CellPress

Review

Evolution and Biological Roles of Alternative 3′UTRs

Christine Mayr[1,*]

More than half of human genes use alternative cleavage and polyadenylation to generate alternative 3′ untranslated region (3′UTR) isoforms. Most efforts have focused on transcriptome-wide mapping of alternative 3′UTRs and on the question of how 3′UTR isoform ratios may be regulated. However, it remains less clear why alternative 3′UTRs have evolved and what biological roles they play. This review summarizes our current knowledge of the functional roles of alternative 3′UTRs, including mRNA localization, mRNA stability, and translational efficiency. Recent work suggests that alternative 3′UTRs may also enable the formation of protein–protein interactions to regulate protein localization or to diversify protein functions. These recent findings open an exciting research direction for the investigation of new biological roles of alternative 3′UTRs.

Evolution of 3′UTR Length and Function

A few years ago, it was found that a large fraction of genes use alternative cleavage and polyadenylation to generate alternative 3′UTRs [1–5]. Initially, most efforts concentrated on establishing sequencing protocols to map alternative 3′UTRs transcriptome wide [6–13] and elucidating how alternative 3′UTR ratios are regulated [14–24], two topics that have been summarized in several recent reviews [25–28]. Here, the functional roles of alternative 3′UTRs are discussed and reasons for why they have evolved are suggested.

It is largely unknown how the biological complexity of organisms is achieved. Most cellular processes are performed by proteins through interactions with other proteins [29,30]. Therefore, when asked what makes humans different from worms, it was initially surprising that the number of protein-encoding genes and the coding region length have remained fairly constant during evolution from worms to humans [31–34]. Intuitively, higher protein diversity would enable more complex biological functions. However, higher protein diversity within a constrained space also increases the number of nonfunctional interactions [35,36]. Therefore, protein concentration and diversity have to be limited to avoid overcrowding. When only a fixed number of proteins are available, biological complexity can still be accomplished through compartmentalization to avoid the coexistence of too many protein types, by the strengthening of specific interactions through cooperativity [35], and by enabling multifunctionality of existing proteins.

Genome size, and, thus, the noncoding part of the genome, has dramatically increased during evolution from worms to humans. The expansion in noncoding sequences includes the 3′UTRs of mRNAs. The number of genes that produce alternative 3′UTRs has doubled and 3′UTR length has increased from a median of 140 nucleotides (nt) in worms to 1200 nt in humans and even to 2300 nt when examining genes that generate alternative 3′UTRs [6,13]. This suggests that more complex organisms have increased post-transcriptional gene regulation mediated by 3′UTR elements. 3′UTRs are well known to control mRNA stability, translational efficiency, and mRNA localization [2,5,37–42]. In addition, 3′UTRs were recently shown to mediate protein–protein interactions [43]. By facilitating alternative protein complex formation, alternative 3′UTRs can

Trends

During animal evolution, the number of protein-encoding genes has remained fairly constant but 3′ untranslated region (3′UTR) length and the fraction of genes expressing alternative 3′UTRs have increased substantially.

3′UTRs mediate protein–protein interactions. Thus, alternative 3′UTRs facilitate the formation of alternative protein complexes, which can perform alternative protein functions. This diversifies proteome function without a change in amino acid sequence.

About 15–35% of alternative 3′UTRs have significantly different half-lives, which may contribute to the transcriptome diversity of single cells.

Translation rates of mRNAs with alternative 3′UTRs can be differentially affected by signaling. Whereas one isoform generates basal protein levels, translation of the other is induced by signaling.

Long 3′UTRs seem to be bound by many RNA-binding proteins (RBPs) and may exert their functions within RNA granules.

[1]Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

*Correspondence: mayrc@mskcc.org (C. Mayr).

