

Реконструкция и компьютерный анализ ассоциативной генной сети человека, контролирующей аппетит

Матросова Екатерина Алексеевна,

магистрант КИБ ФЕН НГУ, направление подготовки 06.04.01

Научный руководитель Игнатьева Е.В.,

к.б.н., с.н.с. лаборатории эволюционной биоинформатики
и теоретической генетики ИЦиГ СО РАН

Новосибирск, 2021 г.

Нарушения аппетита



Нервная
анорексия
(*anorexia
nervosa*)



Гиперфагия
(булемия)

Избыточный
вес, ожирение



Возможные осложнения анорексии:

- дегидратация
- внезапная смерть в результате остановки сердца
- синдром возобновления питания (рефидинг-синдром)

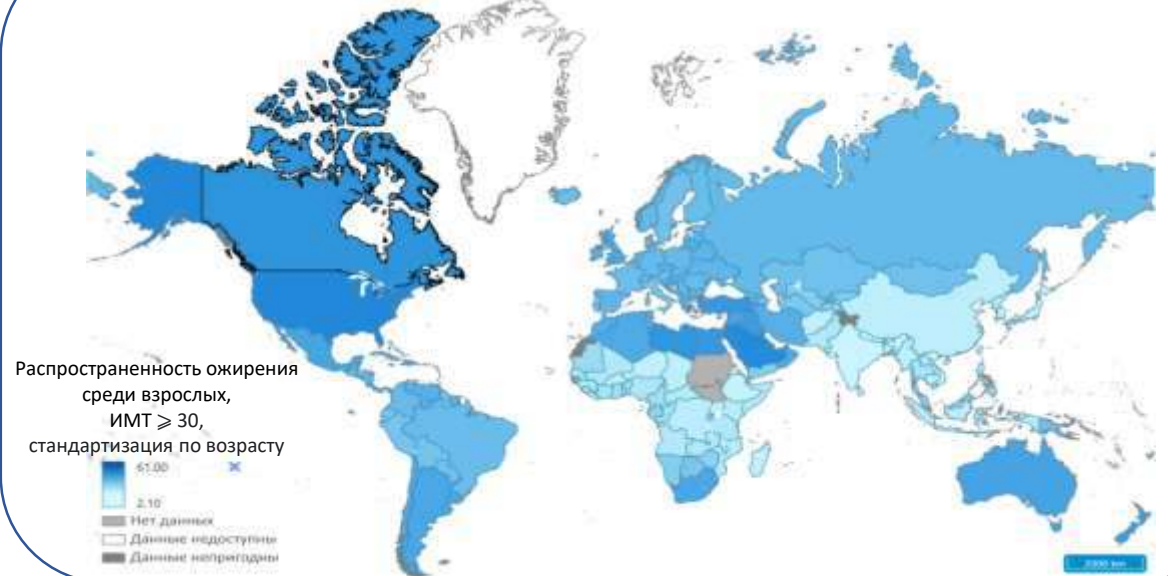
Ожирение – фактор риска для:

- сахарного диабета 2 типа
- сердечно-сосудистых заболеваний
- нарушения опорно-двигательного аппарата
- тяжелая форма COVID-19

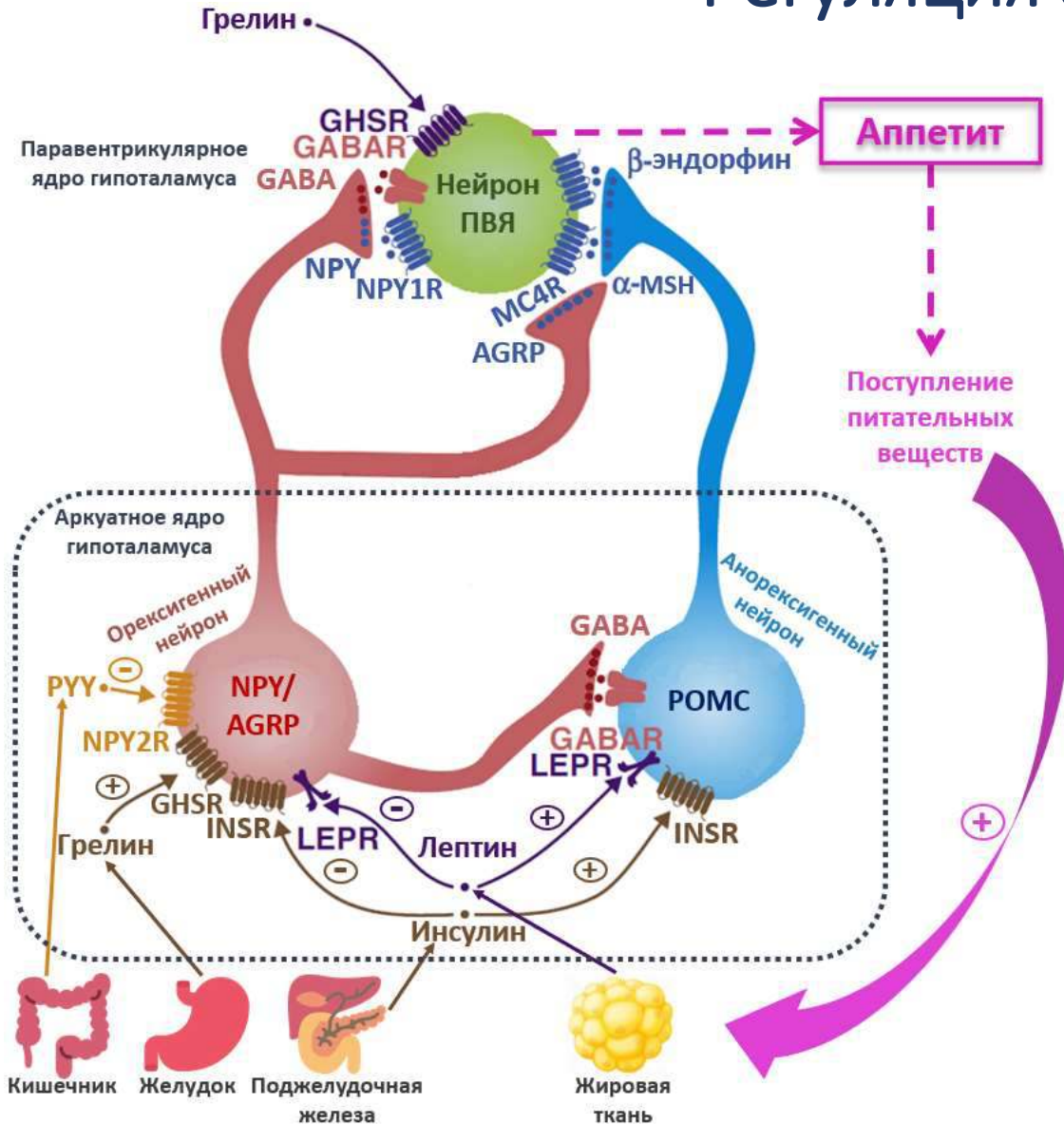
- Распространение нервной анорексии:
0.9-2% среди женщин, 0.1-0.3% среди мужчин.
(Stice E., Bohon C. Child and Adolescent Psychopathology. 2012.)
- Анорексия – третье по распространенности хроническое заболевание после астмы и ожирения среди подростков.
(Eating Disorders Coalition. Facts About Eating Disorders: What the Research Shows.)

- Нервная анорексия повышает риск смерти в 10 раз у молодых людей.
(Fichter M. M., Quadflieg N. International Journal of Eating Disorders. 2016.)

Распространенность ожирения по данным ВОЗ (за 2021 год)



Регуляция аппетита



Нейроны аркуатного ядра гипоталамуса, регулирующие аппетит:

1. **Орксигенные** – стимулируют аппетит, вырабатывая и высвобождая NPY, AGRP и GABA.
2. **Анорксигенные** – подавляют аппетит, вырабатывая POMC.

Гормоны, стимулирующие аппетит:

Адипонектин
Грелин

Гормоны, подавляющие аппетит:

Лептин
Инсулин
Холецистокинин
PYY

Схематичное изображение основных органов и структур, участвующих в регуляции пищевого поведения

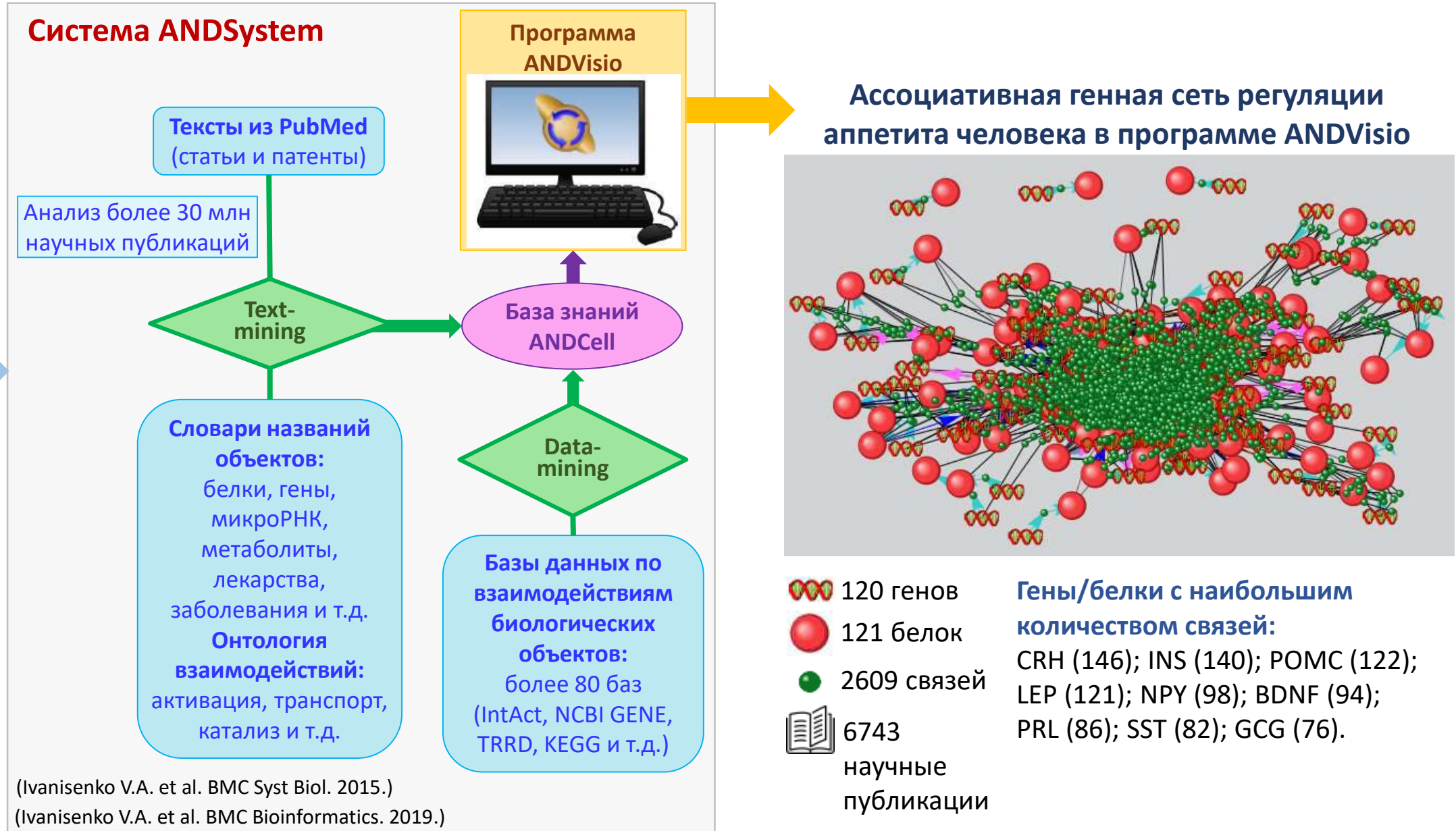
(Heisler L.K., Lam D.D. Current Opinion in Pharmacology. 2017.)

(Нельсон Д., Кокс М. Основы биохимии Ленинджера. 2014.)

Цель и задачи

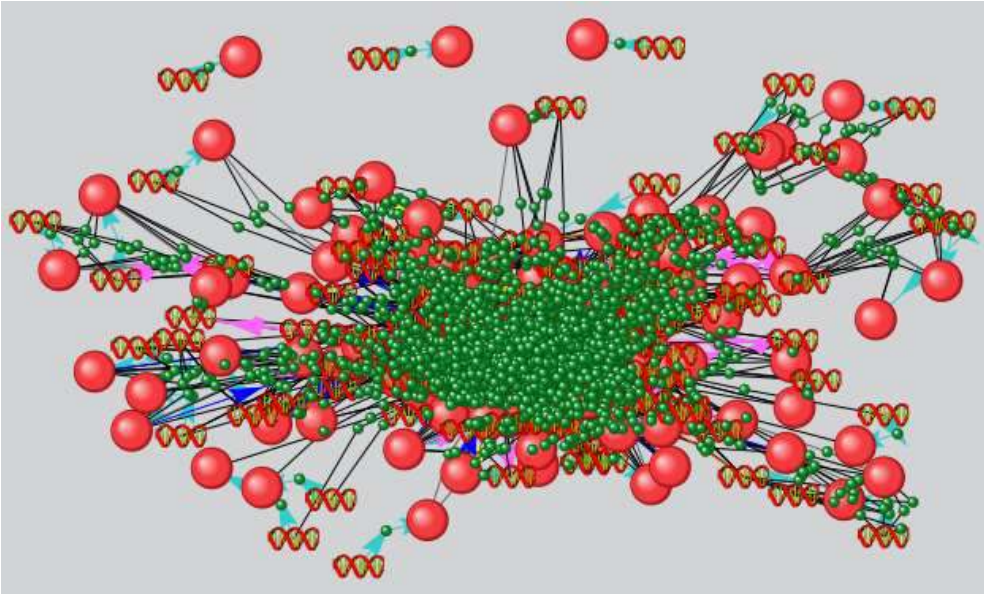
- **Цель:** реконструкция и компьютерный анализ ассоциативной генной сети человека, контролирующей аппетит.
- **Задачи:**
 1. Реконструкция ассоциативной генной сети регуляции аппетита человека с помощью системы ANDSystem.
 2. Компьютерный анализ полученной сети:
 - а) Выявление микроРНК, регулирующих экспрессию белков сети;
 - б) Выявление заболеваний, ассоциированных с генами сети;
 - в) Выявление особенностей эволюции генов сети, основываясь на индексах PAI (индекс филостратиграфического возраста) и DI (индекс эволюционной изменчивости).




