Genomics & Informatics Vol. 5(3) 113-117, September 2007

Fluorescence Quenching Causes Systematic Dye Bias in Microarray Experiments Using Cyanine Dye

Hosang Jeon and Sangdun Choi*

Department of Molecular Science and Technology, Ajou University, Suwon 443-749, Korea

Abstract

The development of microarray technology has facilitated the understanding of gene expression profiles. Despite its convenience, the cause of dye-bias that confounds data interpretation in dual-color DNA microarray experiments is not well known. In order to economize time and money, it is necessary to identify the cause of dye bias, since designing dye-swaps to reduce the dye-specific bias tends to be very expensive. Hence, we sought to determine the reliable cause of systematic dye bias after treating murine macrophage RAW 264.7 cells with 2-keto-3-deoxyoctonate (KDO), interferon-beta (IFN-β), and 8-bromoadenosine (8-BR). To find the cause of systematic dye bias from the point of view of fluorescence quenching, we examined the correlation between systematic dye bias and the proportion of each nucleotide in mRNA and oligonucleotide probe sequence. Cy3-dye bias was highly correlated with the proportion of adenines. Our results support the fact that systematic dye bias is affected by fluorescence quenching of each feature. In addition, we also found that the strength of fluorescence quenching is based on not only dye-dye interactions but also dye-nucleotide interactions as well.

Keywords: DNA microarray, dye bias, fluorescence quenching, cyanine dye, gene expression

Introduction

DNA microarray provides a wealth of information on various condition-specific transcriptional regulation systems. Microarray also yields vast amounts of data at once, and one can interpret the microarray data based on the intensity of dye fluorescence. Therefore, while performing data interpretation, the effect of dyes is considered one of the factors determining accuracy. Dye swap experiments have been done twice in hybridizations in which each dye was used to alternately label the samples in order to normalize the dye-related bias (Tseng *et al.*, 2001). In spite of these preventive measures of dye bias, the Cy3/green channel may appear consistently brighter than the Cy5/red channel for a particular gene, thereby causing dye bias (Dobbin *et al.*, 2005).

Previous reports have suggested various methods to rectify the problems associated with microarray (Dombkowski *et al.*, 2004; Rosenzweig *et al.*, 2004; Han *et al.*, 2005), which confounds 20% of data interpretation (Martin-Magniette *et al.*, 2005). Despite these efforts, the cause of dye bias remains ambiguous. General categories for various dye biases have been set forth by Dobbin *et al.* (2005), which are as follows: (1) dye bias that is similar for all genes on an array, causing one channel to appear brighter overall than the other; (2) dye bias that depends on the overall spot intensity, and is different for bright spots than for dim spots; (3) dye bias that is associated with some subset of genes, but is consistent for the same gene across samples; (4) dye bias that depends on a combination of characteristics of the sample as well as the gene. In our investigation, we further subdivided the dye bias into two major categories: gene-specific dye bias and systematic dye bias.

To remove systematic dye biases from microarray studies, there are currently various normalization methods available (Yang *et al.*, 2002). However, due to limited information on the types of errors, scientists are still far from eliminating dye biases in microarray studies (Uchida *et al.*, 2005). A microarray researcher may think that based on the previous information available, there may be various complicated reasons for dye bias (Martin-Magniette *et al.*, 2005). For example, first of all, the intensity of the signal is related to the amount of transcript synthesis as well as the amount of transcript degradation. Second, amplified RNA degradation is affected not only by the amplification time, but also by the types of bound dyes (Spiess *et al.*, 2003). Apart from these, intensity of the signal is also affected by the strength of bonding between an antisense RNA and a probe on the chip. In addition, it is also known that the nature of the dye can affect melting temperature of the nucleic acid duplex (Randolph *et al.*, 1997). More interestingly, the microarray intensity signal is highly correlated with the fluorescence quenching of dye-dye interactions as a result of dye stacking or aggregate formation during the hybridization of dye-labeled samples

*Corresponding author: E-mail sangdunchoi@ajou.ac.kr
Tel +82-31-219-2600, Fax +82-31-219-1615
Accepted 3 August 2007

(Herz, 1974), which was verified by nuclease treatment experiments (Cox *et al.*, 2004). The fluorescence of commonly-used dyes conjugated to oligodeoxyribonucleotides can also be quenched by the interactions between the dye and nucleobase (Cooper and Hagerman, 1990).

