

Лекция №5

БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ

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

В лекции № 4 была рассмотрена база данных:

EPD Eukaryotic Promoter Database Geneva, Switzerland
- одна из самых старых баз данных, в настоящее время дополнена новой информацией (часть EPDnew)

В лекции № 5 будет дана характеристика баз данных:

TRANSFAC Eukaryotic *cis*-acting regulatory
DNA elements and *trans*-acting factors Germany
(регуляторные элементы и взаимодействующие с ними транскрипционные факторы)

РЕГУЛЯТОРНЫЕ ЭЛЕМЕНТЫ на ДНК

TRED	Transcriptional Regulatory Element Database	Cold Spring Harbor Laboratory USA
ORegAnno	Open REGulatory ANNOtation database	USA
DBTSS	DataBase of Transcriptional Start Sites	Japan
ENCODE	The Encyclopedia of DNA Elements	USA
FANTOM	The Functional Annotation of the Mammalian Genome	Japan + International consortium

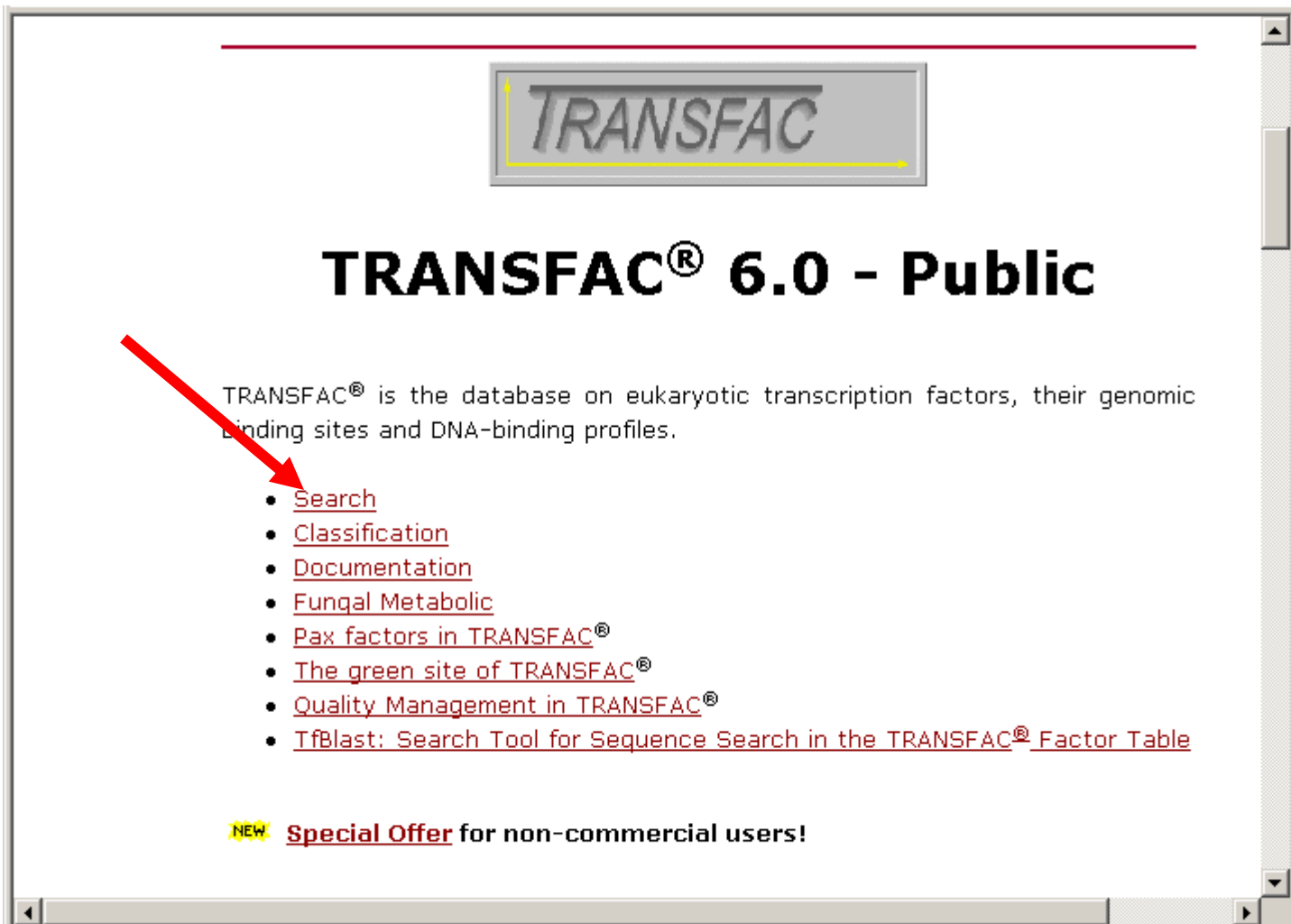
В лекции № 5 будет дана характеристика баз данных
(продолжение):

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

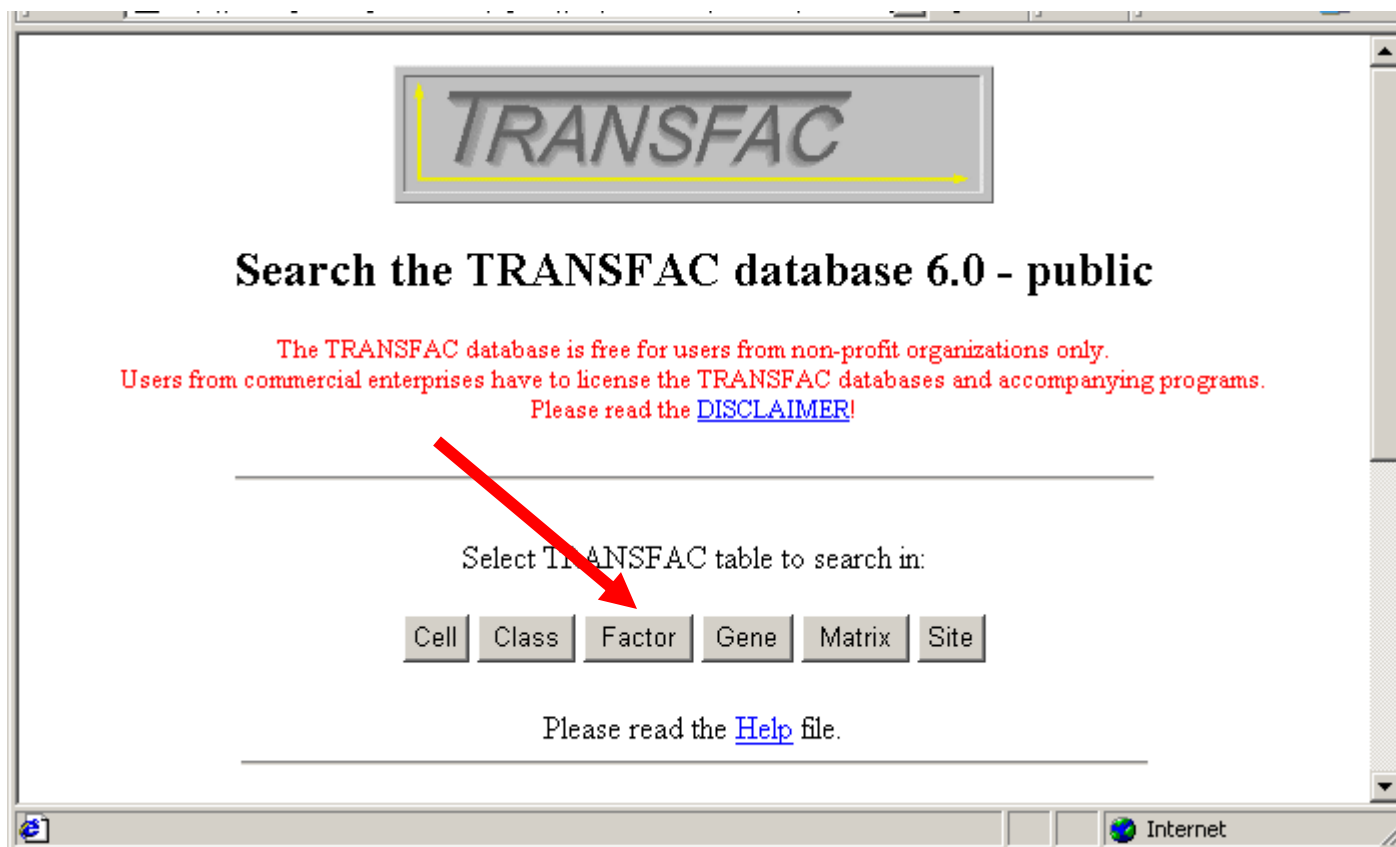
UniProtKB Switzerland	Protein knowledgebase	Geneva,
TFClass	Classification of transcription factors	Germany
AnimalTFDB	Animal Transcription Factor DataBase	Китай

TRANSFAC - информация о транскрипционных факторах и их сайтах связывания на ДНК

<http://www.gene-regulation.com/pub/databases.html#transfac>



Интерфейс для поиска в базе TRANSFAC



ЗАПРОС ПО КЛЮЧЕВОМУ СЛОВУ “HNF4” В БАЗЕ TRANSFAC



Searching in table *Factor*

Search term: [\(Search Hints\)](#)

Case Sensitive Search: ☐ Yes ☒ No

Table field to search in:

Number of hits per page:



The TRANSFAC database is free for for users from non-profit organizations only.
Users from commercial enterprises have to license the TRANSFAC databases and accompanying programs.
Please read the [DISCLAIMER](#) !

TRANSFAC Factor entries - You searched for "***HNF4***" in *Factor Name* - You got **8** entries.

Entry:	Accession No.:	Search Field: <i>Factor Name</i>	Factor Name	Organism Species
1:	T00372	... HNF4 ; HNF-4; hepatocyte ...	HNF-4alpha1	rat, Rattus norvegicus
2:	T02421	... HNF4 ; HNF-4; hepatocyte ...	HNF-4alpha1	human, Homo sapiens
3:	T02424	... HNF4 ; HNF-4; hepatocyte ...	HNF-4alpha1	mouse, Mus musculus
4:	T02429	... HNF4 ; HNF-4; hepatocyte ...	HNF-4alpha1	clawed frog, Xenopus laevis

ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА HNF4 В базе TRANSFAC



TRANSFAC FACTOR TABLE, Release 6.0 - public - 2002-08-01,

AC T00372
XX
ID T00372
XX
DT 26.01.1993 (created); hse
DT 18.04.2000 (updated); rio
CO Copyright (C), Biobase GmbH
XX
FA HNF-4alpha1
XX
SY HNF4; HNF-4; hepatocyte nuclear factor 4; hepatocyte nuclear factor
SY 4alpha1; NR2A1.
XX
OS rat, Rattus norvegicus
OC eukaryota; animalia; metazoa; chordata; vertebrata; tetrapoda; mammalia;
OC eutheria; rodentia; myomorpha; muridae; murinae
XX
CL C0002 CC (rec); 2.1.2.11.1.1.
XX
SZ 455 AA; 50.6 kDa (cDNA), 54 kDa (SDS)
XX
SQ MDMADYSAALDPAYTTLEFENVQVLTMGNDTSPSEGANLNSSNSLGVSAICAICGDRATG
SQ KHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVDKDKRNQCRYCRLKKCFRAGMKKEA
SQ VQNERDRISTRSSYEDSSLPSINALLQAEVLSQQITSPISGINGDIRAKRIASITDVCE
SQ SMKEQLLVLEVWAKYIPAFCELLLDDQVALLRAHAGEHLLLGATKRSMVFKDVLILLGNDY
SQ IVPRHCPELAEMSRVSIIRILDELVLPPFQELQIDDNEYACLKAIFFDPDAKGLSDPGKIK

 Internet

ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА HNF4 в базе TRANSFAC (продолжение)

[SQ](#) LQEMLLGGSASDAPHAHHPLHPHLMQEHMGTNVIVANTMPSHLSNGQMSTPETPQPSPPS

[SQ](#) GSGSESYKLLPGAITTIVKPPSAIPQPTITKQEI

XX

[SC](#) edited SwissProt #P22449

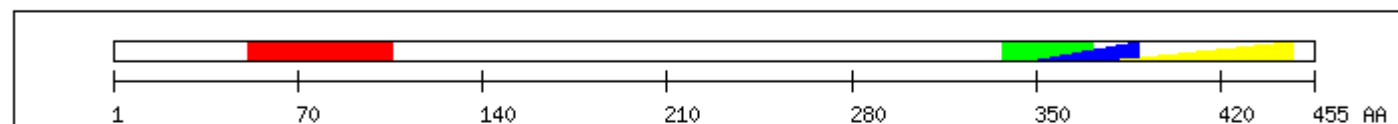
XX

[FT](#) 51 106 zinc finger domain.

[FT](#) 337 371 putative AF-2 region (by homology) [\[2\]](#).

[FT](#) 346 389 essential for trans-activation [\[2\]](#).

[FT](#) 374 447 proline-rich region (14/74).



XX

[CP](#) liver, kidney, intestine

XX

[FF](#) transcriptional activator [\[2\]](#) [\[3\]](#);

[FF](#) may be antagonized by ARP-1 and/or COUP-TF [\[4\]](#);

[FF](#) first step of activation by HNF-4 appears to involve recruitment of TFIIB,

[FF](#) a second step operates through the AF-2-homologous activation domain [\[2\]](#);

[FF](#) can act as transcriptional activator (e. g. the Fabpi promoter) affected by

[FF](#) other upstream elements and transcription factors [\[3\]](#)..

XX

[IN](#) [T02159](#) TFIIB; rat, Rattus norvegicus.

XX

[MX](#) [M00158](#) V\$COUP_01.

[MX](#) [M00134](#) V\$HNF4_01.

[MX](#) [M00411](#) V\$HNF4_01_B.

XX

[BS](#) [R09367](#) AS\$HNF4_01_B_01; Quality: 2.

[BS](#) [R09368](#) AS\$HNF4_01_B_02; Quality: 2.

[BS](#) [R09369](#) AS\$HNF4_01_B_03; Quality: 2.

[BS](#) [R09370](#) AS\$HNF4_01_B_04; Quality: 2.

[BS](#) [R09371](#) AS\$HNF4_01_B_05; Quality: 2.

ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА HNF4 в базе TRANSFAC (окончание)

```
BS R09381 AS$HNF4_01_B_15; Quality: 2.
BS R09382 AS$HNF4_01_B_16; Quality: 2.
BS R09383 AS$HNF4_01_B_17; Quality: 2.
BS R09384 AS$HNF4_01_B_18; Quality: 2.
BS R09385 AS$HNF4_01_B_19; Quality: 2.
BS R02179 HNF4$CONS; Quality: 1.
BS R00114 HS$A1ANTR_01; Quality: 1; alpha1-AT, G000199; human, Homo sapiens.
BS R02657 HS$APOC3_01; Quality: 1; apoCIII, G000206; human, Homo sapiens.
BS R04478 HS$F8_05; Quality: 3; factor VIII, G001026; human, Homo sapiens.
BS R01457 HS$TF_08; Quality: 1; Tf, G000405; human, Homo sapiens.
BS R04770 MOUSE$CRBP2_03; Quality: 2; CRBP2, G001210; mouse, Mus musculus.
BS R03034 MOUSE$TTPA_05; Quality: 6; TTR, G000625; mouse, Mus musculus.
BS R03035 MOUSE$TTPA_06; Quality: 1; TTR, G000625; mouse, Mus musculus.
BS R08877 RAT$FABPI_01; Quality: 1; FABPI, G001738; rat, Rattus norvegicus.
BS R03881 RAT$HNF1_02; Quality: 3; HNF-1, G000756; rat, Rattus norvegicus.
BS R01183 RAT$OTC_05; Quality: 6; OTC, G000781; rat, Rattus norvegicus.
BS R01186 RAT$OTC_08; Quality: 6; OTC, G000781; rat, Rattus norvegicus.
XX
DR TRANSPATH: M0000024884.
DR EMBL: X57133; RNHNF4.
DR PIR: A36471; A36471.
DR TRANSCOMPEL: C00122.
DR TRANSCOMPEL: C00123.
DR TRANSCOMPEL: C00283.
XX
RN [1]
RX MEDLINE: 98256442.
RA Fraser J. D., Martinez V., Straney R., Briggs M. R.
RT DNA binding and transcription activation specificity of hepatocyte nuclear
RT factor 4.
RL Nucleic Acids Res. 26:2702-2707 (1998).
RN [2]
RX MEDLINE: 96239539.
RA Malik S., Karathanasis S. K.
RT TFIIB-directed transcriptional activation by the Orphan nuclear receptor
RT hepatocyte nuclear factor 4.
```

Ссылки на искусственные
последовательности

Ссылки на сайты
связывания в генах
эукариот

ОПИСАНИЕ САЙТА СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННОГО ФАКТОРА HNF4 в базе TRANSFAC

TRANSFAC SITE TABLE, Release 6.0 - public - 2002-08-01, (C) Biobas	
AC	R03881
XX	
ID	RAT\$HNF1_02
XX	
DT	22.10.1994 (created); ewi.
DT	18.04.2000 (updated); rio.
CO	Copyright (C), Biobase GmbH.
XX	
TY	D
XX	
DE	HNF-1 (hepatocyte nuclear factor 1); G000756 .
XX	
SQ	GGCTGAAAGTCCAAAAGTTCAGTC.
XX	
EL	TRH
XX	
SF	-69
ST	-48
XX	
BF	T02758 HNF-4; Quality: 4; Species: human, Homo sapiens.
BF	T00372 HNF-4alpha1; Quality: 3; Species: rat, Rattus norvegicus.
BF	T02429 HNF-4alpha1; Quality: 3; Species: clawed frog, Xenopus laevis.
BF	T02422 HNF-4alpha2; Quality: 3; Species: rat, Rattus norvegicus.
XX	
OS	rat, Rattus norvegicus
OC	eukaryota; animalia; metazoa; chordata; vertebrata; tetrapoda; mammalia;
OC	eutheria; rodentia; myomorpha; muridae; murinae
XX	
SO	0104 HepG2
SO	0289 rl
SO	0399 kidney
XX	
MM	DNase I footprinting
XX	

Идентификатор гена

Название гена

Позиции сайта в промоторе гена

Транскрипционные факторы, взаимодействующие с сайтом

Вид организма

Названия клеток, использованных в экспериментах

База S/MARt DB (часть TRANSFAC)

<http://www.gene-regulation.com/pub/databases/smartdb/toc.html>



S/MARt DB™ 1.1 - Public

S/MARt DB™ collects information about scaffold/matrix attached regions and the nuclear matrix proteins that are supposed be involved in the interaction of these elements with the nuclear matrix. It covers the whole range from yeast to human.

- [Search](#)
- [Browse](#)
- [Documentation](#)
- [Information](#)
- [SbBlast: Search Tool for Sequence Search in the S/MARt Binder Database](#)

Список входов в базе S/MARt DB

S/MARt DB – S/MAR table

Accession number	Name	Species
SM0000001	MOUSE\$kappa-MAR	mouse, Mus spec.
SM0000002	HS\$IFNB1-E	human, Homo sapiens
SM0000003	HS\$IFNA2-SMAR2	human, Homo sapiens
SM0000005	HS\$IFNB1-G	human, Homo sapiens
SM0000006	HS\$IFNB1-K	human, Homo sapiens
SM0000008	HS\$IFNB1-I	human, Homo sapiens
SM0000009	HS\$IFNB1-H	human, Homo sapiens
SM0000010	HS\$IFNB1-D	human, Homo sapiens
SM0000011	HS\$IFNA10-IFNA7(1)	human, Homo sapiens
SM0000012	HS\$IFNA10-IFNA7(2)	human, Homo sapiens
SM0000013	MOUSE\$INT11	mouse, Mus spec.
SM0000014	MOUSE\$INT14	mouse, Mus spec.
SM0000015	MOUSE\$INT19	mouse, Mus spec.
SM0000016	MOUSE\$INT20	mouse, Mus spec.
SM0000017	MOUSE\$INT24	mouse, Mus spec.
SM0000018	MOUSE\$INT25	mouse, Mus spec.
SM0000019	MOUSE\$INT26	mouse, Mus spec.
SM0000020	MOUSE\$INT28	mouse, Mus spec.
SM0000021	HS\$MOA11	human, Homo sapiens
SM0000022	HS\$MOB1	human, Homo sapiens
SM0000023	HS\$MOB2	human, Homo sapiens

Пример записи в базе S/MARt DB

S/MARt DB – S/MAR

```
AC SM00000001
XX
DT 1.1.99 00:00:00 (created); ili
DT 13.12.99 16:08:27 (updated); ili
XX
NA MOUSE$kappa-MAR
XX
OS mouse, Mus spec.
OC eukaryota; animalia; metazoa; chordata; vertebrata;
OC tetrapoda; mammalia; eutheria; rodentia; myomorpha; muridae;
OC murinae
XX
HO human, rabbit [1]
XX
SZ 368 bp
XX
DE G000538; immunoglobulin kappa light chain
DP Direction: 3'; Pos 1: ATG
DN Internal: y;
DC between joining and constant regions [2]; ~200 bp
DC upstream of the kappa enhancer [2]
XX
SQ AGCTTAATGTATATAATCTTTTAGAGGTAAAAATCTACAGCCAGCAAAAAGTCATGGTAAAT
SQ ATTCTTTGACTGAACTCTCACTAAACTCCTCTAAATTATATGTCATATTAACCTGGTTAAA
SQ TTAATATAAAATTTGTGACATGACCTTAACTGGTTAGGTAGGATATTTTCTTCATGCAAAA
SQ AATATGACTAATAATAATTTAGCACAAAAATATTTCCCAATACTTTAATTCTGTGATAGA
SQ AAAATGTTTAACTCAGCTACTATAATCCCATAAATTTTGAAAACTATTTATTAGCTTTTGT
SQ GTTTGACCCTTCCCTGCCAAAGGCAACTATTTAAGGACCCTTTAAAACTCTTGAAACTAC
SQ TTTAGAGT
SC [J. Bode, direct submission]
XX
```

Идентификатор участка S/MAR

Вид организма

Длина участка S/MAR

Идентификатор
близкорасположенного
гена

Нуклеотидная
последовательность
участка S/MAR

http://www.gene-regulation.com/cgi-bin/pub/databases/transcompel/search.cgi



Search the TRANSCompel database 7.0 - public

The data provided in TRANSFAC and TRANSCompel public is only a snapshot from 2005.

For a modest academic/non-profit price, subscription to TRANSFAC Professional provides full access to the most comprehensive, up-to-date content available, as well as more advanced tools and an easy-to-use interface. [Learn more.](#)

Select TRANSCompel table to search in:

Compel

Evidence

TRANSCompel database search engine by Biobase GmbH - Braunschweig 1997-2008

TRANSFAC® Public 版は、啓蒙活動用で、研究用ではありません。有償版 TRANSFAC® Professional には、多くの機能が搭載されており、PWMs に代表されるデータ量は、Public 版の約3倍です。最新のデータを利用できる TRANSFAC® Professional を是非ご検討ください。

両者の比較は [こちら](#)

no: ... mpromdb - ... database of ... mpromdbHg... http://...n.de/ Gene Regul... Search...



Searching in table *Compel*

Search term: ([Search Hints](#))

Case Sensitive Search: ☐ Yes ☒ No

Table field to search in:

Number of hits per page:

Select other TRANSCompel table:

TRANSCompel database search engine by Bio

TRANSCompel Compel entries - You searched for "GATA*" in *Factor name* - You got 7 entries.

Entry:	Accession No.:	Search Field: <i>Factor name</i>
1:	C00026	...; -156 to -151; GATA-2 or GATA-3...
2:	C00088	...156 to -151; GATA-2 or GATA-3. ; -142 to...
3:	C00147	...136 to -131; GATA-2. ; -108 to -102; c-Jun...
4:	C00191	...82 to -62; GATA-4. ; -54 to -41; c-Jun ...
5:	C00220	...SF-1.; -73 to -68; GATA-4...
6:	C00282	...95 to -90; GATA-5. ; -86 to -74; HNF-...
7:	C00292	...85 to -80; GATA-4. ; 36 to 48; HNF-...

TRANSCOMPEL COMPEL TABLE, Release 7.0 - public - 2005-09-30, (C) Biobase

AC C00292
XX
ID C4\$GATA_002
XX
DT 11.09.2001 (created); oke.
DT 28.02.2002 (updated); oke.
CO Copyright (C), Biobase GmbH.
XX
GE [G002684](#); MOUSE\$FTF; Fetoprotein Transcription Factor; mouse, Mus musculus.
XX
SQ TTATCAaccgg...gaaaaAGTGCAGAGTCCA
XX
PS -85 to 48
XX
DR EMBL: [AF239709](#); AF239709; 629.
XX
BS -85 to -80; GATA-4.
BS 36 to 48; HNF-4alpha.
XX
TY synergism
XX
CL tissue-restricted/tissue-restricted.
XX
EV [ev00902](#)
EX Transient co-transfections
CN Functional synergism between factors
CT 0103; Hep3B.
XX
EV [ev00903](#)
EX Site-directed mutagenesis and study of promoter activity
CN Functional synergism between sites
CT 0103; Hep3B.
XX
RN [1]
RX MEDLINE; 21201157.
RA Pare J. F., Roy S., Galarneau L., Belanger L.

GATA-4

HNF-4a

Классификация транскрипционных факторов в базе TRANSFAC

<http://transfac.gbf.de/TRANSFAC/cl/cl.html>

Transcription Factor Classification Last modified 17.02.1999

1 *Superclass*: Basic Domains

1.1 *Class*: [Leucine zipper factors \(bZIP\)](#)

1.1.1 *Family*: AP-1(-like) components

1.1.1.1 *Subfamily*: Jun

1.1.1.1.1 [XBP-1](#)

1.1.1.1.2 [v-Jun](#)

1.1.1.1.3 [c-Jun](#)

1.1.1.1.4 [JunB](#)

1.1.1.1.5 [JunD](#)

1.1.1.1.6 dJRA

1.1.1.2 *Subfamily*: Fos

1.1.1.2.1 [v-Fos](#)

1.1.1.2.2 [c-Fos](#)

1.1.1.2.3 [FosB](#)

1.1.1.2.3.1 FosB1

1.1.1.2.3.2 FosB2

1.1.1.2.4 [Fra-1](#)

1.1.1.2.5 [Fra-2](#)

1.1.1.2.6 [dFRA](#)

1.1.1.2.7 [LRF-1](#)

1.1.1.3 *Subfamily*: Maf

1.1.1.3.1 [v-Maf](#)

TRED

<http://rulai.cshl.edu/TRED>



Transcriptional Regulatory Element Database

HOME

Database Statistics

Browse Database

Retrieve Promoters

Search TF Target Genes

Retrieve TF Motifs

Gene Regulatory Networks

Search Orthologs

Sequence Analysis Tools

- Regular Expression Search
- Matrix Search
- Motif Finding (DWE)
- PromoterWise Alignment
- Palindrome search

Useful Links

MEME

Introduction

In order to understand gene regulation, accurate and comprehensive knowledge of transcriptional regulatory elements is essential. Transcriptional Regulatory Element Database (TRED) has been built in response to increasing needs of an integrated repository for both cis- and trans- regulatory elements in mammals, and the lack of such resources at present.

Genome-wide human, mouse and rat promoter annotation in TRED was realized by an automated pipeline to extract known promoters from databases such as Genbank, EPD and DBTSS, and employ promoter finding program FirstEF combined with mRNA/EST information and cross-species comparisons. We have also carried out hand curation to assess computational prediction and ensure data accuracy. A quality level is assigned to each promoter based on the reliability of the supporting evidence.

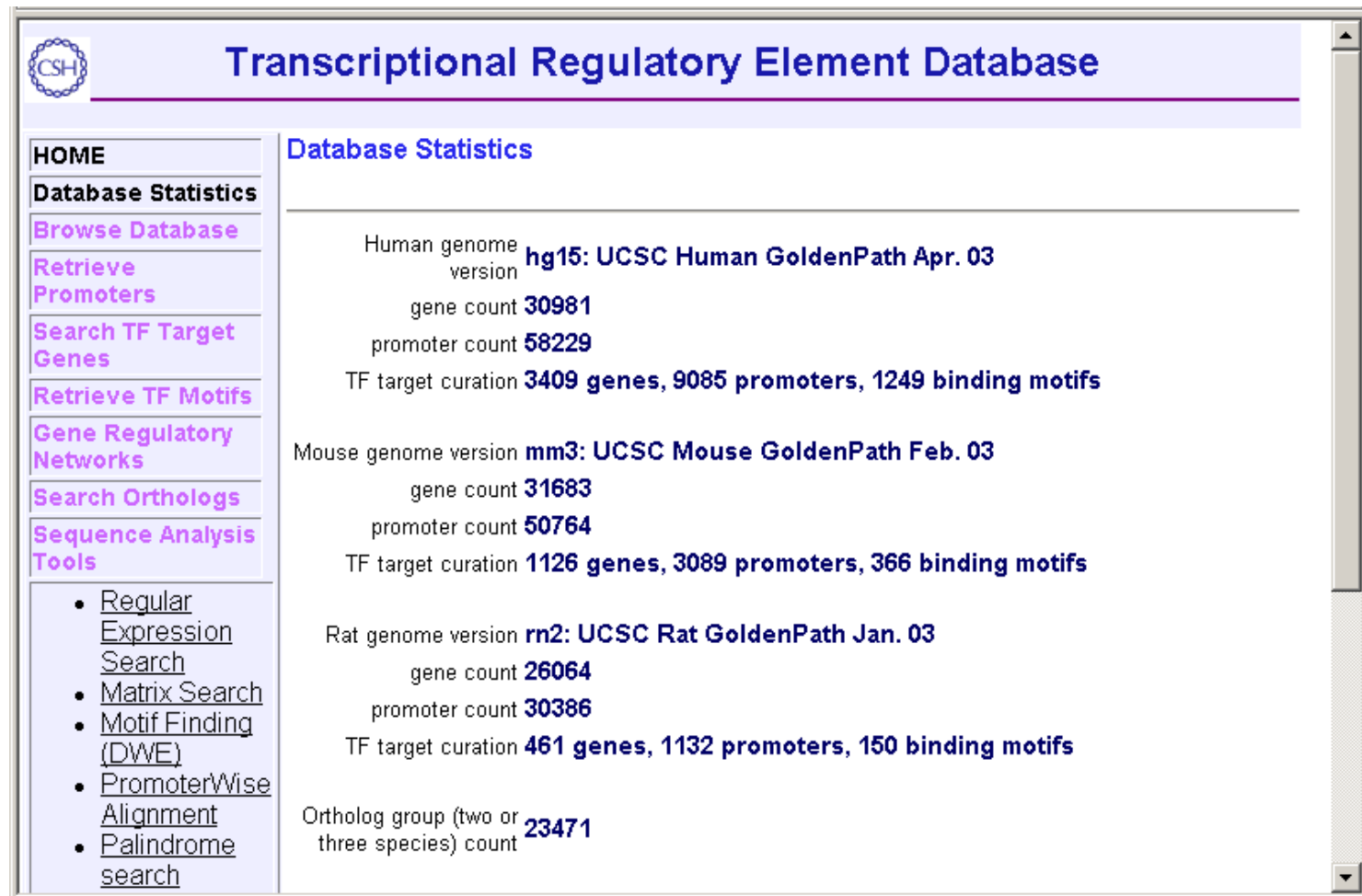
Curation has also been done for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. Binding motifs are mapped on promoters of the corresponding genes and binding quality levels are assigned based on definitiveness of the binding evidence. Curation is currently focusing on target genes of 36 cancer-related TF families.