Реконструкция ассоциативной генной сети регуляции аппетита человека с помощью системы ANDSystem




Выявление микроРНК, регулирующих экспрессию белков сети

Ассоциативная генная сеть регуляции аппетита человека в программе ANDVisio



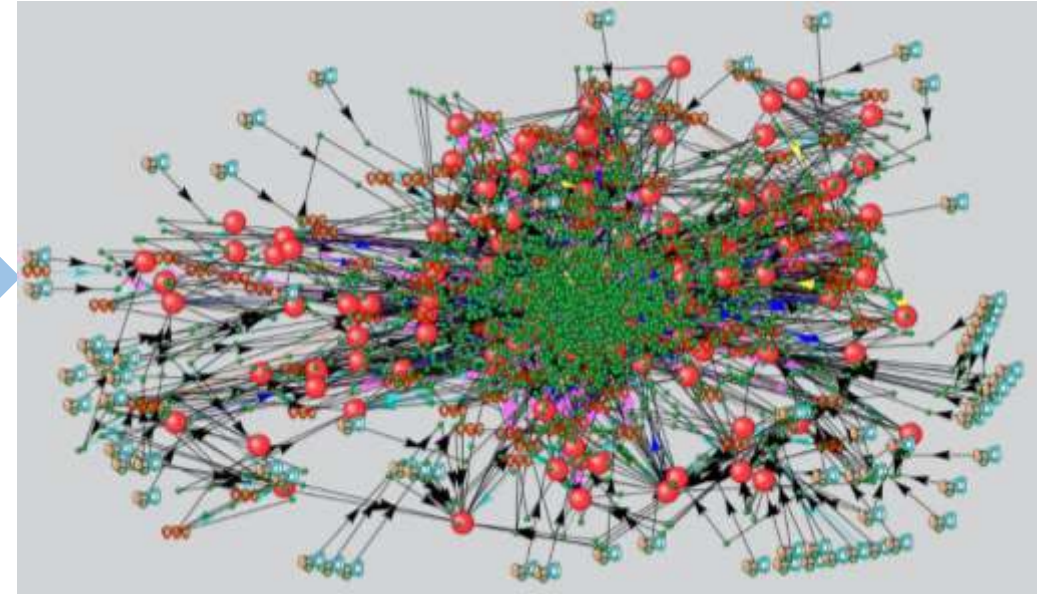
 120 генов
 121 белок
 2609 связей

 6743
 научные
 публикации

Поиск микроРНК,
 регулирующих
 экспрессию белков
 сети, с помощью
 системы ANDSystem



Ассоциативная генная сеть регуляции аппетита человека, расширенная для микроРНК, в программе ANDVisio



64 связи микроРНК с белками сети, экспериментально доказанные для человека:
 54 микроРНК,
 28 регулируемых белков



МикроРНК с наибольшим количеством связей:
 MIR877 (4); MIR215 (3); MIR192 (3)

Визуализация центральной части ассоциативной геной сети регуляции аппетита человека в системе Cytoscape



ANDSystem



Данные о 54 микроРНК (с экспертной проверкой)



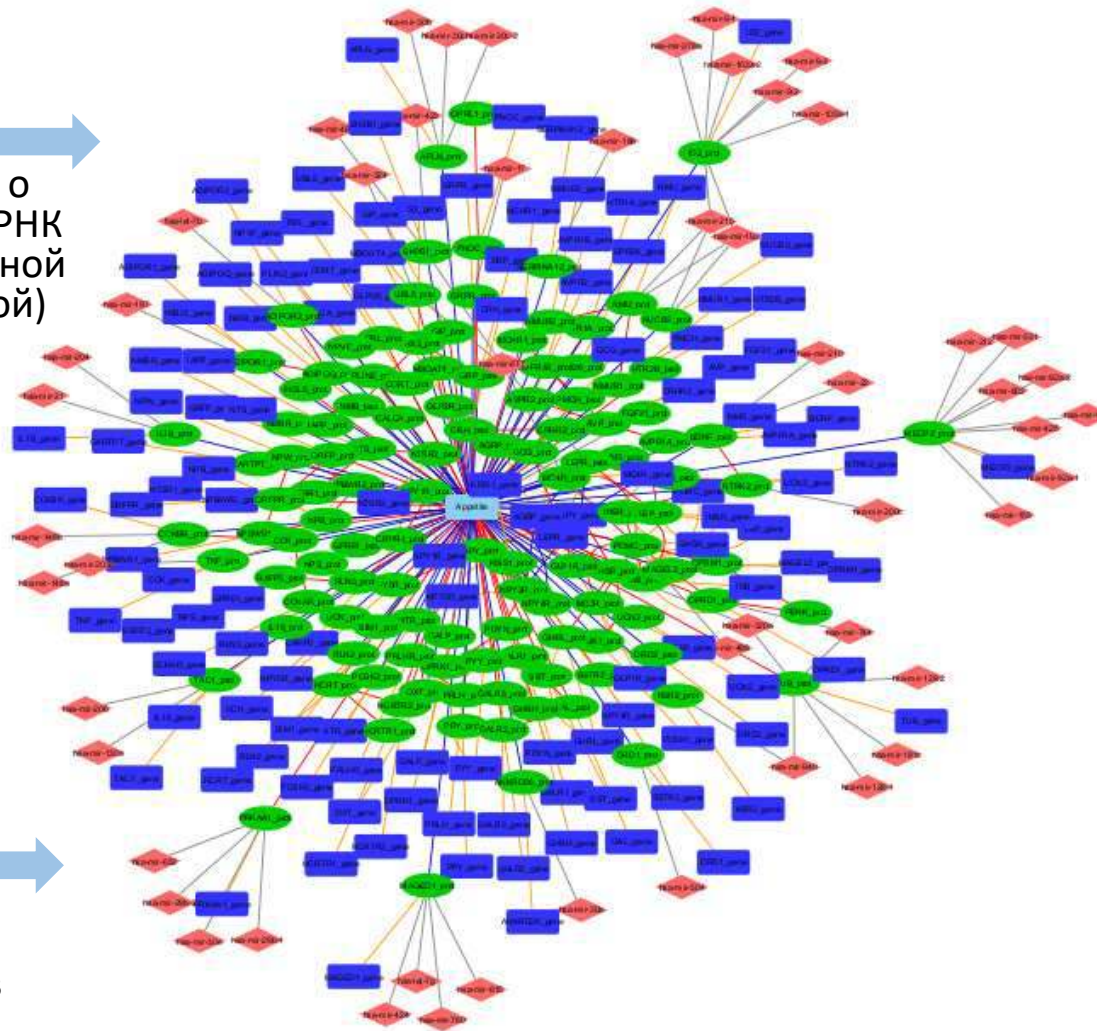
База данных GeneNet (ИЦиГ СО РАН)






Данные о регуляторных связях между генами и белками



Научные публикации

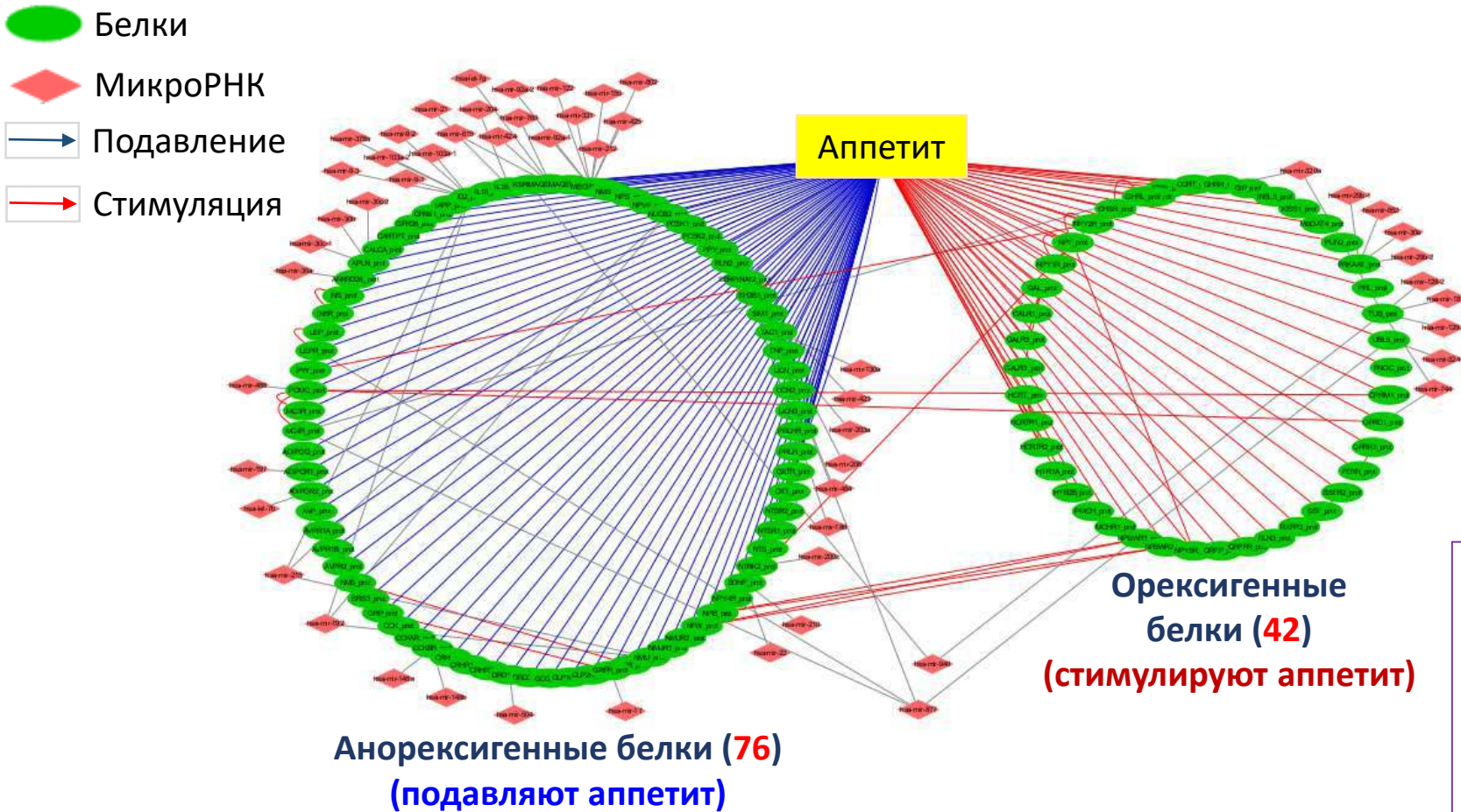
Данные о влиянии 120 генов и белков на аппетит



-  120 генов
-  121 белок
-  54 микроРНК
-  394 связи
-  92 научные публикации

Центральная часть ассоциативной геной сети регуляции аппетита человека, представленная в Cytoscape

Анорексигенные и орексигенные белки в центральной части ассоциативной генной сети регуляции аппетита человека



Распределение белков по эффекту на аппетит

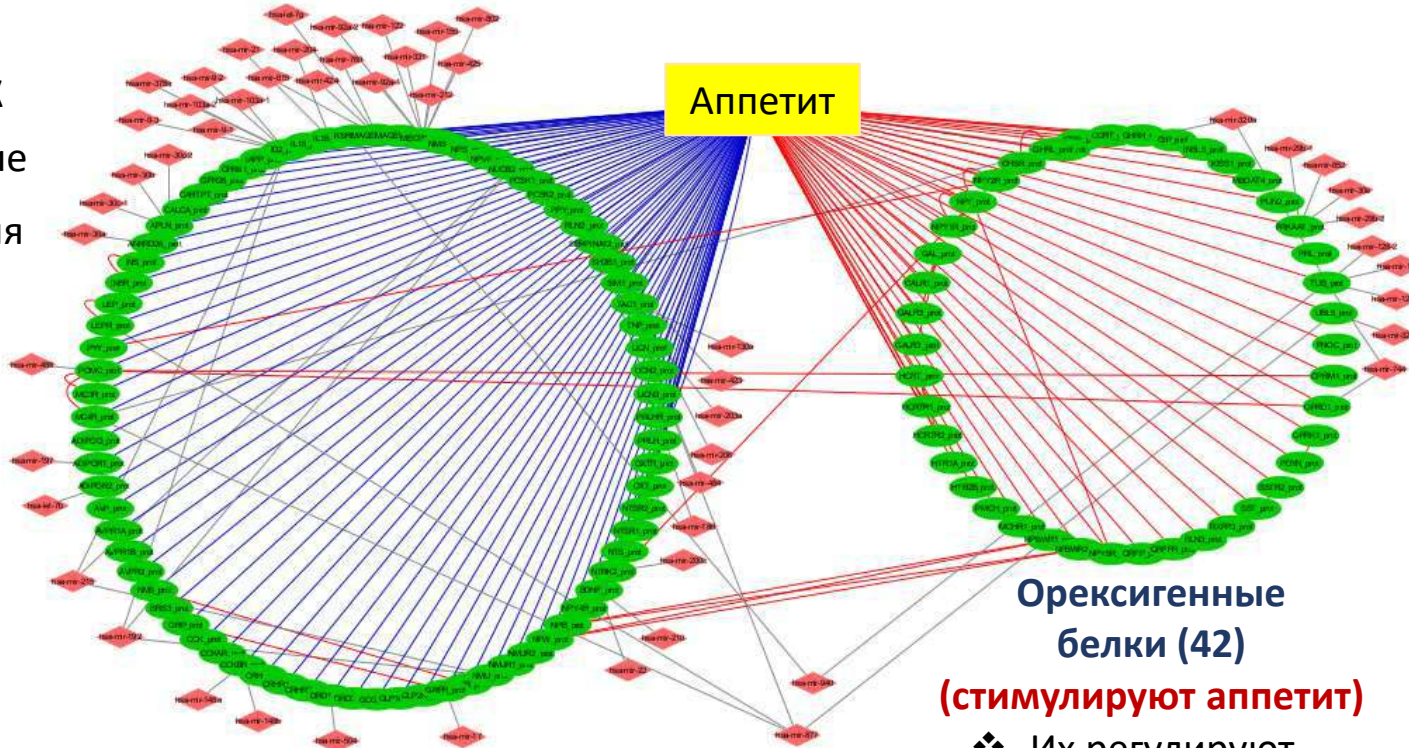


Наблюдаемые количества анорексигенных (76) и орексигенных (42) белков достоверно выше и ниже, соответственно, их количества, ожидаемого при равномерном биномиальном распределении $\text{Bin}(n; p=q=1-p=1/2)$:
уровень статистической значимости $\alpha < 0.01$

	Общее количество (n)	Анорексигенные белки (N_{\downarrow})	Орексигенные белки (N_{\uparrow})	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv q(N_{\uparrow}) \equiv 1/2)$
Ожидаемое	118	$118/2 = 59$	$118/2 = 59$	-
Наблюдаемое	118	76	42	< 0.01

МикроРНК, регулирующие экспрессию анорексигенных и орексигенных белков генной сети

