Integrative Biology

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c4ib00156g

Characterization of the microRNA pool and the factors affecting its regulatory potential†

Kai Cui,‡[ab] Qing Lyu,‡[ab] Naihan Xu,[b] Qing Liu,[ab] Jiarong Zhang,[b] Wei Xing,[b] Linfu Bai,[b] Meijian Liao,[ab] Jie He,[b] Bo Yuan,[c] Deheng Chen,[d] Weidong Xie*[b] and Yaou Zhang*[b]

Received 4th July 2014,
Accepted 6th September 2014

DOI: 10.1039/c4ib00156g

www.rsc.org/ibiology

The regulation of gene expression by microRNAs (miRNAs) is complex due to a number of variables involved. The potential for one miRNA to target many genes, the presence of multiple miRNA response elements (MREs) in one mRNA molecule and the interplay between RNAs that share common MREs each add a layer of complexity to the process; making it difficult to determine how regulation of gene expression by miRNAs works within the context of the system as a whole. In this study, we used luciferase report vectors inserted with different 3′UTR fragments as probes to detect the repressive effect of the miRNA pool on gene expression and uncovered some essential characteristics of gene regulation mediated by the miRNA pool, such as the nonlinear correlative relationship between the regulatory potential of a miRNA pool and the number of potential MREs, the buffering effect and the saturating effect of the miRNA pool, and the restrictive effect caused by the density of MREs. Through expressing gradient concentration of 3′UTR fragments, we indirectly detected the regulatory potential of the competing endogenous RNA (ceRNA) pool and analysed its effect on the regulatory potential of the miRNA pool. Our results provide some new insights into miRNA pool mediated gene regulation.

Insight, innovation, integration

Considerable studies have focused on the regulation of a single miRNA on its multiple target genes. However, it is difficult to measure the effects of the entire cellular miRNA pool on gene expression, due to the complex nature of gene regulation and the lack of suitable research methods. In this study, we generated a simple method by using 3′UTR luciferase report vectors to detect the repressive effect of miRNA pool on gene expression and uncovered some essential characteristics of gene regulation mediated by the miRNA pool. We also indirectly detected the regulative potential of ceRNA pool for some genes and the interplay between miRNA pool and ceRNA pool. Our results provide some new insights about miRNA pool mediated gene regulation.

Introduction

MicroRNAs (miRNAs) are a class of small non-coding RNA molecules that play a major role in the regulation of gene expression. First discovered in *Caenorhabditis elegans* and subsequently found in many other multi-cellular organisms,[1–3] they have been implicated in a broad range of cellular processes; most notably cellular proliferation, differentiation, morphogenesis, development, angiogenesis and tumorigenesis.[4–8]

Computational predictions of potential miRNA binding sites within a genome indicate that one miRNA may target ten to a few hundred genes and one gene may be the target of several or hundreds of miRNAs.[4,7–13] At any given time, an individual gene has the potential to be regulated by all of the accessible miRNAs in which respective miRNA response elements (MREs) are present within this gene. The entire pool of endogenous miRNAs, whether known or unknown, which are expressed in a cell is referred to as the miRNA pool. A considerable body of research exists in which the functional association between an individual miRNA and its mRNA target has been characterized; however, the complex process of how this miRNA pool as a whole regulates gene expression is still relatively unknown. An added layer of complexity comes from recent studies indicating

[a] School of Life Sciences, Tsinghua University, Beijing, 100084, P. R. China
[b] Key Lab in Healthy Science and Technology, Division of Life Science, Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, P. R. China. E-mail: zhangyo@sz.tsinghua.edu.cn, xiewd@sz.tsinghua.edu.cn; Fax: +86 0755 26036884; Tel: +86 0755 26036884, +86 0755 26036458
[c] Intelligent Computing Lab, Division of Informatics, Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, P. R. China
[d] Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, P. R. China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4ib00156g

‡ These authors contributed equally to this work.

that the MRE of RNA also plays a major role in regulating gene expression. RNA molecules that share MREs can modulate each other's expression by competing for available miRNAs and are referred to as competing endogenous RNAs (ceRNAs).[14–20] The pool of all endogenous RNAs, including coding, non-coding, known and unknown RNAs, which share at least one common MRE with a specific gene, is referred to as that gene's ceRNA pool. The interaction between the ceRNA and the miRNA pool and the overall effects on the regulation of gene expression remain poorly understood due to their complexity and a lack of appropriate research methods.

This study focuses on the regulation of gene expression mediated by the miRNA pool and how different factors affect the regulatory potential of the miRNA pool. In an attempt to simplify this complex process, we first examined the effects that the miRNA pool has on a single target. Different sized fragments of 3′UTRs from different genes were linked to a luciferase report vector. The regulatory potential of the miRNA pool to each 3′UTR fragment was detected by measuring the luciferase activities. Then, we studied which factors could have an effect on the regulatory potential of the miRNA pool and analyzed the characterization of the relationship between the affecting factors and the regulatory potential of the miRNA pool. Finally, we investigated how the ceRNA pool affected the regulatory potential of the miRNA pool and measured the maximal regulative potential of a gene's ceRNA pool. Overall, some new essential features of miRNA-mediated gene regulation were uncovered in this investigation.

Results

The regulatory potential of the miRNA pool on gene expression

The overall regulatory potential of the miRNA pool on a given gene is defined as the net repressive effect of gene expression resulting from the direct interaction between 3′UTR and the available miRNAs within the miRNA pool. In this work, we established a system to study the regulatory potential of the miRNA pool. First, a set of constructs were designed in which 3′UTR fragments, taken from 23 different genes, were inserted downstream of the stop codon of the *Renilla* gene in the pRL-TK luciferase report vector (Table S1, ESI†). In order to avoid the influence caused by specific gene function, we chose 3′UTR fragments from genes with diverse cellular functions including those that play a role in the cell cycle, proliferation, angiogenesis, aerobic or anaerobic glycolysis, cytoskeleton, stem cell pluripotency and housekeeping. The lengths of these 3′UTR fragments cover from 131 bp to 3034 bp, and the average length is 962 bp which is similar to the human 3′UTR average length reported to be about 950 bp.[21] The optimal concentration for transfection of the constructs in nasopharyngeal carcinoma (CNE) cells was determined through a dose–response gradient assay in which the output being measured was luciferase activity (Fig. S1, ESI†). The optimal concentration was determined to be 40 ng mL^{-1}, which is far less than the recommend dose (1000 ng mL^{-1}) of a transfection reagent according to the instruction. At this concentration it is unlikely that cellular homeostasis is being drastically perturbed and the luciferase activity was strong enough to detect.[22] The luciferase activity assays showed that the constructs containing different 3′UTR fragments resulted in different levels of measured luciferase activity despite having the same promoter and coding region (Fig. 1A).

MiRNA pool regulation and the buffer effect

The composition of the miRNA pool is an obvious important factor in determining its regulatory potential. Previous results from miRNA microarray assays in CNE cells suggested that culture conditions, such as switching from normoxia to desferrioxamine (DFOM)-induced hypoxia, led to changes in miRNA pool composition (Tables S2 and S3, ESI†).[4] To test the effect that this change in miRNA composition has on its regulatory potential, we compared the luciferase activity of our constructs under each condition. There was no significant change in the luciferase activity between the two conditions tested for most of our constructs, except KLF4, LDHA, PGK1, PTN and VEGFA (Fig. 1A). A Pearson's Chi-square test indicated that there was no significant difference in the composition of the miRNA pool between the two conditions (Fig. 1B). We speculated that the regulatory potential of the miRNA pool tended to be stable if the composition of a miRNA pool does not have a statistically significant change. The result indicated that the miRNA pool may have a buffer effect in keeping cellular homeostasis within a certain extent. Here, the buffer effect is defined as the effect that the miRNA pool can buff the influence of gene expression caused by statistically non-significant change of miRNA pool composition.

