

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

University of California, Santa Cruz

США

ENCODE Encyclopedia of DNA Elements

США

FANTOM The Functional Annotation of Mammalian genome

Rikagaku Kenkyūsho = **RIKEN** (理研)

Япония

+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**

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



UCSC = University of California, Santa Cruz

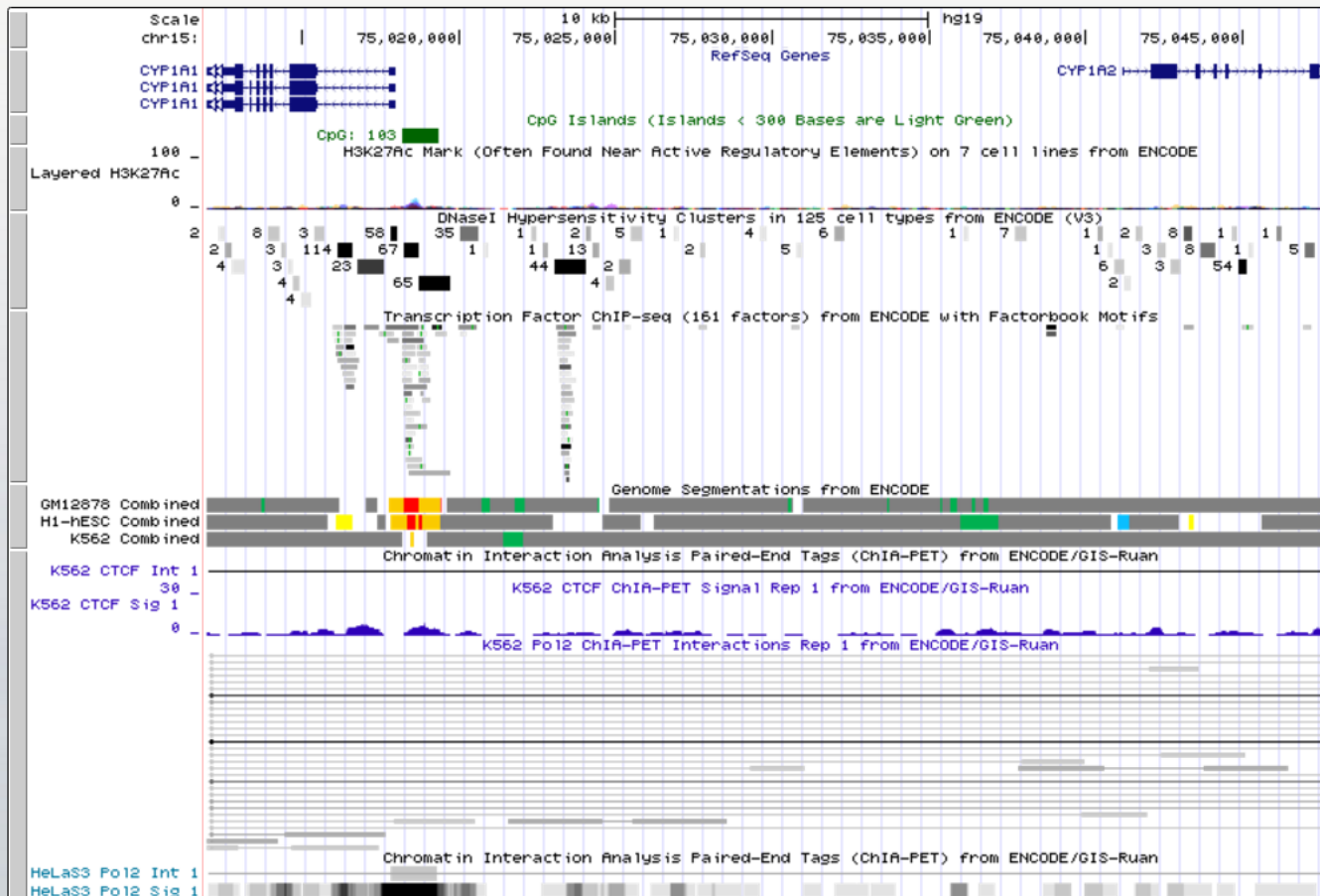
Геномный браузер Университета г.Санта Круз

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr15:75,011,947-75,047,947 36,001 bp. enter position, gene symbol or search terms go

chr15 (q24.1) p13 p12 p11.2 11.212 15q14 21.1 q21.322.2 q23 25.1



СрG – 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)

The image shows the UCSC Genome Browser track configuration interface. It is organized into several sections, each with a collapse/expand button and a 'refresh' button.

- Genes and Gene Predictions** (collapsed):
 - UCSC Genes: hide
 - RefSeq Genes: pack
 - AceView Genes: hide
 - Augustus: hide
 - CCDS: hide
 - CRISPR...: hide
 - Ensembl Genes: hide
 - EvoFold: hide
 - Exoniphy: hide
 - GENCODE...: hide
 - Geneid Genes: hide
 - Genscan Genes: hide
 - H-Inv 7.0: hide
 - IKMC Genes Mapped: hide
 - lincRNAs...: hide
 - LRG Transcripts: hide
 - MGC Genes: hide
 - N-SCAN: hide
 - Old UCSC Genes: hide
 - ORFeome Clones: hide
 - Other RefSeq: hide
 - Pfam in UCSC Gene: hide
 - Retroposed Genes: hide
 - SGP Genes: hide
 - SIB Genes: hide
 - sno/miRNA: hide
 - TransMap...: hide
 - tRNA Genes: hide
 - UCSC Alt Events: hide
 - UniProt: hide
 - Vega Genes: hide
 - Yale Pseudo60: hide
- Phenotype and Literature** (expanded): refresh
- mRNA and EST** (expanded): refresh
- Expression** (expanded): refresh
- Regulation** (expanded): refresh
 - ENCORE Regulation...: show
 - CD34 DnaseI: hide
 - CpG Islands...: show
 - ENC Chromatin...: show
 - ENC DNA Methyl...: hide
 - ENC DNase/FAIRE...: hide
 - ENC Histone...: hide
 - ENC RNA Binding...: hide
 - ENC TF Binding...: hide
 - FSU Repli-chip: hide
 - Genome Segments: full
 - NKI Nuc Lamina...: hide
 - ORegAnno: hide
 - Stanf Nucleosome: hide
 - SUNY SwitchGear: hide
 - SwitchGear TSS: hide
 - TFBS Conserved: hide
 - TS miRNA sites: hide
 - UCSF Brain Methyl: hide
 - UMMS Brain Hist: hide
 - UW Repli-seq: hide
 - Vista Enhancers: hide
- Comparative Genomics** (collapsed): refresh

А вот еще....

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

FANTOM5 disconnect refresh

[FANTOM5 summary..](#) TSS activity(read counts) TSS activity (TPM)

show hide hide

Mapping and Sequencing refresh

[Base Position](#) [P12 Fix Patches](#) [P12 Alt Haplotypes](#) [P12 Assembly](#) [Centromeres](#) [P12 Chromosome Band](#)

dense pack dense hide hide hide

[Clone Ends](#) [FISH Clones](#) [P12 Gap](#) [P12 GC Percent](#) [GRC Contigs](#) [GRC Incident](#)

hide hide hide hide hide hide

[Hg19 Diff](#) [P12 INSDC](#) [LRG Regions](#) [Mappability..](#) [P12 RefSeq Acc](#) [Restr Enzymes](#)

hide hide hide hide hide hide

[Scaffolds](#) [Short Match](#) [STS Markers](#)

hide hide hide

Genes and Gene Predictions refresh

[P12 GENCODE v32](#) [NCBI RefSeq](#) [P12 Other RefSeq](#) [P12 Updated All GENCODE..](#) [P12 AUGUSTUS](#) [CCDS](#)

pack pack hide show hide hide

[CRISPR Targets](#) [Genelid Genes](#) [P12 Genscan Genes](#) [IKMC Genes Mapped](#) [LRG Transcripts](#) [MANE select v0.6](#)

hide hide hide hide hide hide

[P12 MGC Genes](#) [Non-coding RNA..](#) [Old UCSC Genes](#) [P12 ORFeome Clones](#) [P12 Pfam in UCSC Gene](#) [RetroGenes V9](#)

hide hide hide hide hide hide

[SGP Genes](#) [SIB Genes](#) [New TransMap V5..](#) [P12 UCSC Alt Events](#) [UniProt](#)

hide hide hide hide hide

Phenotype and Literature refresh

mRNA and EST refresh

Expression refresh

Regulation refresh

[P12 ENCODE Regulation..](#) [GeneHancer](#) [P12 CpG Islands...](#) [New Hi-C and Micro-C](#) [ORegAnno](#) [RefSeq Func Elems](#)

show hide hide hide hide hide

Comparative Genomics refresh

Variation refresh

[Common SNPs\(151\)](#) [Common SNPs\(150\)](#) [Common SNPs\(147\)](#) [Common SNPs\(146\)](#) [Common SNPs\(144\)](#) [Common SNPs\(142\)](#)

dense hide hide hide hide hide

[Common SNPs\(141\)](#) [All SNPs\(151\)](#) [All SNPs\(150\)](#) [All SNPs\(147\)](#) [All SNPs\(146\)](#) [All SNPs\(144\)](#)

hide hide hide hide hide hide

[All SNPs\(142\)](#) [All SNPs\(141\)](#) [Flagged SNPs\(151\)](#) [Flagged SNPs\(150\)](#) [Flagged SNPs\(147\)](#) [Flagged SNPs\(146\)](#)

hide hide hide hide hide hide

[Flagged SNPs\(144\)](#) [Flagged SNPs\(142\)](#) [Flagged SNPs\(141\)](#) [Mult. SNPs\(151\)](#) [Mult. SNPs\(150\)](#) [Mult. SNPs\(147\)](#)

hide hide hide hide hide hide

[Mult. SNPs\(146\)](#) [Mult. SNPs\(144\)](#) [Mult. SNPs\(142\)](#) [Mult. SNPs\(141\)](#) [DGV Struct Var](#) [New gnomAD Variants](#)

hide hide hide hide hide hide

[New Platinum Genomes](#)

hide

Repeats refresh

refresh

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: genome: assembly:

group: track:

table:

region: ☐ genome ☐ ENCODE Pilot regions ☒ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to ☐ [Galaxy](#) ☐ [GREAT](#) ☐ [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

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 Table Browser 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 Сур1А1 (в границах chr15:75011947-75047947)

The Encyclopedia of DNA Elements (ENCODE)

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

The screenshot displays the ENCODE website interface. On the left is a blue sidebar with navigation links: Human Data at UCSC, Downloads, Experiment Matrix, Search, Genome Browser (hg19), Experiment List, Cell Types, Mouse Data at UCSC, Downloads, Experiment Matrix, Search, Genome Browser (mm9), Experiment List, Cell Types, Metadata Terms, Registered Variables, Antibodies, Other Resources, News Archive, First Production (2007-2012), Pilot (2003-2007), and Contacts. The main content area is titled 'Encyclopedia of DNA Elements at UCSC 2003 - 2012'. It includes an 'About' section explaining the ENCODE Consortium's goal to build a comprehensive list of functional elements in the human genome. Below this are four panels: 'Explore ENCODE data (2003 - 2012) at UCSC' showing a heatmap, 'View ENCODE data (2003 - 2012) in the UCSC Genome Browser' showing a genomic track, 'Search for data (current) at the ENCODE Portal' showing a search results table, and 'Search for ENCODE tracks (2003 - 2012) in the UCSC Browser' showing a search results table.

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).

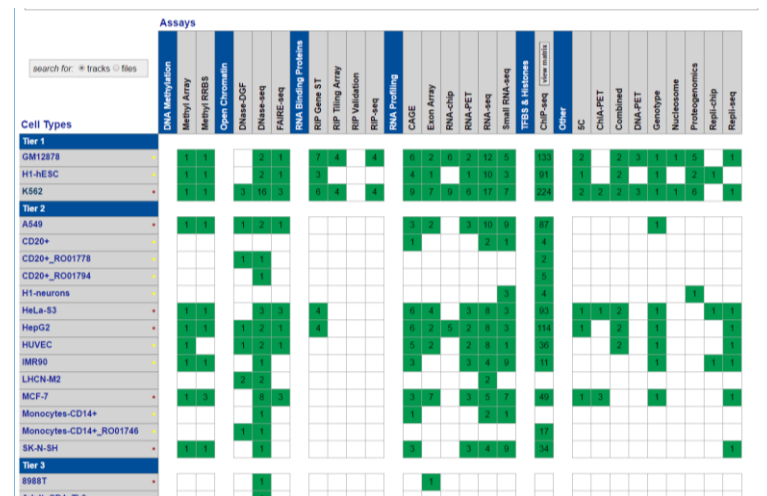
- Результаты, полученные с 2003 по 2012 гг. доступны через геномный браузер Калифорнийского университета (г.Санта Круз)
- Результаты, полученные с 2007 г и позднее доступны на сайте проекта Encode (encodeproject.org).

