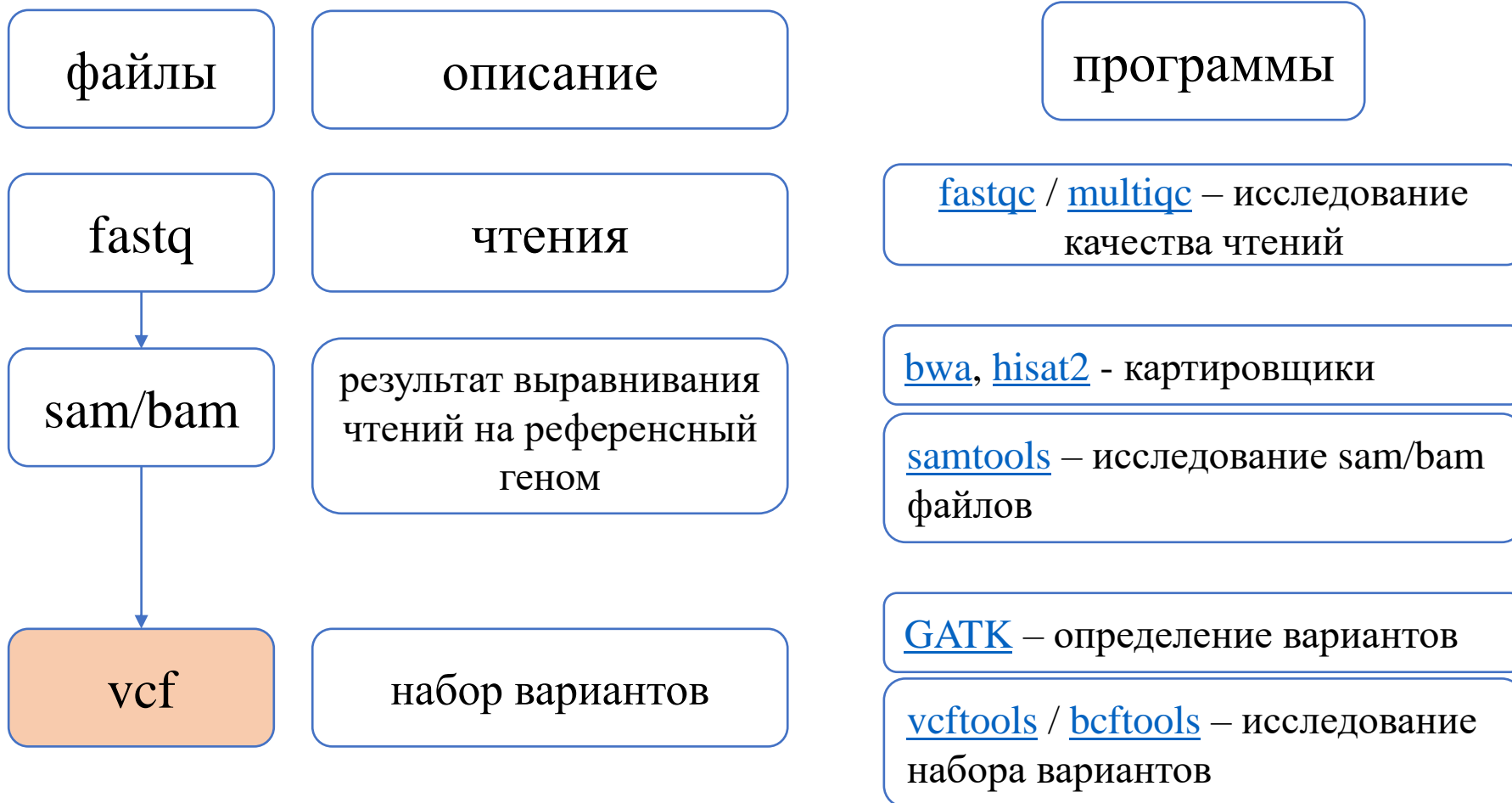


# Введение в анализ NGS

Блок 13 – Занятие 14

Анастасия Жарикова

# Протокол



Все файлы храним и анализируем в архивированном виде!

# Variant Effect Predictor



- [VEP](#)
- На вход можно подать vcf файл с вариантами


A screenshot of the Ensembl Variant Effect Predictor (VEP) web interface. The top navigation bar is dark blue with the 'e!Ensembl' logo on the left and links for 'BLAST/BLAT', 'VEP', 'Tools', 'BioMart', 'Downloads', 'Help &amp; Docs', and 'Blog' on the right. Below the navigation bar, there is a 'VEP' dropdown menu. On the left, a 'Web Tools' sidebar is open, showing a list of tools: 'Web Tools', 'BLAST/BLAT', 'Variant Effect Predictor' (highlighted), 'Linkage Disequilibrium Calculator', 'Variant Recoder', 'File Chameleon', 'Assembly Converter', 'ID History Converter', 'VCF to PED Converter', and 'Data Slicer'. On the right, the main content area is titled 'Variant Effect Predictor ?'. It contains a 'New job' button, a 'Recent jobs' section with a dropdown arrow, and a 'Refresh' button with a circular arrow icon. A red arrow points from the bottom left towards the 'New job' button.

# VEP



## New job

Species:

 Homo\_sapiens X

Assembly: *GRCh38.p14*  
[Change species](#)

Name for this job (optional):

Input data:

Either paste data:

Examples: [Ensembl default](#), [VCF](#), [HGVS notations](#), [SPDI](#)

Or upload file:  Файл не выбран

Or provide file URL:



# VEP



- При загрузке файла создается новый процесс

A screenshot of the VEP web interface. The top navigation bar has 'Analysis' and 'Jobs' tabs. Under 'Jobs', there is a job entry: 'Variant Effect Predictor' followed by a small profile icon, the text 'VEP analysis of doom\_1\_sample\_het\_hom.vcf in Homo\_sapiens', and a grey button labeled 'Queued'.

- Аннотация занимает какое-то время
- Дождитесь статуса **Done**

A screenshot of the VEP web interface showing the same job entry as the previous screenshot, but now the status is 'Done' in a green button. A blue link labeled '[View results]' is visible to the right of the 'Done' button. A red arrow points from the bottom right towards the 'View results' link.