To further investigate the effects that significant changes in miRNA pool composition have on its regulatory potential we also carried out experiments using mouse embryonic stem cells (mESCs). The cells, grown under feeder free conditions, were differentiated using all-*trans*-retinoic acid (RA). The degree of differentiation was determined by measuring the expression of Pou5f1 and Sox2 (Fig. S2A and B, ESI†). Both undifferentiated and differentiated cells were collected and sent for miRNA deep sequencing. Reads higher than 5000 resulted in their inclusion in determining the miRNA pool composition for each population (Tables S4 and S5, ESI†). The composition of the miRNA pools between the undifferentiated and differentiated cell populations was shown by Pearson's Chi-square test to be significantly different (Fig. 1C). Fragments of the 3′UTR from 8 different mouse genes were cloned into our luciferase report vectors (Table S1, ESI†). The optimal concentration for transfection was determined (as previously described) to be 40 ng mL^{-1} (Fig. S2C, ESI†) and both undifferentiated and RA-differentiated mESCs were transfected with each construct (Fig. 1D). Unlike our previous results obtained under normoxic and hypoxic conditions in CNE cells, the luciferase activities measured for most constructs were significantly different under the undifferentiated and differentiated conditions. Our results suggested that the miRNA pool may change its regulatory potential for certain genes to meet the need for certain function when the change of miRNA pool composition is statistically significant.

The factors affecting the regulatory potential of the miRNA pool

As shown above, the composition of the miRNA pool directly affects its regulatory potential, however, it is reasonable to assume that only the miRNAs present within the pool that also

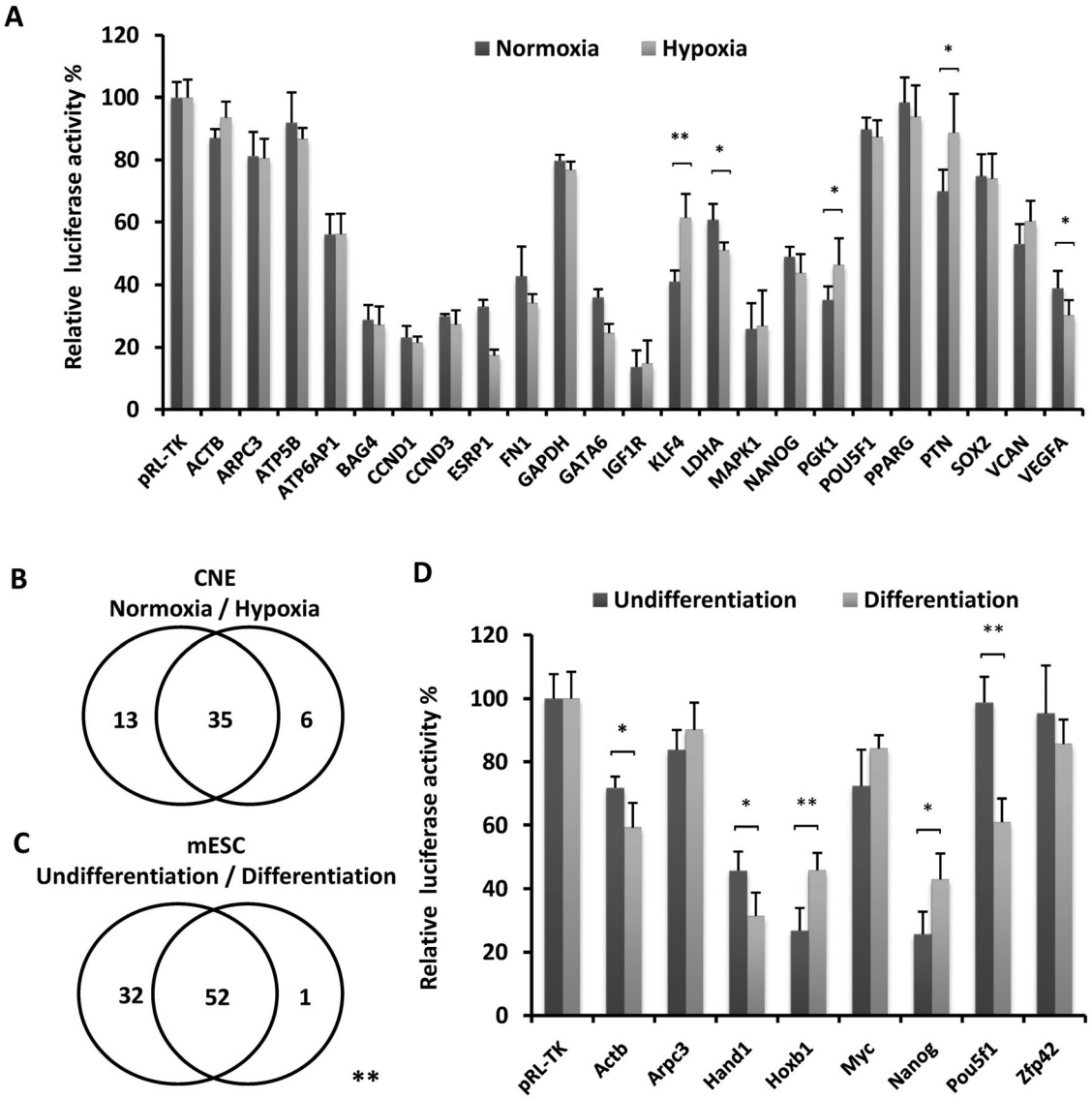


Fig. 1 The regulatory potential of the miRNA pool. (A) 3′UTR fragments from different genes were inserted into the luciferase report vector, pRL-TK, to construct molecular probes. CNE cells were transfected with these constructs under normoxic and DFOM-induced hypoxic conditions. Then, the cells were collected and luciferase activity was measured. *T* test was carried out. $N = 9$, $^{*}P < 0.05$, $^{**}P < 0.01$. (B) Composition comparison of the miRNA pools from normoxic and DFOM-induced hypoxic CNE cells. The comparison was performed using Pearson Chi-square analysis and the result showed no significant difference, $P > 0.05$. (C) Composition comparison of miRNA pools from mESCs with or without RA-induced differentiation conditions using deep sequencing data. Pearson Chi-square analysis was performed, and the result showed a significant difference, $^{**}P < 0.01$. (D) The comparison of luciferase activity between undifferentiated and RA-induced differentiated conditions in mESCs. *T* test was carried out. $N = 9$, $^{*}P < 0.05$, $^{**}P < 0.01$.

have respective MREs present in the 3′UTR will have the ability to directly regulate that gene. Therefore, the presence of an MRE is an obvious pre-requisite for miRNA-mediated regulation. The presence of an MRE does not always imply that an interaction can occur as the location of the MRE can also play an important role. Overlapping seed regions for different MREs would result in an increased probability that only the miRNAs with the most optimal binding would interact, and these MREs were named as potential MREs. Thus, potential miRNA–MRE interactions would out-compete resulting in little to no effect of other miRNAs on the regulation of that gene. The density of potential MREs, in which multiple miRNAs have the potential to bind to the 3′UTR simultaneously, could also affect the regulatory potential of the miRNA pool. The length of 3′UTR fragments and the average free energy of potential MREs may also be affecting factors and deserve to be verified.

