➜ NON-CODING RNA

Pri-miRNA processing: structure is key

> **"** these features cooperate to improve processing efficiency **"**

MicroRNAs (miRNAs) are processed from hairpin-containing primary transcripts (pri-miRNA), but how pri-miRNAs are selected for processing given the ubiquity of hairpin motifs in non-coding transcripts is poorly understood. A new computational study with experimental validation in *Genome Research* has identified several features that are enriched in pri-miRNA hairpins and that seem to govern efficient pri-miRNA processing.

In animal cells, pri-miRNAs are processed into pre-miRNAs by the Microprocessor complex in the nucleus, and are then transported into the cytoplasm to undergo further processing into mature miRNAs. Previous studies, many of which modelled specific miRNAs and their mutants, have identified several features that enable the efficient processing of pri-miRNA, including an optimal hairpin stem length of ~35 nucleotides and apical loop sizes of ~3–23 nucleotides. In addition, the presence of some primary sequence motifs — for example, the CNNC motif and the UG motif — seem to enhance processing.

The study by Roden *et al.* builds on earlier research by using a computation-based approach to compare the structural and sequence features of pri-miRNA hairpins and other hairpin-containing transcripts. The authors identified several hairpin features that are enriched in miRNA hairpins. The most distinguishing feature was an optimal stem length (defined as a length of 33–39 nucleotides). Using an *in vivo* reporter vector for pri-miRNA processing, the team noted that mutant mouse *miR125b-2* transcripts (with stem lengths of 31 or 39 nucleotides) reduced the processing efficiency compared with wild-type *miR125b-2* (stem length of 35 nucleotides), confirming the importance of stem length for pri-miRNA processing.

Roden *et al.* also investigated the role of bulges (that is, sequence mismatches) and their location in pri-miRNA processing. Bulges were identified in a relatively uniform distribution in the stem of non-pri-miRNAs, but were enriched at ~5–9 nucleotides from the stem base in pri-miRNAs. Moreover, the team identified two bulge-depleted regions in pri-miRNAs, located at ~5–9 and ~16–21 nucleotides relative to the apical loop, or ~16–21 and ~28–32 nucleotides relative to the base of the hairpin; they propose that bulges in these regions are more detrimental for efficient processing than those located outside these regions.

In addition, the CNNC primary sequence motif was enriched in pri-miRNA hairpins of optimal length, compared with those of suboptimal lengths. Indeed, the CNNC motif selectively enhanced processing of hairpins with an optimal, or near optimal, stem length, supporting the idea that these features cooperate to improve processing efficiency.

Having established some of the rules governing pri-miRNA processing, the authors set out to determine how these might be affected by genetic variation (that is, single-nucleotide polymorphisms (SNPs)). The researchers identified 17,948 SNPs within 30 bases of human pre-miRNAs and systematically annotated hairpin structural and sequence features on these alleles. Overall, common alleles had fewer detrimental features and more favourable structural features than rare alleles.

Additional analyses led the authors to surmise that changes in stem length and bulge size might alter RNA secondary structure and thereby lead to processing defects that underlie disease associations.

Despite the shortcomings associated with predicting RNA structure using computer-based software, that several of the hairpin features speculated to affect pri-miRNA processing could be validated experimentally is promising. Further studies will be required to establish the molecular mechanisms underlying some of these features.

Louise Adams, Associate Editor,
Nature Reviews Disease Primers

ORIGINAL ARTICLE Roden, C. *et al.* Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. *Genome Res.* http://dx.doi.org/10.1101/gr.208900.116 (2017)