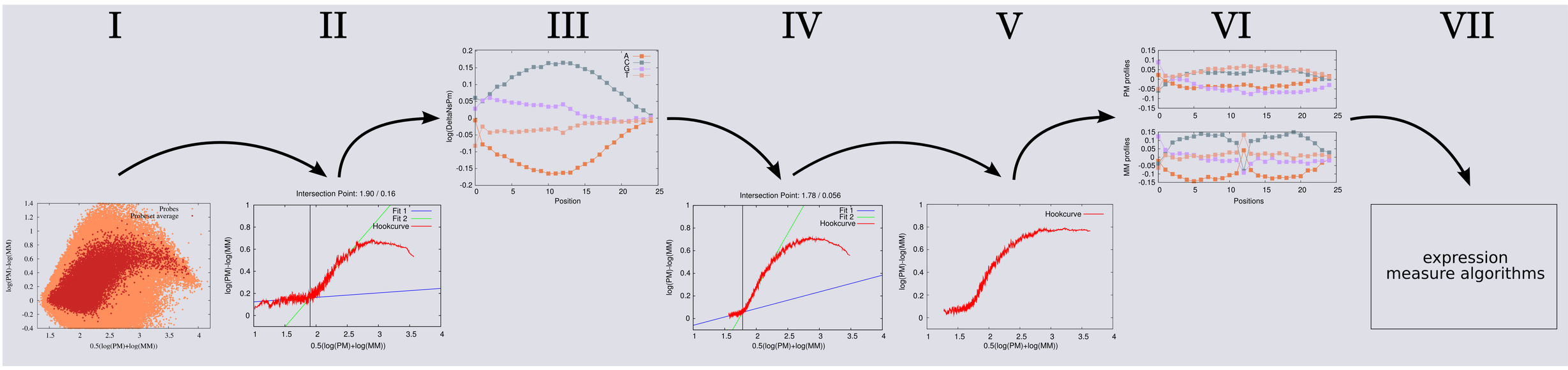
A new program package for model-based microarray analysis

Mario Fasold[1], **Jan Brücker**[2], **Peter F. Stadler**[1,2,3], **Hans Binder**[2], **Jörg Hackermüller**[2,4,*]

Email: *joerg.hackermueller@izi.fraunhofer.de
WWW: http://www.izbi.uni-leipzig.de
[1] Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig, Germany
[2] Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany
[3] Santa Fe Institute, Santa Fe, USA
[4] Fraunhofer Institute for Cell Therapy and Immunology — IZI, Leipzig, Germany

Abstract

Microarrays aim at measuring the expression degree of thousands of genes in a single experiment. The measured intensities of the probe spots are unfortunately affected by parasitic interferences, which in many cases prohibit the direct translation into the target RNA concentrations. The calibration of raw intensity data is therefore a fundamental prerequisite for the analysis of oligonucleotide microarray data. The field of array normalization tools can be separated into tools that use only the perfect match probes (PMs) of one chip, tools that use PMs and mismatch probes (MMs) of one chip and tools that rely on multichip normalization techniques and reference arrays. The best results are yielded by multichip techniques like RMA or GCRMA. However, it is often unwanted to hybridize multiple arrays. We report a method of data adjustment based on a physico-chemical model which corrects raw microarray intensity data of one chip for the effect of (i) sequence-specific affinities, (ii) mismatches, (iii) cross-hybridization and (iv) saturation [3, 1, 2].



The Process:

I Scatterplot.
$\Delta = \log PM - \log MM$ **vs** $\Sigma = \log PM + \log MM$.
bright: The probe intensities.
dark: The probeset averaged intensities.

II Hook Curve.
Extreme smoothing of the scatter plot produces the raw hook curve.
The hook curve gives the opportunity to separate the probes that follow nonspecific binding from those that have specific content.

III Sensitivity Profiles.
A sequence dependent binding model is calculated for the probes that follow non specific binding.

