

Построение и анализ генной сети нарушений слуха человека

Замятин Владимир Игоревич

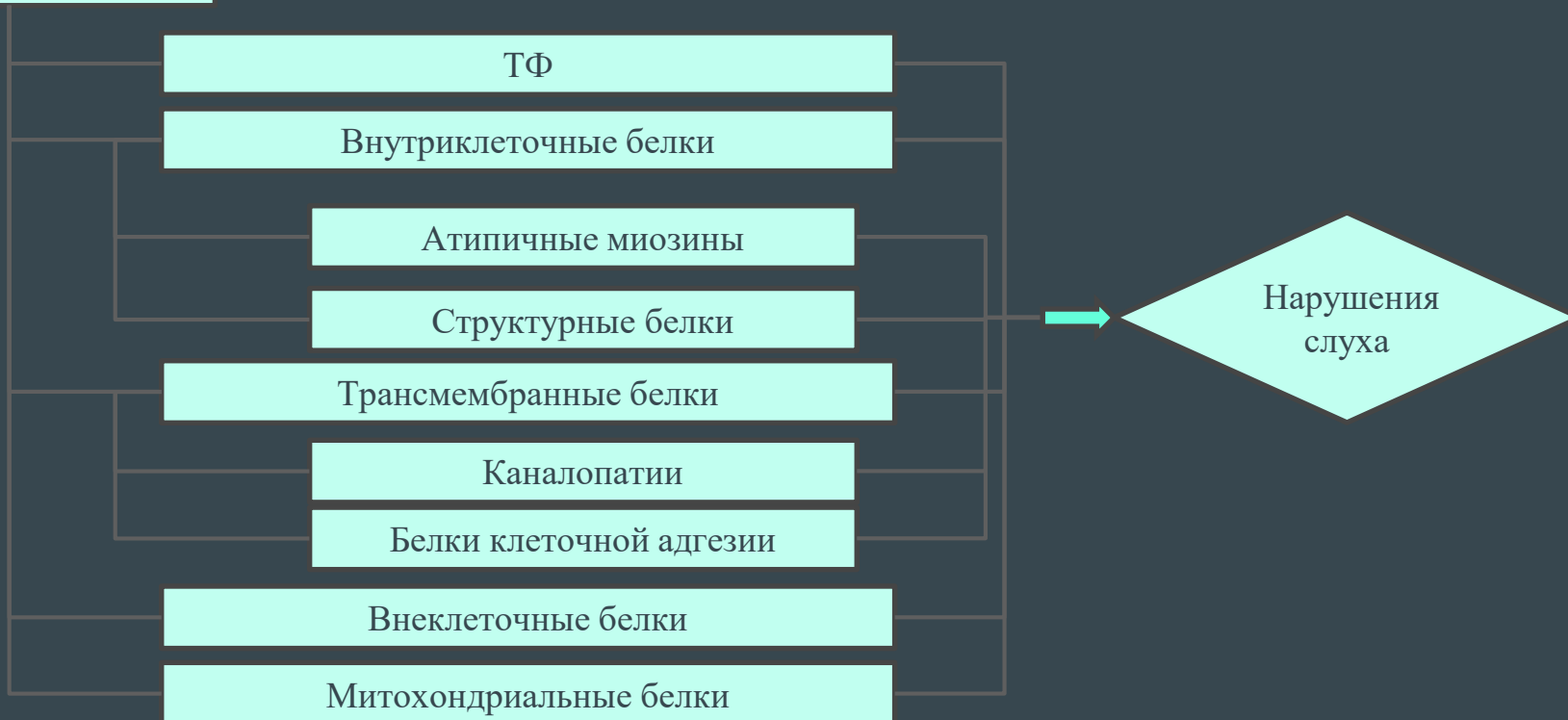
Научный руководитель:
к.б.н. Лашин Сергей Александрович
зав.сек. компьютерного анализа и моделирования биол. систем

Этиология нарушений слуха

Нарушение слуха - это полное или частичное снижение способности обнаруживать и понимать звуки, обнаруживать некоторые (или все) частоты или неспособность различать звуки с низкой амплитудой.



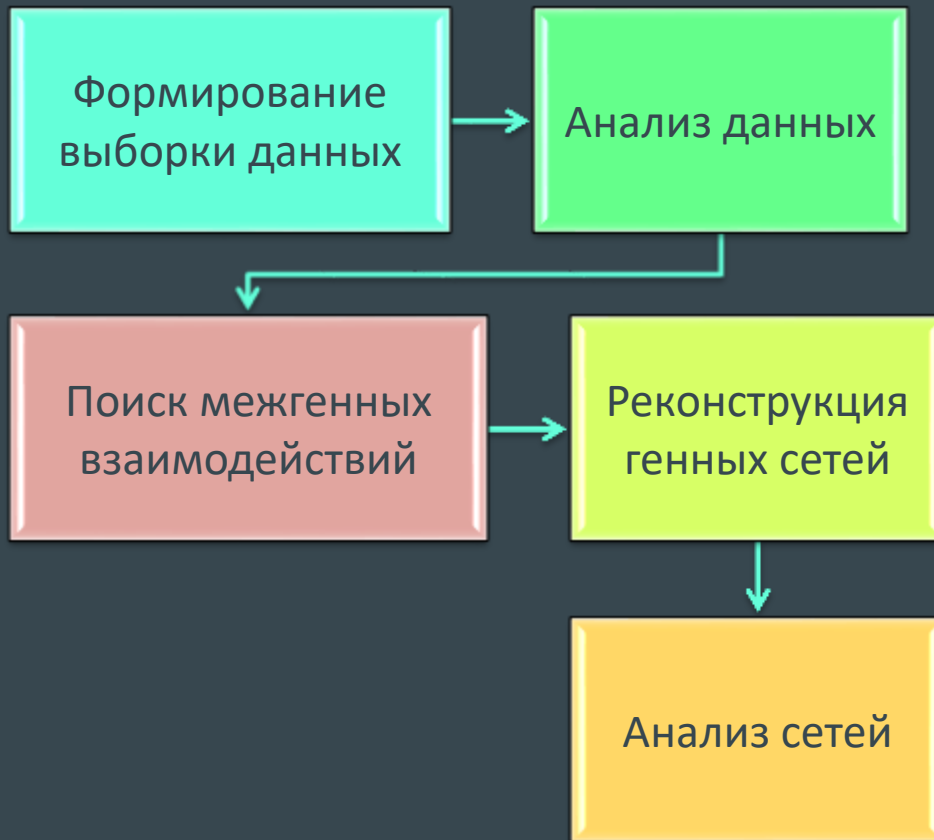
Дефекты в генах

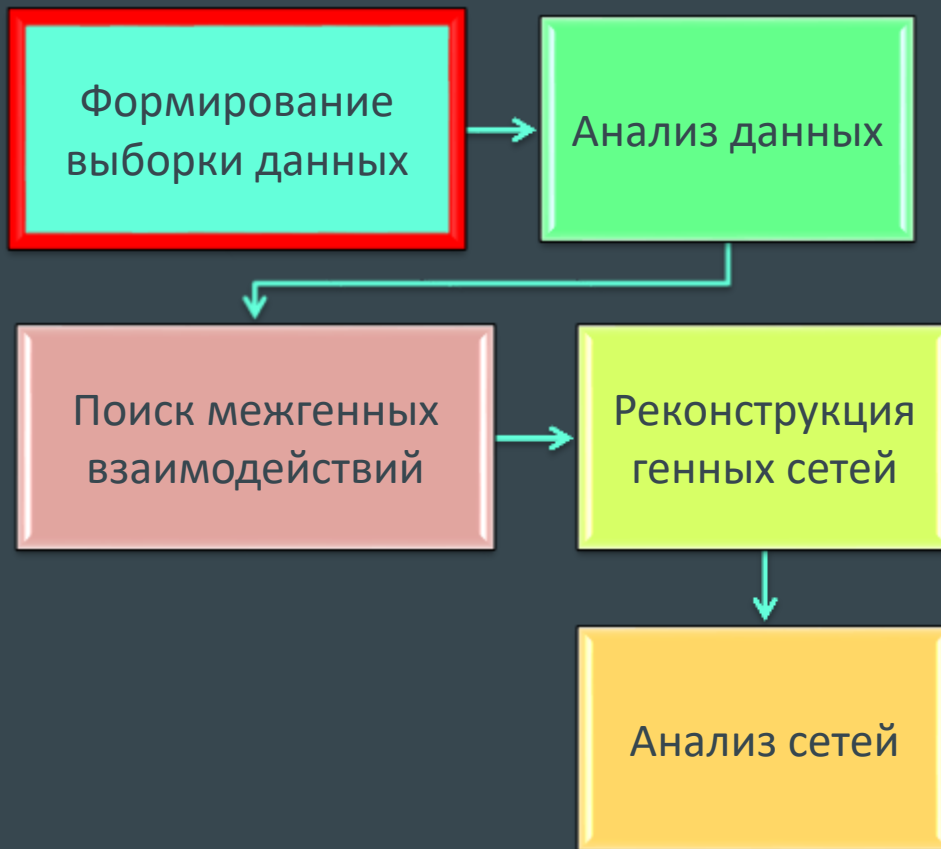


Цель: построение и анализ генных сетей нарушений слуха человека.

Задачи:

1. Используя литературные данные, сформировать выборку генов, ассоциированных с нарушениями слуха.
2. Провести предобработку данных, поиск предполагаемых сайтов связывания транскрипционных факторов для выборки генов.
3. Провести эволюционный анализ и оценку внутрипопуляционной изменчивости.
4. Выявить межгенные взаимодействия на основе данных о транскрипционной регуляции и белок-белковых взаимодействиях.
5. Реконструировать генные сети и провести их анализ.





