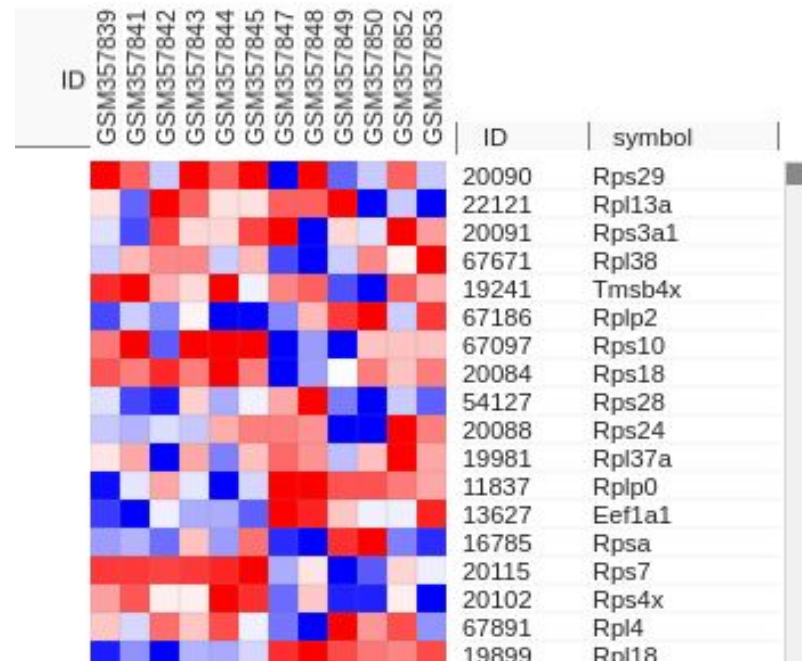


# Реализация эффективного взаимодействия между платформой для анализа экспрессии генов Morpheus и библиотекой вычислительных методов R/Bioconductor

Зенкова Д.М.,  
группа М3436,  
руководитель: Сергушичев А.А., к.т.н.

# Анализ экспрессии генов

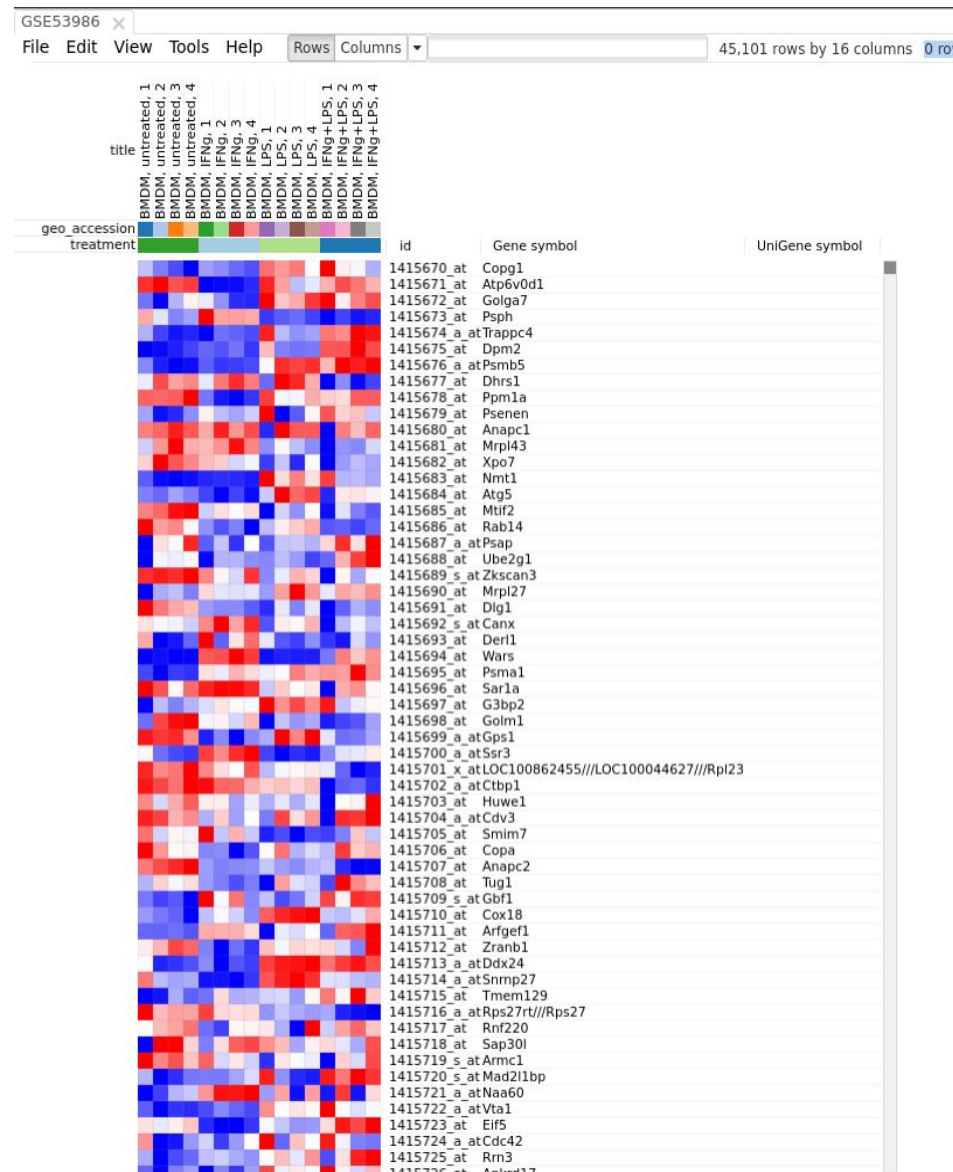
- Экспрессия генов — процесс преобразования наследственной информации от гена в РНК или белок
- Анализ экспрессии — часто используемый метод при исследованиях в биологии



