

Исследование базы данных Immune Epitope Database на предмет необходимой информации для моделирования процессинга антигенов для HLA первого класса

Антон Смирнов

Feb 16, 2023

Оглавление

Введение	2
Версии используемых библиотек	2
R	2
python	3
MySQL	3
Подготовка данных	4
Поиск расположения необходимых данных	4
Получение данных	7
Анализ bind assays	9
Эксперименты	9
Измеряемые метрики	10
Распределение длины эпитопов и аминокислотного состава	13
Виды - источники эпитопов	15
Распределение аллелей HLA	15
Анализ elution assays	16
Эксперименты	16
Половозрастная характеристика испытуемых	17
Информация по антиген-презентирующим клеткам	17
Распределение длины эпитопов и аминокислотного состава	21
Виды - источники эпитопов	23
Распределение аллелей HLA	23
Объединенный датасет	25
Приложение	27
Полная статистика	27

Введение

Immune Epitope Database представляет из себя крупнейшую базу данных и базу знаний о Т- и В- клеточных эпитопах. Она курируется Национальным институтом аллергии и инфекционных болезней США. Количество информации в ней на 16 февраля 2023 года представлено в таблице 1.

Table 1: Метрики БД

Характеристика	Количество
Peptidic Epitopes	1,554,329
Non-Peptidic Epitopes	3,168
T Cell Assays	453,372
B Cell Assays	1,378,855
MHC Ligand Assays	4,697,592
Epitope Source Organisms	4,337
Restricting MHC Alleles	983
References	23,554

Дамп базы данных был скачан с сайта iedb.org и развернут на локальном сервере MySQL.

Версии используемых библиотек

R

[1] "R version 4.2.2 Patched (2022-11-10 r83330)"

[1] "RMySQL 0.10.24"

[1] "dplyr 1.0.10"

[1] "stringr 1.4.1"

[1] "knitr 1.40"

[1] "vroom 1.6.0"

[1] "ggplot2 3.3.6"

python

Python

3.9.12 (main, Apr 5 2022, 07:05:27)

[GCC 7.5.0]

logomaker 0.8

matplotlib 3.5.1

MySQL

mysql Ver 15.1 Distrib 10.6.11-MariaDB, for debian-linux-gnu (x86_64) using EditLine wrapper

Скрипт восстановления БД.

mysql -u stotoshka -p

mysql> create database iedb;

mysql> use iedb;

mysql> source <path to sql_script>;

Подключимся к базе данных

Подготовка данных

Поиск расположения необходимых данных

Для начала посмотрим каких типов эксперименты хранятся в этой базе данных.

```
SELECT DISTINCT category FROM assay_type;
```

Table 2: 4 records

category
T Cell
MHC
B Cell
Naturally Processed

В данный момент нас не интересуют эксперименты на В-лимфоцитах. Посмотрим, какие эксперименты проводят на Т-клетках.

```
SELECT DISTINCT assay_type FROM assay_type WHERE category = 'T Cell';
```

assay_type
RNA/DNA detection
ICS
ELISPOT
cytometric bead array
ELISA
biological activity
in vivo assay

assay_type
x-ray crystallography
binding assay
CFSE
multimer/tetramer
bioassay
51 chromium
3H-thymidine
reporter gene assay
BrdU
surface plasmon resonance (SPR)
in vivo skin test
radio immuno assay (RIA)
intracellular staining
in vitro assay
High throughput multiplexed assay
any method

Эти эксперименты связаны с детекцией распознавания антигена Т-клеточным рецептором. В данный момент они нам скорее всего не подходят, потому что эти эксперименты не ставили задачу определения связывания эпитопа и HLA. Они могут внести сильное смещение в прогноз, так как содержат большое количество положительных примеров.

```
SELECT DISTINCT assay_type FROM assay_type WHERE category = 'MHC';
```

assay_type
purified MHC/competitive/fluorescence
lysate MHC/direct/radioactivity
cellular MHC/competitive/radioactivity
cellular MHC/T cell inhibition
purified MHC/direct/radioactivity
cellular MHC/direct/radioactivity
purified MHC/competitive/radioactivity
purified MHC/direct/fluorescence
x-ray crystallography
any method

assay_type
binding assay
cellular MHC/direct/fluorescence
lysate MHC/direct/fluorescence
cellular MHC/competitive/fluorescence
purified MHC/direct/phage display
cellular MHC
purified MHC
lysate MHC/competitive/radioactivity
High throughput multiplexed assay
lysate MHC

```
SELECT DISTINCT assay_type FROM assay_type WHERE category = 'Naturally Processed';
```

Table 5: 7 records

assay_type
secreted MHC/mass spectrometry
cellular MHC/mass spectrometry
coelution
Edman degradation
T cell recognition
High throughput multiplexed assay
mass spectrometry

Типы экспериментов 'MHC' и 'Naturally Processed', так как методы, содержащиеся в них, направлены на определение связывания эпитопа и MHC.

```
SELECT DISTINCT at2.category,at2.assay_type FROM mhc_elution me
INNER JOIN assay_type at2 ON as_type_id = at2.assay_type_id;
```

category	assay_type
Naturally Processed	cellular MHC/mass spectrometry
Naturally Processed	coelution
Naturally Processed	secreted MHC/mass spectrometry

category	assay_type
Naturally Processed	T cell recognition
Naturally Processed	Edman degradation
Naturally Processed	mass spectrometry

```
SELECT DISTINCT at2.category,at2.assay_type FROM mhc_bind mb
INNER JOIN assay_type at2 ON as_type_id = at2.assay_type_id
```

category	assay_type
MHC	purified MHC/competitive/radioactivity
MHC	lysate MHC/direct/fluorescence
MHC	cellular MHC/competitive/fluorescence
MHC	x-ray crystallography
MHC	lysate MHC/direct/radioactivity
MHC	cellular MHC/direct/fluorescence
MHC	purified MHC/direct/phage display
MHC	cellular MHC/competitive/radioactivity
MHC	cellular MHC/T cell inhibition
MHC	purified MHC/direct/fluorescence
MHC	purified MHC/competitive/fluorescence
MHC	cellular MHC/direct/radioactivity
MHC	purified MHC/direct/radioactivity
MHC	cellular MHC
MHC	purified MHC
MHC	binding assay
MHC	lysate MHC
MHC	High throughput multiplexed assay

Данные типы находятся в таблицах ***mhc_bind*** и ***mhc_elution***.

Получение данных

Скрипт *iedb_create_mhc_tables.sql* содержит описание первичной подготовки таблиц.

Критерии фильтрации следующие:

- mhc_bind и mhc_elution
 - Должен быть ИД публикации
 - Должен быть ИД эпитопа
 - Должен быть исход эксперимента
 - Должен быть ИД аллеля МНС
- для эпитопов
 - Должна быть линейная последовательность эпитопа
 - Эта последовательность не должна содержать модификаций
 - Должна быть последовательность белка-источника
- для аллеля МНС
 - аллель должен быть человеческий
 - аллель должен быть первого класса МНС
 - аллель должен быть записан в формате HLA-(ген)*(аллельная группа):(белковый продукт)

Количество строк в таблицах

mhc_bind

Table 8: 1 records

COUNT(*)
168071

mhc_elution

Table 9: 1 records

COUNT(*)
1099078

Анализ bind assays

Скрипт *iedb_clean_mhc_bind.R* производит вторичную очистку данных, полученных ранее, по следующим критериям:

1. У каждой записи должен быть источник
2. У каждой записи должен быть исход эксперимента
3. Последовательности должны быть записаны однобуквенным кодом для аминокислот
4. Эпитоп должен находиться в референсной последовательности
5. Эпитоп должен быть полным. Это не должен быть участок антигенного сайта или частью большего эпитопа.
6. Информация в комментариях не должна противоречить исходу эксперимента
7. Результаты должны быть получены экспериментально, не быть предсказанными
8. Эпитопы должны быть иммуногенными
9. Не должно быть дефицита по TAP
10. Все HLA аллели должны быть записаны в формате, позволяющем однозначно идентифицировать молекулу

