

Лекция №5

БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ

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

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)

ИНФОРМАЦИЯ ПО БЕЛКАМ, РЕГУЛИРУЮЩИМ ТРАНСКРИПЦИЮ

UniProtKB Protein knowledgebase Geneva,
Switzerland

TFClass Classification of transcription factors Germany

AnimalTFDB Animal Transcription Factor DataBase Китай

РЕГУЛЯТОРНЫЕ ЭЛЕМЕНТЫ на ДНК

TRED Transcriptional Regulatory Element Database Cold Spring Harbor
Laboratory USA

DBTSS DataBase of Transcriptional Start Sites Japan

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

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

← → ↺ 🏠

gene-regulation.com/pub/databases.html#transfac

📖 ☆ ≡ ✎ 🔔 ⋮

gene-regulation.com

sponsored by geneXplain

Home

Databases

Programs

Publications

Resources

TRANSFAC® Professional

Subscribe to TRANSFAC® Professional:

+ 6x more data and matrices

+ >30 million ChIP-seq sites

+ Download option with flat files and command line tools

SUBSCRIBE TODAY

FREE TRIAL

Like

Like us on FACEBOOK for special promotions, news and events!

Public Databases for Academic and Non-profit Organizations

TRANSFAC® 7.0 Public 2005 and TRANSCompel 7.0 Public 2005

TRANSFAC® provides data on eukaryotic transcription factors, their experimentally-proven binding sites, consensus binding sequences (positional weight matrices) and regulated genes. TRANSCompel contains data on eukaryotic transcription factors experimentally proven to act together in a synergistic or antagonistic manner.

The data provided here is only a snapshot from 2005. For a modest academic/non-profit price, subscription to TRANSFAC® Professional provides full access to regularly updated content that goes well beyond the breadth and depth of content offered by others, as well as more advanced tools and an easy-to-use interface. To learn more about TRANSFAC® Professional:

- Compare the public and professional versions
- Watch an introductory video on TRANSFAC® Professional
- Read about recently released features

Access TRANSFAC Public and TRANSCompel Public

- Search the TRANSFAC® Public database
- Search the TRANSCompel Public database
- Browse transcription factors by class
- TfBlast: Search the TRANSFAC® Factor Table by protein sequence
- molwSearch 1.0: Search for TRANSFAC® Factors by molecular weight
- View TRANSFAC® documentation, View TRANSCompel documentation

TRANSPATH® 6.0 Public 2005

TRANSPATH® provides data about protein-protein interactions and directed modification of proteins involved in signal transduction pathways, with a particular focus on signaling cascades that affect the activity of transcription factors.

The data provided here is only a snapshot from 2005. For a modest academic/non-profit price, subscription to TRANSPATH® provides full access to regularly updated content that goes well beyond the depth of content offered by others, as well as more advanced tools and an easy-to-use interface. [Learn more about TRANSPATH®.](#)

- Search the TRANSPATH® Public database

Supplementary resources provided by collaborative research groups

PathoDB® 2.0 Public 2005

PathoDB provides data on pathologically relevant mutated forms of transcription factors and their binding sites.

Database Login

> Name

> Password

Log In

Forgot Password?

New User Registration

Need Help?

Feedback

Contact us



Edgar Wingender

Dr. rer. nat.

Director

Universitätsmedizin

Göttingen , Göttingen ·

Institute of

Bioinformatics

Интерфейс для поиска в базе TRANSFAC



Search the TRANSFAC 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 TRANSFAC table to search in:

Please read the [Help](#) file.

Search TRANSFAC Factors by their molecular weight - [MolwSearch](#)

TRANSFAC 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。](#)

<http://gene-regulation.com/cgi-bin/pub/databases/transfac/search.cgi>

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](#) LQEMLLGGSASDAPHAAHPLHPLHMQEHMGTNVIVANTMPSHLSNGQMSTPETPQSPSPS

[SQ](#) GSGSESYKLLPGAITTIVKPPSAIPQPTITKQEI

XX

[SC](#) edited SwissProt #P22449

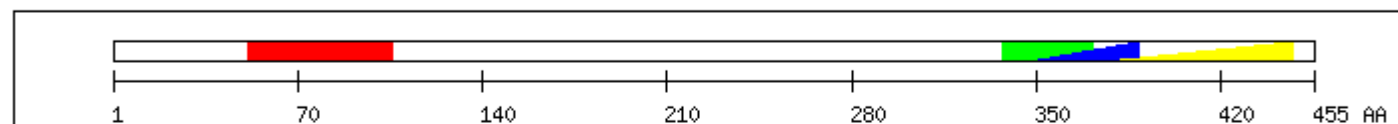
XX

[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 ATTCTTTGACTGAACTCTCACTAAACTCCTCTAAATTATATGTCATATTAACTGGTTAAA
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 Bioinformatics

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

Entry No.:	Accession	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](#)

TFClass –классификация транскрипционных факторов млекопитающих на основе строения их ДНК-связывающих доменов

<http://tfclass.bioinf.med.uni-goettingen.de/index.jsf>

Classification of Transcription Factors in Mammalia

Search in TFClass:

[Go to the search of TRANSFAC](#)

Superclass: ☐, Class: ☐,
Family: ☐, Subfamily: ☐, Genus: ☐,

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

General

Transcription factors, in a somewhat simplified definition, are proteins that regulate transcription by binding to specific sequence elements in regulatory genome regions such as promoters, enhancers etc.

TFClass is a classification of eukaryotic transcription factors based on the characteristics of their DNA-binding domains. It comprises four general levels (superclass, class, family, subfamily) and two levels of instantiation (genus and molecular species). Two of them (subfamily and factor species) are optional. More detailed explanations about the classification scheme and its criteria are given [here](#).

For further information about the latest version, kindly refer to Wingender, E., Schoeps, T., Haubrock, M., Krull, M. and Dönitz, J.: TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 46, D343-D347 (2018).

TFClass: The Schema

Level	Rank denomination	Definition	Example
1	Superclass	General topology of the DBD	Zinc-coordinating DBDs (Superclass 2)
2	Class	Structural blueprint of the DBD	Nuclear receptors with C4 zinc fingers (Class 2.1)
3	Family	Sequence & functional similarity	Thyroid hormone receptor-related factors (NR1) (Family 2.1.2)
4	Subfamily	Sequence-based subgroupings	Retinoic acid receptors (NR1B) (Subfamily 2.1.2.1)
5	Genus	Transcription factor gene	RAR- α (Genus 2.1.2.1.1)
4	Species	TF polypeptide	RAR- α 1 (Species 2.1.2.1.1.1)



Edgar Wingender
Dr. rer. nat.
Director
Universitätsmedizin
Göttingen , Göttingen ·
Institute of Bioinformatics

Wingender E, Schoeps T, Haubrock M, Krull M and Dönitz J.: TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* 46, D343-D347 (2018).

Данные о конкретном белке в базе TFClass

<http://tfclass.bioinf.med.uni-goettingen.de/index.jsf#9606>

Classification of Transcription Factors in Mammalia

Search in TFClass:

[Go to the search of TRANSFAC](#)

Superclass: ☐, Class: ☐,
Family: ☐, Subfamily: ☐, Genus: ☐,

- 1 Basic domains
 - 1.1 Basic leucine zipper factors (bZIP)
 - 1.1.1 Jun-related
 - 1.1.1.1 Jun
 - 1.1.1.1.1 c-Jun**
 - 1.1.1.1.2 JunB
 - 1.1.1.1.3 JunD
 - 1.1.1.2 NFE2
 - 1.1.1.3 ATF2
 - 1.1.2 Fos-related
 - 1.1.3 Maf-related
 - 1.1.4 B-ATF-related
 - 1.1.5 XBP1-related
 - 1.1.6 ATF4-related
 - 1.1.7 CREB-related

all_collections_p....tar.gz ^

[About TFClass](#)

General

Species within this Genus:
28

Species specific reports

Jump to [human](#), [mouse](#) or [rat](#)
Hint: You can rearrange the panels by drag-and-drop. The new arrangement will persist during your session.

Ailuropoda melanoleuca

Synonyms: [giant panda](#)

Outlinks

- Ensembl gene: [ENSAMEG00000000275](#)

Bos taurus

Synonyms: [cattle](#), [domestic cow](#)

Iso

human, man

Synonyms: [human](#), [man](#)

Outlinks

- BioGPS: [ENSG00000177606](#)
- Ensembl gene: [ENSG00000177606](#)
- ProteinAtlas: [ENSG00000177606 \(with antibody\)](#)
- TRANSFAC: [T00133](#)
- UniProt: [P05412](#)

Protein expression pattern: [Show table](#)

Isoforms and DNA binding domain

Ictidomys tridecemlineatus

Synonyms: [thirteen-lined ground squirrel](#)

Outlinks

- Ensembl gene: [ENSSTOG00000028549](#)
- UniProt: [NA](#)

Wingender E, Schoeps T, Haubrock M, Krull M and Dönitz J.: TFClass: expanding the classification of human transcription factors to their mammalian orthologs. Nucleic Acids Res. 46, D343-D347 (2018).