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>

Типы экспериментов

View matrix 

Типы клеток (~ 260)

search for: ☐ tracks ☒ files

Assays

DNA Methylation

Methyl Array

Methyl RRBS

Open Chromatin

DNase-DGF

DNase-seq

FAIRE-seq

RNA Binding Proteins

RIP Gene ST

RIP Tiling Array

RIP Validation

RIP-seq

RNA Profiling

CAGE

Exon Array

RNA-chip

RNA-PET

RNA-seq

Small RNA-seq

TFBS & Histones

ChIP-seq

Other

5C

ChIA-PET

Combined

DNA-PET

Genotype

Nucleosome

Proteogenomics

Repli-chip

Repli-seq

view matrix

Cell Types

Tier 1

GM12878

H1-hESC

K562

Tier 2

A549

CD20+

CD20+_RO01778

CD20+_RO01794

H1-neurons

HeLa-S3

HepG2

HUVEC

IMR90

LHCN-M2

MCF-7

1	1		2	1		7	4		4		6	2	6	2	12	5		133		2		2	3	1	1	5		1
1	1			2	1	3					4	1		1	10	3		91		1		2		1		2	1	
1	1	3	16	3		6	4		4		9	7	9	6	17	7		224		2	2	2	3	1	1	6		1
1	1		1	2	1						3	2		3	10	9		87						1				
											1				2	1		4										
		1	1															2										
				1														5										
					1													4								1		
1	1				3	3		4			6	4		3	8	3		93		1	1	2		1			1	1
1	1		1	2	1		4				6	2	5	2	8	3		114		1		2		1				1
1			1	2	1						5	2		2	8	1		36				2		1				1
1	1			1							3			3	4	9		11					1			1	1	
		2	2												2													
1	3		8	3							3	7		3	5	7		49		1	3			1				1

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)

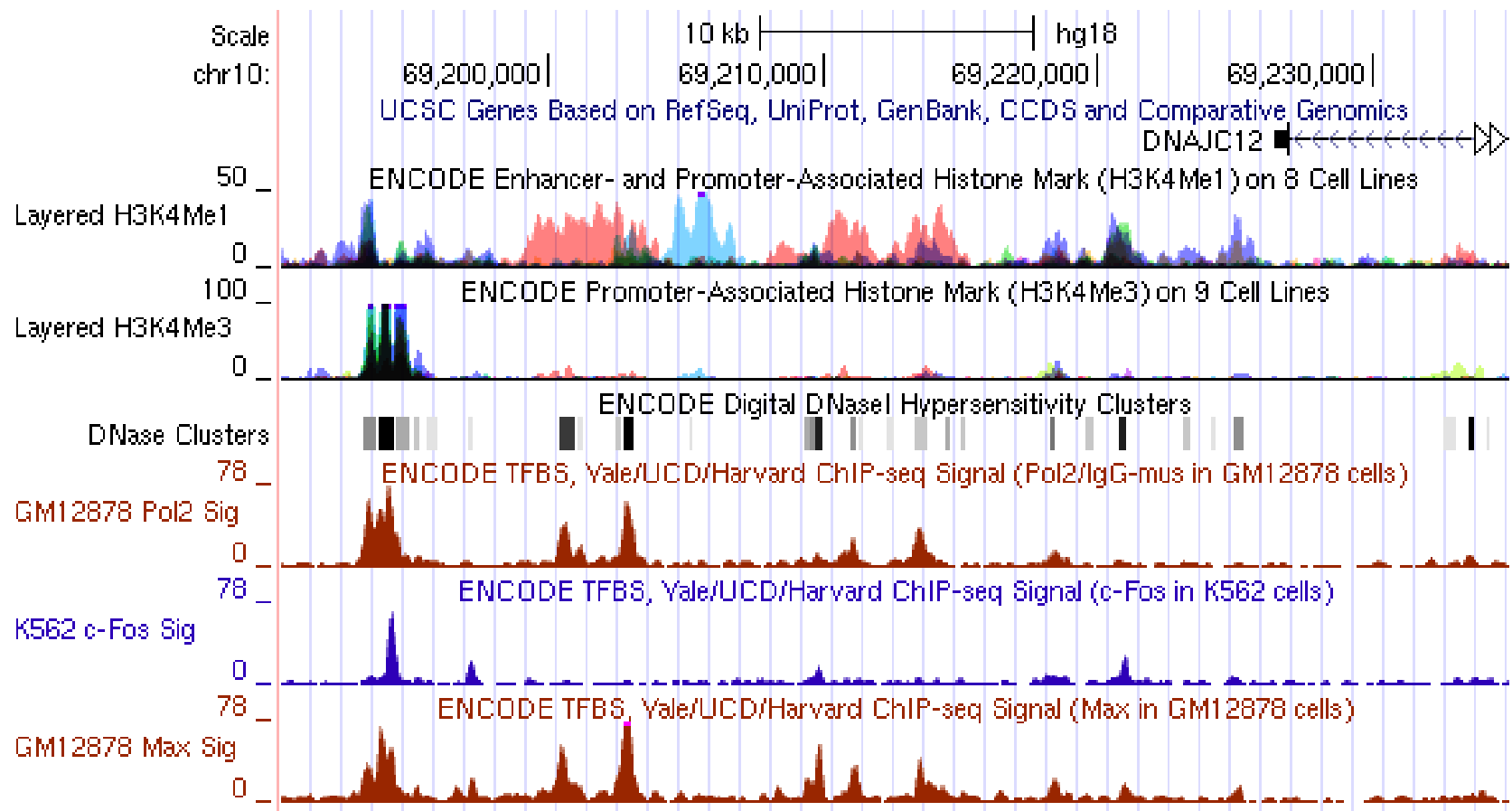
Histone Modification
(12)

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

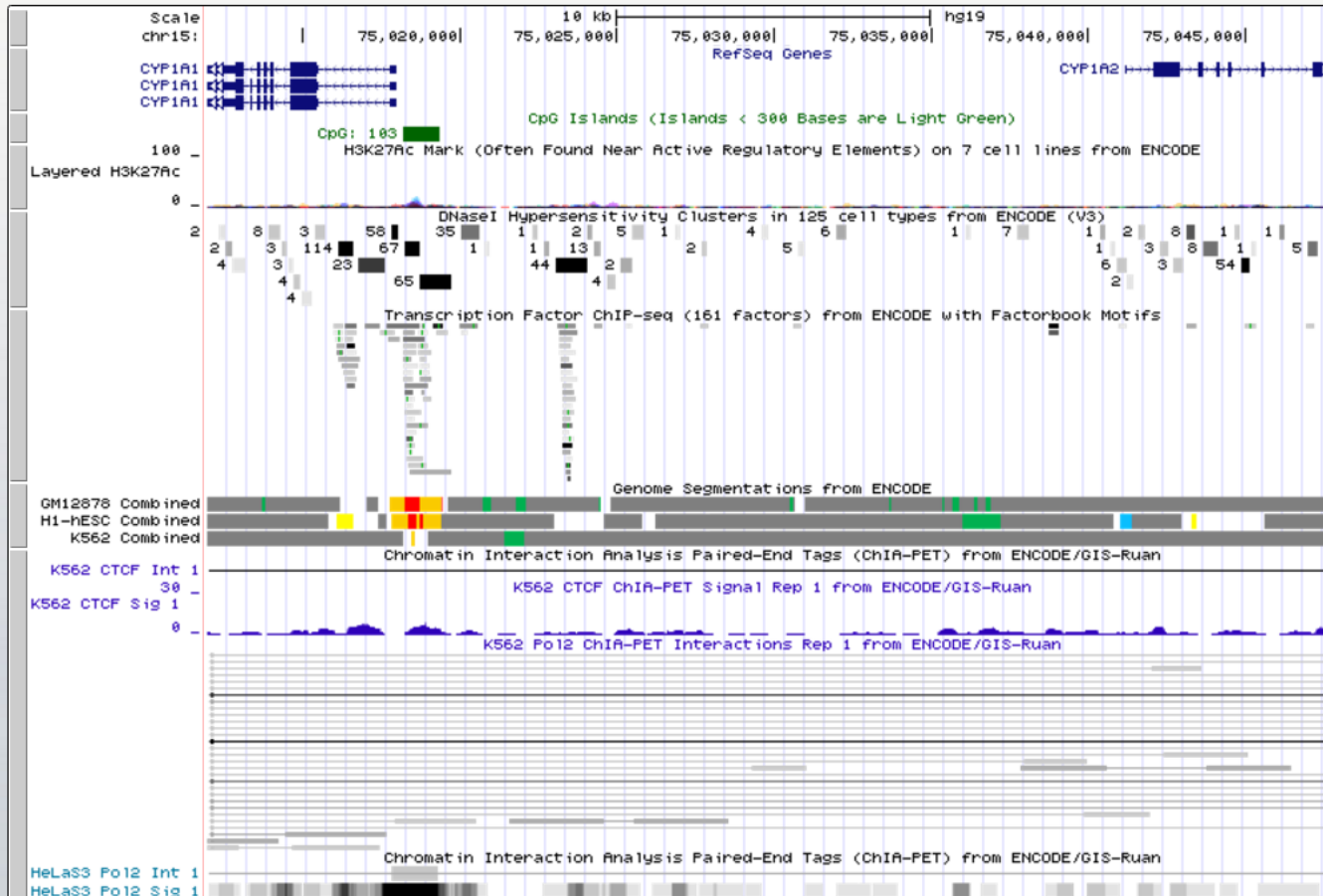
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr15:75,011,947-75,047,947 36,001 bp. enter position, gene symbol or search terms

go

chr15 (q24.1) p13 p12 p11.2 11.212 15q14 21.1 q21.322.2 q23 25.1



СрG – богатые участки

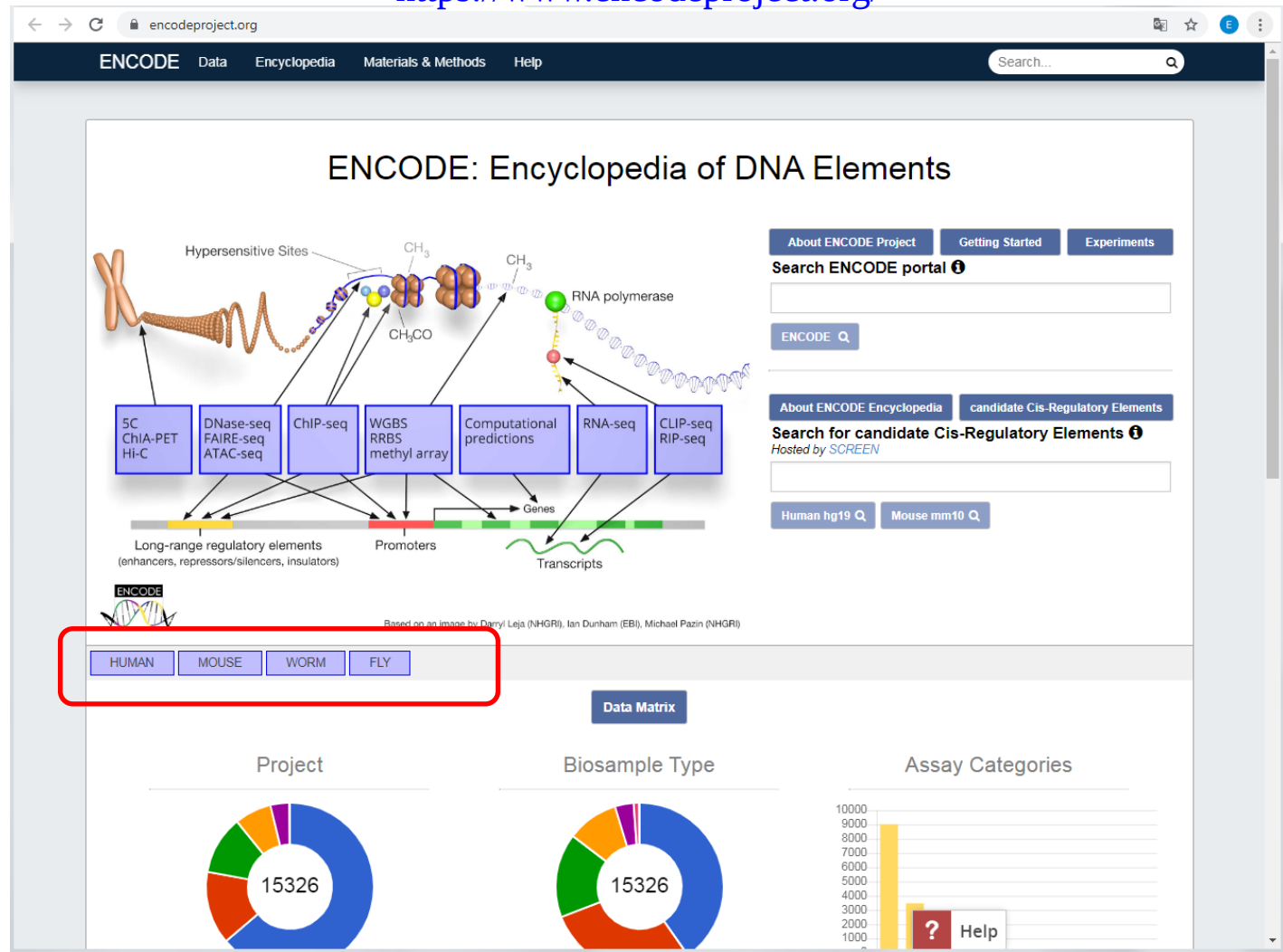
Сайты гиперчувствительности к DNaseI (открытый хроматин)

Районы связывания с различными ТФ по данным экспериментов ChIP-seq

Типы хроматина в различных типах клеток

3D структура хроматина: контакты между удаленными участками генома

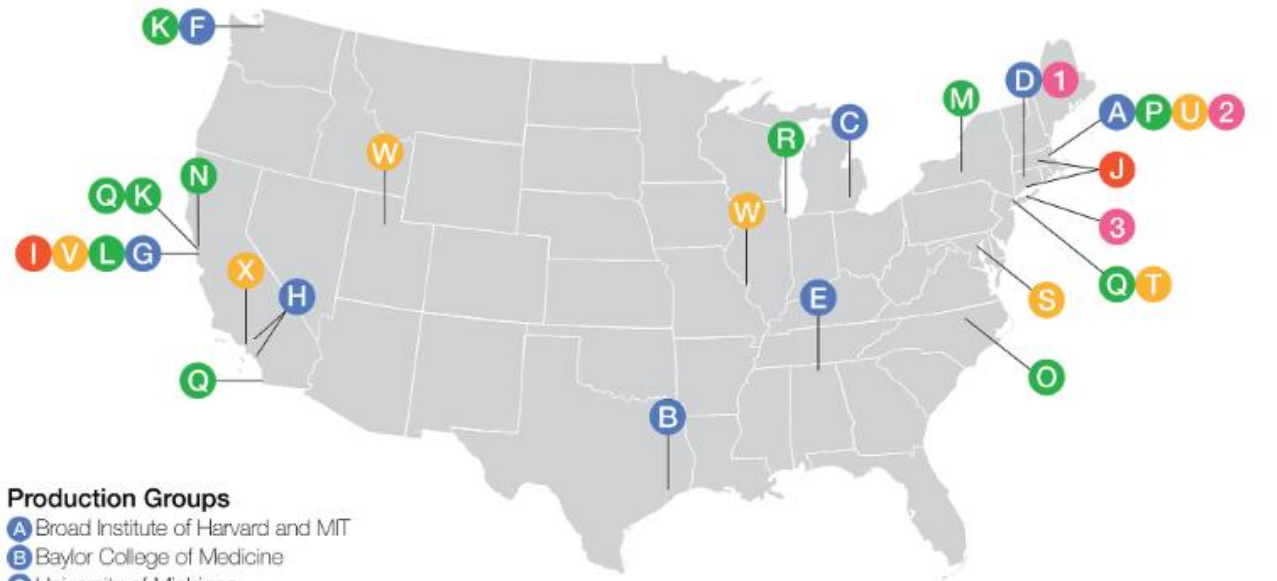
Результаты, полученные в проекте ENCODE с 2007 г и позднее доступны на сайте проекта Encode
<https://www.encodeproject.org/>