# VEP



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

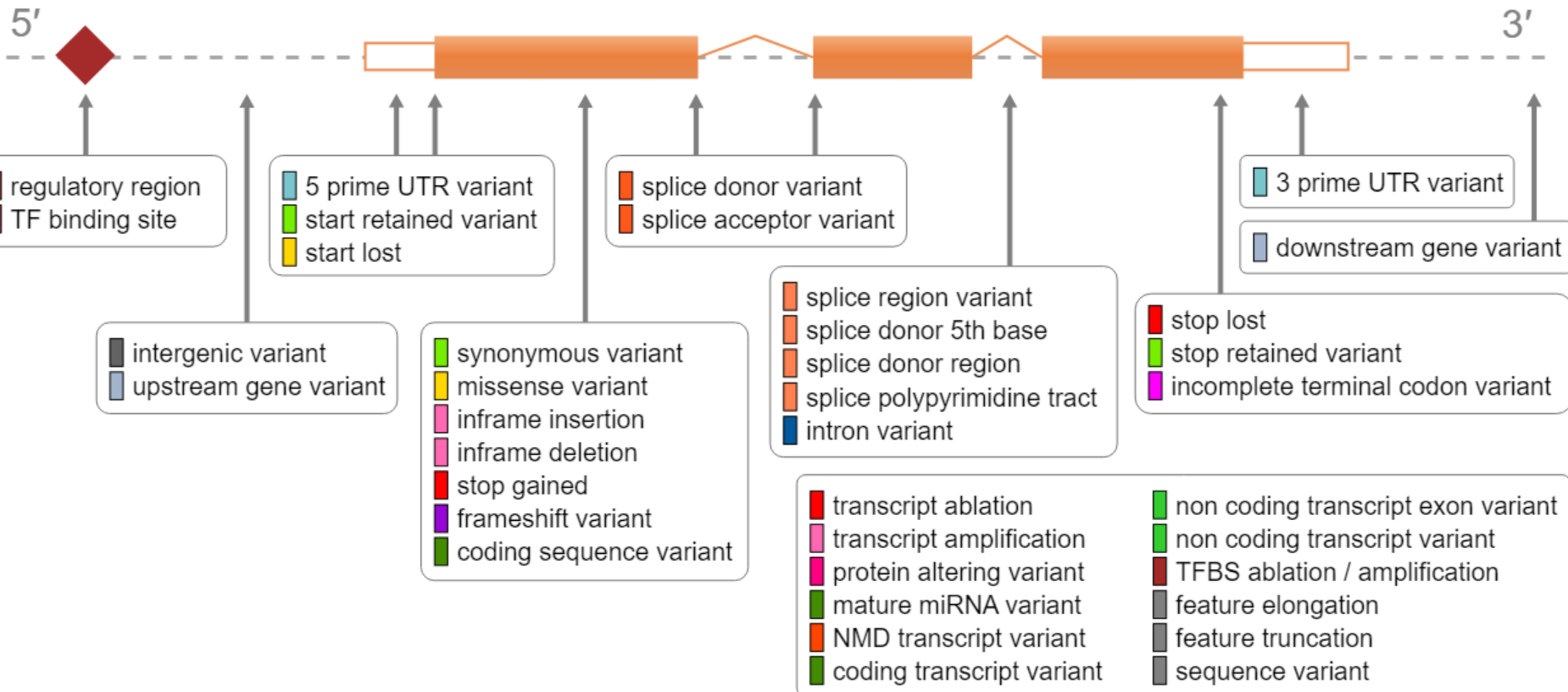
# Фильтрация

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

# VEP



- Consequences



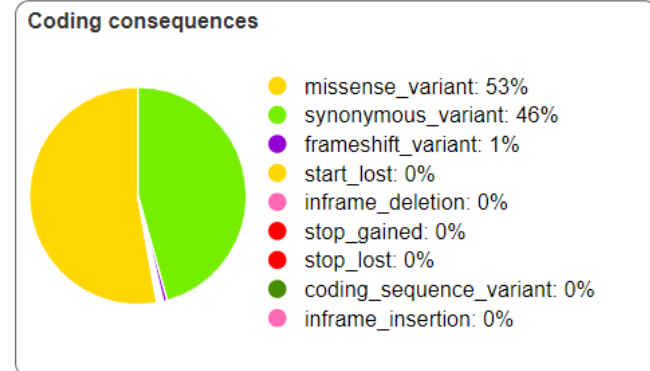
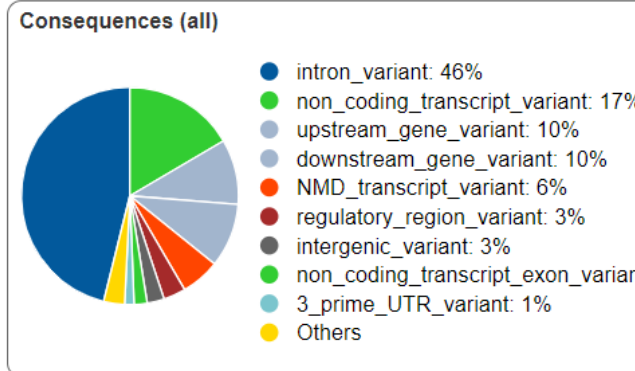


# VEP



- 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



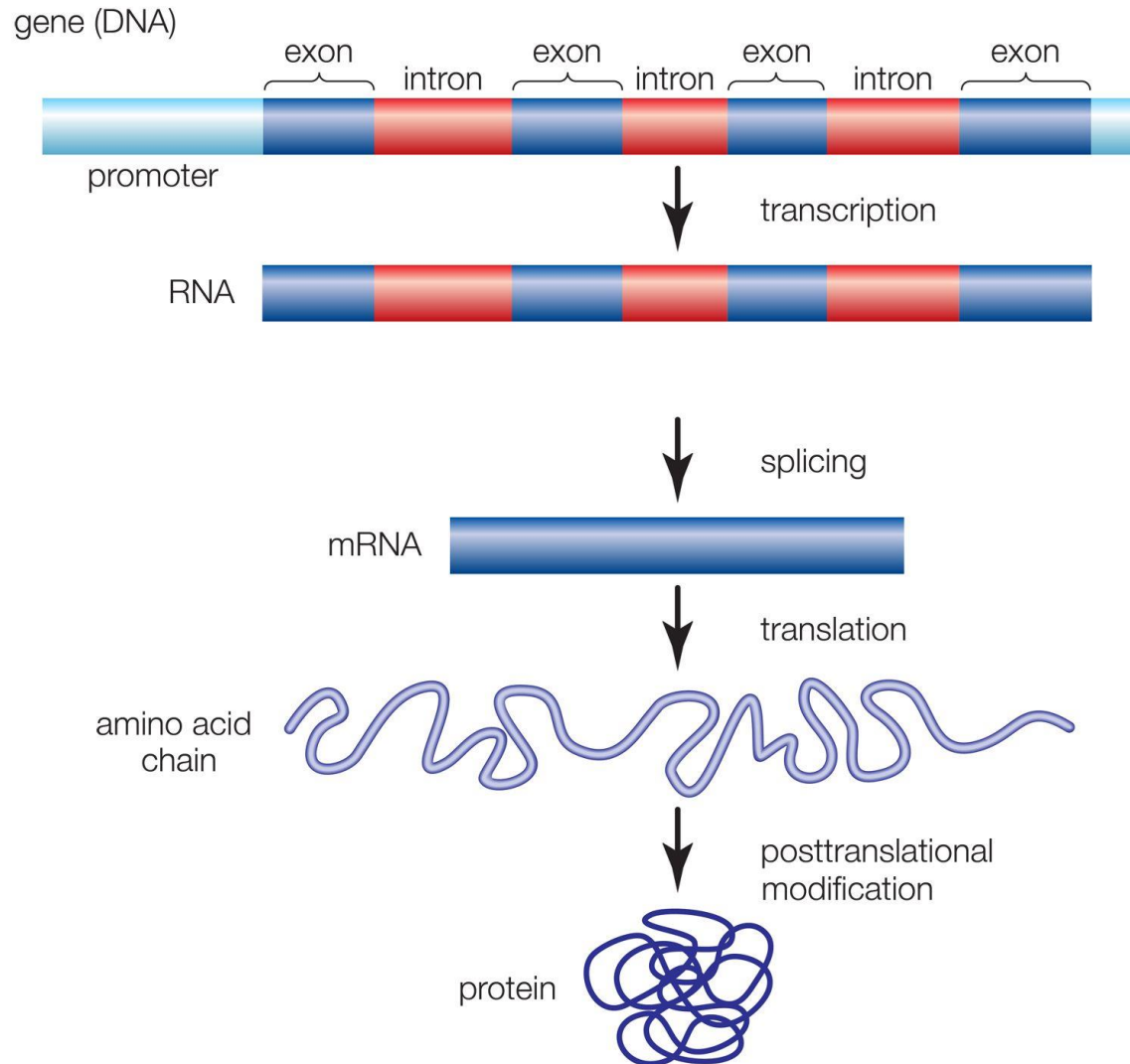
| <i>IMPACT</i> | <i>Consequence examples</i>  | <i>Description</i>  |
|---------------|--|---|
| HIGH          | splice_acceptor_variant,<br>splice_donor_variant,<br>stop_gained, stop_lost,<br>start_lost | The variant is assumed to have high (disruptive) impact in the protein, probably causing protein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

# VEP

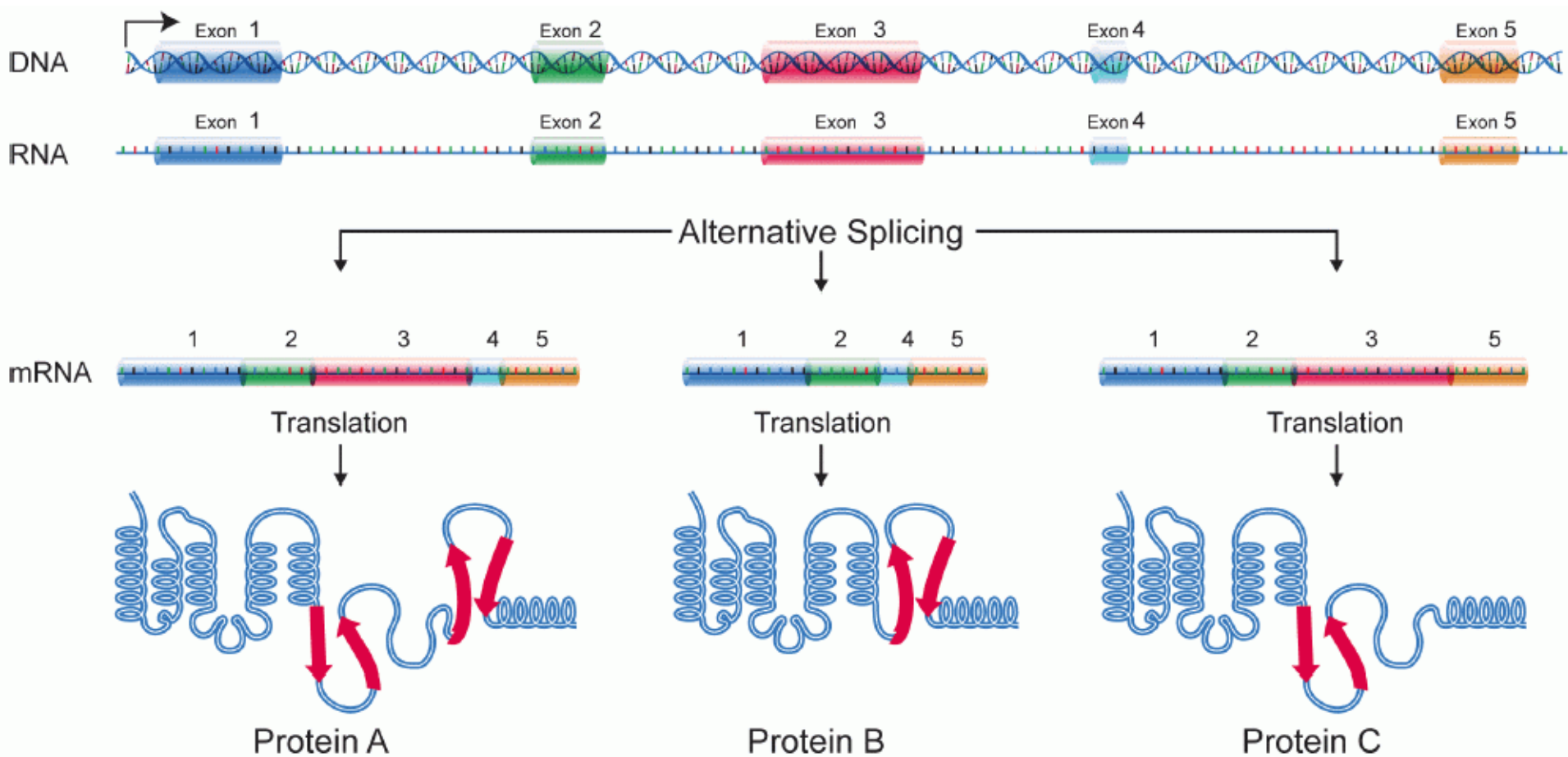


- [HGVC](#)
- Рекомендации по описанию геномных вариантов
- Единая система описания вариантов позволяет присваивать уникальное и однозначное «имя» варианту
  - HGVS<sub>c</sub> - ENST00000320048.1:c.819T>A
  - HGVS<sub>p</sub> - ENSP00000321506.1:p.Tyr273Ter