In order to investigate how MREs can alter the regulatory potential of the miRNA pool, first we determined all of the relevant MREs for each 3′UTR of the genes in the constructs. To reduce false positive predicted MREs present, we integrated three different prediction algorithms; FindTar,[23] miRanda software[24] and microTar,[25] and generated the CoPrediction software (Fig. 2A). The sequences of the 3′UTR fragments and the sequences of the miRNAs expressing in normoxic and hypoxic

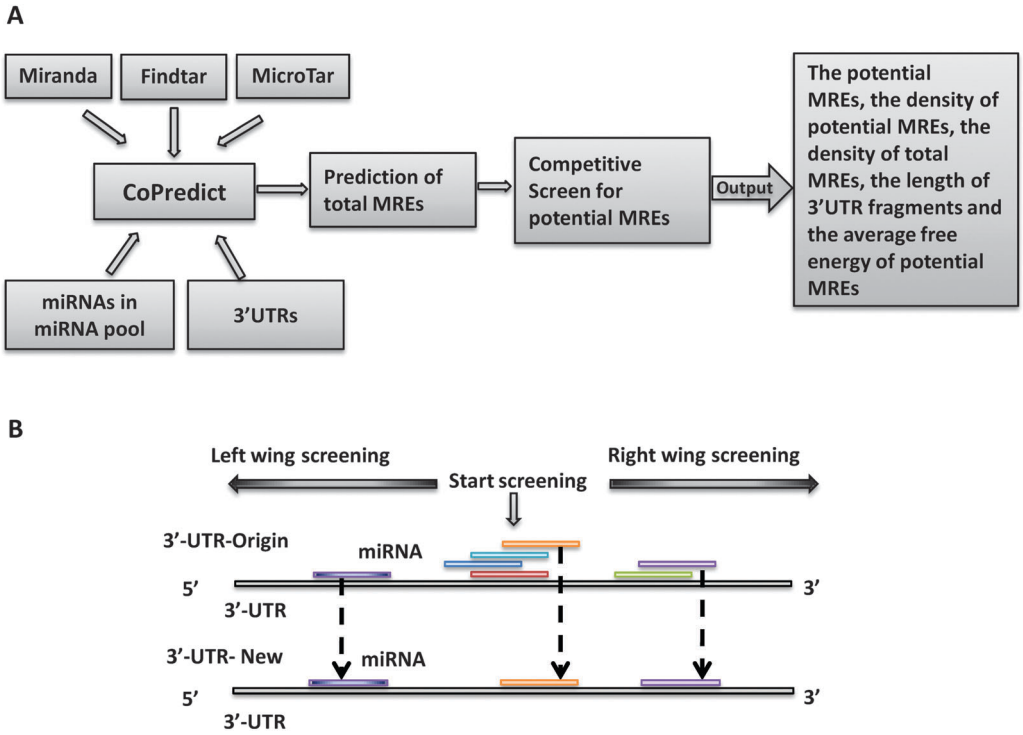


Fig. 2 Prediction of total MREs and potential MREs. (A) Three target prediction algorithms, FindTar, MiRanda and microTar were integrated into CoPrediction. The total MREs, the potential MREs, the density of total MREs, the density of potential MREs, the length of 3′UTR fragments and the average free energy of potential MREs were output from this algorithm. (B) Competitive screen was performed to determine the potential MREs with optimal binding parameters, and the other miRNAs with the seed region overlapping with the potential MREs′ were removed.

CNE cells as well as undifferentiated and differentiated mESCs (Tables S1–S5, ESI†) were uploaded into CoPrediction. MREs that were predicted by FindTar and at least one of the other two prediction algorithms were considered to be part of the total MREs present (Tables S6–S9, ESI†). To exclude the confounding effects of overlapping MREs, a competitive screen was done to determine the potential MREs, and the other miRNAs with the seed region overlapping with the potential MREs′ were removed from further analyses (Fig. 2B and Table S10, ESI†)

Experimental analysis was then focused on how these six factors; the total MREs, the potential MREs, the density of total MREs, the density of potential MREs, the length of 3′UTR fragments and the average free energy of potential MREs, affect the regulatory potential of the miRNA pool (Tables S6–S9, ESI†). The correlative relationship between the six parameters and the fluorescence repression ratio (= 1-relative % of luciferase activity) was determined using a Spearman correlation analysis. The potential MREs (Spearman $r = 0.87$, $P = 3.29 \times 10^{-20}$, $n = 62$) and total MREs (Spearman $r = 0.85$, $P = 1.95 \times 10^{-18}$, $n = 62$) had a strong correlative relationship with the fluorescence repression ratio. A nonlinear regression analysis, using the model $y = a \times x/(b + x)$ which was fitted to the two correlative relationships,[26] indicated that the potential MREs had the highest R^2 value which was determined to be of 0.77 ($n = 62$) (Fig. 3).

Characterization of the regulatory potential of the miRNA pool

The regression curve for the potential MREs and the fluorescence repression ratio suggests that the relationship between the regulatory potential of the miRNA pool and potential MREs is not linear, but rather, curvilinear. To further analyze this possibility, the regression curve was segmented into several fragments (Fig. 4A) and the average repression ratios per potential MRE (the slope rate) were calculated (Fig. 4B). The calculated slope rates for the different fragments varied considerably; with fragment 1 having a slope rate of 5.52 and fragment 5 having a slope rate of 0.34. Fragment 1, with the highest slope rate, was determined to have a regulatory potential of 27.61% and was comprised of genes with 0 to 5 potential MREs. An increase or decrease by 1 MRE resulted in approximately a 5.52% change in the regulatory potential of the miRNA pool. As shown for fragment 2, when the number of potential MREs present was greater than 5, an increase or decrease by 1 MRE only resulted in a 3.15% change in the regulatory potential of the miRNA pool. This change dropped steadily from 1.73 down to 0.34% in fragments 3 to 5. Thus, the data indicate that the regulatory potential of the miRNA pool quickly strengthened when the number of potential MREs increases from one to five, however, the increase gradually slowed down when the potential MREs grew more (Fig. 4A and B).

The relationship between the density of potential MREs and the fluorescence repression ratio per potential MRE was shown to be weakly correlated (Spearman $r = -0.37$, $P = 0.004$, $n = 49$). It is interesting to note that the density of more than 75% of potential MREs present in those 3′UTRs is found to be less than 20 per kb (Fig. 4C). Genes with an MRE density of less than 20 were shown on average to have a higher fluorescence repression ratio per total MRE as compared to genes with an MRE density

