Гены, аннотация вариантов в генах, болезни и фенотипы

Василий Евгеньевич Раменский

Анастасия Александровна Жарикова

НМИЦ Терапии и профилактической медицины

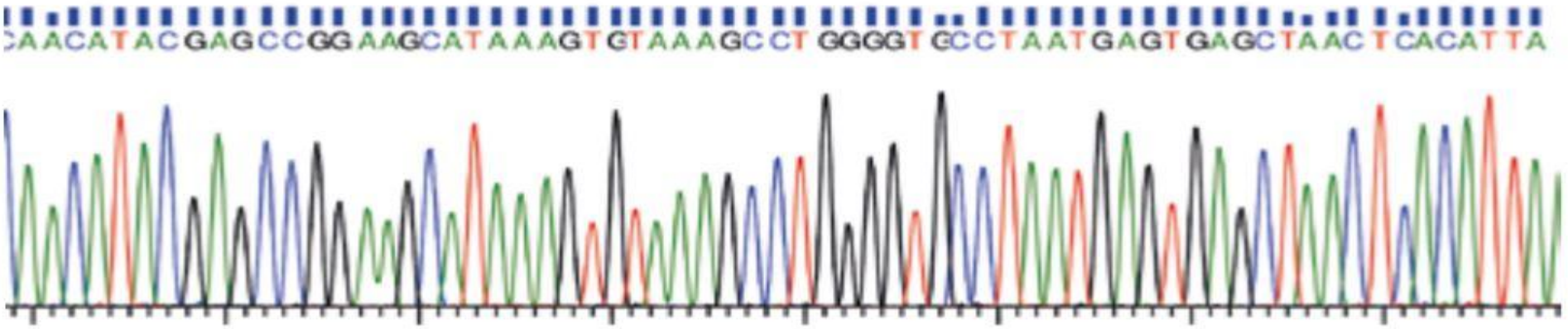
Факультет биоинженерии и биоинформатики МГУ

Институт искусственного интеллекта МГУ

2025

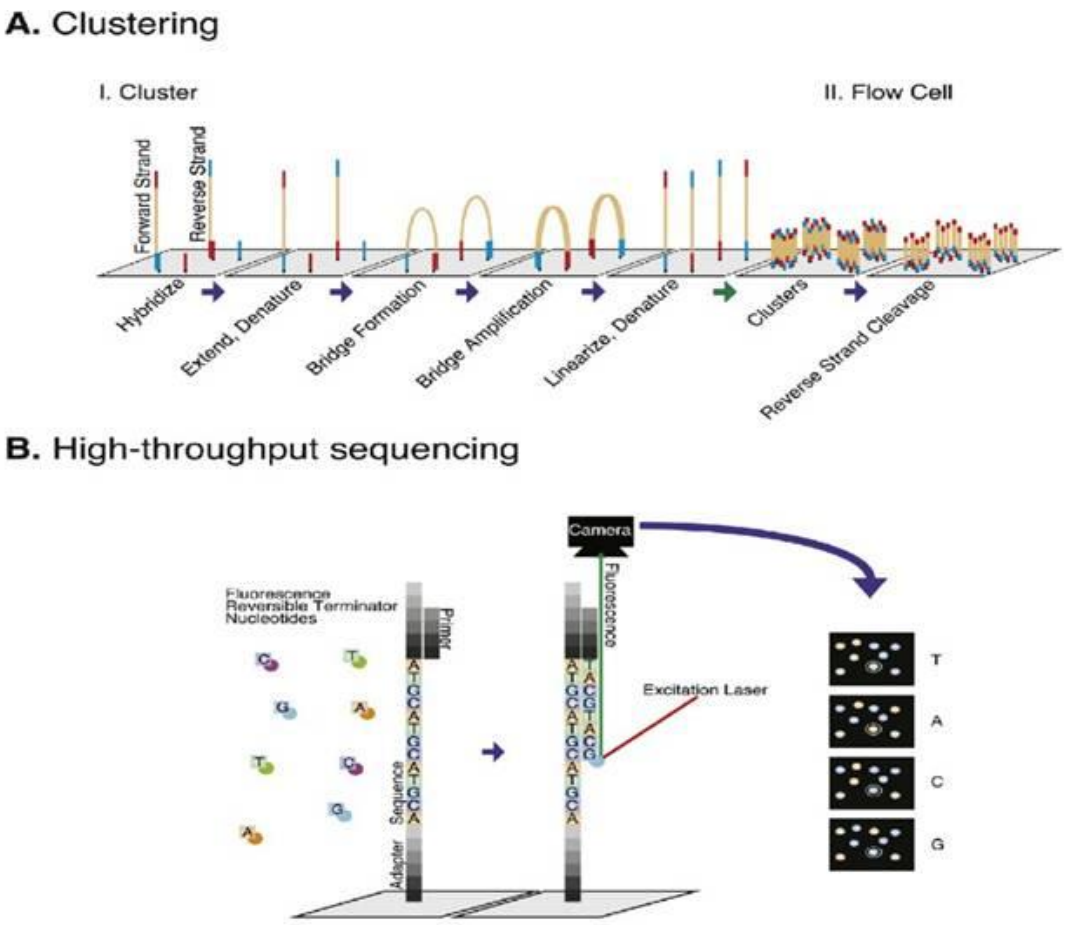
Секвенирование

- Определение последовательности некоторого нерегулярного биологического гетерополимера – белка или нуклеиновой кислоты
- Про белки говорить не будем
- Про РНК тоже не будем



Next-generation sequencing (NGS) - Illumina

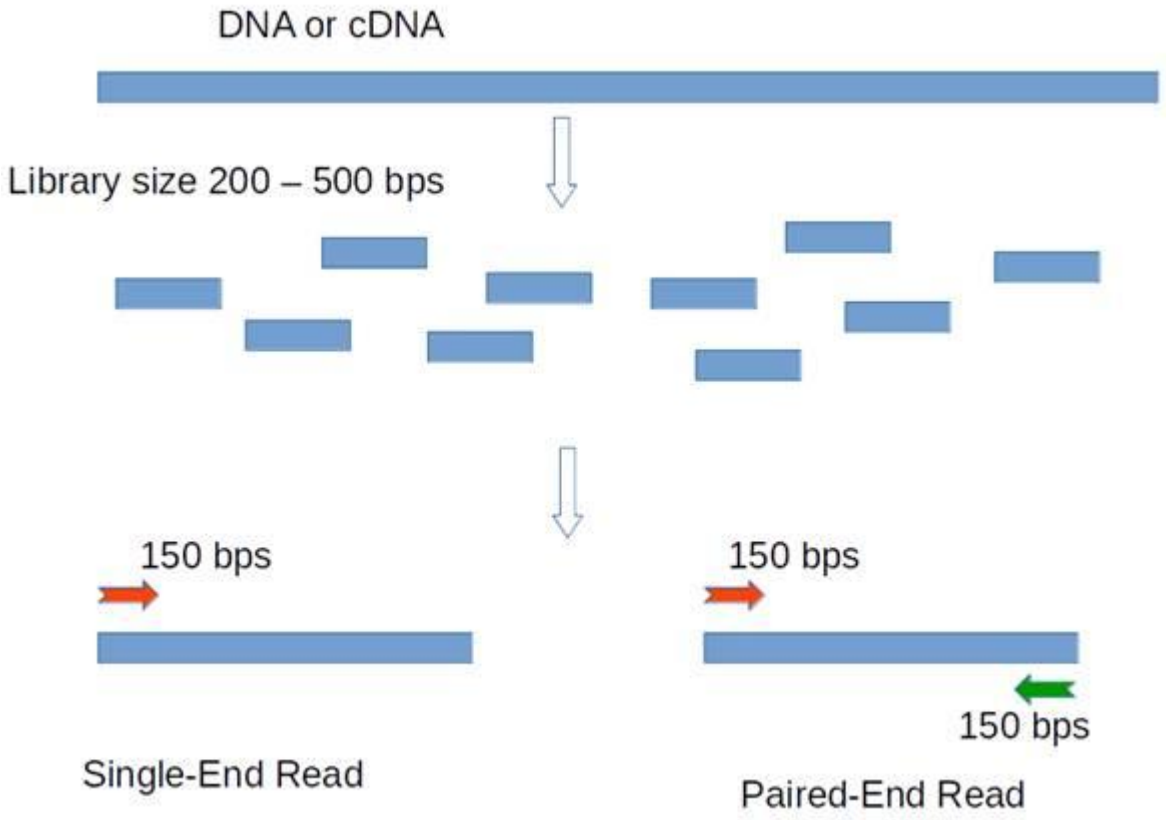
Вспомним, как работает секвенатор компании Illumina



ref

Next-generation sequencing (NGS) - Illumina

Вспомним, что чтения бываю парноконцевыми и одноконцевыми



ref

Секвенирование ДНК

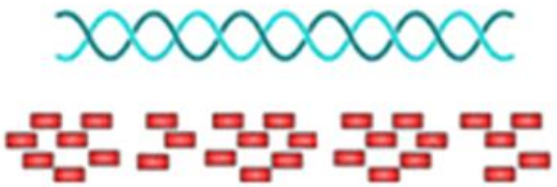
Стратегии секвенирования

- **Полный геном**
- **Экзом** – экзоны белок-кодирующих генов
- **Панели** – набор генов (и\или локусы), варианты в которых интересны при проведении какого-либо исследования или диагностики

- *Вопрос: перечислите плюсы и минусы каждого подхода*

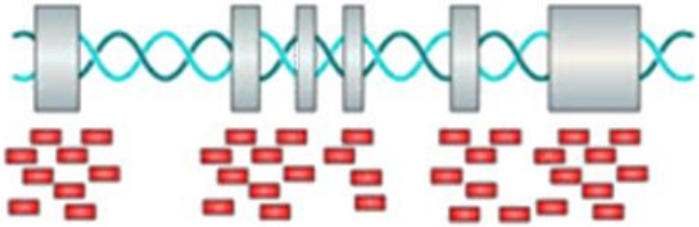
Области секвенирования

Whole genome sequencing



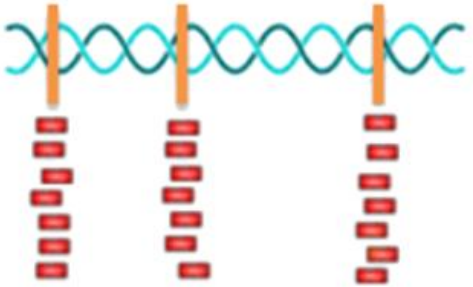
- Sequencing region : whole genome
- Sequencing Depth: >30X
- Covers everything – can identify all kinds of variants including SNPs, INDELs and SV.

Whole exome sequencing



- Sequencing region: whole exome
- Sequencing Depth : >50X ~ 100X
- Identify all kinds of variants including SNPs, INDELs and SV in coding region.
- Cost effective

Targeted sequencing



- Sequencing region: specific regions (could be customized)
- Sequencing Depth : >500X
- Identify all kinds of variants including SNPs, INDELs in specific regions
- Most Cost effective

ref

Какие варианты бывают

- **SNV** - однонуклеотидные варианты, т.е. изменение одного нуклеотида
- **Indels** - короткие вставки и делеции (~ 50 п.н.)
- **CNV** - структурные варианты: инверсии и транслокации
- **Анеуплоидии**: нульсомии, моносомии, трисомии, полисомии
- **Полиплоидизация**

Задача семинара

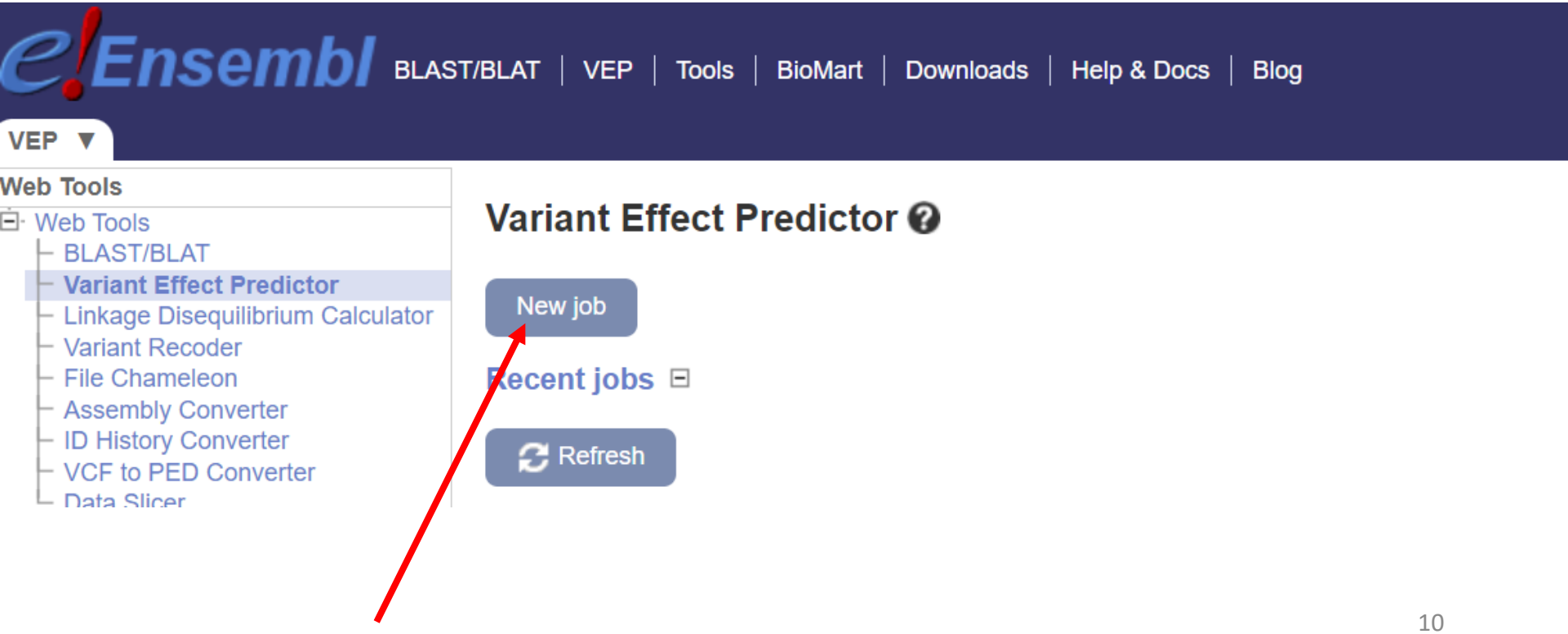
- Проаннотировать набор вариантов в нескольких генах человека с помощью веб-сервиса VEP
- Выстроить систему приоритизации вариантов согласно набору критериев

Дано

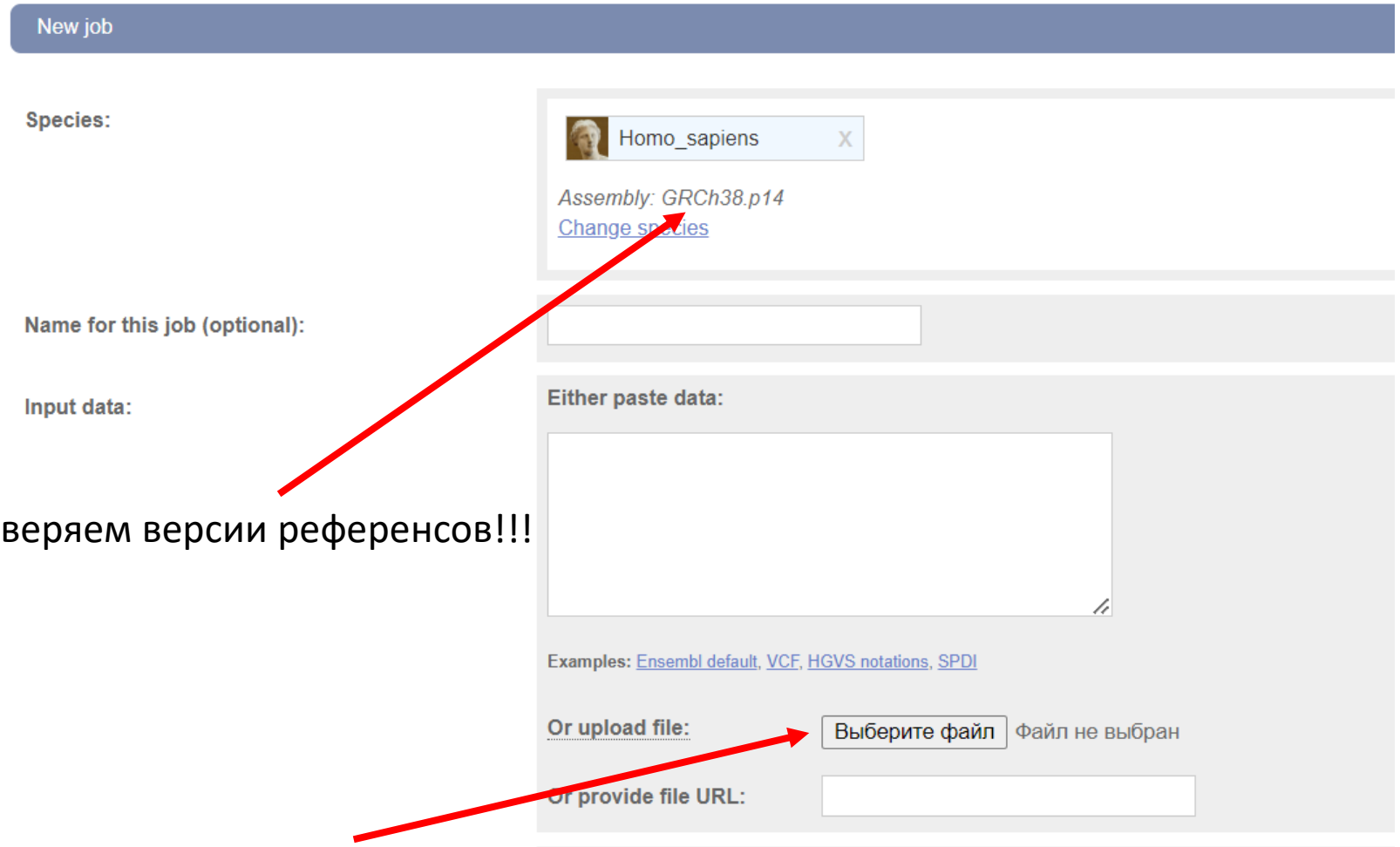
- Набор вариантов в нескольких генах человека
  - Файл с вариантами в формате VCF

Variant Effect Predictor

- VEP
- На вход можно подать vcf файл с вариантами

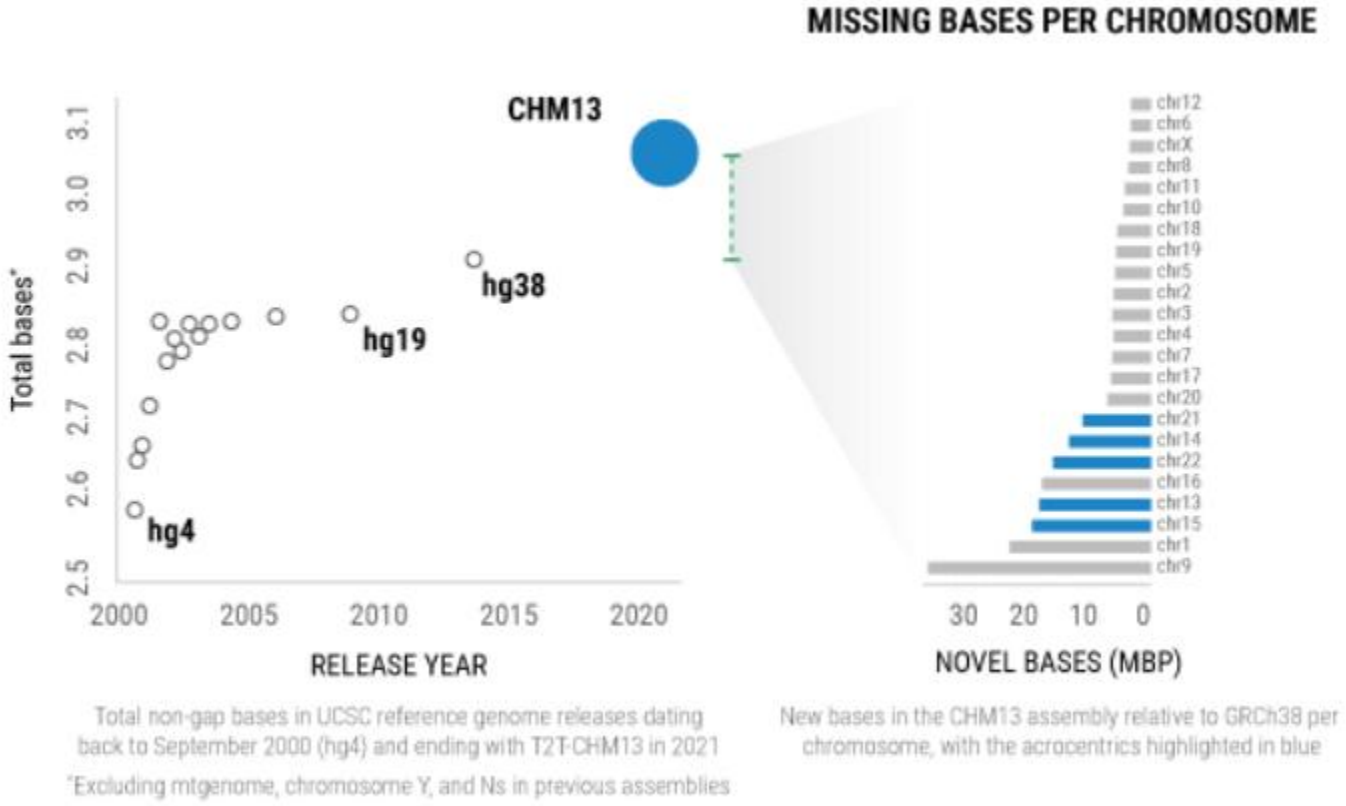


VEP



Проверяем версии референсов!!!

Референсный геном



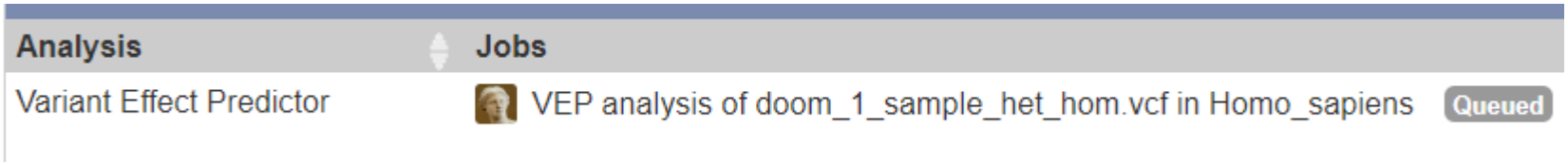
ref

Упражнение

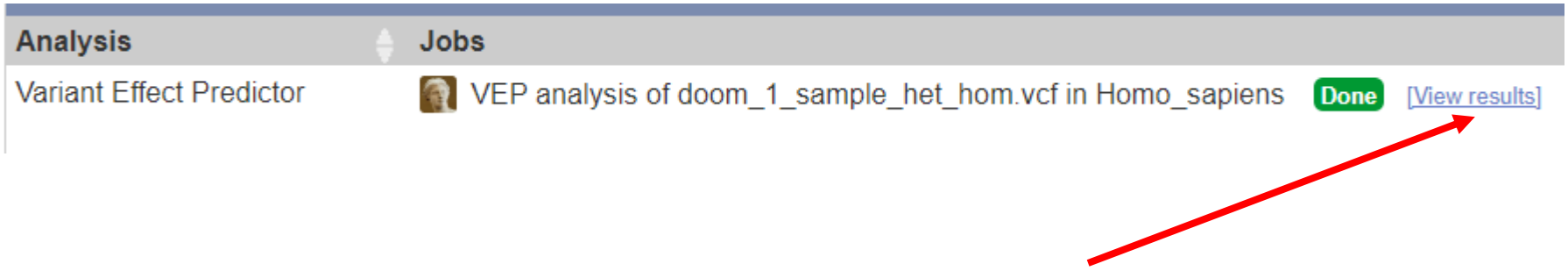
- Загрузите в VEP тестовый файл ***single_sample_het_hom.vcf***
- Ниже есть настройки, добавьте:
  - HGVS
  - UniProt
  - Exon and intron numbers
  - gnomAD exomes
  - gnomAD genomes
  - MANE
  - Identify canonical transcripts

VEP

- При загрузке файла создается новый процесс



- Аннотация занимает какое-то время
- Дождитесь статуса Done



Пока ждем…

… вспомним основные файлы в анализе данных высокопроизводительного секвенирования

Протокол

Все файлы храним и анализируем в архивированном виде!

FASTQ

```
1 @NB501222:13:HY55HBGXY:1:11101:26102:3380 1:N:0:ATGTCA
2 CGTTGGAGAAATAAAATGTGCATAGTGGGGATTTTATTTTAAGTTTGTTGGTTAGGTAGTTGAGGTCTAGGGTTG
3 +
4 AAAAAEEEE6EEEE6EE/EEAEE6/E//EE//EEE//EEE///EEEEAEEEEEEEEA/A//EEE//EAEEA///A
```

Для каждого чтения выделено 4 строки:

1 – идентификатор чтения

2 – нуклеотидная последовательность чтения

3 – строка идентификатора показателя качества (обычно только «+»)

4 – качество каждого нуклеотида

Подробнее

SAM / BAM

```
NB501222:13:HY55HBGXY:1:11101:16088:1242        272     14      49586777        1       75M     *
      0       0       GAAACGGAGCAGGTCAAAACTCCCGTGCTGATCAGTAGTGGGATCGCGCCTGTGAATAGCCACTGCACTCCAGCC
    EEEAEEEEEAEEEEEEEEEAEEAEEEEEEEEAEEEEEEE<EAE/EEEEEEEEEEEEEEEEE6EEEEAEEAAAAA     AS:i:0  ZS:i:0
 XN:i:0  XM:i:0  XO:i:0  XG:i:0  NM:i:0  MD:Z:75 YT:Z:UU XS:A:+  NH:i:2
```

SAM

Col	Field	Type	Regexp/Range	Brief description
1	QNAME	String	`[!-?A-~]{1,254}`	Query template NAME
2	FLAG	Int	$[0, 2^{16} - 1]$	bitwise FLAG
3	RNAME	String	`\*|[:rname:^*=][:rname:]*`	Reference sequence NAME[11]
4	POS	Int	$[0, 2^{31} - 1]$	1-based leftmost mapping POSition
5	MAPQ	Int	$[0, 2^{8} - 1]$	MAPping Quality
6	CIGAR	String	`\*|([0-9]+[MIDNSHPX=])+`	CIGAR string
7	RNEXT	String	`\*|=|[:rname:^*=][:rname:]*`	Reference name of the mate/next read
8	PNEXT	Int	$[0, 2^{31} - 1]$	Position of the mate/next read
9	TLEN	Int	$[-2^{31} + 1, 2^{31} - 1]$	observed Template LENgth
10	SEQ	String	`\*|[A-Za-z=.]+`	segment SEQuence
11	QUAL	String	`[!-~]+`	ASCII of Phred-scaled base QUALity+33

Из особо важного:

- ID чтения
- координаты места картирования чтения на референсный геном
- схема картирования (CIGAR)
- качество картирования
- последовательность в нуклеотидах (аналогично 2ая строка fastq)
- качество нуклеотидов в чтении (аналогично 4ая строка fastq)
- различные флаги (парность чтения, факт картирования, дупликат и пр.)
- различные тэги (количество ошибок, количество мест картирования чтения и др.)

VCF

Спецификация

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4
```

Для работы на семинаре

- В качестве примера будем использовать файл ***single_sample_het_hom.vcf***
- Откройте его любым способом (кроме excel) и рассмотрите из чего этот файл состоит
- Три основные части:
  - Шапка (строки начинаются с ##)
  - Строка (одна) с заголовками столбцов (начинается с #)
  - Информация о вариантах

VCF – столбцы

8 фиксированных колонок (еще)

- CHROM – имя хромосомы
- POS – позиция варианта
- ID – может быть любая информация о варианте, но обычно пустой (.)
- REF – референсная аллель
- ALT – альтернативная аллель
- QUAL – качество варианта (Phred-scaled)
- FILTER – PASS (если ранее была осуществлена маркировка по каким-либо показателям: покрытие, качество и пр)
- INFO – различные характеристики варианта
- FORMAT – список параметров варианта для конкретного образца
- HG00096 – значения параметров, указанных в столбце FORMAT для конкретного образца (в заголовке – ID образца)

VCF – метрики образца

- Колонка FORMAT: GT:AD:DP:GQ:PL
- Колонка ID образца: 0/1:21,4:25:99:1220,108,0

VCF – FORMAT - GT

- Кодирует генотип варианта
- Для диплоидных организмов:
  - 0 – референсный аллель
  - 1 – альтернативный аллель
- Образец по варианту:
  - 0/0 – референсная гомозигота
  - 0/1 – гетерозигота
  - 1/1 – альтернативная гомозигота

VCF – FORMAT – AD и DP

- Отражает покрытие
- AD – количество чтений, которые поддерживают каждую из возможных аллелей; участвуют все чтения, использованные при поиске вариантов
- DP – общее количество чтений, прошедших фильтрацию и поддерживающих каждую из представленных аллелей

VCF – FORMAT – PL и GQ

- Отражает качество генотипа
- PL – нормализованные «вероятности» возможных генотипов (по шкале Phred). Поле содержит 3 числа, что соответствует генотипам 0/0, 0/1, 1/1. PL наиболее вероятного генотипа = 0
- GQ – вычисляется на основании PL, представляет собой разницу «вероятностей» двух наиболее вероятных генотипов (но не более 99). Низкие значения (т.е. << 99) – в генотипе нет уверенности

Упражнение

- Расшифруйте записи

FORMAT	SAMPLE_ID
GT:AD:DP:GQ:PL	0/1:18,15:33:99:393,0,480
GT:AD:DP:GQ:PL	0/1:1,4:6:20:73,0,20
GT:AD:DP:GQ:PL	1/1:0,30:30:89:913,89,0
GT:AD:DP:GQ:PL	0/1:18,17:35:99:660,0,704

Multiple VCF

- В одном VCF файле может быть представлена информация сразу о нескольких образцах
- В конце будут добавлены столбцы на каждый образец
- QUAL – максимальный из возможных

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS     ID        REF    ALT     QUAL FILTER INFO                              FORMAT      NA00001        NA00002        NA00003
20     14370   rs6054257 G      A       29   PASS   NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20     17330   .         T      A       3    q10    NS=3;DP=11;AF=0.017               GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3   0/0:41:3
20     1110696 rs6040355 A      G,T     67   PASS   NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2   2/2:35:4
20     1230237 .         T      .       47   PASS   NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20     1234567 microsat1 GTC    G,GTCT  50   PASS   NS=3;DP=9;AA=G                    GT:GQ:DP    0/1:35:4       0/2:17:2       1/1:40:3
```

VEP

- **Job details** – отображает все настройки и предоставляет команду для аналогичного анализа на вычислительном кластере

Разнообразие данных

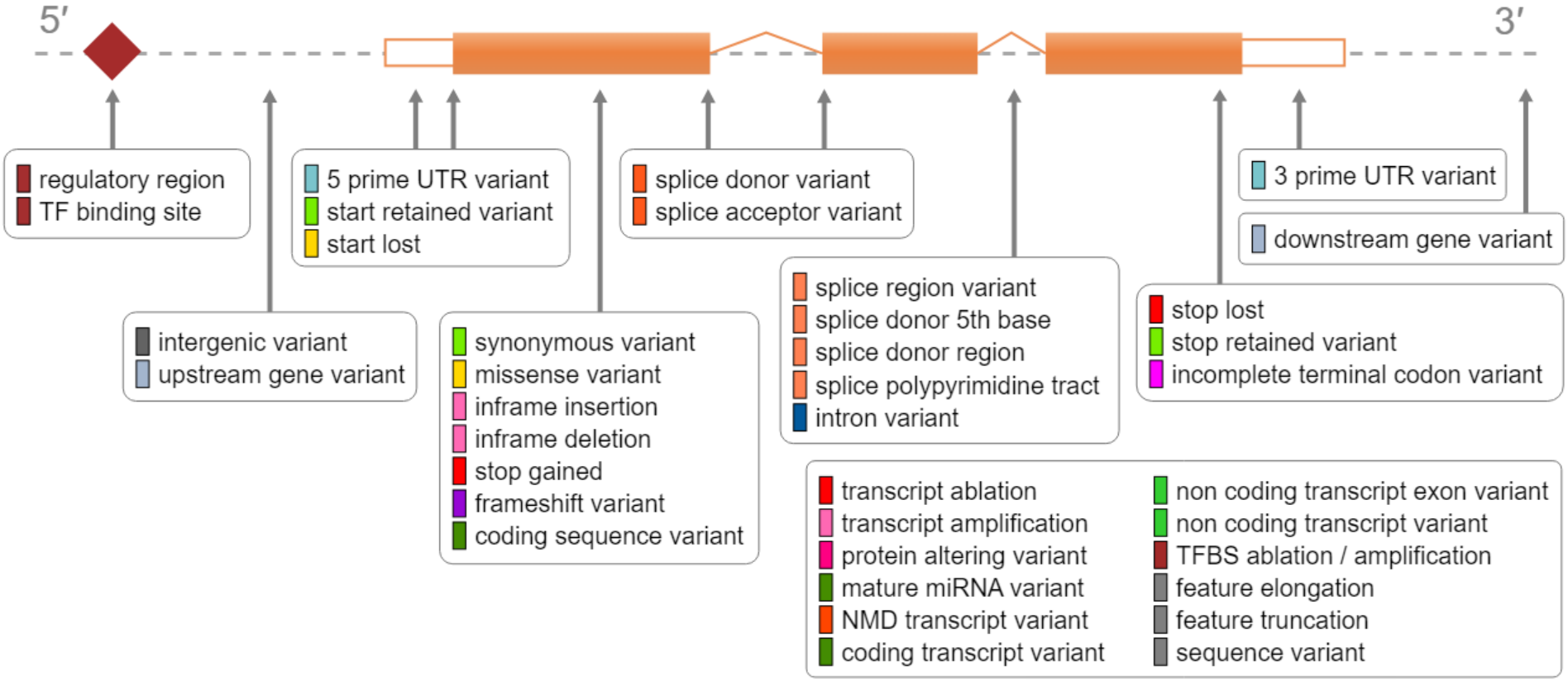
- Загружено 34350 вариантов
- Каждый проаннотирован каким-то образом
- Наша глобальная задача – дать человеку медицинское заключение на основании проведенного генетического исследования
- Нужно ли просматривать 34350 вариантов?

