

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.*

ОСОБЕННОСТИ ТРАНСКРИПЦИИ У ЭУКАРИОТ:

-А) У эукариот транскрипция осуществляется при участии РНК-полимераз трех типов, которые обозначаются как pol I, pol II, и pol III. Можно выделить три группы генов:

- pol I транскрибируемые: гены рРНК (18S, 5.8S и 28S)
- pol II транскрибируемые: все белок-кодирующие гены + гены многих мяРНК (U1, U2, U4, U5, U11, U12 и др.)
- pol III транскрибируемые : гены тРНК + гены низкомолекулярных РНК (5S рРНК, 7SL РНК, U6 РНК, 7SK РНК и др.)

Б) Каждая РНК-полимераза имеет сложное, многосубъединичное строение, например, pol II включает более 10 субъединиц;

В) Для точной посадки РНК-полимераз на участок, прилежащий к старту транскрипции, нужны вспомогательные белки – базальные (или общие) транскрипционные факторы;

Г) Для достижения необходимого уровня транскрипции каждого гена, помимо базальных (общих) транскрипционных факторов, требуются другие белки (транскрипционные факторы, коактиваторы, корепрессоры, медиаторы и др.);

Д) Промоторы генов, транскрибируемых каждым типом полимераз, имеют характерное строение

✓ **Е) Транскрипция генов эукариот зависит от состояния хроматина и плотности нуклеосомной упаковки.**



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**

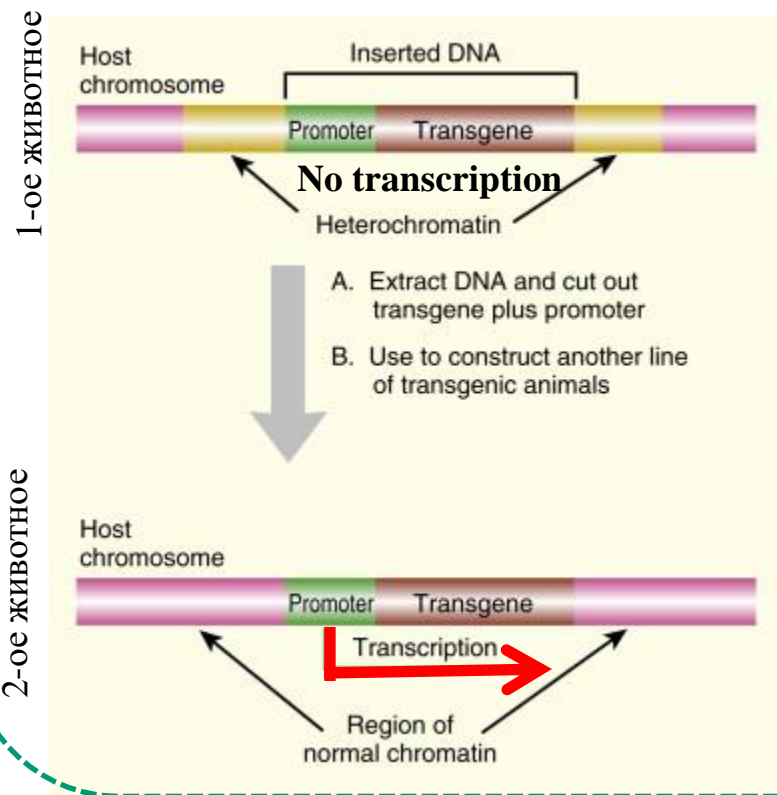
ЛОКУС-КОНТРОЛИРУЮЩИЙ РАЙОН (LCR)

ЛОКУС-КОНТРОЛИРУЮЩИЕ РАЙОНЫ БЫЛИ ИДЕНТИФИЦИРОВАНЫ В ПРОЦЕССЕ ИССЛЕДОВАНИЯ ЭКСПРЕССИИ ТРАНСГЕННЫХ КОНСТРУКЦИЙ.

При работе с трансгенами часто возникали проблемы:

- 1) Трансген не экспрессировался;
- 2) Уровень экспрессии не был пропорционален количеству встроенных копий трансгена.

ПРИМЕР ЭКСПЕРИМЕНТА ПО ТРАНСГЕНЕЗУ



ДНК, несущая трансген, была вставлена в геном другого организма для создания трансгенного животного. В этом случае ДНК была вставлена в область гетерохроматина. Хотя были получены трансгенные животные, трансген не экспрессировался.

Вставленную ДНК выделяли и использовали для получения другого трансгенного животного. Трансген экспрессировался у второго животного, что говорило о том, что трансгенная конструкция не была повреждена. Отсутствие экспрессии у первого животного можно объяснить эффектом положения

Biotechnology (Second Edition) Applying the Genetic Revolution 2016,
David P.ClarkNanette J.Pazdernik

Chapter 16 - Transgenic Animals Pages 493-521

<https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/transgenic-plant>

ЛОКУС-КОНТРОЛИРУЮЩИЙ РАЙОН (LCR)

Впервые LCR был выявлен в кластере бета-глобиновых генов человека

ОПРЕДЕЛЕНИЕ: 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 .

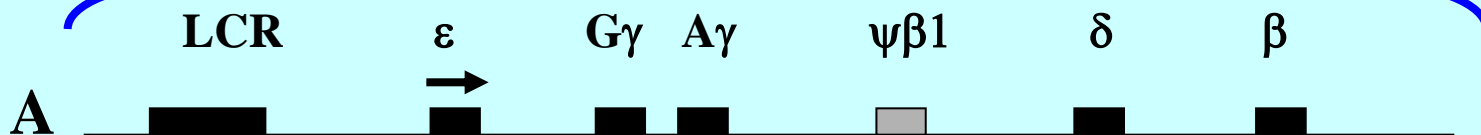
Как правило, локус-контролирующие районы регулируют гены, находящиеся в кластерах и обеспечивают их ткане- и стадие-специфическую экспрессию .

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

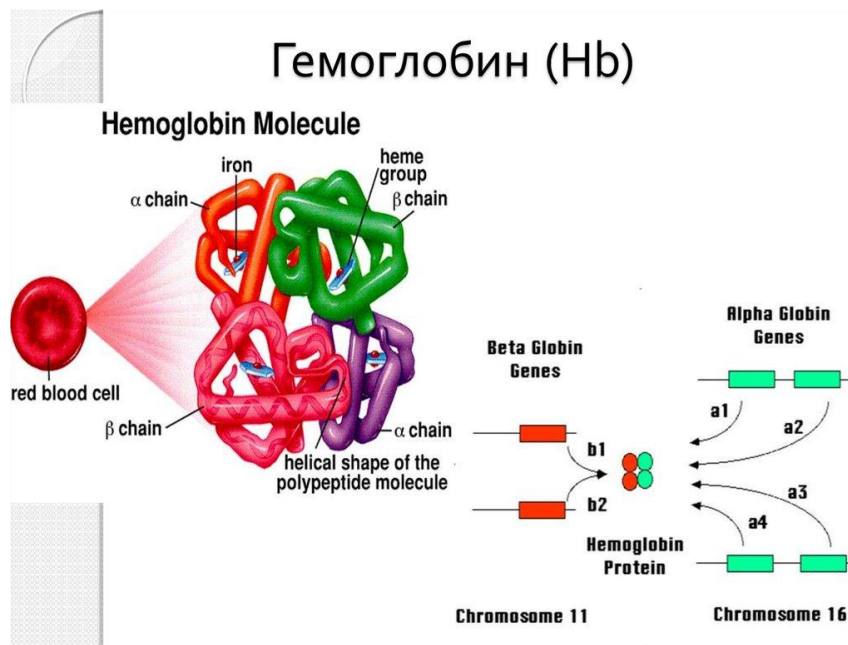
70 тыс. пар оснований



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

Levings PP1, Bungert J The human beta-globin locus control region. Eur J Biochem. 2002 Mar;269(6):1589-99

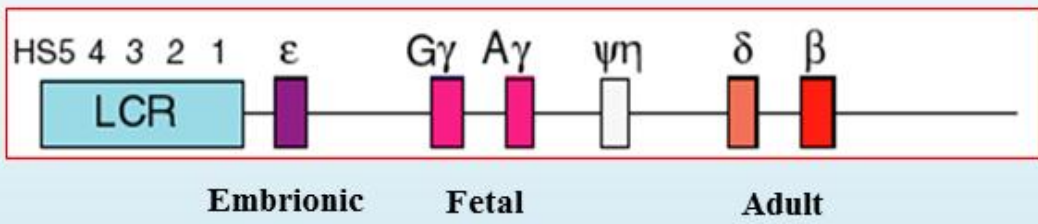
β -ГЛОБИН –
структурная единица
молекулы гемоглобина



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

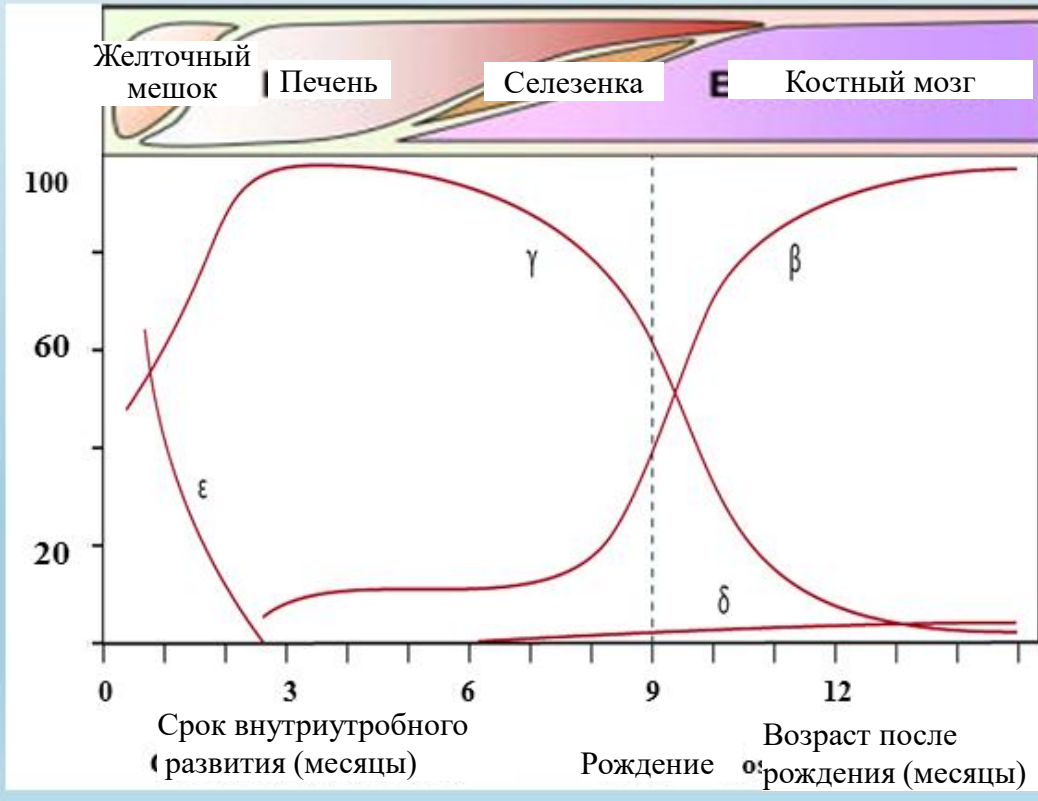
B-Globin locus

Chr11



Орган, где осуществляется эритропоэз

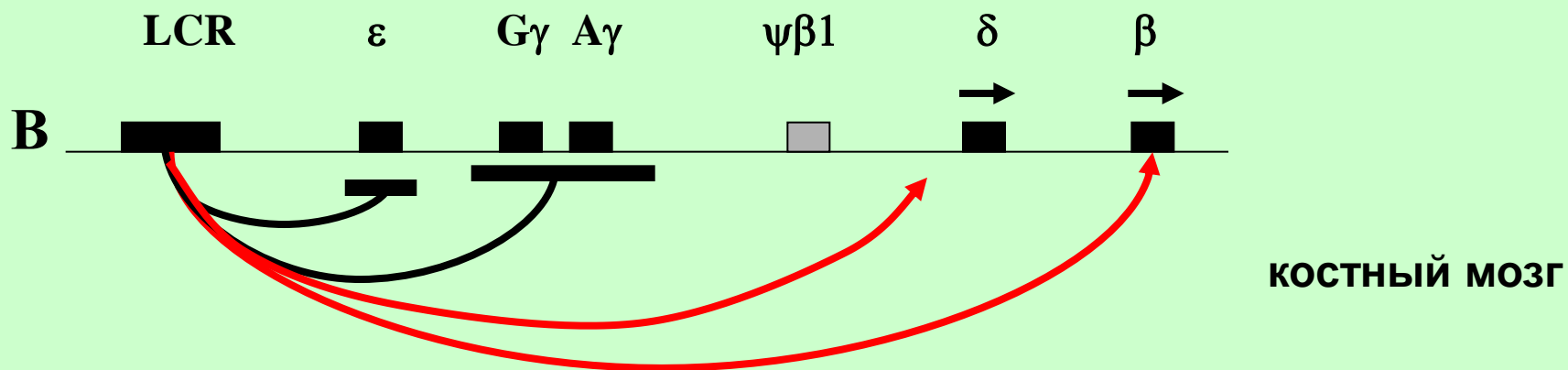
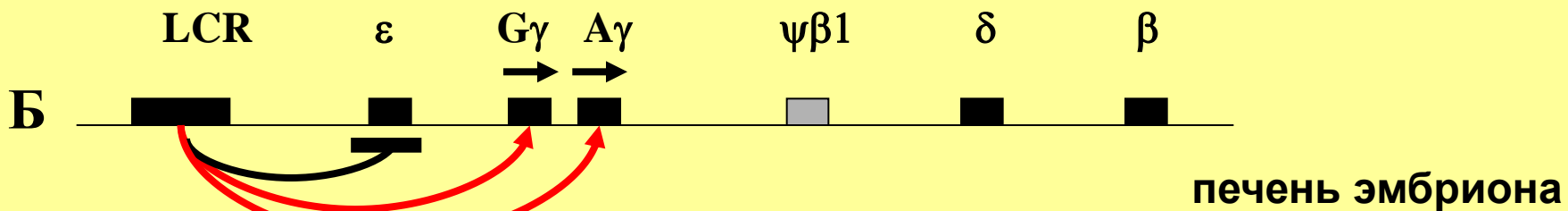
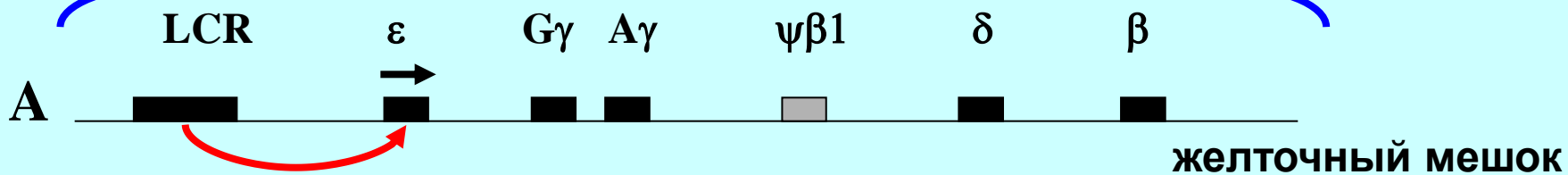
Уровень экспрессии глобинов (%)



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.

