Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation

Christine Roden,[1,2,3,4,9] Jonathan Gaillard,[2,5,9] Shaveta Kanoria,[6] William Rennie,[6] Syndi Barish,[4] Jijun Cheng,[1,2,3] Wen Pan,[1,2,3] Jun Liu,[1,2,3] Chris Cotsapas,[1,7] Ye Ding,[6] and Jun Lu[1,2,3,8]

[1]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA; [2]Yale Stem Cell Center and Yale Cancer Center, Yale University, New Haven, Connecticut 06520, USA; [3]Yale Center for RNA Science and Medicine, New Haven, Connecticut 06520, USA; [4]Graduate Program in Biological and Biomedical Sciences, Yale University, New Haven, Connecticut 06510, USA; [5]School of Medicine, Yale University, New Haven, Connecticut 06510, USA; [6]Wadsworth Center, New York State Department of Health, Albany, New York 12208, USA; [7]Department of Neurology, Yale School of Medicine, New Haven, Connecticut 06511, USA; [8]Yale Cooperative Center of Excellence in Hematology, Yale University, New Haven, Connecticut 06520, USA

Mature microRNAs (miRNAs) are processed from hairpin-containing primary miRNAs (pri-miRNAs). However, rules that distinguish pri-miRNAs from other hairpin-containing transcripts in the genome are incompletely understood. By developing a computational pipeline to systematically evaluate 30 structural and sequence features of mammalian RNA hairpins, we report several new rules that are preferentially utilized in miRNA hairpins and govern efficient pri-miRNA processing. We propose that a hairpin stem length of 36 ± 3 nt is optimal for pri-miRNA processing. We identify two bulge-depleted regions on the miRNA stem, located ~16–21 nt and ~28–32 nt from the base of the stem, that are less tolerant of unpaired bases. We further show that the CNNC primary sequence motif selectively enhances the processing of optimal-length hairpins. We predict that a small but significant fraction of human single-nucleotide polymorphisms (SNPs) alter pri-miRNA processing, and confirm several predictions experimentally including a disease-causing mutation. Our study enhances the rules governing mammalian pri-miRNA processing and suggests a diverse impact of human genetic variation on miRNA biogenesis.

[Supplemental material is available for this article.]

Mature miRNAs are derived from hairpin-containing primary transcripts (pri-miRNAs). For the vast majority of miRNAs, pri-miRNAs are processed in the nucleus by the Microprocessor complex, containing DROSHA, DGCR8, and other proteins (Lee et al. 2003; Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Nguyen et al. 2015; Kwon et al. 2016), into precursor miRNAs (pre-miRNAs), which can be further processed by DICER1 into mature miRNAs (Bernstein et al. 2001). Accurate levels of miRNA expression are required for normal cell functions, the disruption of which is frequently observed in diseased states such as cancer (Lu et al. 2005; Farazi et al. 2011; Iorio and Croce 2012).

Many of the mammalian coding and noncoding transcripts contain predicted hairpin structures. While DROSHA/DGCR8 cleaves a small fraction of non-miRNA transcripts (Karginov et al. 2010; Macias et al. 2012), the vast majority of hairpin-containing transcripts do not enter the pri-miRNA processing pathway. How are pri-miRNAs distinguished from hairpin-containing transcripts in the genome? Previous studies, mostly by modeling specific miRNAs and their mutants, have examined rules governing efficient pri-miRNA processing. For example, several studies have proposed that the Microprocessor complex efficiently processes hairpins with stem lengths of ~33 nucleotides (nt) (Han et al. 2006; Nguyen et al. 2015). Another recent study has used a cell-free screen on more than 210,000 random mutations in three pri-miRNA hairpins and found that the most favored substrates of the Microprocessor complex are hairpins of $\sim 35 \pm 1$ nt in stem length (Fang and Bartel 2015). Apical loop sizes of 3–23 nt have been found to be compatible with processing (Zeng and Cullen 2003), with loop size of ≥ 10 nt proposed to allow efficient processing (Ma et al. 2013). In addition, a bulge was often found near the DROSHA processing site several bases from the base of the hairpin (Han et al. 2006) and, in the context of a GHG motif, can enhance the processing of specific miRNAs (Fang and Bartel 2015).

In addition to structural features, several primary sequence motifs have been identified to regulate pri-miRNA processing.

374 Genome Research
www.genome.org
27:374–384 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/17; www.genome.org

One of the processing-enhancing motifs is the CNNC motif, which can be found ~17–18 nt downstream from 3p mature miRNA in a subset of pri-miRNAs (Auyeung et al. 2013). Both SRSF3 and DDX17 have been reported to bind specific CNNC sites and affect processing (Auyeung et al. 2013; Mori et al. 2014). However, it is unclear whether the functional CNNC motif is indeed degenerate and restrained spatially. Other than the CNNC motif, the UG motif at the basal stem junction and UGU/GUG/UGUG motif in the apical loop can be found in a minor subset of miRNAs (Auyeung et al. 2013; Nguyen et al. 2015). The relationship between structural features and primary sequence motifs has been examined in a recent study, which proposes that a combination of primary sequence motifs (CNNC, basal UG, and apical UGU/GUG/UGUG) enhances the processing of hairpins with nonoptimal lengths (Fang and Bartel 2015). Aside from structural and sequence features in pri-miRNAs, a number of protein factors have been shown to regulate processing of specific miRNAs or miRNA subsets (Denli et al. 2004; Gregory et al. 2004; Yang et al. 2006; Guil and Caceres 2007; Davis et al. 2008, 2010; Newman et al. 2008; Viswanathan et al. 2008; Paroo et al. 2009; Trabucchi et al. 2009; Nam et al. 2011; Piskounova et al. 2011; Tang et al. 2011; Kawahara and Mieda-Sato 2012; Wada et al. 2012; Di Carlo et al. 2013; Cheng et al. 2014). More recently, the role of m6A has been demonstrated to regulate pri-miRNA processing (Alarcon et al. 2015a,b).

The lack of systematic comparisons between pri-miRNA hairpins and other genomic hairpin-containing transcripts raises the possibility that comparative analysis of these two classes of hairpins can both identify additional rules governing pri-miRNA processing and reveal key distinctions between miRNA and non-miRNA hairpins. In this study, based on a new computational pipeline to characterize hairpin features, and with experimental validation, we propose additional and modified rules for efficient pri-miRNA processing, which are preferentially utilized in miRNAs over non-miRNA hairpins. These rules can be further confirmed by human SNPs that lead to the disruption of favorable processing-related features.

Results

Systematic comparison of hairpin features for mammalian pri-miRNAs versus genomic hairpin-containing transcripts

We sought to identify structural and sequence features that distinguish hairpins in pri-miRNAs from those residing in other genomic transcripts. We designed a computational pipeline, "HairpIndex,"

to identify hairpins from computationally folded RNA structures and to annotate such hairpins with a total of 30 structural and sequence features (see Methods) (Fig. 1A). These features can be roughly categorized as secondary structural features (stem length, hairpin pairing, bulge size/position, and apical loop size) and primary sequence (CNNC, basal UG, and apical GUG/UGUG) motifs (Fig. 1B).

