Практическое занятие 2

**Реконструкция сети белок-белковых взаимодействий для набора генов из базы KEGG с помощью программы STRING, их анализ в системе Cytoscape и сопоставление с данными, полученными из GeneMania**

**Цель** – знакомство с интернет-доступными ресурсами KEGG pathway и **STRING**, а также с системой Cytoscape. Освоение методов реконструкции и анализа сетей взаимодействий между генами.

В ходе практического занятия нужно выполнить описанные ниже упражнения (этапы 1, 2 и 3) и приготовить отчет, (см. образец отчета в конце документа).

Отчет нужно отправить на адрес **eignat@bionet.nsc.ru** в приложении к письму в виде файла в MS word.

# Этап 1. Подготовка списка генов для дальнейшего анализа из базы KEGG Pathway

**Цель этапа:** Иметь в наличии список генов, участвующих в метаболическом /сигнальном пути, на основе данных, полученных из базы KEGG pathway.

**Последовательность операций:**

Использовать список генов, полученный на занятии №1: 14 генов, кодирующих субъединицы РНК-полимеразы II. Этот список был получен из базы KEGG pathway (часть 2.1 Transcription, раздел «RNA polymerase»). Были взяты 14 первых генов, имеющих названия вида POLR2\*\*\*.

# Этап 2. Экстракция данных о белок-белковых взаимодействиях с помощью программы STRING

**Цель:** Получить данные о белок-белковых взаимодействиях между белками, кодируемыми генами из списка А (14 генов, кодирующих субъединицы РНК-полимеразы II человека) с помощью программы / ресурса **STRING** (http://string-db.org/). Экстрагировать данные в текстовом виде.

**Последовательность операций:**

1. Открыть программу **STRING** (<http://string-db.org/>), на панели слева выбрать режим “Multiple proteins” (Рисунок 1, скриншот 1).
2. Вставить список генов (список А) в окошко программы **STRING, выбрать вид организма Homo sapiens** (Рисунок 1, скриншот 1), нажать кнопку «Search».
3. В следующем диалоговом окне будет представлено описание генов. Нужно проверить, правильно ли система распознала гены, поданные на вход. Затем нажать кнопку “CONTINUE” (Рисунок 1, скриншот 2).

|  |
| --- |
|  |
| Рисунок 1 Интерфейс системы STRING. Первый этап работы – задан список генов. |

1. Будет получена сеть (Рисунок 2). Выбрать вкладку Data Settings. Выбрать опцию «No more than 5 interactors», остальные опции не менять (использовать опции по умолчанию), и нажать клавишу “Update settings”. Таким образом, в сеть будут добавлены 5 дополнительных генов/белков (этих генов не было в исходном списке).
2. Экспортировать данные в текстовом виде. Для этого воспользоваться опциями Tables/Exports -> Export your current network: ... as simple tabular text output (**download**). Полученные текстовые данные загрузить в Exel.

|  |
| --- |
|  |
| Рисунок 2 Интерфейс системы STRING. Второй этап работы – настраиваем режим построения сети так, чтобы система выдавала 5 добавочных объектов. |

# Этап 3. Анализ сети белок-белковых взаимодействий с помощью системы Cytoscape

**Цель:** Получить изображение сети в программе Cytoscape, а также статистику по сети.

1. На основе данных, полученных из STRING, подготовить таблицу в программе Exel, включающей только данные о связях типа «experimentally\_determined\_interaction» для последующей загрузки в Cytoscape. Для того, чтобы получить такую таблицу, нужно из исходного файла (экспортированного из STRING) удалить колонки с данными (скорами) по всем остальным типам связей. Далее необходимо удалить строчки, содержащие пары объектов, у которых скор белок-белкового взаимодействия (experimentally\_determined\_interaction) равен 0. Пример формата см. в Таблице 1.

