FEBS Letters 587 (2013) 2284–2287

journal homepage: www.FEBSLetters.org

Complex-forming proteins escape the robust regulations of miRNA in human

Jyotirmoy Das, Sandip Chakraborty, Soumita Podder, Tapash Chandra Ghosh *

Bioinformatics Centre, Bose Institute, P 1/12, C.I.T. Scheme VII M, Kolkata 700 054, India

ARTICLE INFO

Article history:
Received 26 March 2013
Revised 21 May 2013
Accepted 29 May 2013
Available online 10 June 2013

Edited by Takashi Gojobori

Keywords:
miRNA hits
Protein complexity
Gene duplication
Evolutionary rate
Protein complex association number
Gene essentiality

ABSTRACT

Most proteins carry out their functions by participating in protein complexes. Recently, miRNAs were identified as promising post-transcriptional regulators that influence a large proportion of genes in higher eukaryotes. We aim to understand the role of miRNAs in the regulation of human proteins that are present in protein complexes. Here, we show that robust regulation by miRNA is absent in human complex-forming proteins. Moreover, the numbers of miRNA hits cannot direct the evolutionary fate of complex-forming proteins independently. However, the duplicated complex-forming proteins having a severe effect on organismal fitness are profoundly targeted by miRNA, probably to reduce the chances of dosage imbalance.

1. Introduction

MicroRNAs (miRNAs) are presently recognized as novel agents exercising post-transcriptional control in most eukaryotic genomes. Recently, miRNA genes and their functional roles have been predicted through the high-throughput experimentation methods [1–4]. Studies on miRNA regulation of Protein–Protein Interaction Networks (PPIN) have revealed that most of the miRNAs target highly connected nodes in a PPIN [5] and it has also been demonstrated that miRNA regulation is more important for the inter-modular hubs having lower clustering coefficient than the intra-modular ones with higher clustering coefficient [5]. Moreover, it has been established that the highly clustered modules are generally represented as the protein complex association [6] and later on it becomes more prominent from the study of Liang and Li [5] that the intra-modular hubs are the subunits of the protein complex and they contain no miRNA target sites, while inter-modular hubs contain several miRNA target sites. Recently, a contradictory report [7] has been published which evidently showed that single miRNA or co-expressed miRNAs frequently target several components of protein complexes.

Afterwards Goh et al. [8] also demonstrated that miRNAs with widely different expression profiles have some impacts on different protein complex formation to govern the biological processes. Now, one protein can participate in a number of protein complexes. Hence, we designate the number of complexes in which a protein participates, as protein complex association number. Earlier, it was reported that proteins that are linked with multiple complexes tend to be more essential than the members of a single protein complex [9]. This indicates that protein complex association number has a vital attribute on the biological system. Here, we test the hypothesis that different complex proteins having distinct coding selective constraints due to their different complex forming ability also facilitate discrete patterns of miRNA regulation. Indeed, we have found that proteins with higher complex association number evade robust miRNA regulation to maintain their co-expressivity in the complex unit. Moreover, miRNA is also incapable to regulate the evolutionary rate of complex-forming proteins independently. The profound action of miRNA only observed for the duplicated proteins in the complex association. Indeed, we have found that proteins with higher complex association number evade robust miRNA regulation to maintain their co-expressivity in the complex unit. Moreover, miRNA is also incapable to regulate the evolutionary rate of complex-forming proteins independently. The profound action of miRNA only observed for the duplicated proteins in the complex association.

