

Лекция 6

Реконструкция и анализ ассоциативных генных сетей

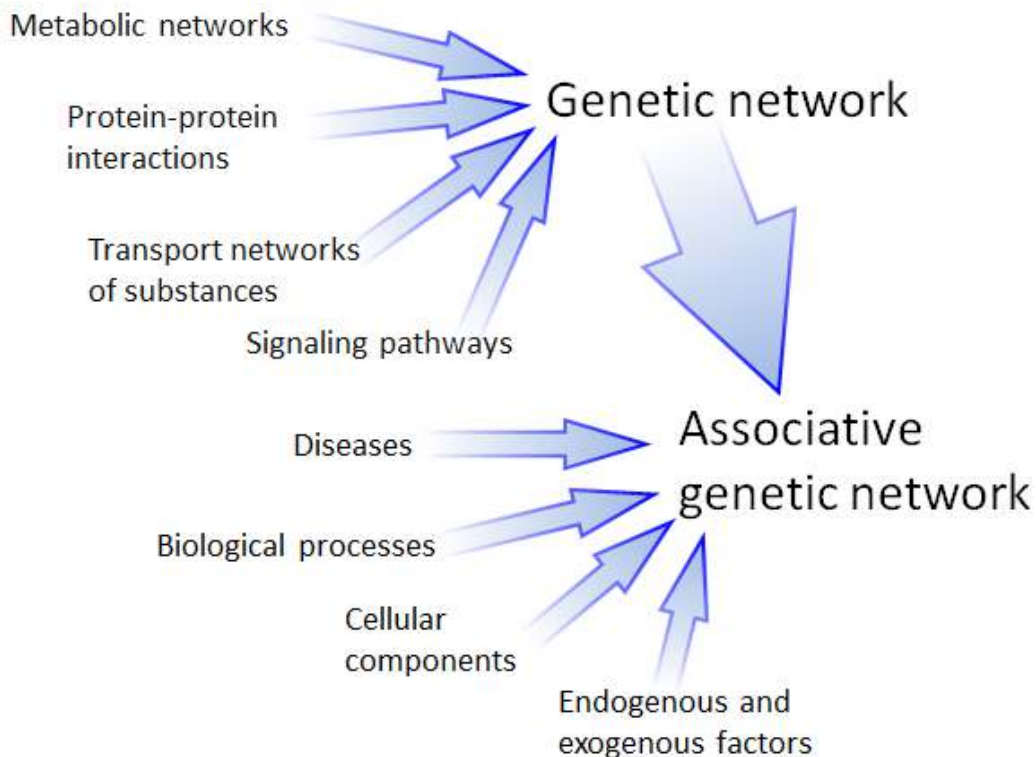
м.н.с. лаб. компьютерной протеомики Тийс Евгений Сергеевич

План лекции

- 1) Введение в реконструкцию ассоциативных сетей
- 2) Инструменты реконструкции основанные на текст-майнинге и базах данных (MeSHOPs, COREMINE, ANDSystem)
- 3) Анализ реконструируемых сетей
- 4) Примеры исследований

Ассоциативные генные сети

Ассоциативные генные сети являются молекулярно-генетическими сетями, ассоциированными с биологическими процессами, фенотипическими признаками или заболеваниями. Введение расширенного состава базовых объектов по сравнению с генными сетями позволяет проводить интеграцию разрозненных локальных генных сетей, которые напрямую не пересекаются между собой по молекулярно-генетическим объектам, но оказываются связанными на основании информации о их связи через какой либо расширенный объект, например фенотипический признак или заболевание.



Что можно поймать в ассоциативные сети?

Механизмы развития фенотипических признаков (в том числе заболеваний)

Новые диагностические маркеры

Новые терапевтические мишени

Механизмы лекарственного воздействия

Токсичность лекарств

Лекарственные взаимодействия

Новые мишени известных лекарств

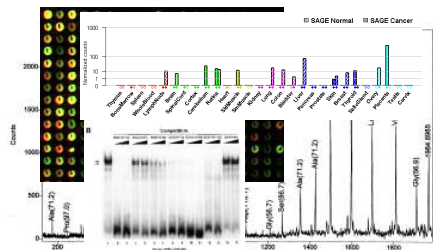
Синтропные, дистропные заболевания

Проведение анализа высокопроизводительных экспериментов

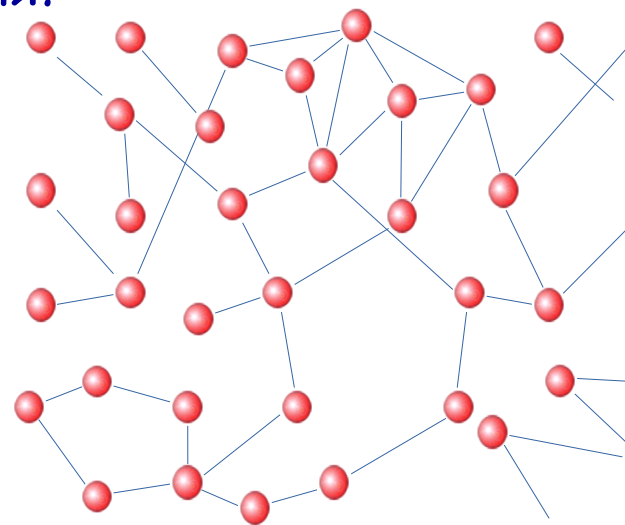
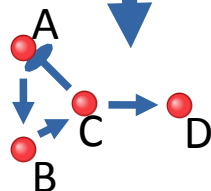
Исследование структурных и функциональных характеристик сетей:
Выявление хабов, функциональных модулей, путей передачи сигнала, обогащённости топологическими паттернами (поиск мотивов).

Моделирование биологических процессов

Что такое реконструкция?



PubMed



Курируемые базы данных



Базы данных, использующие методы автоматического анализа текстов научных публикаций и баз данных

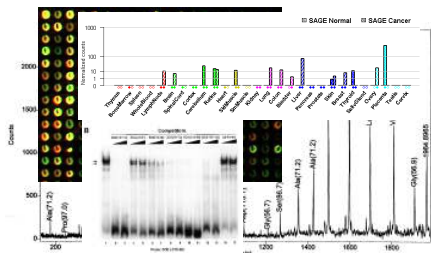


Инструменты анализа

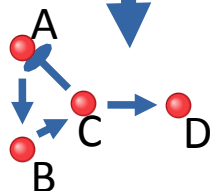
PINA: Protein Interaction Network Analysis platform



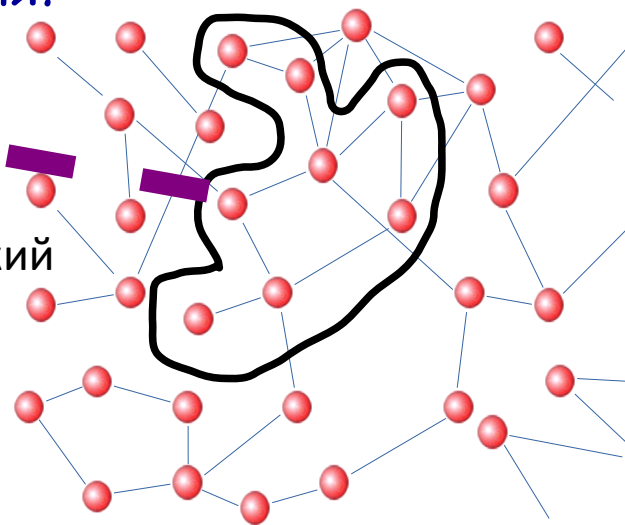
Что такое реконструкция?



PubMed



Целевой биологический процесс



Курируемые базы данных



Базы данных, использующие методы автоматического анализа текстов научных публикаций и баз данных



Инструменты анализа

PINA: Protein Interaction Network Analysis platform



План лекции

- 1) Введение в реконструкцию ассоциативных сетей
- 2) Инструменты реконструкции основанные на текст-майнинге и базах данных (MeSHOPs, COREMINE, ANDSystem)
- 3) Анализ реконструируемых сетей
- 4) Примеры исследований

Medical Subject Headings (MeSH)

Медицинские предметные рубрики (Medical Subject Headings, сокращённо MeSH) — контролируемый словарь, индексирующий журнальные статьи и книги по биомедицине; может также служить в качестве тезауруса, облегчающего поиск информации. Создан и обновляется Национальной медицинской библиотекой США, используется в базах статей Medline и PubMed. На данный момент содержит около 30000 терминов.



Publication type, MeSH terms



У каждой
статьи
в PubMed

Publication type

Research Support, Non-U.S. Gov't

MeSH terms

Apoptosis

Databases, Genetic

Gene Ontology

Gene Regulatory Networks*

Genomics/methods*

Humans

Internet*

Thyroid Neoplasms/genetics

Thyroid Neoplasms/pathology

Anatomy [A] +
Organisms [B] +
Diseases [C] +
Chemicals and Drugs [D] -
 Inorganic Chemicals [D01] +
 Organic Chemicals [D02] +
 Heterocyclic Compounds [D03] -
 Acids, Heterocyclic [D03.066] +
 Alkaloids [D03.132] -
 1-Deoxyxojirimycin [D03.132.014]
 Acridones [D03.132.032] +
 Amaryllidaceae Alkaloids [D03.132.052] +
 Anabasine [D03.132.066]
 Arecoline [D03.132.080]
 Benzophenanthridines [D03.132.089]
 Benzylisoquinolines [D03.132.098] -
 Aporphines [D03.132.098.038] +
 Berberine Alkaloids [D03.132.098.057] +
 Bicuculline [D03.132.098.077]
 Papaverine [D03.132.098.666] +

Корневые категории в MeSH:

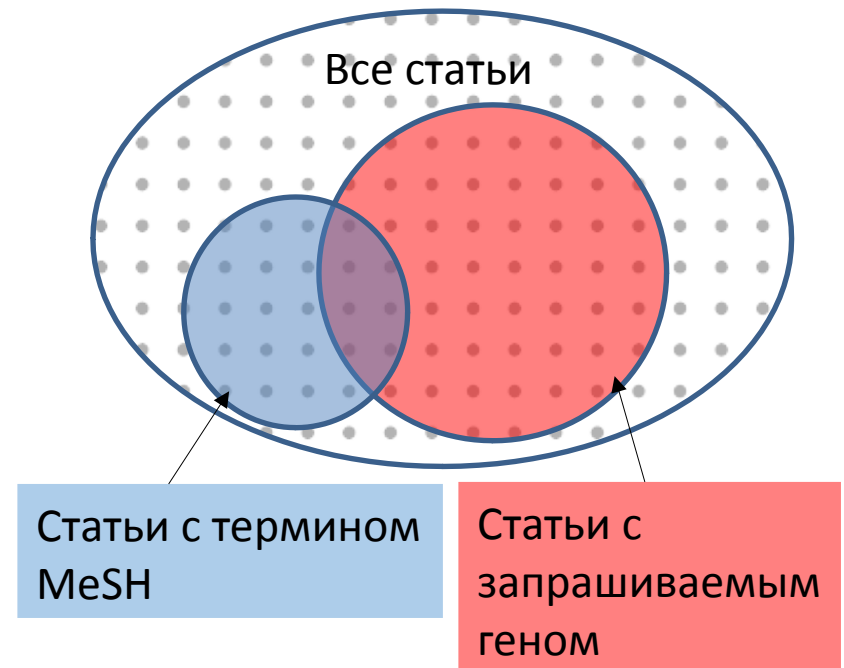
- ✓ Анатомия [A]
- ✓ Организмы [B]
- ✓ Болезни [C]
- ✓ Химические вещества и средства [D]
- ✓ Аналитические, диагностические и терапевтические методы и оборудование [E]
- ✓ Психиатрия и психология [F]
- ✓ Биологические науки [G]
- ✓ Физические науки [H]
- ✓ Антропология, образование, социология и социальные явления [I]
- ✓ Технологии, продукты питания и напитки [J]
- ✓ Гуманитарные науки [K]
- ✓ Информатика [L]
- ✓ Персоналии [M]
- ✓ Здравоохранение [N]
- ✓ Характеристики публикаций [V]
- ✓ Географическая локализация [Z]

Оценка со-встречаемости биологических терминов в научных текстах. Системы MeSHOPs и COREMINE

Анализ профилей перепредставленности MeSH терминов, открывает широкие возможности для обобщения информации посвященной изучаемому биологическому объекту, количественной идентификации меры связанности с ним других биологических терминов и поиска косвенных ассоциаций.

Системы [MeSHOPs](#) и [COREMINE](#) используют технологию оценки статистической значимости со-встречаемости биологических терминов в научных статьях для поиска связей между биологическими объектами.

На вход подается набор статей, связанных с биологической сущностью, из которых извлекаются все MeSH термины. Оценивается частота встречаемости каждого MeSH термина по сравнению с частотой встречаемости этого термина во всей совокупности статей. Для оценки статистической значимости перепредставленности каждого MeSH термина в статьях, связанных с изучаемой биологической сущностью, применяется Точный критерий Фишера.



Система MeSHOPs

<http://meshop.oicr.on.ca/meshop/index.html>

Для каждого биологического термина используется точный критерий Фишера для оценки P-value статистической значимости со-встречаемости биологических терминов в научных статьях.

Составляется таблица сопряженности (2 × 2):

	Статьи с термином pre-eclampsia	Статьи без термина pre-eclampsia	Полное число статей
Число статей с геном MTHFR	40		2426
Число статей без гена MTHFR			
Полное число статей	967	370917	

Look for specific MeSH terms: Pre-Eclampsia

Gene MeSHOP DB: Non-Redundant Gene Profiles (Gene2Pubmed/Human Genes Background)

Sort By: Score

Results: Show Top 500 Results Order: Up Submit

Number of *gene2pubmed* human genes $n=30262$.

Gene: MTHFR (Gene ID: 4524)

MeSH term: Pre-Eclampsia

Returning 1 results

Score	Articles	Gene Refs	Background Term Refs	Background Term Non-Refs
1.16e-19	40	2426	967	370917

Cheung, W. A., Ouellette, B. F., & Wasserman, W. W. (2012). Quantitative biomedical annotation using medical subject heading over-representation profiles (MeSHOPs). *BMC bioinformatics*, 13(1), 249.

Система CoreMine, основанная на анализе со-встречаемости MESH и Gene Ontology (GO) терминов

Система CoreMine доступна по адресу <https://www.coremine.com/medical/#search>.

The screenshot displays the CoreMine web interface. The main area shows a network diagram with a central node 'YY1' and numerous surrounding nodes representing related genes and biological processes. The right-hand panel provides detailed information for the selected focus, 'YY1'. This panel includes a description of YY1 as a transcription factor, a list of synonyms, and a section for 'Extracted associations' which lists various biological categories and their counts. A blue callout box with an arrow points to a bar chart in the 'Extracted associations' section, indicating a significance value of 1.20E-5.

Оценка достоверности со-встречаемости двух биологических терминов через p-value

Дженсен и др. провели автоматический анализ открытых текстов более 10 млн. научных публикаций (абстрактов) из MEDLINE для создания глобальной сети со-встречаемости 13712 именованных генов человека. Гены были аннотированы путем ассоциации с терминами из MeSH и геной онтологии (GO). Полученные сети были подтверждены экспериментами и с помощью анализа двух общедоступных наборов данных микрочипов.

Jenssen, T. K., Lægreid, A., Komorowski, J., & Hovig, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. *Nature genetics*, 28(1), 21-28.

Система для автоматической реконструкции ассоциативных генных сетей ANDSystem, основанная на технологии text-mining

Методы автоматического анализа текстов (text-mining) дают возможность быстрой и эффективной обработки больших объемов текстовой информации. В ИЦИГ СО РАН (Новосибирск) В.А. Иванисенко и П.С. Деменковым с соавторами с 2005 года разрабатывается система **ANDSystem** для автоматической реконструкции и анализа молекулярно-генетических сетей, описывающих биологические процессы как в норме, так и при заболеваниях. **ANDSystem** доступна по адресу <http://www-bionet.sccc.ru/andvisio/>

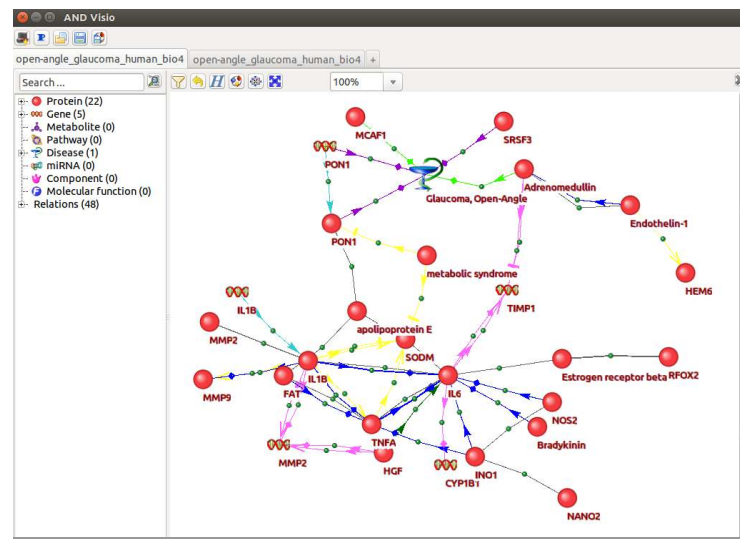
ANDSystem включает:

- модуль автоматического извлечения знаний с помощью семантических шаблонов,
- базу знаний ANDCell,
- программу ANDVisio.

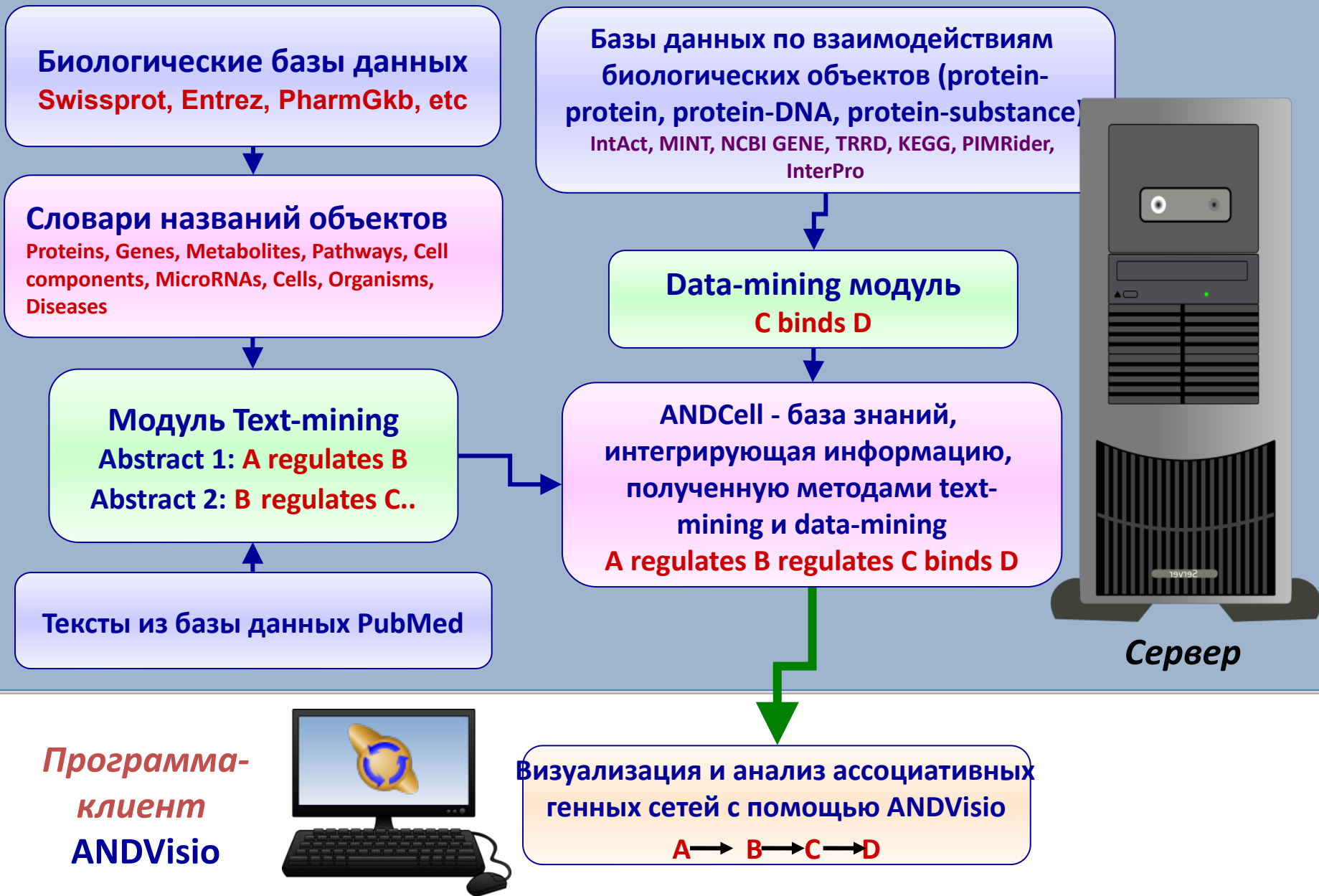
ANDVisio обеспечивает доступ к базе знаний, анализ и визуализацию результатов запроса.

ANDCell содержит более 4 млн фактов о молекулярно-генетических и каталитических взаимодействиях между белками, генами, метаболитами и т.д., извлеченных из более 20 млн абстрактов научных статей, представленных в базе данных PubMed.

Интерфейс системы ANDSystem для автоматической реконструкции ассоциативных генных сетей, реализованный в программе ANDVisio.



Общая схема системы ANDSystem



Источники информации

PubMed

- Pubmed
- PMID 1986
- PMID 1990
- PMID 1991
- PMID 1992
- PMID 1993
- PMID 1994
- PMID 1995
- PMID 1996
- PMID 1997
- PMID 1998
- PMID 1999
- PMID 2000
- PMID 2001
- PMID 2002
- PMID 2003
- PMID 2004
- PMID 2005
- PMID 2006
- PMID 2007
- PMID 2008
- PMID 2009
- PMID 2010
- PMID 2011
- PMID 2012
- PMID 2013
- PMID 2014
- PMID 2015
- PMID 2016
- PMID 2017
- Pubmed

- araport11
- AGI_LocusCode
- AspGD
- CGD
- dictyBase
- ena
- EchoBASE
- EcoGene
- EcoliWiki
- EC
- ECK
- EMBL
- Ensembl
- ENSEMBL_GeneID
- FB
- FlyBase
- GB
- GenBank
- GeneDB
- GR_PROTEIN
- HAMAP
- HGNC
- insdc
- irgsp
- JCVI_CMV
- KEGG_LIGAND
- MaizeGDB_Locus
- MGI
- NCBI_gi
- NCBI_Gene
- NCBI_GP
- protein_id
- PANTHER
- PDB
- Pfam
- PIR
- PomBase
- PR
- PubChem_Compound
- PubChem_Substance
- RefSeq
- RGD
- RGDID
- RNAcentral
- TAIR
- UniPathway
- UniProtKB-KW
- UniProtKB-SubCell
- UniRule
- WB
- WormBase
- ZFIN

Фактографические базы данных по молекулярно-генетическим взаимодействиям (protein-protein, protein-DNA, protein-substance)

IntAct, MINT, NCBI GENE, TRRD, KEGG, PIMRider, InterPro

- Clinvar
- MIM

- Gene - disease interactions
- Gene annotation
- GeneID
- InnateDB
- KEGG REACTION accession
- miRTarBase
- Protein - protein interactions

- innatedb
- innatedb allergy


- afcs
- BioGrid
- CHEBI
- DIP
- emdb
- GO
- HPRD
- imex
- IntAct
- InterPro
- IntEnz
- matrixdb
- mpidb
- MINT
- psi-mi
- rcsb pdb
- SGD
- SWISSPROT_AC
- wwpdb

Онтологии и словари объектов, используемые в ANDSystem.