Фильтрация

- Техническая
  - До аннотации можно удалить варианты
    - С низким покрытием
    - С низким качеством
    - …
- Смысловая
  - Это самое интересное
  - Предложите 5 вариантов приоритизации вариантов

VEP

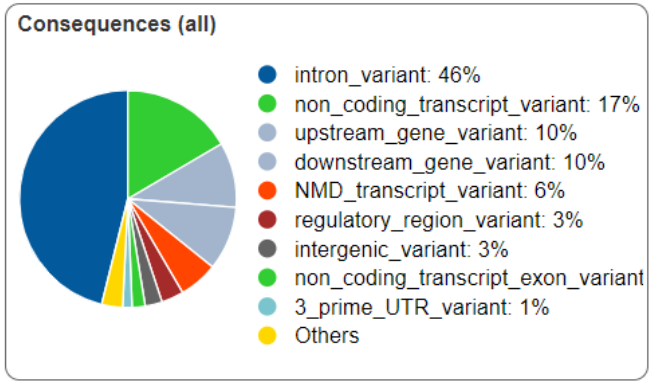
- Consequences

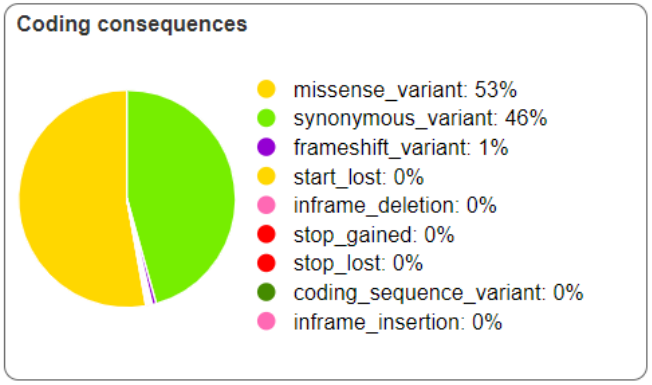


VEP

- Summary statistics

Category	Count
Variants processed	34350
Variants filtered out	0
Novel / existing variants	6 (0.0) / 34344 (100.0)
Overlapped genes	9179
Overlapped transcripts	47336
Overlapped regulatory features	4771





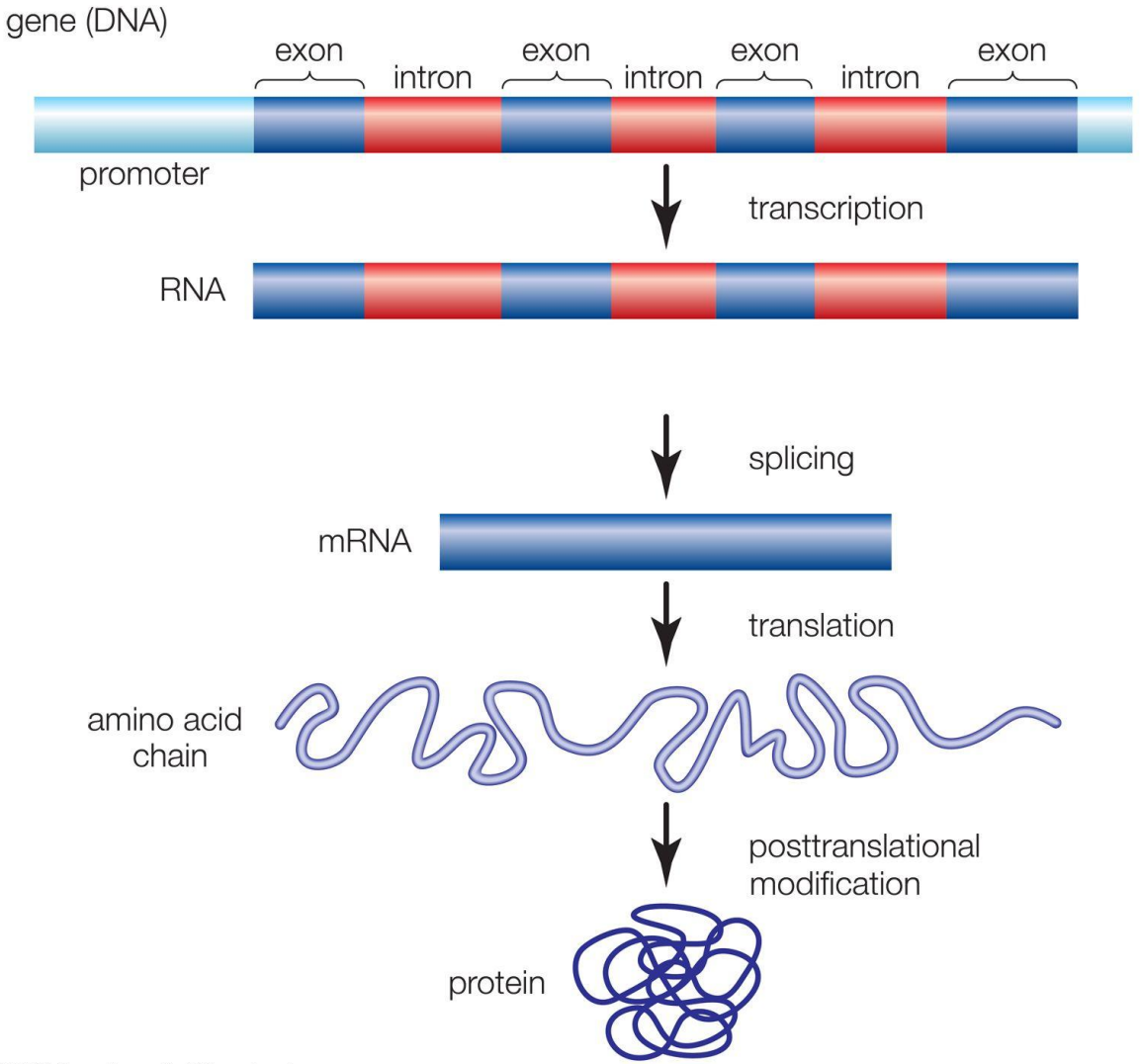
VEP

IMPACT	*Consequence examples*	*Description*
HIGH	splice_acceptor_variant, splice_donor_variant, stop_gained, stop_lost, start_lost	The variant is assumed to have high (disruptive) impact in the protein, probably causing protein truncation, loss of function or triggering nonsense mediated decay
MODERATE	inframe_insertion, inframe_deletion, missense_variant	A non-disruptive variant that might change protein effectiveness
LOW	splice_region_variant, synonymous_variant	A variant that is assumed to be mostly harmless or unlikely to change protein behaviour
MODIFIER	5_prime_UTR_variant, 3_prime_UTR_variant, intron_variant, TFBS_ablation	Usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact

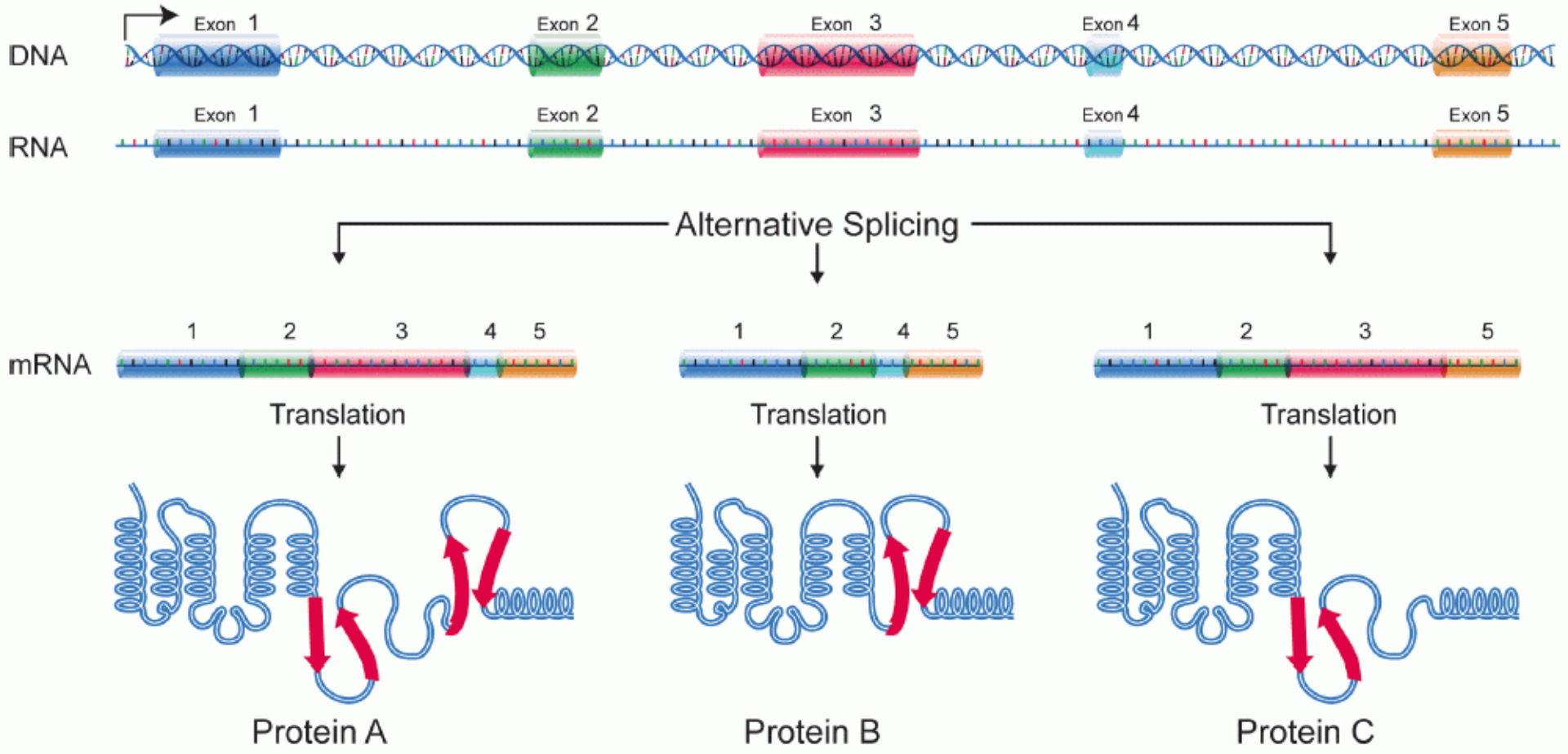
VEP

- HGVC
- Рекомендации по описанию геномных вариантов
- Единая система описания вариантов позволяет присваивать уникальное и однозначное «имя» варианту
  - HGVSc - ENST00000320048.1:c.819T>A
  - HGVSp - ENSP00000321506.1:p.Tyr273Ter

Структура гена



Альтернативный сплайсинг

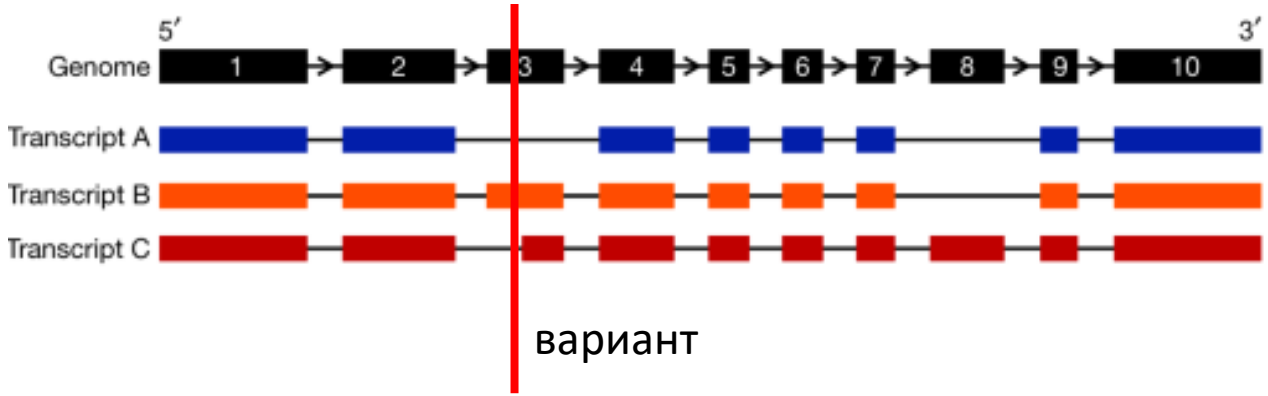


ref

Чем аннотировать варианты?

- Экзон или интрон
- Приводит ли к замене аминокислоты
- Приобретение или потеря STOP-кодона
- Функциональные локусы
  - Сайт сплайсинга
  - Сайт связывания транскрипционного фактора
- …

Экзон или интрон?



Для транскрипта А – интронный вариант
Для транскрипта В – экзонный вариант
Для транскрипта С – сайт сплайсинга

MANE

- Matched Annotation from NCBI and EBI
- Целью аннотации является разрешение проблемы множественных транскриптов
- Для каждого гена представлен один транскрипт, удовлетворяющий ряду условий

Упражнение

- Обсудите результаты, представленные в вкладке Summary statistics
- Есть ли в ваших данных укорачивающие белок варианты? В каких категориях вы будете искать такие варианты?

Упражнение

- Выберите варианты только с высоким импактом
- Сколько их?
- В каких генах они представлены?
- Были ли ранее описаны эти варианты?
- Что указано в колонке MANE?
- Что еще можно сказать об этих вариантах?

Упражнение

- Отберите варианты по частоте представленности в европейской популяции
- gnomADe NFE AF < 0.03
- Сколько таких вариантов?
- Как распределены значения столбцов
  - Consequence
  - Impact

Упражнение

- Повторите предыдущее упражнение, отобрав только частые варианты, представленные в европейской популяции
- Сравните представленность значений Consequence и Impact у частых и редких вариантов

Что из этого название гена?

- A1BG
- alpha-1-B glycoprotein
- ENSG00000121410
- ENST00000263100.8
- NM_130786
- P04217

- Почему так много?!

Номенклатуры

- A1BG - symbol
- alpha-1-B glycoprotein - name
- ENSG00000121410 – Ensembl (gene)
- ENST00000263100.8 – Ensembl (transcript)
- NM_130786 – Refseq
- P04217 – UniProt/Swiss-Prot

HUGO Gene Nomenclature Committee

- Утвержденная номенклатура генов человека

Protein-coding gene	**19392**
Non-coding RNA	**9303**
RNA, cluster	127
RNA, long non-coding	5867
RNA, micro	1970
RNA, misc	29
RNA, ribosomal	60
RNA, small nuclear	58
RNA, small nucleolar	569
RNA, transfer	615
RNA, vault	4
RNA, Y	4
Phenotype	**569**
Pseudogene	**14749**
Immunoglobulin pseudogene	203
Pseudogene	14509
T cell receptor pseudogene	37
Other	**1547**
Complex locus constituent	70
Endogenous retrovirus	117
Fragile site	118
Immunoglobulin gene	230
Readthrough	148
Region	82
T cell receptor gene	206
Transposable element	4
Unknown	564
Virus integration site	8

HGNC

HGNC data for A1BG

Approved symbol	A1BG
Approved name	alpha-1-B glycoprotein
Locus type	gene with protein product
HGNC ID	HGNC:5
Symbol status	Approved
Chromosomal location	19q13.43
Gene groups	Immunoglobulin like domain containing

Gene resources for A1BG

Ensembl ENSG00000121410 Curated
Ensembl region in detail, Ensembl gene sequence

NCBI Gene 1 Curated

UCSC uc002qsd.5

Alliance of Genome Resources HGNC:5

Nucleotide resources for A1BG

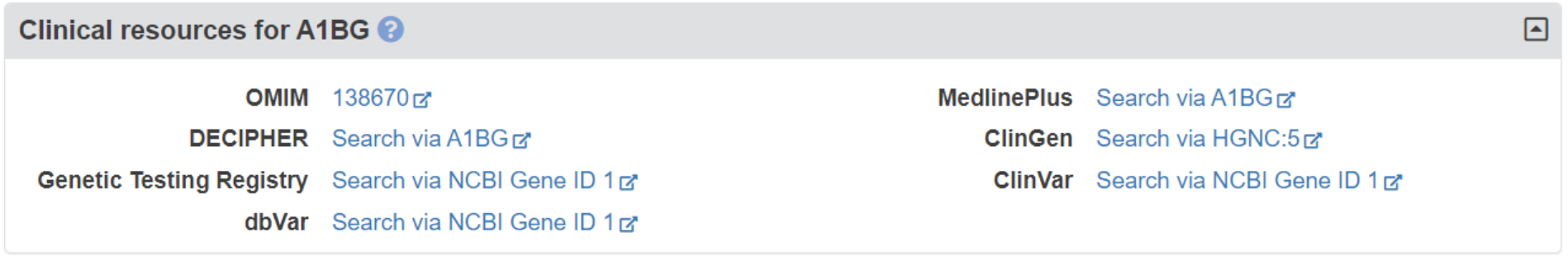
MANE Select NM_130786.4
ENST00000263100.8

RefSeq NM_130786 Curated
NCBI sequence viewer

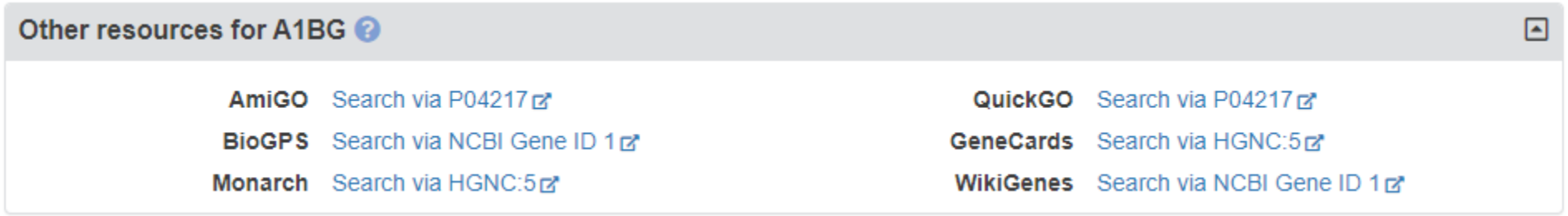
CCDS CCDS12976 Curated

HGNC

- Полезное для медицинской геномики

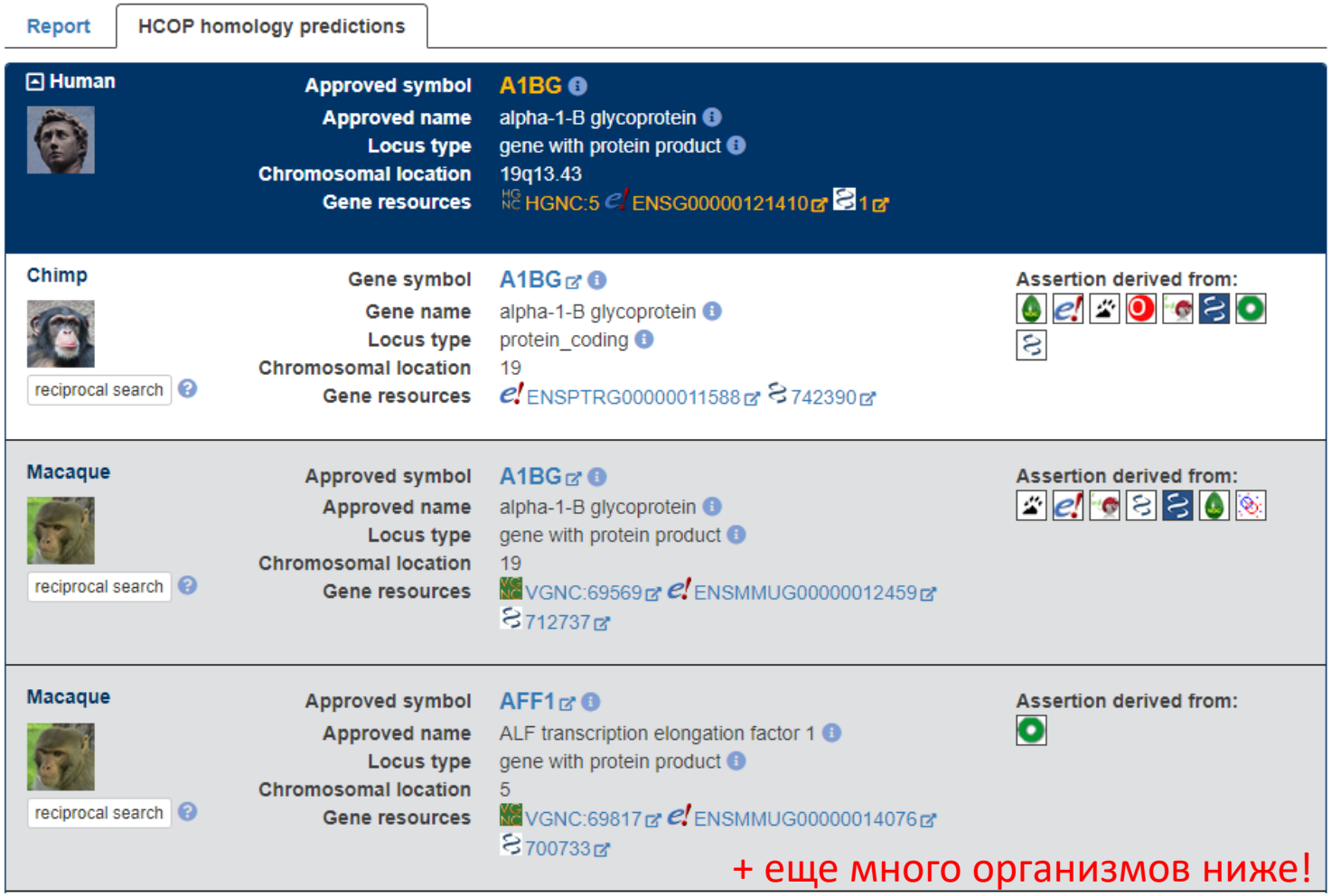


- Справочная информация



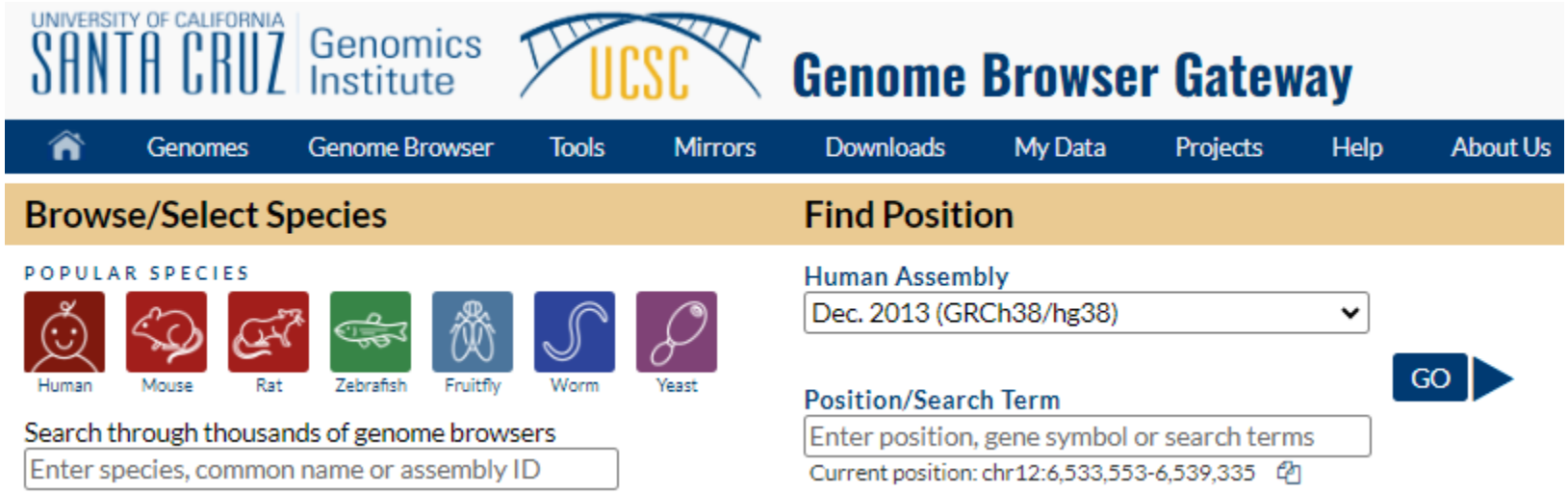
HGNC

- Полезное для филогенетики



Genome Browser

- Геномный браузер

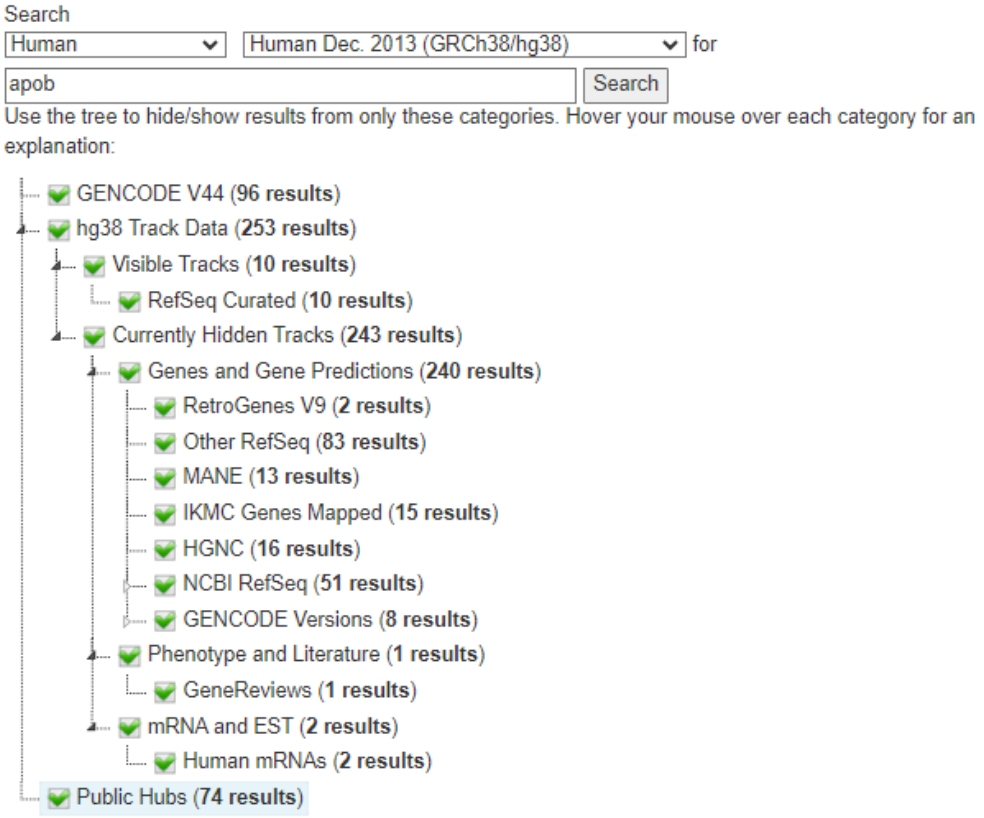


Genome Browser

- Визуализация структуры генов, включая транскрипты, в рамках разных номенклатур
- Большое количество аннотаций локусов:
  - консервативность
  - уровень экспрессии в разных тканях
  - наличие вариантов, представленных в различных клинических базах данных (OMIM, ClinVar, COSMIC и пр.)
  - функциональные участки (сайты связывания, энхансеры и пр.)
  - повторяющиеся элементы
  - многое другое

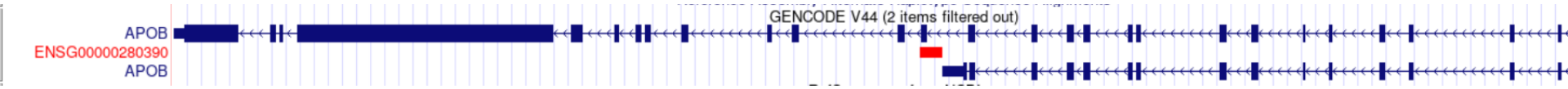
Genome Browser

- Можно подавать на вход ID гена
- Поддерживает множество номенклатур

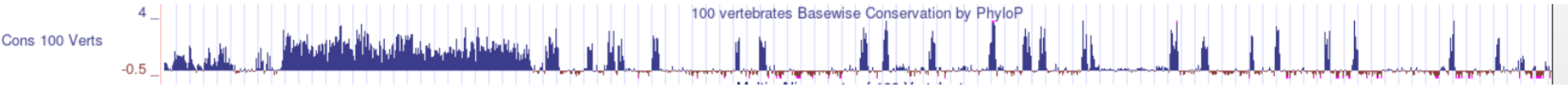


Genome Browser

- В геномном браузере вся информация визуализирована в виде треков
- Разметка генов по версии GENCODE V44; представлено 3 транскрипта



