bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Response to "No evidence of functional co-adaptation between clustered microRNAs"**

2 Yirong Wang[1], Hong Zhang[1], and Jian Lu*[,1]

3 1. State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics,

4 School of Life Sciences, Peking University, Beijing, 100871, China

5

6 *Corresponding author: Jian Lu, luj@pku.edu.cn

7

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Abstract**

2 microRNAs (miRNAs) are a class of endogenously expressed small non-coding RNAs that
3 regulate target genes at the post-transcriptional level. One significant feature of miRNA is that
4 their genomic locations are often clustered together in the genome. In a previous study (Wang,
5 et al. 2016), we proposed a "functional co-adaptation" model to explain how clustering helps
6 new miRNAs survive and develop functions during long-term evolution. In a manuscript
7 recently posted at bioRxiv (doi:10.1101/274811), Marco claimed that he re-analyzed our data
8 and came to a different conclusion. However, we found his analyses were conducted in an
9 inappropriate approach. He also claimed that the absence of substitution in highly conserved
10 miRNAs does not support the "functional co-adaption" model based on the misunderstanding
11 of our model. In summary, the analyses and claims of Marco, which are flawed, do not refute
12 our model.

13

14 Keywords: miRNAs, target, functional co-adaptation, drift-draft

15

16

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 miRNAs are a class of endogenously expressed small noncoding RNAs (~22 nt in length) that
2 down-regulate the expression of target genes at the post-transcriptional level. A salient feature
3 is that many animal miRNAs are clustered into discrete genomic regions (Lagos-Quintana et
4 al. 2001; Lau et al. 2001; Lai et al. 2003; Altuvia et al. 2005; J. Graham Ruby et al. 2007; Marco
5 et al. 2013; Mohammed, Siepel, et al. 2014). The clustering patterns suggest that miRNAs in
6 the same cluster might be co-transcribed (Baskerville and Bartel 2005; Saini et al. 2007;
7 Ozsolak et al. 2008; Wang et al. 2009; Ryazansky et al. 2011) and be functionally related by
8 targeting the same gene or different genes in the same biological pathway (Bartel 2004; Grun,
9 et al. 2005; Kim and Nam 2006; Yu, et al. 2006). For example, the *mir-17~92* cluster plays an
10 important role in mammalian development and tumorigenesis (O'Donnell, et al. 2005; Lu, et al.
11 2007; Ventura, et al. 2008; Xiao, et al. 2008). Gene deletion experiments suggest members in
12 the *mir-17~92* cluster have essential and overlapping functions (Ventura, et al. 2008). The *mir-*
13 *106b~93~25* and *mir-222~221* clusters are upregulated and modulate G1/S phase transition in
14 gastric cancer, and members of the two cluster have functional associations by targeting genes
15 in the Cip/Kip family members of Cdk inhibitors (Kim, et al. 2009). The brain specifically
16 expressed *mir-379~410* cluster is required for the activity-dependent development of
17 hippocampal neurons, and multiple miRNAs from the cluster are necessary for the correct
18 elaboration of the dendritic tree (Fiore, et al. 2009). miRNAs in *mir-23a~27a~24-2* cluster also
19 have cooperative effects in various health and diseased conditions (Chhabra, et al. 2010).
20
21 We recently proposed a "functional co-adaptation" model to systematically investigate the
22 functional relatedness of clustered miRNAs (Wang, et al. 2016). We provided several lines of
23 evidence to support the "functional co-adaptation" model. First, we found the observed number
24 of genes co-targeted by miRNAs in the same cluster but with different seeds are significantly
25 higher than the number obtained by random permutations. Second, we found genes targeted by
26 multiple miRNAs from the same clusters, in general, have lower expression levels than genes
27 targeted by multiple miRNAs from distinct clusters. Third, we show that the miRNAs in the
28 same cluster with different seeds tend to target genes in the same biological pathways. Fourth,
29 we transfected four members of the *mir-17~92* cluster into human 293FT cells individually and
30 quantified the alteration of mRNA abundance with deep-sequencing, which verified the
31 overlapping of target genes experimentally. Fifth, we experimentally determined the target
32 genes of *miR-92a*, the founding member of the *mir-17~92* in *Drosophila*， and examined the
33 relationship between the target genes of *miR-92a* in *Drosophila* and the target genes of the *mir-*
34 *17~92* cluster in humans. Our experimental results well supported the "functional co-adaptation"
35 model. Finally, we also conducted evolutionary analysis to show that positive Darwinian