- Белки
- ◆ МикроРНК
- Подавление
- Стимуляция



**Анорексигенные белки (76)
(подавляют аппетит)**

- ❖ Их регулируют **44** микроРНК

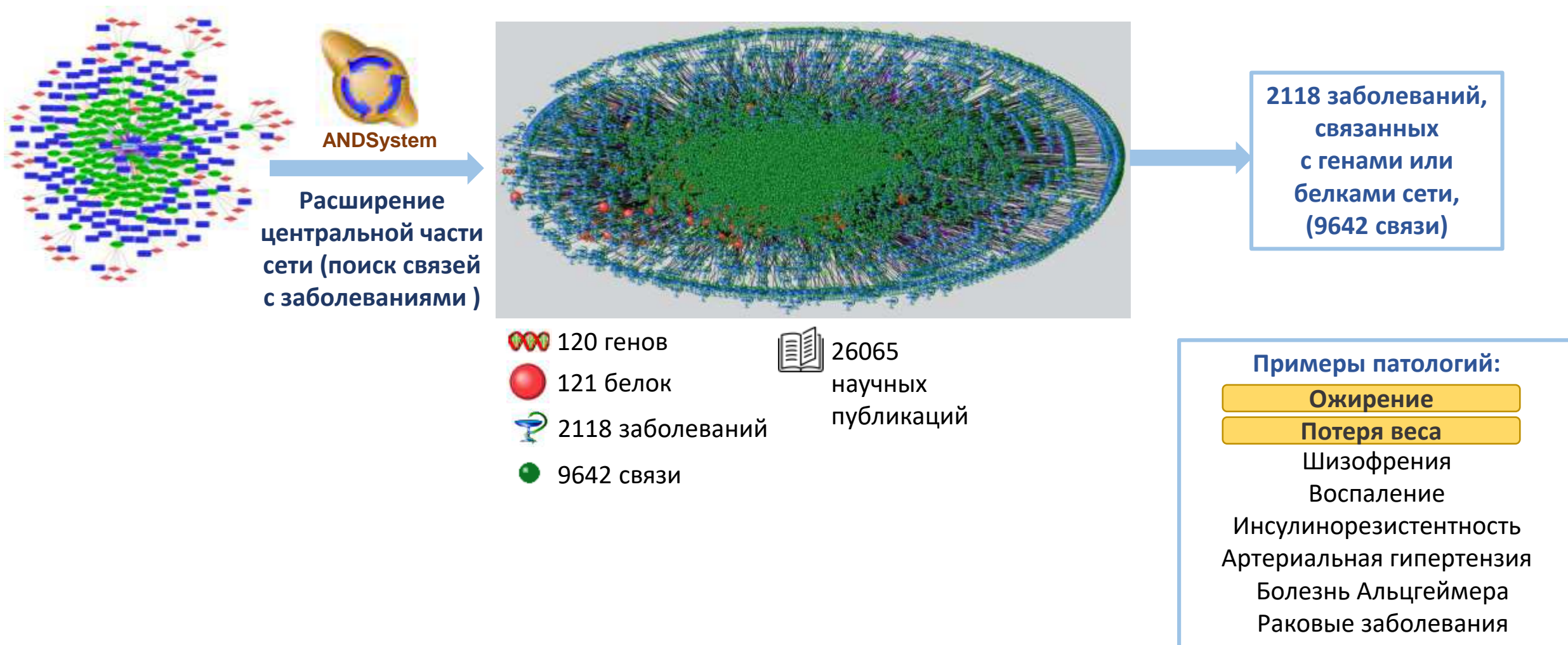
**Орексигенные белки (42)
(стимулируют аппетит)**

- ❖ Их регулируют **12** микроРНК

Наблюдаемые количества микроРНК, регулирующих только анорексигенные (**42**) либо только орексигенные (**10**) белки, достоверно выше и ниже, соответственно, их количества, ожидаемого при биномиальном распределении, учитывающем соотношение общих количеств анорексигенных и орексигенных белков, $\text{Bin}(n; p=76/118*n, q=1-p=42/118*n)$: уровень статистической значимости **$\alpha < 0.01$**

	Общее количество (n)	МикроРНК, регулирующие только анорексигенные белки (N_{\downarrow})	МикроРНК, регулирующие только орексигенные белки (N_{\uparrow})	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv 76*n/118) = \text{Bin}(q(N_{\uparrow}) \equiv 42*n/118)$
Ожидаемое	52	$52*76/(76+42) = 33.49$	$52*42/(76+42) = 18.51$	-
Наблюдаемое	52	$44-2 = \mathbf{42}$	$12-2 = \mathbf{10}$	< 0.01

Выявление заболеваний, ассоциированных с генами или белками сети



Заболевания, связанные с анорексигенными и орексигенными генами сети

	Общее количество заболеваний, связанных со всеми генами сети (n)	Среднее количество заболеваний, связанных			<i>P-value</i> (по критерию χ^2)
		со всеми генами сети	с анорексигенными генами (N_A)	с орексигенными генами (N_O)	
Ожидаемое	9642	9642/120 = 80.35	9642/120 = 80.35	9642/120 = 80.35	-
Наблюдаемое	9642	82	103	44	<<< 0.01

Термины, связанные преимущественно с анорексигенными генами:

Повышенный вес

Гипертония

Воспаление

Сердечная недостаточность

Сепсис

Сахарный диабет

Аутистическое расстройство

Хроническая обструктивная
болезнь лёгких

Почечная недостаточность

Раковые заболевания

Термины, связанные преимущественно с орексигенными генами:

Зависимость от
психоактивных веществ

Эпилепсия

Головная боль

Зависимость от
алкоголя

Невралгия

Нейробластома

Энцефаломиелит

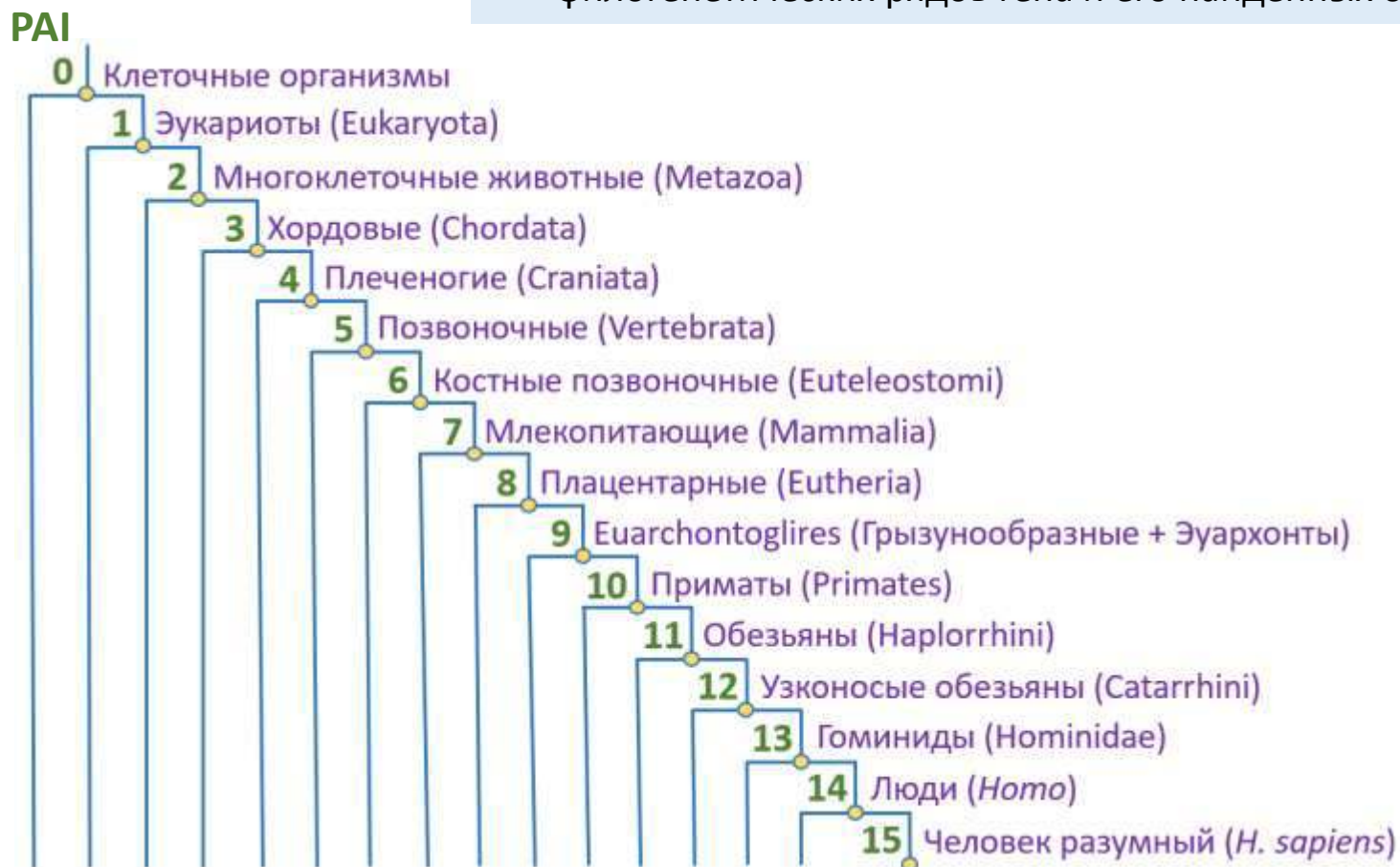
Простатит

Катаплексия

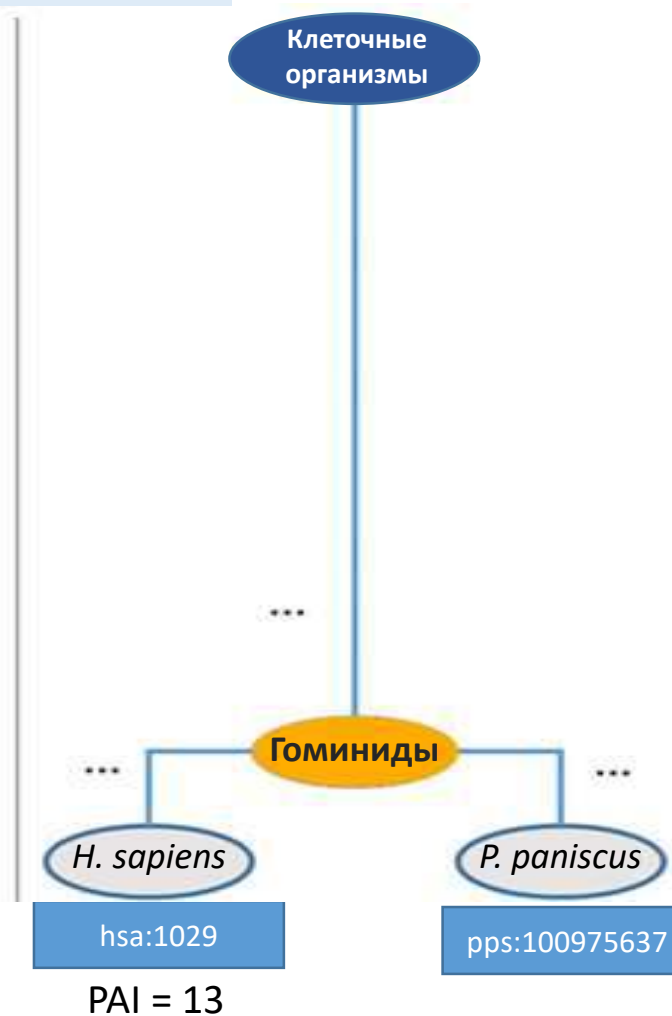
Эволюционный возраст генов

Филостратиграфический индекс гена (PAI = phylostratigraphic age index)

рассчитывается как порядковый номер наиболее удаленного (от корня дерева) таксона, на котором начинаются расхождения филогенетических рядов гена и его найденных ортологов.



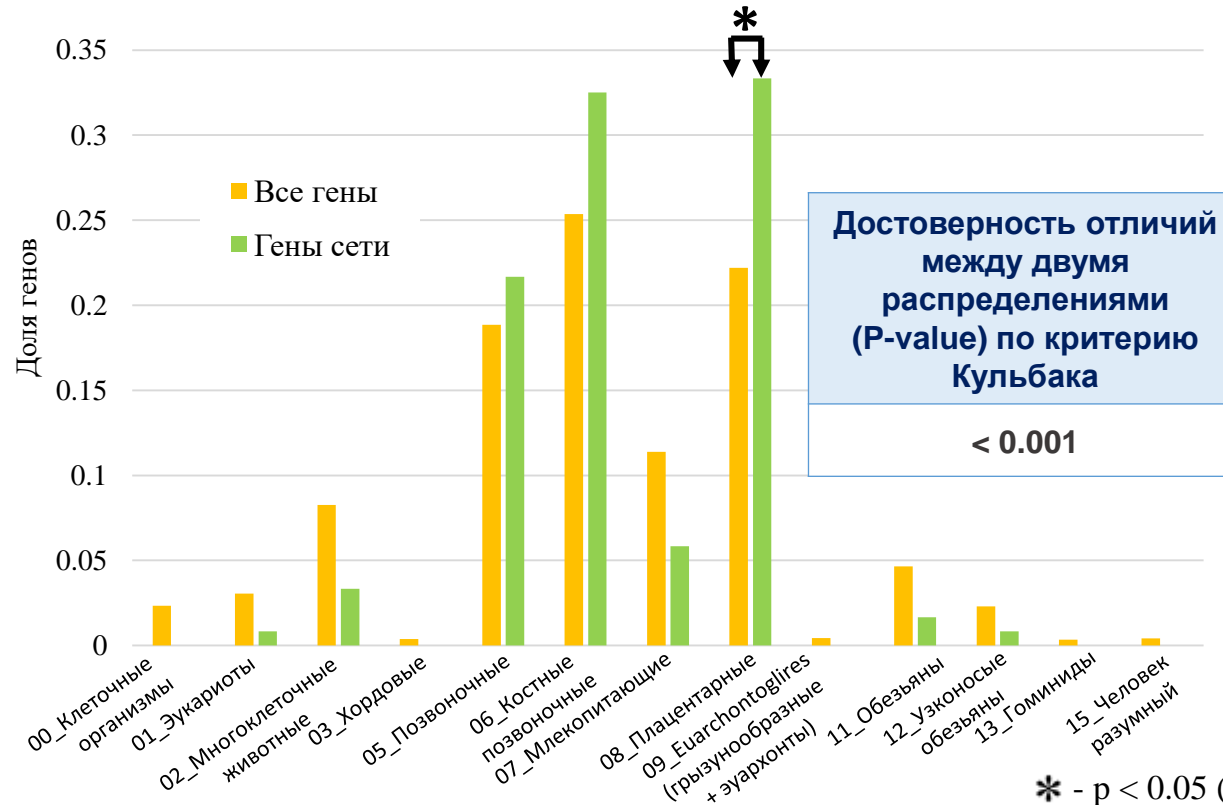
(Mustafin Z.S. et al. BMC Bioinformatics. 2017.)



