

SHORT COMMENTARY

OPEN ACCESS
Full open access to this and
thousands of other papers at
<http://www.la-press.com>.

Microarray Data Analysis of Gene Expression Evolution

Honghuang Lin

Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA. Email: honghuang_lin@dfci.harvard.edu

Abstract: Microarrays are becoming a widely used tool to study gene expression evolution. A recent paper by Wang and Rekaya describes a comprehensive study of gene expression evolution by microarray.¹ The work provides a perspective to study gene expression evolution in terms of functional enrichment and promoter conservation. It was found that gene expression patterns are highly conserved in some biological processes, but the correlation between promoter and gene expression is insignificant. This scope of this work and future improvement to study gene expression evolution will be discussed in this article.

Keywords: microarray, data analysis, gene expression evolution, normalization, function enrichment

Gene Regulation and Systems Biology 2009:3 211–214

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.

The advance of microarray technology enables scientists to monitor the expression profile of thousands of genes simultaneously, making it a possible tool to study transcriptome evolution. Microarrays have been widely used to study expression relationship between humans and other organisms.²⁻⁶ The rationale behind these studies is that orthologous tissues carry out similar physiological functions, which suggests that they are likely to have similar expression profiles. In particular, the expression profile should be conserved for functionally important genes.

A recent paper by Wang and Rekaya describes a comprehensive study of gene expression evolution between humans and mice.¹ Two human/mouse gene expression data sets^{2,7} and one yeast expression data set⁸ were analyzed. The expression similarity was measured by two methods, relative abundance (RA)⁵ and all one-to-one ortholog pairs.⁹ Significant expression conservation was observed between functional related genes in terms of gene ontology (GO). Such conservation could be found in both related species (human vs. mouse) and distant species (human vs. yeast). The authors proposed that events like gene duplication and speciation might result in conservation loss. Expression conservation is not solely dependent on the degree of sequence identity or evolutionary

divergence time.^{1,9} Similar results were also observed in previous studies.⁴⁻⁶ It should be noted that GO is not always be the only or most appropriate source of gene functional annotation. Knowledge from other sources, such as DAVID,¹⁰ Pfam,¹¹ and UniProt,¹² might be adopted in the future study.

Wang and Rekaya also investigated the correlation between promoter sequences and gene expression based on global alignment, local alignment and motif-count. Weak correlation was observed between humans and mice. Such correlation, however, was not observed between humans and yeast, suggesting different regulatory mechanisms might be involved in these two species.¹ Moreover, promoter function is highly context dependent, which limits the capability of homology search for functional annotation.¹³ Duplication and transposition of DNA motifs might also result in promoter mutations together with nucleotide mutations.¹

The expression divergence between species is likely to be overestimated due to various factors. The expression of each gene is usually interrogated by multiple probes called a probeset. The intensity signals from each probe in a probeset are then summarized to obtain the overall expression measurement for the gene.¹⁴⁻¹⁶ Different probesets for the same gene

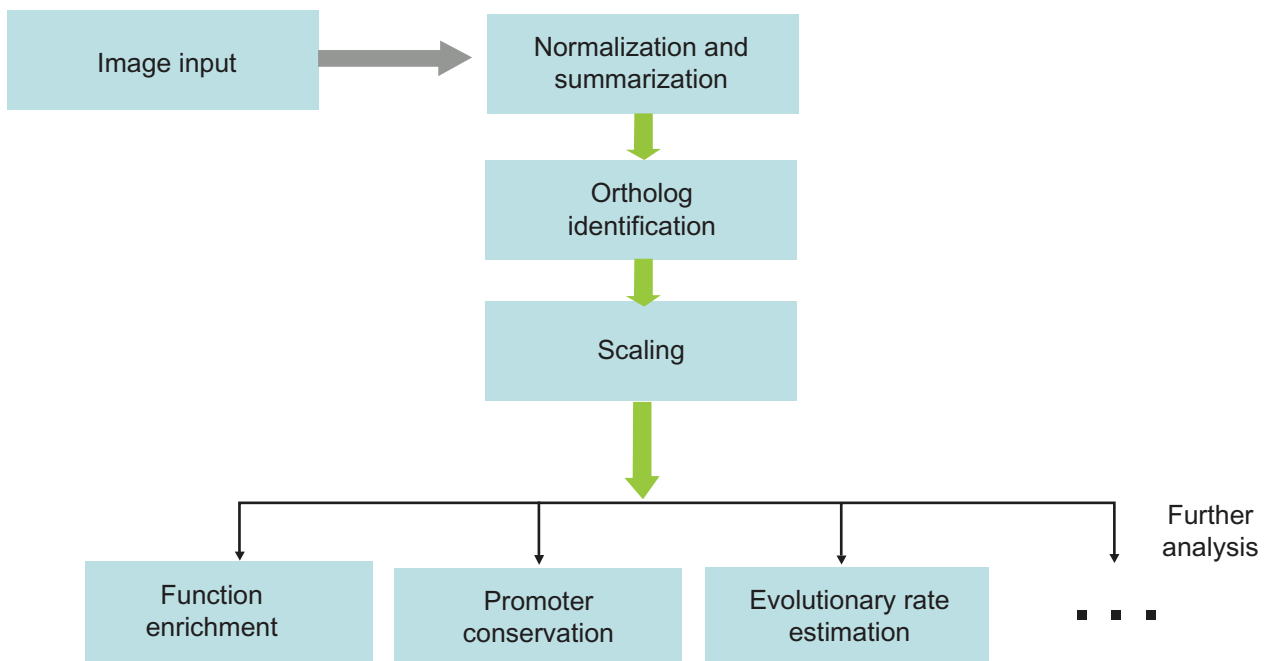


Figure 1. A flowchart of typical steps involving in microarray data analysis of gene expression evolution.



