

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. Инсуляторы;
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)

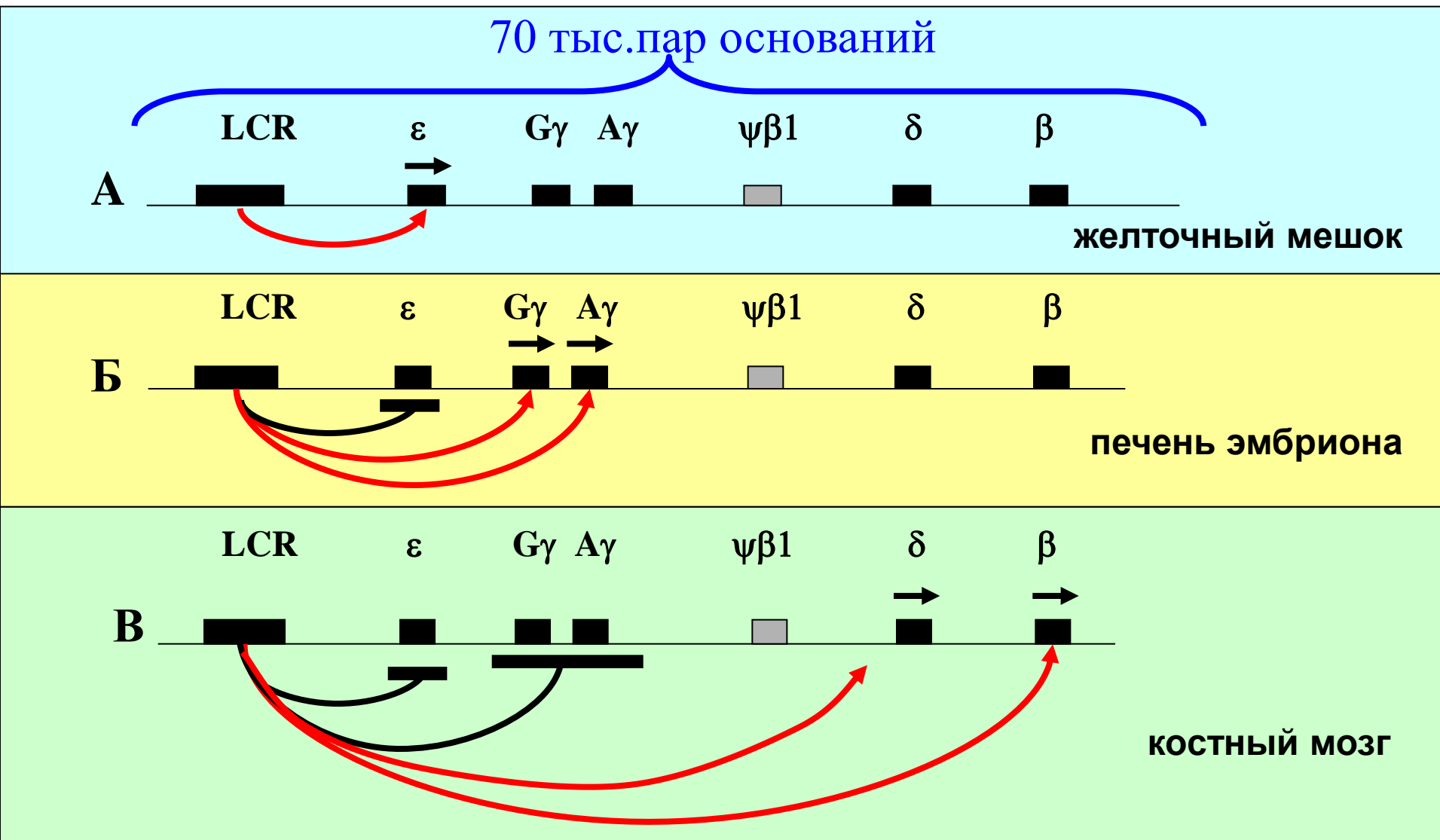
- фрагмент (или группа фрагментов) ДНК, присутствие которых обеспечивает высокий уровень тканеспецифической экспрессии связанного с ним гена, входящего в трансгенную конструкцию,
- пропорционально количеству копий трансгена и
- независимо от места встраивания в геном .

Как правило, локус-контролирующие районы регулируют гены, находящиеся в кластерах и обеспечивают их ткане- и стадие-специфическую 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

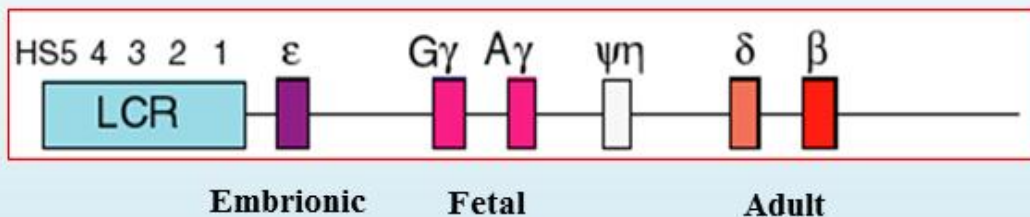


ϵ , $G\gamma$, $A\gamma$, δ , β - гены гемоглобинов; $\psi\beta 1$ - псевдоген

КЛАСТЕР β -ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА И ВРЕМЕННАЯ ДИНАМИКА ЭКСПРЕССИИ В ОНТОГЕНЕЗЕ

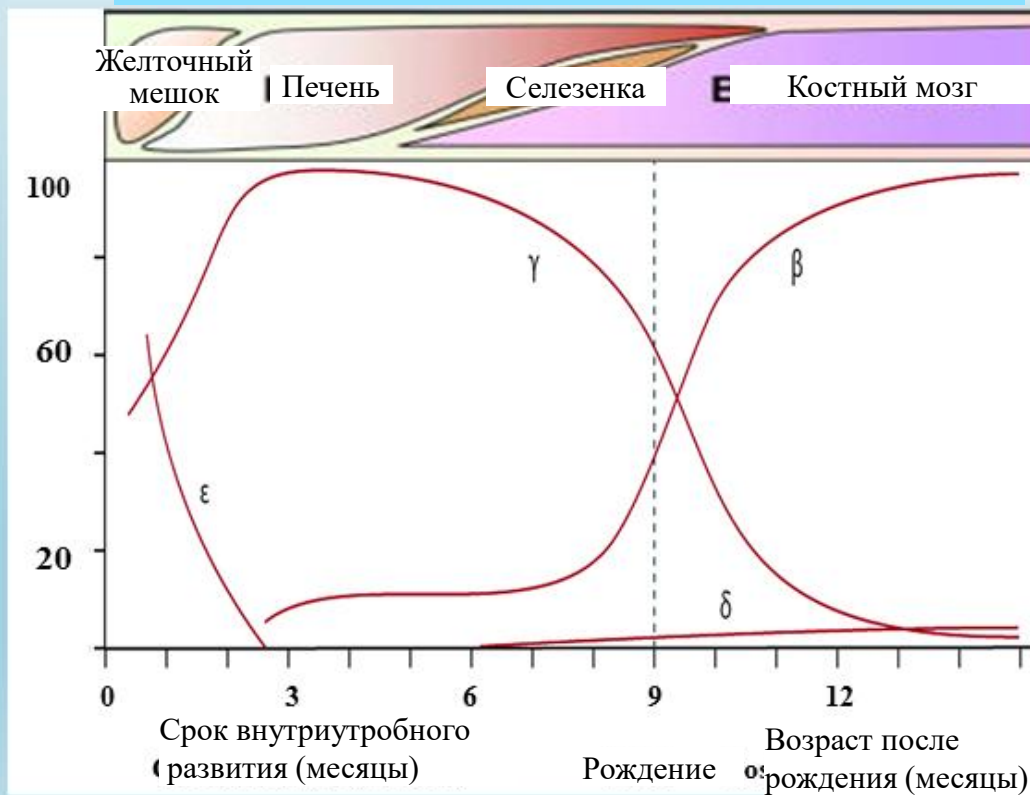
B-Globin locus

Chr11



Орган, где осуществляется эритропоэз

Уровень экспрессии глобинов (%)



Эмбриональный (ϵ) и фетальные ($G\gamma$, $A\gamma$) глобины обеспечивают большее сродство гемоглобина к кислороду, что обеспечивает эффективное снабжение эмбриона и плода кислородом путем экспорта кислорода из крови матери.

Примеры структурно-функциональной организации LCR

TCR α/δ (мышь)

L0009



HS1 - энхансер

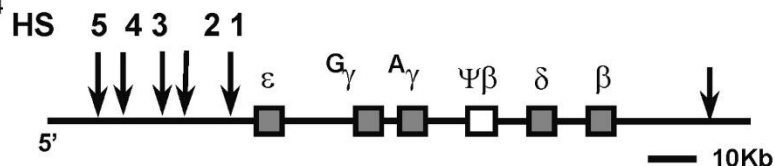
HS1, HS1' район определяющий распространенность в тканях

HS2 - HS6 район открывающий хроматин

Цифрами обозначены экзоны гена *Dad1*

Кластер β глобиновых генов (человек)

L0014

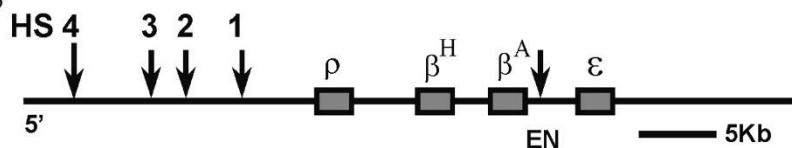


HS5 - инсультатор

HS4 - HS1 - энхансер

Кластер β глобиновых генов (цыпленок)

L0005



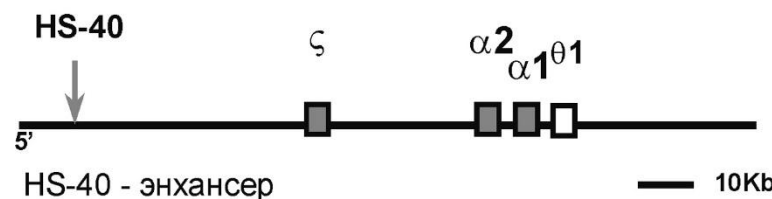
HS4 - инсультатор

HS3, HS2 - энхансер

EN - межгенный энхансер

Кластер α-глобиновых генов (человек)

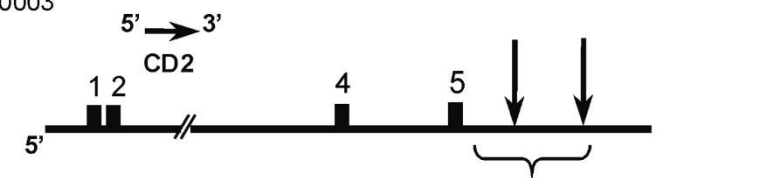
L0006



HS-40 - энхансер

Ген *CD2* (человек)

L0003



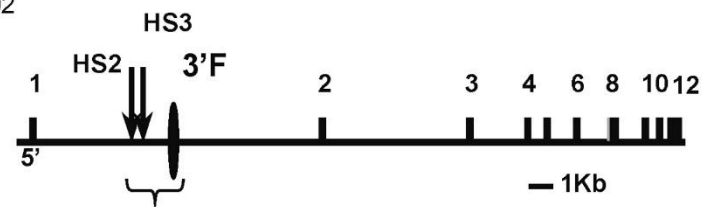
HS1 - энхансер

HS3 - инсультатор

Цифрами обозначены экзоны гена *CD2*

Ген *ADA* (человек)

