HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2019 February 11.

Published in final edited form as:

Cell Rep. 2019 January 08; 26(2): 447–459.e4. doi:10.1016/j.celrep.2018.12.054.

Structural Differences between Pri-miRNA Paralogs Promote Alternative Drosha Cleavage and Expand Target Repertoires

Xavier Bofill-De Ros[1], **Wojciech K. Kasprzak**[2], **Yuba Bhandari**[3], **Lixin Fan**[4], **Quinn Cavanaugh**[1], **Minjie Jiang**[1], **Lisheng Dai**[1], **Acong Yang**[1], **Tie-Juan Shao**[1,5], **Bruce A. Shapiro**[6], **Yun-Xing Wang**[3], and **Shuo Gu**[1,7,*]

[1]RNA Mediated Gene Regulation Section, RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[2]Basic Science Program, RNA Biology Laboratory, Frederick National Laboratory for Cancer Research sponsored by the National Cancer Institute, Frederick, MD 21702, USA

[3]Protein-Nucleic Acid Interaction Section, Structural Biophysics Laboratory, National Cancer Institute, Frederick, MD 21702, USA

[4]Small-Angle X-ray Scattering Core Facility, Center for Cancer Research of the National Cancer Institute, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21702, USA

[5]School of Basic Medicine, Zhejiang Chinese Medical University, Hangzhou, 310053, China

[6]RNA Structure and Design Section, RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702, USA

[7]Lead Contact

SUMMARY

MicroRNA (miRNA) processing begins with Drosha cleavage, the fidelity of which is critical for downstream processing and mature miRNA target specificity. To understand how pri-miRNA sequence and structure influence Drosha cleavage, we studied the maturation of three pri-miR-9 paralogs, which encode the same mature miRNA but differ in the surrounding scaffold. We show that pri-miR-9-1 has a unique Drosha cleavage profile due to its distorted and flexible stem structure. Cleavage of pri-miR-9-1, but not pri-miR-9-2 or pri-miR-9-3, generates an alternative miR-9 with a shifted seed sequence that expands the scope of its target RNAs. Analyses of low-grade glioma patient samples indicate that the alternative-miR-9 has a potential role in tumor

*Correspondence: shuo.gu@nih.gov.

AUTHOR CONTRIBUTIONS

X.B.-D.R. and S.G. designed all experiments and wrote the paper, with contributions from all co-authors. X.B.-D.R. conducted the experiments and analysis, with help from Q.C., M.J., L.D., A.Y., and T.-J.S. Structural analysis and MD were done by W.K.K. and B.A.S. SAXS determination and ensemble analysis were done by Y.B., L.F., and Y.-X.W.

SUPPLEMENTAL INFORMATION

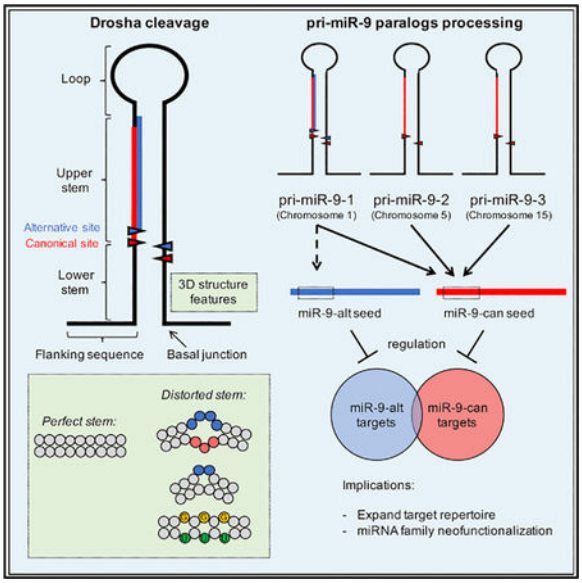
Supplemental Information includes five figures and two tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.12.054.

DECLARATION OF INTERESTS

The authors declare no competing interests.

progression. Furthermore, we provide evidence that distortion of pri-miRNA stems induced by asymmetric internal loops correlates with Drosha cleavage at non-canonical sites. Our studies reveal that pri-miRNA paralogs can have distinct functions via differential Drosha processing.

Graphical Abstract



In Brief

By studying the processing of pri-miR-9 family, Bofill-De Ros et al. demonstrate that the tertiary structure of pri-miRNA triggers alternative biogenesis. Drosha cleaves pri-miR-9-1 at an additional site because of its distorted and flexible lower stem, generating a 5’ isomiR that regulates a distinct set of genes in low-grade glioma.

INTRODUCTION

MicroRNAs (miRNAs) are a class of small non-coding RNAs (~22 nt) that are highly conserved among species (Bartel, 2018; Pasquinelli, 2012). Collectively, miRNAs negatively regulate the expression of >60% of human genes by post-transcriptional mechanisms (Friedman et al., 2009). It is well established that miRNAs play critical roles in the development of many diseases, including cancer (Lin and Gregory, 2015). Under the canonical biogenesis pathway, miRNAs are transcribed as long primary transcripts (pri-miRNAs), the maturation of which requires stepwise cleavage by two RNase III enzymes, Drosha and Dicer (Ha and Kim, 2014). Drosha and its cofactor DGCR8 form the Microprocessor complex in the nucleus, cutting the pri-miRNA to release the precursor miRNA (pre-miRNA) in the form of a short hairpin (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Lee et al., 2003). The pre-miRNA is exported to the cytoplasm, where it is processed further by Dicer (Grishok et al., 2001; Hutvágner et al., 2001; Knight and Bass, 2001). The resulting duplex is loaded onto one of four Argonaute proteins (AGO1–4), forming the RNA-induced silencing complex (RISC). During this process, one of the two strands (miRNA or guide strand) is retained, while the other (passenger strand) is discarded

(Khvorova et al., 2003; Schwarz et al., 2003). In mammals, most miRNAs are partially complementary to their targets, inducing translational repression and/or mRNA decay (Iwakawa and Tomari, 2015; Jonas and Izaurralde, 2015). The nucleotides at positions 2–7, counting from the miRNA 5’ end, are termed the seed region. Base-pairing of this region with its mRNA target is essential and sufficient for many miRNAs to function (Bartel, 2009).

The miRNA biogenesis machinery differentiates a wide variety of pri-miRNA hairpins from numerous other RNAs folded into similar structures. The pri-miRNA hairpin is structurally defined by the presence of a terminal loop, a stem of roughly three helical turns and basal unpaired flanking sequences. The stem is divided into an upper stem where the mature miRNA sequence is embedded and a lower stem next to the flanking region (Nguyen et al., 2015b). Drosha senses the basal junction between the lower stem and the single-stranded basal region, establishing a cleavage site 11 bp away (Han et al., 2006). In parallel, DGCR8 contacts the apical junction between the upper stem and terminal loop to facilitate proper recognition of pri-miRNA (Nguyen et al., 2015b; Zeng et al., 2005). Recent findings indicate that certain sequence motifs in the pri-miRNA scaffold also contribute to efficient processing of pri-miRNA in mammalian cells (Auyeung et al., 2013; Fang and Bartel, 2015).

Although both Drosha and Dicer cleavages define the sequence of mature miRNAs, the former is more critical because it defines the latter: Dicer cleaves at a fixed distance from the terminal ends of pre-miRNA (MacRae et al., 2007; Park et al., 2011); its cut site is determined largely by the Drosha cleavage site. Drosha cleavage at an alternative position generates a miRNA isoform (isomiRs) with a distinct 5’ end and an altered seed sequence. As these isomiRs will have altered target specificity, alternative cleavage by Drosha can profoundly affect miRNA function. Several studies have shown that changes in the relative distances between the expected cleavage site, the basal junction, and the apical junction of a pri-miRNA can all affect Drosha cleavage fidelity (Burke et al., 2014; Ma et al., 2013; Roden et al., 2017). However, despite the importance of Drosha cleavage in dictating miRNA specificity, the mechanisms by which Drosha selects its precise site of cleavage are largely unknown.

A class of pri-miRNAs that may be particularly susceptible to alternate cleavage by Drosha are those encoded by multigene families. Approximately 30%–40% of miRNAs are encoded by multiple loci that derive from gene duplications. Many of the miRNAs encoded by more than one loci are predicted to have identical seed sequences and overall homology (Berezikov, 2011). One example are the miRNAs encoded by the human pri-miR-9 family, which consists of three members encoded on chromosome 1 (pri-miR-9-1), chromosome 5 (pri-miR-9-2), and chromosome 15 (pri-miR-9-3). Despite the complete conservation of the mature miR-9 sequence within the upper stems of all three pri-miRNAs, variations occur at other positions that could potentially affect biogenesis. MiR-9 is highly expressed in brain, where it is involved in neural stem cell self-renewal and differentiation (Zhao et al., 2009) and synaptic plasticity (Sim et al., 2016). Mir-9 levels are also dysregulated in many human cancers (Ma et al., 2010; Nowek et al., 2018). Consistent with a role for Drosha cleavage fidelity in driving function, an isomiR of miR-9 with an altered 5’ end was detected in human embryonic stem cells and neural cells and shown to affect target site selection of

several reporter constructs (Tan et al., 2014.). However, both the mechanism by which this isomiR arises and its biological roles are unknown.

Here, we systematically investigate the biogenesis of miR-9 and its isomiR. We report that although canonical miR-9 is derived from all three pri-miRNAs, the miR-9 5' isomiR is primarily generated from pri-miR-9-1. We show that the distorted and flexible structure of the pri-miR-9-1 lower stem promotes Drosha cleavage at an alternative site, resulting in isomiR generation. Analysis of brain tumors transcriptomes suggests that the new isomiR downregulates specific targets to potentially regulate tumor progression. We provide evidence that non-canonical Drosha cleavage of other pri-miRNAs may similarly be driven by distortion of the pri-miRNA stems. Together, our findings reveal the pri-miRNA structural features that alter the fidelity of Drosha cleavage and provide insights into the mechanisms that result in neofunctionalization of miRNA paralogs.

RESULTS

Pri-miR-9-1 Generates a 5' IsomiR because of Its Unique Drosha Cleavage Profile

To investigate whether all three pri-miR-9 paralogs can produce mature miR-9, we first used a previously established reporter system to measure the efficacy of Drosha cleavage for each paralog (Dai et al., 2016). Briefly, three luciferase reporter constructs harboring each paralog sequence in their 3'UTR were co-transfected into Drosha-knockout cells with either a Drosha-expressing vector or a control plasmid. In this assay, Drosha cleavage of the pri-miRNA sequence triggers degradation of the luciferase transcripts and reduces luciferase activity. Drosha cleavage efficiency can thus be measured by comparing reporter activity in the presence or absence of Drosha (Figure S1A). The three pri-miR-9 paralogs on the reporter undergo Drosha processing with similar efficiencies (Figure 1A), indicating that they are all recognized by Drosha.

To further examine maturation, we cloned the genomic sequence of each pri-miR-9 paralog and its surrounding sequence into a CMV (Pol II)-driven transcription vector. By separately transfecting these plasmids into HEK293T cells, we were able to express each pri-miR-9 paralog independently. Northern blot analysis confirmed the production of mature miR-9 (hsa-miR-9–5p/MIMAT0000441) from all three pri-miR-9 transcripts (Figure 1B). Interestingly, multiple miRNA products varying in length were identified in all cases, suggesting alternative Drosha-Dicer cleavages and/or post-cleavage modifications. Endogenous miR-9 was barely detectable in the northern blot (Figure 1B), indicating that its abundance is marginal in these cells compared with the overexpressed miR-9.

