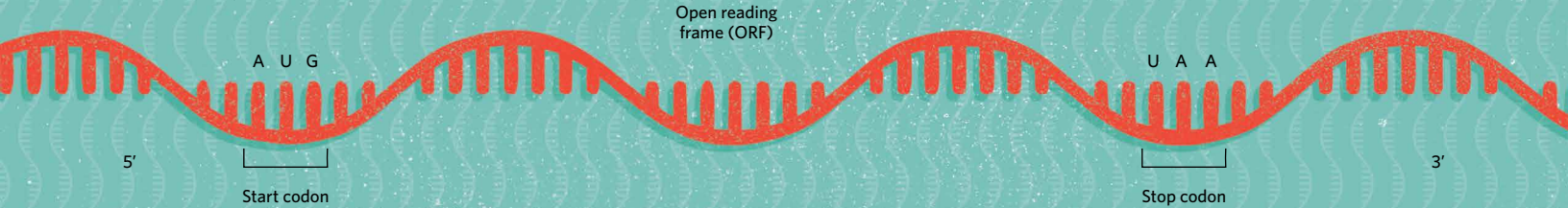
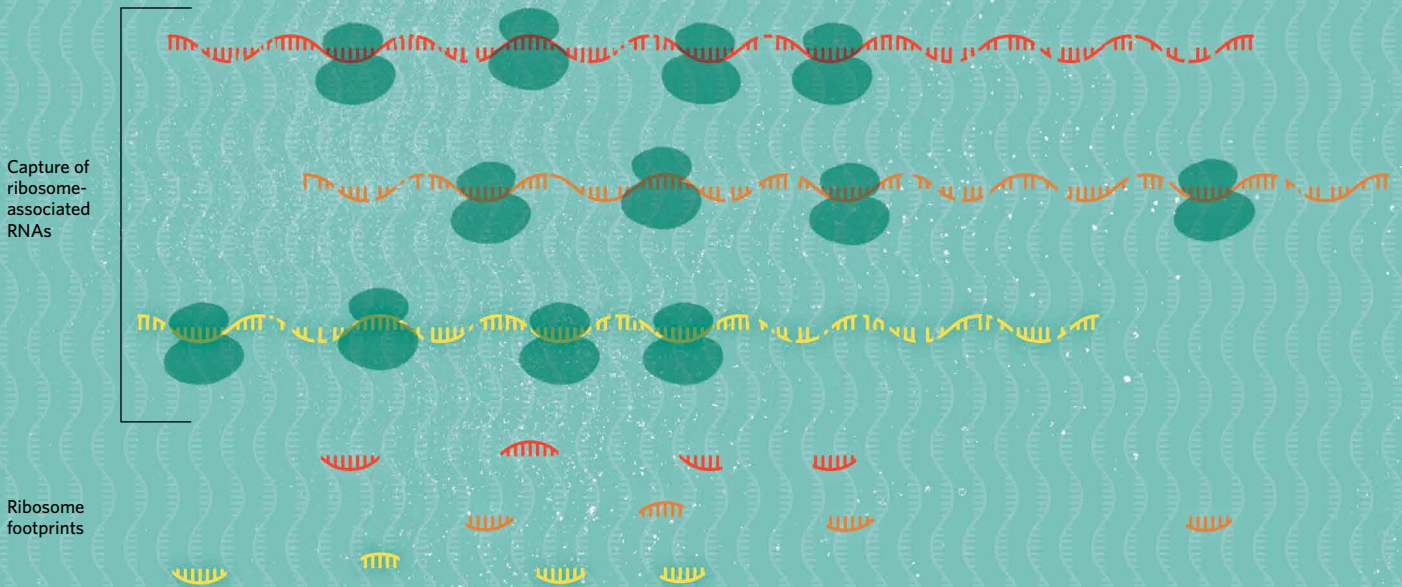


FOLLOWING THE CODE

With the advent of genome sequencing technologies, researchers began combing genomes for open reading frames (ORFs). To enrich for genuine protein-coding ORFs and to eliminate those random sequences that by chance were bookended by start and stop codons, most ORF-finding algorithms ignored any stretches shorter than 300 nucleotides. Unfortunately, this also meant that many short ORFs encoding micropeptides were missed. Now, new techniques are helping scientists identify tiny ORFs within what were presumed to be long noncoding RNAs.



To search for coding RNAs directly, rather than through the genome, researchers turned their attention to translation and implemented a technique known as ribosome footprinting, which involves isolating and digesting ribosome-associated RNAs to leave only those fragments that are protected by the bound ribosomes. Advances in next-generation sequencing technology have allowed researchers to make this process high-throughput, capturing likely translation events across a cell's entire transcriptome.



Not all ribosome-associated RNAs are truly protein coding, however. To identify true protein-coding mRNAs, researchers are now devising analytical techniques such as the ribosome release score (RSS), which assesses the distribution of ribosome-bound fragments along the whole RNA molecule. True mRNAs should have more ribosome-associated regions within the ORF than after the stop codon.

