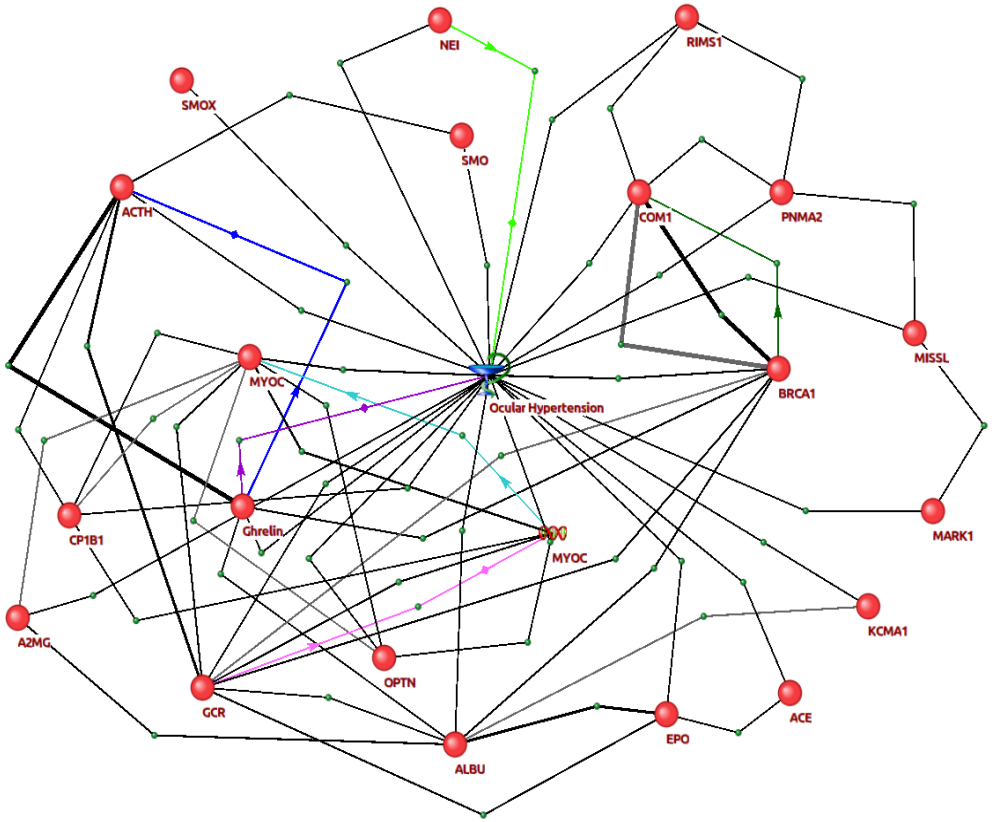
Генные сети: что в них можно поймать?

Сайк О.В., Тийс Е.С.,
Петровский Е.Д.,
Попик О.В.,
Иванисенко Н.В.,
Иванисенко Т.В.,
Деменков П.С.,
Иванисенко В.А.



Лаборатория компьютерной протеомики, ИЦиГ СО РАН.

Новосибирск, 2015

Генная сеть – группа координированно функционирующих генов, взаимодействующих друг с другом как через свои первичные продукты (РНК и белки), так и через разнообразные метаболиты и другие вторичные продукты функционирования генных сетей, которая контролирует какой-либо фенотипический признак организма.

Компоненты генной сети:

1) группа координированно экспрессирующихся генов, составляющая ядро сети;
2) белки, кодируемые этими генами;
3) низкомолекулярные компоненты (гормоны и другие сигнальные молекулы, энергетические компоненты, метаболиты);
4) связи между участниками сети (в том числе отрицательные и положительные обратные связи).

По Колчанову Н.А. с соавт. ГЕННЫЕ СЕТИ // Вавиловский журнал генетики и селекции, 2013, Т. 17:4/2.

A gene network is a mixed graph G : =(V ,U,D) over a set V of nodes, corresponding to gene activities, with unordered pairs U, the undirected edges, and ordered pairs D, the directed edges.

Pinna A., Soranzo N., De La Fuente A. From knockouts to networks: establishing direct cause-effect relationships through graph analysis //PloS one. – 2010. – T. 5. – №. 10. – C. e12912.

Gene network is a graphical illustration for exploring the functional linkages and the potential coordinate regulations of genes.

Wang et al.Gene Network Exploration of Crosstalk between Apoptosis and Autophagy in Chronic Myelogenous Leukemia // BioMed Research International, 2014.

A gene network is a directed labelled graph, where each node represents a gene and each arc represents a relation between the genes.

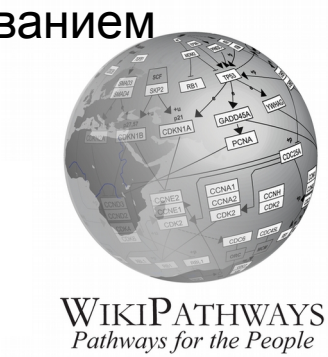
Rung J. et al. Building and analysing genome-wide gene disruption networks //Bioinformatics. – 2002. – T. 18. – №. suppl 2. – C. S202-S210.

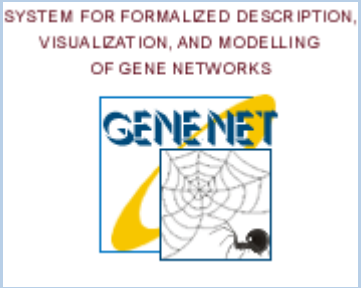
A gene network is a collection of effective interactions, describing the multiple ways through which one gene affects all the others to which it is connected.

Zhu Y., Pan W., Shen X. Support vector machines with disease-gene-centric network penalty for high dimensional microarray data //Statistics and its interface. – 2009. – T. 2. – №. 3. – C. 257.

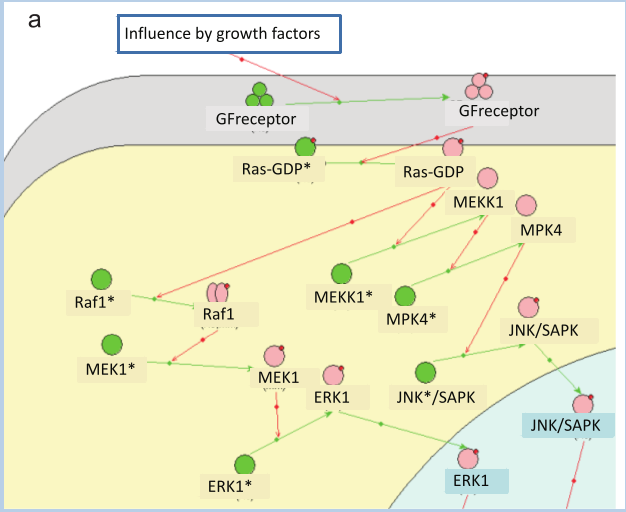
Методы реконструкции генных сетей

1. Ручная реконструкция генных сетей экспертами с использованием специальных программных средств – редакторов генных сетей.





ИЦиГ СО РАН, 1998, N.A. Kolchanov, E.A. Ananko, N.L. Podkolodny, I.L. Stepanenko, E.V. Ignatieva, O.A. Podkolodnaya et al.



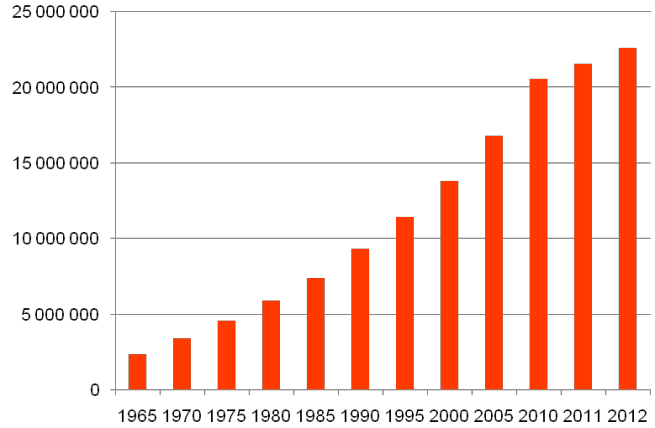
МАР-киназный путь передачи сигнала в ядро клетки, контролирующий процесс клеточного деления, активируемый ростовыми факторами.

2. Автоматическое извлечение знаний о молекулярно-генетических взаимодействиях различного типа из текстов научных публикаций и баз данных компьютерными методами (методы text-mining).

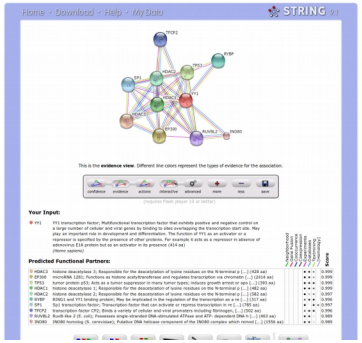
По Колчанову Н.А. с соавт. ГЕННЫЕ СЕТИ // Вавиловский журнал генетики и селекции, 2013, Т. 17:4/2.