[1] "Количество строк 139042"

Эксперименты

Table 10: Распределение исходов экспериментов

Value	Number
Negative	66138
Positive	72904

Table 11: Категория экспериментов

Value	Procent
MHC	100

Table 12: Виды экспериментов

Value	Procent
binding assay	0.14
cellular MHC	0.09
cellular MHC/competitive/fluorescence	3.82
cellular MHC/competitive/radioactivity	0.03
cellular MHC/direct/fluorescence	2.86
cellular MHC/direct/radioactivity	0.01
cellular MHC/T cell inhibition	0.02
lysate MHC/direct/fluorescence	0.00
lysate MHC/direct/radioactivity	0.10
purified MHC	0.69
purified MHC/competitive/fluorescence	3.50
purified MHC/competitive/radioactivity	45.95
purified MHC/direct/fluorescence	37.69
purified MHC/direct/radioactivity	4.77
x-ray crystallography	0.33

Измеряемые метрики

Некоторые эксперименты имеют точно измеряемую метрику. Посмотрим, какие метрики используются, их распределение

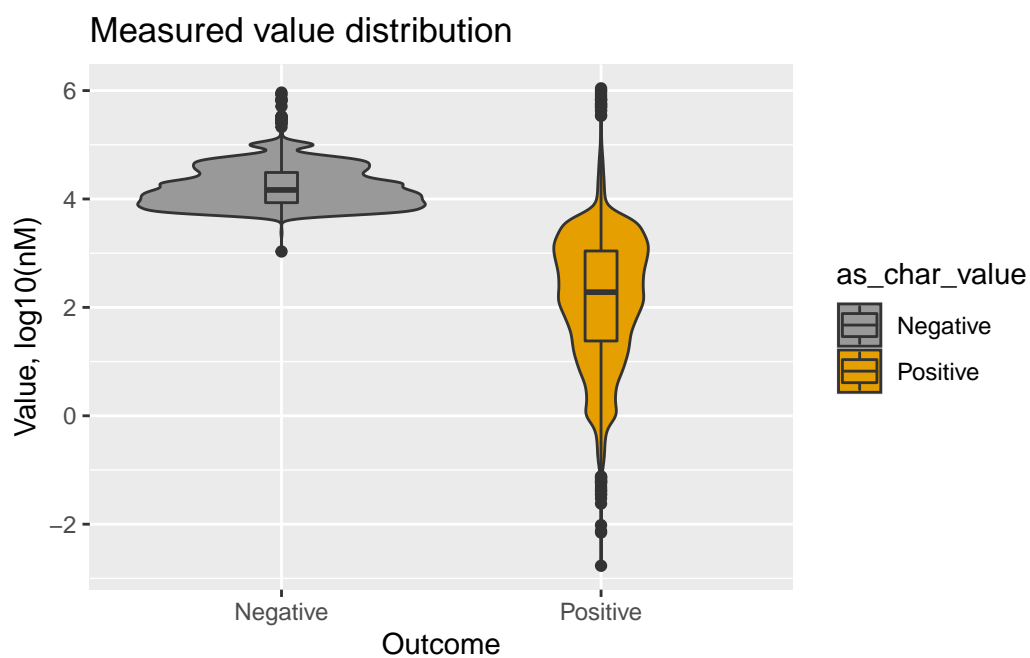
[1] "Количество строк 32269"

Table 13: Ед.измерения

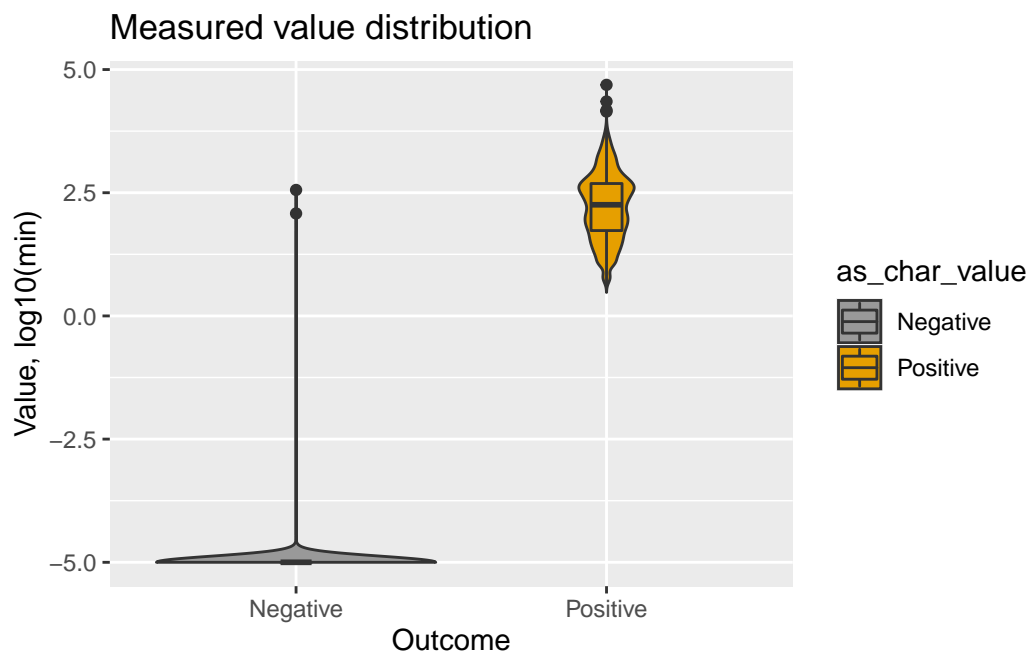
	°C	angstroms	min	nM	NA
cellular MHC/competitive/fluorescence	0	0	0	3861	0

	°C	angstroms	min	nM	NA
cellular MHC/direct/fluorescence	0	0	68	291	54
purified MHC/competitive/fluorescence	0	0	0	693	0
purified MHC/competitive/radioactivity	0	0	0	11001	0
purified MHC/direct/fluorescence	23	0	86	9423	40
purified MHC/direct/radioactivity	0	0	6284	0	0
x-ray crystallography	0	445	0	0	0
NA	0	0	0	0	0

