

Impacts of Pretranscriptional DNA Methylation, Transcriptional Transcription Factor, and Posttranscriptional microRNA Regulations on Protein Evolutionary Rate

Trees-Juen Chuang* and Tai-Wei Chiang

Division of Physical & Computational Genomics, Genomics Research Center, Academia Sinica, Taipei, Taiwan

*Corresponding author: E-mail: trees@gate.sinica.edu.tw.

Accepted: June 5, 2014

Abstract

Gene expression is largely regulated by DNA methylation, transcription factor (TF), and microRNA (miRNA) before, during, and after transcription, respectively. Although the evolutionary effects of TF/miRNA regulations have been widely studied, evolutionary analysis of simultaneously accounting for DNA methylation, TF, and miRNA regulations and whether promoter methylation and gene body (coding regions) methylation have different effects on the rate of gene evolution remain uninvestigated. Here, we compared human–macaque and human–mouse protein evolutionary rates against experimentally determined single base-resolution DNA methylation data, revealing that promoter methylation level is positively correlated with protein evolutionary rates but negatively correlated with TF/miRNA regulations, whereas the opposite was observed for gene body methylation level. Our results showed that the relative importance of these regulatory factors in determining the rate of mammalian protein evolution is as follows: Promoter methylation \approx miRNA regulation $>$ gene body methylation $>$ TF regulation, and further indicated that promoter methylation and miRNA regulation have a significant dependent effect on protein evolutionary rates. Although the mechanisms underlying cooperation between DNA methylation and TFs/miRNAs in gene regulation remain unclear, our study helps to not only illuminate the impact of these regulatory factors on mammalian protein evolution but also their intricate interaction within gene regulatory networks.

Key words: promoter/gene body methylation, transcription factor, microRNA, protein evolutionary rate, comparative genomics.

Introduction

Various factors have been known to control gene expression and form a complex regulatory network. Regulation of gene expression is strongly associated with the maintenance of normal cells and a variety of biological functions. The most prominent gene regulators are DNA methylation, transcription factor (TF), and microRNA (miRNA), which regulate gene expression at the pretranscriptional, transcriptional, and posttranscriptional levels, respectively. DNA methylation is a heritable epigenetic marker that regulates gene expression without altering DNA sequence (Egger et al. 2004). This modification is highly associated with many cellular processes, including transcription, genomic imprinting, suppression of transposons, X-chromosome inactivation, chromatin structure, embryonic development, and carcinogenesis (Li et al. 1993; Heard et al. 1997; Walsh et al. 1998; Reik et al. 2001; Feinberg and Tycko 2004; Laurent et al. 2010). Several studies have indicated that promoter methylation

and gene body methylation exhibit different correlation patterns with gene expression. Promoter methylation is generally associated with transcriptional suppression, whereas gene body methylation is associated with active transcription (Jones and Takai 2001; Hellman and Chess 2007; Ball et al. 2009; Bogdanovic and Veenstra 2009; Lister et al. 2009; Feng et al. 2010; Laurent et al. 2010; Maunakea et al. 2010; Xiang et al. 2010; Zemach et al. 2010; Defossez and Stancheva 2011; Jones 2012); therefore, methylation in promoter and gene body regions may play distinct biological roles.

As for the two *trans*-regulatory factors (TFs and miRNAs), TFs are proteins that bind to specific DNA sequences (so-called “TF binding sites” [TFBSs]) usually located within promoters, through which they facilitate or repress the transcription of their target genes (Elnitski et al. 2006; Vaquerizas et al. 2009); on the other hand, miRNAs are small (~22 nt) noncoding RNAs that target mRNAs and regulate gene expression through mRNA cleavage or translation repression

(Bartel 2004, 2009). It has been reported that miRNAs can form intricate feedback and feed-forward loops with TFs within the context of gene regulatory networks (Hornstein and Shomron 2006; Shalgi et al. 2007; Tsang et al. 2007; Su et al. 2010). Such miRNA–TF networks are important for the stability of gene regulation mechanisms (Shalgi et al. 2007; Yu et al. 2008), and may play crucial roles in diverse biological processes (Marson et al. 2008; Dahan et al. 2011). In contrast, interactions between these two *trans*-regulatory factors (i.e., TFs and miRNAs) and DNA methylation during gene regulation are relatively poorly understood.

In terms of molecular evolution, DNA methylation has been reported to remarkably increase the rate of spontaneous C-to-T mutations at CpG dinucleotides (Ehrlich and Wang 1981; Hwang and Green 2004; Mugal and Ellegren 2011), resulting in enhanced sequence divergence. On the other hand, it has been recently reported that hypermethylated genes are subject to stronger selective constraints than hypomethylated genes (Hunt et al. 2010; Lyko et al. 2010; Park et al. 2011; Sarda et al. 2012; Takuno and Gaut 2012). We previously observed that the first exons of transcripts are more susceptible to mutagenic effects, whereas the internal and last exons are more affected by the regulatory effects of DNA methylation (Chuang et al. 2012). We further indicated that the extent of gene body methylation correlates highly with within-gene variations (e.g., the type of exonic sequences, relative genic position, and degeneracy of coding nucleotides) in evolutionary rates at both the exon and nucleotide levels (Chuang et al. 2012; Chuang and Chen 2014). On the other hand, the evolutionary effects of TFs and miRNAs have been carefully examined, revealing that genes targeted by more different TFs (number of different TFs designated as “ N_{TF} ”) or miRNAs (number of different miRNAs designated as “ N_{miR} ”) tend to evolve more slowly in diverse species (Cheng et al. 2009; Xia et al. 2009; Wang et al. 2010; Chen, Chuang et al. 2011; Chen et al. 2013). These results manifested that all these regulatory factors are important indicators of evolutionary rates, regardless of DNA methylation at the pretranscriptional level, TF regulation at the transcriptional level, or miRNA regulation at the posttranscriptional level. However, to the best of our knowledge, no systematic evolutionary analysis is currently available that simultaneously accounts for these three regulatory factors. The following questions still await exploration: whether promoter and gene body methylation differentially correlated with the evolutionary rates of their target genes, N_{TF} , and N_{miR} ; which of these factors has a greater effect on the rate of mammalian protein evolution; and whether DNA methylation and *trans*-regulations have an interactive influence on protein evolutionary rates.

