

TCGA Mutations MIPT

Виктория Гросс

March 2019

1 Введение в использование базы данных TCGA

The Cancer Genome Atlas (TCGA) — это база данных, содержащая и способная анализировать данные о человеческих опухолях, созданная как инструмент для исследования молекулярных нарушений в организме, приводящих к раку. Существует множество интерфейсов, используемых для получения данных с портала, а также много уже написанных анализаторов. Наиболее известный — портал от National Cancer Institute (portal.gdc.cancer.gov). В разделе Exploration можно подобрать случаи необходимые для исследования: местоположение опухоли, база данных (TCGA или другие подобные), тип рака. Данные можно отфильтровать по характеристикам пациентов (пол, возраст, раса, информация о выживании). Для выбранных случаев можно скачать файлы, содержащие данные об уровне экспрессии различных генов, информацию о мутациях, клинические данные пациента. Затем эти данные в формате csv или json могут быть обработаны стандартными методами обработки данных python или R.

Данные могут экспортированы при помощи специальных библиотек: xenaPython для Python или RTCGA для R.

Кроме того, существует множество уже готовых интерфейсов, позволяющих определить, например, зависимость выживаемости от уровня экспрессии того или иного белка. Например, <http://kmplot.com/analysis/>. Создатели этого сайта, использовали библиотеки языка R для получения данных, нормировки и анализа данных с трех баз данных: TCGA, MetaBricks, Affymetrix. Аналогичными методами, описанными в их статье, можно повторить их результаты для выбранной базы данных.

1.1

2 С помощью базы данных узнать какие мутации чаще всего бывают в комплексе Arp2/3 с помощью базы данных

[!h]		Таблица 1: Mutations in ARP2/3	
bcr patient barcode	dataset	Hugo Symbol	Variant Classification
TCGA-OR-A5KB-01	ACC.mutations	ARPC4	Silent
TCGA-OR-A5KY-01	ACC.mutations	ARPC1B	Silent

