

Курсовая работа по биологии на тему:

Сравнительный анализ транскриптомных паттернов сарком человека

Автор: Корлякова Мария Денисовна, 10 “Н”

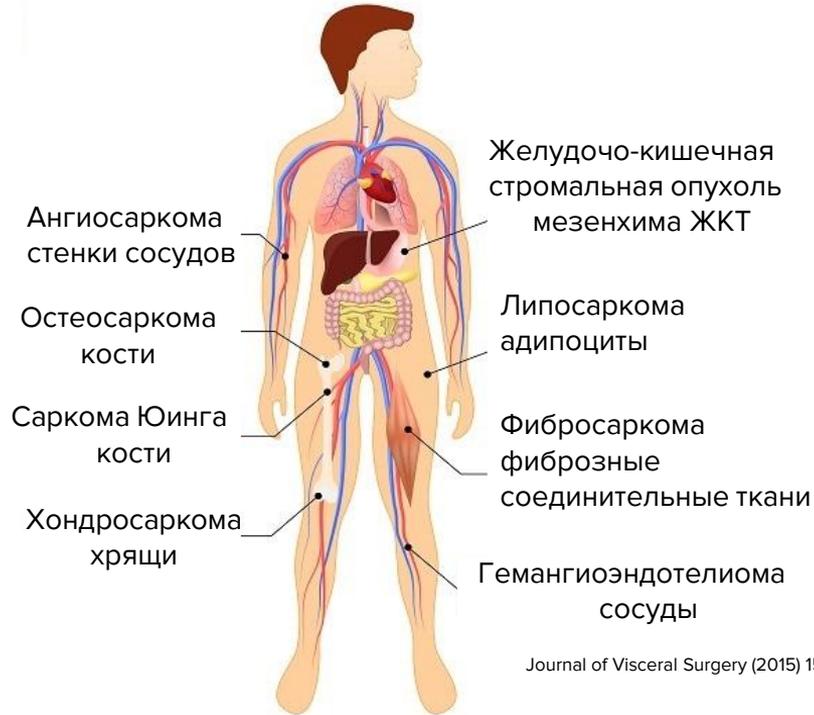
Научный руководитель: Азбукина Надежда Витальевна,
Факультет Биоинженерии и Биоинформатики
МГУ имени Н. В. Ломоносова

Соруководитель: Горбатенко Владислав Олегович,
Факультет Биоинженерии и Биоинформатики
МГУ имени Н. В. Ломоносова

СУНЦ МГУ
май 2019г

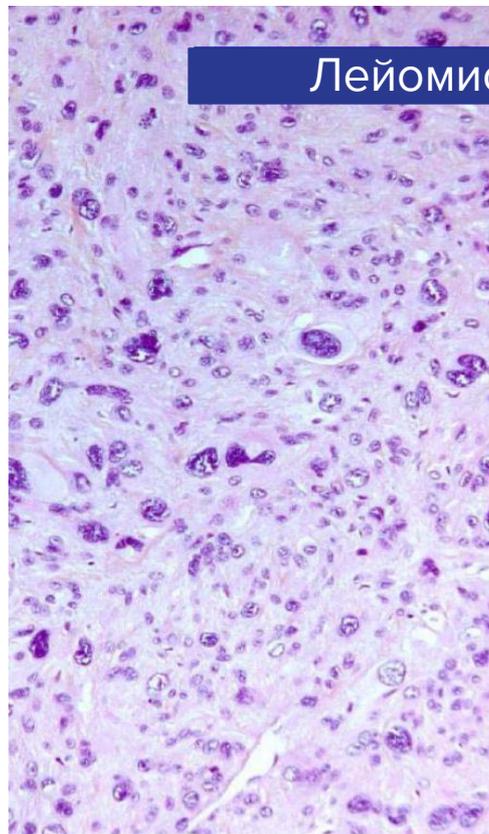
Саркома

тип злокачественного новообразования, локализующийся, в основном, в соединительных и мышечных тканях

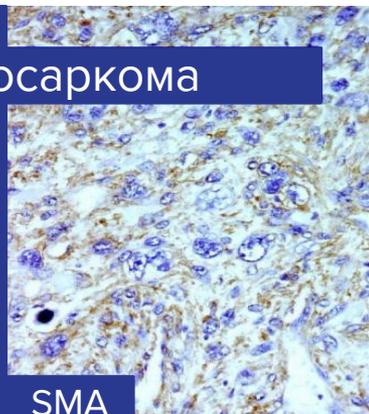


- риск первоначальной диагностической ошибки 10-25%;
- в 4% случаев ошибочное диагностирование саркомы в качестве доброкачественной опухоли;
- в 10% случаев ошибочное диагностирование доброкачественных опухолей в качестве саркомы.

