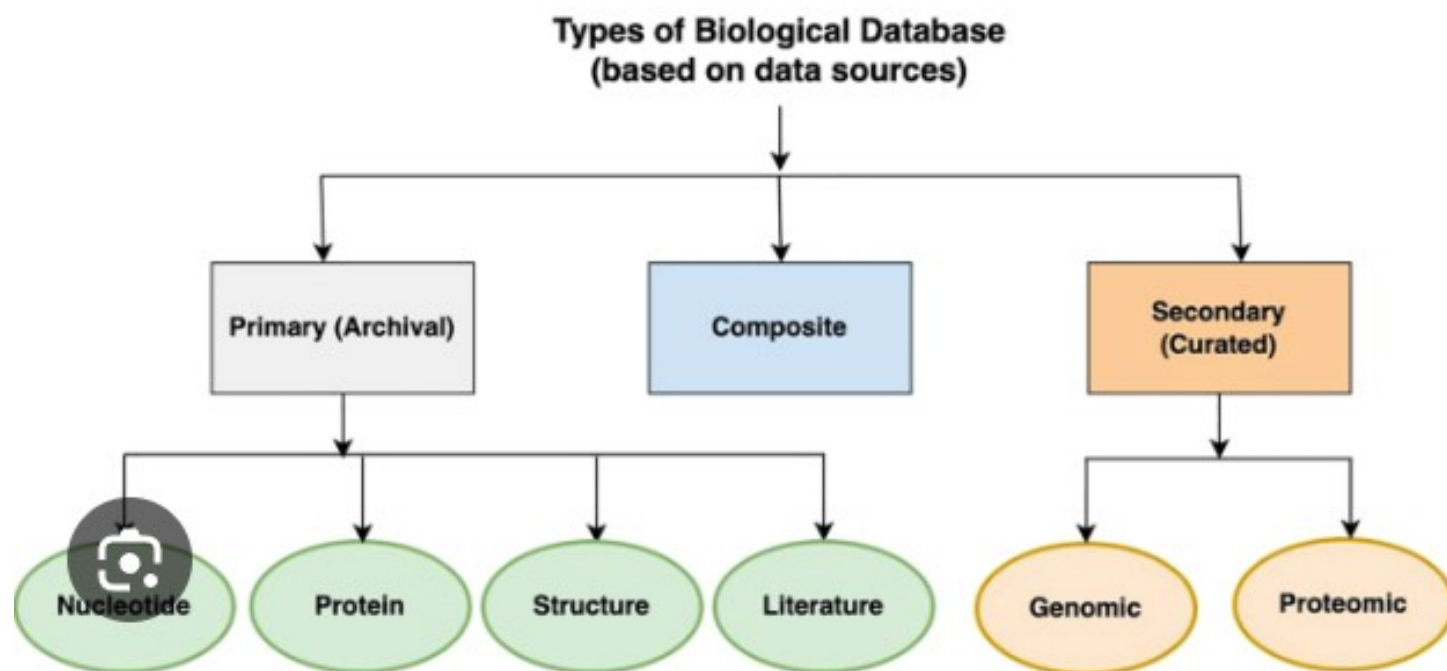


# **Биологиялық деректер және мәліметтер базалары (miRNA target tools)**

**2-Дәріс**

## Биологиялық деректер дегеніміз не?

- Биологиялық деректер – организмнің молекулалық құрылымдары мен процестері туралы ақпарат (ДНҚ, РНҚ, ақуыздар, метаболиттер және т.б.).
- Зерттеу деректері үлкен көлемде сақталып, тиімді талдау үшін мәліметтер базаларына орналастырылады.



**Нуклеотидті.** ДНҚ және РНҚ тізбегі сияқты нуклеотидтер тізбегін сақтайтын негізгі дерекқорларға жатады. Бұл дерекқорларда көбінесе өңделмеген бастапқы деректер бар және олар генетикалық және геномдық зерттеулер үшін өте маңызды.

# Протеомдық деректер

**Протеомдық деректер** – бұл организмдегі барлық ақуыздар туралы мәліметтерді қамтитын биологиялық деректер. Протеомика ақуыздардың құрылымын, функциясын және өзара әрекеттесуін зерттейтін ғылым саласы. Протеомдық деректер ақуыздардың аминқышқыл реттілігін, олардың үшөлшемді құрылымдарын, функционалдық аймақтарын, сондай-ақ олардың жасушалық процестердегі рөлін қамтиды.