Системы для автоматической реконструкции генных сетей

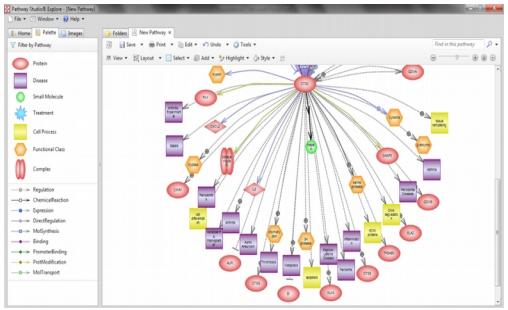
Рост числа публикаций в системе PubMed по годам (более 24 млн статей на 2015 год)



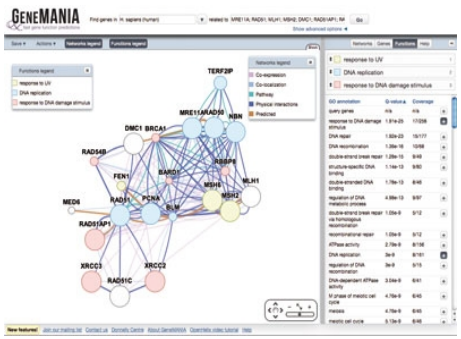
String



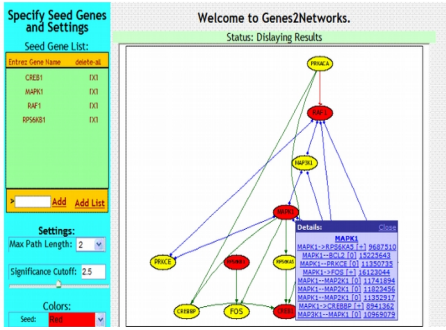
Pathway Studio



GeneMANIA

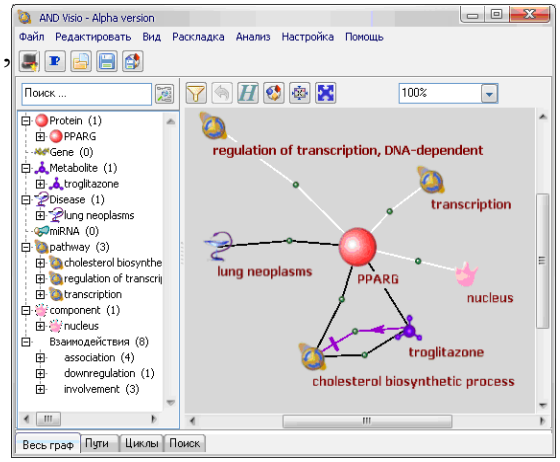


Genes2Networks

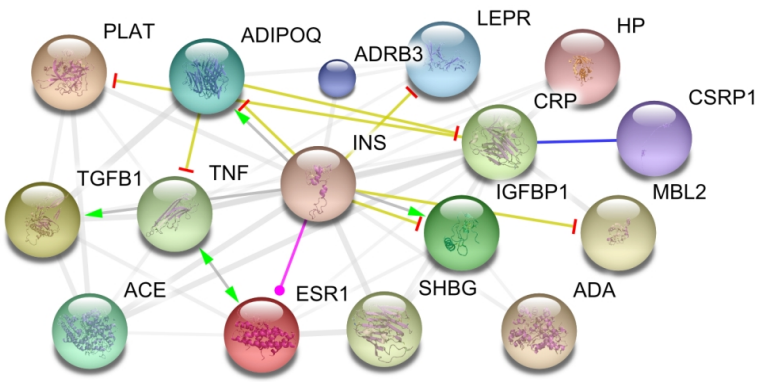


ANDSystem

(Иванисенко В.А., Деменков П.С. с соавт., ИциГ СО РАН).

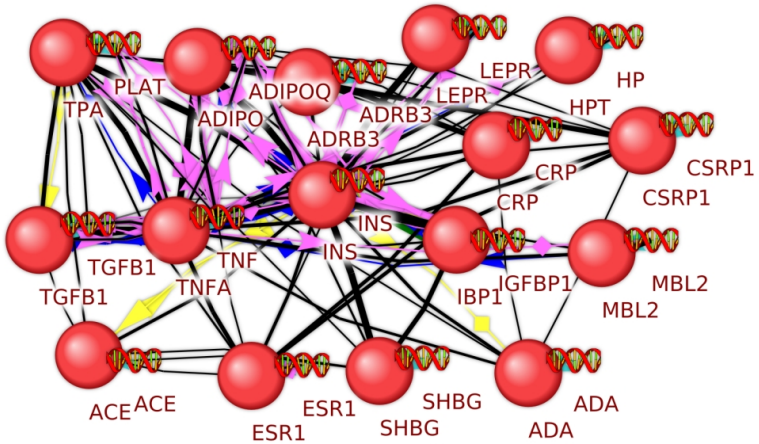


Парадокс недостатка полноты информации о взаимодействиях при избыточности данных о взаимодействиях



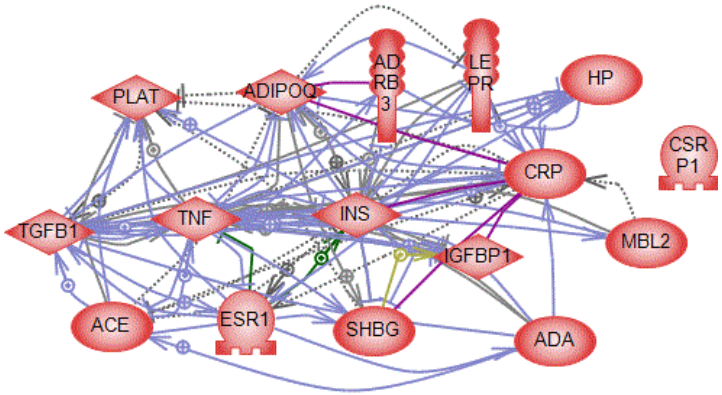
STRING

- Activation
- Fuctional link
- Binding
- Post-translational modification
- Expression with inhibition



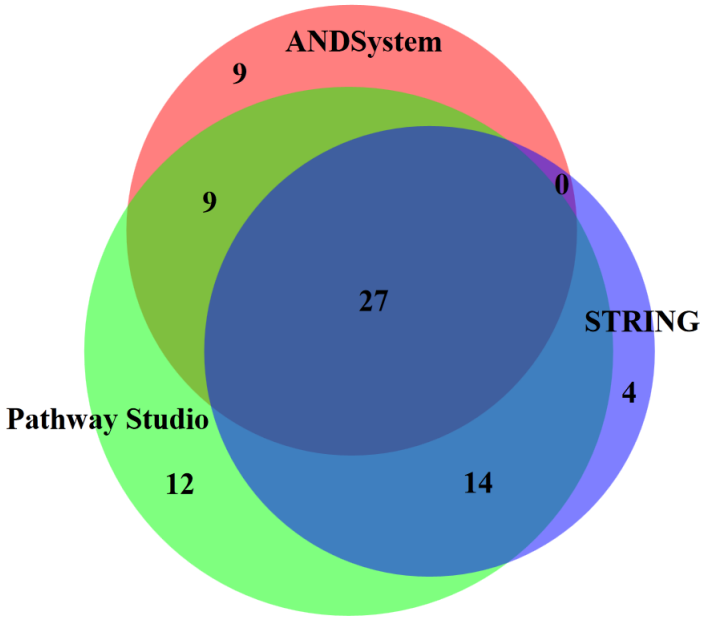
ANDSystem

- Activity downregulation
- Activity regulation
- Activity upregulation
- Catalysis
- Coexpression
- Cleavage
- Expression
- Degradation downregulation
- Degradation upregulation
- Expression downregulation
- Expression upregulation
- Expression regulation
- Transport regulation
- Association



Pathway Studio

- ProtModification
- MolTransport
- PromoterBinding
- Expression
- Regulation
- Binding



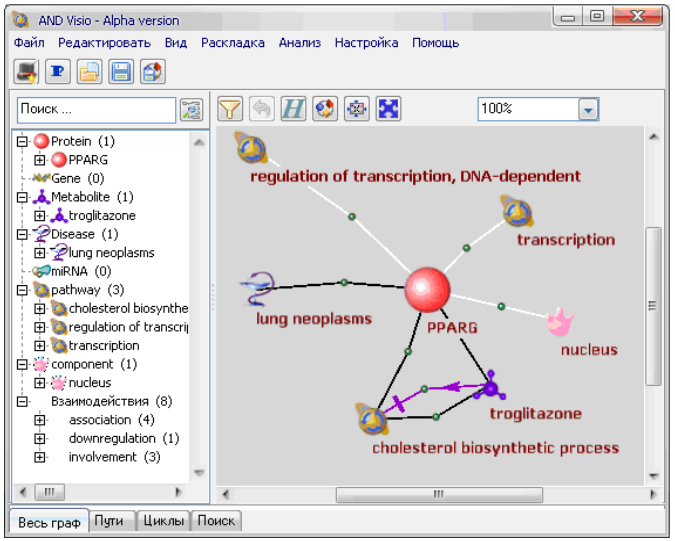
Пересечение связей между генами, реконструированными программами ANDSystem, Pathway Studio и STRING.

