

## **1. Наименование работы**

Аннотация транскриптома паразитического червя *Opistorchis felineus*

## **2. Состав коллектива исполнителей**

Афонников Д.А., Помазной М.П.

## **3. Контактное лицо (ФИО, адрес электронной почты)**

Афонников Дмитрий Аркадьевич, [ada@bionet.nsc.ru](mailto:ada@bionet.nsc.ru)

## **4. Научное содержание работы**

### **1. Постановка задачи**

У человека паразитируют не менее 13 видов червей (гельминтов), в том числе и паразиты печени семейства *Opisthorchiidae*. По различным оценкам в мире от них страдает около 40 млн людей. Члены семейства описторхид вызывают у человека описторхозы – паразитарные заболевания, приводящие к поражению гепатобилиарной системы. Описторхозы характеризуются длительным течением заболевания и тяжелыми последствиями, одними из которых являются онкологические заболевания желчных путей и печени.

Паразит *O.felineus* встречается преимущественно в странах восточной Европы, в том числе в России, Украине и Белоруссии. Также описторхозы широко распространены и среди населения, проживающего в бассейне реки Обь (Западная Сибирь). Лечение же паразитозов, вызываемых трематодами на сегодняшний день сводится в основном к применению единственного препарата, хотя ведутся клинические испытания, связанные с апробацией новых потенциально эффективных препаратов.

В ходе выполнения проекта проведено исследование белок-кодирующей фракции транскриптома *O. felineus* методами биоинформатики с целью выявления возможных генов, кодирующих белки с выраженными антигенными свойствами как возможных кандидатов для разработки методов иммунодиагностики паразитарной инфекции.

### **2. Современное состояние проблемы**

Чтобы получить информацию о транскрипционно активных генах исследуемого организма нужно провести функциональную аннотацию полученных последовательностей. Как правило, на данном этапе исследователь имеет дело с большим массивом данных, который обрабатывается специальными компьютерными программами.

Современные методы биоинформатики позволяют предсказывать иммуногенные свойства белков, используя информацию из баз данных,

содержащих последовательности эпитопов и антигенов, методы распознавания последовательностей, информацию о доменной структуре белков и их локализации. Использование методов биоинформатики становится все более актуальным в последнее время, с появлением возможности полного секвенирования геномов патогенных организмов. Компьютерные методы на основе анализа полного набора последовательностей генов, кодирующих белки, позволяет идентифицировать наиболее вероятные антигены. Такой подход получил название обратной разработки вакцин. Впервые такой подход был использован для разработки вакцин против менингококковой инфекции серогруппы В. В настоящее время для компьютерной идентификации белков, которые являются антигенами, разработано большое число подходов, прежде всего, распознающих последовательности эпитопов в белках-кандидатах. Эти методы основаны на использовании последовательностей белков, а так же их пространственных структур. Тем не менее, для более точного определения белков-кандидатов для антигенного скрининга, может быть важной вся доступная информации о структуре, функции, взаимодействиях белка.

### **3. Полученные результаты**

Для анализа нами были использованы две EST библиотеки *O. felineus*, полученные как коллективом участников проекта в Институте Цитологии и генетики СО РАН. Предварительная обработка нуклеотидных последовательностей включала удаление фрагментов векторных последовательностей из ДНК клонов и других неинформативных участков.

Для установления более отдаленной гомологии и идентификации функциональных доменов транслированные в 6 рамках нуклеотидные последовательности анализировали программой Interproscan. Поиск гомологичных белков в публичной базе Non-redundant при помощи программы BLASTX позволил для 158 уникальных последовательностей библиотеки (60%), представляющих собой контиги и синглеты, получить информацию о предположительной функциональной значимости.

Для нуклеотидных последовательностей, представленных в выборке БКО1-3D, мы разработали алгоритм определения возможной аминокислотной последовательности с использованием выравнивания с гомологичным белком, полученного программой BLASTX. Эта программа в результате выравнивания выдает участок сходства последовательностей, его положение в тестируемой нуклеотидной последовательности, а так же направление и смещение считывания белка (открытую рамку считывания, для которой был идентифицирован гомологичный белок).

Алгоритм идентифицирует положение гомологичного участка в аминокислотной последовательности, транслированной в рамке, для которой

был обнаружен гомолог (рис. 1). Затем последовательность расширяется на С'-конце до первого встретившегося стоп-кодона. Если таковой не находится, последовательность расширяется до правого конца транслированной последовательности.

Далее последовательность расширяется на N' конце рамки влево до первого метионина, следующего за первым встретившимся справа стоп-кодом. Если таковой не встретился, последовательность расширяется до левого конца транслированной последовательности.

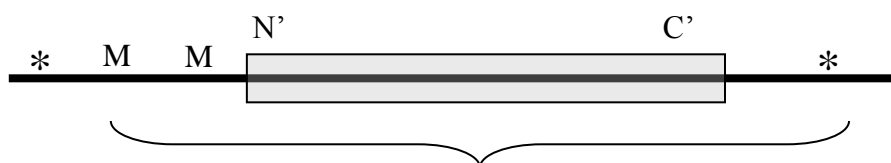


Рисунок 1. Схема алгоритма расширения кодируемой последовательности белка. Транслированная последовательность показана жирной линией. Звездочками показано расположение стоп-кодонов вне участка гомологии. Буквы М обозначают положение метионина (стартового кодона). Серый прямоугольник показывает участок гомологии с известной белковой последовательностью, выявленный программой BLASTX. Фигурной скобкой показана расширенная последовательность белка. N' и C' обозначают соответствующие концы участка гомологии.

