Universality splitting in distribution of number of miRNA co-targets

Mahashweta Basu · Nitai P. Bhattacharyya ·
P. K. Mohanty

Received: 11 November 2013 / Revised: 16 January 2014 / Accepted: 17 January 2014 / Published online: 18 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract In a recent work (Basu et al., in EPL 105:28007, 2014) it was pointed out that the link-weight distribution of microRNA co-target network of a wide class of species are universal up to scaling. The number cell types, widely accepted as a measure of complexity, turns out to be proportional to these scale-factor. In this article we discuss additional universal features of these networks and show that, this universality splits if one considers distribution of number of common targets of three or more number of microRNAs. These distributions for different species can be collapsed onto two distinct set of universal functions, revealing the fact that the species which appeared in early evolution have different complexity measure compared to those appeared late.

Keywords MicroRNA · Network · Co-regulation · Universality

MicroRNAs (miRNAs) are small non-coding single stranded RNAs of about 22 nucleotides long and act as a secondary regulator of gene expression (Liu 2008; Ambros 2004; Bartel 2009). They are transcribed from the DNA from either inter or intra genomic region (Bartel 2004; Clarke et al. 2006) and bind to the 3' untranslated regions (UTR) of messenger RNA (mRNA) via base pairing leading to their degradation or inhibition of the gene expression (Farh et al. 2005). In effect, the respective proteins are

produced less compared to the situation when miRNAs are absent (Clarke et al. 2006; Dong et al. 2013). Even being secondary regulators (transcription factors being the primary ones) miRNAs are seen as potential therapeutic targets for treatment of cancer (Wang and Wu 2009; Wu et al. 2013) and other disease (Hansen and Obrietan 2013; Hu et al. 2010). There are large number of databases (miRBase; Miranda et al. 2006) which predict short genomic sequences which might be acting as a miRNAs. Substantial effort (Bartel 2009; Lewis et al. 2003; miRDB; TargetScan; miRWalk; MicroCosm Targets Version v5) has also been given in predicting the targets of these miRNAs.

A web resource MicroCosm Targets (MicroCosm Targets Version v5) provides computationally predicted targets of microRNAs of several species, where the miRNA sequences are obtained from the miRBase Sequence database (miRBase). There are about 851 miRNAs in *Homo sapiens* which averagely target about 950 mRNAs out of the total 35864 transcripts. Experimental validation of these predictions are, however, largely lacking. It is believed (Yu et al. 2006; Lu et al. 2008; Boross et al. 2009) that specific biological functions and processes are possibly carried out by groups of miRNAs, through noise reduction, than individual ones. It is thus important to look for combinatorial regulation (Ivanovska and Cleary 2008). Recent studies on miRNA (Mookherjee et al. 2009; Xu et al. 2011) co-target network for *H. sapiens* reveal that the groups of miRNA together regulate the functions of protein coding genes acting in combination with their targets. The miRNA co-target network is formed by joining a pair of miRNA by link and associating the number of co-targets as the weight of the link. Community structures of these densely weighted networks are then obtained using the modularization algorithms in Refs. (Mookherjee et al. 2009; Basu et al. 2011).

M. Basu · P. K. Mohanty (✉)
Condensed Matter Physics Division, Saha Institute of Nuclear
Physics, 1/AF Bidhan Nagar, Kolkata 700064, India
e-mail: pk.mohanty@saha.ac.in

N. P. Bhattacharyya
Crystallography and Molecular Biology Division, Saha Institute
of Nuclear Physics, 1/AF Bidhan Nagar, Kolkata 700064, India

Surprisingly the weight distribution of these networks show amazing universal features, which extends over many species classes, families and genera (Basu et al. 2014). In other words the distribution function $P(w)$ of the number of co-targets w is found to be a scaled form of an universal function; the species are characterized by a unique scale factor λ . It was also observed that the scale factor is proportional to the number of cell types of the respective species, and thus, can be considered as a measure of the complexity. A simple random-target model, where miRNAs of a species target a fixed number of mRNAs, could produce the universal function reasonably well. Thus, animal specificity is not resolved through pair-wise co-targets and from these networks one expects to obtain generic functions common to species of wide range.

In this article, we propose that, if one consider number of common targets of three or more miRNAs of a given species, the distribution function show *two* distinct groups of animals. We also observe that this sub-groups are consistent with the natural partition of the species into two groups obtained from the bi-model distribution of number of miRNAs.

