Appendix B: Data from Three Gene Logic Spike-In Experiments

T. M. Therneau K. V. Ballman

July 2005

This appendix contains the complete set of curves for the three Gene Logic spike-in experiments referenced in Technical Report 74 [1].

1 Data

The structure of the experiment is nicely described in a white paper that accompanies the Gene Logic dataset. The dataset can be requested from Gene Logic website, `www.genelogic.com/media/studies`. This experiment is also described by Irizarry et al. [2]. The data consists of three experiments using the U95A and U95Av2 GeneChips. In all three instances, there are 11 spiked in genes. These are bacterial genes normally used as controls in the array process, and each is represented by 20 probe pairs on the array.

Experiment 1 consists of 26 arrays. CreX-3 is at 0 concentration on all arrays and the remaining ten genes are at 0, 0.5, 0.75, 1, 1.5, 2, and 3 pM (one array each), 5 and 100 pM (two arrays each), and 12.5, 25, 50, 75, and 150 pM (three arrays each). All arrays had a common complex cRNA derived from an acute myeloid leukemia (AML) tumor cell line. Experiment 2 used the latin square design. The background is again an AML tumor cell line. Experiment 3 uses a different latin square with concentrations of 0.5, 0.75, 1, 1.5, 2, 3, 5, 12.5, 25, 50, 75, and 100. There are three replicates of 12 patterns giving 36 arrays in total. The background material for this experiment is from a tonsil tissue sample.

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

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

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,

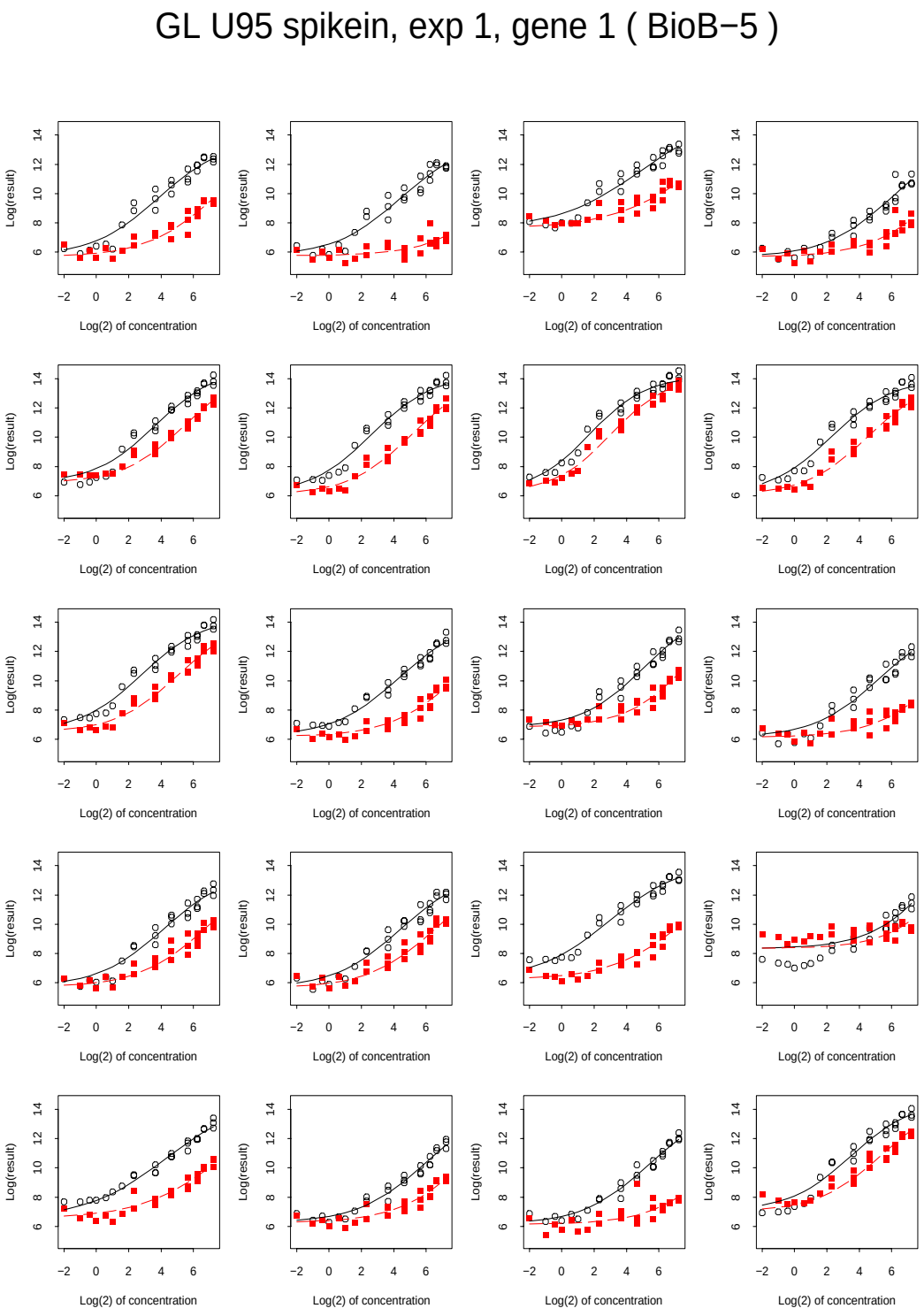
- d is the inflection point for the PM probes, and
- $d + e$ is the inflection point for the MM probes.

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.

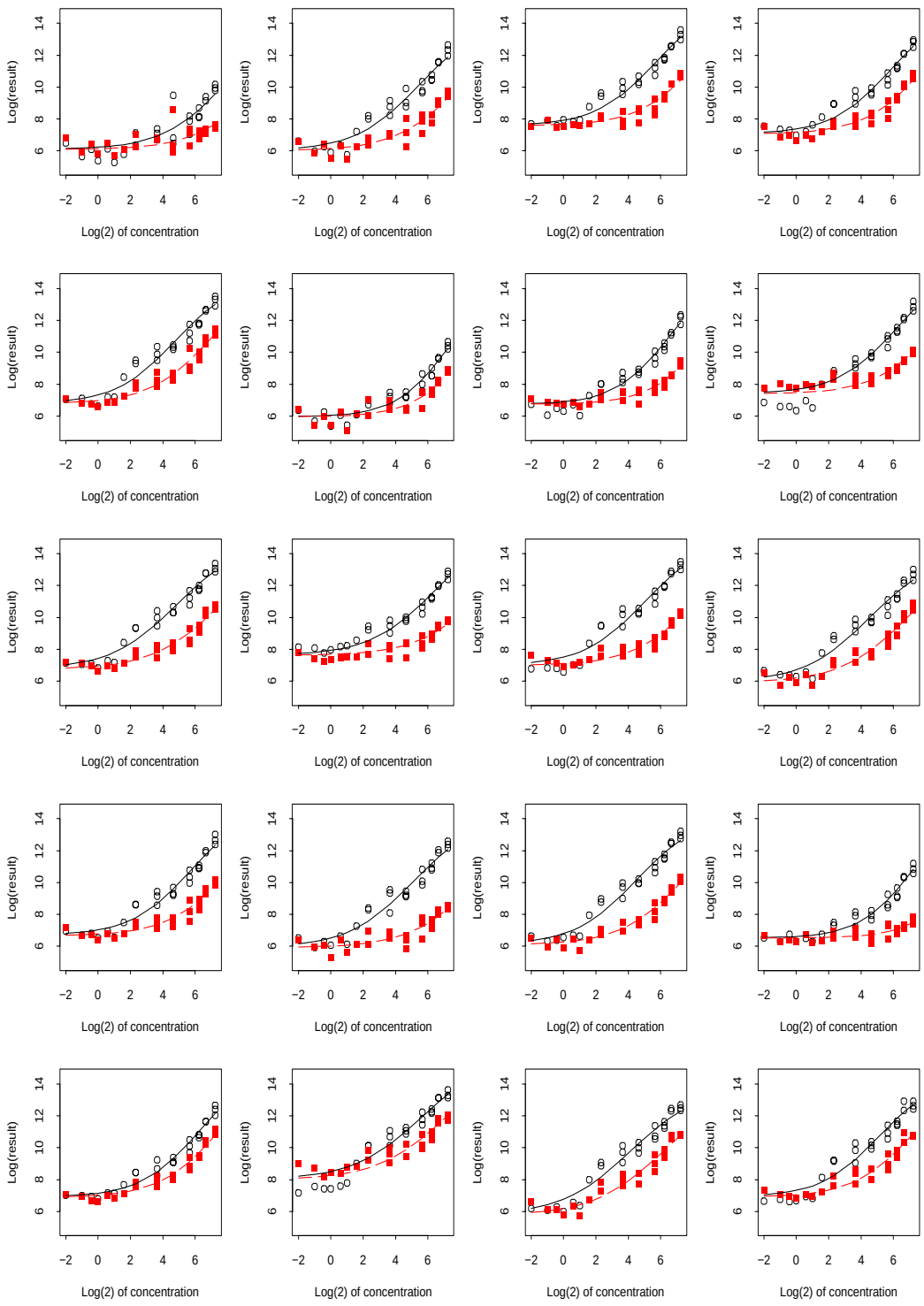
3 The plots

Each page or panel contains 20 plots (one per probeset) and there are 11 panels of plots yielding 220 plots for each of the three experiments. 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.

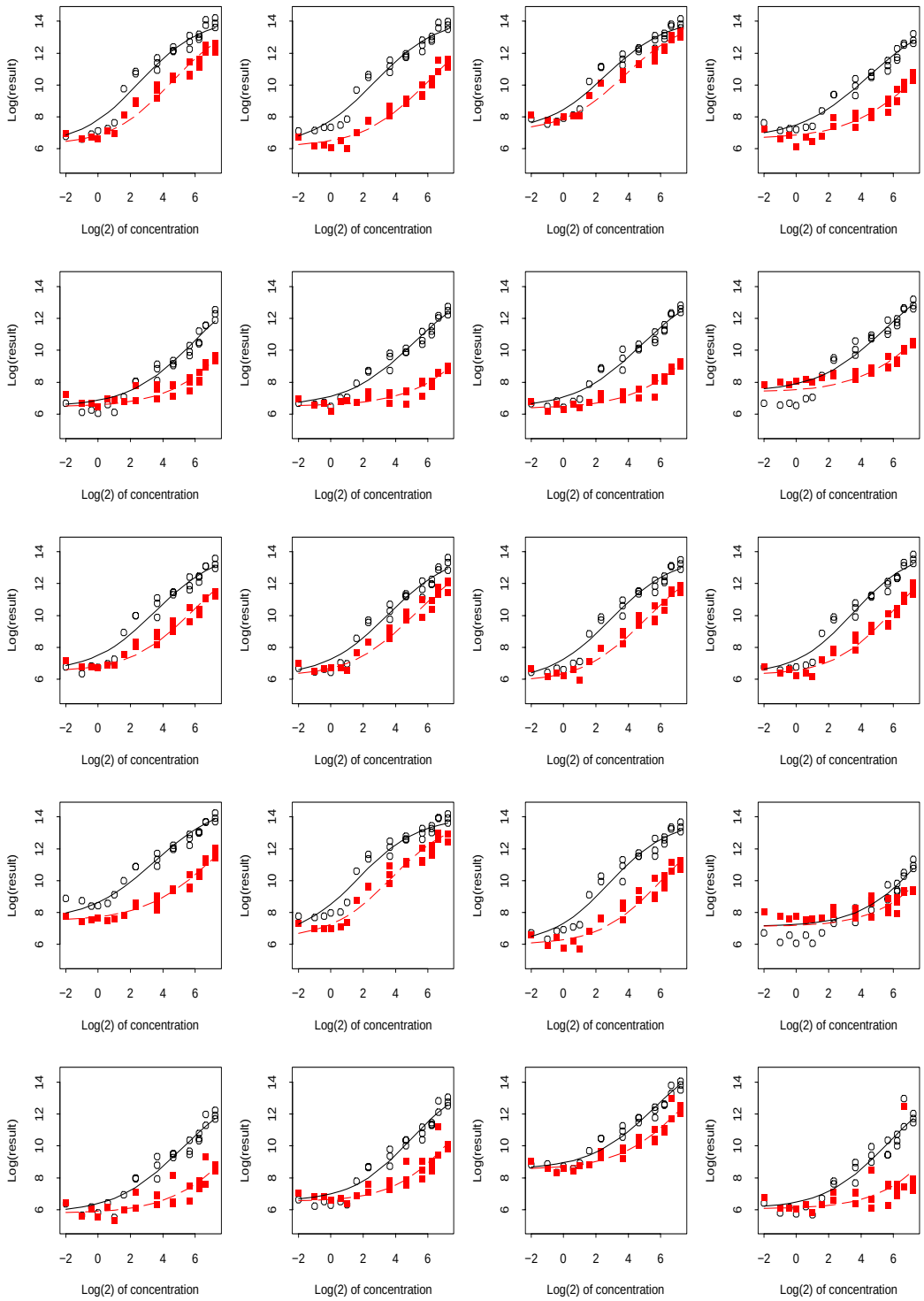
3.1 Experiment 1 plots



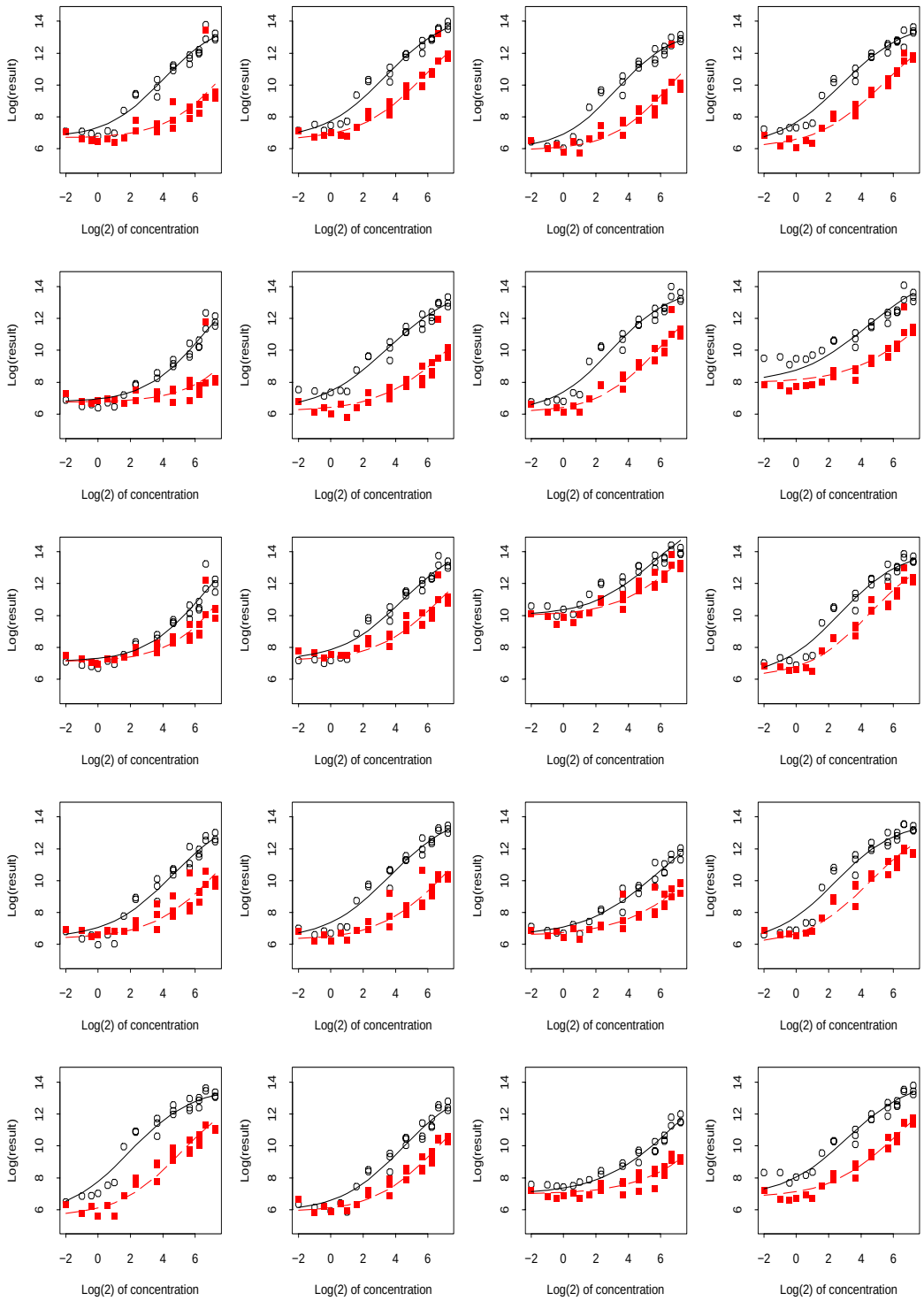
GL U95 spikein, exp 1, gene 2 (BioB−M)



