Практическое занятие 2

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

**Цель** – работа с базой данных TRSIG и обучение поиску участков нуклеотидных последовательностей в мРНК, представляющих собой потенциальный сигнал регуляции экспрессии.

В ходе практического занятия нужно выполнить описанные ниже упражнения (этапы 1-4) и приготовить отчет, включающий:

(а) фамилию,

(б) результаты выполнения этапов 2, 3 и 4 в виде таблиц (образцы таблиц приведены в конце разделов 2, 4). Это будут участки гомологии и интерпретация функциональной роли последовательностей, выявленных на этапах 1 и 2 , а также ответы на вопросы о содержании базы TRANSIG (этап 4)

Отчет нужно отправить на адрес **eignat@bionet.nsc.ru** в виде приложения (файл в MS word).

# Этап 1. Создание выборки нуклеотидных последовательностей 5'-нетранслируемых районов мРНК

**Цель** Создать выборку нуклеотидных последовательностей 5'-нетранслируемых районов мРНК 2-х генов в формате фаста.

В качестве источника данных необходимо использовать GenBank (http://www.ncbi.nlm.nih.gov/nuccore/ ) . В качестве примера необходимо найти последовательности из следующих карточек базы GenBank:

|  |  |
| --- | --- |
| BC062708 | Homo sapiens ferritin, light polypeptide, mRNA (cDNA clone MGC:71996 IMAGE:4774403), complete cds |
| BC016009 | Homo sapiens ferritin, heavy polypeptide 1, mRNA (cDNA clone MGC:27328 IMAGE:4666545), complete cds |
|  |  |
|  |  |

Необходимо скопировать карточки полностью в любой текстовый редактор и вытащить из нуклеотидных последовательностей, приведенных в карточках, те участки, которые соответствуют 5'- нетранслируемым последовательностям (= 5’-UTR = 5’-НТП)

Для выбранных карточек 5’- нетранслируемые последовательности будут найдены, если взять участки нуклеотидной последовательности между 5'- концом мРНК (первый нуклеотид в последовательности) и началом CDS (стартовый кодон трансляции)

Эта информация представлена в поле «Feature Table».

Например, в карточке с номером BC062708 в поле «Feature Table» имеется такая запись:

 “CDS 203..730”

Эта запись означает, что нуклеотидная последовательность с позициями 1 – 202 является 5’-нетранслируемой последовательностью (5’-UTR).

Данные (две последовательности 5’-НТП) для дальнейшего анализа необходимо подготовить в фаста-формате. Фаста-формат представления нуклеотидной последовательности включает две строчки. Первая строчка начинается с символа ">" и содержит комментарий (текстовое описание последовательности). Вторая строчка содержит нуклеотидную последовательность. См. пример ниже

**Example**:

>BC062708; human ferritin light chain

ggggcagttc ggcggtcccg cgggtctgtc tcttgctcca acagtgtttg gacggaacag

atccggggac tctcttccag cctccgaccg ccctccgatt tcctctccgc ttgcaacctc

cgggaccatc ttctcggcca tctcctgctt ctgggacctg ccagcaccgt ttttgtggtt

agctccttct tgccaaccaa cc

>BC016009; human ferritin heavy chain

ggggagacgt tcttcgccga gagtcgtcgg ggtttcctgc ttcaacagtg cttggacgga

acccggcgct cgttccccac cccggccggc cgcccatagc cagccctccg tcacctcttc

accgcaccct cggactgccc caaggccccc gccgccgctc cagcgccgcg cagccaccgc

cgccgccgcc gcctctcctt agtcgccgcc

# Этап 2. Выявление участков гомологии в 5'-нетранслируемых районах мРНК паралогичных генов (ferritin light chain и ferritin heavy chain) с помощью программы BLAST (bl2seq)

1. Зайдите на страничку сервера института NCBI, через которую доступна программа выравнивания нуклеотидных последовательностей BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi )
2. Выберите опцию **Global Align** и перейдите по этой ссылке.
3. Вы получили интерфейс программы, позволяющей строить выравнивание двух последовательностей.
4. Вставьте в пустые окошки для последовательностей полученные на предыдущем шаге 5'-НТП мРНК генов в фаста формате.

