

Лекция 6

Реконструкция и анализ ассоциативных генных сетей

м.н.с. лаб. Компьютерной протеомики Сайк Ольга Владимировна

План лекции

1) Реконструкция генных сетей, основанная на методах автоматического анализа текстов (text-mining).

2) Система ANDSystem – как пример системы, использующей автоматический анализ текстов для реконструкции ассоциативных генных сетей.

- Источники информации.
- Онтологии, используемые в ANDSystem.
- Модуль автоматического анализа текстов в ANDSystem.
- Точность и полнота информации, извлеченной методом text-mining.
- Краткая характеристика интерфейса AndVisio, реализованного в ANDSystem – возможности поиска и фильтрации данных.
- Примеры задач, которые можно решать с помощью ANDSystem.

3) 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 Gene Ontology биологических процессов (ресурс DAVID).
- Приоритезация генов-кандидатов (ресурсы Endeavour, TopGene, DIR).

4) Примеры исследований, проведенных с использованием ANDSystem

Ассоциативные генные сети

Ассоциативные генные сети являются сложными, комплексными молекулярно-генетическими сетями, ассоциированными с какими либо биологическими процессами, фенотипическими признаками или заболеваниями. Типы базовых объектов, которыми оперируют ассоциативные генные сети, подразделяются на несколько классов, согласно их природе:

- (1) **молекулярно-генетические объекты**. К этому классу относятся такие типы объектов, как гены, РНК, белки, метаболиты, клеточные компоненты;
- (2) **биологические процессы и системы** (метаболические пути, пути передачи сигналов, транспортные пути и т.д.);
- (3) **фенотипические признаки**, поведенческие характеристики или функциональные состояния организма или генетических систем, включая заболевания;
- (4) **внутренние и внешние факторы**, воздействующие на систему (мутации, эпигенетический контроль, температура, давление, лекарства и другие химические соединения).

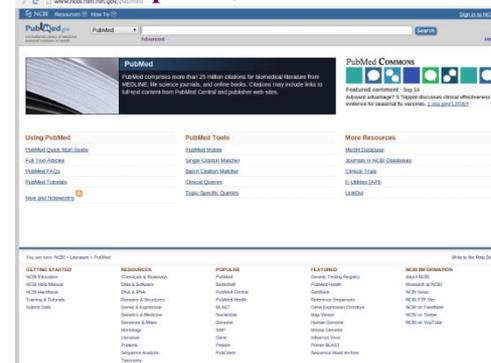
Таким образом, ассоциативные генные сети содержат существенно расширенный состав базовых объектов по сравнению с генными сетями, оперирующими только с молекулярно-генетическими объектами. Введение такого расширенного состава базовых объектов позволяет строить сети, в которых вершинами могут быть любые объекты из перечисленных классов. При этом, молекулярно-генетические объекты могут оказаться связанными с фенотипическими признаками или управляющими факторами. Такое свойство ассоциативных генных сетей 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.

Автоматический анализ текстов позволяет экстрагировать подобную информацию из научных публикаций.

Источники информации

- Важнейшим источником информации по биомедицинской тематике являются научные публикации. В системе **PubMed**, которая является крупнейшей базой данных научных статей, доступных через интернет, собрано более 26 миллионов публикаций. Ежегодно число статей в системе PubMed увеличивается, примерно, на 1 миллион.

Главная страница системы PubMed



- Базы данных. В журнале **Nucleic Acids Research (NAR) database issue** (импакт фактор 8,65) опубликованы статьи по более чем 1900 различным базам данных, содержащих информацию по биомедицинской тематике. Каждая база данных содержит структурированную и формализованную информацию.

Базы данных по участникам генных сетей



Medical Subject Headings (MeSH)

Медицинские предметные рубрики (Medical Subject Headings, сокращённо MeSH) — всеобъемлющий контролируемый словарь, индексирующий журнальные статьи и книги по естественным наукам; может также служить в качестве тезауруса, облегчающего поиск информации. Создан и обновляется Национальной медицинской библиотекой США, используется в базах статей Medline и PubMed.



Дескрипторы (предметные рубрики) сгруппированы в иерархическом порядке. В PubMed каждая журнальная статья индексируется по 10-15 рубрикам или подрубрикам, один или два из которых обозначены как «основные» и помечены звёздочкой. При выполнении поиска через PubMed, входящий термин автоматически переводится на соответствующий дескриптор. Помимо этого, по умолчанию в поиск будут включены все дескрипторы, расположенные ниже него в иерархии.

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]

Система MeSHOPs

Система MeSHOPs доступна по адресу <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 четырехпольная таблица сопряженности 2 × 2:

- 1) Частота встречаемости термина t_i во множестве статей по заданной тематике;
- 2) Все множество статей по заданной тематике за исключением статей, содержащих упоминания термина t_i ;
- 3) Частота встречаемости термина t_i во множестве статей, не затрагивающих заданную тематику;
- 4) Оставшиеся статьи – в которых не встречается термина t_i , и, которые не затрагивают заданную тематику.

Всего в анализе в системе MeSHOPs было использовано 18 млн. публикаций статей из системы MEDLINE.

2 Find Genes Directly Linked to Disease Keywords



	A2M articles	Rest of PubMed (with Disease terms)	Total
Articles with <i>Alzheimer Disease</i>	8	42 112	42 120
Articles without <i>Alzheimer Disease</i>	23	8 743 558	8 743 581
Total	31	8 785 670	8 785 701

- Fisher's Exact Test p-value: 1.97E-12
- 221 170 gene-disease associations examined
- Bonferroni Corrected p-value: 4.36E-07

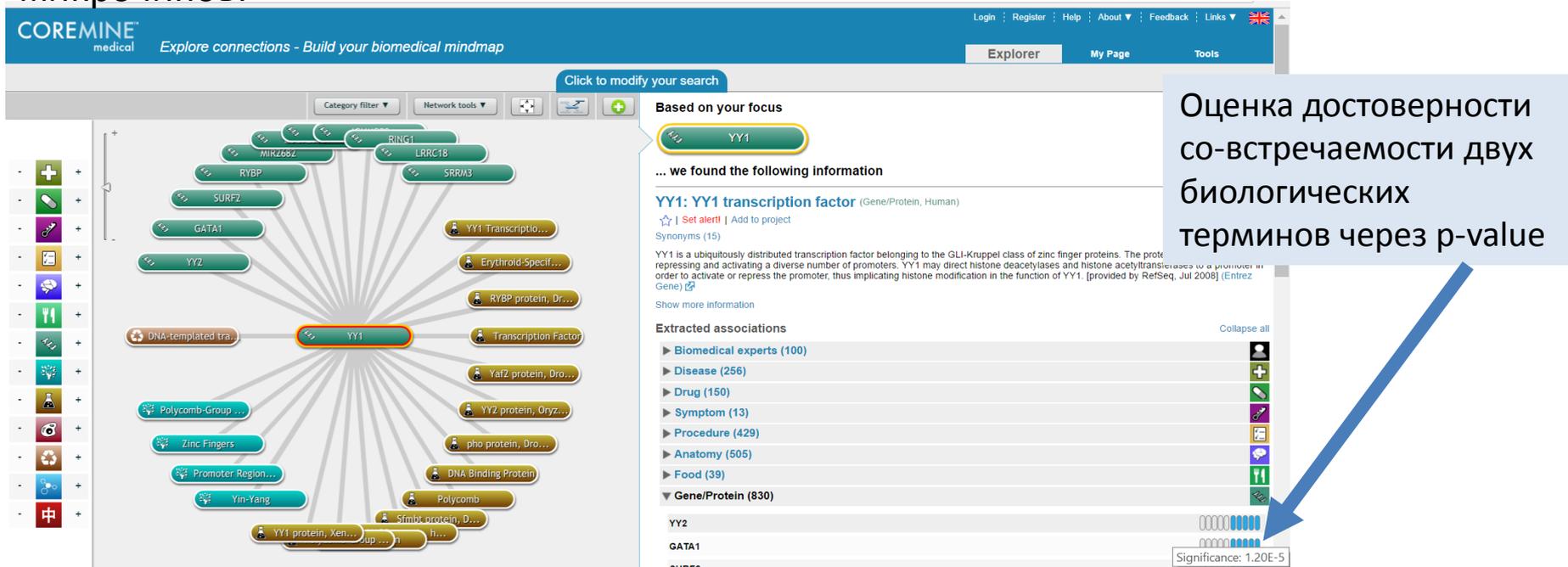
3 Generate a Keyword Profile for Each Gene and Disease

MeSH term profile for the A2M Gene

Acute-Phase Proteins Alpha-Globulins
 Alzheimer Disease Apolipoproteins E Apolipoproteins Apoproteins
 Blood Proteins Brain Diseases Delirium, Dementia, Amnesic, Cognitive Disorders Dementia Endopeptidases
 Galanthus Globulins LDL-Receptor Related Protein 1 LDL-Receptor Related Proteins Macroglobulins Mental Disorders
 Mesna Neurodegenerative Diseases Peptide Hydrolases Proteinase Inhibitory Proteins, Secretory Receptors, LDL Receptors, Lipoprotein Serum Globulins Tauopathies alpha-Macroglobulins

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

Система CoreMine доступна по адресу <https://www.coremine.com/medical/#search>. Авторы провели автоматический анализ открытых текстов 10 млн. научных публикаций (абстрактов) из MEDLINE для создания глобальной сети со-встречаемости 13 712 генов человека. Гены были аннотированы путем ассоциации с терминами из MeSH и генной 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/>

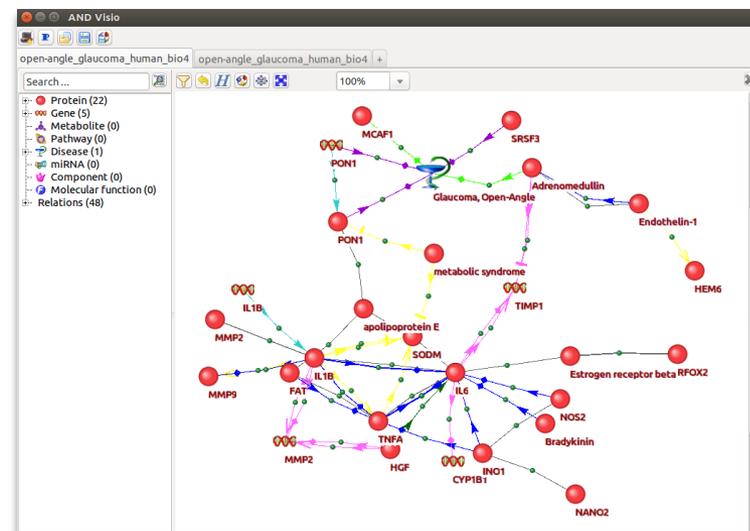
ANDSystem включает:

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

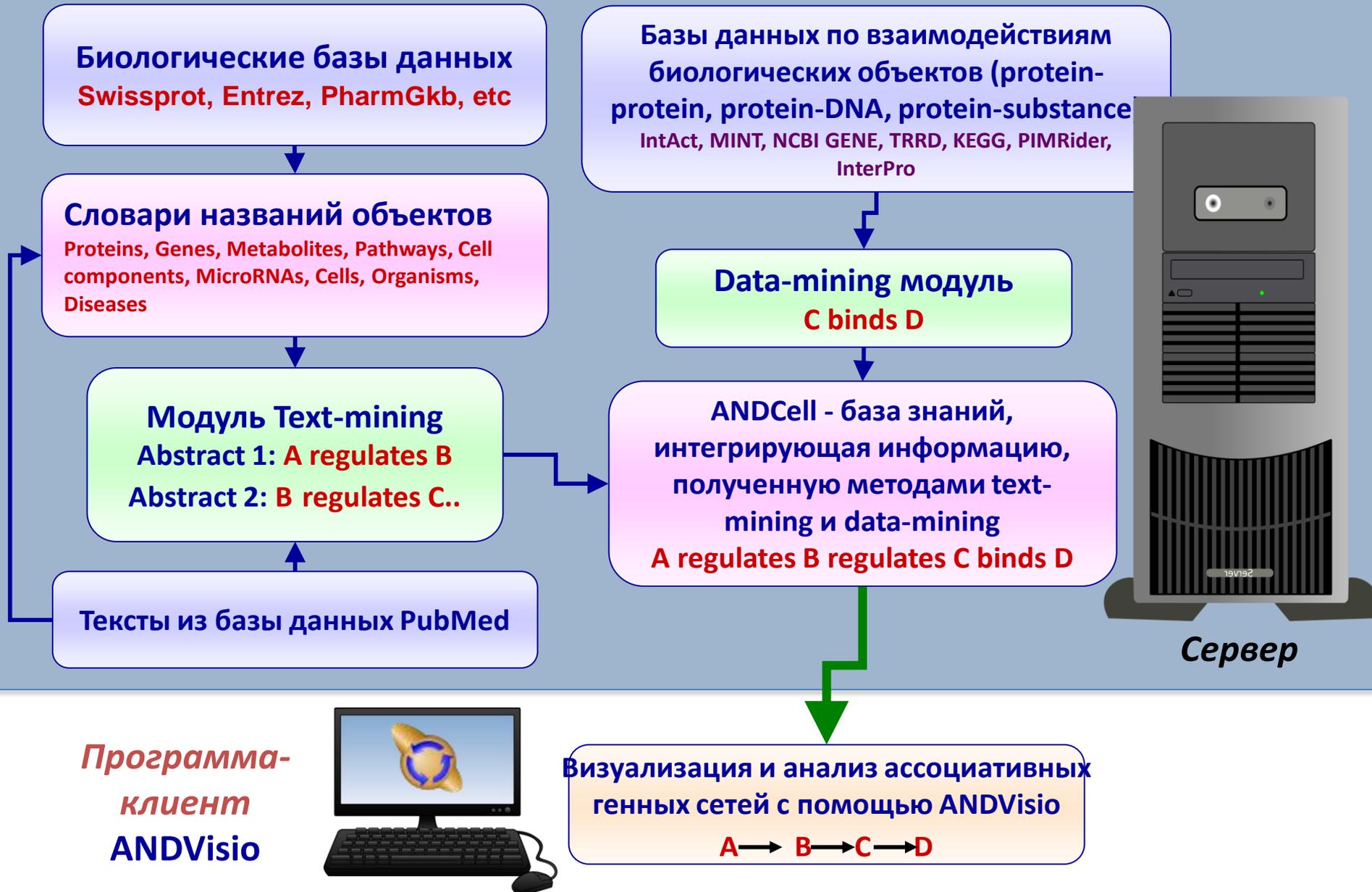
ANDVisio обеспечивает доступ к базе знаний, анализ и визуализацию результатов запроса.

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
 - PMID1986
 - PMID1990
 - PMID1991
 - PMID1992
 - PMID1993
 - PMID1994
 - PMID1995
 - PMID1996
 - PMID1997
 - PMID1998
 - PMID1999
 - PMID2000
 - PMID2001
 - PMID2002
 - PMID2003
 - PMID2004
 - PMID2005
 - PMID2006
 - PMID2007
 - PMID2008
 - PMID2009
 - PMID2010
 - PMID2011
 - PMID2012
 - PMID2013
 - PMID2014
 - PMID2015
 - PMID2016
 - PMID2017
 - 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

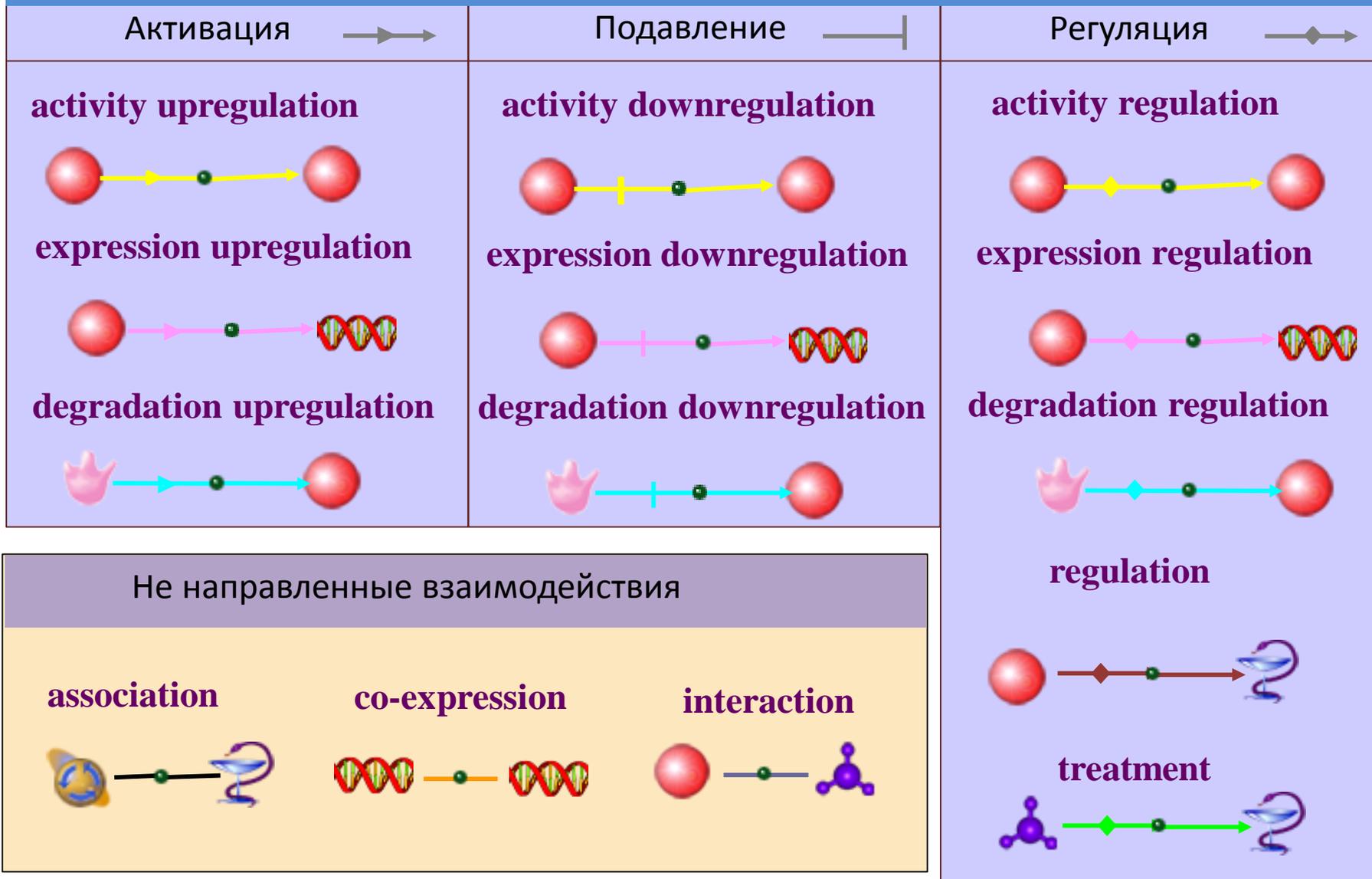
- afcs
- BioGrid
- CHEBI
- DIP
- emdb
- GO
- HPRD
- imex
- IntAct
- InterPro
- IntEnz
- matrixdb
- mpidb
- MINT
- psi-mi
- rcsb pdb
- SGD
- SWISSPROT_AC
- wwpdb

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



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		

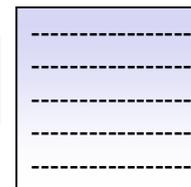
Модуль автоматического анализа текста (text-mining) в системе ANDSystem.

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

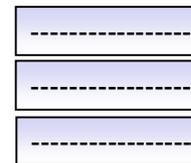


Документы

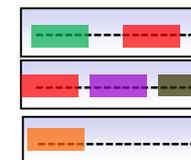
Перевод в стандартизированный текст



Разбивка текста на предложения

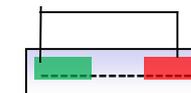


Распознавание объектов в предложениях



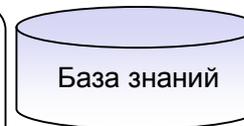
Словари названий объектов

Извлечение взаимоотношений между объектами

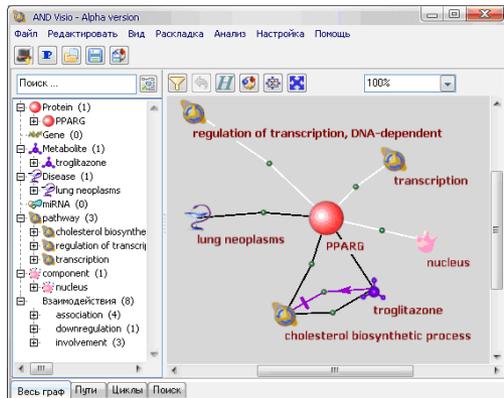


Лингвистические шаблоны

Хранение и визуализация извлеченных взаимоотношений



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: 16138112]

Similarly **MnSOD** specific **activity was increased by TNF** (290% increase) and the amosite-TNF combination (313% increase) but not by amosite alone. [PubMed: 8817862]

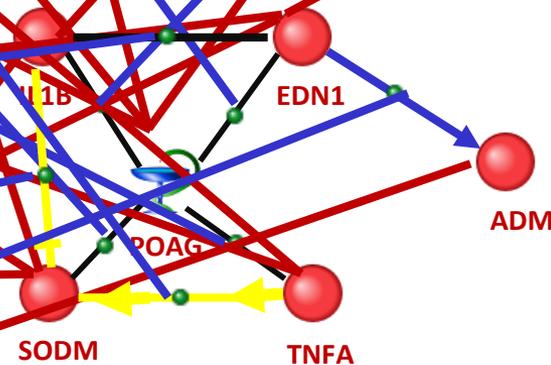
Incubation with **IL-1beta** **increased** concentrations of **ET-1** and pPET-1 relative gene expression. [PubMed: 18588356]

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]

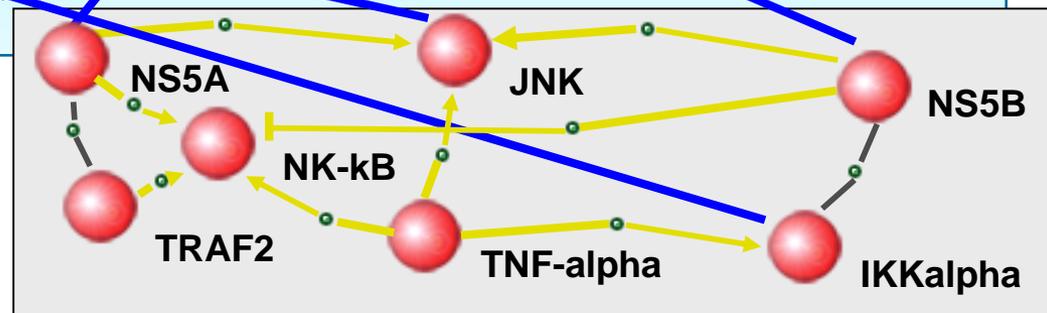
Analysis of aqueous humor from patients with primary **open-angle glaucoma (POAG)** revealed marked **increases** in the content of **endothelin-1 (ET-1)** and transforming growth factor beta (TGF-). [PubMed: 22736605]



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

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

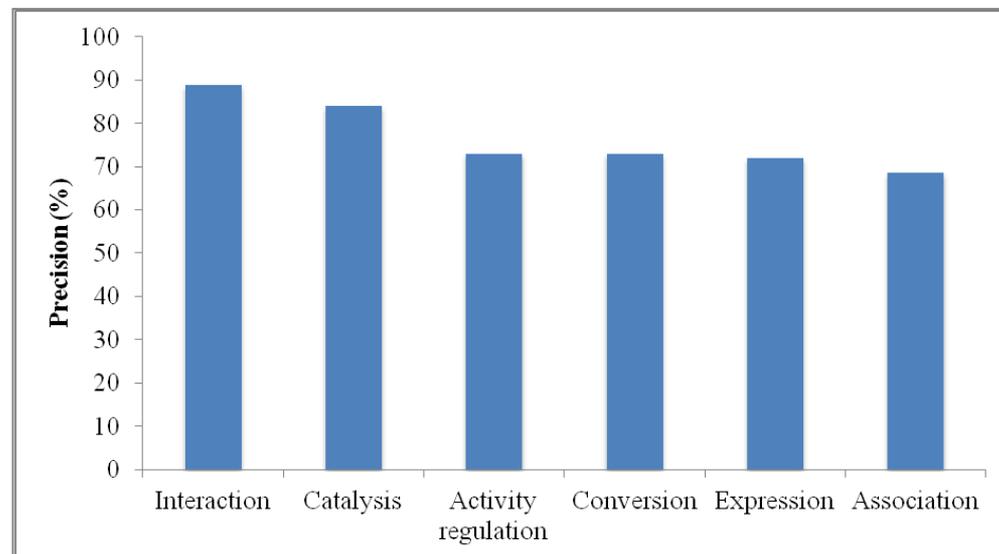
N_{correct} – число верно распознанных взаимодействий.

