

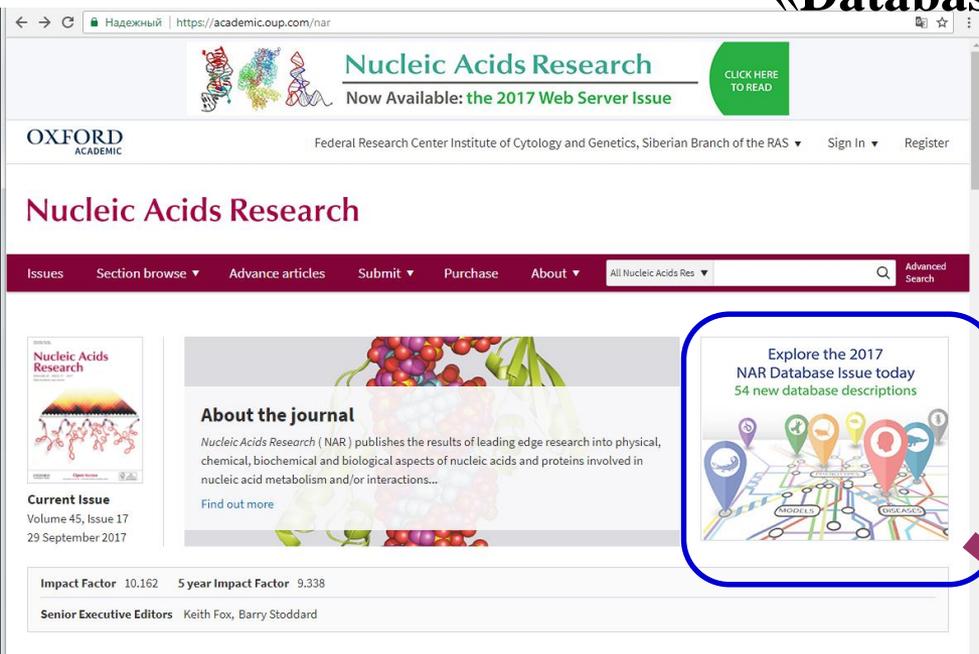
Лекция №5

**БАЗЫ ДАННЫХ ПО
РЕГУЛЯЦИИ
ТРАНСКРИПЦИИ**

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

Журнал NAR (<http://nar.oxfordjournals.org/>) ежегодно публикует информацию о базах данных в специальном выпуске

«Database issue»



Надежный | <https://academic.oup.com/nar>

Nucleic Acids Research
Now Available: the 2017 Web Server Issue

CLICK HERE TO READ

OXFORD ACADEMIC
Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the RAS | Sign In | Register

Nucleic Acids Research

Issues Section browse Advance articles Submit Purchase About All Nucleic Acids Res Advanced Search

About the journal
Nucleic Acids Research (NAR) publishes the results of leading edge research into physical, chemical, biochemical and biological aspects of nucleic acids and proteins involved in nucleic acid metabolism and/or interactions...
[Find out more](#)

Current Issue
Volume 45, Issue 17
29 September 2017

Impact Factor 10.162 5 year Impact Factor 9.338
Senior Executive Editors Keith Fox, Barry Stoddard

Explore the 2017 NAR Database Issue today
54 new database descriptions

Advance articles Submit Purchase About All Nucleic Acids Res Advanced Search

Volume 45, Issue D1, January 2017 [Browse by volume](#)

Database issue

NUCLEIC ACID SEQUENCE, STRUCTURE, AND REGULATION

The 24th annual *Nucleic Acids Research* database issue: a look back and upcoming changes

Michael Y. Galperin; Xosé M. Fernández-Suárez; Daniel J. Rigden

Nucleic Acids Research, Volume 45, Issue D1, 4 January 2017, Pages D1–D11,

<https://doi.org/10.1093/nar/gkw1188>

[Abstract](#) [View article](#)

Database Resources of the National Center for Biotechnology Information

Ссылка на «online Database Collection»

Журнал NAR: online Database Collection

<http://www.oxfordjournals.org/nar/database/a/>

OXFORD ACADEMIC Journals

You are here: [NAR Journal Home](#) » Database Summary Paper Alpha List

NAR Database Summary Paper Alphabetic List

1 2 3 4 5 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

1000 Genomes Selection Browser Engelken, Johannes; Pybus, Marc; Dall'Olio, Giovanni; Luisi, Pierre; Uzkudun, Mani Carreño-Torres, Angel; Pavlidis, Pavlos; Laayouni, Hafid; Bertranpetit, Jaume
Signature of selection in the human genomes
[database](#) [summary](#)

16S and 23S Ribosomal RNA Mutation Database Triman K.L.
16S and 23S ribosomal RNA mutations
[database](#) [summary](#)

2D-PAGE Pleissner, K.-P., Eifert, T., Buettner, S., Knipper, J., Schmelzer, P., Stein, R., Schmidt, F., Mattow, J., Zimny-Arndt, U., Schmid, M., Jungblut, P.R.
Proteome database system for microbial research
[database](#) [summary](#)

2P2Idb Basse, M.J., Betzi, S., Bourgeas, R., Bouzidi, S., Chetrit, B., Hamon, V., Morelli, X., and Roche, P.
2P2Idb - database dedicated to the modulation of protein-protein interactions
[database](#) [summary](#)

3D rRNA modification maps
Locations of modified rRNA nucleotides within the 3D structure of the ribosome

Полный список баз данных,
сгруппированных по категориям .

<http://www.oxfordjournals.org/nar/database/c/>

OXFORD ACADEMIC Journals

You are here: [NAR Journal Home](#) » Database Summary Paper Categories

NAR Database Summary Paper Category List

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

Nucleotide Sequence Databases
RNA sequence databases
Protein sequence databases
Structure Databases
Genomics Databases (non-vertebrate)
Metabolic and Signaling Pathways
Human and other Vertebrate Genomes
Human Genes and Diseases
Microarray Data and other Gene Expression Databases
Proteomics Resources
Other Molecular Biology Databases
Organelle databases
Plant databases
Immunological databases
Cell biology

Информация о базах данных по регуляции транскрипции в «2017 NAR Database Summary Paper Category List»:

<http://www.oxfordjournals.org/nar/database/a/>

Поиск по ключевому слову «virus»

OXFORD
ACADEMIC | Journals

You are here: [NAR Journal Home](#) » Database Summary Paper Search

NAR Database Summary Paper Search

No Search Term Entered

Please enter a 'Search Term', Select criteria from 'Search In' and Click on 'Search'.

Search Term

Search In Title Author Affiliation Paper References All

Search

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

OXFORD
ACADEMIC | Journals

You are here: [NAR Journal Home](#) » Database Summary Paper Search

NAR Database Summary Paper Search

HERVd - Human Endogenous Retrovirus database
Human endogenous retrovirus database
[database](#) [summary](#)

Poxvirus.org
Poxvirus genomic sequences and gene annotation
[database](#) [summary](#)

IVDB - Influenza Virus Database Chang S.^{1,2}, Zhang J.¹, Liao X.¹, Zhu X.^{1,3}, Wang D and Wang J.¹
Annotated sequences and geographic distribution of the influenza virus
[database](#) [summary](#)

Virus Mentha Calderone, Alberto; Licata, Luana; Cesareni, Gianni
Virus-host protein interactions
[database](#) [summary](#)

Papillomavirus Episteme Alison McBride
A database of Papillomaviridae family of viruses
[database](#) [summary](#)

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	Protein knowledgebase	Geneva,
Switzerland		
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>

gene-regulation.com sponsored by geneXplain

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](#)

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 us on FACEBOOK for special promotions, news and events!



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

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 MDMADYSAALDPAYTTLEFENVQVLTMGNDTSPSEGANLNSSNSLGVSAALCAICGDRATG
SQ KHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVDKDKRNQCRYCRLKKCFRAGMKKEA
SQ VQNERDRISTRSSYEDSSLPSINALLQAEVLSQQITSPISGINGDIRAKRIASITDVCE
SQ SMKEQLLVLEWAKYIPAFCELLLDDQVALLRAHAGEHLLLGATKRSMVFKDVLLLGNDY
SQ IVPRHCPELAEMSRVSIIRILDELVLPFQELQIDDNEYACLKAIIFFDPAKGLSDPGKIK

Internet

ОПИСАНИЕ ТРАНСКРИПЦИОННОГО ФАКТОРА HNF4 В базе TRANSFAC (продолжение)

```
SO LQEMLLGGSASDAPHAAHHPLHPLHMQEHMGTNVIVANTMPSHLSNGQMSTPETPQPSPPS
SO GSGSESYKLLPGAITITIVKPPSAIPOPTITKQEAI
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 TFIIIB,
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 TFIIIB; 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.
```

Internet

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 AGCTTAATGTATATAATCTTTTAGAGGTA AAAATCTACAGCCAGCAAAAAGTCATGGTAAAT
SQ ATTCTTTGACTGAACTCTCACTAAAACCTCCTCTAAAATTATATGTCATATTAAC TGGTTAAA
SQ TTAATATAAAATTTGTGACATGACCTTAACTGGTTAGGTAGGATATTTTTCTTCATGC AAA
SQ AATATGACTAATAAATAATTTAGCACAAAAATATTTCCCAATACTTTAATTCTGTGATAGA
SQ AAAATGTTTAACTCAGCTACTATAATCCCATAAATTTTGAAAACTATTTATTAGCTTTTGT
SQ GTTTGACCCTTCCCTGCCAAAAGGCAACTATTTAAAGGACCCTTTAAAACTCTTGAAACTAC
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](#)



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 Biob

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-...



