

# Лекция №8(6)

## Часть 1. БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ (окончание)

## Часть 2. ТРАНСЛЯЦИЯ

к.б.н., с.н.с. лаб. эволюционной биоинформатики  
и теоретической генетики Игнатьева Е.В.

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

<b>UniProtKB</b> Switzerland	Protein knowledgebase	Geneva,
<b>TFClass</b>	Classification of transcription factors	Germany
<b>AnimalTFDB</b>	Animal Transcription Factor DataBase	Китай

**СЕГОДНЯ, в лекции № 8(6) будет дана характеристика баз**

**ДАННЫХ** (продолжение):

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

<b>CREMOFAC</b>	Database of chromatin remodeling factors	Индия
<b>TcoF- DB</b>	Dragon database of transcription co-factors and transcription factor interacting proteins	Королевство Саудовская Аравия

МАТРИЦЫ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

**TRANSFAC Matrix, JASPAR , HOCOMOCO , CIS-BP**

РЕГУЛЯТОРНЫЕ РАЙОНЫ, ССТФ, ТРАНСКРИПЦИОННЫЕ ФАКТОРЫ

<b>TRRD</b>	Transcription Regulatory Regions Database	ИЦиГ СО РАН, Новосибирск, Россия
-------------	---	----------------------------------

**Дополнение:**  
**Интернет-ресурсы по транскрипции у растений:**

PlantPAN 2.0

PlantDHS

PPdb

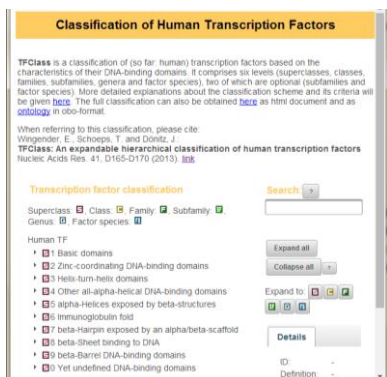
PlantTFDB 3.0

# Повторение слайда из предыдущей лекции: Современные оценки количества транскрипционных факторов в геноме человека

Компьютерная аннотация генома с целью идентификации генов, кодирующих белки, содержащие ДНК связывающие домены.

## Ресурс TFClass

(<http://tfclass.bioinf.med.uni-goettingen.de/>)



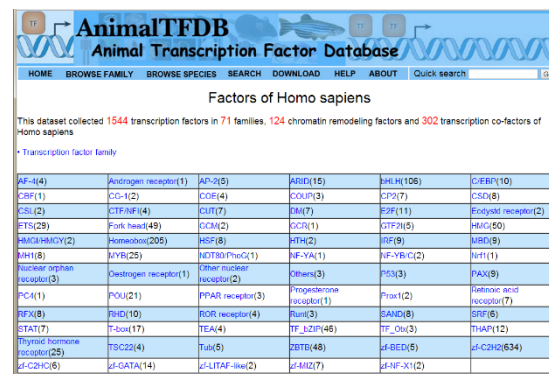
1558 генов,

кодирующих 2904 изоформ белков,  
содержащих ДНК-связывающий домен.

Из них 970 генов (62.3%) кодируют  
экспериментально подтвержденные  
транскрипционные факторы.

## Ресурс AnimalTFDB

(<http://www.bioguo.org/AnimalTFDB/>)



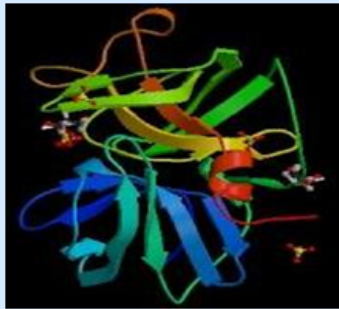
1544 гена,

кодирующих белки,  
содержащие ДНК-связывающий домен.

Zhang H.M. et al., **AnimalTFDB**: a comprehensive animal transcription factor database. Nucleic Acids Res. 2012, 40(Database issue):D144-9.

Wingender E, et al., **TFClass**: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res. 2013 Jan;41(Database issue):D165-70.

<http://www.jncasr.ac.in/cremofac/>



# CREMOFAC

A web-database of  
**Chromatin Remodeling Factors**

[Bioinformatics, 2006; 22: 2940-2944](#)

**Enter**

**Number of Visits**

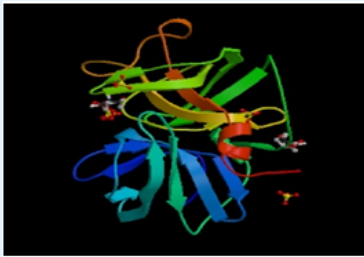
Halloween Costumes

**Developed and maintained by :**  
**Shipra Agrawal<sup>\*</sup>, Chetan Kumar<sup>\*</sup> and M.R.S. Rao**

(\* - Equally contributed to this work)



**Chromatin Biology Lab**  
**Molecular Biology and Genetics Unit**  
**Jawaharlal Nehru Centre for Advanced Scientific Research**



- ▶ Home
- ▶ Useful Links
- ▶ Search Database
- ▶ Remodeling Factors
- ▶ Phylogeny Trees
- ▶ Database Statistics
- ▶ Downloads
- ▶ Literature Sources
- ▶ Admin Login
- ▶ FAQs & User Guide
- ▶ Contact Us
- ▶ JNCASR
- ▶ Chromatin Lab

# CREMOFAC

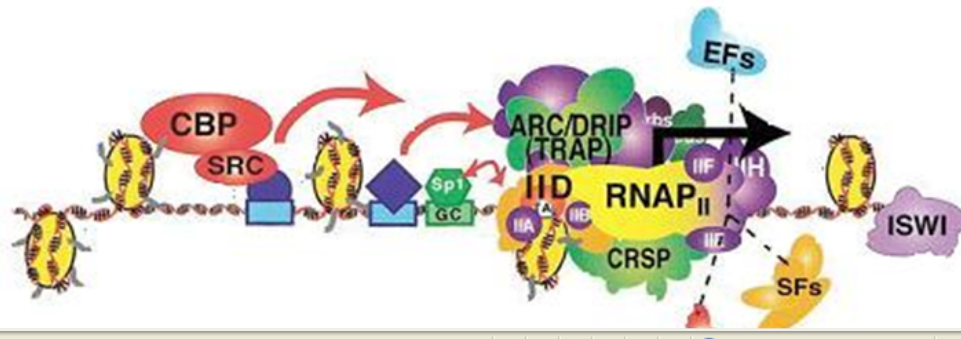
## Chromatin Remodeling Factors

Chromatin Biology Lab, Molecular Biology and Genetics Unit,  
Jawaharlal Nehru Centre For Advanced Scientific Research, Bangalore, India.

### Introduction

**Chromatin remodeling** is an important event in the eukaryotic nucleus rendering nucleosomal DNA accessible for various transaction processes. The dynamic nature of chromatin is facilitated by Remodeling Factors through participation of the collective action of **(a) ATP and (b) Non-ATP dependent factors**. Considering the importance of these factors in mammals and eukaryotes, '**CREMOFAC**', a dedicated and frequently updated web-database for chromatin remodeling factors has been developed.

The database harbors **64 types of remodeling factors** from **49 different organisms** reported in literature and facilitates a comprehensive search for them. In addition, it also provides **in-depth information** for the factors reported in the three widely studied mammals namely, **human, mouse and rat**. Further, information on literature, pathways, and phylogenetic relationships has also been covered.



Данные собраны вручную на основе аннотирования научных публикаций

## CREMOFAC Data Statistics

Number of **redundant** remodeling factor sequences present in the database: **1725**

Number of **non-redundant** remodeling factor sequences present in the database: **720**

List of organisms in which remodeling factors have been reported.

S.No.	Name of Organism	No. of factors found
1.	<a href="#">Aedes aegypti</a>	9
2.	<a href="#">Aegolius funereus</a>	1
3.	<a href="#">Anopheles gambiae</a>	7
4.	<a href="#">Apis mellifera</a>	5
5.	<a href="#">Arabidopsis thaliana</a>	3
6.	<a href="#">Ashbya gossypii ATCC 10895</a>	1
7.	<a href="#">Aspergillus fumigatus Af293</a>	5
8.	<a href="#">Aspergillus nidulans FGSC A4</a>	3
9.	<a href="#">Bos taurus</a>	55
10.	<a href="#">Caenorhabditis briggsae</a>	2
11.	<a href="#">Caenorhabditis elegans</a>	2

Всего 49 видов организмов

# Ресурс CR Cistrome

<http://cistrome.org/cr/>

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

## Chromatin Regulator Cistrome

A knowledgebase for chromatin modifying enzymes and chromatin remodelers

[Home](#)

[Collection Stats](#)

[FAQ](#)

[Quick Start](#)

CR Name:

Species:

Tissue or Cell:

### Regulator Atlas

- Reader
- Writer
- Eraser
- Remodeler

## Welcome to CR Cistrome!

CR Cistrome is a unique database integrating curated information of CRs, CR ChIP-seq datasets, CR related HM ChIP-seq datasets, and analysis of the relationship between CRs and HMs ChIP-seq pairs in human and mouse.

## About CR

CRs are expressed in a tissue-specific manner and play important roles in normal physiology and disease. Hundreds of chromatin regulators (CRs) control chromatin structure and function by catalyzing and binding histone modifications.

## Summary

Chromosome Regulator Number: **36**

Species Number: **2**

Cell line Number: **54**

Cell type Number: **25**

ChIP Seq Data Number: **371**

## Summary

Chromosome Regulator Number: **36**

Species Number: **2**

Cell line Number: **54**

Cell type Number: **25**

ChIP Seq Data Number: **371**

## Regulator Atlas

Reader

Writer

Eraser

Remodeler



# Данные о хроматин-ремоделирующем белке CHD1, представленные в ресурсе CR Cistrome

Home Collection Stats FAQ Quick Start

CR Name: --Select CR Name-- Species: All Tissue or Cell: --All-- Search Reset

## Regulator Atlas

- Reader
- Writer
- Eraser
- Remodeler

## Chromatin Regulator

CHD1

## Alias

DKFZp686E2337

## External Links:

Wiki GeneCards NCBI UniProt

## Related histone modifications:

H3K4me3

## Introduction

Full name: Chromodomain helicase DNA-binding protein 1. CHD1 is a chromatin remodeler belonging to the SNF2 family. It contains two chromodomains in its N-terminus, a central helicase-like ATPase motor, and a C-terminal DNA-binding domain that preferentially binds to A/T-rich sequences. CHD1 can assemble, slide, and space nucleosomes in vitro and is required to maintain the pluripotency of embryonic stem cells (1-4).

## Function and Interaction

CHD1 is essential for assembling H3.3, a conserved histone variant, into chromatin in vivo, suggesting its involvement in replication-independent nucleosome assembly (5). CHD1 has been shown to interact with the transcription elongation modulator Rtf1 and elongation factors Spt4-Spt5 and Spt16-Pob3, indicating that CHD1 participates in the transcription elongation process (6). CHD1 is thought to bind to trimethylated H3K4 in active genes, rather than bivalent regions (which contain both H3K4me3 and H3K27me3), which is accomplished via a mediator through association with the assembly of an active transcription complex (7-10).

## Disease Association

NA

## ChIP-Seq data

SPECIES	CELL LINE	CELL TYPE	TISSUE	DATA	DOWNLOAD	SEND TO CISTROME	ANALYSIS FIGURES	COMPARISON	REFERENCE
Homo sapiens	K562	Erythroblast	Bone Marrow	GSE32509 ,GSM830988,GSM830989	Bed Big wiggle	Send Bed Send Big wiggle	Click Download	Click	22196736
Homo sapiens	H1	Embryonic Stem	Embryo	GSE32509 ,GSM831025,GSM831026	Bed Big wiggle	Send Bed Send Big wiggle	Click Download	Click	22196736
Homo sapiens	H1	Embryonic Stem	Embryo	GSE29611 ,GSM1003444	Bed Big wiggle	Send Bed Send Big wiggle	Click Download	Click	The Encyclopedia Of DNA Elements (ENCODE) Broad
Homo sapiens	K562	Erythroblast	Bone Marrow	GSE29611 ,GSM1003575	Bed Big wiggle	Send Bed Send Big wiggle	Click Download	Click	The Encyclopedia Of DNA Elements (ENCODE) Broad

Ручная аннотация функциональных характеристик белков-регуляторов хроматина + данные ChIP-seq

# Современная оценка количества хроматин-модифицирующих и хроматин-ремоделирующих факторов в геноме человека: подход на основе интеграции данных из нескольких баз.

База EntrezGene: запрос по термину Gene Ontology  
«chromatin modification»  
(аннотация GO терминами ручная и компьютерная)

**Ресурс CREMOFAC**  
(Ручная аннотация статей)

**Ресурс CR Cistrome**  
(Ручная аннотация статей)



Ни один из трех источников не являлся полным, каждый источник содержал уникальные сведения, которых не было в двух других источниках

# МАТРИЦЫ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

**TRANSFAC Matrix,  
JASPAR ,  
HOCOMOCO ,  
CIS-BP**

# Данные по матрицам сайтов связывания в базе TRANSFAC Matrix <http://www.gene-regulation.com/cgi-bin/pub/databases/transfac/search.cgi>



This version of the TRANSFAC database is free for users from non-profit organizations only.  
Users from commercial organizations have to license the TRANSFAC databases and accompanying programs.

[TRANSFAC MATRIX TABLE, Release 7.0 - public - 2005-09-30, \(C\) Biobase GmbH](#)

```

AC M00134
XX
ID V$HNF4_01
XX
DI 22.05.1995 (created); hiwi.
DI 18.10.1995 (updated); ewi.
CO Copyright (C), Biobase GmbH.
XX
NA HNF-4
XX
DE hepatic nuclear factor 4
XX
BF T00372 HNF-4alpha1; Species: rat, Rattus norvegicus.
BF T00373 HNF-4alpha2; Species: human, Homo sapiens.
XX
PO      A      C      G      T
01     10     4      4      6      N
02      6      9      7      5      N
03     12     6      7      6      N
04     12     3     14     3      R
05      2      0     29     1      G
06      5      2     17     8      G
07      3      8     10    11      N
08      1     23     1      7      C
09     27     1      3      1      A
10     29     0      3      0      A
11     26     0      5      1      A
12      3      0     28     1      G
13      3      1     16    12      K
14      2      6      6     18      T
15      0     24     1      7      C
16     22     4      4      2      A
17      9      9      6      6      N
18      7      5     13     5      N
19      8      3      6      7      N
XX
BA 32 binding sites from 24 genes
XX
CC compiled sequences
XX
//
    
```


## TRANSFAC® Professional vs. Public

TRANSFAC® Professional is an internationally unique knowledgebase containing published data on eukaryotic transcription factors and their regulated genes, coupled with transcription factor binding site prediction tools for advanced research. The professional version content is more than six years ahead of the public release and includes data and analysis capabilities not available in the public version.

TRANSFAC®	Professional	Public
<b>Data</b>		
Factors	21,215	6,133
miRNAs	894	n/a
DNA sites	38,283	7,915
mRNA sites	15,895	n/a
Factor DNA	51,769	Yes
miRNA-mRNA site links	49,541	n/a
Genes	79,866	2,397
ChIP-chip/Seq Fragments	2,332,432	n/a
Matrices	5,551	398
References	29,681	Flat file only
Promoter Sequences	277,337	n/a

Пример: две матрицы сайтов связывания CTCF

Version: 5.0\_ALPHA You are using the JASPAR development server: jaspardev.genereg.net Visit the stable JASPAR server: jaspar.genereg.net



The high-quality transcription factor binding profile database

Browse the JASPAR CORE database directly:

- JASPAR CORE Vertebrata
- JASPAR CORE Nematoda
- JASPAR CORE Insecta
- JASPAR CORE Plantae
- JASPAR CORE Fungi
- JASPAR CORE by Structural Class

DOCUMENTATION DOWNLOAD TOOLS CONTACT

Browse/search a JASPAR database

BRCA1 Quick Search

Select a JASPAR database

- JASPAR CORE
- JASPAR Collections
- JASPAR ONE
- JASPAR FAM
- JASPAR PBM
- JASPAR PBM\_HLH

The JASPAR CORE database contains a curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes. The prime difference to similar resources (TRANSFAC, etc) consist of the open data access, non-redundancy and quality. When should it be used? When seeking models for specific factors or structural classes, or if experimental evidence is paramount.

Summary page for ID: MA0139.1 NAME: CTCF from the JASPAR CORE database

DATA	
Name	CTCF
Class	Zinc-coordinating
Family	BetaBetaAlpha-zinc finger
Species	Homo sapiens
Taxon	vertebrates
ACC	P49711
Type	ChIP-seq
MEDLINE	17512414
Pazar ID	TF000607
TFBSShape ID	133
TFencyclopedia ID	-
Comment	-

VERSION INFORMATION

There is only one version of the model.

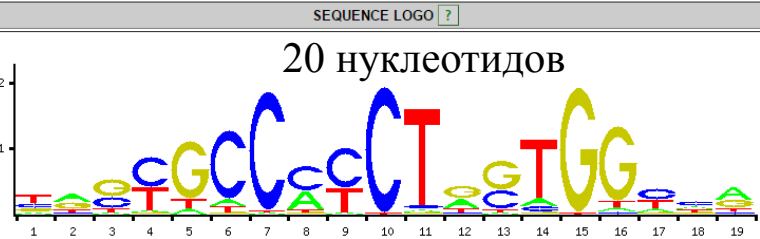
SITES

...as web page ...as fasta file

Show me all the binding sites

SEQUENCE LOGO

20 нуклеотидов



Make a SVG logo

FREQUENCY MATRIX

A	[ 59 396 37 17 67 71 8 341 9 3 18 324 11 91 2 36 134 187 459 ]
C	[ 266 73 307 507 5 775 890 504 566 903 32 48 334 65 0 21 449 414 76 ]
G	[ 181 322 482 13 733 8 0 12 3 0 11 433 528 13 903 800 49 145 291 ]
T	[ 402 117 82 372 104 56 12 54 333 5 851 107 40 744 8 56 281 167 87 ]

Reverse complement

Summary page for ID: MA0531.1 NAME: CTCF from the JASPAR CORE database

DATA	
Name	CTCF
Class	Zinc-coordinating
Family	BetaBetaAlpha-zinc finger
Species	Drosophila melanogaster
Taxon	insects
ACC	Q9VS55
Type	ChIP-chip
MEDLINE	17616980
Pazar ID	-
TFBSShape ID	331
TFencyclopedia ID	-
Comment	-

VERSION INFORMATION

There is only one version of the model.

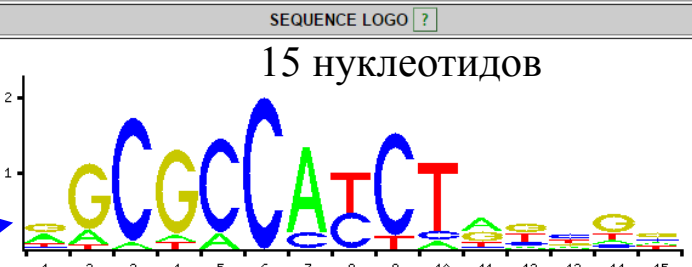
SITES

...as web page ...as fasta file

Show me all the binding sites

SEQUENCE LOGO

15 нуклеотидов



Make a SVG logo

FREQUENCY MATRIX

A	[ 754 301 94 130 248 0 1585 3 0 175 844 92 236 223 317 ]
C	[ 144 8 1807 3 1652 1902 311 912 1700 192 87 350 826 219 403 ]
G	[ 773 1514 0 1645 0 0 4 0 0 1 714 784 383 1147 876 ]
T	[ 231 79 1 124 2 0 2 987 202 1534 257 676 457 313 306 ]

Reverse complement

Графики WEB-logo

# HOCOMOCO (Homo sapiens Comprehensive Model Collection)

<http://autosome.ru/HOCOMOCO/>

The screenshot displays the HOCOMOCO website interface. At the top, there is a navigation menu with options: Home, Human TFs, Mouse TFs, Tools, Downloads, and Help. A search bar is located on the right, with a dropdown menu set to 'Human PWMs' and a 'Search' button. Below the navigation, there is a 'Please cite:' section with a citation link and a 'Show more' button. To the right of the citation is a 'Primary URL' and a 'Mirror' link. The main content area features a large hierarchical diagram of transcription factor binding domains. The diagram starts with a central node 'HOCOMOCO' and branches out into several categories: Basic domains, Zinc-coordinating DNA-binding domains, Helix-turn-helix domains, Other all-alpha-helical DNA-binding domains, alpha-Helices exposed by beta-structures, Immunoglobulin fold, beta-Hairpin exposed by an alpha/beta-scaffold, beta-Sheet binding to DNA, beta-Barrel DNA-binding domains, and Yet undefined DNA-binding domains. Each category is further subdivided into specific domain types, such as Basic leucine zipper factors (bZIP), Basic helix-loop-helix factors (bHLH), and others.

**Please cite:**  
HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models  
[Show more](#)

[Primary URL](#) [Mirror](#)

HOMO sapiens COMPREHENSIVE MOdel COLLECTION (HOCOMOCO) v11 provides transcription factor (TF) binding models for 680 human and 453 mouse TFs.

Since v11, HOCOMOCO is complemented by [MoLoTool](#), an interactive web tool to mark motif occurrences in a given set of DNA sequences.

In addition to basic mononucleotide position weight matrices (PWMs), HOCOMOCO provides dinucleotide position weight matrices based on ChIP-Seq data.

All the models were produced by the [ChIPMunk](#) motif discovery tool. Model quality ratings are results of a comprehensive cross-validation benchmark.