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

Точность 6 основных типов взаимодействий, выявленных системой ANDSystem.

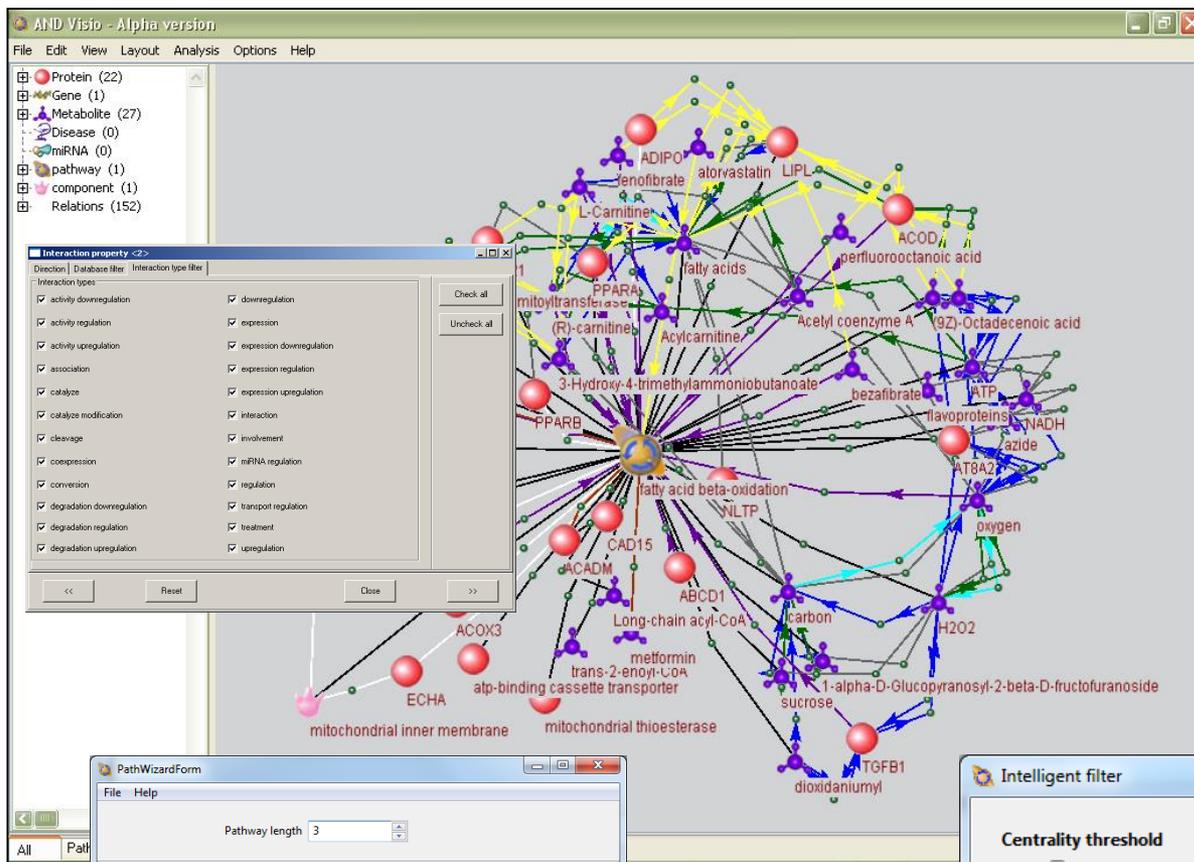
Для взаимодействий типа “interaction” наблюдалась максимальная точность (88.8%), минимальная -- для типа “association” (68.6%).

В среднем точность составила 76.5%.

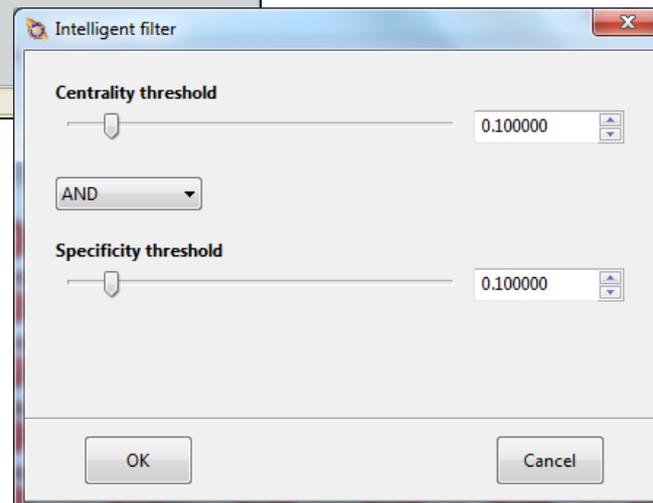
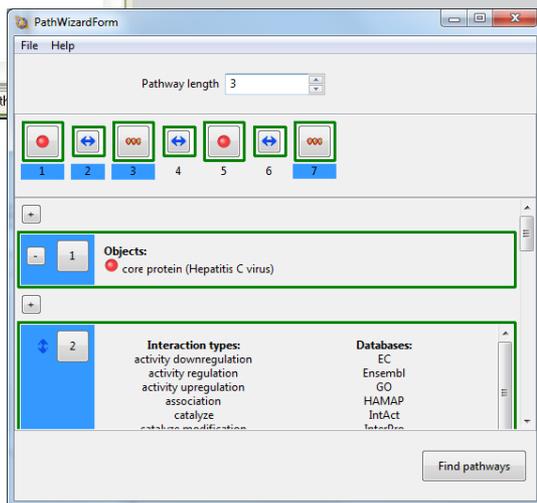


Полнота составила около 54% при оценке на Золотом стандарте, содержащем информацию из базы данных GeneNet, собранную вручную экспертами, о различных типах молекулярно-генетических взаимодействий.

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.

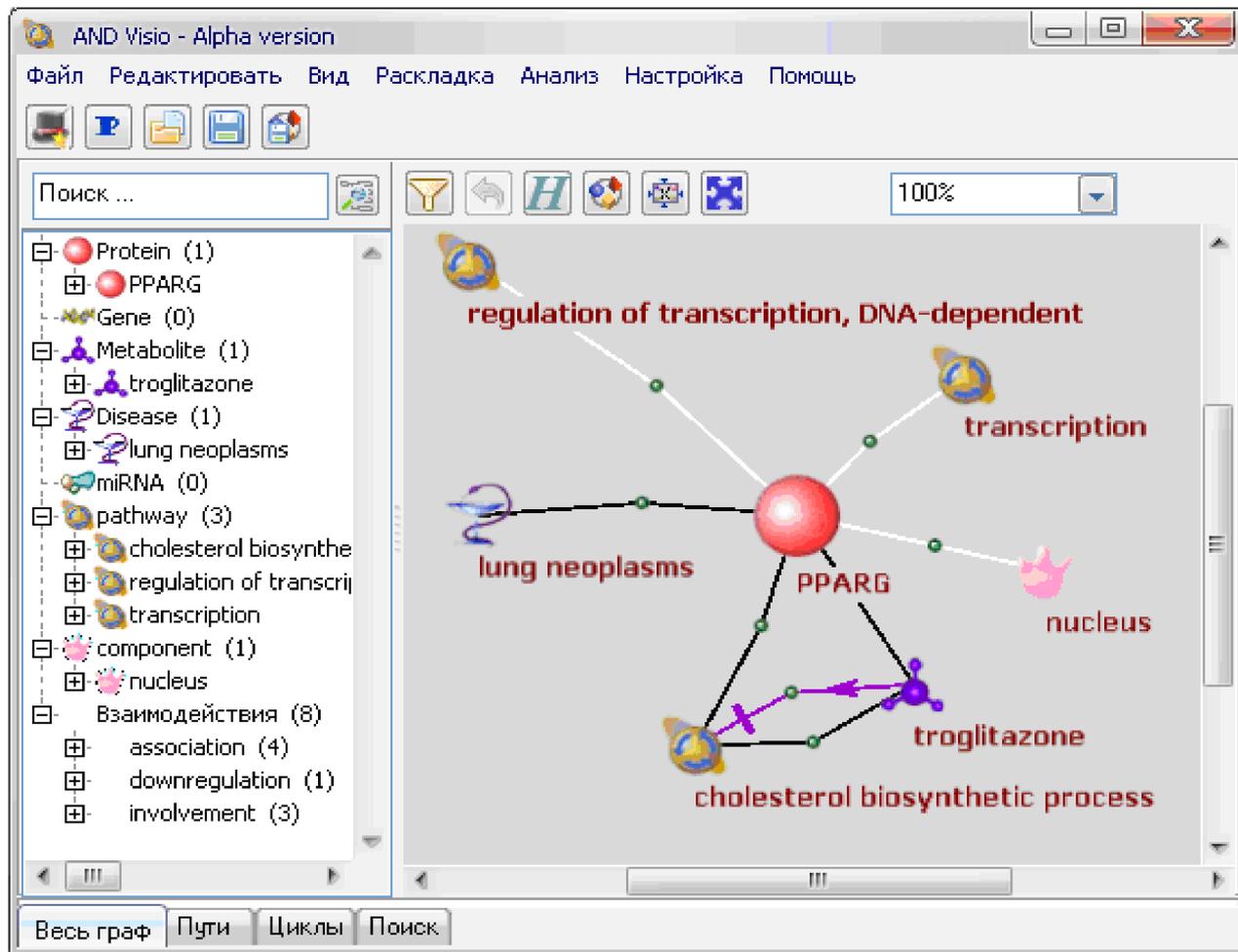


Программа для визуализации ассоциативных сетей ANDVisio предназначена для выполнения следующих основных функций:

- 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

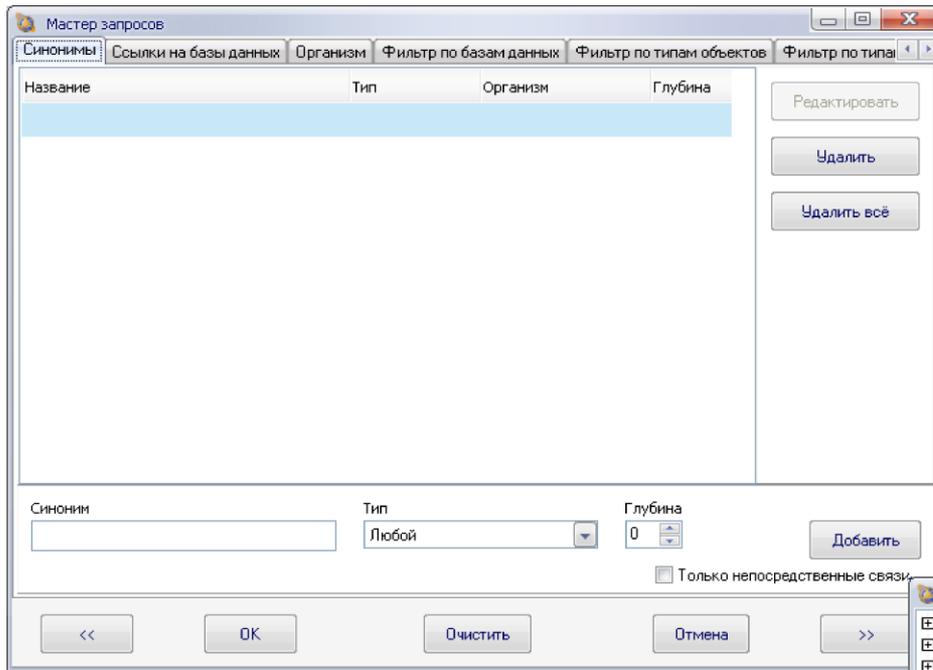
Главное окно программы

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.

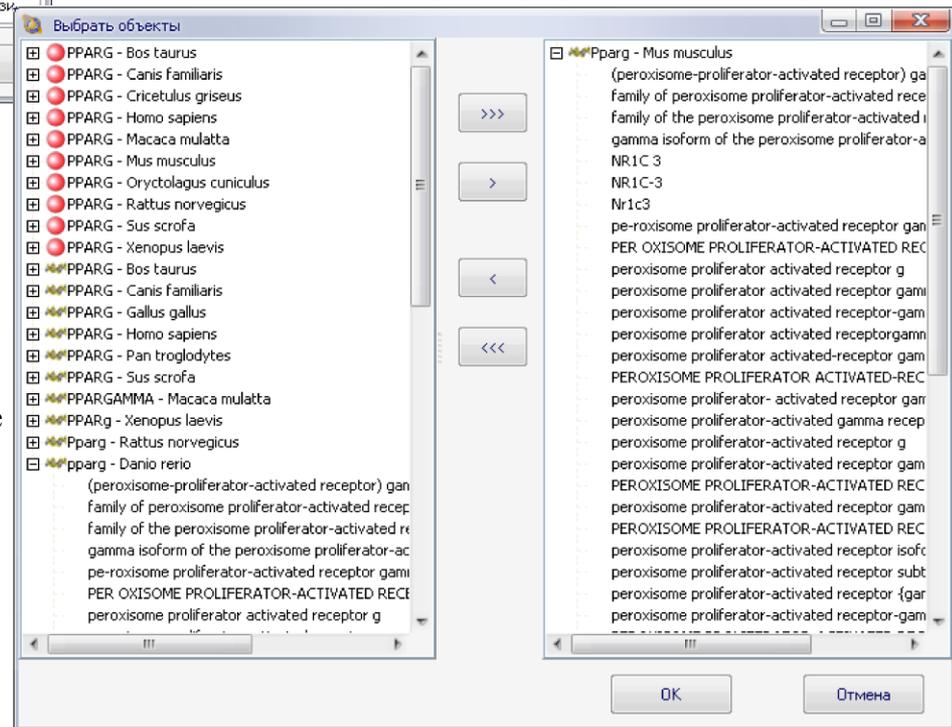
Окно «Мастера запросов»



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».

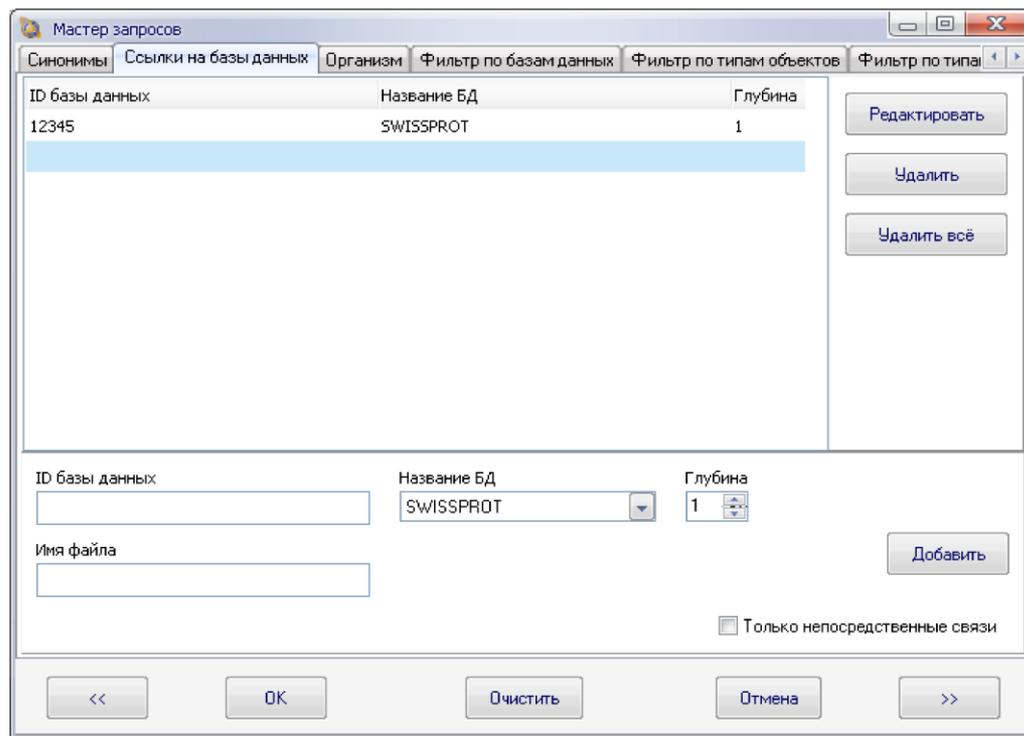
На вкладке "Синонимы" в нижней части формы пользователь может ввести шаблон названия объекта и выбрать тип объекта. После нажатия кнопки "Добавить" производится поиск объектов в базе данных, имеющих совпадающие с указанным шаблоном названия. В случае если обнаружено более 1 объекта, удовлетворяющего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

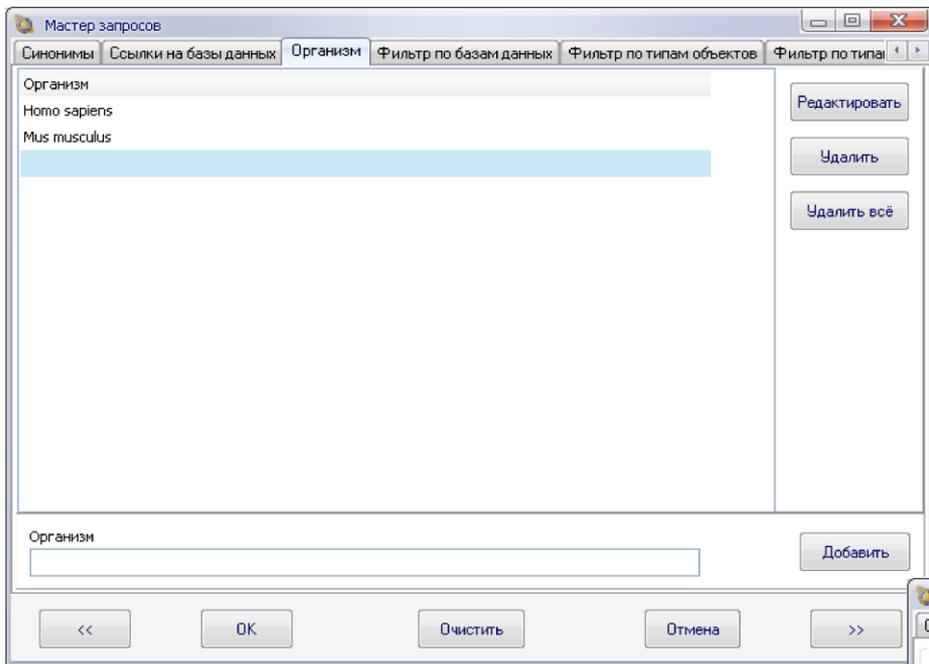


Если название базы данных не известно, то можно попробовать поискать эти идентификаторы по всем доступным базам данных, но идентификаторы многих баз данных пересекаются между собой и соответствуют совершенно разным объектам. После нажатия на кнопку "Добавить" идентификаторы добавятся в список интересующих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

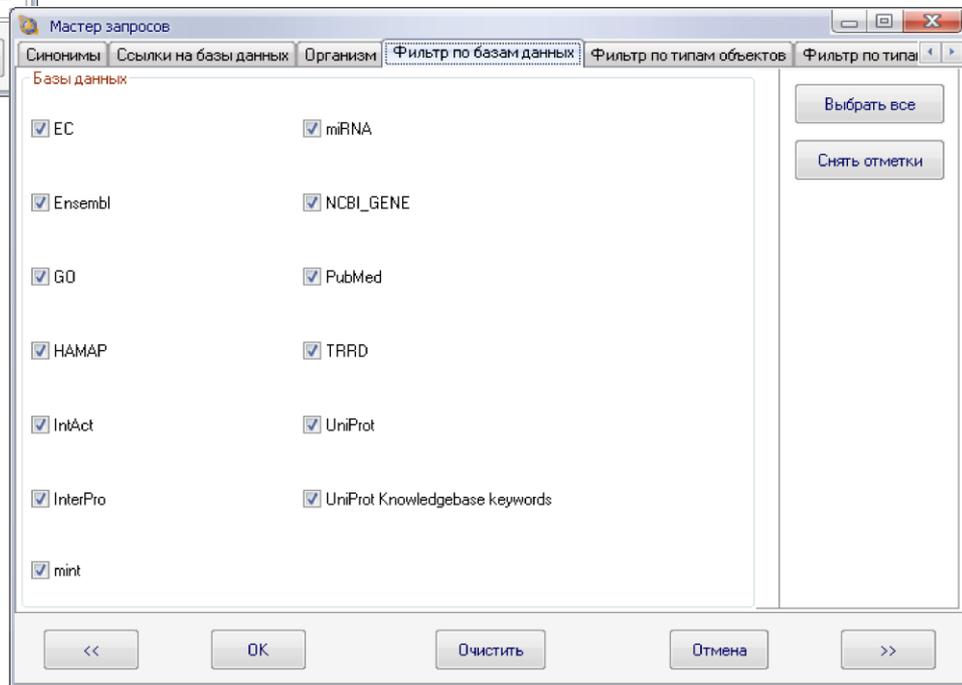
Окно «Мастера запросов» для введения фильтра по организмам



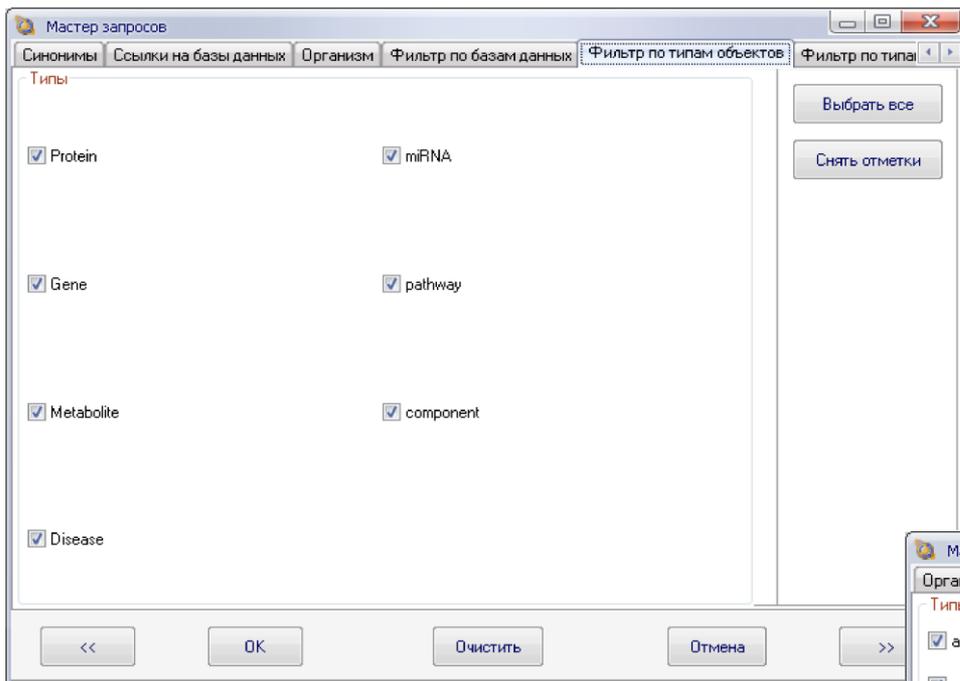
На вкладке "Организм" пользователю предлагается указать список интересующих организмов. В нижней части вкладки в поле "Организм" требуется ввести шаблон названия организма и нажать на кнопку "Добавить". В случае если обнаружено более 1 названия организма, удовлетворяющего условиям, то открывается форма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



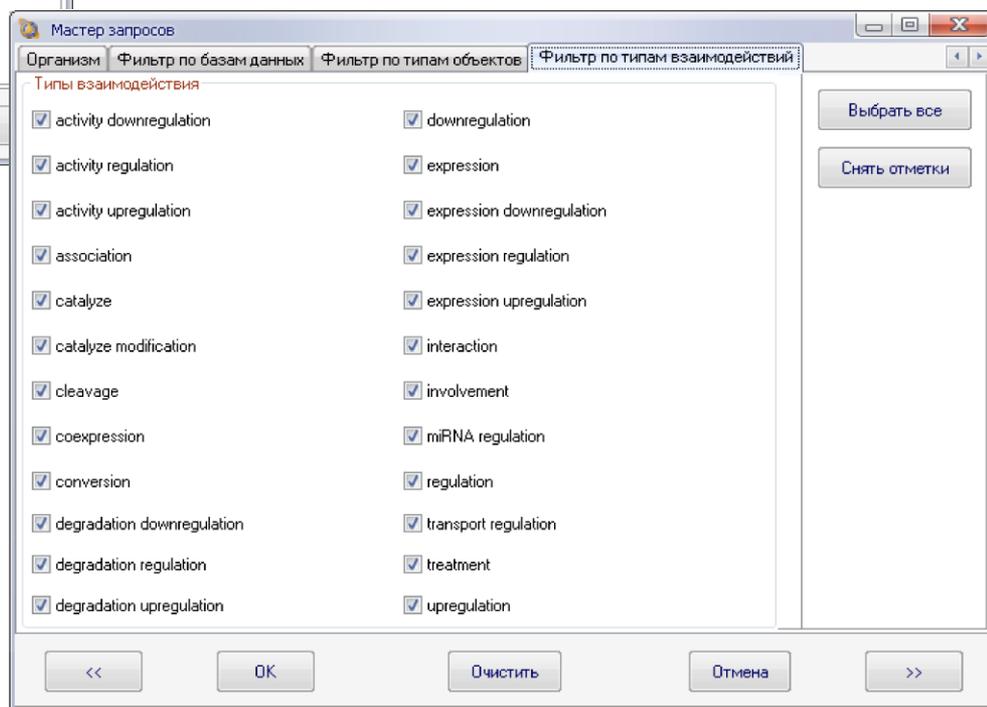
Окно «Мастера запросов» для заполнения фильтра по типам объектов



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.