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

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

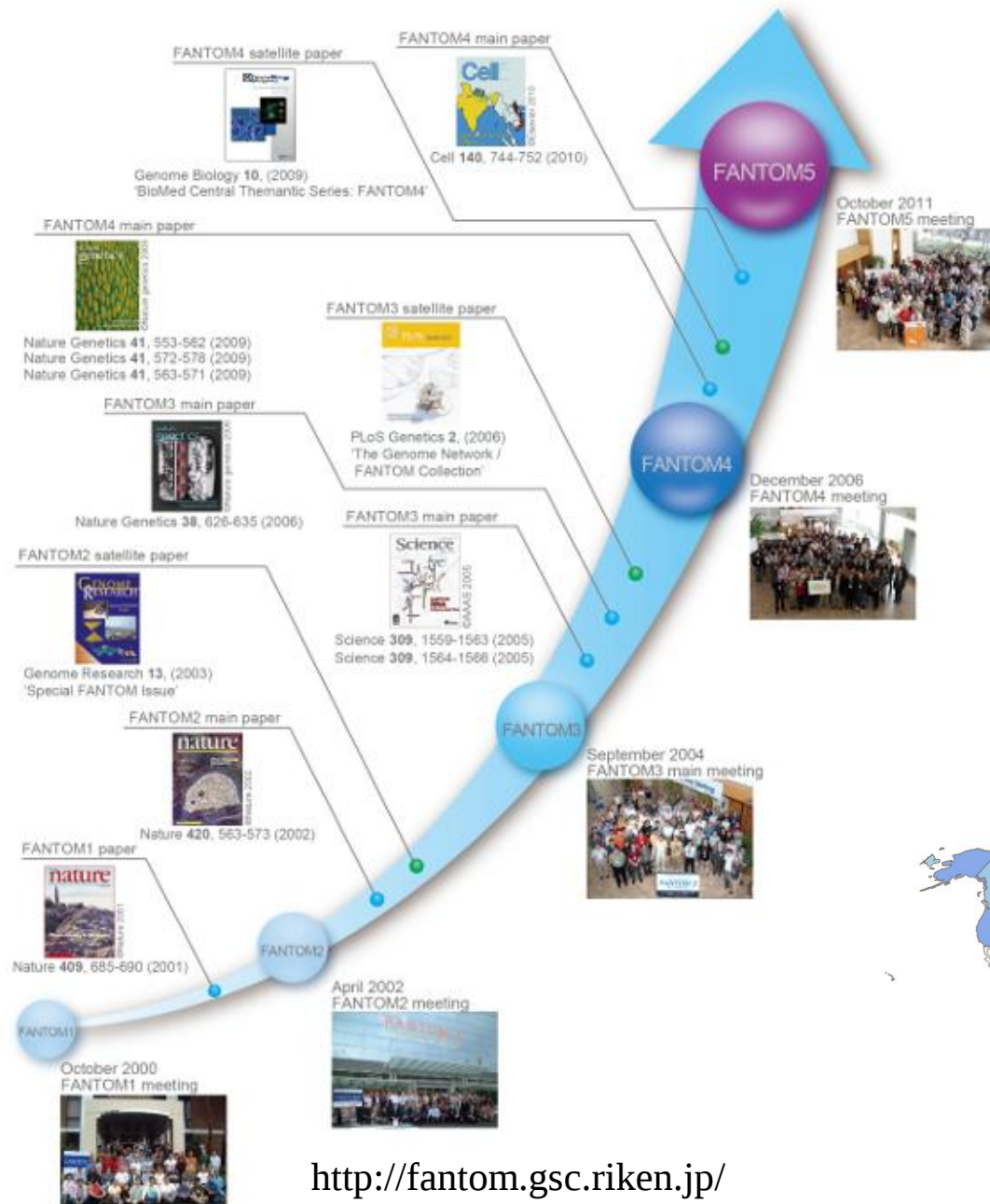
- S** Johns Hopkins University
- T** Memorial Sloan Kettering Cancer Center
- U** Harvard University; Brigham and Women's Hospital
- V** Stanford University
- W** Washington University; University of Utah
- X** University of California, Los Angeles

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.

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)

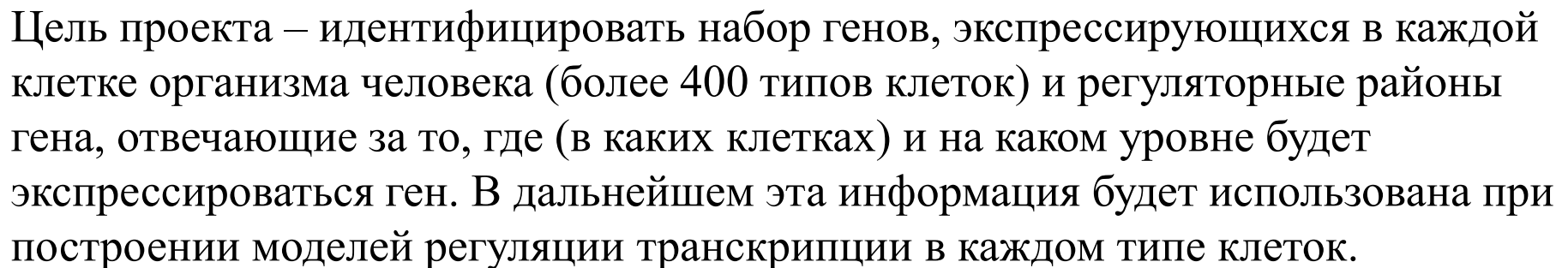
Members

Over 500 FANTOM members from more than 20 countries.



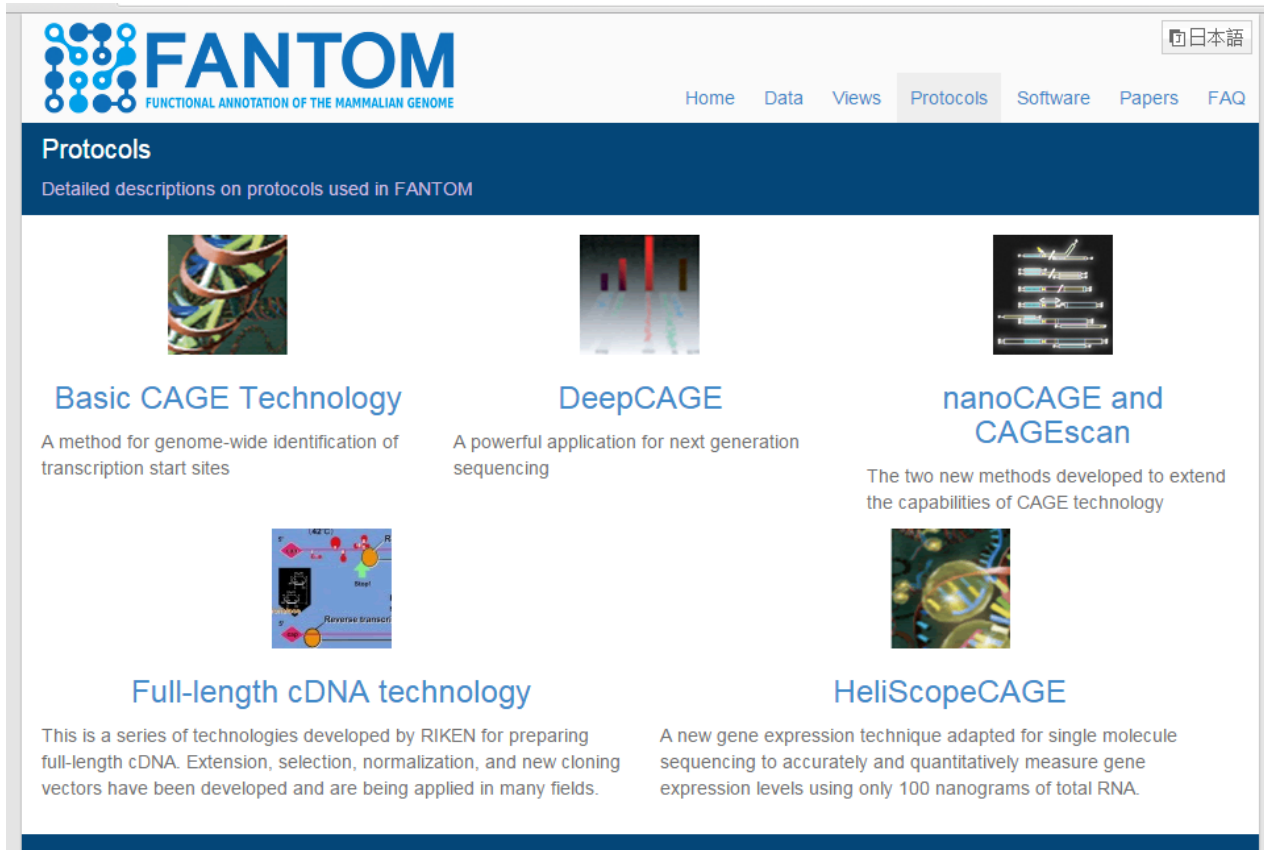
<http://fantom.gsc.riken.jp/>

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



FANTOM5: экспериментальные методы

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



The screenshot shows the FANTOM website's 'Protocols' page. The header includes the FANTOM logo (Functional Annotation of the Mammalian Genome) and navigation links: Home, Data, Views, Protocols (selected), Software, Papers, and FAQ. A language selector for '日本語' is in the top right. The main section is titled 'Protocols' with a subtitle 'Detailed descriptions on protocols used in FANTOM'. It features five protocol cards, each with an icon, title, and description:

- Basic CAGE Technology**: A method for genome-wide identification of transcription start sites. (Icon: DNA double helix)
- DeepCAGE**: A powerful application for next generation sequencing. (Icon: Bar chart)
- nanoCAGE and CAGEscan**: The two new methods developed to extend the capabilities of CAGE technology. (Icon: Microarray)
- Full-length cDNA technology**: This is a series of technologies developed by RIKEN for preparing full-length cDNA. Extension, selection, normalization, and new cloning vectors have been developed and are being applied in many fields. (Icon: Microscope)
- HeliScopeCAGE**: A new gene expression technique adapted for single molecule sequencing to accurately and quantitatively measure gene expression levels using only 100 nanograms of total RNA. (Icon: DNA double helix)

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/>

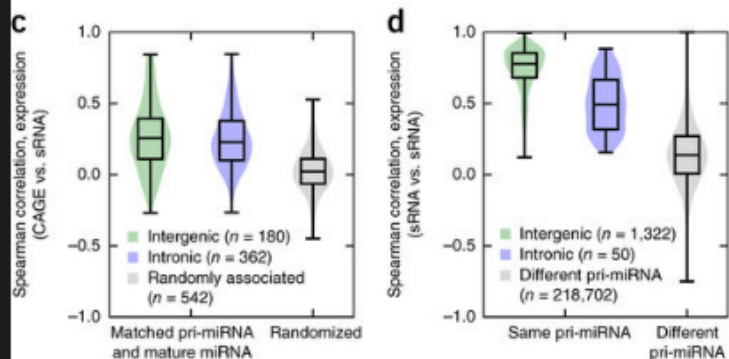


日本語

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/>



FUNCTIONAL ANNOTATION OF THE MAMMALIAN GENOME

HomeDataViewsProtocolsSoftwarePapersFAQ

Data Resources

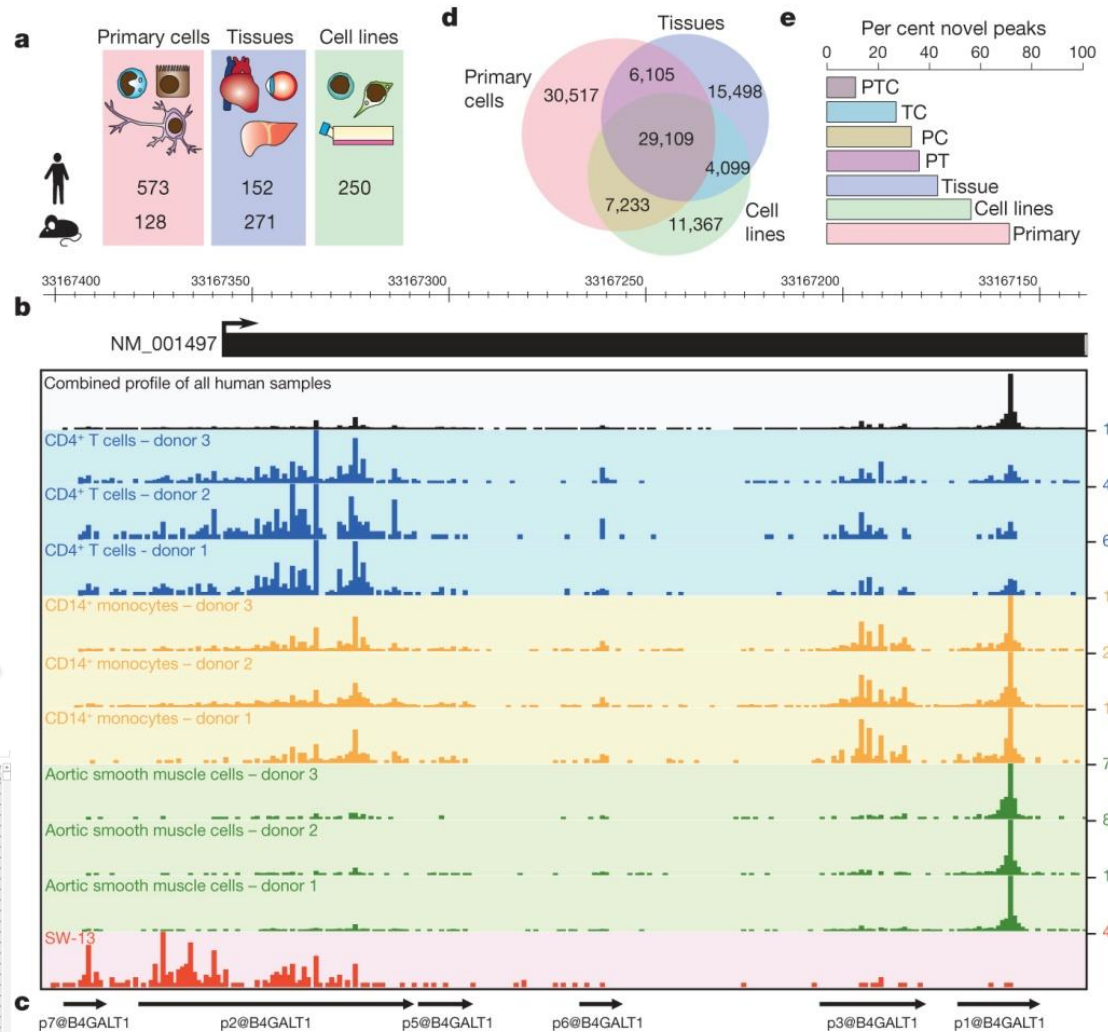
Resources for accessing FANTOM raw and processed data