Физиологический уровень		
Организмы		21 982
Клетки/ткани		644 304
Фенотипический уровень		
Заболевания		15 478
Фенотипические признаки		23 224
Побочные эффекты лекарств		5 226
Молекулярно-генетический уровень		
Гены		7 946 479
Белки		550 657
Метаболиты		42 594
МикроРНК		28 512
Лекарства		2 430
Биологические процессы		122 297
Клеточные компоненты		3 680

Онтология молекулярно-генетических взаимодействий системы ANDSystem

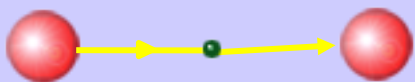
Направленные взаимодействия

Активация 

Подавление 

Регуляция 

activity upregulation



expression upregulation



degradation upregulation



activity downregulation



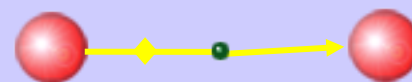
expression downregulation



degradation downregulation



activity regulation



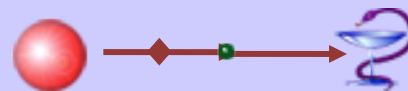
expression regulation



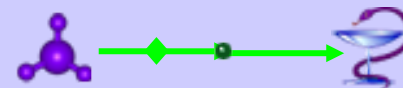
degradation regulation



regulation

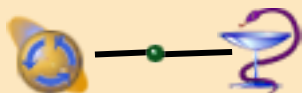


treatment



Не направленные взаимодействия

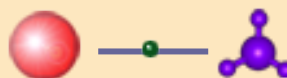
association



co-expression



interaction



Между объектами рассматриваются следующие типы взаимоотношений:

- **ассоциации** между объектами, т.е. связь между объектами не имеет четко определенной семантики, а задает потенциальную связь, смысл которой может быть уточнен в результате дальнейшего анализа;
- **физические взаимодействия** - образование короткоживущих или постоянных молекулярных комплексов. Физические взаимодействия могут связывать два или более белков, белок и низкомолекулярное вещество – лиганд, белок и ген (взаимодействие транскрипционного фактора с промоторным районом гена), белок и клеточную компоненту, два или более низкомолекулярных вещества, две или несколько клеточных компонент;
- **коэкспрессия**, одновременная экспрессия нескольких генов, которая была вызвана общими регуляторными механизмами, активизирующими экспрессию генов при меняющихся условиях в клетке;
- **лечение** – применение молекулярного агента для лечения определенного заболевания. Во взаимодействиях этого типа могут принимать участие белки и низкомолекулярные вещества – лекарства;
- **химические превращения**;
- **регуляторные взаимодействия** - влияние одного объекта на другой;

К химическим взаимодействиям относятся:

- **превращения** одних молекул в другие. Такой тип приписывается реакции в том случае если не указан фермент, катализирующий реакцию, а также если реакция протекает без участия катализатора. В превращениях участвуют только низкомолекулярные вещества. К превращениям относятся также взаимодействие между начальным и конечным продуктом метаболического пути, содержащего несколько промежуточных этапов, не описанных в тексте.
- **каталитические реакции**, в которых участвуют низкомолекулярные вещества в качестве субстратов и продуктов, а также белок в качестве фермента, осуществляющего катализ этой реакции.
- **расщепление** одного белка (субстрата) другим белком (протеолитическим ферментом).

Регуляторные взаимодействия могут быть разделены по типу регуляции на:

- **регуляцию экспрессии генов** транскрипционными факторами, а также их лигандами. В эту категорию входят как прямые регуляторные события, то есть регуляция экспрессии гена транскрипционным фактором, физически взаимодействующим с промотором этого гена, так и опосредованные регуляторные влияния лигандов этих факторов, а также белков, входящих в регуляторный путь, включая рецептор и белки сигнального пути.
- **регуляцию активности или функции** белка, гена, клеточной компоненты или молекулярно-генетического процесса. В качестве регулятора может выступать белок, низкомолекулярное вещество, клеточная компонента.
- **регуляцию транспорта** белков или низкомолекулярных веществ между компартментами клетки, а также секрецию этих молекул из клетки. В качестве регуляторов транспорта могут выступать белок, низкомолекулярное вещество или клеточная компонента.
- **регуляцию стабильности или деградации** молекулярных объектов. Объектом регуляции могут быть молекулярные структуры белков, матричные РНК, считываемые с генов, клеточные компоненты и низкомолекулярные вещества. Регуляторами, как и в предыдущих случаях, могут являться белки, низкомолекулярные вещества, клеточные компоненты и молекулярно-биологические процессы.
- **Регуляцию молекулярно-биологических процессов и заболеваний**. В качестве регуляторов могут выступать белки, низкомолекулярные вещества, гены, процессы и клеточные компоненты.

Кроме того, регуляторные события подразделяются по эффекту, который один объект оказывает на другой, то есть **усиление** или **ослабление** процесса.

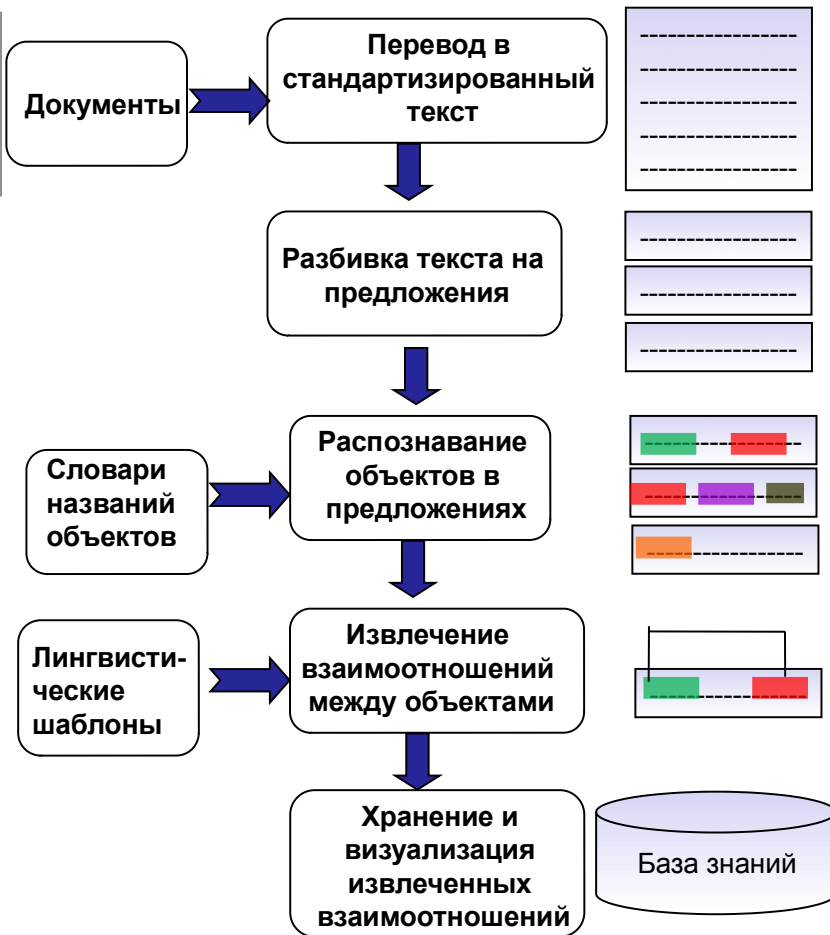
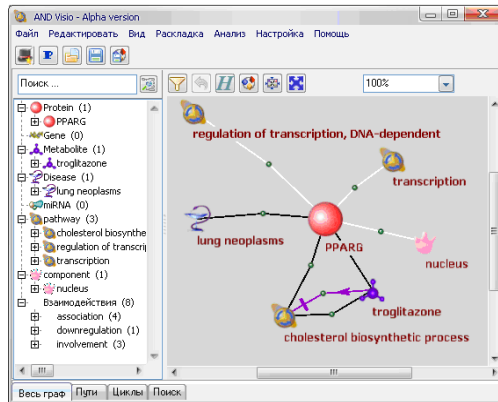
Статистика молекулярно-генетических взаимодействий системы ANDSystem

<i>association</i>	7 767 711	<i>activity regulation</i>	80 576
<i>involvement</i>	4 217 997	<i>activity downregulation</i>	295 160
<i>interaction</i>	625 594	<i>activity upregulation</i>	227 483
<i>expression regulation</i>	559 599	<i>degradation regulation</i>	3 222
<i>expression upregulation</i>	157 533	<i>degradation downregulation</i>	55 215
<i>expression downregulation</i>	129 273	<i>degradation upregulation</i>	34 004
<i>pathway regulation</i>	529 997	<i>miRNA regulation</i>	23 576
<i>pathway upregulation</i>	375 514	<i>coexpression</i>	6 617
<i>pathway downregulation</i>	374 856	<i>cleavage</i>	7 141
<i>transport regulation</i>	382 373	<i>catalyze modification</i>	3 089
<i>treatment</i>	188 264	<i>conversion</i>	27 078
<i>Catalyze</i>	518 114		

Модуль автоматического анализа текста (text-mining) в системе ANDSystem.

Проанализировано 23 млн рефератов из системы PubMed.
Выявлено 39 740 237 взаимодействий

Интерфейс модуля визуализации ANDVisio



Пример автоматической реконструкции генной сети ПОУГ системой ANDSystem

Results: Using binary logistic regression in an additive model, the rs2842980 SNP in **SOD2** was significantly **associated** with **POAG** diagnosis ($p < 0.03$) at a univariate level. [PubMed: 23638916]

Recent studies suggested a role for **tumour necrosis factor-alpha (TNF-alpha)** in the **pathogenesis** of **POAG**. [PubMed: 16133112]

Similarly **MnSOD** specific **activity was increased by TNF** (290% increase) and the amosite-TNF combination (313% increase) but not by amosite alone. [PubMed: 8817062]

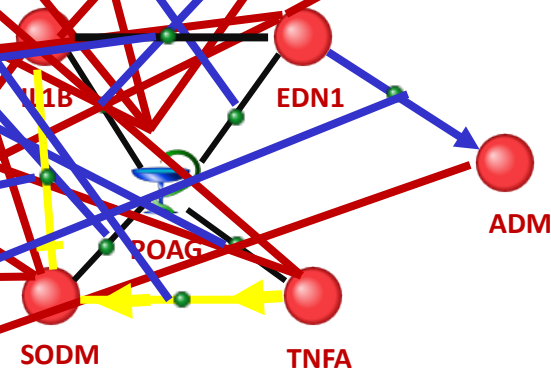
Incubation with **IL-1beta** **increased** concentrations of **ET-1** and PDET-1 relative gene expression. [PubMed: 18588356]

However, when supplemented with 10% FCS, **IL-1** **decreased the activity** of **SOD** after 24 and 72 h of incubation. [PubMed: 10912627]

Exogenous **ET-1** (1 to 100 nM) more potently **stimulated the release** of **ADM** from ET(A)/ET(B) compared with ET(A) isolates. [PubMed: 11588009]

A recent study identified single nucleotide polymorphisms (SNPs) within the **IL-1** gene cluster at chromosomal locus 2q13 that were **associated** with reduced risk for primary **open-angle glaucoma (POAG)** in whites. [PubMed: 17460270]

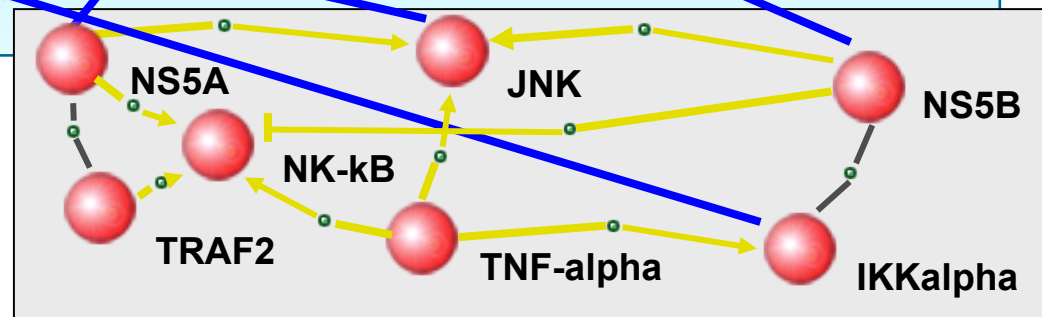
Analysis of aqueous humor from patients with primary **open-angle glaucoma (POAG)** revealed marked **increases** in the content of **endothelin-1 (ET-1)** and transforming growth factor beta (TGF- β). [PubMed: 22736605]



Красные – названия объектов
Синие – взаимодействия

Пример автоматической реконструкции генной сети ВГС-хозяин системой ANDSystem

Hepatitis C virus (HCV) NS5B protein is a membrane-associated phosphoprotein that possesses an RNA-dependent RNA polymerase activity. We recently reported that **NS5A protein** interacts with **TRAF2** and **modulates tumor necrosis factor alpha (TNF-alpha)-induced NF-kappaB** and **Jun N-terminal protein kinase (JNK)**. Since **NS5A** and **NS5B** are the essential components of the **HCV** replication complex, we examined whether **NS5B** could **modulate** TNF-alpha-induced **NF-kappaB** and **JNK** activation. In this study, we have demonstrated that TNF-alpha-induced **NF-kappaB** activation is **inhibited** by **NS5B protein** in HEK293 and hepatic cells. Furthermore, **NS5B protein** **inhibited** both **TRAF2-** and **IKK-induced NF-kappaB** activation. Using coimmunoprecipitation assays, we show that **NS5B** **interacts** with **IKKalpha**. Most importantly, **NS5B protein** in **HCV** subgenomic replicon cells **interacted** with endogenous **IKKalpha**, and then **TNF-alpha-mediated IKKalpha** kinase **activation** was significantly **decreased** by **NS5B**. Using in vitro kinase assay, we have further found that **NS5B protein** synergistically **activated** TNF-alpha-mediated **JNK** activity in HEK293 and hepatic cells. These data suggest that **NS5B protein** **modulates** **TNF-alpha** signaling pathways and may contribute to HCV pathogenesis.



Красные – названия объектов

Синие – ключевые слова
взаимодействий

Зеленые – организмы

Розовые – типы объектов

Оценка точности информации, содержащейся в базе данных ANDCell

$$\text{Precision} = N_{\text{correct}} / N_{\text{total}}$$

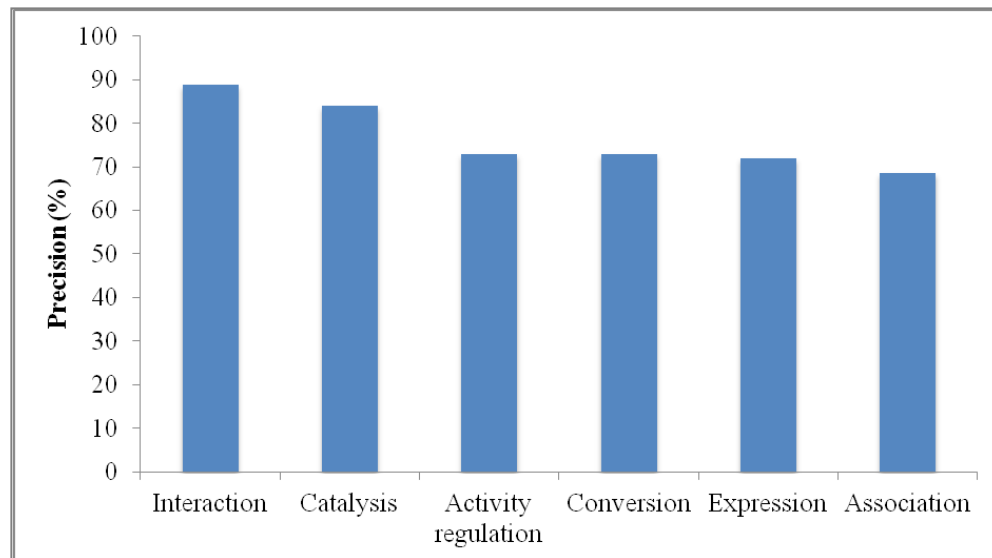
N_{correct} – число верно распознанных взаимодействий.

N_{total} – общее число взаимодействий в тестовом наборе.

Точность 6 основных типов взаимодействий, выявленных системой ANDSystem.

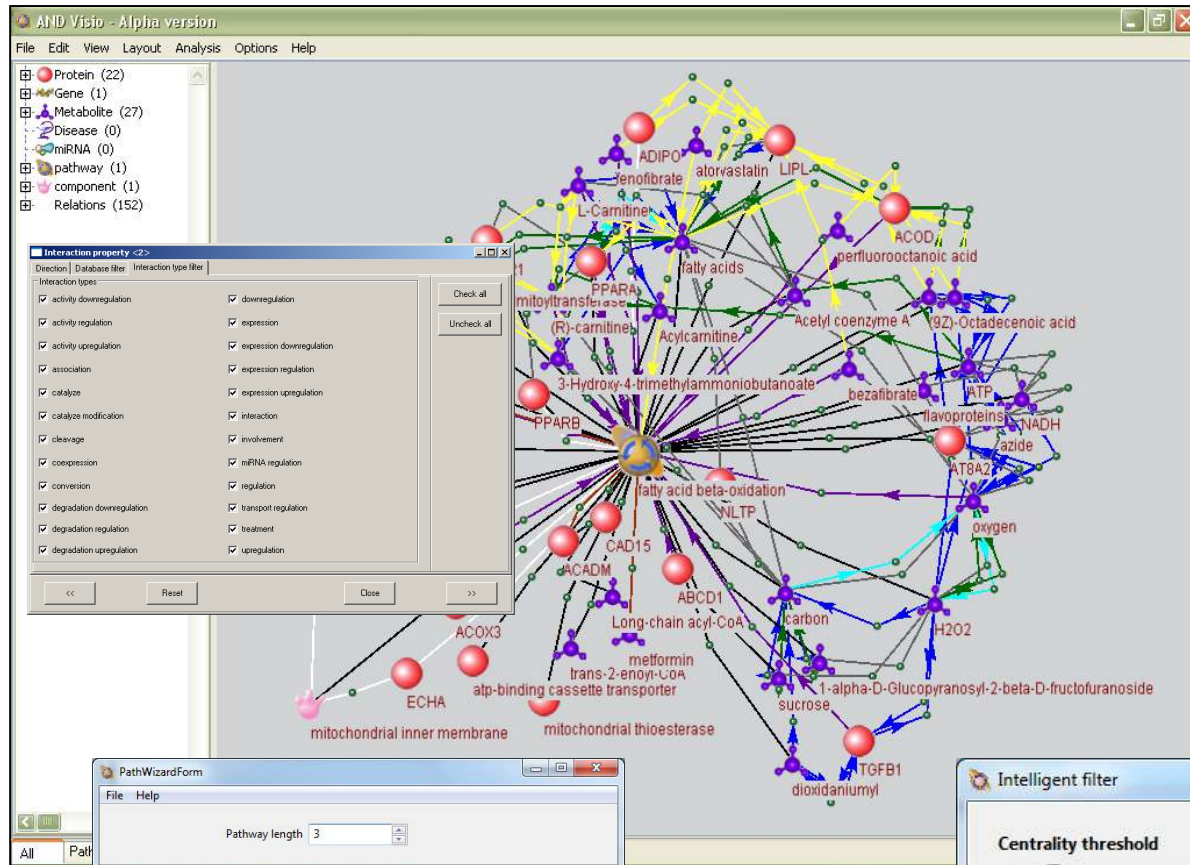
Для взаимодействий типа “interaction” наблюдалась максимальная точность (88.8%), минимальная -- для типа “association” (68.6%).

В среднем точность составила 76.5%.

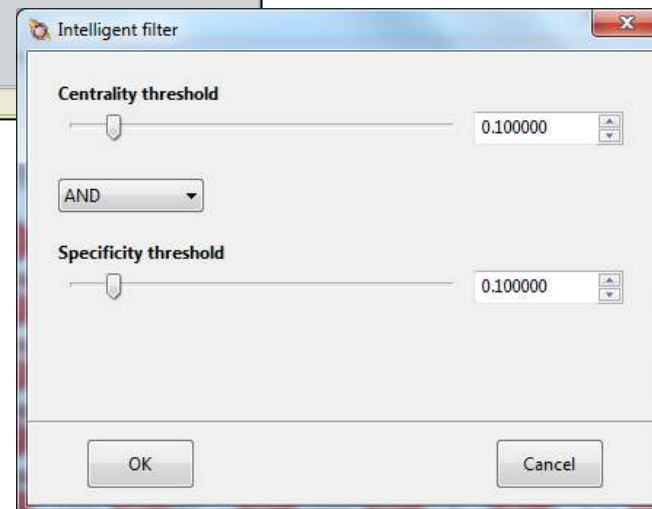
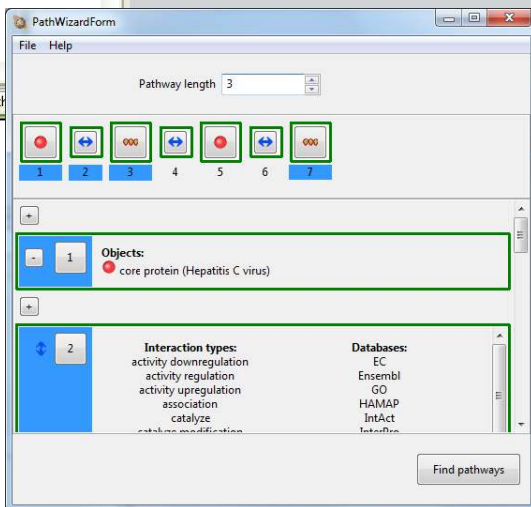


Полнота составила около 54% при оценке на Золотом стандарте, содержащем информацию из базы данных GeneNet, собранную вручную экспертами, о различных типах молекулярно-генетических взаимодействий.

Краткая характеристика интерфейса AndVisio, реализованного в ANDSystem - возможности поиска и фильтрации данных.



Для интерактивной визуализации ассоциативных генных сетей была разработана программа ANDVisio. Программа является графическим интерфейсом к базе данных ANDCell, в которой представлена информация о взаимоотношениях между биологическими объектами.