* Corresponding author. Fax: +91 33 2355 3866.
E-mail address: tapash@jcbose.ac.in (T.C. Ghosh).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.05.062

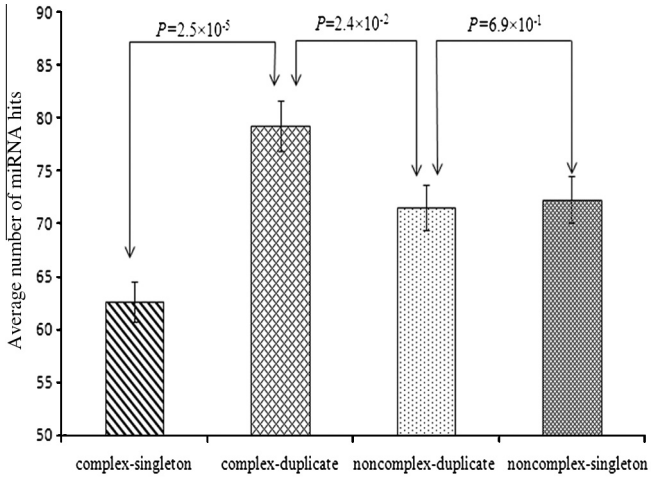

Figure 1. Comparison of average miRNA hits between singleton and duplicated proteins present in complex and non-complex proteins. *P* value denotes the significant level of differences and the error bars represent the 5% error in data.

2. Materials and methods

2.1. Dataset preparation for protein complex and miRNA analysis

A total of 1,610 human (*Homo sapiens*) protein complex data were retrieved from the CORUM database, a manually curated repository of experimentally characterized protein complexes (release February, 2012) (http://mips.helmholtz-muenchen.de/genre/proj/corum) [10], (Supplementary Table S1). We, thereby, obtained protein complexes ranging from 1 to 55 present in the dataset. miRNA dataset for human was downloaded by fetching TargetScan database (http://www.targetscan.org/) [11]. TargetScan is used for its reported accuracy and advantages of seed-pairing mechanisms in miRNA (which is required for mRNA-miRNA bindings) over other miRNA databases [12]. Furthermore, to increase the reliability of our results, we only considered the miRNAs whose target sites are conserved across most mammals (as defined by TargetScan). Thus, a total of 331,470 hits were downloaded from the database. After removing redundant interactions, a total of 4,695 miRNAs were collected for further analysis (Supplementary Table S2). Next, we computed the number of unique miRNA hits per gene, which ranges from 1 to 425.

Next, to confirm our hypothesis, we tested our dataset with experimentally validated data from miRWalk database ((http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/) [13]. A total of 24,442 miRNA hits were downloaded from the database and miRNA hits per gene were calculated.

2.2. Collection of gene expression data

The gene expression information in human was obtained from Microarray data, GNF Gene Atlas (http://biogps.gnf.org) [14]. Genes represented by more than a single probe set were discarded in order to avoid re-counting. Different genes sharing the same probe sets were also excluded (with the exception of splice variants). A single and unique probe set therefore represents each of the 19,860 remaining human genes. Expression values (expression signal intensities in each tissue) of genes were averaged over 19 tissues. We mapped these expression datasets with our existing dataset and a total of 2,394 genes were collected for analysis.

2.3. Evolutionary rate estimation

Protein evolutionary rates [*dN*/*dS*] data for human using 1:1 orthology relationship to Chimpanzee (*Pan troglodytes*) were downloaded from ENSEMBL database (ver. 67) [15]. Next, we mapped the evolutionary rate data to genes and gene encoded proteins with available miRNA hits and protein complex association numbers for further analysis.

2.4. Essential gene dataset collection

The Essential gene dataset was downloaded from Online Gene Essentiality database (OGEE build: 304) [16] (http://ogeedb.embl.de). This set contains 107,071 essential genes from 24 organisms (8 Eukaryotic and 16 Prokaryotic), of which 24,076 are from human. After removing the redundancy, 20,684 genes were obtained and matched with our dataset and finally we collected 4,340 essential genes for analysis in our dataset.

2.5. Identification of paralogs

Human paralogs were obtained from the ENSEMBL database (ver. 67). For identifying them, ENSEMBL follows the steps mainly based on the construction of the gene tree reconciled with the species tree formed by the cluster of aligned sequences obtained from BLASTP [15]. Finally, we considered true paralogs by using 40% similarity and alignable region is >80% between the two sequences [17]. Thus, a total of 12,791 duplicated proteins in humans were identified. Among these 12,791 human paralogs, only 1,992 genes having paralogs are present both in the miRNA targeted genes as well as protein complexes.

