**Organ-specific transcripts as a source of gene multifunctionality: lessons learned from the Drosophila melanogaster sbr (Dm nxf1) gene**

**Abstract**

Analysis of the transcriptomes of different organisms has demonstrated that a single gene can have multiple transcripts. The sources of transcriptional variability are the alternative promoters, polyadenylation sites, splicing, and RNA editing. A comparison of the organisms of different taxa has demonstrated that the complexity of organization during evolution arises not due to an increase in the number of protein-coding genes. The greatest variability of transcripts is specific to the nervous and germinal systems. A variety of mechanisms providing for the complexity of the transcriptome ensures a precise and coordinated regulation of organ-specific functions through a combination of cis-acting elements and trans-acting factors. The D. melanogaster sbr (Dm nxf1) gene has proven to be an excellent model for investigating mechanisms potentially leading to the emergence of multiple products with various functions.

Анализ транскриптомов различных организмов показал, что один ген может иметь несколько транскриптов. Источниками транскрипционной изменчивости являются альтернативные промоторы, сайты полиаденилирования, сплайсинг и редактирование РНК. Сравнение организмов разных таксонов показало, что усложнение организации в ходе эволюции возникает не за счет увеличения числа белок-кодирующих генов. Наибольшая вариабельность транскриптов характерна для нервной и зародышевой систем. Разнообразие механизмов, обеспечивающих сложность транскриптома, обеспечивает точную и скоординированную регуляцию органоспецифических функций за счет сочетания цис-действующих элементов и транс-действующих факторов. Ген D. melanogaster sbr (Dm nxf1) оказался отличной моделью для исследования механизмов, потенциально ведущих к появлению множественных продуктов с различными функциями.

**Contributors to Transcriptome Complexity**

Analysis of the transcriptomes of different organisms has demonstrated that a single gene can have multiple transcripts. In Drosophila melanogaster, 47 genes can potentially code for more than 1000 transcript isoforms each (Brown et al., 2014). The sources of transcriptional variability are the alternative promoters, polyadenylation sites, splicing, and RNA editing. The same DNA sequence can encode both protein coding transcripts and non-coding RNAs, including regulatory antisense transcripts. The so-called nested genes provide an additional opportunity for an increased variability of transcripts that are products of the same genome locus. These can be separate genes located either on the same strand or on opposite strands within the host gene (Kumar, 2009). Such “nested” genes are located in intron 3 of the D. melanogaster sbr (small bristles) gene (Fig. 1). Ten exons and nine introns of the sbr gene with a length of 14341 bp are located in genomic region X:10,832,752..10,847,092 (FlyBase, 2019). Intron 3 (8892 nt) comprises 62 % of this gene sequence. Within this intron, there exist three genes: CG32669, CG15209, and CG15210. Initially they were predicted to be ORFs; their transcriptional activity was demonstrated later. The expression pattern and functions of these genes have been poorly investigated so far. One of them, CG32669, has the same orientation as the sbr gene, and the other two genes — CG15209 and CG15210 — are transcribed in the opposite direction.

Анализ транскриптомов различных организмов показал, что один ген может иметь несколько транскриптов. У Drosophila melanogaster 47 генов потенциально могут кодировать более 1000 изоформ транскриптов каждый (Brown et al., 2014). Источниками транскрипционной изменчивости являются альтернативные промоторы, сайты полиаденилирования, сплайсинг и редактирование РНК. Одна и та же последовательность ДНК может кодировать как транскрипты, кодирующие белок, так и некодирующие РНК, включая регуляторные антисмысловые транскрипты. Так называемые вложенные гены дают дополнительную возможность для повышенной изменчивости транскриптов, являющихся продуктами одного и того же геномного локуса. Это могут быть отдельные гены, расположенные либо на одной цепи, либо на противоположных цепях внутри гена-хозяина (Kumar, 2009). Такие «вложенные» гены расположены в интроне 3 гена sbr (small bristles) D. melanogaster (рис. 1). Десять экзонов и девять интронов гена sbr длиной 14 341 п.н. расположены в геномной области X:10 832 752..10 847 092 (FlyBase, 2019). Интрон 3 (8892 н.) составляет 62 % последовательности этого гена. Внутри этого интрона существует три гена: CG32669, CG15209 и CG15210. Первоначально предполагалось, что они будут ORF; их транскрипционная активность была продемонстрирована позже. Паттерн экспрессии и функции этих генов до сих пор плохо изучены. Один из них, CG32669, имеет ту же ориентацию, что и ген sbr, а два других гена — CG15209 и CG15210 — транскрибируются в противоположном направлении.

Antisense transcripts that are complementary to the coding part of the gene are a common feature of transcriptomes. They include both long non-coding RNAs (lncRNAs) and short RNAs. Comprehensive analysis of the Drosophila transcriptome (Brown et al., 2014) has found both sense and antisense transcripts to be present in the same cell type at the same time. Sometimes antisense lncRNAs overlap with the 3′ and 5′ UTRs of adjacent genes, forming gene chains across contiguously transcribed regions. Antisense transcripts may overlap with both non-coding and coding sequences (CDSs). Moreover, orthologous genes have comparable antisense transcripts, which suggests a conserved regulation of gene expression with antisense transcripts (Brown et al., 2014).

Антисмысловые транскрипты, комплементарные кодирующей части гена, являются общей чертой транскриптомов. Они включают как длинные некодирующие РНК (днРНК), так и короткие РНК. Всесторонний анализ транскриптома дрозофилы (Brown et al., 2014) обнаружил, что как смысловые, так и антисмысловые транскрипты присутствуют в одном и том же типе клеток в одно и то же время. Иногда антисмысловые lncRNAs перекрываются с 3'- и 5'-UTR соседних генов, образуя генные цепи через непрерывно транскрибируемые области. Антисмысловые транскрипты могут перекрываться как с некодирующими, так и с кодирующими последовательностями (CDS). Более того, ортологичные гены имеют сравнимые антисмысловые транскрипты, что предполагает консервативную регуляцию экспрессии генов антисмысловыми транскриптами (Brown et al., 2014).

The presence of paralogous genes with specific functions as a result of duplication followed by subsequent functional divergence may increase the flexibility of the gene expression network and, subsequently, adaptive reaction. The Nxf gene family is a good example of the research of the evolution of paralogous genes. Inside this family, only the Nxf1 gene is homologous for all the species; other genes, even those sharing the same name, may be non-homologous in distant species. For example, the nxf3 of Drosophila is not homologous to the mammalian Nxf3 (Herold et al., 2001; Mamon et al, 2013), and the mouse Nxf3 is not orthologous to the human Nxf3 (Sasaki et al., 2005).

Наличие паралогичных генов со специфическими функциями в результате дупликации с последующей функциональной дивергенцией может увеличить гибкость сети экспрессии генов и, следовательно, адаптивную реакцию. Семейство генов Nxf является хорошим примером исследования эволюции паралогичных генов. Внутри этого семейства только ген Nxf1 гомологичен для всех видов; другие гены, даже те, которые имеют одно и то же имя, могут быть негомологичными у отдаленных видов. Напр., nxf3 дрозофилы не гомологичен Nxf3 млекопитающих (Herold et al., 2001; Mamon et al., 2013), а мышиный Nxf3 не ортологичен человеческому Nxf3 (Sasaki et al., 2005).

In mammals, the Nxf family includes genes with testis-specific and brain-specific expression (Jun et al., 2001; Tretyakova et al., 2005; Zhou et al., 2011). However, in D. melanogaster, the presence of organ-specific transcripts generated by different alternative events (such as the use of alternative promoters, alternative splicing, and polyadenylation) has been demonstrated only for the sbr (Dm nxf1) gene (Ivankova et al., 2010; Ginanova et al., 2016; Ginanova, 2017). An organ-specific pool of significantly different transcripts serves as the basis for the multifunctionality of the sbr gene. It is possible that certain tissue-specific functions are scattered throughout different paralogous genes in mammals.