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](#)

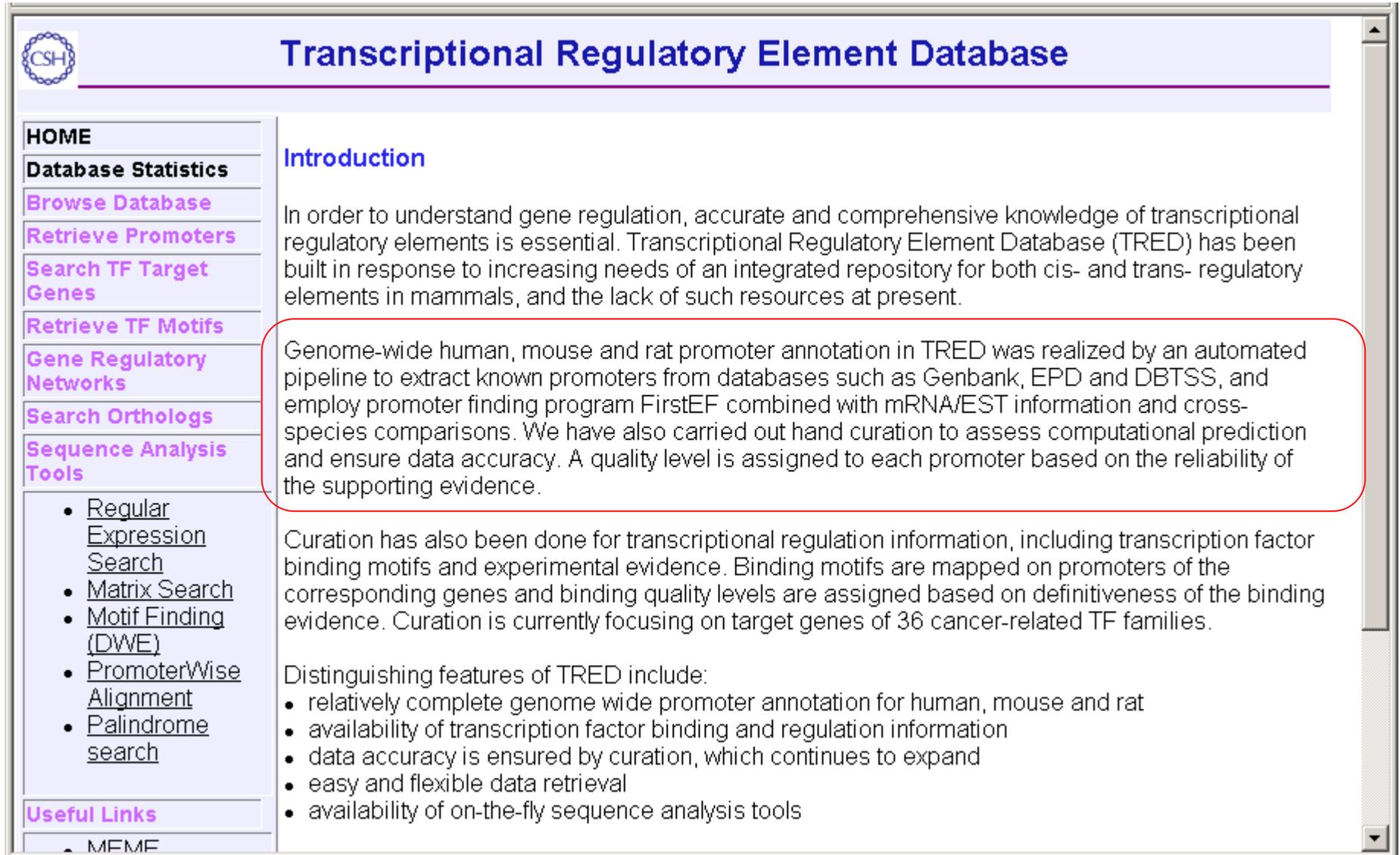
1.1.1.2.7 [LRF-1](#)

1.1.1.3 *Subfamily*: Maf

1.1.1.3.1 [v-Maf](#)

TRED

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



The screenshot shows the TRED website interface. At the top left is the CSH logo. The main header is "Transcriptional Regulatory Element Database". On the left is a navigation menu with categories: HOME, Database Statistics, Browse Database, Retrieve Promoters, Search TF Target Genes, Retrieve TF Motifs, Gene Regulatory Networks, Search Orthologs, and Sequence Analysis Tools. The "Sequence Analysis Tools" category is expanded to show: Regular Expression Search, Matrix Search, Motif Finding (DWE), PromoterWise Alignment, and Palindrome search. Below the menu is a "Useful Links" section with a link to MEME. The main content area is titled "Introduction" and contains three paragraphs of text. The second paragraph is highlighted with a red rounded rectangle.

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.

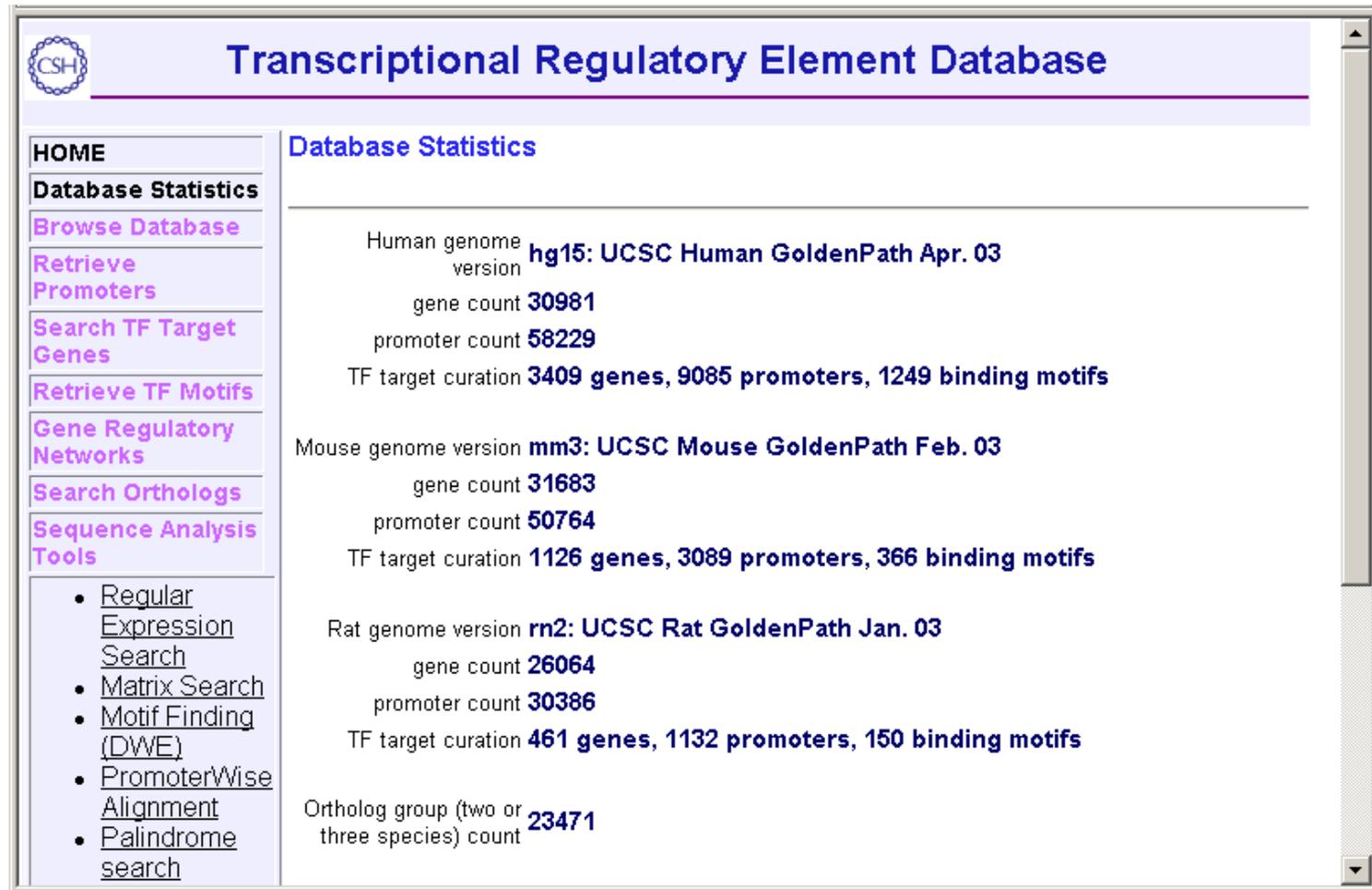
Curation has also been done for transcriptional regulation information, including transcription factor binding motifs and experimental evidence. Binding motifs are mapped on promoters of the corresponding genes and binding quality levels are assigned based on definitiveness of the binding evidence. Curation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

Navigation Menu:
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

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
<input type="checkbox"/>	SYNGR1	human, Homo sapiens	22q13.1	28332	3.1: refseq,predicted	predicted
<input type="checkbox"/>	PCNA	human, Homo sapiens	20pter-12	27147	1: known,curated	known
<input type="checkbox"/>	RIPK2	human, Homo sapiens	8q21	40094	2: known	maybe
<input type="checkbox"/>	OGFR	human, Homo sapiens	20q13.3	26502	2: known	predicted
<input type="checkbox"/>	POLE2	human, Homo sapiens	14q21-22	112824	1: known,curated	known
<input type="checkbox"/>	FLJ35487	human, Homo sapiens	1p22.3	3302	3.2: refseq	maybe
<input type="checkbox"/>	n/a	human, Homo sapiens		5838	4: withRNA	maybe
<input type="checkbox"/>	PTPN6	human, Homo sapiens	12p13	8664	1: known,curated	known
<input type="checkbox"/>	n/a	human, Homo sapiens		11668	2: known	maybe
<input type="checkbox"/>	LIPC	human, Homo sapiens	15q21-23	13325	1: known,curated	known
<input type="checkbox"/>	POLR2C	human, Homo sapiens	16q13-21	15265	1: known,curated	maybe
<input type="checkbox"/>	n/a	human, Homo sapiens		17054	4: withRNA	maybe
<input type="checkbox"/>	ZNF540	human, Homo sapiens	19q13.13	20774	2: known	maybe
<input type="checkbox"/>	SLC16A14	human, Homo sapiens	2q37.1	24581	2: known	maybe
<input type="checkbox"/>	TRPM2	human, Homo sapiens	21q22.3	27513	2: known	maybe
<input type="checkbox"/>	FHIT	human, Homo sapiens	3p14.2	30712	2: known	maybe
<input type="checkbox"/>	POU3F2	human, Homo sapiens	6q16	35658	2: known	maybe
<input type="checkbox"/>	n/a	human, Homo sapiens		39702	6: other	maybe
<input type="checkbox"/>	n/a	human, Homo sapiens		42816	6: other	maybe
<input type="checkbox"/>	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](#)

