**Практическое занятие № 2**

**В КУРСЕ «Регуляторные геномные последовательности»**

**ТЕМА: КОМПЬЮТЕРНЫЕ РЕСУРСЫ, ПОЗВОЛЯЮЩИЕ АНАЛИЗИРОВАТЬ РЕГУЛЯТОРНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ мРНК (5’-НТП И 3’-НТП).**

(проводила Е.В.Игнатьева)

**Цели**

1) Знакомство с содержанием базы TRANSIG (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/index.html>)

2) Анализ потенциальной функциональной значимости фрагментов 5’-НТП мРНК с помощью опции “BLAST-search” по базе TRSIG (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/blast.html>)

3) Анализ последовательности мРНК с помощью программы AUG\_hairpin (http://wwwmgs.bionet.nsc.ru/mgs/programs/aug\_hairpin/)

В ходе практического занятия нужно выполнить задания, ответы занести в таблицы. Методичку с заполненными таблицами отправить на адрес **eignat@bionet.nsc.ru** в виде приложения (файл в MS word). Не забыть указать фамилию!!!

# Этап 1. Знакомство с содержанием базы TRANSIG

База TRANSIG (TRanslational SIGnals) (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/index.html>) – новая версия базы TRSIG. Содержит данные о посттранcкрипционных сигналах экспрессии (=сигналы экспрессии). Эти сигналы (участки последовательности мРНК) могут обеспечивать дополнительный контроль экспрессии трансгена на уровне трансляции. К числу таких сигналов можно отнести трансляционные энхансеры.

Задание 1

На главной станице базы TRANSIG (<http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/index.html>) перейти к гиперссылке TRANSIG\_OBJECT. Будет виден стандартный интерфейс поисковой системы SRS. Перейти к полю «OS» (= objest species) и пролистать (browse) записи в этом поле с помощью опции пролистывания “List values”, как это делалось на предыдущем занятии. Заполнить таблицу 1

**Таблица 1**. Количество входов в базе TRANSIG, относящихся к различным видам организмов.

|  |
| --- |
| В базе TRANSIG найдено  \_\_\_\_\_\_\_\_ входов, относящихся к человеку,  \_\_\_\_\_\_\_\_ входов, относящихся к мыши.  \_\_\_\_\_\_\_\_ входов, относящихся к вирусу гриппа (influenza a virus)  \_\_\_\_\_\_\_\_ входов, относящихся к мягкой пшенице (triticum aestivum)  \_\_\_\_\_\_\_\_ входов, относящихся к гороху (pisum sativum) |

Задание 2

На главной станице базы TRANSIG перейти по ссылке TRANSIG\_ENCHANCER. Перейти к полю «Keywords» и пролистать записи в этом поле с помощью опции пролистывания “List values”, как это делалось в предыдущем задании. Ответить на вопрос: «Сколько записей в базе TRANSIG описывают трансляционные сигналы, реагирующие на тепловой шок (heat shock), гипоксию (hypoxia). Сколько в базе участков внутренней посадки рибосомы (IRES) ??? Заполнить Таблицу 2

**Таблица 2**. Количество некоторых трансляционных сигналов в базе TRANSIG.

|  |
| --- |
| В базе TRANSIG найдено  \_\_\_\_\_\_\_\_ трансляционных сигналов, реагирующих на тепловой шок (heat shock),  \_\_\_\_\_\_\_\_ трансляционных сигналов, реагирующих на гипоксию (hypoxia).  \_\_\_\_\_\_\_\_ участков внутренней посадки рибосомы (IRES) |

# Этап 2. Анализ потенциальной функциональной значимости фрагментов 5’-НТП мРНК с помощью опции “BLAST-search” по базе TRSIG (http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/blast.html)

Задание 1

Подготовить в фаста формате нуклеотидные последовательности 5'-нетранслируемых районов мРНК двух генов

Первый ген

Ген *HSPA1A* (Heat shock protein family A (Hsp70) member 1A) вида *Homo sapiens*.

Для этого необходимо обратиться к карточке гена *HSPA1A* человека в базе EntrezGene (https://www.ncbi.nlm.nih.gov/gene/) (идентификатор Gene ID: 3303).