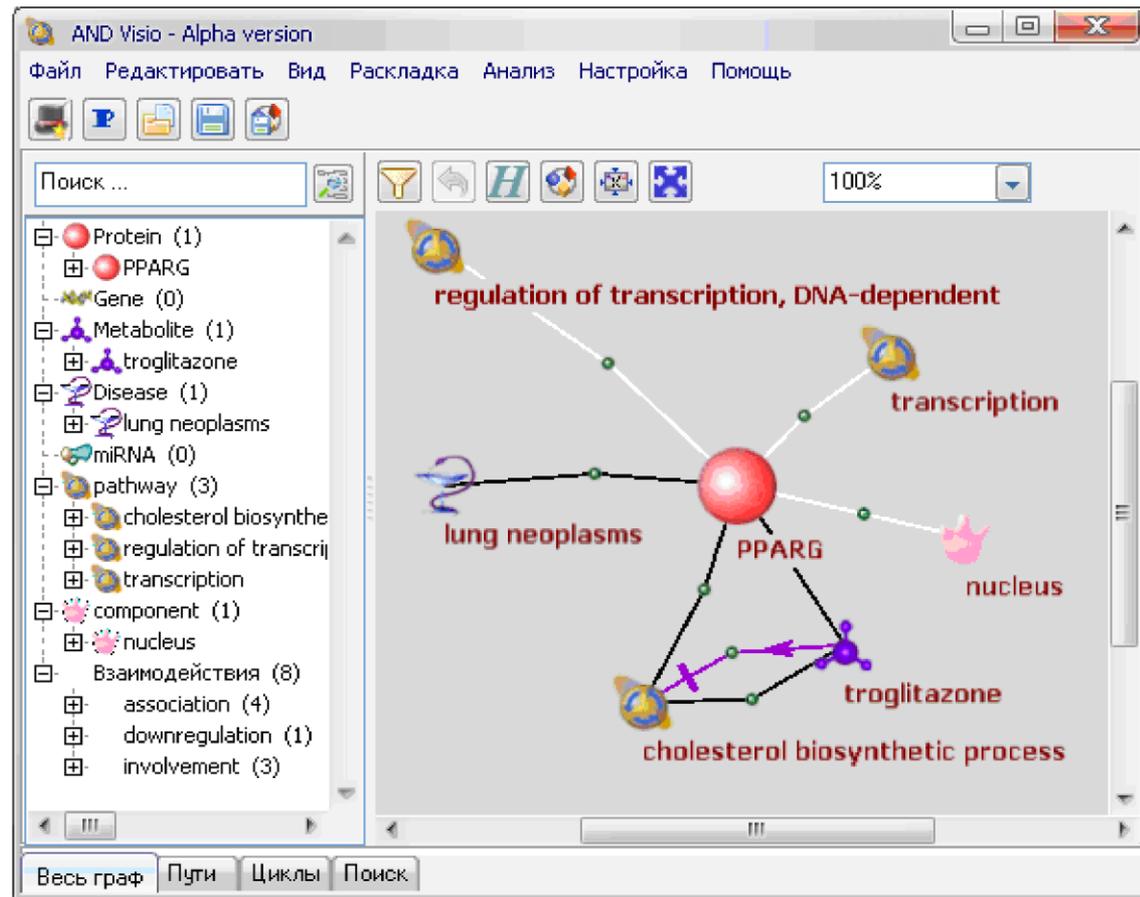
Щелчок левой кнопкой мыши на объекте производит его выделение, при этом подсвечиваются все его связи красным цветом.

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

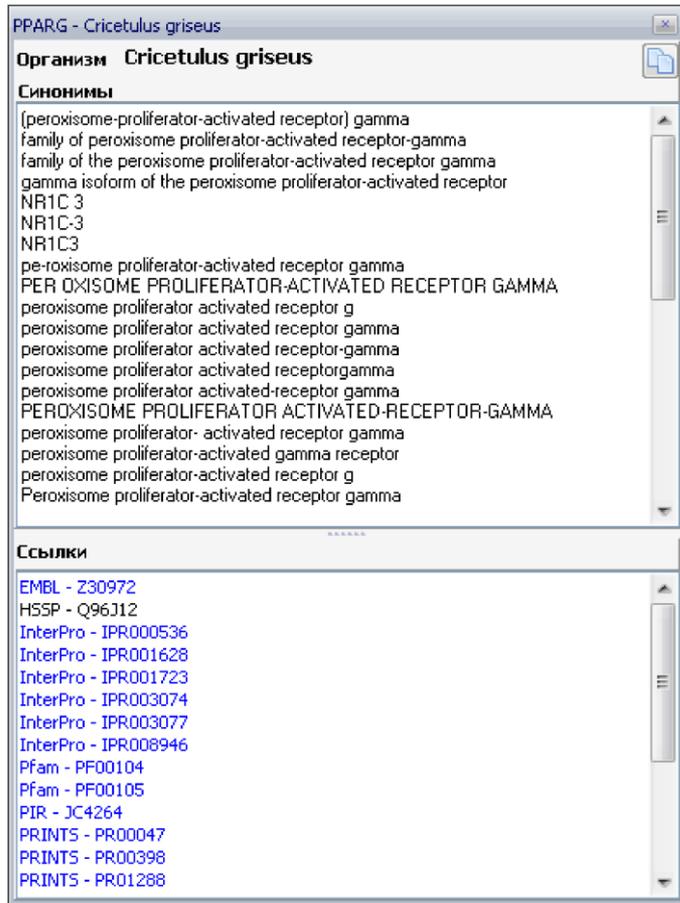
Чтобы уменьшить или растянуть сеть по размеру окна программы следует использовать пункт меню *Вид -> Вписать в окно*.

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

Главное окно программы. Построена сеть первого уровня для белка PPARG



Окно свойств объекта



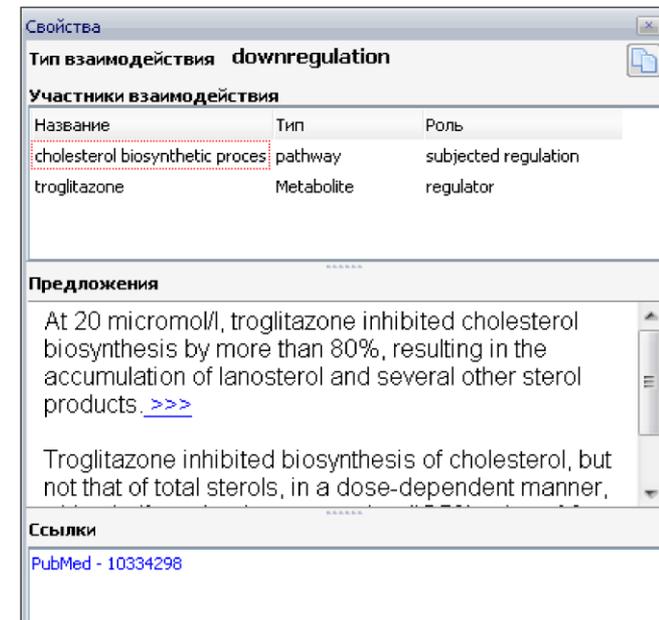
В верхнем поле окна свойств объектов указаны синонимы данного объекта, в нижнем поле --- ссылки на идентификаторы этого объекта в различных базах данных. Щелчок на ссылке позволяет через web-браузер получить доступ к карточке интересующего объекта в выбранной базе данных.

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 -> Переразложить 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 -> 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* 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.

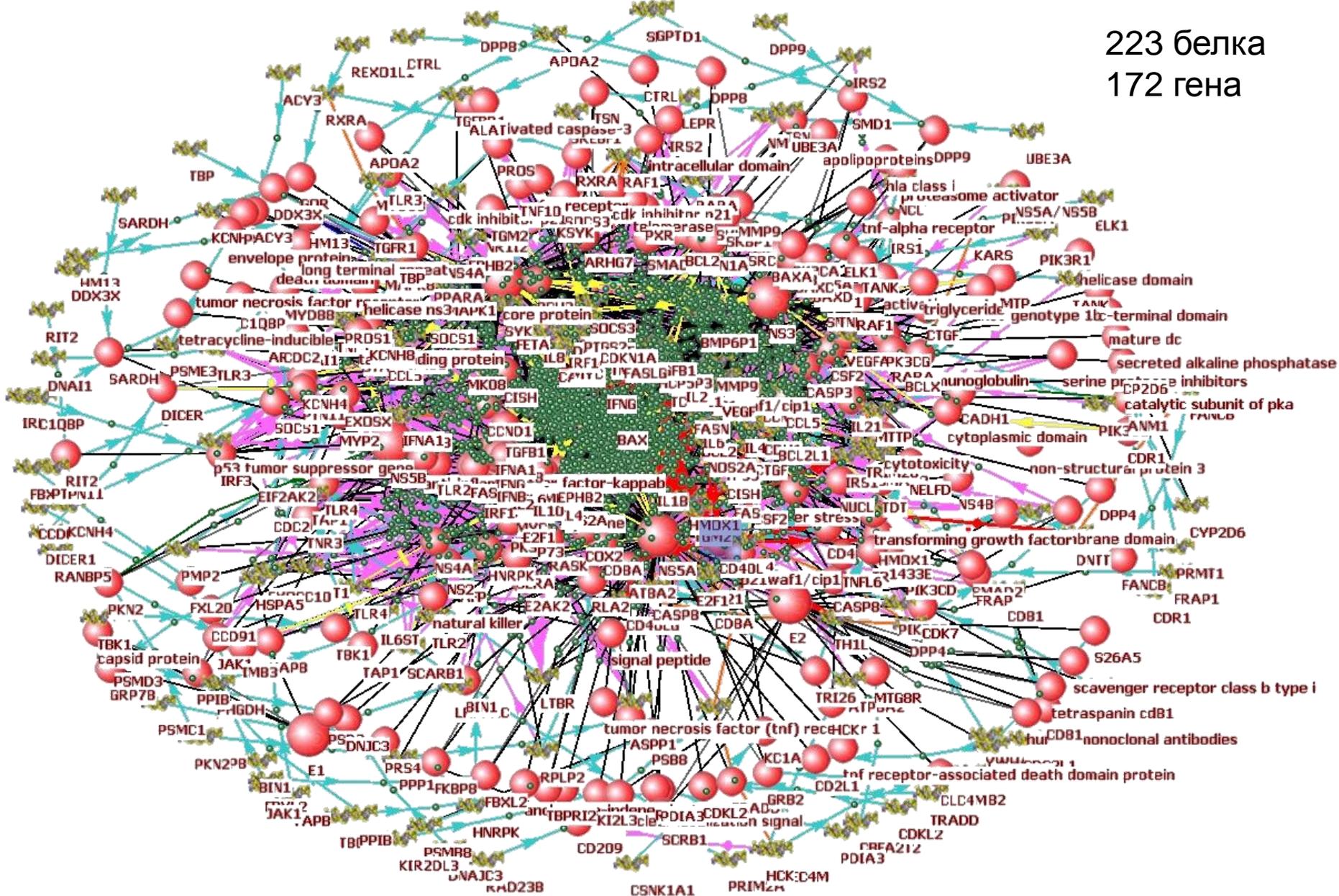
Чтобы расширить сеть новыми объектами, связанными с одним или несколькими объектами из реконструированной сети, выделите эти объекты и выберите пункт *Расширить* в контекстном меню, которое открывается по клику правой кнопкой мыши. Выберите объекты и укажите свойства фильтров как описано в разделе (*Составление 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:

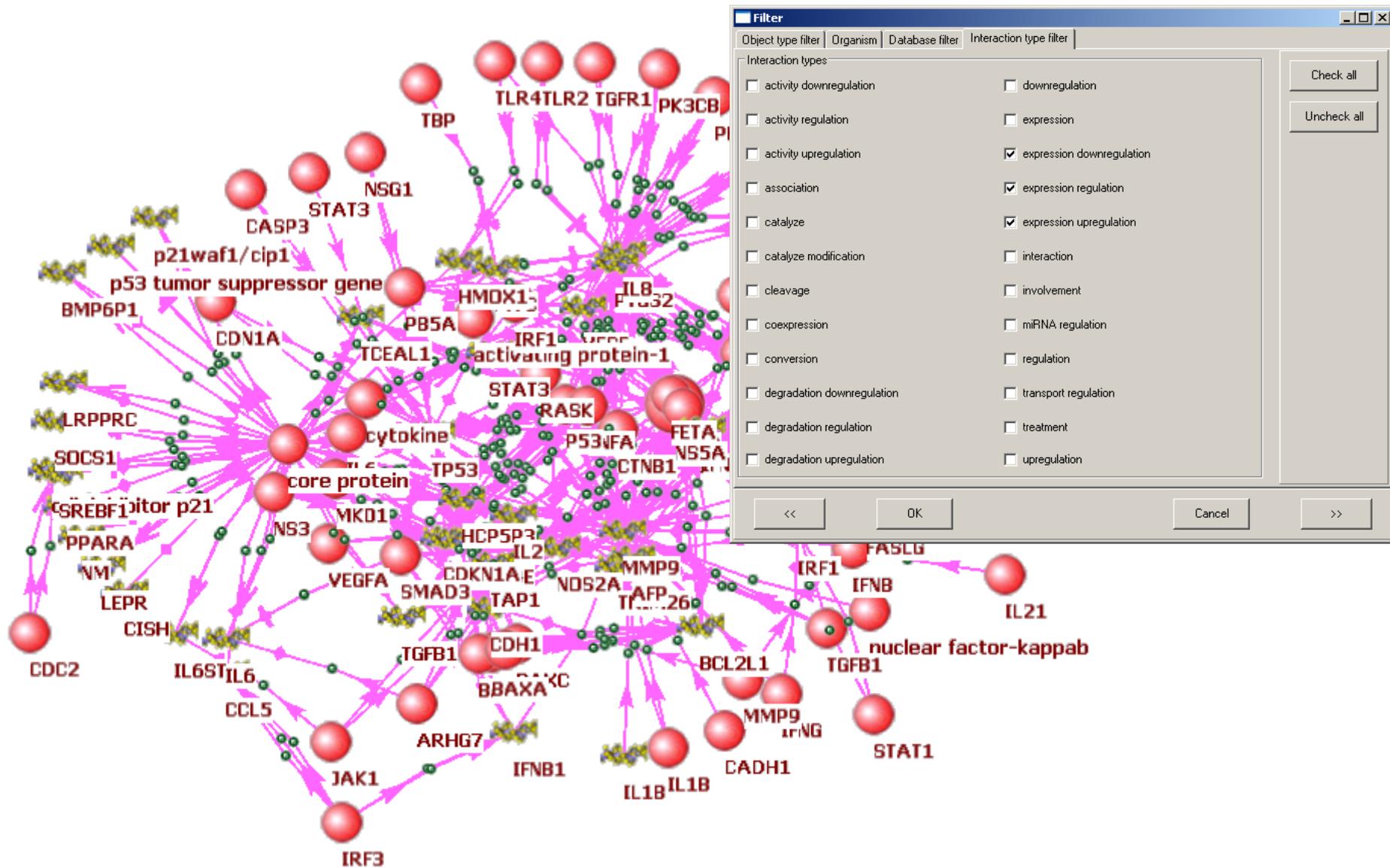
- Реконструкция и анализ ассоциативных генных сетей
- Интерпретация и визуализация экспериментальных данных
- Анализ межорганизменных взаимодействий, в том числе типа «патоген-хозяин»
- Выявление коморбидных взаимоотношений между заболеваниями; реконструкция сетей «заболевание-заболевание»
- Выявление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

Пример ассоциативной генной сети. Сеть белков и генов человека, ассоциированных с хронической инфекцией гепатитом С.

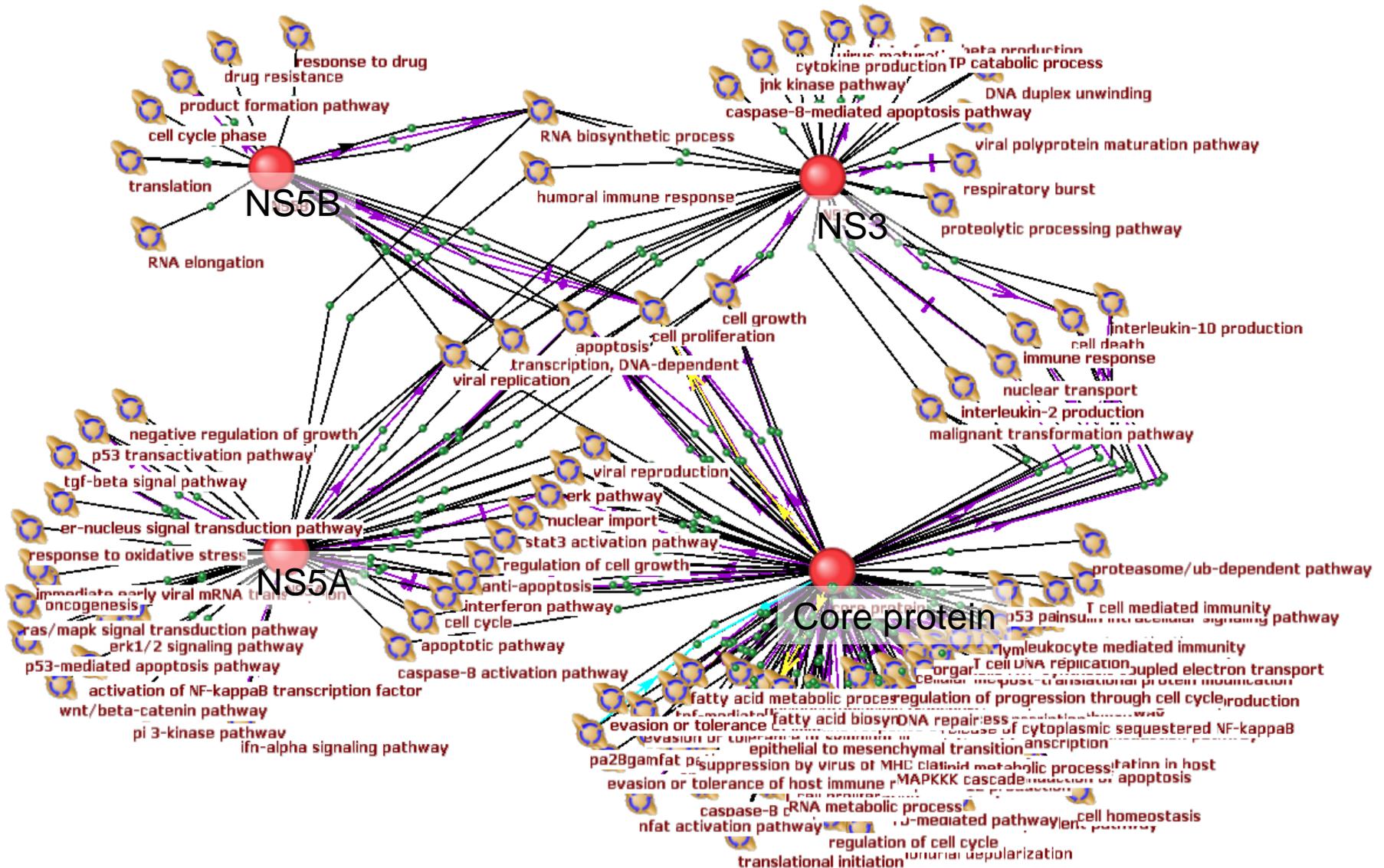
223 белка
172 гена



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

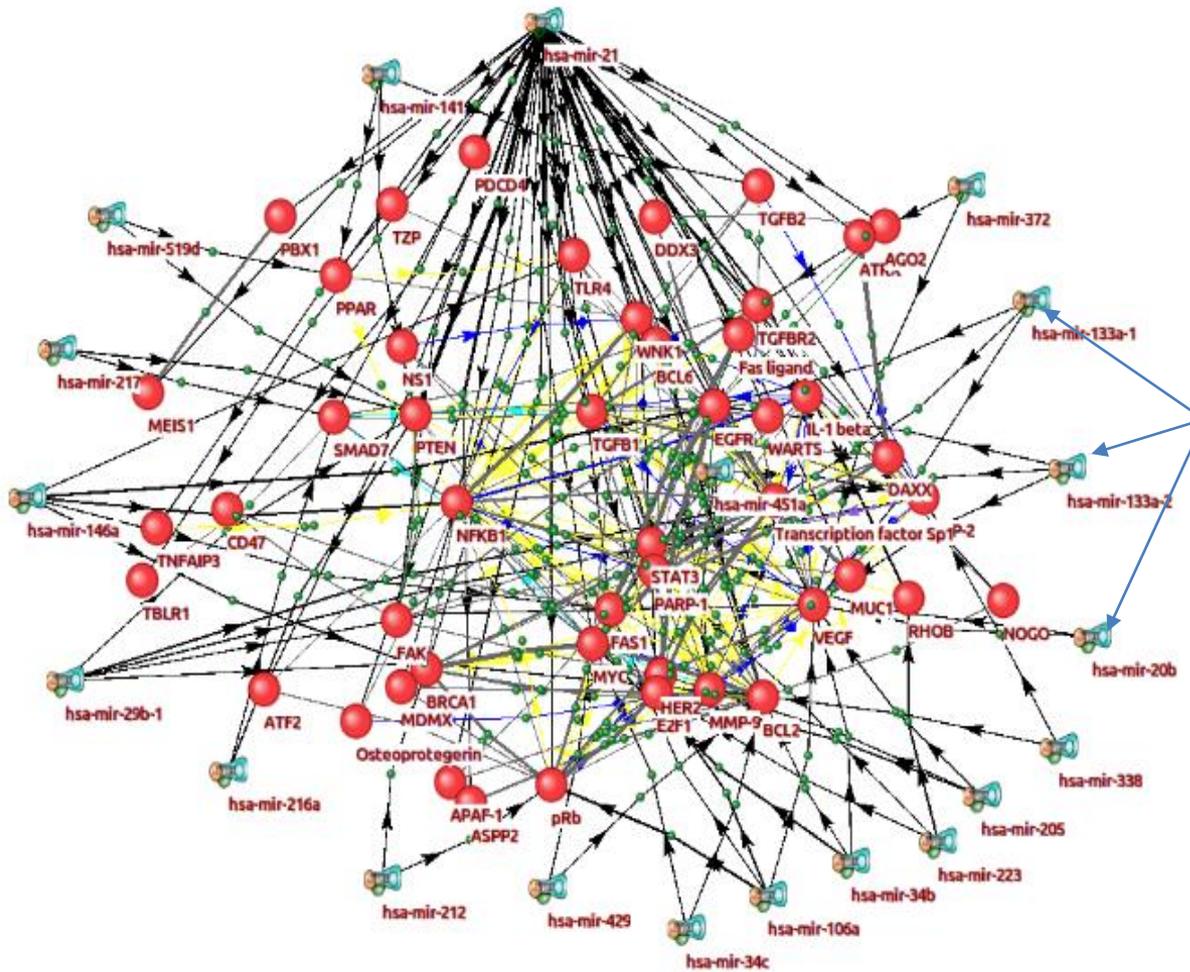


Взаимодействия биологических процессов человека, ассоциированных с воздействием белков вируса гепатита С