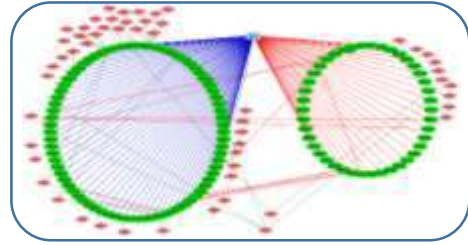
Orthoweb
(ИЦИГ СО РАН)

Филостратиграфические
индексы (PAI) генов
(идентичность = 0.7)

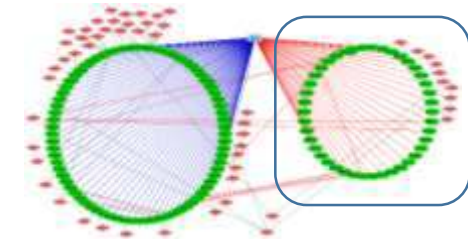
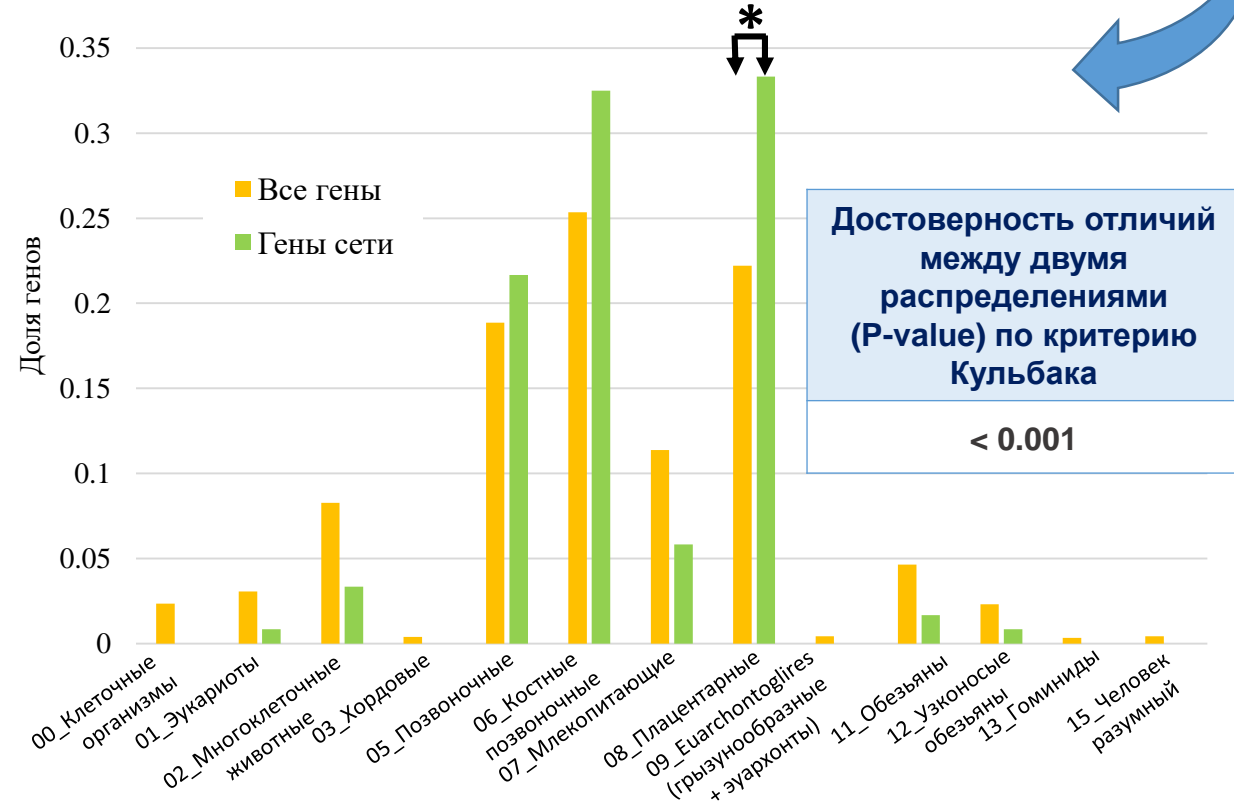
Распределение генов сети (120) и белок-кодирующих
генов генома человека (19556) по индексу PAI



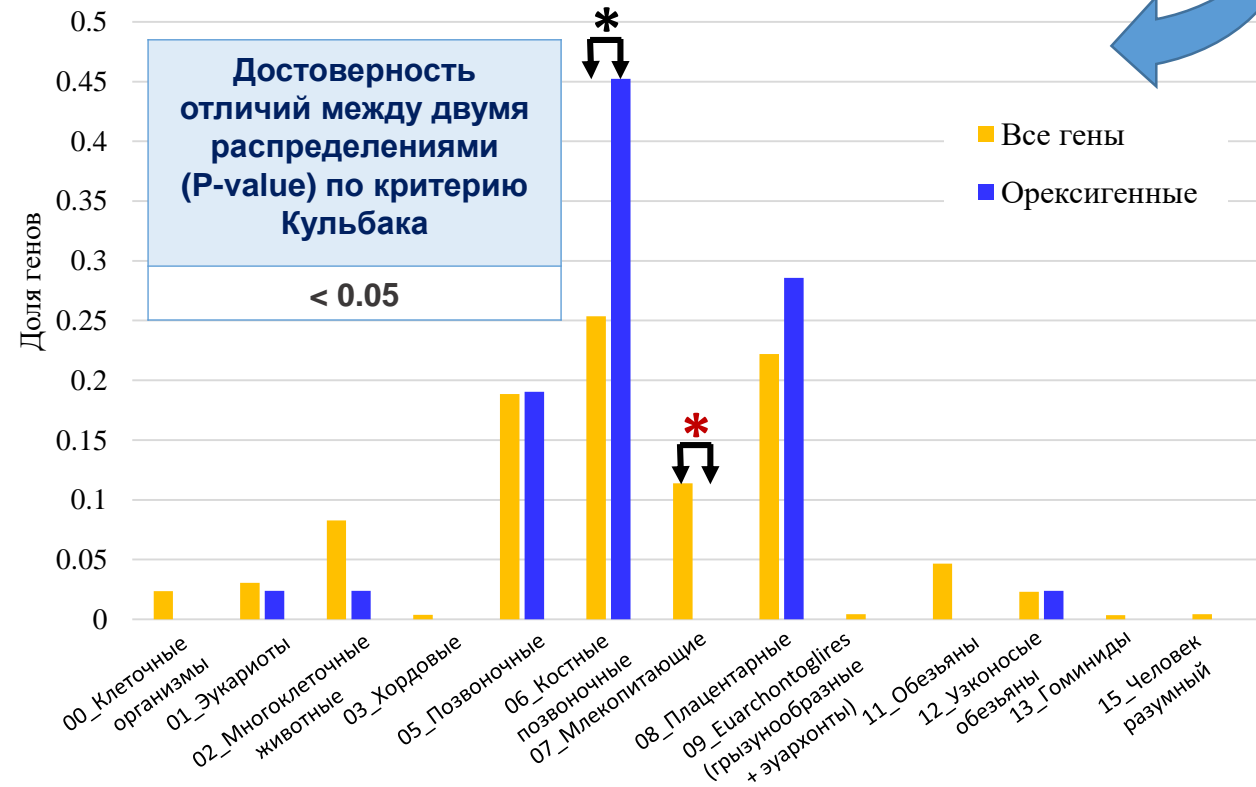
Эволюционный возраст генов генной сети регуляции аппетита



Распределение генов сети (120) и белок-кодирующих генов генома человека (19556) по индексу PAI



Распределение орексигенных генов сети (42) и белок-кодирующих генов генома человека (19556) по индексу PAI

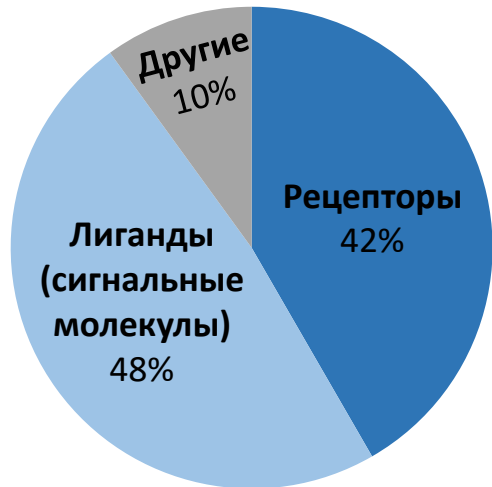


* - $p < 0.05$ (по критерию χ^2 с поправкой Бонферрони)

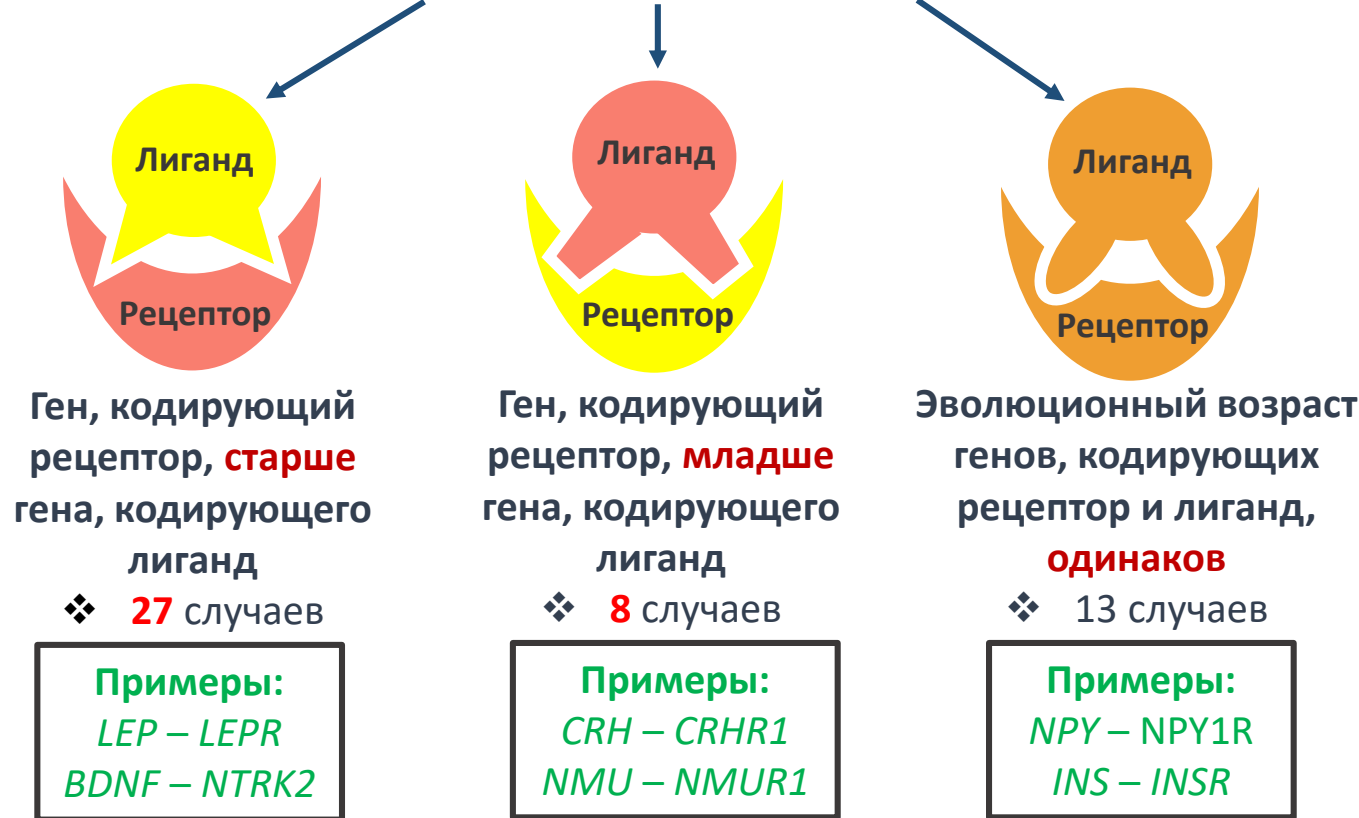
* - $p < 0.05$ (по критерию Кульбака с поправкой Бонферрони)

Эволюционный возраст генов, кодирующих лиганды и рецепторы генной сети

Распределение белков по молекулярным функциям



48 пар лиганд-рецептор в сети



- Молодой ген
- Древний ген
- Ген среднего возраста

Наблюдаемые количества случаев, когда ген, кодирующий рецептор, старше (**27**) либо младше (**8**) гена, кодирующего лиганд, достоверно выше и ниже, соответственно, их количества, ожидаемого при равномерном биномиальном распределении $\text{Bin}(n; p=q=1-p=1/2)$: уровень статистической значимости $\alpha < 0.001$

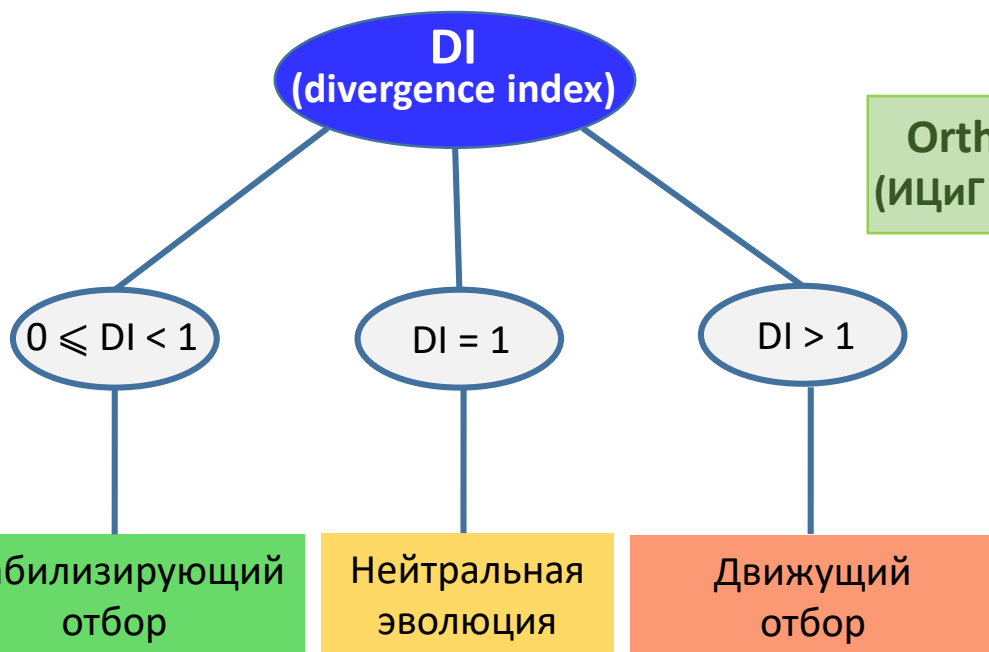
	Общее количество пар лиганд-рецептор с разным возрастом (n)	Лиганды, возраст генов которых меньше возраста генов рецепторов (N_{\downarrow})	Лиганды, возраст генов которых больше возраста генов рецепторов (N_{\uparrow})	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv q(N_{\uparrow}) \equiv 1/2)$
Ожидаемое	35	$35/2 = 17.5$	$35/2 = 17.5$	-
Наблюдаемое	35	27	8	< 0.001

Эволюционная изменчивость генов генной сети регуляции аппетита

Индекс эволюционной изменчивости гена
(DI = divergence index)