Перейти к разделу «mRNA and Protein(s)», перейти по ссылке, обозначающей идентификатор мРНК (NM\_005345.6).

Определить, ориентируясь на описание последовательности в карточке, позицию старта трансляции (первый AUG=ATG кодон). Эта информация представлена в поле «FEATURES». Например, в карточке с NM\_005345.6 в поле «FEATURES CDS» имеется такая запись:

“CDS 215..2140” -- Эта запись означает, что нуклеотидная последовательность с позициями 1 – 214 является 5’-нетранслируемой последовательностью (5’-UTR) для данной мРНК.

Далее перейти в режим FASTA, скопировать последовательность в редактор Word.

Для анализа с помощью программы BLAST-search” по базе TRSIG необходимо подготовить последовательность, включающую **ТОЛЬКО** 5’-нетранслируемую последовательность (то есть убрать все нуклеотиды, начиная со старта трансляции).

Последовательность будет иметь вид:

>NM\_005345.6 Homo sapiens heat shock protein family A (Hsp70) member 1A (HSPA1A), mRNA

AACGGCTAGCCTGAGGAGCTGCTGCGACAGTCCACTACCTTTTTCGAGAGTGACTCCCGTTGTCCCAAGGCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTTTCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCCAGCCCCCAATCTCAGAGCGGAGCCGACAGAGAGCAGGGAACCGGC

Первая строчка начинается с символа ">" и содержит комментарий (текстовое описание последовательности). Вторая строчка содержит нуклеотидную последовательность.

Второй ген

*Rbm3 -* Mus musculus RNA-binding motif protein 3

Последовательность берем из базы Genbank (карточка с идентификатором AY052560.1) по следующему URL:

https://www.ncbi.nlm.nih.gov/nuccore/AY052560

Аналогичным образом (как было описано для первой последовательности) готовим последовательность 5’-НТП в фаста формате:

>AY052560.1 5’UTR Mus musculus RNA-binding motif protein 3

CACGAGGCATTACCCAGTAGACTCTTACACTGAACCTCCCAATCCTAACTTAGTACATGCATACAATAAATGCTCACTAAGTACTGATATATGATTAAAACACTTGAAAACAAACAAACTGTAATGTACATGACATAATCGTGATGAGCAGTAAGTAGGAACATCTTCTTGGCAAAGGAAGGAGTCAAACAATGAGCAGGGGCTGTGAGTGTGGGTCAGTGGTAGCACAGCATGCATGAGGCCCTATATTTGGTCTCTAGCAACACACATAAAAGATGAACAAAACCAGACAGAGGAAGAGGTATGTCGAGAGTCTTAAATAAAACAGGTCAGATCTAAAACAACTTTTAGGAGAAACAGAAAGATACTTACTTTCGGTCACTCCCCCAATAGCAAGAGAAATAATAGCTAAAACGTTCTCACATGCGGAATGATTTATAATTTCTTCTTCCAGAACACCTCTGAAAGCTTGGTCAAGGGTACATTTTTTTTCATTTTCACTGCCAGGTAACTGACTGAAGGCAGTCAACAACGGCTTGATATTTTTGTTATTCAAGGCTTCTCTGGTAGATTTCGTAAATCGCGTCCGCGCCTCCGGCAAACTGAAAAGCGCTTTATCAGGCGTCTTCCCGCGCCGCAGTCTCTCTGTTCTCCCGGTTCCTTCGAGCTCGTCGTCTCTGCCGTCCTCTGACTTTTAATTTCCAGGACTTGCCTTCTGCC

Задание 2

Поиск потенциальных сигналов трансляции в 5’-НТП мРНК с помощью опции “BLAST-search” по базе TRSIG (http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/blast.html)

Эта программа позволяет выявлять в лидерных участках мРНК сходство с трансляционными сигналами, аннотированными в базе TRSIG.