ЛОКУС-КОНТРОЛИРУЮЩИЙ РАЙОН КЛАСТЕРА β-ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА

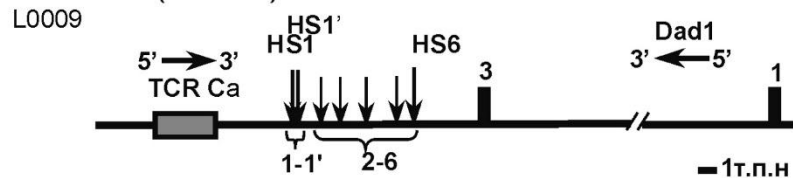
70 тыс. пар оснований



ε, Gγ, Aγ, δ, β - гены гемоглобинов; ψβ1 - псевдоген

Примеры структурно-функциональной организации LCR

TCR α/δ (мышь)



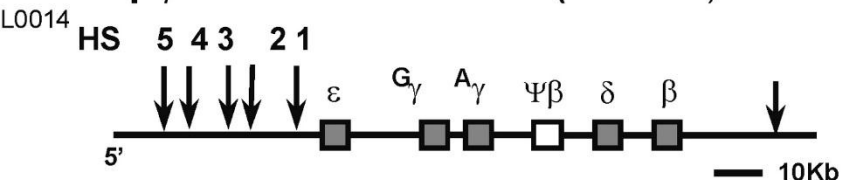
HS1 - энхансер

HS1, HS1' район определяющий распространенность в тканях

HS2 - HS6 район открывающий хроматин

Цифрами обозначены экзоны гена *Dad1*

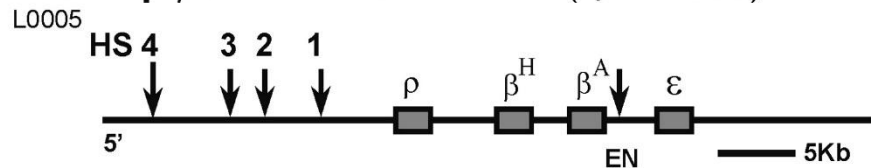
Кластер β глобиновых генов (человек)



HS5 - инсулятор

HS4 - HS1 - энхансер

Кластер β глобиновых генов (цыпленок)

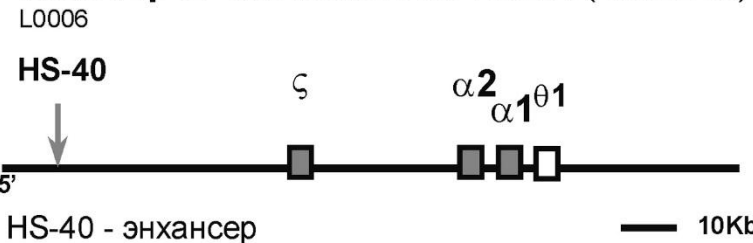


HS4 - инсулятор

HS3, HS2 - энхансер

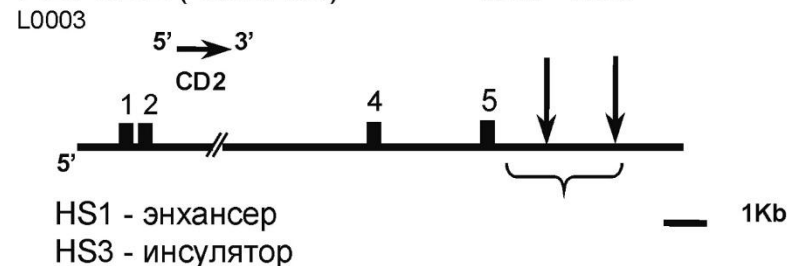
EN - межгенный энхансер

Кластер α -глобиновых генов (человек)



HS-40 - энхансер

Ген *CD2* (человек)

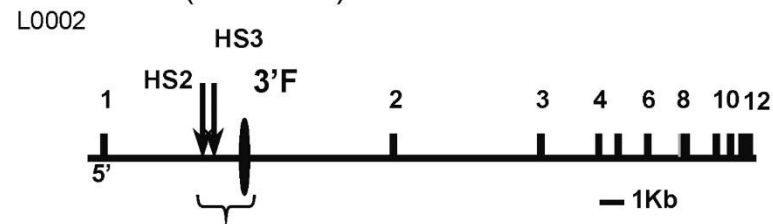


HS1 - энхансер

HS3 - инсулятор

Цифрами обозначены экзоны гена *CD2*

Ген *ADA* (человек)



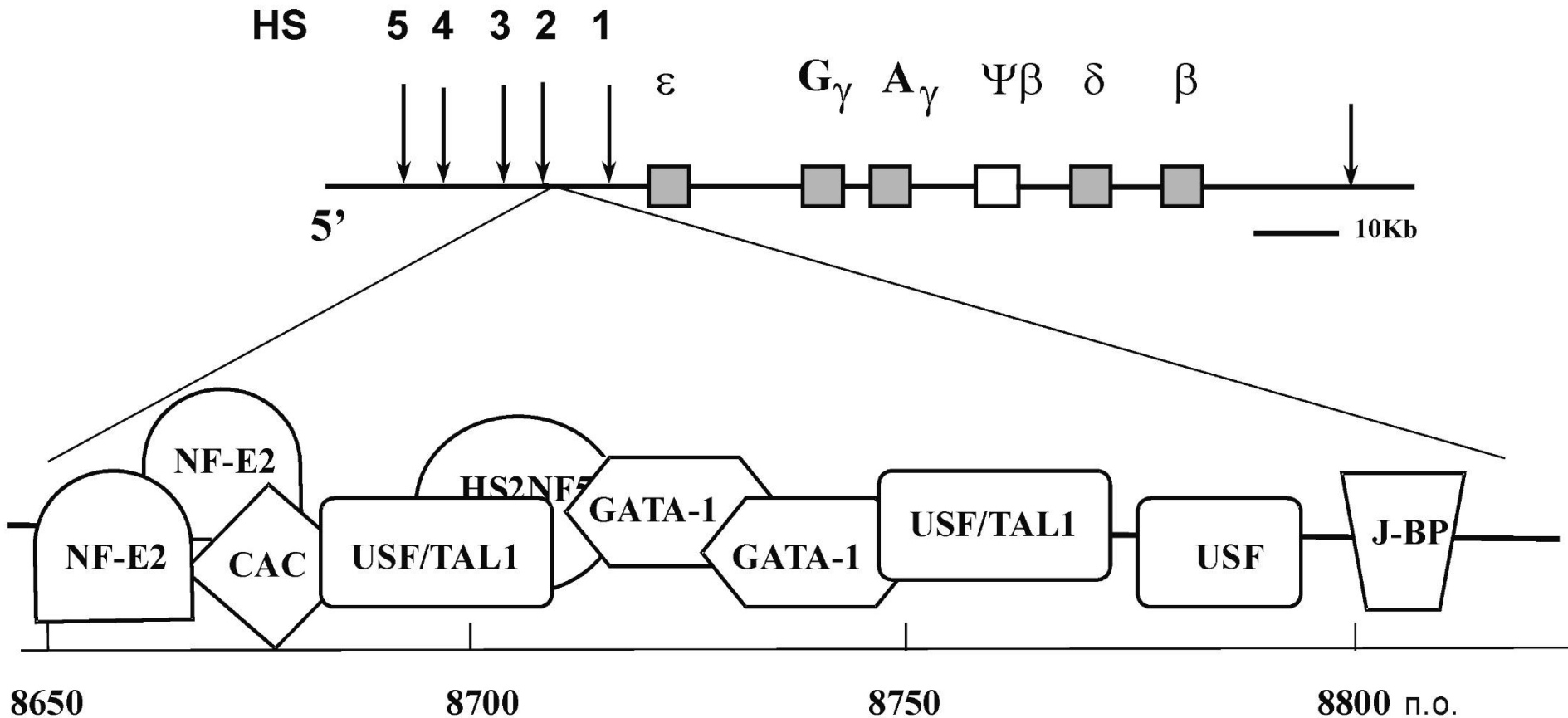
HS2 - 5' вспомогательный элемент

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

ТРЕХМЕРНАЯ СТРУКТУРА ХРОМАТИНА

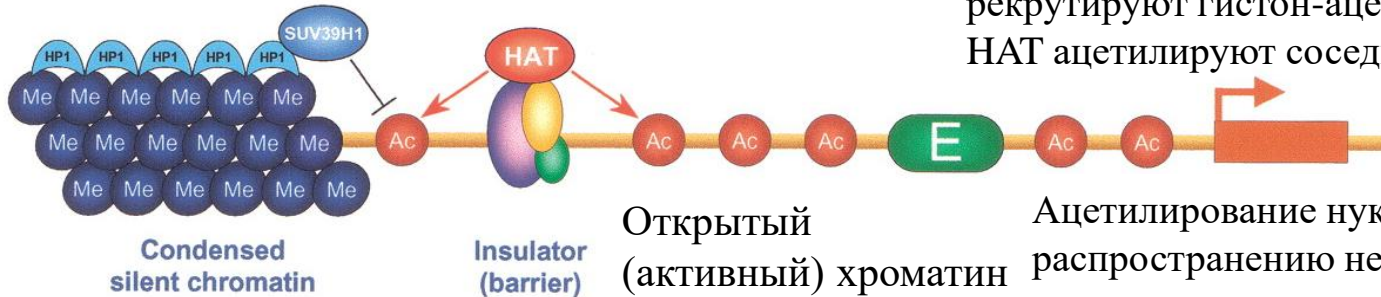
1. LCR – локус-контролирующие районы;
2. Инсуляторы;
3. Транскрипционные факторы CTCF;
4. Когезин (cohesin) – белковый комплекс с кольцевой структурой;
5. Хромосомный оперон, транскрипционные фабрики;
6. Энхансерная РНК (eRNA) и ее регуляторные функции



Инсулятор - участок ДНК, выполняющий специализированную регуляторную функцию

Функции инсулятора:

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



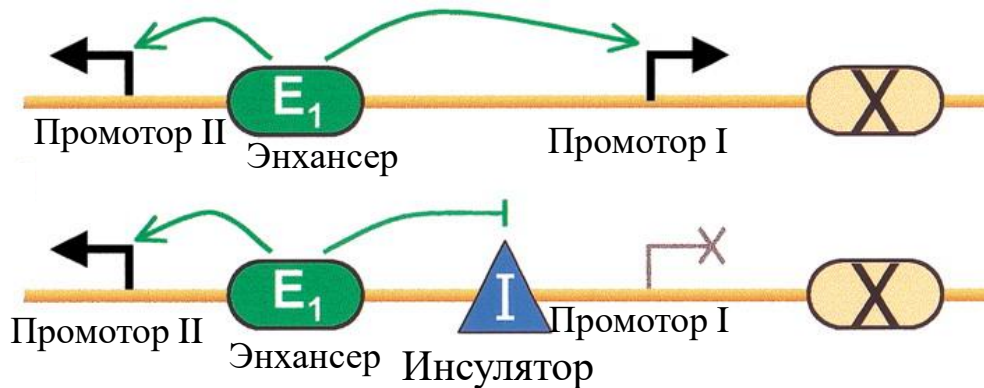
Белки, взаимодействующие с инсулятором рекрутируют гистон-ацетилтрансферазы (НАТ). НАТ ацетилируют соседние нуклеосомы.

Открытый (активный) хроматин

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

Adam G. et al., Insulators: many functions, many mechanisms *Genes & Dev.* 2002, 16, 271-288

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



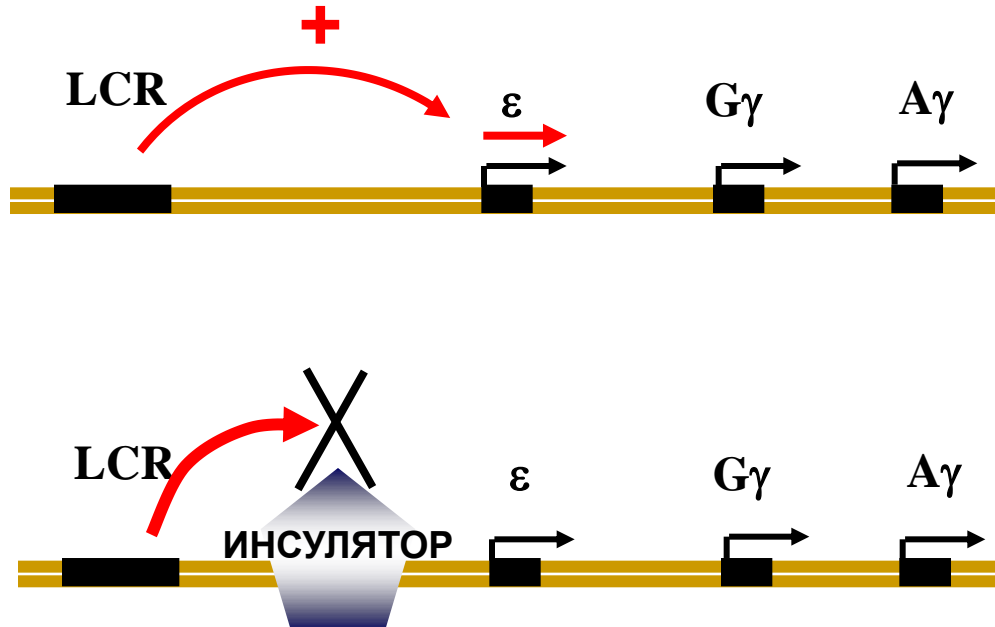
Инсулятор блокирует активность энхансера только по отношению к промотору I. Если вместо инсулятора расположить негативный регуляторный элемент (сайленсер), то блокирующее влияние будет распространяться на оба промотора

Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet.* 2006 Sep;7(9):703-13.

Инсулятор - участок ДНК, выполняющий специализированную регуляторную функцию

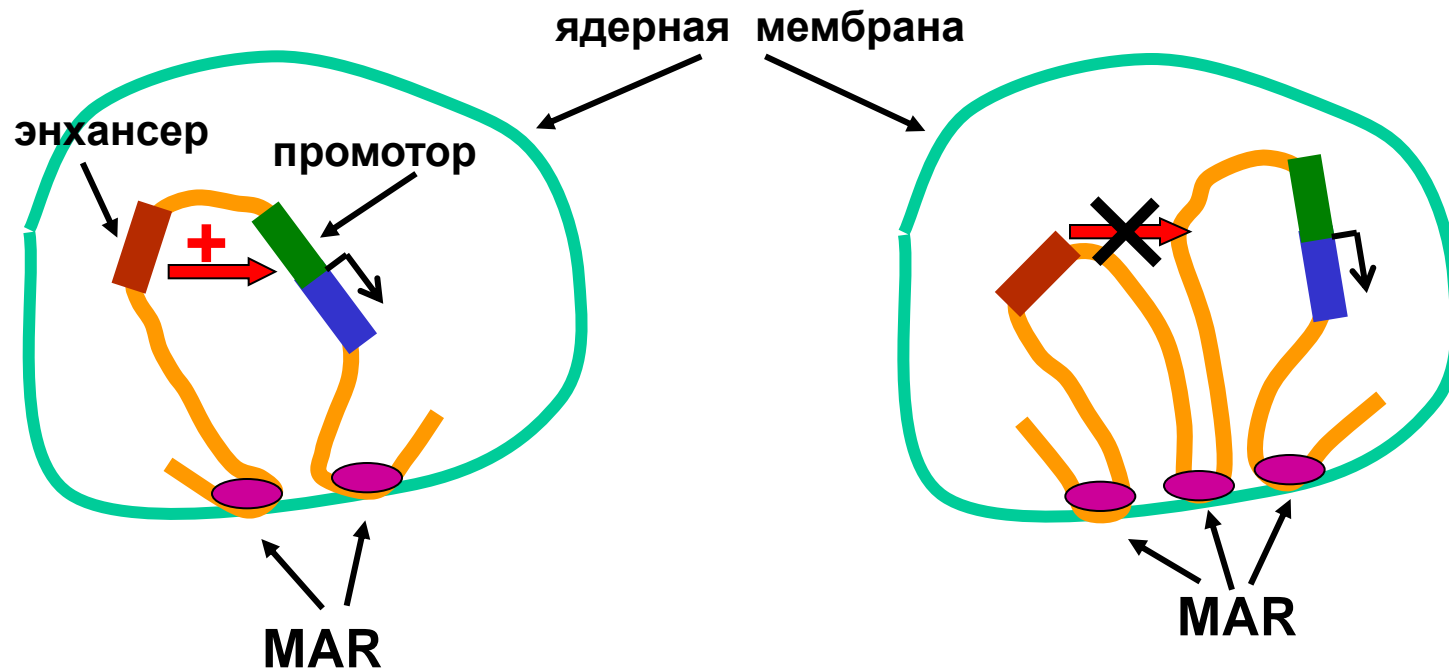
Функции инсулятора:

- 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

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



Роль транскрипционных факторов CTCF при формировании петель ДНК

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

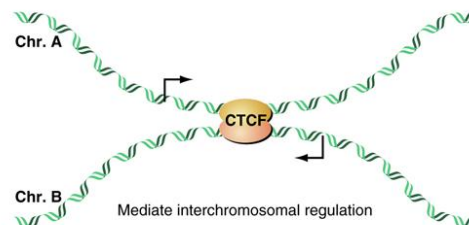
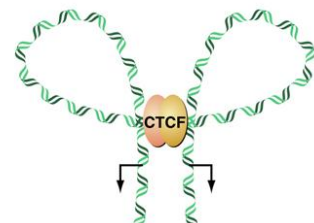
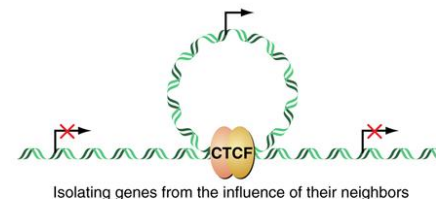


WebLogo моноклеотидной весовой матрицы сайта связывания фактора CTCF из базы HOCOMOCO (CTCF_HUMAN.H11MO.0.A)

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.

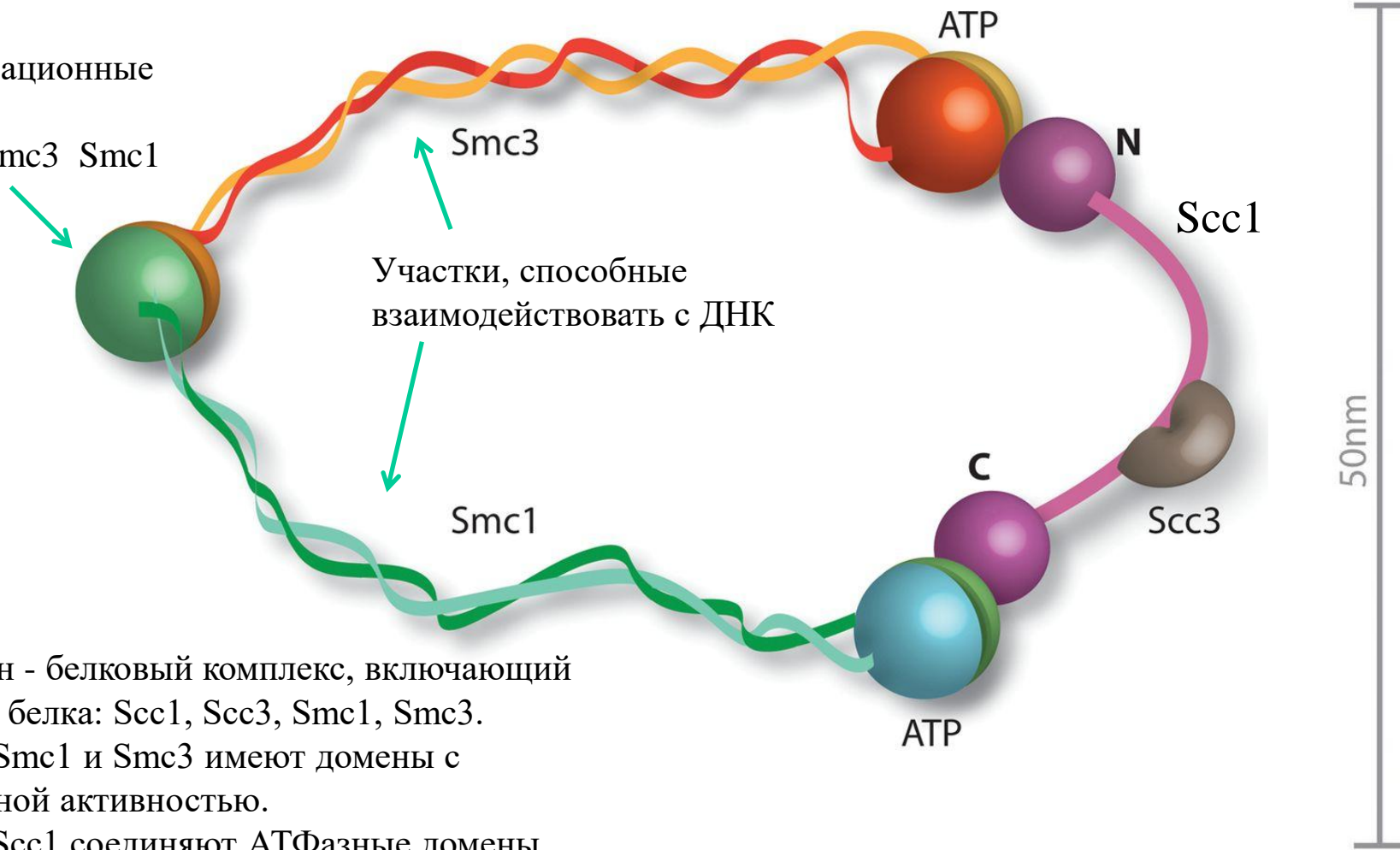
CTCF является маркером участков ДНК, разделяющих активный и репрессированный хроматин

Nunez E., Fu X-D., Rosenfeld M.G. Nuclear organization in the 3D space of the nucleus — cause or consequence? *Current Opinion in Genetics & Development*, 2009, 19, P. 424–436.