FANTOM5 Data

- FANTOM5 Data Summary
Selected data files for FANTOM5 and instructions for bulk download.
- TET
A system for extracting subsets of data from selected FANTOM5 data files including phase 1 and 2 expression tables.
- BioMart
A BioMart instance to download user defined subsets of the FANTOM5 promoters.
- Transcribed Enhancer Atlas
A database describing enhancer regions defined by CAGE tags in the FANTOM5 project, with samples over the whole human body.
- FANTOM CAT
An atlas of human long non-coding RNAs with accurate 5' ends

FANTOM5 Supplemental Data

- Kanamori-Katayama et al. 2011
An automated workflow to sequence CAGE libraries.
- Motakis et al. 2013
Redefinition of the human mast cell transcriptome by deep sequencing
- Kawaji et al. 2013
Comparison of CAGE and RNA-seq transcriptome profiling using a clonally amplified and single molecule next generation sequencing
- Verardo et al. 2014



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Sample type	Species	description	supplier	sample ID/Accession	Catalog number	external URL	Cell number	donor/Cell lot	Sex	Age	DOB	Q20 map		
1	cell line	human squamous carcinoma cell line HCC1306	company: ATCC	10177	CH1306	RC1306	http://www.atcc.org/ATCCData/Accession/10177	2-3	female	80	10			
2	tissue	human achilles tendon, donor2	Collaboration: AM	10292	CH1343	RC1343	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10292	12-40	male	0	2-4			
3	cell line	human acute lymphoblastic leukemia (B-ALL) cell line BAAL-1	company: RIKEN	10453	CH11251	RC10256	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10453	unknown	unknown	0	9-3			
4	cell line	human acute lymphoblastic leukemia (B-ALL) cell line NBAL-8	company: RIKEN	10548	CH11282	RC11333	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10548	unknown	unknown	0	9-3			
5	cell line	human acute lymphoblastic leukemia (T-ALL) cell line HPB-ALL	company: RIKEN	10429	CH10746	RC11335	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10429	unknown	unknown	0	9-3			
6	cell line	human acute lymphoblastic leukemia (T-ALL) cell line Jurkat	company: RIKEN	10464	CH11243	RC10809	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10464	144	9-3					
7	cell line	human acute myeloid leukemia (FAB M0) cell line Kasumi-1	company: JAPAN	10792	CH11241	RC11004	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10792	37	no data					
8	cell line	human acute myeloid leukemia (FAB M0) cell line K562	company: RIKEN	10872	CH13053	RC11366	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10872	59	10					
9	cell line	human acute myeloid leukemia (FAB M1) cell line IMPT-1	company: RIKEN	10828	CH13054	RC11297	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	62	9-4					
10	cell line	human acute myeloid leukemia (FAB M2) cell line Kasumi-1	company: JAPAN	10798	CH13052	RC11003	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10798	7	8-4					
11	cell line	human acute myeloid leukemia (FAB M2) cell line Kasumi-6	company: JAPAN	10792	CH13052	RC11004	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10792	64	9-4					
12	cell line	human acute myeloid leukemia (FAB M2) cell line NBAL-3	company: JAPAN	10792	CH13054	RC11006	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10792	33	9-4					
13	cell line	human acute myeloid leukemia (FAB M3) cell line NBAL-3	company: RIKEN	10828	CH13055	RC11005	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	30	9-4					
14	cell line	human acute myeloid leukemia (FAB M3) cell line NBAL-3	company: RIKEN	10828	CH13053	RC11338	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	40	6-3					
15	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: RIKEN	10828	CH13054	RC11296	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	45	9-4					
16	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: RIKEN	10828	CH13054	RC11296	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	30	10					
17	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: RIKEN	10828	CH13054	RC11296	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	30	10					
18	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: RIKEN	10828	CH13054	RC11296	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10828	30	10					
19	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: JAPAN	10794	CH13050	RC10804	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10794	unknown	unknown	0	10			
20	cell line	human acute myeloid leukemia (FAB M4) cell line NBAL-3	company: JAPAN	10794	CH13051	RC10805	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10794	unknown	unknown	0	10			
21	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
22	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
23	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
24	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
25	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
26	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
27	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
28	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
29	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
30	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
31	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
32	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
33	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
34	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
35	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
36	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
37	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
38	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
39	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
40	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
41	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
42	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
43	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
44	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
45	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
46	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
47	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
48	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
49	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
50	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
51	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
52	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
53	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
54	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
55	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
56	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
57	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
58	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
59	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
60	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
61	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
62	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
63	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
64	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
65	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
66	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
67	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
68	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
69	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
70	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
71	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation	10899	CH10722	RC10722	http://www2.riken.jp/fantom/cell/detail.cgi?cell=10899	1	9-4					
72	cell line	human acute myeloid leukemia (FAB M5) cell line NBAL-3 (fresh)	In-house isolation											

→ ↻ fantom.gsc.riken.jp/6/  ☆ 

 **FANTOM6**
FUNCTIONAL ANNOTATION OF THE MAMMALIAN GENOME

日本語

Home Data Views Protocols Software Papers FAQ

The FANTOM6 project

Functional annotation of long non-coding RNAs in the human genome

Knockdown



285 lncRNAs

194 lncRNAs Successful KD

68%

Cell Phenotype



194 lncRNAs

Growth & Morphology

30%

CAGE transcriptomic profiling



119 lncRNAs

Function

11%

Introduction to FANTOM6

FANTOM (Functional ANnoTation Of the Mammalian genome) is a worldwide collaborative project aiming at identifying all functional elements in mammalian genomes. The goal of the sixth edition of the FANTOM project (FANTOM6) is to systematically elucidate the function of long non-coding RNAs (lncRNAs) in the human genome.

We envision the following as the main experimental strategy for FANTOM6.

- We are generating a reference set of genome-wide profiles to establish the basal state of the transcriptome and epigenome in each cell type.
- We are perturbing a set of lncRNAs in each cell type, and evaluating the molecular phenotype of the perturbation by CAGE profiling followed by bioinformatics analysis. The lncRNAs selected for perturbation are based on published transcripts as well as yet unpublished RNAseq assemblies specific to the cell types in the FANTOM5 collection.
- A selected subset of lncRNAs will be functionally characterized in more detail using other complementary technologies.

Functional annotation of human long noncoding RNAs via molecular phenotyping

We published a pilot study of FANTOM6, the latest edition of the project. Looking at human long non-coding RNAs, which outnumber protein-coding genes in mammals but whose function is still poorly understood. We selectively targeted

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>



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

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

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

TRANSFAC® Professional vs. Public

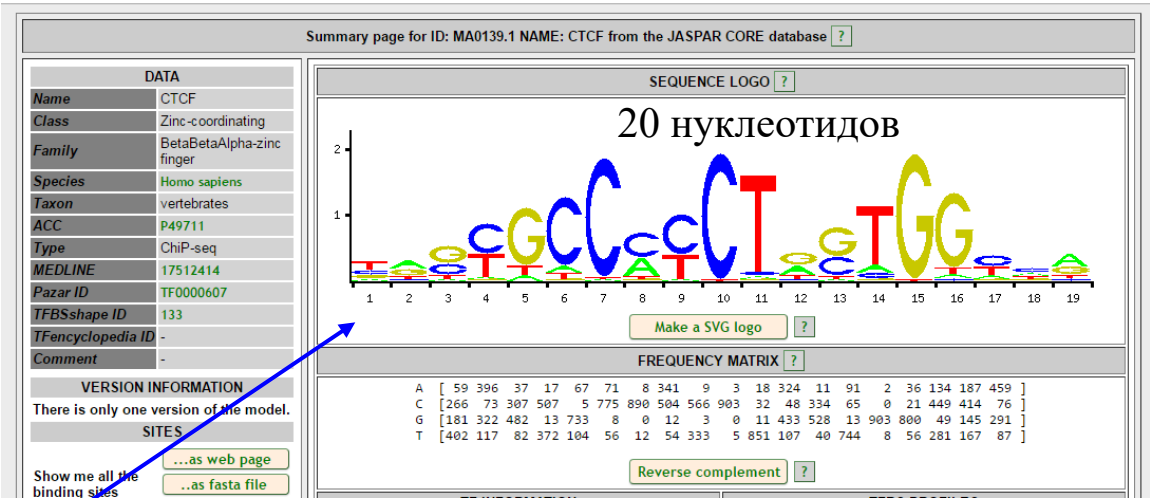
TRANSFAC® Professional is an internationally unique knowledgebase containing published data on eukaryotic transcription factors and their regulated genes, coupled with transcription factor binding site prediction tools for advanced research. The professional version content is more than six years ahead 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

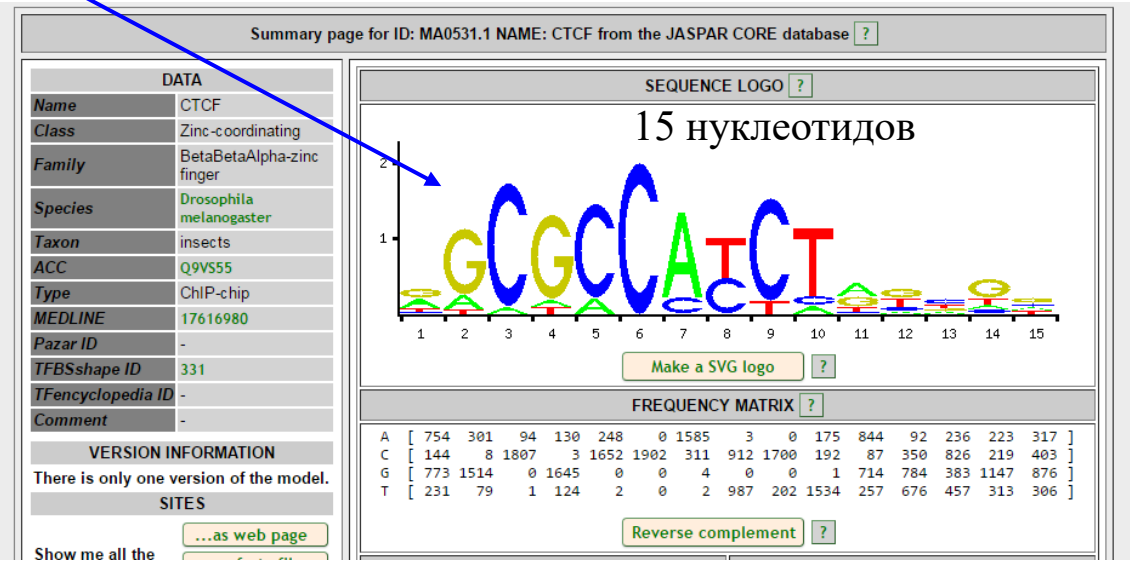
The screenshot displays the JASPAR 2020 website. The top navigation bar includes the JASPAR 2020 logo, a search bar with the text "Search JASPAR database...", and links for "Cart" and "JASPAR Blog". A left sidebar contains a menu with items: Home, About, Search, Browse JASPAR CORE, Unvalidated Profiles, Browse Collections, Tools, RESTful API, Download Data, Matrix Clusters, and Genome Tracks. The main content area features a search bar with examples (SPI1, P17676, ChIP-seq, Homo sapiens) and a "Search" button. Below this is a section titled "Browse JASPAR CORE for six different taxonomic groups" with six image-based buttons: Vertebrata, Nematoda, Insecta, Plantae, Fungi, and Urochordata. To the right is a large banner for "The high-quality transcription factor binding profile database" with a "JASPAR interactive tour" button. Below the banner is a section titled "Citing JASPAR 2020" with a citation for Fornes O, Castro-Mondragon JA, Khan A, et al. (2019) and links to PubMed, NAR, and PDF. At the bottom, there are four colored boxes: "Unvalidated profiles" (blue), "Q&A Forum" (green), "RESTful API" (orange), and "Download" (red).

