Лекция №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.

Прошлая лекция № 5

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

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

СЕГОДНЯ, в лекции № 6 будет дана характеристика баз данных (продолжение):

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

CREMOFAC	Database of chromatin remodeling factors	Индия
TcoF- DB	Dragon database of transcription co-factors and transcription factor interacting proteins	Королевство Сау́довская Ара́вия

МАТРИЦЫ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

TRANSFAC Matrix, JASPAR , HOCOMOCO , CIS-BP

РЕГУЛЯТОРНЫЕ РАЙОНЫ, ССТФ, ТРАНСКРИПЦИОННЫЕ ФАКТОРЫ

TRRD	Transcription Regulatory Regions Database	ИЦиГ СО РАН, Новосибирск, Россия

Дополнение:
Интернет-ресурсы по транскрипции у растений:

PlantPAN 2.0
PlantDHS
PPdb
PlantTFDB 3.0

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

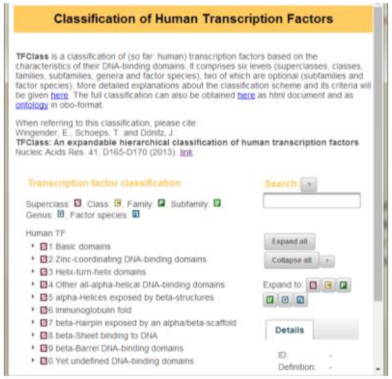
Повторение слайда из предыдущей лекции:

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/)



Classification of Human Transcription Factors

TFClass is a classification of (so far: human) transcription factors based on the characteristics of their DNA-binding domains. It comprises six levels (superclasses, classes, families, subfamilies, genera and factor species), two of which are optional (subfamilies and factor species). More detailed explanations about the classification scheme and its criteria will be given here. The full classification can also be obtained here as html document and as ontology in obo-format.

When referring to this classification, please cite:
Wingender, E., Schoeps, T. and Dönitz, J.:
TFClass: An expandable hierarchical classification of human transcription factors
Nucleic Acids Res. 41, D165-D170 (2013). link

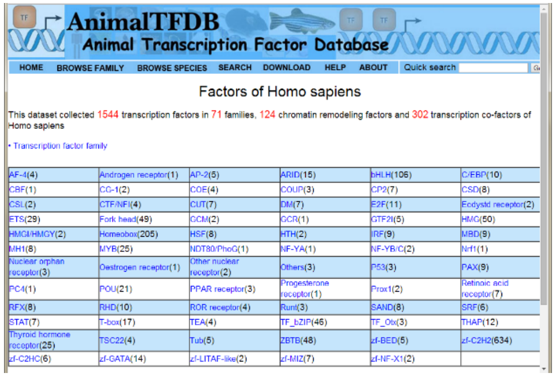
Transcription factor classification

Superclass: S, Class: C, Family: F, Subfamily: F,
Genus: G, Factor species: F

Human TF
- 1 Basic domains
- 2 Zinc-coordinating DNA-binding domains
- 3 Helix-turn-helix domains
- 4 Other all-alpha-helical DNA-binding domains
- 5 alpha-Helices exposed by beta-structures
- 6 Immunoglobulin fold
- 7 beta-Hairpin exposed by an alpha/beta-scaffold
- 8 beta-Sheet binding to DNA
- 9 beta-Barrel DNA-binding domains
- 0 Yet undefined DNA-binding domains

Ресурс AnimalTFDB

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



Factors of Homo sapiens

This dataset collected 1544 transcription factors in 71 families, 124 chromatin remodeling factors and 302 transcription co-factors of Homo sapiens

- Transcription factor family

AF-4(4)	Androgen receptor(1)	AP-2(5)	ARID(15)	bHLH(106)	C/EBP(10)
CBF(1)	CG-1(2)	COE(4)	COUP(3)	CP2(7)	CSD(8)
CSL(2)	CTF/NFI(4)	CUT(7)	DM(7)	E2F(11)	Ecdystd receptor(2)
ETS(29)	Fork head(49)	GCM(2)	GCR(1)	GTF2I(5)	HMG(50)
HMGI/HMGY(2)	Homeobox(205)	HSF(8)	HTH(2)	IRF(9)	MBD(9)
MH1(8)	MYB(25)	NDT80/PhoG(1)	NF-YA(1)	NF-YB/C(2)	Nrf1(1)
Nuclear orphan receptor(3)	Oestrogen receptor(1)	Other nuclear receptor(2)	Others(3)	P53(3)	PAX(9)
PC4(1)	POU(21)	PPAR receptor(3)	Progesterone receptor(1)	Prox1(2)	Retinoic acid receptor(7)
RFX(8)	RHD(10)	ROR receptor(4)	Runt(3)	SAND(8)	SRF(6)
STAT(7)	T-box(17)	TEA(4)	TF_bZIP(46)	TF_Otx(3)	THAP(12)
Thyroid hormone receptor(25)	TSC22(4)	Tub(5)	ZBTB(48)	zf-BED(5)	zf-C2H2(634)
zf-C2HC(6)	zf-GATA(14)	zf-LITAF-like(2)	zf-MIZ(7)	zf-NF-X1(2)	

1558 генов,
кодирующих 2904 изоформ белков,
содержащих ДНК-связывающий домен.

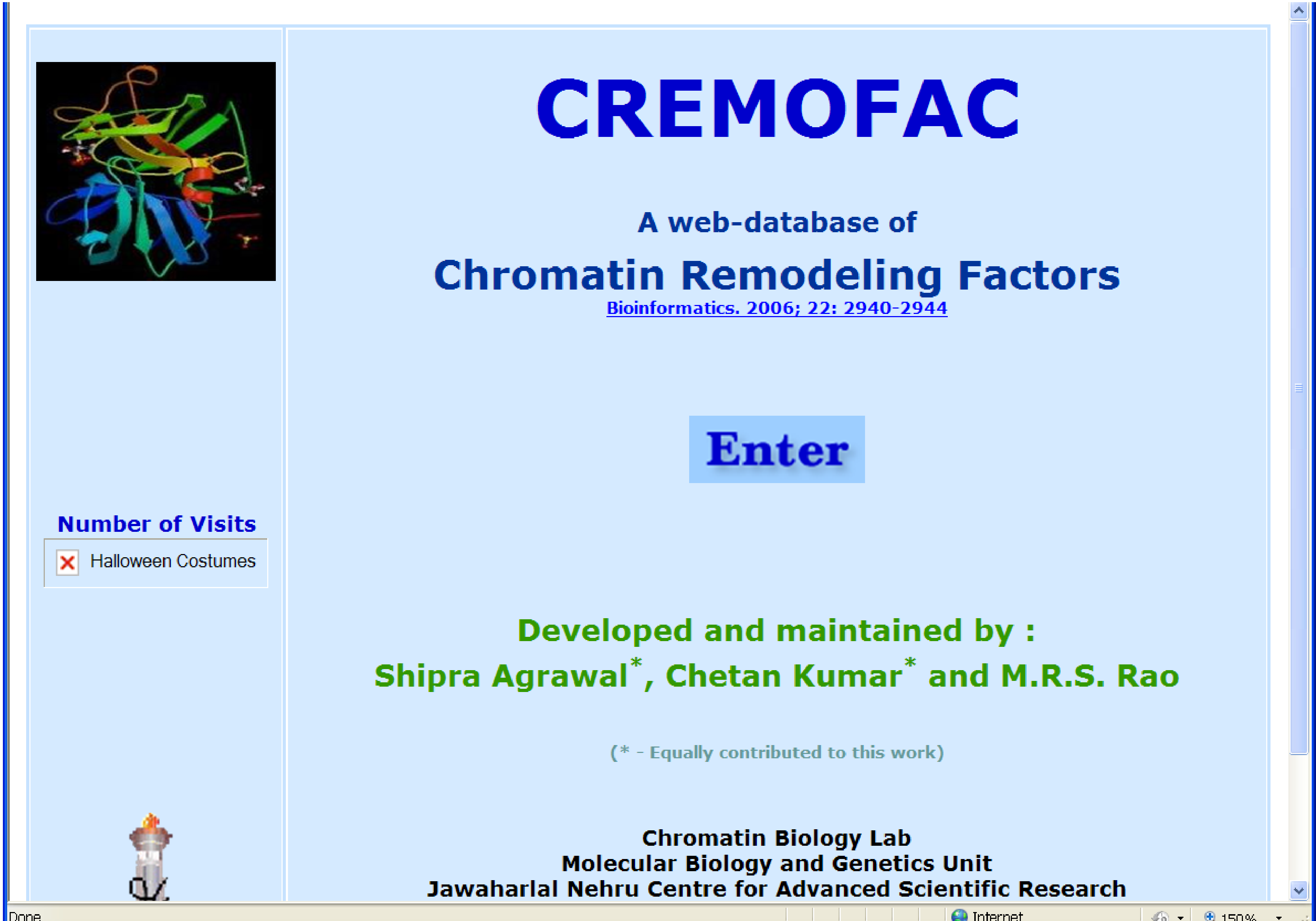
1544 гена,
кодирующих белки,
содержащие ДНК-связывающий домен.

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

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

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

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

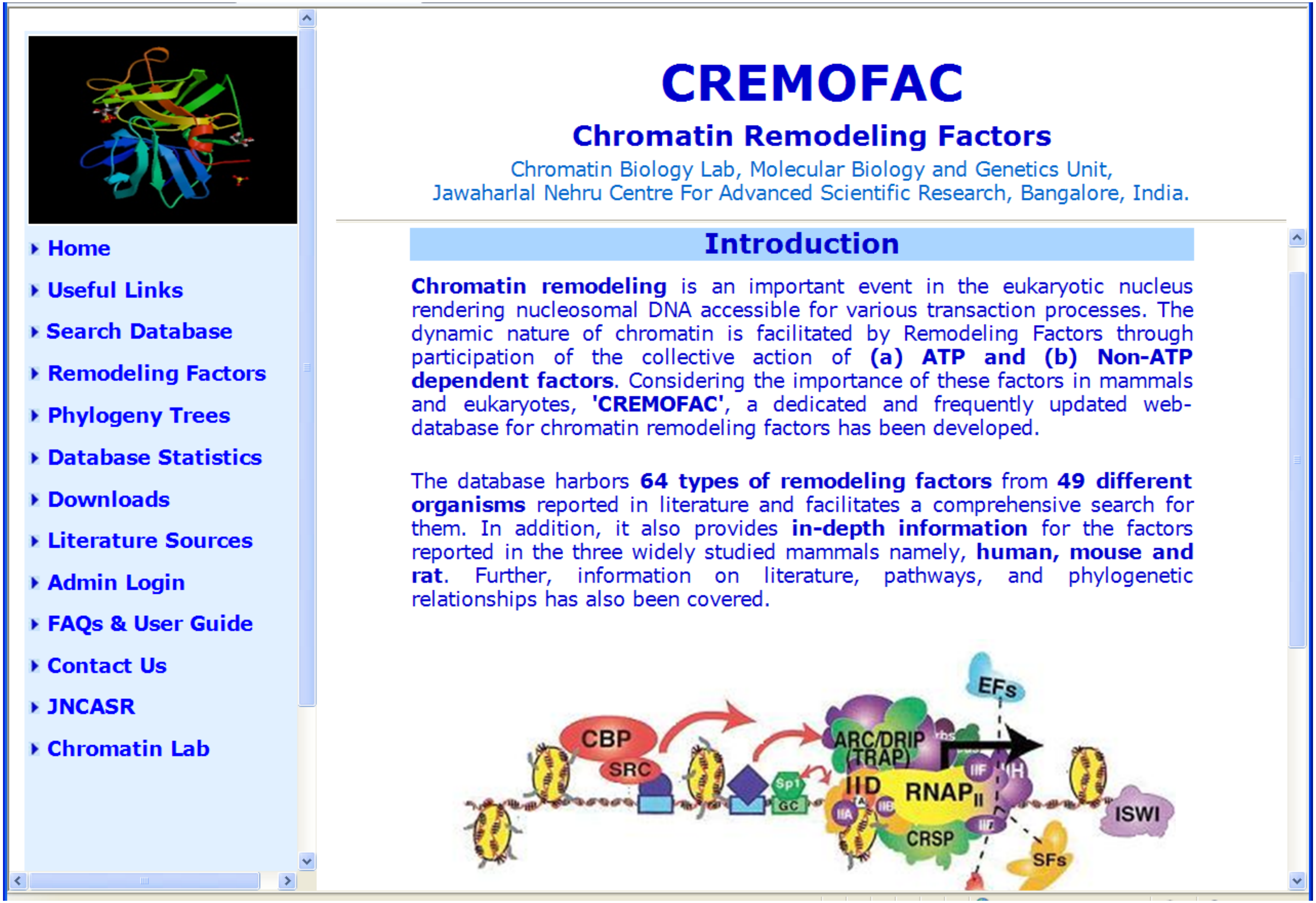
Developed and maintained by :

Shipra Agrawal*, Chetan Kumar* 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

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ



CREMOFAC

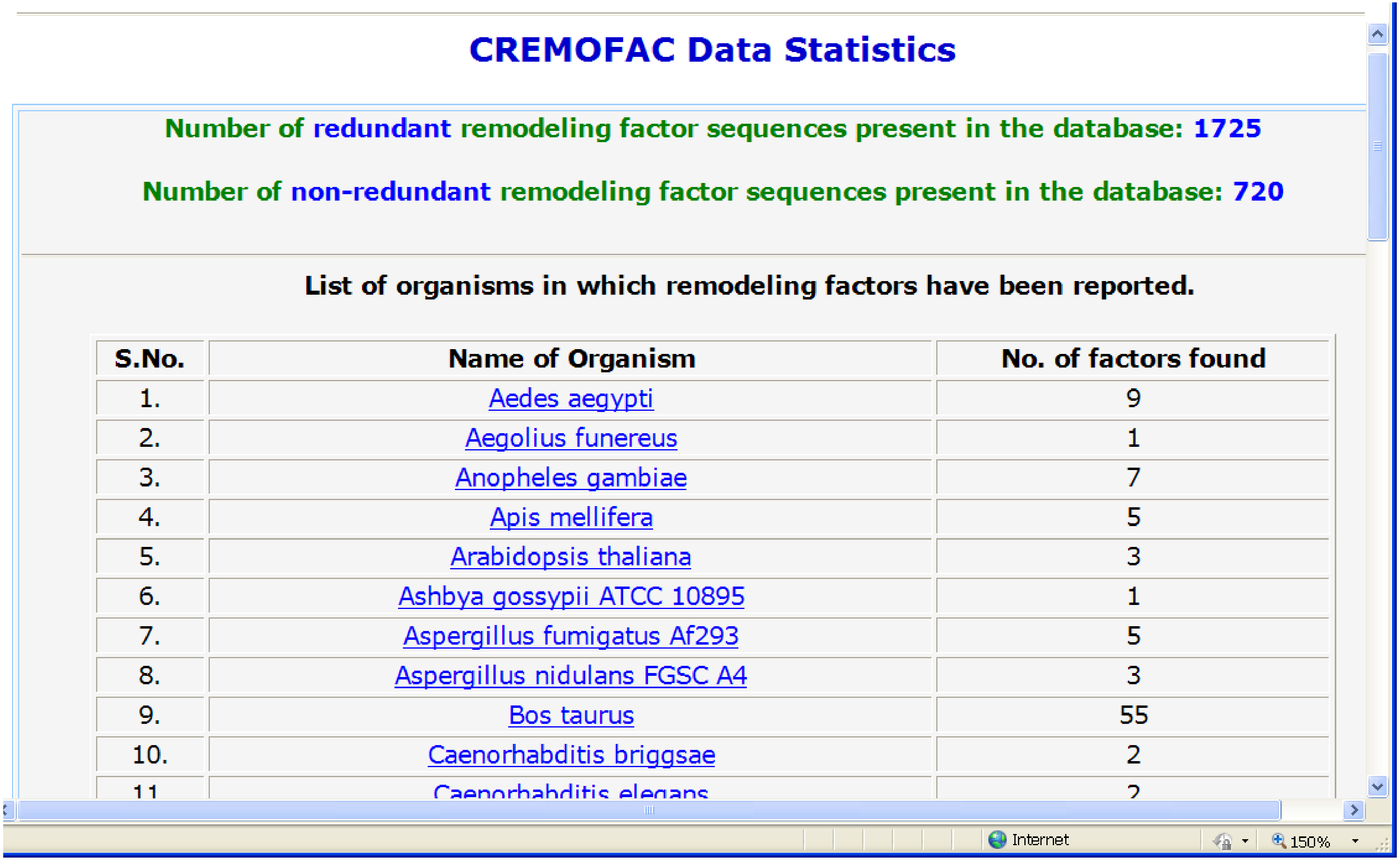
Chromatin Remodeling Factors

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.



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

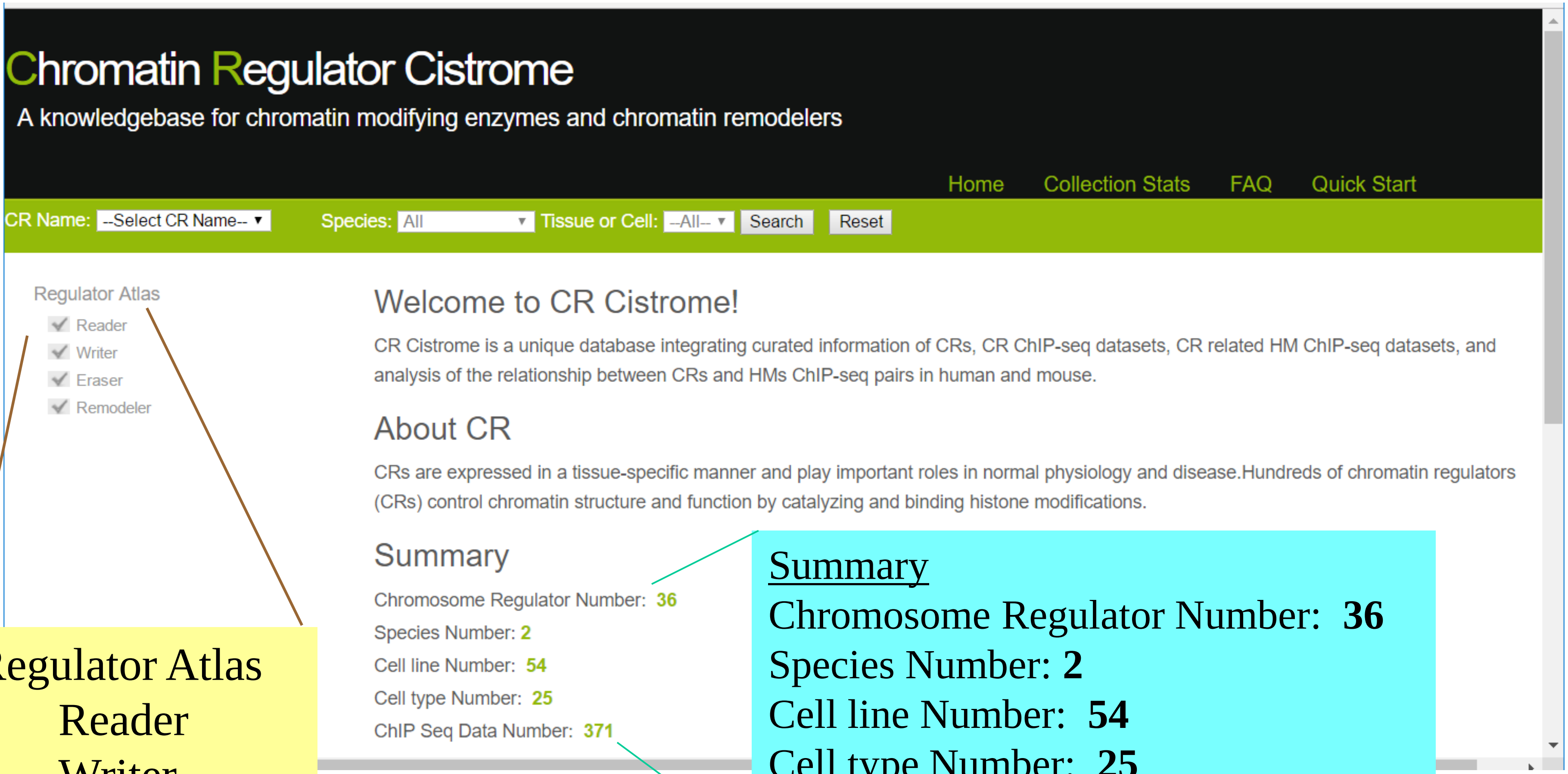
List of organisms in which remodeling factors have been reported.

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

Всего 49 видов организмов

Ресурс CR Cistrome

http://cistrome.org/cr/



Chromatin Regulator Cistrome

A knowledgebase for chromatin modifying enzymes and chromatin remodelers

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

Regulator Atlas

- Reader
- Writer
- Eraser
- Remodeler

<u>Summary</u>

Chromosome Regulator Number: **36**

Species Number: **2**

Cell line Number: **54**

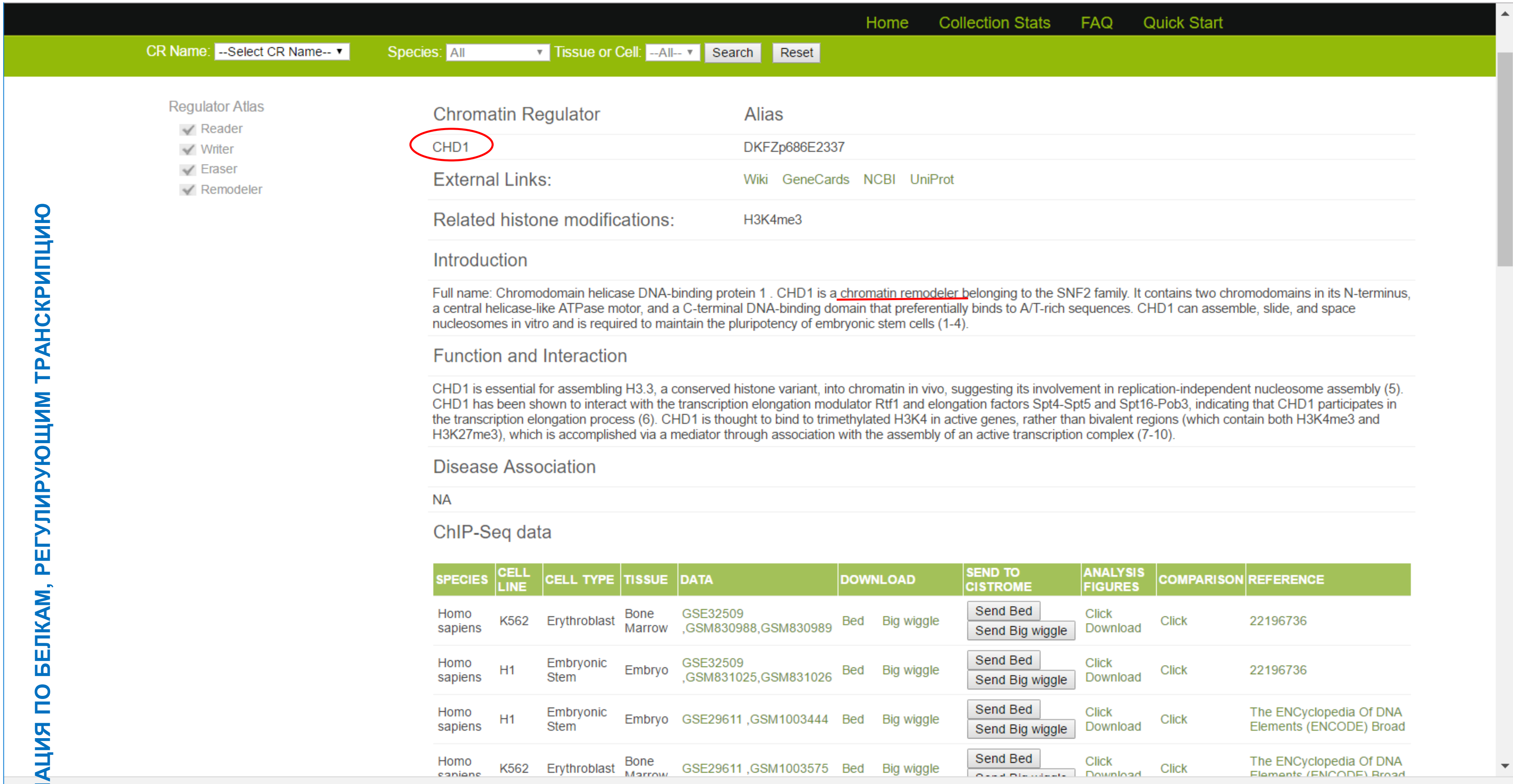
Cell type Number: **25**

ChIP Seq Data Number: **371**

Wang Q. et al., CR Cistrome: a ChIP-Seq database for chromatin regulators and histone modification linkages in human and mouse. Nucleic Acids Res. 2014 Jan;42(Database issue):D450-8. doi: 10.1093/nar/gkt1151. Epub 2013 Nov 18.

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

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ



Chromatin Regulator	Alias
CHD1	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	K562	Erythroblast	Bone	GSE29611 ,GSM1003575	Bed Big wiggle	Send Bed	Click	Click	The ENCyclopedia Of DNA

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)

64 гена

99 генов

23 гена

167 генов, кодирующих белки с хроматин-модифицирующей активностью

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

Матрицы сайтов связывания транскрипционных факторов

JASPAR
http://jaspar.binf.ku.dk/


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:

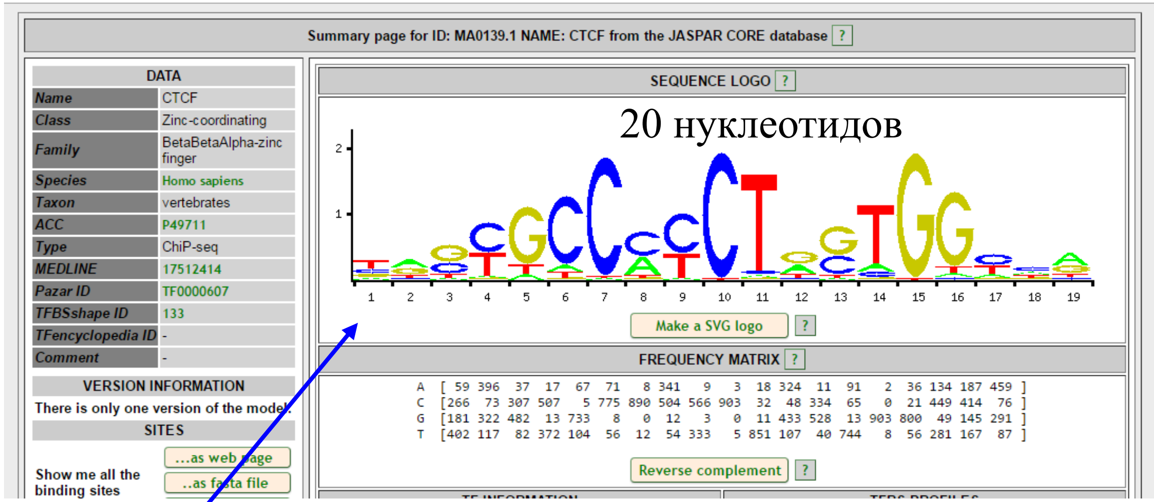
Browse/search a JASPAR database

BRCA1 Quick Search

Select a JASPAR database ?

