Лекция №8(6)

Часть 1.БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ (окончание)

Часть 2. ТРАНСЛЯЦИЯ

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

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 № 8(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 Саудовская Аравия

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:

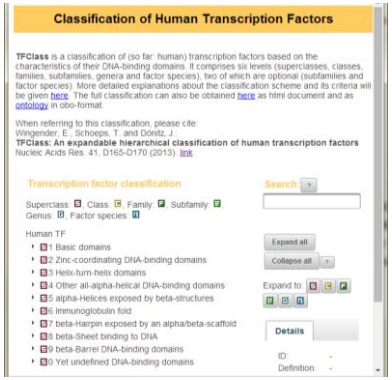
PlantPAN 2.0
PlantDHS
PPdb
PlantTFDB 3.0

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/)

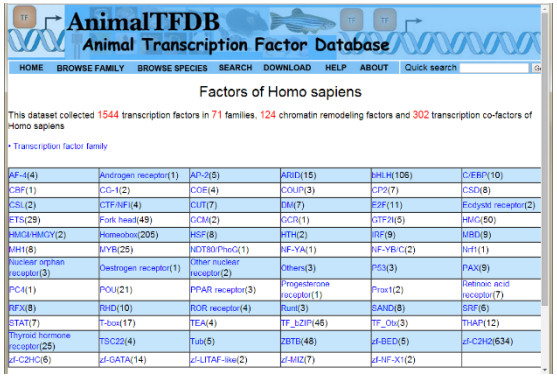


1558 генов,
кодирующих 2904 изоформ белков,
содержащих ДНК-связывающий домен.

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.

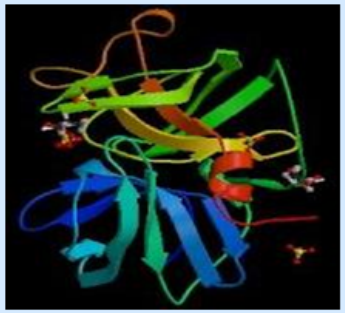
Ресурс AnimalTFDB
(http://www.bioguo.org/AnimalTFDB/)



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

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

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



CREMOFAC

A web-database of

Chromatin Remodeling Factors

Bioinformatics. 2006; 22: 2940-2944

Enter

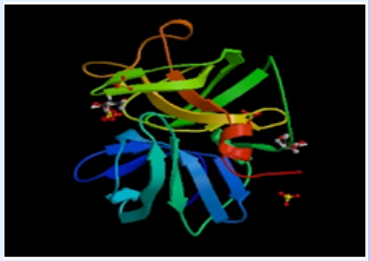
Number of Visits

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



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

CREMOFAC

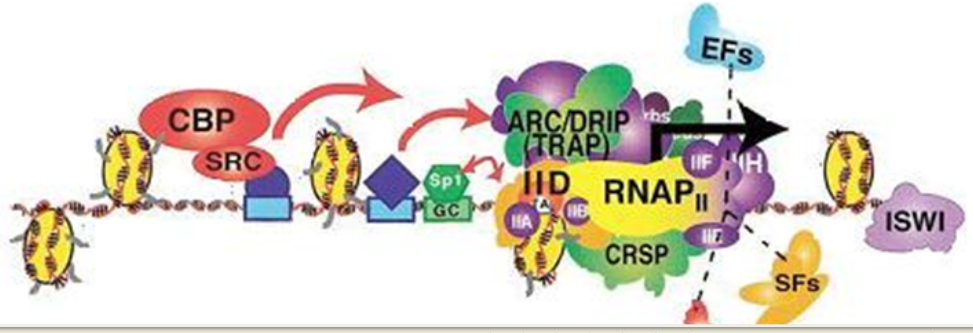
Chromatin Remodeling Factors

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

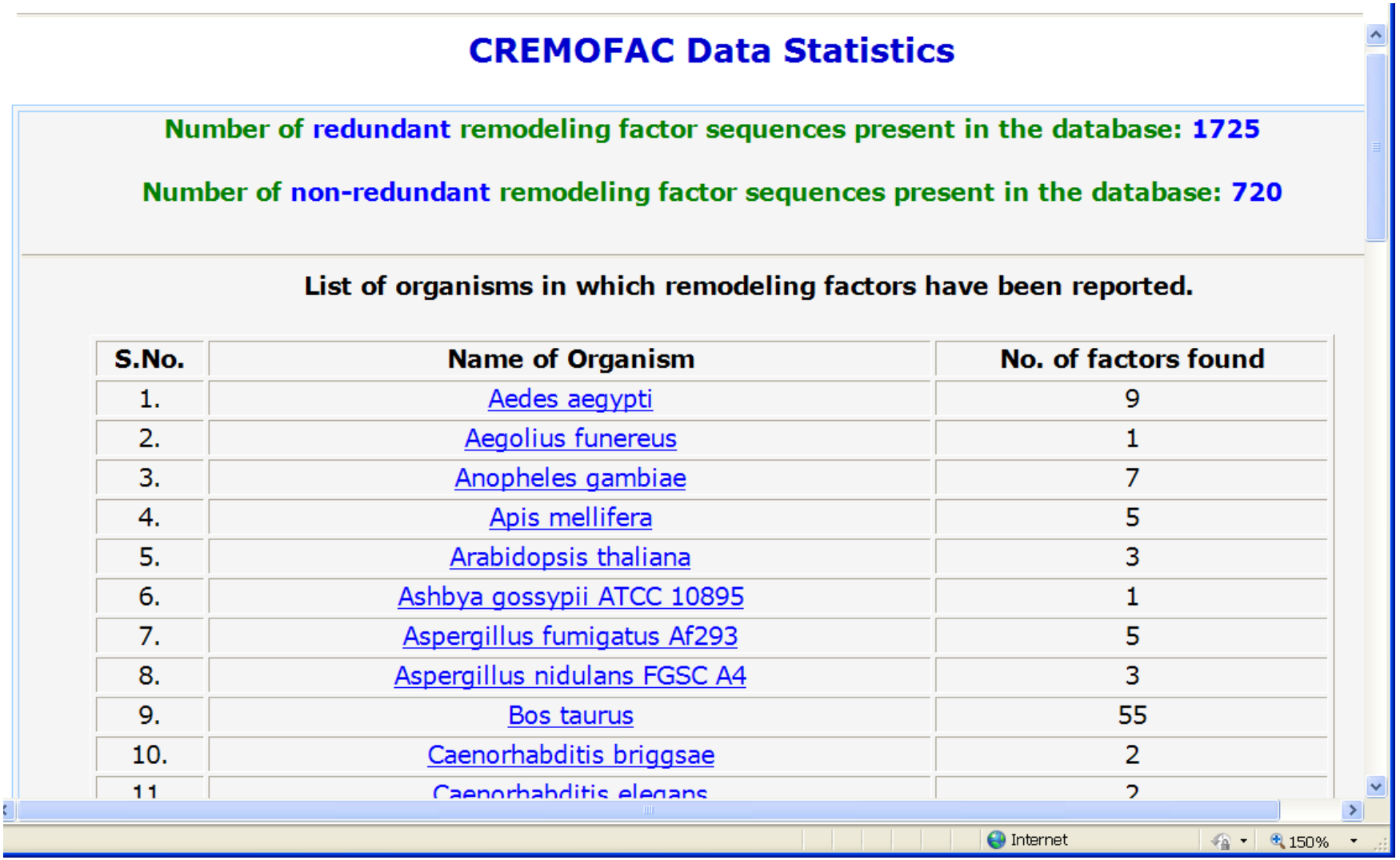
Introduction

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

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



Данные собраны вручную на основе аннотирования научных публикаций



CREMOFAC Data Statistics

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

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

List of organisms in which remodeling factors have been reported.

S.No.	Name of Organism	No. of factors found
1.	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

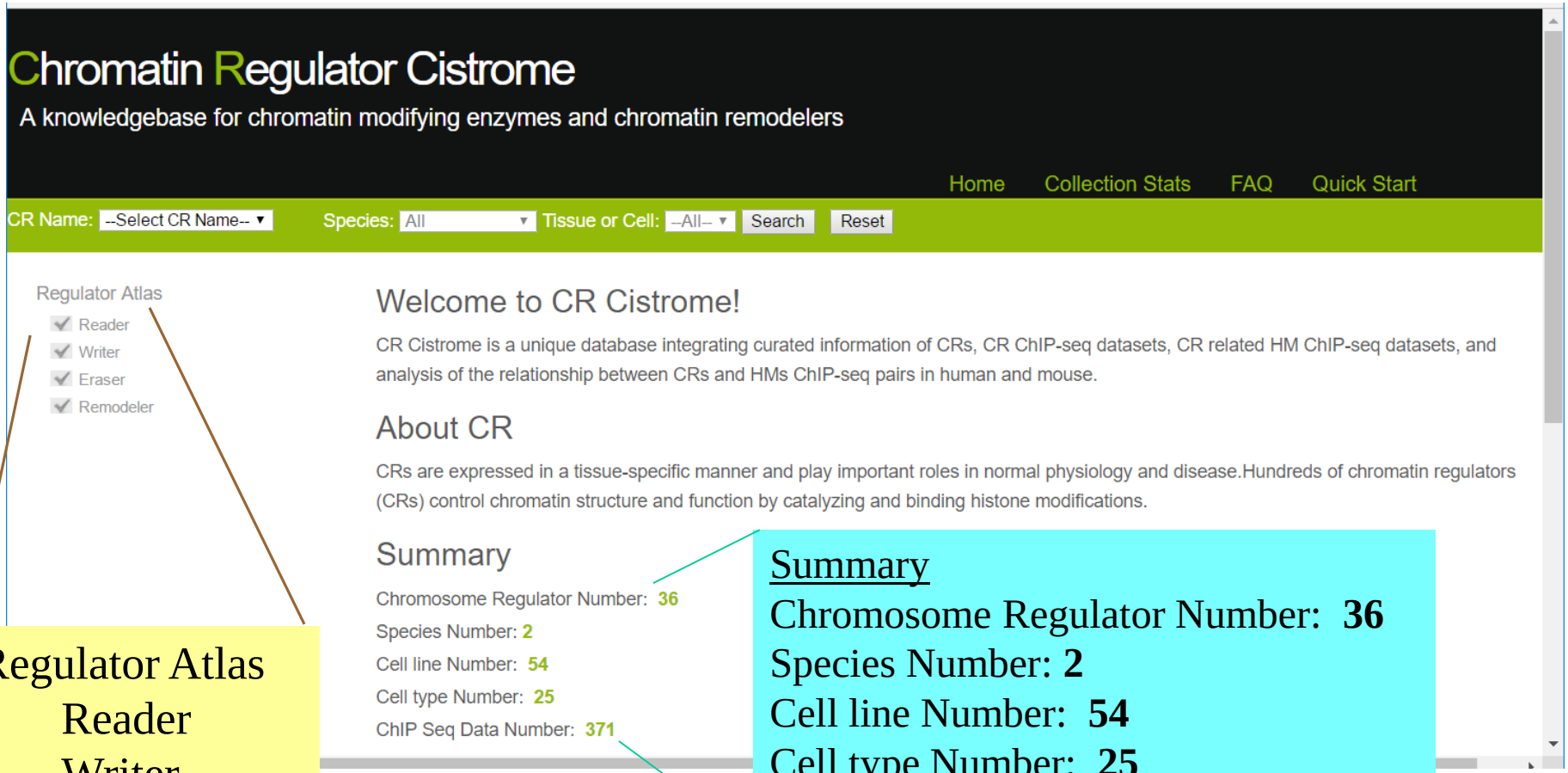
Internet 150%

Всего 49 видов организмов

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

Ресурс CR Cistrome

http://cistrome.org/cr/

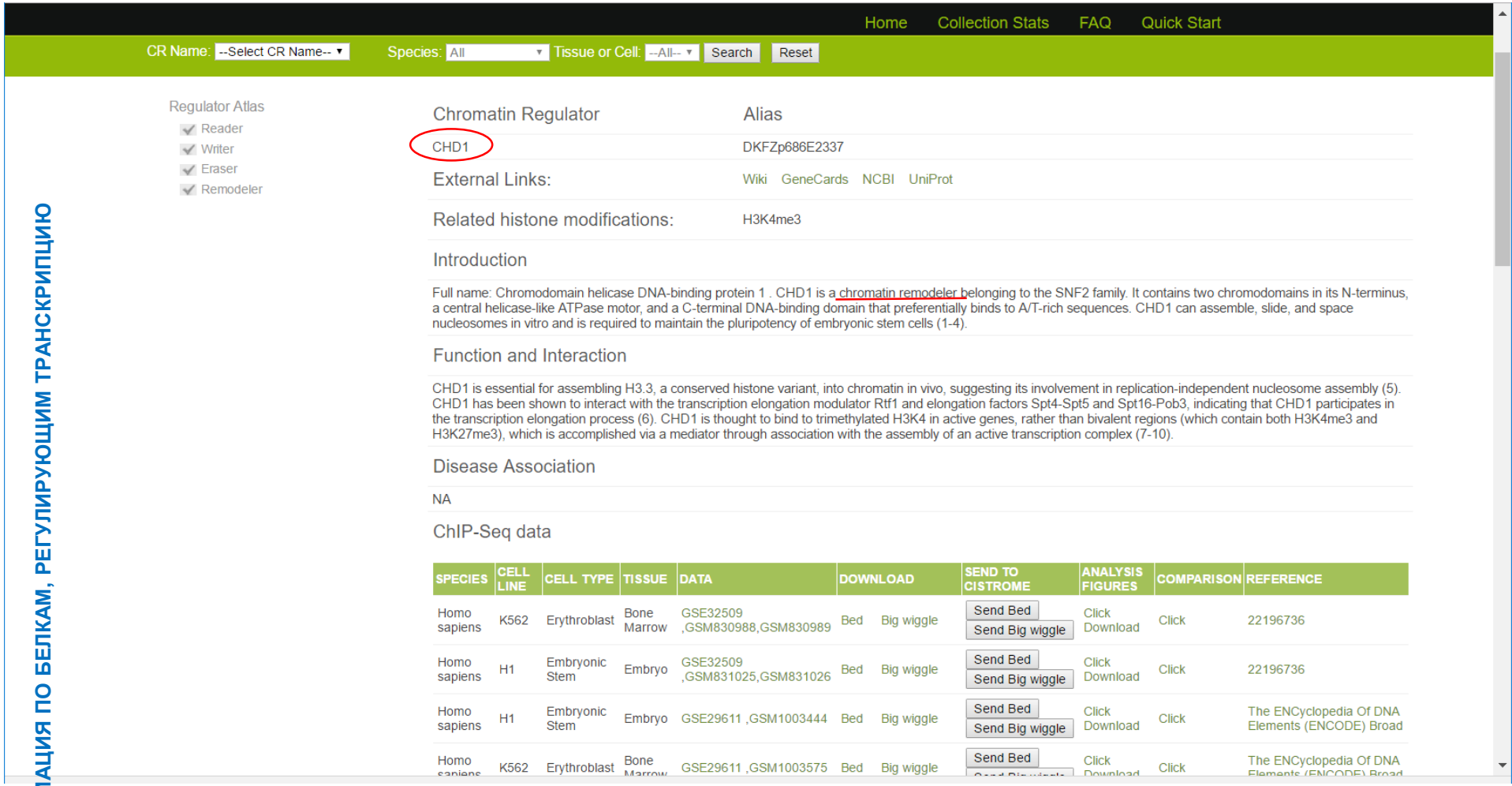


Regulator Atlas
- Reader
- Writer
- Eraser
- Remodeler

Summary
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

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) → 64 гена

99 генов

Ресурс CR Cistrome (Ручная аннотация статей) → 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

