

Appendix A: Affymetrix U95A Experiment

T. M. Therneau K. V. Ballman

July 2005

This appendix contains the complete set of curves for the Affymetrix U95A spike-in data referenced in Technical Report 74 [1].

1 Data

This is a subset of Affymetrix data that was used to develop the MAS 5.0 algorithm, which generates gene expression intensities values from the 16 perfect match (PM) and mismatch (MM) probe pairs. This dataset was created by Affymetrix and is publicly available at their web site www.affymetrix.com; search on the phrase “Latin square data” to find the link to the page containing a description of the experiment and the downloadable files of data. Human cRNA fragments corresponding to 16 probesets on the HGU95A GeneChip were added to the hybridization mixture. Each array started with the same hybridization mixture from a common tissue source and the 16 probesets were spiked in concentrations ranging from 0 to 1024 pM. Specifically, the 16 cRNAs were spiked in according to 14 different concentrations (0, 0.25, 0.5, 1, 2, 4, . . . , 512, 1024 pM). Each concentration pattern of the 16 probesets were replicated three times with the exception of one pattern, which was only replicated twice; there were 59 arrays in this experiment. The cRNAs were spiked in at a different concentration on each array (apart from the replicates) arranged in a cyclic Latin square design. A more detailed description of this experiment is given by Irizarry et al. [2].

2 Plot description

We chose to use the logistic function to fit the data from this spiked-in gene expression experiment, where x represents the level of known spiked-in transcripts concentration and y is the observed intensity value produced by the array.

To fit the perfect match (PM) and mismatch (MM) probes simultaneously, we used a five-parameter version of the model

$$\log(y) = a + bf(c[\log(x) - d - e]) \quad (1)$$

where

- f is the logistic function defined above,
- a is the lower threshold for the probe pair (i.e. the background level),
- b is the range of the curve,
- $a + b$ is the upper threshold for the probe pair (i.e. the saturation level),
- $bc/4$ is the slope of the curve at its inflection point,
- d is the inflection point for the PM probes, and

- $d + e$ is the inflection point for the MM probes.

All the plots are presented on a log base 2 scale. A least squares approach was used to fit the curves to the data. The observed expression values were plotted on the y -axis and the spike-in concentrations were plotted on the x -axis. A plot was made for each probe in the probeset; both the perfect match (PM) and mismatch (MM) values were plotted using different symbols. Plots for all the probe pairs within a probeset, referred to as a panel, were plotted on the same page in the order they appeared on the gene. Fitted logistic calibration curves are superimposed on the data plots.

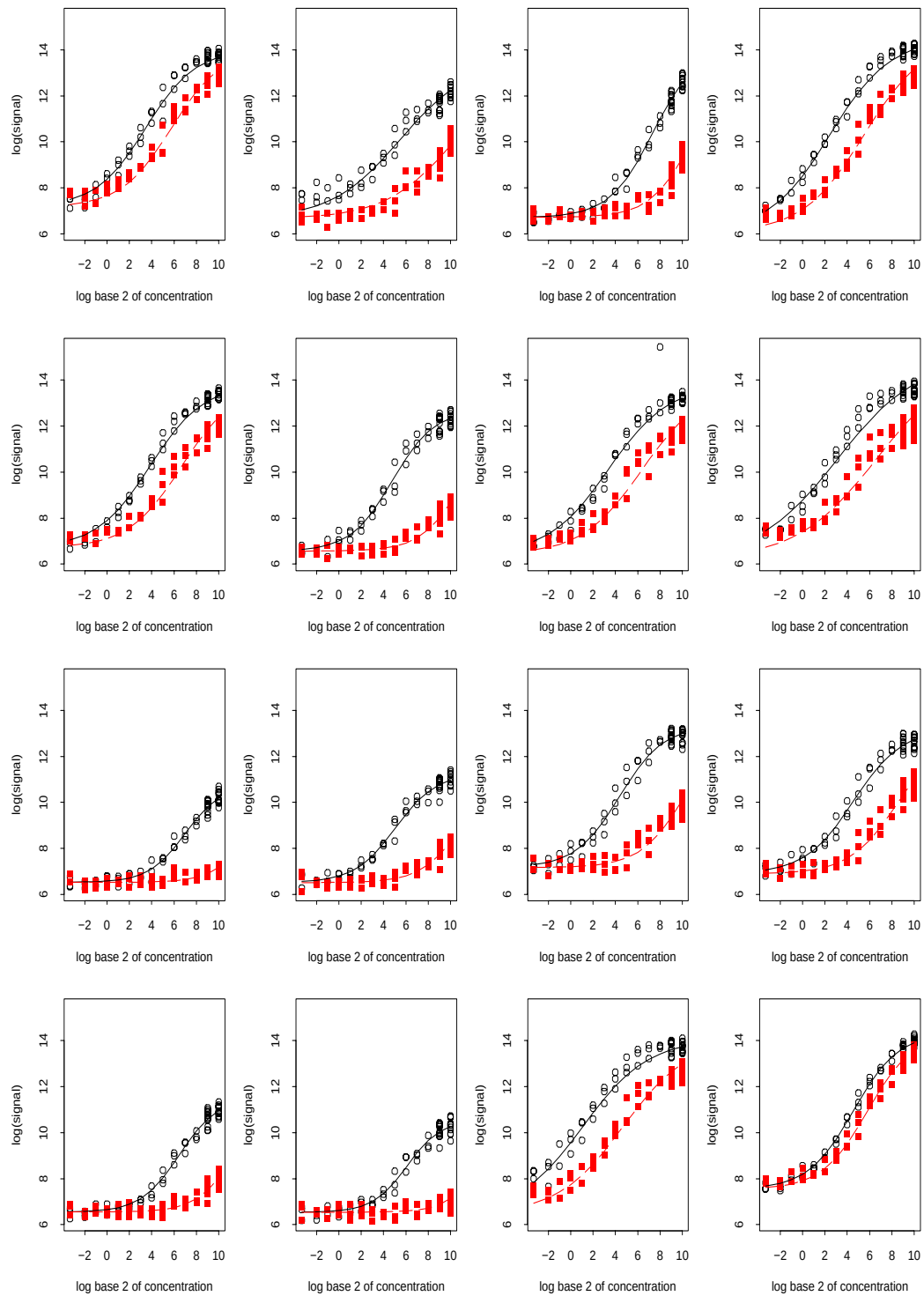
3 The plots

Each page or panel contains 16 plots (one per probeset) and there are 16 panels of plots yielding 256 plots. The black open circles are the values of the PM probes and the red filled squares are the values of the MM probes. The black solid line corresponds to the fitted logistic calibration curve for the PM probes and the red dashed line corresponds to the fitted logistic calibration curves for the MM probes; these curves were the results of a simultaneous fit.