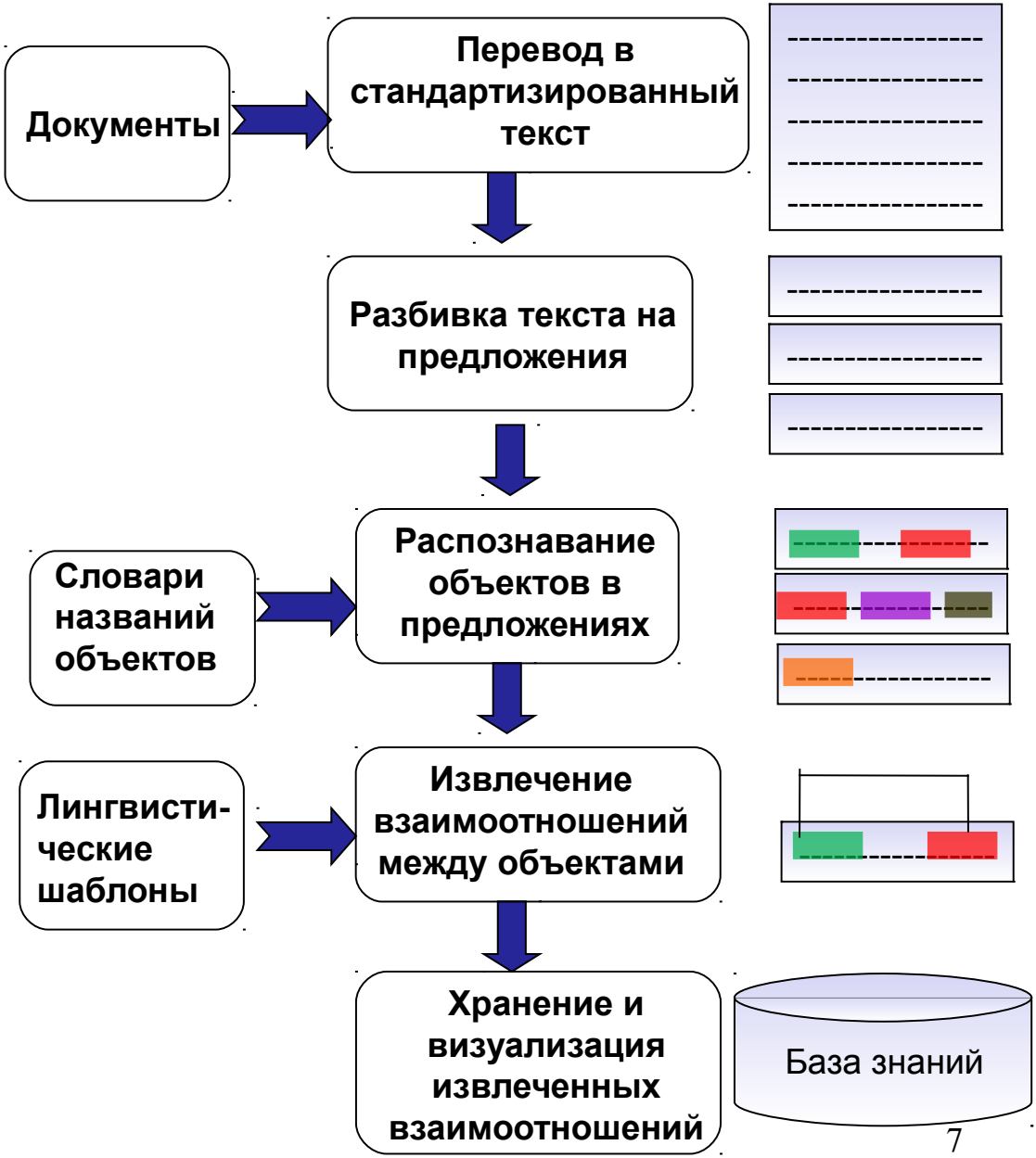
Text-mining. Система ANDSystem.

Проанализировано 23 млн рефератов из системы PubMed.

Выявлено 39 740 237 взаимодействий

Интерфейс модуля визуализации ANDVisio



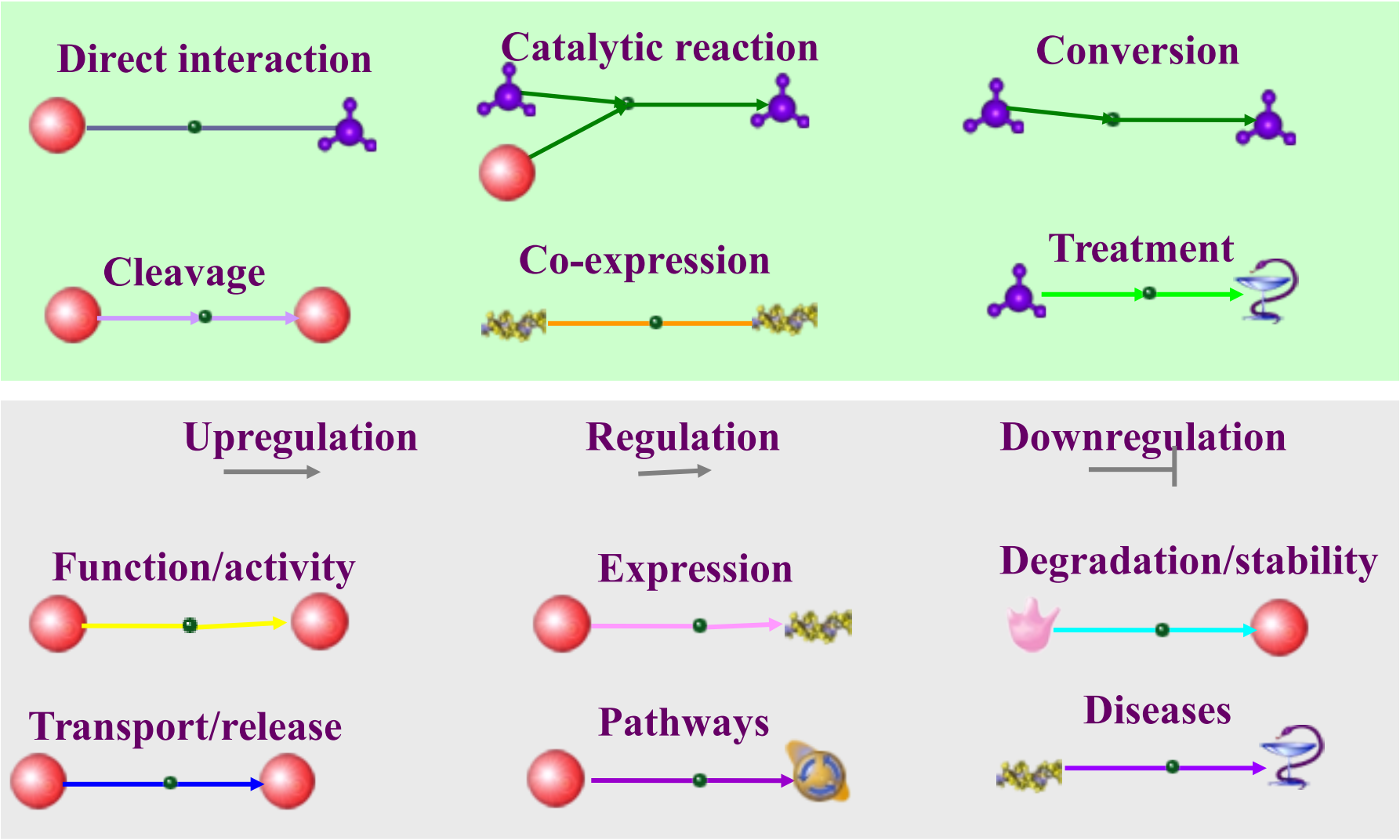


Статистика системы ANDSystem

Protein names	550 657
Gene names	7 946 479
Disease names	4 582
Metabolite names	33 159
Organism names	17 459
MicroRNA names	4 515
Pathway names	21 935
Cell component names	3 017
Cell/tissue names	2 500

Controlled vocabulary of tissues (UniProt, SwissProt Protein Knowledgebase)

Онтология молекулярно-генетических взаимодействий



Оценка точности информации, содержащейся в базе данных ANDCell

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

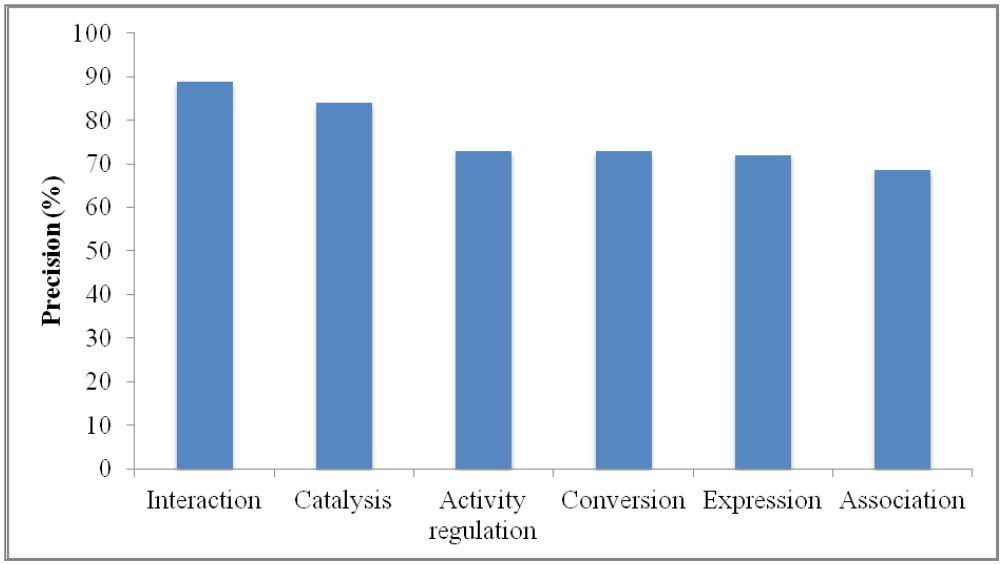
$N_{correct}$ – число верно распознанных взаимодействий.

