

# 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.**

# ОСНОВНЫЕ КЛАССЫ БЕЛКОВ, УЧАСТВУЮЩИХ В РЕГУЛЯЦИИ ТРАНСКРИПЦИИ ГЕНОВ ЭУКАРИОТ

## РНК-полимеразы

Синтез молекулы РНК, комплиментарной и антипараллельной матричной цепи ДНК.

РНК-полимеразы не могут функционировать сами по себе

## Базальные (общие) транскрипционные факторы

ДНК-связывающий домен **имеется** у белка ТВР (компонент TFIID)

Формируют ПИК, обеспечивают точную посадку РНК-полимеразы на участки ДНК, прилежащие к стартам транскрипции

## Транскрипционные факторы

**Имеют** ДНК-связывающий домен, Специфически взаимодействуют с ДНК

Регулируют интенсивность синтеза РНК каждого конкретного гена в соответствии с потребностями организма (типом ткани, стадией развития, воздействиями окружающей среды)

## Белки-медиаторы

**Не имеют** ДНК-связывающего домена, Взаимодействуют с белок-белковыми и ДНК-белковыми комплексами

## Корегуляторные белки (коактиваторы и корепрессоры)

**Определение:**

**Корегуляторные белки (коактиваторы и корепрессоры)** не имеют ДНК-связывающих доменов и участвуют в регуляции транскрипции без непосредственного специфического взаимодействия с ДНК.

Корегуляторные белки, эффект которых состоит в активации транскрипции, называются коактиваторами, а белки, ослабляющие транскрипцию, именуются корепрессорами.

## Классы корегуляторных белков (в соответствии с механизмом их действия):

I. Влияют на структуру хроматина благодаря своей способности ковалентно модифицировать **гистоны**. Наиболее известные представители белков этого класса функционируют в составе **гистон-ацетилазных** (либо **гистон-деацетилазных**) комплексов, ослабляя либо, напротив, усиливая взаимодействие гистоновых белков с ДНК в составе нуклеосом.

II. Влияют на структуру хроматина, осуществляя модификации **цитозина в составе ДНК** (**DNMT-ДНК метилтрансферазы, белки семейства TET, etc.**)

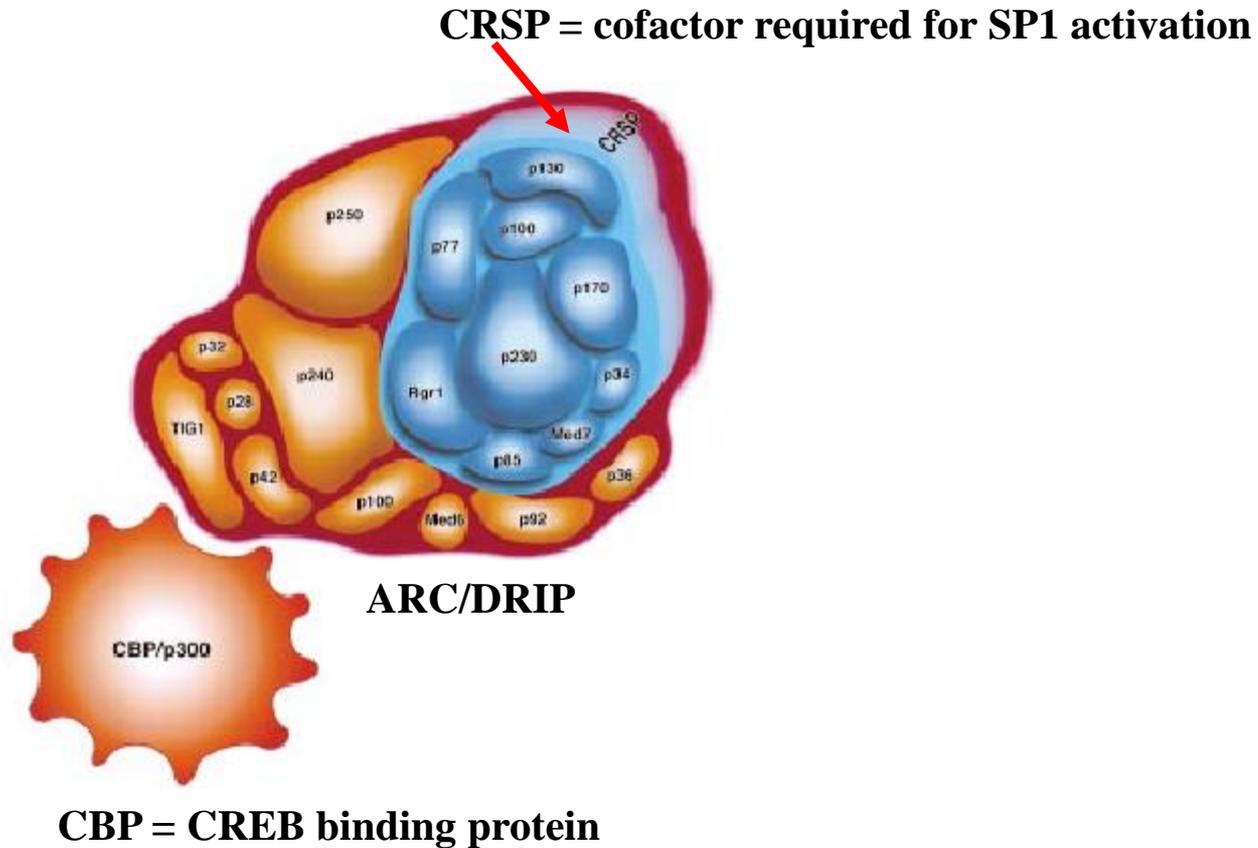
- **Райтеры** – записывающие белки, осуществляют ацетилирование, метилирование и т.д.

- **Ридеры** - белки, имеющие в своем составе домены, способные опознавать определенные модификации хроматина.

- **Эрайзеры** – белки, стирающие метки хроматина

III. Осуществляют АТФ-зависимую реорганизацию (ремоделинг) хроматина, включая ослабление связи ДНК с гистоновыми белками (разрыхление нуклеосомной укладки), а также перемещение нуклеосомы вдоль ДНК (**SWI/SNF комплекс**).