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 acess, non-redundancy and quality.
When should it be used? When seeking models for specific factors or structural classes, or if experimental evidence is paramount.

Пример: две матрицы сайтов связывания CTCF


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	TF0000607
TFBSshape ID	133
TFencyclopedia ID	-
Comment	-

VERSION INFORMATION

There is only one version of the model.

SITES

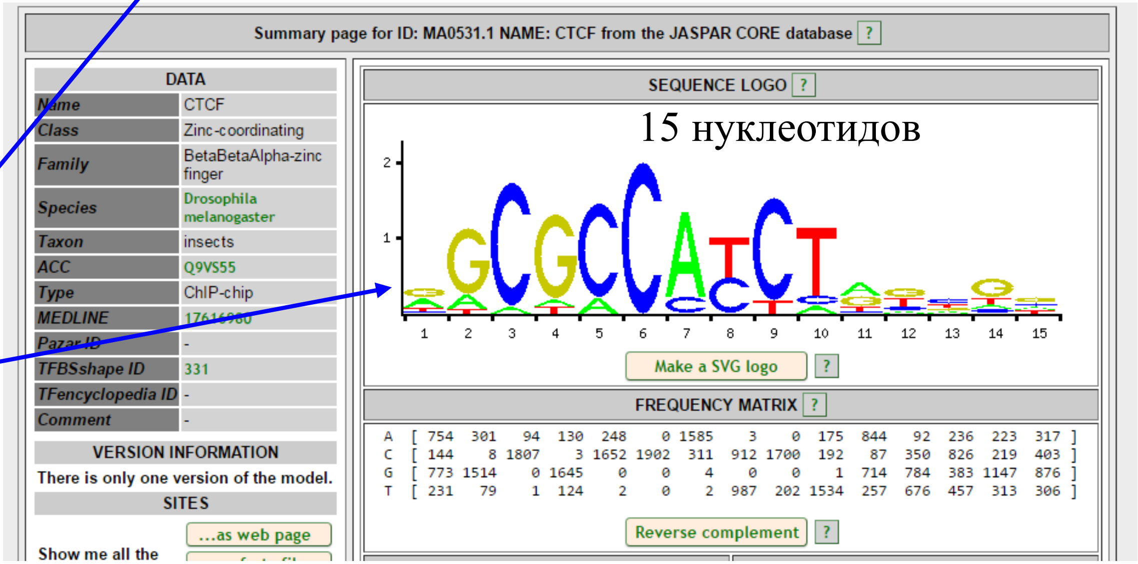
Show me all the binding sites ...as web page ..as fasta file

SEQUENCE LOGO ?

20 нуклеотидов

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 ]
```


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

Show me all the ...as web page

SEQUENCE LOGO ?

15 нуклеотидов

FREQUENCY MATRIX ?

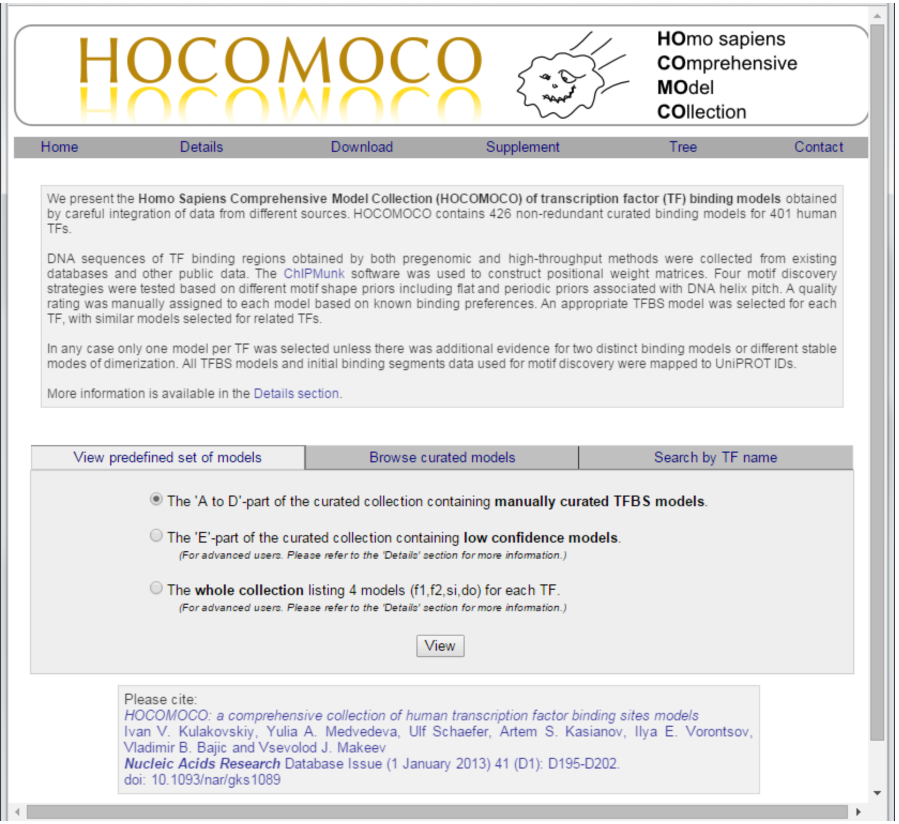
```
A  [ 754  301   94  130  248    0 1585    3    0  175  844   92  236  223  317 ]
C  [ 144    8 1807    3 1652 1902  311  912 1700  192   87  350  826  219  403 ]
G  [ 773 1514    0 1645    0    0    4    0    0    1  714  784  383 1147  876 ]
T  [ 231   79    1  124    2    0    2  987  202 1534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

Графики WEB-logo

Матрицы сайтов связывания транскрипционных факторов

HOCOMOCO

http://autosome.ru/HOCOMOCO/



HOmo sapiens
COmprehensive
MOdel
COllection

We present the **Homo Sapiens Comprehensive Model Collection (HOCOMOCO) of transcription factor (TF) binding models** obtained by careful integration of data from different sources. HOCOMOCO contains 426 non-redundant curated binding models for 401 human TFs.

DNA sequences of TF binding regions obtained by both pregenomic and high-throughput methods were collected from existing databases and other public data. The ChIPMunk software was used to construct positional weight matrices. Four motif discovery strategies were tested based on different motif shape priors including flat and periodic priors associated with DNA helix pitch. A quality rating was manually assigned to each model based on known binding preferences. An appropriate TFBS model was selected for each TF, with similar models selected for related TFs.

In any case only one model per TF was selected unless there was additional evidence for two distinct binding models or different stable modes of dimerization. All TFBS models and initial binding segments data used for motif discovery were mapped to UniPROT IDs.

More information is available in the Details section.

● The 'A to D'-part of the curated collection containing **manually curated TFBS models**.

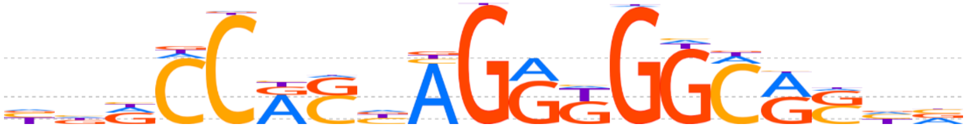
○ The 'E'-part of the curated collection containing **low confidence models**.
(For advanced users. Please refer to the 'Details' section for more information.)

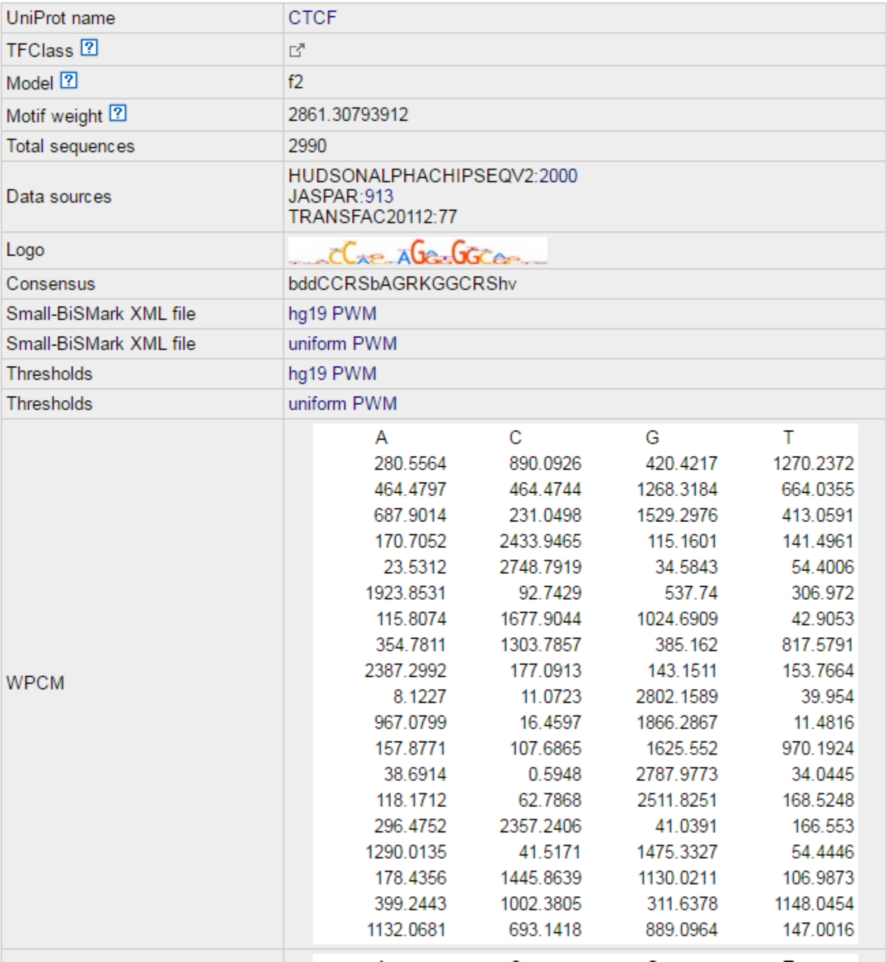
○ The **whole collection** listing 4 models (f1,f2,si,do) for each TF.
(For advanced users. Please refer to the 'Details' section for more information.)

Please cite:
HOCOMOCO: a comprehensive collection of human transcription factor binding sites models
Ivan V. Kulakovskiy, Yulia A. Medvedeva, Ulf Schaefer, Artem S. Kasianov, Ilya E. Vorontsov, Vladimir B. Bajic and Vsevolod J. Makeev
Nucleic Acids Research Database Issue (1 January 2013) 41 (D1): D195-D202.
doi: 10.1093/nar/gks1089

Пример: матрица сайтов связывания CTCF

15 нуклеотидов



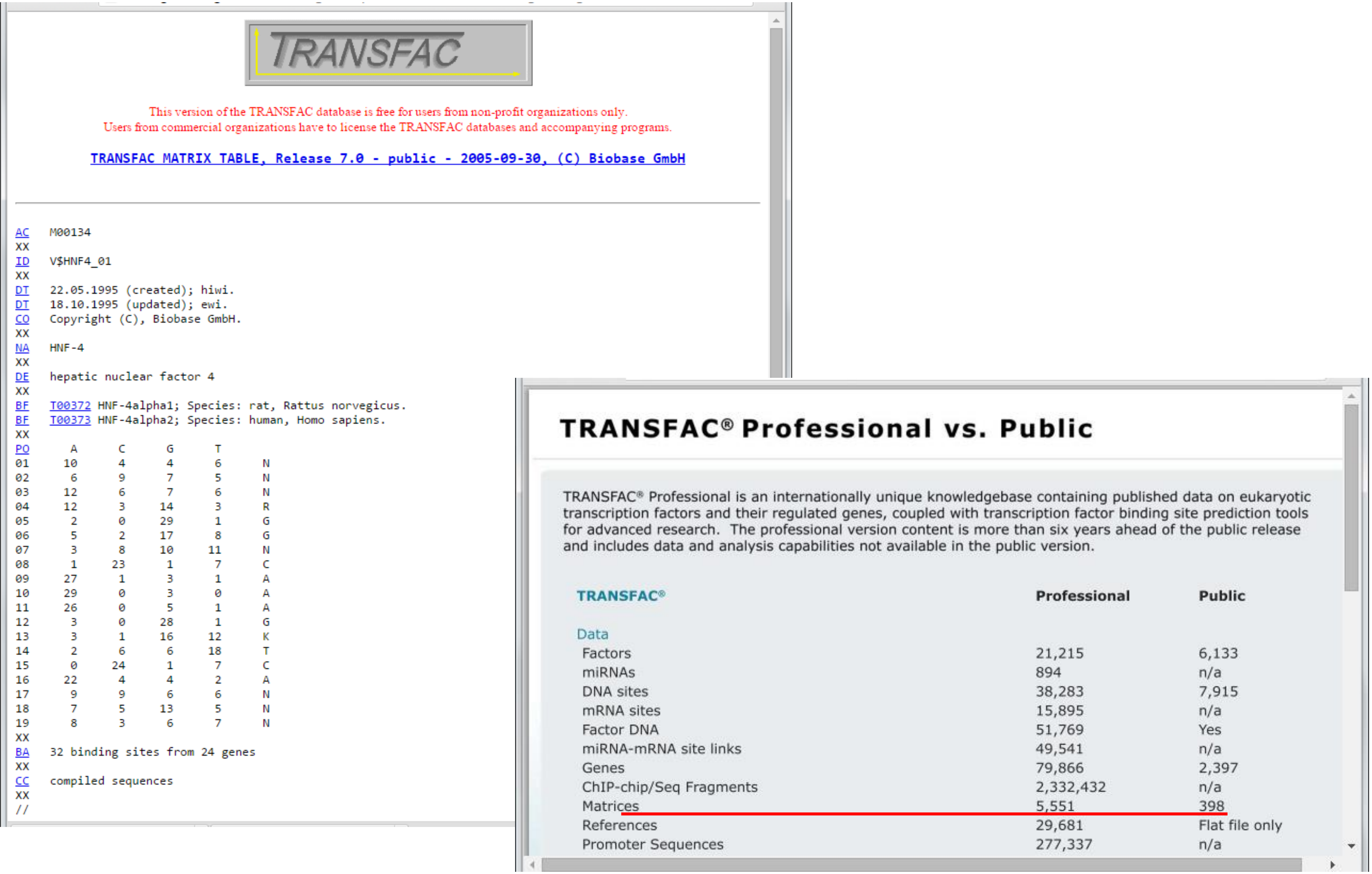


UniProt name	CTCF
TFClass ?	
Model ?	f2
Motif weight ?	2861.30793912
Total sequences	2990
Data sources	HUDSONALPHACHIPSEQV2:2000 JASPAR:913 TRANSFAC20112:77
Logo	
Consensus	bddCCRSbAGRKGGCRShv
Small-BiSMark XML file	hg19 PWM
Small-BiSMark XML file	uniform PWM
Thresholds	hg19 PWM
Thresholds	uniform PWM
WPCM	

A	C	G	T
280.5564	890.0926	420.4217	1270.2372
464.4797	464.4744	1268.3184	664.0355
687.9014	231.0498	1529.2976	413.0591
170.7052	2433.9465	115.1601	141.4961
23.5312	2748.7919	34.5843	54.4006
1923.8531	92.7429	537.74	306.972
115.8074	1677.9044	1024.6909	42.9053
354.7811	1303.7857	385.162	817.5791
2387.2992	177.0913	143.1511	153.7664
8.1227	11.0723	2802.1589	39.954
967.0799	16.4597	1866.2867	11.4816
157.8771	107.6865	1625.552	970.1924
38.6914	0.5948	2787.9773	34.0445
118.1712	62.7868	2511.8251	168.5248
296.4752	2357.2406	41.0391	166.553
1290.0135	41.5171	1475.3327	54.4446
178.4356	1445.8639	1130.0211	106.9873
399.2443	1002.3805	311.6378	1148.0454
1132.0681	693.1418	889.0964	147.0016

HOmo sapiens COmprehensive MOdel COllection (HOCOMOCO) v10 provides transcription factor (TF) binding models for **600 human** and **395 mouse** TFs.

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

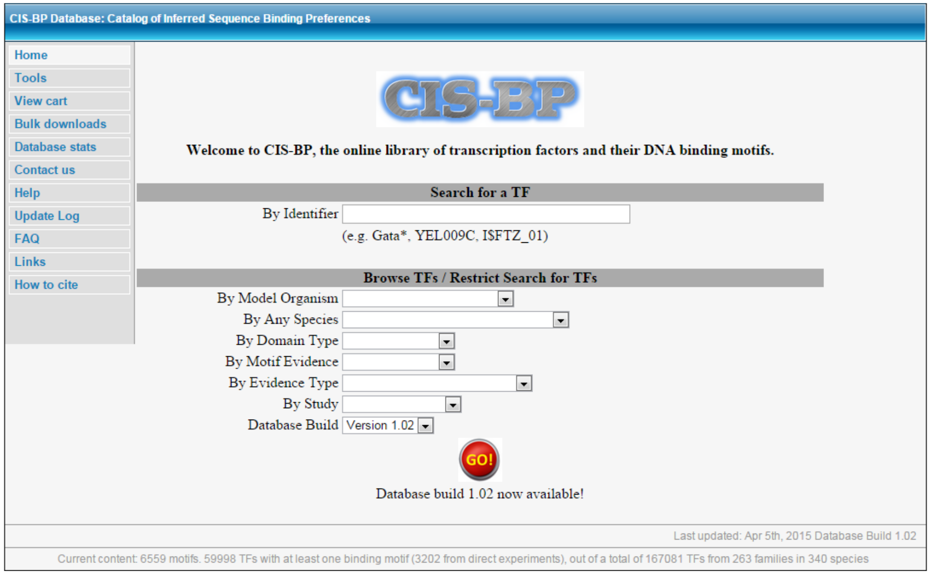
```
AC  M00134
XX
ID  V$HNF4_01
XX
DT  22.05.1995 (created); hiwi.
DT  18.10.1995 (updated); ewi.
CO  Copyright (C), Biobase GmbH.
XX
NA  HNF-4
XX
DE  hepatic nuclear factor 4
XX
BF  T00372 HNF-4alpha1; Species: rat, Rattus norvegicus.
BF  T00373 HNF-4alpha2; Species: human, Homo sapiens.
XX
PO      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
Data		
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

База CIS-BP (http://cisbp.ccbr.utoronto.ca/index.php) - компиляция данных по матрицам из других баз



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, I$FTZ_01)

Browse TFs / Restrict Search for TFs

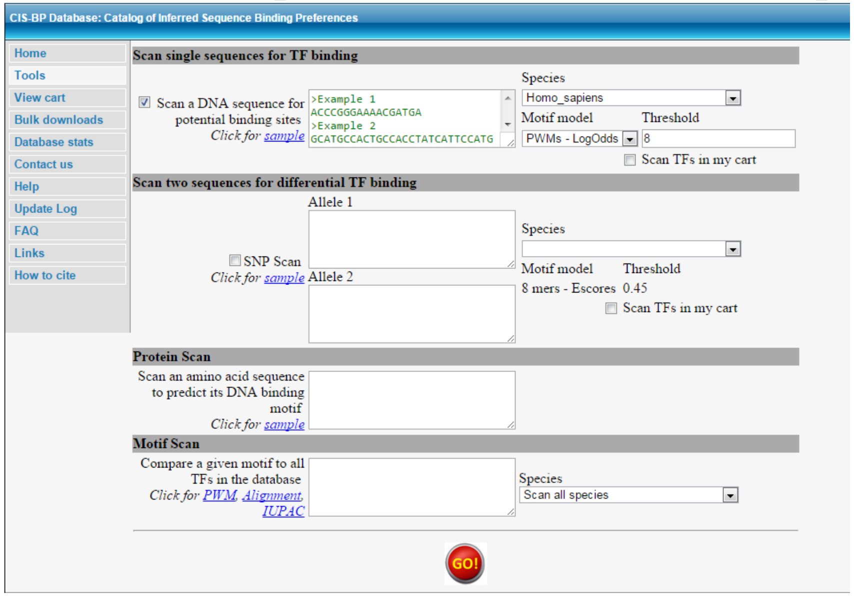
By Model Organism
By Any Species
By Domain Type
By Motif Evidence
By Evidence Type
By Study
Database Build: Version 1.02

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

Возможность предсказания сайтов по матрицам



Scan single sequences for TF binding

☑ Scan a DNA sequence for potential binding sites
Click for sample

