Genome-wide Identification and Analysis of Enhancer Regulated microRNAs Across 31 Human Tissues

Fei Tang[1], Mei Lang[1], Wenzhu Wang[1], Yanjing Li[1], Changying Li[1], Zheng tang Tan[1], Junjie Yue[2]✉, Zhiyun Guo[1]✉

1. School of Life Sciences and Engineering, Southwest Jiaotong University, Chengdu 610031, P.R. China

2. Beijing Institute of Biotechnology, Beijing 100071, P.R. China

✉Corresponding authors: Zhiyun Guo, Email: zhiyunguo@swjtu.edu.cn; Junjie Yue, Email: yue_junjie@126.com

Abstract: Enhancers are cis-regulatory DNA elements that positively regulate the transcription of target genes in a tissue-specific manner and dysregulation in various diseases such as cancer. Recent studies shown that enhancers could regulate miRNAs and participate in the biological synthesis of miRNAs. However, the network of enhancer-regulated miRNAs across multiple cancers are still unclear. Here, we identified a total of 2,418 proximal enhancer-miRNA interactions and 1,280 distal enhancer-miRNA interactions through integration of genomic distance, co-expression and 3D genome data in 31 cancers. The results shown that both proximal and distal interactions exhibit significantly tissue-specific feature and there is a significant positive correlation between the expression of miRNA and the number of regulated enhancers in most tissues. Furthermore, we found that there is a high correlation between the formation of enhancer-miRNA pairs and expression of eRNAs whether in distal or proximal regulation. The characteristics analysis shown that miRes (enhancers that regulated miRNAs) and non-miRes presented significantly different in sequence conservation, GC content and histone modification signatures. Notably, GC content, H3K4me1, H3K36me3 present different between distal regulation and proximal regulation suggesting they may participate in chromosome looping of enhancer-miRNA interactions. Finally, we introduced a case study, enhancer: chr1:1186391-1186507~miR-200a was highly relevant with survival of thyroid cancer patients and a cis-eQTL SNP on enhancer effect the expression TNFRSF18 gene as a tumor suppressor.

Introduction

Enhancers are cis-regulatory DNA regulatory elements that positively regulate the transcription of target genes in a spatiotemporal-specific manner. The dysfunction of enhancer has been considered affect enhancer-promoter communication and cause lots of diseases such as cancer (Thandapani, 2019). Previous studies have shown that enhancer activity is affected by the enhancer RNA (eRNA) which transcribed bidirectionally from active regulatory enhancers, and plays a key role in regulating downstream gene expression. The Functional Annotation of the Mammalian Genome (FANTOM) group used CAGE technology had identified ~ 65,000 active enhancers across multiple tissues, these valuable resources provide important data sources for subsequent research (Andersson et al., 2014). Recently, a large-scale pan-cancer study for TCGA patient samples across 33 cancer types revealed that the enhancer activity affects the expression of a variety of tumor-associated genes and was involved in tumor tumorigenesis (Chen et al.,

2018).MicroRNAs (miRNAs) are a subset of endogenous non-coding RNAs (~22 nucleotides long) which play vital roles in regulating genes expression via targeting the specific sites in 3' untranslated region (3' UTR) of miRNA (Lu & Rothenberg, 2018; Sandiford et al., 2018). In the past years, a large number of literatures has confirmed that miRNAs were involved in almost all known cancers. A recent study shown that miR-24-1 that present in nuclear promote gene expression by targeting enhancers suggesting there was a obviously interaction between enhancer and miRNA. Other recent studies showed that enhancers (including typical enhancers and super enhancers) have been found to regulate miRNA expression and participate in the biological synthesis of miRNAs regulated by Drosha / DGCR8 (Wood et al., 2019; Yun et al., 2018). These studies suggest that enhancers involved in miRNA regulatory networks and play an important role in tumorigenesis and development.