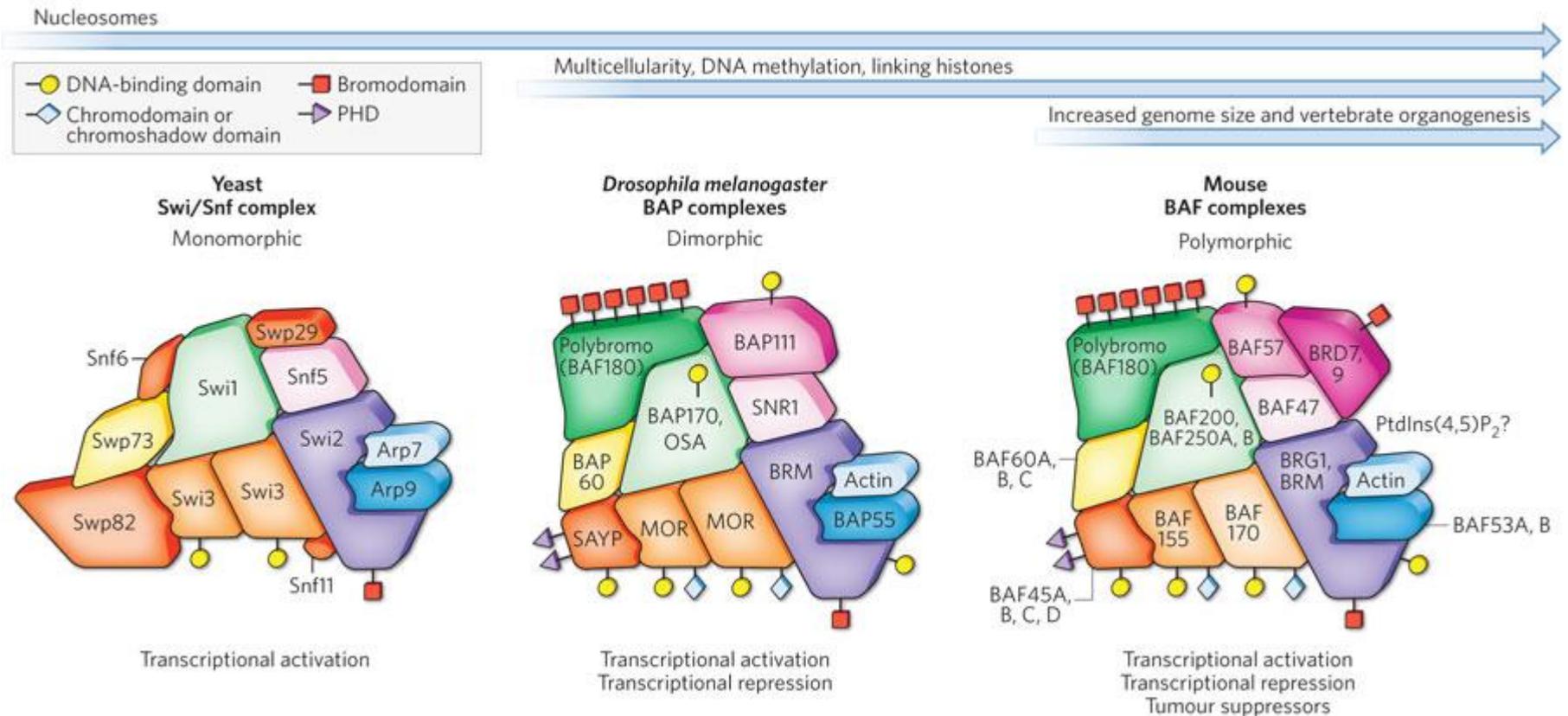
# ПРИМЕР: структура коактиваторного комплекса ARC/DRIP МЛЕКОПИТАЮЩИХ



**Комплекс CRSP опосредует активацию транскрипционным фактором SP1 не зависимо от ARC/DRIP**

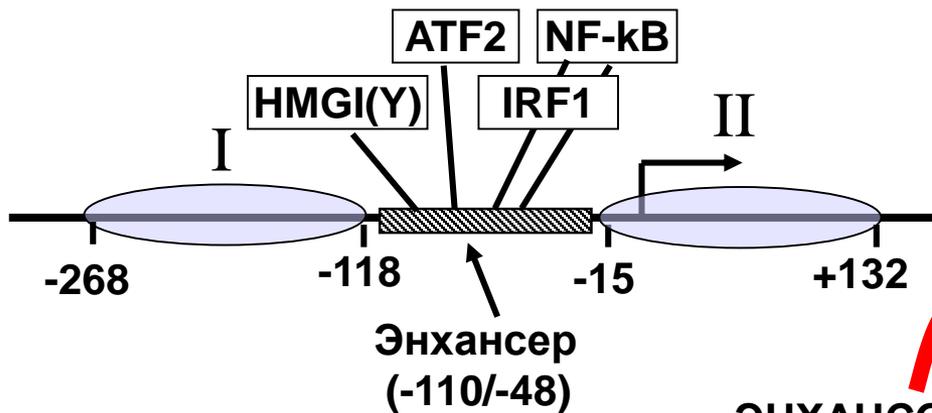
Hampsey M, Reinberg D. RNA polymerase II as a control panel for multiple coactivator complexes. Curr. Opin. Genet. Dev. 1999, 9, 2, 132-139

# Комплекс, осуществляющий АТФ-зависимую реорганизацию (ремоделлинг) хроматина



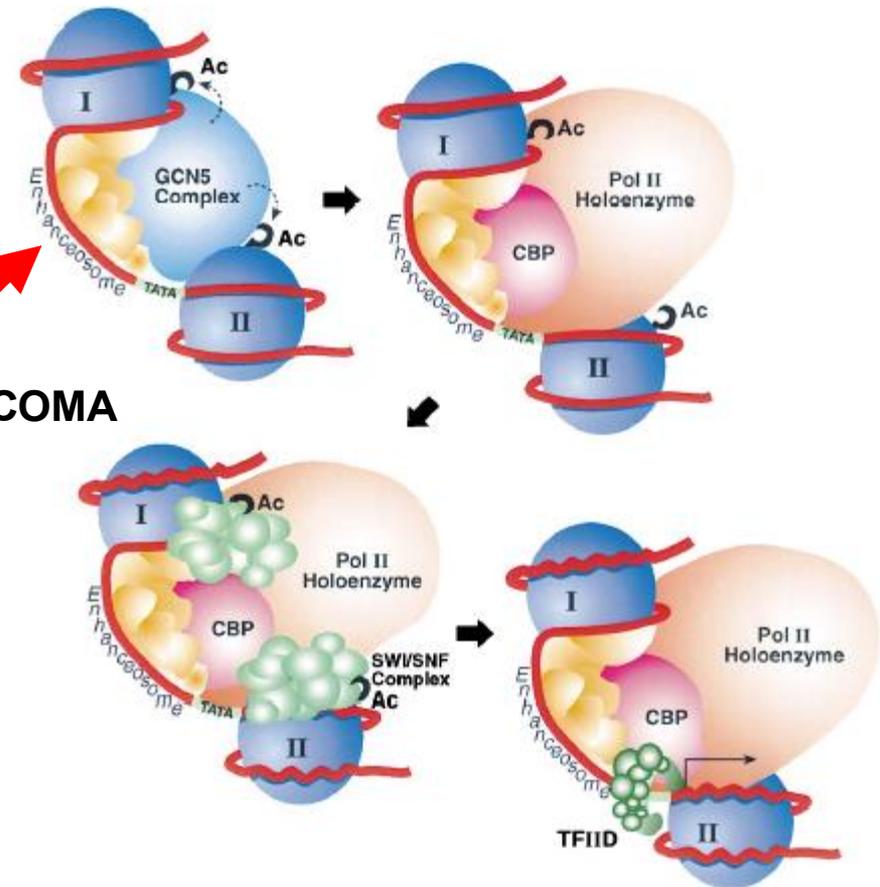
BAF complexes, brahma-associated factor complexes; BAP complex, BRM-associated proteins; BRD, bromodomain-containing protein; BRG1, brahma-related gene 1; MOR, Moira; PHD, plant homeodomain; PtdIns(4,5)P<sub>2</sub>, phosphatidylinositol-4,5-bisphosphate; SAYP, supporter of activation of yellow protein; SNR1, Snf5-related protein 1; **SWI/SNF, SWItch/Sucrose Non-Fermentable**