## **Протеомдық деректердің түрлері:**

- 1.Аминқышқыл реттіліктері:** Ақуыздардың аминқышқылдарынан тұратын реттіліктері.
- 2.Ақпараттық сипаттама:** Ақуыздардың биологиялық рөлі, жасушадағы қызметі.
- 3.Үшөлшемді құрылымдар:** Ақуыздардың кеңістіктегі құрылымы, олардың функционалдық ерекшеліктерін түсінуге көмектеседі.
- 4.Функционалдық аймақтар:** Ақуыздың белсенді аймақтары мен олардың басқа молекулалармен байланысу орындары.
- 5.Ақуыздардың өзара әрекеттесуі:** Ақуыздар арасындағы байланыстар мен олардың бір-бірімен қалай әрекеттесетіні туралы деректер.

- Протеомдық деректер базалары:

## 1. UniProt (Universal Protein Resource)

**1. Мазмұны:** Ақуыз реттіліктері мен олардың функциялары туралы толық ақпарат. Ғылыми қоғамдастыққа ақуыз реттілігі мен функционалдық ақпараттың жан-жақты, жоғары сапалы және еркін қол жетімді ресурсын қамтамасыз ету.

**2. Сипаттама:** UniProt – ақуыз реттіліктері мен олардың сипаттамаларын қамтитын ең ірі және сенімді дерекқор.

**3. Сайт:** [UniProt](https://www.uniprot.org/)



# Protein Data Bank (PDB)

**Мазмұны:** Ақуыздардың үшөлшемді құрылымдары.

**Сипаттама:** PDB ақуыздардың және басқа биомолекулалардың үшөлшемді құрылымдарын сақтайды. Бұл дерекқор зерттеушілерге ақуыздардың құрылымын көруге және талдауға мүмкіндік береді.

**Сайт:** [PDB](https://www.rcsb.org/)

The screenshot shows the PDB website homepage. At the top, there is a navigation bar with logos for PDB-101, PDB, EMDataResource, NAKB, wwPDB Foundation, and PDB-Dev, along with social media icons for Facebook, Twitter, YouTube, and LinkedIn. Below the navigation bar is a teal banner with the text "Access Computed Structure Models (CSMs) of available model organisms" and a "Learn more" button. On the left side, there is a dark blue sidebar menu with the following items: Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area features a paragraph about the PDB's mission: "RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:" followed by two bullet points: "Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive" and "Computed Structure Models (CSM) from AlphaFold DB and ModelArchive". Below this, it states: "These data can be explored in context of external annotations providing a structural view of biology." At the bottom of the main content area, there is a banner for "Explore NEW Features" and "PDB-101 Training Resources". On the right side, there is a section titled "September Molecule of the Month" featuring a 3D molecular model of a protein structure in green.

# PRIDE (PRoteomics IDentifications Database)

- **Мазмұны:** Ақуыздардың масс-спектрометриялық мәліметтері мен олардың экспрессиясы туралы ақпарат.
- **Сипаттама:** PRIDE дерекқоры ақуыздардың масс-спектрометриялық деректерін және олардың тіңдердегі немесе жасушалардағы экспрессия деңгейлерін сақтайды.
- **Сайт:** PRIDE

The image shows a search interface on the PRIDE website. At the top, the word "Search" is displayed. Below it is a search input field containing the text "PXD005011" followed by a small "x" icon and the placeholder text "input here". Below the search field is a filter section with a dropdown menu labeled "Submission Date" and a "Select" button.

# PeptideAtlas

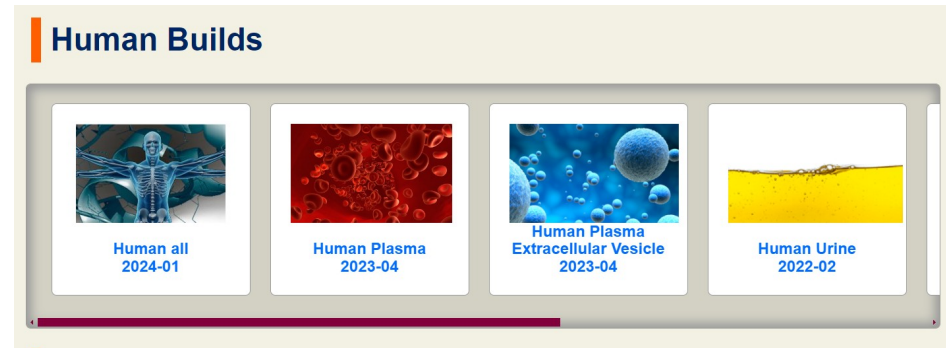
**Мазмұны:** Пептидтер туралы ақпарат.