In our studies, we focused on the fluorescence quenching mediated by dye-dye interactions to determine a reliable cause of systematic dye bias using oligonucleotide microarray data of murine macrophage RAW 264.7 cells treated with 2-keto-3-deoxyoctonate (KDO), interferon-beta (IFN-β), and 8-bromoadenosine (8-BR). We further correlated the degree of labeling with the proportion of adenines in the mRNA sequence. Since the labeled antisense RNAs were synthesized by the coupling reaction of mono-Cy3/mono-Cy5 molecules with aminoallyl- UTPs (aaUTPs) (http://www.signaling-gateway.org/data/cgi-bin/ProtocolFile.cgi; AfCS protocol PP00000184), which are complementary to adenines in mRNA sequences, we also examined the role of the adenine proportion.

Materials and Methods

Cell culture, treatment, and RNA preparation

The RAW 264.7 cell line is an immortalized macrophage clone isolated from BALB/c mice transformed with Abelson leukemia virus. These cells were maintained in RAW 264.7 growth medium (RAWGM; DMEM 0.87X, FBS 10%, HEPES 20 mM, L-Glutamine 2 mM). KDO (100 ng/ml), IFN-β (300 pM), and 8-BR (100 μM) were applied individually and in combination with KDO to stimulate RAW 264.7 cells for 1 h and 2 h, and then their RNA was extracted using TriPure (Roche) according to its protocol. Duplicate experiments were done for each treatment except for KDO, wherein individual treatments were performed in quadruplicate.

Oligoarray fabrication and annotation

Gene expression was analyzed by custom-made 16K mouse oligo arrays with 15,631 65- or 70-bp oligomers printed on them. The oligomers were obtained from Operon and Sigma-Genosys and were inkjet-printed onto glass slides by Agilent Technologies. The gene contents are available through the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) as accession number GPL254.

Aminoallyl-UTP (aaUTP) antisense RNA (aRNA) synthesis and microarray hybridization

Each array was hybridized with the Cy5-labeled antisense RNA prepared from the RNA of ligand-treated cells and the Cy3-labeled antisense RNA prepared from the RNA of time-matched control cells using the aminoallyl method (http://www.signaling-gateway.org/data/cgi-bin/ProtocolFile.cgi; AfCS protocol PP00000184). Dye-swap labeling and array hybridization were performed for each pair of antisense RNA samples.

Microarray fluorescent imaging

The arrays were scanned using the Agilent Scanner G2505A (Agilent Technologies) with maximum red and green (Cy5 and Cy3) fluorescence intensities around 65,000.

Data selection and analysis

For each set of microarray data, we filtered out features based on four criteria: (1) control features, (2) any features with a green or red fluorescence signal that was saturated ("gIsSaturated" or "rIsSaturated" flags), (3) non-uniform features ("gIsFeatNonUnifOL" or "rIsFeatNonUnifOL" flags), and (4) features that had fluorescence signals below background. The expression changes of filtered features are represented as $\log_2$-(Cy5/Cy3), also called log-FC (log-Fold Change).

Selection of gene-specific (non-systematic) dye bias

We considered the features as gene-specific dye bias in which the value of log-FC was higher than 0.378 or lower than -0.378 (1.3-fold change) at the same direction in both the first comparison and the dye-swapped one in 56 independent experiments. as gene-specific dye bias. Gene-specific dye bias was visualized by the Multiple Experiment Viewer (MeV:http:// www.tigr.org/software/tm4/mev.html).

Systematic dye bias

Besides the gene-specific dye bias, there also exist different types of dye bias in which the signal intensities of regular and the dye-swapped results are unreliable. In other words, if the log-FC value of one feature is 2 in a regular experiment, then the log-FC value of dye must be -2 in the dye swap experiment, but some features showed very different values. We have chosen those features that showed p-values lower than 0.001 by two-tailed student's t-test of log-FC on KDO (1 h)-treated experiments. "Degree of Cy3-dye bias" indicates the sum of average log-FC values of regular experiments and those of dye-swapped ones.