To study the processing in detail, we deep-sequenced all miRNAs from the HEK293T cells with or without ectopic expression of pri-miR-9-transcripts and mapped the reads back to the corresponding pri-miR-9 paralog. Consistent with the northern blot result (Figure 1B), the level of endogenous miR-9 (654 ± 137 RPM) was much lower than that of the overexpressed miR-9 (23,034 ± 9,316 RPM) and therefore was neglected in subsequent analyses. Interestingly, although all pri-miR-9 transcripts produce the expected canonical miR-9 sequence (miR-9-can), an alternative miR-9 sequence (miR-9-alt) that begins 1 nt downstream of the miR-9-can is generated exclusively from pri-miR-9-1 (14% of total

reads) (Figure 1C). Both miR-9-can and miR-9-alt are heterogeneous in length because of their 3' tailing and trimming, as described previously (Berezikov et al., 2011; Gu et al., 2012). Thus, each band observed in the northern blot (Figure 1B) is a mixture of miR-9-can and miR-9-alt with various 3' sequence modifications. These results demonstrate that northern blot is insufficient to distinguish miR-9-alt from miR-9-can, which can be measured only by deep sequencing. Similar results were obtained when pri-miR-9 paralogs were expressed in HeLa cells: miR-9-alt (6% of the total reads) is generated from pri-miR-9-1, but not from pri-miR-9-2 or pri-miR-9-3, indicating that our observations are not limited to a single cell type (Figures S1B and S1C).

Given that post-maturation sequence modifications at the 5' of miRNA are extremely rare (Burroughs et al., 2010), the production of miR-9-alt is likely a result of Drosha cleavage at an alternative site. In fact, previous studies have established that the 5' end of mature miRNA is a faithful indicator of the Drosha cleavage site (Gu et al., 2012; Ma et al., 2013). To test this directly, we performed an *in vitro* Drosha cleavage assay on pri-miR-9-1 and pri-miR-9-2 transcripts (Figure 1D). Consistent with the *in vivo* results, Drosha processing of pri-miR-9-1, but not pri-miR-9-2, generates an additional pre-miR-9 product shorter than the expected size, indicating the use of an alternative cleavage site downstream of the canonical one. Moreover, data mining of a recent study in which Drosha cutting sites were mapped by sequencing the endogenous cleavage products (pre-miRNAs) crosslinked to Drosha (Kim et al., 2017) revealed that some pre-miR-9-1 RNAs, but not pre-mir-9-2 RNAs, contained the alternative Drosha 5' cleavage site (Figure S1D). Together, these results indicate that the 5' isomiR miR-9-alt is generated from pri-miR-9-1 because of a unique Drosha cleavage profile that differs from the other two primiR-9 paralog transcripts. Consistent with previous studies (Gu et al., 2012; Lee and Doudna, 2012; Park et al., 2011), this shift of Drosha cleavage site on pri-miR-9-1 also resulted in an alteration of Dicer cleavage pattern and eventually led to variation of strand selection during Argonaute loading (Figure S1E): more miR-9-3p reads (43%) were generated from pri-miR-9-1 than those from pri-miR-9-2 and pri-miR-9-3 (27% and 21%, respectively), demonstrating that the choice of Drosha cleavage site has a profound impact on miRNA biogenesis.

miR-9-alt Regulates a Novel Set of Targets in Low-Grade Glioma

The current model proposes that the seed region plays a critical role in defining miRNA target specificity (Bartel, 2009). Although miR-9-alt and miR-9-can originate from the same parental transcript (pri-miR-9-1), they have different seed sequences and therefore should each repress a unique set of targets. To evaluate this experimentally, we sought to decouple their biogenesis by expressing miR-9-can and miR-9-alt independently. We achieved this goal by constructing two well-designed U6 (Pol III)-driven short hairpin RNAs (shRNAs) (sh-miR-9-can and sh-miR-9-alt). Deep sequencing confirmed that each shRNA generates the corresponding mature miRNA isomiR with a negligible amount (<0.1%) of the other (Figure S2A). Luciferase reporters containing artificial target sites complementary to the seed sequence of either miR-9-can or miR-9-alt were co-transfected with these two shRNAs. As expected, sh-miR-9-can and sh-miR-9-alt only repress efficiently their corresponding target, establishing that we can measure the function of miR-9-can and miR-9-alt separately (Figure 2A). We used Targetscan (Agarwal et al., 2015), a well-established seed-based

program, to predict potential targets of miR-9-can and miR-9-alt. This endeavor identified 539 unique targets of miR-9-alt, thus increasing the number of potential targets of miR-9 paralogs by more than 57% compared with previously predicted targets of miR-9-can alone (Figure 2B). Using luciferase reporters, we validated three known targets of miR-9-can (FOXG1, GABRB2, and HES1) (Bonev et al., 2012; Pietrzykowski et al., 2008; Shibata et al., 2008) as well as three novel targets of miR-9-alt (CSGALNACT1, FOXN3, and PURB) (Figure S2B). Together, these results demonstrate that miR-9-alt has the potential to regulate a distinct set of target genes in cells.

To identify potential consequences of mir-9-alt expression, we searched datasets for cells and tissues in which this miRNA was well expressed. Consistent with previous reports describing the relevance of miR-9 in neuronal development and homeostasis (Coolen et al., 2013), we found that low-grade glioma (LGG) has the highest average expression of miR-9 among tumors documented in The Cancer Genome Atlas (TCGA) (Figure S2C). Further analysis revealed that the abundance of miR-9-alt, albeit only ~10% of miR-9-can, is still higher than the levels of miRNAs with well-established functions such as let-7, miR-30, or miR-21 (Figure 2C).

To investigate whether miR-9-alt regulates predicted miR-9-alt-specific targets in LGG, we took advantage of the RNA sequencing (RNA-seq) and miRNA sequencing (miRNA-seq) data in TCGA. Specifically, we compared mRNA expression profiles between patients with high levels of miR-9-alt (top 5%, 25 samples) and relatively low levels of miR-9-alt (bottom 5%, 25 samples) (Figure S2D). As expected, those mRNAs with one predicted target site of miR-9-alt were repressed in comparison with the whole transcriptome. Increased repression was observed with mRNAs containing multiple target sites, suggesting the repression is specific to miR-9-alt (Figure 2D). It is worth noting that the level of miR-9-can was correlated with that of miR-9-alt, therefore mRNAs targeted by miR-9-can were repressed as well. mRNAs containing target sites of both were further repressed, indicating that there is an additive effect between miR-9-can and miR-9-alt (Figure S2E). Applying 2-fold repression as a threshold, we identified 40 genes that are potential targets of miR-9-alt-specific regulation (Table S1).

Next, we sought to interrogate the potential biological impact of miR-9-alt in LGG. Gene Ontology analysis revealed an enrichment of extracellular matrix remodeling factors among the identified miR-9-alt targets. Many of these factors, including BACE2, COL1A2, and FGL2, have documented roles in promoting tumorigenesis (Balbous et al., 2014; Liu et al., 2010; Oliveras-Ferraros et al., 2014; Shin et al., 2017; Yan et al., 2015). Consistent with this idea, we found that lower levels of BACE2, COL1A2, and FGL2 correlate with increased survival time for the LGG patients (Figure 2E). Results of luciferase reporter as-says confirmed that these targets are specifically inhibited by miR-9-alt (Figure 2F), suggesting that miR-9-alt could function as a tumor suppressor by inhibiting BACE2, COL1A2, and FGL2. Although further studies are required to establish the definite roles miR-9-alt plays, our results demonstrate that miR-9-alt represses a novel set of target genes and correlates with changes in transcriptomic profile in glioma patients. Overall these highlight the importance to investigate the mechanism underlying its biogenesis.

The Lower Stem Plays a Major Role in Determining Cleavage Fidelity

Next, we sought to understand why miR-9-alt is generated exclusively from pri-miR-9-1. In these experiments, we investigated why Drosha processes pri-miR-9-1 and pri-miR-9-2 differently. The pri-miR-9-1 and pri-miR-9-2 scaffolds differ in three regions: the terminal loop (TL), the lower stem (LS), and the 5' and 3' flanking sequences (F) (Figure 3A). To determine which element drives the changes on Drosha cleavage fidelity, we swapped these regions between pri-miR-9-1 and pri-miR-9-2, generating six chimeric pri-miRNAs. These chimeric constructs were expressed individually in HEK293T cells. Drosha cleavage efficiency was measured using the reporter assay and northern blotting. In parallel, cleavage fidelity was determined by deep sequencing. All chimeras were processed with a similar efficiency (Figures S3A and S3B), indicating the overall miR-9 levels did not change. However, the fidelity of Drosha cleavage varied: the alternative cleavage is nearly abolished when the lower stem of pri-miR-9-1 is replaced with that of pri-miR-9-2 (F1-LS2-TL1), whereas exchanging the loop (F1-LS1-TL2) or the flanking sequences (F2-LS1-TL1) only partially reduces use of the alternative cleavage site (Figures 3B and S3C). Consistent with the finding that features of the lower stem are important for fidelity, a pri-miR-9-2 scaffold in which only the lower stem is replaced by that of pri-mir-9-1 stem (F2-LS1-TL2) shows an increase in the rate of alternative Drosha cleavage. In contrast, a pri-mir-9-2 scaffold containing either the pri-mir-9-1 loop (F2-LS2-TL1) or pri-mir-9-1 flanking sequences (F1-LS2-TL2) did not give increased cleavage at the alternative site (Figures 3C and S3D). Thus, the lower stem is the major determinant for inducing alternative cleavage of pri-miR-9-1.

Similar results were obtained when these chimeras were expressed in HeLa cells (Figures S3E and S3F). Taken together, our results demonstrate that the lower stem of pri-miR-9-1 is both necessary and sufficient for the increased level of Drosha alternative cleavage. Interestingly, the lower stem sequence of pri-miR-9-1 is well conserved in vertebrates. In contrast, the lower stem sequences of pri-miR-9-2 and pri-miR-9-3 as well as the terminal loop sequences of all pri-miR-9 paralogs are variable (Figure S3G). This suggests that Drosha alternative cleavage of pri-miR-9-1 and therefore the production of miR-9-alt may be evolutionarily conserved.

The Distorted Lower Stem Structure Drives Alternative Drosha Cleavage

The current model proposes that Drosha cleavage fidelity is determined mainly by the length of the lower stem, while the length of upper stem plays a minor role (Ma et al., 2013). Specifically, Drosha prefers to cleave at a 5' site 13 nt away and a 3' site 11 nt away from the basal junction (Auyeung et al., 2013; Han et al., 2006; Nguyen et al., 2015b). Consistent with this model, we found that the pri-miR-9-1 lower stem is a major determinant for Drosha cleavage fidelity. However, all of the pri-miR-9 paralogs' upper stem lengths are the same, and none of their lower stem lengths are optimal (Figure S4A), suggesting that additional factors underlie the Drosha alternative cleavage of pri-miR-9-1. We speculated that structural features might contribute to the choice of Drosha cleavage site. Indeed, secondary structure predictions revealed that pri-miR-9-1 contains an internal loop near the cut site (Figure S4A). However, pri-miR-9-2 and pri-miR-9-3 also contain internal loops and mismatches in their lower stems (Figure S4A), making an internal loop per se an unlikely cause of Drosha alternative cleavage.