ChIP-Seq data for motif discovery was extracted from [GTRD](#) database of BioUML platform, that also provides an [interface](#) for motif finding (sequence

- Basic domains
  - Basic leucine zipper factors (bZIP)
  - Basic helix-loop-helix factors (bHLH)
  - Basic helix-span-helix factors (bHSH)
  - Nuclear receptors with C4 zinc fingers
  - Other C4 zinc finger-type factors
- Zinc-coordinating DNA-binding domains
  - C2H2 zinc finger factors
    - DM-type intertwined zinc finger factors
    - CXXC zinc finger factors
    - C2HC zinc finger factors
    - C3H zinc finger factors
    - C2CH THAP-type zinc finger factors
  - Homeo domain factors
    - Paired box factors
    - Fork head / winged helix factors
    - Heat shock factors
    - Tryptophan cluster factors
    - TEA domain factors
    - ARID domain factors
    - High-mobility group (HMG) domain factors
    - Heteromeric CCAAT-binding factors
- Helix-turn-helix domains
  - MADS box factors
  - SAND domain factors
  - Rel homology region (RHR) factors
  - STAT domain factors
  - p53 domain factors
  - Runt domain factors
  - T-Box factors
  - NDT80 domain factors
  - Grainyhead domain factors
  - SMAD/NF-1 DNA-binding domain factors
  - GCM domain factors
- Other all-alpha-helical DNA-binding domains
  - TATA-binding proteins
  - A T hook factors
  - Cold-shock domain factors
  - Uncharacterized
  - AXUD/CSRNP domain factors
  - NonO domain factors
  - Leucine-rich repeat flightless-interacting proteins
  - NFX1-type putative zinc finger factors
- alpha-Helices exposed by beta-structures
- Immunoglobulin fold
- beta-Hairpin exposed by an alpha/beta-scaffold
- beta-Sheet binding to DNA
- beta-Barrel DNA-binding domains
- Yet undefined DNA-binding domains

HOCOMOCO v11 provides transcription factor (TF) binding models for **680 human** and **453 mouse** TFs.

# Нomo sapiens CОmprehensive MОdel CОllection (НОCOMOCO) v11

Пример: матрица сайтов связывания STF1\_HUMAN.H11MO.0.B

PCM				PWM					
A	C	G	T	A	C	G	T		
01	67.0	123.0	210.0	100.0	01	-0.613	0.016	0.314	-0.32
02	77.0	196.0	302.0	187.0	02	-1.489	0.448	-0.324	0.399
03	28.0	120.0	14.0	138.0	03	-1.404	0.56	-2.096	0.638
04	3.0	499.0	28.0	10.0	04	-3.325	1.292	-1.454	2.394
05	420.0	58.0	14.0	8.0	05	1.203	-0.754	-2.096	-2.584
06	499.0	2.0	36.0	3.0	06	1.292	-3.573	-1.215	-3.325
07	20.0	1.0	474.0	5.0	07	-1.77	-3.903	1.324	-2.961
08	13.0	20.0	457.0	10.0	08	2.163	-1.77	1.287	-2.394
09	51.0	225.0	19.0	205.0	09	-0.879	0.587	-1.818	0.487
10	11.0	480.0	4.0	20.0	10	-2.311	1.272	-3.326	-1.242
11	378.0	33.0	37.0	52.0	11	1.208	-1.298	-1.189	-0.86

11 нуклеотидов

Пример: матрица сайтов связывания CTCF

PCM				PWM					
A	C	G	T	A	C	G	T		
01	75.0	82.0	199.0	94.0	01	-0.399	-0.311	0.565	-0.177
02	20.0	375.0	33.0	22.0	02	-1.667	1.195	-1.195	-1.578
03	8.0	424.0	12.0	6.0	03	-2.482	1.317	-2.132	-2.718
04	153.0	31.0	168.0	168.0	04	0.304	-1.254	0.397	-0.136
05	56.0	201.0	188.0	5.0	05	-0.684	0.574	0.508	-2.86
06	64.0	230.0	41.0	115.0	06	-0.554	0.708	0.986	0.022
07	393.0	20.0	14.0	23.0	07	1.241	-1.667	-1.994	-1.537
08	4.0	5.0	440.0	1.0	08	-3.027	-2.86	1.354	-3.809
09	104.0	0.0	340.0	6.0	09	-0.077	-4.313	1.097	-2.718
10	36.0	41.0	330.0	43.0	10	-1.111	-0.986	1.067	-0.94
11	4.0	1.0	442.0	3.0	11	-3.027	-3.809	1.358	-3.226
12	5.0	6.0	421.0	18.0	12	-2.86	-2.718	1.31	-1.765
13	37.0	404.0	2.0	7.0	13	-1.085	1.269	-3.476	-2.593
14	95.0	2.0	346.0	7.0	14	-0.167	-3.476	-1.154	-2.593
15	13.0	318.0	93.0	26.0	15	-2.06	1.03	-0.188	-1.421
16	58.0	184.0	35.0	173.0	16	-0.45	0.487	-1.138	0.426
17	119.0	118.0	171.0	42.0	17	0.095	0.047	0.414	-0.963

17 нуклеотидов

Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, Makeev VJ. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. 2016 Jan 4;44(D1):D116-25.



CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

Home  
Tools  
View cart  
Bulk downloads  
Database stats  
Contact us  
Help  
Update Log  
FAQ  
Links  
How to cite

Welcome to CIS-BP, the online library of transcription factors and their DNA binding motifs.

Search for a TF  
By Identifier   
(e.g. Gata\*, YEL009C, ISFTZ\_01)

Browse TFs / Restrict Search for TFs  
By Model Organism   
By Any Species   
By Domain Type   
By Motif Evidence   
By Evidence Type   
By Study   
Database Build  Version 1.02

**GO!**

Database build 1.02 now available!

Last updated: Apr 5th, 2015 Database Build 1.02  
Current content: 6559 motifs, 59998 TFs with at least one binding motif (3202 from direct experiments), out of a total of 167081 TFs from 263 families in 340 species

## Статистика

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

Home	Source	Type	Author	Year	PMD	Num TFs	Pct TFs
Tools	Badis08	PBM	Badis	2008	19111667	110	0.1
View cart	Badis09	PBM	Badis	2009	19443739	105	0.1
Bulk downloads	Berger06	PBM	Berger	2006	16998473	5	0
Database stats	Berger08	PBM	Berger	2008	18585359	170	0.1
Contact us	Campbell10	PBM	Campbell	2010	21060817	17	0
Help	CEPD	PBM	Narasimhan	2015	NULL	129	0.1
Update Log	Chang2013	PBM	Chang	2013	23795294	16	0
FAQ	DeBoer11	DeBoer11	DeBoer	2011	22102575	198	0.1
Links	DeBianco10	PBM	DeBianco	2010	21124806	0	0
How to cite	DeMasi11	PBM	DeMasi	2011	21335608	0	0
	DeSilva08	PBM	DeSilva	2008	18541913	3	0
	DREAM_contest	PBM	Weirauch	2013	23354101	83	0
	ENCODE	ChIP-seq	Gerstein	2012	22955619	70	0
	FlyFactorSurvey	BIH	Zhu	2011	21097781	298	0.2
	Grove09	PBM	Grove	2009	19632181	10	0
	Helfer11	PBM	Helfer	2011	21236673	0	0
	hmCHIP	ChIP-seq	Chen	2011	21450710	18	0
	HocoMoco	HocoMoco	Kulakovskiy	2013	23175603	395	0.2
	JASPAR	JASPAR	Mathelier	2014	24194598	543	0.3
	Jolma	SELEX	Jolma	2013	23332764	453	0.3
	Lam11	PBM	Lam	2011	21321018	23	0
	Lesch09	PBM	Lesch	2009	19204119	1	0
	modENCODE	ChIP-seq	Boyle	2014	25164757	35	0
	Scharer09	PBM	Scharer	2009	19147588	1	0
	SebePedros2013	PBM	SebePedros	2013	24043797	4	0
	Transfac	Transfac	Matys	2006	16381825	1002	0.6
	Wei10	PBM	Wei	2010	20517297	22	0
	Zhu09	PBM	Zhu	2009	19158363	89	0.1
	Zoo_01	PBM	Weirauch	2014	25215497	1017	0.6

## Возможность предсказания сайтов по матрицам

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

Home  
Tools  
View cart  
Bulk downloads  
Database stats  
Contact us  
Help  
Update Log  
FAQ  
Links  
How to cite

Scan single sequences for TF binding

Scan a DNA sequence for potential binding sites  
Click for [sample](#)

Species  Homo\_sapiens  
Motif model  Threshold   
PWMs - LogOdds  8

Scan TFs in my cart

Scan two sequences for differential TF binding

Allele 1   
Species   
 SNP Scan  
Click for [sample](#)

Allele 2   
Motif model  Threshold  8 mers - Scores 0.45  
 Scan TFs in my cart

Protein Scan

Scan an amino acid sequence to predict its DNA binding motif  
Click for [sample](#)

Motif Scan

Compare a given motif to all TFs in the database  
Click for [PWM alignment](#)  
[IUPAC](#)

Species  Scan all species

**GO!**

Weirauch MT et al., . Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014 Sep 11;158(6):1431-43. doi: 10.1016/j.cell.2014.08.009.



# Информационное содержание базы CIS-BP

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences							
	Source	Type	Author	Year	PMID	Num TFs	Pct TFs
Home	Badis08	PBM	Badis	2008	19111667	110	0.1
Tools	Badis09	PBM	Badis	2009	19443739	105	0.1
View cart	Berger06	PBM	Berger	2006	16998473	5	0
Bulk downloads	Berger08	PBM	Berger	2008	18585359	170	0.1
Database stats	Campbell10	PBM	Campbell	2010	21060817	17	0
Contact us	CEPD	PBM	Narasimhan	2015	NULL	129	0.1
Help	Chang2013	PBM	Chang	2013	23795294	16	0
Update Log	DeBoer11	DeBoer11	DeBoer	2011	22102575	198	0.1
FAQ	DelBianco10	PBM	DelBianco	2010	21124806	0	0
Links	DeMasi11	PBM	DeMasi	2011	21335608	0	0
How to cite	DeSilva08	PBM	DeSilva	2008	18541913	3	0
	DREAM_contest	PBM	Weirauch	2013	23354101	83	0
	ENCODE	ChIP-seq	Gerstein	2012	22955619	70	0
	FlyFactorSurvey	B1H	Zhu	2011	21097781	298	0.2
	Grove09	PBM	Grove	2009	19632181	10	0
	Helfer11	PBM	Helfer	2011	21236673	0	0
	hmCHIP	ChIP-seq	Chen	2011	21450710	18	0
	HocoMoco	HocoMoco	Kulakovskiy	2013	23175603	395	0.2
	JASPAR	JASPAR	Mathelier	2014	24194598	543	0.3
	Jolma	SELEX	Jolma	2013	23332764	453	0.3
	Lam11	PBM	Lam	2011	21321018	23	0
	Lesch09	PBM	Lesch	2009	19204119	1	0
	modENCODE	ChIP-seq	Boyle	2014	25164757	35	0
	Scharer09	PBM	Scharer	2009	19147588	1	0
	SebePedros2013	PBM	SebePedros	2013	24043797	4	0
	Transfac	Transfac	Matys	2006	16381825	1002	0.6
	Wei10	PBM	Wei	2010	20517297	22	0
	Zhu09	PBM	Zhu	2009	19158363	89	0.1
	Zoo_01	PBM	Weirauch	2014	25215497	1017	0.6

Weirauch MT et al., . Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014 Sep 11;158(6):1431-43. doi: 10.1016/j.cell.2014.08.009.

# **TRANSCRIPTION REGULATORY REGIONS DATABASE (TRRD)**

*ИЦиГ СО РАН, Новосибирск, Россия*

<http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/>

**РЕГУЛЯТОРНЫЕ РАЙОНЫ, ССТФ, ТРАНСКРИПЦИОННЫЕ ФАКТОРЫ**

# ПУБЛИКАЦИЯ ПО TRRD в Nucleic Acids Research



312–317 *Nucleic Acids Research*, 2002, Vol. 30, No. 1

© 2002 Oxford University Press

## Transcription Regulatory Regions Database (TRRD): its status in 2002

N. A. Kolchanov\*, E. V. Ignatieva, E. A. Ananko, O. A. Podkolodnaya, I. L. Stepanenko, T. I. Merkulova, M. A. Pozdnyakov, N. L. Podkolodny, A. N. Naumochkin and A. G. Romashchenko

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentyevskiy pr. 8/1, Novosibirsk 630090, Russia

Received September 19, 2001; Accepted September 26, 2001

312–317 *Nucleic Acids Research*, 2002, Vol. 30, No. 1

© 2002 Oxford University Press

### Transcription Regulatory Regions Database (TRRD): its status in 2002

N. A. Kolchanov\*, E. V. Ignatieva, E. A. Ananko, O. A. Podkolodnaya, I. L. Stepanenko, T. I. Merkulova, M. A. Pozdnyakov, N. L. Podkolodny, A. N. Naumochkin and A. G. Romashchenko

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrentyevskiy pr. 8/1, Novosibirsk 630090, Russia

Received September 19, 2001; Accepted September 26, 2001

#### ABSTRACT

Transcription Regulatory Regions Database (TRRD) is an informational resource containing an integrated description of the gene transcription regulation. An entry of the database corresponds to a gene and contains the data on localization and functions of the transcription regulatory regions as well as gene expression patterns. TRRD contains only experimental data that are inputted into the database through annotating scientific publications. TRRD release 8.0 comprises the information on 1167 genes, 5537 transcription factor binding sites, 1714 regulatory regions, 14 locus control regions and 5338 expression patterns obtained through annotating 3898 scientific papers. This information is arranged in seven databases: TRRDGENES (general gene description), TRRDLCR (locus control regions); TRRDUNITS (regulatory regions: promoters, enhancers, silencers, etc.); TRRDSITES (transcription factor binding sites); TRRDFACTORS (transcription factors), TRRDEXP (expression patterns) and TRRDBIB (experimental publications). Sequence Retrieval System (SRS) is used as a basic tool for navigating and searching TRRD and integrating it with external informational and software resources. The visualization tool, TRRD Viewer, provides the information representation in a form of maps of gene regulatory regions. The option allowing nucleotide sequences to be searched for according to their homology using BLAST is also included. TRRD is available at <http://www.bionet.nsc.ru/trrd/>.

#### DESCRIPTION OF TRRD

Transcription Regulatory Regions Database (TRRD) has been developed and supported at the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia) since 1993. The main goal while developing TRRD was to provide a most complete and adequate description of the structure-function organization of transcription regulatory regions of eukaryotic genes. Both the

TRRD structure and format were formed to achieve this goal and are still developing. The current TRRD release fully comprises seven databases linked with cross-references: TRRDGENES (general gene descriptions), TRRDLCR (locus control regions), TRRDUNITS (regulatory regions: promoters, enhancers, silencers, etc.), TRRDSITES (transcription factor binding sites), TRRDFACTORS (transcription factors), TRRDEXP (expression patterns) and TRRDBIB (bibliography).

The format of TRRD allows the transcription regulation of the eukaryotic genes transcribed by RNA polymerase II to be described in an integrated manner in all the organs, tissues and cell types of the organism as well as in cell lines. First, TRRD contains the data on structural organization of transcription regulatory regions of the following hierarchical levels: (i) transcription factor binding sites (TRRDSITES); (ii) regulatory units, including promoters, enhancers and silencers (TRRDUNITS); (iii) regulatory regions, including 5' and 3' regulatory regions, exons and introns (TRRDGENES); and (iv) locus control regions (TRRDLCR). Secondly, TRRD accumulates functional characteristics of regulatory elements of all the levels, such as the effect of the gene on transcriptional activity, specific function at a certain stage of the cell cycle or ontogenesis, in particular cell types, tissues or organs, and involvement of a regulatory element in regulation of gene expression in response to various intracellular and external stimuli or influences. Thirdly, TRRD contains the data on patterns of gene expression (TRRDEXP). The interventional links RegLink (RP) and SiteLink (RS) of this database are hyperlinked to textual descriptions of regulatory units (promoters, enhancers and silencers, described in TRRDUNITS) and transcription factor binding sites (TRRDSITES) that realize specific expression features typical of each pattern.

A distinguishing feature of the TRRD database is accumulation of the information confirmed experimentally. The data are inputted into TRRD through annotating publications describing results of experiments of different types, exemplified in Table 1. These experiments may be aimed at (i) detection and primary analysis of extended regulatory regions of genes, (ii) detection of transcription factor binding sites, (iii) confirming the functional importance of the sites, and (iv) identification of DNA-binding proteins. Each type of experiment has its own code (second column of the table). Digital codes together with the information on cell types involved in experiments are given

\*To whom correspondence should be addressed. Tel: +7 3822 333408; Fax: +7 3822 331278; Email: ikol@bionet.nsc.ru

# ГЛАВНАЯ СТРАНИЦА ДЛЯ ВХОДА В TRRD ЧЕРЕЗ ИНТЕРНЕТ <http://www.bionet.nsc.ru/trrd/>



**Gene Networks**  
HOME DNA RNA PROTEIN GENENETWORKS MAP

**TRANSCRIPTION REGULATORY REGIONS DATABASE**

**TRRD**

TRRD is a unique information resource, accumulating information on structural and functional organization of transcription regulatory regions of eukaryotic genes. Only experimental information is included into TRRD.

**ACCESS to TRRD:**

- [SRS ACCESS](#)
- [TRRDGENES](#)
- [TRRDEXP](#)
- [TRRDSITES](#)
- [TRRDFACTORS](#)
- [TRRDBIB](#)
- [TRRDUNITS](#)
- [TRRDLCR](#)
- [TRRD sections \(genes within functional systems\)](#)
- [Blast search TRRD database](#)
- [Browse the TRRD Programs](#)

**General information**

- [How to cite TRRD?](#)
- [TRRD publications](#)
- [The latest report on TRRD](#)
- [TRRD Workgroup](#)
- [Contact us](#)
- [Acknowledgments](#)

**User's guide**

- [Database schema](#)
- [How to search TRRD?](#)
- [Integration with other databases](#)
- [TRRD Viewer](#)
- [FAQ](#)
- [What's new?](#)

**How is TRRD updated ?**

- [Standardization of information input](#)
- [TRRD progress \(from 1996\)](#)

**Current TRRD release**

- [Information contents](#)
- [TRRD statistics](#)

**General information**

- [How to cite TRRD?](#)
- [TRRD publications](#)
- [The latest report on TRRD](#)
- [TRRD Workgroup](#)
- [Contact us](#)
- [Acknowledgments](#)

**User's guide**

- [Database schema](#)
- [How to search TRRD?](#)
- [Integration with other databases](#)
- [TRRD Viewer](#)
- [FAQ](#)
- [What's new?](#)

**Current TRRD release**

- [Information contents](#)
- [TRRD statistics](#)

To obtain further particular information about TRRD database, please, address to the database scientific supervisor Nikolay A. Kolchanov by e-mail [kol@bionet.nsc.ru](mailto:kol@bionet.nsc.ru) or by fax +7-3832-331278

Done Internet

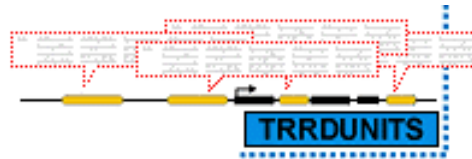
# TRRD ВКЛЮЧАЕТ ДАННЫЕ О:

1) СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ РАЙОНОВ. СТРУКТУРНЫЕ И ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ:

САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ



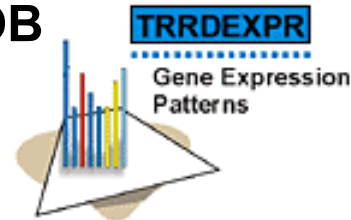
РЕГУЛЯТОРНЫХ РАЙОНОВ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)



ЛОКУС-КОНТРОЛИРУЮЩИХ РАЙОНОВ

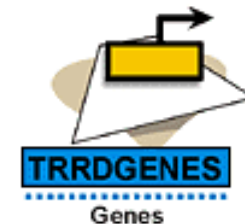
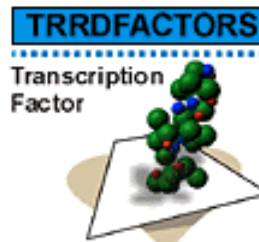


2) ДАННЫЕ О ПАТТЕРНАХ ЭКСПРЕССИИ ГЕНОВ



4) ОБЩУЮ ИНФОРМАЦИЮ О ГЕНАХ ВМЕСТЕ С ПЕРЕЧИСЛЕНИЕМ РЕГУЛЯТОРНЫХ ЭЛЕМЕНТОВ ВСЕХ УРОВНЕЙ

3) ДАННЫЕ О ТРАНСКРИПЦИОННЫХ ФАКТОРАХ



# TRRDGENES

## ОБЩЕЕ ОПИСАНИЕ ГЕНА

ID Hs:AAP

DT 09/02/00

AC A00596

CR Ignatieva E.V., Stepanenko I.L.

OS human, Homo sapiens

SN AAP

NG Alzheimer's disease amyloid A4 precursor protein

SY amyloid beta protein precursor gene

SY amyloid precursor protein gene

SY APP

BI EMBL; [HSPADP](#); [X12751](#); ST:3700

DR GDB; [119692](#); [APP](#)

DR SWISS-PROT; [A4 HUMAN](#); [P05067](#)

KW heat shock-induced, TATA-less promoter, pathogenesis-related protein, multiple transcription initiation sites

CH 21

RG 5'region

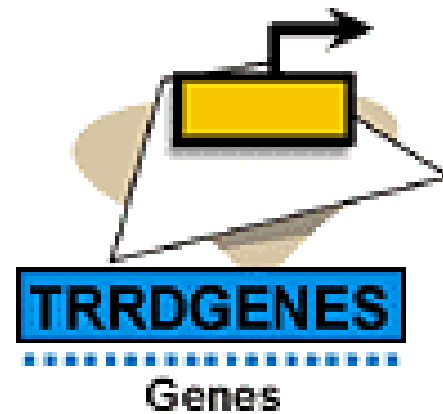
AP [REGULATORY UNIT: P01087](#)

PR regulatory region; ST; -2260 to -1800; [S3936](#), [S3937](#)

AP [REGULATORY UNIT: P00760](#)

PR promoter; ST; -160 to -1; [S2907](#), [S3931](#), [S3932](#), [S3933](#), [S3934](#), [S3935](#)

//



[TRRDGENES4:A00264](#)

# ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)

## [GeneID](#)

Hs:APOA1 ([TRRD Viewer](#))

## Links:

[Binding sites](#)

[Transcription factors](#)

[Gene expression regulation](#)

[Bibliography](#)

## [Updated](#)

04/06/01

## [GeneAC](#)

A00264

## [TransfacLink](#)

[G000203](#)

## [Annotators](#)

Ignatieva E.V., O.A.P., Proscura A.L.

