**ABSTRACT**

One of the features of intron-containing genes of the nxf (nuclear export factor) family in different organisms is the presence of an evolutionarily conserved exon-intron block: exon 110nt-intron-exon 37nt. The intron in this evolutionarily conserved block, which we call a “cassette” intron, can be excised or retained in alternative transcripts of nxf1. It corresponds to intron 10 - 11 in the genes that are orthologous to nxf1 in vertebrates, and intron 5 - 6 in the genes that are orthologous to nxf1 in Drosophilidae. The alignment of sequences of cassette introns in nxf1 genes in vertebrates has revealed four evolutionarily conserved sequences: 1) 5’ flanking sequence, 2) a region containing СТЕ (constitutive transport element), 3) third conserved sequence, and 4) 3’ flanking sequence. Introns 5 - 6 of nxf1 in Drosophilidae have no similar conserved sequences. The results of sequence alignment demonstrate a similarity between cassette introns of nxf1 in Drosophilidae in two poly(A) sequences. The prevalence of Dm nxf1 transcripts containing cassette intron 5 - 6 under completely spliced transcripts in the heads of adult Drosophila melanogaster suggests a functional importance of transcripts that contain a retained intron. Evolutionary conservation, which in Drosophilidae is evident in the presence of poly(A) sequences in cassette introns of the nxf1 genes, is an adaptive feature: the poly(A) sequences are capable of mimicking the 3’-end of transcripts, promote transport from the nucleus to the cytoplasm, or are involved in NMD control. The ability to form characteristic secondary structures is a common feature of nxf1 cassette introns.

Одной из особенностей интронсодержащих генов семейства nxf (nuclear export factor) у разных организмов является наличие эволюционно консервативного экзон-интронного блока: экзон 110 нт-интрон-экзон 37 нт. Интрон в этом эволюционно консервативном блоке, который мы называем «кассетным» интроном, может быть вырезан или сохранен в альтернативных транскриптах nxf1. Он соответствует интрону 10–11 в генах, ортологичных nxf1 у позвоночных, и интрону 5–6 в генах, ортологичных nxf1 у дрозофилид. Выравнивание последовательностей кассетных интронов в генах nxf1 у позвоночных выявило четыре эволюционно консервативные последовательности: 1) 5'-фланкирующая последовательность, 2) область, содержащая СТЭ (конститутивный транспортный элемент), 3) третья консервативная последовательность и 4) 3'-фланкирующая последовательность. последовательность. Интроны 5-6 nxf1 у Drosophilidae не имеют подобных консервативных последовательностей. Результаты выравнивания последовательностей демонстрируют сходство кассетных интронов nxf1 у дрозофилид в двух поли(А)-последовательностях. Преобладание транскриптов Dm nxf1, содержащих кассетный интрон 5-6, среди полностью сплайсированных транскриптов в головах взрослых Drosophila melanogaster указывает на функциональную важность транскриптов, содержащих сохраненный интрон. Эволюционная консервативность, проявляющаяся у дрозофилид в присутствии поли(А)-последовательностей в интронах кассет генов nxf1, является адаптивным признаком: поли(А)-последовательности способны имитировать 3'-конец транскриптов, способствовать транспорту из ядра в цитоплазму или участвуют в контроле NMD. Способность образовывать характерные вторичные структуры является общей чертой интронов кассеты nxf1.

**1. INTRODUCTION**

**1.1. The nxf Gene Family and Functions of the nxf1 Gene**

The nxf (nuclear export factor) gene family was named after the function of the universal gene nxf1 responsible for the nuclear-cytoplasmic transport of most mRNAs [1-3]. This path of mRNA export is RanGTP-independent [4]. Several proteins involved in mRNA export pathways are less conserved and presumably appeared later in the evolution of eukaryotes [5]. The blocking of the transport path enabled by the NXF1 protein results in the accumulation of polyadenylated RNAs within the nucleus [4,6,7].

Семейство генов nxf (nuclear export factor) было названо в честь функции универсального гена nxf1, ответственного за ядерно-цитоплазматический транспорт большинства мРНК [1-3]. Этот путь экспорта мРНК не зависит от RanGTP [4]. Некоторые белки, участвующие в путях экспорта мРНК, менее консервативны и предположительно появились позже в эволюции эукариот [5]. Блокирование транспортного пути, обеспечиваемого белком NXF1, приводит к накоплению полиаденилированных РНК внутри ядра [4,6,7].

Genes of the nxf family have been found in eukaryotic organisms of the Opisthokonta group, and are characterized by evolutionary conservation [1,5,8]. Genomes of various fungi have only one gene, Mex67, which belongs to this family; animals usually have two to five paralogous genes (see Table 1) [1,9-11]. Plants and some protozoa lack genes of the nxf family [5]. nxf1 genes are the most evolutionarily conserved in the nxf family. What is even more interesting is that nxf1 genes across different species of mammals exhibit a much greater degree of similarity than Mm nxf1 and Mm nxf2 or Hs nxf1 and Hs nxf2 exhibit between themselves (Figure 1).