```
>Example 1
ACCCGGGAAAACGATGA
>Example 2
GCATGCCACTGCCACCTATCATTCCATG
```

Species: Homo_sapiens
Motif model: PWMs - LogOdds
Threshold: 8
☐ Scan TFs in my cart

Scan two sequences for differential TF binding

Allele 1
☐ SNP Scan
Click for sample
Allele 2
Species
Motif model: 8 mers - Escores
Threshold: 0.45
☐ Scan TFs in my cart

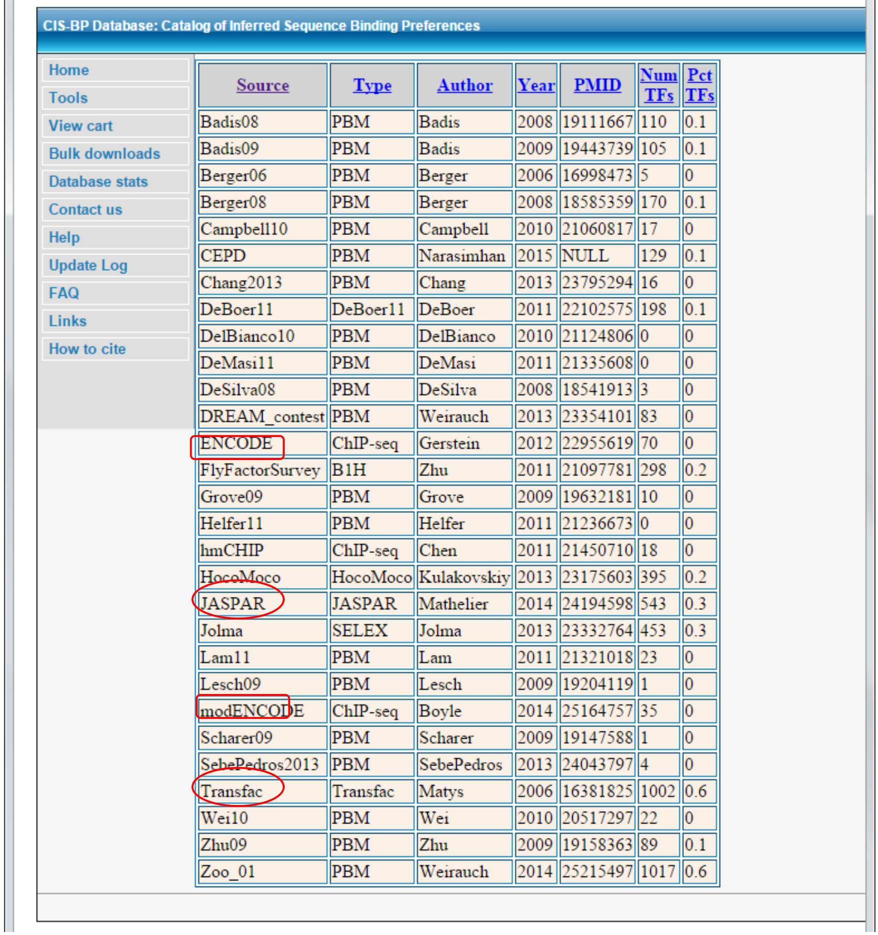
Protein Scan

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

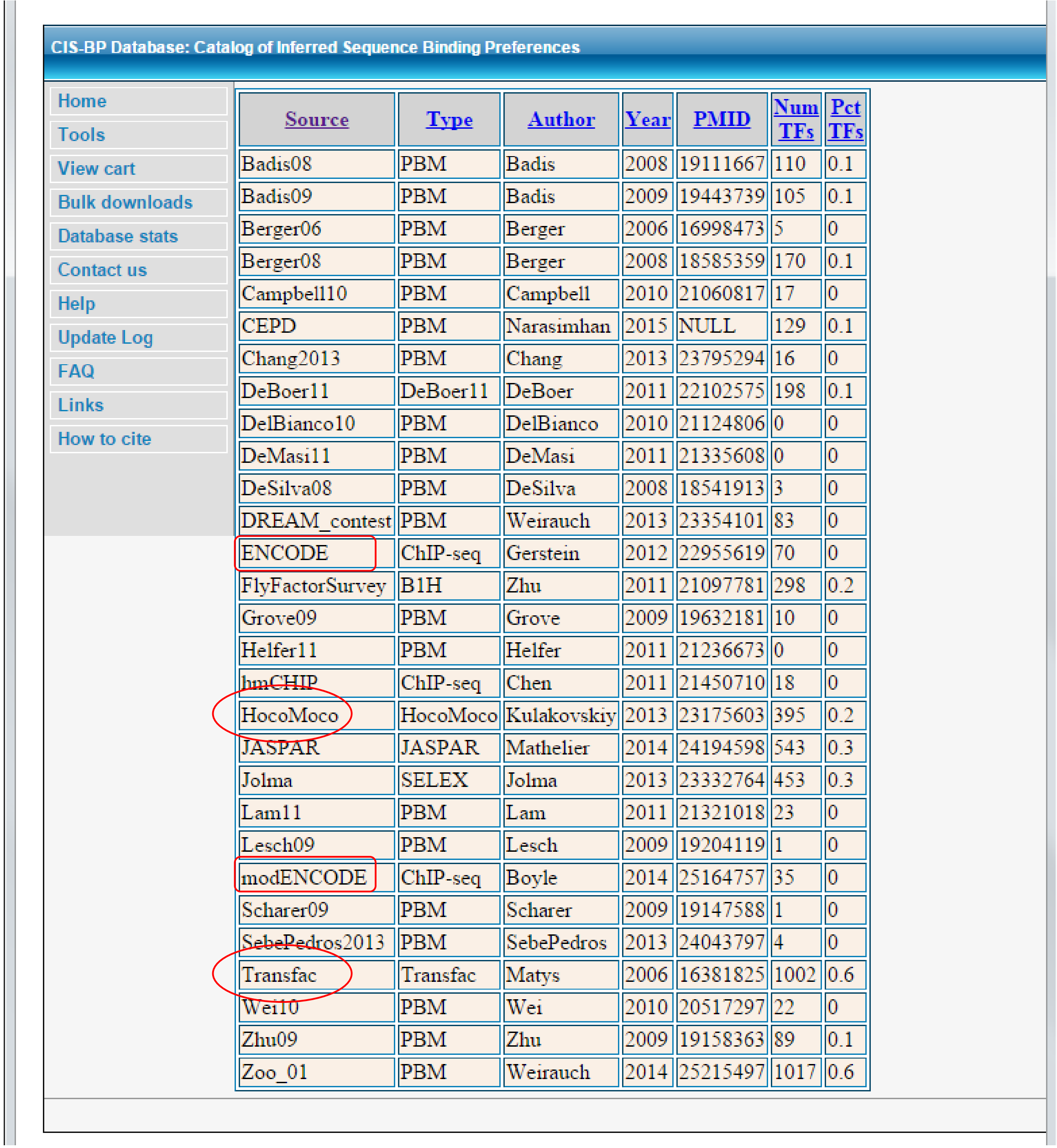
Статистика



Source	Type	Author	Year	PMID	Num TFs	Pct TFs
Badis08	PBM	Badis	2008	19111667	110	0.1
Badis09	PBM	Badis	2009	19443739	105	0.1
Berger06	PBM	Berger	2006	16998473	5	0
Berger08	PBM	Berger	2008	18585359	170	0.1
Campbell10	PBM	Campbell	2010	21060817	17	0
CEPD	PBM	Narasimhan	2015	NULL	129	0.1
Chang2013	PBM	Chang	2013	23795294	16	0
DeBoer11	DeBoer11	DeBoer	2011	22102575	198	0.1
DelBianco10	PBM	DelBianco	2010	21124806	0	0
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	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.

Информационное содержание базы CIS-BP



CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

Source	Type	Author	Year	PMID	Num TFs	Pct TFs
Badis08	PBM	Badis	2008	19111667	110	0.1
Badis09	PBM	Badis	2009	19443739	105	0.1
Berger06	PBM	Berger	2006	16998473	5	0
Berger08	PBM	Berger	2008	18585359	170	0.1
Campbell10	PBM	Campbell	2010	21060817	17	0
CEPD	PBM	Narasimhan	2015	NULL	129	0.1
Chang2013	PBM	Chang	2013	23795294	16	0
DeBoer11	DeBoer11	DeBoer	2011	22102575	198	0.1
DelBianco10	PBM	DelBianco	2010	21124806	0	0
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	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, Lavrenti
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, Lavrentieva 10,
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 publication. TRRD release 6.0 comprises the information on 1167 genes, 5537 transcription factor binding sites, 1714 regulatory regions, 14 locus control regions and 5335 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 (6.0) comprises seven databases linked with cross-references: 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 (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 informational fields RegUnitLink (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 site; 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 3832 333468; Fax: +7 3832 331278; Email: kol@bionet.nsc.ru

ГЛАВНАЯ СТРАНИЦА ДЛЯ ВХОДА В TRRD ЧЕРЕЗ ИНТЕРНЕТ http://www.bionet.nsc.ru/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 statistics
- Information contents

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

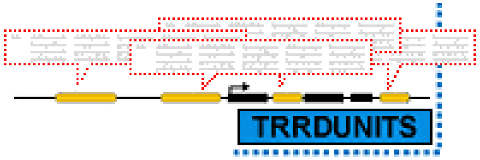
To obtain further particular information about TRRD database, please, address to the database scientific supervisor Nikolay A. Kolchanov by e-mail kol@bionet.nsc.ru or by fax +7-3832-331278

TRRD ВКЛЮЧАЕТ ДАННЫЕ О:

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

САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

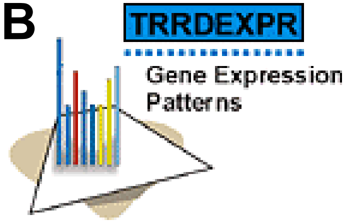
РЕГУЛЯТОРНЫХ РАЙОНОВ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)



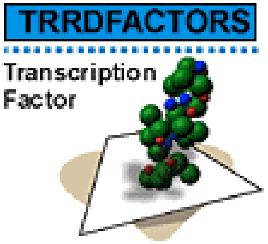
ЛОКУС-КОНТРОЛИРУЮЩИХ РАЙОНОВ

TRRDLCR

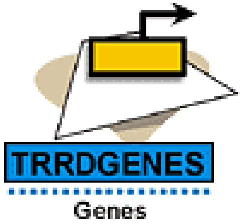
2) ДАННЫЕ О ПАТТЕРНАХ ЭКСПРЕССИИ ГЕНОВ



3) ДАННЫЕ О ТРАНСКРИПЦИОННЫХ ФАКТОРАХ



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

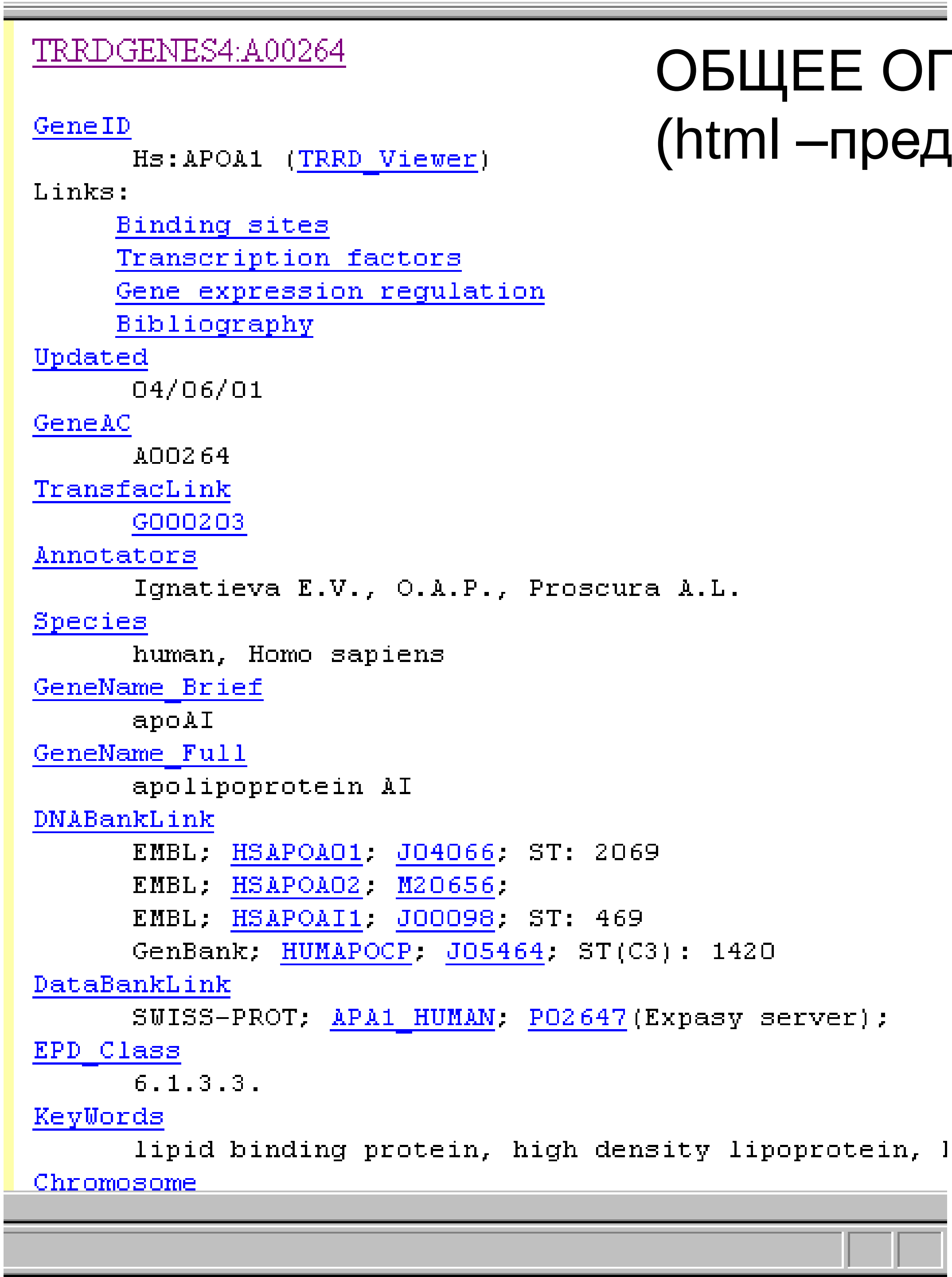


TRRDGENES

ОБЩЕЕ ОПИСАНИЕ ГЕНА

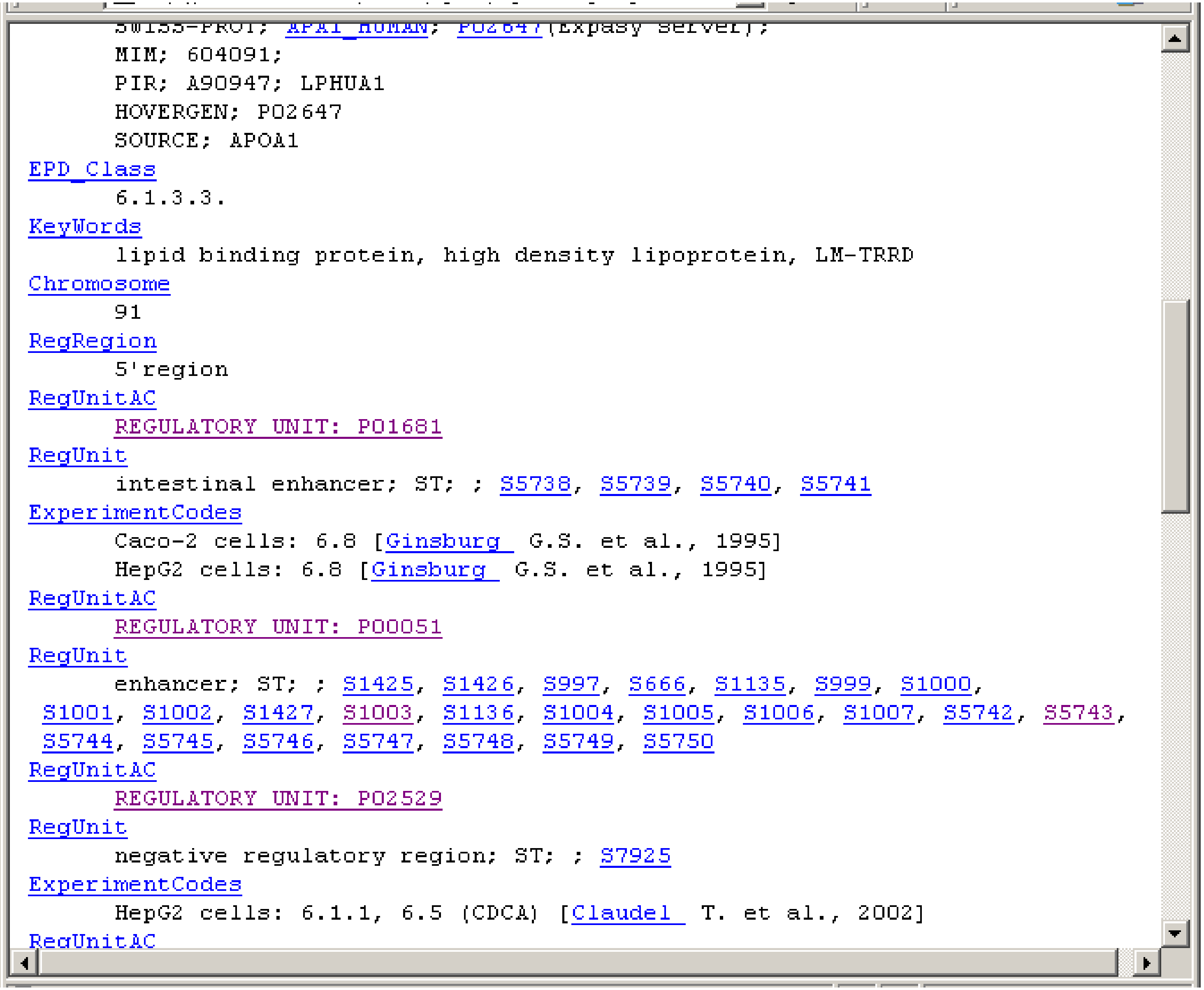
ID Hs:AAP
DT 09/02/00
AC A00596
CR Ignatieva E.V., Stepanenko I.L.
OS human, Homo sapiens
SN AAP
NG Alzheimer's disease amyloid A4 precursor protein
SY amyloid beta protein precursor gene
SY amyloid precursor protein gene
SY APP
BI EMBL;HSPADP;X12751; ST:3700
DR GDB; 119692; APP
DR SWISS-PROT; A4_HUMAN; 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
//

ОБЩЕЕ ОПИСАНИЕ ГЕНА
(html –представление)



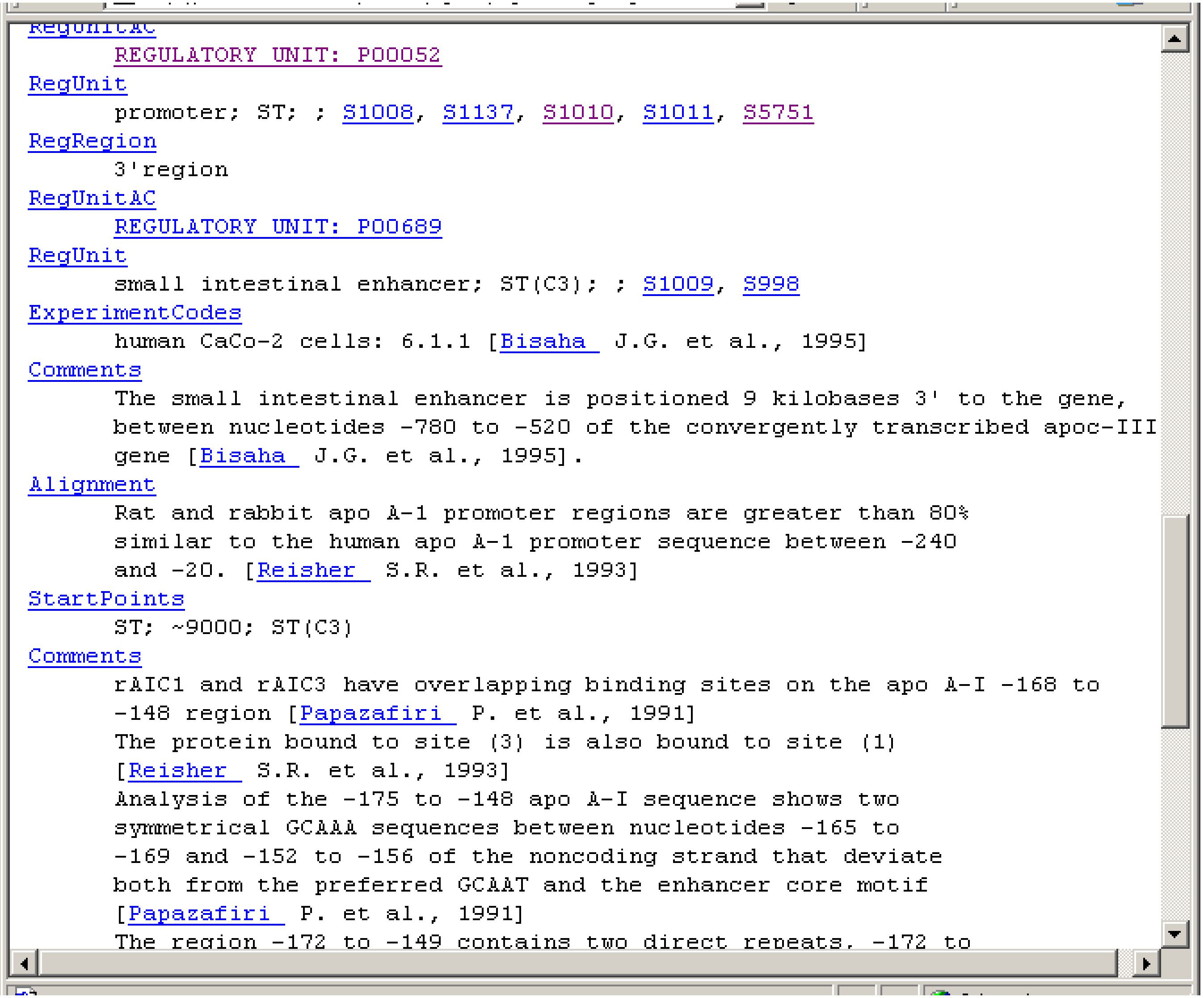
```
TRRDGENES4:A00264

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
       apoAI
GeneName_Full
       apolipoprotein AI
DNABankLink
       EMBL; HSAPOA01; J04066; ST: 2069
       EMBL; HSAPOA02; M20656;
       EMBL; HSAPOAI1; 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,
Chromosome
```

Продолжение:
ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)



```
        SWISS-PROT; APA1_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
```

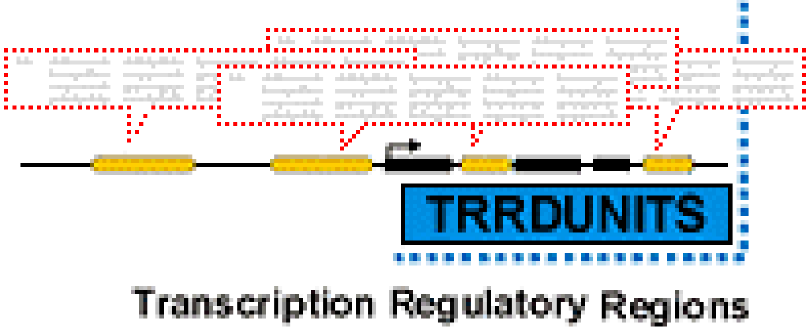
Окончание:
ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)



```
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 GCAAA 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
```

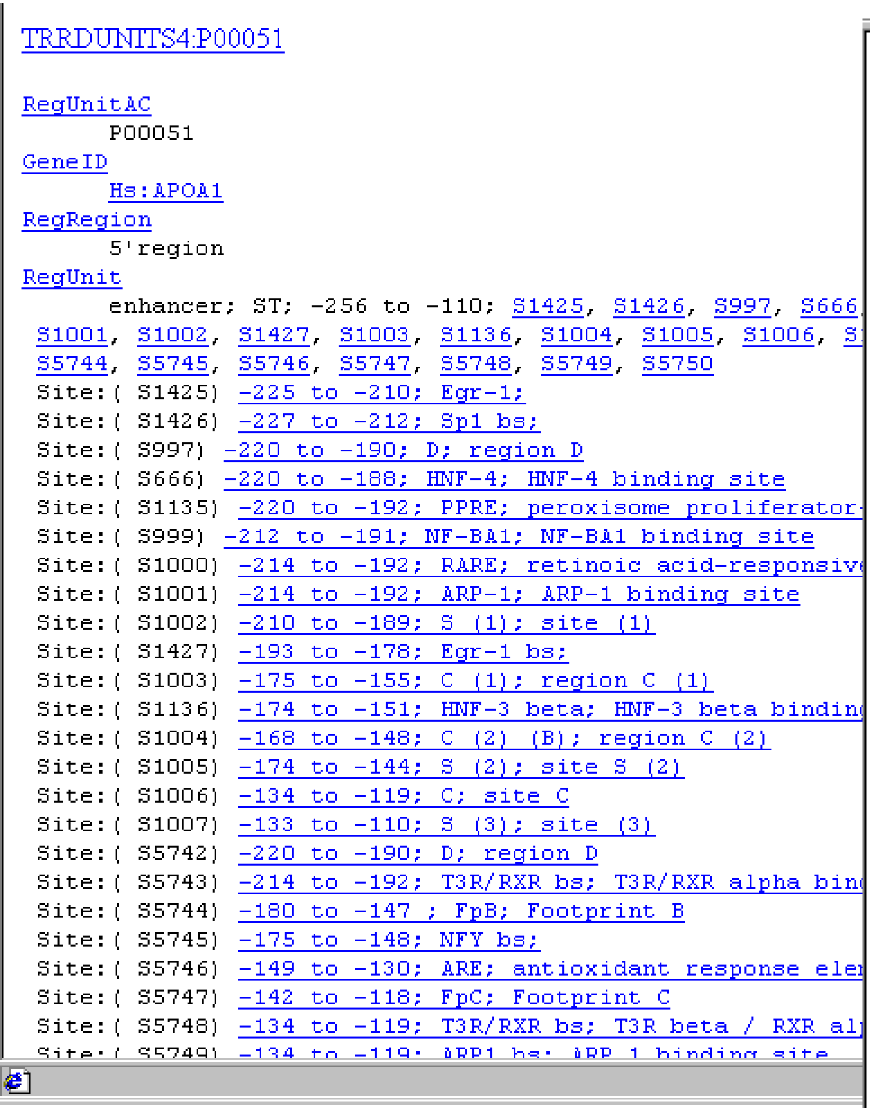
TRRDUNITS

ОПИСАНИЕ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ ЕДИНИЦ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)



RegUnitAC	P00562
GeneID	Rn:D2
RegRegion	5'region
RegUnit	Promoter; ST; -150 to +1; S79, S80
DNA BankLink	EMBL; RND2RPR; X77137; 704 to 855
LeftTrunc	0
RightTrunc	0
SeqLength	152
Sequence	cccaggcccc acagtgcaga gatagttctg gggccctggg tgggtggggc ctctgtacaa ggggcggggt tcccgggcgc ctcgtggcca gggtgacccc gccccctcct cctgcgcagc gctctgattc cgcggagctg tccagcctca gt
PromotTisSp	0
PromotInd	1
ExperimentCodes	6.1.1, 6.8 [Minowa T. 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, S
 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-responsive
 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; HSAPOA01; J04066; 1813 to 1959
LeftTrunc
        0
RightTrunc
        0
SeqLength
        147
Sequence
        ccacccggga gacctgcaag cctgcagcac tcccctcccg cccccactga
        acccttgacc cctgccctgc agcccccgca gcttgctgtt tgcccactct
        atttgcccag ccccagggac agagctgatc cttgaactct taagttc
DNA_BankLink
        EMBL; HSAPOA01; J04066; 1813 to 1959
LeftTrunc
        0
RightTrunc
        0
SeqLength
        147
Sequence
        ccacccggga gacctgcaag cctgcagcac tcccctcccg cccccactga
        acccttgacc cctgccctgc agcccccgca gcttgctgtt tgcccactct
        atttgcccag ccccagggac agagctgatc cttgaactct taagttc
DNA_BankLink
        EMBL; HSAPOAI1; J00098; 215 to 361
LeftTrunc
        0
RightTrunc
        0
SeqLength
        147
Sequence
        cccgggagac ctgcaagcct gcagcactcc cctcccgccc ccactgaacc
        cttgacccct gccctgcacg ccccgcagct tgctgtttgc ccactcctat
        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]

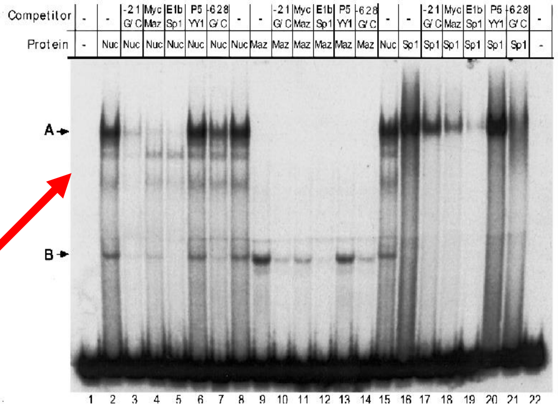
AG rat liver cells: 3.6 [Metzger S. et al., 1993]

AG human HepG2 cells: 6.2 [Metzger S. et al., 1993]

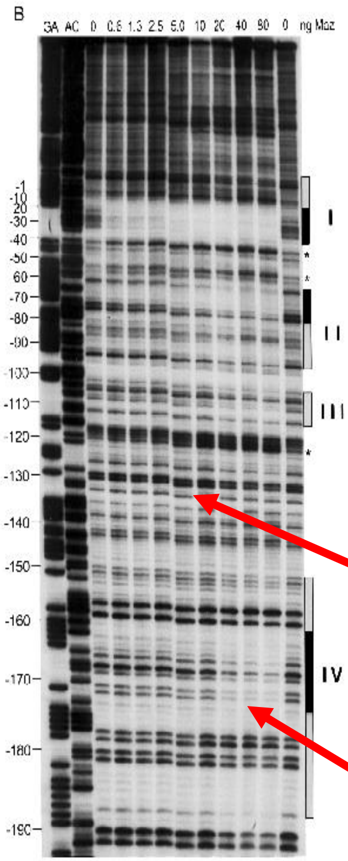
AG HeLa cells: 6.2, 6.6 [Metzger S. et al., 1993]

AG 1.1.5, 3.3, 3.5, 4.2 [Ladias J.A. et al., 1992]

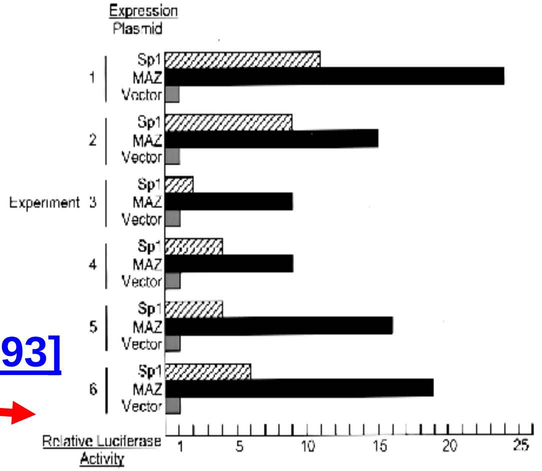
//



GEL-MOBILITY SHIFT ASSAY



DNASE I FOOTPRINTING

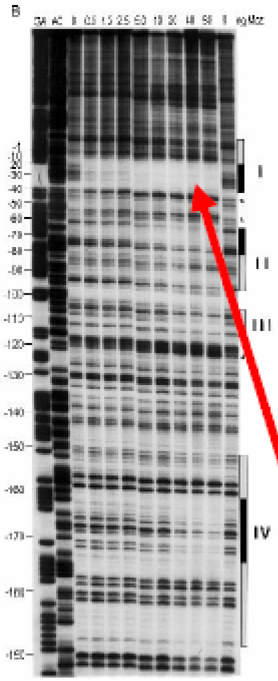


TRANSIENT EXPRESSION ANALYSIS

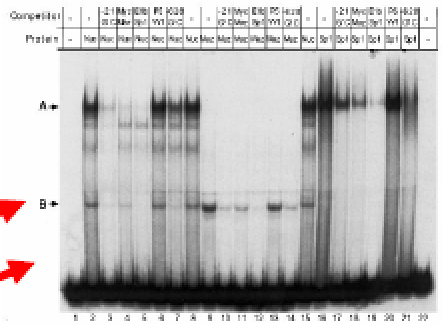
The table TRRDSITES: description of the HNF-4 binding site in the human ApoB gen