# Структура гена



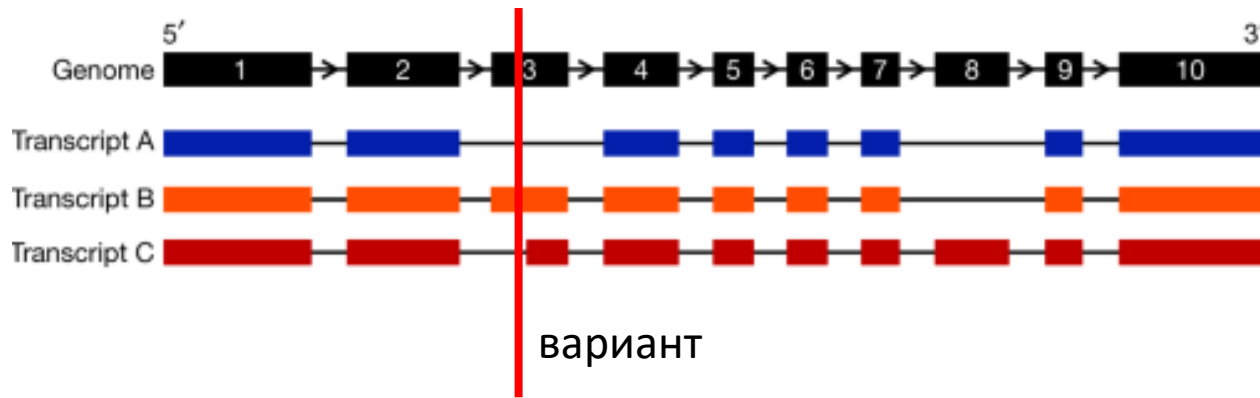
# Альтернативный сплайсинг



# 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



# 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

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

UCSC [uc002qsd.5](#)

NCBI Gene [1](#) [Curated](#)

Alliance of Genome  
Resources [HGNC:5](#)

## Nucleotide resources for A1BG [?](#)

[MANE Select](#) [NM\\_130786.4](#)  
[ENST00000263100.8](#)

CCDS [CCDS12976](#) [Curated](#)

RefSeq [NM\\_130786](#) [Curated](#)  
[NCBI sequence viewer](#)

# HGNC

- Полезное для медицинской геномики

Clinical resources for A1BG ?

|   |  |
|---|--|
| <b>OMIM</b> 138670 <a href="#">↗</a>  | <b>MedlinePlus</b> Search via A1BG <a href="#">↗</a>       |
| <b>DECIPHER</b> Search via A1BG <a href="#">↗</a>                           | <b>ClinGen</b> Search via HGNC:5 <a href="#">↗</a>         |
| <b>Genetic Testing Registry</b> Search via NCBI Gene ID 1 <a href="#">↗</a> | <b>ClinVar</b> Search via NCBI Gene ID 1 <a href="#">↗</a> |
| <b>dbVar</b> Search via NCBI Gene ID 1 <a href="#">↗</a>                    |  |

- Справочная информация





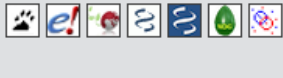


Other resources for A1BG ?

|   |  |
|---|--|
| <b>AmiGO</b> Search via P04217 <a href="#">↗</a>          | <b>QuickGO</b> Search via P04217 <a href="#">↗</a>           |
| <b>BioGPS</b> Search via NCBI Gene ID 1 <a href="#">↗</a> | <b>GeneCards</b> Search via HGNC:5 <a href="#">↗</a>         |
| <b>Monarch</b> Search via HGNC:5 <a href="#">↗</a>        | <b>WikiGenes</b> Search via NCBI Gene ID 1 <a href="#">↗</a> |

# HGNC

- Полезное для филогенетики

Report **HCOP homology predictions**

|  |   |   |
|--|---|---|
|  <p><b>Human</b></p>  | <p><b>Approved symbol</b> <b>A1BG</b> ⓘ</p> <p><b>Approved name</b> alpha-1-B glycoprotein ⓘ</p> <p><b>Locus type</b> gene with protein product ⓘ</p> <p><b>Chromosomal location</b> 19q13.43</p> <p><b>Gene resources</b> <a href="#">HGNC:5</a> <a href="#">e!ENSG00000121410</a> <a href="#">1</a></p>                     |   |
|  <p><b>Chimp</b></p> <p><a href="#">reciprocal search</a> ⓘ</p>     | <p><b>Gene symbol</b> <b>A1BG</b> ⓘ</p> <p><b>Gene name</b> alpha-1-B glycoprotein ⓘ</p> <p><b>Locus type</b> protein_coding ⓘ</p> <p><b>Chromosomal location</b> 19</p> <p><b>Gene resources</b> <a href="#">e!ENSPTRG00000011588</a> <a href="#">742390</a></p>   | <p><b>Assertion derived from:</b></p>    |
|  <p><b>Macaque</b></p> <p><a href="#">reciprocal search</a> ⓘ</p>  | <p><b>Approved symbol</b> <b>A1BG</b> ⓘ</p> <p><b>Approved name</b> alpha-1-B glycoprotein ⓘ</p> <p><b>Locus type</b> gene with protein product ⓘ</p> <p><b>Chromosomal location</b> 19</p> <p><b>Gene resources</b> <a href="#">VGNC:69569</a> <a href="#">e!ENSMMUG00000012459</a> <a href="#">712737</a></p>               | <p><b>Assertion derived from:</b></p>   |
|  <p><b>Macaque</b></p> <p><a href="#">reciprocal search</a> ⓘ</p> | <p><b>Approved symbol</b> <b>AFF1</b> ⓘ</p> <p><b>Approved name</b> ALF transcription elongation factor 1 ⓘ</p> <p><b>Locus type</b> gene with protein product ⓘ</p> <p><b>Chromosomal location</b> 5</p> <p><b>Gene resources</b> <a href="#">VGNC:69817</a> <a href="#">e!ENSMMUG00000014076</a> <a href="#">700733</a></p> | <p><b>Assertion derived from:</b></p>  |

+ еще много организмов ниже!

# Genome Browser

- [Геномный браузер](#)

The screenshot shows the UCSC Genome Browser Gateway interface. At the top, there is a header with the University of California Santa Cruz Genomics Institute logo and the text "UCSC Genome Browser Gateway". Below the header is a navigation bar with links for Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Projects, Help, and About Us. The main content area is divided into two sections: "Browse/Select Species" and "Find Position".

**Browse/Select Species**

POPULAR SPECIES

Human Mouse Rat Zebrafish Fruitfly Worm Yeast

Search through thousands of genome browsers  
Enter species, common name or assembly ID

**Find Position**

Human Assembly  
Dec. 2013 (GRCh38/hg38)

Position/Search Term  
Enter position, gene symbol or search terms  
Current position: chr12:6,533,553-6,539,335

GO

# 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

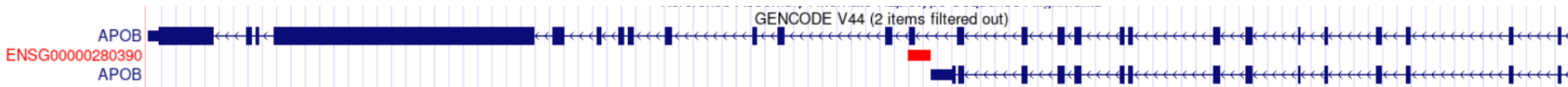
Search  
Human  for

Use the tree to hide/show results from only these categories. Hover your mouse over each category for an explanation:

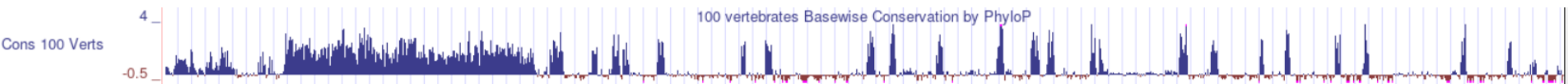
- ✓ GENCODE V44 (96 results)
- ✓ hg38 Track Data (253 results)
  - ✓ Visible Tracks (10 results)
    - ✓ RefSeq Curated (10 results)
  - ✓ Currently Hidden Tracks (243 results)
    - ✓ Genes and Gene Predictions (240 results)
      - ✓ RetroGenes V9 (2 results)
      - ✓ Other RefSeq (83 results)
      - ✓ MANE (13 results)
      - ✓ IKMC Genes Mapped (15 results)
      - ✓ HGNC (16 results)
      - ✓ NCBI RefSeq (51 results)
      - ✓ GENCODE Versions (8 results)
    - ✓ Phenotype and Literature (1 results)
      - ✓ GeneReviews (1 results)
    - ✓ mRNA and EST (2 results)
      - ✓ Human mRNAs (2 results)
  - ✓ Public Hubs (74 results)

# 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

# Genome Browser



Можно в строке поиска ввести локус в формате  
chrN:start-end

# Genome Browser

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

The screenshot displays the Genome Browser interface, organized into three main sections, each with a 'refresh' button in the top right corner.