L0002



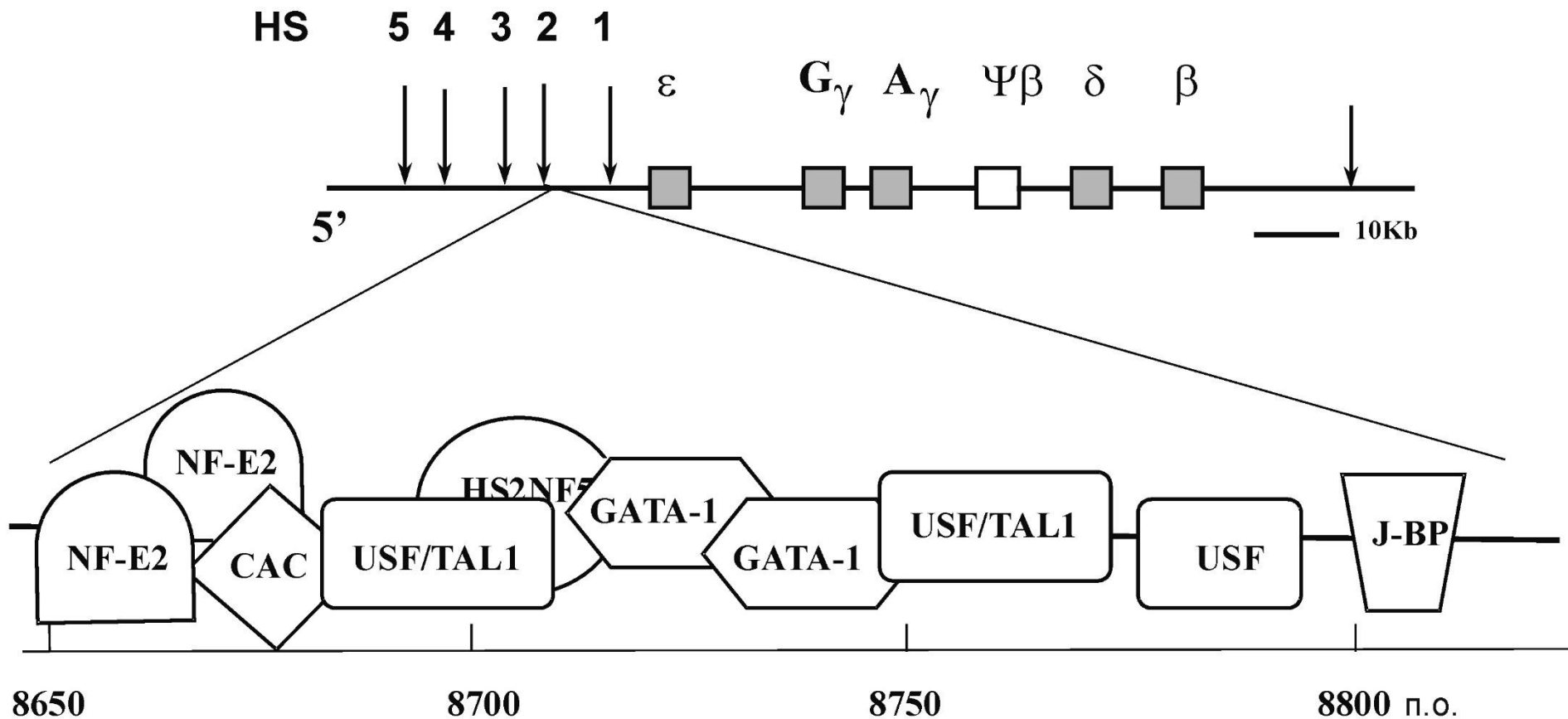
HS2 - 5'вспомогательный элемент

HS3 - энхансер;

3'F - 3'вспомогательный элемент

Цифрами обозначены экзоны гена *ADA*

LCR глобинового локуса человека



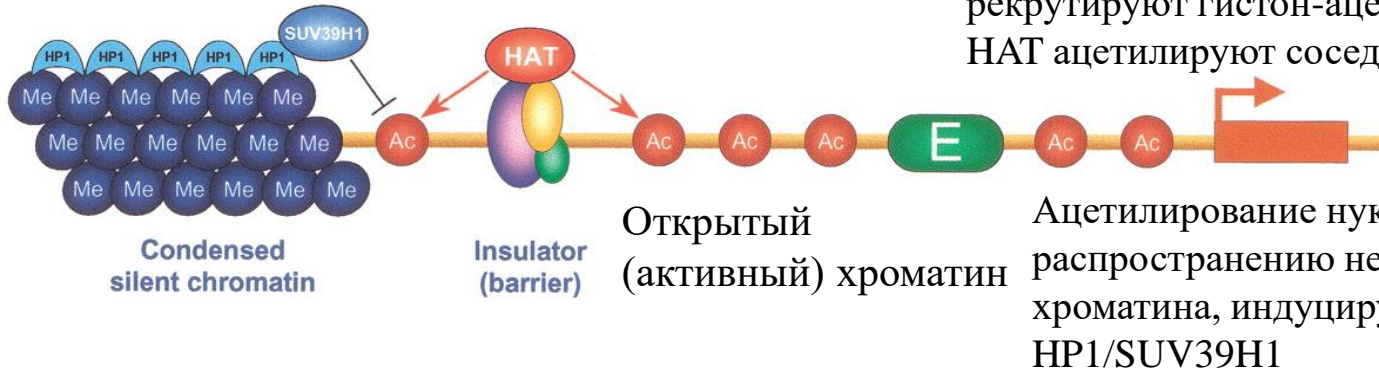
Каждый LCR включает специфическую группу сайтов связывания и располагается иногда на очень большом (до десятков тысяч п.о.) расстоянии от контролируемой кассеты генов

Инсулятор - участок ДНК, выполняющий специализированную регуляторную функцию

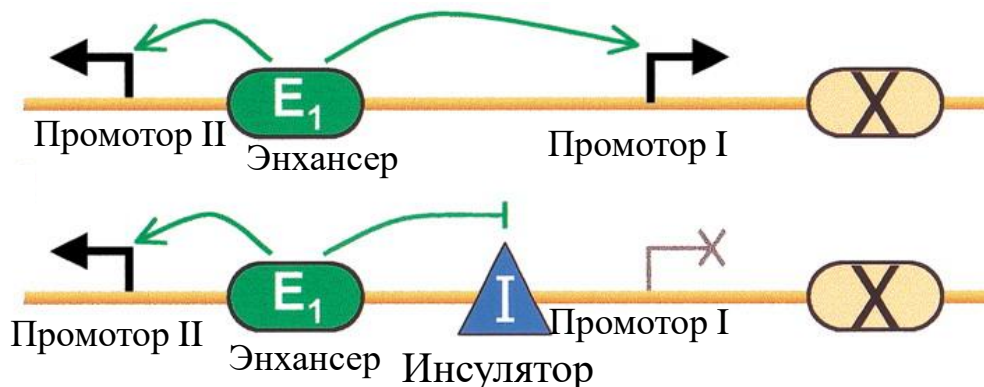
Функции инсулятора:

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

Белки, взаимодействующие с инсулятором рекрутируют гистон-ацетилаз (НАТ). НАТ ацетилируют соседние нуклеосомы.

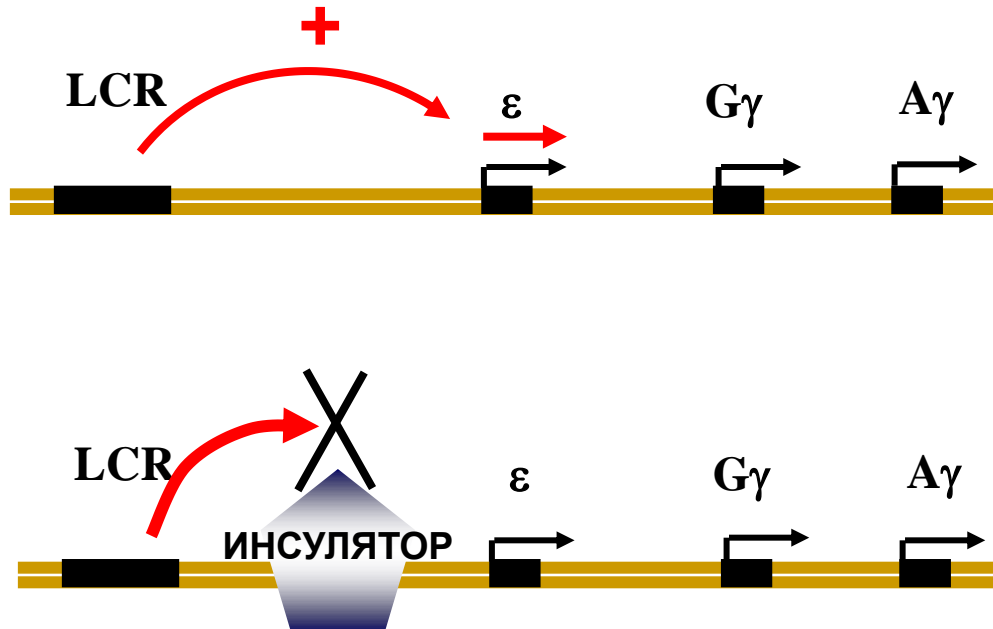


Инсулятор может располагаться между энхансером и промотором и блокировать активирующее влияние энхансера на транскрипцию гена



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

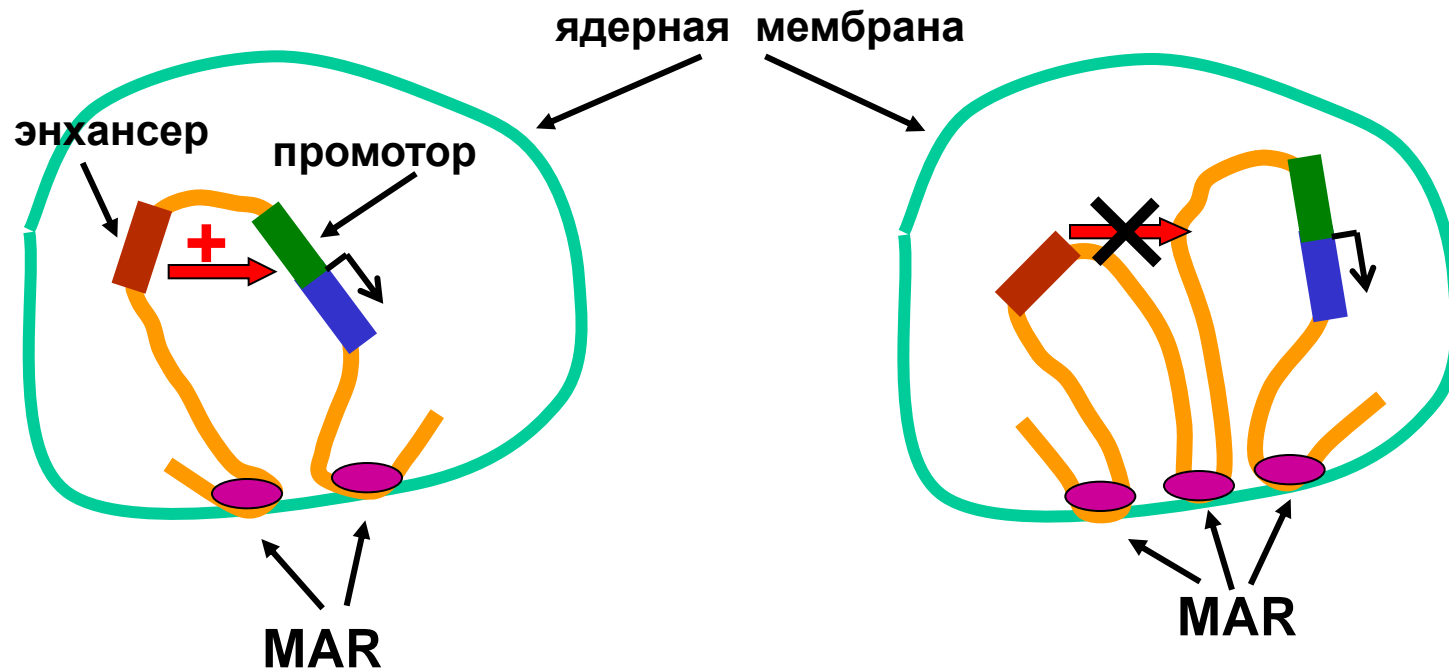
ИНСУЛЯТОР



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

Роль инсультатора может выполнять участок прикрепления к ядерному матриксу (MAR). При включении такого инсультатора 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