To further investigate the structural differences between the three paralogs, we analyzed their tertiary structures with RNAComposer (Popenda et al., 2012). These initial predicted structures were then refined using molecular dynamics. In brief, the position and motion of each atom were calculated every 2 fs over the course of 250 ns. Interestingly, while pri-miR-9-2 and pri-miR-9-3 are predicted to maintain a relatively straight helix during the course of simulation, tertiary structure of pri-miR-9-1 is predicted to be distorted and flexible. The folding structure of pri-miR-9-1 with lowest energy is bent at the lower stem (Figure S4B). To validate these predicted topologies, we used small-angle X-ray scattering (SAXS). SAXS is a solution-based method that does not require crystallization and provides information on overall molecular size, shape, and intermolecular distances and dynamics (Fang et al., 2013). We used ensemble calculations to characterize small- and large-amplitude motions of the three-dimensional structure of the paralog stems. As predicted, the lower stem of pri-miR-9-1 is kinked whereas that of pri-miR-9-2 is not (Figures 4A, 4B, and S4C).

Because the sequence of the lower stem is conserved in vertebrates (Figure S3G), we tested whether the sequence of the internal loop was important by creating two additional primiR-9-1 constructs (pri-miR-9-1-a and pri-miR-9-1-b) with changes in the lower stem bulge nucleotide composition (Figure 4C). Interestingly, although both pri-miR-9-1-a and primiR-9-1-b have the same secondary structure as that of primiR-9-1, their three-dimensional (3D) predicted structures and dynamics are different. In fact, multiple structures with variable degree of distortion at the lower stem were generated for both pri-miR-9-1-a and pri-miR-9-1-b, indicating that their tertiary structures are highly flexible (Figure 4C). Correspondingly, the frequency of Drosha cleavage at the alternative site is much higher (54% and 83%, respectively) in HEK293T cells and similar results were obtained in HeLa cells (Figures 4C, S4D, and S4E). These results suggest that the distorted and flexible lower stem structure, which is apparently a result of the asymmetrical internal loop, correlates with alternative Drosha cleavage. To establish causality, we sought to correct the distortion and flexibility of the lower stem of pri-miR-9-1 without changing its length (Figure S4F). We achieved this goal by either replacing the bulge with a perfect stem (pri-miR-9-1-perfect-stem) or adding a single nucleotide to the bulge (pri-miR-9-1-bulged-stem) (Figure 4D). Molecular dynamics modeling confirmed that these modified structures form a relatively rigid and straight helix at the lower stem (Figure S4G). Deep sequencing analysis revealed that both of the modified pri-miR-9-1 s are nearly free of Drosha-mediated alternative cleavage, instead resulting in a cleavage profile similar to that of pri-miR-9-2 in both HEK293T and HeLa cells (Figures 4D, S4H, and S4I). These results demonstrate that a 3D structural feature of pri-miRNA, specifically a distorted and flexible stem, affects Drosha cleavage fidelity.

Pri-miRNA Tertiary Structure-Based Alternative Drosha Cleavage Is a General Mechanism for IsomiR Production

Finally, we investigated the extent to which tertiary structure affects the choice of Drosha cleavage sites for pri-miRNAs in general. To this end, we measured the Drosha cleavage fidelity of the top 200 miRNAs that are highly expressed in various tissues on the basis of published miRNA-seq results. We focused on high-confidence 5’-arm-miRNAs (5p)

(Fromm et al., 2015) because Drosha cleavages of these miRNAs may produce isomiRs with shifted seed sequences. The weighted average number of isomiRs with distinct 5' ends was calculated for each individual miRNA and used as a measurement of Drosha alternative cleavage. Although the majority of the pri-miRNAs are processed precisely by Drosha, many pri-miRNAs have an alternative Drosha cleavage frequency higher than that of pri-miR-9-1 (Figure 5A). The cleavage profiles of each pri-miRNA are largely consistent among a wide range of tissues, which supports the idea that an intrinsic feature such as tertiary structure is a major determinant for Drosha cleavage fidelity. However, a subset of pri-miRNAs displayed distinct Drosha cleavage profiles between tissues (Figure 5A), suggesting that tissue-specific factors may also affect cleavage fidelity, possibly by modulating pri-miRNA structure.

All pri-miRNAs were ranked by the average number of 5' isomiRs each produces (Figure 5A). The top 50 are classified as the low Drosha cleavage fidelity group, whereas the bottom 50 are classified as the high-fidelity group. To analyze the lower stem length, we aligned all pri-miRNA sequences (miRBase version 21) at the canonical Drosha cleavage site and calculated the percentage of paired bases at each position. The pri-miRNA basal junction can be inferred by a clear transition from unpaired (flanking sequence) to paired regions (lower stem). Consistent with previous studies (Nguyen et al., 2015b), we observed that the average length of the lower stem, defined by the distance between the basal junction and the Drosha cleavage site, is 13 nt on the 5p strand and 11 nt on the 3p strand (Figures 5B and S5A). This result supports the current model in which the Drosha cleavage site is determined mostly by its distance to the basal junction (Nguyen et al., 2015b). However, the same analysis of the low-fidelity group and the high-fidelity group revealed that both groups have better defined basal junctions at expected positions compared with pri-miRNAs on average, suggesting that the distance by itself is insufficient to determine Drosha cleavage fidelity.

In parallel, we tried to analyze the 3D structures of primiRNAs in both groups. Because a high-throughput approach in determining RNA tertiary structure is unavailable, we sought to first measure asymmetrical bulges on pri-miRNA stems, a two-dimensional (2D) feature associated well with the stem distortion and flexibility in the case study of pri-miR-9 family. To this end, we calculated for each pri-miRNA an asymmetrical score, which is the sum of difference between numbers of nucleotides on each side of every bulge along the stem. Interestingly, although the average number of mismatches on pri-miRNA stem is similar (Figure S5B), the low-fidelity group has a higher average asymmetrical score (Figure 5C), indicating that their stem structures are potentially more distorted and flexible. In line with this, most pri-miRNAs with low Drosha cleavage fidelity are predicted to have a distorted structure, whereas the majority of the pri-miRNAs with high Drosha cleavage fidelity are relatively straight (Figure S5C). To measure their tertiary curvature quantitatively, we superimposed each structure with an ideal A-form RNA helix and calculated the root-mean-square deviation (RMSD) between their backbones (Figure 5D). A more distorted stem structure results in a larger RMSD value. As expected, pri-miRNAs in the low-fidelity group have a higher average RMSD than these of the high-fidelity group, implying that the overall distortion of a pri-miRNA stem affects Drosha cleavage fidelity. To validate this model, we selected two pri-miRNAs from the low-fidelity group (pri-miR-411 and pri-miR-10b) and corrected the structural distortions on their lower stems by mutating the lower stem

sequences in a similar strategy as we did with pri-miR-9-1. Specifically, asymmetrical internal loops were removed and the G:U wobble pairs were replaced with Watson-Crick base pairs. Consistent with the results of pri-miR-9-1, such corrections abolished the Drosha alternative cleavage (Figures 5E and 5F).

Together, these results demonstrate that the pri-miRNA tertiary structure, specifically the distorted stem correlates with alternative Drosha cleavage and production of isomiRs with altered seed sequences.

DISCUSSION

The roles of RNase III enzymes in processing structured RNAs and in regulating gene expression are conserved from prokaryotic to eukaryotic cells (Court et al., 2013). Drosha, in particular, is essential for the maturation of most miRNAs. Understanding how Drosha precisely determines its cleavage sites is of great importance because this cleavage is the initial step in miRNA maturation. Downstream events, such as Dicer cleavage, RISC loading, post-maturation tailing-trimming, and eventually miRNA target selection, all depend on the position where Drosha cleaves (Figure S1E). Here, by combining deep sequencing and structural analysis, biochemical assays and functional studies, we demonstrated that pri-miRNA structure affects Drosha cleavage fidelity. Importantly, we demonstrated that miR-9-alt, an isomiR resulting from an apparently low-frequency alternative Drosha cleavage, recognizes a distinct set of targets and is abundant in LGG, where it may function as a tumor suppressor. Thus, alterations in Drosha fidelity may profoundly influence cell function.

Previous models for Drosha cleavage focused on one-dimensional information (sequence motifs) and 2D structural characteristics of pri-miRNAs. Here, our results demonstrate that tertiary structure affects Drosha cleavage fidelity. Specifically, we establish that eliminating distortion and flexibility of the lower stem leads to the diminished Drosha cleavage ambiguity. It is intriguing to speculate about the underlying molecular mechanism: given that RNase III enzymes in bacteria and yeast only accommodate substrates in the shape of A-form helices (Gan et al., 2006), it is reasonable to assume that Drosha has a similar preference based on its known structure (Kwon et al., 2016). Hence, pri-miRNAs with a bent or distorted stem in solution may need to alter its 3D conformation to fit in the Microprocessor complex. One possibility is that tension arising from this process induces Drosha cleavage at an alternative site. This model aligns well with previous studies of the other mammalian RNase III enzyme (Dicer) in which asymmetrical structural motifs in precursor hairpins, which are likely source of tertiary structure bending, induce Dicer alternative cleavages (Starega-Roslan et al., 2011). Alternatively, pri-miRNAs with distorted stems have higher flexibility, allowing them to fold into several distinct structures when complexing with Drosha. In this case, the alternative cleavage site may be a result of different configurations of the catalytic center and substrate. Future high-resolution structures of the ternary complex formed by Drosha, DGCR8 and pri-miRNA should give additional insights into the underlying mechanisms of Drosha alternative cleavage.

Although the existence of miR-9-alt was previously reported (Tan et al., 2014), we have identified a role for this miRNA in LGG tumorigenesis. The host genes of pri-miR-9-1 and primiR-9-2 are highly expressed (fragments per kilobase of transcript per million mapped reads [FPKM] > 10) in LGG (Figure S2F) and both contribute to the production of miR-9-can. miR-9-alt, on the other hand, originates exclusively from pri-miR-9-1 (Figure 1C). In support of a potential tumor suppressor role of miR-9-alt, the level of pri-miR-9-1, but not pri-miR-9-2, is positively correlated with LGG patient survival with statistical significance ($p < 0.01$) (Figures S2G and S2H). Moreover, we also identified the glutamate receptor NMDA2A as a target of miR-9-alt (Figure S2I). Thus, miR-9-alt likely functions in brain physiology as well. The lower stem sequences are more conserved in primiR-9-1 than the other two paralogs (Figure S3G), which is consistent with a potential evolutionarily conserved role in maintaining alternative processing of pri-miR-9-1 to generate miR-9-alt. However, it is puzzling that the first nucleotide of miR-9-alt is a “C,” which should be less favorable for Argonaute binding (Frank et al., 2010). One possibility is that the “C” is preserved as part of the miR-9-can seed sequence, allowing the canonical miR-9 to regulate specific mRNA targets that are conserved in most vertebrates.

We have established the specificity of miR-9-can and miR-9-alt on their cognate targets using reporter assays (Figures 2A, 2F, and S2B). Because the endogenous miR-9-can is normally present at much higher levels than miR-9-alt, we wondered whether the target selectivity was also maintained between the endogenous miR-9 isomiRs. To address this concern, we performed a meta-analysis of the AGO CLIP data on mouse brain (Moore et al., 2015), where the chimeric reads indicate direct interactions between a miRNA and its corresponding targets. Although there is a certain degree of cross-recognition, each isomiR still prefers its cognate targets despite the difference in expression levels between the two isomiRs (Figure S2J). However, we cannot exclude the possibility that part of the observed repression of endogenous miR-9-alt targets was partially mediated by miR-9-can. Nonetheless, given that miR-9-can is far more efficient in inhibiting its own targets than miR-9-alt targets, the fact that miR-9-alt targets were repressed to a similar extent as miR-9-can targets in LGG patients (Figure S2E) indicates that miR-9-can cannot be the sole cause and that miR-9-alt plays a physiological role.

