Лекция №8(6)

БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ (окончание)

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

СЕГОДНЯ, в лекции № 8(6) будет дана характеристика баз данных (информационных ресурсов)

ГЕНОМНЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ

UCSC Genome Browser — **США**
University of California, Santa Cruz

ENCODE **Encyclopedia of DNA Elements** — **США**

FANTOM **The Functional Annotation of Mammalian genome** — **Япония**
Rikagaku Kenkyūsho = **RIKEN (**理研) — **+консорциум**

МАТРИЦЫ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

TRANSFAC Matrix, JASPAR , HOCOMOCO , CIS-BP

РЕГУЛЯТОРНЫЕ РАЙОНЫ, ССТФ, ТРАНСКРИПЦИОННЫЕ ФАКТОРЫ

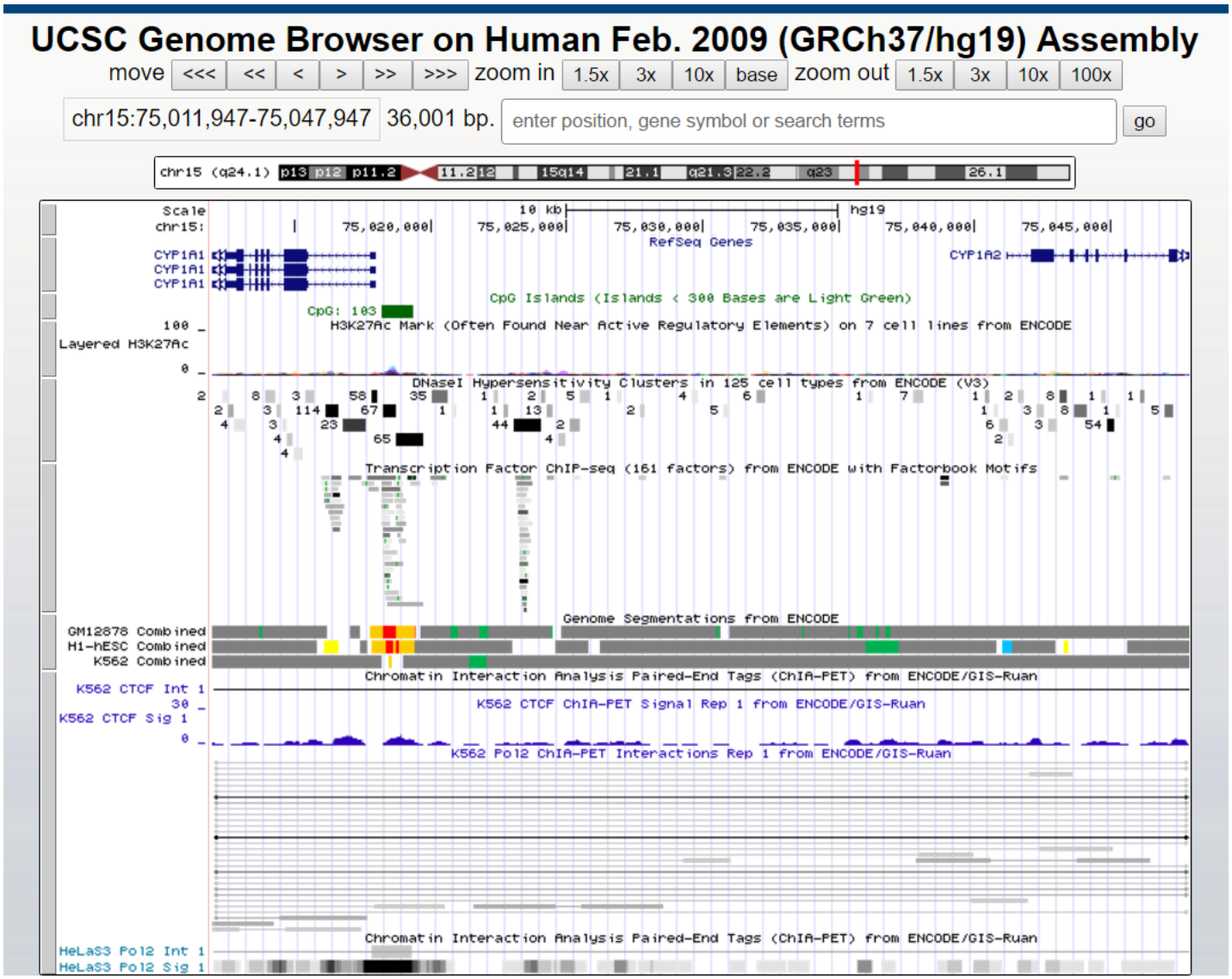
TRRD Transcription Regulatory Regions Database — ИЦиГ СО РАН, Новосибирск, Россия

Геномный браузер Калифорнийского университета в г.Санта Круз

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

UCSC = University of California, Santa Cruz

Геномный браузер Университета г.Санта Круз



CpG – богатые участки

Сайты гиперчувствительности к DNaseI (открытый хроматин)

Районы связывания с различными ТФ по данным экспериментов ChIP-seq

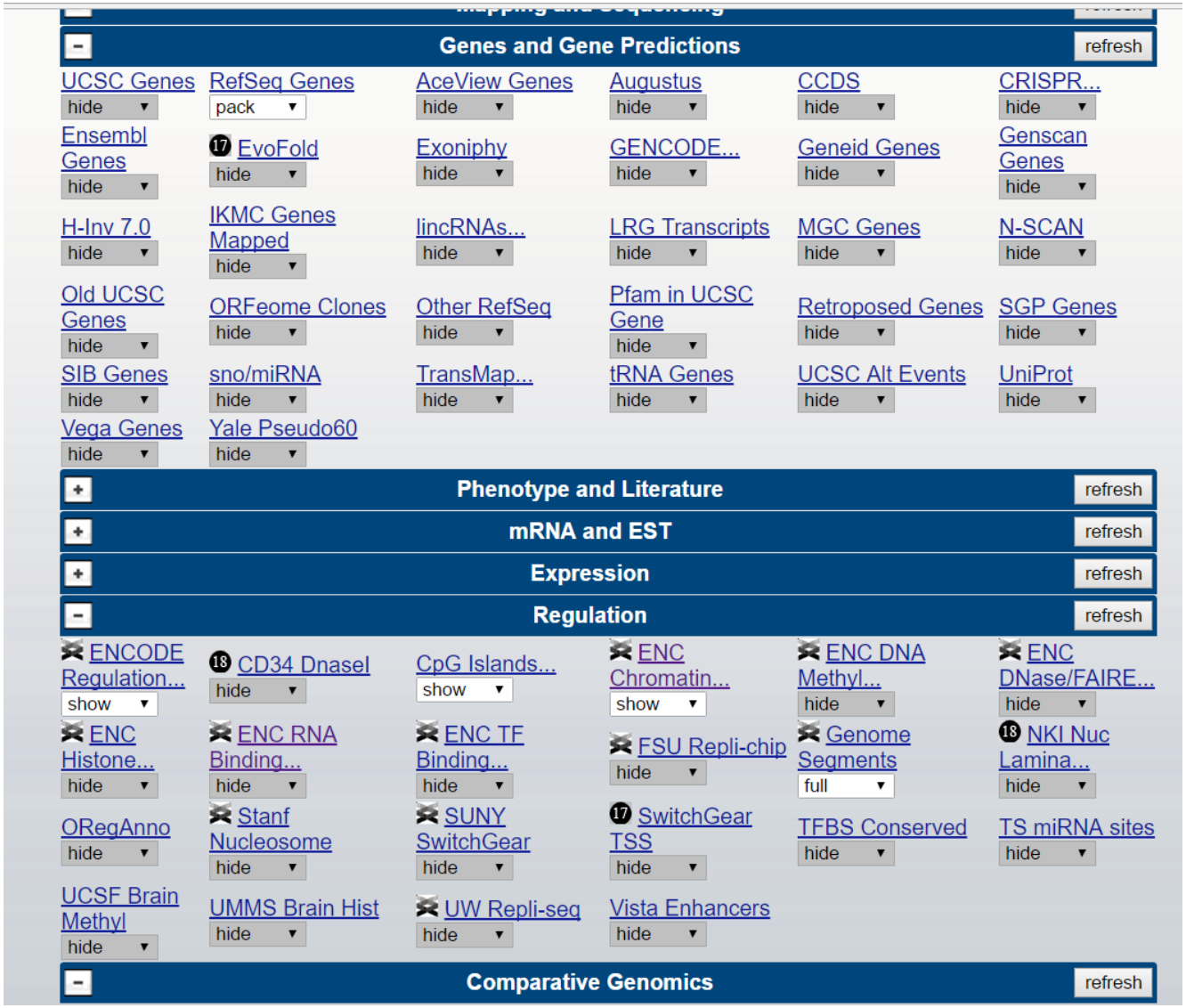
Типы хроматина в различных типах клеток

3D структура хроматина: контакты между удаленными участками генома

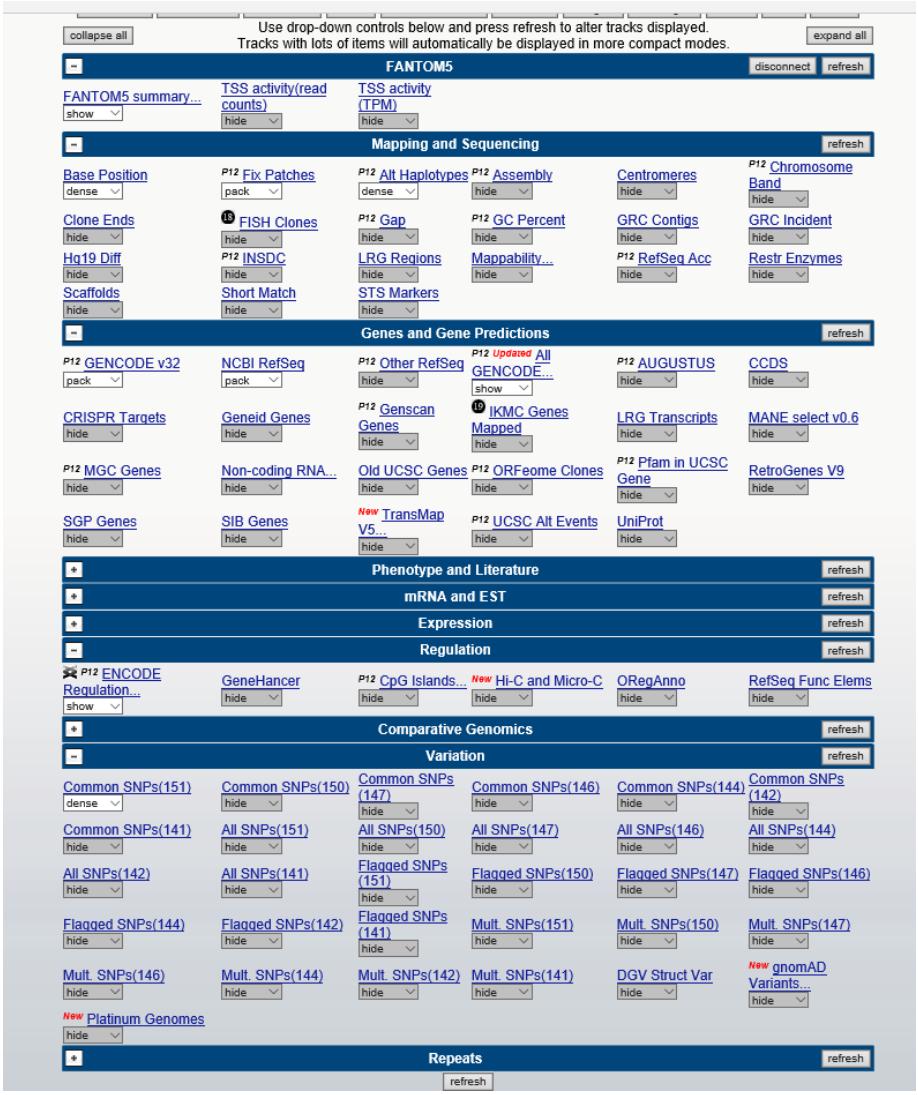
Панель инструментов для настройки отображаемой информации см. на следующем слайде.

Панель инструментов для настройки отображаемой информации в геномном браузере UCSC

(эти настройки использовались, чтобы получить изображение, представленное на предыдущем слайде)



А вот еще….



Панель инструментов для настройки отображаемой информации в геномном браузере UCSC

Геномный браузер Университета г.Санта Круз

Инструмент для загрузки данных в текстовом виде

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Send data to GenomeSpace for use with diverse computational tools. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page.

clade: Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)
group: Regulation **track:** CpG Islands add custom tracks track hubs
table: cpgIslandExt describe table schema
region: genome ENCODE Pilot regions position chr15:75011947-75047947 lookup define regions
identifiers (names/accessions): paste list upload list
filter: create
intersection: create
correlation: create
output format: all fields from selected table Send output to Galaxy GREAT GenomeSpace
output file: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

get output summary/statistics

To reset **all** user cart settings (including custom tracks), click here.

Using the Table Browser

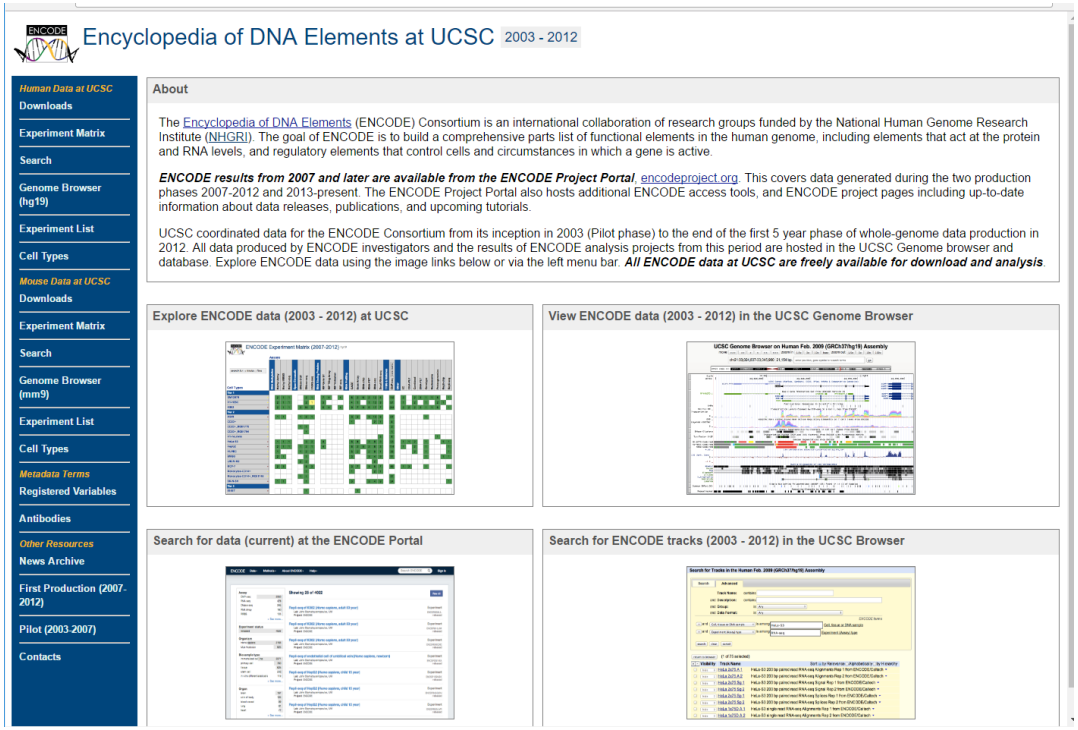
This section provides brief line-by-line descriptions of the Ta
program, see the Table Browser User's Guide.

- **clade:** Specifies which clade the organism is in.

С использованием таких настроек можно получить данные о CpG-богатых участках в районе гена Cyp1A1 (в границах chr15:75011947-75047947)

The Encyclopedia of DNA Elements (ENCODE)

http://genome.ucsc.edu/ENCODE/



Цель проекта - объединение всех данных по функциональным элементам генома человека, включая:
(1) белок- и РНК-кодирующие участки генома,
(2) регуляторные участки

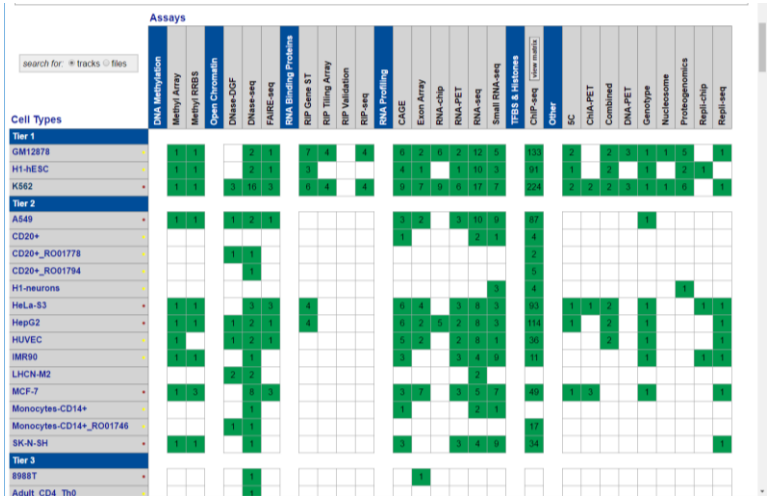
Разрабатывается международным исследовательским консорциумом.
Финансирование идет через National Human Genome Research Institute (NHGRI).

- Результаты, полученные с 2003 по 2012 гг. доступны через геномный браузер Калифорнийского университета (г.Санта Круз)
- Результаты, полученные с 2007 г и позднее доступны на сайте проекта Encode (encodeproject.org).