- Трек консервативности; рассчитан на уровне позвоночных; выше значение – более консервативный локус



Genome Browser

- Треки можно выводить в пяти вариациях:
  - Hide
  - Dense
  - Squish
  - Pack
  - Full

  Для изменения типа представления щелкните по треку правой кнопкой мыши и выберите необходимое представление

Упражнение

- Возьмите ID гена (можно из аннотации vcf файла с помощью VEP)
- Найдите этот ген по ID в геномном браузере

Genome Browser

Можно в строке поиска ввести локус в формате

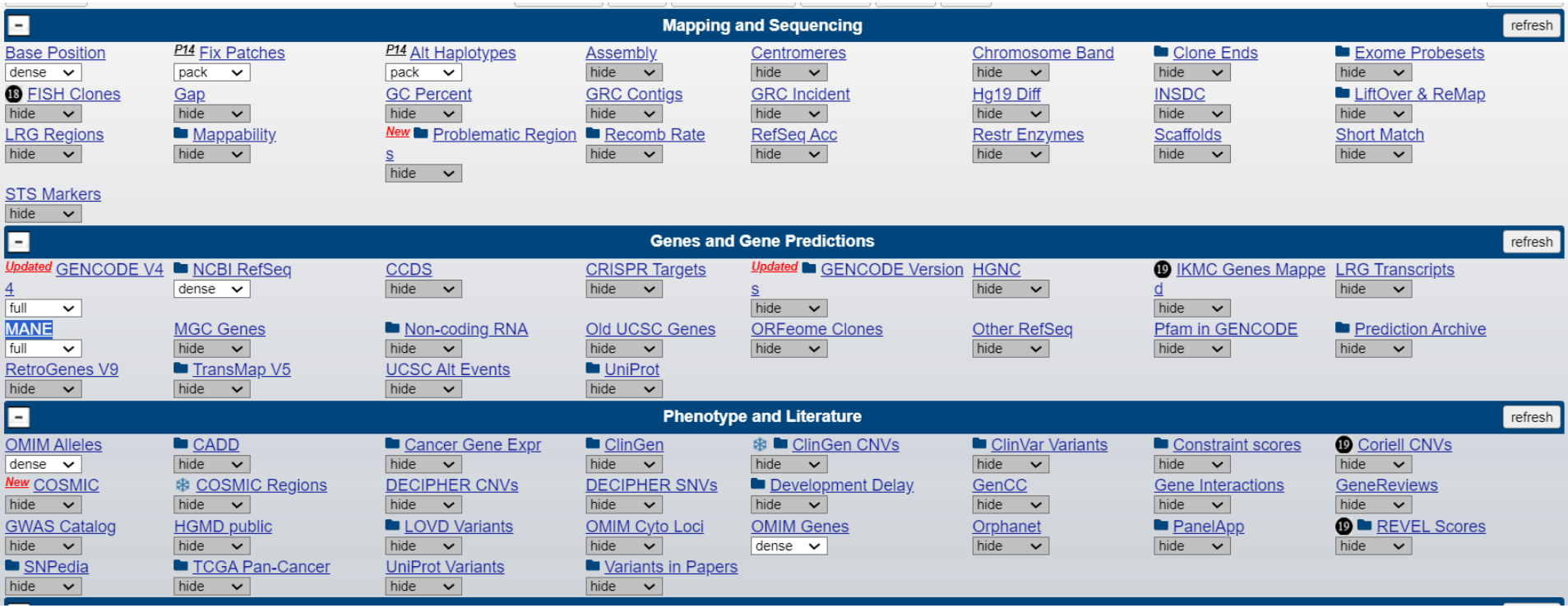
chrN:start-end

Упражнение

- Найдите для своего гена треки:
  - RefSeq Curated
  - OMIM Alleles
  - GTEx RNA-seq
  - Cons 100 Verts (измените тип представления трека)
  - Common dbSNP
  - Repeat Masker
  - CpG Islands

Genome Browser

- Внизу страницы еще есть огромный список спрятанных треков (в представлении hide)



Далее внизу еще много

Genome Browser

- Для отображения нового трека
  - выберите его из списка внизу
  - поменяйте представление трека на необходимое
  - обновите страницу (кнопки refresh)
- Для удаления трека из браузера
  - поменяйте представление трека на hide

GeneCards

- Энциклопедия аннотированных генов человека
- Агрегирует множество информации, баз данных и дополнительных ресурсов
- ~200 источников!!!
- Можно подавать имя гена в любой номенклатуре

Статистика

GeneCards Version 5.18 (Updated: Oct 5, 2023)

Total genes	466,332
HGNC approved	43,718
Disease genes	20,000
Hot genes	500

Category	# of Genes	Example Genes
Protein-coding	21,652	MTOR FGFR2 RET RAF1 MET MAP2K2 MAP2K1
ncRNA genes	291,346	
lncRNAs	130,005	SFTA3 OFCC1 SPATA8 SLC22A18AS HCP5 LINC03040 DLEU1
piRNAs	111,811	piR-52356 piR-30791-073 piR-62069 piR-62060 piR-62024 piR-61955 piR-61945-518
miRNAs	6,903	MIR21 MIR143 MIR140 MIR27A MIR145 MIRLET7D MIRLET7C
rRNAs	1,250	MT-RNR2 MT-RNR1 RNA5S17 RNA5S16 RNA5S15 RNA5S13 RNA5S12
tRNAs	1,158	MT-TL1 MT-TV MT-TT MT-TS1 MT-TF MT-TW MT-TN
snoRNAs	1,904	SNORD89 SNORD3A SNORD118 SNORA73B SNORA64 SNORA62 SCARNA5
SRP_RNAs	9,022	RN7SL2 RN7SL1 RN7SL3 RF00017-7992 RF00017-7752 RF00017-6963 RF00017-6018
circRNAs	120	OP794511 OP794616 OP794610 OP794600 OP794560 OP794534 OP794524
Other ncRNAs	29,173	ADGRF2P TERC ARRDC1-AS1 HCG22 SCARNA7 SCARNA6 RNU4ATAC
Functional elements	128,259	FRAXA HBB-LCR FRAXE H19-ICR LOC111365204 FRA16B FRA11B
Pseudogenes	21,979	BIRC8 SLC26A10P GUCY1B2 GNRHR2 ZNF781 TRIM16L OR10J3
Genetic loci	1,287	ERVE-1 ST2 VIS1 IGKDEL IFNR ERDA1 AZF1
Gene clusters	10	PCDHG@ PCDHB@ IGLV@ IGKV@ HOXD@ HOXA@ HOXB@
Uncategorized	1,799	C20orf181 UGT1A ERVK9-11 ERVH-1 KHDRBS2-OT1 ERVK-28 CCDST

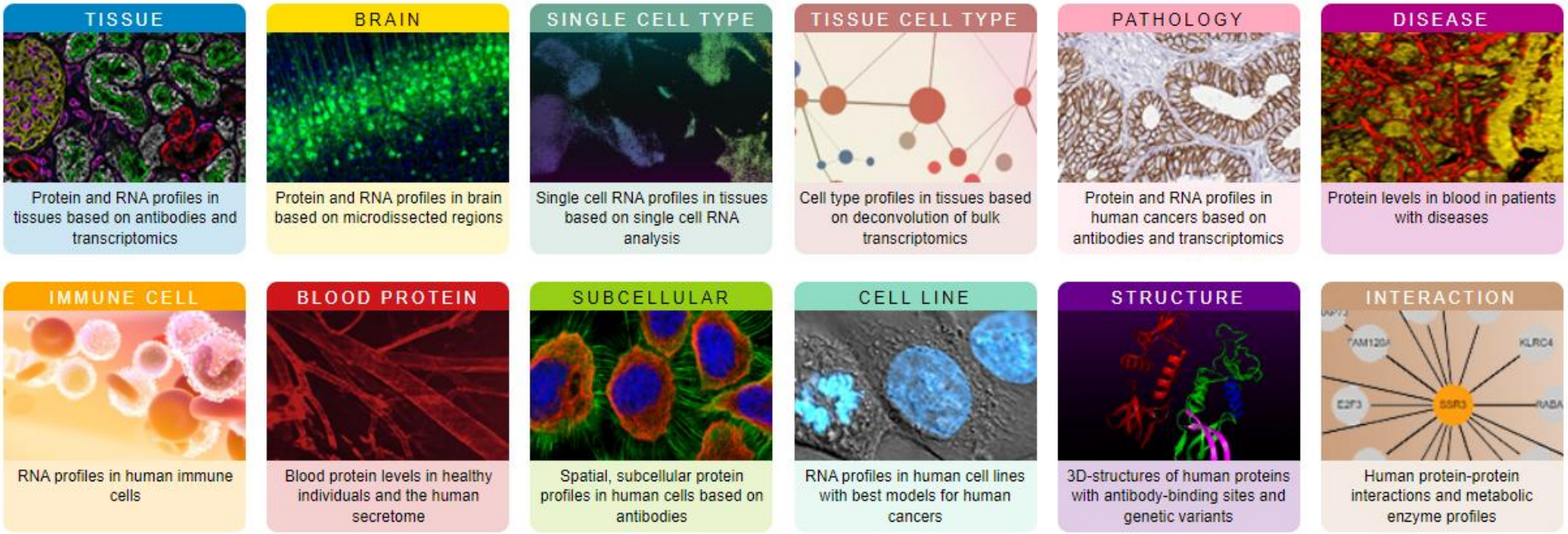
Разделы

GeneCards Sections

- Aliases
- Summaries
- Genomics
- GeneHancer Regulatory Elements
- Proteins
- Domains
- Function
- Localization
- Pathways & Interactions
- Drugs & Chemical Compounds
- Transcripts
- Expression
- Orthologs
- Paralogs
- Variants
- Disorders / Diseases
- Publications
- Products

The human protein atlas

- На вход: ID гена или белка
- 12 секций:



The human protein atlas

Gene

- [x] Gene
- [ ] Gene synonym
- [ ] Ensembl gene id
- [x] Gene description
- [ ] Uniprot accession
- [ ] Chromosome
- [ ] Chromosome position
- [ ] Protein class
- [ ] Biological process
- [ ] Molecular function
- [ ] Disease involvement

Evidence

- [x] Evidence (summary)
- [ ] HPA evidence
- [ ] UniProt evidence
- [ ] NeXtProt evidence

Atlas

- [x] Tissue
- [x] Brain
- [x] Single cell type
- [x] Tissue cell type
- [x] Pathology
- [x] Disease
- [x] Immune cell
- [x] Blood
- [x] subcellular
- [x] Cell line
- [x] Structure
- [x] Interaction

RNA category human

- [ ] RNA tissue specificity
- [ ] RNA tissue distribution
- [ ] RNA tissue specificity score
- [ ] RNA tissue specific nTPM
- [ ] RNA tissue nTPM max in non-specific
- [ ] RNA single cell type specificity
- [ ] RNA single cell type distribution
- [ ] RNA single cell type specificity score
- [ ] RNA single cell type specific nTPM
- [ ] RNA cancer specificity
- [ ] RNA cancer distribution
- [ ] RNA cancer specificity score
- [ ] RNA cancer specific FPKM
- [ ] RNA brain regional specificity
- [ ] RNA brain regional distribution
- [ ] RNA brain regional specificity score
- [ ] RNA brain regional specific nTPM
- [ ] RNA blood cell specificity
- [ ] RNA blood cell distribution
- [ ] RNA blood cell specificity score
- [ ] RNA blood cell specific nTPM
- [ ] RNA blood lineage specificity
- [ ] RNA blood lineage distribution
- [ ] RNA blood lineage specificity score
- [ ] RNA blood lineage specific nTPM
- [ ] RNA cell line specificity
- [ ] RNA cell line distribution
- [ ] RNA cell line specificity score
- [ ] RNA cell line specific nTPM
- [ ] RNA tissue cell type enrichment

RNA category pig/mouse

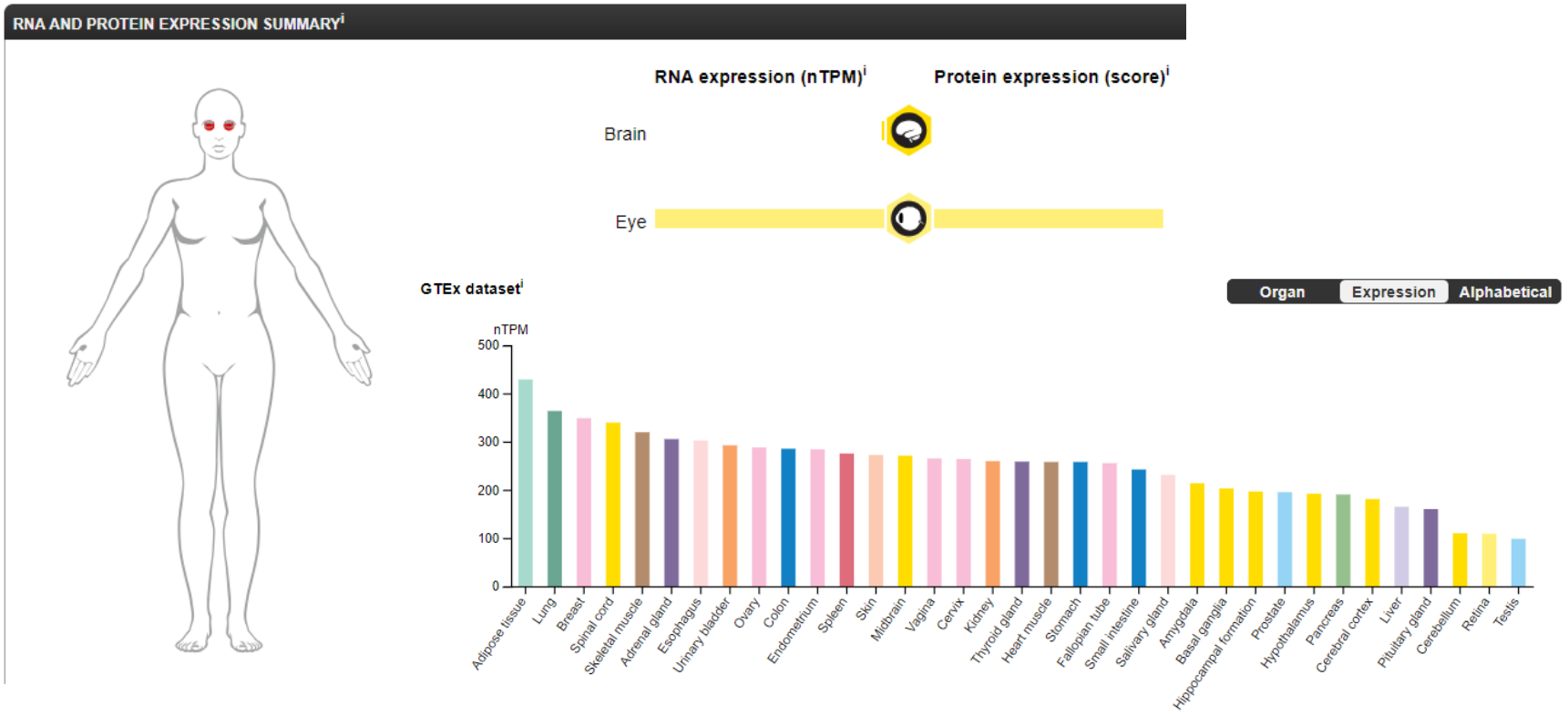
- [ ] RNA mouse brain regional specificity
- [ ] RNA mouse brain regional distribution
- [ ] RNA mouse brain regional specificity score
- [ ] RNA mouse brain regional specific nTPM
- [ ] RNA pig brain regional specificity
- [ ] RNA pig brain regional distribution
- [ ] RNA pig brain regional specificity score
- [ ] RNA pig brain regional specific nTPM

Annotation

- [ ] Antibody ID
- [ ] Reliability (IH)
- [ ] Reliability (Mouse Brain)
- [ ] Reliability (IF)
- [ ] IH abundance (Normal Tissue)
- [ ] Subcellular location
- [ ] Secretome location
- [ ] Secretome function
- [ ] Cell Cycle Dependent Protein
- [ ] Cell Cycle Dependent Transcript
- [ ] Cancer prognostic p-value
- [ ] Blood expression cluster
- [ ] Tissue expression cluster
- [ ] Brain expression cluster
- [ ] Cell line expression cluster
- [ ] Single cell expression cluster
- [ ] Num protein interactions

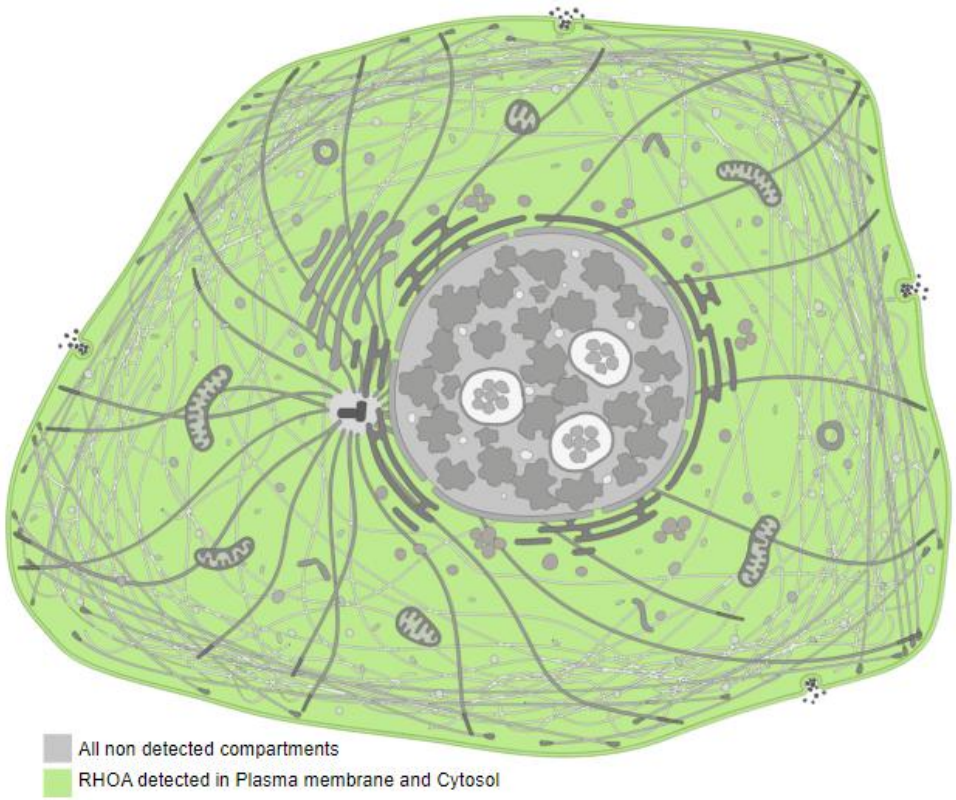
The human protein atlas

- Детекция мРНК и соответствующего белка в различных тканях, типах клеток и клеточных линиях



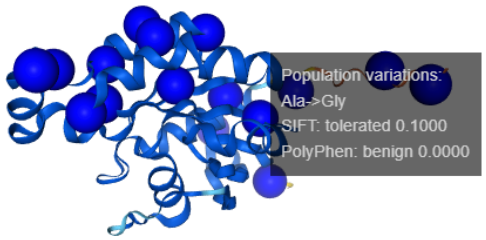
The human protein atlas

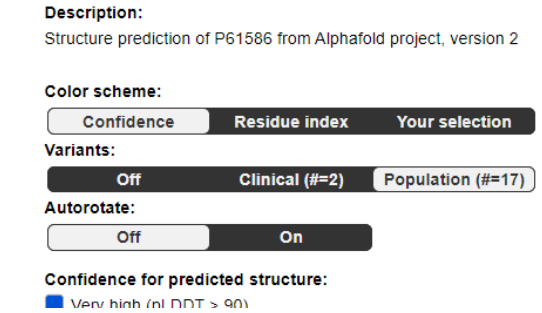
- Субклеточная локализация белка

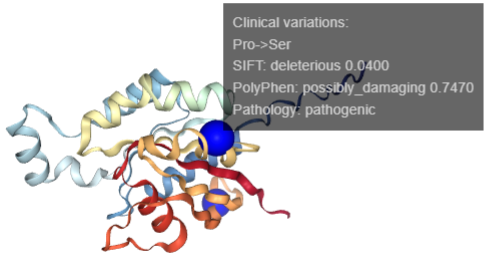


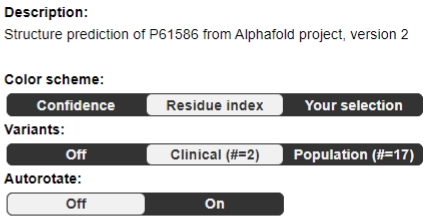
The human protein atlas

- Структура белка с популяционными и клиническими вариациями







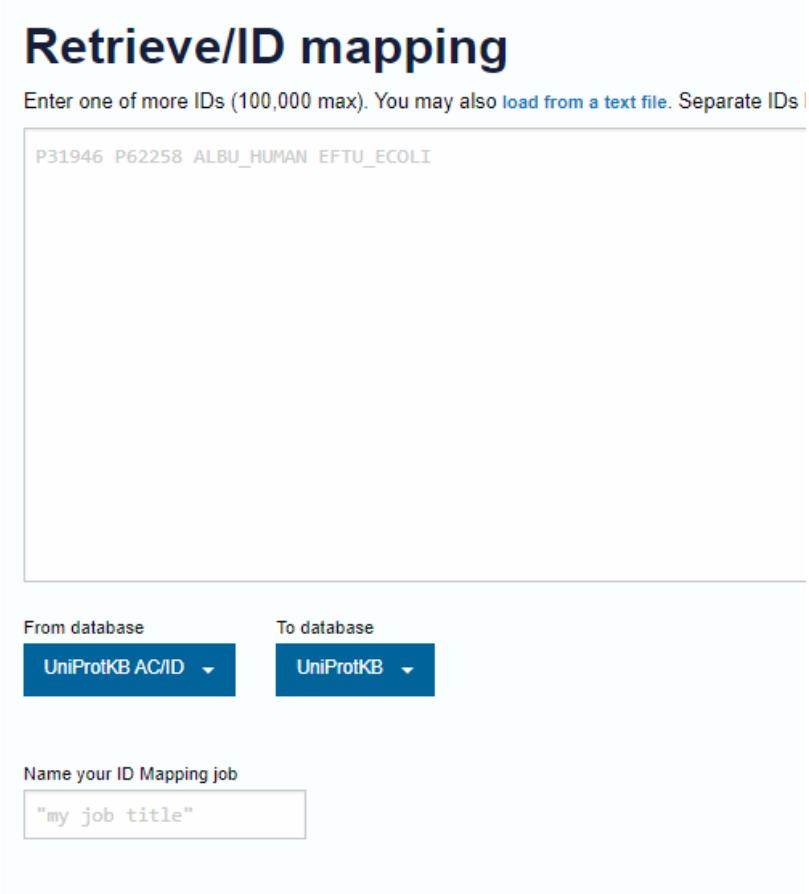


И многое другое!

UniProtKB

Retrieve/ID mapping

Сервис позволяет перевести список ID из одной номенклатуры в другую



Для работы на семинаре

- Остановимся на гене APOB
- Вариант в этом гене:
  - ENST00000233242.5:c.2786del
  - ENSP00000233242.1:p.Pro929GlnfsTer24
  - rs1553385404
  - 2:21022861-21022862
  - frameshift_variant
  - Impact - HIGH

Упражнение

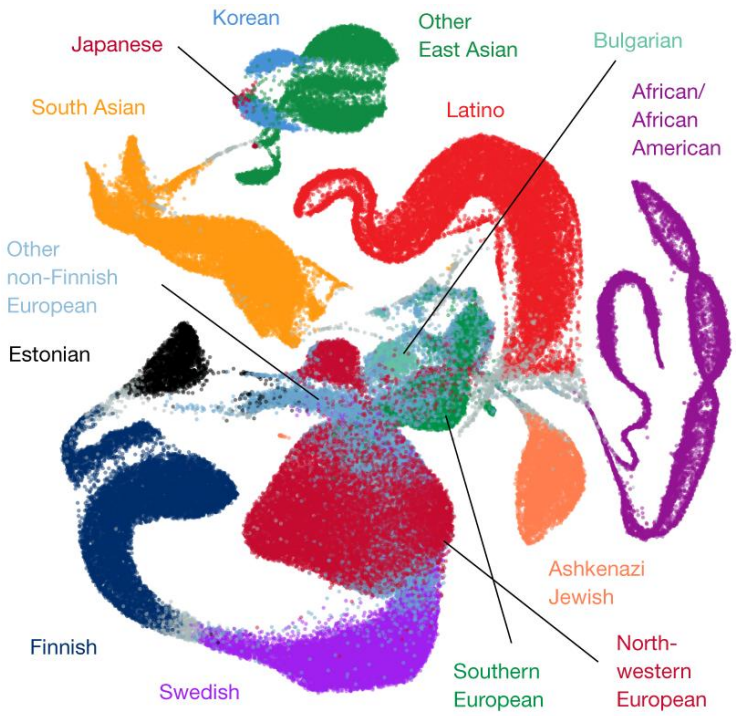
- На предыдущем слайде один вариант охарактеризован 6 способами
- Расшифруйте\объясните каждый из них

Задача семинара

- Описанный выше вариант в гене APOB мы нашли у конкретного человека
- Необходимо описать этот вариант средствами баз данных
  - gnomad
  - ClinVar

gnomAD

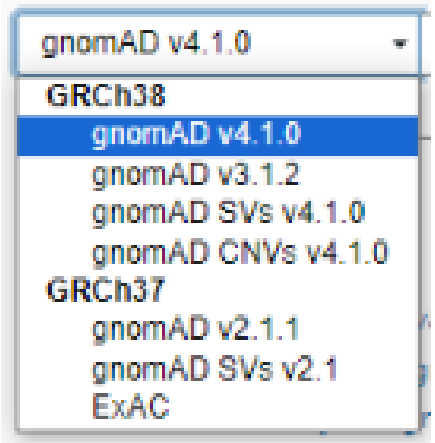
- https://gnomad.broadinstitute.org/



https://www.nature.com/articles/s41586-020-2308-7

gnomAD

- Новая версия v4.1.0
- 1 ноября 2023 (v4.0.0)
- https://gnomad.broadinstitute.org/stats

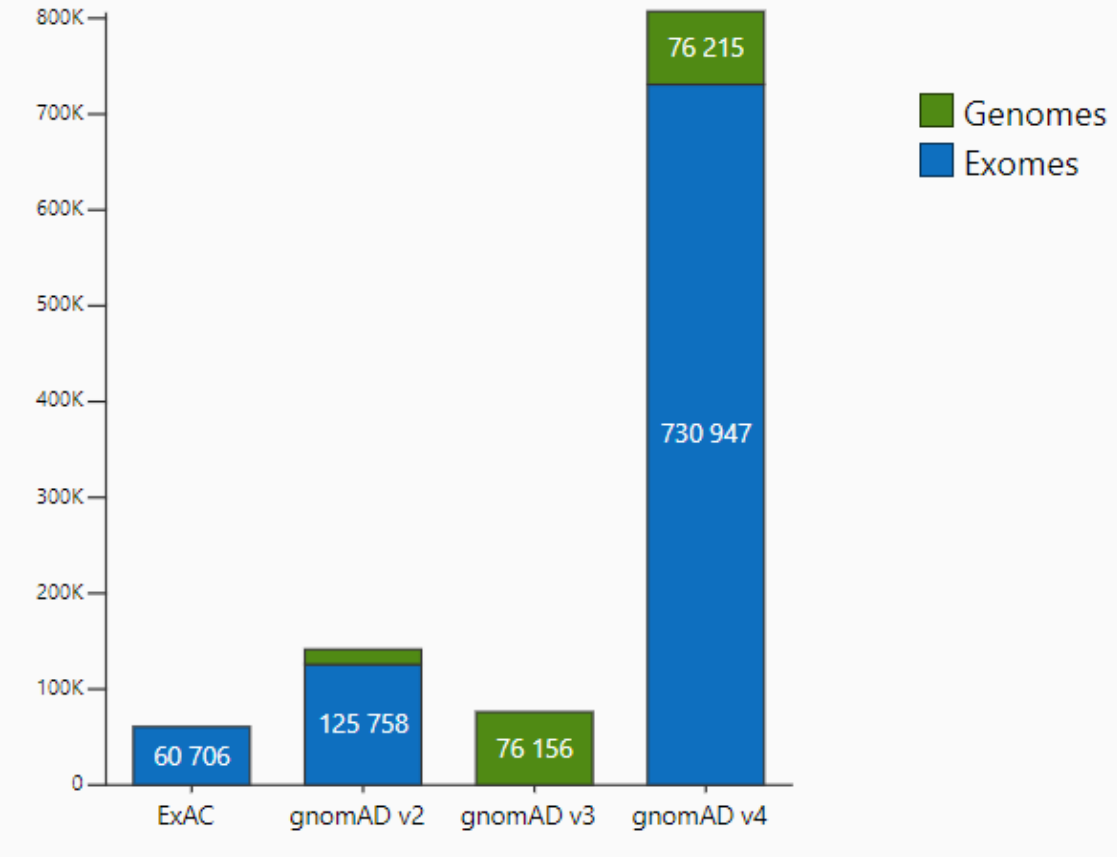


gnomAD

gnomAD
Genome Aggregation Database

gnomAD v4 includes 807,162 individuals

- 730,947 exomes
- 76,215 genomes



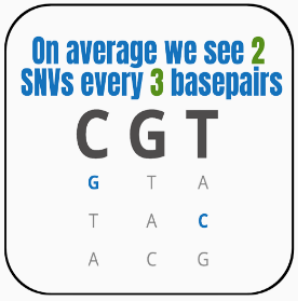
Откуда появилось столько новых вариантов?

gnomAD

gnomAD
Genome Aggregation Database

Short variants

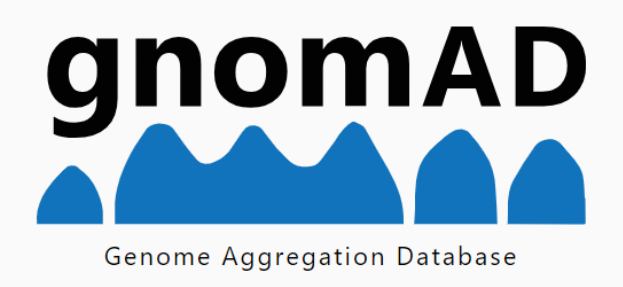
- Total SNVs: 786,500,648
- Total InDels: 122,583,462
- Variant type* counts
  - Synonymous: 9,643,254
  - Missense: 16,412,219
  - Nonsense: 726,924
  - Frameshift: 1,186,588
  - Canonical splice site: 542,514