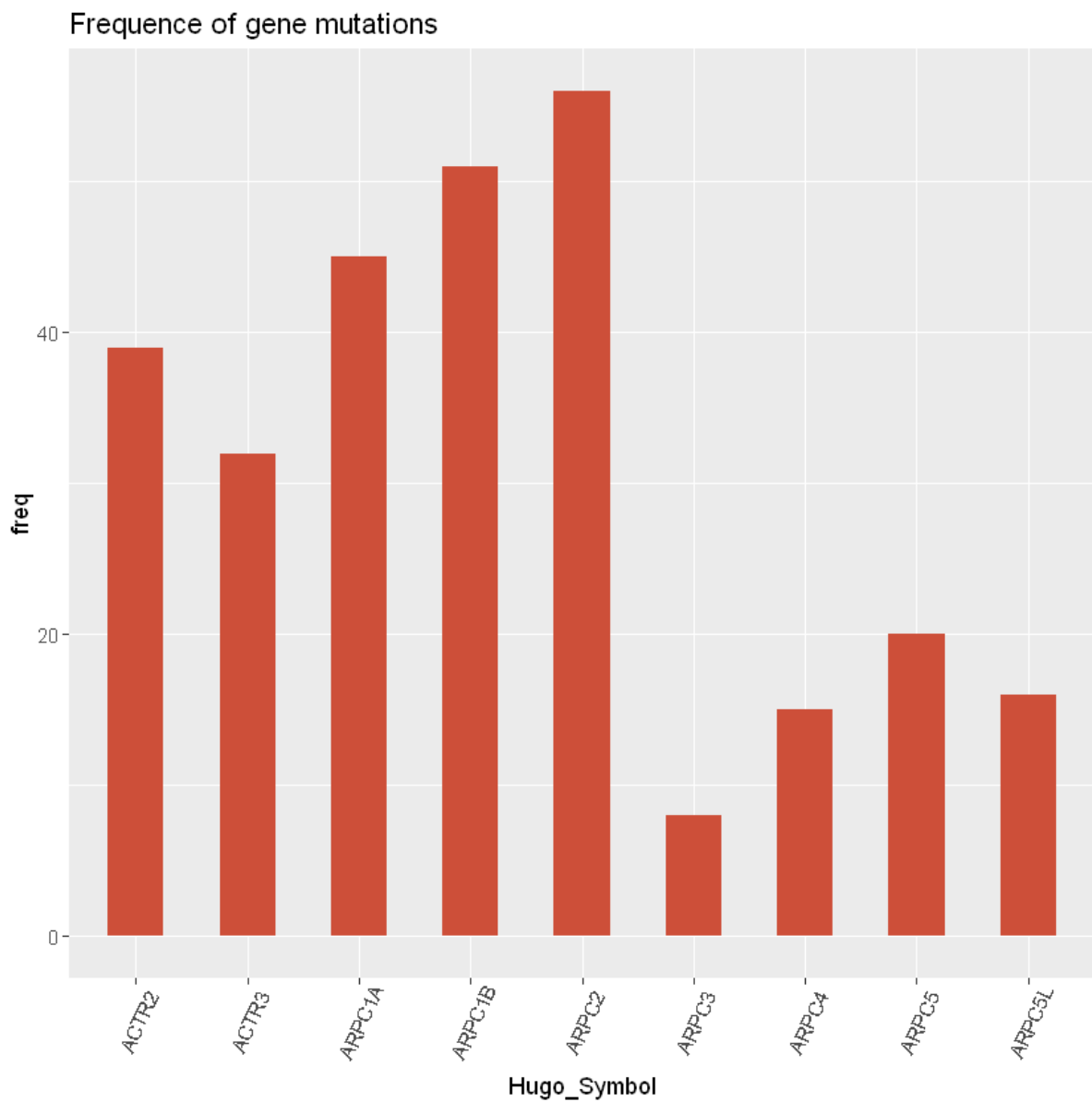
bcr patient barcode	dataset	Hugo Symbol	Variant Classification
TCGA-OR-A5LD-01	ACC.mutations	ARPC2	Frame Shift Ins
TCGA-BT-A3PK-01	BLCA.mutations	ARPC2	Missense Mutation
TCGA-CF-A3MI-01	BLCA.mutations	ARPC5L	Silent
TCGA-DK-A2I6-01	BLCA.mutations	ACTR3	Missense Mutation
TCGA-DK-A3WW-01	BLCA.mutations	ARPC2	Missense Mutation
TCGA-DK-A3X1-01	BLCA.mutations	ARPC2	Missense Mutation
TCGA-FD-A3SO-01	BLCA.mutations	ARPC5	Missense Mutation
TCGA-FD-A3SS-01	BLCA.mutations	ARPC3	Silent
TCGA-K4-A3WS-01	BLCA.mutations	ARPC4	Missense Mutation
TCGA-A2-A0T5-01	BRCA.mutations	ARPC2	RNA
TCGA-A2-A0YK-01	BRCA.mutations	ARPC1A	Silent
TCGA-A7-A26J-01	BRCA.mutations	ARPC5L	In Frame Del
TCGA-A7-A426-01	BRCA.mutations	ARPC1A	Nonsense Mutation
TCGA-A8-A08L-01	BRCA.mutations	ARPC2	Missense Mutation
TCGA-A8-A097-01	BRCA.mutations	ARPC1A	Missense Mutation
TCGA-A8-A09I-01	BRCA.mutations	ACTR2	Silent
TCGA-A8-A09Z-01	BRCA.mutations	ARPC1A	Silent
TCGA-AC-A23H-01	BRCA.mutations	ACTR3	Missense Mutation
TCGA-AC-A23H-01	BRCA.mutations	ARPC2	Missense Mutation
TCGA-AC-A23H-01	BRCA.mutations	ARPC5	Missense Mutation
TCGA-AN-A046-01	BRCA.mutations	ACTR2	Missense Mutation
TCGA-AN-A046-01	BRCA.mutations	ACTR3	Nonsense Mutation
TCGA-AN-A046-01	BRCA.mutations	ARPC2	Nonsense Mutation
TCGA-AN-A046-01	BRCA.mutations	ARPC2	RNA
TCGA-AN-A04D-01	BRCA.mutations	ARPC1A	Silent
TCGA-AN-A0XT-01	BRCA.mutations	ACTR2	RNA
TCGA-AQ-A54O-01	BRCA.mutations	ARPC5	RNA
TCGA-B6-A0IP-01	BRCA.mutations	ACTR2	Silent
TCGA-ZR-A9CJ-01	STES.mutations	ARPC4	Silent
TCGA-4K-AA1H-01	TGCT.mutations	ARPC1B	Missense Mutation
TCGA-XE-AAOC-01	TGCT.mutations	ARPC2	Missense Mutation
TCGA-A5-A0GB-01	UCEC.mutations	ARPC1B	Silent
TCGA-A5-A0VP-01	UCEC.mutations	ARPC1A	Missense Mutation
TCGA-AP-A051-01	UCEC.mutations	ARPC3	Missense Mutation
TCGA-AP-A056-01	UCEC.mutations	ACTR3	Nonsense Mutation
TCGA-AP-A059-01	UCEC.mutations	ACTR2	Missense Mutation
TCGA-AP-A059-01	UCEC.mutations	ARPC1B	Missense Mutation
TCGA-AP-A0LE-01	UCEC.mutations	ARPC1A	Missense Mutation
TCGA-AP-A0LM-01	UCEC.mutations	ACTR2	Silent
TCGA-AP-A0LM-01	UCEC.mutations	ARPC2	Missense Mutation
TCGA-AP-A0LM-01	UCEC.mutations	ARPC1A	Silent
TCGA-AX-A0J1-01	UCEC.mutations	ARPC1A	Missense Mutation
TCGA-B5-A0JR-01	UCEC.mutations	ACTR2	Missense Mutation
TCGA-B5-A0JY-01	UCEC.mutations	ARPC5	Missense Mutation
TCGA-B5-A0JY-01	UCEC.mutations	ARPC1A	Missense Mutation
TCGA-B5-A0JY-01	UCEC.mutations	ARPC1B	Missense Mutation
TCGA-B5-A0K6-01	UCEC.mutations	ARPC1A	Missense Mutation