In this study, we collected single base-resolution DNA methylation data and TF- and miRNA-binding data from human, and systematically examined the correlations between these regulatory factors (i.e., levels of promoter/gene body

methylation, N_{TF} , and N_{miR}) and the evolutionary rates of their target genes (nonsynonymous substitution rate [d_N], synonymous substitution rate [d_S], and d_N/d_S ratio). To control for other confounding factors that may affect the evolutionary rates of protein-coding genes, we also considered the following eight biological factors in our statistical analyses: 1) Protein connectivity (Lemos et al. 2005; Plotkin and Fraser 2007; Xia et al. 2009; Liao et al. 2010; Wang et al. 2010); 2) gene expression level (Liao et al. 2006, 2010; Larracunte et al. 2008; Xia et al. 2009; Chen, Chuang et al. 2011; Yang and Gaut 2011); 3) tissue-specific gene expression (Liao et al. 2006, 2010; Larracunte et al. 2008; Park and Choi 2010; Yang and Gaut 2011); gene compactness in terms of 4) untranslated region (UTR) length (Liao et al. 2006; Cheng et al. 2009; Yang and Gaut 2011), 5) intron length (Marais et al. 2005; Liao et al. 2006, 2010; Yang and Gaut 2011), and 6) intron number (Larracunte et al. 2008; Yang and Gaut 2011); and protein structure in terms of 7) solvent accessibility (Bloom et al. 2006; Lin et al. 2007; Zhou et al. 2008; Franzosa and Xia 2009) and 8) disorder content (Kim et al. 2008; Brown et al. 2010; Chen, Chuang et al. 2011). In this way, we show that the levels of DNA methylation of both promoters and gene bodies are important indicators of protein evolutionary rates (d_N and d_N/d_S) when other confounding factors are controlled. Interestingly, protein evolutionary rates are positively correlated with the level of promoter methylation, but negatively correlated with the level of gene body methylation. Furthermore, promoter methylation is negatively correlated with N_{TF} and N_{miR} , whereas gene body methylation is positively correlated with the two *trans*-regulations. We also report that the level of promoter methylation and N_{miR} have the greatest influence on protein evolutionary rates, and they have a dependent effect on the rate of mammalian protein evolution. This result supports the previously hypothesized potential reciprocal regulation between these two regulatory factors (Taguchi 2013a, 2013b).

Materials and Methods

Collection of Single Base-Resolution DNA Methylation, TFBS, and miRNA Target Data

The human gene annotations, gene orthology assignments, and human–mouse/human–rhesus macaque evolutionary rates (d_N , d_S , and d_N/d_S) were downloaded from the Ensembl database at <http://www.ensembl.org/> (last accessed November 2013) (version 73). We considered only 1:1 human–mouse and human–rhesus macaque orthologs to avoid the confounding factor of gene duplication; furthermore, we considered only the longest isoforms of alternatively spliced genes. The promoter region of a gene was defined as the intergenic region from 8 kb upstream to 2 kb downstream of the Ensembl-annotated gene start position. Genes with genic regions (including exons and introns) ≤ 2 kb in length

were not considered. For accuracy, a gene whose promoter region overlaps with other gene(s) was not considered. The base-resolution DNA methylation data from 12 human cell types (including cultured cells, cells from healthy individuals [S1–S11, table 1], and breast cancer cells [supplementary table S1, Supplementary material online]) were downloaded from NGSmethDB v2 (Geisen et al. 2014) at <http://bioinfo2.ugr.es/NGSmethDB/> (last accessed October 2013). These data sets were generated with bisulfite or MethylC sequencing. CpG dinucleotides with single nucleotide variants were not considered in this study to avoid potential sequencing errors. To ensure the accuracy of the methylome data, only CpG dinucleotides covered by ≥ 5 bisulfite/MethylC reads were considered (such CpG dinucleotides were designated as “sampled CpGs”). A sampled CpG site is regarded as methylated (designated as “mCG”) if $\geq 80\%$ of the mapped reads support the methylation status at the CpG site (Meissner et al. 2008; Laurent et al. 2010). We only considered the genes whose promoter and gene body regions both contained ≥ 10 sampled CpGs to ensure that the examined regions contained sufficient information for estimating the methylation level. Of note, the gene body region of a gene represents the coding sequence, with the exception of regions overlapping with the promoter of the gene (e.g., 2 kb downstream of the Ensemble-annotated gene start position).

TFBS data were obtained by downloading chromatin immunoprecipitation (ChIP) data (which includes 162 human TF ChIP-seq data sets) from the ENCODE project (Bernstein et al. 2012). A given TF was considered to regulate a gene if at least one of its ChIP-seq peaks was located within the promoter region of the gene. Human miRNA target prediction data

(which include 1,267 miRNAs) were downloaded from TargetScan release 6.2 (April 2013) (Ruby et al. 2007; Friedman et al. 2009). For accuracy, we only considered the human miRNA families in which the corresponding target sites were determined to be conserved across mammals using TargetScan (Friedman et al. 2009). The human, rhesus macaque, and mouse genes examined in this study (together with the related information) are available at http://idv.sinica.edu.tw/trees/DM_TF_miRNA/DM_TF_miRNA.html (last accessed June 25, 2014).

Measurement of CpG Dinucleotide Depletion ($CpG_{O/E}$) and the Methylation Level in Promoter and Gene Body Regions

Since a low ratio of observed-to-expected CpG dinucleotides ($CpG_{O/E}$) represents a large fraction of mutated CpG dinucleotides, $CpG_{O/E}$ is a measurement of CpG dinucleotide depletion (Bird and Taggart 1980; Park et al. 2011, 2012). $CpG_{O/E}$ was defined as follows:

$$CpG_{O/E} = \frac{P_{CpG}}{P_C \times P_G} = \frac{\text{number of CpGs} \times \text{length of the examined region}}{\text{number of Cs} \times \text{number of Gs}}$$

where P_{CpG} , P_C , and P_G represent the frequency of CpG dinucleotides, C nucleotides, and G nucleotides, respectively, in the examined promoter/gene body regions. The methylation level of an examined region was measured by calculating the density of mCG per 100 CpG dinucleotides (mCG density). mCG density was defined as

Table 1

Experimentally Determined Single Base-Resolution DNA Methylation Data Used in this Study

Sample	Description (Ref.)	No. of Genes (Sampled #CG ≥ 10)	Average CpG Coverage ^a (Promoter) (%)	Average CpG Coverage ^a (Gene Body) (%)
S1	Peripheral blood B lymphocytes (Hodges et al. 2011)	10,627	76.04	86.69
S2	Peripheral blood hematopoietic stem/progenitor cells (CD133+CD34+CD38–Lin–) (Hodges et al. 2011)	10,451	78.03	89.10
S3	Newborn foreskin fibroblasts (Laurent et al. 2010)	10,388	89.34	93.56
S4	H1 embryonic stem cells (Lister et al. 2009)	10,751	68.54	89.41
S5	Breast cells from adult female (Hon et al. 2012)	10,916	96.78	98.75
S6	Peripheral blood hematopoietic stem/progenitor cells (CD34+CD38–Lin–) (Hodges et al. 2011)	10,722	80.04	89.41
S7	Fetal lung fibroblasts (Lister et al. 2009)	10,848	78.13	93.62
S8	Peripheral blood granulocytic neutrophils (Hodges et al. 2011)	10,081	71.27	82.91
S9	Prefrontal cortex (Zeng et al. 2012)	10,556	80.58	90.38
S10	H9 embryonic stem cells (Laurent et al. 2010)	10,522	91.81	95.31
S11	Fibroblasts derived from H9 embryonic stem cells (Laurent et al. 2010)	10,436	90.01	93.65

^aCoverage of CpG dinucleotides for each promoter/gene body (protein-coding) region = (number of sampled CpG dinucleotides)/(number of sampled CpG dinucleotides + number of nonsampled CpG dinucleotides).

$$\text{mCG density} = \frac{\text{number of mCGs} \times 100}{\text{number of all CpGs sampled}}.$$

Data Retrieval of Protein Connectivity, Gene Expression, Gene Compactness, and Protein Structural Features

