

Лекция 4

Часть 1. Механизмы регуляции
транскрипции

Часть 2. Базы данных по регуляции
транскрипции

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

Часть 1. ТРЕХМЕРНАЯ СТРУКТУРА ХРОМАТИНА

1. LCR – локус-контролирующие районы;
2. Инсулаторы;
3. Транскрипционные факторы CTCF;
4. Когезин (cohesin) – белковый комплекс с кольцевой структурой;
5. Хромосомный оперон, транскрипционные фабрики;
6. Энхансерная РНК (eRNA) и ее регуляторные функции

ЛОКУС-КОНТРОЛИРУЮЩИЙ РАЙОН (LCR)

-фрагмент (или группа фрагментов) ДНК, присутствие которых обеспечивает высокий уровень тканеспецифической экспрессии связанного с ним гена, входящего в трансгенную конструкцию,

- пропорционально количеству копий трансгена и**
- независимо от места встраивания в геном .**

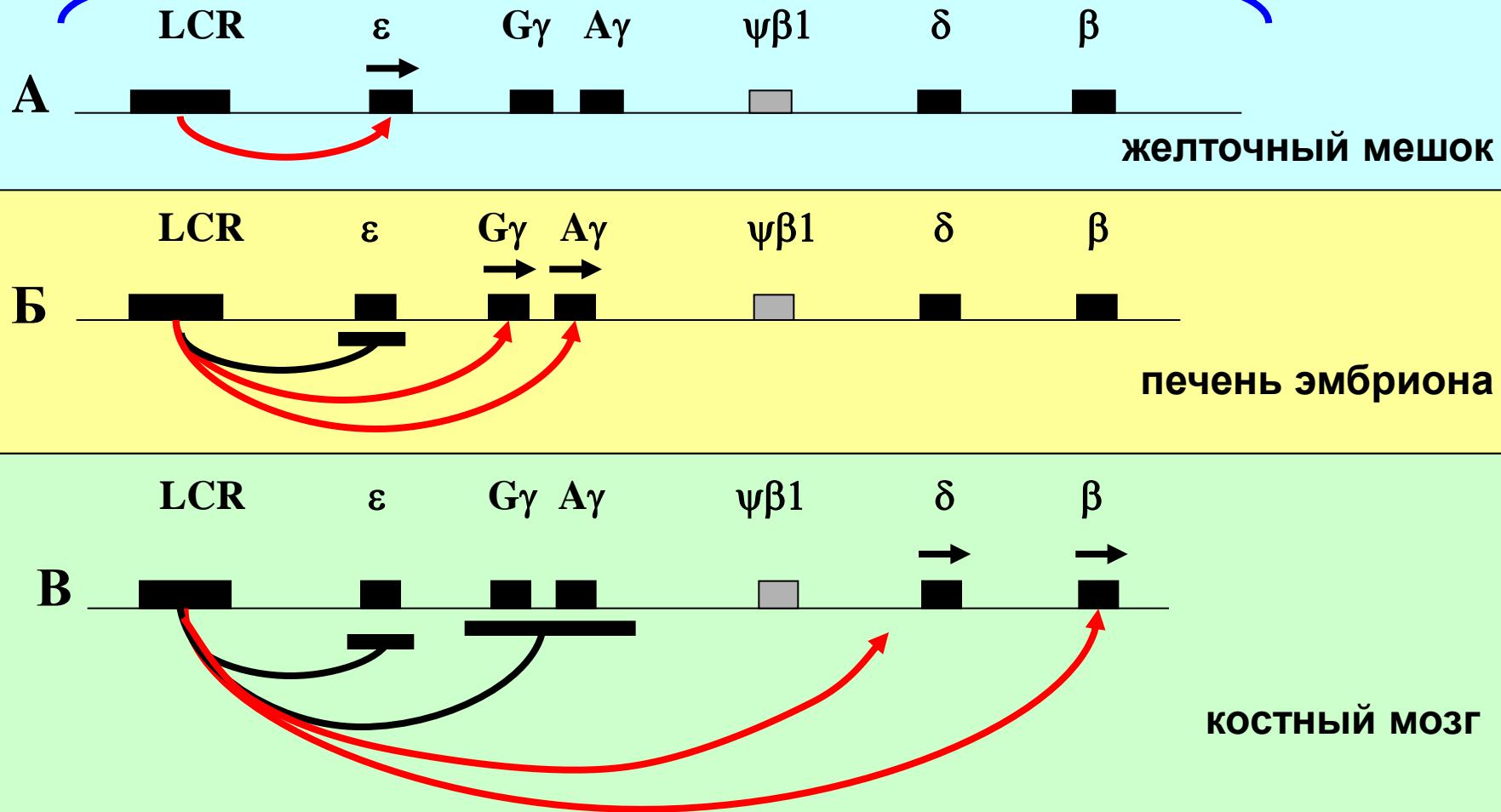
Как правило, локус-контролирующие районы регулируют гены, находящиеся в кластерах и обеспечивают их ткане- и стадиеспецифическую экспрессию .

Каждый LCR включает специфическую группу сайтов связывания и располагается иногда на очень большом (до десятков тысяч п.о.) расстоянии от контролируемой кассеты генов

Locus control regions (LCRs) are operationally defined by their ability to enhance the expression of linked genes to physiological levels in a tissue-specific and copy number-dependent manner at ectopic chromatin sites.

ПРИМЕР: ЛОКУС-КОНТРОЛИРУЮЩИЙ РАЙОН КЛАСТЕРА β-ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА

70 тыс.пар оснований

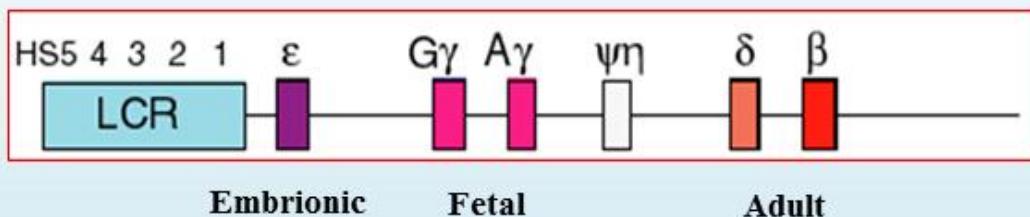


ϵ , $G\gamma$, $A\gamma$, δ , β - гены гемоглобинов; $\psi\beta 1$ - псевдоген

КЛАСТЕР β -ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА и ВРЕМЕННАЯ ДИНАМИКА ЭКСПРЕССИИ в ОНТОГЕНЕЗЕ

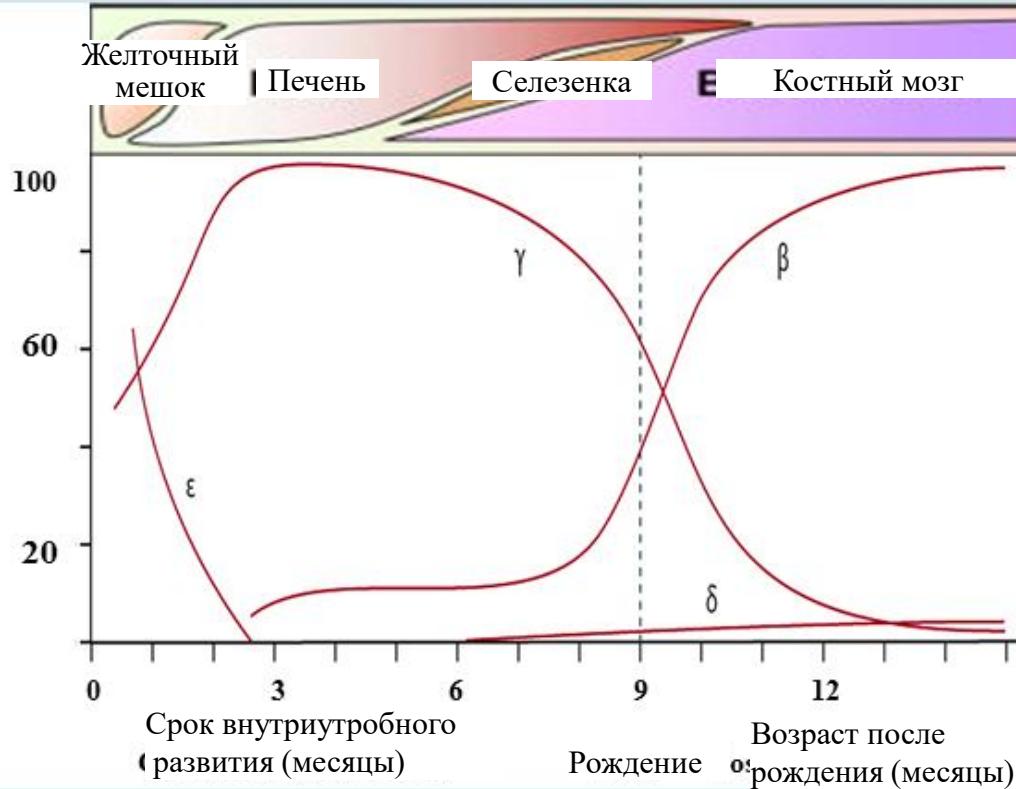
B-Globin locus

Chr11



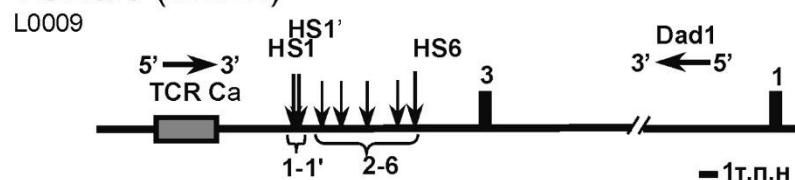
Орган, где осуществляется эритропоэз

Уровень экспрессии глобинов (%)



Примеры структурно-функциональной организации LCR

TCR α/δ (мышь)



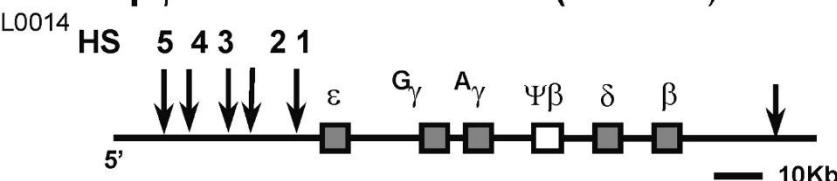
HS1 - энхансер

HS1, HS1' район определяющий распространенность в тканях

HS2 - HS6 район открывающий хроматин

Цифрами обозначены экзоны гена *Dad1*

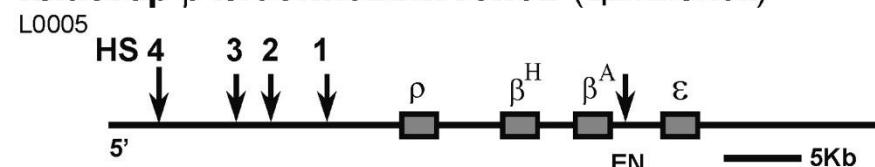
Кластер β глобиновых генов (человек)



HS5 - инсулатор

HS4 - HS1 - энхансер

Кластер β глобиновых генов (цыпленок)

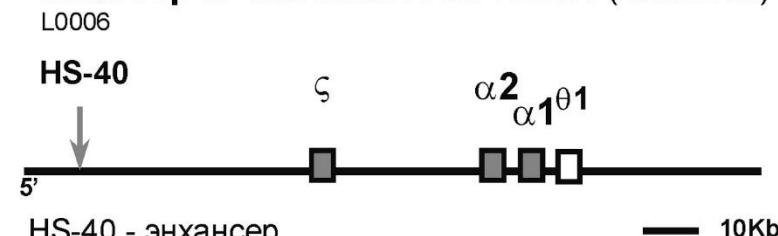


HS4 - инсулатор

HS3, HS2 - энхансер

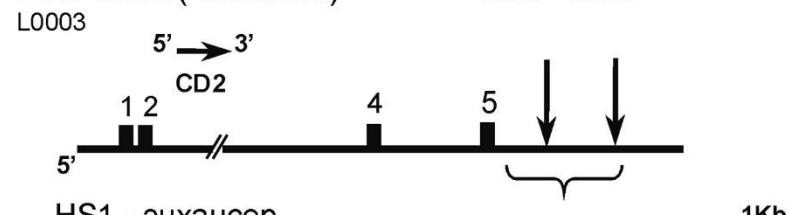
EN - межгенный энхансер

Кластер α -глобиновых генов (человек)



HS-40 - энхансер

Ген CD2 (человек)

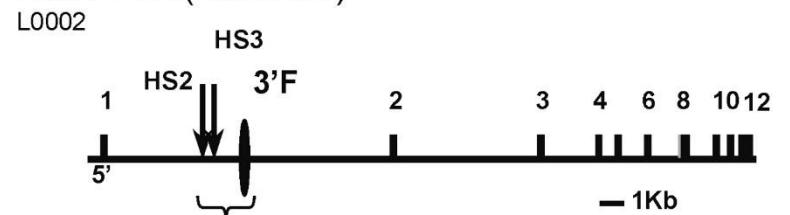


HS1 - энхансер

HS3 - инсулатор

Цифрами обозначены экзоны гена CD2

Ген ADA (человек)