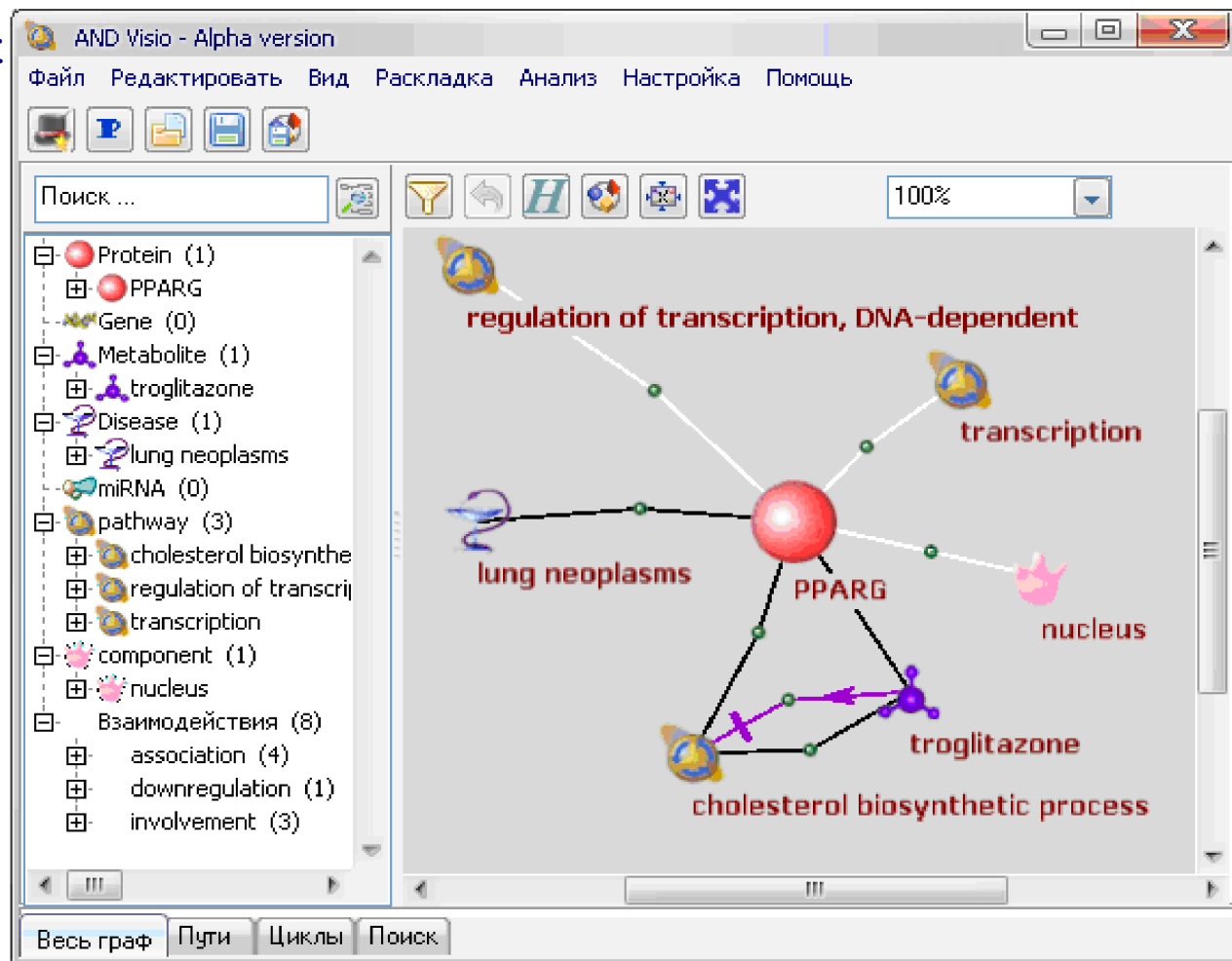


Программа для визуализации ассоциативных сетей ANDVisio предназначена для выполнения следующих основных функций:

- Составление запросов к базе знаний с целью поиска информации о взаимодействиях между белками, генами, метаболитами, микроРНК, заболеваниями, метаболическими процессами и клеточными компонентами
- Навигация по ассоциативной сети, осуществление доступа к первичным источникам информации в сети.
- Раскладка объектов ассоциативной сети на плоскости
- Поиск объектов в ассоциативной сети по синониму
- Редактирование и фильтрация ассоциативной сети, добавление и удаление объектов из сети
- Анализ ассоциативных сетей: поиск путей и замкнутых циклов внутри сетей
- Сохранение ассоциативной сети в формате XML, SVG и TSV

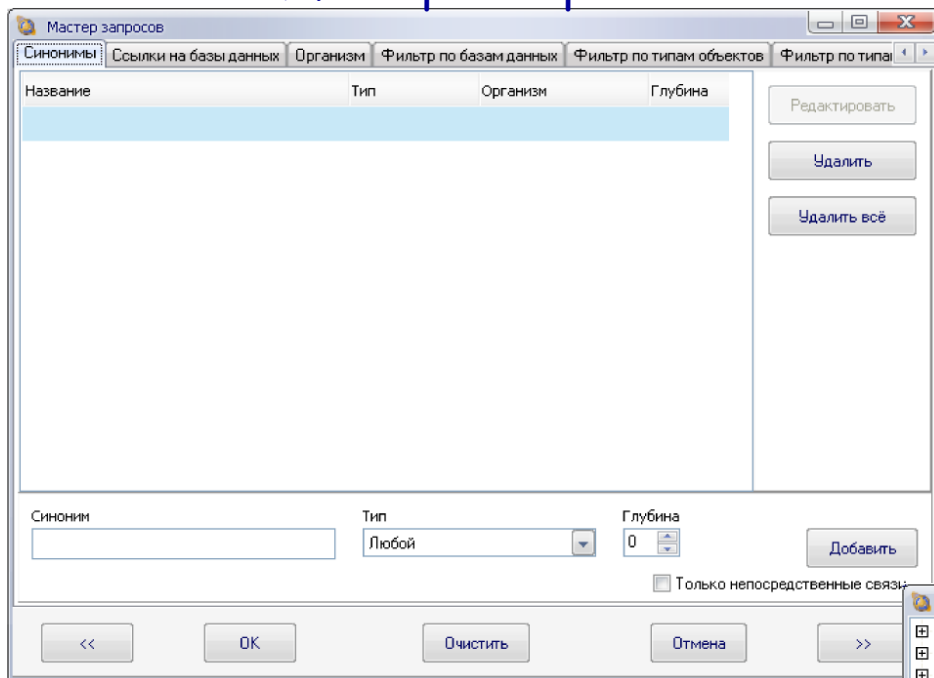
Главное окно программы

Основное окно программы содержит панель главного меню, две панели с кнопками для быстрого доступа к основным функциям, область для представления сетей в виде списка всех участников сети с их свойствами и область для графического представления сетей



Меню файл содержит пункты "Соединение", "Мастер запросов", "Открыть ...", "Сохранить как ...", "Сохранить как изображение ...", "Выход". Пункт "Соединение" позволяет восстановить соединение с базой данных. В случае невозможности восстановления соединения выдаётся предупреждение и блокируются все функции работы с базой данных.

Окно «Мастера запросов»



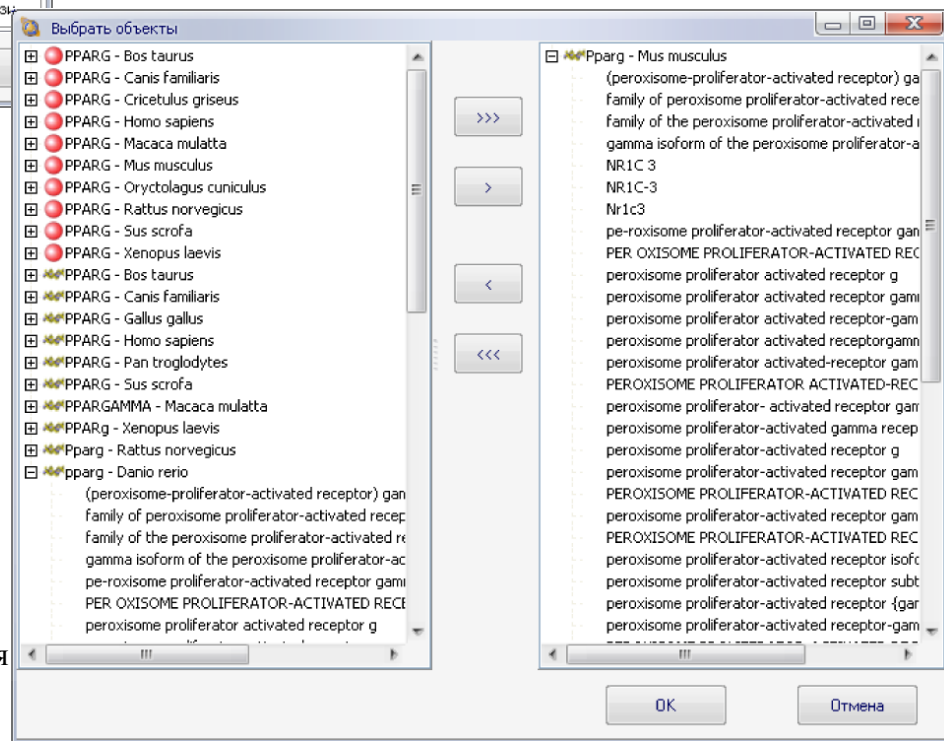
Пункт "Мастер запросов" открывает форму для настройки параметров запроса к базе данных. Форма состоит из нескольких вкладок: "Синонимы", "Ссылки на базы данных", «Организмы», «Фильтр по базам данных», «Фильтр по типам объектов», «Фильтр по взаимодействиям».

На вкладке "Синонимы" в нижней части формы пользователь может ввести шаблон названия объекта и выбрать тип объекта. После нажатия кнопки "Добавить" производится поиск объектов в базе данных, имеющих совпадающие с указанным шаблоном названия. В случае если обнаружено более 1 объекта, удовлетворяющего условиям, то открывается форма для уточнения пользователем интересующих объектов.

Форма «Выбрать объекты» разделена на 2 половины, на левой панели представлены доступные объекты, удовлетворяющие критерию, а на правой список выбранных объектов. Для удобства выбора пользователем интересующих объектов, для каждого объекта приведены все, содержащиеся в базе, его названия, а также принадлежность к организмам.

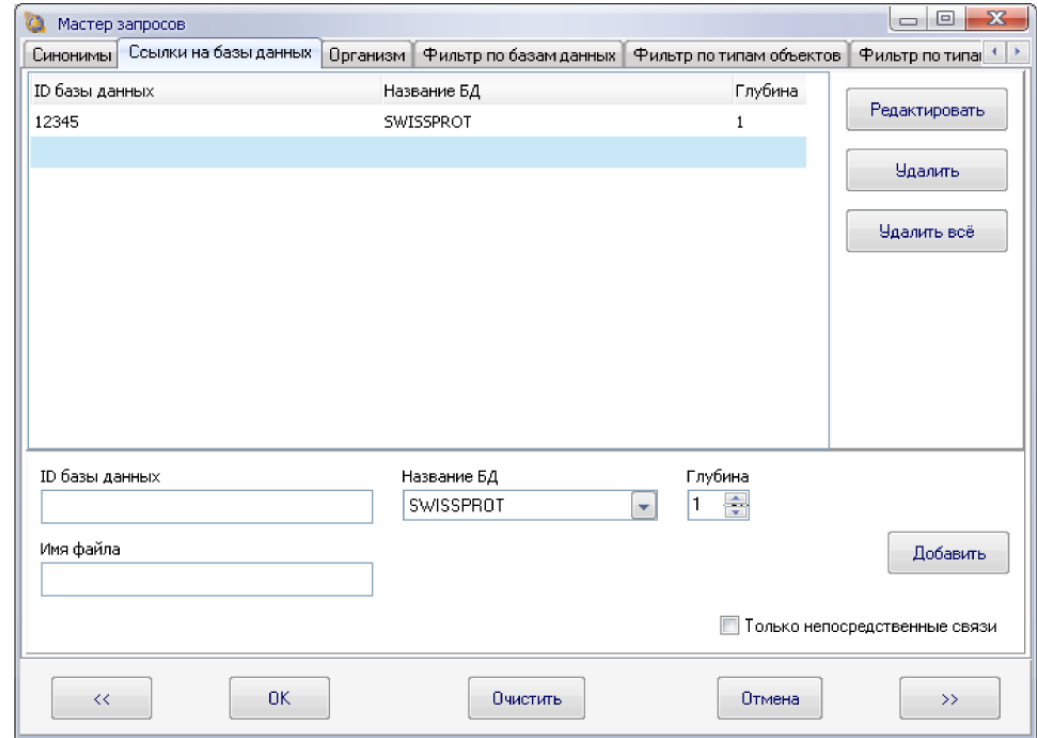
Перемещение объектов с одной панели на другую может быть осуществлено несколькими способами: доступно использование механизма Drag&Drop (объекты можно перетаскивать мышкой), либо можно воспользоваться кнопками, расположенными в центре формы. Кнопки выполняют следующие действия: ">>>>" и "<<<<" позволяют перенести все объекты с одной панели на другую, а ">" и "<" перемещают выделенные объекты.

После подтверждения списка выбранных объектов они заносятся в список в форме "Мастера запросов".



Окно «Мастера запросов» для добавления объектов по идентификаторам баз данных

На вкладке "Ссылки на базы данных" пользователь может указать интересующие его объекты по их идентификаторам во внешних базах данных. Идентификаторы можно добавлять в список как по одному, так и сразу несколько. Для добавления одного идентификатора следует его ввести в поле "ID базы данных". Для занесения сразу нескольких идентификаторов следует их сохранить в текстовый файл (каждый идентификатор должен быть записан в отдельной строке), а затем выбрать созданный файл в поле "имя файла". Далее следует указать название базы данных идентификаторы которой были указаны.

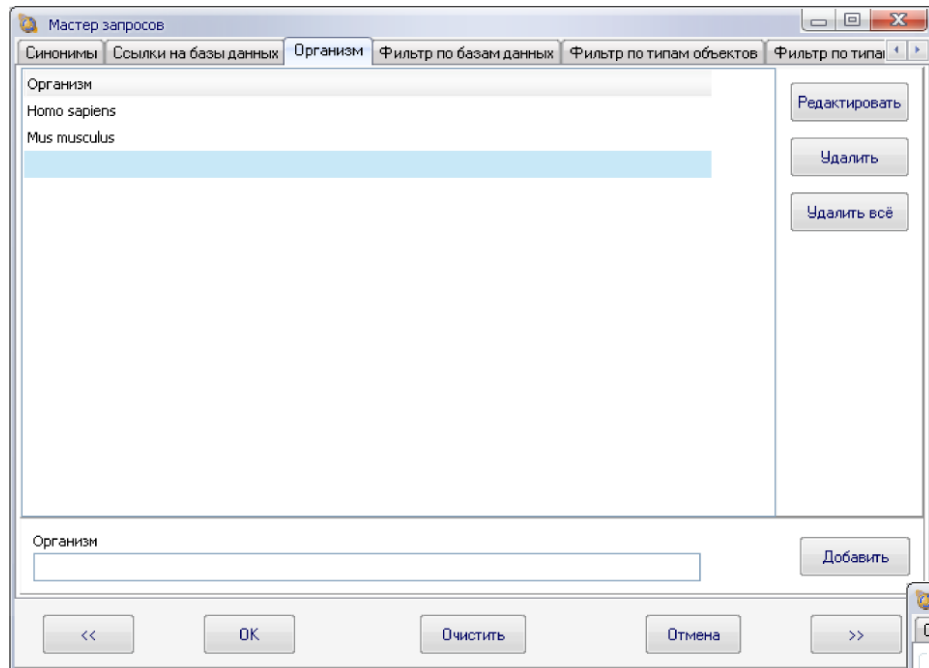


Если название базы данных не известно, то можно попробовать поискать эти идентификаторы по всем доступным базам данных, но идентификаторы многих баз данных пересекаются между собой и соответствуют совершенно разным объектам. После нажатия на кнопку "Добавить" идентификаторы добавятся в список интересующих объектов.

Под уровнем сети мы понимаем максимальную длину пути кратчайшего пути от начальной вершины до всех остальных. Уровень сети можно указать как перед добавлением новых объектов в список, так и после. Уровень может быть задан для каждого объекта в списке индивидуально.

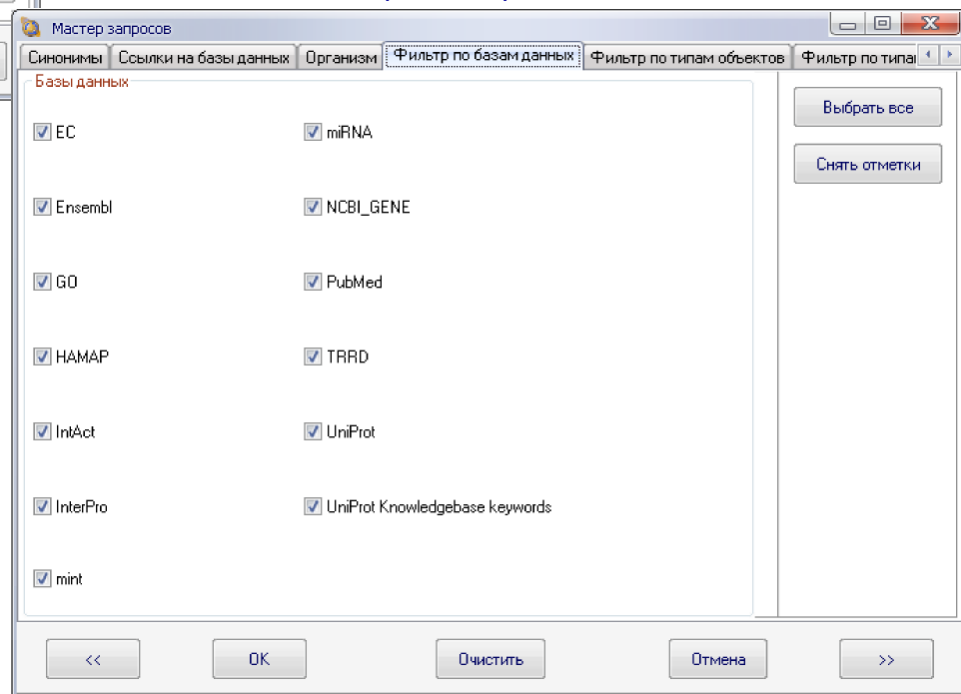
Включение переключателя "Только непосредственные связи" используется в случае, если требуется получить только непосредственно связанные объекты и при этом не требуются взаимосвязи между объектами 1 уровня.

Окно «Мастера запросов» для введения фильтра по организмам



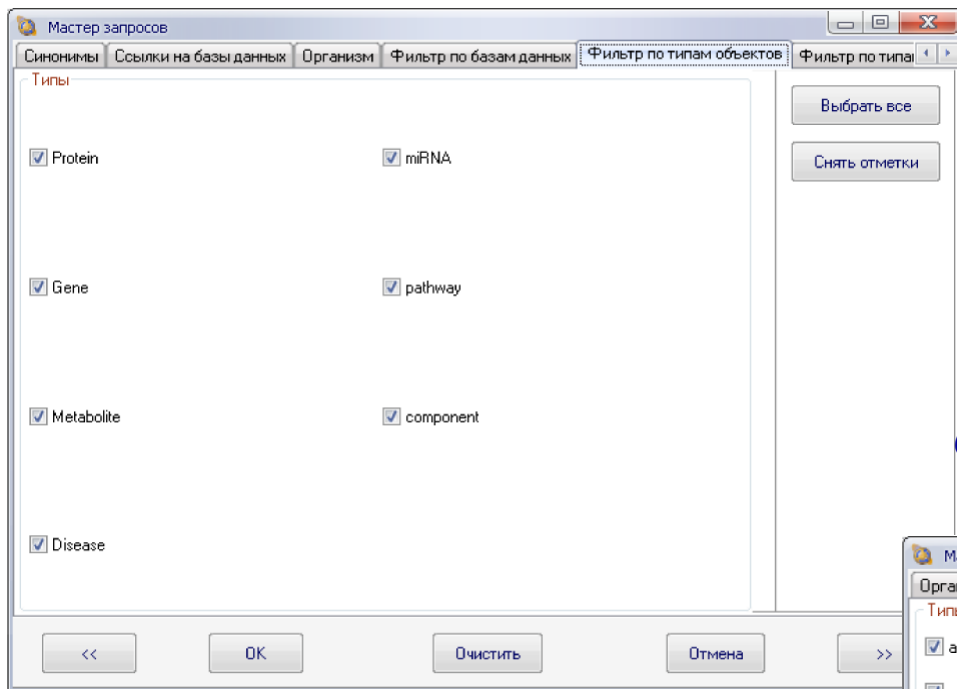
На вкладке "Организм" пользователю предлагается указать список интересующих организмов. В нижней части вкладки в поле "Организм" требуется ввести шаблон названия организма и нажать на кнопку "Добавить". В случае если обнаружено более 1 названия организма, удовлетворяющего условиям, то открывается форма для уточнения пользователем интересующих организмов.

Окно «Мастера запросов» для заполнения фильтра по базам данных



Вкладка "Фильтр по базам данных" позволяет выбрать базы данных из которых будут выбираться связи между объектами.

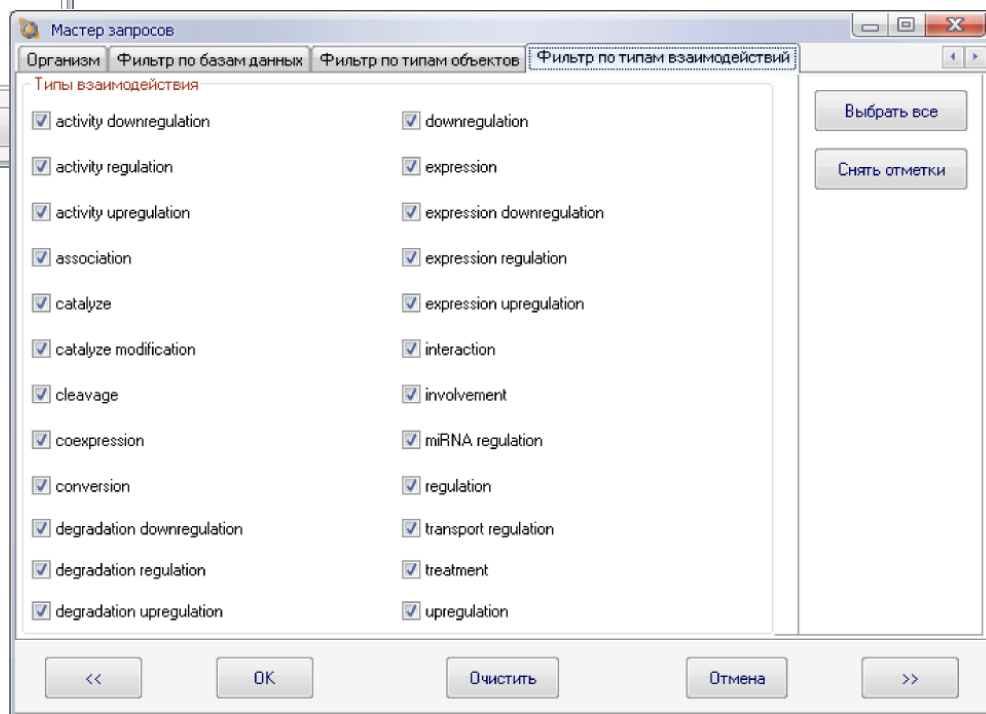
Окно «Мастера запросов» для заполнения фильтра по типам объектов



На вкладке "Фильтр по типам объектов" указываются типы объектов, которые должны быть в реконструированной сети.

На вкладке "Фильтр по типам взаимодействий" указываются типы взаимодействий, которые должны быть в реконструированной сети

Окно «Мастера запросов» для заполнения фильтра по типам взаимодействий



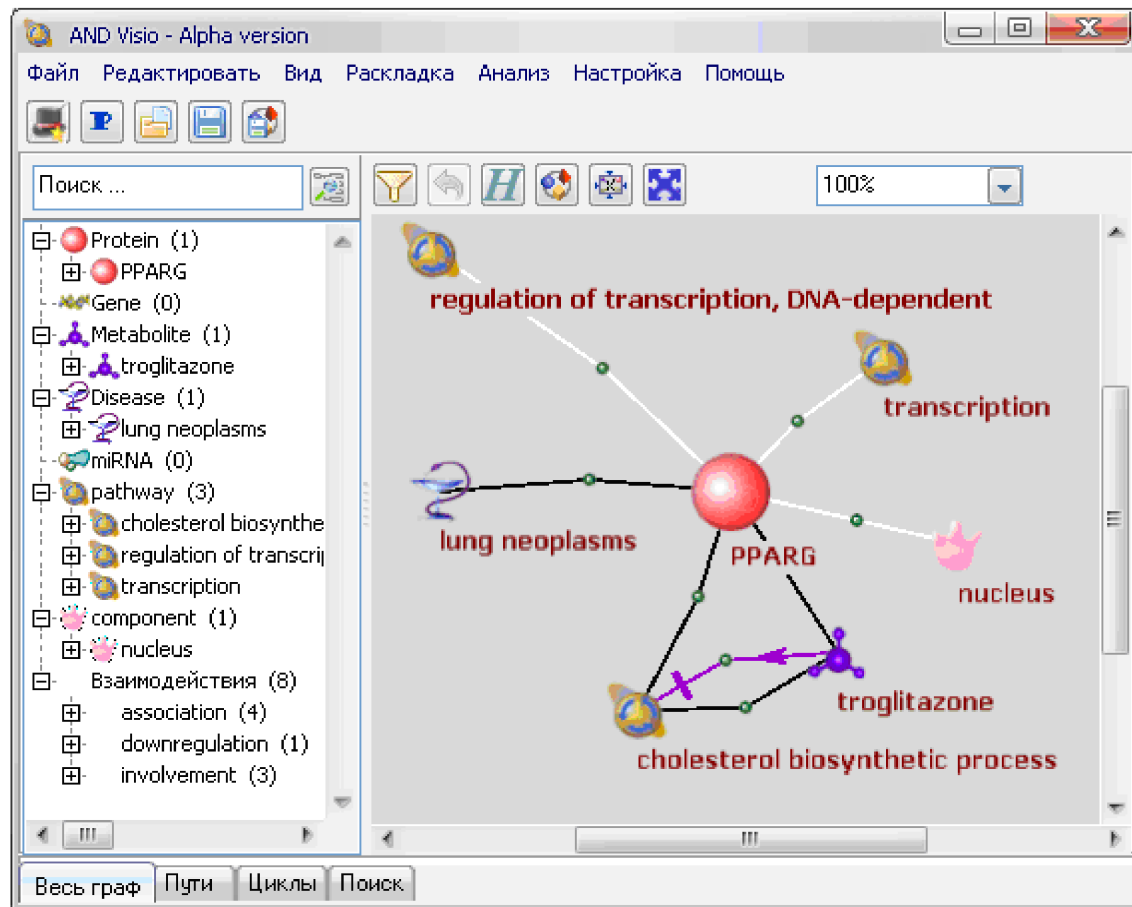
После окончания заполнения формы "Мастер запросов" нажимается кнопка "OK". Результат выполнения запроса на сервере возвращается в программу ANDVisio.

