

Surprisingly many peculiar long introns found in brain genes

9 July 2020



In a recent study of genes involved in brain functioning, their previously unknown features have been uncovered by bioinformaticians from the Moscow Institute of Physics and Technology and the Institute of Mathematical Problems of Biology, RAS. Credit: MIPT

In a recent study of genes involved in brain functioning, their previously unknown features have been uncovered by bioinformaticians from the Moscow Institute of Physics and Technology and the Institute of Mathematical Problems of Biology, RAS. The findings are reported in *PLOS One*.

DNA is the molecule that stores information about the structure and functioning of living organisms. This 'book of life' is a carefully arranged nucleotide-by-nucleotide record on every [protein](#) and RNA synthesized in a cell. Each DNA fragment corresponding to a particular protein is called a gene, and the pattern for translating a DNA sequence into the amino acid sequence of the associated protein is known as the [genetic code](#).

Back in the 1960s, biologists discovered the basic properties of the genetic code, including its so-called triplet nature: Each amino acid is encoded by a codon—a sequence of three nucleotides. For

example, the sequence adenine-thymine-guanine encodes the amino acid called methionine, which usually begins the proteins of all living beings at the stage of synthesis.

Since the genetic code was discovered, scientists have learned a lot about [gene structure](#). For example, they found that a kind of fragmentation was characteristic for the genes of eukaryotes—organisms whose cells have a nucleus. Namely, genes contain noncoding regions referred to as introns. They are removed from the sequence in a process called splicing. The remaining regions that actually encode parts of the protein are termed exons.

Researchers have proposed a number of hypotheses as to how long ago and in what way introns originated, and what their functions are. For one thing, introns enable alternative splicing. This refers to the selective joining of certain exons but not others. The consequence is that more than one [protein sequence](#) can be produced based on the template of a single gene. As a result, the number of distinct proteins in cells is far greater than the number of genes.

Another [intron](#)-enabled process important for gene evolution is exon shuffling. This involves a kind of atypical recombination, where a foreign exon can become incorporated into a gene where it does not belong, giving rise to a new gene.

The currently available full-genome sequences of many organisms have made it possible to study the evolution of introns in detail. They are now known to vary in length from several dozen pairs to 10,000 times as many. Introns are also distinguished by phase, depending on where they occur relative to a codon. Phase 0 introns are found in between codons, whereas phases 1 and 2 occur immediately after the first or second nucleotide in the codon, respectively.

Now a team of bioinformaticians from MIPT and IMPB RAS has examined the relation between intron phase and length in humans and mice.

"No one had thought of investigating a potential link between intron length and phase before us. Common sense says there shouldn't be any connection at all, similarly to how a person's height has nothing to do with their eye color," commented Eugene Baulin, a researcher at the Applied Mathematics Lab at IMPB RAS, and the Algorithms and Programming Technologies Department at MIPT.

To their surprise, the study's authors identified a group of genes containing an unusually large number of phase 1 introns that were over 50,000 nucleotide pairs long. Moreover, these turned out to be genes involved in nerve impulse transmission in the brain.

A detailed analysis of numerous scientific publications enabled the team to put the fragments of knowledge together and arrive at a unified understanding. It turned out that in most cases, the phase 1 introns in the group of [genes](#) in question resulted from the presence of a particular amino acid sequence at the beginning of the protein. This so-called [signal peptide](#) serves to direct the protein to where it should perform its function. In the case of nerve cell receptors, that means to the plasma membrane.

As for the introns being fairly long, this also indirectly has to do with the signal peptide. In such proteins, the signal peptide is always located at the beginning of the molecule, and the DNA region encoding it is found at the start of the gene. And it is precisely there, at the beginning of a gene, that long introns tend to occur, because they contain regulatory DNA sequences important for the protein's synthesis.

The study reveals a clear and complete picture of how exon shuffling works and what role long phase 1 introns play in it. "That mechanism speeds up the evolution of intercellular and membrane proteins in animals, particularly the younger ones [evolutionarily speaking], and these are the proteins that enable nerve impulse transmission in brain

cells," Baulin added.

More information: Eugene F. Baulin et al, Brain-related genes are specifically enriched with long phase 1 introns, *PLOS ONE* (2020). [DOI: 10.1371/journal.pone.0233978](https://doi.org/10.1371/journal.pone.0233978)

Provided by Moscow Institute of Physics and Technology

APA citation: Surprisingly many peculiar long introns found in brain genes (2020, July 9) retrieved 8 September 2022 from <https://medicalxpress.com/news/2020-07-surprisingly-peculiar-introns-brain-genes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.