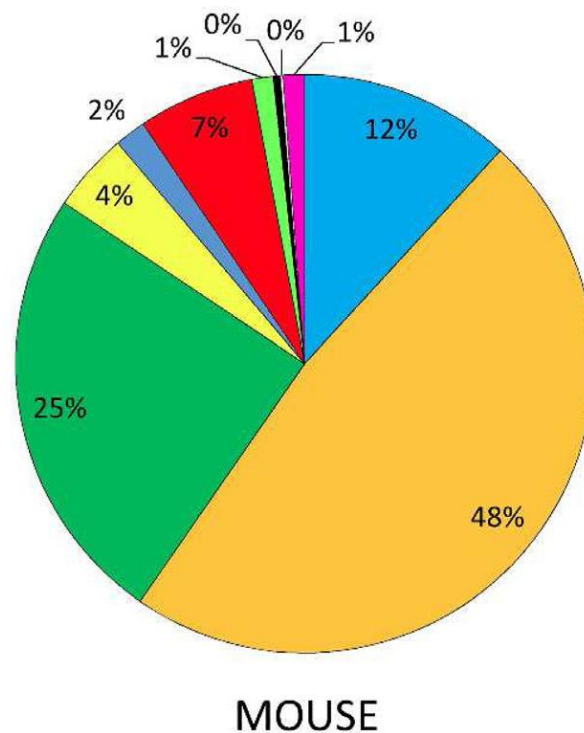
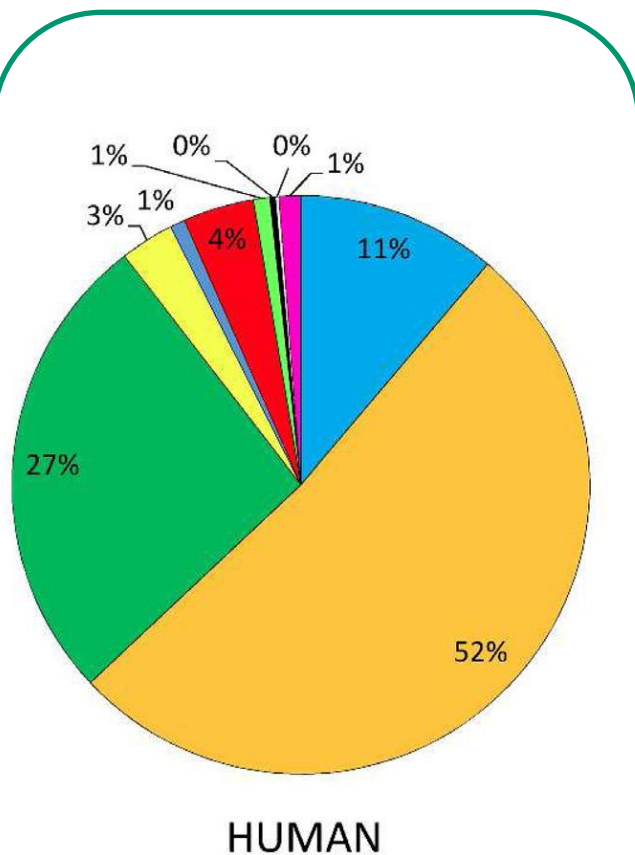
Classification of Human Transcription Factors

- Draft October 05, 2018 -

This is a preliminary draft version, some more rearrangements are about to come!

1 Superclass: Basic domains						TFC			
1.1 Class: Basic leucine zipper factors (bZIP)						TFC			?
1.1.1	Family: Jun-related factors					TGAGTCA			
1.1.1.1	Subfamily: Jun factors					TGAGTCA			
1.1.1.1.1	c-Jun					U			
1.1.1.1.2	JunB					U			
1.1.1.1.3	JunD					U			
1.1.1.2	Subfamily: NF-E2-like factors					GCTGAGTCA			
1.1.1.2.1	NF-E2 p45					ATGAcTCA			
1.1.1.2.2	NF-E2L1 (NRF1)					U			
1.1.1.2.2.1	NF-E2L1, isoform 1					U			
1.1.1.2.2.2	NF-E2L1, isoform 2					U			
1.1.1.2.3	NF-E2L2 (NRF2)					U			
1.1.1.2.3.1	NF-E2L2, isoform 1					U			
1.1.1.2.3.2	NF-E2L2, isoform 2					U			
1.1.1.2.3.3	NF-E2L2, isoform 3					U			
1.1.1.2.4	NF-E2L3 (NRF3)					U			
1.1.1.2.5	BACH1					U			
1.1.1.2.6	BACH2					U			
1.1.1.3	Subfamily: ATF-2-like factors					TGACGTCA			
1.1.1.3.1	ATF-2					U			
1.1.1.3.1.1	ATF-2, isoform 1					U			
1.1.1.3.1.2	ATF-2, isoform 2					U			
1.1.1.3.1.3	ATF-2, isoform 3					U			
1.1.1.3.1.4	ATF-2, isoform 4					U			
1.1.1.3.1.5	ATF-2, isoform 5					U			
1.1.1.3.1.6	ATF-2, isoform 6					U			
1.1.1.3.1.7	ATF-2, isoform 7					U			
1.1.1.3.1.8	ATF-2, isoform 8 (ATF2-small)					U			
1.1.1.3.2	ATF-7					U			
1.1.1.3.2.1	ATF-A, isoform 1					U			
1.1.1.3.2.2	ATF-A δ , isoform 2					U			
1.1.1.3.2.3	ATF-A1, isoform 3					U			
1.1.1.3.2.4	ATF-7A0, isoform 4					U			
1.1.1.3.2.5	ATF-7 (ATF-4), isoform 5					U			
1.1.1.3.2.6	ATF-7, isoform 6					U			
1.1.1.3.3	CREB-5 (CRE-BPa)					U			
1.1.1.3.3.1	CREB-5 α , isoform 1					U			
1.1.1.3.3.2	CREB-5 β , isoform 2					U			
1.1.1.3.3.3	CREB-5 γ , isoform 3					U			
1.1.1.3.3.4	CREB-5 δ , isoform 4					U			

Распределение по суперклассам транскрипционных факторов человека и их мышинных ортологов.



- Basic domain
- Zinc-coordinating domain
- Helix-turn-helix domain
- Other all- α -helical DNA-binding domain
- α -Helices exposed by β -structures
- Immunoglobulin fold
- β -Hairpin exposed by an α/β -scaffold
- β -Sheet binding to DNA
- β -Barrel DNA-binding domain
- Yet undefined DNA-binding domain

1557 генов,
 кодирующих 3268 изоформ
 белков, содержащих ДНК-
 связывающий домен

База данных о белках UniProtKB

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

Swiss Institute of Bioinformatics (Geneva)



UniProtKB ▾

Advanced ▾  Search

BLAST Align Retrieve/ID mapping Peptide search SPARQL 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 (563,082)

 Manually annotated and reviewed.

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

TrEMBL (188,961,949)

 Automatically annotated and not reviewed.

Records that await full manual annotation.

UniRef

Sequence clusters



UniParc

Sequence archive



Proteomes

Proteome sets



Supporting data

Literature citations



Cross-ref. databases



Taxonomy



Diseases

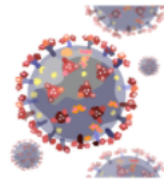
XXX

Subcellular locations



Keywords





New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

View SARS-CoV-2 Proteins and Receptors

News



Forthcoming changes

Planned changes for UniProt

[UniProt release 2020_04](#)
Inflammation: the Good, the Bad and the Ugly | Cross-references to GlyGen and PathwayCommons | New automatic annotation system ARBA (Asso...

[UniProt release 2020_03](#)
Mitochondrial call for help | Cross-references to IDEAL and BioGRID-ORCS | Changes to

 News archive

Getting started

 Text search

Our basic text search allows you to search all the

UniProt data

 Download latest release

Get the UniProt data

Protein spotlight



Either You, Or Me

September 2020



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 2020 году

Имеется ИНФОРМАЦИЯ по ВСЕМ ВИДАМ ОРГАНИЗМОВ. Из них (ручная аннотация)

20 375 входа, описывающих белки человека,

17 046 входа для мыши

8 106 входа для крысы

15 991 - *Arabidopsis thaliana*

6721 - *S. cerevisiae*

UniProtKB



UniProtKB reviewed:yes Advanced Search

BLAST Align Retrieve/ID mapping Peptide search SPARQL Help Contact

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 by:

 **Reviewed (563,082)**
Swiss-Prot

Popular organisms
[Human \(20,375\)](#)
[Mouse \(17,046\)](#)
[A. thaliana \(15,991\)](#)
[Rat \(8,106\)](#)
[S. cerevisiae \(6,721\)](#)
Other organisms
 [Go](#)

[BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#) [Share](#)

1 to 25 of 563,082 Show 25

☐	Entry	Entry name		Protein names	Gene names	Organism	Length	
☐	Q99TG3	ACCA_STAAM		Acetyl-coenzyme A carboxylase carbo...	accA SAV1700	Staphylococcus aureus (strain Mu50 / ATCC 700699)	314	
☐	P49145	5HT1D_RABIT		5-hydroxytryptamine receptor 1D	HTR1D	Oryctolagus cuniculus (Rabbit)	377	
☐	P18271	8513_TRYCR		Sialidase 85-1.3	SA85-1.3	Trypanosoma cruzi	175	
☐	Q3KF15	ACCD_PSEPF		Acetyl-coenzyme A carboxylase carbo...	accD Pfl01_1898	Pseudomonas fluorescens (strain Pf0-1)	306	
☐	P80512	ADH1_NAJNA		Alcohol dehydrogenase 1		Naja naja (Indian cobra)	375	
☐	B3GXY6	AMPA_ACTP7		Probable cytosol aminopeptidase	pepA APP7_1157	Actinobacillus pleuropneumoniae serotype 7 (strain AP76)	499	
☐	Q9AT76	AGL27_ARATH		Agamous-like MADS-box protein AGL27	AGL27 FK1, FLM, MAF1, At1g77080, F22K20.15	Arabidopsis thaliana (Mouse-ear cress)	196	

 all_collections_p....tar.gz ^

[Показать все](#) ×

UniProtKB

В базах **SWISS-PROT** и **TrEMBL** 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 supported by the [National Eye Institute \(NEI\)](#), [National Human Genome Research Institute \(NHGRI\)](#), [National Heart, Lung, and Blood Institute \(NHLBI\)](#), [National Institute on Aging \(NIA\)](#), [National Institute of Allergy and Infectious Diseases \(NIAID\)](#), [National Institute of Diabetes and Digestive and Kidney Diseases \(NIDDK\)](#), [National Institute of General Medical Sciences \(NIGMS\)](#), [National Institute of Mental Health \(NIMH\)](#), and [National Cancer Institute \(NCI\)](#) of the [National Institutes of Health \(NIH\)](#) under grant U24HG007822. Additional support for the EMBL-EBI's involvement in UniProt comes from [European Molecular Biology Laboratory \(EMBL\)](#) core funds, the [British Heart Foundation \(BHF\)](#) (RG/13/5/30112), the [Parkinson's Disease United Kingdom \(PDUK\)](#) GO grant G-1307, the [Alzheimer's Research UK \(ARUK\)](#) grant ARUK-NAS2017A-1, the NIH GO grant U41HG02273 and [Open Targets](#). 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 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).