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	symbol	GSM357839	GSM357841	GSM357842	GSM357843	GSM357844	GSM357845	GSM357847	GSM357848	GSM357849	GSM357850	GSM357852	GSM357853
2	20090	Rps29	16.32141057	16.30051518	16.25429304	16.32141057	16.30051518	16.3214106	16.2104266	16.3214106	16.23220169	16.25429304	16.30051518	16.25429304
3	22121	Rpl13a	16.27241391	16.23220169	16.32141057	16.30051518	16.27241391	16.2724139	16.30051518	16.3005152	16.32141057	16.2104266	16.25429304	16.2104266
4	20091	Rps3a1	16.23220169	16.18528469	16.30051518	16.25429304	16.25429304	16.3005152	16.32141057	16.1622511	16.25429304	16.23220169	16.32141057	16.27241391
5	67671	Rpl38	16.2104266	16.25429304	16.27241391	16.27241391	16.2104266	16.254293	16.16225112	16.1359833	16.2104266	16.27241391	16.23220169	16.32141057
6	19241	Tmsb4x	16.30051518	16.32141057	16.23220169	16.2104266	16.32141057	16.1852847	16.25429304	16.2724139	16.10255505	16.06215078	16.27241391	16.23220169
7	67186	Rplp2	16.16225112	16.2104266	16.18528469	16.23220169	16.13598328	16.1359833	16.18528469	16.254293	16.30051518	16.32141057	16.2104266	16.30051518
8	67097	Rps10	16.10255505	16.12314465	16.05120582	16.12314465	16.12314465	16.1231446	16.03531296	16.0621508	16.03531296	16.08954522	16.08954522	16.08954522
9	20084	Rps18	16.18528469	16.16225112	16.2104266	16.16225112	16.23220169	16.1622511	15.9457813	16.035313	16.08954522	16.16225112	16.12314465	16.16225112
10	54127	Rps28	16.06215078	15.98154012	15.95751874	16.10255505	16.03531296	16.0714312	16.12314465	16.2104266	16.0103654	15.9457813	16.05120582	15.99551072
11	20088	Rps24	15.98154012	15.96795367	15.99551072	15.98154012	16.07143124	16.102555	16.10255505	16.0895452	15.85007321	15.85007321	16.18528469	16.10255505
12	19981	Rpl37a	16.03531296	16.06215078	15.91137633	16.06215078	15.96795367	16.0512058	16.08954522	16.0714312	15.99551072	16.05120582	16.13598328	16.06215078
13	11837	Rplp0	15.9457813	16.07143124	16.13598328	16.07143124	15.93889341	16.0621508	16.23220169	16.2322017	16.18528469	16.18528469	16.16225112	16.13598328
14	13627	Eef1a1	15.96795367	15.9457813	16.03531296	16.0103654	16.0103654	15.9815401	16.13598328	16.1231446	16.06215078	16.03531296	16.03531296	16.12314465
15	16785	Rpsa	16.08954522	16.10255505	16.06215078	16.18528469	16.08954522	16.2322017	16.02104671	15.9955107	16.27241391	16.30051518	16.07143124	16.02104671
16	20115	Rps7	16.13598328	16.13598328	16.12314465	16.13598328	16.16225112	16.2104266	15.7690479	15.9113763	15.54340319	15.66058358	15.93160941	15.85599438
17	20102	Rps4x	15.99551072	16.02104671	15.96795367	15.96795367	16.05120582	16.035313	15.91137633	15.9815401	15.88458581	15.88458581	15.96795367	15.87372876
18	67891	Rpl4	16.05120582	16.0103654	16.08954522	16.05120582	16.10255505	16.0210467	15.96795367	15.9181346	16.13598328	16.07143124	16.10255505	15.98154012

# morpheus.js

- Веб-приложение, созданное для визуализации и анализа числовых матриц
- Создан на JavaScript
- Не требует информатической подготовки
- Имеет реализацию некоторых функций, но их **недостаточно** для полноценного анализа



# R/Bioconductor

- R - язык программирования для статистической обработки данных
- Bioconductor - хранилище готовых реализаций биоинформатических алгоритмов и методов обработки биологических данных
- Для использования Bioconductor необходимо знание R

# Цель работы

Разработать веб-приложение, интегрирующее возможности визуального анализа morpheus.js и методы анализа библиотек Bioconductor