- Mapping and Sequencing:** This section contains tracks for Base Position (dense), P14 Fix Patches (pack), P14 Alt Haplotypes (pack), Assembly (hide), Centromeres (hide), Chromosome Band (hide), Clone Ends (hide), Exome Probesets (hide), FISH Clones (hide), Gap (hide), GC Percent (hide), GRC Contigs (hide), GRC Incident (hide), Hg19 Diff (hide), INSDC (hide), LiftOver & ReMap (hide), LRG Regions (hide), Mappability (hide), Problematic Region (New, hide), Recomb Rate (hide), RefSeq Acc (hide), Restr Enzymes (hide), Scaffolds (hide), and Short Match (hide).
- Genes and Gene Predictions:** This section contains tracks for GENCODE V4 (Updated, dense), NCBI RefSeq (dense), CCDS (hide), CRISPR Targets (hide), GENCODE Version (Updated, s, hide), HGNC (hide), IKMC Genes Mapped (19, d, hide), LRG Transcripts (hide), MANE (full, hide), MGC Genes (hide), Non-coding RNA (hide), Old UCSC Genes (hide), ORFeome Clones (hide), Other RefSeq (hide), Pfam in GENCODE (hide), and Prediction Archive (hide).
- Phenotype and Literature:** This section contains tracks for OMIM Alleles (dense), CADD (hide), Cancer Gene Expr (hide), ClinGen (hide), ClinGen CNVs (hide), ClinVar Variants (hide), Constraint scores (hide), and Coriell CNVs (19, hide), New COSMIC (hide), COSMIC Regions (hide), DECIPHER CNVs (hide), DECIPHER SNVs (hide), Development Delay (hide), GenCC (hide), Gene Interactions (hide), GeneReviews (hide), GWAS Catalog (hide), HGMD\_public (hide), LOVD Variants (hide), OMIM Cyto Loci (hide), OMIM Genes (dense), Orphanet (hide), PanelApp (hide), and REVEL Scores (19, hide), SNPedia (hide), TCGA Pan-Cancer (hide), UniProt Variants (hide), and Variants in Papers (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)

|               |         | Category            | # of Genes | Example Genes   |
|---------------|---------|---------------------|------------|---|
| Total genes   | 466,332 |                     |            |   |
| HGNC approved | 43,718  | Protein-coding      | 21,652     | MTOR FGFR2 RET RAF1 MET MAP2K2 MAP2K1   |
| Disease genes | 20,000  | ncRNA genes         | 291,346    |   |
| Hot genes     | 500     | 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



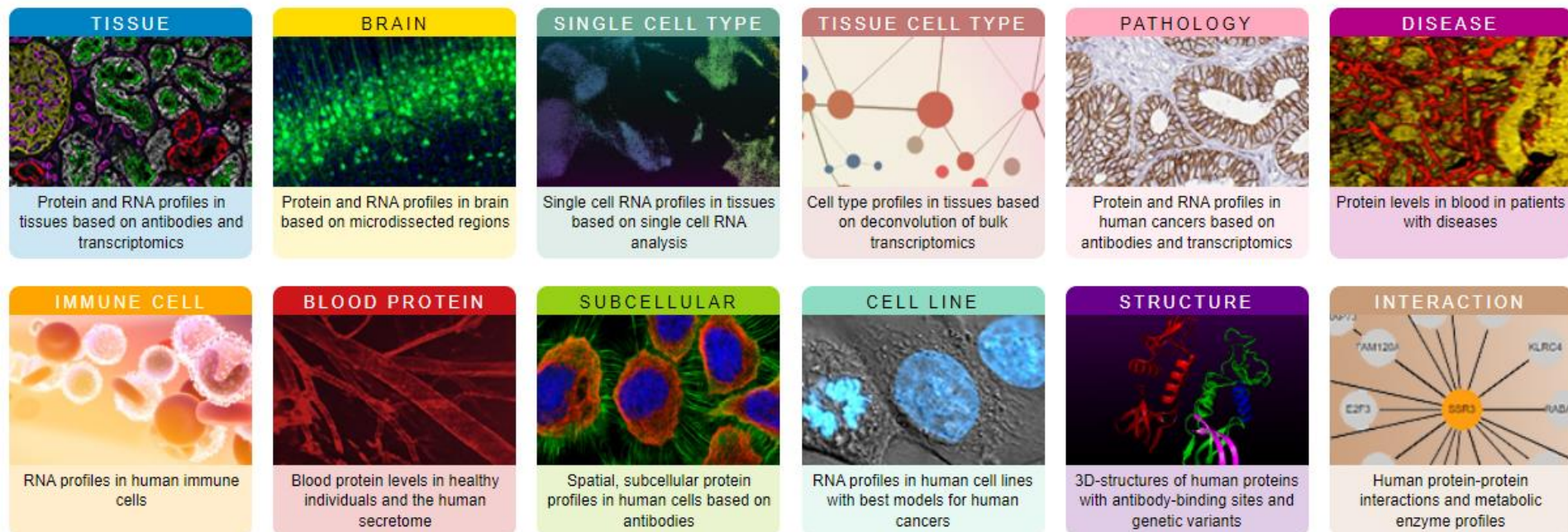
## GeneCards Sections

- 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:



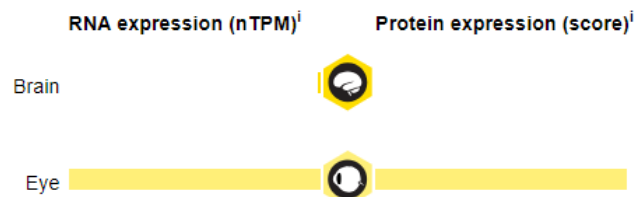
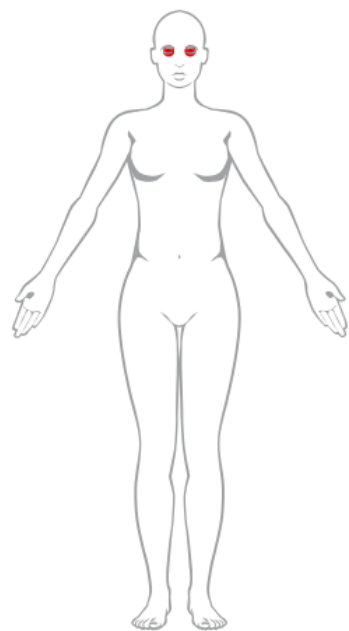
# The human protein atlas