UniProt activities at SIB have benefited from EC support through the FP6 FELICS project (2006-2009, contract number 021902), and the FP7 projects SLING (2009-2012, contract number 226073), [GEN2PHEN](#) (2009-2013, contract number 200754), and [MICROME](#) (2009-2013, contract number 222886-2).

UniProt activities at PIR have benefited from NIH grants HHSN266200400061C (2004-2009), R01GM080646 (2007-2011), R01GM080646-04S2 (2009-2012), R01GM080646-07S1 (2012), and P20RR016472-09S2 (2009-2011), NSF grants DBI-0138188 (2002-2005) and IIS-0430743 (2004-2007), and UNIDEL Foundation award (2010-2012).

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		

AnimalTFDB3.0

Animal Transcription Factor Database

Home Family Species Search Predict TF Predict TFBS Blast Download Document Citation

HumanTFDB old AnimalTFDB



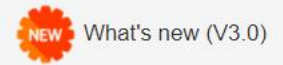
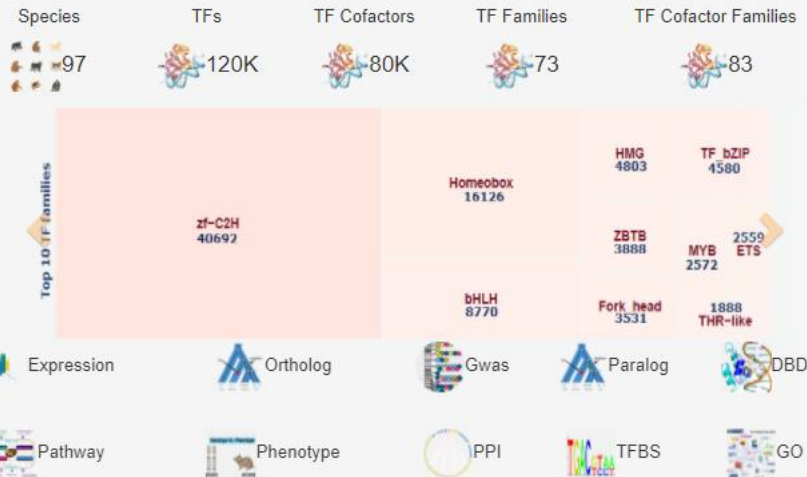
AnimalTFDB is a comprehensive database including classification and annotation of genome-wide transcription factors (TFs), and transcription cofactors in 97 animal genomes. The TFs are further classified into 73 families based on their DNA-binding domain (DBD) and cofactors are classified into 83 families and 6 categories.

You can browse by



For your convenience, we have launched HumanTFDB:

[HumanTFDB portal](#)





(<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 97 видов животных организмов.

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

HumanTFDB

Human Transcription Factor Database

[Home](#) [TF Family](#) [TF Cofactor Family](#) [Search](#) [Predict TF](#) [Predict TFBS](#) [Blast](#) [Download](#) [Document](#) [Citation](#)

[AnimalTFDB3.0](#)



Top 10 TF families

 zf-C2H2 663	 Homeobox 198	 bHLH 109	 TF_bZIP 92	No Image HMG 54
 Fork_head 52	 ZBTB 50	 ETS 49	 MYB 30	 THR-like 27



Welcome to HumanTFDB

TFs	TF Cofactors	TF Families	TF Cofactor Families	
 1665	 1025	 73	 83	
 Expression	 Ortholog	 Gwas	 Paralog	 DBD
 Pathway	 Phenotype	 PPI	 TFBS	 GO

Top 10 TF Cofactor families

Mediator complex 30	GTF 26	Cyclin 21	Lysine demethylase 19
TATA-box 27	Other CRF 23	Tripartite motif 14	Bromodomain 11
		Ring finger protein 13	Histone deacetylase 11

Copyright © Guo Lab, College of Life Science and Technology , HUST, China.

Any comments and suggestions, please [contact us](#)

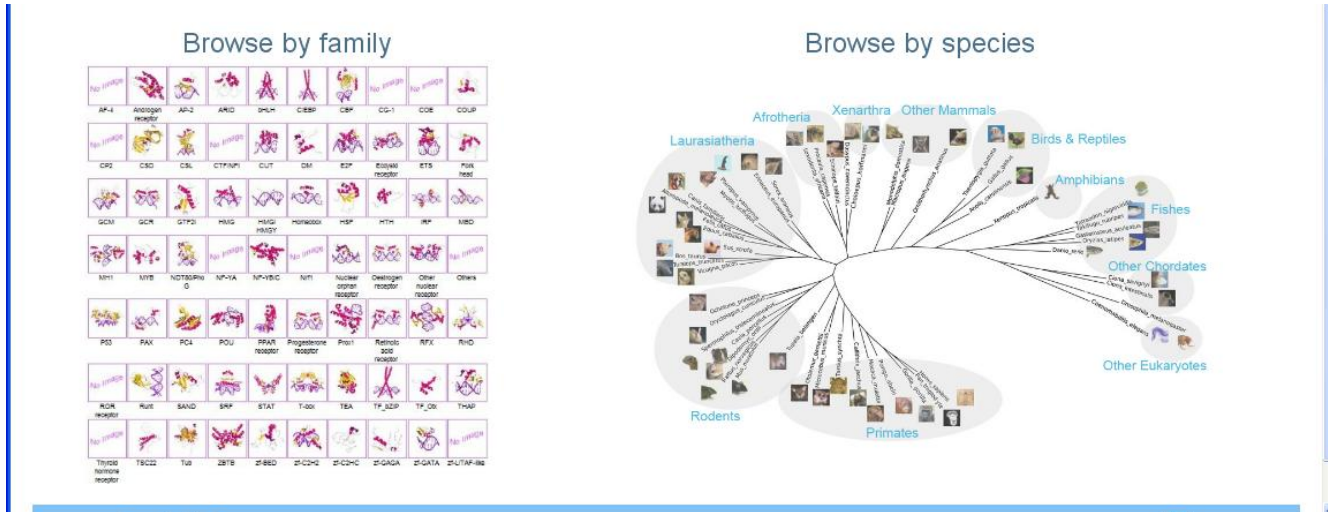
AnimalTFDB – возможности поиска ТФ:

А) по семейству

Б) по виду организма

В) по экспрессии в тканях и клетках

Г) по термину Gene Ontology либо биологическому пути



The figure shows the HumanTFDB website interface. The header includes the text 'HumanTFDB' and 'Human Transcription Factor Database'. The navigation bar contains links: Home, TF Family, TF Cofactor Family, Search, Predict TF, Predict TFBS, Blast, Download, Document, Citation, and AnimalTFDB3.0. The search section is divided into four panels: 'Search by basic information' (with fields for Ensembl Gene ID and ENSG00000072364), 'Search by gene annotation' (with fields for Pathway ID or Description and hsa04931), 'Search by human mRNA expression' (with fields for Source and Threshold), and 'Search by human protein expression' (with fields for Tissue&Cell and Threshold). Each panel has a 'Submit' button.