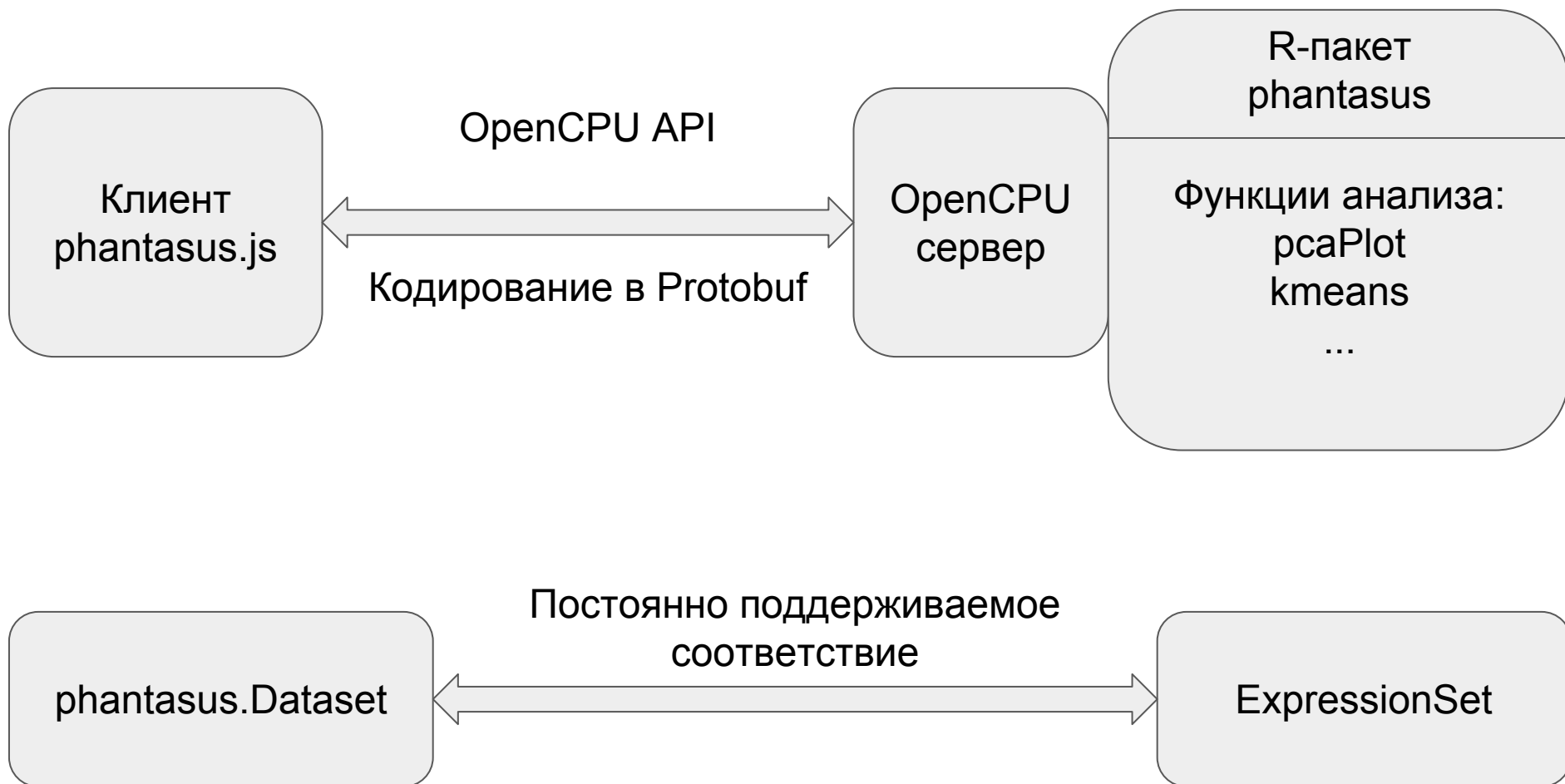
# Задачи

- Разработать способ взаимодействия между js-клиентом и R (и встроить его в morpheus.js)
- Реализовать интерфейс для нескольких стандартных методов анализа
- Соединить все в одном веб-приложении **phantasus**

# Взаимодействие через OpenCPU API

- Система для интеграции R через HTTP API
- Имеет библиотеку `оренспи.js` для работы из JavaScript, реализованную с помощью Ajax
- Для каждого вызова OpenCPU создает новую временную сессию, ключ которой в дальнейшем можно использовать для новых вызовов

# Схема взаимодействия





# Реализованные функции в phantastus

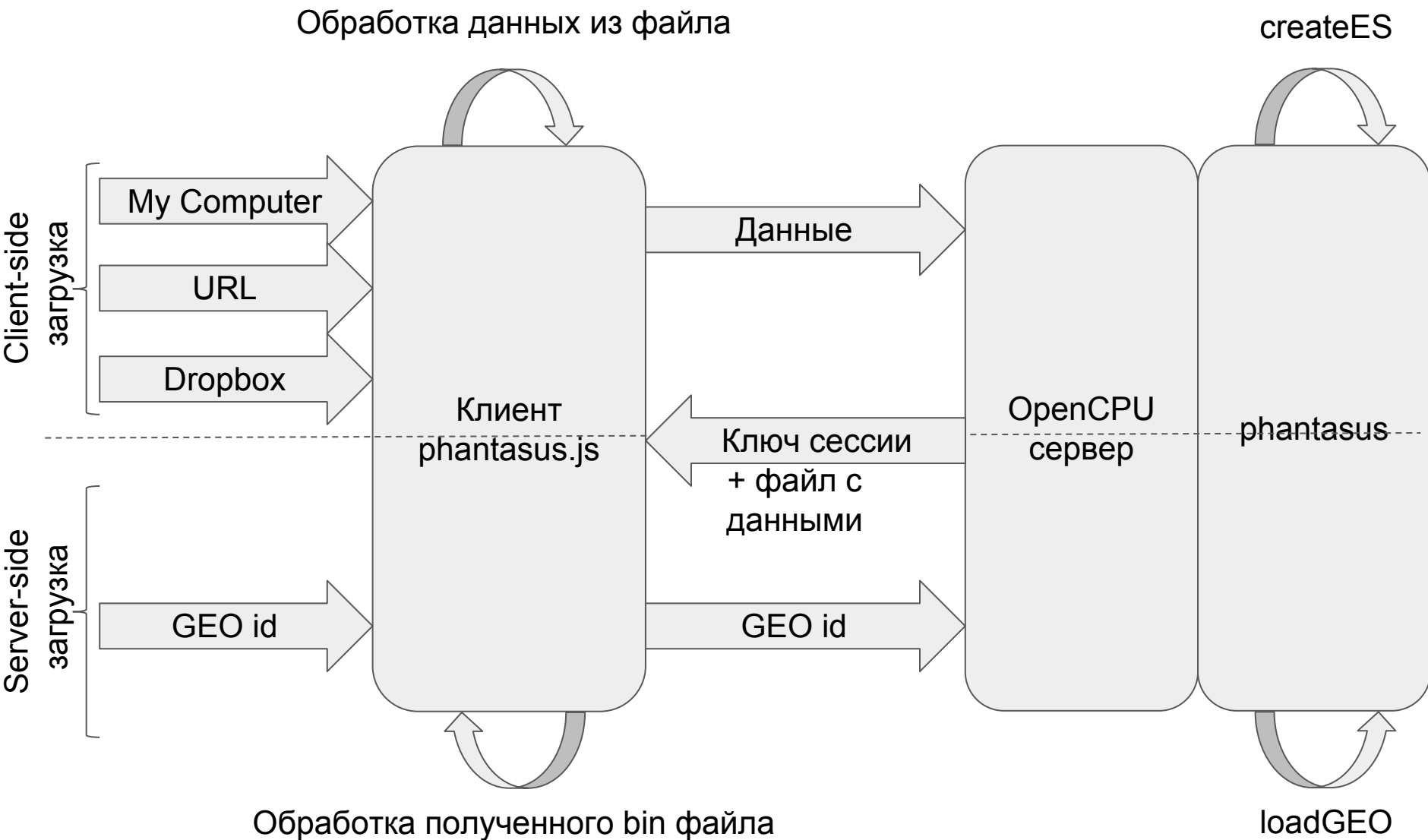
Необходимый минимум для исследования:

- **loadGEO** - загрузка и визуализация данных по идентификатору в репозитории Gene Expression Omnibus (GEO)
- **pcaPlot** - реализация метода главных компонент и визуализация результата в виде интерактивного графика с помощью Plotly.js
- **kmeans** - кластеризация генов