Protein connectivity (PPI) data were downloaded from STRING 9.0 (Szklarczyk et al. 2011). Gene expression data were obtained by downloading normalized expression data sets for 78 nonpathogenic human tissues from BioGPS (Wu et al. 2009). If more than one probe set referred to the same gene, the signals from the relevant probe sets were averaged. The gene expression level was defined as the average signal intensity across the 78 examined tissues. The tissue specificity (τ) of a gene was defined as follows:

$$\tau = \frac{\sum_{i=1}^n \left(1 - \frac{\log S(i)}{\log S_{\max}}\right)}{n - 1},$$

where n , $S(i)$, and S_{\max} denote the number of examined tissues (i.e., $n = 78$), the signal intensity, and the highest signal across all examined tissues, respectively (Yanai et al. 2005). A large τ value indicates high tissue specificity for a gene. To minimize potential noise that might be caused by low signal intensities, values of $S(i)$ less than 100 were set to 100 (Liao and Zhang 2006; Liao et al. 2006; Chen et al. 2010). Gene compactness was described based on intron number and average length of UTR/intron for a given gene. Protein structure was described based on solvent accessibility and disorder content. The solvent accessibility of a protein was calculated from the maximum number of exposed residues interacting with solvent molecules over the protein's length; the exposed residues were determined using ACCPro release 4.1 with the default parameters (Cheng et al. 2005). Only proteins <8,000 amino acids in length were considered due to limitations of ACCPro (Cheng et al. 2005). The disorder content of a protein was defined as the percentage of intrinsically disordered regions, estimated by dividing the number of disordered residues by protein length. The disordered residues were predicted using DISOPRED2 version 2.4 with default parameters (Akgul et al. 2004). To minimize the standard error when calculating disorder content, we only considered proteins ≥ 100 amino acids in length (Chen, Chuang et al. 2011).

Calculation of the Relative Contribution to Variability Explained

The relative contribution to variability explained (RCVE) is used to assess the relative importance of each tested factor, which is defined as follows:

$$\text{RCVE} = \frac{R_{\text{full}}^2 - R_{\text{reduced}}^2}{R_{\text{full}}^2},$$

where R_{full}^2 and R_{reduced}^2 denote the R^2 value (share of variability explained) of the full model (including all of the tested factors) and that of the reduced model (excluding the factor of interest), respectively. R^2 is the square of the coefficient of correlation between the observations and their predicted values in a multiple linear regression model, which is a measure of the proportion of total variation of observed outcomes explained by the model. Accordingly, the RCVE coefficient represents the relative contribution of a factor in the context of all other factors included in the full model, which ranges from 0 to 1 with a higher value indicating a more important contribution of the factor of interest to the regression model (Kvikstad et al. 2007).

Results and Discussion

Promoter and Gene Body Methylation Exhibit Opposite Correlation Patterns with the Protein Evolutionary Rates of Their Target Genes

To investigate the levels of DNA methylation in human promoter and gene body regions, we retrieved single base-resolution DNA methylation data from 11 human cell lines (see Materials and Methods; table 1). It should be noted that the gene body regions examined were coding sequences, with the exception of regions overlapping with the promoter regions. Table 1 shows the number of genes examined in each data set. The level of DNA methylation was measured by the density of mCG (see Materials and Methods). For each gene considered, both the examined promoter and gene body regions contained at least 10 sampled CpGs (see Materials and Methods). On average, over 68% of CpGs were sampled for both examined promoters and gene bodies (table 1), indicating that sufficient CpG dinucleotides were sampled in the examined regions.

To probe the impact of promoter and gene body methylation on evolution, we first examined the correlation between $\text{CpG}_{\text{O/E}}$ (i.e., the ratio of the observed-to-expected number of CpG dinucleotides, see Materials and Methods) and promoter/gene body mCG density. Low $\text{CpG}_{\text{O/E}}$ ratios (indicating a high level of C-to-T mutations) have been reported to be caused primarily by DNA methylation (Bird and Taggart 1980; Park et al. 2011, 2012); therefore, an inverse correlation is observed between $\text{CpG}_{\text{O/E}}$ and DNA methylation (Bird and Taggart 1980; Zemach et al. 2010; Park et al. 2011). As such, the coefficient of correlation between these two measurements can be used to estimate the proportion of methylated CpGs that have undergone mutation (Chuang et al. 2012). As shown in figure 1, both promoter and gene body mCG densities were generally negatively correlated with $\text{CpG}_{\text{O/E}}$. Meanwhile, the absolute values of the Spearman's rank correlation coefficient (ρ) between $\text{CpG}_{\text{O/E}}$ and mCG densities were observed to be generally higher in promoters than in gene bodies (fig. 1), suggesting that

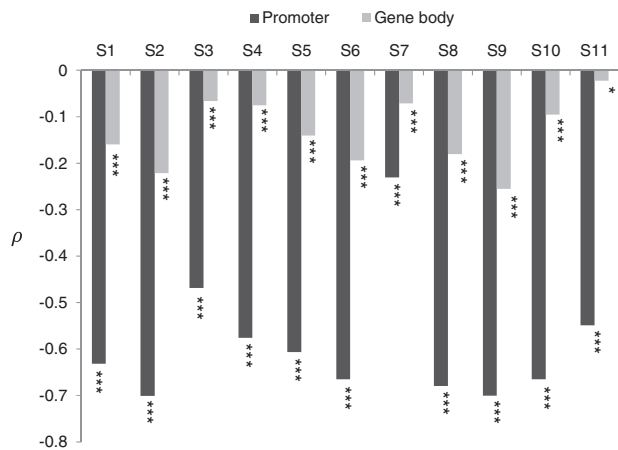


Fig. 1.—Spearman's rank correlation coefficient (ρ) between $CpG_{O/E}$ and mCG density for the promoter and gene body regions in the 11 analyzed methylomes. * $P < 0.05$ and *** $P < 0.001$.

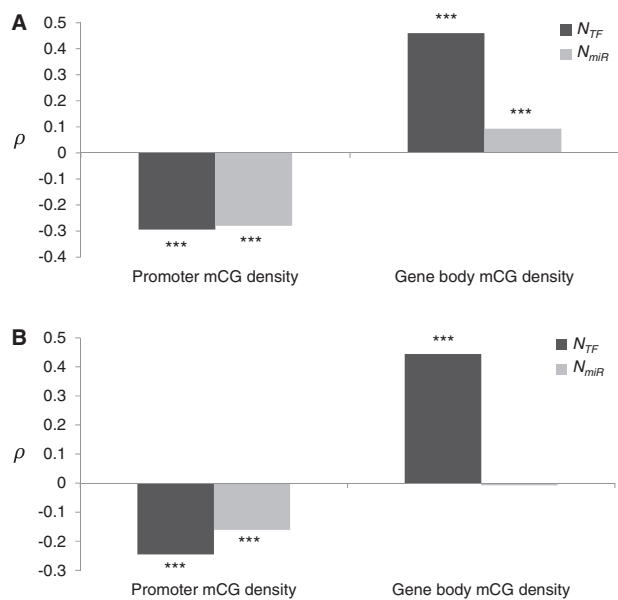


Fig. 2.—Spearman's rank correlation coefficient (ρ) between average promoter/gene body mCG densities for the 11 analyzed samples (table 1) and the 2 regulatory factors (N_{TF} or N_{miR}) (A) before and (B) after controlling for 10 other confounding factors (see text). The analyses were based on 5,418 genes containing all confounding factors examined. *** $P < 0.001$.

promoter regions are more susceptible to the mutagenic effect of DNA methylation than gene bodies. This result also echoes our earlier report that DNA methylation tends to more easily induce C-to-T mutations in the first exons than in internal/last exons (Chuang et al. 2012), and implies differential evolutionary effects of promoter and gene body methylation.