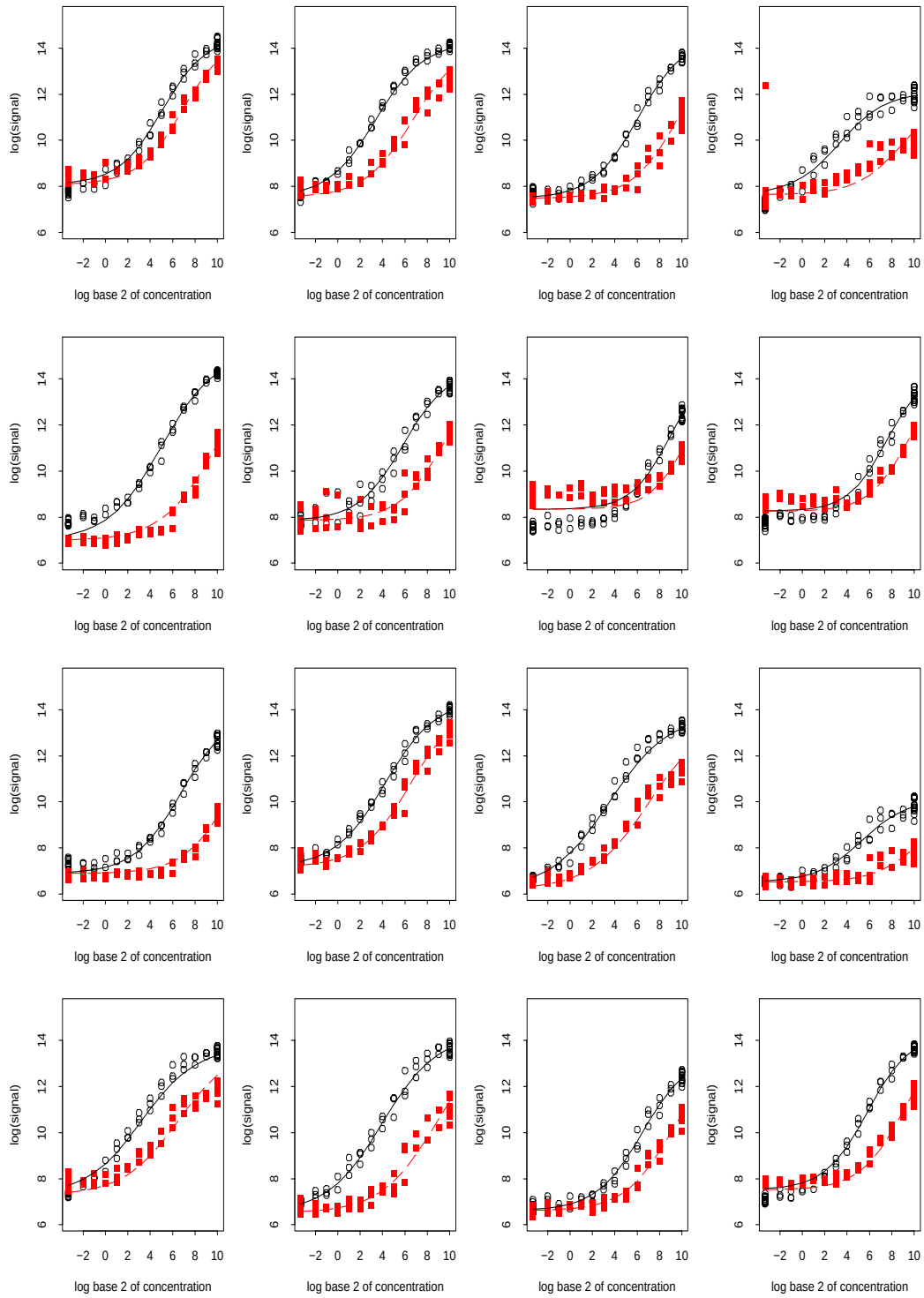
References

- [1] K. V. Ballman and T. M. Therneau. A exploration of affymetrix probe-set intensities in spike-in experiments. Technical Report 74, Mayo Clinic College of Medicine, March 2005.
- [2] R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed. Summaries of affymetrix genechip probe level data. *Nucleic Acids Research*, 31:e15, 2003.

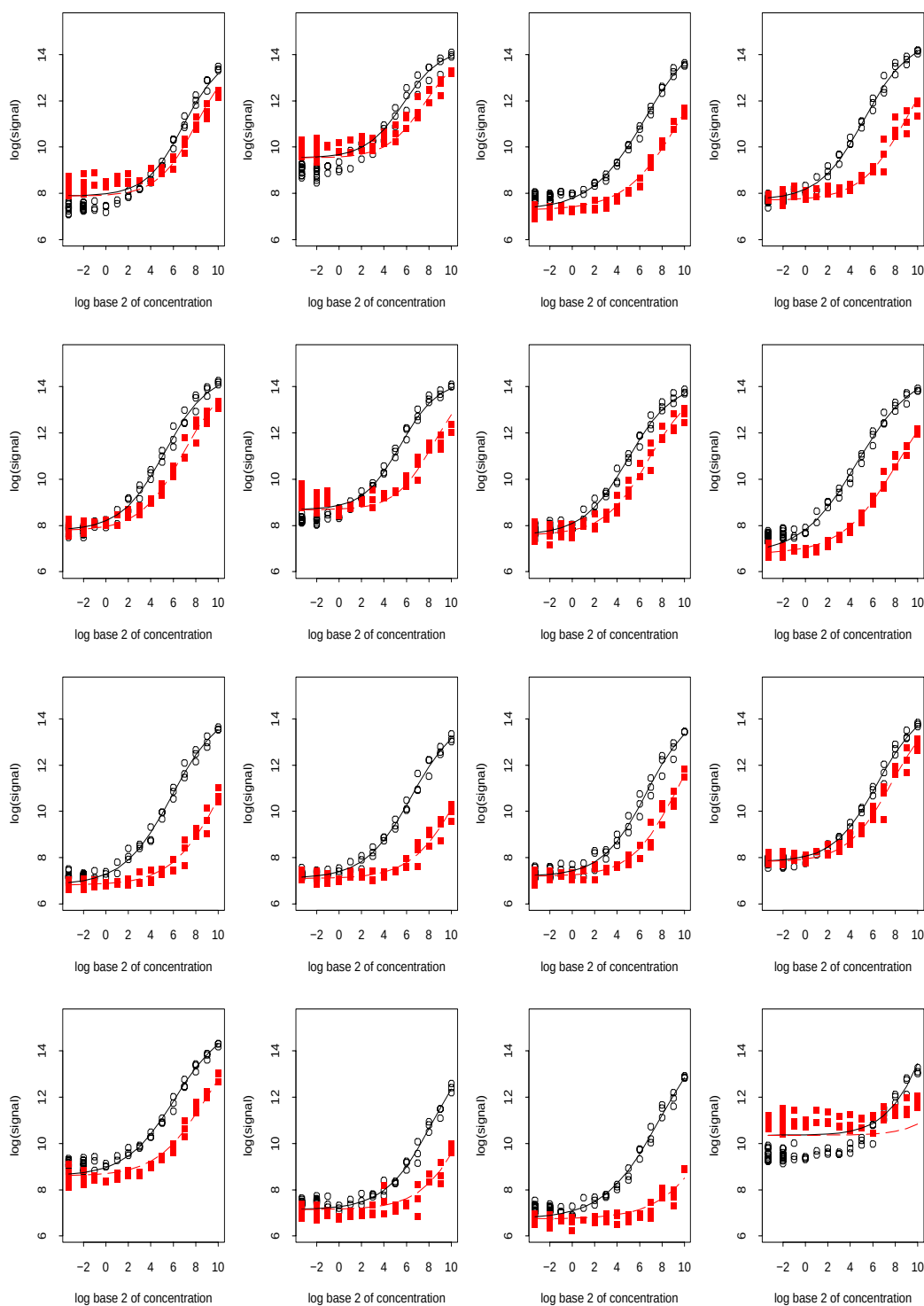
Affy U95 spikein, gene 1 (37777_at)