# Модель сборки комплекса хроматин-модифицирующих и базальных факторов на промоторе гена интерферона- бета человека



ЭНХАНСОСОМА

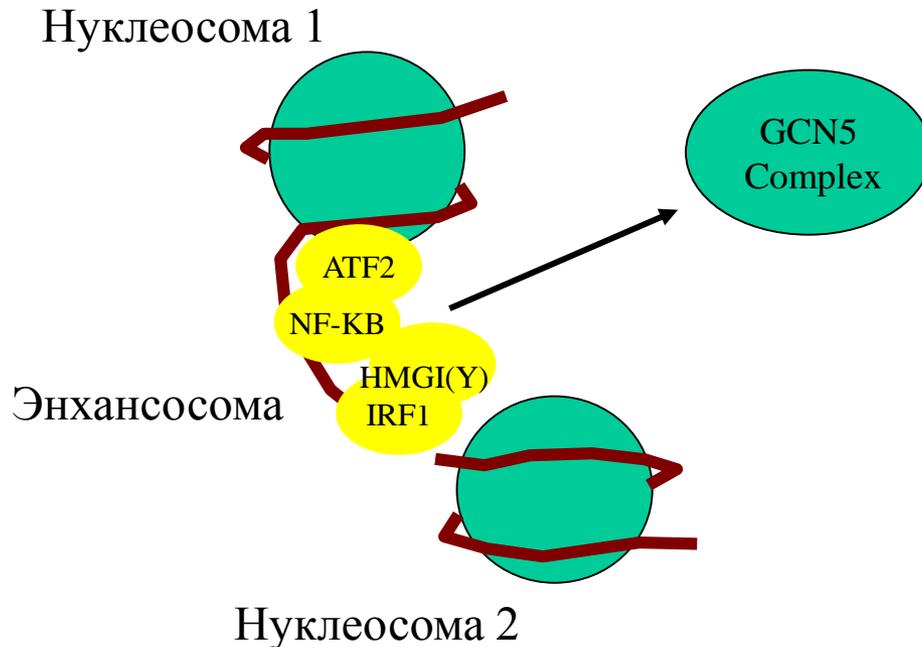
Вирусная инфекция активирует транскрипционные факторы ATF2, NF-KB, IRF1, которые вместе с HMGI(Y) формируют комплекс в районе энхансера – энхансосому. На ее поверхности происходит дальнейшая сборка активаторного комплекса, включающего хроматин-модифицирующие активности



Agalioti T., Lomvardas S., Parekh B., Yie J., Maniatis T., Thanos D. Ordered Recruitment of Chromatin Modifying and General Transcription Factors to the IFN- $\beta$  Promoter. Cell, 2000, Vol. 103, 667–678

# Регуляция транскрипции гена интерферона $\beta$ человека. Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

## Стадия 1: сборка энхансосомы



### Участники:

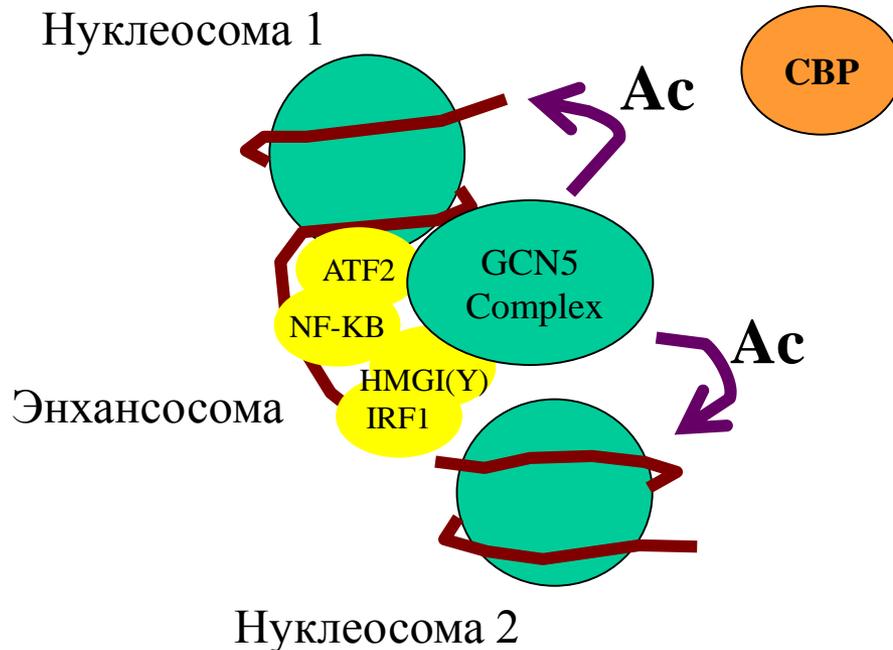
Белки (транскрипционные факторы): ATF2, NF- $\kappa$ B, IRF1, HMGI(Y)

Участок ДНК (энхансер), свободный от нуклеосомной укладки

Результат: образуется энхансосома - ДНК-белковый комплекс, способный притягивать мультибелковый комплекс GCN5

# Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

## Стадия 2: Ацетилирование гистонов с участием комплекса GCN5



### Участники:

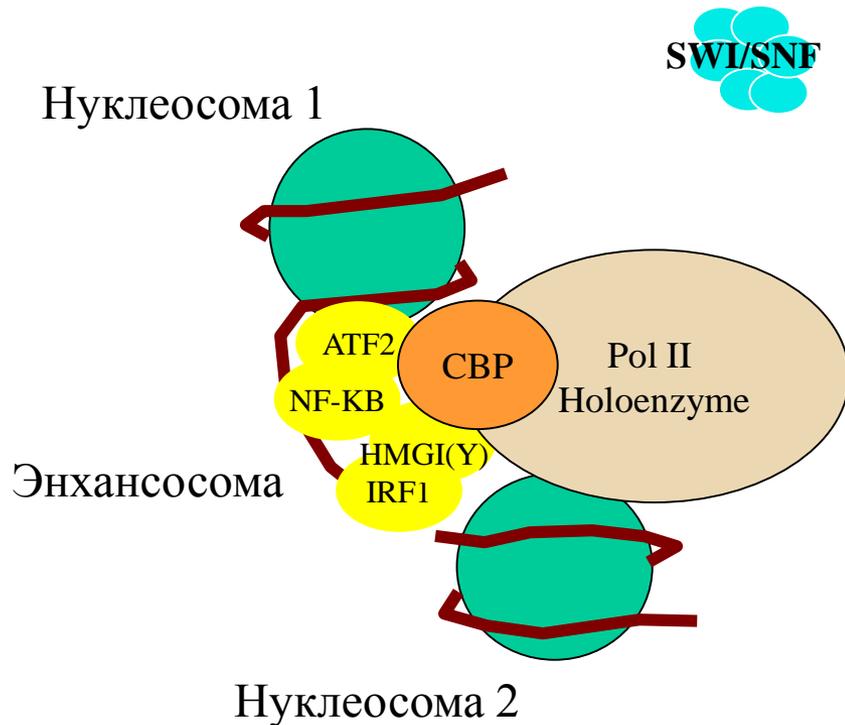
Гистон-ацетилазный комплекс: GCN5

N - концевые участки гистоновых белков

Результат: ДНК-белковый комплекс приобретает конформацию, оптимальную для привлечения белка-коактиватора **CBP** (CREB binding protein)

# Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

## Стадия 3: Привлечение комплекса СВР/ Pol II



Участники:

Комплекс: ДНК / энхансосома

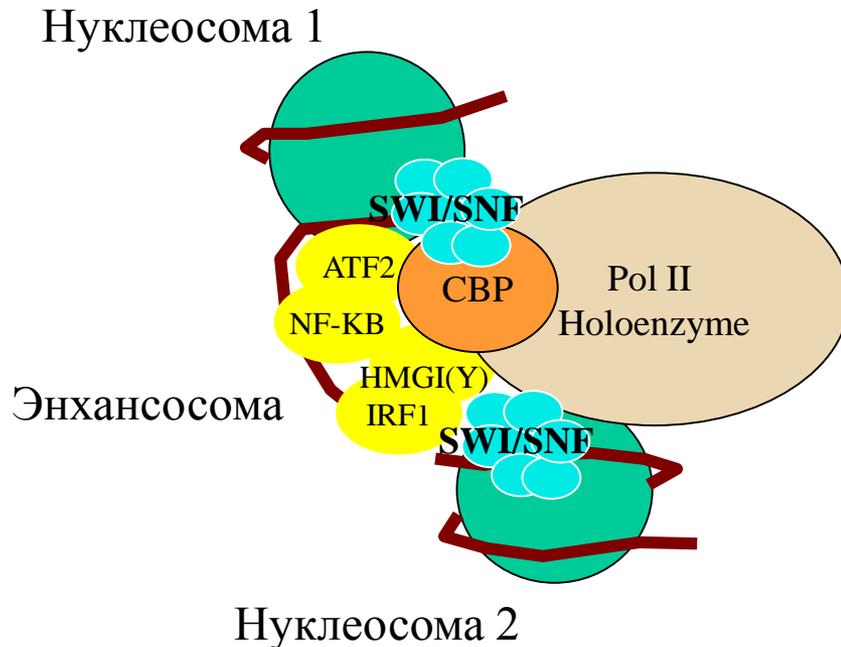
Белок-коактиватор: СВР (CREB binding protein)

Белковая машина: холоэнзим, включающая белок Pol II

Результат: Создается возможность для привлечения SWI/SNF комплекса (хроматин-ремоделирующий комплекс)

# Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

## Стадия 4: Привлечение комплекса SWI/SNF



Участники:

Хроматин-ремоделирующая белковая машина SWI/SNF.