Structural variants

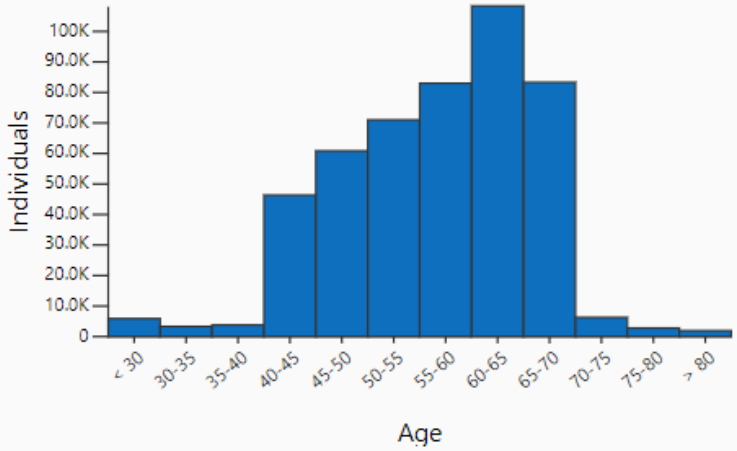
- 1,199,117 genome SVs
  - 627,947 Deletions
  - 258,882 Duplications
  - 711 CNVs
  - 296,184 Insertions
  - 2,185 Inversersions
  - 13,116 Complex
  - 92 Canonical recriprocal translocations
- 66,903 rare (<1% site frequency (SF)) exome CNVs
  - 30,877 Deletions
  - 36,026 Duplications

gnomAD

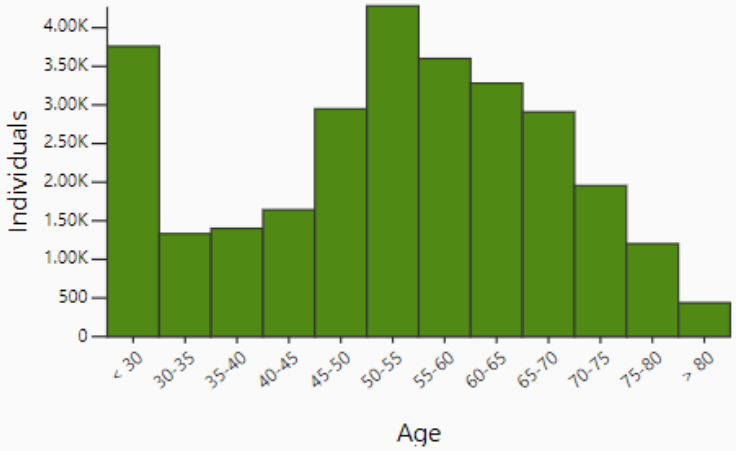


Age

Exomes



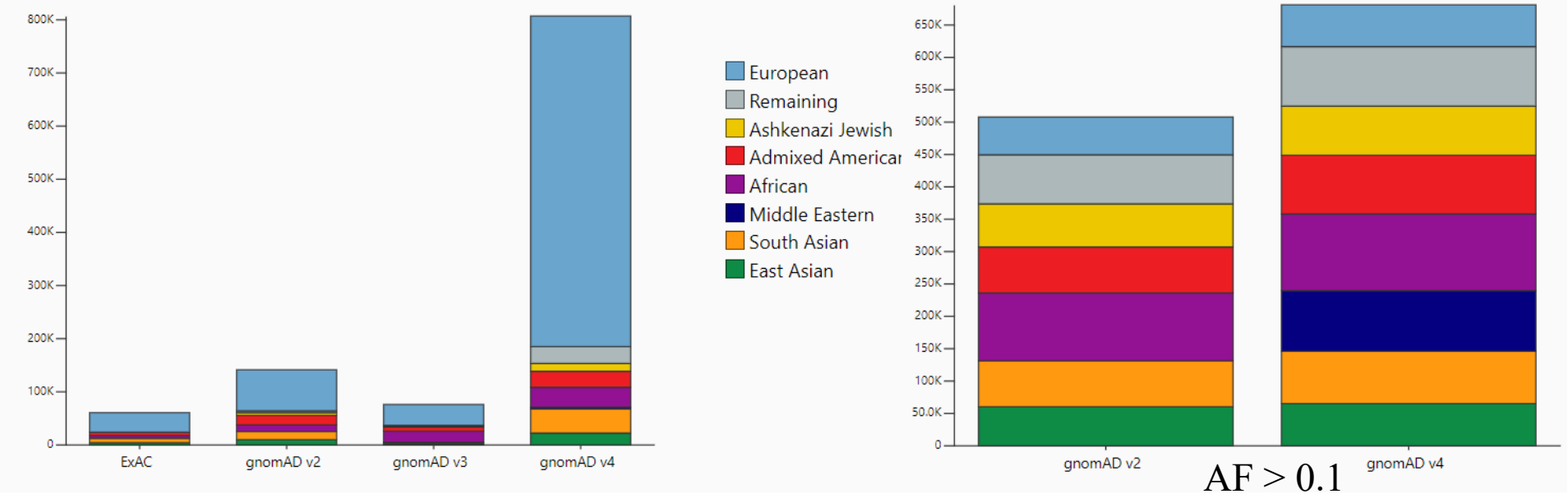
Genomes



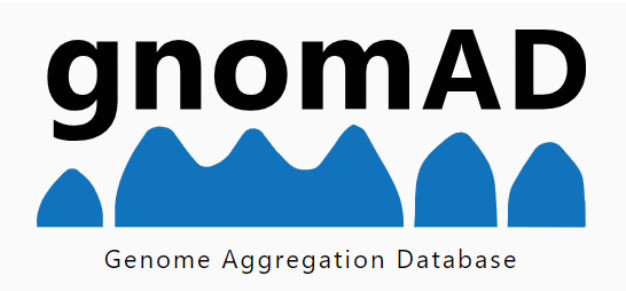
Sex

- 406,265 XX individuals
- 400,897 XY individuals

gnomAD

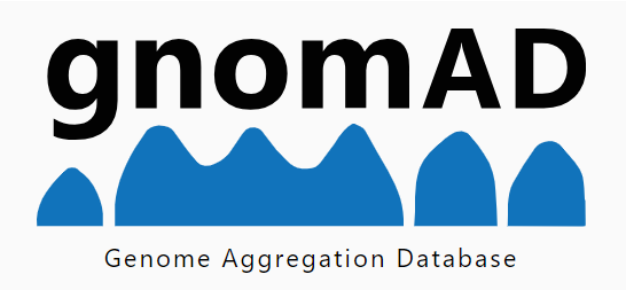

800K
700K
600K
500K
400K
300K
200K
100K
0
ExAC
gnomAD v2
gnomAD v3
gnomAD v4
European
Remaining
Ashkenazi Jewish
Admixed American
African
Middle Eastern
South Asian
East Asian
650K
600K
550K
500K
450K
400K
350K
300K
250K
200K
150K
100K
50.0K
0
gnomAD v2
gnomAD v4
AF > 0.1

gnomAD



	ExAC	gnomAD v2	gnomAD v3	gnomAD v4*		
	#	#	#	#	%	Fold increase from v2
Admixed American	5,789	17,720	7,647	30,019	3.72%	1.7x
African	5,203	12,487	20,744	37,545	4.65%	3x
Ashkenazi Jewish	-	5,185	1,736	14,804	1.83%	2.9x
East Asian	4,327	9,977	2,604	22,448	2.78%	2.3x
European^	36,667	77,165	39,345	622,057	77.07%	8.1x
Middle Eastern	-	-	158	3,031	0.38%	19.2x
Remaining Individuals^	454	3,614	1,503	31,172	3.93%	8.8x
South Asian	8,256	15,308	2,419	45,546	5.64%	3x
Total	**60,706**	**141,456**	**76,156**	**-**	**807,162**	**-**

gnomAD



- Поиск можно проводить, начиная с:
  - Gene: PCSK9
  - Transcript: ENST00000302118
  - Variant: 1-55051215-G-GA
  - Structural variant region: 19-11078371-11144910
  - Copy number variant region: 19-11078371-11144910
  - Mitochondrial variant: M-8602-T-C
  - Short tandem repeat locus: ATXN1
  - Regional missense constraint (gnomAD v2, GRCh37): GRIN2A
  - Variant co-occurrence (gnomAD v2, GRCh37): 1-55505647-G-T and 1-55523855-G-A

Упражнение

- В браузере gnomAD найдите исследуемый вариант по любой подходящей характеристике
- Убедитесь, что при поиске вы используете новую версию v4.1.0

- Далее обсудим выдачу поиска от варианта

gnomAD

- Обратите внимание, что есть возможность увидеть результаты отдельно для экзомов и геномов

	Exomes	Genomes	Total
Filters	Pass	No variant	
Allele Count	1		1
Allele Number	628768 *		628768 *
Allele Frequency	0.000001590		0.000001590
Grpmax Filtering AF (95% confidence)	0		—
Number of homozygotes	0		0
Fraction of individuals with >20x coverage	1.0		

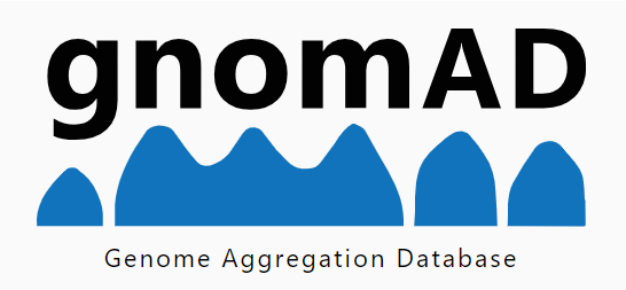
External Resources

- dbSNP (rs1553385404)
- UCSC
- ClinVar (440527)
- All of Us

Feedback

Report an issue with this variant

gnomAD



Genetic Ancestry Group Frequencies

Genetic Ancestry Group	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (non-Finnish)	1	350098	0	0.000002856
Remaining	0	33096	0	0.000
Admixed American	0	43740	0	0.000
European (Finnish)	0	53140	0	0.000
Middle Eastern	0	4148	0	0.000
South Asian	0	69798	0	0.000
Ashkenazi Jewish	0	20984	0	0.000
East Asian	0	36070	0	0.000
African/African American	0	17694	0	0.000
XX	1	286240	0	0.000003494
XY	0	342528	0	0.000
Total	**1**	**628768**	**0**	**0.000001590**

gnomAD

- Related Variants
- Nearby Variants
- View variants located within 20 bases of this variant.

Variant ID	Source	Gene	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance
2-21022841-G-A	E G	APOB	p.Leu936Phe	missense		
2-21022842-C-T	E	APOB	p.Leu935Leu	synonymous		
2-21022843-A-C	E	APOB	p.Leu935Arg	missense		
2-21022845-C-G	E	APOB	p.Lys934Asn	missense		
2-21022849-A-G	E	APOB	p.Val933Ala	missense		
2-21022852-G-C	E	APOB	p.Pro932Arg	missense		
2-21022852-G-A	E	APOB	p.Pro932Leu	missense		
2-21022854-T-C	E	APOB	p.Arg931Arg	synonymous		
2-21022854-T-G	E	APOB	p.Arg931Ser	missense		
2-21022856-T-C	E	APOB	p.Arg931Gly	missense		
2-21022860-TG-T	E	APOB	p.Pro929GlnfsTer24	frameshift		Pathogenic
2-21022862-G-C	E	APOB	p.Pro929Ala	missense		
2-21022863-G-A	E	APOB	p.Ser928Ser	synonymous		
2-21022865-A-G	E	APOB	p.Ser928Pro	missense		
2-21022867-G-A	E	APOB	p.Pro927Leu	missense		
2-21022868-G-T	E G	APOB	p.Pro927Thr	missense		
2-21022869-A-T	E	APOB	p.Ile926Ile	synonymous		
2-21022871-T-A	E	APOB	p.Ile926Phe	missense		
2-21022872-G-A	E	APOB	p.Ile925Ile	synonymous		
2-21022878-C-G	E	APOB	p.Lys923Asn	missense		

gnomAD

gnomAD
Genome Aggregation Database

- Откуда эта информация?

Variant Effect Predictor

This variant falls on 3 transcripts in 1 gene.

Note The gene symbols shown below are provided by VEP and may differ from the symbol shown on gene pages.

frameshift

1. **APOB**
   1. ENST00000233242.5
      MANE Select transcript for APOB
      HGVSp: p.Pro929GlnfsTer24
      Domains: PF09172 (Pfam), and 4 more
      pLoF: ● High-confidence

3' UTR

1. **APOB**
   1. ENST00000673739.1
      HGVSc: c.*2092del
   2. ENST00000673882.1
      HGVSc: c.*2092del

Упражнение

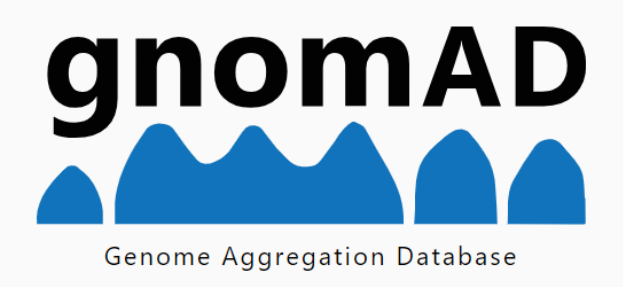
- На слайде выше
  - вариант найден в трех транскриптах
  - вариант отнесен к двум разным категориям замен (frameshift, 3`UTR)
  - у одного из транскриптов указано “MANE”

Обсудите, что это значит и почему так произошло

Упражнение

- С страницы gnomAD с описанием варианта перейдите по ссылке на страницу гена, в котором найден вариант

gnomAD



APOB apolipoprotein B

Dataset gnom

Genome build GRCh38 / hg38
Ensembl gene ID ENSG00000084674.15
MANE Select transcript ENST00000233242.5 / NM_000384.3
Ensembl canonical transcript ENST00000233242.5
Other transcripts ENST00000399256.4, ENST00000673739.1, ENST00000673882.1
Region 2:21001429-21044073
External resources Ensembl, UCSC Browser, and more

Constraint | Variant co-occurrence

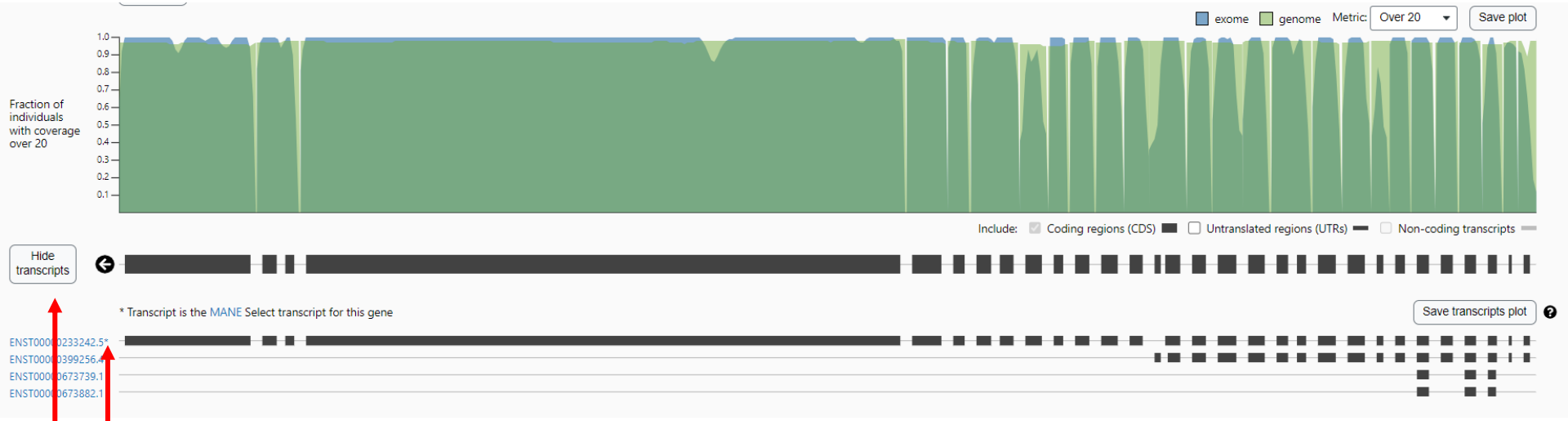
Category	Expected SNVs	Observed SNVs	Constraint metrics
Synonymous	2081.8	2013	Z = 0.82 o/e = 0.97 (0.93 - 1)
Missense	5461	5271	Z = 0.94 o/e = 0.97 (0.94 - 0.99)
pLoF	319.7	150	pLI = 0 o/e = 0.47 (0.41 - 0.54)

Constraint metrics based on MANE Select transcript (ENST00000233242.5).

Упражнение

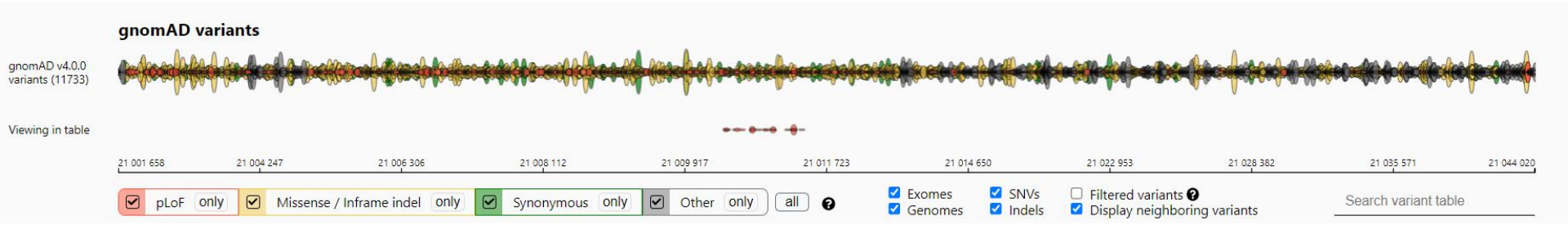
- Посмотрите на выдачу по генам:
  - GAPDH
  - CCK

gnomAD



Для каждого гена есть возможность показать все транскрипты

gnomAD



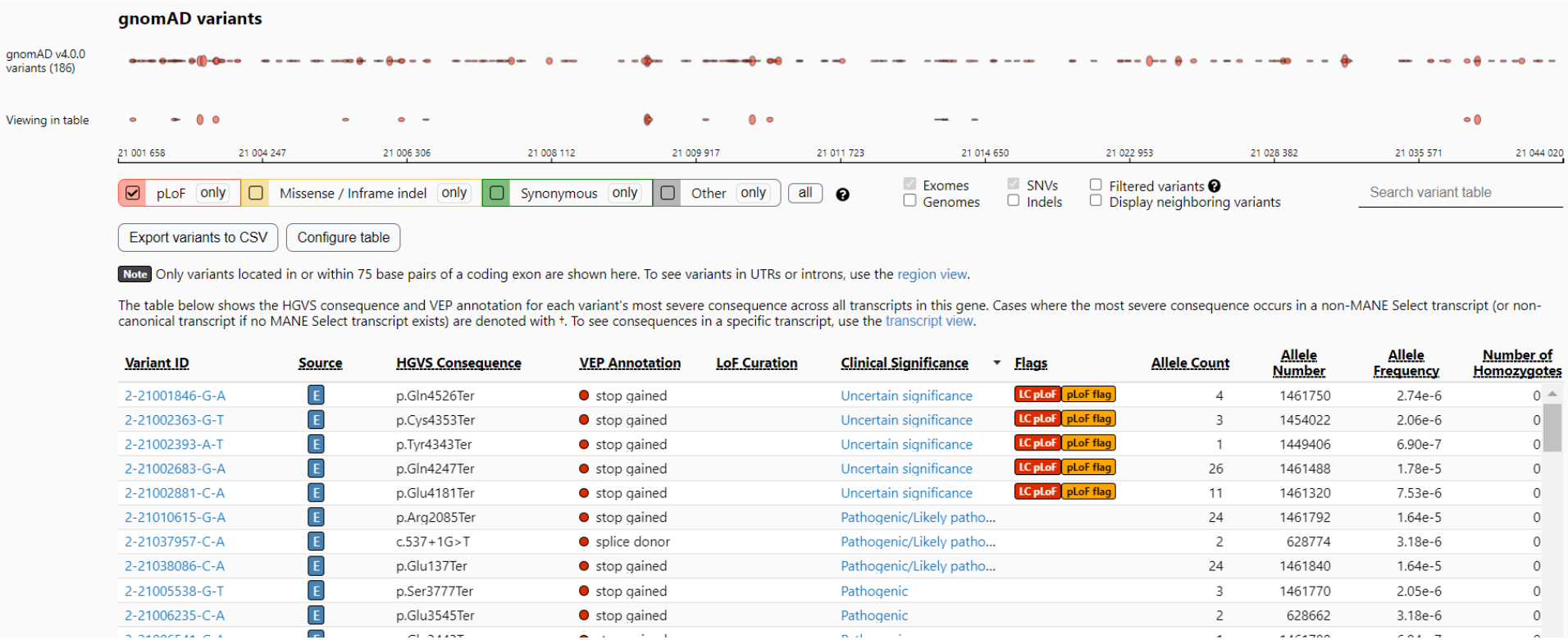
Варианты можно отфильтровать по ряду критериев

Упражнение

- Для вариантов gnomAD оставьте только однонуклеотидные pLoF в экзомах
- Отсортируйте по клинической значимости
- Охарактеризуйте полученные варианты по частоте и аннотации VEP

Упражнение

- Примерное решение



gnomAD variants

gnomAD v4.0.0 variants (186)

Viewing in table

21 001 658 | 21 004 247 | 21 006 306 | 21 008 112 | 21 009 917 | 21 011 723 | 21 014 650 | 21 022 953 | 21 028 382 | 21 035 571 | 21 044 020

pLoF only | Missense / Inframe indel only | Synonymous only | Other only | all

Exomes | Genomes | SNVs | Indels | Filtered variants | Display neighboring variants

Search variant table

Export variants to CSV | Configure table

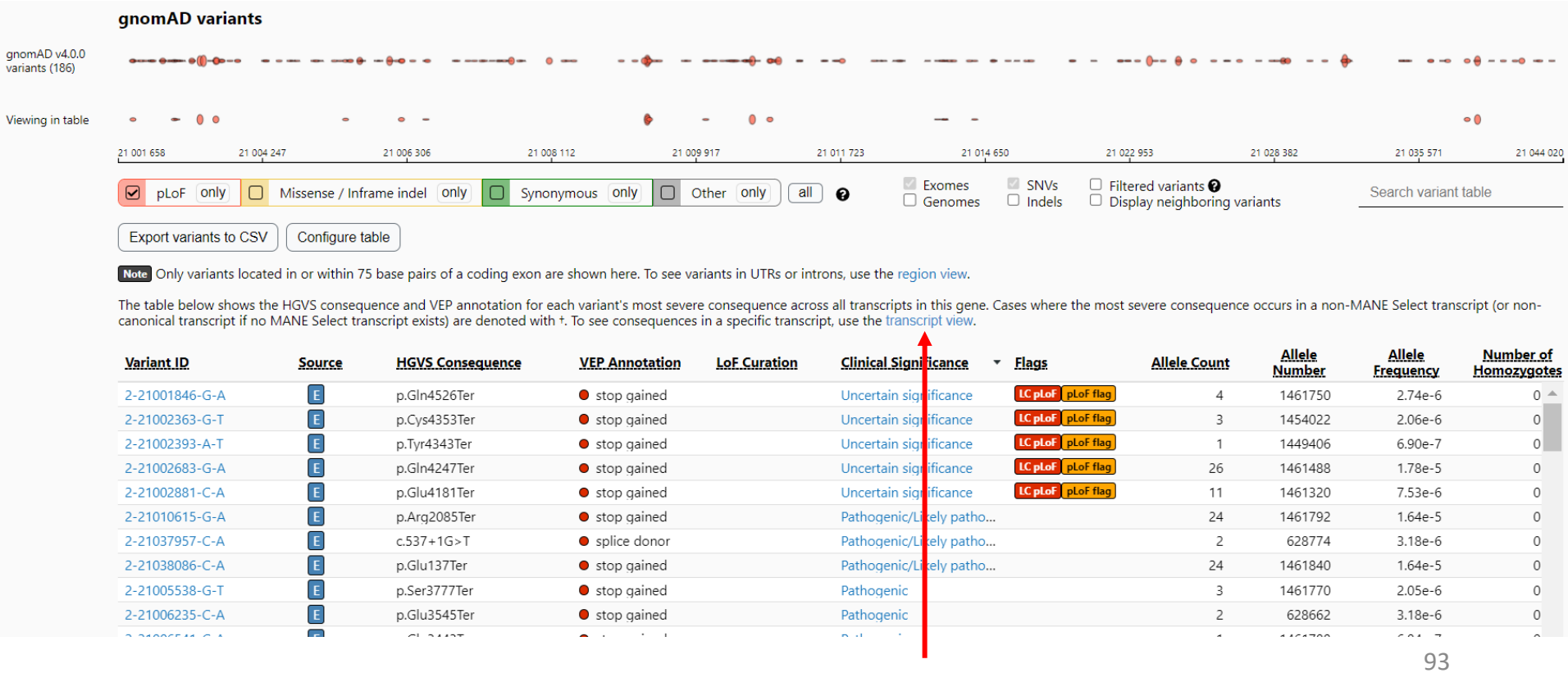
Note Only variants located in or within 75 base pairs of a coding exon are shown here. To see variants in UTRs or introns, use the region view.

The table below shows the HGVS consequence and VEP annotation for each variant's most severe consequence across all transcripts in this gene. Cases where the most severe consequence occurs in a non-MANE Select transcript (or non-canonical transcript if no MANE Select transcript exists) are denoted with †. To see consequences in a specific transcript, use the transcript view.

Variant ID	Source	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance	Flags	Allele Count	Allele Number	Allele Frequency	Number of Homozygotes
2-21001846-G-A	E	p.Gln4526Ter	stop gained		Uncertain significance	LC pLoF pLoF flag	4	1461750	2.74e-6	0
2-21002363-G-T	E	p.Cys4353Ter	stop gained		Uncertain significance	LC pLoF pLoF flag	3	1454022	2.06e-6	0
2-21002393-A-T	E	p.Tyr4343Ter	stop gained		Uncertain significance	LC pLoF pLoF flag	1	1449406	6.90e-7	0
2-21002683-G-A	E	p.Gln4247Ter	stop gained		Uncertain significance	LC pLoF pLoF flag	26	1461488	1.78e-5	0
2-21002881-C-A	E	p.Glu4181Ter	stop gained		Uncertain significance	LC pLoF pLoF flag	11	1461320	7.53e-6	0
2-21010615-G-A	E	p.Arg2085Ter	stop gained		Pathogenic/Likely patho...		24	1461792	1.64e-5	0
2-21037957-C-A	E	c.537+1G>T	splice donor		Pathogenic/Likely patho...		2	628774	3.18e-6	0
2-21038086-C-A	E	p.Glu137Ter	stop gained		Pathogenic/Likely patho...		24	1461840	1.64e-5	0
2-21005538-G-T	E	p.Ser3777Ter	stop gained		Pathogenic		3	1461770	2.05e-6	0
2-21006235-C-A	E	p.Glu3545Ter	stop gained		Pathogenic		2	628662	3.18e-6	0

gnomAD

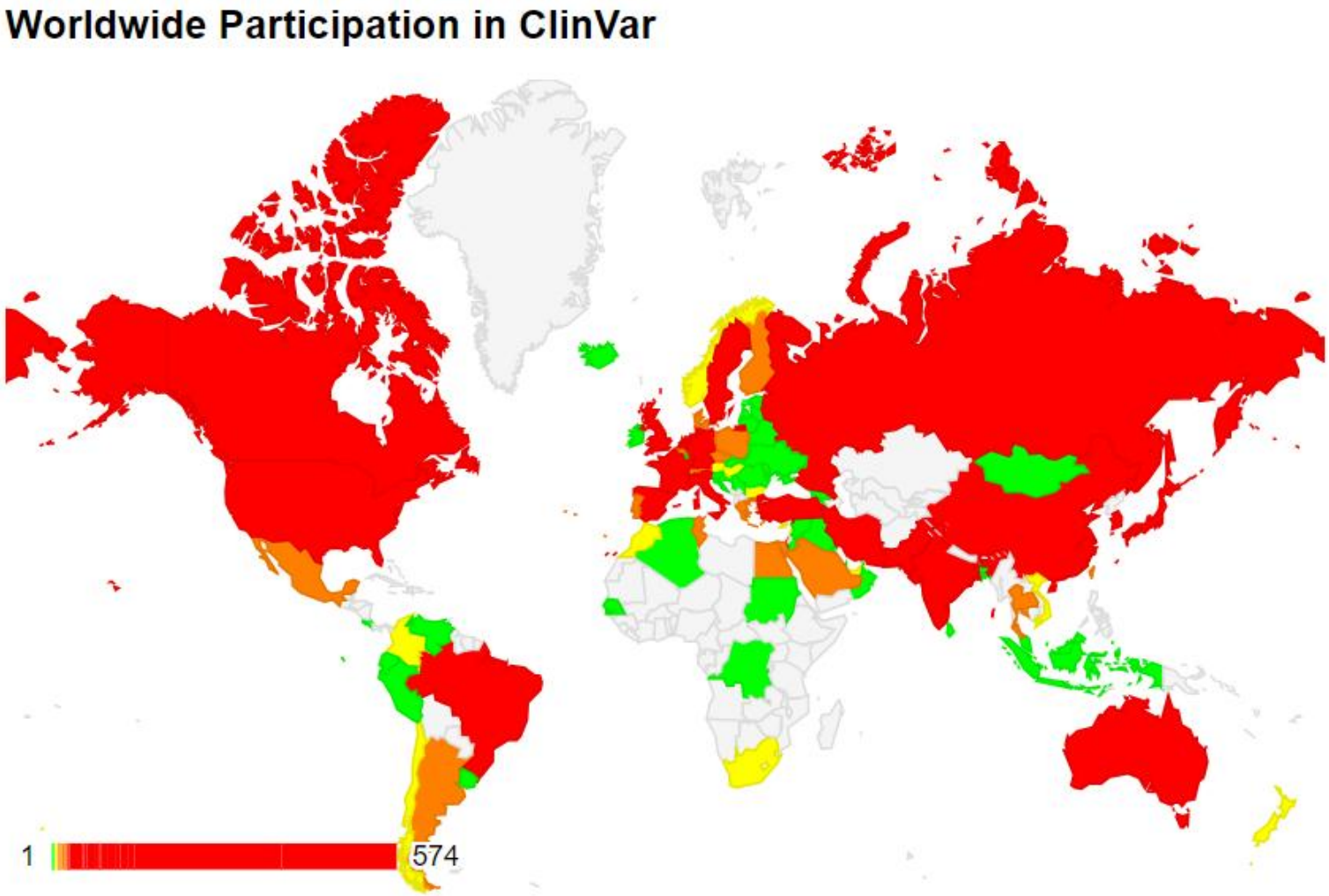
- Обратите внимание, что при исследовании можно сосредоточиться на конкретных транскриптах



ClinVar

- https://www.ncbi.nlm.nih.gov/clinvar/
- ClinVar aggregates information about genomic variation and its relationship to human health

ClinVar



Упражнение

- Найдите в базе ClinVar вариант rs1553385404

ClinVar

- Для каждого варианта получаем характеристику

Interpretation:	**Pathogenic**
Review status:	☆ ☆ ☆ ☆ no assertion criteria provided
Submissions:	1
First in ClinVar:	Oct 1, 2017
Most recent Submission:	Oct 1, 2017
Accession:	VCV000440527.1
Variation ID:	440527
Description:	1bp deletion