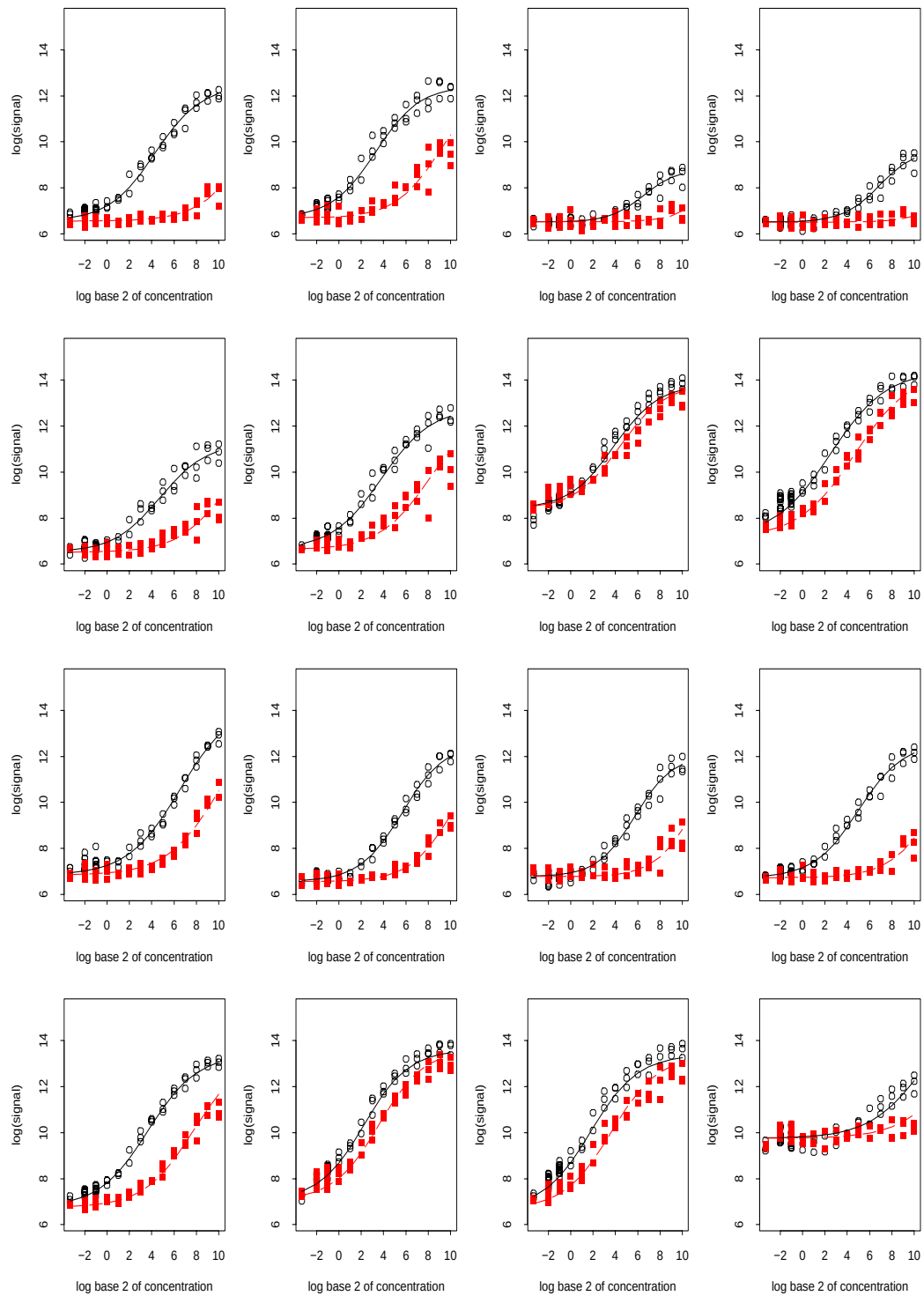
Affy U95 spikein, gene 2 (684_at)



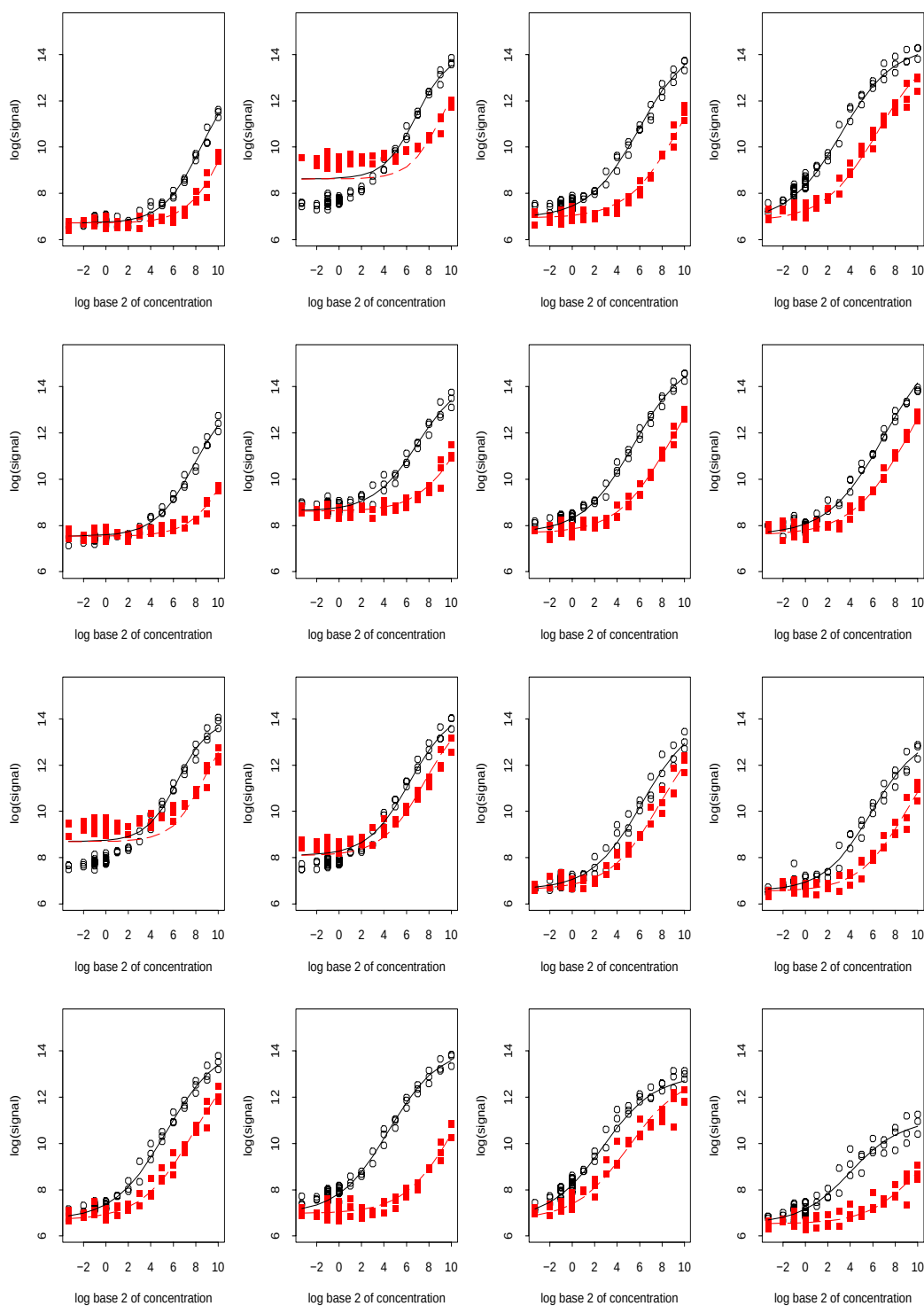
Affy U95 spikein, gene 3 (1597_at)