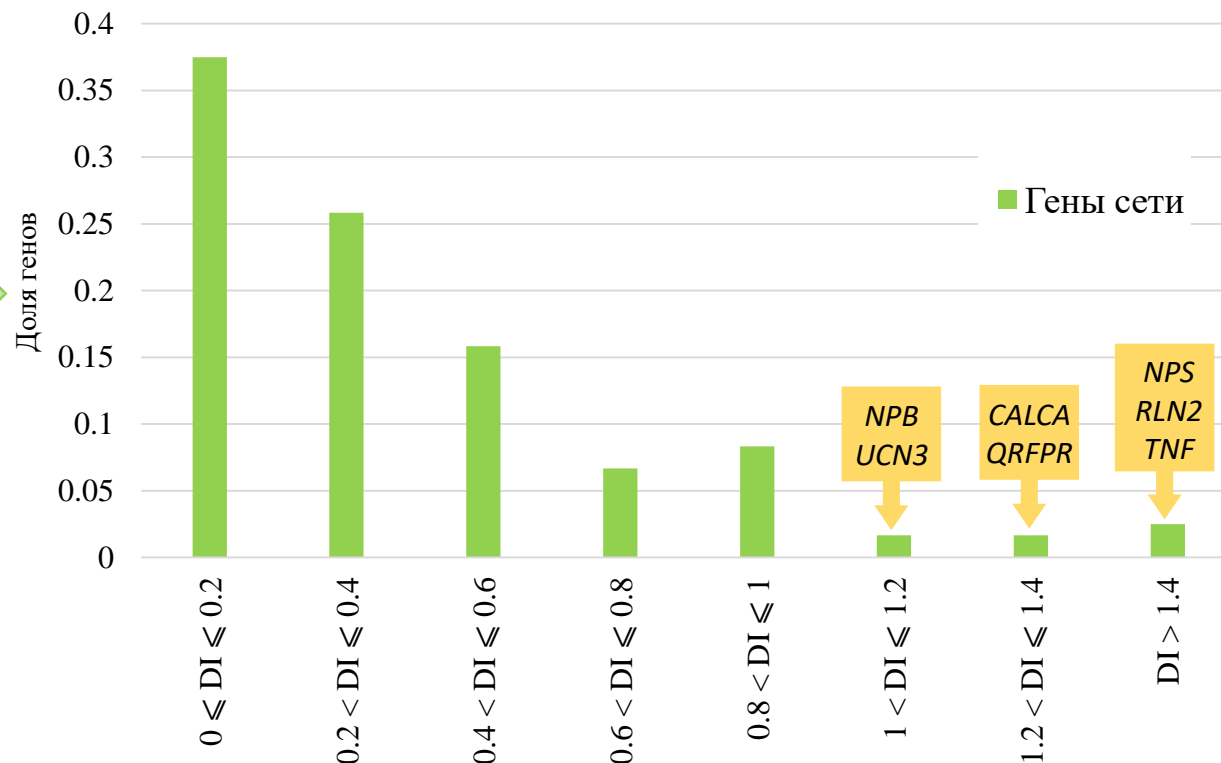
вычисляется на основании отношения dN/dS :

- dN – доля несинонимичных замен;
- dS – доля синонимичных замен.



Orthoweb
(ИЦИГ СО РАН)

Распределение генов сети (120) по индексу DI

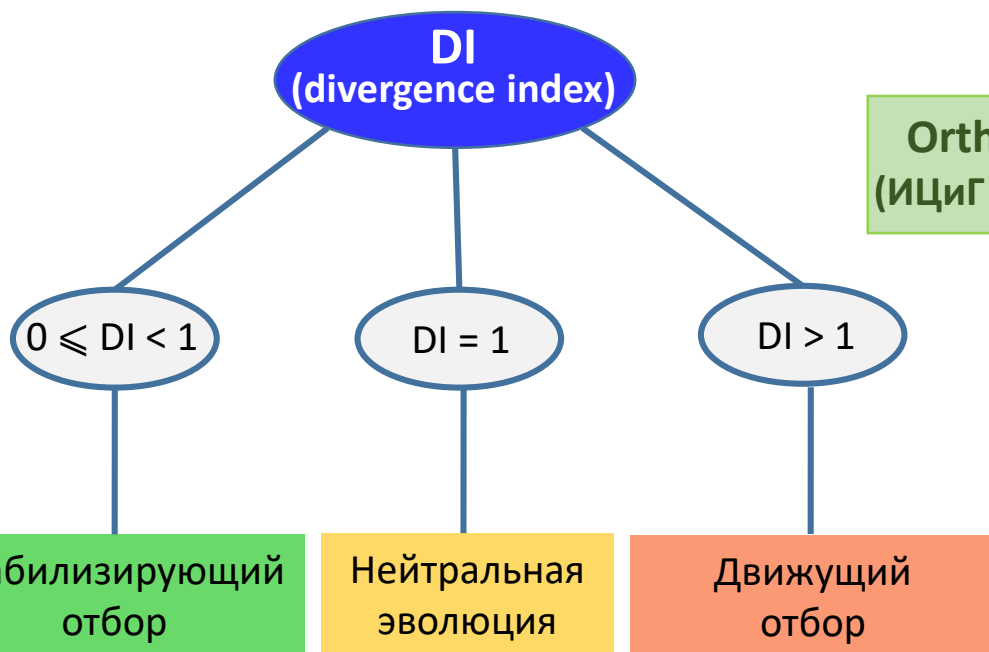


Эволюционная изменчивость генов генной сети регуляции аппетита

Индекс эволюционной изменчивости гена
(DI = divergence index)

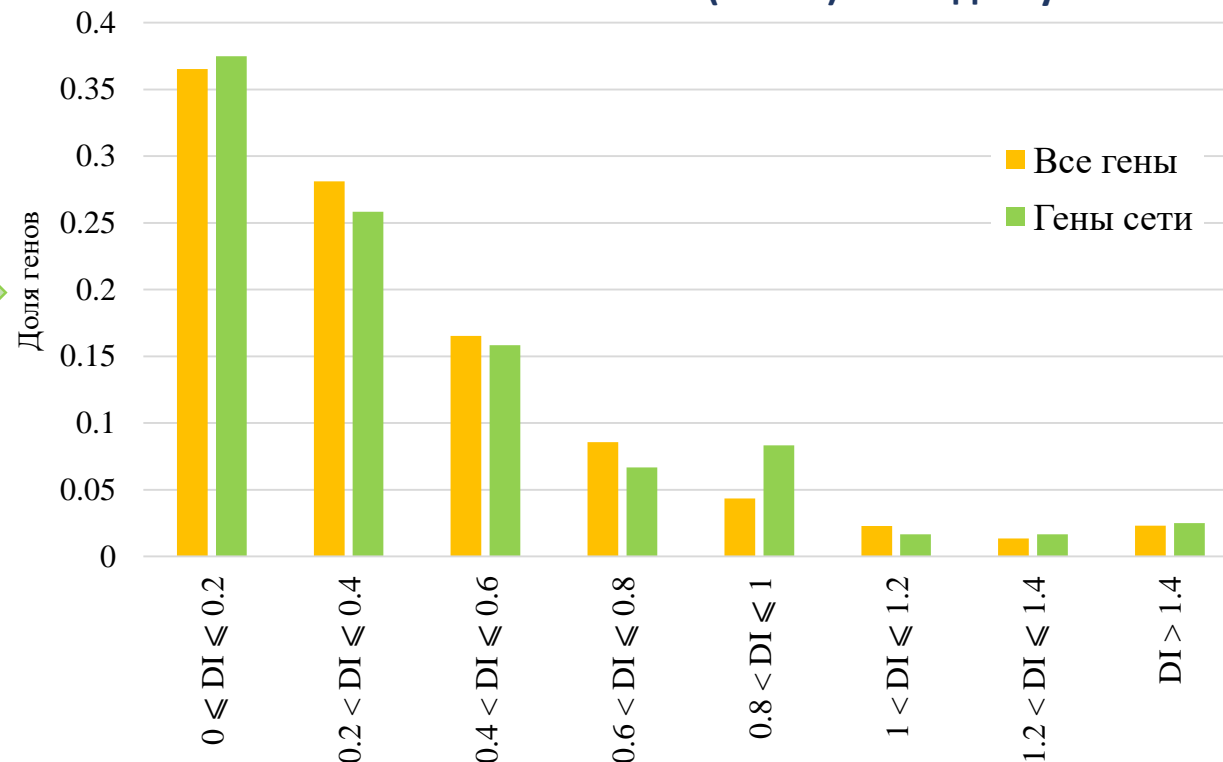
вычисляется на основании отношения dN/dS :

- dN – доля несинонимичных замен;
- dS – доля синонимичных замен.



Orthoweb
(ИЦИГ СО РАН)

Распределение генов сети (120) и белок-кодирующих генов генома человека (19556) по индексу DI



Эволюционная изменчивость генов генной сети регуляции аппетита

Распределение генов сети, кодирующих лиганды, (51) и белок-кодирующих генов генома человека (19556) по индексу DI



Распределение генов сети, кодирующих лиганды (51) и рецепторы (50), по индексу DI



Выводы

1. Построена ассоциативная генная сеть регуляции аппетита человека, включающая 120 генов, 121 белок, 54 микроРНК, 2118 заболеваний и 12315 связей.
2. Система генов, связанных с подавлением аппетита, включает больше элементов, обладает дополнительным механизмом регуляции экспрессии с участием микроРНК и в среднем связана с большим количеством заболеваний по сравнению с системой, связанной с усилением аппетита.
3. Преимущественно с анорексигенными генами связаны такие патологии, как повышенный вес, воспалительные заболевания и гипертония. А в основном с орексигенными генами связаны такие заболевания, как зависимость от психоактивных веществ, эпилепсия, головная боль и алкоголизм.
4. В генной сети регуляции аппетита преобладают гены среднего эволюционного возраста. При этом гены сети содержат значимо более высокую долю (33%) генов, имеющих PAI, равный 8 (Плацентарные), чем все белок-кодирующие гены генома человека (22%). Вместе с тем для орексигенных генов показано, что гены, возникшие на эволюционном уровне 6 (Костные позвоночные), встречаются в этой группе значимо чаще, чем в полном наборе белок-кодирующих генов, а гены, возникшие на эволюционном уровне 7 (Млекопитающие), напротив, не встречаются вовсе.
5. Количество случаев, когда ген, кодирующий рецептор клеточной поверхности, старше гена, кодирующего соответствующий ему лиганд, значимо превышает количество случаев, когда наблюдается обратная ситуация.
6. Большинство генов сети регуляции аппетита подвержены стабилизирующему отбору. Примечательно, что гены, кодирующие лиганды, имеют значимо более высокие индексы DI (то есть менее подвержены стабилизирующему отбору) как по сравнению с генами, кодирующими рецепторы клеточной поверхности, так и по сравнению со всеми белок-кодирующими генами генома человека.

Благодарности

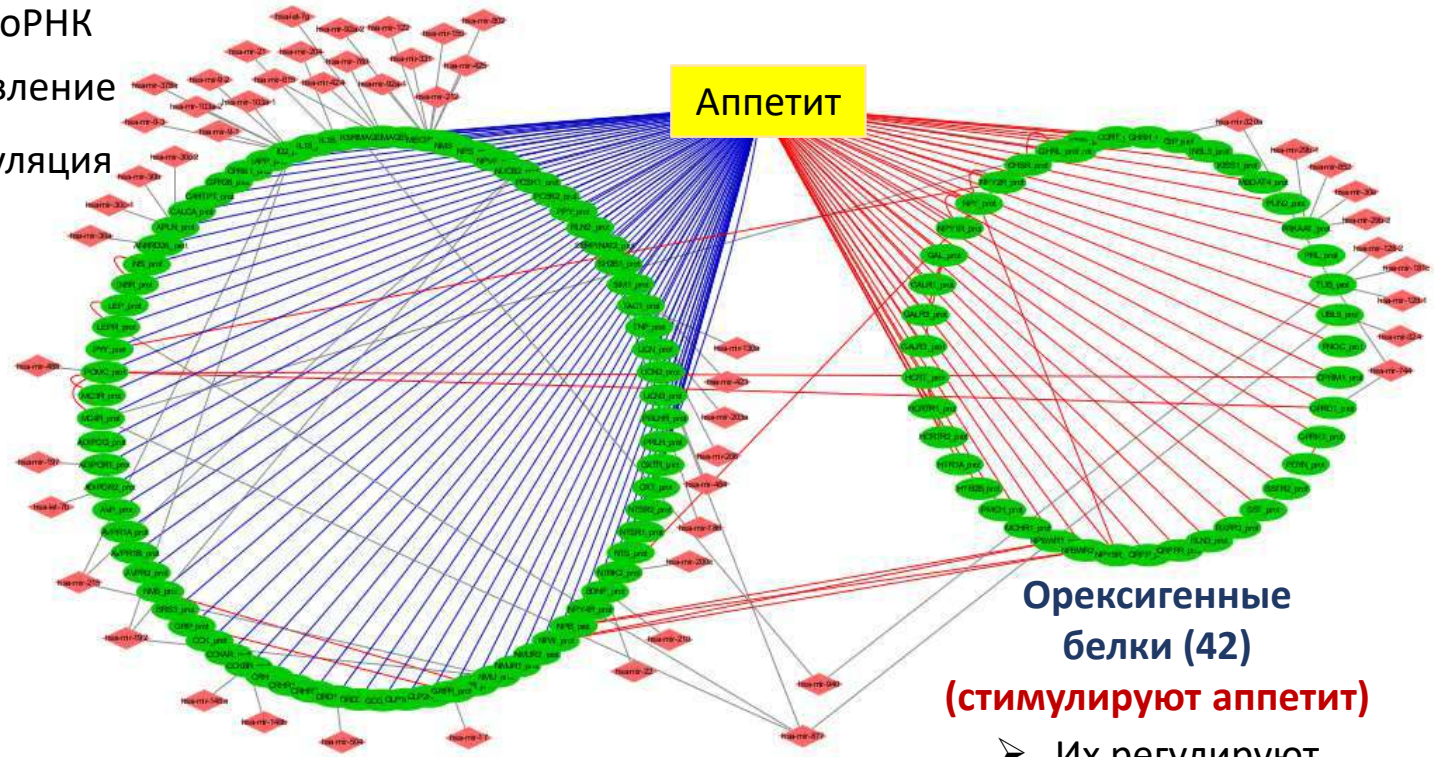
- к.т.н. Деменкову Павлу Сергеевичу
- д.б.н. Ефимову Вадиму Михайловичу
- к.б.н. Лашину Сергею Александровичу
- Мустафину Захару Сергеевичу
- д.б.н. Пономаренко Михаилу Павловичу
- к.б.н. Вишневскому Олегу Владимировичу
- к.ф.-м.н. Титову Игорю Ивановичу
- к.б.н. Левицкому Виктору Георгиевичу
- к.б.н. Иванисенко Владимиру Александровичу
- Подкоłodному Николаю Леонтьевичу
- к.б.н. Подкоłodной Ольге Александровне
- к.б.н. Акбердину Илье Ринатовичу
- рецензенту к.б.н. Юдину Николаю Серафимовичу
- академику РАН, д.б.н. Колчанову Николаю Александровичу



Спасибо за внимание!