Формирование выборки данных

Списки генов

Источники:

1. База данных Hereditary Hearing Loss, **143 гена**
2. Статья Стаматью, Станкович, **450 генов** [1]

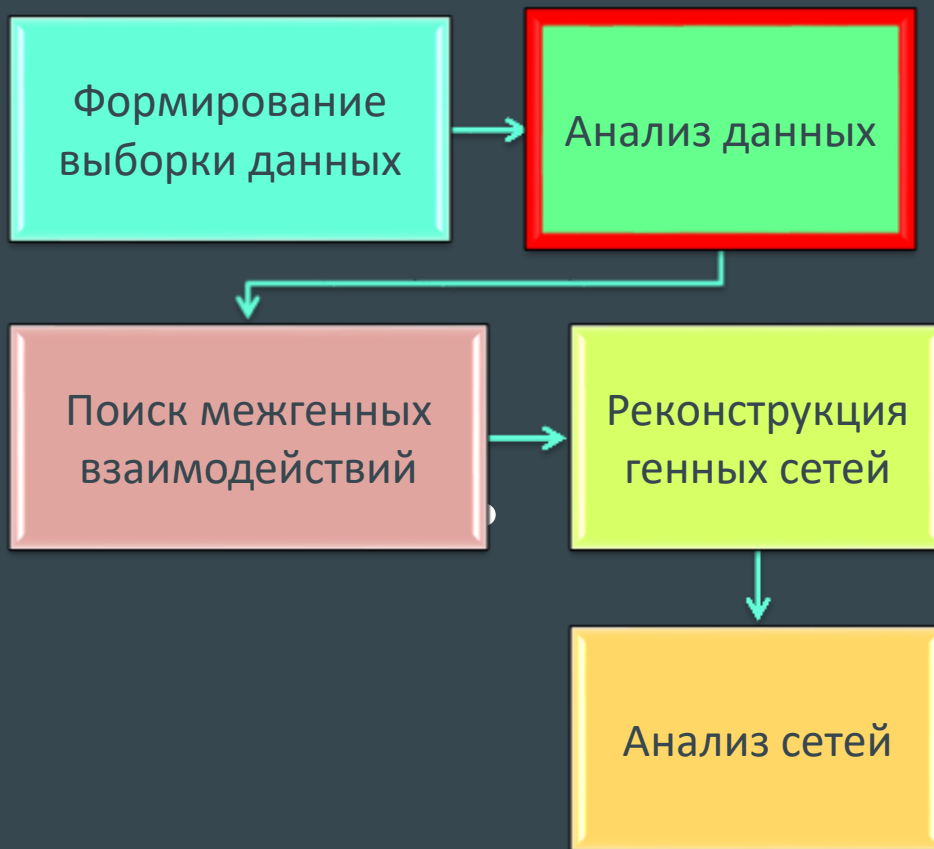
Выделение подмножества ТФ

Пересечение со списком транскрипционных факторов БД AnimalTFDB.
11 и 40 ТФ

Экстракция upstream районов

300, 1000, 2000 п.н. получены с помощью скриптов на **Python** (библиотеки **Biopython, API NCBI, KEGG**)
GRCh37

[1] Stamatiou G.A., Stankovic K.M. A comprehensive network and pathway analysis of human deafness genes // Otol. Neurotol. 2013. Т. 34. № 5. С. 961–970.



Анализ данных

Поиск сайтов связывания транскрипционных факторов (ССТФ)

Оценка внутривидовой изменчивости

Эволюционный анализ

Upstream-районы

Сервис MoLoTool
(Nocomo [4])

Обработка скриптами
Python

Таблицы транскрипционной
регуляции

Проект «1000 геномов человека»

Samtools [10] (экстракция
целевых районов)

Обработка скриптами Python

Таблицы насыщенности SNP

Аннотация SNP в VEP [5]

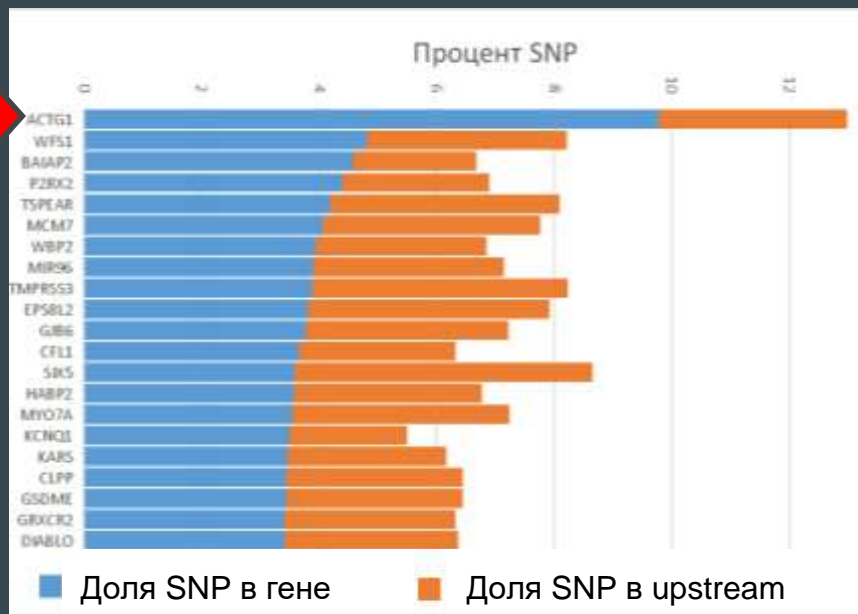
Списки генов из выборок

Программный комплекс
Orthosuit [6]

Таблицы PAI и DI

Анализ данных

Оценка внутривидовой изменчивости

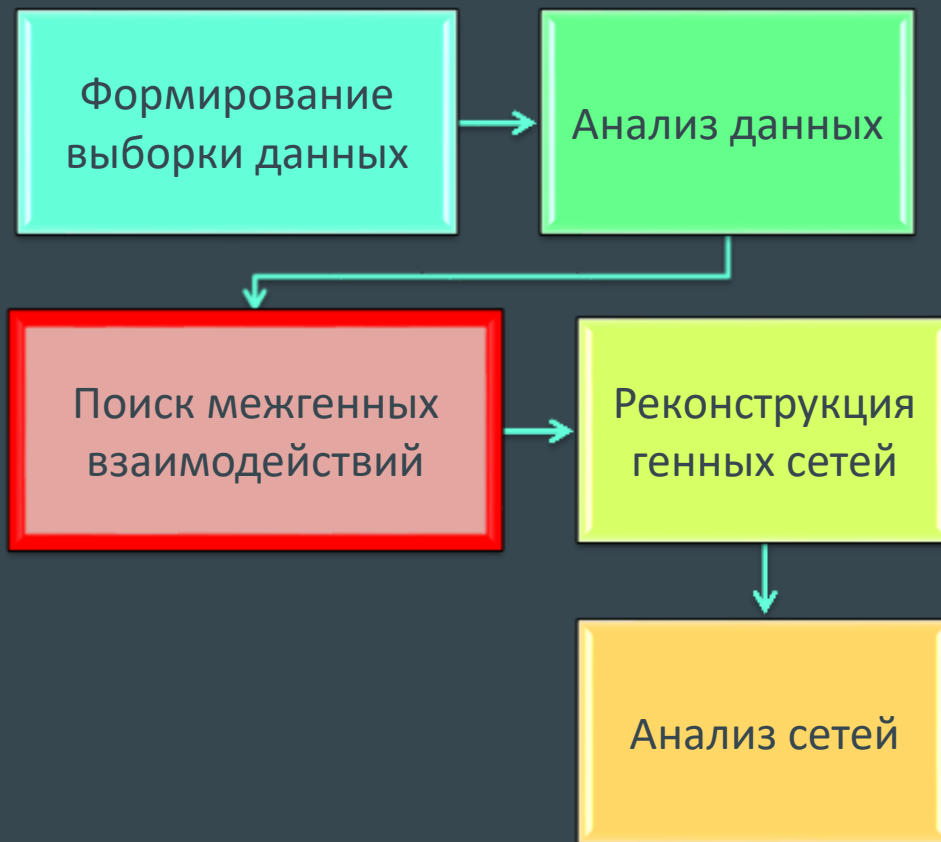


Эволюционный анализ

Gene	DI	Identity									Legend	
		1	0.9	0.8	0.7	0.6	0.5	0.4	0.3			
POU3F4	0.001466	13	6	2	2	2	0	0	0	15	Homo sapiens (human)	
EDNRB	0.001299	13	6	6	5	5	5	0	0	14	Homo	
CAMK4	9.83E-04	15	8	6	5	2	1	1	0	13	Hominidae	
POLR1C	9.29E-04	13	7	6	5	2	0	0	0	12	Catarrhini	
ROR1	8.63E-04	13	6	6	5	5	5	0	0	11	Haplorrhini	
HOMER2	8.53E-04	13	8	6	6	2	2	0	0	10	Primates	
RDX	6.58E-04	12	5	0	0	0	0	0	0	9	Euarchontoglires	
ACTG1	0	6	1	1	1	0	0	0	0	8	Eutheria	
ACVR1	0	12	8	6	5	2	2	0	0	7	Mammalia	
										6	Euteleostomi	
										5	Vertebrata	
										4	Craniata	
										3	Chordata	
										2	Metazoa	
										1	Eukaryota	
										0	Cellular Organisms	

DI – индекс дивергенции, Ka/Ks относительно близкого вида.

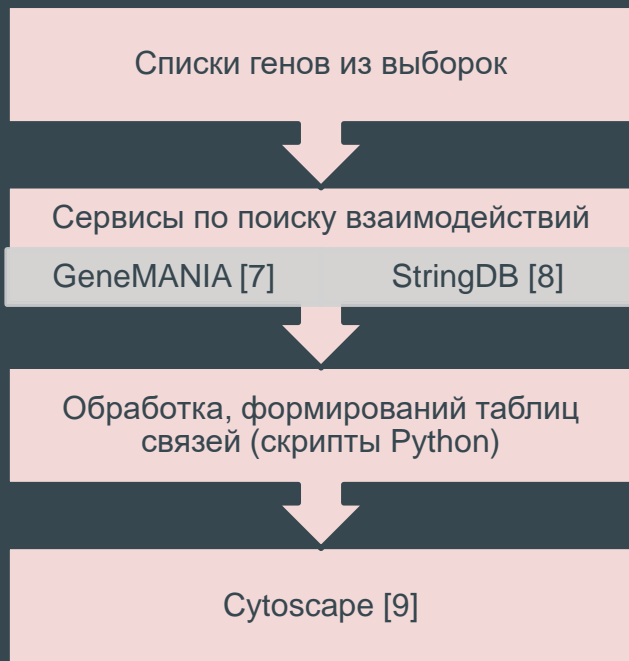
PAI – филогенетический возраст – близость белка к белку определённого системного таксона при данном identity.