Гены семейства nxf обнаружены у эукариотических организмов группы Opisthokonta и характеризуются эволюционной консервативностью [1,5,8]. Геномы различных грибов имеют только один ген Mex67, принадлежащий к этому семейству; животные обычно имеют от двух до пяти паралогичных генов (см. табл. 1) [1,9-11]. У растений и некоторых простейших отсутствуют гены семейства nxf [5]. Гены nxf1 являются наиболее эволюционно консервативными в семействе nxf. Что еще более интересно, так это то, что гены nxf1 у разных видов млекопитающих проявляют гораздо большую степень сходства, чем Mm nxf1 и Mm nxf2 или Hs nxf1 и Hs nxf2 обнаруживают между собой (Рис. 1).

In the S. cerevisiae yeast, the factor Mex67 (mRNA EXport factor of 67 kDa), which is orthologous to other NXF1 proteins of the eukaryotes of the Opisthokonta group, is also involved in the nuclear-cytoplasmic transport of ribosomal RNAs [5,12,13].

У дрожжей S. cerevisiae в ядерно-цитоплазматическом транспорте рибосомных РНК также участвует фактор Mex67 (mRNA EXport factor 67 кДа), ортологичный другим белкам NXF1 эукариот группы Opisthokonta [5,12,13].

Initially, the NXF1 protein in humans was identified as a potential cytoplasmic cofactor for Tip (tyrosine kinase interacting protein) encoded by the herpesvirus saimiri, and was named TAP (Tip-associated protein) [14]. Later, TAP was demonstrated to be involved in the nuclear-cytoplasmic transport of the unspliced or partially spliced RNA of retroviruses. TAP directly recognizes only one sequence—CTE (Constitutive Transport Element)—which was initially discovered in RNAs of retroviruses [15-18]. Adaptor proteins mediate the interaction between NXF1 and cellular mRNAs in metazoan [19-21]. Nuclear mRNA export is connected with transcription, splicing, processing, and mRNA quality control [12].

Первоначально белок NXF1 у людей был идентифицирован как потенциальный цитоплазматический кофактор для Tip (белок, взаимодействующий с тирозинкиназой), кодируемый вирусом герпеса saimiri, и был назван TAP (Tip-associated protein) [14]. Позже было показано, что ТАР участвует в ядерно-цитоплазматическом транспорте несплайсированной или частично сплайсированной РНК ретровирусов. TAP напрямую распознает только одну последовательность — CTE (Constitutive Transport Element), которая была первоначально обнаружена в РНК ретровирусов [15–18]. Белки-адаптеры опосредуют взаимодействие между NXF1 и клеточными мРНК многоклеточных животных [19-21]. Экспорт ядерной мРНК связан с транскрипцией, сплайсингом, процессингом и контролем качества мРНК [12].

**1.2. Modular Principle of Organization of NXF Factors**

Proteins of the NXF family have a modular domain organization consisting of an RNA-binding domain (RBD), four leucine-rich repeats (LRRs), a domain exhibiting a similarity to the nuclear transport factor 2 (NTF2-like domain), and a C-terminal ubiquitin associated (UBA)-like domain (Figure 2) [22,23]. These proteins are considered RNA transport receptors due to the combination of their receptor and transport functions. The N-terminus of the protein is predominantly responsible for interacting with mRNA, while the C-terminus enables transport of the RNP complex through nuclear pores by interacting with partner protein p15 and nucleoporins—proteins of nuclear pore complexes [24,25]. The RBD (RNA-binding domain) belongs to the RRM (RNA recognition motif) family, which has a distinctive βαββαβ structure, characteristic of many RNA-binding proteins, such as РАВ—poly(A)-binding-protein, Sex-lethal protein (Sxl), and U1A and U2B″ splicing factors [16]. The receptor function is much older: RRM have been discovered in proteins of prokaryotes [26]. While the RRM domain alone is sufficient for binding with mRNA in vitro, both RRM and LRR are necessary for exporting mRNA in vivo [16,23].

Белки семейства NXF имеют модульную доменную организацию, состоящую из РНК-связывающего домена (RBD), четырех богатых лейцином повторов (LRR), домена, проявляющего сходство с ядерным транспортным фактором 2 (NTF2-подобный домен), и С-концевой убиквитин-ассоциированный (UBA)-подобный домен (рис. 2) [22,23]. Эти белки считаются рецепторами транспорта РНК из-за сочетания их рецепторных и транспортных функций. N-конец белка преимущественно отвечает за взаимодействие с мРНК, а C-конец обеспечивает транспорт комплекса RNP через ядерные поры за счет взаимодействия с белком-партнером p15 и нуклеопоринами — белками комплексов ядерных пор [24,25]. RBD (РНК-связывающий домен) принадлежит к семейству RRM (РНК-узнаваемый мотив), которое имеет характерную структуру βαββαβ, характерную для многих РНК-связывающих белков, таких как РАВ — поли(А)-связывающий белок, секс-летальный белок (Sxl) и факторы сплайсинга U1A и U2B″ [16]. Рецепторная функция гораздо старше: RRM обнаружены в белках прокариот [26]. В то время как одного домена RRM достаточно для связывания с мРНК in vitro, RRM и LRR необходимы для экспорта мРНК in vivo [16,23].