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

Пример: две матрицы сайтов связывания CTCF

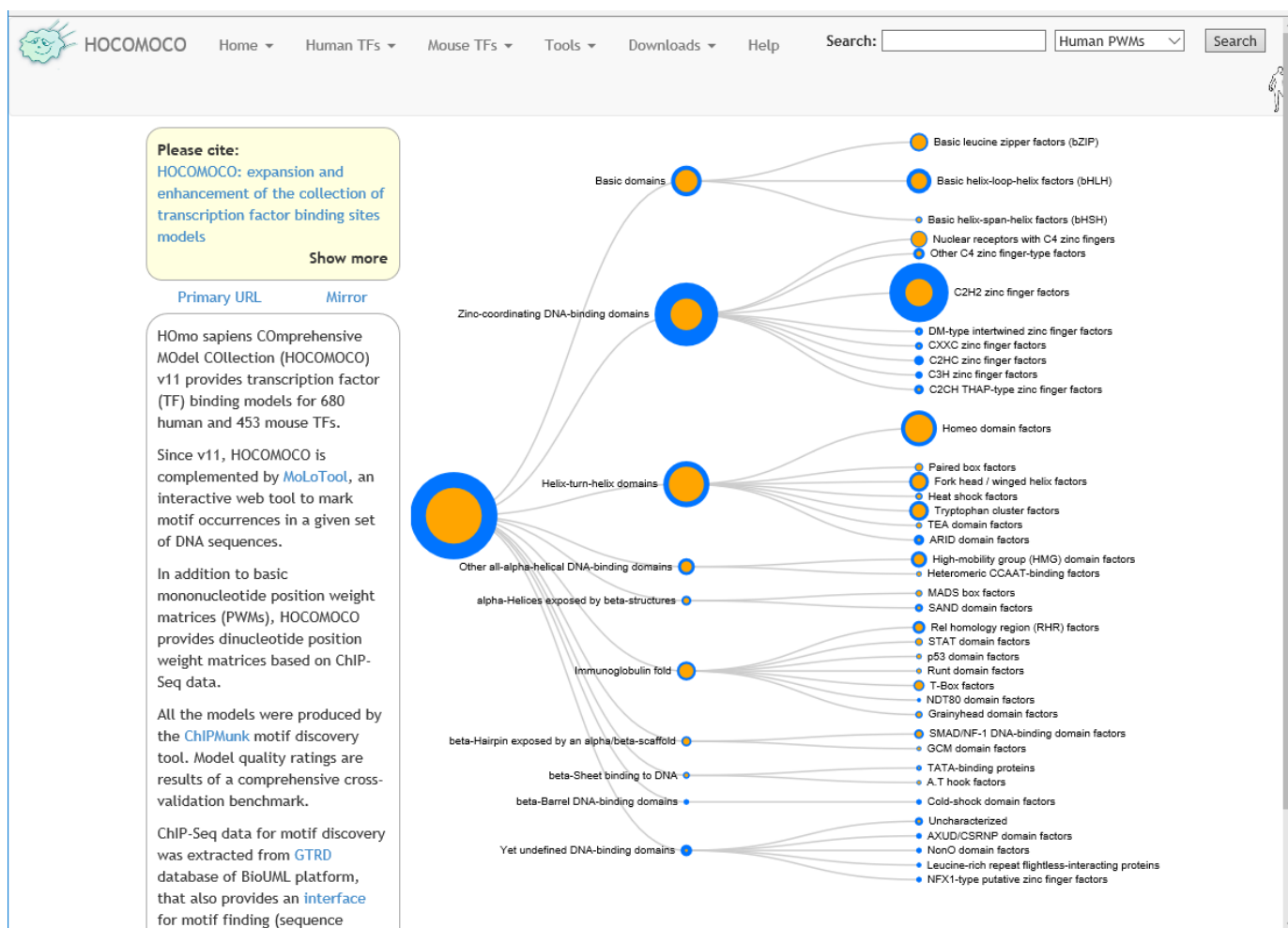


Графики WEB-logo



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.

HOmo sapiens COmprehensive MOdel COllection (HOCOMOCO) v11

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

OCO

Home

Human TFs

Mouse TFs

Tools

Downloads

Help

Search

Human PWM

Model Info

Transcription factor

NFSA1
(GeneCards)

Model

STF1_HUMAN.H11MO.0.B

Model type

Mononucleotide PWM

LOGO

LOGO (reverse complement)

Data source

ChIP-Seq

Model release

HOCOMOCCv11

Model length

11

Quality [↗](#)

8

Motif rank [↗](#)

0

Consensus

NYCAAGGYCW

Best suRCC (human)

0.798521387649958

Best suRCC (mouse)

0.897739768837399

Peak sets to benchmark (human)

1

Peak sets to benchmark (mouse)

2

Aligned words

500

TF family

FTZ-F1-related receptors (NR3(2, 1.5))

TF subfamily

FTZ-F1 (EF-1) (NFSA1(2, 1.5.0, 1))

HGNC

7983

EntrezGene

7516
(OSTATG)

UniProt ID

STF1_HUMAN

UniProt AC

Q31285
(TFClass)

Comment

Downloads

actn

path

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

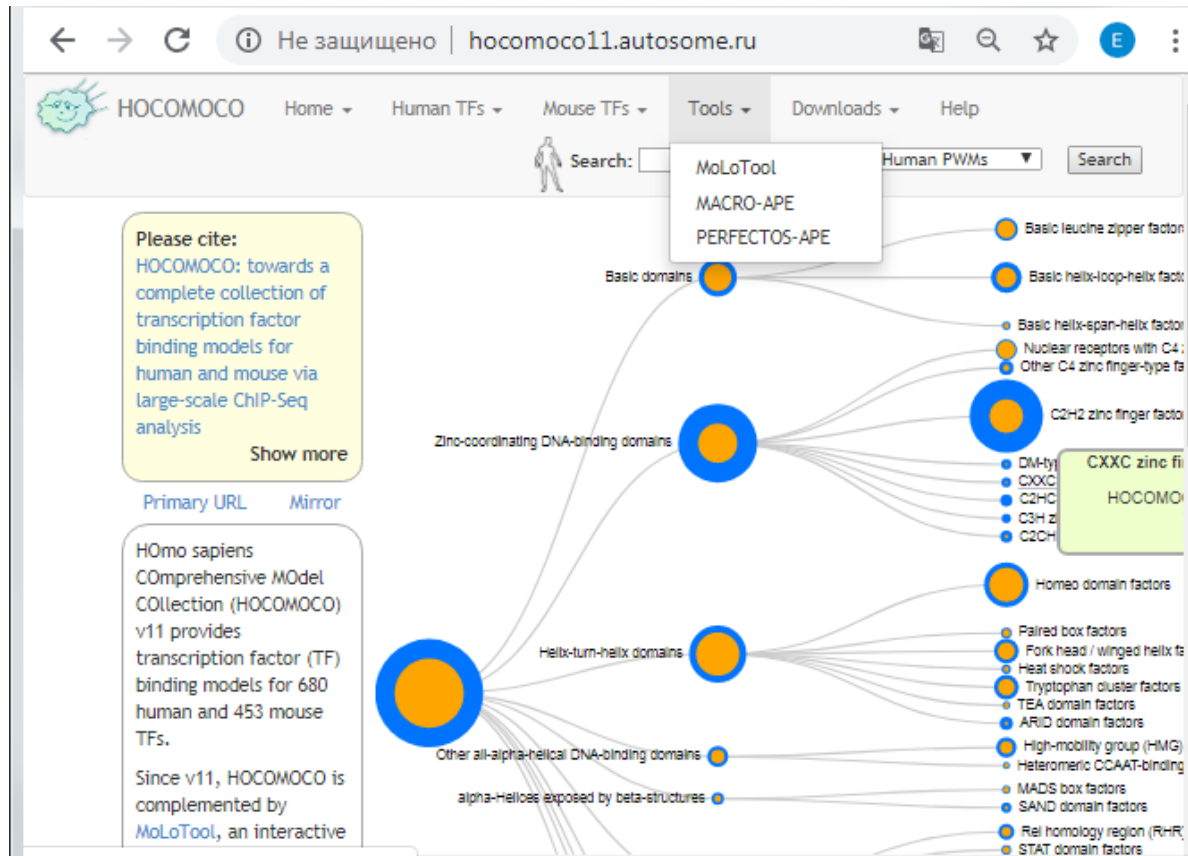
NFSA1 expression

ReMap ChIP-seq dataset list

NFSA1 datasets

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 Funconal Effect by Approximate P-value Estimation**

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

- компиляция данных по матрицам из других баз

Home

Tools

View cart

Bulk downloads

Database stats

Contact us

Help

Update Log

FAQ

Links

How to cite

CIS-BP

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

Search for a TF

By Identifier
(e.g. Gata*, YEL009C, ISFTZ_01)

Browse TFs / Restrict Search for TFs

By Model Organism

By Any Species

By Domain Type

By Motif Evidence

By Evidence Type

By Study

Database BuildVersion 2.00

GO!

Database build 2.00 now available!

Last updated

Current content: 11491 motifs. 165030 TFs with at least one binding motif (4559 from direct experiments), out of a total of 392333 TFs from 3

Home

Tools

View cart

Bulk downloads

Database stats

Contact us

Help

Update Log

FAQ

Links

How to cite

Scan single sequences for TF binding

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

Species

Motif model

Threshold

☐ Scan TFs in my cart

Scan two sequences for differential TF binding

Allele 1
TCCTCACGGGAACC

☒ SNP Scan
Click for [sample](#)

Allele 2
TCCTACCGGAACC

Species

Motif model

Threshold

8 mers - Scores 0.45

☐ Scan TFs in my cart

Protein Scan

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

Motif Scan

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

Species

Scan all species

GO!

Last updated: Jan 8th, 2019 Database Build 2.00

Current database contents: 165030 TFs with at least one binding motif (4559 from direct experiments), out of a total of 392333 Eukaryotic TFs from 321 families in 741 species

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	Source	Type	Author	Year	PMID	Num TFs	Pct TFs
Tools	Transfac	Transfac	Matys	2006	16381825	958	0.2
View cart	HocoMoco	Misc	Kulakovskiy	2013	23175603	747	0.2
Bulk downloads	Yin2017	SELEX	Yin	2017	28473536	541	0.1
Database stats	OMalley2016	Dap-seq	OMalley	2016	27203113	534	0.1
Contact us	JASPAR	JASPAR	Mathelier	2014	24194598	486	0.1
Help	Jolma2013	SELEX	Jolma	2013	23332764	452	0.1
Update Log	FlyFactorSurvey	B1H	Zhu	2011	21097781	297	0.1
FAQ	Nitta2015	SELEX	Nitta	2015	25779349	248	0.1
Links	Barazandeh2018	ChIP-seq_a	Barazandeh	2018	29146583	221	0.1
How to cite	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.

- ВОЗМОЖНОСТЬ предсказания сайтов по матрицам

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

[Home](#) [Tools](#) [View cart](#) [Bulk downloads](#) [Database stats](#) [Contact us](#) [Help](#) [Update Log](#) [FAQ](#) [Links](#) [How to cite](#)

Scan single sequences for TF binding

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

>Example 1
ACCCGGGAAACGATGA
>Example 2
GCATGCCACTGCCACCTATCATTCCATG

Species:

Motif model: Threshold:

☐ Scan TFs in my cart

Scan two sequences for differential TF binding

☐ SNP Scan [Click for sample](#)

Allele 1:

Allele 2:

Species:

Motif model: Threshold:

☐ Scan TFs in my cart

Protein Scan

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

Motif Scan

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

Species:



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

TRANSCRIPTION REGULATORY REGIONS DATABASE (TRRD)

ИЦиГ СО РАН, Новосибирск, Россия

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

РЕГУЛЯТОРНЫЕ РАЙОНЫ, ССТФ, ТРАНСКРИПЦИОННЫЕ ФАКТОРЫ

ПУБЛИКАЦИЯ ПО TRRD в Nucleic Acids Research



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

© 2002 Oxford University Press

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

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

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Lavrent'evskiy pr. 10, 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, Lavrent'evskiy pr. 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 publications. TRRD release 6.0 comprises the information on 1167 genes, 5637 transcription factor binding sites, 1714 regulatory regions, 14 locus control regions and 535 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.), TRRDSTES (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 the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia) since 1993. The main goal while developing TRRD was to provide a unique 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.), TRRDSTES (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 (TRRDSTES); (ii) regulatory units, including promoters, enhancers and silencers (TRRDUNITs); (iii) regulatory regions, including 5'- and 3'-regulatory regions, exon and intron (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 RegLink (RP) and SiteLink (RS) of this database are hyperlinked to textual descriptions of regulatory units (promoters, enhancers and silencers, described in TRRDUNITs) and transcription factor binding sites (TRRDSTES) 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, summarized 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 333408; Fax: +7 3832 333278; Email: kol@bionet.nsc.ru

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



HOME DNA RNA PROTEIN GENENETWORKS MAP

**TRANSCRIPTION REGULATORY
REGIONS DATABASE**



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

ACCESS to TRRD:

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

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

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

General information

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

User's guide

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

How is TRRD updated ?