Виды информации в ENCODE:

ENCODE включает данные об участках генома **человека**, имеющих важные функциональные характеристики:

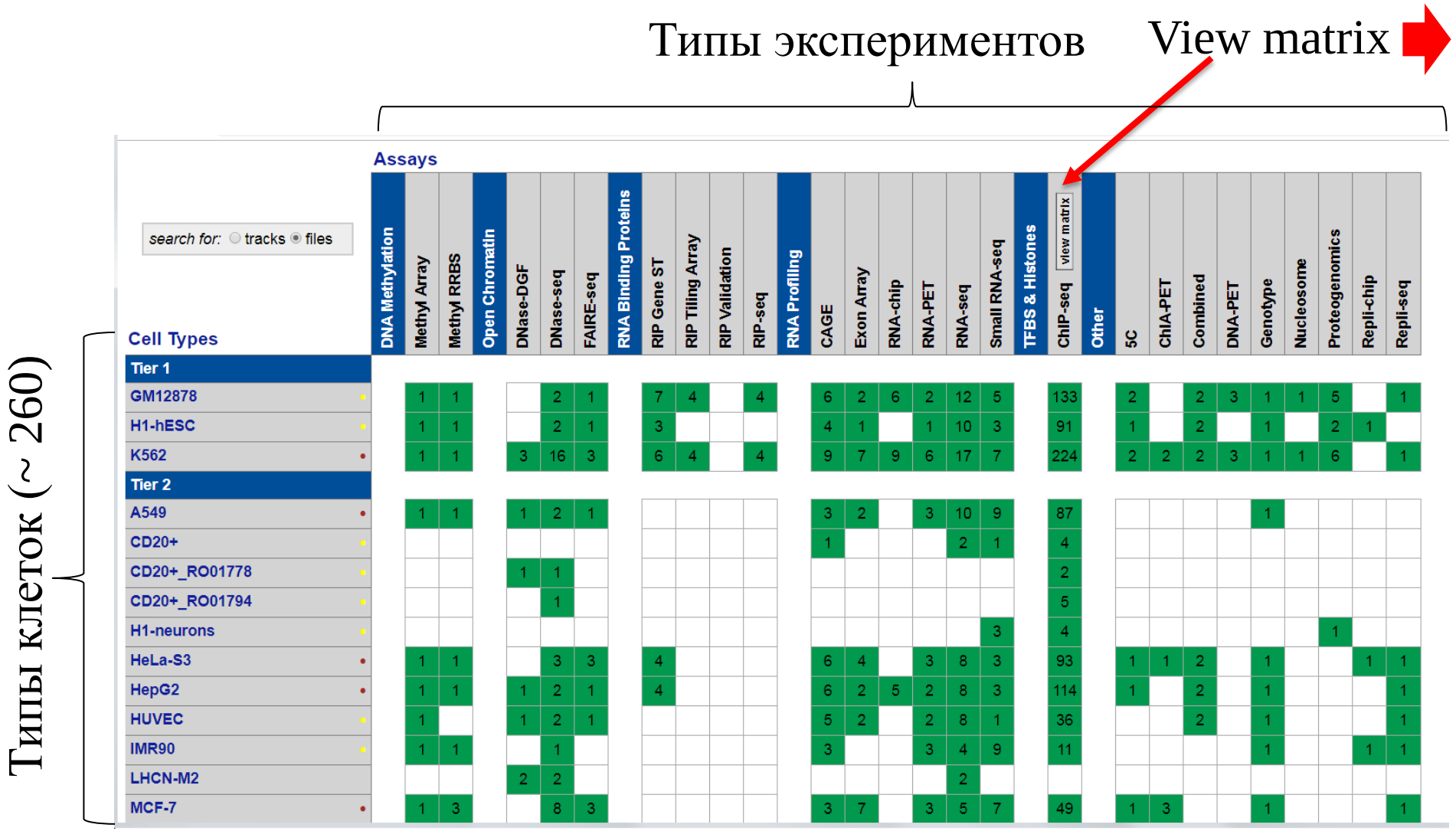
- районы ДНК, связанные с модифицированными гистонами (всего 12 различных типов модификаций в 46 различных типах клеток);
-сайты, гиперчувствительные к действию ДНКазы I (исследовано 125 линий клеток и выявлено 2.89 млн. сайтов) ;
- профили метилирования (1.2 млн. CpGs островов в каждой из 82 клеточных линий либо тканей) ;
-районы взаимодействия с ДНК-связывающими белками и компонентами РНК-полимераз в 72 типах клеток
- (всего исследовано связывание
- 119 различных белков ,
-87 (73%) из которых являются
- транскрипционными факторами).



Доступ к информации через сайт , http://genome.ucsc.edu/ENCODE/, либо через **геномный браузер Университета г.Санта Круз**

Матрица экспериментов проекта ENCODE (September 2007 to July 2012):

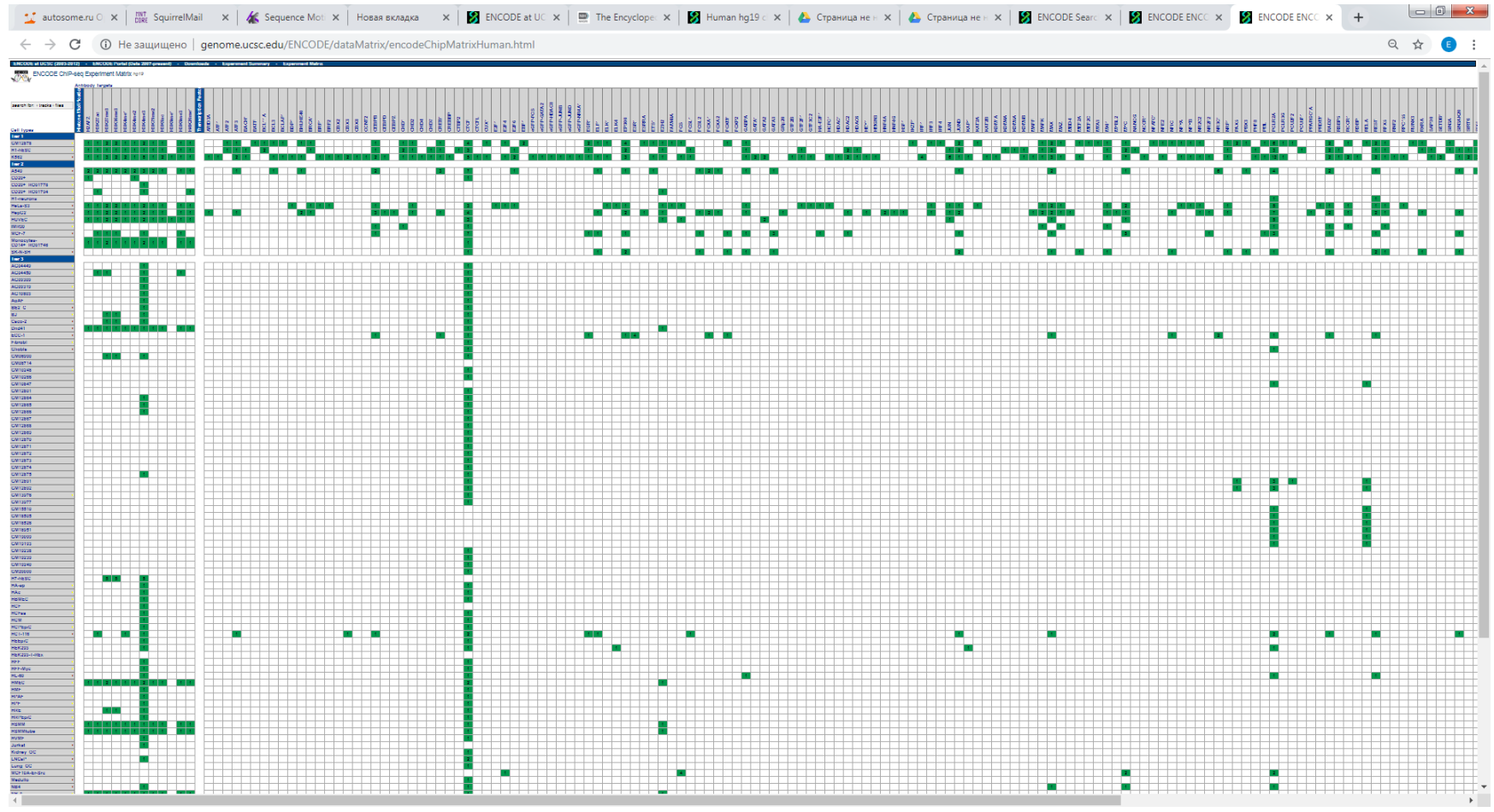
http://genome.ucsc.edu/ENCODE/dataMatrix/encodeDataMatrixHuman.html



Эксперимент определяется как биохимическое исследование и последующий анализ данных, выполненные на одном типе клеток в одной лаборатории. Данные эксперимента представляются в браузере в нескольких видах (enrichment signal graph, peak calls) и доступны для скачивания в разных форматах (. sequence alignments in BAM format, signal graph in bigWig format)

ENCODE ChIP-seq Experiment Matrix *hg19*

http://genome.ucsc.edu/ENCODE/dataMatrix/encodeChipMatrixHuman.html



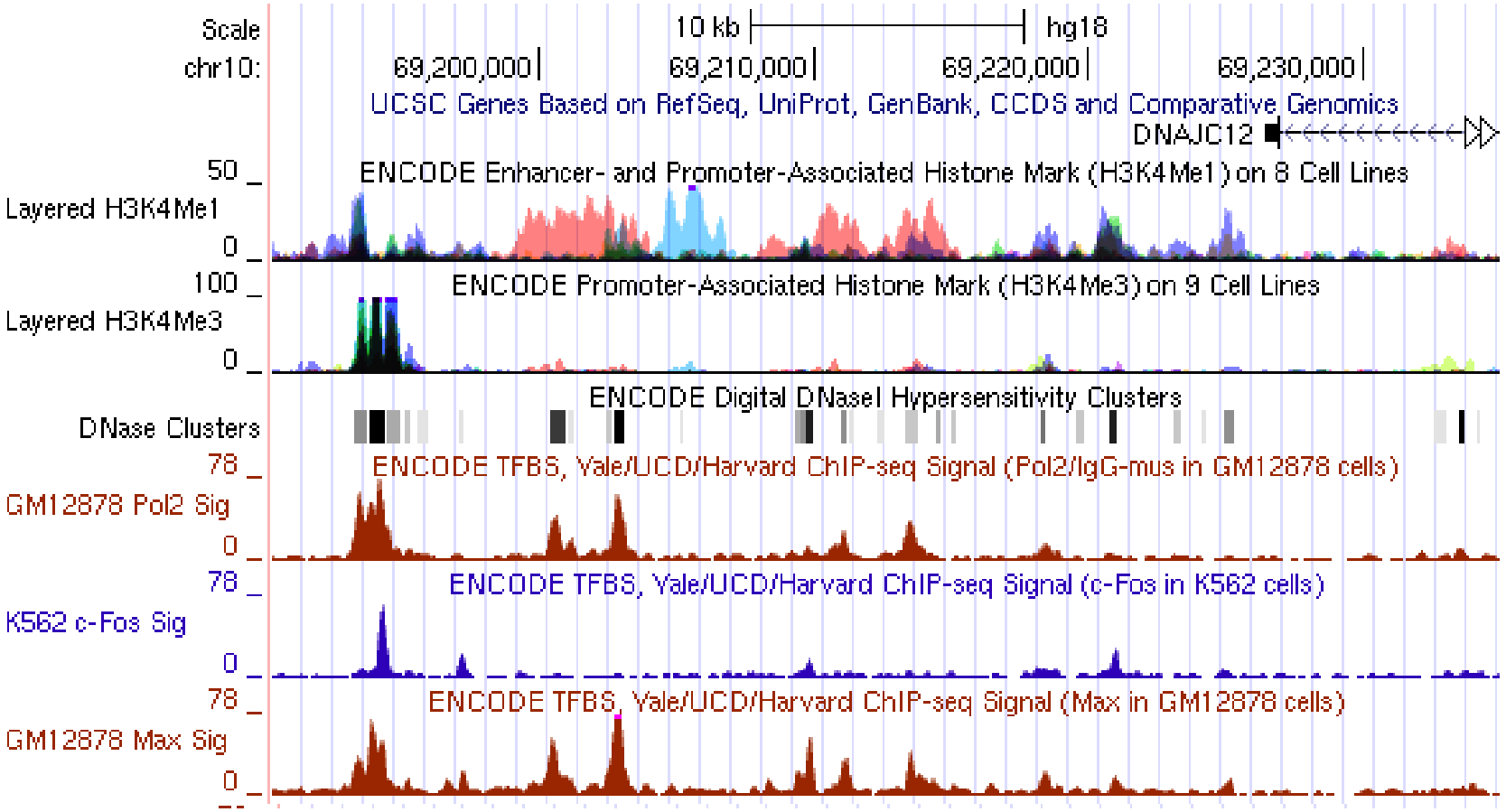
Типы клеток (~ 260)

Histone Modification
(12)

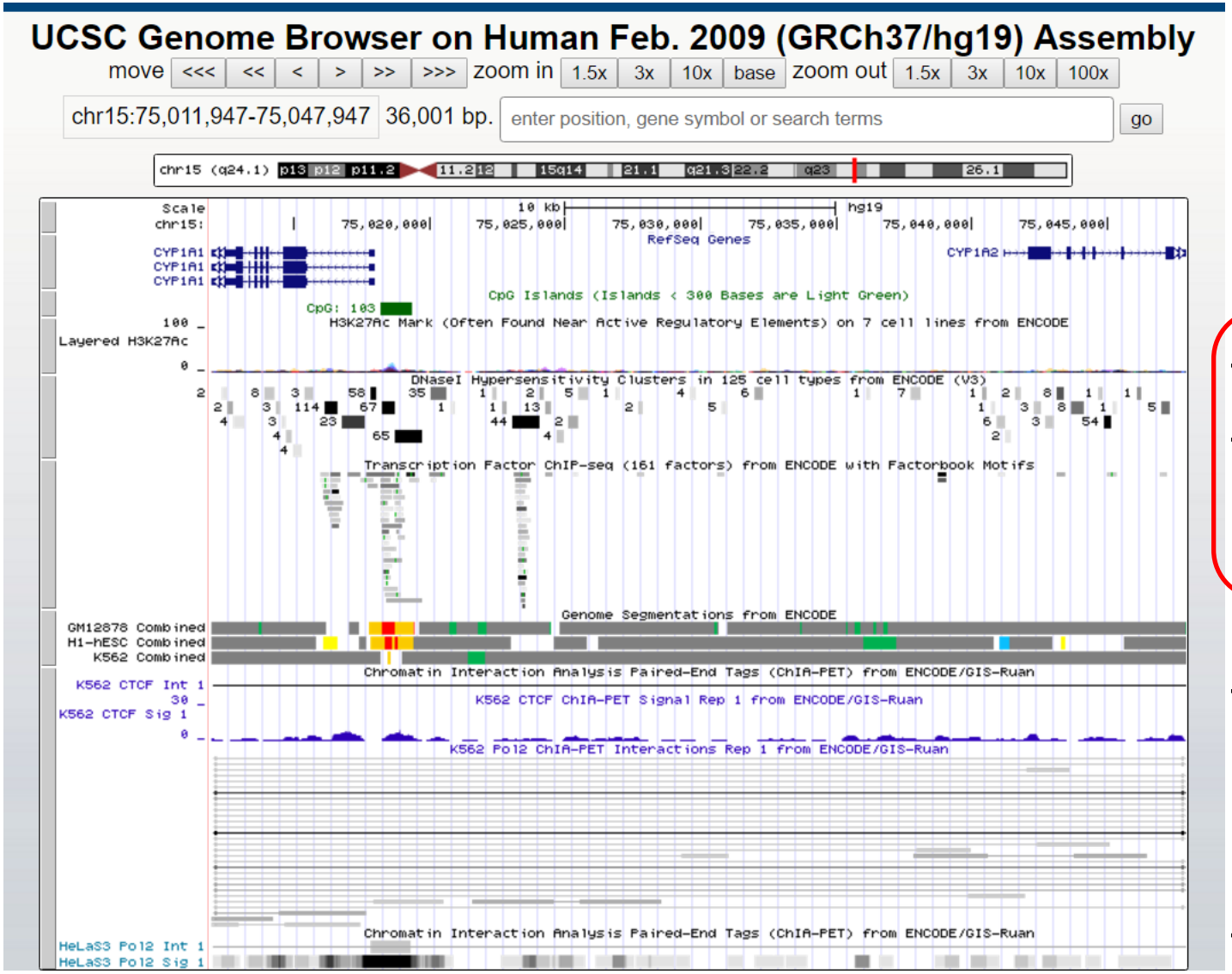
Transcription Factor
(~160)

Визуализация данных проекта ENCODE в геномном браузере Университета г.Санта Круз (UCSC).