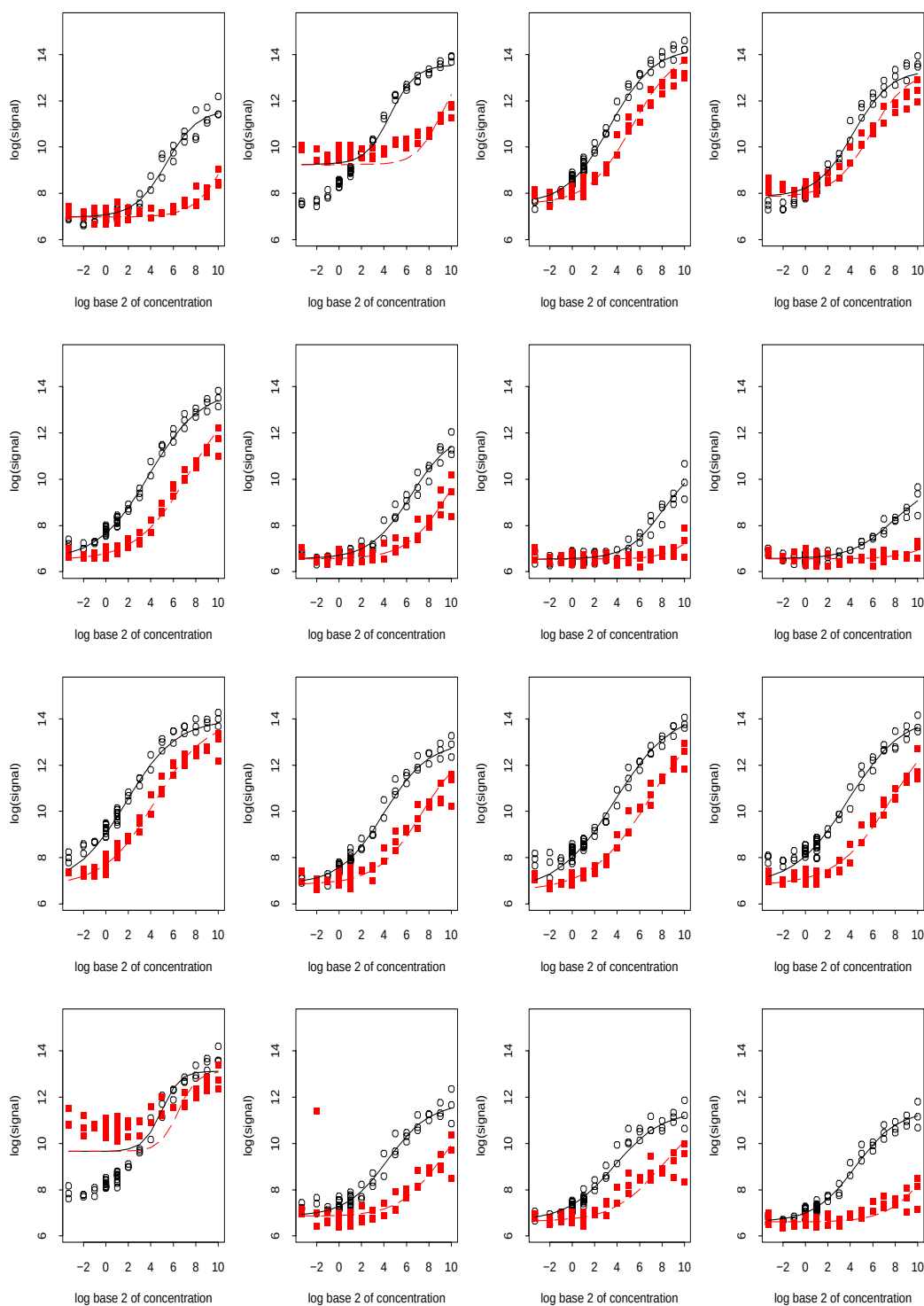
Affy U95 spikein, gene 4 (38734_at)



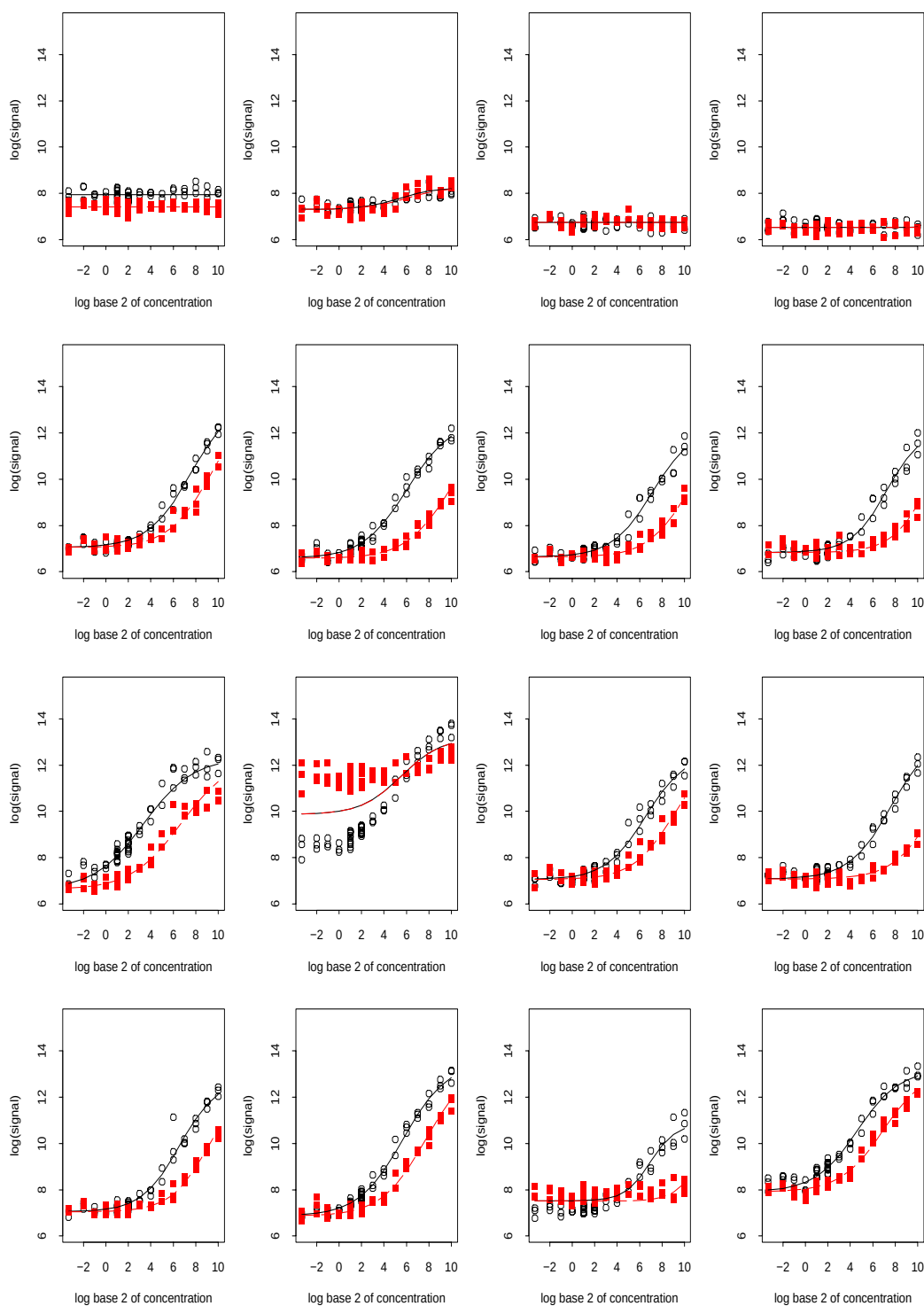
Affy U95 spikein, gene 5 (39058_at)