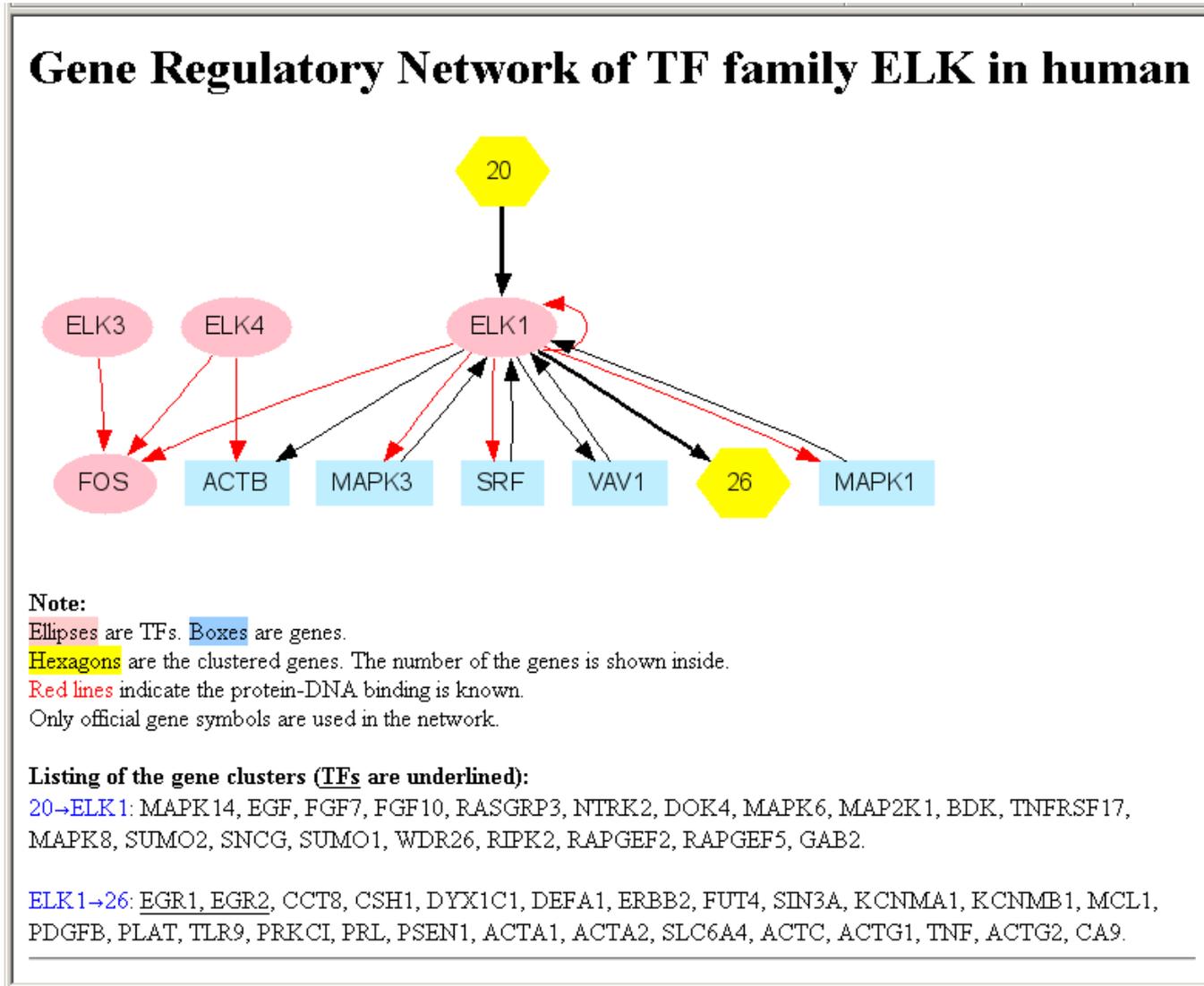
Introduction

In order to understand gene regulation, accurate and comprehensive knowledge of transcriptional 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



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, релиз 10 (<http://dbtss.hgc.jp/>)

Database of Transcriptional Start Sites
DBTSS
DataBase of Encyclopedia of Regulatory Omics
DBKERO

Release 10.0 Updated (September 15, 2017)
Based on UCSC hg38, mm10

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

Top |

Database Search

Species:

Keyword:

Genome browser:

Human Chromatin Features

Search from Genomic Position:

Search from SNP (dbSNP rsID):

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.

DBKERO; Database Encyclopedia of Regulatory Omics

We separated the part of DBTSS, particularly focusing on genomic changes effecting the transcriptional regulations as DBKERO. It is believed that single nucleotide variations (SNVs) in the transcriptional regulatory regions are responsible for many human diseases, including cancers. However, it remains difficult to identify functionally relevant SNVs from those having no explicit biological consequences. In this version of DBKERO, we attempt to associate SNVs with the omics information of the surrounding regions. We used SNVs which we identified from genomic analyses of various types of cancers, including somatic mutations of 100 lung adenocarcinoma and lung small cell carcinoma. For germline variations, we used SNVs in dbSNP as well as our unique dataset of variations in 1000 Japanese individuals. We integrated those SNV information with our original datasets of TSS-seq, RNA-seq, ChIP-seq of representative histone modifications and Bisulfite Sequencing of cytosine methylations of DNA. We further connected the multi-omics data of model organisms by genome-genome alignment.

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/DBKERO is helpful to understand biological consequences of the massively identified TSSs and identify human genetic valuations which are associated with disordered transcriptional regulations.

Tweets by @dbkero

KERO
@dbkero
こんにちは。#はじめでのツイート
Jul 28, 2017

News

15 Sep. 2017: DBTSS version 10.0 released.

1 Sep. 2017: New C1 data of human lung adenocarcinoma cell line (LC-2/ad: (replicate)) are now available (See browser: Single cell -> C1 -> LC-2/ad (replicate)).

TSSs = Transcriptional Start Sites

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

База развивается с 2002г.

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

В 2017 году базу ресурс разделили на две части : DBTSS and DBKERO

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



- Database of Transcriptional Start Sites -
DBTSS

Release 9.0 Updated (July 9, 2015)
Based on UCSC hg38, mm10

Top

Database Search

Species:
H. sapiens
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):
rs375228889

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

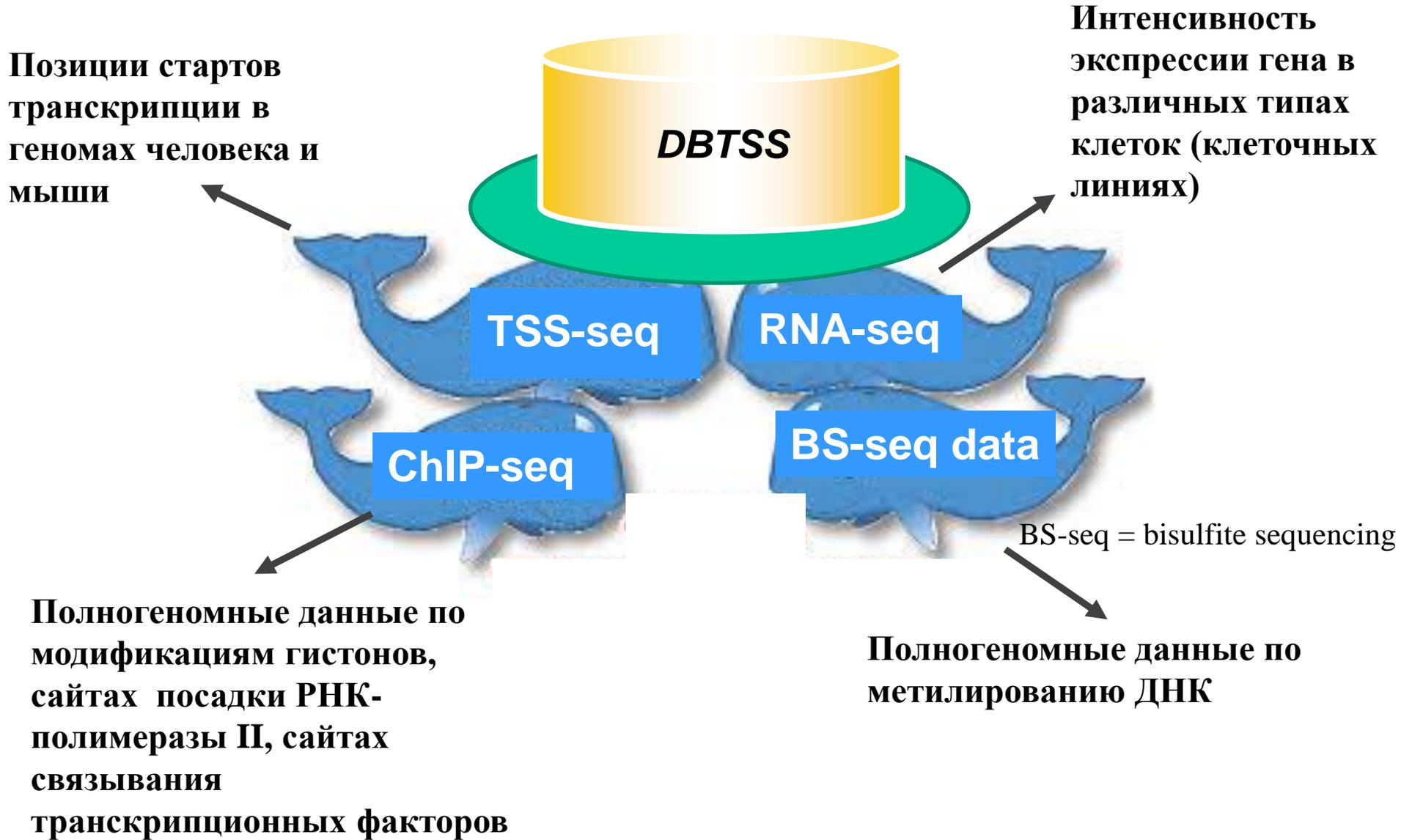
TSSs = Transcriptional Start Sites

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

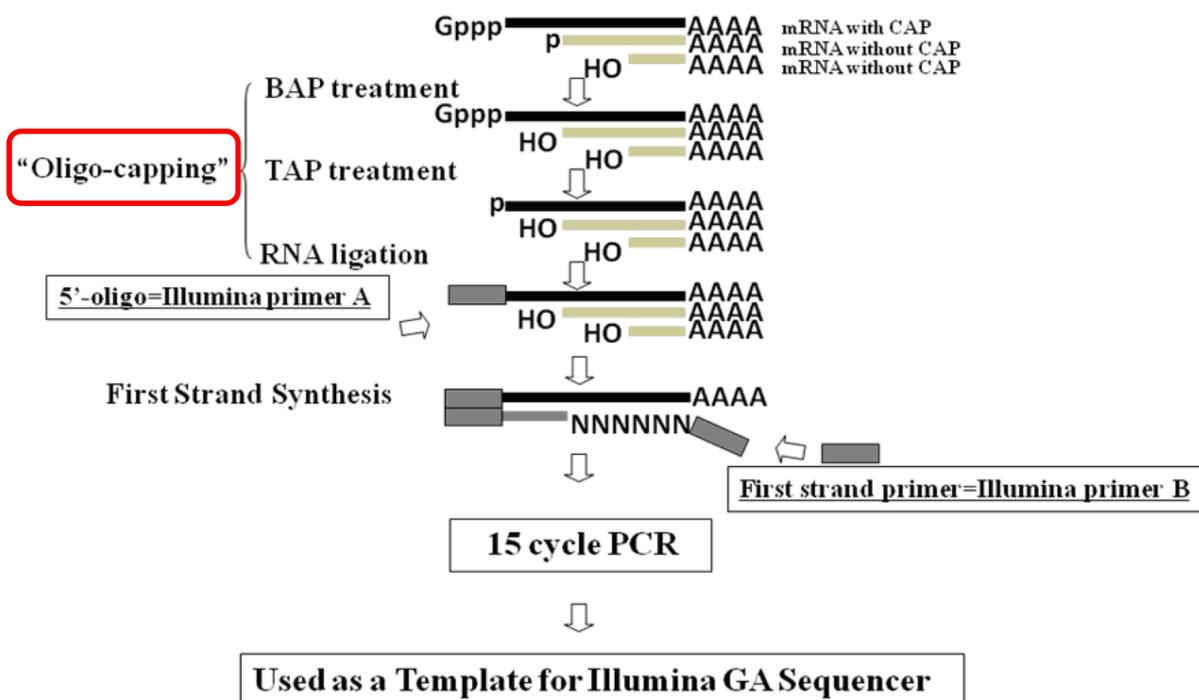


TSS-seq – оригинальная методика, позволяющая определять позиции стартов транскрипции в геноме . Представляет собой сочетание нескольких подходов:

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

Метод oligo-capping:

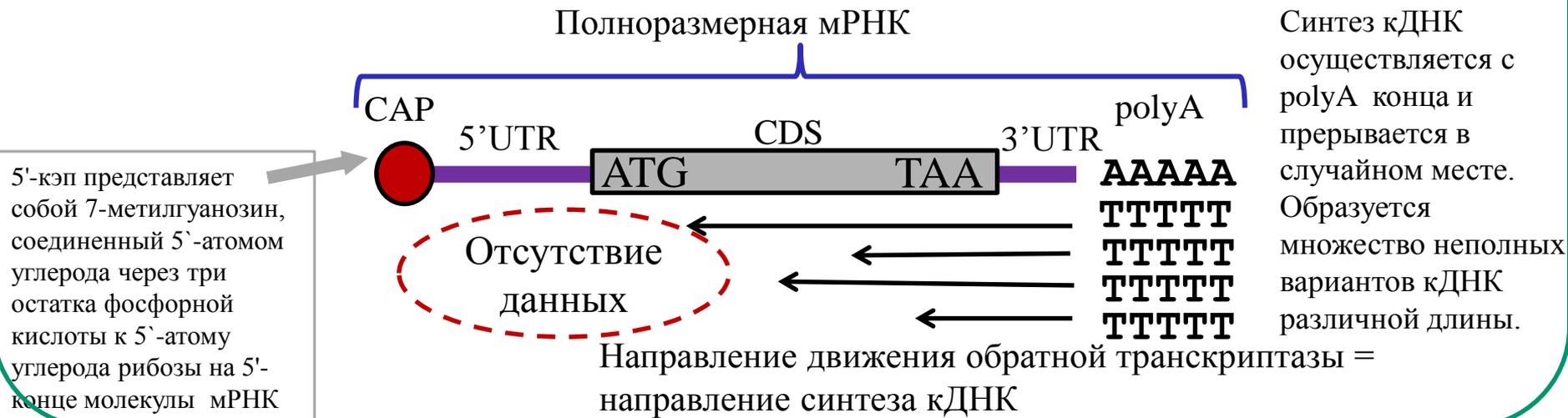
Адапторы (олигонуклеотиды), которые используются в процессе секвенирования на платформе Illumina, пришиваются к кэп-сайту транскриптов. Процесс включает три этапа , на которых используются три разных фермента. Это позволяет пришивать адапторы только к полноразмерным транскриптам (то есть транскриптам,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



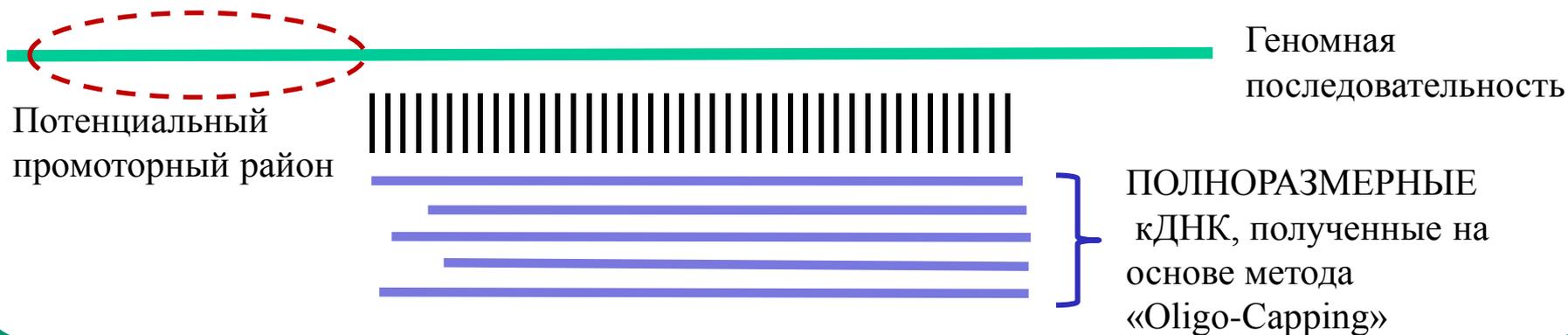
Преимущество «Oligo-Capping method»:

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 ?

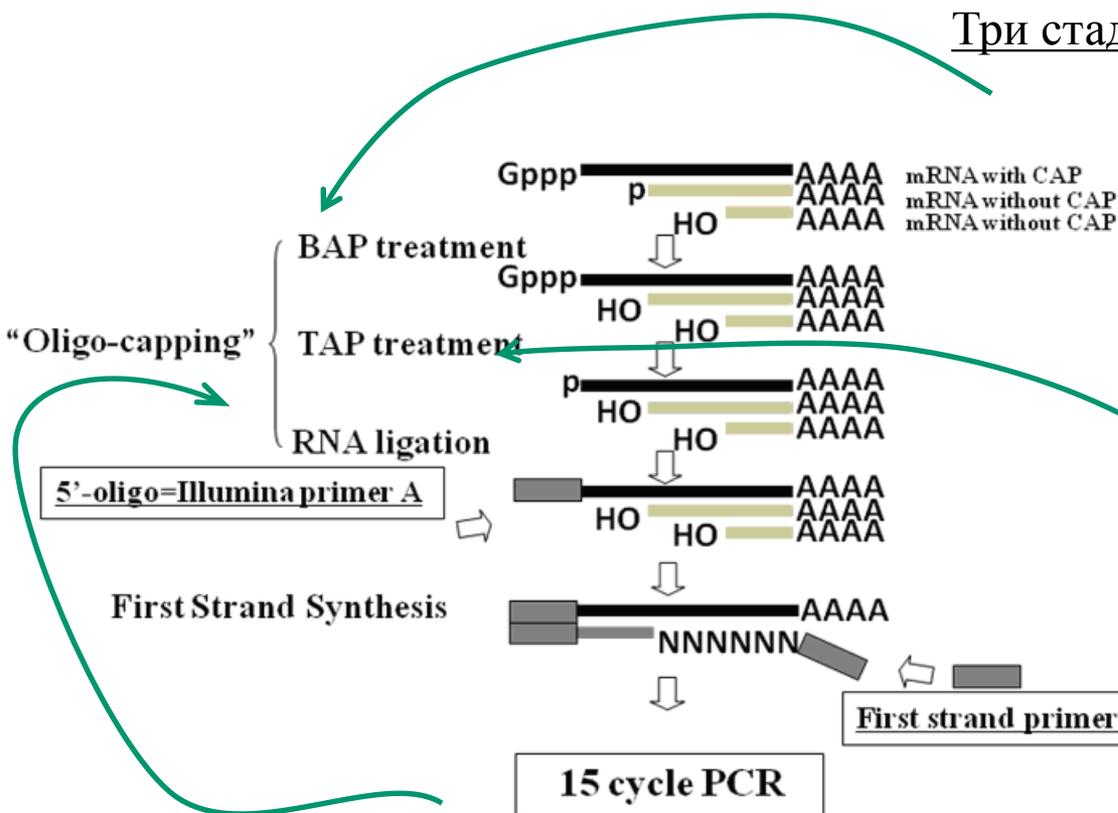
Мечение 5'-концов мРНК, имеющих 5'-кэп*. При этом 5'-кэп замещается на синтетический олигонуклеотид определенной длины и состава.

* 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) **ВАР** (Bacterial alkaline phosphatase) – фосфатаза, которая отщепляет фосфорную группу на конце поврежденной (неполной) мРНК, у которой отсутствует САР

2) **ТАР** (Tobacco acid pyrophosphatase) – пиррофосфатаза, отщепляющая САР, 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 САР

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

Геномный браузер базы DBTSS: TSS viewer

Top

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	Hs.488293	NM_005228	epidermal growth factor receptor isoform a precursor	5600 bases
EGFR (1956)	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 HEPG2 TIG3 HeLa 26 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

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

Adult tissues

- TSS-seq

Представление
гена в геномном
браузере

Представление
информации о
транскриптах в
различных
тканях

Геномный браузер базы DBTSS

The screenshot displays the DBTSS genome browser interface. On the left, there are search and track options. The main area shows a genomic track for chromosome 7 (hg38) with various data layers. The top navigation bar includes 'Database Search' and 'Human Chromatin Features' sections. The 'Track items' section is highlighted with a green border and contains a list of cell lines and tissues. The main track area is titled 'Genome viewer (Multi-omics Data)' and shows a genomic region from 55,019,032 to 55,156,951 on chromosome 7. The tracks include: 'Your Query' (red bar), 'Sequence', 'CpG Island', 'NCBI RefSeq' (EGFR gene models), 'LC2/ad SNV 50%', 'PC-7 SNV 50%', 'Chromatin Hi128' (Heterochromilo), 'Chromatin Hi 78' (Heterochromilo), 'Chromatin HiEC' (Heterochromilo), 'Chromatin HiSM' (Heterochromilo), 'Chromatin HepG2' (Heterochromilo), 'Chromatin Huvcc' (Heterochromilo), 'Chromatin HHEK' (Heterochromilo), 'Chromatin HMLP' (Heterochromilo), 'PC-7 TSS 50%', 'PC-7 RNA RPKM' (EGFR 0.01), 'PC-7 BS 100%' (50%), 'PC-7 Chromatin' (8_Low/no_signal), 'PC-7 H3K4me3' (100), 'PC-7 H3K4me1' (100), 'PC-7 H3K27me3' (100), 'PC-7 H3ac' (100), and 'PC-7 H3K9me3' (100). The tracks are color-coded and show signal intensity across the genomic region.

Database Search

Species:

Keyword:

Search

Genome browser:

Search

Human Chromatin Features

Search from Genomic Position:

Search

Search from SNP (dbSNP rsID):

Search

Search from SNV (COSMIC: somatic mutation):

Track items

- + User data
- + Public data
- + Comparative genome
- + Chromatin map (ENCODE)
- + SNP/SNV public data
- + Lung adenocarcinoma 26 cell lines
- + DLD-1: Colorectal adenocarcinoma cell lines
- + Adult tissue
- + Fetal tissue
- + BEAS-2B: Bronchial epithelial cell line
- + Ramos: Burkitt's lymphoma cell line
- + MCF7: Breast adenocarcinoma cell line
- + HEK293: Embryonic kidney cell line
- + TIG3: Lung fibroblasts cell line
- + HeLa cell
- + UPF1 RNAi

Genome viewer (Multi-omics Data)

Database: DBTSS version 9.0 | Assembly: hg38 | Genomic position: chr7:55 019 032-55 156 951

Width: 750 | Height: 900 | Change size | Show item panel | Browse | Info.

Zoom in: 1.5x | 2x | 5x | Zoom out: 2/3x | 1/2x | 1/5x | Synchronize track settings

chr7:55,019,032-55,156,951

100,000,000

138 kb

55,100,000

Your Query

Sequence

CpG Island

NCBI RefSeq

LC2/ad SNV 50%

PC-7 SNV 50%

Chromatin Hi128 78

Chromatin Hi 78

Chromatin HiEC

Chromatin HiSM

Chromatin HepG2

Chromatin Huvcc

Chromatin HHEK

Chromatin HMLP

PC-7 TSS 50%

PC-7 RNA RPKM

PC-7 BS 100%

PC-7 Chromatin

PC-7 H3K4me3

PC-7 H3K4me1

PC-7 H3K27me3

PC-7 H3ac

PC-7 H3K9me3

Add separator | Clear all tracks | Show basic public data (Sequence+RefSeq+CpG)

Нуклеотидные замены в различных типах клеток

Типы хроматина в различных типах клеток

Уровень метилирования

Модификации гистонов в различных клетках

Регулярное обновление версий базы DBTSS

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

The screenshot shows the left sidebar of the DBTSS website. At the top, it says "Database of Transcriptional Start Sites" and "DBTSS". Below that are language options: "English | [Japanese](#)" and a link to the "old version". The "Database Search" section includes a "Keyword Search" box with fields for "Species:" (set to "H. sapiens"), "Cell:" (set to "TSSseq : DLD1 : clorectal adenocarcino"), "Category:" (set to "RefSeq ID (NM_)"), and "Keyword:" (with an asterisk). There is also a "ppm #(>=):" field set to "5" and a "Search" button. At the bottom of the sidebar, it says "(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 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, 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.