## [Species](#)

human, Homo sapiens

## [GeneName\\_Brief](#)

apoA1

## [GeneName\\_Full](#)

apolipoprotein AI

## [DNABankLink](#)

EMBL; [HSAPOA01](#); [J04066](#); ST: 2069

EMBL; [HSAPOA02](#); [M20656](#);

EMBL; [HSAPOA11](#); [J00098](#); ST: 469

GenBank; [HUMAPOCP](#); [J05464](#); ST(C3): 1420

## [DataBankLink](#)

SWISS-PROT; [APA1\\_HUMAN](#); [P02647](#)(Expasy server);

## [EPD\\_Class](#)

6.1.3.3.

## [KeyWords](#)

lipid binding protein, high density lipoprotein, l

## [Chromosome](#)

Продолжение:

## ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)

```
SWISS-PROT; APOA1\_HUMAN; P02647 (Expasy server);
MIM; 604091;
PIR; A90947; LPHUA1
HOVERGEN; P02647
SOURCE; APOA1

EPD Class
  6.1.3.3.

KeyWords
  lipid binding protein, high density lipoprotein, LM-TRRD

Chromosome
  91

RegRegion
  5' region

RegUnitAC
  REGULATORY UNIT: P01681

RegUnit
  intestinal enhancer; ST; ; S5738, S5739, S5740, S5741

ExperimentCodes
  Caco-2 cells: 6.8 [Ginsburg G.S. et al., 1995]
  HepG2 cells: 6.8 [Ginsburg G.S. et al., 1995]

RegUnitAC
  REGULATORY UNIT: P00051

RegUnit
  enhancer; ST; ; S1425, S1426, S997, S666, S1135, S999, S1000,
  S1001, S1002, S1427, S1003, S1136, S1004, S1005, S1006, S1007, S5742, S5743,
  S5744, S5745, S5746, S5747, S5748, S5749, S5750

RegUnitAC
  REGULATORY UNIT: P02529

RegUnit
  negative regulatory region; ST; ; S7925

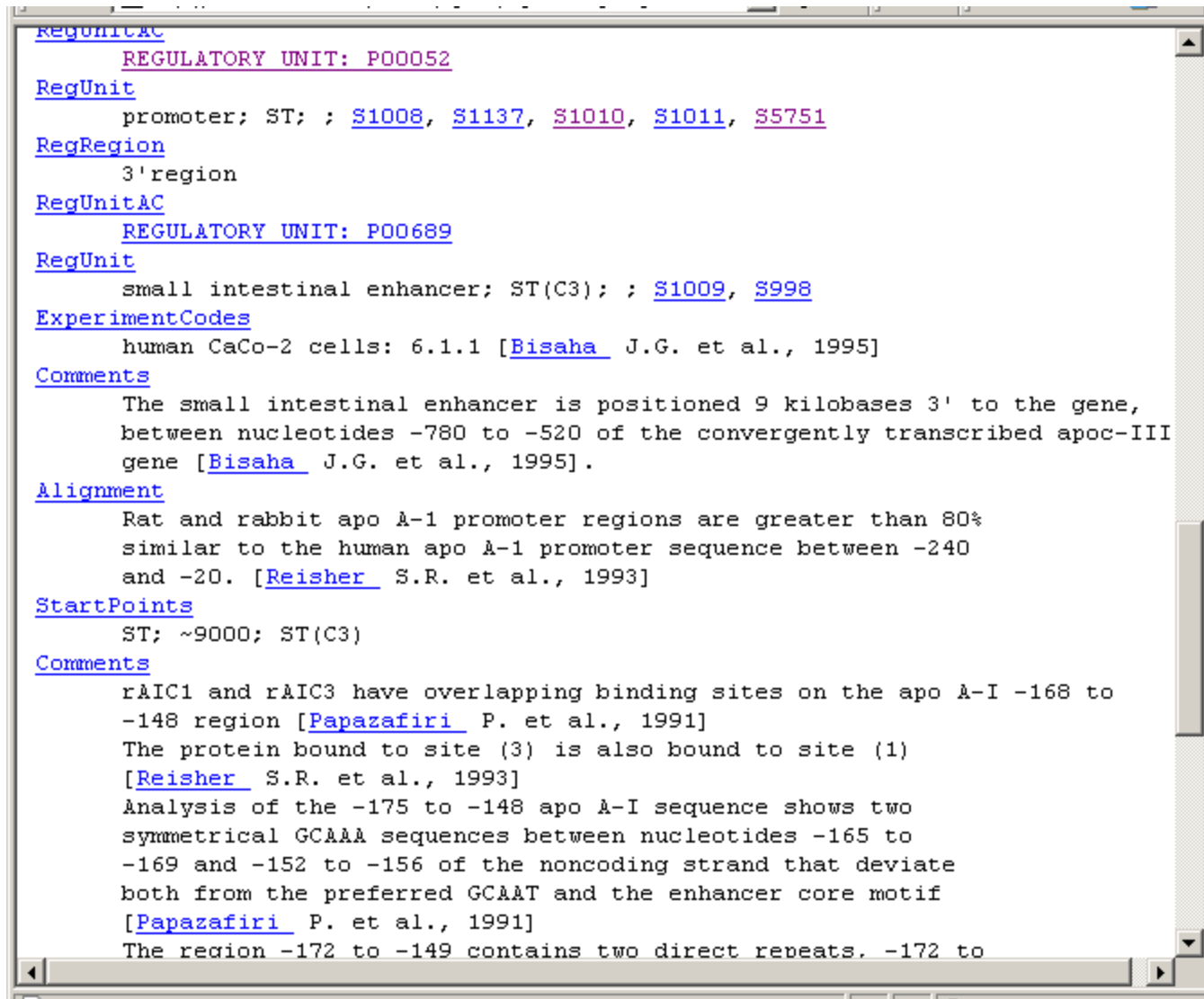
ExperimentCodes
  HepG2 cells: 6.1.1, 6.5 (CDCA) [Claudel T. et al., 2002]

RegUnitAC
```



Окончание:

## ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)



[RegUnitAC](#)  
REGULATORY UNIT: P00052

[RegUnit](#)  
promoter; ST; ; [S1008](#), [S1137](#), [S1010](#), [S1011](#), [S5751](#)

[RegRegion](#)  
3' region

[RegUnitAC](#)  
REGULATORY UNIT: P00689

[RegUnit](#)  
small intestinal enhancer; ST(C3); ; [S1009](#), [S998](#)

[ExperimentCodes](#)  
human CaCo-2 cells: 6.1.1 [[Bisaha](#) J.G. et al., 1995]

[Comments](#)  
The small intestinal enhancer is positioned 9 kilobases 3' to the gene, between nucleotides -780 to -520 of the convergently transcribed apoc-III gene [[Bisaha](#) J.G. et al., 1995].

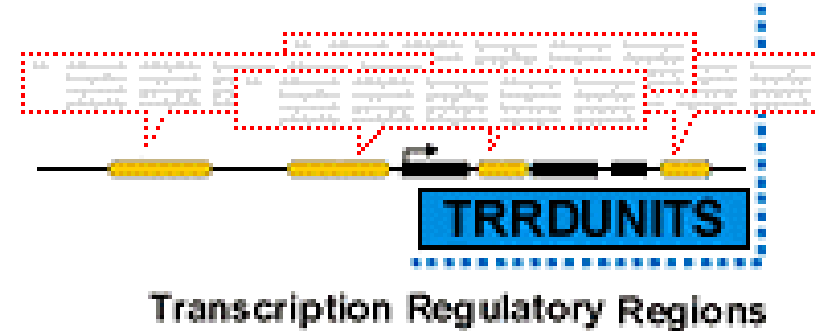
[Alignment](#)  
Rat and rabbit apo A-1 promoter regions are greater than 80% similar to the human apo A-1 promoter sequence between -240 and -20. [[Reisher](#) S.R. et al., 1993]

[StartPoints](#)  
ST; ~9000; ST(C3)

[Comments](#)  
rAIC1 and rAIC3 have overlapping binding sites on the apo A-I -168 to -148 region [[Papazafiri](#) P. et al., 1991]  
The protein bound to site (3) is also bound to site (1) [[Reisher](#) S.R. et al., 1993]  
Analysis of the -175 to -148 apo A-I sequence shows two symmetrical GCAA sequences between nucleotides -165 to -169 and -152 to -156 of the noncoding strand that deviate both from the preferred GCAAT and the enhancer core motif [[Papazafiri](#) P. et al., 1991]  
The region -172 to -149 contains two direct repeats. -172 to

# TRRDUNITS

## ОПИСАНИЕ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ ЕДИНИЦ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)



<a href="#">RegUnitAC</a>	P00562
<a href="#">GeneID</a>	Rn:D2
<a href="#">RegRegion</a>	5' region
<a href="#">RegUnit</a>	Promoter; ST; -150 to +1; S79, S80
<a href="#">DNA BankLink</a>	EMBL; RND2RPR; X77137; 704 to 855
<a href="#">LeftTrunc</a>	0
<a href="#">RightTrunc</a>	0
<a href="#">SeqLength</a>	152
<a href="#">Sequence</a>	cccaggcccc acagtgcaga gatagttctg gggccctggg tgggtggggc ctctgtacaa ggggcggggt tcccgggcgc ctcgtagcca gggtagcccc gccccctcct cctgcgcagc gctctgatte cgcgagctg tccagcctca gt
<a href="#">PromotTisSp</a>	0
<a href="#">PromotInd</a>	1
<a href="#">ExperimentCodes</a>	6.1.1, 6.8 [ <a href="#">Minowa T.</a> et al., 1992]

# ОПИСАНИЕ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ ЕДИНИЦ (html –представление)

[TRRDUNITS4:P00051](#)

[RegUnitAC](#)

P00051

[GeneID](#)

[Hs:APOA1](#)

[RegRegion](#)

5' region

[RegUnit](#)

enhancer; ST; -256 to -110; [S1425](#), [S1426](#), [S997](#), [S666](#)

[S1001](#), [S1002](#), [S1427](#), [S1003](#), [S1136](#), [S1004](#), [S1005](#), [S1006](#), [S5744](#), [S5745](#), [S5746](#), [S5747](#), [S5748](#), [S5749](#), [S5750](#)

Site: ( [S1425](#)) [-225 to -210; Egr-1;](#)

Site: ( [S1426](#)) [-227 to -212; Sp1 bs;](#)

Site: ( [S997](#)) [-220 to -190; D; region D](#)

Site: ( [S666](#)) [-220 to -188; HNF-4; HNF-4 binding site](#)

Site: ( [S1135](#)) [-220 to -192; PPRE; peroxisome proliferator](#)

Site: ( [S999](#)) [-212 to -191; NF-BA1; NF-BA1 binding site](#)

Site: ( [S1000](#)) [-214 to -192; RARE; retinoic acid-responsiv](#)

Site: ( [S1001](#)) [-214 to -192; ARP-1; ARP-1 binding site](#)

Site: ( [S1002](#)) [-210 to -189; S \(1\); site \(1\)](#)

Site: ( [S1427](#)) [-193 to -178; Egr-1 bs;](#)

Site: ( [S1003](#)) [-175 to -155; C \(1\); region C \(1\)](#)

Site: ( [S1136](#)) [-174 to -151; HNF-3 beta; HNF-3 beta bindin](#)

Site: ( [S1004](#)) [-168 to -148; C \(2\) \(B\); region C \(2\)](#)

Site: ( [S1005](#)) [-174 to -144; S \(2\); site S \(2\)](#)

Site: ( [S1006](#)) [-134 to -119; C; site C](#)

Site: ( [S1007](#)) [-133 to -110; S \(3\); site \(3\)](#)

Site: ( [S5742](#)) [-220 to -190; D; region D](#)

Site: ( [S5743](#)) [-214 to -192; T3R/RXR bs; T3R/RXR alpha bin](#)

Site: ( [S5744](#)) [-180 to -147 ; FpB; Footprint B](#)

Site: ( [S5745](#)) [-175 to -148; NFY bs;](#)

Site: ( [S5746](#)) [-149 to -130; ARE; antioxidant response ele](#)

Site: ( [S5747](#)) [-142 to -118; FpC; Footprint C](#)

Site: ( [S5748](#)) [-134 to -119; T3R/RXR bs; T3R beta / RXR al](#)

Site: ( [S5749](#)) [-134 to -119; ARP1 bs; ARP 1 binding site](#)

Site: ( [S5750](#)) [-134 to -119; HNF-4 bs; HNF-4 binding site](#)

[DNA\\_BankLink](#)

EMBL; [HSAPOAO1](#); [J04066](#); 1813 to 1959

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

```
ccacccggga gacctgcaag cctgcagcac tcccctcccg cccccactga
acccttgacc cctgccctgc agcccccgca gcttgctggt tgccccactct
attgcccag ccccagggac agagctgatc cttgaactct taagttc
```

[DNA\\_BankLink](#)

EMBL; [HSAPOAO1](#); [J04066](#); 1813 to 1959

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

```
ccacccggga gacctgcaag cctgcagcac tcccctcccg cccccactga
acccttgacc cctgccctgc agcccccgca gcttgctggt tgccccactct
attgcccag ccccagggac agagctgatc cttgaactct taagttc
```

[DNA\\_BankLink](#)

EMBL; [HSAPOA11](#); [J00098](#); 215 to 361

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

```
cccgggagac ctgcaagcct gcagcactcc cctcccggcc ccaactgaacc
cttgaccctt gccctgcacg ccccgcagct tgctggttgc ccaactctat
ttgcccagtc ccagggacag agctgatcct tgaactctta agttcca
```

# TRRDSITES

## ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА

AN S1160

ID [Gene: Hs:APOB](#)

AP [REGULATORY UNIT: P00670](#)

NM HNF-4 bs; HNF-4 binding site

NY AF-1 binding site

NY BA1 binding site

NS [R01612](#)

TF [HNF-4; hepatic nuclear factor 4](#)

AT increase

SQ cccgggaggCGCCCTTTGGACCTtttg

PQ -88 to -62

PF -82 to -62

BF EMBL: [M15053](#) : 68

AG 1.1.5 3.5 [\[Metzger S. et al., 1993\]](#)

AG rat liver cells: 3.6 [\[Metzger S. et al., 1993\]](#)

AG human HepG2 cells: 6.2 [\[Metzger S. et al., 1993\]](#)

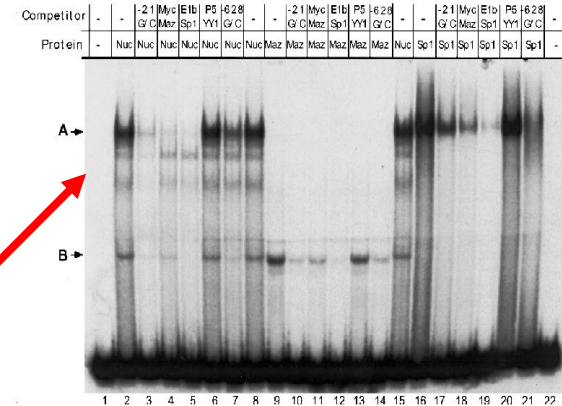
AG HeLa cells: 6.2, 6.6 [\[Metzger S. et al., 1993\]](#)

AG 1.1.5 3.3, 3.5, 4.2 [\[Ldias J.A. et al., 1992\]](#)

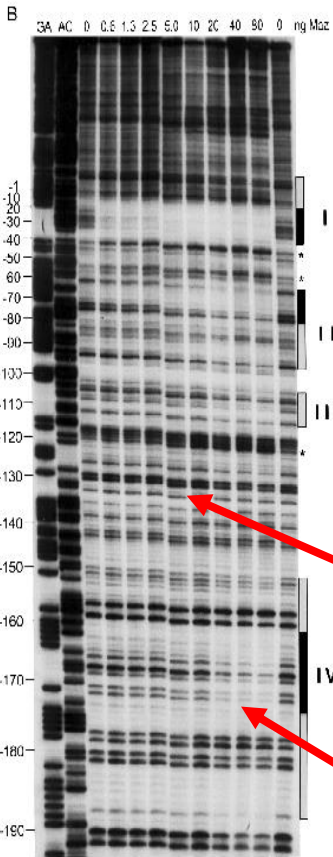
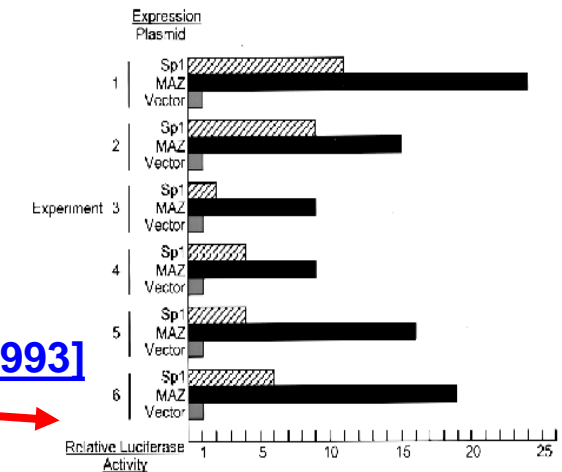
### TRRDSITES



Transcription Factor Binding Sites



GEL-MOBILITY SHIFT ASSAY



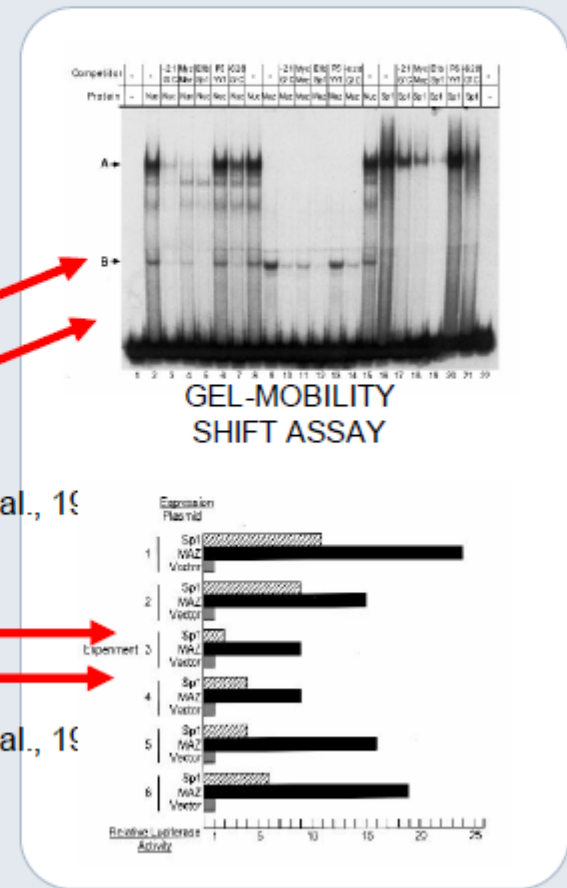
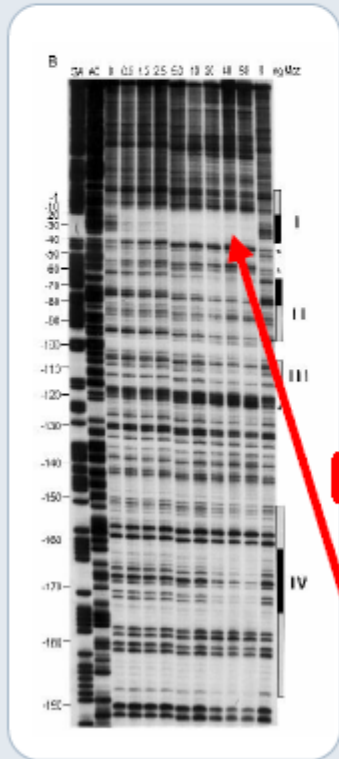
DNASE I FOOTPRINTING

TRANSIENT EXPRESSION ANALYSIS

The table TRRDSITES: description of the HNF-4 binding site in the human ApoB gene

-88 5' cccgggaggCGCCCTTTGGACCTtttg 3' -62  
 3' gggccctccGCGGGAAACCAGGAaac 5'

[SiteAC](#) S1160  
[GeneID](#) Hs:APOB  
[RegUnitAC](#) P00670  
[SiteName](#) HNF-4;  
[PreferredName](#) HNF-4  
[SiteNameSynonym](#) AF-1 binding site  
[SiteNameSynonym](#) BA1 binding site  
[SiteIndex](#) 1  
[FactorName](#) HNF-4; hepatic nuclear factor 4  
[FactorInfluence](#) increase  
[Sequence](#) cccgggaggCGCCCTTTGGACCTtttg  
[SequencePosition](#) -88 to -62  
[FootprintSequencePosition](#) -82 to -62  
[DNA\\_BankLink](#) M15053: 68  
[ImportantPos](#)  
 -----CG--CTTTGGACCT--; HNF4 [Metzger S. et al., 1993]  
[ExperimentCodes](#)  
 3.5 (HNF4) [Metzger S. et al., 1993]  
 rat liver cells: 3.6 (HNF4) [Metzger S. et al., 1993]  
 human HepG2 cells: 6.2 [Metzger S. et al., 1993]  
 Hel a cells: 6.2 6.6 (HNF4) [Metzger S. et al., 1993]  
 1.1.5 (HNF4) 3.3, 3.5 (HNF4), 4.2 (HNF4) [Ladias J.A. et al., 1993]