Когезин – комплекс белков , формирующий кольцеобразную структуру

Димеризационные
домены
белков Smc3 Smc1



Участки, способные
взаимодействовать с ДНК

Smc1

ATP

50nm

Когезин - белковый комплекс, включающий четыре белка: Scc1, Scc3, Smc1, Smc3.

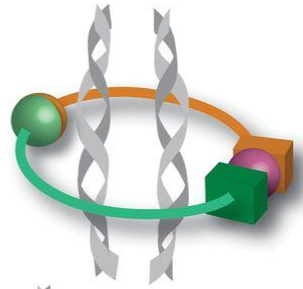
Белки Smc1 и Smc3 имеют домены с АТФазной активностью.

Белок Scc1 соединяют АТФазные домены Smc1 и Smc3, стабилизируя структуру кольца.

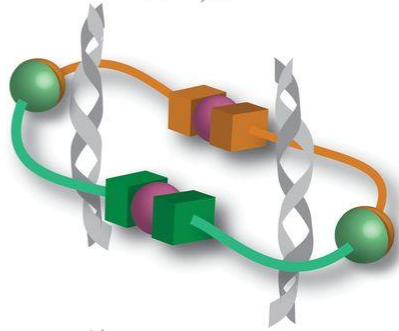
Белок Scc3 связывается с С-терминальным участком белка Scc1

Модели 3-Д структур хроматина с участием когезина

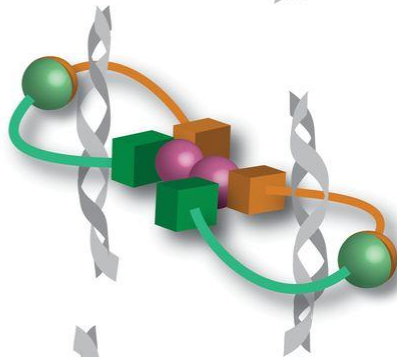
A) *Ring model:*



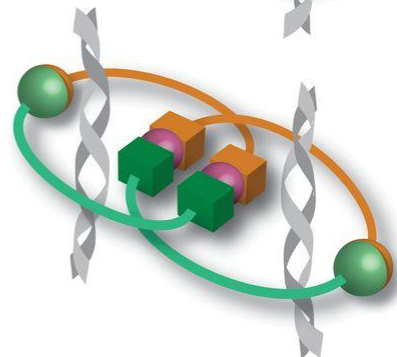
B) *Handcuff model (1):*



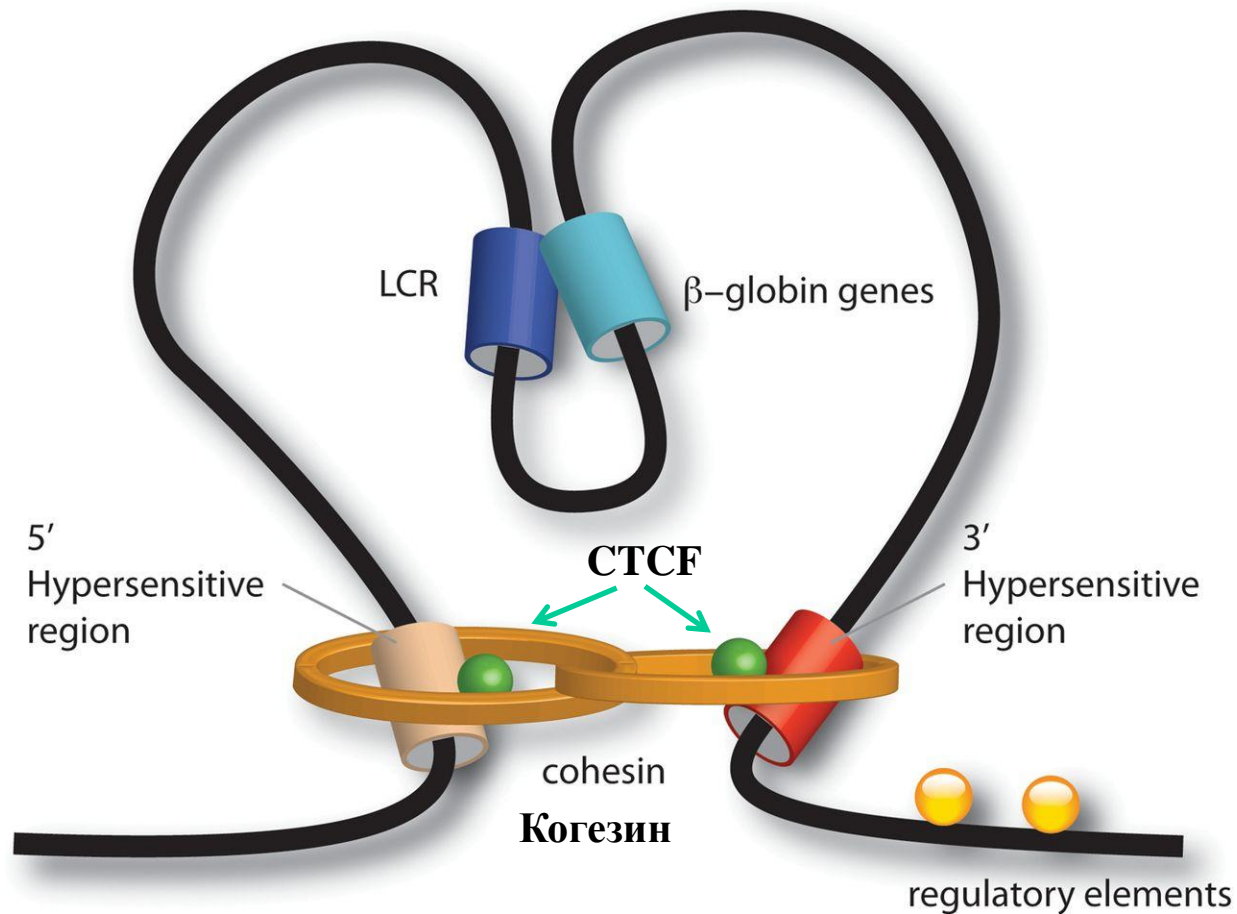
Handcuff model (2):



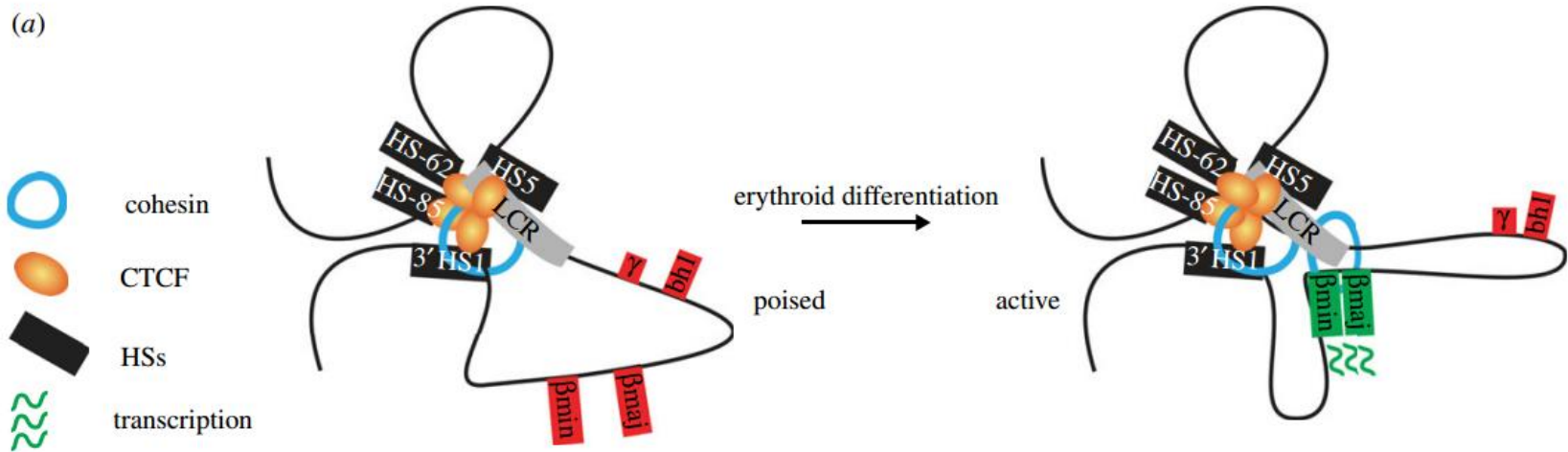
Handcuff model (3):



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

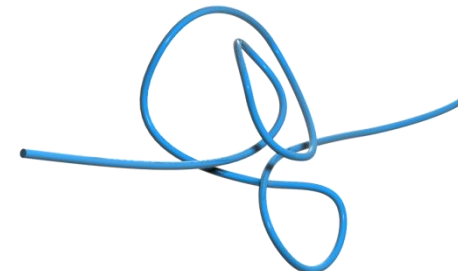
Исследованы контакты в клеточных культурах человека MCF7 и K562



Промоторная модель

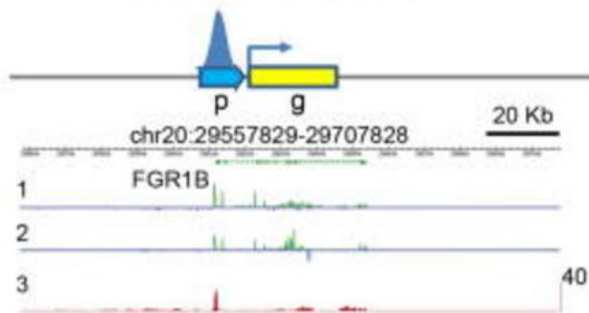


Одногенная модель
(Промотор-Энхансер)
Single-gene interaction model

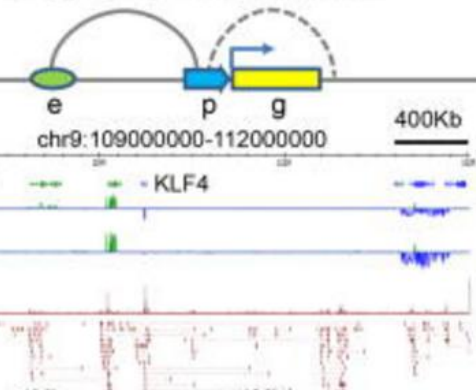


Мультигенная модель (хромосомный оперон - "chromoperon")
Multi-gene interaction model

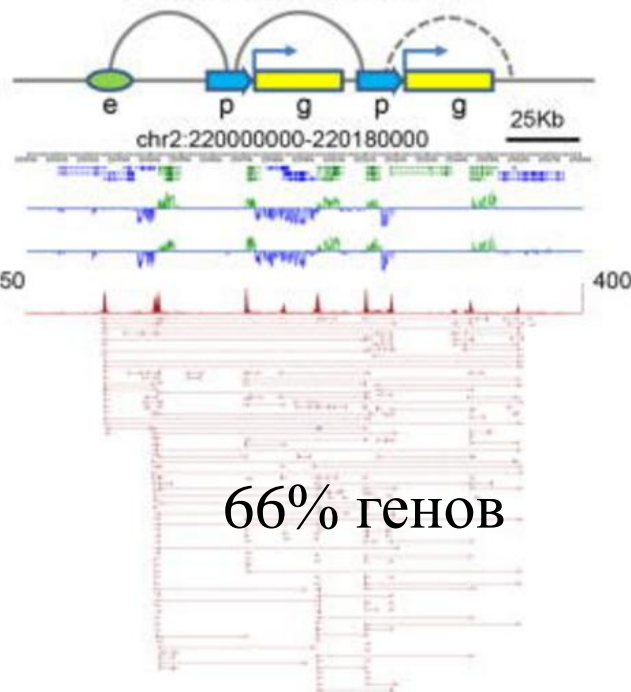
Basal promoter model



26% ГЕНОВ



6% ГЕНОВ

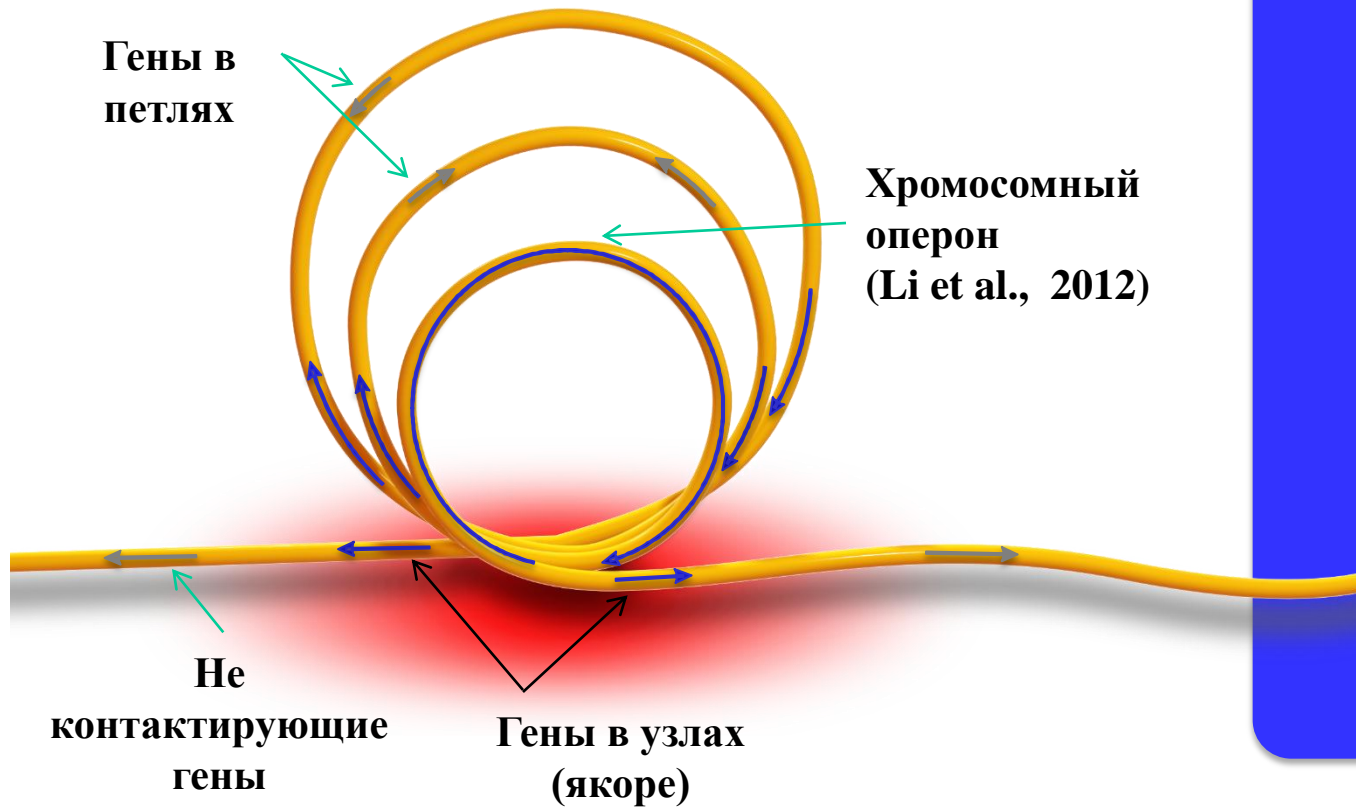


66% ГЕНОВ

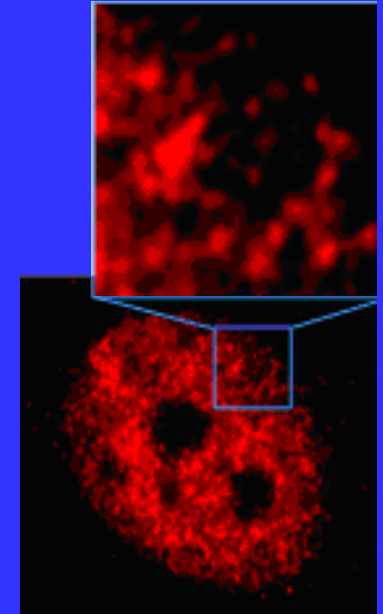
Li G. Et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