Ассоциативная сеть, описывающая регуляцию апоптоза микроРНК-21

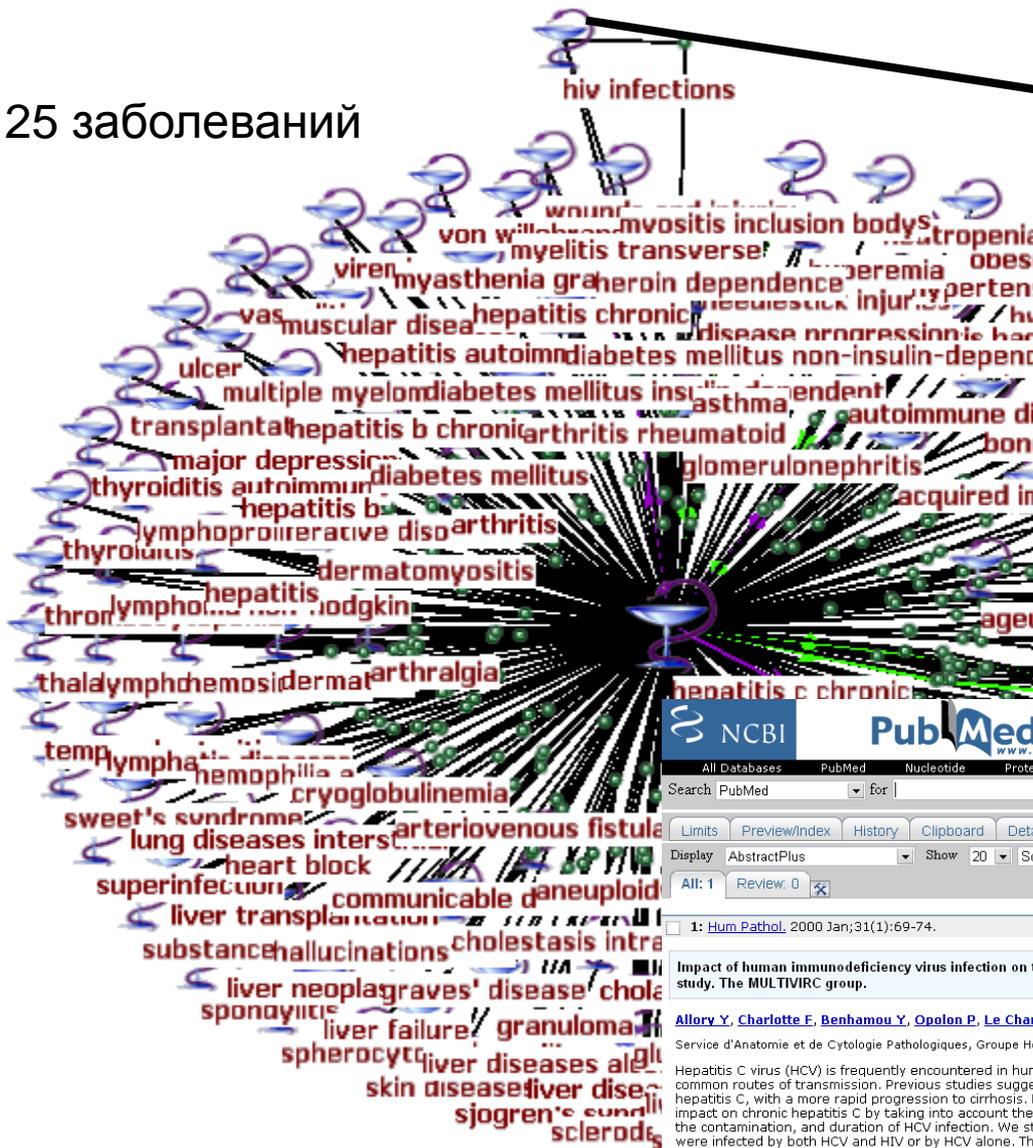
Одной из первых микроРНК, для которой была показана функция в патогенезе рака, стала miRNA21, связанная с глиобластомой и астроцитомой.



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

Сеть связей заболеваний с хроническим гепатитом С

125 заболеваний



Properties

Interaction type: association

Participants

Name	Type	Role
hepatitis c chronic	Disease	undetermined
hiv infections	Disease	undetermined

Sentences

CONCLUSIONS: a) HIV infection seems to adversely affect the clinical course of chronic hepatitis C, b) injected drug use, bi/homosexual practice and genotype-2/3 were more frequently encountered in co-infected patients, c) immunization against HBV should be encouraged in these patients. >>>

Does HIV-infection influence the response of chronic hepatitis C to interferon treatment? >>>

HIV infection has a negative impact on the natural history of chronic hepatitis C, accelerating the progression of liver fibrosis. >>>

Links

- PubMed - 10665916
- PubMed - 10863316
- PubMed - 10898321
- PubMed - 11600830
- PubMed - 12175272
- PubMed - 1348079
- PubMed - 15853988
- PubMed - 16177654
- PubMed - 16381246
- PubMed - 16767310

NCBI PubMed

All Databases PubMed Nucleotide Protein

Search PubMed for

Limits Preview/Index History Clipboard Detail

Display AbstractPlus Show 20 Sort

All: 1 Review: 0

1: Hum Pathol, 2000 Jan;31(1):69-74.

Impact of human immunodeficiency virus infection on the histological features of chronic hepatitis C: a case-control study. The MULTIVIRC group.

Allory Y, Charlotte F, Benhamou Y, Opolon P, Le Charpentier Y, Pognard T.

Service d'Anatomie et de Cytologie Pathologiques, Groupe Hospitalier Pitié-Salpêtrière, Paris, France.

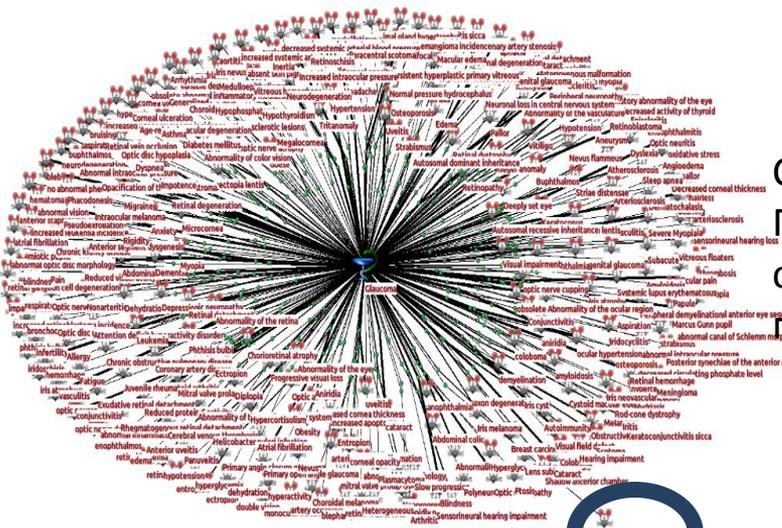
Hepatitis C virus (HCV) is frequently encountered in human immunodeficiency virus (HIV)-infected patients because of common routes of transmission. Previous studies suggested that HIV infection impaired the natural course of chronic hepatitis C, with a more rapid progression to cirrhosis. However, these studies did not assess the HIV infection impact on chronic hepatitis C by taking into account the risk factors for liver fibrosis progression: alcohol, sex, age at the contamination, and duration of HCV infection. We studied liver biopsy specimens of 2 groups of 58 patients that were infected by both HCV and HIV or by HCV alone. The 2 groups were matched according to those risk factors, and liver biopsy responses were evaluated with the METAVIR items. The METAVIR activity was higher in HIV-positive than HIV-negative patients. Cirrhosis was more frequent: (1) in HIV-positive patients with CD4 < or = 200 cells/microL (45%) than in HIV-negative patients (10%) (P = .003), (2) in HIV-positive patients with CD4 < or = 200 cells/microL (45%) than in HIV-positive patients with CD4 > 200 cells/microL (17%) (P = .04). These differences, which were linked to HIV status, might be related to the enhanced HCV replication during HIV infection or other immune mechanisms that need further studies.

PMID: 10665916 [PubMed - indexed for MEDLINE]

- Related Articles**
- Liver fibrosis progression in human immunodeficiency virus-infected patients. The Multivirc
 - Impact of early-unreated HIV infection on liver fibrosis progression: a case-control study.
 - The influence of human immunodeficiency virus infection on liver fibrosis progression: a long-term
 - Progression of chronic hepatitis C in HIV-infected patients: a case-control study.
 - Influence of hepatitis C virus genotype on liver fibrosis progression in chronic hepatitis C. The Hepatitis

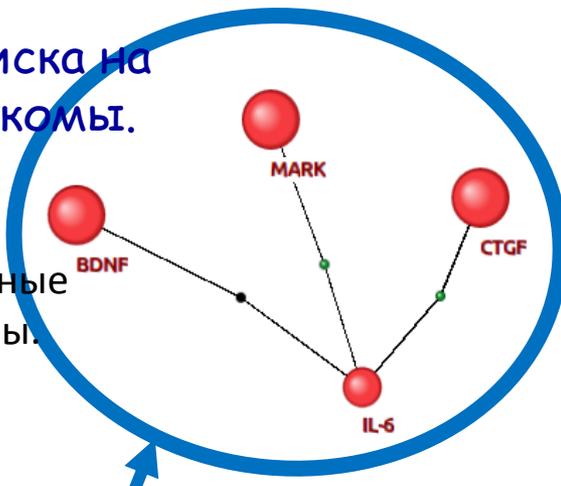
Комбинированные молекулярно-генетические и физиологические сети. Поиск генов-новых факторов риска на основе анализа ассоциом фенотипов заболевания глаукомы.

Ассоциом фенотипов, наблюдаемых при глаукоме

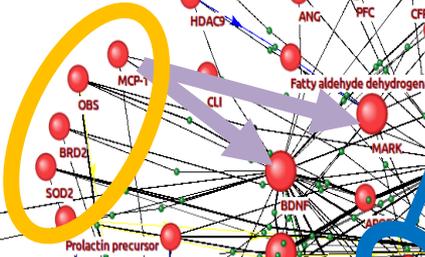


Макулярная дегенерация

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



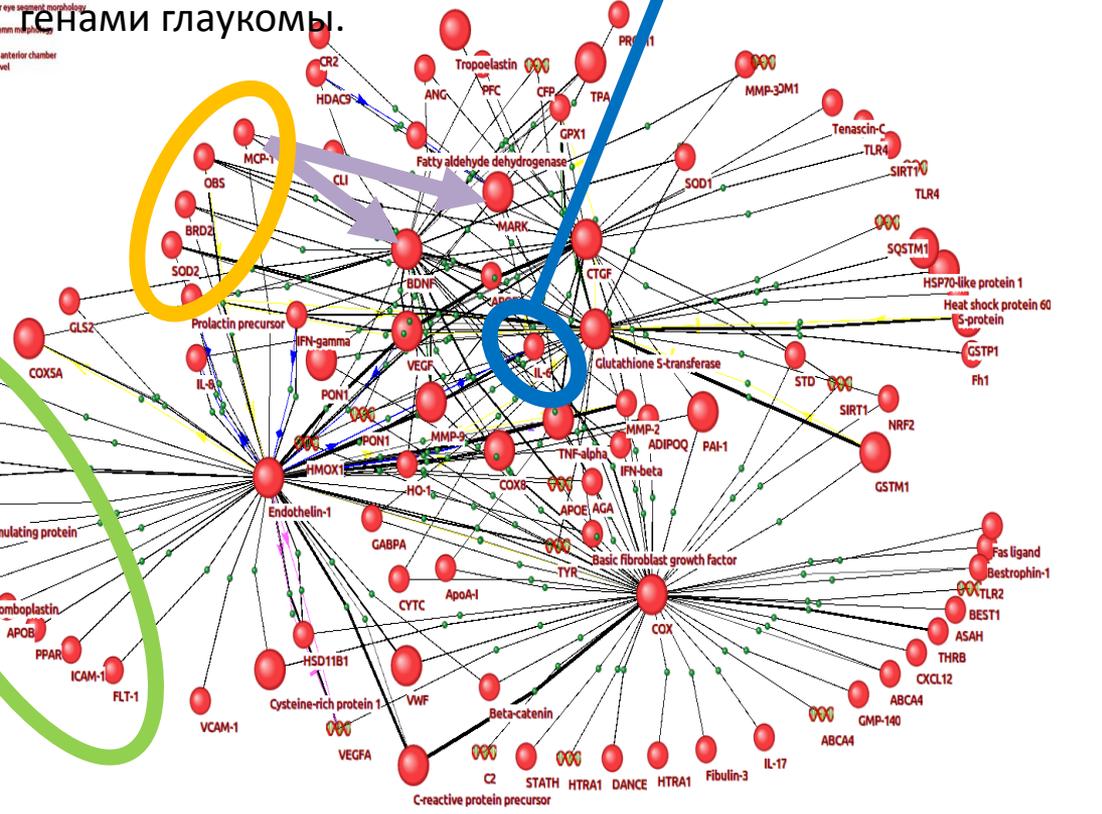
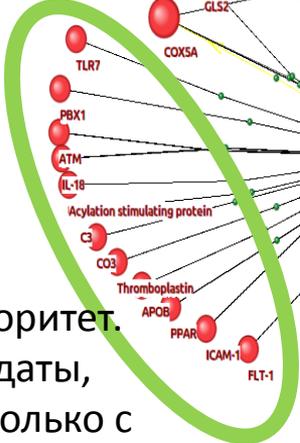
Средний приоритет. Гены-кандидаты, связанные с 2 генами глаукомы.



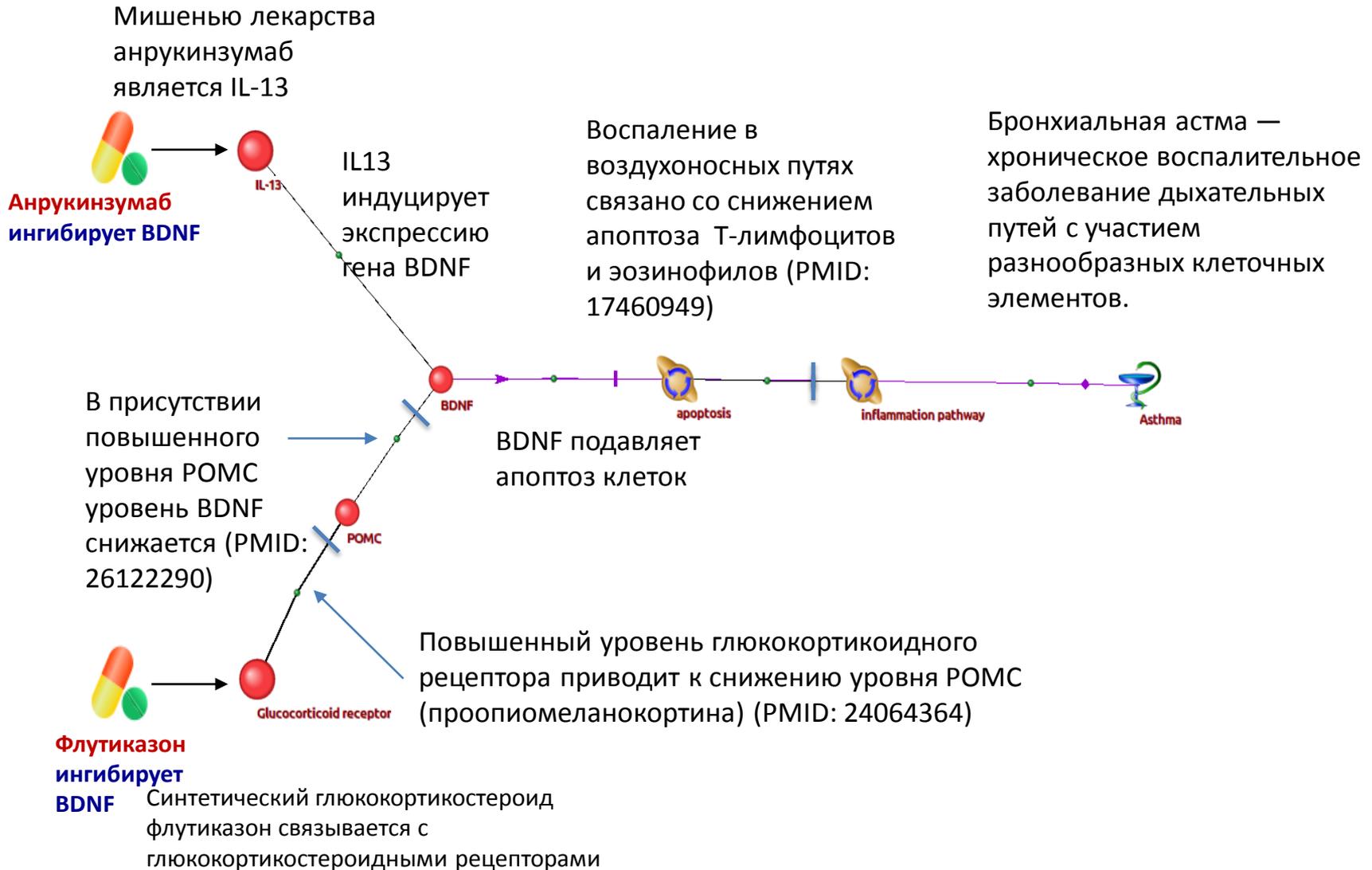
Фрагмент ассоциативной генной сети макулярной дегенерации



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

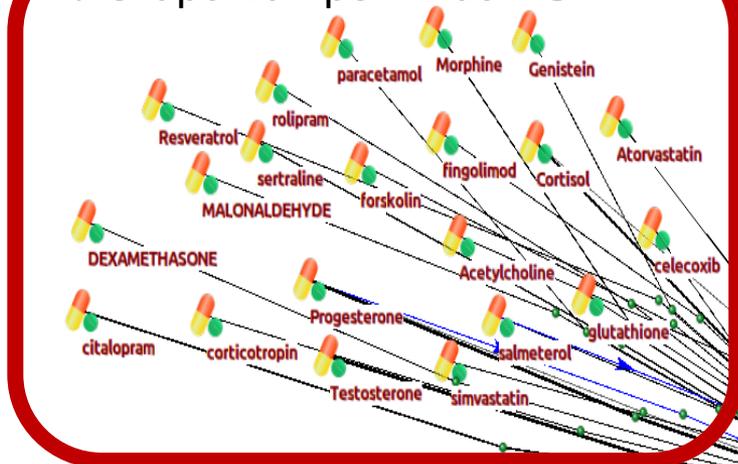


Пример реконструкции молекулярных механизмов действия лекарств.

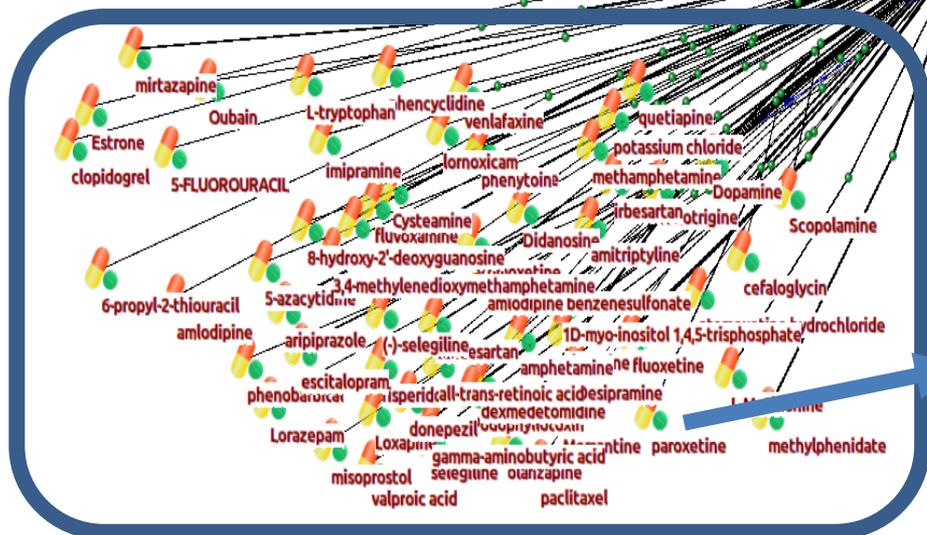


BDNF является мишенью для целого ряда лекарств. Некоторые из этих лекарств могут быть кандидатами для лечения астмы.