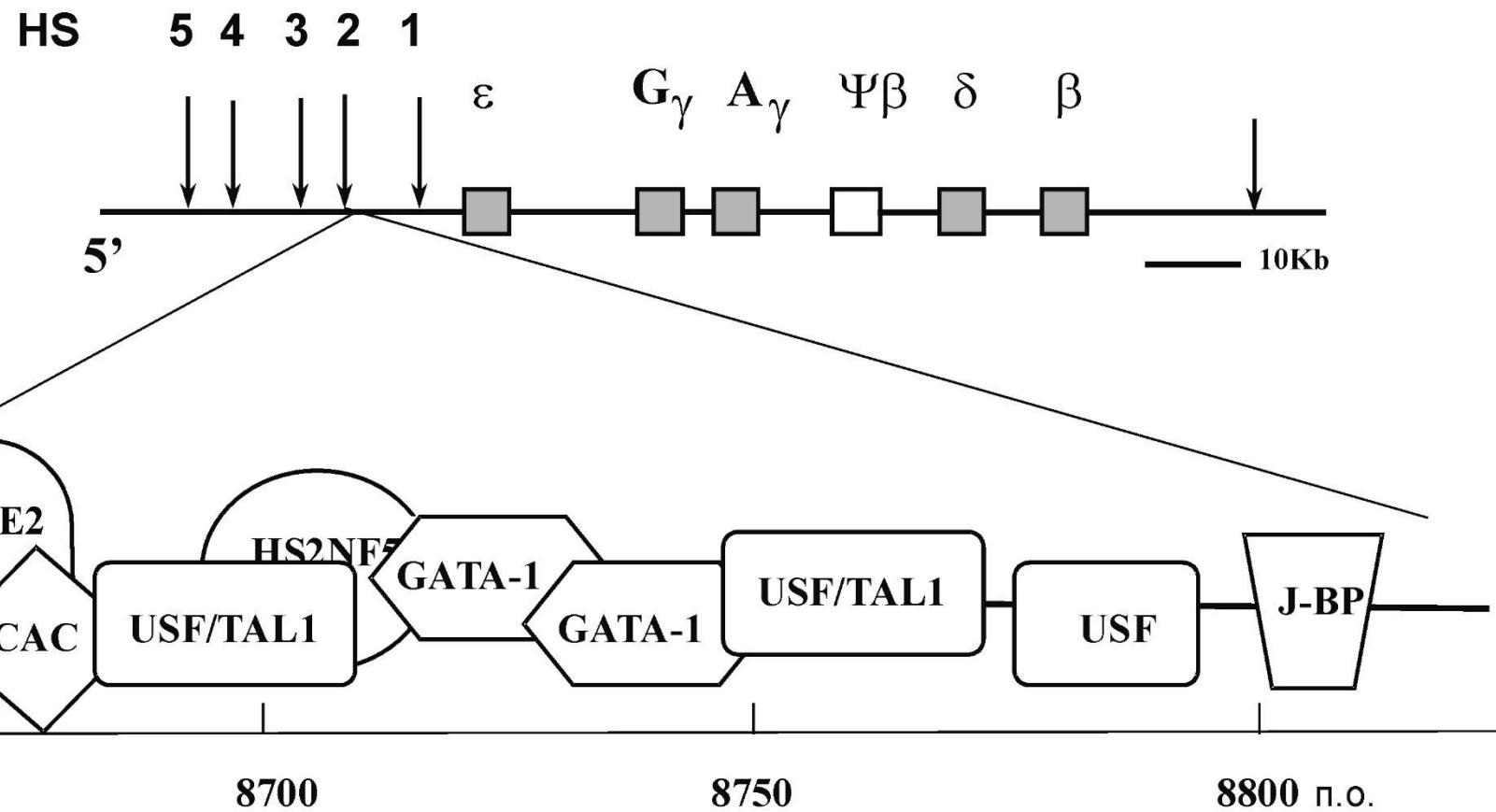
HS2 - 5'вспомогательный элемент

HS3 - энхансер;

3'F - 3'вспомогательный элемент

Цифрами обозначены экзоны гена ADA

LCR глобинового локуса человека

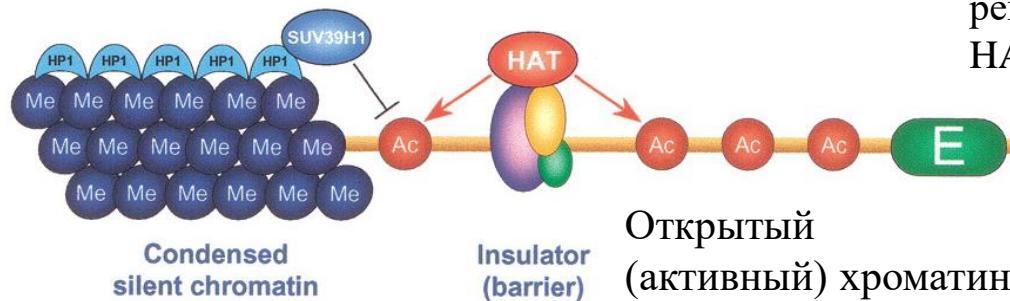


Каждый LCR включает специфическую группу сайтов связывания и располагается иногда на очень большом (до десятков тысяч п.о.) расстоянии от контролируемой кассеты генов

Инсулатор - участок ДНК, выполняющий специализированную регуляторную функцию

Функции инсулатора:

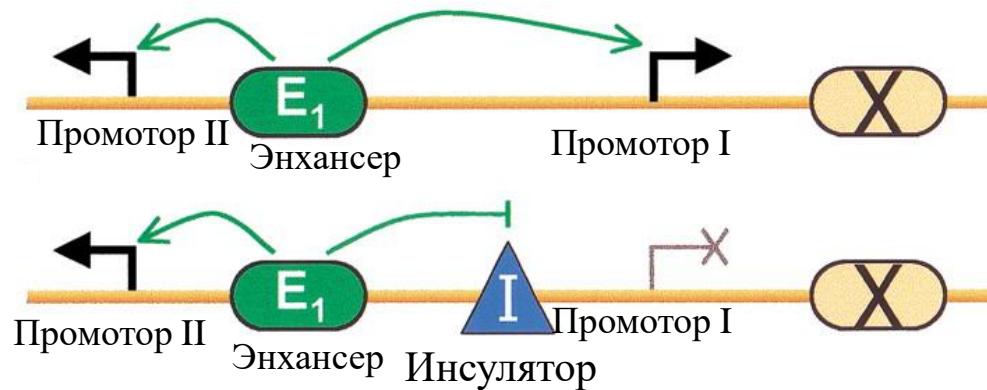
Инсулатор может располагаться на границе между открытым и закрытым хроматином и препятствовать распространению инактивирующего влияния конденсированного хроматина



Белки, взаимодействующие с инсулатором рекрутируют гистон-ацетилтрансферазы (HAT). HAT ацетилируют соседние нуклеосомы.

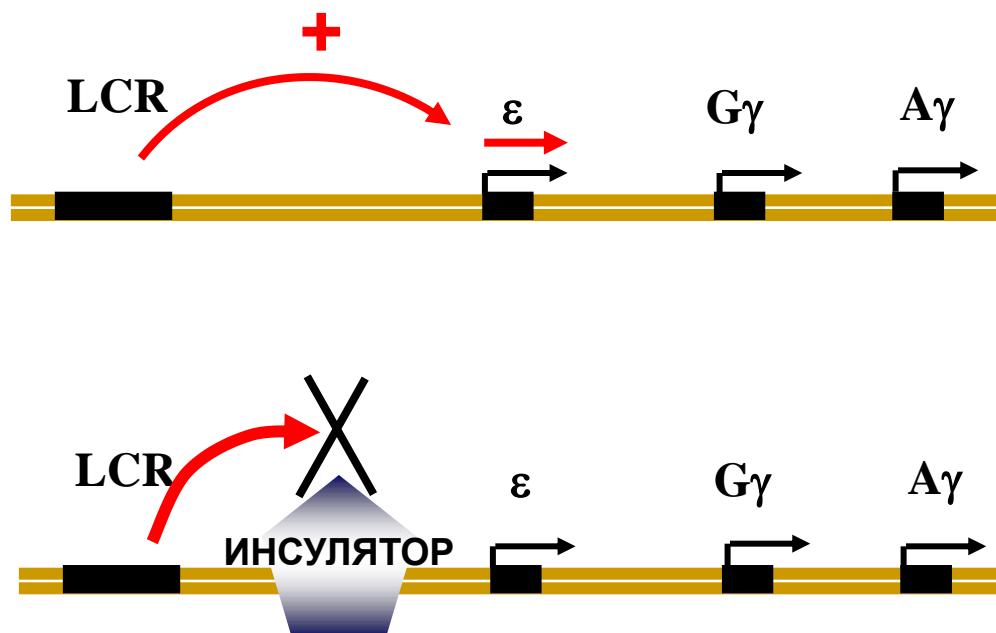
Ацетилирование нуклеосом препятствует распространению неактивного состояния хроматина, индуцируемого комплексом HP1/SUV39H1

Инсулатор может располагаться между энхансером и промотором и блокировать активирующее влияние энхансера на транскрипцию гена



Инсулатор блокирует активность энхансера только по отношению к промотору I. Если вместо инсулатора расположить сайленсер (негативную регуляторную единицу), то блокирующее влияние будет распространяться на оба промотора

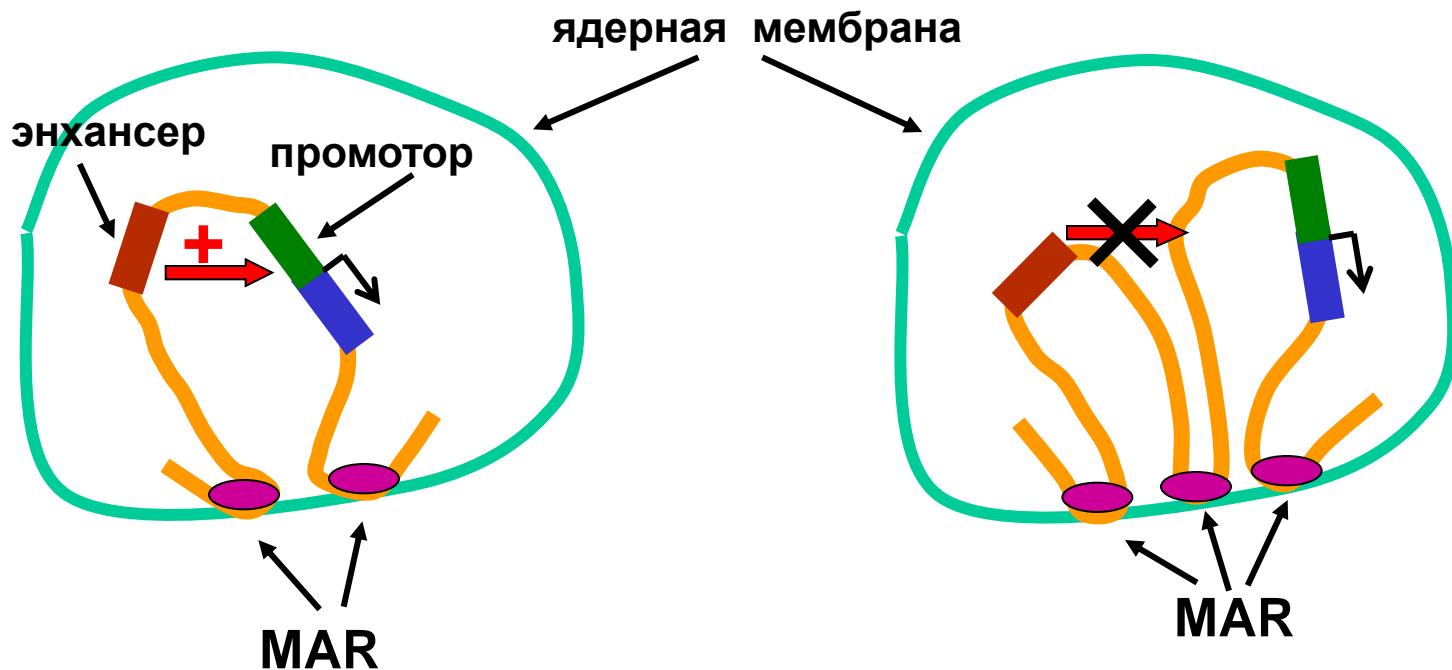
ИНСУЛЯТОР



Инсулатор может быть помещен между локус-контролирующими районом (LCR) и кластером регулируемых им генов. В этом случае регуляторное воздействие локус-контролирующего района будет блокировано

Роль инсулатора может выполнять участок прикрепления к ядерному матриксу (MAR). При включении такого инсулатора два регуляторных района оказываются в различных доменах, и не способны взаимодействовать

MARs =matrix attachment regions



энхансер активирует транскрипцию гена

энхансер не влияет на транскрипцию гена

Инсулатор - участок ДНК, выполняющий специализированную регуляторную функцию. Будучи помещенным между двумя регуляторными элементами может препятствовать активирующему либо подавляющему действию одного элемента на другой.