bcr patient barcode	dataset	Hugo Symbol	Variant Classification
TCGA-B5-A11E-01	UCEC.mutations	ACTR2	Missense Mutation
TCGA-BS-A0UF-01	UCEC.mutations	ACTR2	Missense Mutation
TCGA-BS-A0UF-01	UCEC.mutations	ARPC1B	Missense Mutation
TCGA-D1-A103-01	UCEC.mutations	ACTR2	Missense Mutation
TCGA-D1-A15V-01	UCEC.mutations	ARPC1B	Missense Mutation
TCGA-D1-A17D-01	UCEC.mutations	ARPC1A	Splice Site
TCGA-D1-A17Q-01	UCEC.mutations	ARPC1B	Missense Mutation
TCGA-EY-A1GS-01	UCEC.mutations	ARPC4	Missense Mutation
TCGA-N8-A4PP-01	UCS.mutations	ACTR3	Missense Mutation
TCGA-V4-A9EY-01	UVM.mutations	ARPC2	Missense Mutation
TCGA-VD-AA8O-01	UVM.mutations	ARPC1B	Missense Mutation

Посмотрим частоту встречаемости мутаций в отдельных белках комплекса:

Hugo Symbol	freq
ACTR2	39
ACTR3	32
ARPC1A	45
ARPC1B	51
ARPC2	56
ARPC3	8
ARPC4	15
ARPC5	20
ARPC5L	16