To identify potential differences in miRNA hairpins versus non-miRNA hairpins, we resorted to three sequence data sets, which were obtained for both human and mouse. We first created a control data set for non-miRNA transcripts by using RefSeq sequences (see Methods). Second, we used pre-miRNA from miRBase (miRBase v21) with 30 bases of flanking sequence on each side. Third, because miRBase contains a number of miRNAs whose validity has been questioned (e.g., *MIR4521*) (Guo et al. 2015), we assigned a subset of miRNAs enriched for well-processed
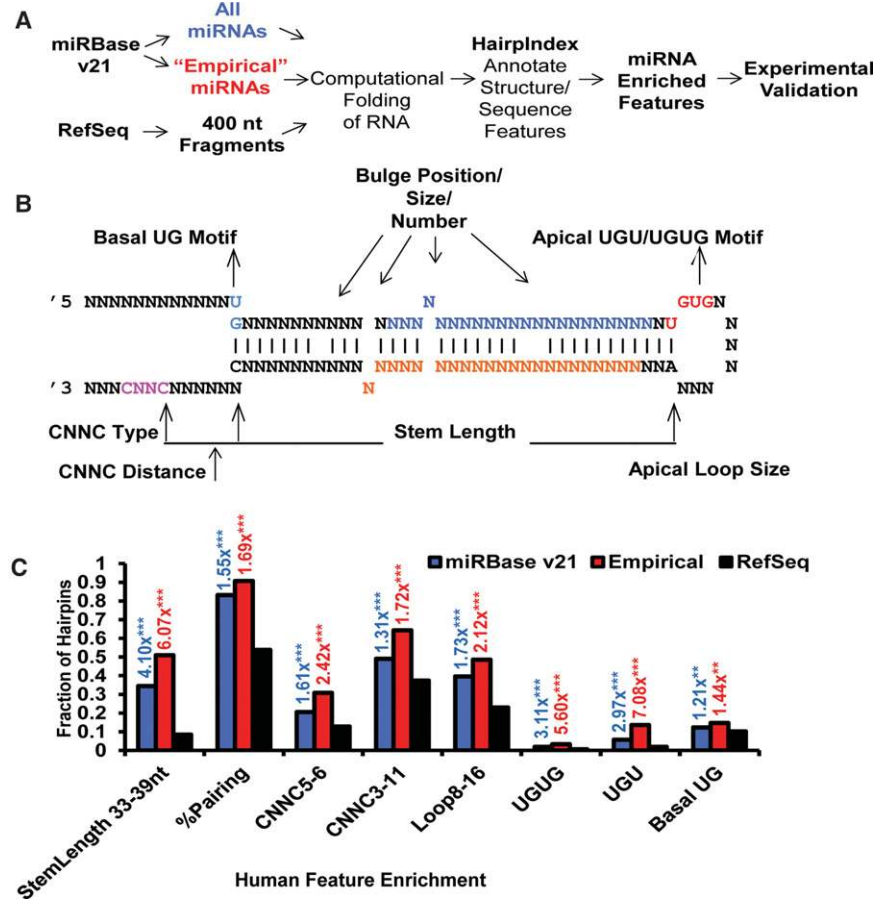
Figure 1. Systematic evaluation of miRNA and non-miRNA hairpins reveal enriched structural and sequence features in pri-miRNAs. (*A*) A schematic of the computational and experimental workflow. (*B*) Diagram of major hairpin features annotated by the HairpIndex pipeline. For pri-miRNA hairpins, blue letters represent 5p mature miRNA; orange letters, 3p mature miRNA. Primary sequence motifs—including CNNC, basal UG, and apical UGU/UGUG—were also highlighted. (*C*) The fractions of hairpins containing each indicated feature were plotted for human pri-miRNAs in miRBase v21, in a subset of "Empirical" miRNAs, or in human RefSeq fragments. The numbers *above* the bars indicate the fold enrichment of the corresponding feature over the RefSeq control. The numbers for stem length indicate the range of stem lengths in nucleotides; % pairing indicates the percentage of paired bases within the stem; the numbers after CNNC indicate the position range of putative CNNC motifs relative to the base of the hairpin, as measured in nucleotides; the numbers after loop indicate the range of the loop sizes; and basal UG, apical UGU, or apical UGUG motifs indicate the presence or absence of such features in hairpins. *P*-values were calculated using Fisher's exact test. (**) $P < 0.005$; (***) $P < 0.0005$.

miRNAs as an "Empirical" set (see Methods). All sequence data sets were then computationally folded and analyzed with HairpIndex (Supplemental Tables S1, S2).

We started by examining previously documented features relevant to miRNA processing, including stem length, apical loop size, the presence of CNNC, apical UGU or UGUG, and basal UG motifs. We also included percentage pairing within the hairpin stem. For some of the features, we assigned optimal enrichment parameters (e.g., optimal stem lengths of 33–39 nt). The choice of optimal feature definitions for stem lengths, stem pairing, and CNNC motif will be discussed in much more detail in the following sections. For apical loop size, we observed that the enriched range is 8–16 nt; apical loop size from 2–7 nt was prevalent in miRNA hairpins but not enriched compared with non-miRNA hairpins (Supplemental Fig. S1A,B).

Overall, the eight measured features above were significantly enriched in miRNA hairpins over RefSeq hairpins (Fig. 1C; Supplemental Fig. S1C). Similar enrichment results could be observed when we used refined sets (v2) of Empirical miRNAs (Supplemental Tables S3, S4), removed a small number of miRNA-containing RefSeq sequences, filtered the RefSeq control data set to result in similar folding energy as the miRNA data set, and/or used different folding algorithms (Supplemental Fig. S1D,E), indicating that the results were insensitive to variations in analysis parameters. Among these enriched features, we observed that the optimal range of stem length is the most distinguishing feature, with a greater than sixfold enrichment for the Empirical miRNA set versus RefSeq and with ~50% of Empirical miRNAs having optimal stem lengths (Fig. 1C; Supplemental Fig. S1C). In contrast, other features were enriched at much lower levels, were present in a small fraction of miRNAs, or both. Taken together, systematic hairpin comparisons between miRNAs and RefSeq revealed a number of enriched features, with optimal stem lengths as a key distinguishing feature.

Optimal stem lengths are required for efficient miRNA processing in cells

We examined stem length by plotting the distribution of hairpins with regard to stem lengths. Bulges were included in stem length calculations (see Methods). Unlike RefSeq hairpins, the distribution

of miRNA hairpins had a peak at 36 nt, which decreases with either increasing or decreasing lengths (Fig. 2A; Supplemental Fig. S2A). As expected, the Empirical miRNA set showed a stronger peak than the miRBase v21 set. Based on these distributions and
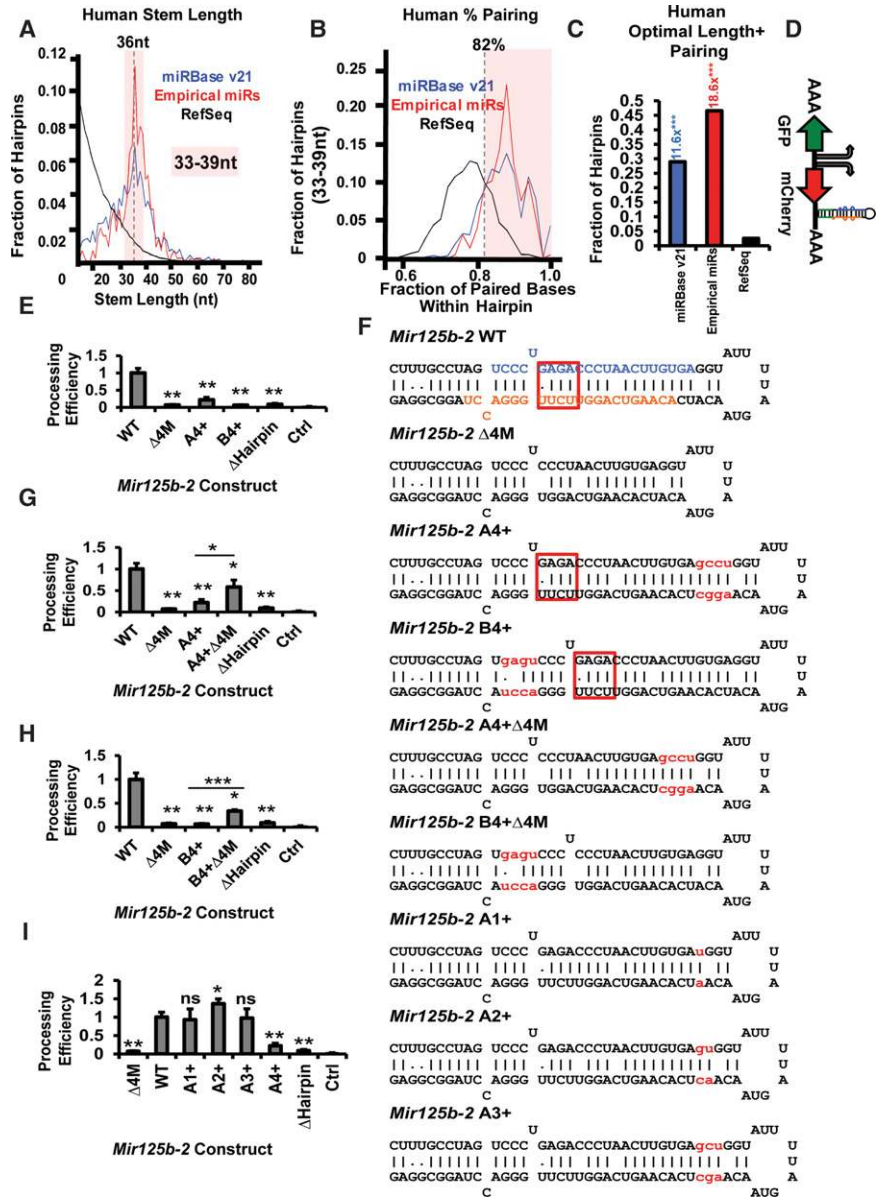
Figure 2. Optimal stem length is required for efficient processing. (*A*) The distributions of the stem lengths of human hairpins were plotted for pri-miRNAs in miRBase, Empirical pri-miRNAs, and RefSeq hairpins. The optimal range of 36 ± 3 nt is highlighted. (*B*) The distributions of percentage pairing within hairpin stem were plotted for human hairpins. A cutoff of 82% pairing was highlighted. (*C*) The fractions of human hairpins with both optimal stem length (33–39 nt) and pairing (≥82%) were plotted, with fold enrichment over RefSeq indicated. (*D*) Diagram of the lentiviral pri-miRNA processing reporter. (*E–I*) The processing of wild-type (WT) and mutant mouse *Mir125b-2* was measured in BaF3 cells. Data were normalized with the processing level for WT mouse *Mir125b-2* construct set to one and the level of an empty vector (Ctrl) set to zero. A construct removing the hairpin in mouse *Mir125b-2* (ΔHairpin) was also used as a control. *N* = 3. *P*-values are annotated *above* the bars for comparison with WT construct. Other *P*-value comparisons are indicated with horizontal bars. (*F*) Sequences and predicted structures of constructs tested in *E* and *G* through *I*. Color-coded elements include 5p mature miRNA (blue), 3p miRNA (orange), insertions (red letters; lowercase), and deletions (red box) that occurred in other related constructs. Watson-Crick pairings are indicated with vertical bars, whereas G:U pairings are indicated with dots. (*) *P* < 0.05; (**) *P* < 0.01; (***) *P* < 0.001; (ns) not significant.