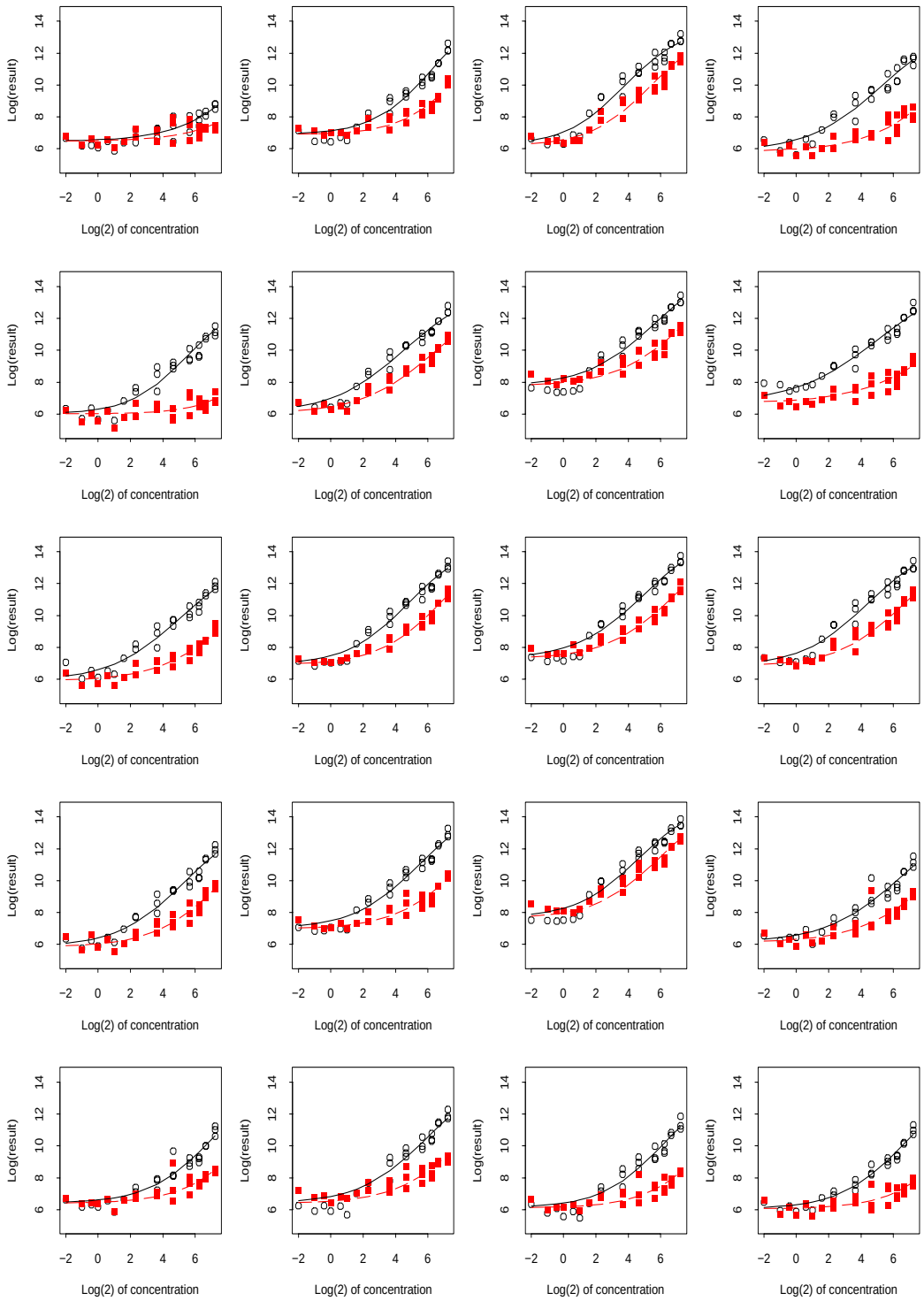
GL U95 spikein, exp 1, gene 3 (BioB−3)



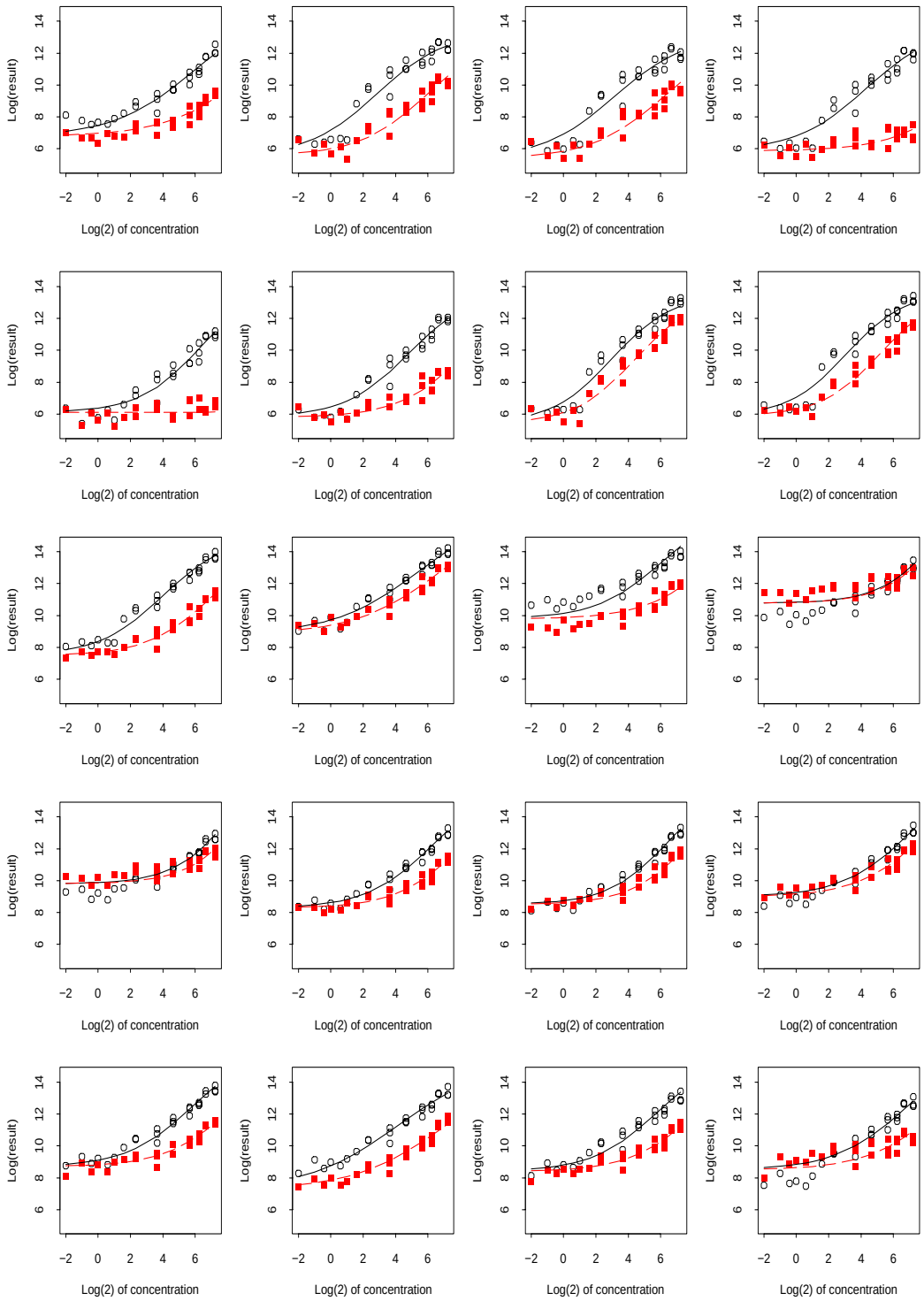
GL U95 spikein, exp 1, gene 4 (BioC−5)



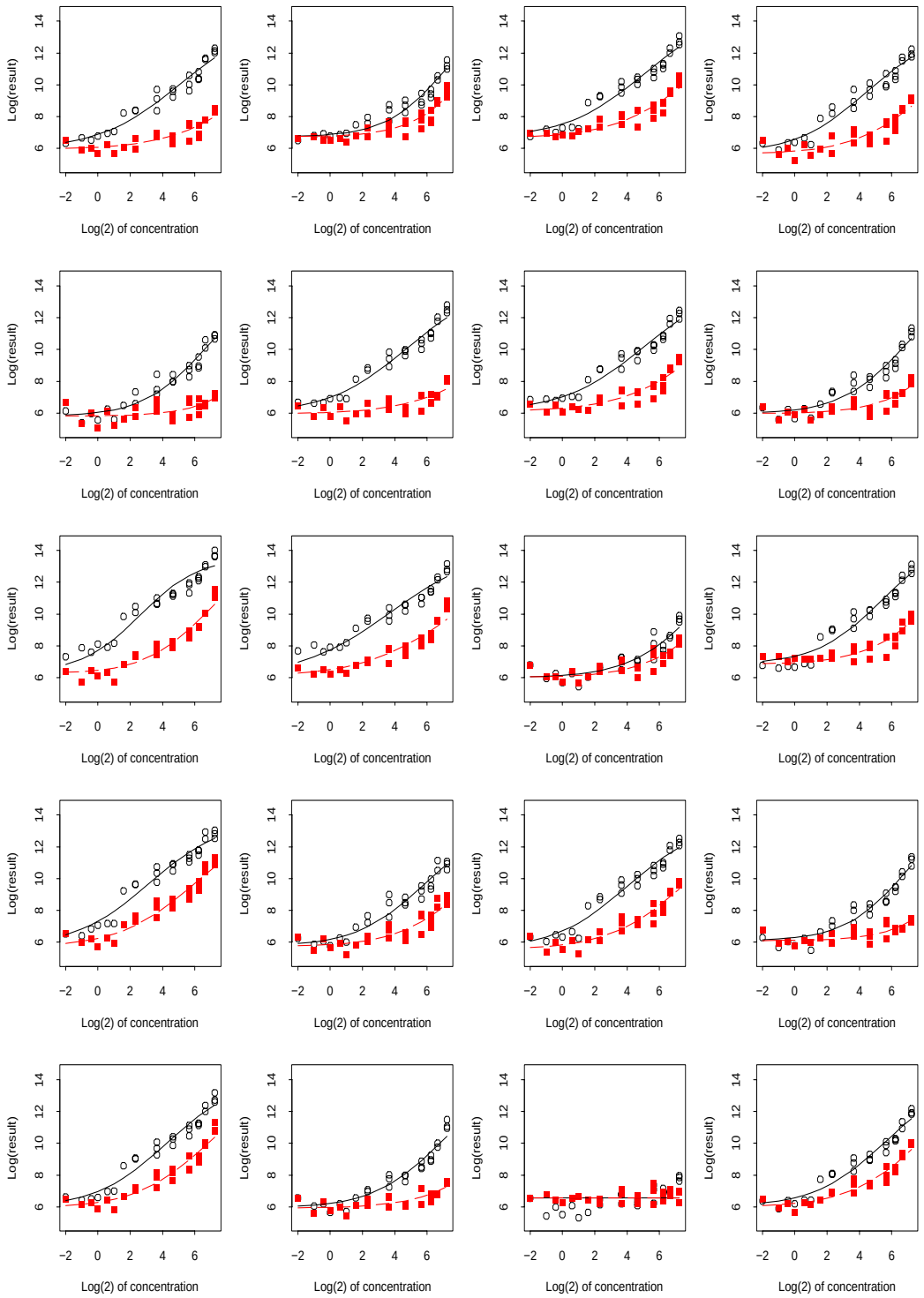
GL U95 spikein, exp 1, gene 5 (BioC−3)



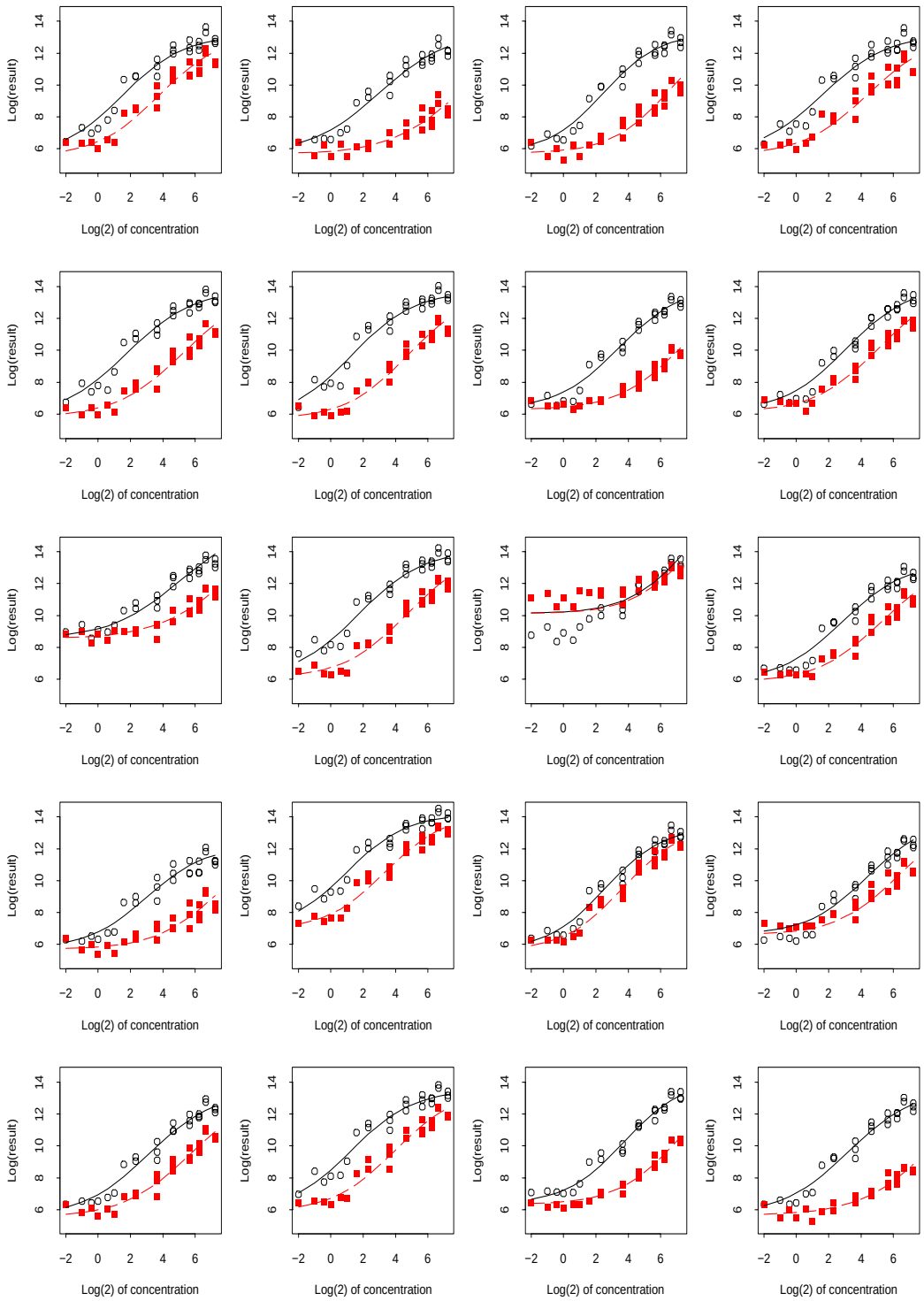
GL U95 spikein, exp 1, gene 6 (BioDn−3)