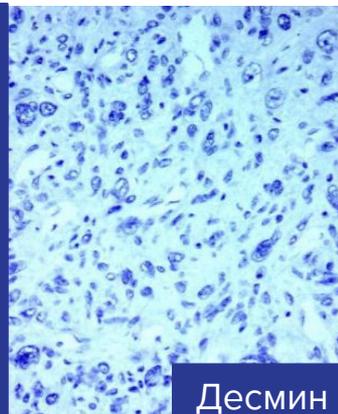
Иммуногистохимия



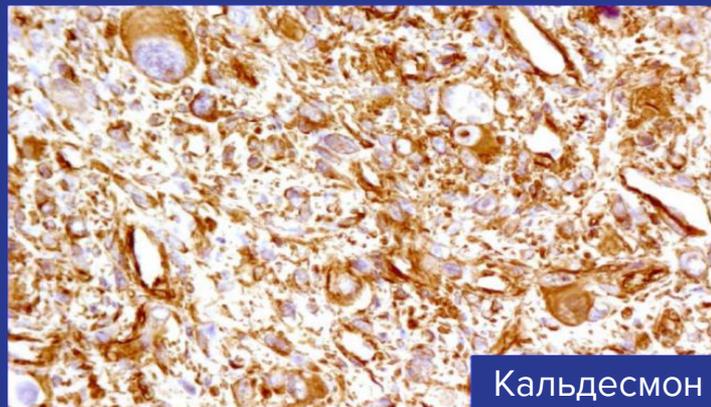
Лейомиосаркома



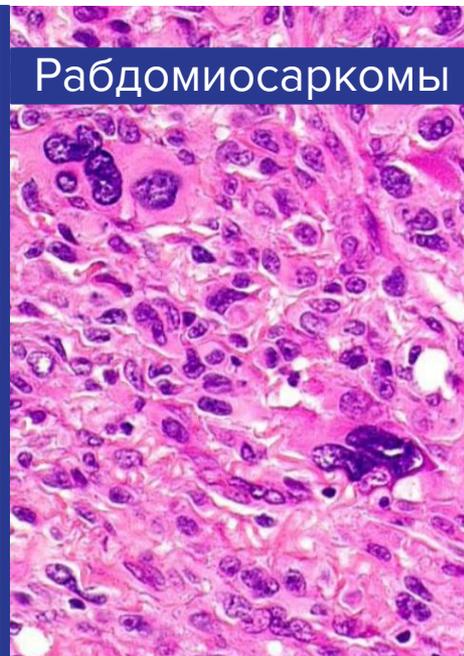
SMA



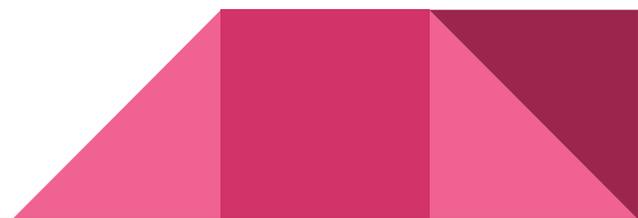
Десмин



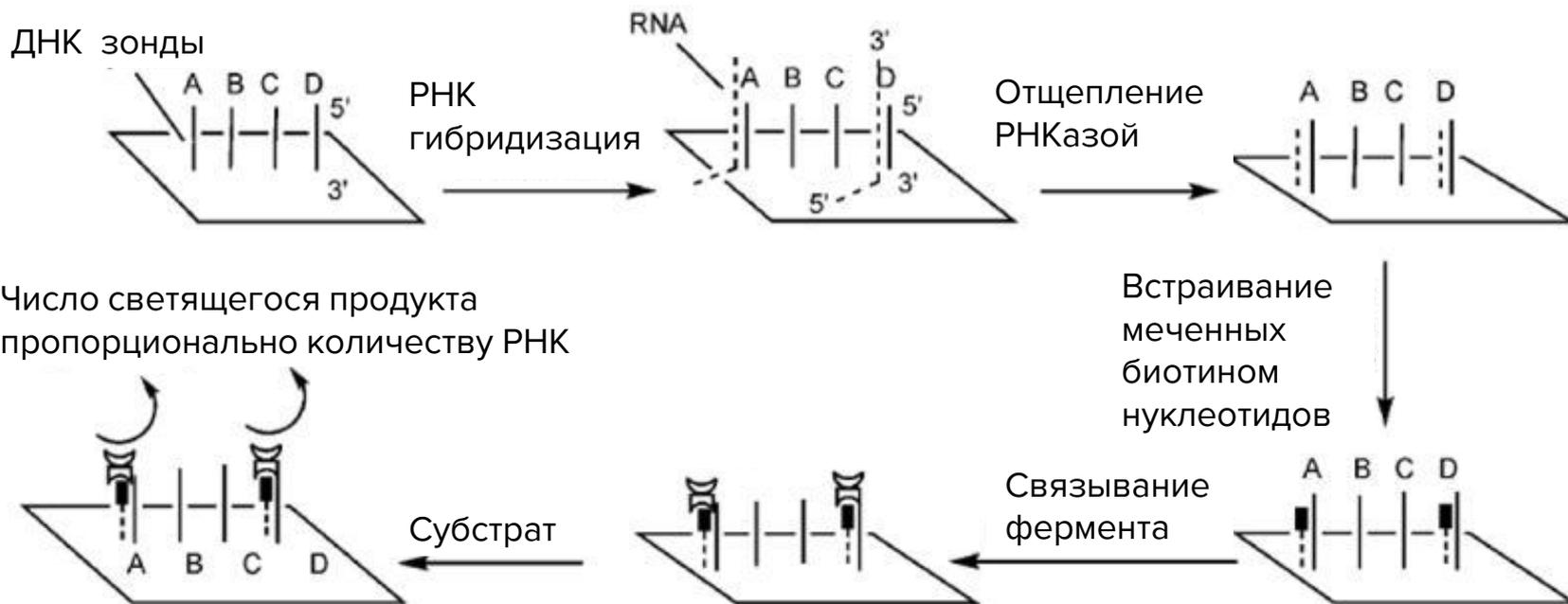
Кальдесмон



Рабдомиосаркомы



Микрочип



Цель исследования

оценить сходство и различие транскриптомных маркёров сарком в зависимости от их гистологического происхождения на примере миосарком.



Задачи исследования

- отобрать датасеты транскриптомов рабдомиосарком, лейомиосарком и липосарком;
 - обработать данные по экспрессии генов (включая нормировку и определение дифференциально экспрессирующихся генов);
 - сравнить полученные паттерны экспрессии между различными саркомами, определить их схожесть и различие в зависимости от ткани, в которых они развивались;
 - проаннотировать выявленные маркеры, описать, какую функцию они выполняют, определить их взаимосвязь друг между другом.
- 

Материалы и методы

Схема эксперимента:

- поиск датасетов содержащих как контрольную, так и исследуемую группу результатов секвенирования с использованием микрочипов с помощью сетевой библиотеки PubMed;
 - проведение теста Стьюдента для каждого датасета при помощи программы MeV (Multiple Array Viewer);
 - определение величины full change для каждого гена, отбор генов с приемлемой величиной;
 - выделить дифференциально экспрессируемые гены;
 - определить различия и совпадения в ДЭГах для разных типов сарком;
 - определить функции ДЭГов.
- 

Пример: до нормирования

	Опухолевые клетки 1	Опухолевые клетки 2	Контроль 1	Контроль 2
ген 1	578	338	941	965
ген 2	202	321	789	608
ген 3	278	455	1189	948
ген 4	663	386	112	79
ген 5	478	534	991	835

Пример: после нормирования

	Опухолевые клетки 1	Опухолевые клетки 2	Контроль 1	Контроль 2
ген 1	0,61	0,35	0,97	1
ген 2	0,26	0,41	1	0,77
ген 3	0,24	0,38	1	0,80
ген 4	1	0,58	0,17	0,10
ген 5	0,48	0,54	1	0,84