Distinguishing features of TRED include:

- relatively complete genome wide promoter annotation for human, mouse and rat
- availability of transcription factor binding and regulation information
- data accuracy is ensured by curation, which continues to expand
- easy and flexible data retrieval
- availability of on-the-fly sequence analysis tools

Разработана в лаборатории «Michael Zhang Lab, Cold Spring Harbor Laboratory»

Информационное содержание базы TRED



Transcriptional Regulatory Element Database

Database Statistics

Human genome version **hg15: UCSC Human GoldenPath Apr. 03**
gene count **30981**
promoter count **58229**
TF target curation **3409 genes, 9085 promoters, 1249 binding motifs**

Mouse genome version **mm3: UCSC Mouse GoldenPath Feb. 03**
gene count **31683**
promoter count **50764**
TF target curation **1126 genes, 3089 promoters, 366 binding motifs**

Rat genome version **rn2: UCSC Rat GoldenPath Jan. 03**
gene count **26064**
promoter count **30386**
TF target curation **461 genes, 1132 promoters, 150 binding motifs**

Ortholog group (two or three species) count **23471**

- [Regular Expression Search](#)
- [Matrix Search](#)
- [Motif Finding \(DWE\)](#)
- [PromoterWise Alignment](#)
- [Palindrome search](#)

TRED содержит данные о промоторах, выявленных различными методами в полных геномах человека, мыши, крысы. Разработчики TRED сфокусировали свое внимание на сборе данных о сайтах связывания транскрипционных факторов, вовлеченных в развитие опухолевых процессов. Для решения поставленной задачи, в TRED занесены **данные о сайтах связывания 36 семейств транскрипционных факторов**, полный список которых доступен для просмотра.

Реализация запроса по базе TRED: поиск промоторов, содержащих сайты связывания E2F-типа

Transcriptional Regulatory Element Database

HOME
Database Statistics
Browse Database
Retrieve Promoters
Search TF Target Genes
Retrieve TF Motifs
Gene Regulatory Networks
Search Orthologs
Sequence Analysis Tools

- Regular Expression Search
- Matrix Search
- Motif Finding (DWE)
- PromoterWise Alignment
- Palindrome search

Transcription Factor Target Gene Page

Type of search key:
Factor Name

Search terms:(separated by commas)
(e.g. E2F,E2F-4,myc / T01546,T00140)
* Currently 36 cancer-related TF families are available. [Click for the complete TF list.](#)

E2F

Target Gene Organism: Human

Promoter Quality:
--All--
1: known, curated
2: known
3.1: refseq, predicted

Binding Quality:
--All--
known
likely
maybe

SEARCH RESET

Так как база содержит данные о ССТФ и промоторах, полученных различными способами (ручная аннотация и предсказанные компьютерно программой), в каждом входе базы имеется указание на способ получения данных

Результат поиска промоторов, содержащих сайты связывания E2F-типа в базе TRED

E2F-4 E2F-1 E2F E2F-2 E2F-3 E2F-6 target genes in human,

	Gene	Species	Map Location	Best Promoter	Best Promoter Quality	Best Binding Quality
☐	SYNGR1	human, Homo sapiens	22q13.1	28332	3.1: refseq,predicted	predicted
☐	PCNA	human, Homo sapiens	20pter-12	27147	1: known,curated	known
☐	RIPK2	human, Homo sapiens	8q21	40094	2: known	maybe
☐	OGFR	human, Homo sapiens	20q13.3	26502	2: known	predicted
☐	POLE2	human, Homo sapiens	14q21-22	112824	1: known,curated	known
☐	FLJ35487	human, Homo sapiens	1p22.3	3302	3.2: refseq	maybe
☐	n/a	human, Homo sapiens		5838	4: withRNA	maybe
☐	PTPN6	human, Homo sapiens	12p13	8664	1: known,curated	known
☐	n/a	human, Homo sapiens		11668	2: known	maybe
☐	LIPC	human, Homo sapiens	15q21-23	13325	1: known,curated	known
☐	POLR2C	human, Homo sapiens	16q13-21	15265	1: known,curated	maybe
☐	n/a	human, Homo sapiens		17054	4: withRNA	maybe
☐	ZNF540	human, Homo sapiens	19q13.13	20774	2: known	maybe
☐	SLC16A14	human, Homo sapiens	2q37.1	24581	2: known	maybe
☐	TRPM2	human, Homo sapiens	21q22.3	27513	2: known	maybe
☐	FHIT	human, Homo sapiens	3p14.2	30712	2: known	maybe
☐	POU3F2	human, Homo sapiens	6q16	35658	2: known	maybe
☐	n/a	human, Homo sapiens		39702	6: other	maybe
☐	n/a	human, Homo sapiens		42816	6: other	maybe
☐	TRAPPC4	human, Homo sapiens	11q23.3	7146	2: known	maybe

Указание на способ получения данных о промоторе и сайте связывания ТФ

Marked promoter sequence from to relative to transcription start site.

Format:

OR



[HOME](#)

[Database Statistics](#)

[Browse Database](#)

[Retrieve Promoters](#)

[Search TF Target Genes](#)

[Retrieve TF Motifs](#)

[Gene Regulatory Networks](#)

[Search Orthologs](#)

[Sequence Analysis Tools](#)

- [Regular Expression Search](#)
- [Matrix Search](#)
- [Motif Finding \(DWE\)](#)
- [PromoterWise Alignment](#)
- [Palindrome search](#)

[Useful Links](#)

- [MEME](#)
- [FootPrinter](#)
- [Gibbs Sampler](#)
- [VISTA](#)
- [PipMaker](#)
- [LAGAN](#)

Introduction

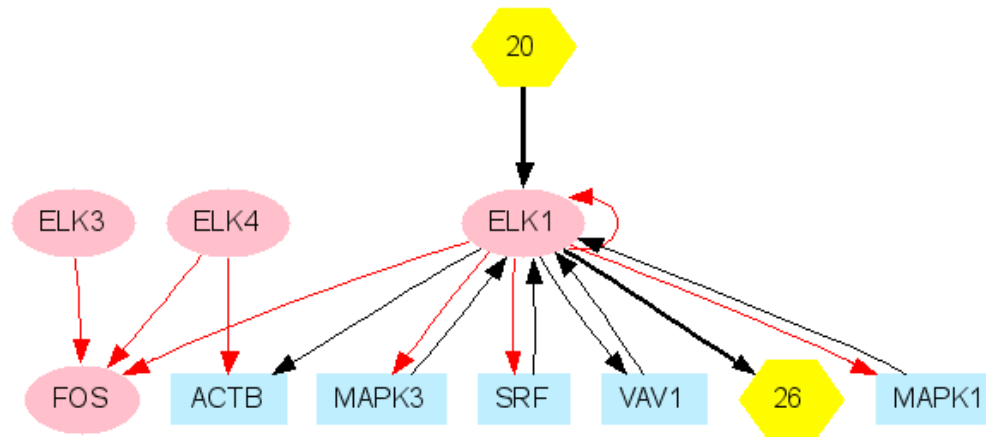
In order to understand gene regulation, accurate and comprehensive knowledge of transcr regulatory elements is essential. Transcriptional Regulatory Element Database (TRED) ha

Gene Regulatory Networks for the 36 TF families in human, mouse, and rat. Click the TF family name.

Family	Full Name	Members (Official Gene Symbols)
AP1	Activator Protein 1	FOS, FOSB, JUN, JUNB, JUND
AP2	Activator Protein 2	TFAP2A, TFAP2B, TFAP2C, TFAP2D, TFAP2E
AR	Androgen Receptor	AR
ATF	Activating Transcription Factor	ATF1 - 7
BCL	B-cell CLL/lymphoma	BCL3, BCL6
BRCA	breast cancer susceptibility protein	BRCA1 - 3
CEBP	CCAAT/enhancer binding protein	CEBPA, CEBPB, CEBPD, CEBPE, CEBPG
CREB	cAMP responsive element binding protein	CREB1 - 5, CREM
E2F	E2F transcription factor	E2F1 - 7
EGR	early growth response protein	EGR1 - 4
ELK	member of ETS oncogene family	ELK1, ELK3, ELK4
ER	Estrogen Receptor	ESR1, ESR2
ERG	ets-related gene	ERG
ETS	ETS-domain transcription factor	ETS1, ETS2, ETV4, SPI1
FLI1	friend leukemia integration site 1	FLI1
GLI	glioma-associated oncogene homolog	GLI1 - 4
HIF	Hypoxia-inducible factor	HIF1A, ARNT, EPAS1, HIF3A

Пример представления регуляторной сети в базе TRED

Gene Regulatory Network of TF family ELK in human



Note:

Ellipses are TFs. Boxes are genes.

Hexagons are the clustered genes. The number of the genes is shown inside.

Red lines indicate the protein-DNA binding is known.

Only official gene symbols are used in the network.

Listing of the gene clusters (TFs are underlined):

20→ELK1: MAPK14, EGF, FGF7, FGF10, RASGRP3, NTRK2, DOK4, MAPK6, MAP2K1, BDK, TNFRSF17, MAPK8, SUMO2, SNCG, SUMO1, WDR26, RIPK2, RAPGEF2, RAPGEF5, GAB2.

ELK1→26: EGR1, EGR2, CCT8, CSH1, DYX1C1, DEFA1, ERBB2, FUT4, SIN3A, KCNMA1, KCNMB1, MCL1, PDGFB, PLAT, TLR9, PRKCI, PRL, PSEN1, ACTA1, ACTA2, SLC6A4, ACTC, ACTG1, TNF, ACTG2, CA9.

В базе TRED имеется возможность визуализации регуляторных сетей для каждого из 36 семейств транскрипционных факторов, объектами которых являются сами факторы, члены их семейств и регулируемые ими гены

The Open REGulatory ANNOtation database (ORegAnno)

<http://py-gi1.stanford.edu:8080/oreganno/>

You are not logged in

ORegAnno v3.0
Open Regulatory Annotation Database

REGULATORY HAPLOTYPE: 7 entries.
OPERON: 4520 entries.
REGULATORY REGION: 39400 entries.
REGULATORY POLYMORPHISM: 185 entries.
miRNA BINDING SITE: 3320 entries.
sRNA BINDING SITE: 53 entries.
TRANSCRIPTION FACTOR BINDING SITE: 1901112 entries.

More details...

Welcome to ORegAnno version 3.0

The Open REGulatory ANNOtation database (ORegAnno) is an open database for the curation of known regulatory elements from scientific literature. Annotation is collected from users worldwide for various biological assays and is automatically cross-referenced against PubMed, Entrez Gene, Ensembl, dbSNP, the eVOC: Cell type ontology, and the Taxonomy database, where appropriate, with information regarding the original experimentation performed (evidence). ORegAnno further provides an open validation process for all regulatory annotation in the public domain. Assigned validators receive notification of new records in the database and are able to cross-reference the citation to ensure record integrity. Validators have the ability to modify any record (deprecating the old record and creating a new one) if an error is found. Further, any contributor to the database can comment on any annotation by marking errors, or adding special reports into function as they see fit. These features of ORegAnno ensure that the collection is of the highest quality and uniquely provides a dynamic view of our changing understanding of gene regulation in the various genomes. As a first step, we recommend reading through our [Help](#) page.

The ORegAnno data and web application are all [LGPL open-source](#) to encourage the development and maintenance of the database to new information and experimentation techniques. Please use our [current citation information](#) when referring to ORegAnno data in publication. We encourage interested contributors to send email to the ORegAnno mailing list at oreganno@bcgsc.ca or to visit the [mailing-list archives](#).

Questions or comments to oreganno@bcgsc.ca. ([Archived here](#))

Washington University in St. Louis

OREG* Wild card searches using * (Note: Boost terms (weight documents, t
Notch^4 RBP Uses AND or OR grouping
9707 AND 10090 Uses AND or OR grouping
20050901 TO 20051231 Use TO to search a range

SEARCH FILTER: ☐ Exclude deprecated records

Search

all
id
stable_id
type
dataset
gene_name
gene_source
gene_id
tf_name
tf_source
tf_id
species
taxon_id
user
date
promoter
reference
pubmed_id
evidence
comment

ORegAnno_Combine...tsv
110/849 MB, Осталось 3...

Все скачанные файлы...

Questions or comments to oreganno@bcgsc.ca

Star MED

+ Есть возможность
скачивания базы в текстовом
формате

Griffith OL1, Montgomery SB, Bernier B, Chu B, Kasaian K, Aerts S, Mahony S, Sleumer MC, Bilenky M, Haeussler M, Griffith M, Gallo SM, Giardine B, Hooghe B, Van Loo P, Blanco E, Ticoll A, Lithwick S, Portales-Casamar E, Donaldson IJ, Robertson G, Wadelius C, De Bleser P, Vlieghe D, Halfon MS, Wasserman W, Hardison R, Bergman CM, Jones SJ; Open Regulatory Annotation Consortium. ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic Acids Res. 2008 Jan;36(Database issue):D107-13. Epub 2007 Nov 15.