Correlation between dye bias and fluorescence quenching

The correlation between systematic dye bias and the proportion of each nucleotide in the mRNA sequence was calculated by Pearson and Spearman correlation coefficients. The degree of Cy3-dye bias was reformed to the rank of

Cy3-dye bias to be used for Spearman correlation.

Results

Gene-specific dye bias

We selected the features showing net fold changes of 1.3 to assess dye bias in more than 70% of the 56 independent microarray experiments. Based on these criteria, 65 features were chosen and we could see intuitively that these gene-specific dye bias features accumulated in four regions on the chip. Fig. 1 shows the location of those features on the microarray as visualized by MeV, where the systematic dye bias spots, always showing green in the scanned image, are represented by their average log-FC values. In MeV visualization, the downregulated features, whose log-FC values were less than zero, are shown in green. On the other hand, red colors indicate the features that show upregulation, and their log-FC values are greater than zero.

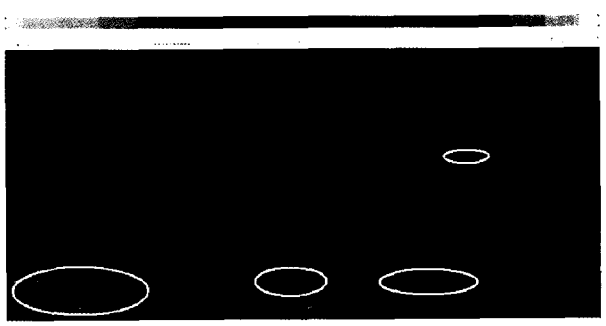

Fig. 1. Gene-specific Cy3-dye bias on microarray. The custom-made inkjet-deposited presynthesized oligo array used in this experiment has 119 rows and 156 columns. A total of 65 spots with gene-specific Cy3-dye bias were concentrated in four regions.

Systematic dye bias and its correlation with fluorescence quenching

Features selected by two-tailed student's t-test were examined with respect to their proportion of adenines in the mRNA sequence. First, we sorted out the features based on their degree of Cy3-dye bias. Since the correlation coefficients can be easily distorted by outliers, features that had more than a 5% difference in the proportion of adenines with respect to their adjacent features were discarded from the studies. This resulted in a total of 295 features that were chosen to analyze the correlation between dye bias and fluorescence quenching. The first 100 features with the highest Cy3-dye bias and the last 100 features with the lowest Cy3-dye bias showed a distinguished proportion of adenines (Table 1).

Note: Top 100, features showing the highest Cy3-dye bias; bottom 100, features that show the lowest Cy3-dye bias. P-value was calculated through two-tailed t-test with Std. Dev (Standard Deviation of adenine proportion) in each feature.

The higher the degree of Cy3-dye bias, the higher the proportion of adenines in the mRNA sequence as well as in the oligomer probe sequence. On top of that, the sum of the adenine proportions in mRNAs and oligomer probe sequences correlated more with the degree of Cy3-dye bias than with any other individual comparison (Fig. 2).

Correlations between the degree of Cy3-dye bias and other nucleotide proportions in mRNAs and oligomer

Table 1. Comparison of adenine proportion with respect to the rank of Cy3-dye bias in 1h KDO treatment experiments

	[A]%	Std. Dev.
Top 100	27.40043942	2.588782
Bottom 100	24.01309951	2.651331
p-value	7.48161E-17	