experiments below and in the following sections, we assigned optimal enriched lengths to be 33–39 nt.

MiRNA hairpins within optimal length range were overall better paired than those from RefSeq, and based on their distributions (Fig. 2B; Supplemental Fig. S2B), we assigned a cutoff of 82%. Combining both optimal stem lengths and optimal percentage pairing within stem, we observed an 18.6-fold enrichment for human Empirical miRNAs in comparison to RefSeq, with 46.6% of Empirical miRNAs having both optimal features (Fig. 2C; Supplemental Fig. S2C for mouse). Furthermore, when we restricted the enrichment analysis to hairpins with both optimal lengths and pairing, we observed improvements in enrichment for other features, including apical loop size, CNNC 5-6, basal UG, and apical UGU (Supplemental Fig. S2D). These data are consistent with optimal stem length/pairing as important determinants of miRNA processing.

The peak of 36 nt for miRNA stem length, and the optimal range of 33–39 nt, predicts that efficient miRNA processing would be impaired when miRNAs are too long or too short. It also predicts that there would be several bases of tolerance for stem lengths. To test these possibilities, we resorted to a widely used in vivo reporter vector for pri-miRNA processing (Fig. 2D; Mori et al. 2014; Weitz et al. 2014). Specifically, miRNA hairpins or their mutants were cloned into the 3′ UTR of mCherry in a dual GFP/mCherry vector so that pri-miRNA processing will destabilize mCherry mRNA, with more efficient processing leading to a higher ratio of GFP/mCherry signals by flow cytometry (see example in Supplemental Fig. S2E). Knocking down DROSHA or DGCR8, but not DICER1, profoundly reduced processing levels measured by this reporter system (Supplemental Fig. S2F,G). In addition, we observed that the reporter activities were inversely correlated with pri-miRNA levels and positively correlated with mature miRNA levels (Supplemental Figs. S2H, S4J). For both human and mouse constructs, similar reporter activities were observed in human and mouse cell lines (HL60 and BaF3) and across two mouse cell lines with distinct tissue origins (BaF3 and NIH/3T3) (Supplemental Fig. S2I). These data support that the processing reporter assay faithfully reflects pri-miRNA processing efficiency.

We used mouse Mir125b-2 (stem length of 35 nt) as a model to test the stem length requirements. Of note, we confirmed that within the predicted folding structure of this miRNA, the last two pairing bases in the stem near the apical loop are functionally important, the unpairing of which reduced processing efficiency (Supplemental Fig. S2J,K). We next created a number of mutants in which the stem lengths were increased or decreased by four bases. Specifically, we inserted four pairs of bases near the apical loop (A4+, 39 nt) or near the basal side (B4+, 39 nt) with B4+ also introducing an extra bulge and decreased percentage pairing. In addition, we removed four pairs of bases in the middle of the hairpin (Δ4M, 31 nt). In each case, four-base alterations strongly reduced processing efficiency (Fig. 2E,F), with the 39-nt A4+ (without extra bulge) being the least affected and retaining ~22% processing efficiency. To confirm that the mutant hairpins (A4+, B4+, Δ4M) reduced processing in a stem length–dependent manner, rather than being solely due to primary sequence alterations, we performed a rescue experiment by combining A4+ and B4+ mutations with the Δ4M mutation (Fig. 2F). Combination mutations (A4+Δ4M and B4+Δ4M) rescued stem length and largely rescued the processing efficiency (Fig. 2G,H). To examine the tolerance of stem length changes, we created additional mutants that extended stem length by one, two, or three bases (A1+, 36 nt; A2+, 37 nt; A3+, 38 nt). Increasing stem length up to 3 nt did not result in significant impairment of processing (Fig. 2I). Taken together, our data above support the existence of a range of optimal stem lengths that are required for efficient pri-miRNA processing. In addition, our data also suggest that stem length by itself is insufficient to account for all requirements for processing, because our A4+Δ4M and B4+Δ4M combination mutants only partially rescued the processing defects.

Bulge positions control hairpin processing

We reasoned the location of bulges might affect processing efficiency. Among hairpins with optimal lengths and pairing, we observed that RefSeq hairpins showed relatively uniform distribution of bulge positions along the stem (Fig. 3A; Supplemental Fig. S3A). In contrast, we observed enrichment of bulges at ~5–9 nt from the base of the stem in miRNAs, which was more prominent when counting from the base of the hairpin (Supplemental Fig. S3B,C). This is consistent with earlier observations from the Kim group and recent findings by the Bartel group (Han et al. 2006; Fang and Bartel 2015). Unexpectedly, we also observed two bulge-depleted regions in miRNA hairpins, located at positions ~5–9 nt and ~16–21 nt relative to apical loop, or ~16–21 nt and ~28–32 nt relative to the base of the hairpin, observable for both arms of the stem (Fig. 3A,B). In contrast to bulge location, the numbers of bulges in miRNA hairpins and non-miRNA hairpins were only mildly different (Supplemental Fig. S3D,E). Thus, bulges are differentially distributed along miRNA hairpins in comparison to RefSeq hairpins.

We then asked whether bulges in bulge-depleted regions are less tolerated during miRNA processing. We first examined the partially rescued mouse Mir125b-2 B4+Δ4M construct, in which the stem length rescue strategy shifted the bulge locations and resulted in an enlarged bulge in one of the bulge-depleted regions compared with the WT construct (19/20 nt from the stem base) (Fig. 3C). Flattening this bulge (B4+Δ4M FB2) significantly enhanced processing relative to B4+Δ4M (Fig. 3D), whereas flattening a bulge outside of the bulge-depleted region (B4+Δ4M FB1, 13/14 nt from stem base) (Fig. 3C) did not significantly improve processing (Fig. 3D). These data suggest that the positions of bulges contribute to pri-miRNA processing.