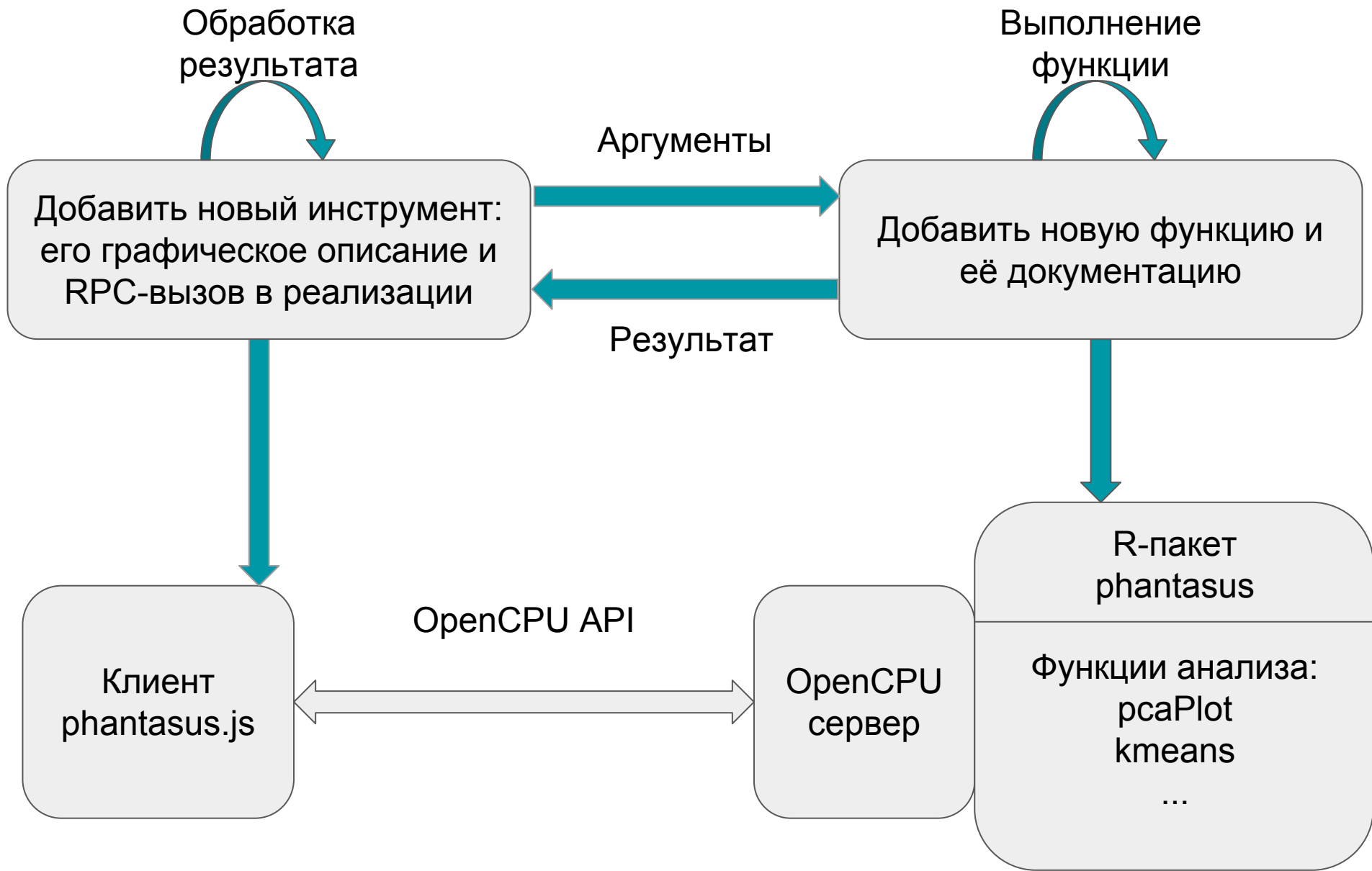
Дополнительные функции:

- **limma** - анализ дифференциальной экспрессии для сравнения образцов

# Схема начала работы с данными



# Добавление нового функционала



# Различия между phantasm.js и morpheus.js

- Добавлена поддержка ProtoBuf со стороны клиента
- Графические интерфейсы для функций pcaPlot, kmeans, limma
- Код для поддержки актуального ключа сессии с ExpressionSet

# Запуск приложения

- Два варианта запуска и использования Web-приложения:
  1. Запустить R-пакет phantusus через R непосредственно
  2. Загрузить Docker-образ и запустить от него Docker-контейнер
- Приложение работает в режиме single-user
- Для запуска параллельных сессий используется balancer в apache2

# Внедрение

- Веб-приложение используется в лаборатории Максима Артемова в Washington University in St. Louis
- Демонстрация приложения входит в программу семинара по системной биологии:
  - Сидней, Австралия, 10-13 апреля 2017 г.  
(<https://register.gimr.garvan.org.au/systemsbiology/>)
  - Санкт-Петербург, 14-19 мая 2017 г.  
(<http://bioinformaticsinstitute.ru/sbw2017>)

# Доступность

- <https://artyomovlab.wustl.edu/phantasus>
- Скоро на genome.ifmo.ru
- Код доступен на github под MIT
  - <https://github.com/ctlab/phantasus> - R-пакет
  - <https://github.com/ctlab/phantasus.js> - js-клиент
- Docker-образ
  - <https://hub.docker.com/r/dzenkova/phantasus>

# Результаты

- Применен способ взаимодействия между js-клиентом и R через OpenCPU
- Реализован интерфейс для функций: pcaPlot, kmeans, limma
- Добавлена возможность загружать данные из GEO
- Все компоненты соединены в приложении phantasus
- Приложение было внедрено в лаборатории, а также представлено на семинаре по системной биологии