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

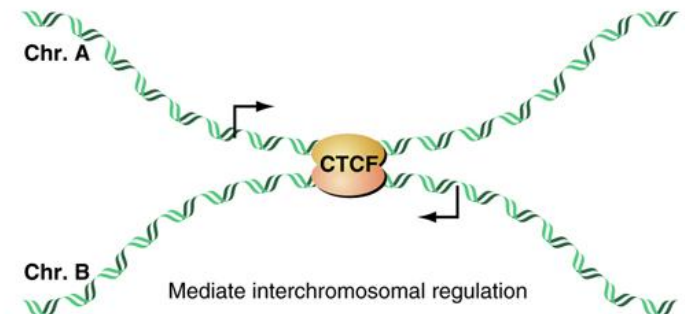
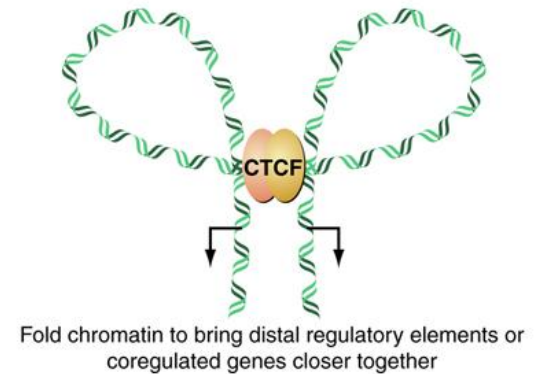
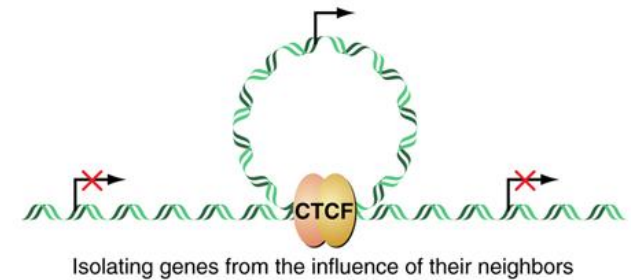
При включении инсулятора два регуляторных района оказываются в различных доменах, и не способны взаимодействовать

Роль транскрипционных факторов CTCF при формировании петель ДНК

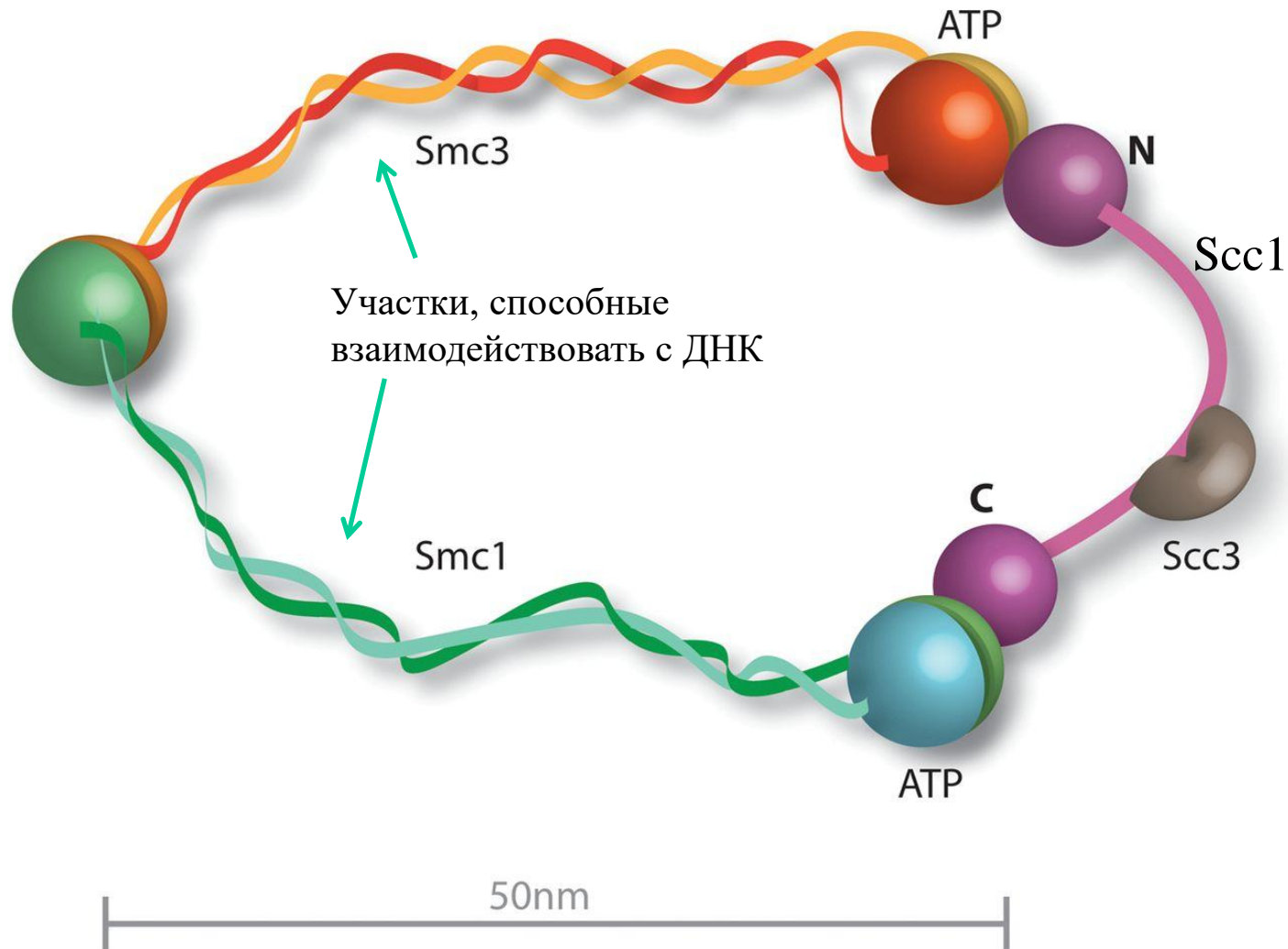
Функционирование инсуляторов тесно связано с наличием сайтов связывания транскрипционного фактора **CTCF** (CCCTC-binding factor).

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.

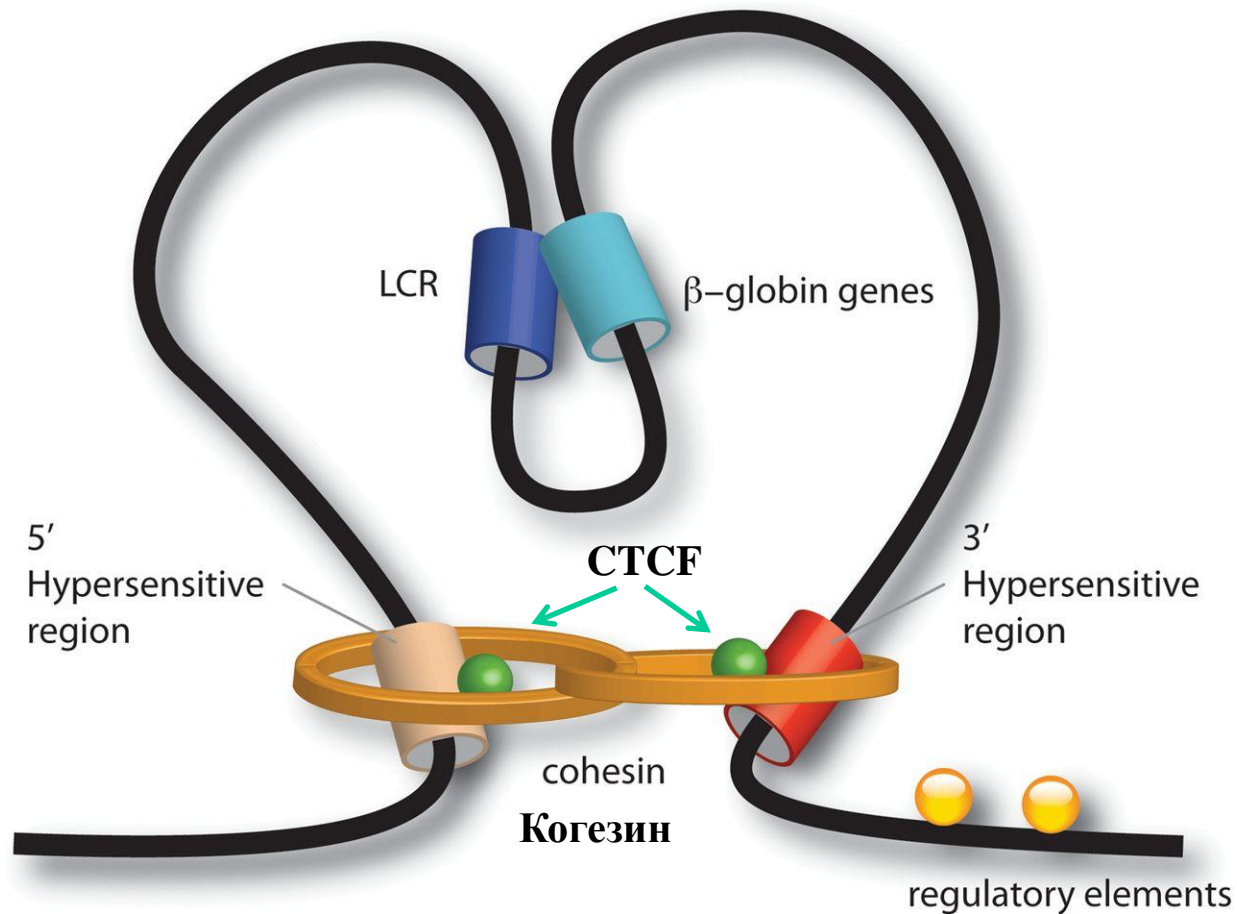
CTCF является маркером участков ДНК, разделяющих активный и репрессированный хроматин



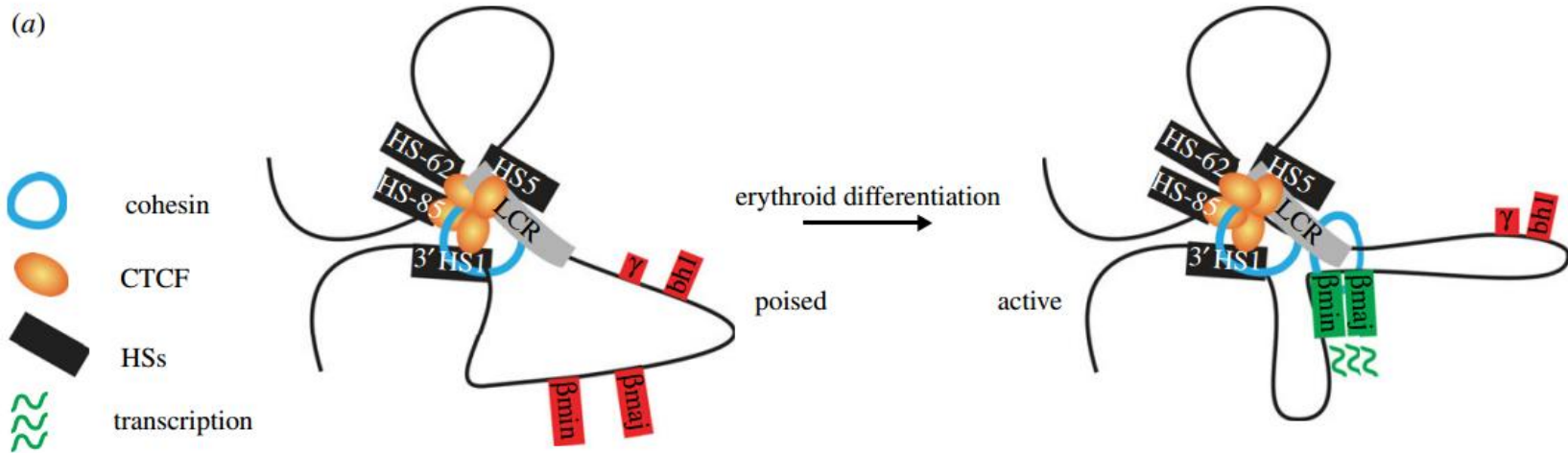
Когезин – комплекс белков , формирующий кольцеобразную структуру



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.