We further reasoned that stronger processing differences could be revealed if more profound changes in bulges were engineered. To test this, we used human MIR16-1 as a model. We applied a similar strategy of stem length rescue by introducing both four pairs of insertion (MIR16-1 A4+) and four pairs of deletion (MIR16-1 Δ4M). Our design of the combination mutant (MIR16-1 A4+ Δ4M) rescued the stem length but moved a bulge into the center of one of the bulge-depleted regions (19 nt from stem base) (Fig. 3E). As we expected, MIR16-1 A4+ and Δ4M each resulted in strong reductions in processing, and the combination mutant MIR16-1 A4+Δ4M failed to rescue the processing (Fig. 3F). To determine whether this failure is due to the bulge in the bulge-depleted region, we removed the bulge-base in the MIR16-1 A4+Δ4M construct (MIR16-1 FBD), which led to a complete rescue of processing efficiency to the WT level (Fig. 3F). In contrast, eliminating a bulge in the bulge-tolerated region (MIR16-1 FBE; 12 nt from base of stem) did not rescue processing (Fig. 3F). Consistent with this result, forcing pairing of that bulge in WT MIR16-1 (MIR16-1 WTFBE) did not improve processing efficiency over WT (Fig. 3F).

We next asked whether shifting the bulge from the center of the bulge-depleted region could lead to enhanced processing. Due to technical difficulty in shifting the bulge at position 19
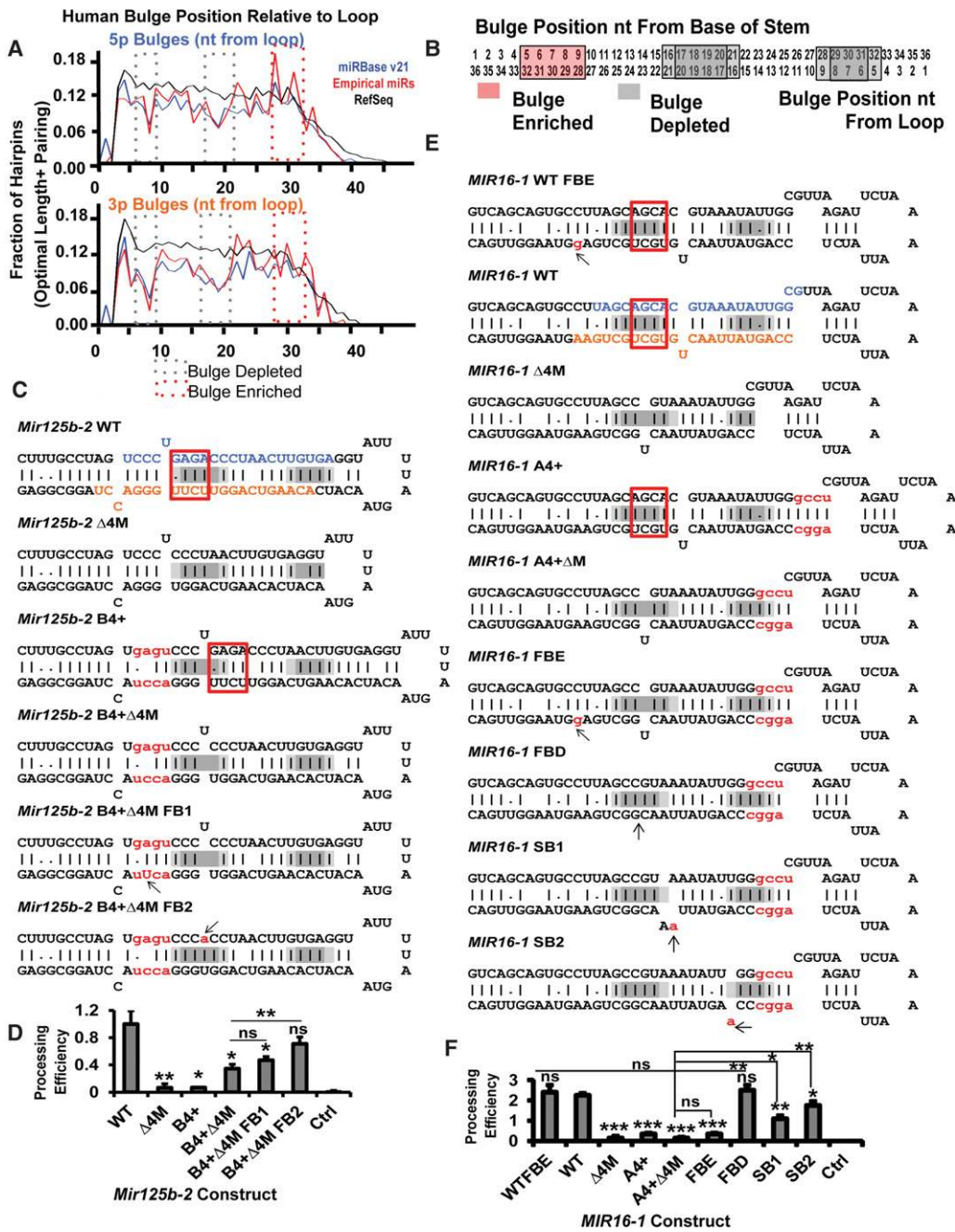
Figure 3. The roles of the bulge-depleted regions in pri-miRNA processing. (*A*) The distributions of bulges along hairpin stem were plotted with distance measured from the junction between apical loop and the stem in nucleotides, for human miRBase v21 hairpins, Empirical miRNA hairpins, or RefSeq hairpins. Hairpins were preselected to have optimal length (33–39 nt) and pairing (≥82%). Bulge distributions along the 5p arm (*top*) or 3p arm (*bottom*) of the hairpin were plotted. The bulge-depleted and bulge-enriched regions were indicated with dashed boxes. (*B*) A diagram for bulge-depleted and bulge-enriched regions was illustrated for hairpins with a stem length of 36 nt. The numbers indicate position from the base of the stem (*left* side) or apical/stem junction (*right* side). The bulge-enriched region was depicted with a pink box, whereas bulge-depleted regions were depicted with gray boxes, with the edge of the depleted regions in lighter gray to reflect less depletion. (*C*) Hairpin structures for WT or mutant mouse *Mir125b-2* hairpins are illustrated. Color-coded elements include 5p mature miRNA (blue), 3p miRNA (orange), insertions (red letters; lowercase), and deletions that occurred in other related constructs (red box). Watson-Crick pairings are indicated with vertical bars, whereas G:U pairings are indicated with dots. The positions of the bulge-depleted regions, as measured from the base of the stem, are shaded in gray. (*D*) The constructs in *C* were subjected to the processing reporter assay in BaF3 cells, with data normalized and the levels of WT mouse *Mir125b-2* set to one and an empty vector (Ctrl) set to zero. *N* = 3. (*E*) Hairpin structures for WT or mutant *MIR16-1* are depicted, with the same color-coding system as in *C*. Arrows point to nucleotide alterations. (*F*) The constructs in *E* were subjected to the processing reporter assay in BaF3 cells, with data normalized and the levels of WT mouse *Mir125b-2* set to one and an empty vector (Ctrl) set to zero. *N* = 3. Error bars, SD. *P*-values are annotated *above* the bars for comparison with WT construct. Other *P*-value comparisons are indicated with horizontal bars. (*) $P < 0.05$; (**) $P < 0.01$; (***) $P < 0.001$; (ns) not significant.

(counting from base) to a bulge-tolerated region without affecting overall folding, we were only able to design two mutants that moved the bulge to the edges of the bulge-depleted regions (*MIR16-1* SB1 and SB2) (Fig. 3E), which are less depleted than the center of the regions (Fig. 3B). In both cases, significant, but partial rescue was observed when comparing to A4+Δ4M construct, although processing was lower than the level of WT (Fig. 3F).