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

TRANSFAC

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

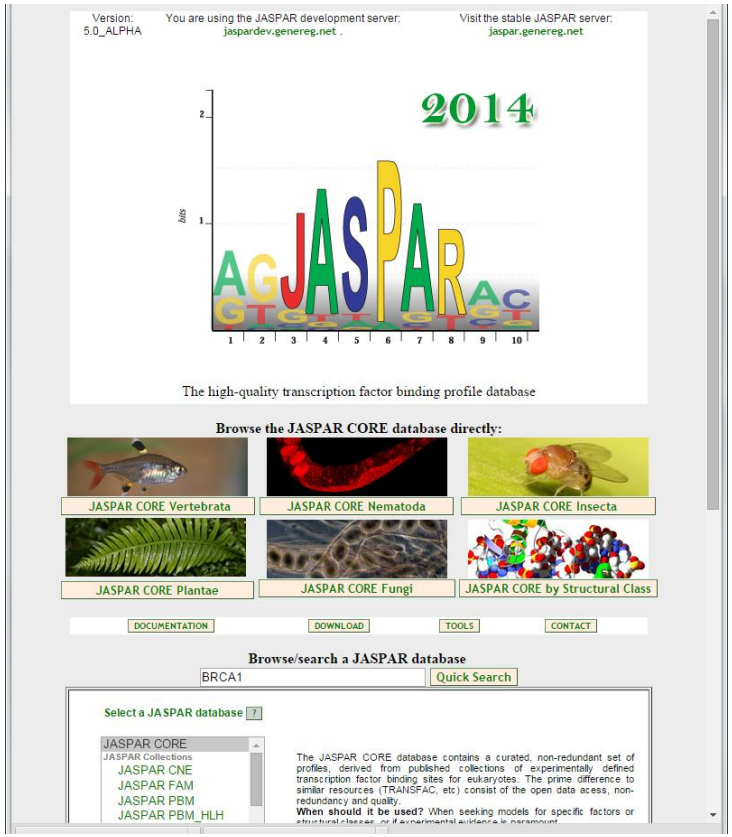
```
AC   M00134
XX
ID   V$HNF4_01
XX
DT   22.05.1995 (created); hiwi.
DT   18.10.1995 (updated); ewi.
CO   Copyright (C), Biobase GmbH.
XX
NA   HNF-4
XX
DE   hepatic nuclear factor 4
XX
BF   T00372 HNF-4alpha1; Species: rat, Rattus norvegicus.
BF   T00373 HNF-4alpha2; Species: human, Homo sapiens.
XX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

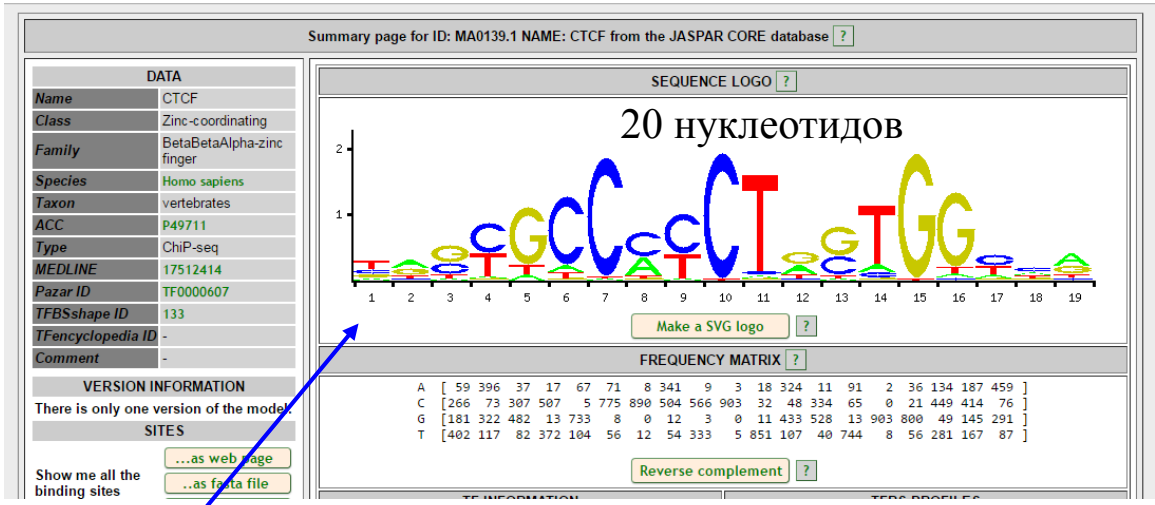
TRANSFAC® Professional is an internationally unique knowledgebase containing published data on eukaryotic transcription factors and their regulated genes, coupled with transcription factor binding site prediction tools for advanced research. The professional version content is more than six years ahead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

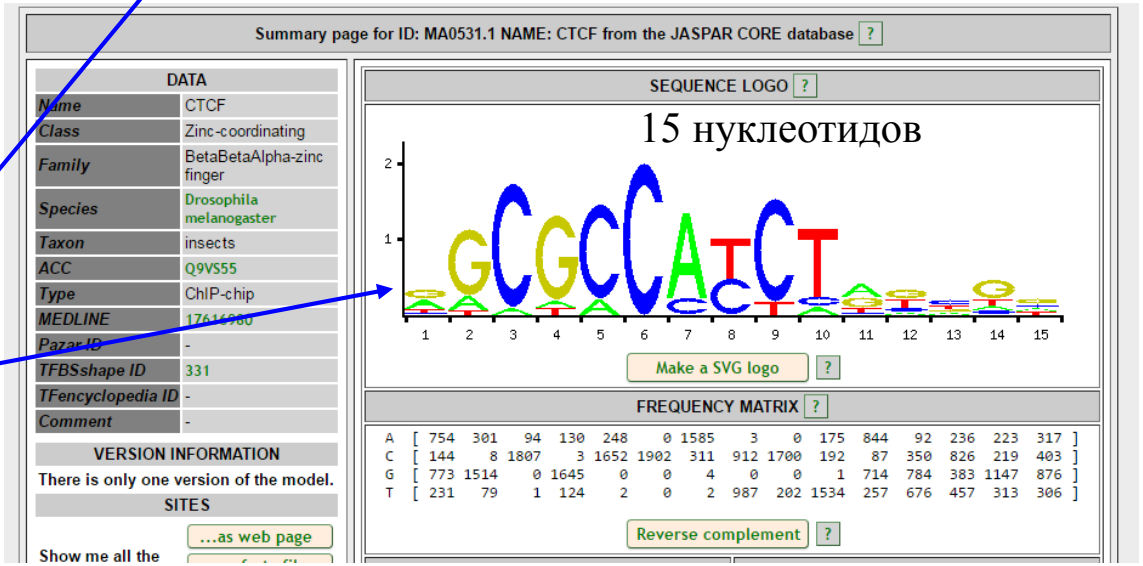
TRANSFAC®	Professional	Public
Data		
Factors	21,215	6,133
miRNAs	894	n/a
DNA sites	38,283	7,915
mRNA sites	15,895	n/a
Factor DNA	51,769	Yes
miRNA-mRNA site links	49,541	n/a
Genes	79,866	2,397
ChIP-chip/Seq Fragments	2,332,432	n/a
Matrices	5,551	398
References	29,681	Flat file only
Promoter Sequences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

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



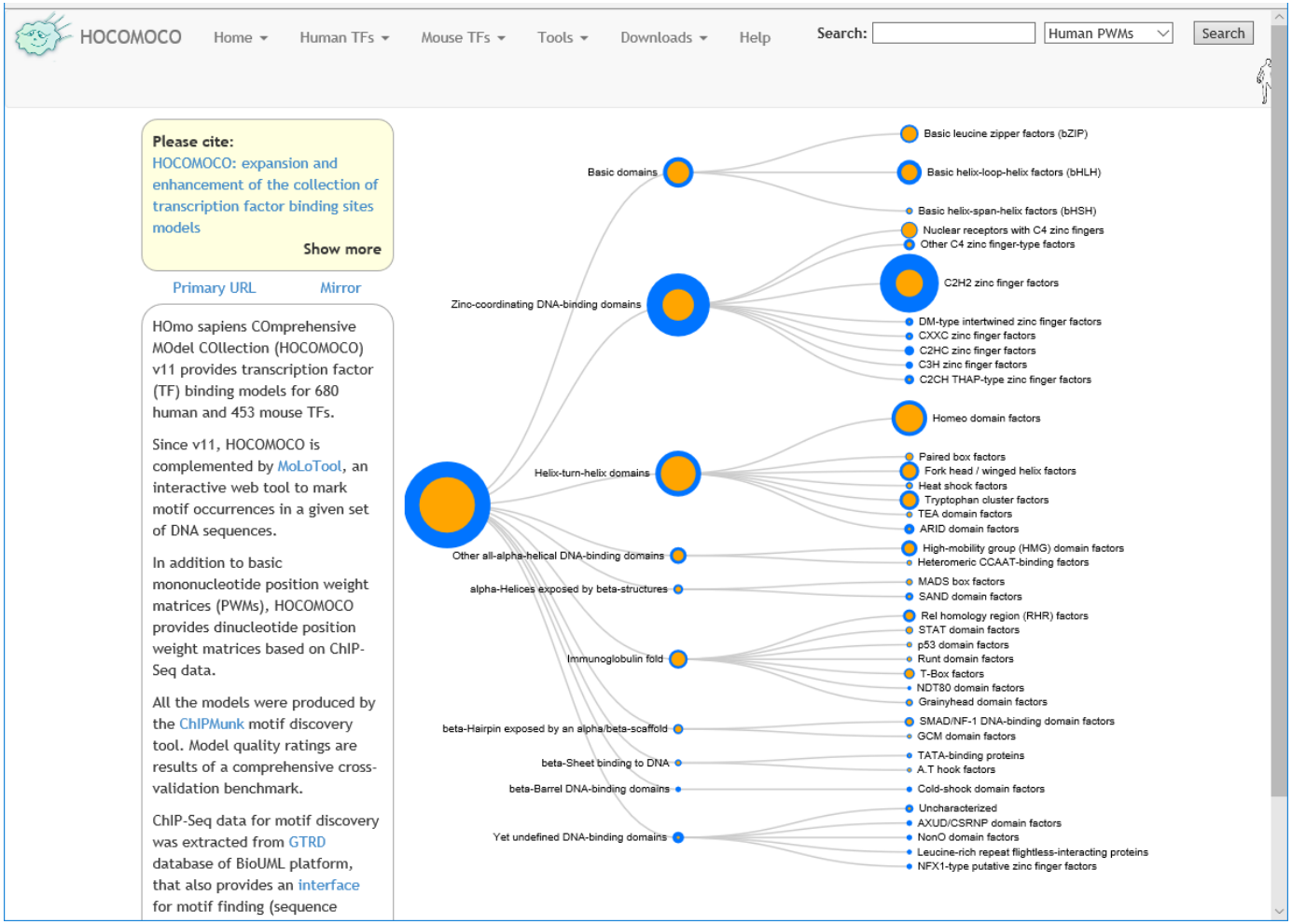
Пример: две матрицы сайтов связывания CTCF





Графики WEB-logo

HOCOMOCO (HOmo sapiens COmprehensive MOdel Collection)
http://autosome.ru/HOCOMOCO/

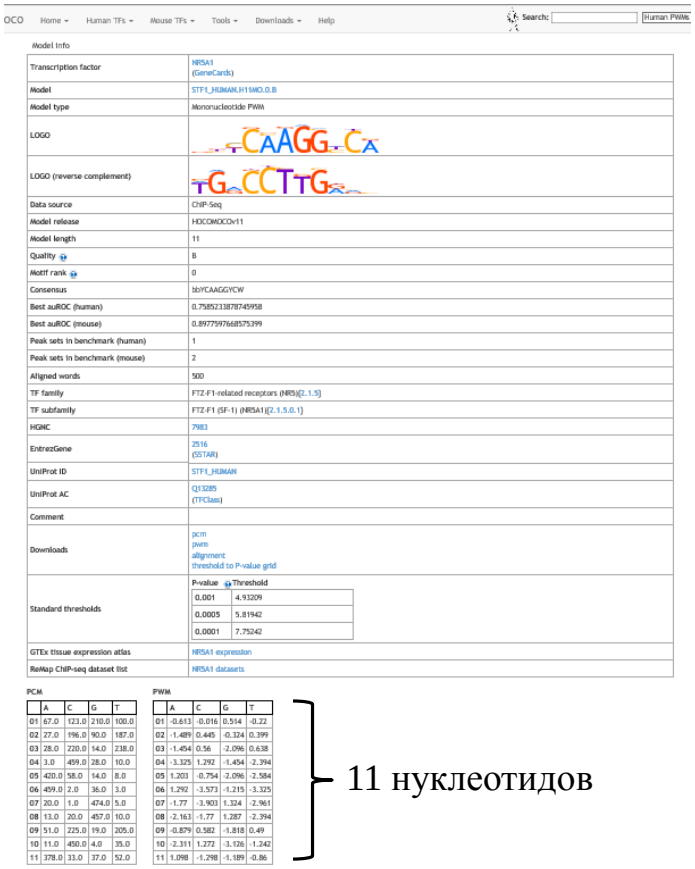


HOCOMOCO v11 provides transcription factor (TF) binding models for **680 human** and **453 mouse** TFs.

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

HOmo sapiens COmprehensive MOdel COllection (HOCOMOCO) v11

Пример: матрица сайтов связывания STF1_HUMAN.H11MO.0.B



Model Info

Transcription factor	NR5A1 (GeneCards)
Model	STF1_HUMAN.H11MO.0.B
Model type	Mononucleotide PWM
LOGO	
LOGO (reverse complement)	
Data source	ChIP-Seq
Model release	HOCOMOCOv11
Model length	11
Quality	B
Motif rank	0
Consensus	bbYCAAGGYCW
Best auROC (human)	0.7585233878745958
Best auROC (mouse)	0.8977597668575399
Peak sets in benchmark (human)	1
Peak sets in benchmark (mouse)	2
Aligned words	500
TF family	FTZ-F1-related receptors (NR5)(2.1.5)
TF subfamily	FTZ-F1 (SF-1) (NR5A1)(2.1.5.0.1)
HGNC	7983
EntrezGene	2516 (SSTAR)
UniProt ID	STF1_HUMAN
UniProt AC	Q13285 (TFClass)
Comment	
Downloads	pcm pwm alignment threshold to P-value grid
Standard thresholds	P-value / Threshold: 0.001 — 4.93209; 0.0005 — 5.81942; 0.0001 — 7.75242
GTEx tissue expression atlas	NR5A1 expression
ReMap ChIP-seq dataset list	NR5A1 datasets

PCM

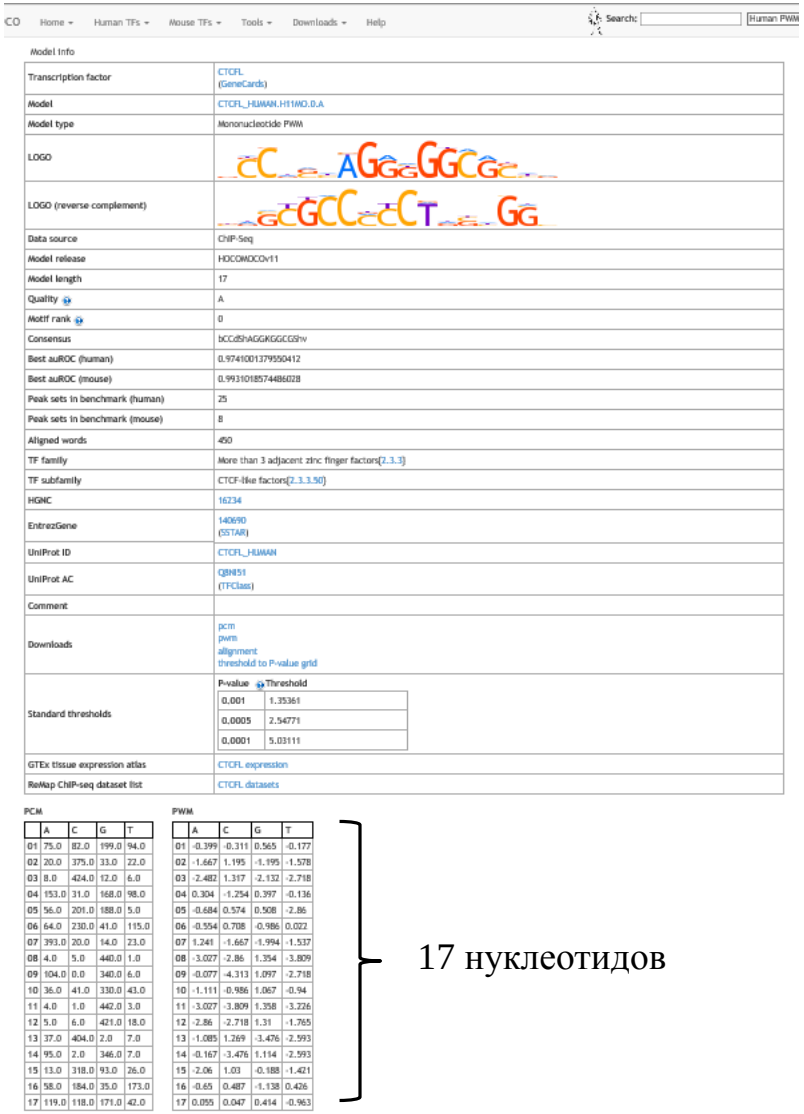
	A	C	G	T
01	67.0	123.0	210.0	100.0
02	27.0	196.0	90.0	187.0
03	28.0	220.0	14.0	238.0
04	3.0	459.0	28.0	10.0
05	420.0	58.0	14.0	8.0
06	459.0	2.0	36.0	3.0
07	20.0	1.0	474.0	5.0
08	13.0	20.0	457.0	10.0
09	51.0	225.0	19.0	205.0
10	11.0	450.0	4.0	35.0
11	378.0	33.0	37.0	52.0

PWM

	A	C	G	T
01	-0.613	-0.016	0.514	-0.22
02	-1.489	0.445	-0.324	0.399
03	-1.454	0.56	-2.096	0.638
04	-3.325	1.292	-1.454	-2.394
05	1.203	-0.754	-2.096	-2.584
06	1.292	-3.573	-1.215	-3.325
07	-1.77	-3.903	1.324	-2.961
08	-2.163	-1.77	1.287	-2.394
09	-0.879	0.582	-1.818	0.49
10	-2.311	1.272	-3.126	-1.242
11	1.098	-1.298	-1.189	-0.86

11 нуклеотидов

Пример: матрица сайтов связывания CTCF



Model Info

Transcription factor	CTCFL (GeneCards)
Model	CTCFL_HUMAN.H11MO.0.A
Model type	Mononucleotide PWM
LOGO	
LOGO (reverse complement)	
Data source	ChIP-Seq
Model release	HOCOMOCOv11
Model length	17
Quality	A
Motif rank	0
Consensus	bCCdShAGGkKGGCGShw
Best auROC (human)	0.9741001379550412
Best auROC (mouse)	0.9931018574486028
Peak sets in benchmark (human)	25
Peak sets in benchmark (mouse)	8
Aligned words	450
TF family	More than 3 adjacent zinc finger factors(2.3.3)
TF subfamily	CTCF-like factors(2.3.3.50)
