

# Method for finding p53 binding sites using tiling array

**Shingo Tsuji**<sup>1,2</sup>

tsuji@genome.rcast.u-tokyo.ac.jp

**Kiyofumi Kaneshiro**<sup>1</sup>

kaneshiro@genome.rcast.u-tokyo.ac.jp

**Shuichi Tsutsumi**<sup>1</sup>

shuichi@genome.rcast.u-tokyo.ac.jp

**Sigeo Ihara**<sup>2</sup>

ihara@genome.rcast.u-tokyo.ac.jp

**Hiroyuki Aburatani**<sup>1</sup>

haburata-ky@umin.ac.jp

Genome Science Division<sup>1</sup> and Dynamical Bioinformatics Division<sup>2</sup>,  
Research Center for Advanced Science and Technology, The University of Tokyo,  
4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan

**Keywords:** ChIPchip, p53, tiling array data analysis, transcription factor binding site, p53MH

## 1 Introduction

Genome tiling arrays, which represent nonrepetitive sequences on human genome have been used to investigate transcription factor binding sites (TFBS) *in vivo* [1]. Namely, not only the upstream genomic regions of gene but also intronic and intergenic regions are identified as TFBS. It is necessary to develop an efficient analysis method to deal with tiling array output.

In combination with chromatin immunoprecipitation (ChIP) assay and Affymetrix's tiling array, we identified p53 binding sites on human chromosome 21 and 22. We present a new method for tiling array data analysis, which involves efficiency test statistics (e.g. Kolmogorov-Smirnov). We compared tiling array output with computational prediction such as P53MH [2] to optimize threshold of p-value of the test statistics.

## 2 Method and Results

Human colorectal cancer cell, HCT116 (p53+/+) was grown to a density of  $0.5 \times 10^6$  cells/ml. HCT116 cells were treated with 375uM of 5-FU for 9h to induce expression of wild-type p53. Chromatin Immunoprecipitation was performed for the cells at 0h and 9h after the treatment using anti-p53 antibody (DO7) and normal mouse IgG (as a control). Chromatin-immunoprecipitated DNA and corresponding control samples were amplified and hybridized to the chromosome 21 and 22 arrays (Affymetrix's high-density oligo array).

Tiling array output data was analyzed with the procedure proposed by Affymetrix. (PM, MM) intensity pairs were mapped to the genome using exact 25-mer matching. For each probe position, a data set was generated with 550bp length sliding window. The average number of probes within the window is 14. A Wilcoxon rank sum test was applied to the treatment and the control data set, and then p-value of test statistics allotted to each probe position. Fig. 1 visualizes an analysis result on chromosome 21 by Integrated Genome Browser [3]. Regions with continuous low p-value were indicated at the upper portion of Fig 1(a), which are considered as p53 binding sites. Middle part of the figure shows RefSeq genes.

Fig. 1(b) and (c) show magnification of a part of Fig 1(a). We call Fig. 1(b) as region B and Fig. 1(c) as region C. Rectangles on upper part of figures are low p-value region. Bar graph at middle part of the figure indicates probe intensities (PM) of p53 arrays, and probe intensities (PM) of IgG (control) bar graph is below.

Intensities of p53 in region B are considerably stronger than those of region C. However, Wilcoxon rank sum test statistics of p53 to IgG showed almost similar p-value in these two regions, where a fold enrichment of p53 to IgG was not observed. To address this problem, we developed a new procedure using

Kolmogorov-Smirnov test, which is sensitive to difference of shape of empirical distribution function.

Analysis results vary with the p-value threshold. With higher p-value, we would get more false positive sites. To determine the appropriate threshold, we need to know the actual p53 binding sites. Because such information is not available, we alternatively used computationally predicted sites as the potential p53 binding sites. The appropriate p-value threshold was determined by applying P53MH [2] algorithm to whole chromosome 21 and 22.

