

RESEARCH ARTICLE

Open Access

Genomic features and computational identification of human microRNAs under long-range developmental regulation

Ying Sheng^{1,2*} and Christopher Previti^{1*}

Abstract

Background: Recent functional studies have demonstrated that many microRNAs (miRNAs) are expressed by RNA polymerase II in a specific spatiotemporal manner during the development of organisms and play a key role in cell-lineage decisions and morphogenesis. They are therefore functionally related to a number of key protein coding developmental genes, that form genomic regulatory blocks (GRBs) with arrays of highly conserved non-coding elements (HCNEs) functioning as long-range enhancers that collaboratively regulate the expression of their target genes. Given this functional similarity as well as recent zebrafish transgenesis assays showing that the miR-9 family is indeed regulated by HCNEs with enhancer activity, we hypothesized that this type of miRNA regulation is prevalent. In this paper, we therefore systematically investigate the regulatory landscape around conserved self-transcribed miRNAs (ST miRNAs), with their own known or computationally inferred promoters, by analyzing the hallmarks of GRB target genes. These include not only the density of HCNEs in their vicinity but also the presence of large CpG islands (CGIs) and distinct patterns of histone modification marks associated with developmental genes.

Results: Our results show that a subset of the conserved ST miRNAs we studied shares properties similar to those of protein-coding GRB target genes: they are located in regions of significantly higher HCNE/enhancer binding density and are more likely to be associated with CGIs. Furthermore, their putative promoters have both activating as well as silencing histone modification marks during development and differentiation. Based on these results we used both an elevated HCNE density in the genomic vicinity as well as the presence of a bivalent promoter to identify 29 putative GRB target miRNAs/miRNA clusters, over two-thirds of which are known to play a role during development and differentiation. Furthermore these predictions include miRNAs of the miR-9 family, which are the only experimentally verified GRB target miRNAs.

Conclusions: A subset of the conserved miRNA loci we investigated exhibits typical characteristics of GRB target genes, which may partially explain their complex expression profiles during development.

Background

MicroRNAs (miRNAs) are small RNAs (~ 22 nt) found in plants, animals, viruses and at least one unicellular organism (*Chlamydomonas reinhardtii*) [1]. They function by binding to target sites in 3' UTRs of messenger RNAs (mRNAs) to repress translation or mediate mRNA degradation, although alternative modes of action have been reported recently, such as direct

transcriptional silencing of *POLR3D* by miR-320 [2]. In animals, the majority of mature miRNAs are synthesized in two processing steps: first, the primary miRNA transcript (pri-miRNA) is cut by the nuclear RNase III enzyme *Drosha* and its cofactor *DGCR8* in the nucleus, generating precursor miRNAs (pre-miRNAs). These are subsequently exported to the cytoplasm via the nuclear transport receptor *exportin-5* and the co-factor *RanGTP*, where they are cleaved by the RNase III enzyme *Dicer* into a double stranded RNA of ~22nt. The strand with the less stable 5' hydrogen bonding is usually selected as the mature miRNA, although both strands can be functional. It is assumed that the pri-miRNAs of most

* Correspondence: Ying.Sheng@medisin.uio.no; Christopher.Previti@uni.no
¹Computational Biology Unit, Bergen Center for Computational Science, and Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, 5008, Norway
Full list of author information is available at the end of the article

intergenic miRNAs are transcribed independently by RNA polymerase II (Pol II) [3], while most intragenic miRNAs are co-transcribed with their host genes. Given the similarity with class II protein-coding genes, their expression may be controlled through a variety of shared regulation pathways. However, few pri-miRNAs have been characterized experimentally to date [3-8], making the localization of their promoters challenging and one of the prime reasons we are only beginning to understand the mechanism by which their expression is regulated. Recently, several studies attempted to predict the primary transcripts of pri-miRNAs by exploring transcription factor binding sites (TFBSs), expressed sequence tags (ESTs), transcription start site (TSS) predictions, and further, complementary genomic data [9-15]. Based on these studies, the length of pri-miRNAs was estimated to range from several hundred to several thousand nucleotides. These studies are useful references to estimate the scale of pri-miRNAs as well as the location of their promoter, in order to investigate miRNA transcriptional regulation.

miRNAs can be expressed in a tissue- and stage-specific manner during development and that they can play key roles in lineage decisions of progenitor cells and organogenesis (reviewed in [16]). Enhancer-mediated long-range regulation is an important mechanism for controlling gene expression during development and has recently been shown to affect miRNAs as well [17]. Many of these enhancers are highly conserved non-coding elements (HCNEs) that collaboratively regulate the specific expression of their respective target genes [18-24]. It has previously been demonstrated that both HCNEs and their target genes are preserved within syntenic blocks in vertebrates and insects during evolution, which has served as the foundation for the concept of the "genomic regulatory block" (GRB). GRBs are functional regulatory units that consist of HCNEs, genes regulated by HCNEs ("target genes") as well as "unrelated" genes ("bystander genes"). Both HCNEs and target genes have coevolved in order to maintain their functional association, while bystander genes can be lost through the time.

Further investigation of the general transcriptional initiation properties of genes in GRBs has shown that the promoters of GRB target genes share common features that can be used to distinguish them from the promoters of bystander genes, which are genes that may be close to, or even harbor, HCNEs but are not under their regulation. Target genes are generally associated with long CpG islands (CGIs) that are not limited to the 5' end of the genes, but also occur in introns or internal exons of the gene [25] and coincide with genomic regions bound by repressor Polycomb group proteins [26]. In addition, they have a higher number and wider

spacing of alternative TSSs, and a distinct composition of TFBSs in their core/proximal promoters [25]. Furthermore, many target gene promoters belong to the class of "bivalent promoters" that display a distinct pattern of both activating and repressing histone modifications in embryonic stem (ES) cell lines [25], which may allow them to be turned on or off quickly during organogenesis [27,28]. The analysis of the mir-9 miRNA family members (*dre-mir-9-5* and *dre-mir-9-1*) in zebrafish has shown that they are regulated by the same type of enhancers as protein-coding GRB target genes [17]. These miRNA target genes are both embedded in areas of conserved syntenicity throughout vertebrates and co-localize with a number of HCNEs that function as long-range enhancers controlling their expression. As a result, they show highly stage- and tissue-specific expression in dorsal telencephalon at 24 h post fertilization, while the expression pattern of the neighboring bystander genes is far less specific. The inspection of other miRNAs in the Ancora genome browser [29] clearly reveals further examples of miRNAs that fall within regions of high HCNE density, such as *hsa-mir-124-2*, indicating that this type of regulation may be prevalent.

Based on their functional similarity and the common transcriptional mechanisms they employ, we hypothesize that miRNAs with complex spatiotemporal expression patterns may be regulated in the same manner as protein-coding GRB target genes and that they share the following genomic properties:

- a genomic neighborhood with a high HCNE density
- a bivalent promoter [28] during development and differentiation
- large CGIs spanning both the promoter and the gene body

In our work we show that both a subset of conserved self-transcribed miRNAs (ST miRNAs) as well as a set of experimentally validated GRB target miRNAs [17] demonstrate these properties. But our ability to use these features for predicting novel GRB target miRNAs was limited, since more than half of the ST miRNAs in our dataset do not have primary transcript predictions and we were therefore not able to accurately compute CpG island-related features for them. As a result, we used an elevated HCNE density in the genomic vicinity as well as the presence of a bivalent promoter to then identify 29 putative GRB target miRNAs/miRNA clusters, over two-thirds of which are known to play a role during development and differentiation. Furthermore these predictions include miRNAs of the miR-9 family, which are the only experimentally verified GRB target miRNA genes.

Results and Discussion

Defining self-transcribed miRNAs and the extent of their pri-miRNAs

Based on the assumption that some miRNAs have co-evolved with their cis-acting regulatory elements, in this study, we focused on conserved human self-transcribed miRNAs (ST miRNAs), which are assumed to be transcribed from their own promoters and have an ortholog in at least one other vertebrate species (Additional file 1, Table S1). These requirements are consistent with the work describing the original definition of a GRB [17] as well as the subsequent analysis of the features of GRB target genes [25]. In general, ST miRNAs include all conserved human intergenic miRNAs except those transcribed by RNA polymerase III (Pol III) [30], as well as known cases of intronic miRNAs whose fate after whole-genome duplication in fish indicates that they have their own promoters, independent of their host genes (see Methods).

Since proximal miRNAs can appear in genomic clusters that share the same promoter and are transcribed as a single transcript, we clustered ST miRNAs according to their genomic vicinity (see Methods) and analyzed only one promoter per cluster,

The annotation of the pri-miRNAs we used in this study was the result of a combinatorial approach by Saini et al. [14] that employed predicted and experimental evidence to identify both transcription start and end sites. One of the most important reasons for choosing this dataset was the fact that they considered a larger range of possible miRNA sizes, thereby avoiding biases introduced by excluding very long transcripts. We estimated the general scale of pri-miRNAs based on these data, where the TSS is within 50 kb upstream of pre-miRNAs and the transcription end is within 20 kb downstream of the pre-miRNAs. This approach was sensitive enough to include around 84% of the pri-miRNAs in the initial dataset (see Figure 1 and Methods).