DBTSS (<http://dbtss.hgc.jp/>)

The screenshot shows the DBTSS website interface. At the top, the title is "DBTSS" with the subtitle "- DataBase of Transcriptional Start Sites -". The release information is "Release 9.0 Updated (July 9, 2015) Based on UCSC hg38, mm10". The main navigation bar includes a "Top" link. The left sidebar contains two sections: "Database Search" and "Human Chromatin Features". The "Database Search" section has a "Species:" dropdown menu with options: H. sapiens (selected), M. musculus, P. falciparum, C. merolae, R. norvegicus, P. troglodytes, and M. fascicularis. There is a "Genome browser:" dropdown menu with "hg38" selected. Below these are input fields for "chr1:99,950,000-100,050,000" and a "Search" button. The "Human Chromatin Features" section has a "Search from Genomic Position:" input field with "chr1:75,787,000" and a "Search" button. Below that is a "Search from SNP (dbSNP rsID):" input field with "rs375229869". The main content area is titled "About this database" and contains a welcome message, a description of the database, and a list of recent updates. The "News" section on the right lists updates from April 2016 to November 2015, including new mouse data and human data accessions.

Database Search

Species:
H. sapiens
M. musculus
P. falciparum
C. merolae
R. norvegicus
P. troglodytes
M. fascicularis

Genome browser:
hg38

chr1:99,950,000-100,050,000

Search

Human Chromatin Features

Search from Genomic Position:
chr1:75,787,000

Search

Search from SNP (dbSNP rsID):
rs375229869

About this database

Welcome to **DBTSS (DataBase of Transcriptional Start Sites)**.

To support transcriptional regulation studies, we have constructed the DBTSS (DataBase of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS tag sequences for collected from a total of 20 tissues and 7 cell cultures. We also integrated our newly generated RNA-seq data of subcellular- fractionated RNAs and ChIP-seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project.

In this update, we further associated those TSS information with public and original SNV data, in order to identify single nucleotide variations (SNVs) in the regulatory regions.

It is believed that single nucleotide variations (SNVs)

News

- 20 Apr. 2016: New HBZ over-expressed mouse data are now available. Raw data accession: [DRA003229](#) and [DRA003744](#) (Genome Shien).
- 27 Jan. 2016: New CD4Tcell/Bcell/EScell data of mouse are now available. Raw data accession: [GSE73825](#) (Genome Shien).
- 30 Nov. 2015: New Germ Cell Methylation data of mouse are now available. Raw data accession: [DRA000484](#) and [DRA000607](#) (Genome Shien).
- 10 Nov. 2015: New granulocyte macrophage progenitors (GMPs) data of mouse are now available. Raw data accession: [DRA000485](#), [DRA000486](#), [DRA000487](#), [DRA000488](#) and

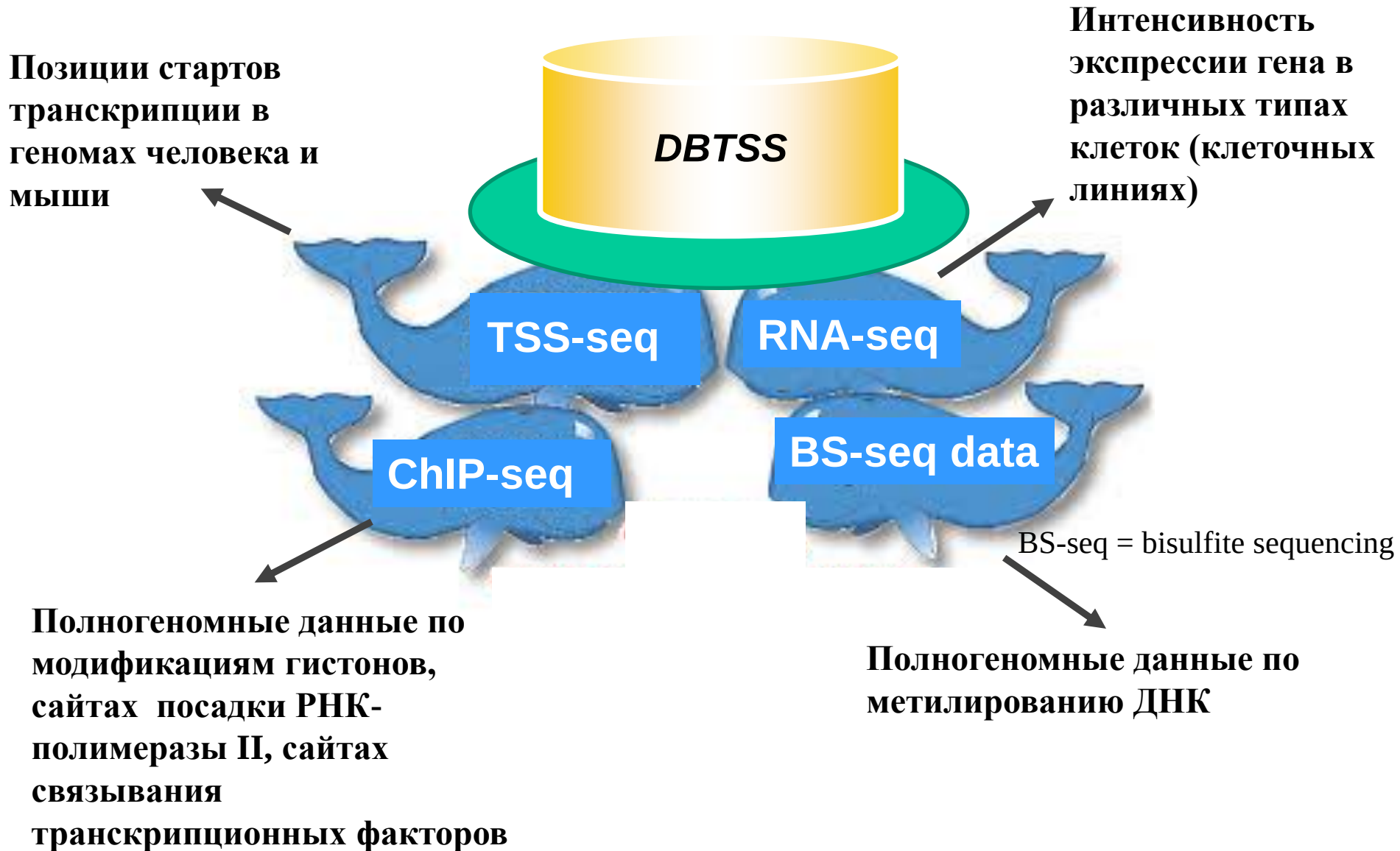
TSSs =Transcriptional Start Sites

Точные позиции TSS в геномах человека, мыши и еще 5 видов организмов

База развивается с 2002г.

Свежий релиз (2015 г.) включил данные по модификациям хроматина и SNVs (single nucleotide variations)

DBTSS (Release 9.0) – экспериментальные методики получения данных



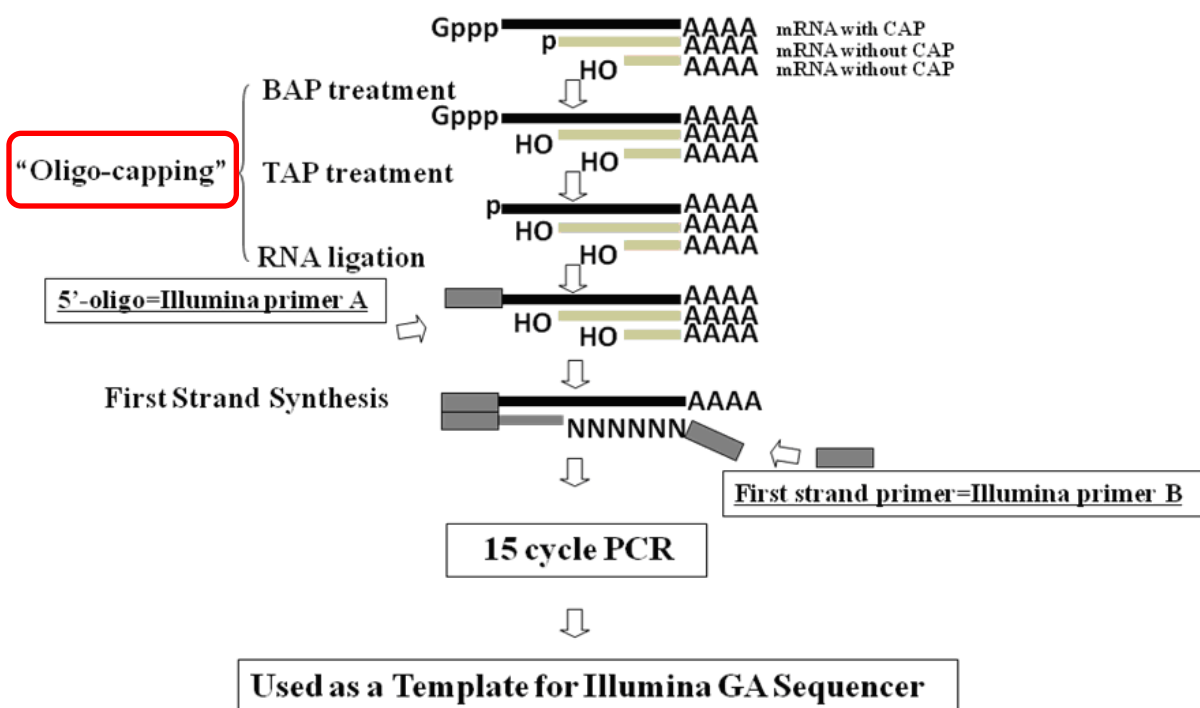
Методика TSS-seq

TSS-seq – оригинальная методика, позволяющая определять позиции стартов транскрипции в геноме . Представляет собой сочетание нескольких подходов:

- Выделение полноразмерной cDNA technology
- Метод oligo-capping (Suzuki and Sugano, Methods in Mol Biol, 2003)
- Секвенирование на платформе Illumina

Метод oligo-capping:

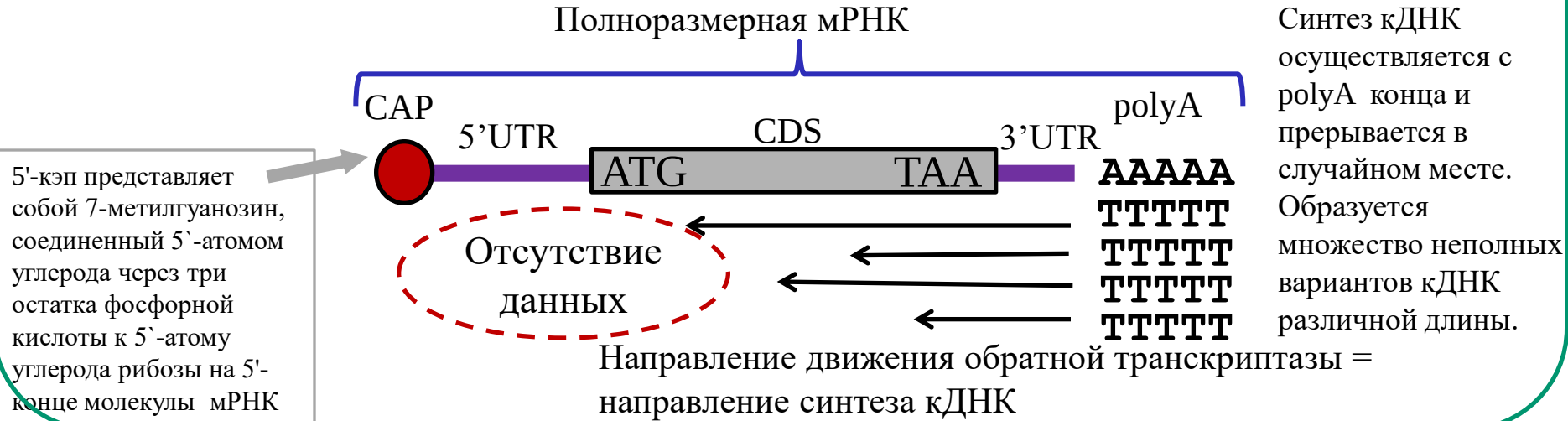
Адапторы (олигонуклеотиды), которые используются в процессе секвенирования на платформе Illumina, пришиваются к кэп-сайту транскриптов. Процесс включает три этапа , на которых используются три разных фермента. Это позволяет пришивать адапторы только к полноразмерным транскриптам (то есть транскриптам, имеющим неповрежденный 5'-конец и имеющим кэп)



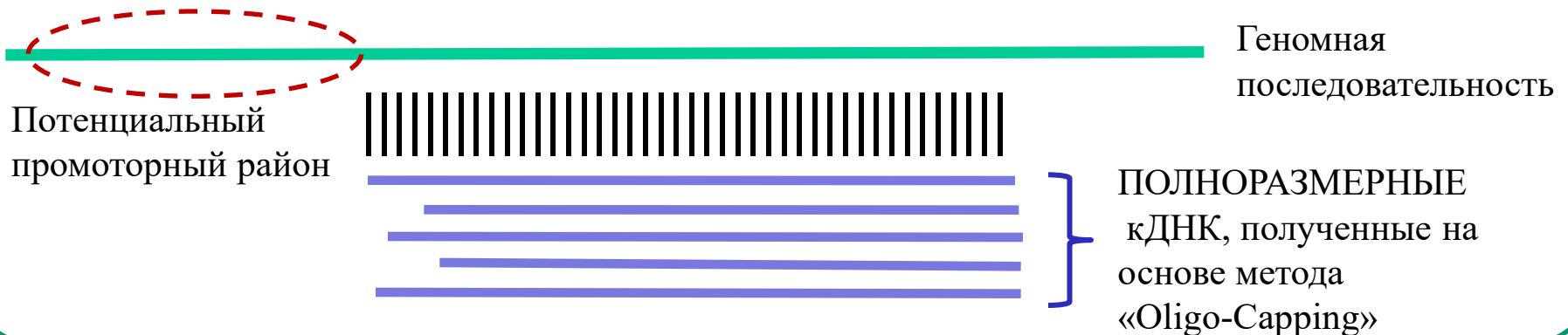
Преимущество «Oligo-Capping method»:

возможность идентификации промоторного района

Традиционные способы получения библиотек кДНК не гарантировали получение полноразмерных продуктов. Участки кДНК, соответствующие 5' – концам транскриптов, заведомо недопредставлены. Точная идентификация старта транскрипции не возможна.



Метод «Oligo-Capping» обеспечивает получение **ПОЛНОРАЗМЕРНЫХ** кДНК. Это позволяет надежно определять позиции стартов транскрипции.



Более подробно: три стадии метода «Oligo-Capping»:

Что делается ?

Мечение 5'-концов мРНК, имеющих 5'-кэп*. При этом 5'-кэп замещается на синтетический олигонуклеотид определенной длины и состава.