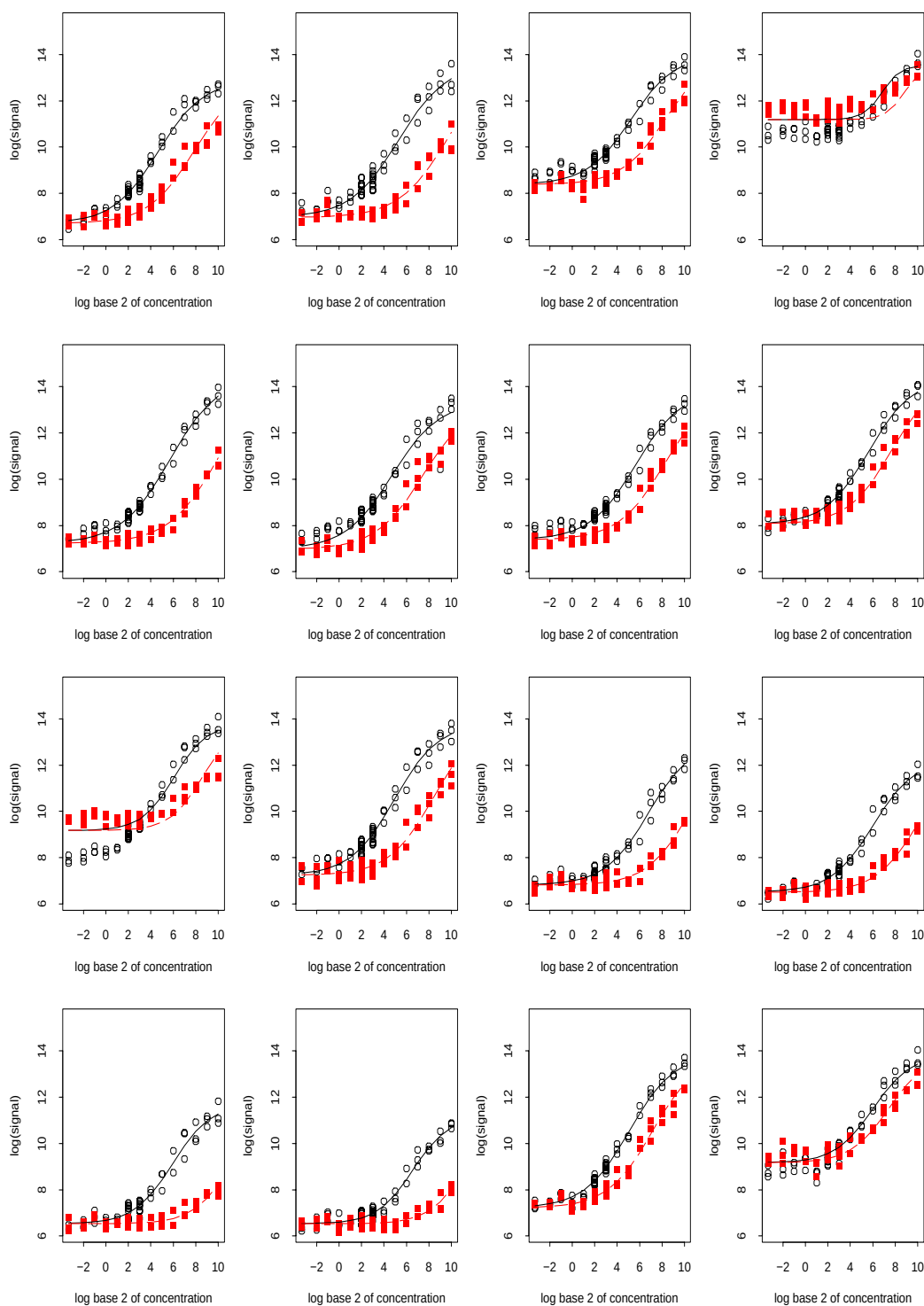
Affy U95 spikein, gene 6 (36311_at)



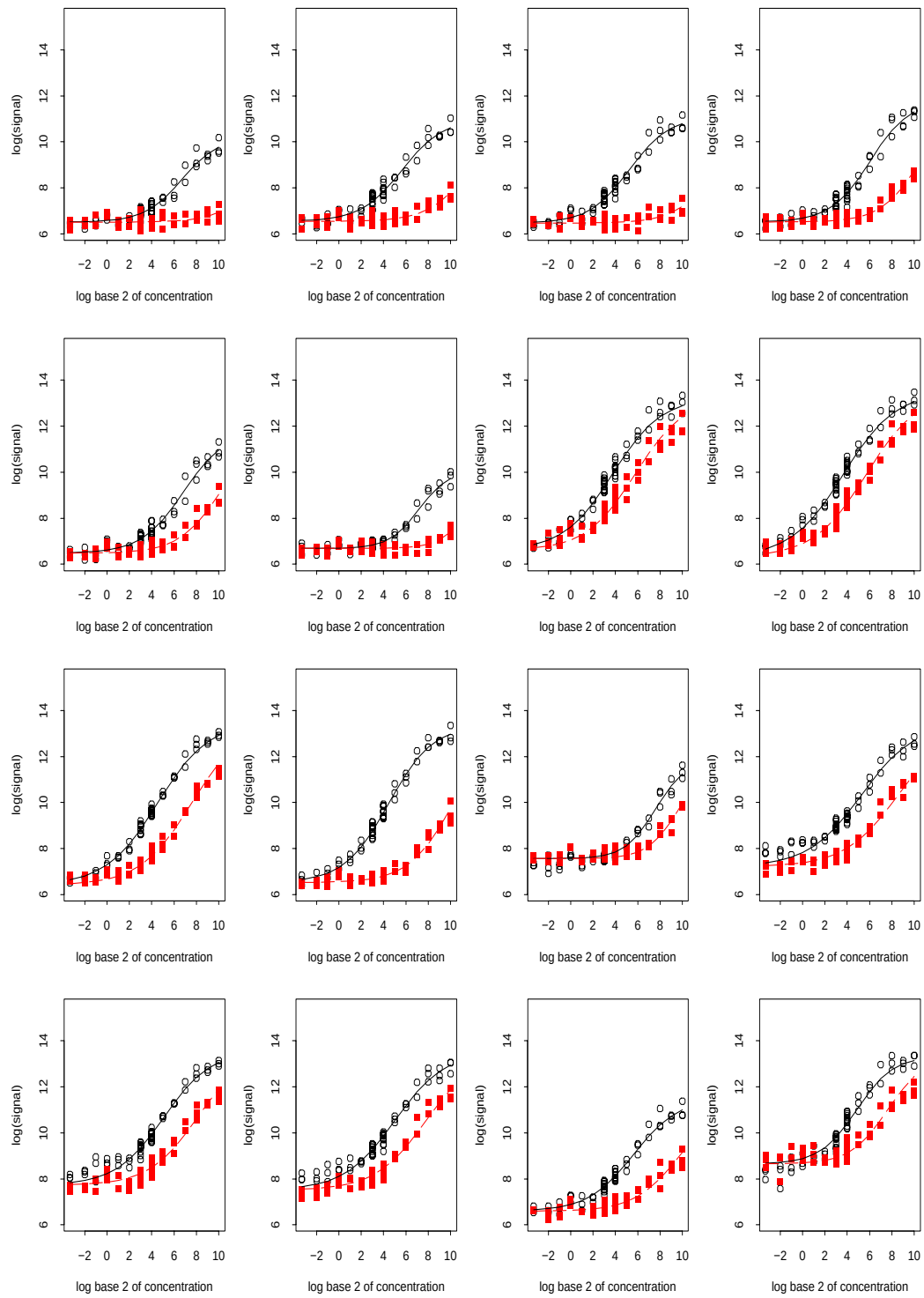
Affy U95 spikein, gene 7 (36889_at)