**Сипаттама:** PeptideAtlas пептидтердің эксперименттік мәліметтерін жинақтайды және ақуыздардың толық сипаттамасын беруге көмектеседі. PeptideAtlas – көп ағзалы, жалпыға қолжетімді пептидтер жинағы, тандемдік масс-спектрометрлік протеомика эксперименттерінің үлкен жинағында анықталған. Масс-спектрометрдің шығыс файлдары адам, тінтуір, ашытқы және бірнеше басқа организмдер үшін жиналады және соңғы іздеу жүйелері мен ақуыз тізбектері арқылы ізделеді.

**Сайт:** [PeptideAtlas](https://www.peptideatlas.org/)



**PeptideAtlas** is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences. All results of sequence and spectral library searching are subsequently processed through the [Trans Proteomic Pipeline](#) to derive a probability of correct identification for all results in a uniform manner to insure a high quality database, along with false discovery rates at the whole atlas level. Results may be queried and browsed at the PeptideAtlas web site. The raw data, search results, and full builds can also be downloaded for other uses.

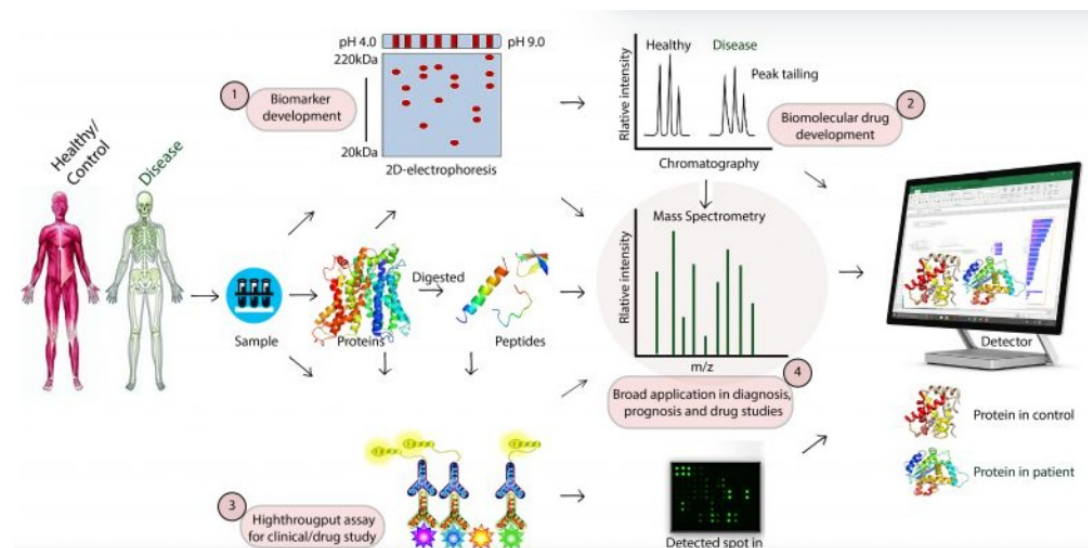




# Протеомдық деректердің маңыздылығы:

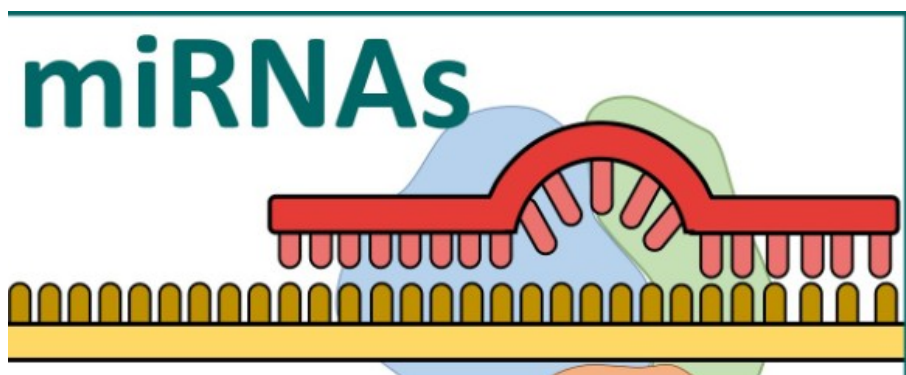
- **Ауруларды зерттеу:** Протеомдық деректер аурулардың дамуын, олардың диагностикасы мен емдеу әдістерін түсінуге көмектеседі.
- **Дәрілерді әзірлеу:** Ақуыздардың құрылымы мен функцияларын білу арқылы дәрілік препараттарды дәл бағыттауға болады.
- **Эволюцияны зерттеу:** Ақуыздардың эволюциялық өзгерістері мен олардың түрлер арасындағы айырмашылықтарын талдау.