We next examined whether promoter and gene body methylation levels are differentially correlated with the evolutionary rates (d_N , d_S , and d_N/d_S) of their target genes. Here, we examined the level of methylation using the average mCG density across the 11 methylomes in table 1. As shown in table 2, the d_N , d_S , and d_N/d_S values of target genes are all positively correlated with average mCG density in promoters, but negatively correlated with that in gene bodies, for both human-macaque and human-mouse comparisons (all P values $< 10^{-7}$). This result reveals that DNA methylation levels of promoters and gene bodies have opposite effects on the evolutionary rates of their target genes. To control for other confounding factors that may affect the evolutionary rates of protein-coding genes, we used partial correlation analyses (Kim and Yi 2007) to simultaneously control for gene body (or promoter) mCG density, *trans*-regulation (N_{TF} and N_{miR}), and eight other confounding factors as stated in Introduction section: Protein connectivity, gene expression (level and tissue specificity), gene compactness (UTR length, intron length, and intron number), and protein structure (solvent accessibility and disorder content). We found that protein evolutionary rates (d_N and d_N/d_S values) remain positively correlated with promoter methylation level and negatively correlated with gene body methylation level (table 2). The results thus indicate that both promoter and gene body methylation levels are important indicators of protein evolutionary rates, even though the evolutionary effect of gene body methylation is more influenced by the aforementioned confounding factors than that of promoter methylation. Of interest, the partial correlation between gene body methylation level and d_S disappears or even becomes positive after the control (table 2), suggesting that highly methylated regions are subject to strong selective pressure at the protein level despite the enhanced mutation rate (resulting in elevated d_S) in gene bodies. Moreover, the absolute correlation coefficient value for promoter mCG density is greater than that for gene body mCG density after the control (table 2). This indicates that d_N and d_N/d_S are more strongly correlated with promoter mCG density than with gene body mCG density. We thus suggest that the rate of mammalian protein evolution may be influenced more by promoter methylation than by gene body methylation. Furthermore, we also examined the correlation between protein evolutionary rates and gene promoter/gene body mCG density in turn by controlling for each of the abovementioned confounding factors. The trends that protein evolutionary rates are positively correlated with promoter methylation level and negatively correlated with gene body methylation level are generally maintained (supplementary table S2, Supplementary Material online). The absolute correlation coefficient values also indicate that the correlations between protein evolutionary rates and promoter/gene body methylation levels are greatly influenced by N_{TF} and N_{miR} , although all the correlations remain significant (supplementary table S2, Supplementary Material online).

Table 2

Spearman's Rank Correlation Coefficient (ρ) between Promoter (or Gene Body) mCG Density and the Evolutionary Rates (d_N , d_S , and d_N/d_S) for Human–Macaque and Human–Mouse Comparisons Before and After Controlling for Gene Body (or Promoter) mCG Density and Ten Other Confounding Factors

	Before Control			After Control		
	d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
Human–macaque ^a						
Promoter methylation	0.2052***	0.0798***	0.1928***	0.1434***	0.0009	0.1601***
Gene body methylation	−0.0975***	−0.0872***	−0.0803***	−0.0607***	−0.0121	−0.0718***
Human–mouse ^b						
Promoter methylation	0.268***	0.1453***	0.2304***	0.1907***	0.0329*	0.1827***
Gene body methylation	−0.096***	−0.0768***	−0.0793***	−0.0289*	0.025	−0.0452***

NOTE.—The ten confounding factors are N_{TF} , N_{miR} , protein connectivity, expression level, tissue specificity, UTR length, intron length, intron number, solvent accessibility, and disorder content.

^aThe analysis was based on 5,128 human genes and their macaque orthologs.

^bThe analysis was based on 5,357 human genes and their mouse orthologs.

* $P < 0.05$ and *** $P < 0.001$.

Promoter and Gene Body Methylation Levels Exhibit Different Correlation Patterns with TF/miRNA Regulations

We proceeded to investigate the relationships between DNA methylation in promoter/gene body regions and the two *trans*-regulations (N_{TF} and N_{miR}); Spearman's correlation analyses revealed that N_{TF} and N_{miR} both exhibit a negative correlation with promoter mCG density, but a positive correlation with gene body mCG density (all P values $< 10^{-8}$; fig. 2A). Because N_{TF} and N_{miR} may also be correlated with other confounding factors as stated in the last section, we in turn performed partial correlation between one of the promoter/gene body mCG densities and one of the two *trans*-regulations (N_{TF} or N_{miR}) by simultaneously controlling for two other factors and the eight abovementioned confounding factors (fig. 2B). After the control, we found that N_{TF} and N_{miR} still exhibit a negative correlation with promoter mCG density, and N_{TF} remains positively correlated with gene body mCG density, whereas the partial correlation between N_{miR} and gene body mCG density disappears (fig. 2B). This suggests that the correlation between N_{miR} and gene body methylation is greatly influenced by other confounding factors. In addition, the absolute correlation coefficient values also suggest that both promoter and gene body mCG densities are more strongly correlated with N_{TF} than with N_{miR} .

Regarding promoter methylation, it was previously reported that the interactions between TFBSs and their corresponding TFs are sensitive to DNA methylation (Chen, Feng et al. 2011), and TFBSs tend to be hypomethylated to prevent destabilization of TF–DNA interactions (Siegfried et al. 1999; Lister et al. 2009; Straussman et al. 2009). In addition, it has been suggested that promoter methylation and miRNA regulation may complement each other's function, and thus the promoters of genes regulated by more miRNAs tend to have a lower level of DNA methylation (Su et al 2011). These interconnections between biological features may lead to such a

negative correlation between promoter mCG density and these two *trans*-regulations.

Regarding gene body methylation, body-methylated genes have been suggested to be functionally important and represent housekeeping functions (Sarda et al. 2012; Takuno and Gaut 2012). Densely methylated genes tend to evolve more slowly than sparsely methylated genes (Hunt et al. 2010; Lyko et al. 2010; Park et al. 2011; Sarda et al. 2012; Takuno and Gaut 2012). These observations indicate a negative correlation between gene body methylation and protein evolutionary rates (d_N and d_N/d_S). As for TF/miRNA regulations, genes targeted by more TFs or miRNAs tend to be under stronger selective constraints (Cheng et al. 2009; Xia et al. 2009; Wang et al. 2010; Chen, Chuang et al. 2011). Such a trend is broadly maintained throughout metazoa (Chen et al. 2013). Thus, both N_{TF} and N_{miR} are negatively correlated with protein evolutionary rates. In addition, a positive correlation was observed between N_{TF} and N_{miR} (Cui et al. 2007; Chen et al. 2013). These findings thus reveal that these three factors (gene body mCG density, N_{TF} , and N_{miR}) are all negatively correlated with d_N and d_N/d_S , implying a positive correlation between gene body mCG density and *trans*-regulations.

