Spot shape modelling and saturated pixels in microarrays

Claus Ekstrøm[*], Charlotte Kristensen[†], Søren Bak[†] & Mats Rudemo[*‡]

Abstract

To be able to study lowly expressed genes in microarray experiments it is useful to increase the gain in the scanning. However, a large gain may cause some pixels for highly expressed genes to become saturated, that is the registered pixel values become censored at the upper limit, which with 16 bit precision is $2^{16} - 1 = 65535$. Techniques for adjustment of highly expressed signal intensities are given in [3] based on a small set of available spot summaries such as spot mean, spot median and spot variance. As mentioned in [3] it should be possible to get more accurate adjustments when all pixel values are available. In the present project we study spatial statistical models for pixel values which should enable such adjustments.

A convenient type of modelling is to transform data to become approximately Gaussian distributed with a mean value function determined by gene intensities and spot shapes and a corresponding covariance function. For such models censored pixel values can be optimally estimated. We study different types of transformations, spot shapes and covariance functions. The transformations include logarithmic and power transforms with an offset and the inverse hyperbolic sine transform of [1]. The spot shapes include three types suggested in [2]: (i) an isotropic 2D Gaussian distribution, (ii) a crater spot distribution consisting of a difference between two scaled isotropic 2D Gaussian distributions and (iii) a plateau spot distribution. For the covariance we assume that locally with cylindrical coordinates (r, ϕ) in the neighbourhood of a pixel center, the covariance function $c(r, r', \phi, \phi')$ can be approximated by a function of form $C(r, r', \phi - \phi')$.

The models are applied to the analysis of a dataset obtained with a specially designed 50mer oligonucleotide microarray. Here 452 selected genes in transgenic Arabidopsis plants are compared to the corresponding genes in wild-type plants. Data include scans with different gains ranging from no saturation to heavy saturation.

References

[1] Wofgang Huber, Anja von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18** Suppl. 1:S96–S104, 2002.

[2] Cristoph K Wierling, Matthias Steinfach, Thorsten Elge, Steffen Schulze-Kremer, Pia Aanstad, Matthew Clark, Hans Lehrach, and Ralf Herwig. Simulation of DNA array hybridization experiments and evaluation of critical parameters during subsequent image and data analysis. *BMC Bioinformatics*, **3**:29, October 2002.

[3] Ernst Wit and John McClure. Statistical adjustment of signal censoring in gene expression experiments. *Bioinformatics*, **19**, 2003. To appear.

[*]Dept. of Mathematics and Physics, The Royal Veterinary and Agricultural University, Denmark
[†]Dept. of Plant Biology, The Royal Veterinary and Agricultural University, Denmark
[‡]rudemo@math.chalmers.se