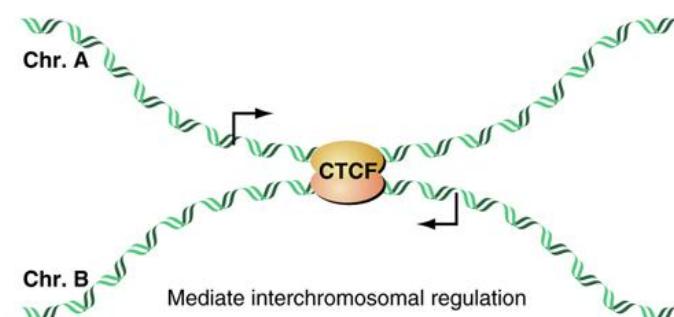
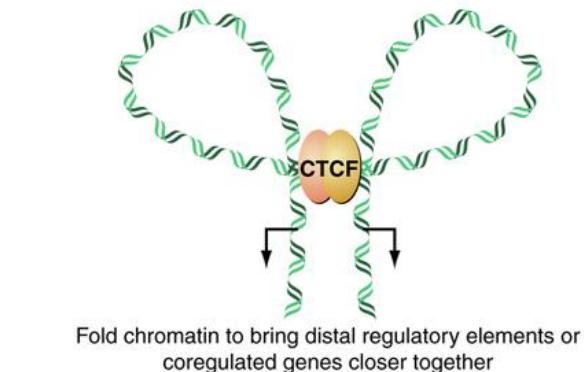
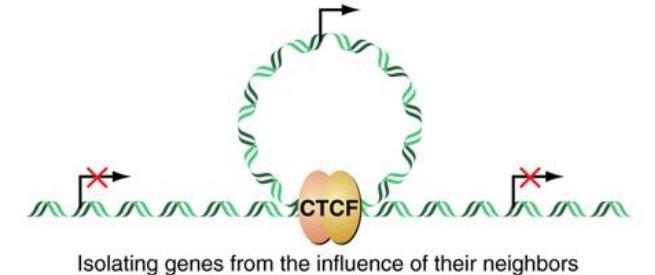
При включении инсулатора два регуляторных района оказываются в различных доменах, и не способны взаимодействовать

Роль транскрипционных факторов CTCF при формировании петель ДНК

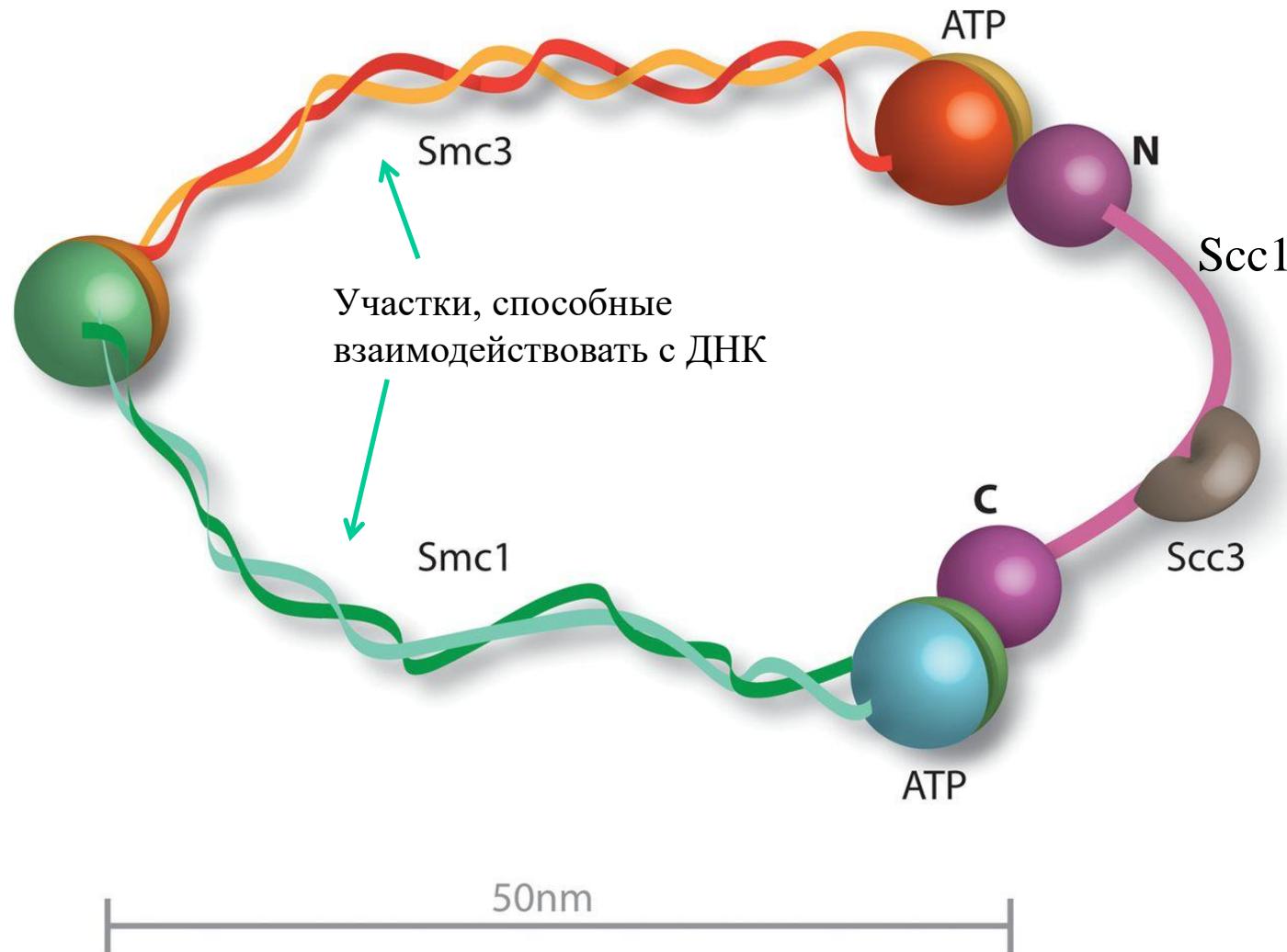
Функционирование инсулиторов тесно связано с наличием сайтов связывания транскрипционного фактора **CTCF** (CCCTC-binding factor).

CTCF - димерный фактор. Имеет пространственную структуру, обеспечивающую возможность взаимодействовать с различными нитями ДНК, за счет чего в ядре клетки могут формироваться петли ДНК либо межхромосомные контакты.

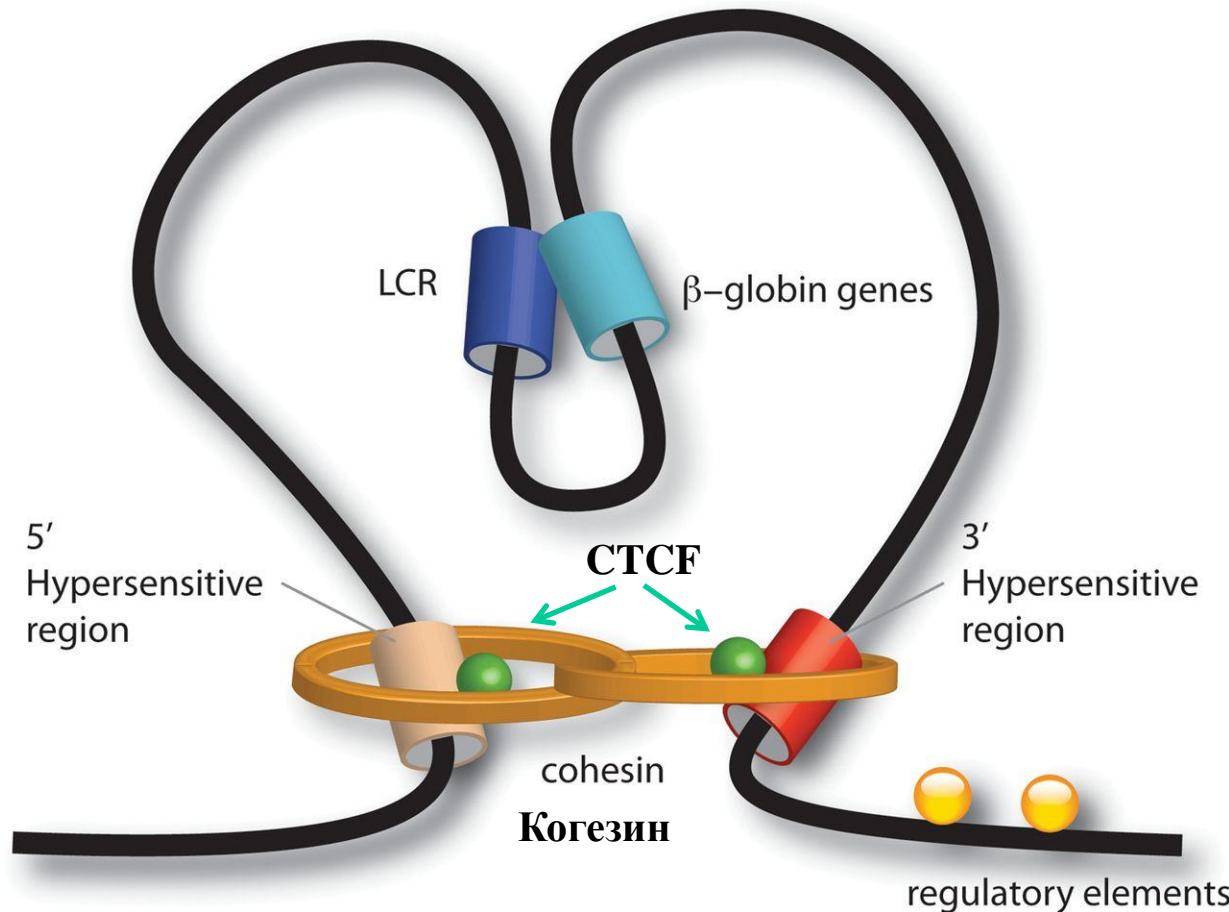
CTCF является маркером участков ДНК, разделяющих активный и репрессированный хроматин



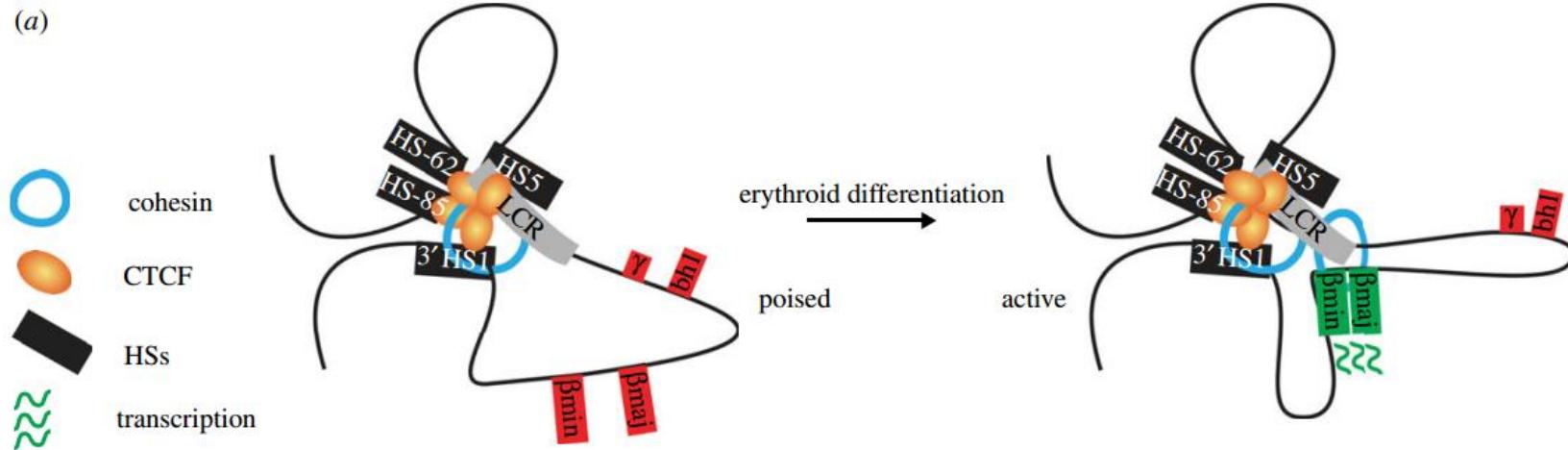
Когезин – комплекс белков , формирующих кольцеобразную структуру



Кластер генов β -глобинов: петлеобразная структура образуется благодаря взаимодействию белков CTCF и когезина



Роль транскрипционных факторов CTCF в регуляции экспрессии генов бета-глобинового кластера цыпленка



Активность генов кластера β -глобинов регулируется локус-контролирующим районом (LCR). На начальной стадии развития факторы CTCF взаимодействуют с ДНК и друг с другом таким образом, что образуется петля, включающая LCR и гены β -глобинов. В ходе дифференцировки клеток по эритроидному типу эритроид-специфичные транскрипционные факторы и белковый комплекс когезин модифицируют петлю ДНК таким образом, что LCR сближается с генами β -глобинов и активирует их транскрипцию.

Holwerda S.J., de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. Philos Trans R Soc Lond B Biol Sci. 2013 May 6;368(1620):20120369.

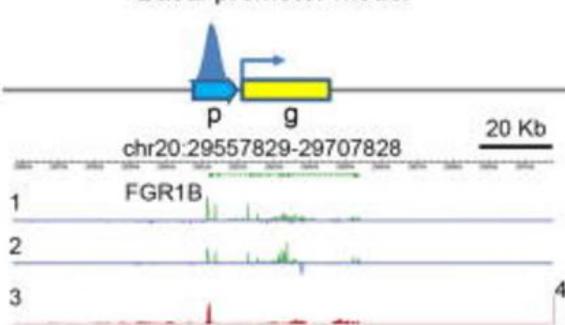
Модели взаимодействия между регуляторными участками генов в геноме человека.