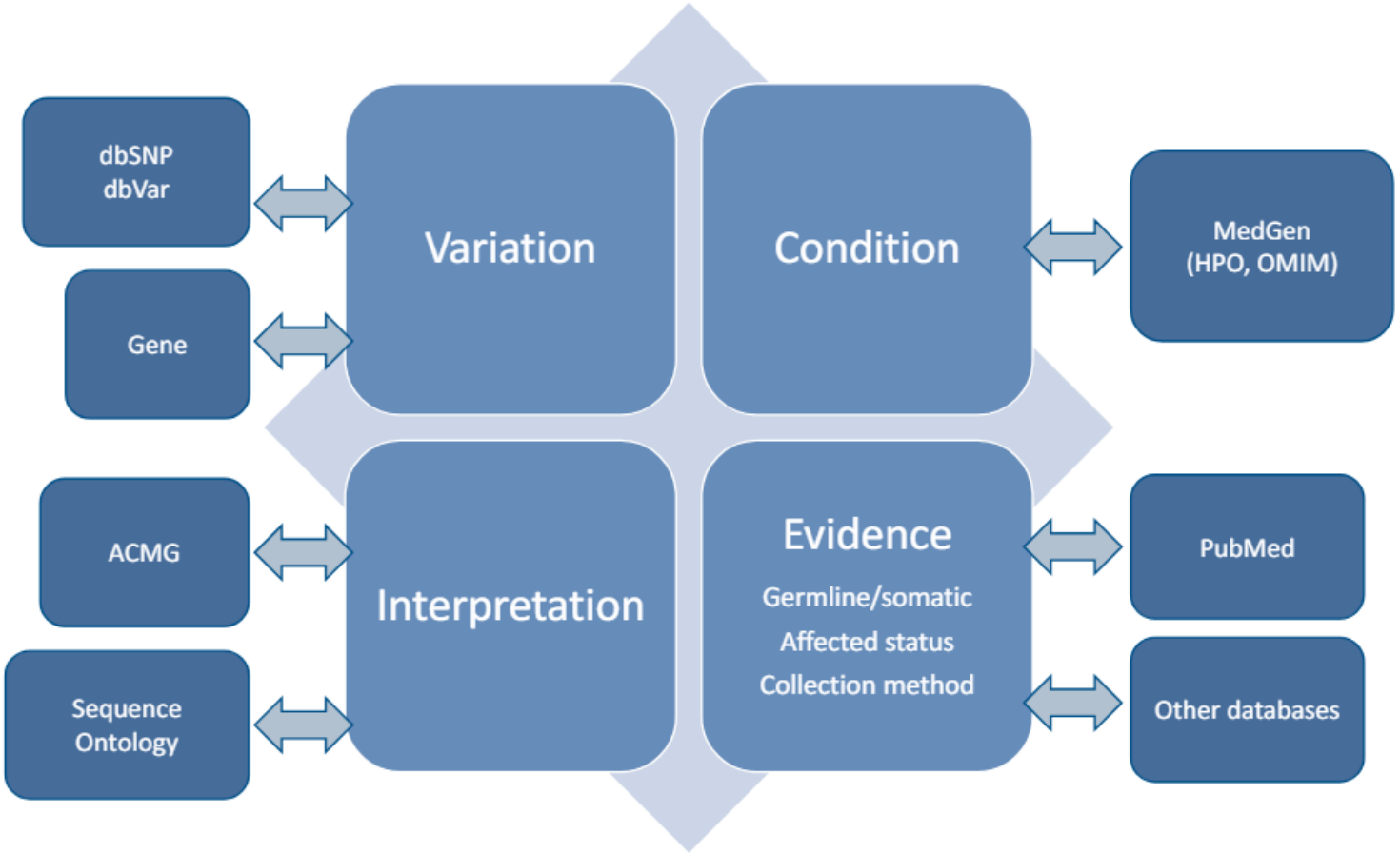
ClinVar

- Review status
- https://github.com/ncbi/clinvar/blob/master/ReviewStatus.md

Number of gold stars	Review status	**Description**
four	practice guideline	practice guideline
three	reviewed by expert panel	reviewed by expert panel
two	criteria provided, multiple submitters, no conflicts	Two or more submitters with assertion criteria and evidence (or a public contact) provided the same interpretation.
one	criteria provided, conflicting interpretations	Multiple submitters provided assertion criteria and evidence (or a public contact) but there are conflicting interpretations. The independent values are enumerated for clinical significance.
one	criteria provided, single submitter	One submitter provided an interpretation with assertion criteria and evidence (or a public contact).
none	no assertion for the individual variant	The allele was not interpreted directly in any submission; it was submitted to ClinVar only as a component of a haplotype or a genotype.
none	no assertion criteria provided	The allele was included in a submission with an interpretation but without assertion criteria and evidence (or a public contact).
none	no assertion provided	The allele was included in a submission that did not provide an interpretation.

Practice guideline

ClinVar



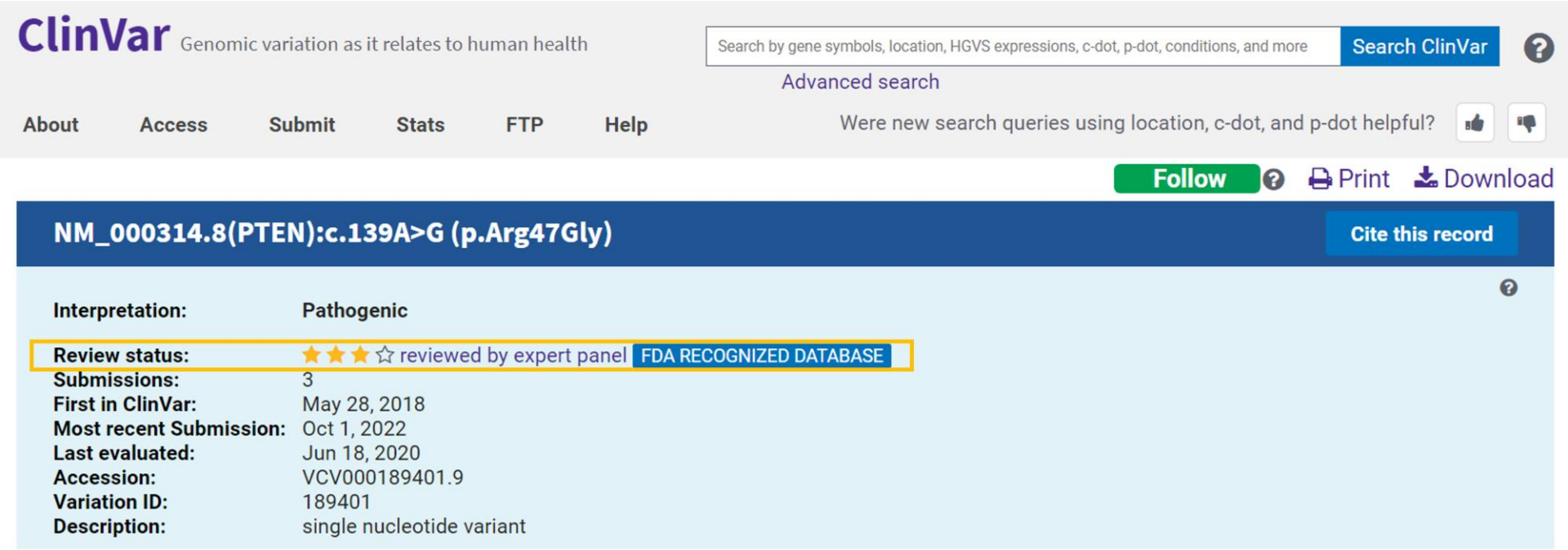
About ClinVar

ClinVar

Category of analysis	Current total (Nov 21, 2023)
Records submitted	3544617
Records with assertion criteria	3237894
Records with an interpretation	3513987
Total genes represented	92077
Unique variation records	2388783
Unique variation records with interpretations	2377607
Unique variation records with assertion criteria	2273062
Unique variation records with practice guidelines (4 stars)	663
Unique variation records from expert panels (3 stars)	15473
Unique variation records with assertion criteria, multiple submitters, and no conflicts (2 stars)	345008
Unique variation records with assertion criteria (1 star)	1805893
Unique variation records with assertion criteria and a conflict (1 star)	106025
Unique variation records with conflicting interpretations	106298
Genes with variants specific to one gene	17213
Genes with variants specific to one protein-coding gene	17053
Genes included in a variant spanning more than one gene	92406
Variants affecting overlapping genes	35479
Total submitters	2686

ClinVar

- Как бы хотелось, чтоб было



Representation of classifications in ClinVar

- https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/

Representation of classifications in ClinVar

- Types of classifications
- Classfication on ClinVar submitted records (SCV)
  - Standards for classification terms
  - Options for germline classification
    - Assertion score
    - Clinical significance and mode of inheritance
  - Options for somatic classification
  - Source of classifications
- Classifications on ClinVar aggregate records (VCV and RCV)
  - Overview
  - Aggregate germline classification
  - Aggregate somatic classification of clinical impact
  - Aggregate somatic classification of oncogenicity
- Terminology

Упражнение

- Поищите в базе ClinVar ген APOB
- Отберите только патогенные варианты, попавшие в сайты сплайсинга
- Выберите один вариант, для которого указано более одного факта регистрации и отсутствие конфликтов в интерпретации
- Изучите выдачу ClinVar

ClinVar

- Результаты поиска гена APOB

Clinical significance
Conflicting interpretations (0)
Benign (0)
Likely benign (0)
Uncertain significance (0)
Likely pathogenic (45)
Pathogenic (115)

Molecular consequence
Frameshift (64)
Missense (18)
Nonsense (48)
Splice site (14)
ncRNA (0)
Near gene (0)
UTR (0)

Variation type
Deletion (60)
Duplication (10)
Indel (2)
Insertion (12)
Single nucleotide (75)

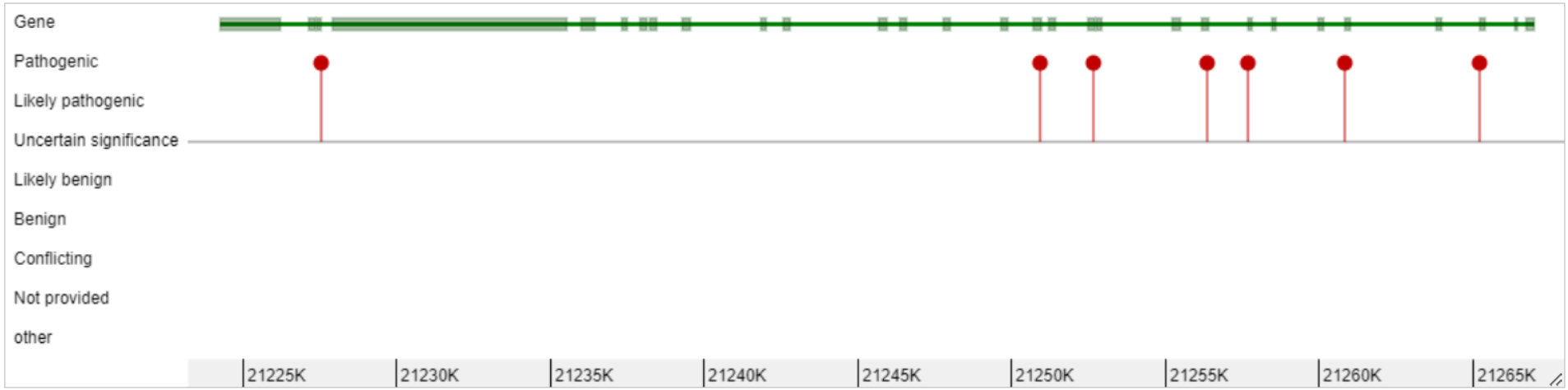
Variation size
Short variant (< 50 bps) (145)
Structural variant (>= 50 bps) (0)

Variant length
< 1kb, single gene (145)
> 1kb, single gene (0)
> 1kb, multiple genes (0)

Review status
Practice guideline (0)
Expert panel (0)
Multiple submitters (22)
Single submitter (97)
At least one star (119)
Conflicting interpretations (0)

Упражнение

- Примерное решение



ClinVar

Interpretation:	**Pathogenic**
Review status:	★★☆☆ criteria provided, multiple submitters, no conflicts
Submissions:	3
First in ClinVar:	Dec 26, 2017
Most recent Submission:	Feb 7, 2023
Last evaluated:	Aug 31, 2021
Accession:	VCV000477783.10
Variation ID:	477783
Description:	single nucleotide variant

ClinVar

NM_000384.3(APOB):c.1830-1G>A

Allele ID: 450492
Variant type: single nucleotide variant
Variant length: 1 bp
Cytogenetic location: 2p24.1
Genomic location: 2: 21028066 (GRCh38) GRCh38 UCSC
2: 21250938 (GRCh37) GRCh37 UCSC

HGVS:

Nucleotide	Protein	Molecular consequence
NM_000384.3:c.1830-1G>A MANE SELECT		splice acceptor
NC_000002.12:g.21028066C>T		
NC_000002.11:g.21250938C>T		
NG_011793.1:g.21008G>A		
NG_011793.2:g.21007G>A		

... less HGVS

Protein change: -
Other names: -
Canonical SPDI: NC_000002.12:21028065:C:T
Functional consequence: -
Global minor allele frequency (GMAF): -
Allele frequency: -
Links: ClinGen: CA346013946
dbSNP: rs1399892057
VarSome

ClinVar

- Информация о фактах регистрации варианта

Interpretation (Last evaluated)	Review status (Assertion criteria)	Condition (Inheritance)	Submitter	More information
Pathogenic (Jul 25, 2017)	criteria provided, single submitter (Invitae Variant Classification Sherloc (09022015)) Method: clinical testing	- Familial hypobetalipoproteinemia 1 Affected status: unknown Allele origin: germline	Invitae Accession: SCV000659242.1 First in ClinVar: Dec 26, 2017 Last updated: Dec 26, 2017	Comment: This sequence change affects an acceptor splice site in intron 13 of the APOB gene. It is expected to disrupt RNA splicing and likely results … (more)
Pathogenic (Aug 29, 2019)	criteria provided, single submitter (ACMG Guidelines, 2015) Method: clinical testing	- Hypobetalipoproteinemia, familial, 1 Affected status: unknown Allele origin: germline	Knight Diagnostic Laboratories, Oregon Health and Sciences University Accession: SCV001448948.1 First in ClinVar: Dec 12, 2020 Last updated: Dec 12, 2020	
Pathogenic (Aug 31, 2021)	criteria provided, single submitter (Invitae Variant Classification Sherloc (09022015)) Method: clinical testing	- Familial hypobetalipoproteinemia 1 - Hypercholesterolemia, autosomal dominant, type B Affected status: unknown Allele origin: germline	Invitae Accession: SCV001581337.3 First in ClinVar: May 10, 2021 Last updated: Feb 07, 2023	Publications: PubMed (3) Comment: This sequence change affects an acceptor splice site in intron 13 of the APOB gene. It is expected to disrupt RNA splicing. Variants that disrupt … (more)

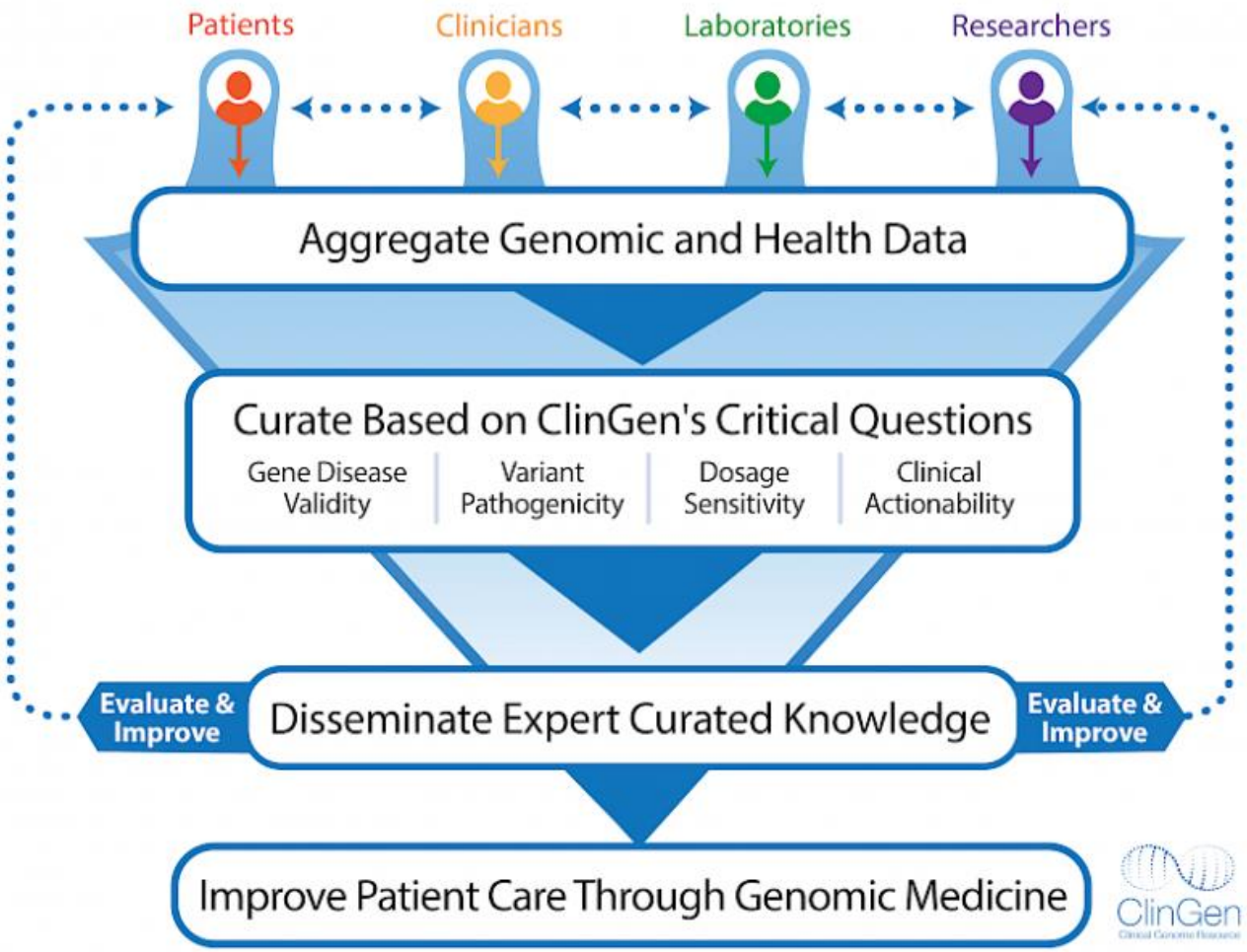
ClinVar

- https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/

Disease name and MIM number	MedGen	Gene via GTR	Variations that may be pathogenic
Adenomatous polyposis coli (MIM 175100)	MedGen	APC (MIM 611731)	ClinVar
Aortic aneurysm, familial thoracic 4 (MIM 132900)	MedGen	MYH11 (MIM 160745)	ClinVar
Aortic aneurysm, familial thoracic 6 (MIM 611788)	MedGen	ACTA2 (MIM 102620)	ClinVar
Arrhythmogenic right ventricular cardiomyopathy, type 5 (MIM 604400)	MedGen	TMEM43 (MIM 612048)	ClinVar
Arrhythmogenic right ventricular cardiomyopathy, type 8 (MIM 607450)	MedGen	DSP (MIM 125647)	ClinVar
Arrhythmogenic right ventricular cardiomyopathy, type 9 (MIM 609040)	MedGen	PKP2 (MIM 602861)	ClinVar
Arrhythmogenic right ventricular cardiomyopathy, type 10 (MIM 610193)	MedGen	DSG2 (MIM 125671)	ClinVar
Arrhythmogenic right ventricular cardiomyopathy, type 11 (MIM 610476)	MedGen	DSC2 (MIM 125645)	ClinVar
Biotinidase deficiency (MIM 253260)	MedGen	BTD (MIM 609019)	ClinVar
Breast-ovarian cancer, familial 1 (MIM 604370)	MedGen	BRCA1 (MIM 113705)	ClinVar

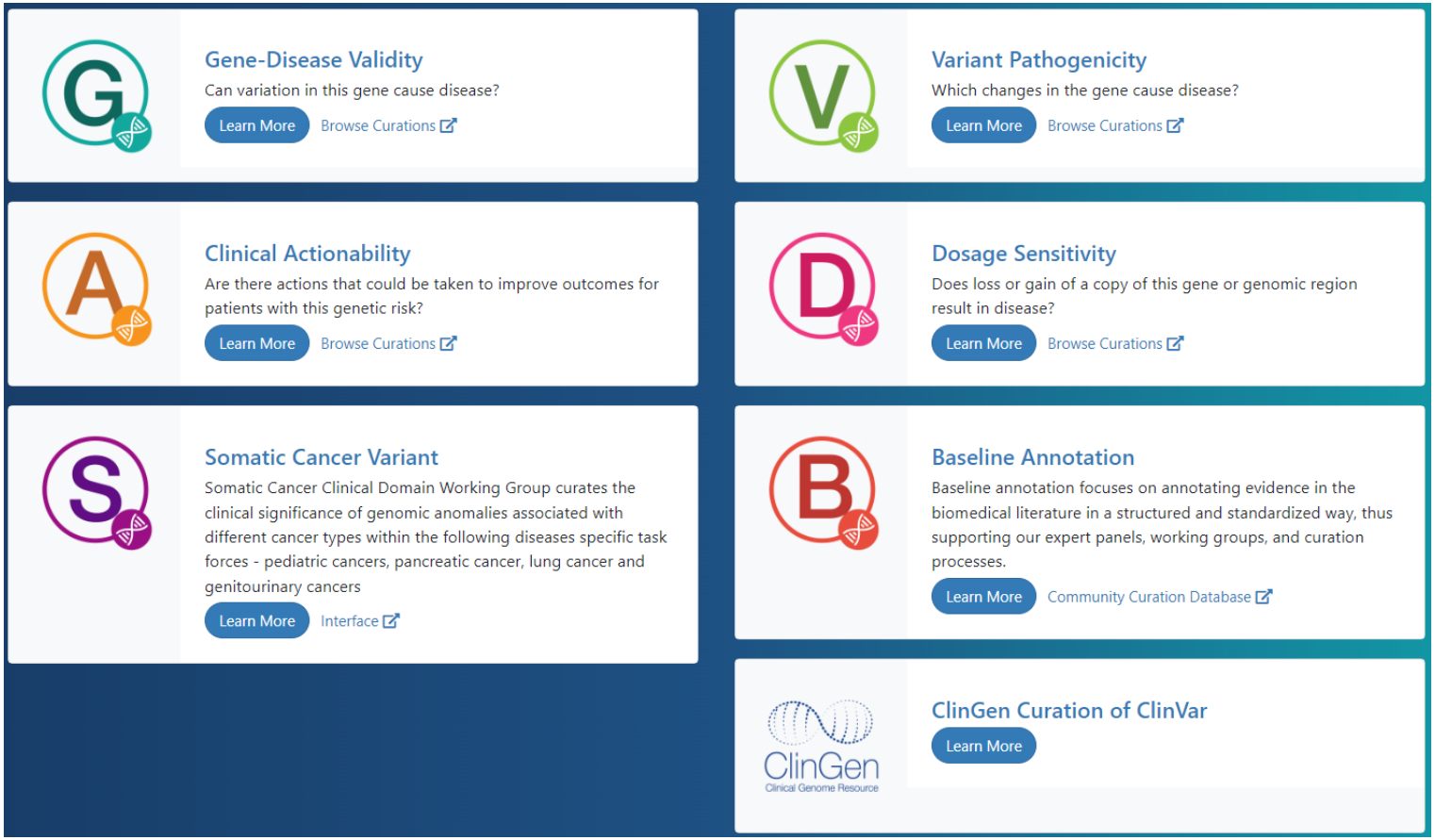
ClinGen

ClinGen
Clinical Genome Resource



ClinGen

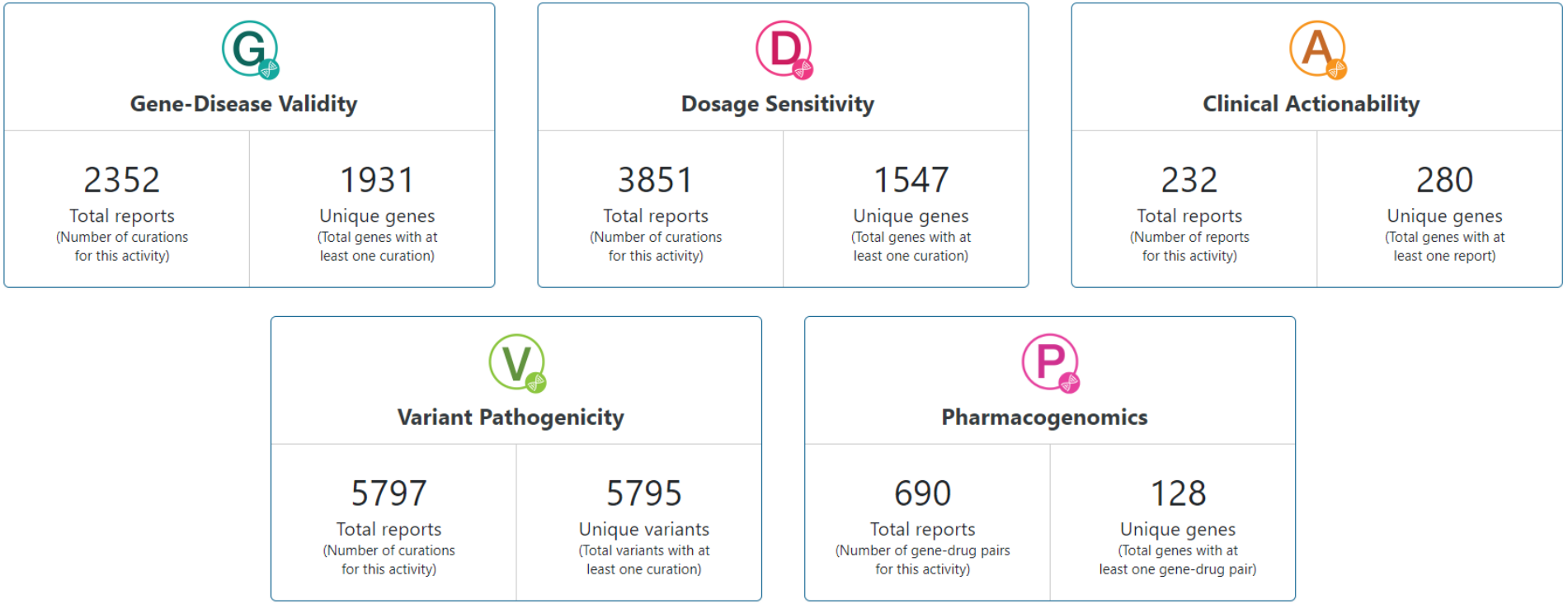
- https://clinicalgenome.org/



ClinGen

ClinGen
Clinical Genome Resource

- https://search.clinicalgenome.org/kb/reports/stats

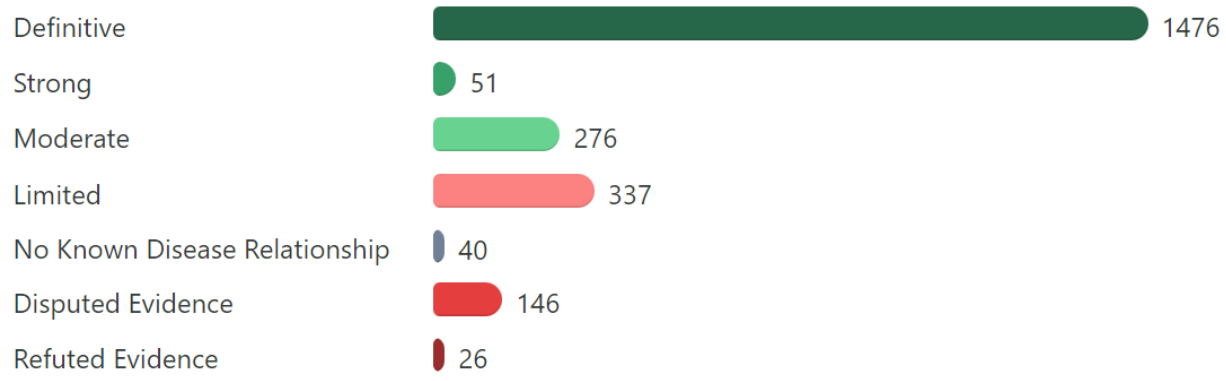


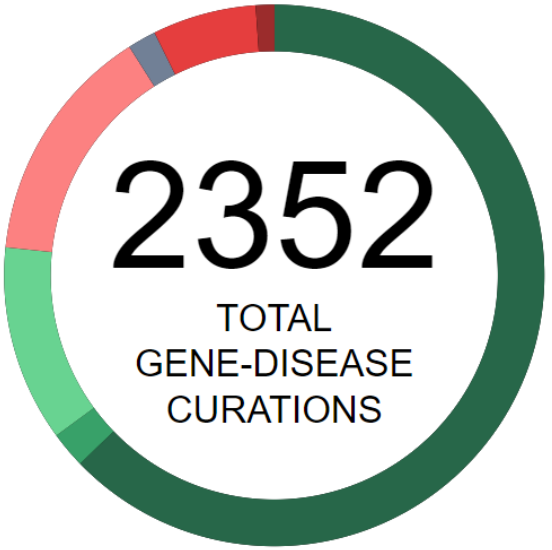
ClinGen

ClinGen
Clinical Genome Resource

Classification Statistics

Gene-Disease Clinical Validity has **2352 curations** encompassing **1931 genes**.