У млекопитающих семейство Nxf включает гены со специфичной для семенников и мозговой экспрессией (Jun et al., 2001; Tretyakova et al., 2005; Zhou et al., 2011). Однако у D. melanogaster наличие органоспецифических транскриптов, генерируемых различными альтернативными событиями (такими как использование альтернативных промоторов, альтернативный сплайсинг и полиаденилирование), показано только для гена sbr (Dm nxf1) (Иванкова и др.). ., 2010; Гинанова и др., 2016; Гинанова, 2017). Органоспецифический пул существенно различающихся транскриптов служит основой полифункциональности гена sbr. Возможно, что определенные тканеспецифические функции разбросаны по разным паралогичным генам у млекопитающих.

**Alternative sources of transcriptome complexity**

**DIFFERENTIAL PROMOTER USAGE**

Core promoters consist of a variety of sequence elements, such as the TATA box, the Initiator (INR), and the downstream promoter element (DPE), recognized by the TATAbinding protein (TBP) and TBP-associated factors of the TFIID complex (Xu et al., 2016). No universal elements have been discovered in all core promoters, and the TATA box is found only in about 10 to 20% of the metazoan core promoters (Kadonaga, 2012). The canonical basal transcription machinery, including RNA polymerase II (RNAPII), interacts with the core promoter, typically located within -40 to +40 of the transcription start site (Kadonaga, 2002). The activity of the basal transcription machinery itself is quite low. Additional factors, such as activators or repressors with a variety of co-regulators (co-activators or co-repressors), are necessary for an effective transcription initiation by the RNAPII basal transcription machinery at the core promoter (Xu et al., 2016). These sequencespecific regulators of transcription are bound to regulatory DNA sequences, located at promoter-proximal or more distal regions. Transcription initiation efficiency of the RNAPII transcription machinery depends on the core promoter elements, combinatorial assortments of a variety of transcriptional activators or repressors, and chromatin modifications that can be navigated by RNAs, including miRNAs (Kadonaga, 2002, 2012).

Основные промоторы состоят из множества элементов последовательности, таких как блок ТАТА, инициатор (INR) и нижний промоторный элемент (DPE), распознаваемых ТАТА-связывающим белком (TBP) и связанными с TBP факторами комплекса TFIID (Xu и др., 2016). Универсальные элементы не были обнаружены во всех основных промоторах, и TATA-бокс обнаруживается только примерно у 10-20% основных промоторов многоклеточных животных (Kadonaga, 2012). Канонический базальный аппарат транскрипции, включая РНК-полимеразу II (RNAPII), взаимодействует с основным промотором, обычно расположенным в пределах от -40 до +40 от сайта начала транскрипции (Kadonaga, 2002). Активность самого базального аппарата транскрипции довольно низкая. Дополнительные факторы, такие как активаторы или репрессоры с различными корегуляторами (коактиваторами или корепрессорами), необходимы для эффективной инициации транскрипции базальным механизмом транскрипции РНКПII на основном промоторе (Xu et al., 2016). . Эти специфичные к последовательности регуляторы транскрипции связаны с регуляторными последовательностями ДНК, расположенными в проксимальных или более дистальных областях промотора. Эффективность инициации транскрипции транскрипционного аппарата RNAPII зависит от основных промоторных элементов, комбинаторных наборов разнообразных активаторов или репрессоров транскрипции и модификаций хроматина, которые могут управляться РНК, включая микроРНК (Kadonaga, 2002, 2012).

Alternative promoter usage is yet another mechanism responsible for the complexity of the transcriptome and proteome (Vacik and Raska, 2017). If alternative promoters are located downstream of the canonical transcription start site, usually in one of the introns, they drive the expression of alternative RNA isoforms without upstream exons. As a result, some important functional domains coded for by the upstream exons will be lost in proteins coded for by alternative mRNA isoforms. Such shortened protein isoforms can be functionally distinct from the full-length protein, coded for by the canonical mRNA isoform.

Использование альтернативного промотора является еще одним механизмом, ответственным за сложность транскриптома и протеома (Vacik and Raska, 2017). Если альтернативные промоторы расположены ниже канонического сайта начала транскрипции, обычно в одном из интронов, они управляют экспрессией альтернативных изоформ РНК без экзонов выше по течению. В результате некоторые важные функциональные домены, кодируемые вышестоящими экзонами, будут потеряны в белках, кодируемых альтернативными изоформами мРНК. Такие укороченные изоформы белка могут функционально отличаться от полноразмерного белка, кодируемого канонической изоформой мРНК.

**ALTERNATIVE SPLICING**

Alternative splicing (AS) is a common mechanism for increasing protein variety in eukaryotes (Black, 2000; Graveley, 2001; Nilsen and Graveley, 2010). Approximately 95 % of human genes are characterized by alternative pre-mRNA splicing (Wang et al., 2008). The major types of alternative splicing are exon “skipping”, the use of exon cassettes that are sets of several adjacent exons with only a single exon from each cassette being chosen during the splicing of pre-mRNA, and the use of different 5′ and 3′ splice sites in exons or introns. A complex interplay of cis- and trans-acting factors promotes or represses the assembly of a splicing complex called a spliceosome at the splice site, and carries out the selection of canonic or an alternative 5′ or 3′ splice site. The spliceosome is a multi-protein-RNA complex, responsible for the precise excision of introns or intron/ exon blocks from the pre-mRNA and the fusion of the remaining exons together.

Альтернативный сплайсинг (AS) является распространенным механизмом увеличения разнообразия белков у эукариот (Black, 2000; Graveley, 2001; Nilsen and Graveley, 2010). Приблизительно 95 % генов человека характеризуются альтернативным сплайсингом пре-мРНК (Wang et al., 2008). Основными типами альтернативного сплайсинга являются «пропуск экзонов», использование кассет экзонов, которые представляют собой наборы из нескольких соседних экзонов, при этом только один экзон из каждой кассеты выбирается во время сплайсинга пре-мРНК, а также использование различных 5'- и 3'-сайты сплайсинга в экзонах или интронах. Комплексное взаимодействие цис- и транс-действующих факторов способствует или подавляет сборку комплекса сплайсинга, называемого сплайсосомой, в сайте сплайсинга и осуществляет выбор канонического или альтернативного 5'- или 3'-сайта сплайсинга. Сплайсосома представляет собой мультибелковый РНК-комплекс, ответственный за точное вырезание интронов или интронно-экзонных блоков из пре-мРНК и слияние оставшихся экзонов вместе.

If the pre-mRNA of a gene undergoes alternative splicing of different types, the number of protein isoforms may exceed the number of coding genes of the respective organism. The Drosophila Dscam (Down syndrome cell adhesion molecule) gene encoding axon guidance receptors are a great example thereof. This gene includes four exon cassettes of varying exon numbers, and can potentially generate over 38000 mRNA splice-isoforms. Different Dscam mRNAs probably label cell types or even single cells in the Drosophila brain (Schmucker, 2000; Graveley, 2005).

Если пре-мРНК гена подвергается альтернативному сплайсингу разных типов, то количество изоформ белка может превышать количество кодирующих генов соответствующего организма. Отличным примером этого является ген Dscam дрозофилы (молекула клеточной адгезии синдрома Дауна), кодирующий рецепторы наведения аксонов. Этот ген включает четыре кассеты экзонов с различным числом экзонов и потенциально может генерировать более 38000 сплайс-изоформ мРНК. Различные мРНК Dscam, вероятно, маркируют типы клеток или даже отдельные клетки в мозге Drosophila (Schmucker, 2000; Graveley, 2005).