However, the network of enhancer-regulated miRNAs across multiple tumors are still unclear. Therefore, we performed pan-cancer study for enhancer-regulated miRNAs across 33 human cancer types in TCGA. According to the distance between enhancer and miRNA, we classify enhancer-miRNA pairs into two types: proximal and distal enhancer-miRNA regulation. A series of enhancer-miRNA regulation were identified through integration of co-expression, distance information and 3D genomes data of enhancers and miRNAs from 8,693 samples in TCGA. GO and KEGG enrichment shown that target genes of enhancer regulated miRNAs were significantly involved in tumor-associated biological processes and signaling pathway. Furthermore, we found that there is a high correlation between the formation of enhancer-miRNA pairs and expression of eRNAs. The results shown that miRes (enhancers that regulated miRNAs) and non-miRes presented significantly different characteristics including sequence conservation, GC content and histone modification signatures. Several histone modifications show significant cancer specificity and enhancer-miRNA spatial distance specificity. Finally, we introduced a case study, enhancer: chr1:1186391-1186507~miR-200a was highly relevant with survival of thyroid cancer patients and a cis-eQTL SNP on enhancer effect the expression TNFRSF18 gene as a tumor suppressor.

Materials and Methods

1.1 Identification of enhancer-miRNA interactions

Enhancer annotations and expression data for TCGA 33 cancers were downloaded from previous study (Chen et al., 2018). The expression data of miRNAs from all paired tumors and eight adjacent normal tissues were downloaded from the TCGA database. Co-expression analysis of enhancers and miRNAs were performed using spearman correlation analysis (correlation coefficient $|R| > 0.1$, p-value<0.05).

According to the distance of the enhancer-miRNA interactions, they could be classified into two types: proximal regulation and distal regulation. Refer to previous study(Suzuki, Young, & Sharp, 2017), proximal enhancer-miRNAs was calculated by the following formula: $S\ (B/A) = (M-G) / (M+G)$. M and G represent the distance from the enhancer to the nearest miRNA gene and the nearest gene, respectively, and parameters A and B correspond to $(G+M) / 2^{1/2}$ and $(G-M) / 2^{1/2}$, respectively. $S < 0.2$ as the threshold to screen the reliable enhancer-miRNA pairs. Distal

regulation of enhancer-miRNA was identified as following the procedure: Firstly, the transcription initiation sites (TSS) of 2,248 miRNAs were downloaded from the FANTOM5 data portal(Abugessaisa, Noguchi, Carninci, & Kasukawa, 2017), 0.5 kb downstream and 1 kb upstream of the TSS of these miRNAs were defined as putative promoter region. A total of 1215 miRNAs were obtained through integrating 1881 miRNAs of TCGA and 2248 miRNAs of FANTOM5. Human chromatin interaction data was downloaded from 4DGenome (Teng, He, Wang, & Tan, 2016). If the enhancer and miRNA promoter regions overlap with the chromatin interaction region of the 4D genome, it is considered that there is a physical interaction between the enhancer and miRNA, and define the pair as distal regulation.

1.2 Characteristics of enhancer-miRNA interactions

Enhancer RNAs (eRNAs) were determined by aligning the RNA transcribed from enhancer with the annotated RNA (GENCODE.v19). The transcripts overlapping protein coding genes were removed. The GC content data was downloaded from the UCSC GC Percent track. The GC content is taken as the average of the regions of the enhancer itself. The PhastCons score was obtained from the UCSC cons100way track. The region of upstream and downstream 1 kb from center of enhancer was considered as the calculation range of conservation.

The nine obtainable histone modification CHIP-Seq data of eight cell lines was downloaded from the ENCODE including H3K4me3, H3K4me1, H3K27ac, H3K9me3, H3K27me3, H3K36me3, H3K4me2, H3K9ac and H2K20me1. The eight cell lines matched type of cancers were A549 (LUAD), HepG2 (LIHC), HELA (CESC), HCT116 (COAD), DOHH2 (DLBC), PC-3 (PRAD), PANC-1 (PAAD), DND-41 (LAML). Signal consistency was considered when it appears in at least five tissues.

1.3 Identification and analysis ubiquitously expressed enhancers

We define the enhancer-miRNA pairs that occur in more than 10 tissues as ubiquitously expressed enhancer-miRNA interaction. In order to investigate the function of the miRNA involved in enhancer-miRNA interaction, we downloaded the experimentally confirmed miRNA target genes from the miRTarbase database. Furthermore, target genes of each miRNA were subjected to GO and KEGG signaling pathway database for functional enrichment analysis using R package clusterProfiler (p-value <0.05). The eQTL data were retrieved from PancanQTL database (Gong et al., 2018), and a high correlation between SNP located on the enhancers and gene could be identified if the q value is lower than 0.05. Next, by the database starBase(Li, Liu, Zhou, Qu, & Yang, 2014), the expression level of the target miRNA inferred for the enhancer in the disease was analyzed by patient survival.