Enhancer density

HCNEs

HCNEs represent putative enhancers of gene expression, whose frequency rises strongly around the conserved developmental genes they usually target [17,22]. We therefore first investigated whether this applied to conserved ST miRNAs as well, by comparing the HCNE density in a 300 kb window centered on the ST miRNA with random coding and non-coding genomic regions with similar gene density (see Methods). In a few cases (~8%), the location of the ST miRNA search window overlapped with a previously defined GRB making it difficult to identify the actual target gene, which could either be the miRNA, the annotated GRB target gene, or both. For this reason, we compared the HCNE densities

both including and excluding cases in which the ST miRNA search window overlapped with a previously defined GRB. As shown in Table 1, the HCNE density was significantly higher around ST miRNAs than the random coding and non-coding regions in both comparisons, a trend that is independent of the lineages compared (p -value ≤ 0.05 , two-sided bootstrapped version of the Kolmogorov-Smirnov test: see Table 1, Figure 2 as well as Additional files 2 and 3). Thus, in support of our original hypothesis, we can conclude that HCNEs are highly over-represented in the genomic vicinity of conserved human ST miRNAs. The HCNE density was also compared between the conserved ST miRNAs and the GRB target genes annotated in Akalin et al. [25]. We found that the HCNE density was significantly lower for the ST miRNAs (p -value ≤ 0.05 , two-sided bootstrapped version of the Kolmogorov-Smirnov test: see Table 1, Figure 2 as well as Additional files 2 and 3). This is due to the fact that the GRB target genes were annotated based on their high HCNE density [25], which was not a prerequisite for constructing our dataset of ST miRNAs.

Experimentally verified enhancers

In addition to HCNEs, we utilized experimental data describing the location of the transcriptional co-activator p300 to analyze the enhancer density around ST miRNAs. This transcription factor is a ubiquitous component of enhancer-associated protein assemblies. It co-localizes with active enhancers and plays a critical role during embryonic development [31-36]. Visel et al. [36] mapped the genome-wide binding of p300 in forebrain, midbrain and limb tissue of developing mouse embryos using ChIP-seq technology. This generated ~5000 p300 binding regions which are associated with active enhancers. Since conservation criteria did not play a role in determining these p300 binding regions, they represent a dataset of experimentally verified enhancers that is completely independent of the HCNEs, allowing us to test our hypothesis in an unbiased manner. We then compared the p300 binding density around ST miRNAs that are conserved between mouse and human (mouse:human) with that of mouse coding and non-coding regions using the same method as the previous HCNE density analysis and found that the density of p300 binding sites around mouse:human ST miRNAs was indeed significantly higher than in the control sets (p -value $< 1e-20$ for both comparisons, two-sided bootstrapped version of the Kolmogorov-Smirnov test) (Figure 3).

These results are independent of HCNE density, since the same trend was still present when we performed the analysis excluding p300 binding sites that overlap with mouse orthologs of our HCNEs (p -value $< 1e-20$ for both comparisons, two-sided bootstrapped version of

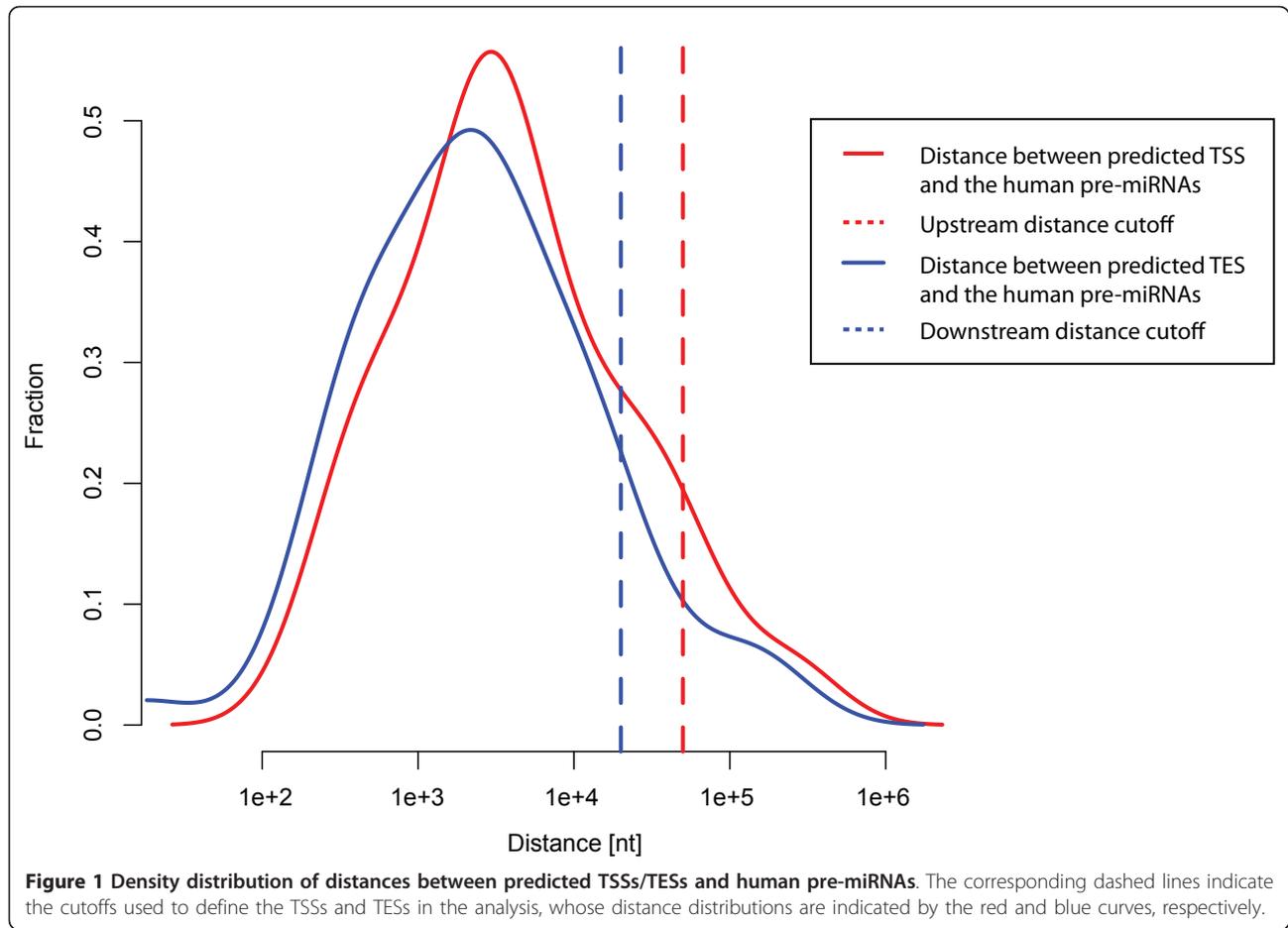


Table 1 HCNE density comparison

Lineage comparison	Comparison between human conserved ST miRNAs and human random coding regions		Comparison between human conserved ST miRNAs and human random non-coding regions		Comparison between human conserved ST miRNAs and human GRB target genes	
	<i>p</i> -value (all regions)	<i>p</i> -value (excluding regions overlapping GRBs)	<i>p</i> -value (all regions)	<i>p</i> -value (excluding regions overlapping GRBs)	<i>p</i> -value (all regions)	<i>p</i> -value (excluding conserved human ST miRNAs overlapping GRBs)
human: mouse	0	0	0	0	0	0
human: dog	0	0	0	0	0	0
human: opossum	0	0	1.0e-04	0	0	0
human: platypus	0	0	0	2.0e-04	0	0
human: chicken	0	0	0	0	0	0
human: frog	0	0	0	0	0	0
human: zebrafish	4.0e-03	8.0e-03	1.8e-03	4.0e-04	0	0

The comparisons represented by each column were performed after selecting HCNEs and ST miRNAs conserved between the lineages shown in the left column (see Methods). The *p*-values were computed using the two-sided bootstrapped version of the Kolmogorov-Smirnov test. All *p*-values lower than 1.0e-20 were set to 0.

