

Добро пожаловать в программу для реконструкции ассоциативных сетей биологических процессов и генетической регуляции в клетке

Associative Network Discovery (ANDCell)

Иванисенко В.А., Деменков П.С., Яркова Е.Э.

Группа компьютерной протеомики Лаборатории теоретической генетики ИЦиГ СО РАН

Компьютерная система ANDCell доступна для пользования только внутри сети Института Цитологии и Генетики. В настоящий момент разработана тестовая версия системы ANDCell.

Большая просьба, все замечания, пожелания и информацию об ошибках сообщать по email: andcell@pbiosoft.ru.

Аннотация

1. Работа с научной литературой необходима для проведения научных исследований в любой области знаний. В настоящее время в базе данных PubMed собрано более 15000000 (15 млн.) научных рефератов. Каждый год это число увеличивается примерно на 1 миллион публикаций. Анализ такого количества литературных данных, включающий в себя поиск источников, установление взаимосвязи между описанными фактами требует значительных временных затрат. Кроме того, современные подходы к анализу литературных данных требуют непременно обращения к другому важному источнику информации – фактографическим базам данных по молекулярной биологии и генетике (GeneBank, Swiss Prot и др.).

2. В Группе компьютерной протеомики ИЦиГ СО РАН разработана компьютерная система автоматической экстракции знаний из текстов рефератов PubMed и баз данных о молекулярно-генетических взаимодействиях, генетической регуляции, взаимосвязи полиморфизмов с заболеваниями, а также других ассоциациях между фактами, и их представления в виде ассоциативных семантических сетей. Экстракция информации проводится с использованием современных информационных технологий text-mining.

3. Создана база знаний, содержащая более 5 миллионов фактов, а также программа доступа к базе знаний и представления результатов в графическом виде, позволяющая строить ассоциативные сети. Вершинами таких сетей являются молекулярно-генетические объекты, заболевания и процессы, а связями между вершинами выступают типы взаимодействий и ассоциаций.

4. Рассматривались следующие объекты: гены, белки, микроРНК, метаболиты, молекулярные процессы, клеточные компоненты. Для молекулярных взаимодействий и ассоциаций представлены данные по типам клеток и организмам. Система обладает дружественным интерфейсом пользователя со многими функциями, включая пересылки на молекулярно-генетические базы данных, а также рефераты статей, из которых была экстрагирована информация.

Требования к оборудованию

Рекомендуемые требования к параметрам компьютера для пользования системой:

Операционные системы: Windows 2000 и выше

Объем оперативной памяти: не менее 512 Мб

Производительность процессора: не менее 1 ГГц

Видеокарта: с поддержкой OpenGL

При работе с системой с помощью портативных компьютеров (ноутбуков) после закрытия окон на основном окне программы могут оставаться «тени» окон, возникающие из-за особенности видеокарт. Эти артефакты никак не влияют на работу системы и их можно убрать, кликнув мышкой в любом месте основного окна программы.

Программные модули системы ANDCell

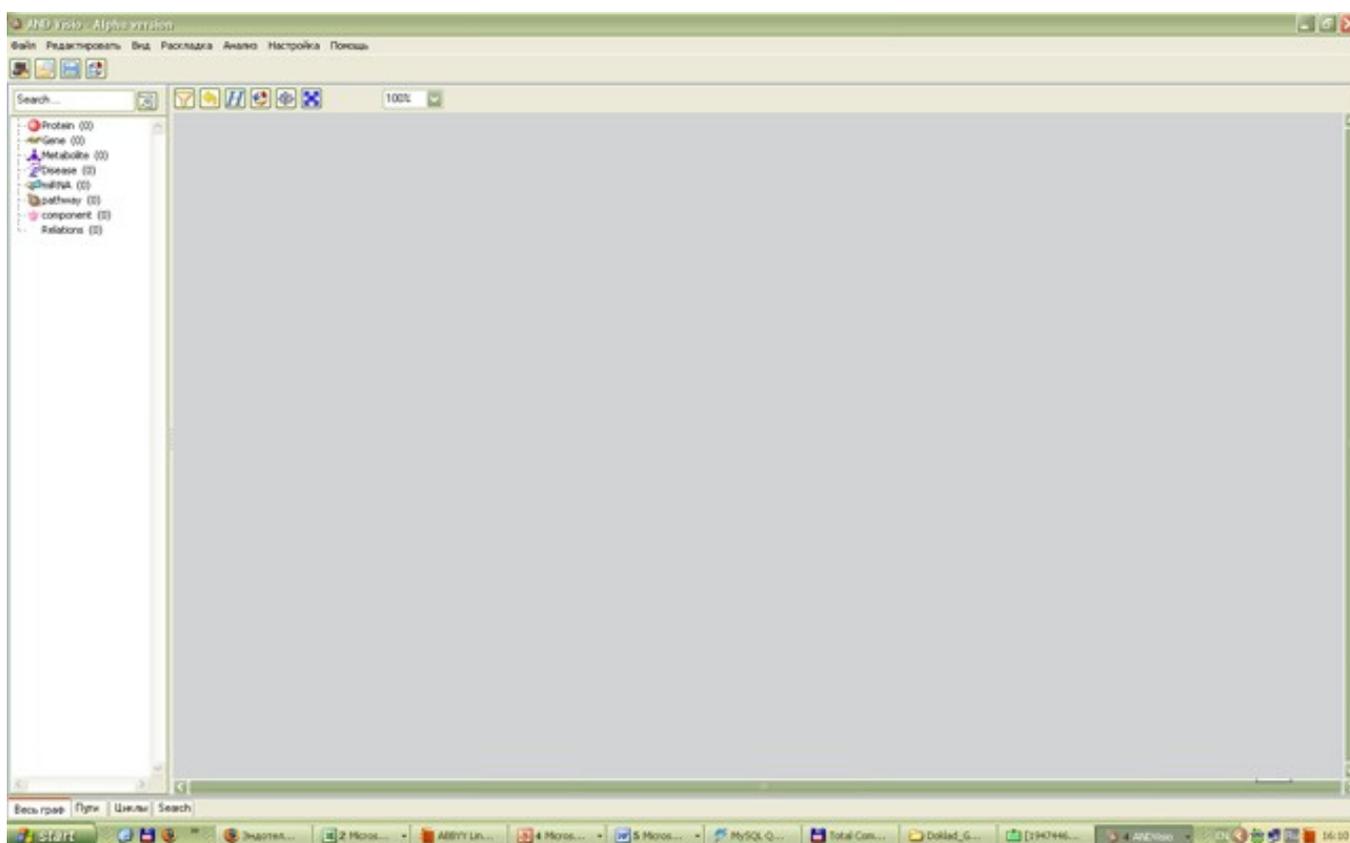
Система ANDCell состоит из базы знаний ANDCell и программы ANDVisio, обеспечивающей обращение к базе знаний и визуализацию результатов поиска по базе в виде ассоциативных сетей. База знаний ANDCell расположена на сервере Лаборатории теоретической генетики. Программа ANDVisio является клиентской частью системы, т.е. она устанавливается на персональном компьютере пользователя.