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

Current TRRD release

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

Done

Internet

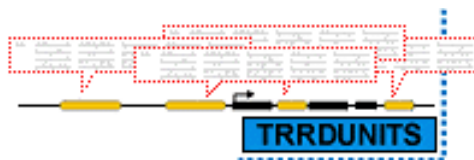
TRRD ВКЛЮЧАЕТ ДАННЫЕ О:

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

САЙТОВ СВЯЗЫВАНИЯ
ТРАНСКРИПЦИОННЫХ
ФАКТОРОВ



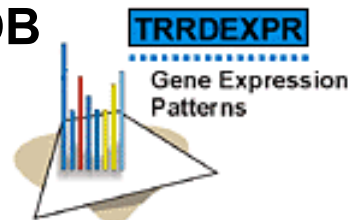
РЕГУЛЯТОРНЫХ РАЙОНОВ
(ПРОМОТОРОВ, ЭНХАНСЕРОВ,
САЙЛЕНСЕРОВ)



ЛОКУС-КОНТРОЛИРУЮЩИХ
РАЙОНОВ

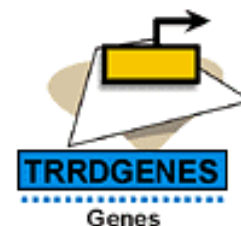
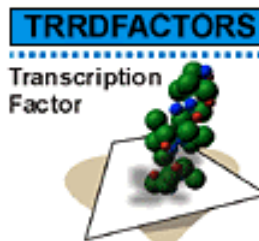
TRRDLCR

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

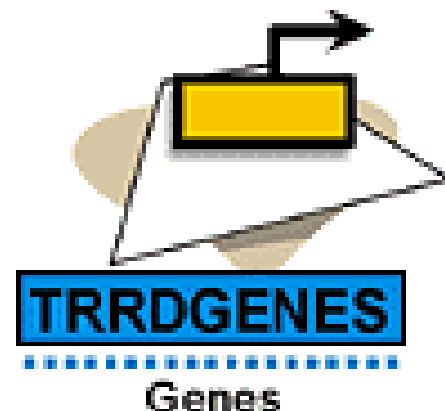
AP [REGULATORY UNIT: P01087](#)

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

AP [REGULATORY UNIT: P00760](#)

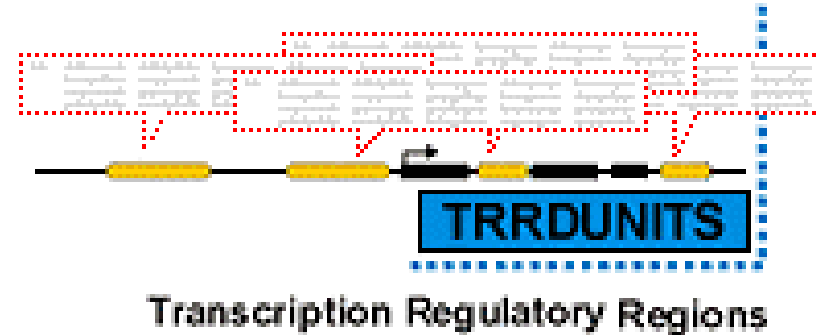
PR promoter; ST; -160 to -1; [S2907](#), [S3931](#), [S3932](#), [S3933](#), [S3934](#), [S3935](#)

//



TRRDUNITS

ОПИСАНИЕ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ ЕДИНИЦ (ПРОМОТОРОВ, ЭНХАНСЕРОВ, САЙЛЕНСЕРОВ)



<u>RegUnitAC</u>	P00562
<u>GeneID</u>	Rn:D2
<u>RegRegion</u>	5' region
<u>RegUnit</u>	Promoter; ST; -150 to +1; S79, S80
<u>DNA BankLink</u>	EMBL; RND2RPR; X77137; 704 to 855
<u>LeftTrunc</u>	0
<u>RightTrunc</u>	0
<u>SeqLength</u>	152
<u>Sequence</u>	cccaggcccc acagtgcaga gatagttctg gggccctggg tgggtggggc ctctgtacaa ggggcggggg tcccgggcgc ctctggcca gggtgacccc gccccctcct cctgcgcagc gctctgattc cgcggagctg tccagcctca gt
<u>PromotTisSp</u>	0
<u>PromotInd</u>	1
<u>ExperimentCodes</u>	6.1.1, 6.8 [<u>Minowa T.</u> et al., 1992]

ОПИСАНИЕ ТРАНСКРИПЦИОННЫХ РЕГУЛЯТОРНЫХ ЕДИНИЦ (html –представление)

[TRRDUNITS4:P00051](#)

[RegUnitAC](#)

P00051

[GeneID](#)

[Hs:APOA1](#)

[RegRegion](#)

5'region

[RegUnit](#)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[DNA BankLink](#)

EMBL; [HSAPOA01](#); [J04066](#); 1813 to 1959

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

ccacccggga gacctgcaag cctgcagcac tccccctcccg cccccactga
acccttgacc cctgccctgc agcccccgca gcttgctgtt tgccccactct
atttgcccag ccccagggac agagctgac cttgaactct taagtgc

[DNA BankLink](#)

EMBL; [HSAPOA01](#); [J04066](#); 1813 to 1959

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

ccacccggga gacctgcaag cctgcagcac tccccctcccg cccccactga
acccttgacc cctgccctgc agcccccgca gcttgctgtt tgccccactct
atttgcccag ccccagggac agagctgac cttgaactct taagtgc

[DNA BankLink](#)

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

[LeftTrunc](#)

0

[RightTrunc](#)

0

[SeqLength](#)

147

[Sequence](#)

ccggggagac ctgcaagcct gcagcactcc cctccccccc ccaactgaacc
cttgaccctt gccctgcacg ccccgcagct tgctgtttgc ccaactctat
ttgcccagtc ccagggacag agctgatcct tgaactctta agttcca

TRRDSITES

ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА

AN S1160

ID [Gene: Hs:APOB](#)

AP [REGULATORY UNIT: P00670](#)

NM HNF-4 bs; HNF-4 binding site

NY AF-1 binding site

NY BA1 binding site

NS [R01612](#)

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

AT increase

SQ cccgggaggCGCCCTTTGGACCTtttg

PQ -88 to -62

PF -82 to -62

BF EMBL: [M15053](#) : 68

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

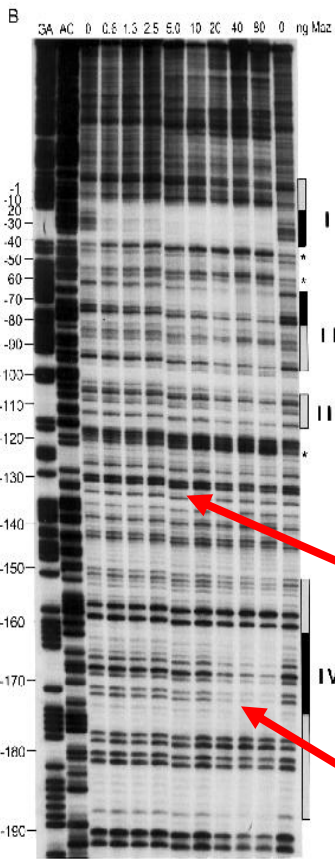
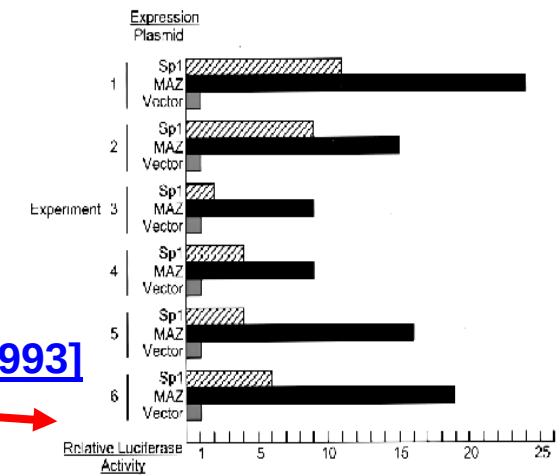
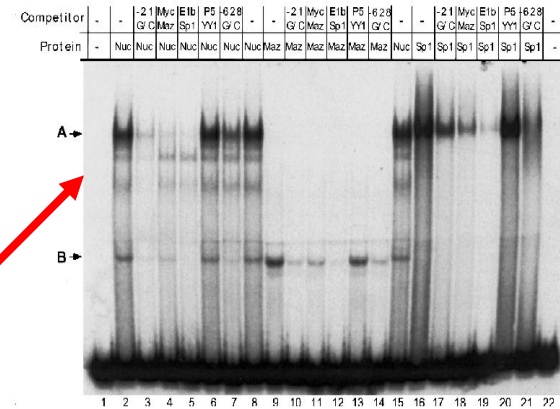
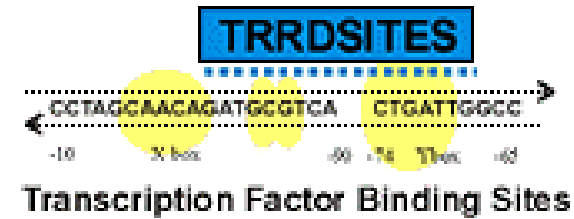
AG rat liver cells: 3.6 [\[Metzger S. et al., 1993\]](#)

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

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

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

//



TRANSIENT EXPRESSION ANALYSIS

Экспериментальные методики, по результатам выполнения которых вносятся 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

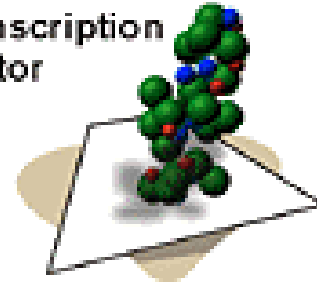
Полный список см. по адресу

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

TRRDFACTORS

TRRDFACTORS

Transcription
Factor



ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА

ID [Hs:APOA1](#)

AN [Site: S1135](#)

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

FS PPARgamma; peroxisome proliferator-activated receptor gamma

TS human

TO in vitro synthesized

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

FS RXRalpha; retinoic X receptor alpha

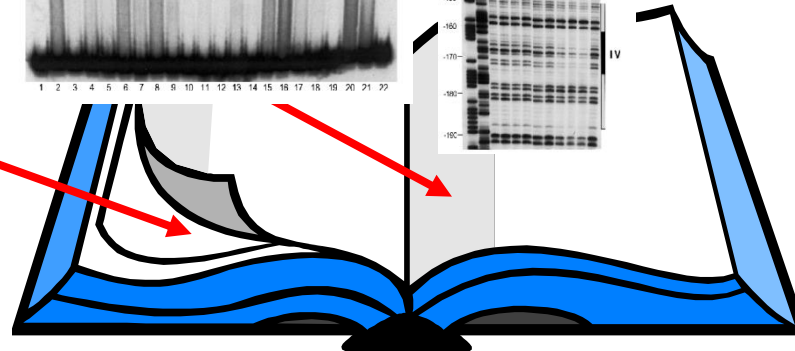
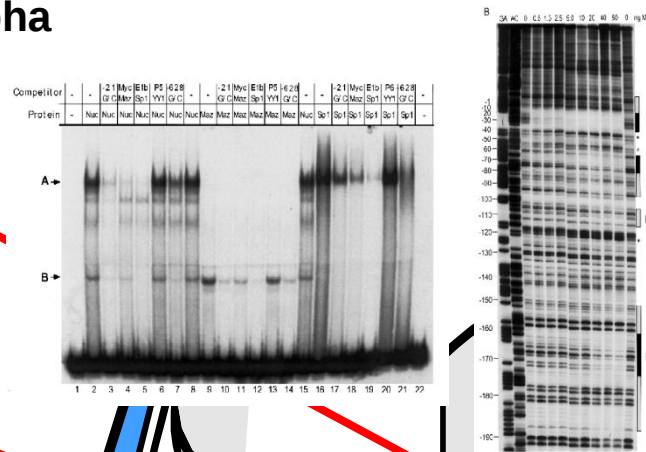
TS mouse

NF [TRANSFAC link: 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

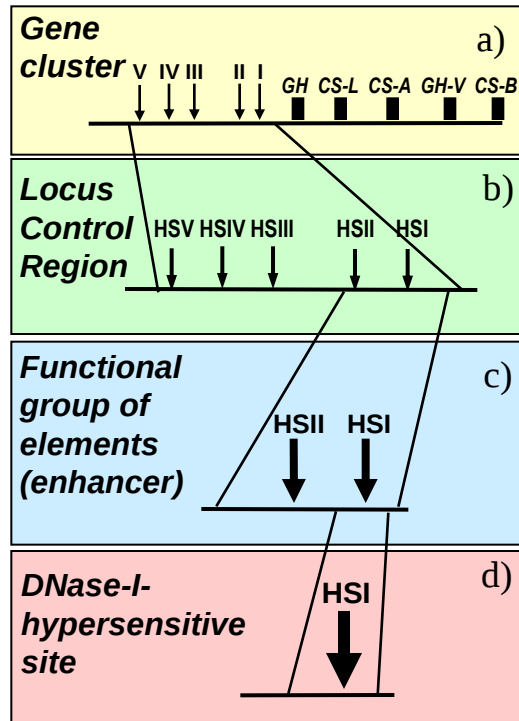
EL LADA:HSSII; LADA:HSSIII; LADA:EFS

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; Catarrh
LO 17q22-q24
LM LGH:HSV;~2000; LGH:HSIV;~3000;~
LGH:HSIII;~11100; LGH:HSII;~800;
LGH:HSI;~14600; ST GH1 gene;;CS
CS2
XX
GI GH1: SOMATOTROPIN (GROWT

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

AR R14
RF
RT pituitary specific
SV

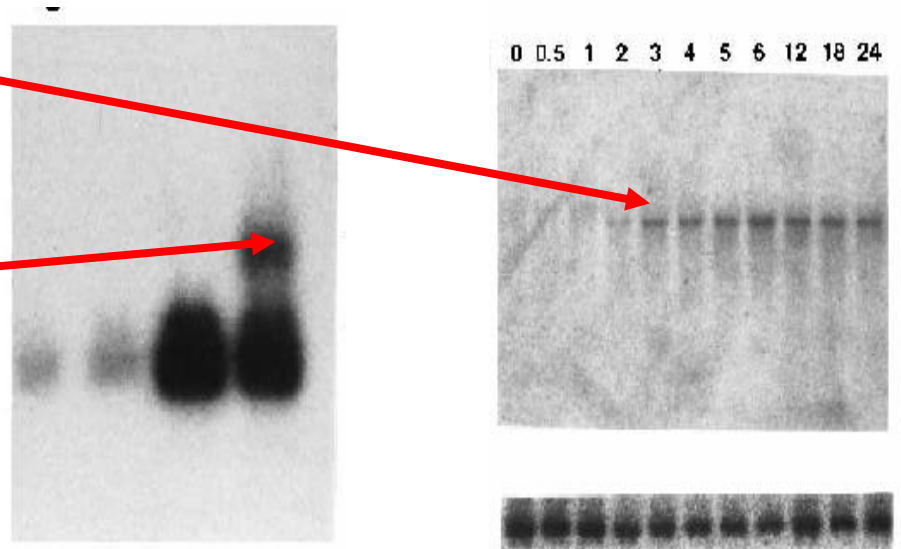
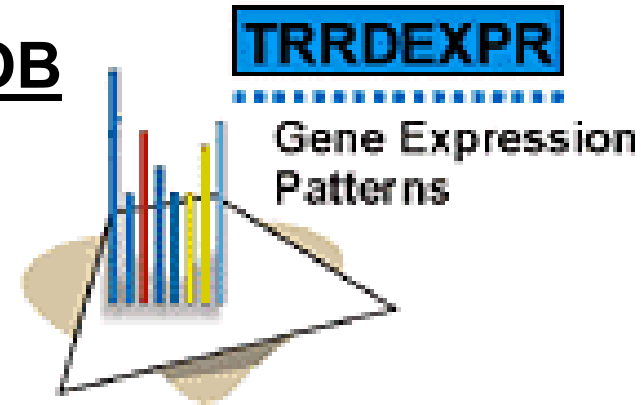
HI LGH:HSV
DE HSV
HT pituitary specific
DP ST of GH-1 gene
PT at -32.5kb
XX
PA positive
OA pituitary
TA GH-secreting adenoma
XX
CA primary pituitary cells
XX
AG Stable Transfection
AG transgenic analysis [Jones B.K. et al., 1995]
AG DNase I hypersensitivity [Jones B.K. et al., 1995]
XX
CA K562
AG Stable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

TRRDEXP

ОСОБЕННОСТИ ЭКСПРЕССИИ ГЕНОВ

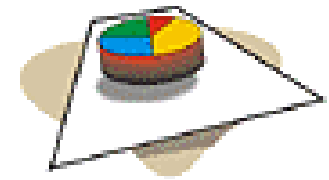
RE A00596.002
ID Hs:AAP
RT mRNA
RN lymphoblastoid cells
RL present
RI heat shock
FF induction
RH 3 h
RR [Abe K. et al., 1991]

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



RNA blot analysis

TRRDIB



TRRDIB

Bibliography

NN 2027

ID [Hs:AAP](#)

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

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

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

SO J.Biol.Chem.