Хромосомные опероны и транскрипционные фабрики

Структуры, найденные с помощью ChIA-PET



Ядро клетки – участки транскрипции под микроскопом



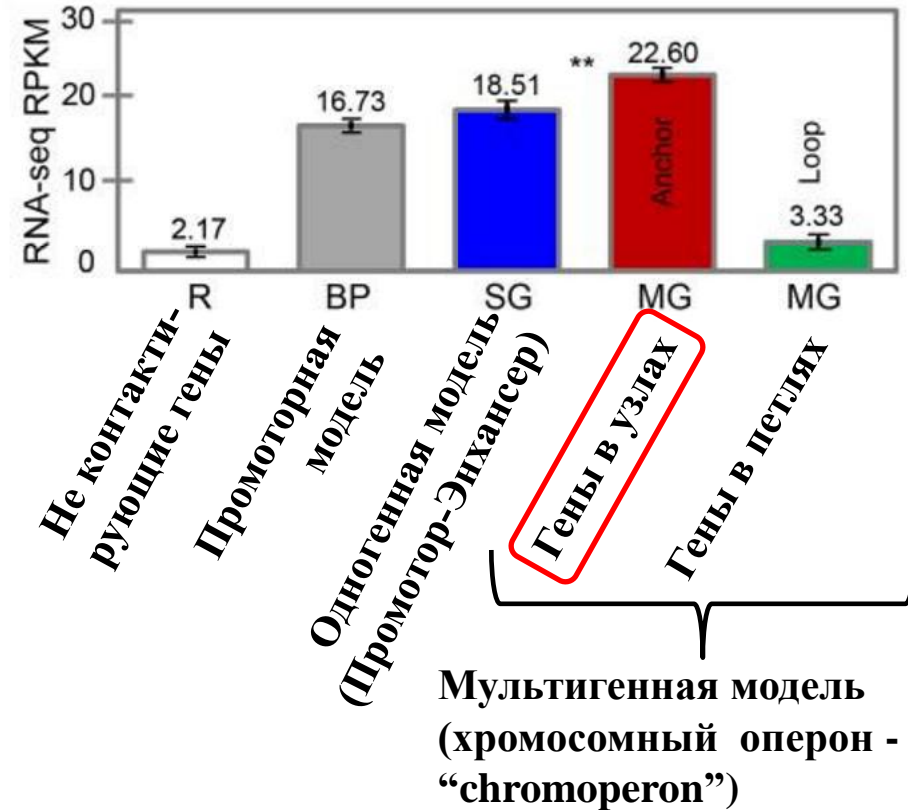
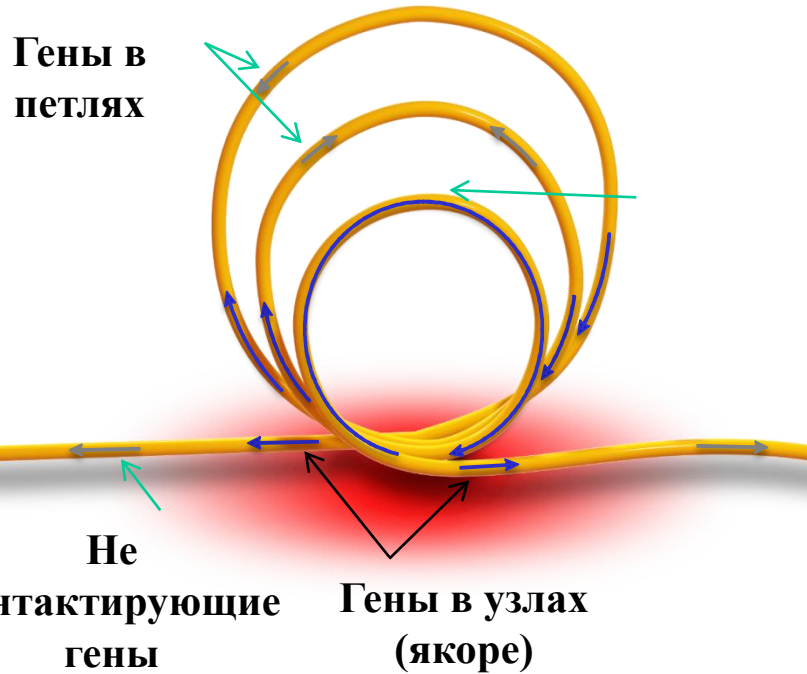
FISH

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

Роль трехмерной структуры хроматина в регуляции транскрипции

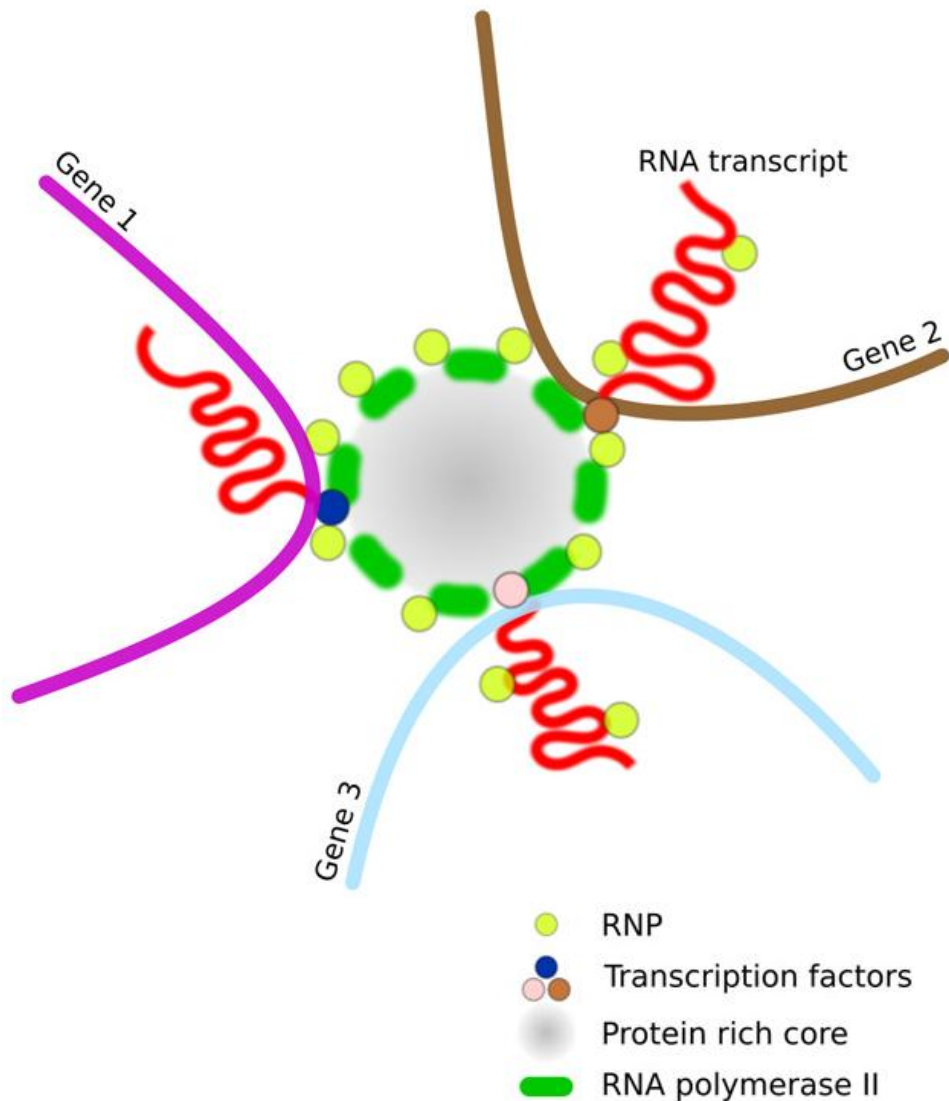
Структуры, найденные с помощью методики ChIA-PET, в MCF7 клетках человека

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:

Промоторный район, энхансеры, сайленсеры - РЕГУЛЯТОРНЫЕ ЕДИНИЦЫ

Энхансеры – регуляторные единицы, активирующие транскрипцию.

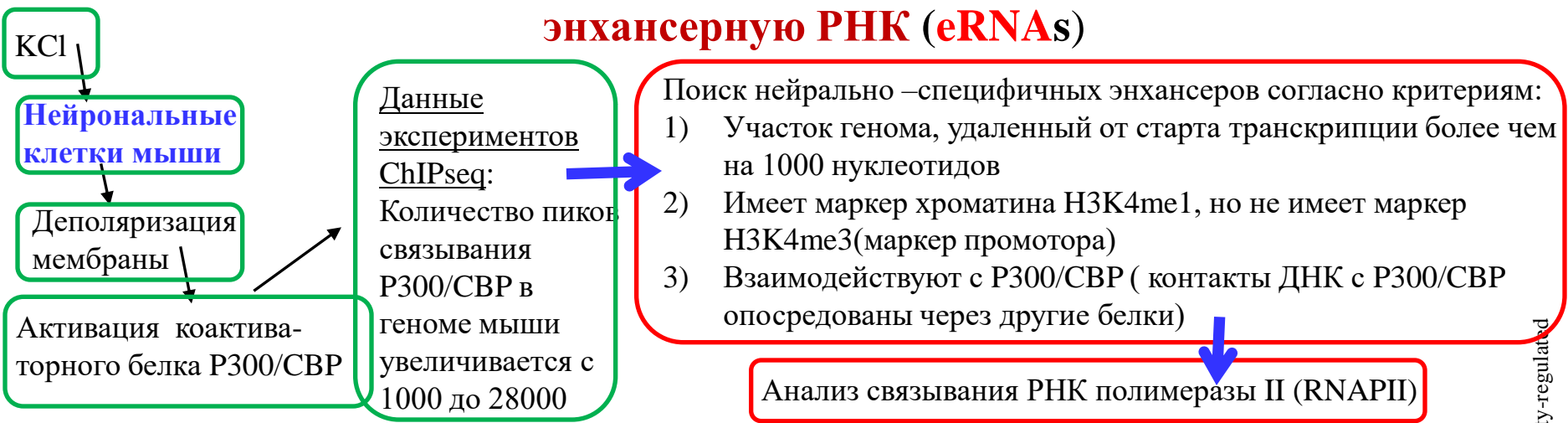
Энхансеры усиливают транскрипцию гена, управляемую определенным промотором, и оказывают свой эффект как в прямой, так и в обратной ориентации и в различной локализации (5' - либо 3' -) по отношению к промотору.

.....

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)



Kim T.K. et al., Widespread transcription at neuronal activity-regulated enhancers Nature, 2010 May 13; 465(7295): 182–187.

Расстояние от центра пика связывания СВР

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:

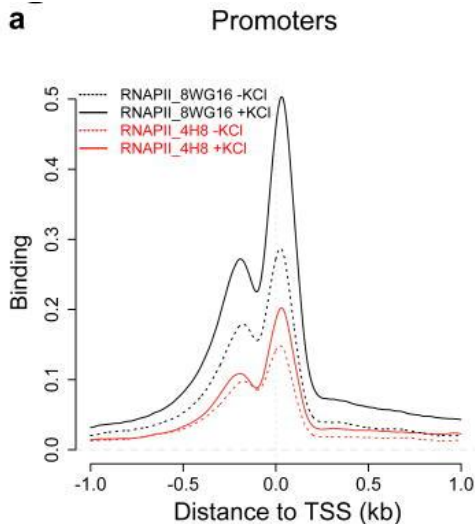
Количество пиков связывания P300/CBP в геноме мыши увеличивается с 1000 до 28000

Поиск нейтрально-специфичных энхансеров согласно критериям:

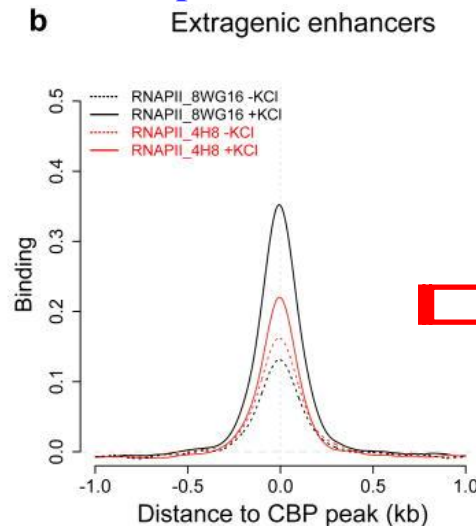
- 1) Участок генома, удаленный от старта транскрипции более чем на 1000 нуклеотидов
- 2) Имеет маркер хроматина H3K4me1, но не имеет маркер H3K4me3 (маркер промотора)
- 3) Взаимодействуют с P300/CBP (контакты ДНК с P300/CBP опосредованы через другие белки)

Анализ связывания РНК полимеразы II (RNAPII)

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

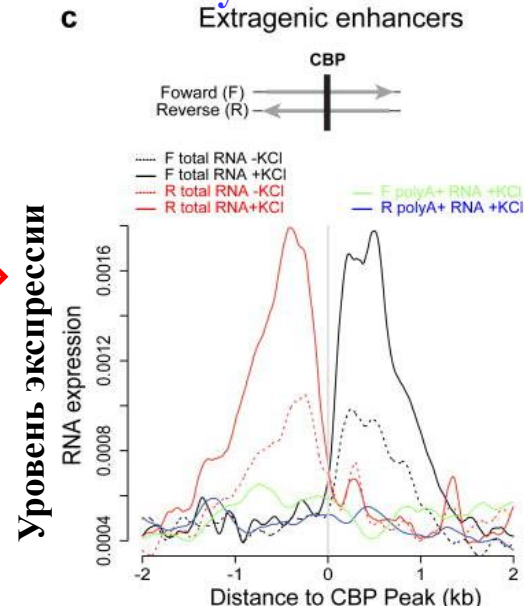


Ассимметричный пик – транскрипция идет в одном направлении



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

Экспрессия РНК в энхансерных участках



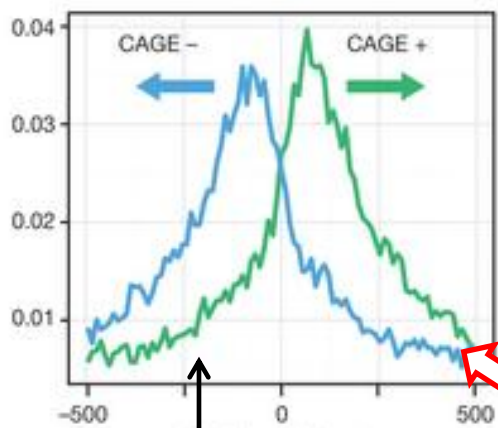
Уровень экспрессии

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)

Уровень экспрессии
= доля энхансеров,
с которых считывается РНК



Позиция относительно центра
(пик связывания белка Р300)

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 Р300 (выявленных методикой ChIP-seq).

Обнаруженные таким образом энхансеры были центрированы относительно сайтов связывания Р300. Для энхансеров были рассчитаны усредненные уровни экспрессии (по данным экспериментов CAGE).

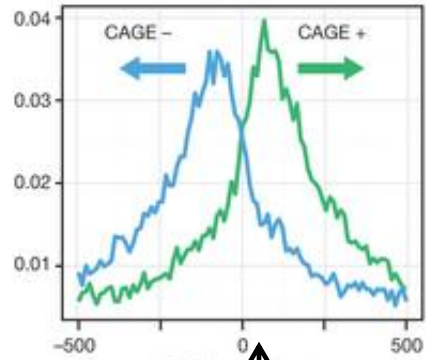
Были выявлены транскрипты (короткие РНК), которые считывались как с прямой (+), так и обратной (-) цепи ДНК в районе энхансера. Расстояние между (+) и (-) пиками соответствовало длине участка ДНК, упакованного в нуклеосому (180 п.о.).