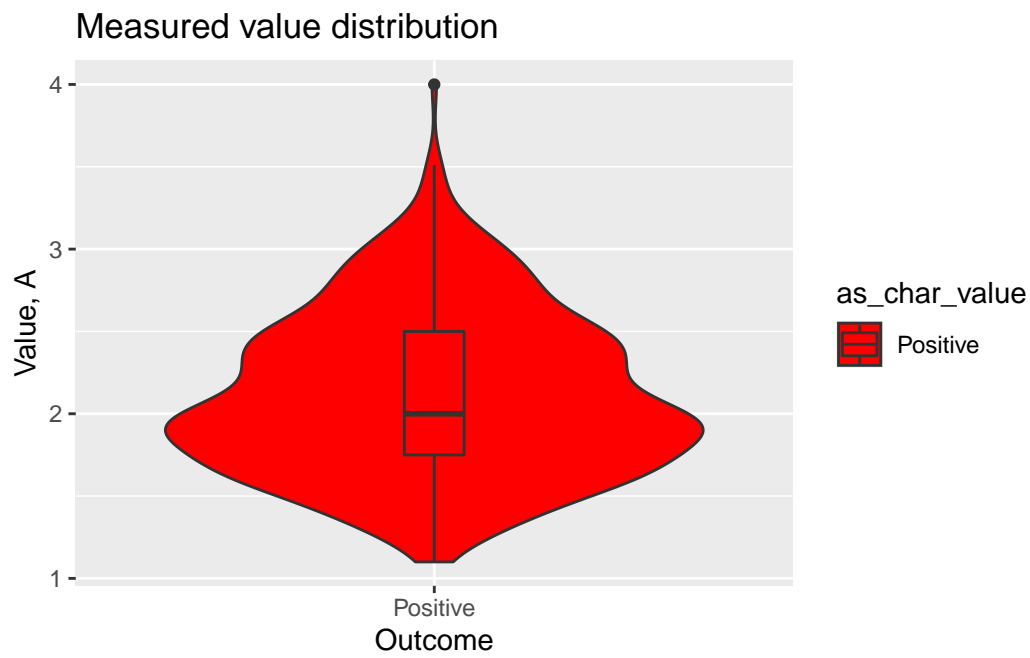
Для нМ



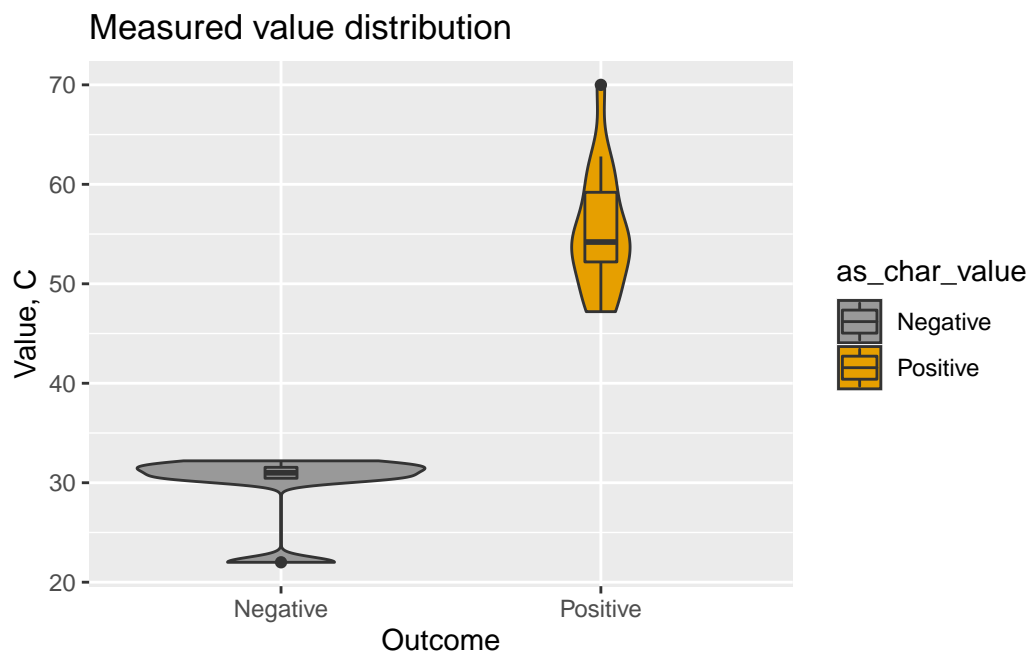
Для минут



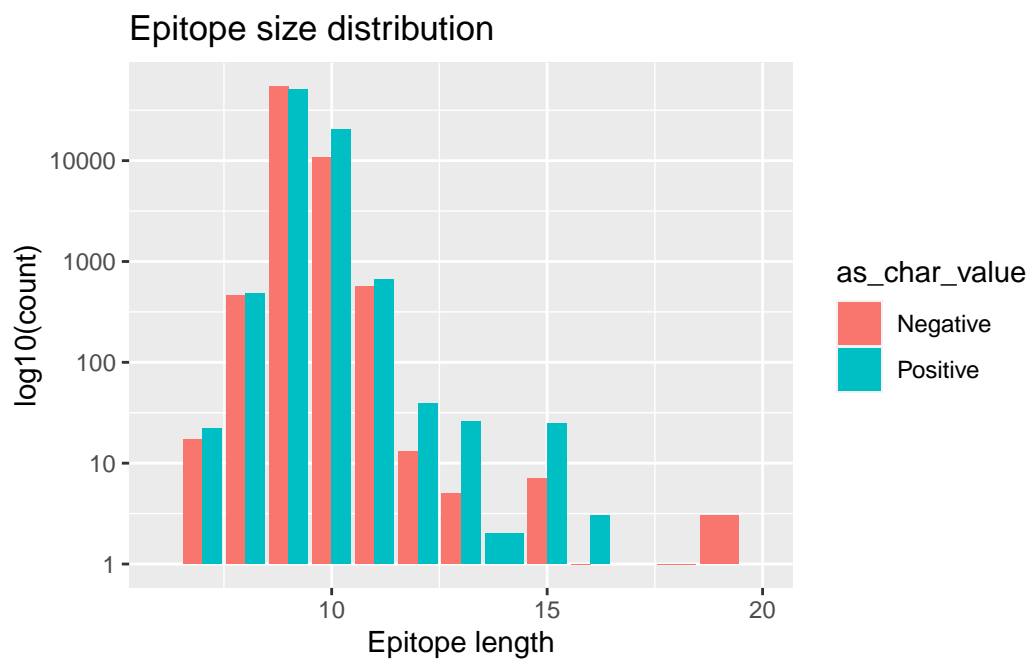
Для ангрём



Для °C

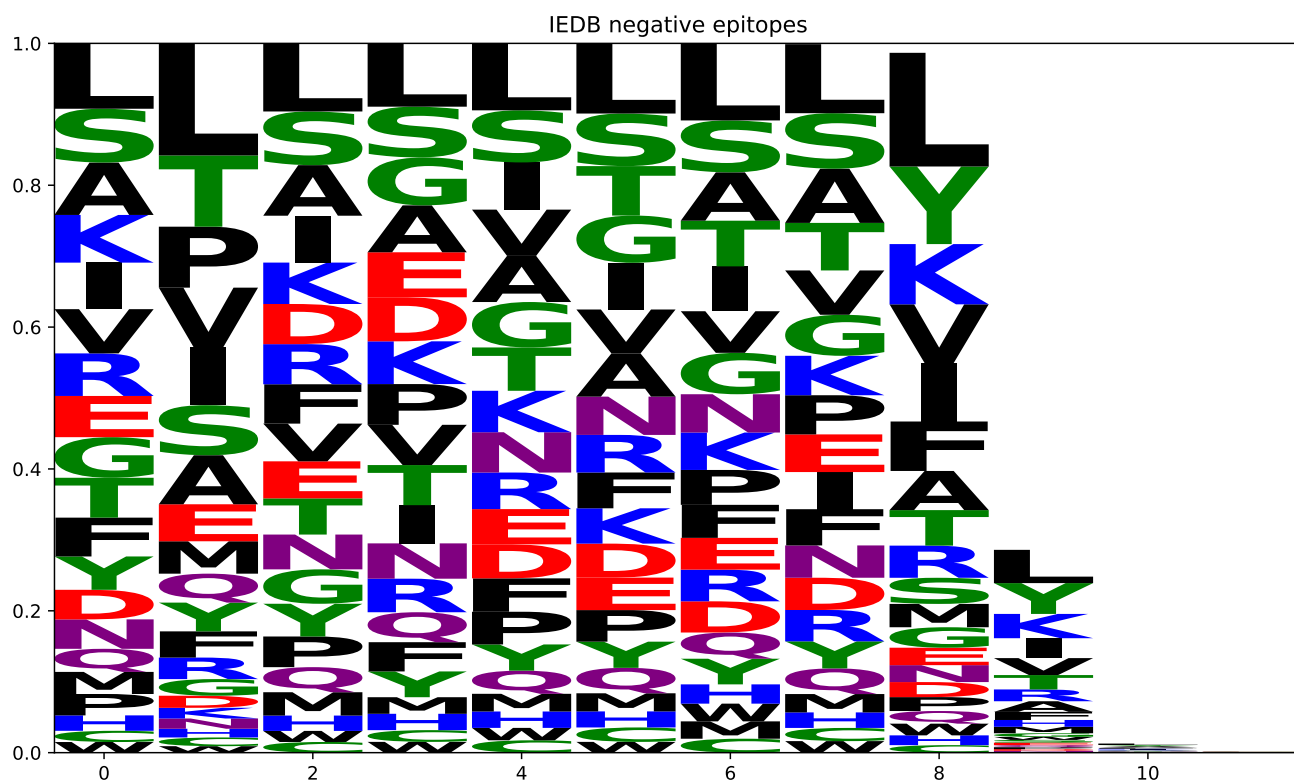
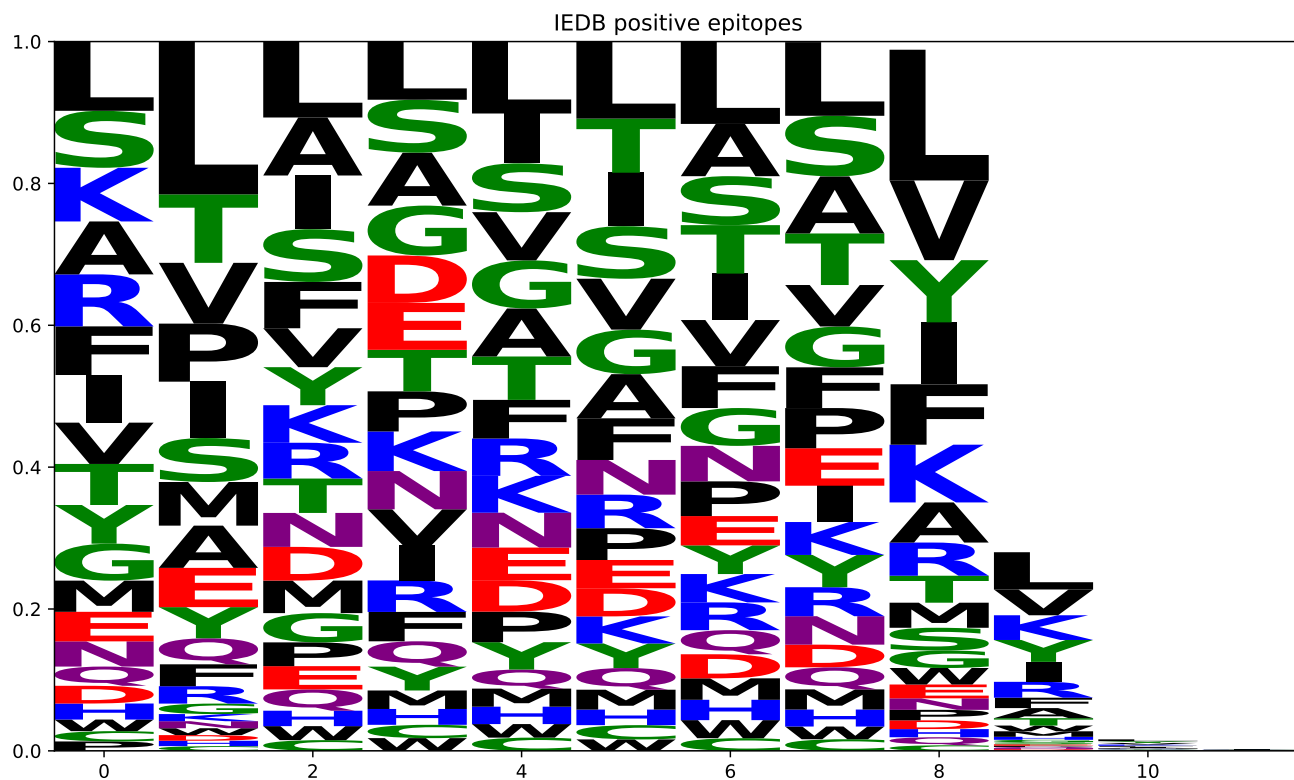


Распределение длины эпитопов и аминокислотного состава



[1] "Положительные случаи, уникальные эпитопы 25743 шт."

[1] "Отрицательные случаи, уникальные эпитопы 14800 шт."



Виды - источники эпитопов

[1] "Всего видов 1207"

Table 14: Топ-10 видов

Вид	Процент
Vaccinia virus WR (Vaccinia virus (strain WR))	18.14
SARS-CoV1	7.27
Homo sapiens (human)	5.55
Vaccinia virus (vaccinia virus VV)	4.38
Mycobacterium tuberculosis	3.43
Zaire ebolavirus (Zaire Ebola virus)	3.42
Giardia lamblia ATCC 50803	2.34
Vaccinia virus Copenhagen (Vaccinia virus (strain Copenhagen))	2.14
Mycobacterium tuberculosis H37Rv (Mycobacterium tuberculosis str. H37Rv)	1.50
Dengue virus 2 Jamaica/1409/1983 (Dengue virus type 2 (strain Jamaica))	1.48

Распределение аллелей HLA

[1] "Всего аллелей 105"

Table 15: Топ-10 аллелей

Аллель	Процент