Taken together, our data above support that bulges in bulge-depleted regions inhibit processing, with stronger effects seen at the center of these regions, whereas bulges outside of such regions are relatively more tolerated.

CNNC distance, but not nucleotide identity, strongly affects pri-miRNA processing

Among the primary sequence motifs, we focused on the CNNC motif because it is the only known primary sequence motif that does not have a fixed location within hairpin. It remains unclear whether the distance of CNNC motif and whether the 16 possible nucleotide combinations of "NN" within CNNC (referred to as CNNC subtypes) functionally affect processing.

We first examined the location of the CNNC motif. Because of the absence of mature miRNAs within RefSeq hairpins, we calculated the CNNC distance based on the position of the first C in CNNC relative to the base of the stem. Comparing between RefSeq and miRNAs hairpins, we observed a broad range of enrichment (Fig. 4A; Supplemental Fig. S4A). Based on these distributions, we used two criteria, CNNC5-6 and CNNC3-11, to refer to CNNC distance. Both CNNC distance criteria were significantly enriched in miRNA hairpins (Fig. 1C).

Mouse *Mir125b-2* harbors two putative CNNC motifs within the enriched range, located with overlap at positions 7 (CNNC1) and 10 (CNNC2), respectively (Fig. 4B). We first asked whether both CNNCs are functional. Surprisingly, mutation of CNNC1 (KCNNC1), which fell closer to the highly enriched 5- to 6-nt range, resulted in a slight increase in processing efficiency (Fig. 4C). In contrast, mutation of CNNC2 (KCNNC2), which was located further away from the peak of enriched positions, strongly reduced the processing of this miRNA (Fig. 4C). Evolutionarily, CNNC2 was more conserved than CNNC1, even though the NN of CNNC2 could be variable (Supplemental Fig. S4B). These data revealed the functional CNNC within mouse *Mir125b-2* and also proved that not all putative CNNC motifs within the enriched distance range are functional.

To test the functional relevance of the distance of CNNC, we made an insertion of 5 nt into our KCNNC1 construct, effectively extending the CNNC2 position to 15 nt from the stem base (SCNNC5), a distance that is out of the range of the enrichment. This mutant resulted in a strong reduction of processing similar to the level seen for KCNNC2 (Fig. 4C), indicating that the distance of CNNC is important for miRNA processing.

We next examined the 16 possible CNNC subtypes by comparing the distributions within miRNA hairpins versus RefSeq hairpins. Although there was strong variability in the occurrence of a CNNC subtype near miRNA hairpins, a similar pattern was seen for RefSeq hairpins (Supplemental Fig. S4C,D) and paralleled the overall nonrandom distribution of CNNC subtypes in RefSeq sequences regardless of their location (Supplemental Fig. S4E). These distributions suggest that the CNNC subtypes do not differ strongly in processing. To test this possibility, we created five mutants by mutating the CNNC2 of *Mir125b-2*, representing six CNNC subtypes with low, medium, and high occurrences

(Supplemental Fig. S4C–E, pink-lettered CNNCs). Despite some mutants producing statistically significant changes in processing efficiency, the levels of such changes were small (Supplemental Fig. S4F). We further asked whether these CNNC mutants could respond to constitutively active YAP1 overexpression (Supplemental Fig. S4G), which can sequester the CNNC-binding protein DDX17 and reduce processing (Mori et al. 2014). YAP1 significantly suppressed all CNNC motif mutants, albeit with some levels of quantitative differences (Supplemental Fig. S4H). Our data above indicate that the CNNC distance strongly affects processing efficiency, whereas alteration of CNNC subtypes has no or mild effect on processing.

The CNNC motif preferentially enhances the processing of hairpins of optimal stem lengths

We observed that the enrichment of putative CNNC motifs was much stronger in miRNA hairpins with optimal length (33–39 nt) than those without (<33 or >39 nt) (Fig. 4D; Supplemental Fig. S4I). These data indicate that putative CNNC motifs preferentially co-occur with optimal length miRNA hairpins, and raise the possibility that the CNNC motif cooperates with optimal stem length to enhance processing.

We addressed this possibility first by performing combination mutations in mouse *Mir125b-2*. As we have demonstrated above, both increasing stem length (A4+) and mutating CNNC (KCNNC2) resulted in a strong reduction of processing efficiency, but not completely, compared with the ΔHairpin level (Fig. 4E). Indeed, the KCNNC2 mutant produced mature mmu-miR-125b to ~12% of WT levels and significantly reduced pri-miRNA levels in comparison to ΔHairpin control (Supplemental Fig. S4J), further indicating that it is not completely inactive for processing. The combination of A4+ and KCNNC2 mutations did not result in a further decrease of processing efficiency, suggesting that CNNC enhances optimal length hairpin.

To further test this notion, we resorted to human *MIR579* WT (44 nt) and *MIR579* M7- mutant (37 nt) with the latter being within the optimal length range. *MIR579* does not contain a CNNC motif within the enriched distances. When we engineered an artificial CNNC motif in WT *MIR579* (579CNNC+), we did not observe any increase in processing (Fig. 4F; Supplemental Fig. S4K). A mild increase was observed when the stem length was reduced to 37 nt (Fig. 4F, *MIR579* M7- mutant). A combination of CNNC and the stem length reduction (*MIR579* M7-CNNC+) resulted in a stronger increase in processing relative to the WT construct (Fig. 4F), supporting cooperation of the engineered CNNC with optimal stem length.

