Варианты в геноме человека

Василий Евгеньевич Раменский

Анастасия Александровна Жарикова и Мария Ильинична Зайченока

НМИЦ Терапии и профилактической медицины

Факультет биоинженерии и биоинформатики МГУ

Институт искусственного интеллекта МГУ

2024

В предыдущих сериях

- В прошлый раз мы отобрали среди набора вариантов всего несколько
- Средствами какого ресурса мы для этого пользовались?
- Какие критерии были применены для отбора вариантов?

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

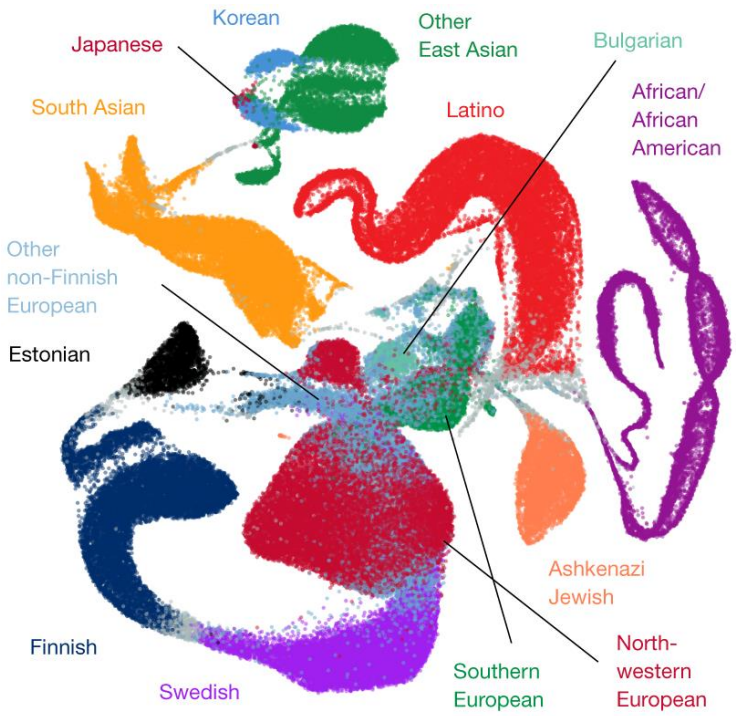
- На предыдущем слайде один вариант охарактеризован 6 способами
- Расшифруйте\объясните каждый из них

Задача семинара

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

gnomAD

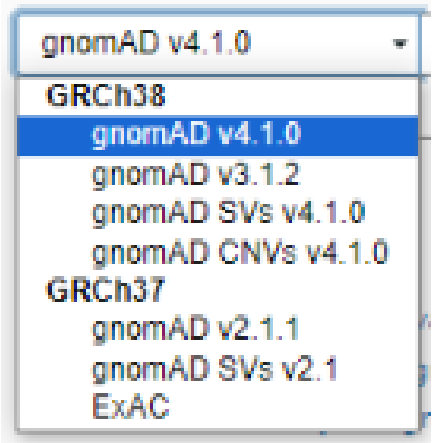
- https://gnomad.broadinstitute.org/



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

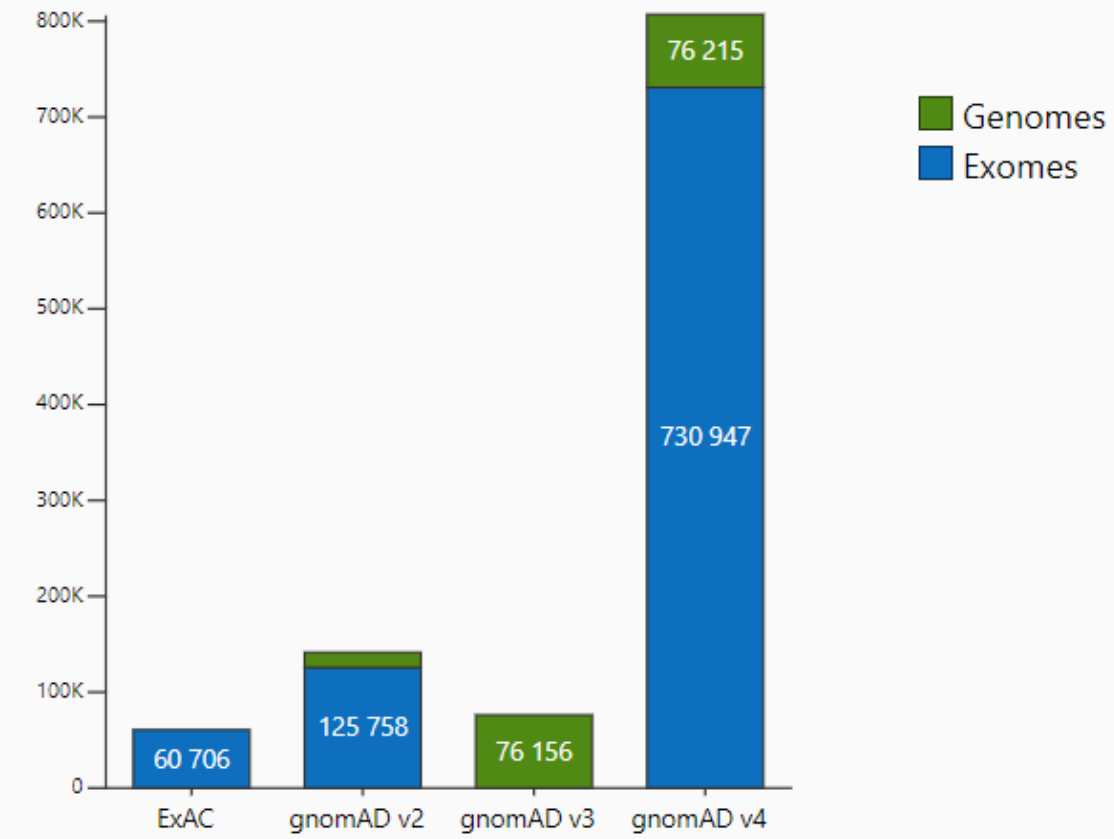


gnomAD

Genome Aggregation Database

gnomAD v4 includes 807,162 individuals

- 730,947 exomes
- 76,215 genomes



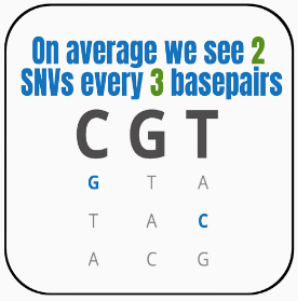
Откуда появилось столько новых вариантов?

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

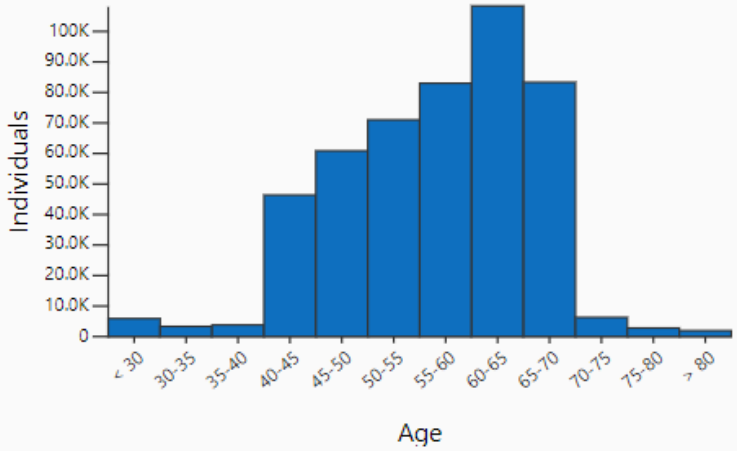
- 1,199,117 genome SVs
  - 627,947 Deletions
  - 258,882 Duplications
  - 711 CNVs
  - 296,184 Insertions
  - 2,185 Inversersions
  - 13,116 Complex
  - 92 Canonical recriprocal translocations
- 66,903 rare (<1% site frequency (SF)) exome CNVs
  - 30,877 Deletions
  - 36,026 Duplications