Results and Discussion

2.1 Genome-wide identification of enhancer-miRNA interactions in 31 cancers

Previous studies have shown that enhancers are involved in the synthesis and regulation of miRNAs (Xiao et al., 2017). To further explore the mechanism of enhancer-miRNA regulation in

caners, we identified a series of enhancer-regulated miRNAs in 33 cancer types. We first analyzed the co-expression between 15,080 enhancers from 8,693 samples and 1,881 miRNAs in 33 cancers. Finally, we obtained all co-expression pairs of enhancer-miRNA in 31 tissues except Uterine Corpus Endometrial Carcinoma (UCES) and Glioblastoma multiforme (GBM) because too few enhancers and miRNA samples in these two tissues. According to the distance between enhancer and miRNA, we divide enhancer-miRNA pairs into two types: proximal and distal enhancer-miRNA regulation. For proximal regulation, we used the method presented in the previous study to calculate enhancer regulated neighbor miRNAs (Suzuki et al., 2017). For distal enhancer-miRNA regulation, the enhancer-miRNA interactions were identified by Hi-C data from 4Dgenome. Finally, we obtained a total of 2,418 proximal enhancer-miRNA pairs and 1,280 distal enhancer-miRNA pairs through integration of genomic distance, co-expression and interaction analysis in 31 cancer (Figure 1A, 1B, Supplementary Table S1 and Table S2) . To investigate whether these enhancer-miRNA interactions are tissue-specific or ubiquitously expressed, we counted the frequency of occurrence of these two types of interactions appear in 31 cancer tissues. The results shown that both proximal and distal interactions exhibit significantly tissue-specific feature, with only a few number of regulations(1.2% and 2.5% in proximal and distal interactions, respectively) was ubiquitously expressed (presented in more than 10 cancer tissues) (Figure 1C).

If the regulatory relationship across large number cancers, it suggests that these regulations are critical for the tumorigenesis and development. To explore the biological functions of these miRNAs which involved in ubiquitously expressed enhancer-miRNA regulation, the GO and KEGG functional enrichment analysis were performed using experimentally verified miRNA target genes. GO analysis indicated that these target genes of miRNAs that regulated by ubiquitously expressed enhancers were significantly involved in tumor-associated biological processes such as cell cycle, cell differentiation, cell growth, metabolic regulation, metastasis, Ras protein catabolic process etc. in distal (Figure 1D) or proximal regulation (Supplementary Figure S1A); KEGG analysis revealed that these miRNA target genes were significantly involved in cancer transcriptional dysregulation signaling pathways, such as FoxO signaling pathway, p53 signaling pathway, MAPK signaling pathway, the P13K-Akt signaling pathway, etc. in distal (Figure 1E) or proximal regulation (Supplementary Figure S1B).