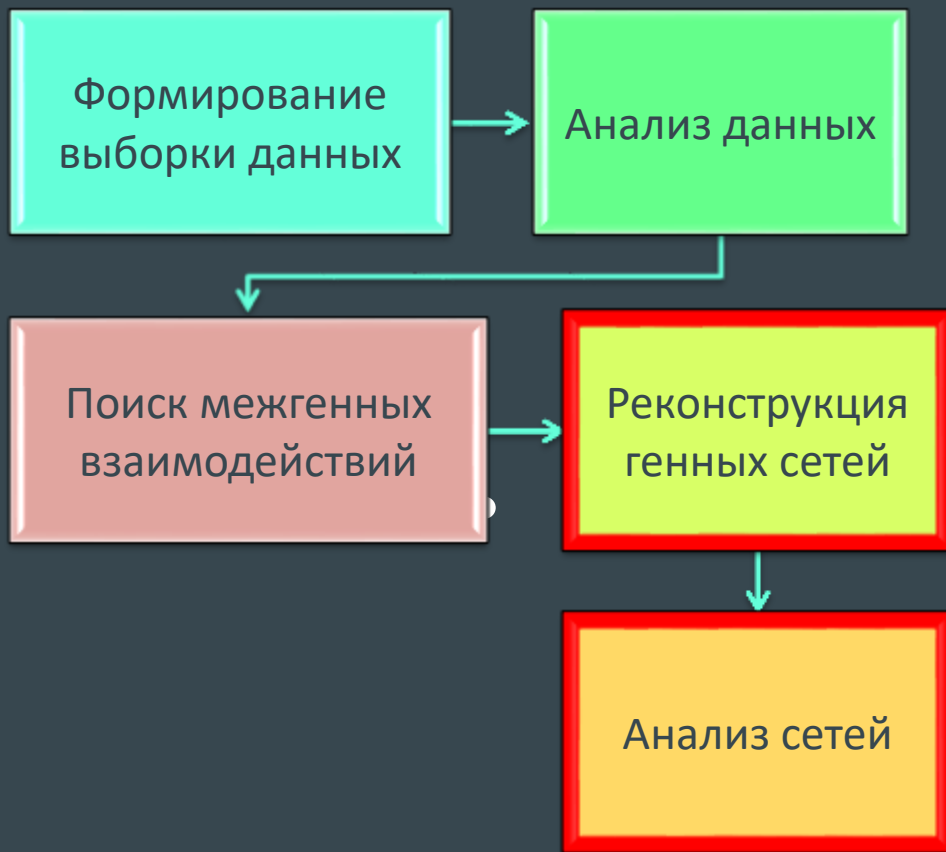


Поиск межгенных взаимодействий

Белок-белковые взаимодействия

Транскрипционная регуляция (ТР)





Реконструкция и анализ сетей

Сеть - управляющее ядро

- Белок-белковые взаимодействия для ТФ из StringDB

Сеть транскрипционной регуляции

- Данные о ССТФ на upstream-районах для ТФ из выборок

Общая сеть белок белковых взаимодействий

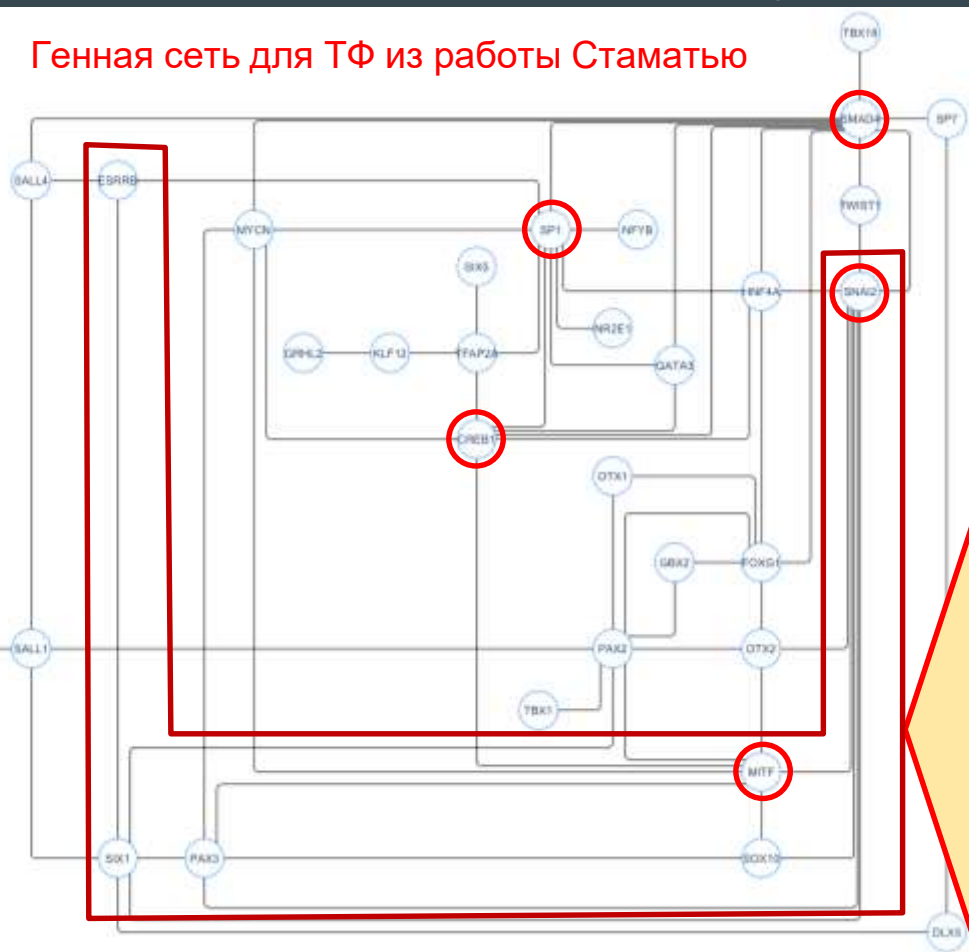
- Белок-белковые взаимодействия между всеми элементами из StringDB и GeneMANIA

Визуализация в Cytoscape

Реконструкция и анализ сетей

Сеть - управляющее ядро

Генная сеть для ТФ из работы Стаматью



Выделены основные ТФ – регуляторы сети:
SMAD4, SNAI2, SP1, CREB1, MITF для Стаматью.
Показано наложение сетей.

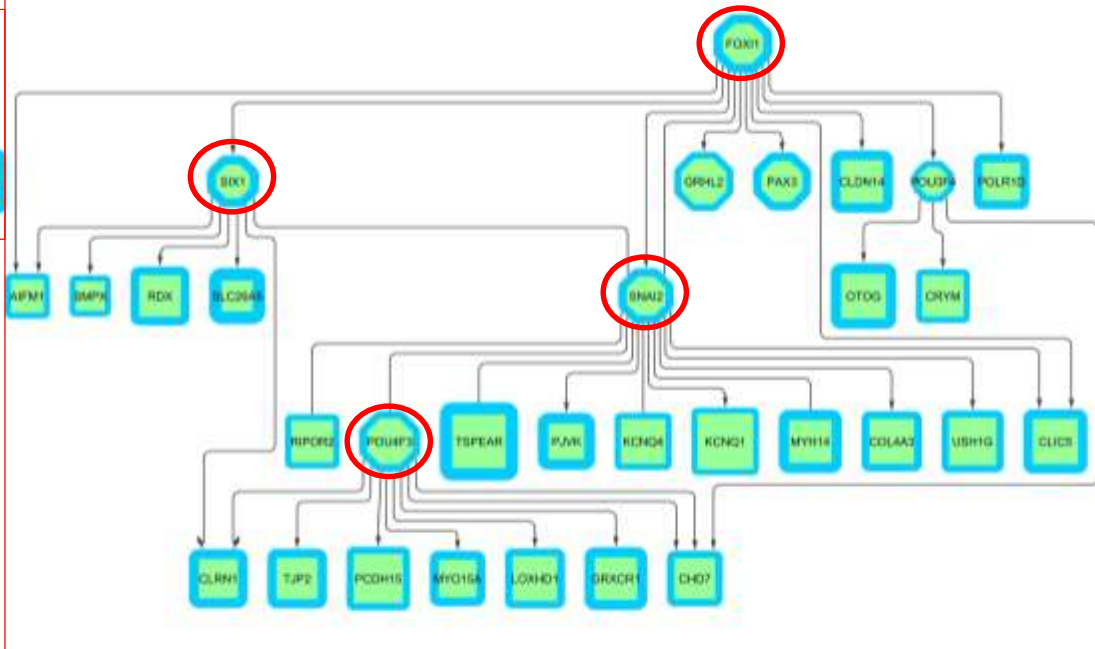
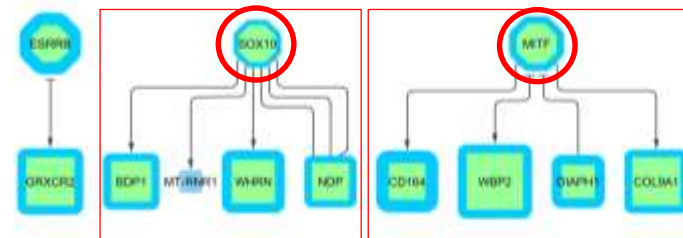
Генная сеть
для ТФ из ННЛ