2.6. Statistical analysis

All statistical analyses except partial correlations were performed using SPSS ver. 20. TANAGRA (ver.1.4) [17,18] was used to determine the partial correlation. Spearman rank correlation coefficient (ρ) was used throughout the manuscript. To find the difference between two datasets, we performed Mann-Whitney U test.

3. Results

3.1. Effect of miRNA on complex-forming proteins

We studied the global correlation between protein complex association number and the number of miRNA target-site types at the 3′ UTRs of the gene encoding the protein. Interestingly, we noticed a significant negative correlation between them ($\rho_{\text{number of miRNA hits vs protein complex association number}} = -0.102$, $P = 2.0 \times 10^{-2}$, $N = 513$ [TargetScan]; $\rho_{\text{number of miRNA hits vs protein complex number}} = -0.191$, $P = 3.8 \times 10^{-2}$, $N = 118$ [miRWalk]) (Supplementary Fig. 1A) which suggests that the proteins with low complex association number could be more targeted by miRNA compared to the proteins with high complex association number. Although the reported effect of miRNA on protein complex is significant, but of a very small magnitude which implies that protein complex association number is not tightly regulated by miRNA target. It was evident that highly sensitive nature of multi-protein complex towards their expression variation [19] may restrain the protein complex to be regulated by miRNA. Likewise, proteins that are shared across different complexes require a highly orchestrate expression pattern, frequent targeting by miRNA may perturb their expression coordination which also can damage their functional activity. Thus proteins, present in a large number of protein complexes turn down the robust regulation of miRNA.

Table 1
Partial correlation of evolutionary rates with number of miRNA hits, protein complex association number and expression level.

Factors[a]	Partial correlation[b] for evolutionary rates	Level of significance[c]
Number of miRNA hits	0.056 (controlling expression level and protein complex association number)	2.5×10^{-1}
Protein complex association number	−0.12 (controlling expression level and number of miRNA hits)	1×10^{-2}
Expression level	−0.19 (controlling protein complex association number and number of miRNA hits)	8×10^{-5}

Note:
[a] Factors are the parameters which are the correlates of evolutionary rates.
[b] Partial correlation column shows the correlation coefficient of the factors.
[c] Level of significance columns lists the p-value of the partial correlation tests.

Previous findings that proteins attached with multiple protein complexes tend to evolve slowly than those with fewer or no complex assembly [20]. In our study, negative correlation between protein complex association number and miRNA regulation thus need clarification because genes with a higher level of miRNA regulation were reported to evolve more slowly [21]. To unravel the related effects of miRNA and protein complex association number, we divided proteins into two groups, complex-forming proteins and non-complex proteins, according to whether they are present in complex assembly or not. For complex-forming proteins, we observed a positive correlation ($\rho_{dN/dS \text{ vs number of miRNA hits}} = 0.102$, $P = 3.5 \times 10^{-2}$, $N = 428$ [TargetScan]; $\rho_{dN/dS \text{ vs number of miRNA hits}} = 0.177$, $P = 6.8 \times 10^{-2}$, $N = 108$ [miRWalk]) (Supplementary Fig. 1B) between miRNA hits and evolutionary rates (dN/dS) which contradicts the notion that genes with higher miRNA target sites evolve more slowly than non-targeted genes [21]. Interestingly, the number of miRNA hits hold a negative correlation with expression level ($\rho_{expression\ level \text{ vs number of miRNA hits}} = -0.216$, $P = 7.0 \times 10^{-6}$, $N = 428$ [TargetScan]; $\rho_{expression\ level \text{ vs number of miRNA hits}} = -0.194$, $P = 4.4 \times 10^{-2}$, $N = 108$ [miRWalk]) (Supplementary Fig. 1C) since expression level is a well-known negative correlate of evolutionary rate [22]. However, for non-complex proteins (which do not take part in protein-complex assemblies), we obtained a negative correlation between miRNA hits and evolutionary rates ($\rho_{dN/dS \text{ vs number of miRNA hits}} = -0.131$, $P = 1.0 \times 10^{-6}$, $N = 2{,}274$ [TargetScan]; $\rho_{dN/dS \text{ vs number of miRNA hits}} = -0.032$, $P = 6.4 \times 10^{-1}$, $N = 209$ [miRWalk]) (Supplementary Fig. 1D). It seems the impact of protein complex association number may obscure the effect of miRNA on protein evolutionary rates. Moreover, we also noticed a significant negative correlation ($\rho_{dN/dS \text{ vs expression level}} = -0.142$, $P = 1.0 \times 10^{-6}$, $N = 2{,}393$ [TargetScan]; $\rho_{dN/dS \text{ vs expression level}} = -0.171$, $P = 2.0 \times 10^{-3}$, N = 311 [miRWalk]) (Supplementary Fig. 1E) between expression level and protein evolutionary rate for the whole dataset including both the complex and non-complex proteins.