For completeness, first we discuss how to obtain the miRNA co-target distributions for different species. We consider all the species whose miRNA targets have been predicted by MicroCosm Targets (MicroCosm Targets Version). The complete list of species, along with the number of miRNAs M and mRNAs N are given in Table 1. Let us denote the miRNAs of a species as $\{m_1, m_2, \dots, m_M\}$. To construct the co-target of miRNA multiplets of size k , we first take k distinct miRNAs $\{m_{i_1}, m_{i_2}, \dots, m_{i_k}\}$ from the set of M . Thus in total, there are C_k^M multiplets. Then from the target database, we find the targets which are common to the first two miRNAs m_{i_1} and m_{i_2} and denote the number of common targets as w_2 . These w_2 targets are then compared with the targets of next miRNA m_{i_3} to find w_3 which is now the number of common targets of the triplet $\{m_{i_1}, m_{i_2}, m_{i_3}\}$. This process is iterated until one obtains w_k . The nonzero w_k s obtained through this process are now considered for obtaining the distribution function $P_k(w) \equiv P(w_k)$. Clearly, for $k = 2$, the distribution $P_2(w)$ is equivalent to $P(w)$, the link-weight distribution of miRNA co-target networks discussed in Ref. (Basu et al. 2014). In Fig. 1a we have shown $P_2(w)$ for four different species, namely *H. sapiens* (Human), *Bos taurus* (bovine), *Xenopus tropicalis* (western clawed frog) and *C. elegans*. Note that, the number of common genes targeted by any miRNA pair does not depend on the total number miRNAs of the species, however depends on the number of genes N . To understand the shift in peak position and the change in width (or variance) of $P_2(w)$, a simple model has been proposed in Ref. (Basu et al. 2014). It was shown that the peak position is proportional to square of the

average number of genes targeted by the miRNAs and inversely proportional to N . Interestingly, $P_2(w)$ s for different species could be collapsed onto each other by aligning their peaks at origin and rescaling the x - and y - axis suitably by a single factor λ which is species dependent. It was observed that the scale factors are proportional to the number of cell-types and serve as an equivalent measure of complexity of the species.

In the following we consider certain other universal features of the miRNA co-target network. Mookherjee et al. (2009) have constructed the miRNA co-target network of *H. sapiens* from a adjacency matrix W with elements w_{ij} same as the number of co-targets of miRNA pair i and j . The human miRNA co-target network was found to be fully connected with large variation in link-weights - some as large as 1282 and as small as 1. They argued that links with small weights are rather un-important and the network can be made simple by erasing links whose weights are smaller than a pre-specified value q . In this case, the network breaks into N_q number of disconnected components; N_q being a non-decreasing function of q with $N_0 = 1$ (as all miRNAs are connected at $q = 0$). The variation of N_q with q for *H. sapiens* and other three species are shown in Fig. 1b. Let us define density of components

$$v_q = \frac{N_q - 1}{M - 1}. \quad (1)$$

Evidently, as q is increased, v_q picks up a non-zero value at some $q = q_c$ (when the network starts breaking up). Mookherjee et al. (2009) have claimed that the optimum network, that does not contain irrelevant links nor looses the network functionality, occurs at a value of $q^* = 103$ where the breaking rate $\frac{dv_q}{dq}$ is maximum. The largest component at $q = q^*$, which contains 429 miRNAs, provide all essential regulations. This group of miRNAs consists of several small clusters which are found to be tissue, pathway, diseases specific.

We revisit co-target networks for human and 21 other species and find that the density of components N_q also shows certain other universal features. Firstly, all the networks are found to be fully connected, with unit clustering coefficient and diameter. Further, v_q for all the species show data collapse, *i.e.* one can write $v_q = \mathcal{F}(q - q_c)$ where q_c is the critical threshold where the network starts breaking into dis-joint components and A is a scale factor. We find that v_q is scale free near $q = q_c$,

$$v_q \sim (q - q_c)^\beta. \quad (2)$$

In Fig. 2 we have plotted v_q as a function of $(q - q_c)$ in log scale where (a) corresponds to species with less than 250 miRNAs and (b) corresponds to the rest. The x -axis is rescaled here to obtained the collapse. In fact figures (a) and (b) could be collapsed onto each other, but they are shown as

