confirming that the marriage between the postgenomic and classical biochemical approaches is essential in future biochemical research.

Acknowledgements
This article is dedicated to Professor Hans Eklund on the occasion of his retirement. This work has been supported by a grant from the Swedish Research Council.

References

1 Blattner, F.R. *et al.* (1997) The complete genome sequence of *Escherichia coli* K12. *Science* 277, 1453–1474
2 Goffeau, A. *et al.* (1996) Life with 6000 genes. *Science* 274, 563–567
3 Van Lanen, S.G. and Shen, B. (2006) Microbial genomics for the improvement of natural product discovery. *Curr. Opin. Microbiol.* 9, 252–260
4 Vaillancourt, F.H. *et al.* (2005) Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* 436, 1191–1194
5 Galonic, D.P. *et al.* (2007) Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* 3, 113–116
6 Loh, K.D. *et al.* (2006) A previously undescribed pathway for pyrimidine catabolism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5114–5119
7 Zimmer, D.P. *et al.* (2000) Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc. Natl. Acad. Sci. U. S. A.* 97, 14674–14679
8 Osterman, A. (2006) A hidden metabolic pathway exposed. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5637–5638
9 Gojkovic, Z. *et al.* (2001) Eukaryotic beta-alanine synthases are functionally related but have high degree of structural diversity. *Genetics* 158, 999–1011
10 Andersen, G. *et al.* (2007) A gene duplication led to specialized gamma-aminobutyrate and beta-alanine aminotransferase in yeast. *FEBS J.* 274, 1804–1817
11 Andersen, G. *et al.* (2006) Catabolism of pyrimidines in yeast: A tool to understand degradation of anticancer drugs. *Nucleosides Nucleotides Nucleic Acids* 25, 991–996
12 Schmitt, U. *et al.* (1996) Purification and characterization of dihydropyrimidine dehydrogenase from *Alcaligenes eutrophus*. *Arch. Biochem. Biophys.* 332, 175–182
13 Soong, C.L. *et al.* (2002) Barbiturase, a novel zinc-containing amidohydrolase involved in oxidative pyrimidine metabolism. *J. Biol. Chem.* 277, 7051–7058
14 Butcher, R.A. *et al.* (2007) The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1506–1509
15 Myllykallio, H. *et al.* (2002) An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 297, 105–107
16 Graziani, S. *et al.* (2006) Catalytic mechanism and structure of viral flavin-dependent thymidylate synthase ThyX. *J. Biol. Chem.* 281, 24048–24057
17 Mathews, C.K. and Van Holde, K.E. (1996) Nucleotide metabolism. In: *Biochemistry* (2nd edn), Benjamin Cummings
18 Wasternack, C. (1980) Degradation of pyrimidines and pyrimidine analogs – pathways and mutual influences. *Pharmacol. Ther.* 8, 629–651
19 Simaga, S. and Kos, E. (1979) Uracil catabolism by *Escherichia coli* K12S. *Z. Naturf. C: Biosci.* 33, 1006–1008
20 Reinbothe, H. (1964) Urea formation from pyrimidines in fruit-bodies of higher basidiomycetes. *Tetrahedron Lett.* 5, 2651–1657
21 Rawls, J.M. (2006) Analysis of pyrimidine catabolism in *Drosophila melanogaster* using epistatic interactions with mutations of pyrimidine biosynthesis and beta-alanine metabolism. *Genetics* 172, 1665–1674
22 West, T.P. (1998) Isolation and characterization of an *Escherichia coli* B mutant strain defective in uracil catabolism. *Can. J. Microbiol.* 44, 1106–1109
23 Kao, C-H. and Hsu, W-H. (2003) A gene cluster involved in pyrimidine reductive catabolism from *Brevibacillus agri* NCHU1002. *Biochem. Biophys. Res. Commun.* 303, 848–854

0168-9525/$ – see front matter.
doi:10.1016/j.tig.2007.05.007

Genome Analysis

MicroRNA regulation and interspecific variation of gene expression

Qinghua Cui*, Zhenbao Yu*, Enrico O. Purisima and Edwin Wang

Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec, H4P 2R2, Canada

MicroRNAs (miRNAs) modulate expression of their target genes in various tissues and at different developmental stages, but it is unclear whether they drive cross-species variation in gene expression. By comparing data from mammal and fly species we found that the cross-species expression variation of miRNA targets is significantly lower than that of other genes. This implies that miRNAs can affect gene expression by reducing stochastic noise, buffering cross-species variation and constraining evolutionary gene expression variation.