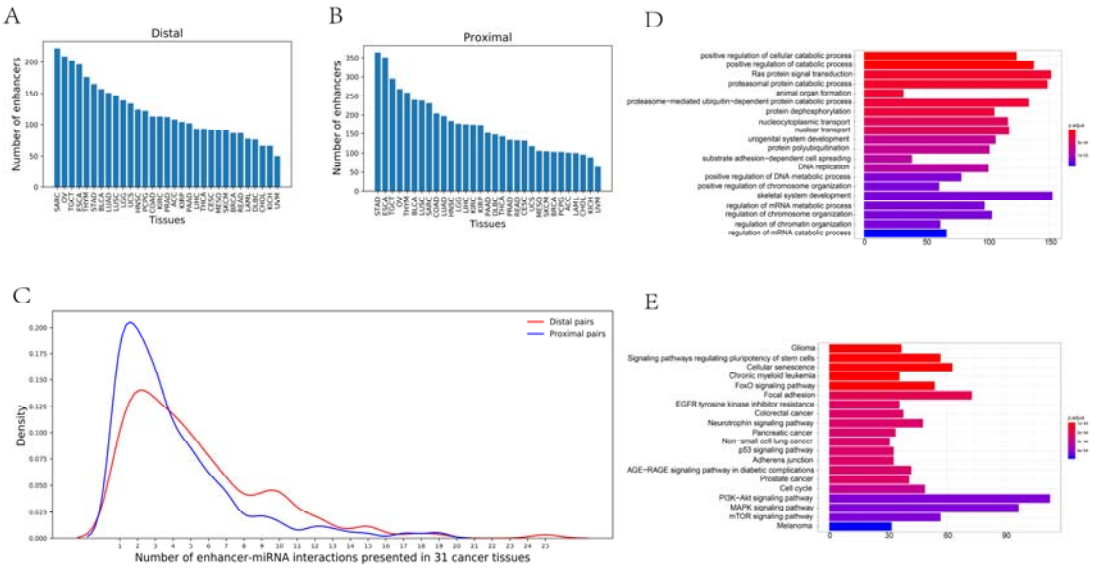

Figure 1 Go and KEGG pathway enrichment analysis of enhancer-miRNA regulation. No. of pairs of proximal regulations (A) or distal regulations (B) presented in each tissue. (C) The frequency of occurrence of enhancer-miRNA interactions appear in 31 cancer tissues. GO (D) and KEGG pathway (E) enrichment analysis of target genes of miRNAs that regulated by ubiquitously expressed enhancers.

2.2 The correlation between the miRNA expression and the number of regulated enhancers

Enhancers often regulate target genes and do not strictly follow one-to-one regulatory relationships. In order to investigate whether there is a correlation between the expression level of miRNAs and the number of miRes regulated these miRNAs, we performed principal component analysis (PCA) of the expression levels of miRNAs regulated by enhancers in 31 tissues. Here, we only analyze the distal regulation because most enhancers-miRNA interactions in proximal regulation following one-to-one regulatory rule according to genomic position restriction. The PCA results shown that the 31 cancers clearly be divided into three groups according to the number of the highly expressed miRNAs that regulated by enhancers as follows: low group(1-3), medium(4-7) and high(>7) (Figure 2A, Supplementary Table S3). For example, miRNAs in PRAD, LUAD, LAML and ESCA were regulated by more than seven enhancers present significantly higher expression compared with the number of miRNAs regulated by enhancers less than seven ($p<0.05$). Interestingly, some similar types of cancers tend to cluster into one group which shared the same enhancer-miRNA regulation pattern. For example, the most highly expression miRNAs in three types kidney cancers (ACC, KIRC, KIRP) tend to be regulated by 4-7 enhancers (Figure 2A). It is noteworthy that there is a significant positive correlation between

the expression of miRNA and the number of regulated enhancers in the Glad urothelial carcinoma (BLCA), Lung squamous cell carcinoma (LUSC), Ovarian serous cystadenocarcinoma (OV) and Testicular Germ Cell Tumors (TGCT) (Figure 2B, Supplementary Figure S2).

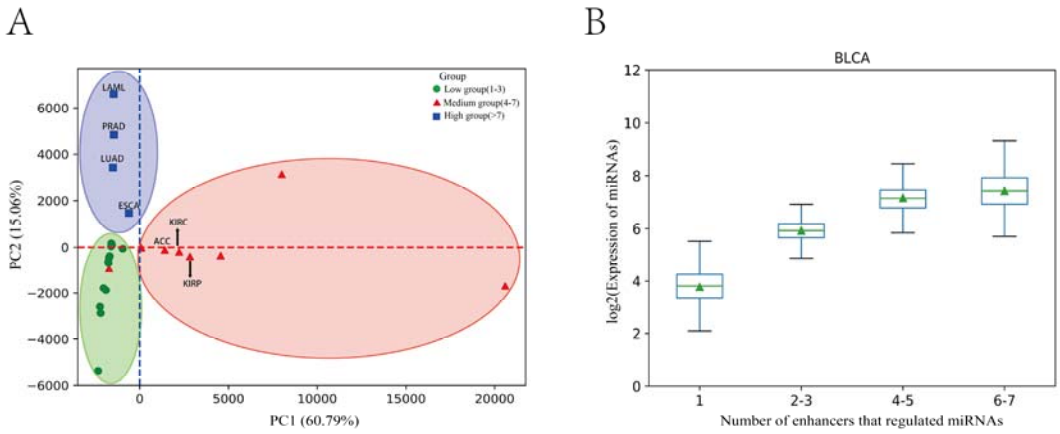

Figure 2 The correlation between the miRNA expression and the number of regulated enhancers. (A) The PCA analysis of the expression levels of miRNAs regulated by enhancers in 31 tissues. (B) The expression of miRNA that regulated by different number of enhancers in BLCA