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

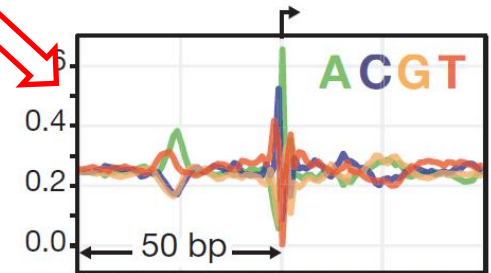


Позиция относительно центра
(пик связывания белка P300)

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

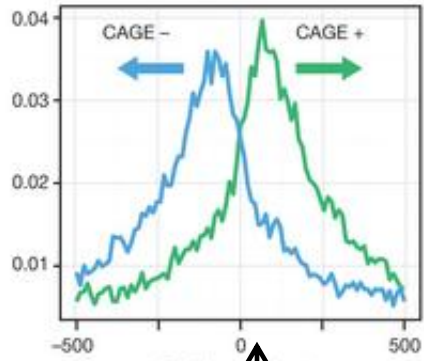
Частота встречаемости нуклеотида



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

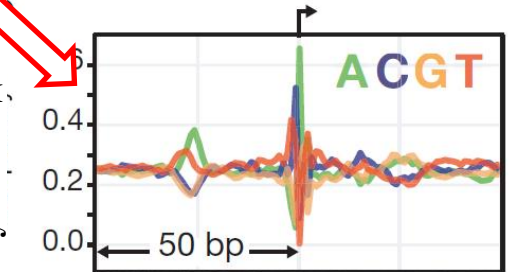


Позиция относительно центра
(пик связывания белка Р300)

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

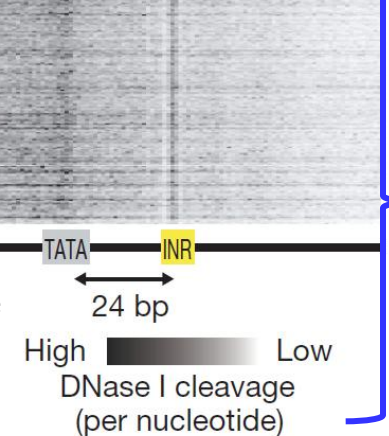
Inr – инициаторный элемент, можно описать консенсусным мотивом YYANWYY, где Y = C либо T

Частота встречаемости нуклеотида



Enhancer
CAGE 5' ends

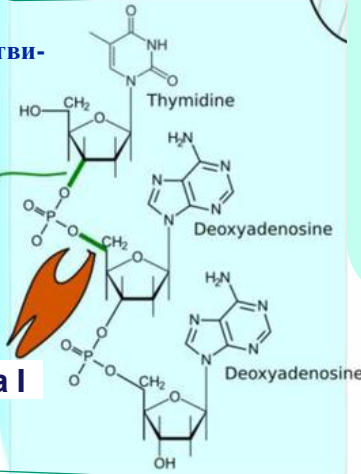
Расщепление ферментом DNase I



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:

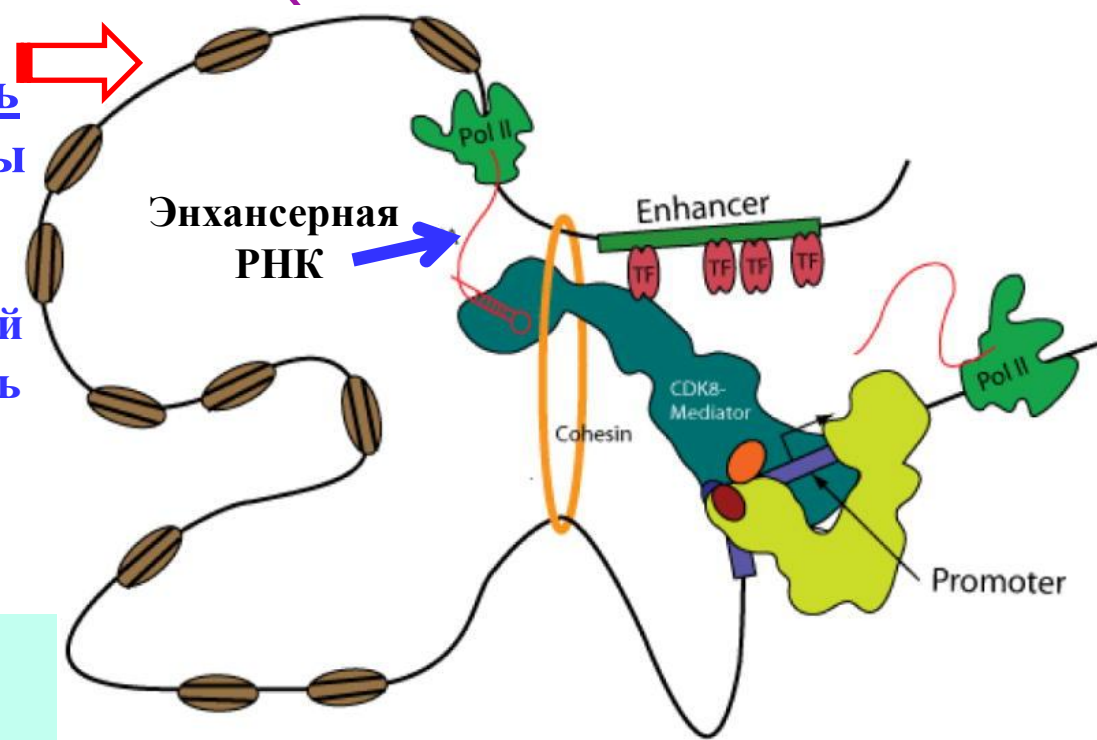
- Облегчает образование петель ДНК, сближающих энхансеры и промоторные районы.
- Способствует освобождению РНК-полимеразы из комплекса , обеспечивающего остановку (паузу) на стадии ранней элонгации
-

Роль энхансерной РНК:

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

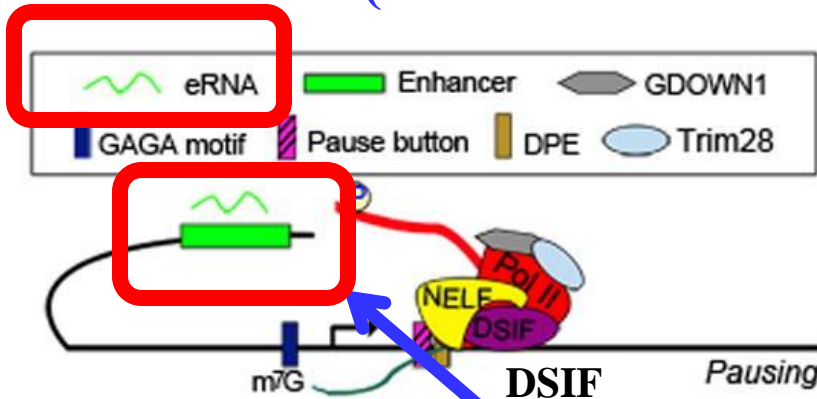
Показано, что в реализации этой функции участвует CDK8 модуль медиаторного комплекса, объединяющий субъединицы MED12, MED13, CCNC, CDK8.

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 полимеразы II связалась с медиаторным комплексом, модуль CDK8 должен отсоединиться от медиаторного комплекса. Медиаторный комплекс (без модуля CDK8) связывается с РНК полимеразой II. А 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:

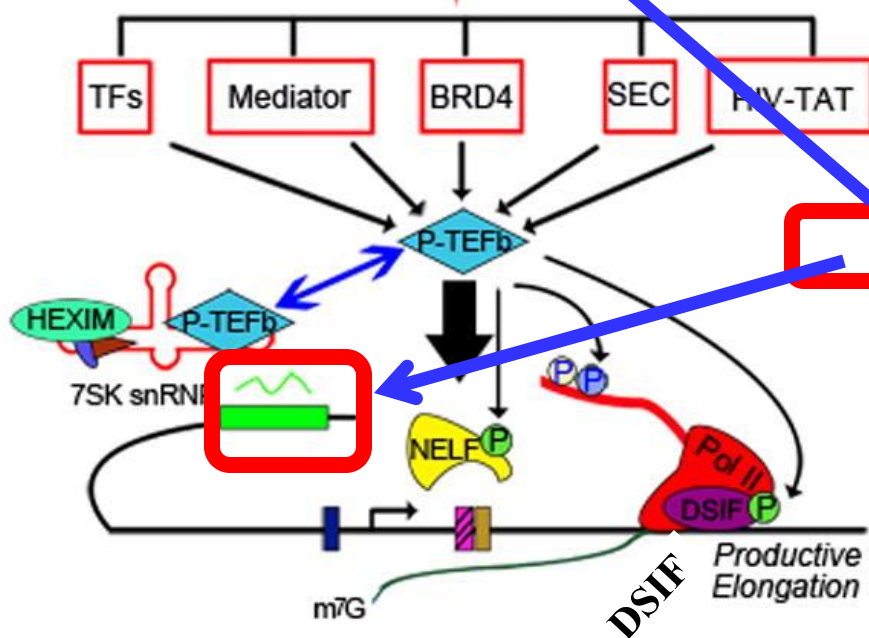
- Тепловом шоке (ТФ)

- Гипоксии (ТФ)

- Воспалению (Nf-kB, BRD4)

- Дифференцировке стволовых клеток (SEC, ТФ)

- Инфекции HIV (HIV-TAT)

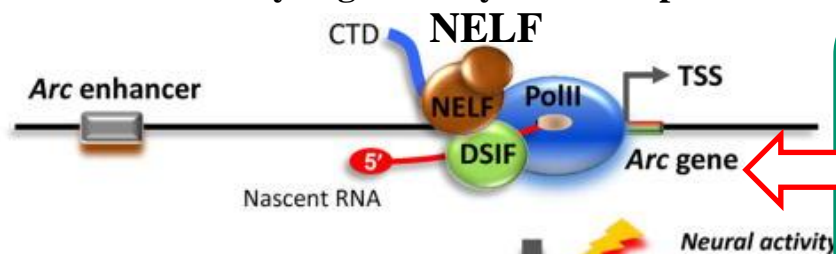


Роль энхансерной РНК:

Взаимодействие энхансерной РНК с ингибитором элонгации NELF

высвобождает РНК-полимеразу II из ингибиторного комплекса и активирует транскрипцию гена Arc

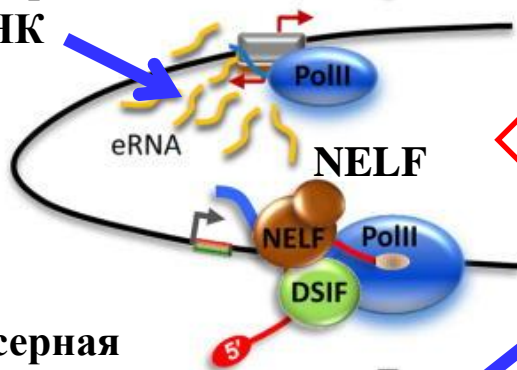
Ген Arc = Activity-regulated cytoskeletal protein



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

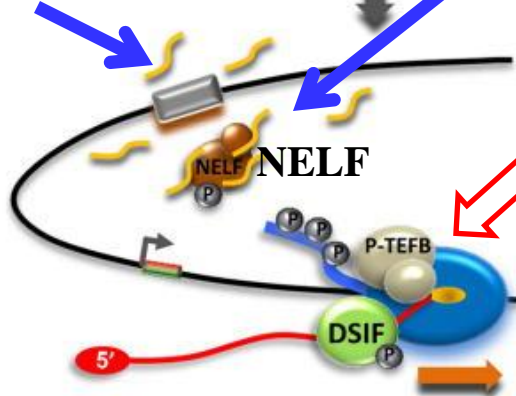
НЕГАТИВНЫЕ РЕГУЛЯТОРЫ ЭЛОНГАЦИИ:
NELF = negative elongation factor
DSIF = DRB Sensitivity Inducing Factor

Энхансерная РНК



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

Энхансерная РНК



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»

OXFORD ACADEMIC

Explore the 2018 NAR Database Issue today
82 new database descriptions

CLICK HERE TO READ

Sign In Register

Nucleic Acids Research

Issues Section browse Advance articles Submit Purchase About

All Nucleic Acids Re

Advanced Search

Impact Factor 11.561
5 year Impact Factor 10.235

Senior Executive Editors
Keith Fox
Barry Stoddard

Latest Issue
Volume 46, Issue 17
28 September 2018

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

SCIENCE & MATHEMATICS
Search opportunities on the Oxford University Press Journals Career Network
Your perfect job is closer than you think

Nucleic Acids Research

Issues Section browse Advance articles Submit Purchase About

All Nucleic Acids Research

Advanced Search

Browse issues

Year 2018 Issue Volume 46, Issue D1, 4 January 2018

Browse by volume

Nucleic Acids Research

Volume 46, Issue D1
4 January 2018
ISSN 0305-1948
EISSN 1362-4962

Editorial

Database Resources
Nucleic acid sequences, structure, and regulation
Protein sequence and structure, motifs, and domains
Metabolic and signaling pathways, proteomics and
Viruses, protozoa and
Human genome, model organisms, comparative genomics

Volume 46, Issue D1, 4 January 2018
Database issue

Page 1 of 2 1 2 Next

EDITORIAL

The 2018 *Nucleic Acids Research* database issue and the online molecular biology database collection

Daniel J Rigden; Xosé M Fernández

Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D1–D7, <https://doi.org/10.1093/nar/gkx1235>

Abstract View article

DATABASE RESOURCES

Database resources of the National Center for Biotechnology Information

NCBI Resource Coordinators

Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D8–D13, <https://doi.org/10.1093/nar/gkx1095>

Email alerts

New issue alert
Advance article alerts
Subject alert

Receive exclusive offers and updates from Oxford Academic

RSS Feeds

RSS Feed - Latest Issue Only
RSS Feed - Advance Articles
RSS Feed - Open Access

Volume 46, Issue D1, 4 January 2018
Database issue

Page 1 of 2 1 2 Next

EDITORIAL

The 2018 *Nucleic Acids Research* database issue and the online molecular biology database collection

Daniel J Rigden; Xosé M Fernández

Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D1–D7, <https://doi.org/10.1093/nar/gkx1235>

Abstract

The 2018 *Nucleic Acids Research* Database Issue contains 181 papers spanning molecular biology. Among them, 82 are new and 84 are updates describing resources that appeared in the Issue previously. The remaining 15 cover databases most recently published elsewhere. Databases in the area of nucleic acids include 3DIV for visualisation of data on genome 3D structure and RNArchitecture, a hierarchical classification of RNA families. Protein databases include the established SMART, ELM and MEROPS while GPCRdb and the newcomer STCRDab cover families of biomedical interest. In the area of metabolism, HMDB and Reactome both report new features while PULDB appears in NAR for the first time. This issue also contains reports on genomics resources including Ensembl, the UCSC Genome Browser and ENCODE. Update papers from the IUPHAR/BPS Guide to Pharmacology and DrugBank are highlights of the drug and drug target section while a number of proteomics databases including proteomicsDB are also covered. The entire Database Issue is freely available online on the *Nucleic Acids Research* website (<https://academic.oup.com/nar>). The NAR online Molecular Biology Database Collection has been updated, reviewing 138 entries, adding 88 new resources and eliminating 47 databases. It is available at <http://www.oxfordjournals.org/nar/database/c/>.

Ссылка на «online Database Collection»

Журнал NAR: online Database Collection

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

OXFORD ACADEMIC Journals

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

NAR Database Summary Paper Alphabetic List

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

- 1000 Genomes Selection Browser Engelken, Johannes; Pybus, Marc; Dall'Olio, Giovanni; Luisi, Pierre; Uzkudun, Mani Carreño-Torres, Angel; Pavlidis, Pavlos; Laayouni, Hafid; Bertranpetit, Jaume
Signature of selection in the human genomes
[database](#) [summary](#)
- 16S and 23S Ribosomal RNA Mutation Database Triman K.L.
16S and 23S ribosomal RNA mutations
[database](#) [summary](#)
- 2D-PAGE Pleissner, K.-P., Eifert, T., Buettner, S., Knipper, J., Schmelzer, P., Stein, R., Schmidt, F., Mattow, J., Zimny-Armdt, U., Schmid, M., Jungblut, P.R.
Proteome database system for microbial research
[database](#) [summary](#)
- 2P2Idb Basse, M.J., Betzi, S., Bourgeas, R., Bouzidi, S., Chetrit, B., Hamon, V., Morelli, X., and Roche, P.
2P2Idb - database dedicated to the modulation of protein-protein interactions
[database](#) [summary](#)
- 3D rRNA modification maps
Locations of modified rRNA nucleotides within the 3D structure of the ribosome

Navigation menu:
▶ [Compilation Paper](#)
▶ [Category List](#)
▶ [Alphabetical List](#)
▶ [Category/Paper List](#)
▶ [Search Summary Papers](#)

Полный список баз данных,
сгруппированных по категориям .

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

OXFORD ACADEMIC Journals

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

NAR Database Summary Paper Category List

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

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

The NAR online Molecular Biology Database Collection has been updated, reviewing 138 entries, adding 88 new resources and eliminating 47 discontinued URLs, bringing the current total to 1737 databases.

Информация о базах данных по регуляции транскрипции в «2019 NAR Database Summary Paper Category List»

OXFORD ACADEMIC Journals

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

NAR Database Summary Paper

Nucleotide Sequence Databases
ARED-Plus
lncRNASNP2
miRTarBase

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
- dbSUPER
- DBTSS
- DBTSS/DBKERO
- DoOP - Databases of Orthologous Promoters
- DPRP
- ECRbase
- Factorbook
- FlyFactorSurvey
- FlyTF
- GeneNet
- GenomeTraFaC
- Greglist
- GTRD
- HOCOMOCO
- HOCOMOCO
- HTPSELEX
- JASPAR
- JASPAR
- MachiBase
- MAPPER

77 строчек

1. 3D-Footprint
2. ABS
3. ACTIVITY
4. AGRIS
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. dbSUPER
17. DBTSS
18. DBTSS/DBKERO
19. DoOP
20. DPRP
21. ECRbase
22. Factorbook
23. FlyFactorSurvey
24. FlyTF
25. GeneNet
26. GenomeTraFaC
27. Greglist
28. GTRD
29. HOCOMOCO
30. HOCOMOCO
31. HTPSELEX
32. JASPAR
33. JASPAR
34. MachiBase
35. MAPPER
36. ODB - Operon database
37. OnTheFly
38. ooTFD
39. ORegAnno
40. PAZAR
41. Plant Stress-Responsive Gene Catalog
42. PlantProm
43. PReMod
44. PRODORIC
45. PRODORIC2
46. REDfly
47. RegPrecise
48. ScerTF
49. SCPD
50. SEA
51. SELEXdb
52. SKY/M-FISH and CGH
53. SNP2TFBS
54. STIFDB2
55. SwissRegulon
56. TcoF-DB
57. Telomerase database
58. TESS
59. TFBSbank
60. TFBSshape
61. TFClass
62. TFClass
63. TiProD
64. TRACTOR db
65. TRANSCompel®
66. TRANSFAC®
67. TransmiR
68. TRANSPATH®
69. TRED
70. TrSDB
71. TTSMI
72. UniPROBE
73. VISTA Enhancer Browser
74. WebGeSTer DB
75. YEASTRACT
76. YEASTRACT
77. YeTFaSCO

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

EPD

Eukaryotic Promoter Database

Geneva, Switzerland
(Швейцария)

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

Check out our short survey [Follow @EPD_SIB](#)

Access EPDnew

H. sapiens
H. sapiens non-coding
M. mulatta
M. musculus
M. musculus non-coding
R. norvegicus
C. familiaris
G. gallus
D. melanogaster
A. mellifera
D. rerio
C. elegans
A. thaliana
Z. mays
S. cerevisiae
S. pombe
P. falciparum

Select / Download
Promoter analysis tools
FTP site

Access EPD

Promoter elements
Select / Download
FTP site

Access MGA Database

MGA-Search
MGA Data Overview
MGA FTP site

Documents

Other Resources

News: 2019-04-26 -- First EPDnew release of non-coding RNA promoter collection for *H. sapiens* [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*: 29598 promoters,
 - Macaca mulatta*: 9575 promoters,
 - Mus musculus*: 25111 promoters,
 - Rattus norvegicus*: 12601 promoters,
 - Gallus gallus*: 6127 promoters,
 - Canis familiaris*: 7545 promoters,
 - Drosophila melanogaster*: 16972 promoters,
 - Apis mellifera*: 6493 promoters,
 - Danio rerio*: 10728 promoters,
 - Caenorhabditis elegans*: 7120 promoters;
- Plants:
 - Arabidopsis thaliana*: 22703 promoters;
 - Zea mays*: 17081 promoters;
- Fungi:
 - Saccharomyces cerevisiae*: 5117 promoters,
 - Schizosaccharomyces pombe*: 4802 promoters;
- Invertebrates:
 - Plasmodium falciparum*: 5597 promoters.