[About ClinVar & ClinGen]

Упражнение

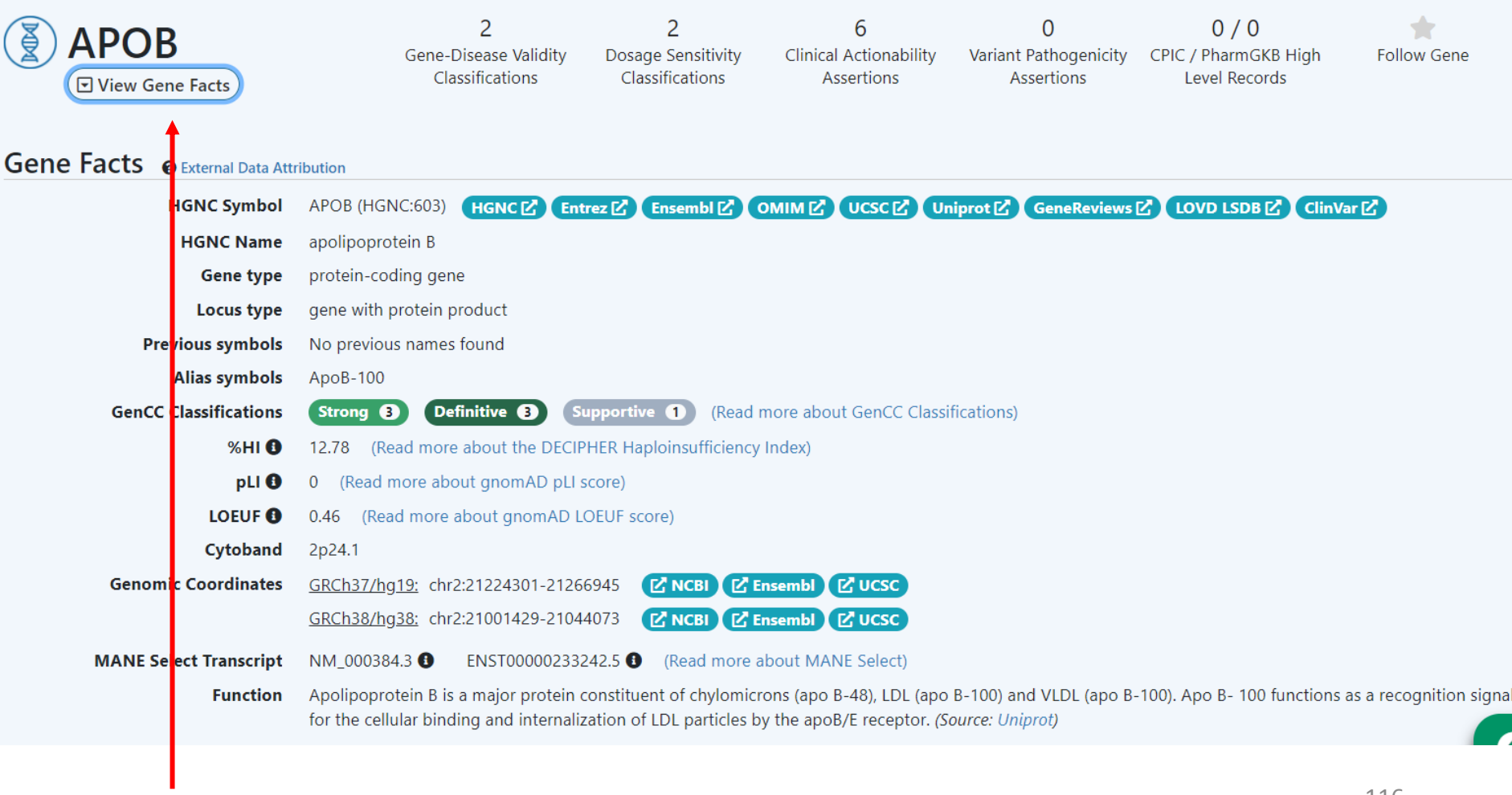
- Найдите в базе ClinGen информацию о гене APOB

ClinGen


APOB - hypercholesterolemia, autosomal dominant, type B
(MONDO:0007751)
Activity
MOI / Report
Expert Panel / Working Group
Classification
Report & Date
Gene-Disease Validity
Autosomal Dominant
General Gene Curation GCEP
Definitive
11/14/2018
Clinical Actionability
Heterozygous Familial Hypercholesterolemia
Pediatric Actionability WG
Pediatric
Strong Actionability
08/03/2020
Homozygous Familial Hypercholesterolemia
Pediatric Actionability WG
Pediatric
Strong Actionability
08/03/2020
Homozygous Familial Hypercholesterolemia
Adult Actionability WG
Adult
Strong Actionability
08/03/2020
Adult Actionability WG
Adult
Definitive Actionability
08/03/2020

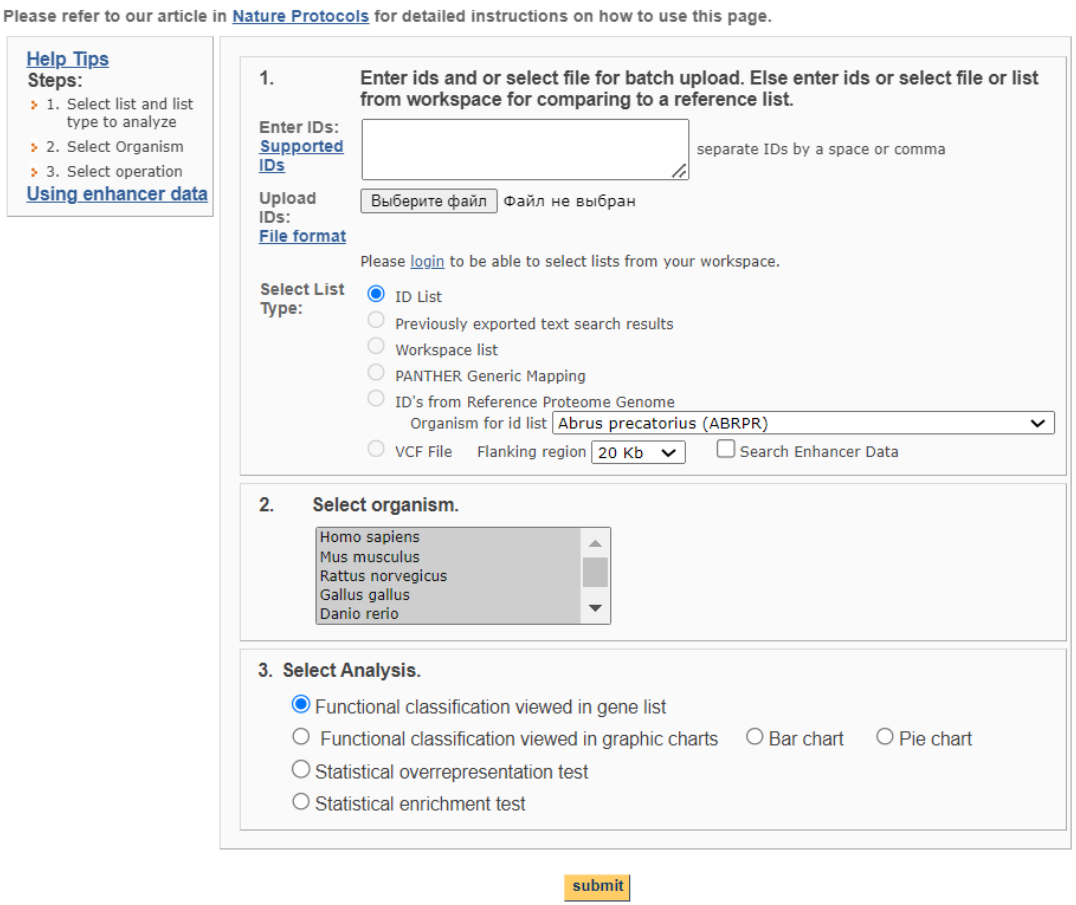
ClinGen

ClinGen
Clinical Genome Resource



PANTHER

- https://www.pantherdb.org/



PANTHER

PANTHER™ website news
September 17, 2023

PANTHER18.0 Released.

- PANTHER18.0 is generated from the 2022_02 and 2023_01 release of ReferenceProteome dataset. Here is the composition of all genomes.
  - 143 total genomes
  - 35 bacteria
  - 8 archaea
  - 15 fungus
  - 40 plants
  - 8 protista and alveolata
  - 3 amoebazoa
  - 15 invertebrate
  - 19 vertebrate
  - 2617023 total genes
- 1968858 genes in PANTHER™ families with phylogenetic trees, multiple sequence alignments and HMMs
  - 15693 PANTHER™ families
  - 125138 subfamilies
  - 177 pathways
  - 3092 pathway components
  - 51031 sequences associated to pathways
  - 5996 references captured for the pathways
- PANTHER17.0 is indexed by PANTHER GO slim and an updated PANTHER Protein Class. PANTHER GO slim is based on Gene Ontology phylogenetic annotations to over 8000 PANTHER™ families. The GO slim ontology contains:
  - 3420 total terms
  - 2282 biological process terms
  - 538 cellular component terms
  - 600 molecular function terms
- PANTHER™ Protein Class contains a total of 210 terms.

PANTHER

PANTHER GENOME INFORMATION

Organism:	Homo sapiens
Version:	Reference Proteome 2022_02
Common Name:	human
Short Name:	HUMAN
Taxonomy Id:	9606
Total number of genes in genome:	20592
Genes assigned to PANTHER families	19446
Genes with Molecular Function annotations:	11161
Genes with Biological Process annotations:	12290
Genes with Cellular Component annotations:	11888
Genes with Protein Class annotations:	13983
Genes with Pathway annotations:	2597

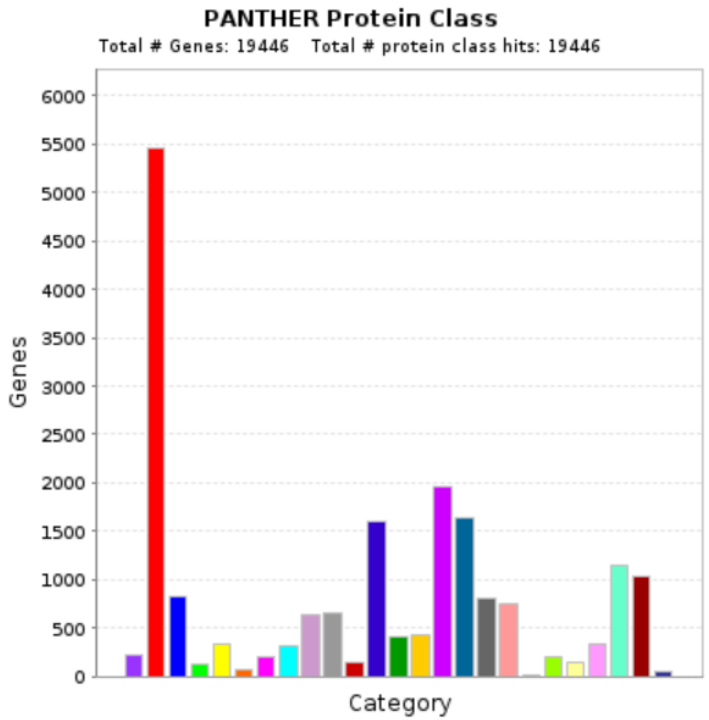
PANTHER

PANTHER CLASSIFICATION DETAILS

Molecular Function | Biological Process | Cellular Component | Protein Class | PANTHER Pathways

Category	Count
DNA metabolism protein (225 genes)	
↳DNA glycosylase (9 genes)	
↳DNA helicase (22 genes)	
↳DNA ligase (2 genes)	
↳DNA methyltransferase (6 genes)	
↳DNA photolyase (2 genes)	
↳DNA polymerase processivity factor (1 gene)	
↳DNA topoisomerase (2 genes)	
↳DNA-directed DNA polymerase (16 genes)	
↳damaged DNA-binding protein (17 genes)	
↳endodeoxyribonuclease (14 genes)	
↳exodeoxyribonuclease (6 genes)	
↳primase (2 genes)	
↳replication origin binding protein (9 genes)	
RNA metabolism protein (817 genes)	
↳DNA-directed RNA polymerase (39 genes)	

PANTHER



- DNA metabolism protein (PC00009)
- No PANTHER category is assigned (UNCLASSIFIED)
- RNA metabolism protein (PC00031)
- calcium-binding protein (PC00060)
- cell adhesion molecule (PC00069)
- cell junction protein (PC00070)
- chaperone (PC00072)
- chromatin/chromatin-binding, or -regulatory protein (PC00077)
- cytoskeletal protein (PC00085)
- defense/immunity protein (PC00090)
- extracellular matrix protein (PC00102)
- gene-specific transcriptional regulator (PC00264)
- intercellular signal molecule (PC00207)
- membrane traffic protein (PC00150)
- metabolite interconversion enzyme (PC00262)
- protein modifying enzyme (PC00260)
- protein-binding activity modulator (PC00095)
- scaffold/adaptor protein (PC00226)
- storage protein (PC00210)
- structural protein (PC00211)
- transfer/carrier protein (PC00219)
- translational protein (PC00263)
- transmembrane signal receptor (PC00197)
- transporter (PC00227)
- viral or transposable element protein (PC00237)

Упражнение

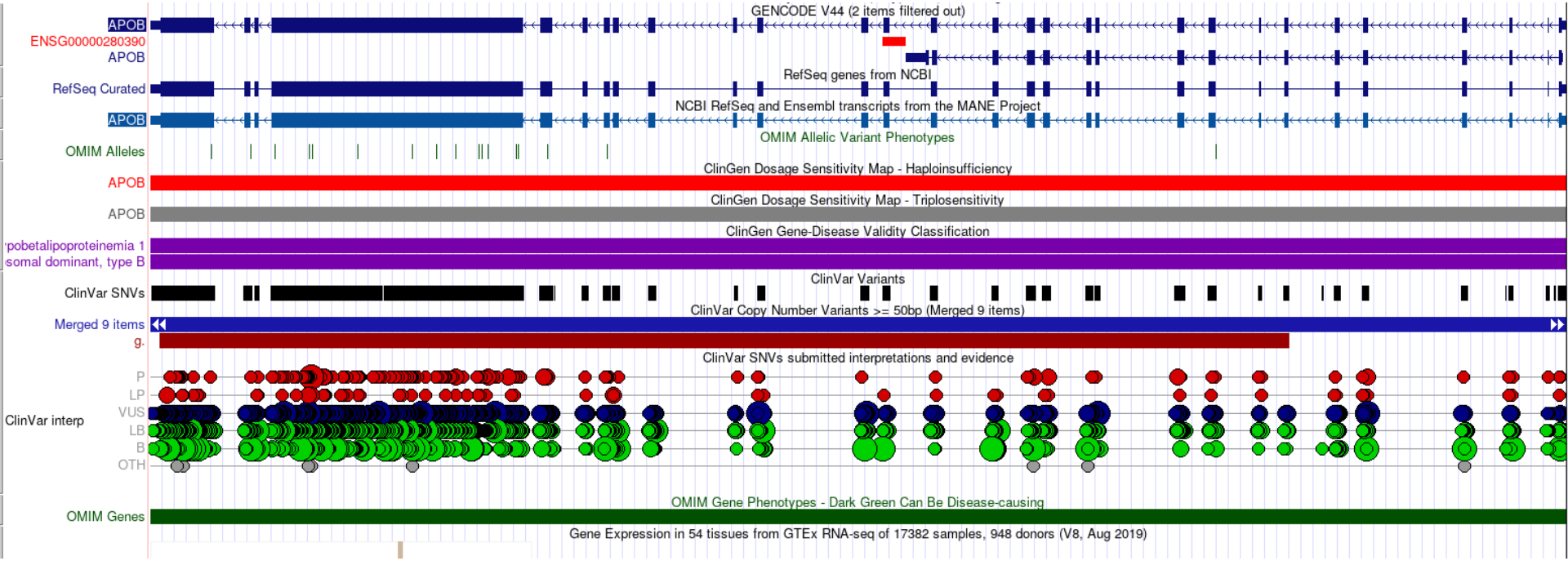
- С помощью возможностей базы Panther.db охарактеризуйте свой белок (APOB)
- Обратите внимание, что при поиске необходимо указать нужный организм

Упражнение

- Воспользуйтесь геномным браузером
- Найдите там ген APOB
- Оставьте только треки, которые несут клинически интересную информацию

Упражнение

- Примерное решение для гена

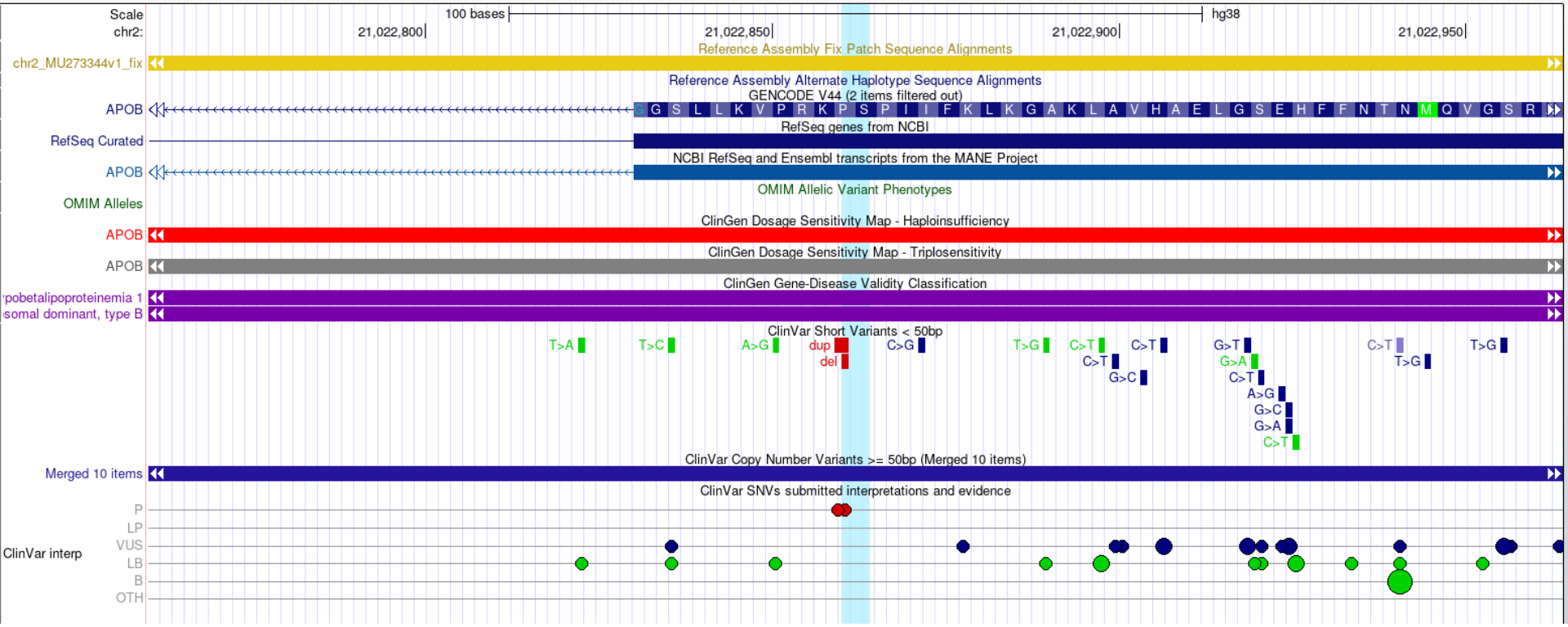


Упражнение

- Воспользуйтесь геномным браузером
- Найдите там вариант, который мы изучали (rs1553385404)
- Оставьте только треки, которые несут клинически интересную информацию

Упражнение

- Примерное решение для варианта

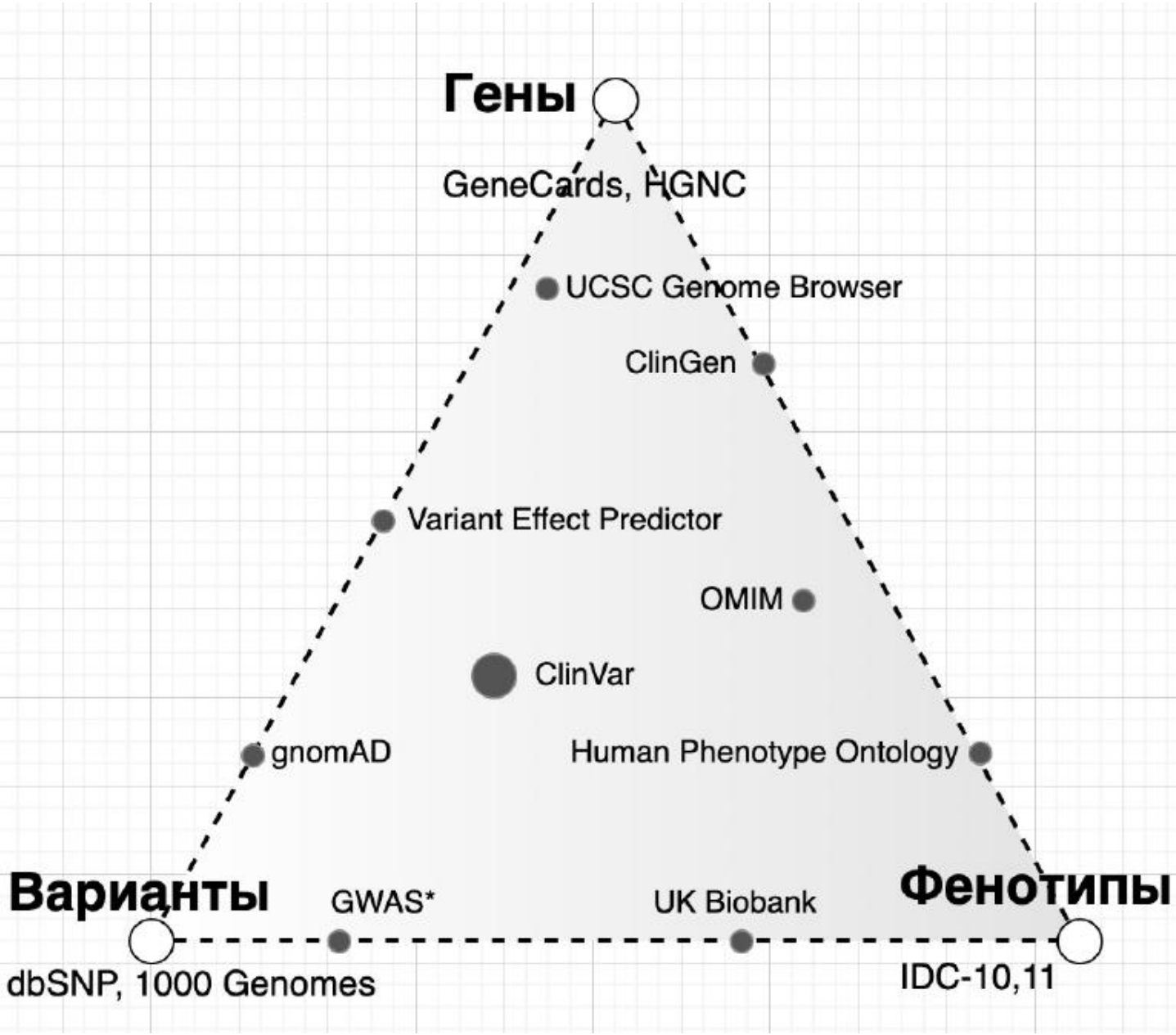


Упражнение

- В gnomAD есть возможность увидеть варианты из базы ClinVar
- Выведите только патогенные pLoF варианты
- Сколько таких вариантов удалось найти?

Упражнение

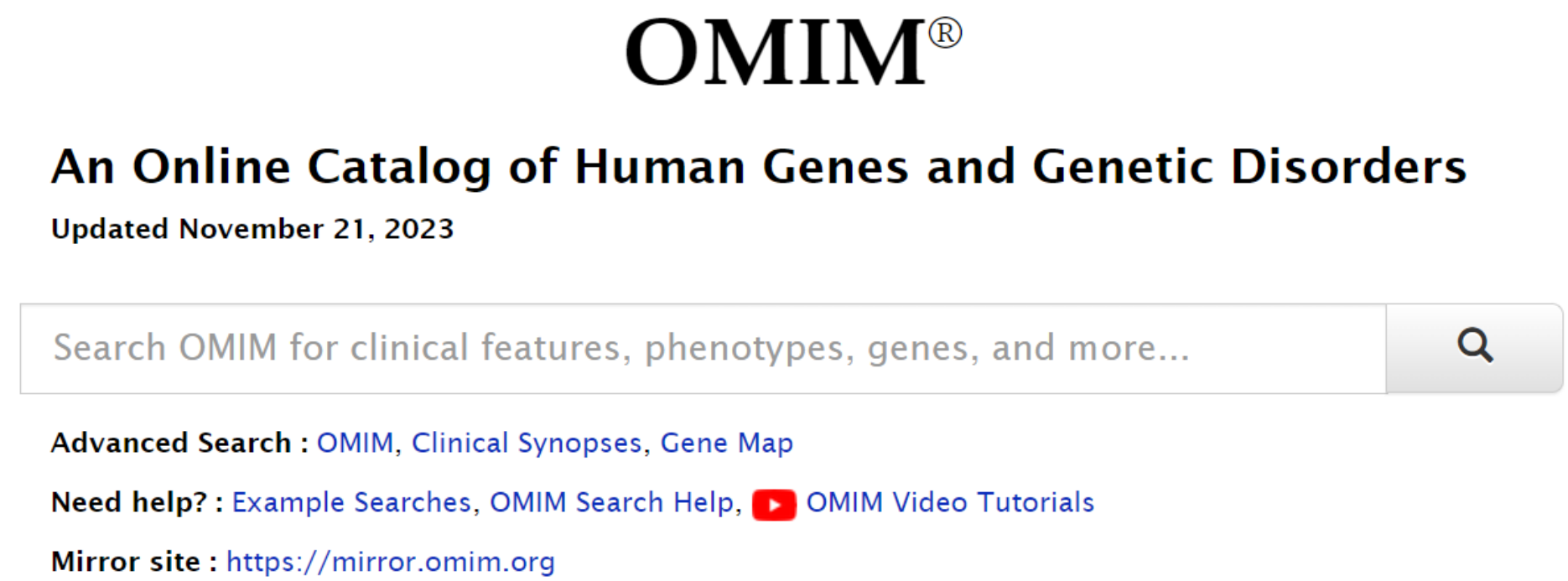
- Воспользуйтесь любой базой данных или web-сервисом
- Найдите для каждого варианта из типов:
  - Stop-gain
  - Synonymous
  - Missense
  - Splice-site
  - Frameshift indel
- Охарактеризуйте эти варианты
  - По координатам
  - В какой ген попал вариант
  - С какой болезнью ассоциирован вариант


Гены
GeneCards, HGNC
UCSC Genome Browser
ClinGen
Variant Effect Predictor
OMIM
ClinVar
gnomAD
Human Phenotype Ontology
Варианты
GWAS*
UK Biobank
Фенотипы
dbSNP, 1000 Genomes
IDC-10,11

OMIM

Online Mendelian Inheritance in Man (**OMIM**)

https://www.omim.org/



Коды MIM

Диапазон кода заболевания зависит от типа наследования:

- 100000—299999 — аутосомные заболевания (создано до 15 мая 1994 года);
- 300000—399999 — X-сцепленные заболевания;
- 400000—499999 — Y-сцепленные заболевания;
- 500000—599999 — Митохондриальные заболевания;
- 600000 и выше — аутосомные заболевания (создано после 15 мая 1994 года).

wiki

OMIM

OMIM Entry Statistics

Number of Entries in OMIM (Updated November 21st, 2023) :

MIM Number Prefix	Autosomal	X Linked	Y Linked	Mitochondrial	Totals
Gene description *	16,285	769	51	37	17,142
Gene and phenotype, combined +	21	0	0	0	21
Phenotype description, molecular basis known #	6,330	381	5	34	6,750
Phenotype description or locus, molecular basis unknown %	1,391	112	4	0	1,507
Other, mainly phenotypes with suspected mendelian basis	1,640	100	3	0	1,743
Totals	25,667	1,362	63	71	27,163

OMIM

OMIM Morbid Map Scorecard (Updated November 21st, 2023) :

Total number of phenotypes* for which the molecular basis is known	7,450
Total number of genes with phenotype-causing mutation	4,859

* Phenotypes include (1) single-gene mendelian disorders and traits; (2) susceptibilities to cancer and complex disease (e.g., BRCA1 and familial breast-ovarian cancer susceptibility, 113705.0001, and CFH and macular degeneration, 134370.0008); (3) variations that lead to abnormal but benign laboratory test values ("nondiseases") and blood groups (e.g., lactate dehydrogenase B deficiency, 150100.0001 and ABO blood group system, 110300.0001); and (4) select somatic cell genetic disease (e.g., GNAS and McCune-Albright syndrome, 139320.0008 and IDH1 and glioblastoma multiforme, 147700.0001.)

OMIM

Distribution of Phenotypes across Genes (Updated November 21st, 2023) :

Number of genes with 1 phenotype	3,410
Number of genes with 2 phenotypes	880
Number of genes with 3 phenotypes	316
Number of genes with 4+ phenotypes	253

OMIM

Dissected OMIM Morbid Map Scorecard (Updated November 21st, 2023) :

Class of phenotype	Phenotype	Gene *
Single gene disorders and traits	6,392	4,495
Susceptibility to complex disease or infection	677	500
"Nondiseases"	151	118
Somatic cell genetic disease	237	131

*Some genes may be counted more than once because mutations in a gene may cause more than one phenotype and the phenotypes may be of different classes (e.g., activating somatic BRAF mutation underlying cancer, 164757.0001. and germline BRAF mutation in Noonan syndrome, 164757.0022.)