2.3 There are significant differences in the sequence characteristics of miRes

It is important to explore the sequence characteristics of the miRes to conduct further identification of enhancer-miRNA interactions. Previously reported that eRNA can be used as a trans-acting element to participate in the regulation of target genes (Zhang et al., 2019). Consequently, we first investigated the transcript types of the distal and proximal regulatory miRes using the human GENCODE annotation. We found that 312 of the 998 (31.34%) enhancers that regulate distal miRNAs could transcribe eRNA, and the largest proportion (70.71%) of RNAs is lincRNA (Figure 3A). Similarly, we also found largest proportion of lincRNA present in enhancers that regulate proximal miRNAs (Supplementary Figure S3A). Moreover, we investigated whether there is correlation between the formation of enhancer-miRNA pairs and expression of eRNAs. The result showed that there is a high correlation between them in distal regulation (chi-square test, p-value<$1.8e^{-3}$) or proximal regulation (chi-square test, p-value<$10e^{-5}$), which suggesting that enhancer may regulate the expression of miRNAs with the participation of eRNAs (Supplementary Table S4 and Table S5).

Continuously, we analyzed the conservation of the enhancer sequence using PhastCons. In distal regulatory, the results showed that the sequence of the enhancer is more conservative than the random sequence (p<$3.2e^{-23}$), and the conserved region of enhancer was mainly located within ±250 bp around the center of enhancer (Figure 3B). Notably, the miRes showed a higher conservation compared with the enhancers that did not regulate miRNAs. The similar results also appeared in proximal regulation (Supplementary Figure S3B). The above results indicate that the functional region of enhancer mainly concentrated near the enhancer

center and that the miRes exhibits greater conservation than non-miRes. Furthermore, the GC content of the distal and proximal regulatory miRes were calculated. The results shown that miRes exhibit significantly higher GC content than the average value of random enhancer sequence in distal regulation(P-value<$2.6e^{-22}$) and the miRe exhibits a higher GC content than non-miRe in each tissue(Figure 3C). Interestingly, there was no significant difference between the GC content of miRes and non-miRes in proximal regulation (P > 0.05) (Supplementary Figure S3C). Therefore, we speculate that the GC content is an inherent property of the enhancer and may have a potential impact on chromosome looping which more necessary in distal regulation than proximal regulation.

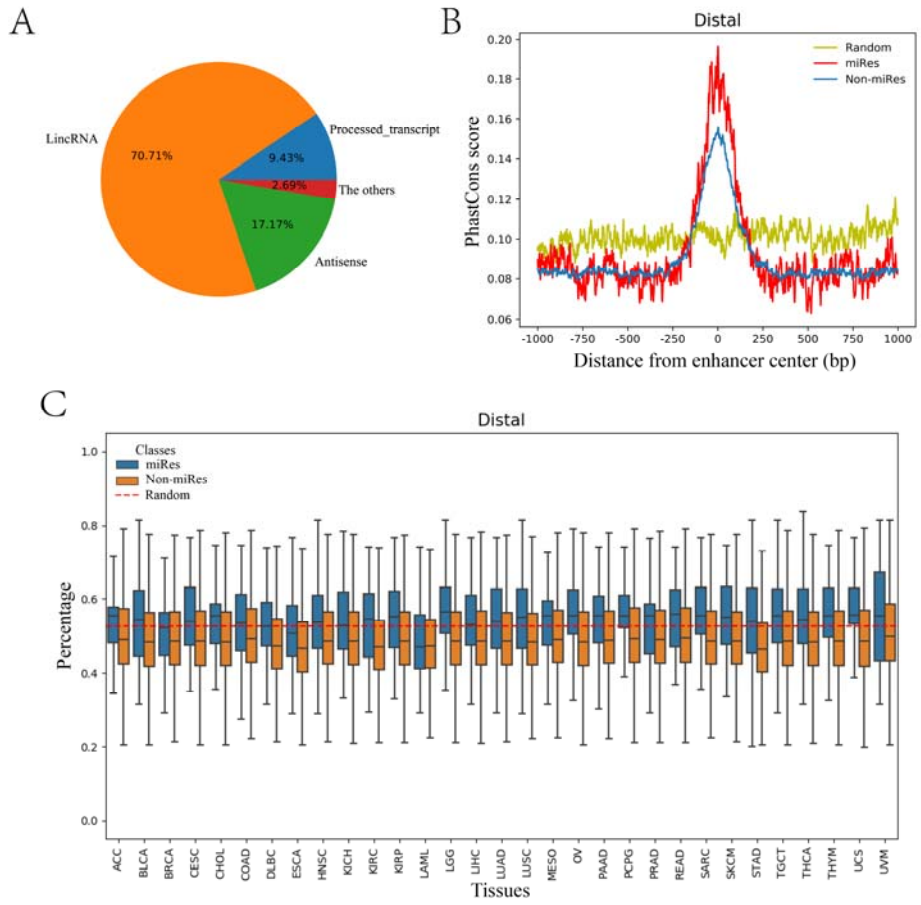

Figure 3 There are significant differences in the sequence characteristics of miRes. (A) Pie chart of all enhancer transcripts in distal regulation. (B) Conservative score of the enhancer sequence using PhastCons in distal regulation. (C) GC content of the enhancer in distal regulation.