gnomAD

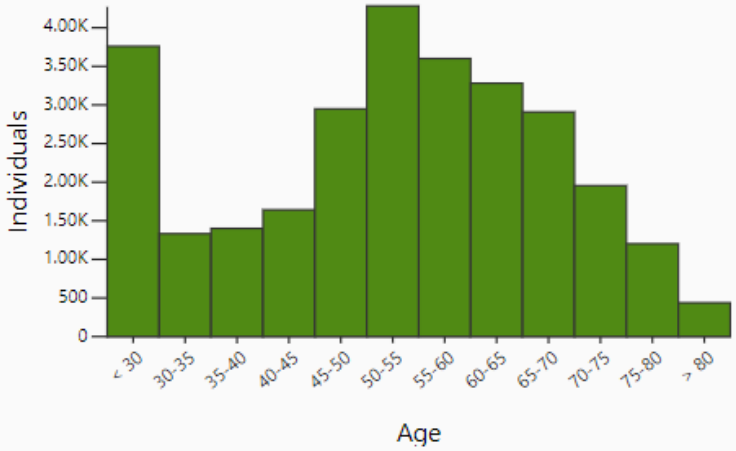
gnomAD
Genome Aggregation Database

Age

Exomes



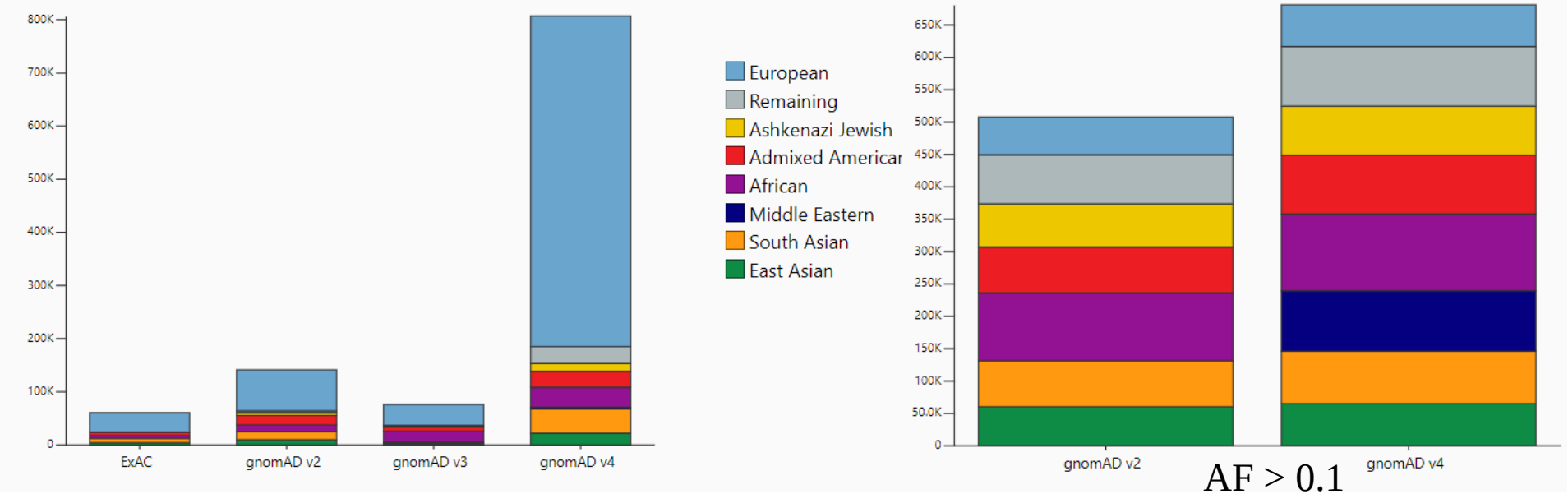
Genomes



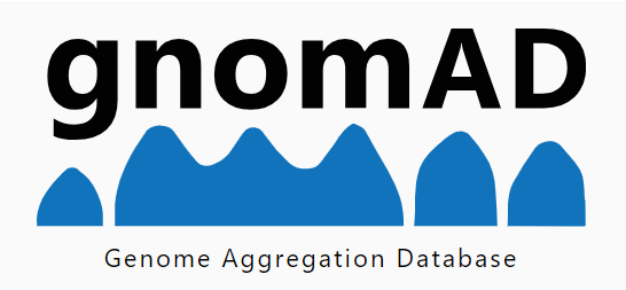
Sex

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

gnomAD

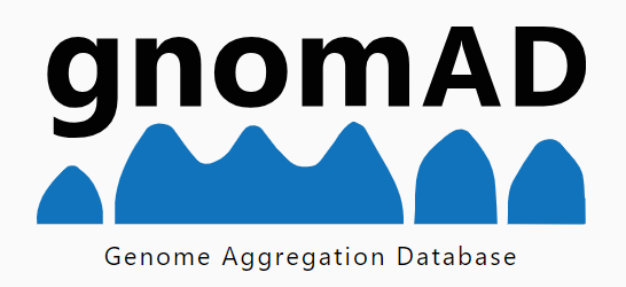

800K
700K
600K
500K
400K
300K
200K
100K
0
ExAC
gnomAD v2
gnomAD v3
gnomAD v4
European
Remaining
Ashkenazi Jewish
Admixed American
African
Middle Eastern
South Asian
East Asian
650K
600K
550K
500K
450K
400K
350K
300K
250K
200K
150K
100K
50.0K
0
gnomAD v2
gnomAD v4
AF > 0.1

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		

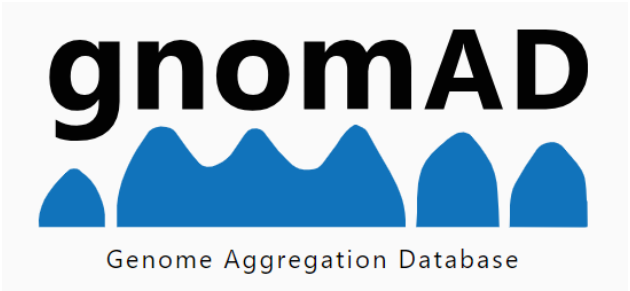
External Resources

- 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

- 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

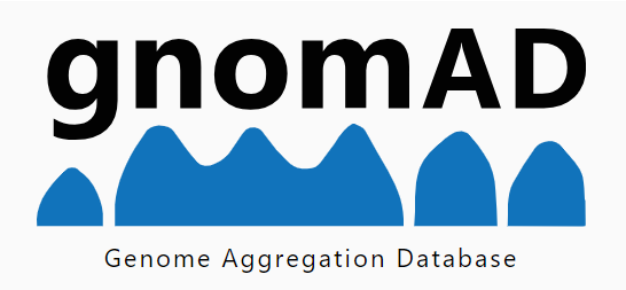
1. **APOB**
   1. ENST00000673739.1
      HGVSc: c.*2092del
   2. ENST00000673882.1
      HGVSc: c.*2092del

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

gnomAD



Site Quality Metrics

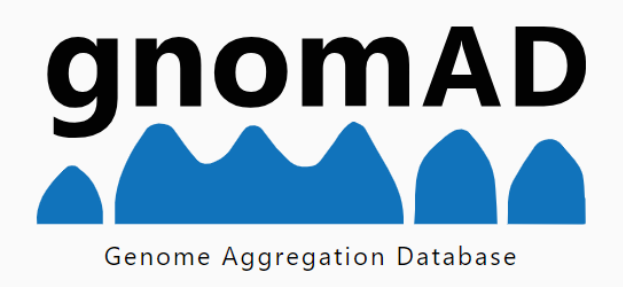
Metric distribution | All metric values

Metric	Exome samples
SiteQuality	2.039e+3
inbreeding_coeff	-1.590e-6
AS_FS	3,108
AS_MQ	60
AS_MQRankSum	0
AS_pab_max	0,738
AS_QUALapprox	1.061e+3
AS_QD	13,263
AS_ReadPosRankSum	-0,608
AS_SOR	1,103
AS_VarDP	80
AS_VQSLOD	3,664