We determined the pri-miRNA structural features that govern the biogenesis of miR-9-alt. In addition, we demonstrated that tertiary structure-based Drosha alternative cleavage is likely to be responsible for the generation of most, if not all, 5p isomiRs. Many factors are known to fine-tune Drosha cleavage efficiency (Ha and Kim, 2014; Jiang et al., 2017). Interestingly, we have shown that Drosha cleavage fidelity can vary among cell lines (Figure S1C) and tissues (Figure 5A), suggesting that it is not merely an intrinsic feature of Drosha but is subject to regulation, perhaps by changing the structure of the RNA substrate. HnRNP A1 was reported to promote Drosha cleavage efficacy of primiR-18a by altering its structure (Michlewski et al., 2008). It is intriguing to hypothesize that other RNA binding proteins and helicases affect Drosha cleavage fidelity via modulating the folding of pri-miRNA. Future studies are required to understand how the cellular environment regulates Drosha cleavage fidelity and how the resulting isomiR changes affect cell physiology and disease development.

Although pri-miRNA paralogs are usually expected to function similarly, a previous report suggested that individual paralogs can have distinct functions (Ruby et al., 2007a). Our demonstration that pri-mir-9-1 is primarily responsible for the production of miR-9-alt has provided direct evidence to support this idea. It was also proposed that during evolution, selective pressures stabilize alternative Drosha cleavage events, leading to them becoming the dominant cutting site and resulting in the formation of new miRNAs (Berezikov, 2011; Ruby et al., 2007a). In line with this, *Drosophila* pri-miR-4 has high sequence homology to the *Drosophila* primiR-9 paralogs (9a, 9b, and 9c) but has a different Drosha cleavage profile. The main cut site of the *Drosophila* pri-miR-4 is the same as the alternative Drosha cleavage site of human pri-miR-9-1, suggesting pri-miR-4 evolved from pri-miR-9 paralogs in flies. Thus, our work provides mechanistic insights into how miRNA paralog genes can be used as substrates to generate novel miRNAs. In this model of neofunctionalization, Drosha plays a central role in miRNA diversification and specificity.

Finally, the studies presented here have implications for shRNA design. A source of off-target effects originates from heterogeneous products of shRNA processing *in vivo*. We previously established the “loop-counting rule” of Dicer processing, which laid the groundwork for designing Pol III-driven pre-miRNA-like shRNAs free of heterogeneous processing (Gu et al., 2012). Second generation pri-miRNA-like shRNAs, which can be expressed from a Pol II promoter and are thus more amenable to transcriptional control, require Drosha processing (Bofill-De Ros and Gu, 2016). Here, we demonstrated that imprecise Drosha cleavage generates miRNA or small interfering RNA (siRNA) with shifted seed sequences, generating undesired off-target repression. Hence, the tertiary structure of shRNA should be taken into consideration to avoid causing Drosha miscleav-ages. Our results provide additional guidelines for designing shRNAs with reduced off-target effects, which can then be used as tools for biological discovery and therapeutics.

STAR★METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Shuo Gu (shuo.gu@nih.gov).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines—HEK293T (Female), HeLa cells (Female) and derived knockout cells were maintained in DMEM high glucose (GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (Hyclone), 100 U/ml penicillin-streptomycin (GIBCO) at 37°C. Cells were tested to be free of mycoplasma contamination. Transfections were performed using PolyJet DNA Transfection Reagent (SignaGen) according to the manufacturer’s instructions.

METHOD DETAILS

***In vivo* Drosha cleavage assay—**Luciferase-based reporters were generated on the psiCHECK-2 vector (Promega). Pri-miR-9-1/−2/−3 inserts containing the premiRNA and ~200 nt flanking sequences at both ends were amplified by PCR from genomic DNA and inserted into the 3’UTR of the *Renilla* luciferase gene. Primers used in cloning are listed in

Table S2. 50 ng of the pri-miRNA reporter plasmids were cotransfected with either 50 ng of Drosha-expressing plasmids or empty vector in Drosha-KO cell lines. Cell lysates were obtained 48h post-transfection. Firefly and *Renilla* enzymatic activity were measured with Dual-Luciferase® Reporter Assay System (Promega) and detected by GloMax®-Multi Luminescence Module (Promega) according to the manufacturer's protocol. Microprocessor cleavage efficiency was calculated as follows:

$$\eta(Microprocessor) = \left(1 - \frac{\left(\frac{RLuc\ Drosha\ Rescued\ cells}{FLuc\ Drosha\ Rescued\ cells}\right)}{\left(\frac{RLuC\ Drosha\ KO\ cells}{FLuc\ Drosha\ KO\ cells}\right)}\right) \cdot 100$$

measured as the percentage of reporter activity in Drosha rescued to that where Drosha is knocked-out. Microprocessor efficiency was calculated as the complement of the Drosha cleavage ratio.

In vitro* Drosha cleavage assay—In vitro* Drosha cleavage assay was performed similarly as previously described (Nguyen et al., 2015b). In brief, FLAG-tagged Drosha and DGCR8 were co-expressed in HEK293T cells. At 48h post-transfection, Microprocessor complex was isolated by anti-FLAG immunoprecipitation. Pri-miRNA transcripts consisted of pre-miRNA and 200 nt flanking sequences on both sides were *in vitro* transcribed, purified, heat to 85°C and then refolded by slow cooling to 25°C for 15 minutes (ramp rate 0.1°C/sec). The microprocessor complex and pri-miRNAs were incubated at 37°C in reaction solution (6.4mM $MgCl_2$ and 40U of RNase inhibitor) for 3h. Cleavage products were recovered using an acid phenol/chloroform extraction. Pre-miRNA products generated in the cleavage assay were detected by Northern blot.

MicroRNA repression assay—Briefly, psiCHECK-2 reporters (Promega) with target sites for miR-9-alt or miR-9-can were inserted into the 3'UTR of the *Renilla* luciferase gene. Both strands of the target sequence were chemically synthesized, phosphorylated and ligated into the linearized psiCHECK-2 vector. Similarly, sh-miR-9-can/alt and sh-control were cloned downstream of a U6 promoter. Primers and oligonucleotides used in cloning are listed in Table S2. 50 ng of the target reporter plasmids were co-transfected with either 50 ng of sh-miR-9-can/alt or sh-control into HEK293T cells. Cell lysates were obtained 48h post-transfection and measured with the Dual-Luciferase® Reporter Assay System (Promega).

Northern blot—Total RNA was isolated using Trizol (Life Technologies) and fractionated in 20% (w/v) acrylamide/8M urea gels. After, RNA was transferred to Hybond-N1 membranes (Amersham Pharmacia Biotech), crosslinked and blocked (PerfectHyb Plus Hybridization Buffer - Sigma). MicroRNA-9 (5p strand) was detected using ^{32}P-labeled oligonucleotides. Images were obtained and analyzed using Amersham Typhoon (GE Healthcare).

Small RNA NGS library preparation—5 μg of total RNA was ligated with RNA 3' adaptor using T4 RNA Ligase 2 - truncated (NEB), in the presence of RNase Inhibitor (NEB). RNA 5' adaptor was ligated using T4 RNA Ligase 1 - high concentration (NEB) and

10 mM ATP. Ligated small RNAs were reverse transcribed using SuperScript® IV Reverse Transcriptase (Thermo-Fisher). Small RNA library cDNA was amplified and indexed using Phusion® High-Fidelity DNA polymerase (NEB). Constructs were purified in a 6% (w/v) native acrylamide gel based on the expected product size and purified by ethanol precipitation. Library quality was assessed by using Qubit dsDNA HS Assay Kit (ThermoFisher) and Agilent High Sensitivity DNA kit (Agilent). Libraries were mixed together and prepared at a final concentration of 12pM and run on MiSeq Reagent Kit v3 (Illumina) according to the manufacturer's specifications. Adaptors, primer sequences and detailed protocol temperatures can be found in Table S2.

Secondary and tertiary structure prediction—The sequences of pri-miR-9-1/−2/−3 was obtained from miRBase v21 (Kozomara and Griffiths-Jones, 2014). Pri-miRNA sequences, excluding nucleotides upstream or downstream of the lower stem (Figure S4A) were subjected to folding with RNAstructure, using the default parameters for RNA to find the minimum free energy structure and close suboptimal solutions (Mathews et al., 2010). Predicted secondary structures were used as input to RNAComposer, a 3D structure prediction software (Popenda et al., 2012). 3D models were subjected to further refinements via molecular dynamics simulations with AMBER 14 package (Assisted Model Building with Energy Refinement). The AMBER force field ff14SB with ff99bsc0 and chi.OL3 parameter refinements for RNA were employed (Zgarbová et al., 2011). Implicit solvent simulations utilizing the Generalized Born model (GB) were performed, utilizing the latest corrections to the intrinsic Born radii parameters (mbondi3) in a GB-neck2 protocol (AMBER flag igb = 8)(Nguyen et al., 2015a; Tsui and Case, 2000–2001). Simulations were run at 310 K, with a 2 fs time step and a Debye-Hückel (monovalent) salt screening concentration of 1.0. No cutoff was imposed on nonbonded interactions (cut = 999). The SHAKE algorithm was used to constrain all hydrogen bonds. The Langevin thermostat was employed with a collision frequency of 1.0 ps^{-1}. A six-step 2.0 ns-long equilibration protocol was used that included energy minimization, heating to the target temperature of 310 K with harmonic restraints of 15 kcal/mol/$Å^2$ on the RNA, followed by short MD stages with harmonic restraints gradually lowered from 10.0, down to0.01 kcal/mol/$Å^2$. Unrestrained (production) MD simulations were calculated every 2 femtoseconds for 250 ns. Post-processing of the MD trajectories (RMSD and average structure calculations) were generated with the cpptraj program of AMBER. Curvature of pri-miRNA tertiary structures was measured by calculating RMSD between pri-miRNA and a perfect A-form RNA helix using Pymol v2.0 (Schrodinger, LLC).

SAXS sample preparation and data collection—Synthetic RNA (IDT) was used to synthesize exclusively the lower-stem, upper-stem and loop of pri-miR-9-1 (87nt) and pri-miR-9-2 (85nt). Concentration series SAXS measurements were carried out in order to remove the scattering contribution due to interparticle interactions and to extrapolate the data to infinite dilution. These RNAs were suspended in DEPC-treated water (Invitrogen) prepared at 3 concentrations (0.75, 1.2 and 1.8 mg/mL) in 50 mM HEPES buffer (GIBCO). The optimization of sample condition and screening of samples were performed on in-house SAXS instrument of NCI SAXS core. The selected samples were then measured at 12-ID-B beamline at the Advanced Photon Source in the Argonne National Laboratory. The

procedures for data collection, processing, and analysis are similar to that previously described (Fang et al., 2013). The buffer background subtraction and intensity extrapolation to infinite dilution were carried out using NCI in-house developed MATLAB script NCI-SAXS Core Facility.