Протеомдық деректер биологиялық зерттеулерде, ауруларды диагностикалауда және жаңа терапияларды әзірлеуде маңызды рөл атқарады.





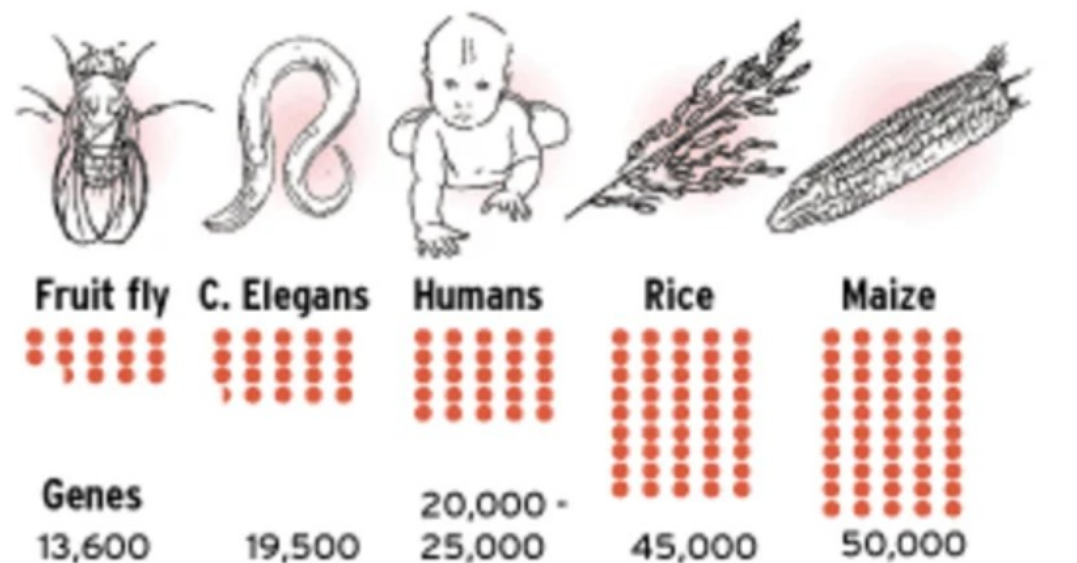
# МикроРНК молекуласының нысанандары: болжау құралдарынан эксперименттік валидацияға дейін (miRNA Targets: From Prediction Tools to Experimental Validation)



**МикроРНК** — ұзындығы шамамен 20-25 нуклеотид болатын қысқа РНК молекулалары, олар мРНК (матрицалық РНК) молекулаларымен байланысып, олардың аудармасын басқарады немесе оларды деградацияға ұшыратады. МиРНК-лар гендердің экспрессиясын реттеуде маңызды рөл атқарады және көптеген биологиялық процестерге қатысады, соның ішінде даму, жасушаның бөлінуі, апоптоз және стресс жағдайларына жауап беру. МиРНК-лар рак, нейродегенеративті аурулар, жүрек-қан тамырлары аурулары сияқты патологиялық жағдайлармен байланысты.

## Humans have fewer genes

In Thursday's issue of the journal *Nature*, researchers who decoded the human genome concluded that people have only 20,000 to 25,000 genes, a drop from the 30,000 to 40,000 estimated in 2001.



SOURCE: *Nature*

AP

*Nature* журналының бейсенбілік санында адам геномын ашқан зерттеушілер адамдарда тек 20 000-нан 25 000-ға дейін ген бар екенін анықтады, бұл 2001 жылы бағаланған 30 000-нан 40 000-ға дейінгі саннан аз.