1. Зайти на главную страницу “BLAST-search” по базе TRSIG <http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/blast.html>
2. Скопировать из текстового файла последовательность 5’-НТП мРНК *HSPA1A* в фаста формате и вставить в окно для поиска. Для того, чтобы снизить уровень ложнопредсказанных результатов, выбрать значение "**word size**" равным **10** nucleotides.
3. Запустить программу, использовав кнопку «Search Blast».
4. Занести название найденной последовательности (сигнала) в таблицу 3 вместе с характеристикой найденного сигнала. Характеристику сигнала можно получить по гиперссылке, соответствующей идентификатору сигнала (в данном случае это будет S0141). А далее надо идти по гиперссылке, находящейся в поле OBJID (в данном случае, HSHSP70UTR5)
5. Аналогичным образом провести поиск с помощью программы “BLAST-search” для второй последовательности (AY052560.1) и занести результат **(только сигналы длиннее 20 нуклеотидов !!!)** в таблицу. Длину (протяженность) сигнала смотрим в части интерфейса, где приведено выравнивание двух последовательностей (участка последовательности мРНК, которую Вы анализируете, и сигнала, найденного в базе).

**Таблица 3**. Результат поиска программой BLAST-search в базе TRSIG

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Идентификатор сигнала | Идентификатор из поля OBJID | Название гена, вид организма | Участок мРНК (написать, где расположен найденный участок в - 5-НТП или в  3-НТП?) | Краткое описание сигнала (индуктора), регулирующего интенсивность трансляции из поля KEYWORD |
| Результат анализа последовательности из карточки NM\_005345.6 | | | | |
| S0141 |  |  |  |  |
|  |  |  |  |  |
| Результат анализа последовательно гена из карточки AY052560.1  **(только сигналы длиннее 20 нуклеотидов !!!)** | | | | |
| S0218 |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# Этап 3. Анализ последовательности мРНК с помощью программы AUG\_hairpin: предсказание шпилечной структуры ниже AUG кодона

Программа AUG\_hairpin осуществляет предсказание шпилечной структуры мРНК ниже AUG кодона. Анализ необходимо провести для мРНК гена *HSPA1A* (карточка NM\_005345.6), которую Вы анализировали на предыдущем этапе.

Задание 1. Перейдите к программе AUG\_hairpin (http://wwwmgs.bionet.nsc.ru/mgs/programs/aug\_hairpin/) и ознакомьтесь с форматом входных данных, нажав кнопку “Example”. Подготовьте фрагменты последовательностей, окружающих старт трансляции (первый ATG, позиция 215 в карточке NM\_005345.6), а также второй AUG кодон (**позицию второго AUG в карточке NM\_005345.6 найдите сами**)

Выполните анализ последовательно, сначала для фрагментов последовательности, окружающих старт трансляции мРНК гена *HSPA1A*, а затем для фрагментов последовательности, окружающих второй AUG кодон.

Результаты анализа вставьте в таблицу 4. (**см. ниже**)

**Таблица 4.** Результаты анализа программой AUG\_hairpin.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Отчет программы (в текстовом виде) | Изображение вторичной структуры мРНК | **Вывод** |
| Example | Energy of secondary structure = -26.8 kcal/mol  Position of Hairpin start: 16  Energy of double strands in Hairpin: -6.8 kcal/mol | http://wwwmgs.bionet.nsc.ru/cgi-bin/mgs/common/gettemp?file=JxDKJ5.gif&mime=image/gif | Шпилечная структура ниже AUG кодона найдена  Суммарная энергия вторичной структуры  -26.8 kcal/mol  Позиция начала шпильки - 16  Энергия участка двойной спирали в шпильке -6.8 kcal/mol |
| Старт трансляции мРНК гена HSPA1A (первый ATG, позиция 215 в карточке NM\_005345.6) |  |  | Шпилечная структура ниже AUG кодона  **найдена (не найдена)**  Суммарная энергия вторичной структуры **- ???**  Позиция начала шпильки **- ???**  Энергия участка двойной спирали в шпильке - ??? |
| Второй AUG кодон мРНК гена HSPA1A (позиция ??? в карточке NM\_005345.6) |  |  | Шпилечная структура ниже AUG кодона  **найдена (не найдена)**  Суммарная энергия вторичной структуры **- ???**  Позиция начала шпильки **- ???**  Энергия участка двойной спирали в шпильке - ??? |