**INTRON RETENTION**

Intron retention (IR) is one of the variants of AS common in mammals. IR is considered to be a mechanism of gene expression regulation (Braunschweig et al., 2014; Jacob and Smith, 2017; Schmitz et al., 2017). IR enhances the complexity of the transcriptome (Schmitz et al., 2017). Within the main open reading frame, IR may result in the emergence of premature termination codons (PTCs), leading to the nonsense-mediated decay of mRNA with a retained intron, or, possibly, to the production of truncated proteins. mRNAs with retained introns give rise to alternative protein isoforms, and are therefore a source for the functional diversity of gene products. The existence of mRNA with an intron is a conservative feature of nxf1 genes in different organisms (Mamon et al., 2013, 2014; Wang et al., 2015). The retained intron may be part of the protein coding sequence, as in the Ce nxf1 mRNA, or it may only code for either the 17 C-terminus amino acids of the mammalian short NXF1 protein (Li et al., 2006) or the 5 C-terminus of the Drosophila short SBR (Mamon et al., 2013, 2014). The use of the PTC in the retained intron leads to the production of truncated proteins. In all cases, IR leads to the emergence of mRNAs with an extended 3′ UTR, including the canonical sequence of the Nxf1 mRNAs downstream of the retained intron. This suggests the presence of special regulatory functions related to the control of the spatial-temporal properties of mRNAs with a retained intron (Schmitz et al., 2017).

Задержка интрона (ИР) — один из вариантов АС, распространенных у млекопитающих. ИР рассматривается как механизм регуляции экспрессии генов (Braunschweig et al., 2014; Jacob, Smith, 2017; Schmitz et al., 2017). IR увеличивает сложность транскриптома (Schmitz et al., 2017). В пределах основной открытой рамки считывания ИР может приводить к появлению кодонов преждевременной терминации (PTC), что приводит к нонсенс-опосредованному распаду мРНК с сохраненным интроном или, возможно, к продукции укороченных белков. мРНК с сохранившимися интронами дают альтернативные изоформы белков и, следовательно, являются источником функционального разнообразия генных продуктов. Наличие мРНК с интроном является консервативным признаком генов nxf1 у разных организмов (Mamon et al., 2013, 2014; Wang et al., 2015). Сохраненный интрон может быть частью последовательности, кодирующей белок, как в мРНК Ce nxf1, или он может кодировать только 17 С-концевых аминокислот короткого белка NXF1 млекопитающих (Li et al., 2006) или 5 аминокислот. С-конец короткого SBR дрозофилы (Mamon et al., 2013, 2014). Использование PTC в сохраненном интроне приводит к продукции укороченных белков. Во всех случаях ИР приводит к появлению мРНК с удлиненной 3'-НТО, в том числе канонической последовательности мРНК Nxf1 ниже оставшегося интрона. Это свидетельствует о наличии особых регуляторных функций, связанных с контролем пространственно-временных свойств мРНК с сохраненным интроном (Schmitz et al., 2017).

**ALTERNATIVE POLYADENYLATION**

The 3′ end mRNA processing is a necessary step of mRNA maturation in eukaryotes, including endonucleolytic cleavage and untemplated polyadenylation. The cleavage and polyadenylation (CPA) machinery is composed of macromolecular complexes, such as the cleavage and polyadenylation stimulatory factor (CPSF), cleavage stimulatory factor (CStF), cleavage factor complexes — Im (CFIm) and IIm (CFIIm), and others (Proudfoot, 2011; Erson-Bensan, 2016). CPSF recognizes the polyadenylation signal (PAS), located ~10–30 nt upstream of the cleavage site. Most eukaryotic genes have multiple PASs used in alternative cleavage and polyadenylation (APA). There are two major PAS hexamers: AAUAAA and AUUAAA. Other weaker signal variants were later detected in different species: UAUAAA, AGUAAA, AAGAAA, AAUAUA, AAUACA, CAUAAA, GAUAAA, AAUGAA, UUUAAA, ACUAAA, AAUAGA, AAAUAA, AUAAAA, AUAAAU, AUAAAG, CAAUAA, UAAUAA, AUAAAC, AAAAUA, AAAAAA, AAAAAG, AACAAA, UUAUAU, AAAAAU, and UUUAUU (Beaudoing et al., 2000; Tian et al., 2005; Derti et al., 2012; Sanfilippo et al., 2017). Such noncanonical PASs may be recognized by specific trans-acting factors and serve for alternative polyadenylation. The choice of the cleavage and polyadenylation site is determined not only by PAS but also by the two motifs adjacent to it: U-rich/UGUA upstream elements (USEs) and the U-/GU-rich downstream element (DSE) (Colgan and Manley, 1997; Neve et al., 2017). The CstF complex interacts with DSE and mediates mRNA cleavage at the polyadenylation (pA) site (Neilson and Sandberg, 2010; Erson-Bensan, 2016). The CFIm complex binds to USE and mediates the cleavage reaction. The CFIIm complex promotes the termination of the RNA polymerase II-mediated transcription, and poly(A) polymerases (PAPs) catalyze the addition of untemplated adenosines downstream of the pA site (Proudfoot, 2011). The assembly process of the core factors of the cleavage and polyadenylation complex, including CPSF, CStF, CFIm, CFIIm, PAP, and others, is complicated by the interconnection between the 3′ end processing with 5′ capping, pre-mRNA splicing, and the transcription in the nucleus (Perales and Bentley, 2009; Szostak and Gebauer, 2012; Neve et al., 2017).

Процессинг 3'-конца мРНК является необходимым этапом созревания мРНК у эукариот, включая эндонуклеолитическое расщепление и полиаденилирование без матрицы. Аппарат расщепления и полиаденилирования (CPA) состоит из макромолекулярных комплексов, таких как фактор стимуляции расщепления и полиаденилирования (CPSF), фактор стимуляции расщепления (CStF), комплексы факторов расщепления — Im (CFIm) и IIm (CFIIm) и др. Праудфут, 2011; Эрсон-Бенсан, 2016). CPSF распознает сигнал полиаденилирования (PAS), расположенный примерно на 10–30 нт выше сайта расщепления. Большинство эукариотических генов имеют несколько PAS, используемых в альтернативном расщеплении и полиаденилировании (APA). Существует два основных гексамера PAS: AAUAAA и AUUAAA. Позже у разных видов были обнаружены другие более слабые варианты сигнала: UAUAAA, AGUAAA, AAGAAA, AAUAUA, AAUACA, CAUAAA, GAUAAA, AAUGAA, UUUAAA, ACUAAA, AAUAGA, AAAUAA, AUAAAA, AUAAAU, AUAAAG, CAAUAA, UAAUAA, AUAAAC, AAAAUA, AAAAAA. , AAAAAG, AACAAA, UUAUAU, AAAAAU и UUUAUU (Beaudoing et al., 2000; Tian et al., 2005; Derti et al., 2012; Sanfilippo et al., 2017). Такие неканонические PAS могут распознаваться специфическими трансакционными факторами и служить для альтернативного полиаденилирования. Выбор сайта расщепления и полиаденилирования определяется не только PAS, но и двумя соседними с ним мотивами: U-богатые/UGUA восходящие элементы (USEs) и U-/GU-богатые нижние элементы (DSE) (Colgan и Мэнли, 1997; Неве и др., 2017). Комплекс CstF взаимодействует с DSE и обеспечивает расщепление мРНК в сайте полиаденилирования (pA) (Neilson and Sandberg, 2010; Erson-Bensan, 2016). Комплекс CFIm связывается с USE и опосредует реакцию расщепления. Комплекс CFIIm способствует терминации опосредованной РНК-полимеразой II транскрипции, а поли(А)-полимеразы (PAPs) катализируют добавление нематричных аденозинов ниже сайта pA (Proudfoot, 2011). Процесс сборки кор-факторов комплекса расщепления и полиаденилирования, включая CPSF, CStF, CFIm, CFIIm, PAP и др., осложнен взаимосвязью процессинга 3'-конца с кэпированием 5', сплайсингом пре-мРНК и транскрипция в ядре (Perales, Bentley, 2009; Szostak, Gebauer, 2012; Neve et al., 2017).

Depending on the location of the cleavage and polyadenylation (pA) site, APA events can be classified into two major groups: coding region APA (CR-APA) and UTR-APA. CR-APA uses pA sites that are located within either CDS or the introns between CDS segments. As a result, alternative mRNA isoforms differ in their coding potential. UTR-APA uses the different pA sites that are located in the 3′ UTR, and the resulting coding potential of alternative mRNA isoforms does not change (Neve et al., 2017).