Для того чтобы отобрать белки-кандидаты в качестве антигенов-мишеней мы использовали следующий подход. При помощи программы netMHCII-2.2 для предсказания пептидов связывающихся с человеческим лейкоцитарным антигеном (MHC class II) мы определяли для каждой белковой последовательности выборки число участков с сильной аффинностью к антигену. Анализировались 28 типов аллелей: H-2-IAb, H-2-IA<sub>d</sub>, HLA-DPA101-DPB10401, HLA-DPA10103-DPB10201, HLA-DPA10201-DPB10101, HLA-DPA10201-DPB10501, HLA-DPA10301-DPB10402, HLA-DPB10301-DPB10401, HLA-DQA10101-DQB10501, HLA-DQA10102-DQB10602, HLA-DQA10301-DQB10302, HLA-DQA10401-DQB10402, HLA-DQA10501-DQB10201, HLA-DQA10501-DQB10301, HLA-DRB10101, HLA-DRB10301, HLA-DRB10401, HLA-DRB10404, HLA-DRB10405, HLA-DRB10701, HLA-DRB10802, HLA-DRB10901, HLA-DRB11101, HLA-DRB11302, HLA-DRB11501, HLA-DRB30101, HLA-DRB40101, HLA-DRB50101.

Затем белки ранжировались по полному числу участков связывания. Среди этих последовательностей были отобраны 30, которые имеют наибольшее число предсказанных пептидов. Полученные в результате последовательности будут в дальнейшем исследованы экспериментальными методами.

Отбирались 30 белков с наибольшим числом участков.

#### 4. Эффект от использования кластера в достижении целей работы

Вычисление на кластере позволило ускорить обработку данных по аминокислотным последовательностям *O. felineus*.

#### 5. Иллюстрации, визуализация результатов

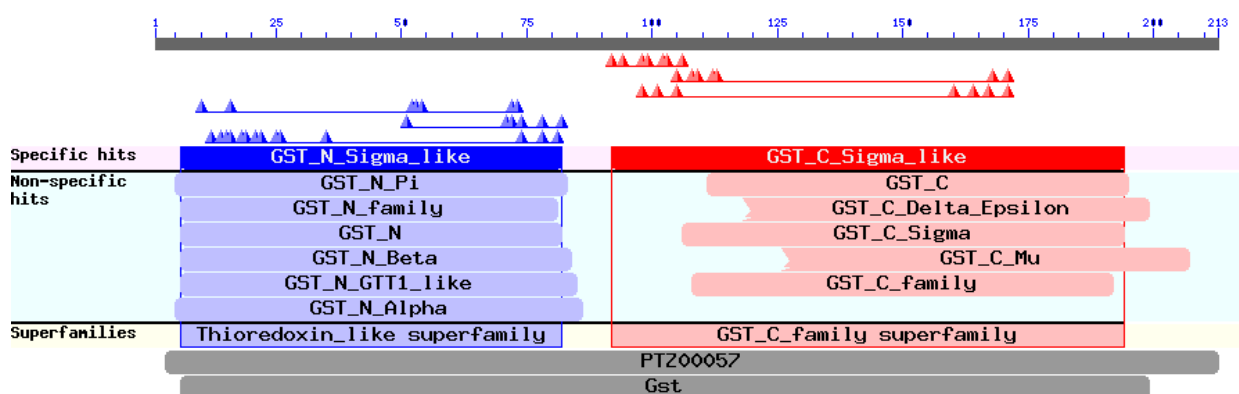


Рисунок 2. Доменная структура аминокислотной последовательности GST\_SIGMA, одного из иммунологически важных белков *O. felineus*. Серой линейкой сверху обозначена последовательность белка с разметкой по номерам позиций. Под ней треугольниками показаны позиции, формирующие активные центры белка в соответствующих доменах (если они известны). Ниже в линейке специфических совпадений (Specific hits), если они были обнаружены, показаны домены в базе данных CDD, имеющие высокую гомологию с последовательностью. Различные домены обозначены различными цветами. Ниже приведены совпадения с доменами, имеющих средний уровень сходства – неспецифические совпадения (Non-specific hits). Они так же раскрашены разными цветами в соответствии с разной функцией доменов. Еще ниже представлены совпадения с известными белковыми суперсемействами (Superfamilies).

#### 5. Перечень публикаций, содержащих результаты работы

Afonnikov D.A., Pomaznoy M.Yu., Katokhin A.V., Mordvinov V.A., Kolchanov N.A., Kostryukova E.S., Levitskii S.A., Selezneva O.V., Chukin M.M., Larin A.K., Lazarev V.N., V.M. Govorun. (2011) Analysis of the transcriptome of the human parasitic trematode *Opisthorchis felineus*. Proc. Of the 5th Moscow Conference on Computational Molecular Biology, Moscow, Jul 21-24, 2011, p. 27.

#### 6. Ваши впечатления от работы вычислительной системы и деятельности ИВЦ НГУ, а также Ваши предложения по их совершенствованию.

Замечаний нет.