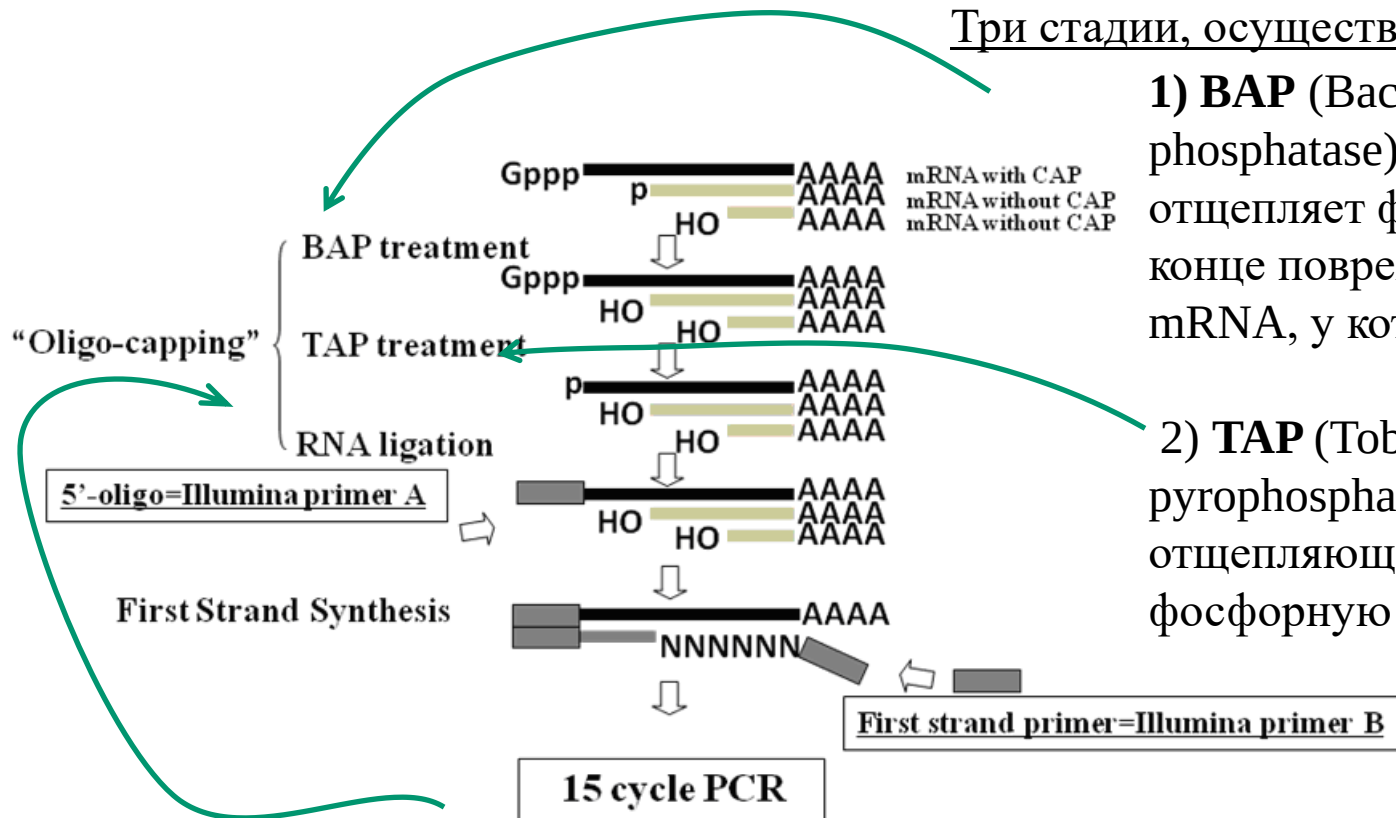
* 5'-кэп представляет собой 7-метилгуанозин, соединенный 5'-атомом углерода через три остатка фосфорной кислоты к 5'-атому углерода рибозы на 5'-конце молекулы мРНК

Три стадии, осуществляемые ферментами:

1) **BAP** (Bacterial alkaline phosphatase) – фосфатаза, которая отщепляет фосфорную группу на конце поврежденной (неполной) mRNA, у которой отсутствует CAP

2) **TAP** (Tobacco acid pyrophosphatase) – пиррофосфатаза, отщепляющая CAP, оставляя фосфорную группу на 5'конце

3) **T4 RNA ligase**, - лигаза, для работы которой необходимо наличие фосфата на 5'конце субстрата, селективно присоединяет синтетический олигонуклеотид к 5'-концу тех продуктов, которые исходно имели CAP



DBTSS (Release 9.0) – информационное содержание

Позиции стартов транскрипции

Интенсивность экспрессии генов

Модификации гистонов, сайты посадки РНК-полимеразы II, сайты связывания транскрипционных факторов

Метилирование ДНК

Типы хроматина

Замены нуклеотидов

Data contents

Number of dataset

Data contents

	Human	Mouse	Malaria	Chyzon	Rat	Chimpanzee	Macaque
TSS-seq	73	7	1	1	1	1	2
RNA-seq	42	36	0	0	0	0	0
ChIP-seq	255	162	0	0	0	0	0
RIP-seq	12	6	0	0	0	0	0
BS-seq	26	16	0	0	0	0	0
ChromHMM	36	0	0	0	0	0	0
SNV	49	0	0	0	0	0	0

[To top](#)

Dataset table

Data contents

Human

		Lung cancer 26 cell-lines	Other cancer cell-lines	Normal cell-lines	Normal tissues	Clinical samples (ICGC, TCGA, etc.)
TSS-seq		↓		↓		-
RNA-seq	total RNA	↓		↓		-
	polysome		↓		-	-
	nucleosome	-		-		-
	cytoplasmic					-
ChIP-seq (transcription factor)	Pol II	↓				-
	HIF1A		↓		-	-
	STAT6	-				-
ChIP-seq (histone modification)	H3K27me3		↓			-
	H3K4me3		↓			-
	H3ac					-
	H3K27ac	↓			-	-
	H3K36me3					-
	H3K4me1		-			-
	H3K9me3					-
Bisulfite sequencing		↓		-		-
SNV		↓		-		↓

Mouse

	3T3	10T1/2	embryo	ATDC5
TSS-seq			↓	

[To top](#)

Геномный браузер базы DBTSS: TSS viewer

Database Search

Species:

Keyword:

Genome browser:

Human Chromatin Features

Search from Genomic Position:

Search from SNP (dbSNP rsID):

Search from SNV (COSMIC: somatic mutation):

Search from SNV-enriched Gene in Cancers:

SNV Summary in Cancers:

Pathway Map

Species:

Documents

- Experimental Procedures

TSS Viewer

Genome Viewer (Multi-omics Viewer)
Chromosome: chr7, Strand: Plus, From: 54,984,552 To: 55,170,743

About this gene

Entrez Gene	Unigene ID	ID of transcript	Product	mRNA Length
EGFR (1956)	Hs.488293	NM_005228	epidermal growth factor receptor isoform a precursor	5600 bases
	Hs.488293	NM_201283	epidermal growth factor receptor isoform c precursor	1572 bases
	Hs.488293	NM_201284	epidermal growth factor receptor isoform d precursor	2865 bases
	Hs.488293	NM_201282	epidermal growth factor receptor isoform b precursor	2239 bases

TSS Seq Data

Tissues ☒ Adult Tissues ☐ Fetal Tissues

Cell Lines ☐ DLD-1 ☐ BEAS-2B ☐ Ramos ☐ MCF7 ☐ HEK293 ☐ TIG3 ☐ HeLa ☐ 25 Lung cancer cell lines

chr7:54,984,552-55,170,743 zoom out: 2/3x 1/2x 1.5x zoom in: 1.5x 2x 5x

chr7:54,984,552-55,170,743

Sequence

Cytos Island

NCBI RefSeq

EGFR (NM_201284)

EGFR (NM_201283)

EGFR (NM_201282)

EGFR (NM_005228)

Slide: <- Mouse hold ->, Zoom: Double click

Adult tissues

- TSS-seq

Adipose

Adrenal

Brain1

Brain2

BrainC1

Breast

Colon

Heart

HeartC1

Kidney

KidneyC1

Liver

Lung

Представление
гена в геномном
браузере

Представление
информации о
транскриптах в
различных
тканях

Геномный браузер базы DBTSS



Нуклеотидные замены в различных типах клеток

Типы хроматина в различных типах клеток

Уровень метилирования

Модификации гистонов в различных клетках

Информация, добавленная в DBTSS в 2015 г. :

Добавлены данные о заменах нуклеотидов (single nucleotide variation = SNV). Опухолевые клетки:

97 lung adenocarcinomas (пациенты –Японцы)

57 lung small cell carcinomas (пациенты –Японцы)

26 линий клеток рака легких

+ данные по раковым клеткам у 11,322 пациентов, которые были собраны из открытых источников по всему миру.

Методы TSS-seq, RNA-seq, ChIP-seq and BS-seq data.

BS-seq (= Bisulfite conversion of genomic DNA combined with next-generation sequencing) is widely used to measure the methylation state of a whole genome, the methylome, at single-base resolution.

Suzuki A et al., DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. Nucleic Acids Res. 2015 Jan;43(Database issue):D87-91.

Регулярное обновление версий базы DBTSS

(<http://dbtss.hgc.jp/index.html>)

Database of Transcriptional Start Sites
DBTSS

English | [Japanese](#)

[Link to old version](#)

Database Search

Keyword Search

Species:
H. sapiens

Cell:
TSSseq : DLD1 : clorectal adenocarcino

Category:
RefSeq ID (NM_)

Keyword:
*

ppm #(>=):
5

(Sep,2011,update)

- DataBase of Transcriptional Start Sites -
DBTSS

Release 8.0 Updated (September 15 2011)
Based on UCSC hg19, mm9

We recommend to use the Internet Explorer 6.0 or later for visiting our database.

About this Database

DBTSS: Database of Transcriptional Start Sites

Current version is based on UCSC hg19, mm9

ABSTRACT

To support transcriptional regulation studies, we have constructed the DBTSS (DataBase of Transcriptional Start Sites) which contains unique positions of transcriptional start sites (TSS) and unique experimentally validated TSS sequences. In this update, we included new TSS data, so that a major portion of embryonic tissues are covered. DBTSS now contains tag sequences for collected from a total of

Доступ к более ранним версиям:

Select DBTSS Version

[DBTSS Version 8.0](#) (Based on UCSC hg19, mm9)

[DBTSS Version 7.0](#) (Based on UCSC hg18, mm9)

[DBTSS Version 6.0](#) (Based on UCSC hg18, mm8)

[DBTSS Version 5.2](#) (Based on UCSC hg17)

[DBTSS Version 5.1](#) (Based on UCSC hg17)

[DBTSS Version 5.0](#) (Based on UCSC hg17)

[DBTSS Version 4.0](#) (Based on UCSC hg16)

[DBTSS Version 3.0](#) (Based on UCSC hg13)

TSS =Transcriptional Start Site
Точные позиции TSS в геномах
Человека, мыши и еще 5 видов
организмов

Информационное содержание и «изюминки» каждого релиза (версии) базы DBTSS (<http://dbtss.hgc.jp/index.html>)

Информация, представленная в DBTSS в 2012 г.:

Точные позиции стартов транскрипции (TSS) в геноме. Данные получены на основании комбинированного экспериментально-теоретического подхода **TSS Seq**.

Данные по тканям взрослого организма и эмбриона:

- Исследовано 20 тканей и 7 клеточных культур;
- 491 млн. TSS tag sequences

-Перессылки на данные из других баз:

-TRANSFAC

-Проект ENCODE (модификации хроматина)

Информационное содержание и «изюминки» каждого релиза (версии) базы DBTSS (<http://dbtss.hgc.jp/index.html>)

DBTSS развивается с **2002** г.

С **2006** г. имеется следующая информация :

	<i>#covered locus</i>	<i>#promoter</i>	<i>#total TSSs</i>	<i>#allocated clones</i>
<i>human</i>	15,262	30,964	425,117	1,359,000
<i>mouse</i>	14,162	19,023	149,876	364,487
<i>zebrafish</i>	3,061	3,382	15,198	32,263
<i>malaria</i>	1,527	N.A.	6,908	10,236
<i>schyzon</i>	3,635	N.A.	14,029	22,923

В **2008** году в базу добавлена новая информация :

	Total no. of mapped sequences	No. of sequences associated with NMs	No. represented NMs	Total no. of putative promoters
MCF7 (Solexa)	11 919 330	10 000 349	12 133	29 210
HEK293 (Solexa)	10 062 560	8 633 345	11 598	41 238
CDNA (Sanger, total)	1 540 411	1 370 985	15 194	32 122

Всего в 2008 году добавлено 19 000 000 5' -концевых участков

В **2009** году добавлена новая информация для различных клеток человека и мыши :

Sample name	Cell type	Tag count
Fetal Heart	Normal fetal tissues	10 182 282
Fetal Kidney	Normal fetal tissues	8 424 482
Fetal Liver	Normal fetal tissues	4 741 889
Fetal Thymus	Normal fetal tissues	7 122 556
Fetal Brain	Normal fetal tissues	11 285 710
Brain	Normal adult tissues	11 561 960
Heart	Normal adult tissues	9 378 901
Kidney	Normal adult tissues	11 196 359
Mouse 3T3	Fibroblast	20 246 303
DLD1	Fibroblast	
Beas2B	Bcell	
Ramos	Bcell	
MCF7	Breast adenocarcinoma	
TIG	Fetal lung	
293	Embryonic kidney	
Total		330 533 354

9 клеточных ситуаций

В норме и при различных воздействиях

Всего в 2009 году добавлено 330 000 000 5' -концевых участков из 31 клеточной ситуации

22 клеточные ситуации

DBTSS: всю информацию можно скачать !!!

Содержание /pub/hgc/db/dbtss/		
Название	Размер	Последнее изменение
 [родительский каталог]		
 MinION_seminor_2016/		21.07.16, 14:29:00
 Yamashita_NAR/		20.09.05, 0:00:00
 Yamashita_et_al/		07.04.08, 1:00:00
 annotaion	1.3 kB	08.07.05, 1:00:00
 data	4.3 kB	08.07.05, 1:00:00
 dbtss_recently	0 B	09.12.14, 23:00:00
 dbtss_ver4/		06.05.05, 1:00:00
 dbtss_ver5/		25.05.05, 1:00:00
 dbtss_ver5_1	0 B	09.12.14, 23:00:00
 dbtss_ver5_2/		07.09.06, 0:00:00
 dbtss_ver6/		15.09.09, 0:00:00
 dbtss_ver7/		20.07.10, 1:00:00
 dbtss_ver8/		12.07.12, 1:00:00
 dbtss_ver9/		15.09.14, 0:00:00
 old/		06.05.05, 1:00:00
 other/		14.05.15, 0:00:00

Ftp – сайт, где доступна информация из всех релизов DBTSS

<ftp://ftp.hgc.jp/pub/hgc/db/dbtss/>



DBTSS – «идеальная» база данных ??

- 1) DBTSS развивается давно, с 2002 г. и регулярно пополняется новой информацией
- 2) Данные полногеномные, по нескольким видам организмов
- 3) Всю информацию можно скачать

DBTSS – не идеальная база данных ((



Типы информации о регуляторных районах ограничены (стартеры транскрипции, единичные сайты связывания ТФ и только по данным ChIP-seq)

The Encyclopedia of DNA Elements (ENCODE)

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

The screenshot displays the ENCODE Project Portal interface. On the left is a dark blue sidebar with white text 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 has a light gray header with the ENCODE logo and the text 'Encyclopedia of DNA Elements at UCSC 2003 - 2012'. Below this is an 'About' section with text about the consortium and data availability. The page is divided into 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 interface, and 'Search for ENCODE tracks (2003 - 2012) in the UCSC Browser' showing a search interface for tracks.