В зависимости от расположения сайта расщепления и полиаденилирования (pA) события APA можно разделить на две основные группы: кодирующая область APA (CR-APA) и UTR-APA. CR-APA использует сайты pA, которые расположены либо внутри CDS, либо в интронах между сегментами CDS. В результате альтернативные изоформы мРНК различаются по своему кодирующему потенциалу. UTR-APA использует разные сайты pA, расположенные в 3'-UTR, и результирующий кодирующий потенциал альтернативных изоформ мРНК не изменяется (Neve et al., 2017).

The majority of genes contain multiple PASs in their 3′ UTRs (Elkon et al., 2012); therefore, 3′ UTR-APA may be a very common event, increasing the variability of transcripts at a transcriptome level. mRNA is usually a part of the ribonucleoprotein (RNP) complex of RNAs and RNA-binding proteins attracted by a combination of cis-acting elements in the 3′ UTR or in another part of mRNA. The longer 3′ UTRs may contain additional regulatory cis-acting elements that can regulate mRNA localization and protein abundance, or create conditions and act a scaffolding for recruiting a protein complex containing RNA-binding proteins (Berkovits and Mayr, 2015; Mitra et al., 2015).

Большинство генов содержат множественные PAS в своих 3'-UTR (Elkon et al., 2012); следовательно, 3'-UTR-APA может быть очень частым событием, повышающим изменчивость транскриптов на уровне транскриптома. мРНК обычно является частью рибонуклеопротеинового (РНП) комплекса РНК и РНК-связывающих белков, привлекаемых комбинацией цис-действующих элементов в 3'-UTR или в другой части мРНК. Более длинные 3'-UTR могут содержать дополнительные регуляторные цис-действующие элементы, которые могут регулировать локализацию мРНК и количество белков, или создавать условия и действовать в качестве каркаса для рекрутирования белкового комплекса, содержащего РНК-связывающие белки (Berkovits and Mayr, 2015; Mitra et al. , 2015).

The use of the proximal PAS in the 3′ UTR may lead to a loss of cis-elements for binding with microRNAs and/or RNA-binding proteins (RBPs) (Lόpez de Salines et al., 2004; Flynt and Lai, 2008; Szostak and Gebauer, 2012; White et al., 2012). RBPs can recognize not only certain sequences, but also a specific secondary structure of the 3′ UTR. RBP target sites in 3′ UTR can overlap and form targets for miRNAs, and vice versa. In each case, mRNA’s fate is determined by the combination of cis-elements in 3′ UTR and the combination of transacting factors (RBP and miRNA). A different length of 3′ UTR can serve as a mechanism for differential regulation of subcellular functions of the same proteins (An et al., 2008). There are two types of mRNA BDNF (brainderived neurotrophic factor) with long 3′ UTR and short 3′ UTR in mice. It was shown that only the long 3′ UTR provides the targeting mRNA BDNF to dendrites, controlling the abundance of the dendritic BDNF protein and the pruning and enlargement of dendritic spines. miRNAs binding sites in the 3′ UTRs are cis-acting elements, which mediate mRNA decay and translational repression in animals (Iwakawa and Tomari, 2015). So, despite no changes in the protein-coding capacity difference in 3′ UTR, it may be functionally significant and can influence the transport, stability, translation efficiency, and subcellular localization of mRNAs (Colgan and Manley, 1997; Hilger et al., 2011).

Использование проксимального PAS в 3'-UTR может привести к потере цис-элементов для связывания с микроРНК и/или РНК-связывающими белками (RBP) (Lόpez de Salines et al., 2004; Flynt and Lai, 2008; Шостак и Гебауэр, 2012; Уайт и др., 2012). RBP могут распознавать не только определенные последовательности, но и специфическую вторичную структуру 3'-UTR. Сайты-мишени RBP в 3'-UTR могут перекрываться и образовывать мишени для miRNAs, и наоборот. В каждом случае судьба мРНК определяется комбинацией цис-элементов в 3'-UTR и комбинацией трансактирующих факторов (RBP и miRNA). Различная длина 3'-UTR может служить механизмом дифференциальной регуляции субклеточных функций одних и тех же белков (An et al., 2008). Существует два типа мРНК BDNF (мозговой нейротрофический фактор) с длинной 3'-UTR и короткой 3'-UTR у мышей. Было показано, что только длинная 3'-UTR обеспечивает нацеливание мРНК BDNF на дендриты, контролируя обилие дендритного белка BDNF, а также обрезку и увеличение дендритных шипиков. Сайты связывания miRNAs в 3'-UTR представляют собой цис-действующие элементы, которые обеспечивают распад мРНК и репрессию трансляции у животных (Iwakawa and Tomari, 2015). Таким образом, несмотря на отсутствие изменений в способности кодировать белок в 3'-UTR, она может быть функционально значимой и может влиять на транспорт, стабильность, эффективность трансляции и субклеточную локализацию мРНК (Colgan and Manley, 1997; Hilger et al., 2011).

The length of 3′ UTRs can change by increasing or decreasing the poly(A)-tail in the cytoplasm allowing for the regulation of the stability of mRNA and the efficiency of its translation (Richter, 1996). The cytoplasmic polyadenylation element (CPE) in 3′ UTR is the binding site for the CPE-binding protein (CPEB) that promotes polyadenylation-induced translation. CPE (UUUUUAU) is located upstream of the nuclear PAS (AAUAAA). In Xenopus laevis, CPEB associates with Maskin, binding the translation initiation factor 4E (Barnard et al., 2005). The interaction of Maskin with eIF4E excludes eIF4G and prevents the formation of the eIF4F initiation complex. Phosphorylation events within the CPE-binding protein complex disrupt its connection with the eIF4F initiation complex and allow the cytoplasmic poly(A) polymerase to elongate the poly(A) tail of mRNA.The elongated poly(A) tail is bound by the poly(A)-binding protein, which in turn binds eIF4G, disrupting the Maskin and eIF4E interaction, thereby initiating translation (Barnard et al., 2005; Richter and Sonenberg, 2005). Such strategy of translation regulation allows for the storage of mRNAs in a state of temporary unavailability for translation and for the activation of mRNA translation with the corresponding signal. This mechanism of translation regulation of specific mRNAs in cell cycle regulation (Barnard et al., 2005; Richter and Sonenberg, 2005) and the synapto-dendritic compartment of neurons (Wells et al., 2000; Du and Richter, 2005) has been found.

Длина 3'-UTR может изменяться за счет увеличения или уменьшения поли(А)-хвоста в цитоплазме, что позволяет регулировать стабильность мРНК и эффективность ее трансляции (Richter, 1996). Цитоплазматический элемент полиаденилирования (CPE) в 3'-UTR представляет собой сайт связывания CPE-связывающего белка (CPEB), который способствует трансляции, индуцированной полиаденилированием. CPE (UUUUUAU) расположен выше по течению от ядерной PAS (AAUAAA). У Xenopus laevis CPEB ассоциирован с Maskin, связывая фактор инициации трансляции 4E (Barnard et al., 2005). Взаимодействие Maskin с eIF4E исключает eIF4G и препятствует образованию комплекса инициации eIF4F. События фосфорилирования в комплексе СРЕ-связывающего белка нарушают его связь с комплексом инициации eIF4F и позволяют цитоплазматической поли(А)-полимеразе удлинять поли(А)-хвост мРНК. Удлиненный поли(А)-хвост связывается с поли(А)-хвостом. A)-связывающий белок, который, в свою очередь, связывает eIF4G, нарушая взаимодействие Maskin и eIF4E, тем самым инициируя трансляцию (Barnard et al., 2005; Richter and Sonenberg, 2005). Такая стратегия регуляции трансляции позволяет сохранять мРНК в состоянии временной недоступности для трансляции и активировать трансляцию мРНК соответствующим сигналом. Этот механизм регуляции трансляции специфических мРНК в регуляции клеточного цикла (Barnard et al., 2005; Richter and Sonenberg, 2005) и в синапто-дендритном компартменте нейронов (Wells et al., 2000; Du and Richter, 2005) был изучен. найденный.