Реконструкция и анализ сетей

Сеть транскрипционной регуляции

Данные для 300п.н. upstream, выборка генов из Hereditary Hearing Loss



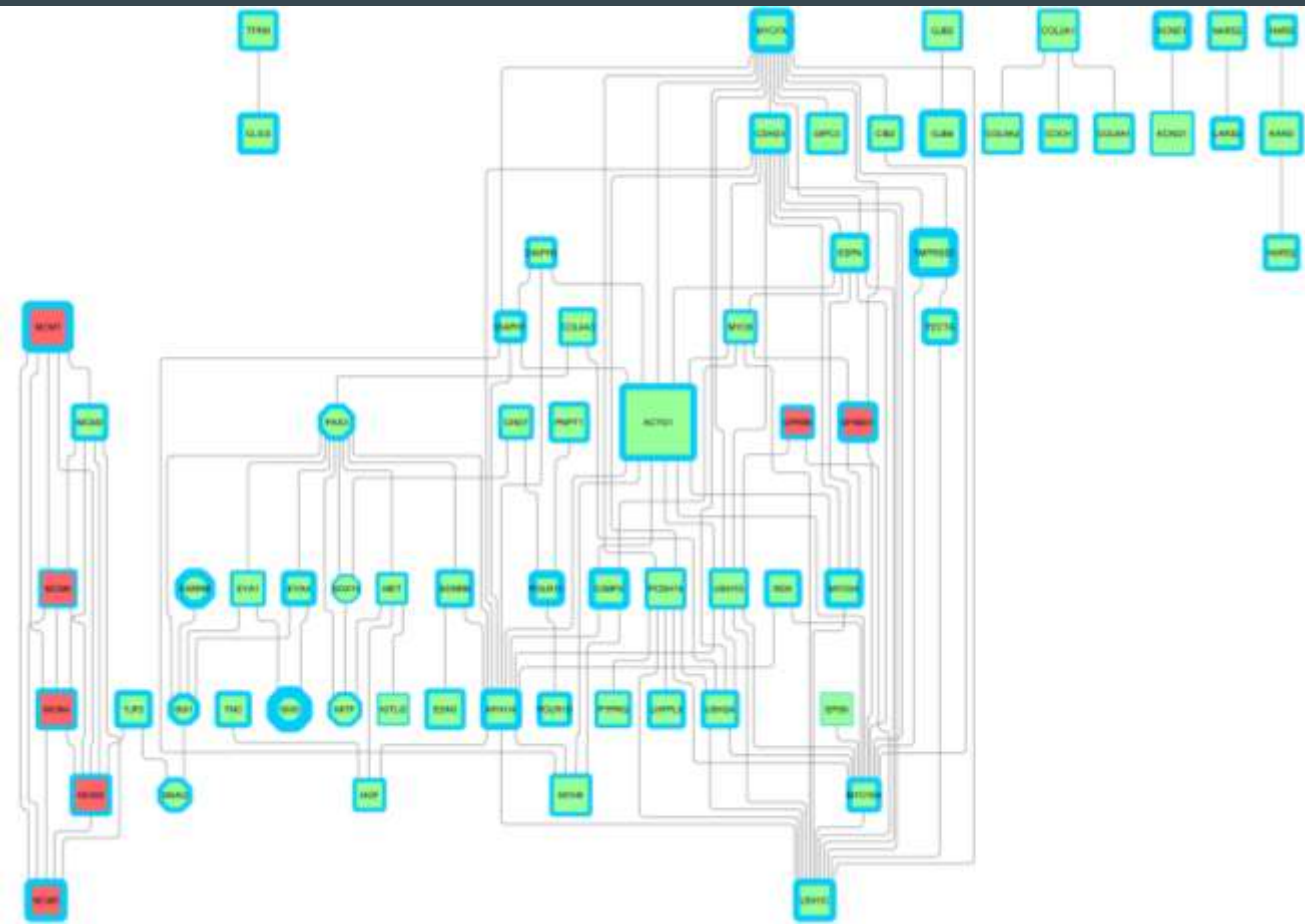
Легенда	
Форма	 Ген
	 ТФ
Толщина рамки	%SNP в промоторе
Размер узла	%SNP в гене
Цвет узла	Ген из выборки


 Транскрипционный фактор ESRRB, из выборки, SNP в гене – 3.1%, в промоторе – 3.9%

Реконструкция и анализ сетей

Сеть белок-белковых взаимодействий (StringDB)

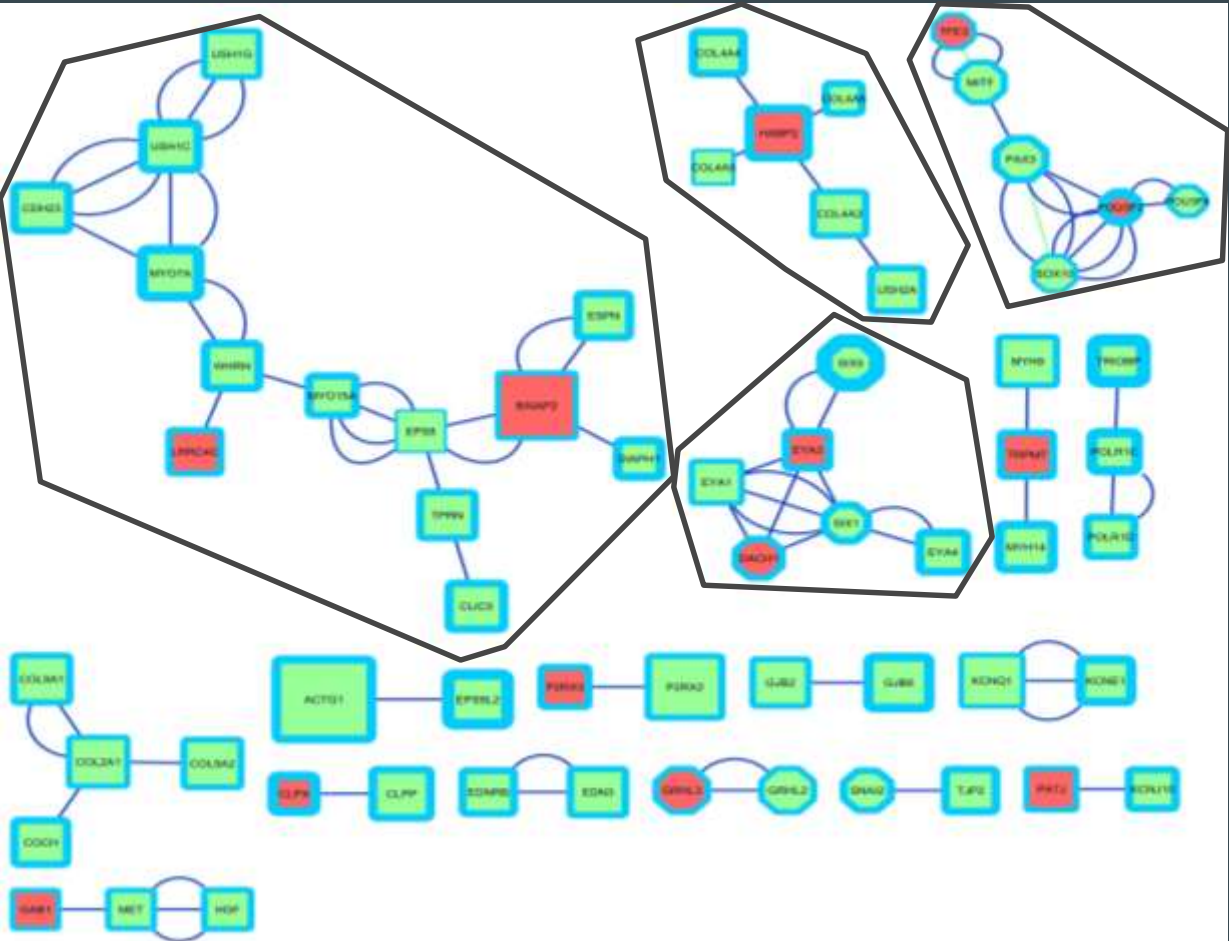
Выборка генов из Hereditary Hearing Loss (String → Cytoscape, experimental)



Ген	Описание
<i>DFNB31</i>	Участвует в формировании стереоцилий и сборке актина. Связан с AP глухотой и синдромом Ушера
<i>GPR98</i>	Кодирует рецептор из суперсемейства связанных с G-белком. Синдром Ушера 2.
<i>MCM3</i> <i>MCM4</i> <i>MCM5</i> <i>MCM6</i> <i>MCM7</i>	Кодируют субъединицы белка, участвующего в инициации репликации.