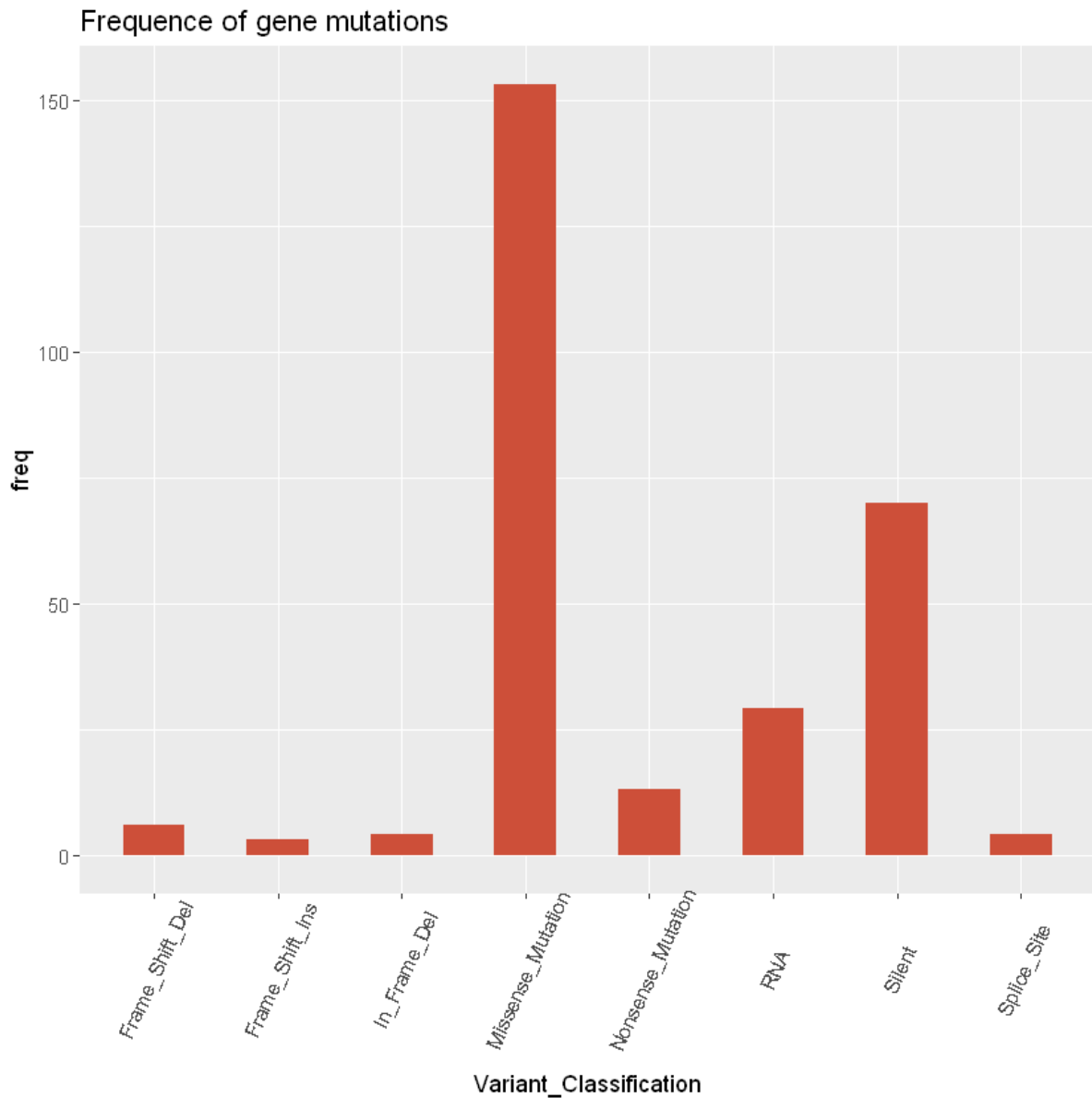
mutations frequency.png



И на частоту встречаемости мутаций различного типа:

Variant Classification	freq
Frame Shift Del	6
Frame Shift Ins	3
In Frame Del	4
Missense Mutation	153
Nonsense Mutation	13
RNA	29
Silent	70
Splice Site	4

type.png



- 3 Отобразить места, где эти мутации встречаются чаще всего, на структуре комплекса Arp2/3
- 4 Проверить - могут ли те или иные места, в которых бывают мутации, быть задействованы в взаимодействии комплекса с лигандами или другими белками