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 selection drives the evolution of the newly formed miRNA clusters in both primates and
2 *Drosophila*.
3
4 In a manuscript recently posted at bioRxiv (Marco, 2018; doi:10.1101/274811), Marco claimed
5 that he re-analyzed our data and found “No evidence of functional co-adaptation between
6 clustered microRNAs”. Marco claimed that the observed overlap of target genes by the
7 clustered miRNAs are mostly caused by the similarity between two seed sequences in the *miR-*
8 *182/183/96* cluster. Marco argued that clustered miRNAs from different miRNA families do
9 not share more targets than expected by chance after correcting for these factors. Marco also
10 argued that our permutation tests by shuffling miRNA-target inactions would lead to spuriously
11 low *P* values. Moreover, Marco also raised a series of other critiques about the “function co-
12 adaptation” model in different versions of his comments. Unfortunately, Marco’s critiques are
13 based on flawed analysis and misled concept of miRNA evolution, which are summarized as
14 follows.
15
16 The major concern Marco raised is whether the observed number of genes targeted by at least
17 two conserved miRNAs with different seeds from the same miRNA clusters is statistically
18 higher than the number obtained under the assumption of randomness. In our previous study
19 (Wang, et al. 2016), to test whether miRNAs in the same clusters tend to regulate overlapping
20 sets of genes, we obtained expression profiles of miRNAs and mRNAs from five tissues of
21 human males as determined in previous study (Brawand, et al. 2011; Meunier, et al. 2013).
22 Since the co-adaption of clustered miRNAs is the result of co-evolution between miRNA and
23 target sites, in the permutation analysis, we first shuffled the co-expressed seed:target pairing
24 (TargetScan $P_{CT} > 0.5$), and then we tested how many genes were targeted by at least two
25 miRNAs (with distinct seeds) in the same clusters. These permutation tests were performed for
26 1,000 replicates. By this way, the conservation level and length of 3' UTR of target mRNAs,
27 the number of miRNAs for each target gene, and the compositions of each miRNA cluster are
28 fully controlled. Applying this procedure to the pooled dataset of miRNA-mRNA co-
29 expression from different tissues, the result of Wang et al. was successfully reproduced (Fig.
30 1). When the co-expression data of each tissue was analyzed individually, the similar pattern
31 was still observed (Fig. 1). Importantly, when the *miR-182/183/96* cluster was excluded, we
32 can still observe similar patterns (Fig. 2).
33
34 By contrast, Marco failed to reproduce our results because he only shuffled the location of the
35 miRNAs and kept the seed: targeting pairing unaltered. Marco obtained a pattern that the
36 observed number slightly but still significantly higher than the expected number under the
37 assumption of randomness (P = 0.0359). Since the difference between the observed and

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 expected numbers are quite smaller obtained by Marco compared to what we obtained (Wang,
2 et al. 2016), Marco argued that our results are mainly caused by the similarity of the targets
3 between two seed sequences of the *miR-182/183/96* cluster, and "the expected high number of
4 common targets between pairs of microRNAs that have a large number of targets each".
5 However, Marco's permutation tests are biased and flawed. Here, the condition to be tested is
6 whether miRNAs with distinct seeds from the same cluster have more common target genes
7 than expected under randomness, rather than to test whether miRNAs are clustered as Marco
8 conducted.
9
10 Despite the numerous exchanges between Marco and us on the detailed permutation test
11 procedures, Marco consistently argued that our permutation tests by shuffling miRNA-target
12 interactions would lead to spuriously low *P* values. To show this point, Marco introduced a
13 hypothetical regulatory network, in which two clustered miRNAs and an additional miRNA
14 target the same gene "a" and each miRNA has *n* additional targets (see Figure 2 of Marco, 2018;
15 doi:10.1101/274811 for details). Marco derived the formula for calculating the probability that
16 a permutation reports a common target between two clustered microRNAs under randomness
17 as

18 $$P = \frac{6(n!)^2(n+1)!}{(3n+3)!} \sim \frac{4\sqrt{3}\pi}{n}(\frac{1}{3})^{3(n+1)}$$

19 based on "standard combinatorics" (Marco, 2018; doi:10.1101/274811).
20
21 Marco argued that based on this formula, one can obtain very small *P* values even when there
22 is no enrichment in common genes targeted by clustered microRNAs. Unfortunately, after
23 careful examination, we found the formula Marco derived is incorrect. The correct formula
24 for *P* value should be:

25 $$P = \frac{4n^2 + 6n + 2}{9n^2 + 9n + 2}$$

26 ,based on basic combinatorics (see Appendix for details).
27
28 In Figure 3, we show that the *P* values calculated by Marco are consistently smaller than the *P*
29 values calculated by the formula we derived. To further support our argument, we also followed
30 the scheme presented in Figure 2 of Marco, 2018 (doi:10.1101/274811) and estimated the
31 empirical *P* value by randomly permutating the miRNA:target pairing for 1,000 times at a given
32 *n*. As we show in Figure 3, The *P* values calculated with the correct formula (red curve) are
33 highly consistent with the empirical *P* values, while the *P* values calculated with Marco's
34 formula are consistently much smaller than the *P* values obtained by simulations. Therefore,