Лекарства против астмы

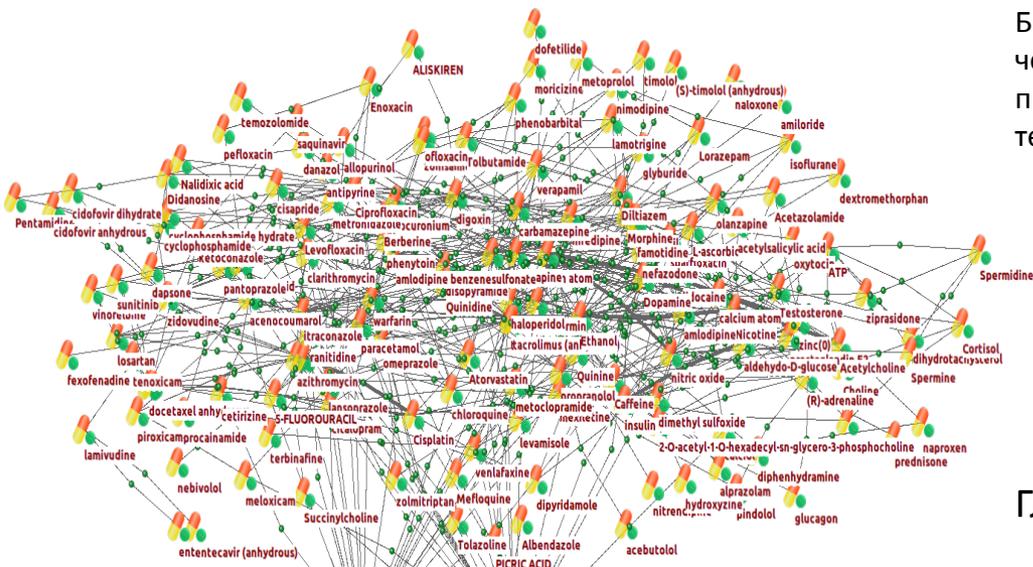


Лекарства против других заболеваний

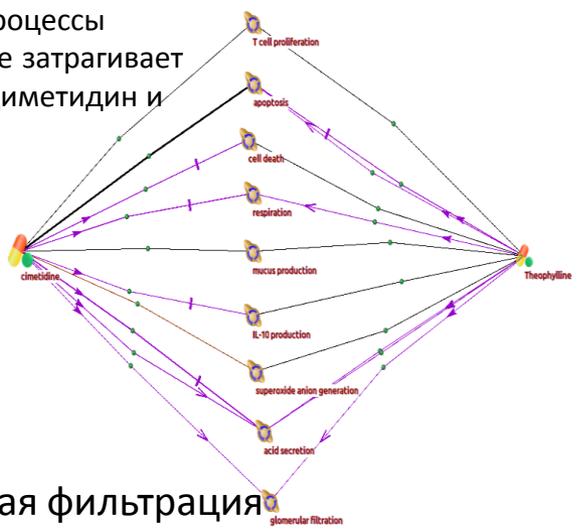


- **Пароксетин** — антидепрессант с сильным противотревожным действием, селективный ингибитор обратного захвата серотонина.

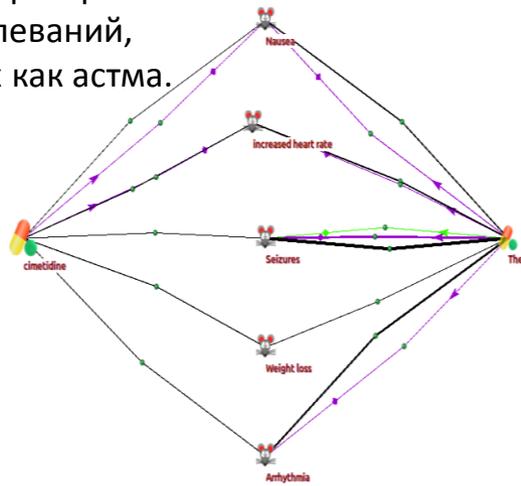
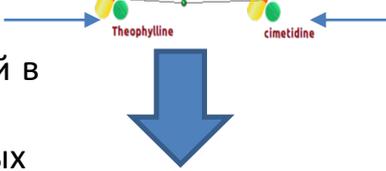
Пример анализа лекарство-лекарство взаимодействий



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



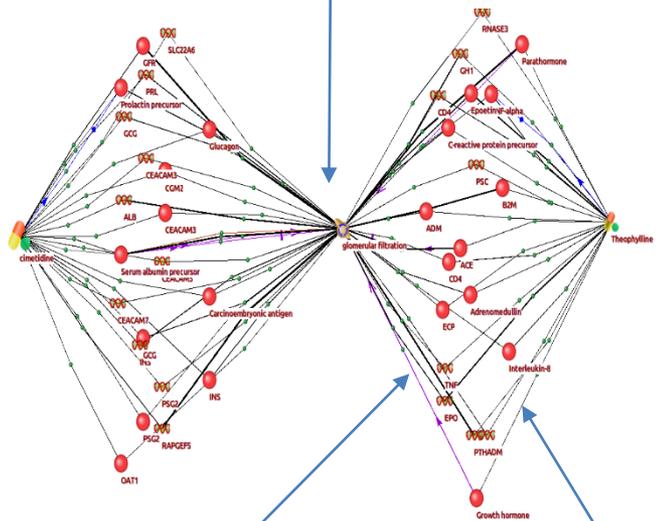
Теофиллин - это препарат, применяемый в терапии респираторных заболеваний, таких как астма.



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.

Гломерулярная фильтрация



Гормоны роста увеличивают **гломерулярную фильтрацию**.

Теофиллин усиливает ответ на **гормоны роста**.

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 Gene Ontology биологических процессов (ресурс DAVID).
- Приоритезация генов-кандидатов (ресурсы Endeavour, TopGene, DIR).

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;
- подграф ассоциативной генной сети выполняет биологически интерпретируемую функцию в рамках функционирования молекулярно-генетической системы, описываемой ассоциативной генной сетью.

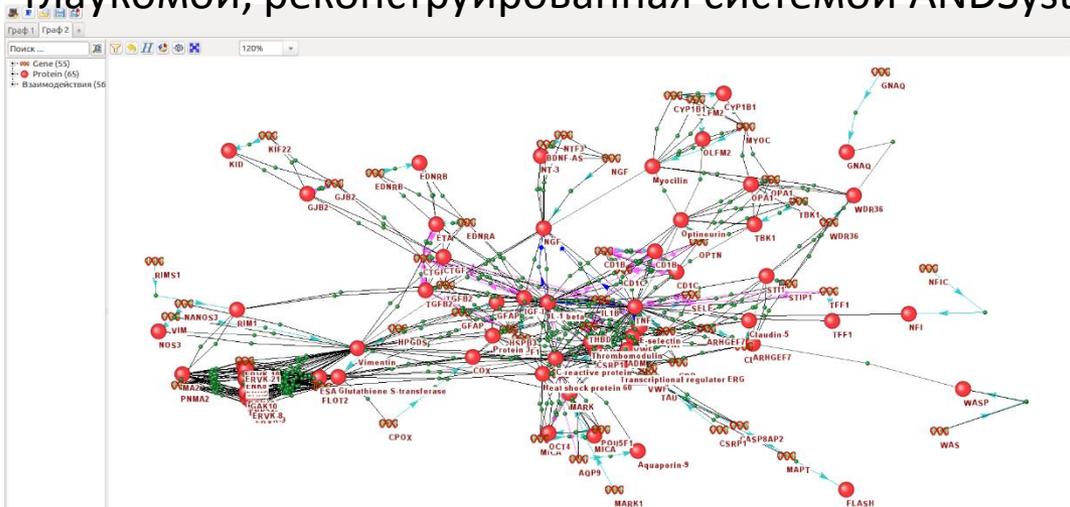
Список плагинов Cytoscape для кластеризации сетей доступен по адресу:

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 indicator (3.0+). The 'ClusterViz' card is highlighted with a red border.

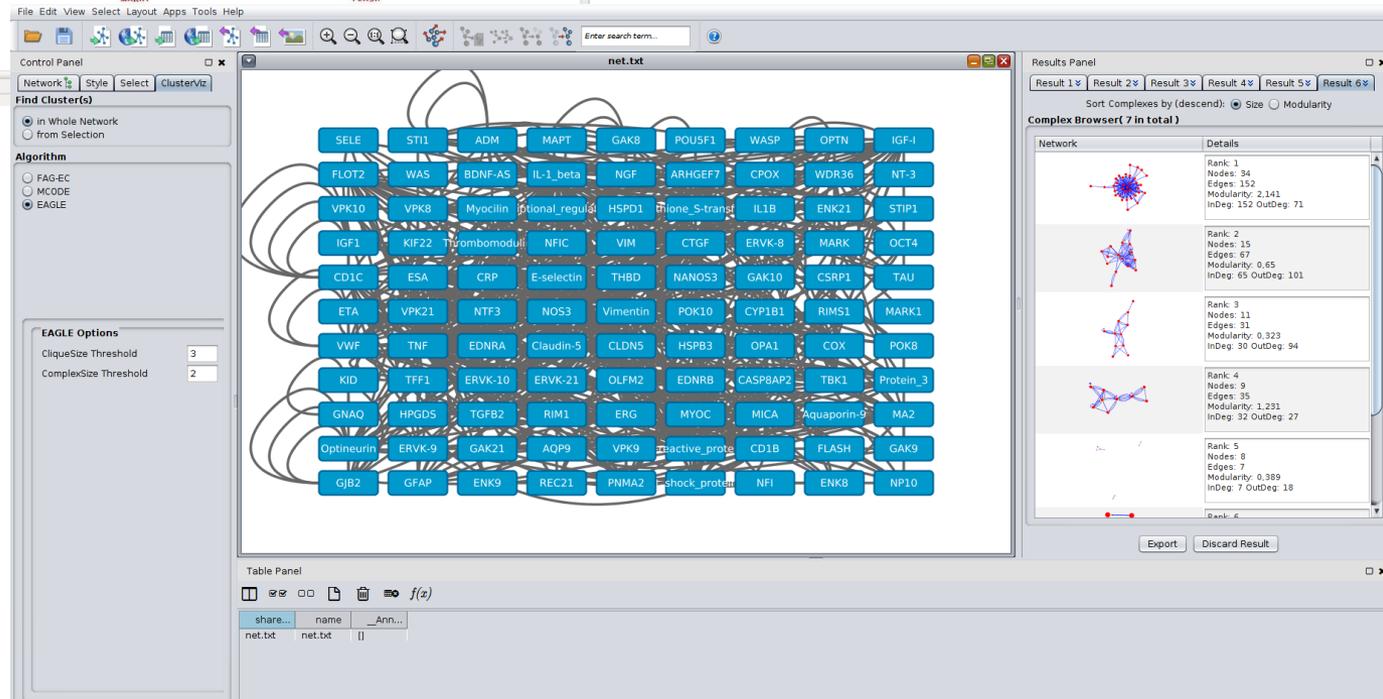
App Name	Description	Version
AutoAnnotate	Finds clusters and visually annotates them with labels and	3.0+
cdtReader	Read Cluster 3 output files into Cytoscape	3.0+
clusterMaker2	Multi-algorithm clustering app for Cytoscape	3.0+
ClusterONE	Finds overlapping protein complexes in a protein interaction	3.0+
ClusterViz	Clustering based on FAG-EC, EAGLE or MCODE. Found cluster	3.0+
CytoCluster	a Cytoscape app for analysis and visualization of clusters from	3.0+
cytoHubba	Predicts and explores important nodes and subnetworks in a given	3.0+
DyNetViewer	Construction, analysis, and visualization of dynamic networks	3.0+
EClizer		3.0+
GIANT	a Cytoscape plugin for network cartography	3.0+
Mass Spec Collection	Collection of apps to facilitate network analysis of mass	3.0+
MCODE	Clusters a given network based on topology to find densely connected	3.0+
ModuLand 2.0	Modularization method family offering modular hierarchies and	3.0+
NCMine	Finds functional modules in gene networks by Near-Clique Mining	3.0+
Omics Analysis Collection	Collection of apps to facilitate network analysis of omic data. By	3.0+
PEmeasure	Compute links weights and assess the reliability of the links in a	3.0+
PEPPER	Find meaningful pathways / complexes connecting a protein set	3.0+
PEWCC	Detect dense sub-network (e.g. protein/gene complexes) in	3.0+
ReactomeFIPlugin	Explore Reactome pathways and search for diseases related	3.0+
SCORE	This is the Cytoscape app implementation for Protein	3.0+

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)



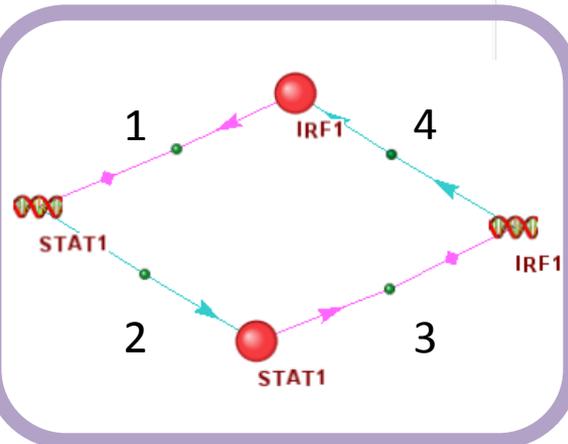
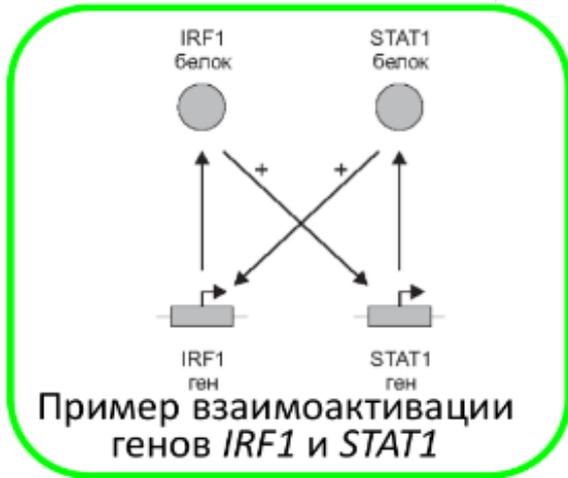
Регуляторные контуры в генных сетях

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.



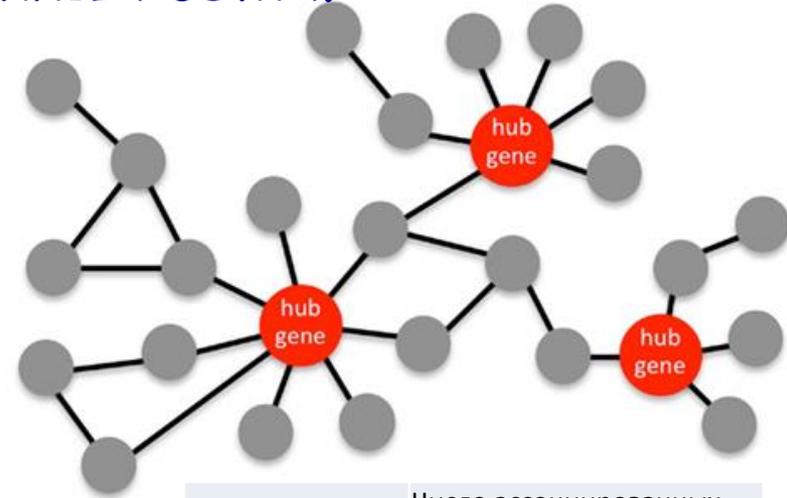
1. Белок **IRF1** -> **регуляция экспрессии** -> ген **STAT1**
Newly synthesized IRF-1 may in turn activate expression of STAT1, resulting in positive feedback regulation of IRF-1 expression (55). Pubmed:23420765.
2. Ген **STAT1** -> **экспрессия** -> белок **STAT1**
3. Белок **STAT1** -> **регуляция 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 дуг 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

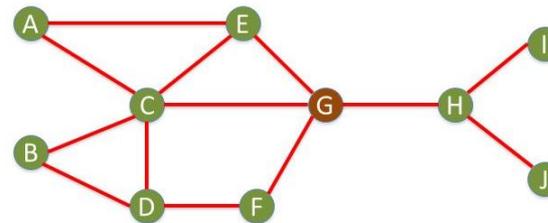
Центральность по степени (degree centrality, DC) - это отношение количества связей определенного узла к общему количеству других узлов в сети. Центральность по степени тем выше, чем больше число связей вершины с другими вершинами в графе. Данная мера принимает значение от 0 до 1 (чем ближе значение к 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) –

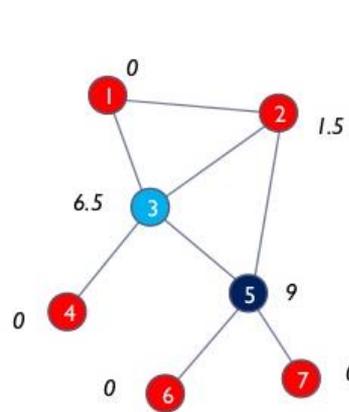
взвешенная сумма всех кратчайших путей между вершинами в графе, проходящих через данную вершину. Этот показатель определяет важность вершины графа с точки зрения ее влияния на пути передачи информации между вершинами графа.

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

V - набор вершин;

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

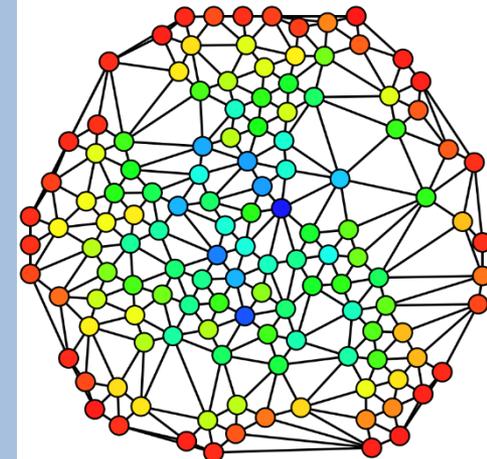
$\sigma(s,t|e)$ - число наикратчайших путей между s и t , проходящих через вершину e .



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

- 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 .
- Нужно сложить полученные значения для всех вершин i и j .



Синие вершины имеют максимальный показатель центральности по посредничеству.

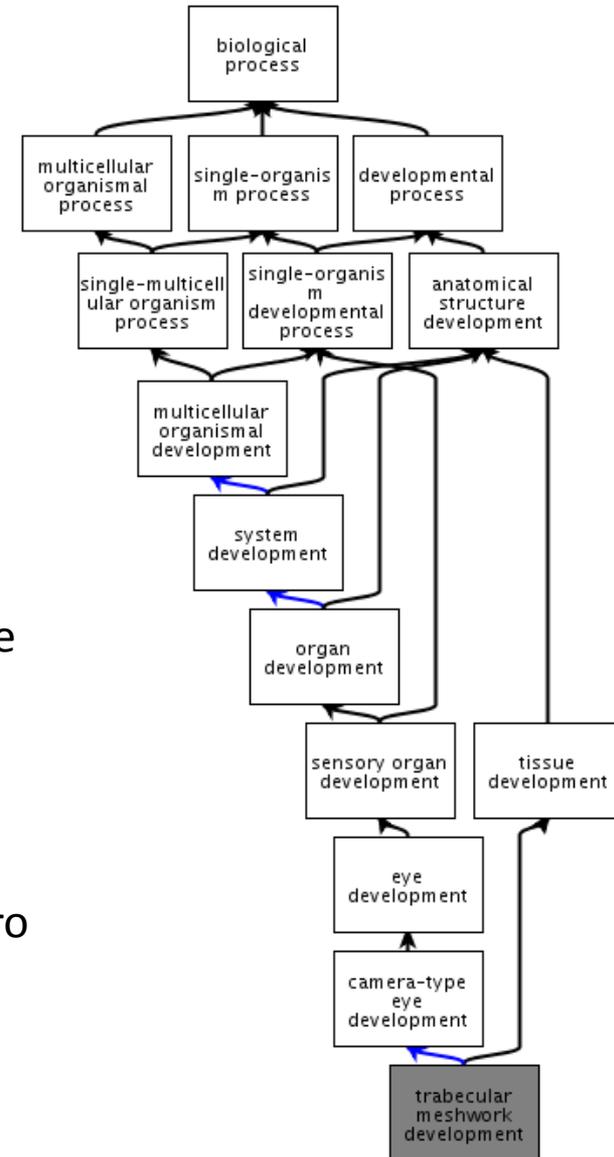
Онтология генов *Gene Ontology (GO)*

Gene Ontology («Генная онтология») — биоинформатический ресурс, содержащий формализованное унифицированное представление молекулярно-генетических и клеточных структур, выполняемых ими элементарных функций, а также процессов, в которые вовлечены эти структуры. С использованием GO онтологии были проаннотированы гены человека и других организмов, в результате чего была создана база данных **Gene Ontology Annotation (GOA)**. Аннотация Gene Ontology Annotation включает разбиение генов на группы согласно их вовлеченности в биологические процессы или клеточные структуры, описанные в онтологии GO.

Онтология доступна по ссылке <http://geneontology.org/>.

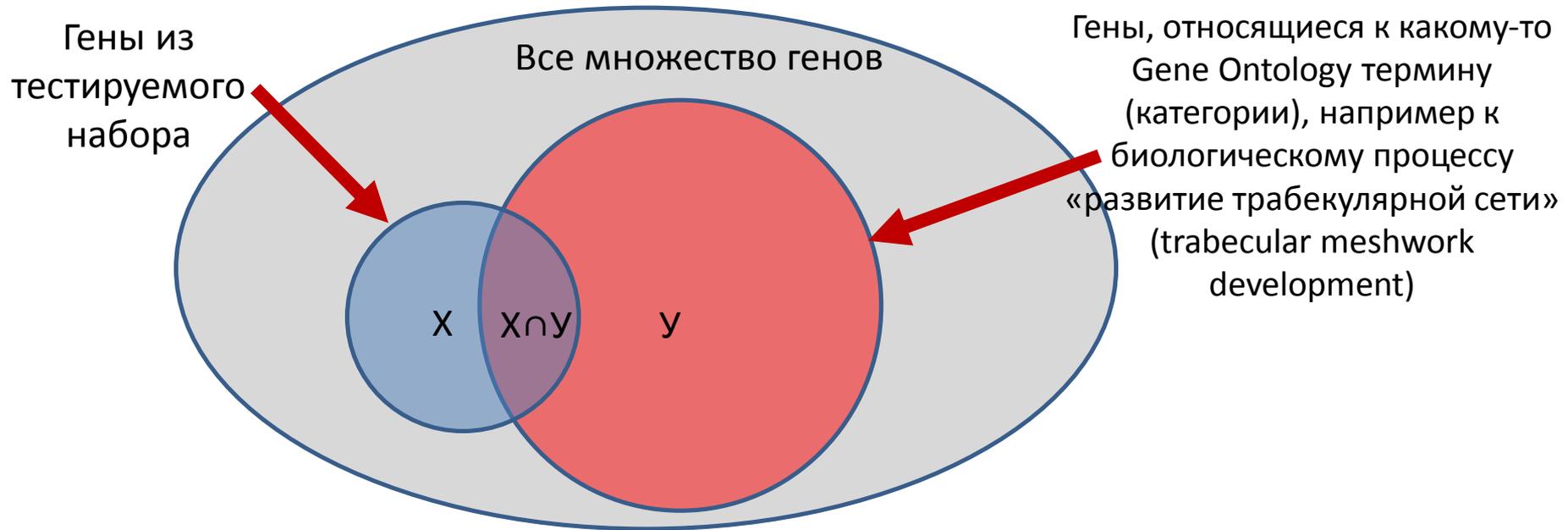
На рисунке приведено графическое представление фрагмента онтологии GO, которое демонстрирует иерархическую структуру онтологии, где на нижнем уровне находятся термины, описывающие наиболее частные биологические процессы (например термин «развитие трабекулярной сети»), а на верхних уровнях расположены термины, описывающие более глобальные процессы (например «развитие глаза», «развитие органов сенсорного восприятия», «развитие органов и систем» и т.д.).

Пример термина GO «развитие трабекулярной сети» (trabecular meshwork development) и его положение в графе онтологии



Определение сверхпредставленных Gene Ontology биологических процессов.

Выявление сверхпредставленных Gene Ontology биологических процессов - это метод интерпретации наборов генов с использованием геной онтологии (Gene Ontology), в которой гены относятся к набору Gene Ontology терминов в зависимости от их функциональных характеристик.