HGNC	16234
EntrezGene	140690 (SSTAR)
UniProt ID	CTCFL_HUMAN
UniProt AC	Q8NI51 (TFClass)
Comment	
Downloads	pcm pwm alignment threshold to P-value grid
Standard thresholds	P-value / Threshold: 0.001 — 1.35361; 0.0005 — 2.54771; 0.0001 — 5.03111
GTEx tissue expression atlas	CTCFL expression
ReMap ChIP-seq dataset list	CTCFL datasets

PCM

	A	C	G	T
01	75.0	82.0	199.0	94.0
02	20.0	375.0	33.0	22.0
03	8.0	424.0	12.0	6.0
04	153.0	31.0	168.0	98.0
05	56.0	201.0	188.0	5.0
06	64.0	230.0	41.0	115.0
07	393.0	20.0	14.0	23.0
08	4.0	5.0	440.0	1.0
09	104.0	0.0	340.0	6.0
10	36.0	41.0	330.0	43.0
11	4.0	1.0	442.0	3.0
12	5.0	6.0	421.0	18.0
13	37.0	404.0	2.0	7.0
14	95.0	2.0	346.0	7.0
15	13.0	318.0	93.0	26.0
16	58.0	184.0	35.0	173.0
17	119.0	118.0	171.0	42.0

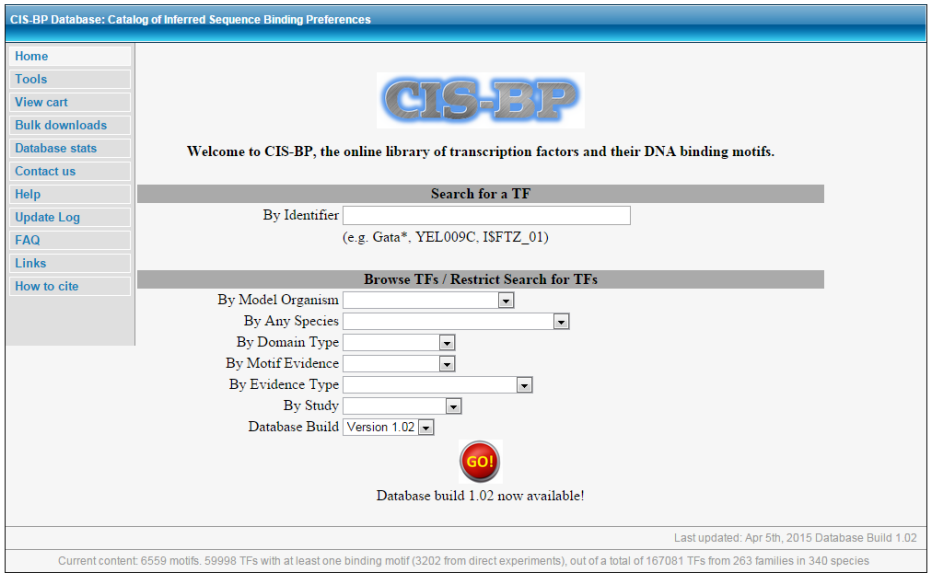
PWM

	A	C	G	T
01	-0.399	-0.311	0.565	-0.177
02	-1.667	1.195	-1.195	-1.578
03	-2.482	1.317	-2.132	-2.718
04	0.304	-1.254	0.397	-0.136
05	-0.684	0.574	0.508	-2.86
06	-0.554	0.708	-0.986	0.022
07	1.241	-1.667	-1.994	-1.537
08	-3.027	-2.86	1.354	-3.809
09	-0.077	-4.313	1.097	-2.718
10	-1.111	-0.986	1.067	-0.94
11	-3.027	-3.809	1.358	-3.226
12	-2.86	-2.718	1.31	-1.765
13	-1.085	1.269	-3.476	-2.593
14	-0.167	-3.476	1.114	-2.593
15	-2.06	1.03	-0.188	-1.421
16	-0.65	0.487	-1.138	0.426
17	0.055	0.047	0.414	-0.963

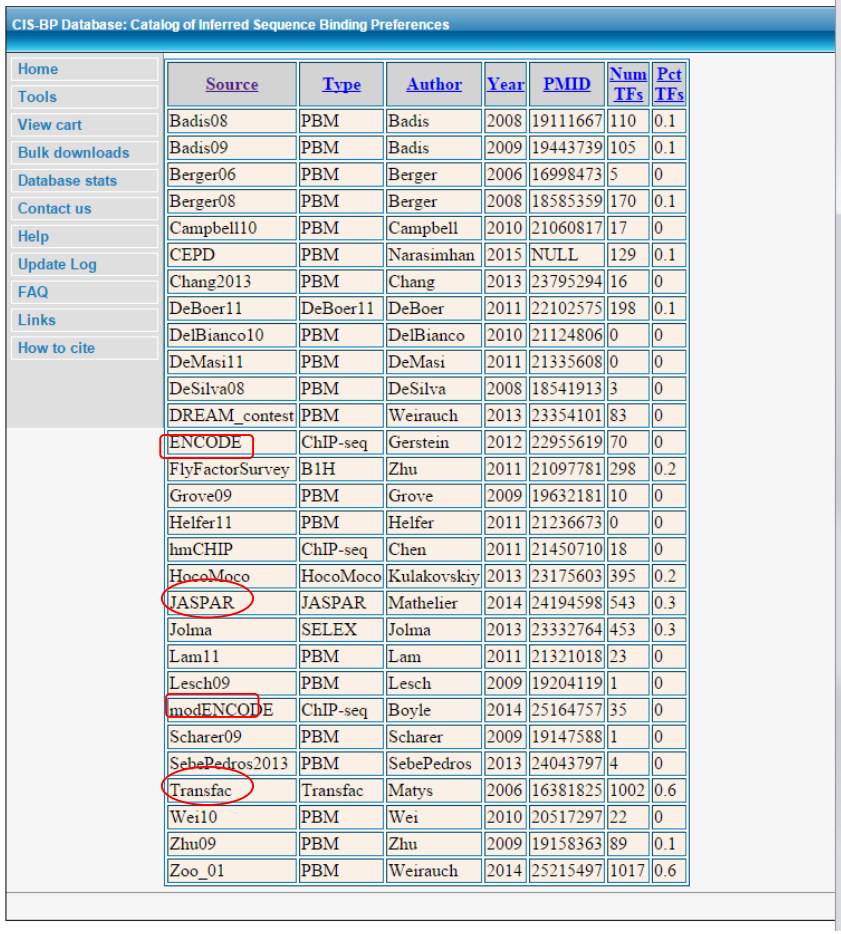
17 нуклеотидов

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

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

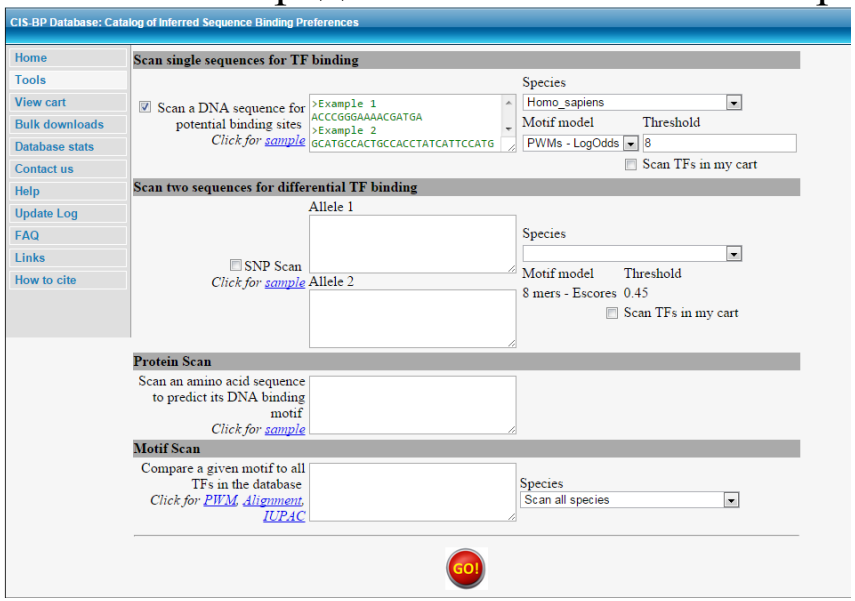


Статистика



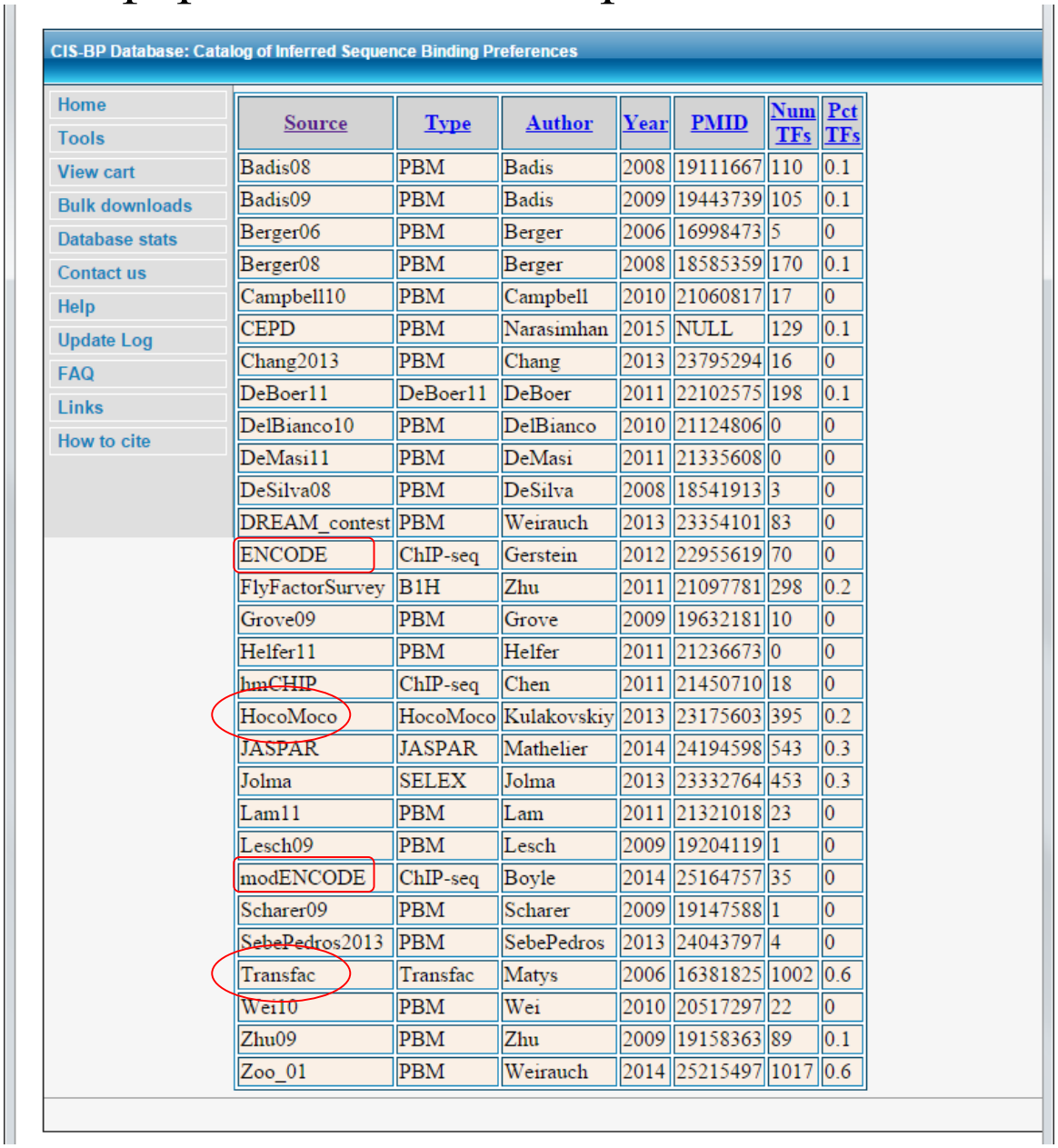
Source	Type	Author	Year	PMID	Num TFs	Pct TFs
Badis08	PBM	Badis	2008	19111667	110	0.1
Badis09	PBM	Badis	2009	19443739	105	0.1
Berger06	PBM	Berger	2006	16998473	5	0
Berger08	PBM	Berger	2008	18585359	170	0.1
Campbell10	PBM	Campbell	2010	21060817	17	0
CEPD	PBM	Narasimhan	2015	NULL	129	0.1
Chang2013	PBM	Chang	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

Возможность предсказания сайтов по матрицам



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

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

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*

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/

TRRD ВКЛЮЧАЕТ ДАННЫЕ О:

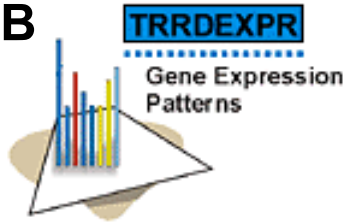
1) СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ РАЙОНОВ. СТРУКТУРНЫЕ И ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ:

САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

РЕГУЛЯТОРНЫХ РАЙОНОВ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)

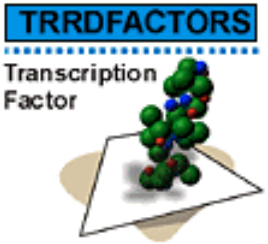
ЛОКУС-КОНТРОЛИРУЮЩИХ РАЙОНОВ

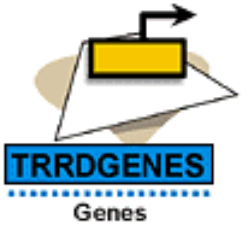
2) ДАННЫЕ О ПАТТЕРНАХ ЭКСПРЕССИИ ГЕНОВ



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

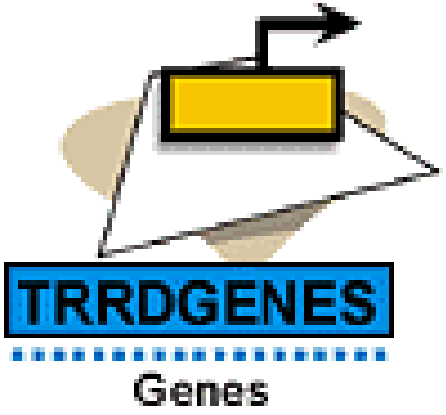
3) ДАННЫЕ О ТРАНСКРИПЦИОННЫХ ФАКТОРАХ





TRRDGENES

ОБЩЕЕ ОПИСАНИЕ ГЕНА



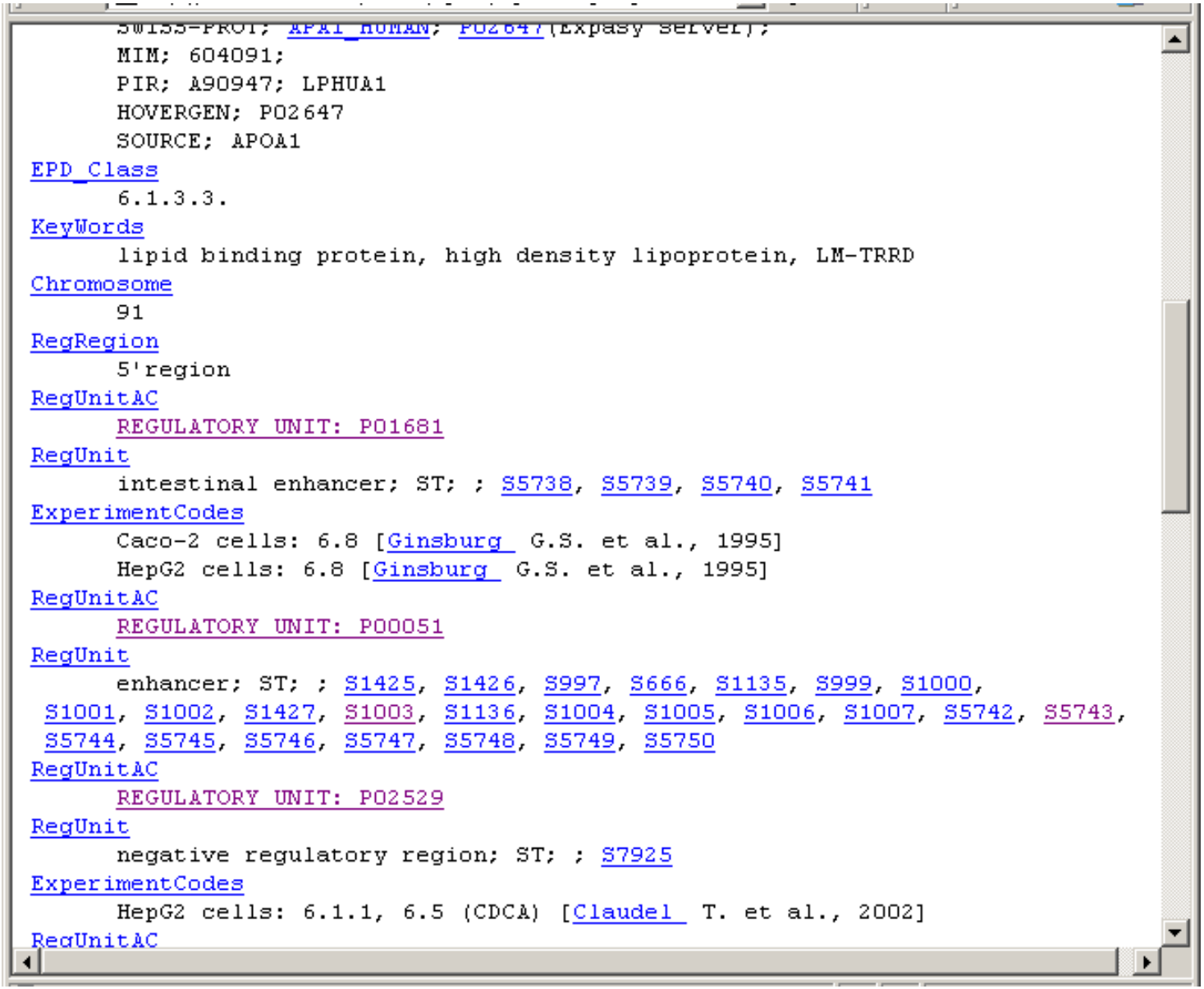
ID Hs:AAP
DT 09/02/00
AC A00596
CR Ignatieva E.V., Stepanenko I.L.
OS human, Homo sapiens
SN AAP
NG Alzheimer's disease amyloid A4 precursor protein
SY amyloid beta protein precursor gene
SY amyloid precursor protein gene
SY APP
BI EMBL;HSPADP;X12751; ST:3700
DR GDB; 119692; APP
DR SWISS-PROT; A4_HUMAN; P05067
KW heat shock-induced, TATA-less promoter, pathogenesis-related protein, multiple transcription initiation sites
CH 21
RG 5'region
AP REGULATORY UNIT: P01087
PR regulatory region; ST; -2260 to -1800; S3936, S3937
AP REGULATORY UNIT: P00760
PR promoter; ST;-160 to -1; S2907, S3931, S3932, S3933, S3934, S3935
//

ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)

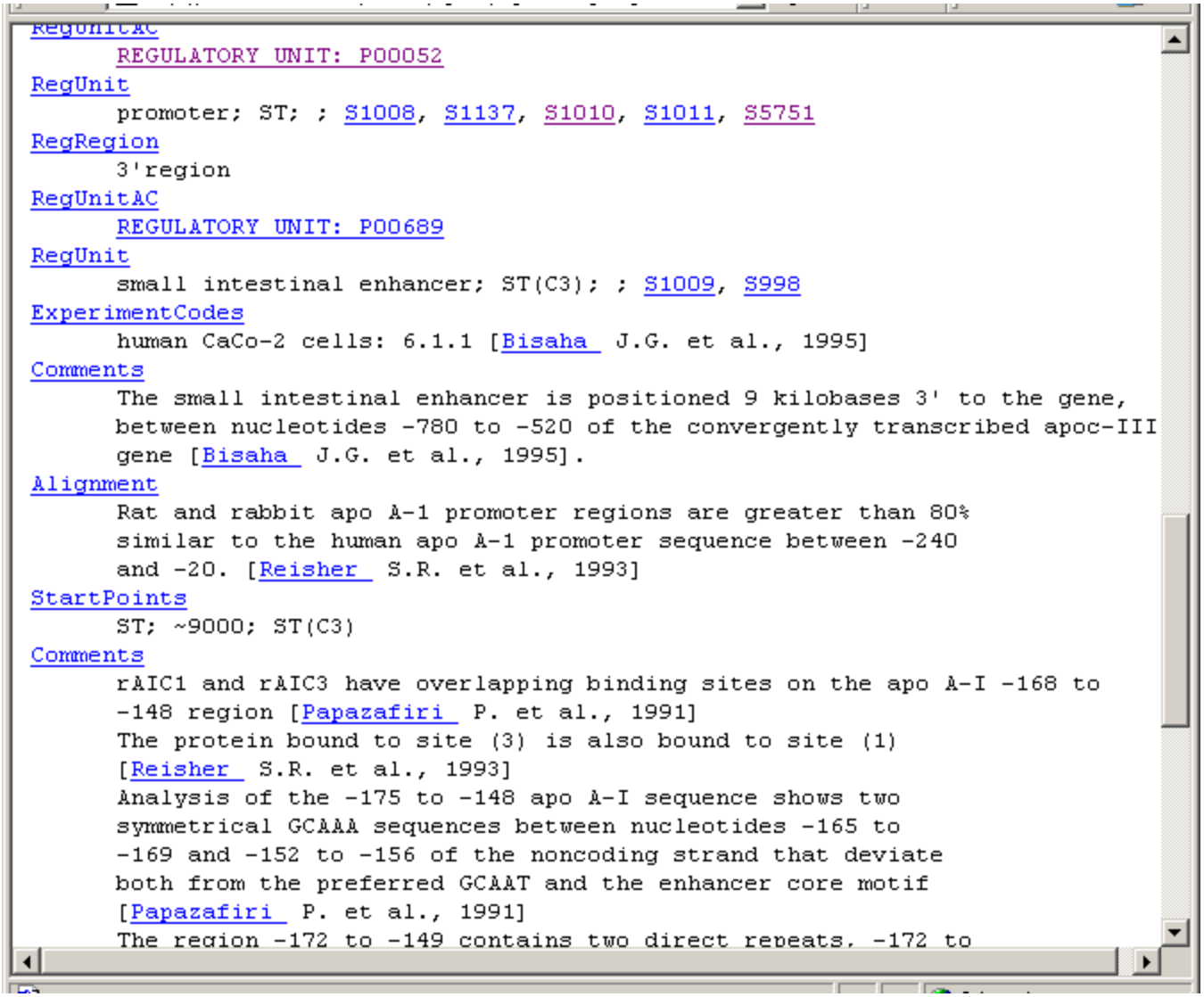
TRRDGENES4:A00264

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)



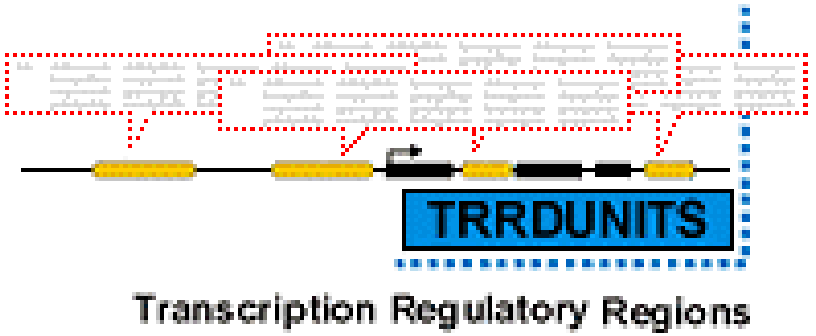
Окончание:
ОБЩЕЕ ОПИСАНИЕ ГЕНА (html –представление)



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)

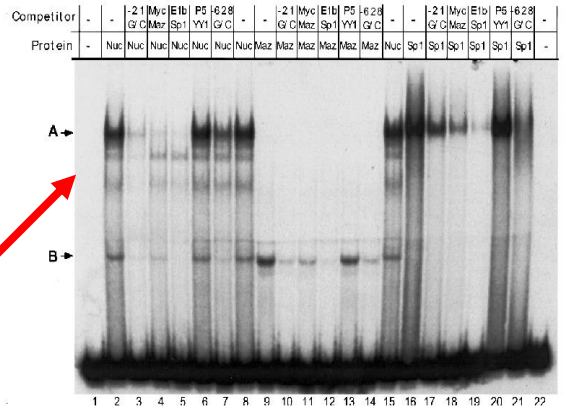
TRRDUNITS4:P00051