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

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

В норме и при различных воздействиях

22 клеточные ситуации

DBTSS, релиз 10

«Июминка» -2017 года !!!!!:

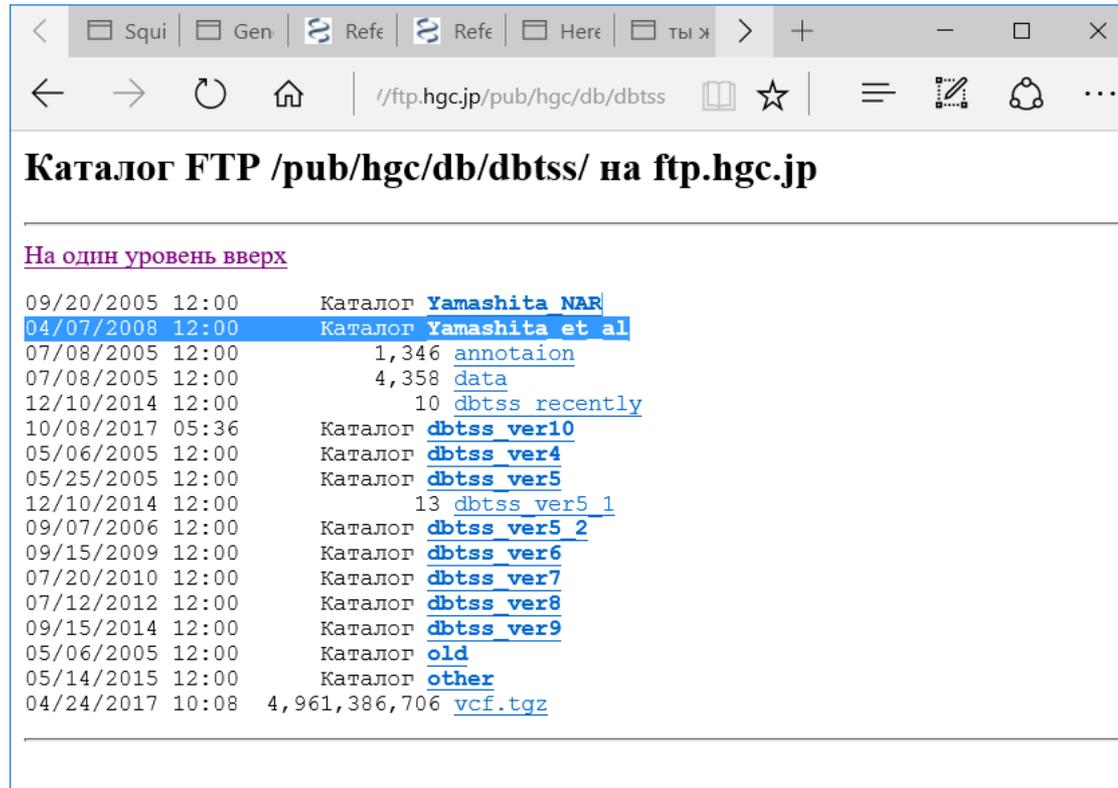
The screenshot shows the website interface for DBTSS and DBKERO. At the top, there is a navigation bar with the URL 'dbtss.hgc.jp' and a search bar. The main header features the logo for 'DBTSS / DBKERO' and the text 'Release 10.0 Updated (September 15, 2017) Based on UCSC hg38, mm10'. Below the header, there is a section titled 'About this database' which provides an overview of the data. The 'Database Search' section includes fields for 'Species' (set to 'H. sapiens'), 'Keyword' (with a placeholder 'keyword (e.g. EGFR, NM)'), and 'Genome browser' (set to 'hg38'). There are also search buttons and a 'Human Chromatin Features' section with a search bar. On the right side, there is a 'Tweets by @dbkero' section and a 'News' section with recent updates.

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: всю информацию можно скачать !!!



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 (старты 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 website interface. On the left is a vertical navigation menu with categories: 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), and First Production (2007-2012), Pilot (2003-2007), and Contacts.

The main content area is titled "Encyclopedia of DNA Elements at UCSC 2003 - 2012" and includes an "About" section. The "About" text states: "The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active." It also notes: "ENCODE results from 2007 and later are available from the ENCODE Project Portal, encodeproject.org. This covers data generated during the two production phases 2007-2012 and 2013-present. The ENCODE Project Portal also hosts additional ENCODE access tools, and ENCODE project pages including up-to-date information about data releases, publications, and upcoming tutorials." A key message is: "UCSC coordinated data for the ENCODE Consortium from its inception in 2003 (Pilot phase) to the end of the first 5 year phase of whole-genome data production in 2012. All data produced by ENCODE investigators and the results of ENCODE analysis projects from this period are hosted in the UCSC Genome browser and database. Explore ENCODE data using the image links below or via the left menu bar. **All ENCODE data at UCSC are freely available for download and analysis.**"

Below the "About" section are four panels:

- Explore ENCODE data (2003 - 2012) at UCSC**: Shows a heatmap visualization of ENCODE Experiment Matrix data (2007-2012).
- View ENCODE data (2003 - 2012) in the UCSC Genome Browser**: Shows a screenshot of the UCSC Genome Browser interface displaying ENCODE tracks for Human Feb. 2009 (GRCh37/hg19) Assembly.
- Search for data (current) at the ENCODE Portal**: Shows a screenshot of the ENCODE Portal search interface.
- Search for ENCODE tracks (2003 - 2012) in the UCSC Browser**: Shows a screenshot of the UCSC Browser search interface for ENCODE 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,

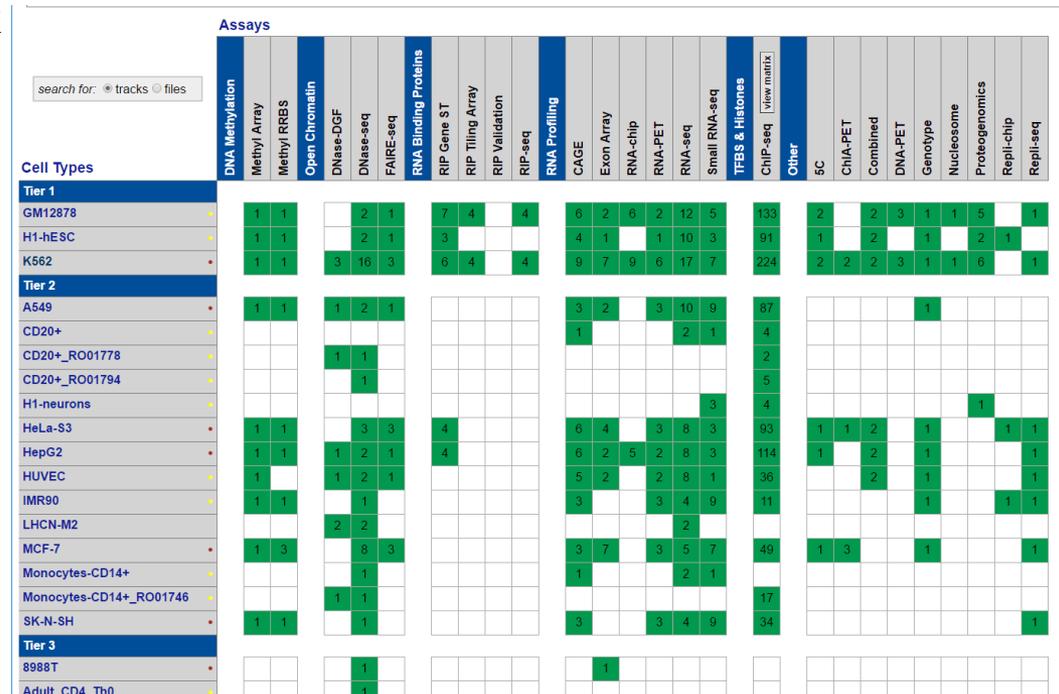
- регуляторные участки

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



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

Scale 100 kb hg19
chr10: 69,450,000 | 69,500,000 | 69,550,000

RefSeq Genes
CTNNA3
CTNNA3
DNAJC12
DNAJC12

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples

move start < 2.0 > move end < 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.

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

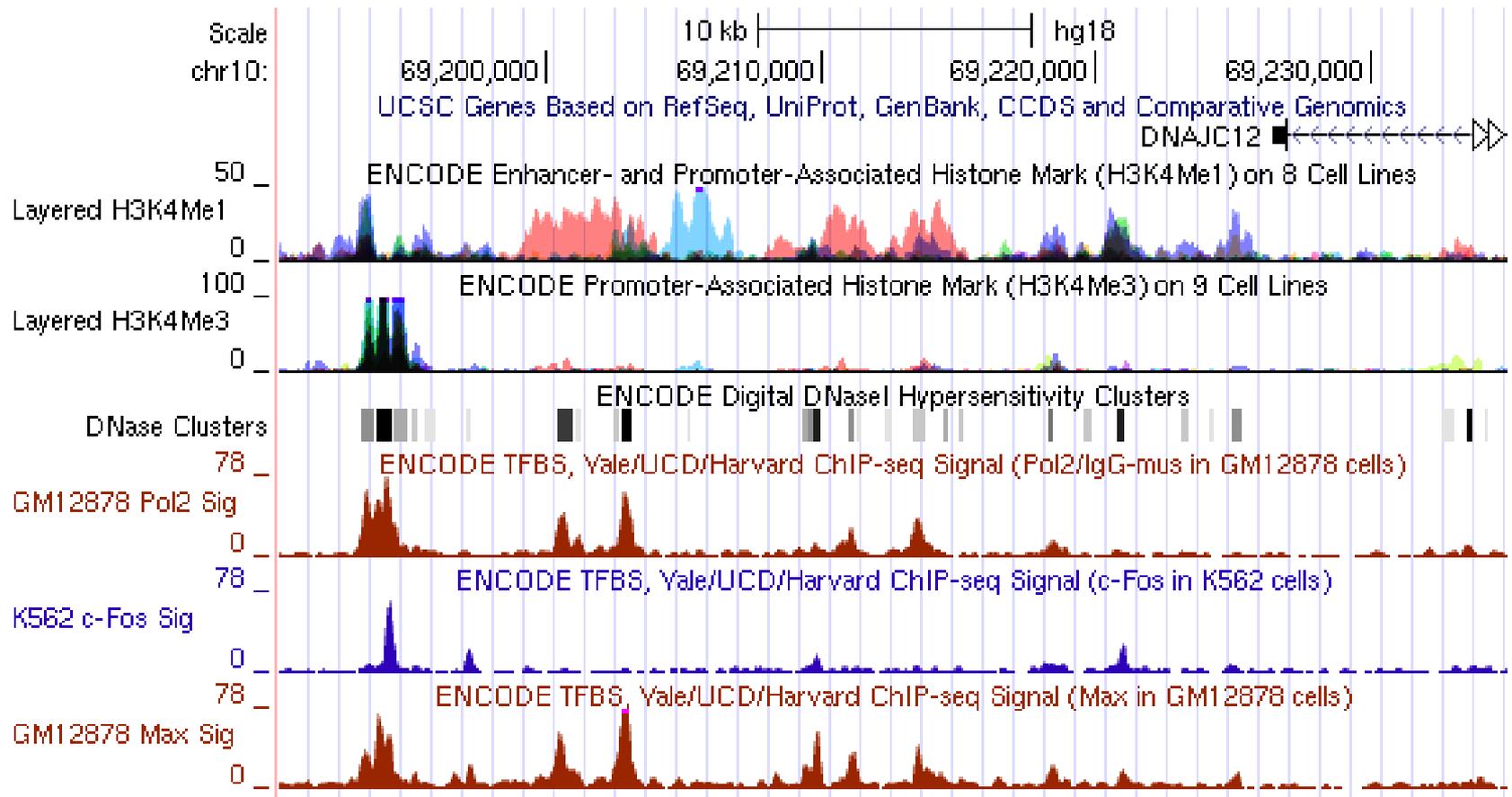
collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