IV Corrected Hook Curve.
The intersection point can be calculated more exactly.
The nonspecific binding model is updated.
If the chip saturates I_{max} can be calculated by fitting a theoretic curve onto the hook curve.

V Desaturation.
With an adequate I_{max} the intensities can be desaturated.
The Langmuir model is transformed into a linear model: $L_p^P = L_{0,p}^{P,S} \cdot \Delta_p^{P,S} + L_{0,p}^{P,NS} \cdot \Delta_p^{P,NS}$

VI Specific Sensitivity Profiles.
The non-specific contant of each probe is subtracted.
Sensitivity profiles are calculated for the left specific contant.

VII Expression Measures.
The corrected intensities are prepared to be processed by different expression measures.

Expression Measures

PM:
$X_{0,p}^{PM,S} = \frac{X_p^{PM,S}}{\delta y_p^{PM,S}}$ **with**
$X_p^{PM,S} = X_p(I_p^{PM}) - \left[X_0^{PM,NS}\right]_{p \in NS} \cdot \delta y_p^{PM,S}$
MM:
$X_{0,p}^{MM,S} = \frac{X_p^{MM,S}}{\delta y_p^{MM,S}}$ **with**
$X_p^{MM,S} = \left(X_p(I_p^{PM}) - \frac{\left[X_0^{PM,NS}\right]_{p \in NS}}{n_0} \cdot \delta y_p^{MM,S}\right)$
PM-MM:
$X_{0,p}^{PM-MM,S} = X_{0,p}^{PM,S} - X_{0,p}^{MM,S}$

Enhanced Measures

Substitute $\left[X_0^{PM,NS}\right]_{p \in NS}$ **by an integration over the distribution of the** NS **probes.**
PM:
$X_p^{PM,S} = \frac{\int_{\mu - xu}^{\mu + xo} p^n(x) glog(\Delta L_p^{PM}(x)) dx}{\int_{\mu - xu}^{\mu + xo} p^n(x) dx}$
MM and PM-MM enhanced are not listed.

Results

The quality of the method was assessed by Affycomp II [4, 5], a Benchmark for Affymetrix GeneChip Expression Measures. Graphics to the right are chosen from the HGU133 spike-in assessment.

a MA plot of gcrma. Log fold change as a function of mean log expression level. Spiked-in genes are symbolized by numbers. Red: non-differentially expressed genes with fold changes larger than 2**.**