Taken together, the trend that promoter and gene body methylation levels have different correlation patterns with the two *trans*-regulations (N_{TF} and N_{miR}) also reflects the different effects of promoter and gene body methylation on the protein evolutionary rates of target genes (table 2). This also implies the existence of a complicated interaction between these two *trans*-regulations and DNA methylation in different regions.

Promoter Methylation and miRNA Regulation Are Major Factors of Protein Evolutionary Rates

The above analyses indicated that both promoter and gene body methylation levels are important indicators of protein evolutionary rates (d_N and d_N/d_S). We next set out to

determine which biological factor(s) is/are the major factor(s) of protein evolutionary rates. We previously reported that of the ten biological factors associated with evolutionary rates of proteins (N_{TF} , N_{miR} , protein connectivity, expression level, tissue specificity, UTR length, intron length, intron number, solvent accessibility, and disorder content), N_{miR} tends to be the most important factor of d_N and d_N/d_S in mammals (Chen et al. 2013). We therefore estimated the relative importance of these ten factors and promoter/gene body methylation levels in determining the rate of mammalian protein evolution. Using partial correlation analysis, we examined the correlations between protein evolutionary rates and each of these 12 factors in turn by simultaneously controlling for the other 11 factors. As shown in figure 3, we found that promoter mCG density and N_{miR} had the greatest influence on d_N and d_N/d_S for both human–macaque and human–mouse comparisons. Gene body mCG density has a relatively lower effect on d_N and d_N/d_S than promoter mCG density and N_{miR} , but a relatively greater effect than N_{TF} (fig. 3). We also performed a linear regression model, RCVE (see Materials and Methods), to measure the relative effect of these 12 factors in determining d_N and d_N/d_S , and showed the similar trends (supplementary fig. S1, Supplementary Material online).

As discussed above, promoter mCG density and N_{miR} have opposite effects on d_N and d_N/d_S . The finding that they are both the major factors of protein evolutionary rates thus raises the question of whether these two rate determinants have an interaction impact (especially mutual impact) on protein evolutionary rates. This will be discussed in the next section.

Promoter Methylation and miRNA Regulation Exhibit Dependent Effects on Protein Evolutionary Rates in Mammals

Recent studies reported that promoter methylation and miRNA may coregulate their target genes. Changes in promoter methylation may affect miRNA targeting, suggesting a mutual correlation between miRNA-mediated regulation of target genes and miRNA-targeting-specific promoter methylation in brain (Taguchi 2013a, 2013b). Promoter methylation of miRNA-targeted genes has also been suggested to be highly correlated with miRNA seed region features (Taguchi 2013b). For example, the promoters of genes targeted by miR-548 tend to be significantly hypomethylated (Taguchi 2013b). Figure 2 also reveals a significantly negative correlation between mCG density and N_{miR} . Therefore, we are curious about whether there is a dependent effect of promoter mCG density and N_{miR} on protein evolutionary rates. To this end, we conducted a stepwise multiple regression analysis including promoter and gene body mCG densities and the aforementioned ten biological factors to examine the interaction of the effects of these factors on d_N/d_S . The stepwise model selection showed that the coefficient of promoter mCG density– N_{miR} interaction terms significantly deviate

from zero in both human–macaque and human–mouse comparisons (supplementary table S3, Supplementary Material online), suggesting the dependence between promoter methylation and miRNA regulations in determining d_N/d_S .

To further probe the interaction impact of promoter methylation and miRNA regulations on mammalian protein evolution, we divided the human protein-coding genes into five groups of similar size, according to the magnitudes of d_N or d_N/d_S , and calculated the median values of promoter mCG density and N_{miR} for each group of genes. As shown in figure 4, the lower the promoter mCG density, the lower the d_N and d_N/d_S values for both human–macaque and human–mouse comparisons; on the other hand, an opposite correlation was observed between N_{miR} and protein evolutionary rates. In other words, genes with hypomethylated promoters and strong miRNA regulation are subject to stringent selective constraints; in contrast, genes with hypermethylated promoters and weak miRNA regulation are subject to relaxed selective constraints. This result indicates that these two factors have a mutual impact on protein evolutionary rates, also reflecting the above observations that d_N and d_N/d_S values are negatively correlated with N_{miR} but positively correlated with promoter mCG density (table 2).

Potential Caveats

In this study, the single base-resolution methylome data were derived from cultured cells or cells from healthy individuals (table 1). It is possible that methylome data from cancerous cell lines may have introduced bias in the trends we observed. To examine this possibility, we extracted single base-resolution methylome data from breast cancer cells (supplementary table S1, Supplementary Material online) (Hon et al. 2012), and performed the same analyses as described above. We found the similar trends as described above: promoter methylation is positively correlated with protein evolutionary rates but negatively correlated with N_{TF} and N_{miR} , whereas the opposite was observed for gene body methylation (supplementary tables S4 and S5 and fig. S2, Supplementary Material online); the relative importance of these regulatory factors in determining the rate of mammalian protein evolution is as follows: promoter methylation \approx miRNA regulation $>$ gene body methylation $>$ TF regulation (supplementary fig. S3, Supplementary Material online); and promoter methylation and miRNA regulation have a significant dependent effect on protein evolutionary rates (supplementary fig. S4 and table S6, Supplementary Material online). In addition, because rodents have a faster molecular clock than primates (Li 1997; Nekrutenko et al. 2003), it is possible that the observed trends may be biased toward the compared species with different molecular clocks. In this study, we performed all statistical analyses on both comparisons between primate and rodent (i.e., human vs. mouse) and those between primates (i.e., human vs. rhesus macaque), and showed that they had the