Multiple Array Viewer

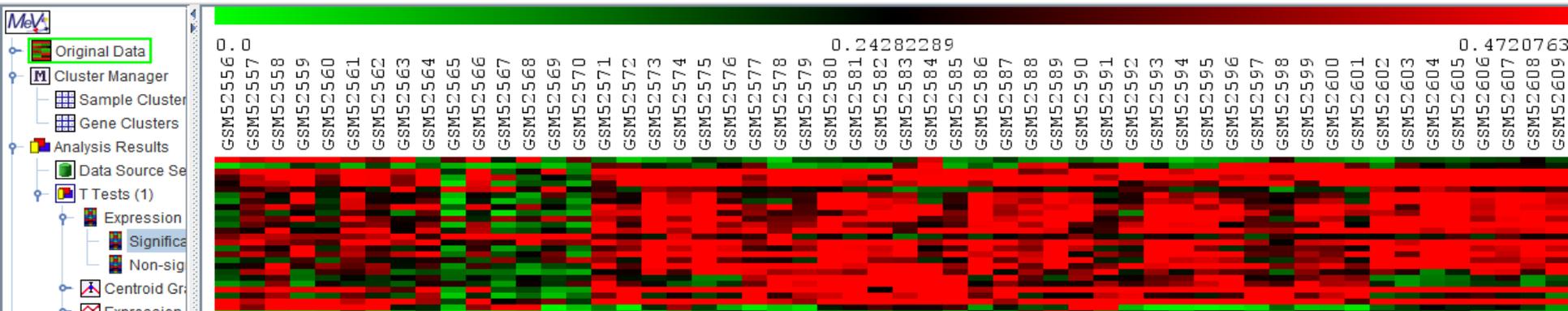
Схема эксперимента:

- Были отобраны датасеты для рабдомиосаркомы, лейомиосаркомы, липосаркомы по следующим критериям: подходящая платформа (affimetryх gpl980), наличие контроля, размер выборки более 10-15 человек;
- нормирование значений датасетов на максимальное значение экспрессии в строке;
- получение значения об экспрессии в виде тепловых карт и таблиц;
- проведение теста Стьюдента.

$$t = \frac{(m_c - m_t)}{\sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t}}}$$

- m — средние значения по выборкам
- s — дисперсия
- n — количество образцов выборки

Multiple Array Viewer



Multiple Array Viewer

File Adjust Data Metrics Analysis Display Utilities

Clustering Statistics Classification Data Reduction Meta Analysis Visualization

Stored Color	Expression line	ID_REF	PROBE_ID	CHR_LOCATION	GroupA mean	GroupA std.dev.	GroupB mean	GroupB std.dev.	Absolute t value	Degree
		DDR1~discoidin do...	DDR1~discoidin do...	chrNA:-1-1	0.36869973	0.25315645	0.14853175	0.07613524	3.7898664	21.0
		MMP14~matrix meta...	MMP14~matrix meta...	chrNA:-1-1	0.13322842	0.08923209	0.31205916	0.23198754	4.0179005	46.0
		EIF4G2~eukaryotic t...	EIF4G2~eukaryotic t...	chrNA:-1-1	0.50293773	0.20490387	0.6550549	0.16173595	2.8401396	32.0
		GDI2~GDP dissocia...	GDI2~GDP dissocia...	chrNA:-1-1	0.4499916	0.1905482	0.6056114	0.17210256	3.002371	36.0
		SEPT2~septin 2~2q...	SEPT2~septin 2~2q...	chrNA:-1-1	0.39954677	0.22261085	0.5991019	0.20158522	3.2927139	36.0
		SLC25A3~solute ca...	SLC25A3~solute ca...	chrNA:-1-1	0.36626744	0.19407189	0.23544896	0.07515531	2.889775	22.0
		RPL6~ribosomal pr...	RPL6~ribosomal pr...	chrNA:-1-1	0.3012068	0.15630388	0.43901652	0.20160007	2.803168	47.0
		RPL17~ribosomal p...	RPL17~ribosomal p...	chrNA:-1-1	0.27563968	0.16950503	0.45194814	0.2291406	3.229247	49.0
		YY1~YY1 transcripti...	YY1~YY1 transcripti...	chrNA:-1-1	0.37607375	0.13620545	0.5015466	0.16537906	3.0149078	46.0
		ZNF146~zinc finger ...	ZNF146~zinc finger ...	chrNA:-1-1	0.28083694	0.21549274	0.49788514	0.20275998	3.652633	38.0
		RPL30~ribosomal p...	RPL30~ribosomal p...	chrNA:-1-1	0.3768114	0.12776902	0.5049399	0.20085675	2.863025	51.0
		NPM1~nucleophos...	NPM1~nucleophos...	chrNA:-1-1	0.30058777	0.18379854	0.45841828	0.17777859	3.0842638	38.0
		IK~IK cytokine, down...	IK~IK cytokine, down...	chrNA:-1-1	0.36437654	0.15813936	0.5449721	0.17398378	3.9032688	43.0
		C2orf24~chromoso...	C2orf24~chromoso...	chrNA:-1-1	0.48241395	0.2251682	0.25016472	0.072001256	4.480011	21.0
		HNRNPd~heteroge...	HNRNPd~heteroge...	chrNA:-1-1	0.45465007	0.14423212	0.5744735	0.17398676	2.7271502	46.0
		RPS7~ribosomal pr...	RPS7~ribosomal pr...	chrNA:-1-1	0.29553792	0.16733332	0.43618438	0.19450158	2.8058145	44.0