В левом окне программы в виде дерева указаны все объекты ассоциативной сети, сгруппированные по типам. Для каждого объекта приведены список синонимов его названий и организм. Двойной щелчок на объекте выделит его и разместит сеть таким образом, чтобы этот объект был в центре окна. А двойной клик на объекте в сети произведёт поиск его в списке на левой панели.

Взаимосвязи на левой панели сгруппированы по их типам. Для каждой взаимосвязи приводятся все её участники с их свойствами.

Щелчок левой кнопкой мыши на объекте производит его выделение, при этом подсвечиваются все его связи красным цветом.

Главное окно программы. Построена сеть первого уровня для белка PPARG

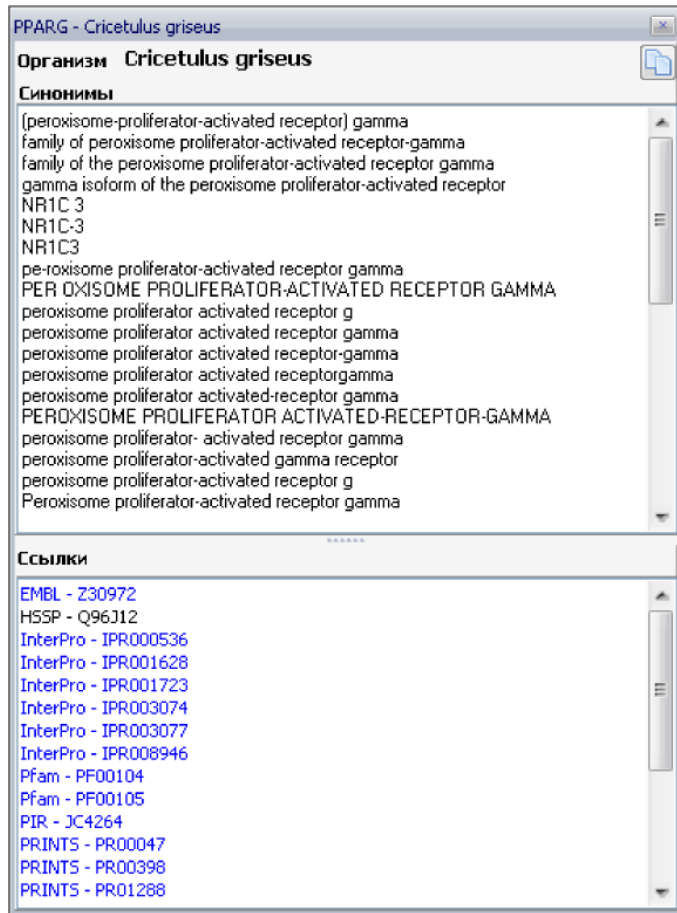


Пункт меню *Вид* -> *Показывать названия* позволяет управлять отображением подписей к объектам.

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

Правый щелчок мыши на объекте открывает всплывающее меню. В нем можно выбрать пункт *Свойства* для получения подробной информации о свойствах данного объекта.

Окно свойств объекта



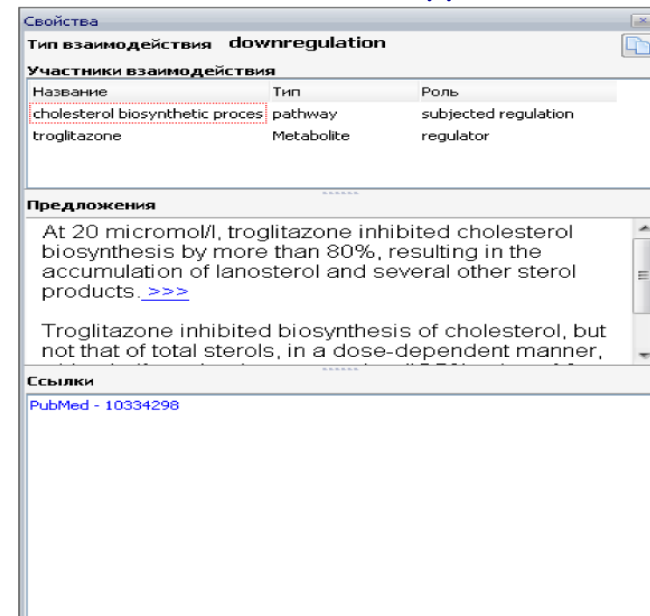
В верхнем поле окна свойств объектов указаны синонимы данного объекта, в нижнем поле --- ссылки на идентификаторы этого объекта в различных базах данных. Щелчок на ссылке позволяет через web-браузер получить доступ к карточке интересующего объекта в выбранной базе данных.

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.

Окно свойств взаимодействия



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

С помощью пункта меню **Раскладка -> Спрямить связи** можно перерасложить 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** или нажмите **Del**.

Чтобы применить фильтр к реконструированной сети, выберите пункт меню **Редактировать -> Фильтр**. Выберите свойства фильтра.

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

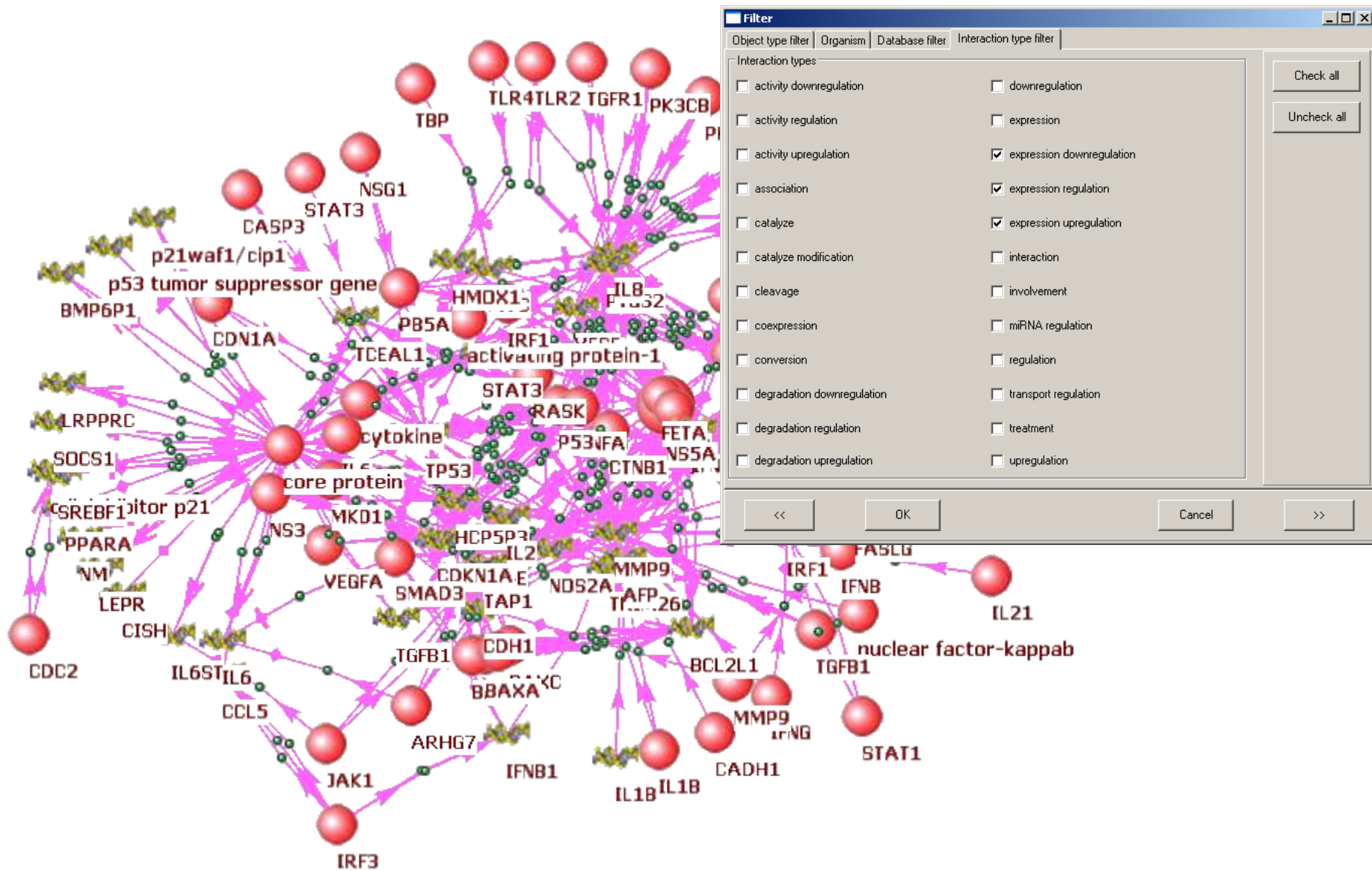
Отметьте галочкой пункт **Не связывать новые объекты с существующими** для того, чтобы не показывать связи между добавленными в сеть объектами и объектами, ранее представленными в сети.

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 разделе (**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.

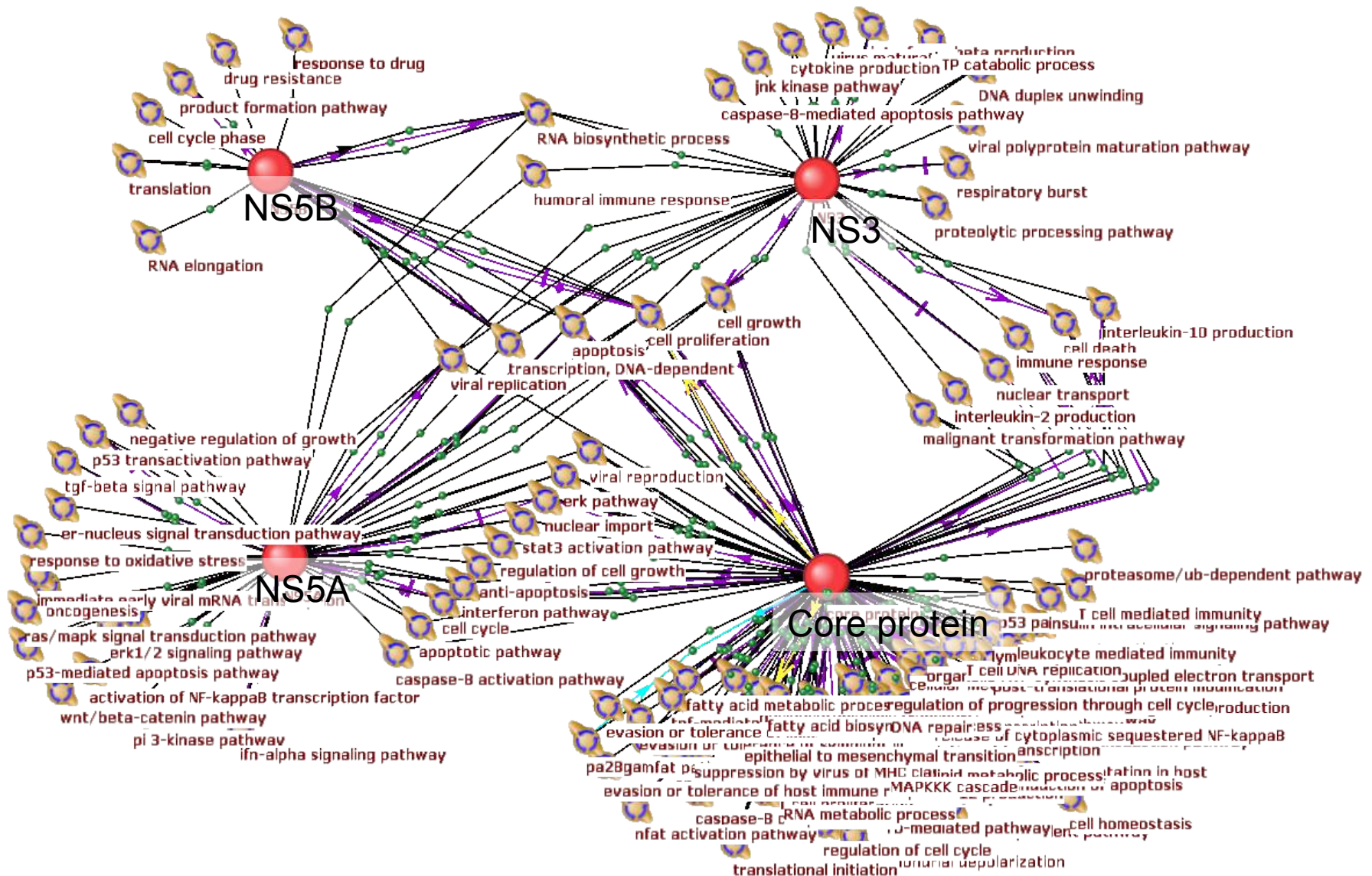
Примеры задач, которые можно решать с помощью системы ANDSystem:

- Реконструкция и анализ ассоциативных генных сетей
- Интерпретация и визуализация экспериментальных данных
- Анализ межорганизменных взаимодействий, в том числе типа «патоген-хозяин»
- Выявление коморбидных взаимоотношений между заболеваниями; реконструкция сетей «заболевание-заболевание»
- Выявление побочных эффектов терапии
- Поиск мишеней для лекарств
- Выявление генов-кандидатов для генотипирования
- Выявление механизмов патогенеза заболеваний
- Выявление механизмов действия лекарств
- Поиск лекарство-лекарство взаимодействий

Подсеть белков и генов человека, ассоциированных с хронической инфекцией гепатитом С, включающая только регуляторные связи.

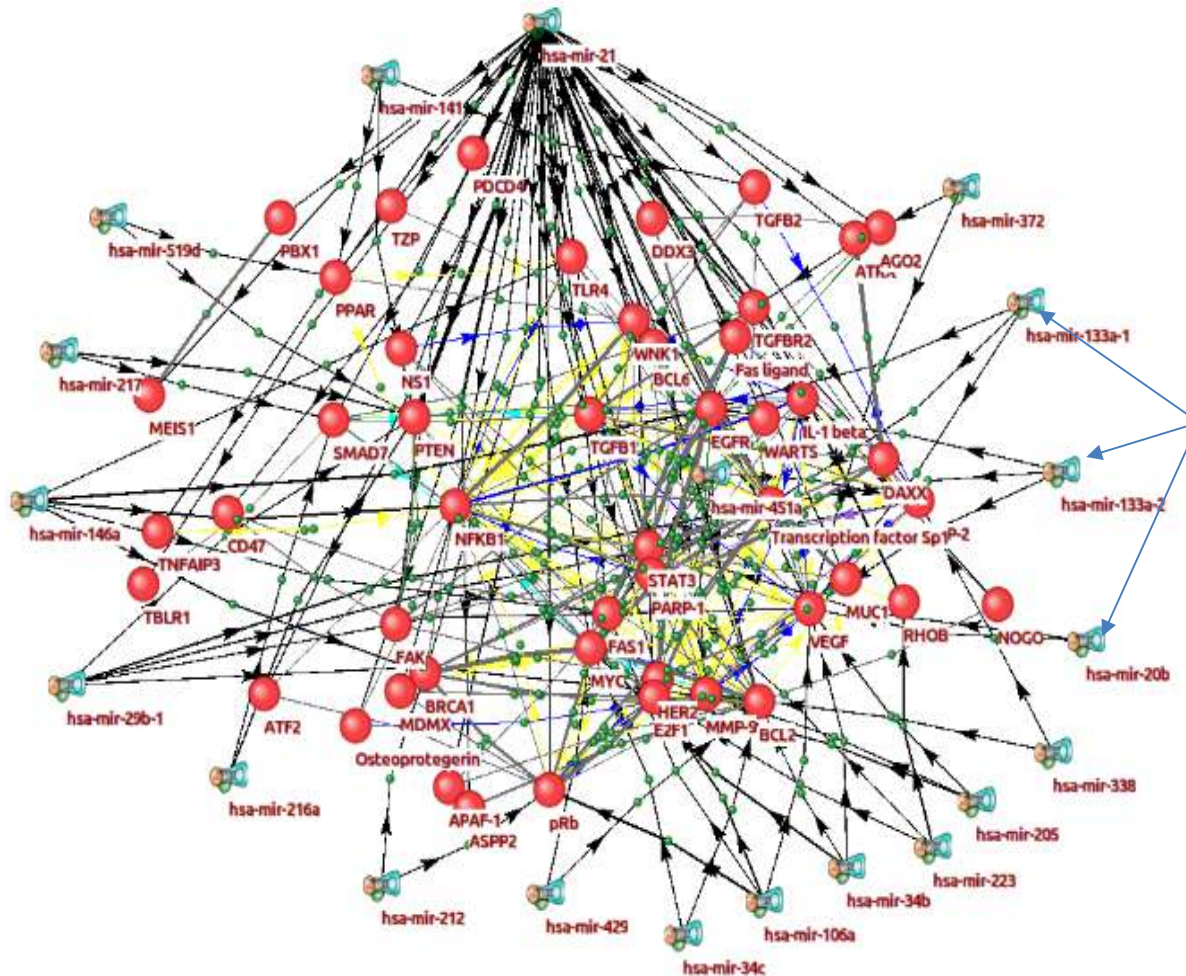


Взаимодействия биологических процессов человека, ассоциированных с воздействием белков вируса гепатита С



Ассоциативная сеть, описывающая регуляцию апоптоза микроРНК-21

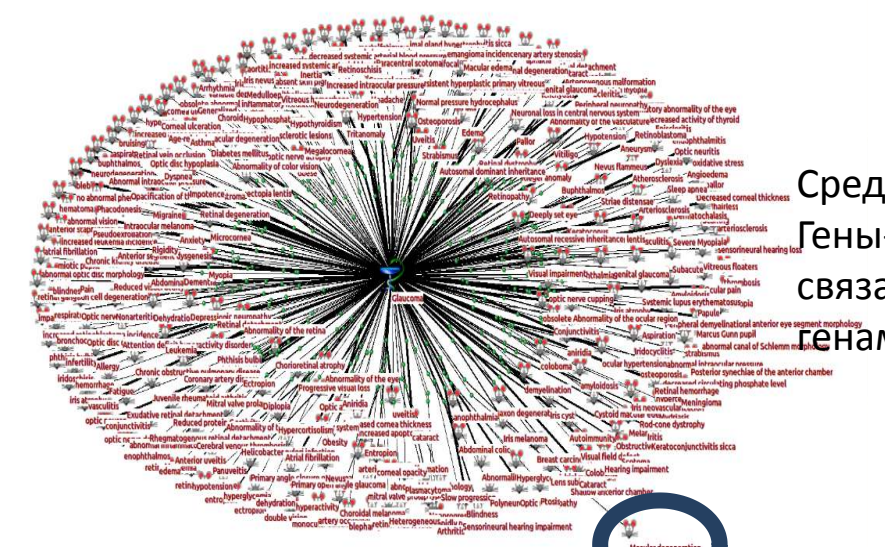
Одной из первых микроРНК, для которой была показана функция в патогенезе рака, стала miRNA21, связанная с глиобластомой и астроцитомой.



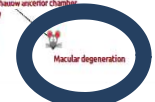
микроРНК, специфически регулирующие гены/белки, вовлеченные в апоптоз.

Комбинированные молекулярно-генетические и физиологические сети. Поиск генов-новых факторов риска на основе анализа ассоциом феноотипов заболевания

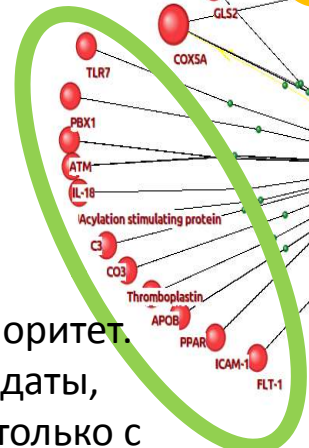
ассоциом феноотипов, наблюдаемых при глаукоме



Макулярная дегенерация

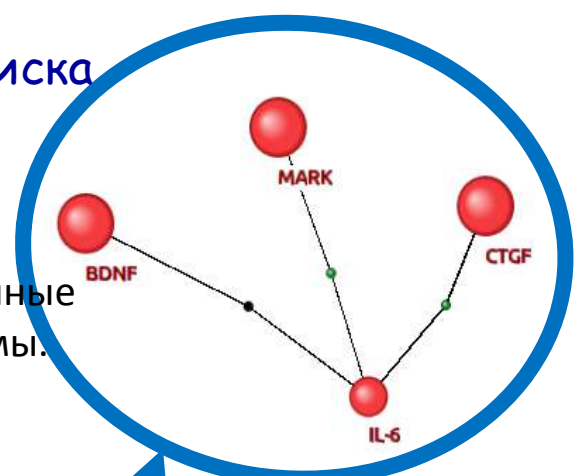


Фрагмент ассоциативной генной сети макулярной дегенерации

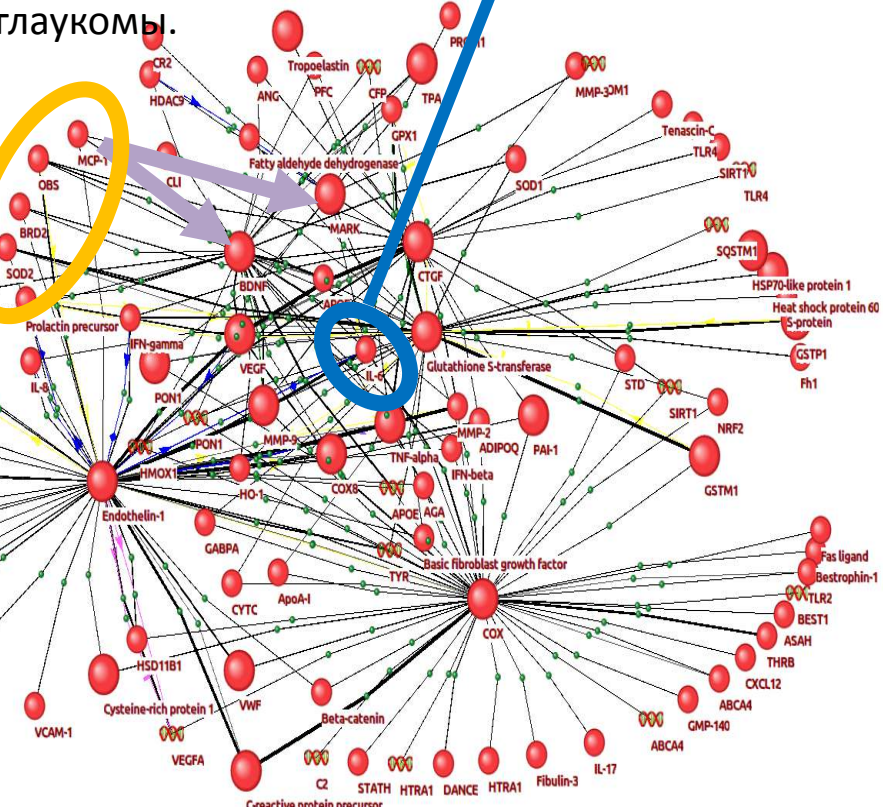
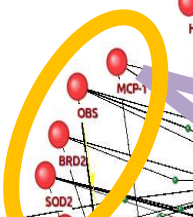


Низкий приоритет. Гены-кандидаты, связанные только с 1 геном глаукомы.

Максимальный приоритет. Гены-кандидаты, связанные с 3 генами глаукомы.