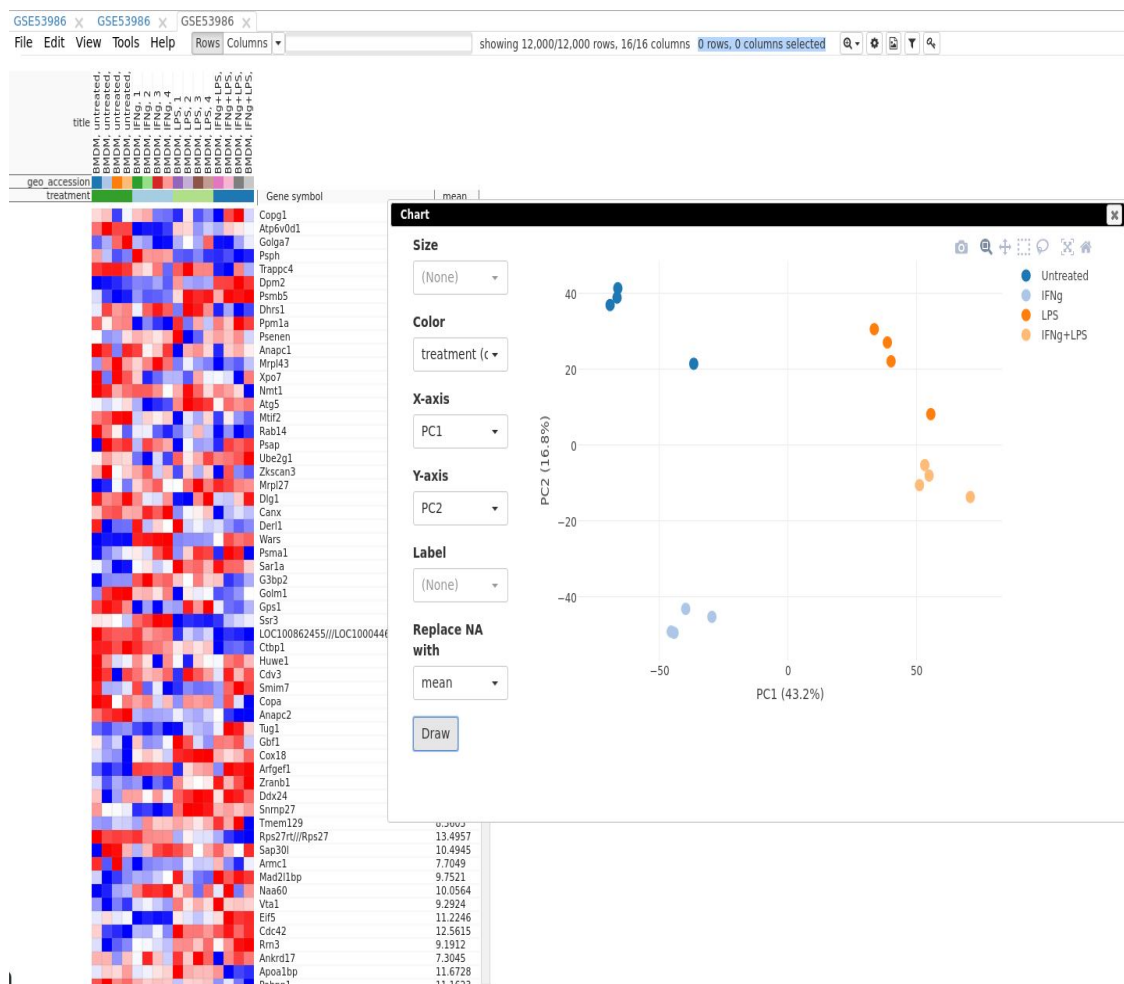
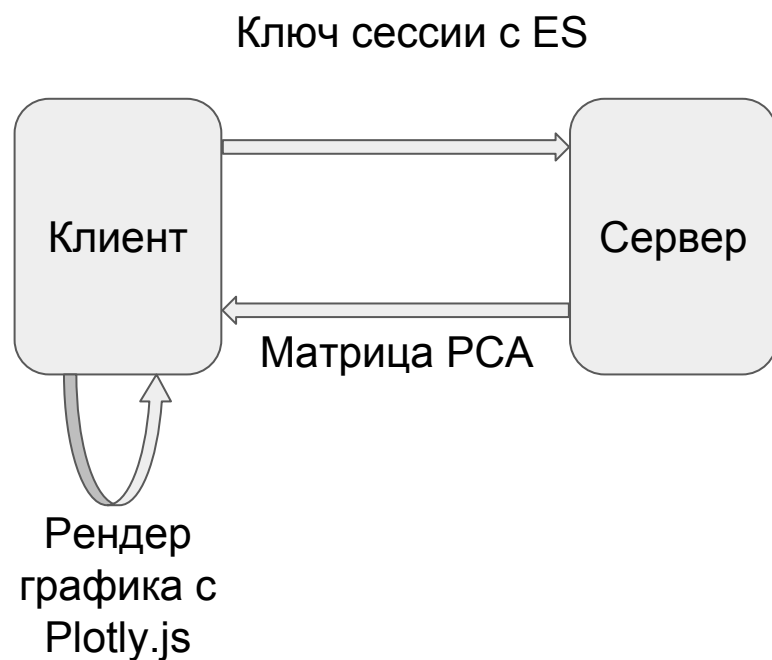
# Дальнейшая деятельность по проекту

- Добавить автоматическое тестирование для JavaScript-кода и R-кода проекта phantasus
- Подробно документировать весь исходный код
- Подготовить проект к добавлению в Bioconductor и подать заявку
- Внедрить Google Analytics для сбора статистики об использовании веб-приложения

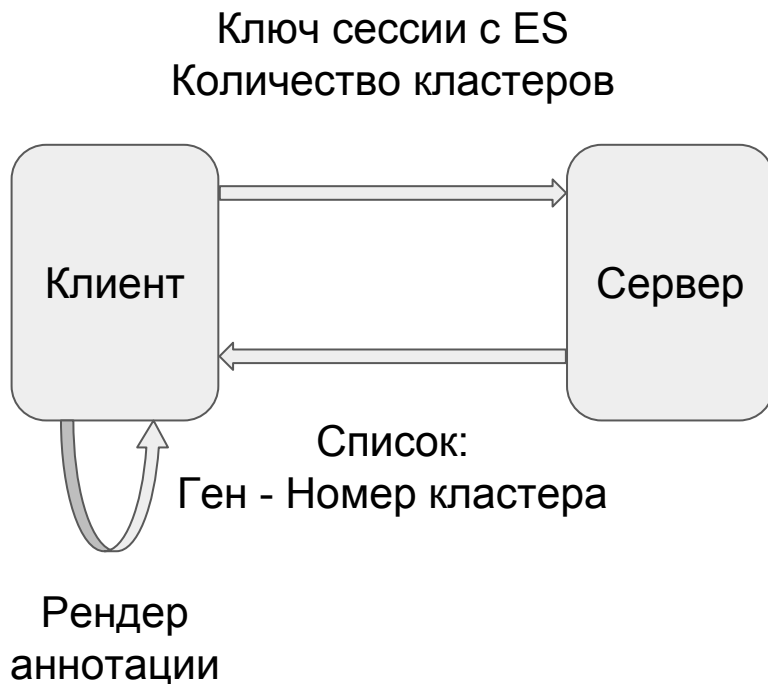
# Реализация эффективного взаимодействия между платформой для анализа экспрессии генов Morpheus и библиотекой вычислительных методов R/Bioconductor

Зенкова Д.М.,  
группа М3436,  
руководитель: Сергушичев А.А., к.т.н.