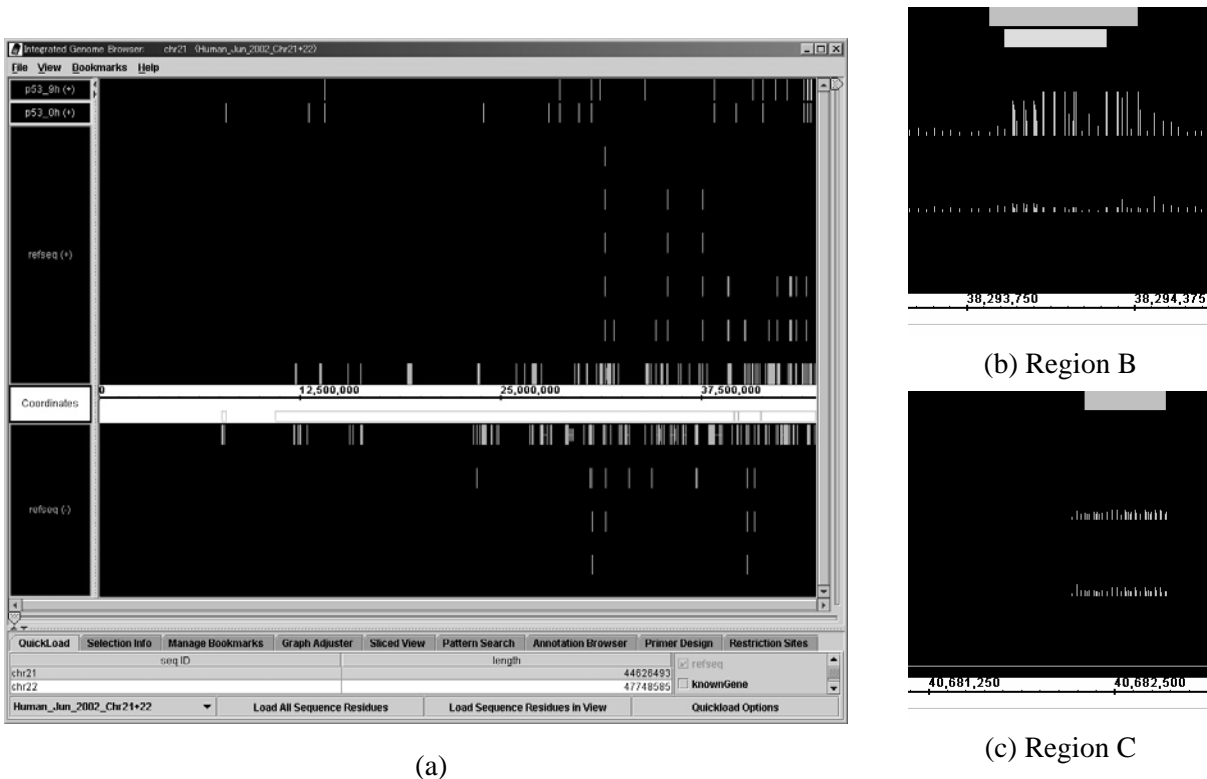


Figure 1: Visualization of analysis result by Integrated Genome Browser

### 3 Discussions

Each chromatin immunoprecipitation assay with different antibodies should have own threshold p-value. Thus, if we don't have supporting tools like P53MH, we have to determine the appropriate p-value threshold for each CHIPchip assay, by using parameters like the whole intensity distribution. We are currently improving the analysis method by accumulating more CHIPchip assay data.

### References

- [1] Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V.I., Cheng, J., Williams, A.J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammanna, H., Helt, G., Struhl, K. and Gingeras, T.R., Unbiased Mapping of Transcription Factor Binding Sites Along Human Chromosomes 21 and 22 Points to Widespread Regulation of Non-Coding RNAs. *Cell*, 116(4):499-509, 2004.
- [2] Hoh J, Jin S, Parrado T, Edington J, Levine AJ, Ott J., The p53MH algorithm and its application in detecting p53-responsive genes, *Proc Natl Acad Sci U S A*, 99(13):8467-72, 2002.
- [3] [http://www.affymetrix.com/support/developer/tools/download\\_igb.affx](http://www.affymetrix.com/support/developer/tools/download_igb.affx)