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-responsiv
Site:(S1001) -214 to -192; ARP-1; ARP-1 binding site
Site:(S1002) -210 to -189; S (1); site (1)
Site:(S1427) -193 to -178; Egr-1 bs;
Site:(S1003) -175 to -155; C (1); region C (1)
Site:(S1136) -174 to -151; HNF-3 beta; HNF-3 beta bindin
Site:(S1004) -168 to -148; C (2) (B); region C (2)
Site:(S1005) -174 to -144; S (2); site S (2)
Site:(S1006) -134 to -119; C; site C
Site:(S1007) -133 to -110; S (3); site (3)
Site:(S5742) -220 to -190; D; region D
Site:(S5743) -214 to -192; T3R/RXR bs; T3R/RXR alpha bin
Site:(S5744) -180 to -147 ; FpB; Footprint B
Site:(S5745) -175 to -148; NFY bs;
Site:(S5746) -149 to -130; ARE; antioxidant response ele
Site:(S5747) -142 to -118; FpC; Footprint C
Site:(S5748) -134 to -119; T3R/RXR bs; T3R beta / RXR al
Site:(S5749) -134 to -119; ARP1 bs; ARP-1 binding site

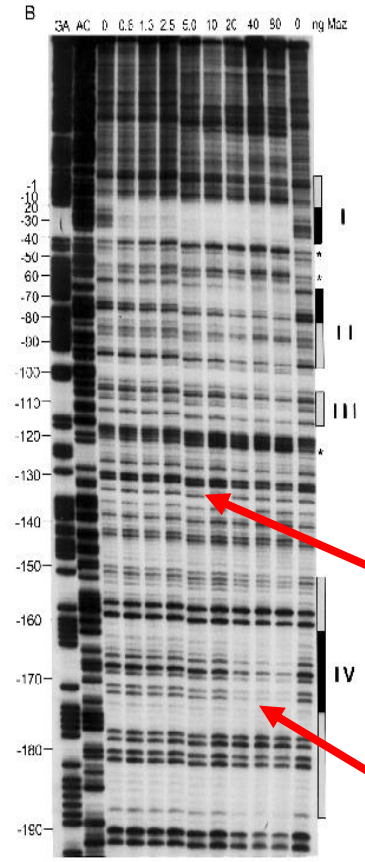
Site:(S5750) -134 to -119; HNF-4 bs; HNF-4 binding site
DNA_BankLink
EMBL; HSAPOA01; J04066; 1813 to 1959
LeftTrunc
0
RightTrunc
0
SeqLength
147
Sequence
ccacccggga gacctgcaag cctgcagcac tcccctcccg cccccactga
acccttgacc cctgccctgc agcccccgca gcttgctgtt tgcccactct
atttgcccag ccccagggac agagctgatc cttgaactct 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

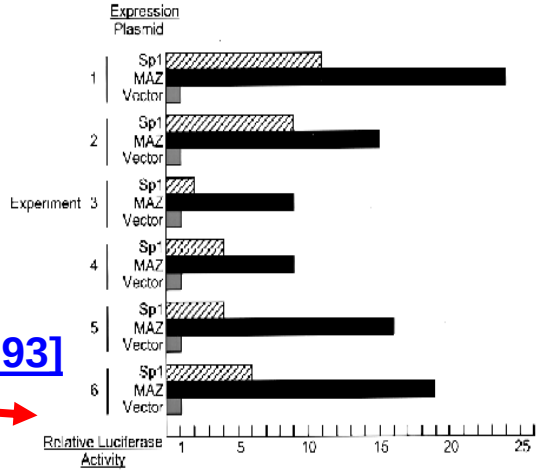
ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА



GEL-MOBILITY SHIFT ASSAY



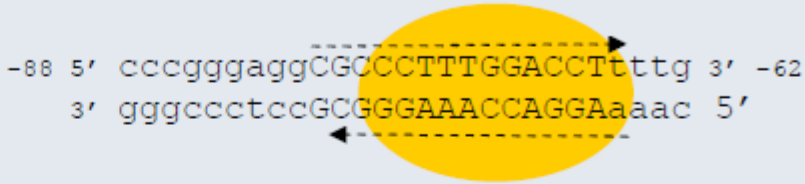
AN S1160
ID Gene: Hs:APOB
AP REGULATORY UNIT: P00670
NM HNF-4 bs; HNF-4 binding site
NY AF-1 binding site
NY BA1 binding site
NS R01612
TF HNF-4; hepatic nuclear factor 4
AT increase
SQ cccgggaggCGCCCTTTGGACCTtttg
PQ -88 to -62
PF -82 to -62
BF EMBL: M15053 : 68
AG 1.1.5, 3.5 [Metzger S. et al., 1993]
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]
//



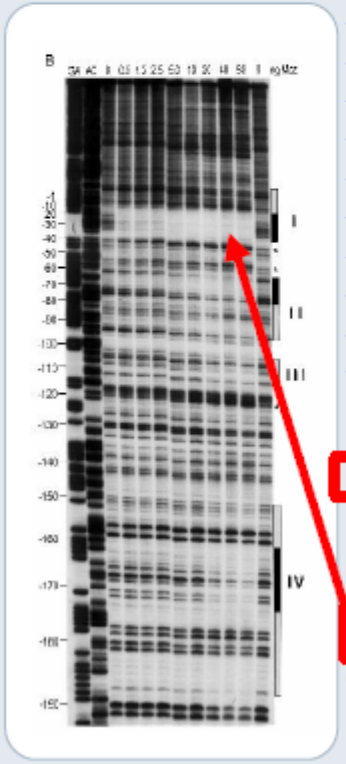
DNASE I FOOTPRINTING

TRANSIENT EXPRESSION ANALYSIS

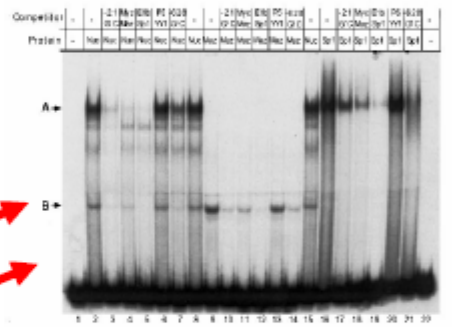
The table TRRDSITES: description of the HNF-4 binding site in the human ApoB ge



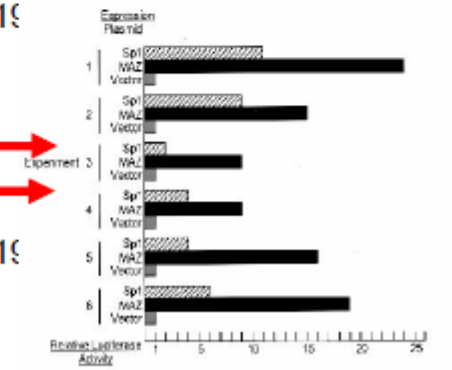
SiteAC S1160
GeneID Hs:APOB
RegUnitAC P00670
SiteName HNF-4;
PreferredName HNF-4
SiteNameSynonym AF-1 binding site
SiteNameSynonym BA1 binding site
SiteIndex 1
FactorName HNF-4; hepatic nuclear factor 4
FactorInfluence increase
Sequence cccgggaggCGCCCTTTGGACCTtttg
SequencePosition -88 to -62
FootprintSequencePosition -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

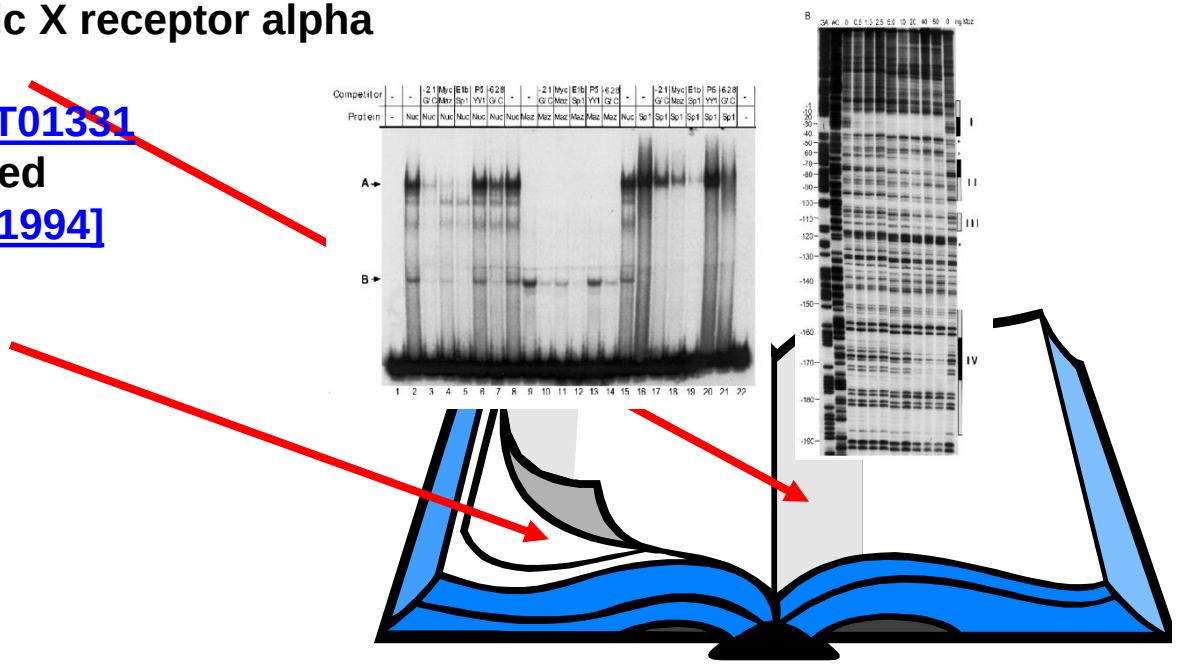
SiteAC
S5743
GeneID
Gene: Hs:APOA1
RegUnitAC
REGULATORY UNIT: P00051
SiteName
T3R/RXR bs; T3R/RXR alpha binding site
NP RXR
NP TR
SiteIndex
2
FactorName
T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer
T3R beta/RXR alpha; T3R beta / retinoic X receptor alpha heterodimer
FactorInfluence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

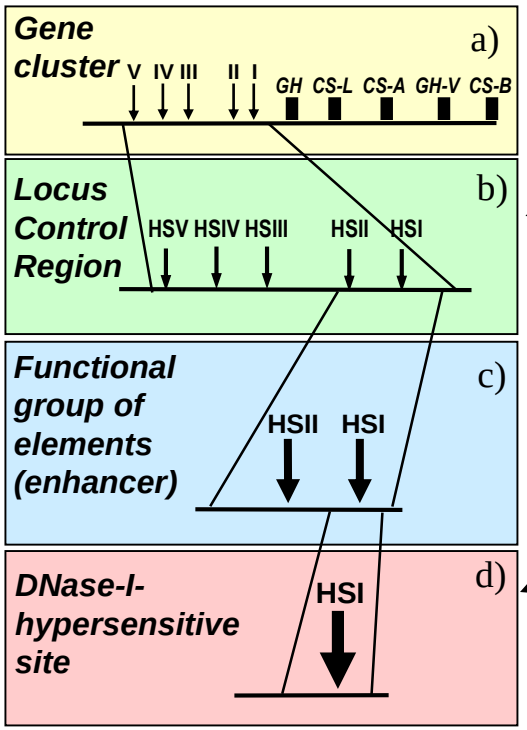
Identifier
F5743.2
GeneID
Hs:APOA1

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; Catarr
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

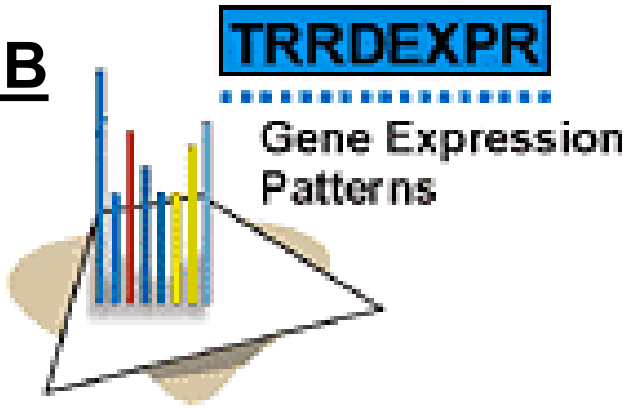
AL A7
LI Hs:LPGH
TL pituitary specific
BI EMBL; AF010280; AF010280
EL LGH:HSV, LGH:HSIII, LGH ENHANCER
.....

AR R14
RF
RT pituitary specific
SV

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

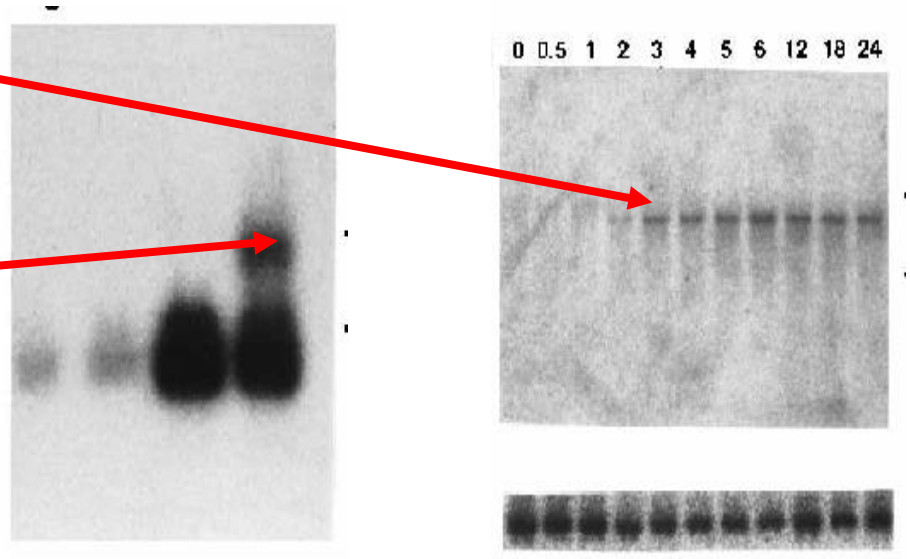
TRRDEXP

ОСОБЕННОСТИ ЭКСПРЕССИИ ГЕНОВ



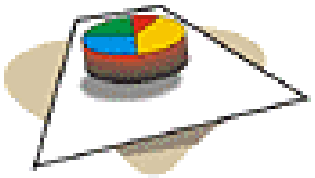
RE A00596.002
ID Hs:AAP
RT mRNA
RN lymphoblastoid cells
RL present
RI heat shock
FF induction
RH 3 h
RR [Abe K. et al., 1991]

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



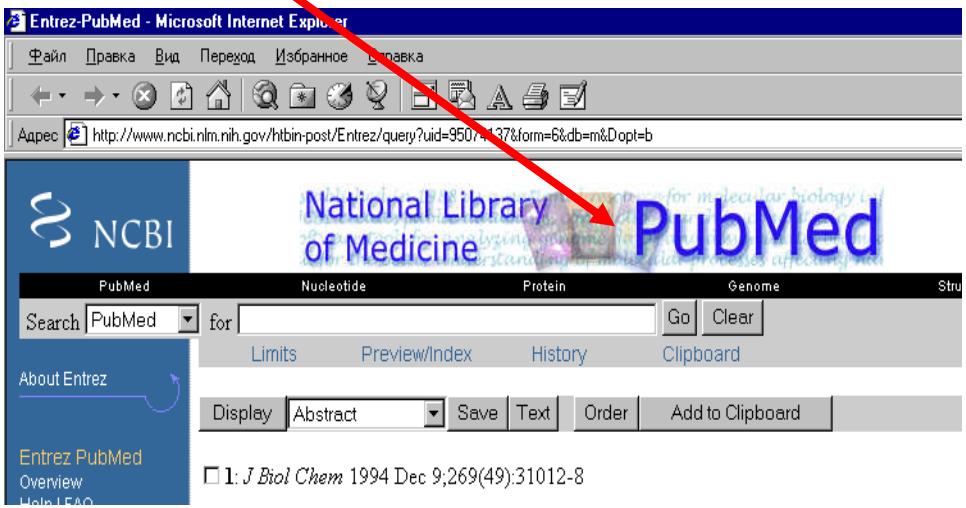
RNA blot analysis

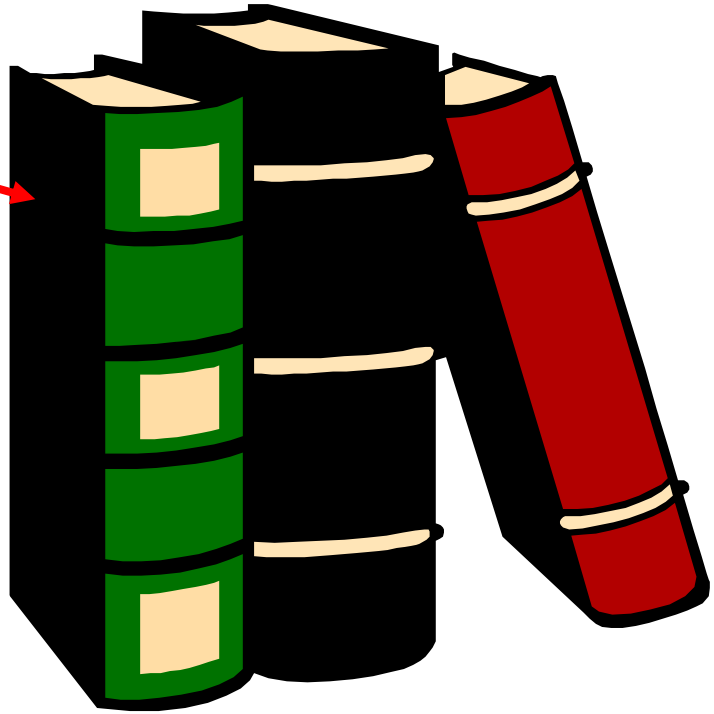
TRRDBIB



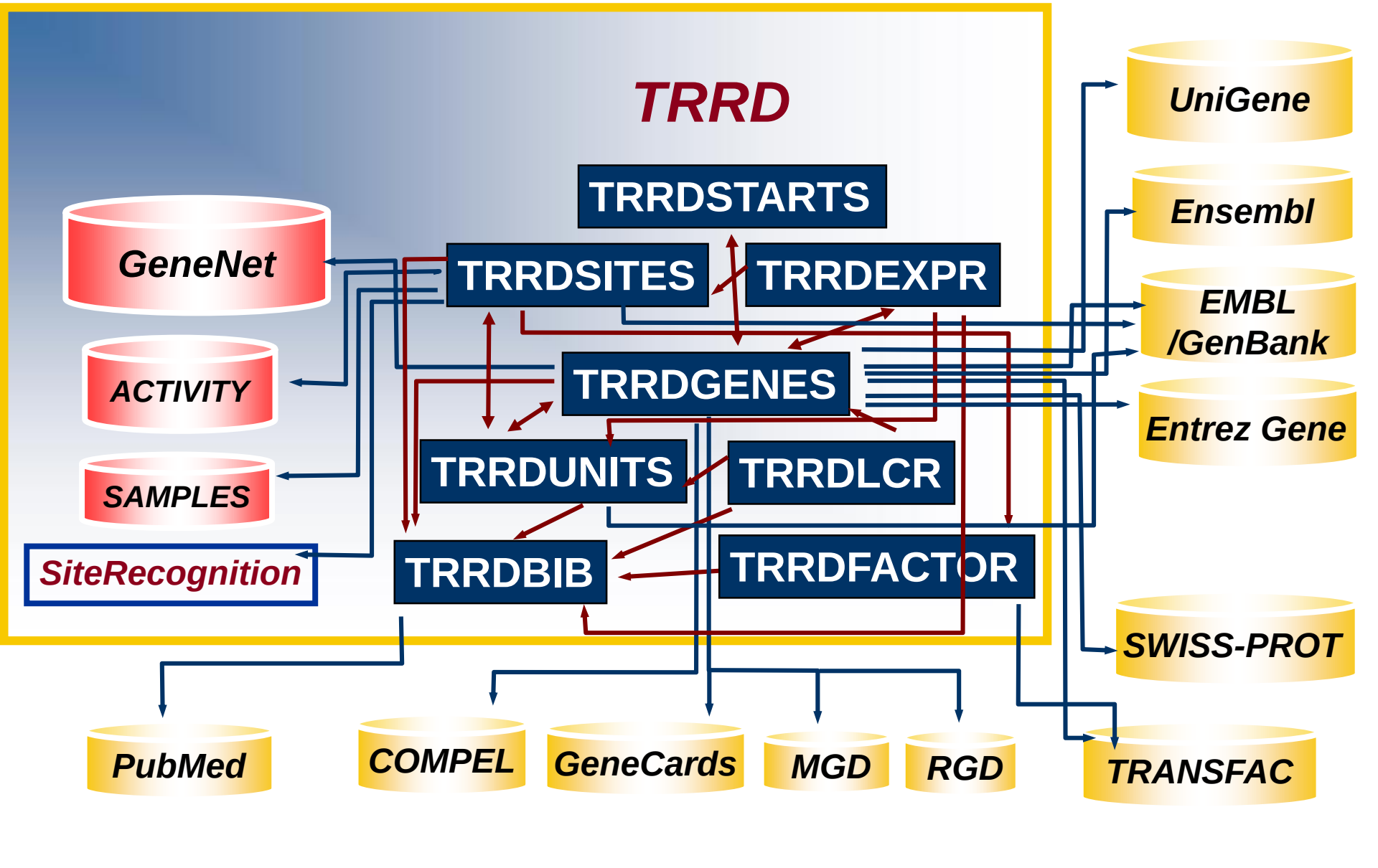
TRRDBIB
Bibliography

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





Линковка между таблицами TRRD и ссылки на внешние ресурсы.



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