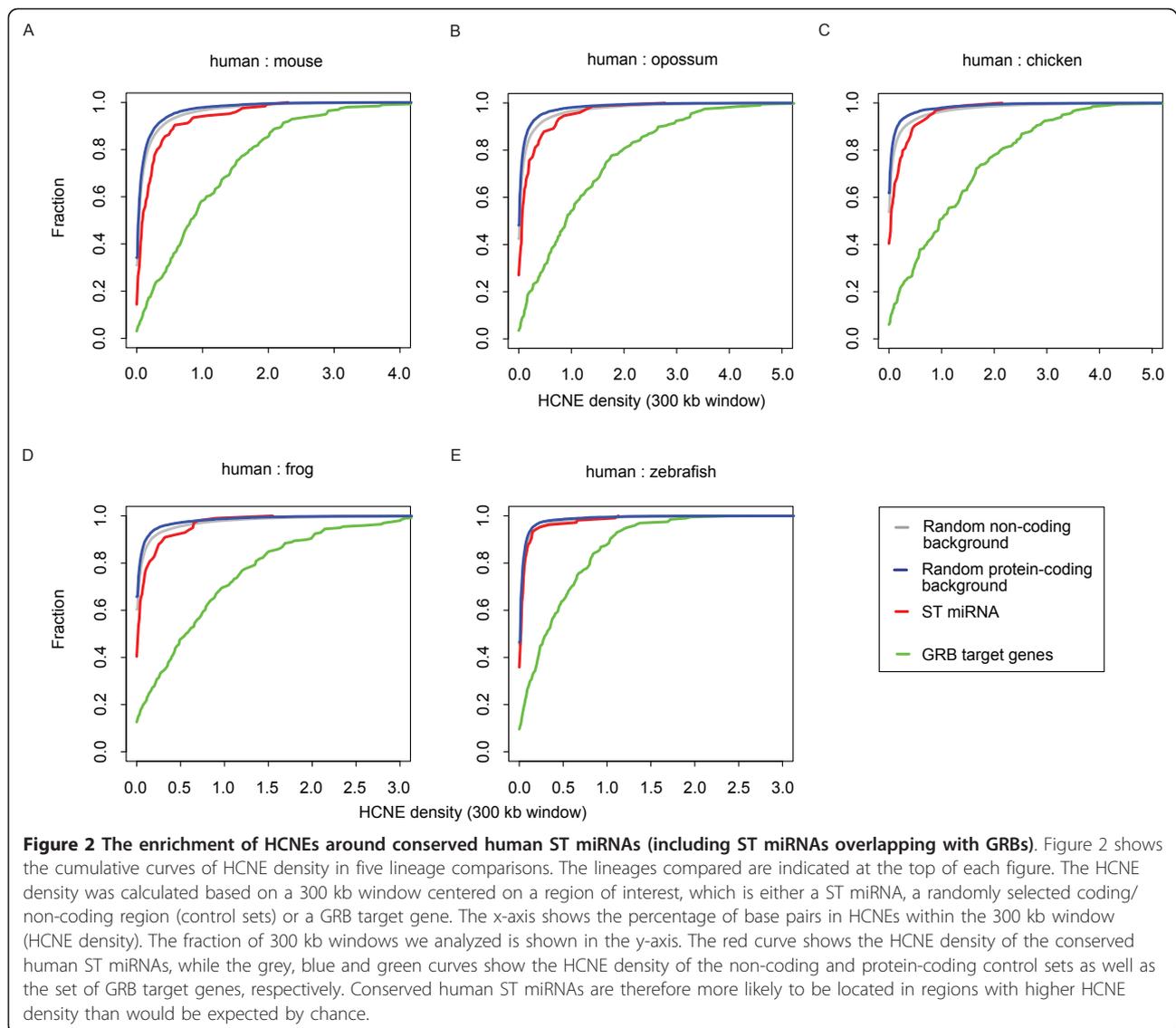


Figure 2 The enrichment of HCNEs around conserved human ST miRNAs (including ST miRNAs overlapping with GRBs). Figure 2 shows the cumulative curves of HCNE density in five lineage comparisons. The lineages compared are indicated at the top of each figure. The HCNE density was calculated based on a 300 kb window centered on a region of interest, which is either a ST miRNA, a randomly selected coding/non-coding region (control sets) or a GRB target gene. The x-axis shows the percentage of base pairs in HCNEs within the 300 kb window (HCNE density). The fraction of 300 kb windows we analyzed is shown in the y-axis. The red curve shows the HCNE density of the conserved human ST miRNAs, while the grey, blue and green curves show the HCNE density of the non-coding and protein-coding control sets as well as the set of GRB target genes, respectively. Conserved human ST miRNAs are therefore more likely to be located in regions with higher HCNE density than would be expected by chance.

the Kolmogorov-Smirnov test) (Figure 3, Additional file 4 and Table S2).

Comparison between enhancer enriched and enhancer poor ST miRNAs

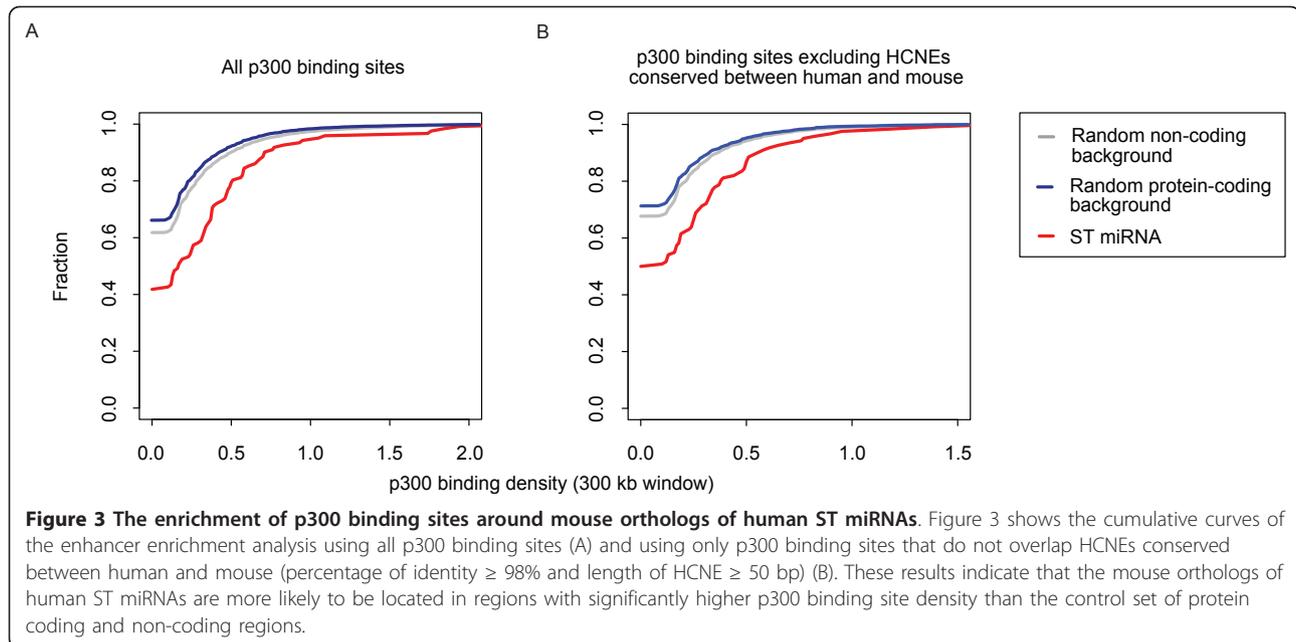
As the density of HCNEs/p300 binding around conserved ST miRNAs is significantly higher than in random non-coding and protein coding regions and there are conserved ST miRNAs that are known developmental regulators similar to GRB target genes, we wanted to see whether there was an association between the individual features of GRB target genes and HCNE/p300 densities. We again limited the analysis to ST miRNAs conserved between human and mouse in order to be able to incorporate the p300 binding data.

We then compared the features of GRB target genes between the HCNE/p300 enriched and HCNE/p300

poor miRNAs excluding intragenic ST miRNAs due to the uncertainty in associating genomic features with the ST miRNA or the host gene.

Identification and analysis of bivalent ST miRNA promoters

The genome-wide mapping of chromatin states by detection of histone marks revealed regions carrying both the H3K4me3 and H3K27me3 histone modifications. Although these so-called bivalent domains [27] are rare within the genome, they are over-represented at the promoters of key developmental genes (bivalent promoters) in embryonic stem cells [28] and were suggested to mediate activation and repression of expression of the genes during lineage commitment by maintaining genes in a poised status [27]. Furthermore, bivalent promoters represent useful marks for the annotation of both protein-coding and miRNA GRB target genes, since around 70% of them had a bivalent promoter in mouse ES cells,



compared to only 13% of the bystander genes [25], and the expression of several lineage-specific miRNAs is correlated with the presence of this mark at their promoters [28].

Both human [37] and mouse bivalent domains [28] were subsequently mapped to the putative promoter regions of HCNE enriched/poor ST miRNAs and p300 enriched/poor ST miRNAs, respectively (see Methods). We found that 24% (15/63) of the HCNE enriched ST miRNAs are associated with bivalent domains within 50 kb upstream in human ES cells, compared to only 8% (5/62) of HCNE poor ST miRNAs (p -value = 0.03, one side Fisher's exact test). However, bivalent domains are less likely to be associated with p300 enriched ST miRNAs compared to those of p300 poor ST miRNAs (10/62 versus 19/61). This difference could be explained by the fact that the data on histone modifications and p300 binding are from different tissues/developmental stages, since the histone modification data are from V6.5 ES cells, hybrid ES cells, Neural Progenitor cells (NPCs) and primary Embryonic Fibroblasts (MEFs) obtained at embryonic day (E) 13.5 [28], while the p300 data is from embryonic forebrain, midbrain and limb tissue of mouse embryos at E 11.5 [36]. As enhancers and histone modifications are both tissue- and stage-specific, it is possible that the p300 data and bivalent domains we used in the analysis regulate transcription in distinct developmental contexts. Furthermore, the bivalent promoters are associated with promoters "poised" for transcription [27] while the p300 marks of active enhancers are expected to co-occur with actively transcribed target genes [38], so they are unlikely to co-occur in the same tissue/stage for a given gene. Fortunately, the detection of HCNEs

does not rely on tissue or developmental stages, but instead is able to detect putative enhancers in all tissues from all stages. Therefore, the analysis of HCNE enriched/poor ST miRNAs is stage/tissue-independent and thus less likely to be biased. For this reason, in the following analysis we only compared features between HCNE enriched and HCNE poor ST miRNAs.