OMIM

OMIM Update List

Updates since the database was placed on the web in December 1995

2023	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
New	54	38	41	53	48	39	20	42	32	37	37
Updated	422	257	531	415	310	380	415	449	316	411	240

2022	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
New	50	42	49	51	44	41	45	44	42	44	40	35
Updated	434	484	634	641	461	447	383	359	479	383	450	339

2021	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
New	30	44	44	50	53	52	52	53	39	50	41	46
Updated	443	469	642	508	372	433	403	593	371	359	355	411

OMIM

OMIM Update List for November 2023

November 21st, 2023

New Entries:

#620610 OOCYTE/ZYGOTE/EMBRYO MATURATION ARREST 21; OZEMA21
#620629 OPTIC ATROPHY 16; OPA16
*620630 TRANSMEMBRANE PROTEIN 170A; TMEM170A

New Clinical Synopses:

#620603 IMMUNODEFICIENCY 114, FOLATE-RESPONSIVE; IMD114
#620609 LONG-OLSEN SYNDROME; LNGOS

Updated Entries:

#145001 HYPERPARATHYROIDISM 2 WITH JAW TUMORS; HRPT2
#146255 HYPOPARATHYROIDISM, SENSORINEURAL DEAFNESS, AND RENAL DYSPLASIA SYNDROME; HDRS
#165500 OPTIC ATROPHY 1; OPA1
#165510 OPTIC ATROPHY 13 WITH RETINAL AND FOVEAL ABNORMALITIES; OPA13
*600424 SOLUTE CARRIER FAMILY 19 (FOLATE TRANSPORTER), MEMBER 1; SLC19A1
*603078 CHECKPOINT KINASE 1; CHEK1
#606159 NEURODEGENERATION WITH BRAIN IRON ACCUMULATION 3; NBIA3
#606593 LIG4 SYNDROME
#606612 MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH OR WITHOUT IMPAIRED INTELLECTUAL DEVELOPMENT), TYPE B, 5; MDDGB5

OMIM

- Поиск можно начинать с
  - Гена
  - Заболевания
  - Варианта (rs121918383) (могут быть не все варианты)

Упражнение

- Найдите в базе данных OMIM заболевание hypercholesterolemia

OMIM

hypercholesterolemia | Options | View Results as: Gene Map Table | Clinical Synopsis

Display: Highlights

Search: 'hypercholesterolemia '

Results: 137 entries. Show 100 | Download As | « First | ‹ Previous | Next › | Last »

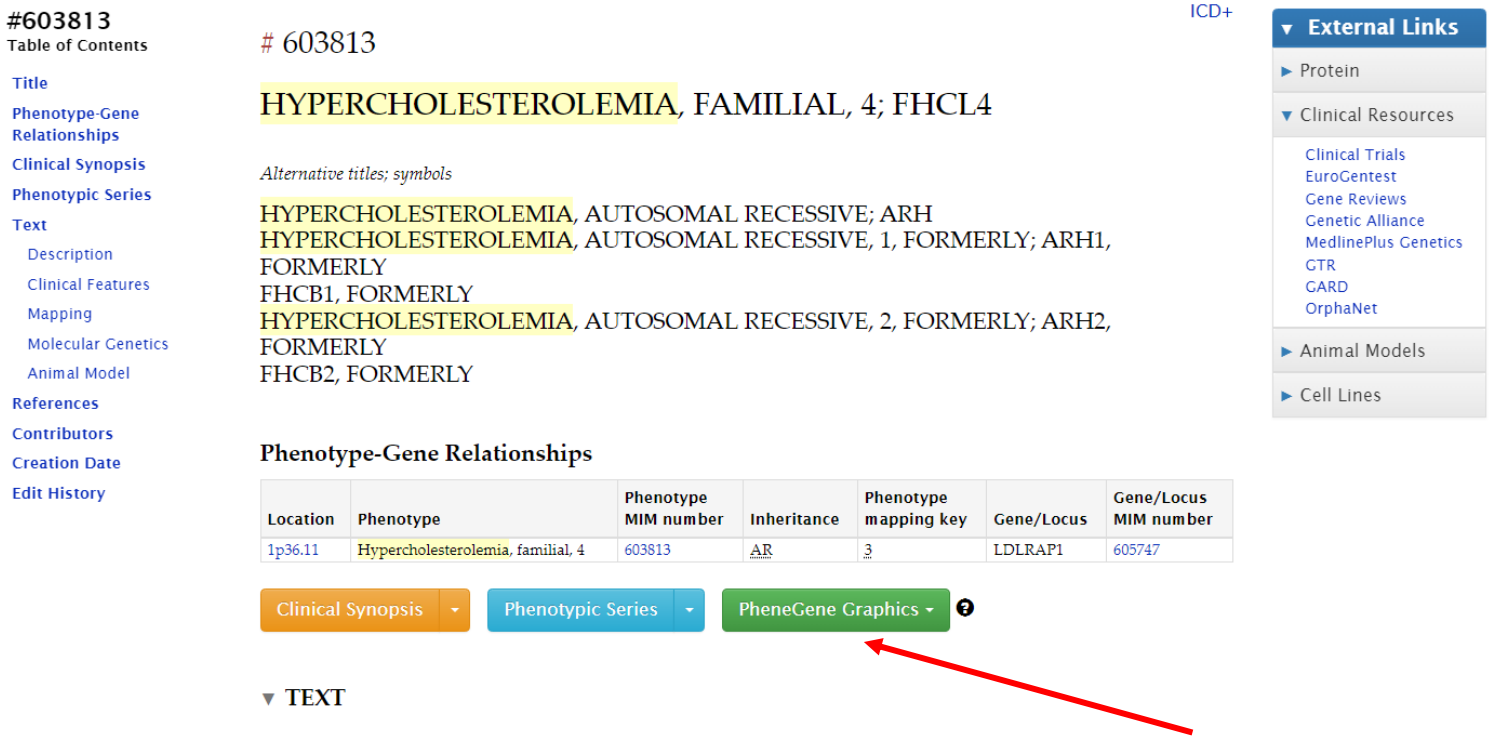
1: # 603813. HYPERCHOLESTEROLEMIA, FAMILIAL, 4; FHCL4
Cytogenetic location: 1p36.11
Matching terms: (hypercholesterolaemia | hypercholesterolemia)
► Phenotype-Gene Relationships ► Phenotypic Series ► ICD+ ► Links

2: # 144010. HYPERCHOLESTEROLEMIA, FAMILIAL, 2; FHCL2
Cytogenetic location: 2p24.1
Matching terms: (hypercholesterolaemia | hypercholesterolemia)
► Phenotype-Gene Relationships ► Phenotypic Series ► ICD+ ► Links

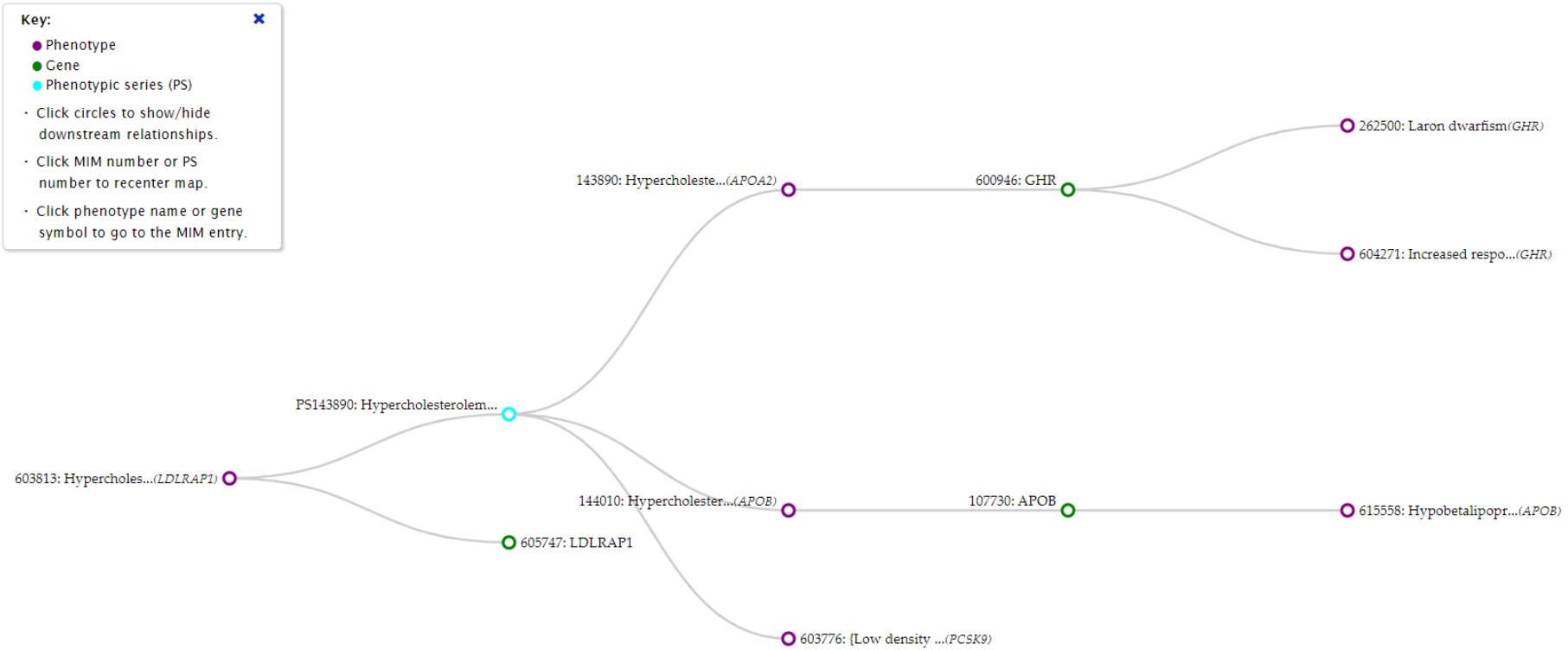
3: # 603776. HYPERCHOLESTEROLEMIA, FAMILIAL, 3; FHCL3
LOW DENSITY LIPOPROTEIN CHOLESTEROL LEVEL QUANTITATIVE TRAIT LOCUS 1, INCLUDED; LDLCQ1, INCLUDED
Cytogenetic locations: 1p32.3,
Matching terms: (hypercholesterolaemia | hypercholesterolemia)
► Phenotype-Gene Relationships ► Phenotypic Series ► Links

OMIM

- Для фенотипа текстовое (!) описание



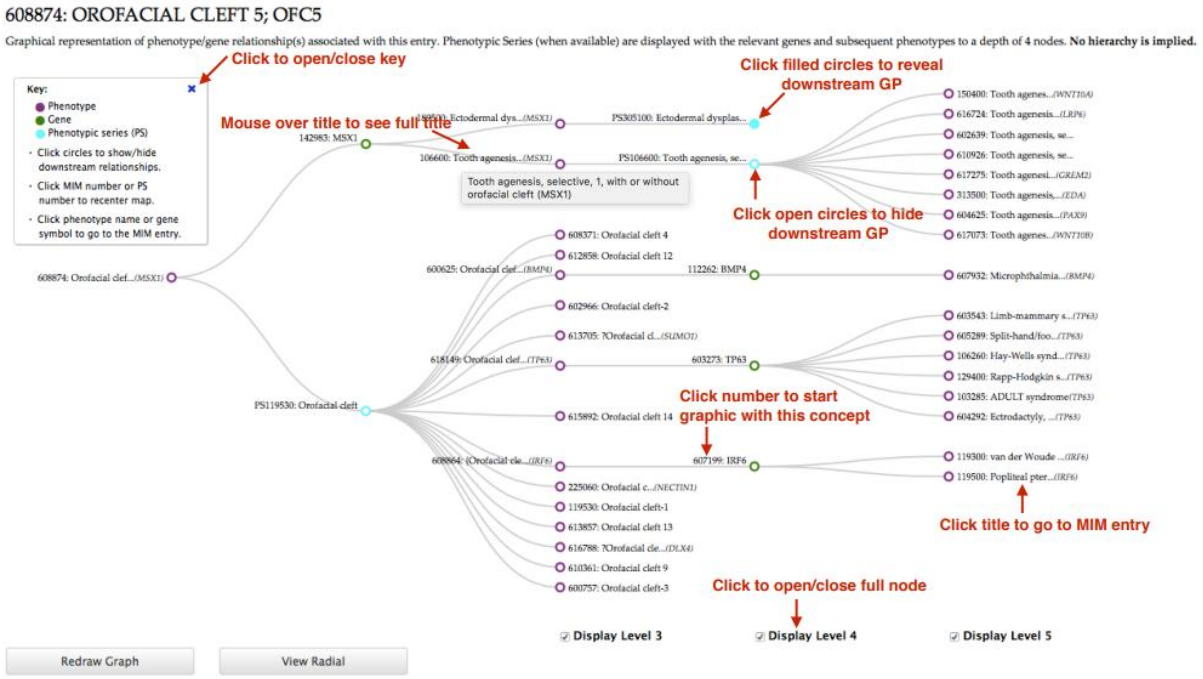
OMIM


Key:
Phenotype
Gene
Phenotypic series (PS)
· Click circles to show/hide downstream relationships.
· Click MIM number or PS number to recenter map.
· Click phenotype name or gene symbol to go to the MIM entry.
603813: Hypercholes...(LDLRAP1)
PS143890: Hypercholesterolem...
143890: Hypercholeste...(APOA2)
600946: GHR
262500: Laron dwarfism(GHR)
604271: Increased respo...(GHR)
144010: Hypercholester...(APOB)
107730: APOB
615558: Hypobetalipopr...(APOB)
605747: LDLRAP1
603776: {Low density ...(PCSK9)

OMIM

- https://www.omim.org/static/omim/pdf/OMIM_graphics.pdf

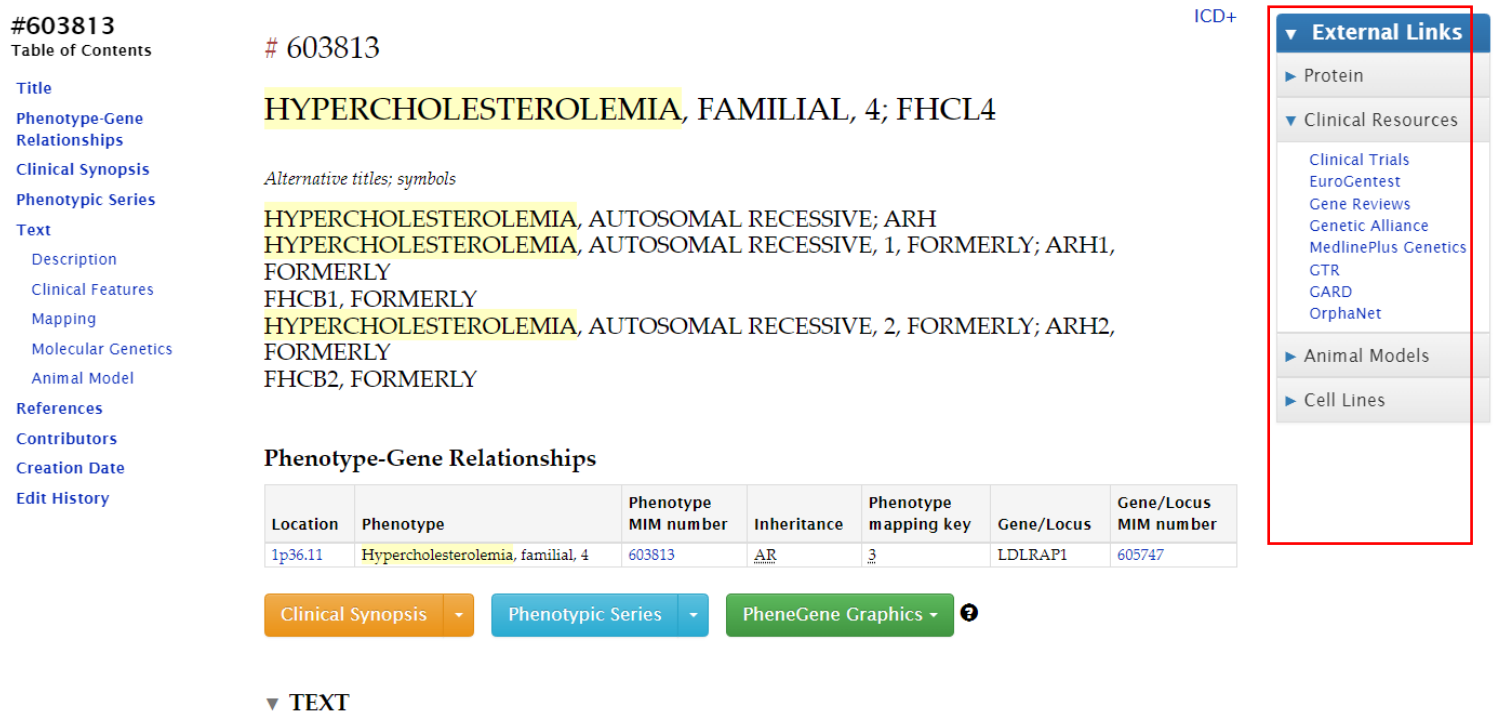
OMIM graphical views of phenotype-gene relationships



Linear graphic

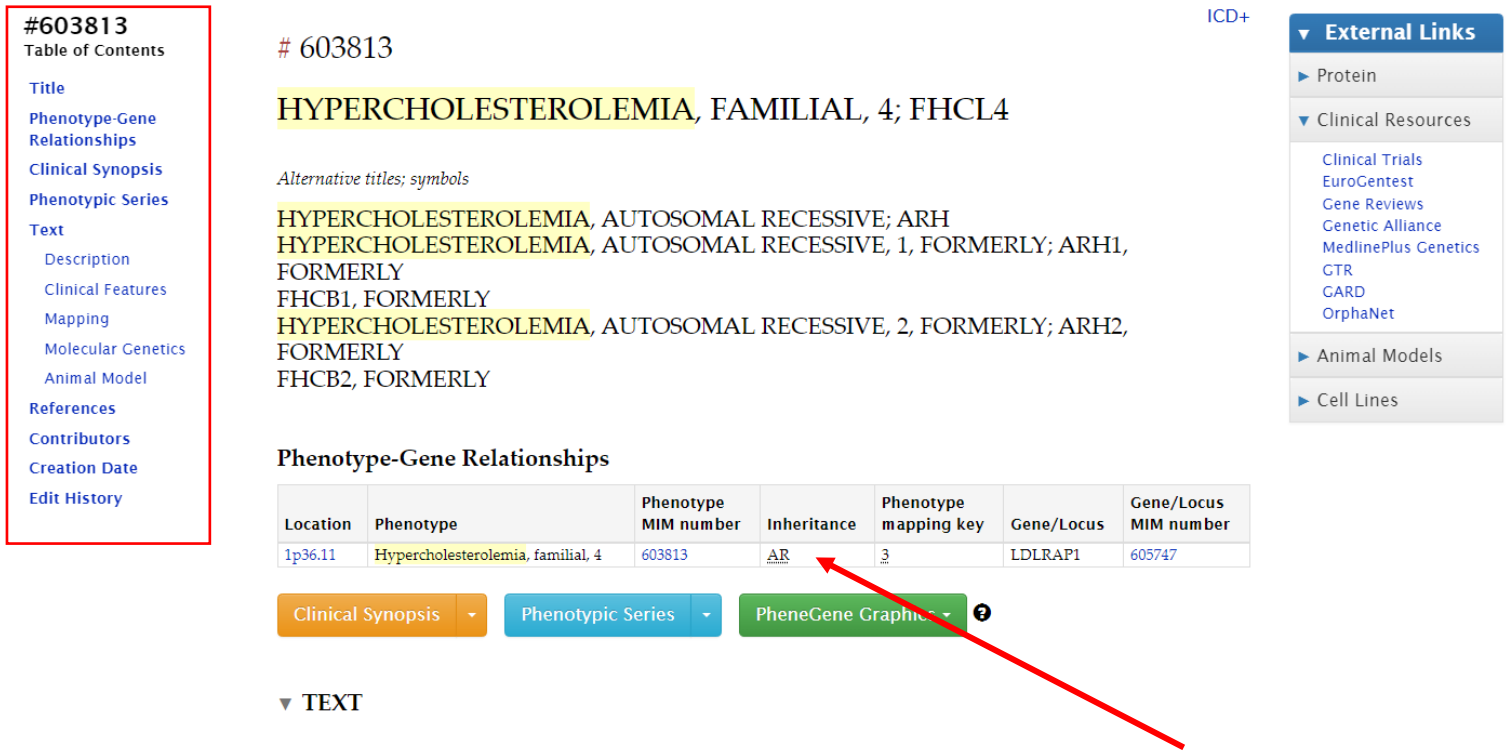
OMIM

- External Links
- Очень много перекрестных ссылок на ресурсы



OMIM

- Для фенотипа текстовое (!) описание



OMIM

Phenotype Mapping Key
1 - The disorder is placed on the map due to its association with a gene, but the underlying defect is not known.
2 - The disorder was placed on the map by statistical methods.
3 - The molecular basis of the disorder is known.
4 - A contiguous gene duplication or deletion syndrome in which multiple genes are involved.

OMIM

- Описание фенотипа

▼ **TEXT**

A number sign (#) is used with this entry because autosomal recessive familial hypercholesterolemia-4 (FHCL4) is caused by homozygous or compound heterozygous mutation in the ARH gene (LDLRAP1; 605747) on chromosome 1p36.

▼ **Description**

Autosomal recessive familial hypercholesterolemia-4 (FCHL4) is a rare monogenic disease characterized by very high levels of low-density lipoprotein (LDL) cholesterol (usually above 400 mg/dl) and increased risk of premature atherosclerotic cardiovascular disease (summary by Sanchez-Hernandez et al., 2018).

OMIM

- Посмотрите, как устроен раздел Clinical Features
- Описания клинических случаев с ссылками на источники

Clinical Features

Zuliani et al. (1995) described a consanguineous Sardinian family in which a brother and sister had a severe form of hypercholesterolemia with the clinical features of familial hypercholesterolemia (FH; 143890) homozygotes, including severely elevated plasma low density lipoprotein (LDL) cholesterol, tuberous and tendon xanthomata, and premature atherosclerosis. However, LDL receptor (LDLR; 606945) activity measured in skin fibroblasts was normal, as was LDL binding ability. Haplotype segregation analysis excluded involvement of the LDLR and apolipoprotein B (APOB; 107730) genes in the pathogenesis of the disorder. Consanguinity, absence of vertical transmission, and bimodal distribution of plasma cholesterol levels in the kindred were consistent with autosomal recessive inheritance. Sitosterolemia (see 210250) and pseudohomozygous hyperlipidemia (see 144250) were ruled out.

OMIM

- Выявление геномного локуса, ассоциированного с фенотипом

Mapping

Eden et al. (2001) performed a genomewide scan with polymorphic genetic markers in the 2 families reported by Norman et al. (1999). In both pedigrees, a single region of approximately 12 cM on 1p36-p35, designated FHCB2, fulfilled the criteria for homozygous inheritance of alleles in the affected offspring but not their unaffected sibs. The combined lod score was 5.3 in these unrelated families.

Using 4 ARH families, including 2 previously studied by Zuliani et al. (1995, 1999), Garcia et al. (2001) mapped the ARH locus to a 1-cM interval on chromosome 1p35 extending from D1S1152 to D1S2885. Garcia et al. (2001) identified 6 mutations in a gene encoding a putative adaptor protein (LDLRAP1; 605747) mapping to this region. They found no linkage to 15q25-q26, the locus that Ciccarese et al. (2000) had found to be associated with ARH using one of the same families.

OMIM

- Исследования молекулярно-генетических аспектов фенотипа

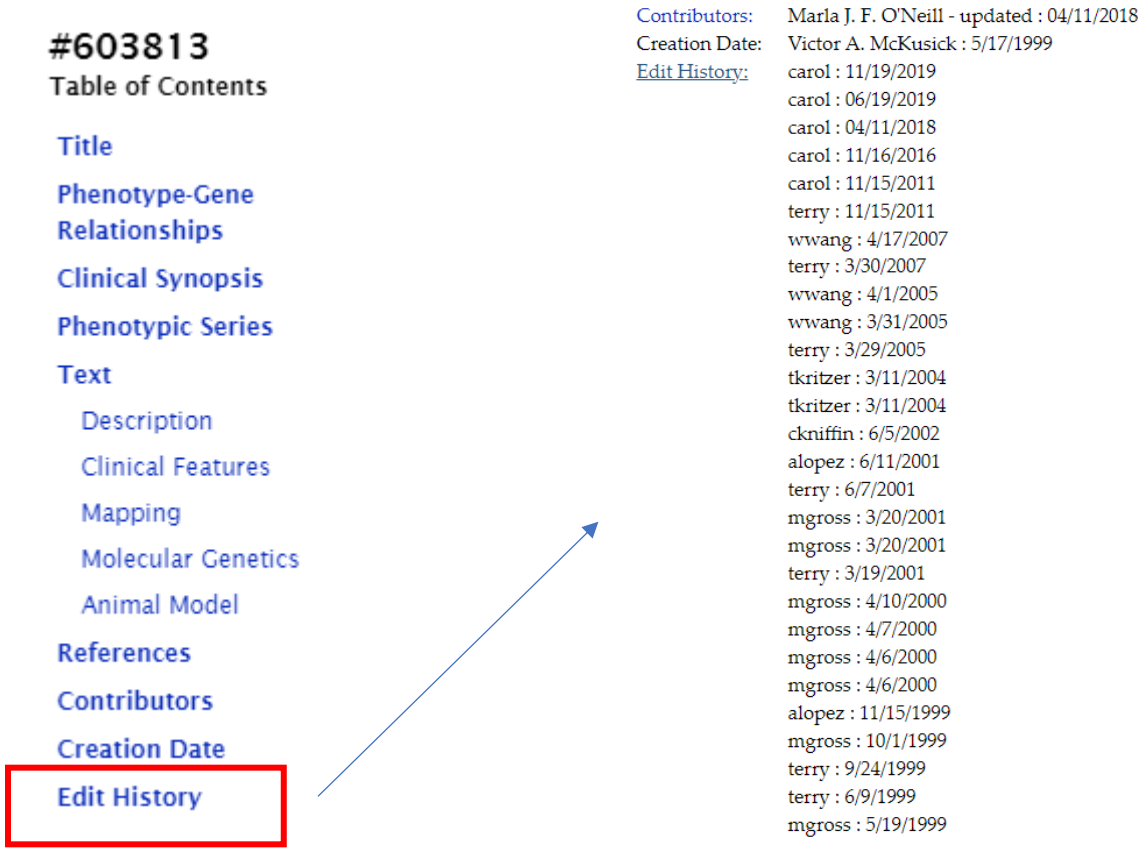
Molecular Genetics

In affected individuals from 6 families with autosomal recessive hypercholesterolemia, including the 2 Sardinian families originally reported by Zuliani et al. (1995) and Zuliani et al. (1999) and a Lebanese family previously described by Khachadurian and Uthman (1973), as well as another Lebanese family, an Iranian family, and an American family, Garcia et al. (2001) identified homozygous mutations in the ARH gene (LDLRAP1; see 605747.0001-605747.0006). The nonsense mutation (W22X; 605747.0001) and 1-bp insertion (605747.0002) that were detected in the 2 original Sardinian families were also identified in homozygosity or compound heterozygosity in 10 additional unrelated Sardinian ARH probands, and neither mutation was found in 50 normolipidemic Sardinians. The authors suggested that the finding of 2 mutations accounting for ARH in 12 Sardinian families represented genetic drift on the island of Sardinia.

Arca et al. (2002) screened the entire coding sequence of LDLRAP1 in 40 unrelated individuals from around the world who had hypercholesterolemia and at least 1 normocholesterolemic parent. They identified 4 Italian probands who were homozygous for the same 1-bp insertion (605747.0002) that had previously been identified in Sardinian patients. No mutations were identified in the other 36 probands.

OMIM

- Информация пополняется новыми исследованиями, наблюдениями и т.п.



#603813
Table of Contents

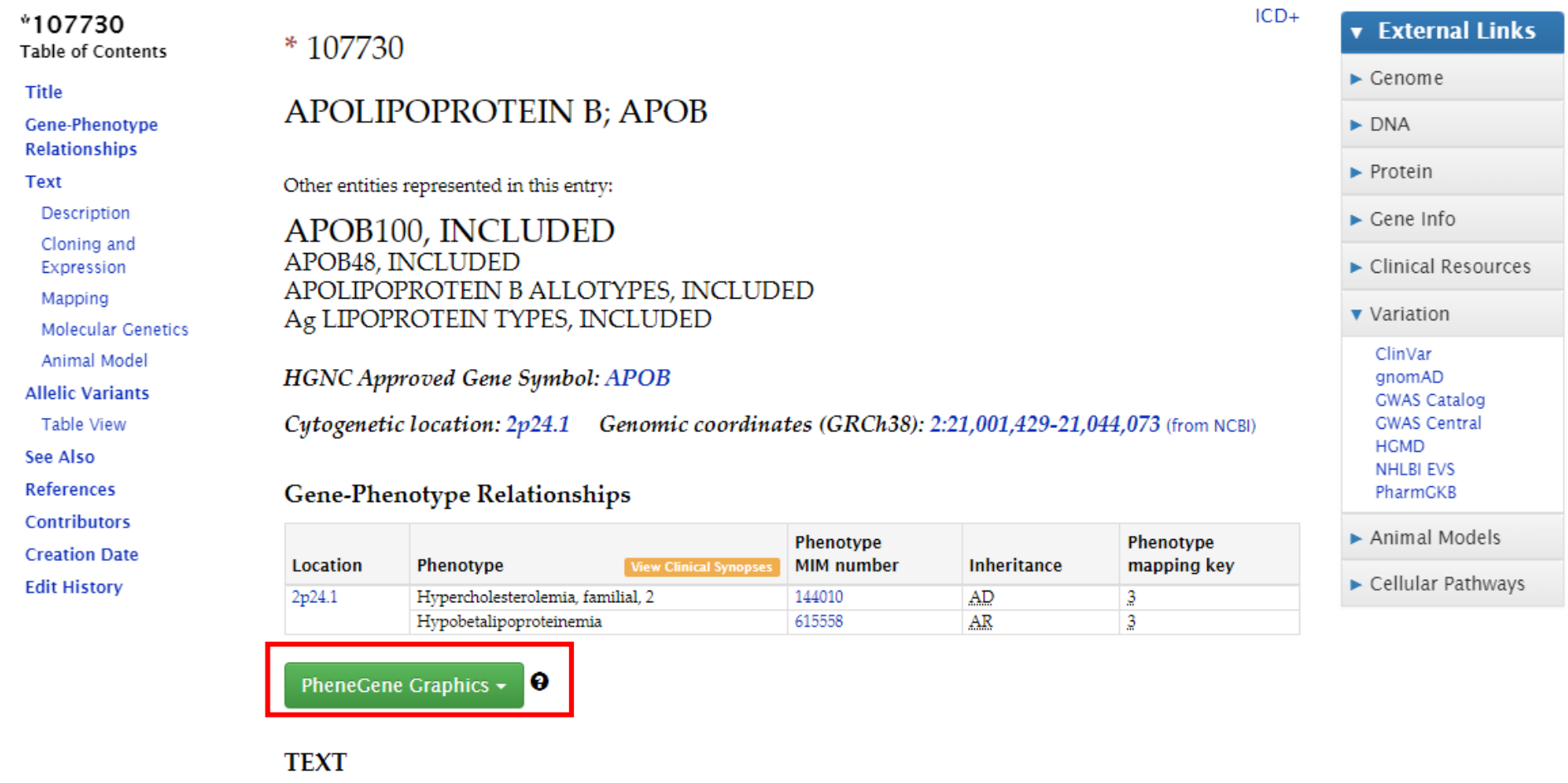
- Title
- Phenotype-Gene Relationships
- Clinical Synopsis
- Phenotypic Series
- Text
  - Description
  - Clinical Features
  - Mapping
  - Molecular Genetics
  - Animal Model
- References
- Contributors
- Creation Date
- Edit History

Contributors: Marla J. F. O'Neill - updated : 04/11/2018
Creation Date: Victor A. McKusick : 5/17/1999
Edit History:
carol : 11/19/2019
carol : 06/19/2019
carol : 04/11/2018
carol : 11/16/2016
carol : 11/15/2011
terry : 11/15/2011
wwang : 4/17/2007
terry : 3/30/2007
wwang : 4/1/2005
wwang : 3/31/2005
terry : 3/29/2005
tkritzer : 3/11/2004
tkritzer : 3/11/2004
ckniffin : 6/5/2002
alopez : 6/11/2001
terry : 6/7/2001
mgross : 3/20/2001
mgross : 3/20/2001
terry : 3/19/2001
mgross : 4/10/2000
mgross : 4/7/2000
mgross : 4/6/2000
mgross : 4/6/2000
alopez : 11/15/1999
mgross : 10/1/1999
terry : 9/24/1999
terry : 6/9/1999
mgross : 5/19/1999

Упражнение

- Найдите в базе данных OMIM ген APOB
- Обсудите структуру результата поиска и записи, аналогичные поиску по заболеваниям

OMIM



*107730
Table of Contents

Title
Gene-Phenotype Relationships
Text
Description
Cloning and Expression
Mapping
Molecular Genetics
Animal Model
Allelic Variants
Table View
See Also
References
Contributors
Creation Date
Edit History

* 107730

ICD+

APOLIPOPROTEIN B; APOB

Other entities represented in this entry:

APOB100, INCLUDED
APOB48, INCLUDED
APOLIPOPROTEIN B ALLOTYPES, INCLUDED
Ag LIPOPROTEIN TYPES, INCLUDED

HGNC Approved Gene Symbol: APOB

Cytogenetic location: 2p24.1 Genomic coordinates (GRCh38): 2:21,001,429-21,044,073 (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype View Clinical Synopses	Phenotype MIM number	Inheritance	Phenotype mapping key
2p24.1	Hypercholesterolemia, familial, 2	144010	AD	3
	Hypobetalipoproteinemia	615558	AR	3