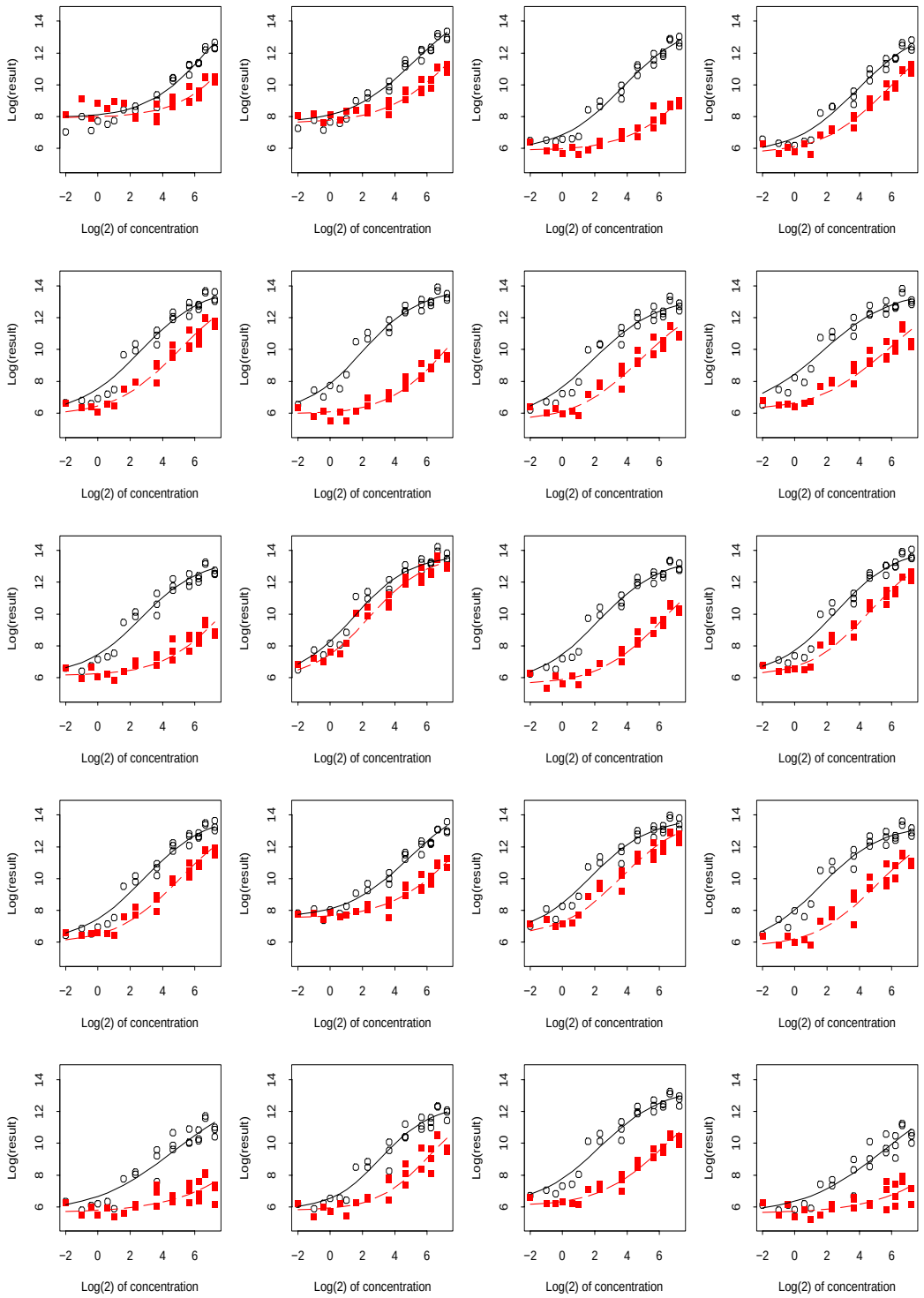
GL U95 spikein, exp 1, gene 7 (DapX−5)



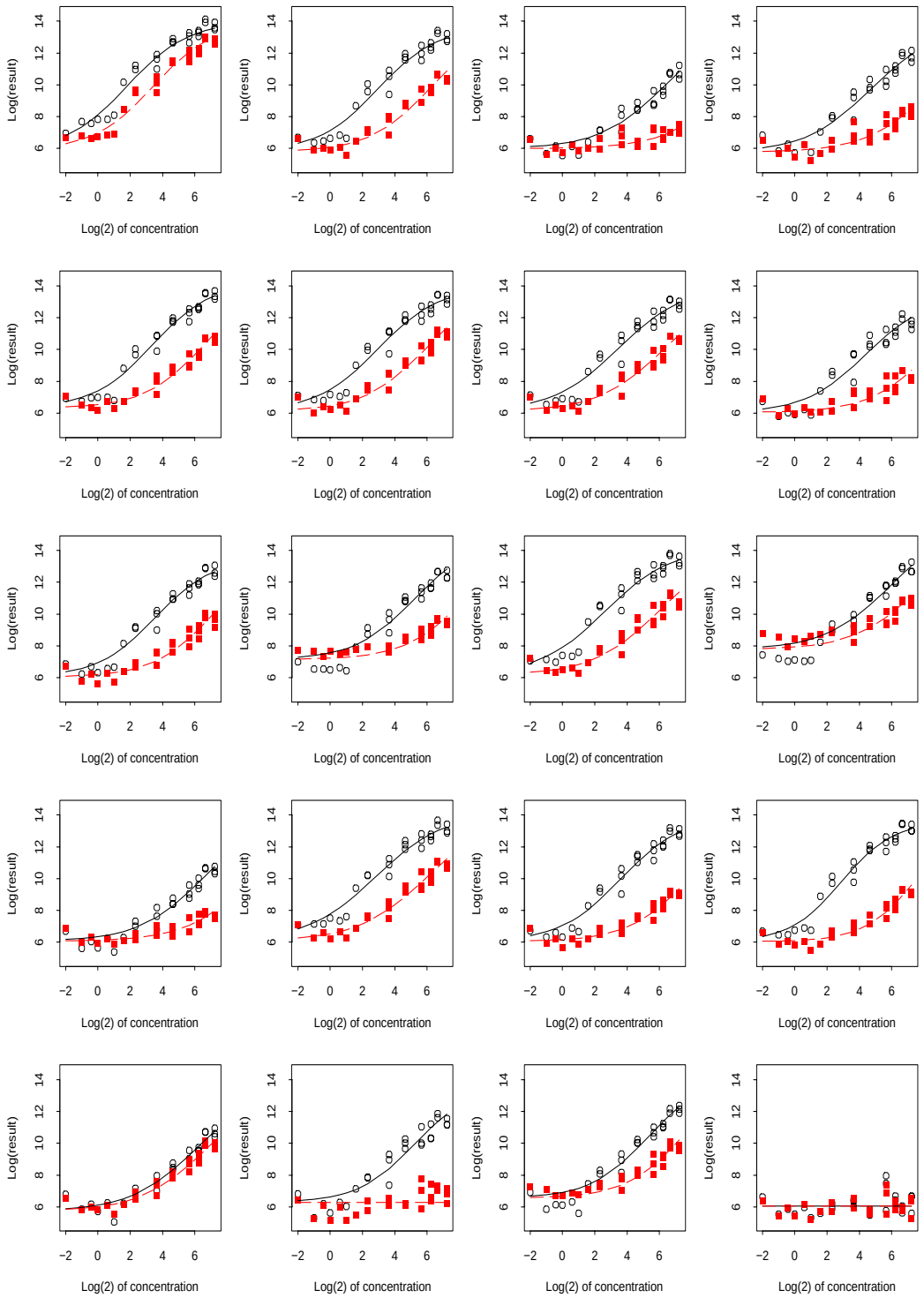
GL U95 spikein, exp 1, gene 8 (DapX−M)



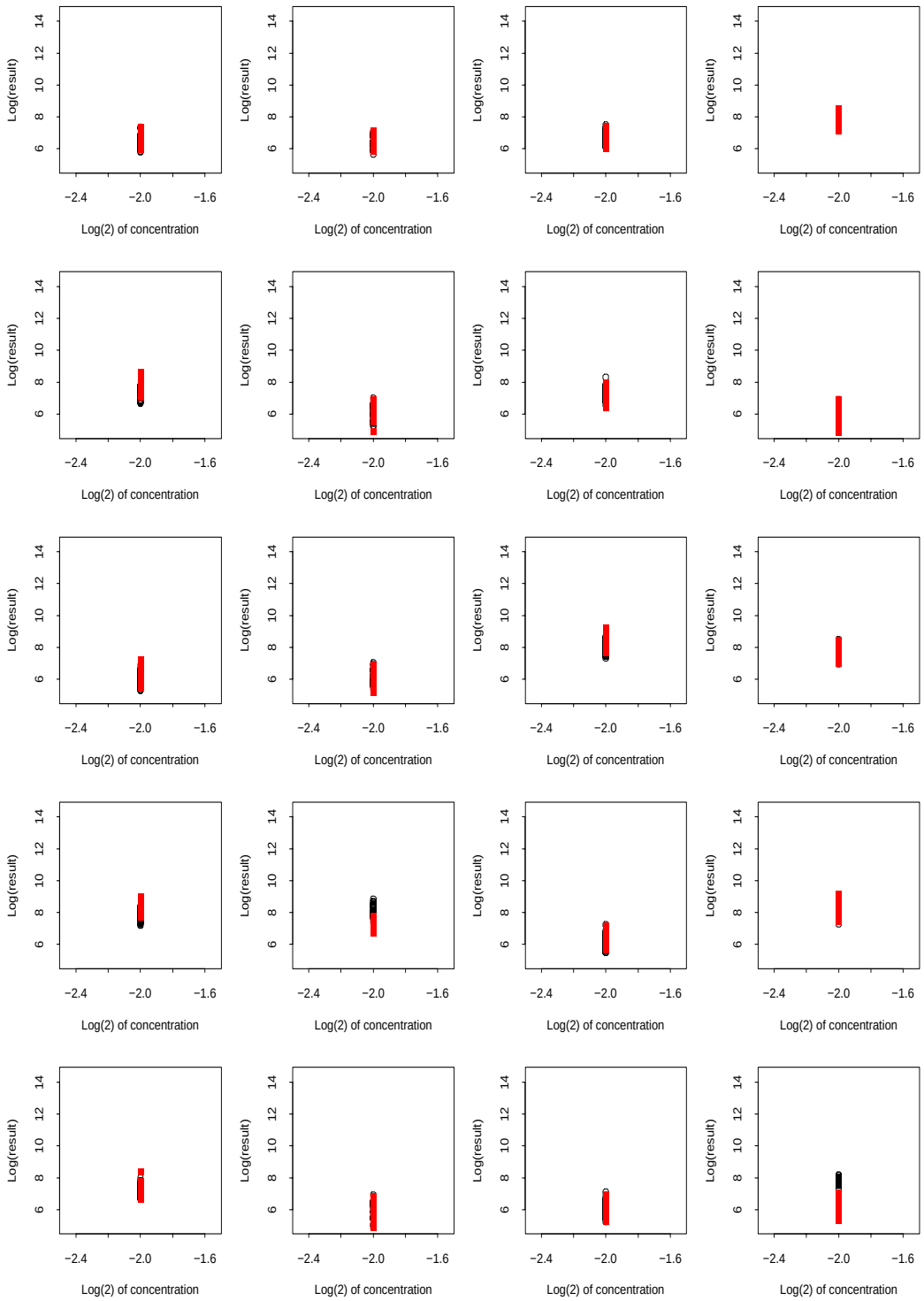
GL U95 spikein, exp 1, gene 9 (DapX−3)



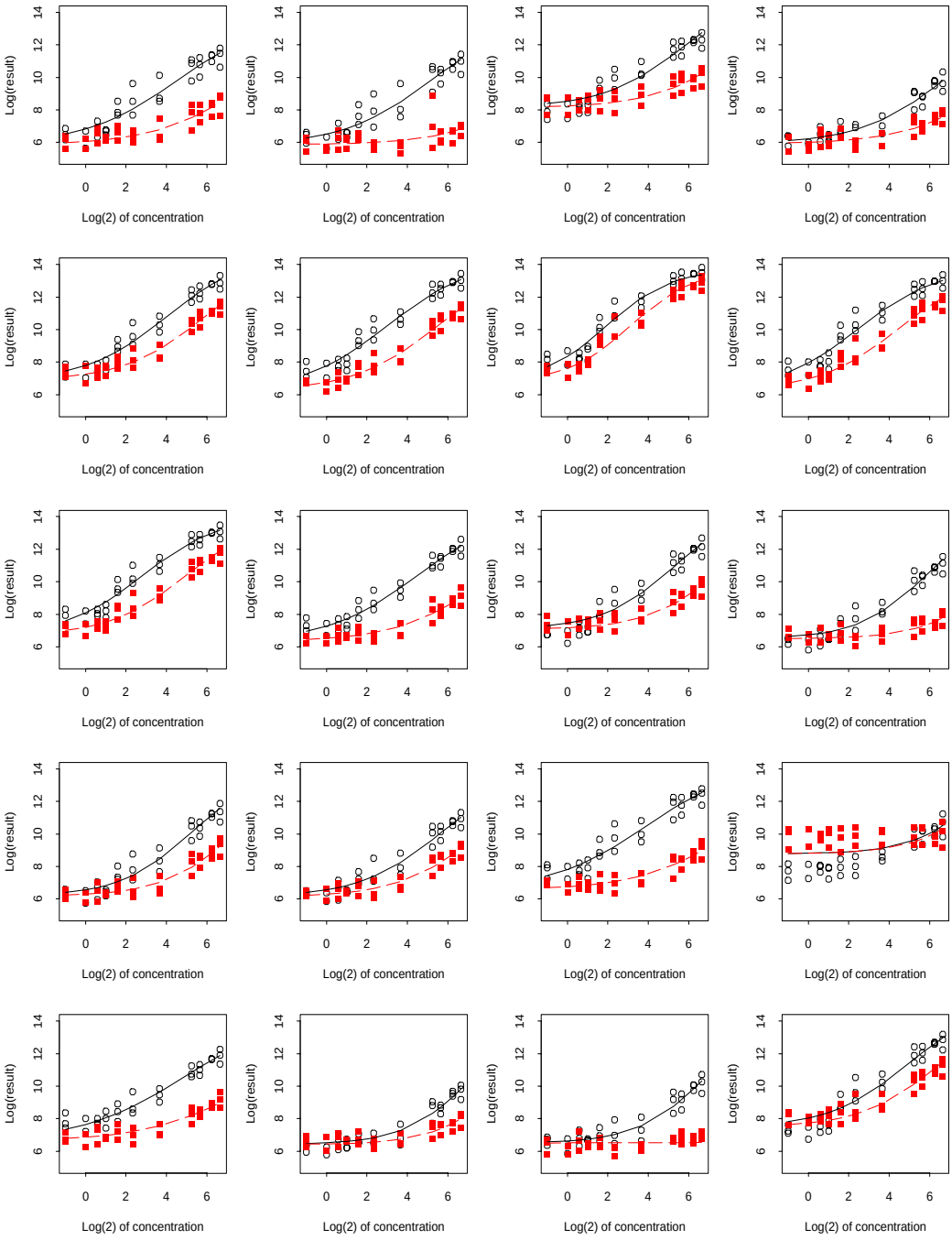
GL U95 spikein, exp 1, gene 10 (CreX−5)