Таблица 1. Вид файла Exel, который будет содержать данные о белок-белковых связях и будет импортирован в Cytoscape.

|  |  |  |  |
| --- | --- | --- | --- |
| #node1 | node2 | Score | Interaction type |
| POLR2J | POLR2C | 0.999 | experimentally\_determined\_interaction |
| POLR2K | POLR2I | 0.954 | experimentally\_determined\_interaction |
| GTF2F1 | POLR2L | 0.89 | experimentally\_determined\_interaction |
| POLR2J3 | POLR2K | 0.722 | experimentally\_determined\_interaction |

1. Загрузить таблицу с данными о белок-белковых взаимодействиях в Cytoscape. Для этого открыть программу Cytoscape. Далее использовать команды File -> Import -> Network -> File -> Ok.
2. Поработайте над дизайном изображения сетей в Cytoscape. Установите стиль «default» либо «Biopax\_SIF», тогда будут видны названия генов.
3. Сохраните сети в формате .cys (предлагается системой Cytoscape).
4. Примените различные режимы расположения объектов на плоскости с помощью команды Layout -> yFilesLayout (Organic, Circular, etc). Убедиться, что все пары объектов соединены не более чем одной связью.
5. Получить статистику для сети. Для этого воспользоваться опцией программы Cytoscape: Tools -> NetworkAnalyser -> NetworkAnalysis -> AnalyseNetwork.
6. Сколько связных вершин в сети? Имелись ли в исходном списке гены, для которых не нашлись белок-белковые взаимодействия ??
7. Обратитесь к нижнему окну интерфейса Cytoscape (Рисунок 3). Там содержатся статистические данные для каждого гена (=вершины сети). Выполните сортировку объектов (вершин) по количеству связей (от большего к меньшему). Для этого нажмите мышкой по заголовку колонки «Degree». (Рисунок 3).

|  |
| --- |
|  |
| Рисунок 3. Интерфейс программы Cytoscape. В нижнем окне рамкой обведены поля, содержащие статистические данные об объектах. Стрелкой выделена колонка Degree, содержащая информацию о количестве связей у каждого белка. Объекты отсортированы по количеству связей. |

1. Перенесите статистику из нижнего окна (Рисунок 3) в файл Exel. Для этого воспользуйтесь простыми командами Shift ↓ ->(выделение) и Ctrl C (копирование в буфер обмена), а затем перейдите в Exel и примените команду Ctrl V (вставить). К сожалению, названия колонок при этом не копируются и в Exel не переносятся. Необходимо присвоить имя «Degree» колонке, в которой содержались данные о количестве связей у каждого объекта и по которой проводилась сортировка. Сохраните файл Exel.
2. Сокращенный вариант таблицы с отсортированным по числу связей списком объектов нужно вставить в отчет. Пример формата таблицы см. ниже в образце отчета (Таблица 1а).
3. Cytoscape позволяет изменять окраску (форму, размер) объектов в зависимости от атрибутов (свойств) объектов. Для того, чтобы разметить двумя разными цветами гены из исходного списка и новые гены, необходимо выполнить следующее. Подготовьте таблицу в Exel, содержащего атрибуты всех генов/белков. Для генов из исходного списка значение атрибута будет равно нулю, а для новых генов – единица.

Таблица 2. Вид файла Exel, который будет содержать данные об атрибутах (свойствах) объектов (вершин) сети, то есть генов. Эти данные будут импортированы Cytoscape.

|  |  |
| --- | --- |
| shared name | Atribute (new gene) |
| POLR2C | 0 |
| POLR2B | 0 |
| POLR2A | 0 |
| POLR2J | 0 |
| И так | 0 |
| далее | 0 |
| Все | 0 |
| гены | 0 |
| Из | 0 |
| исходного | 0 |
| списка | 0 |
| Новый ген | 1 |
| Новый ген | 1 |
| Новый ген | 1 |
| Новый ген | 1 |
| Новый ген | 1 |

Импортируйте таблицу в Cytoscape с помощью команд File -> Import -> Table -> File.

Далее в диалоговом окне рядом с вопросом : «Where to import table data:» нужно выставить ответ «To selected network only».

