A practical system for the detection of chromosomal aberration in tumor cells by array CGH

Junya Futagawa[1] futagawa@hydra.mki.co.jp
Tomohiro Isobe[1] isobe@hydra.mki.co.jp
Atsushi Moroda[1] moroda@hydra.mki.co.jp

Nobutaka Mitsuhashi[1] mituhasi@hydra.mki.co.jp
Ken Aoshima[1] kaku@hydra.mki.co.jp
Yasuhito Arai[3] yarai@ncc.go.jp
Fumie Hosoda[3] fhosoda@ncc.go.jp

Misao Ohki[3] mohki@ncc.go.jp
Issei Imoto[4] issecgen@mri.tmd.ac.jp
Johji Inazawa[4] joinaz.cgen@mri.tmd.ac.jp
Tokuki Sakiyama[2] tsakiyam@ncc.go.jp

[1] R&D Dept. Bioscience Division, Mitsui Knowledge Industry Co., Ltd. Harmony Tower 21st floor, 1-32-2 Honcho, Nakano-ku, Tokyo 164-8721, Japan
[2] Genetics Division, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan
[3] Cancer Genomics Project, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan
[4] Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

Keywords: cancer, chromosomal aberration, array CGH, Affymetrix SNP Chip

1 Introduction

The recent development of microarray technologies made it possible to detect expression level of several tens of thousand of genes thoroughly and systematically. These technologies are also applied to Comparative Genomic Hybridization (CGH) method in order to study solid cancer [1]. Calculating the ratio of two signal intensities derived from tumor and normal cells labeled by different fluorescent dye, we can quantify the chromosomal copy number in the tumor cell at the region identical to the cloned DNA fragment arrayed on the glass slide [2].

However the data obtained from array CGH experiments have several characteristics which are different from those of mRNA expression experiments as follows: (1) The loci closely mapped on the genome are expected to have the same copy number. (2) The copy number of the locus should take on discrete values, e.g. 0 for homozygous deletion, 1 for heterozygous deletion, 2 for normal (no chromosomal aberration), 3 for 1.5-fold amplification. Therefore the base-2 logarithm of the tumor over normal signal intensity ratio (log2 tumor/normal) should comply with the mixture of multiple normal distributions. (3) The samples prepared from cancerous tissues are the random mixture of normal cells and tumor cells. Thus each normal distribution could have the different mean values in every experiments.

Nowadays a variety of software applications to analyze microarray data are available, however, the software tailored to analyze array CGH are very limited. In this study, we developed a software application named "ACUE2" that implements the algorithms dedicated to array CGH analysis.

2 Method

Measurement: In order to gather array CGH data, we used the Affymetrix SNP chip 10K and the made-to-order chip, MCG Genome Wide-4500 (BAC4500 chip), where 4,500 BAC probes were arrayed. We used the SKNAS neuroblastoma cell line as test samples and the normal cell as reference. The labeling and hybridization were performed by the one-color system for the SNP chip and the two-color system for the BAC4500 chip. Chips were imaged on a scanner. And acquired images were analyzed by use of GenePix Pro (Axon Instruments).

Analysis: (a) The values, log2 of test over reference ratio observed at every spots, were mapped on chromosomes by the physical position of the probes arrayed on the spot. Then we conducted the Gaussian

smoothing with window size up to 1000kbp. As a result, we detected the global tendency of chromosomal copy number amplifications and losses by considering the copy number at the adjacent region of the focusing region. (b) The values were fitted to Gaussian mixture model by using EM algorithm, which optimized the mixture rate of tumor and normal cells, the mean values and the deviation values of each normal distribution, and the branching probability. And we transformed each value to z-value based on the distribution which was formed by the values from the region with no chromosomal aberration. We determined the threshold to be ±1.96 z-value, and judged the status (loss, gain, or normal) for all the region.

3 Results and Discussions

Figure1 shows the relationship between standard deviation of base-2 logarithm of the test over reference signal intensity ratio (Sigma) and the window size (WindowSize) when we take the copy number at the adjacent region into account. This relationship is derived under the condition of control: control equal to 1:1. In case the window size is 100kbp, the sigma is approximately 0.4, which determines the false-negative to be almost 0.18. This result is comparable to the same analysis using 5 normal samples versus 5 normal samples without smoothing (data not shown).

Figure2 shows the results in case the values were fitted to the single normal distribution model and the Gaussian mixture model respectively. Bottoms in figure2 show the mapping result on chromosome 1 for both models. SKNAS is known to have a 0.5-fold loss at the short arm on chromosome 1 and 1.5-fold gain at the long arm on the same chromosome. We confirmed the Gaussian mixture model strongly support the well-known fact.

ACUE2 also implements general algorithm to analyze microarray data such as background correction, regression, mask, etc., in addition to the algorithm described above. Figure3 shows the snapshots of ACUE2. ACUE2 is a client-server computer system with sophisticated, easy to use user interfaces and also offers the functionalities of sharing the experiment data among the project team members and the integrated management of user and data.

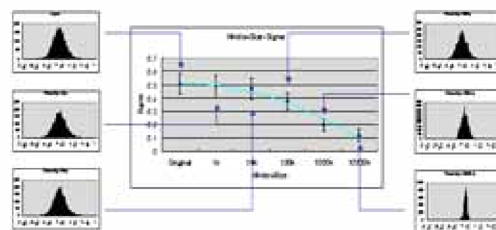

Figure1: The relationship between SD and Window size
The samples from different 26 normal cells were hybridized with Affymetrix SNP chip 10K. The log2 ratio was calculated against two data obtained by bootstrap method.

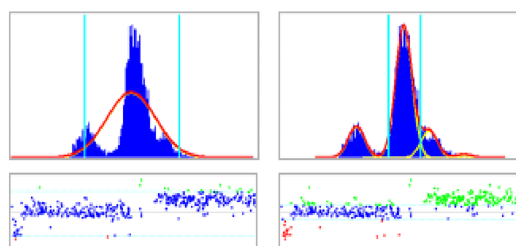

Figure2: The comparison of two models
The samples from normal cell and SKNAS were hybridized with BAC4500 chip and the values of log2 ratio were calculated. The left-top and left-bottom figures are the results for the single normal distribution model and the rest for the Gaussian mixture model.

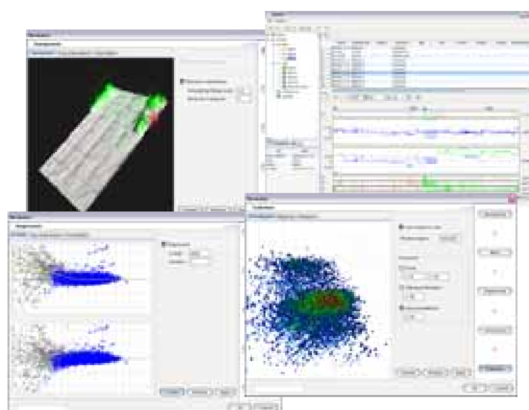

Figure 3：Snapshot of ACUE2 application

Acknowledgements

This study was supported by the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research of Japan.

References

[1] Inazawa, J. and Imoto, I., Comparative genomic hybridization (CGH)-arrays pave the way for identification of novel cancer-related genes, Cancer Sci, vol.95 550-563, July 2004

[2] Pinkel, D. and Albertson, DG., High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays, nature genetics, vol.20 October 1998