1. На первом шаге выявляются гены из тестируемого набора, которые относятся к какому-то Gene Ontology термину.
2. С помощью точного теста Фишера, гипергеометрического распределения или других статистических методов оценивается вероятность того, что такое же или большее число генов из тестового набора оказались связанными с рассматриваемым Gene Ontology термином по случайным причинам. На этом шаге рассчитывается p-value. При этом учитываются такие величины как $X \cap Y$, X, Y и число всех генов.
3. Процедура из пункта 2 повторяется для всех Gene Ontology терминов. Так как всего Gene Ontology терминов много (порядка 10 тысяч), необходимо делать поправку на множественное сравнение при расчете p-value. Часто используются поправка Бонферрони, Бенджамини-Хокберга и FDR (False Discovery Rate).

Определение сверхпредставленных Gene Ontology биологических процессов (ресурс DAVID).

Ресурс DAVID (<https://david-d.ncifcrf.gov/>) предназначен для функциональной аннотации наборов генов, в том числе позволяет выявлять сверхпредставленные Gene Ontology категории.

1. На главной странице выбираем «Functional Annotation».
2. На появившейся странице выбираем вкладку Upload и загружаем анализируемый список генов (можно загружать различные идентификаторы из баз данных).
3. Выбираем тип загруженных идентификаторов. Нажимаем «Gene List» и «Submit List».
4. На выходе получаем таблицу с сверхпредставленными Gene Ontology категориями с указанием p-value.

Functional Annotation Tool
DAVID Bioinformatics Resources 6.7, NIAID/NIH

*** You are currently using DAVID 6.7 which will be available on our development site until at least May 15, 2017. ***
*** If you are looking for DAVID 6.8, please visit our production site. ***

Submit your gene list to start the tool!

Key Concepts:

The DAVID Gene Concept
DAVID 6.7 is designed around the "DAVID Gene Concept", a graph theory evidence-based method to agglomerate species-specific gene/protein identifiers from a variety of public genomic resources including NCBI, EMBL and Uniprot/SwissProt. The DAVID Gene Concept method groups tens of millions of identifiers from over 65,000 species into 1.5 million unique protein/gene records. [More](#)

Term/Gene Co-Occurrence Probability
Ranking functional categories based on co-occurrence with sets of genes in a gene list can rapidly aid in unraveling new biological processes associated with cellular functions and pathways. DAVID 6.7 allows investigators to sort gene categories from dozens of annotation systems. Sorting can be based either the number of genes within each category or by the EASE-score. [More](#)

Gene Similarity Search
Any given gene is associating with a set of annotation terms. If genes share similar set of those terms, they are most likely involved in similar biological mechanisms. The algorithm tries to group those related genes based on the agreement of sharing similar annotation terms by Kappa statistics. [More](#)

Term Similarity Search
Typically, a biological process/term is done by a corporation of a set of genes. If two or more biological processes are done by similar set of genes, the processes might be related in the biological network somehow. This search function is to identify the related biological processes/terms by quantitatively measuring the degree of the agreement how terms share the similar participating genes. [More](#)

DAVID Bioinformatics Resources 6.7
National Institute of Allergy and Infectious Diseases (NIAID), NIH

*** You are currently using DAVID 6.7 which will be available on our development site until at least May 15, 2017. ***
*** If you are looking for DAVID 6.8, please visit our production site. ***

Functional Annotation Chart

Current Gene List: List_1
Current Background: Homo sapiens
89 DAVID IDs
Options

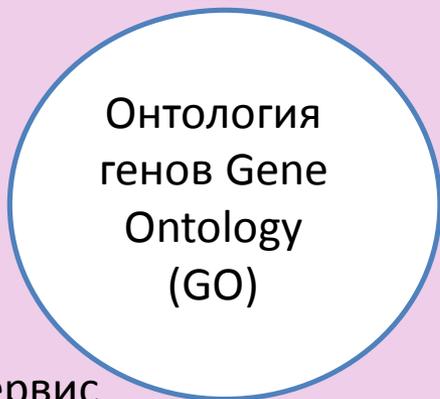
532 chart records

Sublist	Category	Term	RT	Genes	Count	%	P-Value	BenjaminF
GOTERM_BP_FAT	regulation of cell proliferation		RT	22	24.7	2.3E-10	3.9E-7	
GOTERM_BP_FAT	cell migration		RT	14	15.7	1.3E-9	1.2E-6	
GOTERM_BP_FAT	negative regulation of cell proliferation		RT	15	16.9	3.6E-9	2.1E-6	
GOTERM_BP_FAT	localization of cell		RT	14	15.7	4.9E-9	2.1E-6	
GOTERM_BP_FAT	cell motility		RT	14	15.7	4.9E-9	2.1E-6	
GOTERM_BP_FAT	regulation of apoptosis		RT	20	22.5	1.5E-8	5.0E-6	
GOTERM_BP_FAT	regulation of programmed cell death		RT	20	22.5	1.7E-8	4.9E-6	
GOTERM_BP_FAT	regulation of cell death		RT	20	22.5	1.8E-8	4.5E-6	
GOTERM_BP_FAT	regulation of MAPKK cascade		RT	9	10.1	8.5E-8	1.8E-5	
GOTERM_BP_FAT	regulation of phosphorylation		RT	15	16.9	9.0E-8	1.7E-5	
GOTERM_BP_FAT	cell motion		RT	15	16.9	1.1E-7	2.0E-5	
GOTERM_BP_FAT	regulation of phosphate metabolic process		RT	15	16.9	1.5E-7	2.3E-5	
GOTERM_BP_FAT	regulation of phosphorus metabolic process		RT	15	16.9	1.5E-7	2.3E-5	
GOTERM_BP_FAT	positive regulation of transport		RT	11	12.4	2.0E-7	2.9E-5	

Выявление групп генов, статистически достоверно ассоциированных с определёнными биологическими процессами из онтологии генов Gene Ontology (GO)

Гены, ассоциированные с глаукомой:

ADRB2, APOE, BDNF, CAV1, CAV2, CD40LG, CD44, CD79A, CDKN1A, CDKN2B, COL11A1, CYP1B1, EDN1, ESR2, FN1, GLC1B, GLC1C, GLC1D, GPX3, GSTM1, GSTT1, CFH, HGF, HK2, HSPA1A, HSPA1B, IL1B, IL6, KCNA5, KIF2A, KLK2, LMX1B, LOXL1, LTA, LTBP2, MMP2, MMP3, MMP9, MTHFR, MYOC, NOS2, NTF4, OPA1, SIX6, SERPINE1, PON1, SALL1, SRSF3, SIX1, SOD2, SPP1, TGFB2, TIMP1, TIMP2, TLR4, TNF, TP53, TTR, VAV2, VEGFC, NRP1, TCEAL1, ST18, MED12, OPTN, ANGPTL7, VAV3, TXNRD2, CYP46A1, PADI2, ZFPM2, OPTC, NSG1, HPGDS, HOOK2, LRP12, ISYNA1, SRBD1, ELOVL5, CNTNAP4, PCMTD1, WDR36, ASB10, PLEKHA7, NCKAP5, GLC1H, GLC1I, GLC1K, GLC1J, LOC730100, GLC1M, GLC1N, CDKN2B-AS1, GLC1P, LOC102724330, LOC105375094



сервис DAVID



Группа генов	Gene ontology биологические процессы	p-value
TP53, TNF, IL6, IL1B, CDKN1A, CD40LG, CD79A	Иммунные процессы	0.02
TP53, TNF, APOE, IL6, MMP9, IL1B, BDNF, TLR4, CDKN1A, ADRB2, ESR2, CD44, SOD2, HGF, CD40LG, LTA, HSPA1A, TGFB2, HSPA1B, VAV2, VAV3	Регуляция клеточной смерти	2.4E-8
TP53, APOE, IL1B, EDN1, CAV1, SOD2, KCNA5	Гомеостаз ионов	0.004

Домашняя страница сервиса DAVID

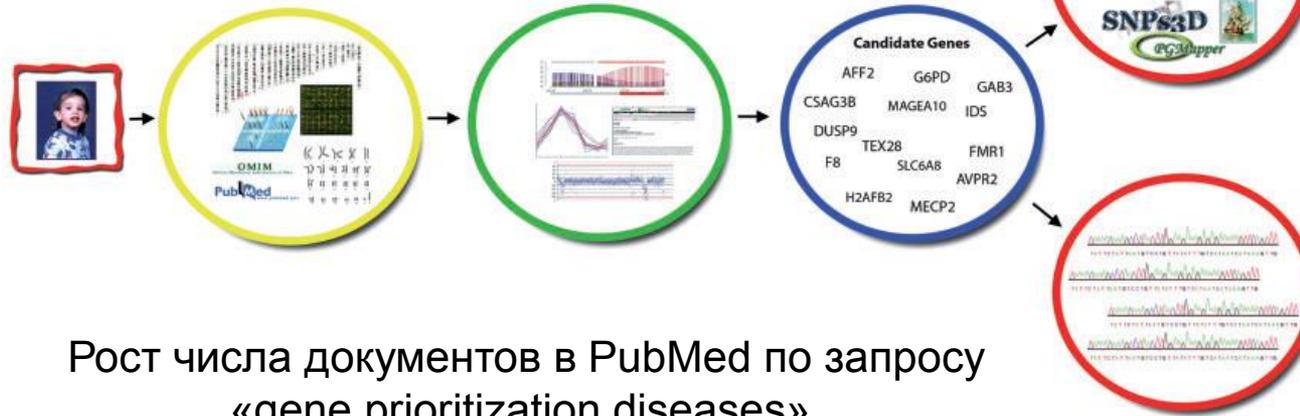
Программа DAVID

(<https://david.ncifcrf.gov/home.jsp>) - один из основных веб-инструментов для поиска сверхпредставленных GO биологических процессов для набора генов.

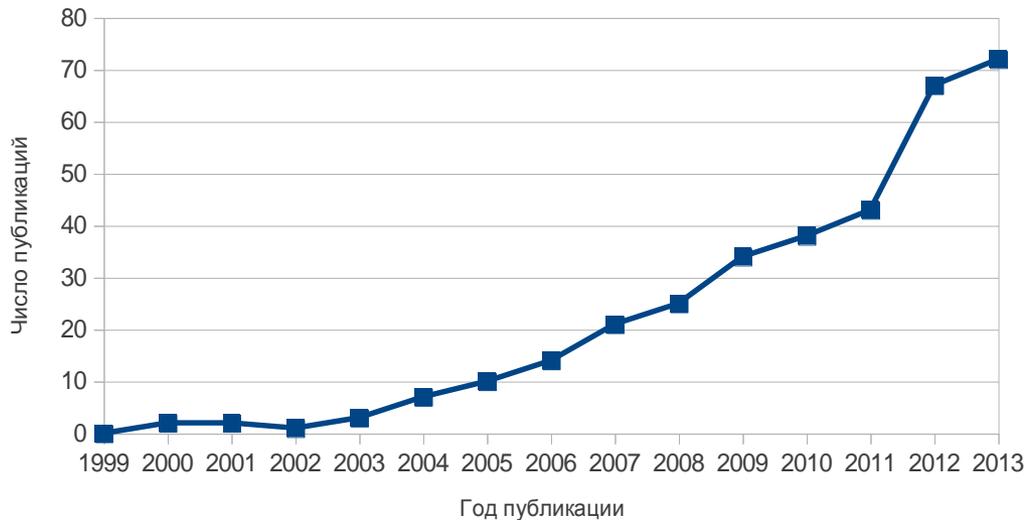
Например в результате анализа с помощью сервиса DAVID был выявлен комплекс из 7 генов (TP53, TNF, IL6, IL1B, CDKN1A, CD40LG, CD79A), каждый из которых ассоциирован с ПОУГ, вовлеченных в функционирование иммунной системы ($p < 0.02$).

Приоритезация генов

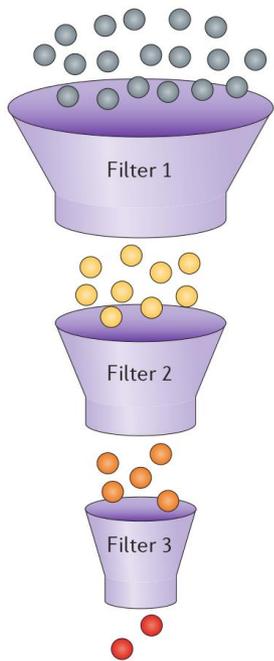
Доказательство причинной связи между геном и болезнью с помощью эксперимента требует много денежных и временных затрат. Всесторонняя приоритезация генов-кандидатов до экспериментальной проверки снижает связанные с этим расходы.



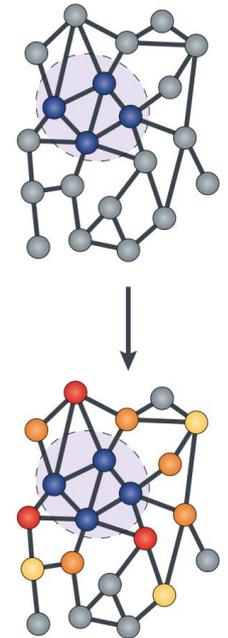
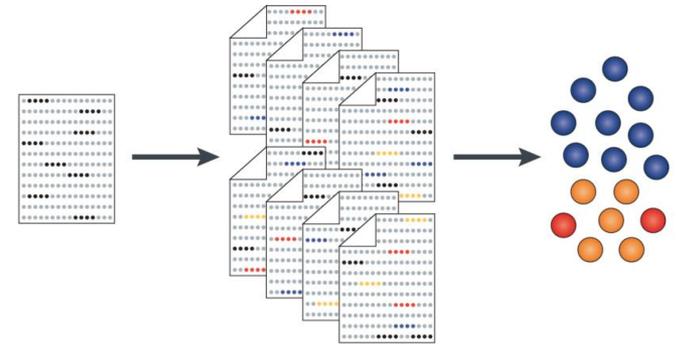
Рост числа документов в PubMed по запросу «gene prioritization diseases».



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.

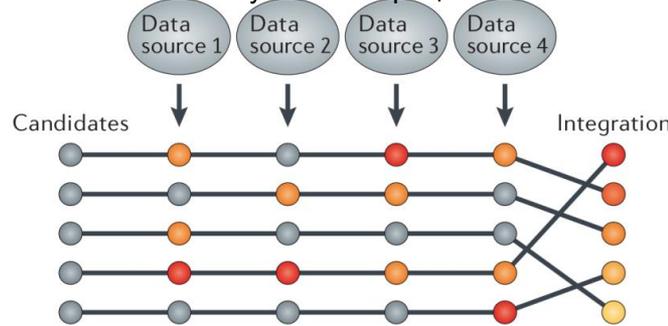


Анализ биологических сетей.

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

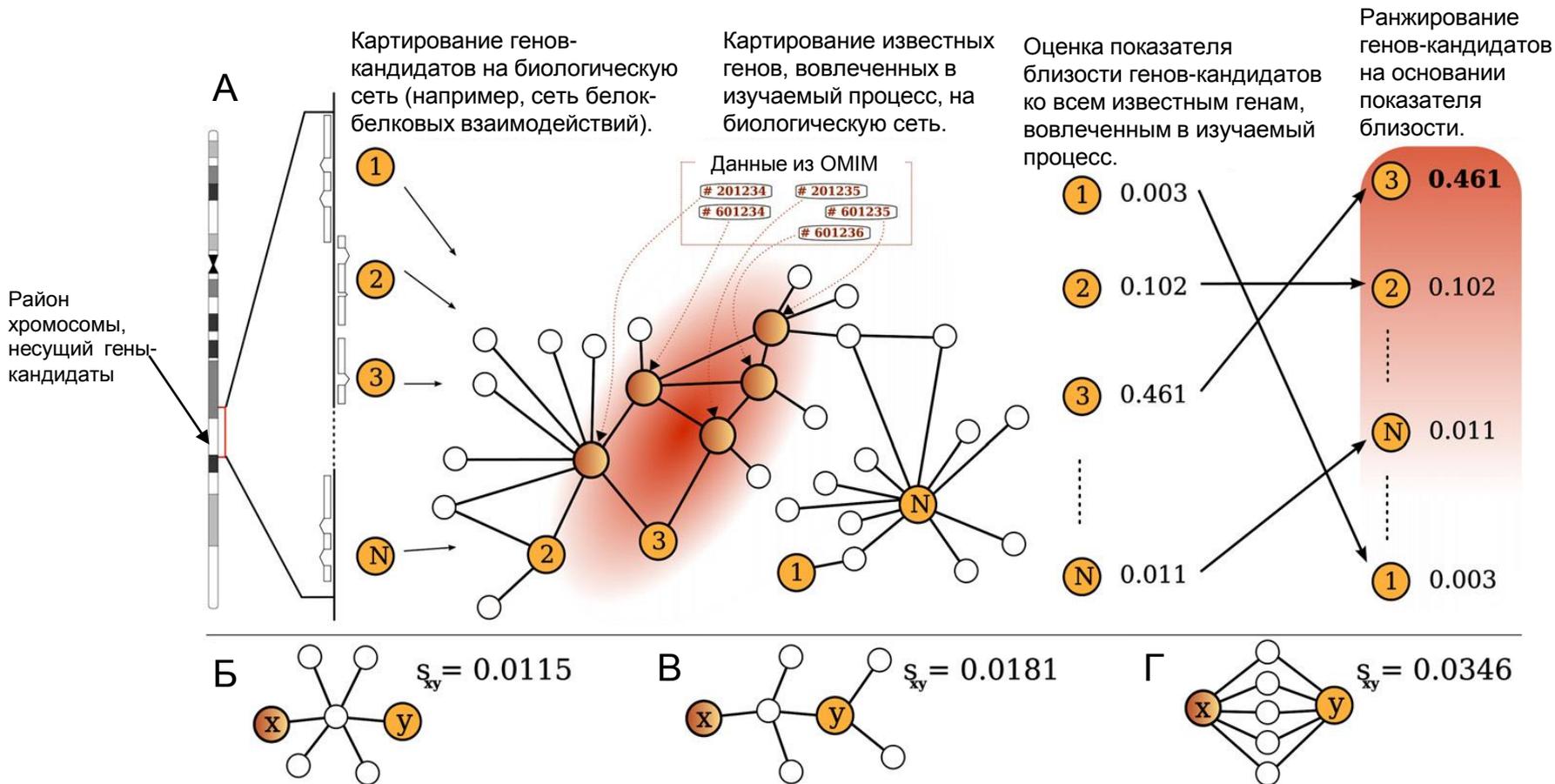
Фильтрация генов-кандидатов на основе их соответствия различным критериям (вовлечение в Gene Ontology процессы, упоминания в GWAS исследованиях и др.). Основным ограничением является то, что строгий процесс фильтрации не позволяет проводить тонкий анализ набора генов-кандидатов. Если ген-кандидат не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

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.



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

Анализ биологических сетей при приоритизации генов

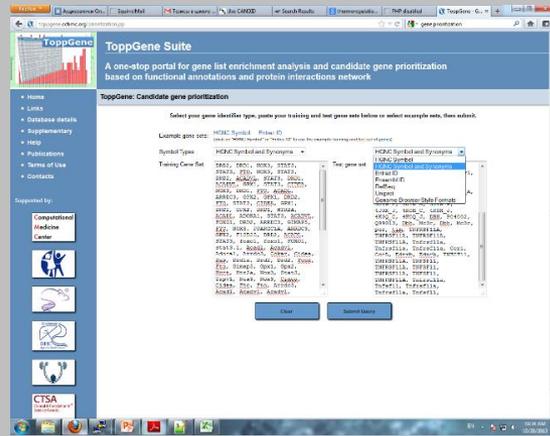


Каждая из трех подсетей **Б**, **В** и **Г** имеет одинаковое количество вершин, но различную конфигурацию и соответственно показатель близости генов **X** и **Y**. В подсети **Б** вершины **X** и **Y** связаны через хаб, имеющий множество связей с другими вершинами, поэтому показатель близости **S_{xy}** 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.

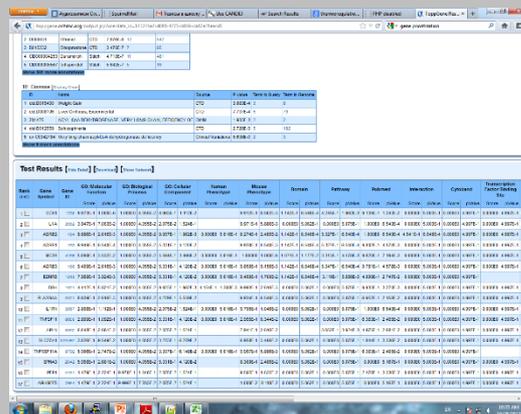
Приоритезация генов-кандидатов (ресурсы Endeavour, ToppGene, DIR).

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

Начальная страница ToppGene

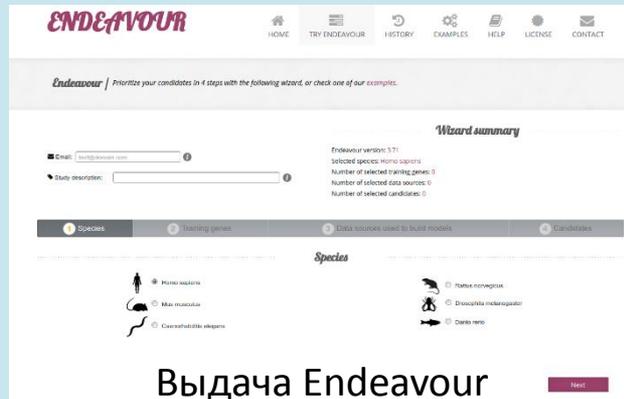


Выдача ToppGene

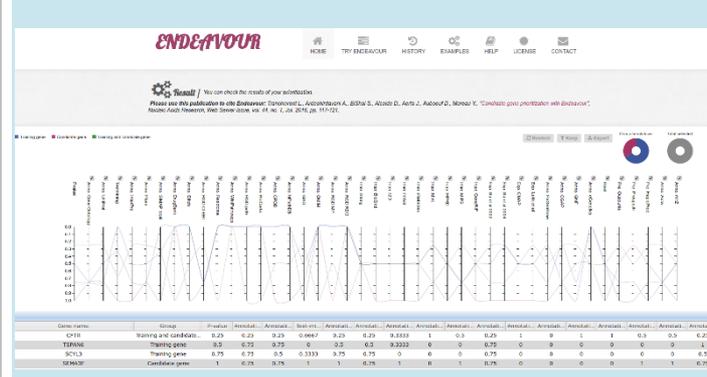


Endeavour (<https://endeavour.esat.kuleuven.be>) используются обширные геномные данные из различных источников, а также информация о связи генов и биологических процессов.