Реконструкция и анализ сетей

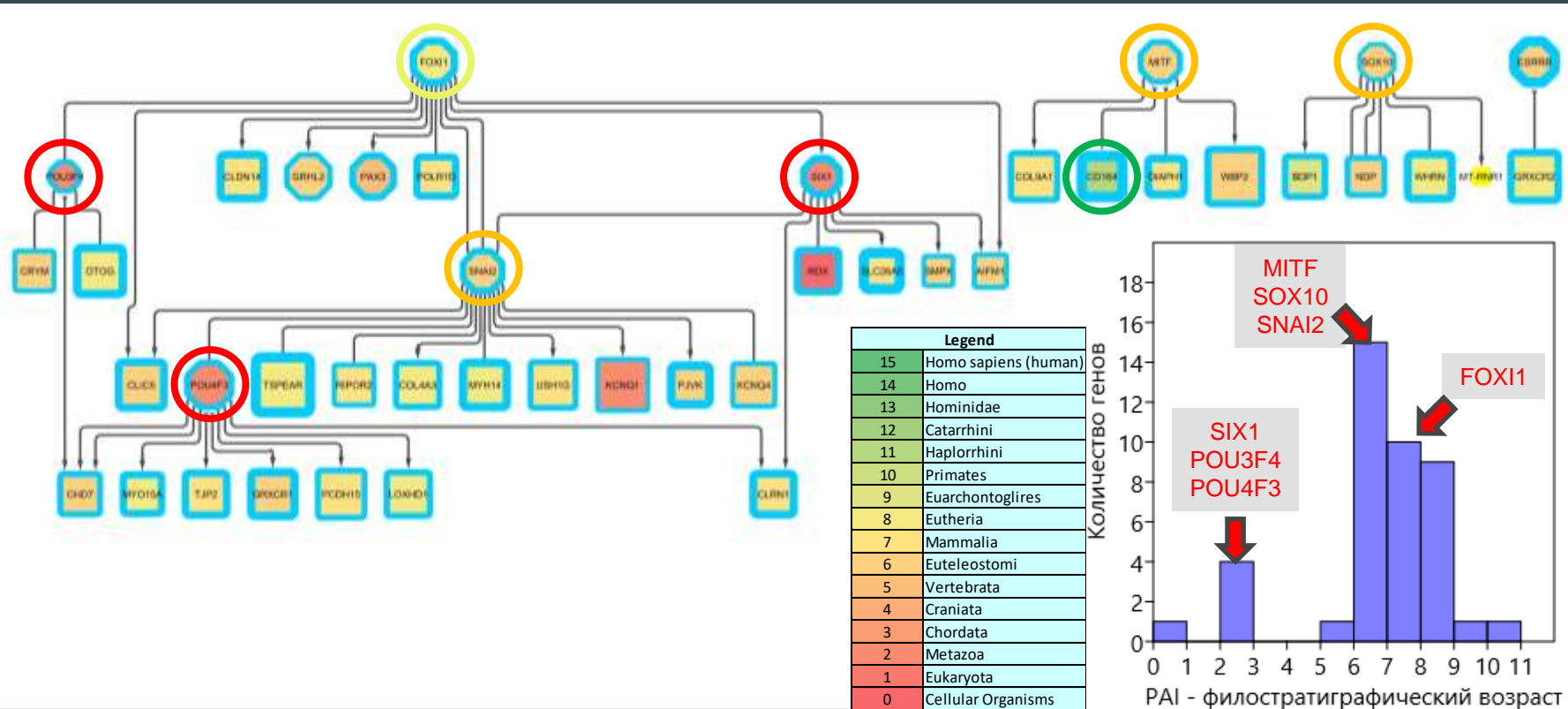
Сеть белок-белковых взаимодействий (GeneMANIA)



Ген	Описание
<i>TFE3</i>	базовый транскрипционный фактор
<i>POU3F2</i>	имунный транскрипционный фактор класса POU-III
<i>TRPM7</i>	ген, кодирующий ионный канал и протеинкиназу
<i>HABP2</i>	пептидаза семейства S1 (сериновые протеазы)
<i>EYA3</i>	белок семейства EYA, транскрипционный коактиватор
<i>DACH1</i>	транскрипционный фактор
<i>BAIAP2</i>	специфичный для мозга белок, кодирующий ингибитор ангиогенеза

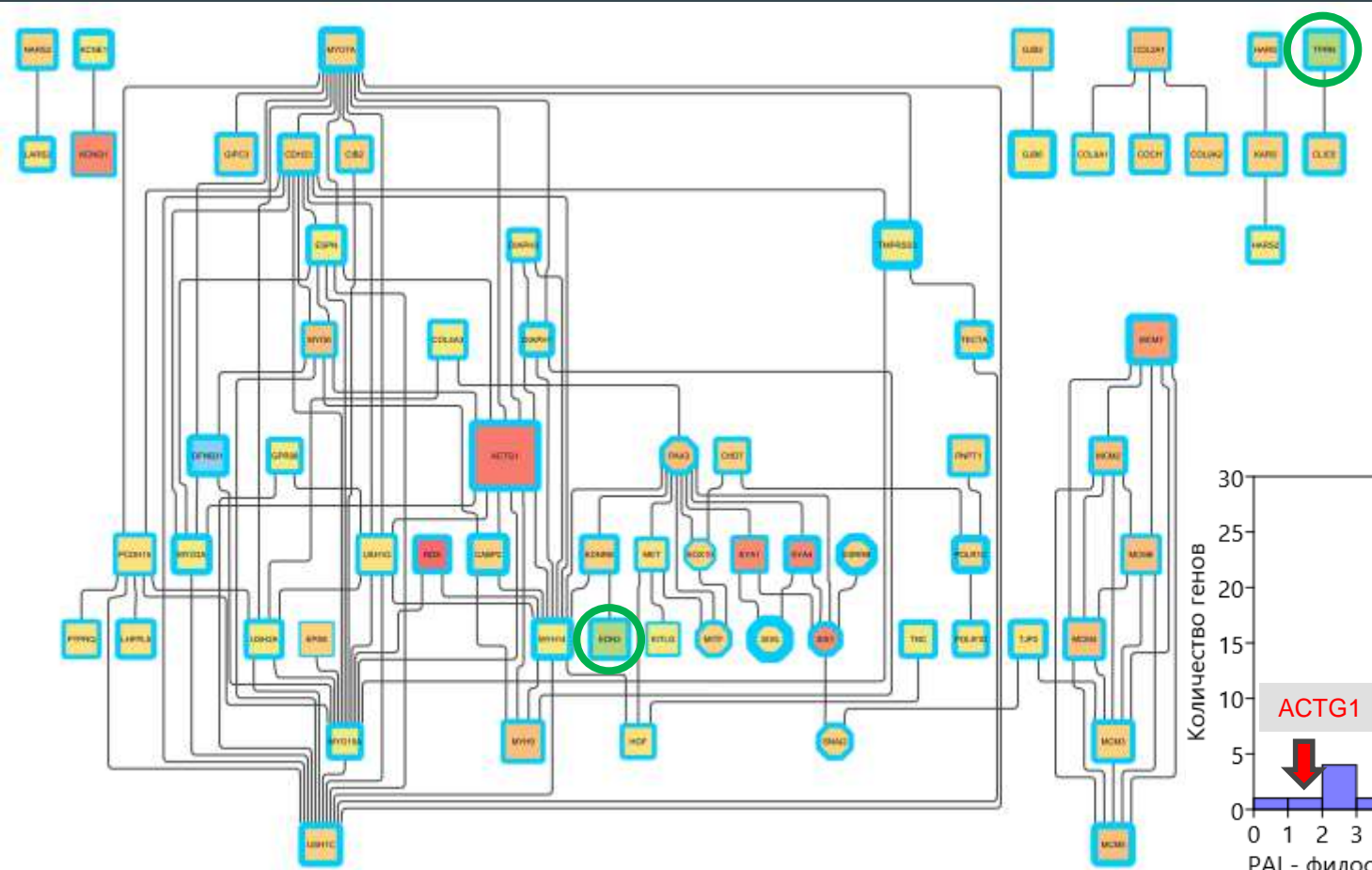
Реконструкция и анализ сетей

Эволюционная карта сети транскрипционной регуляции
Данные для 300п.н. upstream, выборка генов из Hereditary Hearing Loss

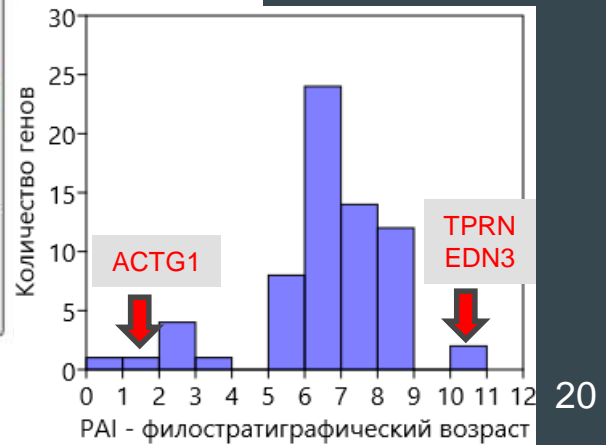


Реконструкция и анализ сетей

Эволюционная карта сети белок-белковых взаимодействий (StringDB)



Legend	
15	Homo sapiens (human)
14	Homo
13	Hominidae
12	Catarrhini
11	Haplorrhini
10	Primates
9	Euarchontoglires
8	Eutheria
7	Mammalia
6	Euteleostomi
5	Vertebrata
4	Craniata
3	Chordata
2	Metazoa
1	Eukaryota
0	Cellular Organisms



Выводы

1. Реконструированы сети, отображающие предположительный характер появлений нарушений слуха: *сети белок-белковых взаимодействий и транскрипционной регуляции*.
2. В сетях выделены отдельные регуляторные контуры, мутации в элементах которых, по всей видимости, приводят к различным механизмам развития отклонений слуха.
3. Показаны транскрипционные факторы, вовлечённые в управление процессами протекания глухоты: **ESRRB, FOXI1, MITF, PAX3, POU3F4, POU4F3, SIX1, SNAI2, SOX10, GRHL2, SIX5.**
4. Обнаружены новые гены в т.ч., гены **ТФ** (**POU3F2, TFE3, DACH1, TRPM7, HABP2, EYA3, BAIAP2** и др.), ассоциированные с заболеванием.
5. Эволюционный анализ позволил выявить «молодые» гены (**TPRN, EDN3, CD164**) – элементы сетей и процессов звуковосприятия и показал преимущественную консервативность рассматриваемых транскрипционных факторов.
6. На основе эволюционного анализа и оценки распределений SNP (данные проекта «1000 геномов человека») обнаружены неоднозначные свойства гена **ACTG1**: несовпадение оценок эволюционной консервативности и насыщенности SNP свидетельствует в пользу прекращения давления стабилизирующего отбора на данный ген в популяциях человека.

Благодарности

Посух О.Л. – биомедицинские консультации

Игнатьева Е.В. – генные сети, ТФ

Мустафин З.С. – Orthosuit, эволюционный анализ

В рамках магистерской диссертации планируется:

1. Более детально рассмотреть распределение SNP и оценить суммарную патогенность в пределах каждого гена;
2. Провести функциональную классификацию генов, в частности, проанализировать гены, задействованные в органогенезе;
3. Проверить гипотезу о корреляции связности гена (количество взаимодействий с другими элементами) и подверженности стабилизирующему отбору;
4. Оценить общую устойчивость узловых элементов относительно более обширных сетей.
5. Попытаться экстраполировать отлаженный конвейер реконструкции и анализа генных сетей на другое заболевание.