VL 269

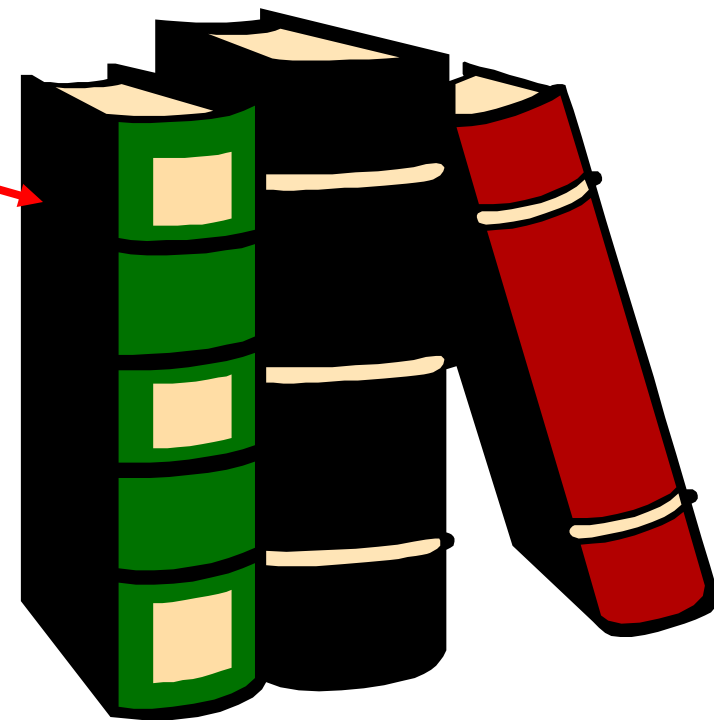
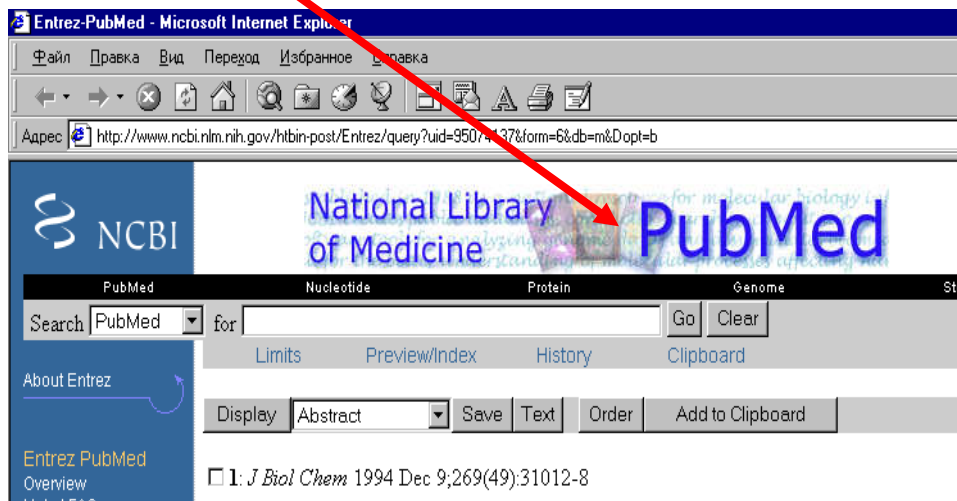
IS 34

YR 1994

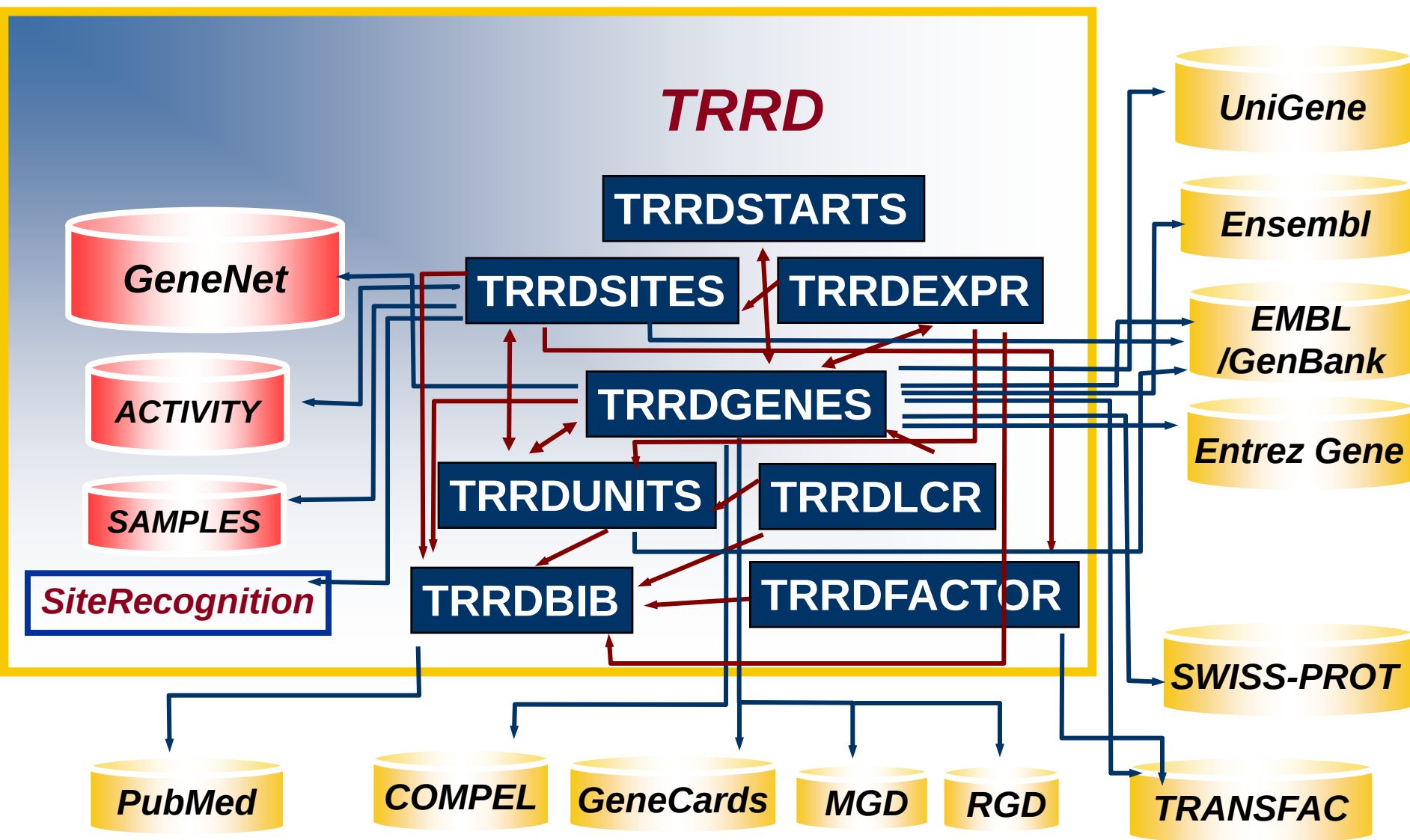
PG 21682-21690

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

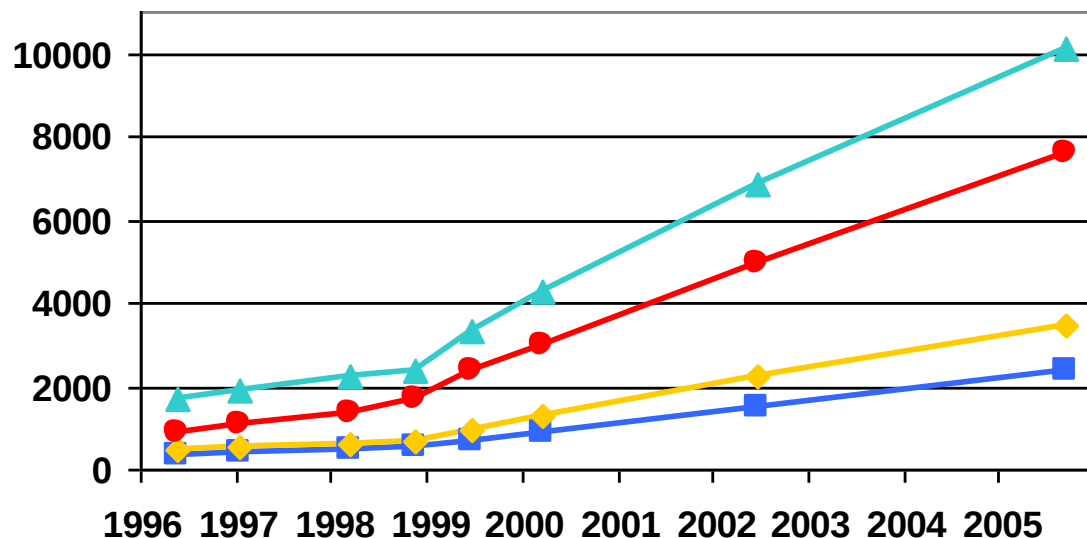
//



Линковка между таблицами TRRD и ссылки на внешние ресурсы.



ДИНАМИКА ПОПОЛНЕНИЯ TRRD

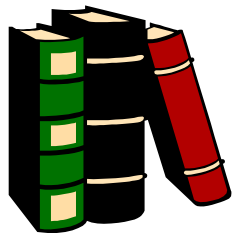


**10135 САЙТОВ СВЯЗЫВАНИЯ
ТРАНСКРИПЦИОННЫХ
ФАКТОРОВ
7609 НАУЧНЫХ ПУБЛИКАЦИЙ**

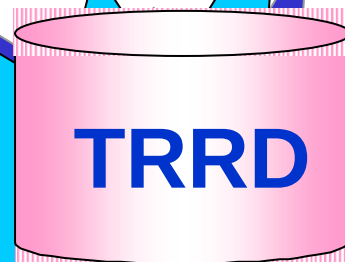
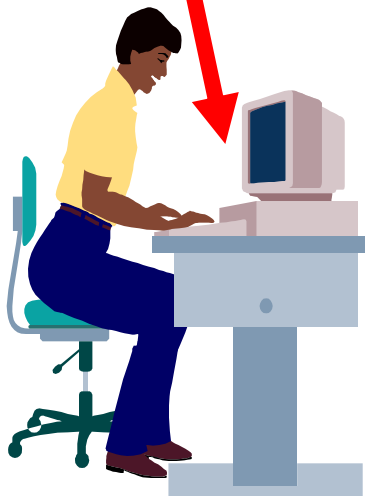
**3400 РЕГУЛЯТОРНЫХ РАЙОНОВ
2344 ГЕНОВ**

	Всего входов в TRRD.	Содержание по видам организмов (%)			
		Человек	Мышь	Крыса	Другие виды
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

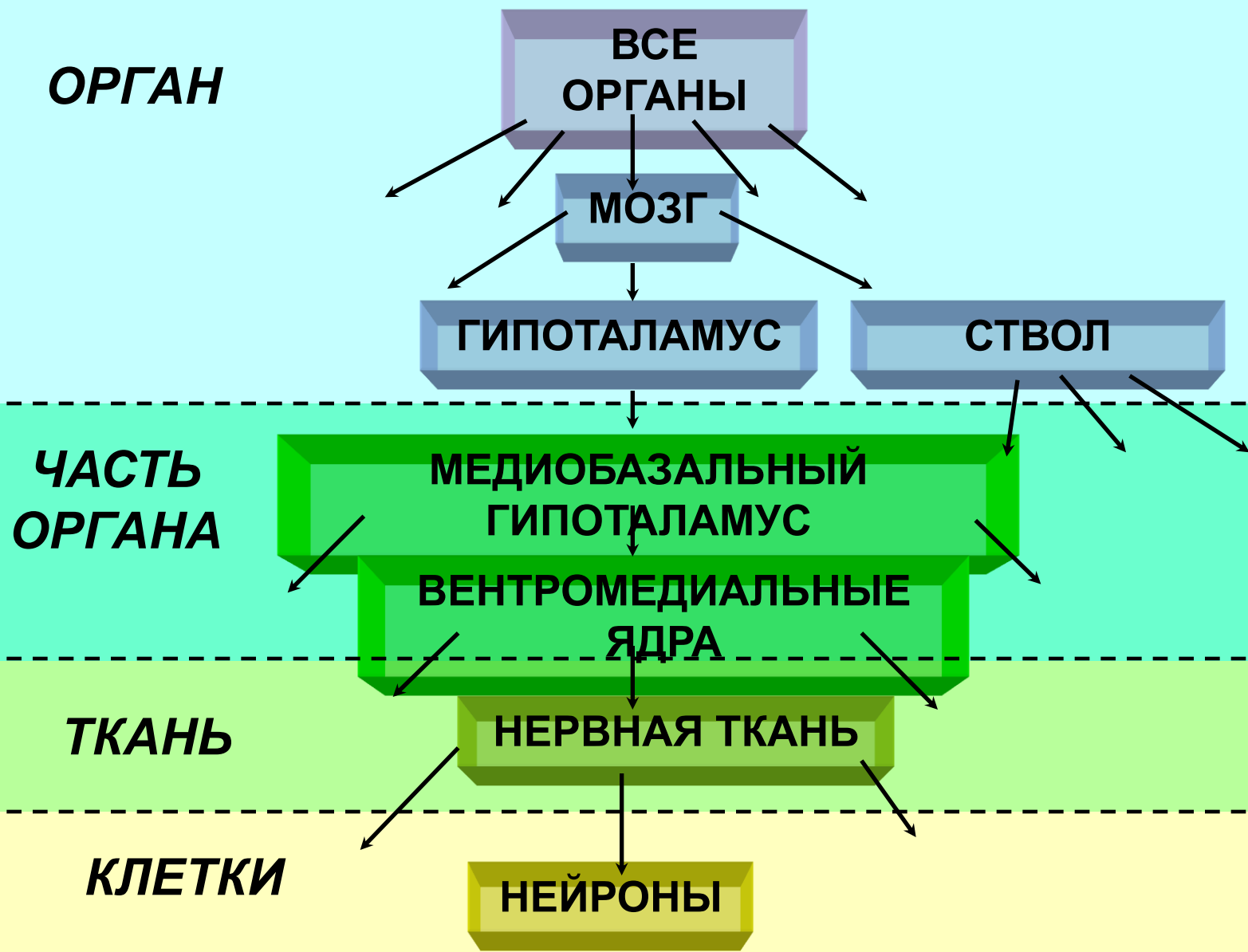


ЭКСПЕРИМЕНТАЛЬНЫЕ
СТАТЬИ



Интернет

ИЕРАРХИЧЕСКАЯ ОРГАНИЗАЦИЯ КОНТРОЛИРУЕМЫХ СЛОВАРЕЙ МОРФОЛОГИЧЕСКИХ ТЕРМИНОВ В TRRD



ПОИСК ДАННЫХ в TRRD: ЗАПРОС НА ОСНОВЕ ТЕЗАУРОСОВ ОРГАНОВ и ТКАНЕЙ МЛЕКОПИТАЮЩИХ

The image shows a composite of three screenshots from the TRRD Gene Express 2.1 website, illustrating a search workflow. Red arrows indicate the sequence of steps: from the main site to the 'Morphology (Mammals)' page, then to the 'KIDNEY' query page, and finally to the search results page.