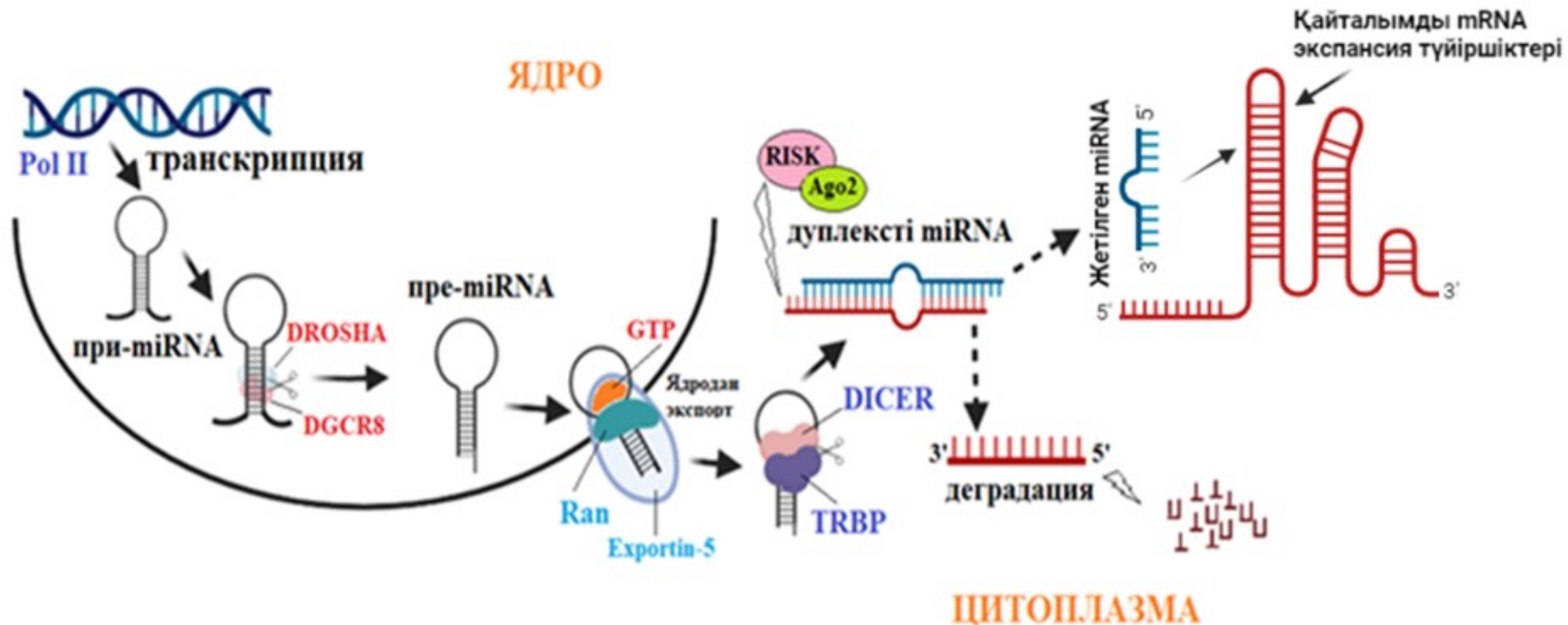
### Гендер:

- Жеміс шыбыны: 13 600
- *C. Elegans* (нематод): 19 500
- Адамдар: 20 000 - 25 000
- Күріш: 45 000
- Жүгері: 50 000

Дереккөз: *Nature*



# miRNA биогенезі және нысаны



# miRNA target prediction



## TargetScan

- Сипаттамасы:** МиРНҚ-дың жануарлар мен адам мРНҚ-ларымен байланысуын болжайтын ең танымал құралдардың бірі. Ол миРНҚ-ның "seed region" комплементарлығын және эволюциялық консервацияны негізге ала отырып, ықтимал мақсатты гендерді анықтайды.
- Сайт:** <http://www.targetscan.org/>

## miRDB

- Сипаттамасы:** МиРНҚ мақсаттарын болжауға арналған дерекқор. Ол эксперименттік деректерге негізделген миРНҚ-ның ықтимал мақсатты гендерін табу үшін ML алгоритмдерін пайдаланады.
- Сайт:** <http://www.mirdb.org/>





## miRTarBase

• **Сипаттамасы:** МиРНҚ-ның мРНҚ-ға әсер етуі туралы эксперименттік дәлелдер жинағы. Бұл дерекқорда миРНҚ-ның мақсатты гендеріне байланысты расталған эксперименттік деректер бар.

• **Сайт:** <http://mirtarbase.mbc.nctu.edu.tw/Diana-microT-CDS>

• **Сипаттамасы:** МиРНҚ-ның мРНҚ-мен өзара әрекеттесуін болжайтын құрал. Ол мРНҚ реттілігінен басқа, кодтау аймағындағы болжамдарды да қамтиды.

• **Сайт:** <http://diana.imis.athena-innovation.gr/DianaTools/index>  
**RNAhybrid**

• **Сипаттамасы:** МиРНҚ және мРНҚ-ның жұптасу энергиясын есептейтін құрал. Ол миРНҚ мен мРНҚ-ның ықтимал байланыстырушы сайттарын бағалауға және олардың термодинамикалық тұрақтылығын анықтауға арналған.

• **Сайт:** <https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid>

• **Сипаттамасы:** МиРНҚ-лардың 3'-UTR, 5'-UTR және кодтау аймақтарымен байланысуын болжауға мүмкіндік беретін кешенді құрал.