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 Marco's argument that our permutation tests by shuffling miRNA-target interactions would
2 lead to spuriously low P values is based on his erroneous formula in calculating the P value. In
3 fact, the correct P value formula indicates shuffling miRNA-target interactions would not
4 generate a significant P value when there is no enrichment in common genes targeted by
5 clustered microRNAs.
6
7 Marco also argued that the differences between clustered and non-clustered miRNAs are not
8 significant when he only focused on the pairs of miRNAs that have fewer than 6 common
9 nucleotides at seed regions in a cluster (the Fig. 1C of (Marco 2018)). He argued that the
10 observed overlap between targets in some clustered miRNAs is actually the random
11 consequence of the similarity between their seed sequences, and is not associated with whether
12 the miRNAs are clustered or not. An inspection of his code indicates his critique is based on
13 incorrect calculation of "the fraction of common targets" for a pair of miRNAs. For example,
14 let us suppose that miRA has x targets, miRB has y targets, and z targets are shared between
15 miRA and miRB. Marco calculated the fraction of common targets as $z/\min(x,y)$ in his Fig. 1C,
16 while calculated the fraction of common targets as $z(x+y)/(2xy)$ in his Fig. 1D. Obviously, using
17 the harmonic mean ($z(x+y)/(2xy)$) is more appropriate to calculate the fraction of common
18 targets, as employed by Marco in his Fig. 1D. Interestingly, when we repeated his analysis in
19 Fig. 1C with the harmonic mean ($z(x+y)/(2xy)$) approach as he did in Fig. 1D, we found a pair
20 of miRNAs with seed similarity less than 6 nucleotides have a significantly higher fraction of
21 common targets if the two paired miRNAs are from the same miRNA cluster ($P = 0.008$, Fig.
22 4). Therefore, the results in the Fig. 1C of (Marco 2018) are based on incorrect analytical
23 procedures.
24
25 Notably, both our original analysis (Wang, et al. 2016) and Marco (2018) only considered the
26 evolutionarily conserved target sites of the evolutionarily conserved miRNAs (only miRNAs
27 that have target sites with TargetScan $P_{CT} > 0.5$ were used in the permutation analyses). Here
28 we considered all the miRNAs that have at least one target site (not necessarily conserved) with
29 weighted context++ score (WCC) < -0.3 (Grimson, et al. 2007; Agarwal, et al. 2015). Target
30 sites with WCC < -0.3 are usually located in optimized genomic context for efficient repression
31 of target mRNAs. With the criteria of WCC < -0.3, only 125 (6%) of 2,081 common targets of
32 clustered miRNAs are contributed by *miR-182/183/96*, which would remarkably reduce the
33 potential bias caused by *miR-182/183/96*. We consistently observed that clustered miRNAs
34 have significantly more common target genes than expected before (empirical $P = 0.0037$, Fig.
35 5A) or after (empirical $P = 0.0088$, Fig. 5B) excluding *miR-182/183/96* cluster in the
36 permutation tests by shuffling miRNA loci as Marco performed. A more significant trend was
37 observed when the permutation tests were performed by shuffling co-expressed seed: target

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 pairs (P < 0.0001 in both Fig. 5C and 5D) as conducted in our original study (Wang, et al.
2 2016). Therefore, we reaffirmed the thesis that clustered miRNAs have more common target
3 genes than expected using miRNA targets with WCC < -0.3.
4
5 Furthermore, it is hard to understand why Marco argued that the targets shared between *miR-*
6 *183* and *miR-96* in the *miR-182/183/96* cluster should be excluded from the analysis. The seeds
7 of *miR-183-5p* and *miR-96-5p* are very similar: AUGGCAC and UUGGCAC for the former
8 and latter, respectively. However, BLAST2SEQ analysis between the precursor sequences of
9 human *mir-183* and *mir-96* does not find significant similarity, suggesting these two miRNA
10 precursors are unlikely to be duplicated miRNAs. Instead, the functional co-adaptation model
11 might well explain the large number of target genes shared between these two miRNAs: During
12 long time evolution, the adaptive changes in miRNA seed region or target sites on mRNAs
13 drive the clustered miRNAs to regulate the same or functionally related genes. Therefore, this
14 cluster serves as a strong evidence that convergent evolution has occurred between the seeds of
15 *miR-183* and *miR-96* due to functional coadaptation.
16
17 Curiously, Marco did not report his re-analysis results of the *miR-17~92* cluster over-expression
18 data we generated (Wang, et al. 2016). Many previous studies have demonstrated that the *mir-*
19 *17~92* cluster plays an important role in tumorigenesis, development of lungs and immune
20 systems (O'Donnell, et al. 2005; Lu, et al. 2007; Ventura, et al. 2008; Xiao, et al. 2008), and
21 deletion of the *mir-17~92* cluster revealed that miRNAs in this cluster have essential and
22 overlapping functions (Ventura, et al. 2008). Furthermore, we found the conserved target genes
23 shared between members of the *mir-17~92* cluster is significantly higher than the simulated
24 ones (Fig. 6). Importantly, we selected four distinct mature miRNAs in the *miR-17~92* cluster
25 (*miR-17, miR-18a, miR-19a,* and *miR-92a*) and transfected each miRNA mimic as well as the
26 miRNA mimic Negative Control (NC) into human 293FT cells (Wang, et al. 2016). With high-
27 throughput mRNA-Seq, we found the predicted target genes (TargetScan P_{CT} > 0.5) of each
28 transfected miRNA are significantly more down-regulated than genes that do not have the target
29 sites (Figure 4 of Wang et al. 2016). We identified 301, 55, 345 and 268 high-confidence target
30 genes (TargetScan P_{CT} > 0.5) for *miR-17*, *18a*, *19a* and *92a* respectively that were down-
31 regulated with log2(FoldChange) < -0.1 in the corresponding miRNA transfection experiments
32 (totally 775 high-confidence genes after removing overlapping genes, Figure 4I of Wang et al.
33 2016). Among these 775 high-confidence target genes, 172 were targeted by at least two out of
34 the four miRNAs, significantly higher than the number obtained by randomness (P < 0.001, see
35 Figure 4I and Table S8 of Wang et al. 2016 for details). These results well support the
36 "functional co-adaptation model" we proposed. If there is really "No evidence of functional co-