SAXS ensemble calculation—The ensemble calculation was performed using an NCI-SAXS-WAXS module deployed in the Xplor-NIH environment as published protocol (Schwieters and Clore, 2007). The NCI-SAXS-WAXS module allowed a simultaneous calculation of fitness between the experimental and back-calculated data in both SAXS regions. The equally sparse SAXS data, with q ranging from 0.004 to0.89$Å^{-1}$ (a total of 72 data points), was used during the SAXS-restrained ensemble calculation. The difference in SAXS curves between experimental and back-calculated data is expressed as χ^2, as defined:

$$x^2 = \frac{1}{N-1}\sum_{i}^{N}\left[\frac{I^{\exp}(q_i) - cI^{cal}(q_i)}{\sigma(q_i)}\right]^2$$

where c is a scaling factor, sigma(q_i) is the experimental error, and $I^{exp}(q_i)$ and $I^{cal}(q_i)$ are the experimental and back-calculated scattering intensities of the i^{th} data point of the total N data points. A weighted harmonic energy potential function was used ($E_{SAXS} = C_{SAXS}X^2$), where C_{SAXS} is a scaling factor. During the calculation, no Rg restraint was applied so that the ensemble outputs were allowed to freely sample the conformational space. Restraints were applied to maintain covalent geometry, prevent atomic overlap, and maintain the provided Watson-Crick and G:U wobble base-pairing in the duplexes. Knowledge-based restraints were applied to nucleic acid torsion angle conformations and to base-base packing (Kuszewski et al., 1997). The ensemble calculation was analyzed using an NCI-SAXS-WAXS data analysis module. All computation modules, scripts that contain all calculation parameters and conditions, and ensemble of the structures and SAXS data used for the calculation are provided upon request to the authors.

MicroRNA expression and 5’ isomiR analysis—MicroRNA expression and 5’isomiR analysis were performed using QuagmiR on Amazon cloud instances through the Seven Bridges Genomics implementation of the NCI Cancer Genomics Cloud. QuagmiR is a customized Python scripts for the motif-based alignment and analysis of miRNA (Bofill-De Ros et al., 2018). The analysis of The Cancer Genome Atlas (TCGA) was also performed using QuagmiR, with a previous conversion of the bam files to fastq files by Picard Sam-to-Fastq. Meta-analysis on Figures S1D and S2J was obtained using custom R scripts (https://github.com/Gu-Lab-RBL-NCI).

Calculation of number of 5’ isomiRs—The weighted average number of 5’ isomiRs was calculated using an inverse Simpson index. This index measures the evenness of the 5’ isomiRs generated by each individual or family of paralog pri-miRNAs. The number of 5’ isomiRs ($1/\lambda$) was calculated as defined:

$$\frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{n} p_i^2}$$

where n is each of the 5’ isomiRs detected and p is the weighted frequency of reads for that given 5’ isomiR in the sample. Normal tissue samples from TCGA (n = 568) were used to generate the weighted average number of 5’ isomiRs in the corresponding tissue. Based on their average expression levels, we selected the top 200 most abundant miRNAs for heat-map plotting.

Lower stem secondary structure analysis and asymmetry score—Genomic sequences of pri-miRNAs were obtained from UCSC Genome Browser. Drosha-independent miRNAs, such as mirtrons (Ruby et al., 2007b), 5’ capped-miRNAs (Xie et al., 2013) and TSS-miRNAs (Zamudio et al., 2014), in addition to low confidence miRNAs lacking documented validations, were excluded in our analyses. High confidence miRNA were annotated based on (Fromm et al., 2015). Each pri-miRNA analyzed consists of the pre-miRNA and flanking sequences of 30 nt on each side. The secondary structure was obtained (Gruber et al., 2008), and the bracket-dot notation of the lower stem was analyzed using custom R scripts (https://github.com/Gu-Lab-RBL-NCI). In brief, we first tested whether a nucleotide in one position is paired with another on the other side of the pre-miRNA. Nucleotides that failed such a test were labeled as unpaired. Then, we aligned all pri-miRNA 5p sequences by the 5’ Drosha cleavage site (5’ end of pre-miRNA) and aligned all 3p sequences by the 3’ Drosha cleavage site (3’ end of pre-miRNA). The fraction of paired nucleotides was calculated for each position and was plotted against its relative distance to the Drosha cleavage site. Similarly, we calculated an asymmetry score for each pri-miRNA. First, every bulge on the stem was analyzed individually and the absolute differences between numbers of nucleotides on each side was calculated. Asymmetry score was defined as the sum of this difference of all bulges along the pri-miRNA stem.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed in GraphPad Prism7 statistical software. p values were calculated using t test, Mann-Whitney U-test or Wilcoxon test, as indicated. p value < 0.05 was considered statistically significant.

DATA AND SOFTWARE AVAILABILITY

The Small RNA-Seq datasets generated in this study are available on NCBI GEO under the accession number GSE108893. Previously published datasets used in this study are summarized in Table S2. Other data and scripts are available upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Drs. Sandra L. Wolin and Donald L. Court for their critical reading of the manuscript and helpful discussions. For the SAXS experiments, we thank the SAXS Core facility of the National Cancer Institute. The SAXS data were collected at beamline under the PUP-24152 agreement with the Argonne National Laboratory. This work was funded by the intramural research program of the National Cancer Institute. This project was funded in part with federal funds from the Frederick National Laboratory for Cancer Research, NIH, under contract HHSN261200800001E, for W.K.K.

REFERENCES

Agarwal V, Bell GW, Nam J-W, and Bartel DP (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife 4.

Auyeung VC, Ulitsky I, McGeary SE, and Bartel DP (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell 152, 844–858. [PubMed: 23415231]

Balbous A, Cortes U, Guilloteau K, Villalva C, Flamant S, Gaillard A, Milin S, Wager M, Sorel N, Guilhot J, et al. (2014). A mesenchymal glioma stem cell profile is related to clinical outcome. Oncogenesis 3, e91. [PubMed: 24637491]

Bartel DP (2009). MicroRNAs: target recognition and regulatory functions. Cell 136, 215–233. [PubMed: 19167326]

Bartel DP (2018). Metazoan MicroRNAs. Cell 173, 20–51. [PubMed: 29570994]

Berezikov E (2011). Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet 12, 846–860. [PubMed: 22094948]

Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung J-H, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R, et al. (2011). Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome Res. 21, 203–215. [PubMed: 21177969]

Bofill-De Ros X, and Gu S (2016). Guidelines for the optimal design of miRNA-based shRNAs. Methods 103, 157–166. [PubMed: 27083402]

Bofill-De Ros X, Chen K, Chen S, Tesic N, Randjelovic D, Skundric N, Nesic S, Varjacic V, Williams EH, Malhotra R, et al. (2018). QuagmiR: a cloud-based application for isomiR big data analytics. Bioinformatics, Published online October 8, 2018. 10.1093/bioinformatics/bty843.

Bonev B, Stanley P, and Papalopulu N (2012). MicroRNA-9 modulates Hes1 ultradian oscillations by forming a double-negative feedback loop. Cell Rep. 2, 10–18. [PubMed: 22840391]

Burke JM, Kelenis DP, Kincaid RP, and Sullivan CS (2014). A central role for the primary microRNA stem in guiding the position and efficiency of Drosha processing of a viral pri-miRNA. RNA 20, 1068–1077. [PubMed: 24854622]

Burroughs AM, Ando Y, de Hoon MJL, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, and Daub CO (2010). A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res. 20, 1398–1410. [PubMed: 20719920]

Coolen M, Katz S, and Bally-Cuif L (2013). miR-9: a versatile regulator of neurogenesis. Front. Cell. Neurosci 7, 220. [PubMed: 24312010]

Court DL, Gan J, Liang Y-H, Shaw GX, Tropea JE, Costantino N, Waugh DS, and Ji X (2013). RNase III: genetics and function; structure and mechanism. Annu. Rev. Genet 47, 405–431. [PubMed: 24274754]

Dai L, Chen K, Youngren B, Kulina J, Yang A, Guo Z, Li J, Yu P, and Gu S (2016). Cytoplasmic Drosha activity generated by alternative splicing. Nucleic Acids Res. 44, 10454–10466. [PubMed: 27471035]

Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, and Hannon GJ (2004). Processing of primary microRNAs by the Microprocessor complex. Nature 432, 231–235. [PubMed: 15531879]

Fang W, and Bartel DP (2015). The menu of features that define primary microRNAs and enable de novo design of microRNA genes. Mol. Cell 60, 131–145. [PubMed: 26412306]

Fang X, Wang J, O'Carroll IP, Mitchell M, Zuo X, Wang Y, Yu P, Liu Y, Rausch JW, Dyba MA, et al. (2013). An unusual topological structure of the HIV-1 Rev response element. Cell 155, 594–605. [PubMed: 24243017]

Frank F, Sonenberg N, and Nagar B (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature 465, 818–822. [PubMed: 20505670]

Friedman RC, Farh KK-H, Burge CB, and Bartel DP (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105. [PubMed: 18955434]

Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, New-comb JM, Sempere LF, Flatmark K, Hovig E, and Peterson KJ (2015). A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. Annu. Rev. Genet 49, 213–242. [PubMed: 26473382]

Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, and Ji X (2006). Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. Cell 124, 355–366. [PubMed: 16439209]

Gregory RI, Yan K-P, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, and Shiekhattar R (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature 432, 235–240. [PubMed: 15531877]

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, and Mello CC (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that controlC. elegans developmental timing. Cell 106, 23–34. [PubMed: 11461699]

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, and Hofacker IL (2008). The Vienna RNA websuite. Nucleic Acids Res. 36, W70–W74. [PubMed: 18424795]

Gu S, Jin L, Zhang Y, Huang Y, Zhang F, Valdmanis PN, and Kay MA (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. Cell 151, 900–911. [PubMed: 23141545]

Ha M, and Kim VN (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol 15, 509–524. [PubMed: 25027649]

Han J, Lee Y, Yeom K-H, Kim Y-K, Jin H, and Kim VN (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 18, 3016–3027. [PubMed: 15574589]

Han J, Lee Y, Yeom K-H, Nam J-W, Heo I, Rhee J-K, Sohn SY, Cho Y, Zhang B-T, and Kim VN (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125, 887–901. [PubMed: 16751099]

Hutvágner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, and Zamore PD (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834–838. [PubMed: 11452083]

Iwakawa H-O, and Tomari Y (2015). The functions of micrornas: mRNA decay and translational repression. Trends Cell Biol. 25, 651–665. [PubMed: 26437588]

Jiang L, Shao C, Wu Q-J, Chen G, Zhou J, Yang B, Li H, Gou L-T, Zhang Y, Wang Y, et al. (2017). NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. Nat. Struct. Mol. Biol 24, 816–824. [PubMed: 28846091]

Jonas S, and Izaurralde E (2015). Towards a molecular understanding of microRNA-mediated gene silencing. Nat. Rev. Genet 16, 421–433. [PubMed: 26077373]

Khvorova A, Reynolds A, and Jayasena SD (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209–216. [PubMed: 14567918]

Kim B, Jeong K, and Kim VN (2017). Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. Mol. Cell 66, 258–269.e5. [PubMed: 28431232]

Knight SW, and Bass BL (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science 293, 2269–2271. [PubMed: 11486053]

Kozomara A, and Griffiths-Jones S (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–D73. [PubMed: 24275495]

Kuszewski J, Gronenborn AM, and Clore GM (1997). Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. J. Magn. Reson 125, 171–177. [PubMed: 9245376]

Kwon SC, Nguyen TA, Choi Y-G, Jo MH, Hohng S, Kim VN, and Woo J-S (2016). Structure of human DROSHA. Cell 164, 81–90. [PubMed: 26748718]

Lee HY, and Doudna JA (2012). TRBP alters human precursor microRNA processing in vitro. RNA 18, 2012–2019. [PubMed: 23006623]

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Rådmark O, Kim S, and Kim VN (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415–419. [PubMed: 14508493]

Lin S, and Gregory RI (2015). MicroRNA biogenesis pathways in cancer. Nat. Rev. Cancer 15, 321–333. [PubMed: 25998712]