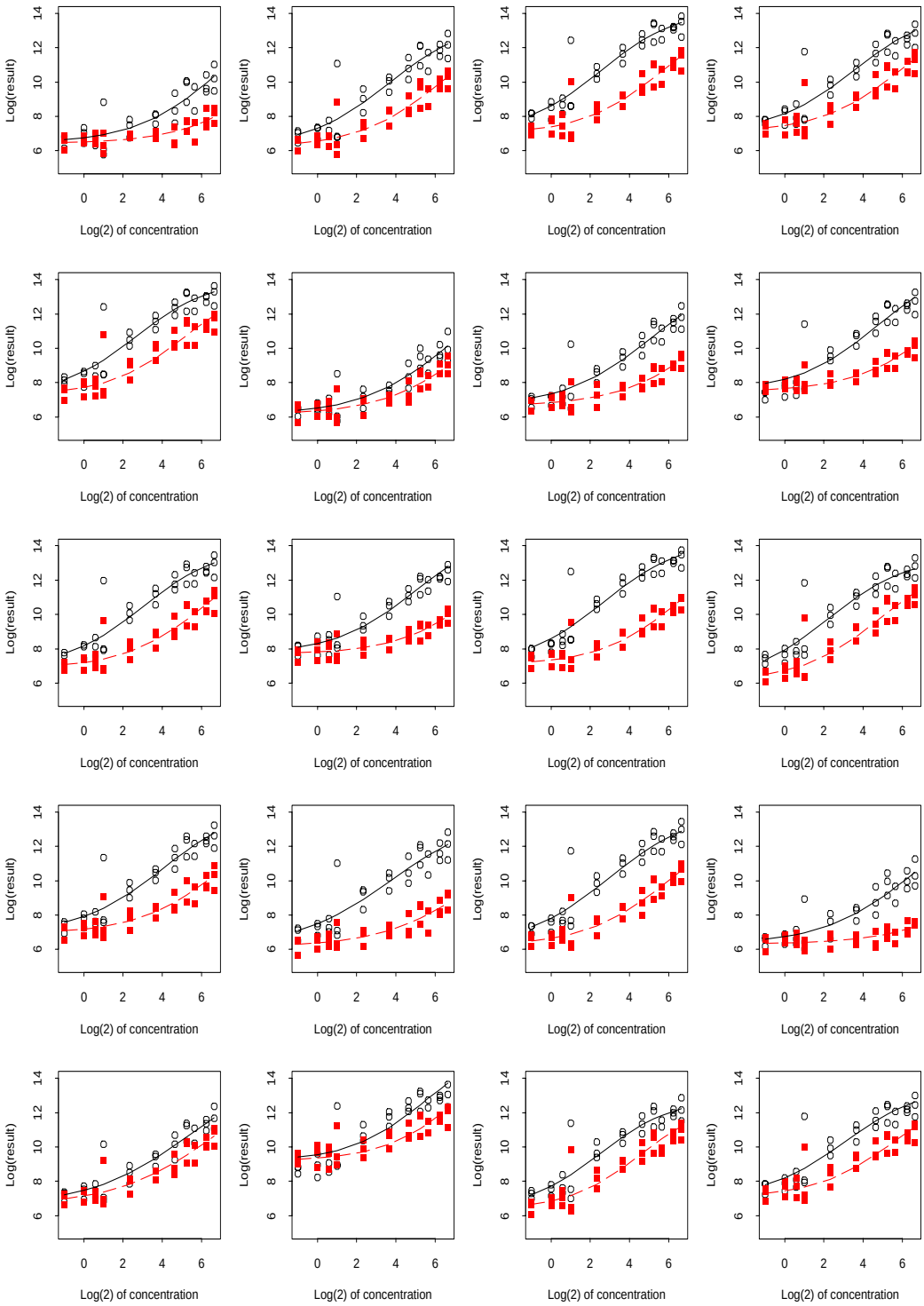
GL U95 spikein, exp 1, gene 11 (CreX−3)



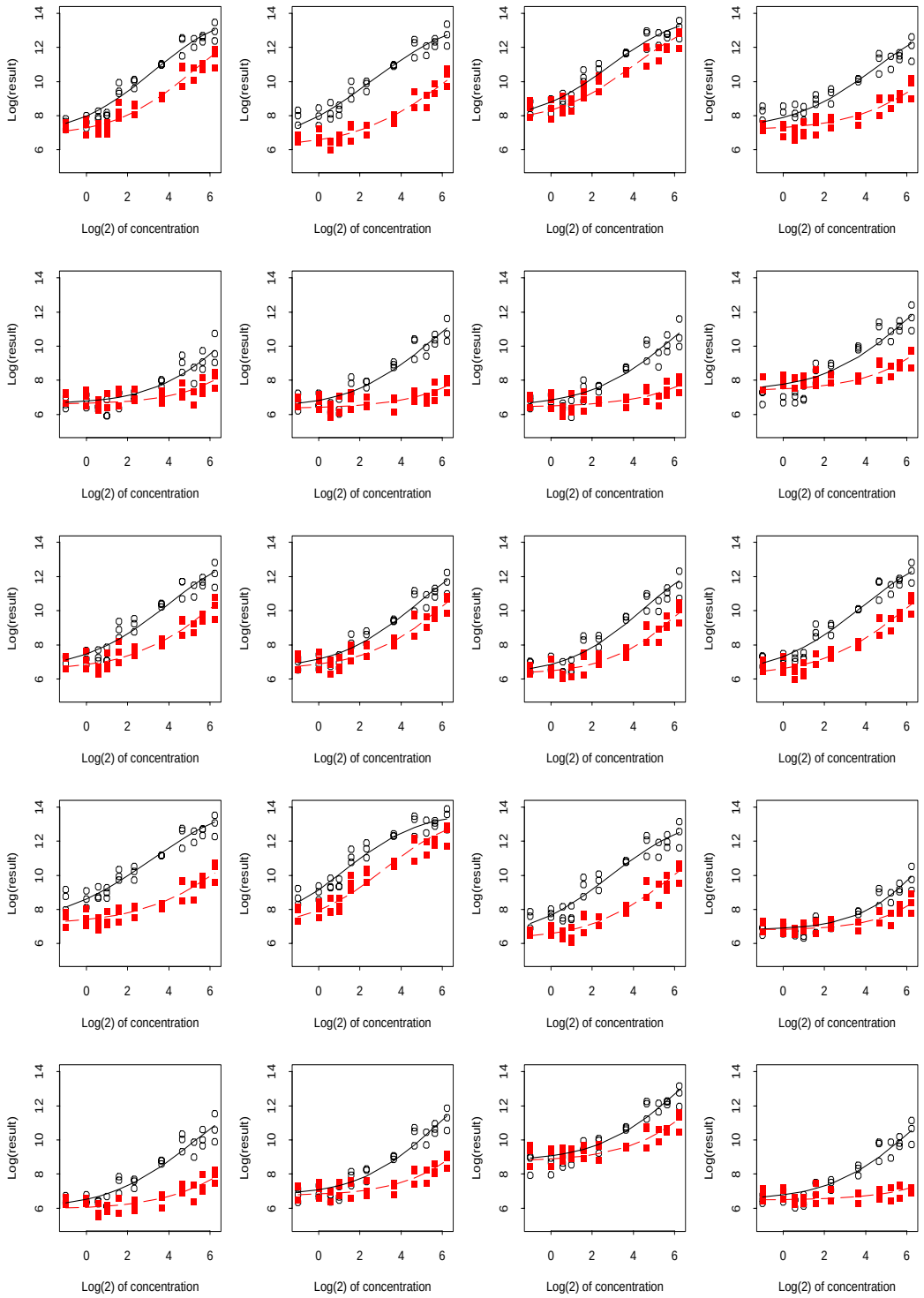
3.2 Experiment 2 plots



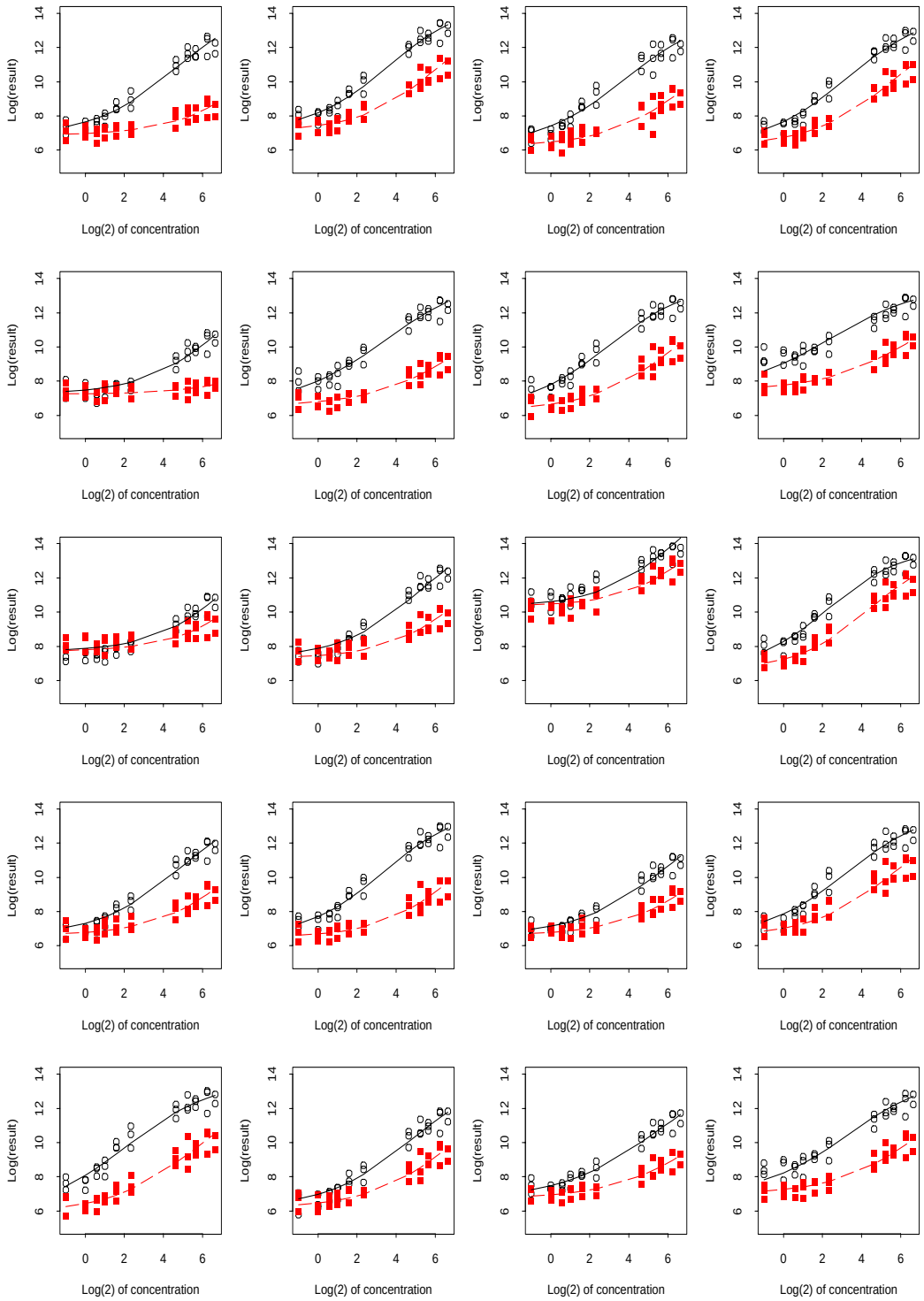
GL U95 spikein, exp 2, gene 2 (BioB−M)



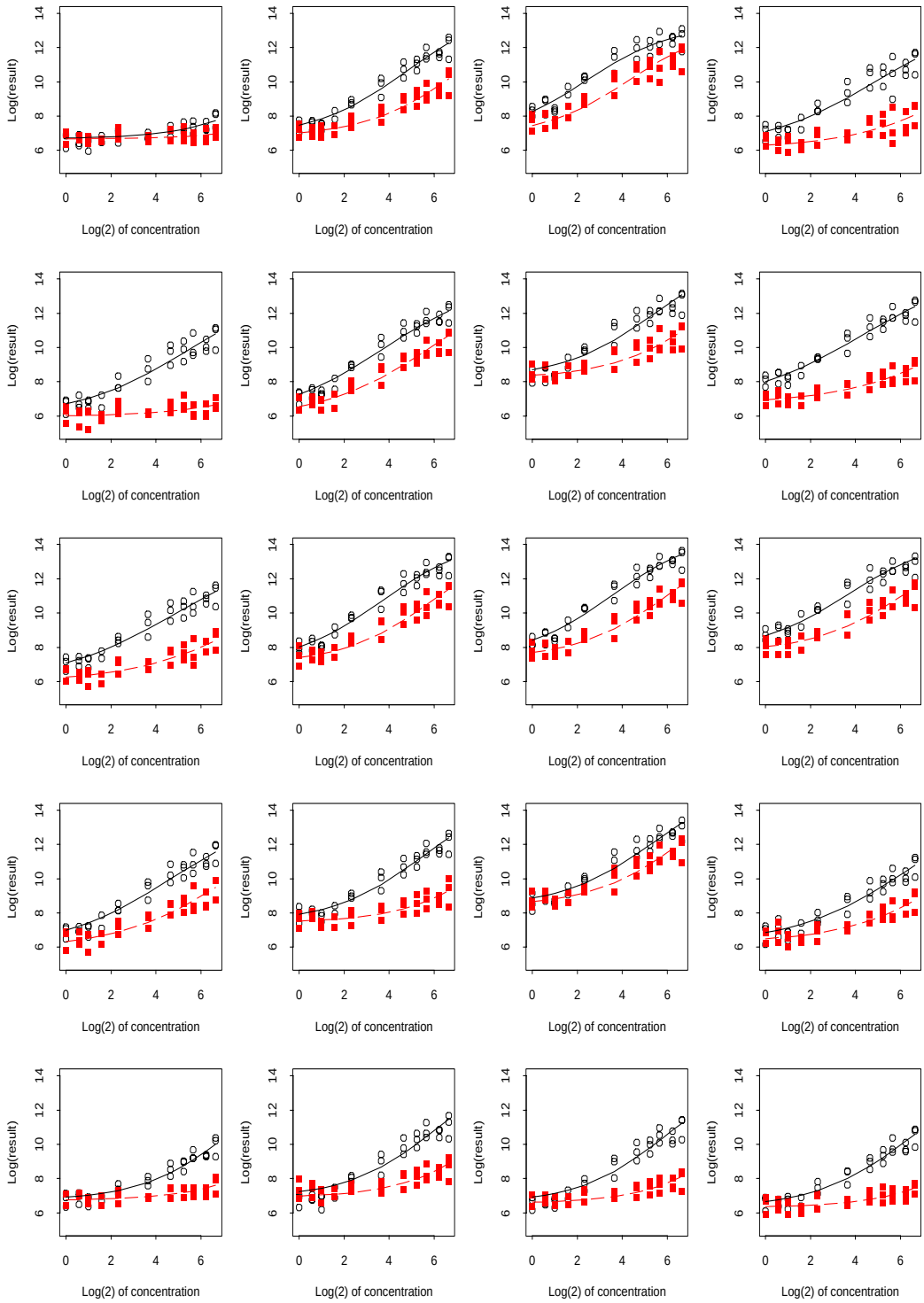
GL U95 spikein, exp 2, gene 3 (BioB−3)