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**.

Latest research from us that uses EPD:
[Influence of Rotational Nucleosome Positioning on Transcription Start Site Selection in Animal Promoters](#), Dreos, R., Ambrosini, G., Bucher, P. PLOS Comp Biol (2016) doi: 10.1371/journal.pcbi.1005144; PMID: [27716823](#)

Please cite us using the following references:

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 interface. At the top, there is a logo for SIB (Russian Academy of Sciences) and the EPD logo, which includes a rabbit and a test tube. Below the logo, the text reads "EPD EUKARYOTIC PROMOTER DATABASE". The navigation bar includes "Computational Cancer Genomics | ExPASy | EPFL". A search bar is present with a dropdown menu set to "All databases" and a "SEARCH" button. The main content area is divided into three sections: "Access EPDnew", "Access EPD", and "Access MGA data". The "Access EPDnew" section lists species: *H. sapiens*, *M. musculus*, *D. melanogaster*, *D. rerio*, and *C. elegans*, along with links for "Standard search", "Select / Download", "Promoter analysis tools", and "FTP site". The "Access EPD" section lists "Promoter elements", "SRS access to EPD", "Select / Download", and "FTP site". The "Access MGA data" section lists "MGA Data Overview" and "MGA FTP site". The main content area features a section titled "EPD database" with a description: "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**." Below this, it states "Current version is based on EMBL 124." and "Collection accessibility" section, which describes four ways to access the database: (1) using the input form, (2) using the download page, (3) using the ftp website, and (4) using the ftp website for bulk download. A "Reference" section provides a link to a detailed description of the principles governing EPD and a citation: "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.

Первая публикация о базе EPD в журнале NAR 1986 год !!!!!

Volume 14 Number 24 1986

Nucleic Acids Research

Compilation and analysis of eukaryotic POL II promoter sequences

Philipp Bucher and Edward N. Trifonov

Department of Polymer Research, The Weizmann Institute of Science, Rehovot, PO Box 26, Israel

Received 20 August 1986; Revised and Accepted 30 October 1986

ABSTRACT

A representative set of 168 eukaryotic POL II promoters has been compiled from the EMBL library and subjected to computer signal search analysis. Application of this technique to *E. coli* promoters as a control ensemble revealed the well known consensus sequences at -35 and -10 which indicates that the methods are adequate to approach problems of this kind. The results obtained from the eukaryotic promoter set can be summarized as follows: (i) Common sequence features are confined to a region between -50 and +10 relative to the transcriptional initiation site. (ii) The only well conserved consensus sequence is TATAAA, centered at -28. (iii) A weak motif, CA followed preferentially by pyrimidines, surrounds the cap-site. (iv) Two pentanucleotides which have been shown by experiments to stimulate transcription of certain genes, GGGCG and CCAAT, are moderately over-represented in the upstream region (between -129 and -50). However, they occur at highly variable distances from the initiation site.

INTRODUCTION

Eukaryotic POL II promoters have been the subject of intense investigation during the

Bucher P, Trifonov EN. Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 1986 Dec 22;14(24):10009-26.

Представление базы EPD в журнале NAR-1986

Nucleic Acids Research

Gene and organism	-40	-30	-20	-10	0	+10
Wheat H3	TCTCGGTGCTCCTCTATTTAACTCCGGCCCGTCCCCTTCTTCCTCTCAAGCCAAATCTC					
Wheat H4	CAACCTCTCGACCCCTTAAAGACGCCCTTCGCCCCACCCAGCAAAATCAAGACACAGAAG					
Maize zein ZA1	AATATTTGAGACCTCAGCTATATAAATAGCTCCGATATCATAGTAGTAACTCCATCCCCAT					
Maize zein 19K	CACAAGGACTGAGATGTGTATAAATATCTTAGATTAGCTAGCTAAATATATCGGCACATA					
Soybean RuBCC SS	ACACAAATCGACACTATTATATATAGCAAGTTTGAGCAGAAGCTGGGTAATCTGGCAGCA					
Soybean Lb I	CTCTTCAAGCCTCTATATAAATAAGATTTGGATGTGAAGTTGTGCAAAATTTGCATTG					
Soybean hb8871	TATATGTCCCTCTACATCAATTTAAATACCCCATGTGCTCTTGAAGACAGATCAACA					
Soybean Le1	AAGTACCAATAATGCTAGTATAAATAGGGGATGACTCCCCATGCACTACAGTGCATTT					
F. v. phaesolin	CTCTCTATATAATACCTATAAATACCTCTAATATCACTCACTTCTTTCATCATCCATCC					
A. t. TDNaO tmr P1	AATGAATTTCAAGGAGACAATATAACCCGCTCTGATAACACAATCTCTAAATATAAAAAAT					
A. t. TDNaO tmr P2	CTGATAACACAATTTCTCTAAATAAAAAATCAGTTGTATTCAATATATCTGCAAAAAATTT					
A. t. TDNaO ocs	TTGCCCATTCATTGATCTATTTAAAGGTGGCCCTCAAGGATAATCGCCCAAAACCATATA					
A. t. TDNaO n	CAAAAAATGCTCCACTGACGTTCCATAAAATCCCTCGGTATCCAATTAGAGTCTCATATT					
A. t. TDNaO tr-7	CGTCCAGCCCGGCATCTATATATAGCGCCAATAGTTTTGTCTTACCAAAAACACACTTC					
CAMV 8s, 35s-major	TTCCGCAAGCCCTCTCTATATAAAGGAAGTTCATTTGATTTGGAGAGGACACGCTGAAA					
CAMV 35s-minor	TAATGACCCCTTATCGATTAAAGAAATAATCCGATAAGCCCGCCCTTAAAAAATTTGGT					
D. m. cut.-protein I	ACTTTGGGTGGCAATCATATAAAAAAGGCTCGCCGACCAATCAAGTATATCGTCAACG					
D. m. cut.-protein I II	TGCATCAGCTTTTGTGATATAAACCAGGATTGAGCATAGATTGTCATCAGTCTTAG					
D. m. sgs4 glue	CGATGGCAAAATCGAGTGGGTATATAAAGAGCATCAAGCGGTATTTGAATTCCTCAAAATCAA					
D. m. 3L 74F	TATGTAACTATAGATCTATAATAAACAAGAAACAACAACTAGTTGTAAGCAAAACAC					
D. m. globin IV	TTCTCAAAATTTTAAAGTATAAATGGAGCACAAAATTCGATAGTAACTCAGTCTCTCAAT					
D. m. YP I	CGCTCAGCGTAAATTTGGTATATAAACCACATCGTTGGATTTGGAAGGCGAGTTCAAC					
D. m. YP II	ATAGACTACCGATTCCAAGGGGTATAAATGCAATGAGTCGGACGATGGGCGATCGAGTA					
D. m. ADH larval	TGCTGTACGGTCTTCTATAAATACGGGGCGACAGAACTGGGAAACCAACCACTAACG					
D. m. ADH adult	CCCCCAGGAGAGACAGTATTTAAGGAGCTGCGAAGTCCAAAGTCAACGATTTATGTCCTC					
D. m. hsp 70K	CGAAAAGAGCCCGGAGTATAAATAGAGCGGCTTCGTCGACGGAGGCTCAATCAATTTCA					
D. m. hsp 22K	TTCTCTCTGTCAAGAGTATAAATAGCCACCGGTTGGACACTACCGCTCTAGTTCAAAAA					
D. m. hsp 23K	TTCCAGCAGCAAGCGGTTGTATAAATATCCGGGCTCTCGTCAACCGGCTCAGTTGAAT					
D. m. hsp 26K	AGAAAAGCTCCAGCGGTTATAAAGCAGCGTCCGTTGACGAACAGAGCAGATCGAATT					
D. m. hsp 27K	TGTGAGCCCGAGCTCAGTATAAAGCCCGGCTCAACGTCGCCCGAGCAGCTAAACTG					
D. m. hsp 68K	TCCCTCCCGGCGACAGAGTATAAATACGGGCGCAAAATTTCCAGAGCGCTACATTTGAAA					
D. m. hsp 83K	TTCCGGTCCGGGTTTTCTATAAAGCAGACCGCGCGGTTTCGGGTTCCGAGTCTTGAA					
D. m. 44D gene H	CACCTTATCGACTAGTATAAAGGCACTGTCAGCTCTCCAGCCCGAAACAAATCGATCAA					
D. m. 44D gene L	CAATGGGAGCGGTATGCTTAAATAGGGGACACCTTTAATCCCTCGGCAATGGCAATCG					
D. m. rp49	TATTTCCAGTGGGTCACTGACCAATAGGCTACACTTGTGTGTCTCCACAGCTTCAAGAT					
B. m. fibroin	AAAACCTCGAAAATTTTTCAGTATAAAGAGGTTCAACTTTTTCAAAATGAGCATCAGTTCCGT					
B. m. Hc-A. 13	GGTGAACATGATTTCTAGTACTATATAAAGCAAGAGTCTTAAGCTTAAAGTATCAAGA					
B. m. Hc-B. 13	ATTTTCAAGAAAATCTGCTCGGTATAAAGAGCTGATGTAGTTTCAGAGTTAACTGATCTGAA					
P. m. early H1	CCACGTCAGCAACCCCGGGGATATAGGTGAGGTTGCGGTGAGGGCCGCTCACTTGTTTTG					
P. m. early H2A	TCCGATCCCGACGTTTGGTATAAATAGCCAGCAAAAAGATAGGTGGTCAACATTTCAAG					
S. p. early H2B	ACGGATCCCGCCCGTGTATAAAGAGAAAGGTTCTCGCTGGCCATTCAGAGTATCCAAA					
S. p. early H3	CCAGGATCCCGCAGCAGATATAAATAGCTGAAAATTTGCCAGTGGTCTCACTTCGCCCGT					
S. p. early H4	CAAGTCCGCAATGGTGTAAACAATACCTGGTGAATCCGGTTGAGGCATCAATCGCTTAGC					
L. p. late H3	CGAGAAGCAGTCTCGAGGTATAAATACGTCGGGTTACTTTGAAAAATTTACAGTTGACT					
L. p. late H4	TAAAGGCTATATATACCCGACCAAGCAGGAGAAATGAGTATCAAGTTGAAATCTCAAAACGG					

10012

Nucleic Acids Research

Exp. def.	Expression/Regulation	References for initiation site	EMBL Sequence Ref.
3	proliferating tissues	MGG196:397	TAHI02 1+ 186
3	proliferating tissues	NAR11:5885	TAHI01 1+ 669
3	endosperm	EMBOJ1:1589	ZMZE05 1+ 148
4	endosperm	Cell129:1015	ZMZE01 1+ 888
3	leaves, +light	JMAG1:483	GMRUBP 1+ 241
3	root nodules	PNAS79:4055	GMGL04 1+ 144
3	root e.g., +heatschock	EMBOJ3:2491	GMHSP2 1+ 492
4	cotyledon	Cell134:1023	GMLEA 1+ 942
4,6	cotyledon	PNAS80:1897	PVPHASL 1+ 101
3	plant tumor	NAR11:6211, JMAG2:354	ATACH5 1+ 8729
3	plant tumor	NAR11:6211, JMAG2:354	ATACH5 1+ 8760
3	plant tumor	JMAG1:499	ATACH5 1- 13658
3	plant tumor	NAR11:369, JMAG1:561	ATNOPA 1+ 550
3	plant tumor	EMBOJ2:419	ATACH5 1- 3303
3,8	infected leaves	Cell130:763	CAMV62 0+ 7435
3,8	infected leaves	Cell130:763	CAMV62 0+ 8017
6	third instar larva	Cell129:1027	DMCUT1 1- 760
6	third instar larva	Cell129:1027	DMCUT2 1+ 2606
3,7	larva; salivary glands	Cell129:1041, Cell134:74	DMSG54 1+ 52
4	larva; salivary glands	EMBOJ3:289	DM74EF 1+ 401
4	larva; fat body	Nature310:795	CTGL01 1+ 260
3,7	puppa; ovary, fat body	NAR10:2261	DMYOLK1 1- 225
3,7	puppa; ovary, fat body	NAR10:2261	DMYOLK1 1+ 1447
4,5	larva; fat body, gut	Cell133:125	DMADH1 1+ 974
4,5	adult	Cell133:125	DMADH1 1+ 267
4,5	+heatschock	NAR8:3105, Cell121:669, EMBOJ1:1583	DMHSP1 1+ 717
4,8	+heatschock	NAR9:1627	DMHS08 1+ 514
4,8	+heatschock	NAR9:1627	DMHS09 1+ 320
3,8	+heatschock	NAR9:1627, PNAS78:3775	DMHS10 1+ 470
4,8	+heatschock	NAR9:1627	DMHS11 1+ 290
3	+heatschock	PNAS78:3775	DMHSP68 1+ 158
3	+heatschock	NAR11:7011, PNAS78:3775	DMHS83 1+ 878
3	larva, adult	JMB166:101	DMCUT3 1- 3169
3	larva, adult	JMB166:101	DMCUT3 1- 9158
3	housekeeping gene	NAR12:5495	DMRP49 1+ 411
1,3,6	larva; silk gland	Cell116:425, Cell118:591	BMFIBR 1+ 551
(3 or 4)	eggshell, late	PNAS81:4452, JME20:265	BMCH01 1- 248
(3 or 4)	eggshell, late	PNAS81:4452, JME20:265	BMCH01 1+ 514
3	early blastula	Nature285:147, Nature288:100	PMHIS7 0+ 4860
3	early blastula	Nature285:147, Nature288:100	PMHIS7 0+ 3614
5	early blastula	Nature279:737, PNAS77:1265	SPHIS1 1+ 170
5	early blastula	PNAS77:1265	SPHIS1 1+ 1341
1,5	early blastula	Bioch20:1216, PNAS77:1265	SPHIS4 1+ 165
3	late blastula	Cell131:383, PNAS81:2411	LPHISL34 1+ 1487
3	late blastula	Cell131:383, PNAS81:2411	LPHISL34 1- 724

10013

Компиляция 168 промоторов генов эукариот, транскрибируемых РНК-полимеразой II (показаны первые две журнальные страницы)

Bucher P, Trifonov EN. Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 1986;14(24):10009-26.

Пример карточки из EPD

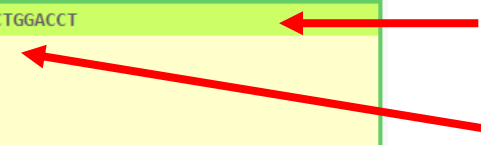
General information about the entry	
Entry name	HS_AK2
Entry type	standard
Promoter type	region
Accession number	EP74037
Description of the gene	Adenylate kinase 2.
Creation date	10-JAN-2003 (Rel. 73)
Last annotation	06-APR-2006 (Rel. 86)
Taxonomic division	VRT
Organism	Homo sapiens (human)
Keywords	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 DDBJ
SWISSPROT	P54819 KAD2_HUMAN
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	tgtgagcggcagtgaggacgtgcgtggcgtgcgtgacctgggaaGCACTGGACCT
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

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)

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

```
>EP33026 (-) Bt TP2; range -200 to 50.  
AGCTCCACCCGACCTGAGGGGCTGCTCTCAACCCACAGACACGCCCTTTGAAAGCTGCC  
CACCTGCGTGTTAGGATGAGGGCAGAGGGCTTTGTTCCCGTTGGGCCACATCTGTTACA  
TACCCCTGTGGCCAGTGCCATCACAATCGGGCCAACTATATAACCAGGGGGCTGCCAG  
GGCCTCTGTGAAGCTGGGTCTGCCAGAAGAGGAGGAGGAGGC GGCGGCCCTGCCCTCT  
AAGCGAGGCCG  
>EP28006 (+) Bt protamine P1^1.1; range -200 to 50.  
GGCCCCACCCACACACATCACAGCCCCACCCCTGCACATCACAGCCCCGCCCTCCCTC  
ACCAAGCACCTCCCACATGCCATATATGGGCATGATTTGGGCAGCTCTGACCCTGGTCT  
GTGAGGTCTGGGTCTCTGTGACCTCACAATGACCAGGACCCTGCCCGGGTCTATATAAGA  
GGCCGGGAAGTCGGCCCCTGTCACAGCCACAAATCCACCTGCTCACAGTTGGCTGGC  
TCAACCAAGGC  
>EP15026 (+) Bt cytokeratin Ia; range -200 to 50.  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNCATGGTGGACGGCAAGTTATTCAAAGGAGAGCGGACC GGAGGTTGGCGGAA  
ACGCGGAGACCTTCTGAAGCTCTGGGCCAGAGGTGGCGCTTATATAGGGCTGGGAGCTTG  
GCTGGCTGGCG  
>EP15027 (+) Bt cytokeratin Ib; range -200 to 50.  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCATGG  
TGAGTGAAGTCGGTGAAGGGAAAGTGCAAGAGTTAAGCAGGGCCACTCAGAGCCAGAGGA  
AGAGAGAAGCAAACCTGAAGACCTGTGCAGGATAAATCCCTTATATACATCTAGGAGGCT  
GCTGGGCTCAA  
>EP15028 (+) Bt cytokeratin IV*; range -200 to 50.  
CAAGGCTAGAAGCCAGAAGAATTTCTCCATGACTAAAGGAAACCAAAGAAGCAATATTCA  
TACTTCATACCTTTCTAGAGGCAGGGGGTGATCTCACTATTTGTAAAGCCAGCCCTTTC  
TAATCTGCAGGCTCACCTCCAGGACTGAGCCCGGCCATTTTTCCATATATAAGCTGC  
TGCCGGGCCGCCCTCTATAGATCTGTTCTTAGCTCTGCTTCCACCTCTCACACCCTTC  
TCAACCTATTC  
>EP11000 (+) Bt cytokeratin VTh; range -200 to 50.
```



1-ая строка - общая информация о последовательности
2-ая строка - последовательность ДНК

EPDnew - вторая часть объединенного ресурса

Информационное
содержание:
15 видов
организмов

Computational Cancer Genomics | ExPASy | EPFL

Check out our short survey [Follow @EPDnew](#)

Access EPDnew

H. sapiens
H. sapiens non-coding
M. mulatta
M. musculus
M. musculus non-coding
R. norvegicus
C. familiaris
G. gallus
D. melanogaster
A. mellifera
D. rerio
C. elegans
A. thaliana
Z. mays
S. cerevisiae
S. pombe
P. falciparum

Select / Download
Promoter analysis tools
FTP site

Access EPD

Promoter elements
Select / Download
FTP site

Access MGA Database
MGA-Search

News: 2019-04-26 -- First EPDnew release of non-coding RNA promoter collection for *H. sapiens* [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*: 29598 promoters,
 - Macaca mulatta*: 9575 promoters,
 - Mus musculus*: 25111 promoters,
 - Rattus norvegicus*: 12601 promoters,
 - Gallus gallus*: 6127 promoters,
 - Canis familiaris*: 7545 promoters,
 - Drosophila melanogaster*: 16972 promoters,
 - Apis mellifera*: 6493 promoters,
 - Danio rerio*: 10728 promoters,
 - Caenorhabditis elegans*: 7120 promoters;
- Plants:
 - Arabidopsis thaliana*: 22703 promoters;
 - Zea mays*: 17081 promoters;
- Fungi:
 - Saccharomyces cerevisiae*: 5117 promoters,
 - Schizosaccharomyces pombe*: 4802 promoters;
- Invertebrates:
 - Plasmodium falciparum*: 5597 promoters.