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 adaptation between clustered microRNAs" as Marco argued, how can one explain these observed
2 patterns?
3
4 Moreover, Marco did not fully understand the "functional co-adaption" model, which led him
5 to make the argument that "microRNAs in a cluster are primarily under positive selection"
6 (April/20/2018 version of doi:10.1101/274811). Although our population genetic analysis
7 suggests Darwinian selection drives the evolution of the newly formed miRNA clusters in
8 primates and in *Drosophila*, our model does not necessarily suggest all the clustered miRNAs
9 are driven by positive selection (Wang, et al. 2016). What we proposed is that, new miRNAs
10 originating nearby a pre-existed miRNA would have a higher chance to be maintained in the
11 initial stage of cluster formation due to the tight genetic linkage. Then positive Darwinian
12 selection might drive the newly emerged miRNAs to develop functions related to the pre-
13 existing miRNAs in the same cluster or drive the evolution of all the new miRNAs in the same
14 cluster to develop related functions during the long-term evolution. Once the cluster is fully
15 established, the miRNAs in the same cluster will be maintained by purifying selection and
16 become highly conserved after that (Wang, et al. 2016). Thus, one could not expect to observe
17 signature of ongoing positive selection in the well-established clusters such as the *miR-17~92*
18 or the *miR-182/183/96* cluster which are ancient and conserved after the establishment, as
19 Marco did. Marco's observation that "both seed sequences (of *miR-183-5p* and *miR-96-5p*)
20 have been conserved since their origin and, therefore, there is no evidence of substitutions
21 happening in the seed of these microRNAs for the last 600 million years" could not refute our
22 model. Marco also used the deep conservation of the clustered miRNAs in other clusters (*mir-*
23 *106b~25* cluster, *mir-23b~24* cluster, and *mir-379~410* cluster of Fig. S1 in the April/20/2018
24 version of his manuscript) to argue against the "functional co-adaptation" model. The deep
25 conservation of the seed sequences as Marco showed can only suggest these miRNAs are
26 conserved due to extremely strong selective constraints during vertebrate evolution. These
27 observations do not provide any evidence to defy the "functional co-adaptation" model since
28 one cannot tell whether the miRNAs have changed since emergence as no outgroup sequence
29 available.
30
31 Based on the observation that *Drosophila* new miRNAs often arose around the pre-existing
32 ones to form clusters, Marco and colleagues proposed a "drift-draft" model which suggests that
33 the evolution of miRNA clusters was influenced by tight genetic linkage and largely non-
34 adaptive (Marco, et al. 2013). Under such a model, the motifs of the pre-existing miRNAs
35 would protect new miRNAs to be transcribed and processed properly since those motifs were
36 already interacting with the miRNA processing machinery. Thus, the *de novo* formed new

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 miRNAs are sheltered by the established ones in the same cluster because mutations that abolish
2 the transcription or processing of the new miRNA will affect the pre-existing ones as well and
3 are hence selected against. On the other hand, if a *de novo* formed miRNA is located in a
4 discrete locus, it will have a higher probability to degenerate, either by mutations abolishing its
5 transcription or by mutations impairing its processing. Although Marco argued the "drift-draft"
6 and "functional co-adaptation" models are mutually exclusive, we did not think the "functional
7 co-adaptation" we proposed is strictly "an alternative to the drift-draft model".
8
9 Our previous results and others suggest that many newly-emerged miRNAs are evolutionarily
10 transient, with a high birth-and-death rate (Berezikov et al. 2006; Rajagopalan et al. 2006; Lu,
11 Shen, et al. 2008; Lu et al. 2010). Therefore, it is possible that the newly emerged miRNAs in
12 the clusters would be sheltered by the pre-existing established miRNAs. However, the
13 protection effect alone cannot explain why miRNAs in the same cluster have significantly
14 higher numbers of overlapping target genes. Moreover, many *de novo* formed novel miRNAs
15 will degenerate even after they are fixed in the populations if they are not maintained by
16 functional constraints (Berezikov et al. 2006; Lu, Shen, et al. 2008). Thus developing functions
17 related to the pre-existing miRNAs will help the novel miRNAs to survive and stabilize. The
18 "functional co-adaptation" model we proposed well accounts for the evolution and function of
19 *de novo* formed new miRNAs in the clusters (Figure 2D of Wang et al. 2016). Since miRNAs
20 in the same clusters are usually co-transcribed temporally or spatially (see below for details),
21 the newly formed miRNAs might gradually develop functions to target genes that are related
22 to the pre-existing miRNAs in the same cluster; or multiple *de novo* formed new miRNAs in
23 the same cluster interplay to regulate overlapping sets of target genes. Therefore, although
24 miRNAs in the same cluster have independent origins, they might regulate overlapping sets of
25 target genes through convergent evolution. After that, the clustering patterns of miRNAs and
26 the modular regulation of target genes will be stabilized by natural selection during long-term
27 evolution. Of course, the evolutionary process of miRNAs is also companied by the co-
28 evolution of the target sites, which might also potentially affect the base compositions or even
29 the length of a 3' UTR. In a separate study, we also showed the target sites of miRNAs also
30 experienced frequent births and deaths (Luo, et al. 2018). But the evolution of the target sites
31 alone would not cause the clustering pattern of miRNAs. A plausible scenario is that after a
32 new miRNA originates in a cluster, the substitutions that change the sequences and expression
33 of the new miRNAs, the interactions between miRNAs in the same cluster, and the co-evolution
34 between miRNAs and the target sites or the 3' UTRs, jointly affect the evolution of the
35 clustering pattern of miRNAs.
36
37 Understanding the molecular mechanisms and evolutionary principles of the miRNA clustering

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

would deepen our understanding of the regulatory roles of miRNAs in various biological processes or diseases. The "functional co-adaptation" model we propose is well supported by evolutionary and functional genomic data.

References

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. Elife 4.

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

Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343-348.

Chhabra R, Dubey R, Saini N. 2010. Cooperative and individualistic functions of the microRNAs in the miR-23a~27a~24-2 cluster and its implication in human diseases. Mol Cancer 9:232.