Установка системы ANDCell на персональном компьютере

Для установки программы ANDVisio требуется загрузить на свой компьютер архив с программой, находящийся по адресу <http://demosis.bionet.nsc.ru/andcell/ANDVisio.zip>. Далее следует распаковать содержимое архива. Запуск производится путем старта на исполнение файла ANDVisio.exe.

Работа в системе ANDCell

Для начала работы с программой запустите ANDVisio.exe.

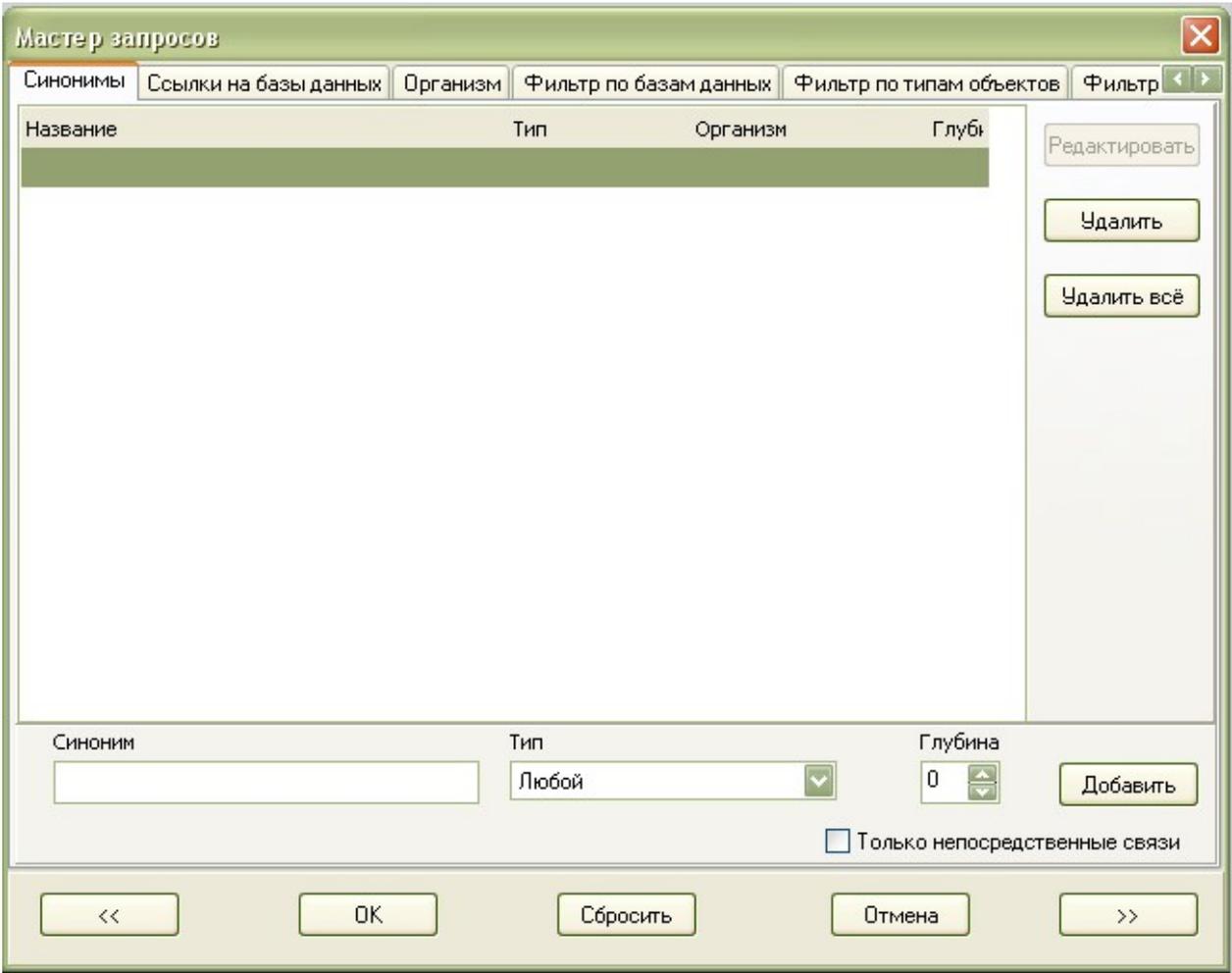


В левой панели обозначены все типы объектов, которые представлены в базе данных ANDCell. Если список объектов отсутствует, проверьте настройки программы (меню **Настройка**) В поле **Хост** должен быть указан адрес *samurai.bionet.nsc.ru*. Если в этом поле указан другой адрес, вручную укажите в нем *samurai.bionet.nsc.ru*, сохраните внесенные изменения и установите соединение с базой данных (меню **Файл** -> **Соединение**). Если список объектов в левой панели по-прежнему отсутствует, обратитесь к администратору базы данных ANDCell.