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**.

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 (2014)

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}, Rouayda Cavin Périet² 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 – информационное содержание (2017)

Table 1.

Current contents of EPDnew

Organism, version	Promoters, genes ^a	TSS libraries ^b	Chromatin data MNase–Dnase ^c	ChIP-seq samples histones–PIC–TFs ^d
<i>H. sapiens</i> (4)	25 503, 17 785 (95%)	1088	23–998	2231–491–3794
<i>M. musculus</i> (2)	21 239, 17 565 (90%)	339	4–0	174–60–384
<i>D. melanogaster</i> (2)	15 073, 12 603 (92%)	57	6–23	29–12–189
<i>D. rerio</i> (1)	10 728, 10 235 (43%)	12	4–4	12–3–1
<i>C. elegans</i> (1)	7120, 6363 (32%)	8	6–6	2–1–3
<i>A. mellifera</i> (1)	6493, 5712 (53%)	16	0–0	0–0
<i>A. thaliana</i> (1)	10 229, 10 177 (37%)	1	0–0	0–0–32
<i>Z. mays</i> (1)	17 081, 15 828 (59%)	8	0–0	8–0–0
<i>S. cerevisiae</i> (2)	5117, 5110 (88%)	19	1–27	0–8–17
<i>S. pombe</i> (1)	3440, 3438 (67%)	1	8–8	6–0–51

^aIn parenthesis is indicated the percentage of genes coverage.

^bCAGE, GRO-cap and TSS-seq samples used to build the relative database.

^cMNase-seq and DNase-seq samples that are present in the MGA repository.

^dChIP-seq samples for histone marks and variants (such as H3K4me3, H2A.Z, H3), components of the PIC (such as Pol-II, TFIID, TFIIB, TBP, etc.) and Transcription Factors that are present in the MGA repository.

Источники информации и процедура получения данных о стартах транскрипции в базе HS_EPDnew

TSS assembly pipeline for Hs_EPDnew_006

Introduction

This document provides a technical description of the transcription start site assembly pipeline that was used to generate EPDnew version 006 for *H. sapiens*.

Source Data

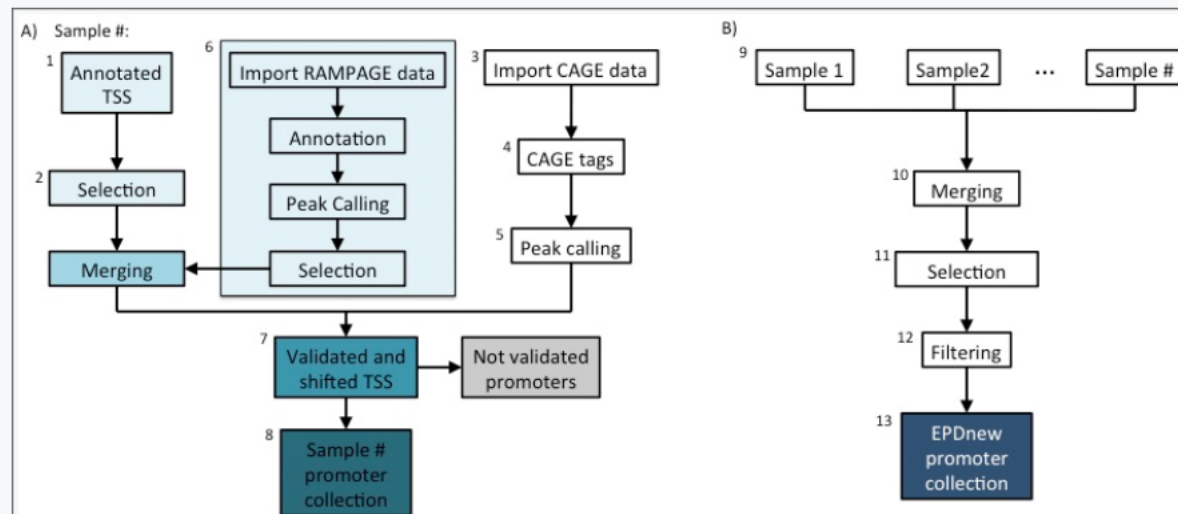
Promoter collection:

Name	Genome Assembly	Promoters	Genes	PMID	Access data
Gencode	Dec 2013 GRCh38/hg38	35320	17056	27250503	SOURCE DOC DATA

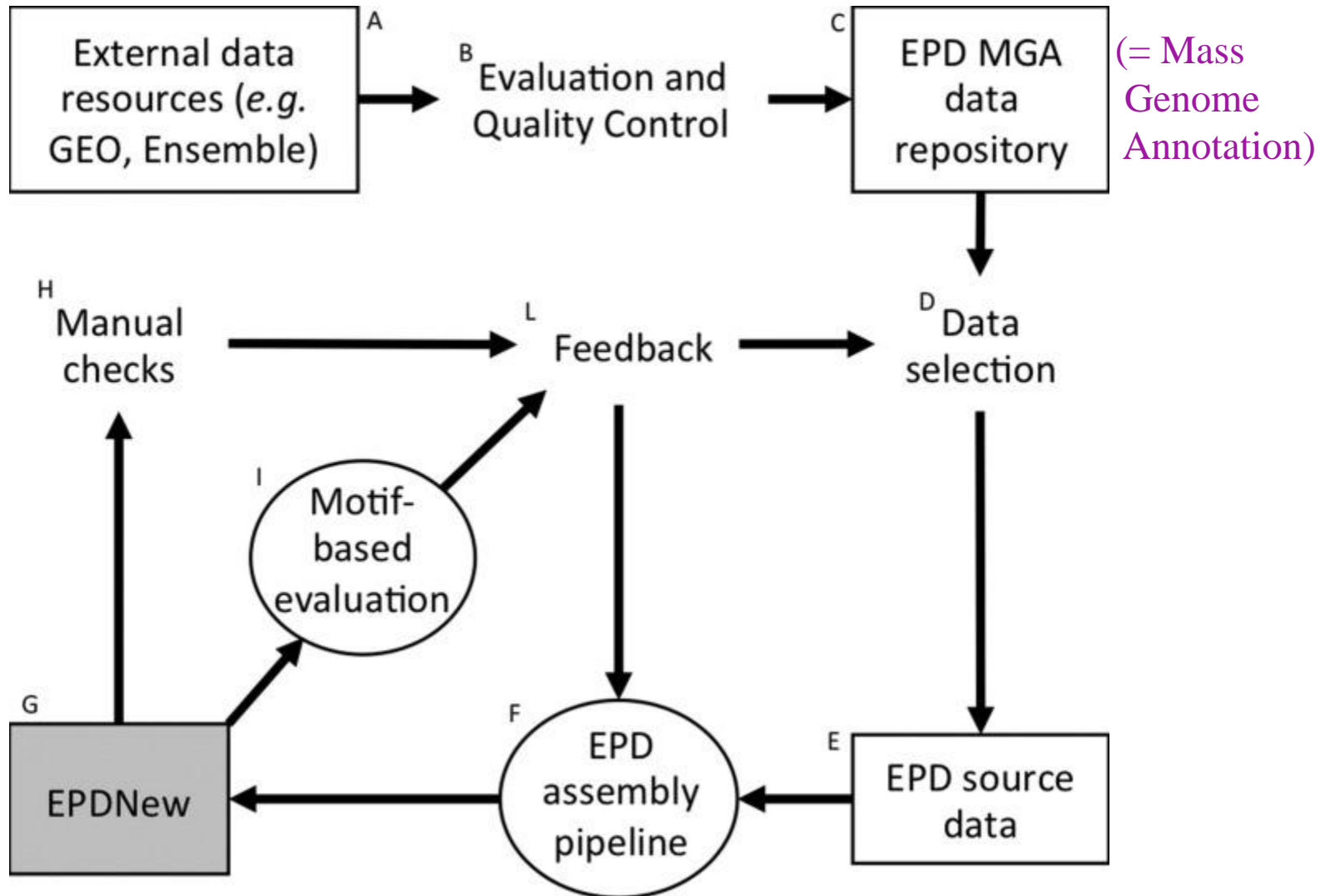
Experimental data:

Name	Type	Samples	Tags	PMID	Access data
FANTOM5	CAGE	941	18,244,201,540	24670764	SOURCE DOC DATA
ENCODE	CAGE	145	7,134,200,060	22955620	SOURCE DOC DATA
ENCODE	RAMPAGE	225	13,540,041,874	22936248	SOURCE DOC DATA

Assembly pipeline overview



Процедура получения данных о стартах транскрипции в базе EPDnew



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

Общая информация о гене

Последовательность ДНК - 50 / +10

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

Характеристики генома в окрестностях старта транскрипции гена MAPK1

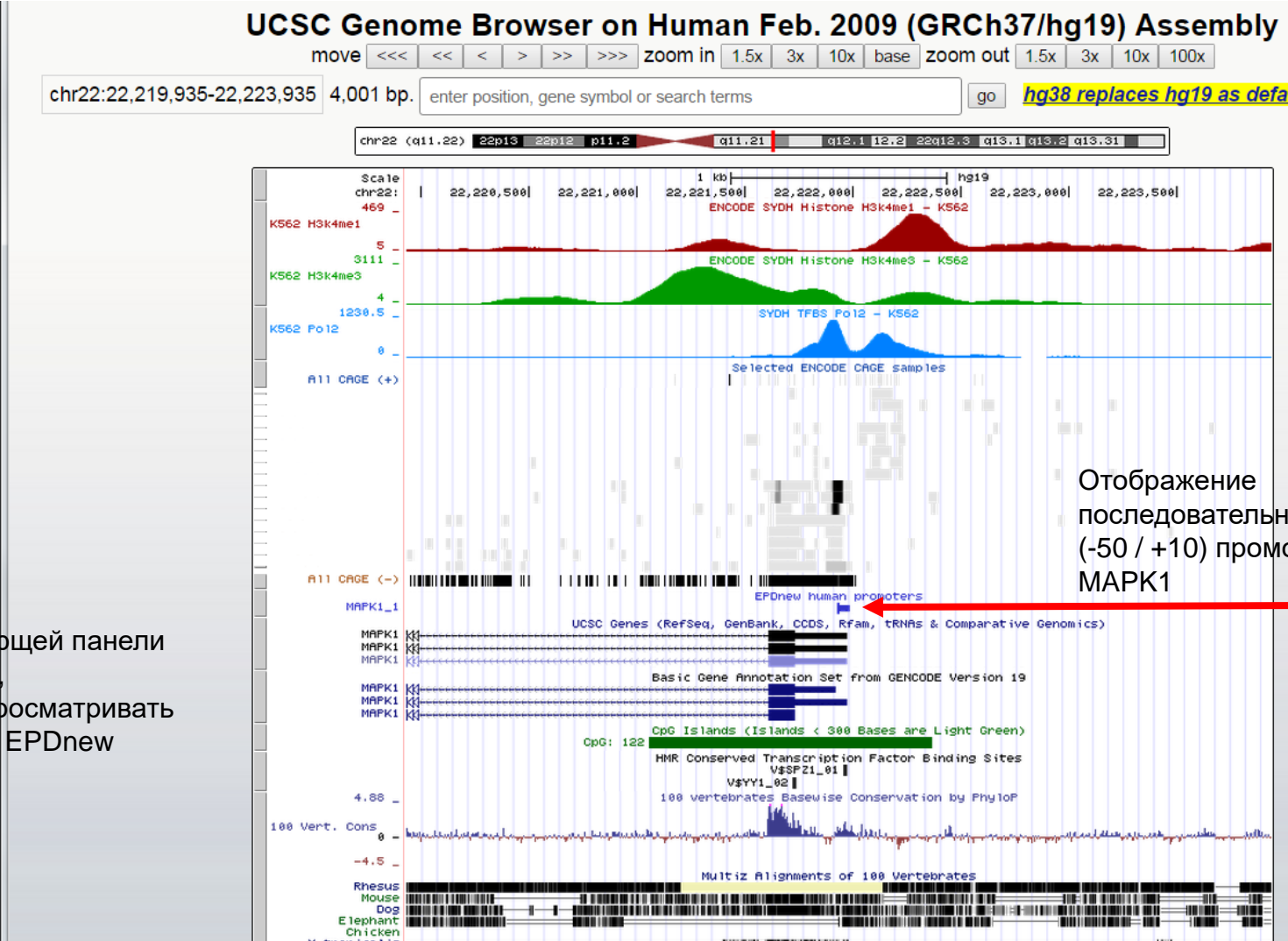
Отображение последовательности ДНК -50 / +10

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: cctctcgtggcgcgcgccctcctcgcgccgcgccgcgcctcAGTCTGGCAGG
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: 1 kb hg19
chr22: 22,220,500 | 22,221,000 | 22,221,500 | 22,222,000 | 22,222,500 | 22,223,000 | 22,223,500
K562 H3k4me1
K562 H3k4me3
K562 Po12
A11 CAGE (+)
A11 CAGE (-)

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

EPD viewer hub refresh

Histone Marks K562 Pol2 All CAGE (+) CAGE All CAGE (-) EPDnew

full full full full full full

move end < 2.0 >

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 ChIP-seq 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.
- [QProf](#): 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: [QProf](#), [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

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

AND average expression of at least tags

AND expressed in at least samples

Database:

Database: human_epdnew

Assembly: hg38

Selection Parameters

TATA-box: with

Initiator: with

CCAAT-box: with

GC-box: all

Marked as: single

Average expression: 5

Expressed in: 3

Results: 15 promoters selected

[SGA file](#)

[FPS file](#)

[BED file](#)

LiftOver options

Sequence Extraction Tool (FASTA format)

From: To:

Downstream Analysis

Motif Enrichment ?

Motif Discovery ?

Chromatin analysis ?