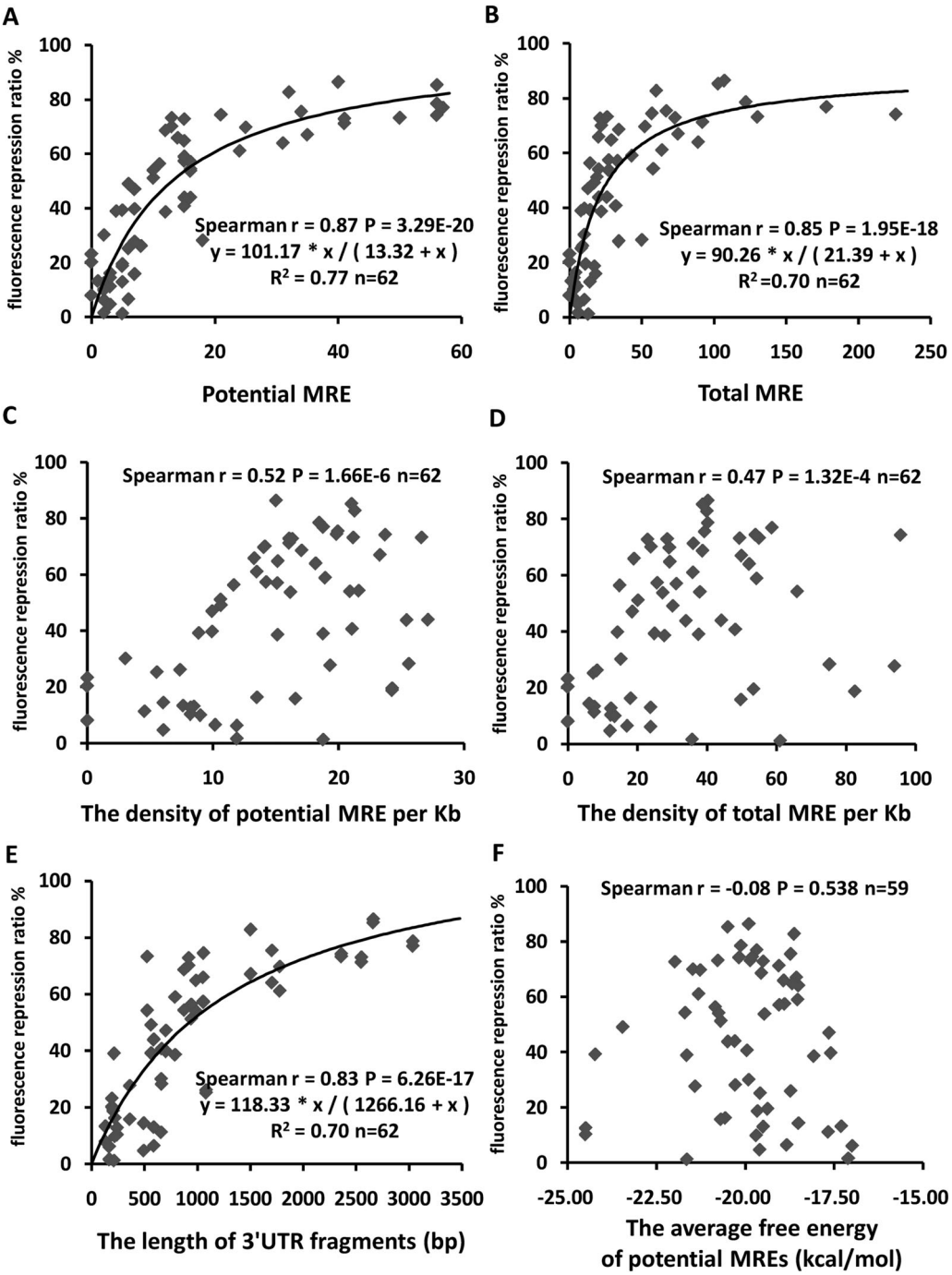


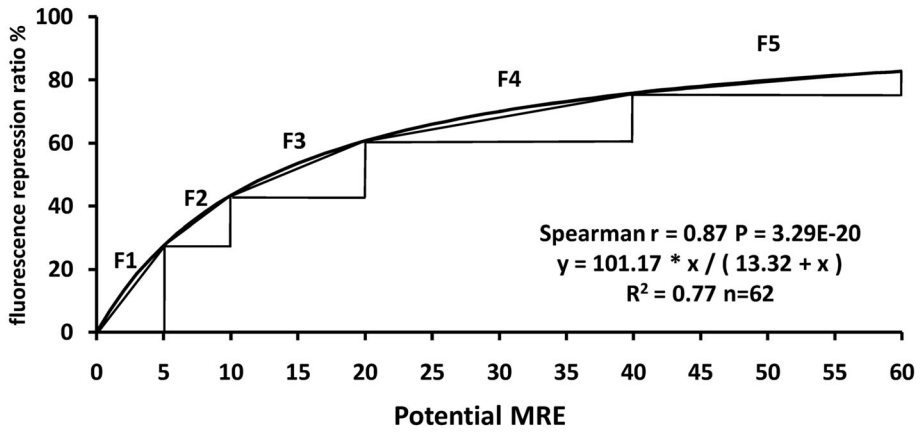
Fig. 3 The factors affecting the regulation of the miRNA pool and their relationship with the regulatory potential of the miRNA pool. The factors are (A) the potential MREs, (B) the total MREs, (C) the density of potential MRE, (D) the density of total MRE, (E) the length of 3′UTR fragments and (F) the free energy of potential MREs.

of more than 20, which had a strong correlative relationship with the fluorescence repression ratio per total MRE (Spearman $r = -0.67$, $P = 1.08 \times 10^{-8}$, $n = 58$) (Fig. 4D).

Effects of the ceRNA pool on the regulatory potential of the miRNA pool

RNA molecules that share common MREs, referred to as ceRNAs, can modulate each other's expression by directly competing for miRNA binding, thus also playing an important role in the regulation of gene expression.[14–20] In order to study the effects of the ceRNA pool on the regulatory potential of the miRNA pool, exogenous luciferase report constructs linked to fragments of different 3′UTRs were expressed in cells to compete with the respective gene's endogenous 3′UTR and its ceRNA pool. Previous microarray and real-time PCR data validated that the following genes: ACTB, ATP6AP1, ARPC3, CCND1, CCND3, GAPDH, LDHA,

A



B

Fragment	Interval potential MREs	Fluorescence repression ratio of each fragment (%)	Average fluorescence repression ratio per potential MRE (%) (the slope rate)	P value
1	0-5	27.61	5.52	/
2	5-10	15.77	3.15	0.276
3	10-20	17.34	1.73	0.0787
4	20-40	15.17	0.76	0.0277*
5	40-60	6.89	0.34	9.97E-03**

C D

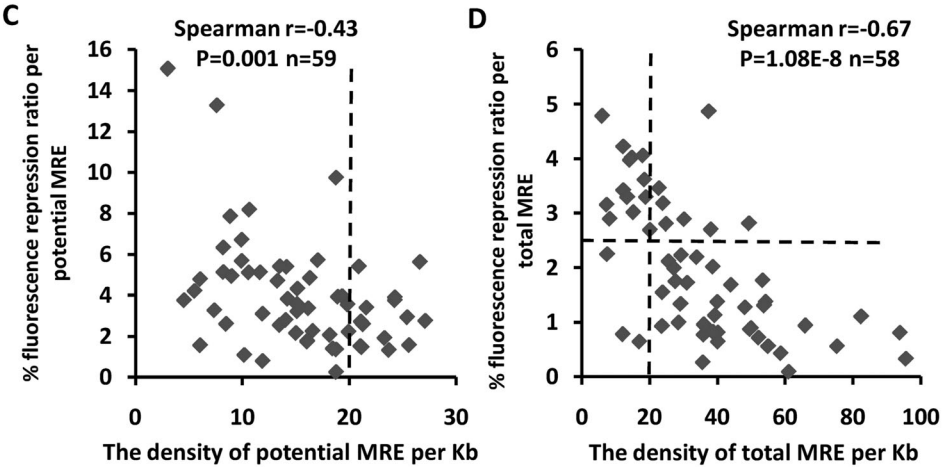


Fig. 4 Analysis of the characteristics of gene regulation mediated by the miRNA pool. (A) The relationship between regulatory potential of the miRNA pool and the potential MREs is curvilinear. The curve was divided into several fragments to calculate the average fluorescence repression ratio per potential MRE (the slope rate). (B) The repression ratio and the average repression ratio per potential MRE (the slope rate) of each fragment in the curve are shown. *$P < 0.05$, **$P < 0.01$. (C) The analysis of the relationship between the fluorescence repression ratio per potential MRE and the density of potential MREs. (D) The analysis of the relationship between the fluorescence repression ratio per total MRE and the density of total MREs.

PGK1, POU5F1 and VEGFA were all endogenously expressed in CNE cells (Table S11 and Fig. S3, ESI†); therefore, fragments of the 3′UTR from these genes were selected. Increasing concentrations of each construct were transfected into CNE cells and the effects that these constructs had on the endogenous expression of their respective gene were determined. The concentrations used ranged from 40 to 480 ng mL^{-1}; this range was shown to have a positive correlative relationship between the luciferase activity and luciferase mRNA levels (Fig. S1 and S4, ESI†). Increasing the exogenous expression of the 3′UTR fragment for a given gene was shown to gradually increase the endogenous expression of that respective gene. Endogenous gene expression was shown to plateau when the exogenous 3′UTR was transfected at concentrations between 160 and 400 ng mL^{-1} (Fig. 5A–I). These results suggest that increasing the abundance of a 3′UTR decreases the regulatory potential of the miRNA pool as there may be more targets to sequester miRNAs thus leading to an increase in expression of that respective gene; however, the increase cannot be sustained when the 3′UTR concentration is high enough. The results also suggest that in the