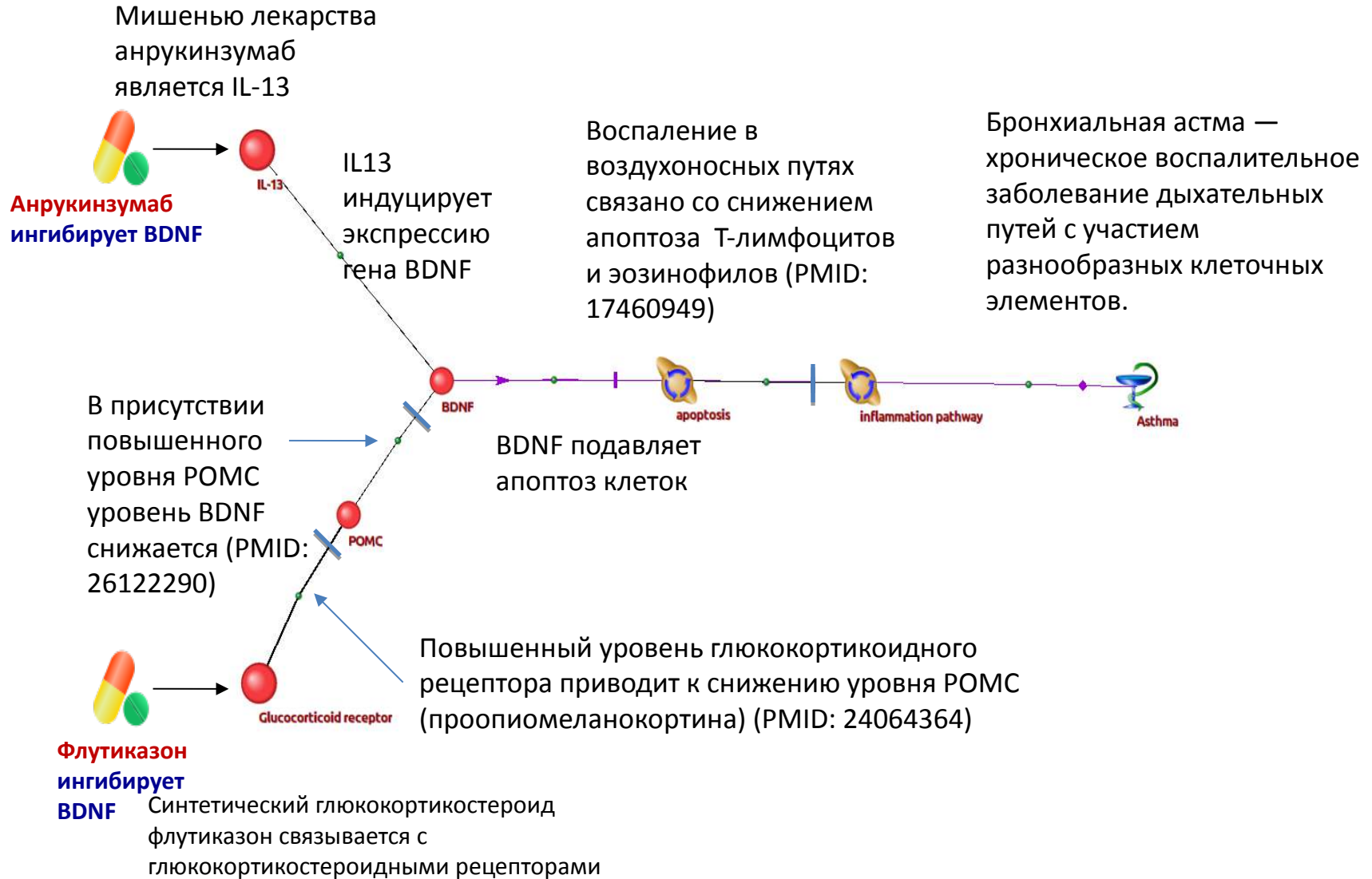


Средний приоритет. Гены-кандидаты, связанные с 2 генами глаукомы.



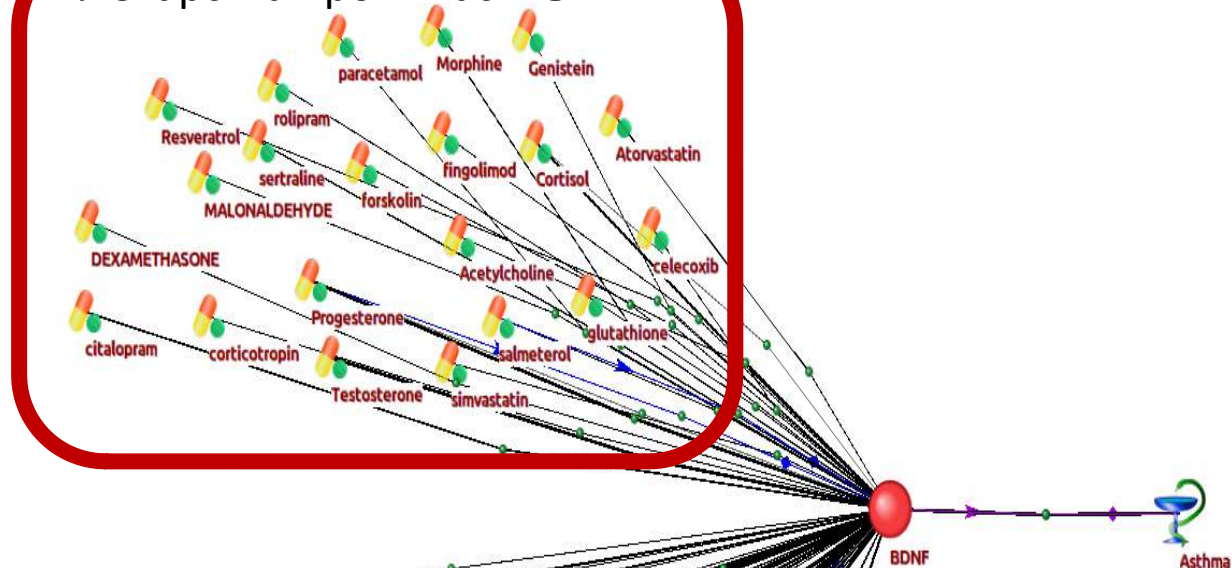
C-reactive protein precursor

Пример реконструкции молекулярных механизмов действия лекарств.

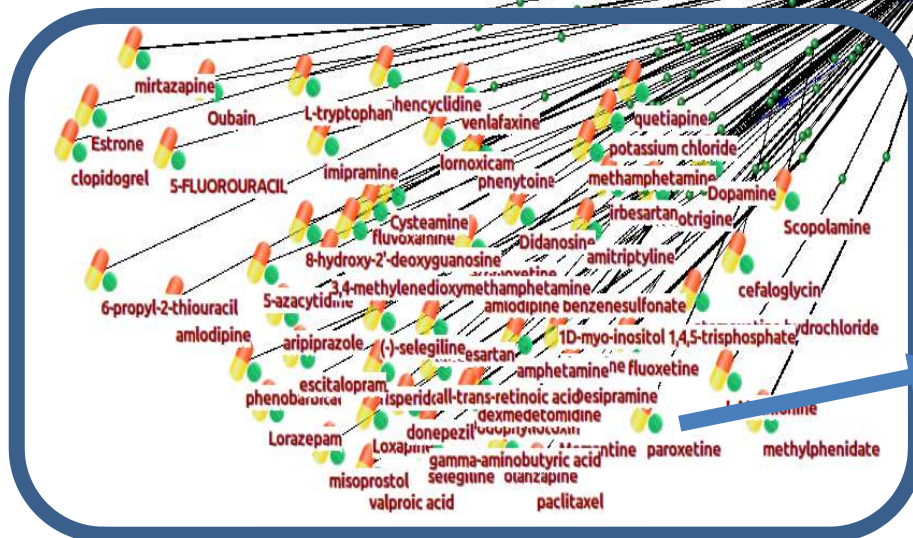


BDNF является мишенью для целого ряда лекарств. Некоторые из этих лекарств могут быть кандидатами для лечения астмы.

Лекарства против астмы



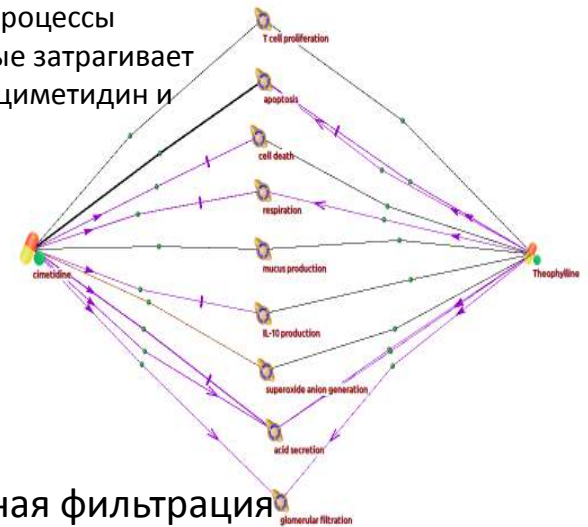
Лекарства против других заболеваний



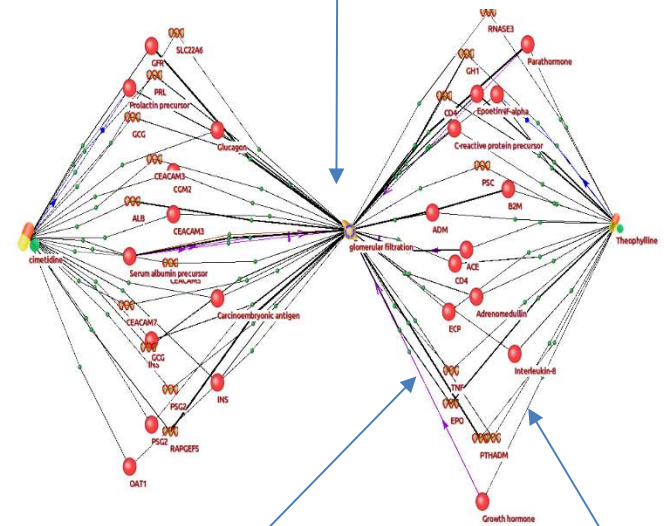
- **Пароксетин** — антидепрессант с сильным противотревожным действием, селективный ингибитор обратного захвата серотонина.

Пример анализа лекарство-лекарство взаимодействий

Биологические процессы человека, которые затрагивает прием лекарств циметидин и теофиллин.

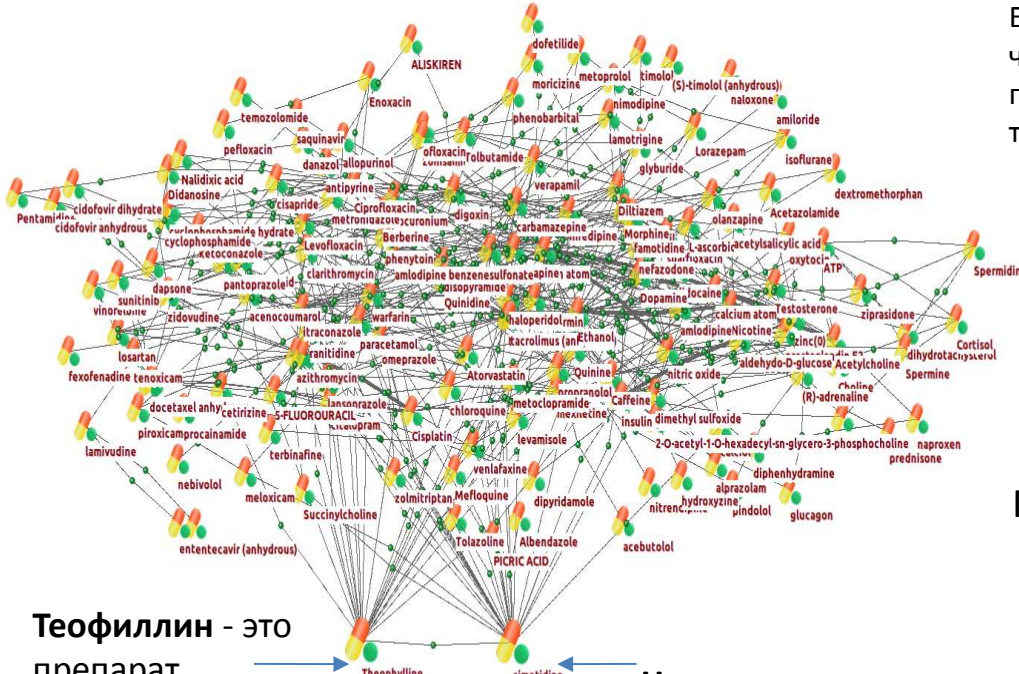


Гломерулярная фильтрация



Гормоны роста увеличивают гломерулярную фильтрацию.

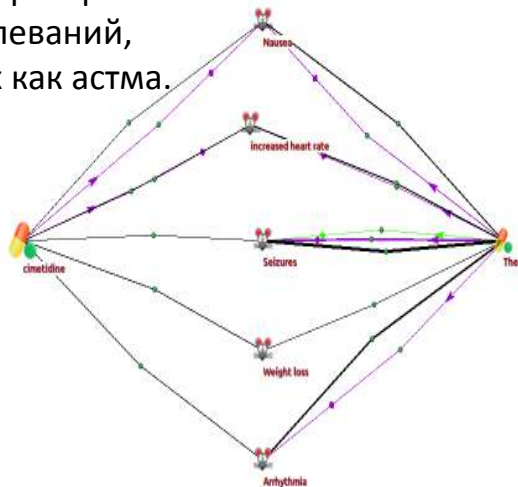
Теофиллин усиливает ответ на гормоны роста.



Теофиллин - это препарат, применяемый в терапии респираторных заболеваний, таких как астма.

Циметидин - препарат, применяемый при терапии язвенной болезни.

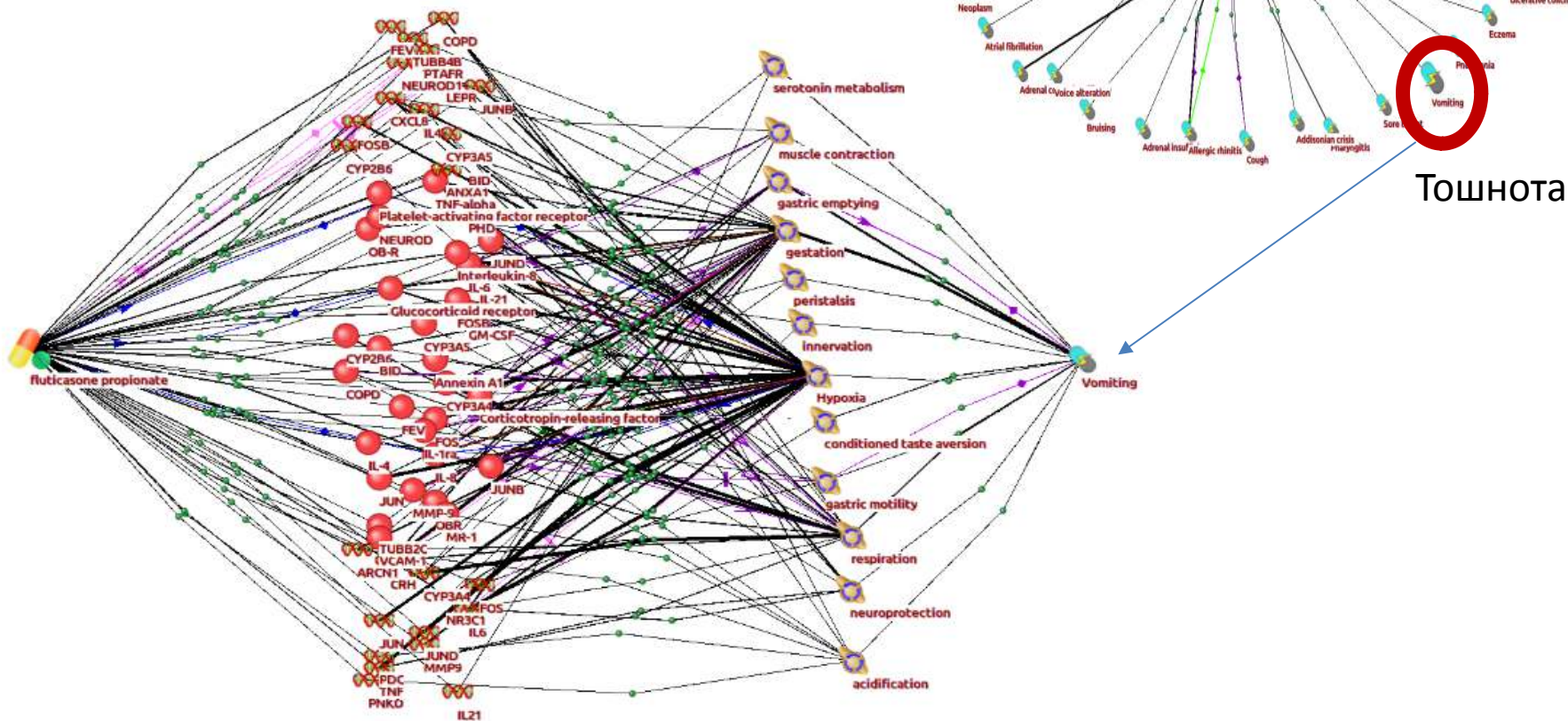
Эти лекарства могут взаимодействовать и вызывать побочные эффекты, такие как потеря аппетита, тошнота, рвота, нервозность, возбуждение, головная боль, нерегулярное сердцебиение или судороги.



Реконструкция молекулярных механизмов, вовлеченных в развитие тошноты - побочного эффекта от приема флутиказона (лекарства от астмы)

Побочные эффекты от приема флутиказона

Сеть молекулярно-генетических взаимодействий, вовлеченных в развитие тошноты при приеме флутиказона.



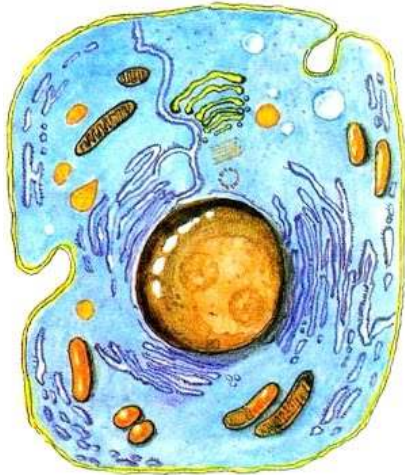
План лекции

- 1) Введение в реконструкцию ассоциативных сетей
- 2) Инструменты реконструкции основанные на текст-майнинге и базах данных (MeSHOPs, COREMINE, ANDSystem)
- 3) Анализ реконструируемых сетей
- 4) Примеры исследований

Выявление структурно-функциональных особенностей ассоциативных генных сетей.

- Выявление кластеров в генных сетях.
- Поиск регуляторных контуров.
- Идентификация ключевых генов, хабов.
- Выявление центральных вершин в генных сетях, расчет показателей центральности вершин.
- Определение сверхпредставленных биологических процессов Gene Ontology (сетевой вариант, ресурс EviNet)
- Приоритезация генов-кандидатов (ресурсы Endeavour, TopGene, DIR).

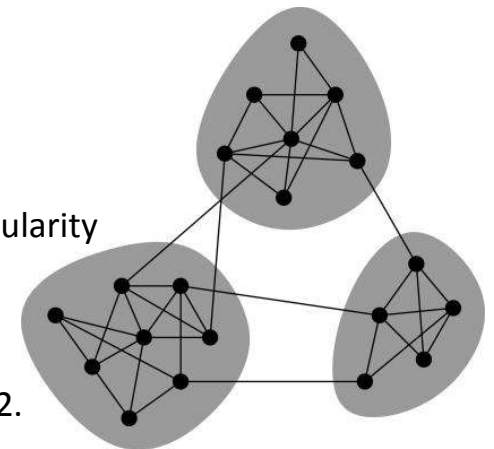
Почему анализ структурно-функциональных характеристик ассоциативных генных сетей имеет смысл?



- Schwikowski et al., 2000 – 63% взаимодействий между генами, имеющими общую функцию
- Grigoriev, 2001 – коэкспрессирующиеся гены часто взаимодействуют
- Ideker, 2002 – взаимодействующие гены часто коэкспрессируются
- Rung *et al.*, 2002 – поиск модульности в сетях
- Rives and Galitski, 2003 – объяснение низкой ассортативности
- Lee *et al.*, 2004 – соответствие сетевых кластеров функциям
- Adamcsek et al., 2006 – поиск сильно связанных подграфов

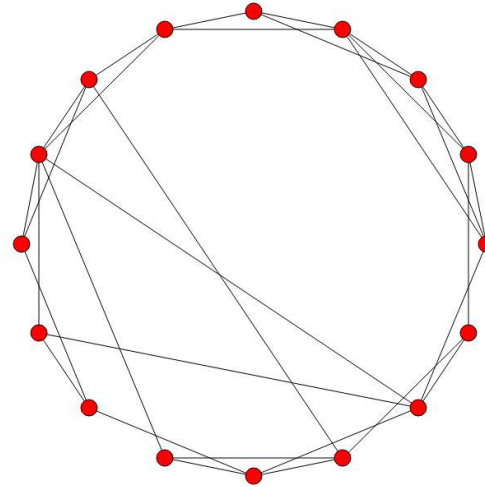
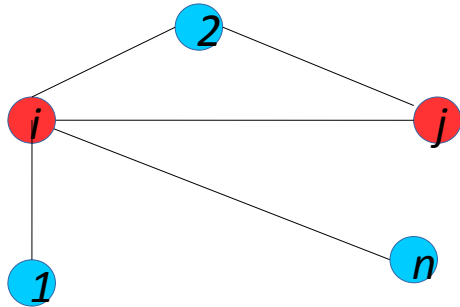
Hartwell, L. H.; Hopfield, J. J.; Leibler, S. & Murray, A. W. (1999), 'From molecular to modular cell biology.', *Nature* **402**(6761 Suppl), C47--C52.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23), 8577-8582.

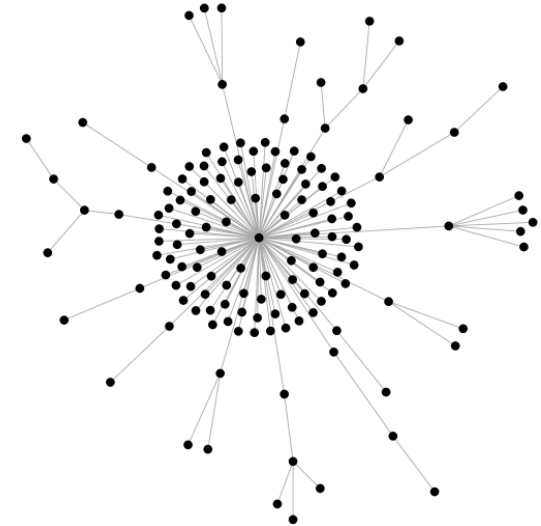


Работы по случайным сетям

- Kauffman, 1969
- Модель Ердёша, Реньи:
 - $P(\{i,j\}) = \text{const}$
- Watts, 1998
- Регулярные сети со случайной перестановкой рёбер

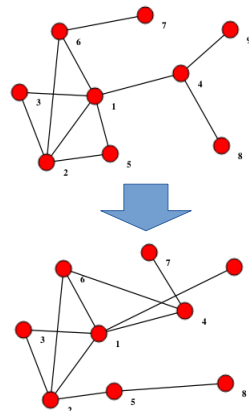


- Jeong *et al.*, 2000
- Fell and Wagner, 2000
- Безмасштабные сети

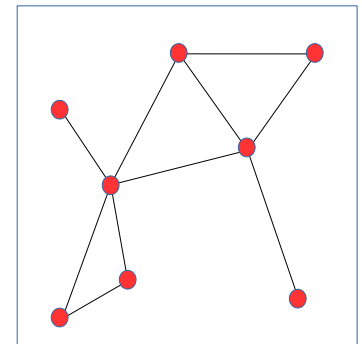


Рандомизация реальных сетей

- Guelzim *et al.*, 2002
- Shen-Orr *et al.*, 2002
- Milo *et al.*, 2002
- Maslov and Sneppen, 2002
- Newman, 2002
- Giot *et al.*, 2003
- Milo *et al.*, 2003



- Przulj *et al.*, 2004, 2007,
- Kuchaiev *et al.*, 2009



Выявление кластеров в генных сетях. Кластеризация генной ассоциативной сети с помощью плагинов для Cytoscape

Структурным модулем или кластером ассоциативной генной сети называется подграф ассоциативной генной сети, обладающий следующими свойствами:

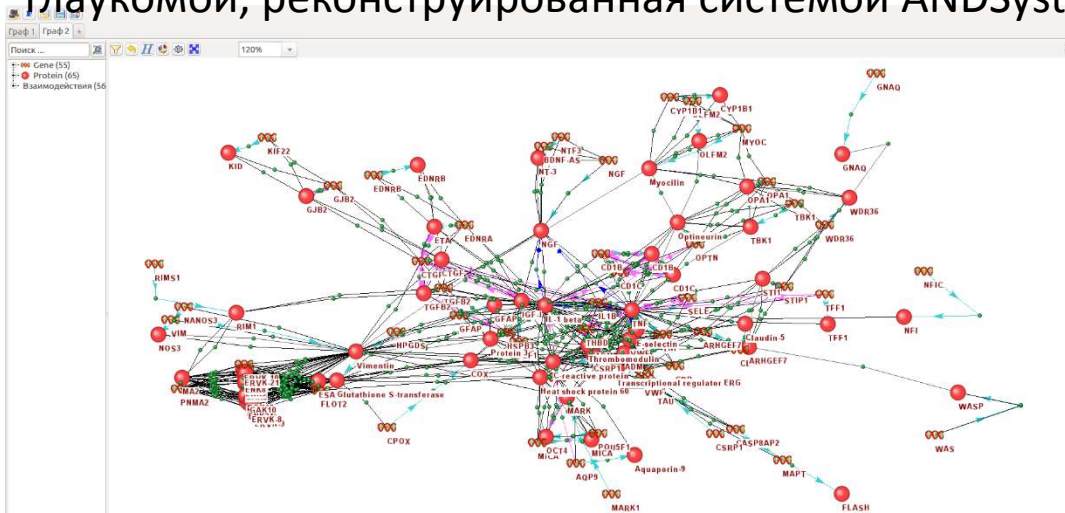
- элементы подграфа сильно связаны между собой и слабо связаны с элементами ассоциативной генной сети не входящими в данный подграф;
- молекулярно-генетической система, описываемая подграфом ассоциативной генной сети, выполняет биологическую функцию.