Role of miRNAs in gene expression

Striking differences in phenotypes exist in humans and other organisms. Traditionally we have assumed that DNA sequences that encode proteins contribute to the adaptive changes and the phenotypic differences between organisms. Recently, microarrays have been used to estimate the genome-wide variation in gene expression across multiple species [1–9]. These studies have identified numerous genes that are expressed differently in closely related species and suggest that natural variation in gene expression rather than structural changes in gene products probably accounts for a substantial part of phenotypic variation [1,3–7,10]. However, the molecular basis of cross-species differences in gene expression is still unclear.

Polymorphism and divergence of gene expression are influenced by a variety of factors, including gene regulation

Corresponding author: Wang, E. (edwin.wang@cnrc-nrc.gc.ca)..
* These authors contributed equally to this work.
Available online 4 May 2007.

Table 1. Average interspecies expression variations (EVs) of miRNA target genes and non-miRNA-target genes in four groups of species

	Human cf. chimpanzee	Human cf. mouse	*Drosophila melanogaster* cf. *D. simulans*
miRNA targets[a]	−0.127	−0.097	−0.105
Non-miRNA-target genes[a]	0.091	0.075	0.033
P value[b]	2.2×10^{-16}	8.3×10^{-15}	1.3×10^{-9}

[a]The values are averages of *Z*-scores and were calculated as described in section M2 of supplementary material online.
[b]Based on *t* test.

by *trans* factors on the *cis* elements located in the regulatory regions of the genes. In the past few years, microRNAs (miRNAs) have been found to have a significant role in regulating gene expression at post-transcriptional and translational levels by base-pairing with the *cis* elements located in the 3′-terminus of target mRNAs. It is currently estimated that miRNAs account for ~1% of predicted genes in higher eukaryotic genomes and that up to one-third of human genes might be regulated by miRNAs. MiRNAs are known to be involved in a broad spectrum of biological processes and different diseases [11–13], and miRNAs and their targets seem to form a complex regulatory network. Thus, we asked whether miRNA-mediated post-transcriptional regulation contributes to driving cross-species divergence of gene expression and the evolution of transcriptomes.

Cross-species variation of expression of miRNA targets is significantly lower than that of other genes

We analyzed the relationship between the distribution of miRNA targets and cross-species gene expression variation using several public datasets. We first calculated the average cross-species expression variation (EV) of each gene in five different tissues between humans and chimpanzees using a public dataset that presents the expression levels of the genes in human and chimpanzee tissues [1]. The definition and calculation of EV are based on the method of Tirosh *et al.* [14] and are described in the supplementary material online. The human miRNA targets were computationally predicted by Krek *et al.* based on the 3′ untranslated region (UTR) conserved among five mammals [15]. Of the human genes that were used to predict miRNA targets [15], we used only those that had human and chimpanzee orthologs in the microarray dataset [1]. We divided the resulting set of genes into two groups and calculated the average EV of the genes in each group. One group contained all miRNA targets (4462 genes), whereas the other group contained the rest (6341 genes). We found that the cross-species expression variation of the genes in the miRNA target group (average EV = −0.127) was significantly lower than that in the non-miRNA-target gene group (average EV = 0.091, $P < 2.2 \times 10^{-16}$, *t* test; Table 1). For a more detailed analysis, we ranked and subgrouped the genes according to their cross-species expression variation and calculated the ratio of miRNA targets to the total genes in each group, termed miRNA target rate here (see the supplementary material online for more details). As shown in Figure 1a, the miRNA target rate rapidly decreased (Pearson correlation coefficient $r = -0.90$, $P < 2.2 \times 10^{-16}$) with increasing EV values. Taken together, these results suggest that the fraction of miRNA targets in gene groups negatively correlates with gene expression divergence between human and chimpanzee.