Составление запросов к базе знаний

Для запроса к базе данных ANDCell вызовите мастер запросов: меню **Файл ->Мастер**

запросов или иконка .



Мастер запросов

Синонимы Ссылки на базы данных Организм Фильтр по базам данных Фильтр по типам объектов Фильтр

Название	Тип	Организм	Глубина
----------	-----	----------	---------

Редактировать
Удалить
Удалить всё

Синоним:
Тип: Любой
Глубина: 0

Добавить

Только непосредственные связи

<< ОК Сбросить Отмена >>

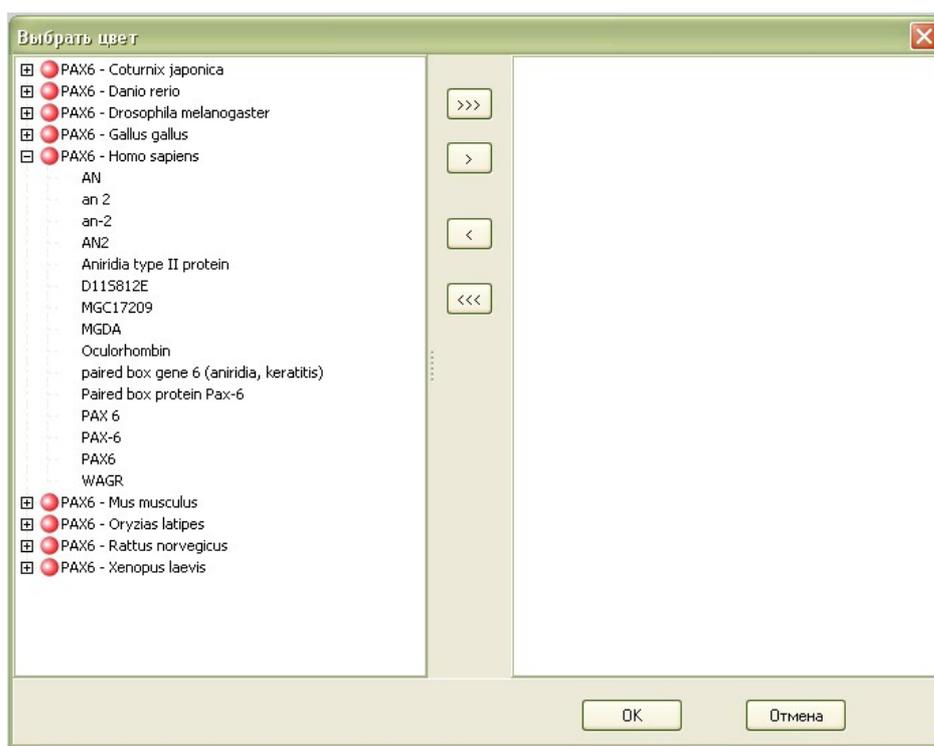
Для формирования запроса введите названия интересующего Вас объекта в поле **Синоним**. Символ * заменяет последовательность любых символов. Он используется в тех случаях, когда есть сомнения относительно точного имени. Так, например, для поиска факторов роста введите в поле **Синоним** строку: *growth factor*. Вам будет предоставлен весь список имен, в которых присутствует строка growth factor. Символ * может использоваться также между двумя терминами, тогда в список отобранных имен войдут все имена, которые содержат любые слова вместо звездочки.

В поле **Тип** определите тип интересующего объекта.

В поле **Глубина** определите глубину связей в сети. Глубина 0 обозначает, что в сети появятся только объекты, выбранные в запросе и связи между ними. Глубина 1 обозначает, что в сети появятся объекты, выбранные в запросе и все объекты, непосредственно связанные с ними. Глубина 2 обозначает, что в сети появятся все объекты из сети глубины 1, а также все объекты непосредственно связанные с ними и т.д. Так как от глубины сети экспоненциально зависит количество объектов и связей, реконструкция сетей глубины 2 и более возвращает слишком большое количество объектов, отображение которых будет неудобно для восприятия и анализа. Поэтому для работы с ассоциативными сетями рекомендуется использовать глубину 1 или 0.

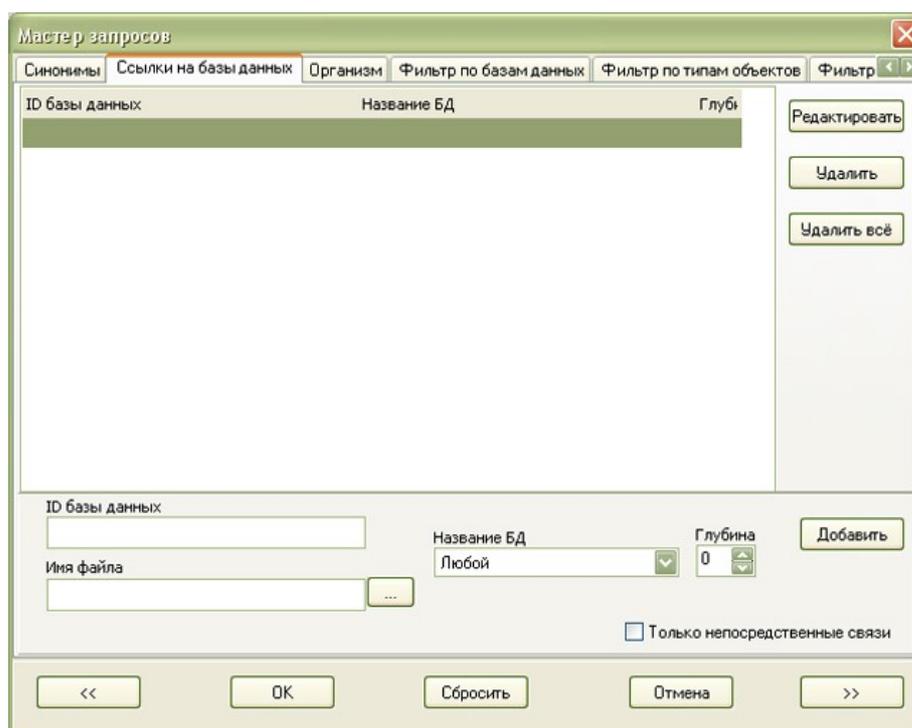
При глубине 1 в сети будут отображаться не только связи объектов 1-го порядка с объектами, выбранными в запросе, но и все связи между объектами 1-го порядка. Выберите пункт **Только непосредственные связи** для отображения только связей между объектами 1-го порядка и объектами, выбранными в запросе.