Mapping and Sequencing Tracks refresh

<u>Base Position</u> dense ▾	<u>Chromosome Band</u> hide ▾	<u>STS Markers</u> hide ▾	18 <u>FISH Clones</u> hide ▾	<u>Recomb Rate</u> hide ▾	18 <u>deCODE Recomb</u> hide ▾
<input checked="" type="checkbox"/> <u>ENCODE Pilot</u> hide ▾	<u>Map Contigs</u> hide ▾	<u>Assembly</u> hide ▾	<u>GRC Map Contigs</u> hide ▾	<u>Gap</u> hide ▾	<u>Publications</u> hide ▾

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

Визуализация данных проекта ENCODE в геномном браузере Университета г.Санта Круз (UCSC). (участок 10 хромосомы (chr10:69,190,281-69,235,000))



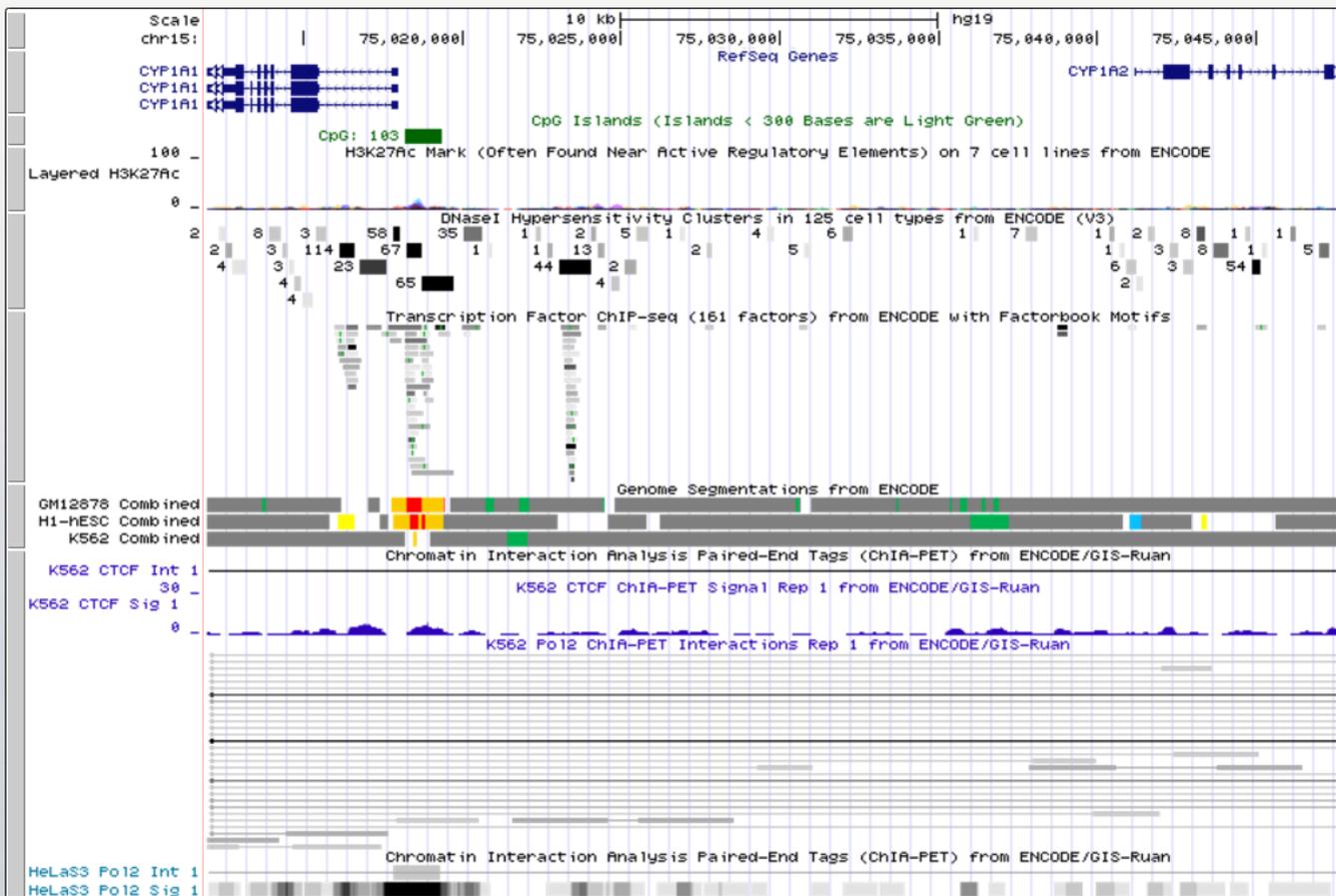
Геномный браузер Университета г.Санта Круз

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 26.1



CpG – богатые участки

Сайты гиперчувствительности к DNaseI (открытый хроматин)

Районы связывания с различными ТФ по данным экспериментов ChIP-seq

Типы хроматина в различных типах клеток

3D структура хроматина: контакты между удаленными участками генома

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

Панель инструментов для настройки отображаемой информации в геномном браузере UCSC

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

The image shows a screenshot of the UCSC Genome Browser's tool panel. The panel is organized into several sections, each with a header and a 'refresh' button. The sections are:

- Genes and Gene Predictions** (collapsed): This section contains a grid of tracks for various gene prediction sources. Each track has a 'hide' dropdown menu. The tracks include: UCSC Genes, RefSeq Genes (set to 'pack'), AceView Genes, Augustus, CCDS, CRISPR..., Ensembl Genes, 17 EvoFold, Exoniphy, GENCODE..., Geneid Genes, Genscan Genes, H-Inv 7.0, IKMC Genes Mapped, lincRNAs..., LRG Transcripts, MGC Genes, N-SCAN, Old UCSC Genes, ORFeome Clones, Other RefSeq, Pfam in UCSC Gene, Retroposed Genes, SGP Genes, SIB Genes, sno/miRNA, TransMap..., tRNA Genes, UCSC Alt Events, UniProt, Vega Genes, and Yale Pseudo60.
- Phenotype and Literature** (expanded): This section is currently expanded and contains a 'refresh' button.
- mRNA and EST** (expanded): This section is currently expanded and contains a 'refresh' button.
- Expression** (expanded): This section is currently expanded and contains a 'refresh' button.
- Regulation** (expanded): This section is currently expanded and contains a 'refresh' button. It includes tracks for: ENCODE Regulation... (set to 'show'), 18 CD34 Dnasel, CpG Islands... (set to 'show'), ENC Chromatin... (set to 'show'), ENC DNA Methyl... (set to 'hide'), ENC DNase/FAIRE..., ENC Histone... (set to 'hide'), ENC RNA Binding..., ENC TF Binding..., FSU Repli-chip (set to 'hide'), Genome Segments (set to 'full'), 18 NKI Nuc Lamina..., ORegAnno (set to 'hide'), Stanf Nucleosome (set to 'hide'), SUNY SwitchGear (set to 'hide'), 17 SwitchGear TSS (set to 'hide'), TFBS Conserved (set to 'hide'), TS miRNA sites (set to 'hide'), UCSF Brain Methyl (set to 'hide'), UMMS Brain Hist (set to 'hide'), UW Repli-seq (set to 'hide'), and Vista Enhancers (set to 'hide').
- Comparative Genomics** (collapsed): This section is currently collapsed and contains a 'refresh' button.

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](#).

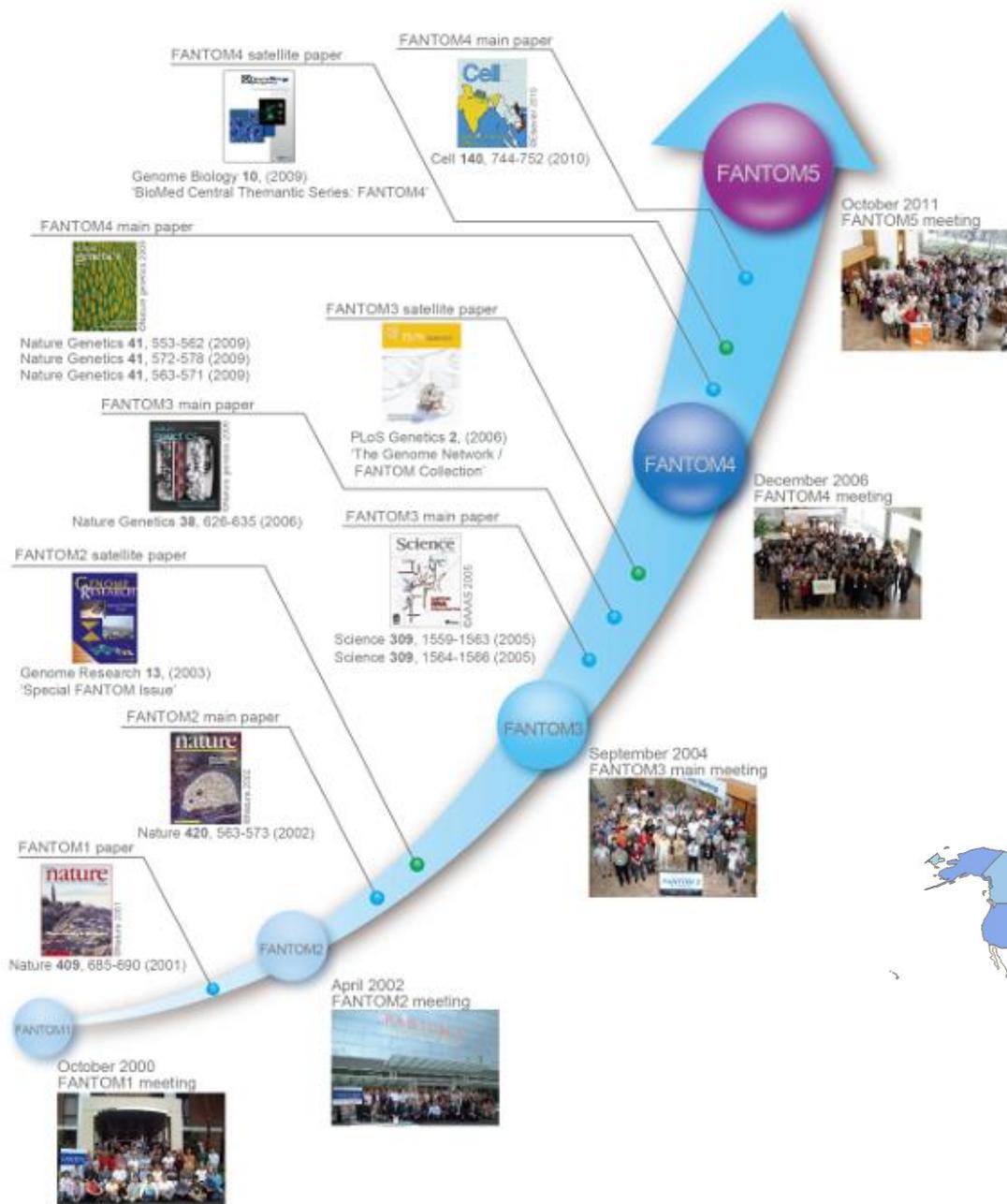
Using the Table Browser

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 *Cyp1A1* (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)