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

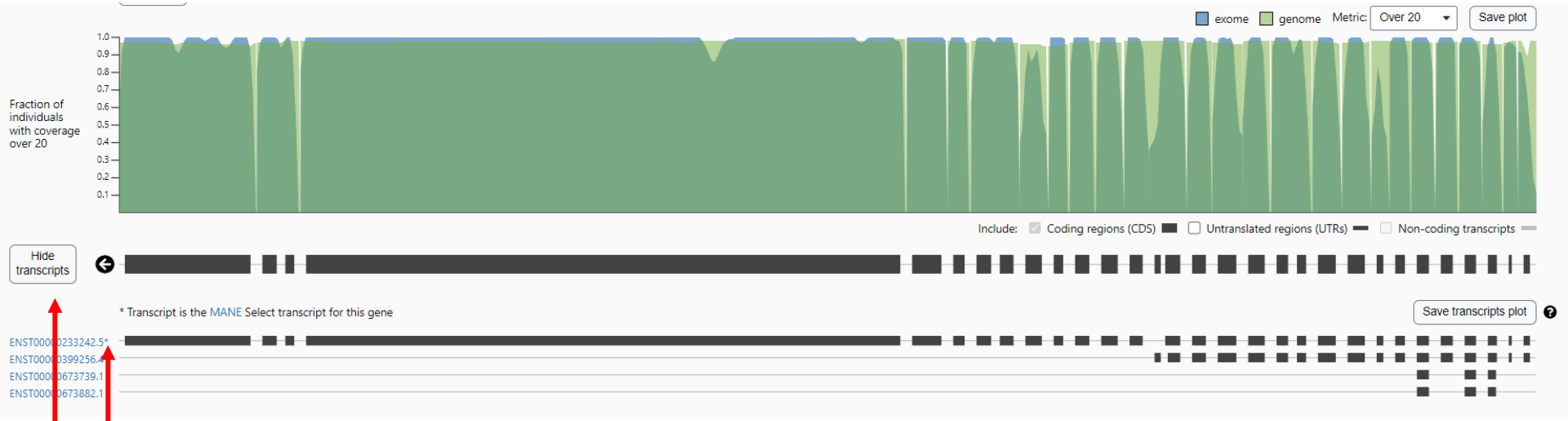
Category	Expected SNVs	Observed SNVs	Constraint metrics
Synonymous	2081.8	2013	Z = 0.82 o/e = 0.97 (0.93 - 1)
Missense	5461	5271	Z = 0.94 o/e = 0.97 (0.94 - 0.99)
pLoF	319.7	150	pLI = 0 o/e = 0.47 (0.41 - 0.54)

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

Упражнение

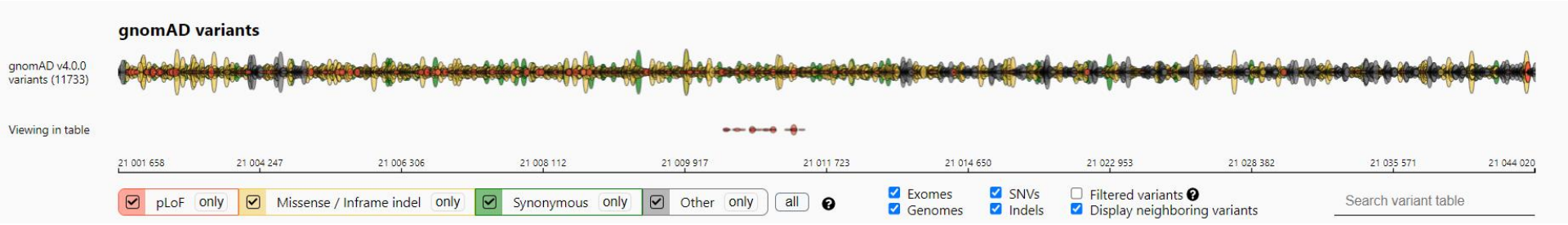
- Посмотрите на выдачу по генам:
  - GAPDH
  - CCK

gnomAD



Для каждого гена есть возможность показать все транскрипты

gnomAD



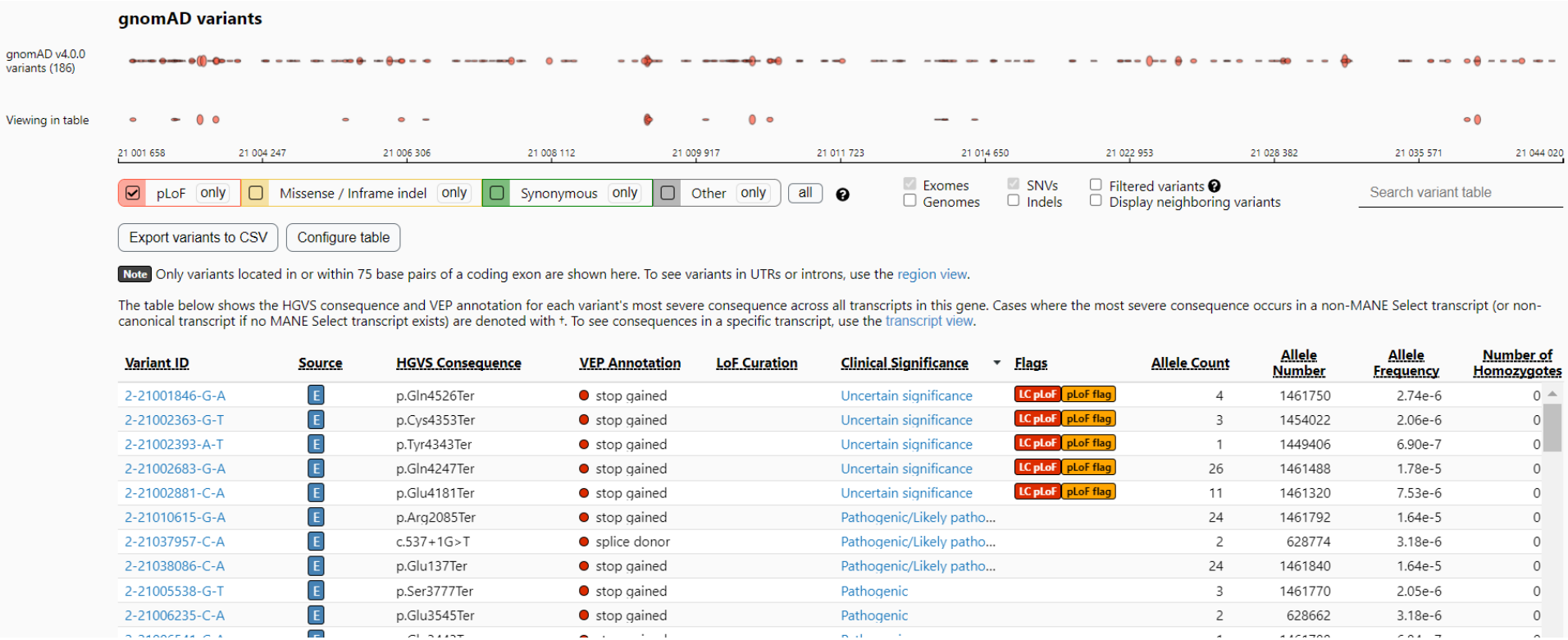
Варианты можно отфильтровать по ряду критериев

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

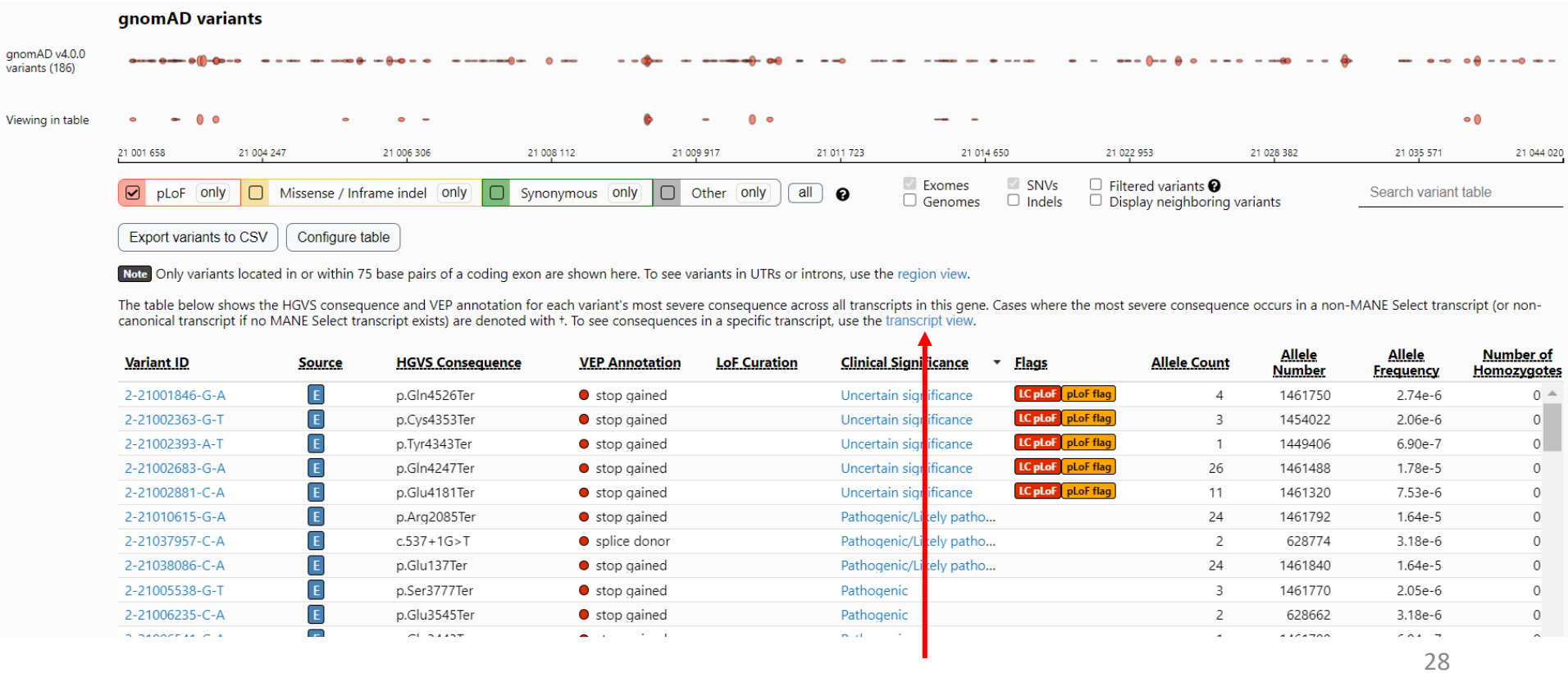
Note Only variants located in or within 75 base pairs of a coding exon are shown here. To see variants in UTRs or introns, use the region view.