```
-88 5′ cccgggaggCGCCCTTTGGACCTttg 3′ -62
    3′ gggccctccGCGGGAAACCAGGAaaac 5′
```

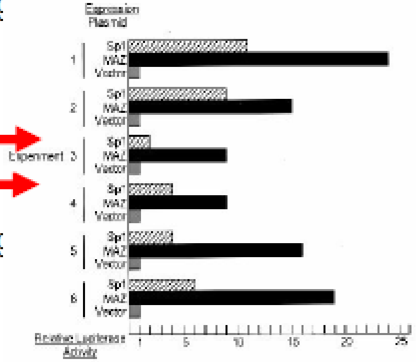
SiteAC S1160
GeneID Hs:APOB
RegUnitAC P00670
SiteName HNF-4;
PreferredName HNF-4
SiteNameSynonym AF-1 binding site
SiteNameSynonym BA1 binding site
SiteIndex 1
FactorName HNF-4; hepatic nuclear factor 4
FactorInfluence increase
Sequence cccgggaggCGCCCTTTGGACCTtttg
SequencePosition -88 to -62
FootprintSequencePosition -82 to -62
DNA_BankLink M15053: 68
ImportantPos
----------CG--CTTTGGACCT-t-; HNF4 [Metzger S. et al., 19
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]
HeLa 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., 19
//



DNASE I FOOTPRINTING



GEL-MOBILITY SHIFT ASSAY



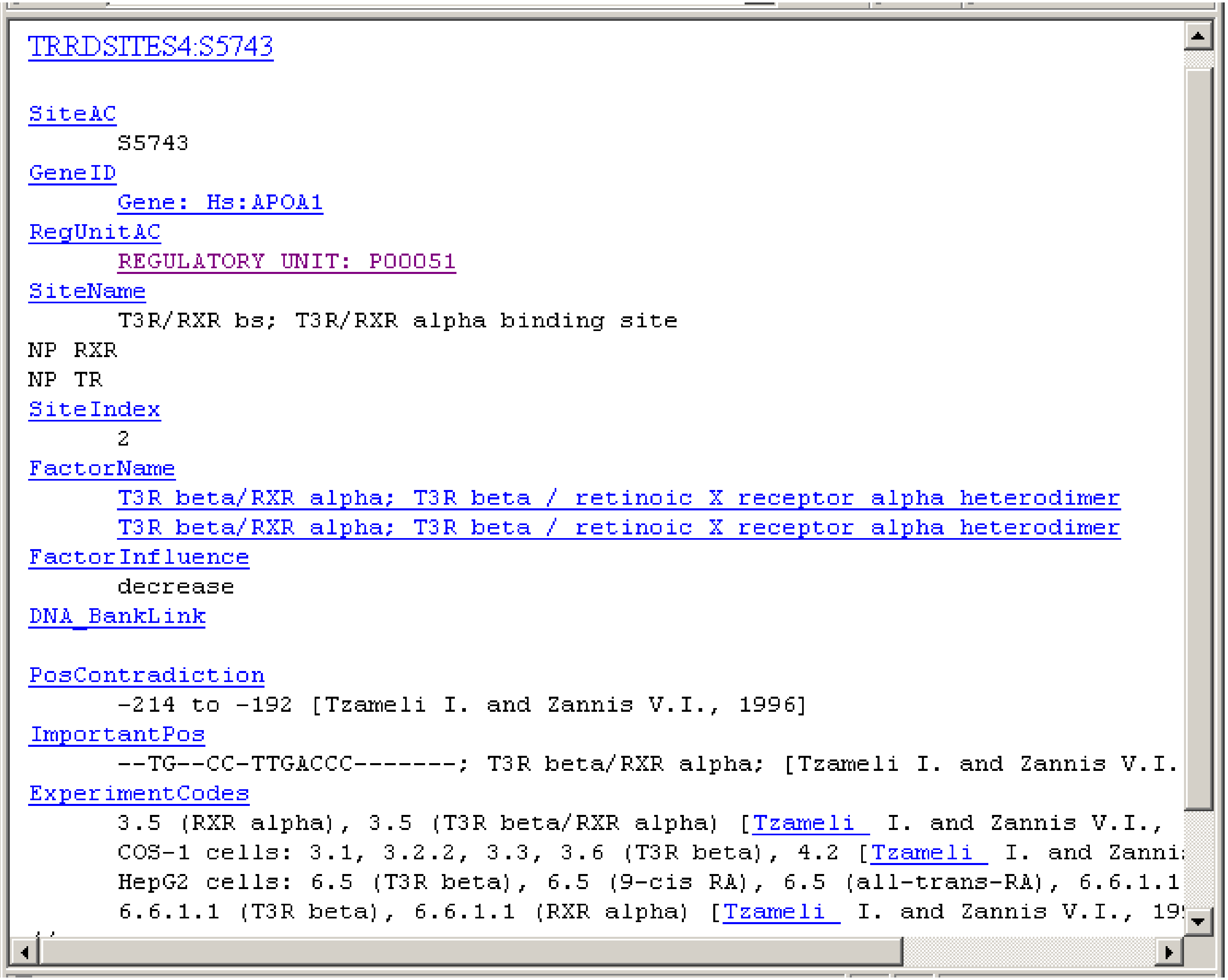
TRANSIENT EXPRESSION ANALYSIS

Экспериментальные методики, по результатам выполнения которых заносятся данные в базу TRRD

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

Полный список см. по адресу
http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/digcodes.shtml

ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА (html –представление)

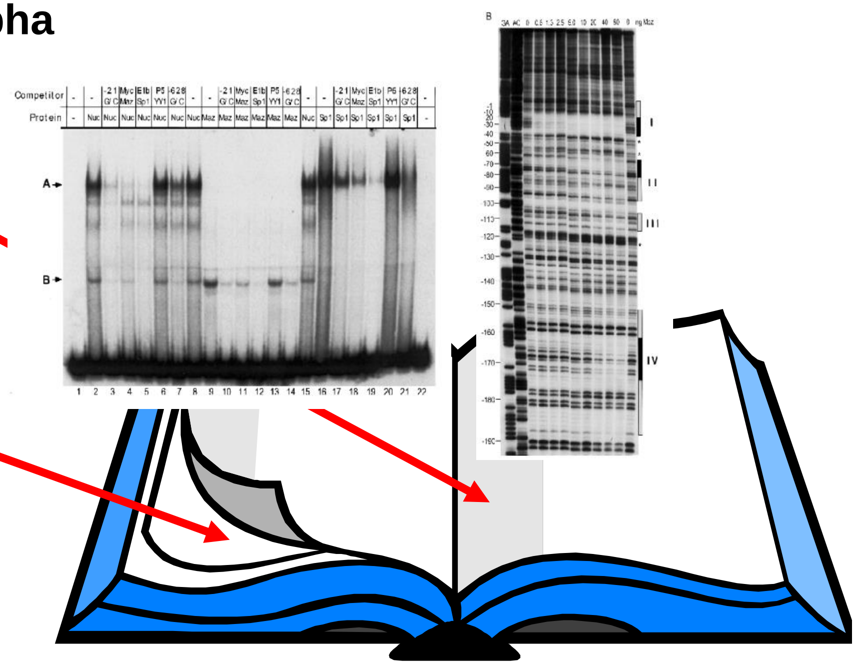


```
TRRDSITES4:S5743

SiteAC
        S5743
GeneID
        Gene: Hs:APOA1
RegUnitAC
        REGULATORY UNIT: P00051
SiteName
        T3R/RXR bs; T3R/RXR alpha binding site
NP RXR
NP TR
SiteIndex
        2
FactorName
        T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer
        T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer
FactorInfluence
        decrease
DNA_BankLink

PosContradiction
        -214 to -192 [Tzameli I. and Zannis V.I., 1996]
ImportantPos
        --TG--CC-TTGACCC-------; T3R beta/RXR alpha; [Tzameli I. and Zannis V.I.
ExperimentCodes
        3.5 (RXR alpha), 3.5 (T3R beta/RXR alpha) [Tzameli  I. and Zannis V.I.,
        COS-1 cells: 3.1, 3.2.2, 3.3, 3.6 (T3R beta), 4.2 [Tzameli  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) [Tzameli  I. and Zannis V.I., 19
```

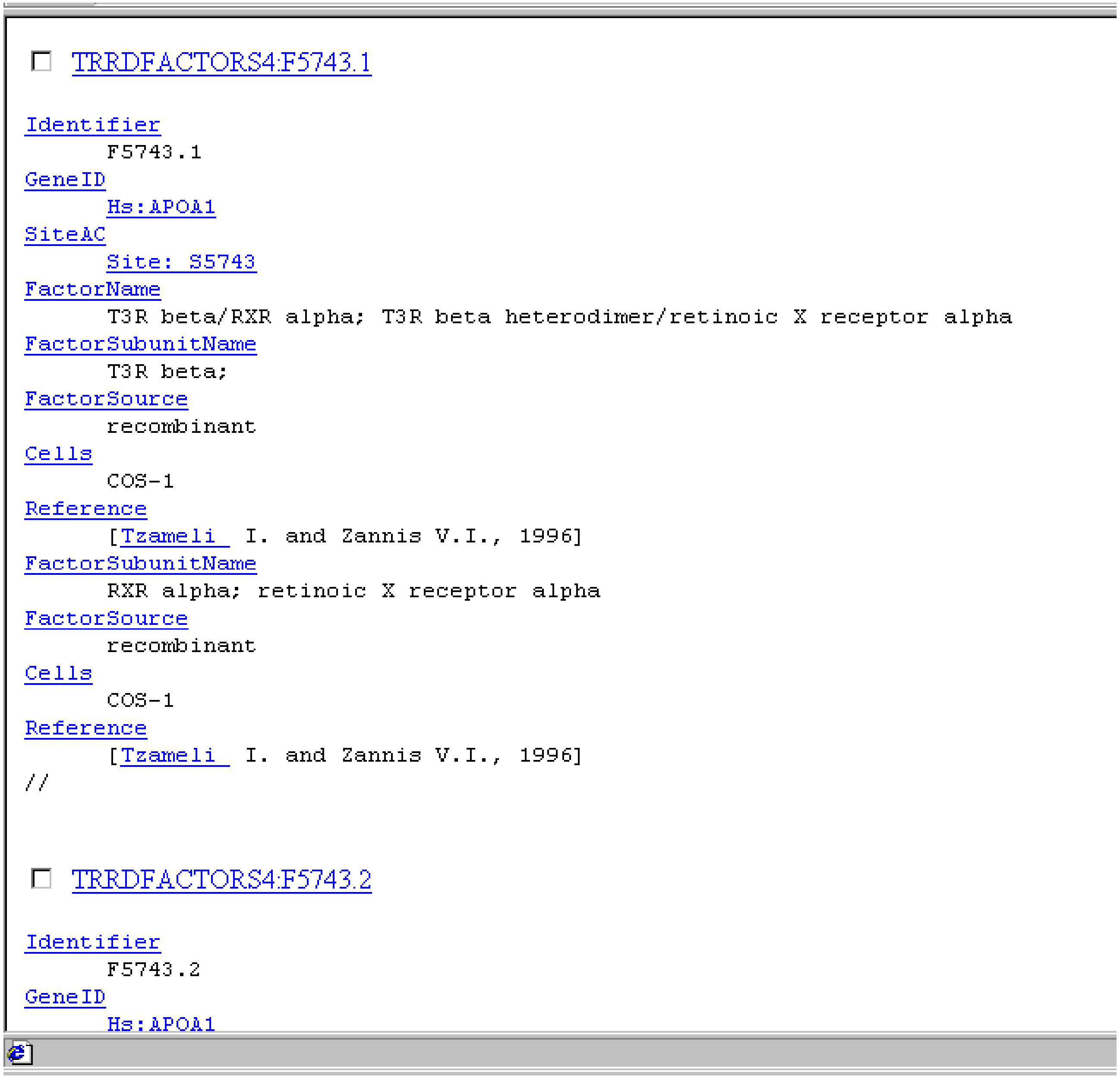
TRRDFACTORS

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: T01331
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

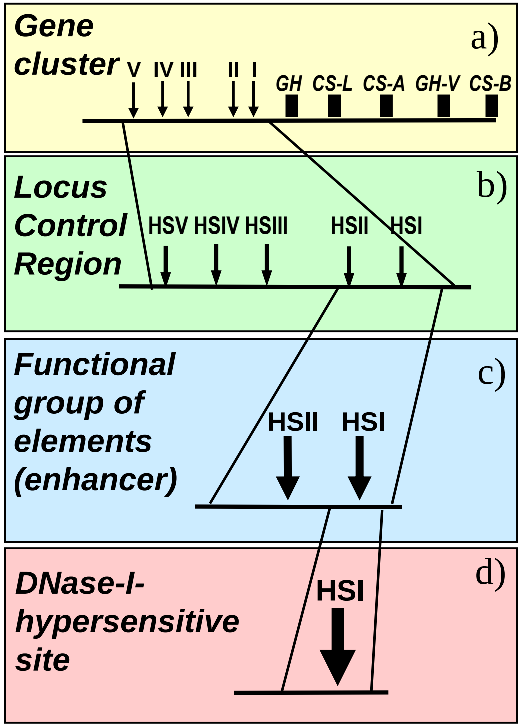
```
Identifier
        F5743.2
GeneID
        Hs:APOA1
```

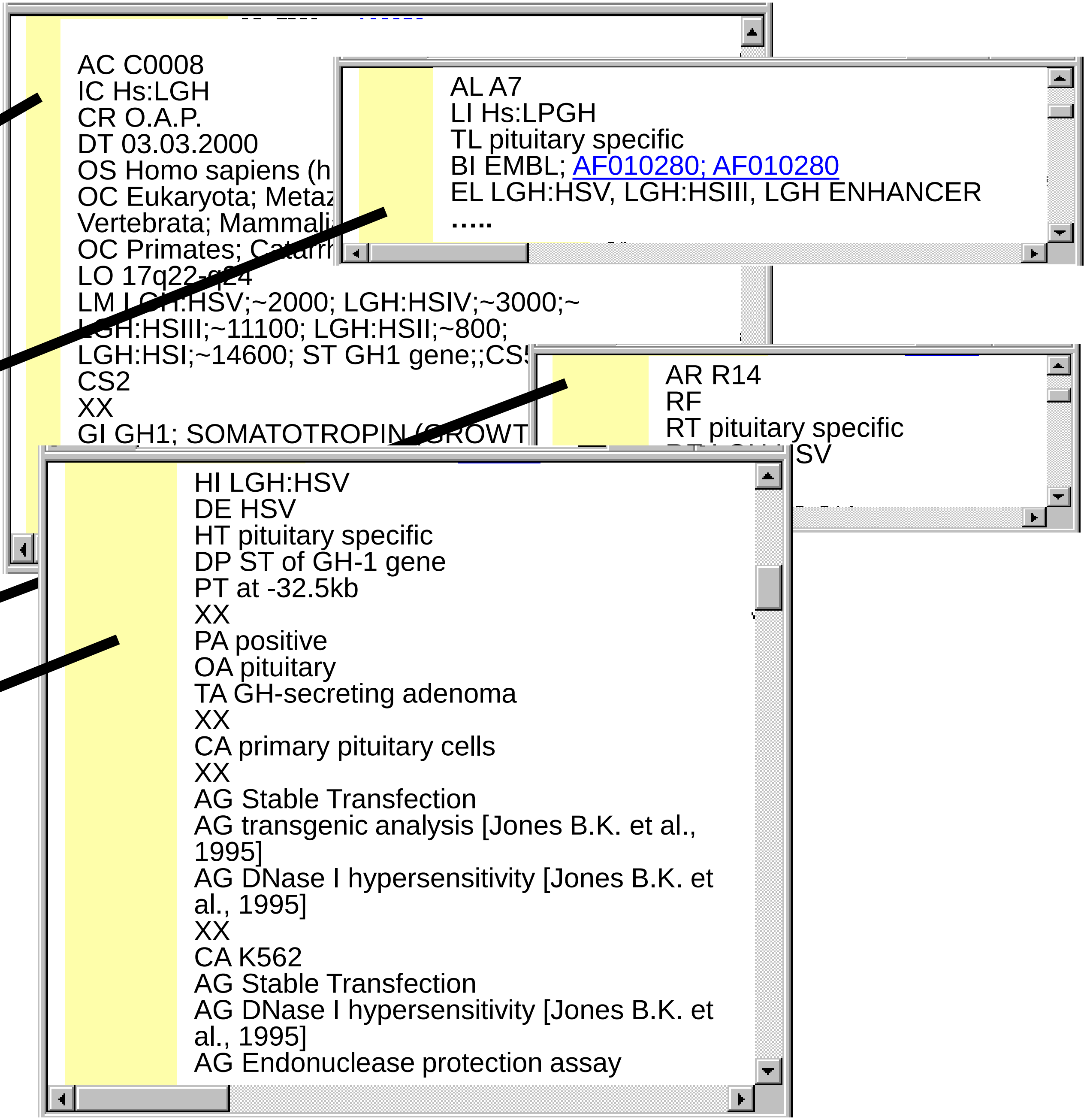
TRRDLCR

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

AL A7
LI Hs:LPGH
TL pituitary specific
BI EMBL; AF010280; AF010280
EL LGH:HSV, LGH:HSIII, LGH ENHANCER
…..

AR R14
RF
RT pituitary specific

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

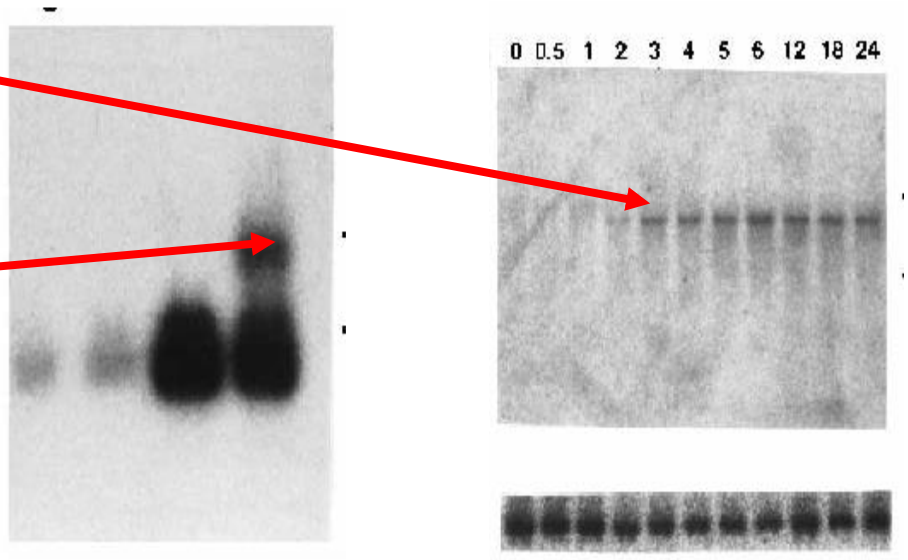
TRRDEXPR

Gene Expression Patterns

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]

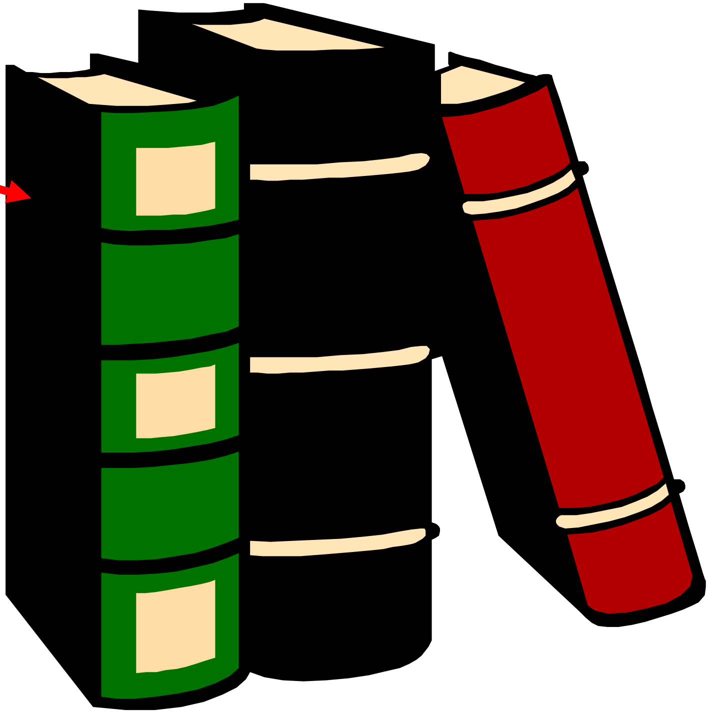
RE A00596.001
ID Hs:AAP
RT mRNA
RN astrocytes
RL present
RR [Amara F.M., et al., 1999]

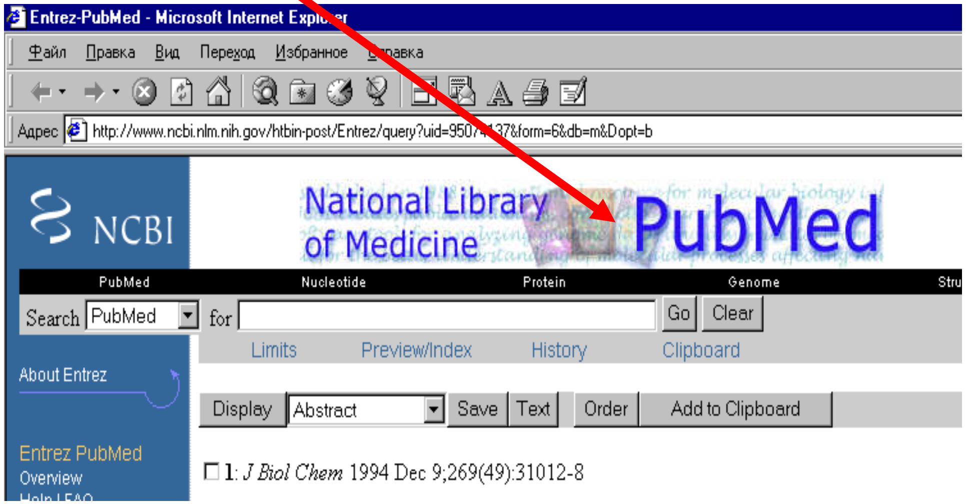


RNA blot analysis

TRRDBIB

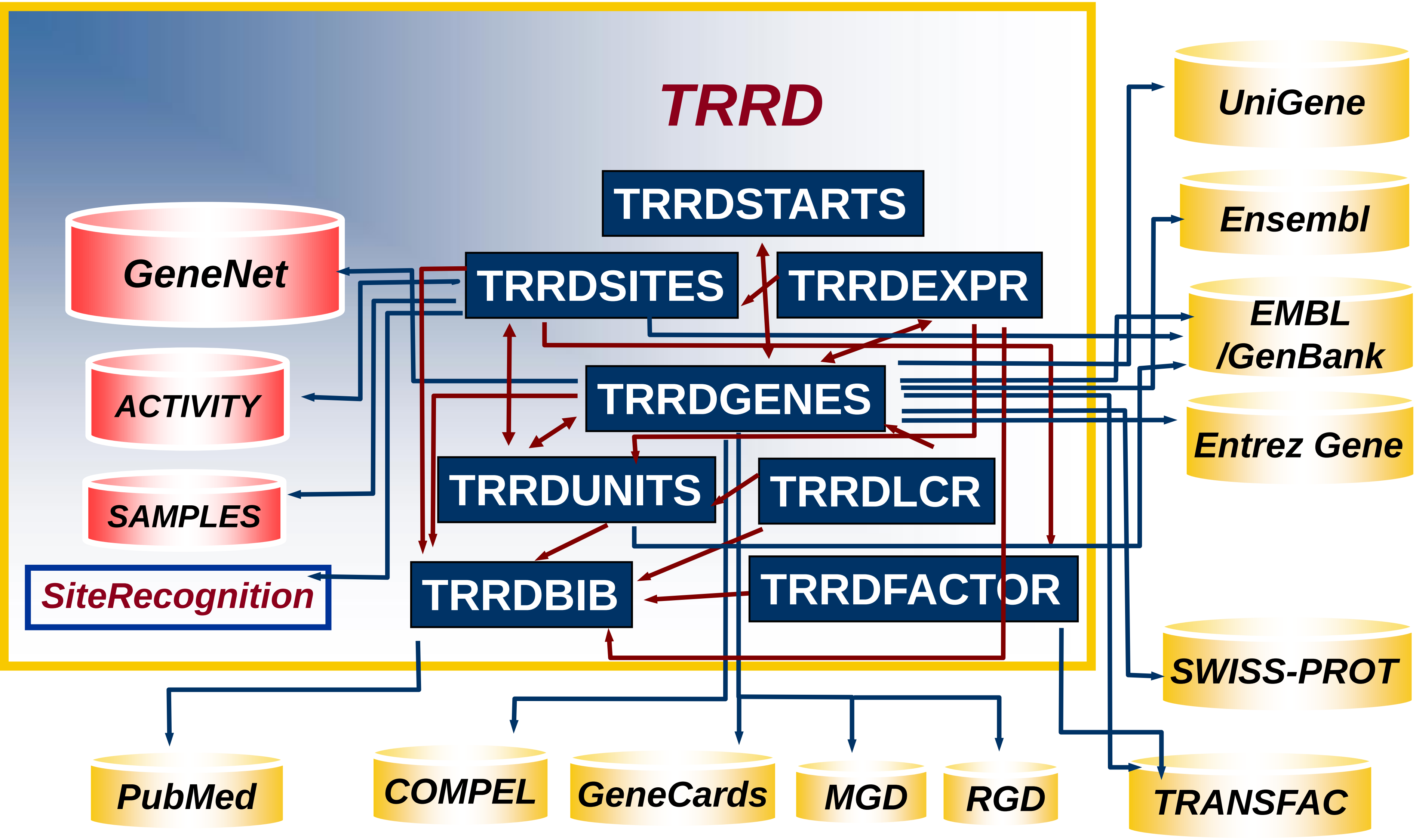
NN 2027
ID [Hs:AAP]
AU Trejo J., Massamiri T., Deng T., Dewji N.N., Bayney R.M., Brown J.H.
TI A direct role for protein kinase C and the transcription factor Jun/AP-1
TI in the regulation of the Alzheimer's beta-amyloid precursor protein gene.
SO J.Biol.Chem.
VL 269
IS 34
YR 1994
PG 21682-21690
ML MEDLINE:[94342362], [See Related Articles]
//





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

Линковка между таблицами TRRD и ссылки на внешние ресурсы.



TRRD

TRRDSTARTS

TRRDSITES

TRRDEXPR

TRRDGENES

TRRDUNITS

TRRDLCR

TRRDBIB

TRRDFACTOR

GeneNet

ACTIVITY

SAMPLES

SiteRecognition

UniGene

Ensembl

EMBL /GenBank

Entrez Gene

SWISS-PROT

TRANSFAC

PubMed

COMPEL

GeneCards

MGD

RGD

ГИПЕРССЫЛКИ НА БАЗЫ ДАННЫХ И ПРОГРАММНЫЕ МОДУЛИ СИСТЕМЫ GENEEXPRESS из БАЗЫ TRRDSITES