The C-terminal end of NXF proteins is represented by NTF2-like and UBA-like domains. The NTF2-like domain is thought to be derived from a prokaryotic precursor [27]. The C-terminal end in proteins paralogous to NXF often differs from the corresponding parts in NXF1 proteins. Consequently, many paralogous proteins, such as Dm NXF3, Dm NXF4, Ce NXF2, Hs NXF3, Hs NXF5, Mm NXF3, and Mm NXF7 are incapable of binding with nucleoporins [1,8,11,28-30]. In most cases this is due to the loss of a UBA-like domain in paralogous proteins, or the absence of both domains (as in the case of the Dm NXF4 protein) [1,7,31].

С-конец белков NXF представлен NTF2-подобным и UBA-подобным доменами. Предполагается, что NTF2-подобный домен происходит от прокариотического предшественника [27]. С-конец в белках, паралогичных NXF, часто отличается от соответствующих частей в белках NXF1. Следовательно, многие паралогичные белки, такие как Dm NXF3, Dm NXF4, Ce NXF2, Hs NXF3, Hs NXF5, Mm NXF3 и Mm NXF7, не способны связываться с нуклеопоринами [1,8,11,28–30]. В большинстве случаев это связано с потерей UBA-подобного домена в паралогичных белках или отсутствием обоих доменов (как в случае белка Dm NXF4) [1,7,31].

The unification of the transport and receptor functions probably took place simultaneously with the origin of the eukaryotic cell. In eukaryotes, transcription and mRNA processing are physically and temporally separated from translation by the nuclear membrane, which required the formation of mechanisms for the active transport of macromolecules, including mRNAs, from the nucleus to the cytoplasm. Most transcripts begin translation soon after they exit the nucleus. The specific class of transcripts capable of staying for extended periods of time in the cytoplasm in a state unavailable for translation presumably required specialization of transport receptors. There are paralogous genes that, unlike the universal gene nxf1, usually exhibit an organ-specific character of expression in mammals [1,8]. The products of paralogous genes are usually localized in the cytoplasm [8,11, 28,30,32].

Объединение транспортной и рецепторной функций, вероятно, произошло одновременно с возникновением эукариотической клетки. У эукариот транскрипция и процессинг мРНК физически и во времени отделены от трансляции ядерной мембраной, что потребовало формирования механизмов активного транспорта макромолекул, в том числе мРНК, из ядра в цитоплазму. Большинство транскриптов начинают трансляцию вскоре после выхода из ядра. Специфический класс транскриптов, способных длительное время находиться в цитоплазме в состоянии, недоступном для трансляции, по-видимому, обусловлен специализацией транспортных рецепторов. Существуют паралогичные гены, которые, в отличие от универсального гена nxf1, обычно проявляют органоспецифический характер экспрессии у млекопитающих [1,8]. Продукты паралогичных генов обычно локализованы в цитоплазме [8,11, 28,30,32].

Several processes of transformation of genetic material underlie the formation of gene families: duplication, separation of certain genes, or fusion of different parts of genes [33-36]. The existence of clusters of linked genes in different organisms seems to support the hypothesis that the nxf gene family was formed via duplication [1,8].

Integration of cDNA copies of transcripts into a genome may also produce gene copies that contain no introns [36-38]. Perhaps this mechanism eventually produced genes Dm nxf3 and Dm nxf4, which contain no introns in D. melanogaster. It is important to note, however, that the lack of introns in genes that are orthologous to Dm nxf4 is not characteristic of all species of the genus Drosophila for which it has been described (FlyBase, 2012).

LLR and NTF2-like domains responsible for the interaction of NXFs with various partner proteins are the most highly conserved [39].

В основе образования генных семейств лежит несколько процессов трансформации генетического материала: дупликация, разделение определенных генов или слияние разных частей генов [33-36]. Существование кластеров сцепленных генов у разных организмов, по-видимому, подтверждает гипотезу о том, что семейство генов nxf образовалось путем дупликации [1,8].

Интеграция кДНК-копий транскриптов в геном может также давать копии генов, не содержащие интронов [36-38]. Возможно, этот механизм в конечном итоге привел к образованию генов Dm nxf3 и Dm nxf4, не содержащих интронов у D. melanogaster. Однако важно отметить, что отсутствие интронов в генах, ортологичных Dm nxf4, характерно не для всех видов рода Drosophila, для которых оно описано (FlyBase, 2012).

Наиболее консервативны LLR и NTF2-подобные домены, ответственные за взаимодействие NXF с различными белками-партнерами [39].

**1.3. Interaction of NXF1 Proteins with RNA**

Nuclear mRNAs are exported as RNP-complexes [40]. NXF1 interacts with mRNA by means of adaptor proteins, which are part of mRNP complexes during transcription, processing, and transport [25,41-43]. During mRNA export, NXF1 partners with proteins E1B-AP5, RAE1, and members of evolutionarily conserved families of proteins—Yra1p in yeast and REF in mammals [24,44,45], as well as components of the TREX (TRanscription-EXport) complex [40,43].