Список плагинов Cytoscape для кластеризации сетей доступен по адресу:
http://apps.cytoscape.org/apps/with_tag/clustering

The image shows a grid of Cytoscape application cards. Each card includes an icon, the app name, a brief description, and a version number (3.0). The 'ClusterViz' app is highlighted with a red border. The apps listed are:

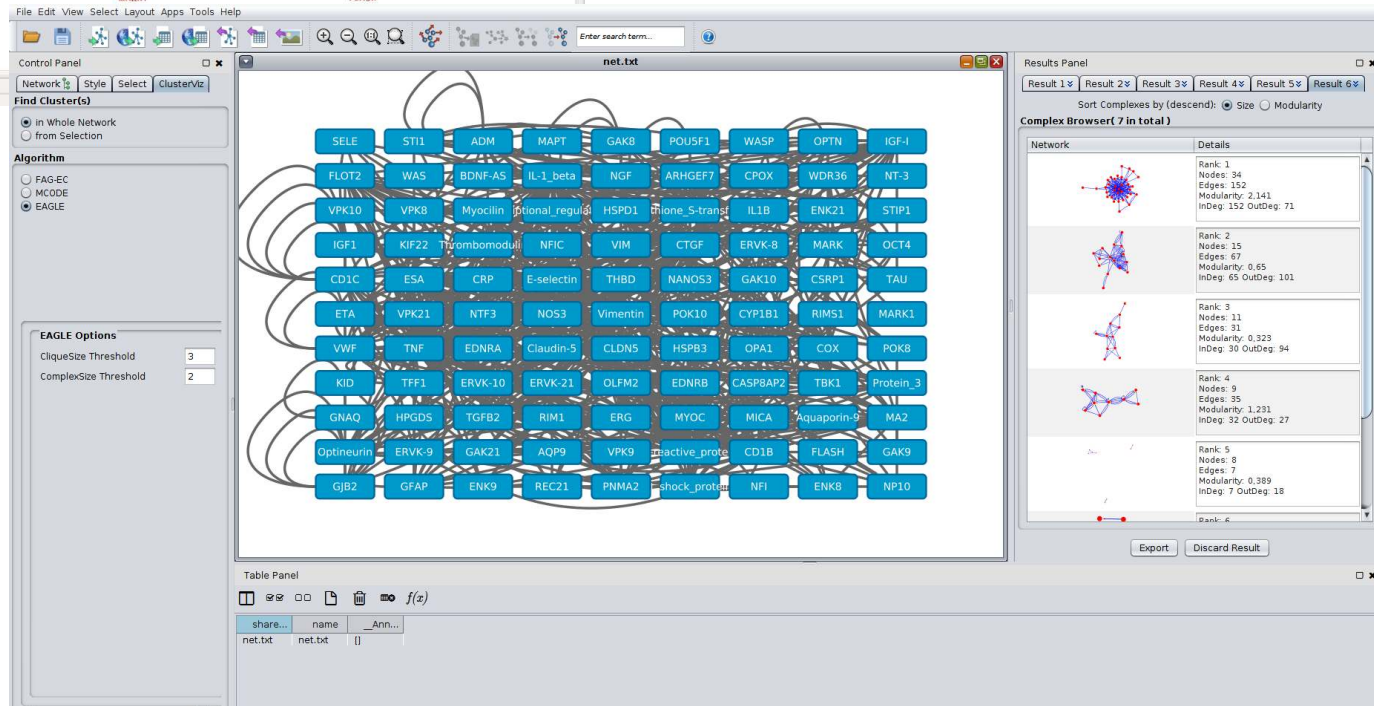
- AutoAnnotate**: Finds clusters and visually annotates them with labels and
- cdtReader**: Read Cluster 3 output files into Cytoscape
- clusterMaker2**: Multi-algorithm clustering app for Cytoscape
- ClusterONE**: Finds overlapping protein complexes in a protein interaction
- ClusterViz**: Clustering based on FAG-EC, EAGLE or MCODE. Found cluster
- CytoCluster**: a Cytoscape app for analysis and visualization of clusters from
- cytoHubba**: Predicts and explores important nodes and subnetworks in a given
- DyNetViewer**: Construction, analysis, and visualization of dynamic networks
- EClerize**
- GIANT**: a Cytoscape plugin for network cartography
- Mass Spec Collection**: Collection of apps to facilitate network analysis of mass
- MCODE**: Clusters a given network based on topology to find densely connected
- ModuLand 2.0**: Modularization method family offering modular hierarchies and
- NCMine**: Finds functional modules in gene networks by Near-Clique Mining
- Omics Analysis Collection**: Collection of apps to facilitate network analysis of omic data. By
- PEmeasure**: Compute links weights and assess the reliability of the links in a
- PEPPER**: Find meaningful pathways / complexes connecting a protein set
- PEWCC**: Detect dense sub-network (e.g. protein/gene complexes) in
- ReactomeFPlugin**: Explore Reactome pathways and search for diseases related
- SCODE**: This is the Cytoscape app implementation for Protein

Ассоциативная генная сеть, ассоциированная с глаукомой, реконструированная системой ANDSystem.



Кластеризация ассоциативной генной сети, ассоциированной с глаукомой, с помощью плагина ClusterViz для Cytoscape.

- 1) Импорт сети из системы ANDSystem в Cytoscape через *File -> import -> network -> file*
- 2) Запуск ClusterViz через *Apps*
- 3) Выбор оптимальных настроек для кластеризации (эмпирическим путем)



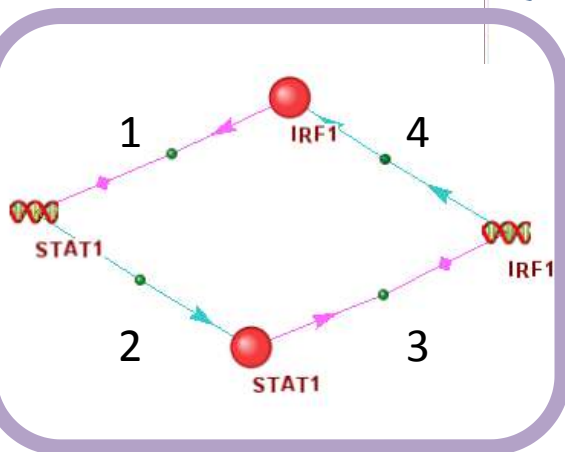
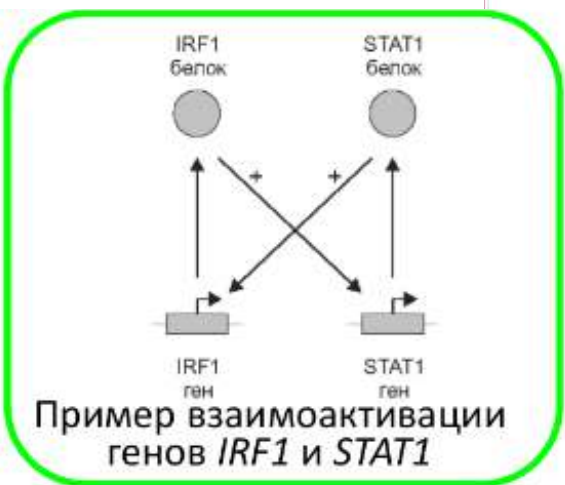
Регуляторные контуры в генных сетях

Авторегуляция генной сети осуществляется благодаря функционированию **регуляторных контуров с положительными и отрицательными обратными связями**, и обеспечивает поддержание определенного, характерного для каждой генной сети типа динамики.

Характерная особенность генных сетей, контролирующих дифференцировку клеток, морфогенез тканей и органов - наличие регуляторных контуров с положительными обратными связями. **Функция контуров с положительной обратной связью** состоит в максимально эффективном отклонении величин контролируемых параметров от их исходных значений. Положительные обратные связи обеспечивают быструю активацию элементов эндокринной и нервной систем (что важно, в частности, для осуществления стрессовых ответов), а также играют ключевую роль в регуляции процессов морфогенеза, роста и развития организмов, по своему смыслу представляющих быстрый уход от имеющегося начального состояния. Характерным примером такого типа генных сетей является генная сеть созревания и дифференцировки эритроцитов.

Контуры с отрицательной обратной связью обеспечивают гомеостатирование внутренней среды организмов, то есть поддерживают контролируемые ими параметры вблизи величин, оптимальных для данных средовых условий.

Поиск регуляторных контуров в ассоциативных генных сетях с помощью функции Pathway Wizard системы ANDSystem.



Мастер путей ...
Длина пути 4

1 Objects: IRF1 (Homo sapiens)

2 Interaction types: Expression Databases: afcs, AGI_LocusCode, AspGD, BioGrid, CGD

3 Objects: IRF1 (Homo sapiens)

4 Interaction types: expression downregulation, expression regulation, expression upregulation Databases: afcs, AGI_LocusCode, AspGD, BioGrid, CGD

Поиск путей

Interaction property <6>

Типы взаимодействия

- activity downregulation
- activity regulation
- activity upregulation
- association
- catalyze
- catalyze modification
- cleavage
- coexpression
- conversion
- degradation downregulation
- degradation regulation
- degradation upregulation
- downregulation
- Expression
- expression downregulation
- expression regulation
- expression upregulation
- interaction
- involvement
- miRNA regulation
- regulation
- transport regulation
- treatment
- upregulation

Выбрать все
Снять отметки

Очистить Закрыть

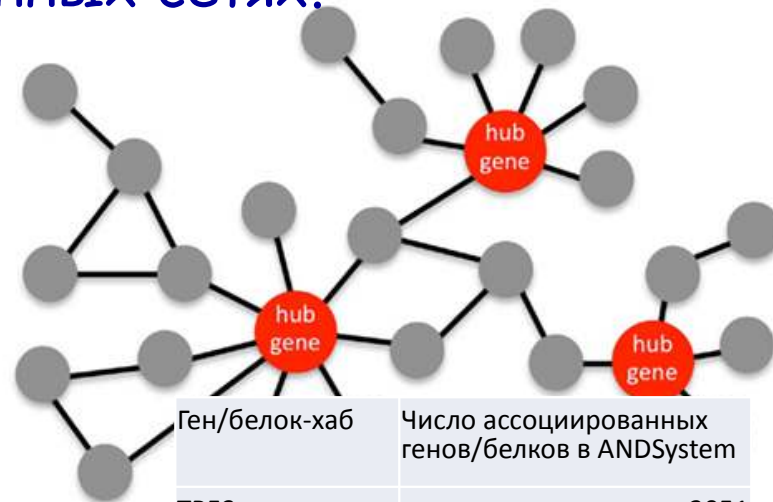
1. Белок *IRF1* -> регуляция экспрессии -> ген *STAT1*
Newly synthesized IRF-1 may in turn activate expression of STAT1, resulting in positive feedback regulation of IRF-1 expression (55). Pubmed:23420765.
2. Ген *STAT1* -> экспрессия -> белок *STAT1*
3. Белок *STAT1* -> регуляция экспрессии -> ген *IRF1*
Further, STAT1 induces the expression of pro-inflammatory genes such as IRF1...Pubmed:25414701
4. Ген *IRF1* -> экспрессия -> белок *IRF1*

Разработанный алгоритм **Pathway Wizard** системы ANDSystem для поиска путей в графах сетей, отвечающих заданным свойствам, основан на переборе с ограничениями, задаваемыми правилами или шаблоном. Путем является линейная цепочка объектов в графе. Шаблон однозначно задает начальный и конечный объекты для пути, а также длину пути или количество входящих в путь объектов. Промежуточные объекты могут задаваться типом, либо может указываться конкретное имя некоего объекта. Связи между объектами могут задаваться фиксированного типа, либо набором возможных типов по правилу ИЛИ.

Выявление центральных вершин в генных сетях. Идентификация хабов.

Хабы – вершины с максимальным количеством рёбер в генных сетях. Хабы обычно соответствуют белкам – регуляторным молекулам.

Структурный анализ ассоциативных или генных сетей имеет прямое практическое применение, особенно в фармакологии. На его основе проводят поиск белков-мишеней для создания новых лекарственных препаратов. Для этого необходимо выявление ключевых узлов с большим количеством связей как потенциальных мишеней. Анализ связи между белками, компонентами генной сети полезен для изучения последствий ингибирования или, напротив, сверхэкспрессии белка-мишени, которые могут сказываться на поведении всей системы в целом.



Ген/белок-хаб	Число ассоциированных генов/белков в ANDSystem
TP53	2051
TNF	1634
EGFR	1347
IL6	1282
NFKB1	1163
INS	1087
MAPK1	1061
TGFB1	897

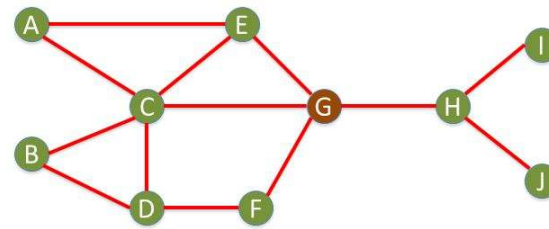
Центральность по степени (degree centrality, DC) - это отношение количества связей определенного узла к общему количеству других узлов в сети. Центральность по степени тем выше, чем больше число связей вершины с другими вершинами в графе. Данная мера принимает значение от 0 до 1 (чем ближе значение к единице, тем больше доля вершин в сети, с которыми напрямую взаимодействует рассматриваемая вершина). Узел с высоким значением DC взаимодействует с большим числом других узлов в сети и, тем самым, может получить доступ к большому числу узлов и повлиять на них.

В случае направленной сети существует две отдельные меры DC: входящая (in-degree) и исходящая (out-degree). Первая показывает общее число вершин, с которыми рассматриваемая вершина связана входящими в эту вершину дугами (числом связей, направленных к узлу); вторая показывает количество вершин, связанных с рассматриваемой вершиной исходящими из нее дугами (числом связей, направленных от узла). Если $DC = 1$, это указывает на то, что определённый узел связан со всеми остальными узлами сети, в то время как $DC = 0$ указывает на то, что узел изолирован.

Расчет показателей центральности вершин. Центральность вершины графа по близости.

Центральность вершины графа по близости (closeness centrality)

пропорциональна сумме кратчайших расстояний от этой вершины до других вершин в графе. Этот показатель определяет важность конкретной вершины графа для быстрой передачи информации в графе. Чем больше показатель центральности по близости, тем короче средний путь до любой вершины графа (ближе узел к другим вершинам).



G→A: 2	H→A: 3
G→B: 2	H→B: 3
G→C: 1	H→C: 2
G→D: 2	H→D: 3
G→E: 1	H→E: 2
G→F: 1	H→F: 2
G→H: 1	H→G: 1
G→I: 2	H→I: 1
G→J: 2	H→J: 1

Total=14 Total=18

Closeness centrality G: $1/14=0,071$ ⁷⁸

Closeness centrality H: $1/18=0,055$

$$C(x) = \frac{1}{\sum_y d(y, x)}$$

$d(y,x)$ – расстояние между вершинами y и x .

Расчет показателей центральности вершин. Центральность вершины по посредничеству.

Центральность вершины по посредничеству

(betweenness centrality)

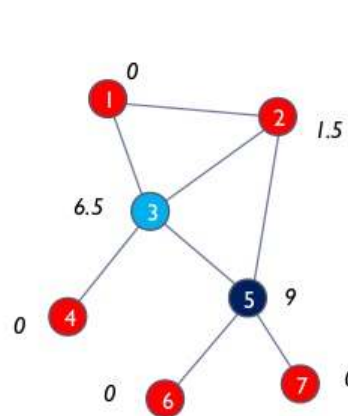
– взвешенная сумма всех кратчайших путей между вершинами в графе, проходящих через данную вершину. Этот показатель определяет важность вершины графа с точки зрения ее влияния на пути передачи информации между вершинами графа.

$$c_B(e) = \sum_{s,t \in V} \frac{\sigma(s,t|e)}{\sigma(s,t)}$$

V- набор вершин;

$\sigma(s,t)$ – число наикратчайших путей между s и t;

$\sigma(s,t|e)$ - число наикратчайших путей между s и t, проходящих через вершину e.



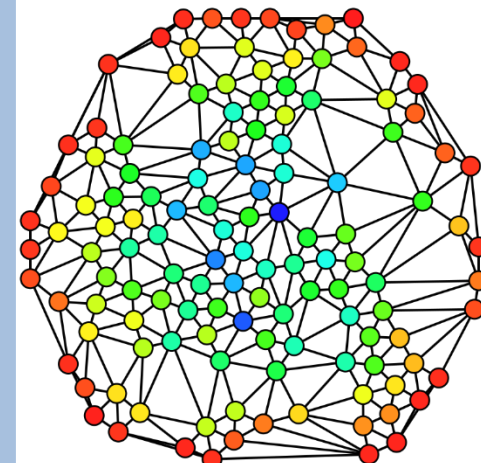
2	3	5
Путь: показатель	Путь: показатель	Путь: показатель
1->5: 0,5	1->4: 1	1->7: 1
1->7: 0,5	1->5: 0,5	1->6: 1
1->6: 0,5	1->6: 0,5	3->6: 1
Сумма: 1,5	1->7: 0,5	3->7: 1
	4->5: 1	2->6: 1
	4->2: 1	2->7: 1
	4->7: 1	4->7: 1
	4->6: 1	4->6: 1
	Сумма: 6,5	6->7: 1
		Сумма: 9

Центральность вершины 2 по посредничеству: 1,5

Центральность вершины 3 по посредничеству: 6,5

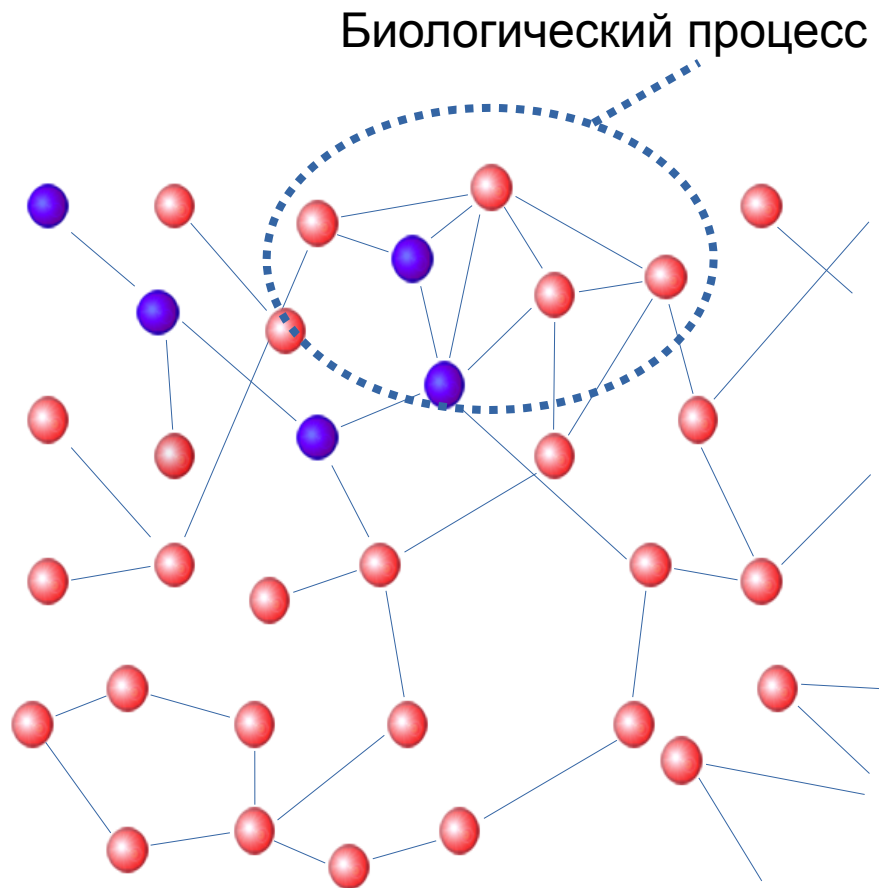
Центральность вершины 5 по посредничеству: 9

- 1) Для вершины v надо рассчитать число наикратчайших путей между вершинами i и j, которые проходят через вершину v и разделить это значение на число всех наикратчайших путей между вершинами i и j.
- 2) Нужно сложить полученные значения для всех вершин i и j.



Синие вершины имеют максимальный показатель центральности по посредничеству.

Сверхпредставленность биологических процессов с учётом топологии сети (EviNet)



● Экспериментально полученные гены

EviNet: evidence-based network enrichment analysis

Project: [-- project ID --]

1: Altered gene sets 2: Network 3: Functional gene sets

Select experimentally derived list(s) of genes/proteins that you want to change

Genes File Venn diagram

Paste a list of IDs:

The screenshot shows the EviNet web interface. At the top, there is a blue header with the text 'EviNet: evidence-based network enrichment analysis'. Below the header, there is a navigation bar with four tabs: '1: Altered gene sets', '2: Network', '3: Functional gene sets', and '4: ...'. The '1: Altered gene sets' tab is selected. Below the navigation bar, there is a text input field with the placeholder text 'Paste a list of IDs:'. Above the input field, there are three tabs: 'Genes', 'File', and 'Venn diagram', with 'Genes' selected. To the right of the input field, there is a 'Project' dropdown menu with the text '[-- project ID --]'. There are also several icons, including a question mark and a 'SHOW ME HOW' icon.

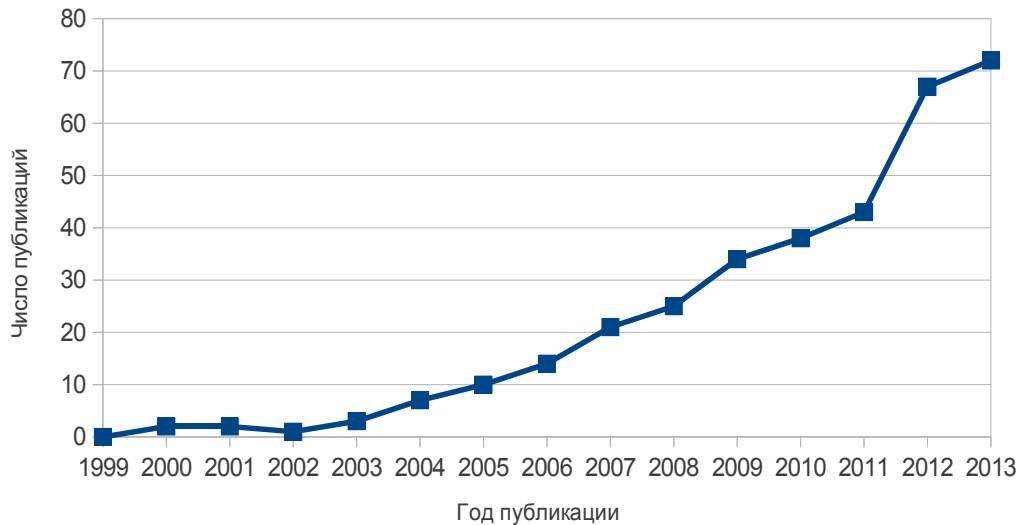
Jeggari A. et al. EviNet: a web platform for network enrichment analysis with flexible definition of gene sets //Nucleic acids research. – 2018. – T. 46. – №. W1. – C. W163-W170.

Приоритезация генов

Доказательство причинной связи между геном и болезнью с помощью эксперимента требует много денежных и временных затрат. Всесторонняя приоритезация генов-кандидатов до экспериментальной проверки снижает связанные с этим расходы.