| Gene  | RNA category human  | RNA category pig/mouse  | Annotation  |
|---|---|---|---|
| <input checked="" type="checkbox"/> Gene <sup>i</sup>               | <input type="checkbox"/> RNA tissue specificity <sup>i</sup>            | <input type="checkbox"/> RNA mouse brain regional specificity <sup>i</sup>  | <input type="checkbox"/> Antibody ID <sup>i</sup>                     |
| <input type="checkbox"/> Gene synonym <sup>i</sup>                  | <input type="checkbox"/> RNA tissue distribution <sup>i</sup>           | <input type="checkbox"/> RNA mouse brain regional distribution <sup>i</sup> | <input type="checkbox"/> Reliability (IH) <sup>i</sup>                |
| <input type="checkbox"/> Ensembl gene id <sup>i</sup>               | <input type="checkbox"/> RNA tissue specificity score                   | <input type="checkbox"/> RNA mouse brain regional specificity score         | <input type="checkbox"/> Reliability (Mouse Brain) <sup>i</sup>       |
| <input checked="" type="checkbox"/> Gene description <sup>i</sup>   | <input type="checkbox"/> RNA tissue specific nTPM                       | <input type="checkbox"/> RNA mouse brain regional specific nTPM             | <input type="checkbox"/> Reliability (IF) <sup>i</sup>                |
| <input type="checkbox"/> Uniprot accession                          | <input type="checkbox"/> RNA tissue nTPM max in non-specific            | <input type="checkbox"/> RNA pig brain regional specificity <sup>i</sup>    | <input type="checkbox"/> IH abundance (Normal Tissue) <sup>i</sup>    |
| <input type="checkbox"/> Chromosome                                 | <input type="checkbox"/> RNA single cell type specificity <sup>i</sup>  | <input type="checkbox"/> RNA pig brain regional distribution <sup>i</sup>   | <input type="checkbox"/> Subcellular location <sup>i</sup>            |
| <input type="checkbox"/> Chromosome position <sup>i</sup>           | <input type="checkbox"/> RNA single cell type distribution <sup>i</sup> | <input type="checkbox"/> RNA pig brain regional specificity score           | <input type="checkbox"/> Secretome location <sup>i</sup>              |
| <input type="checkbox"/> Protein class <sup>i</sup>                 | <input type="checkbox"/> RNA single cell type specificity score         | <input type="checkbox"/> RNA pig brain regional specific nTPM               | <input type="checkbox"/> Secretome function <sup>i</sup>              |
| <input type="checkbox"/> Biological process <sup>i</sup>            | <input type="checkbox"/> RNA single cell type specific nTPM             |   | <input type="checkbox"/> Cell Cycle Dependent Protein <sup>i</sup>    |
| <input type="checkbox"/> Molecular function <sup>i</sup>            | <input type="checkbox"/> RNA cancer specificity <sup>i</sup>            |   | <input type="checkbox"/> Cell Cycle Dependent Transcript <sup>i</sup> |
| <input type="checkbox"/> Disease involvement <sup>i</sup>           | <input type="checkbox"/> RNA cancer distribution <sup>i</sup>           |   | <input type="checkbox"/> Cancer prognostic p-value <sup>i</sup>       |
|   | <input type="checkbox"/> RNA cancer specificity score                   |   | <input type="checkbox"/> Blood expression cluster                     |
|   | <input type="checkbox"/> RNA cancer specific FPKM                       |   | <input type="checkbox"/> Tissue expression cluster                    |
| <b>Evidence</b>   | <input type="checkbox"/> RNA brain regional specificity <sup>i</sup>    |   | <input type="checkbox"/> Brain expression cluster                     |
| <input checked="" type="checkbox"/> Evidence (summary) <sup>i</sup> | <input type="checkbox"/> RNA brain regional distribution <sup>i</sup>   |   | <input type="checkbox"/> Cell line expression cluster                 |
| <input type="checkbox"/> HPA evidence                               | <input type="checkbox"/> RNA brain regional specificity score           |   | <input type="checkbox"/> Single cell expression cluster               |
| <input type="checkbox"/> UniProt evidence                           | <input type="checkbox"/> RNA brain regional specific nTPM               |   | <input type="checkbox"/> Num protein interactions                     |
| <input type="checkbox"/> NeXtProt evidence                          | <input type="checkbox"/> RNA blood cell specificity <sup>i</sup>        |   |   |
| <b>Atlas</b>  | <input type="checkbox"/> RNA blood cell distribution <sup>i</sup>       |   |   |
| <input checked="" type="checkbox"/> Tissue <sup>i</sup>             | <input type="checkbox"/> RNA blood cell specificity score               |   |   |
| <input checked="" type="checkbox"/> Brain <sup>i</sup>              | <input type="checkbox"/> RNA blood cell specific nTPM                   |   |   |
| <input checked="" type="checkbox"/> Single cell type <sup>i</sup>   | <input type="checkbox"/> RNA blood lineage specificity <sup>i</sup>     |   |   |
| <input checked="" type="checkbox"/> Tissue cell type <sup>i</sup>   | <input type="checkbox"/> RNA blood lineage distribution <sup>i</sup>    |   |   |
| <input checked="" type="checkbox"/> Pathology <sup>i</sup>          | <input type="checkbox"/> RNA blood lineage specificity score            |   |   |
| <input checked="" type="checkbox"/> Disease <sup>i</sup>            | <input type="checkbox"/> RNA blood lineage specific nTPM                |   |   |
| <input checked="" type="checkbox"/> Immune cell <sup>i</sup>        | <input type="checkbox"/> RNA cell line specificity <sup>i</sup>         |   |   |
| <input checked="" type="checkbox"/> Blood <sup>i</sup>              | <input type="checkbox"/> RNA cell line distribution <sup>i</sup>        |   |   |
| <input checked="" type="checkbox"/> subcellular <sup>i</sup>        | <input type="checkbox"/> RNA cell line specificity score                |   |   |
| <input checked="" type="checkbox"/> Cell line <sup>i</sup>          | <input type="checkbox"/> RNA cell line specific nTPM                    |   |   |
| <input checked="" type="checkbox"/> Structure <sup>i</sup>          | <input type="checkbox"/> RNA tissue cell type enrichment                |   |   |
| <input checked="" type="checkbox"/> Interaction <sup>i</sup>        |   |   |   |

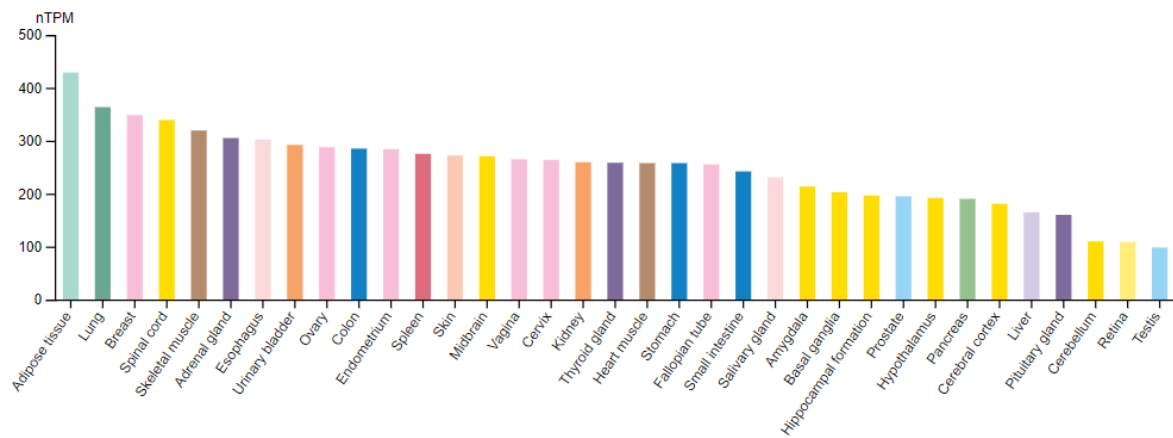
# The human protein atlas

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

## RNA AND PROTEIN EXPRESSION SUMMARY<sup>i</sup>

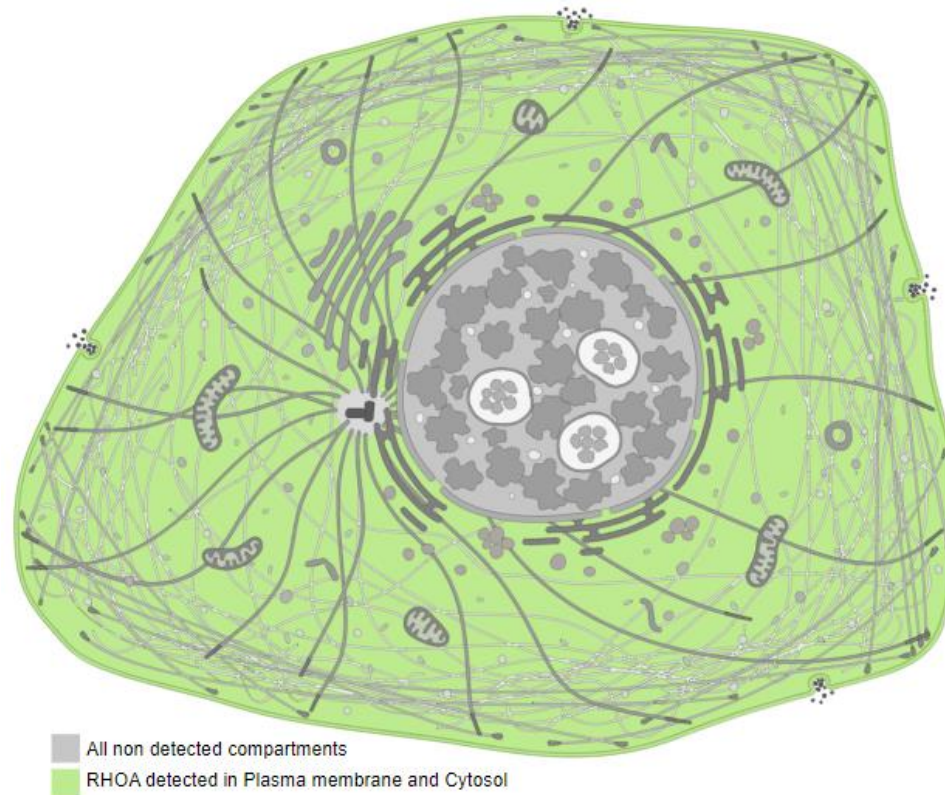


## GTEx dataset<sup>i</sup>



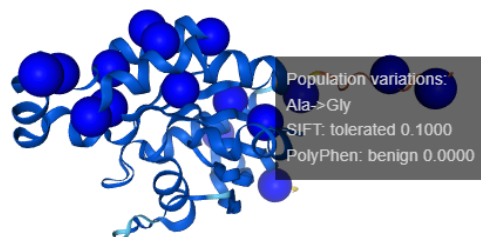
# The human protein atlas

- Субклеточная локализация белка



# The human protein atlas

- Структура белка с популяционными и клиническими вариациями



#### Description:

Structure prediction of P61586 from AlphaFold project, version 2

#### Color scheme:

Confidence  Residue index  Your selection

#### Variants:

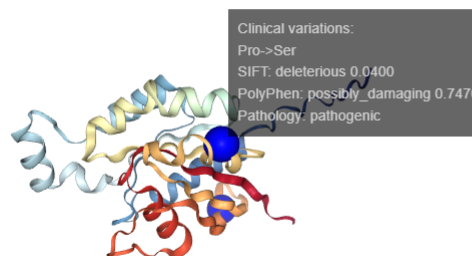
Off  Clinical (#=2)  Population (#=17)

#### Autorotate:

Off  On

#### Confidence for predicted structure:

Very high (nl DDT > 90)



#### Description:

Structure prediction of P61586 from AlphaFold project, version 2

#### Color scheme:

Confidence  Residue index  Your selection

#### Variants:

Off  Clinical (#=2)  Population (#=17)

#### Autorotate:

Off  On

И многое другое!

