Gene expression

Software package for automatic microarray image analysis (MAIA)

Eugene Novikov* and Emmanuel Barillot
Service Bioinformatique, Institut Curie, 26 Rue d’Ulm, 75248 Paris Cedex 05, France

Received on May 11, 2006; revised on December 1, 2006; accepted on December 18, 2006
Advance Access publication January 19, 2007
Associate Editor: Alfonso Valencia

ABSTRACT

Summary: Although various software solutions are currently available for microarray image analysis, one would still expect to develop algorithms ensuring higher level of intelligence and robustness. We present a fully functional software package for automatic processing of the two-color microarray images including spot localization, quantification and quality control. The developed algorithms aim at making ratio estimates more resistant to array contamination and offer automatic tools to evaluate spot quality.

Availability: A demo version of the software can be downloaded from http://bioinfo.curie.fr/projects/maia. A full version is freely available to non-commercial users upon request from the authors.

Contact: eugene.novikov@curie.fr

Two-color (typically, Cy3/green and Cy5/red) comparative microarray experiment is a key point of cDNA (Hedge et al., 2000), CGH (comparative genome hybridization, Pinkel et al., 1998) and, more recently, protein (Eckel-Passow et al., 2005) microarray technologies. In a conventional two-color experiment, one obtains two images (one for each color) containing thousands of bright spots, distributed over the array according to a special pattern. The first step in microarray data analysis consists of image processing intending to estimate the ratio of the fluorescence intensities in two color channels at each spot. This ratio reflects differential gene (cDNA) or protein expression or change in DNA copy number (CGH) between two compared samples. Although various academic and commercial software solutions for microarray image analysis are currently available (Jain et al., 2002; Molecular Devices, 2005; White et al., 2005; Yang et al., 2002), further efforts will perhaps never be exhausted because the existing solutions are not ideal and offer some room for further improvements.

In this note, we present a fully functional software package (MAIA) for automatic processing of the microarray two-color images including spot localization, quantification and quality control (Fig. 1).

The spot localization module (i) identifies the position of each spot on the array to associate it with the spotted clone; and (ii) establishes the borders between the neighboring spots to allow further independent data processing (extracting quantitative information) for each spot. We deal with the most widespread, orthogonal, spot localization pattern. In this pattern, the spots are aligned horizontally and vertically and can be arranged in blocks containing different numbers of spot rows and spot columns (Fig. 1). The spot localization algorithm (Novikov and Barillot, 2006) is fully automatic and robust with respect to deviations from perfect spot alignment and contamination. As an input, it requires only the common array design parameters: number of blocks and number of spots in the $x$ and $y$ directions of the array. Although fully automatic, there is no guarantee that it will perform well for any array. Therefore, we offer some interactive tools to repair grid in case if it is erroneous.

The spot quantification module generates two sets of quantitative parameters for each spot: ratio estimates and spot quality characteristics. Two algorithms for ratio estimation were implemented (Novikov and Barillot, 2005a). The first is a direct arithmetic ratio of the background-corrected fluorescence intensity estimates in the two color channels. This approach requires the identification of both the foreground—the measured spot—and the background—typically the level of non-specific hybridization. The second ratio estimate is the slope of the linear regression plot of the pixel intensities in one color channel versus another one. This approach does not require spot segmentation, however unavoidable presence of some aberrant or outlier pixels, occurring even in small quantities, can distract the regression line and strongly bias the regression ratio. Therefore we have developed a statistical procedure (Novikov and Barillot, 2005a), which systematically searches and removes aberrant or outlier pixels (Fig. 1). This procedure guarantees a higher level of confidence in the ratio estimates obtained using linear regression approach. Moreover, after aberrant pixels are removed the segmentation algorithm also ensures more robust estimates, and there is a greater agreement in the ratio values obtained by both methods.

The spot quality module provides a value of spot quality reflecting the level of confidence in the obtained ratio estimate at each spot. The unique spot quality value is derived from a set of nine marginal quality parameters characterizing certain features of the spot (Novikov and Barillot, 2005b). The contribution of each quality parameter in the overall quality is automatically evaluated based on the visual classification of the spots, or using information available from the replicated spots, located on the same array (Fig. 1) or over a set of replicated arrays (Novikov and Barillot, 2005b).

*To whom correspondence should be addressed.
the developed procedure allows us not only to quantify spot quality, but also to identify different types of spot deficiency occurring in microarray technology. The quality values can be used either directly to flag out some spots with the quality lower than the user-defined threshold, or in the follow-up analysis as a weight controlling the contribution/influence of the obtained ratio estimates.

**Image simulation.** We have developed a special software component for Monte-Carlo simulation of the microarray images (Novikov and Barillot, 2005a). It takes into account the statistical noise and different types of contamination like non-specific hybridization and dust. Since in a simulation experiment the true values of the ratios are precisely known, it allows us to evaluate, to test and to compare different algorithms of microarray image analysis objectively.

**Software.** MAIA runs on 95/98/Me/NT/2000/XP Windows platforms and needs the Java Runtime Environment. Complete analysis of one 4 Mb image pair (Cy3/Cy5, ~7300 spots; each spot is ~45 pixels) takes ~5 s on 3.00 GHz Pentium® 4 CPU with 1 GB of RAM.

The validity of the algorithms has been tested and confirmed on a large set of experimental images from the different microarray platforms used within our Institute and on a representative selection of CGH images obtained from the UCSF Cancer Center.

**ACKNOWLEDGEMENTS**

We would like to thank our colleagues from the different laboratories of the Institute Curie: (F. Radvanyi, UMR 144 CNRS/IC: Compartimentation et dynamique cellulaires; O. Delattre, INSERM 509: Pathologie moléculaire des cancers; M. Dutreix, UMR 2027 CNRS/IC: Génotoxicologie et cycle cellulaire) and Prof D. Pinkel (UCSF Comprehensive Cancer Center), who has provided numerous microarray images allowing considerable improvement of the algorithms.

**Conflict of Interest:** none declared.

**REFERENCES**