During the process of CR-APA, the use of an alternative PAS in an intron leads to the formation of a protein isoform without the C-terminus of a full-length protein. The U1snRNP actively suppresses PASs in introns. These PASs are called cryptic (Neve et al., 2017). However, these intronic PASs can become available during increased proliferation (Elkon et al., 2012). mRNAs with APA in the intron are a source of truncated proteins that are incapable of performing the functions of fulllength proteins that are coded by the same gene. Identification of trans-acting APA regulators and cis-acting regulatory elements may promote understanding of the mechanisms of APA (White et al., 2012).

В процессе CR-APA использование альтернативной PAS в интроне приводит к образованию изоформы белка без С-конца полноразмерного белка. U1snRNP активно подавляет PAS в интронах. Эти PAS называются криптическими (Neve et al., 2017). Однако эти интронные PAS могут стать доступными во время повышенной пролиферации (Elkon et al., 2012). мРНК с АРА в интроне являются источником укороченных белков, неспособных выполнять функции полноразмерных белков, кодируемых тем же геном. Идентификация транс-действующих регуляторов АРА и цис-действующих регуляторных элементов может способствовать пониманию механизмов АРА (White et al., 2012).

The majority of the sources of transcriptome complexity is the cornerstone of the origin of the sbr organspecific transcripts.

Большинство источников сложности транскриптома является краеугольным камнем происхождения органоспецифических транскриптов sbr.

**ORGAN-SPECIFIC TRANSCRIPTS**

The analysis of transcriptomes has made it possible to conclude that the greatest variability of transcripts is specific to the nervous and germinal systems. Genes expressed in the testes, brain, and ovaries have properties that promote the variability of transcriptomes.

Анализ транскриптомов позволил заключить, что наибольшая вариабельность транскриптов специфична для нервной и зародышевой систем. Гены, экспрессируемые в семенниках, головном мозге и яичниках, обладают свойствами, способствующими изменчивости транскриптомов.

**BRAIN-SPECIFIC TRANSCRIPTS**

The presence of mRNAs with an extended 3′ UTR is a conserved feature of the nervous system in contrast to the testis transcripts (Miura et al., 2013, 2014; Hilger et al., 2011; Smibert et al., 2012). In the nervous system, hundreds of genes have mRNAs being processed with the use of increasingly distant PASs, and their 3′ UTRs reach tens of kb in length (Hilgers, 2015). Specific RNA-binding proteins inhibit cleavage and polyadenylation (CPA) at proximal sites. In Drosophila, the ELAV (embryoniclethal abnormal visual system) protein is known to play the role of APA regulators. However, no binding sites for ELAV have been identified in the extended 3′ UTRs (Hilger et al., 2012; Smibert et al., 2012). It was suggested that the distant APA depends on specific sequences in the promoters of genes producing mRNAs with an extended 3′ UTR (Hilgers et al., 2012; Hilgers, 2015; Oktaba et al., 2015). A search of such sequences by means of computational analysis has revealed that the GAGA element is frequently found in the promoter regions of the genes producing mRNAs with extended 3′ UTR (Li and Gilmour, 2013; Oktaba et al., 2015). The GAGA element signals pausing to the RNA Polymerase II (Pol II). This confirms the hypothesis that there exists interrelation between the processes of transcription initiation and CPA.

Наличие мРНК с удлиненной 3'-UTR является консервативным признаком нервной системы, в отличие от транскриптов семенников (Miura et al., 2013, 2014; Hilger et al., 2011; Smibert et al., 2012). В нервной системе сотни генов имеют процессинг мРНК с использованием все более удаленных PAS, а их 3'-UTR достигают в длину десятков т.п.н. (Hilgers, 2015). Специфические РНК-связывающие белки ингибируют расщепление и полиаденилирование (CPA) в проксимальных участках. Известно, что у дрозофилы белок ELAV (эмбрионлетальная аномальная зрительная система) играет роль регуляторов АРА. Однако сайты связывания ELAV не были идентифицированы в удлиненных 3'-UTR (Hilger et al., 2012; Smibert et al., 2012). Предполагается, что отдаленная АРА зависит от специфических последовательностей в промоторах генов, продуцирующих мРНК с удлиненной 3'-НТО (Hilgers et al., 2012; Hilgers, 2015; Oktaba et al., 2015). Поиск таких последовательностей с помощью компьютерного анализа показал, что элемент GAGA часто встречается в промоторных областях генов, продуцирующих мРНК с удлиненной 3'-НТО (Li, Gilmour, 2013; Oktaba et al., 2015). Элемент GAGA сигнализирует о паузе для РНК-полимеразы II (Pol II). Это подтверждает гипотезу о наличии взаимосвязи между процессами инициации транскрипции и ЦПА.

CStF-64 is a brain-specific CStF. CStF-64 is a splice variant of CStF and is found in all regions of the brain and the peripheral nervous system (MacDonald and McMahon, 2010). CStF-64 may play a role in determining the polyadenylation site in brain-specific mRNAs.

CStF-64 представляет собой CStF, специфичный для головного мозга. CStF-64 представляет собой сплайс-вариант CStF и обнаруживается во всех областях мозга и периферической нервной системы (MacDonald and McMahon, 2010). CStF-64 может играть роль в определении сайта полиаденилирования в мРНК, специфичных для мозга.

Interestingly, the hippocampus exhibits the largest number of 3′ UTR extensions in comparison with all other tissues (Miura et al., 2013). 3′ UTR lengthening may increase the number of cis-acting elements used as target sites for neural-specific miRNAs.

Интересно, что гиппокамп демонстрирует наибольшее количество расширений 3'-UTR по сравнению со всеми другими тканями (Miura et al., 2013). Удлинение 3'-UTR может увеличивать количество цис-действующих элементов, используемых в качестве сайтов-мишеней для нейрон-специфических микроРНК.

There are transcripts that undergo post-transcriptional cleavage to release specific fragments, which then function independently (Tuck and Tollervey, 2011), and 3′ UTR extensions may be a source of ncRNAs after posttranscriptional cleavage (Miura et al., 2013). A significant increase in the 3′ UTR length is observed in the sbr mRNA with retained intron 5. This intron was dubbed a cassette intron because it is part of the evolutionarily conserved cassette exon — 110 bp — intron — exon — 37 bp, which is found in the Nxf1 genes in different organisms. mRNAs with a cassette intron are also found in organisms of different taxa (Mamon et al., 2013, 2014; Wang et al., 2015). Cassette introns contain evolutionarily conserved motifs. In Drosophilidae they are evident in the form of two extended poly(A) sequences. Such poly(A)-tracts are known to facilitate the inclusion of corresponding introns into the processed RNA (Jacob and Smith, 2017).

Существуют транскрипты, которые подвергаются посттранскрипционному расщеплению с высвобождением специфических фрагментов, которые затем функционируют независимо (Tuck and Tollervey, 2011), а удлинения 3'-UTR могут быть источником нкРНК после посттранскрипционного расщепления (Miura et al., 2013). Значительное увеличение длины 3'-НТО наблюдается в мРНК sbr с сохраненным интроном 5. Этот интрон был назван кассетным интроном, поскольку он входит в состав эволюционно консервативной кассеты экзон — 110 п.н. — интрон — экзон — 37 п.н., т.е. обнаружены в генах Nxf1 у разных организмов. мРНК с кассетным интроном также встречаются у организмов разных таксонов (Mamon et al., 2013, 2014; Wang et al., 2015). Кассетные интроны содержат эволюционно консервативные мотивы. У дрозофилид они проявляются в виде двух протяженных поли(А) последовательностей. Известно, что такие поли(А)-тракты облегчают включение соответствующих интронов в процессируемую РНК (Jacob, Smith, 2017).