```
AN S161
ID Gene: Hs:PAI1
AP REGULATORY UNIT: T00150
NM USF bs;
WW http://wwwmgs.bionet.nsc.ru/Programs/acts2/AFF_USF.htm
DR ACTIVITY ; A00DF026;
DR SAMPLES; USF;
TF USF; upstream sequence factor
AT increase
SQ gaca[illegible]tcacgtggct
PQ -574 to -559
PF -574 to -559
BF EMBL: X13323 : 236
AG 1.1.1, 1.1.2, 3.1, 3.3, 6.1, 6.2, 6.3 [Riccio A. et al.,
199[illegible]]
//
```

TRRDGENE

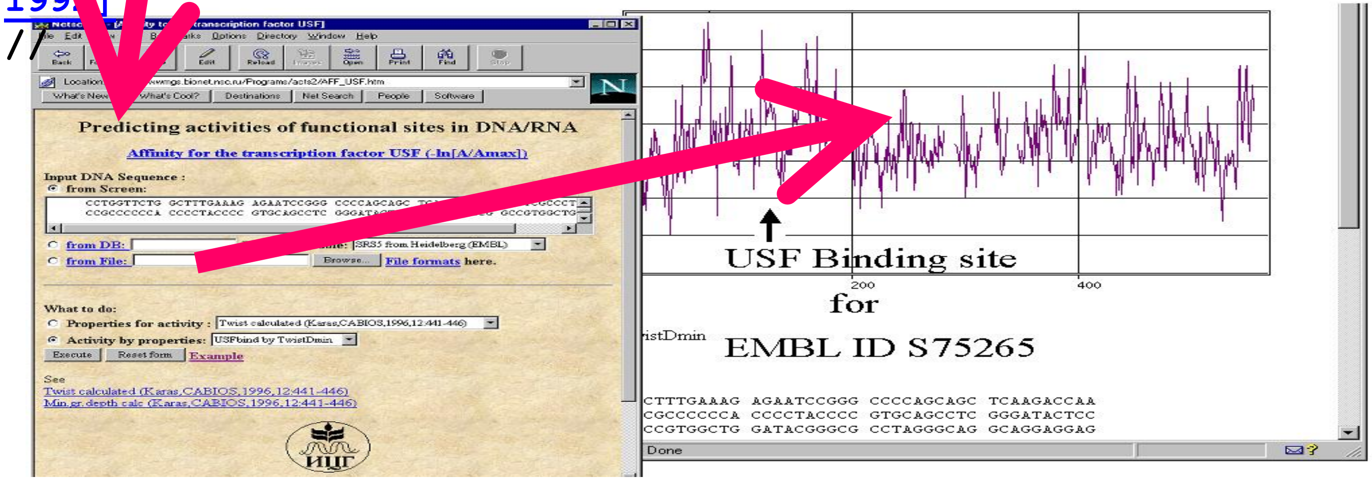
ACTIVITY

SAMPLES

TRRDFACTORS

EMBL

TRRDBIB



Predicting activities of functional sites in DNA/RNA

Affinity for the transcription factor USF (-ln[A/Amax])

Input DNA Sequence :

● from Screen:

```
CCTGGTTCTG GCTTTGAAAG AGAATCCGGG CCCCAGCAGC TC...GCCCT
CCGCCCCCCA CCCCTACCCC GTGCAGCCTC GGGATA... GCCGTGGCTG
```

○ from DB: [ ] ...me: SRS5 from Heidelberg (EMBL)

○ from File: [ ] Browse... File formats here.

What to do:

○ Properties for activity : Twist calculated (Karas,CABIOS,1996,12:441-446)

● Activity by properties: USFbind by TwistDmin

Execute | Reset form | Example

See

Twist calculated (Karas,CABIOS,1996,12:441-446)

Min.gr.depth calc (Karas,CABIOS,1996,12:441-446)

USF Binding site

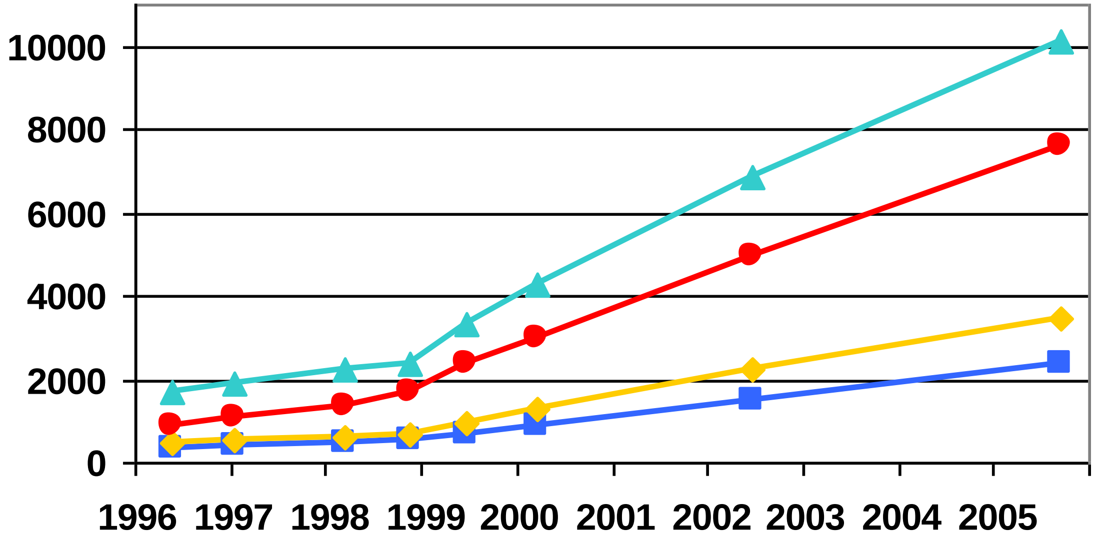
for

...istDmin

EMBL ID S75265

```
CTTTGAAAG AGAATCCGGG CCCCAGCAGC TCAAGACCAA
CGCCCCCCA CCCCTACCCC GTGCAGCCTC GGGATACTCC
CCGTGGCTG GATACGGGCG CCTAGGGCAG GCAGGAGGAG
```

ДИНАМИКА ПОПОЛНЕНИЯ TRRD



10135 САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

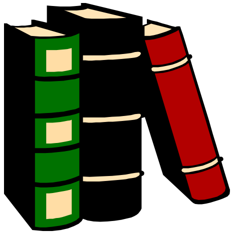
7609 НАУЧНЫХ ПУБЛИКАЦИЙ

3400 РЕГУЛЯТОРНЫХРАЙОНОВ

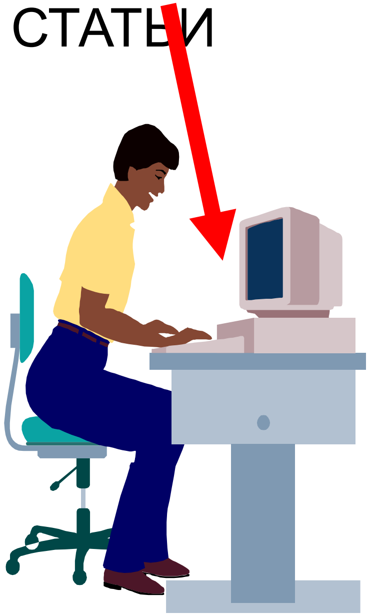
2344 ГЕНОВ

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

ПРОЦЕСС ВВОДА ДАННЫХ В TRRD



ЭКСПЕРИМЕНТАЛЬНЫЕ СТАТЬИ



22 контролируемых словаря (>3000 слов)

СТАДИИ РАЗВИТИЯ

КЛЕТКИ

ТКАНИ

ОРГАНЫ

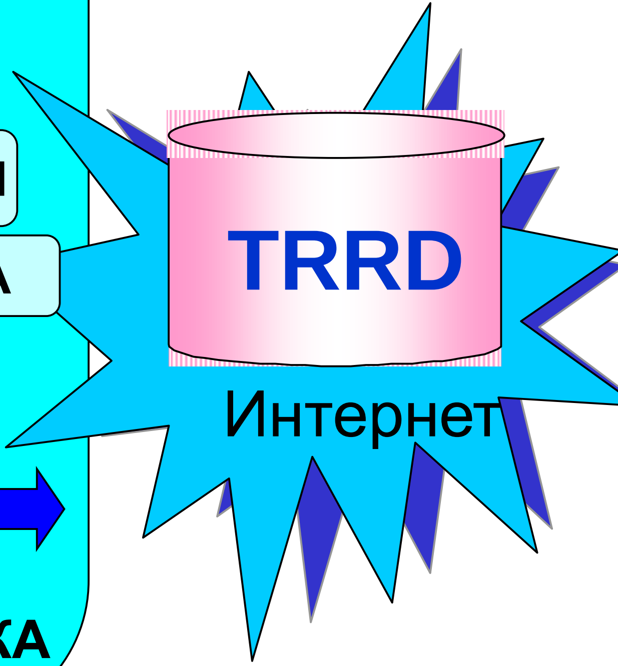
КЛЮЧЕВЫЕ СЛОВА

ВНЕШНИЕ СТИМУЛЫ

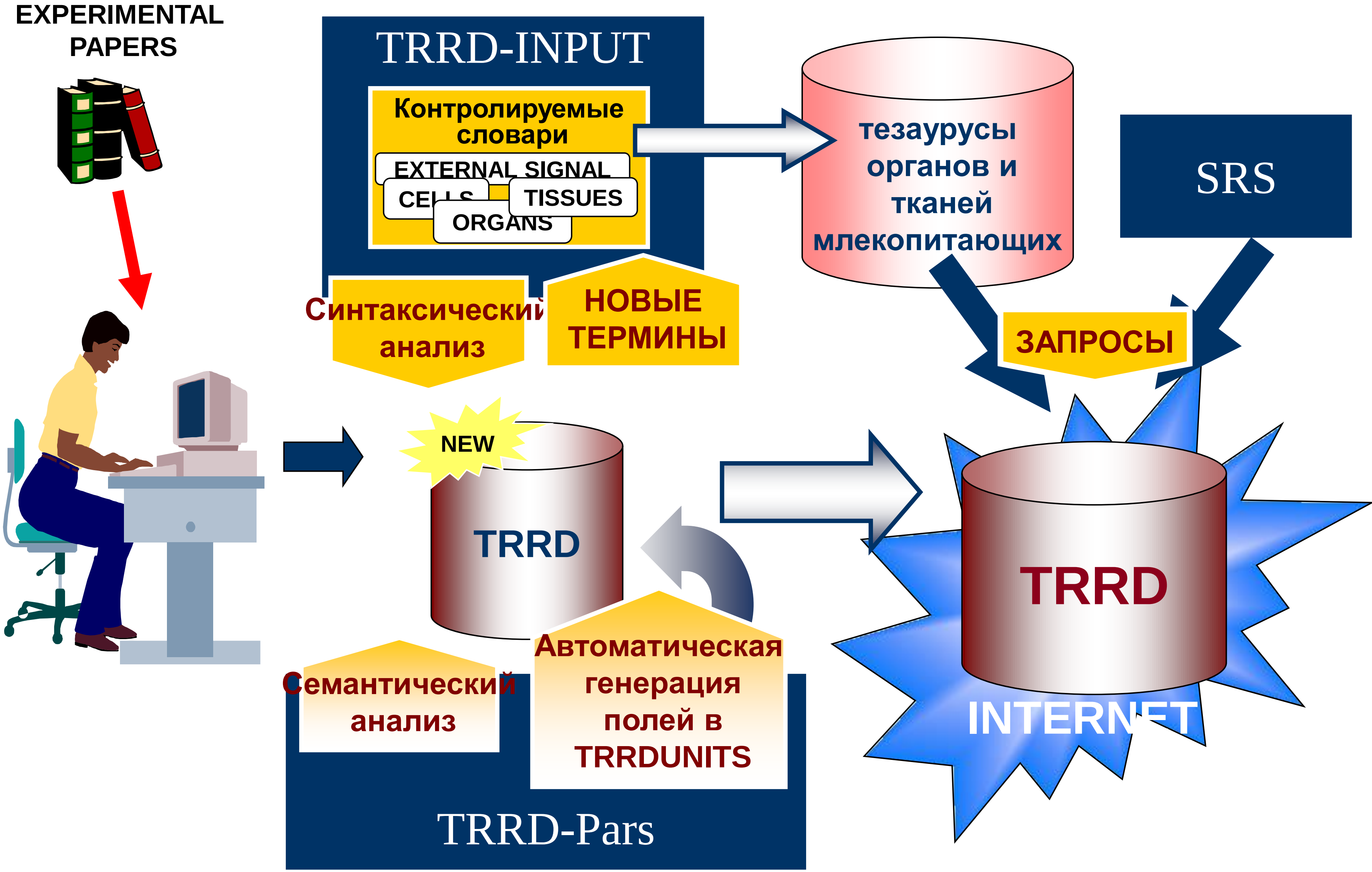
СТАДИИ КЛЕТОЧНОГО ЦИКЛА

КОДЫ ЭКСПЕРИМЕНТОВ

СИНТАКСИЧЕСКАЯ и СЕМАНТИЧЕСКАЯ ПРОВЕРКА



ВВОД и ОБРАБОТКА ДАННЫХ



EXPERIMENTAL PAPERS

TRRD-INPUT

Контролируемые словари

EXTERNAL SIGNAL

CELLS

TISSUES

ORGANS

тезаурусы органов и тканей млекопитающих

SRS

Синтаксический анализ

НОВЫЕ ТЕРМИНЫ

ЗАПРОСЫ

NEW

TRRD

TRRD

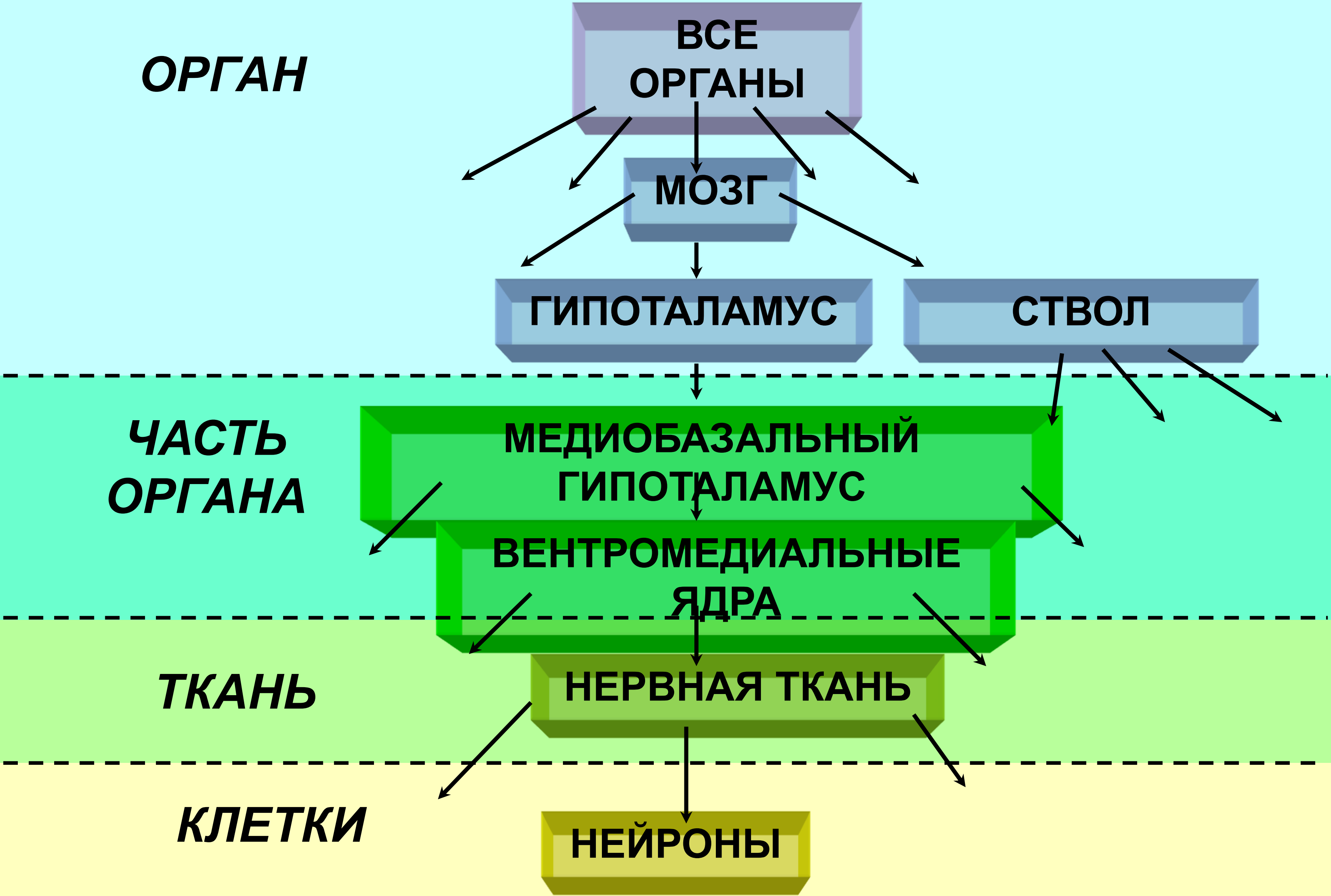
INTERNET

Семантический анализ

Автоматическая генерация полей в TRRDUNITS

TRRD-Pars

ИЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ КОНТРОЛИРУЕМЫХ СЛОВАРЕЙ МОРФОЛОГИЧЕСКИХ ТЕРМИНОВ В TRRD



ОРГАН

ВСЕ ОРГАНЫ

МОЗГ

ГИПОТАЛАМУС

СТВОЛ

ЧАСТЬ ОРГАНА

МЕДИОБАЗАЛЬНЫЙ ГИПОТАЛАМУС

ВЕНТРОМЕДИАЛЬНЫЕ ЯДРА

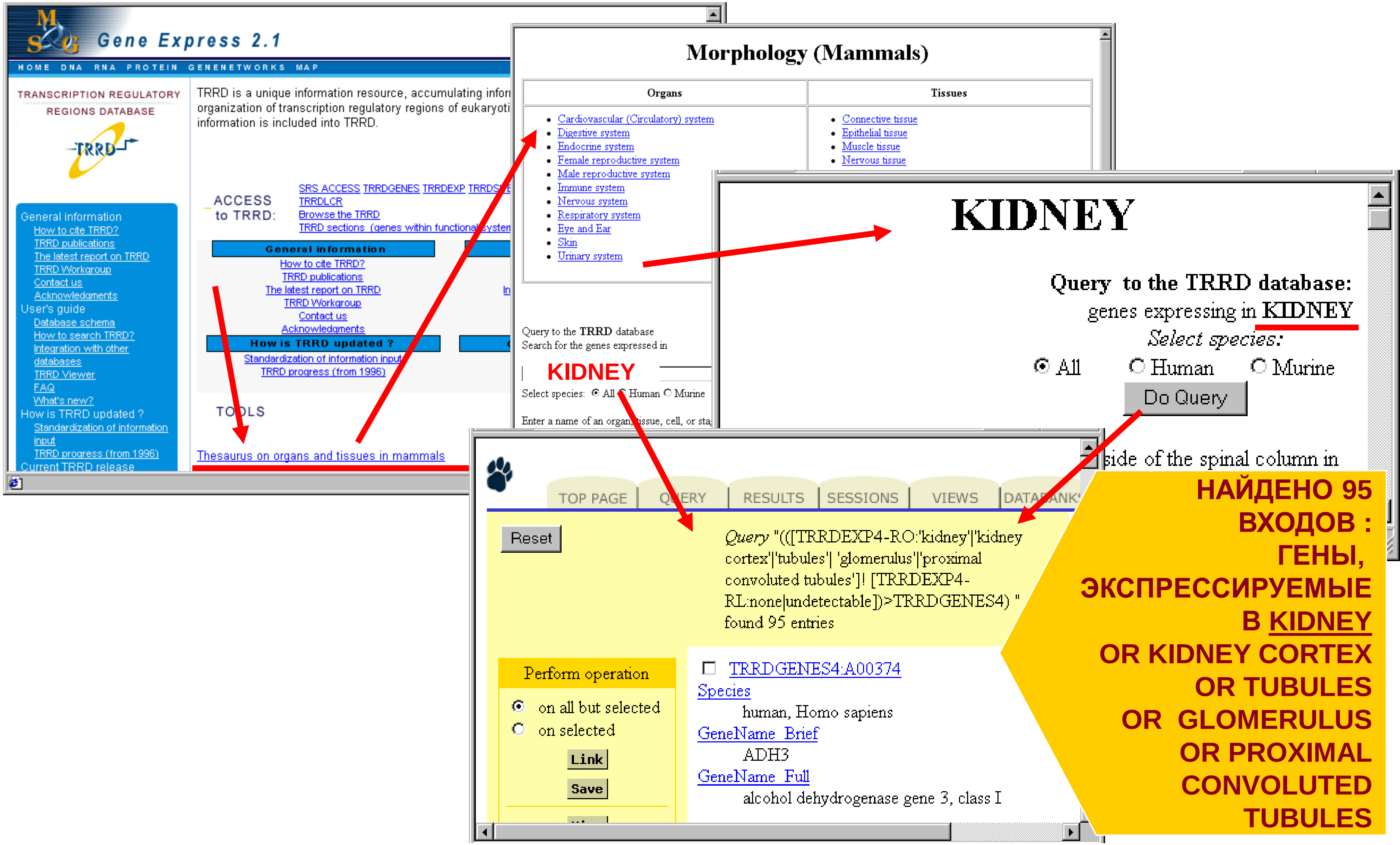
ТКАНЬ

НЕРВНАЯ ТКАНЬ

КЛЕТКИ

НЕЙРОНЫ

ПОИСК ДАННЫХ вTRRD: ЗАПРОС НА ОСНОВЕ ТЕЗАУРОСОВ ОРГАНОВ и ТКАНЕЙ МЛЕКОПИТАЮЩИХ



TRRD is a unique information resource, accumulating infor… organization of transcription regulatory regions of eukaryoti… information is included into TRRD.

Morphology (Mammals)

Organs	Tissues
- Cardiovascular (Circulatory) system - Digestive system - Endocrine system - Female reproductive system - Male reproductive system - Immune system - Nervous system - Respiratory system - Eye and Ear - Skin - Urinary system	- Connective tissue - Epithelial tissue - Muscle tissue - Nervous tissue

Query to the **TRRD** database
Search for the genes expressed in

KIDNEY

Select species: ● All ○ Human ○ Murine

Enter a name of an organ, tissue, cell, or sta…

KIDNEY

Query to the TRRD database:
genes expressing in KIDNEY
Select species:
● All ○ Human ○ Murine

…side of the spinal column in

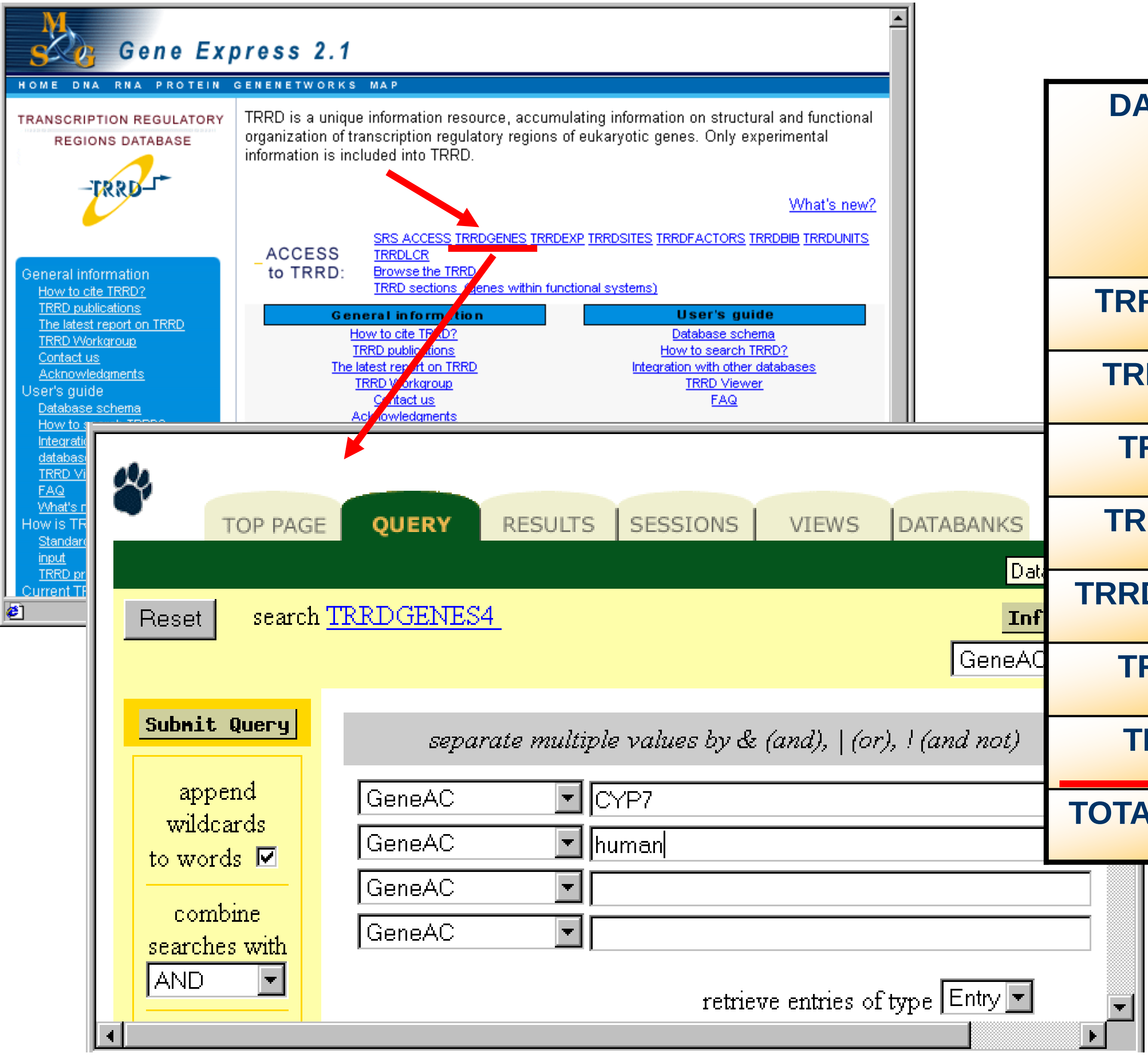
Query "(([TRRDEXP4-RO:'kidney'|'kidney cortex'|'tubules'| 'glomerulus'|'proximal convoluted tubules']! [TRRDEXP4-RL:none|undetectable])>TRRDGENES4) " found 95 entries

Perform operation
● on all but selected
○ on selected

☐ TRRDGENES4:A00374
Species
human, Homo sapiens
GeneName_Brief
ADH3
GeneName_Full
alcohol dehydrogenase gene 3, class I

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

ПОИСК ДАННЫХ в TRRD: SRS



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 TRRDLCR
Browse the TRRD
TRRD sections (genes within functional systems)

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

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 s
Integrati
databas
TRRD Vi
FAQ
What's n
How is TR
Standard
input
TRRD pr
Current TR

search TRRDGENES4

GeneAC

append wildcards to words ☑

combine searches with AND

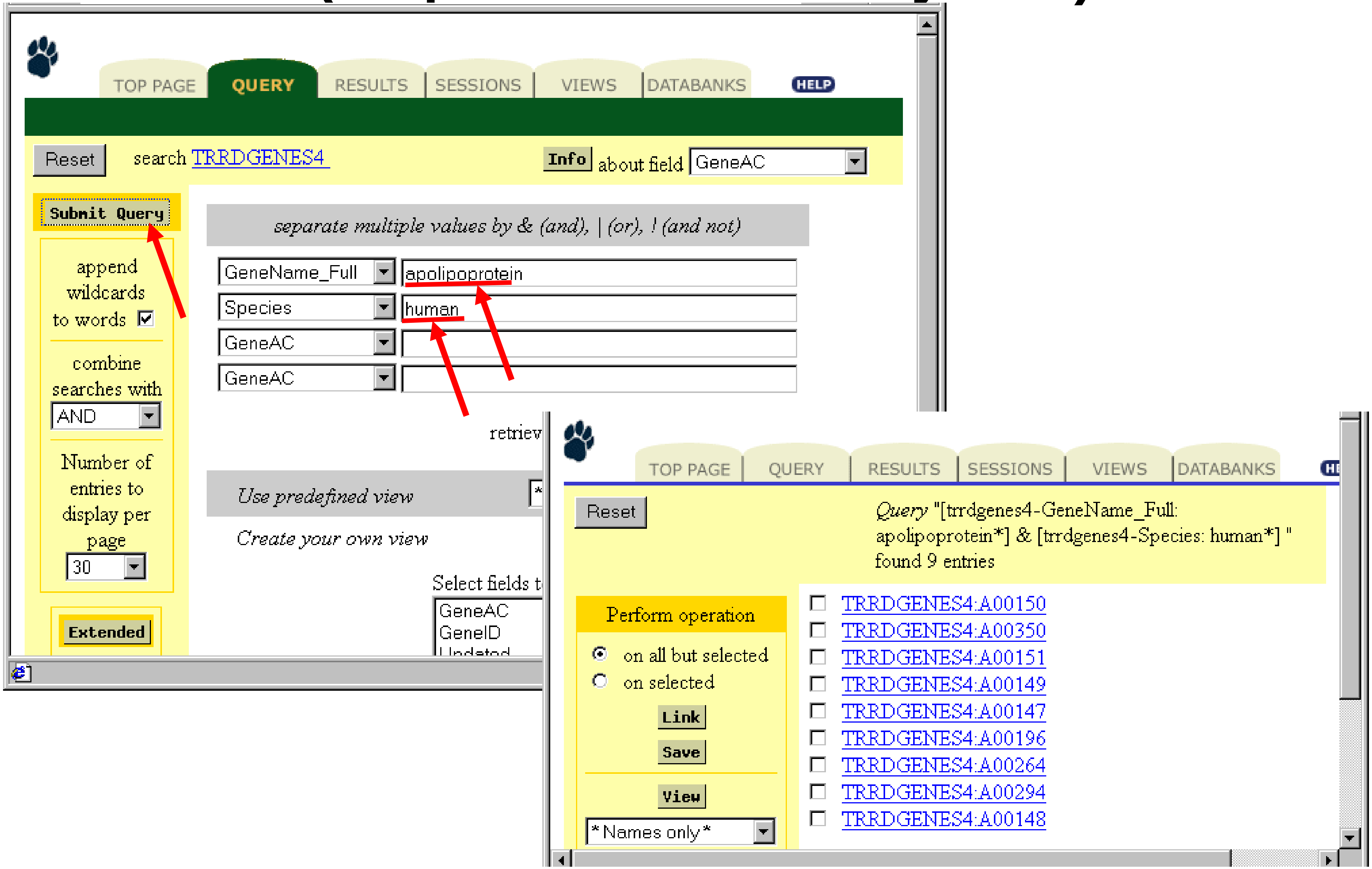
separate multiple values by & (and), | (or), ! (and not)

GeneAC CYP7
GeneAC human
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
TOTAL NUMBER	131

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



search TRRDGENES4

about field GeneAC

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

GeneName_Full	apolipoprotein
Species	human
GeneAC	
GeneAC	

append wildcards to words ☑

combine searches with AND

Number of entries to display per page 30

Use predefined view

Create your own view

Select fields t

GeneAC
GeneID

Query "[trrdgenes4-GeneName_Full: apolipoprotein*] & [trrdgenes4-Species: human*] " found 9 entries

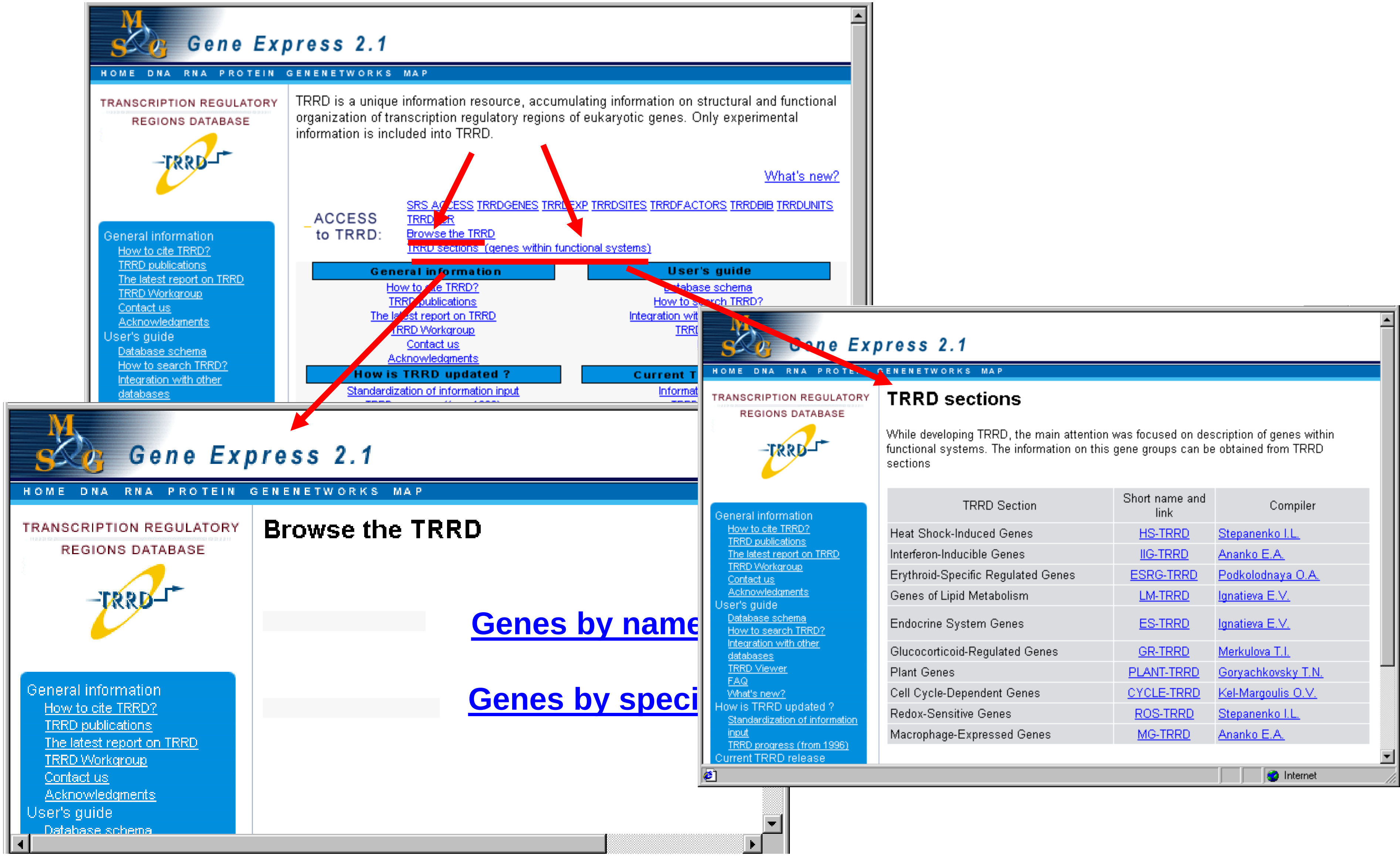
Perform operation

● on all but selected
○ on selected

Names only

☐ 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



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
TRRD ER
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 wit
TRRI

How is TRRD updated ?
Standardization of information input

Current T
Informat

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

Browse the TRRD

Genes by name

Genes by speci

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	HS-TRRD	Stepanenko I.L.
Interferon-Inducible Genes	IIG-TRRD	Ananko E.A.
Erythroid-Specific Regulated Genes	ESRG-TRRD	Podkolodnaya O.A.
Genes of Lipid Metabolism	LM-TRRD	Ignatieva E.V.
Endocrine System Genes	ES-TRRD	Ignatieva E.V.
Glucocorticoid-Regulated Genes	GR-TRRD	Merkulova T.I.
Plant Genes	PLANT-TRRD	Goryachkovsky T.N.
Cell Cycle-Dependent Genes	CYCLE-TRRD	Kel-Margoulis O.V.
Redox-Sensitive Genes	ROS-TRRD	Stepanenko I.L.
Macrophage-Expressed Genes	MG-TRRD	Ananko E.A.

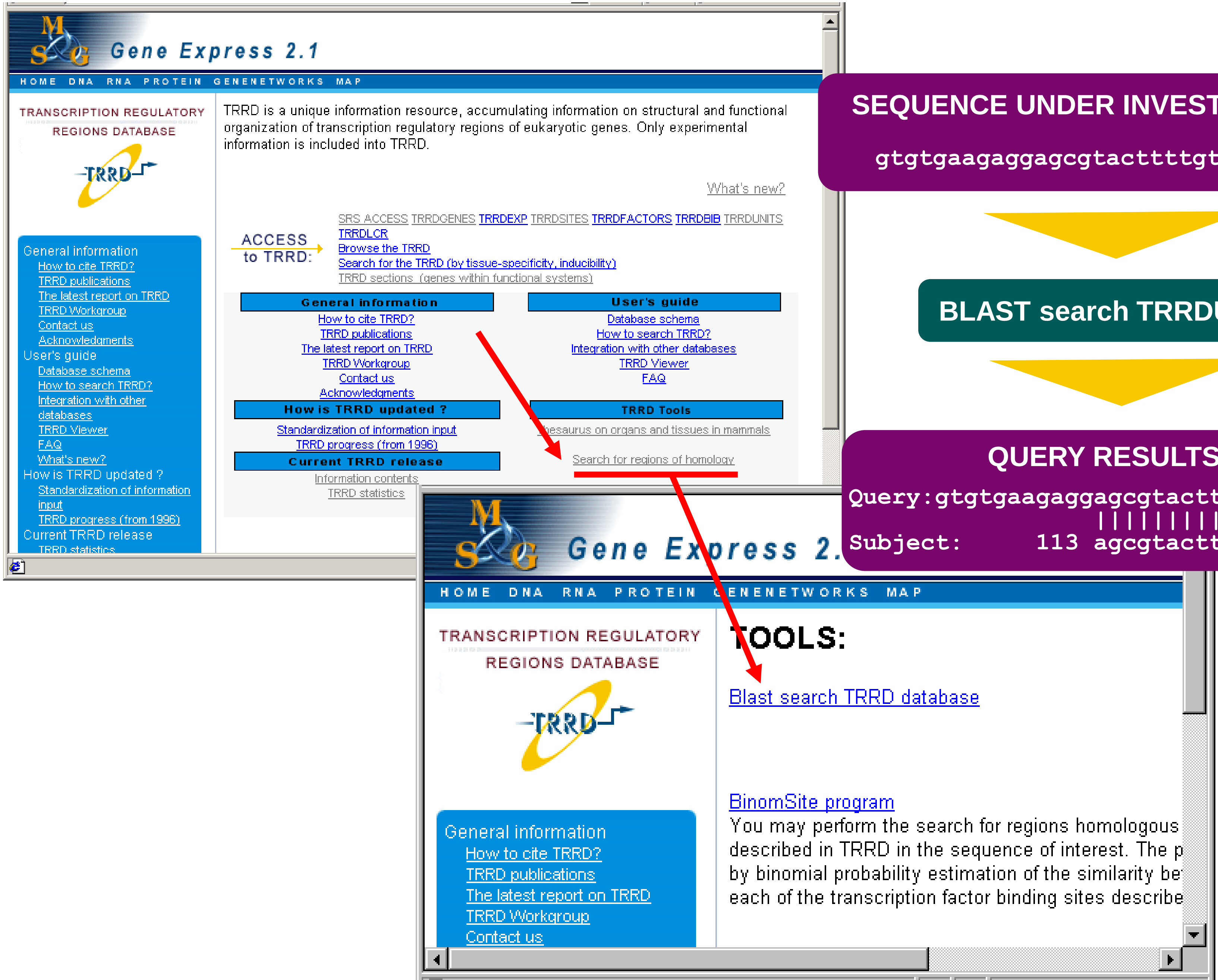
General information
How to cite TRRD?
TRRD publications
The latest report on TRRD
TRRD Workgroup
Contact us
Acknowledgments
User's guide
Database schema
How to search TRRD?
Integration with other databases
TRRD Viewer
FAQ
What's new?
How is TRRD updated ?
Standardization of information input
TRRD progress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

Тематические секции TRRD

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

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

ПОИСК ДАННЫХ в TRRD: BLAST



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 TRRDLCR
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)

TRRD Tools
Thesaurus on organs and tissues in mammals
Search for regions of homology

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?
How is TRRD updated ?
Standardization of information input
TRRD progress (from 1996)
Current TRRD release
TRRD statistics

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 be 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

SEQUENCE UNDER INVESTIGATION:

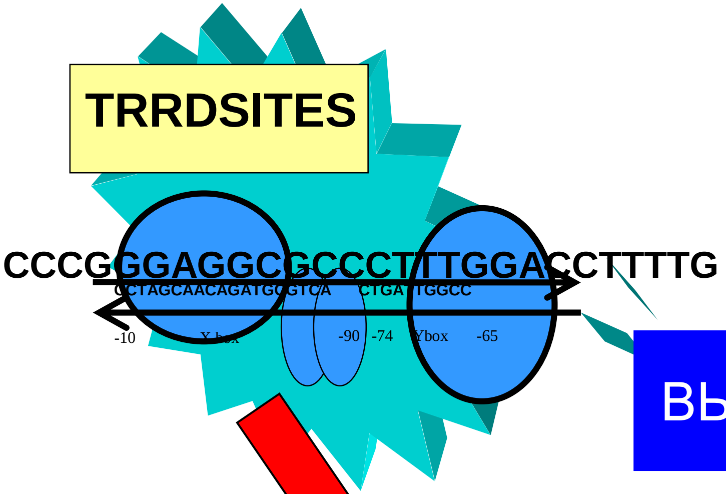
gtgtgaagaggagcgtacttttgtgtgtgaac

BLAST search TRRDUNITS

QUERY RESULTS:

```
Query:gtgtgaagaggagcgtacttttgtgtgtga
                 ||||||||||
Subject:       113 agcgtacttttt 103
```

TRRD – информационная основа для создания выборок

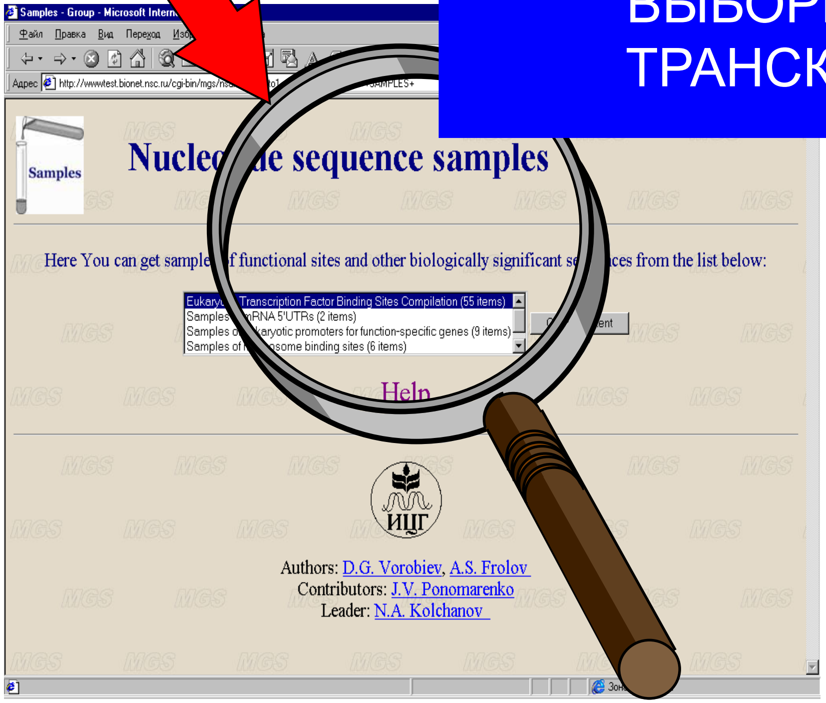


```
ID    es_250_1; DNA
AC    es_250_1
CC    DE   adrenodoxin gene
OS    Homo sapiens (human)
OC    Eukaryota; Metazoa; Chordata; Craniata; Verteb
OC    Mammalia; Eutheria; Primates; Catarrhini; Homi
DR    EMBL; M23665; HSADRDO01; ; join(133..382)
```

ВЫБОРКИ ЭУКАРИОТИЧЕСКИХ ПРОМОТОРОВ

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

ВЫБОРКИ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ



Nucle de sequence samples

Here You can get sample f functional sites and other biologically significant se ces from the list below:

- Eukaryo Transcription Factor Binding Sites Compilation (55 items)
- Samples mRNA 5'UTRs (2 items)
- Samples o karyotic promoters for function-specific genes (9 items)
- Samples of osome binding sites (6 items)

Authors: D.G. Vorobiev, A.S. Frolov
Contributors: J.V. Ponomarenko
Leader: N.A. Kolchanov

```
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata
Mammalia; Eutheria; Primates; Catarrhi
EMBL; M23665; HSADRDO01; ; join(133..3
ST (EMBL/GENBANK) 333
TRRDGENES; A00860; Hs:ADX; 4.2;
{0,0} [1;250]; EXP
ctttcaaaat attttgtttc tgcacggcaa cttca
tccagcttac aacggaacct ggagggttgg taaag
cccgccccat gggaccgggc ggcgtgggcg tgaga
gctctgcttg ccaatgtctt tataggtcac ccgga
cggcgcggtg cttccagcag ggtctctccg ccact
```

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 ?);

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.

Нет 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.

Нет 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.

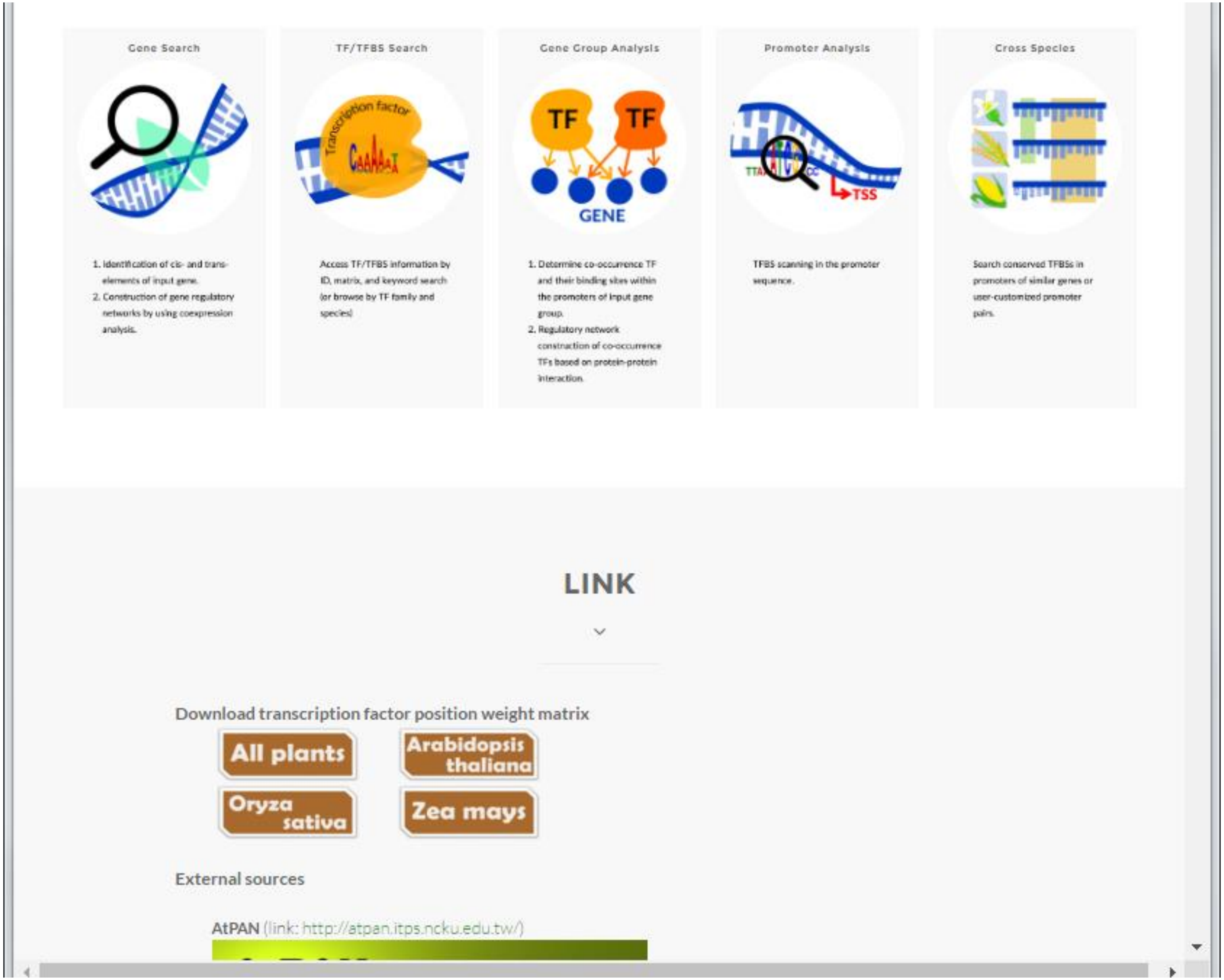
Нет Triticum

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/)



Gene Search

1. Identification of cis- and trans-elements of input gene.
2. Construction of gene regulatory networks by using coexpression analysis.

TF/TFBS Search

Access TF/TFBS information by ID, matrix, and keyword search (or browse by TF family and species)

Gene Group Analysis

1. Determine co-occurrence TF and their binding sites within the promoters of input gene group.
2. Regulatory network construction of co-occurrence TFs based on protein-protein interaction.

Promoter Analysis

TFBS scanning in the promoter sequence.

Cross Species

Search conserved TFBSs in promoters of similar genes or user-customized promoter pairs.

Download transcription factor position weight matrix

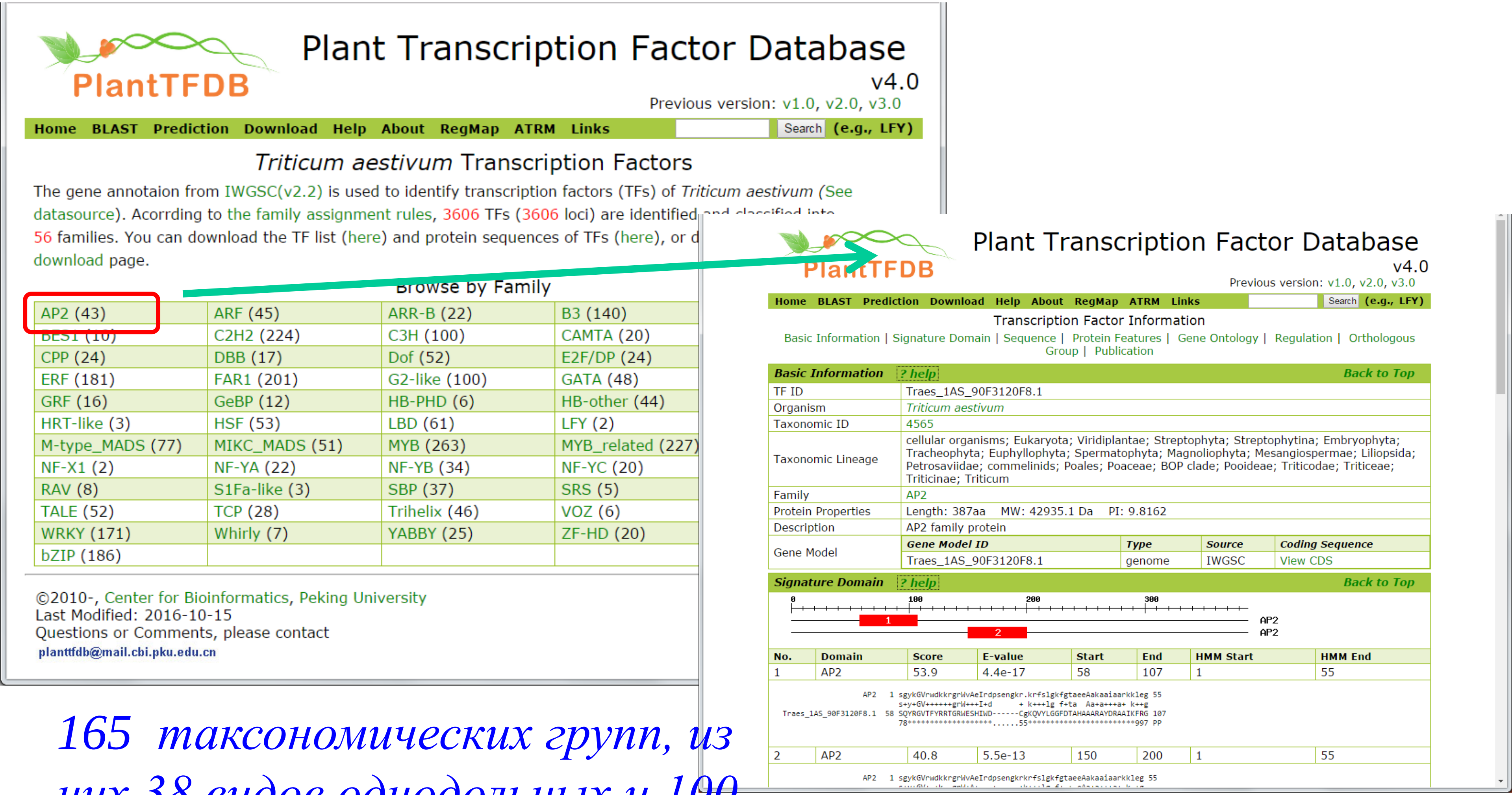
External sources

AtPAN (link: http://atpan.itps.ncku.edu.tw/)

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

PlantTFDB



Plant Transcription Factor Database
v4.0
Previous version: v1.0, v2.0, v3.0

Triticum aestivum Transcription Factors

The gene annotaion from IWGSC(v2.2) is used to identify transcription factors (TFs) of *Triticum aestivum* (See datasource). Acorrding 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 d download page.

Browse by Family

AP2 (43)	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
planttfdb@mail.cbi.pku.edu.cn

Plant Transcription Factor Database
v4.0
Previous version: v1.0, v2.0, v3.0

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	*Triticum aestivum*
Taxonomic ID	4565
Taxonomic Lineage	cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliophyta; Mesangiospermae; Liliopsida; Petrosaviidae; commelinids; Poales; Poaceae; BOP clade; Pooideae; Triticodae; Triticeae; Triticinae; Triticum
Family	AP2
Protein Properties	Length: 387aa MW: 42935.1 Da PI: 9.8162
Description	AP2 family protein
Gene Model	**Gene Model ID**: Traes_1AS_90F3120F8.1; **Type**: genome; **Source**: IWGSC; **Coding Sequence**: View CDS

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 sgykGVrwdkkrgrWvAeIrdpsengkr.krfslgkfgtaeeAakaaiaarkkleg 55
                         s+y+GV++++++grW+++I+d      + k+++lg f+ta  Aa+a+++a+ k++g
Traes_1AS_90F3120F8.1  58 SQYRGVTFYRRTGRWESHIWD------CgKQVYLGGFDTAHAAARAYDRAAIKFRG 107
                         78*******************......55************************997 PP
```

```
                 AP2   1 sgykGVrwdkkrgrWvAeIrdpsengkrkrfslgkfgtaeeAakaaiaarkkleg 55
```

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

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.