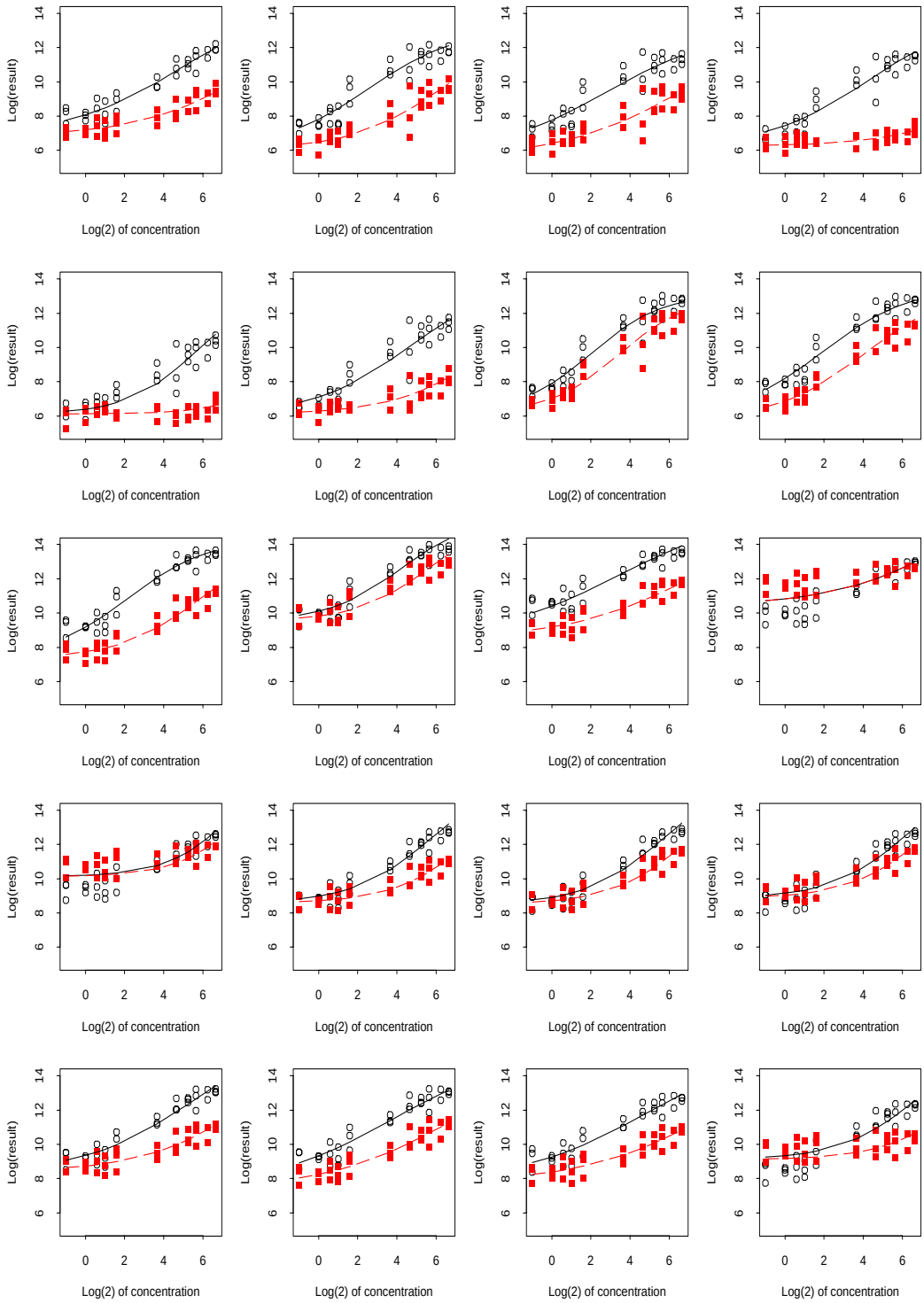
GL U95 spikein, exp 2, gene 4 (BioC−5)



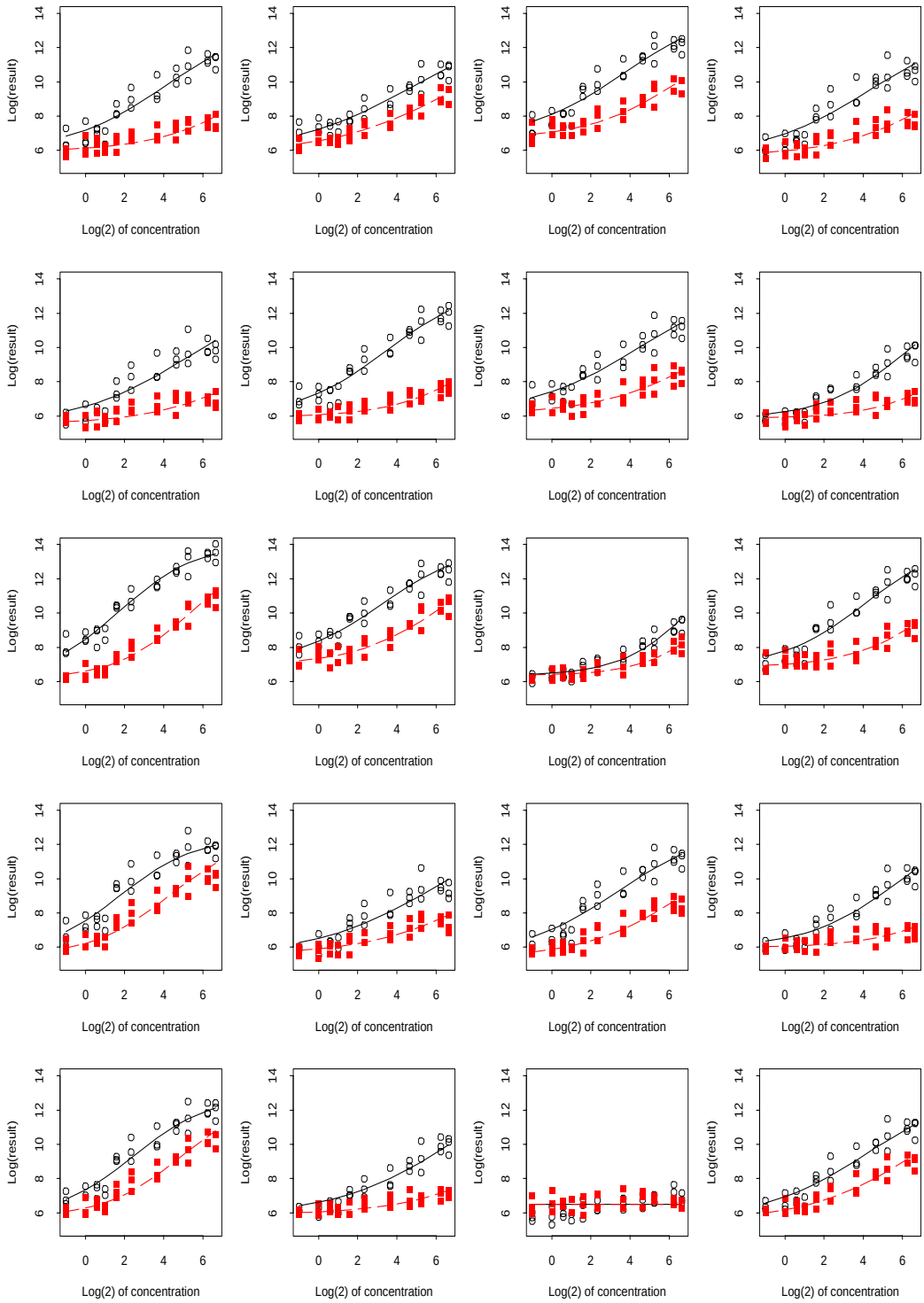
GL U95 spikein, exp 2, gene 5 (BioC−3)



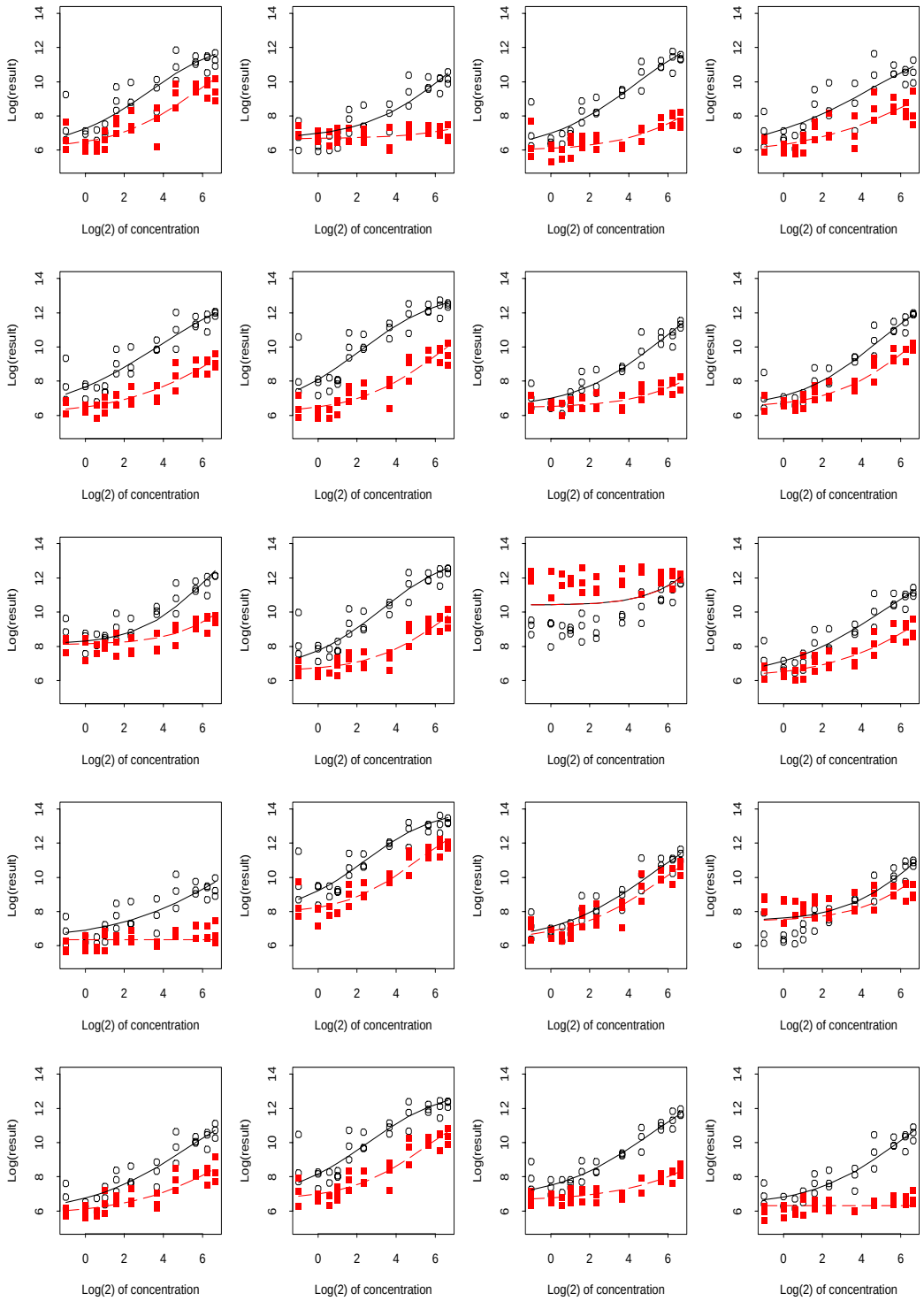
GL U95 spikein, exp 2, gene 6 (BioDn−3)



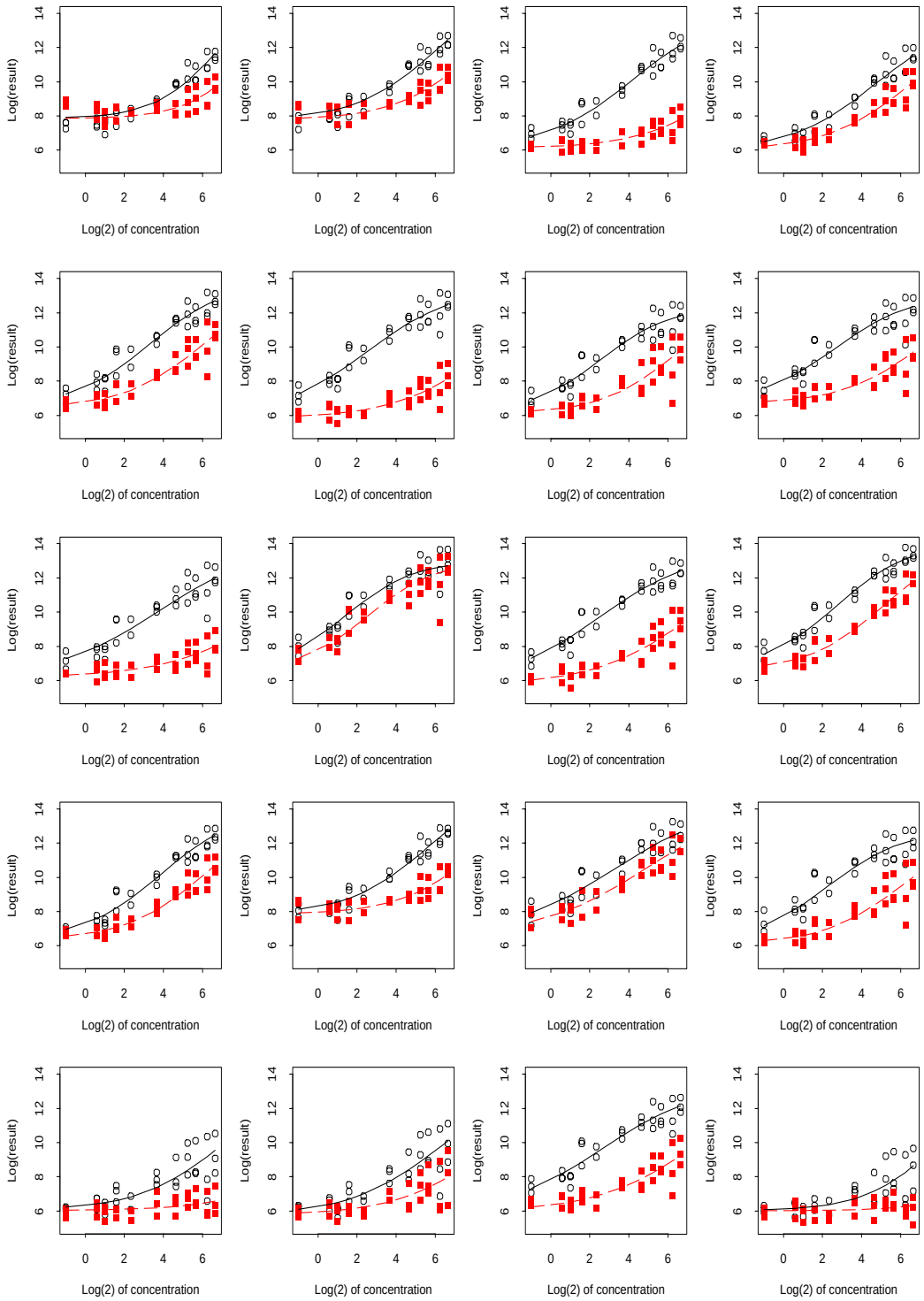
GL U95 spikein, exp 2, gene 7 (DapX−5)



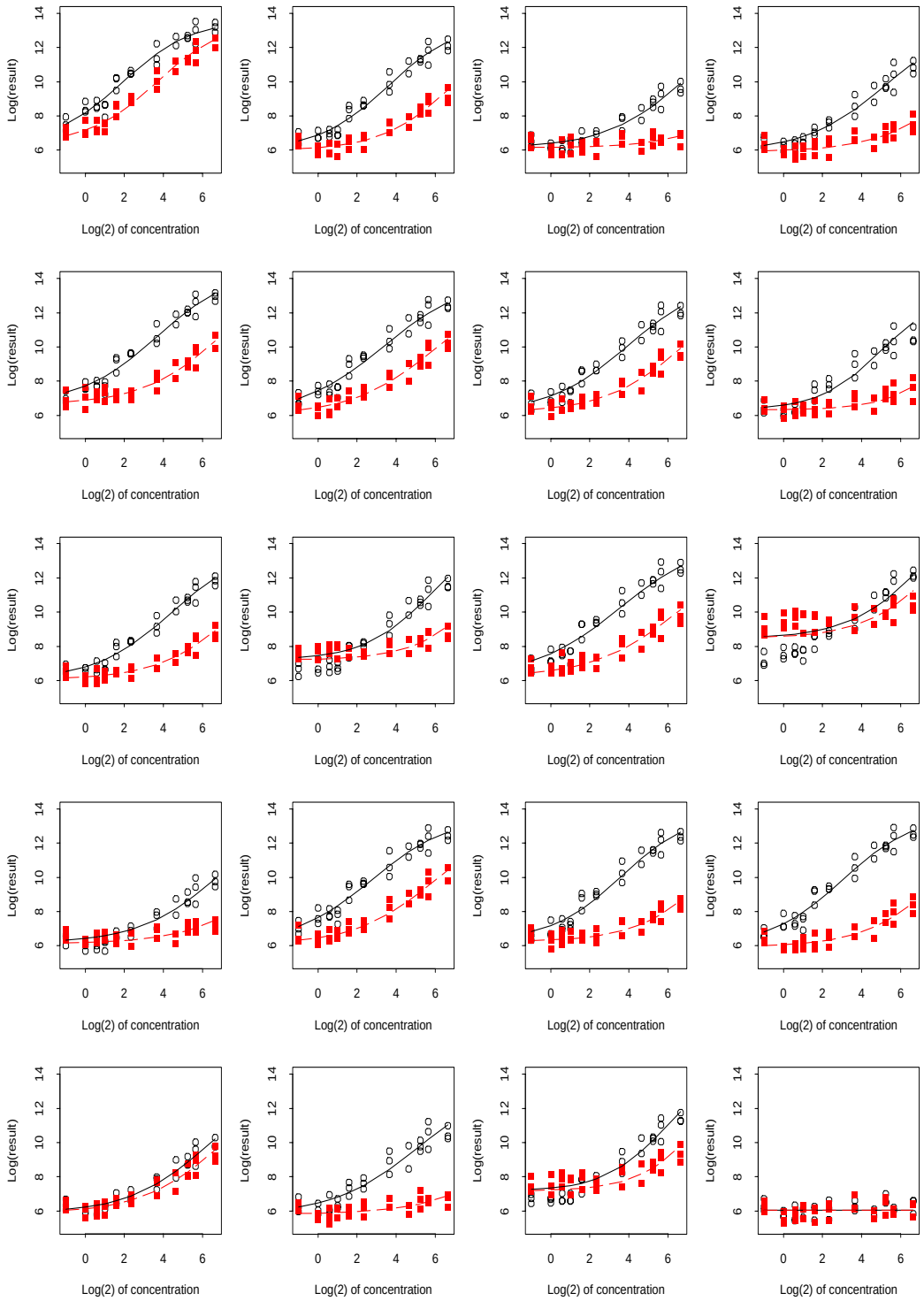
GL U95 spikein, exp 2, gene 8 (DapX−M)



