40%) and X (70%), while chromosomal deletions were on chromosome 1 (60%), 2 (70%), 4 (50%), 5 (40%), 14 (53.33%), 15 (46.66%), 19 (40%), Y (40%). The aberrations were unequally distributed among the patients with 4 patients having less than 10 aberrations. The number of aberrations doesn't seem to depend on the age of the patients. In the female dataset we found 738 loss, 804 gains, 228 high copy gains and 5 homozygous copy losses. Average of 110.93 aberrations were found in each patient, the amplification were more recurrent on chromosome 1 (85.5%), 2 (68.75%), 3 (68.75%), 8 (43.75%), 11 (56.25%), 17 (56.25%), 20(50%), while deletions in chromosome 2 (43.75%), 3 (62.5%), 7 (50%), 15 (81.25%), 16 (50%), 17 (68.75%). The genomic aberration profile is quite different among the two datasets with very few common regions among male and female. In our experience, the modified CBS algorithm together with the statistical Significance Testing for Aberrant Copy number can be used as a valid evaluation method in copy number variations detection methods.

Contact e-mail

s.tommasi@oncologico.bari.it