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

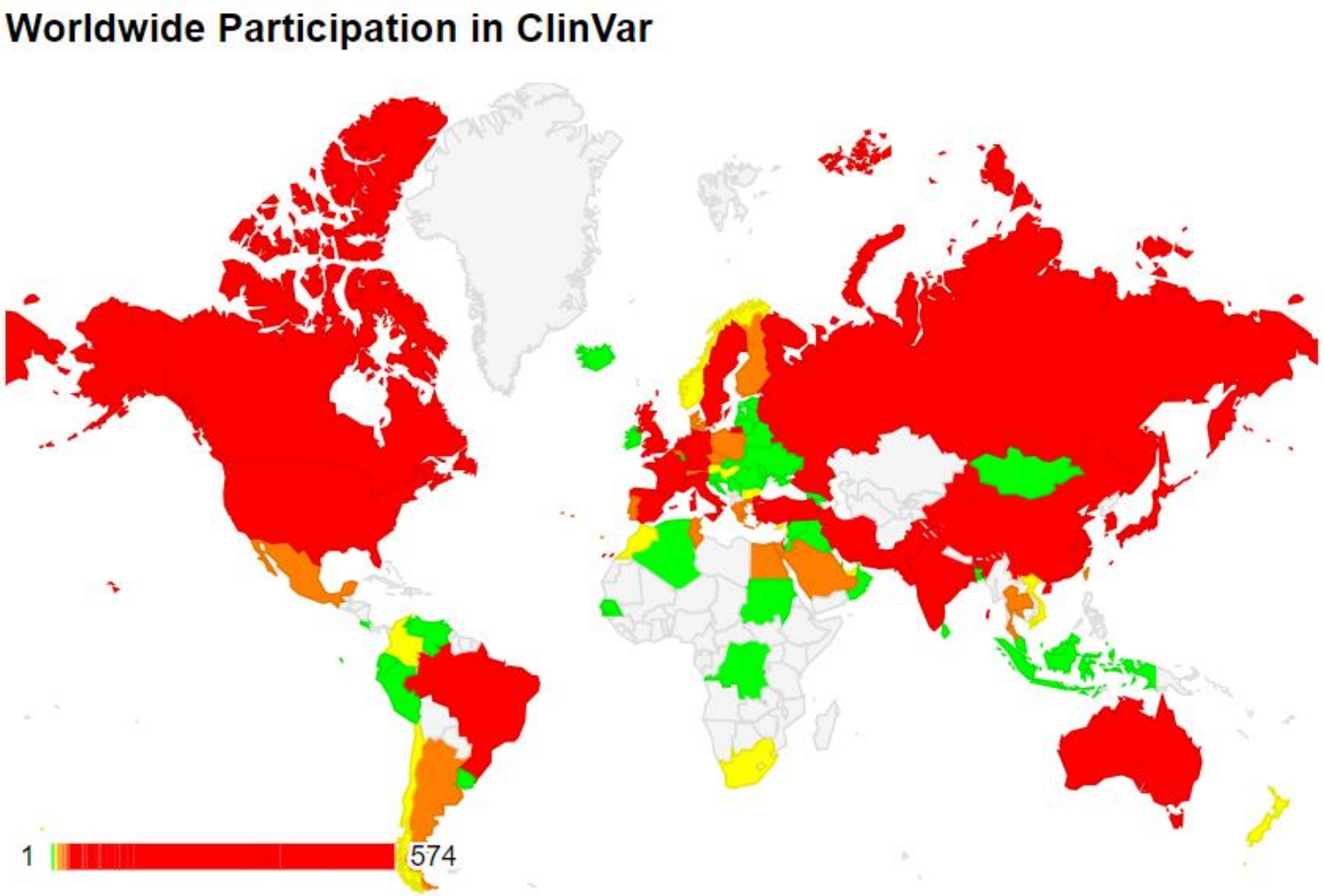
- Обратите внимание, что при исследовании можно сосредоточиться на конкретных транскриптах