HLA-A*02:01	16.17
HLA-A*03:01	5.77
HLA-A*11:01	5.59
HLA-B*07:02	4.46
HLA-A*68:02	4.12
HLA-A*24:02	3.96
HLA-A*02:03	3.94
HLA-A*02:06	3.94
HLA-B*15:01	3.82
HLA-A*01:01	3.78

Анализ elution assays

Скрипт *iedb_clean_mhc_elution.R* производит вторичную очистку данных, полученных ранее, по тем же критериям. Однако помимо этого обнаружили эксперименты, где проводился нокаут ERAP1 с помощью миРНК. Эти эксперименты были выделены в отдельный датасет.

[1] "Количество строк 993567"

Эксперименты

Table 16: Распределение исходов экспериментов

Value	Number
Negative	10669
Positive	982898

Table 17: Категория экспериментов

Value	Procent
Naturally Processed	100

Table 18: Виды экспериментов

Value	Procent
cellular MHC/mass spectrometry	76.77
coelution	0.00
Edman degradation	0.02
mass spectrometry	20.13

Value	Procent
secreted MHC/mass spectrometry	3.07
T cell recognition	0.00

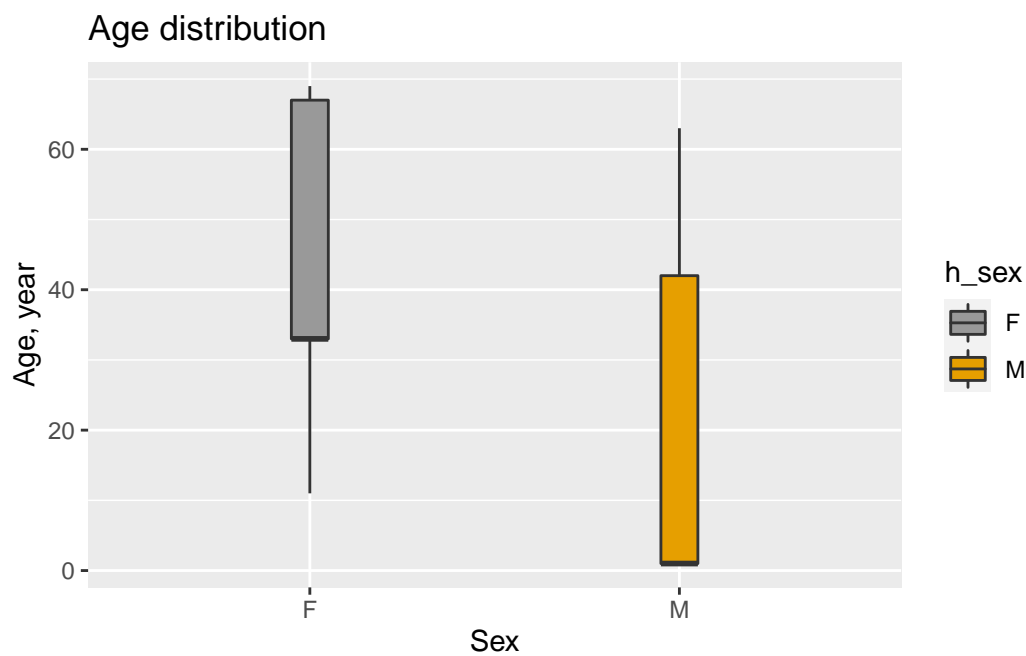
Половозрастная характеристика испытуемых

У этих экспериментов отсутствует какое-либо замерянное значение. Однако есть некоторые данные по полу и возрасту испытуемых.