Метод ChIA-PET (Chromatin Interaction Analysis by Paired-End-Tag sequencing) позволяет выявить контакты между участками хромосом
Исследованы контакты в клеточных культурах человека MCF7 и K562



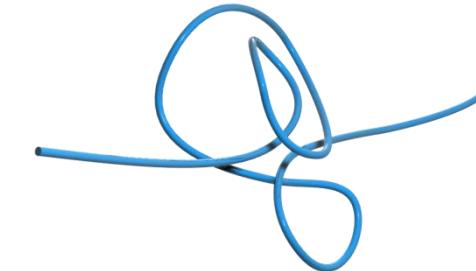
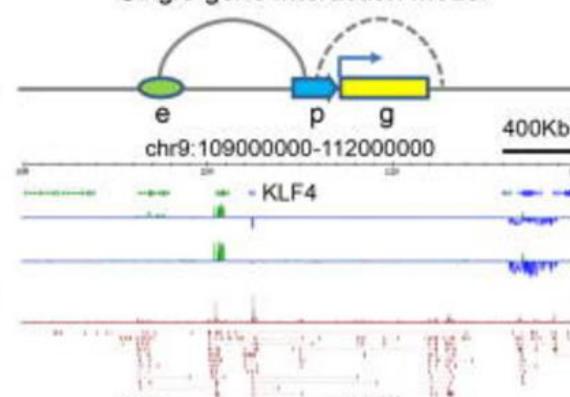
Промоторная модель

Basal promoter model



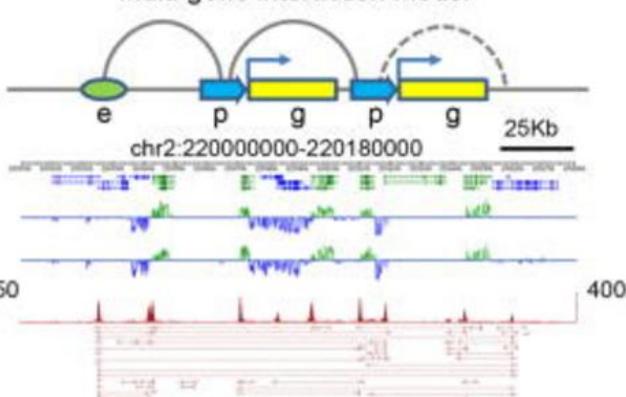
Одногенная модель
(Промотор-Энхансер)

Single-gene interaction model



Мультигенная модель (хромосомный оперон - “chromoperон”)

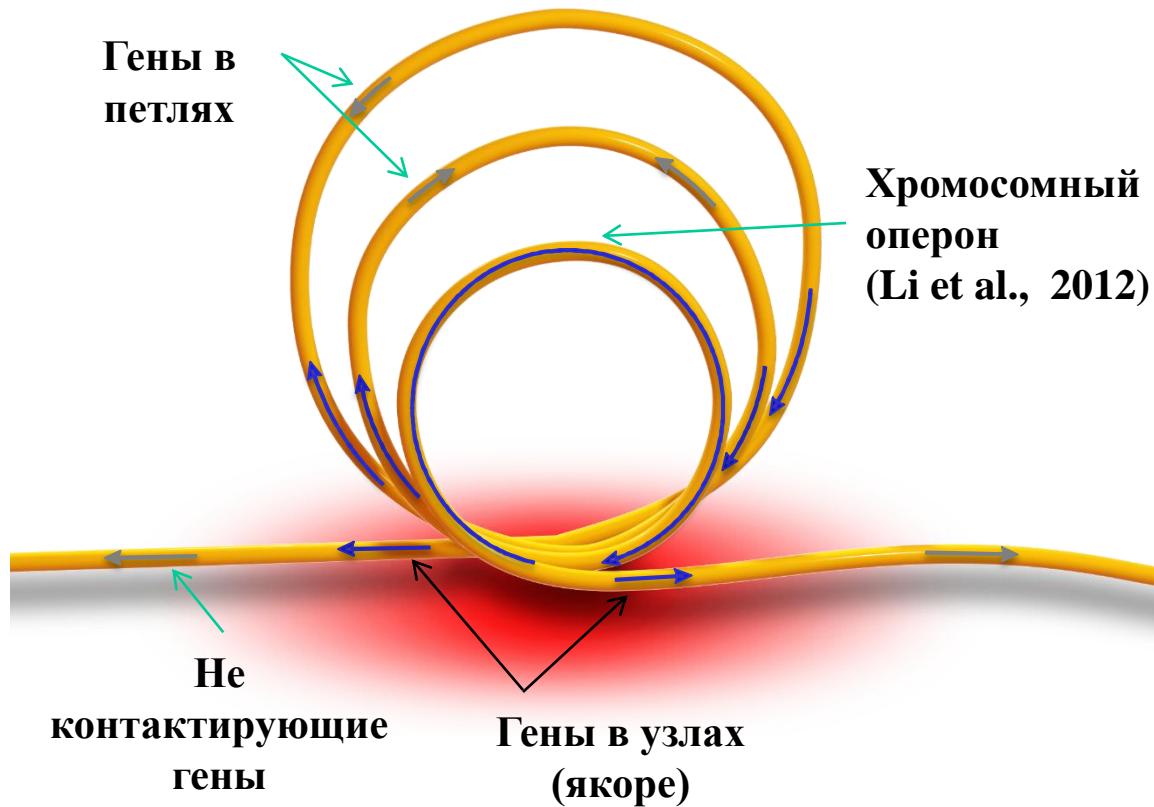
Multi-gene interaction model



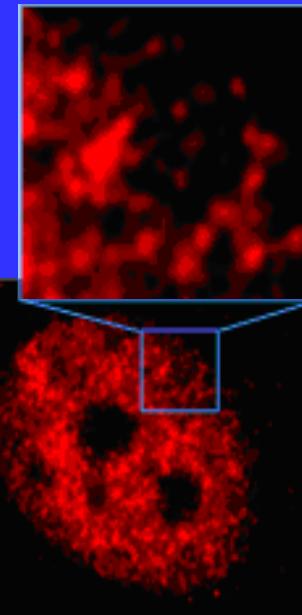
Li G. Et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

Хромосомные опероны и транскрипционные фабрики

Структуры, найденные с помощью ChIA-PET



Ядро клетки – участки транскрипции под микроскопом

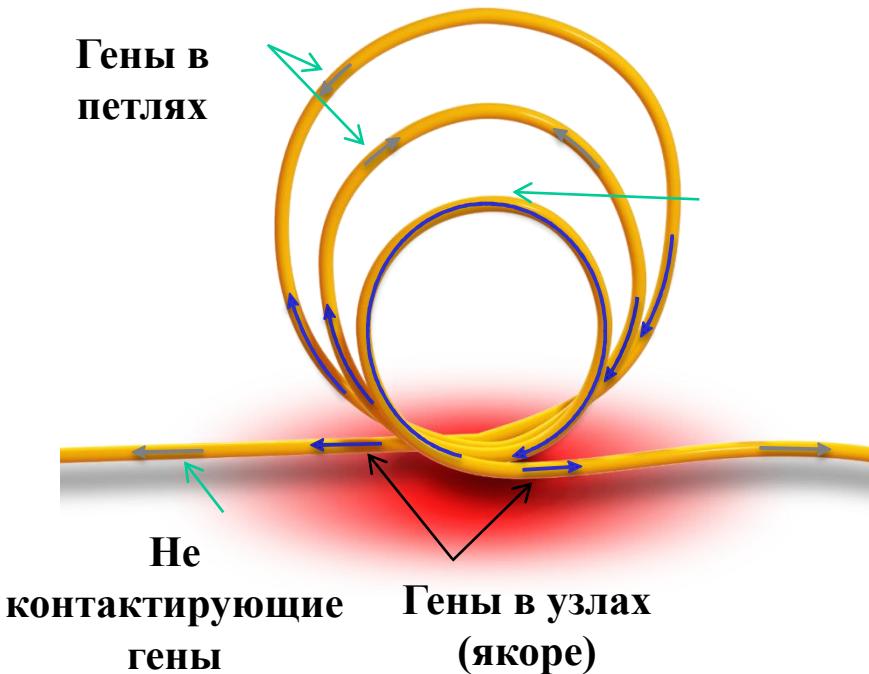


FISH

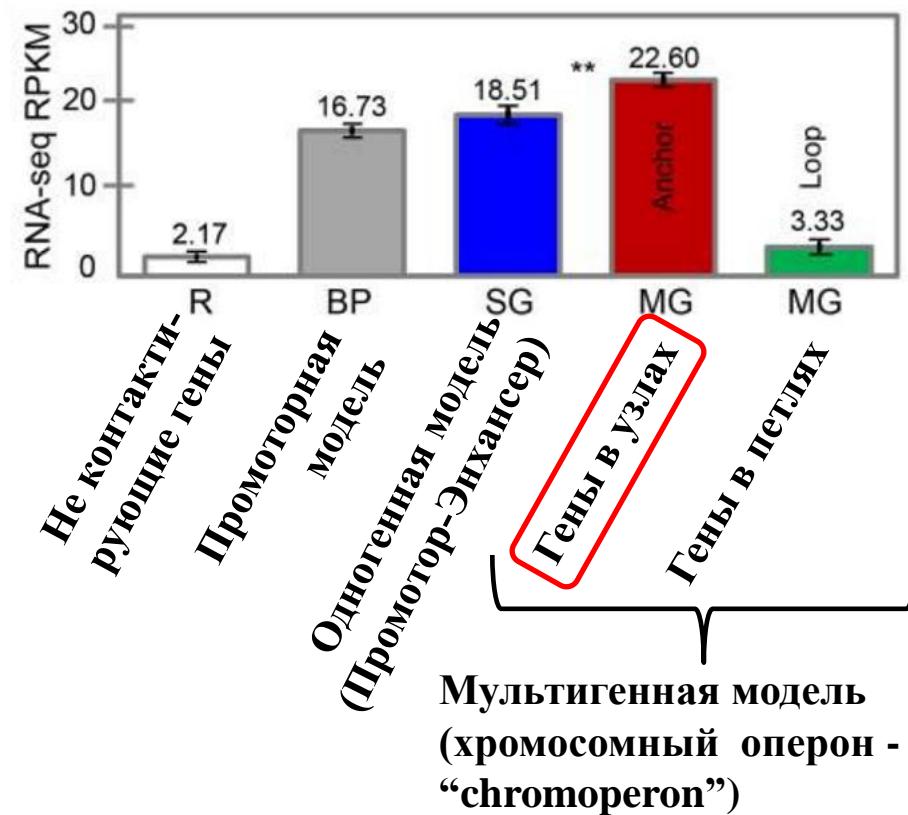
Хромосомный оперон – структура, в которой гены сближены благодаря петлеобразной укладке хроматина . Методами FISH можно пометить расположение транскрипционного комплекса в ядре клетке эукариот, тогда под микроскопом такие структуры в ядре выявляются как «транскрипционные фабрики»

Роль трехмерной структуры хроматина в регуляции транскрипции

Структуры, найденные с помощью ChIA-PET, в MCF7 клетках человека

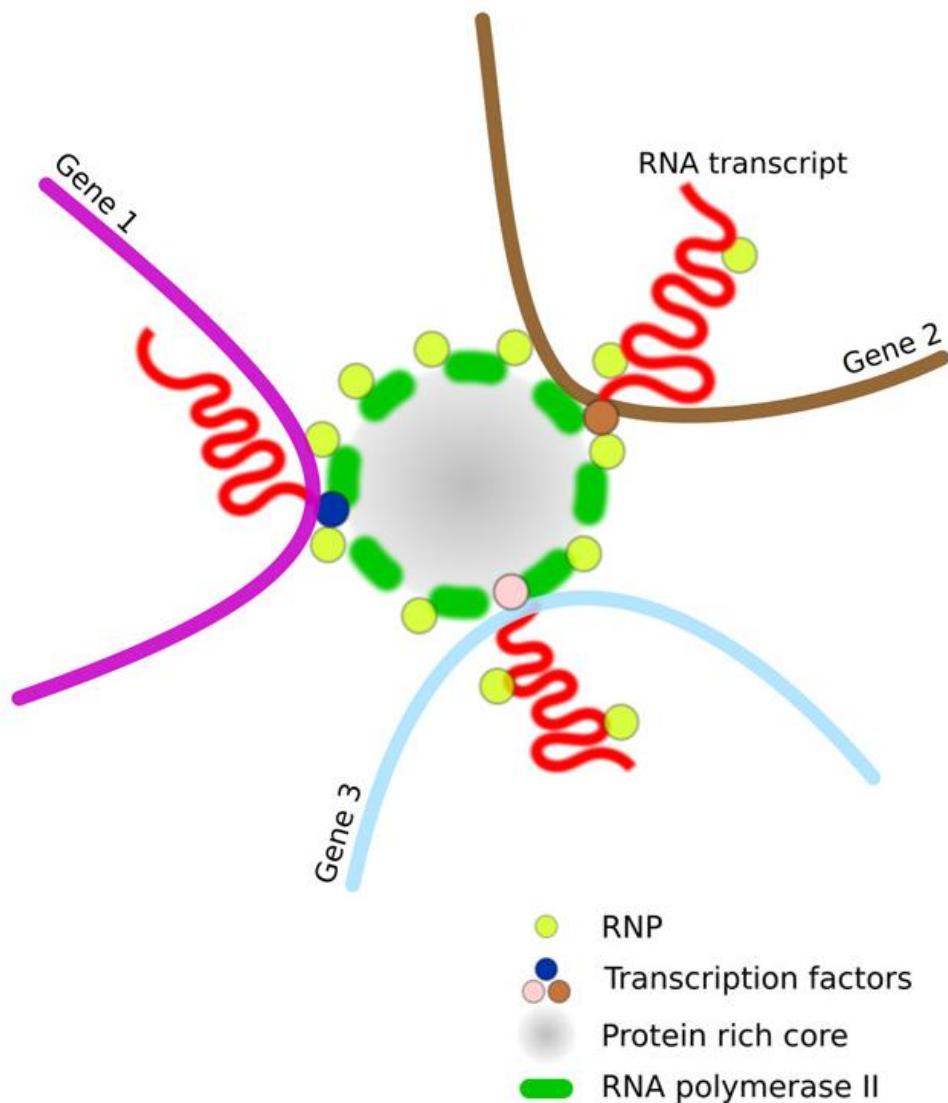


Транскрипционная активность генов в MCF7 клетках человека



Гены, расположенные в узлах хромосомных оперонов (мультигенная модель контактов), имеют достоверно более высокий уровень экспрессии по сравнению с генами из других групп (одногенная модель, промоторная модель и не контактирующие гены)

Транскрипционная фабрика (transcription factory)



Каждая «транскрипционная фабрика» может включать от 4 до 30 молекул РНК- полимеразы II, локализованных на поверхности белкового кора ($d \sim 87$ nm, in HeLa). Белковый кор фабрики содержит множество белков, участвующих в регуляции транскрипции: коактиваторы, белки, ремоделирующие хроматин, транскрипционные факторы, ферменты, модифицирующие гистоны, частицы RNP (рибонуклеопротеины), РНК-геликазы, факторы сплайсинга и процессинга.

При участии одной фабрики может осуществляться транскрипция нескольких генов. Размер фабрики может варьировать от 40 до 198 нм в зависимости от типа клеток, типа фабрики, и экспериментальных методов детекции и измерения.