Ядерные мРНК экспортируются в виде RNP-комплексов [40]. NXF1 взаимодействует с мРНК с помощью адапторных белков, которые входят в состав комплексов мРНП при транскрипции, процессинге и транспорте [25,41-43]. Во время экспорта мРНК NXF1 сотрудничает с белками E1B-AP5, RAE1 и членами эволюционно консервативных семейств белков — Yra1p у дрожжей и REF у млекопитающих [24,44,45], а также с компонентами TREX (TRanscription-EXport) комплекс [40,43].

Most genes of eukaryotes have introns that are then deleted from pre-mRNA during splicing. As a rule, cellular mRNAs with introns do not leave the nucleus [16,46]. Translation of incompletely spliced mRNA due to the presence of a premature termination codon may produce truncated proteins, which are potentially deleterious to a cell. In the nucleus, transcripts undergo quality control, and normally only completely spliced mRNAs are capable of export [47,48].

Большинство генов эукариот имеют интроны, которые удаляются из пре-мРНК во время сплайсинга. Как правило, клеточные мРНК с интронами не покидают ядра [16,46]. Трансляция неполностью сплайсированной мРНК из-за наличия кодона преждевременной терминации может привести к образованию укороченных белков, потенциально вредных для клетки. В ядре транскрипты проходят контроль качества, и в норме только полностью сплайсированные мРНК способны к экспорту [47,48].

Genes that correspond to transcripts with intron retention are not uncommon in humans [49]. The expression of intron-containing messages has been shown to occur in a variety of diseases including several cancers [50], and also as a response to vascular injury in rats [51]. Transcripts with retained introns may serve as sources of alternative protein products with an independent function that may, among other things, influence the function of a full-length product [52].

Гены, соответствующие транскриптам с задержкой интрона, не редкость у человека [49]. Было показано, что экспрессия интронсодержащих сообщений происходит при различных заболеваниях, включая несколько видов рака [50], а также в ответ на повреждение сосудов у крыс [51]. Транскрипты с сохраненными интронами могут служить источниками альтернативных белковых продуктов с независимой функцией, которая может, среди прочего, влиять на функцию полноразмерного продукта [52].

If mRNAs with retained introns are abundant in the cytoplasm, intron retention is probably regulated by factors involved in both splicing and mRNA export. Some of the retroviruses have been shown to export unspliced RNA by means of cis-acting RNA elements, termed constitutive transport elements (CTEs), which interact directly with cellular export proteins [53]. Export of mRNA with retained introns in simple retroviruses is carried out with the help of NXF1 (TAP), which binds directly to the CTE sequence in the genome of retroviruses [15-18,24,53,54]. Microinjection experiments performed on Xenopus oocytes have demonstrated that TAP (Hs NXF1) directly interacts with the CTE, allowing the export of CTE-containing RNAs [15]. Moreover, TAP remains bound to CTE-containing RNAs in polyribosomes and may be present inside the nucleus or in the nuclear rim, as well as in the cytoplasm [55].

Если в цитоплазме много мРНК с сохраненными интронами, задержка интронов, вероятно, регулируется факторами, участвующими как в сплайсинге, так и в экспорте мРНК. Было показано, что некоторые из ретровирусов экспортируют несплайсированную РНК с помощью цис-действующих элементов РНК, называемых конститутивными транспортными элементами (CTE), которые непосредственно взаимодействуют с клеточными экспортными белками [53]. Экспорт мРНК с сохраненными интронами у простых ретровирусов осуществляется с помощью NXF1 (TAP), который напрямую связывается с последовательностью СТЕ в геноме ретровирусов [15-18,24,53,54]. Эксперименты с микроинъекциями, проведенные на ооцитах Xenopus, показали, что TAP (Hs NXF1) напрямую взаимодействует с CTE, позволяя экспортировать CTE-содержащие РНК [15]. Более того, TAP остается связанным с CTE-содержащими РНК в полирибосомах и может присутствовать внутри ядра или в ядерном ободе, а также в цитоплазме [55].

**2. RESULTS AND DISCUSSIONS**

**2.1. Alternative Intron-Retaining Transcripts of nxf1 Genes**

Both the domain structure and the intron-exon structure of the NXF family proteins are evolutionarily conserved. Among the known nxf genes, most have an intron-exon structure. Mex67 in fungi, nxf4 in some species of the genus Drosophila, and Dm nxf3 in D. melanogaster, have no introns.

Как доменная структура, так и интрон-экзонная структура белков семейства NXF эволюционно консервативны. Среди известных генов nxf большинство имеют интронно-экзонную структуру. Mex67 у грибов, nxf4 у некоторых видов рода Drosophila и Dm nxf3 у D. melanogaster не имеют интронов.