Analysis of CpG islands associated with ST miRNAs

GRB target genes are often associated with higher ratios between CGI length and transcript length (CpG-to-gene ratio), and the overlapping CGIs map not only to the promoter, as in most other genes, but also introns, internal exons, and in some cases, even cover the entire target gene [25]. It was shown recently that some of these CGIs coincide within genomic regions bound by repressive Polycomb Group proteins (PcG-proteins) [26]. A recent study also found that 21 human miRNAs co-localized with multiple CGIs within their 10 kb flanking regions, and that 25 pre-miRNAs were completely embedded in CGIs [12]. In addition, Juan et al. [39] showed that the expression of miR-199/214 might be regulated by PcG-proteins during skeletal muscle cell differentiation. In another example shown by Wang et al. miR-29 is repressed by NF-kappaB acting through YY1 and the PcG-proteins [40]. Based on these findings, we compared the CpG-to-gene ratio associated with HCNE enriched and HCNE poor ST miRNAs as well as the gene sets used in the previous GRB target gene study by Akalin et al [25] (see Methods).

The results show that the CpG-to-gene ratio of the HCNE enriched ST miRNAs is significantly higher than those of bystander, transcription factor and CGI-

associated genes, while there is no significant difference in the comparison with the GRB target genes (Table 2). Furthermore, we find no significant differences in the CpG-to-gene ratios between HCNE poor ST miRNAs and bystander genes or other transcription factors. In addition, the CpG-to-gene ratios of HCNE poor ST miRNAs is significantly lower when compared to GRB target genes/other CGI genes. This indicates that the distribution of CpG-to-gene ratios of HCNE enriched ST miRNAs is more similar to GRB target genes than that of HCNE poor ST miRNAs. However, we did not find significant differences in the CpG-to-gene ratios between HCNE enriched ST miRNAs and HCNE poor ST miRNAs, which could be explained by the small sample size. Nevertheless, there is still a common trend showing that HCNE enriched ST miRNAs have higher CpG-to-gene ratios than HCNE poor ST miRNAs, since their median CpG-to-gene ratio is fivefold higher (Table 2).

GRB target gene features can identify ST miRNAs subject to long-range regulation

The miR-9 family of miRNAs is a known, experimentally verified GRB target [17] and therefore a prime example for illustrating how the genomic features we analyzed could serve to annotate miRNAs under long-range regulation. This family is specifically expressed in brain and affects the fate of ES cell-derived neural precursor cells differentiating along the glial or neuronal pathways [41]. The expression of the miR-9 family of genes is dynamically regulated [42] during differentiation and development, and the human miR-9 family has three members: *hsa-mir-9-1*, *-2*, and *-3*. In the human genome, *hsa-mir-9-2* and *hsa-mir-9-3* are intergenic miRNAs, and *hsa-mir-9-1* is found in the second intron of *C1orf61*. Since the zebrafish ortholog of *hsa-mir-9-1* is located in an intergenic region, it qualifies as a ST miRNA based on our definition. Moreover, *hsa-mir-9-1* and *hsa-mir-9-2* are paralogs that were most likely separated by the whole-genome duplication at the root of jawed vertebrates [43-45], which suggests that their common ancestor was intronic, but that the host gene is not required for their transcription [17]. Therefore, all miR-9 family members can be classified as ST miRNAs

and are likely to have their own promoters. A detailed examination of their genomic environments showed that all of them share the features of GRB target genes (Figure 4), since they map to genomic regions with high HCNE densities, the putative promoters of their mouse orthologs map to bivalent domains and they are all associated with several proximal CGIs.

We further examined the annotation of each gene within the investigated region around the miRNAs. Since there is no other putative GRB target gene within the region of *hsa-mir-9-3*, we conclude that *hsa-mir-9-3* is most likely the only target of long-range enhancers in that region. In the neighborhood of *hsa-mir-9-1* and *hsa-mir-9-2*, we find myocyte enhancer factors *MEF2D* and *MEF2C*, respectively. Both of these genes are transcription factors that play a role in myogenesis and are therefore plausible GRB target gene candidates. However, it was shown that two regions with enhancer activity located ~10 kb downstream of *dre-mir-9-1*, the zebrafish ortholog of *hsa-mir-9-1*, and ~100 kb downstream of *dre-mir-9-5*, the zebrafish ortholog of *hsa-mir-9-2*, gave the reporter gene an expression pattern similar to that of zebrafish miR-9, but not the zebrafish myocyte enhancer factors [17]. Thus, it may very well be that both miRNAs and the myocyte enhancer factors are regulated by distinct elements in the regions, and that there is a boundary between the two, similar to that of neighboring GRBs targeting *PAX6* and *WT1* genes [46]. In summary, the genomic features of regions around members of the miR-9 family display characteristics equivalent to those of protein-coding GRB target genes, lending further support to the use of these features for predicting novel miRNA targets of long-range regulation.

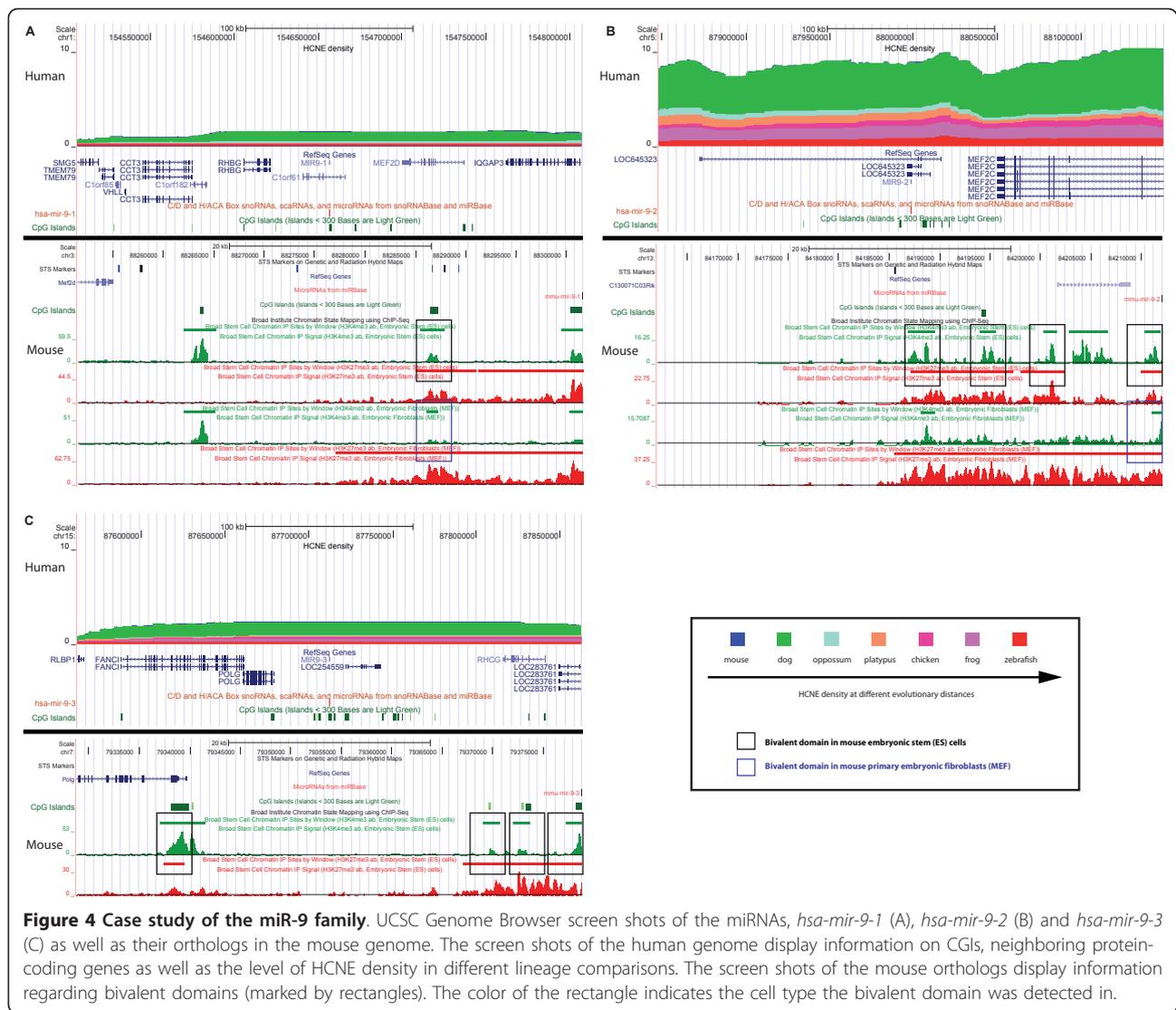
Genome-wide identification of putative GRB target miRNAs

Taken together, the results of our analyses show that HCNE enriched ST miRNA genes are more likely to be associated with the features of GRB target genes than HCNE poor ST miRNA genes, which leads us to believe that we can utilize them for the prediction of GRB target miRNAs.