Фабрики могут иметь специализацию, благодаря обогащению определенным транскрипционным фактором. Тогда они пространственно объединяют вместе несколько генов, регулируемых одним транскрипционным фактором.

Энхансеры и энхансерная РНК:

Определения и предыдущих лекций

Промоторный район, энхансеры, сайленсеры - РЕГУЛЯТОРНЫЕ ЕДИНИЦЫ

Энхансеры – регуляторные единицы, активирующие транскрипцию.

Энхансеры усиливают транскрипцию гена, управляемую определенным промотором, и оказывают свой эффект как в прямой, так и в обратной ориентации и в различной локализации (5'- либо 3'-) по отношению к промотору.

.....

Участок энхансера может транскрибироваться с образованием энхансерной РНК (eRNA)

Еще в 2010 году было показано, что....

Энхансеры связывают РНК-полимеразу II и продуцируют энхансерную РНК (eRNAs)



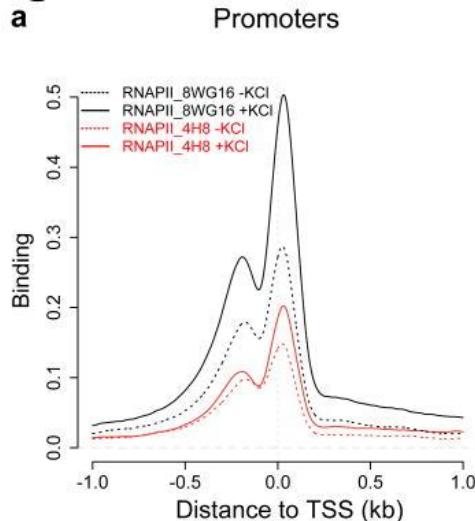
Данные экспериментов ChIPseq:
Количество пиков связывания P300/CBP в геноме мыши увеличивается с 1000 до 28000

Поиск нейрально –специфичных энхансеров согласно критериям:

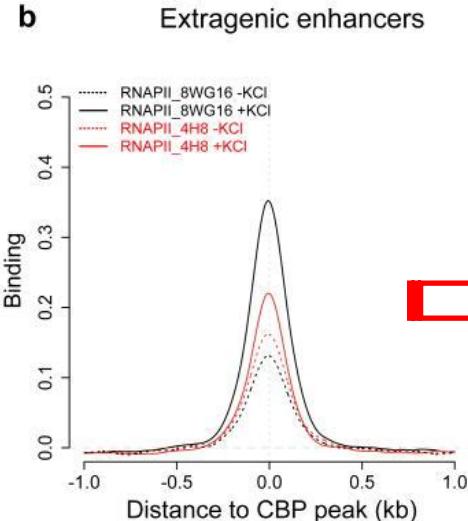
- 1) Участок генома, удаленный от старта транскрипции более чем на 1000 нуклеотидов
- 2) Имеет маркер хроматина H3K4me1, но не имеет маркер H3K4me3(маркер промотора)
- 3) Взаимодействуют с P300/CBP (контакты ДНК с P300/CBP опосредованы через другие белки)

Анализ связывания РНК полимеразы II (RNAPII)

Связывание РНК полимеразы II (RNAPII) в районах промоторов и энхансеров

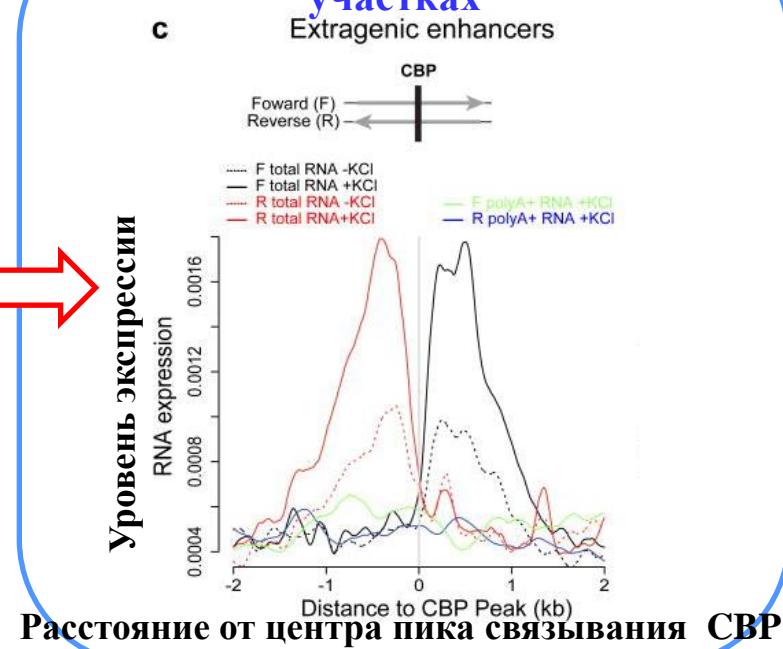


Ассиметричный пик – транскрипция идет в одном направлении



Симметричный пик – транскрипция идет в обоих направлениях (с + и – цепи ДНК)

Экспрессия РНК в энхансерных участках

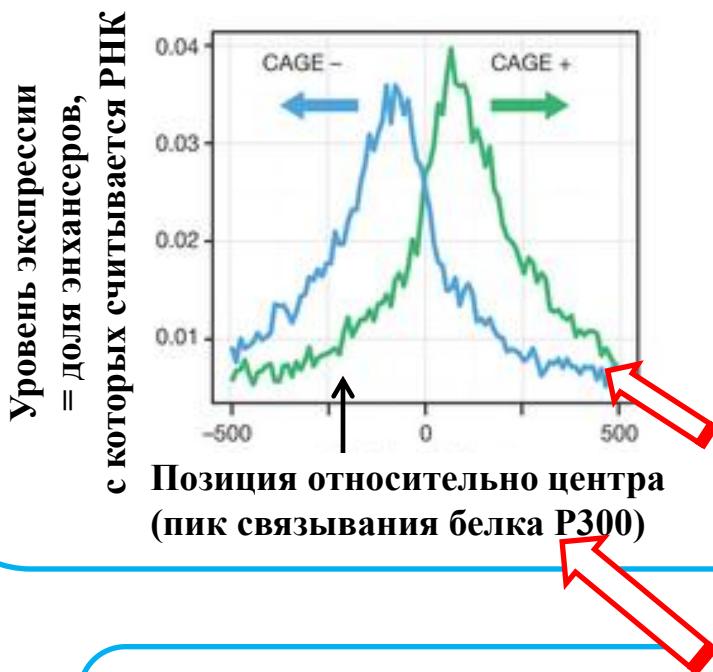


Уровень экспрессии

Расстояние от центра пика связывания CBP

Результаты проекта FANTOM (2014 г.), исследование проведено на клетках HeLa человека

Энхансеры инициируют экспрессию коротких (до 350 п.н) несплайсируемых РНК (eRNAs)



В геноме человека были определены участки, соответствующие энхансерам.

Критерием для выявления энхансеров были:

- (1) совместная встречаемость маркеров хроматина H3K27ac и H3K4me1 и
- (2) пиков связывания коактиваторного белка P300 (выявленных методикой ChIP-seq).

Обнаруженные таким образом энхансеры были центрированы относительно сайтов связывания P300. Для энхансеров были рассчитаны усредненные уровни экспрессии (по данным экспериментов CAGE).

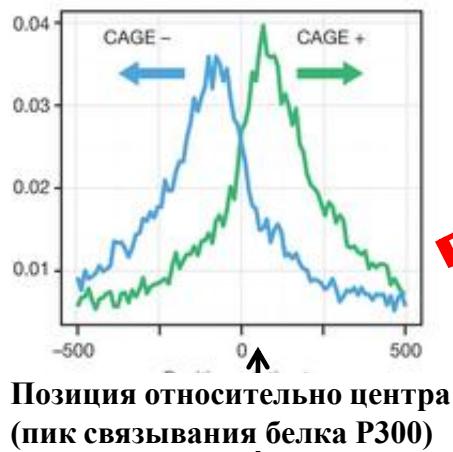
Были выявлены транскрипты (короткие РНК), которые считывались как с прямой (+), так и обратной (-) цепи ДНК в районе энхансера. Расстояние между (+) и (-) пиками соответствовало длине участка ДНК, упакованного в нуклеосому (180 п.о.).

В клетках HeLa человека выявлена энхансерная РНК (эРНК)

5'-фланкирующие участки транскрибуемых энхансеров содержат ТАТА боксы и INR элементы

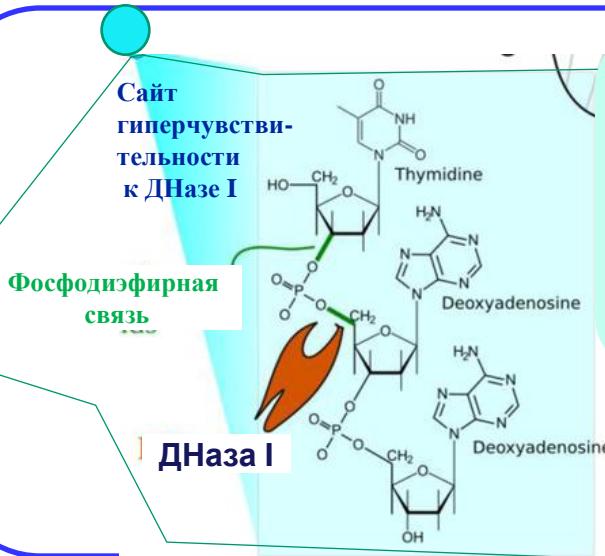
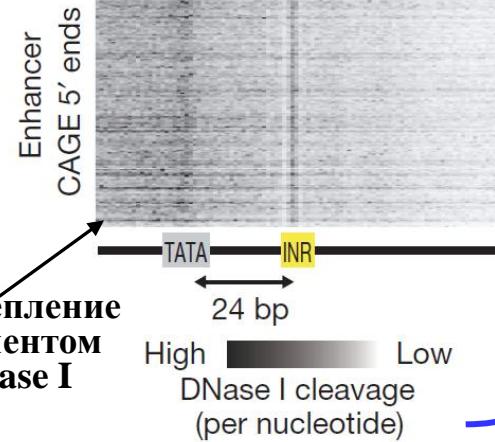
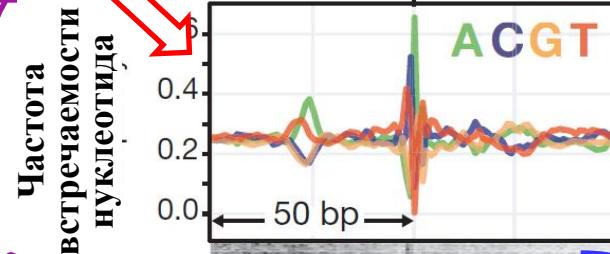
(Результаты проекта FANTOM, исследование проведено на клетках HeLa человека)

Уровень экспрессии
= доля энхансеров,
с которых считывается РНК



Анализ нуклеотидного контекста на 5'-фланкирующих участках транскрибуемых энхансеров выявил наличие ТАТА боксов и INR элементов.

INR – инициаторный элемент, можно описать консенсусным мотивом YYANWYY, где Y = C либо T



DNase I =
Дезоксирибонуклеаза I (ДНКаза I)
Наличие участков ДНК, гиперчувствительных к ДНКазе I, является характеристикой открытого хроматина

Позиции ТАТА боксов и INR элементов характеризовались наибольшей чувствительностью к DNase I (менее плотная нуклеосомная упаковка, наибольшая доступность для контакта с белками), что подтверждает их функциональную значимость.

Регуляторные функции энхансерной РНК:

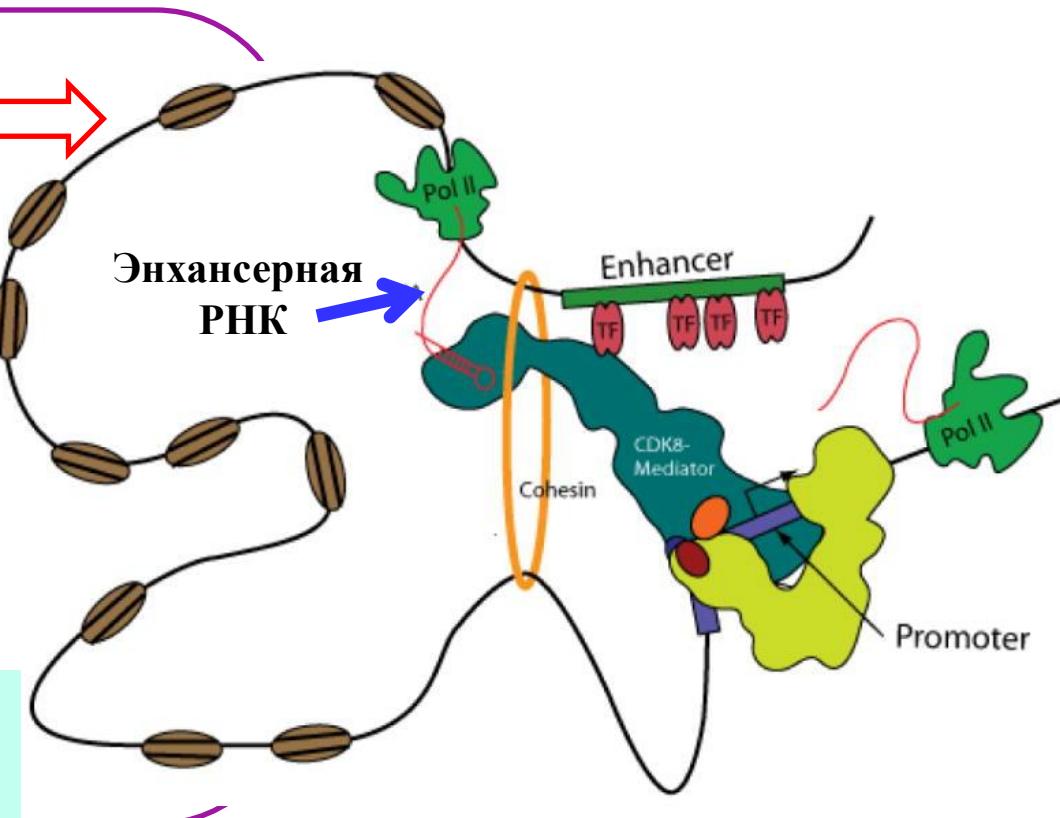
- Облегчает образование петель ДНК, сближающих энхансеры и промоторные районы.
- Способствует освобождению РНК-полимеразы из комплекса, обеспечивающего остановку (паузу) на стадии ранней элонгации
-

Роль энхансерной РНК:

Энхансерная РНК может облегчать образование петель ДНК, сближающих энхансеры и промоторные районы.