DNase I FOOTPRINTING

TRANSIENT EXPRESSION ANALYSIS

## **Экспериментальные методики, по результатам выполнения которых вносятся данные в базу TRRD**

<b>Type of experiment</b>	<b>Assay code in TRRD</b>
<b>Detection of transcription factor binding sites</b>	
<b>DNase I footprinting with nuclear extract</b>	<b>1.1.1</b>
<b>DNase I footprinting with purified or recombinant protein</b>	<b>1.1.5</b>
<b>Genomic footprinting</b>	<b>1.5</b>
<b>Methylation protection assay</b>	<b>4.1</b>
<b>Methylation interference assay</b>	<b>4.2</b>
<b>Electrophoretic mobility shift assay (EMSA) with nuclear extract</b>	<b>3.1</b>
<b>EMSA performed in the presence of competitive oligonucleotides</b>	<b>3.2</b>
<b>EMSA performed with mutant probes or competitors</b>	<b>3.3</b>
<b>Identification of DNA-binding proteins</b>	
<b>DNase I footprinting with purified or recombinant protein</b>	<b>1.1.5</b>
<b>DNase I footprinting with nuclear extract and specific antibodies</b>	<b>1.1.6</b>
<b>EMSA with purified or recombinant protein</b>	<b>3.5</b>
<b>EMSA with nuclear extract and specific antibodies</b>	<b>3.6</b>
<b>Confirming the functional importance of the site</b>	
<b>Insertion of isolated site 5' of homologous or heterologous promoter</b>	<b>6.3.2</b>
<b>Comprehensive mutant analysis</b>	<b>6.2</b>
<b>Trans-activation of a reporter gene by overexpression of a distinct transcription factor</b>	<b>6.6</b>
<b>Genomic footprinting</b>	<b>1.5</b>

**Полный список см. по адресу**

**<http://www.mgs.bionet.nsc.ru/mgs/gnw/trrd/digcodes.shtml>**

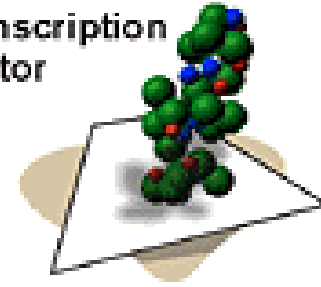
# ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА (html –представление)

<a href="#">TRRDSITES4.S5743</a>
<a href="#">SiteAC</a> S5743
<a href="#">GeneID</a> <a href="#">Gene: Hs:APOA1</a>
<a href="#">RegUnitAC</a> <a href="#">REGULATORY UNIT: P00051</a>
<a href="#">SiteName</a> T3R/RXR bs; T3R/RXR alpha binding site
NP RXR NP TR
<a href="#">SiteIndex</a> 2
<a href="#">FactorName</a> <a href="#">T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer</a> <a href="#">T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer</a>
<a href="#">FactorInfluence</a> decrease
<a href="#">DNA_BankLink</a>
<a href="#">PosContradiction</a> -214 to -192 [Tzameli I. and Zannis V.I., 1996]
<a href="#">ImportantPos</a> --TG--CC-TTGACCC-----; T3R beta/RXR alpha; [Tzameli I. and Zannis V.I.
<a href="#">ExperimentCodes</a> 3.5 (RXR alpha), 3.5 (T3R beta/RXR alpha) [ <a href="#">Tzameli</a> I. and Zannis V.I., COS-1 cells: 3.1, 3.2.2, 3.3, 3.6 (T3R beta), 4.2 [ <a href="#">Tzameli</a> I. and Zanni: HepG2 cells: 6.5 (T3R beta), 6.5 (9-cis RA), 6.5 (all-trans-RA), 6.6.1.1 6.6.1.1 (T3R beta), 6.6.1.1 (RXR alpha) [ <a href="#">Tzameli</a> I. and Zannis V.I., 19

# TRRDFACTORS

TRRDFACTORS

Transcription  
Factor



## ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА

ID [Hs:APOA1](#)

AN [Site: S1135](#)

TF PPARgamma/RXRalpha; PPARgamma and retinoic X receptor alpha heterodimer

FS PPARgamma; peroxisome proliferator-activated receptor gamma

TS human

TO in vitro synthesized

TR [\[Vu-Dac N. et al., 1994\]](#)

FS RXRalpha; retinoic X receptor alpha

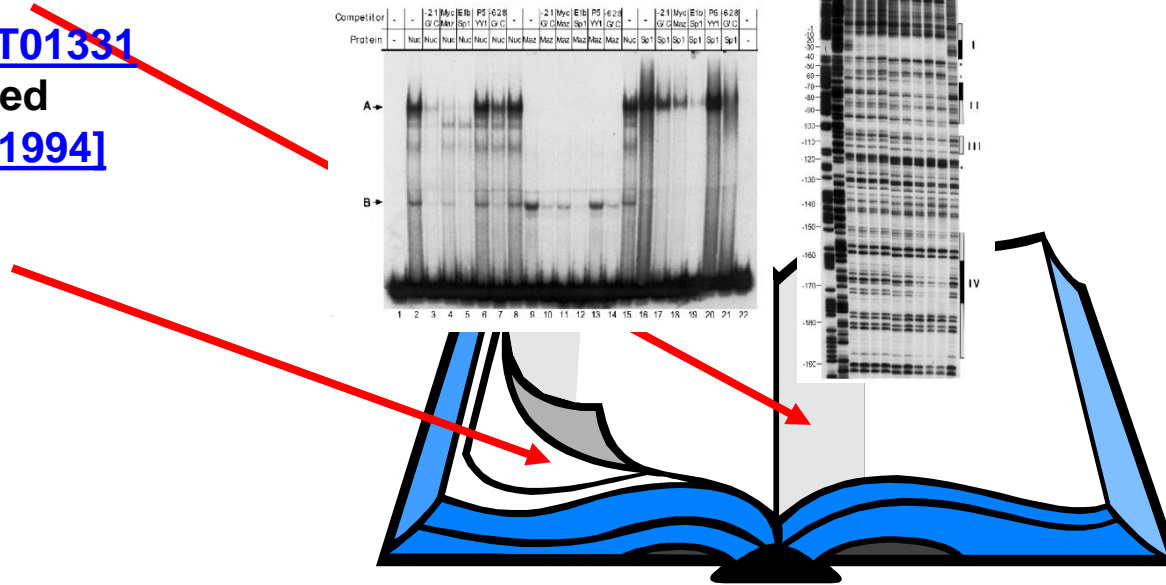
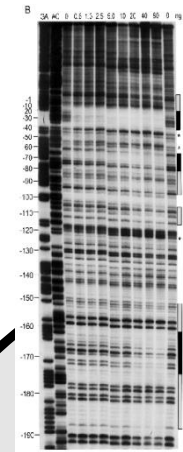
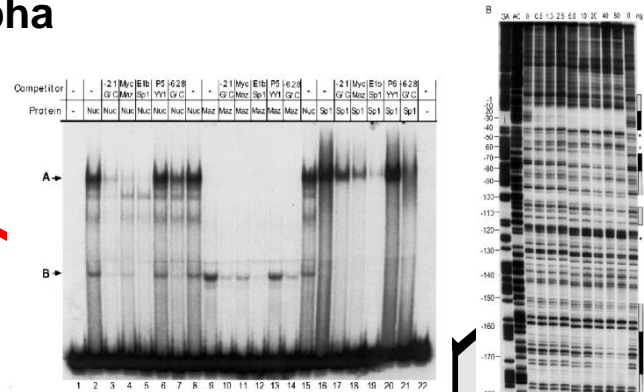
TS mouse

NF [TRANSFAC link: T01391](#)

TO in vitro synthesized

TR [\[Vu-Dac N. et al., 1994\]](#)

//





# ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА (html –представление)

[TRRDFACTORS4.F5743.1](#)

[Identifier](#)

F5743.1

[GeneID](#)

Hs:APOA1

[SiteAC](#)

Site: S5743

[FactorName](#)

T3R beta/RXR alpha; T3R beta heterodimer/retinoic X receptor alpha

[FactorSubunitName](#)

T3R beta;

[FactorSource](#)

recombinant

[Cells](#)

COS-1

[Reference](#)

[[Tzameli](#) I. and Zannis V.I., 1996]

[FactorSubunitName](#)

RXR alpha; retinoic X receptor alpha

[FactorSource](#)

recombinant

[Cells](#)

COS-1

[Reference](#)

[[Tzameli](#) I. and Zannis V.I., 1996]

//

[TRRDFACTORS4.F5743.2](#)

[Identifier](#)

F5743.2

[GeneID](#)

Hs:APOA1



# TRRDLCR

## ОПИСАНИЕ ЛОКУС- КОНТРОЛИРУЮЩЕГО РАЙОНА

AC C0002

IC Hs:ADA

CR O.A.P.

DT 15.02.2000

OS Homo sapiens (human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;  
Mammalia; Eutheria;

OC Primates; Catarrhini; Hominidae; Homo.

LO 20

LM ST; ;Exon1;;LADA:HSSII;; LADA:HSSIII;; LADA:EFS;;

LADA:Exon2

XX

GI ADA; adenosine deaminase

DR TRRD;; [A00862](#)

DR SWISS-PROT; [ADA\\_HUMAN](#); [P00813](#)(Expasy server)

XX

AL L1

LI Hs:LADA

TL thymus-specific

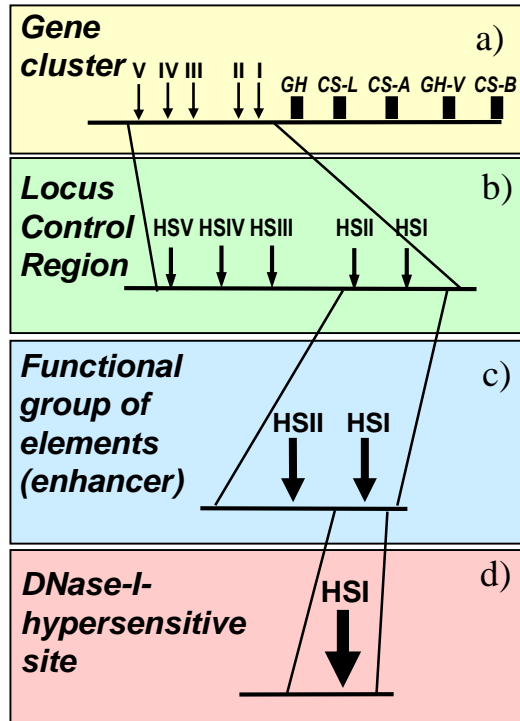
EL LADA:HSSII; LADA:HSSIII; LADA:EFS

BI EMBL; [HSADAG](#); [M13792](#)

.....

//

# TRRDLCR: ИЕРАРХИЧЕСКОЕ ОПИСАНИЕ



AC C0008  
 IC Hs:LGH  
 CR O.A.P.  
 DT 03.03.2000  
 OS Homo sapiens (h  
 OC Eukaryota; Metaz  
 Vertebrata; Mammali  
 OC Primates; Catarrh  
 LO 17q22-q24  
 LM LGH:HSV;~2000; LGH:HSIV;~3000;~  
 LGH:HSIII;~11100; LGH:HSII;~800;  
 LGH:HSI;~14600; ST GH1 gene;;CS:  
 CS2  
 XX  
 GI GH1; SOMATOTROPIN (GROWT

ALA7  
 LI Hs:LPGH  
 TL pituitary specific  
 BI EMBL; [AF010280](#); [AF010280](#)  
 EL LGH:HSV, LGH:HSIII, LGH ENHANCER  
 .....

AR R14  
 RF  
 RT pituitary specific  
 SV

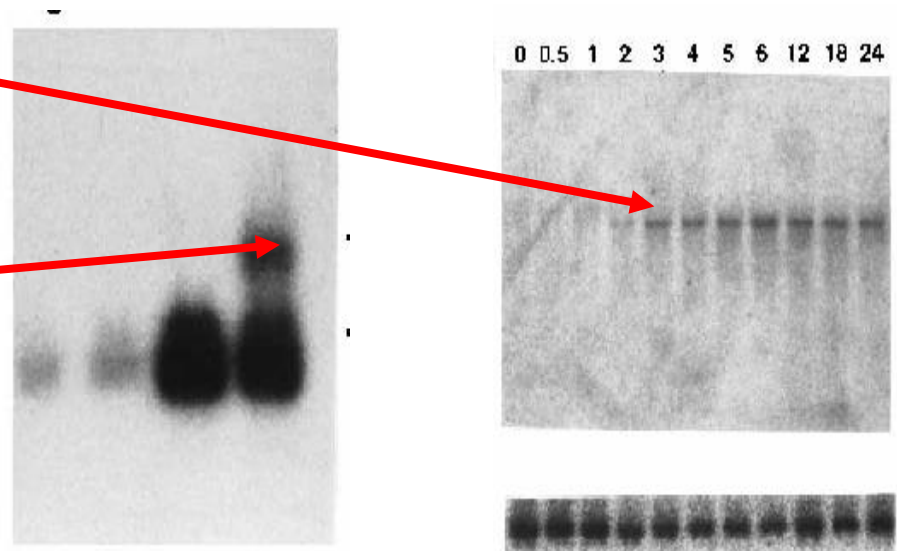
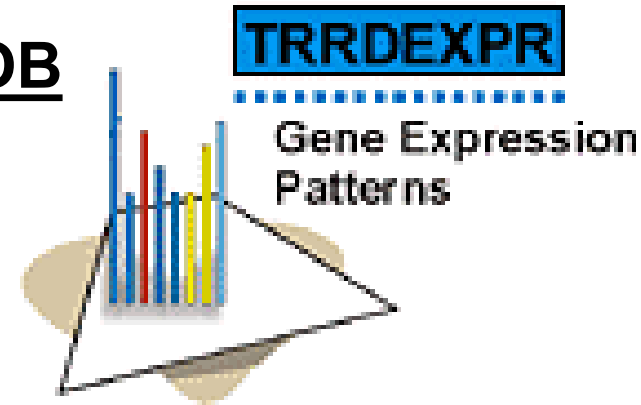
HI LGH:HSV  
 DE HSV  
 HT pituitary specific  
 DP ST of GH-1 gene  
 PT at -32.5kb  
 XX  
 PA positive  
 OA pituitary  
 TA GH-secreting adenoma  
 XX  
 CA primary pituitary cells  
 XX  
 AG Stable Transfection  
 AG transgenic analysis [Jones B.K. et al.,  
 1995]  
 AG DNase I hypersensitivity [Jones B.K. et  
 al., 1995]  
 XX  
 CA K562  
 AG Stable Transfection  
 AG DNase I hypersensitivity [Jones B.K. et  
 al., 1995]  
 AG Endonuclease protection assay

# TRRDEXP

## ОСОБЕННОСТИ ЭКСПРЕССИИ ГЕНОВ

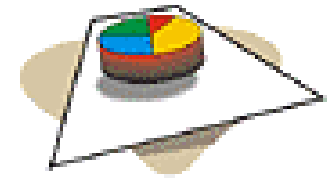
RE A00596.002  
ID Hs:AAP  
RT mRNA  
RN lymphoblastoid cells  
RL present  
RI heat shock  
FF induction  
RH 3 h  
RR [Abe K. et al., 1991]

RE A00596.001  
ID Hs:AAP  
RT mRNA  
RN astrocytes  
RL present  
RR [Amara F.M., et al., 1999 ]



RNA blot analysis

# TRRDIB



TRRDIB

Bibliography

NN 2027

ID [Hs:AAP](#)

AU Trejo J., Massamiri T., Deng T., Dewji N.N., Bayney R.M., Brown J.H.

TI A direct role for protein kinase C and the transcription factor Jun/AP-1

TI in the regulation of the Alzheimer's beta-amyloid precursor protein gene.

SO J.Biol.Chem.

VL 269

IS 34

YR 1994

PG 21682-21690

ML MEDLINE:[94342362](#), [\[See Related Articles\]](#)

//

Entrez-PubMed - Microsoft Internet Explorer

Файл Правка Вид Переход Избранное Справка

Адрес <http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=95074137&form=6&db=m&Dopt=b>

NCBI National Library of Medicine PubMed

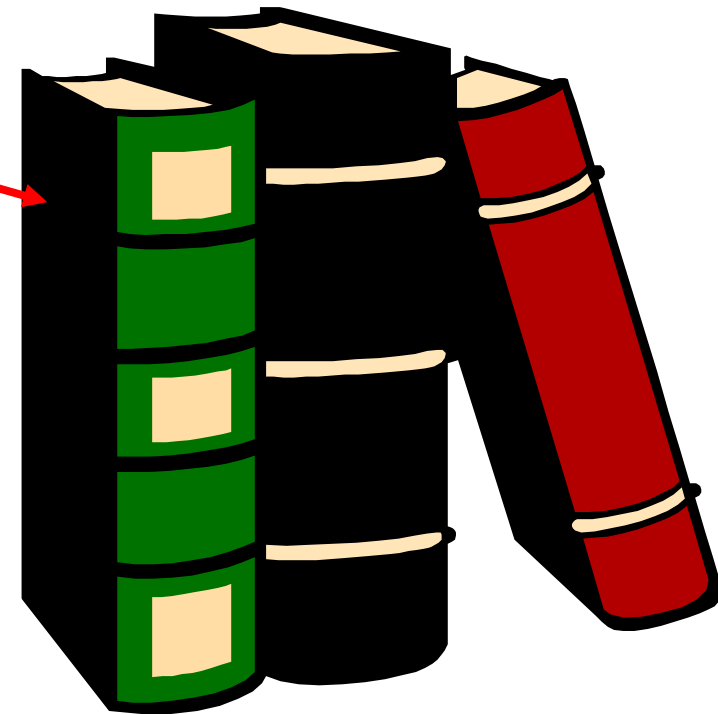
PubMed Nucleotide Protein Genome Stru

Search PubMed for Go Clear

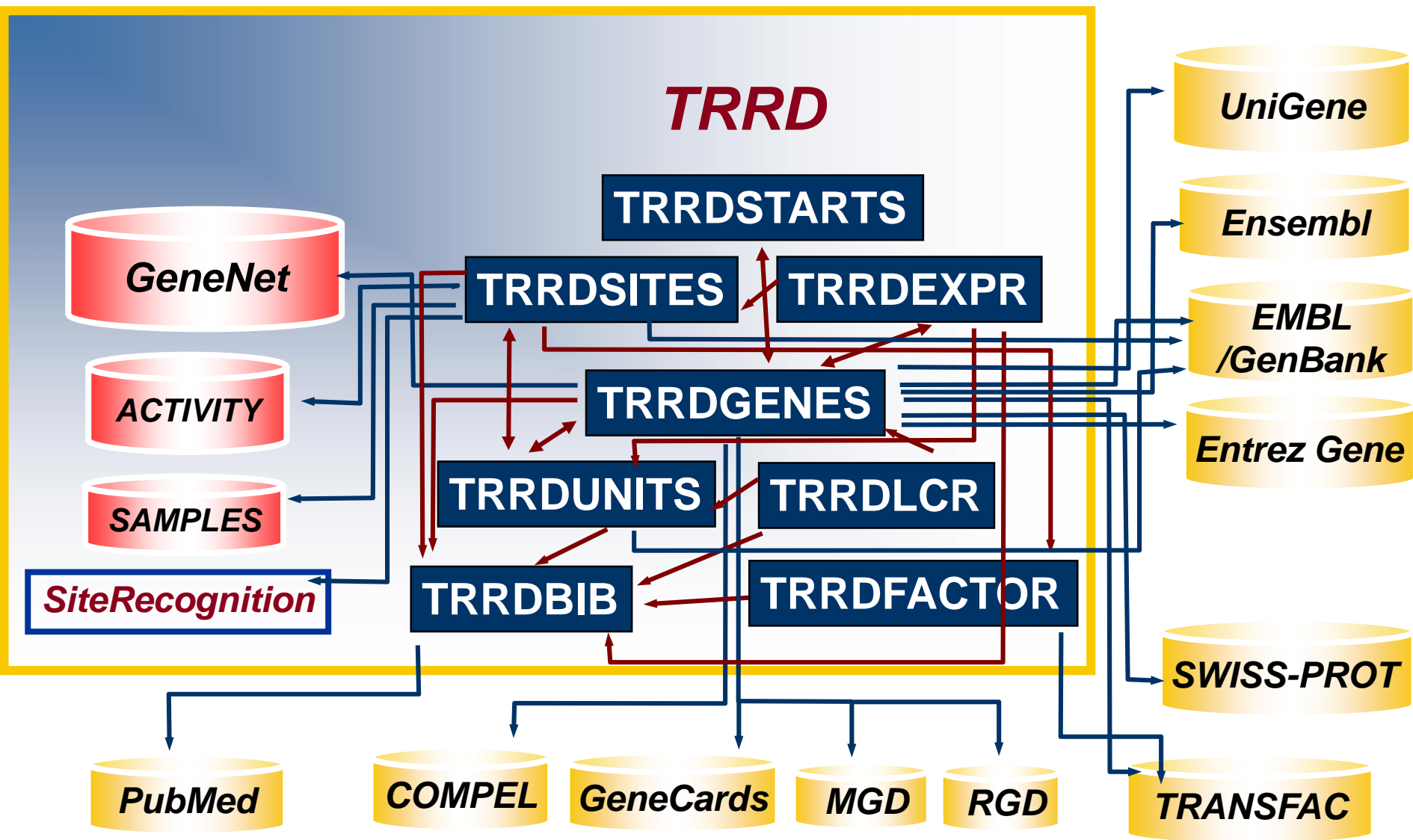
Limits Preview/Index History Clipboard

Display Abstract Save Text Order Add to Clipboard

1: *J Biol Chem* 1994 Dec 9;269(49):31012-8



# Линковка между таблицами TRRD и ссылки на внешние ресурсы.



# ГИПЕРССЫЛКИ НА БАЗЫ ДАННЫХ И ПРОГРАММНЫЕ МОДУЛИ СИСТЕМЫ GENEEXPRESS ИЗ БАЗЫ TRRDSITES

```

AN S161
ID Gene: Hs:PAI1
AP REGULATORY UNIT: 100150
NM USF bs;
WW http://www.mgs.bionet.nsc.ru/Programs/acts2/AFF\_USF.htm
DR ACTIVITY ; A00P026;
DR SAMPLES; USF;
TF USF; upstream sequence fac
AT increase
SQ gacatcacgtggct
PQ -574 to -559
PF -574 to -559
BF EMBL : X13523 : 236
AG 1.1 1, 1.1.2, 3.1, 3.3, 6.1, 6.2, 6.3 [Riccio A. et al., 1991]
    
```