Первое окошко: BC062708; human ferritin light chain

Второе окошко: BC016009; human ferritin heavy chain

1. Запустите программу, нажав кнопку «BLAST»
2. Перенесите результат (участок гомологии) в текстовый файл

**Результат нужно вставить в отчет в текстовом виде здесь:**

**Таблица 0.** Районы гомологии в последовательностях мРНК человека

|  |
| --- |
|  |

# Этап 3. Интерпретация функции участков гомологии в 5'-нетранслируемых районах мРНК

Анализ потенциальной функциональной значимости фрагментов 5’-НТП будем проводить 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 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 одну из полученных последовательностей районов гомологии, из числа выявленных на предыдущем этапе и в фаста формате вставить в окно для поиска. **Рекомендуется выбрать участок мРНК ferritin light chain, human (позиции 28-58).** Для того, чтобы снизить уровень ложнопредсказанных результатов, выбрать значение "**word size**" равным **10** nucleotides .
3. Запустить программу, использовав кнопку «Search Blast».
4. Занесите результат поиска в отчет (см. шаблон таблицы ниже) вместе с характеристикой найденного сигнала.

Для того, чтобы охарактеризовать выявленные сигналы, необходимо воспользоваться сходством c коллекцией сигналов, имеющихся в базе TRSIG. Для этого необходимо воспользоваться ссылкой, предлагаемой записью «sequence ID» (имеет вид S0153, S0222 и т.д.), имеющейся на странице с результатами поиска по BLAST. Затем перейти по ссылке Object ID (OBJID) на входной странице базы данных с нуклеотидными последовательностями. Будет получено описание сигнала. Сигналом может быть свет, окислительный стресс и т.д. Перейти по ссылке Experiment database ID (имеет вид E0086 E0087 E0102 E0103). Будут получены описания экспериментов, с помощью которого выявлен сигнал.

1. Занесите результат поиска в отчет вместе с характеристикой найденного сигнала.
2. Оформить результат выполнения этапа 3 в виде таблицы (таблица в конце методички)
3. Проанализируйте с помощью программы “BLAST-search” базы TRSIG полную последовательность из карточки базы GenBank с идентификатором AH002239.2. Эта карточка содержит Rattus norvegicus preproparathyroid hormone gene, complete cds . **Вырезать последовательность 5’ нетранслируемого района (5’-НТП) в этом случае не нужно, подавайте на анализ программой “BLAST-search” всю последовательность из карточки AH002239.2.** Результатом поиска будут несколько сигналов.Обратите внимание, что первый (по порядку представления в выдаче программы) сигнал существенно отличается от остальных (в лучшую сторону) по значениям скора (Score) и уровня значимости (E-Value). Внесите данные о первом найденном сигнале в отчет по форме, представленной в таблице 1.

# Этап 4. Знакомство с содержанием базы TRANSIG

TRANSIG – одна из версий базы TRSIG, по которой шел поиск с помощью программы “BLAST-search” на предыдущем этапе сегодняшнего занятия.

1. Зайти на страницу http://srs6.bionet.nsc.ru/srs6bin/cgi-bin/wgetz?-page+top+-id+ZGe1t\_67G
2. Выбрать базу «Translation Signals». Раскрыть, нажав на значок «+»
3. Перейти по ссылке TRANSIG\_ENH
4. Перейти к полю «Keywords» и пролистать записи в этом поле с помощью опции пролистывания “List values”, как это делалось на предыдущем занятии.
5. 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) ??? Результат занести в отчет (Таблица 2)

**Отчет**

**Таблица 1.** Результаты поиска сигналов в базе TRSIG с помощью опции “BLAST-search”

Выявлены сигналы (последовательности 5-НТП либо 3-НТП), гомологичные фрагментам последовательностей, поданных на вход.

|  |  |
| --- | --- |
| Последовательность, которая подавалась на вход (Query) | Результат поиска по BLAST в базе TRSIG (Sbjct) |
| Название гена, для которого проводился анализ 5-НТП | Вид организма | Позиции  | После-дова-тельность | Иден-тификатор сигнала | Идентификатор из поля OBJID | Ген, вид организма | Участок (5-НТП или 3-НТП) | Краткое описание сигнала (индуктора), регулирующего интенсивность трансляции \* |
| ferritin light chain | human | 29-58 |  |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Rattus norvegicus preproparathyroid hormone gene, complete cds | rat |  |  |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

\* - Получив запись с идентификатором типа «ID S0155» необходимо перейти по ссылке OBJID, и далее взять информацию из поля KEYWORD

**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) |