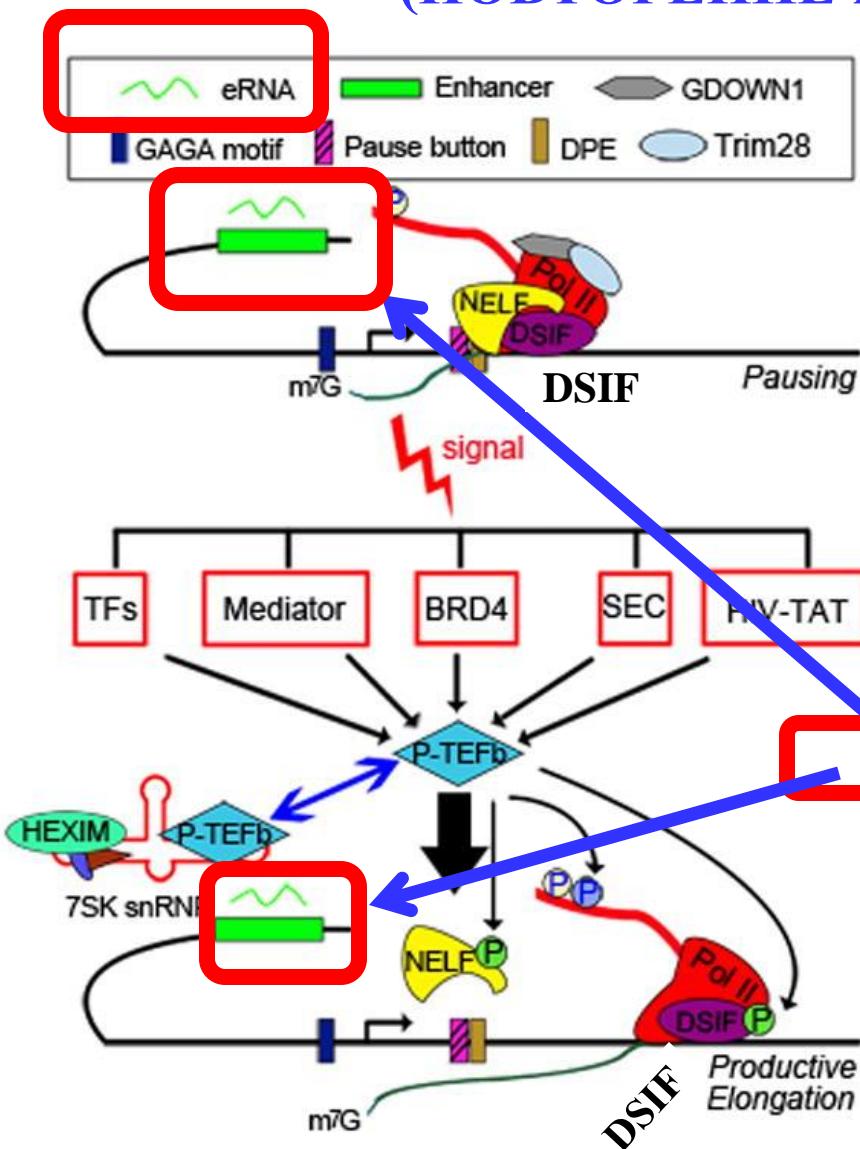
Показано, что в реализации этой функции участвует CDK8 модуль медиаторного комплекса, объединяющий субъединицы MED12, MED13, CCNC, CDK8.

Медиаторный комплекс включает у эукариот 31 субъединицу, которые обозначаются MED1, MED2, ...MED31, CCNC, CDK8. Является коактиватором транскрипции, взаимодействует с прединициаторным комплексом и транскрипционными факторами.



Для того, чтобы РНК полимераза II связалась с медиаторным комплексом, модуль CDK8 должен диссоциировать. При этом модуль CDK8 остается вблизи старта транскрипции, будучи связанным с энхансерной РНК. Это облегчает последующую реинициацию транскрипции (скаффолд комплекс)

Остановка РНК-полимеразы на стадии элонгации и ее освобождение регулируется множеством сигнальных путей и многими факторами (ПОВТОРЕНИЕ из предыдущих лекций)



У дрозофил пауза возникает особенно часто в генах, содержащих такие **регуляторные элементы**:

DPE - downstream promoter element

pause button

GAGA – сайт связывания GAGA-фактора (ТФ)

P-TEFb - positive transcription elongation factor

Может высвобождаться из
7SK-HEXIM inhibitory complex

P-TEFb активируется при участии:

BRD4 - Bromodomain-containing protein 4

SEC - super elongation complex

HIV-TAT - HIV-encoded transcriptional transactivator protein Tat

eRNAs – энхансерные транскрипты

Механизм участия эРНК см. на следующем слайде

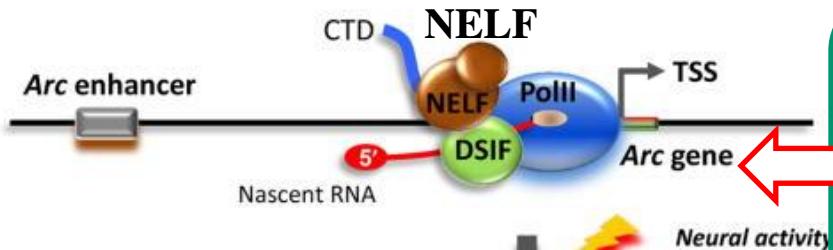
Окончание паузы происходит при:

- Тепловом шоке (ТФ)
- Гипоксии (ТФ)
- Воспалении (Nf-кB, BRD4)
- Дифференцировке стволовых клеток (SEC, ТФ)
- Инфекции HIV (HIV-TAT)

Роль энхансерной РНК:

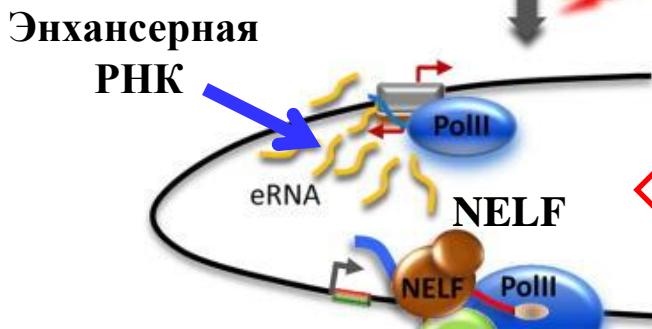
Взаимодействие энхансерной РНК с ингибитором элонгации NELF высвобождает РНК-полимеразу II из ингибиторного комплекса и активирует транскрипцию гена Arc

Ген Arc = Activity-regulated cytoskeletal protein

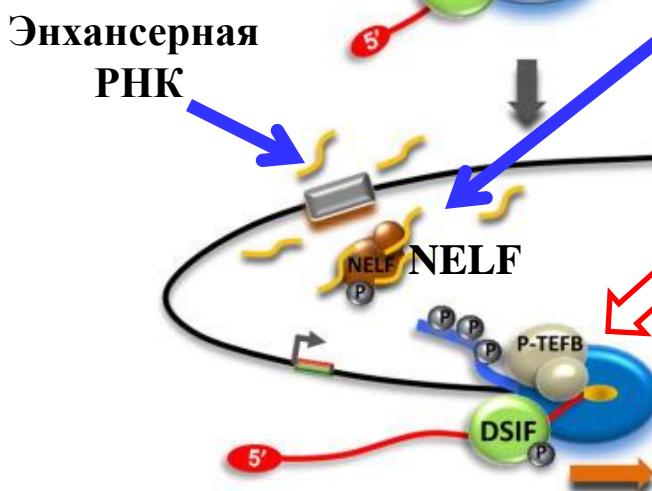


На стадии ранней элонгации происходит остановка РНК-полимеразы II.
Остановка происходит в результате связывания полимеразы с NELF и DSIF

НЕГАТИВНЫЕ РЕГУЛЯТОРЫ ЭЛОНГАЦИИ:
NELF = negative elongation factor
DSIF = DRB Sensitivity Inducing Factor



Энхансер гена Arc находится вблизи промотора.
При его активации нарабатывается энхансерная РНК.
эРНК взаимодействует с Е-субъединицей ингибитора элонгации NELF (с NELF-E) и вытесняет РНК полимеразу II из ингибиторного комплекса



Также рекрутируется фактор P-TEFb, который фосфорилирует РНК-полимеразу II и ингибиторные комплексы DSIF и NELF. Остановка РНК-полимеразы II завершается, транскрипция продолжается

АКТИВАТОР ЭЛОНГАЦИИ:
P-TEFb - positive transcription elongation factor

Лекция 4, Часть 2.

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»

The screenshot shows the Nucleic Acids Research journal website. At the top, there is a banner for the '2015 Web Server Issue'. Below the banner, there are sections for 'READ THIS JOURNAL' and 'THE JOURNAL'. A red arrow points from the text 'Ссылка на «online Database Collection»' to the '2015 Database Issue' link in the 'THE JOURNAL' section. Another red arrow points from the text 'Ссылка на «online Database Collection»' to the 'Database issue' section on the right. The 'Database issue' section features a large image of a DNA molecule and text about 56 new database descriptions and 115 updates. The URL for the 'Database issue' is <http://nar.oxfordjournals.org/volume/43/issue/D1>.

nar.oxfordjournals.org

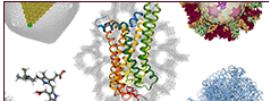
OXFORD JOURNALS

CONTACT US MY BASKET N

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Science & Mathematics > Nucleic Acids Research

 Nucleic Acids Research
Now available: the 2015 Web Server Issue

Click here to start reading

READ THIS JOURNAL

View Current Issue (Volume 43 Issue 17 30 September 2015)
Advance Access
Browse the Archive

NAR is an Open Access journal

NAR's Top Articles are updated monthly and show recent articles that have been most often accessed in HTML and PDF formats in the specified month and category.

NAR's Breakthrough Articles represent the very best papers in terms of originality, significance and scientific excellence.

NAR's F1000 Articles: the Faculty of 1000 provides evaluation and ratings of highly significant research articles.

2015 Web Server Issue

The 2015 Web Server Issue reports on 97 web servers, with a special emphasis on tools for emerging technologies such as synthetic biology, network and pathway analysis, analysis of high-throughput sequencing data and innovative visualizations. Other major categories include function prediction for genes and proteins; DNA and RNA related topics including sequence and structure, alignment, similarity search and motifs; and coding RNA, protein structure, protein domains and ligand binders.

 Nucleic Acids Research

THE JOURNAL

> About this journal
> NAR Methods online
> 2015 Database Issue **View this issue**

> NAR Special Collection
> Reference Form
> Rights & permissions
> Dispatches
> This journal's Impact Factor
> Publication Ethics
> view Recent Content
> Oxford Journals > Science & Mathematics > Nucleic Acids Research > Volume 43, Issue D1
> We are mobile
> Journals Careers

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

56 new database descriptions
115 updates

Database issue

Volume 43 Issue D1 28 January 2015

For checked items
 view abstracts download to citation manager

I. Nucleic acid sequence, structure, and regulation

Michael Y. Galperin, Daniel J. Rigden, and Xosé M. Fernández-Suárez
The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection
Nucl. Acids Res. (28 January 2015) 43 (D1): D1-D5 doi:10.1093/nar/ku1241

» Abstract » FREE Full Text (HTML) » FREE Full Text (PDF) » Database Summaries
» Permissions

« Previous | Next Issue »

This Issue
28 January 2015 43 (D1)

 Nucleic Acids Research

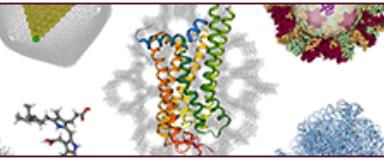
Open Access

Ссылка на «online Database Collection»

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Science & Mathematics > Nucleic Acids Research > Volume 43, Issue D1 > Article



Nucleic Acids
Now available: th

The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection

The NAR online Molecular Biology Database Collection in 2015
Michael Y. Galperin, Daniel J. Rigden, Xosé M. Fernández-Suárez

The collection includes databases previously described in *NAR*, but not represented in full articles in this issue, as well as selected other databases that are relevant to biologists. Hot-links are provided to all of the databases included in the compilation, as well as brief summaries of the content of each database.

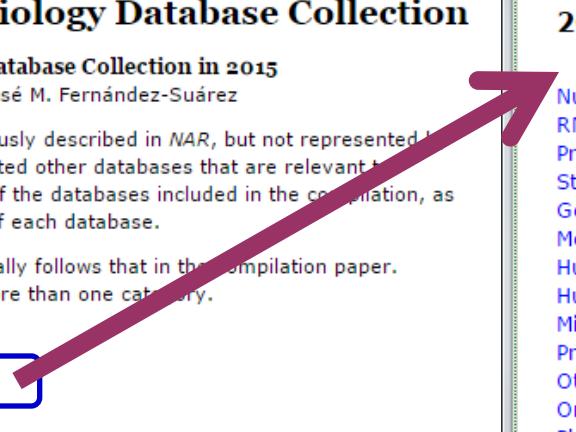
The category and database order generally follows that in the compilation paper. However, many databases appear in more than one category.