TRRDGENE

ACTIVITY

SAMPLES

TRRDFACTORS

TRRDBIB

EMBL

Predicting activities of functional sites in DNA/RNA  
Affinity for the transcription factor USF (-ln(A/Am<sub>max</sub>))

Input DNA Sequence :  
 from Screen:  
 CTTTGAAG AGAATCCGGG CCCAGCAGC TCAAGACCAA  
 CGCCCCCA CCCCTACCC GTGCAGCTC GGGTACTCC  
 from DB:   [SPSS from Heidelberg (EMBL)]  
 from File:

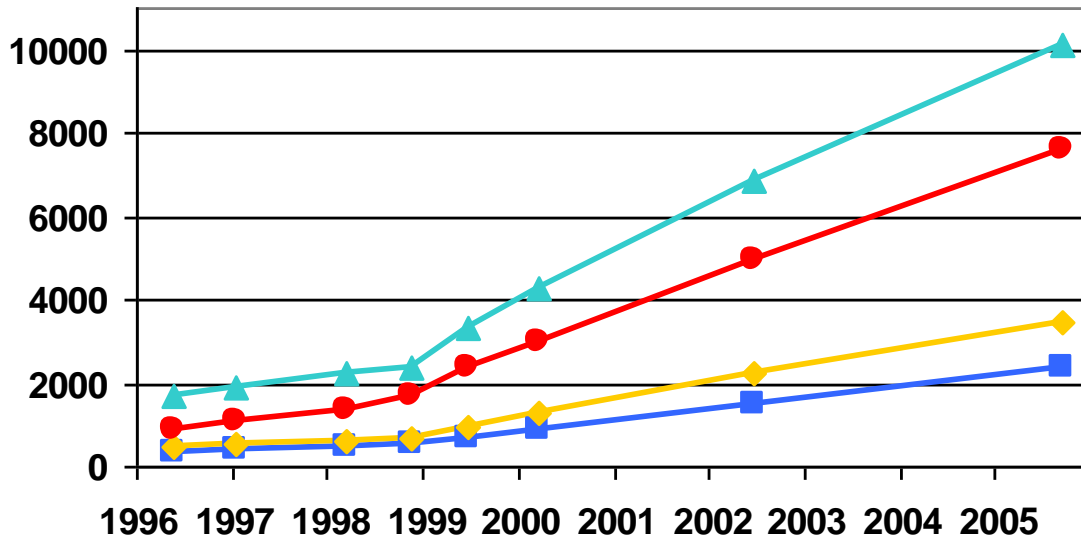
What to do:  
 Properties for activity: [Twist calculated (Karas,CABIOS,1996,12:441-446)]  
 Activity by properties: [USFbind by TwistDmax]

See  
 Twist calculated (Karas,CABIOS,1996,12:441-446)  
 Mo, sr, depth, calc. (Karas,CABIOS,1996,12:441-446)

USF Binding site  
for  
EMBL ID S75265

CTTTGAAG AGAATCCGGG CCCAGCAGC TCAAGACCAA  
 CGCCCCCA CCCCTACCC GTGCAGCTC GGGTACTCC  
 CCGTGGCTG GATACGGGCG CCTAGGGCAG GCAGGAGGAG

# ДИНАМИКА ПОПОЛНЕНИЯ TRRD



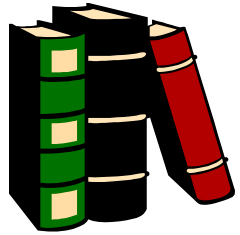
**10135 САЙТОВ СВЯЗЫВАНИЯ  
ТРАНСКРИПЦИОННЫХ  
ФАКТОРОВ  
7609 НАУЧНЫХ ПУБЛИКАЦИЙ**

**3400 РЕГУЛЯТОРНЫХ РАЙОНОВ  
2344 ГЕНОВ**

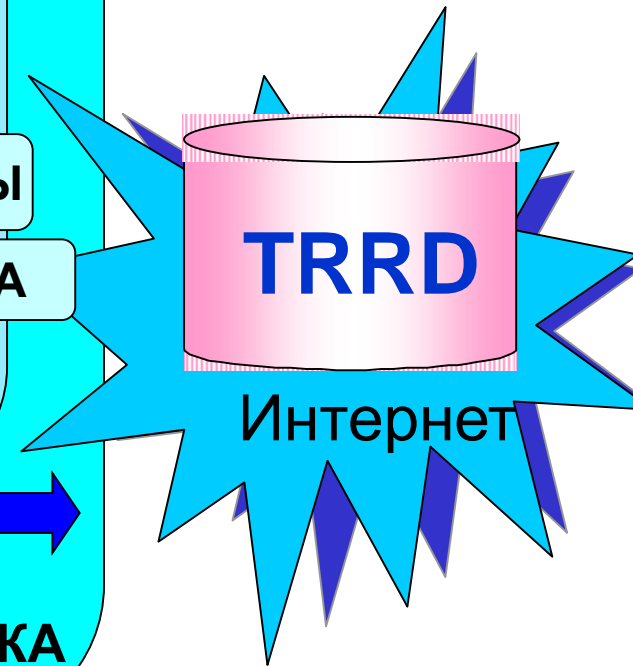
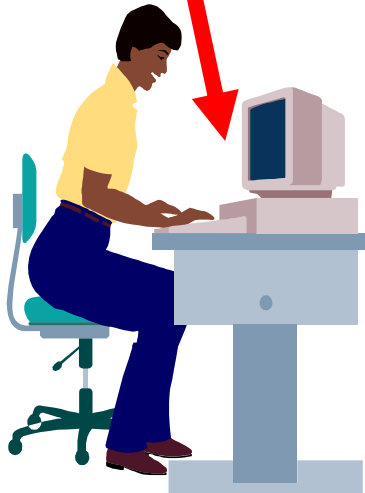
	Всего входов в TRRD.	Содержание по видам организмов (%)			
		Человек	Мышь	Крыса	Другие виды
<b>Гены</b>	<b>2344</b>	<b>32%</b>	<b>22%</b>	<b>15%</b>	<b>31%</b>
<b>Регуляторные единицы</b>	<b>3400</b>	<b>36%</b>	<b>19%</b>	<b>14%</b>	<b>31%</b>
<b>Сайты связывания транскрипционных факторов</b>	<b>10 135</b>	<b>36%</b>	<b>18%</b>	<b>14%</b>	<b>32%</b>



# ПРОЦЕСС ВВОДА ДАННЫХ В TRRD



ЭКСПЕРИМЕНТАЛЬНЫЕ СТАТЬИ



# ВВОД и ОБРАБОТКА ДАННЫХ

EXPERIMENTAL PAPERS

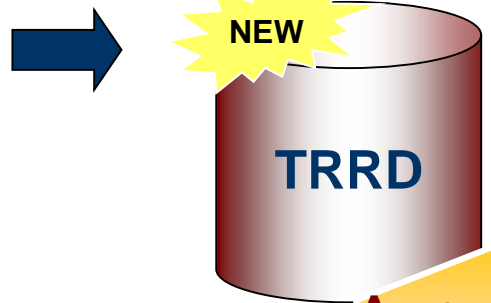


Синтаксический анализ

НОВЫЕ ТЕРМИНЫ

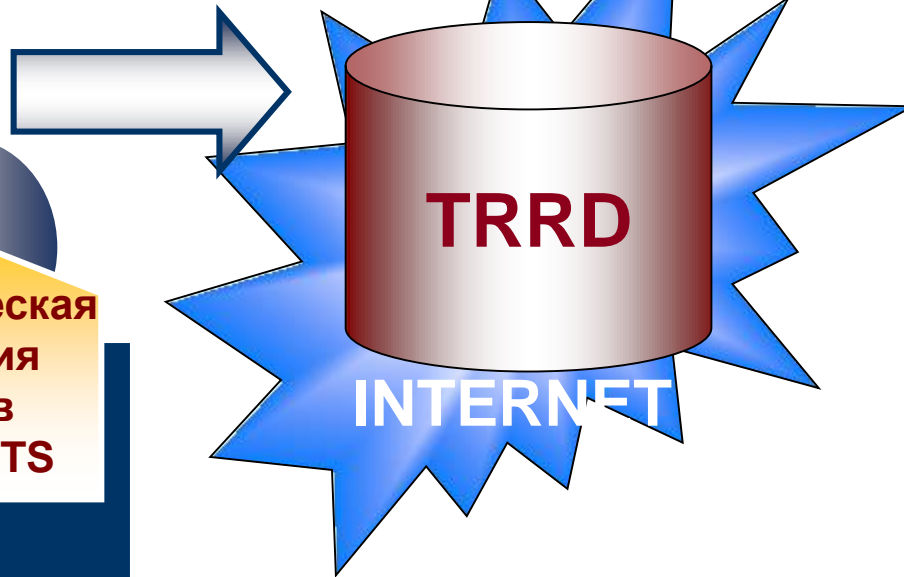


ЗАПРОСЫ



Семантический анализ

Автоматическая генерация полей в TRRDUNITS



# ИЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ КОНТРОЛИРУЕМЫХ СЛОВАРЕЙ МОРФОЛОГИЧЕСКИХ ТЕРМИНОВ В TRRD

*ОРГАН*

ВСЕ  
ОРГАНЫ

МОЗГ

ГИПОТАЛАМУС

СТВОЛ

*ЧАСТЬ  
ОРГАНА*

МЕДИОБАЗАЛЬНЫЙ  
ГИПОТАЛАМУС

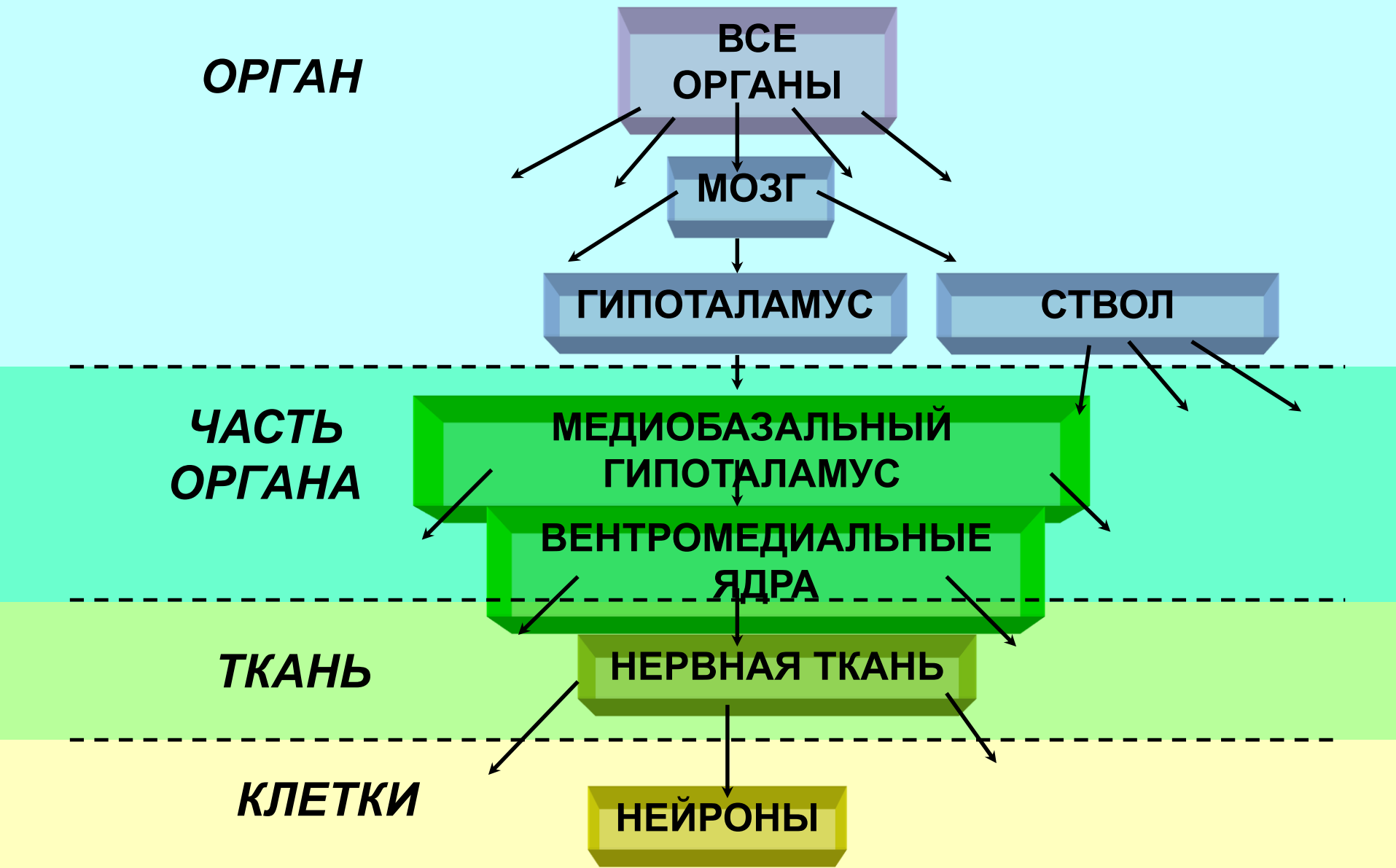
ВЕНТРОМЕДИАЛЬНЫЕ  
ЯДРА

*ТКАНЬ*

НЕРВНАЯ ТКАНЬ

*КЛЕТКИ*

НЕЙРОНЫ



# ПОИСК ДАННЫХ В TRRD: ЗАПРОС НА ОСНОВЕ ТЕЗАУРОСОВ ОРГАНОВ И ТКАНЕЙ МЛЕКОПИТАЮЩИХ

**НАЙДЕНО 95 ВХОДОВ : ГЕНЫ, ЭКСПРЕССИРУЕМЫЕ В KIDNEY OR KIDNEY CORTEX OR TUBULES OR GLOMERULUS OR PROXIMAL CONVOLUTED TUBULES**

# ПОИСК ДАННЫХ В TRRD: SRS

Gene Express 2.1

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

TRRD is a unique information resource, accumulating information on structural and functional organization of transcription regulatory regions of eukaryotic genes. Only experimental information is included into TRRD.

ACCESS to TRRD: [SRS ACCESS](#) [TRRDGENES](#) [TRRDEXP](#) [TRRDSITES](#) [TRRDFACTORS](#) [TRRDBIB](#) [TRRDUNITS](#)

[What's new?](#)

[Browse the TRRD](#)

[TRRD sections \(genes within functional systems\)](#)

**General information**

How to cite TRRD?  
TRRD publications  
The latest report on TRRD  
TRRD Workgroup  
Contact us  
Acknowledgments

**User's guide**

Database schema  
How to  
Integrati  
databas  
TRRD Vi  
FAQ  
What's r  
How is TR  
Standar  
input  
TRRD pr  
Current TR

TOP PAGE **QUERY** RESULTS SESSIONS VIEWS DATABANKS

Reset search TRRDGENES4

submit **Submit Query**

append wildcards to words

combine searches with **AND**

separate multiple values by & (and), | (or), ! (and not)

GeneAC

GeneAC

GeneAC

GeneAC

retrieve entries of type **Entry**

DATABASE	КОЛИЧЕСТВО ПОЛЕЙ, ПРОИНДЕКСИРОВАННЫХ для РЕЛИЗА 6.01
TRRDGENES	24
TRRDUNITS	11
TRRDEXP	17
TRRDSITES	16
TRRDFACTORS	14
TRRDLCR	40
TRRDBIB	9
<b>TOTAL NUMBER</b>	<b>131</b>

# ГЛАВНОЕ ОКНО ДЛЯ ПОИСКА В TRRDGENES ЧЕРЕЗ ПОИСКОВУЮ СИСТЕМУ SRS (Sequence Retrieval System)

The screenshot displays the TRRDGENES search interface, which is part of the SRS (Sequence Retrieval System). The interface is divided into two main sections: the query setup and the search results.

**Query Setup Section:**

- Navigation:** TOP PAGE, QUERY, RESULTS, SESSIONS, VIEWS, DATABANKS, HELP.
- Search Bar:** Reset, search TRRDGENES4, Info about field GeneAC.
- Submit Query:** A button to execute the search.
- Options:**
  - append wildcards to words
  - combine searches with AND
  - Number of entries to display per page: 30
  - Extended
- Query Fields:**
  - GeneName\_Full: apolipoprotein
  - Species: human
  - GeneAC: (empty)
  - GeneAC: (empty)
- Instructions:** separate multiple values by & (and), | (or), ! (and not)
- Views:** Use predefined view, Create your own view.
- Select fields to display:** GeneAC, GeneID, Updated.

**Search Results Section:**

- Navigation:** TOP PAGE, QUERY, RESULTS, SESSIONS, VIEWS, DATABANKS, HELP.
- Reset:** A button to clear the search.
- Query:** Query "[trrdgenes4-GeneName\_Full: apolipoprotein\*] & [trrdgenes4-Species: human\*]" found 9 entries.
- Perform operation:**
  - on all but selected
  - on selected
  - Buttons: Link, Save, View.
  - Filter: \*Names only\*
- Results List:**
  - [TRRDGENES4:A00150](#)
  - [TRRDGENES4:A00350](#)
  - [TRRDGENES4:A00151](#)
  - [TRRDGENES4:A00149](#)
  - [TRRDGENES4:A00147](#)
  - [TRRDGENES4:A00196](#)
  - [TRRDGENES4:A00264](#)
  - [TRRDGENES4:A00294](#)
  - [TRRDGENES4:A00148](#)

# ПОИСК ДАННЫХ В TRRD: БРАУЗЕРЫ И ТЕМАТИЧЕСКИЕ СЕКЦИИ

**Gene Express 2.1**

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

TRRD

General information  
How to cite TRRD?  
TRRD publications  
The latest report on TRRD  
TRRD Workgroup  
Contact us  
Acknowledgments  
User's guide  
Database schema  
How to search TRRD?  
Integration with other databases

TRRD is a unique information resource, accumulating information on structural and functional organization of transcription regulatory regions of eukaryotic genes. Only experimental information is included into TRRD.

ACCESS to TRRD:

- SRS ACCESS TRRDGENES TRRD EXP TRRDSITES TRRDFACTORS TRRDBIB TRRDUNITS TRRD FOR
- Browse the TRRD
- TRRD sections (genes within functional systems)

General information  
How to cite TRRD?  
TRRD publications  
The latest report on TRRD  
TRRD Workgroup  
Contact us  
Acknowledgments

User's guide  
Database schema  
How to search TRRD?  
Integration with other databases

How is TRRD updated ?  
Standardization of information input

Current TRRD release  
Information

What's new?

**Gene Express 2.1**

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

TRRD

General information  
How to cite TRRD?  
TRRD publications  
The latest report on TRRD  
TRRD Workgroup  
Contact us  
Acknowledgments  
User's guide  
Database schema

**Browse the TRRD**

Genes by name

Genes by species

**Gene Express 2.1**

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

TRRD

General information  
How to cite TRRD?  
TRRD publications  
The latest report on TRRD  
TRRD Workgroup  
Contact us  
Acknowledgments  
User's guide  
Database schema  
How to search TRRD?  
Integration with other databases  
TRRD Viewer  
FAQ  
What's new?  
How is TRRD updated ?  
Standardization of information input  
TRRD progress (from 1996)  
Current TRRD release

**TRRD sections**

While developing TRRD, the main attention was focused on description of genes within functional systems. The information on this gene groups can be obtained from TRRD sections

TRRD Section	Short name and link	Compiler
Heat Shock-Induced Genes	<a href="#">HS-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Interferon-Inducible Genes	<a href="#">IIG-TRRD</a>	<a href="#">Ananko E.A.</a>
Erythroid-Specific Regulated Genes	<a href="#">ESRG-TRRD</a>	<a href="#">Podkolodnaya O.A.</a>
Genes of Lipid Metabolism	<a href="#">LM-TRRD</a>	<a href="#">Ignatieva E.V.</a>
Endocrine System Genes	<a href="#">ES-TRRD</a>	<a href="#">Ignatieva E.V.</a>
Glucocorticoid-Regulated Genes	<a href="#">GR-TRRD</a>	<a href="#">Merkulova T.I.</a>
Plant Genes	<a href="#">PLANT-TRRD</a>	<a href="#">Goryachkovsky T.N.</a>
Cell Cycle-Dependent Genes	<a href="#">CYCLE-TRRD</a>	<a href="#">Kel-Margoulis O.V.</a>
Redox-Sensitive Genes	<a href="#">ROS-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Macrophage-Expressed Genes	<a href="#">MG-TRRD</a>	<a href="#">Ananko E.A.</a>

# Тематические секции TRRD

<http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/sections1.shtml>

TRRD Section	Short name and link	Compiler
Heat Shock-Induced Genes	<a href="#">HS-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Interferon-Inducible Genes	<a href="#">IIG-TRRD</a>	<a href="#">Ananko E.A.</a>
Genes Expressed in B cells	<a href="#">B-TRRD</a>	<a href="#">Ananko E.A.</a>
Genes Related to EBV Infection and EBV Transformation	<a href="#">EBV-TRRD</a>	<a href="#">Ananko E.A.</a>
Erythroid-Specific Regulated Genes	<a href="#">ESRG-TRRD</a>	<a href="#">Podkolodnaya O.A.</a>
Genes of Lipid Metabolism	<a href="#">LM-TRRD</a>	<a href="#">Ignatieva E.V.</a>
Endocrine System Genes	<a href="#">ES-TRRD</a>	<a href="#">Ignatieva E.V.</a>
Glucocorticoid-Regulated Genes	<a href="#">GR-TRRD</a>	<a href="#">Merkulova T.I.</a>
Plant Genes	<a href="#">PLANT-TRRD</a>	<a href="#">Goryachkovsky T.N.</a>
Cell Cycle Genes	<a href="#">CCG-TRRD</a>	<a href="#">Turnaev I.I.</a>
Redox-Sensitive Genes	<a href="#">ROS-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Genes Expressed in Endocrine Pancreas	<a href="#">EP-TRRD</a>	<a href="#">Ignatieva E.V.</a>
Macrophage-Expressed Genes	<a href="#">MG-TRRD</a>	<a href="#">Ananko E.A.</a>
Genes, controlling blood coagulation and fibrinolysis	<a href="#">BCF-TRRD</a>	<a href="#">Khlebodarova T.M., Podkolodnaya O.A.</a>
Apoptosis Genes	<a href="#">Apoptosis-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Hepatitis C virus-induced Genes	<a href="#">HCV-TRRD</a>	<a href="#">Stepanenko I.L.</a>
Genes, controlling circadian rhythm, and genes with circadian expression	<a href="#">CLOCK-TRRD</a>	<a href="#">Khlebodarova T.M.</a>
Genes encoding proteins involved in the Fe metabolism	<a href="#">FM-TRRD</a>	<a href="#">Mischenko E.L. , Podkolodnaya O.A.</a>



# ПОИСК ДАННЫХ В TRRD: BLAST

**M S G Gene Express 2.1**  
HOME DNA RNA PROTEIN GENENETWORKS MAP

**TRANSCRIPTION REGULATORY REGIONS DATABASE**  


TRRD is a unique information resource, accumulating information on structural and functional organization of transcription regulatory regions of eukaryotic genes. Only experimental information is included into TRRD.