Thus, we performed a partial correlation analysis with all potential correlates of evolutionary rate (expression level, protein complex association number, miRNA hits) by considering complex and non-complex proteins together and found that protein evolutionary rates are negatively associated with protein complex association number when miRNA hits and expression level are controlled, but the correlation between miRNA hits and rate of protein evolution has disappeared at the 95% level of confidence when the protein complex association numbers and expression level were controlled (Table 1). This result suggests that for complex-forming proteins, miRNA does not play any significant role in guiding protein evolutionary rates.

3.2. Effect of miRNA regulation on complex-forming proteins involved in gene duplication

The recent developments in the analysis of protein complexes suggest that the internal subunit arrangement in complexes is crucial for their more detailed functional understanding. If any subunit of the protein complex loses its function, it is necessary for the complex to keep duplicate copies of that gene to restore the lost function [23,24]. Concomitantly, it was also reported that duplicated genes are more enriched by the miRNA target sites than the singleton genes to maintain the gene expression level [12]. Thus, we have investigated miRNA regulation of complex-forming proteins in duplicated genes. We observed that only 33.45% complex-forming proteins are involved in gene duplication whereas for non-complex proteins it is 66.55% (Two sided Fisher's exact test: $P = 1.0 \times 10^{-6}$). These results are in agreement with the findings of Papp et al. (2003) [25] that duplicated genes are rarely act as a complex-forming subunit to prevent themselves from dosage imbalance. We have also noticed that 36.18% protein is essential among duplicated complex-forming proteins whereas, in singleton complex-forming proteins, it is only 28.94% (Two sided Fisher's exact test: $P = 2.0 \times 10^{-3}$). However, the proportion of essential genes is statistically similar in both duplicated and singleton genes for non-complex proteins (duplicated-essential = 19.7%; singleton-essential = 18.9%). So, it would be interesting to investigate the role of miRNA if any, in regulating the paralog numbers of duplicated complex-forming proteins. Indeed we have observed that they are more targeted by miRNA when compared with singleton complex-forming proteins and both the duplicate and singleton non-complex proteins. (Fig. 1). Moreover, among the duplicated complex-forming proteins, we have found that the essential proteins hold a significant positive correlation between miRNA hits and paralog numbers ($\rho_{\text{paralog numbers vs number of miRNA hits}} = 0.146$, $P = 4.2 \times 10^{-2}$, $N = 194$ [TargetScan]; $\rho_{\text{paralog numbers vs number of miRNA hits}} = 0.490$, $P = 3.3 \times 10^{-3}$, $N = 19$ [miRWalk]) (Supplementary Fig. 1F) whereas it bears an insignificant correlation ($P = 4.9 \times 10^{-1}$) for the non-essential parts of complex-forming proteins. This result highlights that paralogous copies of the complex-forming proteins related to the fitness effect of the organism are strongly regulated by miRNA to circumvent the risk of dosage imbalance.