A feature specific to nxf1 genes is the existence of alternative transcripts with a retained intron flanked by evolutionarily conserved exons: 110 nt upstream and 37 nt downstream of the respective intron. This intron from the evolutionarily conserved block, which can be excised or retained in alternative transcripts of nxf1, we named a “cassette” intron. Such an evolutionarily conserved exonintron block is characteristic of most of the nxf family genes that exhibit an intron-exon structure. There exist nxf genes in which the sequences of exons 110 nt and 37 nt are not separated by an intron, and are represented by exon 147 nt: Dm nxf2 and Ce nxf2, for example. Transcripts of nxf1 genes with a cassette intron between exons 110 nt and 37 nt have been shown for nxf1 genes in M. musculus [8], H. sapiens [56], and D. melanogaster [57]. Such nxf1 transcripts have been found in other species, as well. ESTs (expressed sequence tags) include a portion of one of the aforementioned exons and a portion of the cassette intron (Genbank, UCSC) (see Table 1).

Характерной особенностью генов nxf1 является существование альтернативных транскриптов с сохраненным интроном, фланкированным эволюционно консервативными экзонами: 110 н. выше и 37 н. ниже соответствующего интрона. Этот интрон из эволюционно консервативного блока, который может быть вырезан или сохранен в альтернативных транскриптах nxf1, мы назвали «кассетным» интроном. Такой эволюционно консервативный экзонинтронный блок характерен для большинства генов семейства nxf, обладающих интрон-экзонной структурой. Существуют гены nxf, в которых последовательности экзонов 110 н. и 37 н. не разделены интроном, а представлены экзоном 147 н., например, Dm nxf2 и Ce nxf2. Транскрипты генов nxf1 с кассетным интроном между экзонами 110 и 37 н. показаны для генов nxf1 у M. musculus [8], H. sapiens [56] и D. melanogaster [57]. Такие транскрипты nxf1 были обнаружены и у других видов. EST (теги экспрессированной последовательности) включают часть одного из вышеупомянутых экзонов и часть кассетного интрона (Genbank, UCSC) (см. Таблицу 1).

We have demonstrated that the transcript with cassette intron 5 - 6 in the head tissues of adult fruit flies exceeds the universal, completely spliced transcript of the gene Dm nxf1, in its relative content (Figure 3) [57]. Usually, intron-containing transcripts stay in the nucleus [46,48]. Even if they enter the cytoplasm, they are subject to the cytoplasmic mRNA quality control mechanism (surveillance) termed NMD (nonsense-mediated mRNA decay) [58,59]. The large number of transcripts of nxf1 genes with a retained intron in different organisms raises the following questions:

1) How do intron-containing transcripts manage to exit the nucleus, bypassing the mRNA quality control mechanism?

2) What makes the intron-containing transcripts immune to NMD in the cytoplasm?

3) What evolutionarily conserved features characterize cassette introns in alternative intron-containing transcripts of nxf1 genes?

Нами показано, что транскрипт с кассетным интроном 5-6 в тканях головы взрослых плодовых мушек превосходит по относительному содержанию универсальный, полностью сплайсированный транскрипт гена Dm nxf1 (рис. 3) [57]. Обычно интронсодержащие транскрипты остаются в ядре [46,48]. Даже если они попадают в цитоплазму, они подвергаются механизму контроля качества цитоплазматической мРНК (надзору), называемому NMD (бессмысленно-опосредованный распад мРНК) [58,59]. Большое количество транскриптов генов nxf1 с сохраненным интроном у разных организмов ставит следующие вопросы:

1) Каким образом интронсодержащим транскриптам удается покинуть ядро, минуя механизм контроля качества мРНК?

2) Что делает интронсодержащие транскрипты невосприимчивыми к NMD в цитоплазме?

3) Какие эволюционно консервативные признаки характеризуют кассетные интроны в альтернативных интронсодержащих транскриптах генов nxf1?

Analyzing specific features of the intron retained in alternative transcripts of nxf1 genes may help answer these questions. The existence of a sequence with a length of about 100 nt, which resembles the CTE of the MPMV virus in cassette introns 10 - 11, represented in genes Hs nxf1 and Mm nxf1, has been observed [56]. This sequence is highly conserved in cassette introns of nxf1 in various animals (Figure 4). Hs nxf1 mRNA with the retained intron is exported from the nucleus and is represented in a polysomic fraction [56]. Due to the presence of a premature stop-codon at the beginning of the cassette intron 10 - 11, translation results in a truncated protein. A similar truncated protein has been discovered in human cells [56]. The CTE sequence is an important element affecting the expression of intron-retaining mRNAs in mammals [60,61]. The CTE has not been discovered in the corresponding introns of paralogous genes of the nxf family in mice and humans.

Ответить на эти вопросы может помочь анализ особенностей интрона, сохраняющегося в альтернативных транскриптах генов nxf1. Обнаружено существование последовательности длиной около 100 н., напоминающей CTE вируса MPMV в кассетных интронах 10–11, представленной в генах Hs nxf1 и Mm nxf1 [56]. Эта последовательность высоко консервативна в кассетных интронах nxf1 у различных животных (рис. 4). мРНК Hs nxf1 с сохранившимся интроном экспортируется из ядра и представлена полисомной фракцией [56]. Из-за наличия преждевременного стоп-кодона в начале кассетного интрона 10-11 трансляция приводит к укороченному белку. Аналогичный укороченный белок был обнаружен в клетках человека [56]. Последовательность CTE является важным элементом, влияющим на экспрессию интрон-сохраняющих мРНК у млекопитающих [60,61]. CTE не обнаружен в соответствующих интронах паралогичных генов семейства nxf у мышей и человека.