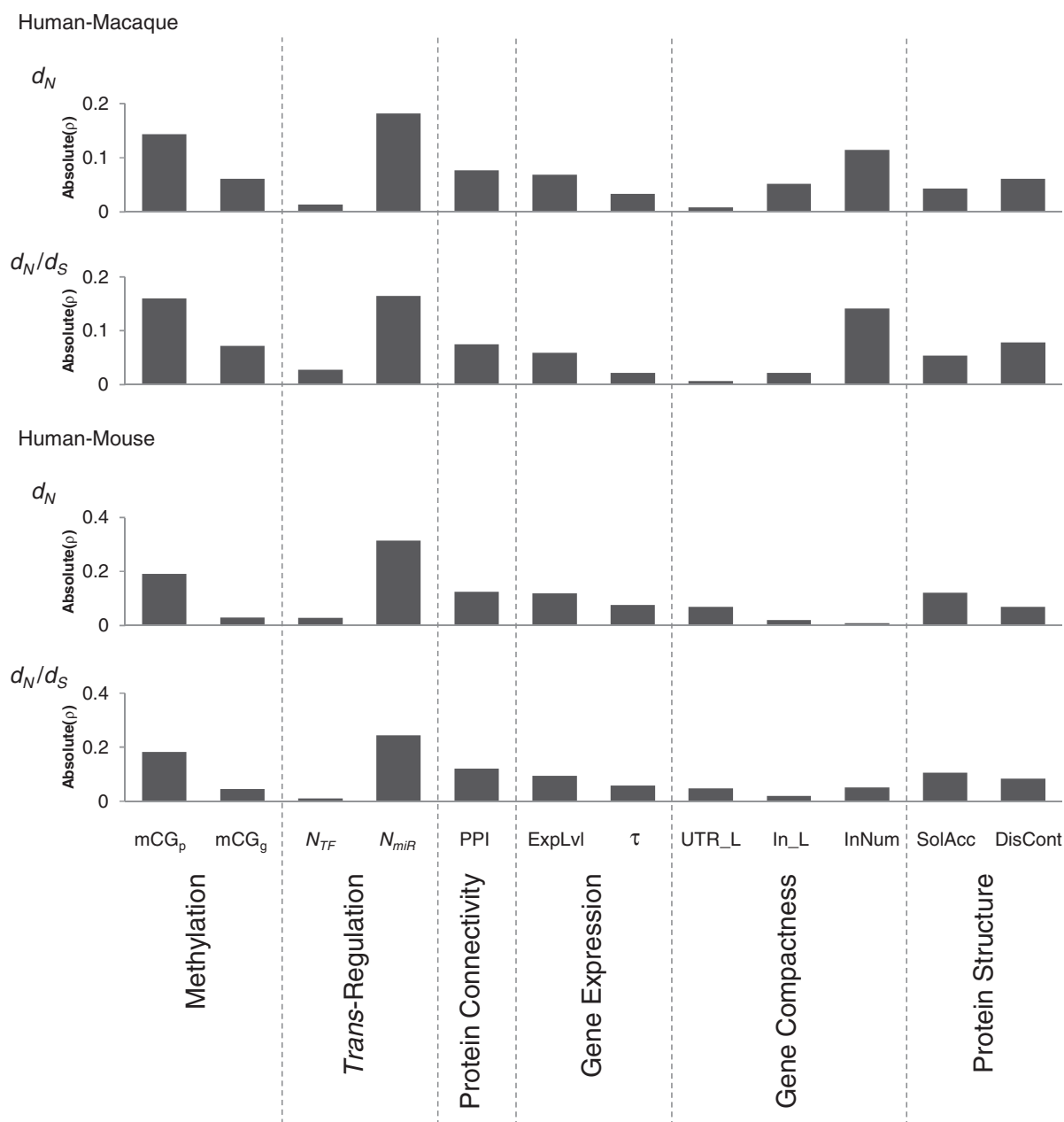


FIG. 3.—Absolute values of Spearman's rank correlation coefficients (ρ) between protein evolutionary rates (d_N and d_N/d_S) and one of the indicated factors (promoter/gene body mCG density [mCG_p and mCG_g], N_{TF} , N_{miR} , protein connectivity [PPI], expression level [ExpLvl], tissue specificity [τ], UTR length [UTR_L], intron length [In_L], intron number [InNum], solvent accessibility [SolAcc], or disorder content [DisCont]), while simultaneously controlling for the other 11 factors, for both human-macaque (5,128 genes) and human-mouse (5,357 genes) orthologs.

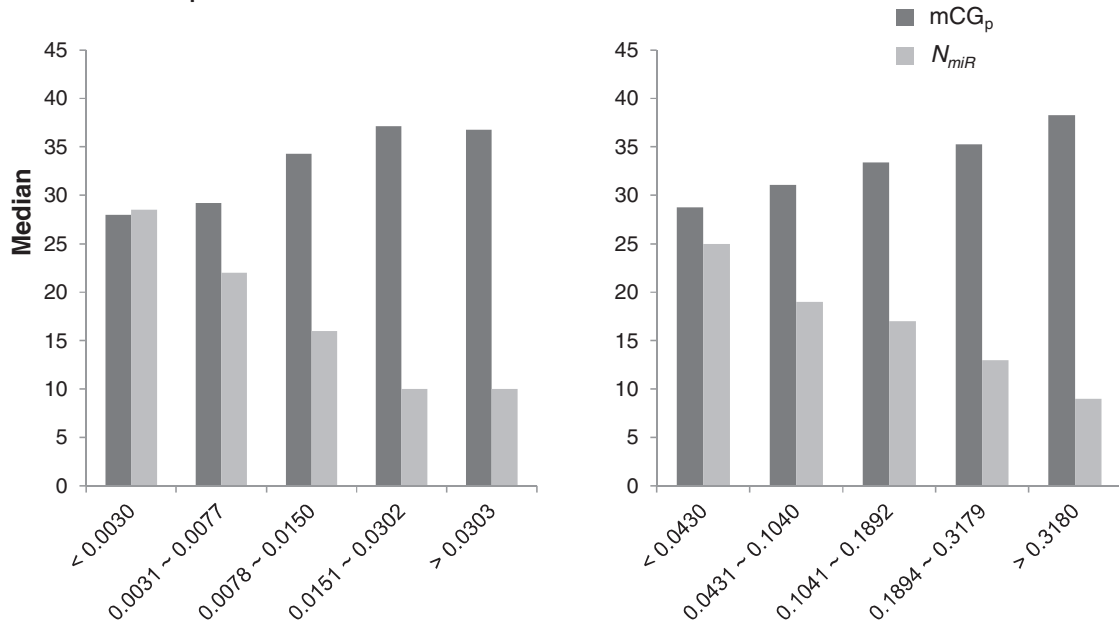
same tendencies. We thus suggest that the observed trends are not biased toward different molecular clocks.

Conclusions

In this study, we examined the impacts of promoter/gene body methylation, TF regulation, and miRNA regulation (which act before, during, and after transcription, respectively)

on the evolutionary rates of the target protein-coding genes. We made several findings. First, promoter and gene body methylation levels exhibit opposite correlation patterns with protein evolutionary rates (d_N and d_N/d_S): the former exhibits a positive correlation with d_N and d_N/d_S , whereas the latter exhibits an inverse correlation with these two evolutionary rates (table 2). On the basis of partial correlation analysis, we emphasize that these correlations are maintained after