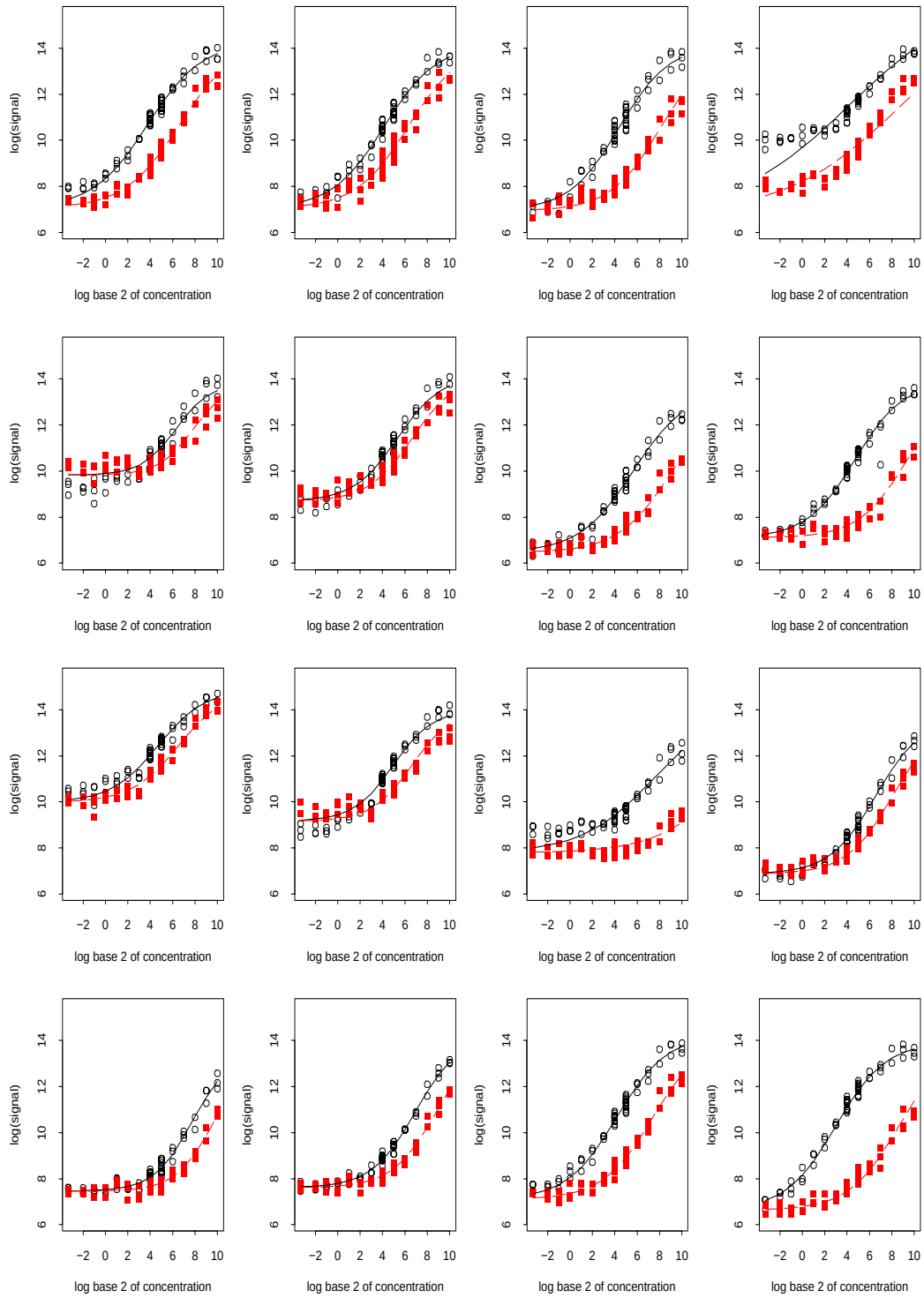
Affy U95 spikein, gene 8 (1024_at)



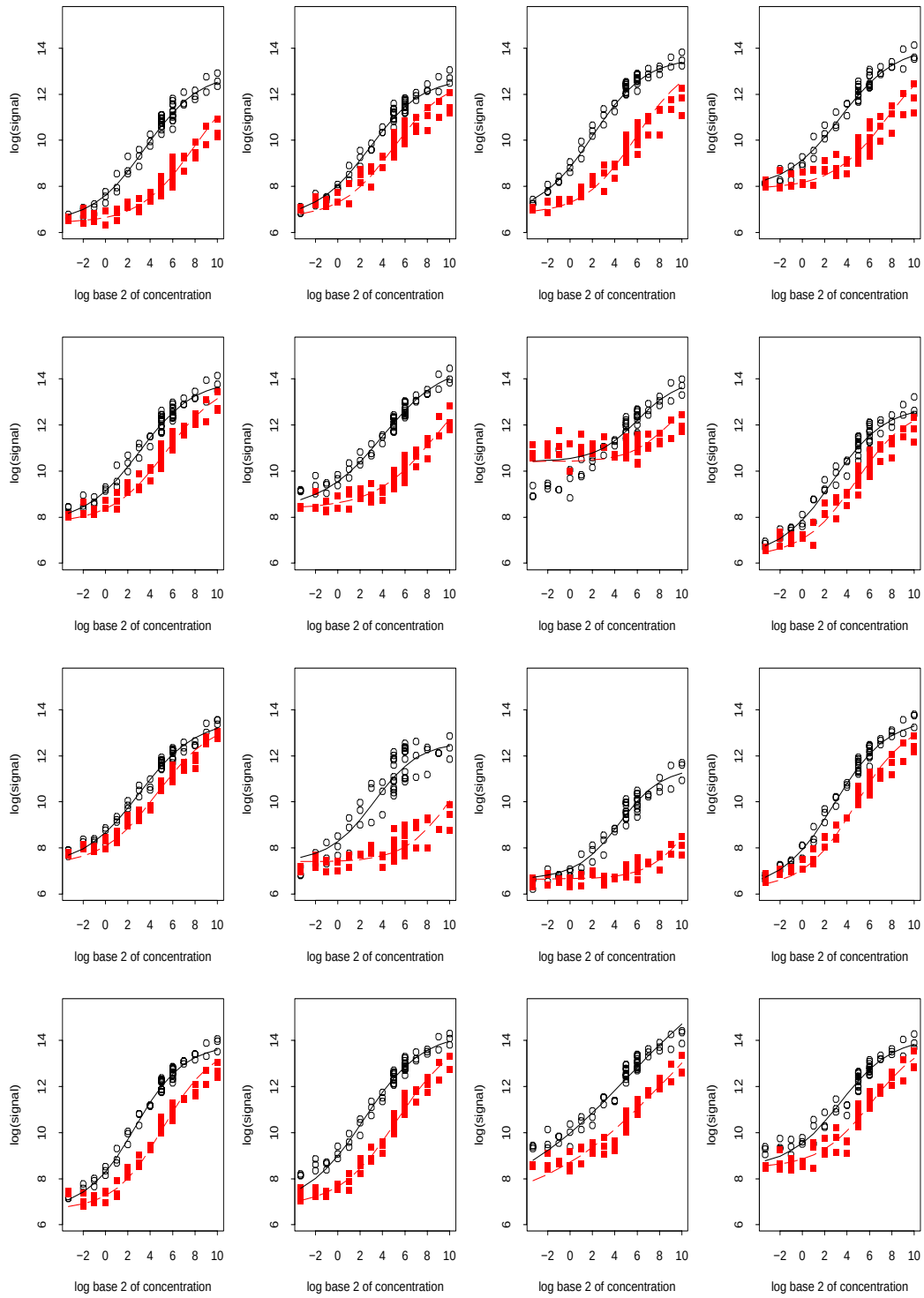
Affy U95 spikein, gene 9 (36202_at)