Fiore R, Khudayberdiev S, Christensen M, Siegel G, Flavell SW, Kim TK, Greenberg ME, Schratt G. 2009. Mef2-mediated transcription of the miR379-410 cluster regulates activity-dependent dendritogenesis by fine-tuning Pumilio2 protein levels. EMBO J 28:697-710.

Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Molecular Cell 27:91-105.

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS Comput Biol 1:e13.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10:126-139.

Kim VN, Nam JW. 2006. Genomics of microRNA. Trends Genet 22:165-173.

Lu Y, Thomson JM, Wong HY, Hammond SM, Hogan BL. 2007. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. Dev Biol 310:442-453.

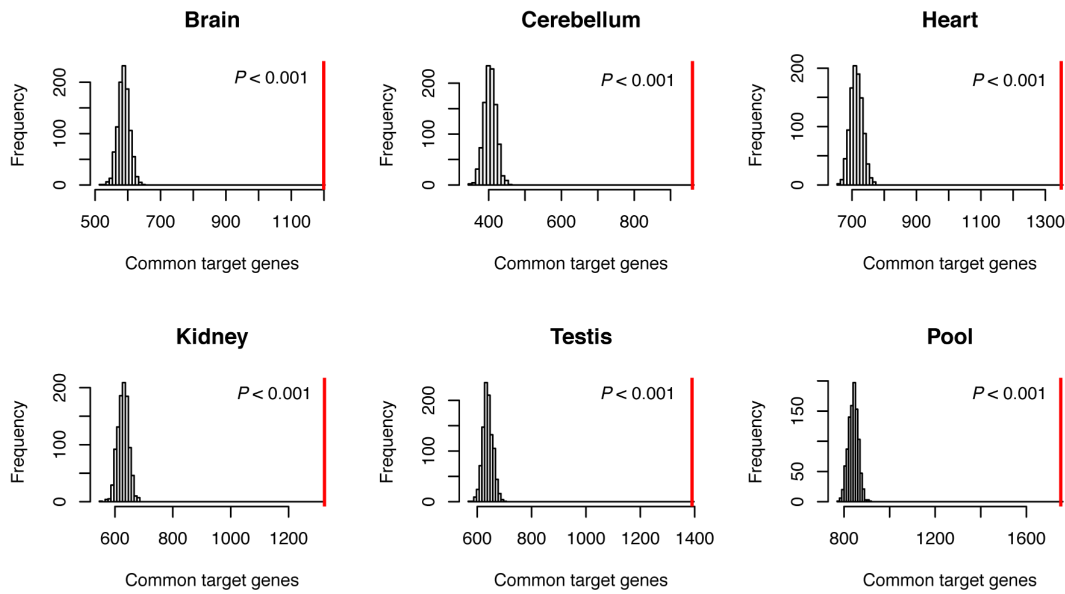
Luo J, Wang Y, Yuan J, Zhao Z, Lu J. 2018. MicroRNA duplication accelerates the recruitment of new targets during vertebrate evolution. RNA.

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 Marco A. 2018. No evidence of functional co-adaptation between clustered microRNAs.
2 bioRxiv. DOI: 10.1101/274811.
3 Marco A, Ninova M, Ronshaugen M, Griffiths-Jones S. 2013. Clusters of microRNAs emerge
4 by new hairpins in existing transcripts. Nucleic Acids Res 41:7745-7752.
5 Meunier J, Lemoine F, Soumillon M, Liechti A, Weier M, Guschanski K, Hu H, Khaitovich P,
6 Kaessmann H. 2013. Birth and expression evolution of mammalian microRNA genes. Genome
7 Res 23:34-45.
8 O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. 2005. c-Myc-regulated
9 microRNAs modulate E2F1 expression. Nature 435:839-843.
10 Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J,
11 Bronson RT, Crowley D, Stone JR, et al. 2008. Targeted deletion reveals essential and
12 overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell 132:875-886.
13 Wang Y, Luo J, Zhang H, Lu J. 2016. microRNAs in the Same Clusters Evolve to Coordinately
14 Regulate Functionally Related Genes. Mol Biol Evol 33:2232-2247.
15 Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, Henderson JM, Kutok JL,
16 Rajewsky K. 2008. Lymphoproliferative disease and autoimmunity in mice with increased
17 miR-17-92 expression in lymphocytes. Nat Immunol 9:405-414.
18 Yu J, Wang F, Yang GH, Wang FL, Ma YN, Du ZW, Zhang JW. 2006. Human microRNA
19 clusters: genomic organization and expression profile in leukemia cell lines. Biochem Biophys
20 Res Commun 349:59-68.
21
22

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Figures**



2

3 **Figure 1.** Permutation analysis of common target genes of miRNAs from the same cluster in
4 each tissue or pooled data by shuffling co-expressed seed:target pairs. 1,000 replicates were
5 performed for each panel. The observed number of common targets was indicated with
6 vertical red line and the proportion of simulations yielding a number larger than observed
7 value was shown at top right.