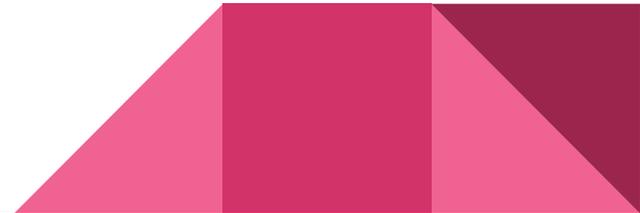
Обработка данных при помощи таблиц Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1						1 и 3 таблицы совпадения						впр
2	SDC2	3,900071881	SDC2	0,248492941	0,063714964	#N/A	P4htm	0,521900354	P4htm	0,31207373	0,59795654	#N/A
3	SDC2	3,588949368	SDC2	0,248492941	0,069238352	#N/A	Dgka	0,523041168	Dgka	0,35429196	0,677369166	#N/A
4	SDC2	2,601351498	SDC2	0,248492941	0,095524554	#N/A	Dnajc4	0,532480414	Dnajc4	0,470391803	0,883397381	#N/A
5	SYT11	3,404610376	SYT11	0,397660513	0,116800594	#N/A	Smarcc1	2,838555799	Smarcc1	2,594980278	0,914190335	#N/A
6	SNAPC1	2,875865608	SNAPC1	0,509530688	0,177174721	#N/A	Nup93	3,08750789	Nup93	3,341924827	1,082402036	#N/A
7	KPNB1	2,976176536	KPNB1	1,796196114	0,603524721	0,585298196	Ppp1ca	1,550118098	Ppp1ca	1,776364576	1,145954349	#N/A
8	KPNB1	2,463032669	KPNB1	1,796196114	0,729261994	0,585298196	Whsc1	2,542478775	Whsc1	3,394833624	1,335245611	#N/A
9	KPNB1	2,370156539	KPNB1	1,796196114	0,7578386	0,585298196	Chd1	2,093924131	Chd1	3,008212028	1,436638503	#N/A
10	CHEK1	4,187182094	CHEK1	3,338854333	0,797398885	#N/A	Olfm1	0,210098462	Olfm1	0,407778684	1,940893236	#N/A
11	PLXNB2	0,477892319	PLXNB2	0,392838208	0,822022436	#N/A	Chd1	1,511948716	Chd1	3,008212028	1,989625703	#N/A
12	RRM1	3,479513619	RRM1	3,495719064	1,004657388	#N/A	Prpf4	0,545571099	Prpf4	2,055030561	3,76675114	#N/A
13	KPNB1	1,548931697	KPNB1	1,796196114	1,159635456	0,585298196	Immt	0,621090258	Immt	2,521648618	4,060035693	#N/A
14	KIF20A	5,052606686	KIF20A	6,257323091	1,23843463	#N/A	E2f6	0,396954449	E2f6	3,059046903	7,706292013	#N/A
15	PTTG1	5,242732597	PTTG1	6,579017659	1,254883315	#N/A	E2f6	0,362138248	E2f6	3,059046903	8,447179825	#N/A
16	TNFRSF10I	0,409712963	TNFRSF10I	0,531009797	1,2960532	#N/A	E2f6	0,357952832	E2f6	3,059046903	8,54594972	#N/A
17	RRM1	2,606000699	RRM1	3,495719064	1,341411407	#N/A						
18	RNASEH2A	3,519962842	RNASEH2A	5,176111954	1,470501874	#N/A						
19	KIF4A	3,830120399	KIF4A	8,379415729	2,187768231	#N/A						
20												
21		липосаркома				лейомиосаркома						
22	KPNB1	0,585298196	KPNB1	1,993512938	KPNB1	1,548931697	KPNB1	1,796196114	1,159635			
23												
24												
25												
26	CHEK1	checkpoint kinase 1										
27	KIF20A	kinesin family member 20A										
28	KPNB1	karyopherin subunit beta 1										
29	PLXNB2	plexin B2										
30	KIF4A	kinesin family member 4A										
31	PTTG1	pituitary tumor-transforming 1										
32	RNASEH2A	ribonuclease H2 subunit A										
33	RRM1	ribonucleotide reductase catalytic subunit M1										

- для поиска одинаковых генов в разных датасетах использовалась функция ВПР (VLOOKUP)

Результаты

- Из всех проанализированных данных было определено: по 13 ДЭГов для лейомиосаркомы и рабдомиосаркомы, 60 ДЭГов для липосаркомы;
- у липосаркомы и лейомиосаркомы совпал один ДЭГ – KPNB1 (karyopherin subunit beta 1), его экспрессия в большинстве клеток повысилась;
- KPNB1 участвует в импорте белка в ядро (некоторых рибосомных белков, гистонов);
- остальные ДЭГи были распределены по функциональным группам.