# UniProtKB

## Retrieve/ID mapping

Сервис позволяет перевести список ID из одной номенклатуры в другую

### Retrieve/ID mapping

Enter one or more IDs (100,000 max). You may also [load from a text file](#). Separate IDs |

P31946 P62258 ALBU\_HUMAN EFTU\_ECOLI

From database

UniProtKB AC/ID ▾

To database

UniProtKB ▾

Name your ID Mapping job

"my job title"

# Типы РНК в клетке

- Какие вы знаете?
- По какому принципу может идти деление?



# Типы РНК в клетке

- Тотальная РНК
- полиА
- Без фракции рРНК
- По размеру
  - Малые: микроРНК, пиРНК, мяРНК, ...
- По внутриклеточной локализации
  - Ядерные
  - Цитоплазматические
  - Ассоциированные с хроматином

# Задача

- Секвенировать транскриптом
- Но в секвенатор нужно подавать ДНК-библиотеку
- Как быть?

# Процесс

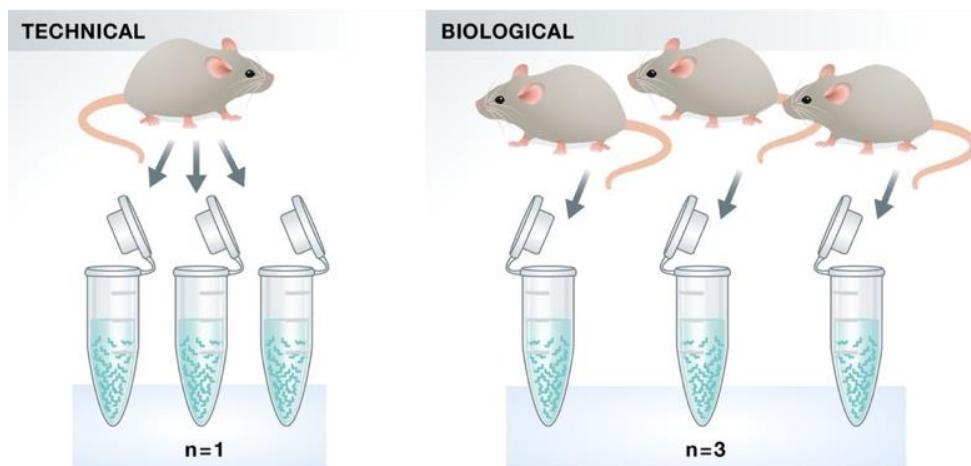
- Подготовка нужной фракции РНК
- Проверка качества РНК
- Обратная транскрипция => кДНК
- Фрагментация (~200-300 нукл)
- Секвенирование (чем глубже, тем лучше)

# Реплики

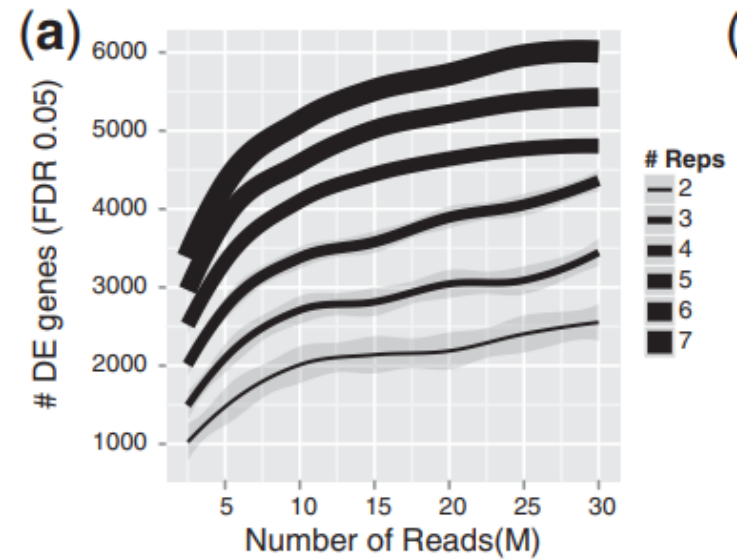
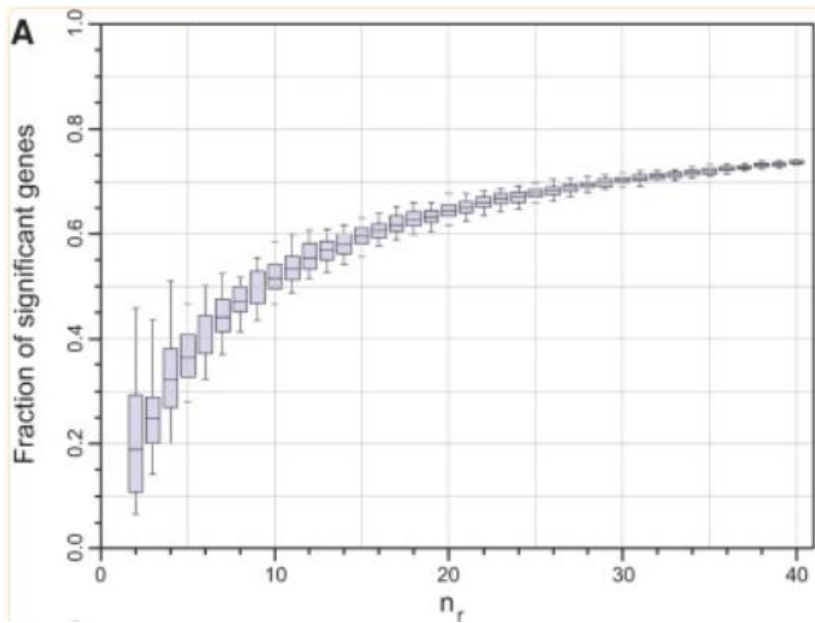
- Что это такое?
- Какие бывают?
- Зачем нужны?

# Реплики

- Технические реплики – повторный анализ одного и того же образца
- Биологические реплики – такое же исследование аналогичного образца