To more comprehensively examine the functional interaction between the CNNC motif and the stem length, we obtained the mature miRNA expression data (Mori et al. 2014) for which the CNNC-binding protein DDX17 has been knocked down in HaCaT cells. For all mature miRNAs derived from hairpins containing putative CNNC motifs within the enriched distance range, we plotted the stem length and the level of mature miRNA reduction upon *DDX17* knockdown. Strikingly, we observed that expression reductions were associated with stem length within or very close to the optimal range of 33–39 nt (Fig. 4G). When we examined all miRNAs with greater than twofold reduction upon *DDX17* knockdown, there were significantly more miRNAs with both putative CNNC and optimal stem length than those without putative CNNC, without optimal stem length, or without both (Fig. 4H). These data support that miRNAs with both optimal stem length
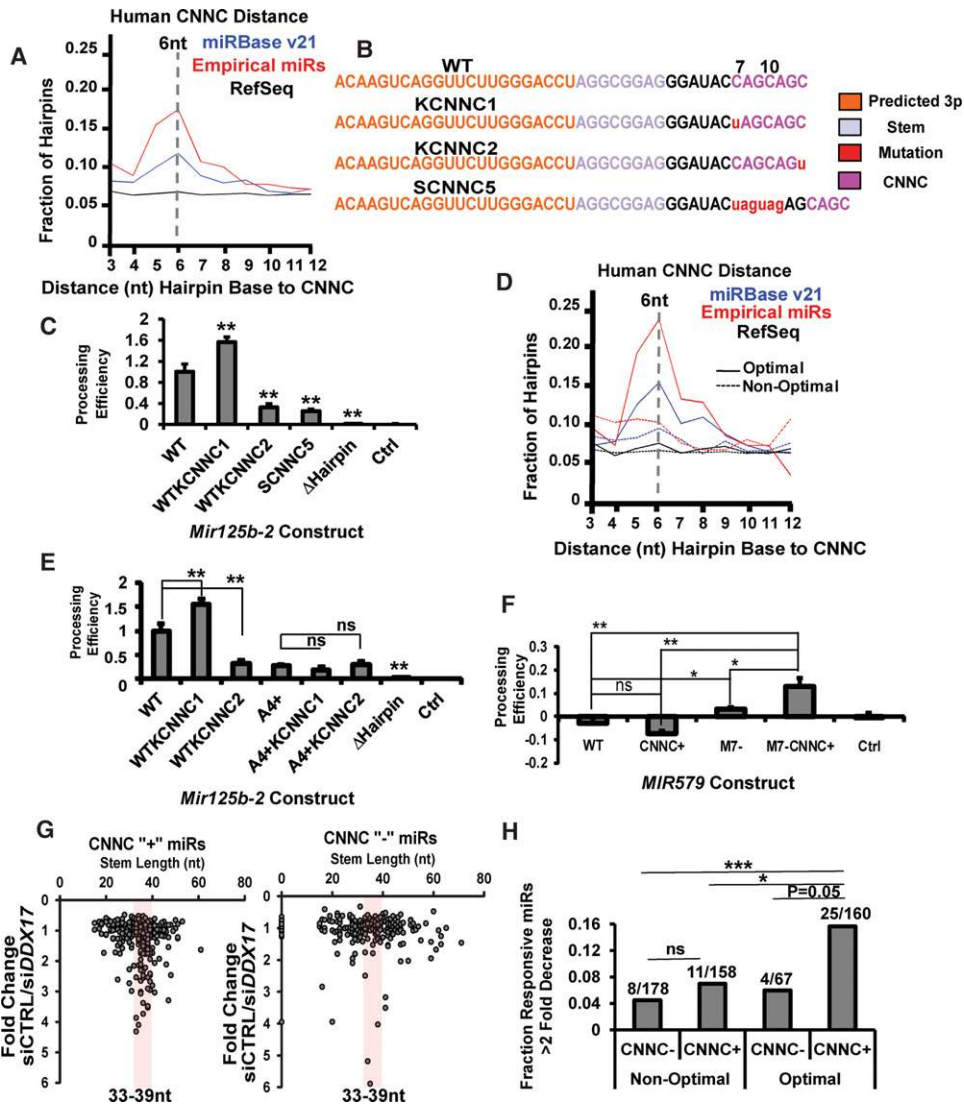
Figure 4. The roles of the CNNC motif in pri-miRNA processing. (*A*) The distributions of putative CNNC motifs were plotted for human hairpins from miRBase, Empirical miRNA set, or RefSeq. The distance of CNNC motif was measured in nucleotides from the base of the stem to the first C in the motif. The most enriched position (6 nt) was highlighted. (*B,C*) Processing of WT mouse *Mir125b-2* or its CNNC mutants was assayed. The sequences of CNNC mutants were depicted in *B*. Color-coded elements include the 3p miRNA shown in orange, nucleotides in the stem shown in lavender, mutated or inserted nucleotides shown in red, putative CNNC motifs in magenta, and positions relative to stem base indicated by numbers. (*C*) BaF3 cells were transduced with the processing reporters, and the processing efficiencies were measured, with the level for WT mouse *Mir125b-2* set to one and the level of an empty vector (Ctrl) set to zero. A construct removing the hairpin in mouse *Mir125b-2* (ΔHairpin) was also used as a control. *N* = 3. *P*-values were annotated for comparison with WT construct. (*D*) Distributions of the CNNC motifs, among miRNA or non-miRNA hairpins that were of optimal (33–39 nt) or nonoptimal (<33 or >39 nt) length. CNNC motifs preferentially co-occur with optimal length miRNA hairpins. (*E*) WT mouse *Mir125b-2* or combination mutants containing stem length alterations (A4+) (Fig. 2) and CNNC mutations (see *C*) were measured in processing reporter assays in BaF3 cells. *N* = 3. (*F*) Processing efficiencies for WT *MIR579*, which is both long (44 nt of stem length) and without a CNNC motif, and its mutants were measured in BaF3 cells. Mutant structures are illustrated in Supplemental Figure S4K, with the CNNC+ mutant containing an engineered putative CNNC motif 8 nt from stem base of *MIR579*, the M7-mutant shortening the stem length by 7 nt, and the M7-CNNC+ mutant having both shortened stem length and a CNNC motif. Data were normalized the same way as in *B*. *N* = 3. (*G*) The fold-decrease of mature miRNA expression upon siRNA knockdown of *DDX17* versus a control siRNA (siCtrl) was plotted against the stem length. Each dot represents a single miRNA, with those containing putative CNNC motifs (CNNC+) and without (CNNC−) plotted separately. The optimal stem length range (33–39 nt) was highlighted. (*H*) Data from *G* were plotted to quantify the number of miRNAs with decreased expression (greater than twofold) upon *DDX17* knockdown, for miRNAs with optimal or nonoptimal stem length, and with putative CNNC or without. Numbers *above* the bars indicate the number of decreased miRNAs out of all miRNAs in the indicated category. Error bars, SD. (*) *P* < 0.05; (**) *P* < 0.01; (***) *P* < 0.001; (ns) not significant.

and putative CNNC motif(s) are selectively sensitive to the loss of the CNNC-binding protein DDX17.

We also reasoned that if optimal stem length, CNNC, and the combination of both are important for miRNA processing, we would expect an increased mature miRNA–to–pri-miRNA ratio from endogenous cellular transcripts for miRNAs with these features. To test this, we obtained RNA-seq data for pri-miRNA expression levels in four human cell lines (MCF7, HCT116, 293T, and

HepG2) published by the Mendell laboratory (Chang et al. 2015), as well as publicly available mature miRNA sequencing data from other laboratories (Bogerd et al. 2014; Cao et al. 2016; Hannafon et al. 2016). By examining miRNAs with a higher mature–to–pri-miRNA ratio versus those with a lower ratio, we observed significant enrichment for optimal stem length or CNNC, and a stronger enrichment when both features were present (Supplemental Fig. S5A–D). Taken together, our data above support a model that the CNNC motif preferentially co-occurs with and enhances the processing of hairpins with optimal stem lengths.

Systematic evaluation of the effects of human SNPs on pri-miRNA processing

With the known and new rules of pri-miRNA processing obtained above, we applied these rules to human SNPs. A total of 17,948 SNP alleles (dbSNP human Build 142) were found within 30 bases of human pre-miRNAs. We systematically annotated hairpin structural and sequence features on these alleles using our HairpIndex pipeline and compared such features of minor alleles to those of the major alleles. Results are summarized in Supplemental Table S5.

Overall, we found 0.4%–2.7% of SNPs in each of the feature categories, which were predicted to be favorable or detrimental to miRNA processing (for details, see Methods) (Fig. 5A; Supplemental Table S5). Due to limited numbers (135 favorable, 177 detrimental) of SNP minor alleles that were within the optimal stem length range and having optimal pairing, when we compared SNP alleles of different minor allele frequencies (MAFs), we did not observe many statistically significant changes. Nevertheless, common minor alleles (MAF ≥ 0.1) tend to be depleted for detrimental structural changes (hairpin number—meaning failure to identify hairpin, and percentage stem pairing) and enriched for favorable structural features (optimal stem length), compared with low MAF minor alleles (Supplemental Fig. S6A,B).

To confirm some of our predicted changes in secondary structure features, we took three SNPs in two pri-miRNAs for validation. For SNP rs371589474, which is located within the stem of MIR126, we predicted a reduction of stem length from 32 nt in the major "G" allele to 16 nt in the minor "A" allele. Indeed, this minor allele led to an approximately fivefold reduction in processing efficiency. In contrast, another G → A SNP

Figure 5. The effects of human SNPs on pri-miRNA processing. (A) Human SNPs in dbSNP human Build 142 that are located close to human pri-miRNAs were evaluated for their impact on hairpin structure and sequence features. For each SNP, hairpin features for minor and major alleles were compared. The fractions of minor alleles that have predicted favorable (gray) or detrimental (blue) impact (relative to the major allele) on pri-miRNA processing were plotted. (B) Predicted hairpin structures of major and minor alleles tested in C through E. Color-coded elements include 5p mature miRNA (blue), 3p miRNA (orange), major allele base (red letter; uppercase), and minor allele base (red letter; lowercase). Watson-Crick pairings are indicated with vertical bars, whereas G:U pairings are indicated with dots. The positions of the bulge-depleted regions, as measured from the base of the stem, are shaded in gray. Arrow and yellow-highlighted red text indicate the sequence variation. (C–E) The indicated pri-miRNA constructs were subjected to processing reporter assay in BaF3 cells. Processing efficiencies were normalized to mouse Mir125b-2 WT (set to one) and an empty vector control (Ctrl; set to zero). N = 3. Error bars, SD. (*) P < 0.05; (**) P < 0.01; (ns) not significant.

