

Analyzing the Layout of Affymetrix Microarrays

Sérgio A. de Carvalho Jr.¹ & Sven Rahmann²

Keywords: microarray layout, placement algorithms, border length minimization.

1 Motivation

Affymetrix microarrays have their probes synthesized in parallel, on the chip, in a series of repetitive steps. Each step appends the same nucleotide to selected regions of the chip. Selection occurs by exposure to light with the help of a photolithographic mask that allows or obstructs the passage of light accordingly. Due to diffraction or internal reflection, untargeted spots can sometimes be accidentally activated. The problem is more likely to occur near the borders between masked and unmasked spots.

The issue was first addressed by Hannenhalli *et al.* [1]. They formulated the Border Length Minimization Problem, which aims at minimizing the number of border conflicts during mask exposure steps. Kahng *et al.* [2] later extended the concept with two simple yet important practical considerations: a) stray light might activate probes that lie as far as three cells away from the targeted spot; and b) imperfections in the middle of a probe are more harmful than in its extremities.

2 Evaluating Affymetrix Microarrays

We define the conflict index $\kappa(s)$ of a spot s whose probes of length ℓ_s are synthesized in μ masking steps as follows. First we define a distance-dependent weighting function, $\delta(s, m, n)$, that accounts for observation a) above:

$$\delta(s, m, n) := \begin{cases} 0 & \text{if spot } n \text{ is masked at step } m, \\ \frac{1}{(d(s, n))^2} & \text{otherwise,} \end{cases} \quad (1)$$

where $d(s, n)$ is the Euclidian distance between spots s and n . We also use position-dependent weights (similar to those suggested in [2]) to account for observation b):

$$\omega(s, m) := \begin{cases} 0 & \text{if spot } s \text{ is unmasked at step } m, \\ \sqrt{\min(i_{s, m} + 1, \ell_s - i_{s, m} + 1)} & \text{otherwise,} \end{cases} \quad (2)$$

where $i_{s, m}$ denotes the number of nucleotides synthesized at spot s up to and including step m . We now define the conflict index of a spot s as

$$\kappa(s) := \sum_{m=1}^{\mu} \left(\omega(s, m) \sum_n \delta(s, m, n) \right), \quad (3)$$

where n ranges over all neighboring spots of s .

We examined the layout of a number of Affymetrix GeneChip[®] arrays and computed the conflict index of their spots. We could distinguish several placement strategies employed

¹Graduiertenkolleg Bioinformatik, Bielefeld University, D-33594 Bielefeld, Germany. E-mail: Sergio.Carvalho@cebitec.uni-bielefeld.de

²Algorithms and Statistics for Systems Biology group, Genome Informatics, Faculty of Technology, Bielefeld University, D-33594 Bielefeld, Germany. E-mail: rahmann@cebitec.uni-bielefeld.de

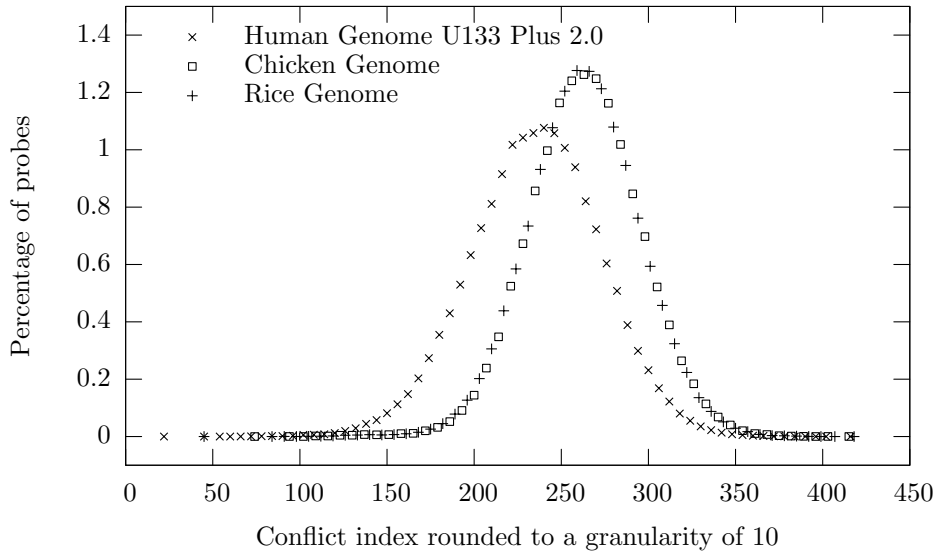


Figure 1: Distribution of conflict index values for three GeneChip arrays.

by Affymetrix. Some of them were visibly better than others when it came to minimizing spot conflict.

Figure 1 shows the distribution of conflict indices for three of the largest GeneChip arrays. The layout of the Human Genome chip has undoubtedly reduced the amount of conflicts. This chip has around 10.4% of empty spots which are evenly distributed throughout the array. On the other hand, the Chicken chip (with as much as 11.9% of empty spots) and the Rice chip (with 6.5% of empty spots) have their probes concentrated on the upper part of the array with the lower part consisting of a large empty area. Clearly, they could have benefited from having a more uniform distribution of empty spots. This came as a surprise given that they are among the latest GeneChip arrays available.

3 Future Work

We are currently analyzing the performance of several placement algorithms that have been suggested lately with our definition of conflict index. We are also developing new placement strategies to reduce the overall spot conflict of microarrays.

References

- [1] Hannenhalli, S., Hubell, E., Lipshutz, R. and Pevzner, P. 2002. Combinatorial algorithms for design of DNA arrays. In *Adv Biochem Eng Biotechnol* 77:1–19.
- [2] Kahng, A. B., Mandoiu, I., Pevzner, P., Reda, S. and Zelikovsky, A. 2003. Engineering a scalable placement heuristic for DNA probe arrays. In *Proceedings of the Seventh Annual International Conference on Computational Molecular Biology (RECOMB 2003)* Berlin: ACM. pp. 148–83.