To examine whether this observation is valid in other mammalian species, we analyzed the relationship of miRNA target distribution with cross-species gene expression variation between human and mouse using a gene expression microarray dataset reported by Su *et al.* [16]. We calculated the average EV between human and mouse for each gene in 26 different tissues. We found that the average EV of the miRNA target genes (−0.097) was lower than that of non-miRNA-target genes (0.075) between human and mouse ($P < 8.3 \times 10^{-15}$, *t* test; Table 1). Similarly, the gene-ranking analysis further confirmed the negative correlation between miRNA target genes and

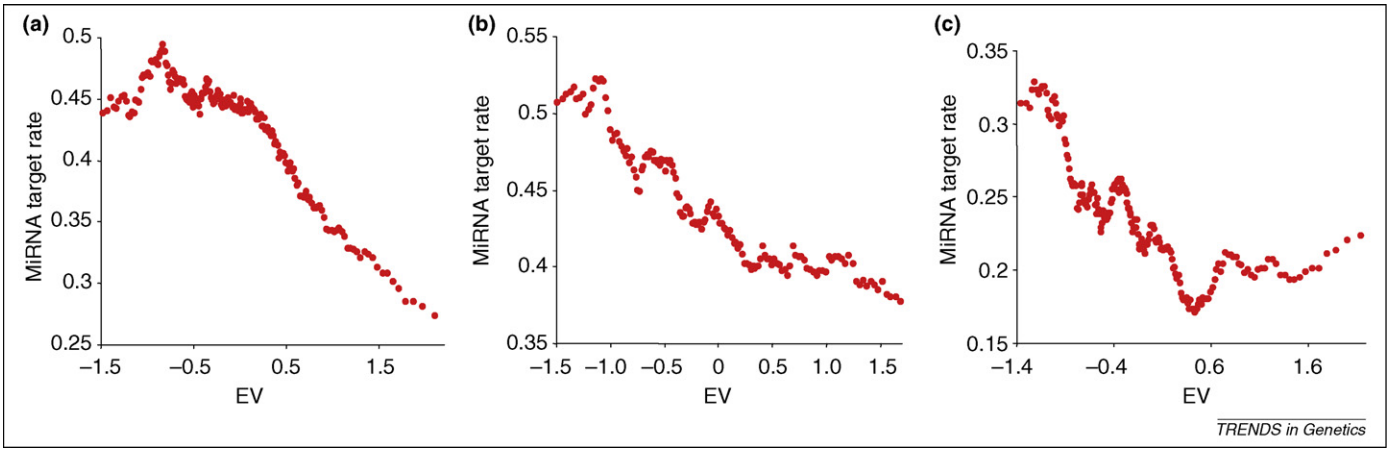

Figure 1. Distribution of miRNA targets among genes with different values for cross-species gene expression variation (EV). The EV value for each gene was calculated as described in the supplementary material online. Genes were grouped according to the EV value. The ratio of miRNA targets to total genes in each group (miRNA target rate) was calculated. **(a)** Human and chimpanzee; **(b)** human and mouse; **(c)** *D. melanogaster* and *D. simulans*.

gene expression divergence between mouse and human (Figure 1b).

To determine whether the same phenomena could be observed in non-mammals, we studied the relationship of miRNA targets [17] with gene expression variation between two *Drosophila* species, *D. melanogaster* and *D. simulans* [3]. We also used the *D. melanogaster* genes that were used for miRNA target prediction, as background genes (see the supplementary material online). We found that the interspecies gene expression variation of miRNA targets (–0.105) was significantly lower than that of the non-miRNA-target genes (0.033, $P < 1.3 \times 10^{-9}$, t test; Table 1). Furthermore, the miRNA target rate dramatically decreased with the increase of interspecies gene expression variation (Figure 1c).