What's new?

**ACCESS to TRRD:**  
[SRS ACCESS](#) [TRRDGENES](#) [TRRDEXP](#) [TRRDSITES](#) [TRRDFACTORS](#) [TRRDBIB](#) [TRRDUNITS](#)  
[TRRDLOR](#)  
[Browse the TRRD](#)  
[Search for the TRRD \(by tissue-specificity, inducibility\)](#)  
[TRRD sections \(genes within functional systems\)](#)

**General information**  
[How to cite TRRD?](#)  
[TRRD publications](#)  
[The latest report on TRRD](#)  
[TRRD Workgroup](#)  
[Contact us](#)  
[Acknowledgments](#)

**User's guide**  
[Database schema](#)  
[How to search TRRD?](#)  
[Integration with other databases](#)  
[TRRD Viewer](#)  
[FAQ](#)

**How is TRRD updated ?**  
[Standardization of information input](#)  
[TRRD progress \(from 1996\)](#)

**Current TRRD release**  
[Information contents](#)  
[TRRD statistics](#)


**TRRD Tools**  
[Glossary on organs and tissues in mammals](#)  
[Search for regions of homology](#)

**SEQUENCE UNDER INVESTIGATION:**  
gtgtgaagaggagcgtactttttgtgtgtgaac

**BLAST search TRRDUNITS**

**QUERY RESULTS:**  
Query : gtgtgaagaggagcgtactttttgtgtgtga  
| | | | | | | | | |  
Subject : 113 agcgtactttt 103

**M S G Gene Express 2.1**  
HOME DNA RNA PROTEIN GENENETWORKS MAP

**TRANSCRIPTION REGULATORY REGIONS DATABASE**  


**TOOLS:**  
[Blast search TRRD database](#)  
[BinomSite program](#)

You may perform the search for regions homologous described in TRRD in the sequence of interest. The p by binomial probability estimation of the similarity between each of the transcription factor binding sites describe

**General information**  
[How to cite TRRD?](#)  
[TRRD publications](#)  
[The latest report on TRRD](#)  
[TRRD Workgroup](#)  
[Contact us](#)

# TRRD – информационная основа для создания выборок

```
ID es_250_1; DNA
AC es_250_1
CC DE adrenodoxin gene
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea
DR EMBL; M23665; HSADRDO01; ; join(133..382)
```

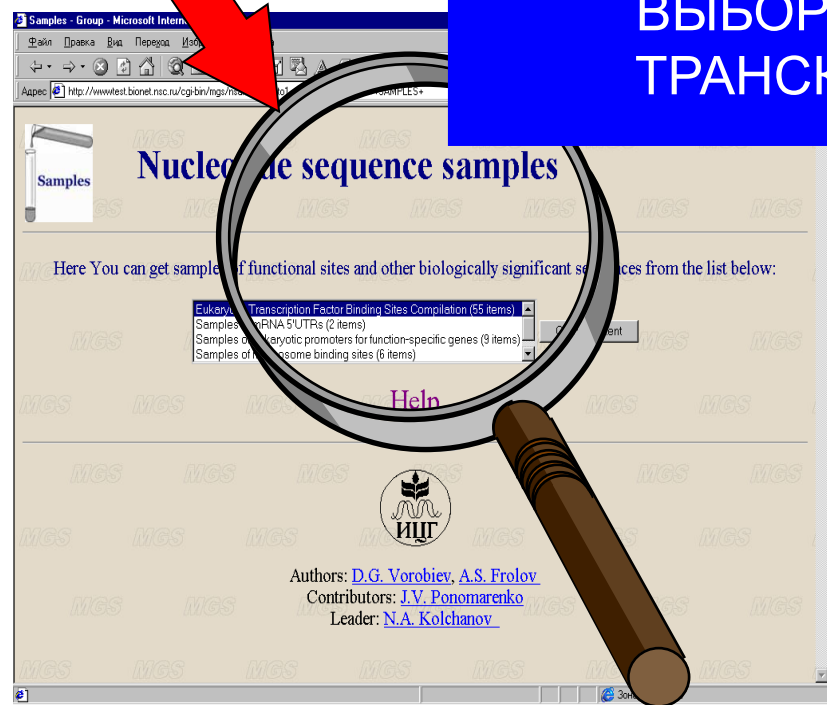
TRRDSITES

CCCGGGAGGGCCCTTTGGACCTTTTG  
CTAGCAACAGATGCGTCA CTGATGGCC  
-10 Y-box -90 -74 Y-box -65

ВЫБОРКИ ЭУКАРИОТИЧЕСКИХ ПРОМОТОРОВ

```
SQ ctttcaaaat attttgtttc tgcacggcaa cttcagccgc ta
tccagcttac aacggaacct ggagggttgg taaaggcccc ct
```

ВЫБОРКИ САЙТОВ СВЯЗЫВАНИЯ  
ТРАНСКРИПЦИОННЫХ ФАКТОРОВ



```
HOMO SAPIENS (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea
EMBL; M23665; HSADRDO01; ; join(133..382)
ST (EMBL/GENBANK) 333
TRRDGENES; A00860; Hs:ADX; 4.2;
{0,0} [1;250]; EXP
ctttcaaat attttgtttc tgcacggcaa cttcagccgc ta
tccagcttac aacggaacct ggagggttgg taaaggcccc ct
cccgcccat gggaccgggc ggcgtgggcg tgagagcttgc
gctctgcttg ccaatgtctt tataggctac ccggagcttgc
cggcgcggtg cttccagcag ggtctctccg ccaacttctccg
```

# Интернет-ресурсы по транскрипции у растений

## ПРОГРАММА

Chow CN et al., **PlantPAN 2.0**: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1154-60.

*Hem Triticum*

*ЕСТЬ: Arabidopsis thaliana, Brachypodium distachyon, Chlamydomonas reinhardtii, Glycine max, Malus domestica, Oryza sativa, Populus trichocarpa, Sorghum bicolor, Volvox carteri, Zea mays*

Zhang T, Marand AP, Jiang J. **PlantDHS**: a **database** for DNase I hypersensitive sites in **plants**. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1148-53.

*Hem Triticum*

*ЕСТЬ Arabidopsis, Rice, Brachypodium*

Hieno A, et al., **PPdb**: plant promoter database version 3.0. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D1188-92.

*Hem Triticum*

*ЕСТЬ Arabidopsis thaliana, Oryza sativa, Physcomitrella patens, Poplar*

Jin J, et al., **PlantTFDB 3.0**: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D1182-7.

*165 таксономических групп, из них 38 видов однодольных и 100 видов двудольных*

# PlantPAN 2.0 (<http://plantpan2.itps.ncku.edu.tw/>)

The screenshot shows the PlantPAN 2.0 website. At the top, there is a green navigation bar with the text "PlantPAN 2.0" and links for "HOME", "LINK", "GUIDE", and "ABOUT". Below the navigation bar, the title "PlantPAN 2.0" is centered. A paragraph describes the tool: "The Plant Promoter Analysis Navigator (PlantPAN; <http://plantpan2.itps.ncku.edu.tw/>) provides an informative resource for detecting transcription factor binding sites (TFBSs), corresponding TFs, and other important regulatory elements (CpG islands and tandem repeats) in a promoter or a set of promoters in plants. The current PlantPAN release (version 2.0) contains 16,960 TFs and 1,143 matrices of TF binding sites among 76 plant species." Below this is a highlighted orange box containing the text "★★★★ PlantPAN 2.0 TF position weight matrix download is available. ★★★★★" and a green button labeled "Go To Download". A "Citation:" section follows, listing the authors and their 2015 publication in Nucleic Acids Res. Below the citation are five feature cards: "Gene Search", "TF/TFBS Search", "Gene Group Analysis", "Promoter Analysis", and "Cross Species", each with an icon and a brief description of its function. At the bottom, a "LINK" section contains a "Download transcription factor position weight matrix" area with four buttons for "All plants", "Arabidopsis thaliana", "Oryza sativa", and "Zea mays". Below that is an "External sources" section with a link to "AtPAN" and its URL.

Chow CN et al., PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1154-60.

*Hem Triticum*



## Plant Transcription Factor Database

v4.0

Previous version: v1.0, v2.0, v3.0

Home BLAST Prediction Download Help About RegMap ATRM Links  Search (e.g., LFY)

### *Triticum aestivum* Transcription Factors

The gene annotation from IWGSC(v2.2) is used to identify transcription factors (TFs) of *Triticum aestivum* (See [datasource](#)). According to the family assignment rules, 3606 TFs (3606 loci) are identified and classified into 56 families. You can download the TF list ([here](#)) and protein sequences of TFs ([here](#)), or download page.

#### Browse by Family

<b>AP2 (43)</b>	ARF (45)	ARR-B (22)	B3 (140)
BES1 (10)	C2H2 (224)	C3H (100)	CAMTA (20)
CPP (24)	DBB (17)	Dof (52)	E2F/DP (24)
ERF (181)	FAR1 (201)	G2-like (100)	GATA (48)
GRF (16)	GeBP (12)	HB-PHD (6)	HB-other (44)
HRT-like (3)	HSF (53)	LBD (61)	LFY (2)
M-type_MADS (77)	MIKC_MADS (51)	MYB (263)	MYB_related (227)
NF-X1 (2)	NF-YA (22)	NF-YB (34)	NF-YC (20)
RAV (8)	S1Fa-like (3)	SBP (37)	SRS (5)
TALE (52)	TCP (28)	Trihelix (46)	VOZ (6)
WRKY (171)	Whirly (7)	YABBY (25)	ZF-HD (20)
bZIP (186)			

©2010-, Center for Bioinformatics, Peking University  
 Last Modified: 2016-10-15  
 Questions or Comments, please contact  
[plantfdb@mail.cbi.pku.edu.cn](mailto:plantfdb@mail.cbi.pku.edu.cn)

165 таксономических групп, из них 38 видов однодольных и 100 видов двудольных



## Plant Transcription Factor Database

v4.0

Previous version: v1.0, v2.0, v3.0

Home BLAST Prediction Download Help About RegMap ATRM Links  Search (e.g., LFY)

### Transcription Factor Information

Basic Information | Signature Domain | Sequence | Protein Features | Gene Ontology | Regulation | Orthologous Group | Publication

#### Basic Information [? help](#) [Back to Top](#)

TF ID	Traes_1AS_90F3120F8.1		
Organism	<i>Triticum aestivum</i>		
Taxonomic ID	4565		
Taxonomic Lineage	cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; Mesangiospermae; Liliopsida; Petrosavidae; commelinids; Poales; Poaceae; BOP clade; Pooidae; Triticeae; Triticeae; Triticeae; Triticeae		
Family	AP2		
Protein Properties	Length: 387aa	MW: 42935.1 Da	PI: 9.8162
Description	AP2 family protein		
Gene Model	<b>Gene Model ID</b>	<b>Type</b>	<b>Source</b> <b>Coding Sequence</b>
	Traes_1AS_90F3120F8.1	genome	IWGSC <a href="#">View CDS</a>

#### Signature Domain [? help](#) [Back to Top](#)

No.	Domain	Score	E-value	Start	End	HMM Start	HMM End
1	AP2	53.9	4.4e-17	58	107	1	55
2	AP2	40.8	5.5e-13	150	200	1	55

```

AP2 1 sgykGVndkkgrIVaEIndpsengkr.krfs1gkfgtaeeAakaalaank1leg 55
syykGVndkkgrIVaEIndpsengkr.krfs1gkfgtaeeAakaalaank1leg 55
Traes_1AS_90F3120F8.1 58 SQYRVTYVYRTRGRHESHIDV-----CgkQVYLGFDTAHAARAYDRAAIKFRG 107
78*****55*****997 PP
    
```

Orthologous Group  
 Best hit in *Arabidopsis thaliana*

Jin J, Zhang H, Kong L, Gao G, Luo J. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 2014 Jan;42(Database issue):D1182-7.

## Характеристики баз данных = «Критерии полезности баз данных»:

- Сущности, представленные в БД (промоторы, старты транскрипции, сайты связывания, MAR/SAR элементы и т.д.)
- Способы наполнения БД (ручная аннотация, объединение данных из разных источников (интеграция), компьютерные методы предсказания сайтов и промоторов, обработка широкомасштабных экспериментов и т.д.);
- Достоверность информации;
- Виды организмов, представленные в БД;
- Количество данных, имеющихся в базе (полногеномный уровень ?);
- Программные средства для поиска, анализа и экстракции данных

# Часть 2. ТРАНСЛЯЦИЯ

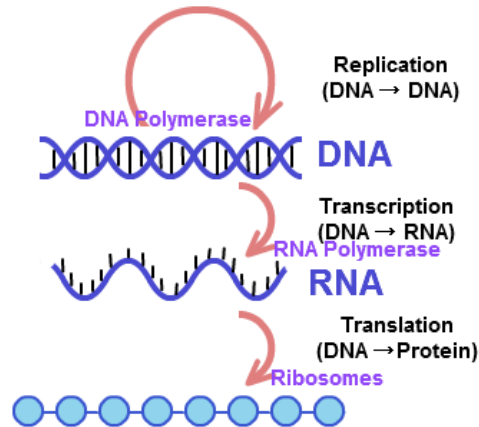
к.б.н., с.н.с. лаб. эволюционной биоинформатики  
и теоретической генетики Игнатьева Е.В.

Эта часть лекции подготовлена с использованием материалов,  
любезно предоставленных член-корр. РАН, д.б.н. Кочетовым А.В.

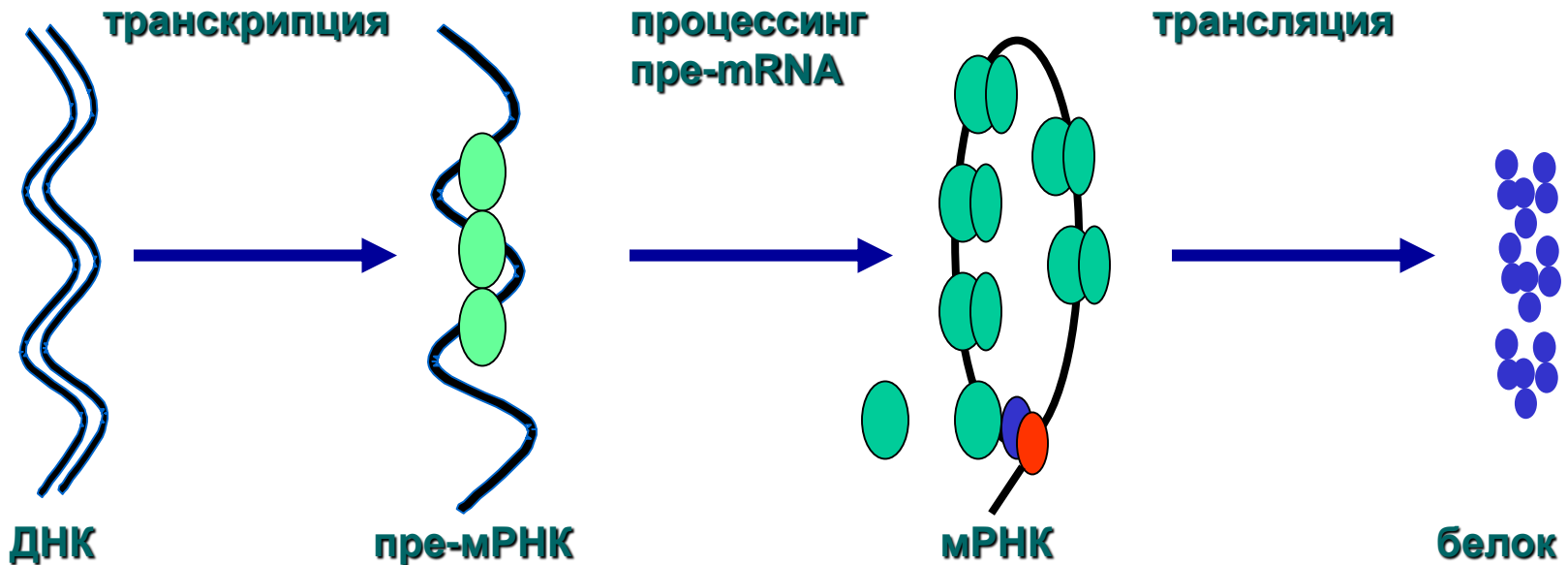


Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

# Трансляция – один из фундаментальных биологических процессов



**Трансляция** – это процесс перехода генетической информации от мРНК к белку = процесс синтеза белка из аминокислот на матрице информационной (матричной) РНК, осуществляемый рибосомой.



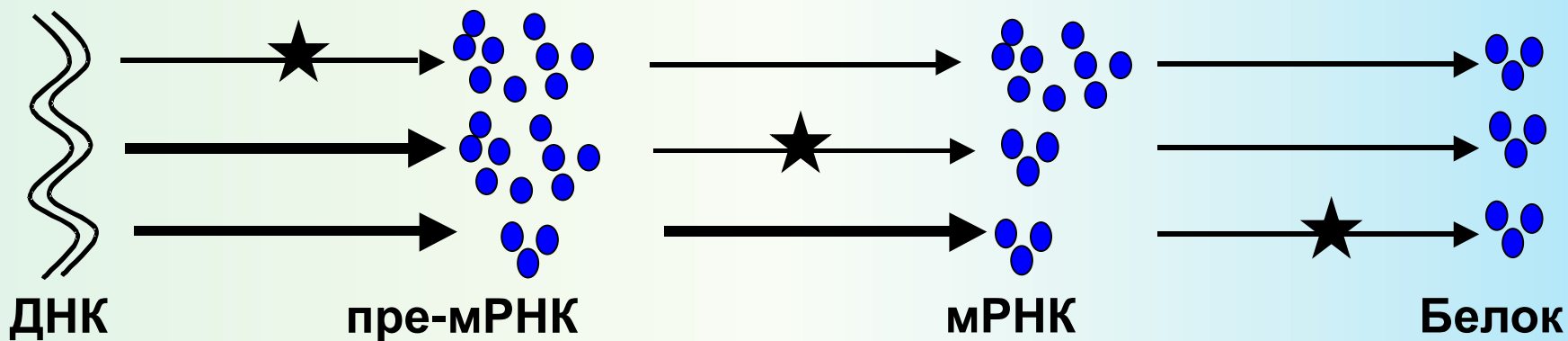


# Концепция «лимитирующего звена» в приложении к процессу экспрессии

*Высокий уровень экспрессии: высокоэффективны все стадии*



Низкий уровень экспрессии (★ - низкоэффективные стадии)

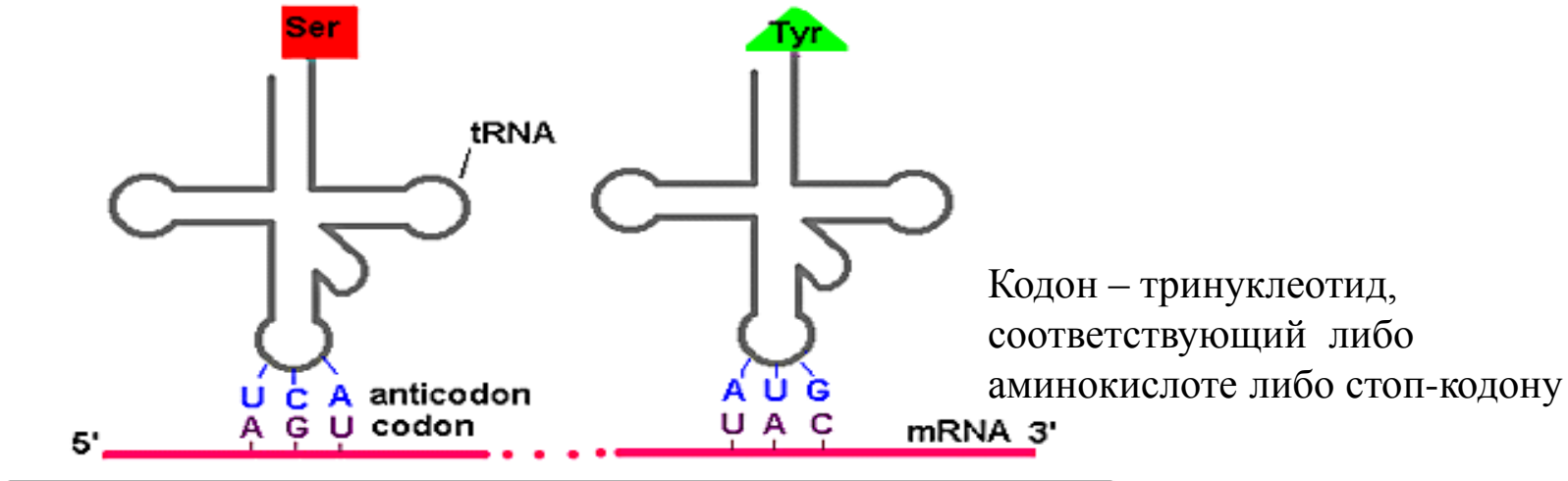


# Основные участники процесса трансляции

- 1) мРНК
- 2) рибосома – сложный РНК-белковый комплекс
- 3) транспортные РНК (тРНК) – посредники между пулом свободных аминокислот и трансляционным аппаратом
- 4) аминоацил тРНК-синтетазы – распознают «свои» тРНК, соответствующие каждой аминокислоте, и осуществляют связывание аминокислот и тРНК,
- 5) Регуляторные белки – факторы инициации, элонгации, терминации и др.