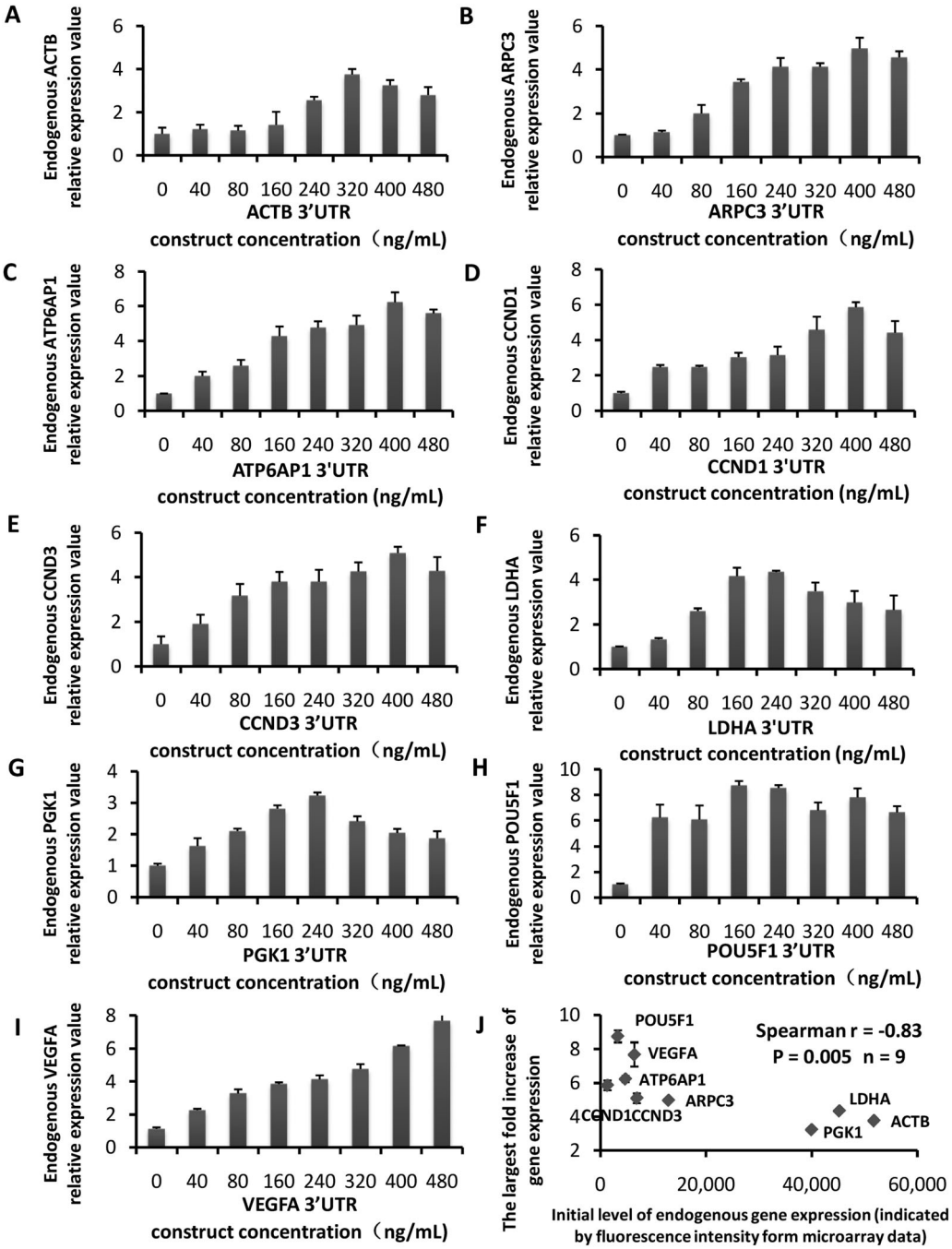


Fig. 5 The regulatory potential of the ceRNA pool on gene expression. (A–I) CNE cells were collected and the expression of endogenous genes which have the same 3′UTRs inserted into the pRL-TK plasmids was determined using RT-PCR; the relative expression was normalized by B2M. (J) The analysis of the correlative relationship between the largest fold change in the up-regulation of gene expression and the basal level of expression of the endogenous gene (indicated by the fluorescence intensity taken from microarray data).

absence of the positive regulatory effects of the ceRNAs pool the miRNA pool would exert greater repressive effects on the expression of these genes.

Exogenous expression of each of the 3′UTR constructs was shown to lead to an increase in the expression of their respective gene, with the exception of GAPDH. GAPDH has a short 3′UTR with no predicted MREs; therefore, it is not surprising that even upon increasing the exogenously expressed levels of the luciferase-GAPDH 3′UTR report construct there was no subsequent change in the endogenous GAPDH expression (Fig. S5A and B, ESI†). Although an increase in endogenous gene expression was observed for all other genes, the amount of up-regulation varied between genes; ranging from 3.75 to 8.74 times greater than basal levels (Fig. 5A–I). Further analysis was done to determine the relationship between the greatest change in endogenous gene expression and the basal levels of gene expression based on our

microarray data (Table S11, ESI†).[4] To investigate the cause of this phenomenon, we hypothesized that genes with low abundance may be sensitive to the expression changes. As we expected, a strong correlative relationship was found to exist between the initial level of endogenous gene expression and the largest fold change in gene expression mediated by the presence of increasing 3′UTR constructs (Spearman $r = -0.83$, $P = 0.005$, $n = 9$) (Fig. 5J and Table S12, ESI†). Our results indicated that a lower initial level of endogenous gene expression led to greater fold changes of its expression in the presence of competition for miRNAs in the miRNA pool. This conclusion is also supported by the dilution effect.[26]

Discussion

Establishing a method with 3′UTR constructs to detect the regulatory potential of the miRNA pool

The regulation of gene expression by miRNAs is a complex process. One miRNA may have the potential to regulate as few as ten and as many as a few hundred different genes; whereas one gene may also have the potential to be targeted by multiple miRNAs.[4,7–13] Studying the effects of a single miRNA on the regulation of gene expression has been investigated utilizing genome-wide approaches such as cDNA microarrays and proteomic analyses.[9,10,27,28] While focusing on studying single miRNA–RNA interactions has resulted in a wealth of information regarding the pathway of miRNA mediated gene regulation, these approaches fail to capture the effects that the entire miRNA pool has on the regulation of gene expression within the cell. These effects have been difficult to determine due to the complex nature of gene regulation and the lack of suitable research methods.

In this investigation, we established a method to study the regulatory potential of the miRNA pool and how other cellular factors affect this regulatory potential. Considering that MREs are predominantly located within 3′UTRs,[3] we designed reporter plasmids in which the luciferase gene was linked to fragments of different 3′UTRs that were amplified from our genes of interest. These 3′UTR reporter plasmids allow for the investigation and comparison of the regulatory effects that the same cellular miRNA pool has on different genes and can also be used to determine how altering the composition of the miRNA pool affects its regulatory potential.

To rule out any artefacts due to the regulation of reporter expression, the coding sequence for the luciferase gene as well as the promoter was exactly the same between reporter plasmids. Thus the only difference between the reporter plasmids was the presence of varying 3′UTR fragments. Also, any effects relating to the interaction between the 3′UTR and the polyadenylation signal related protein complex should be a common feature shared among the reporters because they all have the same SV40 late poly A tract. Other RNA-binding proteins could be preferentially interacting with specific 3′UTRs;[29–31] such as the Zipcode binding protein-1 and AU-rich element (ARE) binding proteins, which could have the potential to affect our results. However, the Zipcode binding protein-1 interacts preferentially with ACTB mRNA and is unlikely to be affecting other genes. Our results did show that the expression of each of the endogenous genes studied gradually increased and reached a plateau with increasing concentrations of our exogenously expressed 3′UTR reporter plasmids with the exception of the VEGFA gene (Fig. 5A–I). This could be due to the presence of AU-rich elements in the 3′UTR of this gene. Even taking into consideration the existence of certain caveats, this experimental design has the power to detect the regulatory potential of the miRNA pool. Utilizing this method, we not only confirmed existing theories on the regulatory potential of the miRNA pool and also revealed some novel characteristics of miRNA pool mediated gene regulation.