Представление базы EPD в журнале NAR-1986

Nucleic Acids Research

Gene and organism	-40	-30	-20	-10	0	+10
Wheat H3	TCTCGGTGCTCCTCTATTTAACTCCGGCCCGTCCCCTTCTTCCTCCTAGCCCAATCTC					
Wheat H4	CAACCTCGACCCCTTTAAGACGCCCTTCGCCCCACCCAGCAAAATCAAGACCAAGAG					
Maize zein ZA1	AATATTTGAGACCTCAGCTATATAAATAGCTCCGATATCAGTAGTTAATCCATCCCCAT					
Maize zein 19K	CACAAGGACTGAGATGTGTATAAATATCTTAGATTAGCTAGCTAAATATATCGGCACATA					
Soybean RuBCC SS	ACACAAATCGACACTATTATATATAGCAAGTTGAGCAGAGCTGGGTAATCTGGCAGCA					
Soybean Lb I	CTCTTCAAGCCTCTATATAAATAAGATTGGATGTGAAGTTGTGGCAAAATTTGCATTG					
Soybean hb8871	TATATGTCCCTCCTACATCATTTTAAATACCCCATGTGCTCTTTGAGAGCACTACAGA					
Soybean Le1	AAGTACCCAATAATGCTAGTATAAATAGGGGCATGACTCCCCATGCATCAAGTGCAATT					
F. v. phaseolin	CTCTCTATATAATACCTATAAATACCTCTAATATCACTCACTTCTTTCATCATCCATCC					
A. t. TDNaO tmr P1	AATGAAATTTCAAGGAGACAATATAACCCGCTCTGATAACCAAAATCTCTAAATATAAAAA					
A. t. TDNaO tmr P2	CTGATAACCAAAATCTCTAAATATAAAAAATCAGTTGTATTCAATATATCGCAAAAATCT					
A. t. TDNaO ocs	TTGCCCATTCATGTATCTTTTAAAGGTGGCCCTCAAGGATAAGTCCCAAAACCATATA					
A. t. TDNaO ncs	CAAAAATGCTCCACTGACGTCCATAAAATCCCTCGGTATCCAATTAGAGTCTCATATT					
A. t. TDNaO tr-7	CGTCCCAGCCCGCATCTATATATAGCGCCAATAGTTTTGTCTTACCAAAAACACACTTC					
CANV 8s, 35s-major	TTCCGCAAGCCCTTCTCTATATAAGGAAGTTCATTTGATTTGGAGAGGACACGCTGAAA					
CANV 35s-minor	TAATGACCCCTTATCGATTATAAGAAATAATCCGCATAAGCCCGCCCTTAAAATAATGGT					
D. m. cut.-protein I	ACTTTGGGTGGCAATCATATAAAAAGGCTCGCCGACCAAAATCAAGTATATCAAGCAAG					
D. m. cut.-protein II	TGCATCAGCTTTTGTATATAAACCAGGATTTGAGCATAGATTGTCATCAGTCTTAG					
D. m. sgs4 glue	CGATGGCAAATCGAGTGGGTATATAAAGAGCATCAAGCGGTATTTGAATTCCAAATGCAA					
D. m. 3L 74F	TATGTAACTATAGATTCTATAAATAAACAAGAAACAAACATAGTTGTAAGAACAACAC					
D. m. globin IV	TTCTCAAAAATTTAAGTATAAATGGAGCACAAAATTCGATAGTAAATCAGTCTCTCAAT					
D. m. YP I	CGCTCAGCGTAAATTTGGTATATAAACCACATCGTTGGATTTGGAAAGGCGAGTTCAAC					
D. m. YP II	ATAGACTACCGATCCAAAGGGGTATAAAATGCATTTGAGTCGGACGATGGCGATCGAGTA					
D. m. ADH larval	TGCTGTACGGATCTTCTATAAATACGGGGCGACAGAACTGGGAACCAAACTAAGC					
D. m. ADH adult	CCCCCAGGAGAGACAGTATTTAAGGAGCTGCGAAGTCCAAGTCAACGATTTATGTCCTC					
D. m. hsp 70K	CGAAAAGAGCGCCGGAGTATAAATAGAGCGGCTTCGTCGACGGAGCGCTCAATCAATTC					
D. m. hsp 22K	TTCTCTCTGTCAAGAGTATAAATAGCCACGGTTGGACACTACCGCTCAGATTCAAAAA					
D. m. hsp 23K	TTCCAGCAGCAAGCGGTTGTATAAATATCCGGCATCTTCGTCGCAACGGCGTCAAGTGAAT					
D. m. hsp 26K	AGAAAAGCTCCAGCGGTATAAAGCAGCGCTCGTTGACGAACAGAGCCAGATCGAATT					
D. m. hsp 27K	TGTGAGCCCGAGCTCAGTATAAAGCCGGCTCAACGTCGCCCGAGCCAGTCTAAACTG					
D. m. hsp 66K	TCCCTCCCGGCGACAGAGTATAAATACGGGGCGAAAATTTCCACAGCGCTACATTTGAAA					
D. m. hsp 83K	TTCCGGTCCGGGTTTTCTATAAAGCAGACCGCGCGGTTTCGCGGTTCGAGTCTTGAA					
D. m. 44D gene H	CACCTTATCGACTAGTATAAAGGCACTGCAGCTCTCCAGCCCGAAACAAAATCGATCAA					
D. m. 44D gene L	CAATGGGAGCGGTATGCTTAAATAGGGCACCTTTAATCCCTCGGCAATGGCAATCG					
D. m. rp49	TATTTCCAGTGGTTCAGTGCACATTAAGGTACACTTGTGTGCTCCTACAGCTTCAAGAT					
B. m. fibroin	AAAACCTCGAAAATTTTCAGTATAAAGAGGTTCAACTTTTTCAAAATCAGCATCAGTTCCGT					
B. m. Hc-A.13	GGTGAACATGATTTCTAGTTACTATATAAAGAACAAGTCTTAAGCTTTAAGTATCAAGA					
B. m. Hc-B.13	ATTTTCAAGAAAATCTGCTCGGTATAAAGAGCTGATGTAGTTTCAGAGTTAAGTATCTGAA					
P. m. early H1	CCACGTCAGCAACCCGGGGATATAGGTGAGTTGCGGTGAGGGCCGCTCACTTGTTTTG					
P. m. early H2A	TCCGATCCCGACGTTTGGTATAAATAGCCAGCAAAAAGATAGGTGGTCAACATTTCAAG					
S. p. early H2B	ACGGATCCGGCCCGTGTATAAAGAGAAAGGTTCTCGCTGGCCATTCAGAGTATCCAAA					
S. p. early H3	CCAGGATCCCGCAGCAGTATAAATAGCTGAAAATTTCCAGTGGTCTCACTTCGCTCCCGT					
S. p. early H4	CAAGTCCGCAATGGTGTAAACAATACTCGGTGCAATCCGGTTGAGGCATCAATCGCTTAGC					
L. p. late H3	CGAAGCAGTCTCGAGGTATAAATACGTCGGCGTTCCTTTGAAAATTTAICAGTTGACT					
L. p. late H4	TAAAGGCTATATATACCCGACCAAGCAGGAGAAATGAGTATCAAGTTTGAATCTCAAAACGG					

10012

Nucleic Acids Research

Exp. def.	Expression/Regulation	References for initiation site	EMBL Sequence Ref.
3	proliferating tissues	MGG196:397	TAHI02 1+ 186
3	proliferating tissues	NAR11:5885	TAHI01 1+ 669
3	endosperm	EMBOJ1:1589	ZMZE05 1+ 148
4	endosperm	Cell129:1015	ZMZE01 1+ 888
3	leaves, +light	JMAG1:483	GMRUBP 1+ 241
3	root nodules	PNAS79:4055	GMGL04 1+ 144
3	root e.g., +heatschock	EMBOJ3:2491	GMHSP2 1+ 492
4	cotyledon	Cell134:1023	GMLEA 1+ 942
4,6	cotyledon	PNAS80:1897	PVPHASL 1+ 101
3	plant tumor	NAR11:6211, JMAG2:354	ATACH5 1+ 8729
3	plant tumor	NAR11:6211, JMAG2:354	ATACH5 1+ 8760
3	plant tumor	JMAG1:499	ATACH5 1- 13658
3	plant tumor	NAR11:369, JMAG1:561	ATNOPA 1+ 550
3	plant tumor	EMBOJ2:419	ATACH5 1- 3303
3,8	infected leaves	Cell130:763	CAMV22 0+ 7435
3,8	infected leaves	Cell130:763	CAMV22 0+ 8017
6	third instar larva	Cell129:1027	DMCUT1 1- 760
6	third instar larva	Cell129:1027	DMCUT2 1+ 2606
3,7	larva; salivary glands	Cell129:1041, Cell134:74	DMSG54 1+ 52
4	larva; salivary glands	EMBOJ3:289	DM74EF 1+ 401
4	larva; fat body	Nature310:795	CTGL01 1+ 260
3,7	puppa; ovary, fat body	NAR10:2261	DMYOLK1 1- 225
3,7	puppa; ovary, fat body	NAR10:2261	DMYOLK1 1+ 1447
4,5	larva; fat body, gut	Cell133:125	DMADH1 1+ 974
4,5	adult	Cell133:125	DMADH1 1+ 267
4,5	+heatschock	NAR8:3105, Cell121:669, EMBOJ1:1583	DMHSP1 1+ 717
4,8	+heatschock	NAR9:1627	DMHS08 1+ 514
4,8	+heatschock	NAR9:1627	DMHS09 1+ 320
3,8	+heatschock	NAR9:1627, PNAS78:3775	DMHS10 1+ 470
4,8	+heatschock	NAR9:1627	DMHS11 1+ 290
3	+heatschock	PNAS78:3775	DMHSP68 1+ 158
3	+heatschock	NAR11:7011, PNAS78:3775	DMHS83 1+ 878
3	larva, adult	JMB166:101	DMCUT3 1- 3169
3	larva, adult	JMB166:101	DMCUT3 1- 9158
3	housekeeping gene	NAR12:5495	DMRP49 1+ 411
1,3,6	larva; silk gland	Cell116:425, Cell118:591	BMFIBR 1+ 551
(3 or 4)	eggshell, late	PNAS81:4452, JME20:265	BMCH01 1- 248
(3 or 4)	eggshell, late	PNAS81:4452, JME20:265	BMCH01 1+ 514
3	early blastula	Nature285:147, Nature288:100	PMHIS7 0+ 4860
3	early blastula	Nature285:147, Nature288:100	PMHIS7 0+ 3614
3	early blastula	Nature279:737, PNAS77:1265	SPHIS1 1+ 170
5	early blastula	PNAS77:1265	SPHIS1 1+ 1341
1,5	early blastula	Bioch20:1216, PNAS77:1265	SPHIS4 1+ 165
3	late blastula	Cell131:383, PNAS81:2411	LPHISL34 1+ 1487
3	late blastula	Cell131:383, PNAS81:2411	LPHISL34 1- 724

10013

Компиляция 168 промоторов генов эукариот, транскрибируемых РНК-полимеразой II (показаны первые две журнальные страницы)

Table 1. Characterization of Constraint Regions by Over-Represented Gapped Trinucleotides.

(from Bucher P, Trifonov EN. Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 1986;14(24):10009-26.

Eukaryotic Promoters											
TATA-box region (from -35 to -16)					Cap-site region (from -9 to +10)						
-30	-25	-20	Occurrence frequency		-5	0	5	Occurrence frequency			
< ---TA-A---			70.1 % (117/167)		< --C--CA---			31.0 % (52/168)			
< ---ATA---			66.5 % (111/167)		< -CA---C-->			28.6 % (48/168)			
< --A-A-A---			63.5 % (106/167)		< ---CA-T---			28.0 % (47/168)			
< ---AT-A---			63.5 % (106/167)		< --CA-C---			27.4 % (46/168)			
< ---TAA---			61.1 % (102/167)		< --CA--T-->			25.6 % (43/168)			
< --T-A-A---			60.5 % (101/167)		< -G---CA-->			25.6 % (43/168)			
< ---AA---			58.1 % (97/167)		< ---A-CA---			24.4 % (41/168)			
< --T-AA---			56.3 % (94/167)		< ---TCA---			24.4 % (41/168)			
< ---TA-A---			55.7 % (93/167)		< ---CAG---			23.8 % (40/168)			
< ---T-TA---			55.4 % (92/168)		< ---CA-C---			23.2 % (39/168)			
< ---AT-A---			53.3 % (89/167)		< -C---CA-->			22.6 % (38/168)			
< ---TA-A---			53.3 % (89/167)		< ---CAC---			22.6 % (38/168)			
< ---TAT---			53.0 % (88/166)		< --CA--C-->			22.6 % (38/168)			
< ---T-AA---			51.5 % (86/167)		< -CA---A-->			22.0 % (37/168)			
< ---AAA---			50.9 % (85/167)		< G-----CA-			21.4 % (36/168)			
< --T-T-A---			50.9 % (85/167)		< A-----CA-->			20.8 % (35/168)			
< ---A-AA---			49.1 % (82/167)		< --G---CA-->			20.8 % (35/168)			
< ---A-AAA---			47.9 % (80/167)		< -G---CA-->			20.8 % (35/168)			
< --T-T-A---			47.3 % (79/167)		< --A---CA-->			20.8 % (35/168)			
< -A-A---G-->			46.1 % (77/167)		< --CA--T---			20.8 % (35/168)			
< -A-A---G-			46.1 % (77/167)								
TATAAA					CA-yyy						
Prokaryotic Promoters											
-35 region (from -45 to -26)						-10 region (from -19 to 0)					
-40	-35	-30	Occurrence frequency		-15	-10	-5	Occurrence frequency			
< ---TTG---			52.7 % (43/81)		< --TA---T-->			71.4 % (80/112)			
< ---T-GA---			42.7 % (58/110)		< --A-A-T---			56.3 % (63/112)			
< ---TT-A---			40.0 % (47/110)		< --A--AT---			50.9 % (57/112)			
< ---TT-C---			39.1 % (44/110)		< ---TA-A---			46.4 % (52/112)			
< ---T-TT---			38.5 % (43/109)		< --T--A-T---			46.4 % (52/112)			
< ---TGA---			37.3 % (42/110)		< ---T-AT---			44.6 % (50/112)			
< --T-AC---			35.5 % (41/110)		< ---AT-T---			42.9 % (48/112)			
< TT-----T>			34.6 % (39/110)		< ---AAT---			39.3 % (44/112)			
< ---TG-C---			34.6 % (38/110)		< -G--A--T-->			39.3 % (44/112)			
< --T-G-C---			34.6 % (38/110)		< ---TAA---			39.3 % (44/112)			
< ---TT-A---			34.6 % (38/110)		< ---TA-A---			38.4 % (43/112)			
< --T-T-A---			34.6 % (38/110)		< T--A--T-->			38.4 % (43/112)			
< -A-----TT-			33.9 % (38/109)		< TA-----C>			37.5 % (42/112)			
< -A-----TT-->			33.9 % (37/109)		< T-----A-T-->			37.5 % (42/112)			
< --T-T-G---			33.7 % (37/110)		< ---TA-T---			36.6 % (41/112)			
< --A---TT---			32.1 % (36/109)		< A---T---C-->			35.7 % (40/112)			
< ---TG-A---			31.8 % (35/110)		< --A--TA---			35.7 % (40/112)			
< ---T-AC---			31.8 % (35/110)		< ---T-AT---			35.7 % (40/112)			
< --T-TG---			31.8 % (35/110)		< --A---TA-->			35.7 % (40/112)			
					< ---T-AT---			35.7 % (40/112)			
t-TTGACa					TATAAT---c						

Результаты анализа данных из первого релиза базы EPD (Eukaryotic Promoter Database)

Публикации по базе EPD в журнале NAR

NCBI Resources How To Sign in to NCBI

PubMed ("Nucleic Acids Res"[Journal]) AND (EPD[Title] OR The eukaryotic promoter) Search

Create RSS Create alert Advanced Help

Первая публикация о базе EPD в журнале NAR -1986

Volume 14 Number 24 1986

Nucleic Acids Research

Compilation and analysis of eukaryotic POL II promoter sequences

Philipp Bucher and Edward N. Trifonov

Department of Polymer Research, The Weizmann Institute of Science, Rehovot, PO Box 26, Israel

Received 20 August 1986; Revised and Accepted 30 October 1986

ABSTRACT

A representative set of 168 eukaryotic POL II promoters has been compiled from the EMBL library and subjected to computer signal search analysis. Application of this technique to *E. coli* promoters as a control ensemble revealed the well known consensus sequences at -35 and -10 which indicates that the methods are adequate to approach problems of this kind. The results obtained from the eukaryotic promoter set can be summarized as follows: (i) Common sequence features are confined to a region between -50 and +10 relative to the transcriptional initiation site. (ii) The only well conserved consensus sequence is TATAAAA, centered at -28. (iii) A weak motif, CA followed preferentially by pyrimidines, surrounds the cap-site. (iv) Two pentanucleotides which have been shown by experiments to stimulate transcription of certain genes, GGGCG and CCAAT, are moderately over-represented in the upstream region (between -129 and -50). However, they occur at highly variable distances from the initiation site.

INTRODUCTION

Eukaryotic POL II promoters have been the subject of intense investigation during the



- [The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms.](#)
Dreos R, Ambrosini G, Groux R, Cavin Périer R, Bucher P.
Nucleic Acids Res. 2017 Jan 4;45(D1):D51-D55. doi: 10.1093/nar/gkw1069. Epub 2016 Nov 28.
PMID: 27899657 [Free PMC Article](#)
[Similar articles](#)
- [The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools.](#)
Dreos R, Ambrosini G, Périer RC, Bucher P.
Nucleic Acids Res. 2015 Jan;43(Database issue):D92-6. doi: 10.1093/nar/gku1111. Epub 2014 Nov 6.
PMID: 25378343 [Free PMC Article](#)
[Similar articles](#)
- [EPD and EPDnew: high-quality promoter resources in the next-generation sequencing era.](#)
Dreos R, Ambrosini G, Cavin Périer R, Bucher P.
Nucleic Acids Res. 2013 Jan;41(Database issue):D157-64. doi: 10.1093/nar/gks1233. Epub 2012 Nov 27.
PMID: 23193273 [Free PMC Article](#)
[Similar articles](#)
- [EPD in its twentieth year: towards complete promoter coverage of selected model organisms.](#)
Schmid CD, Perier R, Praz V, Bucher P.
Nucleic Acids Res. 2006 Jan 1;34(Database issue):D82-5.
PMID: 16381980 [Free PMC Article](#)
[Similar articles](#)
- [The Eukaryotic Promoter Database EPD: the impact of in silico primer extension.](#)
Schmid CD, Praz V, Delorenzi M, Périer R, Bucher P.
Nucleic Acids Res. 2004 Jan 1;32(Database issue):D82-5.
PMID: 14681364 [Free PMC Article](#)
[Similar articles](#)
- [The Eukaryotic Promoter Database EPD: new entry types and links to gene expression data.](#)
Praz V, Périer R, Bonnard C, Bucher P.
Nucleic Acids Res. 2002 Jan 1;30(1):322-4.
PMID: 11752326 [Free PMC Article](#)
[Similar articles](#)
- [The eukaryotic promoter database \(EPD\).](#)
Périer RC, Praz V, Junier T, Bonnard C, Bucher P.
Nucleic Acids Res. 2000 Jan 1;28(1):302-3.
PMID: 10592254 [Free PMC Article](#)
[Similar articles](#)
- [The Eukaryotic Promoter Database \(EPD\): recent developments.](#)
Périer RC, Junier T, Bonnard C, Bucher P.
Nucleic Acids Res. 1999 Jan 1;27(1):307-9.
PMID: 9847211 [Free PMC Article](#)
[Similar articles](#)
- [The Eukaryotic Promoter Database EPD.](#)
Cavin Périer R, Junier T, Bucher P.
Nucleic Acids Res. 1998 Jan 1;26(1):353-7.
PMID: 9399872 [Free PMC Article](#)
[Similar articles](#)

Конец 4-ой лекции