# Визуализация метода главных компонент: pcaPlot



# Кластеризация генов: kmeans

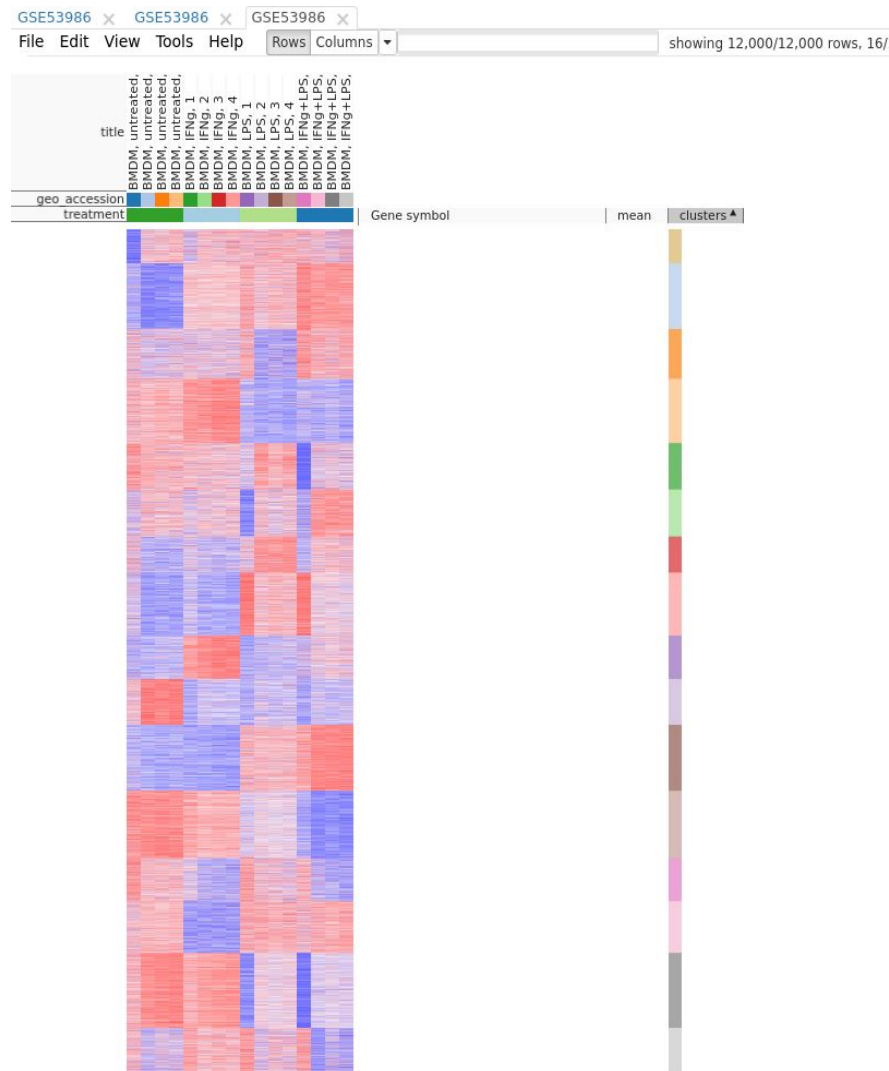


k-means

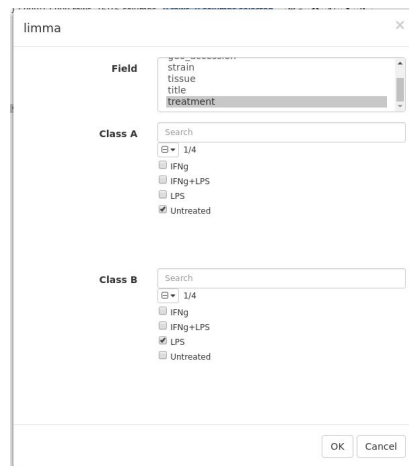
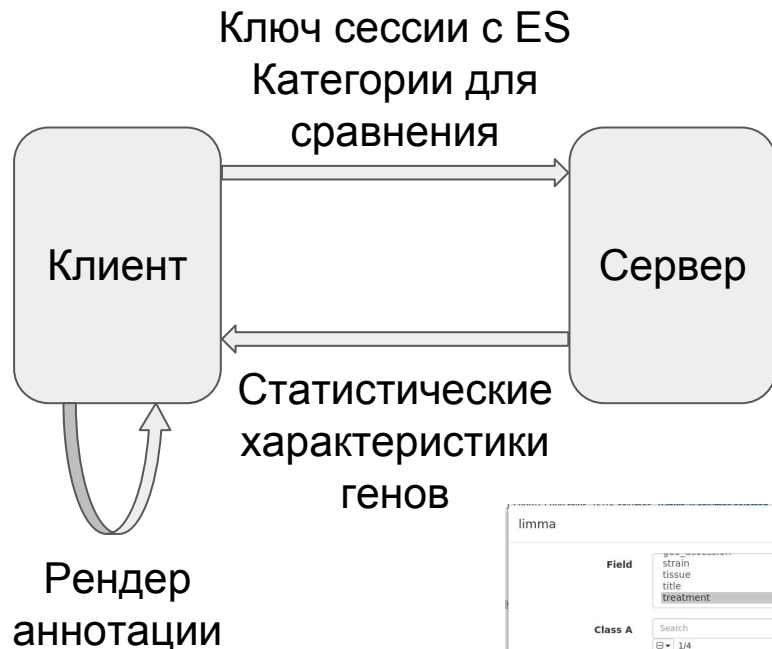
Number of clusters