Liu Y, Carson-Walter EB, Cooper A, Winans BN, Johnson MD, and Walter KA (2010). Vascular gene expression patterns are conserved in primary and metastatic brain tumors. J. Neurooncol 99, 13–24. [PubMed: 20063114]

Ma L, Young J, Prabhala H, Pan E, Mestdagh P, Muth D, Teruya-Feldstein J, Reinhardt F, Onder TT, Valastyan S, et al. (2010). miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. Nat. Cell Biol 12, 247–256. [PubMed: 20173740]

Ma H, Wu Y, Choi J-G, and Wu H (2013). Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. Proc. Natl. Acad. Sci. U S A 110, 20687–20692. [PubMed: 24297910]

MacRae IJ, Zhou K, and Doudna JA (2007). Structural determinants of RNA recognition and cleavage by Dicer. Nat. Struct. Mol. Biol 14, 934–940. [PubMed: 17873886]

Mathews DH, Moss WN, and Turner DH (2010). Folding and finding RNA secondary structure. Cold Spring Harb. Perspect. Biol 2, a003665. [PubMed: 20685845]

Michlewski G, Guil S, Semple CA, and Cáceres JF (2008). Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol. Cell 32, 383–393. [PubMed: 18995836]

Moore MJ, Scheel TKH, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM, and Darnell RB (2015). miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. Nat. Commun 6, 8864. [PubMed: 26602609]

Nguyen H, Pérez A, Bermeo S, and Simmerling C (2015a). Refinement of generalized born implicit solvation parameters for nucleic acids and their complexes with proteins. J. Chem. Theory Comput 11, 3714–3728. [PubMed: 26574454]

Nguyen TA, Jo MH, Choi Y-G, Park J, Kwon SC, Hohng S, Kim VN, and Woo J-S (2015b). Functional anatomy of the human microprocessor. Cell 161, 1374–1387. [PubMed: 26027739]

Nowek K, Wiemer EAC, and Jongen-Lavrencic M (2018). The versatile nature of miR-9/9* in human cancer. Oncotarget 9, 20838–20854. [PubMed: 29755694]

Oliveras-Ferraros C, Vazquez-Martin A, Cuyàs E, Corominas-Faja B, Rodríguez-Gallego E, Fernández-Arroyo S, Martin-Castillo B, Joven J, and Menendez JA (2014). Acquired resistance to metformin in breast cancer cells triggers transcriptome reprogramming toward a degradome-related metastatic stem-like profile. Cell Cycle 13, 1132–1144. [PubMed: 24553122]

Park J-E, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, and Kim VN (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature 475, 201–205. [PubMed: 21753850]

Pasquinelli AE (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nat. Rev. Genet 13, 271–282. [PubMed: 22411466]

Pietrzykowski AZ, Friesen RM, Martin GE, Puig SI, Nowak CL, Wynne PM, Siegelmann HT, and Treistman SN (2008). Posttranscriptional regulation of BK channel splice variant stability by miR-9 underlies neuroadaptation to alcohol. Neuron 59, 274–287. [PubMed: 18667155]

Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, and Adamiak RW (2012). Automated 3D structure composition for large RNAs. Nucleic Acids Res. 40, e112. [PubMed: 22539264]

Roden C, Gaillard J, Kanoria S, Rennie W, Barish S, Cheng J, Pan W, Liu J, Cotsapas C, Ding Y, and Lu J (2017). Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. Genome Res. 27, 374–384. [PubMed: 28087842]

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, and Lai EC (2007a). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 17, 1850–1864. [PubMed: 17989254]

Ruby JG, Jan CH, and Bartel DP (2007b). Intronic microRNA precursors that bypass Drosha processing. Nature 448, 83–86. [PubMed: 17589500]

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, and Zamore PD (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208. [PubMed: 14567917]

Schwieters CD, and Clore GM (2007). A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. Biochemistry 46, 1152–1166. [PubMed: 17260945]

Shibata M, Kurokawa D, Nakao H, Ohmura T, and Aizawa S (2008). MicroRNA-9 modulates Cajal-Retzius cell differentiation by suppressing Foxg1 expression in mouse medial pallium. J. Neurosci 28, 10415–10421. [PubMed: 18842901]

Shin J, Shim HG, Hwang T, Kim H, Kang S-H, Dho Y-S, Park S-H, Kim SJ, and Park C-K (2017). Restoration of miR-29b exerts anti-cancer effects on glioblastoma. Cancer Cell Int. 17, 104. [PubMed: 29176935]

Sim S-E, Lim C-S, Kim J-I, Seo D, Chun H, Yu N-K, Lee J, Kang SJ, Ko H-G, Choi J-H, et al. (2016). The brain-enriched microRNA miR-9-3p regulates synaptic plasticity and memory. J. Neurosci 36, 8641–8652. [PubMed: 27535911]

Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, Sobczak K, and Krzyzosiak WJ (2011). Structural basis of microRNA length variety. Nucleic Acids Res. 39, 257–268. [PubMed: 20739353]

Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, Robinson S, Zhang S, Ellis P, Langford CF, et al. (2014). 5′ isomiR variation is of functional and evolutionary importance. Nucleic Acids Res. 42, 9424–9435. [PubMed: 25056318]

Tsui V, and Case DA (2000–2001). Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 56, 275–291.

Xie M, Li M, Vilborg A, Lee N, Shu M-D, Yartseva V, Šestan N, and Steitz JA (2013). Mammalian 5′-capped microRNA precursors that generate a single microRNA. Cell 155, 1568–1580. [PubMed: 24360278]

Yan J, Kong L-Y, Hu J, Gabrusiewicz K, Dibra D, Xia X, Heimberger AB, and Li S (2015). FGL2 as a multimodality regulator of tumor-mediated immune suppression and therapeutic target in gliomas. J. Natl. Cancer Inst 107, djv137. [PubMed: 25971300]

Zamudio JR, Kelly TJ, and Sharp PA (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. Cell 156, 920–934. [PubMed: 24581493]

Zeng Y, Yi R, and Cullen BR (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. EMBO J. 24, 138–148. [PubMed: 15565168]

Zgarbová M, Otyepka M, Sponer J, Mládek A, Banáš P, Cheatham TE, 3rd, and Jurecka P (2011). Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theory Comput 7, 2886–2902. [PubMed: 21921995]

Zhao C, Sun G, Li S, and Shi Y (2009). A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. Nat. Struct. Mol. Biol 16, 365–371. [PubMed: 19330006]

Highlights

- Pri-miRNA paralogs can adopt new functions due to distinctDrosha processing
- A pri-miR-9-1-derived isomiR regulates a novel set of targets in low-grade glioma
- The 3D structure of pri-miR-9-1 triggers Drosha cleavage at an alternative position
- Structure-induced alternative Drosha cleavage is a general way to produce isomiRs

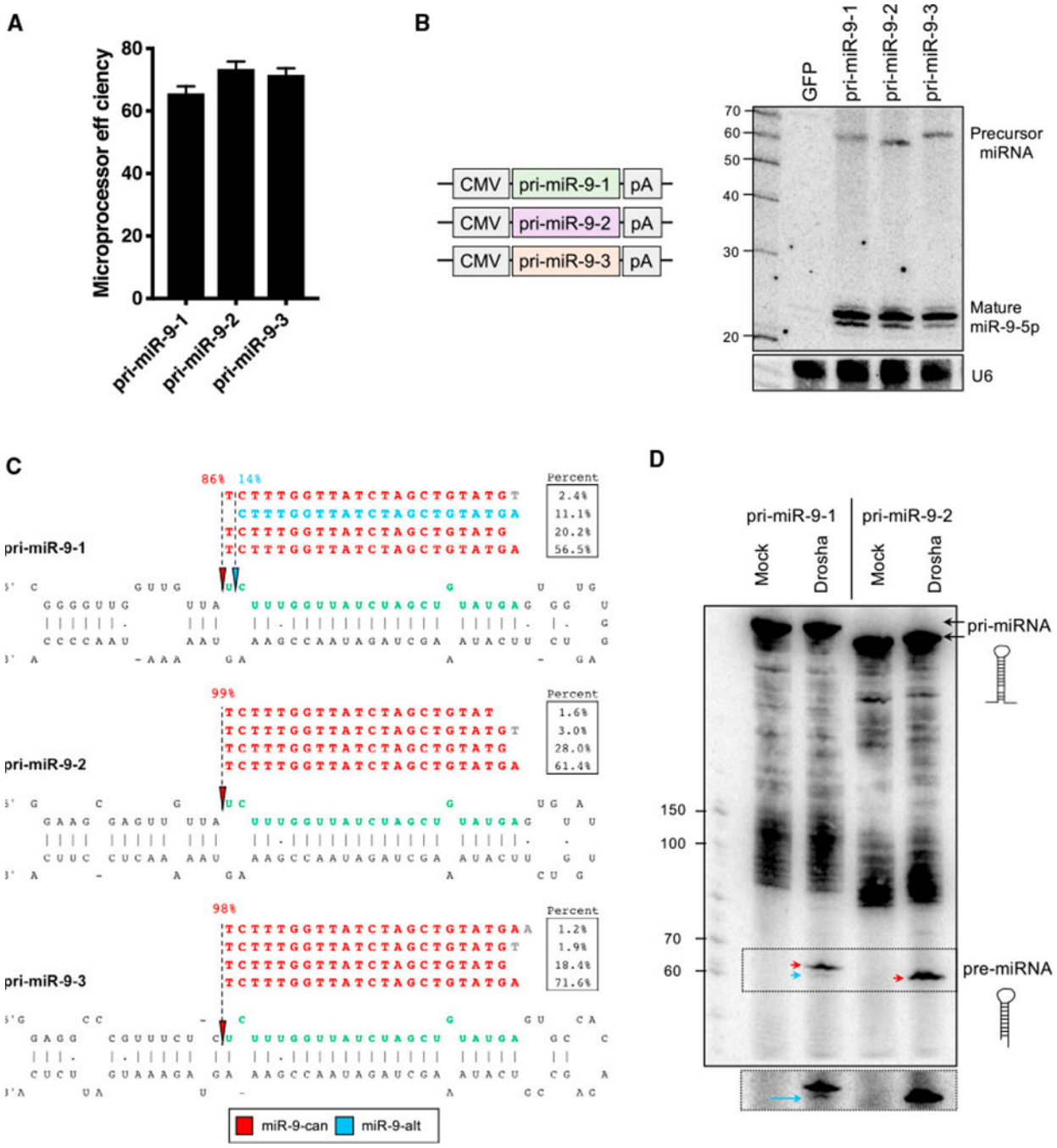


Figure 1. Pri-miR-9-1 Has a Unique Drosha Cleavage Profile

(A) *Renilla* luciferase reporters containing pri-miR-9 family members in the 3'UTR were transfected into HEK293T Drosha knockout (KO) cells with or without the co-expression of Drosha. Dual-luciferase assays were performed 48 hr post-transfection. Error bars represent SEM of three biological replicates.