(участок 10 хромосомы (chr10:69,190,281-69,235,000))



Геномный браузер Университета г.Санта Круз



CpG – богатые участки

Сайты гиперчувствительности к DNaseI (открытый хроматин)

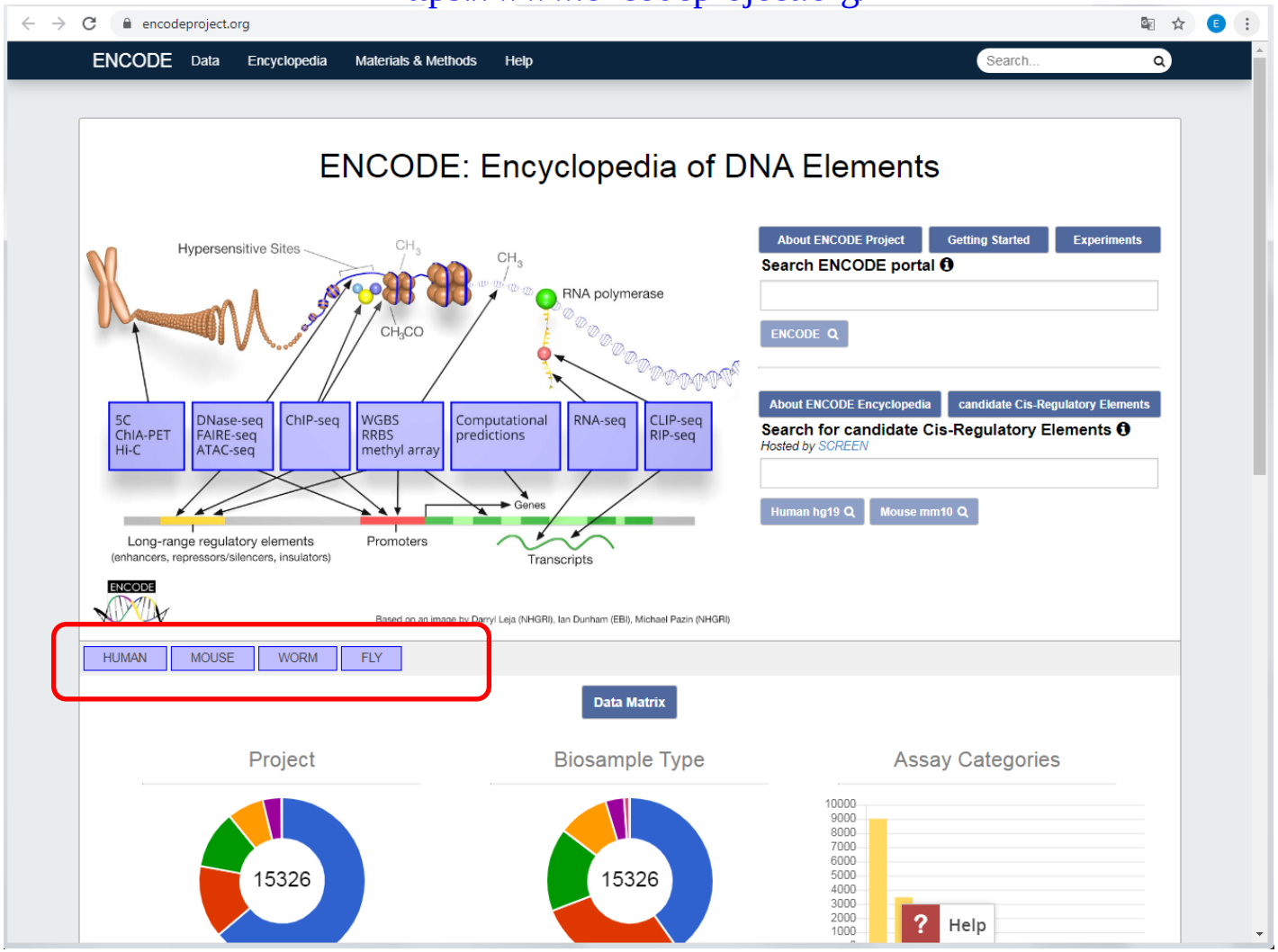
Районы связывания с различными ТФ по данным экспериментов ChIP-seq

Типы хроматина в различных типах клеток

3D структура хроматина: контакты между удаленными участками генома

Результаты, полученные в проекте ENCODE с 2007 г и позднее доступны на сайте проекта Encode

https://www.encodeproject.org/

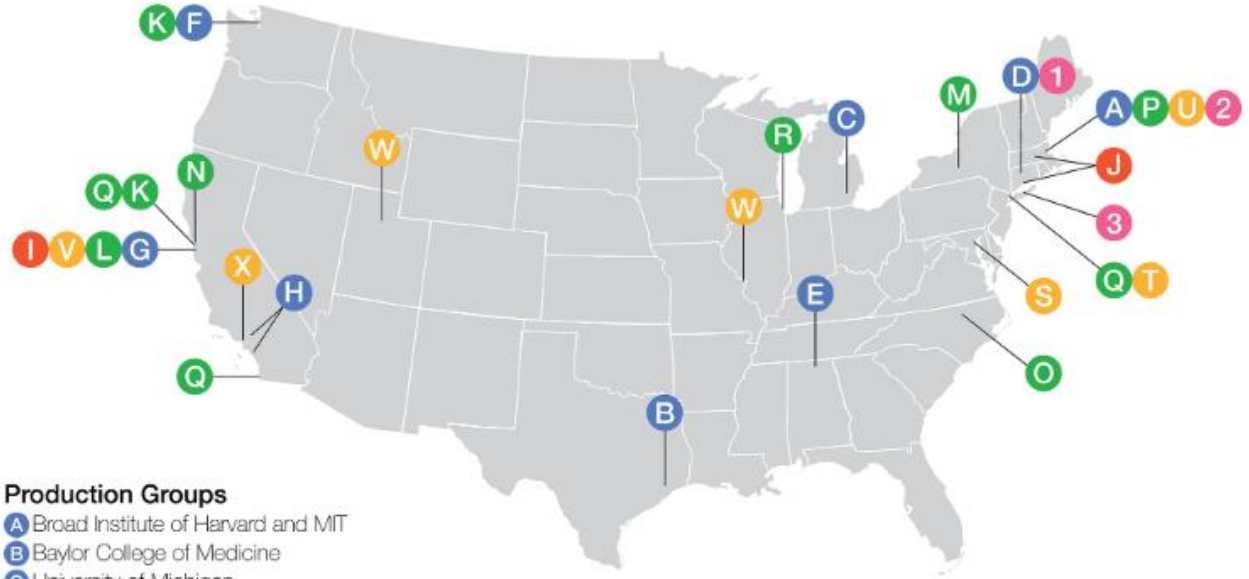


Есть еще проект ModENCODE. Он нацелен на исследование функциональных элементов *C. elegans* and *D. melanogaster*.

https://www.encodeproject.org/help/project-overview/

Current ENCODE Consortium Participants and Projects

The current ENCODE Consortium (ENCODE 4) is composed primarily of scientists funded under Requests for Application (RFAs) released by NHGRI in 2011. NHGRI provides a list of current ENCODE projects and participants. The ENCODE External Consultants Panel oversees the activities of the Consortium and provides advice and feedback on the Consortium's goals, progress and membership.



Production Groups

- A Broad Institute of Harvard and MIT
- B Baylor College of Medicine
- C University of Michigan
- D The Jackson Laboratory
- E HudsonAlpha Institute for Biotechnology, University of Alabama in Huntsville
- F Altius Institute for Biomedical Sciences
- G Stanford University
- H California Institute of Technology, University of California, Irvine

Data Coordination Center

- I Stanford University

Data Analysis Center

- J University of Massachusetts Medical School; Yale University

Characterization Centers

- K University of California, San Francisco; University of Washington
- L Stanford University
- M Cornell University
- N Lawrence Berkeley National Laboratory
- O Duke University
- P Broad Institute of Harvard and MIT
- Q University of California, San Francisco; University of California, San Diego; Ludwig Institute for Cancer Research
- R University of Chicago

Computational Analysis Groups

- S Johns Hopkins University
- T Memorial Sloan Kettering Cancer Center
- U Harvard University; Brigham and Women's Hospital
- V Stanford University
- W Washington University; University of Utah
- X University of California, Los Angeles

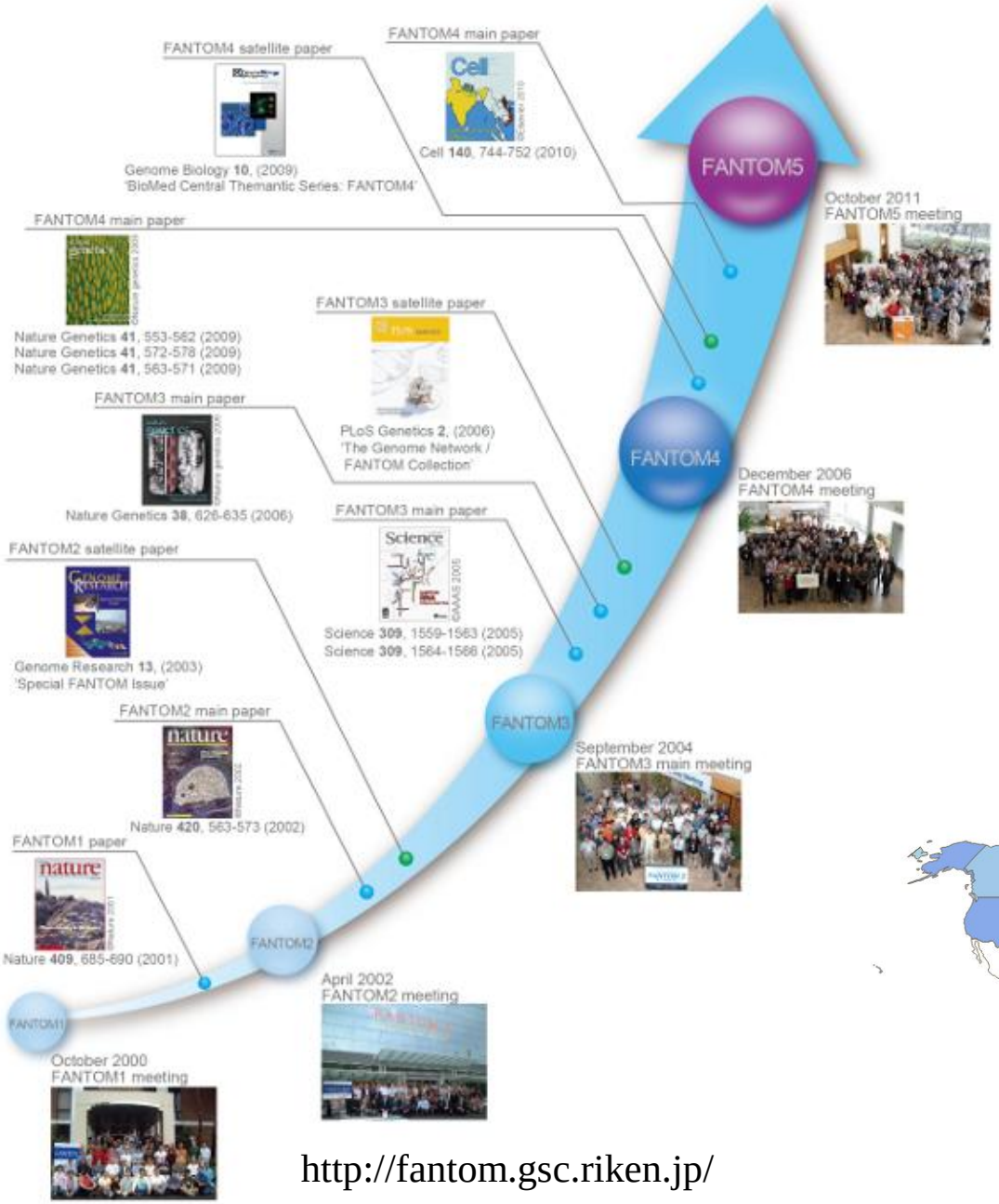
Affiliated Groups

- 1 University of Connecticut Health Center
- 2 Dana-Farber Cancer Institute
- 3 Cold Spring Harbor Laboratory

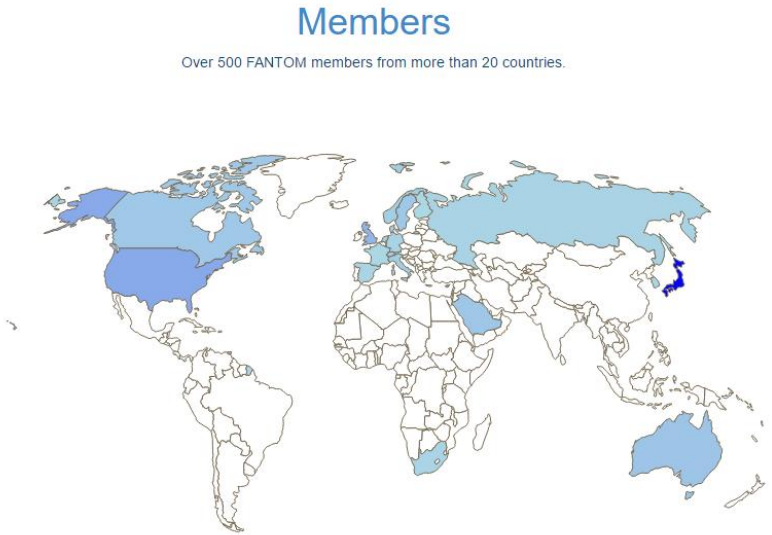
Map of participating investigators and affiliates for ENCODE 4.

? He

FANTOM = The Functional Annotation of Mammalian genome



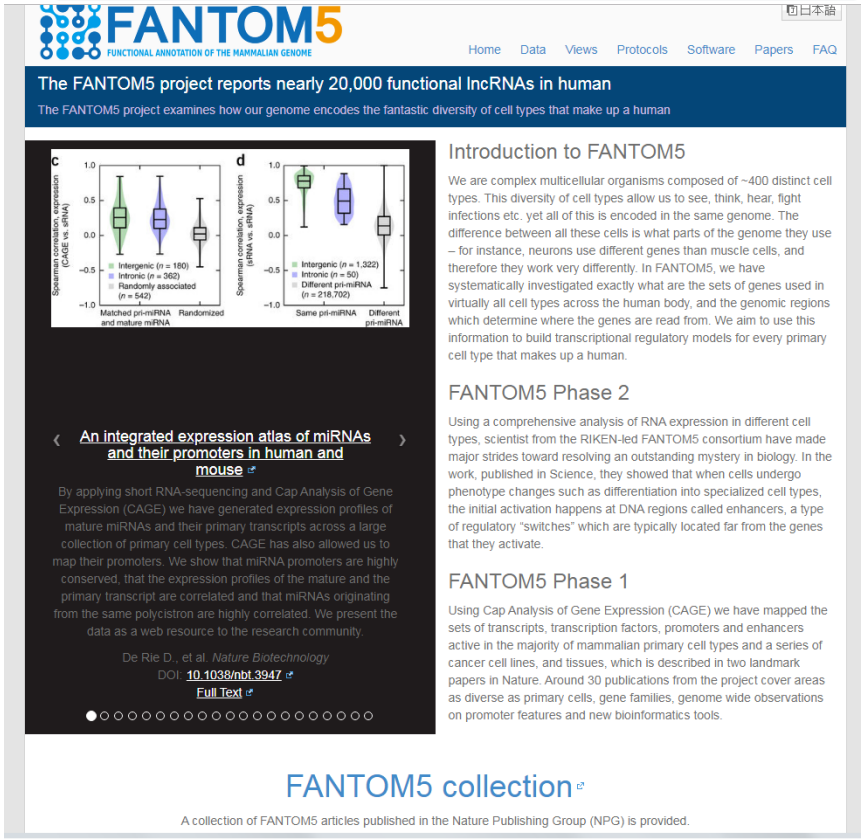
Консорциум был организован в 2000 году. В 2009 году 2009 консорциум объединял работу 51 института (**Япония** + другие страны)



http://fantom.gsc.riken.jp/

FANTOM5

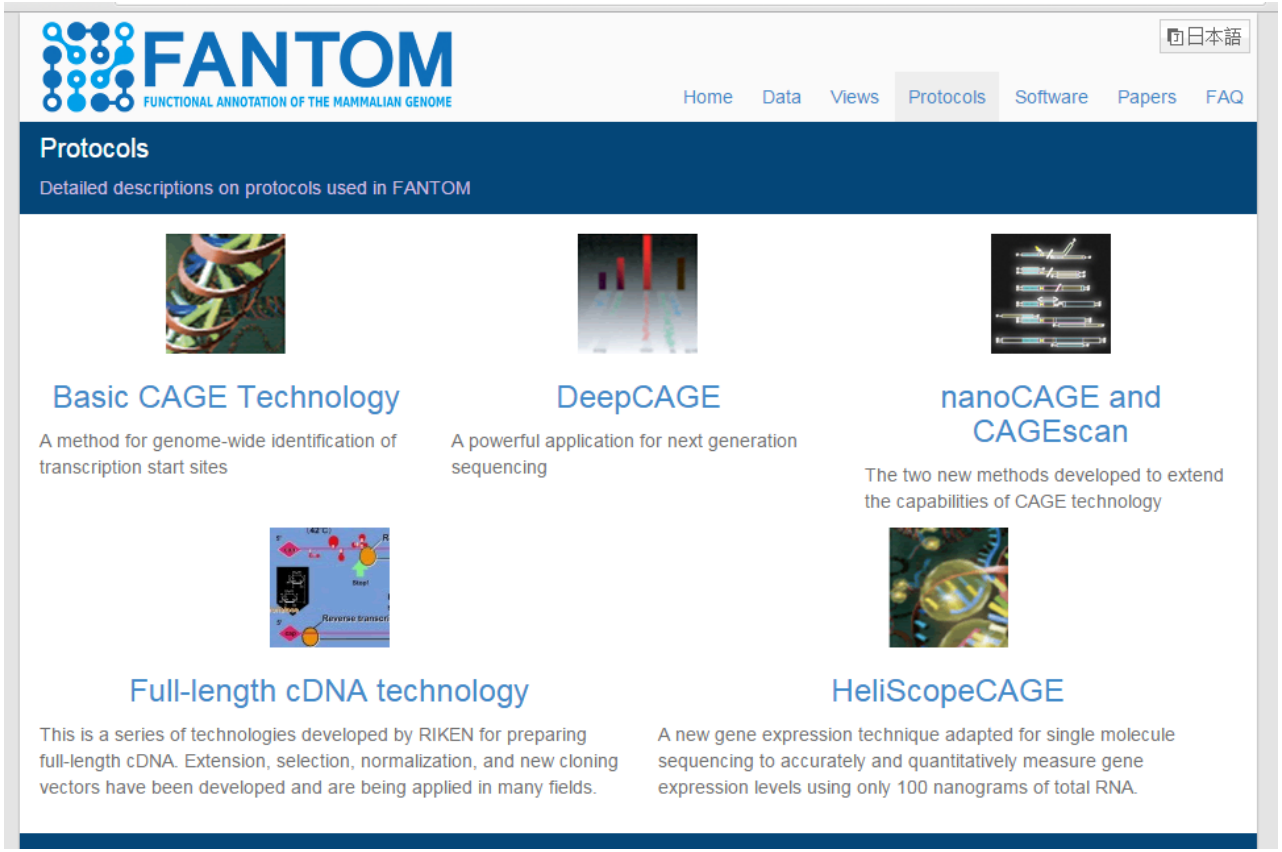
http://fantom.gsc.riken.jp/5/



Цель проекта – идентифицировать набор генов, экспрессирующихся в каждой клетке организма человека (более 400 типов клеток) и регуляторные районы гена, отвечающие за то, где (в каких клетках) и на каком уровне будет экспрессироваться ген. В дальнейшем эта информация будет использована при построении моделей регуляции транскрипции в каждом типе клеток.

FANTOM5: экспериментальные методы

http://fantom.gsc.riken.jp/protocols/



CAGE = Cap analysis gene expression - методика, которая позволяет идентифицировать короткие фрагменты 5’-концов транскриптов. При этом появляется возможность определять старт транскрипции и интенсивность экспрессии транскрипта. Данные, полученные на основе методики CAGE, позволяют исследовать закономерности организации регуляторных районов генов, обеспечивающие их тканеспецифическую экспрессию (Результат можно видеть, например, здесь: https://www.nature.com/articles/nature13182/figures/10).

Simultaneously with producing data, FANTOM established the FANTOM database and the FANTOM full-length cDNA clone bank, which are available worldwide.