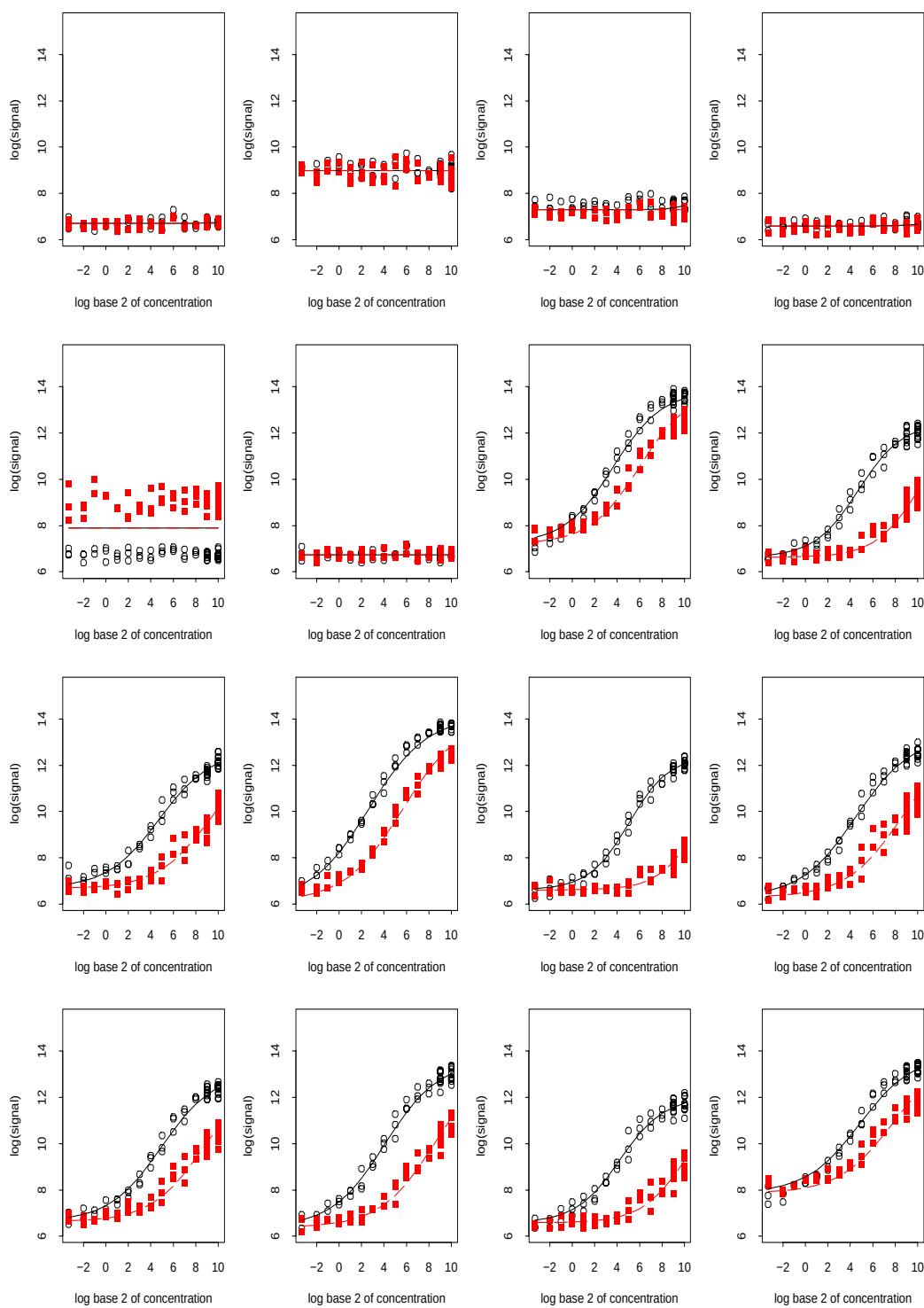
Affy U95 spikein, gene 10 (36085_at)



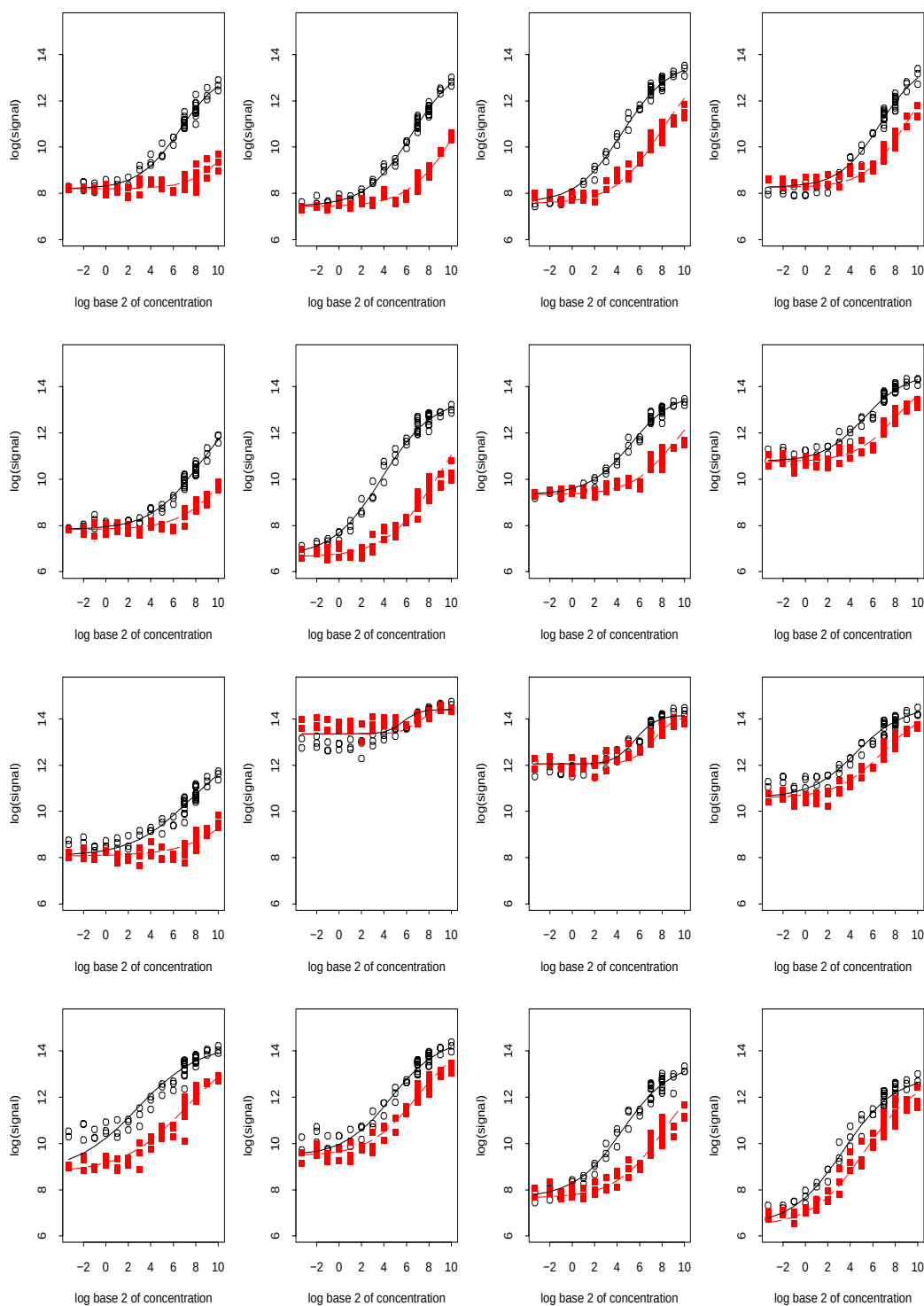
Affy U95 spikein, gene 11 (40322_at)