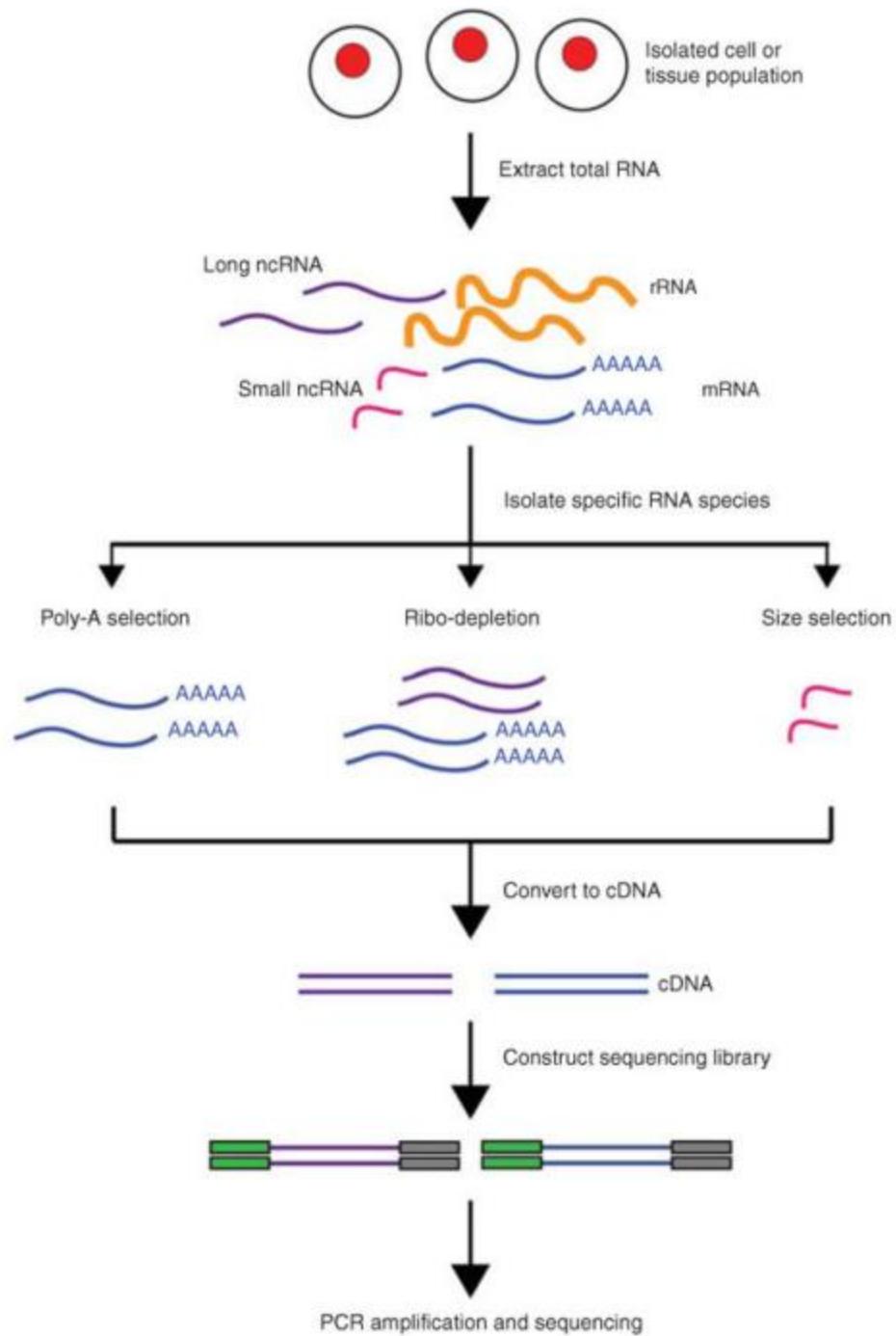


# Сколько нужно реплик?

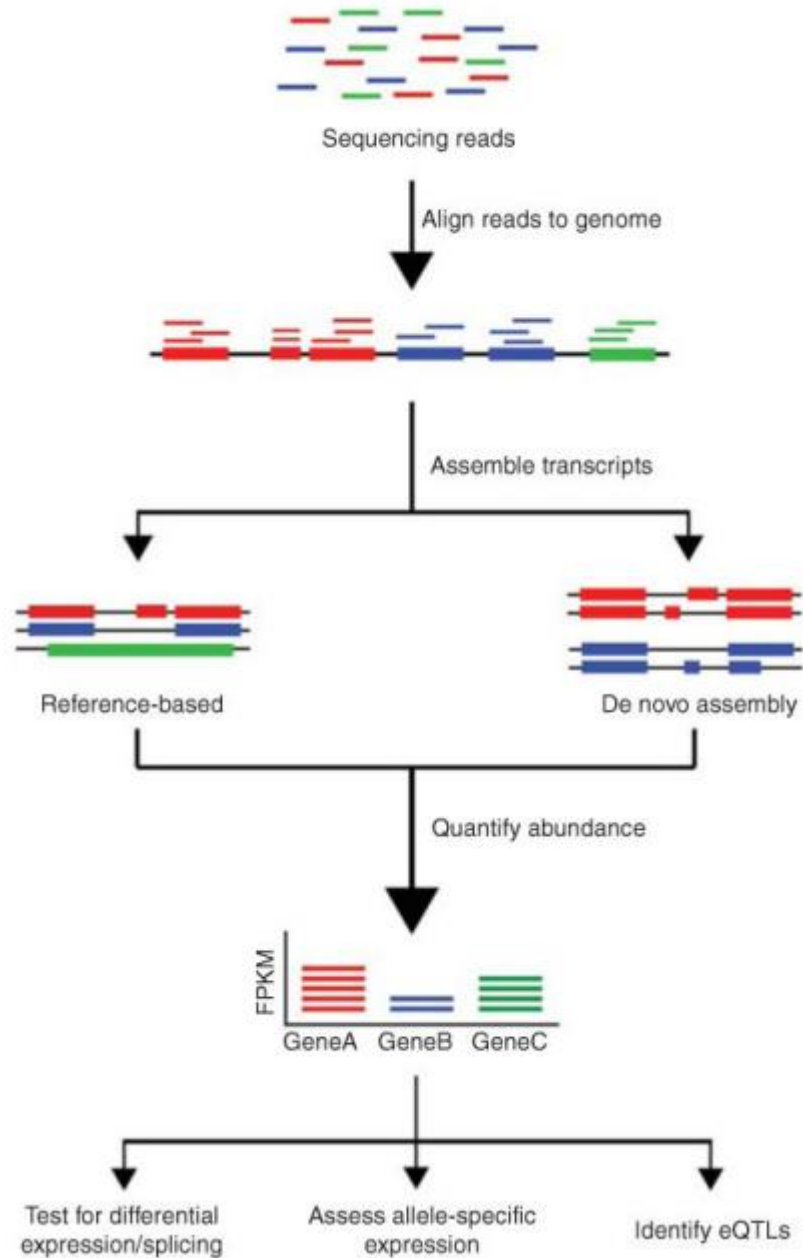


[ref](#)

[ref](#)



[ref](#)



[ref](#)

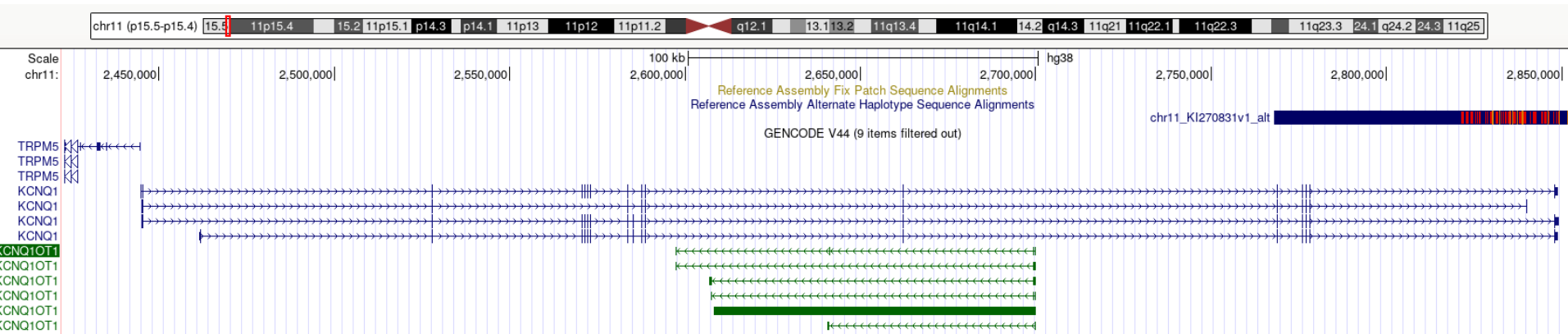


# Цепь-специфичные библиотеки

- При секвенировании сохраняем информацию о том, с какой цепи ДНК шла транскрипция
- ЗАЧЕМ?

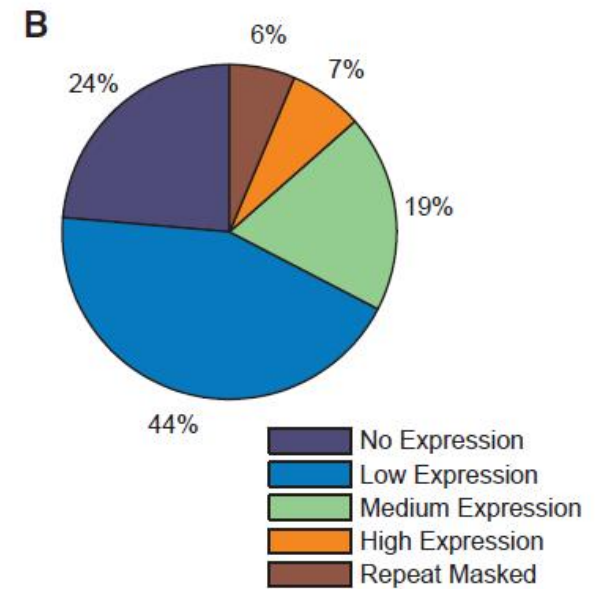
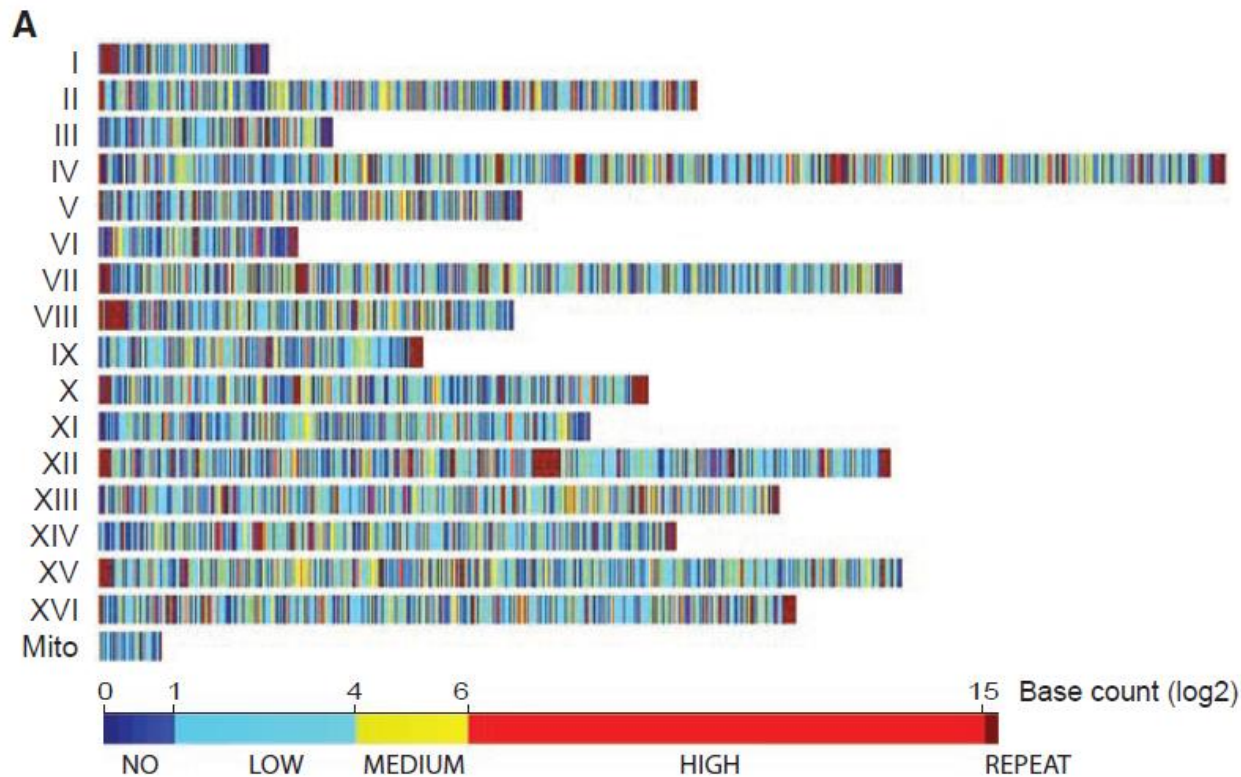
# Цепь-специфичность