# Транспортные РНК (тРНК)

распознают кодоны в белок-кодирующих последовательностях мРНК с помощью комплементарных взаимодействий, в которых участвуют антикодоны.



## Генетический код – соответствие кодонов и аминокислот

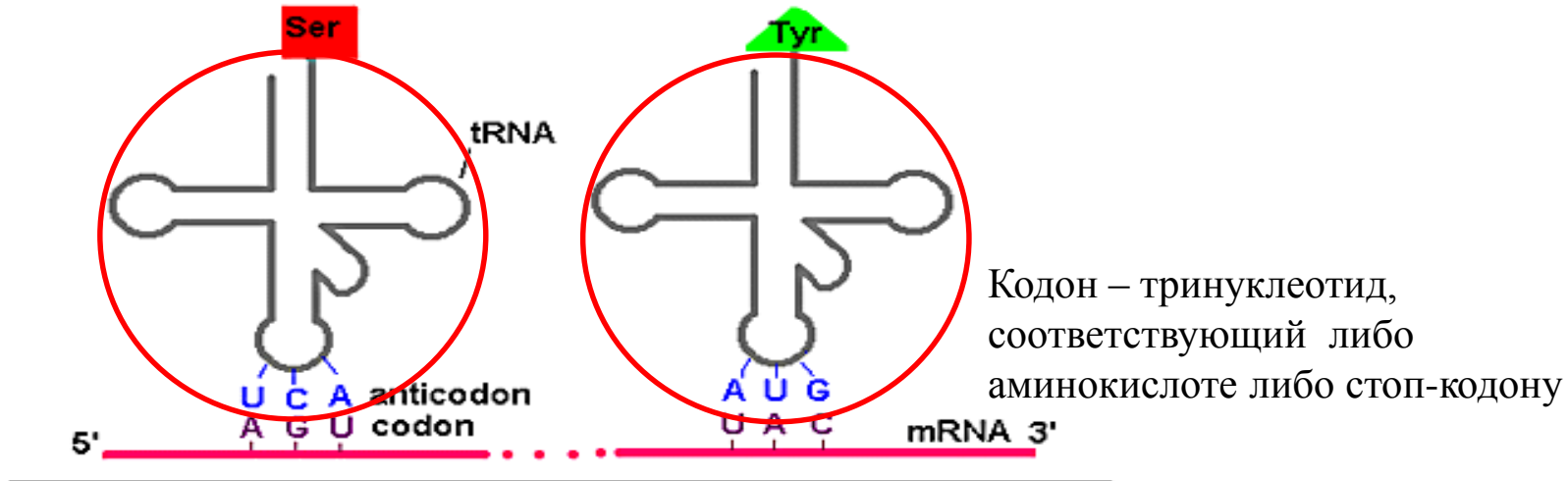
		2nd base in codon				
		U	C	A	G	
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr <b>STOP</b> <b>STOP</b>	Cys Cys <b>STOP</b> Trp	U C A G
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

3rd base in codon

- Число тринуклеотидных комбинаций из 4 нуклеотидов = равно 64 ( $4^3$ )
- Три кодона являются нонсенс-кодонами и выполняют функцию терминаторов трансляции
- Каждой из 20 аминокислот соответствует от одного до шести кодонов.
- Кодоны, кодирующие одну аминокислоту, они называются синонимическими

# Транспортные РНК (тРНК)

распознают кодоны в белок-кодирующих последовательностях мРНК с помощью комплементарных взаимодействий, в которых участвуют антикодоны.

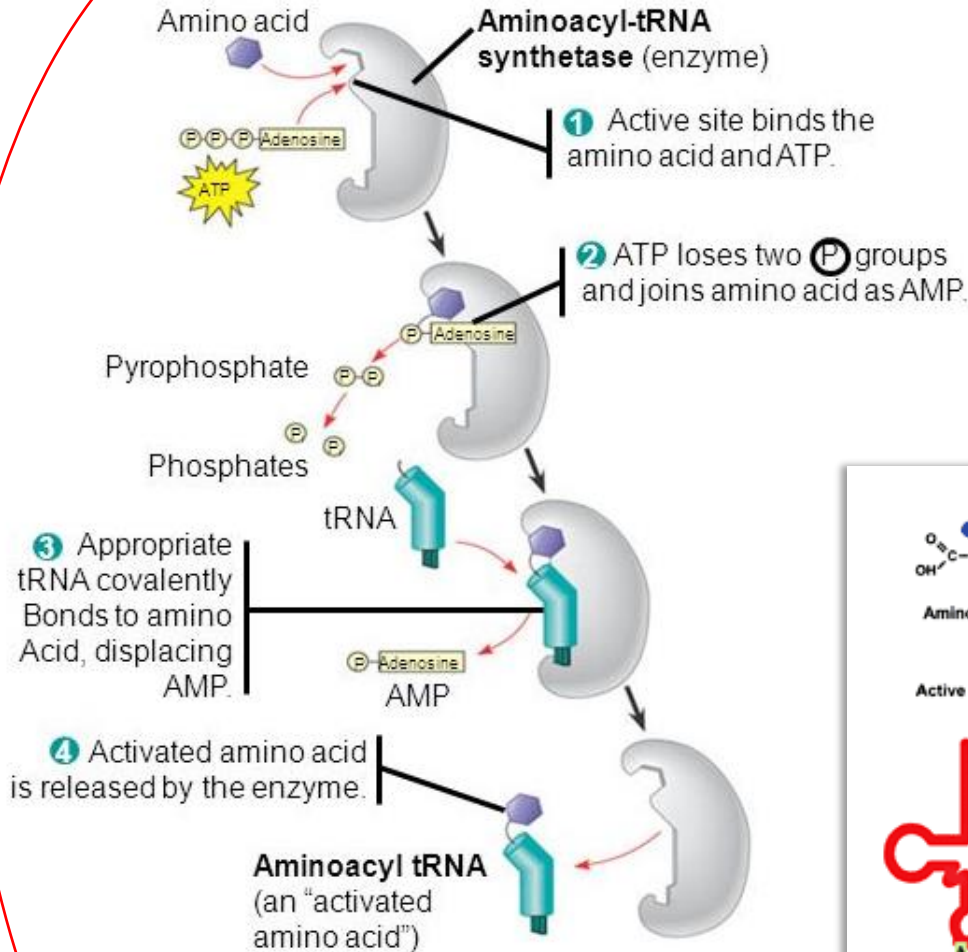


## Генетический код – соответствие кодонов и аминокислот

		2nd base in codon				
		U	C	A	G	
1st base in codon	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr <b>STOP</b> <b>STOP</b>	Cys Cys <b>STOP</b> Trp	U C A G
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

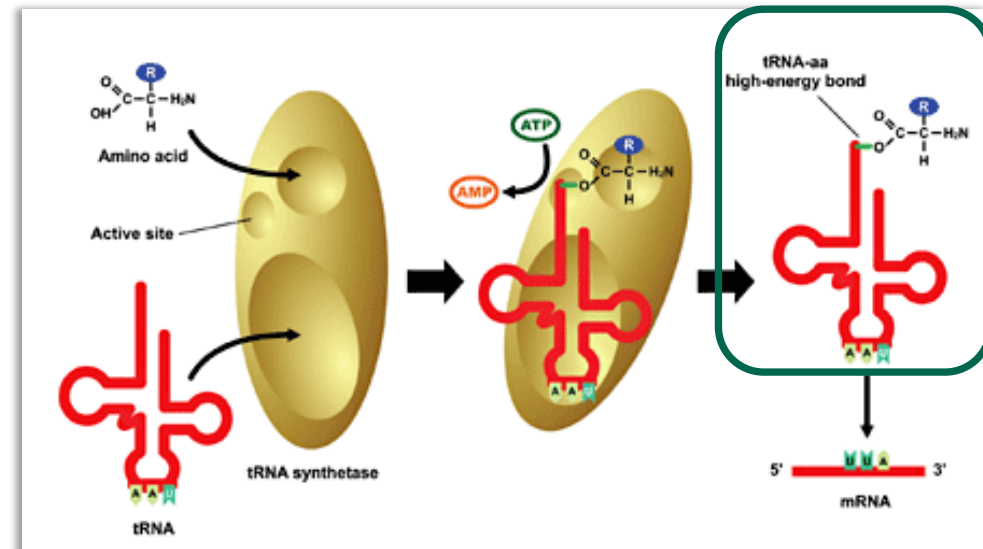
- Число тринуклеотидных комбинаций из 4 нуклеотидов = равно 64 ( $4^3$ )
- Три кодона являются нонсенс-кодонами и выполняют функцию терминаторов трансляции
- Каждой из 20 аминокислот соответствует от одного до шести кодонов.
- Кодоны, кодирующие одну аминокислоту, они называются синонимическими

# Функция фермента аминокил-тРНК-синтетазы:



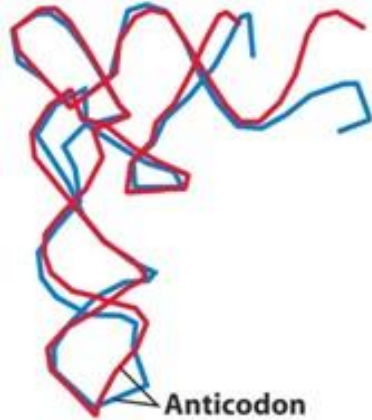
## Аминоацил-тРНК

= Аминокислота прикреплена сложноэфирной связью своей карбоксильной группой к 2' или 3'-гидроксильной группе рибозы 3'-концевого нуклеотида тРНК (это всегда аденин)



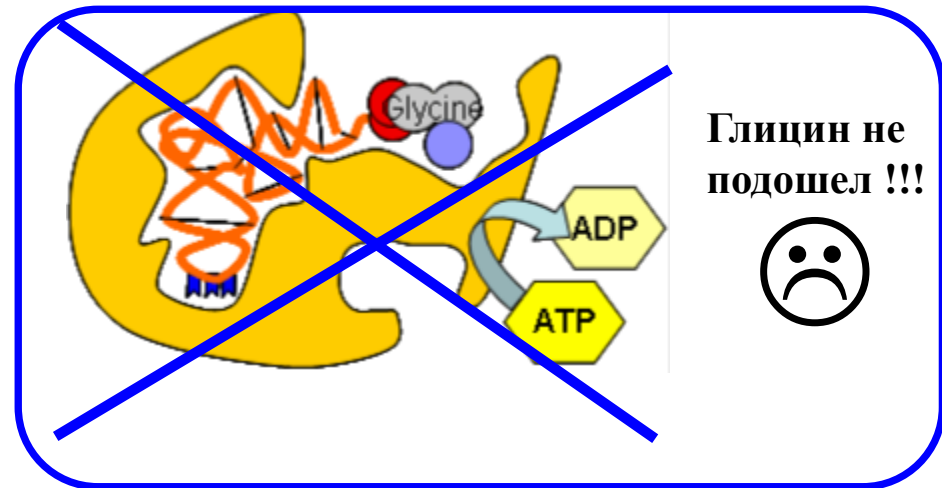
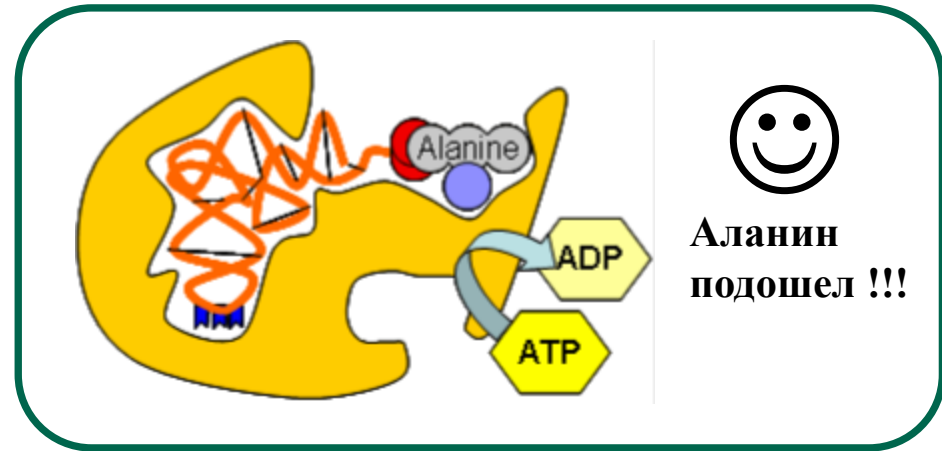
Аминоацил тРНК-синтетазы – распознают «свои» тРНК, соответствующие каждой аминокислоте, и осуществляют связывание аминокислот и тРНК, в силу чего такие аминокислотированные тРНК несут аминокислоты. В этом заключается связь между генетическим кодом – представленным антикодоном тРНК и аминокислотой, с этой тРНК сцепленной

# Аминоацил-тРНК-синтетаза: специфичность



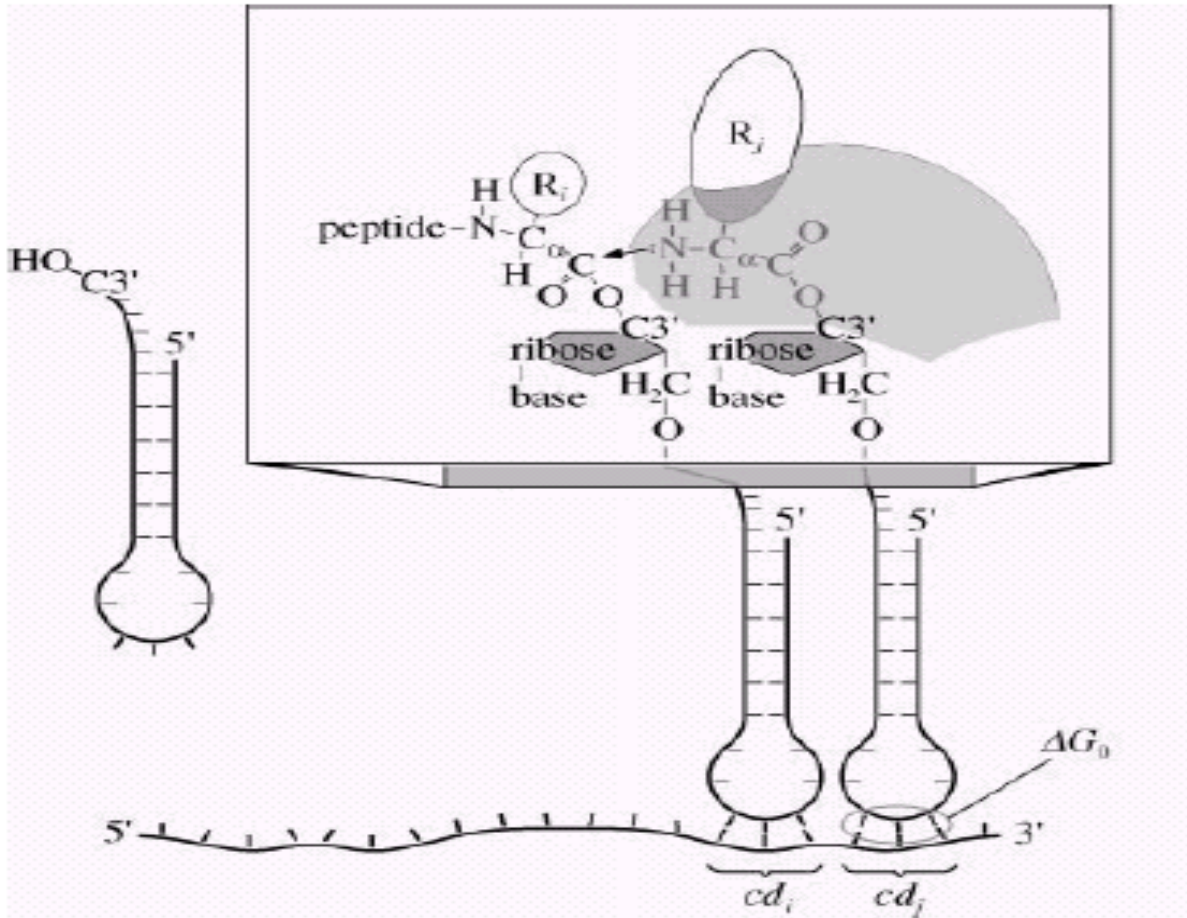
Каждая тРНК имеет уникальную третичную структуру.

Фермент **аминоацил-тРНК-синтетаза** распознает определенную тРНК и специфичен к «своей» аминокислоте. Всего насчитывается (по крайней мере) **20** **аминоацил-тРНК-синтетаз** (по числу аминокислот). Каждый фермент распознает одну аминокислоту и все тРНК, к которым эта аминокислота может быть присоединена.



# Рибосомы

Функция - обеспечение правильного контакта между кодонами на мРНК и антикодонами соответствующих тРНК, а также собственно в синтезе белка из аминокислот



Рибосома эукариот 80 S

↓ ↑

Большая субъединица 60 S

+

Малая субъединица 40 S

- *Lehmann et al., Physico-chemical Constraints Connected with the Coding Properties of the Genetic System J. Theor. Biol. (2000) 202, 129-144*



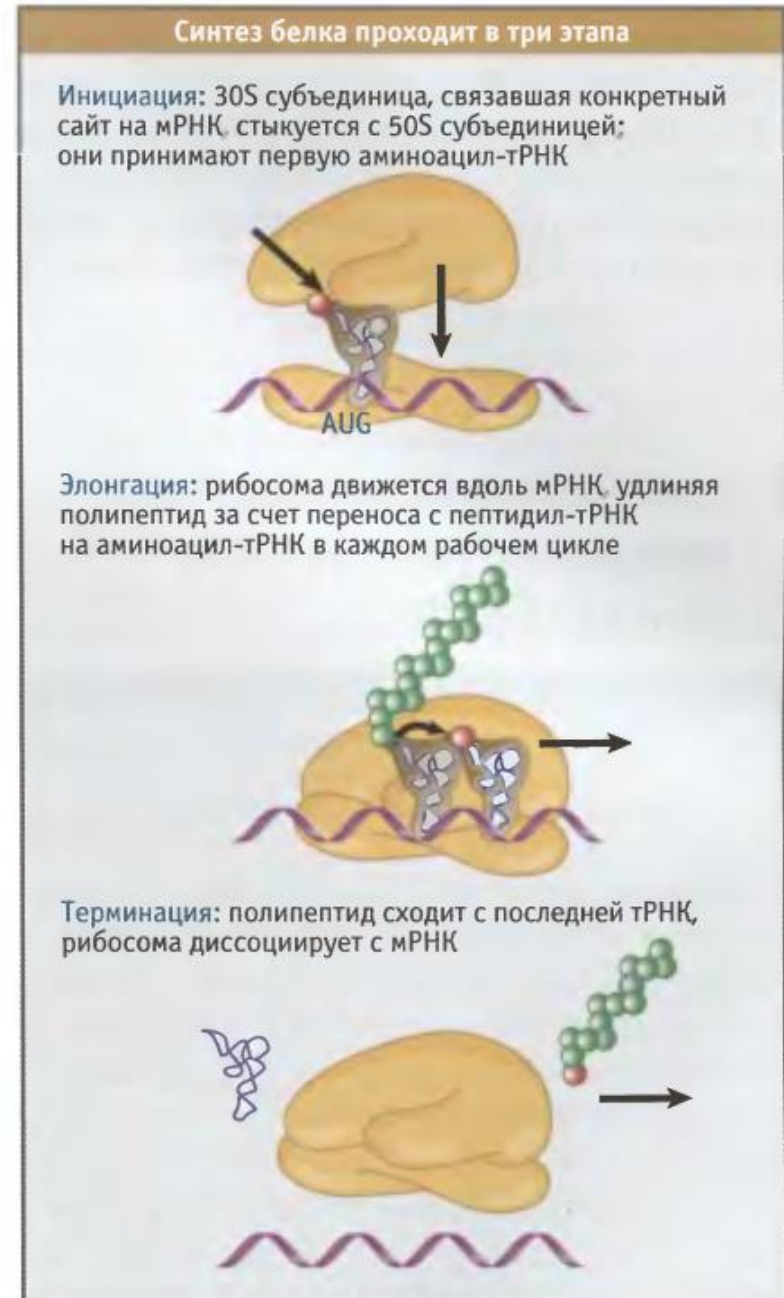
# Синтез белка проходит в три этапа: схематическое представление (прокариоты)

Инициация – система реакций, ведущих к образованию пептидной связи между первыми двумя аминокислотными остатками нового полипептида

Элонгация – все, что происходит с растущим пептидом за время образования всех пептидных связей. На каждом шаге элонгации рибосома, совершая один цикл, добавляет к растущему пептиду одно новое звено

Терминация – все реакции, направленные на освобождение синтезированного пептида и освобождению рибосомы с мРНК