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

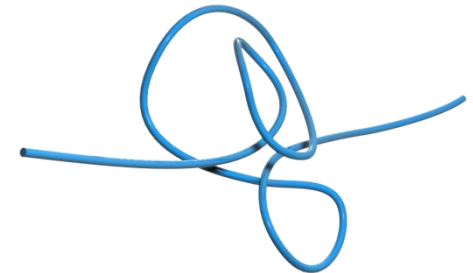
Исследованы контакты в клеточных культурах человека MCF7 и K562



Промоторная модель

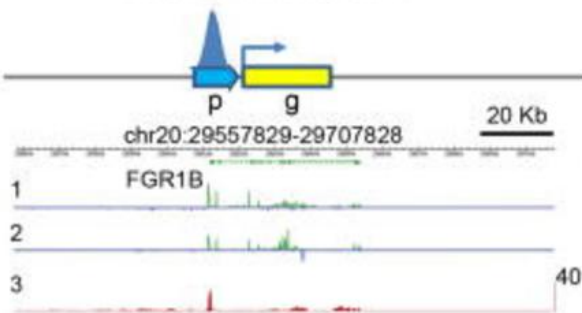


Одногенная модель
(Промотор-Энхансер)

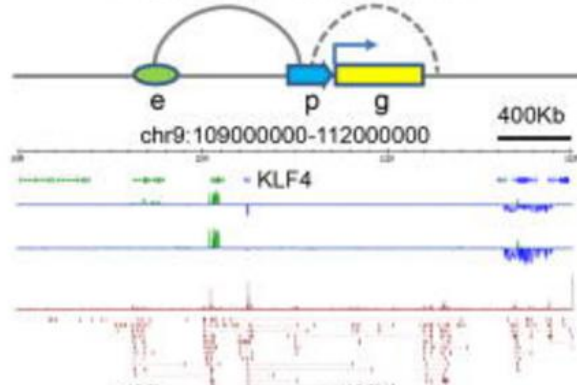


Мультигенная модель (хромосомный оперон - “chromoperon”)

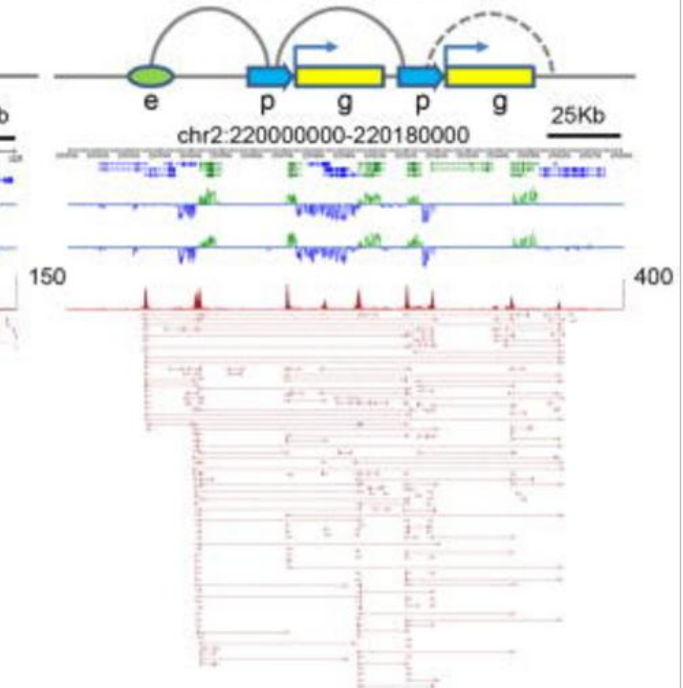
Basal promoter model



Single-gene interaction model



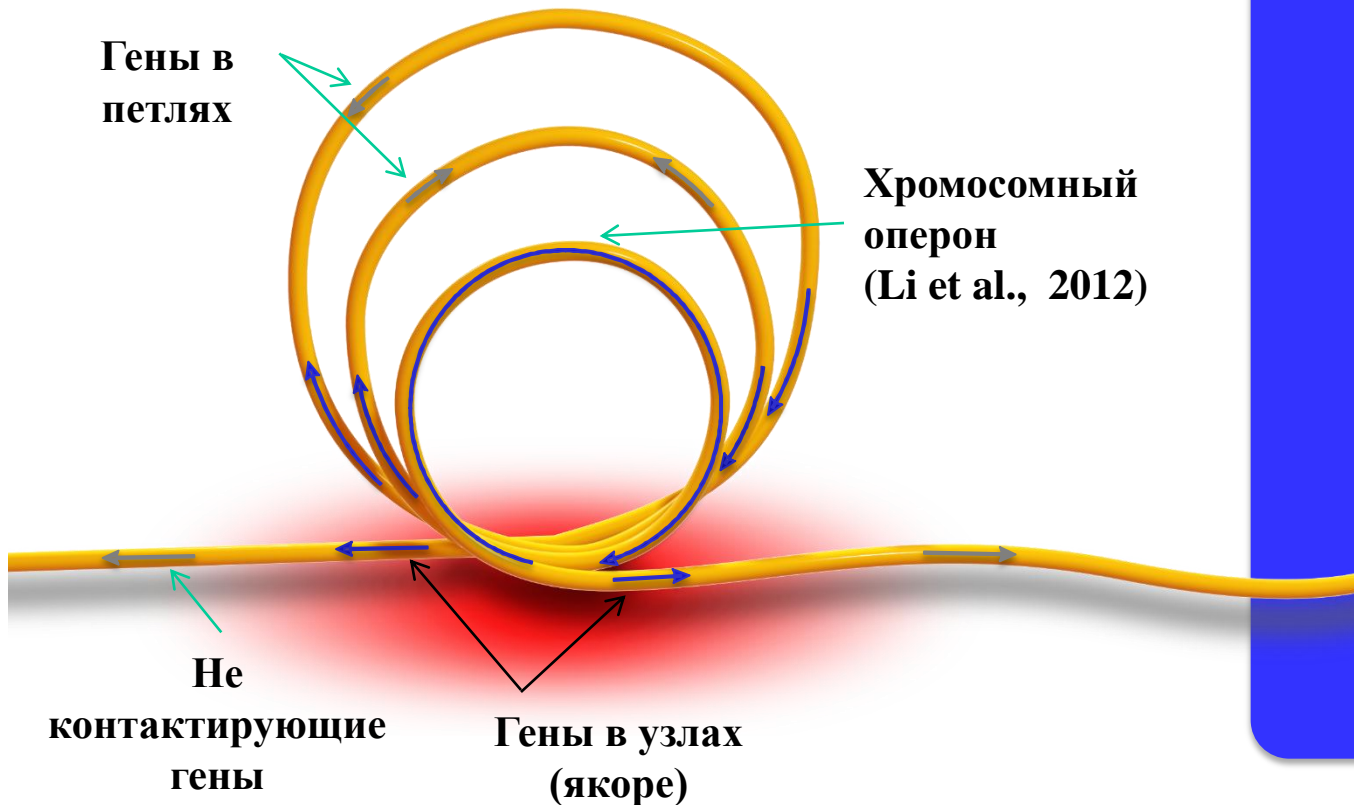
Multi-gene interaction model



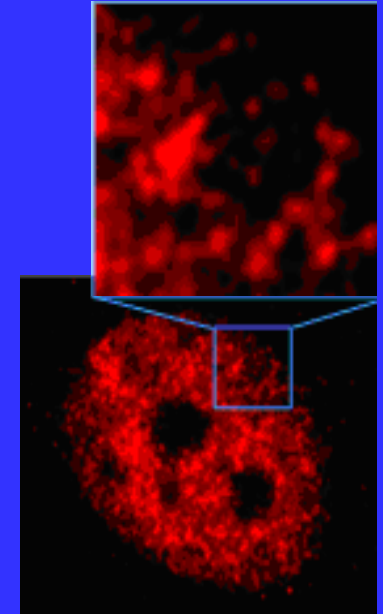
Li G. Et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

Хромосомные опероны и транскрипционные фабрики

Структуры, найденные с помощью ChIA-PET



Ядро клетки – участки транскрипции под микроскопом



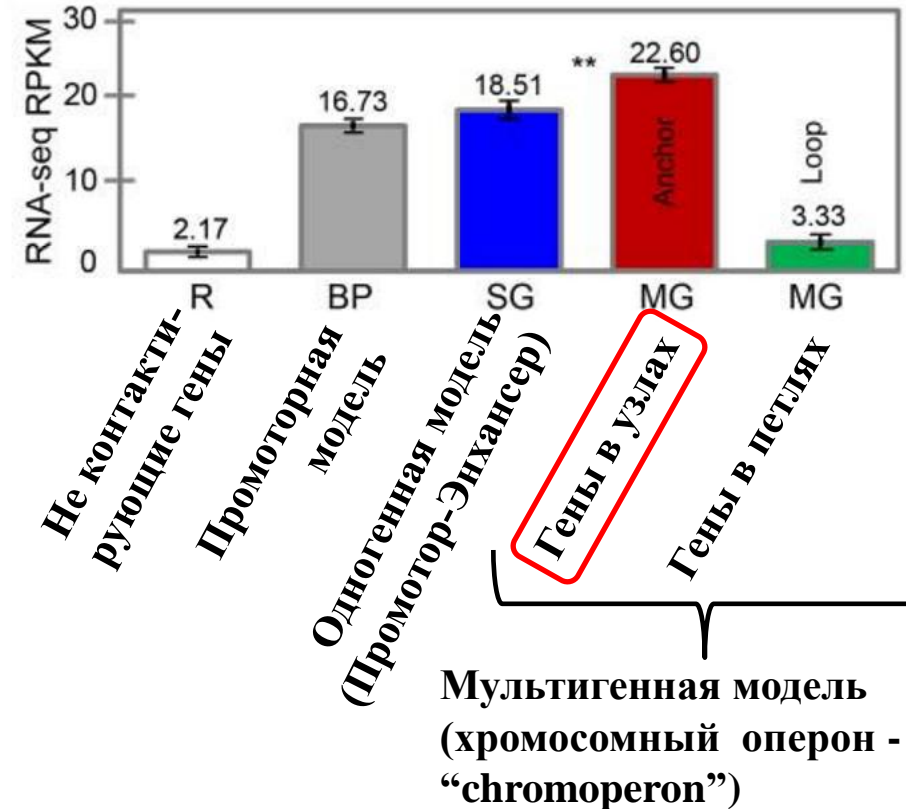
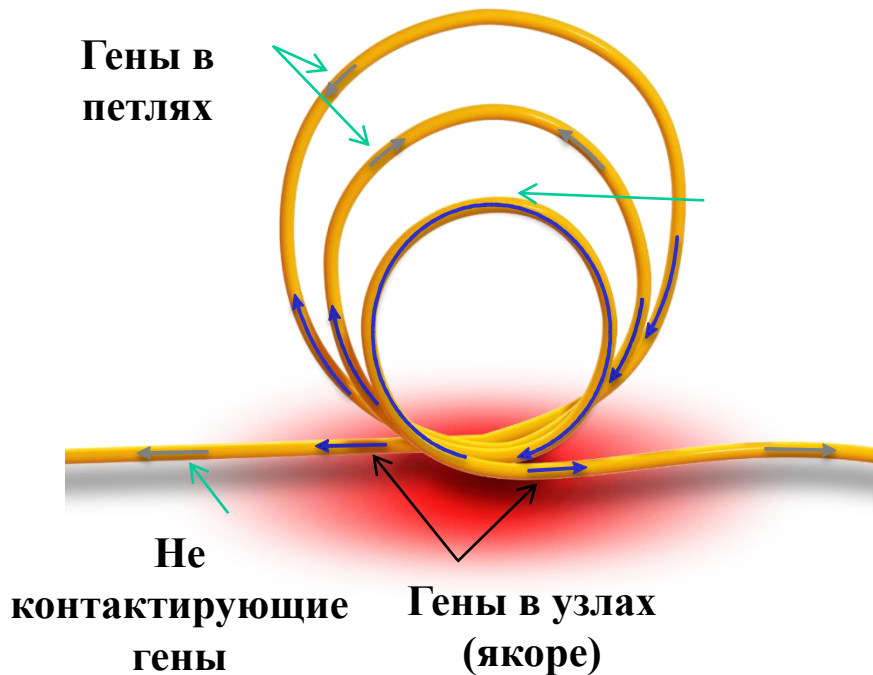
FISH