4. Discussion

miRNAs are believed to regulate different types of genes through post-transcriptional gene regulation and have the potential to silence gene expression. Our analyses revealed an interesting scenario of miRNA regulations for complex-forming proteins. Though a negative correlation achieved between protein complex association number and miRNA targets, the small magnitude of correlation suggested that the overall effect of miRNAs on protein complex association number is intriguingly marginal. One possible scenario is that the subunits of a protein complex show correlated patterns of expression over a time course [26] for which they have relatively less pressure to achieve expression diversity. Therefore, the members of the same protein complex are mostly targeted by single types of miRNA or several co-expressed miRNA to retain the co-expressivity of the subunits [7]. If one member of the protein complex is attached with a large number of complexes, then it also requires all the complexes with which it attaches will be targeted by same miRNA. But this incidence is biologically not favor-

able. There may be some other regulatory machinery, except miRNA, responsible to modulate the protein complex activity.

From an evolutionary point of view it is also speculated that proteins with higher complex association number are targeted by more miRNAs since both the protein complex association number and number of miRNA targets are negatively correlated with evolutionary rates. However, the independent effect of miRNA on evolutionary rate was not observed in our case. One plausible reason behind this is the co-expression nature of complex-forming proteins which could be impaired to a large extent, if they are targeted by miRNA. Alternatively, this facet could be explained in this way that complex-forming proteins are generally highly expressed since a positive correlation exhibits between expression level and protein complex association number ($\rho_{\text{expression level vs protein complex association number}} = 0.149$, $P = 2.0 \times 10^{-3}$ [TargetScan]; $\rho_{\text{expression level vs protein complex association numbers}} = 0.204$, $P = 3.4 \times 10^{-2}$, $N = 107$ [miRWalk]) (Supplementary Fig. 1G), whereas several animal miRNAs have been reported to regulate their target mRNAs by aiding mRNA cleavage [27]. So, to prevent them from rapid mRNA decay for retaining high expression level, it is also a prerequisite for complex-forming proteins to avoid miRNA target. For non-complex proteins, miRNA shows negative correlation with evolutionary rates as observed earlier [21]. The non-complex proteins which are targeted by more miRNAs and evolve slowly than the rest may be the subset of pleiotropic genes that requires complex regulation of miRNA [28].

An interesting scenario was observed for the duplicate proteins which take part in the protein complex. The duplicates are bound to be controlled by miRNA repression to synchronize the expression variation among their paralogous copies since duplication of a subunit in a protein complex might cause dosage imbalance if rapid sub-functionalization or neo-functionalization does not occur to the newly arisen genes [29]. In case of essential duplicates, they always want to keep a backup copy to withstand the harmful mutations, so the tendency of neo-functionalization is much lower for them [30]. The strong positive correlation between the paralogs number and miRNA hits of the essential duplicates ensures the urge of miRNA regulation to control their paralogs number. Whereas for non-essestial duplicates, the chances of neo-functionalization may lower the possibility of dosage imbalance and thus miRNA regulations become relax for them. The knowledge gleaned from our study is important to apprehend the pattern and basis of miRNA regulation for protein complex in an evolutionary landscape.

Acknowledgements

J.D. is a recipient of DST-INSPIRE (IF-110696), Government of India. Financial help from the Department of Biotechnology, Government of India (No. BT/PR692/BID/7/369/2011) is duly acknowledged. We are also thankful to our anonymous reviewer for his/her valuable suggestions and critical review.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2013.05.062.