When comparing introns 5 - 6 of nxf1 genes in different species of Drosophila, we have not discovered extended homologous sequences (such as cassette introns in genes orthologous to the nxf1 of vertebrates). However, we have found a common feature of introns 5 - 6 of the nxf1 gene in different species of Drosophila: the presence of two poly(A) sequences, each with a length of around 100 nt (Figure 5). This feature accounts for the difference of cassette introns 5 - 6 of nxf1 genes in Drosophilidae from cassette introns 10 - 11 of nxf1 genes in vertebrates.

При сравнении интронов 5-6 генов nxf1 у разных видов дрозофил мы не обнаружили протяженных гомологичных последовательностей (таких как кассетные интроны в генах, ортологичных nxf1 позвоночных). Однако мы обнаружили общую черту интронов 5-6 гена nxf1 у разных видов дрозофилы: наличие двух поли(А)-последовательностей, каждая длиной около 100 н. (рис. 5). Эта особенность объясняет отличие кассетных интронов 5-6 генов nxf1 у Drosophilidae от кассетных интронов 10-11 генов nxf1 позвоночных.

The cassette introns of nxf1 in Drosophilidae form complex secondary structures. Poly(A) sequences of introns 5 - 6 of nxf1 in Drosophilidae is usually located in loops (Figure 6). Because the same RNA sequence can form several secondary structure variants, Figure 6 depicts only the most probable structures of introns 5 - 6 of nxf1 genes in some Drosophila species. The choice of species was made in accordance with the relationships between Drosophila species, taking into consideration the phylogenetic tree constructed with the divergence of sequences of intron 5 - 6 (Figure 7).

Кассетные интроны nxf1 у Drosophilidae образуют сложные вторичные структуры. Поли(А) последовательности интронов 5-6 nxf1 у дрозофилид обычно располагаются в петлях (рис. 6). Поскольку одна и та же последовательность РНК может образовывать несколько вариантов вторичной структуры, на рис. 6 показаны только наиболее вероятные структуры интронов 5-6 генов nxf1 у некоторых видов дрозофилы. Выбор видов был сделан в соответствии с родственными отношениями между видами дрозофилы с учетом филогенетического дерева, построенного с дивергенцией последовательностей интронов 5-6 (рис. 7).

What happens to the poly(A) RNA sequence? The presence of a poly(A) tail is an important element in RNA export [62]. Export efficiency depends on the length of this sequence. The poly(A) sequence is responsible for binding with corresponding proteins [63]. The poly(A) sequence is believed to be a recognition target for export factors [62].

Что происходит с последовательностью поли(А)РНК? Наличие поли(А)-хвоста является важным элементом экспорта РНК [62]. Эффективность экспорта зависит от длины этой последовательности. Последовательность поли(А) отвечает за связывание с соответствующими белками [63]. Считается, что поли(А)-последовательность является мишенью для распознавания экспортных факторов [62].

Taking into account the secondary structure of cassette introns of nxf1 genes in Drosophilidae, poly(A) sequences are probably open for interaction with proteins that may recognize them. Poly(A) sequences may not only serve as nuclear export markers, but may also protect the corresponding transcript from degradation in the cytoplasm. It has been demonstrated that the protein PABPC1 (cytoplasmic poly(A)-binding protein) suppresses NMD in D. melanogaster [64].

Учитывая вторичную структуру кассетных интронов генов nxf1 у дрозофилид, поли(А)-последовательности, вероятно, открыты для взаимодействия с белками, которые могут их узнавать. Последовательности поли(А) могут не только служить маркерами ядерного экспорта, но и защищать соответствующий транскрипт от деградации в цитоплазме. Показано, что белок PABPC1 (цитоплазматический поли(А)-связывающий белок) подавляет NMD у D. melanogaster [64].

**2.2. Specific Features of Retained Introns in Alternative Transcripts of nxf Genes**

Cassette introns of nxf1 genes in different species of mammals form complex secondary structures, as well as cassette introns in Drosophilidae nxf1 (Figure 8). It is possible that this feature of cassette introns determines the fate of intron-retaining transcripts in the cytoplasm. It is also possible that cassette introns of nxf1 genes have an independent function both inside the intron-retained transcript, and as the product of splicing of the premRNA of the nxf1. RNAs transcribed from introns are known to participate in a number of processes related to post-transcriptional control of gene expression [65].

Кассетные интроны генов nxf1 у разных видов млекопитающих образуют сложные вторичные структуры, как и кассетные интроны nxf1 у Drosophilidae (рис. 8). Возможно, эта особенность кассетных интронов определяет судьбу интрон-удерживающих транскриптов в цитоплазме. Возможно также, что кассетные интроны генов nxf1 выполняют самостоятельную функцию как внутри удерживаемого интроном транскрипта, так и как продукт сплайсинга премРНК nxf1. Известно, что РНК, транскрибируемые с интронов, участвуют в ряде процессов, связанных с посттранскрипционным контролем экспрессии генов [65].