Разрабатывается международным исследовательским консорциумом.
Финансирование идет через National Human Genome Research Institute (NHGRI).
Цель проекта - объединение всех данных по функциональным элементам генома человека, включая:

- белок- и РНК-кодирующие участки генома,
- регуляторные участки

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

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

3C

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

Monocytes-CD14+

Monocytes-CD14+_RO01746

SK-N-SH

Tier 3

8988T

Adult CD4 Th0

1

1

1

2

1

3

16

3

7

4

4

6

2

6

2

12

5

133

2

2

2

3

1

1

5

1

1

1

1

2

1

6

1

1

2

1

1

2

1

3

10

9

87

4

2

5

4

93

114

36

11

49

17

34

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

1

Геномный браузер Университета г.Санта Круз

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[About Us](#)[View](#)[Help](#)

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

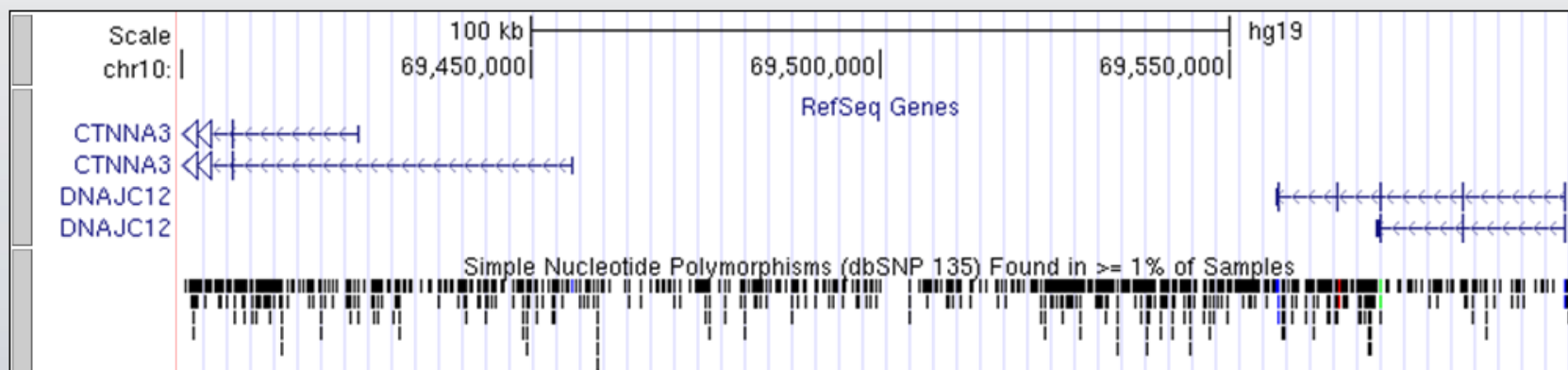
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr10:69,400,000-69,600,000 200,001 bp.

enter position, gene symbol or search terms

go

chr10 (q21.3) p14 p13 q21.1 21.1 23.1 25.1



move start

< 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

move end

< 2.0 >

track search

default tracks

default order

hide all

add custom tracks

track hubs

configure

reverse

resize

refresh

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

Mapping and Sequencing Tracks

refresh

Base Position

dense

Chromosome Band

hide

STS Markers

hide

18 FISH Clones

hide

Recomb Rate

hide

18 deCODE Recomb

hide

ENCODE Pilot

hide

Map Contigs

hide

Assembly

hide

GRC Map Contigs

hide

Gap

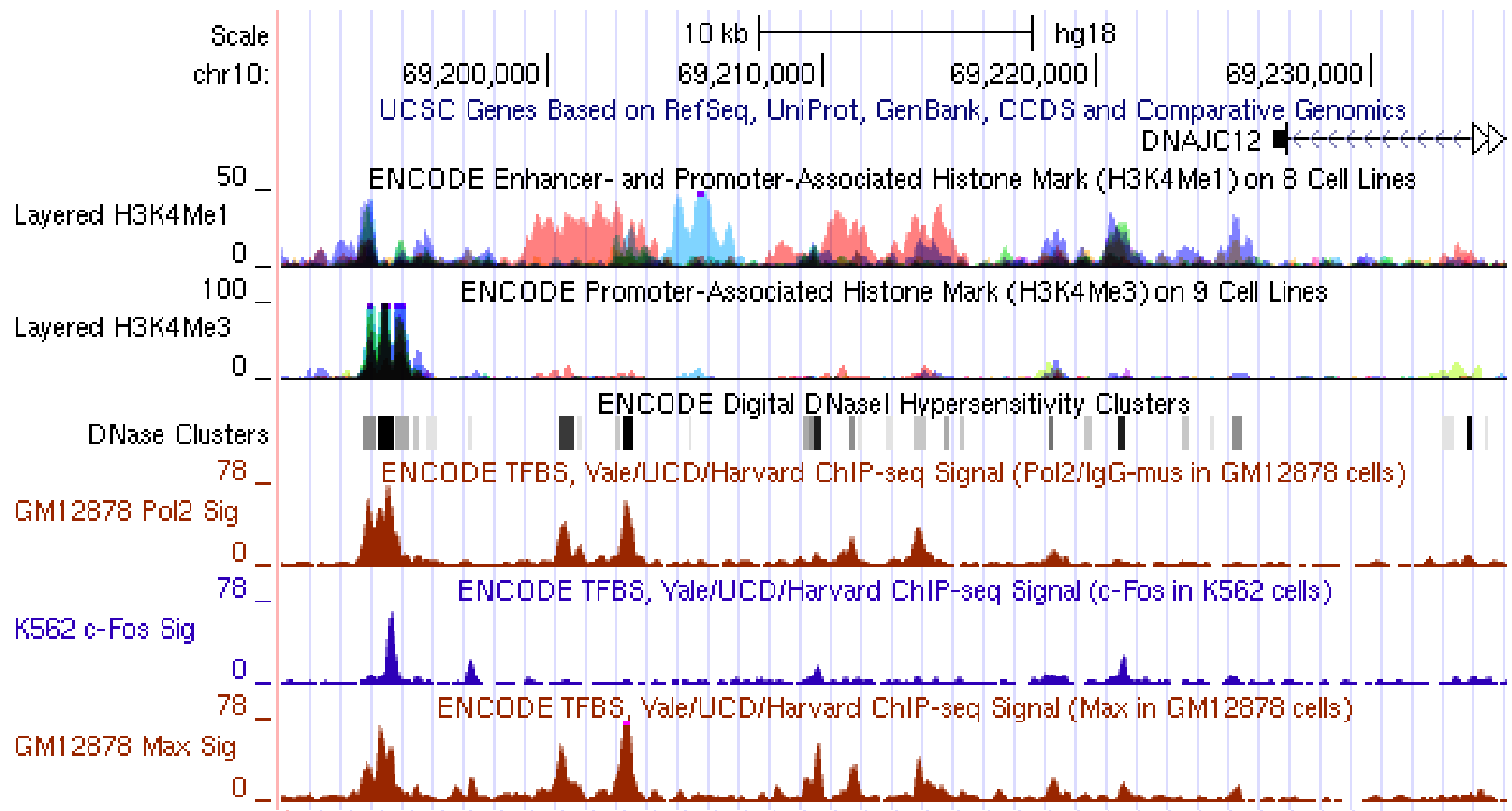
hide

Publications

hide

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

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



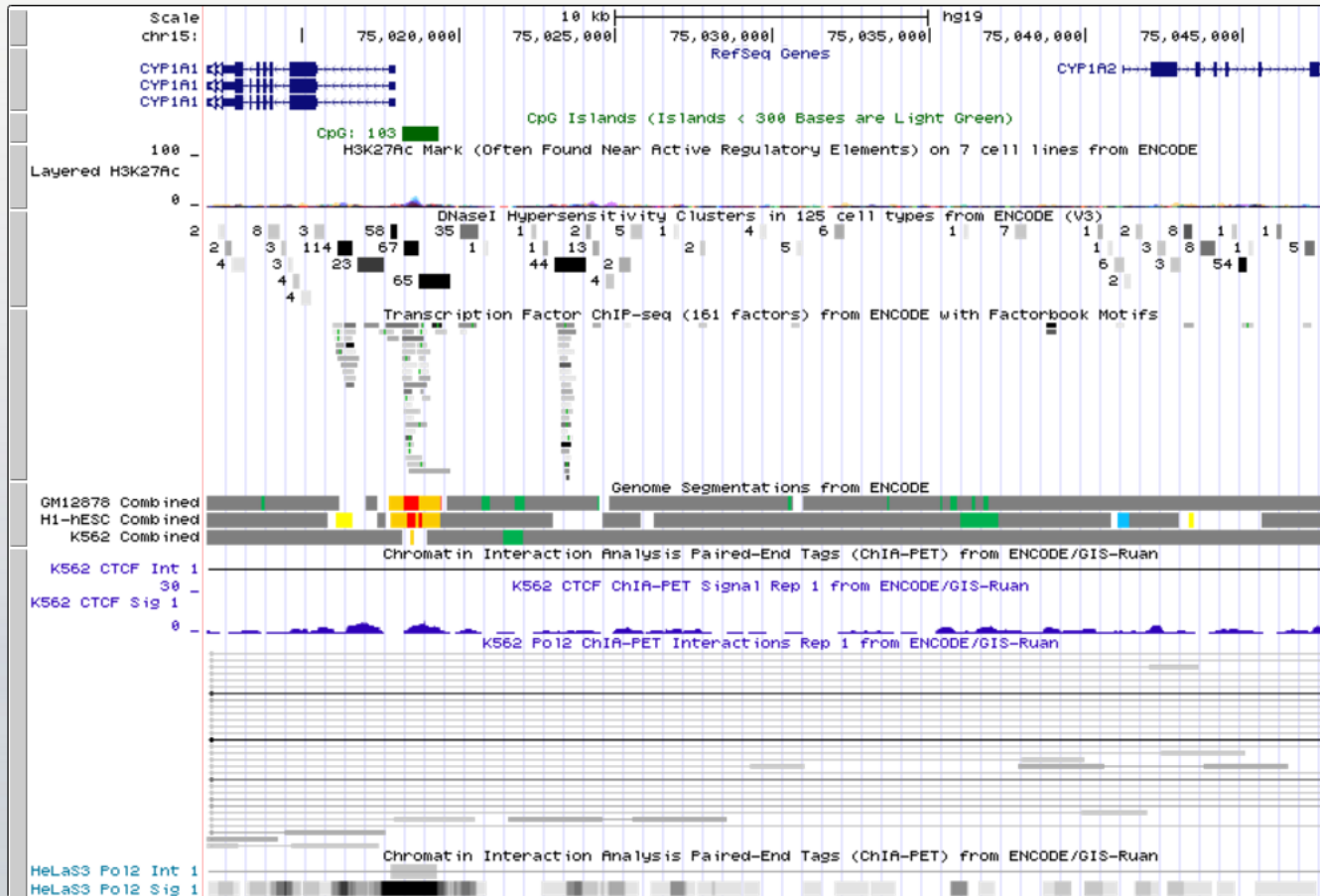
Геномный браузер Университета г.Санта Круз

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

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

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](#).

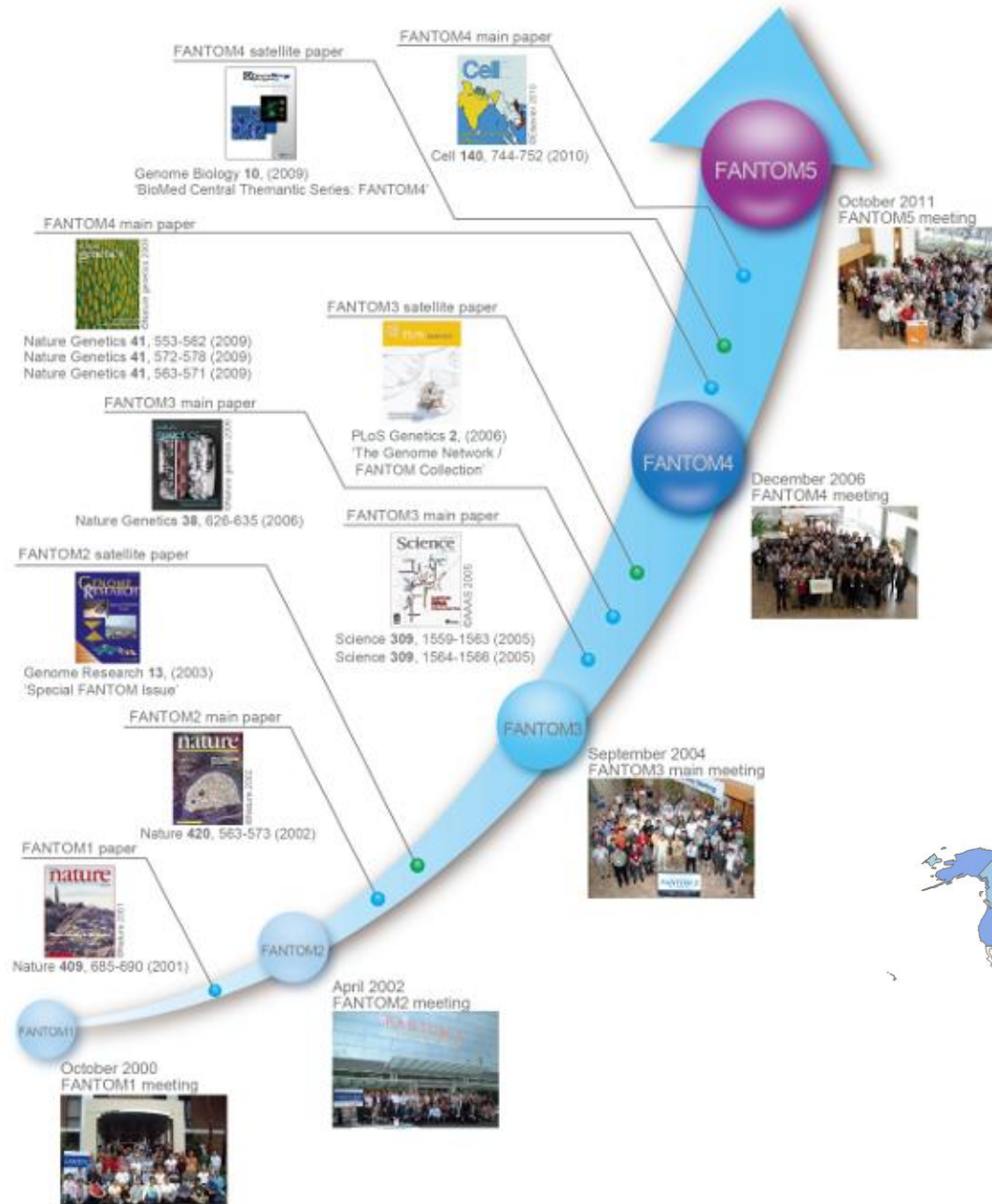
Using the Table Browser

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)

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)