PheneGene Graphics

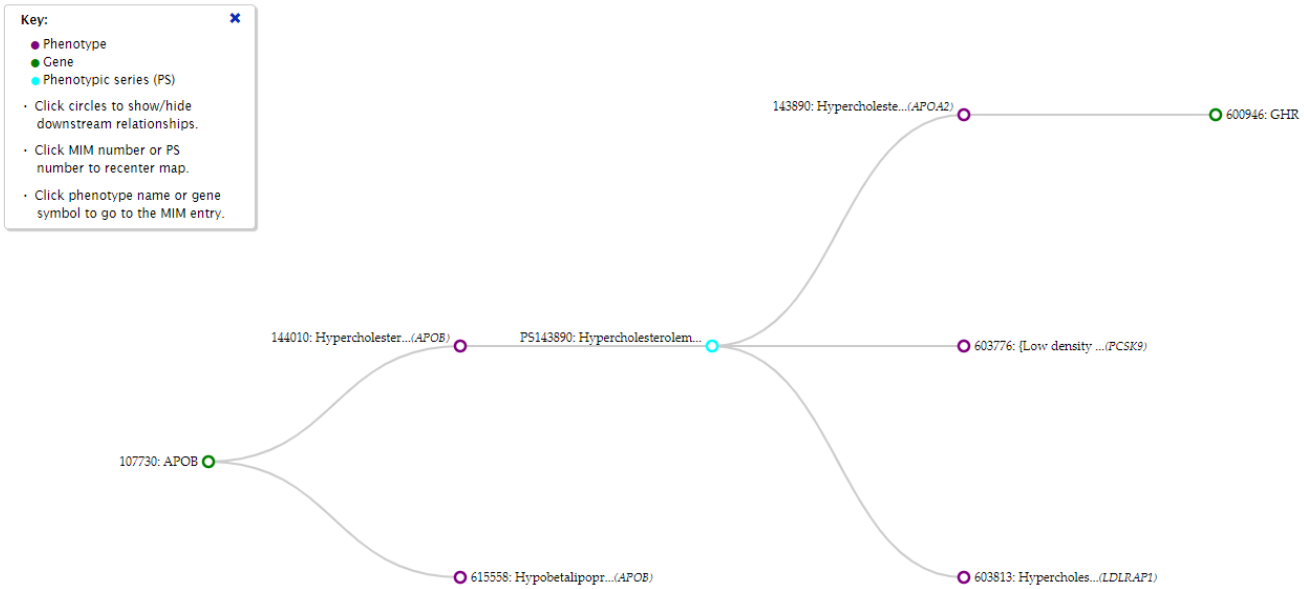
TEXT

External Links
Genome
DNA
Protein
Gene Info
Clinical Resources
Variation
ClinVar
gnomAD
GWAS Catalog
GWAS Central
HGMD
NHLBI EVS
PharmGKB
Animal Models
Cellular Pathways

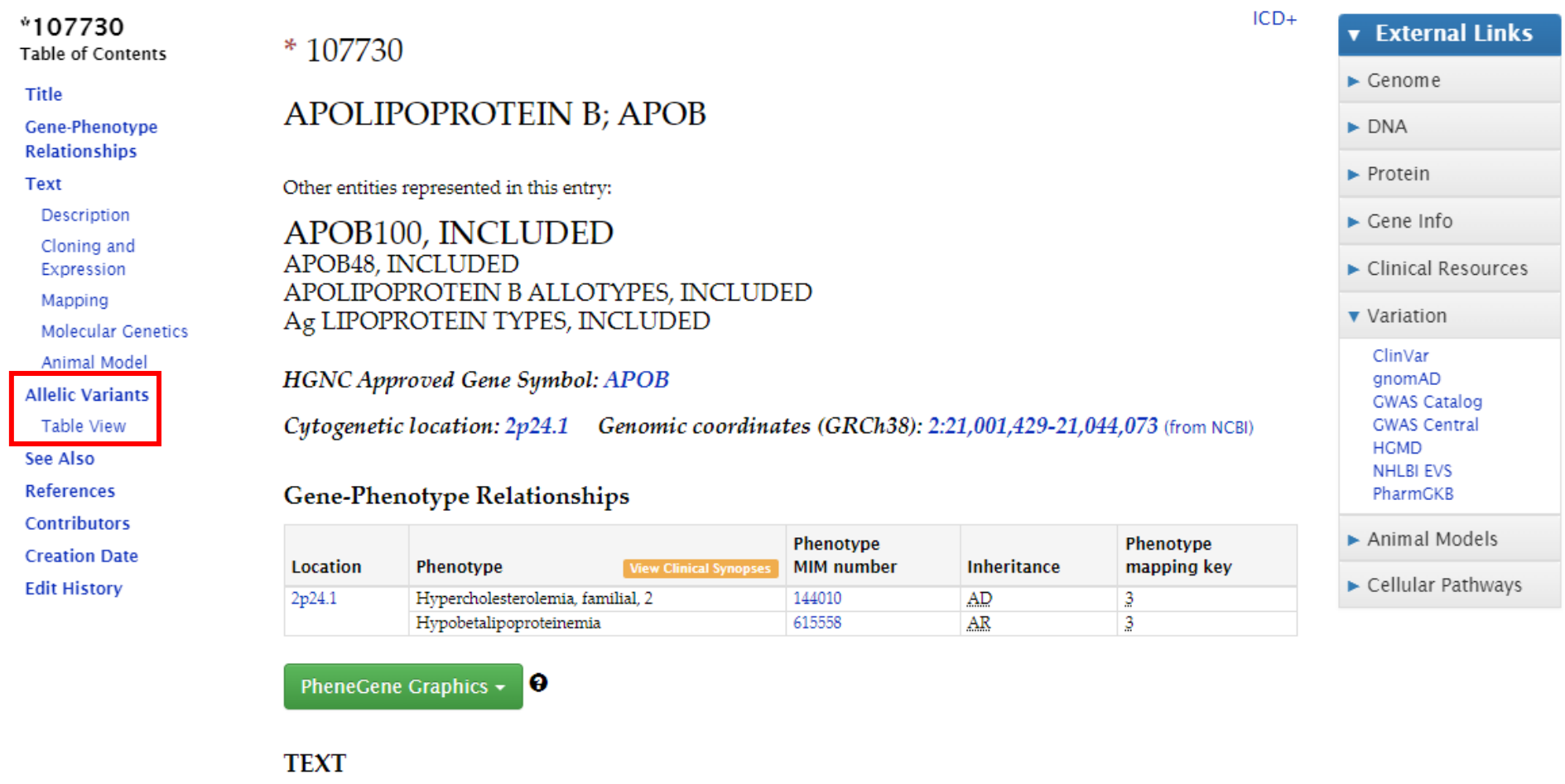
OMIM

107730:

Graphical representation of phenotype/gene relationship(s) associated with this entry. Phenotypic Series (when available) are displayed with the relevant genes and subsequent phenotypes to a depth of 4 nodes. A quick reference overview and guide (PDF). **No hierarchy is implied.** Feedback



OMIM



*107730
Table of Contents

- Title
- Gene-Phenotype Relationships
- Text
  - Description
  - Cloning and Expression
  - Mapping
  - Molecular Genetics
  - Animal Model
- Allelic Variants
  - Table View
- See Also
- References
- Contributors
- Creation Date
- Edit History

ICD+

* 107730

APOLIPOPROTEIN B; APOB

Other entities represented in this entry:

APOB100, INCLUDED
APOB48, INCLUDED
APOLIPOPROTEIN B ALLOTYPES, INCLUDED
Ag LIPOPROTEIN TYPES, INCLUDED

HGNC Approved Gene Symbol: APOB

Cytogenetic location: 2p24.1 Genomic coordinates (GRCh38): 2:21,001,429-21,044,073 (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype (View Clinical Synopses)	Phenotype MIM number	Inheritance	Phenotype mapping key
2p24.1	Hypercholesterolemia, familial, 2	144010	AD	3
	Hypobetalipoproteinemia	615558	AR	3

PheneGene Graphics

TEXT

External Links

- Genome
- DNA
- Protein
- Gene Info
- Clinical Resources
- Variation
  - ClinVar
  - gnomAD
  - GWAS Catalog
  - GWAS Central
  - HGMD
  - NHLBI EVS
  - PharmGKB
- Animal Models
- Cellular Pathways

OMIM

Каждая запись – текстовое описание варианта с ссылкой на публикацию и перекрестными ссылками на записи в OMIM (например, для фенотипа)

APOLIPOPROTEIN B; APOB

Allelic Variants (22 Selected Examples) :

All ClinVar Variants

Number	Phenotype	Mutation	SNP	gnomAD	ClinVar
.0001	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, 4-BP DEL, NT5391	rs281865425	-	RCV000019470...
.0002	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB39	APOB, 1-BP DEL, FS1799TER	rs397514255	rs397514255	RCV000019471
.0003	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, ARG1306TER	rs121918383	rs121918383	RCV000019472...
.0004	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB40	APOB, VAL1829CYS	rs121918384	rs121918384	RCV000019473...
.0005	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB90 OR APOB89	APOB, GLU4034ARG	rs121918385	-	RCV000019474
.0006	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB46	APOB, ARG2058TER	rs121918386	rs121918386	RCV000019476...
.0007	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB87	APOB, 1-BP DEL, 12032G	rs387906569	rs387906569	RCV000019477
.0008	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB31	APOB, 1-BP DEL, 1425G	rs397514256	rs397514256	RCV000019478
.0009	HYPERCHOLESTEROLEMIA, FAMILIAL, 2	APOB, ARG3500GLN	rs5742904	rs5742904	RCV000019479...
.0010	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, EX21DEL	-	-	RCV000019475
.0011	APOB POLYMORPHISM IN SIGNAL PEPTIDE	APOB, INS AND DEL	-	-	RCV000251913...
.0012	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, LEU3041TER	rs121918387	-	RCV000019481...
.0013	HYPOBETALIPOPROTEINEMIA, NORMOTRIGLYCERIDEMIC	APOB, GLN2252TER	rs121918388	rs121918388	RCV001837438
.0014	HYPOBETALIPOPROTEINEMIA, FAMILIAL, ASSOCIATED WITH APOB32	APOB, GLN1450TER	rs121918389	rs121918389	RCV000019483
.0015	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, ARG2495TER	rs121918390	rs121918390	RCV000019484...
.0016	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, 1-BP DEL, NT11840	rs587776852	rs587776852	RCV000019485...
.0017	HYPERCHOLESTEROLEMIA, FAMILIAL, 2	APOB, ARG3531CYS	rs12713559	rs12713559	RCV000019486...
.0018	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, IVS7AS, A-G, -2	rs1572800245	-	RCV000019487...
.0019	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, 1-BP DEL, 4432T	-	-	RCV000019488
.0020	HYPOBETALIPOPROTEINEMIA, NORMOTRIGLYCERIDEMIC	APOB, 4-BP DEL, NT36491	-	-	RCV001837441
.0021	HYPOBETALIPOPROTEINEMIA, NORMOTRIGLYCERIDEMIC	APOB, TYR1173TER	rs121918391	-	RCV001837442...
.0022	HYPOBETALIPOPROTEINEMIA, FAMILIAL	APOB, 2-BP INS, 825GG	rs606231236	-	RCV000032601

OMIM

- Сравнить несколько записей

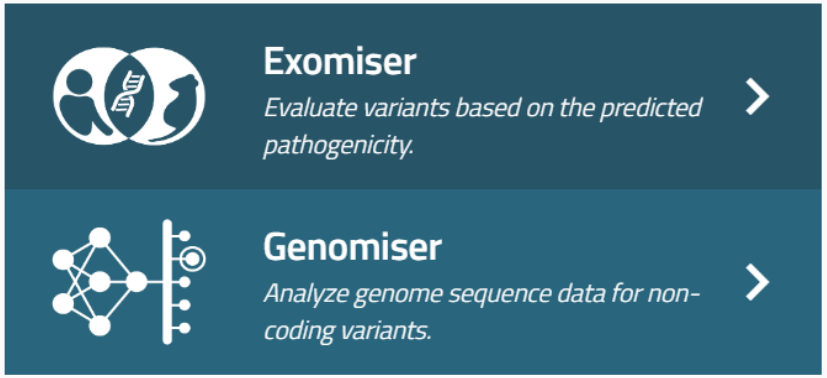
OMIM

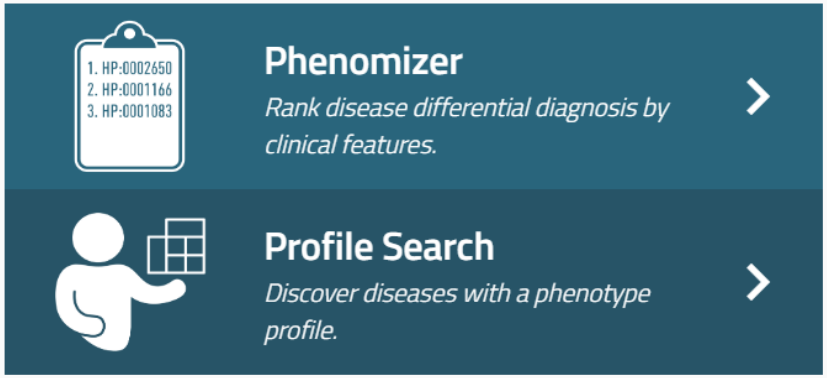
- Результат сравнения

NUMBER	# 603813 ▼	# 144010 ▼	# 603776 ▼
TITLE	HYPERCHOLESTEROLEMIA, FAMILIAL, 4; FHCL4	HYPERCHOLESTEROLEMIA, FAMILIAL, 2; FHCL2	HYPERCHOLESTEROLEMIA, FAMILIAL, 3; FHCL3
GENE	*LDLRAP1* - 605747	*APOB* - 107730	*PCSK9* - 607786
INHERITANCE (in 3/3)	- Autosomal recessive	- Autosomal dominant	- Autosomal dominant
HEAD & NECK (in 2/3) ▼		*Eyes* - Corneal arcus - Xanthelasma	*Eyes* - Arcus corneae
CARDIOVASCULAR (in 2/3) ▼		*Heart* - Coronary artery disease	*Heart* - Coronary artery disease
SKIN, NAILS, & HAIR (in 3/3) ▼	*Skin* - Xanthomas	*Skin* - Tendinous xanthomas - Planar xanthomas (in homozygotes)	*Skin* - Xanthelasmas - Tendinous xanthomata
LABORATORY ABNORMALITIES (in 3/3) ▼	- Hypertriglyceridemia - Very high low-density lipoprotein (LDL) cholesterol (>400 mg/dL) - High total cholesterol (>600 mg/dL)	- Hypercholesterolemia - Abnormal LDL	- High total cholesterol High LD cholesterol
MISCELLANEOUS (in 1/3) ▼			- Elevated cholesterol levels evident before age 20
MOLECULAR BASIS (in 3/3) ▼	- Caused by mutation in the low density lipoprotein receptor adaptor protein 1 gene (LDLRAP1, 605747.0001)	- Caused by mutation in the apolipoprotein B gene (APOB, 107730.0001)	- Caused by mutation in the proprotein convertase, subtilisin/kexin-type, 9 gene (PCSK9, 607786.0001)

HPO

- https://hpo.jax.org/app/





HPO

- Поиск от
  - Фенотипа
  - Заболевания
  - Гена

Упражнение

- Найдите в базе HPO гиперхолестеринемию по OMIM ID: 603813

HPO

Hypercholesterolemia, autosomal recessive OMIM:603813 MONDO:0011374

An autosomal recessive condition caused by mutation(s) in the LDLRAP1 gene, encoding low density lipoprotein receptor adaptor protein 1. The phenotype is similar to that of familial hypercholesterolemia, but generally considered to be a milder form of hypercholesterolemia.

Export Associations | Report Entry Issue

HPO Associations | Gene Associations

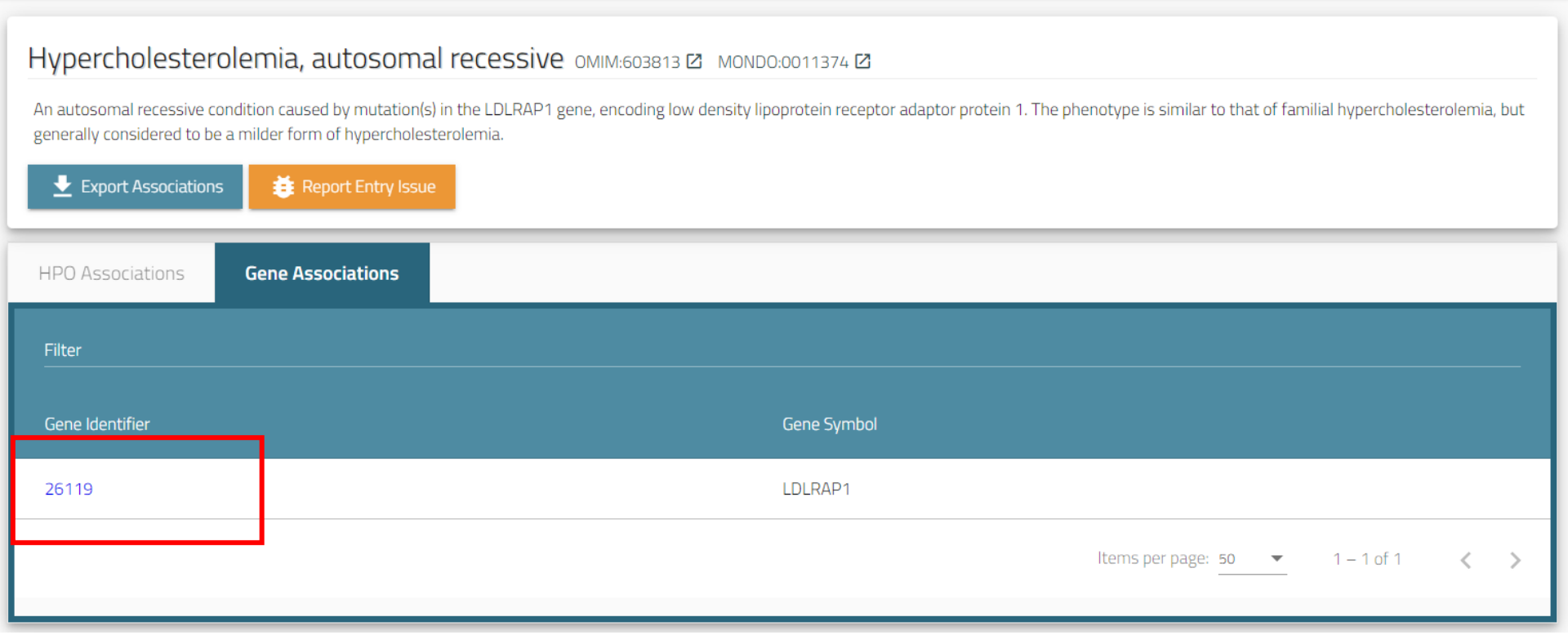
Inheritance [1 annotation]

Term Identifier	Term Name	Onset	Frequency	Source(s)
HP:0000007	Autosomal recessive inheritance			PubMed

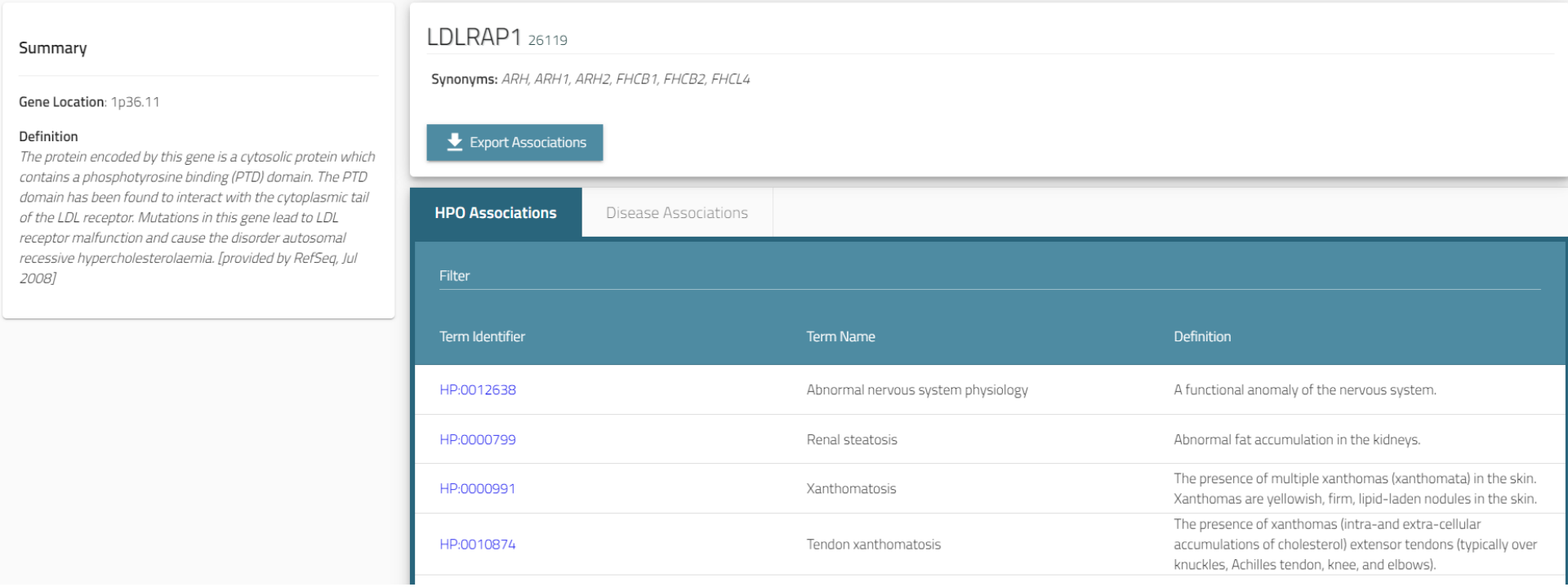
Cardiovascular [1 annotation]

Term Identifier	Term Name	Onset	Frequency	Source(s)
HP:0002621	Atherosclerosis			OMIM

HPO


Hypercholesterolemia, autosomal recessive OMIM:603813 MONDO:0011374
An autosomal recessive condition caused by mutation(s) in the LDLRAP1 gene, encoding low density lipoprotein receptor adaptor protein 1. The phenotype is similar to that of familial hypercholesterolemia, but generally considered to be a milder form of hypercholesterolemia.
Export Associations
Report Entry Issue
HPO Associations
Gene Associations
Filter
Gene Identifier
Gene Symbol
26119
LDLRAP1
Items per page: 50
1 – 1 of 1

HPO



Summary

Gene Location: 1p36.11

Definition

The protein encoded by this gene is a cytosolic protein which contains a phosphotyrosine binding (PTD) domain. The PTD domain has been found to interact with the cytoplasmic tail of the LDL receptor. Mutations in this gene lead to LDL receptor malfunction and cause the disorder autosomal recessive hypercholesterolaemia. [provided by RefSeq, Jul 2008]

LDLRAP1 26119

Synonyms: *ARH, ARH1, ARH2, FHCB1, FHCB2, FHCL4*

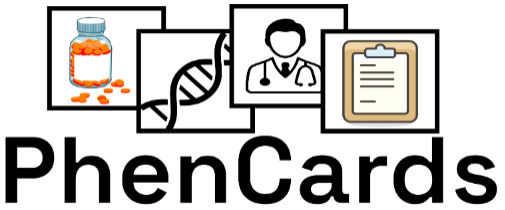
Export Associations

HPO Associations | Disease Associations

Filter

Term Identifier	Term Name	Definition
HP:0012638	Abnormal nervous system physiology	A functional anomaly of the nervous system.
HP:0000799	Renal steatosis	Abnormal fat accumulation in the kidneys.
HP:0000991	Xanthomatosis	The presence of multiple xanthomas (xanthomata) in the skin. Xanthomas are yellowish, firm, lipid-laden nodules in the skin.
HP:0010874	Tendon xanthomatosis	The presence of xanthomas (intra-and extra-cellular accumulations of cholesterol) extensor tendons (typically over knuckles, Achilles tendon, knee, and elbows).

PhenCards



- https://phencards.org/
- PhenCards is a web server for linking human phenotype information to biomedical knowledge

Упражнение

- Найдите в PhenCards ген, заболевание или фенотип
- Обратите внимание на выдачу
- Обсудите, ссылки на какие ресурсы выведены в результате поиска

Mondo

- https://www.ebi.ac.uk/ols4/ontologies/mondo
- слайды

Mondo Disease Ontology

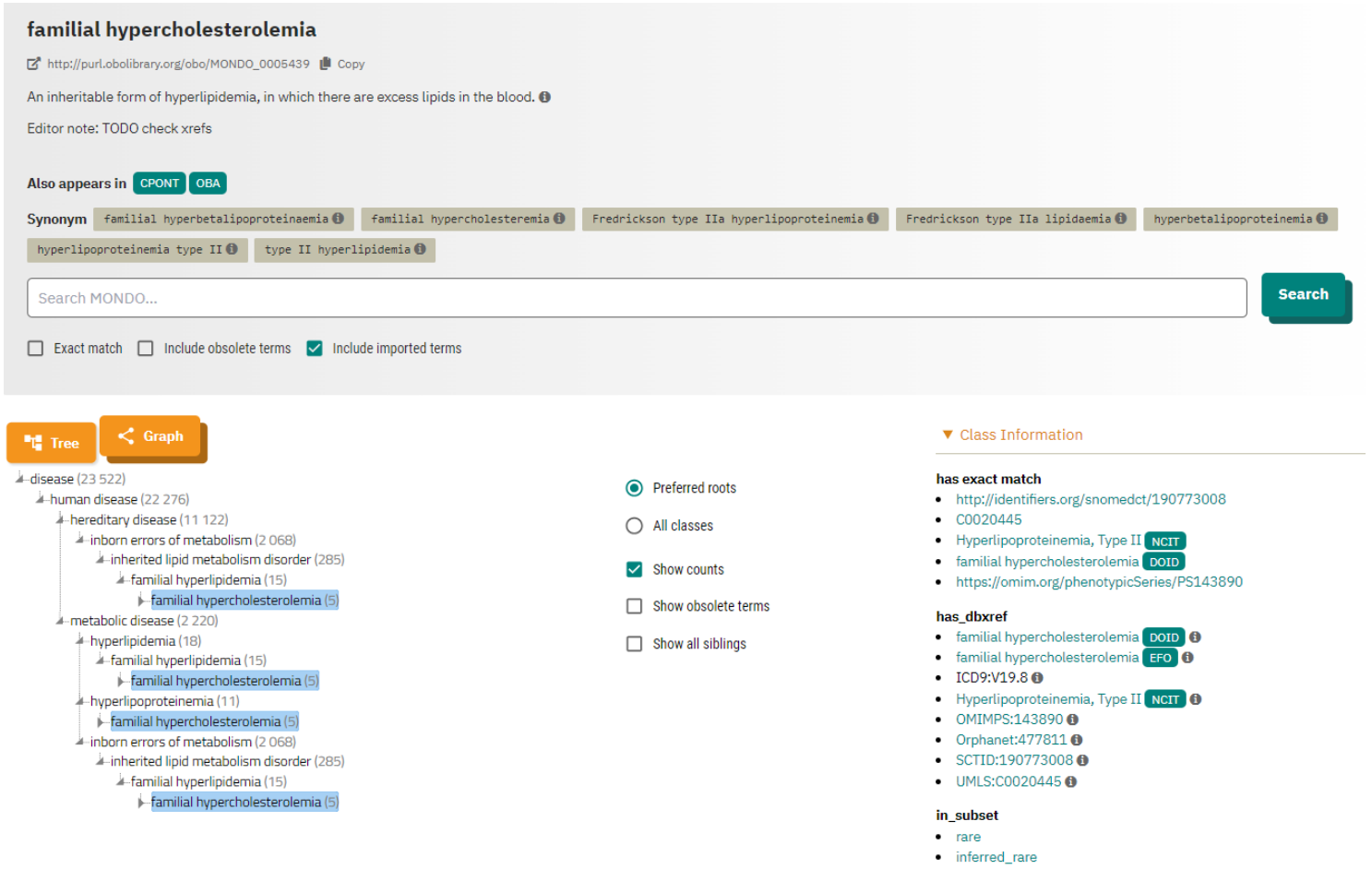
Version 2024-02-06

A semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology.

Disease term feature	Count
Total number of terms	22,157
Database cross references	104,479
Term definitions	15,443
Exact synonyms	66,247
Related synonyms	30,661
Narrow (more specific) synonyms	2,214
Broad (more general) synonyms	847

Disease type	Count (Concepts)
Rare diseases	10,443
Infectious diseases	1,240
Cancers (including neoplasms)	4,298
Mendelian diseases	11,380

Mondo


familial hypercholesterolemia
http://purl.obolibrary.org/obo/MONDO_0005439 Copy
An inheritable form of hyperlipidemia, in which there are excess lipids in the blood.
Editor note: TODO check xrefs
Also appears in CPONT OBA
Synonym familial hyperbetalipoproteinaemia familial hypercholesteremia Fredrickson type IIa hyperlipoproteinemia Fredrickson type IIa lipidaemia hyperbetalipoproteinemia hyperlipoproteinemia type II type II hyperlipidemia
Search MONDO...
Search
Exact match
Include obsolete terms
Include imported terms
Tree
Graph
disease (23 522)
human disease (22 276)
hereditary disease (11 122)
inborn errors of metabolism (2 068)
inherited lipid metabolism disorder (285)
familial hyperlipidemia (15)
familial hypercholesterolemia (5)
metabolic disease (2 220)
hyperlipidemia (18)
familial hyperlipidemia (15)
familial hypercholesterolemia (5)
hyperlipoproteinemia (11)
familial hypercholesterolemia (5)
inborn errors of metabolism (2 068)
inherited lipid metabolism disorder (285)
familial hyperlipidemia (15)
familial hypercholesterolemia (5)
Preferred roots
All classes
Show counts
Show obsolete terms
Show all siblings
Class Information
has exact match
http://identifiers.org/snomedct/190773008
C0020445
Hyperlipoproteinemia, Type II NCIT
familial hypercholesterolemia DOID
https://omim.org/phenotypicSeries/PS143890
has_dbxref
familial hypercholesterolemia DOID
familial hypercholesterolemia EFO
ICD9:V19.8
Hyperlipoproteinemia, Type II NCIT
OMIMPS:143890
Orphanet:477811
SCTID:190773008
UMLS:C0020445
in_subset
rare
inferred_rare

Mondo

Mondo uses the following ontologies as sources or for cross-references (xrefs)/alignments.

Source	ID Space/URI Prefix	Role	Website
OMIM Phenotypes	OMIM	Source	www.omim.org
OMIM Phenotypic Series	OMIMPS	Source	www.omim.org
Orphanet	Orpha	Source	https://www.orpha.net/consor/cgi-bin/index.php
SNOMED (disorder subset)	SCTID	xref/Alignments	www.snomed.org
National Cancer Institute Thesaurus (disease/disorder subset)	NCIT	Source	https://ncit.nci.nih.gov/ncitbrowser/
Genetic and Rare Diseases Information Center	GARD	Source	https://rarediseases.info.nih.gov/
Medical Subject Headings	MESH	xref/Alignments	https://id.nlm.nih.gov/mesh/
Unified Medical Language System	UMLS	xref/Alignments	https://www.nlm.nih.gov/research/umls/index.html
ICD - ICD-9 - International Classification of Diseases	ICD9	xref/Alignments	https://www.cdc.gov/nchs/icd/icd9.htm
ICD - ICD-10 - International Classification of Diseases	ICD10	xref/Alignments	https://www.cdc.gov/nchs/icd/icd10cm.htm
Experimental Factor Ontology	EFO	xref/Alignments*	https://www.ebi.ac.uk/efo/
Disease Ontology	DO	Source	http://www.obofoundry.org/ontology/doid.html
Mental Functioning Ontology	MF	Source	http://www.obofoundry.org/ontology/mf.html
MedGen	MEDGEN	xref/Alignments	https://www.ncbi.nlm.nih.gov/medgen/
Ontology for General Medical Science	OGMS	xref/Alignments	https://github.com/OGMS/ogms
Medical Dictionary for Regulatory Activities	MeDRA	xref/Alignments	https://www.meddra.org/
OncoTree	ONCOTREE	xref/Alignments	http://oncotree.mskcc.org/#/home