A detailed analysis confirmed that the observed correlations between miRNA targets and cross-species variation of gene expression do not depend on specific tissues, miRNAs or gene function categories. For example, the negative correlations were found in each of the tissues analyzed (Figure S1a,b in supplementary material online). Similarly, for most miRNAs, the average EV of their targets was lower than that of the total genes (Figure S2a–c in supplementary material online). In addition, we found that the average EVs of miRNA targets were lower than those of the genes that are not miRNA targets in most gene ontology (GO) groups (Figure S3a–c in supplementary material online).

Similar results were obtained (Tables S1–S3 in supplementary material online) when we performed the same analysis using the experimentally validated miRNA targets, the miRNA targets of tissue-specifically expressed miRNAs, and the nonconserved miRNA targets, respectively, suggesting that the negative correlation between miRNA regulation and gene expression divergence does not depend on specific gene groups.

Recently, Tirosh *et al.* reported that TATA box-containing genes have an increased cross-species gene expression variability, suggesting that the TATA box amplifies fluctuations in promoter activation and then increases gene expression divergence [14]. We asked whether a correlation of gene expression regulation through the TATA box and miRNA targeting exists. We found that the contribution of miRNAs to cross-species gene expression divergence is not mediated by the TATA box (see the supplementary material online).

Possible mechanisms of miRNA regulation of cross-species gene expression variation

Wittkopp *et al.* recently reported that cross-species gene expression differences between *D. melanogaster* and *D. simulans* could be caused by changes in *cis* regulation [18]. Although miRNA binding sites are *cis* elements located in the 3′-terminal regulatory regions of mRNAs, our studies suggest that mutation of miRNAs and their binding sites contributes little to increased divergence of gene expression. This is consistent with recent observations showing that mutations at miRNA binding sites undergo a strong negative selection [19] and that sequences at miRNA binding sites are greatly conserved across species [15].

On the contrary, miRNAs might provide a genetic buffer to constrain gene expression divergence. Indeed, Hornstein and Shomron [20] recently proposed a hypothesis of genetic buffering by miRNAs. Genetic buffering can be achieved by miRNAs that repress target genes that have low abundance in the tissues where the miRNA is expressed. However, in many situations, miRNA and its targets are coexpressed at intermediate levels [21,22]. Hornstein and Shomron proposed that in such situations miRNA has the role of buffering fluctuations in target gene expression through a feed-forward loop architecture in which a transcription factor regulates both miRNAs and the targets of the miRNAs [20]. A recent report provides an example of miRNAs buffering the fluctuation of gene expression: miR-9a serves to buffer the stochastically fluctuating expression of genes by establishing 'thresholds' for a positive regulatory loop in fly development [23,24]. We previously showed that miRNAs preferentially regulate positive regulatory loops [11], possibly providing a common mechanism in buffering gene expression noise.

Concluding remarks

In summary, we found that cross-species expression variation of miRNA targets is significantly lower than that of non-miRNA-targets in the datasets we examined in mammals and flies. These results suggest that miRNA-mediated post-transcriptional regulation might contribute to buffering the divergence of gene expression across species, and have a role in evolutionary constraints on gene expression variation. Genetic buffering can be achieved by miRNAs that repress target genes that have low abundance in the tissues where the miRNA is expressed. Alternatively, miRNAs could buffer stochastically fluctuating expression of genes by establishing 'thresholds' for positive regulatory loops.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.04.003.

References

1 Khaitovich, P. *et al.* (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309, 1850–1854
2 Enard, W. *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296, 340–343
3 Ranz, J.M. *et al.* (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300, 1742–1745
4 Gilad, Y. *et al.* (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440, 242–245
5 Rifkin, S.A. *et al.* (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438, 220–223
6 Rifkin, S.A. *et al.* (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* 33, 138–144
7 Denver, D.R. *et al.* (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* 37, 544–548
8 Brem, R.B. *et al.* (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
9 Bergmann, S. *et al.* (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol.* 2, E9
10 Hsieh, W.P. *et al.* (2003) Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* 165, 747–757