Table 2 Comparison of CpG-to-gene ratios between different gene sets

	HCNE enriched miRNAs	HCNE poor miRNAs	Known GRB target genes	Known bystander genes	Other transcription factors	Other CpG island genes
Median CpG-to-gene ratio	0.1703	0.0238	0.2032	0.0100	0.0339	0.0280
<i>p</i> -values:						
HCNE enriched miRNAs	-	0.2931	0.1185	2.7450e-4	0.0124	2.7800e-3
HCNE poor miRNAs	0.2931	-	0.0158	0.1846	0.0966	1.4200e-3

Significant differences in the median CpG-to-gene ratio between the gene sets were determined using the two-sided bootstrapped version of the Kolmogorov-Smirnov test. Significant *p*-values (<0.05) are shown in bold. The first row indicates the median CpG-to-gene ratio for each gene set and the second and third rows contain the *p*-values of the comparisons between the CpG-to-gene ratios of the HCNE enriched or HCNE poor ST miRNAs and other control gene sets.



In order to identify GRB target miRNA genes on a genome-wide scale, we selected all HCNE enriched ST miRNA genes across different lineage comparisons, which were associated with bivalent domains in at least one developmental cell type (Table 3). Our predictions included 29 ST miRNA genes/miRNA gene clusters, 19 of which have known functions in development (Additional file 5 and Table S3) as well as the miR-9 family, which are the validated GRB target miRNAs. We did not use CGI-related features in the prediction process, since most of the ST miRNAs in our dataset do not have primary transcript predictions. We were therefore unable to accurately assign CGIs to them. Nevertheless, the majority of them are associated with at least one CGI within the estimated pri-miRNAs (regions within 50 kb up- and 20 kb downstream of the pre-miRNAs).

Conclusions

While it has previously been demonstrated that a subset of miRNA genes are under the same type of extreme long-range transcriptional regulation typical for developmental transcription factors [17], we were now able to identify further putative GRB target miRNAs on a genome-wide scale based on the same characteristics defining protein-coding GRB target genes. These features are their localizations in regions with a high HCNE density and the fact that many of them have “bivalent” promoters before or during differentiation.

Furthermore, the majority of our predictions is known to play a role during development and is associated with one or more CGIs that are distributed along the miRNA primary transcript, indicating that these ST miRNAs may be subject to the same type of Polycomb-mediated repression seen in protein coding developmental genes.

Table 3 Annotation of ST miRNA candidates under putative long-range developmental regulation

Name	Intergenic/Intragenic	Cell type of bivalent promoter ¹	Potential alternative GRB target gene ²	CGIs ³
<i>hsa-mir-196b</i>	Intergenic	mES, MEF	HOXA cluster	0
<i>hsa-mir-132~212</i>	Intergenic	hES, mES	-	1
<i>hsa-mir-196a-2</i>	Intergenic	MEF	HOXC cluster	2
<i>hsa-mir-9-1</i>	Intron of <i>C1orf61</i>	mES, MEF	-	-
<i>hsa-mir-9-3</i>	Intergenic	hES, mES	-	6
<i>hsa-mir-9-2</i>	Intergenic	hES	<i>MEF2C</i>	8
<i>hsa-mir-10a</i>	Intergenic	hES, mES, MEF	HOXB cluster	1
<i>hsa-mir-196a-1</i>	Intergenic	hES	HOXB cluster	6
<i>hsa-mir-137</i>	Intergenic	hES	-	2
<i>hsa-mir-375</i>	Intergenic	hES, mES	<i>FEV, INN</i>	2
<i>hsa-mir-124-2</i>	Intergenic	hES, mES	-	3
<i>hsa-mir-542~450b</i>	Intergenic	hES, mES	-	3
<i>hsa-mir-219-2</i>	intergenic	hES, mES	-	1
<i>hsa-mir-708</i>	Intergenic	hES	-	1
<i>hsa-mir-365-2</i>	Intergenic	hES, mES	-	3
<i>hsa-mir-193a</i>	Intergenic	hES, mES	-	3
<i>hsa-mir-129-1</i>	Intergenic	hES, mES, MEF	-	2
<i>hsa-mir-129-2</i>	intergenic	hES, mES, MEF	-	3
<i>hsa-mir-124-1</i>	Intergenic	hES, mES, MEF	-	3
<i>hsa-mir-146b</i>	Intergenic	hES	<i>NFKB2</i>	0
<i>hsa-mir-370</i>	Intergenic	hES	-	0
<i>hsa-mir-124-3</i>	Intergenic	mES, MEF	-	3
<i>hsa-mir-17~92a-1</i>	Intergenic	MEF	-	1
<i>hsa-mir-182~183</i>	Intergenic	hES, MEF	-	4
<i>hsa-mir-1-1</i>	Intron of <i>C20orf166</i>	mES	<i>GATA5</i>	-
<i>hsa-mir-133a-2</i>	Intron of <i>C20orf166</i>	mES	<i>GATA5</i>	-
<i>hsa-mir-203</i>	Intergenic	mES, MEF	-	4
<i>hsa-mir-16-1~15a</i>	Intron of <i>DLEU2</i>	mES	-	-
<i>hsa-let-7a-3~let-7b</i>	Intergenic	mES	<i>WNT7B</i>	3

List of conserved human ST miRNAs that have been annotated as targets of putative long-range regulation during development and differentiation. The mir-9 family of miRNAs is highlighted since it contains know examples of GRB target miRNAs that were captured using our two prediction features: 1) localization in regions of high HCNE density and 2) association with a bivalent promoter.

¹The cell type in which the promoter of the miRNA is predicted to be a bivalent promoter. hES and mES represent embryonic stem cells of human and mouse, respectively. MEFs are mouse embryonic fibroblasts and NPCs are mouse neural progenitor cells.

²Further annotated GRB target genes within the 300kb region that was analyzed. The annotation of the GRB target genes was performed as described in the Methods section.

³Number of CpG islands overlapping with pri-miRNAs. The pri-miRNA is defined as the region 50 kb upstream to 20 kb downstream of the pre-miRNAs. If the defined pri-miRNA overlapped known protein-coding genes (the gene itself plus 1 kb up- and downstream of it), it was truncated to exclude the overlapping gene (see Methods). We did not count CGIs for intragenic miRNAs (denoted as '-').

Our results provide an important new resource for the analysis of miRNA regulation and greatly increase the number of miRNA genes under putative long-range transcriptional control, indicating that many miRNAs are subject to strict regulatory constraints whenever they are required to establish complex spatiotemporal developmental patterns.

Methods

Assemblies and annotation

The following genome assemblies were used: human NCBI 36, mouse NCBI 37 and mouse NCBI 36, chicken v2.1, *Xenopus tropicalis* v4.1 (US DoE JOINT genome

Institute), zebrafish Zv7 (The Wellcome Trust Sanger Institute). Genomic coordinates and sequences of known miRNAs were obtained from miRBase (Version 12)[1] and the genomic coordinates of HCNEs were made available from the Ancora database ([29]; <http://ancora.genereg.net>; Additional file 6, Table S4). CGI annotations, Broad Institute Chromatin State Mapping using ChiP-Seq and genomic sequences were extracted from the UCSC Genome Browser Database ([47]; <http://genome.ucsc.edu>) and gene annotations for human, mouse and zebrafish were obtained from Ensembl ([48]; <http://www.ensembl.org>; version 52) using Biomart ([49]; <http://www.biomart.org>). In addition, because of

the limited annotation of the zebrafish genome, we also included zebrafish orthologs of human genes predicted based on the similarity of exonic sequences and gene structures between zebrafish and human genomes [50]. The list of putative GRB target genes was constructed by inspecting HCNE density peaks, tracing the teleost gene/HCNE synteny, and analyzing functional gene annotation [25]. GRBs are human-zebrafish synteny blocks overlapping with GRB target genes. Synteny blocks were defined through the joined zebrafish to human high scoring net alignments downloaded from the UCSC Genome Browser [47].

miRNA classification

Human and zebrafish miRNAs were classified into intergenic and intragenic miRNAs based on the overlap of the pre-miRNA with known protein coding genes (at least 1 bp overlapping). Our ST miRNAs included both human intergenic miRNAs and human intragenic miRNAs that were intergenic in zebrafish genome, since these are likely to be self-transcribed. We removed the Pol III-transcribed miRNAs mentioned by Borchert et al [30] and subsequently retrieved orthologs of human ST miRNAs conserved at different evolutionary distances. For each lineage comparison we clustered human ST miRNAs into genomic clusters if they were less than 10 kb apart.

Identification of orthologous miRNAs

Orthologs of human ST pre-miRNAs were identified in seven other species (mouse, dog, opossum, platypus, chicken, frog and zebrafish). We used BLAST with default parameters, with the exception of word size, which was set to six, to compare all known pre-miRNAs and mature miRNAs between all species. The best reciprocal BLAST hits highlighted miRNA pairs that were chosen as putative orthologs. In cases where we could not find orthologs at the pre-miRNA level, we used the mature miRNA instead. For precursor orthologs, we required that the length of the aligned region was equal to or greater than 56% of the query length and that the gap-size was no greater than 10.

For orthologs identified via the mature miRNA, we required that the length of the aligned region was equal to or greater than 80% of the query length and that the number of non-matching bases is was no greater than three. Furthermore, the first two to eight bases had to be conserved.