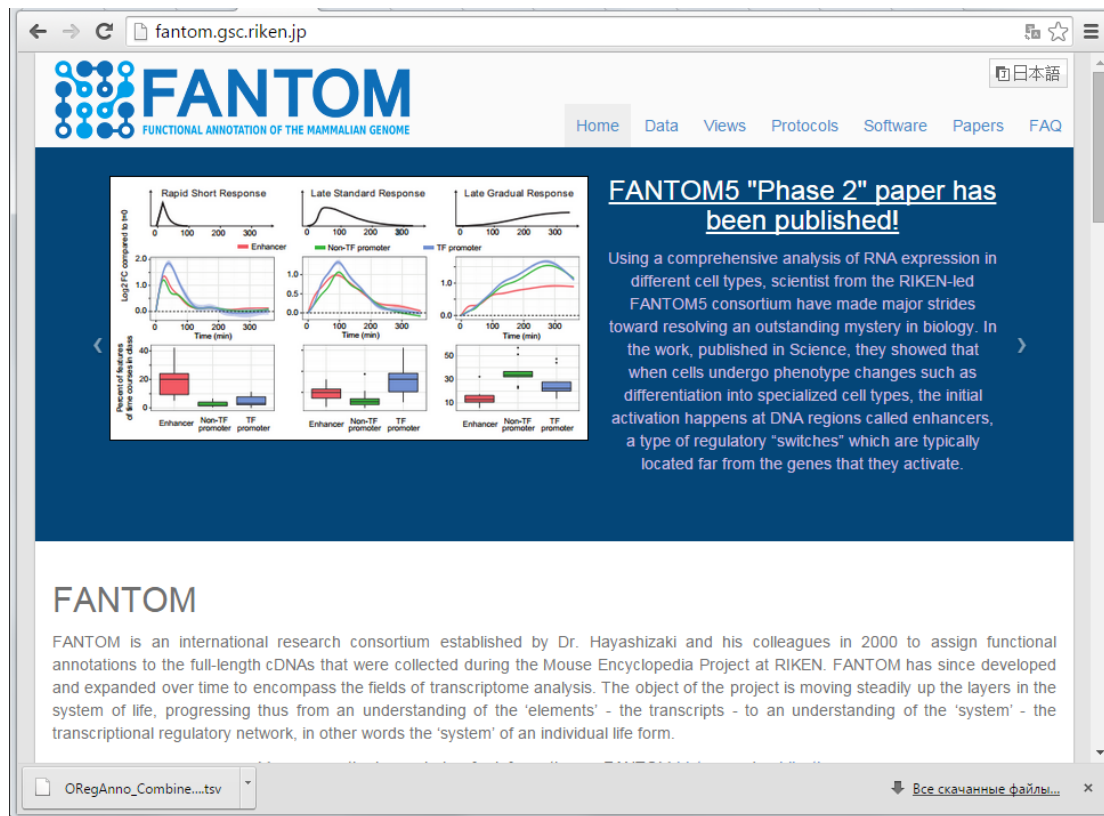
Members

Over 500 FANTOM members from more than 20 countries.



FANTOM5

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



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



日本語

[Home](#)

[Data](#)

[Views](#)

[Protocols](#)

[Software](#)

[Papers](#)

[FAQ](#)

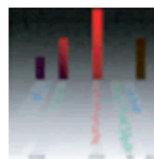
Protocols

Detailed descriptions on protocols used in FANTOM



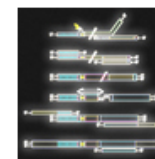
Basic CAGE Technology

A method for genome-wide identification of transcription start sites



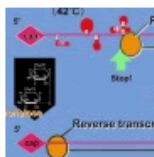
DeepCAGE

A powerful application for next generation sequencing



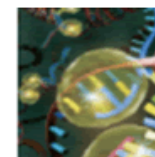
nanoCAGE and CAGEscan

The two new methods developed to extend the capabilities of CAGE technology



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.



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.

БАЗЫ ДАННЫХ, СОДЕРЖАЩИЕ ИНФОРМАЦИЮ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

UniProtKB Switzerland	Protein knowledgebase	Geneva,
TFClass	Classification of transcription factors	Germany
AnimalTFDB	Animal Transcription Factor DataBase	Китай
CREMOFAC	Database of chromatin remodeling factors	Индия
TcoF- DB	Dragon database of transcription co-factors and transcription factor interacting proteins	Королевство Саудовская Аравия

База данных о белках UniProtKB

<http://www.uniprot.org/>

Swiss Institute of Bioinformatics (Geneva)



UniProtKB ▾

Advanced ▾  Search

BLAST Align Retrieve/ID mapping Help Contact

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt Knowledgebase

Swiss-Prot (549,215)

 Manually annotated and reviewed.

TrEMBL (50,825,784)

 Automatically annotated and not reviewed.

UniRef

Sequence clusters



UniParc

Sequence archive



Proteomes



Supporting data

Literature citations 

Cross-ref. databases 

Taxonomy 

Diseases 

Subcellular locations 

Keywords 

News



Forthcoming changes

Planned changes for UniProt

UniProt release 2015_09

Life (and death) in 2D | 27 new species in variation files

UniProt release 2015_08

Pseudo-allergy, real progress | Programmatic access to UniProt with sparql.uniprot.org | Addition of [UniProt release 2015_08](#)

 [News archive](#)

Getting started

 [Text search](#)

Our basic text search allows you to



UniProt data

 [Download latest release](#)

Get the UniProt data

Protein spotlight



UniProtKB

UniProtKB results

 Basket ▾

UniProtKB consists of two sections:



Reviewed (Swiss-Prot) - Manually annotated

Records with information extracted from literature and curator-evaluated computational analysis.



Unreviewed (TrEMBL) - Computationally analyzed

Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

 Help

 UniProtKB help video

 Other tutorials and videos

 Downloads

Filter byⁱ



Reviewed
(549,215)
Swiss-Prot



Unreviewed
(50,825,784)
TrEMBL

Popular organisms

Human (148,284)

Rice (99,786)

Mouse (77,279)

Zebrafish
(58,647)

A. thaliana
(52,169)

Other organisms

 BLAST

 Align

 Download

 Add to basket

◀ 1 to 25 of 51,374,999 ▶ Show 25 ▾

 Columns	Entry ▾	Entry name ▾		Protein names ▾	Gene names ▾	Organism ▾	Length ▾	
☐	Q6GZX3	002L_FRG3G		Uncharacterized protein 002L	FV3-002L	Frog virus 3 (isolate Goorha) (FV-3)	320	
☐	Q6GZX4	001R_FRG3G		Putative transcription factor 001R	FV3-001R	Frog virus 3 (isolate Goorha) (FV-3)	256	
☐	Q197F7	003L_IIV3		Uncharacterized protein 003L	IIV3-003L	Invertebrate iridescent virus 3 (IIV-3) (Mosquito iridescent virus)	156	
☐	Q197F8	002R_IIV3		Uncharacterized protein 002R	IIV3-002R	Invertebrate iridescent virus 3 (IIV-3) (Mosquito iridescent virus)	458	

UniProtKB

В базах **SWISS-PROT (549 215)** и **TrEMBL (50 825 784)** содержится описание структурно-функциональной организации белков, в числе которых белки - транскрипционные регуляторы. Обязательно приводятся аминокислотные последовательности и доменная организация. Для многих белков приводится информация о ткане-специфической экспрессии.

Funding

UniProt is mainly supported by the [National Institutes of Health \(NIH\)](#) grant U41HG007822. Additional support for the EMBL-EBI's involvement in UniProt comes from [European Molecular Biology Laboratory \(EMBL\)](#), the [British Heart Foundation \(BHF\)](#) (RG/13/5/30112), the [Parkinson's Disease United Kingdom \(PDUK\)](#) GO grant G-1307, and the NIH GO grant U41HG02273. UniProt activities at the SIB are additionally supported by the Swiss Federal Government through the [State Secretariat for Education, Research and Innovation SERI](#). PIR's UniProt activities are also supported by the NIH grants R01GM080646, G08LM010720, and P20GM103446, and the [National Science Foundation \(NSF\)](#) grant DBI-1062520.

Past funding

UniProt has been mainly supported by the NIH grants U01HG02712 (2002-2010) and U41HG006104 (2010-2014).

UniProt activities at EMBL-EBI have benefited from the FP7 SLING project (2009-2012, contract number 226073) and a British Heart Foundation grant (SP/07/007/23671).



Разработчики базы UniProtKB/Swiss-Prot (UniProt consortium)

- Швейцарский институт биоинформатики (Swiss Institute of Bioinformatics, SIB)
- Европейский Институт биоинформатики (European Bioinformatics Institute, EBI)
- Американский информационный ресурс по белкам (the Protein Information Resource, PIR).



Информационное наполнение UniProtKB/Swiss-Prot в 2015 году
Имеется ИНФОРМАЦИЯ по ВСЕМ ВИДАМ ОРГАНИЗМОВ. Из них
(ручная аннотация)

20 204 входа, описывающих белки человека,

16 719 входа для мыши

7 928 входа для крысы.

ПРИМЕР ОПИСАНИЯ БЕЛКА CREB в базе UNIPROT/SWISS-PROT

P16220 (CREB1_HUMAN)★ Reviewed, UniProtKB/Swiss-Prot

Last modified September 18, 2013. Version 168.  [History...](#)

[Contribute](#)

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

 Clusters with [100%](#), [90%](#), [50%](#) identity |  [Documents \(6\)](#) |  [Third-party data](#)

[text](#) [xml](#) [rdf/xml](#) [gff](#) [fasta](#)

 [Names](#) · [Attributes](#) · [General annotation](#) · [Ontologies](#) · [Interactions](#) · [Alt products](#) · [Sequence annotation](#) · [Sequences](#) · [References](#) · [Cross-refs](#) · [Entry info](#) · [Documents](#) [Customize order](#)

Names and origin

Protein names	<i>Recommended name:</i> Cyclic AMP-responsive element-binding protein 1 Short name=CREB-1 Short name=cAMP-responsive element-binding protein 1
Gene names	Name: CREB1
Organism	Homo sapiens (Human) [Reference proteome]
Taxonomic identifier	9606 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Euarchontoglires › Primates › Haplorrhini › Catarrhini › Hominidae › Homo 

Protein attributes

Sequence length	341 AA.
Sequence status	Complete.
Protein existence	Evidence at protein level

ПРИМЕР ОПИСАНИЯ БЕЛКА CREB в базе UNIPROT/SWISS-PROT (продолжение)

Names · Attributes · General annotation · Ontologies · Interactions · Alt products · Sequence annotation · Sequences · References · Cross-refs · Entry info · Documents [Customize order](#)

General annotation (Comments)

Function	Phosphorylation-dependent transcription factor that stimulates transcription upon binding to the DNA cAMP response element (CRE), a sequence present in many viral and cellular promoters. Transcription activation is enhanced by the TORC coactivators which act independently of Ser-133 phosphorylation. Involved in different cellular processes including the synchronization of circadian rhythmicity and the differentiation of adipose cells.
Subunit structure	Interacts with PPRC1. Binds DNA as a dimer. This dimer is stabilized by magnesium ions. Interacts, through the bZIP domain, with the coactivators TORC1/CRTC1, TORC2/CRTC2 and TORC3/CRTC3. When phosphorylated on Ser-133, binds CREBBP By similarity . Interacts with CREBL2; regulates CREB1 phosphorylation, stability and transcriptional activity By similarity . Interacts (phosphorylated form) with TOX3. Interacts with ARRB1. Binds to HIPK2. Interacts with SGK1. Ref.9 Ref.10 Ref.15 Ref.17 Ref.18 Ref.19 Ref.20 Ref.21 Ref.26 Ref.27
Subcellular location	Nucleus Ref.16 .
Post-translational modification	Stimulated by phosphorylation. Phosphorylation of both Ser-133 and Ser-142 in the SCN regulates the activity of CREB and participates in circadian rhythm generation. Phosphorylation of Ser-133 allows CREBBP binding By similarity. CREBL2 positively regulates phosphorylation at Ser-133 thereby stimulating CREB1 transcriptional activity By similarity. Phosphorylated upon calcium influx by CaMK4 and CaMK2 on Ser-133. CaMK4 is much more potent than CaMK2 in activating CREB. Phosphorylated by CaMK2 on Ser-142. Phosphorylation of Ser-142 blocks CREB-mediated transcription even when Ser-133 is phosphorylated. Phosphorylated by CaMK1 By similarity. Phosphorylation of Ser-271 by HIPK2 in response to genotoxic stress promotes CREB1 activity, facilitating the recruitment of the coactivator CBP. Phosphorylated at Ser-133 by RPS6KA3, RPS6KA4 and RPS6KA5 in response to mitogenic or stress stimuli. Ref.11 Ref.12 Ref.13 Ref.14 Ref.20 Ref.26 Sumoylated with SUMO1. Sumoylation on Lys-304, but not on Lys-285, is required for nuclear localization of this protein. Sumoylation is enhanced under hypoxia, promoting nuclear localization and stabilization. Ref.16

ПРИМЕР ОПИСАНИЯ БЕЛКА CREB в базе UNIPROT/SWISS-PROT (продолжение)

Involvement in disease	Angiomatoid fibrous histiocytoma (AFH) [MIM:612160]: A distinct variant of malignant fibrous histiocytoma that typically occurs in children and adolescents and is manifest by nodular subcutaneous growth. Characteristic microscopic features include lobulated sheets of histiocyte-like cells intimately associated with areas of hemorrhage and cystic pseudovascular spaces, as well as a striking cuffing of inflammatory cells, mimicking a lymph node metastasis. <u>Note</u>: The gene represented in this entry may be involved in disease pathogenesis. A chromosomal aberration involving CREB1 is found in a patient with angiomatoid fibrous histiocytoma. Translocation t(2;22)(q33;q12) with CREB1 generates a EWSR1/CREB1 fusion gene that is most common genetic abnormality in this tumor type. A CREB1 mutation has been found in a patient with multiple congenital anomalies consisting of agenesis of the corpus callosum, cerebellar hypoplasia, severe neonatal respiratory distress refractory to surfactant, thymus hypoplasia, and thyroid follicular hypoplasia (Ref.29).
------------------------	---

Sequence similarities	Belongs to the bZIP family. Contains 1 bZIP (basic-leucine zipper) domain. Contains 1 KID (kinase-inducible) domain.
-----------------------	---

Ontologies

Keywords

Biological process	Differentiation Host-virus interaction Transcription Transcription regulation
Cellular component	Nucleus
Coding sequence diversity	Alternative splicing

ПРИМЕР ОПИСАНИЯ БЕЛКА CREB в базе UNIPROT/SWISS-PROT (продолжение)

Sequence annotation (Features)

	Feature key	Position (s)	Length	Description	Graphical view	Feature identifier
Molecule processing						
☐	Chain	1 – 341	341	Cyclic AMP-responsive element-binding protein 1		PRO_0000076597
Regions						
☐	Domain	101 – 160	60	KID		
☐	Domain	283 – 341	59	bZIP		
☐	Region	284 – 309	26	Basic motif By similarity		
☐	Region	311 – 332	22	Leucine-zipper By similarity		
Sites						
☐	Site	314	1	Required for binding TORCs		
Amino acid modifications						
☐	Modified residue	133	1	Phosphoserine; by CaMK1, CaMK2, CaMK4, PKB/AKT1 or		

TFClass is a classification of (so far: human) transcription factors based on the characteristics of their DNA-binding domains.