Cassette introns of nxf1 genes in vertebrates have four evolutionarily conserved regions (Figure 4). The first, with a length of around 130 nt, is the 5’-end of cassette introns (Figure 4). This sequence includes 17 complete codons that continue the open reading frame of the preceding exon 10. The last 17 amino acids in the short protein that corresponds to the Hs nxf1 transcript with retained intron 10 - 11 [56] is evolutionarily conserved in truncated proteins corresponding to the intron-containing transcripts in other vertebrates (Figure 9). The first stop-codon in intron 10 - 11 is duplicated by the next one, 12 nt down. Evolutionary conservation of the first region (Figure 4(a)) suggests the functional significance of the truncated NXF1. The only species that is not present in the list is the M. musculus. Deleting one of the five nucleotides (C) at the position 40 - 44 of Mm nxf1 intron shifts the open reading frame compared to other species, and the C-terminus of the supposed truncated protein corresponds to the 80th full codon in intron 10 - 11 of Mm nxf1.

Кассетные интроны генов nxf1 у позвоночных имеют четыре эволюционно консервативных участка (рис. 4). Первый, длиной около 130 нуклеотидов, представляет собой 5'-конец кассетных интронов (рис. 4). Эта последовательность включает 17 полных кодонов, продолжающих открытую рамку считывания предыдущего экзона 10. Последние 17 аминокислот в коротком белке, соответствующем транскрипту Hs nxf1 с сохраненным интроном 10–11 [56], эволюционно консервативны в укороченных белках, соответствующих к интрон-содержащим транскриптам у других позвоночных (рис. 9). Первый стоп-кодон в интронах 10-11 дублируется следующим на 12 нуклеотидов ниже. Эволюционная консервация первой области (рис. 4(а)) указывает на функциональное значение усеченного NXF1. Единственный вид, которого нет в списке, — это M. musculus. Удаление одного из пяти нуклеотидов (С) в положении 40-44 интрона Mm nxf1 сдвигает открытую рамку считывания по сравнению с другими видами, а С-конец предполагаемого укороченного белка соответствует 80-му полному кодону в интроне 10-11. ход nxf1.

The first conserved region of cassette introns (Figure 4(a)) is partially complimentary to the fourth conserved region (Figure 4(d)), which is at the 3’-end of the intron. In secondary RNA structures the first and fourth regions form a “stem”, thus “closing” the secondary structure of cassette introns 10 - 11 (Figure 8).

Первая консервативная область кассетных интронов (рис. 4(а)) частично комплементарна четвертой консервативной области (рис. 4(г)), которая находится на 3’-конце интрона. Во вторичных структурах РНК первый и четвертый участки образуют «стебель», «замыкая» таким образом вторичную структуру кассетных интронов 10–11 (рис. 8).

The second evolutionarily conserved sequence (Figure 4(b)) contains CTE and has been identified by Li with co-authors [56]. Whereas this sequence does not contain an ORF (open reading frame), the third evolutionarily conserved sequence (Figure 4(c)) in introns 10 - 11 of the genes Hs nxf1 and Mm nxf1 is a long open reading frame.

Вторая эволюционно консервативная последовательность (рис. 4(b)) содержит CTE и была идентифицирована Li с соавторами [56]. Поскольку эта последовательность не содержит ОРС (открытой рамки считывания), третья эволюционно консервативная последовательность (рис. 4(с)) в интронах 10-11 генов Hs nxf1 и Mm nxf1 представляет собой длинную открытую рамку считывания.

The supposed proteins, which correspond to the open reading frame in the third evolutionarily conserved sequence of introns 10 - 11 of the genes Hs nxf1 and Mm nxf1, exhibit a great degree of similarity (data not provided). If the open reading frame is significant, it may be preserved in other cassette introns of nxf1 genes in vertebrates. This, however, we did not observe when evaluating the third evolutionarily conserved region of nxf1 genes in other vertebrates for the presence of a long evolutionarily conserved ORF. In all the depicted secondary structures of cassette introns, region three is a hairpin (Figure 8), which supports the assumption of a structural role of this evolutionarily conserved region.

Предполагаемые белки, соответствующие открытой рамке считывания в третьей эволюционно консервативной последовательности интронов 10–11 генов Hs nxf1 и Mm nxf1, обнаруживают большую степень сходства (данные не приводятся). Если открытая рамка считывания значительна, она может сохраняться в интронах других кассет генов nxf1 у позвоночных. Этого, однако, мы не наблюдали при оценке третьей эволюционно консервативной области генов nxf1 у других позвоночных на наличие длительной эволюционно консервативной ORF. Во всех изображенных вторичных структурах кассетных интронов область 3 представляет собой шпильку (рис. 8), что подтверждает предположение о структурной роли этой эволюционно консервативной области.

Long non-coding RNAs often contain ORF. It is speculated that such sequences can be translated, but with very low efficiency, or only at a specific stage of development [66]. Little is known about ncRNAs that code for functionally significant short proteins [67]. Potentially, the possibility for synthesizing short peptides also exists for cassette introns of nxf1 genes. Short peptides, according to some researchers [68], may be a new class of bioactive signaling molecules.