The FANTOM resources have been used in several important research projects. For instance, the full-length cDNA database was used in a computer prediction of the genomic position (transcriptional unit) of genes by the International Human Genome Sequencing Consortium. Also they have been used by a research group led by Dr. Shinya Yamanaka at Kyoto University, Japan, for establishing Induced pluripotent stem (iPS) cells. In the study, 24 transcription factors were selected from FANTOM database as candidate initiation factors. Furthermore, the Allen Institute for Brain Science in the United State has created a digital atlas that encompasses the whole brain, and has made it publicly available. The atlas graphically illustrates the expression of genes within the mouse brain using Informatix software. This project has also made use of the FANTOM database.

https://fantom.gsc.riken.jp/

Доступ к различным данным на сайте FANTOM5

http://fantom.gsc.riken.jp/5/

FANTOM5

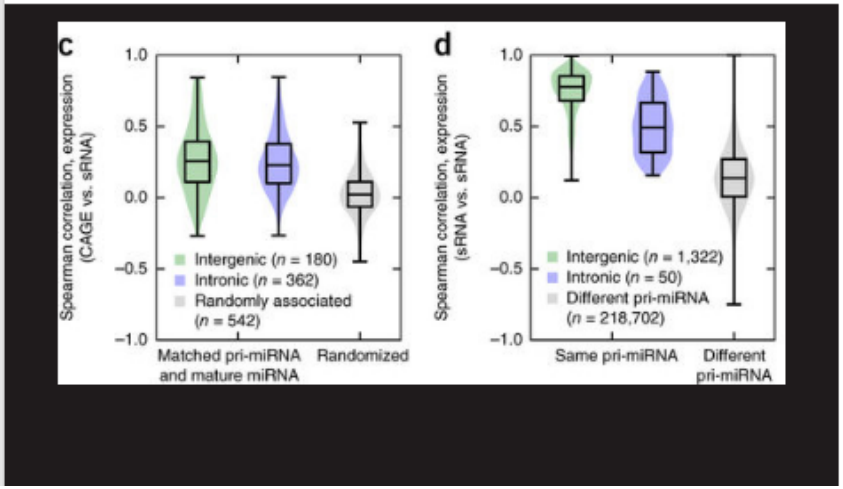
FUNCTIONAL ANNOTATION OF THE MAMMALIAN GENOME

日本語

Home Data Views Protocols Software Papers FAQ

The FANTOM5 project reports nearly 20,000 functional lncRNAs in human

The FANTOM5 project examines how our genome encodes the fantastic diversity of cell types that make up a human

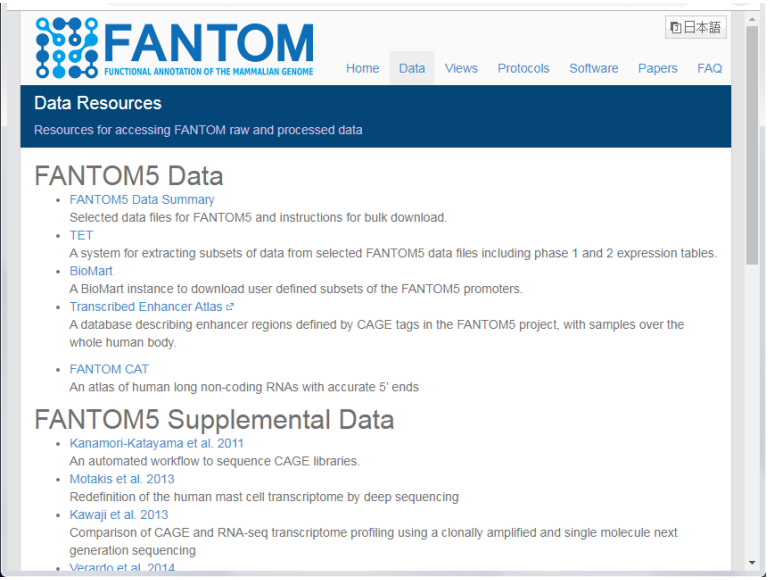


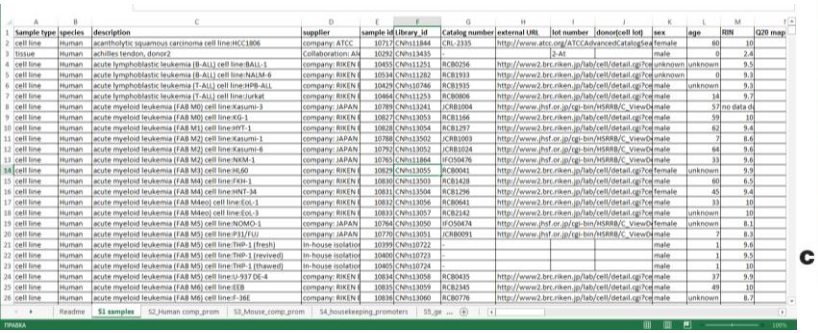
Introduction to FANTOM5

We are complex multicellular organisms composed of ~400 distinct cell types. This diversity of cell types allow us to see, think, hear, fight infections etc. yet all of this is encoded in the same genome. The difference between all these cells is what parts of the genome they use – for instance, neurons use different genes than muscle cells, and therefore they work very differently. In FANTOM5, we have systematically investigated exactly what are the sets of genes used in virtually all cell types across the human body, and the genomic regions which determine where the genes are read from. We aim to use this information to build transcriptional regulatory models for every primary cell type that makes up a human.

Закладка «Данные» на сайте FANTOM5

http://fantom.gsc.riken.jp/data/





ARR Forrest *et al. Nature* **507**, 462-470 (2014)

https://fantom.gsc.riken.jp/6/

Проект направлен на исследование функций длинных некодирующих РНК (lncRNAs) в геноме человека