Characteristics of miRNA pool mediated gene regulation

Previous genome wide studies demonstrated that a single miRNA can repress hundreds of proteins; however, this is a subtle effect that rarely exceeds a 2-fold change from basal levels.[9,10] Results from our study indicate that miRNA pool mediated repression of gene expression is easy to exceed this value (Fig. 1A). This suggests that miRNA regulation to a single gene might rarely be carried out by a single miRNA but should be a multiple regulation executed by total putative regulatory miRNAs in a cellular miRNA pool.

In this article, we uncovered four novel characteristics of miRNA pool mediated gene regulation, including the nonlinear correlative relationship between the regulatory potential of the miRNA pool and the number of potential MREs, the buffering effect, the saturating effect, and the restrictive effect caused by the density of MREs.

Zhen Yang *et al.* reported the buffering role of miRNAs, but it was referred to that miRNAs can buffer the fluctuation of gene expression.[32] The buffering effect of the miRNA pool reported in this investigation means that the regulatory potential of the miRNA pool for gene expression tended to be stable when the change of miRNA pool composition is not statistically significant. We speculate that the buffer effect is quite important. When the changes in an internal or an external environment of cells are not significant enough to cause large difference of the composition of the miRNA pool, the regulatory potential of the miRNA pool will remain relatively stable because of the buffer effect, which let the miRNA pool gets the ability to play an important role in keeping cellular homeostasis within a certain extent.

The physiological significance of nonlinear correlative relationship between the regulatory potential of the miRNA pool and the number of potential MREs is not very clear, but it may be related to a saturating effect.[33,34] The regulatory potential of the miRNA pool quickly increases when the number of potential MRE rises from 0 to 5, and the increase gradually slows down when the number of potential exceeds 5.

Our results also demonstrated that the density of potential MREs in the 3′UTR was on average less than 20 per kb for the majority of genes. Genes with an MRE density less than 20 per kb typically had a higher repression ratio per total MRE. This is most likely due to the occupancy restrictions of the Ago–mRNA complex. The Ago footprint on mRNA has been determined to

span approximately 46–62 nucleotides;[35] thus a 3′UTR 1 kb in length could only interact with 16–21 RISC complexes simultaneously. This suggests that the upper limit for an optimal MRE density is 16–21 MREs per kb and that the average repression ratio per MRE may decrease with the addition of MREs beyond the limit.

The effect of ceRNAs on the regulatory potential of the miRNA pool

As previously mentioned, miRNA-targeted RNA transcripts that share common MREs have the ability to modulate each other's expression. Through this MRE language, the communication within the network could influence miRNA pool mediated gene regulation by sequestering miRNAs in the cellular miRNA pool.[36–38] However, it is still unclear how to measure the effect ceRNA pool has on the gene regulation mediated by the miRNA pool. Our results demonstrated that increasing concentrations of 3′UTR constructs led to a gradual increase in the expression of the corresponding endogenous gene and that this expression would eventually plateau; meanwhile, the regulatory potential of the miRNA pool would decrease most likely because the availability of free miRNAs subsequently decreased and finally exhausted. This can be described with the following model: 'Y = F(M) + G(M)', where the plus sign '+' does not indicate the addition of these two variables, instead referring to the outcome of the interaction between F(M) and G(M). ''Y'' is defined as the observed level of endogenous X gene's expression; M is defined as the concentration of the X genes 3′UTR report construct. The value of F(M) represents the positive regulatory potential of X gene as mediated through the ceRNA pool and G(M) represents the negative regulatory potential of X gene by the miRNA pool. At a certain concentration of M, referred to as M_1, the value of $F(M_1)$ increases enough to theoretically allow $G(M_1)$ to reach zero. When $G(M_1) = 0$, $Y = F(M_1)$; where $F(M_1)$ is the maximal regulative potential of the ceRNA pool, and in this case the value of $F(M_1)$ is equivalent to the value of Y (Fig. 6). For example, when the 3′UTR reporter containing the 3′UTR from PGK1 reached a concentration of 240 ng mL^{-1}; the expression of endogenous PGK1 peaked and was calculated to be about 3.25 times greater than its basal expression. This indicates that the ceRNA pool only has the potential to vary the expression of PGK1 3.25 times its basal

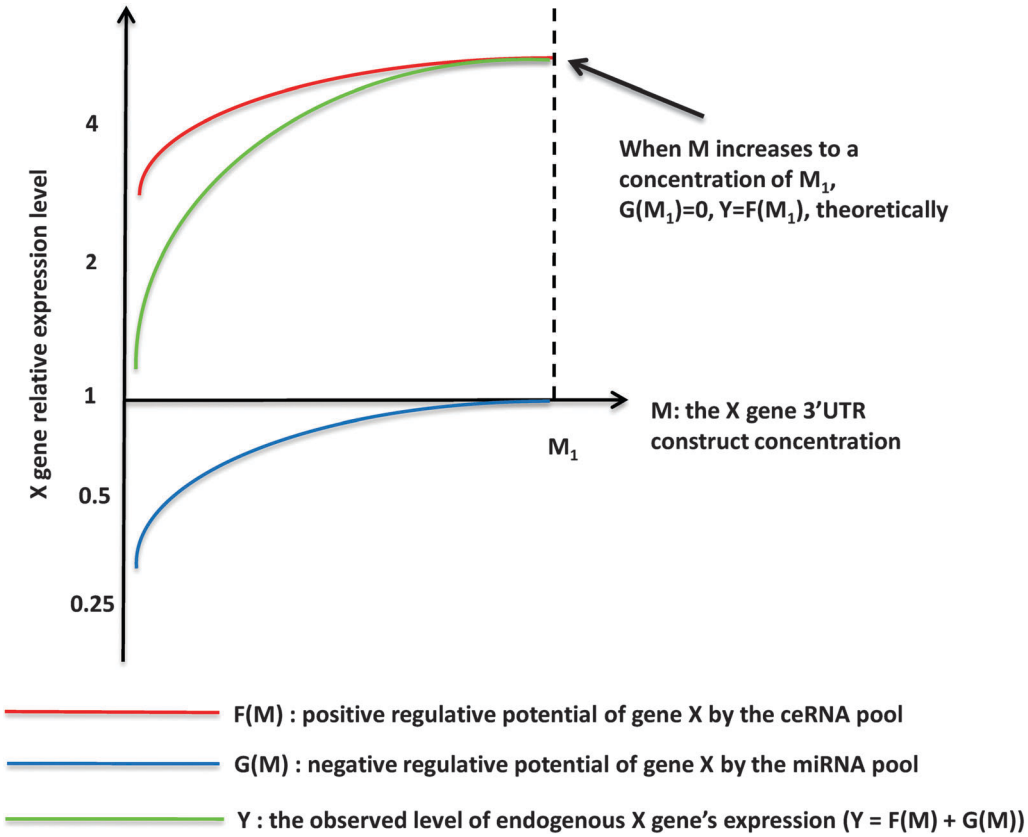


Fig. 6 A mathematical model regarding gene regulation mediated through the interaction between the miRNA pool and the ceRNA pool. The model 'Y = F(M) + G(M)' is used to describe the gene regulation mediated by the interaction between the miRNA pool and the ceRNA pool. ''Y'' is defined as the observed level of endogenous X gene's expression; M is defined as the X gene 3′UTR construct concentration. F(M) represents the positive regulative potential of X gene by the ceRNA pool and G(M) represents the negative regulative potential of X gene by the miRNA pool. When M increases to a concentration of M_1, the amount of $F(M_1)$ is large enough, and $G(M_1)$ theoretically reaches zero. When $G(M_1) = 0$, $Y = F(M_1)$. Where $F(M_1)$ represents the maximal regulative potential of X gene by its ceRNA pool, the value of $F(M_1)$ could be determined by the detection the value of Y. It is assumed that the transcriptional regulation is constant as M changes.

level of expression. However, it is important to note that this model assumes that the transcriptional regulation remains constant as M changes.

