Практическое занятие 1

**Реконструкция сети белок-белковых взаимодействий для набора генов из базы KEGG с помощью программы GeneMANIA и последующий анализ программой Cytoscape.**

**Цель** – знакомство с интернет-доступными ресурсами KEGG pathway и GeneMANIA, 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. Экстракция списка генов из базы KEGG pathway для дальнейшего анализа.

**Цель этапа:** Сгенерировать список генов, участвующих в метаболическом /сигнальном пути, на основе данных, полученных из базы **KEGG** pathway (http://www.genome.jp/kegg/pathway.html).

**Последовательность операций:**

1. В качестве объекта изучения предлагается использовать информацию из раздела «RNA polymerase», который можно найти в части 2.1 базы KEGG pathway (2.1 Transcription) (Рисунок 1, скриншот1).
2. После перехода к диаграмме «RNA polymerase» необходимо выбрать вид организма *Homo Sapiens* (Рисунок 1, скриншот 2)
3. Перейти по ссылке «Pathway entry» (Рисунок 1, скриншот 3). В список для дальнейшего анализа (**список А**) необходимо включить гены, кодирующие субъединицы РНК-полимеразы II (названия генов вида POLR2\*\*\*). Должно получиться 14 генов. См. 14 первых генов на скриншоте 3 (Рисунок 1).

|  |
| --- |
|  |
| Рисунок 1. Экстракция списка генов из базы KEGG |

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

1. **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 GeneMANIA (http://genemania.org/). Экстрагировать данные в текстовом виде.

**Последовательность операций:**

1. Подготовить список официальных символов генов в столбик, каждый ген на отдельной строке:

POLR2B

POLR2A

POLR2C

POLR2J и т.д., всего 14 генов

1. Вставить список генов (список А) в окошко программы GeneMania (Рисунок 2, скриншот 1)
2. Выбрать вид организма «человек»
3. Раскрыть меню, позволяющее регулировать параметры поиска (кнопка с тремя точками справа от окошка со списком генов). Убедиться, что все типы связей будут присутствовать в поиске (галочки в квадратах слева) (Рисунок 2, скриншот 2)
4. Выбрать опцию «Customize advanced options» (Рисунок 2, скриншот 2)
5. Установить параметр «Max resultant attributes» равный 0 (нулю)
6. Установить параметр «Max resultant genes» равный 5 (Рисунок 2, скриншот 3). Это означает, что в результате поиска будут присутствовать пять дополнительных генов (этих генов не было в исходном списке).
7. Остальные опции не менять (использовать опции по умолчанию)

|  |
| --- |
|  |
| Рисунок 2. Последовательность действий при выполнении запроса к системе GeneMania |

1. Нажать кнопку ПОИСК (лупа)
2. Убедиться, что получено нужное количество объектов. Для этого воспользоваться режимом просмотра под номером два (кнопка с двумя стрелками, смотрящими вниз, Рисунок 3). Данная команда позволяет выдать результат в виде двух столбиков – гены из исходного списка (первый столбик), добавочные гены (второй столбик) (Рисунок 3).
3. Сохраните результат запроса в текстовом виде (Рисунок 3, нижняя кнопка с изображением дискеты), далее во всплывающем меню надо выбрать опцию «Network» (Рисунок 3)
4. Текстовые данные вставить в Exel и сохранить.
5. Для дальнейшей работы использовать связи, имеющие тип «Physical interactions»

|  |
| --- |
|  |
| Рисунок 3. Интерфейс GeneMania, отображающий результат запроса (сеть) и опция для экстракции данных о связях между парами объектов в текстовом виде. |

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

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

**Последовательность операций:**

1. Подготовить таблицу в программе Exel следующего вида (см. ниже, Таблица 1) в двух вариантах:

Файл А – включить все строчки, содержащие тип связей «Physical interactions»

Файл Б – за основу взять файл А. В программе Exel отсортировать данные сначала по столбцу 1, потом по столбцу 2, потом по столбцу «Weight». Вы увидите, что многие пары объектов содержатся в нескольких строчках, это значит, что в базе системы GeneMania между некоторыми объектами содержалось несколько связей с разным скором. Необходимо для каждой пары объектов оставить по одной связи с максимальным скором. Для этого после того как в программе Exel была выполнена сортировка сначала по столбцу 1, потом по столбцу 2, потом по столбцу «Weight», необходимо применить команду «удалить дубликаты по двум первым столбцам».