Откройте в программе Cytoscape вкладку Style (Рисунок 4, позиция 1). Нажатием на панель (Рисунок 4, позиция 2) выберите любой стиль, где видны названия объектов (например, BioPAX\_SIF). В строчке «Fill Color» нажмите на клетку во второй колонке (Map) (Рисунок 4, позиция 3). Нужно выбрать информационное поле (колонку), содержащую нужный атрибут, в соответствии с которым будет осуществляться раскраска объектов. В нашем случае нужно выбрать атрибут с названием «new gene» (Рисунок 4, позиция 4). Далее необходимо выбрать способ раскраски : “Discrete mapping”. Присвойте объектам, имеющим атрибут 0 любой цвет (например, желтый). Присвойте объектам с атрибутом 1 любой другой цвет (например, синий). В результате будет получена «двухцветная» сеть, как это показано на рисунке 4.

|  |
| --- |
|  |
| Рисунок 4. Интерфейс программы Cytoscape, позволяющий задавать цвет (форму, размер) объектов, согласно их атрибутам (свойствам). |

1. Экспортируйте и сохраните изображение сети в формате png. Для этого воспользоваться опцией программы Cytoscape: File -> Export -> Network View as Graphics
2. Сравните характеристики сети, полученной на сегодняшнем занятии (занятие №2) с характеристиками сети, полученной на предыдущем занятии (занятие №1). Результат сравнения представьте в виде таблицы (образец см. в конце методички - Таблица 2а)

Образец отчета практическому занятию №2

(должен быть представлен в формате MS Word)

Таблица 1а. Список генов/белков в сети, построенной по данным системы STRING, отсортированный по количеству связей

Пояснение: Таблица должна содержать данные как о 19 генах (14 генов из запроса + 5 добавочных гена). Колонку 4 необходимо заполнить вручную.

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 2 | 3 | 4 |
| Порядковый номер гена | Ген/белок  (колонка shared name или name) | Количество связей в сети (колонка  «Degree») | 0 – ген из исходного списка, 1- новый ген |
|  | POLR2C | 9 | 0 |
|  | POLR2B | 8 | 0 |
|  | POLR2A | 7 | 0 |
|  | POLR2J | 6 | 0 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  | 0 (не было в сети) | 0 |
|  |  |  | 1 |
|  |  |  | 0 |
|  | ELL | 4 | 1 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  |  | 0 |
|  |  |  | 1 |
|  |  |  | 1 |
|  |  |  | 1 |

Изображение сети

Пояснение: Необходимо вставить сохраненное Вами изображение сети (формат png) в документ MS Word и сделать соответствующую подпись. Ниже приведено изображение абстрактной сети. Полученная Вами сеть должна содержать гораздо больше вершин и не иметь дублирующийся ребер (так как в выдаче системы STRING такая информация не представлена). Кроме того, цветом должны быть выделены вершины, соответствующие пяти новым генам.

|  |
| --- |
|  |
| Рисунок 1а. Сеть белок-белковых взаимодействий между генами/белками, субъединицами РНК полимеразы II человека, а также добавочными 5-ю генами/белками, выявленными системой STRING. Гены, кодирующие субъединицы РНК полимеразы II, обозначены ??? цветом. Добавочные пять генов, выявленные системой STRING, обозначены ??? цветом. |

**Таблица 2а** Основные характеристики сетей белок-белковых взаимодействий, построенных на основе данных двух систем STRING и GeneMania.

|  |  |  |
| --- | --- | --- |
| Характеристика сети | GeneMania | STRING |
| Два белка/гена, имеющие наибольшее количество связей (в порядке убывания количества связей) |  |  |
| Добавочные пять генов, то есть генов, которых не было в исходном списке. Привести в порядке убывания количества связей. |  |  |

**Вывод по таблице (выбрать нужный вариант ответа):**

А) «Приведенные в таблице 2а характеристики сетей одинаковые»

Б) «Приведенные в таблице 2а характеристики сетей различны, STRING и GeneMania выдают разную информацию о белок-белковых взаимодействиях»