BMC Bioinformatics

Correction

Open Access

Correction: Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment

Tomokazu Konishi*

Address: Faculty of Bioresource Sciences, Akita Prefectural University, Akita 010-0195, Japan

Email: Tomokazu Konishi* - konishi@akita-pu.ac.jp

* Corresponding author

Published: 28 June 2004

BMC Bioinformatics 2004, **5**:82 doi:10.1186/1471-2105-5-82

Received: 28 June 2004

Accepted: 28 June 2004

This article is available from: http://www.biomedcentral.com/1471-2105/5/82

Correction to formulae in methods section [1]

The lognormal distribution model and estimation of the parameters

The method assumes that the original intensity data, (r_i) for i = 1,2...n, obey a lognormal distribution. The probability density function of the intensity data used was:

$f(r_i) = [k/\{(2\pi)^{1/2}\,\sigma(r_i - \gamma)\}]\,\exp[-\{\log(r_i - \gamma) - \mu\}^2/2\sigma^2]$ for $r_i > \gamma$,

where k is a compensation constant (k = log*e* = 0.4343), σ and μ are the shape and scale parameters for log(r_i - γ), respectively.

The threshold parameter, γ, was found through trial and improvement calculation processes; in the trial, the distribution of log(r_i - γ) was checked by normal probability plotting, and the value that gave the best fit to the model was selected for γ. The fitness was evaluated by the sum of absolute differences between the model and log(r_i - γ), within the interquartile range of data. The parameter μ was found as the median of log(r_i - γ), and the parameter σ was found from the interquartile range of log(r_i - γ); these are known as robust alternatives for the arithmetic mean and standard deviation, respectively. Parameters μ and σ were found for each data grid, a group of data for DNA spots that were printed by an identical pin in order to avoid divergences caused by pin-based differences. *Z*-normalization was carried out for each datum as

$Z_{ri} = \{\log(r_i - \gamma) - \mu\}/\sigma$.

Intensity data (r_i) less than γ were treated as "data not detected", since such data might contain negative noise larger than the signal (see Results).

References

1. Tomokazu Konishi : **Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment.** *BMC Bioinformatics* 2004, **5**:5.