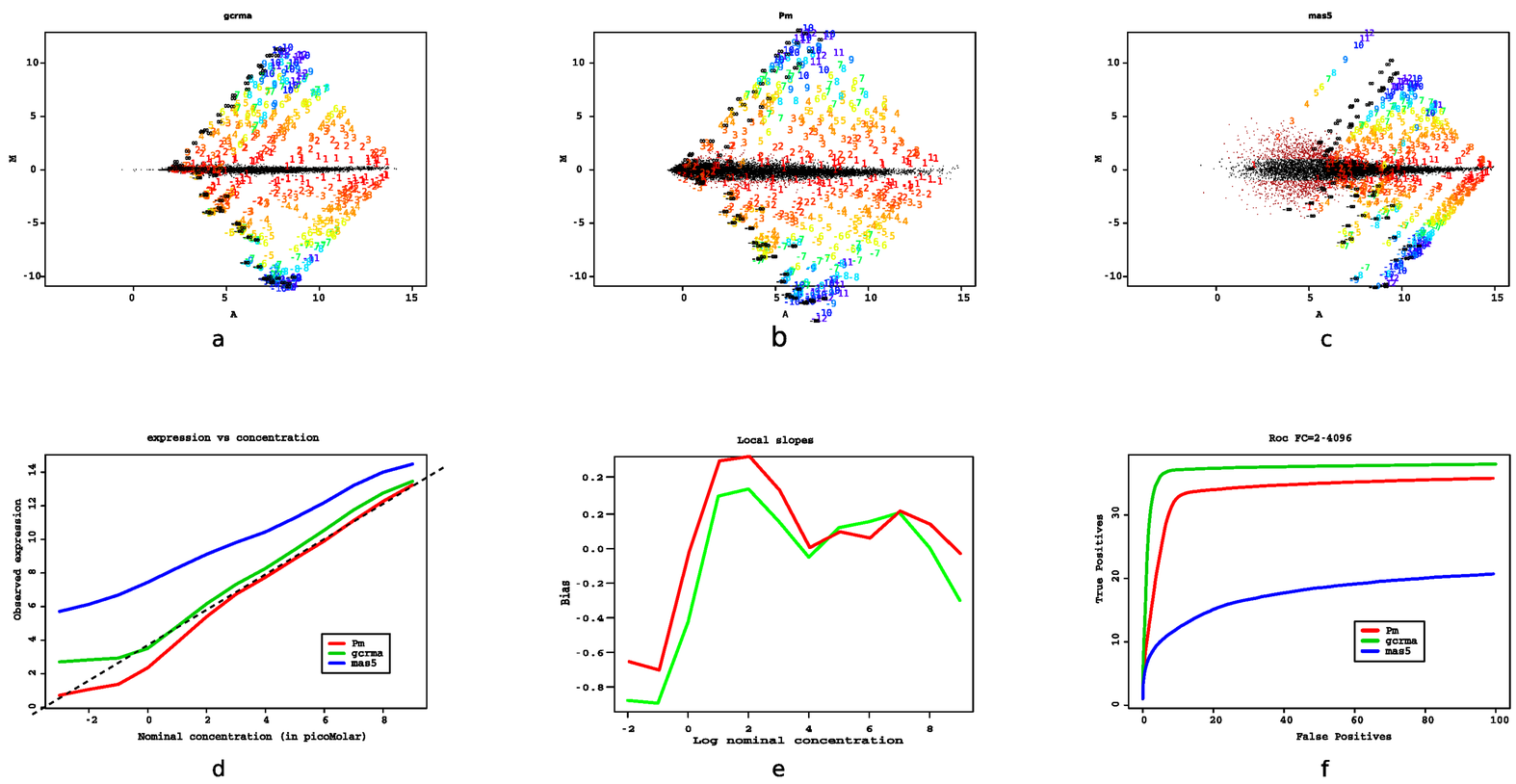
b MA plot of the hook curve method with the enhanced PM expression measure.

c MA plot of MAS 5.0.

d Average observed log_2 **intensity plotted against nominal** log_2 **concentration. the ideal slope of** 1 **is given as a dotted line.**

e Local slopes.

f ROC curves based on comparisons with nominal fold changes ranging from 2 **to** 4096**.**



References

[1] H. Binder and S. Preibisch. Specific and nonspecific hybridization of oligonucleotide probes on microarrays. *Biophys. J.*, 89(1):337–352, 2005.

[2] H. Binder and S. Preibisch. Genechip microarrays—signal intensities, rna concentrations and probe sequences. *Journal of Physics: Condensed Matter*, 18(18):S537–S566, 2006.

[3] H. Binder, S. Preibisch, and T. Kirsten. Base pair interactions and hybridization isotherms of matched and mismatched oligonucleotide probes on microarrays. *Langmuir*, 21(20):9287–9302, Sep 2005.

[4] L. M. Cope, R. A. Irizarry, H. A. Jaffee, Z. Wu, and T. P. Speed. A benchmark for affymetrix genechip expression measures. *Bioinformatics*, 20(3):323–331, Feb 2004.

[5] R. A. Irizarry, Z. Wu, and H. A. Jaffee. Comparison of affymetrix genechip expression measures. *Bioinformatics*, 22(7):789–794, Apr 2006.

Printed by Universitätsrechenzentrum Leipzig