2.4 Histone modification showing cancer and miRes specific features

Previous studies have shown that the H3K27ac, H3K4me1 and H3K4me3 signal were key histone modification features for the activity of enhancers (Rada-Iglesias et al., 2012). To determine whether there are differences activity between miRes and non-miRes, we analyzed

available H3K27ac, H3K4me1 and H3K4me3 ChIP-seq data in eight cancer tissues. Not surprisingly, as an example shown in Figure 4A-F, all of the enhancers in distal regulation pairs and proximal regulation pairs have an enrichment of H3K27ac, H3K4me3 and H3K4me1 signal in the range of 1 kb upstream and downstream from the center of enhancer, and present significantly higher signal in cancers than in normal tissues. It is notable that the signal of H3K27ac and H3K4me3 of miRes were significantly higher than that non-miRes in most tumor tissues (Supplementary Figure S4-7). Conversely, there is no significant difference in normal tissues. Interesting, H3K4me1 show the difference between the miRe and non-miRe signals only in distal regulation (Figure 4E, Supplementary Figure S8) not in proximal regulation (Figure 4F, Supplementary Figure S9). This result is consistent with previous study showed that enhancer activation of adjacent genes does not require H3K4me1 enrichment (Dorighi et al., 2017).

In addition, we ask if there are other histone modifications in addition to the above signals have a significant difference between miRe and non-miRe. Therefore, we downloaded six histone modification data from ENCODE, including H3K4me2, H3K9ac, H3K20me1, H3K9me3, H3K27me3 and H3K36me3. We found that H3K9me3 and H3K36me3 in distal and proximal pairs were significantly different between the miRes and non-miRes in at least five cancer tissues (Supplementary Figure S10-13). Among them, H3K9me3 showed lower enrichment in miRes compared with in non-miRes probably due to this histone modification were the marker of heterochromatin (Supplementary Figure S10-11) (Becker, Nicetto, & Zaret, 2016). This result is consistent with our previous suppose that the transcripts of enhancers have a positive effect on expression of miRNA that enhancer regulated. H3K36me3, a marker for transcription extension, that showing a high enrichment in the miRes in distal interaction pairs not in the proximal interaction pairs. According to previous study (Heinz et al., 2018), transcriptional elongation has an effect on the spatial structure of chromatin, and this may have more influences on distal regulation than proximal regulation (Supplementary Figure S12-13).