Для добавления объекта в запрос нажмите **Добавить**.

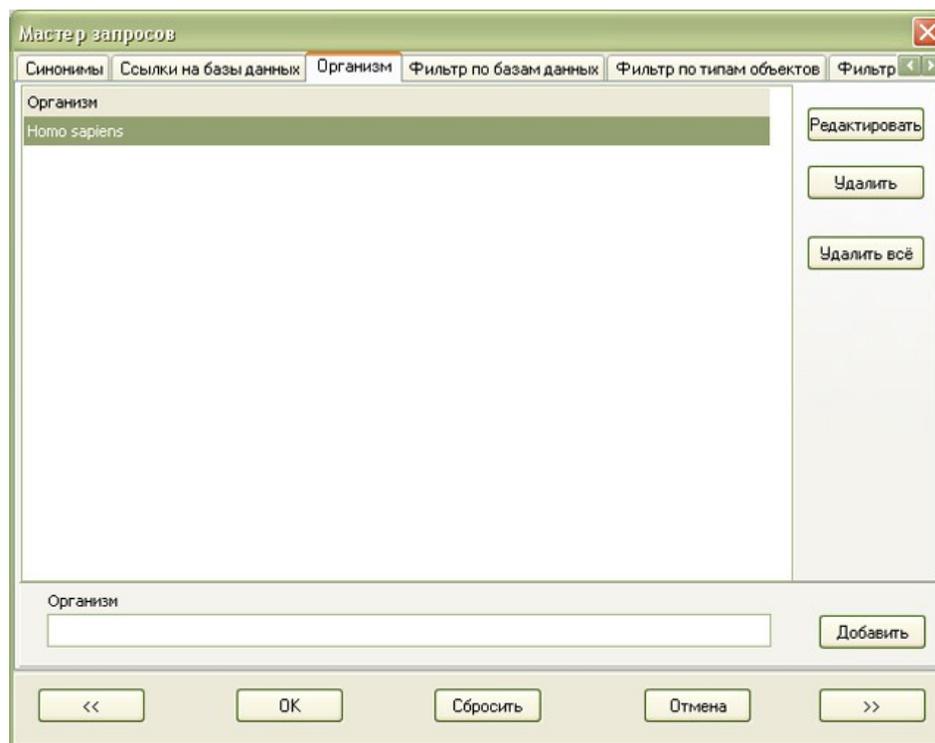


В левом списке выберите объекты, которые Вас интересуют и добавьте их в правое окно с помощью кнопки >. Чтобы удалить объект из запроса, нажмите <. Для добавления всех объектов из списка, нажмите >>. Для удаления всех объектов из запроса нажмите <<. Кликните на значок + рядом с символом объекта, для того чтобы увидеть список синонимов этого объекта.

Если вместо названий объектов Вы хотите использовать идентификаторы из баз данных, перейдите во вкладку **Ссылки на базы данных** окна **Мастер запросов**.

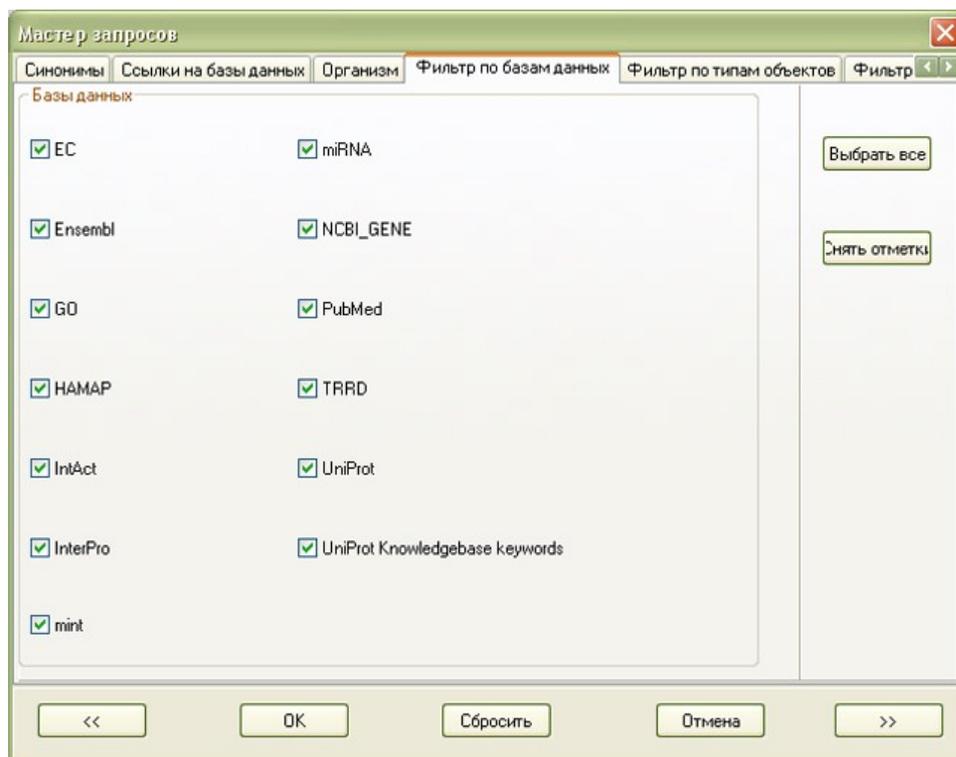


В поле **ID базы данных** введите идентификатор объекта, выберите название базы данных из списка и укажите глубину сети. Если Вы хотите задать список идентификаторов из одной базы данных, вы можете загрузить его из текстового файла через поле **Имя файла**. В текстовом файле идентификаторы должны быть перечислены в столбец, то есть разделены переносом строки. Для добавления объектов в запрос нажмите **Добавить**. С помощью кнопок **Редактировать**, **Удалить** и **Удалить все**, представленных в окне **Мастер запросов** можно удалять или редактировать добавленные в запрос объекты. Для применения к сети фильтра по организму, которому принадлежат объекты запроса, перейдите во вкладку **Организм** окна **Мастер запросов**