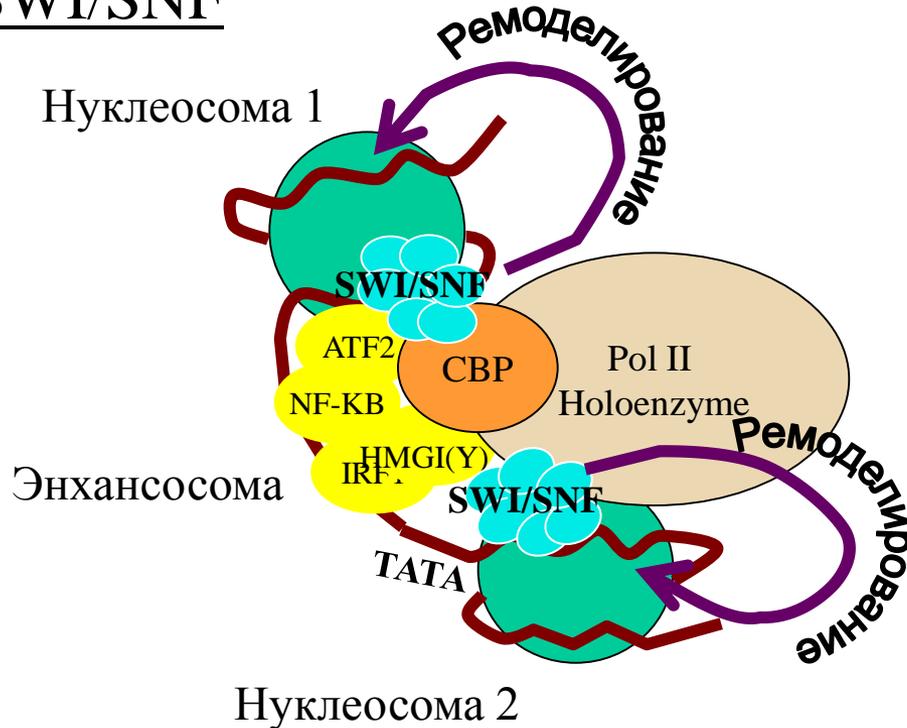
Комплекс ДНК/энхансосома/ CBP

Результат: Создается возможность для функционирования белковой машины SWI/SNF

# Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

Стадия 5: Ремоделирование хроматина (нуклеосомной укладки) с участием хроматин-ремоделирующей белковой машины

## SWI/SNF



Участники:

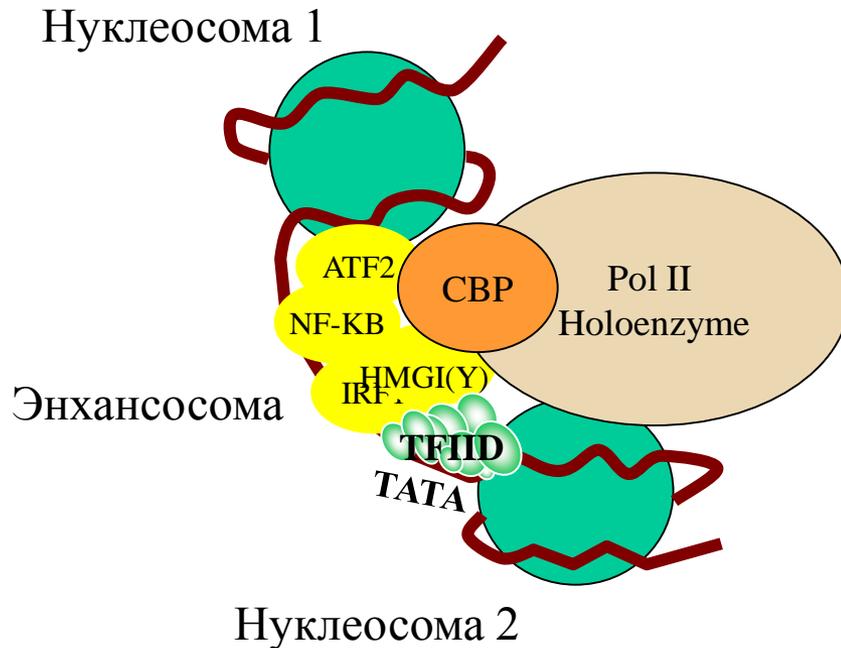
Хроматин-ремоделирующая белковая машина SWI/SNF.

Нуклеосомы

Результат: Нуклеосомы разрыхляются, ТАТА бокс становится доступным для взаимодействия с TFIID.

# Подготовка корового промотора гена интерферона $\beta$ человека к контакту с компонентами ПИК

## Стадия 6: Привлечение белка TFIID



### Участники:

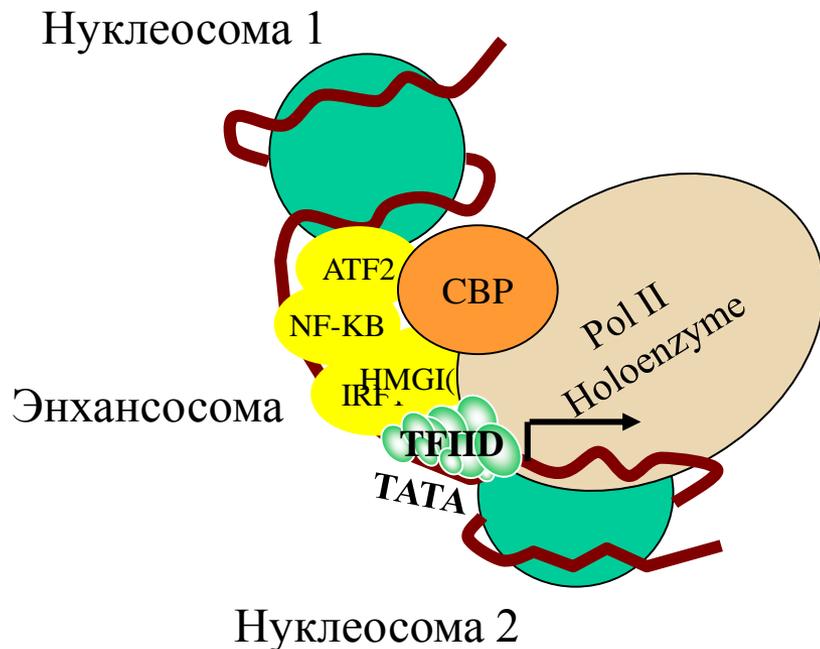
Промотор гена, включающий ТАТА бокс  
Базальный транскрипционный фактор TFIID.