```
AN S161
ID Gene: Hs:PAI1
AP REGULATORY UNIT:
NM USF bs;
WW http://wwwmgs.bionet.nsc.ru/Programs/acts2/AFF_USF.htm
DR ACTIVITY ; A00BF026;
DR SAMPLES; USF;
TF USF; upstream sequence fac
AT increase
SQ gacattcacgtggct
PQ -574 to -559
PF -574 to -559
BF EMBL : X13523 : 236
AG 1.1.1, 1.1.2, 3.1, 3.3, 6.1, 6.2, 6.3 [Riccio A. et al.,
199
//
```

TRRDGENE

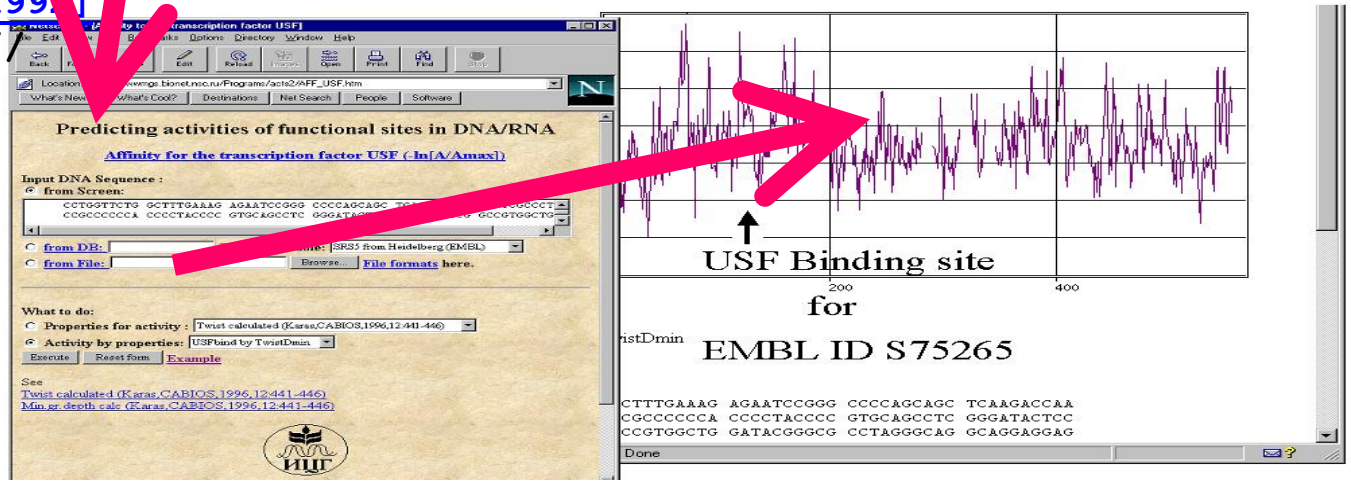
ACTIVITY

SAMPLES

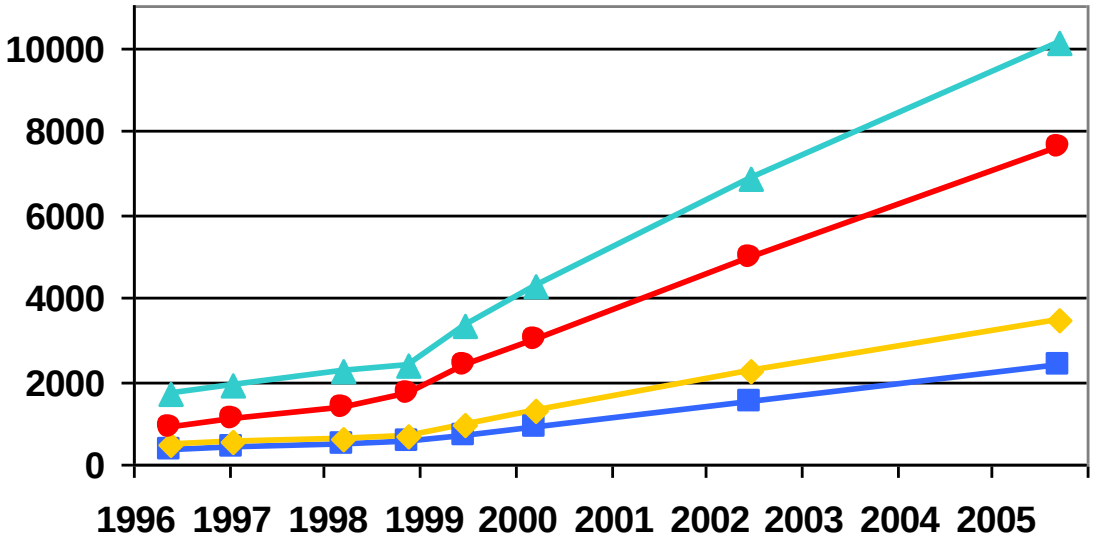
TRRDFACTORS

EMBL

TRRDBIB



ДИНАМИКА ПОПОЛНЕНИЯ TRRD



10135 САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

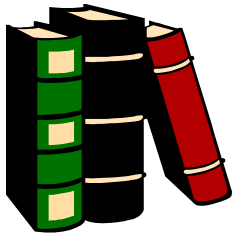
7609 НАУЧНЫХ ПУБЛИКАЦИЙ

3400 РЕГУЛЯТОРНЫХРАЙОНОВ

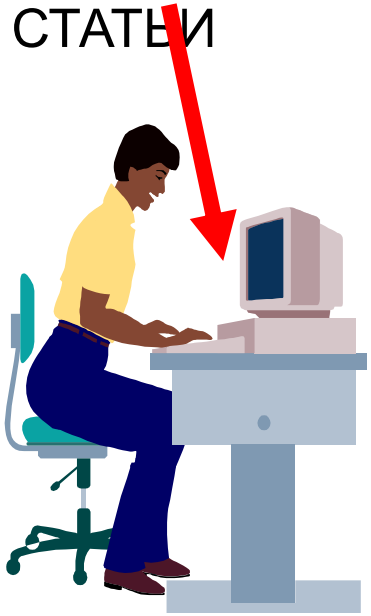
2344 ГЕНОВ

	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)

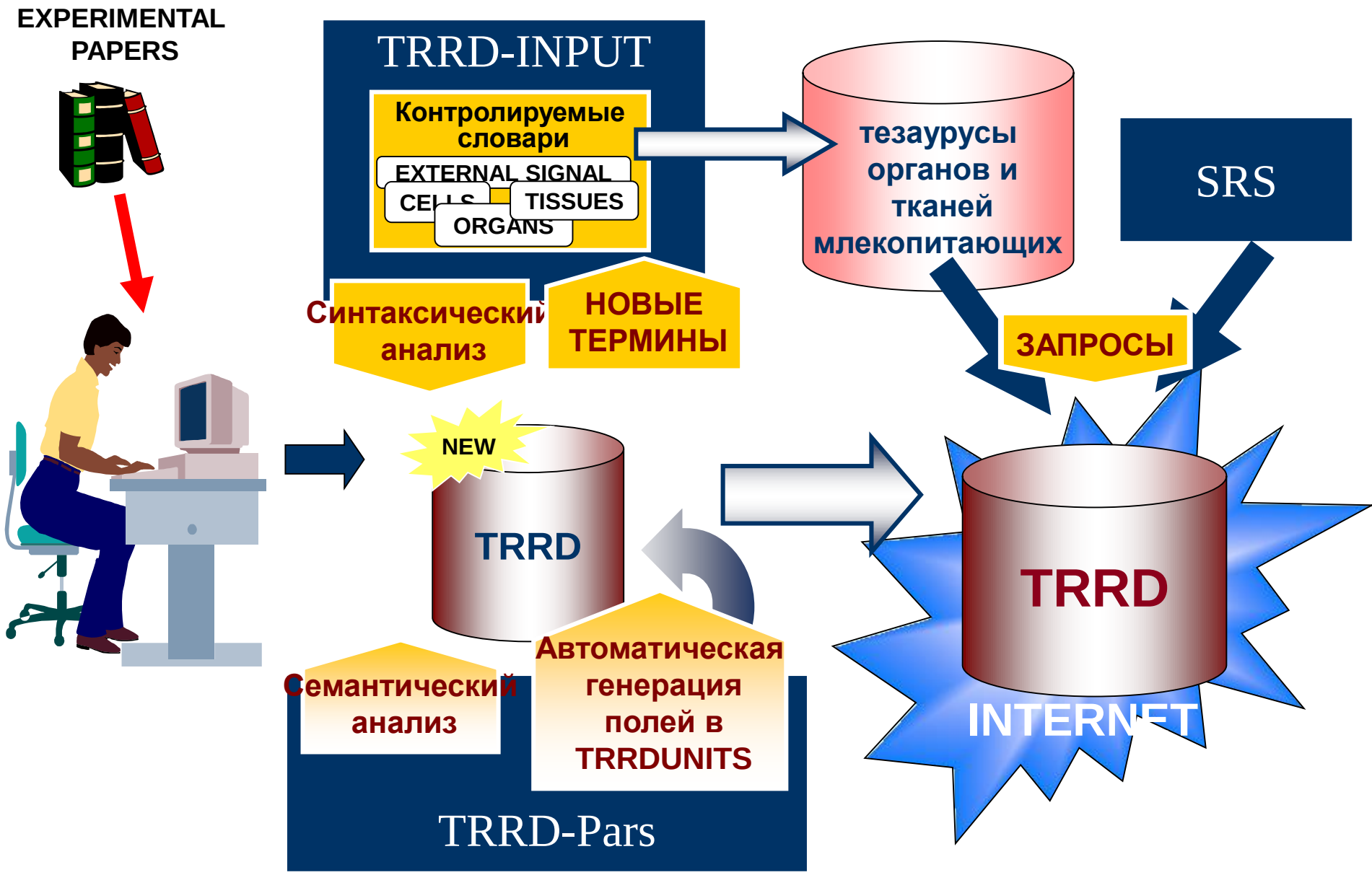
- СТАДИИ РАЗВИТИЯ
- КЛЕТКИ
- ОРГАНЫ
- ТКАНИ
- КЛЮЧЕВЫЕ СЛОВА
- ВНЕШНИЕ СТИМУЛЫ
- СТАДИИ КЛЕТОЧНОГО ЦИКЛА
- КОДЫ ЭКСПЕРИМЕНТОВ

СИНТАКСИЧЕСКАЯ и СЕМАНТИЧЕСКАЯ ПРОВЕРКА

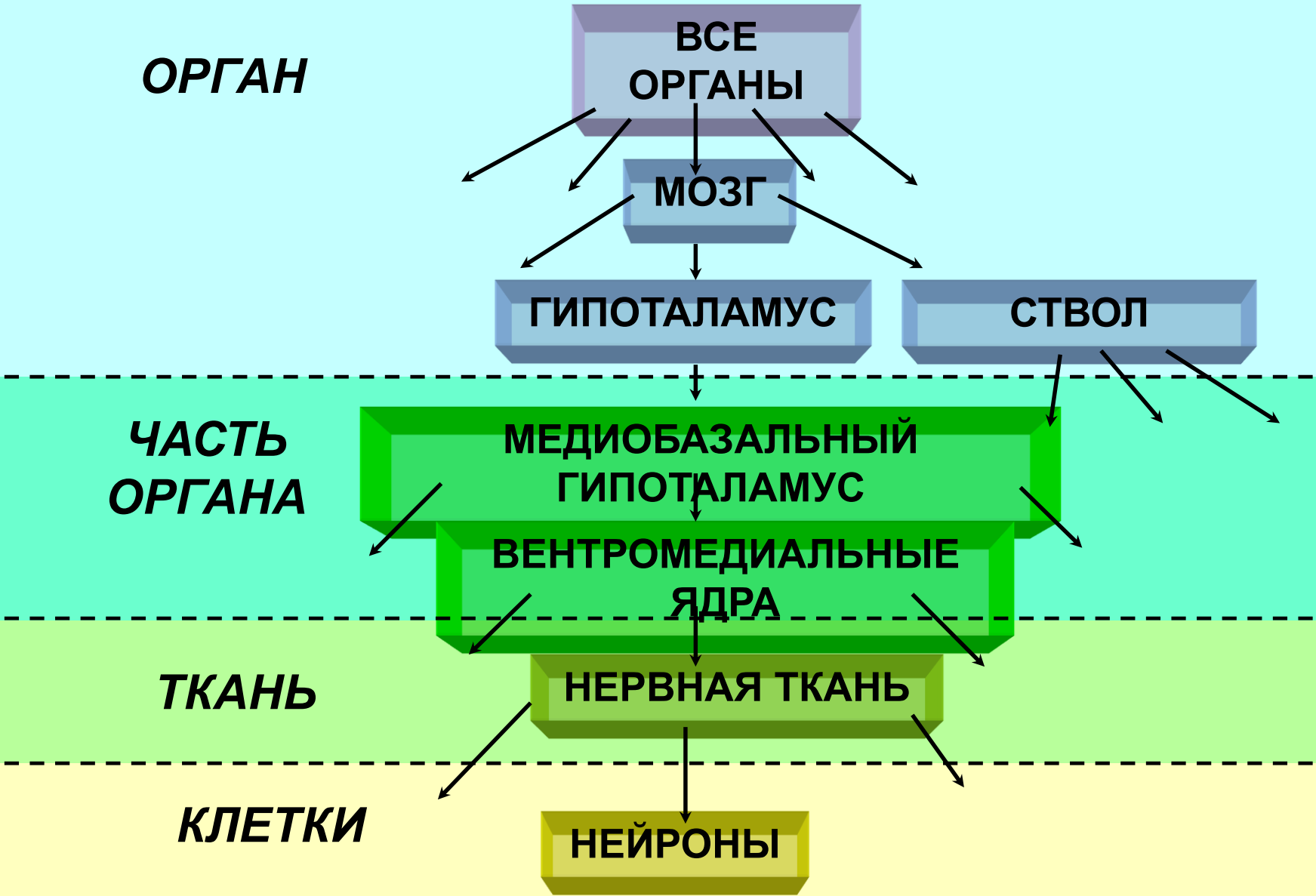
TRRD

Интернет

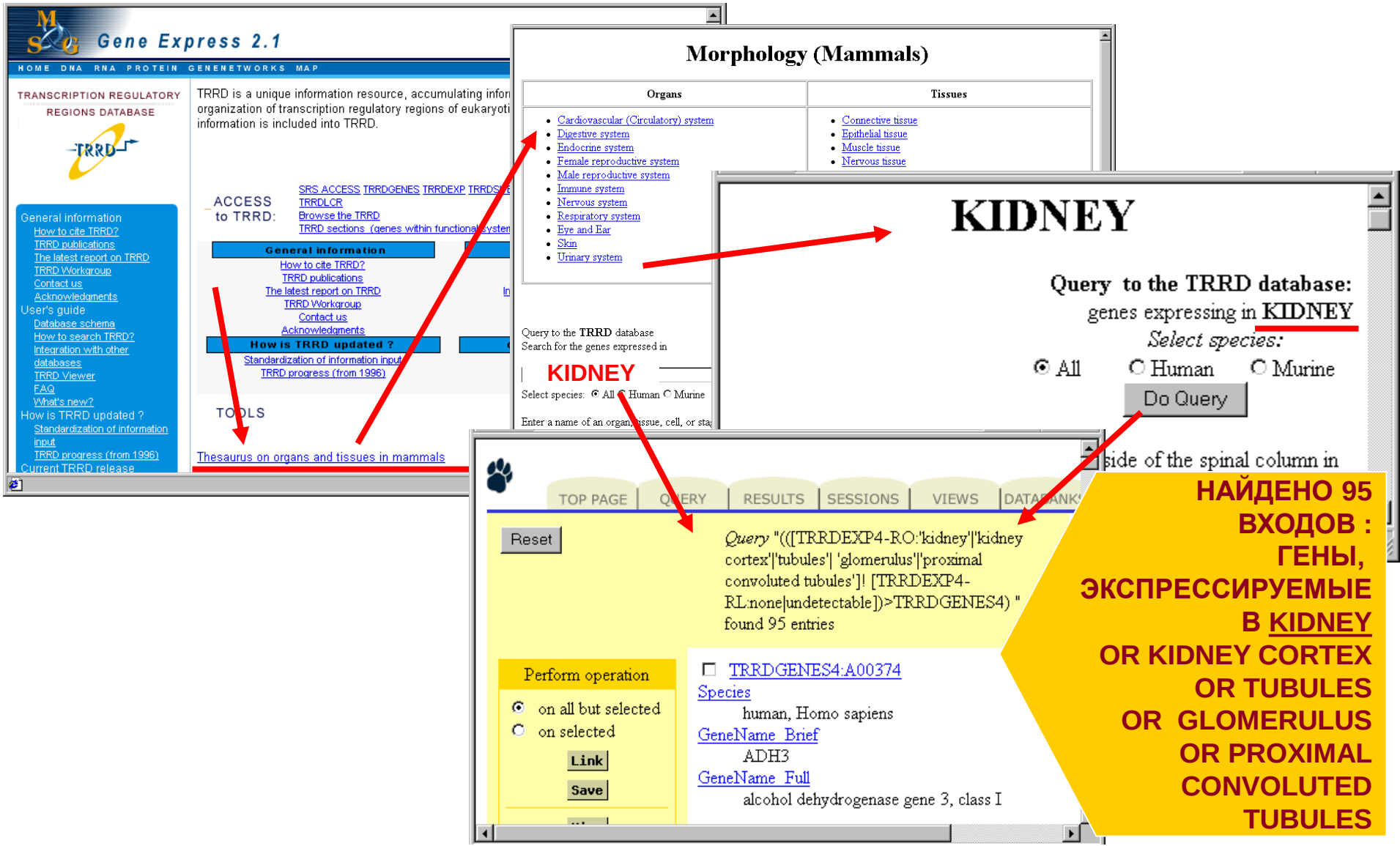
ВВОД и ОБРАБОТКА ДАННЫХ



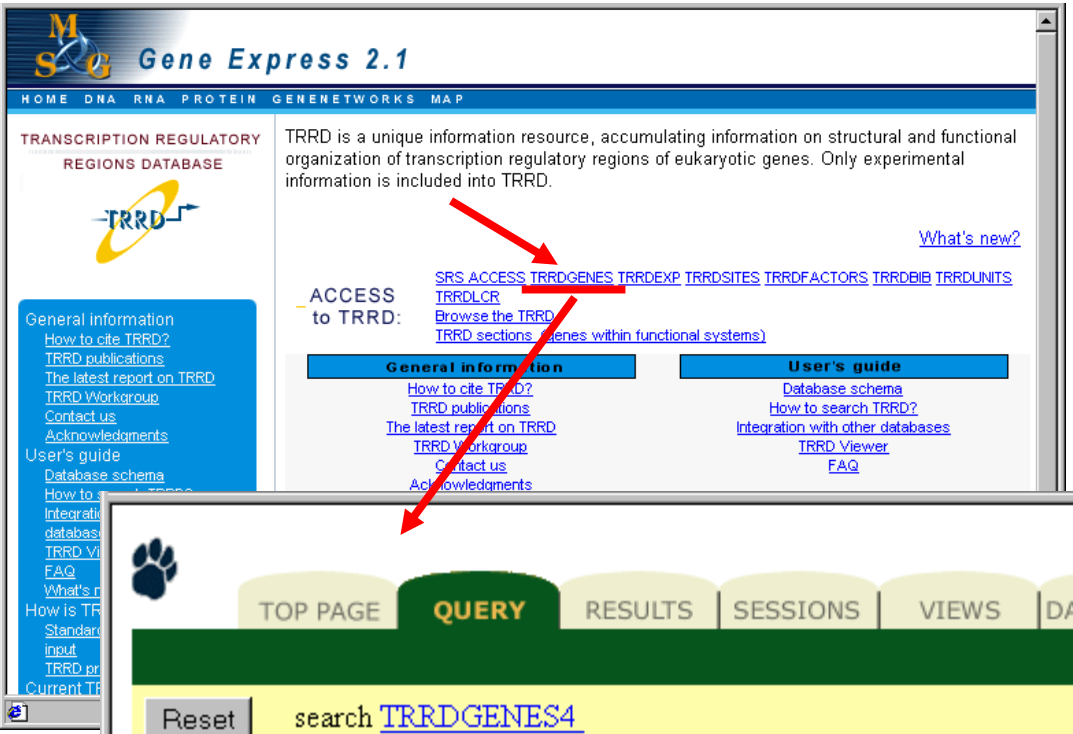
ИЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ КОНТРОЛИРУЕМЫХ СЛОВАРЕЙ МОРФОЛОГИЧЕСКИХ ТЕРМИНОВ В TRRD