in different species might have different sensitivity, which might result in low correlation of expression profiles for between-species comparison.⁵ It was estimated that the measurement error is likely to be attributable to the majority of expression divergence observed in microarray data.⁵ Liao et al. introduced relative abundance (RA) to measure the relative expression level of a gene in a given tissue among the sampled tissues, which showed better performance than using gene measurement alone.⁵ The method was also adopted in Wang and Rekaya's study, and succeeded in identifying highly conserved functional groups. Other factors, such as DNA methylation, RNA alternative splicing, and transcription factor co-evolution, could also affect gene expression.^{13,17} Cross-hybridization is another cause attributable to the inaccurate signal measurement. Some studies found that excluding suboptimal probesets would reduce the effects of cross-hybridization,¹⁸ although its significance is still controversial.⁵ Gene expression profiling is usually studied under different experimental conditions, cell types, and development stages, resulting in divergent sets of genes expressed. A subset, such as a pathway, could be studied, instead of the whole sets of unrelated microarray data, to avoid the overall complexity.^{4,19}

Systematic bias might be introduced during the preparation of sample libraries, hybridization, or image scanning. Proper normalization is thus an essential step in gene expression evolution study. The simplest normalization method is to adjust array signals according to the global signal median, which would, on the other hand, result in local intensity bias. Lowes normalization is a widely used normalization method. It applies a locally weighted linear regression to eliminate intensity-dependent local biases, making it robust to outliers.²⁰ Quantile method normalizes the distribution of probe intensities across different arrays to a baseline, usually the sample with median intensities. In practice, quantile normalization is recommended to be used for gene expression evolution due to its low variance and bias.²¹ A flow-chart of typical steps involving in microarray data analysis of gene expression evolution is shown in Figure 1.

Overall, the work by Wang and Rekaya provides a functional significance approach to investigating gene expression evolution between humans and mice.

Coupled with technologies to alleviate the negative effects from experimental variation, cross hybridization and systematic bias, microarray would become a powerful tool to study gene expression evolution.

Disclosure

The author reports no conflicts of interest.

References

1. Wang Y, Rekaya R. A Comprehensive Analysis of Gene Expression Evolution Between Humans and Mice. *Evolutionary Bioinformatics*. 2009;5: 81–90.
2. Su AI, Cooke MP, Ching KA, et al. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A*. 2002;99(7):4465–70.
3. Yanai I, Graur D, Ophir R. Incongruent Expression Profiles between Human and Mouse Orthologous Genes Suggest Widespread Neutral Evolution of Transcription Control. *OMICS: A Journal of Integrative Biology*. 2004;8(1):15–24.
4. McCarroll SA, Murphy CT, Zou S, et al. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat Genet*. 2004;36(2):197–204.
5. Liao BY, Zhang J. Evolutionary conservation of expression profiles between human and mouse orthologous genes. *Mol Biol Evol*. 2006;23(3):530–40.
6. Tirosh I, Barkai N. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. *Genome biology*. 2007;8(4):R50.
7. Su AI, Wiltshire T, Batalov S, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A*. 2004;101(16): 6062–7.
8. Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular biology of the cell*. 1998;9(12): 3273–97.
9. Dutilh BE, Huynen MA, Snel B. A global definition of expression context is conserved between orthologs, but does not correlate with sequence conservation. *BMC genomics*. 2006;7:10.
10. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 2009;4(1):44–57.
11. Finn RD, Tate J, Mistry J, et al. The Pfam protein families database. *Nucleic Acids Res*. 2008;36(Database issue):D281–8.
12. Wu CH, Apweiler R, Bairoch A, et al. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*. 2006;34(Database issue):D187–91.
13. Wray GA, Hahn MW, Abouheif E, et al. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*. 2003;20(9):1377–419.
14. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*. 2001; 98(1):31–6.
15. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*. 2003;4(2):249–64.
16. Zhou L, Rocke DM. An expression index for Affymetrix GeneChips based on the generalized logarithm. *Bioinformatics*. 2005;21(21):3983–9.
17. Zhang Z, Gu J, Gu X. How much expression divergence after yeast gene duplication could be explained by regulatory motif evolution? *Trends Genet*. 2004;20(9):403–7.
18. Huminięki L, Wolfe KH. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res*. 2004;14(10A):1870–9.
19. Alter O, Brown PO, Botstein D. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(6):3351–6.



20. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 2002;30(4):e15.
21. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England).* 2003;19(2):185–93.

Publish with Libertas Academica and every scientist working in your field can read your article

“I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely.”

“The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I’ve never had such complete communication with a journal.”

“LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought.”

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>