Table 1 List of species and corresponding parameters

Species (short name)	Group-I	M	N	MYA	q_c	Species (short name)	Group-II	M	N	MYA	q_c
<i>Aedes aegypti</i> (aee)		82	16,059	285	14	<i>Monodelphis domestica</i> (mdo)		644	26013	–	26
<i>Anopheles gambiae</i> (aga)		82	12,708	–	13	<i>Gallus gallus</i> (gga)		651	20842	55	30
<i>Drosophila pseudoobscura</i> (dps)		88	12,416	–	14	<i>Macaca mulatta</i> (mml)		656	32302	6.5	57
<i>Drosophila melanogaster</i> (dme)		93	15,416	375	17	<i>Pan troglodytes</i> (ptr)		662	29355	2.7	19
<i>Caenorhabditis briggsae</i> (cbr)		135	13,785	–	6	<i>Canis familiaris</i> (cfa)		668	23628	5	32
<i>Caenorhabditis elegans</i> (cel)		136	24,728	415	12	<i>Ornithorhynchus anatinus</i> (oan)		668	23097	115	21
<i>Gasterosteus aculeatus</i> (gac)		172	26,423	–	31	<i>Bos taurus</i> (bta)		676	25759	15	40
<i>Oryzias latipes</i> (ola)		172	23,514	–	25	<i>Rattus norvegicus</i> (rno)		698	30421	55	42
<i>Takifugu rubripes</i> (tru)		173	21,972	–	31	<i>Mus musculus</i> (mmu)		793	30484	55	52
<i>Tetraodon nigroviridis</i> (tni)		174	28,005	420	34	<i>Homo sapiens</i> (hsa)		851	35864	0.2	54
<i>Xenopus tropicalis</i> (xtr)		199	24,272	360	26						
<i>Danio rerio</i> (dre)		233	28,744	420	30						

M No. of miRNA, N No. of target mRNAs, MYA Million years ago (appeared), q_c Critical threshold

Fig. 1 **a** Link-weight distribution $P_2(w)$ and **b** number of components N_q are shown for four different species (*C. elegans*, *Xenopus tropicalis*, *Bos taurus* and *Homo sapiens*)

Fig. 2 v_q versus $(q - q_c)/q_c$ for **a** Group-I and **b** Group-II, when scaled suitably, show data collapse. The scaling function is scale free near $q = q_c$ with exponent $\beta = 2$ (solid line). Inset of (b) shows that the q_c is proportional to the number of cell types of the respective species; the proportionality constant, from the best fitted line, is 3.76

separate figures only emphasize that the data for species with small number of miRNAs are comparatively noisy. Our best estimate is $\beta = 2$; a straight line with slope $\beta = 2$ is drawn in both figures for comparison. Interestingly, we find that critical threshold q_c is also proportional to the number of cell-types (see inset of Fig 2b) and thus, it can also be considered as an equivalent measure of complexity.

Upto this point, we have discussed that the network formed by the common targets of miRNA-pairs are universal in many ways: the link weight distribution $P_2(w)$ and the density of components v_q near the breakdown point $q = q_c$ for different species across a wide class are only scaled forms of respective universal functions. It is rather surprising that the species specificity show up as a scale

Fig. 3 Distribution functions **a** $P_3(w)$ and **b** $P_4(w)$ of number of miRNA co-targets for different species. In **b**, first 6 species of group-I are not shown as the data is noisy (for M being small). The y -axis in both plots are scaled here for obtaining the data-collapse

factors. One thus expects that biological functions co-regulated by miRNA pairs are possibly less specific and occurs widely across many species. To reveal more specific functions, which might be strongly species dependent, we try to find out common targets of more number of miRNAs, by taking $k > 2$. Note that for $k > 2$ the number of common targets can not be simply interpreted as the the weights of some network (as they carry three or more index referring generically to tensors which, unlike matrices, does not have a network representation). In the following we study the distribution of number of common targets for $k > 2$ number of miRNAs.

First $k = 3$. For each species there are C_3^M triples and the number of common targets of any three different miRNAs $i, j > i$ and $k > j$ is denoted as $w_{i,j,k}$. We find out these numbers from the miRNA target database (MicroCosm Targets Version), using a in house code and obtain the distribution of these numbers, denoted as $P_3(w)$. The same way one can obtain the distribution $P_4(w)$ of number of common targets of $k = 4$ miRNAs. These distribution functions are shown in Fig. 3a, b respectively. In these plots only the y axis is scaled for obtaining a data collapse. However all the data could not be collapsed on a single function; rather they split into two different scaling functions. The splitting is clearly visible for $P_3(w)$. For $P_4(w)$ we find that the data is too noisy for species with small number of miRNAs (first 6 species of group-I, in Table 1) and it was not clear whether they just represent noise due to small number of miRNAs and targets, or there are further sub-classes (splitting). We have not shown $P_4(w)$ for these species as they obstruct visibility of the other two collapsed-curves. Note that the plot also does not contain the distribution functions for *Pan troglodytes* as its number of targets is much lower compare to other species having nearly same number of miRNAs.