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%).

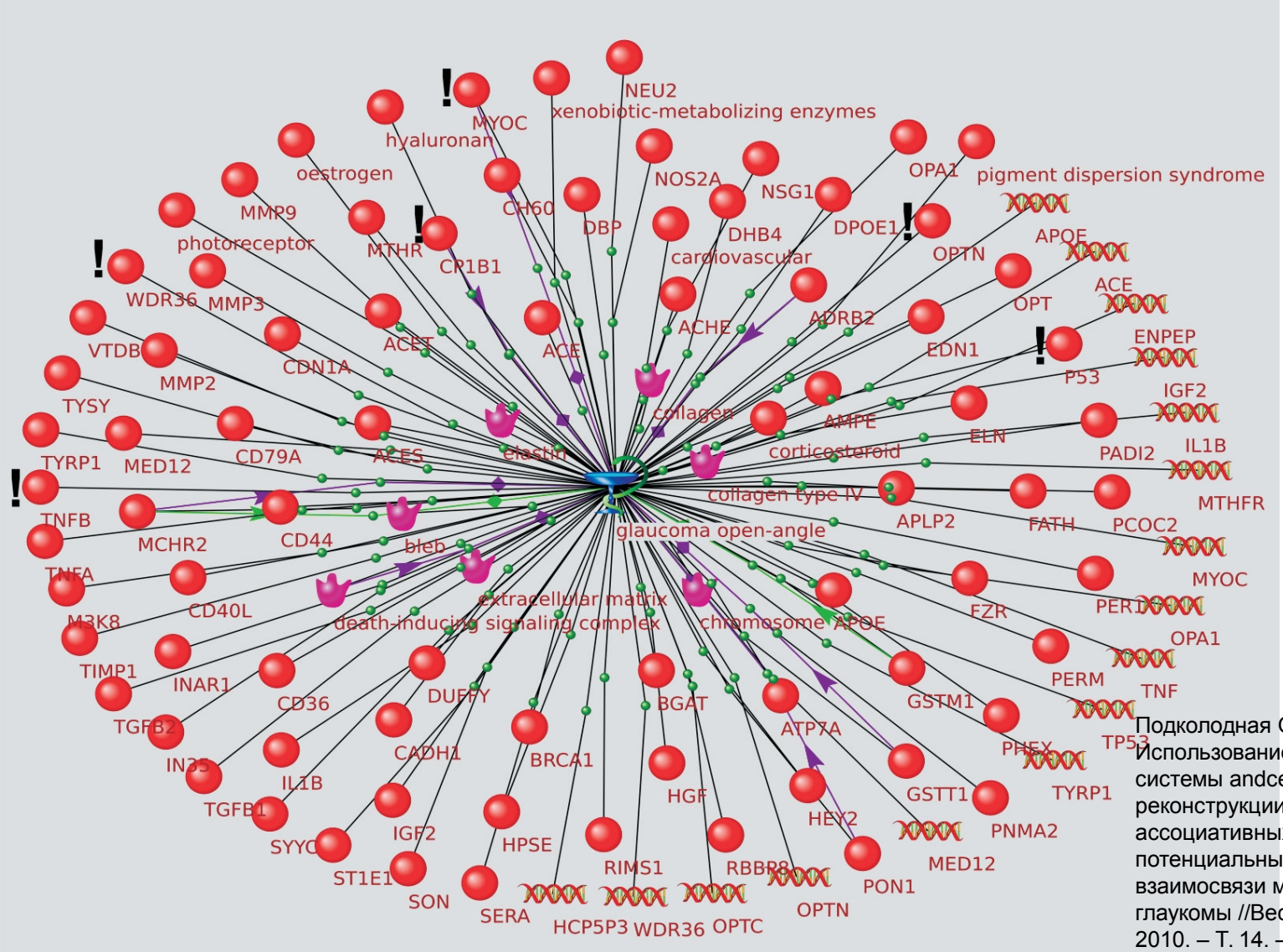
В среднем точность составила 76.5%.



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

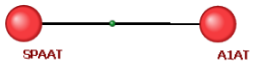
Исследование механизмов формирования фенотипических признаков (в том числе заболеваний)

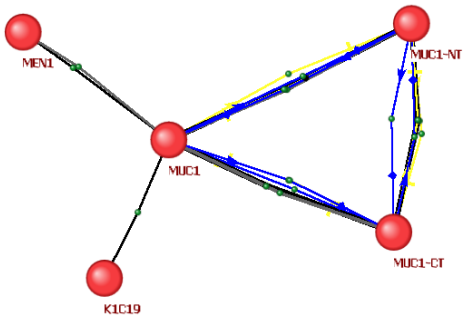
Ассоциативная сеть первичной открытоугольной глаукомы



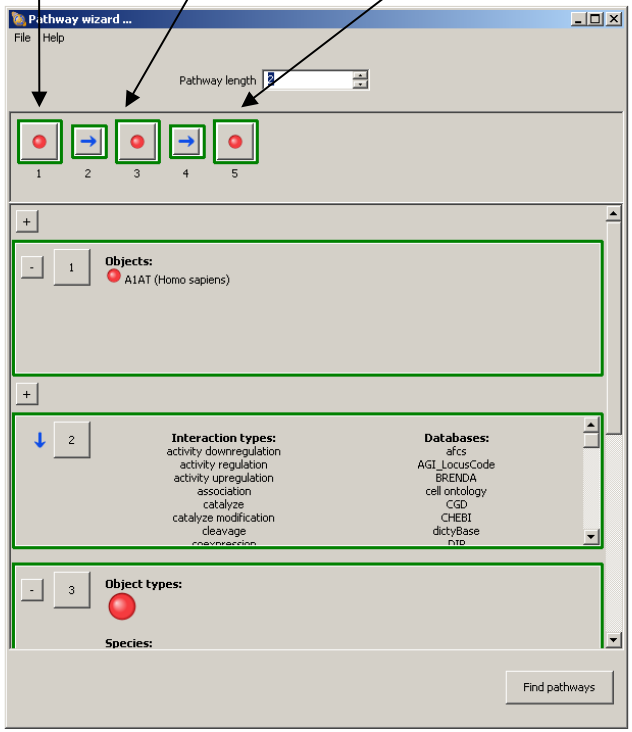
Подколодная О. А. и др. Использование компьютерной системы andcell для реконструкции и анализа ассоциативных сетей потенциальных механизмов взаимосвязи миопии и глаукомы //Вестник ВОГиС. – 2010. – Т. 14. – №. 1.

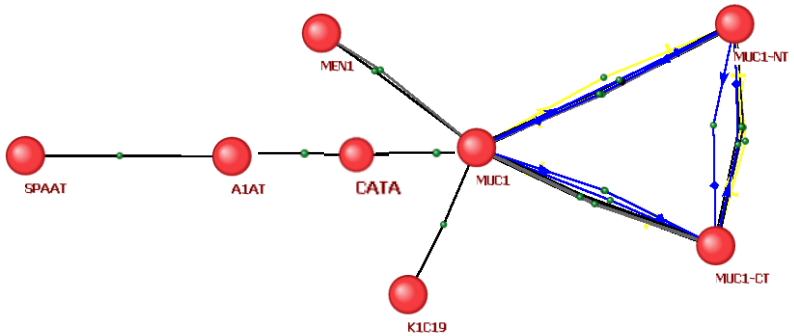
Поиск путей передачи сигнала между 2мя биологическими сетями, с помощью инструмента «Pathway Wizard» системы ANDSystem.