• **Сайт:** <http://mirwalk.umm.uni-heidelberg.de/>



# RNAhybrid

miRWalk



## News and Updates:

- Aug/2024 - **New GSEA Module** - new gsea module is online.
- Jan/2024 - **Migration** - miRWalk database migration to a new server is completed.
- Oct/2023 - **bed\_format** - In addition to the csv format, the results can now also be saved in bed format.
- Jan/2022 - **release\_2022\_01** - 2022 release update with new features. Detail view on miRNA-Gene-Duplex. Disease ontology search.
- Jan/2021 - **new\_update\_2021** - Annual update for the year 2021 was completed. At the request of numerous users, the duplex information has been integrated and can now be saved.

[read more...](#)

## Migration Report

miRWalk is online again. The migration is almost complete. Some functions have not yet been activated, but will be available in the course of next week. Our team is working flat out to finalise the migration. We are also working on the destination prediction tools, which should be available in the course of the year. If any errors occur, please let us know.

Kind regards, the miRWalk team.

## Search for a single gene or miRNA

miRNAs: miRNA names (e.g. hsa-miR-214-3p) or Accession numbers (e.g. MIMAT0000271) based on current miRBase. While searching single miRNAs, also short names or family miRNA(e.g. let-7) belongs to several miRNAs are also acceptable. A list of miRNAs will be shown. mRNAs: Official Genesymbols (e.g. GAS2), EntrezIDs (e.g. 10608), Ensembl-IDs (e.g. ENSG00000148935 or ENST00000454584) and RefseqIDs (e.g. NM\_001143830) were accepted.

species

human



Gene

HTT

miRNA

search



## HTT

### Details

<b>Entrezid</b>	<a href="#">3064</a>
<b>Genesymbol</b>	<a href="#">HTT</a>
<b>Alias</b>	HD;IT15;LOMARS
<b>Description</b>	Homo sapiens huntingtin (HTT), transcript variant 1, mRNA.
<b>Chromosome</b>	4
<b>Ensemblid</b>	ENSG00000197386

### Transcripts

<a href="#">Refseq</a>	<a href="#">Ensemblid</a>	<a href="#">Length</a>	<a href="#">Startcds</a>	<a href="#">Endcds</a>
NM_001388492	ENST00000355072	13472	146	9574

Mirna <sup>△</sup>	Refseqid <sup>△</sup>	Genesymbol <sup>△</sup>	Duplex <sup>△</sup>	Score <sup>△</sup>	Position <sup>△</sup>	Binding Site <sup>△</sup>	Au <sup>△</sup>	Me <sup>△</sup>	N Pairings <sup>△</sup>	Targetscan <sup>△</sup>	Mirdb <sup>△</sup>	Mirtarbase <sup>△</sup>
hsa-let-7a-5p	NM_001388492	HTT	<a href="#">details</a>	1.00	CDS	3100,3118	0.66	-8.107	17	—	—	—
hsa-let-7a-5p	NM_001388492	HTT	<a href="#">details</a>	0.92	CDS	3773,3794	0.54	-6.251	18	—	—	—
hsa-let-7a-2-3p	NM_001388492	HTT	<a href="#">details</a>	1.00	3UTR	9804,9825	0.4	-5.692	18	—	—	—
hsa-let-7a-2-3p	NM_001388492	HTT	<a href="#">details</a>	1.00	3UTR	13013,13035	0.35	-13.114	17	—	—	—
hsa-let-7a-2-3p	NM_001388492	HTT	<a href="#">details</a>	1.00	CDS	3747,3771	0.47	-6.45	18	—	—	—
hsa-let-7b-5p	NM_001388492	HTT	<a href="#">details</a>	0.81	CDS	7047,7068	0.41	-7.128	19	—	—	<b>MIRT052338</b>
hsa-let-7b-3p	NM_001388492	HTT	<a href="#">details</a>	0.92	3UTR	11783,11846	0.47	-3.793	16	—	—	—
hsa-let-7b-3p	NM_001388492	HTT	<a href="#">details</a>	0.85	3UTR	12249,12264	0.57	-20.281	12	—	—	—
hsa-let-7c-5p	NM_001388492	HTT	<a href="#">details</a>	0.85	CDS	3084,3118	0.66	-6.515	19	—	—	—
hsa-let-7c-3p	NM_001388492	HTT	<a href="#">details</a>	0.95	CDS	6492,6507	0.48	-11.501	14	—	—	—