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

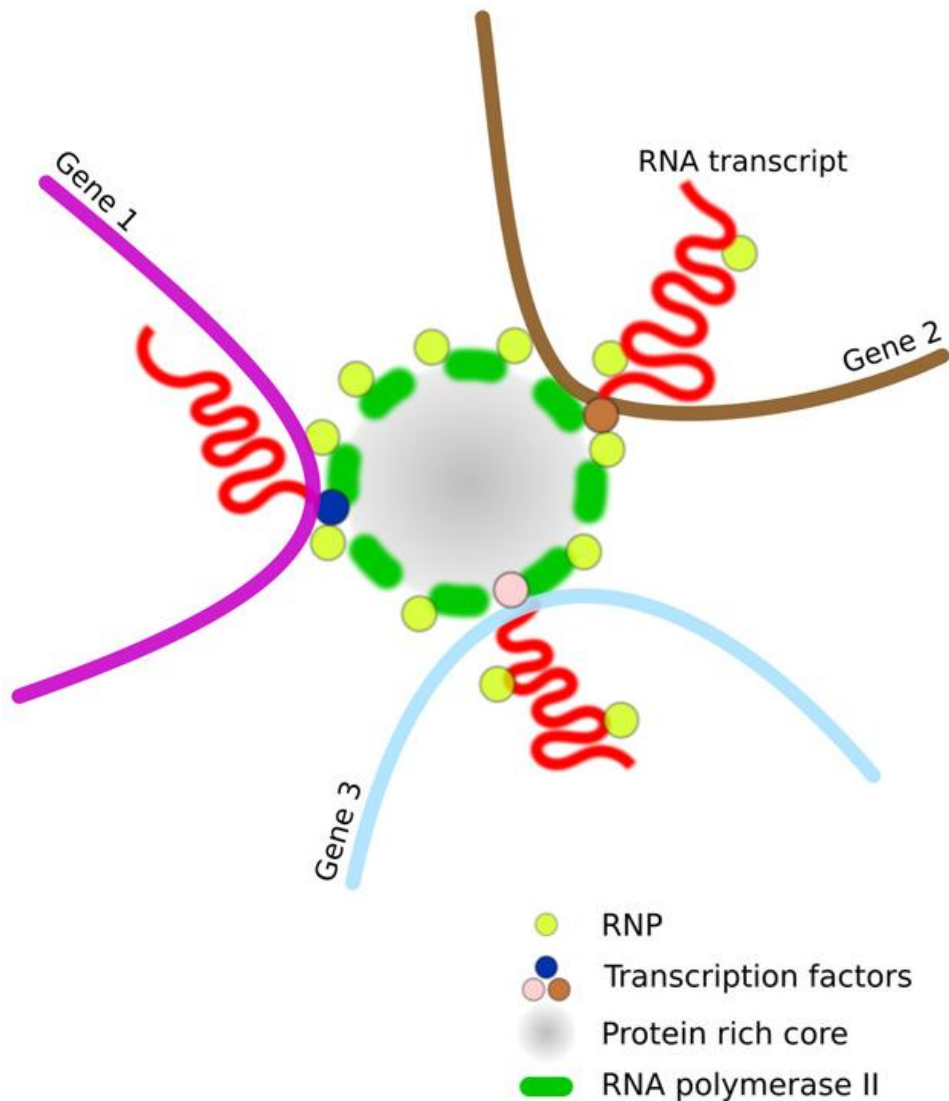
Структуры, найденные с помощью ChIA-PET, в MCF7 клетках человека

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)

KCl

Нейрональные
клетки мыши

Деполяризация
мембраны

Активация коактиваторного белка P300/CBP

Данные
экспериментов
ChIPseq:

Количество пиков связывания P300/CBP в геноме мыши увеличивается с 1000 до 28000

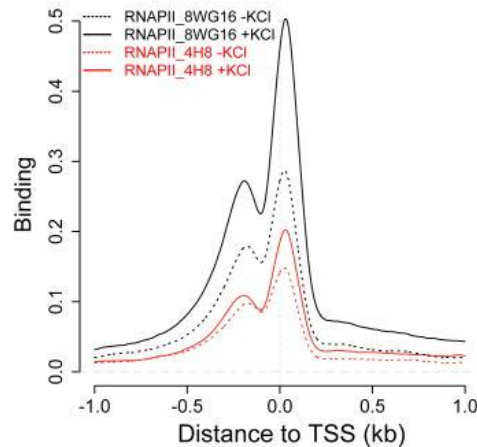
Поиск нейтрально – специфичных энхансеров согласно критериям:

- 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)

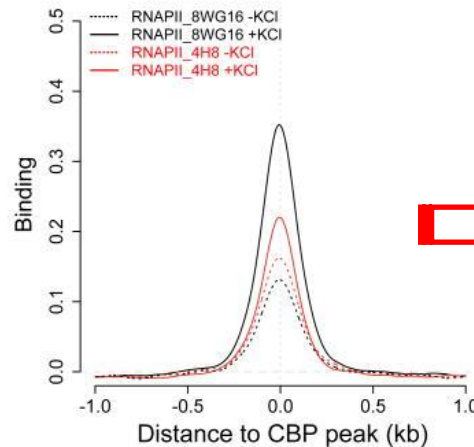
Связывание РНК полимеразы II (RNAPII) в районах промоторов и энхансеров

Promoters



Ассиметричный пик – транскрипция идет в одном направлении

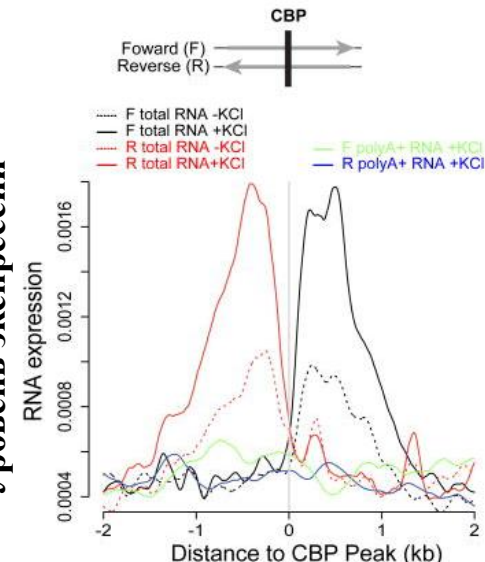
Extragenic enhancers



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

Экспрессия РНК в энхансерных участках

Extragenic enhancers



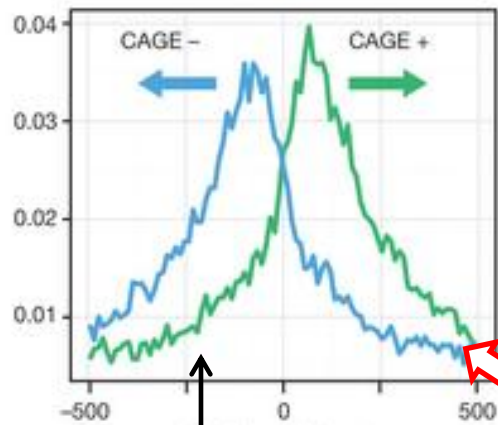
Уровень экспрессии

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)

Уровень экспрессии
= доля энхансеров,
с которых считывается РНК



Позиция относительно центра
(пик связывания белка P300)

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).

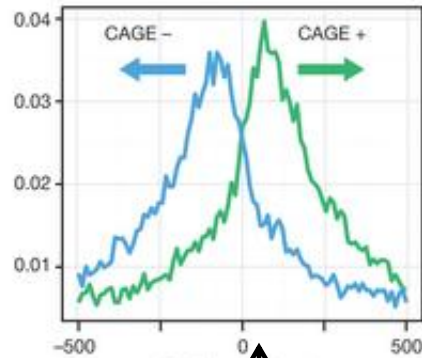
Были выявлены транскрипты (короткие РНК), которые считывались как с прямой (+), так и обратной (-) цепи ДНК в районе энхансера. Расстояние между (+) и (-) пиками соответствовало длине участка ДНК, упакованного в нуклеосому (180 п.о.).

В клетках HeLa человека выявлена энхансерная РНК (эРНК)

5'-фланкирующие участки транскрибируемых энхансеров содержат TATA боксы и INR элементы

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

Уровень экспрессии
= доля энхансеров,
с которых считывается РНК

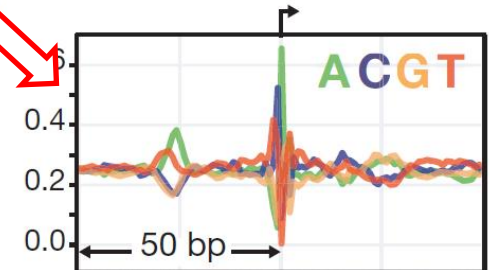


Позиция относительно центра
(пик связывания белка P300)

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

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

Частота встречаемости нуклеотида



Enhancer
CAGE 5' ends

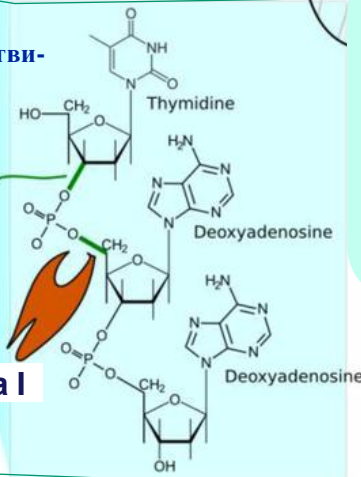
Расщепление ферментом DNase I

High Low
DNase I cleavage
(per nucleotide)

Сайт гиперчувствительности к DNase I

Фосфодиэфирная связь

DNase I



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

Позиции TATA 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:

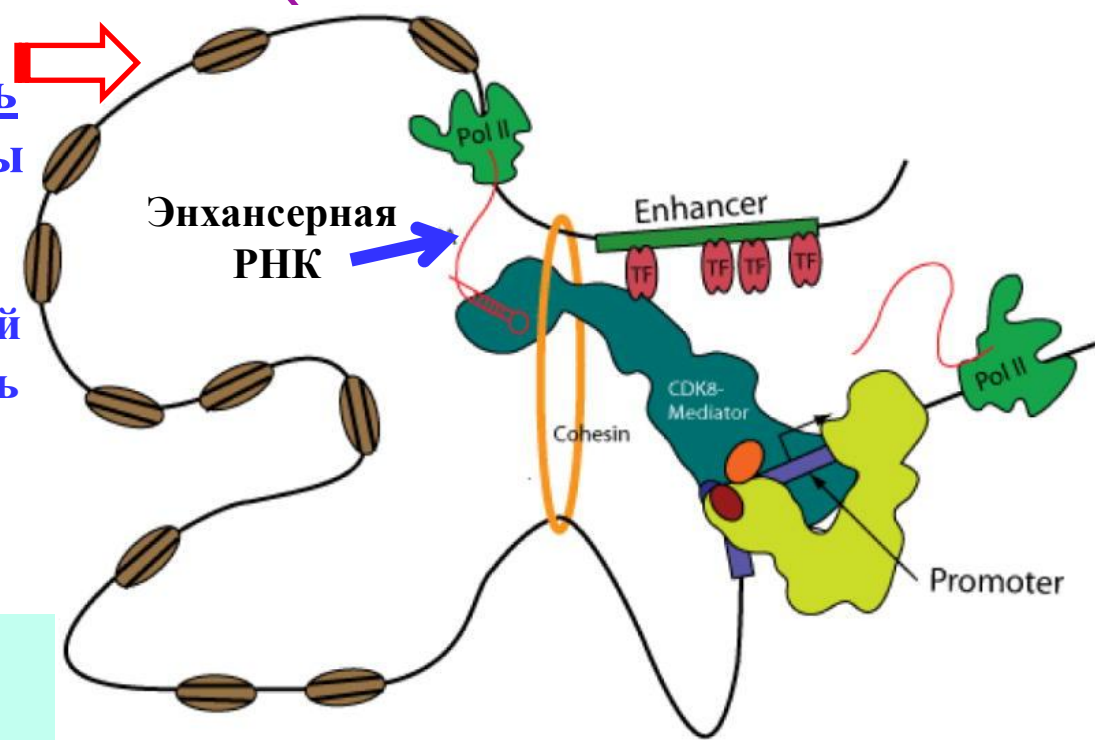
- Облегчает образование петель ДНК, сближающих энхансеры и промоторные районы.
- Способствует освобождению РНК-полимеразы из комплекса , обеспечивающего остановку (паузу) на стадии ранней элонгации
-

Роль энхансерной РНК:

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

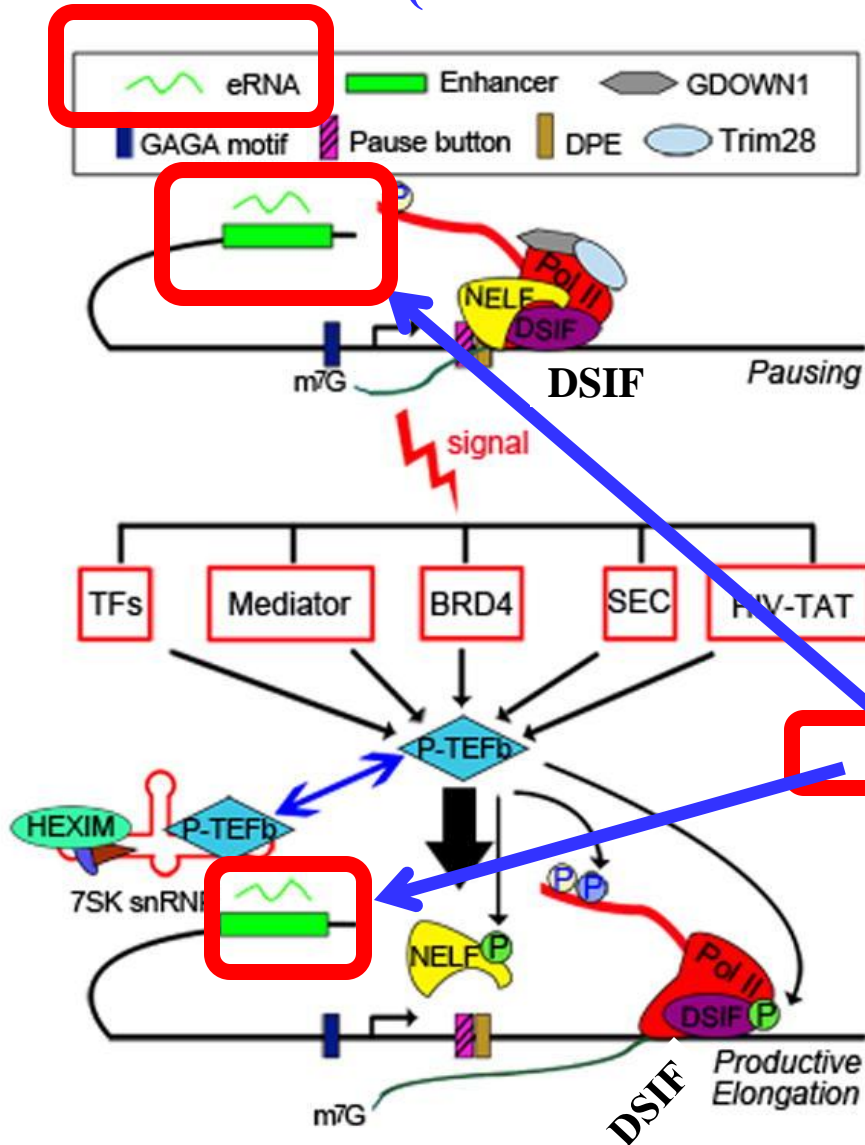
Показано, что в реализации этой функции участвует CDK8 модуль медиаторного комплекса, объединяющий субъединицы MED12, MED13, CCNC, CDK8.

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

Окончание паузы происходит при:

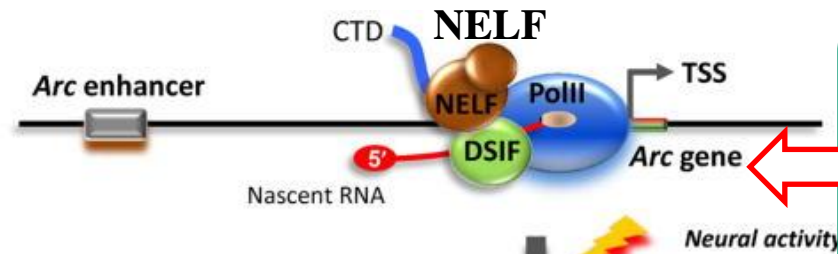
- Тепловом шоке (ТФ)
- Гипоксии (ТФ)
- Воспалении (Nf-kB, BRD4)
- Дифференцировке стволовых клеток (SEC, ТФ)
- Инфекции HIV (HIV-TAT)

Механизм участия эРНК см. на следующем слайде

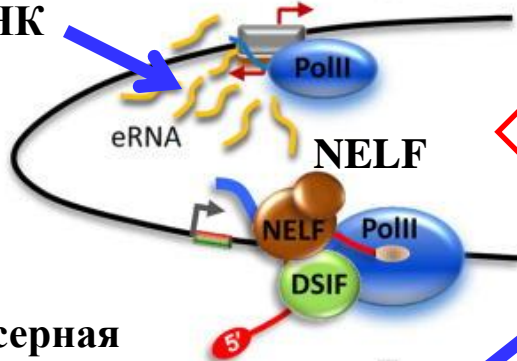
Роль энхансерной РНК:

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

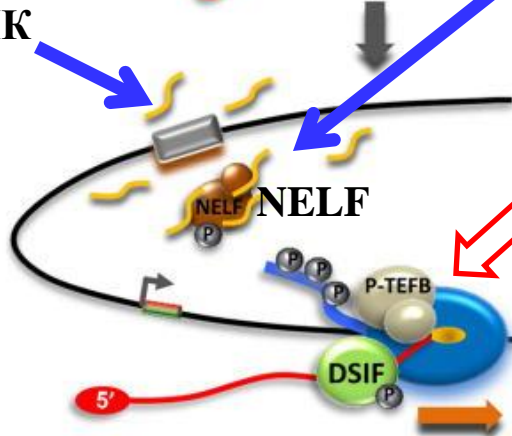
Ген Arc = Activity-regulated cytoskeletal protein



Энхансерная РНК



Энхансерная РНК



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 E-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.

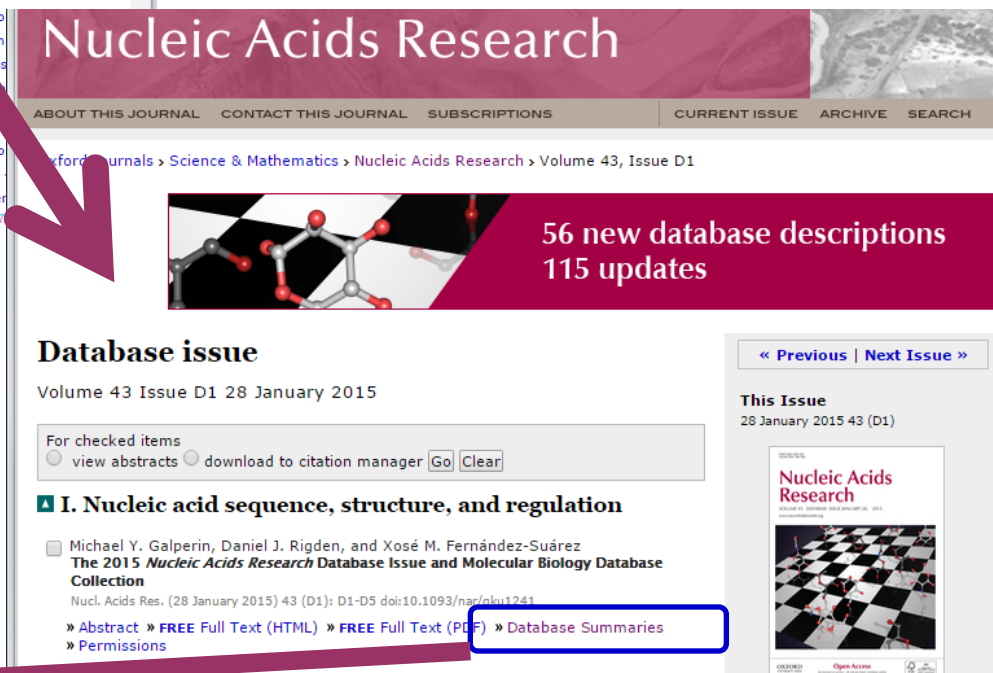
БАЗЫ ДАННЫХ ПО РЕГУЛЯЦИИ ТРАНСКРИПЦИИ

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 NAR journal homepage. At the top, there's a navigation bar with 'OXFORD JOURNALS', 'CONTACT US', and 'MY BASKET'. Below this is a large banner for 'Nucleic Acids Research' with a sub-header 'Now available: the 2015 Web Server Issue'. A sidebar on the left contains links like 'ABOUT THIS JOURNAL', 'CONTACT THIS JOURNAL', 'SUBSCRIPTIONS', 'CURRENT ISSUE', 'ARCHIVE', and 'SEARCH'. The main content area features a 'READ THIS JOURNAL' section with links to 'View Current Issue (Volume 43 Issue 17 30 September 2015)', 'Advance Access', and 'Browse the Archive'. A blue box highlights the '2015 Database Issue' link in the 'THE JOURNAL' section. A large purple arrow points from this link towards the 'Database issue' page shown in the next block.



The screenshot shows the 'Database issue' page for Volume 43, Issue D1, dated 28 January 2015. The page features a large banner with the text '56 new database descriptions' and '115 updates'. Below this, there's a section titled 'Database issue' with a search bar and a list of articles. A blue box highlights the 'Database Summaries' link in the article list. A large purple arrow points from the 'Database Summaries' link towards the text at the bottom left of the image.

