

Genetics and population analysis

## ArrayCyGHt: a web application for analysis and visualization of array-CGH data

Su Young Kim<sup>1,2</sup>, Suk Woo Nam<sup>1,2</sup>, Sug Hyung Lee<sup>1,2</sup>, Won Sang Park<sup>1,2</sup>,  
Nam Jin Yoo<sup>1,2</sup>, Jung Young Lee<sup>1,2</sup> and Yeun-Jun Chung<sup>3,\*</sup>

<sup>1</sup>Microdissection Genomics Research Center, <sup>2</sup>Department of Pathology and <sup>3</sup>Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul, 137-701, Republic of Korea

Received on October 15, 2004; revised on February 7, 2005; accepted on February 24, 2005  
Advance Access publication March 3, 2005

### ABSTRACT

**Summary:** ArrayCyGHt is a web-based application tool for analysis and visualization of microarray-comparative genomic hybridization (array-CGH) data. Full process of array-CGH data analysis, from normalization of raw data to the final visualization of copy number gain or loss, can be straightforwardly achieved on this arrayCyGHt system without the use of any further software. ArrayCyGHt, therefore, provides an easy and fast tool for the analysis of copy number aberrations in any kinds of data format.

**Availability:** ArrayCyGHt can be accessed at <http://genomics.catholic.ac.kr/arrayCGH/>

**Contact:** yejun@catholic.ac.kr

**Supplementary information:** Technical documentation is available. See <http://genomics.catholic.ac.kr/arrayCGH/>

### INTRODUCTION

The advantages of comparative genomic hybridization (CGH) have been accentuated by combining with microarray technology, array-CGH (Pinkel *et al.*, 1998). It is now commonly used for investigating DNA copy number aberrations in various diseases.

Analysis of microarray data requires sophisticated processing steps to make reliable interpretation. Therefore it is crucial to streamline the processes and provide integrated environment for array-CGH data. Recently, several algorithms for array-CGH data analysis have been developed (Hupe *et al.*, 2004; Jong *et al.*, 2004; Lingjaerde *et al.*, 2005). With these technical improvements, recognition of the copy number alteration breakpoints became more accurate and convincing. However, applications supporting raw array-CGH data processing are uncommon. In this study, we intended to develop more user-friendly and feasible web-based tool for array-CGH data analysis with any set of BAC clone chips and with any format of raw data file produced by GenePix (Axon Inc., CA), UCSC SPOT (Jain *et al.*, 2002), Imagine (Biodiscovery Inc., CA) or QuantArray (Packard BioChip Technology).

### SERVER OVERVIEW

ArrayCyGHt is available on the web. To develop arrayCyGHt we have used R (<http://www.r-project.org>), Apache (<http://www.apache.org>), PHP (<http://www.php.net>) and Perl (<http://www.perl.org>) which are well-known softwares available on the Internet.

This web application provides basic functions for filtering, normalizing, analyzing and plotting array-CGH data. Since arrayCyGHt is running on a server, a web browser is only a requirement to use the server.

ArrayCyGHt has the following features:

- data import from tab-delimited text files,
- data preprocessing (filtering, normalization, log transformation, realigning the raw data to linear mapping order),
- graphical representation of copy number gain or loss under user guided threshold and
- direct linking of the identified regions to public database.

### STREAMLINED ANALYSIS STEPS

#### Loading data

Dataset to be analyzed should be given in tab-delimited text files. Chromosome mapping information and raw signal of every individual spot can be uploaded as two separate files or as a single file. Data files consist of a header row followed by rows of data.