ДЭГи лейомиосаркомы

- участвующие в клеточном цикле (CHEK1; RRM1; TNFRSF10B; CCNB1)
- цитоскелет (KIF20A; KIF4A)
- сигналинг (PLXNB2; SDC2)
- участвующие в транскрипции (PTTG1; RNASEH2A; SNAPC1)

ДЭГи рабдомиосаркомы

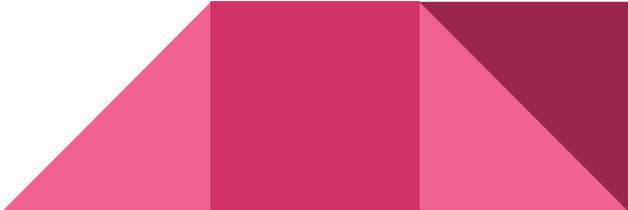
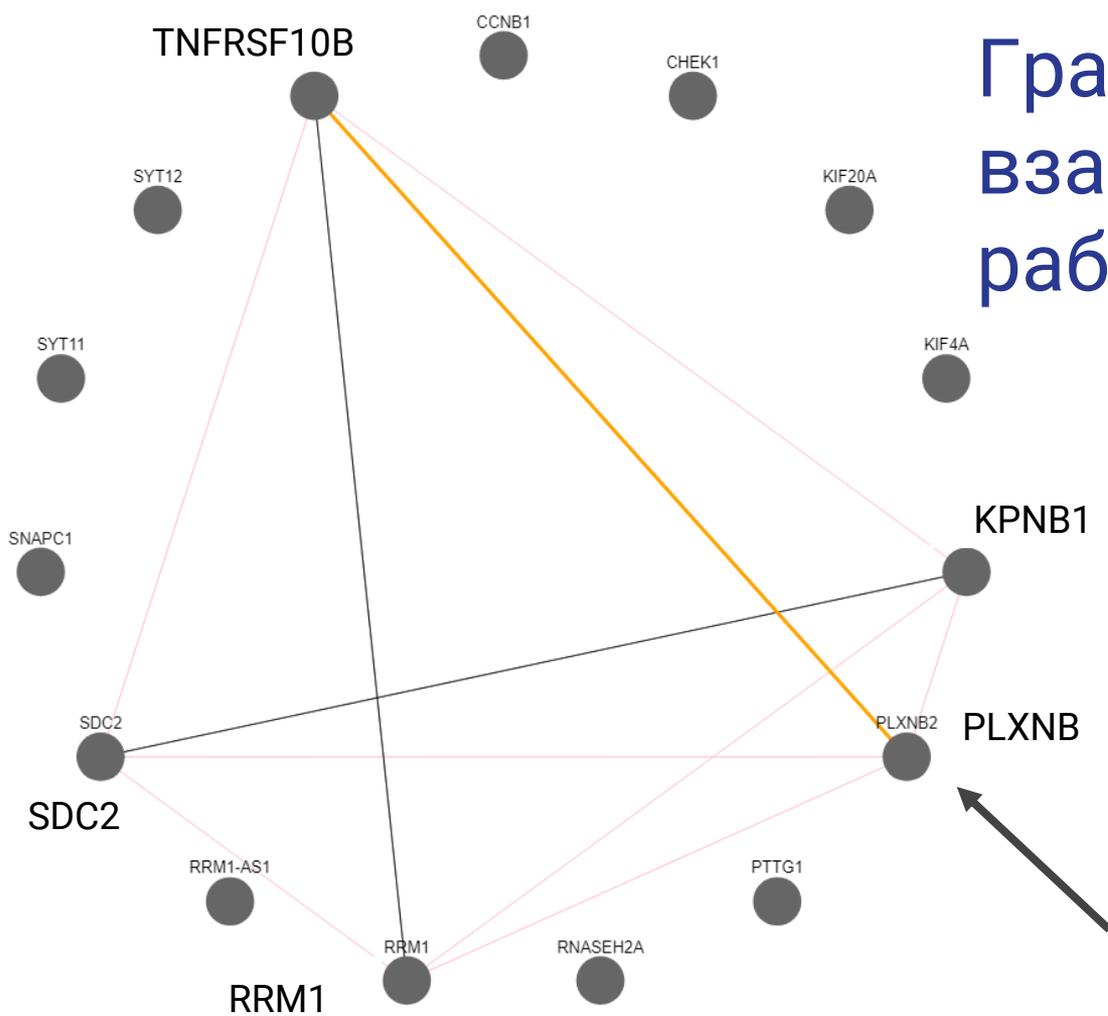
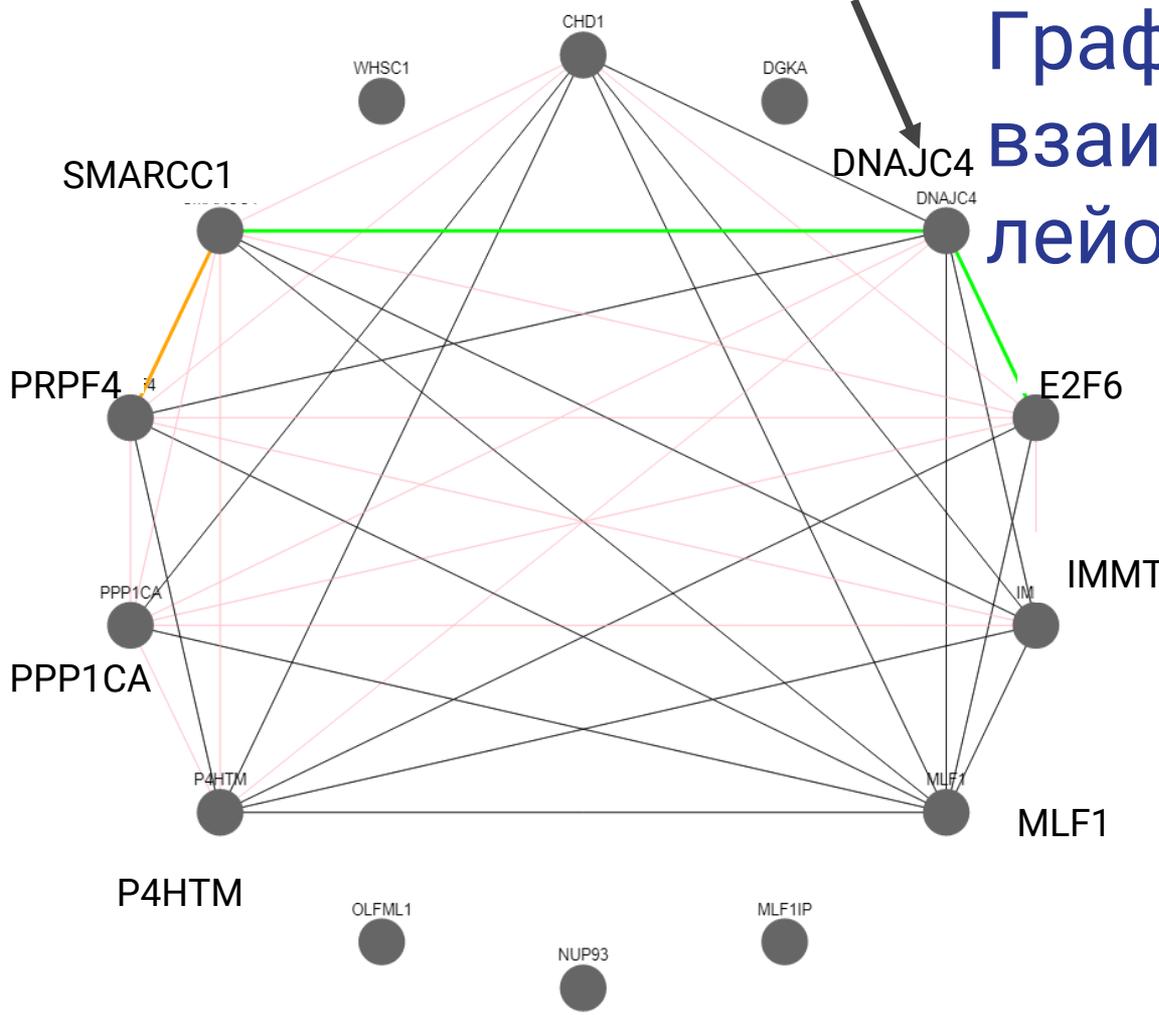
- участвующие в регуляции других клеточных процессов (Nup93; Ppp1ca; Dgka; Prpf4; Dnajc4)
 - участвующие в транскрипции (Smarcc1; Whsc1; Chd1; E2f6; Mlf1)
 - участвующие в метаболизме (Immt; P4htm)
- 