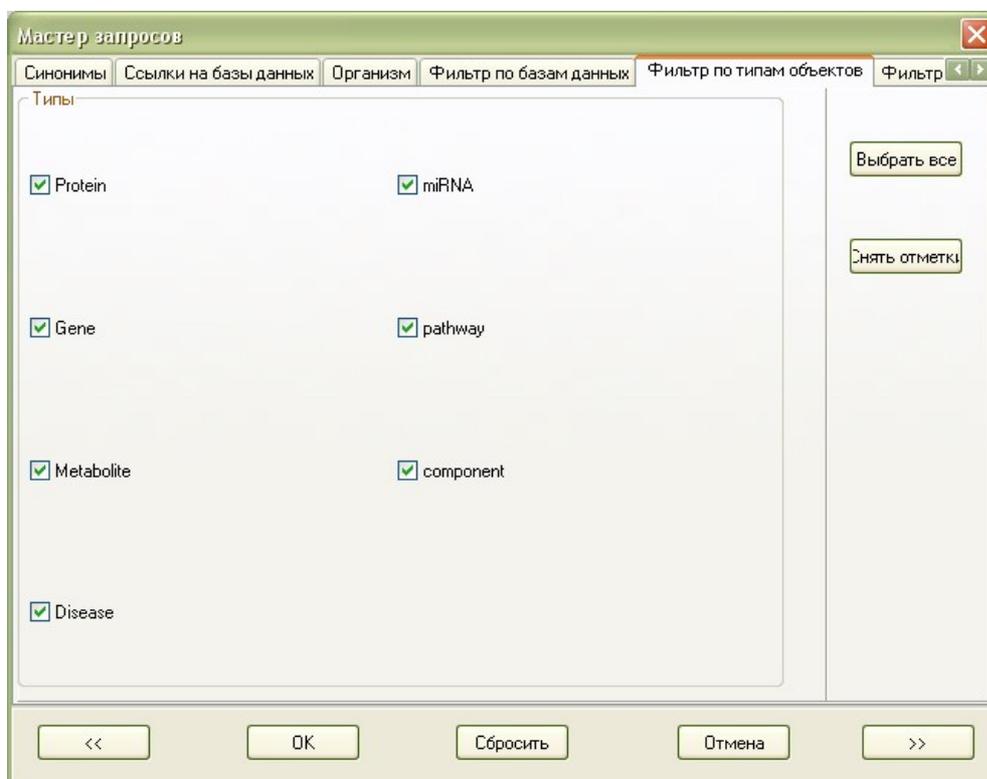
В фильтр по организмам автоматически добавляются те организмы, которым принадлежат выбранные Вами в запросе объекты. Для добавления дополнительного организма, введите его полное латинское или тривиальное название в поле **Организм** и нажмите **Добавить**. С помощью кнопок **Редактировать**, **Удалить** и **Удалить все** Вы можете редактировать или удалять фильтры по организмам.

Для применения к сети фильтра по базам данных, которые являлись источником информации для реконструкции сети, перейдите во вкладку **Фильтр по базам данных** окна **Мастер запросов**.



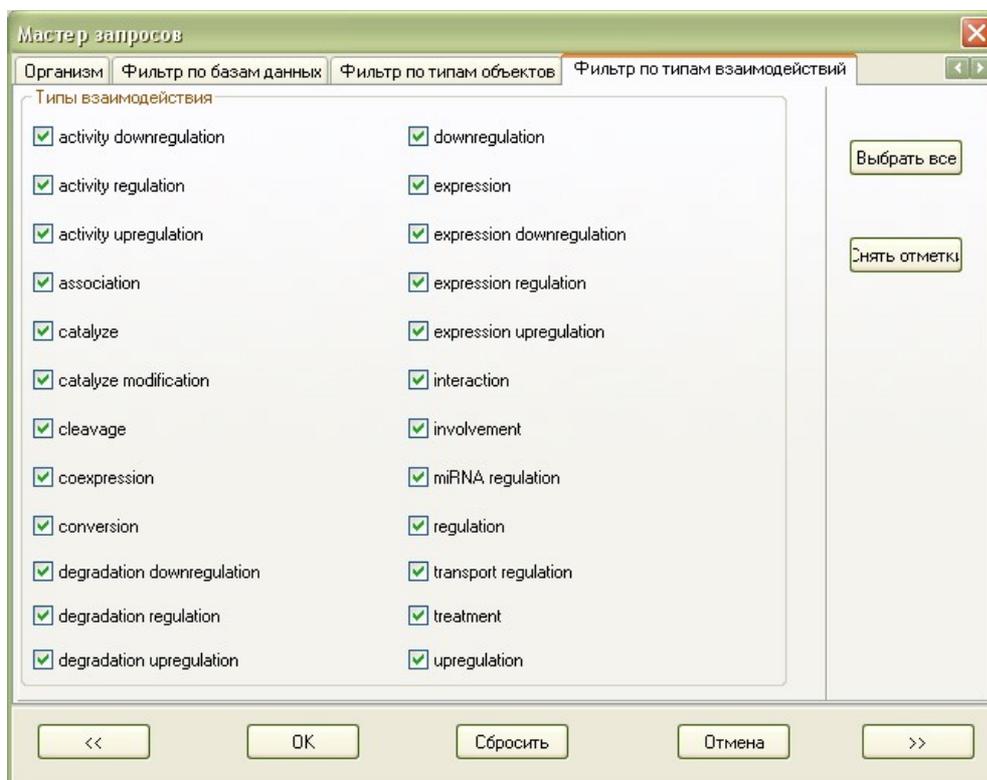
Отметьте галочками базы данных, информация из которых должна быть в реконструированной сети.