Estimating the scale of pri-miRNAs

We used the annotation produced by Saini et al. [14] to examine the distance between putative miRNA transcription start sites/transcription ends and their associated

pre-miRNAs. We found that 93% of the TSSs were within 50 kb upstream of the pre-miRNAs and 90% of the miRNA transcription ends lay within 20 kb downstream of the pre-miRNAs (Figure 1). These results were used to define the general span of miRNA transcripts and extend the analysis window accordingly.

For the analysis of miRNAs with primary transcript predictions, i.e. the study of CGIs and bivalent promoters, we applied a search window that was the same size as the predictions from the study mentioned above [14]. Otherwise, the search window was 50 kb upstream and 20 kb downstream of the pre-miRNAs in the CGI analysis and only the region 50 kb upstream of the pre-miRNA in the bivalent promoter analysis. If the search window overlapped known/annotated protein-coding gene regions (the gene itself plus 1 kb up- and downstream of it) it was truncated to exclude the overlapping gene(s).

Analysis of enhancer density

Enhancer (HCNE or p300 binding site) densities were scanned using a 300 kb sliding window as described in Engström et al.[29]. For the density of the region around a sequence feature of interest (TSS for protein-coding genes), we used the HCNE density value in a 300 kb window centered on the feature.

We constructed 10000 non-coding/200 protein-coding control sets by randomly selecting the same number of coding and non-coding regions in the same genome as the conserved ST miRNAs. They had distributions of length, intergenic/intragenic ratio and gene density equal to that of the conserved ST miRNAs. For the comparison with GRB target genes, we used the whole set of GRB target genes.

Separation of ST miRNAs based on HCNE or p300 density

In the analysis of p300 density, we first ordered the human:mouse conserved ST miRNAs by human:mouse HCNE densities and mouse:human conserved ST miRNAs by p300 binding densities. Then we evenly divided human and mouse conserved ST miRNAs into two groups. The top half is located in regions enriched in HCNE/p300 binding (HCNE/p300 enriched miRNAs), whereas the remaining ST miRNAs were classified as being located in regions that are depleted in HCNE/p300 binding (HCNE/p300 poor miRNAs).

For predicting ST miRNAs that are located in regions of high HCNE density on a genome-wide scale, we ordered the conserved ST miRNAs by their HCNE densities, using HCNEs conserved between human and one of the seven vertebrate species shown in Table S4 and dividing them into HCNE enriched and HCNE poor ST miRNAs in the same way as mentioned above.

Identification of bivalent domains and bivalent promoters

Bivalent domains are regions that are enriched in both H3K4me3 and H3K27me3 histone modifications at the same developmental stage and the same tissue. We only predicted bivalent promoters for intergenic miRNAs since it was impossible to distinguish whether the features were associated with the ST miRNA or the host gene. Mikkelsen et al. [28] made the mouse histone modification information available and the human histone modification information was originally from Pan et al. [37]. We predicted bivalent promoters of protein-coding genes and miRNAs using their respective definitions.

Comparison of CpG island features

We compared the CpG island features between HCNE enriched and HCNE poor miRNAs as well as protein-coding GRB target genes, bystander genes and two other control sets, that were comprised of known transcription factors and CGI-associated genes. These gene sets were taken from the previous study of GRBs by Akalin et al. [25]. Contrary to this prior work, we only used the CpG-to-gene ratio in our analyses and not CGI length and count, since the length of the putative primary ST miRNAs is much shorter than that of protein-coding genes (p -value = $1e-03$; one-sided Wilcoxon test). Furthermore, we focused only on miRNAs that have primary transcript predictions in order to ensure a high level of accuracy.

Additional material

Additional file 1: Number of human ST miRNAs conserved in different vertebrate lineages. The number of human ST miRNAs investigated in our study that is conserved between human and different vertebrate lineages.

Additional file 2: Enrichment of HCNEs around human ST miRNAs, conserved in dog (A) and platypus (B). As in Figure 2, the graphs are cumulative HCNE density curves for conserved human ST miRNAs. The HCNEs shown here are conserved in dog (A) and platypus (B) as indicated at the top of each figure pair. In keeping with our results, conserved human ST miRNAs are more likely to be located in regions with higher HCNE density than would be expected by chance and this association extends to the entire vertebrate lineage.

Additional file 3: Enrichment of HCNEs around conserved human ST miRNAs (excluding ST miRNAs overlapping with GRBs). Each sub-figure shows the cumulative HCNE density curves for conserved human ST miRNAs in distinct lineage comparisons, but this time, excluding ST miRNAs overlapping with known GRBs. These results confirm that conserved human ST miRNAs are also more likely to be located in regions with higher HCNE density than would be expected by chance, independently of their association with any known GRBs and are therefore most likely the actual target of long-range regulation.

Additional file 4: Number of p300 binding regions overlapping mouse HCNEs. Total number of p300 binding regions and those overlapping with mouse HCNEs (with percentage of identity larger than 98% and length longer than 50 bp) in three mouse embryonic tissues.

Additional file 5: Functional annotation of miRNAs under putative long-range regulation. Functional information regarding the miRNA candidates we annotated as being under putative long-range regulation.

Additional file 6: The percentage of identity and length cutoffs of HCNEs. The lineage comparisons, percentage of identity and length cutoffs used to determine the HCNEs in our study.

List of abbreviations

CGI: CpG island; ES cell: Embryonic stem cell; EST: Expressed sequence tags; GRB: Genomic regulatory block; HCNE: Highly conserved non-coding element; MEF: Primary embryonic fibroblast; miRNA: microRNA; NPC: Neural progenitor cell; Pol II: RNA Polymerase II; Pol III: RNA Polymerase III; ST miRNA: Self-transcribed miRNA; TFBS: Transcription factor binding site; TSS: Transcription start site

Acknowledgements

We would like to thank Boris Lenhard, Altuna Akalin, David Fredman, Pär Engström and Xianjun Dong for all the valuable discussions and sharing their resources. In particular, we would like to thank Boris Lenhard and David Fredman for their invaluable comments on the manuscript. YS and CP were supported by grants awarded to Boris Lenhard by the Bergen Research Foundation (BFS) and the Young Future Research Leaders (YFF) program of the Norwegian Research Council (NFR), respectively.

Author details

¹Computational Biology Unit, Bergen Center for Computational Science, and Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, 5008, Norway. ²Department of Medical Genetics, Oslo University Hospital, Oslo, 0407, Norway.

Authors' contributions

YS and CP conceived and designed the experiments. YS performed the experiments. YS and CP wrote the manuscript. Both authors critically read and approved the final version.

Received: 23 February 2011 Accepted: 27 May 2011

Published: 27 May 2011

References

- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**(36 Database):D154-158.
- Kim DH, Saetrom P, Snove O Jr, Rossi JJ: **MicroRNA-directed transcriptional gene silencing in mammalian cells.** *Proc Natl Acad Sci USA* 2008, **105**(42):16230-16235.
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II.** *Embo J* 2004, **23**(20):4051-4060.
- Cai X, Hagedorn CH, Cullen BR: **Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs.** *Rna* 2004, **10**(12):1957-1966.
- Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE: **Trans-splicing and polyadenylation of let-7 microRNA primary transcripts.** *Rna* 2004, **10**(10):1586-1594.
- Houbaviy HB, Dennis L, Jaenisch R, Sharp PA: **Characterization of a highly variable eutherian microRNA gene.** *Rna* 2005, **11**(8):1245-1257.
- Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH, Feldmann G, Yamakuchi M, Ferlito M, Lowenstein CJ, et al: **Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis.** *Mol Cell* 2007, **26**(5):745-752.
- Woods K, Thomson JM, Hammond SM: **Direct regulation of an oncogenic micro-RNA cluster by E2F transcription factors.** *J Biol Chem* 2007, **282**(4):2130-2134.
- Gu J, He T, Pei Y, Li F, Wang X, Zhang J, Zhang X, Li Y: **Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences.** *Mamm Genome* 2006, **17**(10):1033-1041.
- Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE: **Chromatin structure analyses identify miRNA promoters.** *Genes Dev* 2008, **22**(22):3172-3183.
- He S, Su H, Liu C, Skogerbo G, He H, He D, Zhu X, Liu T, Zhao Y, Chen R: **MicroRNA-encoding long non-coding RNAs.** *BMC Genomics* 2008, **9**:236.
- Saini HK, Griffiths-Jones S, Enright AJ: **Genomic analysis of human microRNA transcripts.** *Proc Natl Acad Sci USA* 2007, **104**(45):17719-17724.