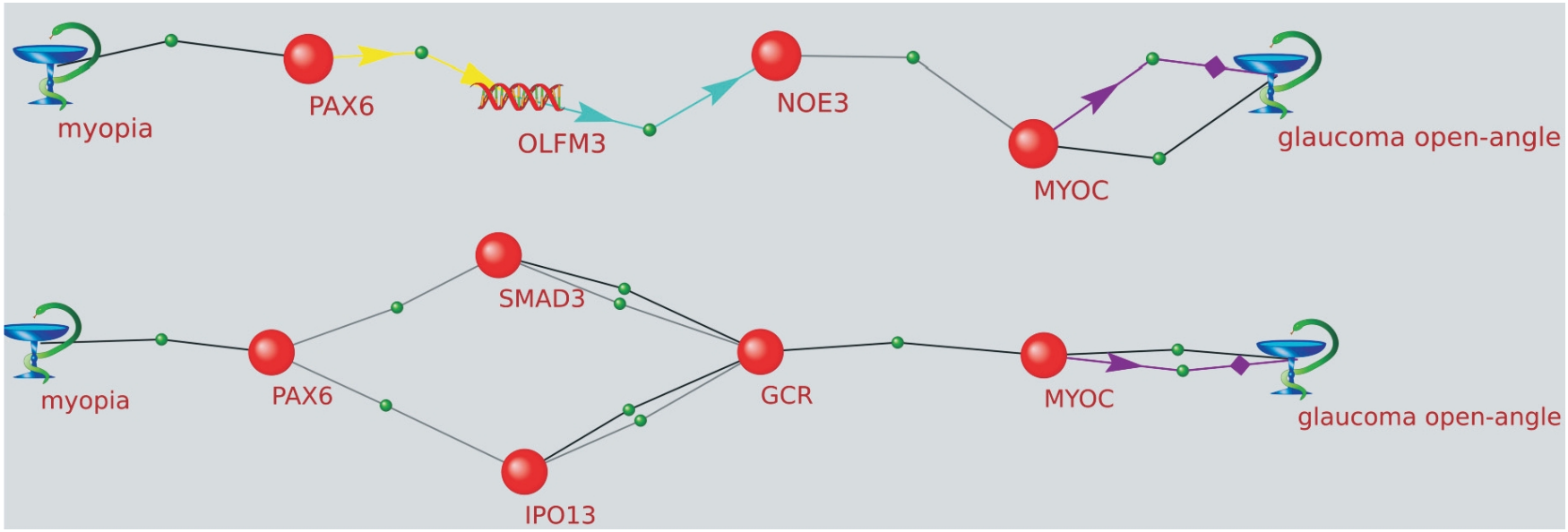


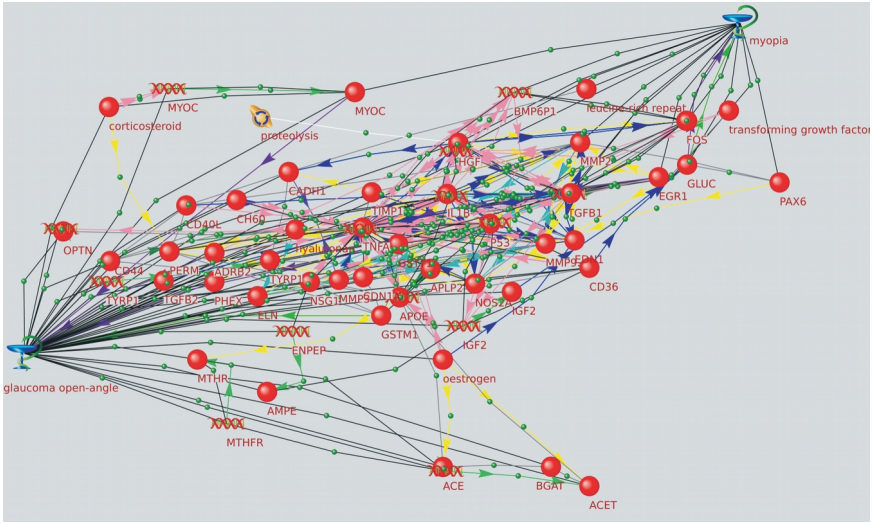
A1AT Any protein MUC1





Молекулярно-генетические пути, ассоциированные с открытоугольной глаукомой и миопией.





Подколодная О. А. и др. Использование компьютерной системы andcell для реконструкции и анализа ассоциативных сетей потенциальных механизмов взаимосвязи миопии и глаукомы //Вестник ВОГиС. – 2010. – Т. 14. – №. 1.

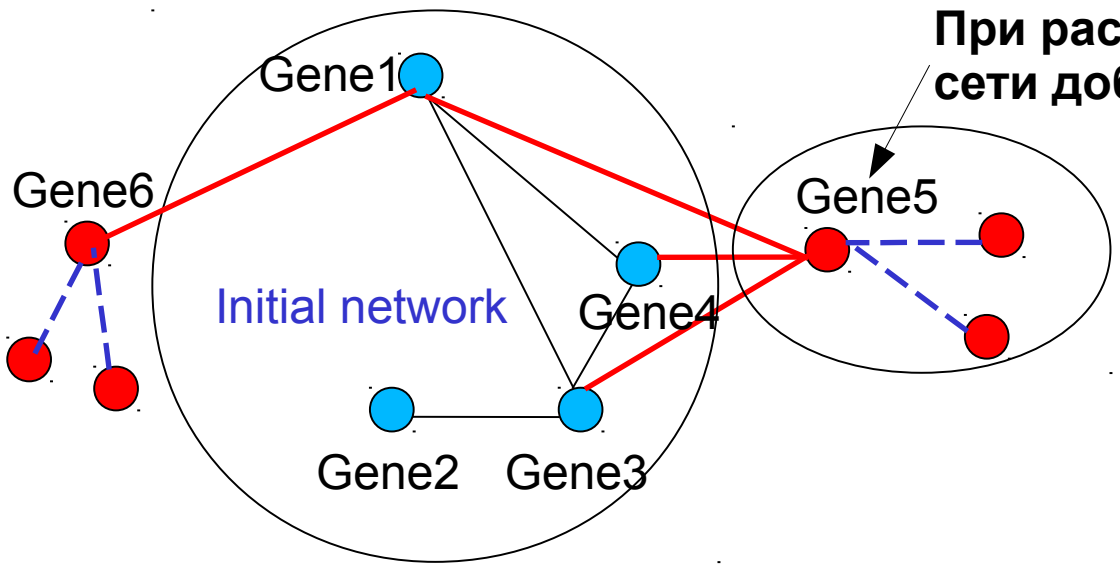
Расширение генной сети с учетом специфичности и центральности

Специфичность гена к реконструируемой сети:

$S(gene_i)$= Число взаимодействий i-ого гена с генами из начальной сети / Число всех известных связей i-ого гена

Центральность гена в реконструируемой сети:

$C(gene_i)$= Число взаимодействий i-ого гена с генами из начальной сети / Число генов в начальной сети



При расширении к начальной сети добавиться ген 5.

S(gene5)=3/5
C(gene5)=3/4

S(gene6)=1/3
C(gene6)=1/4

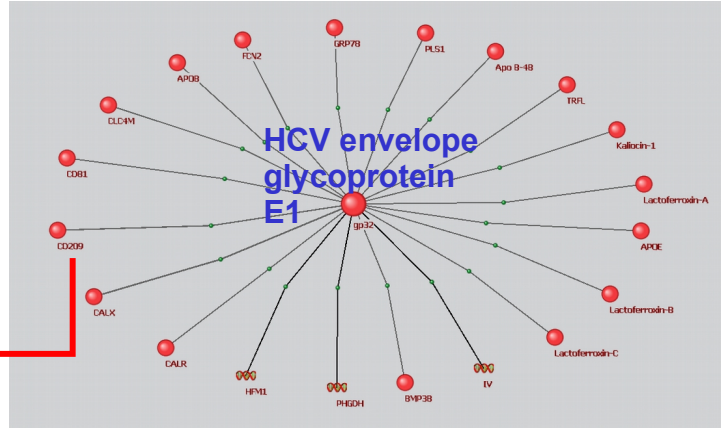
При пороге центральности и специфичности генов-кандидатов равном ½

Ген 5 будет добавлен в расширенную сеть, так как его специфичность и центральность превосходят порог. Ген 6 – добавлен не будет.

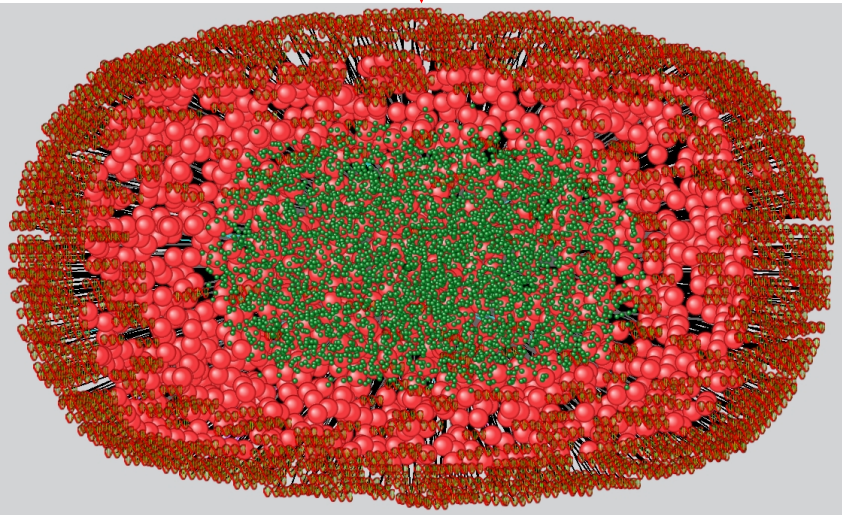
Тийс Е.С., Деменков П.С., Иванисенко В.А., ИциГ СО РАН

Расширение сети взаимодействий белка E1 вируса гепатита C и белков человека.

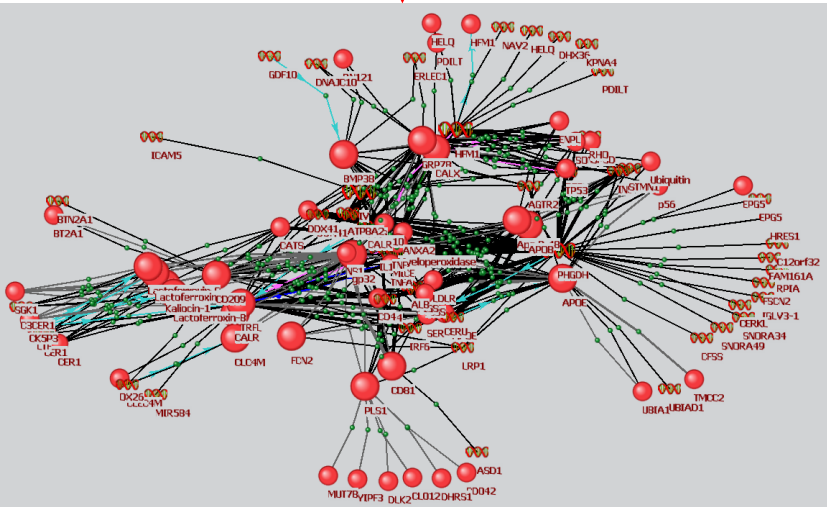
В расширенную сеть добавлены все гены/белки человека, взаимодействующие напрямую с генами/белками из исходной сети.