Рост числа документов в PubMed по запросу

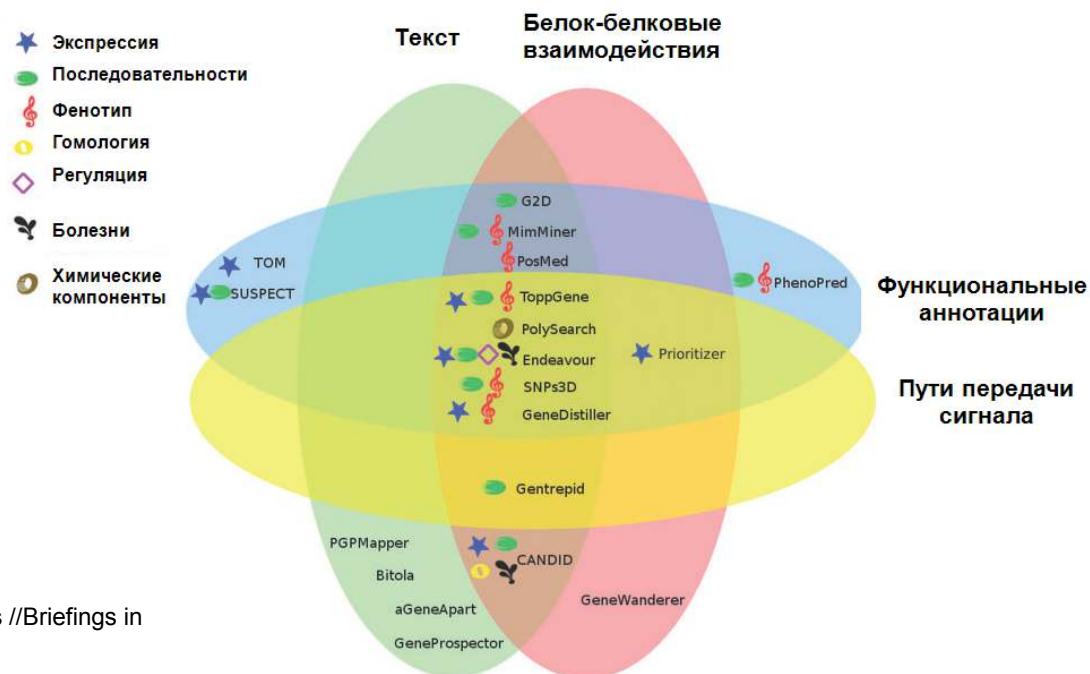


Перечень некоторых инструментов приоритезации генов

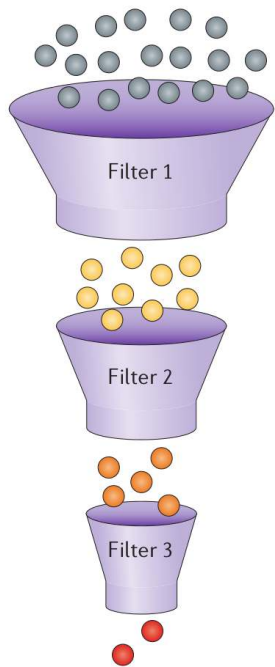
Инструменты	Входные параметры					Выходные параметры		
	Тренировочные Данные		Гены-кандидаты			Ранжирование	Выбранные кандидаты	Статистика
	Известные гены	Ключевые слова	Регион	DEG	Геном			
SUSPECT		x	x	x		x		
ToppGene	x			x		x		x
PolySearch		x			x	x		x
MimMiner		x			x	x		
PhenoPred		x			x	x		
PGMapper		x	x	x		x		
Endeavour	x	x	x	x	x	x		x
G2D	x	x	x			x		x
TOM	x		x				x	
SNPs3D		x			x	x		
GenTrepid		x	x	x		x	x	
GeneWanderer	x		x	x		x		x
Bitola		x	x		x	x	x	
CANDID		x				x		x
aGeneApart		x			x	x		x
GeneProspector		x			x	x		
PosMed		x	x		x	x	x	x
GeneDistiller	x	x	x			x		

Источники данных.

Для повышения качества приоритезации генов в настоящее время активно используется интеграция разнородных данных из множественных источников, в том числе данные по экспрессии, гомологии последовательностей, полиморфизмам, белок-белковым взаимодействиям, публикациям и др.

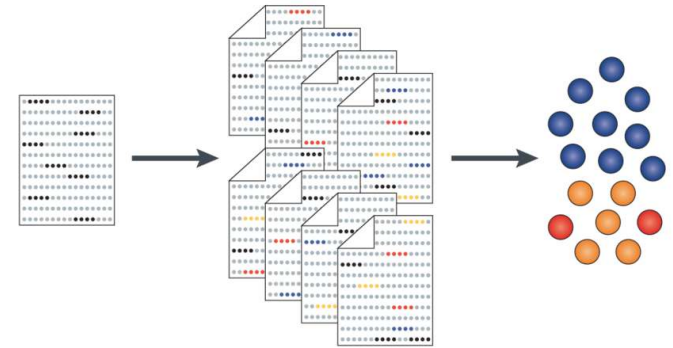


Стратегии приоритезации генов






При **text-mining** используется набор ключевых слов для поиска документов, которые имеют отношение к изучаемому процессу. Автоматическими методами извлекаются гены, встречающиеся в этих документах. Проводится статистический анализ извлеченной информации, для ранжирования генов-кандидатов.

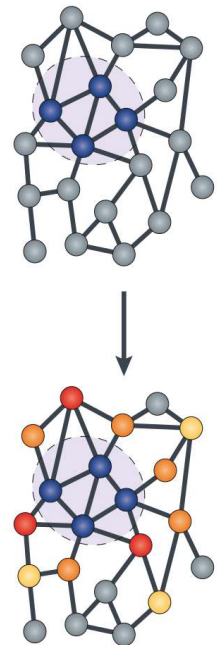
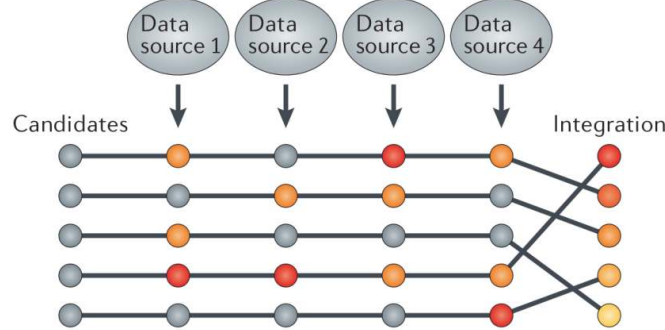
Недостатком метода **text-mining** является то, что плохо изученные гены, которые редко упоминаются в публикациях, выпадают из анализа.



Фильтрация генов-кандидатов на основе их соответствия различным критериям (вовлечение в Gene Ontology процессы, упоминания в GWAS исследованиях и др.). Основным ограничением является то, что строгий процесс фильтрации не позволяет проводить тонкий анализ набора генов-кандидатов. Если ген-кандидат не отвечает хотя бы одному из критериев, он отфильтровывается, что может привести к ложно отрицательному результату.

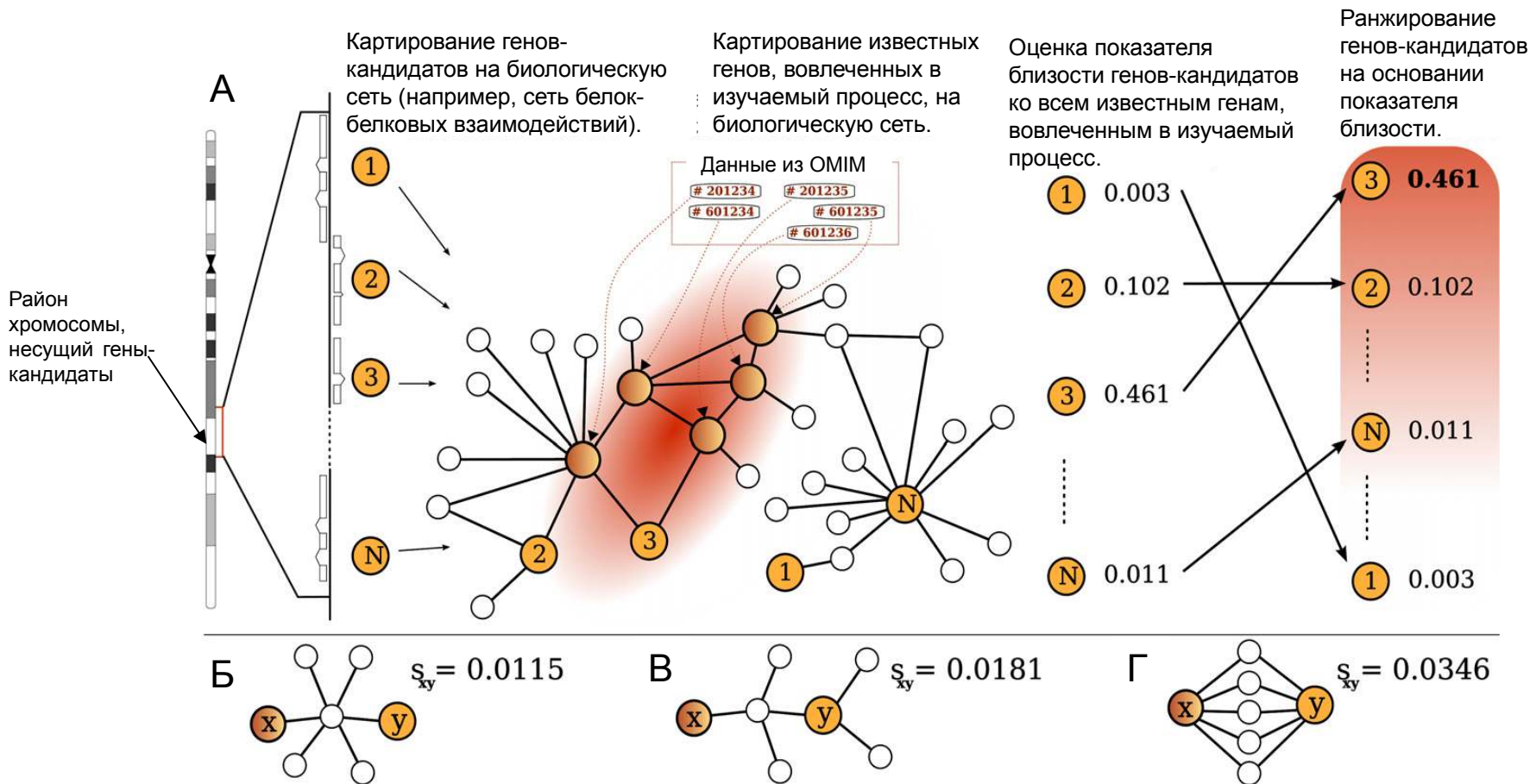
-  Известные гены, вовлеченные в изучаемый процесс
-  Гены-кандидаты
-  Ранжированные гены-кандидаты

Методы оценки сходства различных биологических профилей определяют наиболее перспективных генов-кандидатов в соответствии с их сходством по различным параметрам (Gene Ontology, профили экспрессии, сходство последовательностей, данные по модельным организмам и др.) с уже известными генами, вовлеченными в изучаемый процесс.



Анализ биологических сетей. В сети находят известные гены, связанные с изучаемой областью (например болезнь). Затем выбираются гены-кандидаты на основе расстояния от известных генов.

Анализ биологических сетей при приоритизации генов

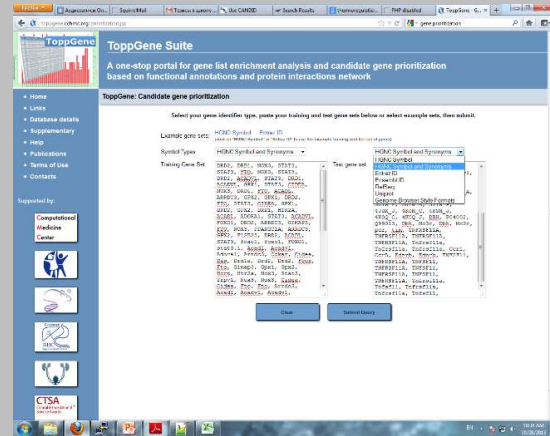


Каждая из трех подсетей **Б**, **В** и **Г** имеет одинаковое количество вершин, но различную конфигурацию и соответственно показатель близости генов **X** и **Y**. В подсети **Б** вершины **X** и **Y** связаны через хаб, имеющий множество связей с другими вершинами, поэтому показатель близости s_{xy} имеет более низкое значение, чем в случае **В**, где вершины **X** и **Y** связаны через вершину, имеющую малое количество связей. Наибольший показатель близости имеет подсеть **Г**, где вершины **X** и **Y** связаны сразу через несколько вершин. Если не учитывать топологию сети, то эти три случая становятся не различимыми, так как расстояние между **X** и **Y** одинаково во всех трех случаях.

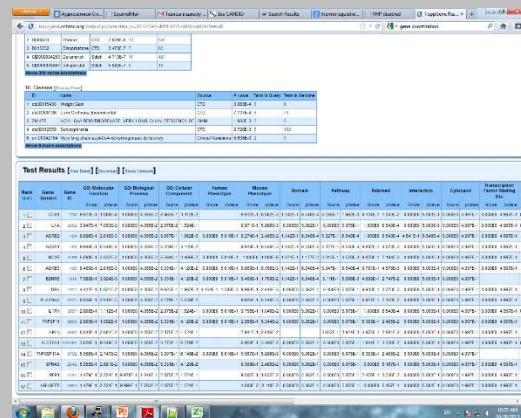
Приоритезация генов-кандидатов (ресурсы Endeavour, ToppGene, DIR).

ToppGene (<https://toppgene.cchmc.org>) используется информация о фенотипах мыши и человека, а также аннотация генов из литературы.

Начальная страница ToppGene

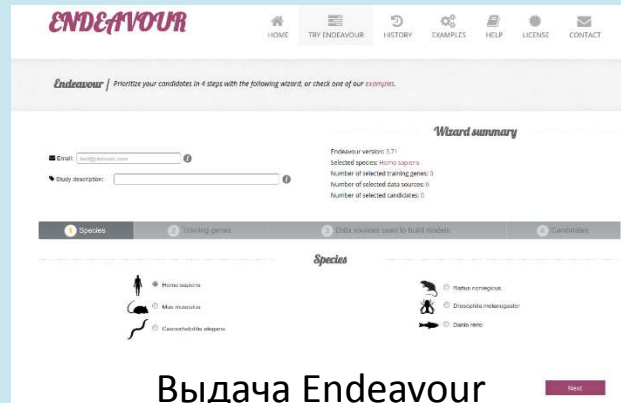


Выдача ToppGene

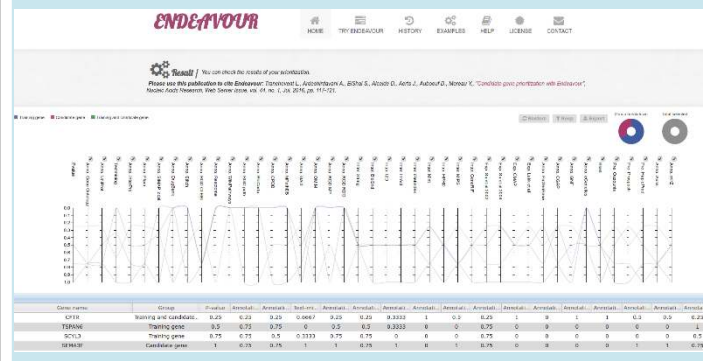


Endeavour (<https://endeavour.esat.kuleuven.be>) используются обширные геномные данные из различных источников, а также информация о связи генов и биологических процессов.

Начальная страница Endeavour



Выдача Endeavour



DIR (<http://cbc.case.edu/dir/>) используется анализ геномных сетей и сетей ген-заболевание в дополнение к геномной информации из различных источников.

Выдача DIR
Informativity of the Disease Genes

Network	Informativeness
ppi	0.922350
exp	0.092889
pwy	3.470162

The yellow rows indicate the the values above the threshold of: 3.876563

Rank	Gene	Score
1	6870	4.049330
2	2895	4.041079
3	5602	2.231536
4	6622	2.079720
5	658	2.073226
6	5530	1.994358

Endeavour

1) Выбор организма

ENDEAVOUR

HOME TRY ENDEAVOUR HISTORY EXAMPLES HELP LICENSE CONTACT

Endeavour | Prioritize your candidates in 4 steps with the following wizard, or check one of our examples.

Wizard summary

Email:
Study descriptor:

Endeavour version: 3.71
Selected species: Homo sapiens
Number of selected training genes: 0
Number of selected data sources: 0
Number of selected candidates: 0



2) Загрузка тренировочного набора генов

Wizard summary

Endeavour version: 3.71
Selected species: Homo sapiens
Number of selected training genes: 3
Number of selected data sources: 0
Number of selected candidates: 0

1 Species 2 Training genes 3 Data sources used to build models 4 Candidates

Species

Homo sapiens
 Mus musculus
 Caenorhabditis elegans

Status response
 Disruptible mechanism
 Data type

Next

Training genes

Gene Chromosome regions Gene Ontology OMIM List

Choose gene: Add

Gene (Reference ID)	Alias	Description	Clear	Remove selected item
ENSG00000000003	TSPAN6	tetraspanin 6 [Source:HGNC Symbol;Acc:11858]	<input type="checkbox"/>	<input type="checkbox"/>
ENSG00000001626	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Source:HGNC Symbol;Acc:1884]	<input type="checkbox"/>	<input type="checkbox"/>
ENSG00000000457	SCYL3	SCY1-like 3 (S. cerevisiae) [Source:HGNC Symbol;Acc:19285]	<input type="checkbox"/>	<input type="checkbox"/>



3) Выбор источников аннотации генов для приоритезации

1 Species 2 Training genes 3 Data sources used to build models 4 Candidates

Data Sources

Select all Select none

Gene and protein function

Annotation - Gene Ontology
 Annotation - Pfam

Chemical information

Annotation - DrugBank

Bio-molecular pathways

Annotation - Reactome
 Annotation - CPDB

Phenotypic information

Annotation - GAD

Interaction networks

Interaction - String
 Interaction - BioGRID

Annotation - UniProt
 Annotation - SIMAP (localization)

Annotation - Stitch

Annotation - WikiPathways
 Annotation - hiPathDB

Annotation - OMIM

Interaction - BioGRID
 Interaction - STRING

Text-mining

Annotation - RGD CHEBI

Annotation - RGD pathways

Annotation - RGD MP

Interaction - I2D
 Interaction - LMPD

Annotation - InterPro

Annotation - BioCarta

Annotation - RGD RDO

Interaction - IntAct
 Interaction - LMPD



5) Приоритизированный список генов-кандидатов

ENDEAVOUR

HOME TRY ENDEAVOUR HISTORY EXAMPLES HELP LICENSE CONTACT

Read this first publication on the Endeavour Wizard: ["Prioritization of Candidates for Gene Identification"](#).
Muller, Anshu; Neuman, PhD; Stone, et al. (4, 16, 1, 28, 2012, pp. 11-22).

Endeavour | Prioritize your candidates in 4 steps with the following wizard, or check one of our examples.

4) Загрузка генов-кандидатов

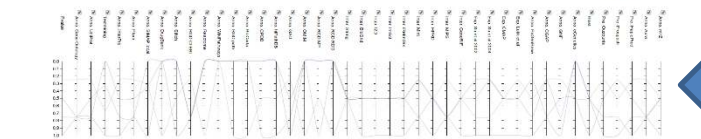
Candidates

1 Species 2 Training genes 3 Data sources used to build models 4 Candidates

Gene Chromosome regions Gene Ontology OMIM List

Choose gene: Add

Gene (Reference ID)	Alias	Description	Clear	Remove selected item
ENSG00000001617	SEMA3F	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3F [Source:HGNC Symbol;Acc:10728]	<input type="checkbox"/>	<input type="checkbox"/>
ENSG00000001626	CFTR	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) [Source:HGNC Symbol;Acc:1884]	<input type="checkbox"/>	<input type="checkbox"/>



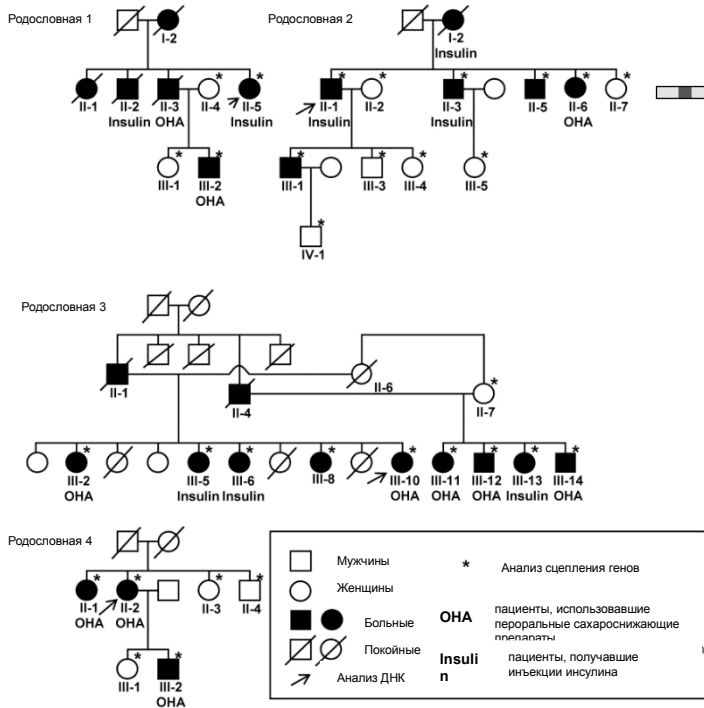
Gene name	Gene	Strand	Start	End	Score	Rank	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation	Annotation
CFTR	Training gene	+	8,25	8,25	0,4667	9,25	0,25	0,3333	1	0,3	0,25	1	1	1	0,3	0,3	0,25		
SEMA3F	Training gene	+	3,5	3,5	0,75	0	0,5	0,25	0	0,75	0	0	0	0	0	0	0	0	0
SCYL3	Training gene	+	0,75	0,75	0,5	0,333	0,75	0,75	0	0,75	0	0	0	0	0	0	0	0	0,75
SEMA3F	Candidate gene	+	1	1	0,75	1	1	0,75	1	1	0,75	0	0	0	1	1	1	1	0,75

План лекции

- 1) Введение в реконструкцию ассоциативных сетей
- 2) Инструменты реконструкции основанные на текст-майнинге и базах данных (MeSHOPs, COREMINE, ANDSystem)
- 3) Анализ реконструируемых сетей
- 4) Примеры исследований

Выявление мутаций в гене *GCKR* у японцев, страдающих диабетом 2 типа.

Родословные 4 японских семей, страдающих сахарным диабетом.



Хромосома 2



Район хромосомы 2p25-22, для которого ранее была показана связь с сахарным диабетом 2 типа (Mori et al., 2002; Iwasaki et al., 2003; Nawata et al., 2004), содержащий 106 генов.

Приоритизация 106 генов из района 2p25-22 с помощью сервиса **Endeavour** на наборе обучающих генов (*HNF4A*, *GCK*, *HNF1A*, *PDX1*, *HNF1B*, *NEUROD1*).

Топовые гены-кандидаты

Ранг	Ген	Название гена
1	<i>GCKR</i>	Glucokinase regulatory protein
2	<i>NCOA1</i>	Nuclear receptor coactivator 1
3	<i>FOSL2</i>	Fos-related antigen 2
4	<i>APOB</i>	Apolipoprotein B-100 precursor
5	<i>MYCN</i>	N-myc proto-oncogene protein
6	<i>RBKS</i>	Ribokinase
7	<i>XDH</i>	Xanthine dehydrogenase/oxidase
8	<i>KHK</i>	Ketohexokinase
9	<i>NRBP1</i>	Nuclear receptor-binding protein
10	<i>CAD</i>	Glutamine-dependent carbamoyl-phosphate synthase
11	<i>RDH14</i>	Retinol dehydrogenase

Мутации *g.-689G>A*, *g.-299G>A* и *g.6859C>G* в гене ***GCKR*** статистически более часто встречались у членов семей, страдающих диабетом 2 типа.

Экспериментальный поиск мутаций в генах-кандидатах



Врожденная генерализованная липодистрофия (CGL4), вызванная мутацией в гене PTRF-CAVIN.