Лекция №6

Часть 2. ТРАНСЛЯЦИЯ

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

Регуляция экспрессии генов на уровне трансляции

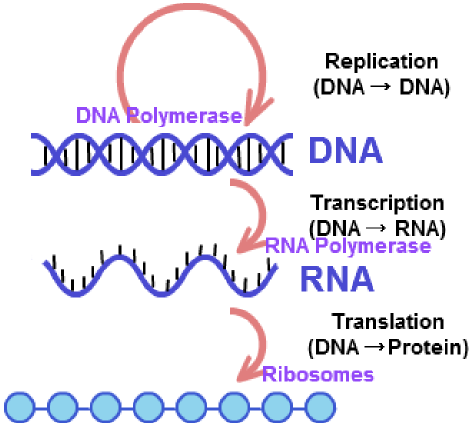
Игнатьева Е.В.
к.б.н., с.н.с. лаб. эволюционной биоинформатики и теоретической генетики

Лекция подготовлена с использованием материалов,
любезно предоставленных д.б.н. А.В. Кочетовым

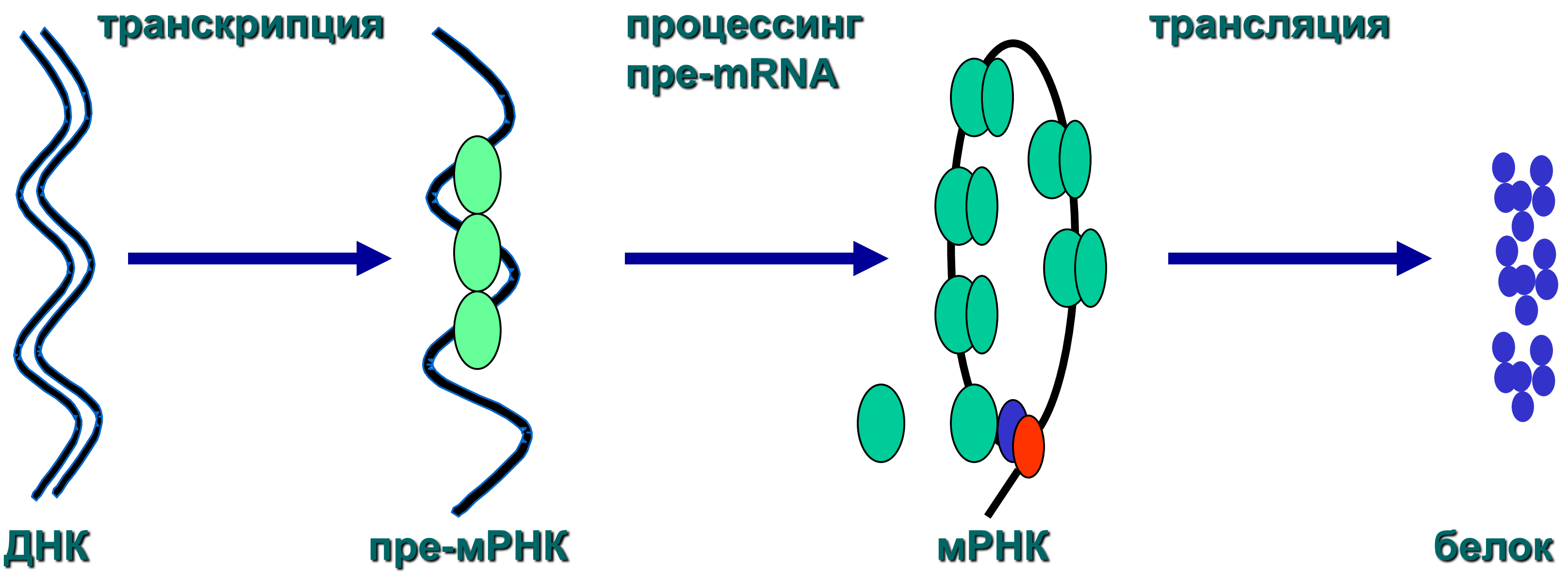
Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук"

Новосибирск - 2016

Трансляция – один из фундаментальных биологических процессов



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



транскрипция

процессинг пре-mRNA

трансляция

ДНК

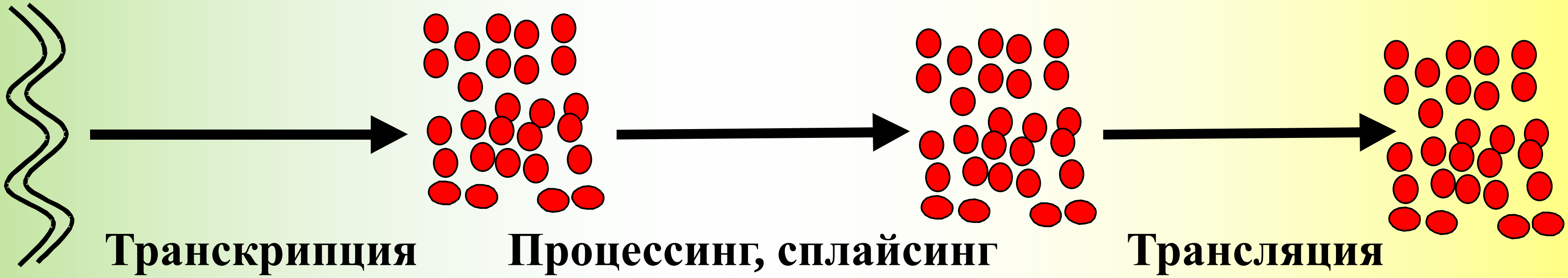
пре-мРНК

мРНК

белок

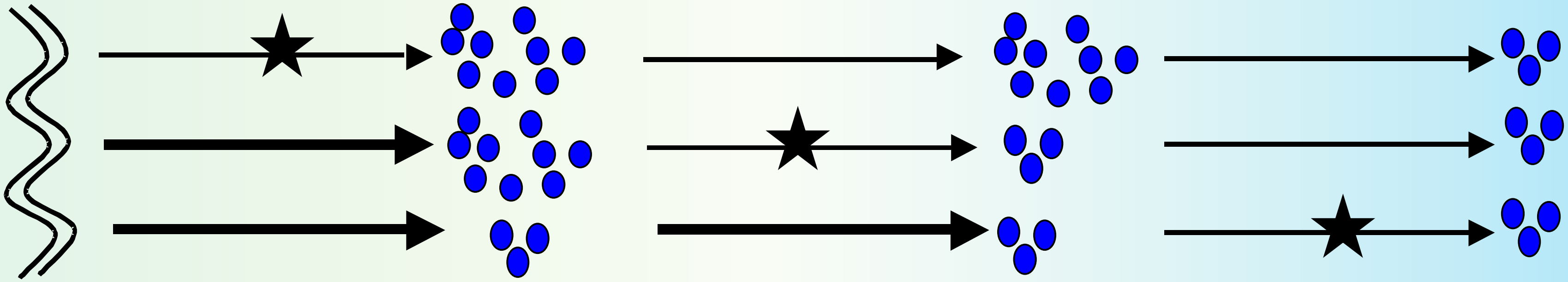
Концепция «лимитирующего звена» в приложении к процессу экспрессии

Высокий уровень экспрессии: высокоэффективны все стадии



Транскрипция **Процессинг, сплайсинг** **Трансляция**

Низкий уровень экспрессии (★- низкоэффективные стадии)



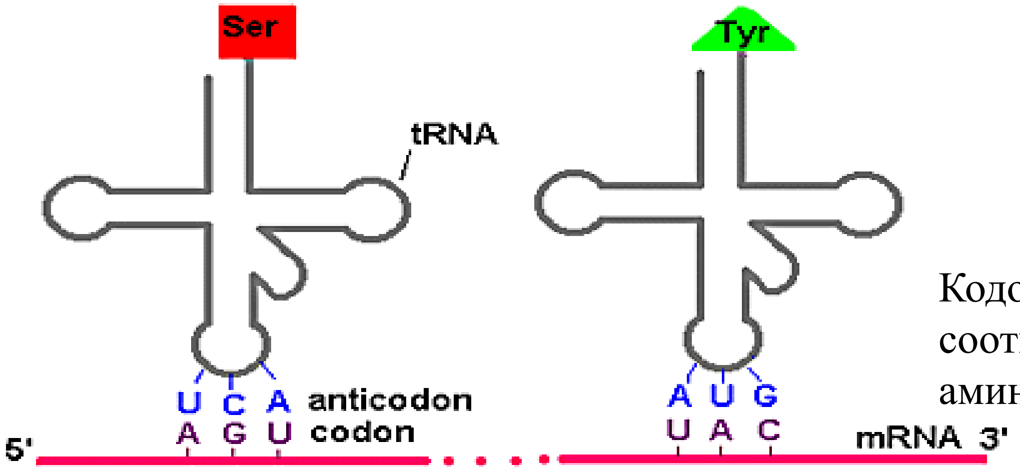
ДНК **пре-мРНК** **мРНК** **Белок**

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.

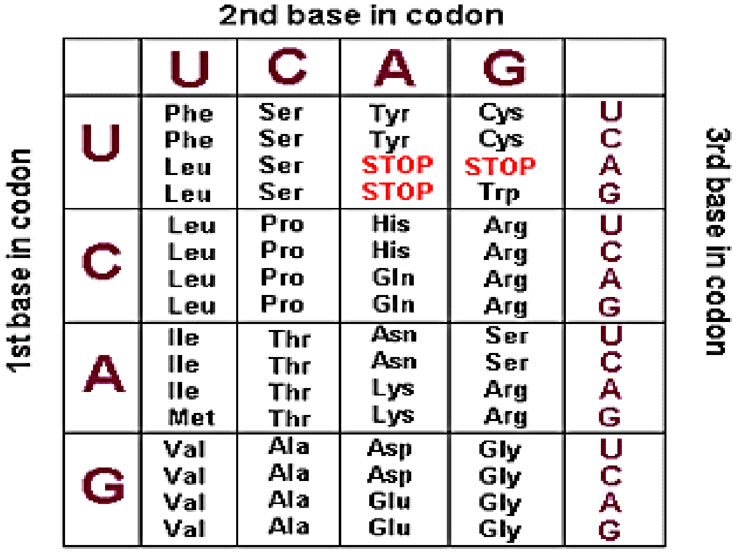
Транспортные РНК (тРНК)

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



Кодон – тринуклеотид, соответствующий либо аминокислоте либо стоп-кодону

Генетический код – соответствие кодонов и аминокислот



2nd base in codon

1st base in codon

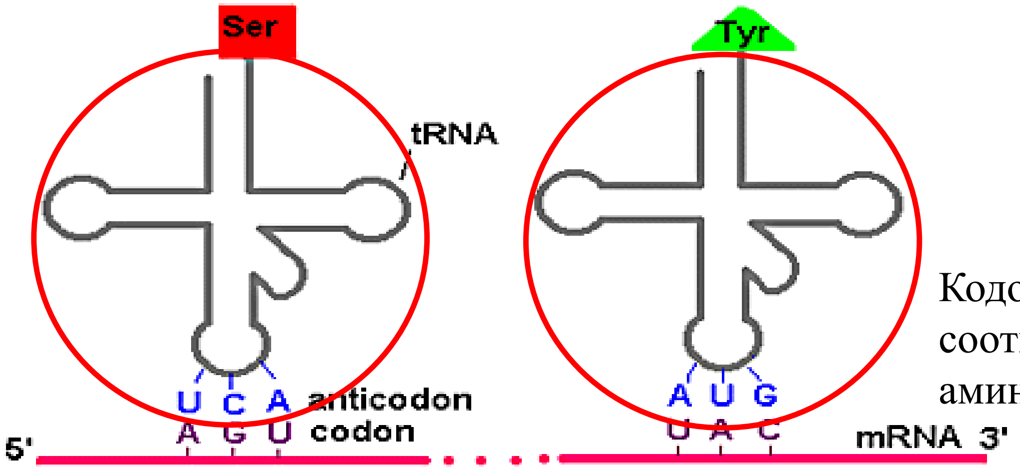
3rd base in codon

	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP 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

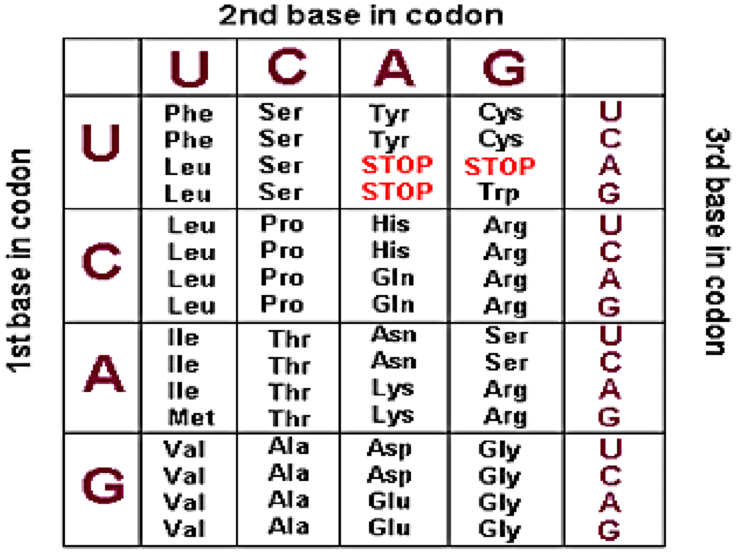
Транспортные РНК (тРНК)

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



Кодон – тринуклеотид, соответствующий либо аминокислоте либо стоп-кодону

Генетический код – соответствие кодонов и аминокислот

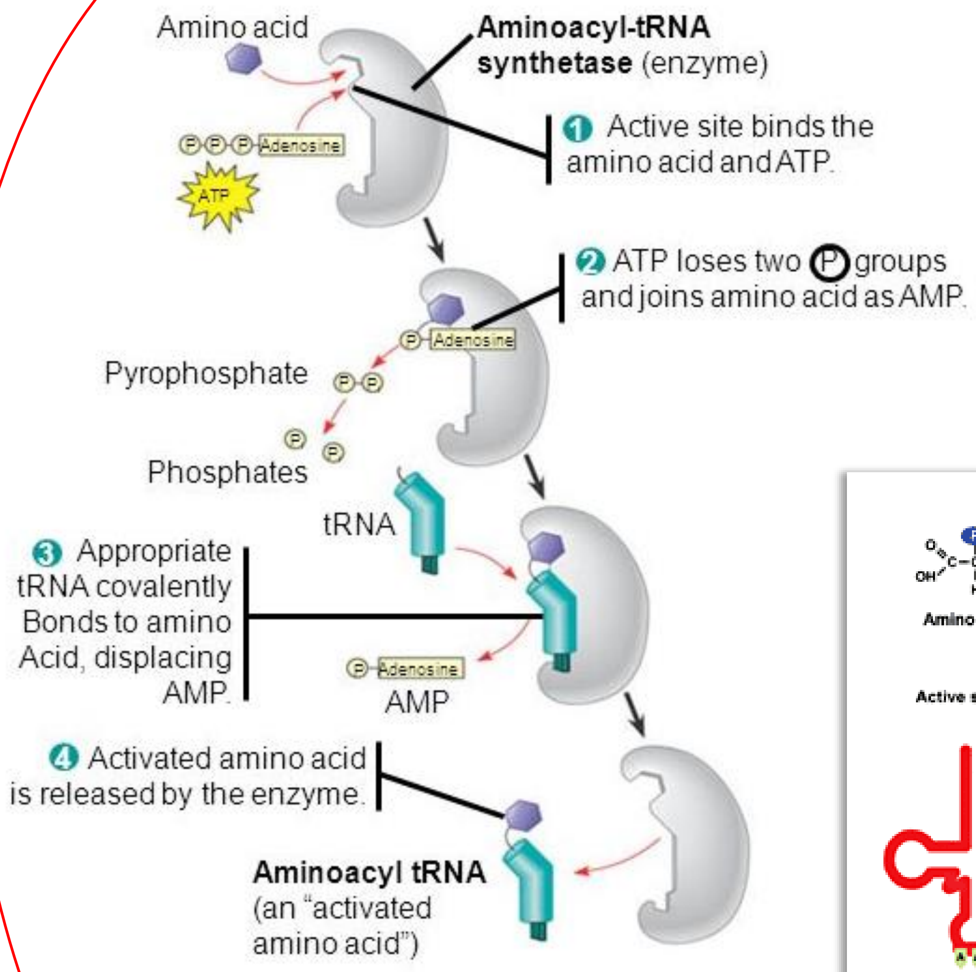


2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

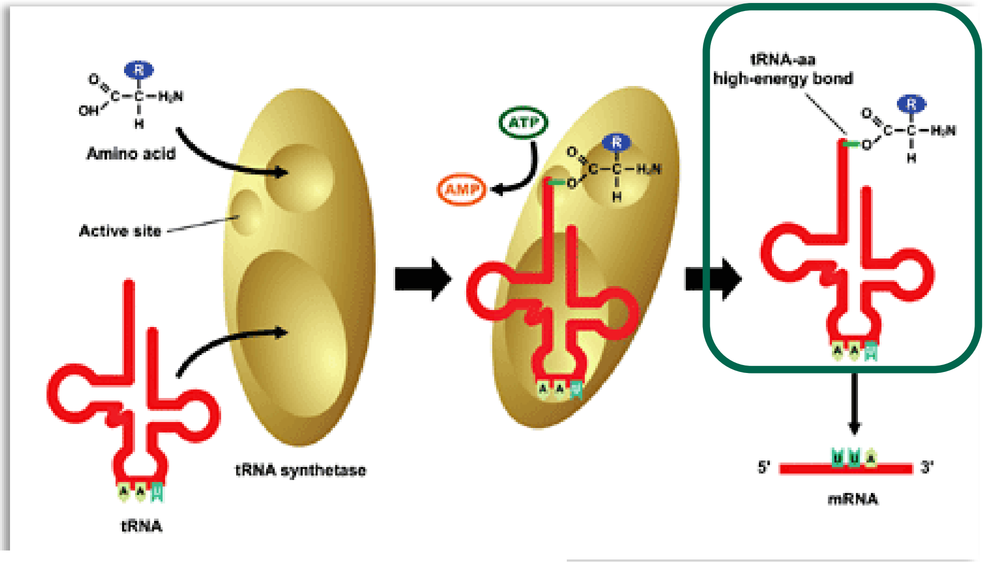
- Число тринуклеотидных комбинаций из 4 нуклеотидов = равно 64 (4^3)
- Три кодона являются нонсенс-кодонами и выполняют функцию терминаторов трансляции
- Каждой из 20 аминокислот соответствует от одного до шести кодонов.
- Кодоны, кодирующие одну аминокислоту, они называются синонимическими

Функция фермента аминоацил-тРНК-синтетазы:



Аминоацил-тРНК

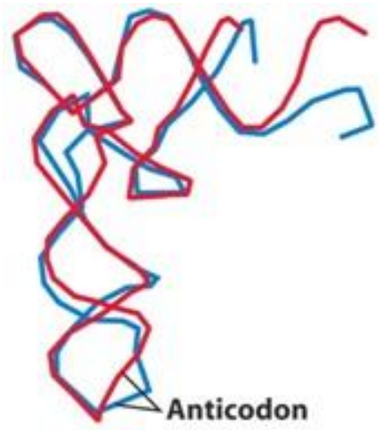
= Аминокислота прикреплена сложноэфирной связью своей карбоксильной группой к 2’ или 3’-гидроксильной группе рибозы 3’-концевого нуклеотида тРНК (это всегда аденин)



Аминокислота + тРНК + АТФ -> Аминоацил-тРНК + АМФ + 2Р

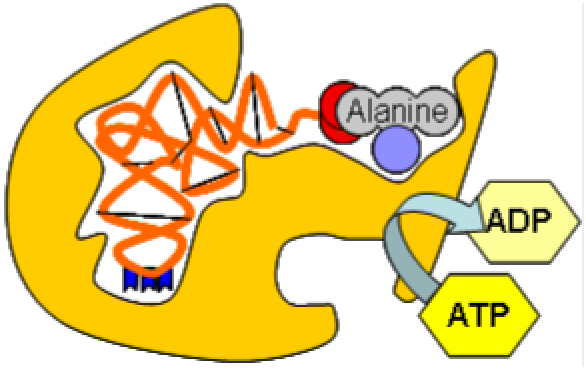
Аминоацил тРНК-синтетазы – распознают «свои» тРНК, соответствующие каждой аминокислоте, и осуществляют связывание аминокислот и тРНК, в силу чего такие аминоацилированные тРНК несут аминокислоты. В этом заключается связь между генетическим кодом – представленным антикодоном тРНК и аминокислотой, с этой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

Аминоацил-тРНК-синтетаза: специфичность



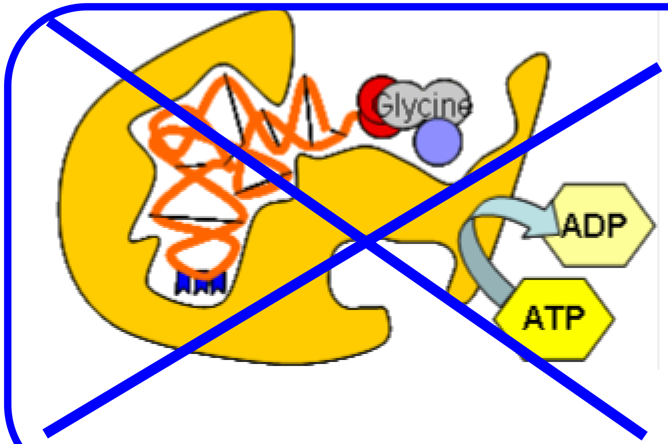
Каждая тРНК имеет уникальную третичную структуру.

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 амнокислота может быть присоединена.**



☺

Аланин подошел !!!

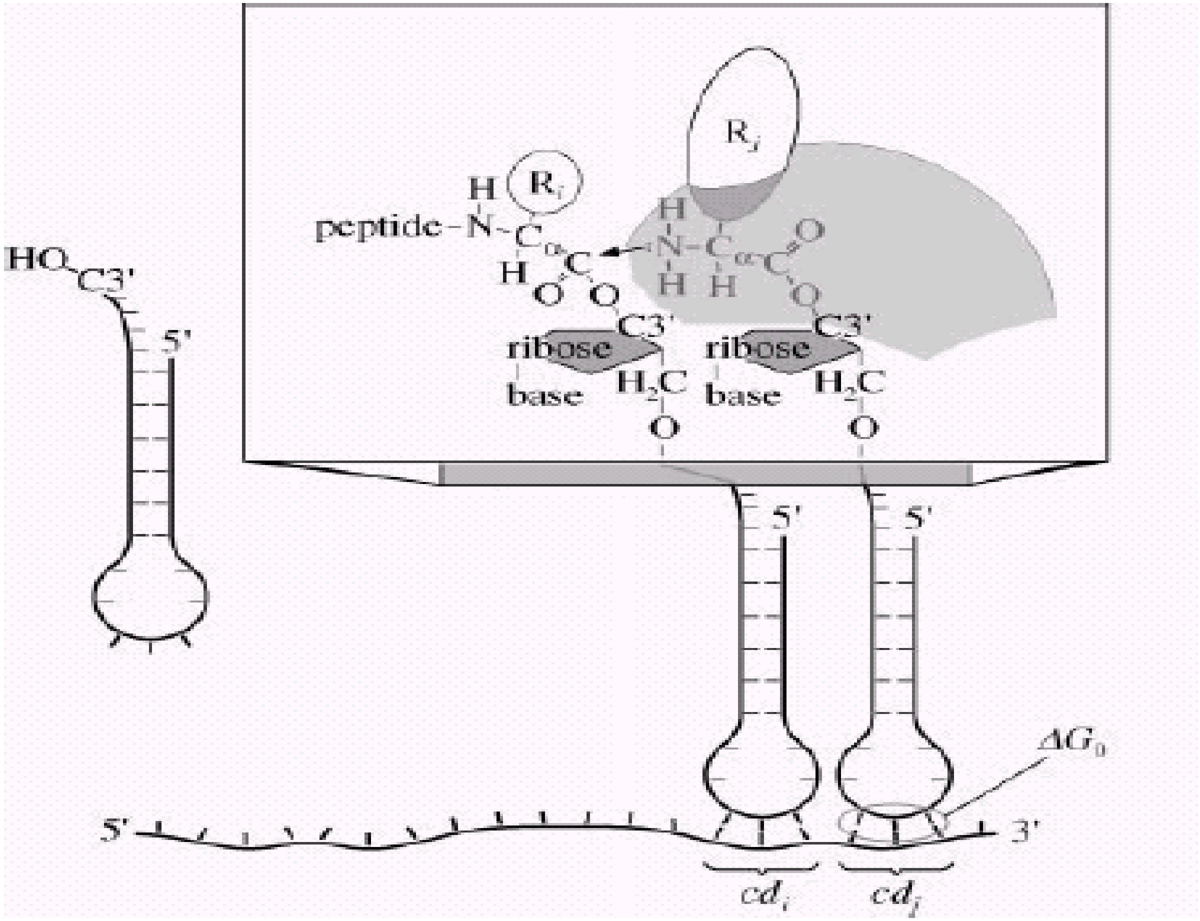


Глицин не подошел !!!