AnimalTFDB –поиск по ключевому слову (Symbol)



AnimalTFDB

Animal Transcription Factor Database

HOME BROWSE FAMILY BROWSE SPECIES SEARCH DOWNLOAD HELP ABOUT

Quick search

Search by Basic Information

Symbol	PPAR*	reset	search
EnsemblID		reset	search
Gene ID		reset	search
Transcript ID		reset	search
Protein ID		reset	search
Symbol		reset	search
Map location		reset	search
Description		reset	search
Pathway:		reset	search
Ortholog:		reset	search
Paralog:		reset	search
Signature Domain:		reset	search

show records per page:

Species	☐ select all	☐ reverse select	Reset	Search
☐ 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

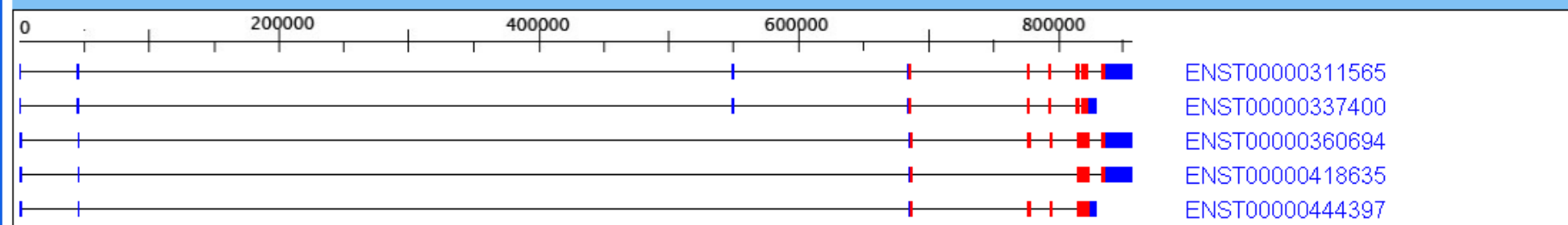
Gene Length : 85621

Gene Position : 6 : 35310335-35395955

Transcripts : There are 5 transcripts

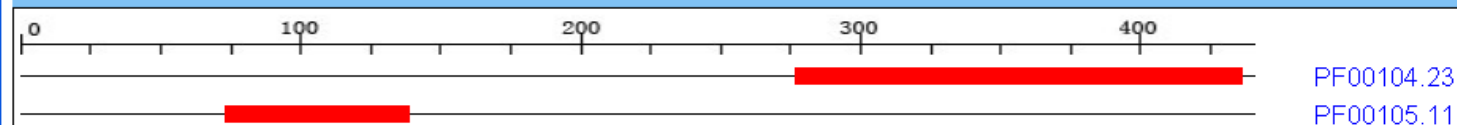
Transcript ID	Length (bp)	Protein ID	Length (aa)
ENST00000311565	3774	ENSP00000310928	441
ENST00000337400	2041	ENSP00000337063	361
ENST00000360694	3747	ENSP00000353916	441
ENST00000418635	3453	ENSP00000413314	343
ENST00000444397	2002	ENSP00000410837	361

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



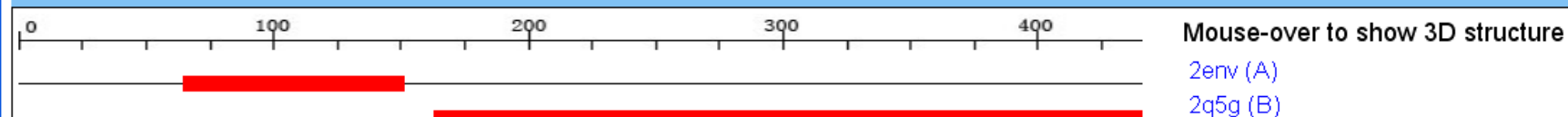
AnimalTFDB – результат поиска, фактор PPARD (продолжение)

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	proteolysis 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
-------------------	----------------	---------	-------------

Новые возможности базы

<http://bioinfo.life.hust.edu.cn/AnimalTFDB/>



What's new

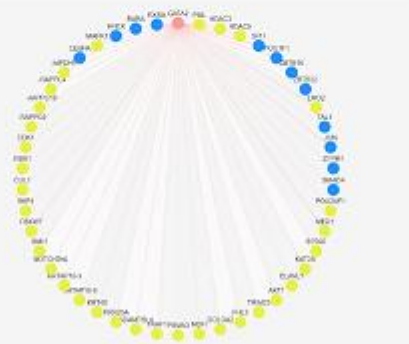
Species



Family



TF related GWAS phenotype



PPI network visualization



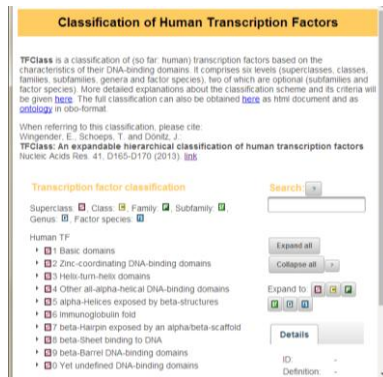
TFBS (Transcription Factor Binding Site)

Современные оценки количества транскрипционных факторов в геноме человека

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

Ресурс TFClass

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



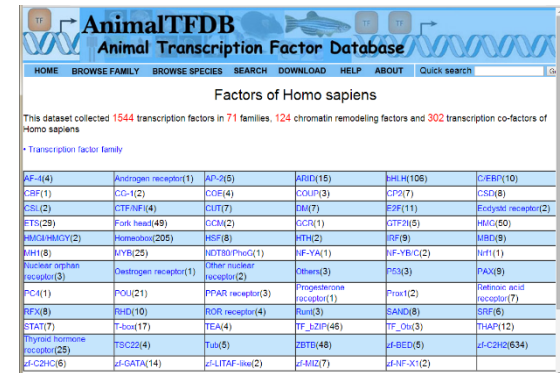
1558 генов,

кодирующих 2904 изоформ белков, содержащих ДНК-связывающий домен.

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

Ресурс AnimalTFDB

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

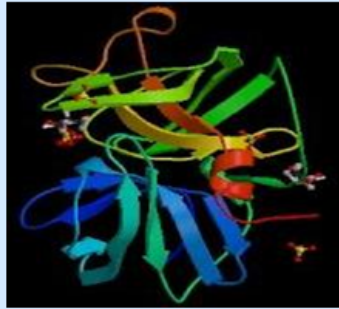


1544 гена,

кодирующих белки, содержащие ДНК-связывающий домен.

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

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



CREMOFAC

A web-database of
Chromatin Remodeling Factors

[Bioinformatics. 2006; 22: 2940-2944](#)

Enter

Number of Visits

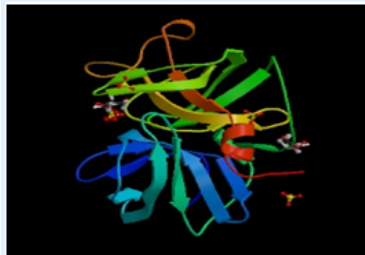
☒ Halloween Costumes

**Developed and maintained by :
Shipra Agrawal*, Chetan Kumar* and M.R.S. Rao**

(* - Equally contributed to this work)



**Chromatin Biology Lab
Molecular Biology and Genetics Unit
Jawaharlal Nehru Centre for Advanced Scientific Research**



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

CREMOFAC

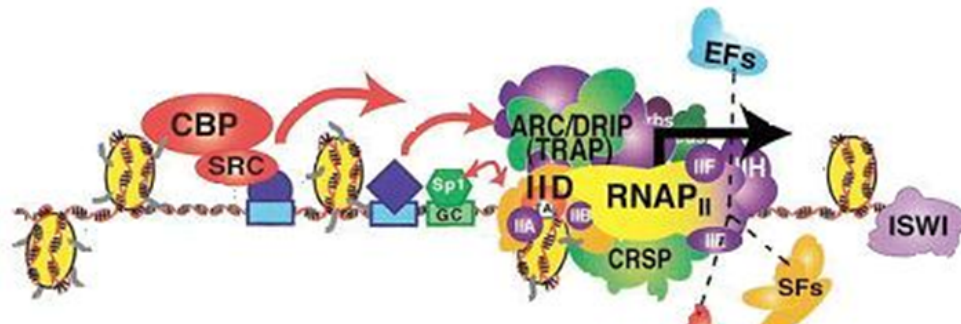
Chromatin Remodeling Factors

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

Introduction

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

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



Данные собраны вручную на основе аннотирования научных публикаций

CREMOFAC Data Statistics

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

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

List of organisms in which remodeling factors have been reported.

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

Homo sapiens

64

Всего 49 видов организмов

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
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



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.