Для применения к сети фильтра по типам объектов, которые должны быть в реконструированной сети, перейдите во вкладку **Фильтр по типам объектов** окна **Мастер запросов**.



Отметьте галочками типы объектов, информация о которых должна быть в реконструированной сети.

Для применения к сети фильтра по типам взаимодействий, которые должны быть в реконструированной сети, перейдите во вкладку **Фильтр по типам взаимодействий** окна **Мастер запросов**.



Отметьте галочками типы взаимодействий, информация о которых должна быть в реконструированной сети.

Для отправки запроса в базу данных нажмите **ОК**.

Ассоциативные сети

В ассоциативных сетях представлены следующие типы объектов:



Белки



Гены



Заболевания



Низкомолекулярные вещества (Метаболиты)



МикроРНК



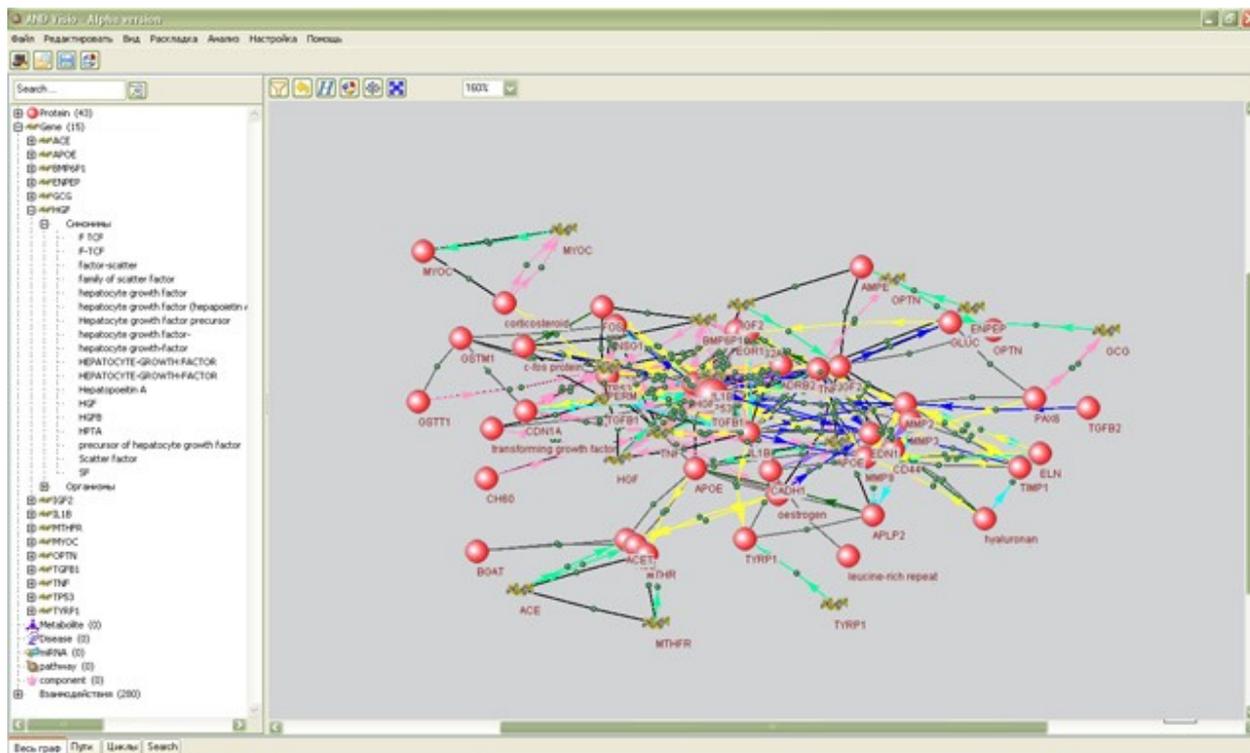
Биологические процессы



Клеточные компоненты



Взаимодействия



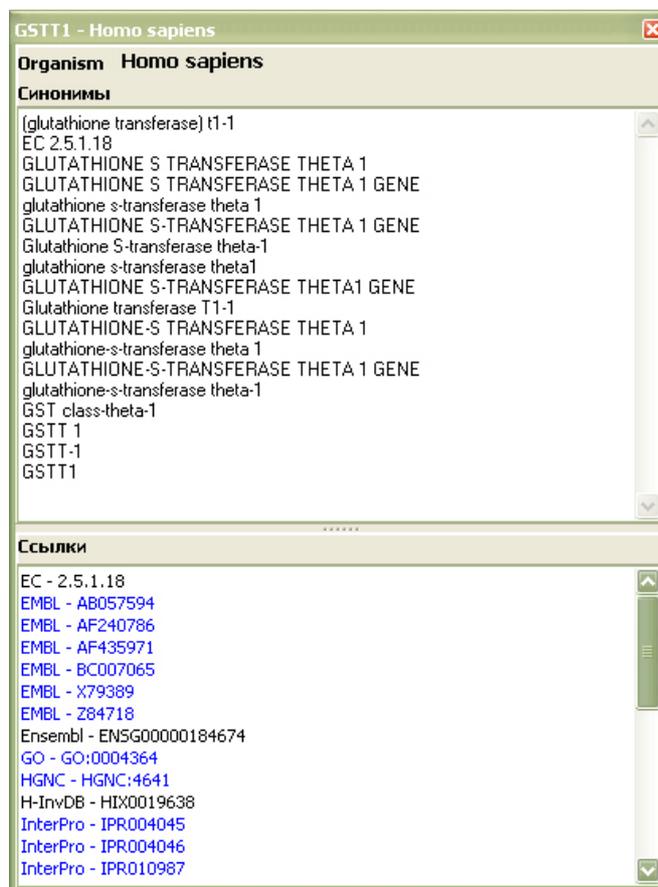
В левом окне программы в виде дерева указаны все объекты ассоциативной сети, сгруппированные по типу объектов. Нажмите + рядом с символом объекта для просмотра списка синонимов и организмов этого объекта. Дважды кликните на название объекта в левом окне для того чтобы выделить объект в сети и центрировать сеть относительно этого объекта. Дважды кликните на объекте в сети, чтобы найти этот объект в списке в левом окне программы.