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

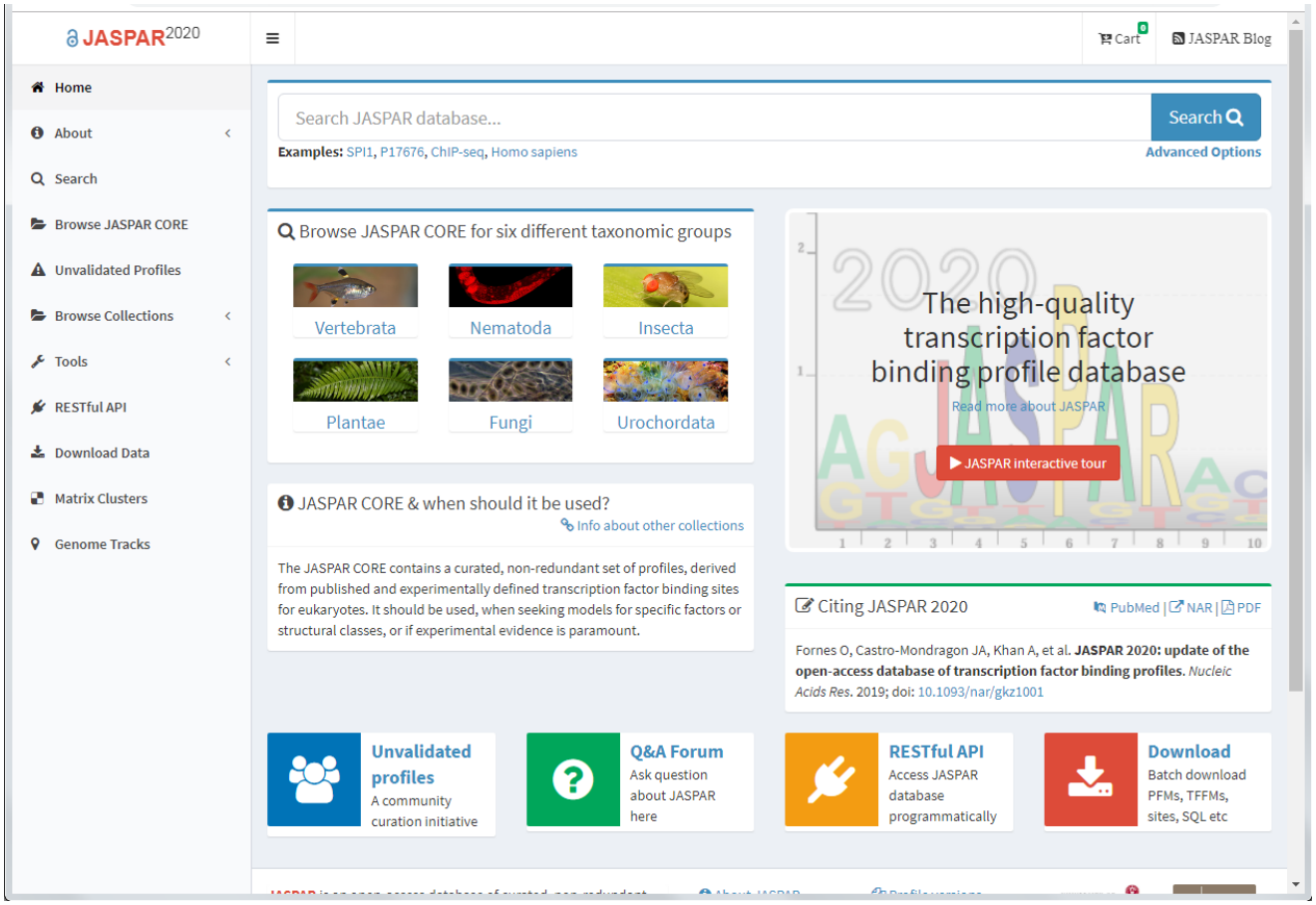
```
AC   M00134
XX
ID   V$HNF4_01
XX
DT   22.05.1995 (created); hiwi.
DT   18.10.1995 (updated); ewi.
CO   Copyright (C), Biobase GmbH.
XX
NA   HNF-4
XX
DE   hepatic nuclear factor 4
XX
BF   T00372 HNF-4alpha1; Species: rat, Rattus norvegicus.
BF   T00373 HNF-4alpha2; Species: human, Homo sapiens.
XX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

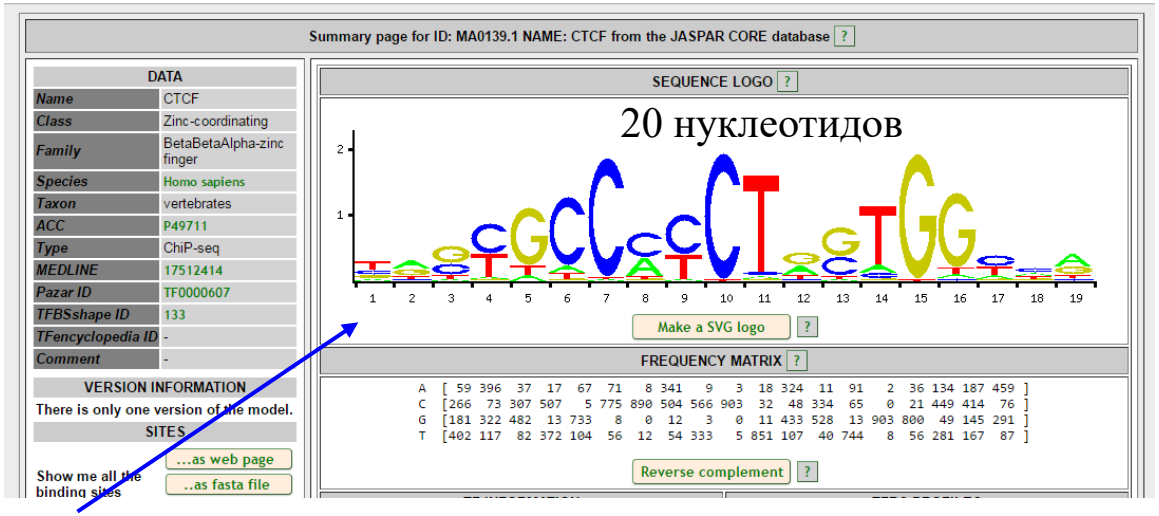


Fornes O, Castro-Mondragon JA, Khan A, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2019

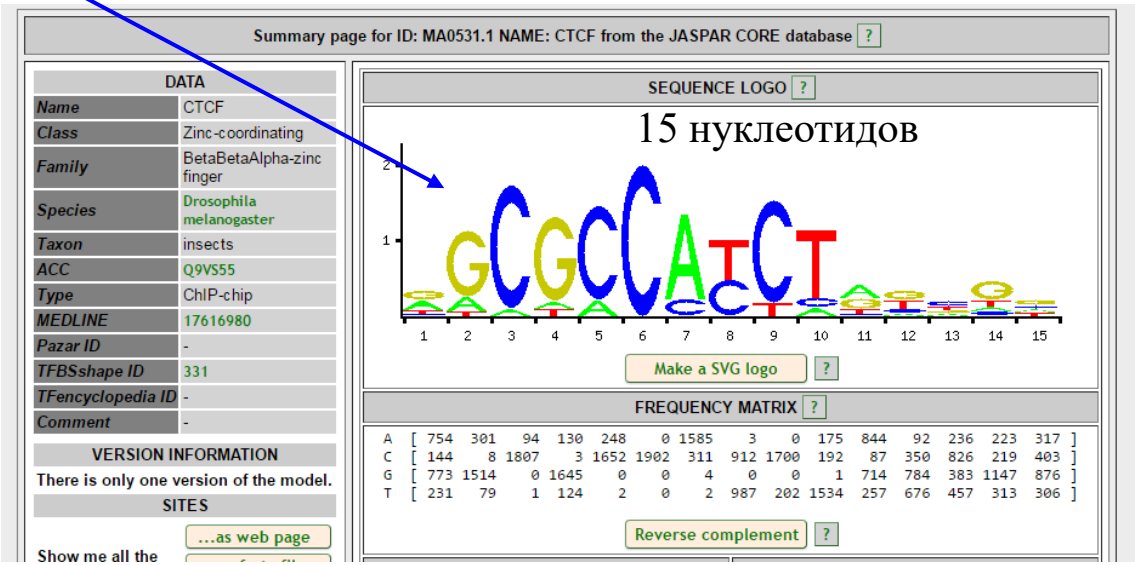
JASPAR

http://jaspar.genereg.net/

Пример: две матрицы сайтов связывания CTCF

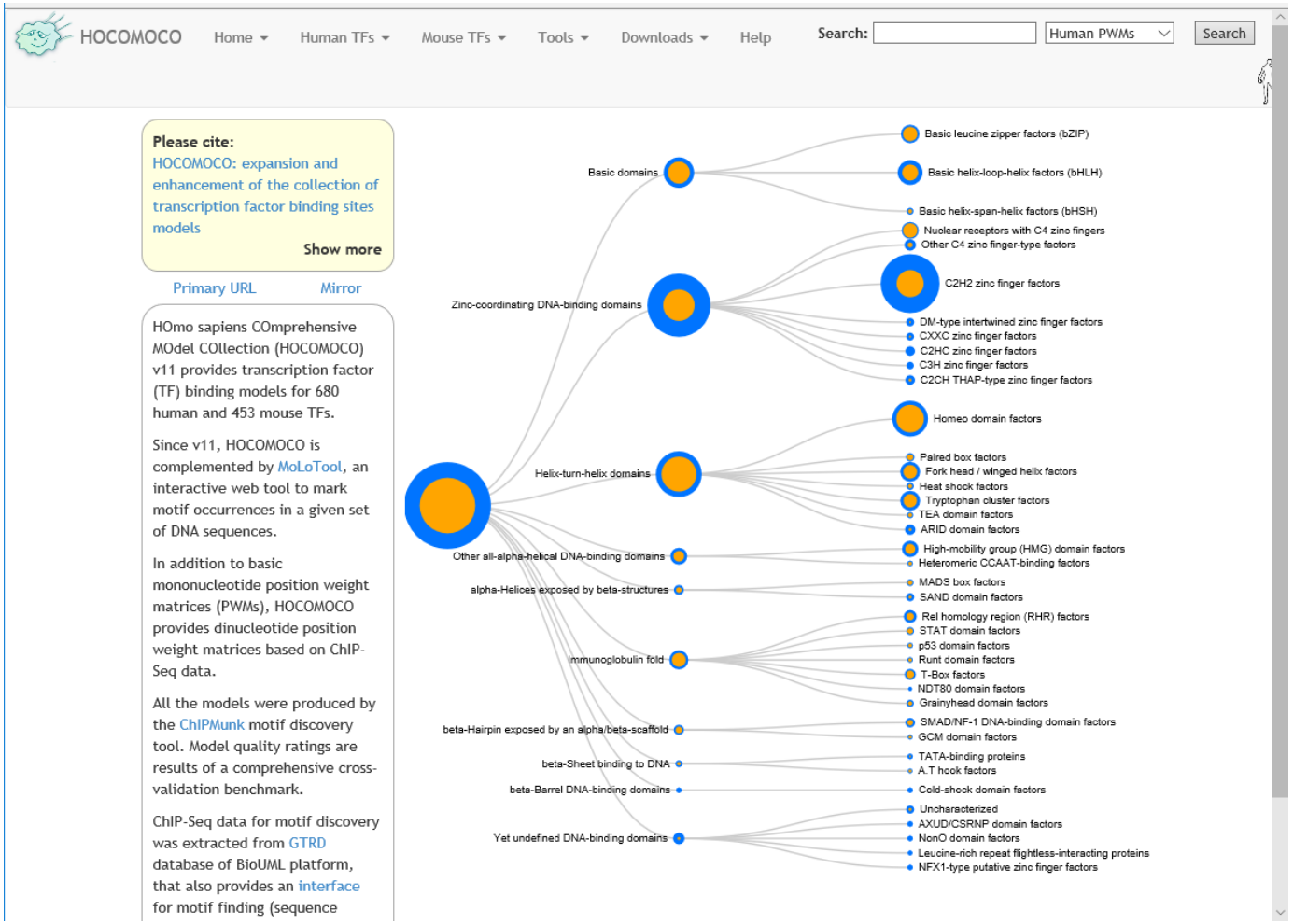


Графики WEB-logo



HOCOMOCO (HOmo sapiens COmprehensive MOdel Collection)

http://hocomoco11.autosome.ru/

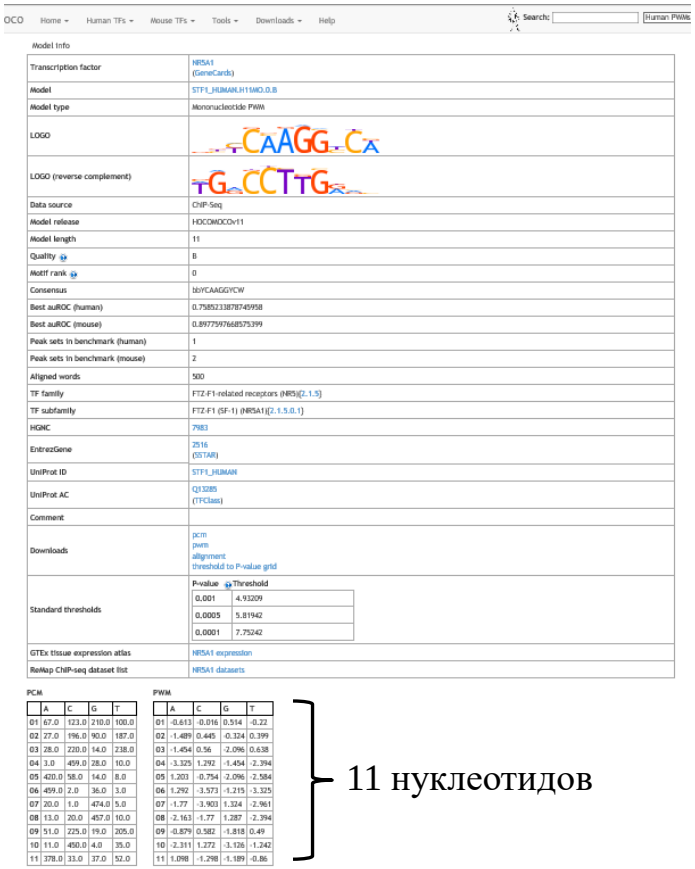


HOCOMOCO v11 provides transcription factor (TF) binding models for **680 human** and **453 mouse** TFs.

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 (STAR)
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

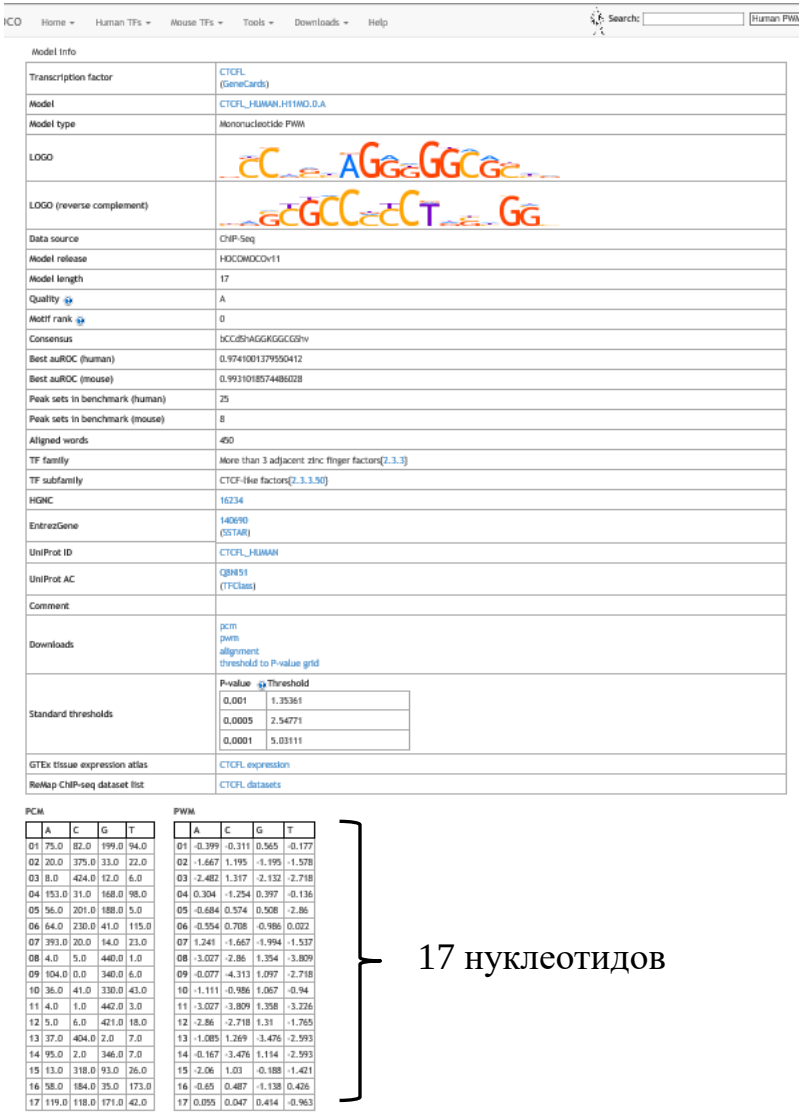
	A	C	G	T
01	67.0	123.0	210.0	100.0
02	27.0	196.0	90.0	187.0
03	28.0	220.0	14.0	238.0
04	3.0	459.0	28.0	10.0
05	420.0	58.0	14.0	8.0
06	459.0	2.0	36.0	3.0
07	20.0	1.0	474.0	5.0
08	13.0	20.0	457.0	10.0
09	51.0	225.0	19.0	205.0
10	11.0	450.0	4.0	35.0
11	378.0	33.0	37.0	52.0

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	bCCdShAGGKGGCGShv
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

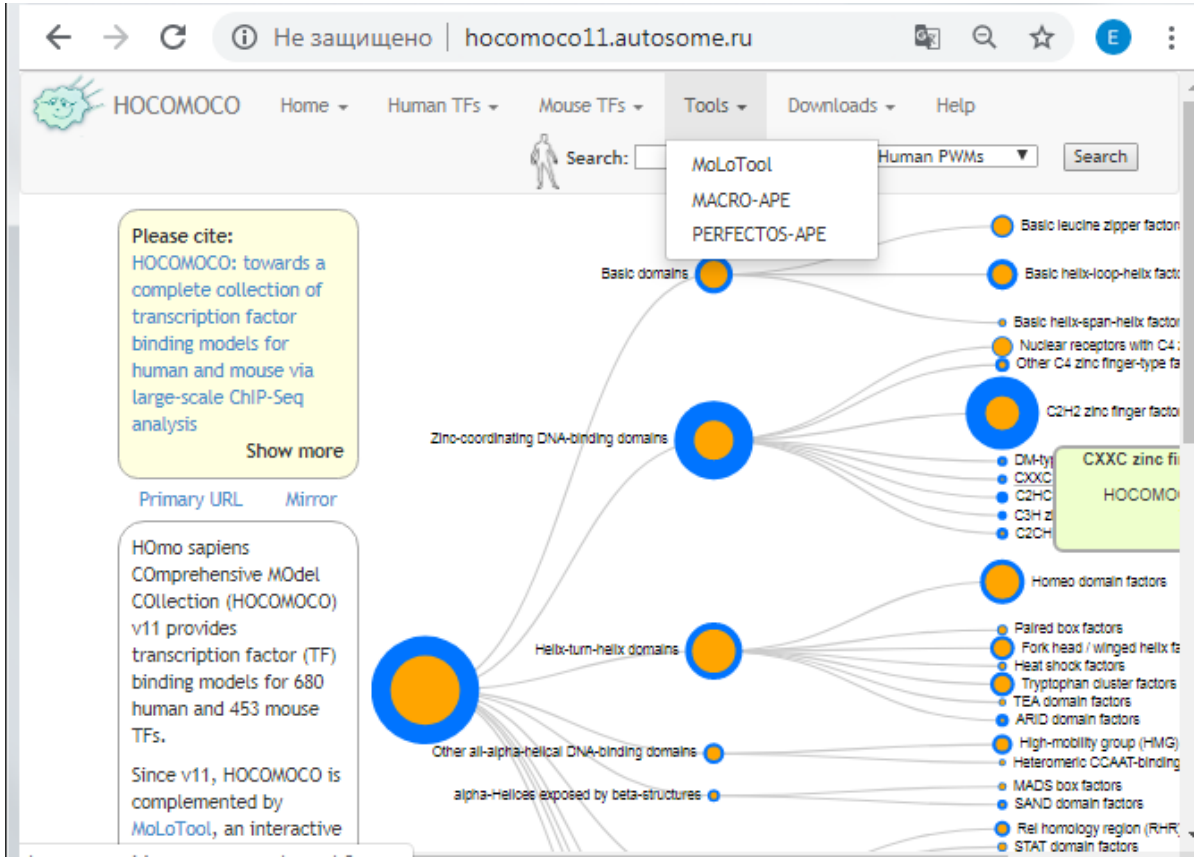
	A	C	G	T
01	-0.399	-0.311	0.565	-0.177
02	-1.667	1.195	-1.195	-1.578
03	-2.482	1.317	-2.132	-2.718
04	0.304	-1.254	0.397	-0.136
05	-0.684	0.574	0.508	-2.86
06	-0.554	0.708	-0.986	0.022
07	1.241	-1.667	-1.994	-1.537
08	-3.027	-2.86	1.354	-3.809
09	-0.077	-4.313	1.097	-2.718
10	-1.111	-0.986	1.067	-0.94
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.

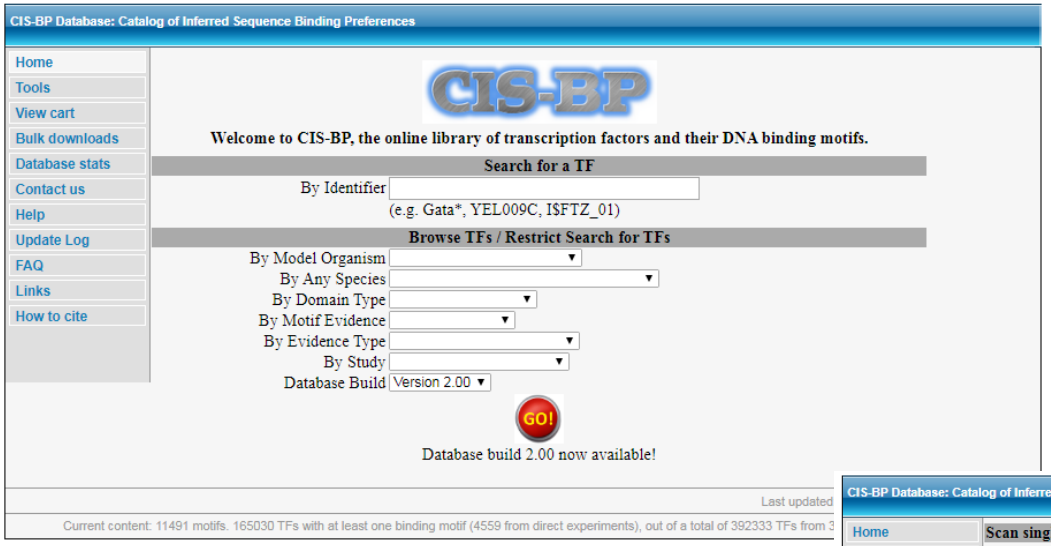
Компьютерные программы, использующие данные из HOCOMOCO

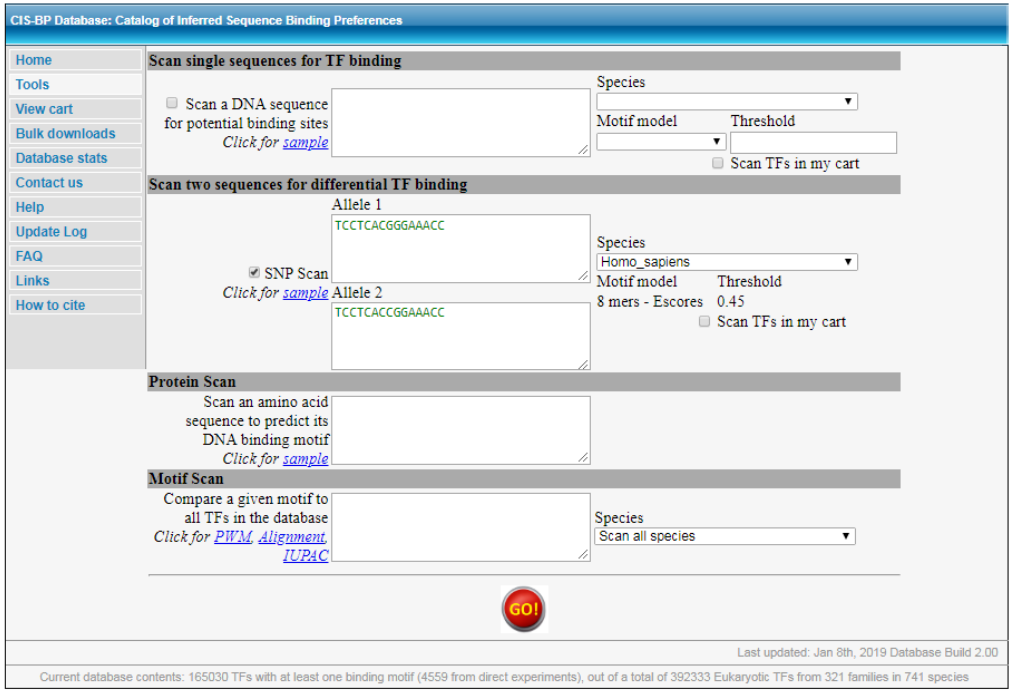
http://hocomoco11.autosome.ru/



- **MoLoTool: Transcription Factor Motif Location Toolbox**
- **MACRO-APE: MAtrix CompaRisOn by Approximate P-value Estimation**
- **PERFECTOS-APE: PrEdicting Regulatory Functional Effect by Approximate P-value Estimation**

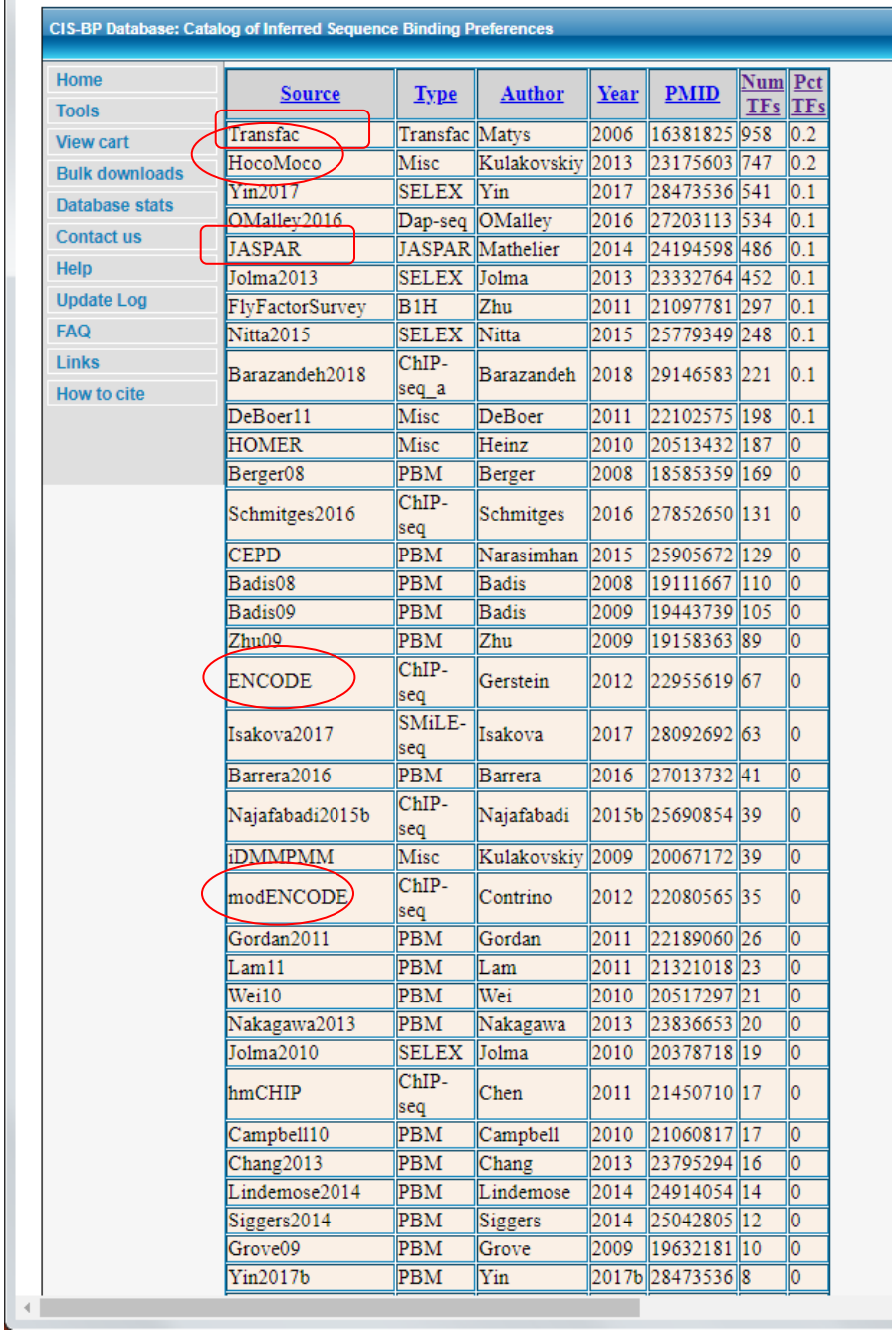