Human-Macaque



Human-Mouse

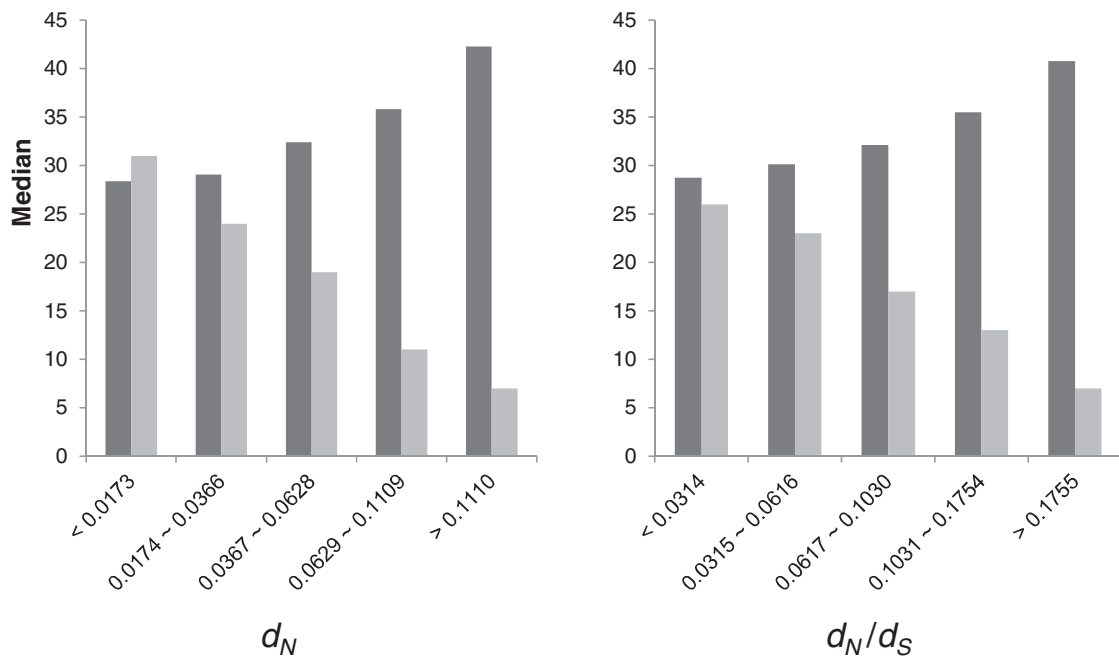


FIG. 4.—The median values of promoter mCG density (mCG_p) and N_{miR} in five groups of human protein-coding genes of similar size (divided according to the magnitudes of d_N and d_N/d_S , from low to high) for both human-macaque (5,128 genes) and human-mouse (5,357 genes) comparisons.

excluding the effect of other confounding factors (table 2), indicating that both promoter and gene body methylation levels are important indicators of protein evolutionary rates of their target genes. We also demonstrated that protein evolutionary rates are more strongly correlated with promoter

methylation level than with gene body methylation level. Second, promoter and gene body methylation levels also exhibit different correlation patterns with the two *trans*-regulations (N_{TF} and N_{miR}); the former is negatively correlated with N_{TF} and N_{miR} , whereas the latter is positively correlated

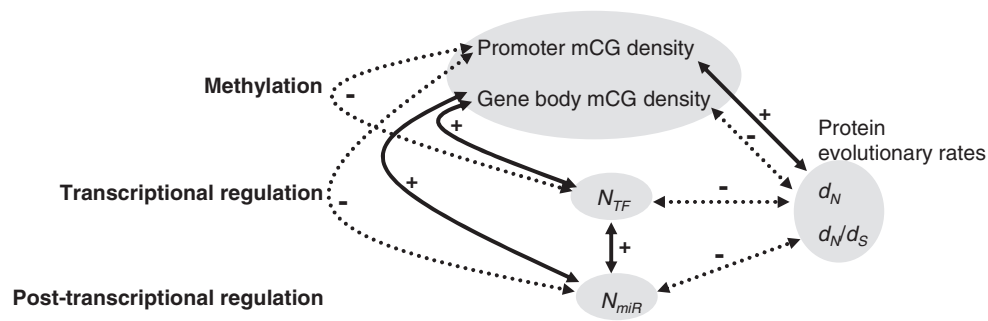


FIG. 5.—The correlations between promoter and gene body mCG densities, N_{TF} , N_{miR} , and protein evolutionary rates (d_N and d_N/d_S). Solid and dotted lines represent positive and negative correlations, respectively.

with these *trans*-regulations (fig. 2). Because N_{TF} and N_{miR} have been previously reported to be positively correlated with each other but inversely correlated with protein evolutionary rates (Chen et al. 2013), the correlations between promoter/gene body mCG density, N_{TF} , N_{miR} , and protein evolutionary rates can be summarized in figure 5. Third, we established that the relative importance of these regulatory factors in determining the protein evolutionary rates is as follows: promoter mCG density $\approx N_{miR} >$ gene body mCG density $> N_{TF}$. We further determined that, among the 12 biological factors considered, promoter methylation and miRNA regulation are generally the major factors in determining d_N and d_N/d_S . Finally, we demonstrated that these two factors have a dependent effect on protein evolutionary rates, and they have a mutual impact on protein evolutionary rates. In summary, our results indicate the complicated effects of natural selection on protein evolution, and the intricate relationships between regulatory systems acting before, during, and after transcription. This study thus increases our understanding of DNA methylation, TF, and miRNA regulations in evolutionary biology.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S6 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

This work was supported by the Genomics Research Center, Academia Sinica, Taiwan, and the National Science Council, Taiwan (under contract NSC102-2621-B-001-003). The authors particularly thank Chia-Ying Chen for assistance in statistical analysis.

Literature Cited

Akgul C, Moulding DA, Edwards SW. 2004. Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol Life Sci*. 61:2189–2199.

Ball MP, et al. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*. 27:361–368.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297.

Bartel DP. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233.

Bernstein BE, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74.

Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res*. 8:1485–1497.

Bloom JD, Drummond DA, Arnold FH, Wilke CO. 2006. Structural determinants of the rate of protein evolution in yeast. *Mol Biol Evol*. 23:1751–1761.

Bogdanovic O, Veenstra GJ. 2009. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* 118:549–565.

Brown CJ, Johnson AK, Daughdrill GW. 2010. Comparing models of evolution for ordered and disordered proteins. *Mol Biol Evol*. 27:609–621.

Chen FC, Chen CJ, Li WH, Chuang TJ. 2010. Gene family size conservation is a good indicator of evolutionary rates. *Mol Biol Evol*. 27:1750–1758.

Chen PY, Feng S, Joo JW, Jacobsen SE, Pellegrini M. 2011. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol*. 12:R62.

Chen SC, Chuang TJ, Li WH. 2011. The relationships among microRNA regulation, intrinsically disordered regions, and other indicators of protein evolutionary rate. *Mol Biol Evol*. 28:2513–2520.

Chen YC, Cheng JH, Tsai ZT, Tsai HK, Chuang TJ. 2013. The impact of trans-regulation on the evolutionary rates of metazoan proteins. *Nucleic Acids Res*. 41:6371–6380.

Cheng C, Bhardwaj N, Gerstein M. 2009. The relationship between the evolution of microRNA targets and the length of their UTRs. *BMC Genomics* 10:431.

Cheng J, Randall AZ, Sweredoski MJ, Baldi P. 2005. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res*. 33:W72–W76.

Chuang TJ, Chen FC. 2014. DNA methylation is associated with an increased level of conservation at nondegenerate nucleotides in mammals. *Mol Biol Evol*. 31:387–396.

Chuang TJ, Chen FC, Chen YZ. 2012. Position-dependent correlations between DNA methylation and the evolutionary rates of mammalian coding exons. *Proc Natl Acad Sci U S A*. 109:15841–15846.

Cui Q, Yu Z, Pan Y, Purisima EO, Wang E. 2007. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. *Biochem Biophys Res Commun*. 352:733–738.