Результат: Становится возможным формирование прединициаторного комплекса



# Формирование прединициаторного комплекса (ПИК) на промоторе гена интерферона $\beta$ человека

## Стадия 1: Закрытый комплекс



Участники:

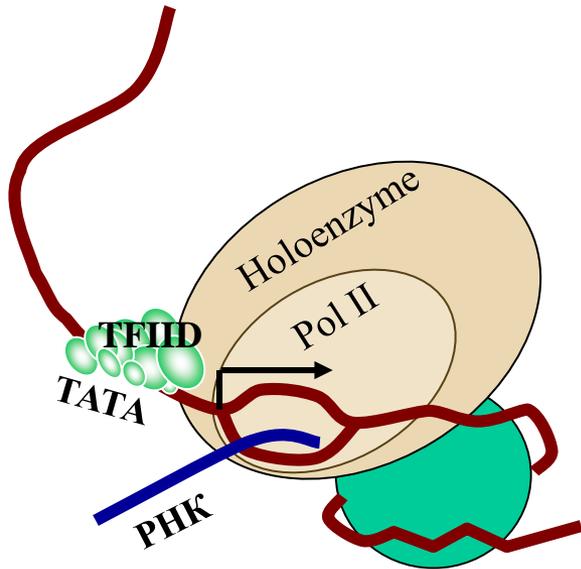
Комплекс ДНК/белок: ТАТА бокс/  
TFIID

Белковая машина: холоэнзим

Результат: Холоэнзим плотнее контактирует с ДНК. РНК-полимераза связывается с промотором. Образуется «закрытый комплекс»

# Инициация транскрипции гена интерферона $\beta$ человека

## Стадия 2: Открытый комплекс -> инициация транскрипции



### Участники:

РНК полимера Pol II

Матричная цепь ДНК

Результат: Инициация транскрипции проходит стадию открытого комплекса. Синтезируются первые 2-9 нуклеотидов РНК

# ЭТАПЫ ТРАНСКРИПЦИИ

Эукариоты

-Подготовка корового промотора к контакту с компонентами ПИК

-Посадка РНК-полимераз в комплексе со вспомогательными белками на ДНК в районе старта транскрипции, что у эукариот означает формирование прединициаторного комплекса (ПИК)

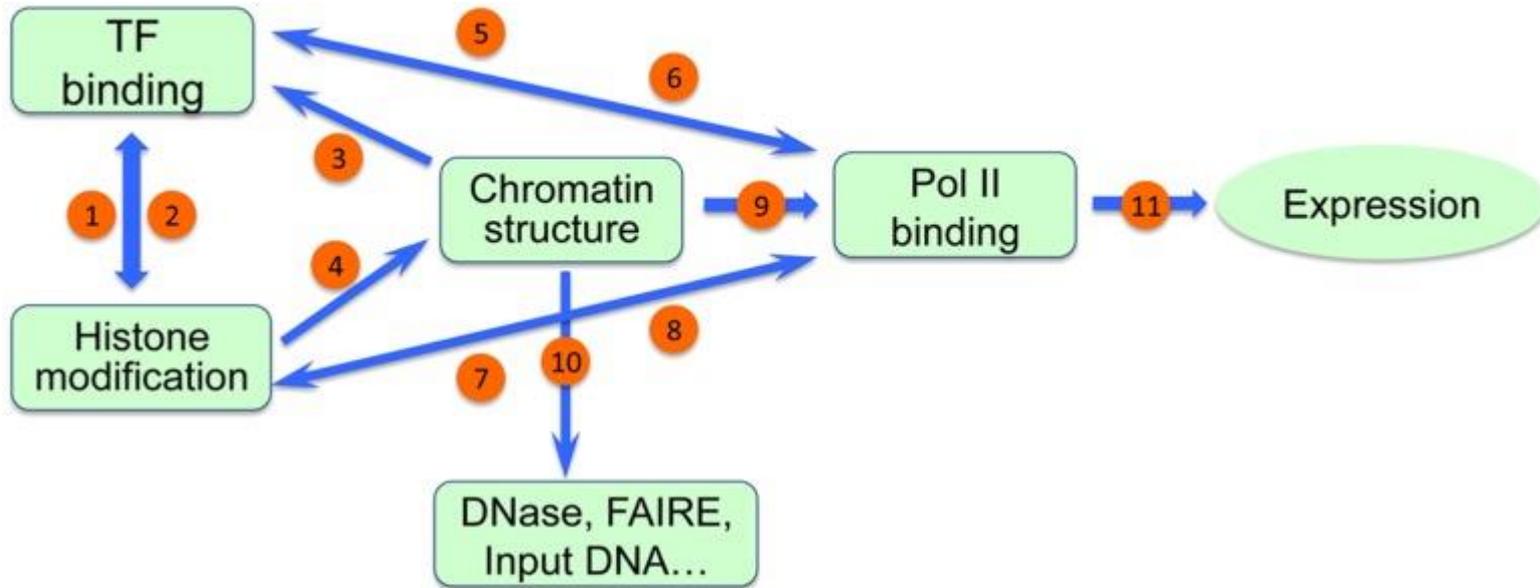
-Инициация транскрипции

-Элонгация РНК

-Терминация транскрипции

Прокариоты, Эукариоты

# Дополнительный слайд для самостоятельного ознакомления



- (1) Recruiting histone modifiers
- (2) Recruiting TFs
- (3) Accessibility
- (4) Remodeling
- (5) Recruiting general TFs
- (6) Interacting with TFs

- (7) Recruit general TFs
- (8) Interacting with histone modifiers
- (9) Accessibility
- (10) Accessibility
- (11) Transcription

Cheng C. et al., Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* 2012 Sep;22(9):1658-67.

# 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)

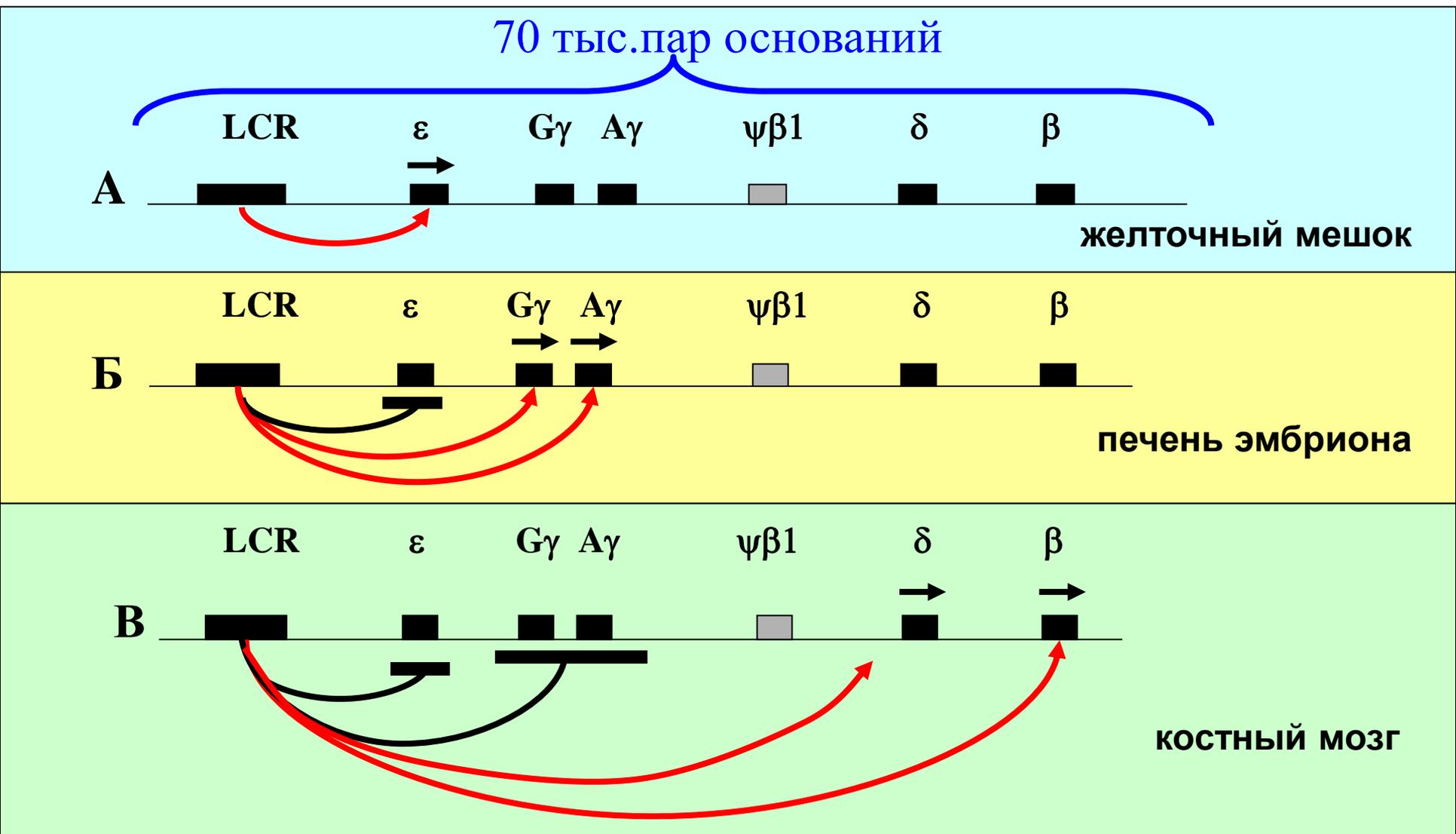
- фрагмент (или группа фрагментов) ДНК, присутствие которых обеспечивает высокий уровень тканеспецифической экспрессии связанного с ним гена, входящего в трансгенную конструкцию,
- пропорционально количеству копий трансгена и
- независимо от места встраивания в геном .