МикроРНК, регулирующие экспрессию анорексигенных и орексигенных белков генной сети

- Белки
- ◆ МикроРНК
- Подавление
- Стимуляция



**Анорексигенные белки (76)
(подавляют аппетит)**

➤ Их регулируют **44** микроРНК

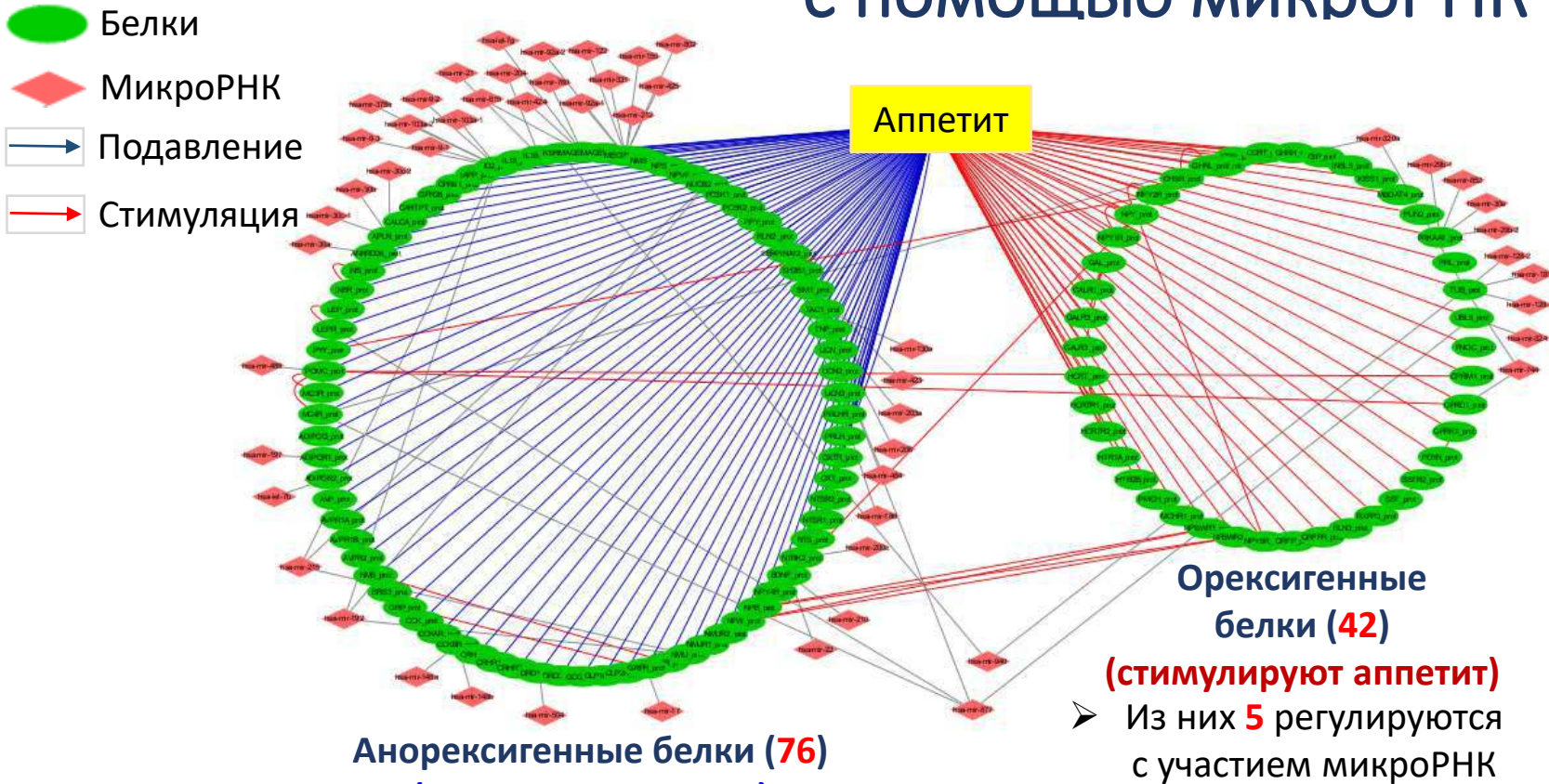
**Орексигенные белки (42)
(стимулируют аппетит)**

➤ Их регулируют **12** микроРНК

Наблюдаемые количества микроРНК, регулирующих только анорексигенные (**42**) либо только орексигенные (**10**) белки, достоверно выше и ниже, соответственно, их количества, ожидаемого при равномерном биномиальном распределении $\text{Bin}(n; p=q=1-p=1/2)$: уровень статистической значимости **$\alpha \lll 0.001$**

	Общее количество (n)	МикроРНК, регулирующие только анорексигенные белки (N_{\downarrow})	МикроРНК, регулирующие только орексигенные белки (N_{\uparrow})	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv q(N_{\uparrow}) \equiv 1/2)$
Ожидаемое	52	$52/2 = 26$	$52/2 = 26$	-
Наблюдаемое	52	$44-2 = \mathbf{42}$	$12-2 = \mathbf{10}$	$\lll 0.001$

Количество анорексигенных и орексигенных белков сети, для которых была выявлена регуляция экспрессии с помощью микроРНК



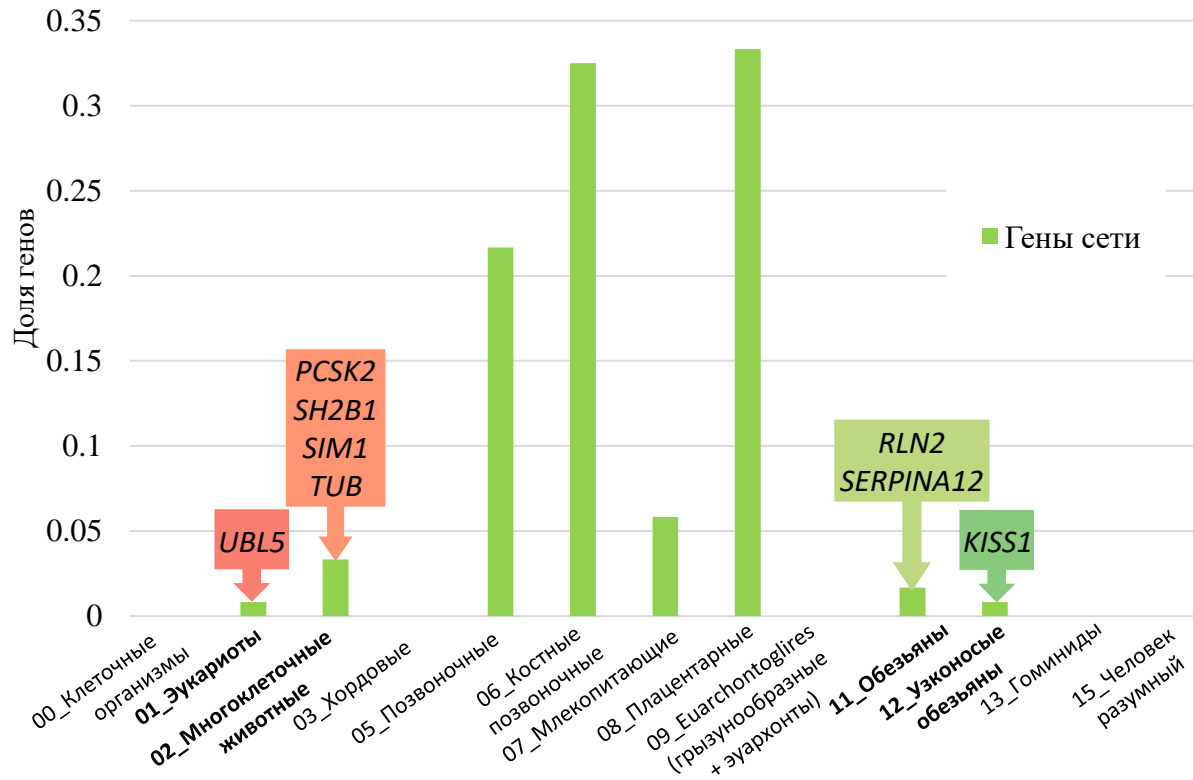
Наблюдаемые количества анорексигенных (**23**) и орексигенных (**5**) белков, для которых была выявлена регуляция с помощью микроРНК, достоверно выше или ниже, соответственно, их количества, ожидаемого при биномиальном распределении

$\text{Bin}(n; p=76/118*n, q=1-p=42/118*n)$:
уровень статистической значимости $\alpha < 0.05$.

	Общее количество белков, регулируемых микроРНК (n)	Анорексигенные белки (N_{\downarrow}), регулируемые микроРНК	Орексигенные белки (N_{\uparrow}), регулируемые микроРНК	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv 76*n/118) = \text{Bin}(q(N_{\uparrow}) \equiv 42*n/118)$
Ожидаемое	28	$28*76/(76+42) = 18.03$	$28*42/(76+42) = 9.97$	-
Наблюдаемое	28	23	5	< 0.05

Эволюционный возраст генов генной сети регуляции аппетита

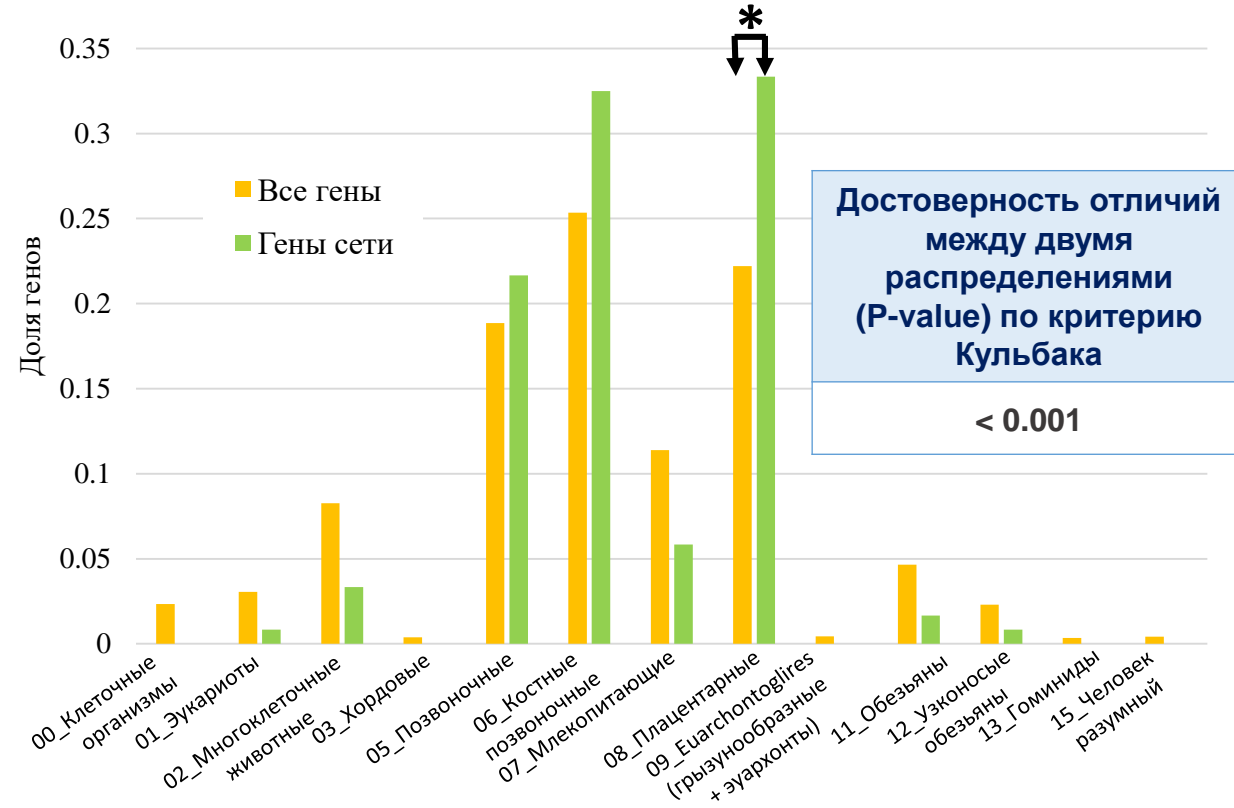
Распределение генов сети (120) по индексу PAI



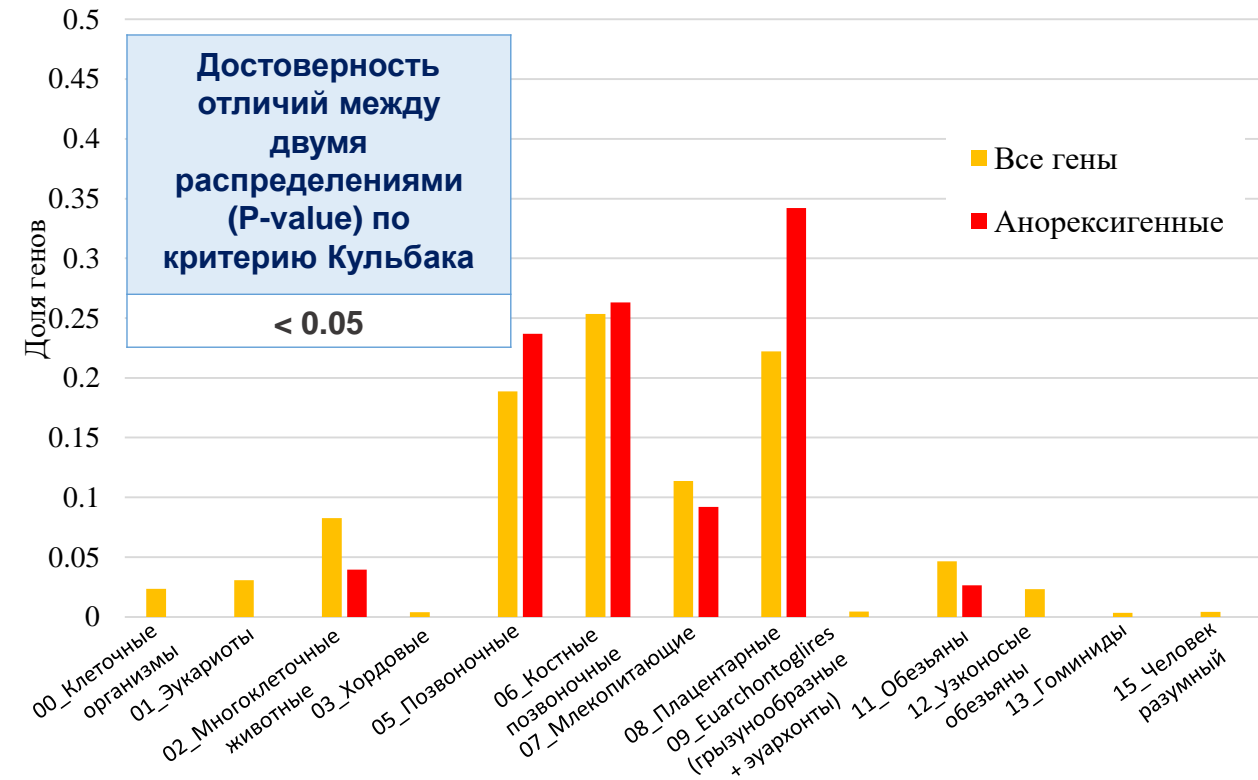
	Название гена	Значение индекса PAI	Молекулярная функция кодируемого белка
Низкие значения PAI (древние гены)	<i>UBL5</i>	1	Убиквитин-подобный белок
	<i>PCSK2</i>	2	Протеаза
	<i>SH2B1</i>	2	Цитоплазматический адаптерный белок
	<i>SIM1</i>	2	Транскрипционный фактор
	<i>TUB</i>	2	Транскрипционный фактор
Высокие значения PAI (молодые гены)	<i>KISS1</i>	12	Секретируемый сигнальный пептид
	<i>SERPINA12</i>	11	Гормон (адипокин)
	<i>RLN2</i>	11	Гормон