Начальная страница Endeavour



Выдача Endeavour



DIR (<http://cbc.case.edu/dir/>) используется анализ геновых 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
pwpy	3.470162

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

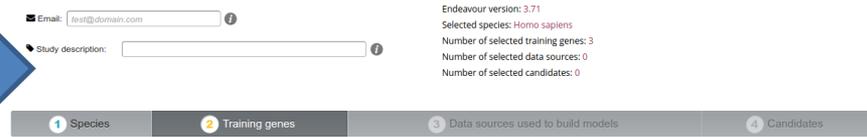
Rank	Gene	Score
1	6870	4.049330
2	2895	4.041079
3	5602	2.231536
4	6622	2.079720
5	658	2.073226
6	5530	1.994358
7	6696	1.693941
8	5473	1.688436
9	4790	1.679600
10	8649	1.670543

Endeavour

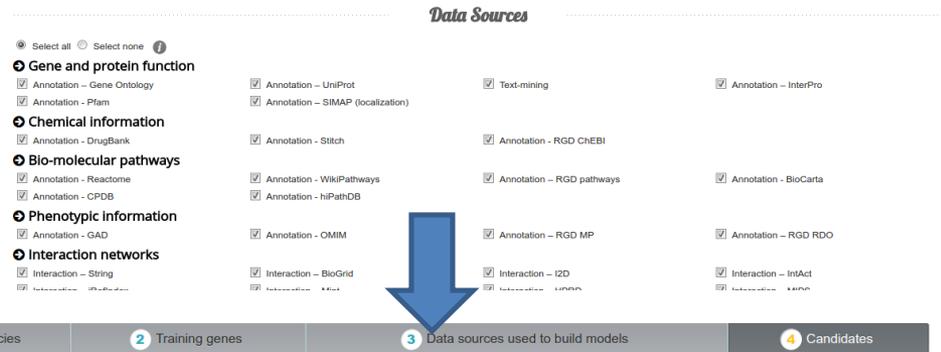
1) Выбор организма



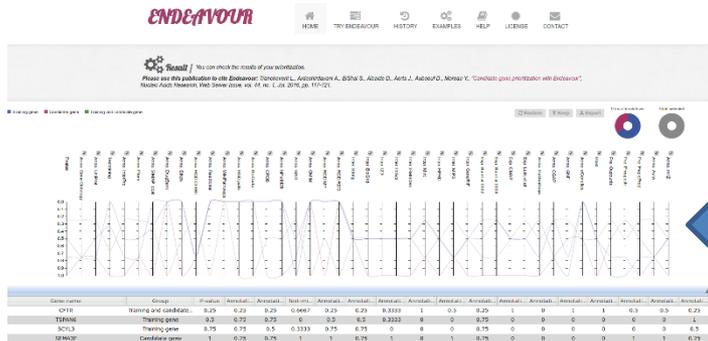
2) Загрузка тренировочного набора генов



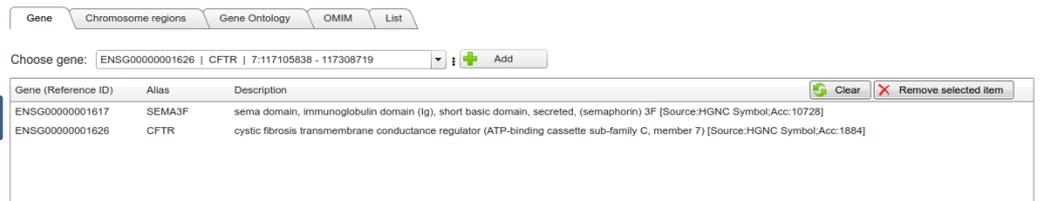
3) Выбор источников аннотации генов для приоритезации



5) Приоритизированный список генов-кандидатов

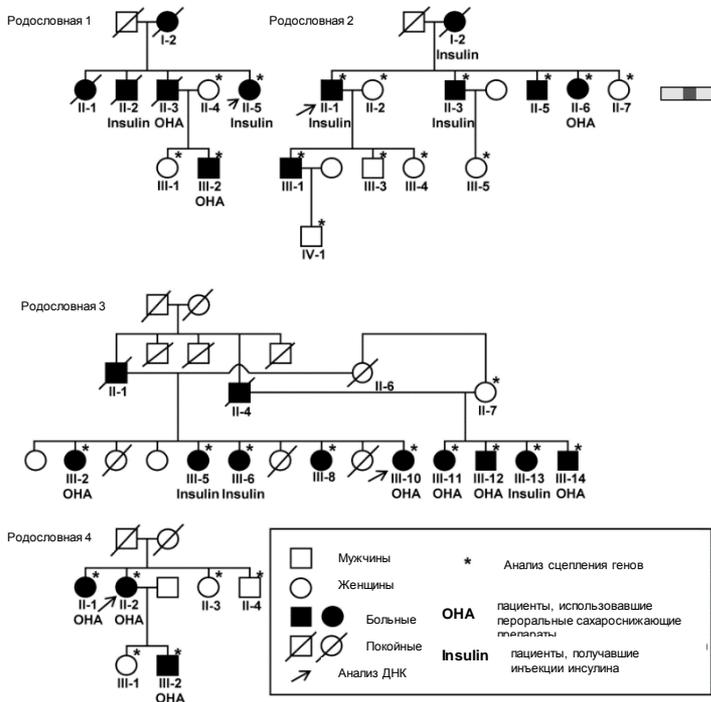


4) Загрузка генов-кандидатов



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.

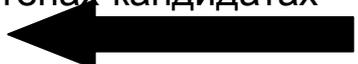
Приоритезация 106 генов из района 2p25-22 с помощью сервиса **Endeavour** на наборе обучающих генов (HNF4A, GSK, HNF1A, PDX1, HNF1B, NEUROD1).

Топовые гены-кандидаты

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

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

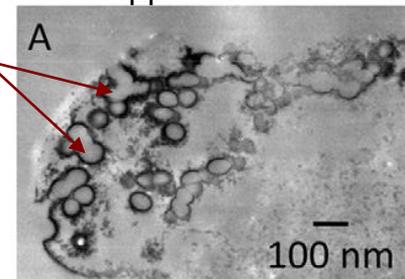
Вестерн-блот анализ



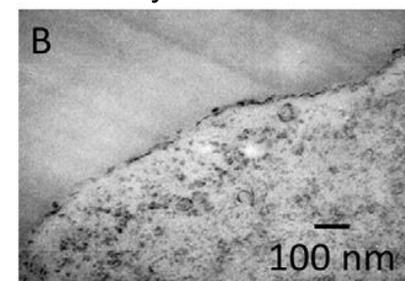
Кавеолы, в образовании которых участвует продукт гена **PTRF-CAVIN**.

Электронные микрофотографии (x 37.000) клеточной поверхности фибробласта

Дикий тип



Мутантный тип



Примеры исследований, проведенных с использованием ANDSystem. Реконструкция некоторых молекулярно-генетических механизмов патогенеза первичной открыто-угольной глаукомы (ПООГ) с помощью системы ANDSystem.

Системная биология и моделирование
ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

Вавиловский журнал генетики и селекции. 2016;20(6):840-847
DOI 10.18699/VJ16.200

Молекулярно-генетические механизмы взаимодействия процессов ответа клетки на механический стресс и нейронального апоптоза при первичной открытоугольной глаукоме

О.В. Саик¹, Н.А. Коновалова², П.С. Демеников¹, Н.В. Иванисенко^{1,3}, Т.В. Иванисенко¹, Д.Е. Иваношчук¹, М.Н. Пономарева², О.С. Коновалова², О.А. Подколodная⁴, И.Н. Лаврик^{1,4}, Н.А. Колчанов¹, В.А. Иванисенко¹

¹ Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Института цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный медицинский университет» Министерства здравоохранения Российской Федерации, Томск, Россия

³ Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет», Новосибирск, Россия

⁴ Магдебургский университет имени Отто фон Гуерике, Магдебург, Германия

Глаукома – хроническое, прогрессирующее заболевание, которым страдают более 60 млн человек в мире. Первичная открытоугольная глаукома (ПООГ) занимает одно из первых мест по распространенности среди различных форм глаукомы. Например, в 2011 г. эта форма заболевания наблюдалась более чем у 2,7 млн человек в США. В настоящее время ПООГ является основной причиной необратимой потери зрения. У больных с открытоугольной глаукомой риск слепоты достигает 27 %, даже несмотря на проводимое лечение. Известно, что гибель клеток зрительного нерва может быть спровоцирована механическим стрессом, вызванным повышенным внутриглазным давлением, наблюдающимся при ПООГ, индуцирующим нейрональный апоптоз. В настоящее время существует огромное количество научных публикаций, описывающих белки и гены, которые участвуют в патогенезе ПООГ, в том числе в нейрональном апоптозе и клеточном ответе на механический стресс. Тем не менее молекулярно-генетические механизмы, лежащие в основе патофизиологии ПООГ, до сих пор плохо изучены. Реконструкция ассоциативных генных сетей, описывающих функциональные взаимодействия между этими генами/белками, включая биохимические реакции, регуляторные взаимодействия, события транспорта и т. д., требует использования методов автоматизированного извлечения знаний из текстов научных публикаций. В работе проанализированы ассоциативные сети, описывающие молекулярно-генетические взаимодействия между белками и генами, вовлеченными в ответ клетки на механический стресс (ОКМС), нейрональный апоптоз и патогенез ПООГ. Показано, что гены, ассоциированные с ПООГ, статистически значимо чаще, чем ожидалось по случайным причинам, представлены среди генов, участвующих во взаимодействии между ОКМС и нейрональным апоптозом, что может быть важным фактором, влияющим на гибель ганглиозных ретинальных клеток при ПООГ.

Ключевые слова: апоптоз; нейрональный апоптоз; ответ клетки на механический стресс; первичная открытоугольная глаукома; ПООГ; ANDSystem; ассоциативные генные сети.

Molecular-genetic mechanisms of the interaction between processes of cell response to mechanical stress and neuronal apoptosis in primary open-angle glaucoma

O.V. Saik¹, N.A. Konovalova², P.S. Demenikov¹, N.V. Ivanisenko^{1,3}, T.V. Ivanisenko¹, D.E. Ivanoshchuk¹, M.N. Ponomareva², O.S. Konovalova², O.A. Podkolodnaya⁴, I.N. Lavrik^{1,4}, N.A. Kolchanov¹, V.A. Ivanisenko¹

¹ Institute of Cytology and Gene SB RAS, Novosibirsk, Russia

² Tyumen State Medical University, Tyumen, Russia

³ Novosibirsk State University, Novosibirsk, Russia

⁴ Institute of Experimental Internal Medicine, Otto von Guericke University, Magdeburg, Germany

Glaucoma is a chronic and progressive disease, which affects more than 60 million people worldwide. Primary open-angle glaucoma (POAG) is one of the most common forms of glaucoma. For example, about 2.7 million people in the USA had primary open-angle glaucoma (POAG) in 2011. Currently POAG is a major cause of irreversible vision loss. In patients with treated open-angle glaucoma the risk of blindness reached to be about 27 %. It is known that the death of optic nerve cells can be triggered by mechanical stress caused by increased intraocular pressure, which induces neuronal apoptosis and is observed in patients with POAG. Currently, there is a large number of scientific publications describing proteins and genes involved in the pathogenesis of POAG, including neuronal apoptosis and the cell response to mechanical stress. However, the molecular-genetic mechanisms underlying the pathophysiology of POAG are still poorly understood. Reconstruction of associative networks describing the functional interactions between these genes/proteins, including biochemical reactions, regulatory interactions, transport, etc., requires the use of methods of automated knowledge

Molecular associations of Primary Open-Angle Glaucoma with potential comorbid diseases (POAG-associome)

Olga V Saik¹, Natalia A Konovalova², Pavel S Demenikov¹, Timofey V Ivanisenko^{1,3}, Evgeny D Petrovskiy¹, Nikita V Ivanisenko^{1,3}, Dinara E Ivanoshchuk¹, Maria N Ponomareva², Olga S Konovalova², Inna N Lavrik^{1,4}, Nikolay A Kolchanov¹, Vladimir A Ivanisenko¹

¹ The Federal Research Center Institute of Cytology and Genetics, The Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentyeva 10, 630090, Novosibirsk, Russia

² Federal State Budgetary Educational Institution of Higher Education "Tyumen State Medical University" of the Ministry of Healthcare of the Russian Federation, Odessa street, 54, 625023, Tyumen, Russia

³ Novosibirsk State University, Novosibirsk State University, Pirogova street 2, 630090, Novosibirsk, Russia

⁴ Department of Translational Inflammation, Institute of Experimental Internal Medicine, Otto von Guericke University, Leipziger Straße 44, D-39120, Magdeburg, Germany

saik@bionet.nsc.ru

ABSTRACT

Glaucoma is the leading cause of irreversible vision loss, which is caused by death of the retinal ganglion cells. Currently, glaucoma affects over 60 million people worldwide with primary open-angle glaucoma (POAG) being one of the most common forms of the disease. Despite the large amount of research devoted to glaucoma, molecular and genetic mechanisms of its development are still poorly understood. Thus, the aim of the present study was prediction of new potentially comorbid diseases of POAG, based on analysis of associative gene networks describing disease-disease interactions. Application of enrichment analysis to associative networks, constructed with the ANDSystem for 31 diseases that are comorbid to POAG according to the literature data, revealed that 10 diseases had a statistically significant overlap of proteins/genes with the POAG associative network (p -value < 0.01). Comparison of POAG with over 4000 diseases with the aid of the ANDSystem showed that there was a statistically significant overrepresentation of proteins/genes in the POAG associative network for more than 100 diseases. Analysis of Gene Ontology (GO) biological processes showed the importance of apoptosis-related and endothelium-related processes for the formation of comorbid conditions of POAG with cancer and cardiovascular diseases, among others.

Keywords: primary open-angle glaucoma, POAG, comorbid diseases, apoptosis, endothelial dysfunction, ANDSystem, gene networks

Biotechnologia Aplicada 2016;33:3201-3206

RESUMEN

Asociaciones moleculares del glaucoma primario de ángulo abierto (POAG) con enfermedades comórbidas potenciales (asocioma de POAG). El glaucoma es la principal causa de la pérdida irreversible de la visión, que ocurre por la muerte de las células ganglionares de la retina. El glaucoma afecta a 60 millones de personas en el mundo y la forma conocida como glaucoma primario de ángulo abierto (POAG) es una de las más frecuentes. A pesar de los abundantes estudios sobre el glaucoma, aun se comprenden poco los mecanismos moleculares y genéticos que motivan su desarrollo. Por tales razones, el propósito de este estudio consistió en predecir cuáles son las enfermedades que muestran comorbilidad con el POAG, mediante el análisis de las redes de asociación de genes que describen las interacciones entre enfermedades. Diez enfermedades mostraron un solapamiento significativo de proteínas y genes con la red asociativa del POAG ($p < 0.01$). Ello fue posible tras el análisis de enriquecimiento de redes asociativas, construidas con el empleo del sistema ANDSystem, a partir de 31 enfermedades para las que se había descrito comorbilidad con el POAG en la literatura científica. La comparación del POAG con 4000 enfermedades, con la ayuda del sistema ANDSystem, mostró una representación incrementada significativa de proteínas y genes en la red asociativa de POAG para más de 100 enfermedades. El análisis de procesos biológicos en la Ontología de Genes (GO) permitió determinar la importancia de los procesos vinculados a la apoptosis y al endotelio, para la formación de condiciones de comorbilidad del POAG con el cáncer y las enfermedades cardiovasculares, entre otras.

Palabras clave: glaucoma primario de ángulo abierto primario, POAG, enfermedades comórbidas, apoptosis, disfunción endotelial, ANDSystem, redes de genes

Introduction

Glaucoma is a neurodegenerative disease characterized by the progressive loss of retinal ganglion cells, followed by a distinctive visual field constriction, and eventually by loss of vision [1]. In most cases, glaucoma is caused by an increase in intraocular

pressure, although it may also develop under normal eye physiology, and primary open-angle glaucoma (POAG) is one of the commonest [2]. Therefore, current therapies for the disease are directed at reducing intraocular pressure; methods include laser or surgical

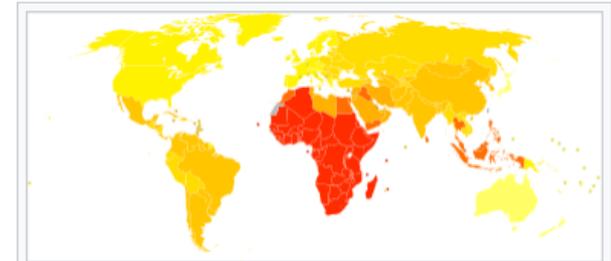
1. Sándor V. Glaucoma: an extension of various chronic neurodegenerative disorders. Mol Neurobiol. 2013;48(1):186-9.

2. Quigley HA. Glaucoma. Lancet. 2011;377(9744):1367-77.

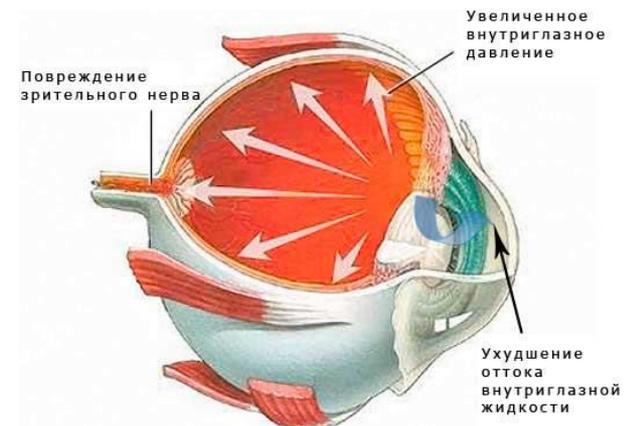
Первичная открыто-угольная глаукома (ПОУГ)

Глаукома — группа глазных заболеваний, характеризующаяся постоянным или периодическим повышением внутриглазного давления с последующим развитием типичных дефектов поля зрения, снижением остроты зрения и атрофия зрительного нерва.

В мире от ПОУГ страдает порядка 44 млн. человек (Tham, Cheng, 2016).



Число потерянных лет жизни (DALY) из-за глаукомы на 100 000 жителей в 2004 году^[60]



Одной из основных причин необратимой потери зрения является первичная открыто-угольная глаукома (ПОУГ). Гибель клеток зрительного нерва может провоцироваться механическим стрессом, вызванным повышенным внутриглазным давлением, наблюдающимся при ПОУГ, индуцирующим нейрональный апоптоз.

Целью работы был анализ ассоциативных сетей, описывающих молекулярно-генетические взаимодействия между белками и генами, вовлеченными в патогенез ПОУГ.

Потеря полей зрения при глаукоме



Нормальное зрение

Центральное зрение

План исследования

- 1) Поиск генов, ассоциированных с ПОУГ:
 - в компьютерных базах данных **OMIM, ClinVar, SNPedia, GWAS catalog**
 - базе знаний ANDCell системы **ANDSystem**
- 2) Выявление **сверхпредставленных Gene Ontology биологических процессов** в наборе генов, ассоциированных с ПОУГ, выявленные системой **DAVID**.
- 3) **Реконструкция ассоциативной генной сети** ПОУГ с использованием **ANDSystem**.
- 4) Выявление структурно-функциональных особенностей ассоциативной генной сети ПОУГ: расчет показателя **центральности по посредничеству** (betweenness centrality) для каждой вершины ассоциативной генной сети ПОУГ.
- 5) **Поиск регуляторных путей** в ассоциативной генной сети ПОУГ с помощью функции **Pathway Wizard** системы ANDSystem.

Базы данных, содержащие информацию о связи полиморфизмов в генах с глаукомой

В настоящее время разработан ряд баз данных, в том числе OMIM, ClinVar, GWAS catalog, SNPedia, которые содержат экспериментально подтвержденные факты, касающиеся связи мутаций в генах с заболеваниями.

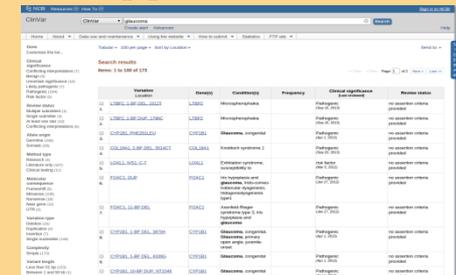
Домашняя страница базы данных OMIM



База данных **OMIM** (<http://www.omim.org/>) разрабатывается медицинским университетом Джонса Хопкинса (JHM, США) еще с 1960 годов. Изначально база включала информацию о мутациях в генах, определяющих развитие "Менделевских заболеваний", то есть являющихся результатом нарушения экспрессии единичного гена, которые наследуются согласно законам Менделя. На данный момент в базе данных содержится информация о 15013 генах и 4544 фенотипах.

База данных **ClinVar** (<http://www.ncbi.nlm.nih.gov/clinvar/>) разрабатывается Национальным центром биотехнологической информации NCBI (США) и содержит информацию о связи 125296 вариаций в 26376 генах с различными патологическими фенотипами человека.

Интерфейс базы данных



Домашняя страница базы данных SNPedia



База данных **SNPedia** (<http://www.snpedia.com/>) содержит информацию о связи 73661 мутаций в 1532 генах с 344 заболеваниями, извлеченную из научных публикаций в ручную.

Help!

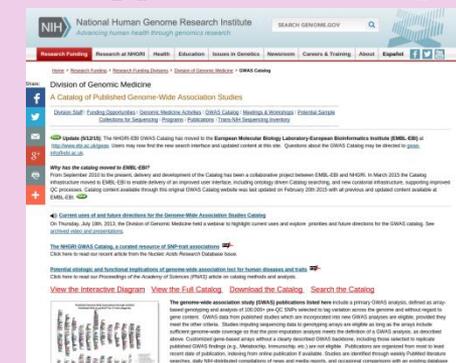
- look at the example rs1234
- learn more about SNPs
- browse
 - genes
 - genomes
 - genotypes
 - medicines
 - medical conditions
 - topics

Popular

- rs53776 in the oxytocin receptor influences social behavior and personality
- rs181815789 muscle performance
- rs17412 and rs4293558 can raise the risk of Alzheimer's disease by more than 10x
- rs1352 can influence baldness
- rs333 resistance to HIV
- rs1800497 in a dopamine receptor may influence the sense of pleasure
- rs1800007 determines red hair and sensitivity to anesthetics
- rs9939609 triggers obesity and type-2 diabetes
- rs682799 prevents weight gain from high fat diets
- rs17495174 green eye color and rs12913832 for blue eye color
- rs17903146 in 3% of the population greatly increases the risk of type-2 diabetes
- rs12255372 linked to type-2 diabetes and breast cancer
- rs1799971 makes alcohol cravings stronger
- rs17225931 determines earwax, sweating and body odor
- rs4680 varied cognitive effects

GWAS Catalog (<https://www.genome.gov/26525384>) разрабатывается Национальным Научно-исследовательским Институтом Генома Человека (NHGRI) и Европейским Институтом Биоинформатики (EMBL-EBI) с 2008 года. Каталог содержит информацию о связи мутаций в генах с фенотипами человека, извлеченную вручную из публикаций, посвященных полногеномному поиску ассоциаций (GWAS).

Интерфейс GWAS catalog



Поиск генов, ассоциированных с ПОУГ в компьютерных базах данных OMIM, ClinVar, SNPedia, GWAS catalog и базе знаний ANDCell системы ANDSystem.

Всего в анализе было найдено 96 генов, ассоциированных с ПОУГ, по данным разных источников информации.