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

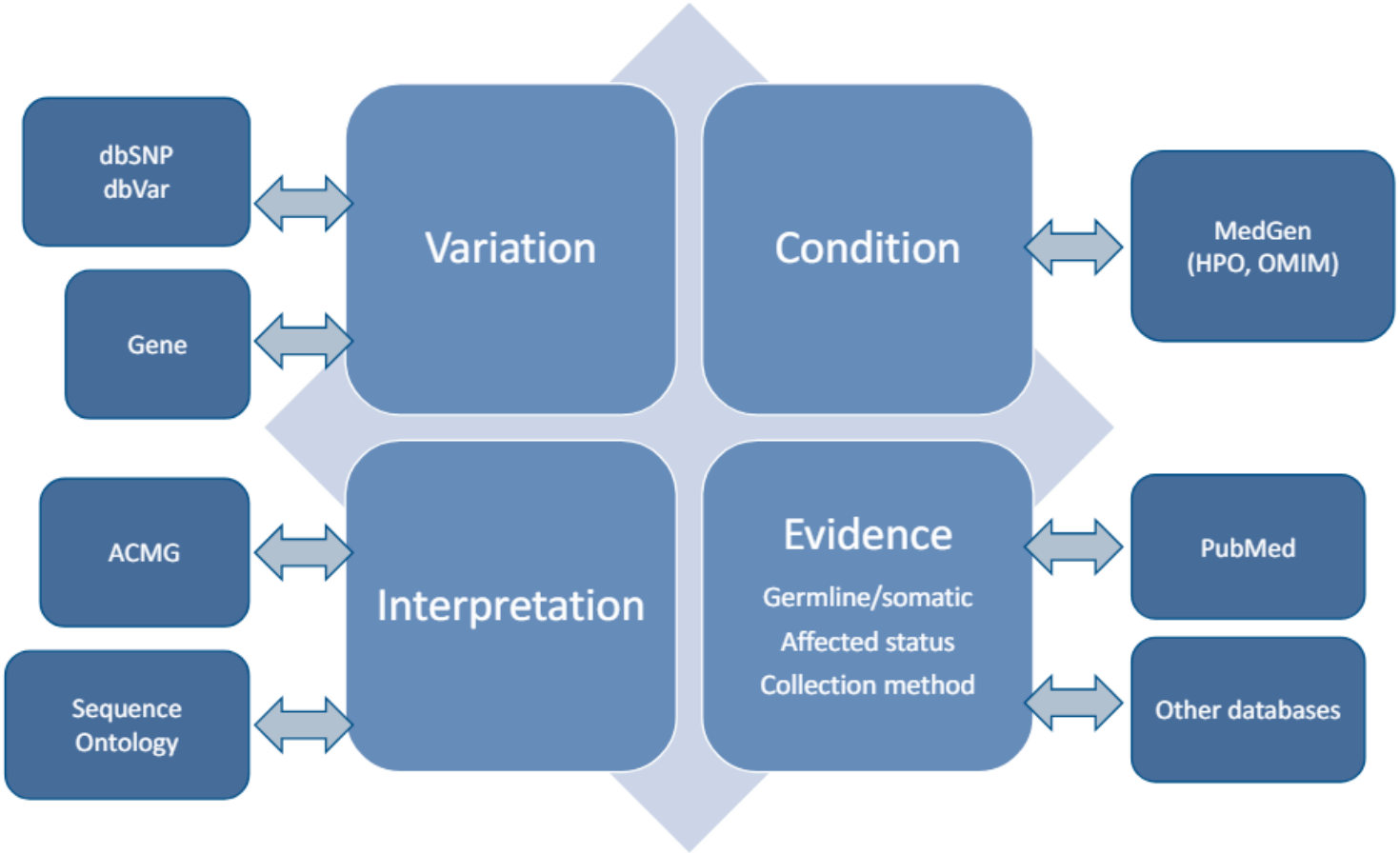
ClinVar

- 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



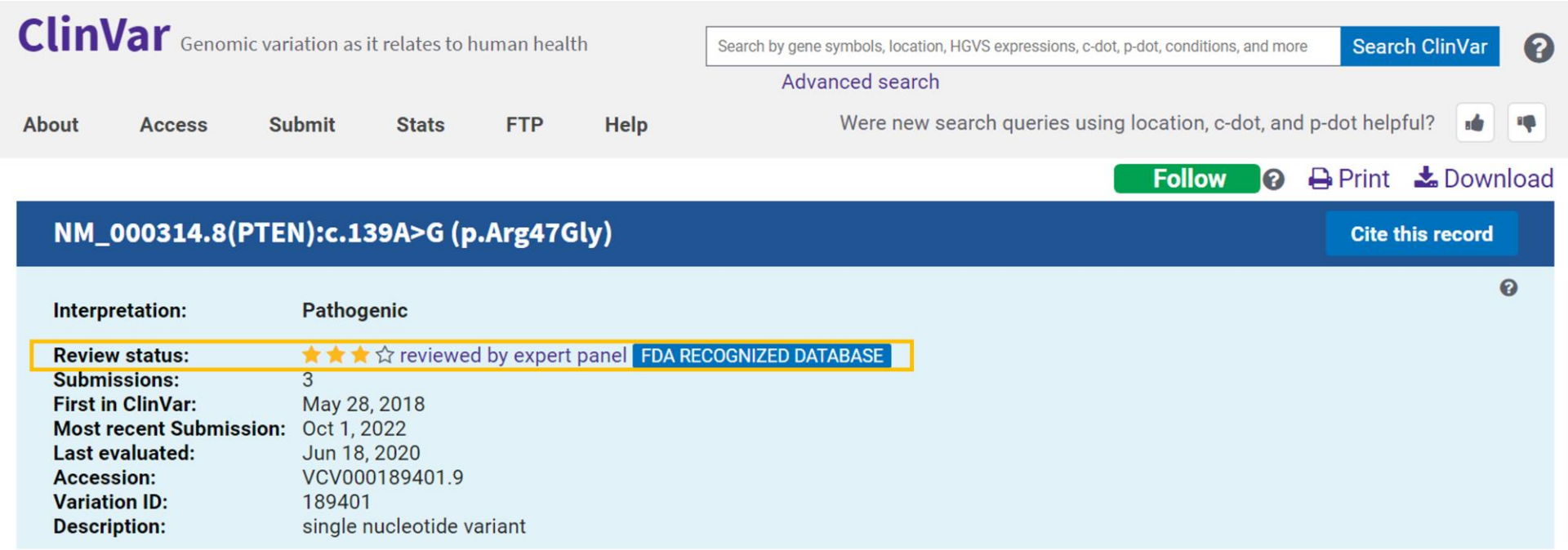
About ClinVar

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)

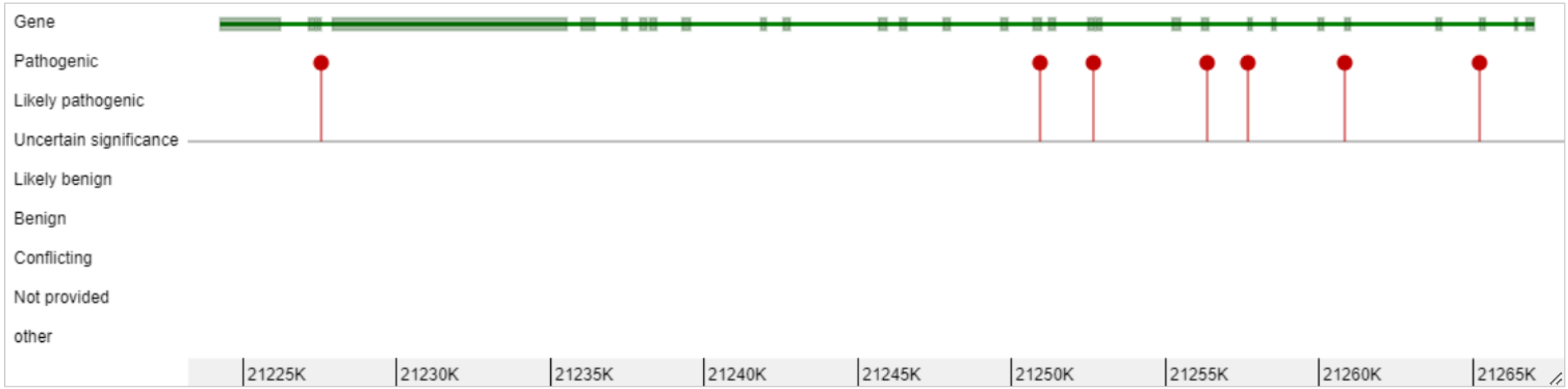
Variation size
Short variant (< 50 bps) (145)
Structural variant (>= 50 bps) (0)

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)