8

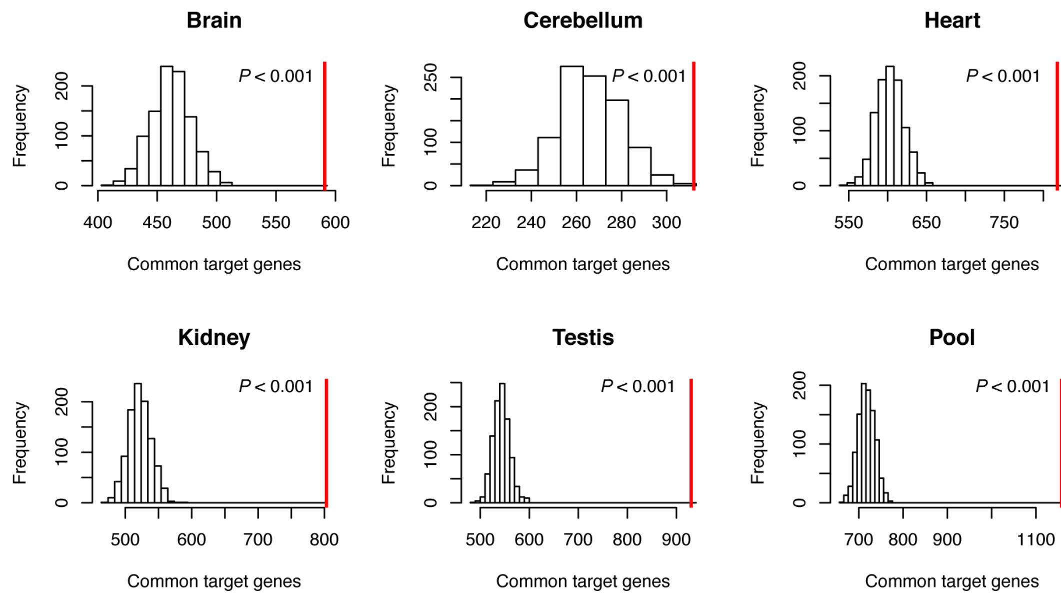
9

10

11

12

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



1

2 **Figure 2.** Permutation analysis of common target genes of miRNAs from the same cluster in
3 each tissue or pooled data by shuffling co-expressed seed:target pairs. Similar as Fig. 1 but
4 *miR-182/183/96* cluster was excluded in the analysis.

5

6

7

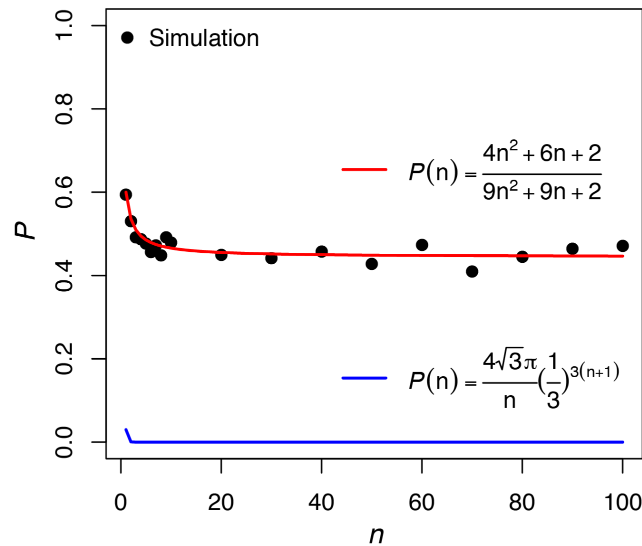
8

9

10

11

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



1

2 **Figure 3.** The comparison between the *P* values (y axis) obtained by Marco's formula (blue
3 line) and the formula derived by us (red line) and the simulation results (black points). In a
4 hypothetical regulatory network, two clustered miRNAs and an additional miRNA target the
5 same gene "a" and each miRNA has *n* (*x-axis*) additional targets, and the *P* value is the
6 probability that a permutation reports a common target between two clustered microRNAs
7 under randomness (see Figure 2 of Marco, 2018; doi:10.1101/274811 for details). The *P* value
8 calculated with Marco's formula are consistently smaller (< 0.05) even when there is no
9 enrichment in common genes targeted by clustered microRNAs. However, the *P* values
10 calculated with the correct formula (red curve) indicates shuffling miRNA-target interactions
11 would not generate a significant *P* value when there is no enrichment in common genes targeted
12 by clustered microRNAs. The black points are the empirical *P* value obtained by randomly
13 permutating the miRNA:target pairing for 1,000 times at a given *n*. Note our formula (red line)
14 and simulation results (black points) are highly consistent.