It should be noted that the sbr gene’s intron containing mRNA is the most abundant among the sbr transcripts in adult fly heads (Fig. 4). This is also supported by the results of a northern blot analysis, and suggests neurospecificity of the mRNAs with the retained intron (Ivankova et al., 2010). The sbr mRNA with the retained intron gives rise to a truncated protein (unpublished). The investigation of the role of the truncated SBR protein in the formation and function of the nervous system is currently underway.

Следует отметить, что мРНК, содержащая интрон гена sbr, является самой распространенной среди транскриптов sbr в головах взрослых мух (рис. 4). Это также подтверждается результатами нозерн-блоттинга и свидетельствует о нейроспецифичности мРНК с сохраненным интроном (Ivankova et al., 2010). мРНК sbr с сохранившимся интроном дает укороченный белок (неопубликовано). В настоящее время проводится исследование роли укороченного белка SBR в формировании и функционировании нервной системы.

**TESTIS-SPECIFIC TRANSCRIPTS**

Translational regulation is a fundamental characteristic of gene expression in mammalian testes (Kleene, 2001, 2003). Long-living mRNAs that are synthesized long before translation have been found. They are stored in a translationally inactive state for several days, and are then translated in transcriptionally inactive elongated spermatids during the formation of spermatozoa from spermatids (Hawthorne et al., 2006). The widespread use of the alternative promoter, upstream PAS, and alternative splicing are features of gene expression in spermatogenic cells (Kleene, 2001, 2003). In spermatogenic cells, the transcripts of some housekeeping genes encode truncated proteins that cannot have the same functions as the full-length products of the same genes in somatic cells (Kleene, 2001). Alternative promoter usage forms an alternative 5′ UTR determining the efficiency of mRNA translation. Translation is repressed with the upstream reading frames (uORFs) in the 5′ UTR before the start codon (AUG) of the reading frame encoding the “functional” protein in the mRNA (Child et al., 1999). The presence of uORFs in mRNA’s 5′ UTR is an additional mechanism of the translation regulation in spermatogenesis (Kleene, 2003). Paralogous genes within gene families with only testes-specific expression are yet another feature of spermatogenesis. In mice, round spermatids express a testisspecific isoform of poly(A) polymerase, which is localized in the cytoplasm in contrast to the somatic isoform with the nuclear localization (Kashiwabara et al., 2000, 2002). Thus, cytoplasmic polyadenylation is an additional way of translation regulation of mRNAs during spermatogenesis (Kashiwabara et al., 2002; Kleene, 2003). There are testisspecific translation repressors, such as Y-box proteins and the testis-brain RNA binding protein (TB-RBP) in mammals (Kleene, 2003). It is an additional source allowing each mRNA to use its own mechanism of translational regulation. The specifics of a testis-specific transcriptome consist of transcripts of many genes, which are expressed in both somatic and spermatogenic cells, and are modified by alternative transcription start sites, splicing, and upstream polyadenylation sites, that form the atypical patterns of gene expression in spermatogenical cells (Kleene, 2001, 2003).

Трансляционная регуляция является фундаментальной характеристикой экспрессии генов в семенниках млекопитающих (Kleene, 2001, 2003). Обнаружены долгоживущие мРНК, которые синтезируются задолго до трансляции. Они сохраняются в трансляционно-неактивном состоянии в течение нескольких дней, а затем транслируются в транскрипционно-неактивные удлиненные сперматиды при образовании сперматозоидов из сперматид (Hawthorne et al., 2006). Широкое использование альтернативного промотора, восходящей PAS и альтернативного сплайсинга является особенностями экспрессии генов в сперматогенных клетках (Kleene, 2001, 2003). В сперматогенных клетках транскрипты некоторых генов домашнего хозяйства кодируют укороченные белки, которые не могут выполнять те же функции, что и полноразмерные продукты тех же генов в соматических клетках (Kleene, 2001). Использование альтернативного промотора формирует альтернативную 5'-UTR, определяющую эффективность трансляции мРНК. Трансляция репрессируется восходящими рамками считывания (uORF) в 5'-UTR перед стартовым кодоном (AUG) рамки считывания, кодирующим «функциональный» белок в мРНК (Child et al., 1999). Присутствие uORFs в 5'UTR мРНК является дополнительным механизмом регуляции трансляции в сперматогенезе (Kleene, 2003). Паралогичные гены в семействах генов с экспрессией, специфичной только для семенников, являются еще одной особенностью сперматогенеза. У мышей круглые сперматиды экспрессируют специфичную для семенников изоформу поли(А)-полимеразы, локализованную в цитоплазме, в отличие от соматической изоформы с ядерной локализацией (Kashiwabara et al., 2000, 2002). Таким образом, цитоплазматическое полиаденилирование является дополнительным способом регуляции трансляции мРНК во время сперматогенеза (Kashiwabara et al., 2002; Kleene, 2003). Существуют семенниково-специфические репрессоры трансляции, такие как белки Y-box и РНК-связывающий белок семенников-мозга (TB-RBP) у млекопитающих (Kleene, 2003). Это дополнительный источник, позволяющий каждой мРНК использовать свой механизм регуляции трансляции. Специфика специфичного для семенников транскриптома состоит из транскриптов многих генов, которые экспрессируются как в соматических, так и в сперматогенных клетках и модифицируются альтернативными сайтами начала транскрипции, сплайсинга и вышестоящими сайтами полиаденилирования, которые формируют атипичные паттерны экспрессии генов в сперматогенные клетки (Kleene, 2001, 2003).

Mammalian and invertebrate testes use many mRNAs with shorter 3′ UTRs compared to other tissues (Liu et al., 2007; Smibert et al., 2012; Neve et al., 2017). 3′ UTR shortening promotes mRNA stability and translational effectivity. It has been demonstrated that in mice, shortening the 3′ UTR eliminates destabilizing elements, such as AU-rich elements and transposable elements located downstream of testis-specific APA (tsAPA) in 3′ UTR (Li W. et al., 2016).

Семенники млекопитающих и беспозвоночных используют многие мРНК с более короткими 3'-UTR по сравнению с другими тканями (Liu et al., 2007; Smibert et al., 2012; Neve et al., 2017). Укорочение 3'UTR способствует стабильности мРНК и эффективности трансляции. Было продемонстрировано, что у мышей укорочение 3'-UTR устраняет дестабилизирующие элементы, такие как элементы, богатые AU, и мобильные элементы, расположенные ниже по течению от специфичной для семенников APA (tsAPA) в 3'-UTR (Li W. et al., 2016). .

Since tsAPA is not used efficiently in somatic cells, it has been suggested that there are testis-specific factors involved in the cleavage and polyadenylation process in non-canonical proximal APAs, such as the testis-specific form of CStF-64 (Wallace et al., 1999). The 64-kDa subunit of the CStF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences the cleavage site choice (MacDonald et al., 1994; MacDonald and McMahon, 2010). Somatic and testisspecific forms of CStF-64 have been found (Wallace et al., 1999). The CStF-64 gene for the somatic form is encoded on the X chromosome in both mice and humans. This suggests that the somatic form of CStF-64 is absent in meiotic cells because of the X chromosome inactivation, whereas the autosomal CStf2t gene encodes the testis-specific form τCStF-64 (Das et al., 2007). The τCStF-64 is found in spermatocytes and early spermatids (Wallace et al., 1999).

Поскольку tsAPA не используется эффективно в соматических клетках, было высказано предположение, что существуют специфичные для семенников факторы, участвующие в процессах расщепления и полиаденилирования в неканонических проксимальных APA, такие как специфичная для семенников форма CStF-64 (Wallace et al. ., 1999). Субъединица 64-кДа фактора полиаденилирования CStF связывается с пре-мРНК ниже сайта расщепления и влияет на выбор сайта расщепления (MacDonald et al., 1994; MacDonald and McMahon, 2010). Были обнаружены соматические и тестис-специфические формы CStF-64 (Wallace et al., 1999). Ген CStF-64 соматической формы кодируется на Х-хромосоме как у мышей, так и у людей. Это предполагает, что соматическая форма CStF-64 отсутствует в мейотических клетках из-за инактивации Х-хромосомы, тогда как аутосомный ген CStf2t кодирует специфичную для семенников форму τCStF-64 (Das et al., 2007). τCStF-64 обнаруживается в сперматоцитах и ранних сперматидах (Wallace et al., 1999).