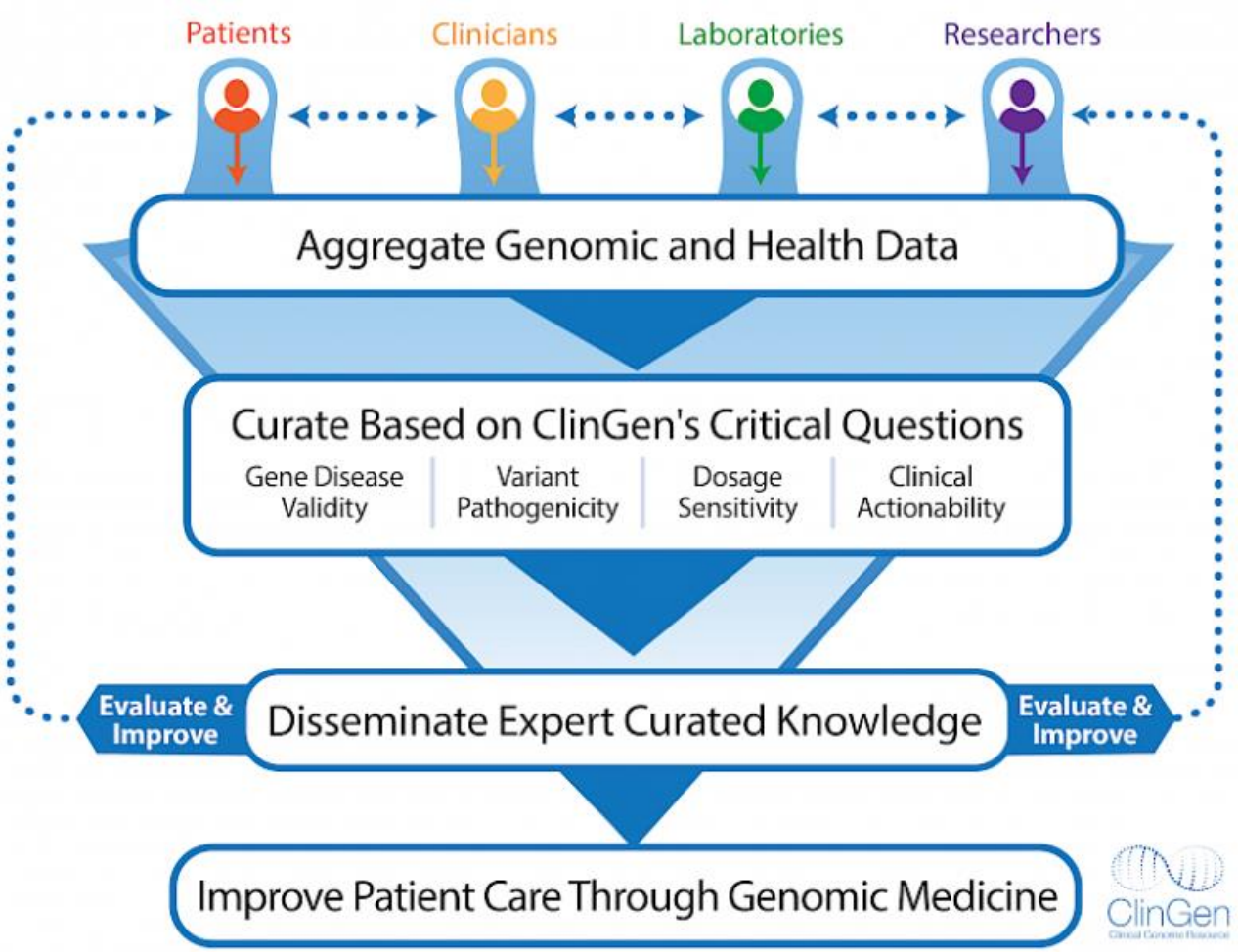
ClinVar

- 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

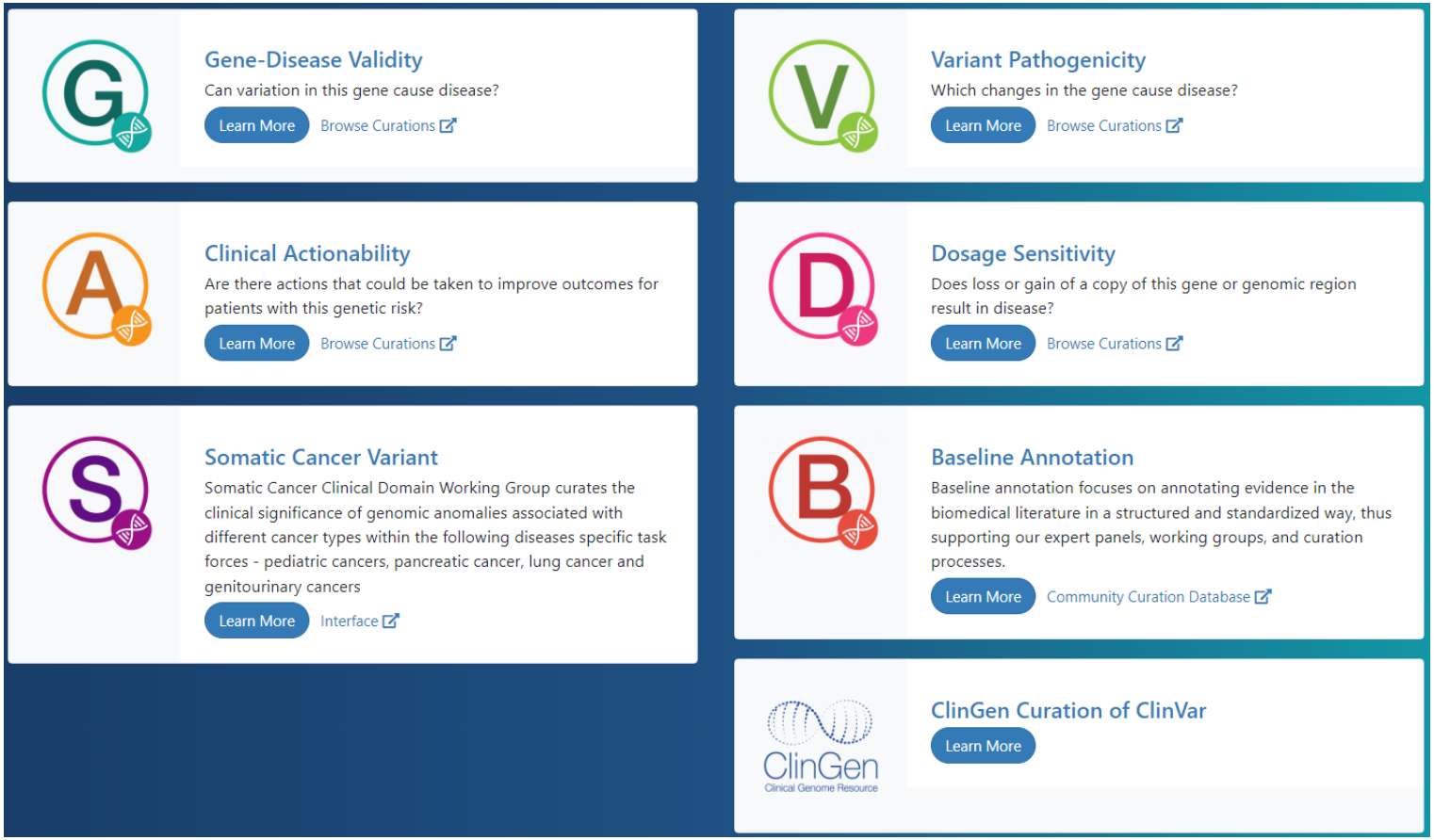
ClinGen
Clinical Genome Resource



ClinGen

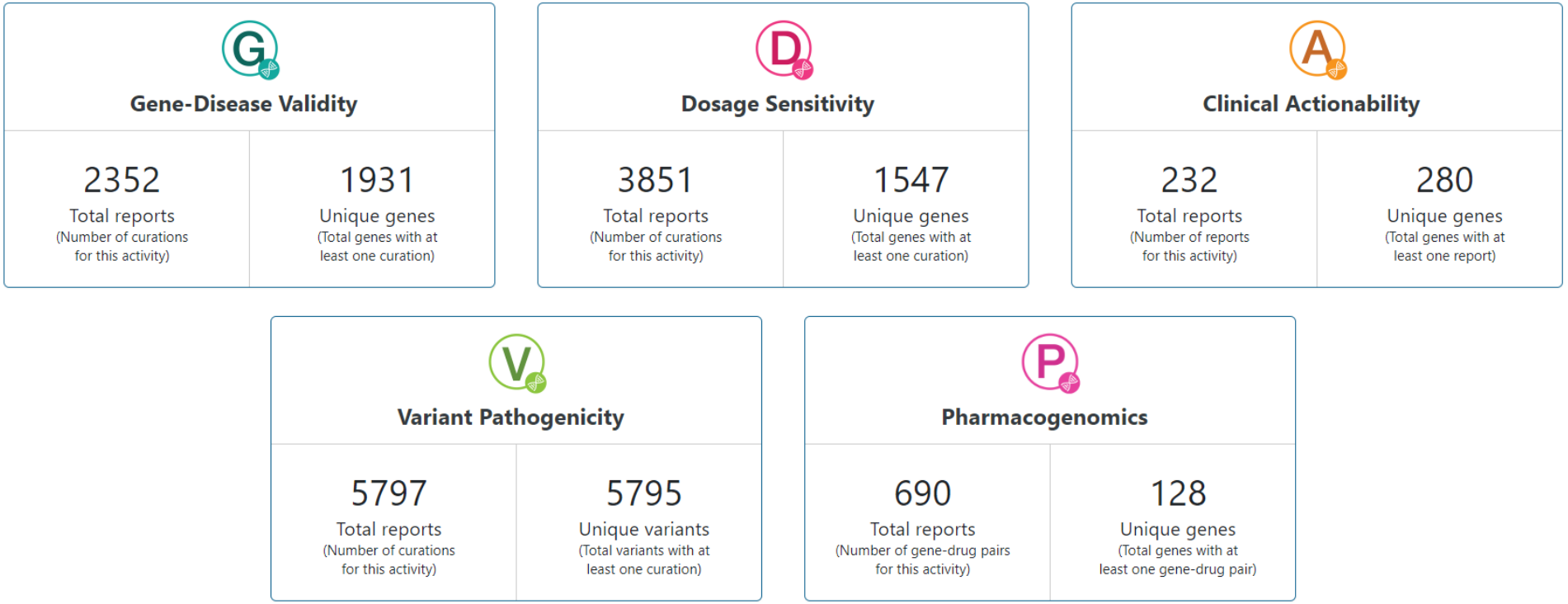
ClinGen
Clinical Genome Resource

- https://clinicalgenome.org/



ClinGen