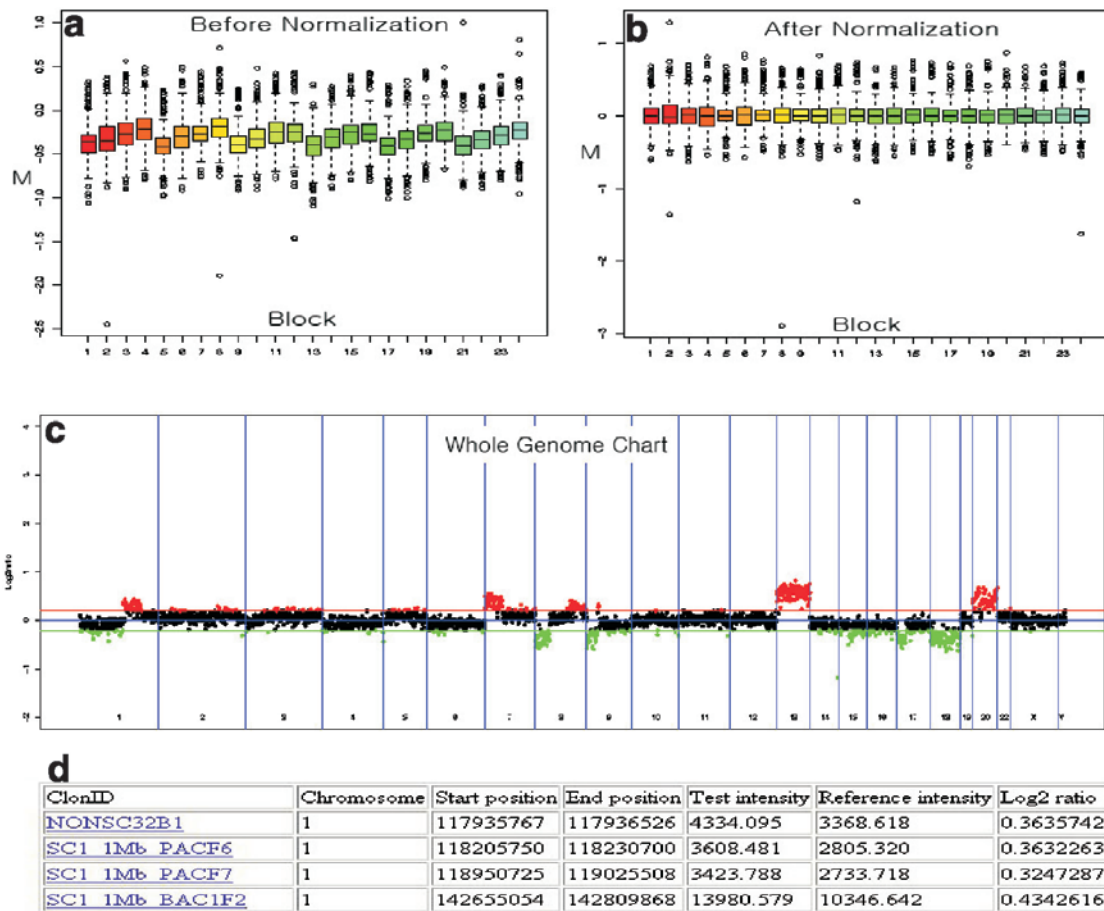
#### Normalization

When users upload array layout, the server shows several data normalization options available to users. After normalization, the server illustrates reconstructed array data, MA plots and box plots representing distribution of both submitted and normalized data (Fig. 1a and b). These plots indicate the quality of submitted data and effects of normalization. For this process, limma (Smyth, 2004) and sma (Yang *et al.*, 2002) packages of R are used.

#### Analysis

If spots are multiplied on a chip, their signal intensities are averaged. The averaged data will be transformed to log<sub>2</sub> scale and realigned according to the linear chromosomal mapping order. Then plots of whole genome are presented. Figure 1c illustrates the example of genome-wide array-CGH plot of primary colon cancer (unpublished data). All of the plot images can be downloaded. ArrayCyGHt defines the clones of copy number alterations by user guided threshold, and subsequently, table of outliers

\*To whom correspondence should be addressed.



**Fig. 1.** Box plots of cy5–cy3 ratio distribution in individual blocks before (a) and after (b) normalization. (c) Genome-wide copy number profile of a primary colon cancer (DUC17) using whole-genome human array-CGH chip. Red and green dots indicate the clones above or below the threshold,  $\pm 0.2$  in  $\log_2$  ratio, respectively. (d) Table of outlier clone list. They are linked to Human Genome Browser Gateway.

is presented (Fig. 1d). The outlier clones are directly linked to the Human Genome Browser Gateway (<http://genome.ucsc.edu/cgi-bin/hgGateway>). When we compared ArrayCyGHt with several other array-CGH analysis programs using published array-CGH data, the expected alterations were concordantly identified by ArrayCyGHt which means that the reliability and accuracy of this application are comparable with other applications. These comparative study results are available at ArrayCyGHt website.

**OUTLOOK**

Although arrayCyGHt is not an all-in-one tool for analysis of array-CGH data, this web-based tool can be a valuable resource for analyzing the constitutional chromosome aberrations with any set of BAC clone chip and for using data files from any kind of software. To our knowledge, this is the first web-based tool for array-CGH data analysis and visualization. ArrayCyGHt is the only application until now that supports the normalization and realigning of the signal intensity ratios in a user-friendly format. However, the comparison of multiple experiments and hence the determination of recurrent copy number alterations is not available in this application and will be implemented in the next version. This temporary shortcoming could be easily overcome by uploading the normalized data to other

software packages that focus mainly on smoothing, clustering and analysis of multiple array-CGH data.

**ACKNOWLEDGEMENT**

This work was supported by BumSuk Academic Research Fund of 2004.

**REFERENCES**

HuPe,P. *et al.* (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, **20**, 3413–3422.  
 Jain,A.N. *et al.* (2002) Fully automatic quantification of microarray image data. *Genome Res.*, **12**, 325–332.  
 Jong,K. *et al.* (2004) Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics*, **20**, 3636–3637.  
 Lingjaerde,O.C. *et al.* (2005) CGH-Explorer: a program for analysis of array-CGH data. *Bioinformatics*, **21**, 821–822.  
 Pinkel,D. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **20**, 207–211.  
 Smyth,G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**(1) Article 1.  
 Yang,Y.H. *et al.* (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.*, **30**, e15.