[1] "Количество записей 170505"

Table 19: Распределение по полу

Var1	Freq
F	70
M	30



Информация по антиген-презентирующим клеткам

Table 20: Типы АПК

Value	Procent
293-T-Epithelial cell	0.02
A549-Epithelial cell	0.58
AML14-Myeloid cell	1.21
B cell	30.78
BT-549-Epithelial cell	0.00
BV-173-Myeloblast	1.10
C1R cells-B cell	17.88
CaSki-Epithelial cell	0.00
Cell found in tissue	15.97
Dendritic cell	0.02
Endothelial cell	0.02
Epithelial cell	0.41
Fibroblast	0.08
Glial cell	0.35
HCC1143-Epithelial cell	0.00
HCC1187-Epithelial cell	0.06
HCC1395-Epithelial cell	0.08
HCC1419-Epithelial cell	0.00
HCC1428-Epithelial cell	0.00
HCC1500-Epithelial cell	0.00
HCC1806-Epithelial cell	0.07
HCC1954-Epithelial cell	0.12
HCT 116-Epithelial cell	0.00
HCT 15-Epithelial cell	0.22
HeLa cells-Epithelial cell	2.42
HROG02-Glial Cell	0.60
HROG17-Glial Cell	0.75
JY cells-B cell	0.73
LM-MEL-33-Melanocyte	1.84
LM-MEL-44-Melanocyte	4.04
Lymph node cells	0.53
Lymphoblast	5.59
Lymphocyte	0.00
Macrophage	0.03

Value	Procent
MCF-7/LY2-Epithelial cell	0.07
MDA-MB-231-Epithelial cell	0.04
MDA-MB-361-Epithelial cell	0.01
MDA-MB-415-Epithelial cell	0.08
Melanocyte	0.05
MM.1S	0.00
Mononuclear cell	0.04
MZ-MEL-5-Melanocyte	0.08
Other	0.87
P815-Mast cell	0.00
Pancreatic beta cell	0.01
PBMC	1.09
Platelet	0.00
RPML-8226-Lymphoblast	0.06
Splenocyte	7.29
SUM159PT-Epithelial cell	0.04
SUP-B15-B cell	0.00
T2 cells-Lymphoblast	0.02
THP-1-Monocyte	2.84
Thymocyte	1.87
U266B1	0.01
UACC-812-Epithelial cell	0.01
Unknown/Unspecified	0.02

Table 21: Ткань

Value	Procent
Adrenal Gland	0.84
Aorta	0.41
Bladder	0.43
Blood	33.96
Bone	0.81
Bone Marrow	1.09
Brain	0.31
Breast	0.68

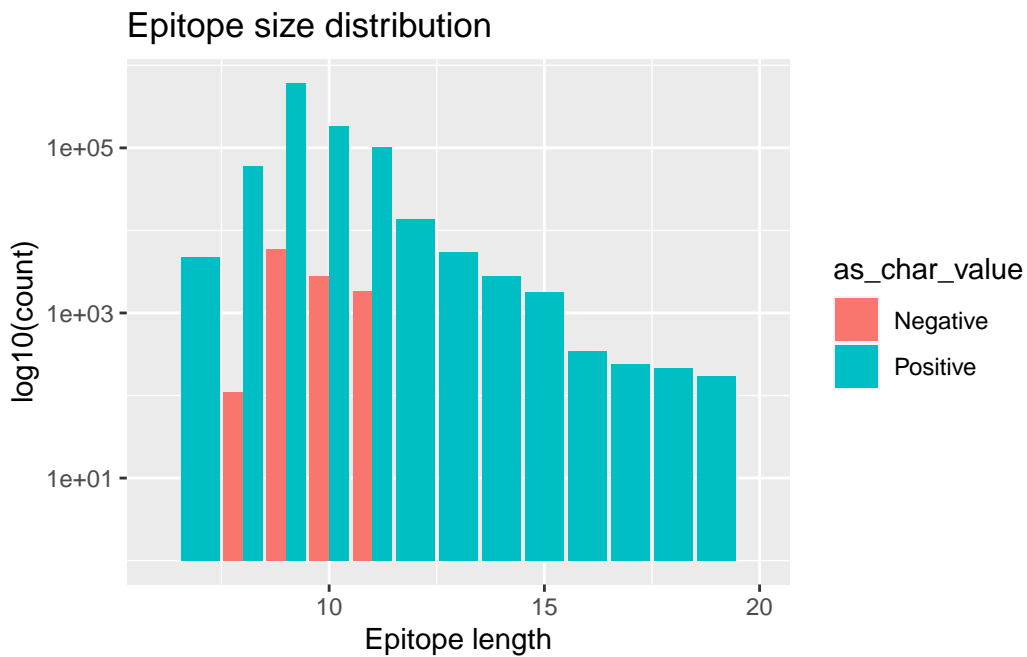
Value	Procent
Central nervous system (CNS)	1.71
Cerebellum	0.44
Colon	0.85
Embryo	0.00
Esophagus	0.68
Gallbladder	0.08
Gastrointestinal Tract	0.01
Heart	0.46
Kidney	1.27
Liver	1.27
Lung	2.94
Lymph Node	6.41
Lymphoid	27.25
Muscle	0.23
Other	0.03
Ovary	1.00
Pancreas	0.17
Prostate	0.11
Skin	0.56
Small Intestine	1.10
Spinal Cord	0.02
Spleen	7.29
Testis	0.51
Thymus	1.87
Thyroid	1.03
Tongue	0.67
Trachea	0.50
Umbilical Cord Blood	0.10
Uterine Cervix	2.54
Uterus	0.36