<http://tfclass2.sybig.de/tfclass/>

Classification of Human Transcription Factors

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

Transcription factory classification

Superclass: , Class: , Family: , Subfamily: , Genus: , Factor species: 

Human TF

- ▼  1 Basic domains
 - ▶  1.1 Basic leucine zipper factors (bZIP)
 - ▶  1.2 Basic helix-loop-helix factors (bHLH)
 - ▶  1.3 Basic helix-span-helix factors (bHSH)
- ▼  2 Zinc-coordinating DNA-binding domains
 - ▶  2.1 Nuclear receptors with C4 zinc fingers
 - ▶  2.2 Other C4 zinc finger-type factors
 - ▶  2.3 C2H2 zinc finger factors
 - ▶  2.4 C6 zinc cluster factors
 - ▶  2.5 DM-type intertwined zinc finger factors

Search: 

Expand all

Collapse all



Expand to:



Details

Protein expression pattern

ID: 1.1

Definition: TRANSFAC class description C0008: A DNA-binding basic region is followed by a leucine zipper. The leucine zipper consists of repeated leucine residues at every seventh position and mediates protein dimerization as a prerequisite for DNA-binding. The leucines are directed towards one side of an alpha-helix. The leucine side chains of two

Данные о конкретном белке в базе TFClass

<http://tfclass2.sybig.de/tfclass/>

Classification of Human Transcription Factors

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

Transcription factory classification

Superclass: , Class: , Family: , Subfamily: , Genus: , Factor species: 

Human TF

- ▼  1 Basic domains
 - ▶  1.1 Basic leucine zipper factors (bZIP)
 - ▼  1.2 Basic helix-loop-helix factors (bHLH)
 - ▼  1.2.1 E2A-related factors
 -  1.2.1.0.3 HTF-4 (TCF-12, HEB)
 - ▶  1.2.1.0.2 SEF2 (E2-2, TCF-4, ITF-2)
 - ▶  1.2.1.0.1 E2A (TCF-3, ITF-1)
 - ▶  1.2.2 MyoD / ASC-related factors
 - ▶  1.2.3 Tal-related factors
 - ▶  1.2.4 Hairy-related factors
 - ▶  1.2.5 PAS domain factors
 - ▶  1.2.6 bHLH-ZIP factors
 - ▶  1.2.8 HLH domain only
 - ▶  1.3 Basic helix-span-helix factors (bHSH)
- ▼  2 Zinc-coordinating DNA-binding domains

Search: 

Expand all

Collapse all



Expand to:



Details

Protein expression pattern

The table below summarizes the protein expression data from the [Protein Atlas](#). The tissues and cell types are linked to the [Cytochrome ontology](#).

 Tissue	 Cell type	 Expression level Select 	 Type	 Reliability
adrenal gland	glandular tissue	High	APE	Low
appendix	glandular tissue	Medium	APE	Low
appendix	lymphoid tissue	Medium	APE	Low
bone marrow	hematopoietic stem cell	High	APE	Low
breast	glandular tissue	High	APE	Low
bronchi	ciliated cell with propulsive function of	High	APE	Low



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

Характеристика всех

транскрипционных факторов,

кофакторов и

белков с хроматин-моделирующей активностью, которые удалось выявить разработчикам базы, используя компьютерный анализ геномов 50 видов животных организмов.

Для систематизации информации по транскрипционным факторам использована оригинальная классификация. Эта классификация построена на основе анализа научных публикаций и включает 72 семейства.

В) поисковая система

No image								No image	No image	
AF-4	Androgen receptor	AR-2	ARID	DMLH	C/EBP	CBF	CG-1	COE	COUP	
No image				No image						
CP2	CSO	CSL	CTF/NFI	CUT	DM	E2F	Ets/tyk receptor	ETS	Fox head	
										
GCM	GCM	GTF2I	HMG	HMG1/HMG2	Hmbox3	HSP	HTH	IRF	NBD	
			No image		No image					
Maf1	MYB	NDR30/Pho G	NP-1YA	NP-VBIC	Nrf1	Nuclear orphan receptor	Oestrogen receptor	Other nuclear receptor	Others	
										
P53	PAK	PCL	POU	RAR receptor	Progesterone receptor	Prlr1	Retinoid acid receptor	RFX	RHD	
No image										
ROR nuclear receptor	RuvE	SAND	SRF	STAT	T-bcl1	TEA	TF_2BP1	TF_100	T-HAP	
No image										No image
TSC22	Tus	ZBTB	ZF-BBD	ZF-C2H2	ZF-C2HC	ZF-GAGA	ZF-GATA	ZF-LTAF-like		

A circular phylogenetic tree of life. The tree is rooted in the center and branches outwards to various groups. Each group is labeled with a text label and accompanied by several small images of representative animals. The groups and their associated labels are: Afrotheria, Xenarthra, Other Mammals, Birds & Reptiles, Amphibians, Fishes, Other Chordates, Other Eukaryotes, Primates, Rodents, and Laurasiatheria. The tree shows the evolutionary relationships between these groups, with branches indicating common ancestry.

AnimalTFDB –поиск по ключевому слову (Symbol)



Search by Basic Information

Search by Basic Information

Symbol

EnsemblID
Gene ID
Transcript ID
Protein ID
Symbol
Map location
Description

notation

(e.g. 1nvp, 1nvpB, 1nvpC, or 1nvpD)

PPI: protein-protein interaction, Fill in a symbol (e.g. MSH6)

KEGG and Biocarta pathway ID or title (e.g. hsa04630, 635, Jak-STAT signaling pathway)

Homologous genes in different species, fill in a EnsemblID (e.g. ENSG00000081189)

Homologous genes in the same species, fill in a EnsemblID (e.g. ENSG00000081189)

Pfam protein domain id or name (e.g. PF00319.11 or SRF-TF)

show records per page:

Species ☐ select all ☐ reverse select

☐ Ailuropoda melanoleuca	☐ Anolis carolinensis	☐ Bos taurus
☐ Caenorhabditis elegans	☐ Callithrix jacchus	☐ Canis familiaris
☐ Cavia porcellus	☐ Choloepus hoffmanni	☐ Ciona intestinalis
☐ Ciona savignyi	☐ Danio rerio	☐ Dasypus novemcinctus
☐ Dipodomys ordii	☐ Drosophila melanogaster	☐ Echinops telfairi
☐ Equus caballus		
☐ Gallus gallus		
☒ Homo sapiens		
☐ Macropus eugenii		

No.	EnsemblID	Species	Family
1	ENSG00000109819	Homo sapiens	Transcription co-factors
2	ENSG00000112033	Homo sapiens	PPAR receptor
3	ENSG00000132170	Homo sapiens	PPAR receptor
4	ENSG00000155846	Homo sapiens	Transcription co-factors
5	ENSG00000186951	Homo sapiens	PPAR receptor

AnimalTFDB – результат поиска, фактор PPARΔ

Homo sapiens PPAR receptor family

Basic information

EnsemblID : [ENSG00000112033](#)

GeneID : [5467](#)

Symbol : PPARΔ

Alias : FAAR;MGC3931;NR1C2;NUC1;NUCI;NUCII;PPARB

Full name : peroxisome proliferator-activated receptor delta

Other designations : OTTHUMP00000016256;OTTHUMP00000016257;PPAR-beta;PPAR-delta;nuclear hormone receptor 1;nuclear receptor subfamily 1 group C member 2;peroxisome proliferator-activated receptor beta

Chr map location : 6p21.2

Gene Orientation : forward

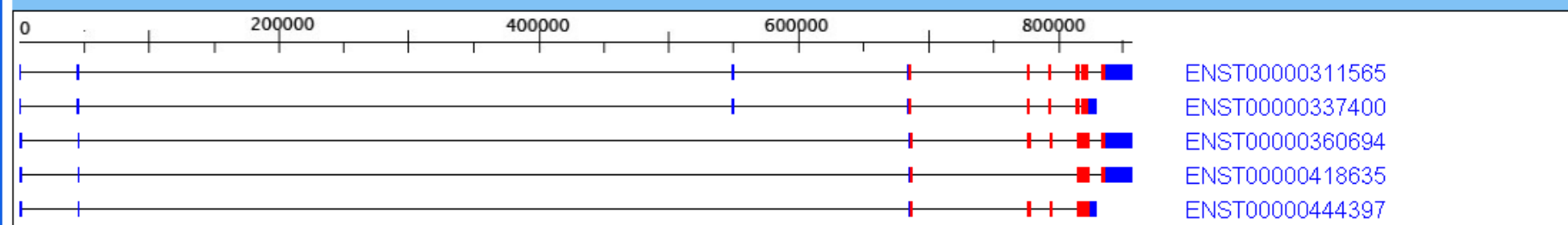
Gene Length : 85621

Gene Position : 6 : 35310335-35395955

Transcripts : There are 5 transcripts

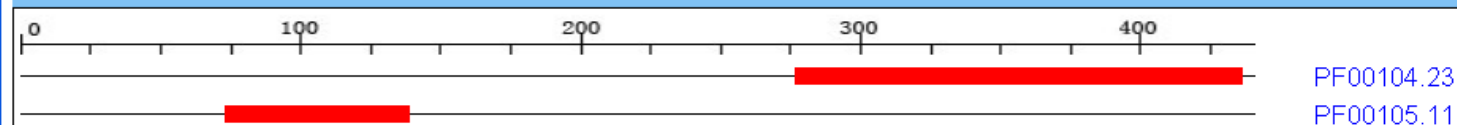
Transcript ID	Length (bp)	Protein ID	Length (aa)
ENST00000311565	3774	ENSP00000310928	441
ENST00000337400	2041	ENSP00000337063	361
ENST00000360694	3747	ENSP00000353916	441
ENST00000418635	3453	ENSP00000413314	343
ENST00000444397	2002	ENSP00000410837	361

Gene structure(for all transcripts) (red: CDS exon, blue: UTR)



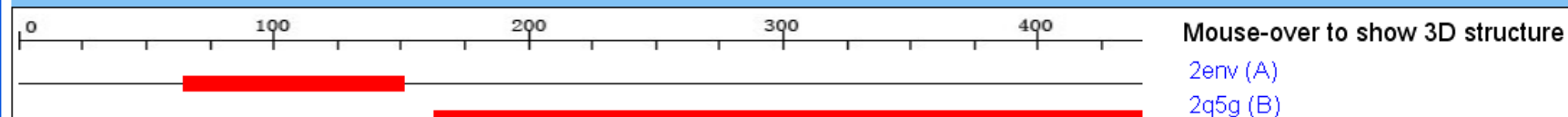
AnimalTFDB – результат поиска, фактор PPARD (продолжение)

Function Domain (for the longest protein)



No.	Domain	Entry	Score	E-value	Start	End
1	Hormone_recep	PF00104.23	88	3.2e-25	277	436
2	zf-C4	PF00105.11	97.2	3.7e-28	73	139

3D structure Hit (for the longest protein)



PDB ID (chain)	Score	E-value	Identity	Start	End
2env (A)	433	0	90.8	65	151
2q5g (B)	1489	0	99.2	163	441

Gene ontology

GO ID	Go Term	Category	Evidence
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	Biological Process	ISS
GO:0001890	placenta development	Biological Process	IEA
GO:0003677	DNA binding	Molecular Function	ISS
GO:0003700	sequence-specific DNA binding transcription factor activity	Molecular Function	IDA NAS
GO:0003707	steroid hormone receptor activity	Molecular Function	IEA
GO:0003713	transcription coactivator activity	Molecular Function	IEA
GO:0004879	ligand-dependent nuclear receptor activity	Molecular Function	IDA
GO:0005504	fatty acid binding	Molecular Function	NAS
GO:0005634	nucleus	Cellular Component	NAS
GO:0005654	nucleoplasm	Cellular Component	TAS
GO:0006006	glucose metabolic process	Biological Process	NAS
GO:0006029	proteoglycan metabolic process	Biological Process	IEA

AnimalTFDB – результат поиска, фактор PPAR δ (продолжение 2)

Pathway information

PathwayID	Description	Source	All gene in this pathway
hsa03320	PPAR signaling pathway	KEGG	Show detail
hsa04310	Wnt signaling pathway	KEGG	Show detail
hsa05200	Pathways in cancer	KEGG	Show detail
hsa05221	Acute myeloid leukemia	KEGG	Show detail
156	WNT Signaling Pathway	Biocarta	Show detail
708	Nuclear Receptors in Lipid Metabolism and Toxicity	Biocarta	Show detail
84	Basic mechanism of action of PPAR α , PPAR δ and PPAR γ and effects on gene expression	Biocarta	Show detail

Protein protein Interaction

Interactors	GeneID	Experimental	Source	PMID
GADD45G	10912	Reconstituted Complex;in vitro	BioGRID HPRD	10872826
Gadd45b	17873	Reconstituted Complex;in vitro	BioGRID HPRD	10872826
HDAC4	9759	Reconstituted Complex	BioGRID	12943985

Paralog

ENSG00000132170	ENSG00000186951		
---------------------------------	---------------------------------	--	--

Ortholog group

Loxodonta africana	ENSLAFG00000017670	Macaca mulatta	ENSMUG00000015904
Macropus eugenii	ENSMEUG00000013000	Microcebus murinus	ENSMICG00000000028
Monodelphis domestica	ENSMODG00000013778	Mus musculus	ENSMUSG00000002250
Myotis lucifugus	ENSMLUG00000010594	Ochotona princeps	ENSOPRG00000015187
Oryctolagus cuniculus	ENSOCUG00000007732	Oryzias latipes	ENSORLG00000006636
Otolemur aarnettii	ENSOGAG00000000630	Pan troglodytes	ENSPTRG00000018082

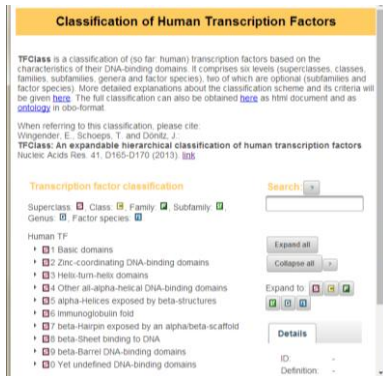
Targets & TFBS

Number of targets	Number of TFBS	TFBS ID	Data Source
-------------------	----------------	---------	-------------

факторов в геноме человека

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

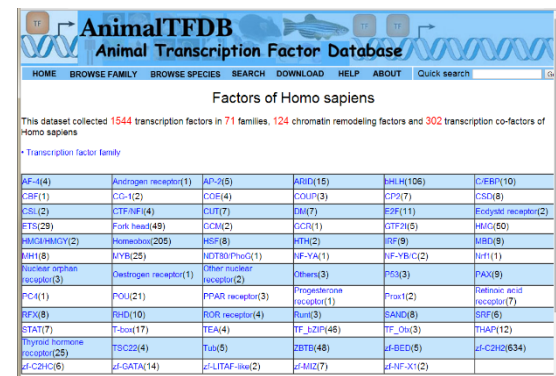
(<http://tfclass.bioinf.med.uni-goettingen.de/>)



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

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

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



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

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

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

Конец 5-ой лекции