11 Cui, Q. *et al.* (2006) Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.* 2, 46
12 Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297
13 Alvarez-Garcia, I. and Miska, E.A. (2005) MicroRNA functions in animal development and human disease. *Development* 132, 4653–4662
14 Tirosh, I. *et al.* (2006) A genetic signature of interspecies variations in gene expression. *Nat. Genet.* 38, 830–834
15 Krek, A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500
16 Su, A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6062–6067
17 Grun, D. *et al.* (2005) microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS. Comput. Biol.* 1, e13
18 Wittkopp, P.J. *et al.* (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430, 85–88
19 Chen, K. and Rajewsky, N. (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.* 38, 1452–1456
20 Hornstein, E. and Shomron, N. (2006) Canalization of development by microRNAs. *Nat. Genet.* 38 (Suppl.), S20–S24
21 Farh, K.K. *et al.* (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* 310, 1817–1821
22 Sood, P. *et al.* (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2746–2751
23 Li, Y. *et al.* (2006) MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev.* 20, 2793–2805
24 Cohen, S.M. *et al.* (2006) Denoising feedback loops by thresholding – a new role for microRNAs. *Genes Dev.* 20, 2769–2772

0168-9525/$ – see front matter.

doi:10.1016/j.tig.2007.04.003

Gene essentiality, gene duplicability and protein connectivity in human and mouse

Han Liang and Wen-Hsiung Li

Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA

It has previously been found that, in yeast, gene essentiality is positively correlated with protein connectivity (number of interaction partners) but negatively correlated with the existence of gene duplicates and that highly connected proteins tend to have a low gene duplicability. Using data from human and mouse, we show here that, in mammals, the first of these relationships holds true, but unlike the second relationship in yeast, highly connected mammalian proteins tend to have a high gene duplicability, and there is no correlation between gene essentiality and gene duplication in mammals.

Introduction

There has been much interest in the relationships among gene function, phenotypic effect of gene deletion or knockout, and gene duplication at the genomic level [1–9]. For this purpose, three terms are often used: (i) protein connectivity, which is defined as the number of links that a protein node has to other nodes in the protein interaction network; (ii) gene essentiality, which is defined using words such as 'the deletion of a gene from the genome has a lethal effect or causes infertility' [10,11]; and (iii) gene duplicability, which describes the likelihood of a gene having one or more paralogs [8]. So far, however, most of our knowledge about the relationships among these three factors comes from yeast. In yeast, a protein that is highly 'connected' to other proteins (i.e. that interacts with many other proteins) tends to result in the death of the organism if it is deleted from the genome [3,12,13]. This is commonly known as the 'centrality–lethality' rule, which either reflects the crucial role of hub proteins (i.e. highly connected proteins) in the architecture of the network [3] or is simply because hub proteins have a higher probability of engaging in essential protein–protein interactions [14]. Furthermore, the proportion of essential (deletion-lethal) genes is significantly higher among singletons than among duplicates, and the deletion of a duplicate gene is, on average, less severe than the deletion of a singleton [2]. Recent studies indicated a negative correlation between protein connectivity and gene duplicability, which implies that genes with a higher protein connectivity tend to have fewer duplicate genes in the yeast genome [15]. Do these relationships hold true in such complex organisms as mammals?

Relationships among gene essentiality, gene duplicability and protein connectivity

First, the available mouse targeted knockout phenotypic annotations were extracted from the Mouse Genome Database (MGD; http://www.informatics.jax.org/) [16], and mouse genes and their orthologous human genes (annotated by MGD) were classified as essential or non-essential genes. Here, we defined an essential gene as a gene whose knockout phenotype is annotated as lethality (including embryonic, perinatal and postnatal lethality) or infertility [10,11]. Second, protein connectivity was calculated based on the human protein–protein interaction data (including both yeast two-hybrid and literature-curated interactions) from the study by Rual *et al.* [17]. Finally, gene family information was obtained (i.e. gene family IDs) in the human and mouse genomes, according to the annotation in the Ensembl Genome Database [18,19].

Corresponding author: Li, W.-H. (whli@uchicago.edu).
Available online 21 May 2007.