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

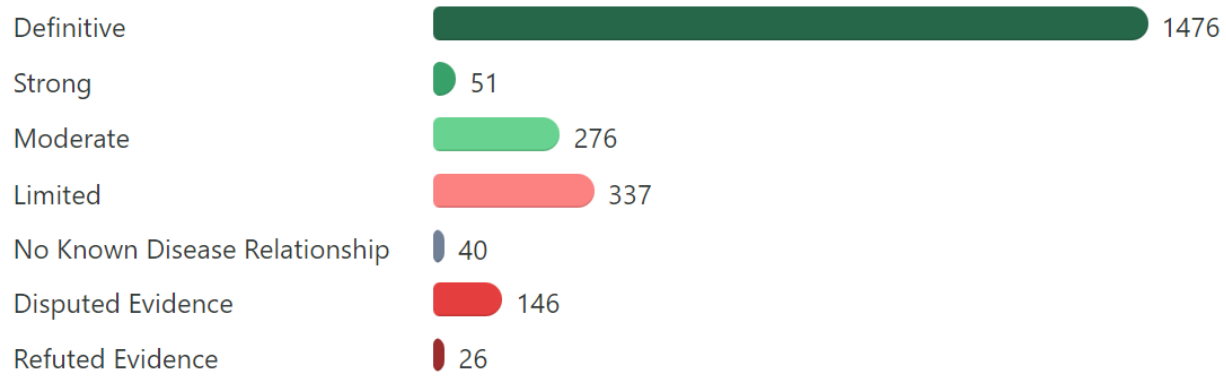


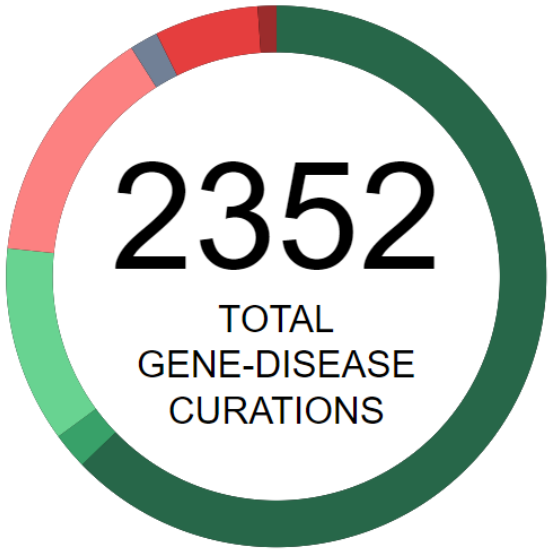
ClinGen

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

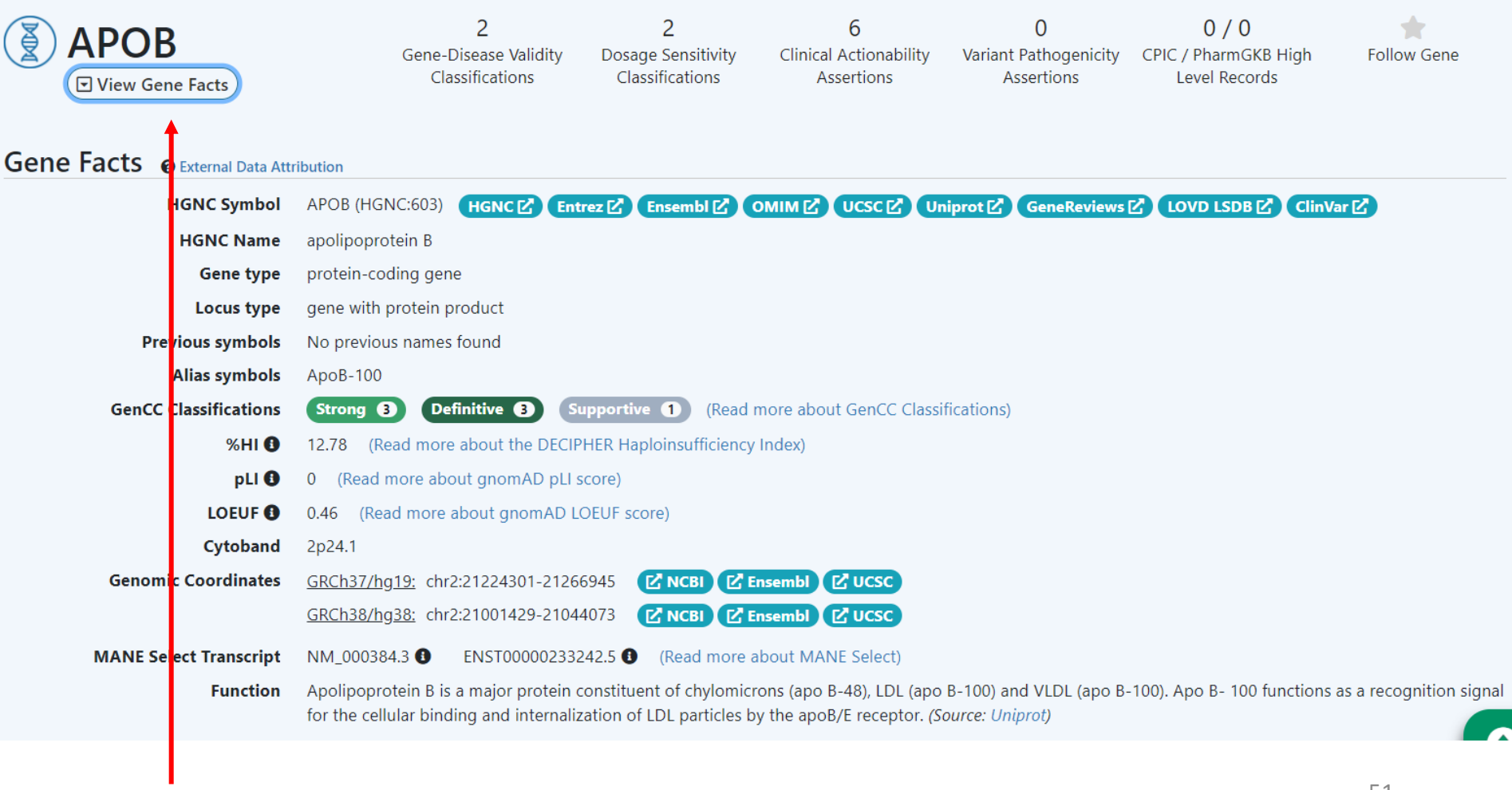
APOB **- hypercholesterolemia, autosomal dominant, type B**