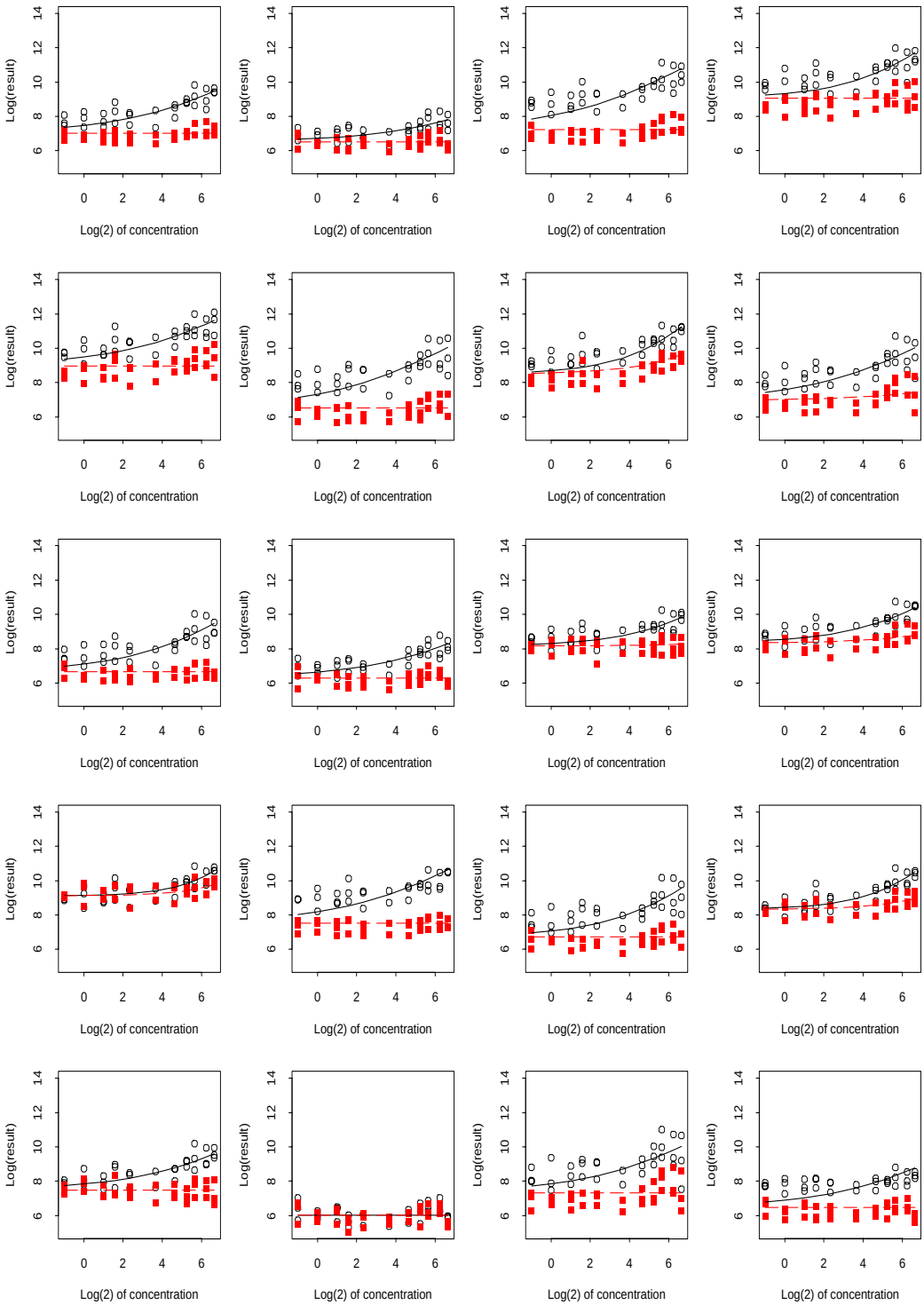
GL U95 spikein, exp 2, gene 9 (DapX−3)



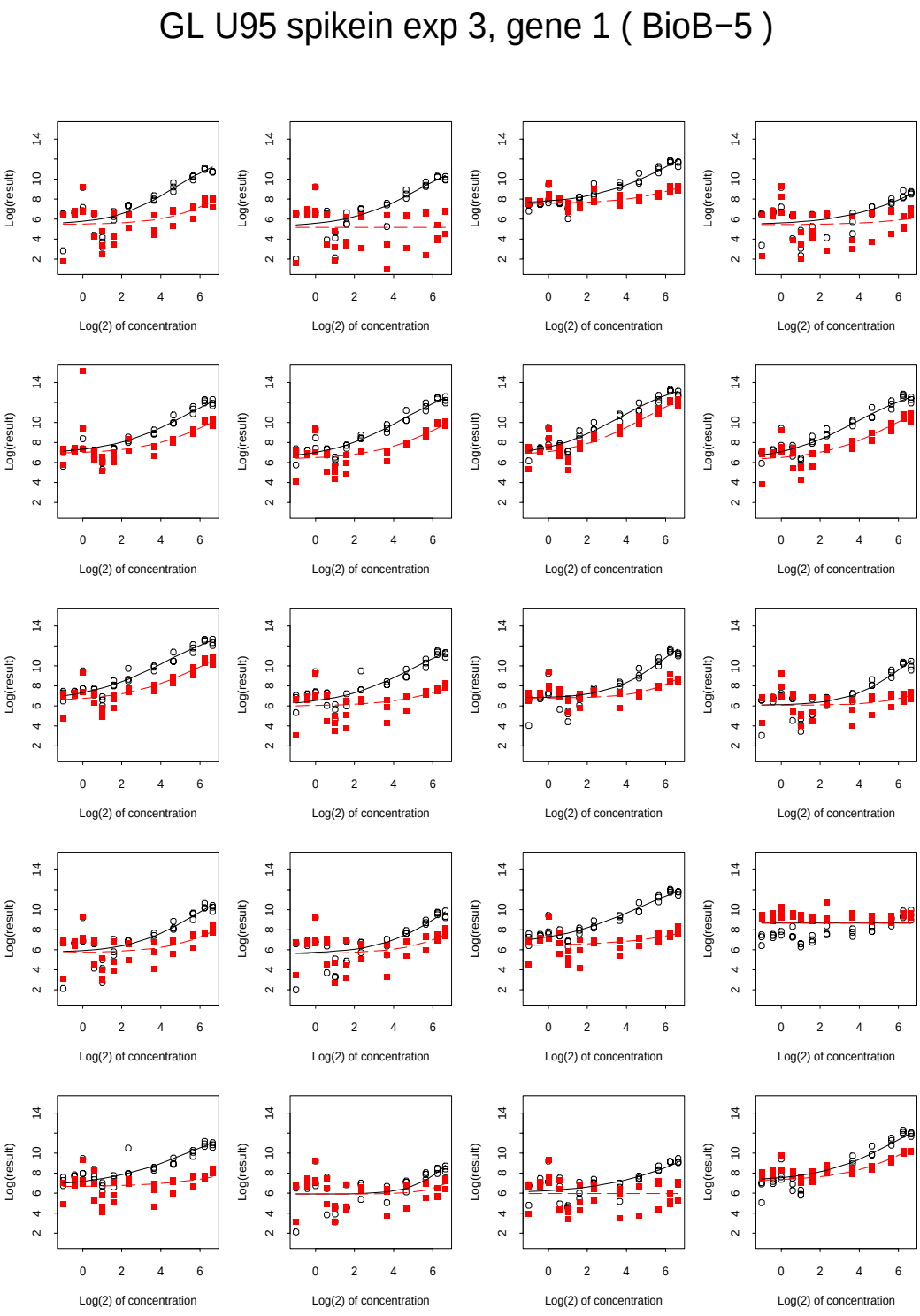
GL U95 spikein, exp 2, gene 10 (CreX−5)



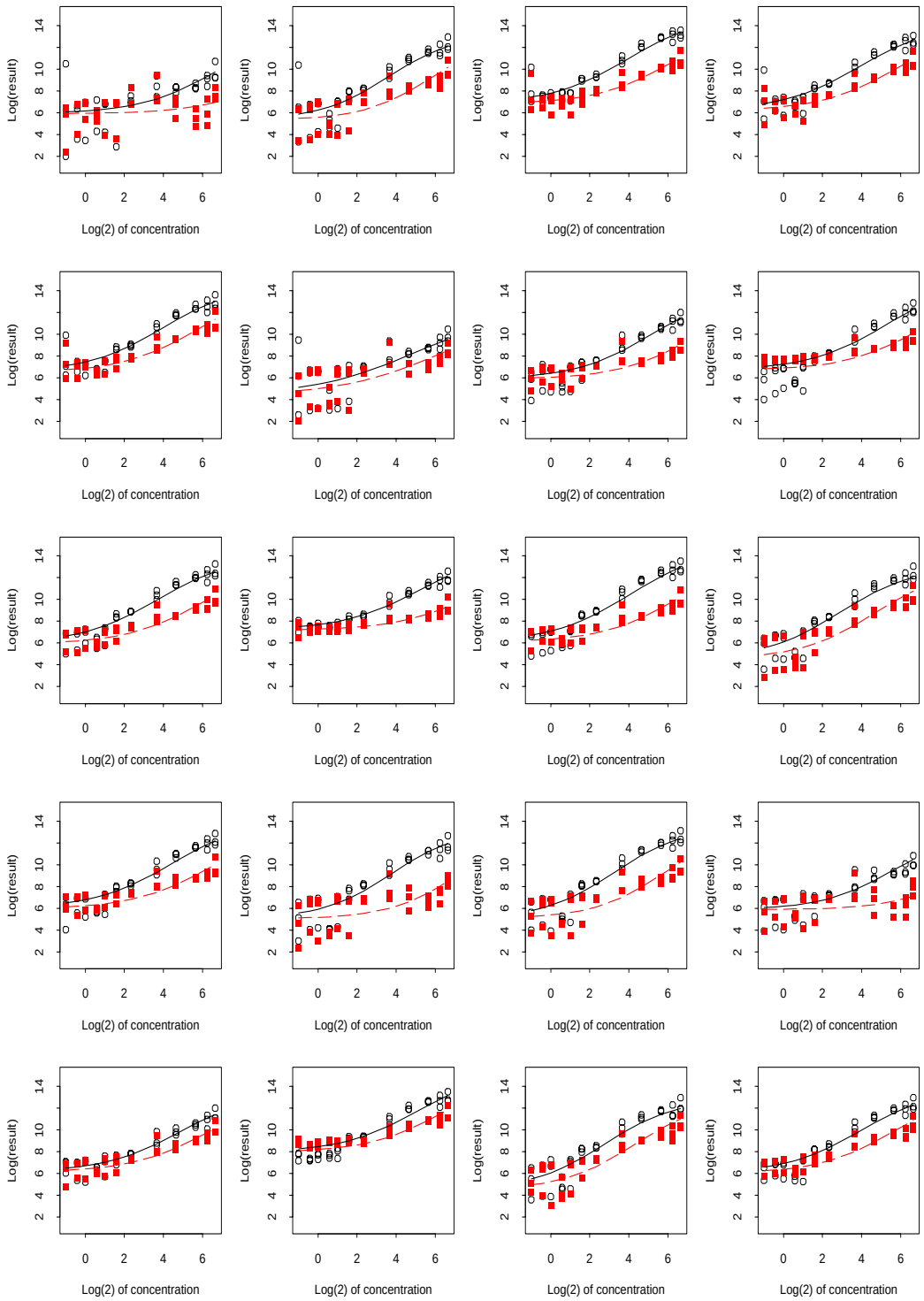
GL U95 spikein, exp 2, gene 11 (CreX−3)