базовые клеточные функции

Распределение генов сети (120) и белок-кодирующих генов генома человека (19556) по индексу PAI



Распределение анорексигенных генов сети (76) и белок-кодирующих генов генома человека (19556) по индексу PAI



* - $p < 0.05$ (по критерию χ^2 с поправкой Бонферрони)

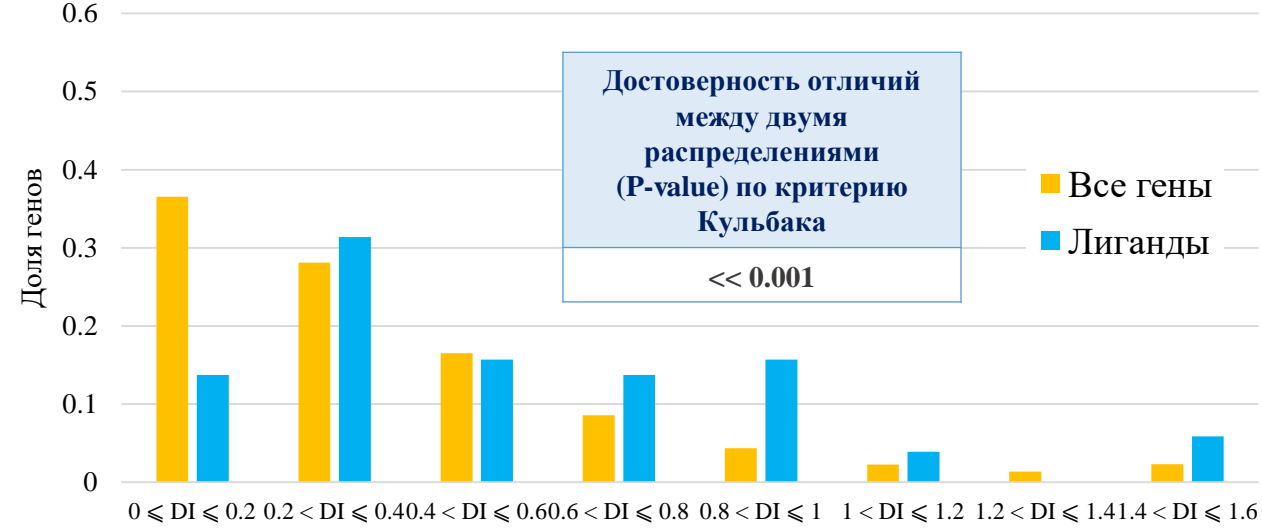
Эволюционный возраст генов, кодирующих лиганды и рецепторы генной сети

Пороговое значение уровня идентичности последовательностей генов-ортологов, при котором рассчитывали величину индекса PAI	Вид данных	Общее количество пар лиганд-рецептор с разным возрастом (n)	Лиганды, возраст генов которых меньше возраста генов рецепторов, (N_{\downarrow})	Лиганды, возраст генов которых больше возраста генов рецепторов, (N_{\uparrow})	$\alpha = \text{Bin}(p(N_{\downarrow}) \equiv q(N_{\uparrow}) \equiv 1/2)$
0.5	Ожидаемое	35	$35/2 = 17.5$	$35/2 = 17.5$	-
	Наблюдаемое	35	27	8	< 0.001
0.6	Ожидаемое	33	$33/2 = 16.5$	$33/2 = 16.5$	-
	Наблюдаемое	33	29	4	<< 0.001
0.7	Ожидаемое	33	$33/2 = 16.5$	$33/2 = 16.5$	-
	Наблюдаемое	33	28	5	< 0.001
0.8	Ожидаемое	31	$31/2 = 15.5$	$31/2 = 15.5$	-
	Наблюдаемое	31	26	5	< 0.001
0.9	Ожидаемое	42	$42/2 = 21$	$42/2 = 21$	-
	Наблюдаемое	42	31	11	< 0.01
1.0	Ожидаемое	26	$26/2 = 13$	$26/2 = 13$	-
	Наблюдаемое	26	13	13	-

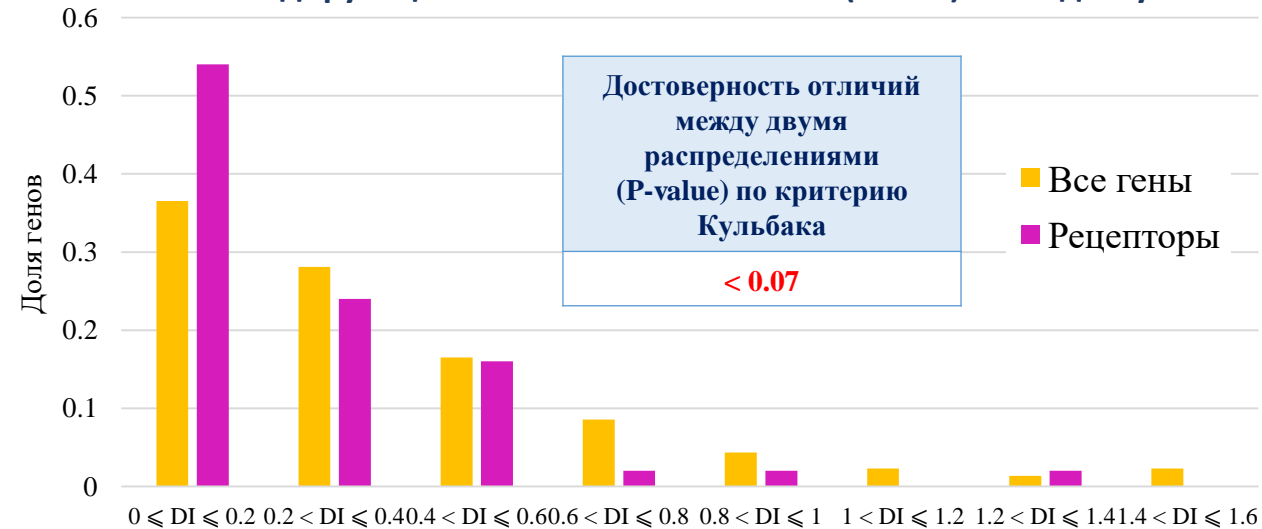
α - уровень статистической значимости того, что наблюдаемые количества случаев, когда ген, кодирующий рецептор, старше либо младше гена, кодирующего лиганд, достоверно выше и ниже, соответственно, их количества, ожидаемого при равномерном биномиальном распределении $\text{Bin}(n; p=q=1-p=1/2)$

Эволюционная изменчивость генов генной сети регуляции аппетита

Распределение генов сети, кодирующих лиганды, (51) и белок-кодирующих генов генома человека (19556) по индексу DI



Распределение генов сети, кодирующих рецепторы, (50) и белок-кодирующих генов генома человека (19556) по индексу DI



Распределение генов сети, кодирующих лиганды (51) и рецепторы (50), по индексу DI



Критерий Кульбака

- Критерий Кульбака – информационная мера. В отличие от критерия χ^2 она зависит не от значений, которые есть в выборке, а от частот этих значений. Поэтому критерий менее уязвим при работе с маленькими значениями в выборке, чем критерий χ^2 .
- Формула для подсчёта критерия Кульбака:

$$I(p^* : \hat{p}) = I(H_1 : H_2) = \sum x_i \log x_i + \sum y_i \log y_i - \sum (x_i + y_i) \log (x_i + y_i) + (N_1 + N_2) \log (N_1 + N_2) - N_1 \log N_1 - N_2 \log N_2$$

- Величина $2 * I(p^* : \hat{p})$ имеет асимптотически χ^2 -распределение с $(r-1) * (c-1)$ степенями свободы

Сравнение распределений двух групп генов по PAI

i	x_i	y_i
Значение индекса PAI	Количество генов из 1-й группы	Количество генов из 2-й группы
0	x_0	y_0
1	x_1	y_1
2	x_2	y_2
	$N_1 = x_0 + x_1 + x_2$	$N_2 = y_0 + y_1 + y_2$

$r = 2$

$c =$ число индексов PAI

Попарное сравнений двух групп генов при определенном значении индекса PAI

	Количество генов из группы, имеющих определенное значение индекса PAI i	Количество генов из группы, исключая гены, имеющие определенное значение индекса PAI i	Количество генов в группе
1 группа	x_{1i}	$x_{2i} = N_1 - x_{1i}$	N_1
2 группа	y_{1i}	$y_{2i} = N_2 - y_{1i}$	N_2

$c = 2$

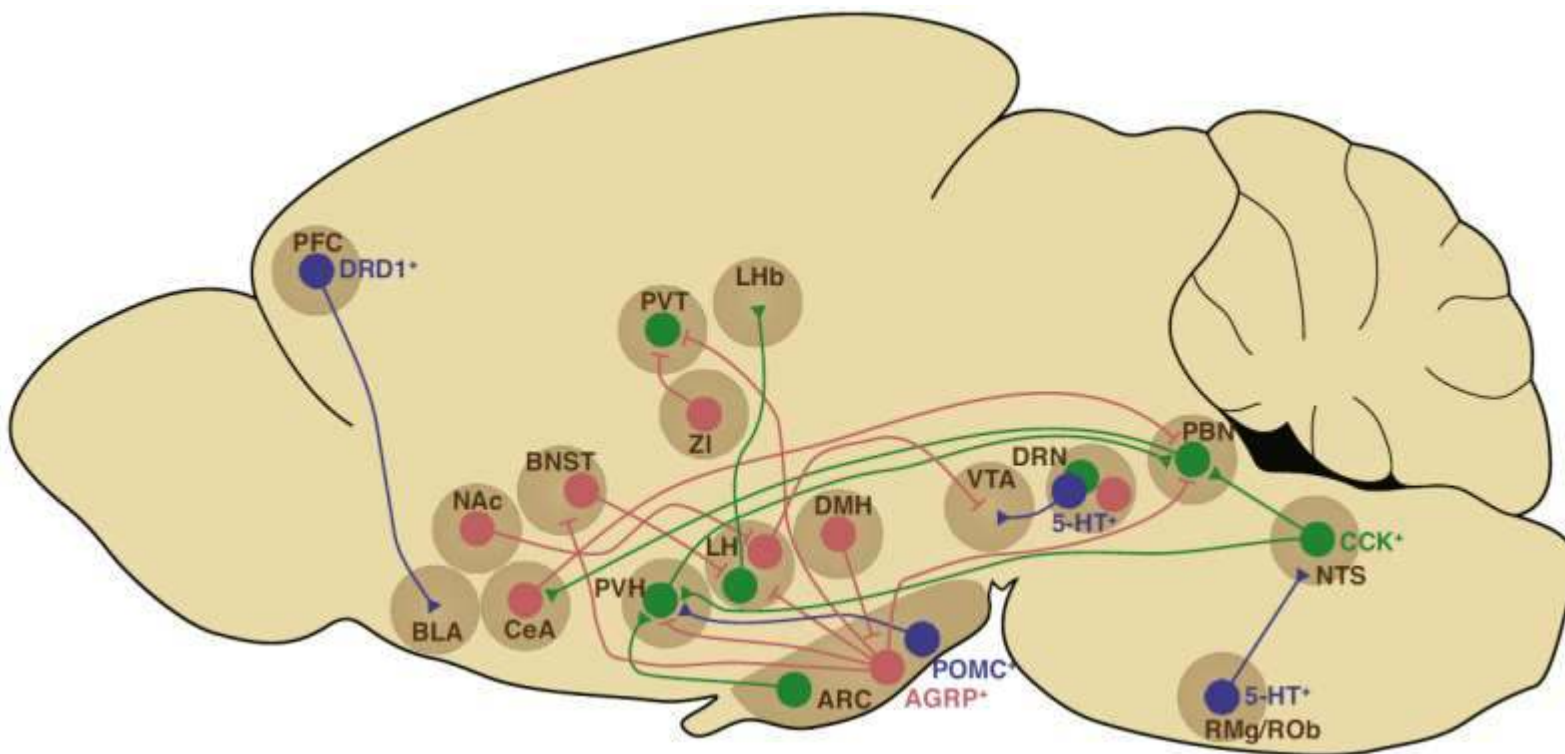
$r = 2$

Критерий χ^2

Попарное сравнений двух групп генов при определенном значении индекса PAI

	Наблюдаемое количество генов	Ожидаемое количество генов
Количество генов из первой группы генов, имеющих определенный i	N_{i1}	$M_{i1} = N_1 * N_{i2} / N_2$ где $N_2 = \sum_{i=1}^c N_{i2}$, N_2 – общее количество генов во второй группе генов, а N_{i2} – количество генов из второй группы генов, имеющих определенный i
Количество генов из первой группы генов, исключая гены, имеющие определенный i	$N_1 - N_{i1}$	$N_1 - M_{i1}$
Общее количество генов в первой группе	$N_1 = \sum_{i=1}^c N_{i1}$	N_1