Льюин Б. Гены, 2012 г. С.163

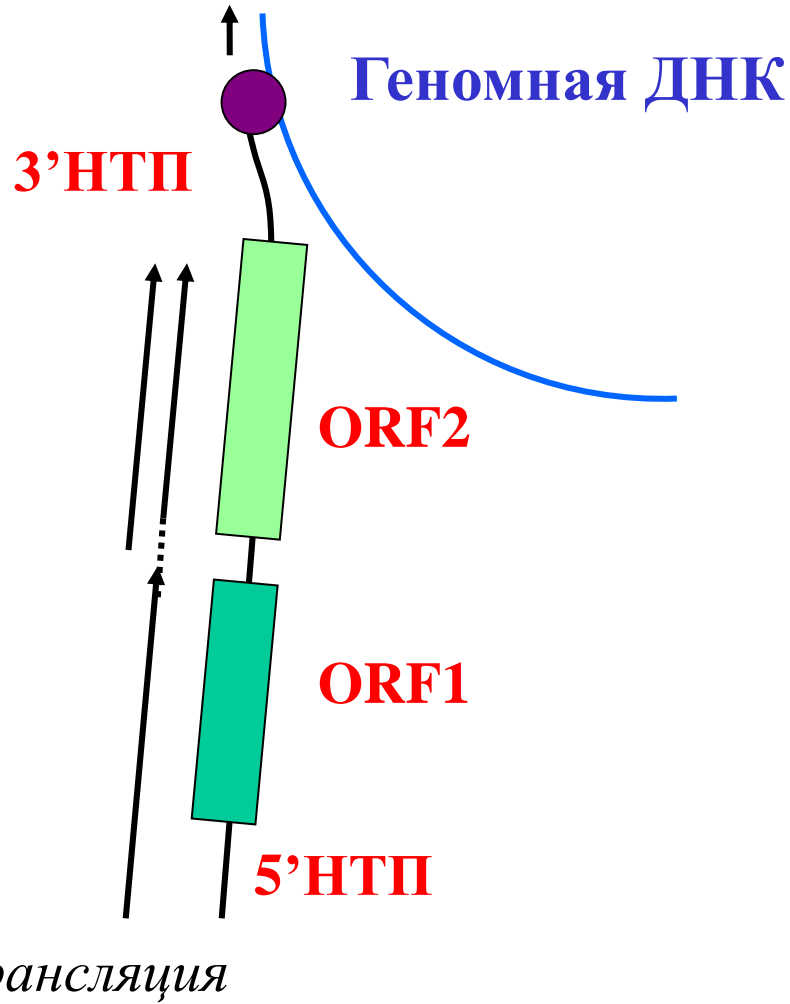




**Структуры мРНК и процесс трансляции у прокариот и эукариот имеют свои отличительные особенности.**

**Функционально наиболее сильно отличается процесс **инициации** трансляции**

# Функциональные различия между мРНК прокариот и эукариот



ядро

Транскрипция, пр-мРНК процессингу

цитоплазма

5'НТП

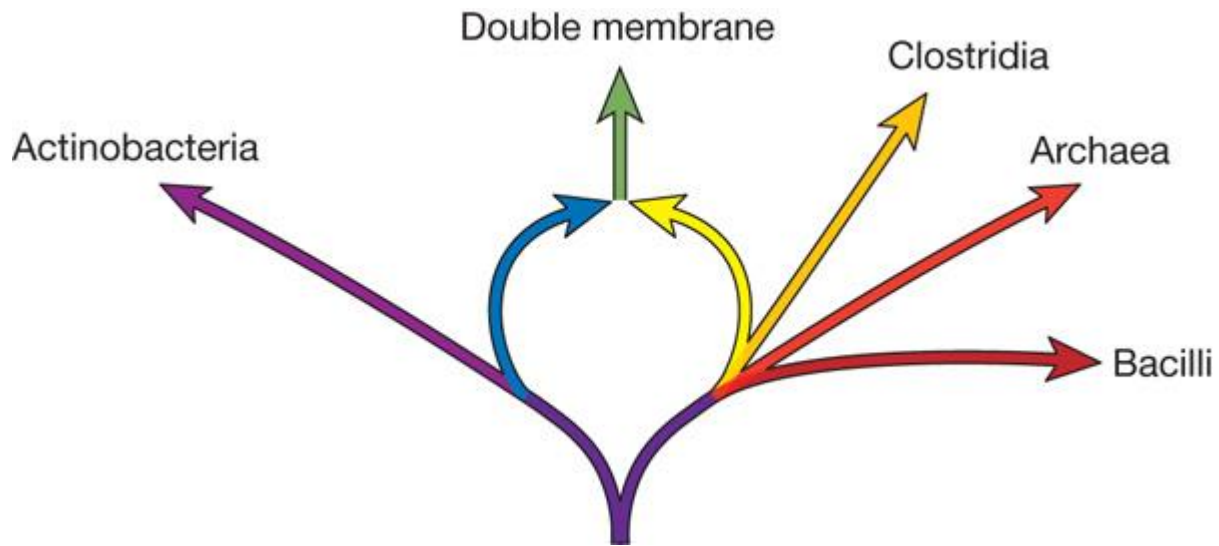
3'НТП р(А)

белок-кодирующая часть (ORF)

трансляция

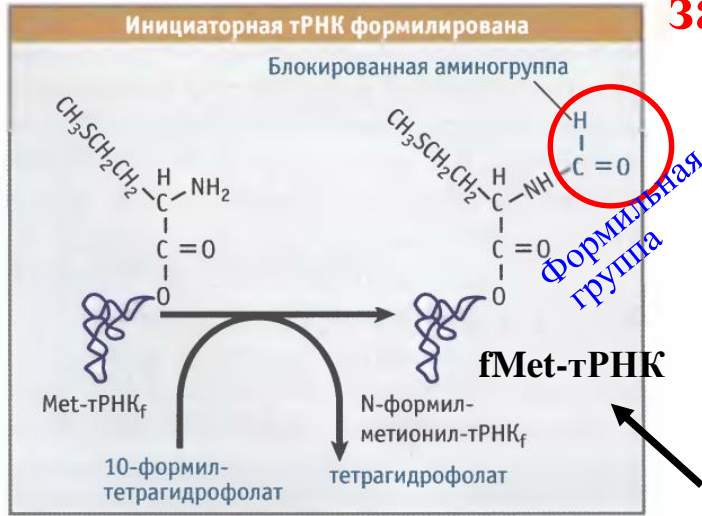


# Прокариоты



# Прокариоты: особая инициаторная тРНК (fMet-тРНК)

закладывает первое звено полипептида



Синтез всех белков начинается с одной и той же аминокислоты – метионина. Сигналом к началу синтеза полипептидной цепи служит иницирующий кодон **AUG**, обозначающий начало открытой рамки считывания. У бактерий, помимо **AUG**, используются также триплеты **GUG** и **UUG**

В инициации и элонгации участвуют разные метиониновые тРНК.

У бактерий, а также в органеллах эукариот метионин инициаторной тРНК имеет формилированную аминокетильную группу. Этот комплекс обозначается как **fMet-тРНК**

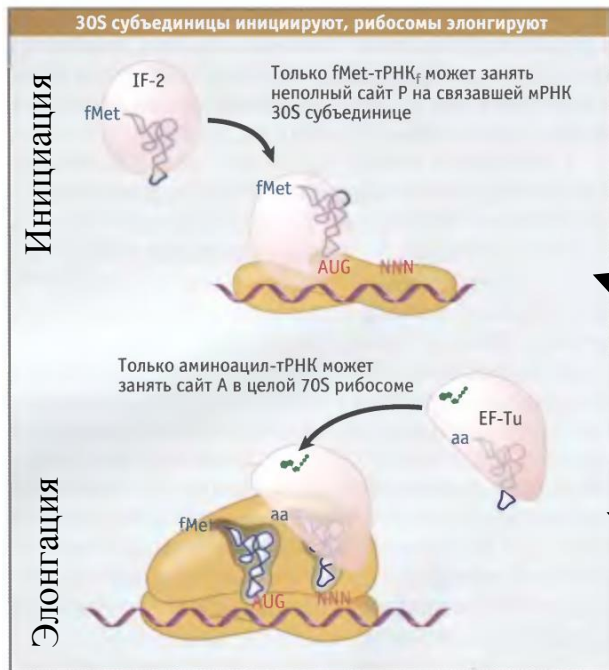
## Прокариоты: факторы инициации

У бактерий есть три фактора инициации: **IF-1**, **IF-2**, **IF-3**. Без этих факторов ни мРНК, ни тРНК, не могут вступить в комплекс инициации.

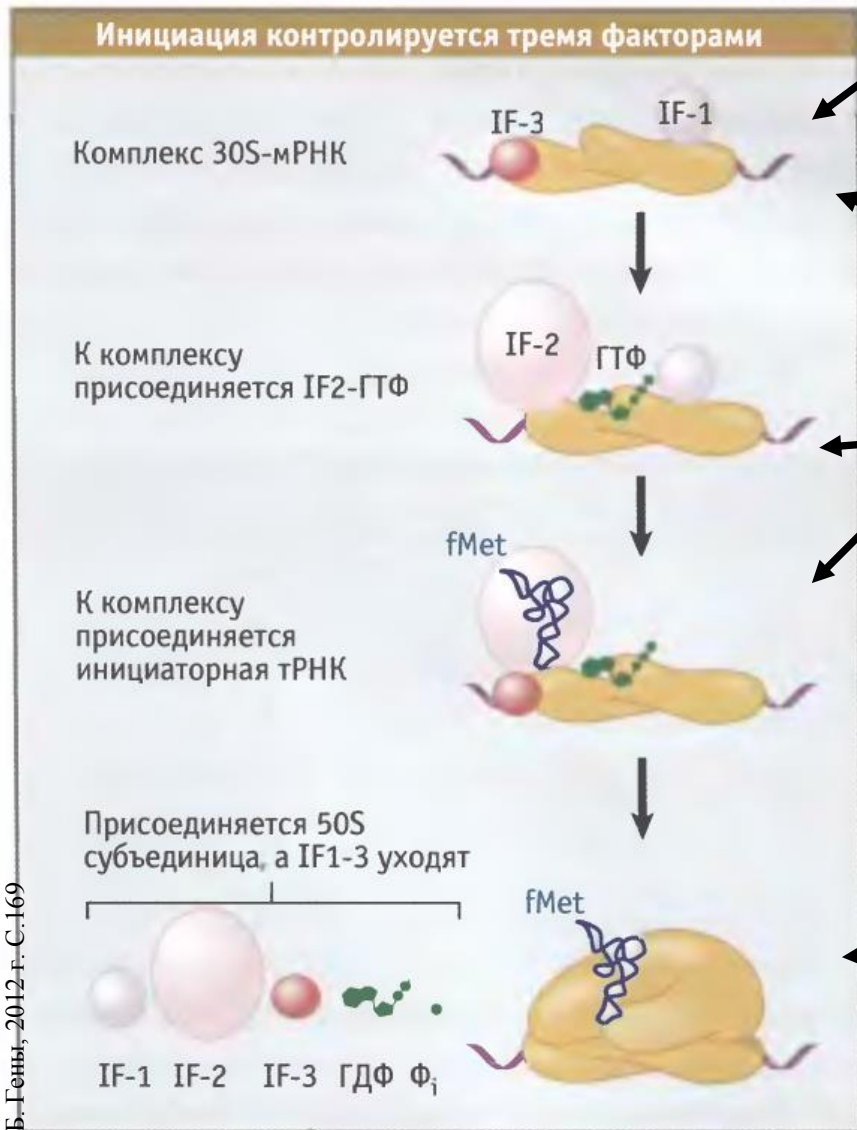
**IF-2** связывает инициаторную fMet-тРНК и позволяет ей занять неполный сайт Р на малой субъединице рибосомы.

**Напротив, в ходе элонгации**

**EF-Tu**, который помещает аминокетильную тРНК в сайт А, совсем не умеет связывать fMet-тРНК, что исключает ее использование.



# Прокариоты: роль факторов инициации (IF-1, IF-2, IF-3)



**IF-3** связывается с поверхностью 30S субъединицы неподалеку от сайта А. Он необходим для связывания 30S субъединицы с мРНК и противодействует ассоциации 30S и 50S субъединиц рибосом.

**IF-1** связывается с 30S субъединицами в уже сформированном комплексе 30S-мРНК. **IF-1** связывается с сайтом А, не позволяя аминокил-тРНК занять этот сайт.

**IF-2** приносит особую инициаторную тРНК (fMet-тРНК) и контролирует ее вхождение в рибосому. **IF-2** обладает ГТФазной активностью (но только в комплексе с рибосомой). ГТФ гидролизуется в тот момент, когда 50S субъединица присоединяется к комплексу инициации с образованием полной рибосомы, что приводит к изменению конформации субъединиц, способствуя их превращению в активную рибосому.

После того, как к комплексу присоединяется 50S субъединица, все факторы инициации уходят, а ГТФ расщепляется.

# Особенности трансляции у прокариот: сайт Шайна-Дальгарно (SD) в комбинации с AUG кодоном определяют старт трансляции

- Прокариотическая полицистронная мРНК

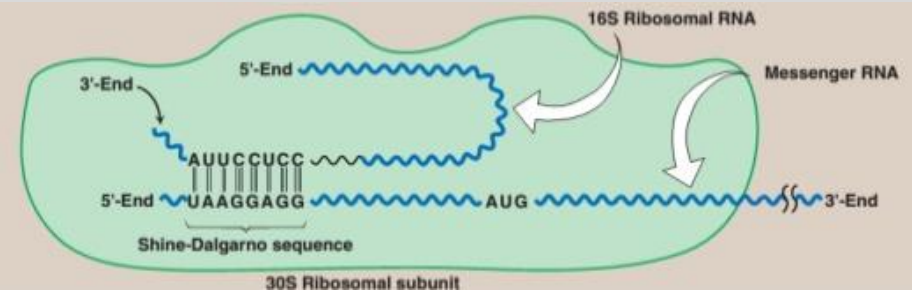
- Ш-Д                      ORF1                      ORF2                      ORF3



- **GGAGGA(N)<sub>8-12</sub>AUG**

- сайт Шайна-Дальгарно (Ш-Д) в комбинации с AUG кодоном определяют старт трансляции, возможна реинициация

У прокариот в составе нуклеотидной последовательности консервативного 3'-конца 16S рибосомной РНК, входящей в состав 30S субъединицы рибосомы, есть комплементарный участок, который и выполняет роль детектора сигнала Шайна-Дальгарно. Именно комбинация AUG кодона и сайта Шайна-Дальгарно обозначает для прокариотической рибосомы начало трансляции. Старт трансляции распознается комплексом 30S субъединицы рибосомы и метиониновой тРНК, антикодон которой является в данном случае детектором кодона AUG.



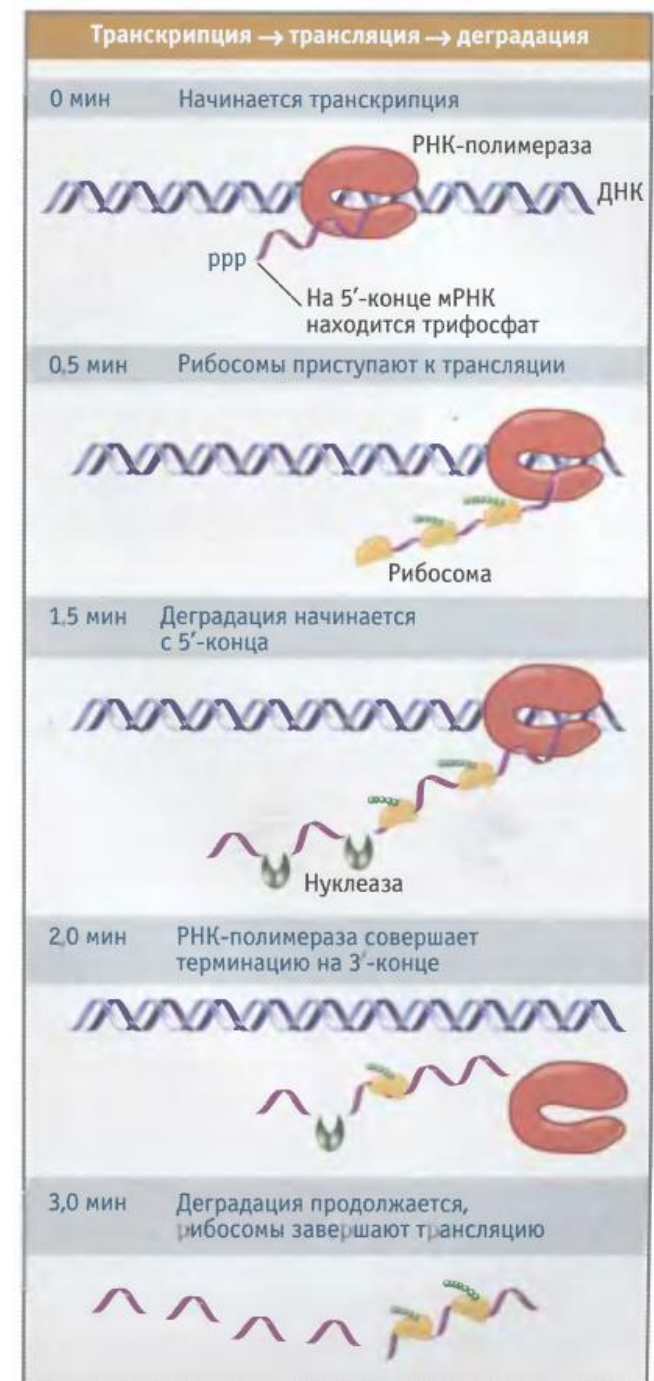
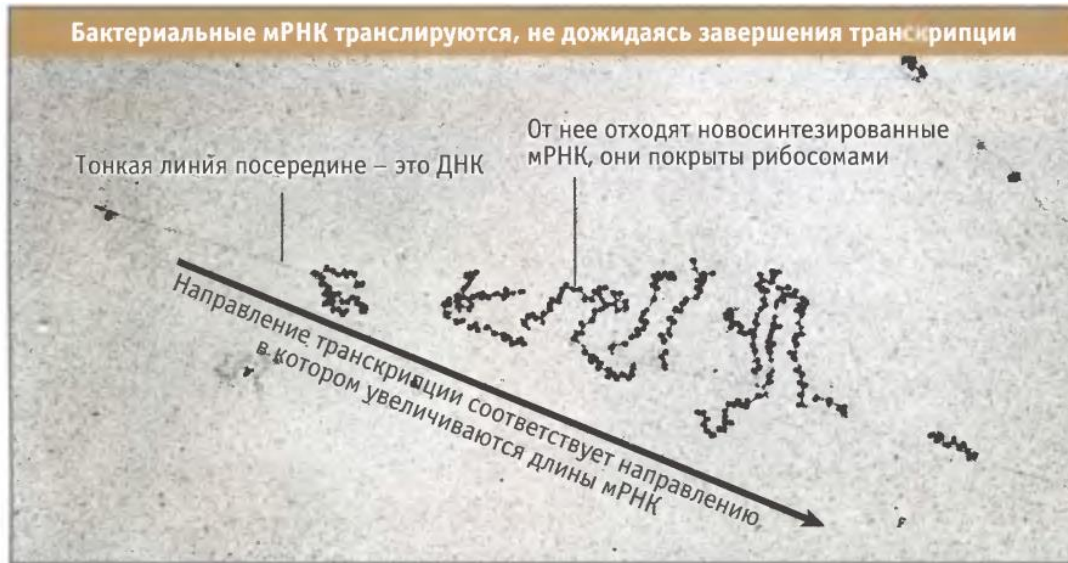
У эукариот нет сайта Ш-Д. У эукариот 40S субъединица рибосомы связывается с 5'-кэпированным концом мРНК, который находится в комплексе с eIF-4 и начинает передвигаться вдоль мРНК до тех пор, пока не встретит AUG кодон.



# Особенности трансляции у прокариот: транскрипция и трансляция сопряжены

Как только появляется мРНК, рибосома прикрепляется к 5'-концу и начинает трансляцию еще до того, как заканчивается синтез оставшейся части РНК. Рибосомы продолжают транслировать мРНК, пока она сохраняет свою целостность.

Дегградация мРНК начинается сразу же после трансляции и, скорее всего, начинается в течение минуты после начала транскрипции. 5-конец мРНК начинает дегградировать еще до того, как 3-конец был синтезирован, либо транслирован. Скорость дегградации примерно в два раза ниже скорости транскрипции или трансляции



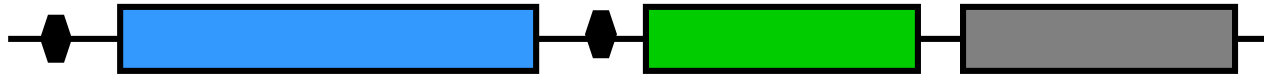
Единицы транскрипции у бактерий можно увидеть

Льюин Б. Гены, 2012 г. С.143-144.

# Особенности трансляции у прокариот: оперонная структура мРНК

- Прокариотическая полицистронная мРНК

- Ш-Д                      ORF1                      Ш-Д                      ORF2                                      ORF3



- в составе многих мРНК находятся две и более открытые рамки считывания, кодирующие аминокислотные последовательности разных белков. Рибосома распознает сайт инициации трансляции в начале матрицы (комбинацию сайта Ш-Д и AUG кодона) и транслирует проксимальную белок-кодирующую последовательность, затем часть рибосом диссоциирует с матрицы, а некоторая часть может реиницировать трансляцию на следующей кодирующей последовательности.

Альтернативно, в межцистронном промежутке может располагаться независимый сайт инициации трансляции – то есть выше стартового кодона AUG второй белок-кодирующей последовательности расположен сайт Шайна-Дальгарно. Тогда часть рибосом будут садиться во внутреннем участке (межцистронном промежутке) и транслировать второй кодирующий участок.

мРНК, в составе которых содержатся несколько белок-кодирующих последовательностей, называются **полицистронными**.

Элемент генома, в составе которого несколько белок-кодирующих последовательностей расположены под транскрипционным контролем одного промотора, называется **опероном**.



# Особенности процесса трансляции у прокариот

1. Особая инициаторная тРНК (fMet-тРНК) закладывает первое звено полипептида
2. Участие факторов инициации (IF-1, IF-2, IF-3)
3. Сайт Шайна-Дальгарно (SD) в комбинации с AUG кодоном определяют старт трансляции
4. Транскрипция и трансляция сопряжены
5. мРНК имеет оперонную структуру

Конец лекции