Export BED

Export CSV

Filter

## Details of Duplex

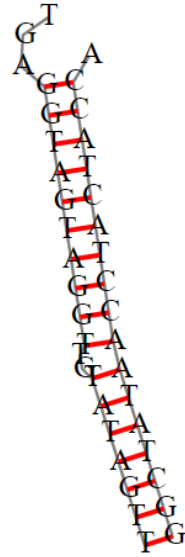
2D Structure dot-bracket from RNA duplex"

```
TGAGGTAGTAGGTTGTATAGTT&GGCTATAACCTACTACCA  
...(((((((((((...(((((((&)))))))))))))))).
```

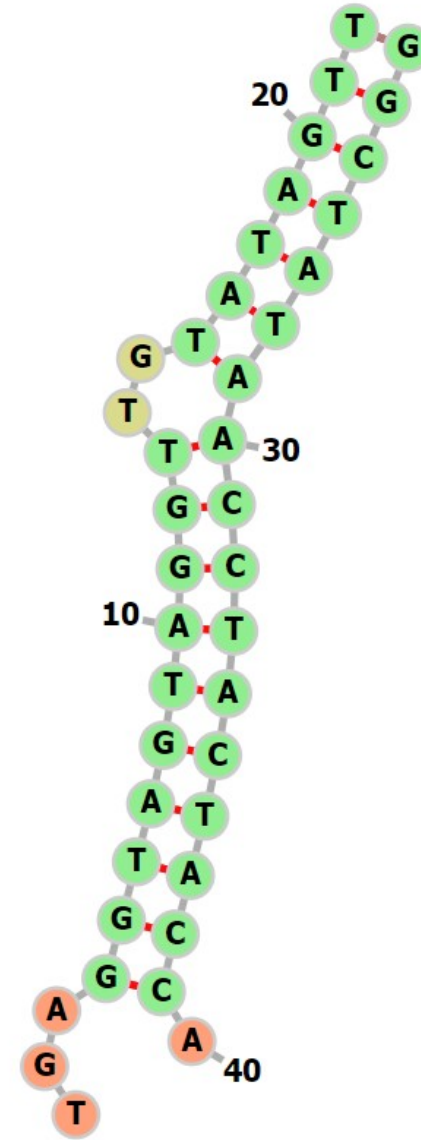
### Details

MIRNA	hsa-let-7a-5p
RefseqID	NM_001388492
GeneSymbol	<a href="#">HTT</a>
Binding site	3100,3118
Score	1.0
Seed	0
Au	0.662
Energy	-21.9
Binding len	18
Me	-8.107
Phylostem	0.0
Phyloflank	0.0
Longest consecutive pairings	17
Position of longest consecutive pairings	1
Pairings in 3prime end	7
Position	CDS

### 2D Structure



[show with forna](#)



## 2D Structure dot-bracket from RNAduplex"

TGAGGTAGTAGGTTGTATAGTT&GGCTATAACCTACTACCA  
...((((((((((..((((((&)))))))))))))))).