Ген	Описание	Ген	Описание	Ген	Описание
ADRB2	adrenoceptor beta 2, surface	GPX3	lysyl oxidase-like 1	NSG1	optineurin
ANGPTL7	apolipoprotein E	GSTM1	lymphotoxin alpha	NTF4	angiopoietin-like 7
APOE	brain-derived neurotrophic factor	GSTT1	latent transforming growth factor beta binding protein 2	OPA1	vav 3 guanine nucleotide exchange factor
ASB10	caveolin 1, caveolae protein, 22kDa	HGF	matrix metalloproteinase 2	OPTC	thioredoxin reductase 2
BDNF	caveolin 2	HK2	matrix metalloproteinase 3	OPTN	cytochrome P450, family 46, subfamily A, polypeptide 1
CAV1	CD40 ligand	HOOK2	matrix metalloproteinase 9	PADI2	peptidyl arginine deiminase, type II
CAV2	CD44 molecule (Indian blood group)	HPGDS	methylenetetrahydrofolate reductase (NAD(P)H)	PCMTD1	zinc finger protein, FOG family member 2
CD40LG	CD79a molecule, immunoglobulin-associated alpha	HSPA1A	myocilin, trabecular meshwork inducible glucocorticoid response	PLEKHA7	opticin
CD44	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	HSPA1B	nitric oxide synthase 2, inducible	PON1	neuron specific gene family member 1
CD79A	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	IL1B	neurotrophin 4	SALL1	hematopoietic prostaglandin D synthase
CDKN1A	collagen, type XI, alpha 1	IL6	optic atrophy 1 (autosomal dominant)	SERPINE1	hook microtubule-tethering protein 2
CDKN2B	cytochrome P450, family 1, subfamily B, polypeptide 1	ISYNA1	SIX homeobox 6	SIX1	low density lipoprotein receptor-related protein 12
CDKN2B-AS1	endothelin 1	KCNA5	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	SIX6	inositol-3-phosphate synthase 1
CFH	estrogen receptor 2 (ER beta)	KIF2A	paraoxonase 1	SOD2	S1 RNA binding domain 1
CNTNAP4	fibronectin 1	KLK2	spalt-like transcription factor 1	SPP1	ELOVL fatty acid elongase 5
COL11A1	glaucoma 1, open angle, B (adult-onset)	LMX1B	serine/arginine-rich splicing factor 3	SRBD1	contactin associated protein-like 4
CYP1B1	glaucoma 1, open angle, C	LOC102724330	SIX homeobox 1	SRSF3	protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1
CYP46A1	glaucoma 1, open angle, D (adult-onset)	LOC105375094	superoxide dismutase 2, mitochondrial	ST18	WD repeat domain 36
EDN1	glutathione peroxidase 3	LOC730100	secreted phosphoprotein 1	TCEAL1	ankyrin repeat and SOCS box containing 10
ELOVL5	glutathione S-transferase mu 1	LOXL1	transforming growth factor, beta 2	TGFB2	pleckstrin homology domain containing, family A member 7
ESR2	glutathione S-transferase theta 1	LRP12	TIMP metalloproteinase inhibitor 1	TIMP1	NCK-associated protein 5
FN1	complement factor H	LTA	TIMP metalloproteinase inhibitor 2	TIMP2	glaucoma 1, open angle, H (adult-onset)
GLC1B	hepatocyte growth factor (hepapoietin A; scatter factor)	LTBP2	toll-like receptor 4	TLR4	glaucoma 1, open angle, I
GLC1C	hexokinase 2	MED12	tumor necrosis factor	TNF	glaucoma 1, open angle, K (juvenile-onset)
GLC1D	heat shock 70kDa protein 1A	MMP2	tumor protein p53	TFP53	glaucoma 1, open angle, J (juvenile-onset)
GLC1H	heat shock 70kDa protein 1B	MMP3	transferrin	TTR	uncharacterized LOC730100
GLC1I	interleukin 1, beta	MMP9	vav 2 guanine nucleotide exchange factor	TXNRD2	glaucoma 1, open angle, M (juvenile-onset)
GLC1J	interleukin 6	MTHFR	vascular endothelial growth factor C	VAV2	glaucoma 1, open angle, N (juvenile-onset)
GLC1K	potassium channel, voltage gated shaker related subfamily A, member 5	MYOC	neuropilin 1	VAV3	CDKN2B antisense RNA 1
GLC1M	kinesin heavy chain member 2A	NCKAP5	transcription elongation factor A (SII)-like 1	VEGFC	glaucoma 1, open angle, P
GLC1N	kallikrein-related peptidase 2	NOS2	suppression of tumorigenicity 18, zinc finger	WDR36	uncharacterized LOC102724330
GLC1P	LIM homeobox transcription factor 1, beta	NRP1	mediator complex subunit 12	ZFPM2	uncharacterized LOC105375094

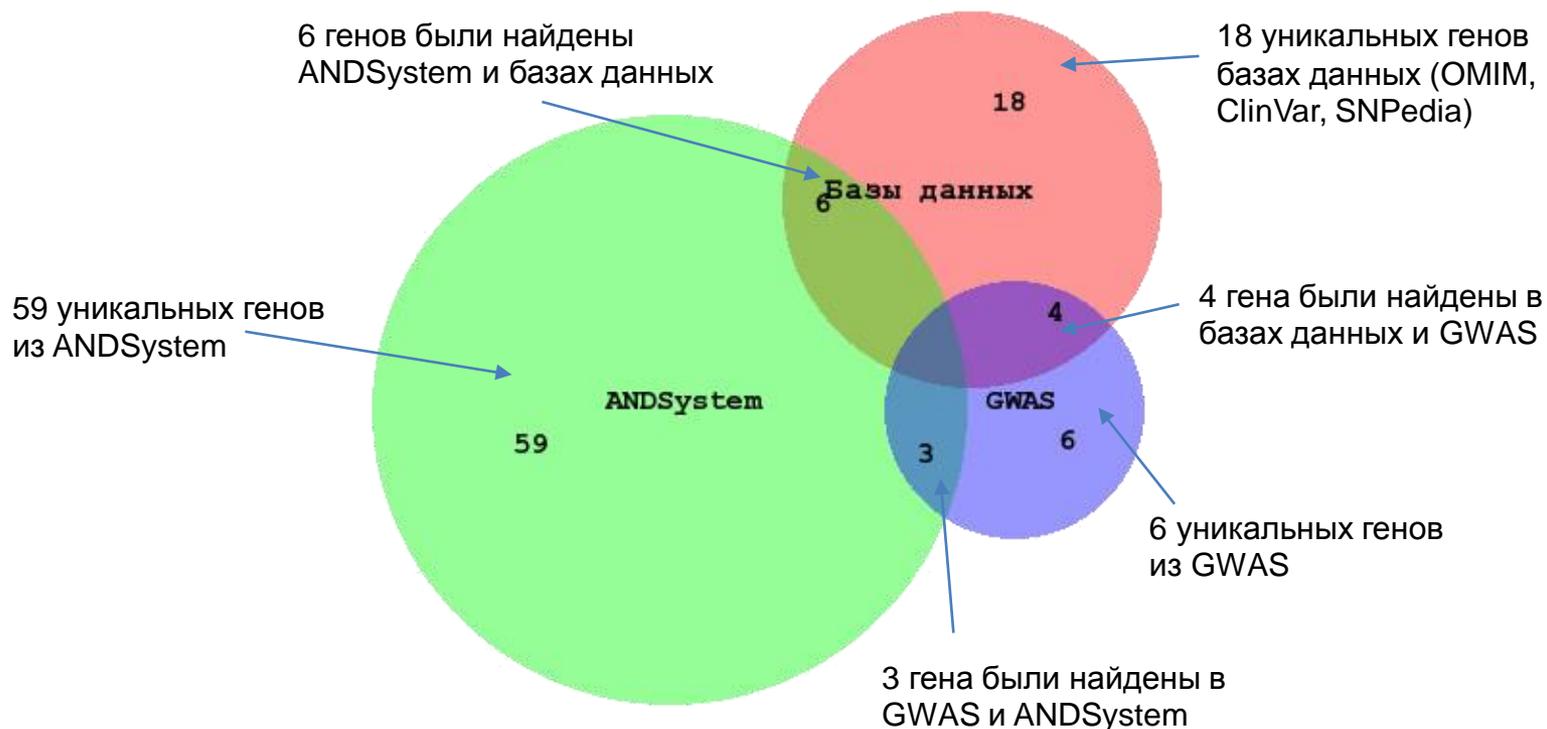
Наиболее полным источником информации для проведенного анализа оказалась база знаний ANDCell системы ANDSystem.

В ANDCell было идентифицировано 68 генов, ассоциированных с ПОУГ.

В базах данных (OMIM, ClinVar, SNPedia) было найдено 28 генов, ассоциированных с ПОУГ.

GWAS catalog (genome-wide association study, содержащем результаты полногеномного анализа полиморфизмов человека) дали информацию о 13 генах. Следует отметить, что один и тот же ген мог быть идентифицирован в нескольких источниках.

Пересечение списков генов, ассоциированных с ПОУГ, составленных на основании информации баз данных, геномных исследований GWAS и системы ANDSystem.



Основные группы сверхпредставленных Gene Ontology биологических процессов в наборе генов, ассоциированных с ТПОУГ, выявленные системой DAVID.

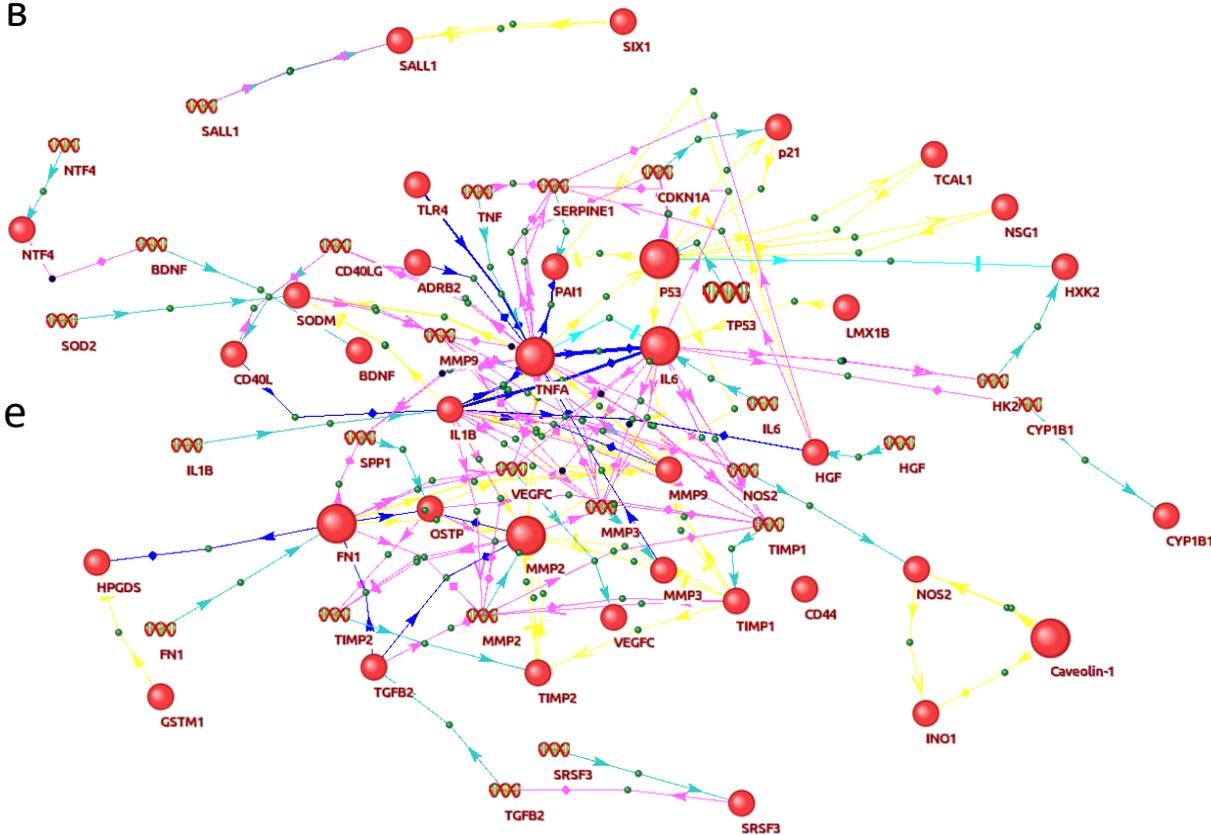
Gene ontology биологические процессы	Гены	p-value
Организменный уровень		
Сенсорное восприятие	CYP1B1, OPA1, COL11A1, WDR36, SIX6	0.02
Регуляция аксоногенеза	APOE, SPP1, NRP1	0.03
Васкулярные процессы	APOE, ADRB2, EDN1, CAV1, SOD2, CAV2	1.3E-5
Гемостаз	IL6, SERPINE1, CD40LG	0.01
Морфогенез трубчатых структур	CD44, EDN1, NRP1, SIX1	8.7E-6
Развитие уrogenитальной системы	BDNF, CD44, SIX1, SALL1	0.02
Развитие мышечной ткани	CAV1, TGFB2, SIX1, COL11A1, CAV2, ZFPM2	4.2E-4
Развитие мезенхимы	EDN1, HGF, TGFB2	0.03
Иммунные процессы	TP53, TNF, IL6, IL1B, CDKN1A, CD40LG, CD79A	0.02
Клеточный уровень		
Регуляция клеточной смерти	TP53, TNF, APOE, IL6, MMP9, IL1B, BDNF, TLR4, CDKN1A, ADRB2, ESR2, CD44, SOD2, HGF, CD40LG, LTA, HSPA1A, TGFB2, HSPA1B, VAV2, VAV3	2.4E-8
Регуляция митоза	TNF, IL1B, EDN1, LTA, CAV2	2.1E-4
Миграция клеток	TNF, IL6, MMP9, IL1B, ESR2, CD44, EDN1, FN1, VEGFC, TGFB2, NRP1, SIX1, CAV2, LMX1B, VAV2, VAV3	0.001
Процессы метаболизма коллагена	MMP2, MMP3, MMP9	0.004
Регуляция организации цитоскелета	EDN1, CAV1, CAV2	0.02
Молекулярный уровень		
Регуляция активности транскрипционных факторов	TP53, TNF, IL6, IL1B, TLR4, ADRB2, ESR2, MED12, SIX1, LMX1B, ZFPM2, TCEAL1, ST18	0.05
Регуляция фосфорилирования	TNF, APOE, IL6, IL1B, TLR4, CDKN1A, ADRB2, EDN1, CAV1, HGF, VEGFC, CDKN2B, TGFB2, HK2, CAV2, VAV2, VAV3	1.3E-7
Регуляция транспорта белков	TP53, TNF, IL6, IL1B, CD40LG, LTA, VEGFC, TGFB2, OPTN, LTBP2	3.1E-5
Ядерный транспорт	TP53, TNF, TGFB2	0.01
Гомеостаз ионов	TP53, APOE, IL1B, EDN1, CAV1, SOD2, KCNA5	0.004
Регуляция активности киназ	APOE, IL1B, CDKN1A, CAV1	0.01
Регуляция протеолиза	TNF, APOE, IL1B, TIMP1	0.002
Метаболические процессы с участием кофакторов	GSTT1, SOD2, GPX3	0.009
Метаболизм жирных кислот	EDN1, HPGDS, ELOVL5	0.06
Сборка белковых комплексов в клетке	ADRB2, CAV1, CAV2	0.02
Не классифицированные	MTHFR, GSTM1, PON1, CFH, NOS2, TTR, TIMP2, MYOC, LOXL1, CDKN2B-AS1, CYP46A1, KLK2, SRSF3, NTF4, PADI2, TXNRD2, KIF2A, OPTC, PLEKHA7, ANGPTL7, ISYNA1, HOOK2, ASB10, CNTNAP4, NSG1, NCKAP5, LRP12, SRBD1, GLC1C, PCMTD1, GLC1B, GLC1M, GLC1N, GLC1I, GLC1H, GLC1P, GLC1J, GLC1K, GLC1D, LOC730100, LOC105375094, LOC102724330	

Реконструкция ассоциативной генной сети ПОУГ с использованием ANDSystem.

Важность генов для функционирования биологических процессов может быть оценена исходя из величины их центральности в генных сетях, описывающих данные процессы (Yu et al., 2007; Özgür et al., 2008). Для анализа центральности генов, ассоциированных с ПОУГ, с использованием ANDSystem (Ivanisenko et al., 2015) была реконструирована ассоциативная генная сеть ПОУГ.

Сеть содержала 1289 связей 14 разных типов, представленных в системе ANDSystem, включая регуляторные связи, коэкспрессию, белок-белок взаимодействия и др. Большое количество связей, присутствующих в генной сети ПОУГ, представляет затруднение для ее графического представления, поэтому для визуализации сети была построена (отфильтрована) подсеть, включающая только регуляторные типы связей.

Ассоциативная регуляторная сеть ПОУГ.

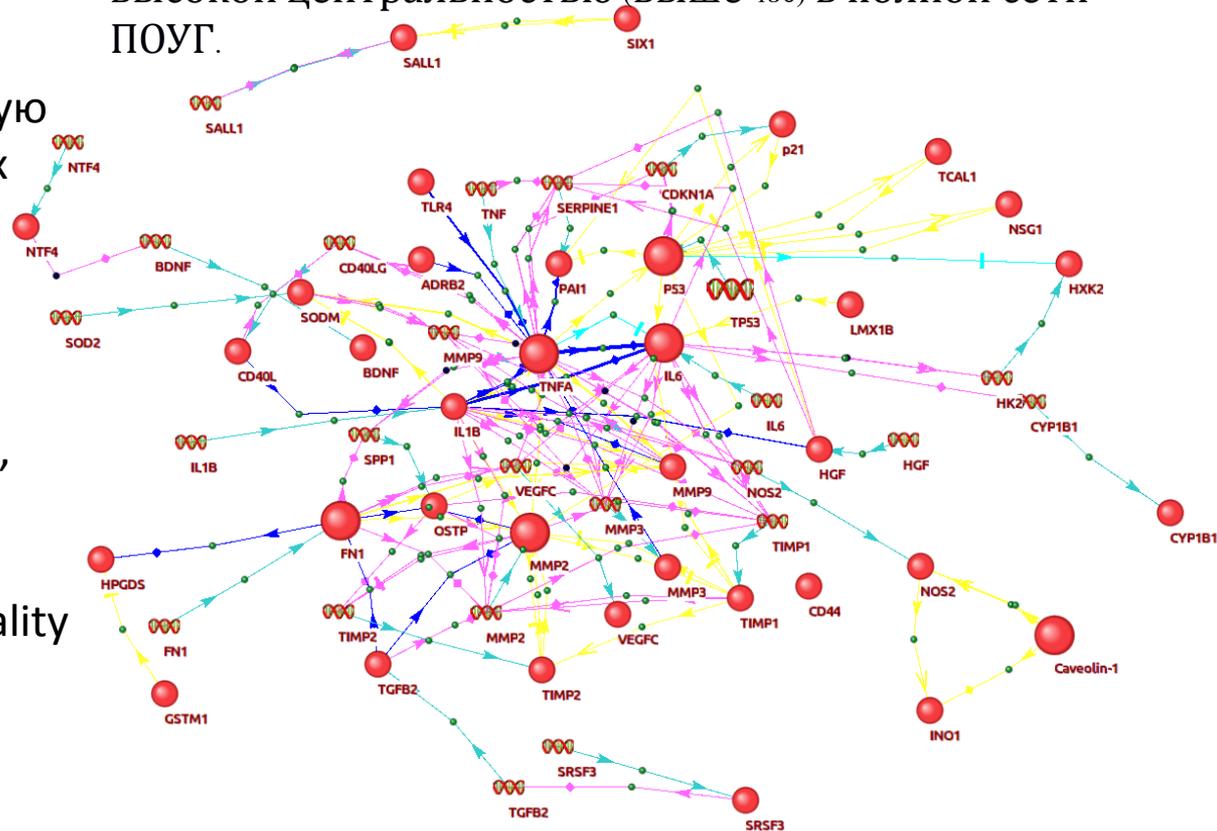


Расчет показателя *betweenness centrality* для каждой вершины ассоциативной геной сети ПОУГ.

Для каждой из вершин этой сети был рассчитан показатель *betweenness centrality* (BC). Расчет показателя *betweenness centrality* для каждой вершины анализируемой ассоциативной геной сети проводился с помощью пакета *igraph*, реализованного на языке программирования R (Csardi, Nepusz, 2006). Средняя центральность для вершин составила $BC=113.46$.

Белок p53 имеет наибольший показатель *betweenness centrality*. Известно, что p53 является хабом и играет важную роль во многих биологических процессах. Наиболее тесно функция p53 связана с апоптозом. Кроме того гены, вовлеченные в Gene Ontology категорию «apoptotic process», имеют статистически значимо более высокий средний показатель *betweenness centrality* равный 174.73, чем средний показатель всех вершин сети равный 88.84 ($p\text{-value}<10^{-4}$).

Ассоциативная регуляторная сеть ПОУГ. Крупными иконками выделены белки и гены, обладающие высокой центральностью (выше 450) в полной сети ПОУГ.



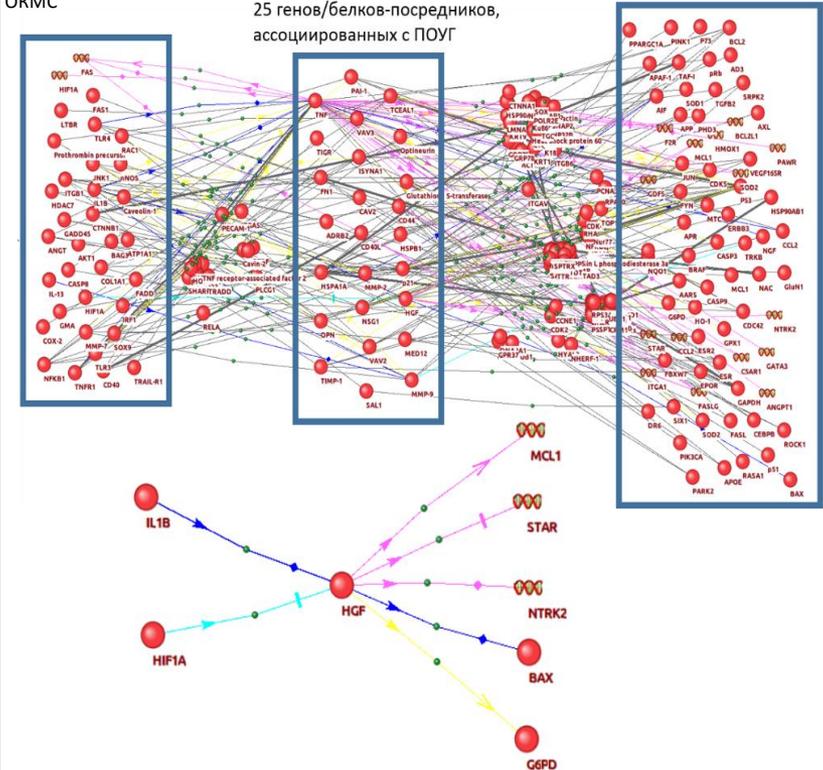
Реконструкция молекулярно-генетических механизмов взаимодействия процессов ответа клетки на механический стресс (ОКМС) и нейронального апоптоза при первичной открытоугольной глаукоме с помощью функции Pathway Wizard системы ANDSystem.

- Известно, что интерлейкин-1 β (IL1B), который вовлечен в ОКМС, активирует секрецию белка HGF, ассоциированного с ПОУГ (Baroni et al., 1998; Paduch et al., 2010).
- Другой белок, индуцируемый гипоксией фактор 1 α (HIF1A) и также участвующий в ОКМС, способен повышать стабильность HGF (Chu et al., 2009).
- В свою очередь HGF оказался связанным с двумя белками и тремя генами, ассоциированными с нейрональным апоптозом.
- Согласно данным работ (Au et al., 2009; Lu et al., 2015), HGF активирует экспрессию генов, кодирующих рецептор 2 типа нейротрофической тирозинкиназы (NTRK2) и белок Mcl1.
- Экспрессия другого гена стероидогенного острого регуляторного белка (STAR), вовлеченного в нейрональный апоптоз, отрицательно регулируется HGF (Taniguchi et al., 2004).
- HGF также задействован в регуляции активности белка G6PD (Shao et al., 1996) и транслокации белка Bax (Nakagami et al., 2002), участвующих в нейрональном апоптозе.

Белки/гены, вовлеченные в ОКМС

25 генов/белков-посредников, ассоциированных с ПОУГ

Белки/гены, вовлеченные в нейрональный апоптоз



Ассоциативная сеть, описывающая потенциальные взаимодействия между ОКМС и нейрональным апоптозом через молекулярно-генетические цепочки. а) Общая сеть, включающая 25 генов/белков-посредников, ассоциированных с ПОУГ. б) Пример цепочек, проходящих через фактор роста гепатоцитов (HGF), выступающий в качестве посредника, ассоциированного с ПОУГ.

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). *Biotechnología 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 Seleksii*=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 Mozgovaya, 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.

Что можно поймать в генные 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

Интерпретация данных высокопроизводительных экспериментов

Исследование структурных и функциональных характеристик генных сетей

Выявление хабных белков, функциональных модулей, путей передачи сигнала

Моделирование биологических процессов

Спасибо за внимание!