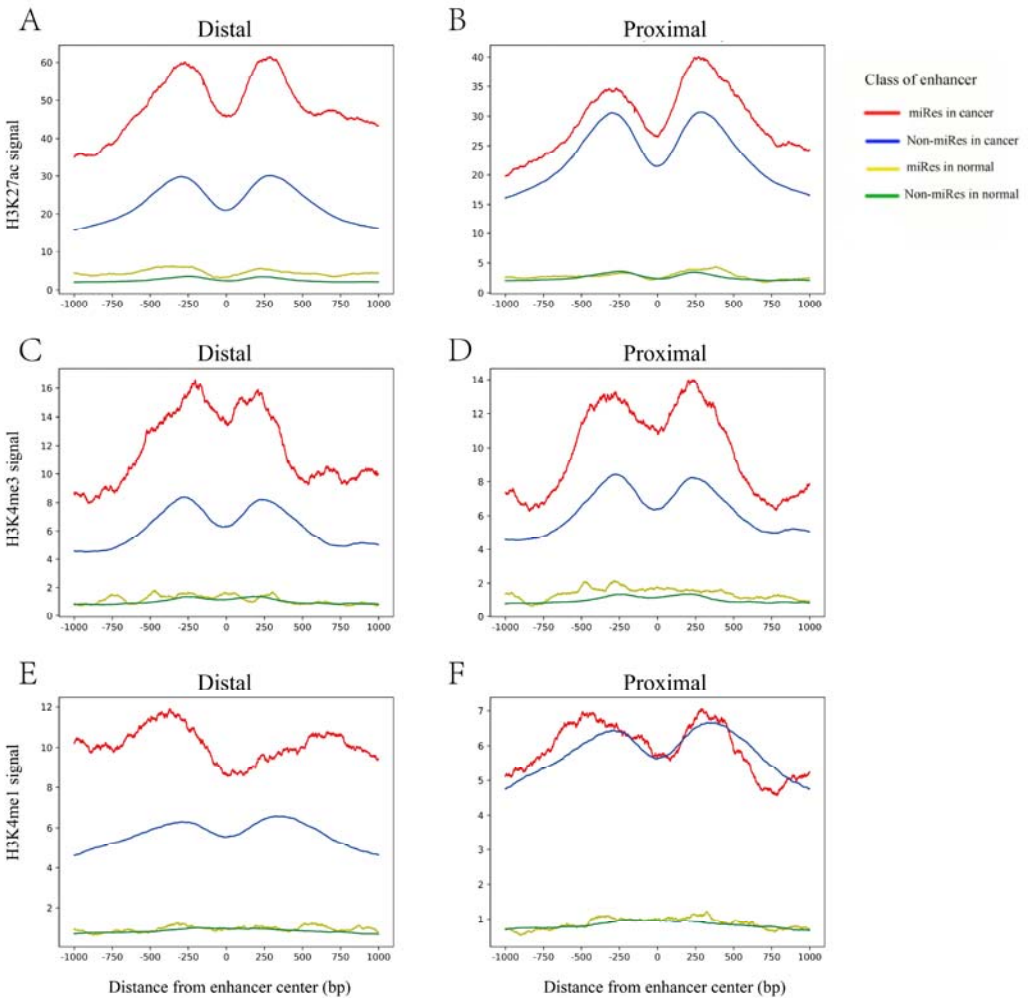

Figure 4 The signal of H3K27ac, H3K4me1 and H3k4me3 within ±1 kb surrounding the center of the enhancer center in LUAD.

2.5 A case study of miRe in thyroid carcinoma

To investigate miRes identified as described above, here we introduced a case study about an enhancer: chr1:1186391-1186507 and its target miRNA: miR-200a in thyroid carcinoma(THCA).

We identified a cis-eQTL SNP (rs6603785) on enhancer: chr1:1186391-1186507 was located close to the transcription start site (TSS) of TNFRSF18 gene (Xiong, Wang, & You, 2019), which acts as a tumor suppressor, and mainly occurs when the base A is mutated to T (Figure 5A). There was a significant difference in the expression levels of samples of different genotypes ($p<1.76e^{-4}$) (Supplementary Figure S14). In addition, miR-200a as a target of enhancer highly relevant with survival of thyroid cancer patients (Figure 5B). Previous study shown that miR-200 regulated the epithelial stromal transformation of thyroid cancer through EGF/EGFR signal (Xue et al., 2015) and it is a key factor in the epithelial phenotype and a tumor suppressor in thyroid carcinoma (Wang et al., 2018). In addition, the survival analysis shown that low expression of miR-200a patients had a lower survival time.

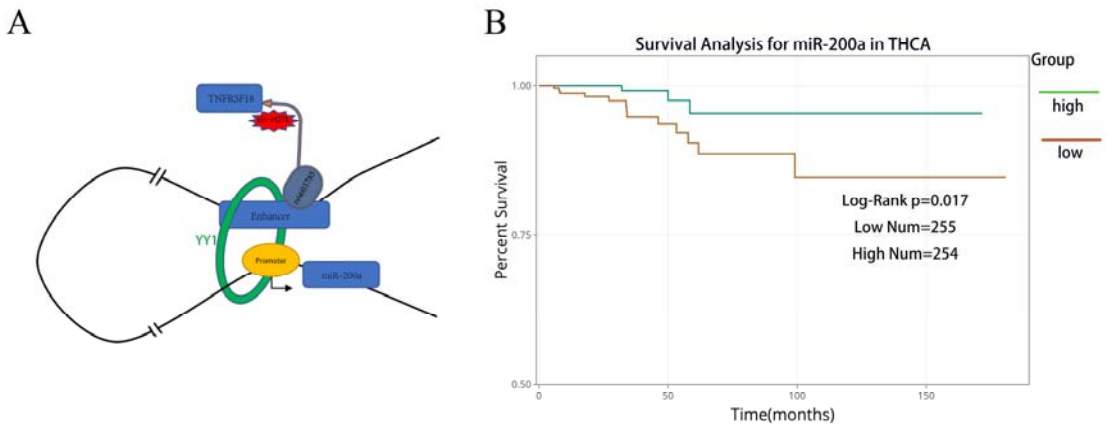

Figure 5 A case study of miRes in thyroid carcinoma. (A) An example of enhancer-miRNA regulation interaction (chr1:1186391-1186507~miR-200a). (B) The survival curve of hsa-mir-200a in thyroid cancer.