References

[1] Bonci, D. et al. (2008) The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. Nat. Med. 14, 1271–1277.
[2] Mendell, J.T. (2008) MiRiad roles for the miR-17-92 cluster in development and disease. Cell 133 (2), 217–222.
[3] Nakada, C. et al. (2008) Genome-wide microRNA expression profiling in renal cell carcinoma: significant down-regulation of miR-141 and miR-200c. J. Pathol. 216 (4), 418–427.
[4] Ambros, V. (2008) The evolution of our thinking about microRNAs. Nat. Med. 14, 1036–1040.
[5] Liang, H. and Li, W.-H. (2007) MicroRNA regulation of human protein–protein interaction network. RNA 13 (9), 1402–1408.
[6] Spirin, V. and Mirny, L.A. (2003) Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci. USA 100, 12123–12128.
[7] Sass, S. et al. (2011) MicroRNAs coordinately regulate protein complexes. BMC Syst. Biol. 5, 136.
[8] Bin Goh, W., Oikawa, H., Sng, J., Sergot, M. and Wong, L. (2012) The role of miRNAs in complex formation and control. Bioinformatics 28, 453–456.
[9] Pereira-Leal, J., Levy, E. and Teichmann, S. (2006) The origins and evolution of functional modules: lessons from protein complexes. Philos. Trans. R. Soc. Bio. 361, 507–517.
[10] Ruepp, A. et al. (2010) CORUM: the comprehensive resource of mammalian protein complexes-2009. Nucleic Acids Res. 38, D497–D501.
[11] Friedman, R., Farh, K., Burge, C. and Bartel, D. (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105.
[12] Li, J., Musso, G. and Zhang, Z. (2008) Preferential regulation of duplicated genes by microRNAs in mammals. Genome Biol. 9, R132.
[13] Dweep, H., Sticht, C., Pandey, P. and Gretz, N. (2011) MiRWalk – Database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J. Biomed. Inform. 44, 839–847.
[14] Su, A. et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA 101, 6062–6067.
[15] Flicek, P. et al. (2011) Ensemble 2011. Nucleic Acids Res. 39, D800–D806.
[16] Chen, W., Minguez, P., Lercher, M. and Bork, P. (2012) OGEE: an online gene essentiality database. Nucleic Acids Res. 40, D901–D906.
[17] Bhattacharya, T. and Ghosh, T. (2010) Protein connectivity and protein complexity promotes human gene duplicability in a mutually exclusive manner. DNA Res. 17, 261–270.
[18] Marcoulides, G. (2004) The elements of statistical learning: data mining, inference and prediction. Struct. Equation Model. 11, 150–151.
[19] Fraser, H., Hirsh, A., Giaever, G., Kumm, J. and Eisen, M. (2004) Noise minimization in eukaryotic gene expression. PLoS Biol. 2, 834–838.
[20] Chakraborty, S., Kahali, B. and Ghosh, T. (2010) Protein complex forming ability is favored over the features of interacting partners in determining the evolutionary rates of proteins in the yeast protein–protein interaction networks. BMC Syst. Biol. 4, 155.
[21] Cheng, C., Bhardwaj, N. and Gerstein, M. (2009) The relationship between the evolution of microRNA targets and the length of their UTRs. BMC Genomics 10, 431.
[22] Pal, C., Papp, B. and Lercher, M. (2006) An integrated view of protein evolution. Nat. Rev. Genet. 7, 337–348.
[23] Dziembowski, A. and Seraphin, B. (2004) Recent developments in the analysis of protein complexes. FEBS Lett. 556, 1–6.
[24] Lin, Y., Hwang, J. and Li, W. (2007) Protein complexity, gene duplicability and gene dispensability in the yeast genome. Gene 387, 109–117.
[25] Papp, B., Pal, C. and Hurst, L. (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424, 194–197.
[26] Jansen, R., Greenbaum, D. and Gerstein, M. (2002) Relating whole-genome expression data with protein–protein interactions. Genome Res. 12, 37–46.
[27] Mok, Y., Park, S. and Choi, S. (2012) Comparative analysis of the structural and expressional parameters of microRNA target genes. Gene 497, 103–109.
[28] Chen, S., Chuang, T. and Li, W. (2011) The relationships among MicroRNA regulation, intrinsically disordered regions, and other indicators of protein evolutionary rate. Mol. Biol. Evol. 28, 2513–2520.
[29] Makova, K.D. and Li, W.H. (2003) Divergence in the spatial pattern of gene expression between human duplicate genes. Genome Res. 13, 1638–1645.
[30] Gu, Z., Steinmetz, L., Gu, X., Scharfe, C., Davis, R. and Li, W. (2003) Role of duplicate genes in genetic robustness against null mutations. Nature 421, 63–66.