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Entity 1 | Entity 2 | Network group | Weight | Network |
| ELL | POLR2A | Physical interactions | 0.034499 | IREF-HPRD |
| ELL | POLR2B | Physical interactions | 0.042011 | IREF-SMALL-SCALE-STUDIES |
| ELL | POLR2C | Physical interactions | 0.019547 | IREF-SMALL-SCALE-STUDIES |
| ELL | POLR2D | Physical interactions | 0.060648 | IREF-SMALL-SCALE-STUDIES |

1. Открыть программу Cytoscape.

Загрузить табличный файл А в Cytoscape. Для этого использовать команды File -> Import -> Network -> File -> Ok.

В диалоговом окне (Рисунок 4) выбрать колонки из файла А(Exel), в которых содержатся названия вершин сети. Для этого присвоить колонкам из файла Файла А(Exel) следующие значения (= роли = смыслы):

Колонка «Entity 1» - «Source node» = узел(вершина), из которого идет связь

Колонка «Entity 2» - «Target node» = узел(вершина), на которую направлена связь

Колонка «Network group» - «Edge Attribute» = атрибут (свойство, характеристика) связи

Колонка «Weight» -«Edge Attribute» = атрибут (свойство, характеристика) связи

Колонка «Network» -«Edge Attribute» = атрибут (свойство, характеристика) связи

|  |
| --- |
|  |
| Рисунок 4. Диалоговое окно программы Cytoscape, позволяющее выбирать колонки, содержащие названия узлов, а также колонки, содержащие атрибуты (свойства) связей (ребер). |

1. Поработайте над дизайном изображения сетей в Cytoscape. Установите стиль «default» либо «Biopax\_SIF», тогда будут видны названия генов.
2. Сохраните сеть в формате .cys (предлагается системой Cytoscape).
3. Посмотреть на сеть, полученную в результате импорта файла А. Применить различные режимы расположения объектов на плоскости с помощью команды Layout -> yFilesLayout (Organic, Circular, etc). Убедиться, что многие пары объектов соединены несколькими связями.
4. Аналогичным образом загрузить табличный файл Б в Cytoscape. Сохранить.
5. Посмотреть на сеть, полученную в результате импорта файла Б. Применить 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.
6. Получить статистику для сети (из файла Б). Для этого воспользоваться опцией программы Cytoscape: Tools -> NetworkAnalyser -> NetworkAnalysis -> AnalyseNetwork.
7. Сколько связных вершин в сети? Имелись ли в исходном списке гены, для которых не нашлись белок-белковые взаимодействия ??
8. Обратитесь к нижнему окну интерфейса Cytoscape (Рисунок 5). Там содержатся статистические данные для каждого гена (=для каждой вершины сети). Выполните сортировку объектов (вершин) по количеству связей (от большего к меньшему). Для этого нажмите мышкой по заголовку колонки «Degree». (Рисунок 5).

|  |
| --- |
|  |
| Рисунок 5. 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 5) 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. Сколько вершин (генов/белков) представлено в данной таблице ? С помощью интерфейса GeneMania определите гены/белки, не имеющие белок-белковых взаимодействий. Внесите эти объекты в таблицу отчета (Таблица 1а), присвоив им значение ноль в колонке 3 (количество связей в сети)
4. Экспортируйте и сохраните изображение сети в формате png. Для этого воспользоваться опцией программы Cytoscape: File -> Export -> Network View as Graphics

Образец отчета практическому занятию №1

(должен быть представлен в формате MS Word)

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

Пояснение: Таблица должна содержать данные 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 |
|  |  |  | 1 |
|  |  |  | 0 |
|  | ELL | 4 | 1 |
|  |  |  | 0 |
|  |  | 0 (не было в сети) | 0 |
|  |  | 0 (не было в сети) | 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 ее построение велось на основе списка Б).

|  |
| --- |
|  |
| Рисунок 1а. Сеть белок-белковых взаимодействий между генами/белками, субъединицами РНК полимеразы II человека, а также добавочными 5-ю генами/белками, выявленными системой GeneMania. |