Table 22: Происхождение клеток

Value	Procent
Blast Activated	0.00

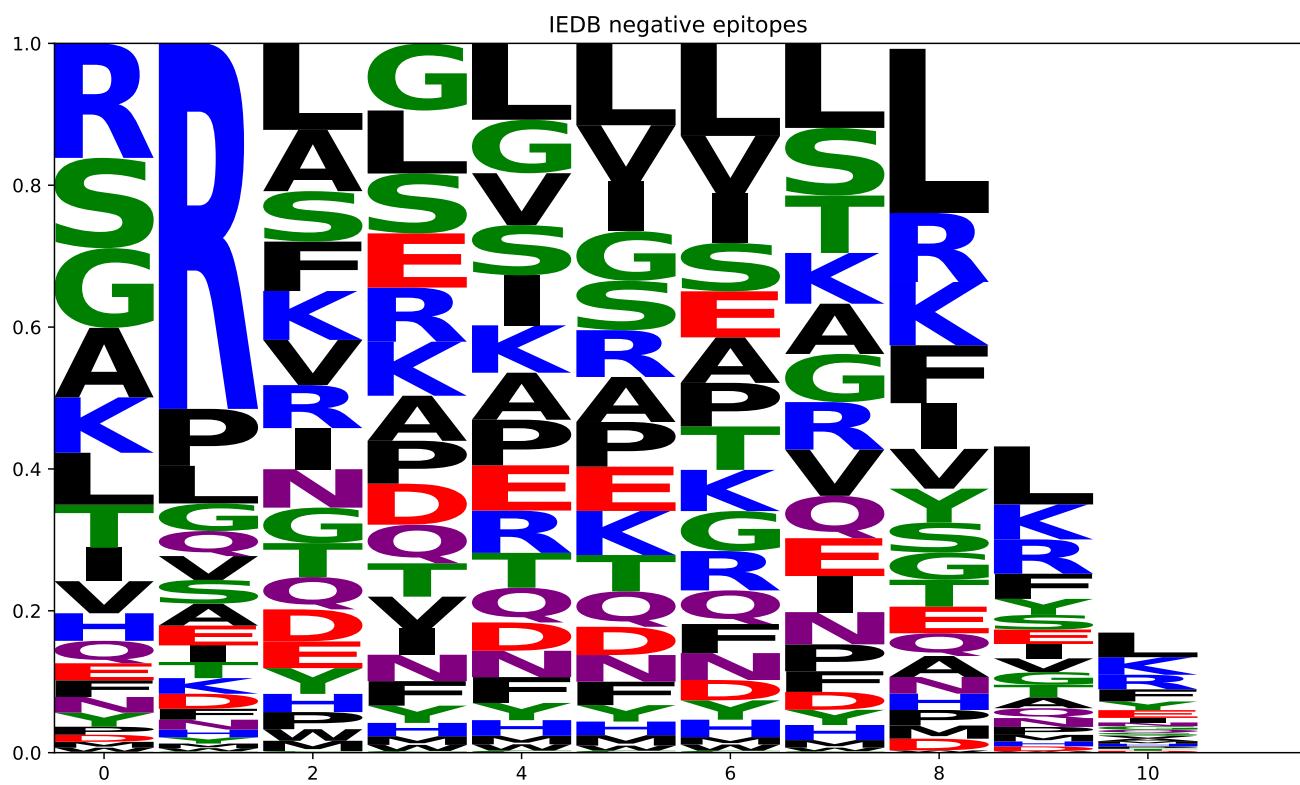
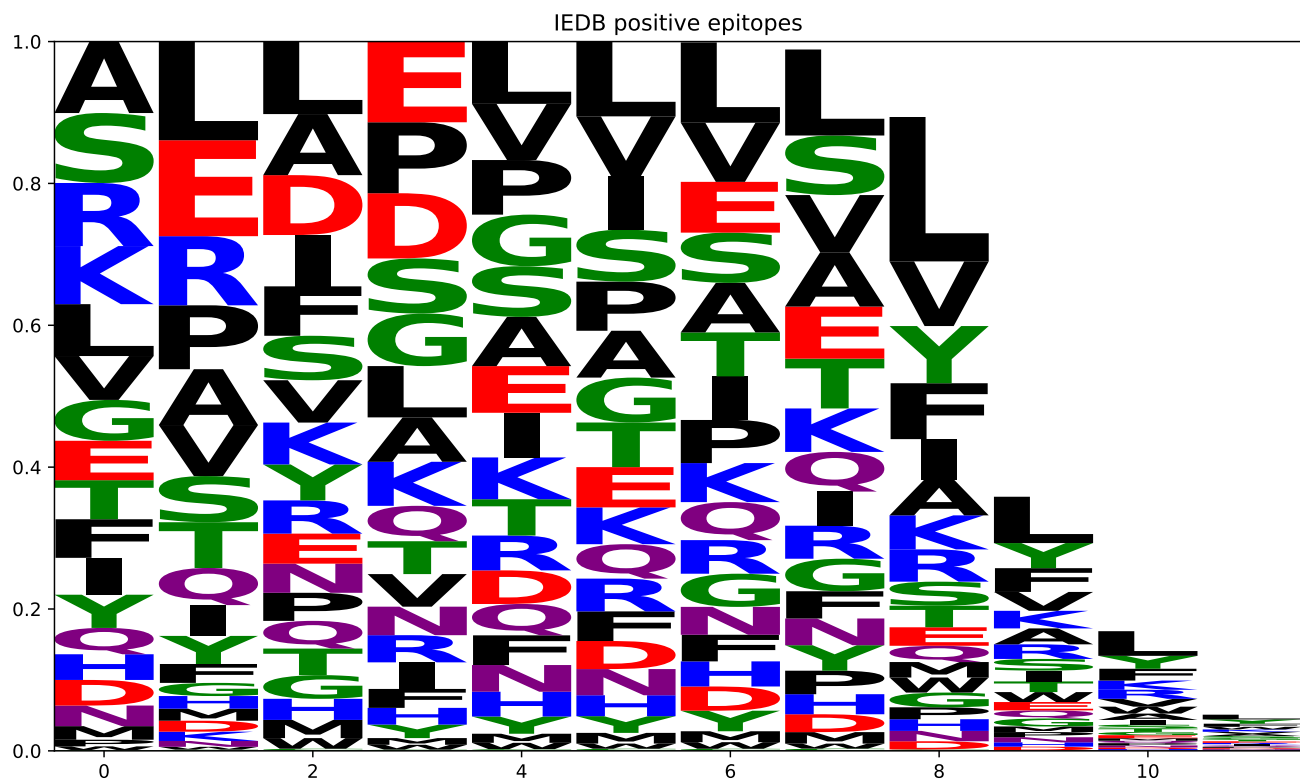
Value	Procent
Cell Line / Clone	45.85
Cell Line / Clone (EBV transformed, B-LCL)	28.20
Direct Ex Vivo	25.95
Short Term Restimulated	0.00

Распределение длины эпитопов и аминокислотного состава



[1] "Положительные случаи, уникальные эпитопы 324092 шт."

[1] "Отрицательные случаи, уникальные эпитопы 9489 шт."



Виды - источники эпитопов

[1] "Всего видов 156"

Table 23: Топ-10 видов

Вид	Процент
Homo sapiens (human)	95.14
Rattus norvegicus (Norway rat)	4.58
Vaccinia virus WR (Vaccinia virus (strain WR))	0.05
Human herpesvirus 5 (Human cytomegalovirus)	0.03
Toxoplasma gondii GT1	0.02
Vaccinia virus Copenhagen (Vaccinia virus (strain Copenhagen))	0.02
Fusobacterium nucleatum subsp. nucleatum ATCC 25586 (Fusobacterium nucleatum subsp. nucleatum str. ATCC 25586)	0.01
Human herpesvirus 5 strain AD169 (Human cytomegalovirus (strain AD169))	0.01
Influenza A virus (A/X-31(H3N2)) (Influenza A virus (A/X31(H3N2)))	0.01
Influenza B virus (B/Malaysia/2506/2004)	0.01

Распределение аллелей HLA

[1] "Всего аллелей 161"

Table 24: Топ-10 аллелей

Аллель	Процент
HLA-A*02:01	13.11
HLA-B*27:05	9.59
HLA-B*07:02	4.20
HLA-A*11:01	3.88
HLA-B*15:01	3.84
HLA-A*01:01	3.63
HLA-B*57:01	3.47
HLA-B*40:02	3.11
HLA-B*44:02	2.92

Аллель	Процент
HLA-A*24:02	2.70

Объединенный датасет

[1] "Количество строк 1132609"

[1] "Количество уникальных эпитопов 354072"

[1] "Количество уникальных HLA аллелей 170"