Replace NA with

OK Cancel



# Анализ дифференциальной экспрессии: limma



Gene symbol	mean	Gene.symbol	logFC	AveExpr	P.Value	adj.P.Val	B
Ptgs1	2.153446	AU019752	-4.7299	6.8897	-33.7916	1.42e-10	0.00000279
Mrc1	4.400463	Hebp1	-4.6402	10.3005	-26.7063	9.95e-9	0.00000543
Smagp	2.616073	Cacng8	-3.5729	8.5854	-26.6364	0.079e-9	0.00000543
Pik	4.289832	Gprc5c	-3.4328	7.1617	-22.3601	6.63e-8	0.00001262
Wwp1	2.918765	Cttnb	-2.7907	11.4047	-21.3242	3.75e-8	0.00001583
Nrpl	11.361985	4933430A20Rik	-2.6727	10.2323	-20.9312	2.75e-8	0.00001737
Sesn1	3.523221	Elfn2	-2.9477	9.8373	-18.7456	5.58e-8	0.00002811
Ank	4.407593	Hes6	-3.5725	9.5909	-17.7051	1.027e-7	0.0000385
Cd300b	4.188422	Gm7040	-2.8749	11.9937	-16.167	2.091e-7	0.00005979
S100a4	2.495393	C1qc	-2.9479	13.0832	-16.0822	2.178e-7	0.00005979
Agmo	3.578526	Em1	-3.3263	6.7793	-16.0672	1.94e-7	0.00005979
Camk1	11.06421	4930553M12Rik	-1.589	11.0224	-16.0225	2.242e-7	0.00005979
Igf1bp4	1.750764	Ang2	-4.2903	8.2702	-15.605	2.755e-7	0.00007033
Arl6ip5	10.81545	4930442L01Rik	-1.6204	12.1591	-15.2903	2.28e-7	0.00007309
Sic9a9	3.524297	Elk3	-2.8912	5.7782	-14.8354	0.083e-7	0.00008124
Ccsap	2.559363	C78516	-2.6111	6.2685	-14.814	4.130e-7	0.00008124
Nckip5d	12.07178	9230110K08Rik	-1.4478	8.5503	-14.7304	0.314e-7	0.0000835
1700047117Rik	1.479253	Gm6086	-1.4294	7.9434	-14.6994	0.387e-7	0.00008356
Mex3b	3.861639	Gats3	-2.4208	6.5764	-14.4934	0.894e-7	0.00008765
Sic25a23	12.12861	9330198N18Rik	-1.7916	5.6677	-14.4475	0.016e-7	0.00008823
Cp183	3.877619	Gd9	-3.4126	6.176	-14.4265	0.073e-7	0.00008823
Ugt1a1	11.24634	4933406P04Rik	-1.6889	12.1842	-14.275	5.215e-7	0.00008941
Ccna2	3.922701	Gira4	-6.8163	6.1045	-14.1046	0.041e-7	0.00009176
Rasgrp3	3.913857	Gk2	-2.4973	7.6857	-14.0476	0.234e-7	0.00009236
D030029J20Rik	3.834699	Gabpb1	-2.2223	11.6108	-14.0026	0.739e-7	0.00009307
2810417H13Rik	12.20312	94330081G06Rik	-2.0882	6.0373	-13.9466	0.591e-7	0.00009307
Klhl6	3.914175	Gk5	-1.6982	9.2855	-13.9256	0.670e-7	0.00009307
Gcnt1	4.365015	H2-Q6	-1.961	8.0979	-13.8377	0.005e-7	0.00009552
Cmb1	4.468338	Hoxb6	-4.8535	6.0207	-13.7127	0.514e-7	0.00009682
Dhcr24	11.40410	4933435C09Rik	-2.5654	6.3182	-13.675	0.767e-7	0.00009682
Rabb6	3.664375	Fam222b	-2.3533	5.5656	-13.6407	0.826e-7	0.00009682
Cdkn2c	10.56447	3632451O06Rik	-2.375	9.0756	-13.590	0.805e-7	0.00009758
Sqle	10.06858	1700025H01Rik	-2.8813	8.622	-13.5348	0.312e-7	0.00009893
Rassf5	1.929832	Arhgap39	-1.6545	9.7071	-13.414	0.902e-7	0.0001033
C79468	4.839185	Kim29	-2.5699	8.1857	-13.400	0.976e-7	0.0001033
Tie1	1.970265	Ar144ep	-2.15	8.059	-13.309	0.492e-7	0.0001033
Smim14	10.51009	2900069G24Rik	-1.3446	11.922	-13.301	0.950e-7	0.0001033
Pla2g15	1.841938	Ap3b1	-2.0014	9.2548	-13.2919	0.559e-7	0.0001033
Galnt12	3.902892	Gins3	-2.6857	7.8858	-13.121	0.000001056	0.0001094
Fam214a	3.779709	Fndc1	-3.2895	6.6865	-13.1180	0.000001057	0.0001094
Hnmt	11.12652	4930577N17Rik	-1.7813	10.1849	-12.897	0.000001205	0.000117
Mfsd6	12.19238	9430051O21Rik	-2.2356	8.8062	-12.879	0.000001218	0.000117
Arap3	1.286180	AA409749	-3.8325	6.0096	-12.874	0.000001220	0.000117
Vsir	10.42422	2810030D12Rik	-2.2133	11.6636	-12.865	0.000001229	0.000117
2610507B11Rik	2.868703	Cer1	-1.4922	10.2759	-12.549	0.000001487	0.0001332
Atp8b4	4.007970	Gm15834	-2.0103	8.0846	-12.4930	0.000001539	0.0001348
Rpl10a	3.377841	Dnajc16	-1.1929	13.7835	-12.380	0.00000165	0.0001395
Ptchd1	4.185098	Gm6682	-3.4829	6.1504	-12.3590	0.000001671	0.0001395
Gmnn	11.00079	4930524008Rik	-1.1411	10.5664	-12.1620	0.00000189	0.0001486
Cdk1	4.259852	Gpalp1	-2.0061	8.7983	-12.0980	0.000001967	0.0001523
Bmyc	3.129025	Cttnl1	-3.0108	7.5333	-11.986	0.000002114	0.0001595
Gas6	1.092159	4930502E18Rik	-2.3928	8.3302	-11.853	0.000002301	0.0001694
Tpcn1	3.667151	Fam2234a	-1.9105	6.3194	-11.7950	0.000002380	0.0001724
Frmf6	4.461250	Hornd2	-2.0006	9.4002	-11.7790	0.000002413	0.0001724
Fam105a	3.713896	Fbxo31	-1.7173	8.6763	-11.7790	0.000002410	0.0001724

# Gene Expression Omnibus

Series **GSE14308** [Query DataSets for GSE14308](#)

Status Public on Jan 08, 2009  
Title Epigenetic Mechanisms Underlie T Cell Plasticity  
Organism [Mus musculus](#)  
Experiment type Expression profiling by array  
Summary Multipotential naïve CD4+ T cells differentiate into distinct lineages including T helper 1 (Th1), Th2, Th17, and inducible T regulatory (iTreg) cells. The remarkable diversity of CD4+ T cells begs the question whether the observed changes reflect terminal differentiation with heritable epigenetic modifications or plasticity in T cell responses. We generated genome-wide histone H3 lysine 4 (H3K4) and lysine 27 (H3K27) trimethylation maps in naïve, Th1, Th2, Th17, iTreg, and natural (n)Treg cells. We found that although modifications of signature cytokine genes (Ifng, Il4, and Il17) partially conform to the expectation of lineage commitment, critical transcription factors such as Tbx21 exhibit a broad spectrum of epigenetic states, consistent with our demonstration of T-bet and IFN-gamma induction in nTreg cells. Our data suggest an epigenetic mechanism underlying the specificity and plasticity of effector and regulatory T cells and also provide a framework for understanding complexity of CD4+ T helper cell differentiation.