Table 1. Differences between Single and Multi-UTR Genes

	Single-UTR	Multi-UTR
3′UTR length (nt)	625	2323
Transcription unit length (bp)	20 629	40 519
N expressed in one tissue	1637 (23.7%)	630 (11.1%)
N expressed in two to five tissues	2414 (34.9%)	1854 (32.6%)
N expressed in six or seven tissues	2861 (41.4%)	3189 (56.2%)

diversify protein functions. This may increase biological complexity by implementing multi-functionality of existing proteins.

Single and Multi-UTR Genes Represent Two Distinct Classes of Genes

3′-end sequencing methods have revealed that at least half of human genes generate alternative 3′UTR isoforms [9,12,13]. Transcriptome-wide analyses of alternative 3′UTRs across several normal human tissues and cell lines showed the presence of two classes of genes. One class produces mRNAs with only one 3′UTR, whereas the other class generates alternative 3′UTR isoforms [13]. Single- and multi-UTR genes differ in their genomic architecture, their functions, and their mode of regulation of tissue-specific expression (Table 1).

The median 3′UTR length of single-UTR genes is ~600 nt, but it is ~2300 nt for multi-UTR genes. The increase in 3′UTR length is associated with longer transcription units, which are about twice as long in multi-UTR genes (Table 1). The tissue-specific expression patterns of the two classes of genes also differ significantly. Whereas most genes expressed in only one tissue have single 3′UTRs, more than half of ubiquitously transcribed genes are multi-UTR genes (Table 1) [13]. Furthermore, 3′UTR length is tissue dependent: single- or multi-UTR genes specifically expressed in testis or liver have much shorter 3′UTRs than genes expressed only in the brain (Figure 1) [44–46]. However, ubiquitously transcribed multi-UTR genes have the longest 3′UTRs. They are even longer than the 3′UTRs of brain-specific genes and are about

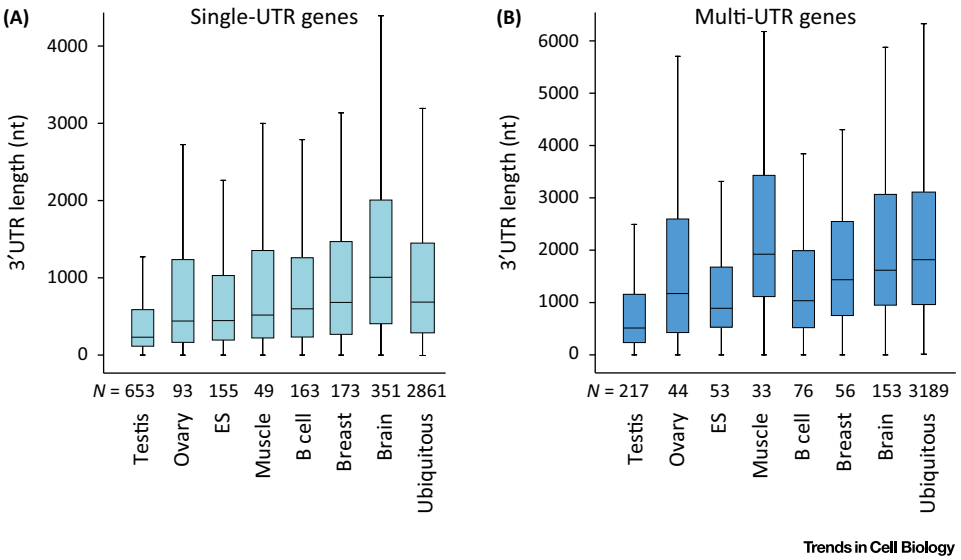

Figure 1. 3′ Untranslated Region (3′UTR) Length of Ubiquitously Transcribed and Tissue-Restricted Genes. (A) Single-UTR genes. Here, ubiquitously expressed genes are defined to be expressed in at least six of seven human tissues comprising testis, ovary, embryonic stem (ES) cells, B cells, muscle, breast, and brain. *N*, number of genes in each category. (B) Multi-UTR genes. As in (A).