Длинные некодирующие РНК часто содержат ORF. Предполагается, что такие последовательности могут транслироваться, но с очень низкой эффективностью или только на определенной стадии развития [66]. Мало что известно о нкРНК, кодирующих функционально значимые короткие белки [67]. Потенциально возможность синтеза коротких пептидов существует и для кассетных интронов генов nxf1. Короткие пептиды, по мнению некоторых исследователей [68], могут быть новым классом биоактивных сигнальных молекул.

Many transcripts, including those that code for a protein, also carry out regulatory functions. They are capable of interactions via a specific nucleotide sequence, thus playing a structural role or serving as catalysts [66,69]. Evolutionarily conserved sequences in the cassette intron of nxf1 genes raise the questions of what adaptive advantages these sequences bring, and what the functional significance of intron-retained transcripts is. The evolutionary path of forming adaptive features may vary depending on the characteristics of a particular taxon.

Многие транскрипты, в том числе кодирующие белок, также выполняют регуляторные функции. Они способны взаимодействовать через определенную последовательность нуклеотидов, таким образом, играя структурную роль или выступая в качестве катализаторов [66,69]. Эволюционно консервативные последовательности в кассетном интроне генов nxf1 поднимают вопрос о том, какие адаптивные преимущества приносят эти последовательности и каково функциональное значение сохраненных в интроне транскриптов. Эволюционный путь формирования адаптивных признаков может различаться в зависимости от особенностей того или иного таксона.

The existence of sequences that may facilitate export of transcripts containing these introns from the nucleus to the cytoplasm is a feature shared between introns 10 - 11 in nxf1 genes in vertebrates and introns 5 - 6 in nxf1 genes of different species of Drosophila. This, along with the ability of these introns to form secondary structures, suggests the functional significance of such transcripts. It is possible that the retention of the cassette intron in transcripts of nxf1 genes facilitates the synthesis of the truncated protein NXF1. We have shown that in the heads of adult flies the transcript with cassette intron 5 - 6 is more abundant than completely spliced transcripts (Figure 3) [57]. It would be interesting to test the hypothesis about the existence of the truncated protein Dm NXF1 specific to the brain tissues of the fruit fly.

Существование последовательностей, которые могут способствовать экспорту транскриптов, содержащих эти интроны, из ядра в цитоплазму, является общей чертой интронов 10-11 в генах nxf1 у позвоночных и интронов 5-6 в генах nxf1 у разных видов дрозофилы. Это, наряду со способностью этих интронов образовывать вторичные структуры, предполагает функциональное значение таких транскриптов. Возможно, сохранение кассетного интрона в транскриптах генов nxf1 способствует синтезу укороченного белка NXF1. Мы показали, что в головах взрослых мух транскрипты с кассетными интронами 5-6 встречаются чаще, чем полностью сплайсированные транскрипты (рис. 3) [57]. Было бы интересно проверить гипотезу о существовании укороченного белка Dm NXF1, специфичного для тканей мозга плодовой мушки.

**3. CONCLUSION**

A great variety of alternative transcripts, including intron-retained and non-coding RNAs, is characteristic of the nervous system. Thus, it is not surprising that transcripts with cassette intron 5 - 6 in Dm nxf1 are found primarily in the head of fruit flies. Insects that are characterized by complex behavior are model organisms, which facilitate the study of the molecular mechanisms of the nervous system. The large number of transcripts with retained intron 5 - 6 of the Dm nxf1 gene in the head of adult fruit flies points to the functional significance of these transcripts. The existence of evolutionarily conserved sequences in the cassette introns of nxf1 genes in animals within some taxonomic groups can be regarded as the acquisition of adaptive properties of the corresponding intron-containing transcripts. These sequences may provide benefits in nuclear-cytoplasmic transport, resistance to NMD, and a possible involvement in the regulation of gene expression. The ability of cassette introns to form complex secondary structures suggests that these introns may have an independent, possibly structural, function and also be a source of non-coding RNAs.

Для нервной системы характерно большое разнообразие альтернативных транскриптов, включая удерживаемые интроном и некодирующие РНК. Таким образом, неудивительно, что транскрипты с кассетным интроном 5—6 в Dm nxf1 обнаруживаются преимущественно в голове дрозофилы. Насекомые, характеризующиеся сложным поведением, являются модельными организмами, что облегчает изучение молекулярных механизмов нервной системы. Большое количество транскриптов с сохранившимся интроном 5—6 гена Dm nxf1 в голове взрослых плодовых мушек указывает на функциональное значение этих транскриптов. Существование эволюционно консервативных последовательностей в кассетных интронах генов nxf1 у животных некоторых таксономических групп можно рассматривать как приобретение адаптивных свойств соответствующими интронсодержащими транскриптами. Эти последовательности могут обеспечить преимущества в ядерно-цитоплазматическом транспорте, устойчивости к NMD и возможном участии в регуляции экспрессии генов. Способность кассетных интронов образовывать сложные вторичные структуры предполагает, что эти интроны могут иметь независимую, возможно, структурную функцию, а также быть источником некодирующих РНК.