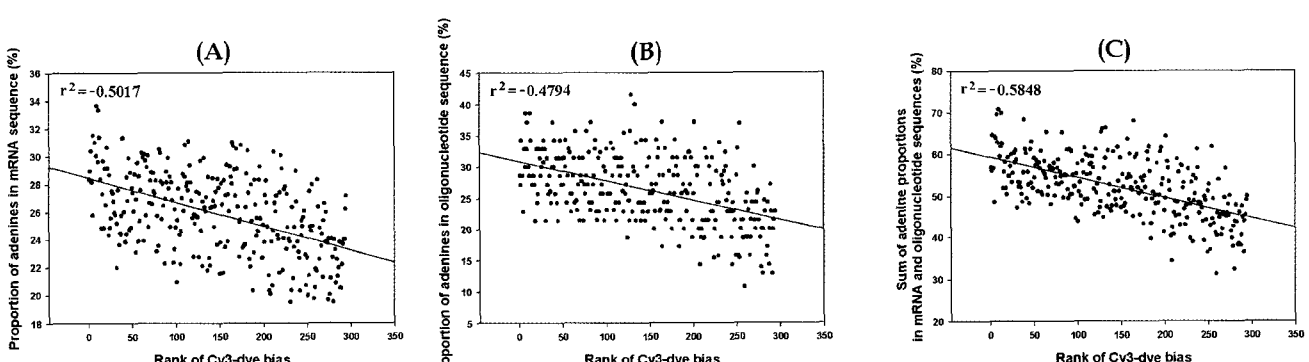

Fig. 2. Co-effect of adenine proportions in mRNAs and oligonucleotide probe sequences. Dye bias is affected not only by the proportion of adenines in mRNA sequences but also by the proportion of adenines in oligonucleotide probe sequences. (A) Correlation between adenine proportion in mRNA sequences and Cy3-dye bias, r^2 = -0.5017. (B) Correlation between adenine proportion in oligonucleotide probe sequences and Cy3-dye bias, r^2 = -0.4794. (C) Correlation between the sum of adenine proportions in mRNAs and oligonucleotide probe sequences and Cy3-dye bias, r^2 = -0.5848