Why do we see universality splitting for mRNAs which are targeted by larger number of miRNAs? Those mRNAs which can be regulated by more number of miRNAs take part in larger number of biological functions or pathways providing possibility of more complex gene-regulation. In this regard, it is quite possible that universality splitting

reflects these complexity. The complexity of species belonging to one specific scaling function could be strikingly different from those belonging to the other scaling function. To verify, if this is indeed the case, in the following, we try to find out other measures which show the same kind of division.

First we look at the distribution of number of miRNAs $P_m(M)$. Since, there are only a few species, we represent the distribution by a histogram by counting the number of species having $M - 25$ to $M + 25$ species for $M = 25, 75, \dots, 825$ (refer to Fig. 4a). Clearly they show *two* distinct peaks centered about 175 and 700. It is rather natural that less complex species, those who originate in early evolution (see Table 1), have smaller number of miRNAs compared to those which are more complex. But the reason for the bi-modal structure in $P_m(N)$ is not clear. Again, existence of these two groups can also be seen from the plot of total number of miRNAs M versus N in Fig. 4b. Here M varies linearly with total number of target mRNAs N with a slope 0.007; however the y -intercept for group-I (2.3) is different from that of group-II (509.8). The natural division of these groups are consistent with the group of species belonging to two different scaling functions. In other words, one may say that the universality splitting is an indication that there are intrigue regulatory mechanisms associated with more complex species.

It is thus natural to ask whether these complexity structures are also hidden somewhere in the co-target network of miRNA *pairs*. In the following we show that, even though v_q show universal features, it does not capture all the underlying correlations of the network. This can be done by comparing the number of components of a species with an equivalent random graph. First note that N_q and $P_2(w)$ are related and N_q can be calculated from $P_2(w)$. Let

$$\rho(q) = \int_q^\infty P_2(w) dw \quad (3)$$

be the cumulative distribution of $P_2(w)$, which is same as the probability that the link-weight is larger than q . Since

Fig. 4 **a** Distribution of number of miRNAs $P_m(M)$. **b** The numbers of miRNAs (M) as a function of the number of targets (N). Two distinct peaks in **a** and two different linear relations in **b**, with same slope 0.007 (within error limits) but different intercepts 2.3 and 509.8

obtained from the best fitted line, are clear indications that the species under investigation form two different groups with respect to the number of miRNAs they have

Fig. 5 **a** Link-weight distribution $P_2(w)$ for *Homo sapiens* ($M = 851$) integrated for $w > q$ and the resulting $\rho(q)$ is plotted in **b**. **c** Number of disjoint components $n(p)$ for a random network (851 nodes and connection probability p). **d** $n(\rho(q))$ for the random network is compared with N_q . Dashed-lines here describes how to obtain $n(\rho(q))$ for a given q (see text for details)

the number of components N_q is obtained by erasing all the links with weight $w_{ij} < q$ and assigning unit weight to all other links ($w_{ij} > q$), the effective network has link density $\rho(q)$. We construct a random network of N nodes where the pairs are connected with probability $\rho(q)$ and compare the number of disjoint components of this network with N_q . Let the average number of components of a random network with connection probability p and number of nodes N nodes be $n(p)$. Then, if the miRNA co-target networks were uncorrelated, it is expected that

$$N_q = n(\rho(q)). \quad (4)$$

In Fig. 5d we plot N_q for *H. sapiens* ($N = 851$) along with $n(\rho(q))$. The construction procedure is demonstrated

in this figure. The $P_2(w)$ for *H. sapiens* (Fig. 5a) is integrated for $w > q$ (shaded region) to obtain $\rho(q)$ (Fig. 5b). Figure 5c shows $n(p)$ for a random network of $N = 851$ nodes. To obtain N_q , one starts from a given q (shown as an open circle) in Fig. 5b, obtain $\rho(q)$ and read out the number of components from (c) following the dashed line, and get the data point $(q, n(\rho(q)))$ which is shown as a solid circle in Fig. 5d. Repeating this for different values of q we obtain the expected N_q (red-line (color online) in Fig. 5d) for a random network. Clearly, this curve is substantially different from the actual N_q versus q curve for *H. sapiens*, indicating presence of correlation in miRNA co-target network. Thus the co-target network is not just another random network with a specific link-weight distribution

$P_2(w)$. These hidden correlations are uncovered in a way when one considers co-targets of three or more miRNAs.