Materials and methods

Cell culture

CNE cells from a human nasopharyngeal carcinoma cell line (Kunming Cell Bank, Kunming, China) were cultured in high glucose DMEM medium containing 10% FBS at 37 °C with 5% CO_2. Hypoxia was induced with DFOM (D9533, Sigma-Aldrich Co., MO, USA) at a final concentration of 130 μM. Cells without DFOM treatment were used as normoxia controls.

R1 Mouse embryonic stem cells (mESCs) (Cyagen Co. Guangzhou, China) were grown on Mitomycin (Roche Co. USA) treated MEF feeders and in the presence of 1000 U mL^{-1} LIF (ESG1106, Millipore, CA, USA). The differentiated mESCs were maintained for 72 hours in media containing 10^{-7} mol L^{-1} all-*trans*-retinoic acid (RA) (R2625, Sigma-Aldrich Co., MO, USA) without LIF.[39]

miRNA microarray and deep sequencing experiments

Total RNA from CNE cells was sent to CapitalBio (CapitalBio Corp., Beijing, China) for miRNA microarray experiments. Procedures were performed as described in detail on the website of CapitalBio (http://www.capitalbio.com). Briefly, 50–100 mg of total RNA was used for miRNA extraction using an miRNA Isolation Kit (AM1560, Ambion Inc. Texas). Fluorescein-labeled miRNA was used for the hybridization on each miRNA microarray chip containing 509 probes in triplicate,[40] corresponding to 435 human (including 122 predicted miRNAs), 261 mouse, and 196 rat miRNAs found in the miRNA Registry (http://microrna.sanger.ac.uk/sequences/; accessed Oct. 2005) or collected from a published paper.[41] Raw data were normalized and analyzed using GenePix Pro 4.0 software (Axon Instruments, PA, USA). Expression data were median centered by using the global median normalization of the BIOCONDUCTOR package (http://www.bioconductor.org). Statistical comparisons were done using the SAM software (SAM version 2.1, http://www-stat.stanford.edu/,tibs/SAM/index.html).[42,43] Some results of this microarray data were confirmed by real-time PCR (Fig. S3, ESI†).

Total RNA from mESCs was sent to Beijing Genomics Institute (Beijing, China) for miRNA deep sequencing experiments. RNAs 18–30 nucleotides in length were isolated from electrophoresis gel. 5′ and 3′ adapters were sequentially ligated to these RNAs. Reverse transcription PCR was performed to generate the first cDNA chain. The small RNA digitalization analysis based on solexa high-throughput sequencing was taken when the second cDNA chain was generated. After discarding the 5′ adaptor, contaminant and low quality reads, the final reads were annotated.

3′UTR datasets and microRNA datasets

3′UTR sequences of human (Homo sapiens) and mouse (Mus musculus) were retrieved using the Ensembl Data base (http://www.ensembl.org) (version 73 from September 2013). Human and mouse microRNA sequences (miRBase 20) were retrieved from the miRBase website (http://www.mirbase.org/).[44,45]

Construct generation

Luciferase report vectors were constructed by inserting full length 3′UTRs or 3′UTR fragments from different genes downstream of the stop codon in the pRL-TK plasmid (E2241, Promega Corp., Madison, WI, USA). Briefly, Total RNA was isolated from CNE cells or mESCs with RNA isoplus (9109, TaKaRa Bio Inc, Japan) according to the manufacturer's protocol. Residual DNA was degraded with Recombinant DNase I (2270A, TaKaRa Bio Inc, Japan). RT-PCR was performed with the TaKaRa one step RNA PCR kit (RR024A, TaKaRa Bio Inc, Japan). 3′UTRs were amplified using the primers listed in Table S1 (ESI†).

Luciferase assays and RT-PCR

CNE cells were seeded onto 24-well plates. The constructs were transfected the following day using Lipofectamine 2000 according to the manufacturer's instructions (11668019, Life technology). Cell lysates were collected 30 h post transfection. *Renilla* luciferase levels were measured using a Luciferase Reporter Assay System (E1960, Promega, USA) and each experiment was repeated in triplicate. 10 ng pGL3-Basic Vector (E1751, Promega, USA) was co-transfected in each assay and the firefly luciferase activity was used to normalize the *Renilla* luciferase activity.

To detect the endogenous gene expression, CNE cells were seeded onto 6-well plates. Plasmid transfection was done as described above. Total RNAs were harvested 36 hours after changing transfection media. RT-PCR was performed as previously described. We used SYBR Green real-time PCR (QPK-212, TOYOBO, OSAKA, Japan) to quantify the endogenous gene expression; the primes are listed in Table S13 (ESI†). The assays done using mESCs were the same as those done using the CNE cells, except that the transfection reagent used was X-tremeGENE HP (06366236001, Roche, Co. USA).

MRE prediction

To reduce the false-positive rate of MRE prediction, we generated a software CoPrediction (Fig. 2A), through the integration of three target prediction algorithms, miRanda software (http://www.microrna.org/microrna/home.do),[24] microTar (http://mirna.imbb.forth.gr/microinspector/),[25] and FindTar, a bioinformatics approach designed by our Lab (http://bio.sz.tsinghua.edu.cn/findtar/).[23] MREs predicted by FindTar and one of the other two bioinformatics algorithms were selected as total MREs. To avoid the overlap of total target miRNA seed regions, we did a competitive screen for potential MRE using the following methods: (1) we generated a 3′-UTR-origin map with all of the MREs and their parameters, including the minimum free energy, dynamic programming scores and central loop scores; (2) then we removed the MREs from the 3′-UTR origin to generate a new map, named 3′-UTR-new; (3) we screened all of the MREs of the 3′-UTR-origin to find the best MRE with the highest parameter score; (4) projected the best MRE into the same location of 3′-UTR-new. (5) Removed the other MREs whose seed sequence

overlapped with the best MRE on the 3′-UTR-origin; (6) the left side and the right side of the UTR of the best MRE were renamed UTR-origin-left and UTR-origin-left; (7) the above steps were repeated on the UTR-origin-left and UTR-origin-left until no MRE could be screened; (8) all the potential MREs were taken from the 3′-UTR-new map (Fig. 2B).

Statistical analysis

Pearson's Chi-square test was performed to compare the miRNA composition of different miRNA pools under different cellular conditions. A *t*-test analysis was also done.[46] The correlative relationship between the parameters related to 3′UTR regulation and the regulatory potential of the miRNA pool for a gene was analyzed using the Charles Spearman correlation.[47] Nonlinear Regression analysis was performed using SPSS statistics 20. A *p*-value less than 0.05 is considered statistically significant. A *p*-value less than 0.01 is considered statistically highly significant. A strong correlative relationship exists when the absolute value of the Spearman correlation coefficient is between 0.7 and 1.0. A weak correlative relationship is considered when the absolute value of the Spearman correlation coefficient is between 0.4 and 0.7.

Conclusions

In this study, we used luciferase report vectors inserted with different 3′UTR fragments as probes to detect the repressive effect of the miRNA pool on gene expression and uncovered some essential characteristics of gene regulation mediated by the miRNA pool, including nonlinear correlative relationship between the regulatory potential of a miRNA pool and the number of potential MREs, the buffer effect and the saturation effect of the miRNA pool, and the restrictive effect of density of MREs. Through expressing gradient concentration 3′UTR fragments, we indirectly detected the regulative potential of the ceRNA pool for some genes and analysed its effect on the regulatory potential of the miRNA pool. In short, our data suggest that the regulative potential of the miRNA pool is strongly affected by the number of potential MREs in 3′UTRs. The gene regulation mediated by the miRNA pool has the buffering and saturating effects, and is negatively affected by the ceRNA pool.