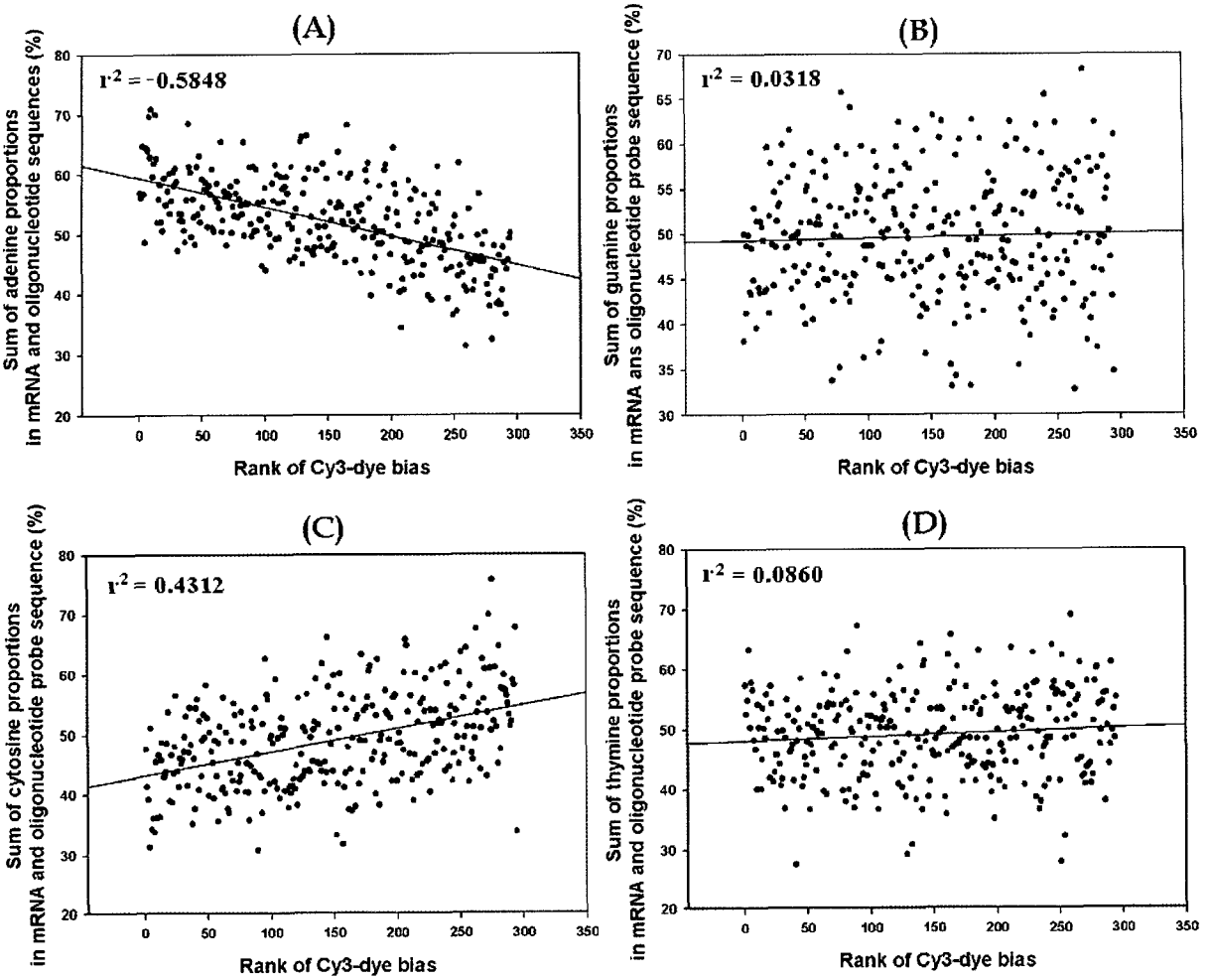

Fig. 3. Correlation between Cy3-dye bias and individual proportions of nucleotides (A, G, C, T) calculated by Spearman correlation coefficients. (A) Correlation between the Cy3-dye bias and proportion of adenines, r^2 = -0.5848. (B) Correlation between the Cy3-dye bias and proportion of guanines, r^2 = 0.0318. (C) Correlation between the Cy3-dye bias and proportion of cytosines, r^2 = 0.4312. (D) Correlation between the Cy3-dye bias and proportion of thymines, r^2 = 0.0860.

probe sequences are shown in Fig. 3. The correlation coefficient values of both guanines and thymines were less than 0.1, but the proportions of adenines and cytosines were highly correlated with the degree of Cy3-dye bias.

Discussion

Theoretically speaking, the intensity of fluorescently-labeled nucleic acid should increase linearly with the number of dye molecules bound. However, recent studies indicated that the fluorescence of Cy3/Cy5-labeled DNA has a non-linear relationship irrespective of the degree of labeling due to the effects of fluorescence quenching. According to Cox *et al.* (2004), if the degree of labeling increases, it can effectively quench the fluorescence of Cy5-labeled nucleic acid than that of Cy3-labeled nucleic acid. This phenomenon causes an increase in the fluorescence of Cy3-labeled nucleic acid than that of Cy5-labeled, resulting in Cy3-dye bias. We confirmed this phenomenon through the correlation between the degree of Cy3-dye bias and the proportion of adenines in mRNA and oligonucleotide sequences in comparison with other nucleotide proportions (Fig. 3).

If the degree of labeling in antisense RNA is the key factor for systematic dye bias, we can expect the same bias aspects in different features designated for a particular gene. We encountered various features for a given gene that did not show same aspects of dye bias. By observing the correlation between the dye bias and the proportion of adenines in oligonucleotide probe sequences, we found out the proportion of adenines in oligonucleotide sequences also affected dye bias. In other words, among mRNA

sequences, the adenine proportion located adjacent to the oligonucleotide probe sequence was found to have a major effect on the signal intensity that caused the systematic dye bias (Fig. 2).

Fig. 3 shows that the proportions of guanines and thymines have no correlation with the Cy3-dye bias. On the other hand, the proportion of cytosines correlated somewhat with systematic dye bias. It can be explained by fluorescence quenching that resulted from the interactions between dye and nucleobase (Fukui *et al.*, 1999; Heinlein *et al.*, 2003; Marras *et al.*, 2002; Walter and Burke, 1997).

Nucleotide quenching seems to affect dyes that emit green and yellow wavelengths more than the ones that emit blue and red wavelengths (Marras *et al.*, 2002). Since guanine is the most oxidizable nucleobase and exhibits good electron-donating properties, it was proposed that fluorescence quenching of dye-nucleobase interactions is dependent on the distance between a dye and its adjacent guanines, the complementary nucleobase of cytosines, in mRNA sequences (Atherton and Harriman, 1993; Seidel *et al.*, 1996; Steenken and Jovanovic, 1997).