Чтобы выделить объект кликните на нем левой кнопкой мыши. При выделении объекта все его связи подсвечиваются красным цветом. Для отключения подсветки снимите галочку в пункте меню **Вид -> Подсвечивать связи**.

Для того чтобы названия объектов не показывались, снимите галочку в пункте меню **Вид -> Показывать названия**.

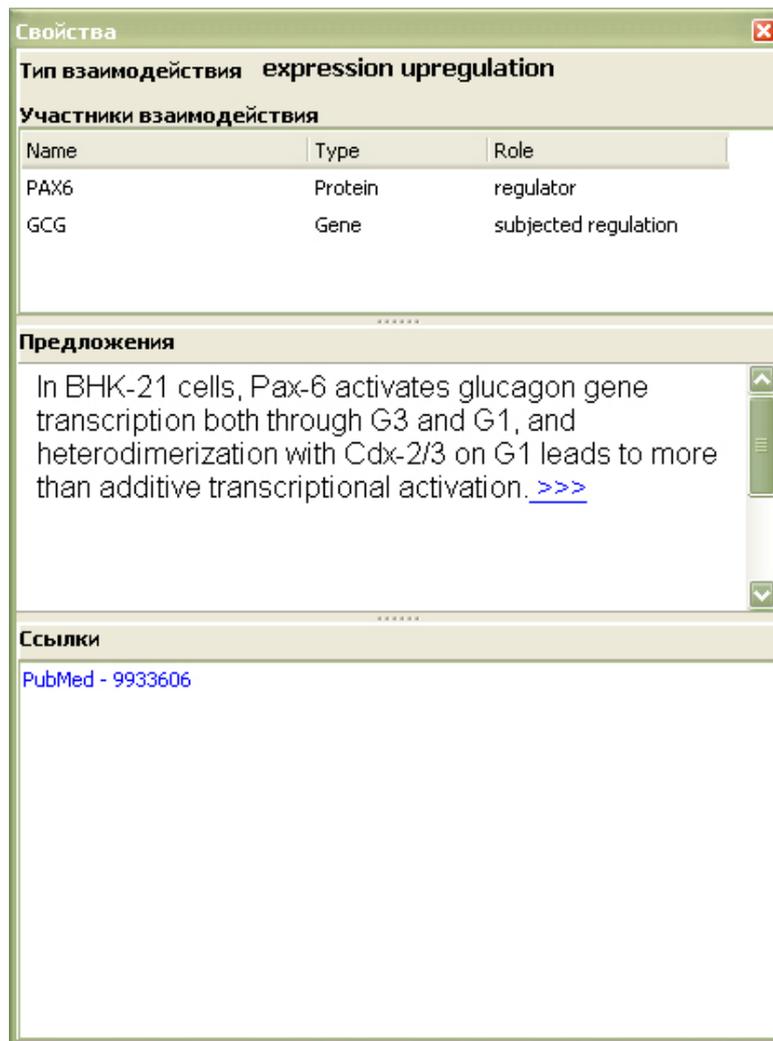
Чтобы уменьшить или растянуть сеть по размеру окна программы выберите пункт меню **Вид -> Вписать в окно**.

Кликните правой кнопкой на символе объекта в сети и выберите в контекстном меню пункт **Свойства** для получения информации о свойствах данного объекта.



В верхнем поле окна свойств объектов указаны синонимы данного объекта, в нижнем поле – ссылки на идентификаторы этого объекта в различных базах данных. Кликнув на ссылке, Вы можете через web-браузер получить доступ к карточке интересующего объекта в выбранной базе данных.

Кликните правой кнопкой на символе взаимодействия и выберите в контекстном меню пункт **Свойства** для получения информации о свойствах данного взаимодействия.



В верхнем поле окна свойств взаимодействий указан тип взаимодействия и его участники (названия, типы объектов, роли объектов во взаимодействии)

В нижнем поле указаны интерактивные ссылки на базу данных, которая была источником информации о взаимодействии.

В том случае, если информация о взаимодействии была извлечена из текстов рефератов PubMed с помощью text-mining, в среднем поле окна свойств взаимодействий указывается предложение, из которого извлечена информация о взаимодействии, а также ссылка на реферат PubMed, который содержит это предложение.

Раскладка сети на экране монитора

При выполнении запроса сети небольших размеров автоматически раскладываются на плоскости с целью наиболее удобной для пользователя визуализации связей. Для того чтобы повторить раскладку сети после перемещения объектов выберите пункт меню **Раскладка -> Перерасложить граф.**

В программе ANDVisio доступны два типа раскладки. Быстрая раскладка (**Раскладка -> Выбор типа раскладки -> Быстрая раскладка**) предназначена для черновой раскладки больших графов. Медленная раскладка (**Раскладка -> Выбор типа раскладки -> Медленная раскладка**) ориентирована на более тщательную раскладку графов небольшого размера.

Если Вам необходимо изменить раскладку фрагмента сети, выделите объекты, входящие в этот фрагмент, и выберите пункт меню **Раскладка -> Переразложить выделенные**.

С помощью пункта меню **Раскладка -> Спрямить связи** можно переразложить взаимодействия, не изменяя положения объектов таким образом, что взаимодействия будут находиться между объектами, которые в них участвуют.