Базы данных, литература, инструменты:

[1] <http://hereditaryhearingloss.org/>

[2] Stamatou G.A., Stankovic K.M. A comprehensive network and pathway analysis of human deafness genes // Otol. Neurotol. 2013. Т. 34. № 5. С. 961–970.

[3] <http://bioinfo.life.hust.edu.cn/AnimalTFDB/>

[4] Kulakovskiy I. V. и др. HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis // Nucleic Acids Res. 2018. Т. 46. № D1. С. D252–D259

[5] McLaren W. и др. The Ensembl Variant Effect Predictor // Genome Biol. 2016. Т. 17. № 1

[6] Mustafin Z.S. и др. Orthoscape: A cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles // BMC Bioinformatics. 2017. Т. 18

[7] Zuberi K. и др. GeneMANIA prediction server 2013 update. // Nucleic Acids Res. 2013. Т. 41. № Web Server issue

[8] Mering C. von и др. STRING: A database of predicted functional associations between proteins // Nucleic Acids Res. 2003. Т. 31. № 1. С. 258–261

[9] Smoot M.E. и др. Cytoscape 2.8: New features for data integration and network visualization // Bioinformatics. 2011. Т. 27. № 3. С. 431–432

[10] Li H. и др. The Sequence Alignment/Map format and SAMtools // Bioinformatics. 2009. Т. 25. № 16. С. 2078–2079

Формирование выборки данных, первичный анализ

Наборы генов, ассоциированных с нарушениями слуха, их промоторные области (2000п.н., RefSeq)

Выделение подмножества ТФ (AnimalTFDB). Поиск предполагаемых ССТФ (Hocomoco/MoLoTool)

Анализ данных

Анализ внутривидовой изменчивости генов (1000 Genomes Project, SNP)

Эволюционный анализ (PAI, DI). Кластерный анализ - паттерны корегуляции.

Поиск межгенных взаимодействий

Транскрипционная регуляция (данные о ССТФ)

Белок-белковые взаимодействия (StringDB, GeneMANIA)

Реконструкция генных сетей (Cytoscape)

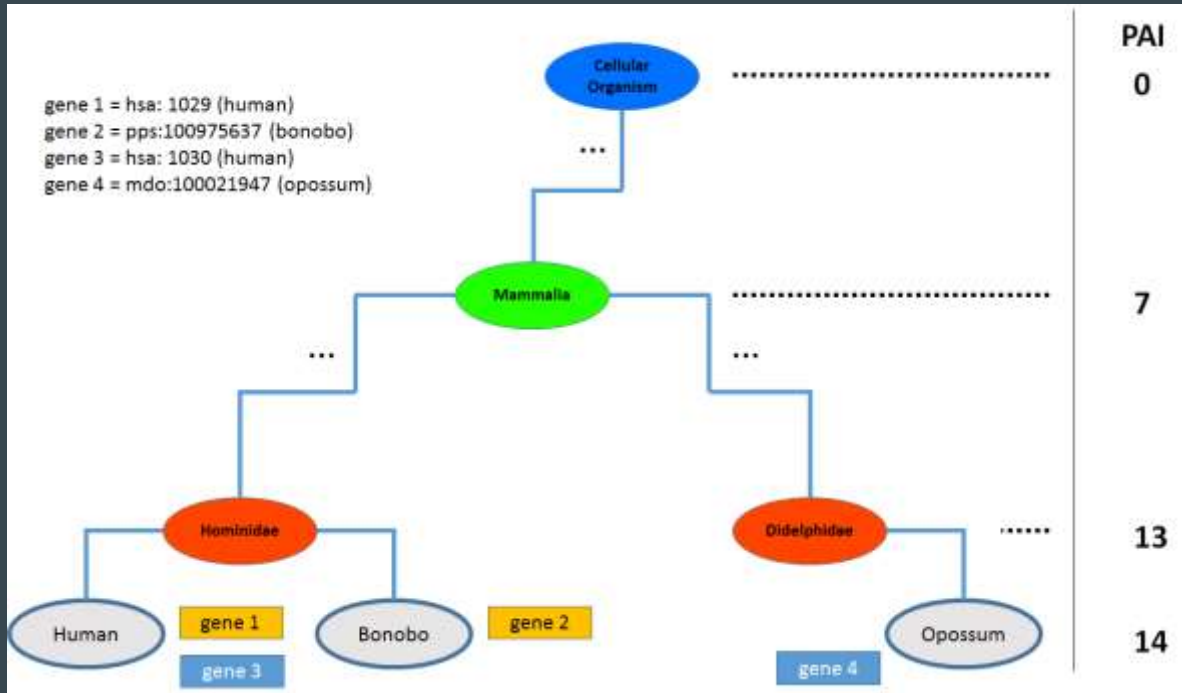
Управляющее ядро (ТФ, StringDB)

Сеть транскрипционной регуляции (ТФ и управляемые гены)

Сеть белок-белковых взаимодействий

Анализ сетей

Филостратиграфический возраст гена (PAI)



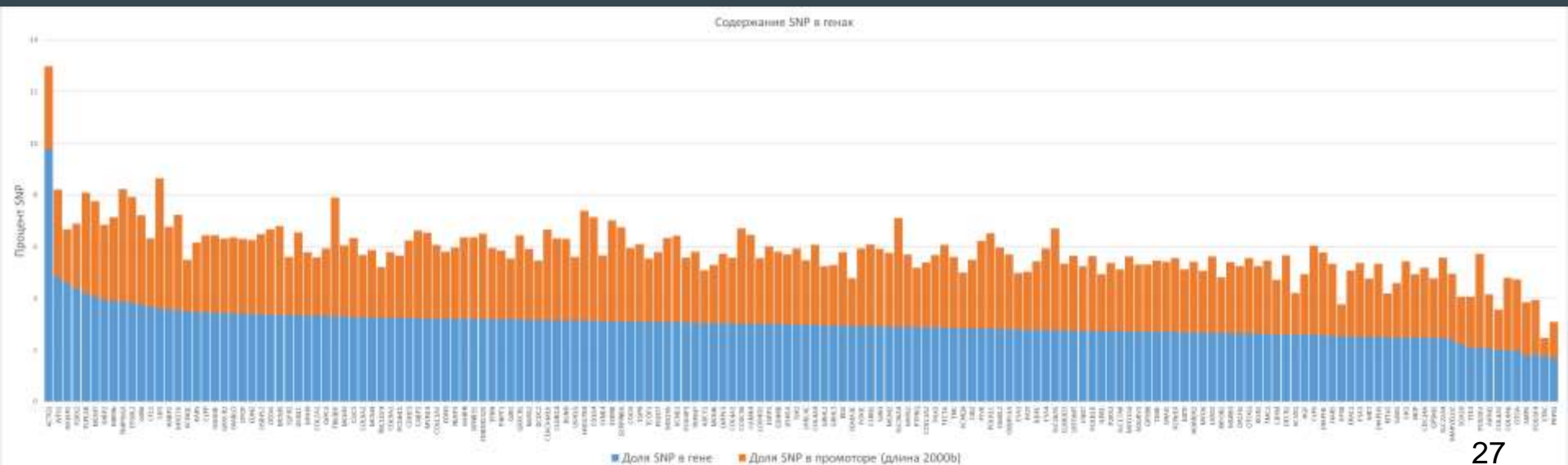
PAI рассчитывается как индекс наиболее удаленного от корня дерева узла, на котором сходятся таксономические ряды гена и всех его ортологов
14 – самый молодой
0 – самый древний

$$\text{Gene Set PAI} = \frac{\sum_{\text{genes}} \text{PAI}}{N}$$

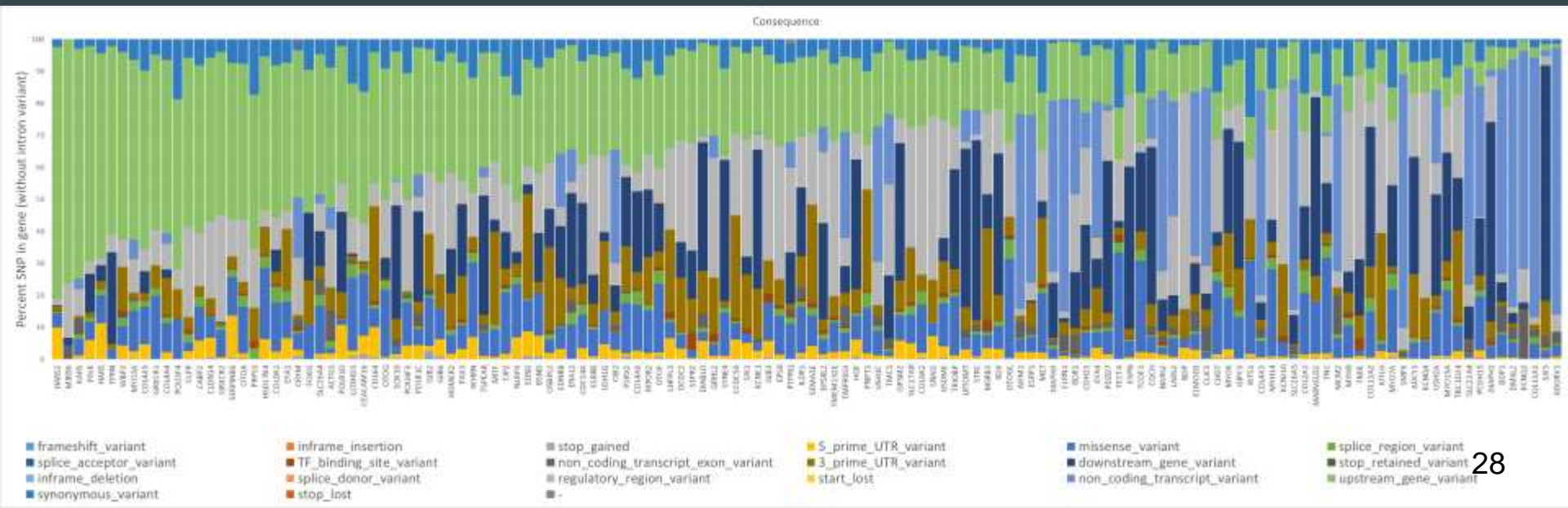
$$\text{Network PAI} = \frac{\sum_{\text{nodes}} (\text{PAI} * d)}{2N}$$