Category List
[Summary Paper List](#)
[Complete Category/Summary Paper List](#)
[Search Summary Papers](#)

Oxford Index BETA

About the Index

Имеется ссылка на полный список баз данных, сгруппированных по категориям .



Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper Category List

2015 NAR Database Summary Paper Category List

- [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](#)

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

Oxford University Press is not responsible for the content of external internet sites

Online ISSN 1362-4962 - Print ISSN 0305-1048

Информация о базах данных по регуляции транскрипции в «2015 NAR Database Summary Paper Category List»

The screenshot shows a list of 80 databases, each with a blue bracket on the left and the text '80 строкочек' (80 lines) written vertically on the right side of the list.

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS C

Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper

2015 NAR Database Summary Paper

Nucleotide Sequence Databases

- International Nucleotide Sequence Database Collaboration
- Coding and non-coding DNA
- Gene structure, introns and exons, splice sites
- Transcriptional regulator sites and transcription factors

- 3D-Footprint
- ABS
- ACTIVITY
- AGRIS - *Arabidopsis* Gene Regulatory Information Server
- AnimalTFDB
- ASPD
- BloodChIP
- ChIPBase
- cisRED
- CMGSDB
- CollecTF
- CoryneRegNet
- COXPRESdb
- CTCF Binding Site Database
- DBD
- DBTBS
- DBTSS
- DoOP - Databases of Orthologous Promoters
- DPInteract
- DPRP
- ECRbase
- EPD
- Factorbook
- FlyFactorSurvey
- FlyTF
- GeneNet
- GenomeTraFaC
- Greglist
- HOCOMOCO
- HTPSELEX
- JASPAR
- MachiBase
- MAPPER
- MPromDB
- ODB - Operon database
- OnTheFly
- ooTFD
- ORegAnno
- PAZAR
- PLACE
- Plant Stress-Responsive Gene Catalog
- PlantCARE
- PlantProm
- PRemod
- PRODORIC
- PromEC
- ProTISA
- QuadBase
- REDfly
- RegPrecise
- RegulonDB
- rSNP Guide
- ScerTF
- SCPD - *Saccharomyces cerevisiae* promoter database
- SELEXdb
- SKY/M-FISH and CGH
- STIFDB2
- SwissRegulon
- TcoF-DB
- Telomerase database
- TESS
- TFBSshape
- TFClass
- TiProD
- TRACTOR db
- TRANSCompel®
- TRANSFAC®
- TransfactomeDB
- TransmiR
- TRANSPATH®
- Transterm
- TRED - Transcriptional Regulatory Element Database
- TRRD
- TrSDB
- TTSMI
- UniPROBE
- VISTA Enhancer Browser
- WebGeSTER DB
- YEASTRACT
- YeTFaSCo

1. 3D-Footprint
2. ABS
3. ACTIVITY
4. AGRIS - *Arabidopsis* Gene Regulatory Information Server
5. AnimalTFDB
6. ASPD
7. BloodChIP
8. ChIPBase
9. cisRED
10. CMGSDB
11. CollecTF
12. CoryneRegNet
13. COXPRESdb
14. CTCF Binding Site Database
15. DBD
16. DBTBS
17. DBTSS
18. DoOP - Databases of Orthologous Promoters
19. DPInteract
20. DPRP
21. ECRbase
22. EPD
23. Factorbook
24. FlyFactorSurvey
25. FlyTF
26. GeneNet
27. GenomeTraFaC
28. Greglist
29. HOCOMOCO
30. HTPSELEX
31. JASPAR
32. MachiBase
33. MAPPER
34. MPromDB
35. ODB - Operon database
36. OnTheFly
37. ooTFD
38. ORegAnno
39. PAZAR
40. PLACE
41. Plant Stress-Responsive Gene Catalog
42. PlantCARE
43. PlantProm
44. PRemod
45. PRODORIC
46. PromEC
47. ProTISA
48. QuadBase
49. REDfly
50. RegPrecise
51. RegulonDB
52. rSNP Guide
53. ScerTF
54. SCPD - *Saccharomyces cerevisiae* promoter database
55. SELEXdb
56. SKY/M-FISH and CGH
57. STIFDB2
58. SwissRegulon
59. TcoF-DB
60. Telomerase database
61. TESS
62. TFBSshape
63. TFClass
64. TiProD
65. TRACTOR db
66. TRANSCompel®
67. TRANSFAC®
68. TransfactomeDB
69. TransmiR
70. TRANSPATH®
71. Transterm
72. TRED - Transcriptional Regulatory Element Database
73. TRRD
74. TrSDB
75. TTSMI
76. UniPROBE
77. VISTA Enhancer Browser
78. WebGeSTER DB
79. YEASTRACT
80. YeTFaSCo

В лекции № 4 (часть 2) будут рассмотрена база данных:

EPD

Eukaryotic Promoter Database

Geneva, Switzerland

EPD = Eukaryotic Promoter Database

<http://epd.vital-it.ch/#>

(объединенный ресурс, включающий две базы)



Philipp BUCHER (Swiss Institute for Experimental Cancer Research (ISREC), Lausanne, CH): Senior scientist and head of the bioinformatics group at the Swiss Institute for Experimental Cancer Research (ISREC)

The screenshot shows the EPD website interface. At the top, there are logos for SIB (Swiss Institute of Bioinformatics) and EPD (Eukaryotic Promoter Database). Below the header, there are links for Computational Cancer Genomics, ExPASy, and EPFL. A search bar is present with a dropdown menu set to "All databases" and a "SEARCH" button. A red banner on the left says "Access EPDnew". Below it is a list of organisms: *H. sapiens*, *M. musculus*, *D. melanogaster*, *A. mellifera*, *D. rerio*, *C. elegans*, *A. thaliana*, *Z. mays*, *S. cerevisiae*, *S. pombe*, Standard search, Select / Download, Promoter analysis tools, and an FTP site. Another red banner says "Access EPD". Below it are links for Promoter elements, Select / Download, and an FTP site. A third red banner says "Access MGA data". Below it are links for MGA Data Overview, MGA FTP site, Documents, Other Resources, and References. On the right, there is a news banner for 10-10-2016: "New liftOver option added to 'Select / Download' page". A detailed description of EPDnew follows, mentioning it is a collection of databases for selected model organisms (Human, Mouse, Fruitfly, Zebrafish, Worm, ThaleCress, Yeast, S. pombe) derived from TSS-mapping and CAGE/oligocapping. It lists the number of promoters for each organism: *Homo sapiens* (25503), *Mus musculus* (21239), *Drosophila melanogaster* (15073), *Apis mellifera* (6493), *Danio rerio* (10728), *Caenorhabditis elegans* (7120). A large red vertical bar on the right is labeled "EPDnew". At the bottom, a summary of the EPD database is provided, stating it is an annotated collection of eukaryotic promoters with determined transcription start sites, available via pointers to nucleotide sequence entries. It contains 4806 promoters from several species. A large red vertical bar on the right is labeled "EPD".

EPD разработана объединенными усилиями трех Швейцарских институтов. Ресурс включает базы EPD (Eukaryotic Promoter Database) и EPDnew.

EPD – первая часть объединенного ресурса

EPD (Eukaryotic Promoter Database) содержит информацию о 4806 промоторах генов эукариот, транскрибуемых РНК-полимеразой II. Описание промотора включает, помимо последовательности, ссылки на другие базы данных (в том числе, содержащие данные по экспрессии генов), название метода, с помощью которого идентифицирован промотор, информацию о наличии альтернативных промоторов, а также библиографические ссылки.

EPD: информационное содержание (download page)

Download the **complete promoter collection** for the following databases:

Download EPD ([refine selection](#))

- All promoters (4809)
 - Plant promoters (198)
 - Chromosomal genes (186)
 - Zea mays (maize) (21)
 - Prokaryotic plasmid DNA (8)
 - Viral genes (4)
 - Nematode promoters (26)
 - Arthropode promoters (2000)
 - Chromosomal genes (1991)
 - Drosophila melanogaster (fruit fly) (1926)
 - Transposable elements and retroviruses (5)
 - Viral genes (5)
 - Mollusc promoters (3)
 - Echinoderm promoters (44)
 - Vertebrate promoters (2540)
 - Chromosomal genes (2383)
 - Xenopus laevis (African clawed frog) (28)
 - Gallus gallus (chicken) (72)
 - Mus musculus (mouse) (196)
 - Rattus norvegicus (rat) (119)
 - Bos taurus (cattle) (24)
 - Homo sapiens (man) (1871)
 - Transposable elements and retroviruses (28)
 - Viral genes (129)
 - EBV (Human Epstein-Barr virus) (23)
 - HSV-1 (Human herpes simplex virus type 1) (48)

Preliminary EPD entries:

- Oryza sativa (rice) (13046)

+ 13046 промоторов риса (с пометкой «предварительные»)

4809 промоторов, аннотированных практически вручную на основании чтения статей . Наиболее полно представлены промоторы насекомых (из них 1926 входов для дрозофилы) и позвоночных (из которых 1871 входов для генов человека, а также 196, 119, 72, входов соответственно для мыши, крысы и цыпленка).

from to as [DOWNLOAD](#)

Опция для загрузки данных в фасти-формате (либо в формате EMBL)

Пример карточки из EPD

General information about the entry	
Entry name	HS_AK2
Entry type	standard
Promoter type	region
Accession number	EP74037
Description of the gene	Adenylate kinase 2.
Creation date	10-JAN-2003 (Rel. 73)
Last annotation	06-APR-2006 (Rel. 86)
Taxonomic division	VRT
Organism	Homo sapiens (human)
Keywords	Transferase, Kinase, ATP-binding, Mitochondrion, Alternative splicing.
Similarities with other entries	
Homology group	none.
Alternative promoter	none.
Neighbouring promoter(s)	none.
Cross References	
GENOME	NT_032977.8
CLEANEX	HS_AK2
DNA References	AL020995.14
SWISSPROT	P54819
REFSEQ	NM_001625 [DBTSS] NM_013411 [DBTSS]
MIM	103020
References	
[1]	MEDLINE= 10521335 Strausberg RL., Feingold EA., Klausner RD., Collins FS. The mammalian gene collection Science 286:455-457(1999).
Promoter-specific information	
Sequence	tgtgagccggcagttggacgtgcgtggcgtgcgtgcgttgcacctggaaGCACTGGACCT
Method(s)	Mammalian gene collection (MGC) full-length cDNA cloning [1].
Taxonomy	6. Vertebrate promoters 6.1. Chromosomal genes 6.1.7. Unclassified
Supplementary information Expression/Regulation:	
5-prime end distribution	

Общая информация о гене

Переcсылки на входы из других баз данных, содержащих информацию об этом гене.

Последовательность ДНК перед стартом транскрипции и до +10

Метод идентификации старта транскрипции

Текстовая выдача, полученная в ответ на запрос: описание промоторов и их нуклеотидные последовательности в фасти формате

```
>EP33026 (-) Bt TP2; range -200 to 50.
AGCTCCACCCGACCTGAGGGGCTGCTCTCAACCCACAGACACGCCCTTGAAAGCTGCC
CACCTGCGTGTAGGATGAGGGCAGAGGGCTTGTCCCCTGGGCCACATCTGTTACA
TACCCCCCTGTGGCCAGTGCCATCACAATCGGGCCAACATATAACCAGGGCTGCCAG
GGCCTCTGTGAAGCTGGGTCTGCCAGAACAGAGGAGGAGGAGGCGGGCCCTGCCCTCT
AAGCGAGGCCG
>EP28006 (+) Bt protamine P1^1.1; range -200 to 50.
GGCCCCACCCCCACACACATCACAGCCCCACCCCTGCACATCACAGCCCCGCCCTCCCTC
ACCAAGCACCTCCCACATGCCATATATGGGCATGATTGGCAGCTCTGACCCCTGGTCT
GTGAGGTCTGGGTCTGTGACCTCACAAATGACCAGGACCCCTGCCCGGGTCTATATAAGA
GGCCGGGAAGTCGGCCCTGTCACAGCCCACAAATTCCACCTGCTCACAGGTTGGCTGGC
TCAACCAAGGC
>EP15026 (+) Bt cytokeratin Ia; range -200 to 50.
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNCATGGTGGACGGCAAGTTATTCAAAGGAGAGCGGGACCGGGAGGTTGGCGGAA
ACGCAGGAGACCTCTGAAGCTCTGGCCAGAGGTGGCGTTATATAGGGCTGGGAGCTTG
GCTGGCTGGCG
>EP15027 (+) Bt cytokeratin Ib; range -200 to 50.
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TGAGTGAAGTCGGTGAAGGGAAAGTGCAAGAGTTAACAGGGCCACTCAGAGCCAGAGGA
AGAGAGAAGCAAACGTAAAGACCTGTGCAGGGATAAAATCCCTTATATACATCTAGGAGGCT
GCTGGGCTCAA
>EP15028 (+) Bt cytokeratin IV*; range -200 to 50.
CAAGGCTAGAAGCCAGAAGAATTCTCCATGACTAAAGGAAACCAAGAAGCAATATTCA
TACTTCATACCTTCTAGAGGCAGGGGGTGTCTCACTATTGTAAAGCCCAGCCCTTC
TAATCTGCAGGCTCACCTCCAGGACTGAGCCCGGGCCATTTCATATATAAGCTGC
TGCCGGGCCGCCCCCTATAGATCTGTTCTTAGCTCTGCTTCCACCTCTCACACCCCTC
TCAACCTATT
```

1-ая строка - общая
информация о
последовательности

2-ая строка –
последовательность ДНК

EPDnew - вторая часть объединенного ресурса

The screenshot shows the EPDnew database homepage. On the left, a sidebar lists species: *H. sapiens*, *M. musculus*, *D. melanogaster*, *A. mellifera*, *D. rerio*, *C. elegans*, *A. thaliana*, *Z. mays*, *S. cerevisiae*, and *S. pombe*. Below this are buttons for 'Access EPD', 'Promoter elements', 'Select / Download', and 'FTP site'. The main content area is titled 'EPDnew databases' and describes the database as a series of species-specific databases of experimentally validated promoters. It mentions that 10 species are supported: *H. sapiens*, *M. musculus*, *D. melanogaster*, *A. mellifera*, *C. elegans* and *D. rerio*, 2 plants (*A. thaliana* and *Z. mays*) and 2 fungus (*S. cerevisiae* and *S. pombe*). Evidence comes from TSS-mapping from high-throughput experiments such as CAGE and Oligocapping. Below this is a row of buttons for 'EPD new Human', 'EPD new Mouse', 'EPD new Fruitfly', 'EPD new Zebrafish', 'EPD new Worm', 'EPD new ThaleCress', 'EPD new Yeast', and 'EPD new S. pombe'. Further down are sections for 'Collection accessibility' and 'Reference', with a link to a detailed description of the principles governing EPDnew.