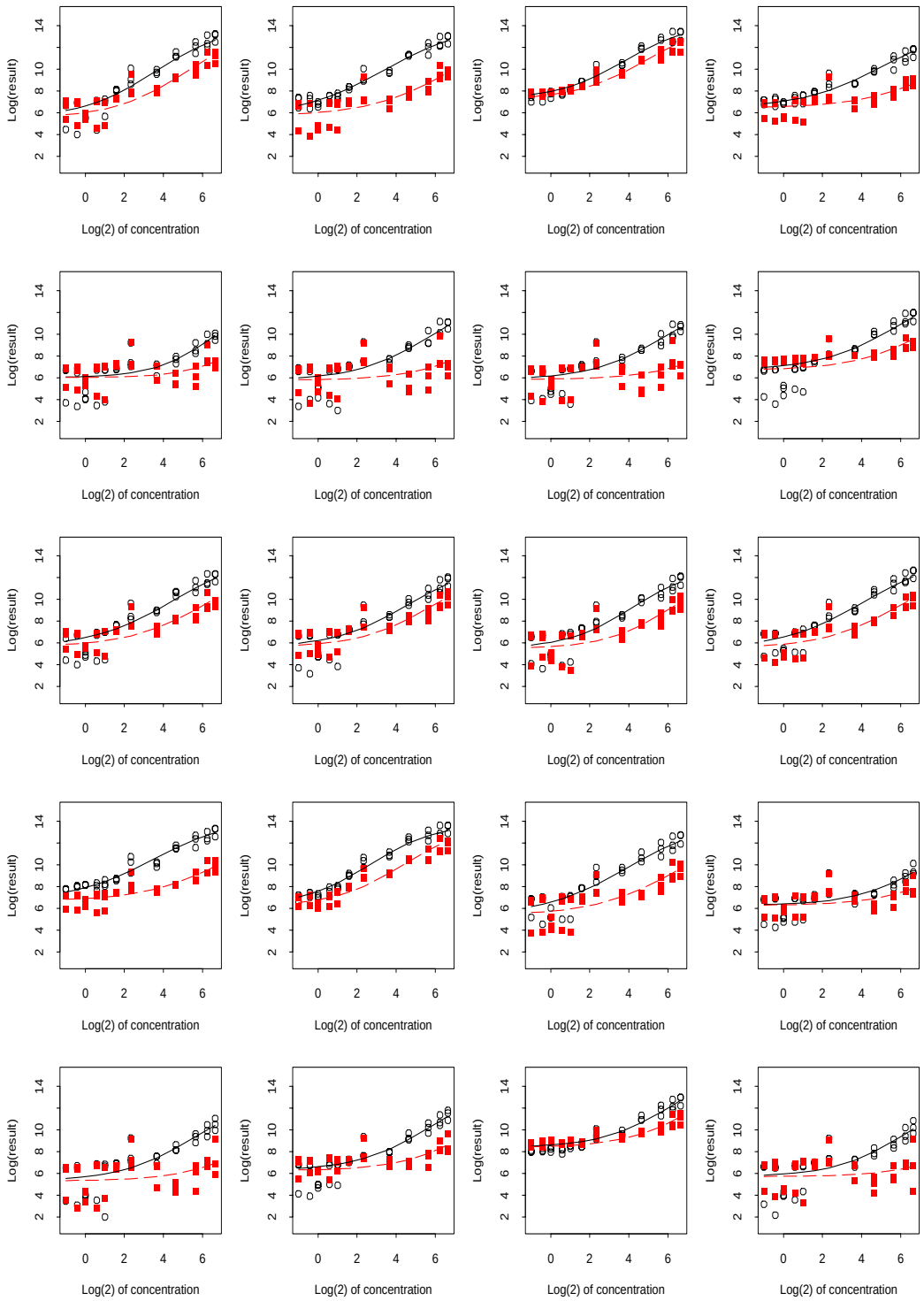
3.3 Experiment 3 plots



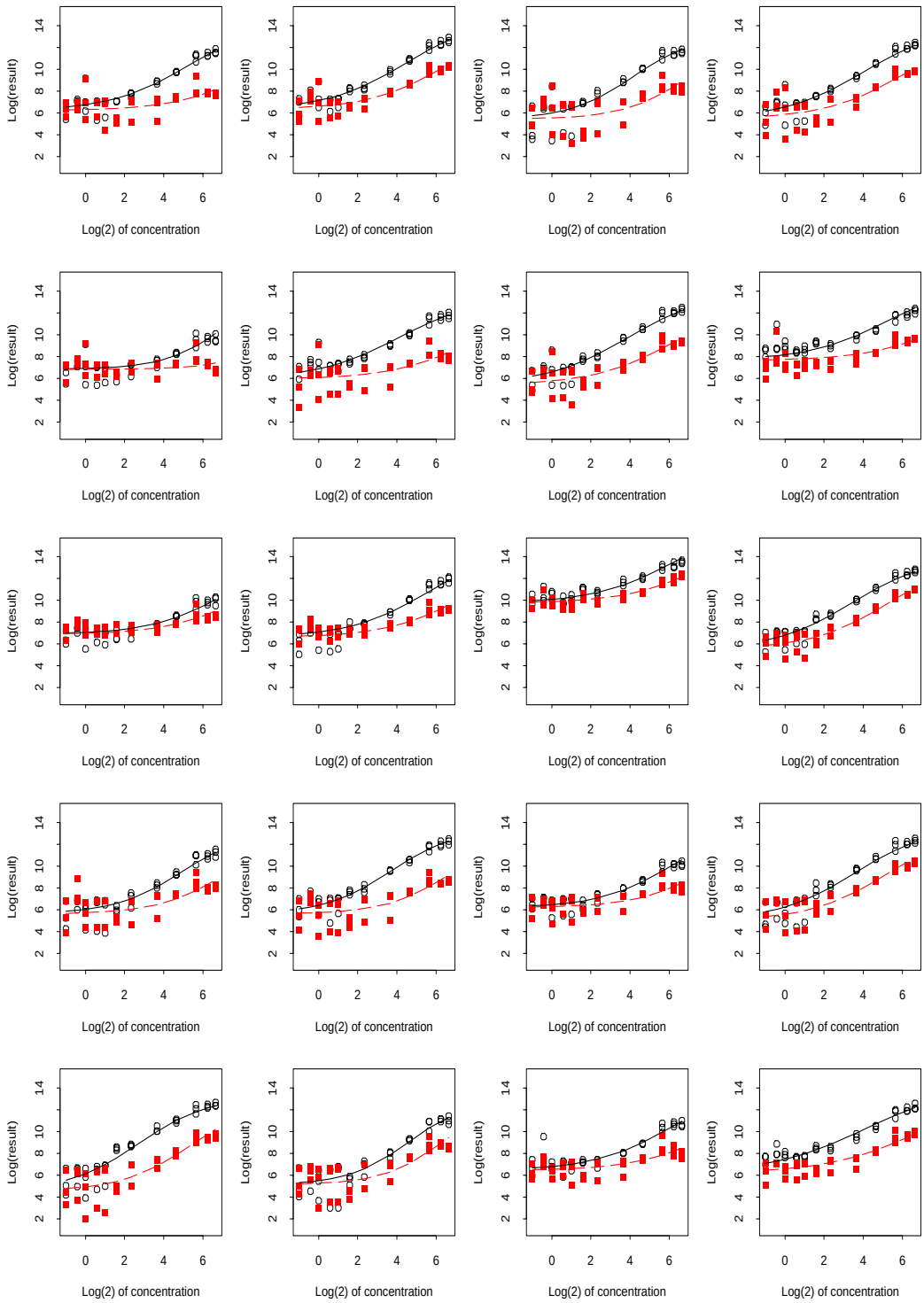
GL U95 spikein exp 3, gene 2 (BioB−M)



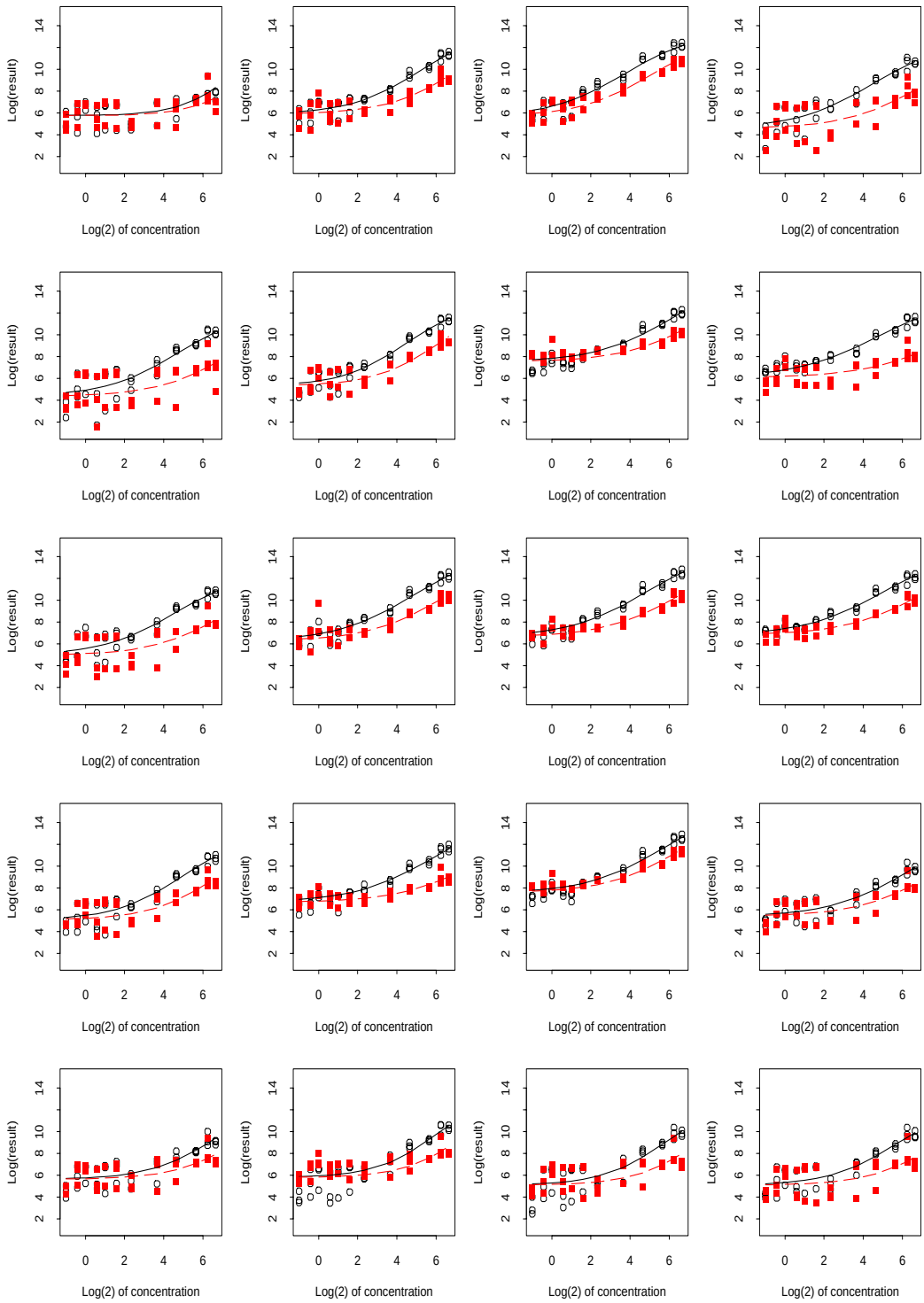
GL U95 spikein exp 3, gene 3 (BioB−3)



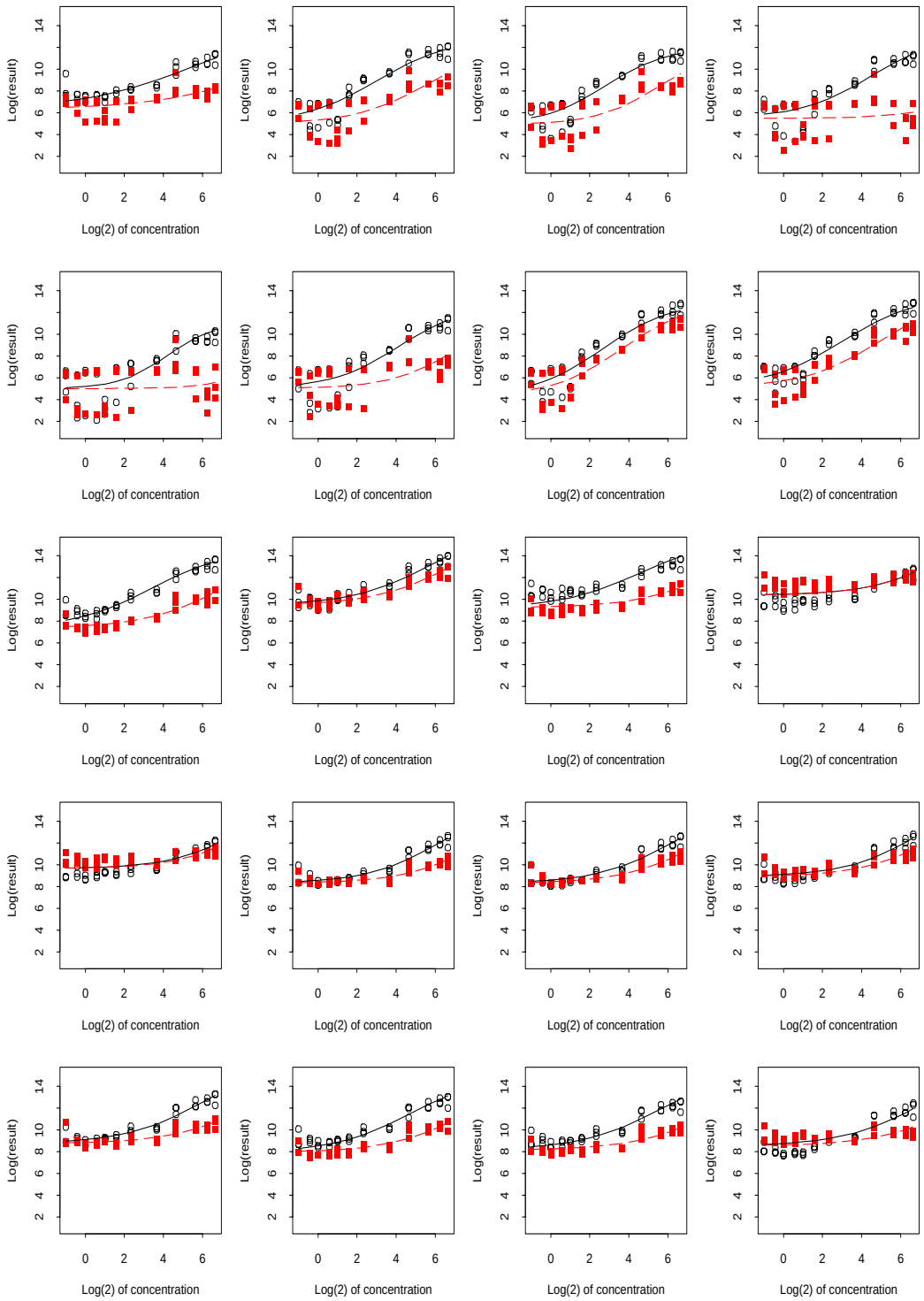
GL U95 spikein exp 3, gene 4 (BioC−5)



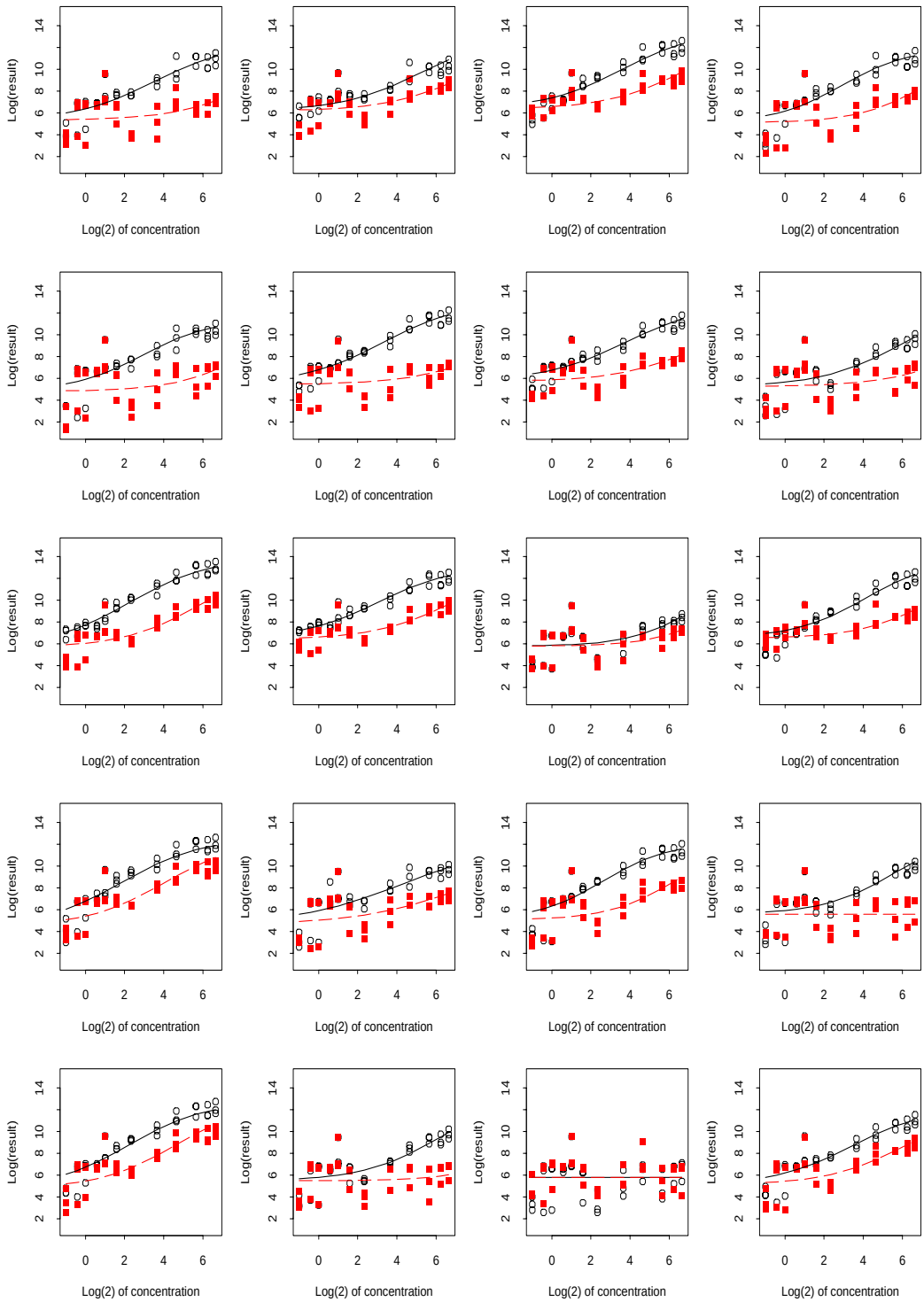
GL U95 spikein exp 3, gene 5 (BioC−3)