References

Abugessaisa, I., Noguchi, S., Carninci, P., & Kasukawa, T. (2017). The FANTOM5 Computation Ecosystem: Genomic Information Hub for Promoters and Active Enhancers. *Methods Mol Biol, 1611*, 199-217.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature, 507*(7493), 455-461.

Becker, J. S., Nicetto, D., & Zaret, K. S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet, 32*(1), 29-41.

Chen, H., Li, C., Peng, X., Zhou, Z., Weinstein, J. N., & Liang, H. (2018). A Pan-Cancer Analysis of Enhancer Expression in Nearly 9000 Patient Samples. *Cell, 173*(2), 386-399.e312.

Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., et al. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Mol Cell, 66*(4), 568-576.e564.

Gong, J., Mei, S., Liu, C., Xiang, Y., Ye, Y., Zhang, Z., et al. (2018). PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res, 46*(D1), D971-d976.

Heinz, S., Texari, L., Hayes, M. G. B., Urbanowski, M., Chang, M. W., Givarkes, N., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. *Cell, 174*(6), 1522-1536.e1522.

Li, J. H., Liu, S., Zhou, H., Qu, L. H., & Yang, J. H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res, 42*(Database issue), D92-97.

Lu, T. X., & Rothenberg, M. E. (2018). MicroRNA. *J Allergy Clin Immunol, 141*(4), 1202-1207.

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S. A., Swigut, T., & Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell, 11*(5), 633-648.

Sandiford, O. A., Moore, C. A., Du, J., Boulad, M., Gergues, M., Eltouky, H., et al. (2018). Human Aging and Cancer: Role of miRNA in Tumor Microenvironment. *Adv Exp Med Biol, 1056*, 137-152.

Suzuki, H. I., Young, R. A., & Sharp, P. A. (2017). Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell, 168*(6), 1000-1014.e1015.

Teng, L., He, B., Wang, J., & Tan, K. (2016). 4DGenome: a comprehensive database of

chromatin interactions. *Bioinformatics, 32*(17), 2727.

Thandapani, P. (2019). Super-enhancers in cancer. *Pharmacol Ther, 199*, 129-138.

Wang, X., Huang, S., Li, X., Jiang, D., Yu, H., Wu, Q., et al. (2018). A potential biomarker hsa-miR-200a-5p distinguishing between benign thyroid tumors with papillary hyperplasia and papillary thyroid carcinoma. *PLoS One, 13*(7), e0200290.

Wood, C. D., Carvell, T., Gunnell, A., Ojeniyi, O. O., Osborne, C., & West, M. J. (2019). Correction for Wood et al., "Enhancer Control of MicroRNA miR-155 Expression in Epstein-Barr Virus-Infected B Cells". *J Virol, 93*(3).

Xiao, M., Li, J., Li, W., Wang, Y., Wu, F., Xi, Y., et al. (2017). MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol, 14*(10), 1326-1334.

Xiong, D., Wang, Y., & You, M. (2019). Tumor intrinsic immunity related proteins may be novel tumor suppressors in some types of cancer. *Sci Rep, 9*(1), 10918.

Xue, L., Su, D., Li, D., Gao, W., Yuan, R., & Pang, W. (2015). MiR-200 Regulates Epithelial-Mesenchymal Transition in Anaplastic Thyroid Cancer via EGF/EGFR Signaling. *Cell Biochem Biophys, 72*(1), 185-190.

Yun, M. R., Lim, S. M., Kim, S. K., Choi, H. M., Pyo, K. H., Kim, S. K., et al. (2018). Enhancer Remodeling and MicroRNA Alterations Are Associated with Acquired Resistance to ALK Inhibitors. *Cancer Res, 78*(12), 3350-3362.

Zhang, Z., Lee, J. H., Ruan, H., Ye, Y., Krakowiak, J., Hu, Q., et al. (2019). Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun, 10*(1), 4562.