Информационное содержание

(10 видов организмов)

Homo sapiens

Mus musulus

D. melanogaster

C. elegans

Apis mellifera (honey bee)

D. rerio (Zebrafish)

Животные

A. thaliana

Z. Mays

Растения

S. Cerevisiae

S. pombe

Дрожжи

Данные о позициях промоторов получены на основе высокопроизводительных методик
CAGE (cap analysis gene expression)
= Кэн-анализ экспрессии генов)
и Oligocapping

Данные экстрагированы из Интернет-доступных ресурсов:

- Научных публикаций с сапплиментами
- Геномного браузера UCSC,
- Проекта FANTOM5

EPDnew – ИСТОЧНИКИ ДАННЫХ

Table 2. Source data

EPDnew database	Source data: type, reference or source repository	# of libraries	total tags (millions)
<i>H. sapiens</i>	CAGE from ENCODE/RIKEN, downloaded from UCSC genome browser database (12)	148	3841
<i>M. musculus</i>	CAGE from FANTOM5 (http://fantom.gsc.riken.jp/5/)	339	6236
<i>D. melanogaster</i>	CAGE from modENCODE (ftp://data.modencode.org/) TSS-seq from Machibase (13)	57	646
<i>D. rerio</i>	CAGE from Nepal et al. (14), downloaded from SRA (8), ID SRA055273	12	65
<i>C. elegans</i>	GRO-cap from Kruesi et al. (15)	8	236

EPDnew, the *Homo sapiens* (human) curated promoter database

Overview

Version: 004

Coverage: 25503 promoters, 17785 genes

Genome assembly: GRCh37 / hg19

Gene annotation UCSC known genes (30-Jun-2013)

Based on data from: Riken/ENCODE CAGE data downloaded from UCSC, FANTOM5 data, EPD (old)

EPDnew, the *Danio rerio* (zebrafish) curated promoter database

Overview

Version: 001

Coverage: 10728 promoters, 10235 genes

Genome assembly: Zv9 / DanRer7

Gene annotation UCSC known genes (30-Jun-2013)

Based on data from: Nepal et al, Genome Res. 2013, PubMed PMID: 24002785, EPD (old)

Oxford Journals > Science & Mathematics > Nucleic Acids Research > Volume 43, Issue D1 > Pp. D92-D96.



The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools

René Dreos¹, Giovanna Ambrosini^{1,2}, Rouyda Cavin Périer² and Philipp Bucher^{1,2,*}

¹ Author Affiliations

² To whom correspondence should be addressed. Tel: +41 21 6930956; Fax: +41 21 693 1850; Email: philipp.bucher@epfl.ch

« Previous | Next Article »
Table of Contents

This Article

Nucl. Acids Res. (28 January 2015) 43 (D1): D92-D96.
doi: 10.1093/nar/gku1111

First published online: November 6, 2014

This article appears in: Database issue

EPDnew: возможность просмотра данных о промоторе гена в графическом виде (EPDnew viewer) на примере промотора гена MAPK1_1:

General information:

Entry name: MAPK1_1
Promoter type: region
Organism: Homo sapiens (Human)
Gene Symbol: MAPK1
Description of the gene: mitogen-activated protein kinase 1
Sequence: cctctggccggccctccctccggccggccggccgtcAGTCTGGCAGG
Position in the genome: Chromosome [NC_000022.10]; Strand [-]; Position [22221935]
Ensembl: ENSG00000100030
RefSeq: NM_002745
NCBI Gene: MAPK1
GeneCards: MAPK1

External resources:

EPD view at UCSC (hg19): [chr22_22219935-22223935](#)
EPD view at UCSC (hg18): [chr22_20549935-20553935](#)
SwissRegulon: [chr22_22219935-22223935](#)
FANTOM4: [chr22_22219935-22223935](#)
MPromDB: [chr22_22219935-22223935](#)

Promoter image:

Scale: chr22: 22,220,500 22,221,000 22,221,500 1 kb 22,222,000 22,222,500 hg19 22,223,000 22,223,500

469 K562 H3K4me1

3111 K562 H3K4me3

1230.5 K562 Pol2

2320 A11 CAGE (+)

12 (+) A549 CAGE
(+/-) CD29 CAGE
(+) GM12878 CAGE
(+) H1hesc CAGE
(+) HEK293 CAGE
(+) HEK293 CAGE
(+) HUVEC CAGE
(+) K562 CAGE
(-) A549 CAGE
(-) CD29 CAGE
(-) GM12878 CAGE
(-) H1hesc CAGE
(-) HUVEC CAGE
(-) K562 CAGE

-11 A11 CAGE (-)

-46760 MAPK1_1

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

CpG Islands (Islands < 300 Bases are Light Green)
HMR Conserved Transcription Factor Binding Sites
VSP21_81
MvY1_82
100 vertebrates Basewise Conservation by PhyloP

4.88

Общая информация о гене

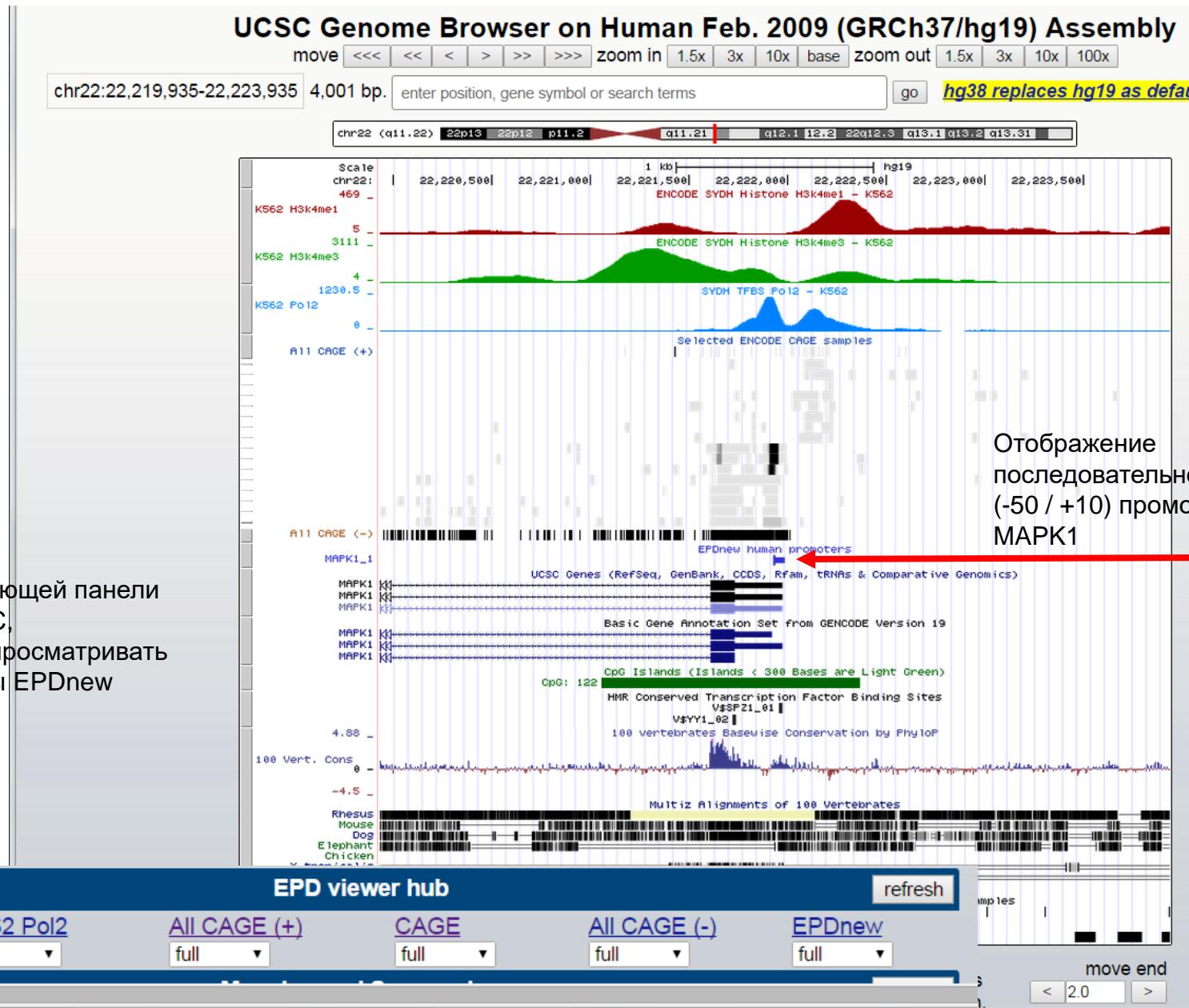
Последовательность ДНК - 50 / +10

Переcсылки на входы из других баз данных, содержащих информацию об этом гене

Характеристики генома в окрестностях старта транскрипции гена MAPK1

Отображение последовательности ДНК -50 / +10

Исходный вариант представления данных о промоторе гена МАРК1 в геномном браузере UCSC (Калифорнийский университет, г.СантаКруз)



Опции управляющей панели браузера UCSC, позволяющие просматривать данные из базы EPDnew

EPDnew: опции поиска и анализа данных



EPD
EUKARYOTIC PROMOTER DATABASE

Computational Cancer Genomics | ExPASy | EPFL

Access EPDnew

[H. sapiens](#)
[M. musculus](#)
[D. melanogaster](#)
[D. rerio](#)
[C. elegans](#)
[A. thaliana](#)
[Standard search](#)
[Select / Download](#)
[Promoter analysis tools](#)
[FTP site](#)

Access EPD

[Promoter elements](#)
[SRS access to EPD](#)
[Select / Download](#)
[FTP site](#)

Access MGA data

[MGA Data Overview](#)
[MGA FTP site](#)

Documents

Other Resources

References

What is new

EPDNew human version 003

Overview

Coverage: 23360 promoters, 16599 genes
Genome assembly: GRCh37 / hg19
Gene annotation: UCSC known genes (30-Jun-2013)
Based on data from: Riken/ENCODE CAGE data downloaded from UCSC
EPD (old)
Documentation files: [Promoter assembly pipeline description](#)
[Statistics and quality control report](#)

Promoter Selection and Analysis tools

Various tools allow you to analyse promoters from EPD and/or to select subsets of promoters. In order to analyze the complete EPD promoter set, go directly to one of the analysis pages. If you prefer to first select a subset of promoters, go to one of the selection pages. From the output of the selection pages you can then directly navigate to one of the analyses pages, or you can continue with another selection page to refine your promoter selection.

Selection tools

- [EPD selection tool](#): Promoter subset selection based on EPD-supplied annotation.
- [ChIP-Cor](#): Promoter subset selection based on experimental data or genome annotation from the MGA repository. Example: select promoters that have more than 100 H3K4me3 between -100 and +100 relative to the TSS.
- [FindM](#): Promoter subset selection based on DNA motif occurrences. Example: select promoters that have (or don't have) a c-Myc binding site between -100 and +100 relative to the TSS.

Analysis tools

- [ChIP-Cor](#): Generation of an aggregation plot (feature correlation plot) for a specific chromatin feature or genome annotation. Example: Distribution of nucleosomes (MNase-seq tags) near promoters, e.g. from -1000 to +1000 relative to the TSS.
- [OProf](#): Generate a motif occurrence profile around TSS positions. Example: Generate a plot showing the occurrence frequency of TATA-boxes between -100 to +100 relative to the TSS.
- [FindM](#): Extract DNA motif positions near transcription start sites. Example: extract coordinates of CCAAT-boxes located between -150 and -50 relative to a TSS. The output is a set of CCAAT-box positions that can be further analysed in the same way as a set of TSS positions.

Поиск по названию гена

Выбор группы генов по
характеристикам промотора: (1)
задокументированным в EPD; (2)
Экспериментальным данным Chip-
seq; (3) присутствию мотивов
(сайтов связывания)

Опции для
анализа

Поиск промоторов человека с заданными свойствами:

EPDnew selection tool

Use this tool to **select all promoters** (leaving all 'Optional criterias' blank) or restrict them based on all or some of their genomic contexts (such as presence of core promoter elements) or expression levels. After selection, you can **download** them in various format (for example in FASTA, BED, etc..) or **liftOver** them to a different assembly or use them to perform **further analysis** such as motif enrichment/search and chromatin status.

Select promoters for **H. sapiens**

Optional criteria:

- with TATA-box motif
- AND with Initiator motif
- AND with CCAAT motif
- AND with GC motif
- AND with CpG island
- AND marked as single
- AND average expression of at least tags
- AND expressed in at least samples

Select

Database:		Selection Parameters
Database: human_epdnew		TATA-box: with
Assembly: hg19		Initiator: with
		CCAAT-box: with
		GC-box: with
		Marked as: single
		Average expression:
		Expressed in:

Results: 3 promoters selected [SGA file](#) [EPS file](#) [BED file](#)

LiftOver options: hg38 (Dec 2013 GRCh38) **Submit**

Sequence Extraction Tool (FASTA format)

From: To:
Submit

Downstream Analysis

Motif Enrichment [OProf](#) 
Motif Discovery [FindM](#) 
Chromatin analysis [ChIP-Cor](#) 

Конец 4-ой лекции