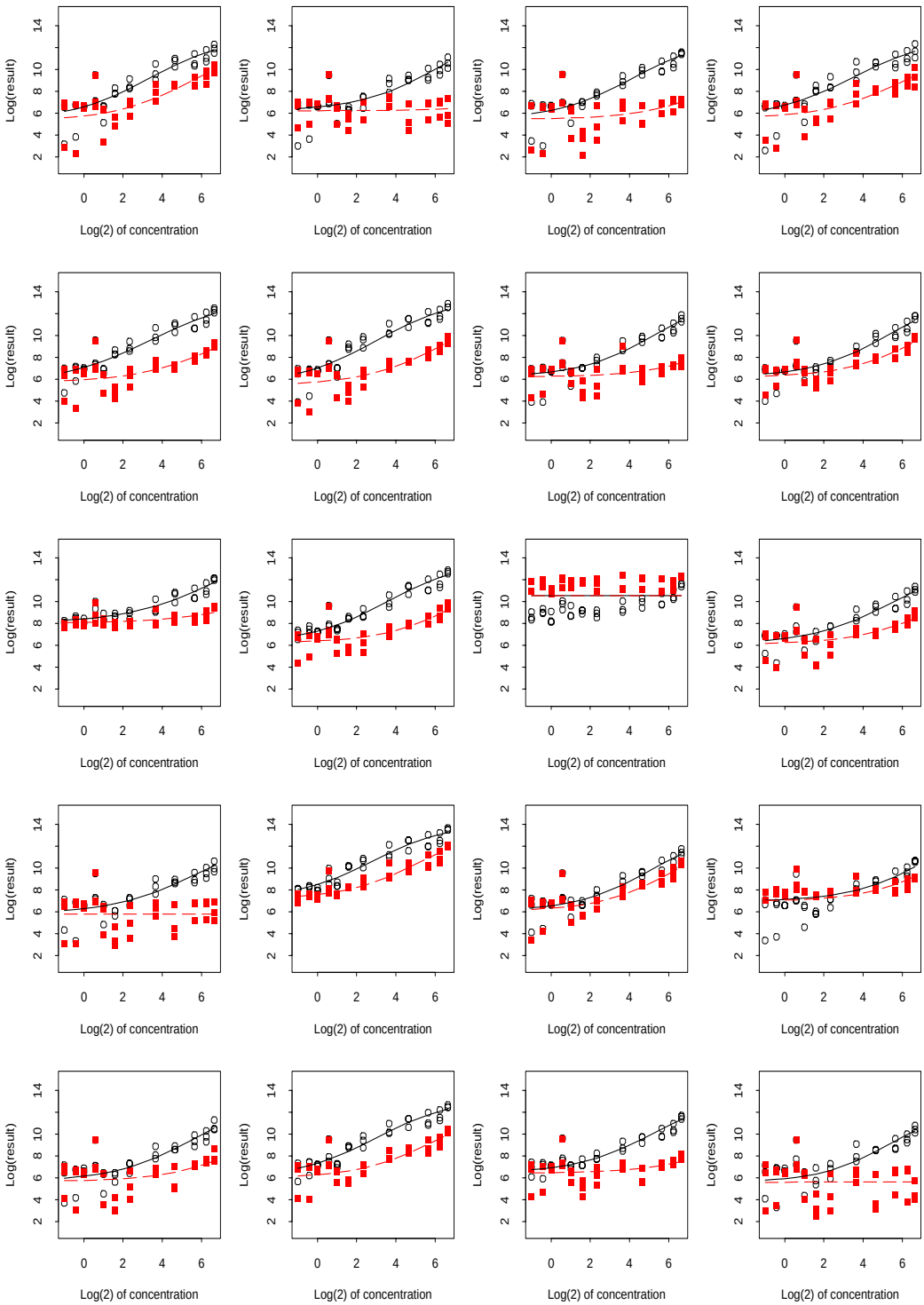
GL U95 spikein exp 3, gene 6 (BioDn−3)



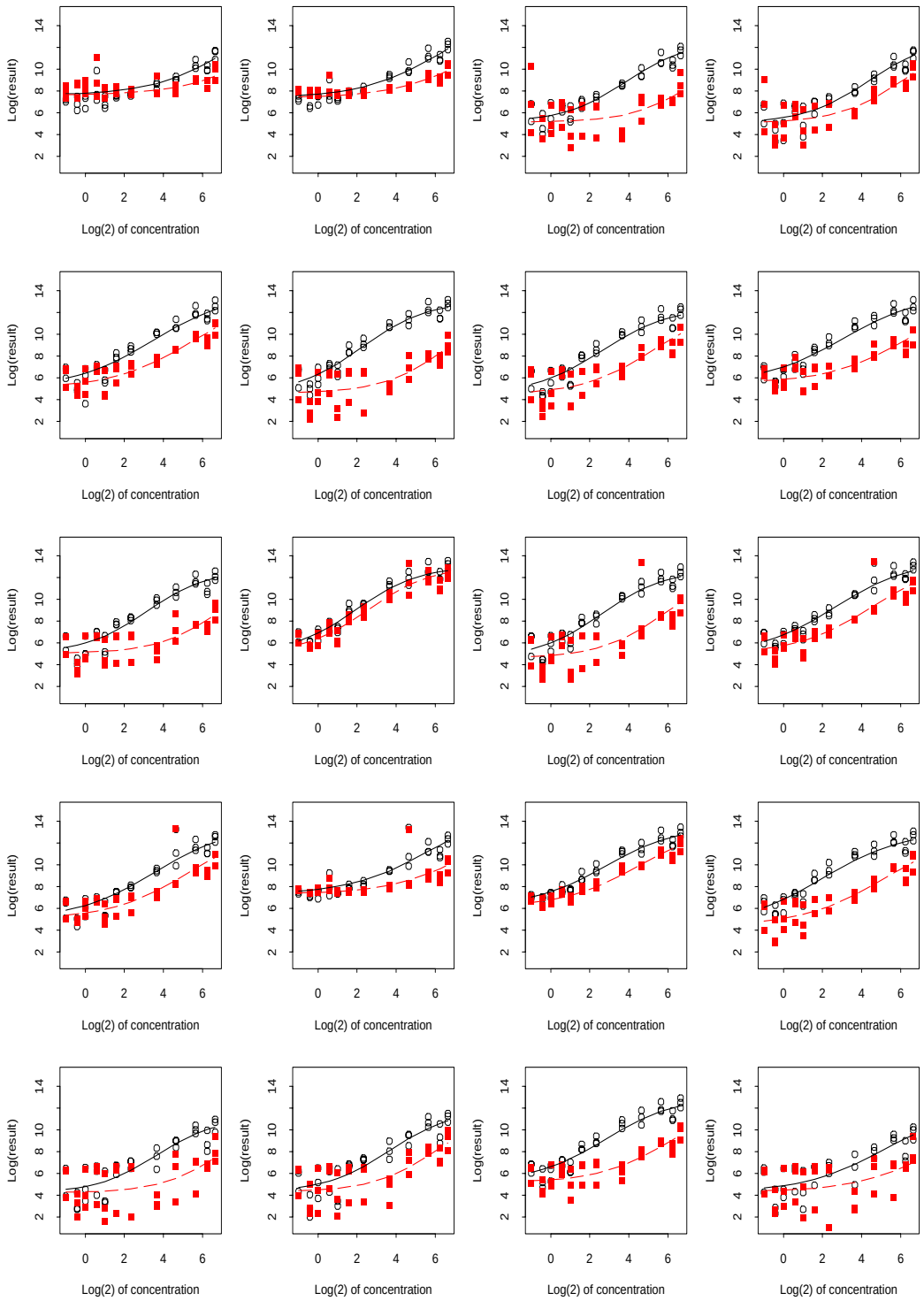
GL U95 spikein exp 3, gene 7 (DapX−5)



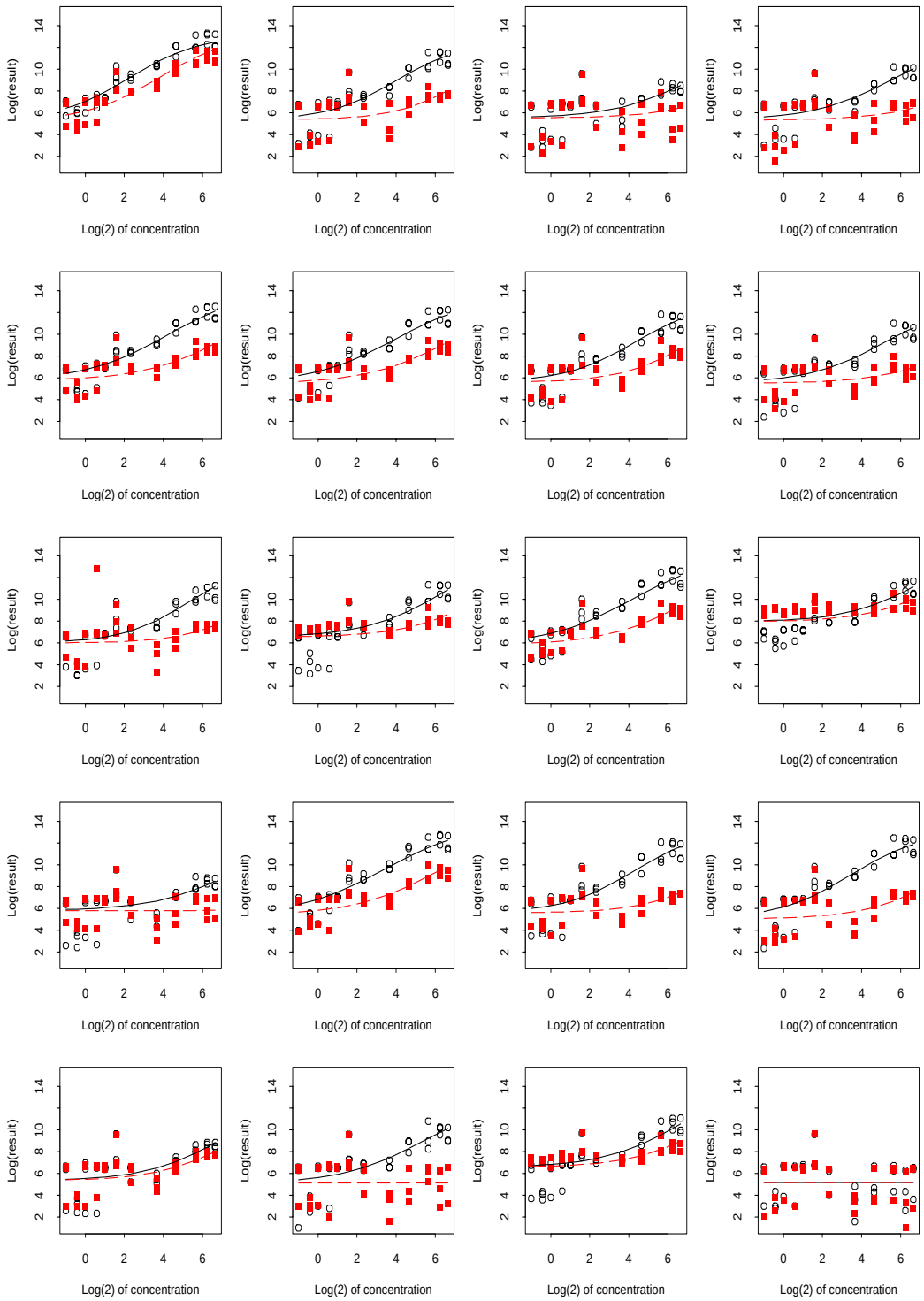
GL U95 spikein exp 3, gene 8 (DapX−M)



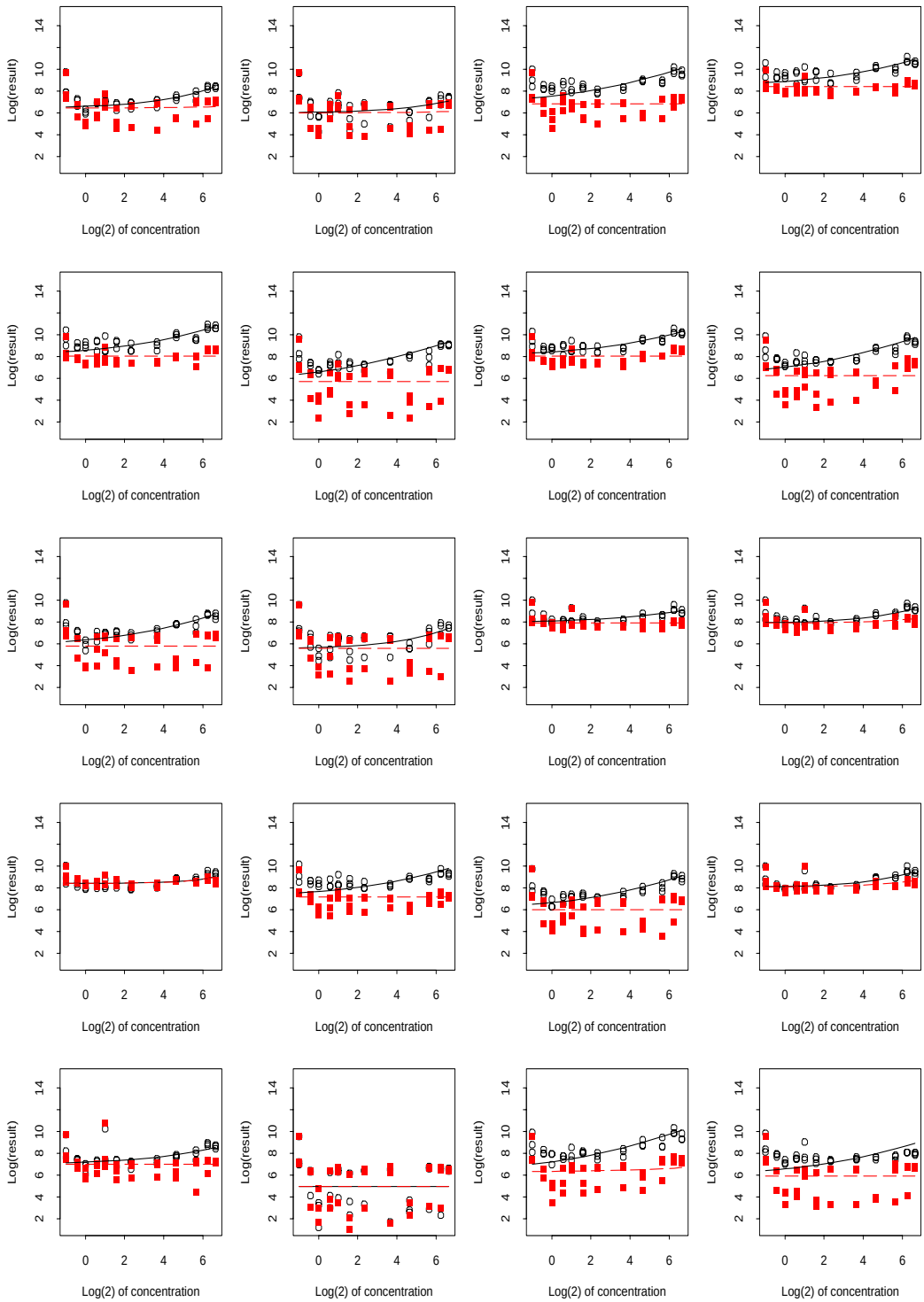
GL U95 spikein exp 3, gene 9 (DapX−3)



GL U95 spikein exp 3, gene 10 (CreX−5)



GL U95 spikein exp 3, gene 11 (CreX−3)



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.