Alternative promoters in the intron 3 of the Drosophila melanogaster sbr (Dm nxf1) gene drive the expression of two testis-specific mRNA isoforms without exons 1, 2, and 3. 5′ UTRs of these mRNAs isoforms contain either uORFs, or a small intron (Ginanova et al., 2016). Both features of the 5′ UTR are known to control translation efficiency of the corresponding mRNAs in mammals (Mededbach et al., 2011; Bicknell et al., 2012); their functionality in Drosophila is unknown. The testisspecific sbr mRNAs initiate the synthesis of the testisspecific short protein (tSBR), which excludes nuclear localization signals (NLS) present in the canonical SBR protein (Fig. 3). This suggests that tSBR is unable to get into the nucleus by itself and carries out some functions in the cytoplasm, likely participating in the biogenesis of long-living mRNAs during meiosis and spermiogenesis (Ginanova et al., 2016).

Альтернативные промоторы в интроне 3 гена Drosophila melanogaster sbr (Dm nxf1) управляют экспрессией двух специфичных для семенников изоформ мРНК без экзонов 1, 2 и 3. 5'-UTR этих изоформ мРНК содержат либо uORF, либо небольшой интрон (Гинанова и др., 2016). Известно, что обе особенности 5'-UTR контролируют эффективность трансляции соответствующих мРНК у млекопитающих (Mededbach et al., 2011; Bicknell et al., 2012); их функциональность у дрозофилы неизвестна. Специфичные для семенников мРНК sbr инициируют синтез специфического для семенников короткого белка (tSBR), который исключает сигналы ядерной локализации (NLS), присутствующие в каноническом белке SBR (Fig. 3). Это свидетельствует о том, что tSBR не может самостоятельно проникнуть в ядро и выполняет некоторые функции в цитоплазме, вероятно, участвуя в биогенезе долгоживущих мРНК в ходе мейоза и спермиогенеза (Ginanova et al., 2016).

All sbr (Dm nxf1) mRNAs have a shortened 3′ UTR in testes (Ginanova et al., 2016) (Fig. 2). The 3′ UTR of the full-length sbr mRNAs has several cis-acting elements, such as AU-rich, alternative PASs, CPE, and the predicted target-sites for miRNAs. All of them are located downstream of the cleavage and polyadenylation sites, which are used during the processing of the testisspecific sbr mRNAs. The poly(U)-tract is among the ciselements in the 3′ UTR of sbr (Dm nxf1) mRNAs, and remains in the shortened 3′ UTR of sbr (Dm nxf1) mRNAs. The polypyrimidine-tract-binding (PTB) protein 2 regulates meiotic male germ cell mRNAs in mice (Iguchi et al., 2006). The dmPTB plays an important role during spermatogenesis in Drosophila (Sridharan et al., 2016).

Все мРНК sbr (Dm nxf1) имеют укороченную 3'-UTR в семенниках (Ginanova et al., 2016) (рис. 2). 3'-UTR полноразмерных sbr мРНК имеет несколько цис-действующих элементов, таких как AU-богатые, альтернативные PAS, CPE и предполагаемые сайты-мишени для miRNAs. Все они расположены ниже сайтов расщепления и полиаденилирования, которые используются при процессинге специфичных для семенников мРНК sbr. Поли(U)-тракт входит в число cis-элементов в 3'-UTR мРНК sbr (Dm nxf1) и остается в укороченной 3'-UTR мРНК sbr (Dm nxf1). Белок 2 polypyrimidine-tract-binding (PTB) регулирует мРНК мейотических мужских зародышевых клеток у мышей (Iguchi et al., 2006). dmPTB играет важную роль во время сперматогенеза у Drosophila (Sridharan et al., 2016).

Testis-specific transcripts are unknown for the mammalian Nxf1 gene. There are testis-specific paralogous genes among Nxf gene family in mice (Sasaki et al., 2005) and humans (Yang et al., 2001).

Транскрипты, специфичные для семенников, неизвестны для гена Nxf1 млекопитающих. Среди генов семейства Nxf у мышей (Sasaki et al., 2005) и людей (Yang et al., 2001) есть паралогичные гены, специфичные для семенников.

**TRANSCRIPTS IN OOGENESIS**

A subset of maternal mRNAs and proteins are synthesized during oogenesis and are retained in the oocyte to direct the first mitotic divisions and specify the patterning of the embryo. Early embryogenesis passes through a stage when developmental control is handed from maternally provided gene products to those synthesized from (by) the zygotic genome (Tadros and Lipshitz, 2009). The oocyte-to-embryo transition (OET) can be subdivided into two interrelated processes: first, a subset of maternal mRNAs and proteins is eliminated; then the transcription of the zygotic genome begins.

Подмножество материнских мРНК и белков синтезируются во время оогенеза и сохраняются в ооците, чтобы направлять первые митотические деления и определять формирование паттерна эмбриона. Ранний эмбриогенез проходит стадию, когда контроль над развитием передается от материнских генных продуктов к продуктам, синтезированным из зиготического генома (Tadros and Lipshitz, 2009). Переход от ооцита к эмбриону (OET) можно разделить на два взаимосвязанных процесса: во-первых, элиминируется подмножество материнских мРНК и белков; затем начинается транскрипция зиготического генома.

There are several pathways of maternal mRNA degradation. One of them is the ARE (AU-rich element)- mediated pathway known in Xenopus laevis (Voeltz and Steitz, 1998). The Embryonic Deadenylation Element Binding Protein (EDEN-BP) triggers the deadenylation of maternal transcripts upon fertilization via the recognition of AU-rich cis-elements. The activity of EDEN-BP is regulated by phosphorylation (Detivaud et al., 2003). This may explain the different activity of EDEN-BP in oocytes and early embryos. Another way to mediate the maternal mRNA degradation is accomplished through the binding of a miRNA (Guo et al., 2008; Lund et al., 2009).

Существует несколько путей деградации материнской мРНК. Одним из них является путь, опосредованный ARE (AU-rich element), известный у Xenopus laevis (Voeltz and Steitz, 1998). Белок, связывающий элемент эмбрионального деаденилирования (EDEN-BP), запускает деаденилирование материнских транскриптов при оплодотворении посредством распознавания цис-элементов, богатых AU. Активность EDEN-BP регулируется фосфорилированием (Detivaud et al., 2003). Это может объяснить различную активность EDEN-BP в ооцитах и ранних эмбрионах. Др. способ опосредования деградации материнской мРНК осуществляется посредством связывания miRNA (Guo et al., 2008; Lund et al., 2009).

Yet another way of post-transcriptional regulation of the maternal mRNAs in early embryogenesis is the CPE-dependent mRNA polyadenylation and its consequent translation. CPEB-mediated mRNA silencing/ re-activation is essential to oogenesis (Tay and Richter, 2001) and in developing hippocampal neurons (Kundel et al. 2009) in mice. Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and embryogenesis in Drosophila (Coll et al., 2010; Cui et al., 2013).

Еще одним путем посттранскрипционной регуляции материнских мРНК в раннем эмбриогенезе является CPE-зависимое полиаденилирование мРНК и ее последующая трансляция. CPEB-опосредованное замалчивание/реактивация мРНК важно для оогенеза (Tay and Richter, 2001) и для развития нейронов гиппокампа (Kundel et al. 2009) у мышей. Цитоплазматическое полиаденилирование является основным регулятором мРНК во время оогенеза и эмбриогенеза у Drosophila (Coll et al., 2010; Cui et al., 2013).

The sequence motifs in the 3′ UTRs are the “combinatorial code” allowing for a precise spatial-temporal regulation of mRNA translation during OET (Evsikov et al. 2006), and the keys to a specific code combination are provided by a combination and/or modifications of the corresponding trans-acting factors with protective or destructive functions (Svoboda et al., 2015; Schultz et al., 2018).