15

16

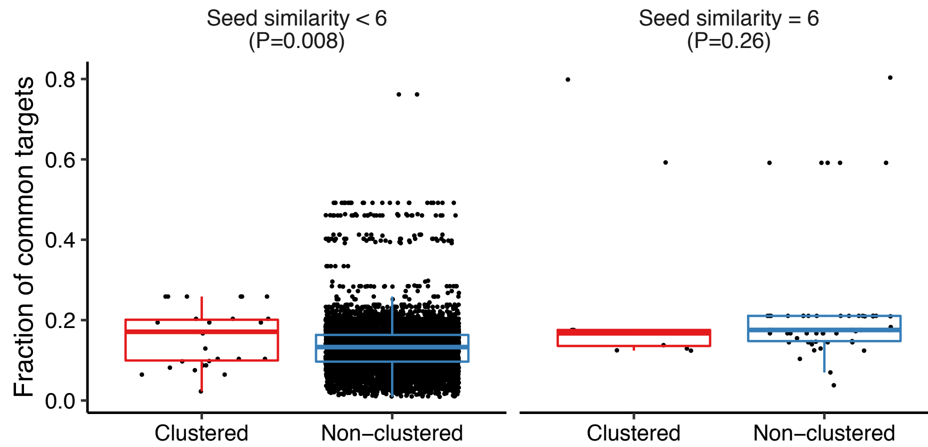
17

18

19

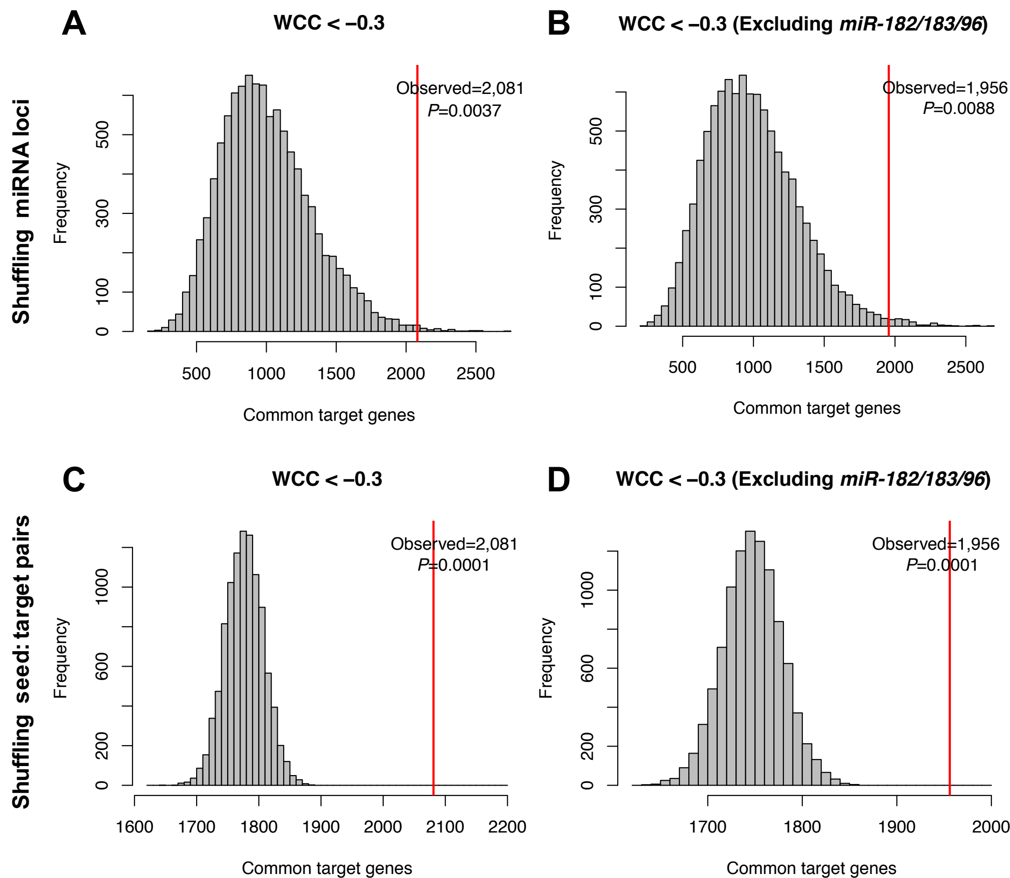
20

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



1

2 **Figure 4**. The distribution of the fraction of common targets for a pair of miRNAs that are from
3 the same cluster (Clustered) or not (Non-clustered). Paired miRNAs with a seed similarity less
4 than or equal to six nucleotides were compared separately with the Wilcoxon rank-sum tests.

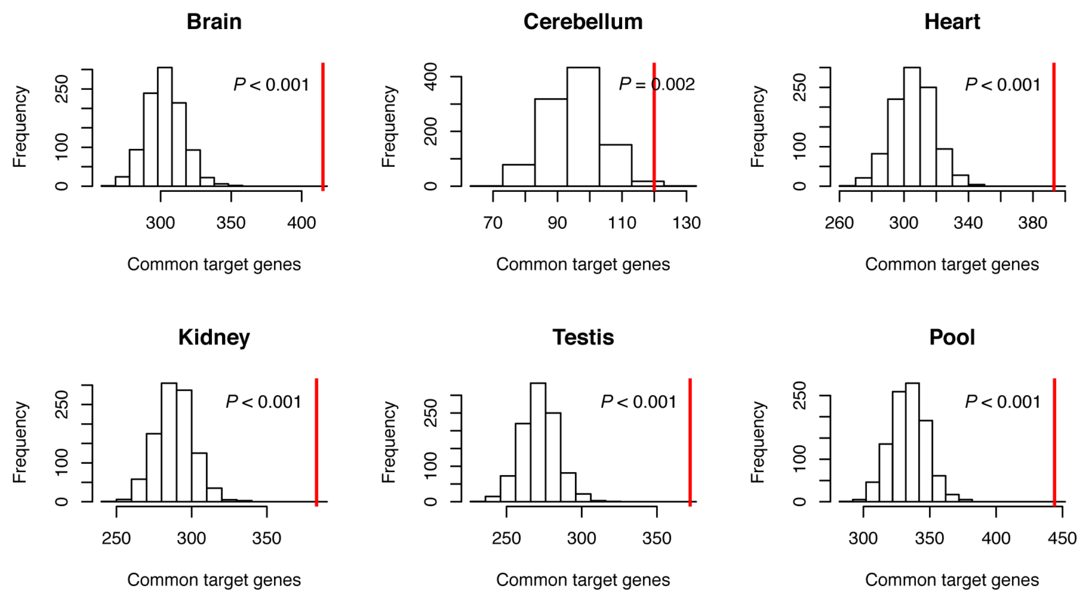


5

6 **Figure 5.** Permutation analysis of common target genes of clustered miRNAs. (**A** and **B**)
7 Permutation analysis by shuffling miRNA loci using miRNA target genes with weighted