Ссылка на «online Database Collection»

Журнал NAR: 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



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 by 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 About the Index

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CUI

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](#)

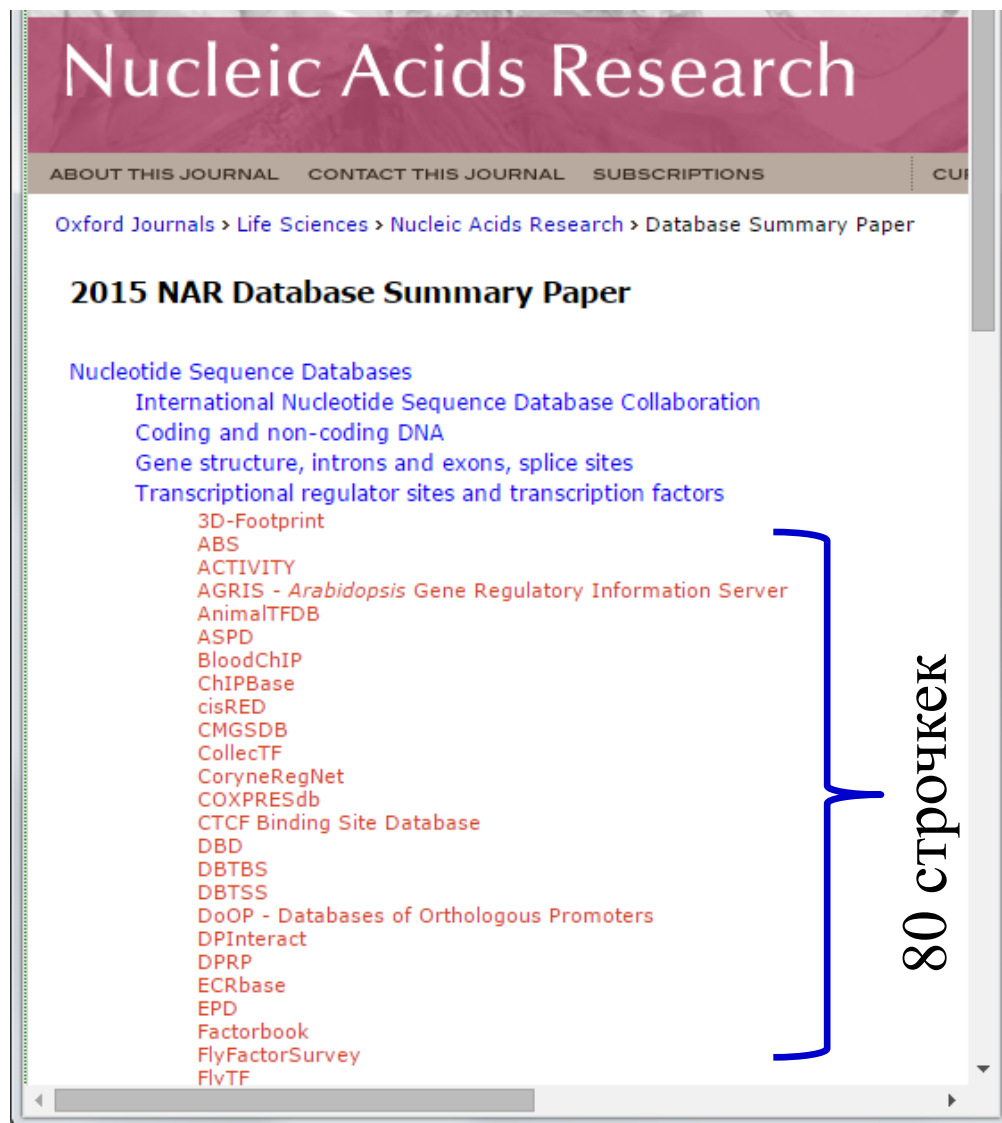
- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

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

Online ISSN 1362-4962 - Print ISSN 0305-1048

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

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



- 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
- DPIInteract
- 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®
- Tranterm
- TRED - Transcriptional Regulatory Element Database
- TRRD
- TrSDB
- TTSMI
- UniPROBE
- VISTA Enhancer Browser
- WebGeSTer DB
- YEASTRACT
- 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)

Computational Cancer Genomics | ExPASy | EPFL

Access EPDnew

- 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
FTP site

Access EPD

- Promoter elements
- Select / Download
- FTP site

Access MGA data

- MGA Data Overview
- MGA FTP site

Documents

Other Resources

References

News: 10-10-2016 -- New liftOver option added to "Select / Download" page [more](#)

This resource allows the access to several databases of experimentally validated promoters: EPD and EPDnew databases. They differ by the validation technique used and the coverage. EPD is a collection of eukaryotic promoters derived from published articles. Instead, the EPDnew databases (HT-EPD) are the result of merging EPD promoters with in-house analysis of promoter-specific high-throughput data for selected organisms only. This process gives EPDnew [high precision and high coverage](#).

EPDnew is a collection of databases of experimentally validated promoters for selected model organisms. Evidence comes from TSS-mapping from high-throughput experiments such as CAGE and Oligocapping. The resulting databases are the following:

- Animals:
 - Homo sapiens*: 25503 promoters,
 - Mus musculus*: 21239 promoters,
 - Drosophila melanogaster*: 15073 promoters,
 - Apis mellifera*: 6493 promoters,
 - Danio rerio*: 10728 promoters,
 - Caenorhabditis elegans*: 7120 promoters;
- Plants:
 - Arabidopsis thaliana*: 10229 promoters;
 - Zea mays*: 17081 promoters;
- Fungi:
 - Saccharomyces cerevisiae*: 5117 promoters,
 - Schizosaccharomyces pombe*: 3440 promoters.

EPD

The **Eukaryotic Promoter Database** is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries. The annotation part of an entry includes description of the initiation site mapping data, cross-references to other databases, and bibliographic references. EPD is structured in a way that facilitates dynamic extraction of biologically meaningful promoter subsets for comparative sequence analysis. This database contains **4806 promoters from several species**.

EPDnew

EPD

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

EPD – первая часть объединенного ресурса

The screenshot shows the EPD website with the following elements:

- Header:** SIB logo, a rabbit icon, and the text "EPD EUKARYOTIC PROMOTER DATABASE".
- Navigation Bar:** "Computational Cancer Genomics | ExPASy | EPFL".
- Search Bar:** A text input field followed by "in" and a dropdown menu set to "All databases", and a green "SEARCH" button.
- Left Sidebar:**
 - Access EPDnew** (red button):
 - H. sapiens*
 - M. musculus*
 - D. melanogaster*
 - D. rerio*
 - C. elegans*
 - A. thaliana*
 - Standard search
 - Select / Download
 - Promoter analysis tools
 - FTP site
 - Access EPD** (red button):
 - Promoter elements
 - SRS access to EPD
 - Select / Download
 - FTP site
 - Access MGA data** (red button):
 - MGA Data Overview
 - MGA FTP site
- Main Content Area:**
 - EPD database**

The Eukaryotic Promoter Database is an annotated non-redundant collection of eukaryotic POL II promoters, for which the transcription start site has been determined experimentally. Access to promoter sequences is provided by pointers to positions in nucleotide sequence entries. The annotation part of an entry includes description of the initiation site mapping data, cross-references to other databases, and bibliographic references. EPD is structured in a way that facilitates dynamic extraction of biologically meaningful promoter subsets for comparative sequence analysis. This database contains **4806 promoters from several species**.

Current version is based on **EMBL 124**.
 - Collection accessibility**

EPD database is accessible in different ways: (1) using the input form in the header, searching for single gene symbol, gene description or ENSEMBL / RefSeq gene IDs; (2) using the [download page](#) for selecting specie-specific promoters and downloading them in various formats or (4) through an [ftp website](#) for bulk download of the whole database in various formats (SGA, BED, ...).
 - Reference**

A detailed description of the principles governing EPD can be found here: [EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era](#), Dreos, R., Ambrosini, G., Périer, R., Bucher, P. Nucleic Acids Res. (2013) 41(Database issue):D157-64; PUBMED [23193273](#)

EPD (Eukaryotic Promoter Database) содержит информацию о 4806 промоторах генов эукариот, транскрибируемых 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)


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
 ▼

+ 13046 промоторов риса (с пометкой «предварительные»)

Опция для загрузки данных в фаста-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	[3474400, -70361425]	HapMap
CLEANEX	HS_AK2		
DNA References	AL020995.14	[46401, -104596]	ENA GenBank
SWISSPROT	P54819	KAD2_HUMAN	DBJ
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	tgtgagcggcgagtgggacgtgcgtggcgtgcgttgacctgggaaGCACTGGACCT		
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

Пересылки 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 фаста формате

```
>EP33026 (-) Bt TP2; range -200 to 50.  
AGCTCCACCCGACCTGAGGGGCTGCTCTCAACCCACAGACACGCCCTTTGAAAGCTGCCC  
CACCTGCGTGTTAGGATGAGGGCAGAGGGCTTTGTTCCCGTTGGGCCACATCTGTTACA  
TACCCCTGTGGCCAGTGCCATCACAATCGGGCCAACTATATAACCAGGGGGTGCCAG  
GGCCTCTGTGAAGCTGGGTCTGCCAGAAGAGGAGGAGGAGGCGGCGGCCCTGCCCTCT  
AAGCGAGGCCG  
>EP28006 (+) Bt protamine P1^1.1; range -200 to 50.  
GGCCCCACCCCCACACACATCACAGCCCCACCCCTGCACATCACAGCCCCGCCCTCCCTC  
ACCAAGCACCTCCCACATGCCCATATATGGGCATGATTTGGGCAGCTCTGACCCTGGTCT  
GTGAGGTCTGGGTCTCTGTGACCTCACAATGACCAGGACCCTGCCCGGGTCTATATAAGA  
GGCCGGGAAGTCGGCCCCTGTCACAGCCCAAAATTCACCTGCTCACAGGTTGGCTGGC  
TCAACCAAGGC  
>EP15026 (+) Bt cytokeratin Ia; range -200 to 50.  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNCATGGTGGACGGCAAGTTATTCAAAGGAGAGCGGACCGGGAGGTTGGCGGAA  
ACGCGGAGACCTTCTGAAGCTCTGGGCCAGAGGTGGCGCTTATATAGGGCTGGGAGCTTG  
GCTGGCTGGCG  
>EP15027 (+) Bt cytokeratin Ib; range -200 to 50.  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCATGG  
TGAGTGAAGTCGGTGAAGGGAAAGTGCAAGAGTTAAGCAGGGCCACTCAGAGCCAGAGGA  
AGAGAGAAGCAAACCTGAAGACCTGTGCAGGATAAATCCCTTTATATACATCTAGGAGGCT  
GCTGGGCTCAA  
>EP15028 (+) Bt cytokeratin IV*; range -200 to 50.  
CAAGGCTAGAAGCCAGAAGAATTTCTCCATGACTAAAGGAAACCAAGAAGCAATATTCA  
TACTTCATACCTTTCTAGAGGCAGGGGGTGATCTCACTATTTGTAAAGCCCAGCCCTTTC  
TAATCTGCAGGCTCACCTCCAGGACTGAGCCCGGCCCATTTTTTCCATATATAAGCTGC  
TGCCGGGCCGCCCTCTATAGATCTGTTCTTTAGCTCTGCTTTCACCTCTCACACCCTTC  
TCAACCTATTC
```

1-ая строка - общая информация о последовательности

2-ая строка - последовательность ДНК

EPDnew - вторая часть объединенного ресурса

Информационное содержание (10 видов организмов)

Homo sapiens
Mus musculus
D. melanogaster
C. elegans
Apis mellifera (honey bee)
D. rerio (Zebrafish)