$PAI(\text{gene1}) = PAI(\text{gene2}) = 13$ (“Hominidae” – «молодые» гены)
 $PAI(\text{gene3}) = PAI(\text{gene4}) = 7$ (“Mammalia” – гены «среднего возраста»)

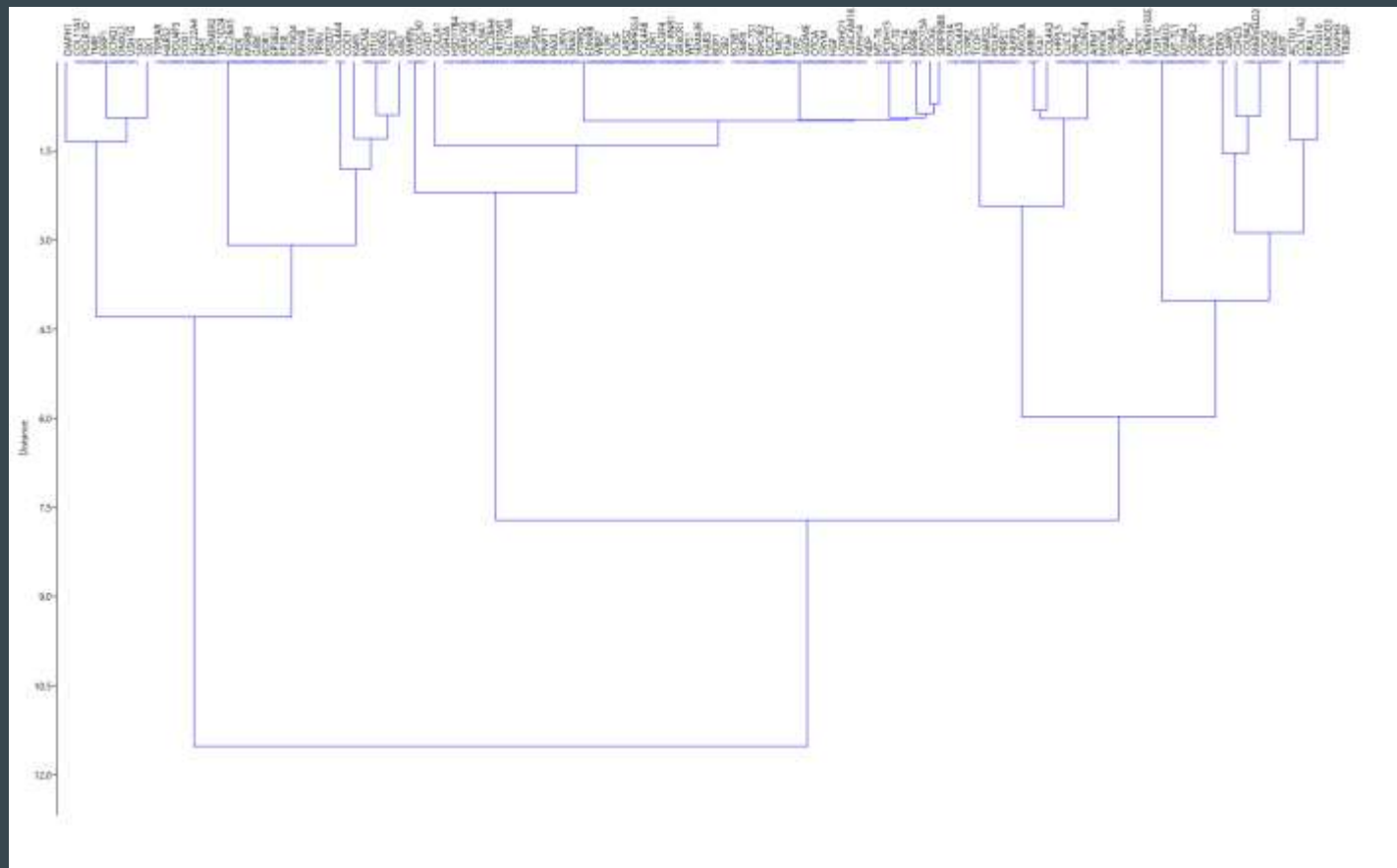
Анализ SNP, доля в гене и промоторе (single nucleotide polymorphism) Выборка генов из Hereditary Hearing Loss



Анализ SNP, аннотация Variants Effect Predictor Ensembl (single nucleotide polymorphism) Выборка генов из Hereditary Hearing Loss

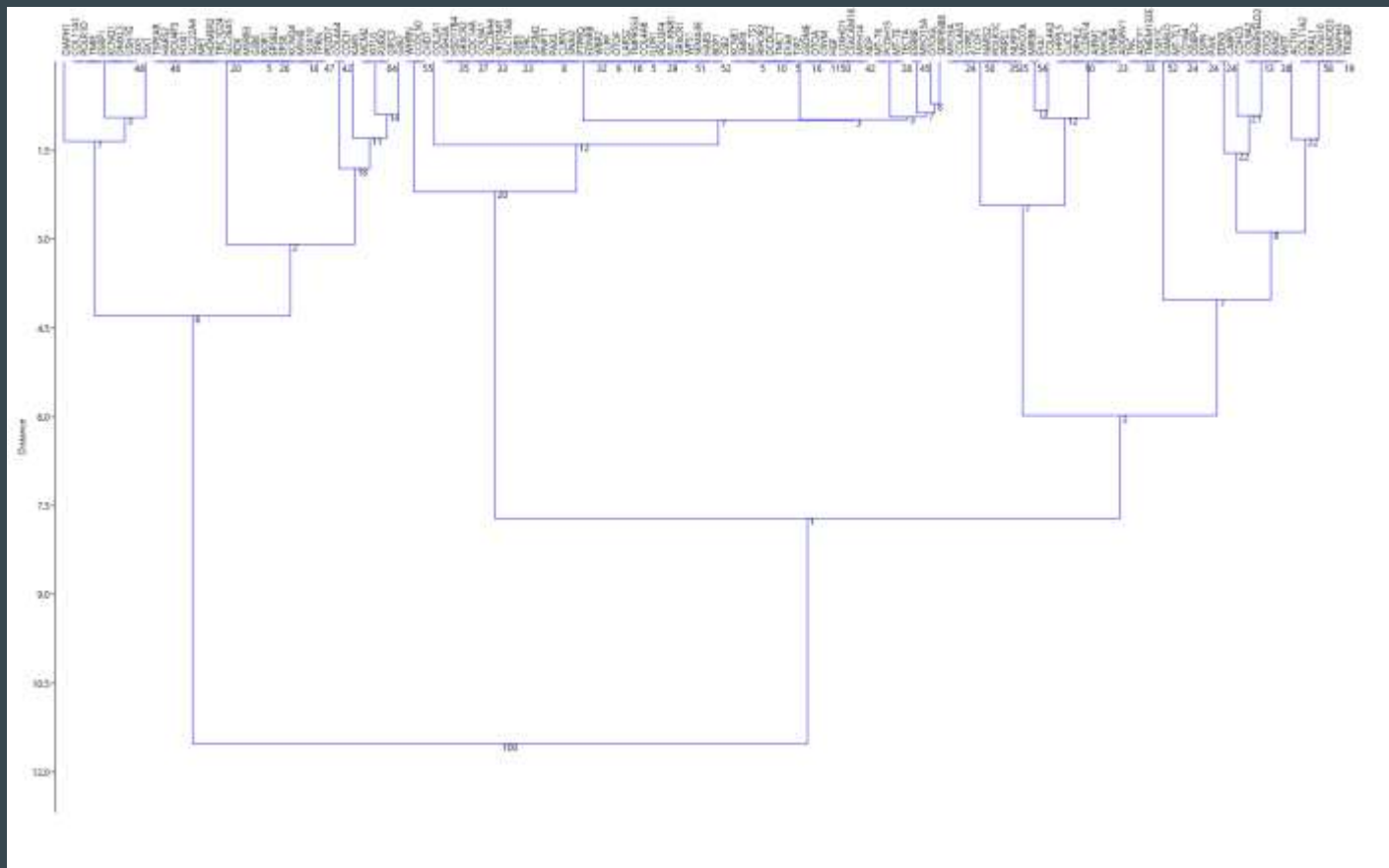


Корегуляция



Данные для 300b upstream, выборка генов из Hereditary Hearing Loss

Корегуляция

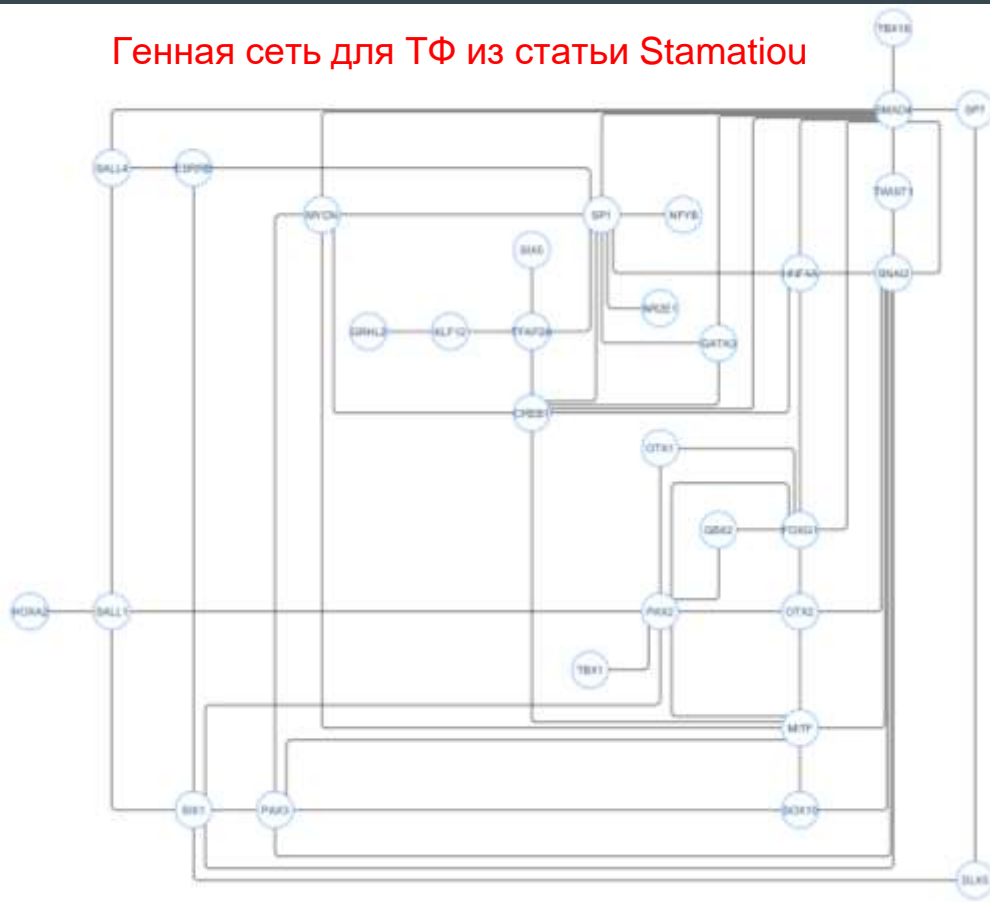


Данные для 300b upstream, выборка генов из Hereditary Hearing Loss

Генная сеть – коровое ядро регуляции (ТФ)