☹

Рибосомы

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



Бактериальная рибосома 70 S

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*

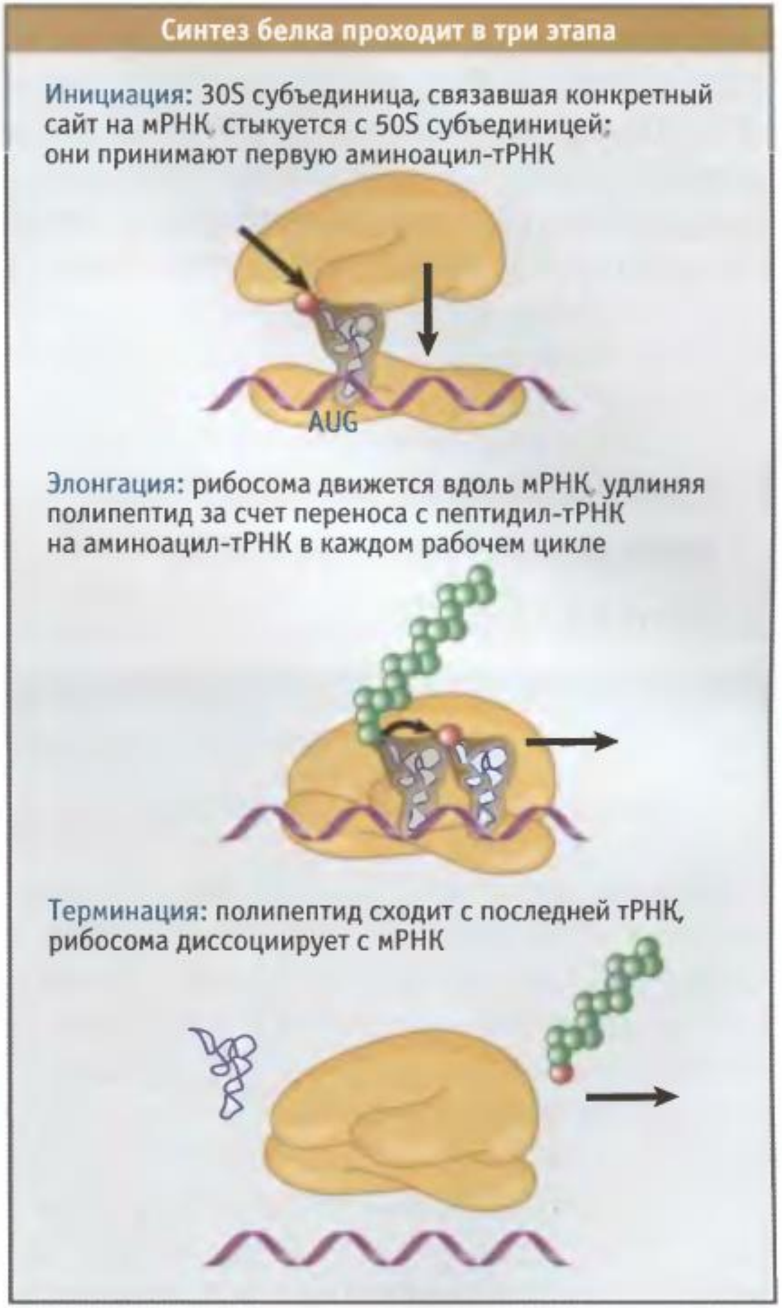
Синтез белка проходит в три этапа: схематическое представление (прокариоты)

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 одни 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



Синтез белка проходит в три этапа

Инициация: 30S субъединица, связавшая конкретный сайт на мРНК, стыкуется с 50S субъединицей: они принимают первую аминоацил-тРНК

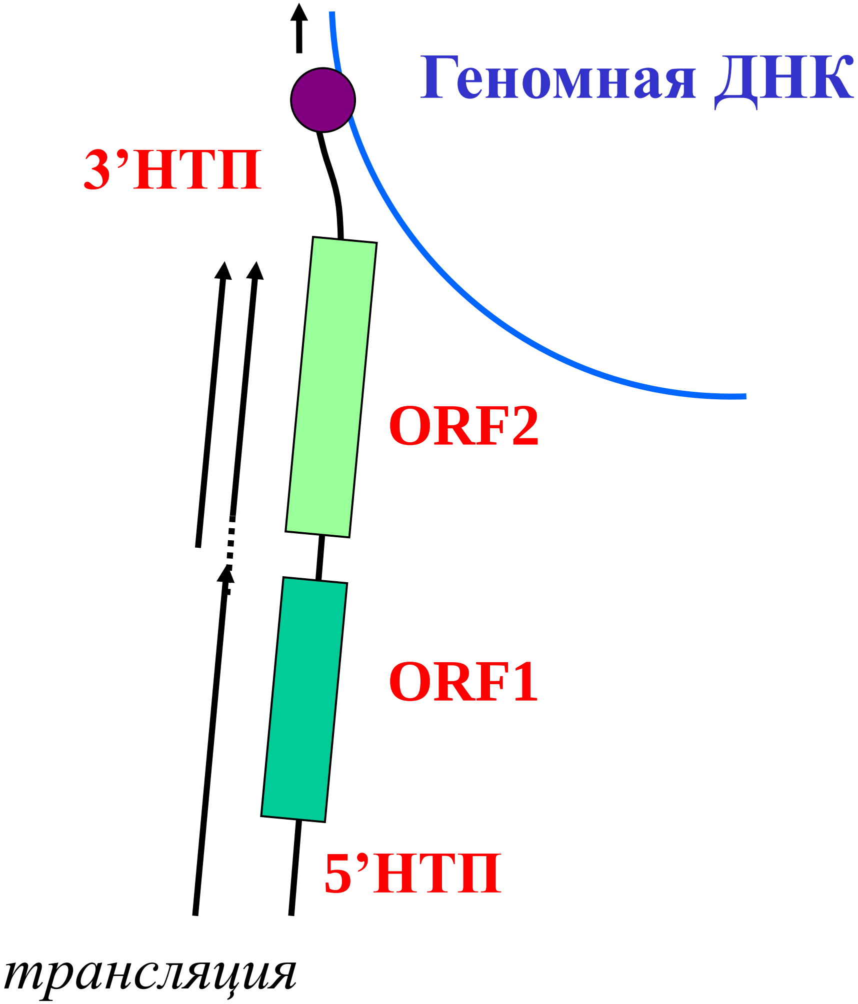
Элонгация: рибосома движется вдоль мРНК, удлиняя полипептид за счет переноса с пептидил-тРНК на аминоацил-тРНК в каждом рабочем цикле

Терминация: полипептид сходит с последней тРНК, рибосома диссоциирует с мРНК

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



Геномная ДНК

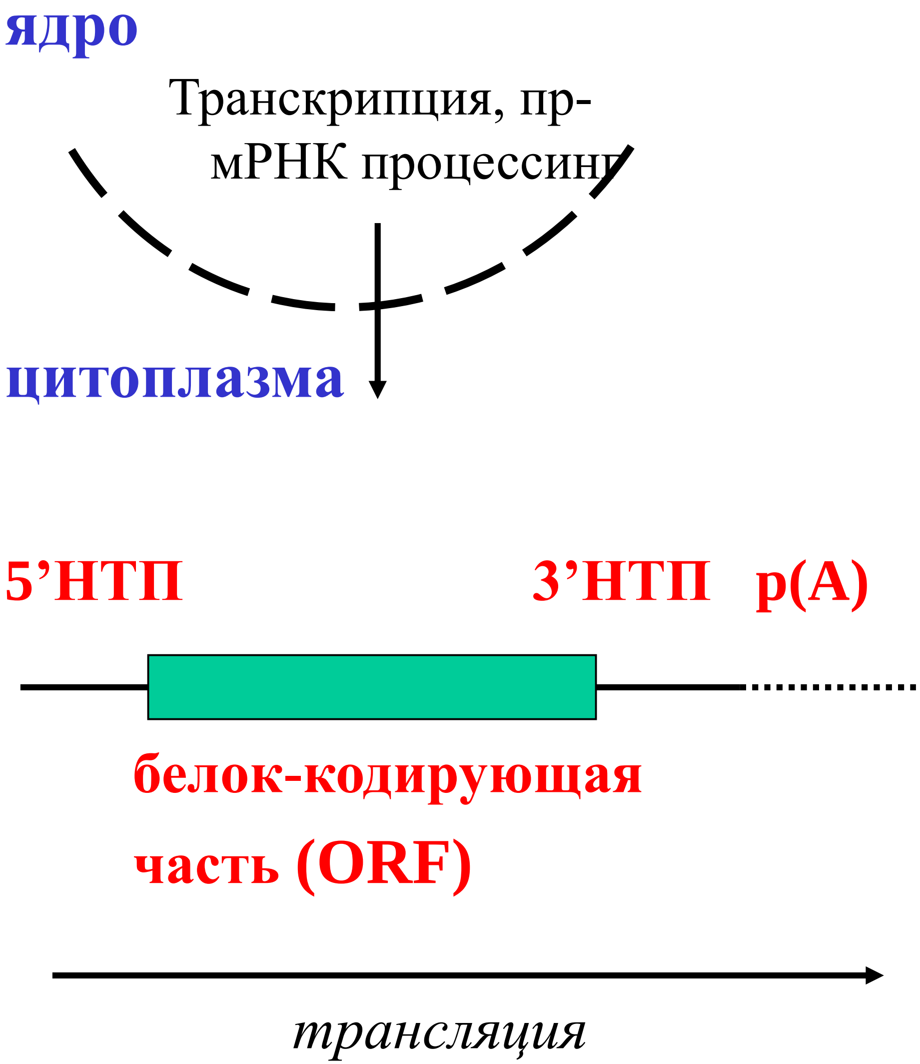
3’НТП

ORF2

ORF1

5’НТП

трансляция



ядро

Транскрипция, пр-мРНК процессинг

цитоплазма

5’НТП

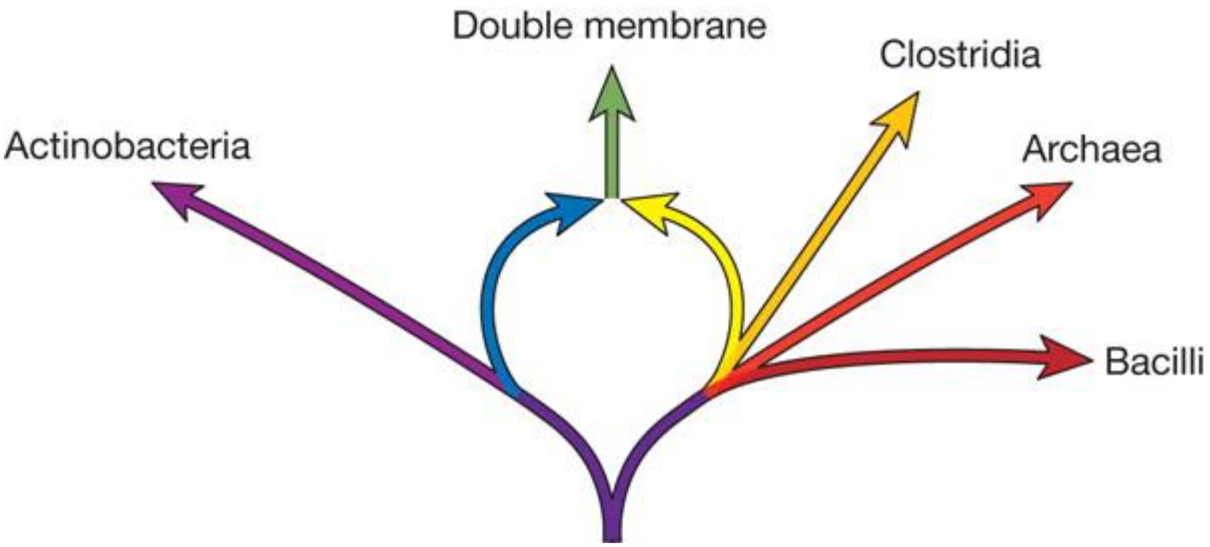
3’НТП

p(A)

белок-кодирующая часть (ORF)

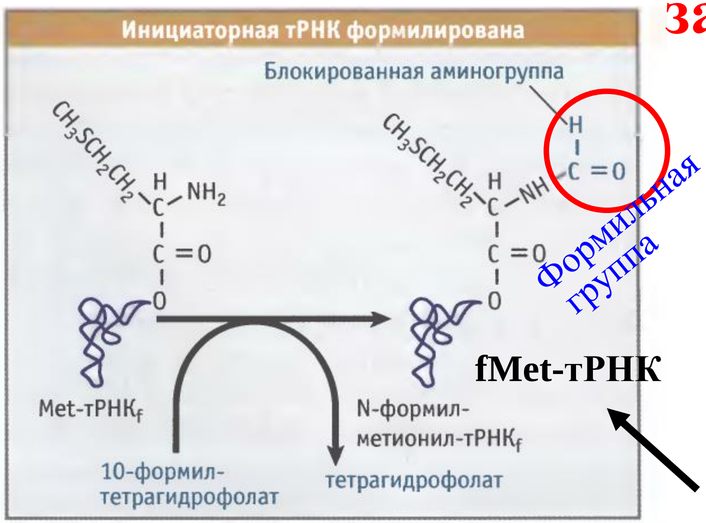
трансляция

Прокариоты



Double membrane
Clostridia
Actinobacteria
Archaea
Bacilli

Прокариоты: особая инициаторная тРНК (fMet-тРНК) закладывает первое звено полипептида

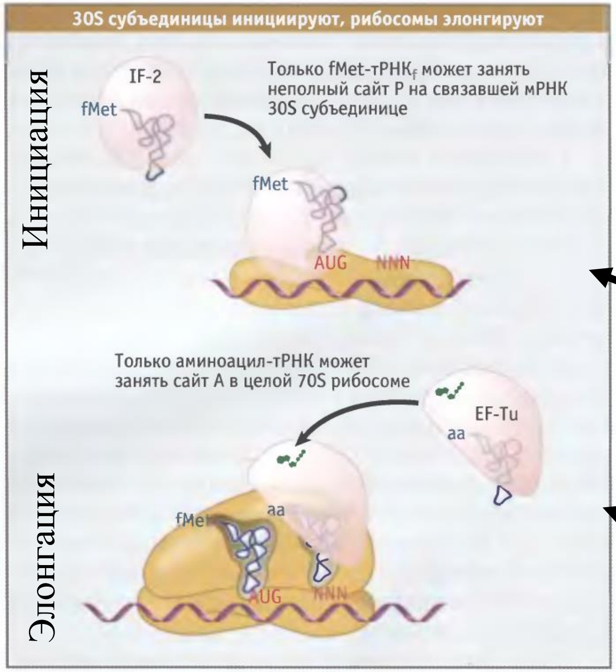


Инициаторная тРНК формилирована

Синтез всех белков начинается с одной и тоже аминокислоты – метионина. Сигналом к началу синтеза полипептидной цепи служит инициирующий 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 аминогруппу. Этот комплекст обознаяается как **fMet-тРНК**



30S субъединицы инициируют, рибосомы элонгируют

Только fMet-тРНК$_f$ может занять неполный сайт P на связавшей мРНК 30S субъединице

Только аминоацил-тРНК может занять сайт A в целой 70S рибосоме

Льюин Б. Гены, 2012 г. С.167, 168

Прокариоты: факторы инициации

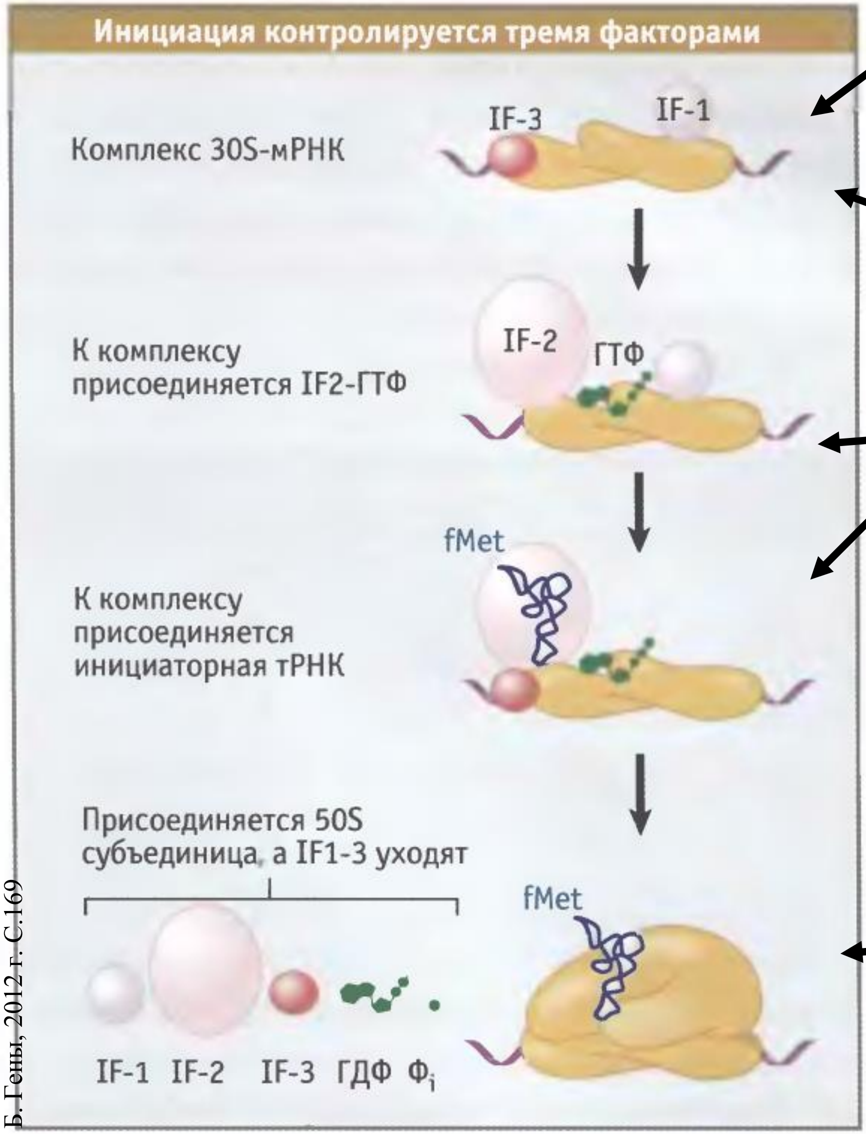
У бактерий есть три фактора инициации: **IF-1, IF-2, IF-3**. Без этих факторов ни мРНК, ни тРНК, не могут вступить в комплекс инициации.

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

Напротив, в ходе элонгации

EF-Tu, который помещает аминоацил-тРНК в сайт A, 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 аминоацил-тРНК занять этот сайт.

IF-2 приносит особую инициаторную тРНК (fMet-тРНК) и контролирует ее вхождение в рибосому. **IF-2 о**бладает 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.

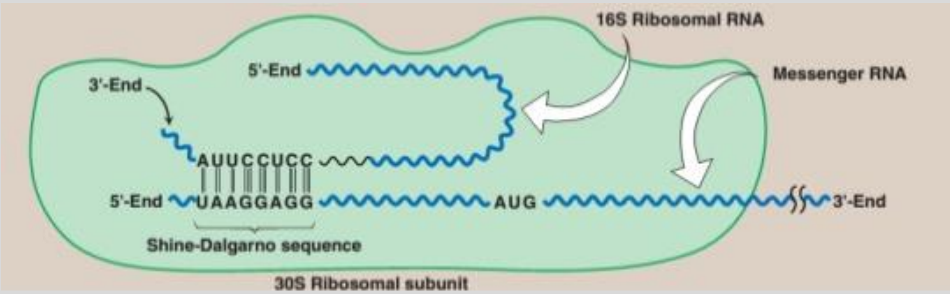
Льюин Б. Гены, 2012 г. С.169

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)$_{8-12}$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.

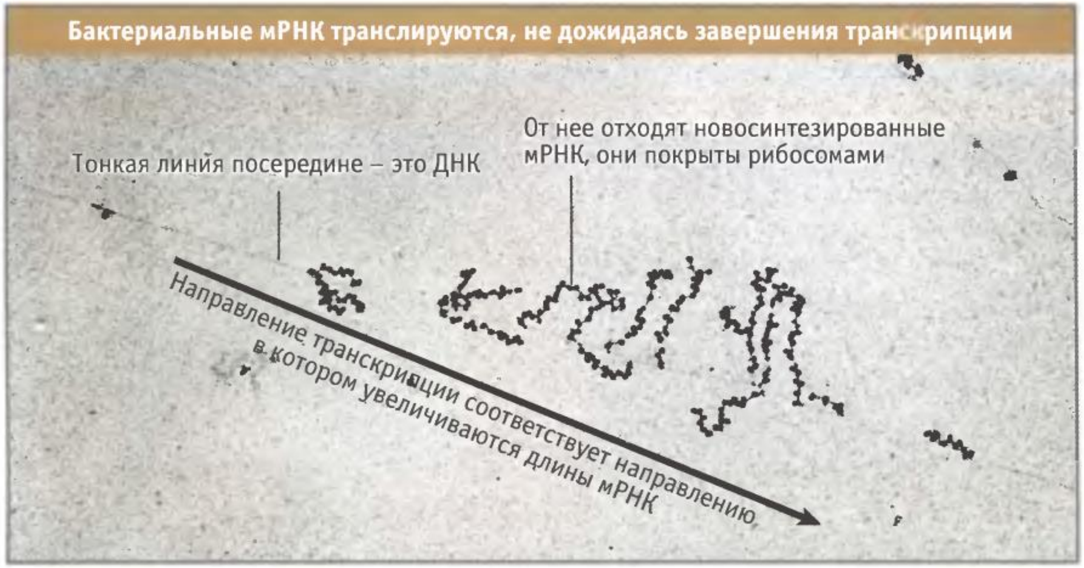


У эукариот нет сайта Ш-Д. У эукариот 40S субъединица рибосомы связывается с 5’-кэпированным концом мРНК, который находится в комплексе с eIF-4 и начинает передвигаться вдоль мРНК до тех пор, пока не встретит AUG кодон.

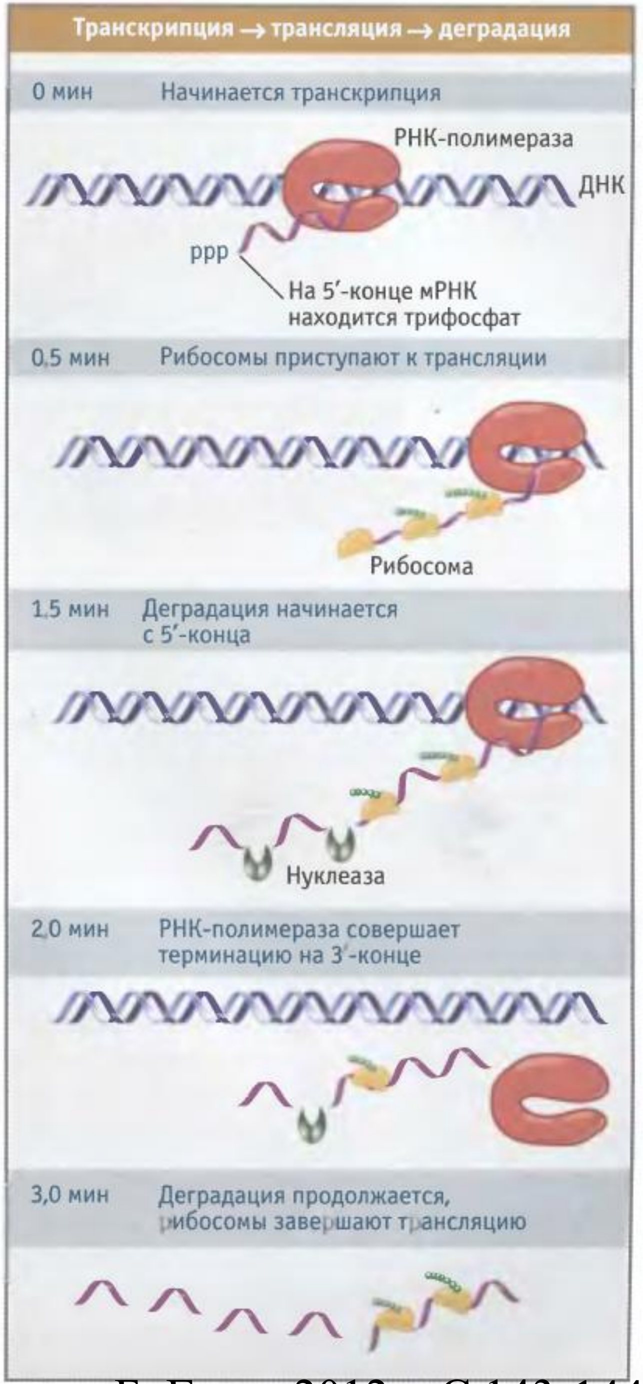
Особенности трансляции у прокариот: транскрипция и трансляция сопряжены

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.

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



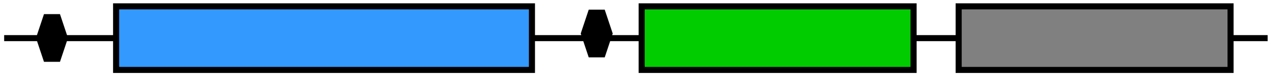
Единицы транскрипции у бактерий можно увидеть



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

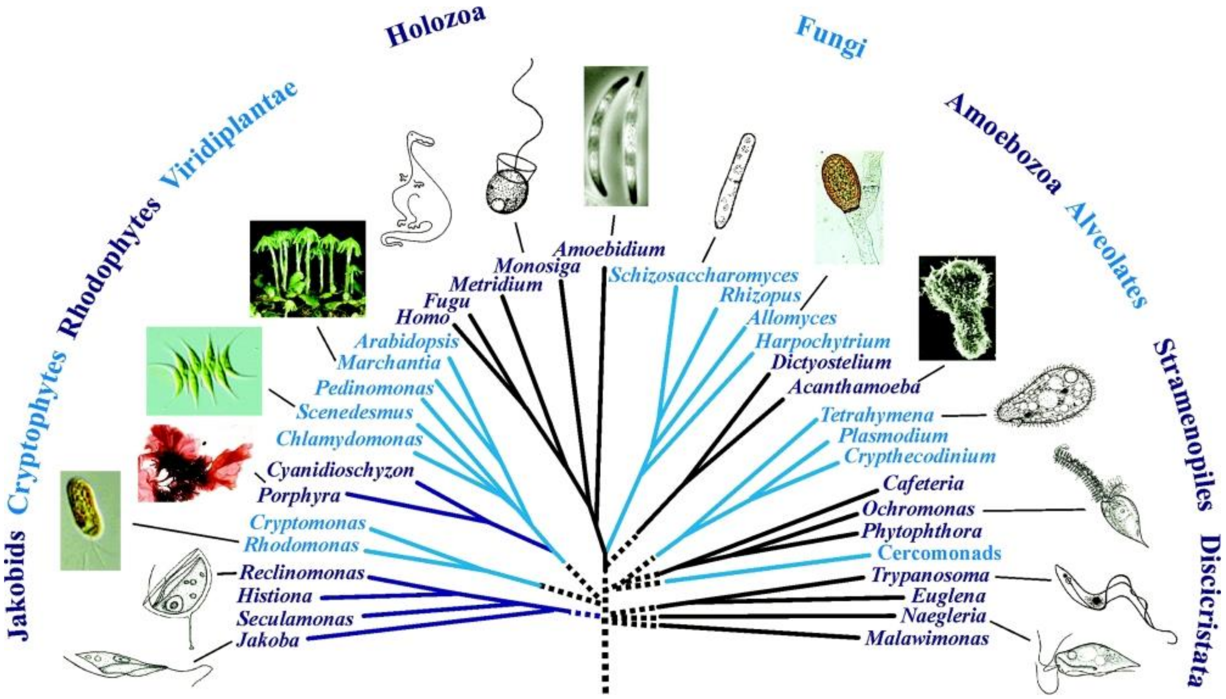


- в составе многих мРНК находятся две и более открытые рамки считывания, кодирующие аминокислотные последовательности разных белков. Рибосома распознает сайт инициации трансляции в начале матрицы (комбинацию сайта Ш-Д и AUG кодона) и транслирует проксимальную белок-кодирующую последовательность, затем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

Эукариоты


Jakobids
Cryptophytes
Rhodophytes
Viridiplantae
Holozoa
Fungi
Amoebozoa
Alveolates
Stramenopiles
Discicristata
Amoebidium
Monosiga
Metridium
Fugu
Homo
Schizosaccharomyces
Rhizopus
Allomyces
Harpochytrium
Dictyostelium
Acanthamoeba
Arabidopsis
Marchantia
Pedinomonas
Scenedesmus
Chlamydomonas
Cyanidioschyzon
Porphyra
Cryptomonas
Rhodomonas
Reclinomonas
Histiona
Seculamonas
Jakoba
Tetrahymena
Plasmodium
Crypthecodinium
Cafeteria
Ochromonas
Phytophthora
Cercomonads
Trypanosoma
Euglena
Naegleria
Malawimonas

Особенности трансляции у эукариот

- Эукариотическая моноцистронная мРНК

- **cap** **AUG** **UGA** **poly(A)**
- транскрипция и трансляция разобщены (ядро – цитоплазма)
- мРНК имеет 5’-кэп
- мРНК имеет поли-А тракт (100-200 оснований поли(А) на 3’-конце)
- Число вспомогательных факторов больше
- Другой способ стыковки малых субъединиц рибосом со своими сайтами связывания в мРНК
- рибосомы связываются с 5’-концом мРНК и движутся вдоль 5’-НТП в поиске подходящего стартового кодона трансляции

Конец 6-ой лекции