Как правило, локус-контролирующие районы регулируют гены, находящиеся в кластерах и обеспечивают их ткане- и стадии-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 $\beta$ -ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА

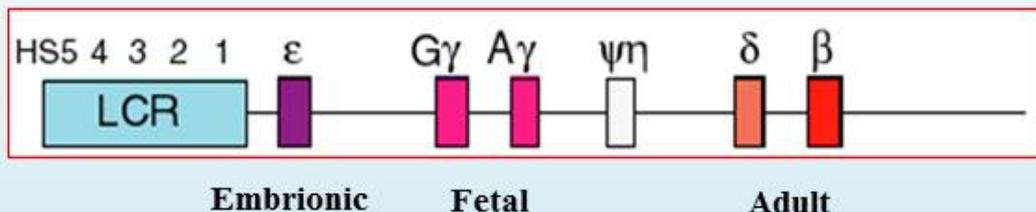


$\epsilon$ ,  $G\gamma$ ,  $A\gamma$ ,  $\delta$ ,  $\beta$  - гены гемоглобинов;  $\psi\beta 1$  - псевдоген

# КЛАСТЕР $\beta$ -ГЛОБИНОВЫХ ГЕНОВ ЧЕЛОВЕКА И ВРЕМЕННАЯ ДИНАМИКА ЭКСПРЕССИИ В ОНТОГЕНЕЗЕ

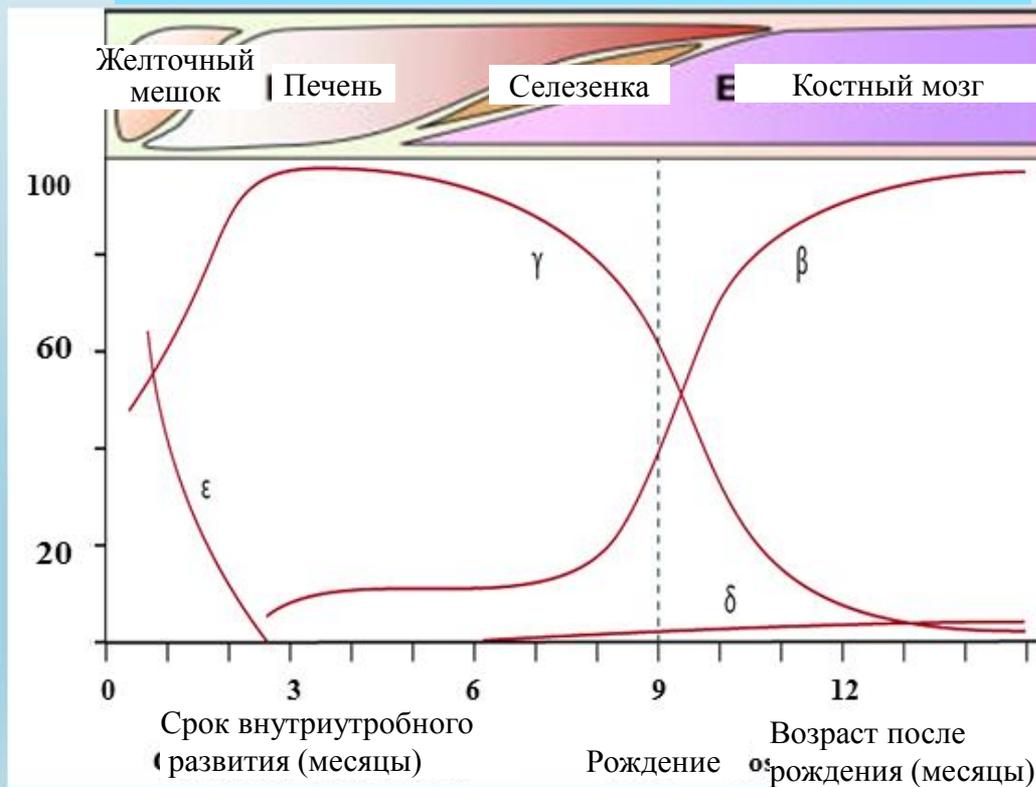
B-Globin locus

Chr11



Орган, где осуществляется эритропоэз

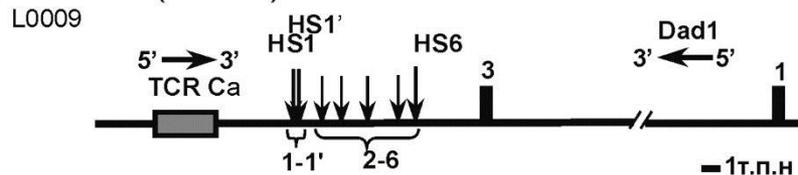
Уровень экспрессии глобинов (%)



Эмбриональный ( $\epsilon$ ) 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 $\alpha/\delta$* (мышь)



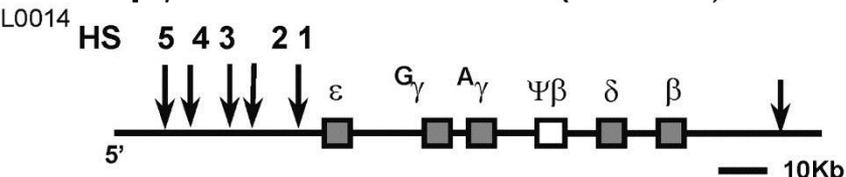
HS1 - энхансер

HS1, HS1' район определяющий распространенность в тканях

HS2 - HS6 район открывающий хроматин

Цифрами обозначены экзоны гена *Dad1*

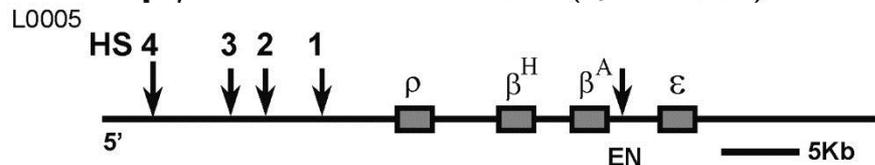
## Кластер $\beta$ глобиновых генов (человек)