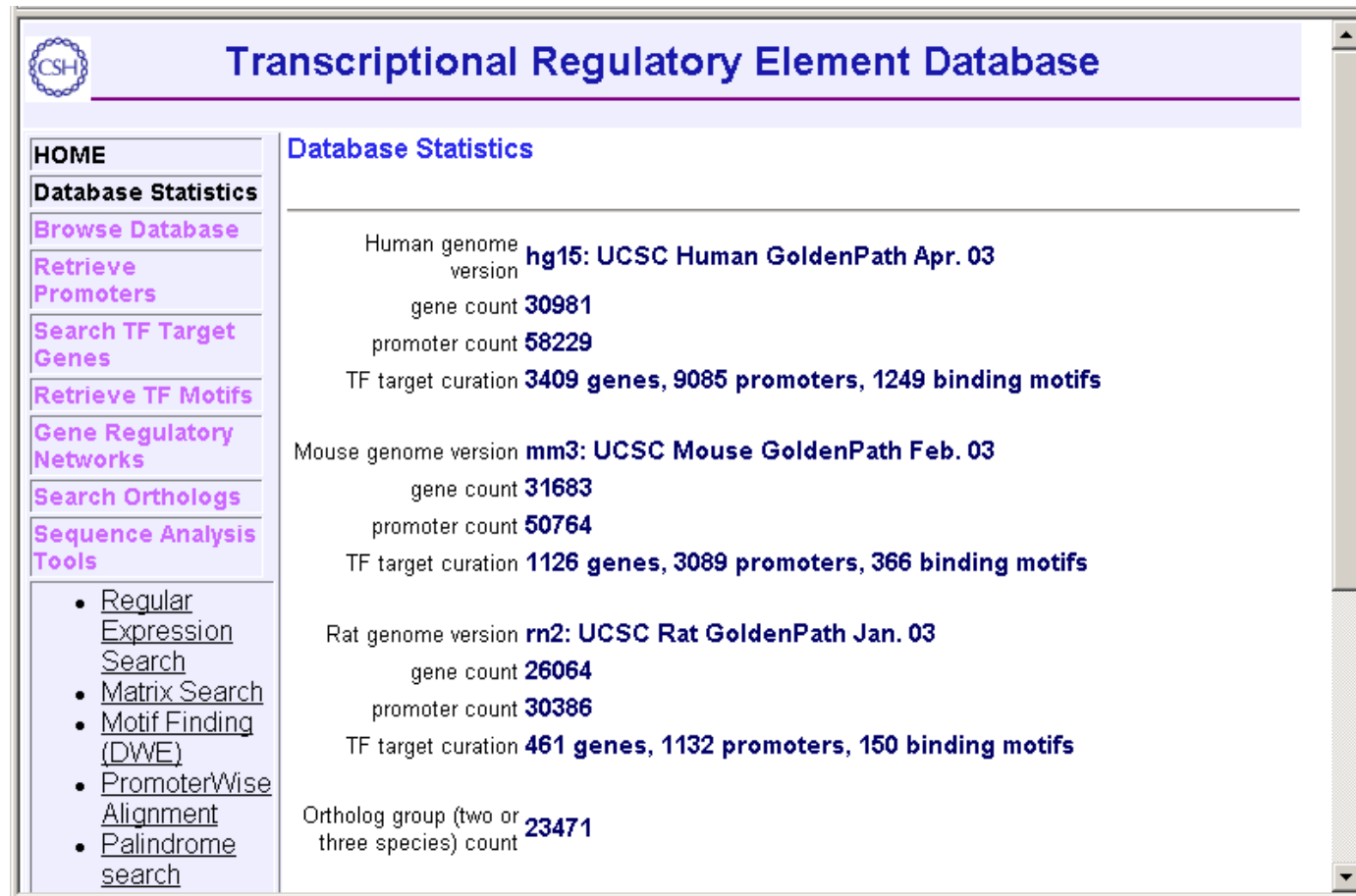
Curation has also been done for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. Binding motifs are mapped on promoters of the corresponding genes and binding quality levels are assigned based on definitiveness of the binding evidence. Curation 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.

Jiang C, Xuan Z, Zhao F, Zhang MQ TRED: a transcriptional regulatory element database, new entries and other development. Nucleic Acids Res. 2007;35(Database issue):D137-40..

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](#)

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.

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

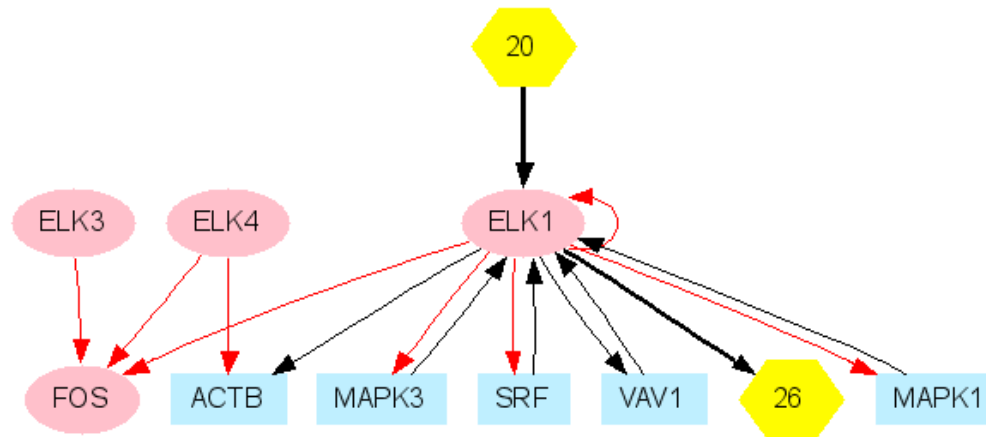
Useful Links

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

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

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

Release 10.1 Updated (December 21, 2017)
Based on UCSC hg38, mm10

We recommend to use Edge (V40 above), Google Chrome (V61 above) or Firefox (V56 above) for the DBTSS browsing. Internet Explorer has not been supported.

Top |

Database Search

Species:
H. sapiens

Keyword:
keyword (e.g. EGFR, NM_*)

Search

Genome browser:
hg38

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

Search

Human Chromatin Features

Search from Genomic Position (hg38):
chr1:75,787,000

Search

Search from SNP (dbSNP rsID):
rs375229869

Search

Search from SNV (COSMIC: somatic mutation):
BRAF

About this database

DBTSS; DataBase of Transcriptional Start Sites.

To support transcriptional regulation studies, we have constructed the DBTSS, 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 DBTSS 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.

Taken together, we provide a unique data resource to investigate what genomic features are observed in a particular genomic coordinates in a wide variety of samples. We believe new DBTSS is helpful to understand biological consequences of the massively identified TSSs and identify human genetic valuations which are associated with disordered transcriptional regulations.

Funding:

[DBTSS](#) is supported by JSPS under the grant ID 221S0002

References

[Ayako Suzuki, Shin Kawano, Toutai Mitsuyama, Mikita Suyama, Yae Kanai, Katsuhiko Shirahige, Hiroyuki Sasaki, Katsushi Tokunaga, Katsuya Tsuchihara, Sumio Sugano, Kenta Nakai, Yutaka Suzuki. DBTSS/DBKERO for integrated analysis of transcriptional regulation. *Nucleic Acids Res.* 2018 \(Database issue\).](#)

[Suzuki A, Mimaki S, Yamane Y, Kawase A, Matsushima K, Suzuki M, Goto K, Sugano S, Esumi H, Suzuki Y, Tsuchihara K. Identification and characterization of cancer mutations in Japanese lung adenocarcinoma without sequencing of normal tissue counterparts. *PLoS One.* 2013 Sep 12;8\(9\).](#)

[Yamashita R, Sathira NP, Kanai A, Tanimoto K, Arauchi T, Tanaka Y, Hashimoto S, Sugano S, Nakai K, Suzuki Y. \(2011\) Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis.](#)

Tweets by @dbkero

KERO @dbkero

DBKERO (疾患マデルミクスデータベース) の利用者アンケート調査を行っています。以下から回答にご協力をよろしくお願い申し上げます。 research.nttcoms.com/r/1803as16QT/1

回答期限: 4月18日 (水) 10:00まで

KERO @dbkero

2017/12/6(水)-9(土)、ConBio2017のBioDB コーナーブースへの出展と、生命科学のデータベース活用法のフォーラム発表を行いました。ブースや発表を聞きに来て下さった方、どうもありがとうございました。

KERO @dbkero

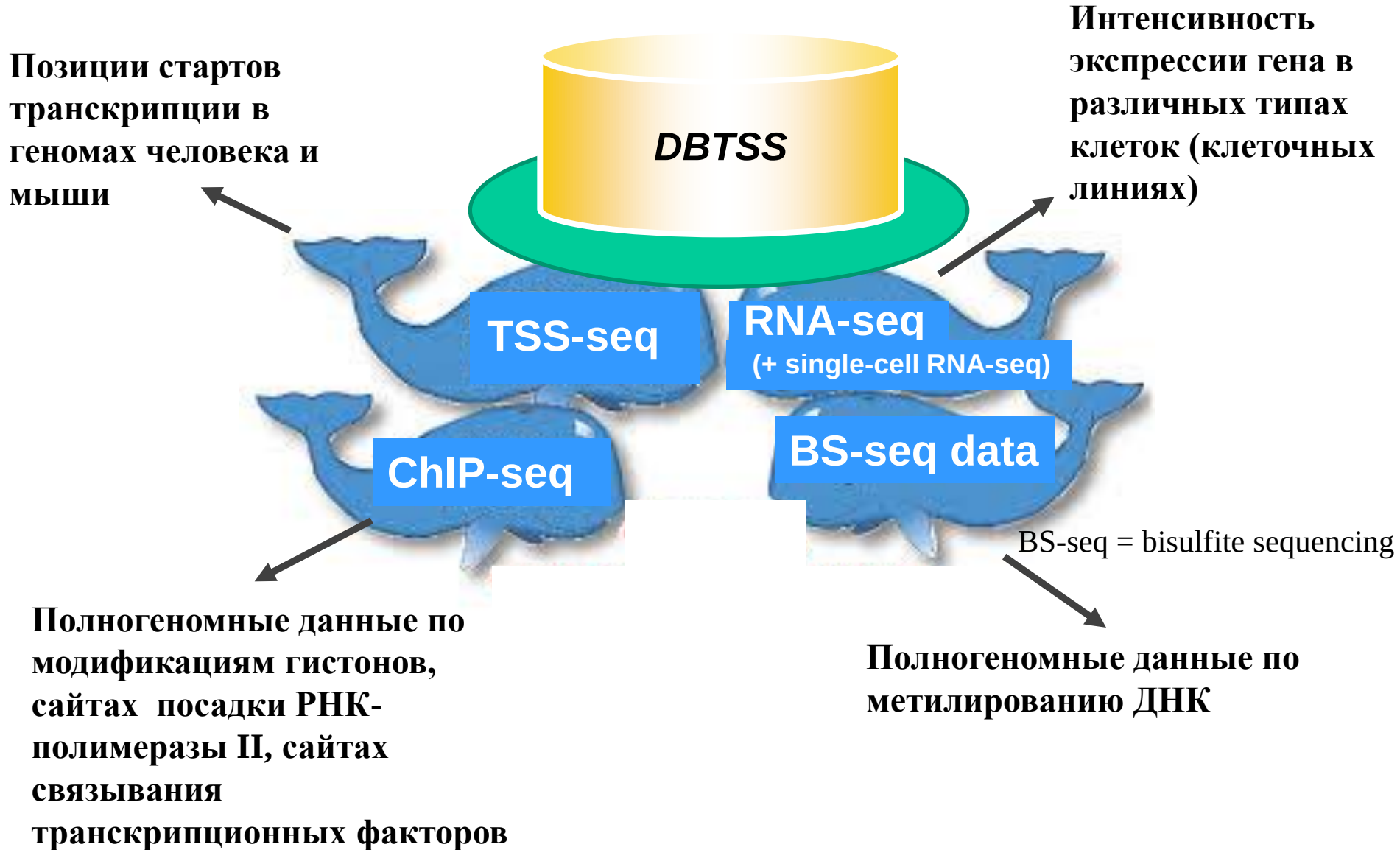
こんにちは。新しいサイトの

TSSs =Transcriptional Start Sites

Точные позиции TSS в геномах человека, мыши и еще 5 видов организмов
База развивается с 2002г.