(B) CMV-driven pri-miR-9 cassettes were expressed individually in HEK293T cells. Pre-miR-9 and mature miR-9 isoforms were detected by northern blotting with probe-miR-9. U6

snRNA was detected as an internal control. Note that pre-miR-9-2 (59 nt) is shorter than pre-miR-9-1 and pre-miR-9-3 (both 61 nt) because of differences in the loop sequence.
(C) Small RNAs from HEK293T cells transfected with pri-miR-9-expressing plasmids were subjected to deep sequencing. After being mapped to the corresponding pri-miR-9 paralog, only the four most abundant sequences were labeled in the figure along with their percent abundance relative to all pri-miR-9 reads. The percentage of sequences starting at a position relative to the total number of miR-9 reads was used to infer the Drosha cleavage percentage, which was labeled with a small solid arrow and a dotted line. miR-9-can and its tailed or trimmed isomiRs are in red; miR-9-alt is in blue.
(D) *In vitro* transcribed pri-miR-9-1 or pri-miR-9-2 transcripts were incubated with mock or purified Microprocessor complex. Precursor miRNAs (pre-miR-9-1 and pre-miR-9-2) were detected by northern blot (red). A shorter product (blue) was observed only with pre-miR-9-1 but not with pre-miR-9-2. A longer exposure for a better visualization of the pre-miRNAs is also presented.
See Figure S1.

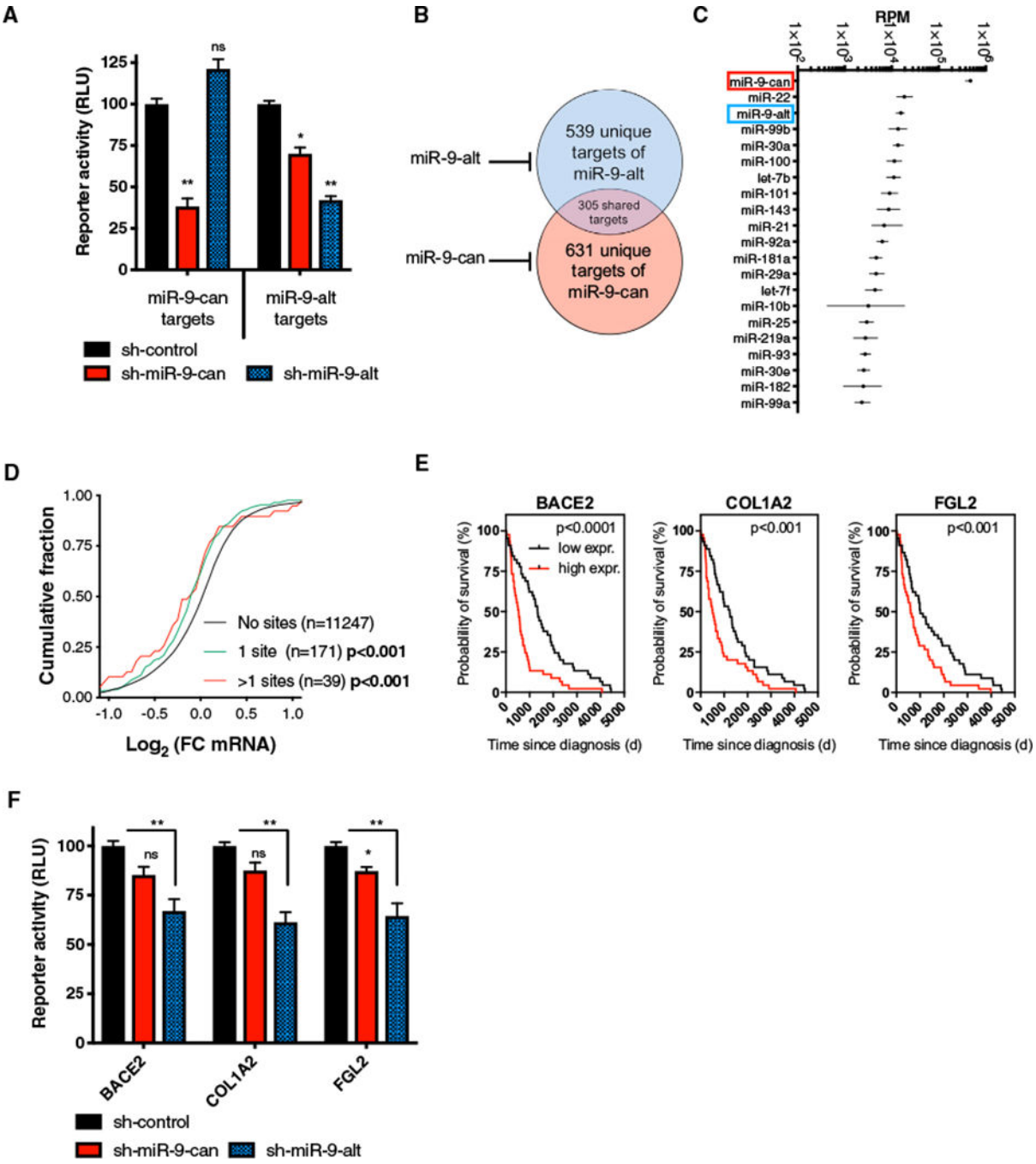


Figure 2. miR-9-alt Regulates a Novel Set of Targets in Low-Grade Glioma

(A) Repression of miR-9 isomiRs was measured using dual-luciferase reporter assay. The psiCHECK2 vector with two tandem target sites in the 3'UTR and DNA plasmids expressing shRNA were co-transfected into HEK293 cells. These target sites were designed with a 7-mer-A1 seed region plus additional base pairs on the 3' end of the miRNA. Dual-luciferase assays were performed 48 hr post-transfection. *Renilla* luciferase activities were normalized, and the percentage of relative enzyme activity compared with the negative

control (sh-control) was plotted. Error bars represent the SEM from three biological replicates (t test with multiple test corrections).

(B) Venn diagram of predicted targets of miR-9-can and miR-9-alt by TargetScan.

(C) Average expression level (RPM) of miR-9-can and miR-9-alt among top expressed miRNAs in LGG tumors. Error bars represent the SD over 525 LGG patient samples documented in TCGA.

(D) Cumulative fraction plot of fold-change in expression of mRNAs between the top and low levels of the miR-9-alt in patients from LGG. Targets were classified according to the number of target sites as one target site (conserved 8-mer seed) or more than one target site (conserved 8-mer, 7-mer-A1, or 7-mer-m8). Shift toward the left indicates repression (one-way ANOVA with multiple test corrections).

(E) Kaplan-Meier survival curve of LGG patients with high (red line) and low (black line) levels of the target gene.

(F) Predicted miR-9-alt target sites from each gene were cloned individually in the 3'UTR of the luciferase reporter. Dual-luciferase assays were performed at 48 hr post-transfection as described previously. Error bars represent the SEM from three biological replicates. *$p < 0.05$ and **$p < 0.01$; n.s., non-significant (t test with multiple test corrections).

See Figure S2.

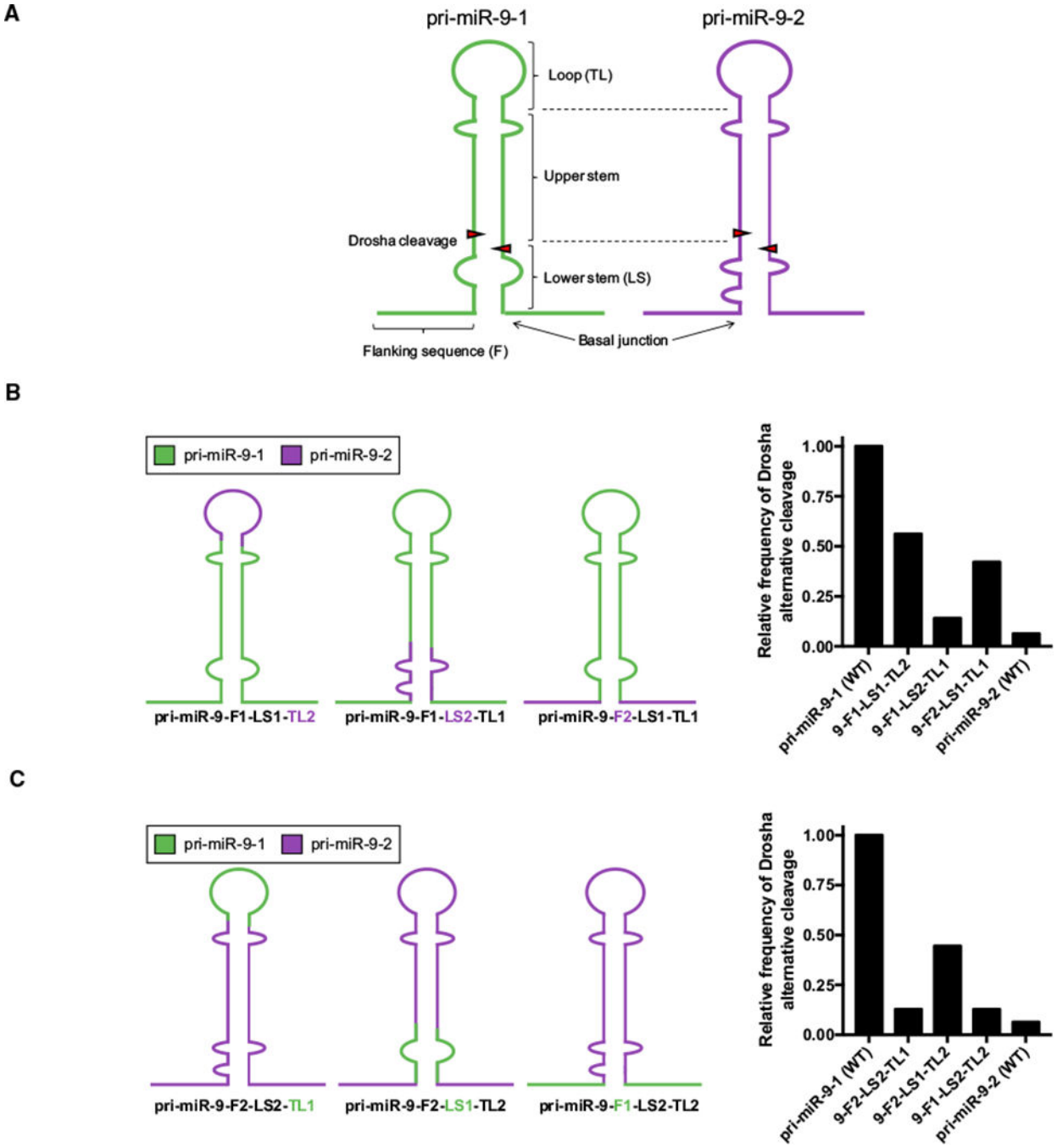


Figure 3. The Lower Stem Plays a Major Role in Determining Cleavage Fidelity

(A) Schematic representation of pri-miRNA structural elements.

(B) Schematic representation of chimeric constructs where structural elements of pri-miR-9-1 have been replaced for their equivalents in pri-miR-9-2. Note that flanking sequences (F), lower stem (LS), and terminal loop (TL) from pri-miR-9-1 are depicted in green, while these from pri-miR-9-2 are in purple. All chimeras were expressed in HEK293T cells. Relative Drosha alternative cleavage frequencies on each chimeric construct were measured by deep sequencing and plotted here.

(C) Schematic representation of chimeric constructs where structural elements of pri-miR-9-2 have been substituted for their equivalents in pri-miR-9-1. Relative Drosha alternative cleavage frequencies on each chimeric construct in HEK293T cells.
Note that in both (B) and (C), Drosha alternative cleavage frequency on wild-type pri-miR-9-1 (14%) was set as 1 in the figure.
See Figure S3.