HS5 - инсулятор

HS4 - HS1 - энхансер

## Кластер $\beta$ глобиновых генов (цыпленок)

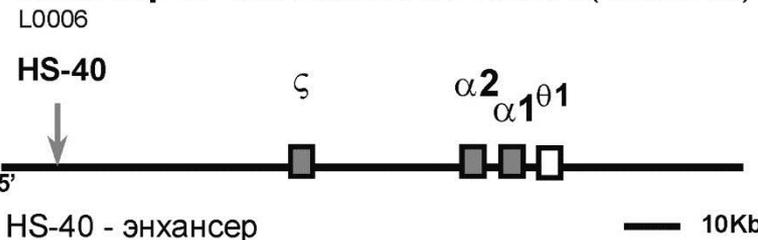


HS4 - инсулятор

HS3, HS2 - энхансер

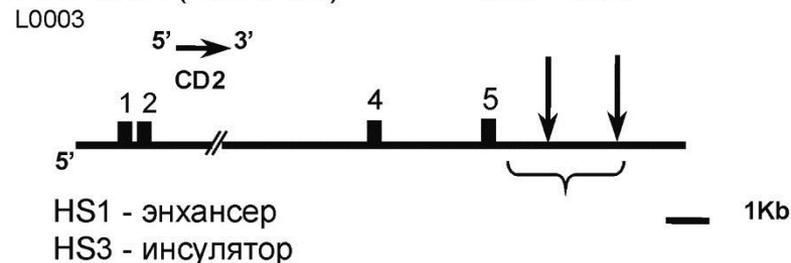
EN - межгенный энхансер

## Кластер $\alpha$ -глобиновых генов (человек)



HS-40 - энхансер

## Ген *CD2* (человек)

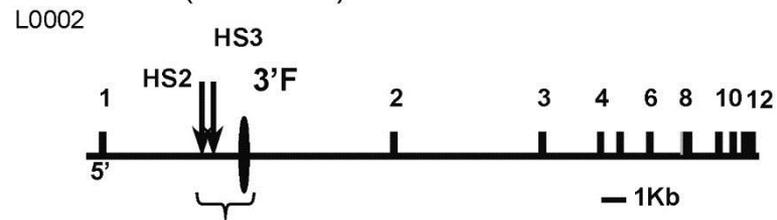


HS1 - энхансер

HS3 - инсулятор

Цифрами обозначены экзоны гена *CD2*

## Ген *ADA* (человек)



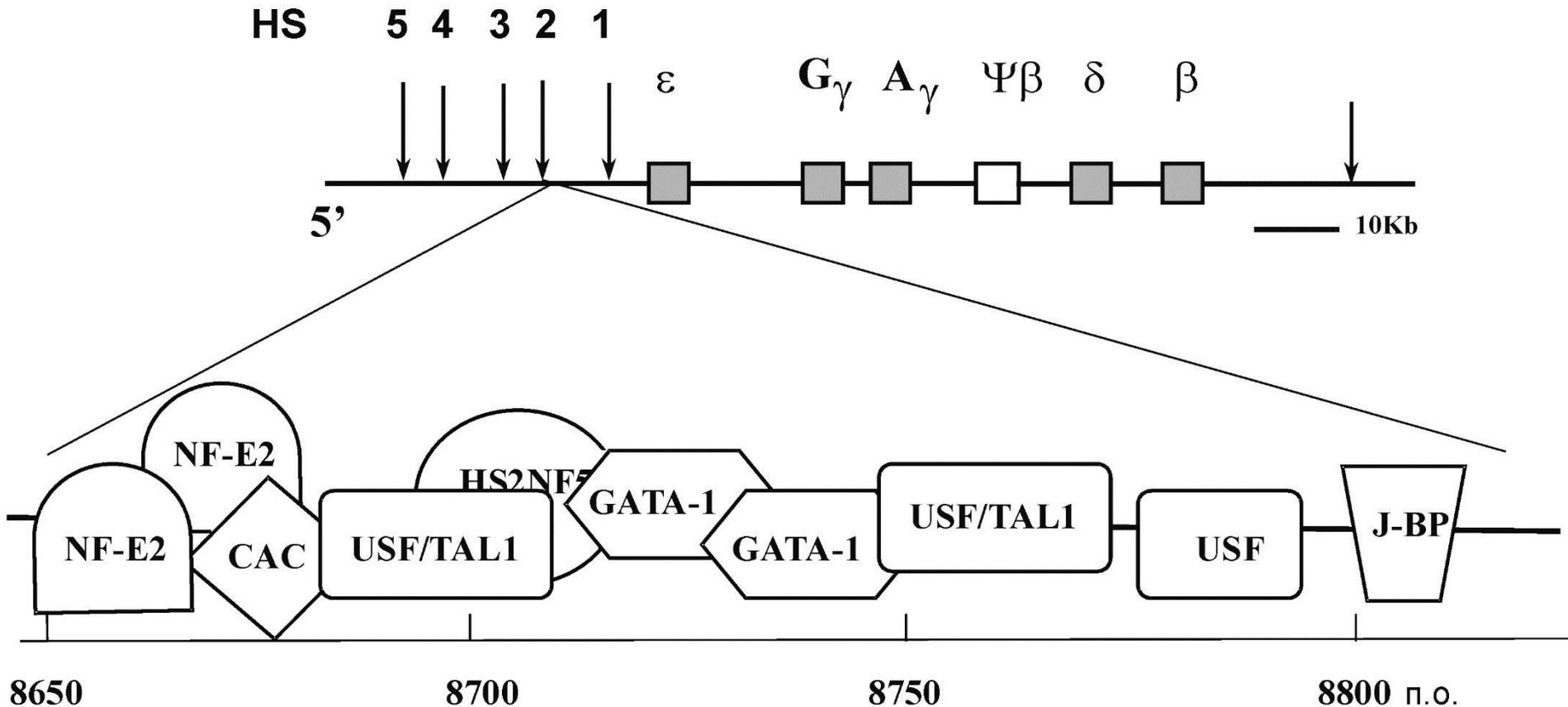
HS2 - 5'вспомогательный элемент

HS3 - энхансер;

3'F - 3'вспомогательный элемент

Цифрами обозначены экзоны гена *ADA*

# LCR глобинового локуса человека



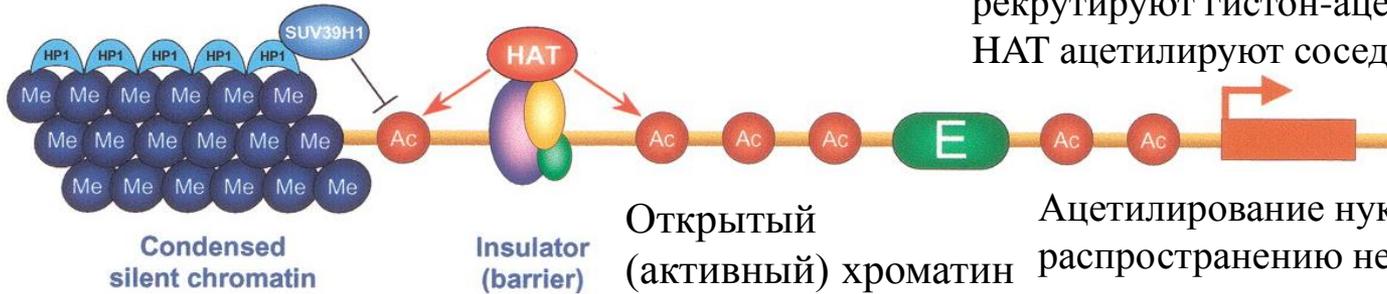
Каждый LCR включает специфическую группу сайтов связывания и располагается иногда на очень большом (до десятков тысяч п.о.) расстоянии от контролируемой кассеты генов