Мотивы последовательности в 3'-UTR представляют собой «комбинаторный код», позволяющий осуществлять точную пространственно-временную регуляцию трансляции мРНК во время OET (Евсиков и др., 2006), а ключи к конкретной кодовой комбинации даются комбинацией и/ или модификации соответствующих транс-действующих факторов с защитной или деструктивной функцией (Свобода и др., 2015; Шульц и др., 2018).

The greatest variety of the sbr mRNAs is found in the ovaries (Fig. 5). Moreover, the canonical isoform is not the most numerous. Several different promoters are used as transcription starts for different mRNAs sbr. Alternative promoters are located in exons 2, 3, and 4 in addition to the canonical one. The sbr expression patterns in early embryos (0–2 h) and in the ovaries are similar, according to the northern blot analysis (Ivankova et al., 2010). One may hypothesize that the majority of sbr mRNA isoforms are transcribed during oogenesis, and the maternal products of the sbr gene are necessary during early development of the embryo. The mutations of the sbr gene have a maternal effect on early development. The SBR protein is abundant in early embryos and marks the spindles of nuclear divisions (Golubkova et al., 2015). The fraction of the shorter sbr mRNAs that is present in embryos 1–2 h of age disappears when zygotic transcription begins (Ivankova et al., 2010). These results imply an essential role of the sbr gene in the early embryogenesis of D.melanogaster. Since the early embryogenesis of D.melanogaster occurs in the absence of a transcription, the functions of SBR (Dm NXF1) as a factor of nuclear export of mRNAs are unnecessary. Nuclear division in syncytial embryos of D.melanogaster is characterized by semiclosed mitosis within the nuclear membranes, which are only disrupted at the poles adjacent to the centrioles (Stafstrom and Staehelin, 1984). The NPC (nuclear pore complex) assembly and disassembly at the site of fusion between the mitotic nuclear envelope and the overlying spindle membrane are detected during early synchronous mitosis in D.melanogaster embryos (Kiseleva et al., 2001). The predicted truncated products of the sbr gene during oogenesis can have alternative functions. The sbr gene products participate in complicated processes that can be relatively independent from transcription. These processes are provided by the cyclic translation of mRNAs of factors that are involved in DNA replication, nuclear divisions, and cytoskeleton dynamics.

Наибольшее разнообразие мРНК sbr обнаружено в яичниках (рис. 5). При этом каноническая изоформа не самая многочисленная. Несколько разных промоторов используются в качестве стартов транскрипции для разных мРНК sbr. Альтернативные промоторы расположены в экзонах 2, 3 и 4 в дополнение к каноническому. Паттерны экспрессии sbr у ранних эмбрионов (0–2 ч) и в яичниках сходны, по данным нозерн-блоттинга (Ivankova et al., 2010). Можно предположить, что большинство изоформ мРНК sbr транскрибируются во время оогенеза, а материнские продукты гена sbr необходимы на раннем этапе развития эмбриона. Мутации гена sbr оказывают материнское влияние на раннее развитие. Белок SBR в изобилии встречается у ранних эмбрионов и маркирует веретена ядерных делений (Голубкова и др., 2015). Фракция более коротких sbr мРНК, присутствующая у эмбрионов 1-2-часового возраста, исчезает с началом зиготической транскрипции (Ivankova et al., 2010). Эти результаты указывают на существенную роль гена sbr в раннем эмбриогенезе D. melanogaster. Поскольку ранний эмбриогенез D. melanogaster протекает в отсутствие транскрипции, функции SBR (Dm NXF1) как фактора ядерного экспорта мРНК не нужны. Деление ядра у синцитиальных зародышей D. melanogaster характеризуется полузакрытым митозом внутри ядерных оболочек, которые разрываются только на полюсах, прилегающих к центриолям (Stafstrom, Staehelin, 1984). Сборка и разборка NPC (nuclear Pore Complex) в месте слияния между митотической ядерной оболочкой и вышележащей мембраной веретена обнаруживаются во время раннего синхронного митоза у эмбрионов D. melanogaster (Kiseleva et al., 2001). Предсказанные укороченные продукты гена sbr во время оогенеза могут иметь альтернативные функции. Продукты гена sbr участвуют в сложных процессах, которые могут быть относительно независимыми от транскрипции. Эти процессы обеспечиваются циклической трансляцией мРНК факторов, участвующих в репликации ДНК, ядерных делениях и динамике цитоскелета.

A variety of SBR isoforms allow each of them to be part of a particular RNP complex that includes mRNAs for nuclear divisions in early embryos. mRNAs with different functional and regulatory capabilities require specialized RBPs. SBR (Dm NXF1) interacts with the subunits of the origin recognition complex, which associates with replication origins and initiates the pre-replication complex assembly (Kopitova et al., 2016). The sbr gene is essential for nuclear divisions (Golubkova and Mamon, 2012) and the cytoskeleton (Mamon et al., 2017). Since the RNA-binding protein SBR is found not only in the nucleus or nuclear envelope but also in the cytoplasm, it may be part of RNP complexes storing organ-specific long-living mRNAs. The nucleoporin-binding domains in the truncated SBR isoforms allow for the localization of the corresponding RNP complexes in the nuclear envelope (Mamon et al., 2017).

Разнообразие изоформ SBR позволяет каждой из них быть частью определенного комплекса RNP, включающего мРНК для ядерных делений у ранних эмбрионов. мРНК с различными функциональными и регуляторными возможностями требуют специализированных RBP. SBR (Dm NXF1) взаимодействует с субъединицами комплекса распознавания ориджинов, который связывается с ориджинами репликации и инициирует сборку пререпликационного комплекса (Копитова и др., 2016). Ген sbr необходим для ядерных делений (Голубкова, Мамон, 2012) и цитоскелета (Мамон и др., 2017). Поскольку РНК-связывающий белок SBR находится не только в ядре или ядерной оболочке, но и в цитоплазме, он может входить в состав РНП-комплексов, хранящих органоспецифические долгоживущие мРНК. Нуклеопорин-связывающие домены в усеченных изоформах SBR позволяют локализовать соответствующие RNP-комплексы в ядерной оболочке (Mamon et al., 2017).

The sbr mRNA diversity in ovaries reflects the specificity of the early embryonic development of D. melanogaster. Various maternal mRNAs are involved in control of the synchronous nuclear divisions. The specific RNAbinding proteins are needed for functioning of different mRNA. This also applies to the SBR proteins.

Разнообразие мРНК sbr в яичниках отражает специфику раннего эмбрионального развития D. melanogaster. Различные материнские мРНК участвуют в контроле синхронных ядерных делений. Специфические РНК-связывающие белки необходимы для функционирования различных мРНК. Это также относится к белкам SBR.

**Conclusion**

A comparison of the organisms of different taxa has demonstrated that the complexity of organization during evolution arises not due to an increase in the number of protein-coding genes. One of the reasons is alternative transcript processing, which generates functional diversity from the same or similar number of genes. A variety of mechanisms providing for the complexity of the transcriptome ensures a precise and coordinated regulation of organ-specific functions through a combination of cis-acting elements and trans-acting factors. The D. melanogaster sbr (Dm nxf1) gene has proven to be an excellent model for investigating these mechanisms, potentially leading to the emergence of multiple products with various functions.

Сравнение организмов разных таксонов показало, что усложнение организации в ходе эволюции возникает не за счет увеличения числа белок-кодирующих генов. Одной из причин является альтернативный процессинг транскриптов, который создает функциональное разнообразие из того же или сходного количества генов. Разнообразие механизмов, обеспечивающих сложность транскриптома, обеспечивает точную и скоординированную регуляцию органоспецифических функций за счет сочетания цис-действующих элементов и транс-действующих факторов. Ген sbr (Dm nxf1) D. melanogaster оказался отличной моделью для исследования этих механизмов, что потенциально может привести к появлению множества продуктов с различными функциями.