- Dahan O, Gingold H, Pilpel Y. 2011. Regulatory mechanisms and networks couple the different phases of gene expression. *Trends Genet.* 27: 316–322.
- Defossez PA, Stancheva I. 2011. Biological functions of methyl-CpG-binding proteins. *Prog Mol Biol Transl Sci.* 101:377–398.
- Egger G, Liang G, Aparicio A, Jones PA. 2004. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457–463.
- Ehrlich M, Wang RY. 1981. 5-Methylcytosine in eukaryotic DNA. *Science* 212:1350–1357.
- Elnitski L, Jin VX, Farnham PJ, Jones SJ. 2006. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res.* 16:1455–1464.
- Feinberg AP, Tycko B. 2004. The history of cancer epigenetics. *Nat Rev Cancer.* 4:143–153.
- Feng S, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A.* 107: 8689–8694.
- Franzosa EA, Xia Y. 2009. Structural determinants of protein evolution are context-sensitive at the residue level. *Mol Biol Evol.* 26:2387–2395.
- Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19: 92–105.
- Geisen S, Barturen G, Alganza AM, Hackenberg M, Oliver JL. 2014. NGSmethDB: an updated genome resource for high quality, single-cytosine resolution methylomes. *Nucleic Acids Res.* 42:D53–D59.
- Heard E, Clerc P, Avner P. 1997. X-chromosome inactivation in mammals. *Annu Rev Genet.* 31:571–610.
- Hellman A, Chess A. 2007. Gene body-specific methylation on the active X chromosome. *Science* 315:1141–1143.
- Hodges E, et al. 2011. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell.* 44:17–28.
- Hon GC, et al. 2012. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 22:246–258.
- Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet.* 38(Suppl), S20–S24.
- Hunt BG, Brisson JA, Yi SV, Goodisman MA. 2010. Functional conservation of DNA methylation in the pea aphid and the honeybee. *Genome Biol Evol.* 2:719–728.
- Hwang DG, Green P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci U S A.* 101:13994–14001.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 13:484–492.
- Jones PA, Takai D. 2001. The role of DNA methylation in mammalian epigenetics. *Science* 293:1068–1070.
- Kim PM, Sboner A, Xia Y, Gerstein M. 2008. The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol.* 4:179.
- Kim SH, Yi SV. 2007. Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica* 131:151–156.
- Kvikstad EM, Tyekucheva S, Chiaromonte F, Makova KD. 2007. A macaque's-eye view of human insertions and deletions: differences in mechanisms. *PLoS Comput Biol.* 3:1772–1782.
- Larracuent AM, et al. 2008. Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24:114–123.
- Laurent L, et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome Res.* 20:320–331.
- Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol.* 22:1345–1354.
- Li E, Beard C, Jaenisch R. 1993. Role for DNA methylation in genomic imprinting. *Nature* 366:362–365.
- Li W-H. 1997. Rates and patterns of nucleotide substitutions. Sunderland (MA): Sinauer Associates, p.
- Liao BY, Scott NM, Zhang J. 2006. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol.* 23:2072–2080.
- Liao BY, Weng MP, Zhang J. 2010. Impact of extracellularly on the evolutionary rate of mammalian proteins. *Genome Biol Evol.* 2: 39–43.
- Liao BY, Zhang J. 2006. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol Biol Evol.* 23:1119–1128.
- Lin YS, Hsu WL, Hwang JK, Li WH. 2007. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol.* 24:1005–1011.
- Lister R, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322.
- Lyko F, et al. 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* 8:e1000506.
- Marais G, Nouvellet P, Keightley PD, Charlesworth B. 2005. Intron size and exon evolution in *Drosophila*. *Genetics* 170:481–485.
- Marson A, et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134: 521–533.
- Maunakea AK, et al. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253–257.
- Meissner A, et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766–770.
- Mugal CF, Ellegren H. 2011. Substitution rate variation at human CpG sites correlates with non-CpG divergence, methylation level and GC content. *Genome Biol.* 12:R58.
- Nekrutenko A, Chung WY, Li WH. 2003. An evolutionary approach reveals a high protein-coding capacity of the human genome. *Trends Genet.* 19:306–310.
- Park J, et al. 2011. Comparative analyses of DNA methylation and sequence evolution using *Nasonia* genomes. *Mol Biol Evol.* 28: 3345–3354.
- Park J, Xu K, Park T, Yi SV. 2012. What are the determinants of gene expression levels and breadths in the human genome? *Hum Mol Genet.* 21:46–56.
- Park SG, Choi SS. 2010. Expression breadth and expression abundance behave differently in correlations with evolutionary rates. *BMC Evol Biol.* 10:241.
- Plotkin JB, Fraser HB. 2007. Assessing the determinants of evolutionary rates in the presence of noise. *Mol Biol Evol.* 24:1113–1121.
- Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian development. *Science* 293:1089–1093.
- Ruby JG, et al. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* 17:1850–1864.
- Sarda S, Zeng J, Hunt BG, Yi SV. 2012. The evolution of invertebrate gene body methylation. *Mol Biol Evol.* 29:1907–1916.
- Shalgi R, Lieber D, Oren M, Pilpel Y. 2007. Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol.* 3:e131.
- Siegfried Z, et al. 1999. DNA methylation represses transcription in vivo. *Nat Genet.* 22:203–206.
- Straussman R, et al. 2009. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol.* 16: 564–571.
- Su N, Wang Y, Qian M, Deng M. 2010. Combinatorial regulation of transcription factors and microRNAs. *BMC Syst Biol.* 4:150.
- Su Z, Xia J, Zhao Z. 2011. Functional complementation between transcriptional methylation regulation and post-transcriptional microRNA regulation in the human genome. *BMC Genomics* 12(Suppl. 5), S15.

- Szklarczyk D, et al. 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39:D561–D568.
- Taguchi YH. 2013a. Correlation between miRNA-targeted-gene promoter methylation and miRNA regulation of target genes. *F1000Research* 2:21.
- Taguchi YH. 2013b. MicroRNA-mediated regulation of target genes in several brain regions is correlated to both microRNA-targeting-specific promoter methylation and differential microRNA expression. *BioData Min.* 6:11.
- Takuno S, Gaut BS. 2012. Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol Biol Evol.* 29: 219–227.
- Tsang J, Zhu J, van Oudenaarden A. 2007. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell.* 26:753–767.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet.* 10:252–263.
- Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet.* 20: 116–117.
- Wang Y, Franzosa EA, Zhang XS, Xia Y. 2010. Protein evolution in yeast transcription factor subnetworks. *Nucleic Acids Res.* 38: 5959–5969.
- Wu C, et al. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10:R130.
- Xia Y, Franzosa EA, Gerstein MB. 2009. Integrated assessment of genomic correlates of protein evolutionary rate. *PLoS Comput Biol.* 5: e1000413.
- Xiang H, et al. 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol.* 28:516–520.
- Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–659.
- Yang L, Gaut BS. 2011. Factors that contribute to variation in evolutionary rate among *Arabidopsis* genes. *Mol Biol Evol.* 28:2359–2369.
- Yu X, Lin J, Zack DJ, Mendell JT, Qian J. 2008. Analysis of regulatory network topology reveals functionally distinct classes of microRNAs. *Nucleic Acids Res.* 36:6494–6503.
- Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.
- Zeng J, et al. 2012. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am J Hum Genet.* 91:455–465.
- Zhou T, Drummond DA, Wilke CO. 2008. Contact density affects protein evolutionary rate from bacteria to animals. *J Mol Evol.* 66:395–404.

Associate editor: Hidemi Watanabe