World Academy of Science, Engineering and Technology
International Journal of Biological and Ecological Engineering
Vol:12, No:04, 2018

Persistent Ribosomal In-Frame Mis-Translation of Stop Codons as Amino Acids in Multiple Open Reading Frames of a Human Long Non-Coding RNA

Authors : Leonard Lipovich, Pattaraporn Thepsuwan, Anton-Scott Goustin, Juan Cai, Donghong Ju, James B. Brown

Abstract : Two-thirds of human genes do not encode any known proteins. Aside from long non-coding RNA (lncRNA) genes with recently-discovered functions, the ~40,000 non-protein-coding human genes remain poorly understood, and a role for their transcripts as de-facto unconventional messenger RNAs has not been formally excluded. Ribosome profiling (Riboseq) predicts translational potential, but without independent evidence of proteins from lncRNA open reading frames (ORFs), ribosome binding of lncRNAs does not prove translation. Previously, we mass-spectrometrically documented translation of specific lncRNAs in human K562 and GM12878 cells. We now examined lncRNA translation in human MCF7 cells, integrating strand-specific Illumina RNAseq, Riboseq, and deep mass spectrometry in biological quadruplicates performed at two core facilities (BGI, China; City of Hope, USA). We excluded known-protein matches. UCSC Genome Browser-assisted manual annotation of imperfect (tryptic-digest-peptides)-to-(lncRNA-three-frame-translations) alignments revealed three peptides hypothetically explicable by 'stop-to-nonstop' in-frame replacement of stop codons by amino acids in two ORFs of the lncRNA MMP24-AS1. To search for this phenomenon genomewide, we designed and implemented a novel pipeline, matching tryptic-digest spectra to wildcard-instead-of-stop versions of repeat-masked, six-frame, whole-genome translations. Along with singleton putative stop-to-nonstop events affecting four other lncRNAs, we identified 24 additional peptides with stop-to-nonstop in-frame substitutions from multiple positive-strand MMP24-AS1 ORFs. Only UAG and UGA, never UAA, stop codons were impacted. All MMP24-AS1-matching spectra met the same significance thresholds as high-confidence known-protein signatures. Targeted resequencing of MMP24-AS1 genomic DNA and cDNA from the same samples did not reveal any mutations, polymorphisms, or sequencing-detectable RNA editing. This unprecedented apparent gene-specific violation of the genetic code highlights the importance of matching peptides to whole-genome, not known-genes-only, ORFs in mass-spectrometry workflows, and suggests a new mechanism enhancing the combinatorial complexity of the proteome. Funding: NIH Director’s New Innovator Award 1DP2-CA196375 to LL.

Keywords : genetic code, lncRNA, long non-coding RNA, mass spectrometry, proteogenomics, ribo-seq, ribosome, RNAseq
Conference Title : ICRNAB 2018 : International Conference on RNA Biology
Conference Location : Kyoto, Japan
Conference Dates : April 26-27, 2018