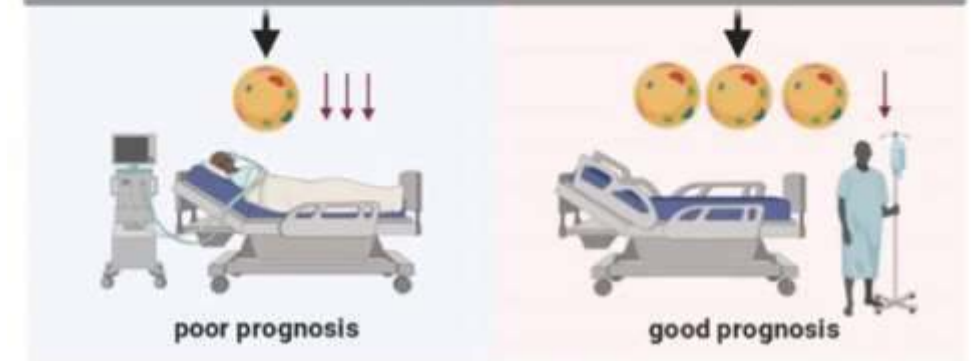
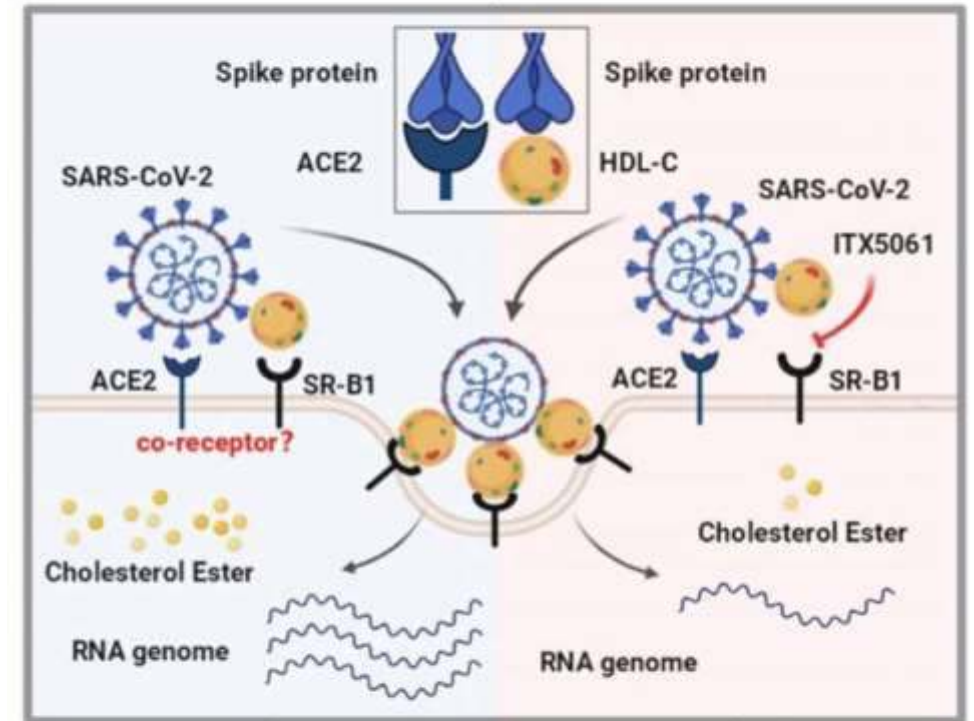
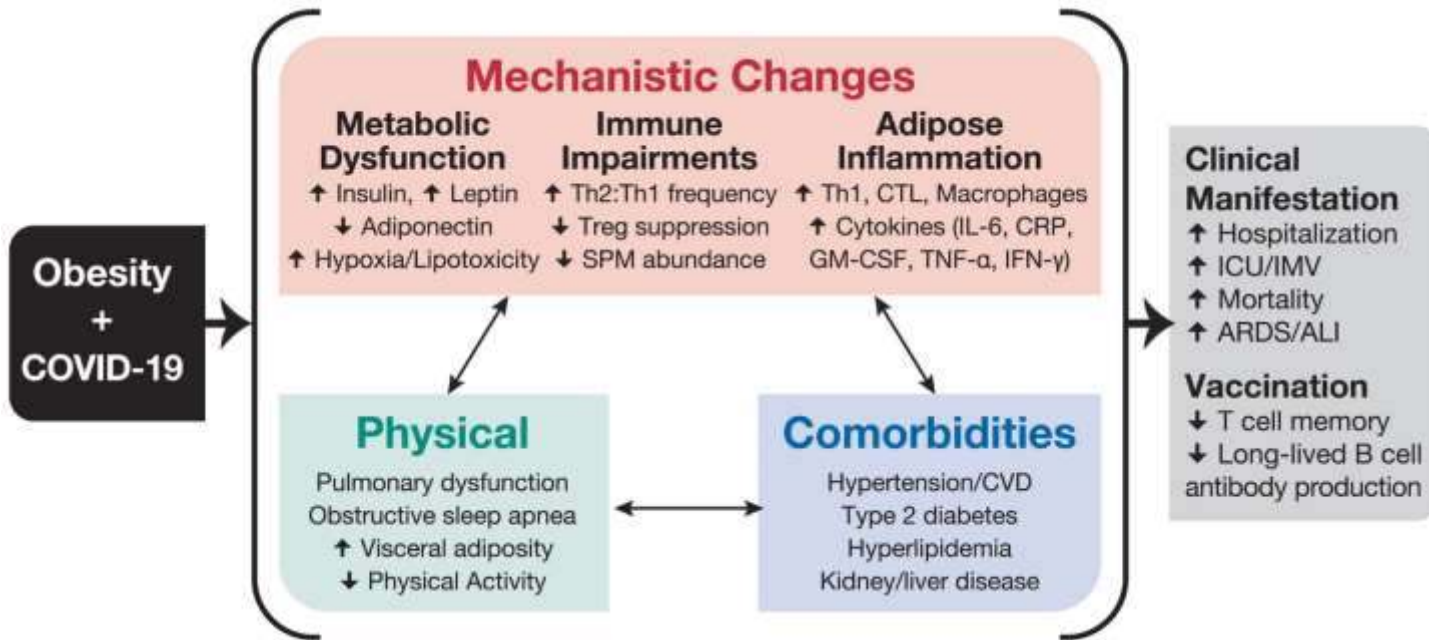
Ключевые ядра нейронов, регулирующие аппетит



- ARC – аркуатное ядро гипоталамуса
- BLA – базолатеральная миндалина амигдалы
- BNST – ядро ложа терминальной полоски
- CeA – центральная миндалина
- DMH – дорсомедиальное ядро гипоталамуса
- DRN – ядро дорсального шва продолговатого мозга
- LH – боковой гипоталамус
- LHb – боковая габенула
- NAc – прилежащее ядро
- NTS – ядро солитарного тракта
- PBN – парабрахияльное ядро ствола мозга
- PFC – префронтальная кора
- PVH – паравентрикулярное ядро гипоталамуса
- PVT - паравентрикулярное ядро таламуса
- VTA – вентральная область покрышки среднего мозга

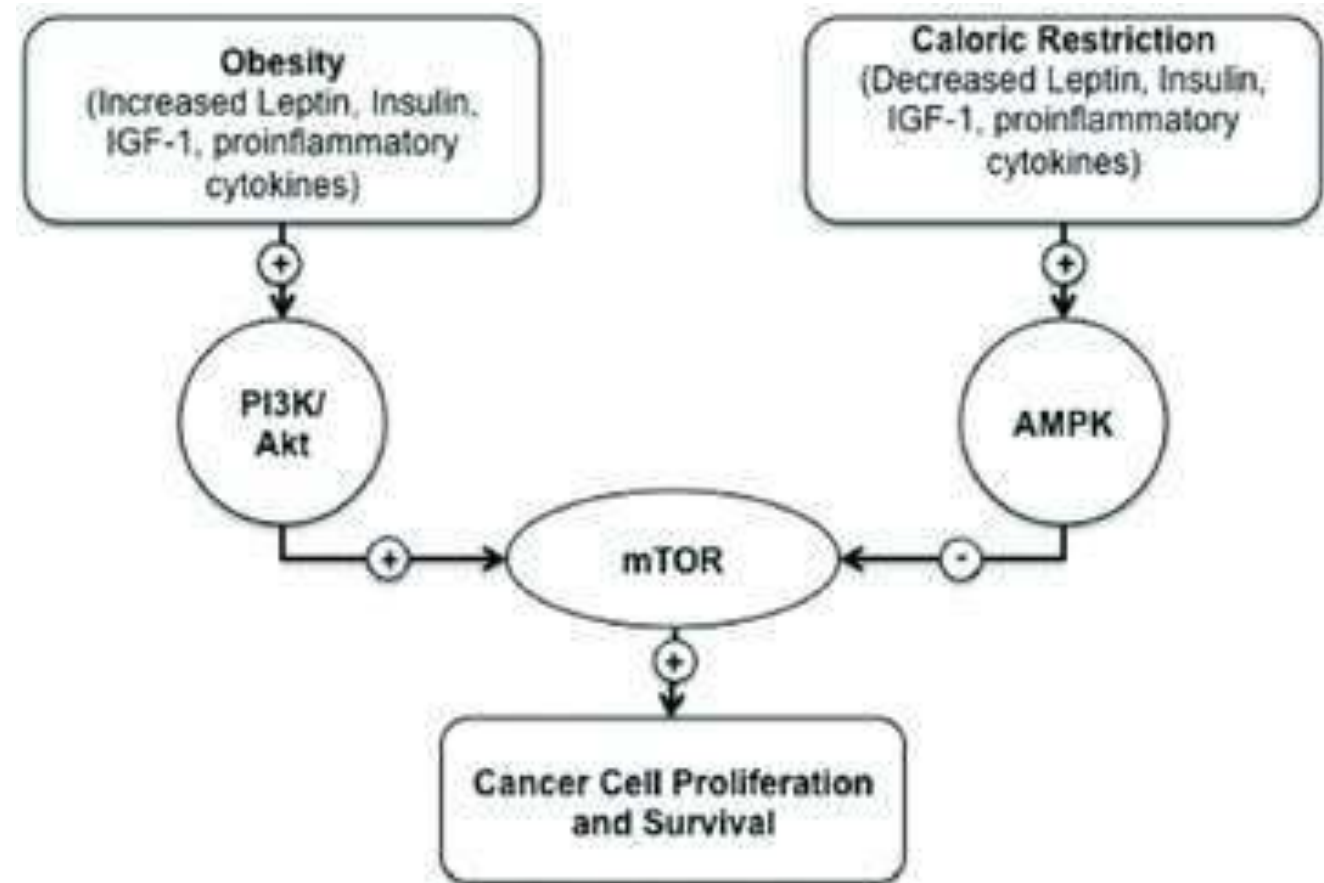
Связь ожирения и COVID-19

Предполагаемый механизм влияния ожирения на риск развития тяжелой формы COVID-19

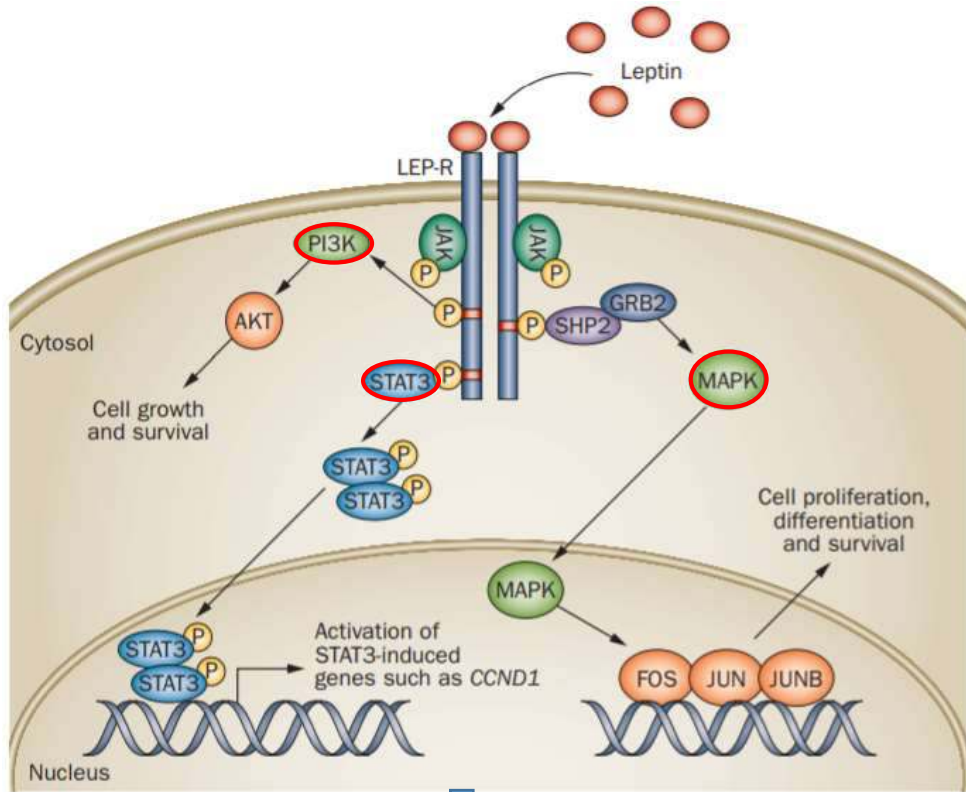


Ожирение является фактором риска развития нескольких типов рака, включая рак толстой кишки, молочной железы, эндометрия, печени, почек, пищевода, желудка, поджелудочной железы, желчного пузыря и лейкемии, а также может приводить к снижению эффективности лечения и увеличению смертности.

Связь ожирения и развития рака

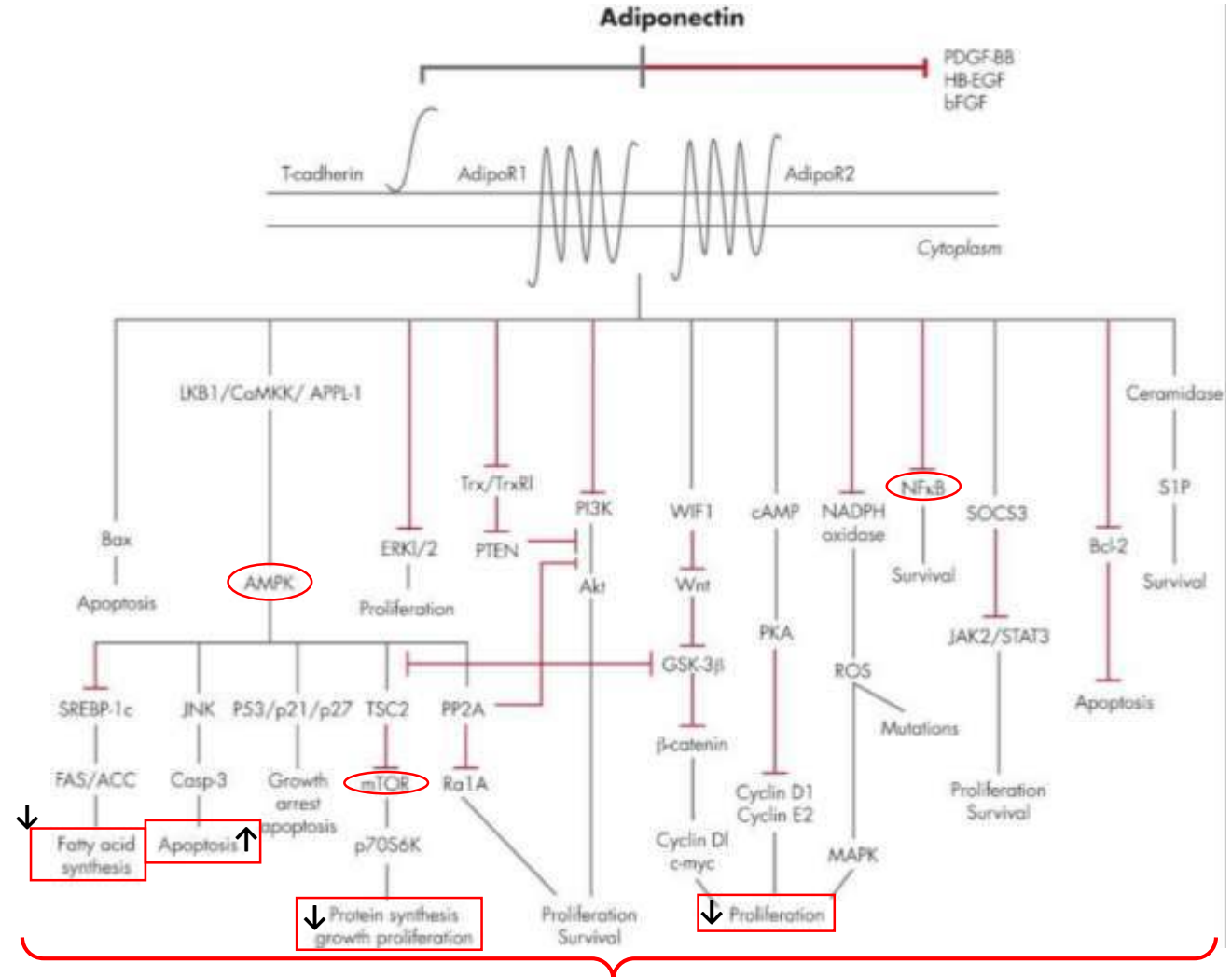


Сигнальные пути, связывающие лептин с канцерогенезом



↓
прогрессирование рака

Сигнальные пути, связывающие адипонектин с канцерогенезом



↑
противораковое действие