(rs4636297) in the same miRNA that was not predicted to alter stem length did not alter processing (Fig. 5B,C). Another example is SNP rs543412 located within the MIR100 hairpin flanking region but outside of the predicted stem (Fig. 5B). The minor "A" allele resulted in predicted secondary changes that propagated into the stem to result in a shortening of stem length from 35 nt to 25 nt.

Consistent with this prediction, the minor allele significantly reduced the processing efficiency by 35% (Fig. 5D). Given that different molecules of the same RNA sequence may adopt multiple structures in vivo, we further modeled this using the Sfold program to draw 1000 structures from the Boltzmann distribution. Analyses of *MIR126* and *MIR100* SNPs resulted in alterations in the fractions of structures with optimal or near optimal stem lengths with similar trends as the experimental data (Supplemental Fig. S6C,D). These data indicate that single-nucleotide changes within stem or outside of stem can lead to alterations in miRNA secondary structure and result in alterations in processing efficiency.

To validate predictions on bulges, we examined a "C" mutation present at position 57 in the stem of *MIR96*, which is associated with deafness, identified in an Italian family. Previously, it was speculated that this minor allele affects DICER1 processing of the miRNA (Solda et al. 2012). We predicted that there was an increase of the bulge size in a bulge-depleted region, which should reduce pri-miRNA processing (Fig. 5B). Indeed, by measuring the pri-miRNA processing assay, this minor allele resulted in approximately twofold reduction of pri-miRNA processing efficiency (Fig. 5E). These data provide support that human SNPs leading to alterations in bulges can impact pri-miRNA processing, and suggest that these processing defects may contribute to mechanisms underlying disease associations.
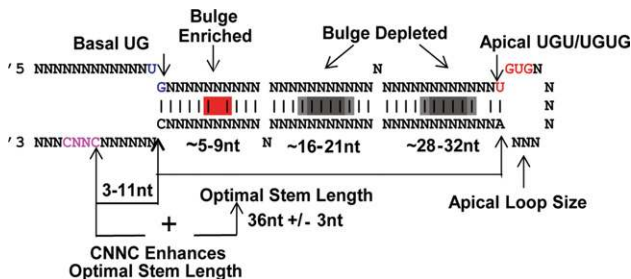
Discussion

In this study, we systematically examined 30 features of mammalian miRNA hairpins in comparison to predicted hairpins in RefSeq sequences. These analyses, coupled with experimental validation, led to a set of previously unrecognized rules that govern efficient mammalian pri-miRNA processing (summarized in Fig. 6). We believe that our approach is complementary to previous studies utilizing random mutagenesis of a few miRNAs (Fang and Bartel 2015). Given that our computational comparisons were made on human and mouse sequences and that experimental validations were performed on a small set of mammalian miRNAs, whether these rules can be extrapolated to other species requires further studies. We also cannot exclude the possibility that unknown protein factors may be involved in the processing of the experimentally tested miRNAs in this study.

One of our findings is the presence of two bulge-depleted regions in miRNA hairpin stem. We propose that bulges located

Figure 6. Summary of hairpin features influencing processing. Diagram of hairpin features that influence processing. An optimal stem length of 36 ± 3 nt is favorable for processing, and the appearance of bulges in the two bulge-depleted regions inhibits processing. The CNNC motif preferentially enhances the processing of optimal length pri-miRNAs and has to locate within a distance limit from the base of the hairpin. In addition, other primary sequence motifs and structural features, including basal UG, apical UGU/UGUG, loop size, and a bulge-enriched region, are highlighted.

within the bulge-depleted regions are more detrimental for miRNA processing than those outside of these regions. Interestingly, the two bulge-depleted regions were offset roughly by one helical turn, suggesting that these regions could be protein-interacting surfaces, possibly with DGCR8. Note that since the positions of the bulge-depleted regions were derived from computational analysis of optimal length hairpins and validated on miRNAs within optimal stem length range, caution should be applied for hairpins of longer or shorter lengths. Previously, Fang and Bartel (2015) have used a random mutagenesis approach on three miRNAs and concluded that bulges, other than a bulge-enriched region (confirmed in our study), decrease pri-miRNA processing regardless of location. In contrast, our findings support that the detrimental effects of bulges are dependent on location. Although we do not fully understand the reasons for the differential findings, we notice several technical differences. First, the comparison to non-miRNA hairpins in our study may have assisted the observation of the depletion signals. Second, Fang and Bartel (2015) have assumed a local effect of a single unpaired base when considering bulges, which may not be applicable in all cases. Third, Fang and Bartel (2015) utilized a cell-free system, whereas our findings were confirmed in vivo, and it is possible that an unknown factor present in the cell is responsible for this difference.

We propose that the optimal length range for miRNA stem is 36 ± 3 nt (based on counting rules in the Methods), which should be considered as an average for all miRNAs. Our proposed range differs from the previously proposed ~33 nt and 35 ± 1 nt for miRNA hairpins (Han et al. 2006; Fang and Bartel 2015). One possibility of the differences of the peak of stem length could be due to the differences in approach, with the previous studies focusing on several miRNAs and their variants and with our study examining many more miRNAs and thus potentially more diverse hairpin backbones. Another possibility is that these peak differences are simply due to different counting methods, especially given that previous studies did not explicitly specify the rules of counting stem length. Beyond the differences in the peak of stem length, we show that for mouse *Mir125b-2*, it can tolerate at least three additional paired bases without major changes in processing, which is more than the previous 35 ± 1-nt proposal (Fang and Bartel 2015). We demonstrated evidence with stem length rescue that the mutant phenotypes were majorly driven by stem length rather than primary sequence alterations, but we cannot completely exclude the possibility that sequence alterations may also contribute to the changes. We further demonstrate that the CNNC motif selectively enhances the processing of hairpins that are within or close to the optimal length range. Previously, Fang and Bartel (2015) have examined the combined effect of CNNC, basal UG, and apical UGU/UGUG motifs and concluded that the combination of all these features enhances the processing of nonoptimal length hairpins, but the relationship between CNNC and stem length has not been thoroughly examined separately. Future experiments can be directed to elucidate the molecular mechanisms of optimal stem lengths and their relationship with the CNNC motif.

Aided with existing and new rules for efficient miRNA processing, we surveyed and predicted the potential effects of human SNPs on pri-miRNA processing. We further validated several examples experimentally, which support that small sequence alterations, in most cases single-nucleotide substitutions, can lead to changes in pri-miRNA structure in violation of miRNA processing rules and consequently decrease processing activity. These findings not only lend support for the processing rules but also reveal

previously unappreciated molecular alterations driven by human SNPs. While the prediction of RNA secondary structures by RNA folding programs may not always be accurate, experimental validation in this study does support that SNPs can reduce pri-miRNA processing when rules are violated. We speculate that careful modeling based on a statistical sample of RNA secondary structures generated from the Boltzmann structure ensemble (Ding and Lawrence 2003) may lead to improvement in prediction and possibly new rules underlying pri-miRNA processing in the future.

Methods

Computational analysis of hairpin structures and features

Sources of miRNA and other sequence data

MiRNA sequences were obtained from miRBase (Kozomara and Griffiths-Jones 2014); RefSeq sequences were downloaded from UCSC Genome Browser (hg38 and mm10), and human SNPs were obtained from dbSNP human Build 142 for hg38. In addition, a set of manually curated human mutations not present in the SNP data set was included in the analysis (see Supplemental Table S6). For additional details on source data, see Supplemental Methods.

RNA folding and HairpIndex pipeline

Sequences were folded using the Sfold program (Ding and Lawrence 2003; Ding et al. 2005) and RNAfold program version 2.1.9 (Gruber et al. 2008). Folding results were then analyzed using the HairpIndex program.

Cell culture / reporter assay

All cell lines were from ATCC, and culture conditions are described in the Supplemental Methods. Production of retro or lentiviruses was performed following our published protocols (Lu et al. 2008; Adams et al. 2012; Guo et al. 2012). Additional information regarding viral production, infection, and selection can be found in the Supplemental Methods. Cells were infected by the lentiviral processing reporter or control reporters and harvested for flow cytometry analysis (GFP and mCherry fluorescence) on LSRII (BD Biosciences). Details of infection procedure can be found in the Supplemental Methods.