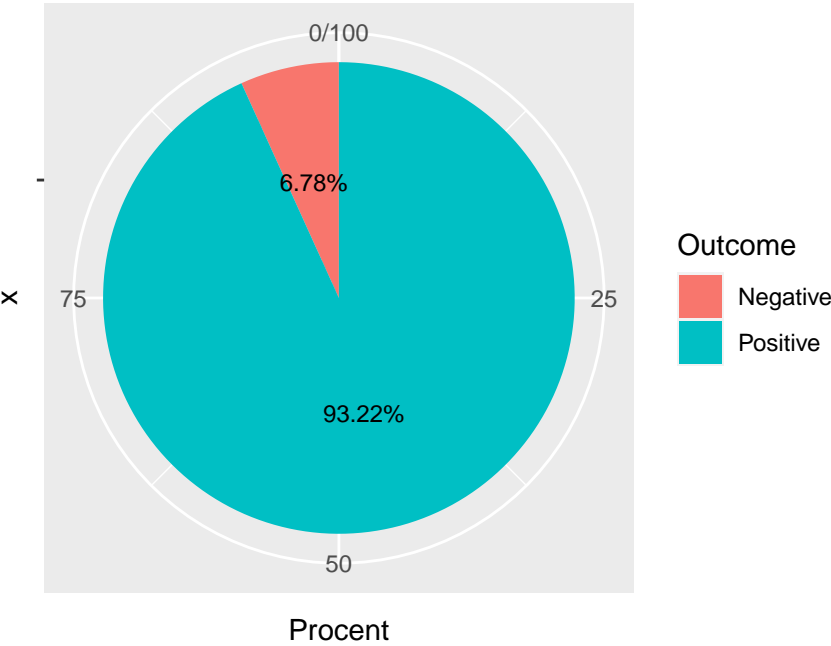


Table 25: Топ-10 представленных аллелей

chain_i_name	n_epi	unique_epi	unique_species
HLA-A*02:01	152716	54901	759
HLA-B*27:05	98187	39287	311
HLA-B*07:02	47915	22812	504
HLA-A*11:01	46326	14110	476
HLA-B*15:01	43414	15856	523

chain_i_name	n_epi	unique_epi	unique_species
HLA-A*01:01	41328	15800	425
HLA-B*57:01	35914	17283	186
HLA-A*24:02	32375	15962	386
HLA-B*40:02	31434	16546	47
HLA-B*44:02	30685	11167	207

Table 26: 10 наименее представленных аллелей

chain_i_name	n_epi	unique_epi	unique_species
HLA-A*03:07	1	1	1
HLA-A*26:02	1	1	1
HLA-A*29:13	1	1	1
HLA-C*08:03	1	1	1
HLA-C*12:15	1	1	1
HLA-A*02:08	2	2	1
HLA-A*02:09	2	2	1
HLA-A*32:02	2	2	1
HLA-B*27:16	2	2	1
HLA-C*15:13	2	2	1

Приложение

Полная статистика

`summarise()` has grouped output by 'chain_i_name'. You can override using the
`.groups` argument.

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-A*01:01	Negative	3674	2896	311
HLA-A*01:01	Positive	37654	12987	237
HLA-A*02:01	Negative	8388	5694	394
HLA-A*02:01	Positive	144328	50212	586
HLA-A*02:02	Negative	1309	747	71
HLA-A*02:02	Positive	6754	5062	135
HLA-A*02:03	Negative	1309	766	75
HLA-A*02:03	Positive	8516	5143	148
HLA-A*02:04	Negative	1	1	1
HLA-A*02:04	Positive	4817	2392	3
HLA-A*02:05	Negative	10	10	7
HLA-A*02:05	Positive	5708	3556	20
HLA-A*02:06	Negative	1711	1017	80
HLA-A*02:06	Positive	5936	4397	134
HLA-A*02:07	Negative	49	49	9
HLA-A*02:07	Positive	8563	4080	9
HLA-A*02:08	Negative	2	2	1
HLA-A*02:09	Positive	2	2	1
HLA-A*02:11	Negative	1	1	1
HLA-A*02:11	Positive	1911	1900	2
HLA-A*02:14	Negative	2	2	1

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-A*02:14	Positive	7	7	2
HLA-A*02:17	Negative	42	42	7
HLA-A*02:17	Positive	514	513	13
HLA-A*03:01	Negative	3850	3037	382
HLA-A*03:01	Positive	22311	8081	384
HLA-A*03:02	Positive	16	16	1
HLA-A*03:07	Positive	1	1	1
HLA-A*11:01	Negative	3368	2648	288
HLA-A*11:01	Positive	42958	11713	322
HLA-A*11:02	Positive	2471	2465	2
HLA-A*11:10	Positive	6	6	1
HLA-A*23:01	Negative	606	562	35
HLA-A*23:01	Positive	5996	3590	74
HLA-A*24:02	Negative	2360	1935	149
HLA-A*24:02	Positive	30015	14539	308
HLA-A*24:03	Negative	409	210	70
HLA-A*24:03	Positive	343	248	99
HLA-A*24:06	Positive	405	405	1
HLA-A*24:07	Positive	1228	1224	1
HLA-A*24:13	Positive	251	250	1
HLA-A*25:01	Negative	30	30	9
HLA-A*25:01	Positive	4309	2494	2
HLA-A*26:01	Negative	1845	1326	157
HLA-A*26:01	Positive	3086	2384	212
HLA-A*26:02	Negative	1	1	1
HLA-A*26:08	Positive	414	411	1
HLA-A*29:02	Negative	609	567	72
HLA-A*29:02	Positive	14772	8237	40
HLA-A*29:13	Positive	1	1	1
HLA-A*30:01	Negative	1085	1085	87
HLA-A*30:01	Positive	4874	3490	69
HLA-A*30:02	Negative	763	724	39
HLA-A*30:02	Positive	3351	3070	45
HLA-A*30:03	Negative	9	9	4
HLA-A*30:03	Positive	14	10	5

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-A*30:04	Positive	5	5	1
HLA-A*31:01	Negative	2915	2519	265
HLA-A*31:01	Positive	5256	3577	127
HLA-A*32:01	Negative	294	294	22
HLA-A*32:01	Positive	7242	4698	28
HLA-A*32:02	Positive	2	2	1
HLA-A*33:01	Negative	1620	1561	50
HLA-A*33:01	Positive	3453	3379	53
HLA-A*33:03	Negative	10	10	5
HLA-A*33:03	Positive	6706	5100	6
HLA-A*34:01	Positive	2076	2072	1
HLA-A*34:02	Positive	3150	3142	2
HLA-A*36:01	Positive	1771	1766	1
HLA-A*66:01	Negative	9	9	3
HLA-A*66:01	Positive	3113	2573	2
HLA-A*66:02	Positive	29	29	1
HLA-A*68:01	Negative	1196	1165	42
HLA-A*68:01	Positive	9771	4837	71
HLA-A*68:02	Negative	2450	1604	174
HLA-A*68:02	Positive	15746	5084	117
HLA-A*69:01	Negative	1846	1433	230
HLA-A*69:01	Positive	3509	1006	100
HLA-A*74:01	Positive	2241	2230	1
HLA-A*80:01	Negative	245	245	76
HLA-A*80:01	Positive	67	67	33
HLA-B*07:02	Negative	3389	2616	297
HLA-B*07:02	Positive	44526	20393	358
HLA-B*07:04	Positive	1337	1334	1
HLA-B*07:06	Positive	16	15	1
HLA-B*08:01	Negative	1753	1094	193
HLA-B*08:01	Positive	13893	6991	126
HLA-B*13:01	Positive	3626	3570	1
HLA-B*13:02	Positive	6860	4097	8
HLA-B*14:02	Negative	44	43	11
HLA-B*14:02	Positive	6857	2079	18