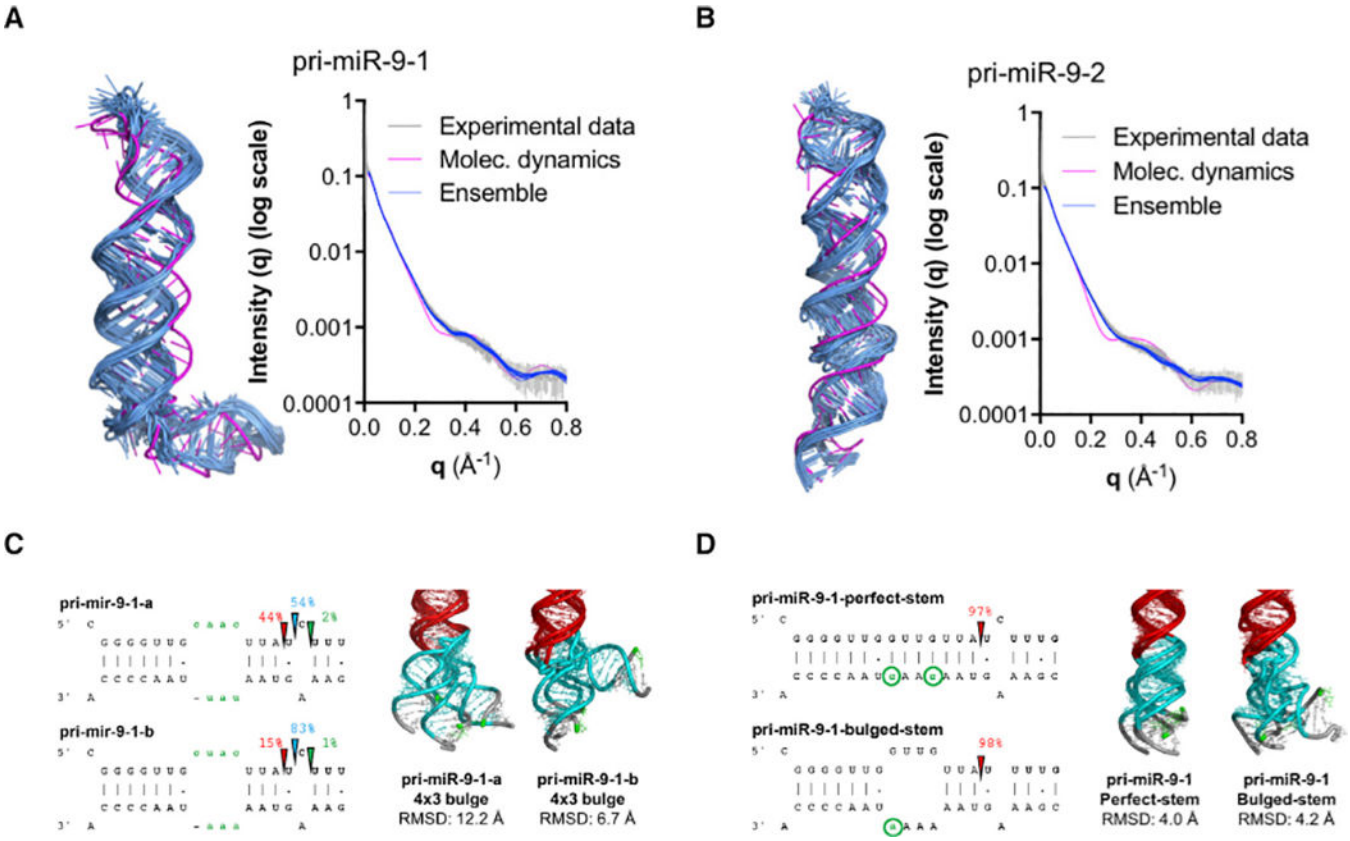


Figure 4. The Tertiary Structure of the Lower Stem Drives Alternative Drosha Cleavage

(A and B) Three-dimensional structures of pri-miR-9-1 (A) and pri-miR-9-2 (B) obtained after molecular dynamics refinement (magenta) and top ten ensemble structures derived from the SAXS data (blue). Experimental SAXS data (gray) plotted as scattering intensity versus momentum transfer q (Å $^{-1}$). The plot also displays the calculated SAXS curves of the molecular dynamics structure (magenta) and the intensity (q) values obtained in the top ten ensemble structures (blue). (C and D) Secondary structure of the lower stem of pri-miR-9-1 mutants pri-mir-9-1-a and pri-mir-9-1-b (C) as well as pri-miR-9-1-perfect-stem and pri-miR-9-1-bulged-stem (D). Green labeling illustrates the nucleotides changed in each case. The arrows and the corresponding numbers indicate the inferred Drosha cleavage sites and their relative percentage. Local tertiary structure motions of pri-miR-9-1 mutants: lower stem (cyan); upper stem (red). The distortion is evaluated by RMSD during molecular dynamics against the backbone of a perfect RNA A-form helix.

See Figure S4.

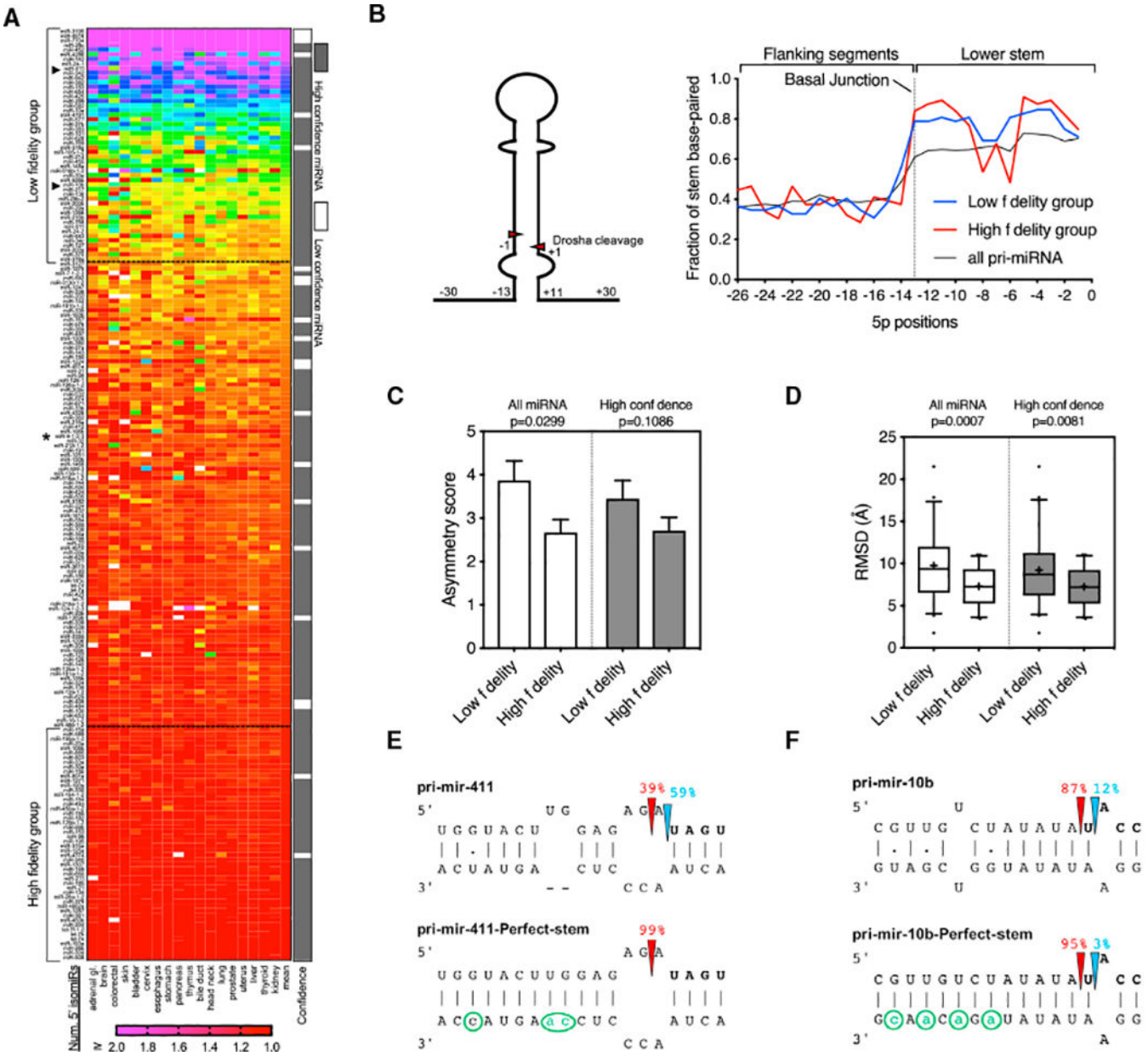


Figure 5. Pri-miRNA Tertiary Structure-Based Alternative Drosha Cleavage Is a General Mechanism for isomiR Production

(A) Heatmap of the weighted average number of 5' isomiRs generated from the top 200 5p miRNAs expressed in various tissues (TCGA normal tissue samples). High-confidence miRNAs were annotated on the basis of a previous report (Fromm et al., 2015). White box in heatmap indicates data unavailable.

(B) Plot shows the fraction of pri-miRNAs that present a stem base-pairing on each position relative the Drosha cleavage site. Positions are numbered on the basis of their distances to the Drosha cleavage site, upstream of the 5p cleavage site. Black solid line shows the stem base-pairing of all the miRNA reported on miRBase version 21.

(C) Plot of the average asymmetry score of pri-miRNAs with high (top 50) or low (bottom 50) Drosha cleavage fidelity (Mann-Whitney U test).

(D) Boxplot of the RMSD of pri-miRNAs with high (top 50) or low (bottom 50) Drosha cleavage fidelity (Mann-Whitney U test).
(E and F) Secondary structure of the lower stem of pri-miR-411 (E), pri-miR-10b (F), and their corresponding mutants. Green denotes the nucleotides changed in each case. The arrows and the corresponding numbers indicate the inferred Drosha cleavage sites and their relative percentage.
See Figure S5.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: 293T cells	ATCC	Cat# CRL-3216
Human: HeLa cells	ATCC	Cat# CCL-2
Human: 293T Drosha-KO cells	(Dai etal., 2016)	N/A
Chemicals, Peptides, and Recombinant Proteins		
DMEM High glucose	GIBCO	Cat# 11965092
Penicillin-Streptomycin (10,000 U/mL)	GIBCO	Cat# 15140163
HyClone Fetal Bovine Serum	GE Healthcare	Cat# SH30070
PolyJet DNA *In Vitro* Transfection Reagent	SignaGen	Cat# SL100688
PerfectHyb Plus Hybridization Buffer	Sigma	Cat# H7033–1L
Dual-Luciferase® Reporter Assay System	Promega	Cat# E1980
T4 RNA Ligase 2, Truncated	NEB	Cat# M0242S
T4 RNA Ligase 1, High Concentration	NEB	Cat# M0437M
RNase inhibitor, Murine	NEB	Cat# M0314L
SuperScript ® IV Reverse Transcriptase	ThermoFisher	Cat# 18090010
Phusion® High-Fidelity DNA polymerase	NEB	Cat# M0530S
Qubit dsDNA HS Assay Kit	ThermoFisher	Cat# Q32854
Agilent High Sensitivity DNA kit	Agilent	Cat# 5067–4626
MiSeq Reagent Kit v3 (150-cycle)	Illumina	Cat# MS-102–3001
MEGAscript T7 Transcription Kit	ThemoFisher	Cat# AMB13345
Anti-FLAG® M2 Magnetic Beads	Sigma-Aldrich	Cat# M8823; RRID:AB_2637089
Plasmids		
Drosha expression plasmid	(Dai etal., 2016)	N/A
psiCHECK-2 Vector	Promega	Cat# C802A
Deposited Data		
FASTQ files	This paper	GEO: GSE108893