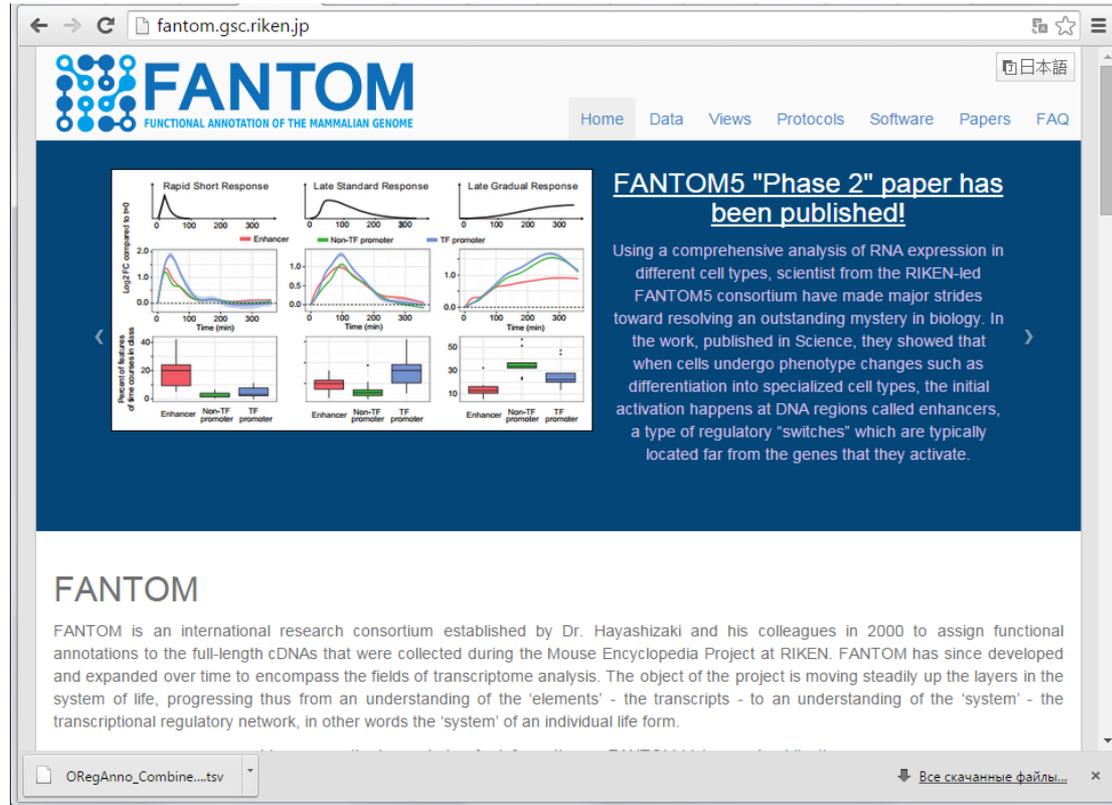
Members

Over 500 FANTOM members from more than 20 countries.



FANTOM5

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



The screenshot shows the FANTOM5 website interface. At the top, there is a navigation menu with links for Home, Data, Views, Protocols, Software, Papers, and FAQ. A language selector for Japanese is visible. The main content area features a blue banner with the headline "FANTOM5 'Phase 2' paper has been published!". Below the headline, there is a summary of the research findings, mentioning that scientists from the RIKEN-led FANTOM5 consortium have made major strides in resolving an outstanding mystery in biology. The text explains that when cells undergo phenotype changes such as differentiation into specialized cell types, the initial activation happens at DNA regions called enhancers, which are typically located far from the genes they activate. To the left of the text, there are six line graphs and three box plots. The line graphs show Log2FC compared to H0 over time (0 to 300 minutes) for three response types: Rapid Short Response, Late Standard Response, and Late Gradual Response. Each graph compares Enhancer (red), Non-TF promoter (green), and TF promoter (blue) regions. The box plots show the Percent of features of the same class in cells for each response type, comparing Enhancer, Non-TF promoter, and TF promoter regions.

FANTOM5 "Phase 2" paper has been published!

Using a comprehensive analysis of RNA expression in different cell types, scientist from the RIKEN-led FANTOM5 consortium have made major strides toward resolving an outstanding mystery in biology. In the work, published in Science, they showed that when cells undergo phenotype changes such as differentiation into specialized cell types, the initial activation happens at DNA regions called enhancers, a type of regulatory "switches" which are typically located far from the genes that they activate.

FANTOM

FANTOM is an international research consortium established by Dr. Hayashizaki and his colleagues in 2000 to assign functional annotations to the full-length cDNAs that were collected during the Mouse Encyclopedia Project at RIKEN. FANTOM has since developed and expanded over time to encompass the fields of transcriptome analysis. The object of the project is moving steadily up the layers in the system of life, progressing thus from an understanding of the 'elements' - the transcripts - to an understanding of the 'system' - the transcriptional regulatory network, in other words the 'system' of an individual life form.

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](#)

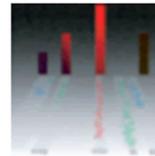
Protocols

Detailed descriptions on protocols used in FANTOM



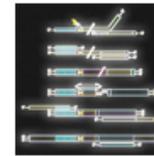
Basic CAGE Technology

A method for genome-wide identification of transcription start sites



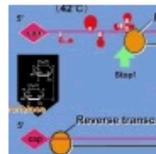
DeepCAGE

A powerful application for next generation sequencing



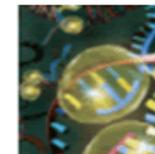
nanoCAGE and CAGEscan

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)

The screenshot shows the UniProt website homepage. At the top left is the UniProt logo. A search bar contains 'UniProtKB' and has a search button. Navigation links include 'BLAST', 'Align', 'Retrieve/ID mapping', 'Help', and 'Contact'. A mission statement follows: '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.' Below this are four main service tiles: UniProtKB (Knowledgebase), UniRef (Sequence clusters), UniParc (Sequence archive), and Proteomes. A 'Supporting data' section lists literature citations, taxonomy, subcellular locations, cross-ref. databases, diseases, and keywords. A 'News' section features social media icons and two news items: 'Forthcoming changes' and 'UniProt release 2015_09'. At the bottom, there are sections for 'Getting started' (with a YouTube icon), 'UniProt data' (with a download icon), and 'Protein spotlight' (with a protein image).

UniProt

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

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

News

[BLOG](#) [Twitter](#) [Facebook](#) [RSS](#)

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

[News archive](#)

Getting started **UniProt data**

[Text search](#)
Our basic text search allows you to

[Download latest release](#)
Get the UniProt data

Protein spotlight



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.

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 by ⁱ

 **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	
<input type="checkbox"/>	Q6GZX3	002L_FRG3G	 Uncharacterized protein 002L	FV3-002L	Frog virus 3 (isolate Goorha) (FV-3)	320	
<input type="checkbox"/>	Q6GZX4	001R_FRG3G	 Putative transcription factor 001R	FV3-001R	Frog virus 3 (isolate Goorha) (FV-3)	256	
<input type="checkbox"/>	Q197F7	003L_IIV3	 Uncharacterized protein 003L	IIV3-003L	Invertebrate iridescent virus 3 (IIV-3) (Mosquito iridescent virus)	156	
<input type="checkbox"/>	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).

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

Done

Internet

150%

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. Note: 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						
<input type="checkbox"/>	Chain	1 – 341	341	Cyclic AMP-responsive element-binding protein 1		PRO_0000076597
Regions						
<input type="checkbox"/>	Domain	101 – 160	60	KID		
<input type="checkbox"/>	Domain	283 – 341	59	bZIP		
<input type="checkbox"/>	Region	284 – 309	26	Basic motif By similarity		
<input type="checkbox"/>	Region	311 – 332	22	Leucine-zipper By similarity		
Sites						
<input type="checkbox"/>	Site	314	1	Required for binding TORCs		
Amino acid modifications						
<input type="checkbox"/>	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/>



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

Classification of Human Transcription Factors

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

Transcription factory classification

Superclass: S, Class: C, Family: F, Subfamily: SF, Genus: G, Factor species: FS

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:

- S C F SF G FS

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

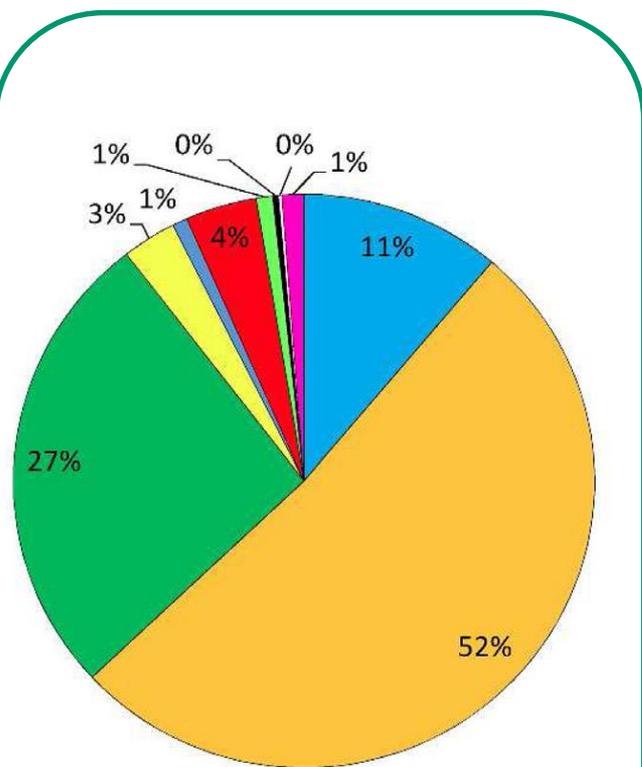
Protein expression pattern

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

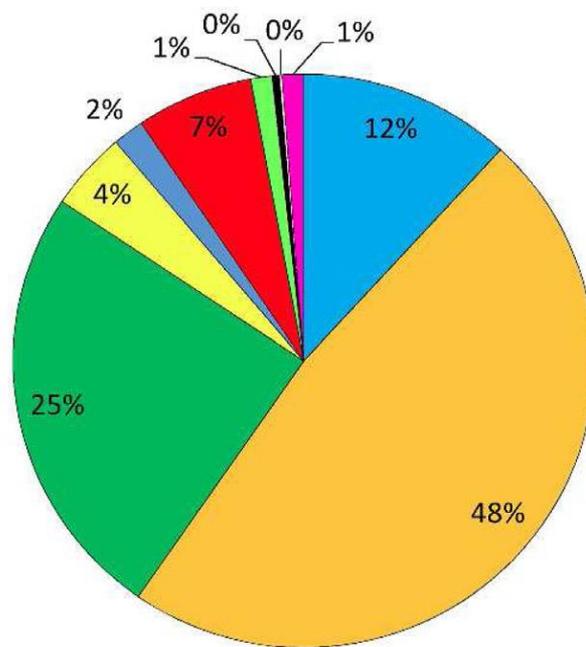
Готово

Распределение по суперклассам транскрипционных факторов человека и их мышинных ортологов.