Пациенты, страдающие врожденной генерализованной липодистрофией (CGL4).



Хромосома 17



Район хромосомы, гомозиготный у всех семи обследуемых пациентов с врожденной генерализованной липодистрофией (CGL4) по данным микрочипового анализа (Affymetrix GeneChip SNP, HomozygosityMapper), содержащий 74 гена.

Приоритезация 74 генов с помощью сервиса **GeneDistiller** по экспрессии в адипоцитах, гладких и сердечных мышцах.

Ген-кандидат **PTRF-CAVIN**

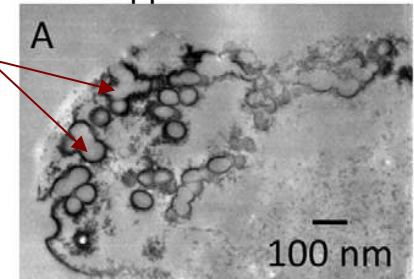
Секвенирование гена PTRF-CAVIN

Вестерн-блот анализ

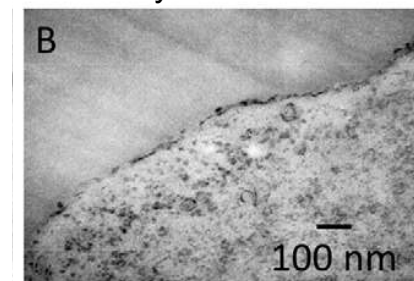
Кавеолы, в образовании которых участвует продукт гена **PTRF-CAVIN**.

Электронные микрофотографии (x 37.000) клеточной поверхности фибробласта

Дикий тип



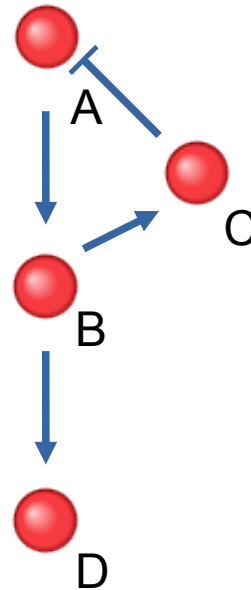
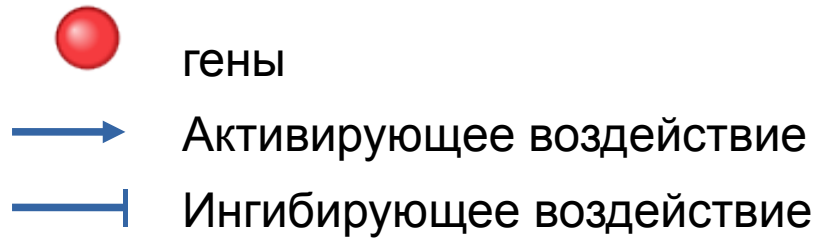
Мутантный тип



Мутация **c.160delG** в гене **PTRF-CAVIN**, приводящая к образованию дополнительного стоп-кодона, присутствует у всех 7 обследованных пациентов.



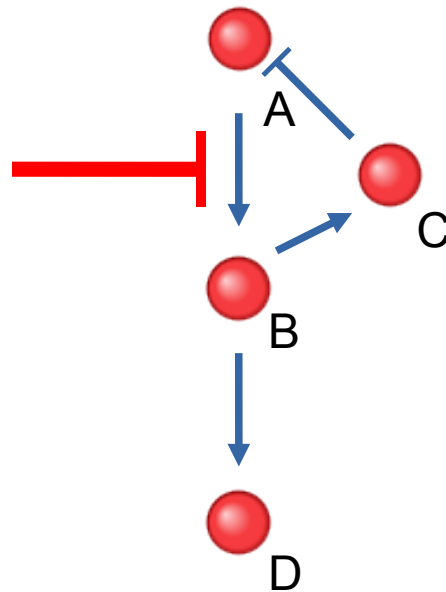
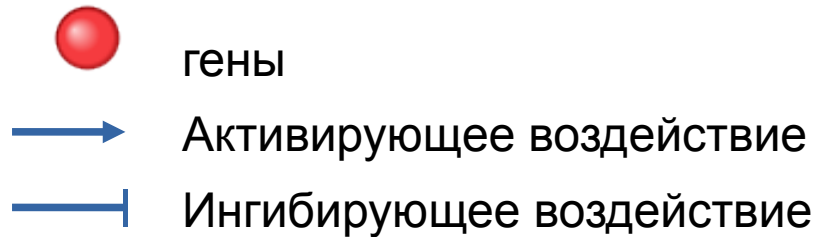
Поиск оптимальных мишеней лекарств



Sridhar, P., Kahveci, T., & Ranka, S. (2007). An iterative algorithm for metabolic network-based drug target identification. In *Biocomputing 2007* (pp. 88-99).

Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nature reviews Drug discovery*, 6(3), 202.

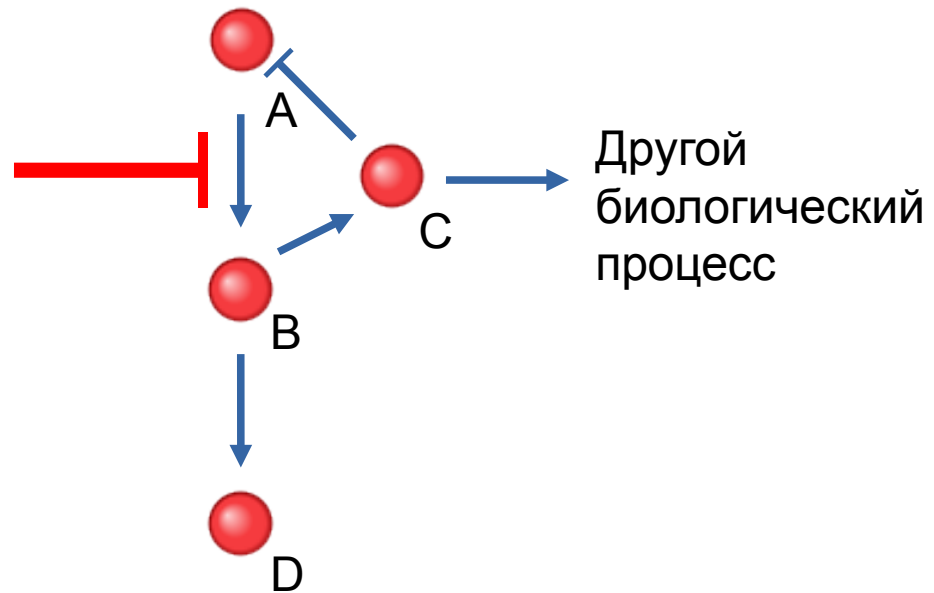
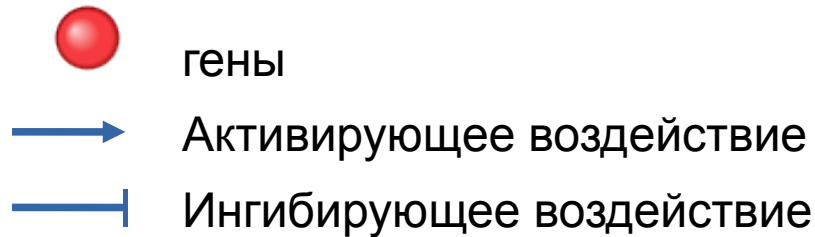
Поиск оптимальных мишеней лекарств



Sridhar, P., Kahveci, T., & Ranka, S. (2007). An iterative algorithm for metabolic network-based drug target identification. In *Biocomputing 2007* (pp. 88-99).

Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nature reviews Drug discovery*, 6(3), 202.

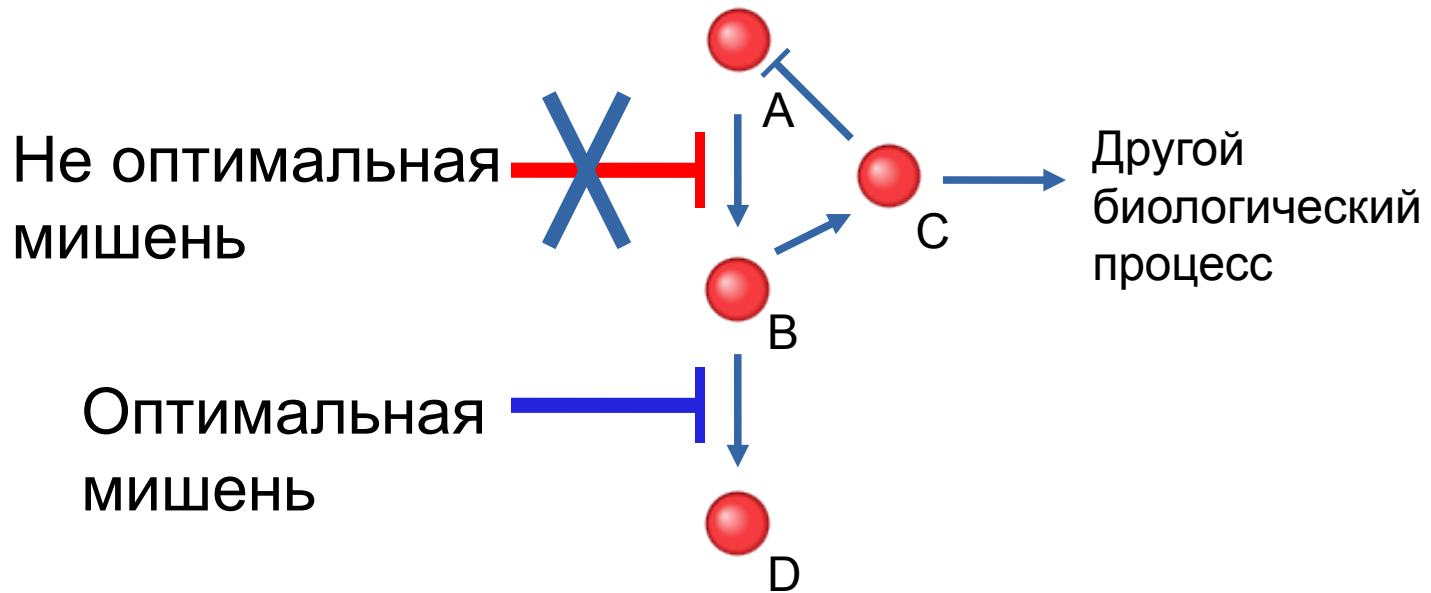
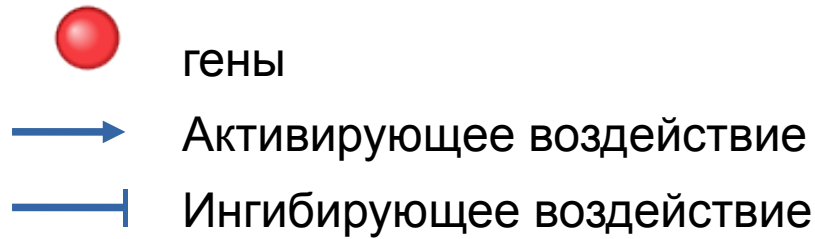
Поиск оптимальных мишеней лекарств



Sridhar, P., Kahveci, T., & Ranka, S. (2007). An iterative algorithm for metabolic network-based drug target identification. In *Biocomputing 2007* (pp. 88-99).

Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nature reviews Drug discovery*, 6(3), 202.

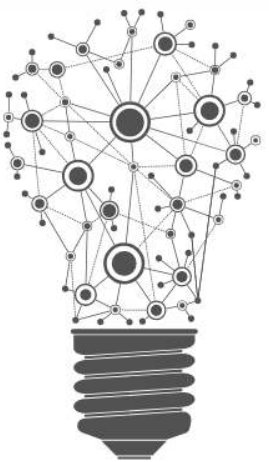
Поиск оптимальных мишеней лекарств



Sridhar, P., Kahveci, T., & Ranka, S. (2007). An iterative algorithm for metabolic network-based drug target identification. In *Biocomputing 2007* (pp. 88-99).

Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nature reviews Drug discovery*, 6(3), 202.

Выявление генетического базиса синтропии/дистропии заболеваний



Fun Gene Net

Gene Network parameters

Select a source of Network data

ANDSYSTEM

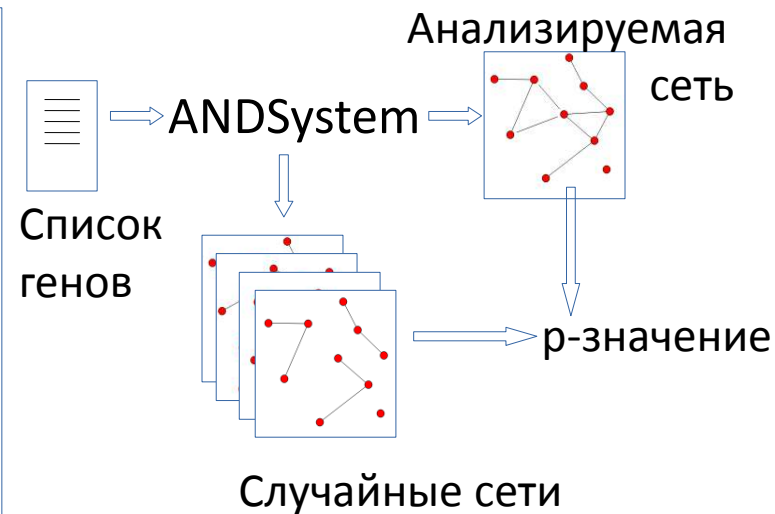
Homo sapiens

'all types'

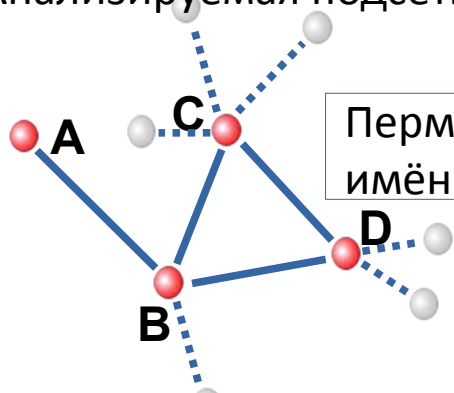
SWISSPROT_ID (ID example: ALBU_HUMAN)

Choose a file...

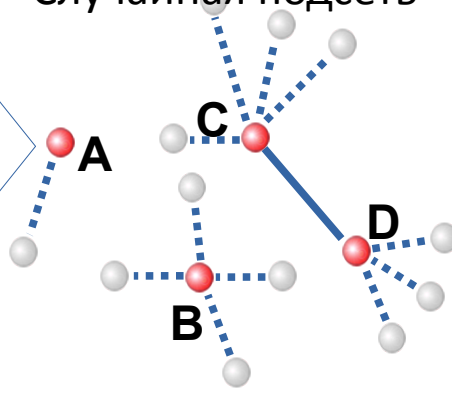
UPLOAD GENE/PROTEIN LIST ...



Анализируемая подсеть



Случайная подсеть



Пермутация имён вершин

- Исходный ген
- Ген, не включённый в сеть
- Связи в рассматриваемой сети
- - - Связи глобальной сети, не вошедшие в рассматриваемую сеть

Выявление генетического базиса синтропии/дистропии заболеваний

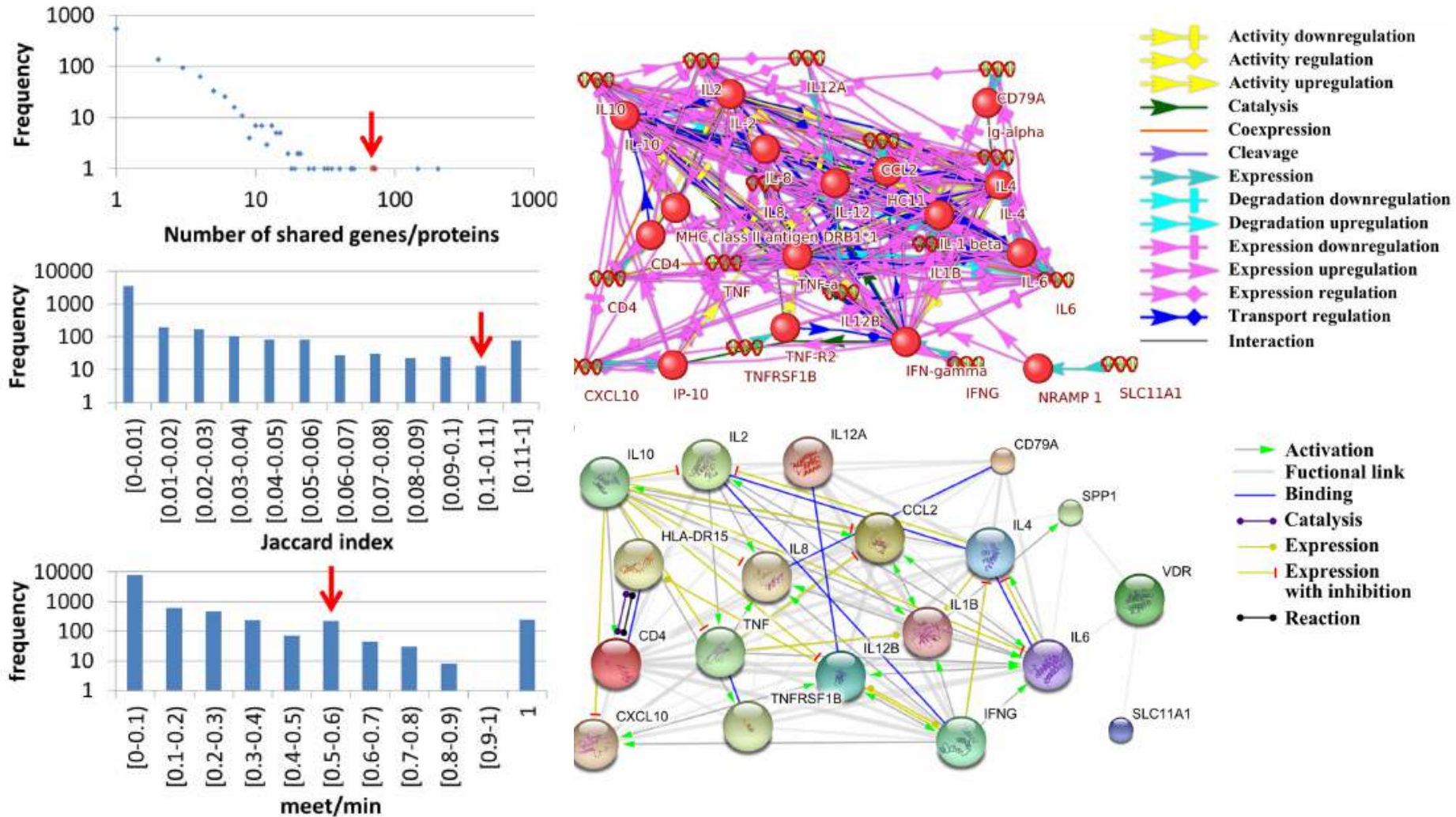


Fig. 2 Distribution of random pairs of diseases according to connectivity measures. The arrows show localization of BA/TB pair

Bragina, E. Y.; Tiys, E. S.; Freidin, M. B.; Koneva, L. A.; Demenkov, P. S.; Ivanisenko, V. A.; Kolchanov, N. A. & Puzyrev, V. P. (2014), 'Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis.', *Immunogenetics* **66**(7-8), 457--465.

Примеры исследований, проведенных с использованием ANDSystem

1. Saik, O. V., Demenkov, P. S., Ivanisenko, T. V., Bragina, E. Y., Freidin, M. B., Goncharova, I. A., ... & Rogaev, E. I. (2018). Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Medical Genomics*, 11(1), 15.
2. Сайк, О. В., Деменков, П. С., Иванисенко, Т. В., Колчанов, Н. А., & Иванисенко, В. А. (2017). Разработка методов автоматического извлечения знаний из текстов научных публикаций для создания базы знаний *Solanum tuberosum*. *Сельскохозяйственная биология*, 52(1).
3. Брагин, А. О., Сайк, О. В., Чадаева, И. В., Деменков, П. С., Маркель, А. Л., Орлов, Ю. Л., ... & Иванисенко, В. А. (2018). Роль генов апоптоза в контроле агрессивного поведения, выявленная с помощью комбинированного анализа ассоциативных генных сетей, экспрессионных и геномных данных по серым крысам с агрессивным поведением. *Вавиловский журнал генетики и селекции*, 21(8), 911-919.
4. Olga V Saik, Natalia A Konovalova, Pavel S Demenkov, Timofey V Ivanisenko, Evgeny D Petrovskiy, Nikita V Ivanisenko, Dinara E Ivanoshchuk, Maria N Ponomareva, Olga S Konovalova, Inna N. Lavrik, Nikolay A Kolchanov, Vladimir A Ivanisenko. Molecular associations of Primary Open-Angle Glaucoma with potential comorbid diseases (POAG-associome). *Biotechnologia Aplicada* 2016;33:3201-3206.
5. Saik O.V., Konovalova N.A., Demenkov P.S., Ivanisenko N.V., Ivanisenko T.V., Ivanoshchuk D.E., Konovalova O.S., Podkolodnaya O.A., Lavrik I.N., Kolchanov N.A., Ivanisenko V.A. Molecular-genetic mechanisms of the interaction between processes of cell response to mechanical stress and neuronal apoptosis in primary open-angle glaucoma. *Vavilovskii Zhurnal Genetiki i Selekcii=Vavilov Journal of Genetics and Breeding*. 2016;20(6):840-847. DOI 10.18699/VJ16.200
6. Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Systems Biology* 2015, 9(Suppl 2):S2.
7. Andrey S Glotov, Evgeny S Tiys, Elena S Vashukova, Vladimir S Pakin, Pavel S Demenkov, Olga V Saik, Timofey V Ivanisenko, Olga N Arzhanova, Elena V Mozhgovaya, Marina S Zainulina, Nikolay A Kolchanov, Vladislav S Baranov, Vladimir A Ivanisenko Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia asociome). *BMC Systems Biology* 2015, 9(Suppl 2):S4.
8. Larina IM, Pastushkova LKh, Tiys ES, Kireev KS, Kononikhin AS, Starodubtseva NL, Popov IA, Custaud MA, Dobrokhotov IV, Nikolaev EN, Kolchanov NA, Ivanisenko VA. Permanent proteins in the urine of healthy humans during the Mars-500 experiment. *J Bioinform Comput Biol*. 2015 Feb;13(1):1540001. doi: 10.1142/S0219720015400016.
9. Popik O.V., Saik O.V., Petrovskiy E.D., Sommer B., Hofestädt R., Lavrik I.N., Ivanisenko V.A. Analysis of signaling networks distributed over intracellular compartments based on protein-protein interactions. *BMC Genomics*, 2014, 15(Suppl 12):S7
10. Binder H., Wirth H., Arakelyan A., Lembcke K., Tiys E.S., Ivanishenko V.A., Kolchanov V.A., Kononikhin A., Popov I., Nikolaev E.N., Pastushkova L., Larina I.M. Time-course human urine proteomics in space-flight simulation experiments—A high resolution and personalized machine learning analysis. *BMC Genomics* 2014, 15(Suppl 12):S2
11. Larina IM, Ivanisenko VA, Nikolaev EN, Grigorev AI. The Proteome of a Healthy Human during Physical Activity under Extreme Conditions. *Acta Naturae*. 2014 Jul;6(3):66-75.
12. Bragina EY, Tiys ES, Freidin MB, Koneva LA, Demenkov PS, Ivanisenko VA, Kolchanov NA, Puzyrev VP. Insights into pathophysiology of dystrophy through the analysis of gene networks: An example of bronchial asthma and tuberculosis. *Immunogenetics*. 2014;66(7-8):457-65.
13. Pastushkova LK, Kireev KS, Kononikhin AS, Tiys ES, Popov IA, Dobrokhotov IV, Custaud M-, Ivanisenko VA, Kolchanov NA, Nikolaev EN, Pochuev VI, Larina IM. Permanent proteins in healthy human's urine in the experiment with 520-day isolation. *Aviakosmicheskaya i Ekologicheskaya Meditsina*. 2014;48(1):48-54.
14. Pastushkova LKh, Kireev KS, Kononikhin AS, Tiys ES, Popov IA, Starodubtseva NL, Dobrokhotov IV, Ivanisenko VA, Larina IM, Kolchanov NA, Nikolaev EN. Detection of renal tissue and urinary tract proteins in the human urine after space flight. *PLoS One*. 2013 Aug 13;8(8):e71652.