База CIS-BP (http://cisbp.ccbr.utoronto.ca/index.php) - компиляция данных по матрицам из других баз





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



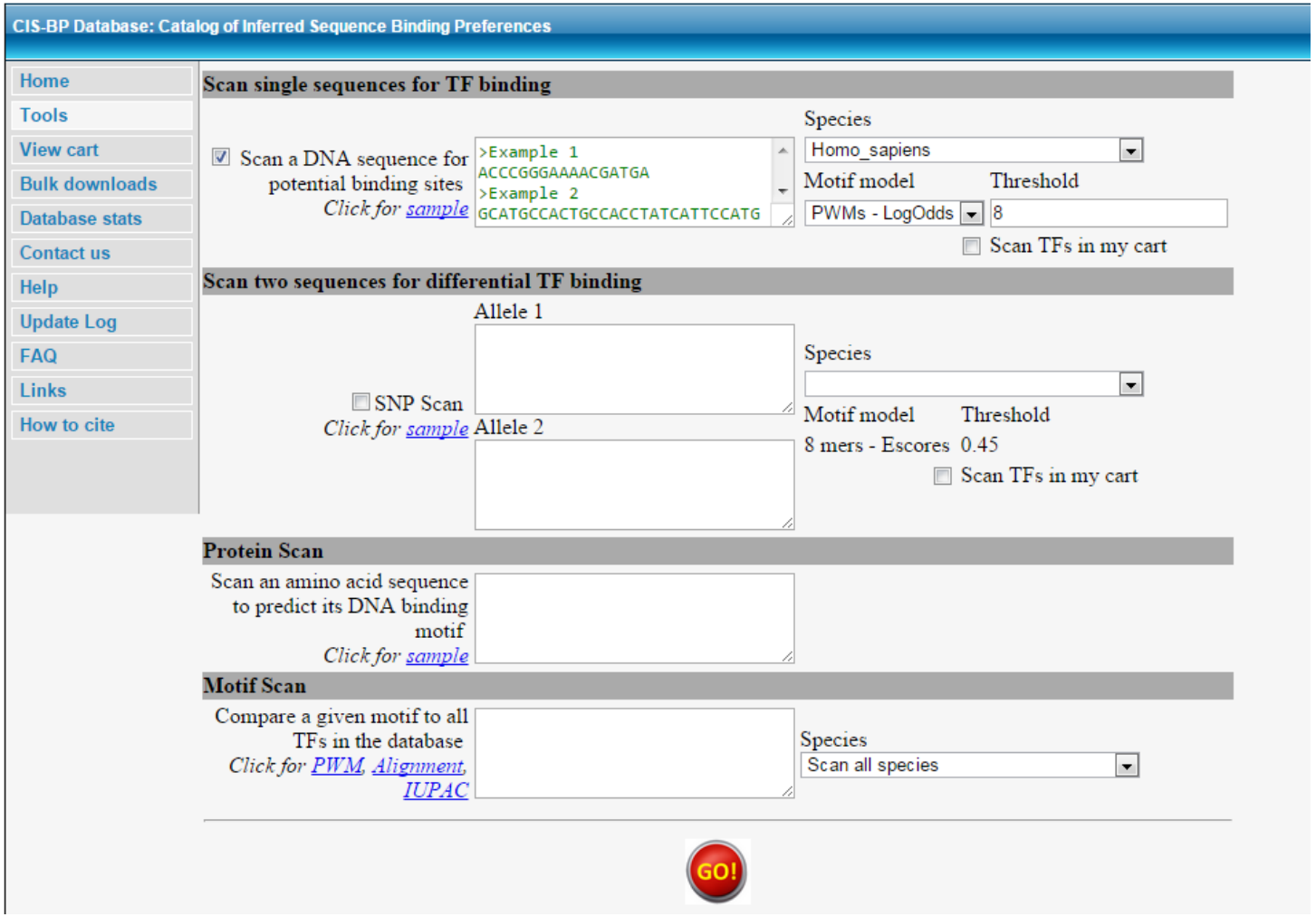
CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

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
Transfac	Transfac	Matys	2006	16381825	958	0.2
HocoMoco	Misc	Kulakovskiy	2013	23175603	747	0.2
Yin2017	SELEX	Yin	2017	28473536	541	0.1
OMalley2016	Dap-seq	OMalley	2016	27203113	534	0.1
JASPAR	JASPAR	Mathelier	2014	24194598	486	0.1
Jolma2013	SELEX	Jolma	2013	23332764	452	0.1
FlyFactorSurvey	B1H	Zhu	2011	21097781	297	0.1
Nitta2015	SELEX	Nitta	2015	25779349	248	0.1
Barazandeh2018	ChIP-seq_a	Barazandeh	2018	29146583	221	0.1
DeBoer11	Misc	DeBoer	2011	22102575	198	0.1
HOMER	Misc	Heinz	2010	20513432	187	0
Berger08	PBM	Berger	2008	18585359	169	0
Schmitges2016	ChIP-seq	Schmitges	2016	27852650	131	0
CEPD	PBM	Narasimhan	2015	25905672	129	0
Badis08	PBM	Badis	2008	19111667	110	0
Badis09	PBM	Badis	2009	19443739	105	0
Zhu09	PBM	Zhu	2009	19158363	89	0
ENCODE	ChIP-seq	Gerstein	2012	22955619	67	0
Isakova2017	SMiLE-seq	Isakova	2017	28092692	63	0
Barrera2016	PBM	Barrera	2016	27013732	41	0
Najafabadi2015b	ChIP-seq	Najafabadi	2015b	25690854	39	0
iDMMPMM	Misc	Kulakovskiy	2009	20067172	39	0
modENCODE	ChIP-seq	Contrino	2012	22080565	35	0
Gordan2011	PBM	Gordan	2011	22189060	26	0
Lam11	PBM	Lam	2011	21321018	23	0
Wei10	PBM	Wei	2010	20517297	21	0
Nakagawa2013	PBM	Nakagawa	2013	23836653	20	0
Jolma2010	SELEX	Jolma	2010	20378718	19	0
hmCHIP	ChIP-seq	Chen	2011	21450710	17	0
Campbell10	PBM	Campbell	2010	21060817	17	0
Chang2013	PBM	Chang	2013	23795294	16	0
Lindemose2014	PBM	Lindemose	2014	24914054	14	0
Siggers2014	PBM	Siggers	2014	25042805	12	0
Grove09	PBM	Grove	2009	19632181	10	0
Yin2017b	PBM	Yin	2017b	28473536	8	0

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.

База CIS-BP (http://cisbp.ccbr.utoronto.ca/index.php) - 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.

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

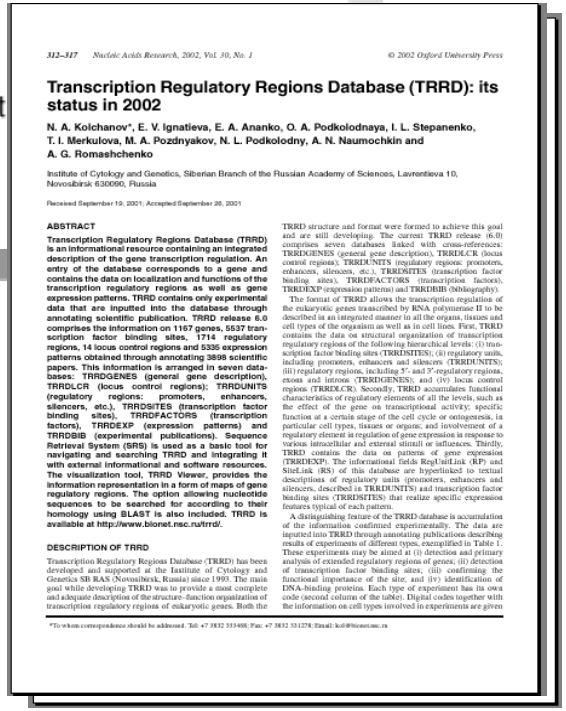
312–317 *Nucleic Acids Research, 2002, Vol. 30, No. 1*

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, Lavrent
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 5336 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:

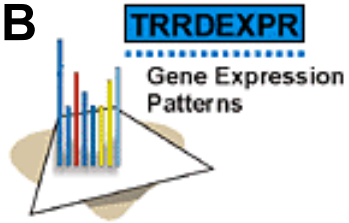
1) СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ ОРГАНИЗАЦИИ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ РАЙОНОВ. СТРУКТУРНЫЕ И ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ:

САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

РЕГУЛЯТОРНЫХ РАЙОНОВ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)

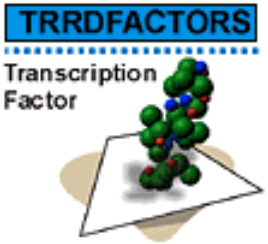
ЛОКУС-КОНТРОЛИРУЮЩИХ РАЙОНОВ

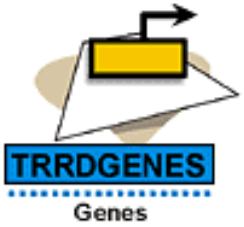
2) ДАННЫЕ О ПАТТЕРНАХ ЭКСПРЕССИИ ГЕНОВ



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

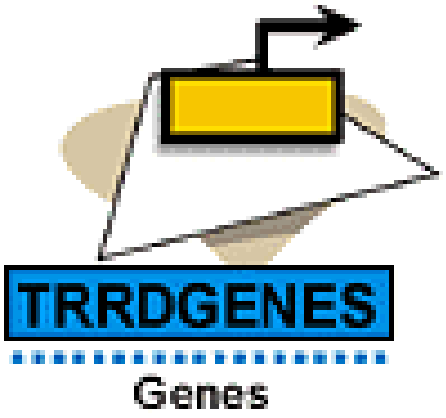
3) ДАННЫЕ О ТРАНСКРИПЦИОННЫХ ФАКТОРАХ





TRRDGENES

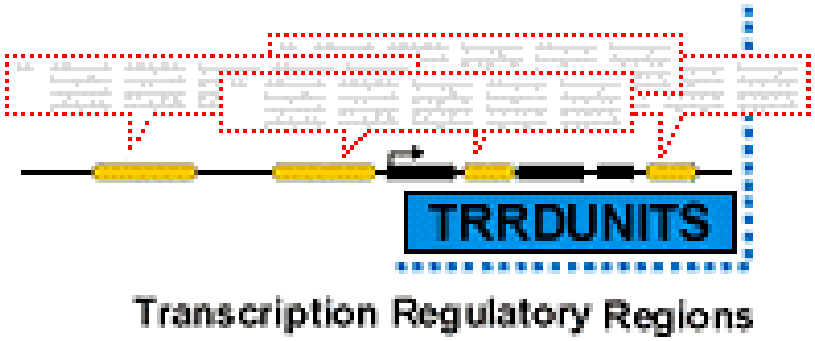
ОБЩЕЕ ОПИСАНИЕ ГЕНА



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
//

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)

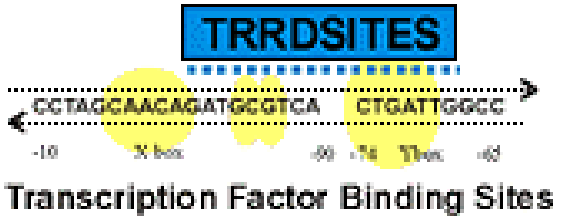
TRRDUNITS4:P00051

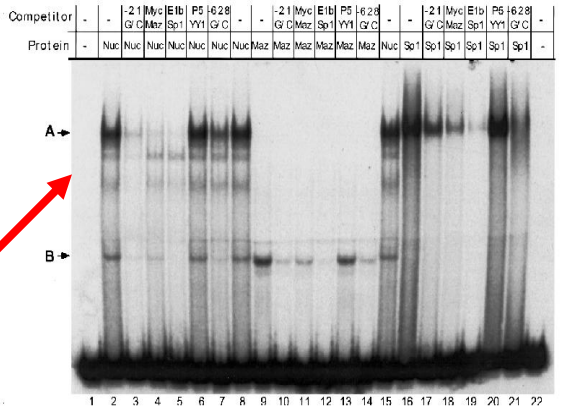
RegUnitAC
P00051
GeneID
Hs:APOA1
RegRegion
5'region
RegUnit
enhancer; ST; -256 to -110; S1425, S1426, S997, S666,
S1001, S1002, S1427, S1003, S1136, S1004, S1005, S1006, S
S5744, S5745, S5746, S5747, S5748, S5749, S5750
Site:(S1425) -225 to -210; Egr-1;
Site:(S1426) -227 to -212; Sp1 bs;
Site:(S997) -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

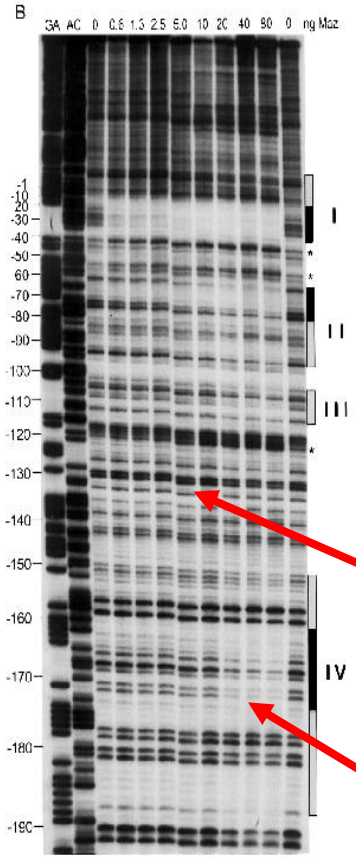
ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА



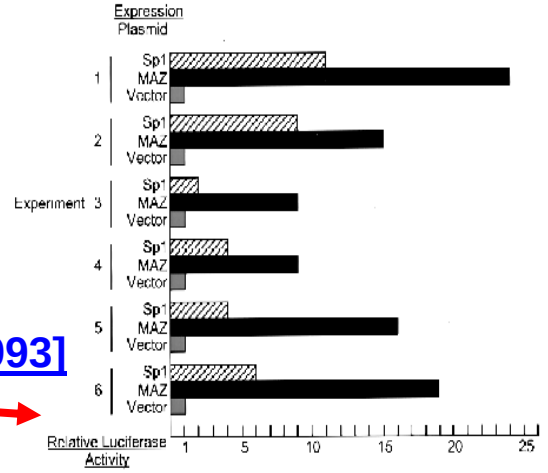


GEL-MOBILITY SHIFT ASSAY

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

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

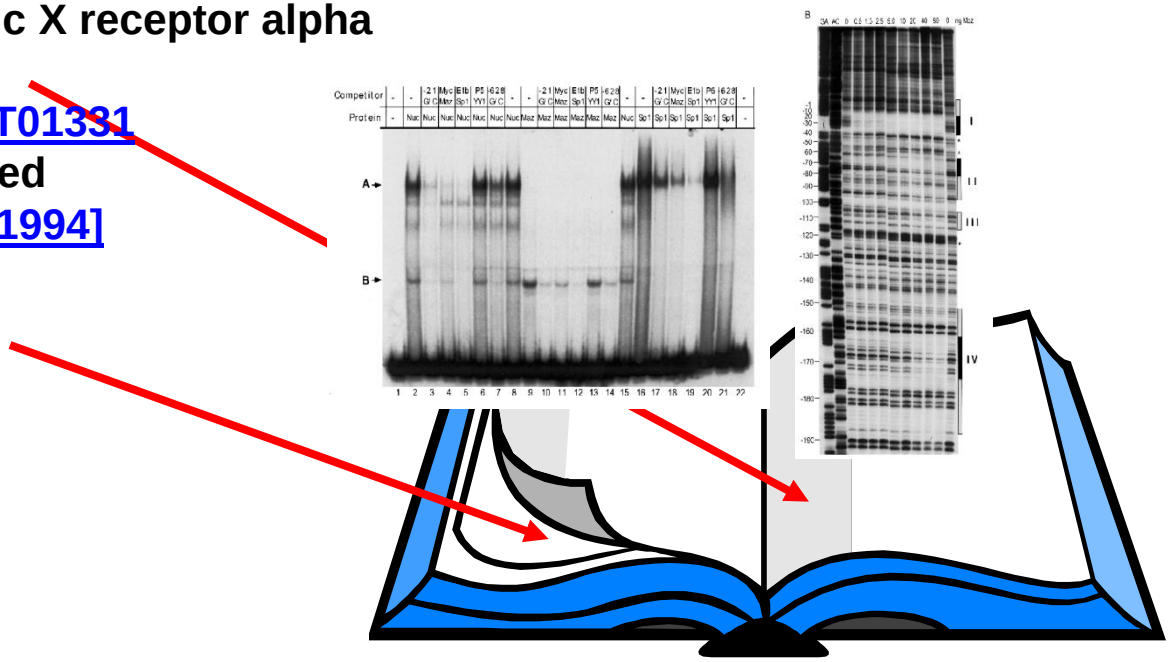
Type of experiment	Assay code in TRRD
Detection of transcription factor binding sites	
DNase I footprinting with nuclear extract	1.1.1
DNase I footprinting with purified or recombinant protein	1.1.5
Genomic footprinting	1.5
Methylation protection assay	4.1
Methylation interference assay	4.2
Electrophoretic 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

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]
//

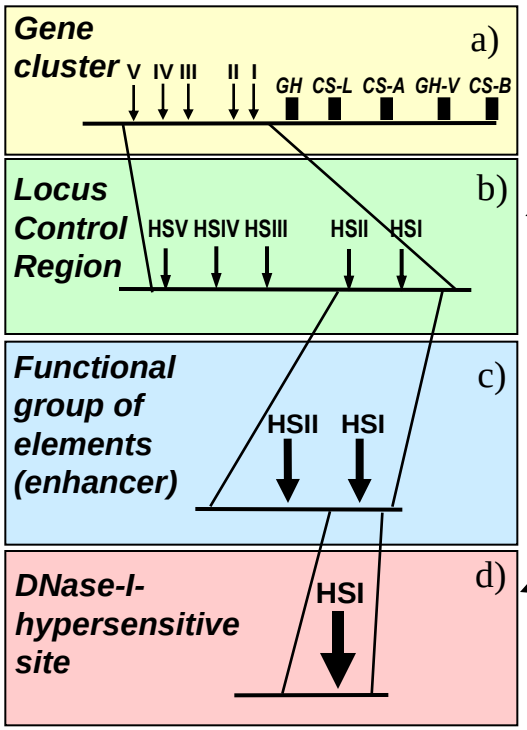


TRRDLCR

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;;CS5
CS2
XX
GI GH1; SOMATOTROPIN (GROWT

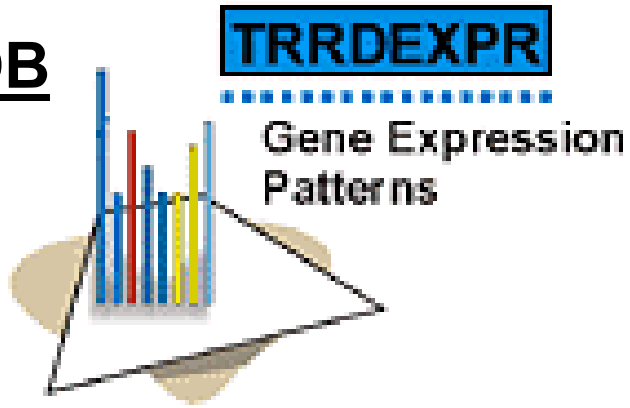
AL A7
LI Hs:LPGH
TL pituitary specific
BI EMBL; AF010280; AF010280
EL LGH:HSV, LGH:HSIII, LGH ENHANCER
.....

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

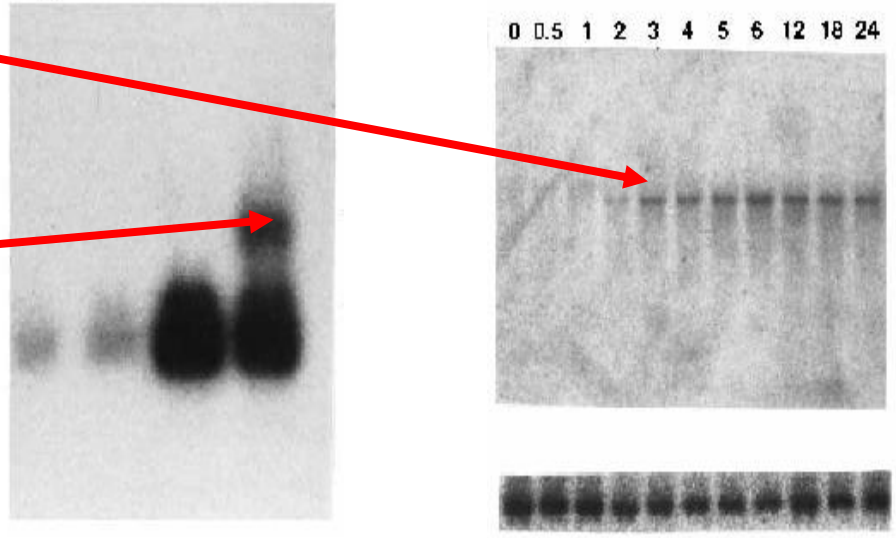
TRRDEXP

ОСОБЕННОСТИ ЭКСПРЕССИИ ГЕНОВ



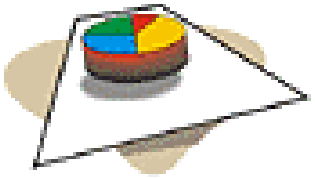
RE A00596.002
ID Hs:AAP
RT mRNA
