ABSTRACT

Summary: We developed an algorithm named GEAR (genomic enrichment analysis of regional DNA copy number changes) for functional interpretation of genome-wide DNA copy number changes identified by array-based comparative genomic hybridization. GEAR selects two types of chromosomal alterations with potential biological relevance, i.e. recurrent and phenotype-specific alterations. Then it performs functional enrichment analysis using a priori selected functional gene sets to identify primary and clinical genomic signatures. The genomic signatures identified by GEAR represent functionally coordinated genomic changes, which can provide clues on the underlying molecular mechanisms related to the phenotypes of interest. GEAR can help the identification of key molecular functions that are activated or repressed in the tumor genomes leading to the improved understanding on the tumor biology.

Availability: GEAR software is available with online manual in the website, http://www.systemsbiology.co.kr/GEAR/.

Contact: yejun@catholic.ac.kr

1 BACKGROUND

Recently developed array-based comparative genomic hybridization (array-CGH) is one of the most advanced genome-wide screening technologies of chromosomal alterations (Albertson and Pinkel, 2003). In spite of the promising potential, the complex and prevalent nature of chromosomal alterations often makes it difficult to identify biologically relevant changes and to draw a functional interpretation. Although there have been relative successes to search master genes from local chromosomal alterations, it is highly challenging to develop a more integrative and function-oriented analytic method for high-throughput data sets (Rhodes and Chinnaiyan, 2005). From this perspective, one promising method is functional enrichment or pathway analysis using a priori selected functional gene sets, by which the large-scale gene expression profiles are interpreted as function-related coordinated expression changes (Curtis et al., 2005). Theoretically, this analytical method could be applied to other types of high-throughput data such as genome-wide DNA copy number alterations of array-CGH. Therefore, we developed a novel algorithm, GEAR (genomic enrichment analysis of regional DNA copy number changes), that selects biologically relevant copy number changes and performs gene functional enrichment analysis using predefined functional gene sets.

2 DESCRIPTION

The GEAR algorithm can be applied to both large insert clone array and oligo-array based whole-genome copy number data. GEAR defines two types of chromosomal alterations; one is a recurrent alteration shared by a significant number of samples and the other one is a phenotype-specific alteration that occurs more often in a certain phenotypic subclass. Then, GEAR measures the significance for the enrichment of the functional gene sets in those two types of alterations. The significantly enriched gene sets, in recurrent and phenotype-specific alterations are termed primary- and clinical-genomic signatures, respectively. The major functions of GEAR algorithm are implemented in VB.NET program running on Microsoft Windows machine with user-friendly graphic interface. The major steps of the GEAR algorithm are as follows;

- Determination of significantly recurrent chromosomal alterations.
- Determination of phenotype-specific chromosomal alterations.
- Mapping genes on respective probes or genomic regions.
- Enrichment analysis using functionally-annotated gene sets.
- Identification of primary and clinical genomic signatures.

(1) The recurrent chromosomal alterations are determined by two distinct methods; one is based on the alteration profile of individual probes and the other based on SW-ARRAY algorithm. In individual probe-based method, GEAR first determines chromosomal gains or losses for each probe using user-defined cutoff values.

*To whom correspondence should be addressed.
The recurrent chromosomal alterations are assigned to the probes whose alteration frequencies exceed the position-independent threshold of alteration frequency. The threshold can be determined arbitrarily by user or calculated under the null hypothesis assuming that the observed alteration frequencies are position-independent and constant over the genome, as described previously (Bilke et al., 2005). GEAR also uses SW-ARRAY algorithm that has been recently proposed as robust and reliable method for identifying genomic alterations (Price et al., 2005).

(2) Using the clinical information of samples, GEAR defines the phenotype-specific chromosomal alterations that occur more frequently in one of two dichotomous clinical conditions (i.e. good versus poor survival). The phenotype-specific alterations are determined using hypergeometric distribution under the user-defined significance level (Bilke et al., 2005) or SW-ARRAY algorithm.

(3) Genomic signatures are identified by functional enrichment analysis based on hyper-geometric distribution. Genes in a functional gene set are mapped on the respective probes in the array. The significantly enriched gene sets in the alterations of interest are identified as genomic signature.

(4) The gene sets significantly enriched in recurrent and phenotype-specific chromosomal alterations are termed primary and clinical signatures, respectively. This functional enrichment analysis algorithm can be also applied to user-defined set of alterations.

(5) GEAR also provides options for multiple tests adjustment of the significance level. The calculated significance can be adjusted using Bonferroni correction or false discovery rate (FDR) controlling method (Rhodes et al., 2004).

(6) GEAR algorithm is implemented in graphic interface to assure user-friendliness (Fig. 1). A detailed description of the entire procedures along with statistical consideration is provided in the online manual available in the website; http://www.systemsbiology.co.kr/GEAR/GEARmanual.pdf.

ACKNOWLEDGEMENTS
This work is supported by FG06-12-01 of the 21C Frontier Functional Human Genome Project from the Ministry of Science and Technology in Korea and a grant of the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (0405-BC02-0604-0004). We thank Seon Hee Yim for critical review of the manuscript.

Conflict of Interest: none declared.

REFERENCES