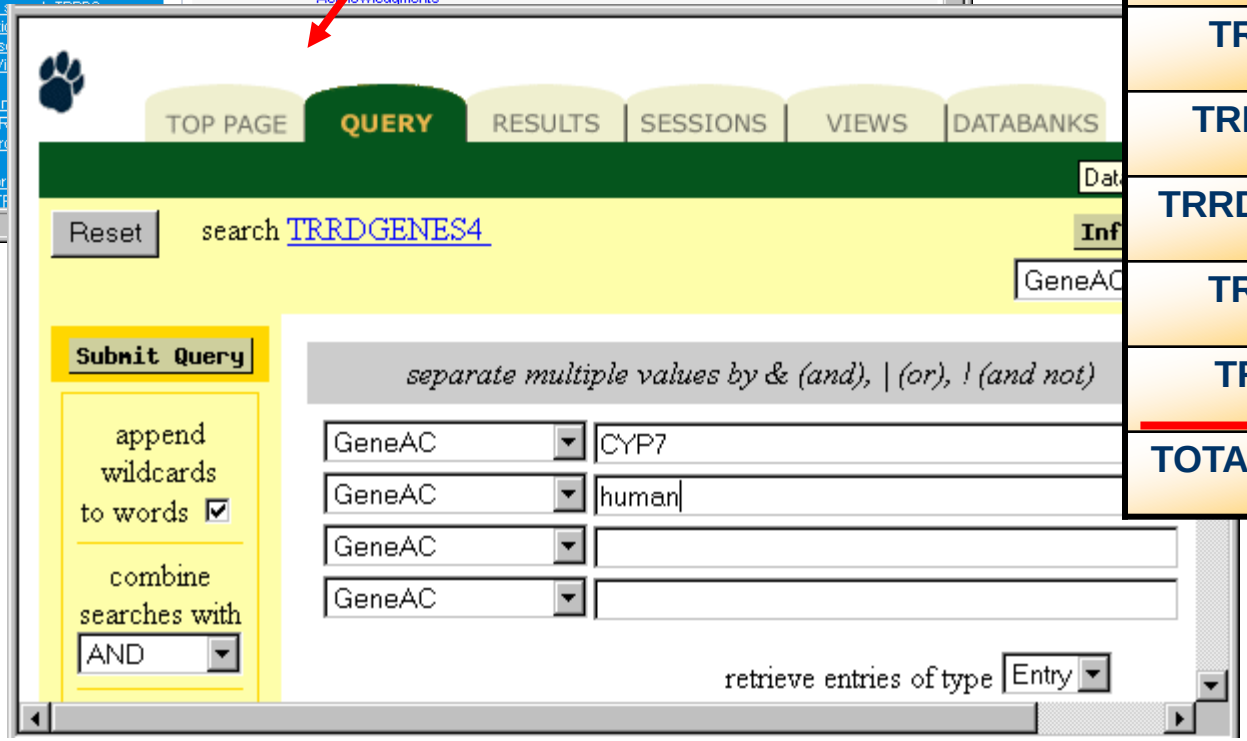


ПОИСК ДАННЫХ вTRRD: ЗАПРОС НА ОСНОВЕ ТЕЗАУРОСОВ ОРГАНОВ и ТКАНЕЙ МЛЕКОПИТАЮЩИХ



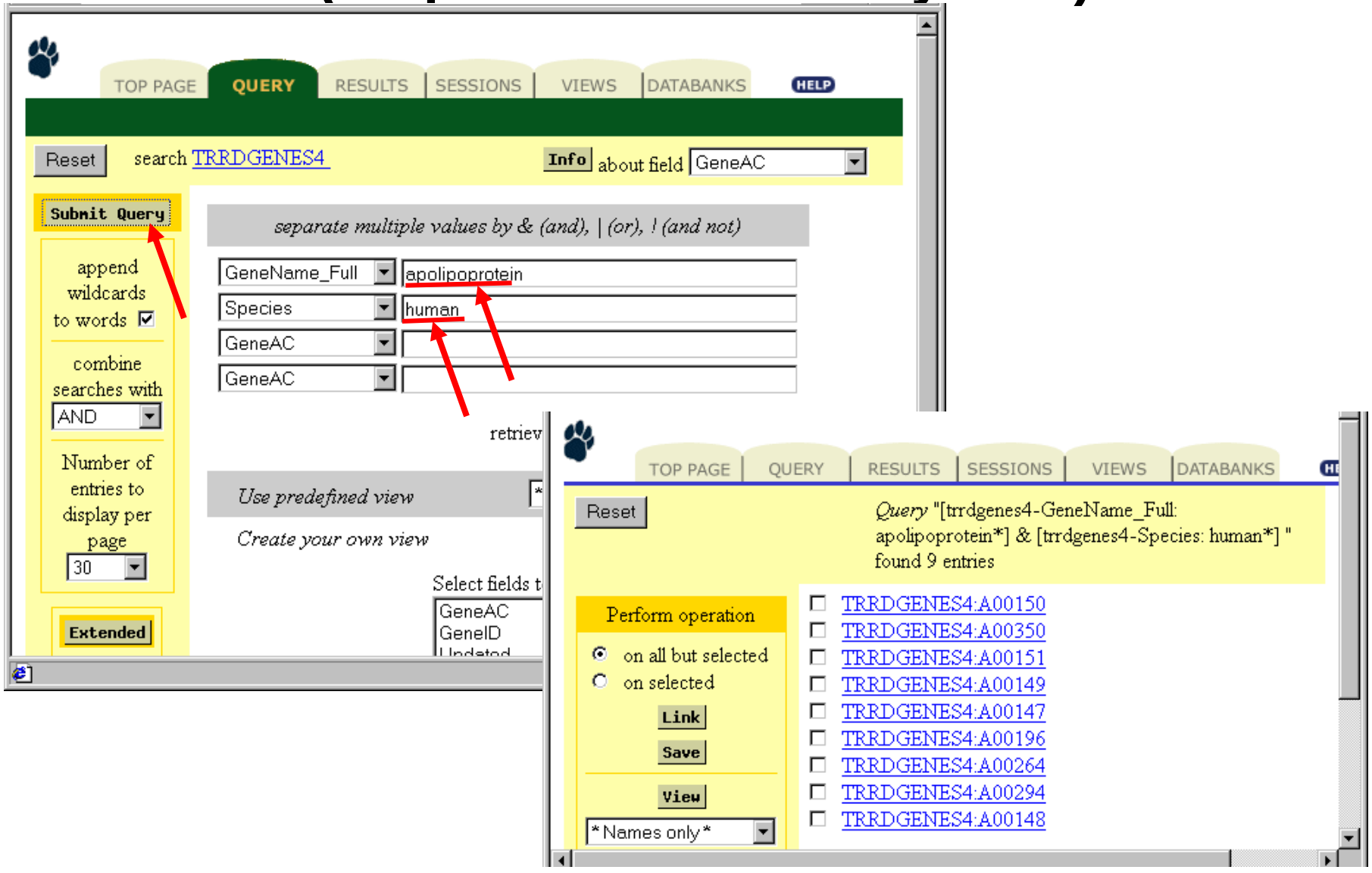
ПОИСК ДАННЫХ в TRRD: SRS



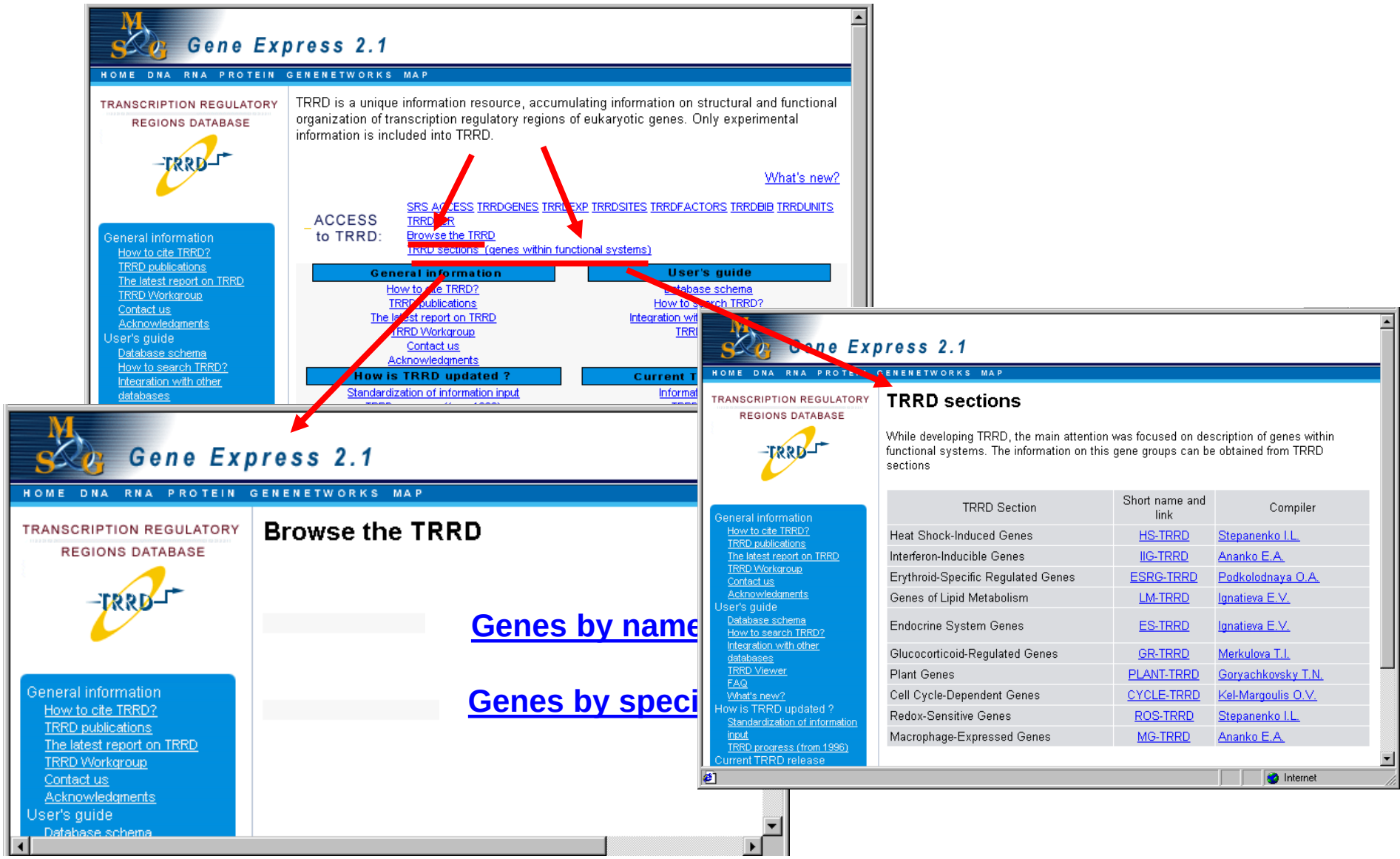


DATABASE	КОЛИЧЕСТВО ПОЛЕЙ, ПРОИНДЕКСИРОВАННЫХ для РЕЛИЗА 6.01
TRRDGENES	24
TRRDUNITS	11
TRRDEXP	17
TRRDSITES	16
TRRDFACTORS	14
TRRDLCR	40
TRRDBIB	9
TOTAL NUMBER	131

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



ПОИСК ДАННЫХ В TRRD: БРАУЗЕРЫ и ТЕМАТИЧЕСКИЕ СЕКЦИИ

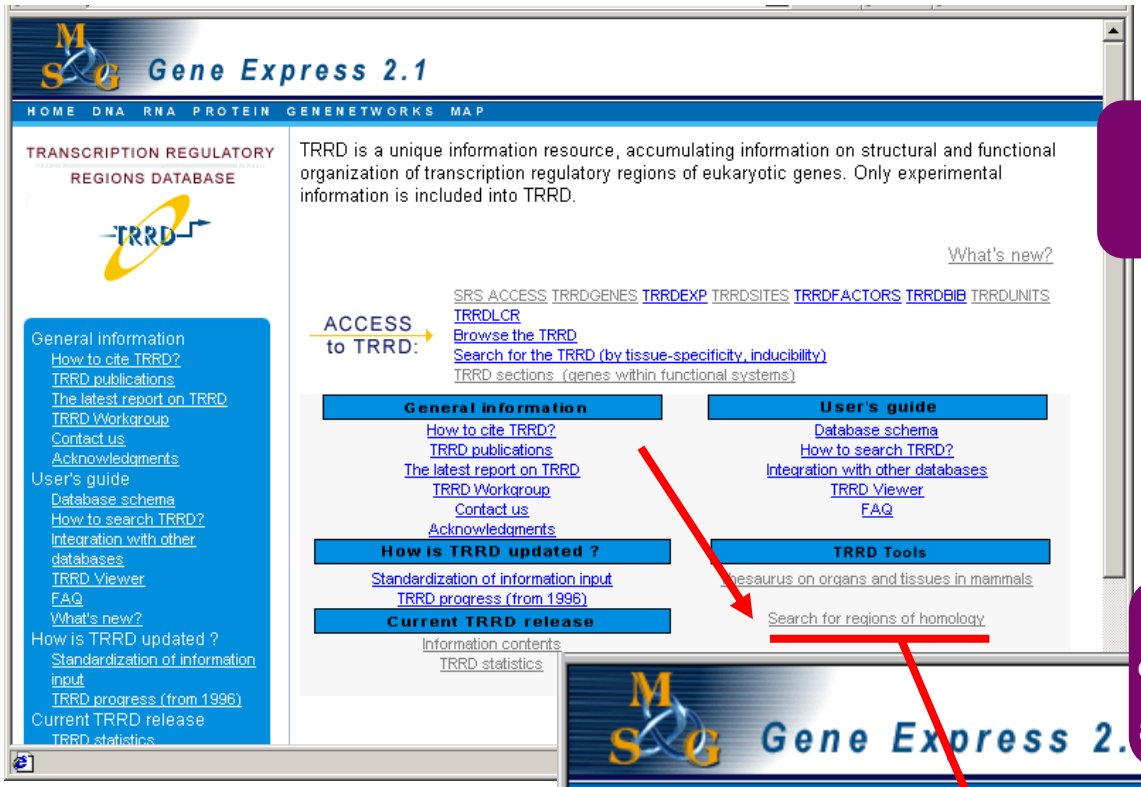


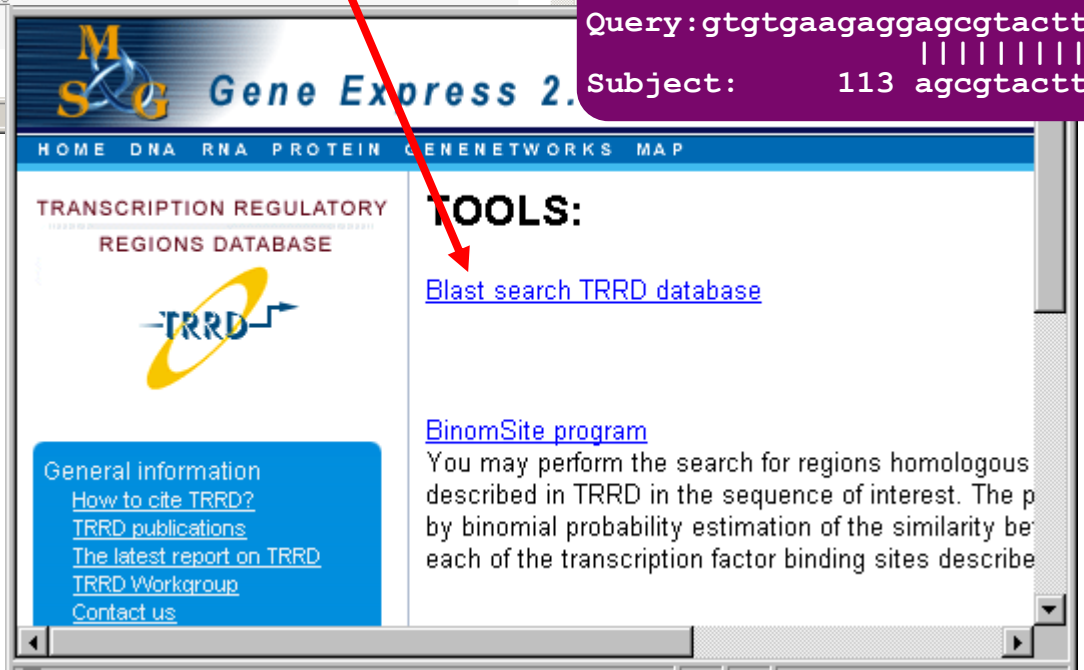
Тематические секции TRRD

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





SEQUENCE UNDER INVESTIGATION:

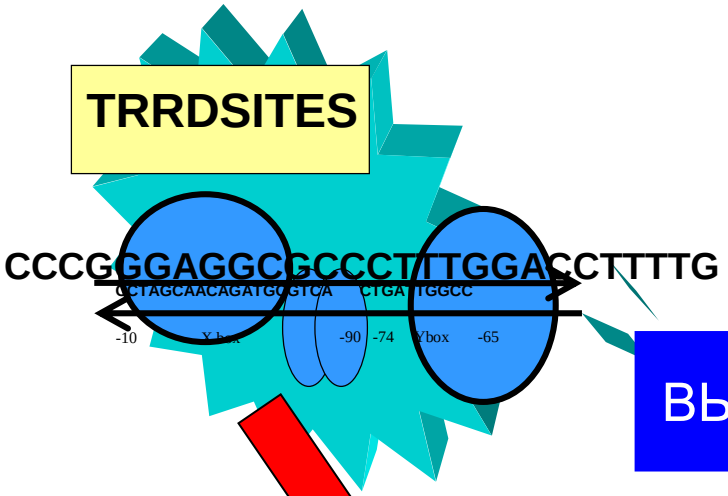
`gtgtgaagaggagcgtacttttgtgtgtgaac`

BLAST search TRRDUNITS

QUERY RESULTS:

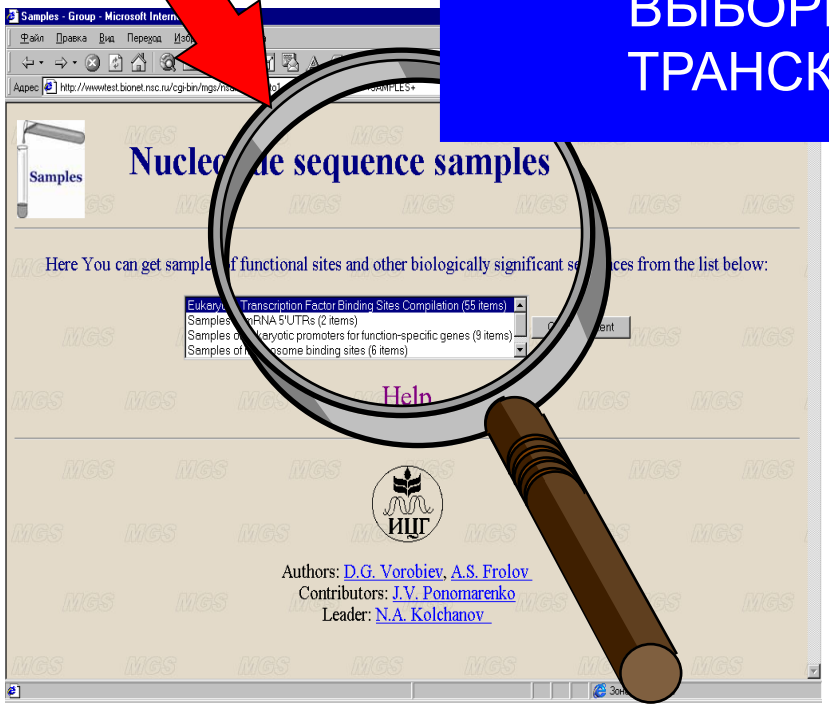
```
Query:gtgtgaagaggagcgtacttttgtgtgtga
                  ||||||||||
Subject:      113 agcgtactttt 103
```

TRRD – информационная основа для создания выборок



ВЫБОРКИ ЭУКАРИОТИЧЕСКИХ ПРОМОТОРОВ

ВЫБОРКИ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ




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