String -> Cytoscape

Генная сеть для ТФ из статьи Stamatiou

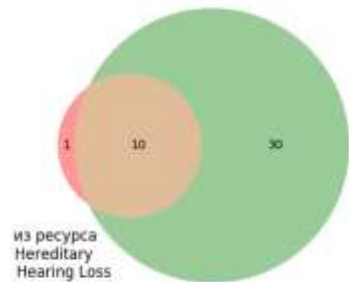


В кружках – официальные названия генов, связи обозначены линиями; взаимодействия – combined score из методов String (включая текст-майнинг)
Явно выделяются основные ТФ – регуляторы сети: SMAD4, SNAI2, SP1, CREB1, MITF для Stamatiou.

Генная сеть для ТФ из NHL



Пересечение ТФ выборок






из ресурса
Hereditary
Hearing Loss

из статьи Stamatiou

Генная сеть регуляции (ТФ-гены)

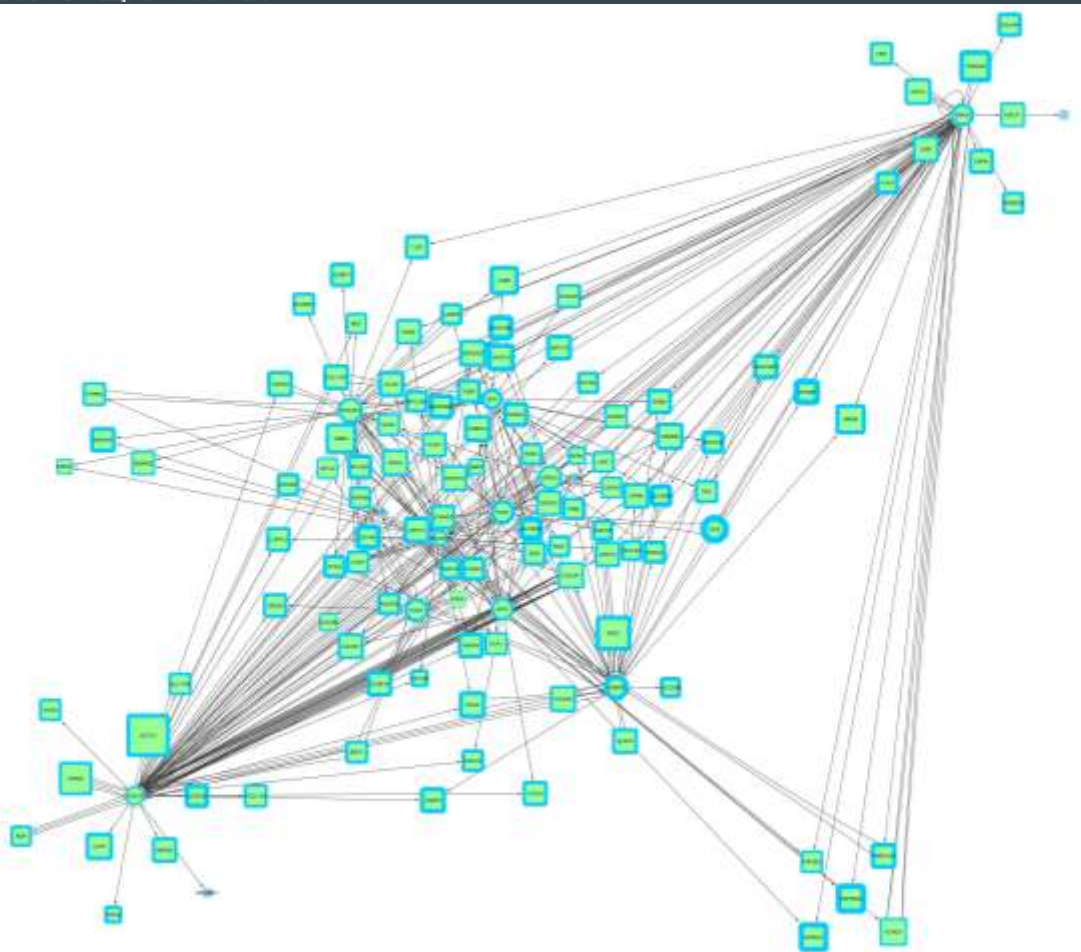
Данные для 1000 п.н. upstream, выборка генов из Hereditary Hearing Loss, $p < 5 \cdot 10^{-5}$






Легенда		
Форма		Ген
		ТФ
Толщина рамки	%SNP в промоторе	
Размер узла	%SNP в гене	
Цвет узла		Ген из выборки

Генная сеть регуляции (ТФ-гены)

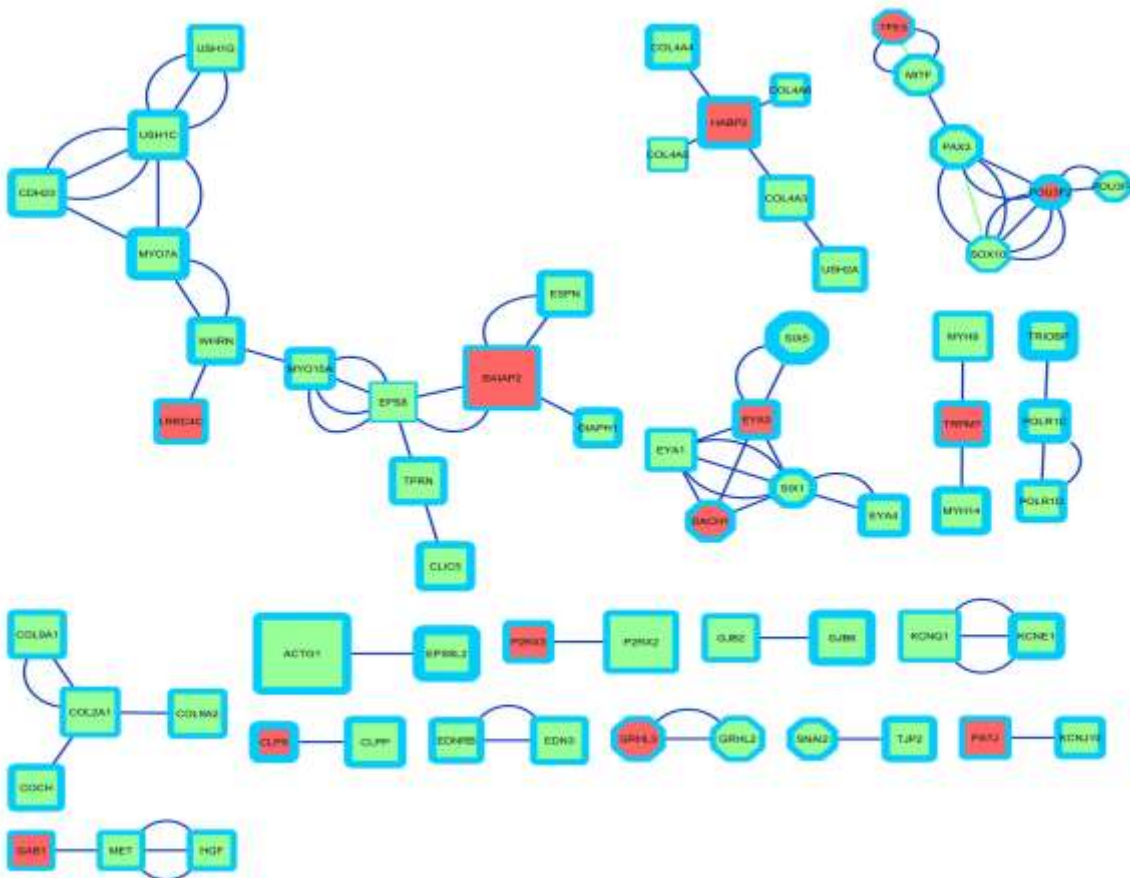
Данные для 2000 п.н. upstream, выборка генов из Hereditary Hearing Loss, $p < 5 \cdot 10^{-5}$



Легенда		
Форма		Ген
		ТФ
Толщина рамки	%SNP в промоторе	
Размер узла	%SNP в гене	
Цвет узла		Ген из выборки

Сеть белок-белковых взаимодействий

Выборка генов из Hereditary Hearing Loss, (GeneMANIA -> Cytoscape), weight ≥ 0.005



Легенда		
форма		Ген
		ТФ
Толщина рамки	%SNP в промоторе	
Размер узла	%SNP в гене	
Цвет узла		Ген из выборки
		Новый ген

Ген BAIP2, не из выборки, SNP в гене – 4.6%, в промоторе – 2.1%

Транскрипционный фактор SIX5, из выборки, SNP в гене – 3.6%, в промоторе – 5.1%