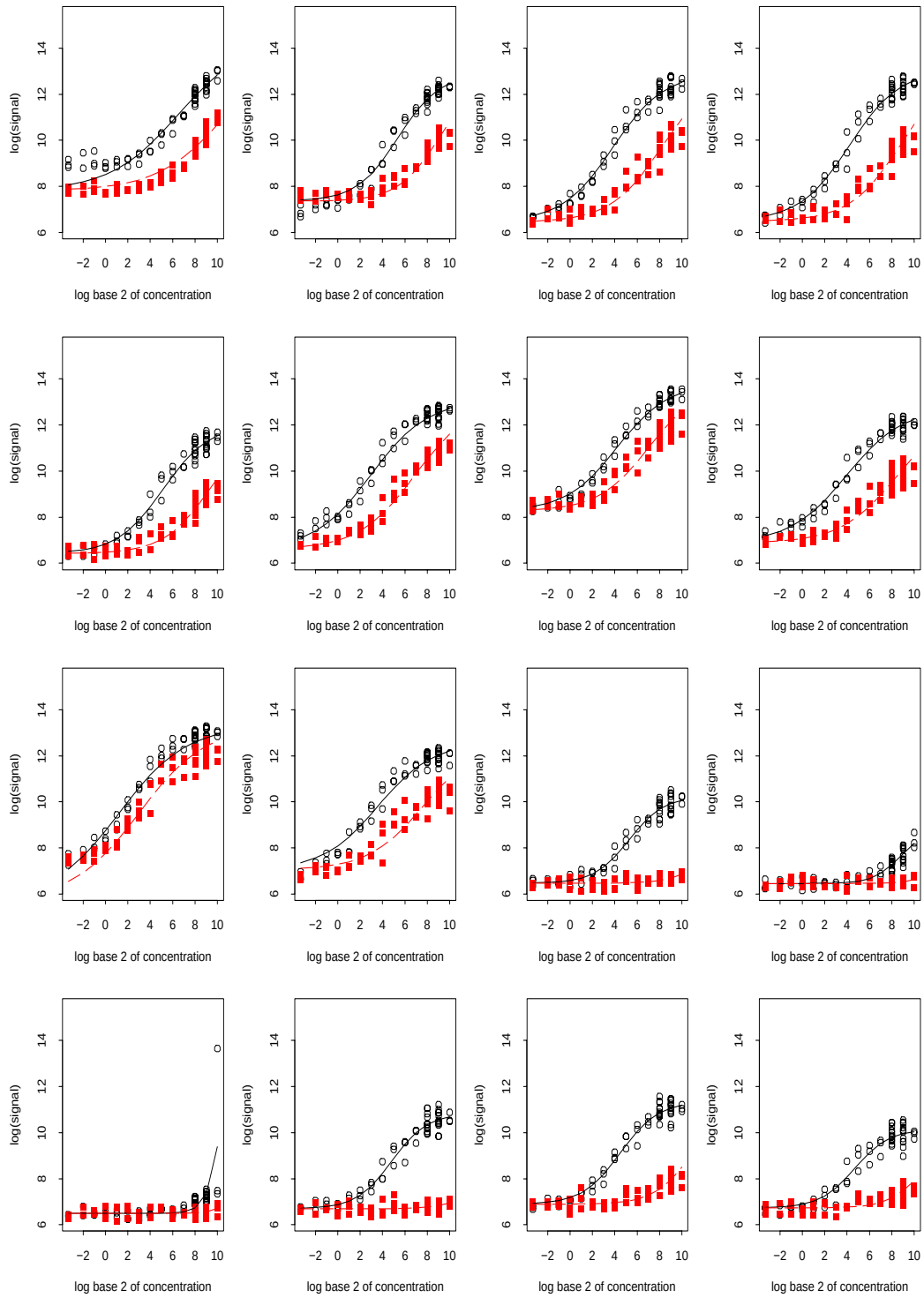
Affy U95 spikein, gene 12 (407_at)



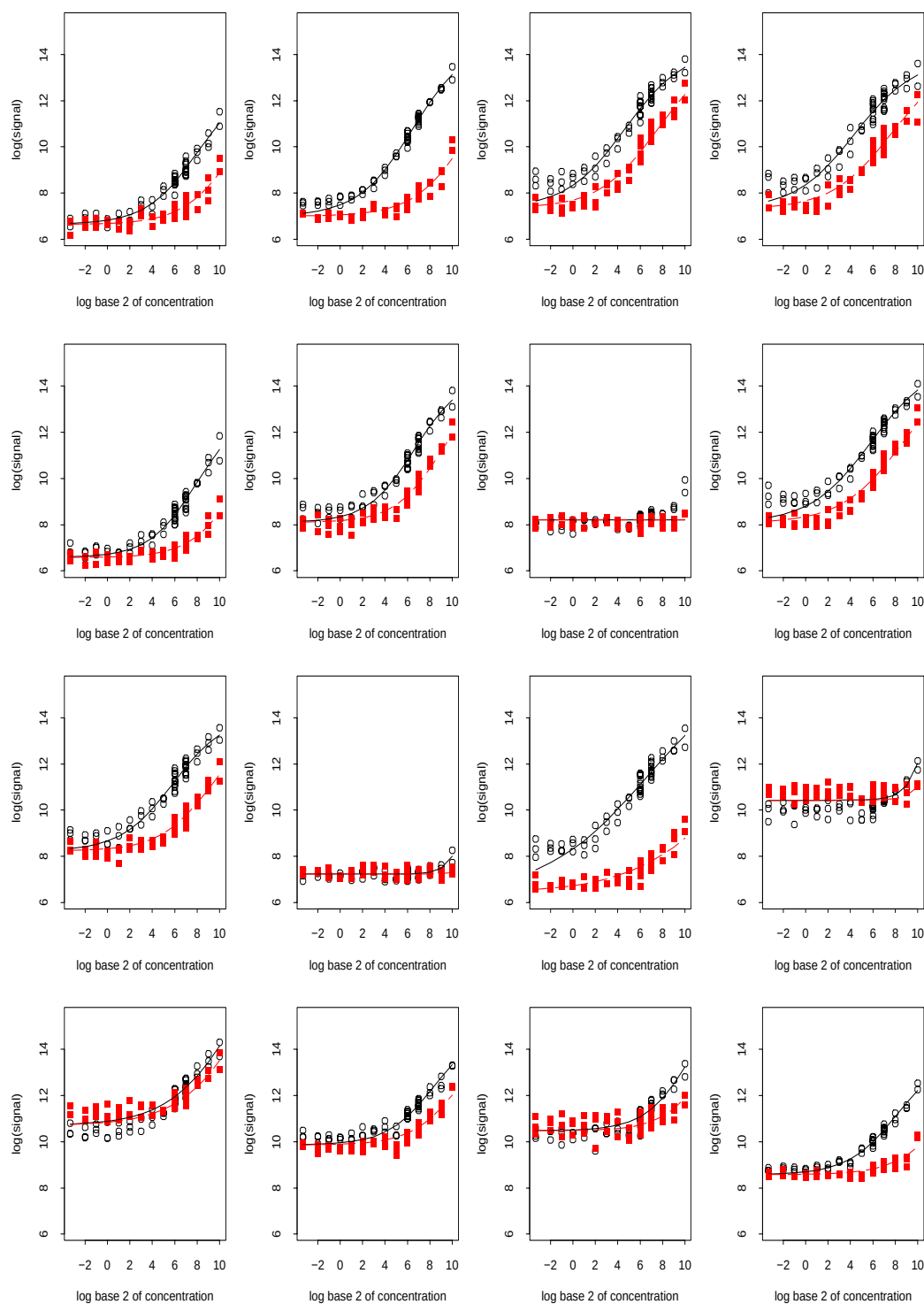
Affy U95 spikein, gene 13 (1091_at)



Affy U95 spikein, gene 14 (1708_at)



Affy U95 spikein, gene 15 (33818_at)



Affy U95 spikein, gene 16 (546_at)