```
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
```

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

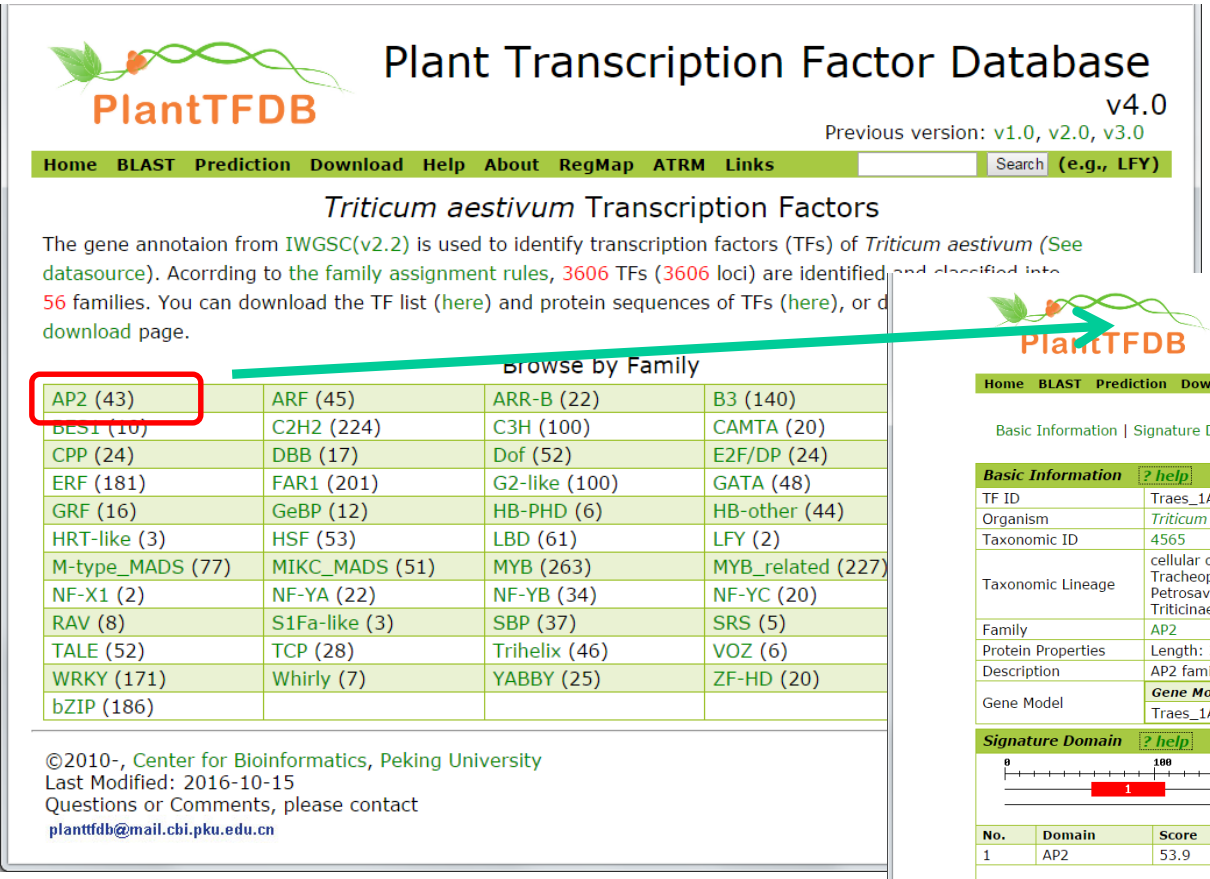
Jin J, et al., **PlantTFDB 3.0:** a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014 Jan;42(Database issue):D1182-7.
165 таксономических групп, из них 38 видов однодольных и 100 видов двудольных

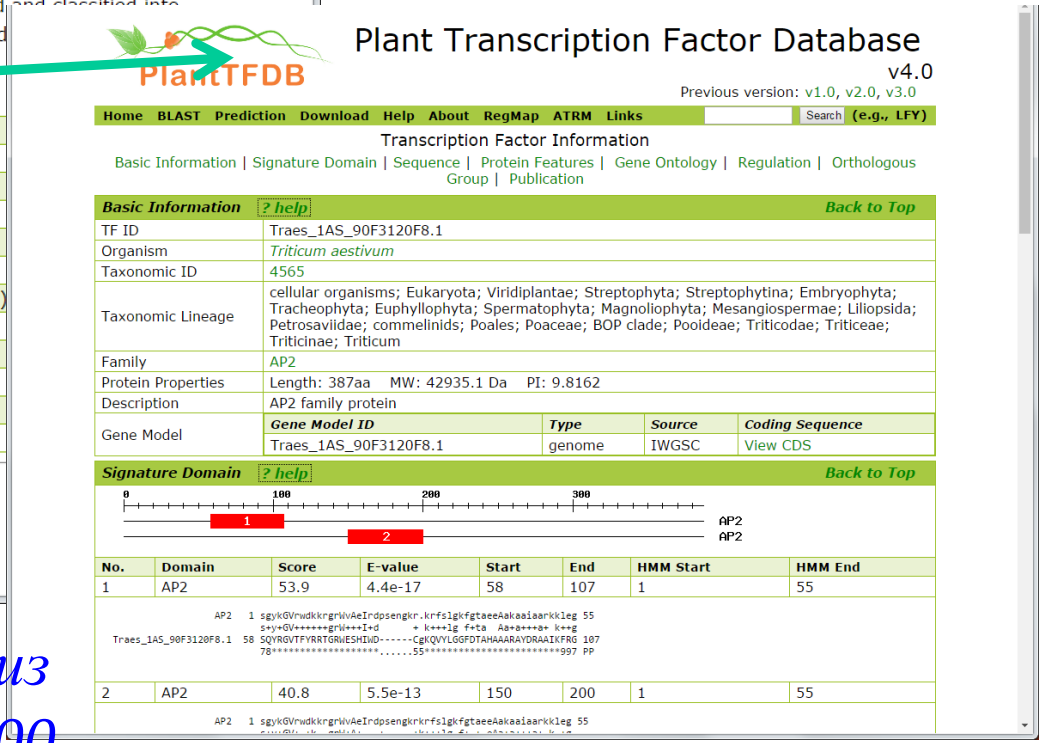
PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw/)

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





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.

Характеристики баз данных = «Критерии полезности баз данных»:

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

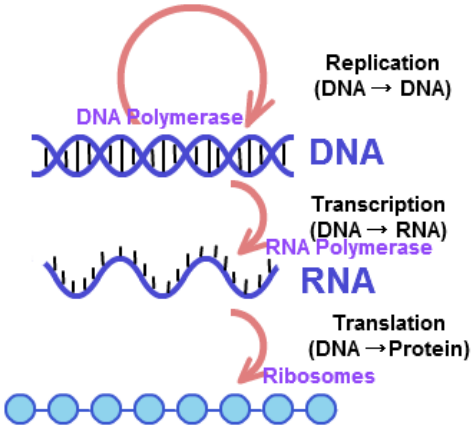
Часть 2. ТРАНСЛЯЦИЯ

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

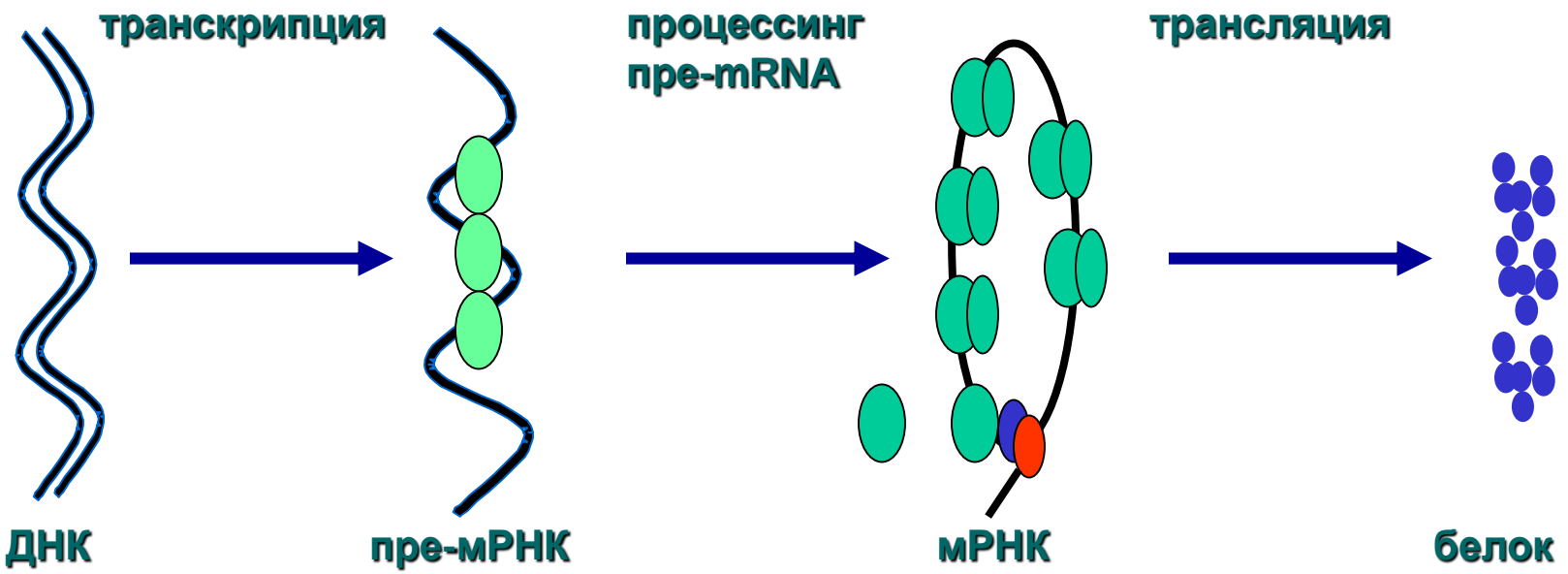
Эта часть лекции подготовлена с использованием материалов,
любезно предоставленных член-корр. РАН, д.б.н. Кочетовым А.В.

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

Трансляция – один из фундаментальных биологических процессов

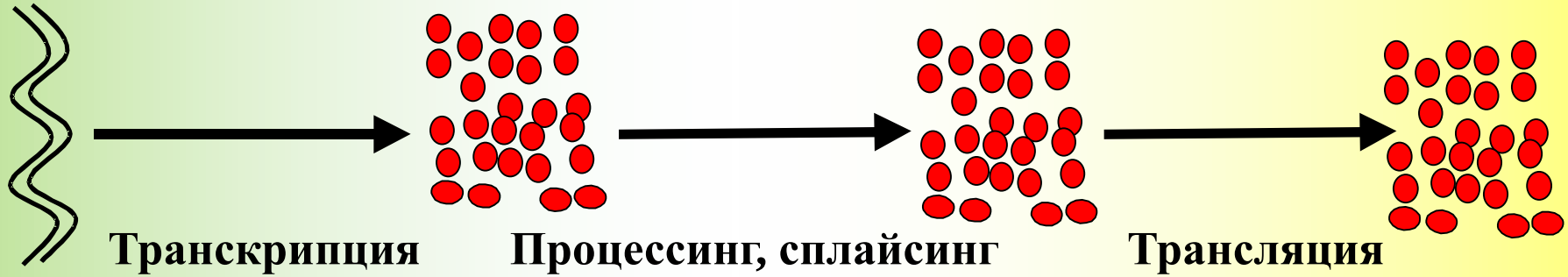


Трансляция –
это процесс перехода генетической информации от мРНК к белку
= процесс синтеза белка из аминокислот на матрице информационной (матричной) РНК, осуществляемый рибосомой.

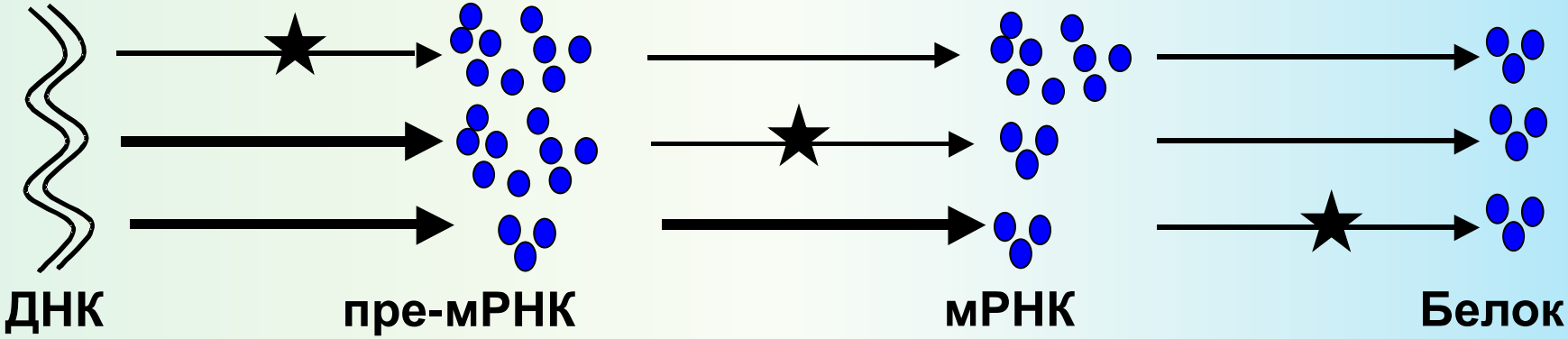


Концепция «лимитирующего звена» в приложении к процессу экспрессии

Высокий уровень экспрессии: высокоэффективны все стадии



Низкий уровень экспрессии (★- низкоэффективные стадии)

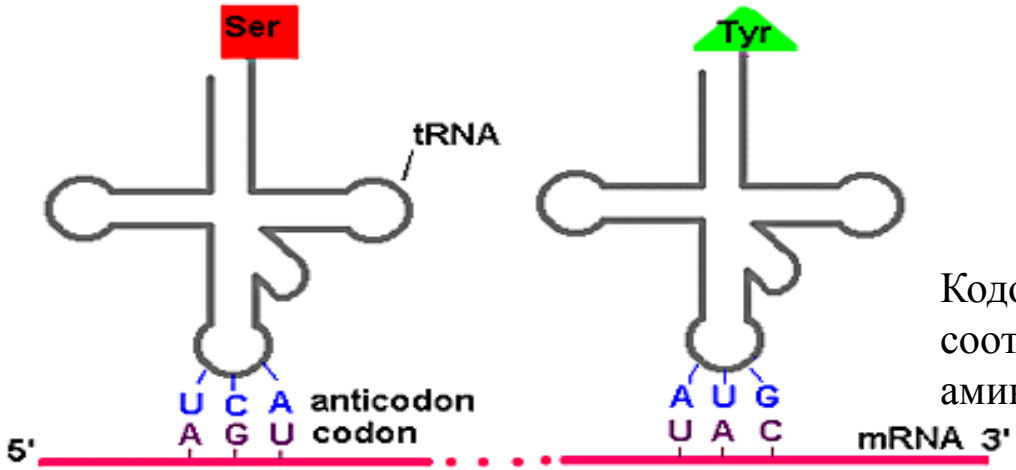


Основные участники процесса трансляции

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

Транспортные РНК (тРНК)

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



Кодон – тринуклеотид, соответствующий либо аминокислоте либо стоп-кодону

Генетический код – соответствие кодонов и аминокислот

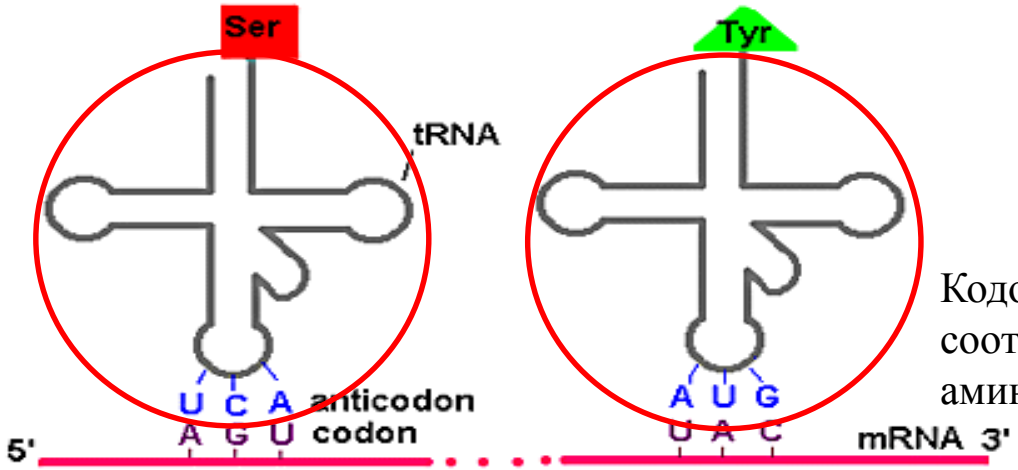
2nd base in codon

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

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

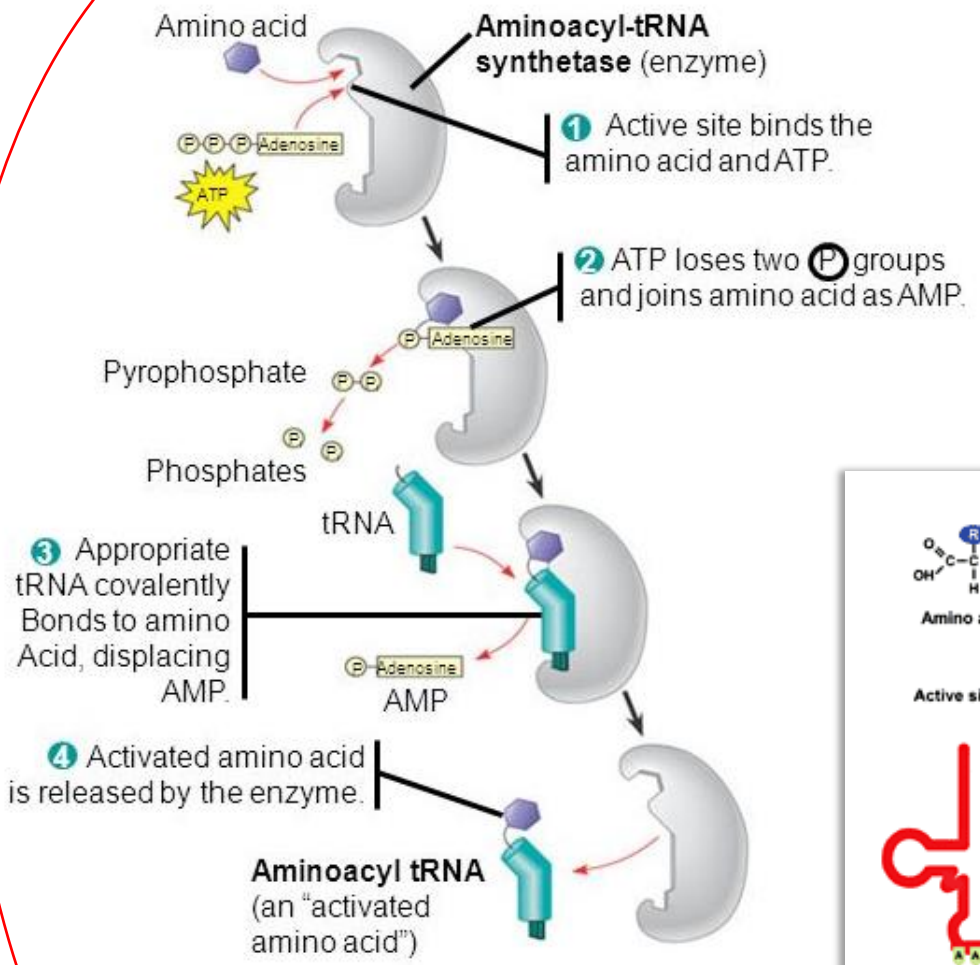
Генетический код – соответствие кодонов и аминокислот

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

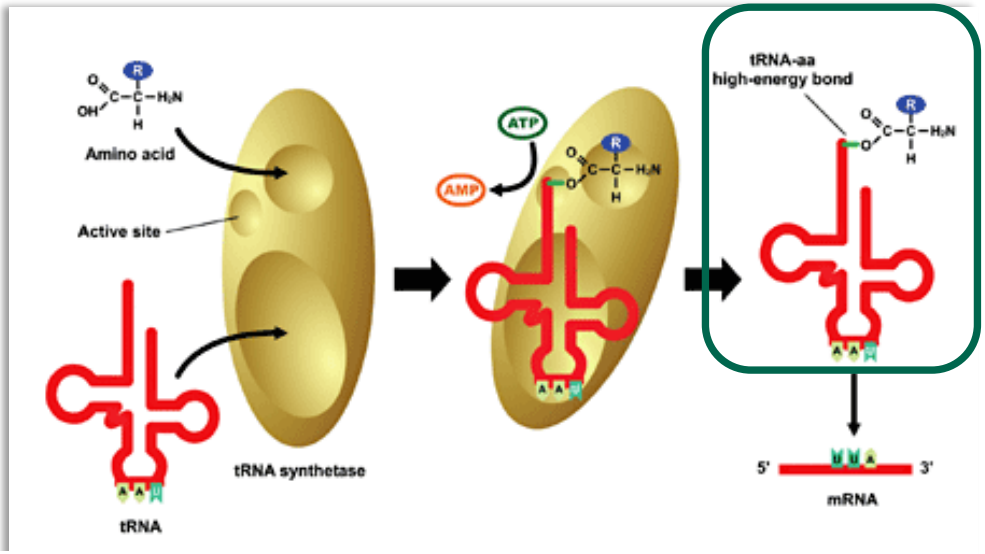
- Число тринуклеотидных комбинаций из 4 нуклеотидов = равно 64 (4^3)
- Три кодона являются нонсенс-кодонами и выполняют функцию терминаторов трансляции
- Каждой из 20 аминокислот соответствует от одного до шести кодонов.
- Кодоны, кодирующие одну аминокислоту, они называются синонимическими

Функция фермента аминоацил-тРНК-синтетазы:



Аминоацил-тРНК

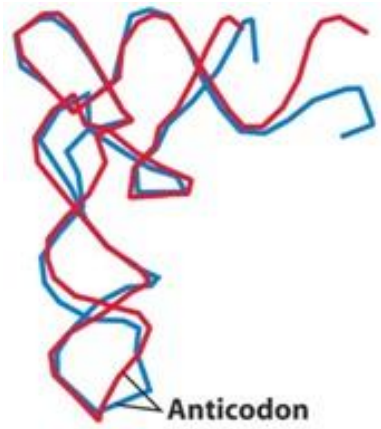
= Аминокислота прикреплена сложноэфирной связью своей карбоксильной группой к 2’ или 3’-гидроксильной группе рибозы 3’-концевого нуклеотида тРНК (это всегда аденин)

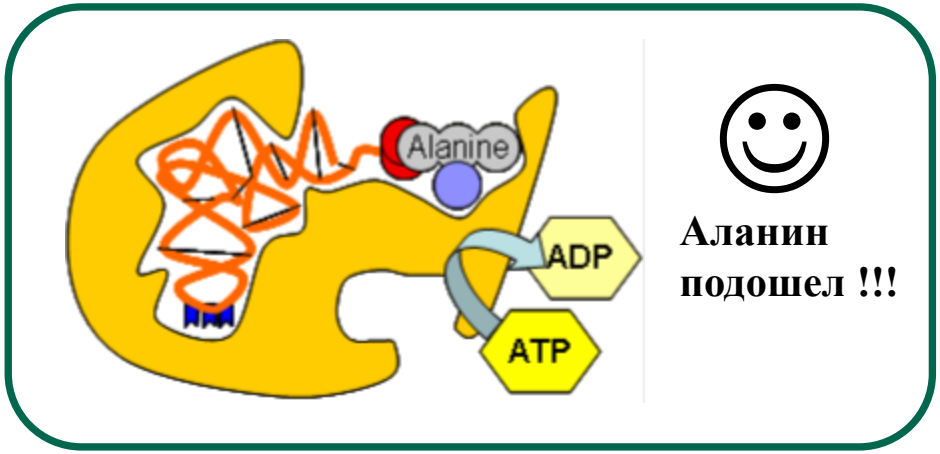


Аминокислота + тРНК + АТФ -> Аминоацил-тРНК + АМФ + 2Р

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

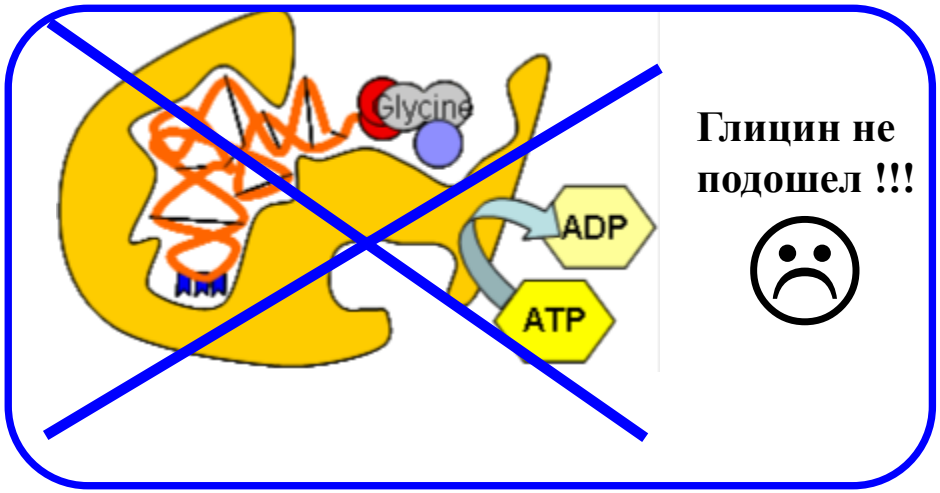
Аминоацил-тРНК-синтетаза: специфичность