Cloning and constructs

Pri-miRNA processing reporters were cloned into our pri-miRNA processing vector, which is described in detail in our recent study (Cheng et al. 2016). Details of cloning procedure can be found in the Supplemental Methods. The specific designs for each pri-miRNA sequence, as well as their mutants, are detailed in Supplemental Table S7. ShRNA target sequences can be found in the Supplemental Methods.

Real-time PCR analysis

Detailed protocol and specific primers are provided in the Supplemental Methods. Total RNA was extracted from puromycin-selected cells. CDNAs were synthesized using a reverse transcription kit and random primers. QPCR analysis was performed using SYBR green. Differential expression was calculated using the $\Delta\Delta CT$ method.

Statistical analysis

Student's t-test was used for analyzing statistical significance of experimental data, except for those specified otherwise.

Software availability

Details of the HairpIndex program, source code, and enrichment calculations for analysis can be found in the Supplemental Material.

References

Adams BD, Guo S, Bai H, Guo Y, Megyola CM, Cheng J, Heydari K, Xiao C, Reddy EP, Lu J. 2012. An in vivo functional screen uncovers miR-150-mediated regulation of hematopoietic injury response. *Cell Rep* **2:** 1048–1060.

Alarcon CR, Goodarzi H, Lee H, Liu X, Tavazoie S, Tavazoie SF. 2015a. HNRNPA2B1 is a mediator of m6A-dependent nuclear RNA processing events. *Cell* **162:** 1299–1308.

Alarcon CR, Lee H, Goodarzi H, Halberg N, Tavazoie SF. 2015b. N^6-methyladenosine marks primary microRNAs for processing. *Nature* **519:** 482–485.

Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. 2013. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152:** 844–858.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409:** 363–366.

Bogerd HP, Whisnant AW, Kennedy EM, Flores O, Cullen BR. 2014. Derivation and characterization of Dicer- and microRNA-deficient human cells. *RNA* **20:** 923–937.

Cao B, Wang K, Liao JM, Zhou X, Liao P, Zeng SX, He M, Chen L, He Y, Li W, et al. 2016. Inactivation of oncogenic cAMP-specific phosphodiesterase 4D by miR-139-5p in response to p53 activation. *eLife* **5:** e15978.

Chang TC, Pertea M, Lee S, Salzberg SL, Mendell JT. 2015. Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res* **25:** 1401–1409.

Cheng TL, Wang Z, Liao Q, Zhu Y, Zhou WH, Xu W, Qiu Z. 2014. MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. *Dev Cell* **28:** 547–560.

Cheng J, Roden CA, Pan W, Zhu S, Baccei A, Pan X, Jiang T, Kluger Y, Weissman SM, Guo S, et al. 2016. A Molecular Chipper technology for CRISPR sgRNA library generation and functional mapping of noncoding regions. *Nat Commun* **7:** 11178.

Davis BN, Hilyard AC, Lagna G, Hata A. 2008. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454:** 56–61.

Davis BN, Hilyard AC, Nguyen PH, Lagna G, Hata A. 2010. Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. *Mol Cell* **39:** 373–384.

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432:** 231–235.

Di Carlo V, Grossi E, Laneve P, Morlando M, Dini Modigliani S, Ballarino M, Bozzoni I, Caffarelli E. 2013. TDP-43 regulates the microprocessor complex activity during in vitro neuronal differentiation. *Mol Neurobiol* **48:** 952–963.

Ding Y, Lawrence CE. 2003. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* **31:** 7280–7301.

Ding Y, Chan CY, Lawrence CE. 2005. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11:** 1157–1166.

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

Farazi TA, Spitzer JI, Morozov P, Tuschl T. 2011. miRNAs in human cancer. *J Pathol* **223:** 102–115.

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235–240.

Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA websuite. *Nucleic Acids Res* **36**(Web Server issue): W70–W74.

Guil S, Caceres JF. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* **14:** 591–596.

Guo S, Bai H, Megyola CM, Halene S, Krause DS, Scadden DT, Lu J. 2012. Complex oncogene dependence in microRNA-125a–induced myeloproliferative neoplasms. *Proc Natl Acad Sci* **109:** 16636–16641.

Guo Y, Liu J, Elfenbein SJ, Ma Y, Zhong M, Qiu C, Ding Y, Lu J. 2015. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res* **43:** 2326–2341.

Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18:** 3016–3027.

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125:** 887–901.

Hannafon BN, Trigoso YD, Calloway CL, Zhao YD, Lum DH, Welm AL, Zhao ZJ, Blick KE, Dooley WC, Ding WQ. 2016. Plasma exosome microRNAs are indicative of breast cancer. *Breast Cancer Res* **18:** 90.

Iorio MV, Croce CM. 2012. microRNA involvement in human cancer. *Carcinogenesis* **33:** 1126–1133.

Karginov FV, Cheloufi S, Chong MM, Stark A, Smith AD, Hannon GJ. 2010. Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol Cell* **38:** 781–788.

Kawahara Y, Mieda-Sato A. 2012. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc Natl Acad Sci* **109:** 3347–3352.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**(Database issue): D68–D73.

Kwon SC, Nguyen TA, Choi YG, Jo MH, Hohng S, Kim VN, Woo JS. 2016. Structure of human DROSHA. *Cell* **164:** 81–90.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415–419.

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435:** 834–838.

Lu J, Guo S, Ebert BL, Zhang H, Peng X, Bosco J, Pretz J, Schlanger R, Wang JY, Mak RH, et al. 2008. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Dev Cell* **14:** 843–853.

Ma H, Wu Y, Choi JG, Wu H. 2013. Lower and upper stem–single-stranded RNA junctions together determine the Drosha cleavage site. *Proc Natl Acad Sci* **110:** 20687–20692.

Macias S, Plass M, Stajuda A, Michlewski G, Eyras E, Caceres JF. 2012. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat Struct Mol Biol* **19:** 760–766.

Mori M, Triboulet R, Mohseni M, Schlegelmilch K, Shrestha K, Camargo FD, Gregory RI. 2014. Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer. *Cell* **156:** 893–906.

Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. 2011. Molecular basis for interaction of *let-7* microRNAs with Lin28. *Cell* **147:** 1080–1091.

Newman MA, Thomson JM, Hammond SM. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14:** 1539–1549.

Nguyen TA, Jo MH, Choi YG, Park J, Kwon SC, Hohng S, Kim VN, Woo JS. 2015. Functional anatomy of the human microprocessor. *Cell* **161:** 1374–1387.

Paroo Z, Ye X, Chen S, Liu Q. 2009. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell* **139:** 112–122.

Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, Iliopoulos D, Gregory RI. 2011. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147:** 1066–1079.

Solda G, Robusto M, Primignani P, Castorina P, Benzoni E, Cesarani A, Ambrosetti U, Asselta R, Duga S. 2012. A novel mutation within the *MIR96* gene causes non-syndromic inherited hearing loss in an Italian family by altering pre-miRNA processing. *Hum Mol Genet* **21:** 577–585.

Tang X, Li M, Tucker L, Ramratnam B. 2011. Glycogen synthase kinase 3 β (GSK3β) phosphorylates the RNAase III enzyme Drosha at S300 and S302. *PLoS One* **6:** e20391.

Trabucchi M, Briata P, Garcia-Mayoral M, Haase AD, Filipowicz W, Ramos A, Gherzi R, Rosenfeld MG. 2009. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459:** 1010–1014.

Viswanathan SR, Daley GQ, Gregory RI. 2008. Selective blockade of microRNA processing by Lin28. *Science* **320:** 97–100.

Wada T, Kikuchi J, Furukawa Y. 2012. Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. *EMBO Rep* **13:** 142–149.

Weitz SH, Gong M, Barr I, Weiss S, Guo F. 2014. Processing of microRNA primary transcripts requires heme in mammalian cells. *Proc Natl Acad Sci* **111:** 1861–1866.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* **13:** 13–21.

Zeng Y, Cullen BR. 2003. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9:** 112–123.