Overall design Different T helper subsets are profiled for mRNA expression.

Contributor(s) [Wei L, Wei G, Zhu J, Hu-Li J, O'Shea JJ, Zhao K](#)

Citation(s) Wei G, Wei L, Zhu J, Zang C et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity* 2009 Jan 16;30(1):155-67. PMID: [19144320](#)

Submission date Jan 06, 2009

Last update date Apr 17, 2017

Contact name Lai Wei

E-mail [weil2@mail.nih.gov](mailto:weil2@mail.nih.gov)

Phone 3014961480

Organization name NIH/NEI/NCCAM

Street address 10 Center Dr. Room 2B47

City Bethesda

State/province MD

ZIP/Postal code 20892

Country USA

Platforms (1) [GPL1261](#) [Mouse430\_2] Affymetrix Mouse Genome 430 2.0 Array

Samples (12) [GSM357839](#) Th2-1

[More...](#)

[GSM357841](#) Th2-2

[GSM357842](#) Th1-2

Samples (12)

[Less...](#)

[GSM357839](#) Th2-1

[GSM357841](#) Th2-2

[GSM357842](#) Th1-2

[GSM357843](#) Th17-1

[GSM357844](#) Th1-1

[GSM357845](#) Th17-2

[GSM357847](#) Naive-1

[GSM357848](#) Naive-2

[GSM357849](#) iTreg-1

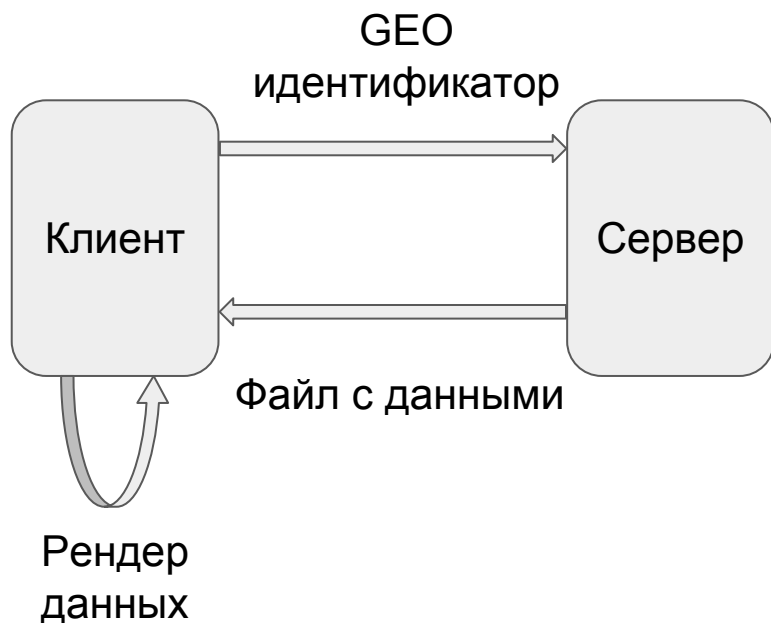
[GSM357850](#) iTreg-2

[GSM357852](#) nTreg-1

[GSM357853](#) nTreg-2

<https://www.ncbi.nlm.nih.gov/geo>

# Получение данных из GEO: loadGEO



GSE14308 x

File Edit View Tools Help Rows Columns showing 45,101/45,101 rows, 12/12

title	geo_accession	characteristics_ch1	id	Gene symbol	UniGene symbol
GSM357839Th2-1	GSM357839		1415670_at	Copg1	
GSM357841Th2-2	GSM357841		1415671_at	Atp6v0d1	
GSM357842Th1-2	GSM357842		1415672_at	Golga7	
GSM357843Th1-1	GSM357843		1415673_at	Psph	
GSM357844Th1-1	GSM357844		1415674_a_at	Trappc4	
GSM357845Th1-2	GSM357845		1415675_at	Dpm2	
GSM357847Naive-1	GSM357847		1415676_a_at	Psmb5	
GSM357848Naive-2	GSM357848		1415677_at	Dhrs1	
GSM357849Treg-1	GSM357849		1415678_at	Ppm1a	
GSM357850Treg-2	GSM357850		1415679_at	Psenen	
GSM357852nTreg-1	GSM357852		1415680_at	Anapc1	
GSM357853nTreg-2	GSM357853		1415681_at	Mrpl43	
			1415682_at	Xpo7	
			1415683_at	Nmt1	
			1415684_at	Atg5	
			1415685_at	Mtif2	
			1415686_at	Rab14	
			1415687_a_at	Psap	
			1415688_at	Ube2g1	
			1415689_s_at	Zkscan3	
			1415690_at	Mrpl27	
			1415691_at	Dlg1	
			1415692_s_at	Canx	
			1415693_at	Derl1	
			1415694_at	Wars	
			1415695_at	Psmal1	
			1415696_at	Sar1a	
			1415697_at	G3bp2	
			1415698_at	Golm1	
			1415699_a_at	Gps1	
			1415700_a_at	Ssr3	
			1415701_x_at	LOC100862455//LOC100044627//Rpl23	
			1415702_a_at	Ctbp1	
			1415703_at	Huwl1	
			1415704_a_at	Cdv3	
			1415705_at	Smim7	
			1415706_at	Copa	
			1415707_at	Anapc2	
			1415708_at	Tug1	
			1415709_s_at	Gbf1	
			1415710_at	Cox18	
			1415711_at	Arfgef1	
			1415712_at	Zranb1	
			1415713_a_at	Ddx24	
			1415714_a_at	Snmp27	
			1415715_at	Tmem129	
			1415716_a_at	Rps27rt//Rps27	
			1415717_at	Rnf220	
			1415718_at	Sap30l	
			1415719_s_at	Armcl	
			1415720_s_at	Mad2l1bp	
			1415721_a_at	Naa60	
			1415722_a_at	Vtal	
			1415723_at	Elf5	
			1415724_a_at	Cdc42	
			1415725_at	Rm3	
			1415726_at	Ankrd17	