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-B*14:03	Positive	21	21	1
HLA-B*15:01	Negative	2329	1698	242
HLA-B*15:01	Positive	41085	14424	418
HLA-B*15:02	Negative	17	14	10
HLA-B*15:02	Positive	21897	19414	13
HLA-B*15:03	Negative	21	21	2
HLA-B*15:03	Positive	4711	3631	5
HLA-B*15:08	Positive	20	18	1
HLA-B*15:09	Positive	11	8	1
HLA-B*15:10	Positive	4538	3507	4
HLA-B*15:11	Positive	148	138	5
HLA-B*15:13	Positive	6	6	1
HLA-B*15:15	Positive	5	5	2
HLA-B*15:16	Negative	15	15	3
HLA-B*15:16	Positive	27	17	1
HLA-B*15:17	Negative	299	299	57
HLA-B*15:17	Positive	1933	1923	79
HLA-B*15:18	Positive	5	5	1
HLA-B*15:42	Positive	5	5	1
HLA-B*18:01	Negative	926	924	151
HLA-B*18:01	Positive	8459	4861	89
HLA-B*18:03	Positive	217	215	1
HLA-B*27:01	Negative	18	18	7
HLA-B*27:01	Positive	3735	3704	5
HLA-B*27:02	Negative	7	7	5
HLA-B*27:02	Positive	3557	3336	20
HLA-B*27:03	Negative	11	9	6
HLA-B*27:03	Positive	1218	1004	10
HLA-B*27:04	Negative	6	5	4
HLA-B*27:04	Positive	1535	1319	8
HLA-B*27:05	Negative	12829	11131	253
HLA-B*27:05	Positive	85358	36577	150
HLA-B*27:06	Negative	8	7	4
HLA-B*27:06	Positive	1436	1248	8
HLA-B*27:07	Positive	2187	2124	1

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-B*27:08	Positive	2273	2140	1
HLA-B*27:09	Negative	4	3	2
HLA-B*27:09	Positive	9138	6868	12
HLA-B*27:10	Negative	1	1	1
HLA-B*27:10	Positive	12	12	1
HLA-B*27:16	Positive	2	2	1
HLA-B*35:01	Negative	870	862	144
HLA-B*35:01	Positive	11763	4586	300
HLA-B*35:02	Positive	38	38	2
HLA-B*35:03	Negative	8	8	4
HLA-B*35:03	Positive	4780	2350	6
HLA-B*35:04	Positive	12	12	1
HLA-B*35:06	Positive	10	10	1
HLA-B*35:07	Positive	1674	1671	1
HLA-B*35:08	Negative	4	4	1
HLA-B*35:08	Positive	1628	738	7
HLA-B*37:01	Negative	26	26	8
HLA-B*37:01	Positive	3353	1726	6
HLA-B*38:01	Negative	2	2	1
HLA-B*38:01	Positive	2190	2173	8
HLA-B*38:02	Positive	2615	2602	1
HLA-B*39:01	Negative	54	49	28
HLA-B*39:01	Positive	1783	1572	54
HLA-B*39:05	Positive	8	8	1
HLA-B*39:06	Negative	7	7	2
HLA-B*39:06	Positive	48	27	2
HLA-B*39:09	Negative	8	8	1
HLA-B*39:09	Positive	26	20	3
HLA-B*39:10	Positive	5	5	1
HLA-B*39:24	Positive	262	261	1
HLA-B*40:01	Negative	2546	1918	277
HLA-B*40:01	Positive	24402	12628	241
HLA-B*40:02	Negative	155	153	31
HLA-B*40:02	Positive	31279	16395	36
HLA-B*40:06	Positive	2388	2219	1

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-B*40:32	Positive	7	7	6
HLA-B*41:01	Positive	1777	1265	1
HLA-B*41:02	Positive	20	19	1
HLA-B*41:03	Positive	82	81	2
HLA-B*41:04	Positive	53	52	1
HLA-B*41:05	Positive	40	40	1
HLA-B*41:06	Positive	21	21	1
HLA-B*42:01	Negative	1	1	1
HLA-B*42:01	Positive	3119	3097	5
HLA-B*42:02	Positive	5	2	1
HLA-B*44:02	Negative	1073	1069	186
HLA-B*44:02	Positive	29612	10102	66
HLA-B*44:03	Negative	368	367	48
HLA-B*44:03	Positive	16122	6924	48
HLA-B*44:05	Positive	12	5	5
HLA-B*44:08	Positive	49	47	2
HLA-B*44:09	Positive	207	206	1
HLA-B*44:27	Positive	202	202	2
HLA-B*44:28	Positive	129	117	1
HLA-B*45:01	Negative	222	221	21
HLA-B*45:01	Positive	5090	2407	31
HLA-B*46:01	Negative	785	785	151
HLA-B*46:01	Positive	3135	2706	30
HLA-B*47:01	Positive	27	27	1
HLA-B*48:01	Negative	714	714	118
HLA-B*48:01	Positive	77	77	39
HLA-B*49:01	Positive	14369	5411	2
HLA-B*50:01	Positive	3938	1520	2
HLA-B*51:01	Negative	787	775	99
HLA-B*51:01	Positive	14207	8089	70
HLA-B*51:02	Positive	6	6	1
HLA-B*51:08	Positive	625	624	1
HLA-B*52:01	Negative	14	13	2
HLA-B*52:01	Positive	1467	1455	6
HLA-B*53:01	Negative	443	436	56

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-B*53:01	Positive	2602	2577	58
HLA-B*54:01	Negative	302	295	34
HLA-B*54:01	Positive	2557	1601	37
HLA-B*55:01	Negative	1	1	1
HLA-B*55:01	Positive	1405	1396	3
HLA-B*55:02	Negative	19	19	3
HLA-B*55:02	Positive	1363	1352	3
HLA-B*56:01	Positive	2361	1923	2
HLA-B*57:01	Negative	899	898	146
HLA-B*57:01	Positive	35015	16390	94
HLA-B*57:02	Positive	67	66	1
HLA-B*57:03	Positive	7856	5210	2
HLA-B*58:01	Negative	2803	2234	257
HLA-B*58:01	Positive	9628	5174	156
HLA-B*58:02	Positive	909	899	2
HLA-B*73:01	Negative	31	31	19
HLA-B*73:01	Positive	190	188	9
HLA-B*81:01	Negative	8	8	1
HLA-B*81:01	Positive	16	12	2
HLA-C*01:02	Negative	1	1	1
HLA-C*01:02	Positive	3643	2548	8
HLA-C*02:02	Positive	3624	2596	3
HLA-C*03:02	Positive	1191	1187	1
HLA-C*03:03	Negative	31	31	23
HLA-C*03:03	Positive	4689	3592	94
HLA-C*03:04	Negative	23	23	5
HLA-C*03:04	Positive	12179	7260	28
HLA-C*03:13	Positive	9	9	3
HLA-C*04:01	Negative	135	135	82
HLA-C*04:01	Positive	8026	3545	32
HLA-C*04:03	Positive	967	964	1
HLA-C*05:01	Negative	76	76	53
HLA-C*05:01	Positive	16501	10755	66
HLA-C*06:02	Negative	75	75	56
HLA-C*06:02	Positive	4969	3482	81

chain_i_name	as_char_value	n_epi	unique_epi	unique_species
HLA-C*07:01	Negative	418	417	5
HLA-C*07:01	Positive	2355	1885	13
HLA-C*07:02	Negative	26	26	14
HLA-C*07:02	Positive	3095	2224	92
HLA-C*07:04	Positive	1219	1110	1
HLA-C*07:06	Positive	68	68	1
HLA-C*08:01	Negative	24	24	3
HLA-C*08:01	Positive	1606	1592	6
HLA-C*08:02	Negative	49	46	39
HLA-C*08:02	Positive	7405	5242	41
HLA-C*08:03	Positive	1	1	1
HLA-C*12:02	Negative	6	6	2
HLA-C*12:02	Positive	11603	10409	9
HLA-C*12:03	Negative	11	11	11
HLA-C*12:03	Positive	6706	4884	97
HLA-C*12:04	Positive	3	3	1
HLA-C*12:15	Positive	1	1	1
HLA-C*14:02	Negative	15	15	9
HLA-C*14:02	Positive	6307	5150	96
HLA-C*14:03	Positive	2677	2667	1
HLA-C*15:02	Negative	72	72	55
HLA-C*15:02	Positive	5473	4290	68
HLA-C*15:05	Positive	94	93	1
HLA-C*15:13	Positive	2	2	1
HLA-C*16:01	Positive	6805	5072	2
HLA-C*16:02	Positive	10	10	1
HLA-C*17:01	Positive	1620	1356	1