В 2017 г был организован гибридный информационный ресурс совместно с DBKERO (Database of Kashiwa Encyclopedia for human genome mutations in Regulatory regions and their Omics contexts)

DBTSS (Release 1.0) – экспериментальные методики получения данных



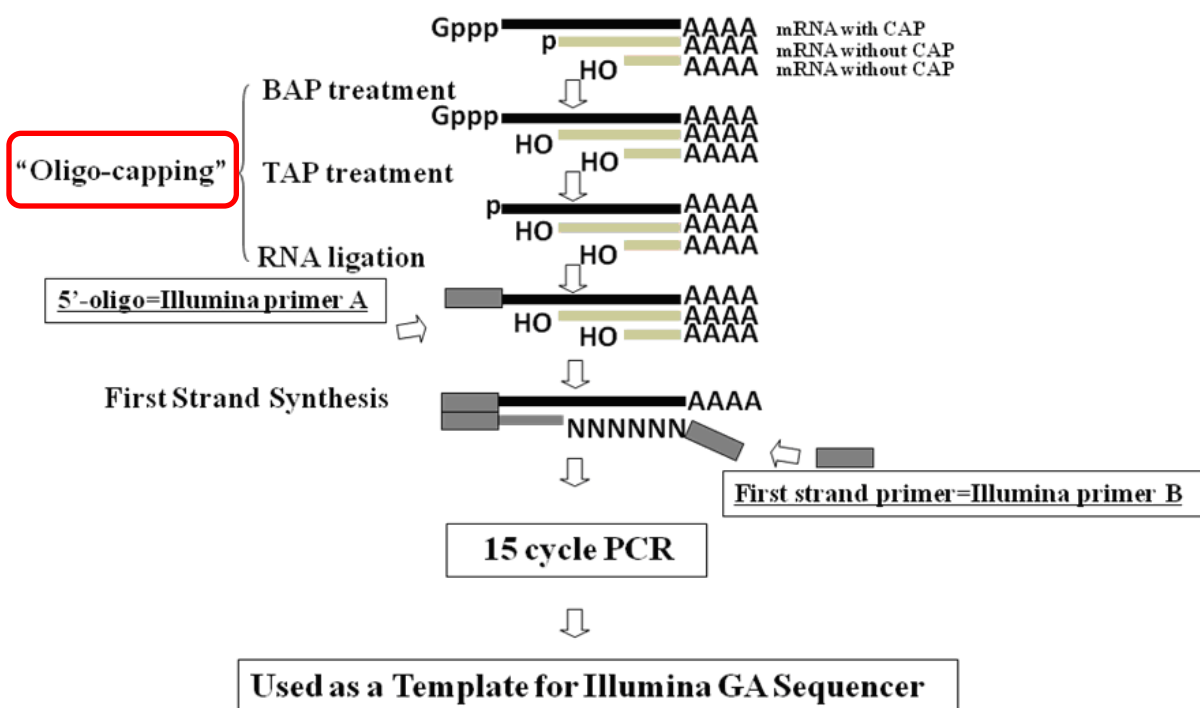
Методика TSS-seq

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:

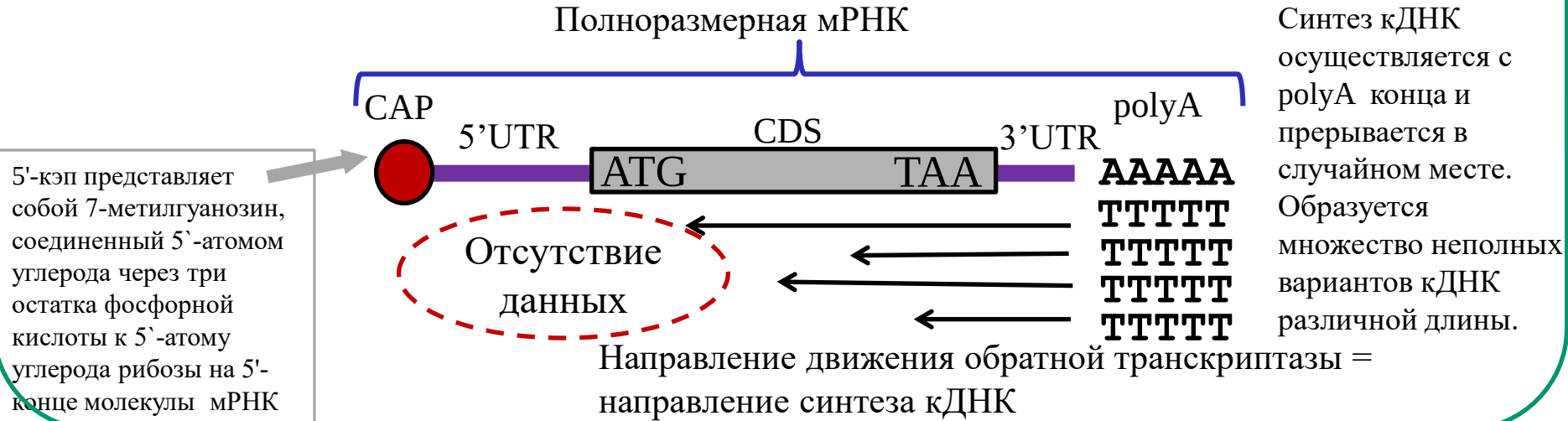
Адапторы (олигонуклеотиды), которые используются в процессе секвенирования на платформе Illumina, пришиваются к кэп-сайту транскриптов. Процесс включает три этапа , на которых используются три разных фермента. Это позволяет пришивать адапторы только к полноразмерным транскриптам (то есть транскриптам,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



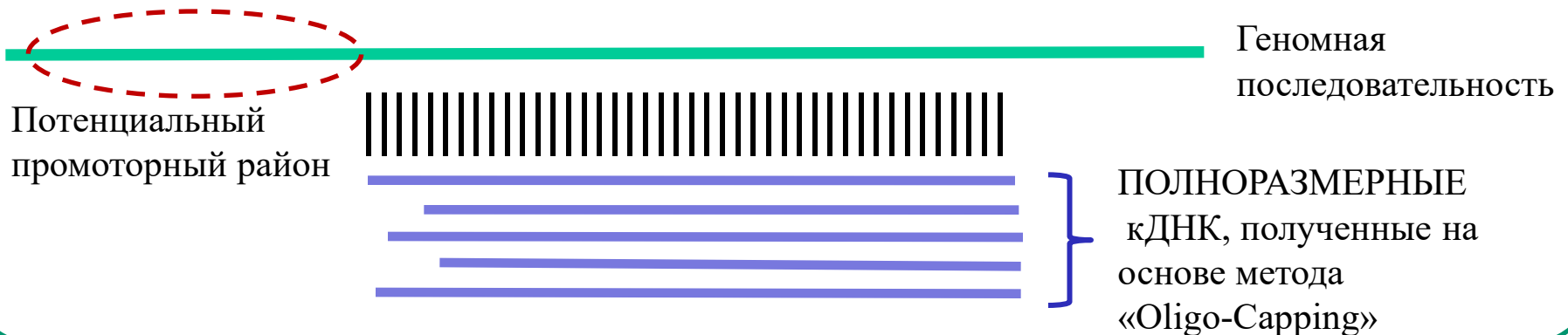
Преимущество «Oligo-Capping method»:

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.