13. Zhou X, Ruan J, Wang G, Zhang W: **Characterization and identification of microRNA core promoters in four model species.** *PLoS Comput Biol* 2007, **3**(3):e37.
14. Saini HK, Enright AJ, Griffiths-Jones S: **Annotation of Mammalian Primary microRNAs.** *BMC Genomics* 2008, **9**(1):564.
15. Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ: **High-resolution human core-promoter prediction with CoreBoost_HM.** *Genome Res* 2008.
16. Erson AE, Petty EM: **MicroRNAs in development and disease.** *Clin Genet* 2008, **74**(4):296-306.
17. Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, et al: **Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates.** *Genome Res* 2007, **17**(5):545-555.
18. de la Calle-Mustienes E, Feijoo CG, Manzanares M, Tena JJ, Rodriguez-Seguel E, Letizia A, Allende ML, Gomez-Skarmeta JL: **A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts.** *Genome Res* 2005, **15**(8):1061-1072.
19. Loots GG, Kneissel M, Keller H, Baptist M, Chang J, Collette NM, Ovcharenko D, Plajzer-Frick I, Rubin EM: **Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease.** *Genome Res* 2005, **15**(7):928-935.
20. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al: **Highly conserved non-coding sequences are associated with vertebrate development.** *PLoS Biol* 2005, **3**(1):e7.
21. Jeong Y, El-Jaick K, Roessler E, Muenke M, Epstein DJ: **A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers.** *Development* 2006, **133**(4):761-772.
22. Sandelin A, Bailey P, Bruce S, Engstrom PG, Klos JM, Wasserman WW, Ericson J, Lenhard B: **Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes.** *BMC Genomics* 2004, **5**(1):99.
23. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al: **In vivo enhancer analysis of human conserved non-coding sequences.** *Nature* 2006, **444**(7118):499-502.
24. Li Q, Ritter D, Yang N, Dong Z, Li H, Chuang JH, Guo S: **A systematic approach to identify functional motifs within vertebrate developmental enhancers.** *Dev Biol* 2010, **337**(2):484-495.
25. Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B: **Transcriptional features of genomic regulatory blocks.** *Genome Biol* 2009, **10**(4):R38.
26. Tanay A, O'Donnell AH, Damelin M, Bestor TH: **Hyperconserved CpG domains underlie Polycomb-binding sites.** *Proc Natl Acad Sci USA* 2007, **104**(13):5521-5526.
27. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells.** *Cell* 2006, **125**(2):315-326.
28. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**(7153):553-560.
29. Engstrom PG, Fredman D, Lenhard B: **Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes.** *Genome Biol* 2008, **9**(2):R34.
30. Borchert GM, Lanier W, Davidson BL: **RNA polymerase III transcribes human microRNAs.** *Nat Struct Mol Biol* 2006, **13**(12):1097-1101.
31. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome.** *Annu Rev Genomics Hum Genet* 2006, **7**:29-59.
32. Eckner R, Ewen ME, Newsome D, Gerdes M, DeCaprio JA, Lawrence JB, Livingston DM: **Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor.** *Genes Dev* 1994, **8**(8):869-884.
33. Eckner R, Yao TP, Oldread E, Livingston DM: **Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation.** *Genes Dev* 1996, **10**(19):2478-2490.
34. Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R: **Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300.** *Cell* 1998, **93**(3):361-372.
35. Merika M, Williams AJ, Chen G, Collins T, Thanos D: **Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription.** *Mol Cell* 1998, **1**(2):277-287.
36. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al: **ChIP-seq accurately predicts tissue-specific activity of enhancers.** *Nature* 2009, **457**(7231):854-858.
37. Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, Jonsdottir GA, Stewart R, Thomson JA: **Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells.** *Cell Stem Cell* 2007, **1**(3):299-312.
38. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nat Genet* 2007, **39**(3):311-318.
39. Juan AH, Kumar RM, Marx JG, Young RA, Sartorelli V: **Mir-214-dependent regulation of the polycomb protein Ezh2 in skeletal muscle and embryonic stem cells.** *Mol Cell* 2009, **36**(1):61-74.
40. Wang H, Garzon R, Sun H, Ladner KJ, Singh R, Dahlman J, Cheng A, Hall BM, Qualman SJ, Chandler DS, et al: **NF-kappaB-YY1-miR-29 regulatory circuitry in skeletal myogenesis and rhabdomyosarcoma.** *Cancer Cell* 2008, **14**(5):369-381.
41. Krichevsky AM, Sonntag KC, Isacson O, Kosik KS: **Specific microRNAs modulate embryonic stem cell-derived neurogenesis.** *Stem Cells* 2006, **24**(4):857-864.
42. Coolen M, Bally-Cuif L: **MicroRNAs in brain development and physiology.** *Curr Opin Neurobiol* 2009, **19**(5):461-470.
43. Liu N, Okamura K, Tyler DM, Phillips MD, Chung WJ, Lai EC: **The evolution and functional diversification of animal microRNA genes.** *Cell Res* 2008, **18**(10):985-996.
44. Meyer A, Van de Peer Y: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27**(9):937-945.
45. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF: **The expansion of the metazoan microRNA repertoire.** *BMC Genomics* 2006, **7**:25.
46. Navratilova P, Fredman D, Hawkins TA, Turner K, Lenhard B, Becker TS: **Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes.** *Dev Biol* 2009, **327**(2):526-540.
47. Kuhn RM, Karolchik D, Zweig AS, Wang T, Smith KE, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pheasant M, et al: **The UCSC Genome Browser Database: update 2009.** *Nucleic Acids Res* 2009, **37** Database: D755-761.
48. Flicke P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Coates G, Fairley S, et al: **Ensembl's 10th year.** *Nucleic Acids Res* 2010, **38** Database: D557-562.
49. Cochrane G, Akhtar R, Aldebert P, Althorpe N, Baldwin A, Bates K, Bhattacharya S, Bonfield J, Bower L, Browne P, et al: **Priorities for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database.** *Nucleic Acids Res* 2008, **36** Database: D5-12.
50. Dong X, Fredman D, Lenhard B: **Synorth: exploring the evolution of syntenic and long-range regulatory interactions in vertebrate genomes.** *Genome Biol* 2009, **10**(8):R86.
51. Chopra VS, Mishra RK: **"Mir"acles in hox gene regulation.** *Bioessays* 2006, **28**(5):445-448.
52. McGlenn E, Yekta S, Mansfield JH, Soutschek J, Bartel DP, Tabin CJ: **In ovo application of antagomiRs indicates a role for miR-196 in patterning the chick axial skeleton through Hox gene regulation.** *Proc Natl Acad Sci USA* 2009, **106**(44):18610-18615.
53. Sehm T, Sachse C, Frenzel C, Echeverri K: **miR-196 is an essential early-stage regulator of tail regeneration, upstream of key spinal cord patterning events.** *Dev Biol* 2009, **334**(2):468-480.
54. Hou W, Tian Q, Zheng J, Bonkovsky HL: **MicroRNA-196 represses Bach1 protein and hepatitis C virus gene expression in human hepatoma cells expressing hepatitis C viral proteins.** *Hepatology* 2009.
55. Edbauer D, Neilson JR, Foster KA, Wang CF, Seeburg DP, Batterton MN, Tada T, Dolan BM, Sharp PA, Sheng M: **Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132.** *Neuron* 2010, **65**(3):373-384.