(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

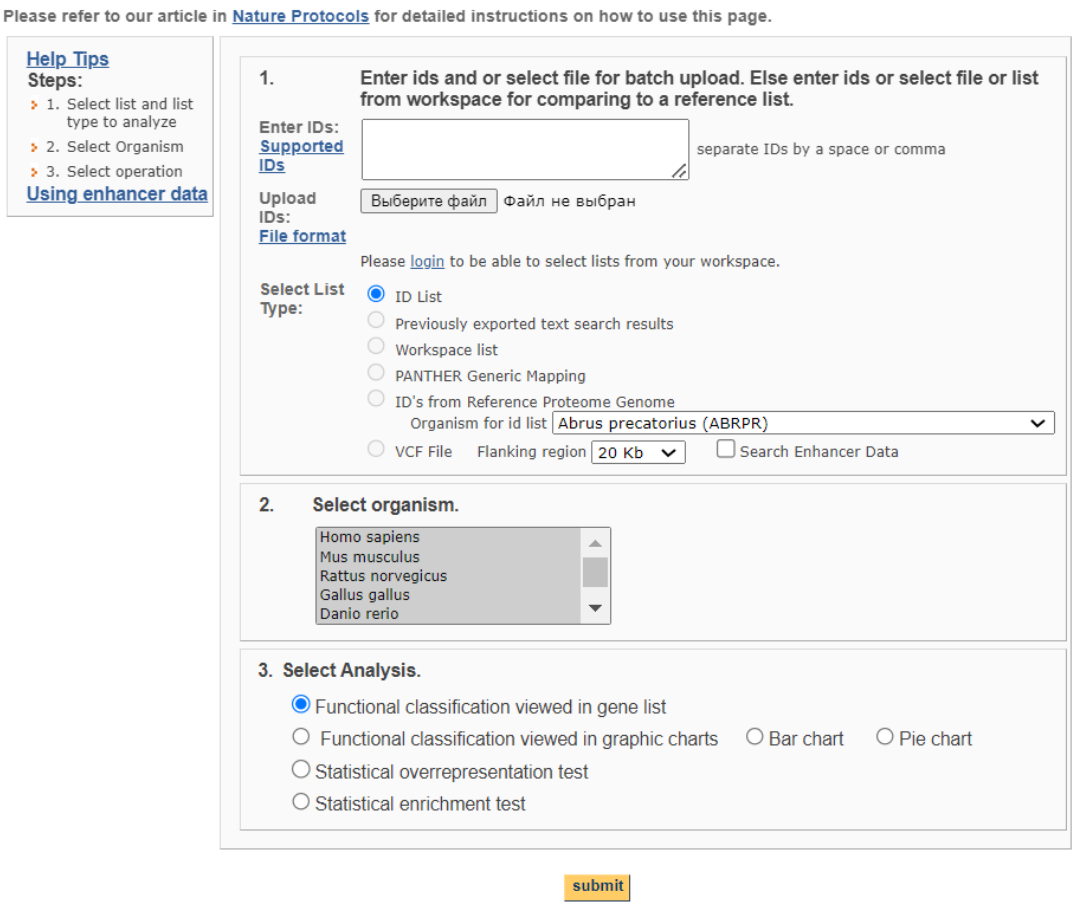
ClinGen

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

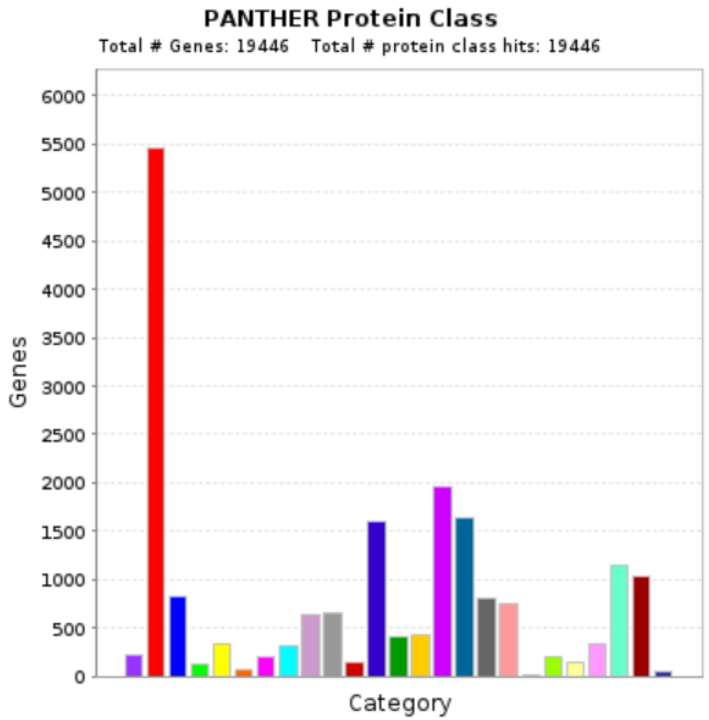
PANTHER

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

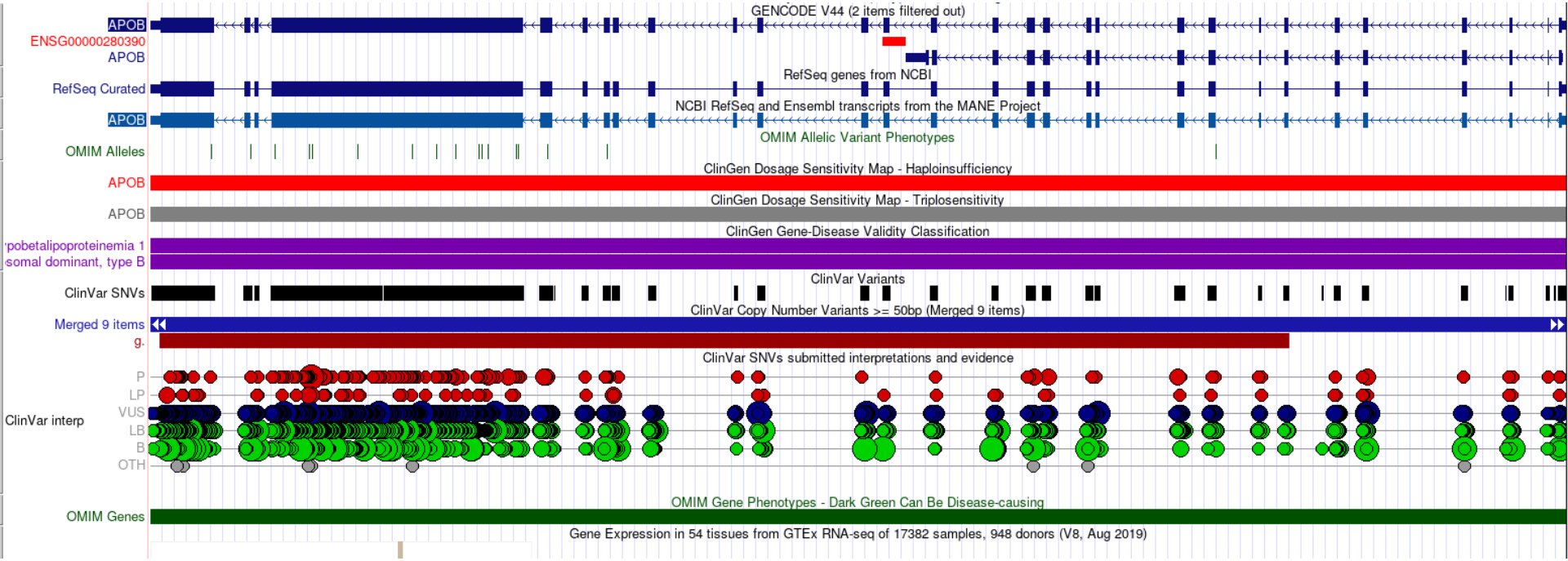
- С помощью возможностей базы Panther.db охарактеризуйте свой белок (APOB)
- Обратите внимание, что при поиске необходимо указать нужный организм

Упражнение

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

Упражнение

- Примерное решение для гена

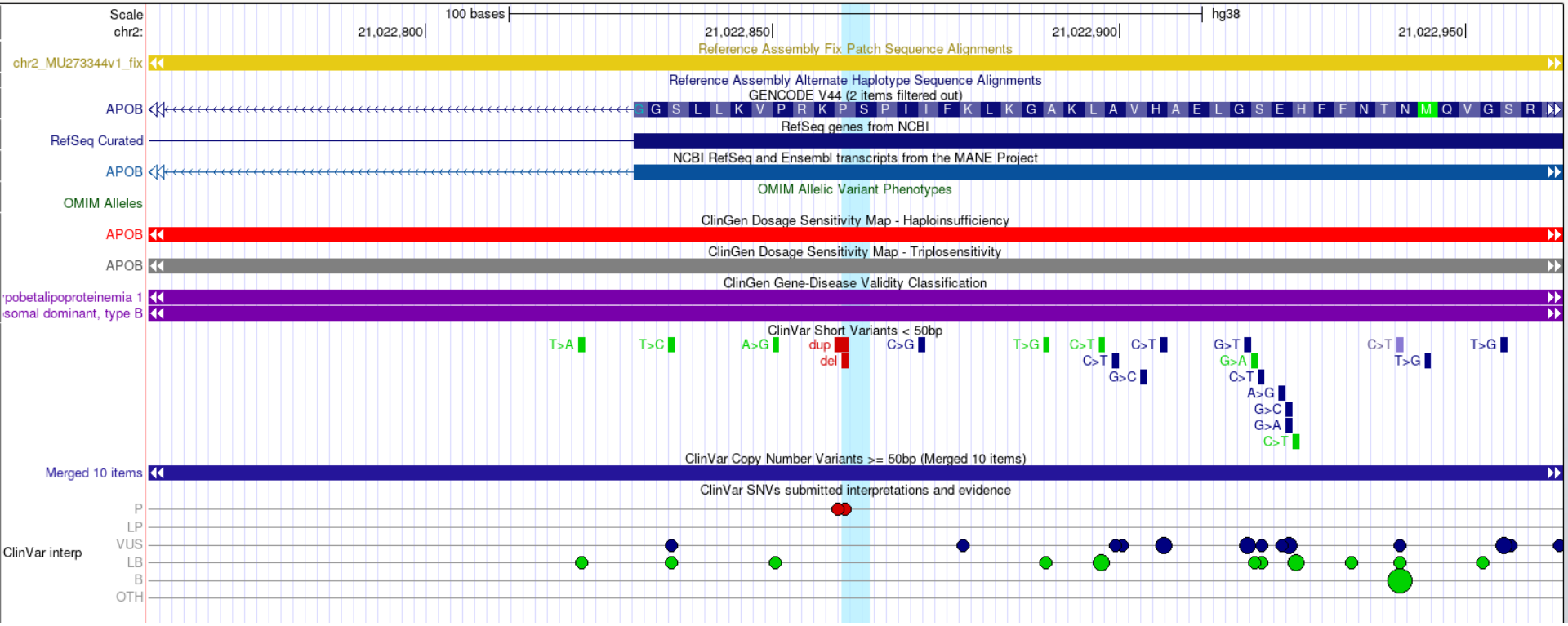


Упражнение

- 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