В расширенную сеть были добавлены только гены/белки, значения специфичности и центральности которых превыщают заданные пороги (C>0.2 or S>0.2).

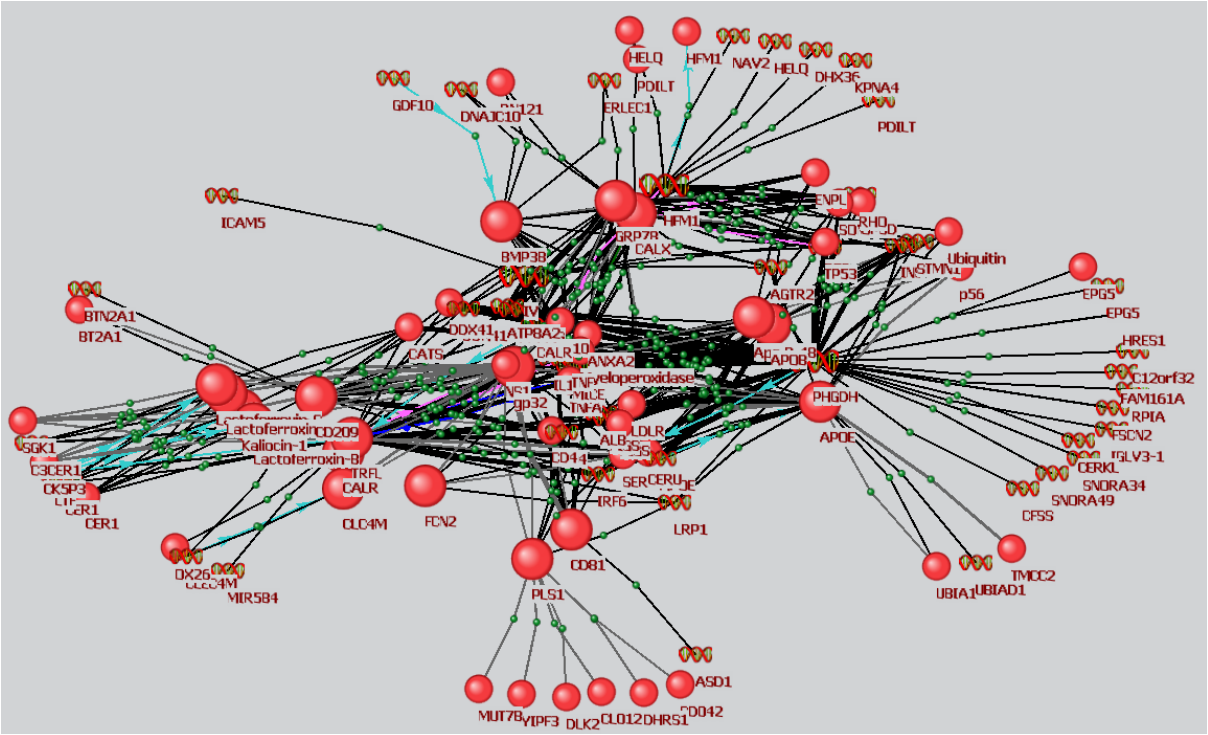


Галактика
1536 белков
1217 генов



54 белков
49 генов

Анализ сверхпредставленных GO биологических процессов в сети белка E1 вируса гепатита C, расширенной с использованием критериев центральности и специфичности



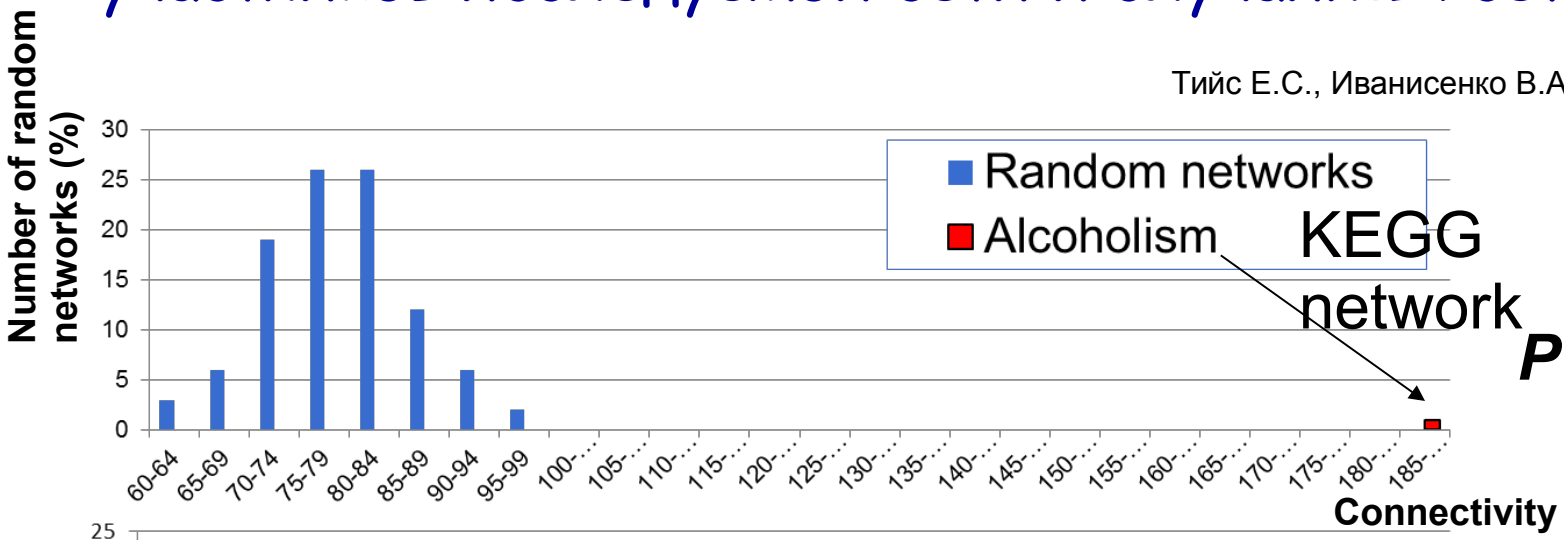
Белок E1 играет важную роль при входе вирусных частиц в клетку, обладает активностью для связывания с поверхностными белками клетки хозяина.

Gene Ontology biological process	PValue
GO:0043066 negative regulation of apoptosis	3.5E-7
GO:0060548 negative regulation of cell death	4.1E-7
GO:0019059 initiation of viral infection	6.1E-5
GO:0019062 virion attachment to host cell surface receptor	1.8E-4

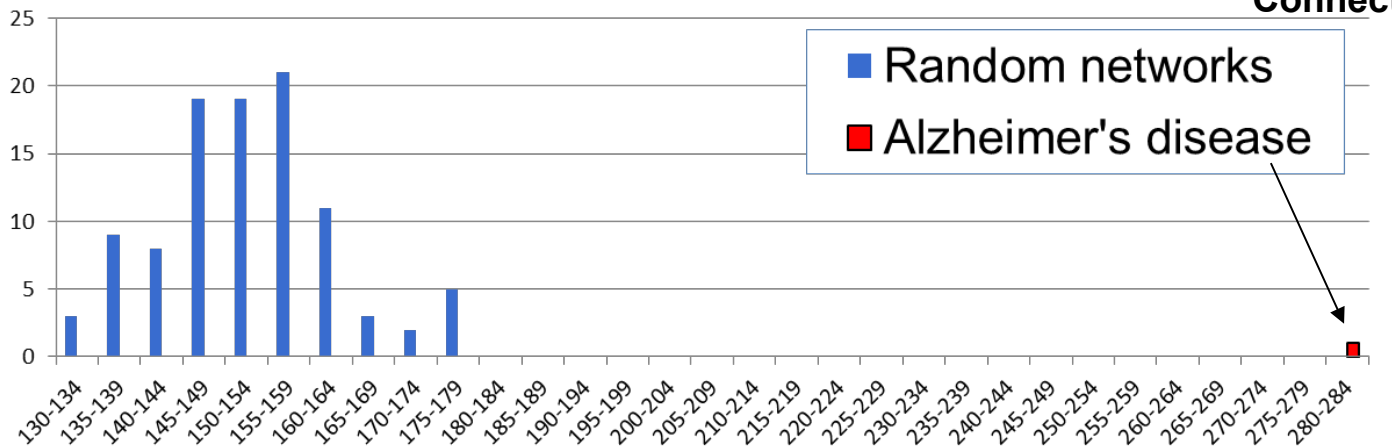
****

Выявление статистически значимого отличия связности участников исследуемой сети и случайных сетей.