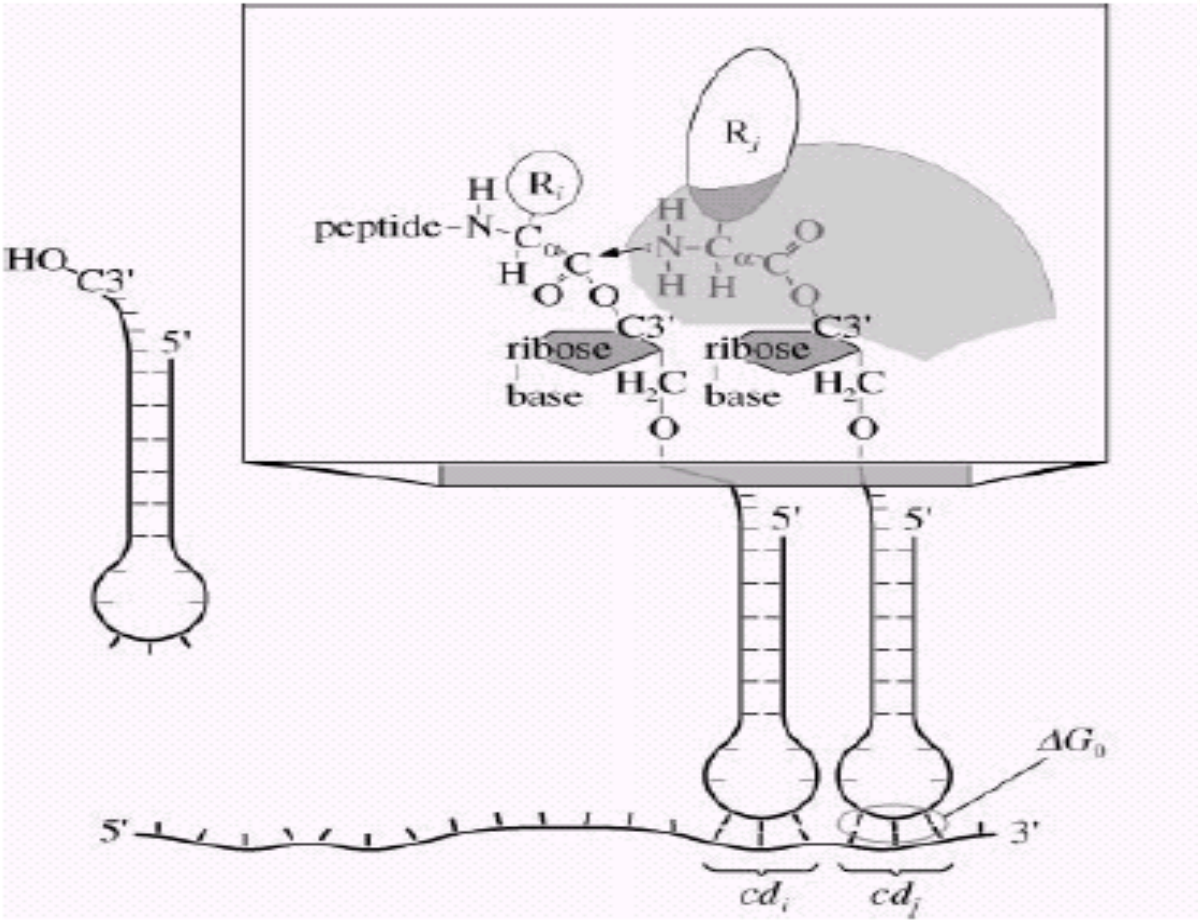
Каждая тРНК имеет уникальную третичную структуру.

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.



Рибосомы

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



Рибосома эукариот 80 S

Большая субъединица 60 S

\+

Малая субъединица 40 S

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

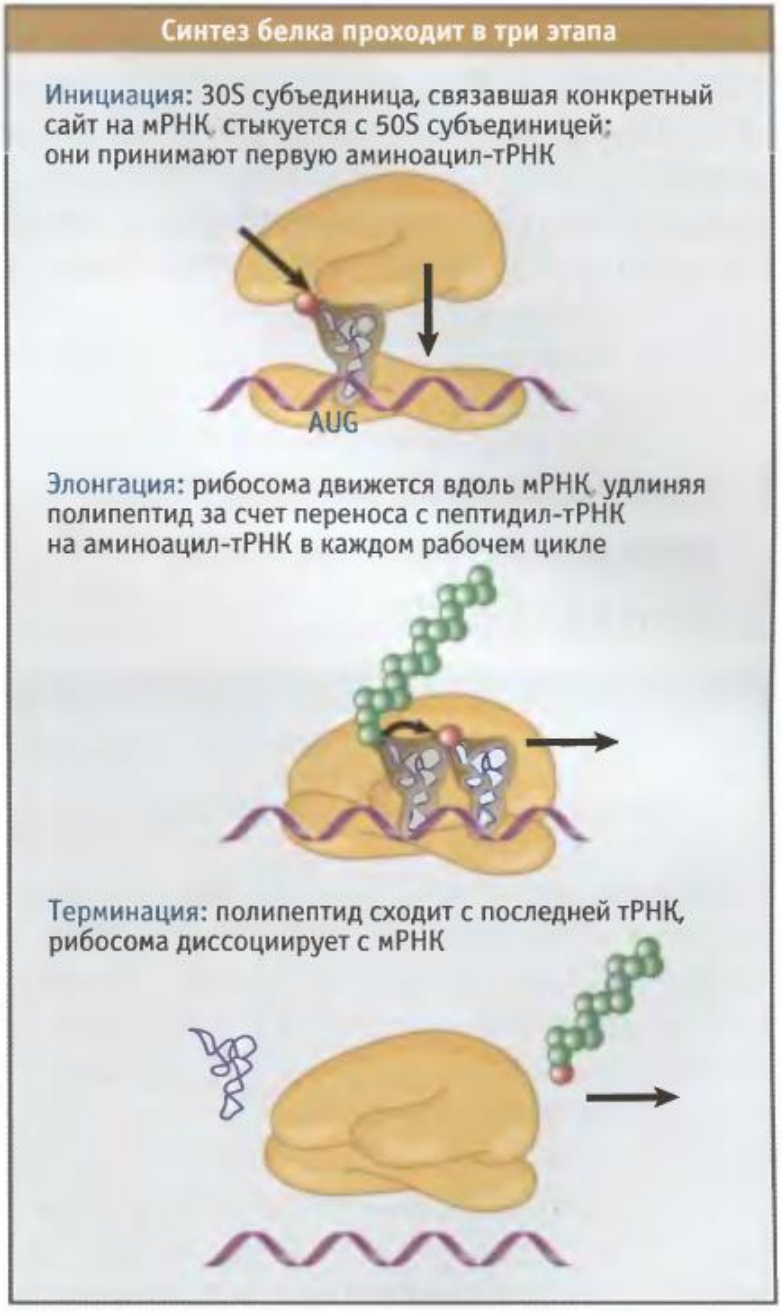
Синтез белка проходит в три этапа: схематическое представление (прокариоты)

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

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

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

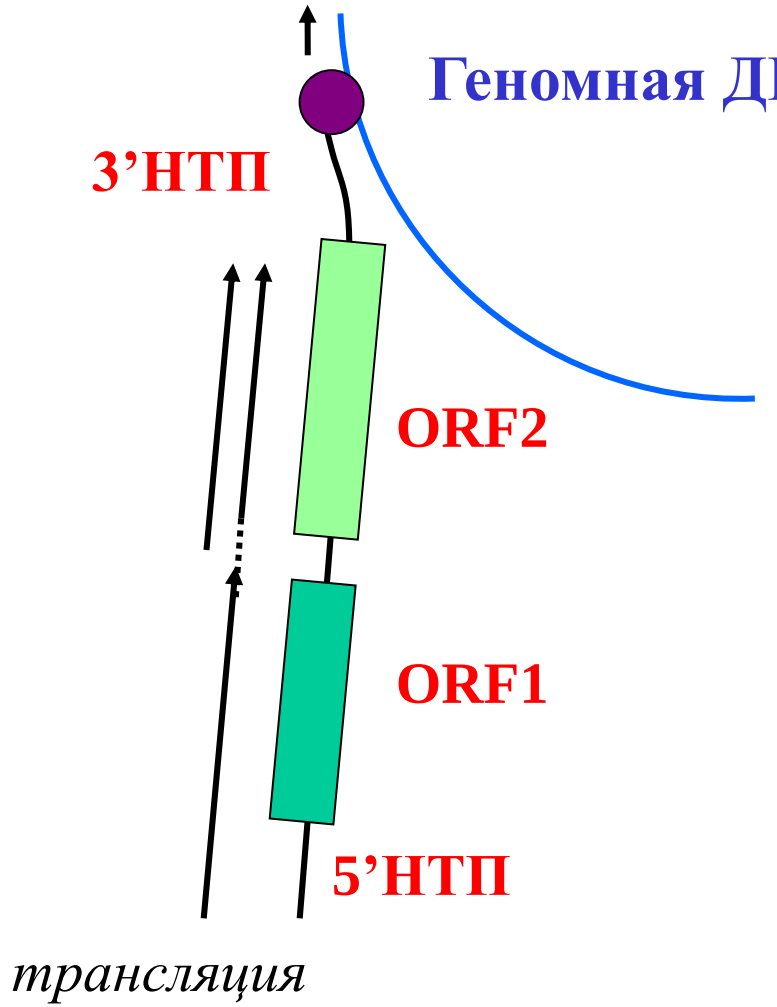
Льюин Б. Гены, 2012 г. С.163

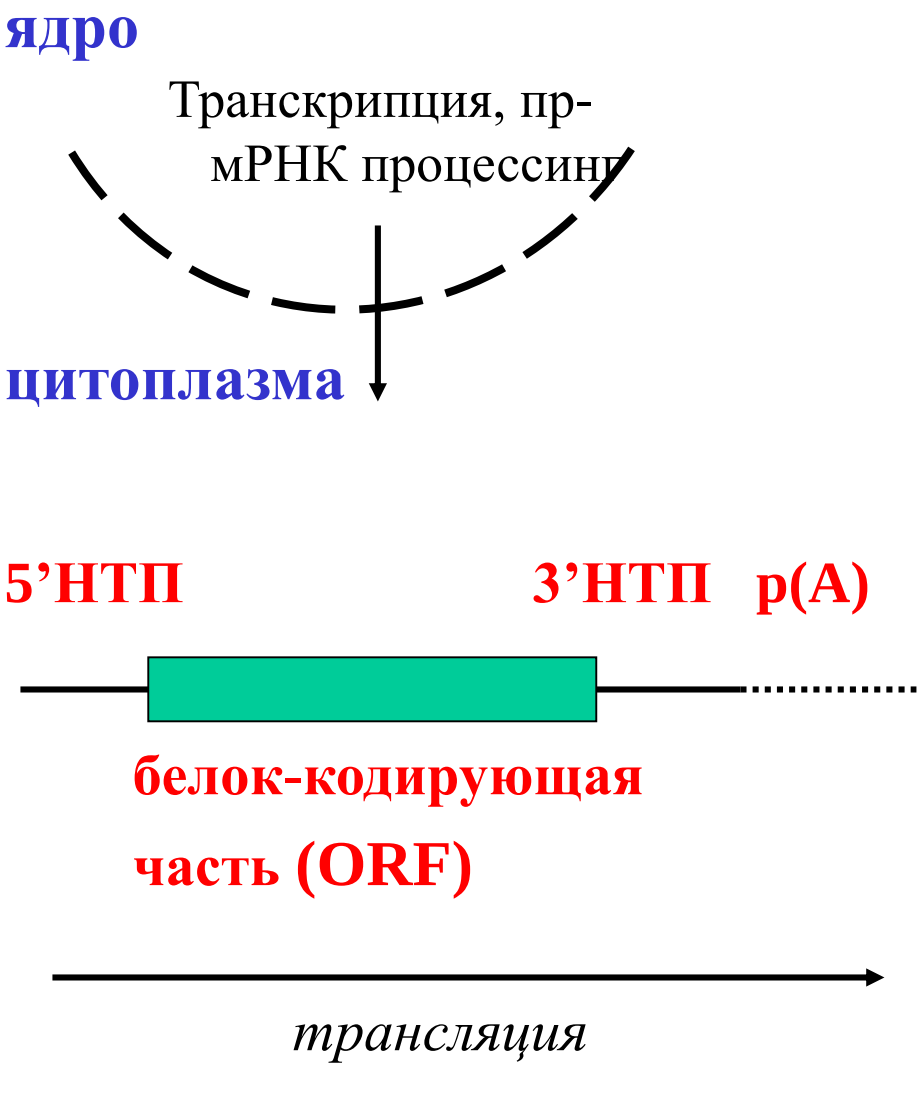


Структуры мРНК и процесс трансляции у прокариот и эукариот имеют свои отличительные особенности.

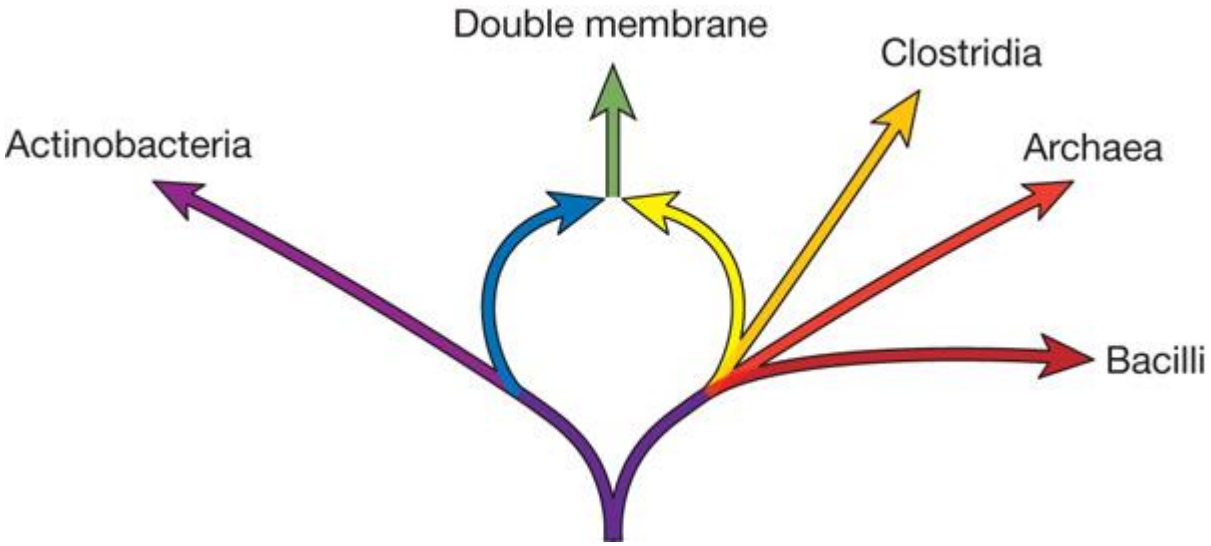
Функционально наиболее сильно отличается процесс инициации трансляции

Функциональные различия между мРНК прокариот и эукариот

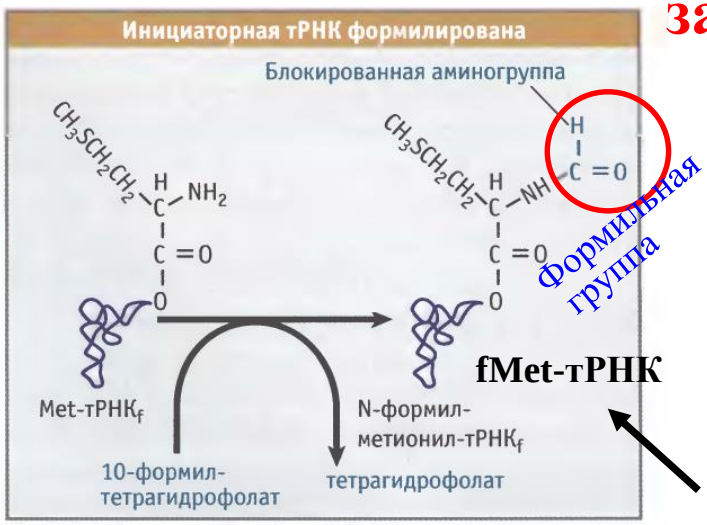




Прокариоты


Double membrane
Clostridia
Actinobacteria
Archaea
Bacilli

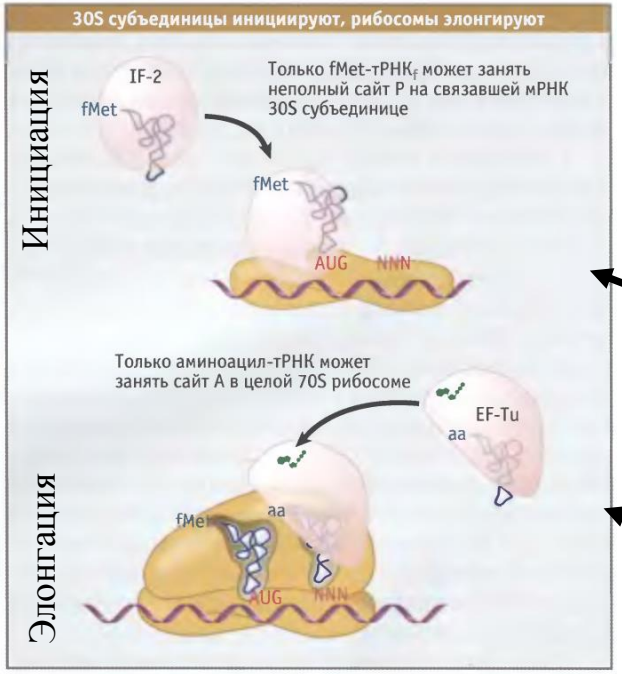
Прокариоты: особая инициаторная тРНК (fMet-тРНК) закладывает первое звено полипептида



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 обозначается как **fMet-тРНК**



Прокариоты: факторы инициации

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

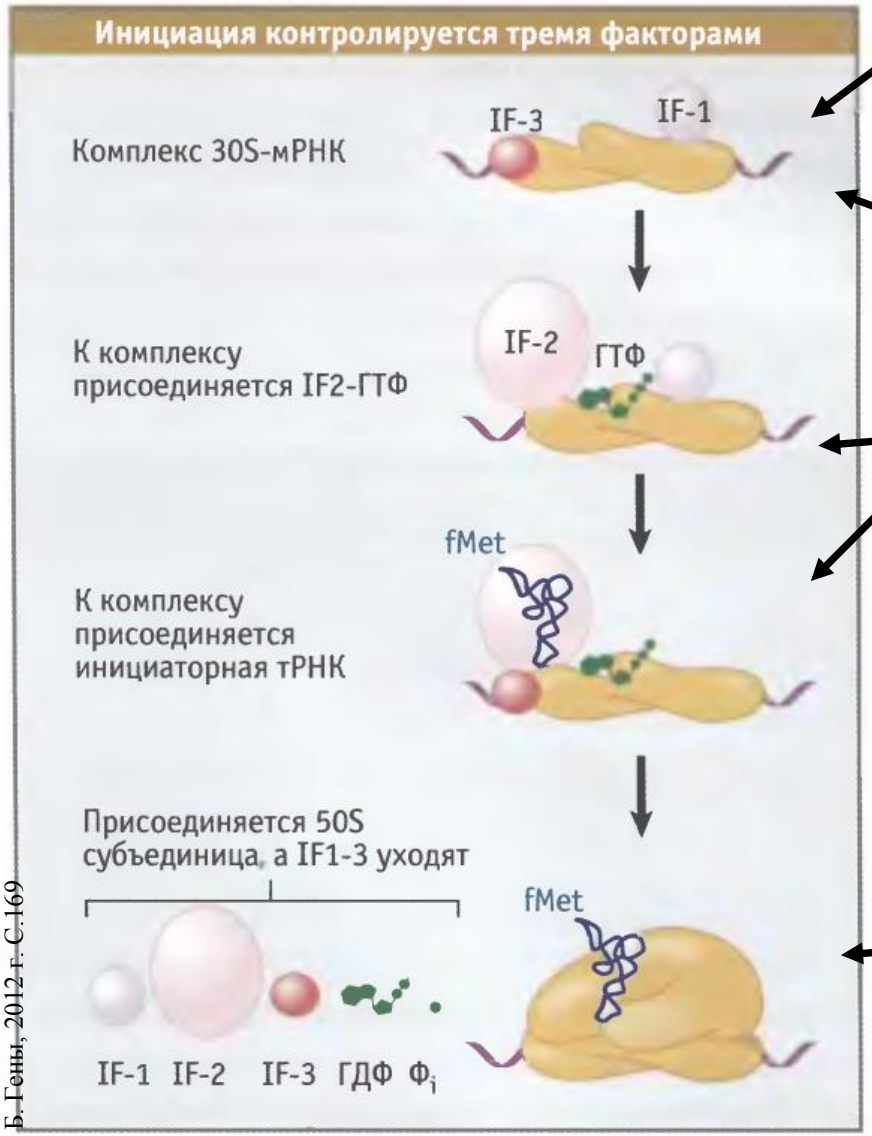
IF-2 связывает инициаторную fMet-тРНК и позволяет ей занять неполный сайт P на малой субъединице рибосомы.

Напротив, в ходе элонгации

EF-Tu, который помещает аминоацил-тРНК в сайт A, совсем не умеет связывать fMet-тРНК , что исключает ее использование.

Льюин Б. Гены, 2012 г. С.167, 168

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.

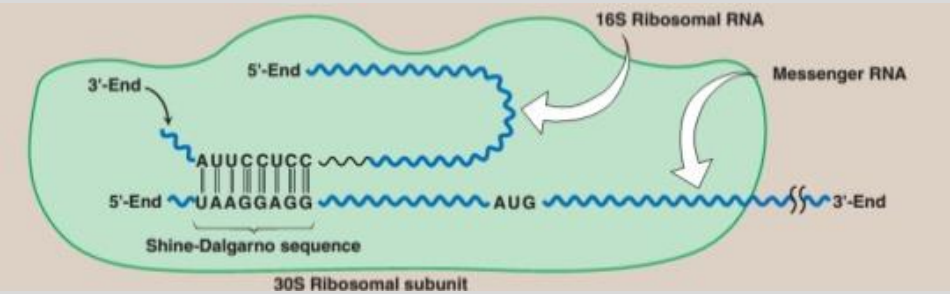
Льюин Б. Гены, 2012 г. С.169

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

- **<u>Прокариотическая</u>** 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

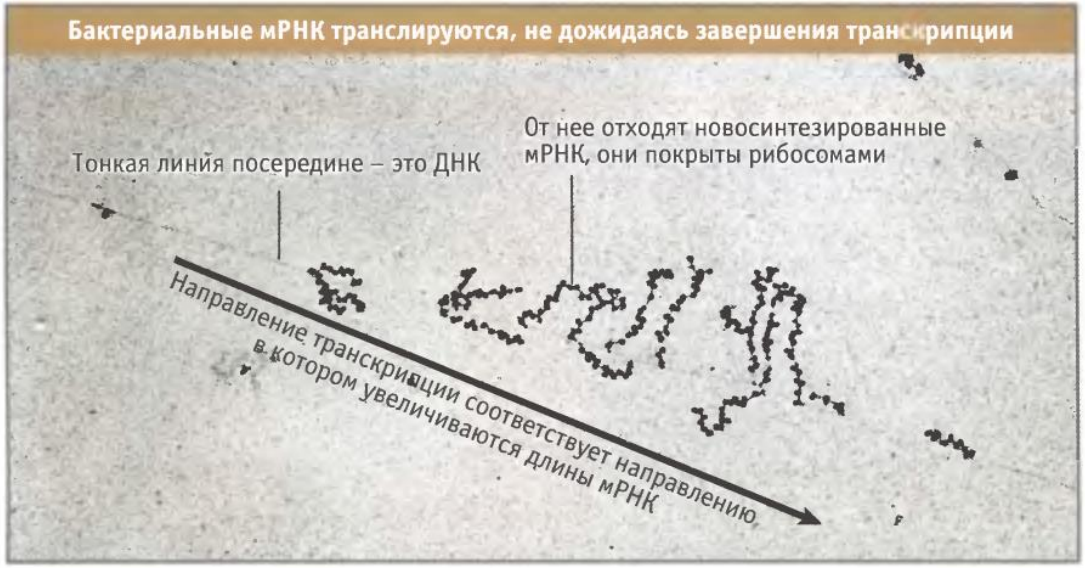
<u>У прокариот</u> в составе нуклеотидной последовательности консервативного 3’-конца 16S рибосомной РНК, входящей в состав 30S субъединицы рибосомы, есть комплементарный участок, который и выполняет роль детектора сигнала Шайна-Дальгарно. Именно комбинация AUG кодона и сайта Шайна-Дальгарно обозначает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



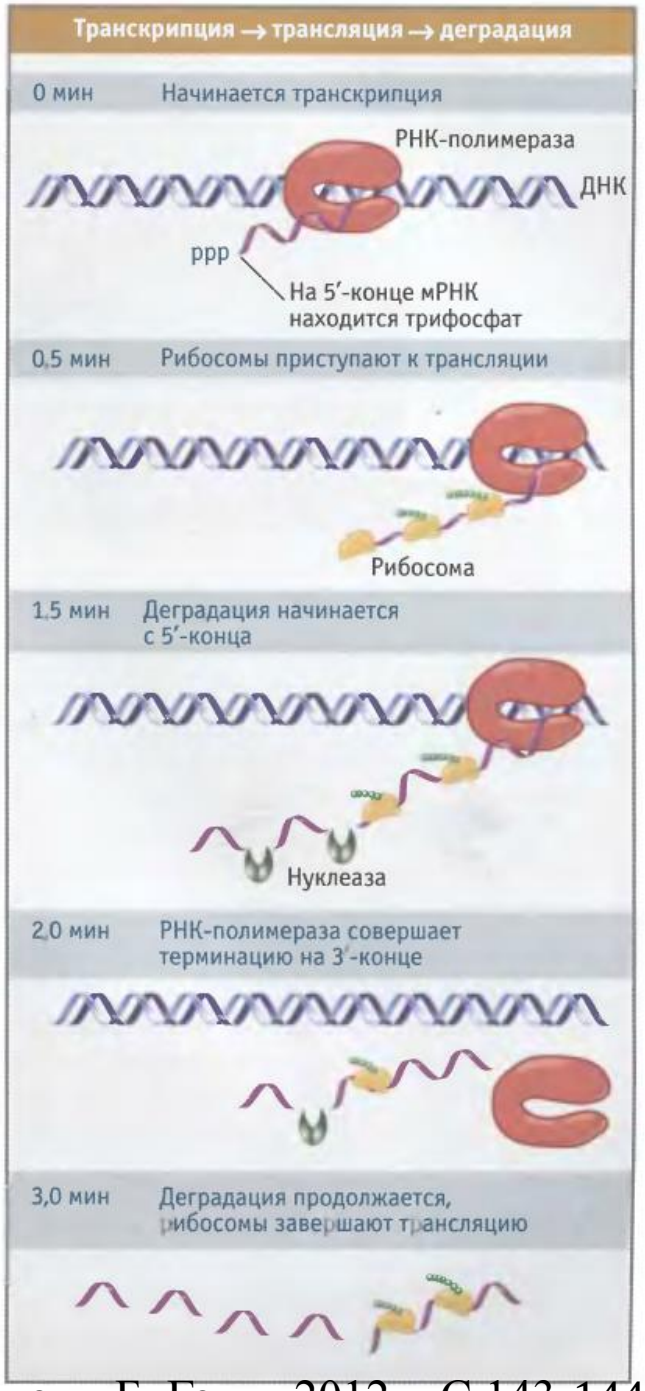
<u>У эукариот</u> нет сайта Ш-Д. У эукариот 40S субъединица рибосомы связывается с 5’-кэпированным концом мРНК, который находится в комплексе с eIF-4 и начинает передвигаться вдоль мРНК до тех пор, пока не встретит AUG кодон.

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



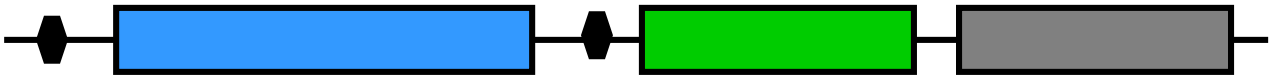
Единицы транскрипции у бактерий можно увидеть



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.**

Особенности процесса трансляции у прокариот

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

Конец лекции