

‘Exon’ is not the same as ‘Protein-Coding Region’

Jiuyong Xie*

Department of Physiology & Pathophysiology, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, CANADA

While writing a short essay, I run into the terms ‘exome’ and ‘protein-coding’ regions and was surprised to see how often they are used mistakenly in the industry as well as scientific communities online. Some official websites or even academic papers took for granted or implied exons simply as ‘protein-coding regions’, which is wrong, for example, “exonic (or protein-coding) regions”.

Exons are ‘expressed regions’ versus the ‘intragenic regions’ (introns) of a gene’s DNA (Gilbert, 1978) or precursor messenger RNA (pre-mRNA). Exons of the pre-mRNA transcripts are ligated together with introns removed during splicing, to generate mRNA. The mRNA is typically comprised of a 5’ untranslated region (UTR), protein-coding region, 3’ UTR and an added poly(A) tail. The length of each mRNA region varies from gene to gene or transcript to transcript, and each region including the UTRs could be comprised of multiple exons. Therefore, exons could be part of either UTR or protein-coding regions of mRNA transcripts but are not equal to their ‘protein-coding regions’. Moreover, by definition, exons are specified by the fate of a DNA or pre-mRNA region to be included in the mature mRNA after pre-mRNA processing but ‘protein-coding regions’ are defined by the usage or function of the mRNA regions to be translated into proteins.

The discovery in more recent years of many intron-containing non-coding RNA and peptide-coding so-called ‘non-coding’ RNA warrant even more attention to avoid confusions when the above terms are used.

One instance where these terms are often mistaken is in the description of whole exome sequencing (WES) with ‘exonic’ as ‘protein-coding’ regions. Actually, to what extent each region is represented in number or length depends on the platform used for WES, where UTRs are definitely not ‘protein-coding’ regions of genes. This has led to a proposed change of WES to CES if coding regions are used for WES (Aspden J., 2023). Readers are referred to this Perspective for more details. In other occasions, mRNA or reverse-transcribed complementary DNA have also been mistakenly described as equal to ‘protein-coding’ regions in a whole.

- May 10, 2023, BRC, Monrovia
- Edited Aug. 18, 2023

References

Aspden J., W.E., and Whiffin N. (2023). Not all exons are protein coding: Addressing a common misconception. *Cell Genomics* 3, 100296.
Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501.

*Correspondence Email: xiej@umanitoba.ca