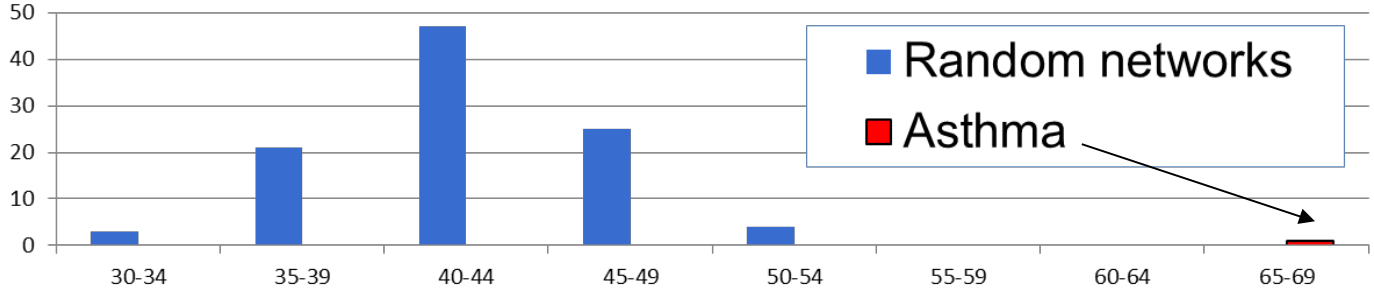
Тийс Е.С., Иванисенко В.А., ИциГ СО РАН



P<0.001



P<0.001



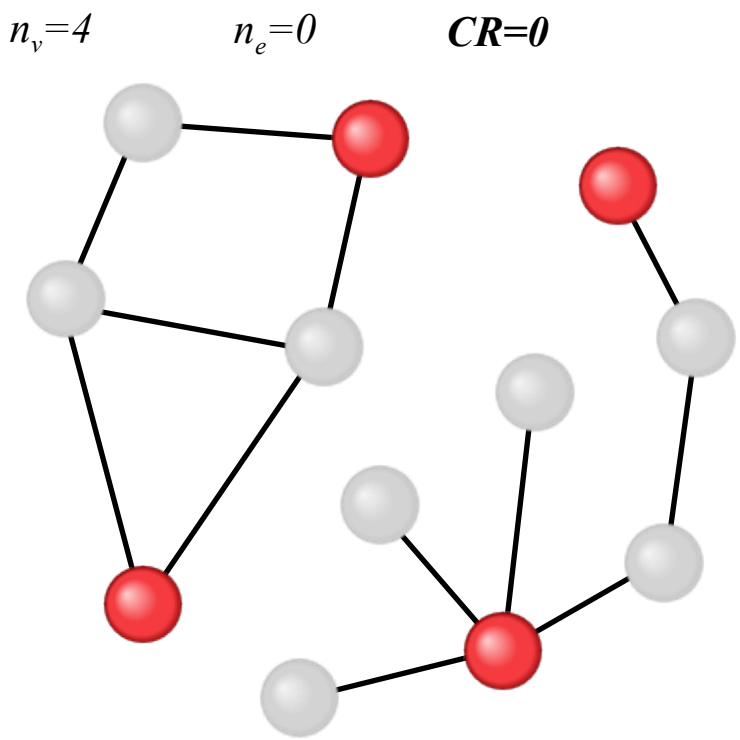
P<0.01

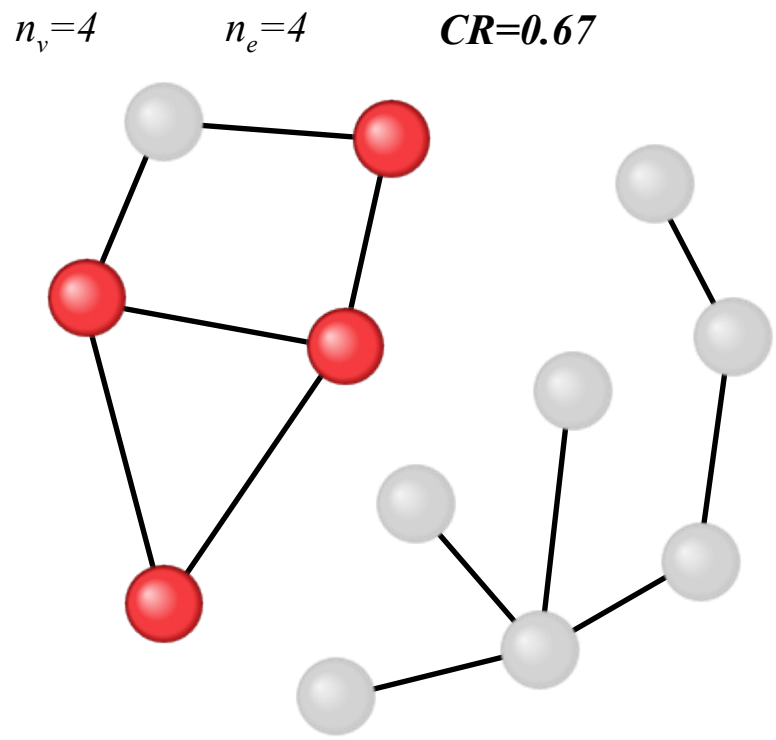
Ранжирование сверхпредставленных Gene Ontology биологических процессов с учетом связности наблюдаемых в эксперименте генов.

CR – Connectivity rate (мера связности)
n_e – число связей
n_v – число вершин

$$CR = \frac{n_e}{C_{n_v}^2}$$

Наблюдаемые в эксперименте гены/белки





Тийс Е.С., Иванисенко В.А., ИциГ СО РАН

Топ-10 сверхпредставленных GO биологических процессов для белков, ассоциированных с преэклампсией, гестационным диабетом, сахарным диабетом и ожирением.

Before ranking by CR

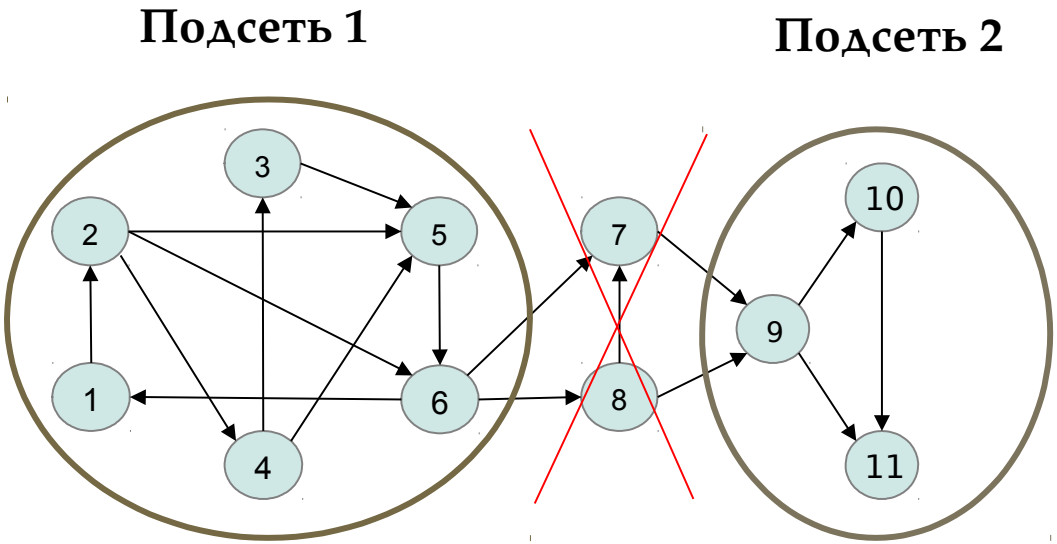
Name	p-value	CR	n	N
response to oxygen-containing compound	3.54E-07	0.711111	10	1579
regulation of biological quality	3.54E-07	0.551282	13	4149
negative regulation of multicellular organismal process	2.13E-06	0.666667	7	557
acute-phase response	4.00E-06	0.833333	4	52
response to hypoxia	4.00E-06	0.8	6	365
response to decreased oxygen levels	4.00E-06	0.8	6	369
response to oxygen levels	5.87E-06	0.8	6	404
response to wounding	9.86E-06	0.678571	8	1321
acute inflammatory response	9.86E-06	0.833333	4	80
positive regulation of response to stimulus	9.86E-06	0.611111	9	1957