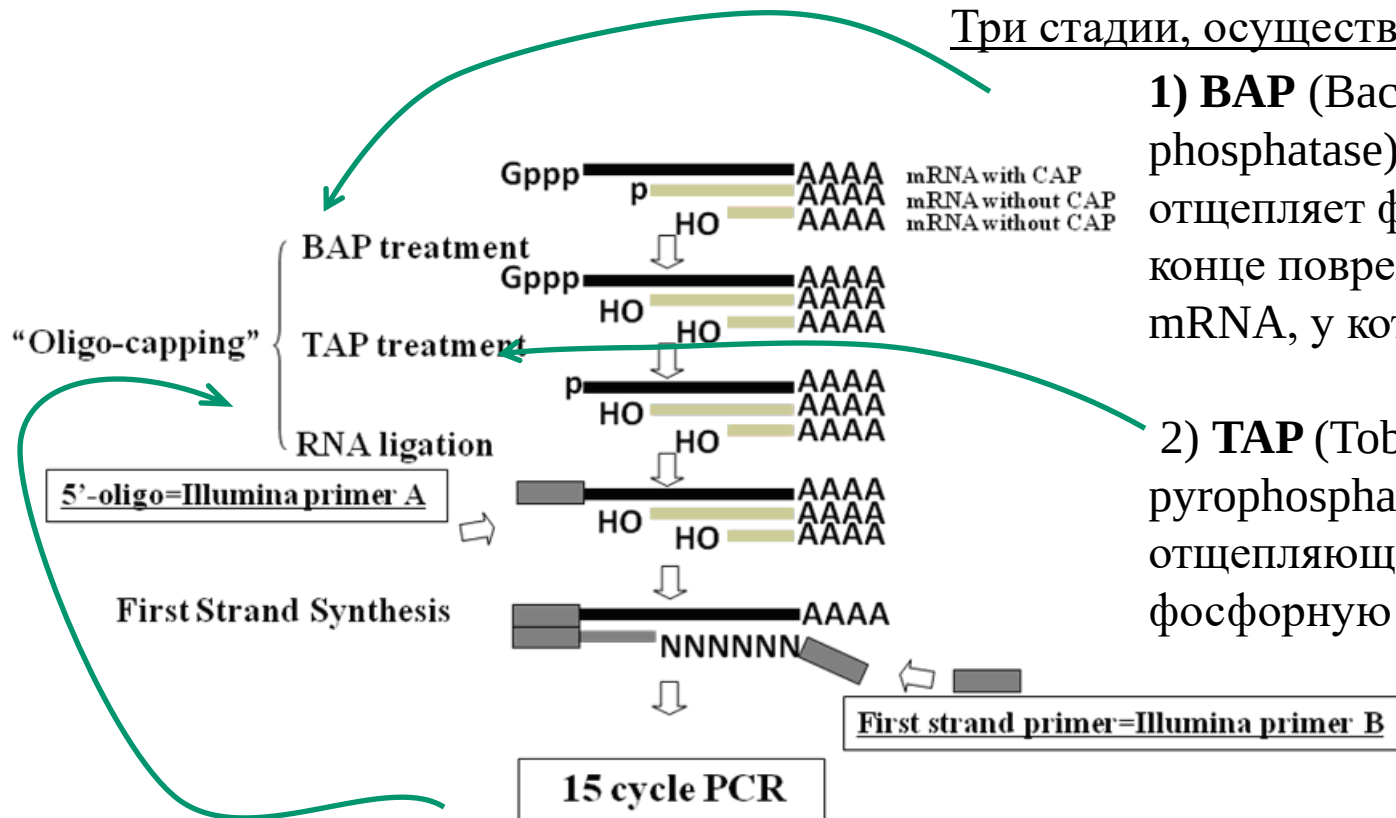
* 5'-кэп представляет собой 7-метилгуанозин, соединенный 5'-атомом углерода через три остатка фосфорной кислоты к 5'-атому углерода рибозы на 5'-конце молекулы мРНК

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



Регулярное обновление версий базы DBTSS

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


Release 10.1 Updated (December 21, 2017)
Based on UCSC hg38, mm10

We recommended to use Edge (V40 above), Google Chrome (V61 above) or Firefox (V56 above) for the DBTSS browsing. Internet Explorer has not been supported.

Top |

Database Search

Species:

Keyword:

Genome browser:

Human Chromatin Features

Search from Genomic Position (hg38):

Previous version

Previous version

[DBTSS Release 8.0](#)

[DBTSS Release 9.0](#)

genome browser © 2018 [KERO-browser](#)

科研費
KAKENHI

DBTSS is financially supported with a Grant-in-Aid for Publication of Scientific Research Results (Databases) by Japan Society for the Promotion of Science, a Grant-in-aid for Scientific Research on Innovative Areas 'Genome Science' [221S0002] from the Ministry of Education, Culture, Sports, Science and Technology of Japan

Copyright © 2001-2018 Yutaka Suzuki, the University of Tokyo, Japan. All rights reserved.
No reproduction or republication without permission.

Доступ к более ранним версиям:

Каталог FTP /pub/hgc/db/dbtss/ на ftp.hgc.jp

[На один уровень вверх](#)

09/20/2005 12:00	Каталог Yamashita NAR
04/07/2008 12:00	Каталог Yamashita et al
07/08/2005 12:00	1,346 annotation
07/08/2005 12:00	4,358 data
03/14/2018 12:00	11 dbtss recently
10/08/2017 12:00	Каталог dbtss ver10
05/06/2005 12:00	Каталог dbtss ver4
05/25/2005 12:00	Каталог dbtss ver5
12/10/2014 12:00	13 dbtss ver5 1
09/07/2006 12:00	Каталог dbtss ver5 2
09/15/2009 12:00	Каталог dbtss ver6
07/20/2010 12:00	Каталог dbtss ver7
07/12/2012 12:00	Каталог dbtss ver8
09/15/2014 12:00	Каталог dbtss ver9
05/06/2005 12:00	Каталог old
05/14/2015 12:00	Каталог other
04/24/2017 12:00	4,961,386,706 vcf.tgz

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

DBTSS (Release 9.0) – информационное содержание

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](#)

Позиции стартов транскрипции

Интенсивность экспрессии генов

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

Метилирование ДНК

Типы хроматина

Замены нуклеотидов

DBTSS (Release 10.1) – информационное содержание

[to top](#)

Inhouse data			Data contents
Type	Project	Number of samples	detail information (stats and registered Acc#)
human lung adenocarcinoma cell line 26 cell line	WGS	26	WGS_detail
	RNA-seq	27	RNA_detail
	TSS-seq	27	TSS_detail
	ChIP-seq	27	ChIP_detail
	Bisulfite-seq	26	BS_detail
	single cell RNA-seq (beads-seq)	5	Beads_scRNA_detail
	single cell RNA-seq (10x chromium)	5	Chromium_scRNA_detail
	single cell RNA-seq (C1)	3	C1_scRNA_detail
	long read (10x gemcode)	24	10x_gemcode_linked_read_detail
	long read (MinION WGS)	4	MinION_WGS_detail
	Drug pertubation compound 23 RNA-seq	5	Drug23_RNA_detail
	Drug pertubation compound 23 ATAC-seq	5	Drug23_ATAC_detail
	Drug pertubation compound 95 RNA-seq	23	Drug95_RNA_detail
	Drug pertubation compound 95 ATAC-seq	23	Drug95_ATAC_detail
human cell line or tissue	RNA-seq	2	human_cell_tissue_detail
	TSS-seq	43	
	ChIP-seq	10	
mouse normal cell line or tissue	RNA-seq	2	mouse_cell_detail
	TSS-seq	5	
	ChIP-seq	2	
MEXT KAKENHI Genome Support (collaborator's data)	RNA-seq	58	genomesupport_data_detail
	BS-seq	11	
	ChIP-seq	104	
	Other	34	

20 типов тканей
и 7 линий клеток

5 линий клеток

[To top](#)

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.

Suzuki A, et al., DBTSS/DBKERO for integrated analysis of transcriptional regulation. Nucleic Acids Res. 2018 Jan 4;46(D1):D229-D238.

DBTSS – возможности поиска данных

- DataBase of Transcriptional Start Sites -
DBTSS

We recommend to use Edge (V40 above), Google Chrome (V61 above) or Firefox (V56 above)

Top |

Database Search

Species:

H. sapiens ▼

Keyword:

keyword (e.g. EGFR, NM_*)

Search

Genome browser:

hg38 ▼

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

Search

About this database

DBTSS; DataBase of Transcriptional Start Sites.

To support transcriptional regulation studies, we have constructed the DBTSS transcriptional start sites (TSSs) in the genome based on our unique experimental TSS-seq. This database includes DBTSS data of a major part of human adult now contains 491 million TSS tag sequences for collected from a total of 20 our newly generated RNA-seq data of subcellular- fractionated RNAs and C polymerase II and several transcriptional regulatory factors in cultured cell line external epigenomic data, such as chromatin map of the ENCODE project.

Taken together, we provide a unique data resource to investigate what genes genomic coordinates in a wide variety of samples. We believe new DBTSS of the massively identified TSSs and identify human genetic variations which regulations.

Funding:

[DBTSS](#) is supported by JSPS under the grant ID 221S0002

Pathway Map

Species:

H. sapiens ▼

Search

DBTSS Home



dbtss.hgc.jp

Human Chromatin Features

Search from Genomic Position (hg38):

chr1:75,787,000

Search

Search from SNP (dbSNP rsID):

rs375229869

Search

Search from SNV (COSMIC: somatic mutation):

BRAF

Search

Search from SNV-enriched Gene in Cancers:

Lung adenocarcinoma ... ▼

Search

SNV Summary in Cancers:

NM_*

Search

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

Cytosine 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

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

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

- Пересылки 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>)

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.