График функциональной взаимосвязи ДЭГов рабдомиосаркомы



- стрелка указывает на ген, понижающий свою экспрессию, остальные гены повышают экспрессию

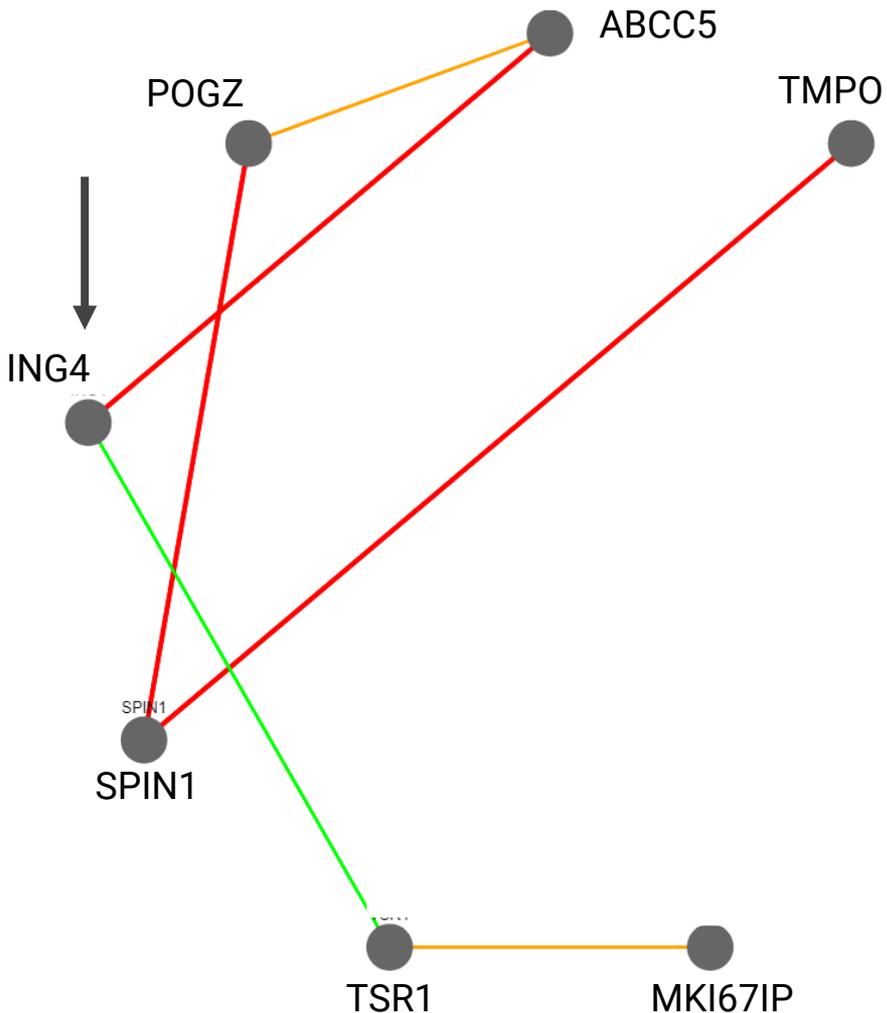
График функциональной взаимосвязи ДЭГов лейомиосаркомы



● стрелка указывает на ген, понижающий свою экспрессию, остальные гены повышают экспрессию

График функциональной взаимосвязи ДЭГов липосаркомы

- стрелка указывает на ген, понижающий свою экспрессию, остальные гены повышают экспрессию



Выводы

- найденные гены можно использовать в качестве маркеров для дифференциальной диагностики лейомиосаркомы и рабдомиосаркомы;
- в лейомиосаркоме основные изменения в экспрессии претерпевают гены, связанные с процессом транскрипции, клеточным циклом и работой цитоскелета;
- в рабдомиосаркоме основные изменения в экспрессии претерпевают гены, связанные с регуляцией метаболизма, транспорта, фолдинга белков, сплайсинга РНК, а также с процессом транскрипции;
- в липосаркоме основные изменения претерпевают гены связанные с метаболизмом.