GGCTATAACCTACTACCAAGCATAACAGACGTCAG  
TGCAGCAGTTTCTCATGAACTAATCACATCAACCA  
AAGCTTTGTGTCTTCTTTCCACTGCCTTCCCAGTT  
GGAGTGCCTCCACTGAGTGCCTCAGATGAGTCTAC  
AATGATTCTGACCCTGCTCTCGTCAGCTTGGTTCC  
CTTTGATTTTGGCCGAAACTTGCTTGAGCCAGT  
GCCTCTGAAGAAGAAGCCAACCCAGCAGCCACCAZ  
GGACCGGGCCCTGGTGCCTCATGGTGGAGCAGCTC  
TTTGTGCCACGTCCTGGATGACGTGGCTCCTGGZ  
CTAACAAACCCCTTCTCTAAGTCCCATCCGACCA  
ACAAGCATCTGTACCGTTGAGTCCAAGAAAGGCAGT  
CTGATACCTCAGGTCTTGTACAACAAGTAAATCCTC  
CCTTCATACCTCAAACCTGCATGATGTCCTGAAAG  
GGATCTTCAGAACAGCACGGAAAAAGTTTGGAGG  
CTCAGATACTAGAGCTGGCCACACTGCAGGACAT  
TACCTGAAATCCTGCTTTAGTCGAGAACCAATG  
GAAGACTCTCTTTGGCACAACCTTGGCTCCCAG  
AGTCACAAGGCCGAGCACAGCGCCTTGGCTCCT  
TGCTTCATGGCCCCGTACACCCACTTCACCCAGG  
GGTGCAGGCGGAGCAGGAGAACGACACCTCGGG  
CCCAGTTGAAGACAAACCTCACGAGTGTCAAAA  
AATCACATTCGTTTGTGAACTCTTGTATAAAA  
ATGTGTGCAGTTACAGAAGCAGGTTTTAGATTT  
ATTACTGTCTTCTGGATTCAGATCAGGTGTTT  
ATTGAAGTGGGCCAGTTCAGGGAATCAGAGGCA  
ATTACTATCTTATGAACGCTATCATTCAAAAC  
TCTGTGATGGCATCATGGCCAGTGGAAAGGCT  
CCCATAGTCCACGACCTCTTTGTATTAAGAGGA  
AGCATAACAGACGTCAGTGCAGCTTCTAGACAAT  
CTGATACCTCAGGTCTTGTACAACAAGTAAATC  
CCTTCATACCTCAAACCTGCATGATGTCCTGAA  
GGATCTTCAGAACAGCACGGAAAAAGTTTGGAG  
CTCAGATACTAGAGCTGGCCACACTGCAGGACA  
TACCTGAAATCCTGCTTTAGTCGAGAACCAAT  
GAAGACTCTCTTTGGCACAACCTTGGCTCCCAG  
AGTCACAAGGCCGAGCACAGCGCCTTGGCTCCT  
TGCTTCATGGCCCCGTACACCCACTTCACCCAG  
GGTGCAGGCGGAGCAGGAGAACGACACCTCGGG  
CCCAGTTGAAGACAAACCTCACGAGTGTCAAAA  
AATCACATTCGTTTGTGAACTCTTGTATAAAA  
ATGTGTGCAGTTACAGAAGCAGGTTTTAGATTT  
ATTACTGTCTTCTGGATTCAGATCAGGTGTTT  
ATTGAAGTGGGCCAGTTCAGGGAATCAGAGGCA  
ATTACTATCTTATGAACGCTATCATTCAAAAC  
TCTGTGATGGCATCATGGCCAGTGGAAAGGCT  
CCCATAGTCCACGACCTCTTTGTATTAAGAGGA

Найти ✕

Что:

Только слово целиком

С учетом регистра

# HTT huntingtin [ *Homo sapiens* (human) ]

[Download Datasets](#)

Gene ID: 3064, updated on 8-Sep-2024

## Summary

<b>Official Symbol</b>	HTT provided by <a href="#">HGNC</a>
<b>Official Full Name</b>	huntingtin provided by <a href="#">HGNC</a>
<b>Primary source</b>	<a href="#">HGNC:HGNC:4851</a>
<b>See related</b>	<a href="#">Ensembl:ENSG00000197386</a> <a href="#">MIM:613004</a> ; <a href="#">AllianceGenome:HGNC:4851</a>
<b>Gene type</b>	protein coding
<b>RefSeq status</b>	REVIEWED
<b>Organism</b>	<a href="#">Homo sapiens</a>
<b>Lineage</b>	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
<b>Also known as</b>	HD; IT15; LOMARS
<b>Summary</b>	<p>Huntingtin is a disease gene linked to Huntington's disease, a neurodegenerative disorder characterized by loss of striatal neurons. This is thought to be caused by an expanded, unstable trinucleotide repeat in the huntingtin gene, which translates as a polyglutamine repeat in the protein product. A fairly broad range of trinucleotide repeats (9-35) has been identified in normal controls, and repeat numbers in excess of 40 have been described as pathological. The huntingtin locus is large, spanning 180 kb and consisting of 67 exons. The huntingtin gene is widely expressed and is required for normal development. It is expressed as 2 alternatively polyadenylated forms displaying different relative abundance in various fetal and adult tissues. The larger transcript is approximately 13.7 kb and is expressed predominantly in adult and fetal brain whereas the smaller transcript of approximately 10.3 kb is more widely expressed. The genetic defect leading to Huntington's disease may not necessarily eliminate transcription, but may confer a new property on the mRNA or alter the function of the protein. One candidate is the huntingtin-associated protein-1, highly expressed in brain, which has increased affinity for huntingtin protein with expanded polyglutamine repeats. This gene contains an upstream open reading frame in the 5' UTR that inhibits expression of the huntingtin gene product through translational repression. [provided by RefSeq, Jul 2016]</p>
<b>Expression</b>	Ubiquitous expression in brain (RPKM 9.1), skin (RPKM 8.2) and 25 other tissues <a href="#">See more</a>
<b>Orthologs</b>	<a href="#">mouse</a> <a href="#">all</a>



# Нуклеотидті қайталану бұзылыстарының дамуы