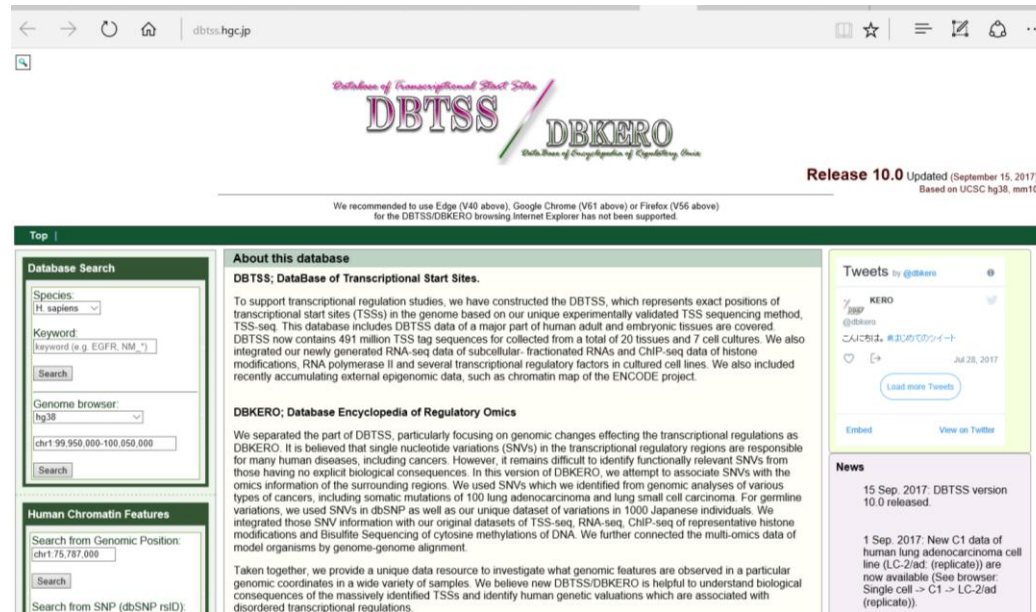
Методы TSS-seq, RNA-seq, ChIP-seq and BS-seq data.

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.

DBTSS, релиз 10

«Июминка» -2017 года !!!!!:



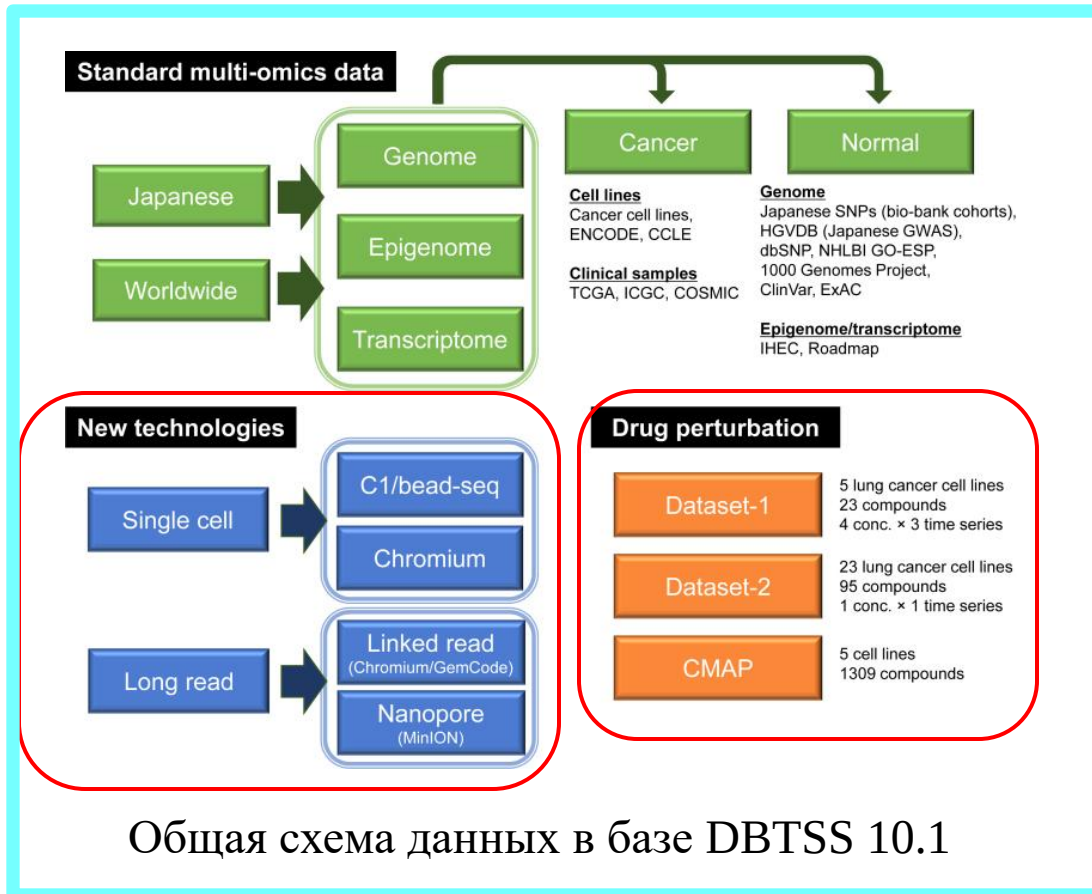
DBKERO – часть ресурса, включающая данные по SNV (single nucleotide variation). Данные о позициях SNV экстрагированы из dbSNP, а также взяты оригинальные данные разработчиков базы , выявленные у 1000 японцев, в также в 100 линиях клеток клеток рака легких.

Данные по SNV интегрированы с другими оригинальными данными , полученными методами TSS-seq, RNA-seq, ChIP-seq (модификации гистонов) и Bisulfite Sequencing (метилирование цитозина в составе ДНК).

Выполнено выравнивание генома человека на геномы модельных видов, за счет чего есть возможность получать информацию о свойствах соответствующего района генома у другого вида.

DBTSS, релиз 10.1

«Июминка» -2018 года !!!!!:



В релиз 10.1 включены новые типы данных

- о влиянии лекарств на клеточные линии (преимущественно раковые)

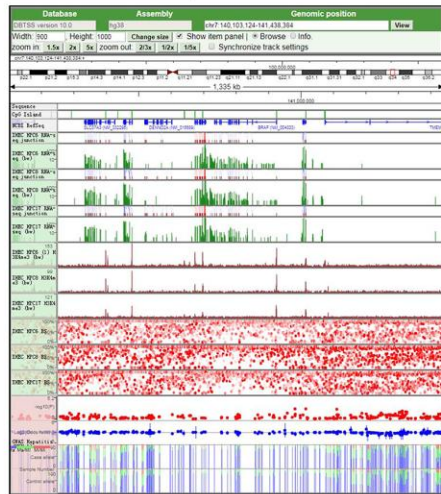
Statistics of the drug perturbation datasets

	Dataset-1	Dataset-2
Sample	5 lung cancer cell lines	23 lung cancer cell lines
Number of compounds	23 + DMSO control	95 + DMSO control
Condition	4 concentration points; 24, 48, 72 h	1 concentration point; 24h
Total number of datasets (RNA-seq)	1299	2011
Total number of datasets (ATAC-seq)	1316	2077

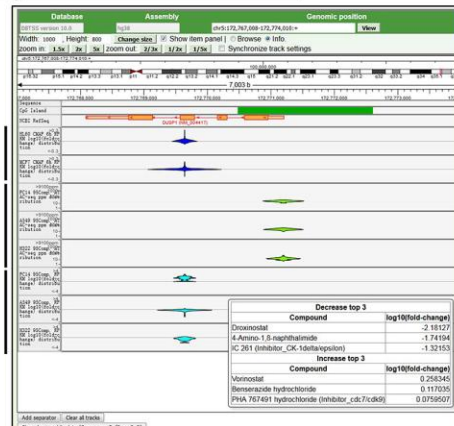
- данные прочитаны на основе анализа одной клетки (single cell thechnology)
- данные получены с помощью технологий, позволяющих читать длинные ряды (Chromium/GemCode (whole-exome & regulatory); Nanopore/MinION (whole-genome))

DBTSS, релиз 10.1 - интерфейсы

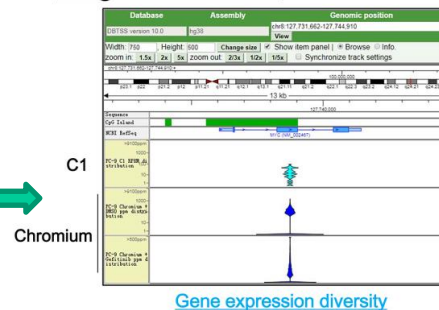
A Genome Browser



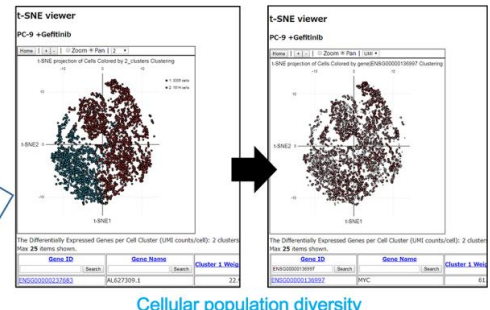
C Drug Perturbation viewer



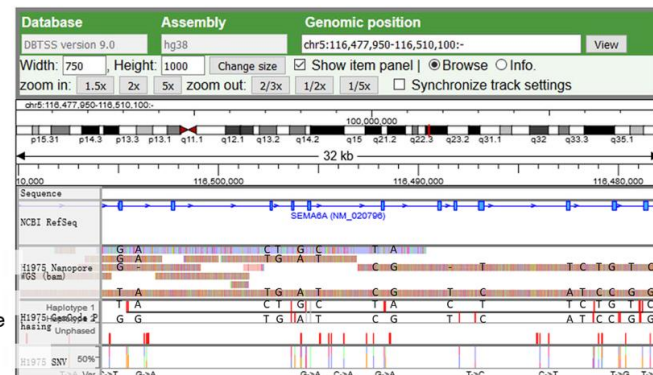
D Single Cell Viewer



link



E Long Read Viewer



Haplotype 1
Haplotype 2

Распределение
клеток по уровню
экспрессии
конкретного гена
(Myc)



Suzuki A, et al., DBTSS/DBKERO for integrated analysis of transcriptional regulation. Nucleic Acids Res. 2018 Jan 4;46(D1):D229-D238.



DBTSS – «идеальная» база данных ??

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

DBTSS – не идеальная база данных ??



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

Конец 5-ой лекции