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 context score++ (WCC) < -0.3 before (A) or after (B) excluding *miR-182/183/96* cluster. (**C**
2 and **D**) Permutation analysis by shuffling co-expressed seed: target pairs using miRNA target
3 genes with WCC < -0.3 before (C) or after (D) excluding *miR-182/183/96* cluster. All the
4 human miRNAs and their targets with WCC < -0.3 were analyzed. Only co-expressed
5 miRNA:target pairs in human tissues were considered. 10,000 replicates were performed for
6 each permutation test. The observed numbers of common targets were indicated with red lines.
7 The empirical *P* value was calculated as the fraction of simulations yielding a number of
8 common targets larger than the observed one.



9

10 **Figure 6.** The number of observed (red) and simulated common target genes of conserved
11 miRNAs in *mir-17~92* cluster in the permutation analysis of Fig. 1.

12

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 **Appendix**

2 In a hypothetical regulatory network where two clustered miRNAs (miR-A and miR-B) and an
3 additional miRNA (miR-C) target the same gene "a" and each miRNA has *n* additional targets
4 (see Figure 2 of Marco, 2018 doi:10.1101/274811 for detailed description), the total number of
5 all possible pairing between the three miRNAs and $3n + 1$ targets is:

6 $$N = \binom{3n+3}{n+1}\binom{2n+2}{n+1}\binom{n+1}{n+1}$$

7 $$= \binom{3n+3}{n+1}\binom{2n+2}{n+1}$$

8 $$= \frac{(3n+3)!}{(2n+2)!\,(n+1)!}\frac{(2n+2)!}{(n+1)!\,(n+1)!}$$

9 $$= \frac{(3n+3)!}{(n+1)!\,(n+1)!\,(n+1)!}$$

10 "a" is the only gene that could be targeted by more than one miRNA since other targets only
11 interact with a single miRNA. Since "a" has three interactions, there are three possible scenarios
12 where both the clustered miR-A and miR-B could target gene "a": (1) two interactions of gene
13 "a" are taken by miR-A, and the remaining one is taken by miR-B; (2) one interaction of gene
14 "a" is taken by miR-A, and the remaining 2 are taken by miR-B; (3) Both miR-A and miR-B
15 takes one interaction of gene "a", and the remaining one is taken by "miR-C". We can obtain
16 the number of possible combinations under each of the three scenarios as:

17 $$x_{(1)} = \binom{3}{2}\binom{3n}{n-1}\binom{1}{1}\binom{2n+1}{n}\binom{n+1}{n+1} = 3\binom{3n}{n-1}\binom{2n+1}{n}$$

18 $$x_{(2)} = \binom{3}{1}\binom{3n}{n}\binom{2}{2}\binom{2n}{n-1}\binom{n+1}{n+1} = 3\binom{3n}{n}\binom{2n}{n-1}$$

19 $$x_{(3)} = \binom{3}{1}\binom{3n}{n}\binom{2}{1}\binom{2n}{n}\binom{n+1}{n+1} = 6\binom{3n}{n}\binom{2n}{n}$$

20 The total number of combinations where gene "a" is targeted by both miR-A and miR-B is:

21 $$X = x_{(1)} + x_{(2)} + x_{(3)}$$

22 $$= 3\binom{3n}{n-1}\binom{2n+1}{n} + 3\binom{3n}{n}\binom{2n}{n-1} + 6\binom{3n}{n}\binom{2n}{n}$$

23 $$= 3\frac{(3n)!}{(2n+1)!\,(n-1)!}\frac{(2n+1)!}{(n+1)!\,n!} + 3\frac{(3n)!}{(2n)!\,n!}\frac{(2n)!}{(n+1)!\,(n-1)!} + 6\frac{(3n)!}{(2n)!\,n!}\frac{(2n)!}{n!\,n!}$$

bioRxiv preprint doi: https://doi.org/10.1101/313817; this version posted April 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1 $$= 6\frac{(3n)!}{(n+1)!\,n!\,(n-1)!} + 6\frac{(3n)!}{n!\,n!\,n!}$$

2 $$= 6(3n)!\left[\frac{1}{(n+1)!\,n!\,(n-1)!} + \frac{1}{n!\,n!\,n!}\right]$$

3 Therefore, the probability that both miR-A and miR-B target gene "a" is:

4 $$P = \frac{X}{N} = 6\frac{(3n)!}{(3n+3)!}\left[\frac{(n+1)!\,(n+1)!\,(n+1)!}{(n+1)!\,n!\,(n-1)!} + \frac{(n+1)!\,(n+1)!\,(n+1)!}{n!\,n!\,n!}\right]$$

5 $$= 6\frac{(3n)!}{(3n+3)!}[(n+1)(n+1)n + (n+1)(n+1)(n+1)]$$

6 $$= 6\frac{(3n)!}{(3n+3)!}(n+1)(n+1)(2n+1)$$

7 $$= 6\frac{(n+1)(n+1)(2n+1)(3n)!}{(3n+3)(3n+2)(3n+1)(3n)!} = 6\frac{(n+1)(n+1)(2n+1)}{3(n+1)(3n+2)(3n+1)}$$

8 $$= \frac{2(2n+1)(n+1)}{(3n+2)(3n+1)} = \frac{4n^2+6n+2}{9n^2+9n+2}$$