HUMAN

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



MOUSE

- 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



(<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.

AnimalTFDB – возможности поиска ТФ:

А) по семейству

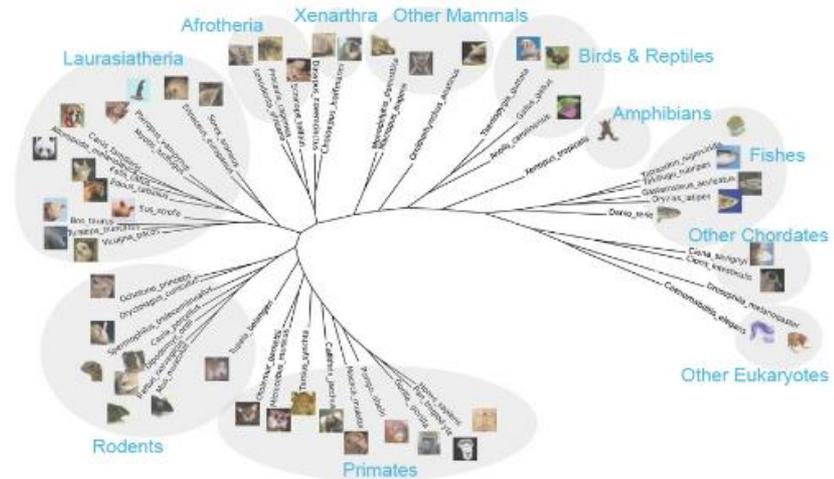
Б) по виду организма

В) поисковая система

Browse by family

No image							No image	No image			
AP-1	Antigen receptor	AP-2	AP-2	AP-2	DHL-1	C-EBP	CBF	CG-1	COE	COUP	
No image				No image							
CP2	CSO	CSL	CTF/NFI	CLT	DM	E2F	E2f8/8	ETS	Fox head		
GCM	GCM	GTF2I	HMG	HMG1/HMG2	HNF1b	HNF1	HNF1	HTH	IRF	MEO	
		No image			No image						
MYB	MYB	NDT80/Pho4	NF-YA	NF-YB/C	NFI	Nuclear orphan receptor	Oestrogen receptor	Other nuclear receptor	Others		
P03	PAX	PC4	POU	PPAR	Progesterone receptor	Pou1	Retinoid acid receptor	RFX	RHD		
No image											
ROR	ROR	Runt	SAND	SRF	STAT	T-box	TEA	TF_2BP	TF_O2	THAP	
No image										No image	
Thyroid hormone receptor	TSC22	Tud	ZBTB	ZF-EBD	ZF-C2H2	ZF-C2H4	ZF-GAGA	ZF-GATA	ZF-UTAF-like		

Browse by species



AnimalTFDB –поиск по ключевому слову (Symbol)



AnimalTFDB
Animal Transcription Factor Database

HOME BROWSE FAMILY BROWSE SPECIES SEARCH DOWNLOAD HELP ABOUT Quick search Go!

Search by Basic Information

Symbol

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

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

Transcript ID

Protein ID

Symbol

Map location

Description

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

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

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

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

show records per page:

Species select all reverse select

<input type="checkbox"/> Ailuropoda melanoleuca	<input type="checkbox"/> Anolis carolinensis	<input type="checkbox"/> Bos taurus
<input type="checkbox"/> Caenorhabditis elegans	<input type="checkbox"/> Callithrix jacchus	<input type="checkbox"/> Canis familiaris
<input type="checkbox"/> Cavia porcellus	<input type="checkbox"/> Choloepus hoffmanni	<input type="checkbox"/> Ciona intestinalis
<input type="checkbox"/> Ciona savignyi	<input type="checkbox"/> Danio rerio	<input type="checkbox"/> Dasyopus novemcinctus
<input type="checkbox"/> Dipodomys ordii	<input type="checkbox"/> Drosophila melanogaster	<input type="checkbox"/> Echinops telfairi
<input type="checkbox"/> Equus caballus		
<input type="checkbox"/> Gallus gallus		
<input checked="" type="checkbox"/> Homo sapiens		
<input type="checkbox"/> 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 PPARD

Homo sapiens PPAR receptor family

Basic information

EnsemblID : [ENSG00000112033](#)

GeneID : [5467](#)

Symbol : PPARD

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

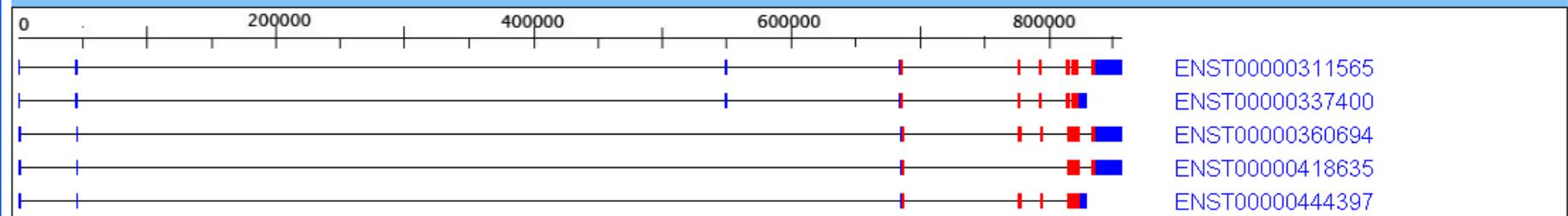
Gene Length : 85621

Gene Position : 6 : 35310335-35395955

Transcripts : There are 5 transcripts

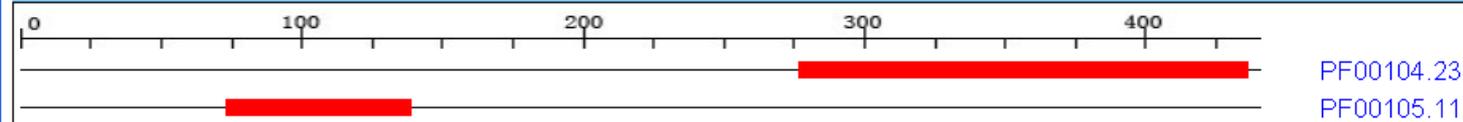
Transcript ID	Length (bp)	Protein ID	Length (aa)
ENST00000311565	3774	ENSP00000310928	441
ENST00000337400	2041	ENSP00000337063	361
ENST00000360694	3747	ENSP00000353916	441
ENST00000418635	3453	ENSP00000413314	343
ENST00000444397	2002	ENSP00000410837	361

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



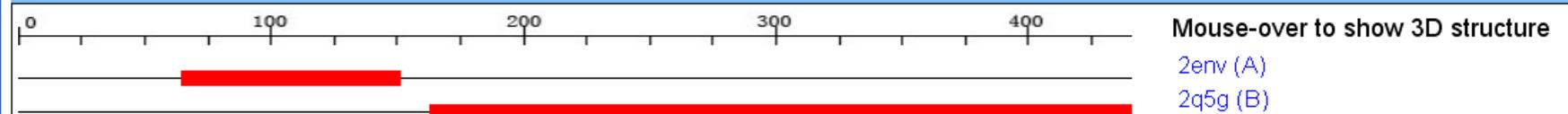
AnimalTFDB – результат поиска, фактор PPARD (продолжение)

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 β (δ) and PPAR γ and effects on gene expression	Biocarta	Show detail

Protein protein Interaction

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

Paralog

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

Ortholog group

Loxodonta africana	ENSLAFG00000017670	Macaca mulatta	ENSMMUG00000015904
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 carnettii	ENSOGAG00000000630	Pan troglodytes	ENSPTRG00000018082

Targets & TFBS

Number of targets	Number of TFBS	TFBS ID	Data Source

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/>)

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.

When referring to this classification, please cite:
Wingender, E., Schoepfs, T. and Donitz, J.
TFClass: An expandable hierarchical classification of human transcription factors
Nucleic Acids Res. 41, D165-D170 (2013). [DOI](#)

Transcription factor classification

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

Human TF

- 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
- 10 Yet undefined DNA-binding domains

Search:

Expand all Collapse all

Expand to:

Details

ID:

Definition:

Ресурс AnimalTFDB

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

AnimalTFDB
Animal Transcription Factor Database

HOME BROWSE FAMILY BROWSE SPECIES SEARCH DOWNLOAD HELP ABOUT Quick search

Factors of Homo sapiens

This dataset collected 1544 transcription factors in 71 families, 124 chromatin remodeling factors and 302 transcription co-factors of Homo sapiens

• Transcription factor family

AR(4)	Androgen receptor(1)	AR(5)	ARID(15)	ARH(108)	C/EBP(10)
CBF(1)	CG-1(2)	COE(4)	COUP(3)	CP2(7)	CSD(8)
CSL(2)	CTF/NF(4)	CIIT(7)	DM(7)	EPF(11)	Endyash receptor(2)
ETS(28)	Fork head(48)	OCM(2)	GCR(1)	CTF2(5)	HMG(50)
HLA/HLA-CY(2)	Homeobox(205)	HSF(8)	HTH(2)	IRF(8)	MBD(8)
IR1(8)	MYB(25)	MDM1/ProC(1)	NF-YA(1)	NF-YB/C(2)	Nr1(1)
Interleukin receptor(3)	Oestrogen receptor(1)	Other nuclear receptor(2)	Others(3)	PS3(3)	PAZ(9)
PC1(1)	POU(21)	PPAR receptor(3)	Progesterone receptor(1)	Prx1(2)	Retinoic acid receptor(7)
RF1(8)	RHD(10)	ROR receptor(4)	Runt(3)	SAND(8)	SRF(8)
STAT(7)	T-box(17)	TEA(4)	TF_2DP(46)	TF_Ou(3)	THAP(12)
Thyroid hormone receptor(25)	TSC22(4)	TuJ(5)	ZBT(48)	zBED(5)	zC2H2(634)
zC2H2(6)	zGATA(14)	zLITAF-like(2)	zMIZ(7)	zNF-X(2)	

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

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

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

Конец 5-ой лекции