Поиск объектов в сети

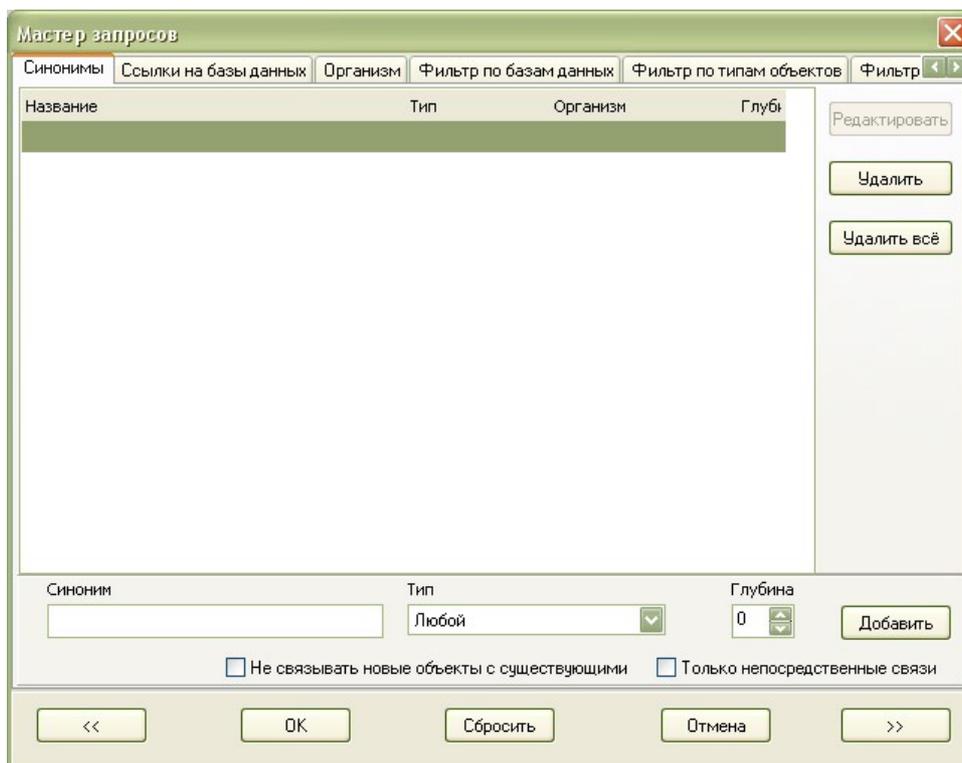
Для того чтобы найти объект в сети по синониму, введите его в поле **Search** над левым окном программы с деревом объектов. Результаты поиска будут показаны отдельным деревом во вкладке **Search** левого окна программы. Для возврата к полному дереву объектов перейдите во вкладку **Весь граф** левого окна программы.

Редактирование сети

Для того чтобы удалить объект или группу объектов из сети выделите их и выберите пункт меню **Редактировать -> Скрыть выделенные**.

Чтобы применить фильтр к реконструированной сети, выберите пункт меню **Редактировать -> Фильтр**. Выберите свойства фильтра как описано в параграфе *Составление запросов к базе данных ассоциативных сетей*.

Чтобы добавить объект в реконструированную сеть выберите пункт меню **Редактировать -> Добавить объекты**. Выберите объекты и укажите свойства фильтров как описано в параграфе *Составление запросов к базе данных ассоциативных сетей*.



Отметьте галочкой пункт **Не связывать новые объекты с существующими** для того, чтобы не показывать связи между добавленными в сеть объектами и объектами, ранее представленными в сети.

Чтобы расширить сеть новыми объектами, связанными с одним или несколькими объектами из реконструированной сети, выделите эти объекты и выберите пункт **Расширить** в контекстном меню, которое открывается по клику правой кнопкой мыши. Выберите объекты и укажите свойства фильтров как описано в параграфе *Составление запросов к базе данных ассоциативных сетей*. Отметьте галочкой пункт **Не связывать новые объекты с существующими** для того, чтобы не показывать связи между добавленными в сеть объектами и объектами, ранее представленными в сети.

Анализ ассоциативных сетей

Программа ANDVisio позволяет производить поиск замкнутых циклов и наикратчайших путей между двумя объектами в ассоциативной сети.

Для поиска кратчайшего пути выделите в сети два объекта (выделите один объект и, удерживая кнопку Shift, выделите второй). Выберите пункт меню **Анализ -> Поиск кратчайших путей**. Список найденных путей появится во вкладке **Пути** левого окна программы. Для просмотра участников пути нажмите на + рядом с названием пути. Чтобы

выделить путь на сети цветом и жирной линией кликните правой кнопкой на названии пути и выберите пункт **Подсветить** в контекстном меню. Выберите тип и цвет линии для подсветки пути. Для использования этой опции рекомендуется отключить подсветку связей выделяемого объекта, сняв галочку в пункте меню **Вид -> Подсвечивать связи**.

Для поиска кратчайших циклов, в которые входит определенный объект, выделите его и выберите пункт меню **Анализ -> Поиск циклов**. Список найденных циклов появится во вкладке **Циклы** левого окна программы. Для просмотра участников цикла нажмите на + рядом с названием цикла. Подсветка цикла возможна тем же способом что и подсветка пути.

Для поиска всех циклов внутри сети выберите пункт **Анализ -> Поиск фундаментальных циклов**. Список циклов появится во вкладке **Циклы** левого окна программы. Подсветка цикла возможна тем же способом что и подсветка пути.

Сохранение сетей

Чтобы сохранить построенную сеть, выберите пункт меню **Файл -> Сохранить как...** Сохранение возможно в 3 форматах: *AND Native* (используется по-умолчанию), *XML (ANDVisio)*, *XML (GeneNet)*.

Чтобы сохранить построенную сеть как картинку, выберите пункт меню **Файл -> Сохранить как изображение...** Сохранение возможно в форматах BMP, PNG и JPG.

Для открытия в программе ANDVisio, сохраненного ранее файла в формате *AND Native* или *XML (ANDVisio)*, выберите пункт меню **Файл -> Открыть...**