Acknowledgements

This work was supported by China-Canada Joint Health Research Grant (from the National Natural Science Foundation of China and Canadian Institutes of Health Research to YZ and BBY, No. 81261120556) and Key Lab Promotion Grant of Shenzhen (ZDSY20120619141025668). The funding institutes had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Notes and references

1 E. G. Moss, R. C. Lee and V. Ambros, *Cell*, 1997, **88**, 637–646.
2 C. T. Lee, T. Risom and W. M. Strauss, *DNA Cell Biol.*, 2007, **26**, 209–218.
3 D. P. Bartel, *Cell*, 2009, **136**, 215–233.
4 Z. Hua, Q. Lv, W. Ye, C. K. Wong, G. Cai, D. Gu, Y. Ji, C. Zhao, J. Wang, B. B. Yang and Y. Zhang, *PLoS One*, 2006, **1**, e116.
5 S. W. Shan, D. Y. Lee, Z. Deng, T. Shatseva, Z. Jeyapalan, W. W. Du, Y. Zhang, J. W. Xuan, S. P. Yee, V. Siragam and B. B. Yang, *Nat. Cell Biol.*, 2009, **11**, 1031–1038.
6 A. L. Kasinski and F. J. Slack, *Nat. Rev. Cancer*, 2011, **11**, 849–864.
7 J. F. Zhang, W. M. Fu, M. L. He, W. D. Xie, Q. Lv, G. Wan, G. Li, H. Wang, G. Lu, X. Hu, S. Jiang, J. N. Li, M. C. Lin, Y. O. Zhang and H. F. Kung, *RNA Biol.*, 2011, **8**, 829–838.
8 J. He, J. Wu, N. Xu, W. Xie, M. Li, J. Li, Y. Jiang, B. B. Yang and Y. Zhang, *Nucleic Acids Res.*, 2013, **41**, 498–508.
9 D. Baek, J. Villen, C. Shin, F. D. Camargo, S. P. Gygi and D. P. Bartel, *Nature*, 2008, **455**, 64–71.
10 M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin and N. Rajewsky, *Nature*, 2008, **455**, 58–63.
11 S. Gu, L. Jin, F. Zhang, P. Sarnow and M. A. Kay, *Nat. Struct. Mol. Biol.*, 2009, **16**, 144–150.
12 J. He, J. F. Zhang, C. Yi, Q. Lv, W. D. Xie, J. N. Li, G. Wan, K. Cui, H. F. Kung, J. Yang, B. B. Yang and Y. Zhang, *PLoS One*, 2010, **5**, e13558.
13 S. L. Ameres and P. D. Zamore, *Nat. Rev. Mol. Cell Biol.*, 2013, **14**, 475–488.
14 J. M. Franco-Zorrilla, A. Valli, M. Todesco, I. Mateos, M. I. Puga, I. Rubio-Somoza, A. Leyva, D. Weigel, J. A. Garcia and J. Paz-Ares, *Nat. Genet.*, 2007, **39**, 1033–1037.
15 H. Seitz, *Curr. Biol.*, 2009, **19**, 870–873.
16 D. Cazalla, T. Yario and J. A. Steitz, *Science*, 2010, **328**, 1563–1566.
17 L. Poliseno, L. Salmena, J. Zhang, B. Carver, W. J. Haveman and P. P. Pandolfi, *Nature*, 2010, **465**, 1033–1038.
18 J. Wang, X. Liu, H. Wu, P. Ni, Z. Gu, Y. Qiao, N. Chen, F. Sun and Q. Fan, *Nucleic Acids Res.*, 2010, **38**, 5366–5383.
19 Z. Jeyapalan, Z. Deng, T. Shatseva, L. Fang, C. He and B. B. Yang, *Nucleic Acids Res.*, 2011, **39**, 3026–3041.
20 Y. Tay, L. Kats, L. Salmena, D. Weiss, S. M. Tan, U. Ala, F. Karreth, L. Poliseno, P. Provero, F. Di Cunto, J. Lieberman, I. Rigoutsos and P. P. Pandolfi, *Cell*, 2011, **147**, 344–357.
21 P. Sood, A. Krek, M. Zavolan, G. Macino and N. Rajewsky, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 2746–2751.
22 D. Y. Lee, T. Shatseva, Z. Jeyapalan, W. W. Du, Z. Deng and B. B. Yang, *PLoS One*, 2009, **4**, e4527.
23 W. Ye, Q. Lv, C. K. Wong, S. Hu, C. Fu, Z. Hua, G. Cai, G. Li, B. B. Yang and Y. Zhang, *PLoS One*, 2008, **3**, e1719.
24 A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander and D. S. Marks, *Genome Biol.*, 2003, **5**, R1.
25 M. Maragkakis, P. Alexiou, G. L. Papadopoulos, M. Reczko, T. Dalamagas, G. Giannopoulos, G. Goumas, E. Koukis, K. Kourtis, V. A. Simossis, P. Sethupathy, T. Vergoulis, N. Koziris, T. Sellis, P. Tsanakas and A. G. Hatzigeorgiou, *BMC Bioinf.*, 2009, **10**, 295.
26 A. Arvey, E. Larsson, C. Sander, C. S. Leslie and D. S. Marks, *Mol. Syst. Biol.*, 2010, **6**, 363.

27 L. P. Lim, N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley and J. M. Johnson, *Nature*, 2005, **433**, 769–773.
28 A. Grimson, K. K. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim and D. P. Bartel, *Mol. Cells*, 2007, **27**, 91–105.
29 C. H. de Moor, H. Meijer and S. Lissenden, *Semin. Cell Dev. Biol.*, 2005, **16**, 49–58.
30 M. J. Moore, *Science*, 2005, **309**, 1514–1518.
31 W. Zhao, D. Blagev, J. L. Pollack and D. J. Erle, *Proc. Am. Thorac. Soc.*, 2011, **8**, 163–166.
32 Z. Yang, D. Dong, Z. Zhang, M. J. Crabbe, L. Wang and Y. Zhong, *BMC Genomics*, 2012, **13**(suppl 7), S14.
33 M. S. Ebert, J. R. Neilson and P. A. Sharp, *Nat. Methods*, 2007, **4**, 721–726.
34 B. Haley and P. D. Zamore, *Nat. Struct. Mol. Biol.*, 2004, **11**, 599–606.
35 S. W. Chi, J. B. Zang, A. Mele and R. B. Darnell, *Nature*, 2009, **460**, 479–486.
36 L. Salmena, L. Poliseno, Y. Tay, L. Kats and P. P. Pandolfi, *Cell*, 2011, **146**, 353–358.
37 M. Cesana and G. Q. Daley, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7112–7113.
38 P. Sumazin, X. Yang, H. S. Chiu, W. J. Chung, A. Iyer, D. Llobet-Navas, P. Rajbhandari, M. Bansal, P. Guarnieri, J. Silva and A. Califano, *Cell*, 2011, **147**, 370–381.
39 Y. Tay, J. Zhang, A. M. Thomson, B. Lim and I. Rigoutsos, *Nature*, 2008, **455**, 1124–1128.
40 J. M. Thomson, J. Parker, C. M. Perou and S. M. Hammond, *Nat. Methods*, 2004, **1**, 47–53.
41 X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander and M. Kellis, *Nature*, 2005, **434**, 338–345.
42 H. He, K. Jazdzewski, W. Li, S. Liyanarachchi, R. Nagy, S. Volinia, G. A. Calin, C. G. Liu, K. Franssila, S. Suster, R. T. Kloos, C. M. Croce and A. de la Chapelle, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 19075–19080.
43 V. G. Tusher, R. Tibshirani and G. Chu, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5116–5121.
44 S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman and A. J. Enright, *Nucleic Acids Res.*, 2006, **34**, D140–D144.
45 A. Kozomara and S. Griffiths-Jones, *Nucleic Acids Res.*, 2011, **39**, D152–D157.
46 M. P. Fay and M. A. Proschan, *Stat. Surv.*, 2010, **4**, 1–39.
47 S. Lovie and P. Lovie, *Int. J. Epidemiol.*, 2010, **39**, 1151–1153.