RN lymphoblastoid cells
RL present
RI heat shock
FF induction
RH 3 h
RR [Abe K. et al., 1991]

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



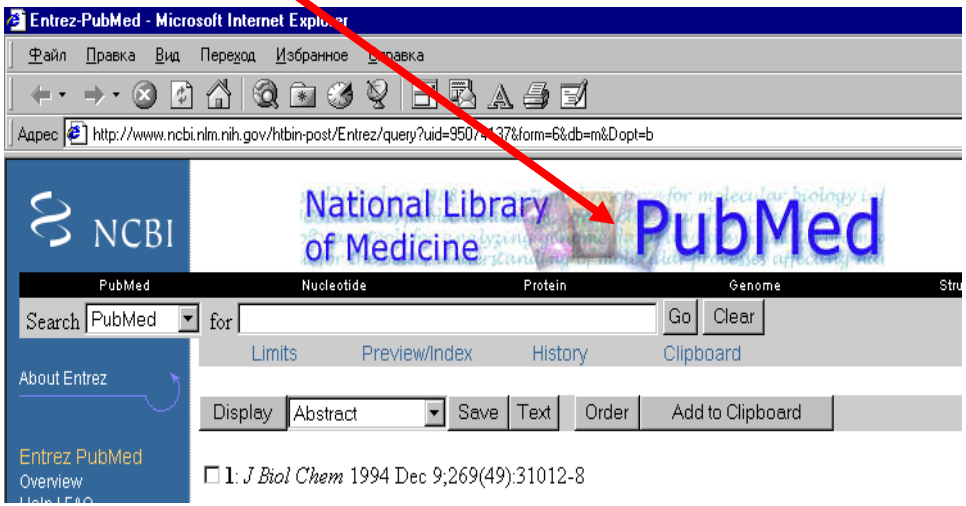
RNA blot analysis

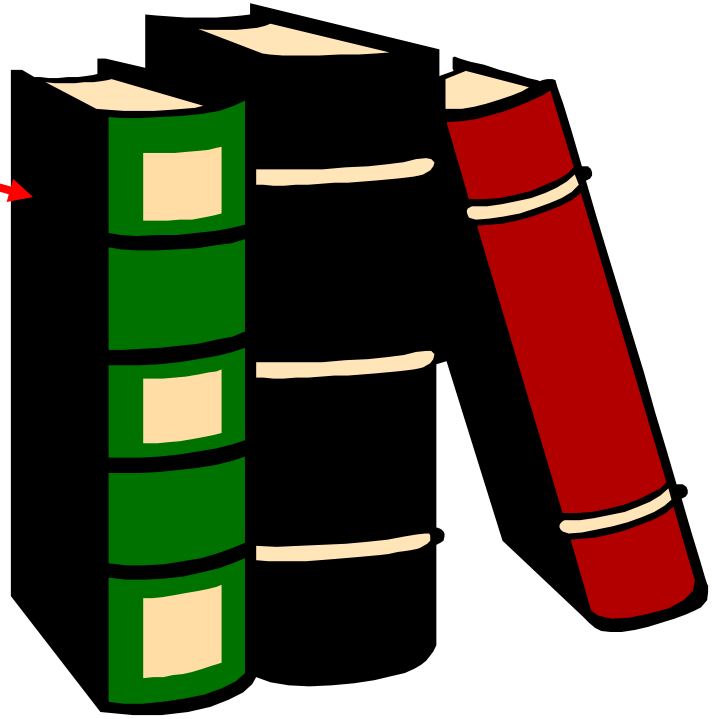
TRRDBIB



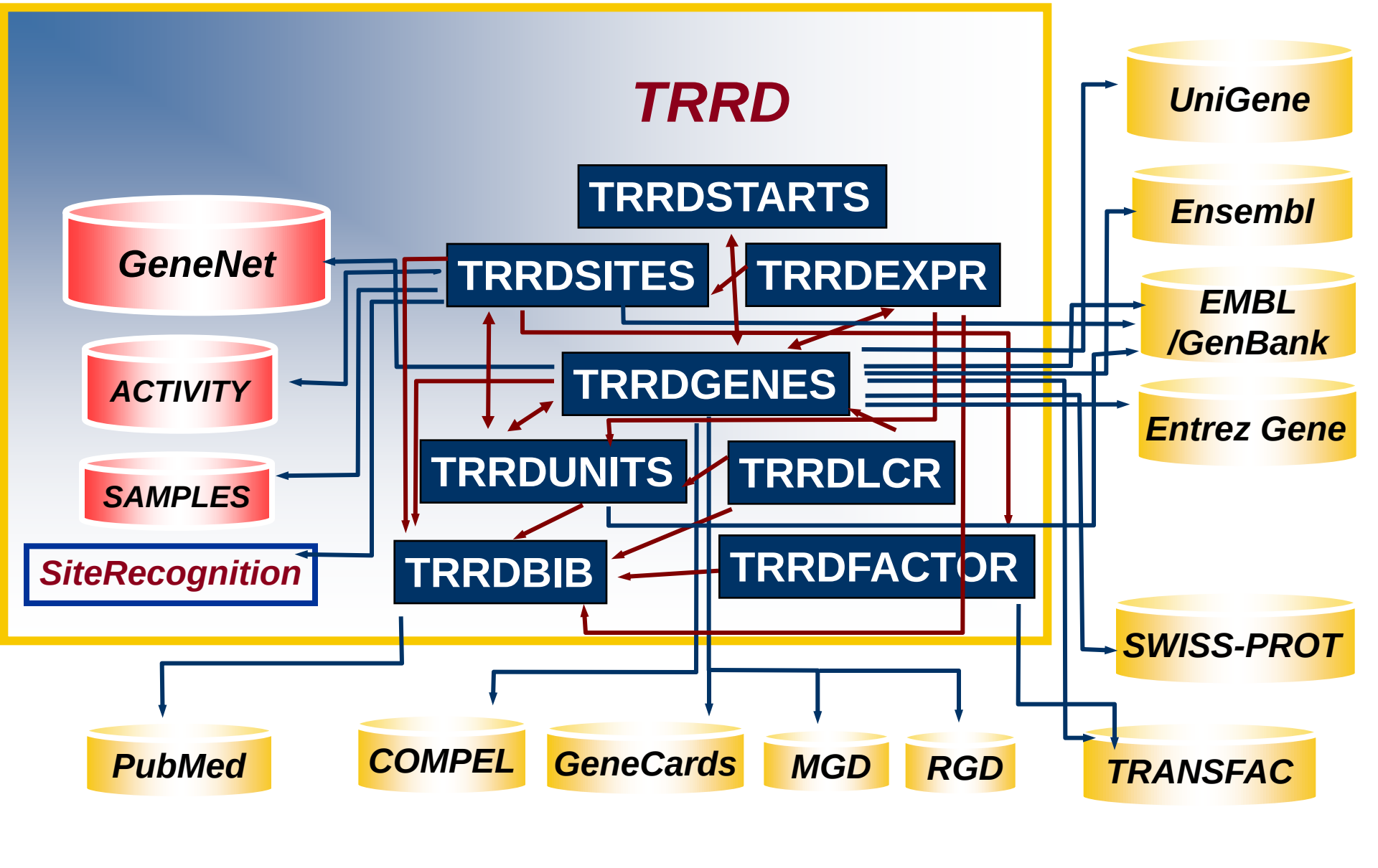
TRRDBIB
Bibliography

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]
//

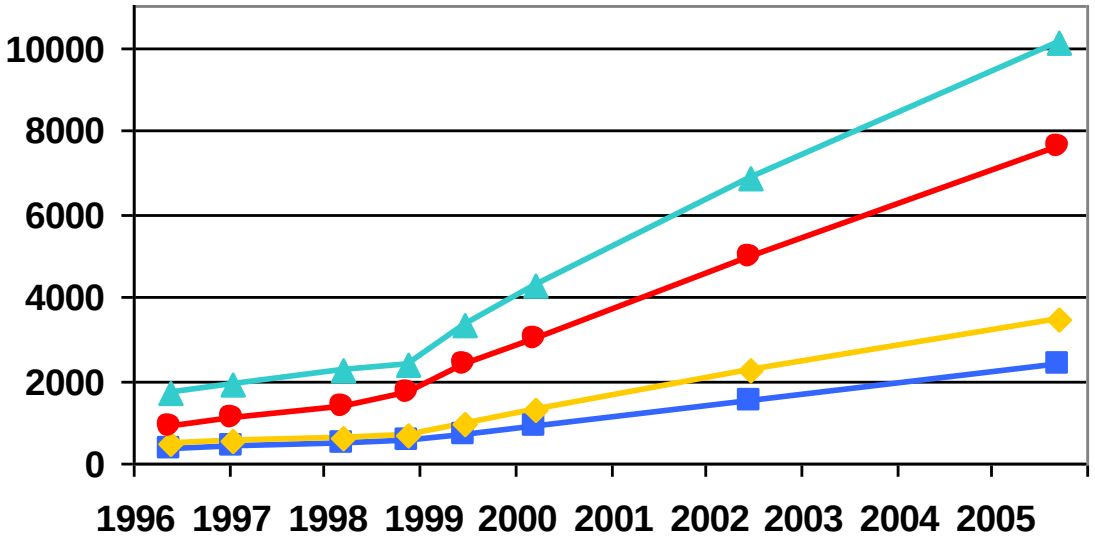




Линковка между таблицами TRRD и ссылки на внешние ресурсы.



ДИНАМИКА ПОПОЛНЕНИЯ TRRD



10135 САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ

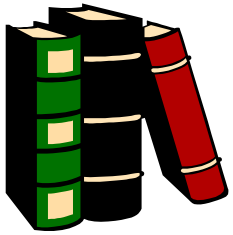
7609 НАУЧНЫХ ПУБЛИКАЦИЙ

3400 РЕГУЛЯТОРНЫХРАЙОНОВ

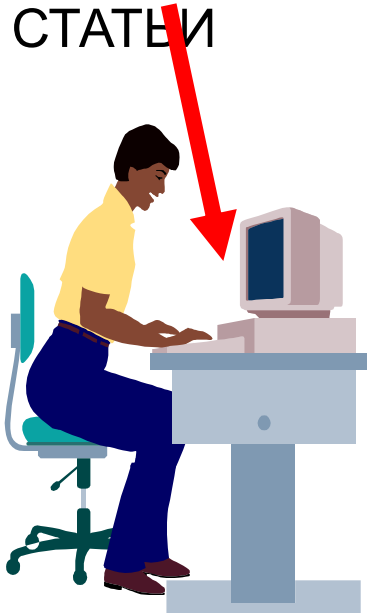
2344 ГЕНОВ

	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)

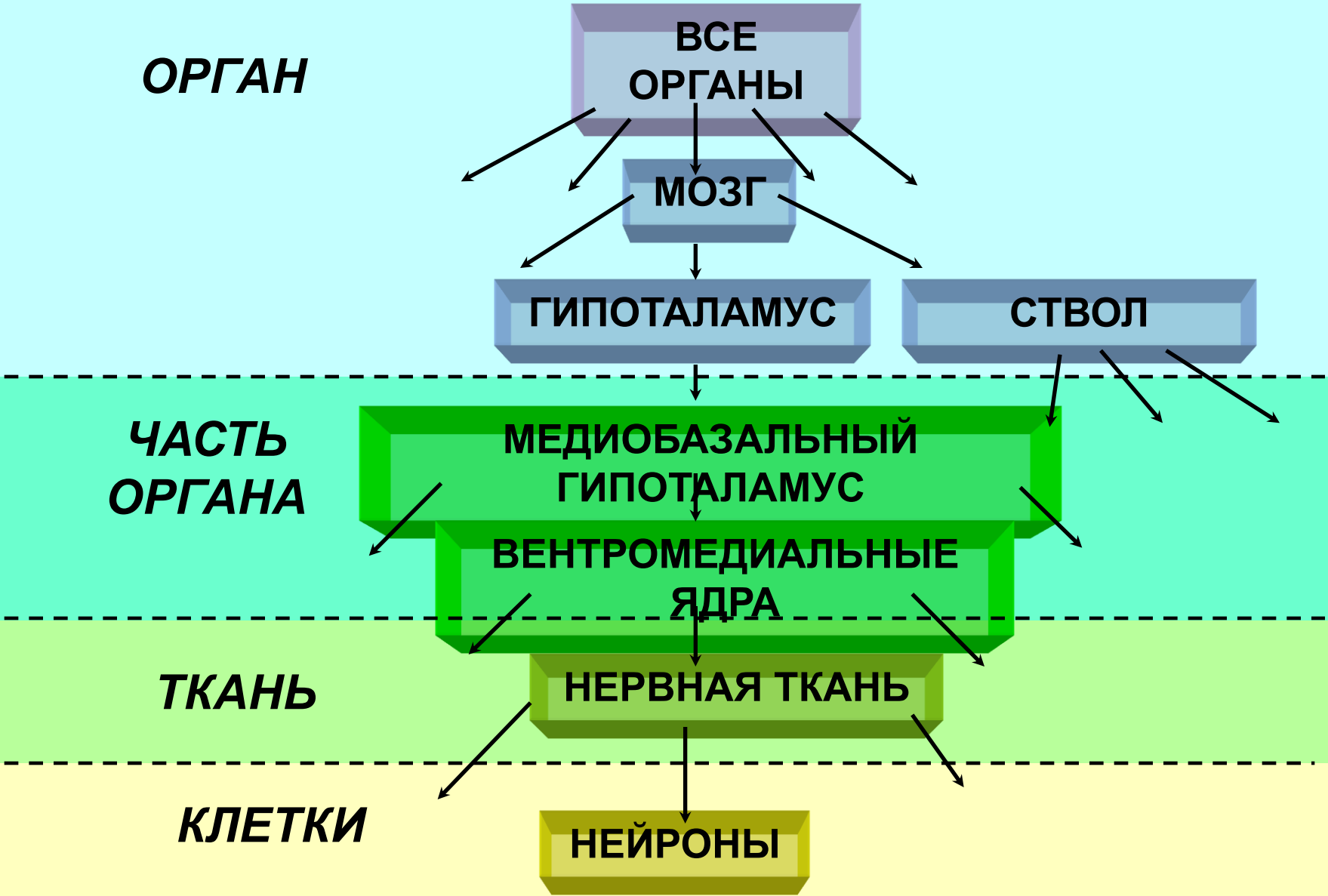
- СТАДИИ РАЗВИТИЯ
- КЛЕТКИ
- ОРГАНЫ
- ТКАНИ
- КЛЮЧЕВЫЕ СЛОВА
- ВНЕШНИЕ СТИМУЛЫ
- СТАДИИ КЛЕТОЧНОГО ЦИКЛА
- КОДЫ ЭКСПЕРИМЕНТОВ

СИНТАКСИЧЕСКАЯ и СЕМАНТИЧЕСКАЯ ПРОВЕРКА

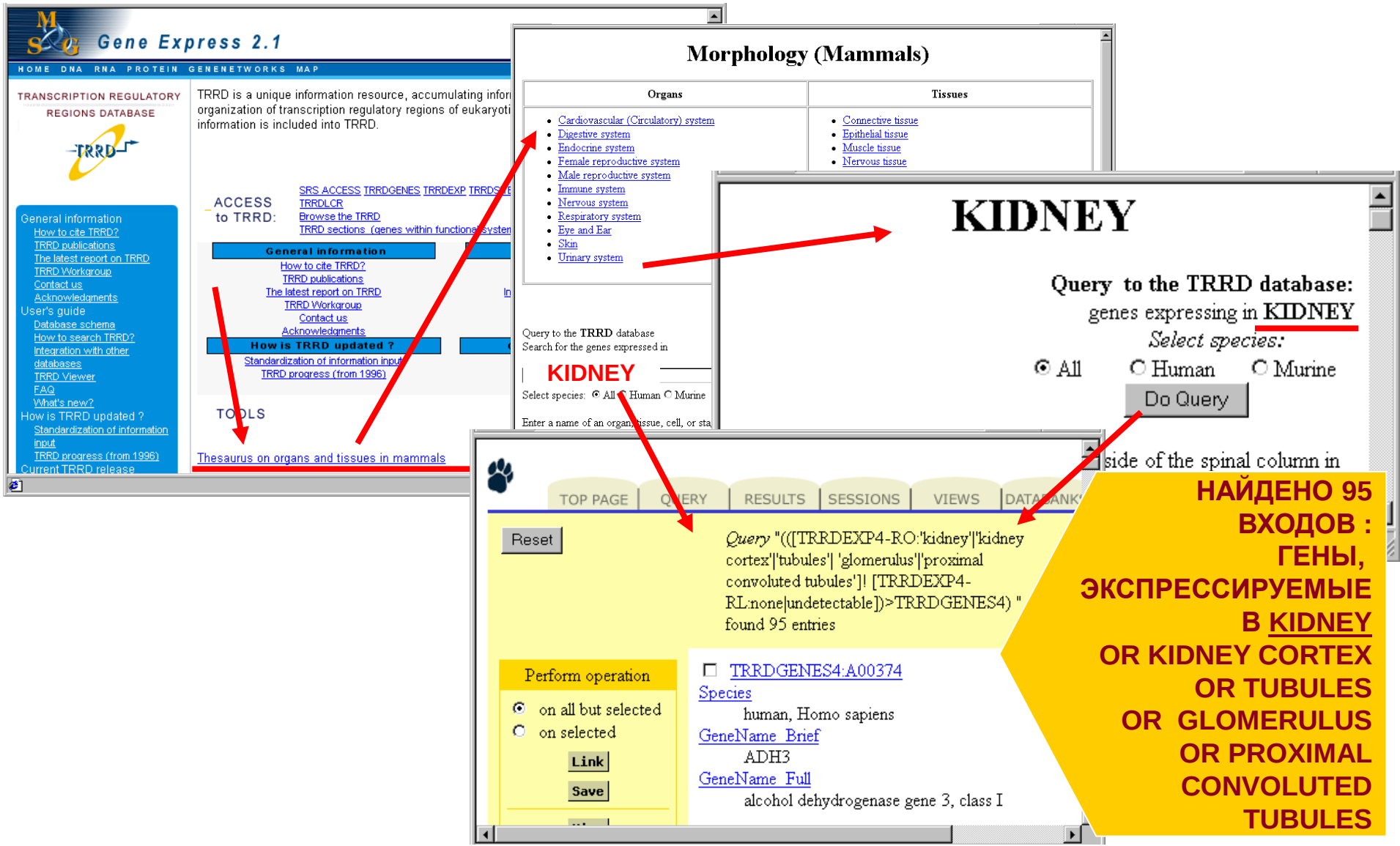
TRRD

Интернет

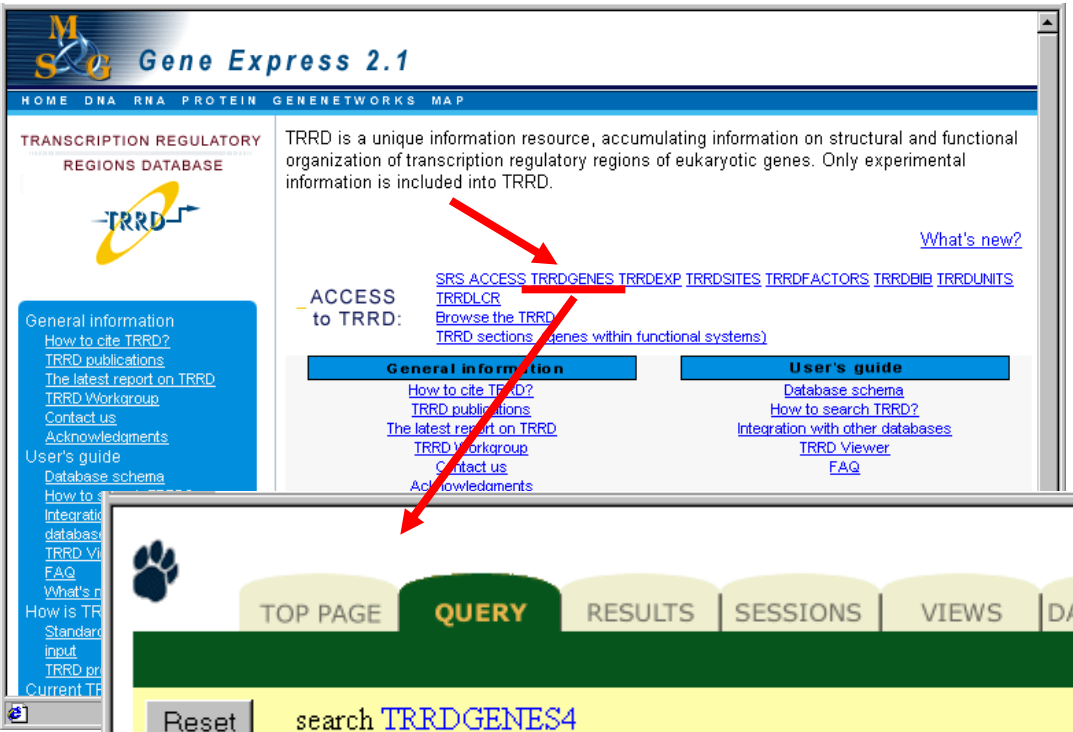
ИЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ КОНТРОЛИРУЕМЫХ СЛОВАРЕЙ МОРФОЛОГИЧЕСКИХ ТЕРМИНОВ В TRRD

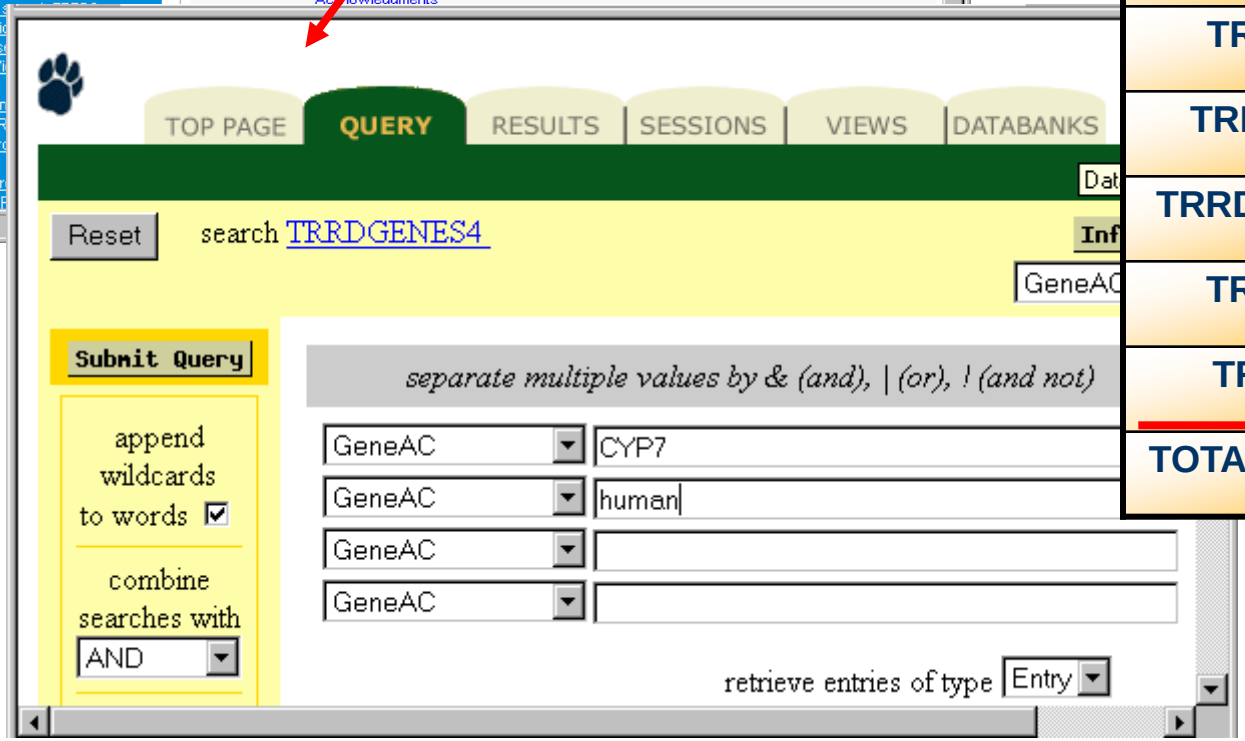


ПОИСК ДАННЫХ вTRRD: ЗАПРОС НА ОСНОВЕ ТЕЗАУРОСОВ ОРГАНОВ и ТКАНЕЙ МЛЕКОПИТАЮЩИХ



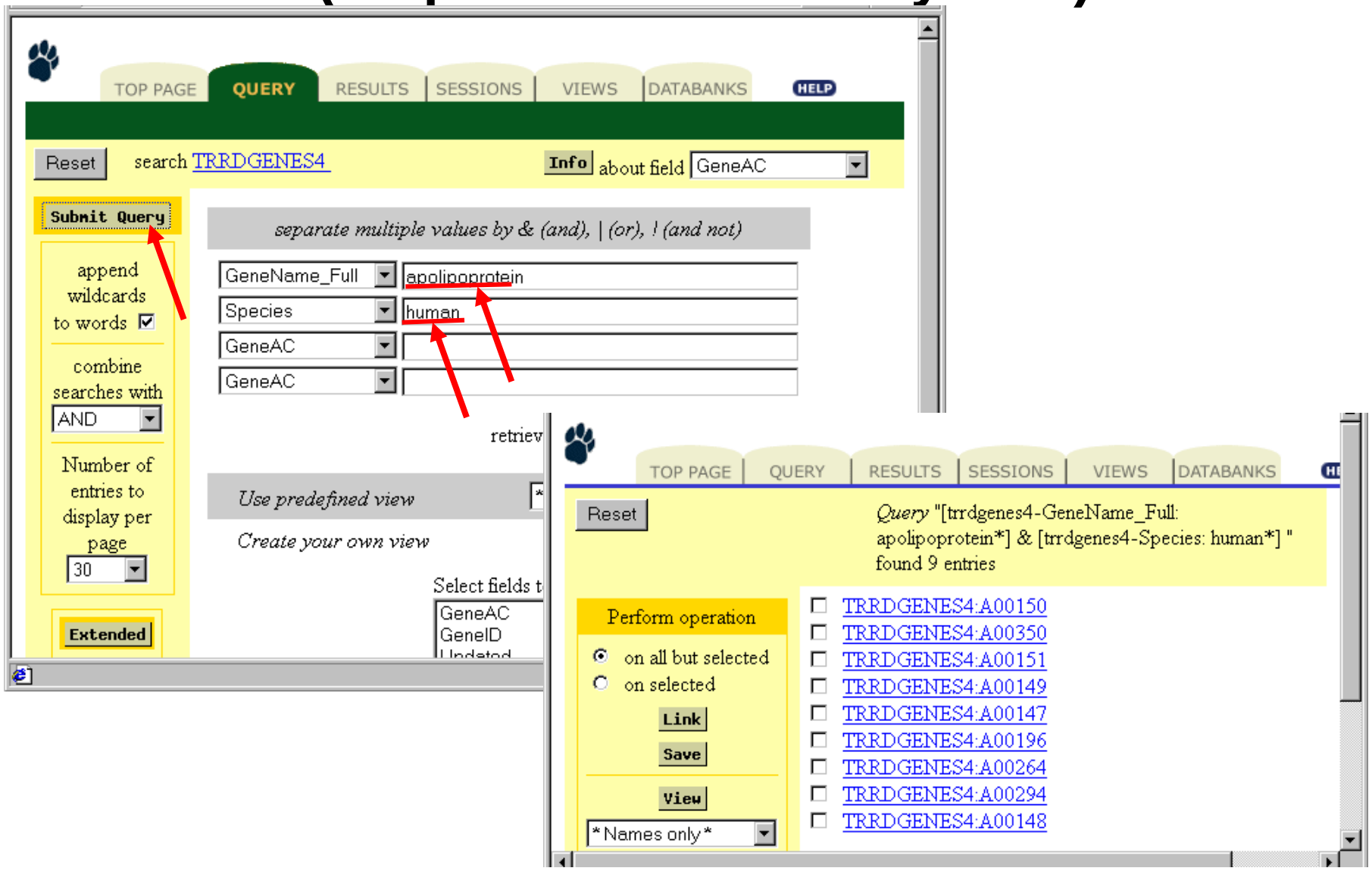
ПОИСК ДАННЫХ в TRRD: SRS



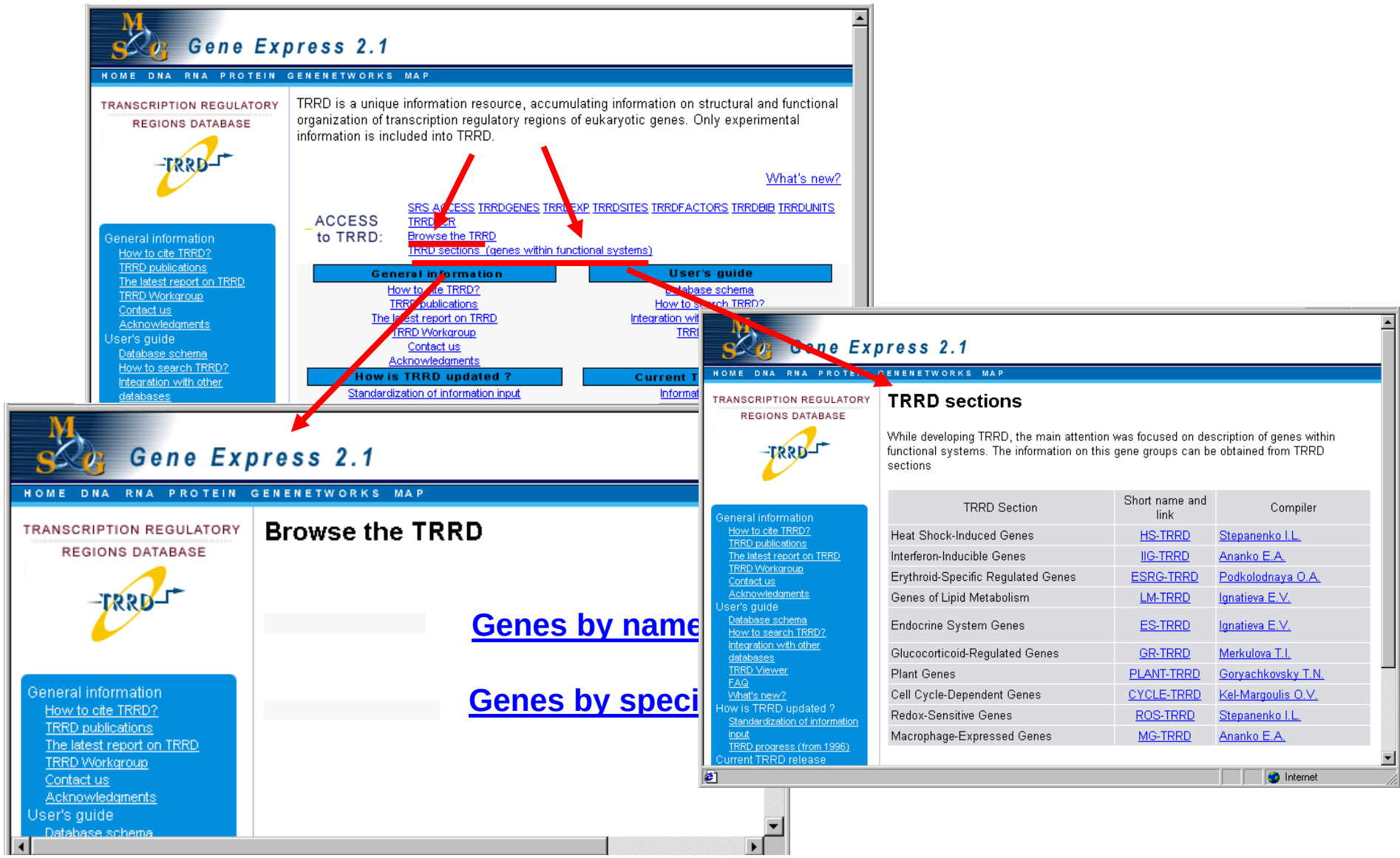


DATABASE	КОЛИЧЕСТВО ПОЛЕЙ, ПРОИНДЕКСИРОВАННЫХ для РЕЛИЗА 6.01
TRRDGENES	24
TRRDUNITS	11
TRRDEXP	17
TRRDSITES	16
TRRDFACTORS	14
TRRDLCR	40
TRRDBIB	9
TOTAL NUMBER	131

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



ПОИСК ДАННЫХ В TRRD: БРАУЗЕРЫ и ТЕМАТИЧЕСКИЕ СЕКЦИИ

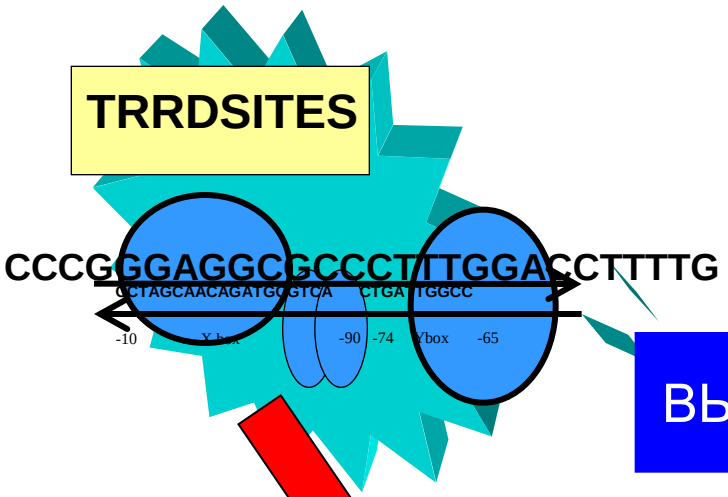


Тематические секции TRRD

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

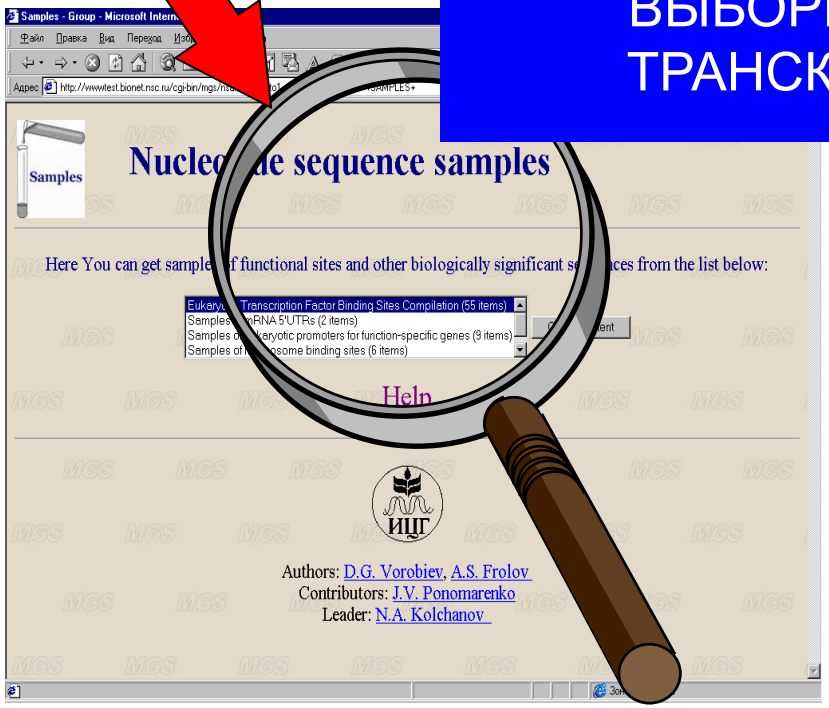
TRRD Section	Short name and link	Compiler
Heat Shock-Induced Genes	HS-TRRD	Stepanenko I.L.
Interferon-Inducible Genes	IIG-TRRD	Ananko E.A.
Genes Expressed in B cells	B-TRRD	Ananko E.A.
Genes Related to EBV Infection and EBV Transformation	EBV-TRRD	Ananko E.A.
Erythroid-Specific Regulated Genes	ESRG-TRRD	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.

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

Характеристики баз данных = «Критерии полезности баз данных»:

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

Конец лекции !!!