TRRD Gene Express 2.1

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

TRRD is a unique information resource, accumulating information on the organization of transcription regulatory regions of eukaryotic genomes. Information is included into TRRD.

ACCESS to TRRD:

- SRS ACCESS
- TRRDGENES
- TRRDEXP
- TRRDLCL
- Browse the TRRD
- TRRD sections (genes within functions system)

General information

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

User's guide

- Database schema
- How to search TRRD?
- Integration with other databases
- TRRD Viewer
- FAQ
- What's new?
- How is TRRD updated?
- Standardization of information input
- TRRD progress (from 1996)
- Current TRRD release

TOOLS

[Thesaurus on organs and tissues in mammals](#)

Morphology (Mammals)

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

KIDNEY

Query to the TRRD database:
genes expressing in KIDNEY

Select species:

☒ All ☐ Human ☐ Murine

Do Query

Query to the TRRD database
Search for the genes expressed in

KIDNEY

Select species: ☒ All ☐ Human ☐ Murine

Enter a name of an organ, tissue, cell, or stage

side of the spinal column in

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

TOP PAGE QUERY RESULTS SESSIONS VIEWS DATABASE

Reset

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

Perform operation

☒ on all but selected ☐ on selected

Link

Save

☐ [TRRDGENES4:A00374](#)

Species

human, Homo sapiens

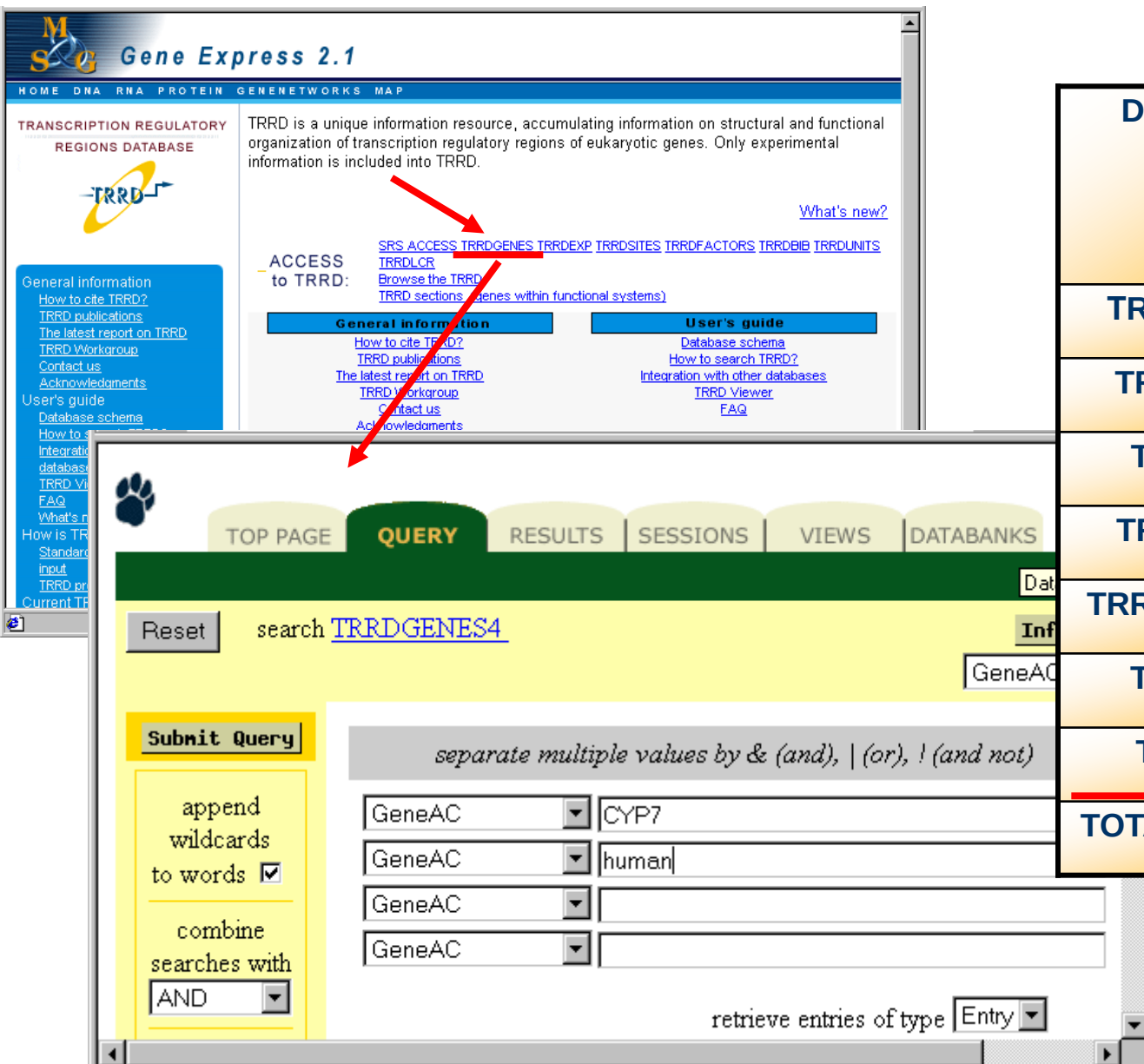
GeneName Brief

ADH3

GeneName Full

alcohol dehydrogenase gene 3, class I

ПОИСК ДАННЫХ В TRRD: SRS



Gene Express 2.1

HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

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

[What's new?](#)

ACCESS to TRRD: [SRS ACCESS TRRDGENES](#) [TRRDEXP](#) [TRRDSITES](#) [TRRDFACTORS](#) [TRRDBIB](#) [TRRDUNITS](#) [TRRDLCR](#) [Browse the TRRD](#) [TRRD sections](#) [genes within functional systems](#)

General information

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

User's guide

Database schema
How to search TRRD?
Integration with other databases
TRRD Viewer
FAQ

QUERY RESULTS SESSIONS VIEWS DATABANKS

Reset search TRRDGENES4

Submit Query

append wildcards to words ☒

combine searches with **AND**

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

GeneAC CYP7

GeneAC human

GeneAC

GeneAC

retrieve entries of type **Entry**

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

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

The screenshot shows the main search window of the TRRDGENES SRS system. At the top, there is a navigation bar with tabs: TOP PAGE, QUERY (highlighted), RESULTS, SESSIONS, VIEWS, DATABANKS, and a HELP button. Below the navigation bar, there is a search bar with the text "search TRRDGENES4" and a "Reset" button. To the right of the search bar is an "Info" button and a dropdown menu labeled "about field" with "GeneAC" selected. Below the search bar, there is a "Submit Query" button highlighted with a red arrow. To the left of the "Submit Query" button, there are several options: "append wildcards to words" with a checked checkbox, "combine searches with" with a dropdown menu set to "AND", and "Number of entries to display per page" with a dropdown menu set to "30". Below these options is an "Extended" button. In the center, there is a section titled "separate multiple values by & (and), | (or), ! (and not)". Below this title, there are four input fields: "GeneName_Full" with the value "apolipoprotein", "Species" with the value "human", "GeneAC", and another "GeneAC" field. Red arrows point to the "apolipoprotein" and "human" values. Below the input fields, there is a "retrieve" button. At the bottom, there is a section titled "Use predefined view" with a dropdown menu set to "Use predefined view" and a "Create your own view" button. To the right of this section, there is a "Select fields to" dropdown menu with options: "GeneAC", "GeneID", and "Updated".

The screenshot shows the results page of the TRRDGENES SRS system. At the top, there is a navigation bar with tabs: TOP PAGE, QUERY, RESULTS (highlighted), SESSIONS, VIEWS, DATABANKS, and a HELP button. Below the navigation bar, there is a "Reset" button. To the right of the "Reset" button, there is a query summary: "Query '[trrdgenes4-GeneName_Full: apolipoprotein*] & [trrdgenes4-Species: human*]' found 9 entries". Below the query summary, there is a section titled "Perform operation" with two radio buttons: "on all but selected" (selected) and "on selected". Below the radio buttons are three buttons: "Link", "Save", and "View". Below the "Perform operation" section, there is a dropdown menu labeled "* Names only *". To the right of the dropdown menu, there is a list of 9 entries, each with a checkbox and a link: [TRRDGENES4:A00150](#), [TRRDGENES4:A00350](#), [TRRDGENES4:A00151](#), [TRRDGENES4:A00149](#), [TRRDGENES4:A00147](#), [TRRDGENES4:A00196](#), [TRRDGENES4:A00264](#), [TRRDGENES4:A00294](#), and [TRRDGENES4:A00148](#).

ПОИСК ДАННЫХ В TRRD: БРАУЗЕРЫ И ТЕМАТИЧЕСКИЕ СЕКЦИИ

Gene Express 2.1
HOME DNA RNA PROTEIN GENENETWORKS MAP

TRANSCRIPTION REGULATORY REGIONS DATABASE

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

ACCESS to TRRD:

- [SRS ACCESS](#)
- [TRRDGENES](#)
- [TRRD EXP](#)
- [TRRDSITES](#)
- [TRRDFACTORS](#)
- [TRRDBIB](#)
- [TRRDUNITS](#)
- [Browse the TRRD](#)
- [TRRD sections \(genes within functional systems\)](#)

General information

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

User's guide

- [Database schema](#)
- [How to search TRRD?](#)
- [Integration with other databases](#)

How is TRRD updated?

- [Standardization of information input](#)

Current TRRD release

TRRD sections

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

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

Browse the TRRD

[Genes by name](#)

[Genes by species](#)

General information

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

User's guide

- [Database schema](#)
- [How to search TRRD?](#)
- [Integration with other databases](#)
- [TRRD Viewer](#)
- [FAQ](#)
- [What's new?](#)
- [How is TRRD updated?](#)
- [Standardization of information input](#)
- [TRRD progress \(from 1996\)](#)
- [Current TRRD release](#)

Тематические секции TRRD

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

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

TRRD – информационная основа для создания выборок

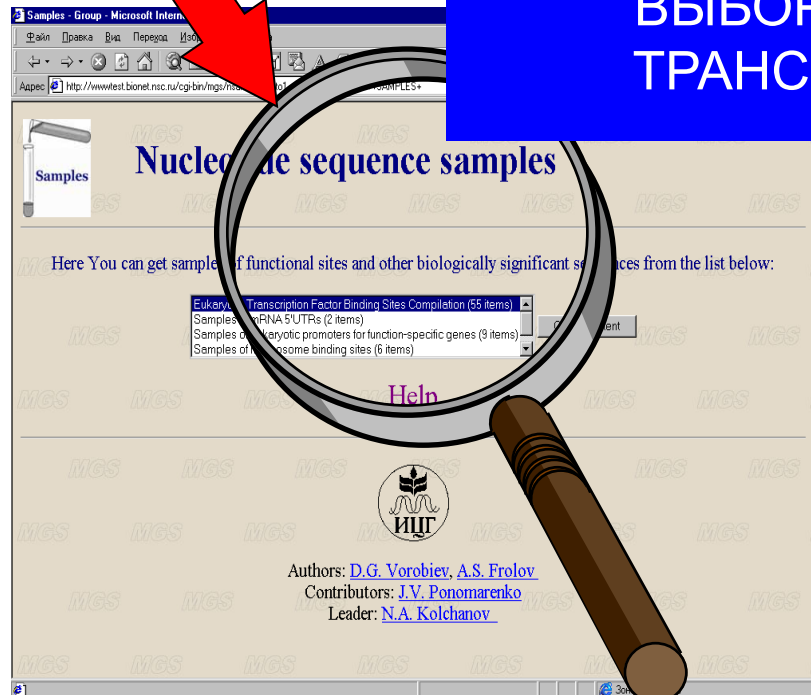
TRRDSITES

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

ВЫБОРКИ ЭУКАРИОТИЧЕСКИХ ПРОМОТОРОВ

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

ВЫБОРКИ САЙТОВ СВЯЗЫВАНИЯ
ТРАНСКРИПЦИОННЫХ ФАКТОРОВ



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

Интернет-ресурсы по транскрипции у растений

ПРОГРАММА

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

Hem Triticum

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

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

Hem Triticum

ECTb Arabidopsis, Rice, Brachypodium

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

Hem Triticum

Есть Arabidopsis thaliana, Oryza sativa, Physcomitrella patens, Poplar

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

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

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 !!!