- [МУЛЬТИК](#)

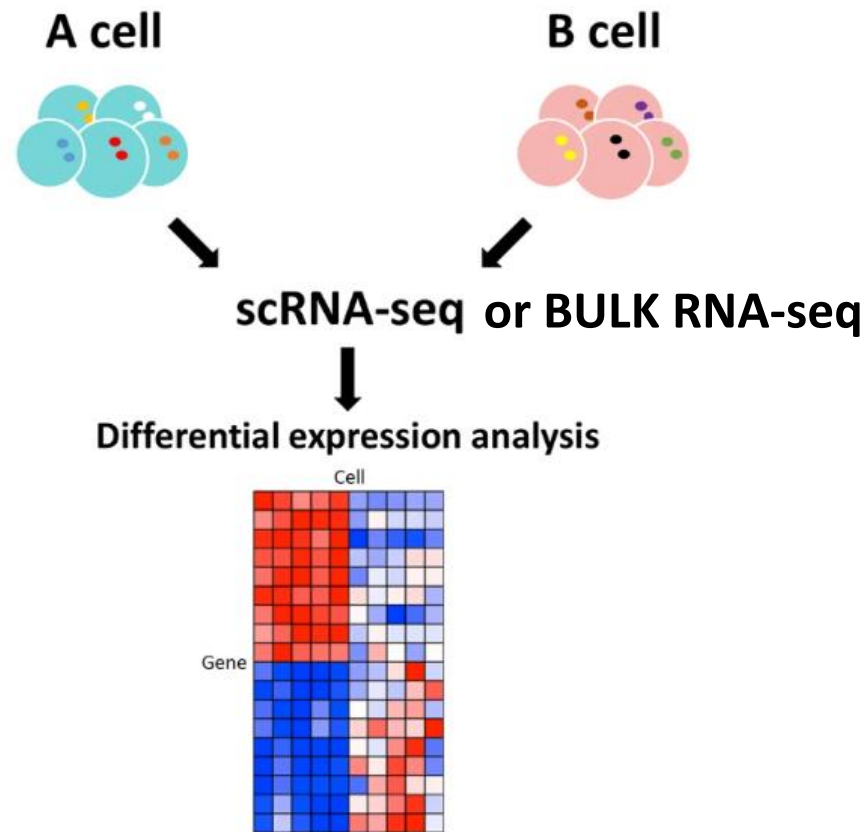


# Анализ транскриптомов

- Первые работы – 2008 год - дрожжи



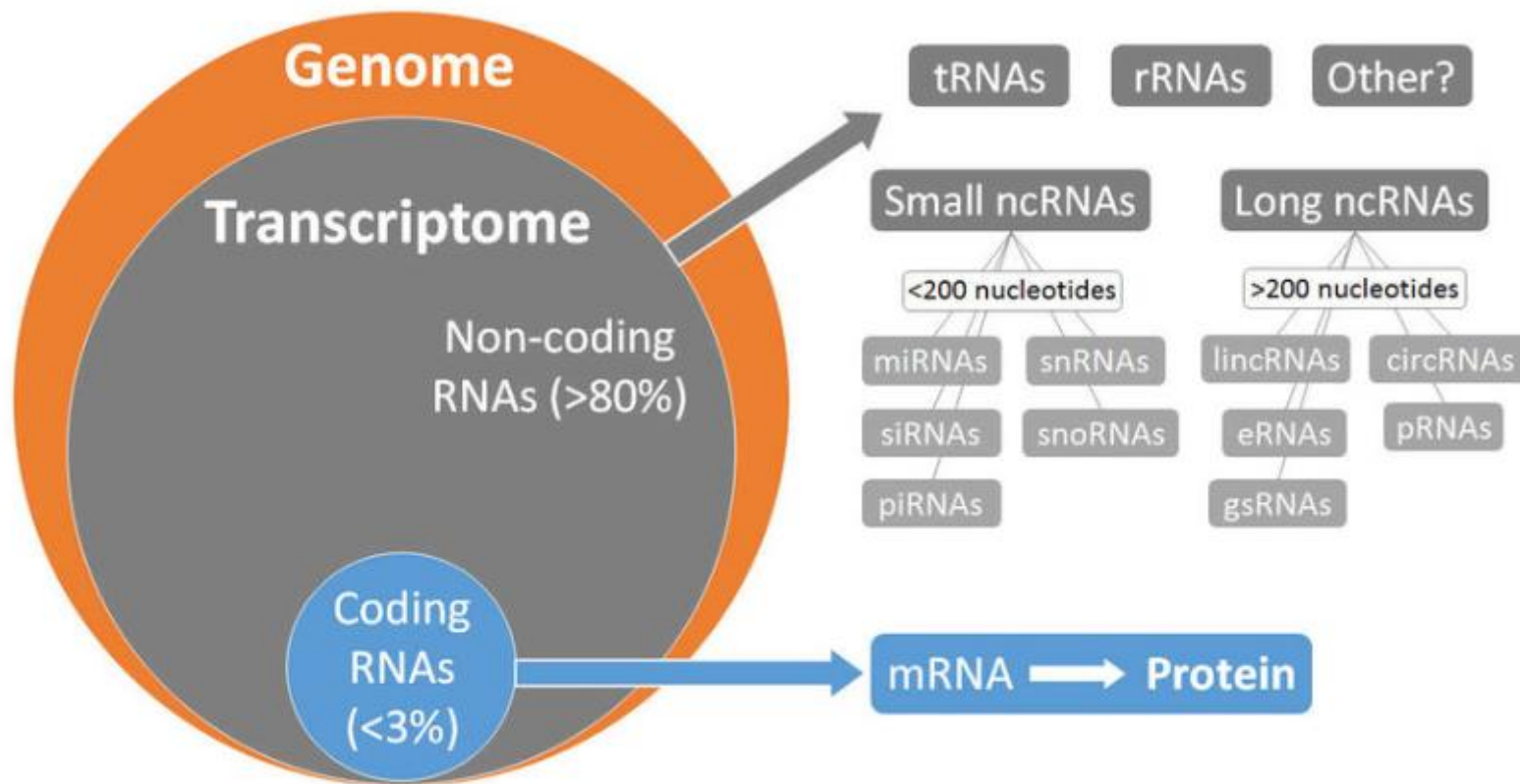
# Дифференциальная экспрессия



# Альтернативный сплайсинг

- АРОВ-100 – длинный транскрипт, синтезируется в печени
- АРОВ-48 – короткий транскрипт – синтезируется в кишечнике
- Синтезирующиеся белки входят в состав разных групп липопротеинов, которые затем идут каждый своим путем метаболизма

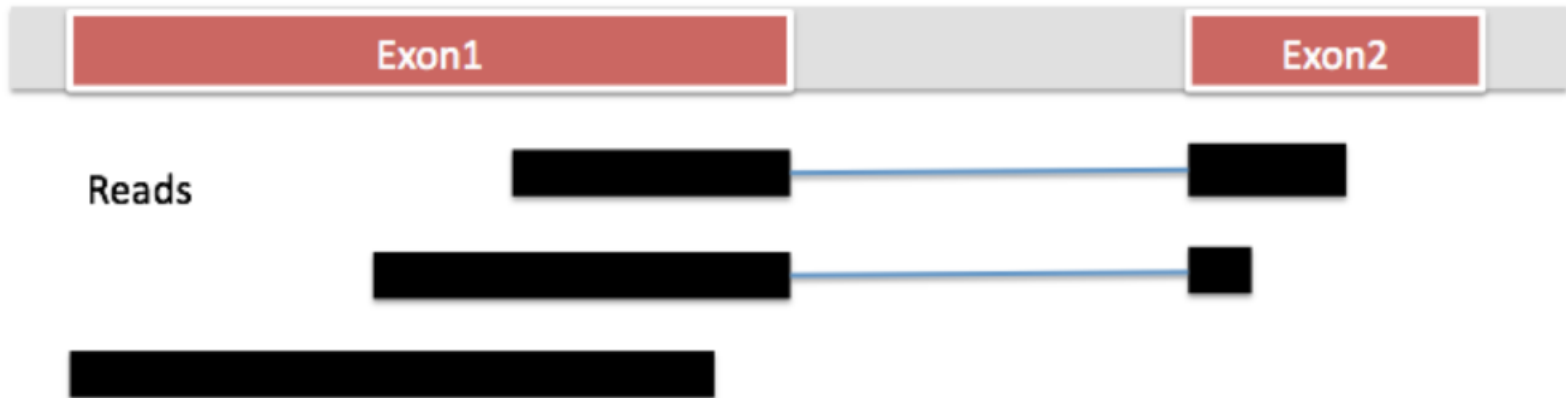
# Некодирующие РНК



# Картирование

- Чем принципиально отличается картирование чтений экзомного секвенирования и РНК-секвенирования?

# HISAT2





# Аннотация

- Нужно каждому гену приписать количество чтений, которые на него попали

- Книжка