After ranking by CR

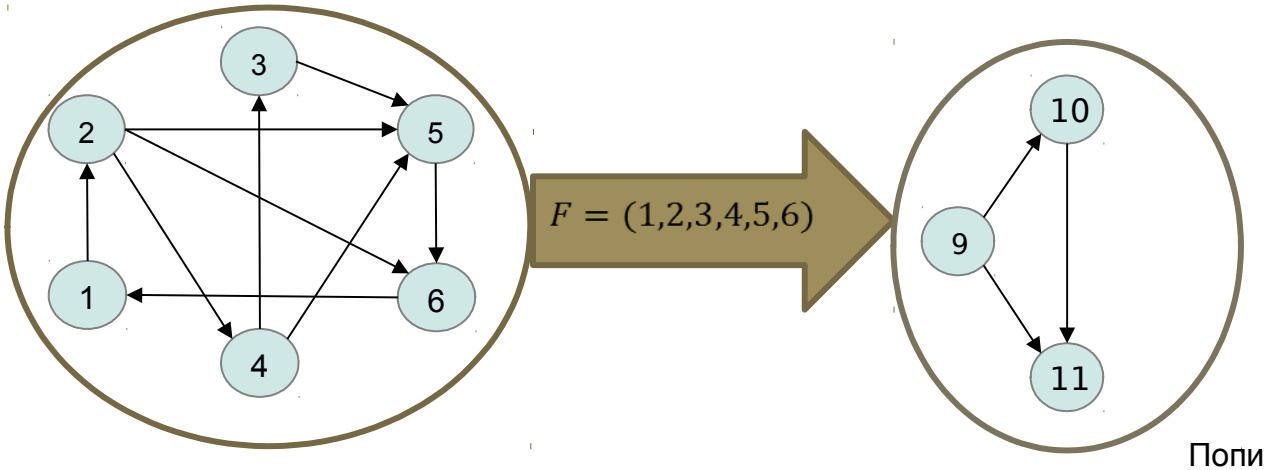
positive regulation of monooxygenase activity	1.65E-05	1	3	22
regulation of fat cell differentiation	1.95E-05	1	4	112
positive regulation of lipid metabolic process	2.98E-05	1	4	133
negative regulation of carbohydrate metabolic process	3.79E-05	1	3	37
positive regulation of nitric oxide biosynthetic process	3.89E-05	1	3	38
positive regulation of oxidoreductase activity	3.89E-05	1	3	38
negative regulation of fat cell differentiation	7.23E-05	1	3	52
regulation of monooxygenase activity	8.42E-05	1	3	57
regulation of nitric oxide biosynthetic process	9.52E-05	1	3	60
defense response to Gram-positive bacterium	1.14E-04	1	3	65

Тийс Е.С., Иванисенко В.А., ИциГ СО РАН

Интеграция моделей генных сетей с помощью линейного функционала.

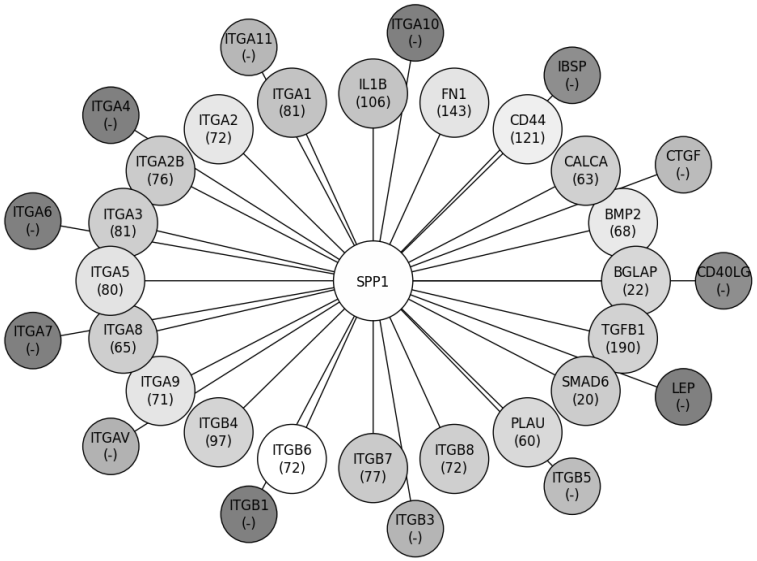


Воздействие подсети 1 на подсеть 2 может моделироваться линейным функционалом, зависящим от концентраций элементов первой подсети.



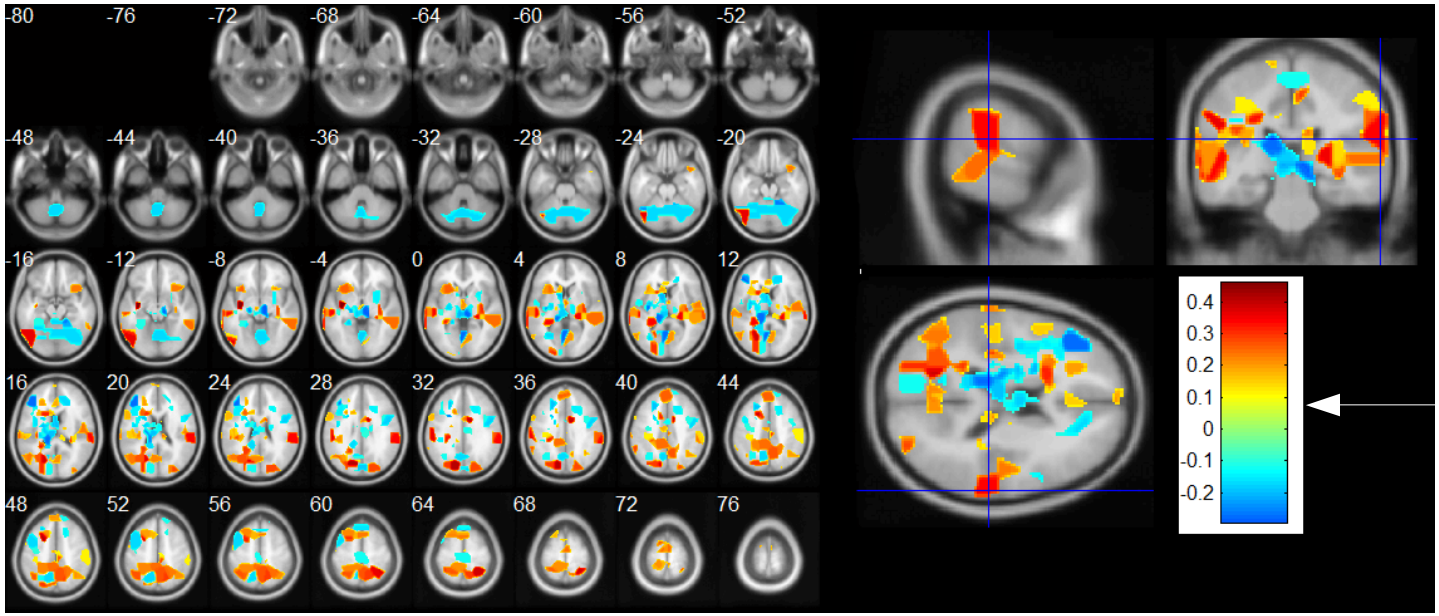
Попик О.В., Иванисенко В.А., ИциГ СО РАН

Моделирование распространения сигнала от нокаута гена в различных структурах мозга методами машинного обучения.



Генная сеть белка SPP1 вовлеченного в клеточную адгезию и развитие глиомы.

Петровский Е. Д., Колчанов Н. А., Иванисенко В. А. Моделирование пространственного распределения эффекта нокаута генов, связанных с агрессивностью глиомы низкой степени злокачественности, в тканях мозга человека с помощью методов машинного обучения //Поиск. – 2014. – Т. 9. – №. 2.



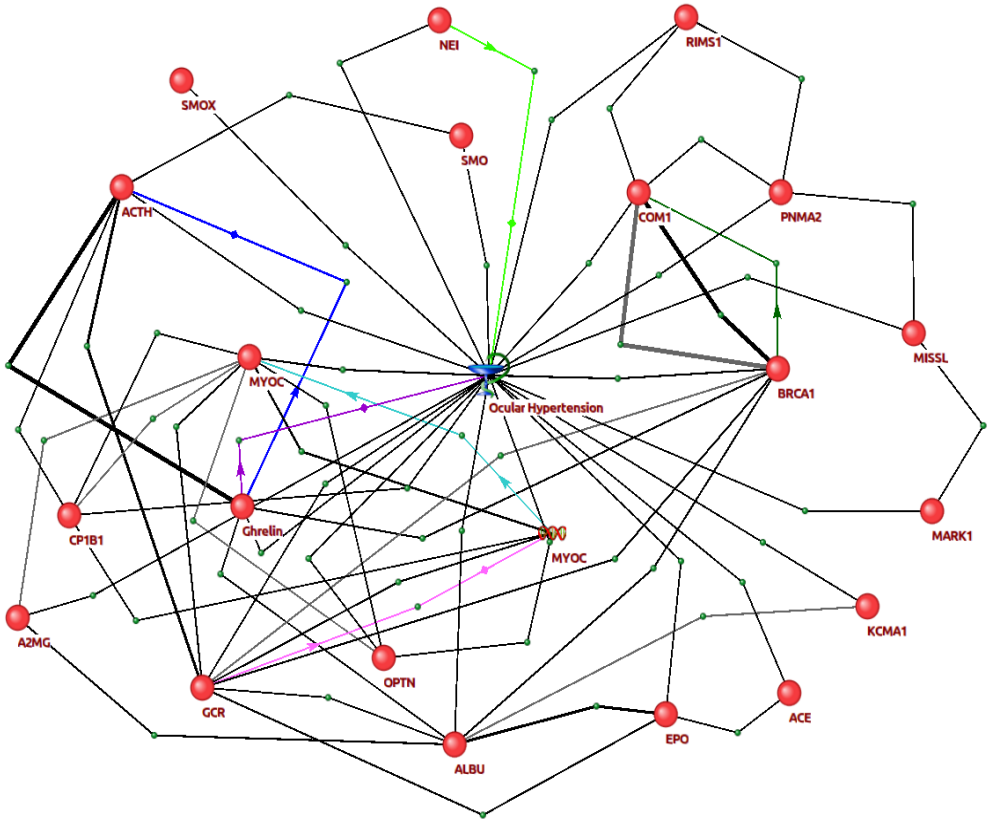
Отклонение усредненного уровня экспрессии соседей первого уровня белка SPP1.

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!