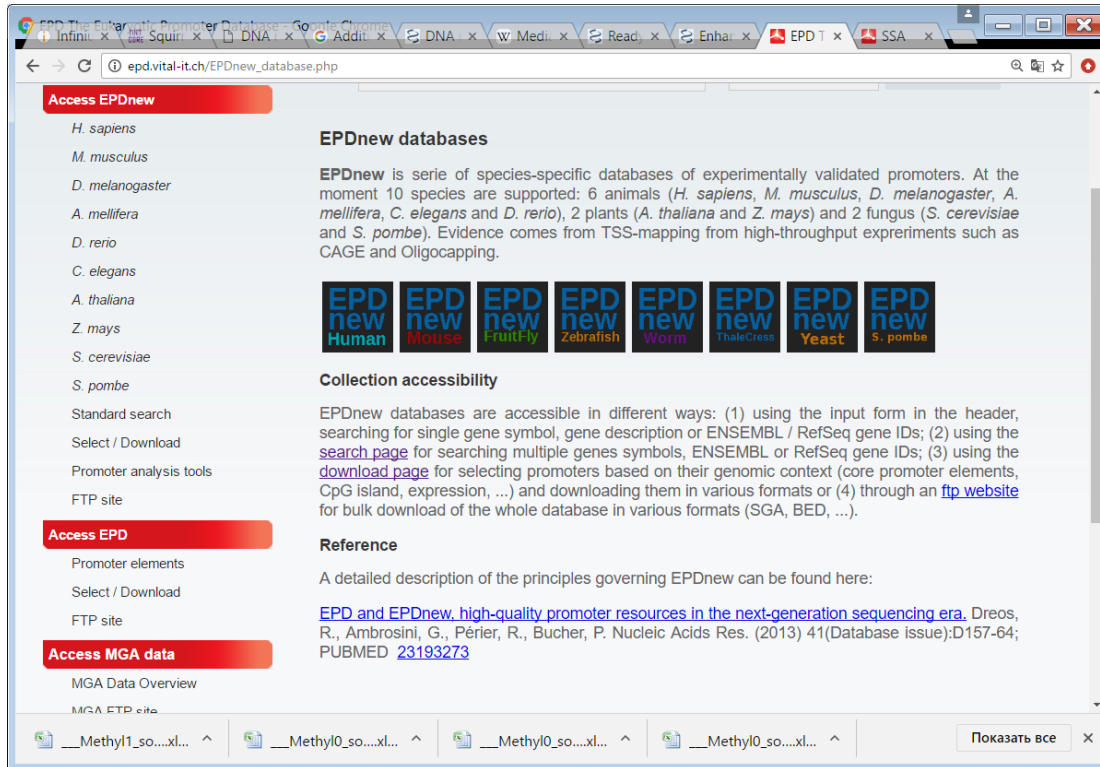
Животные

A. thaliana
Z. Mays

Растения

S. Cerevisiae
S. pombe

Дрожжи



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

Данные экстрагированы из Интернет-доступных ресурсов:

- Научных публикаций с саплиментами
- 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.

 FOLLOW US @NAR_OPEN

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

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

 Author Affiliations

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

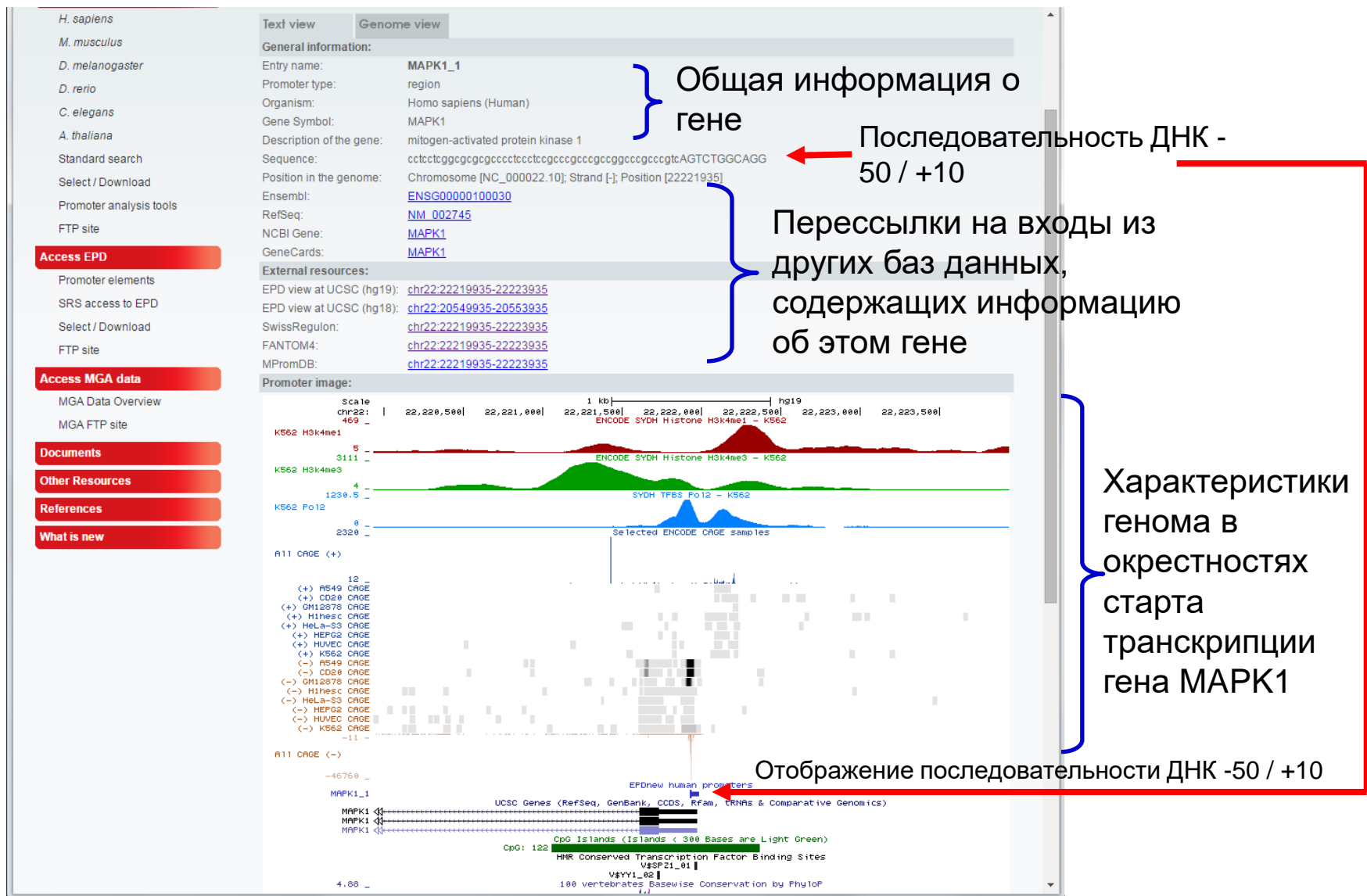
Received September 17, 2014

[« Previous](#) | [Next Article »](#)
[Table of Contents](#)

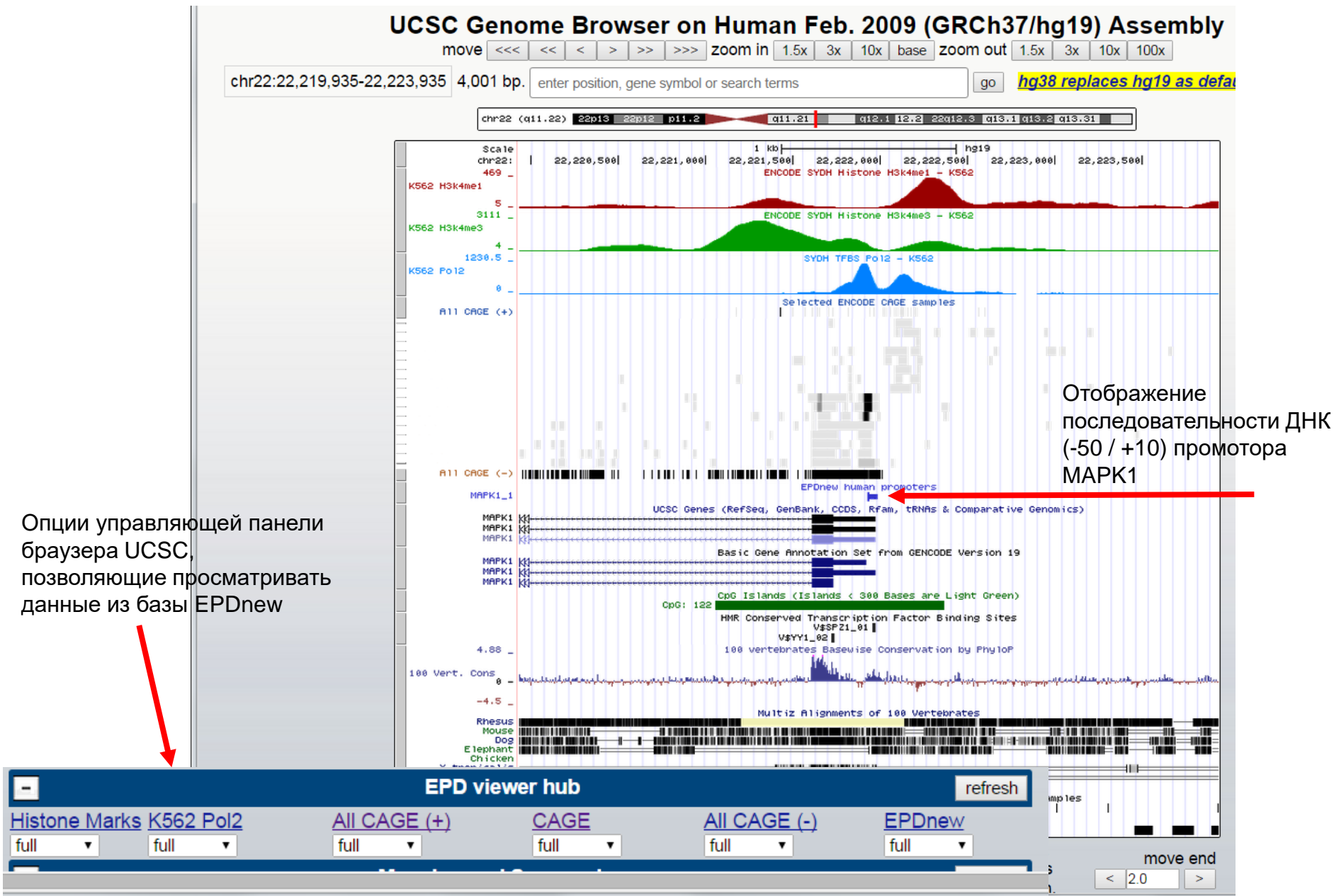
This Article

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:





Исходный вариант представления данных о промоторе гена MAPK1 в геномном браузере UCSC (Калифорнийский университет, г.СантаКруз)



Отображение последовательности ДНК (-50 / +10) промотора MAPK1

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

in All databases SEARCH

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 signal 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 or genome annotation features. 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 set of CCAAT-box positions that can be further analysed in the same way as a set of TSS positions.

How-To Documentation: [OProf](#), [FindM](#) and [ChIP-Cor](#)

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) restrict them based on all or some of their genomic contexes (such : presence of core promoter elements) or expression levels. After selection, you can **download** them in various format (for example in FASTA, BED, etc..) **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 5 tags

☐ AND expressed in at least 3 samples

Select

Database:

Database: human_epdnew

Assembly: hg19

Selection Parameters

TATA-box: with
Initiator: with
CCAAT-box: with
GC-box: with
Marked as: single
Average expression:
Expressed in:

Results: 3 promoters selected

[SGA file](#)

[FPS file](#)

[BED file](#)

LiftOver options hg38 (Dec 2013 GRCh38)

Submit

Sequence Extraction Tool (FASTA format)

From: -499 To: 100

Submit

Downstream Analysis

Motif Enrichment OProf ?
Motif Discovery FindM ?
Chromatin analysis ChIP-Cor ?

Конец 4-ой лекции