56. Lusardi TA, Farr CD, Faulkner CL, Pignataro G, Yang T, Lan J, Simon RP, Saugstad JA: **Ischemic preconditioning regulates expression of microRNAs and a predicted target, MeCP2, in mouse cortex.** *J Cereb Blood Flow Metab* 2009.
57. Shaked I, Meerson A, Wolf Y, Avni R, Greenberg D, Gilboa-Geffen A, Soreq H: **MicroRNA-132 potentiates cholinergic anti-inflammatory signaling by targeting acetylcholinesterase.** *Immunity* 2009, **31**(6):965-973.
58. Nudelman AS, Dirocco DP, Lambert TJ, Garelick MG, Le J, Nathanson NM, Storm DR: **Neuronal activity rapidly induces transcription of the CREB-regulated microRNA-132, in vivo.** *Hippocampus* 2009.
59. Yuen T, Ruf F, Chu T, Sealon SC: **Microtranscriptome regulation by gonadotropin-releasing hormone.** *Mol Cell Endocrinol* 2009, **302**(1):12-17.
60. Cheng HY, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, Nakazawa T, Shimizu K, Okamura H, Impey S, et al: **microRNA modulation of circadian-clock period and entrainment.** *Neuron* 2007, **54**(5):813-829.
61. Lukiw WJ: **Micro-RNA speciation in fetal, adult and Alzheimer's disease hippocampus.** *Neuroreport* 2007, **18**(3):297-300.
62. Haller F, von Heydebreck A, Zhang JD, Gunawan B, Langer C, Ramadori G, Wiemann S, Sahin O: **Localization- and mutation-dependent microRNA (miRNA) expression signatures in gastrointestinal stromal tumours (GISTs), with a cluster of co-expressed miRNAs located at 14q32.31.** *J Pathol* 2010, **220**(1):71-86.
63. Lund AH: **miR-10 in development and cancer.** *Cell Death Differ* 2010, **17**(2):209-214.
64. Choong ML, Yang HH, McNiece I: **MicroRNA expression profiling during human cord blood-derived CD34 cell erythropoiesis.** *Exp Hematol* 2007, **35**(4):551-564.
65. Huang ZM, Yang J, Shen XY, Zhang XY, Meng FS, Xu JT, Zhang BF, Gao HJ: **MicroRNA expression profile in non-cancerous colonic tissue associated with lymph node metastasis of colon cancer.** *J Dig Dis* 2009, **10**(3):188-194.
66. Silber J, Lim DA, Petritsch C, Persson AI, Maunakea AK, Yu M, Vandenberg SR, Ginzinger DG, James CD, Costello JF, et al: **miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells.** *BMC Med* 2008, **6**:14.
67. Bemis LT, Chen R, Amato CM, Classen EH, Robinson SE, Coffey DG, Erickson PF, Shellman YG, Robinson WA: **MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines.** *Cancer Res* 2008, **68**(5):1362-1368.
68. Baroukh NN, Van Obberghen E: **Function of microRNA-375 and microRNA-124a in pancreas and brain.** *Febs J* 2009, **276**(22):6509-6521.
69. Avissar M, Christensen BC, Kelsey KT, Marsit CJ: **MicroRNA expression ratio is predictive of head and neck squamous cell carcinoma.** *Clin Cancer Res* 2009, **15**(8):2850-2855.
70. Mathe EA, Nguyen GH, Bowman ED, Zhao Y, Budhu A, Schetter AJ, Braun R, Reimers M, Kumamoto K, Hughes D, et al: **MicroRNA expression in squamous cell carcinoma and adenocarcinoma of the esophagus: associations with survival.** *Clin Cancer Res* 2009, **15**(19):6192-6200.
71. Avissar M, McClean MD, Kelsey KT, Marsit CJ: **MicroRNA expression in head and neck cancer associates with alcohol consumption and survival.** *Carcinogenesis* 2009, **30**(12):2059-2063.
72. Hui AB, Lenarduzzi M, Krushel T, Waldron L, Pintilie M, Shi W, Perez-Ordenez B, Jurisica I, O'Sullivan B, Waldron J, et al: **Comprehensive MicroRNA profiling for head and neck squamous cell carcinomas.** *Clin Cancer Res* 2010, **16**(4):1129-1139.
73. Ladeiro Y, Couchy G, Balabaud C, Bioulac-Sage P, Pelletier L, Rebouissou S, Zucman-Rossi J: **MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations.** *Hepatology* 2008, **47**(6):1955-1963.
74. Avnit-Sagi T, Kantorovich L, Kredon-Russo S, Hornstein E, Walker MD: **The promoter of the pri-miR-375 gene directs expression selectively to the endocrine pancreas.** *PLoS One* 2009, **4**(4):e5033.
75. Visvanathan J, Lee S, Lee B, Lee JW, Lee SK: **The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development.** *Genes Dev* 2007, **21**(7):744-749.
76. Tang X, Tang G, Ozcan S: **Role of microRNAs in diabetes.** *Biochim Biophys Acta* 2008, **1779**(11):697-701.
77. Toloubeydokhti T, Pan Q, Luo X, Bukulmez O, Chegini N: **The expression and ovarian steroid regulation of endometrial micro-RNAs.** *Reprod Sci* 2008, **15**(10):993-1001.
78. Zhang MC, Lv Y, Qi YT, Zhang Z, Fu XN, Yuan CG, Lai LH: **[Knockdown and overexpression of miR-219 lead to embryonic defects in zebrafish development].** *Fen Zi Xi Bao Sheng Wu Xue Bao* 2008, **41**(5):341-348.
79. Schotte D, Chau JC, Sylvester G, Liu G, Chen C, van der Velden VH, Broekhuis MJ, Peters TC, Pieters R, den Boer ML: **Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia.** *Leukemia* 2009, **23**(2):313-322.
80. Ohlsson Teague EM, Van der Hoek KH, Van der Hoek MB, Perry N, Wagaarachchi P, Robertson SA, Print CG, Hull LM: **MicroRNA-regulated pathways associated with endometriosis.** *Mol Endocrinol* 2009, **23**(2):265-275.
81. Decembrini S, Bressan D, Vignali R, Pitto L, Mariotti S, Rainaldi G, Wang X, Evangelista M, Barsacchi G, Cremisi F: **MicroRNAs couple cell fate and developmental timing in retina.** *Proc Natl Acad Sci USA* 2009, **106**(50):21179-21184.
82. Liu SP, Fu RH, Yu HH, Li KW, Tsai CH, Shyu WC, Lin SZ: **MicroRNAs regulation modulated self-renewal and lineage differentiation of stem cells.** *Cell Transplant* 2009, **18**(9):1039-1045.
83. Katada T, Ishiguro H, Kuwabara Y, Kimura M, Mitui A, Mori Y, Ogawa R, Harata K, Fujii Y: **microRNA expression profile in undifferentiated gastric cancer.** *Int J Oncol* 2009, **34**(2):537-542.
84. Dyrskjot L, Ostenfeld MS, Bramsen JB, Silahatoglu AN, Lamy P, Ramanathan R, Fristrup N, Jensen JL, Andersen CL, Zieger K, et al: **Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro.** *Cancer Res* 2009, **69**(11):4851-4860.
85. Bandres E, Agirre X, Bitarte N, Ramirez N, Zarate R, Roman-Gomez J, Prosper F, Garcia-Foncillas J: **Epigenetic regulation of microRNA expression in colorectal cancer.** *Int J Cancer* 2009, **125**(11):2737-2743.
86. Ogawa R, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Katada T, Harata K, Tanaka T, Fujii Y: **Expression profiling of micro-RNAs in human esophageal squamous cell carcinoma using RT-PCR.** *Med Mol Morphol* 2009, **42**(2):102-109.
87. Merkerova M, Vaskova A, Belickova M, Bruchova H: **MicroRNA expression profiles in umbilical cord blood cell lineages.** *Stem Cells Dev* 2010, **19**(1):17-26.
88. Schmidt WM, Spiel AO, Jilma B, Wolzt M, Muller M: **In vivo profile of the human leukocyte microRNA response to endotoxemia.** *Biochem Biophys Res Commun* 2009, **380**(3):437-441.
89. Taganov KD, Boldin MP, Chang KJ, Baltimore D: **NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses.** *Proc Natl Acad Sci USA* 2006, **103**(33):12481-12486.
90. Patnaik SK, Kannisto E, Knudsen S, Yendamuri S: **Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection.** *Cancer Res* 2010, **70**(1):36-45.
91. Hussein K, Theophile K, Dralle W, Wiese B, Kreipe H, Bock O: **MicroRNA expression profiling of megakaryocytes in primary myelofibrosis and essential thrombocythemia.** *Platelets* 2009, **20**(6):391-400.
92. Melkamu T, Zhang X, Tan J, Zeng Y, Kassie F: **Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol.** *Carcinogenesis* 2010, **31**(2):252-258.
93. Raponi M, Dossey L, Jatko T, Wu X, Chen G, Fan H, Beer DG: **MicroRNA classifiers for predicting prognosis of squamous cell lung cancer.** *Cancer Res* 2009, **69**(14):5776-5783.
94. Iliopoulos D, Drosatos K, Hiyama Y, Goldberg IJ, Zannis VI: **MicroRNA-370 controls the expression of microRNA-122 and Cpt1{alpha} and affects lipid metabolism.** *J Lipid Res* 2010.
95. Xu CF, Yu CH, Li YM: **Regulation of hepatic MicroRNA expression in response to ischemic preconditioning following ischemia/reperfusion injury in mice.** *Omic* 2009, **13**(6):513-520.
96. Dixon-Mclver A, East P, Mein CA, Cazier JB, Molloy G, Chaplin T, Andrew Lister T, Young BD, Debernardi S: **Distinctive patterns of microRNA expression associated with karyotype in acute myeloid leukaemia.** *PLoS One* 2008, **3**(5):e2141.
97. Meng F, Wehbe-Janek H, Henson R, Smith H, Patel T: **Epigenetic regulation of microRNA-370 by interleukin-6 in malignant human cholangiocytes.** *Oncogene* 2008, **27**(3):378-386.

98. Volinia S, Galasso M, Costinean S, Tagliavini L, Gamberoni G, Drusco A, Marchesini J, Mascellani N, Sana ME, Abu Jarour R, et al: **Reprogramming of miRNA networks in cancer and leukemia.** *Genome Res* 2010, **20**(5):589-599.
99. Sun Y, Fang R, Li C, Li L, Li F, Ye X, Chen H: **Hsa-mir-182 suppresses lung tumorigenesis through down regulation of RGS17 expression in vitro.** *Biochem Biophys Res Commun* 2010.
100. Ivey KN, Muth A, Arnold J, King FW, Yeh RF, Fish JE, Hsiao EC, Schwartz RJ, Conklin BR, Bernstein HS, et al: **MicroRNA regulation of cell lineages in mouse and human embryonic stem cells.** *Cell Stem Cell* 2008, **2**(3):219-229.
101. Yi R, Poy MN, Stoffel M, Fuchs E: **A skin microRNA promotes differentiation by repressing 'stemness'.** *Nature* 2008, **452**(7184):225-229.
102. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, et al: **Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.** *Proc Natl Acad Sci USA* 2002, **99**(24):15524-15529.
103. Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, Petrelli NJ, Dunn SP, Krueger LJ: **MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells.** *Cancer Res* 2007, **67**(20):9762-9770.
104. Schultz J, Lorenz P, Gross G, Ibrahim S, Kunz M: **MicroRNA let-7b targets important cell cycle molecules in malignant melanoma cells and interferes with anchorage-independent growth.** *Cell Res* 2008, **18**(5):549-557.

doi:10.1186/1471-2164-12-270

Cite this article as: Sheng and Previti: Genomic features and computational identification of human microRNAs under long-range developmental regulation. *BMC Genomics* 2011 **12**:270.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