The systematic dye bias in dual-color microarray experiments using cyanine dyes basically comes from the difference between fluorescence-quenching degrees of the two dye molecules. This difference in fluorescence quenching is caused not only by dye-dye interactions but also due to dye-nucleobase interactions. Because the excessive degree of labeling causes bigger differences in fluorescence quenching of two different dye molecules, fitting the degree of labeling to an appropriate level can reduce systematic dye bias.

Acknowledgments

We thank Dr. Jayalakshmi Krishnan for critical reading of the manuscript. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2006-311-C00482), and a 2006 Ajou University Internal Research Grant to S.C.

References

Atherton, S.J. and Harriman, A. (1993). Photochemistry of intercalated methylene blue: Photoinduced hydrogen atom abstraction from guanine and adenine. *J. Am. Chem. Soc.* 115, 1816-1822.

Cooper, J.P. and Hagerman, P.J. (1990). Analysis of fluorescence energy transfer in duplex and branched DNA molecules. *Biochem.* 29, 9261-9268.

Cox, W.G., Beaudet, M.P., Agnew, J.Y., and Ruth, J.L. (2004). Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. *Anal. Biochem.* 331, 243-254.

Dobbin, K.K., Kawasaki, E.S., Petersen, D.W., and Simon, R.M. (2005). Characterizing dye bias in microarray experiments. *Bioinformatics* 21, 2430-2437.

Dombkowski, A.A., Thibodeau, B.J., Starcevic, S.L., and Novak, R.F. (2004). Gene-specific dye bias in microarray reference designs. *Federation of European Biochemical Societies Letters* 560, 120-124.

Fukui, K. *et al.* (1999). Distance dependence of electron transfer in acridine-intercalated DNA. *J. Photochem. Photobiol. B: Biol.* 50, 18-27.

Han, J., Lee, H., Nguyen, N.Y., Beaucage, S.L., and Puri, R.K. (2005). Novel multiple 5'-amino-modified primer for DNA microarrays. *Genomics* 86, 252-258.

Heinlein, T. *et al.* (2003). Photoinduced electron transfer between fluorescent dyes and guanosine residues in DNA-hairpins. *J. Phys. Chem.* 107, 7957-7964.

Herz, A.H. (1974). Dye-dye interactions of cyanines in solution and at AgBr surfaces. *Photogr. Sci. Eng.* 18, 323-335.

Marras, S.A.E. *et al.* (2002). Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.* 30, e122

Martin-Magniette, M.L. *et al.* (2005). Evaluation of the gene-specific dye bias in cDNA microarray experiments. *Bioinformatics* 21, 1995-2000.

Randolph, J.B. and Waggoner, A.S. (1997). Stability, specificity and fluorescence brightness of multiply-labeled fluorescent DNA probes. *Nucleic Acids Research.* 25, 2923-2929.

Rosenzweig, B.A. *et al.* (2004). Dye-bias correction in dual-labeled cDNA microarray gene expression measurements. *Environmental Health Perspectives* 112, 480-487.

Seidel, C.A.M. *et al.* (1996). Nucleobase-specific quenching of fluorescent dyes. 1. Nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies., *J. Phys. Chem.* 100, 5541-5553.

Spiess, A.N. *et al.* (2003). Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. *BMC Genomics* 4, 44.

Steenken, S. and Jovanovic, S.V. (1997). How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solutions. *J. Am. Chem. Soc.* 119, 617-618.

Tseng, G.C. *et al.* (2001). Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations, and assessment of gene effects. *Nucleic Acids Res.* 29, 2549-2557.

Uchida, S. *et al.* (2005). Detection and normalization of biases present in spotted cDNA microarray data: a composite method addressing dye, intensity-dependent, spatially-dependent, and print-order biases. *DNA Res.* 12, 1-7.

Walter, N.G. and Burke, J.M. (1997). Real-time monitoring of hairpin ribozyme kinetics through base-specific quenching of fluorescein-labeled substrates. *RNA* 3, 392-404.

Yang, Y.H. *et al.* (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation *Nucleic Acids Res.* 30, e15.