In conclusion we have studied the distribution of number of co-targets of k number of miRNAs. For $k = 2$ these numbers can be interpreted as the link-weight distribution $P_2(w)$ of miRNA co-target network, which is known to have universal features. In this case, $P_2(w)$ s for different species are only a scaled form of an universal scaling function, and the scale-factor is a measure of complexity. We show that when links of small weights (less than q) are erased these networks breaks into several components. At the breakdown point we find an additional universal feature; the number of components show a scale free behaviour $N_q \sim (q - q_c)^2$ and could be collapsed onto each other by rescaling of only x -axis. For $k > 2$, the number of co-targets does not have an graphical representation, but their distribution could also be collapsed. Surprisingly, $P_k(w)$ for $k > 2$ studied for 22 species show universality splitting, *i.e.* $P_k(w)$ for a group of species collapse onto one scaling function where as the others belong to a different scaling function. The universality spiting is consistent with the bimodal distribution of number of miRNAs and the linear dependence of number of miRNAs on the number of mRNAs (two groups have different y -intercept). The two different scaling functions are thus associated with two different class of animal, the early ones which have less number of miRNAs and less complex and the late ones which are more complex. It remains to study, if complex regulation occurring due to the fact that if mRNAs, which are targeted by more number of miRNAs contribute to the possibility of complex regulation and new biological functions.

References

- Ambros V (2004) The functions of animal microRNAs. *Nature* 431:350–355
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2):281–297
- Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136(2):215–233
- Basu M, Bhattacharyya NP, Mohanty PK (2011) Modules of human micro-RNA co-target network. *J Phys Conf Ser* 297:012002
- Basu M, Bhattacharyya NP, Mohanty PK (2014) Link-weight distribution of microRNA co-target networks exhibit universality. *EPL* 105:28007
- Boross G, Orosz K, Farkas IJ (2009) Human microRNAs co-silence in well-separated groups and have different predicted essentialities. *Bioinformatics* 25(8):1063–1069
- Clarke NJ, Sanseau P (eds) microRNA: biology, function and expression, 1st edn. Dna Press, May 2, 2006
- Dong H, Lei J, Ding L, Wen Y, Ju H, Zhang X (2013) MicroRNA: function, detection, and bioanalysis. *Chem Rev* 113(8):6207–6233
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* 310:1817–1821
- Hansen KF, Obrietan K (2013) MicroRNA as therapeutic targets for treatment of depression. *Neuropsychiatr Dis Treat* 9:1011–1021
- Hu S, Huang M, Li Z, Jia F, Ghosh Z, Lijkwan MA, Fasanaro P, Sun N, Wang X, Martelli F, Robbins RC, Wu JC (2010) MicroRNA-210 as a novel therapy for treatment of ischemic heart disease. *Circulation* 122:S124–S131
- Ivanovska I, Cleary MA (2008) Combinatorial microRNAs: working together to make a difference. *Cell Cycle* 7(20):3137–3142
- Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* 115(7):787–798
- Liu J (2008) Control of protein synthesis and mRNA degradation by microRNAs. *Curr Opin Cell Biol* 20(2):214–221
- Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q (2008) An analysis of human microRNA and disease associations. *PLoS ONE* 3(10):e3420
- MicroCosm Targets Version 5. <http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5>
- Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126(6):1203–1217
- miRBase database. <http://www.mirbase.org>
- miRDB database. <http://www.mirdb.org/miRDB/>
- miRWalk database. <http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/>
- Mookherjee S, Sinha M, Mukhopadhyay S, Bhattacharyya NP, Mohanty PK (2009) MicroRNA Interaction network in human: implications of clustered microRNA in biological pathways. *Online J Bioinform* 10:280
- TargetScan database. <http://www.targetscan.org/>
- Wang V, Wu W (2009) MicroRNA-based therapeutics for cancer. *BioDrugs* 23(1):15–23
- Wu Y, Crawford M, Mao Y, Lee RJ, Davis IC, Elton TS, Lee LJ, Nana-Sinkam SP (2013) Therapeutic delivery of MicroRNA-29b by cationic lipoplexes for lung cancer. *Mol Ther Nucleic Acids* 2:e84
- Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, Xu LD, Wang YY, Du L, Zhang YP, Jiang W, Li CQ, Xiao Y, Li X (2011) MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res* 39(3):825–836
- Yu J, Wang F, Yang GH, Wang FL, Ma YN, Du ZW, Zhang JW (2006) Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochem Biophys Res Commun* 349:59–68