# 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-ацетилазасы (НАТ). НАТ ацетилируют соседние нуклеосомы.

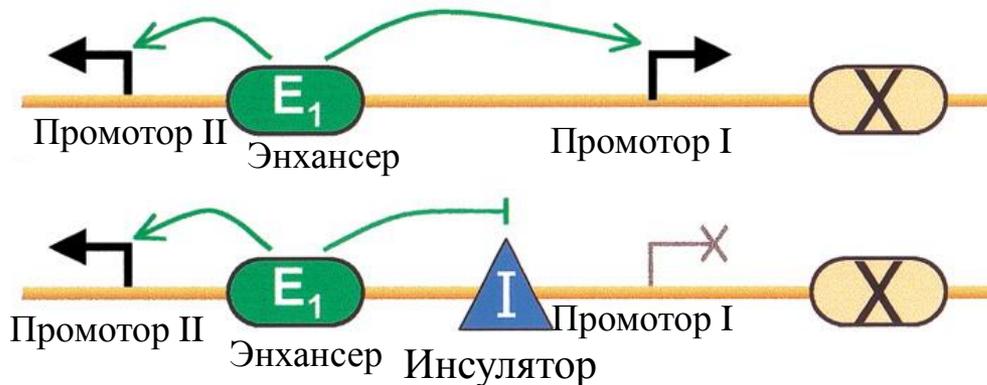


Открытый (активный) хроматин

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

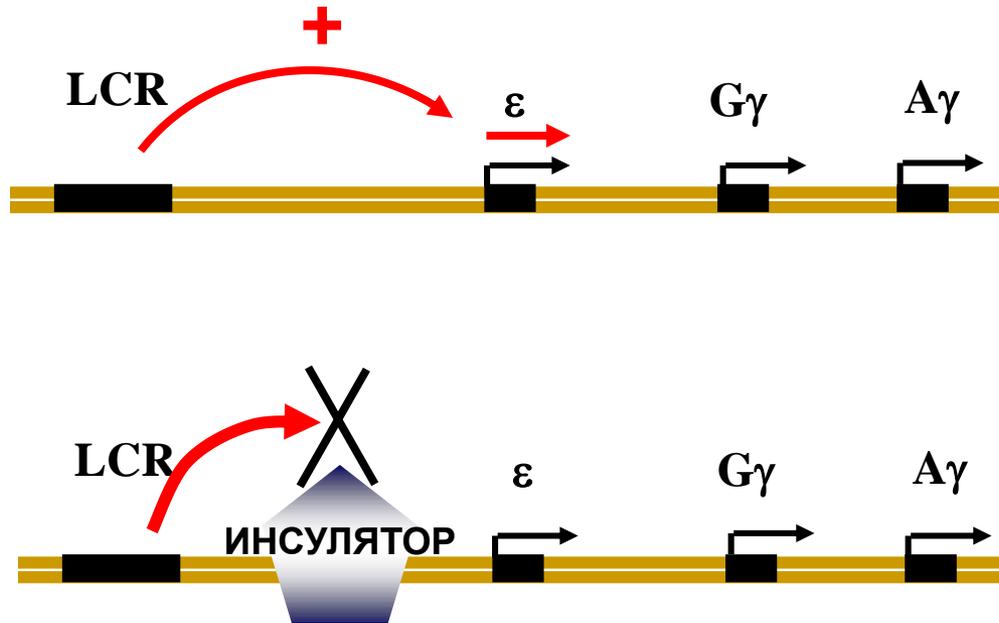
Инсулятор может располагаться между энхансером и промотором и блокировать активирующее влияние энхансера на транскрипцию гена



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.

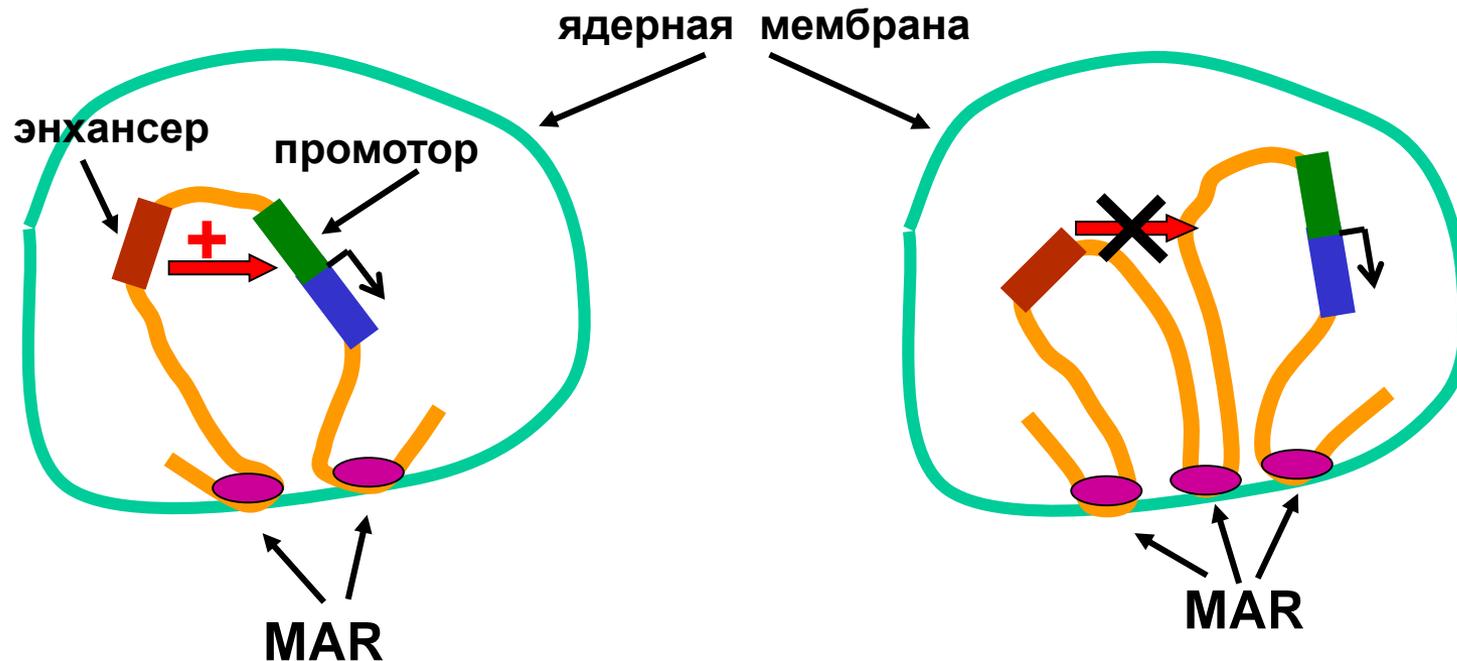
# ИНСУЛЯТОР



**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.**

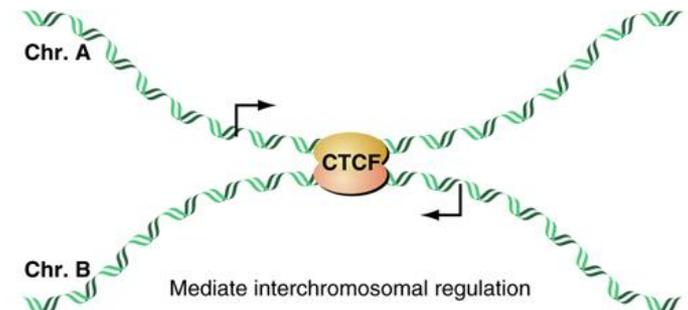
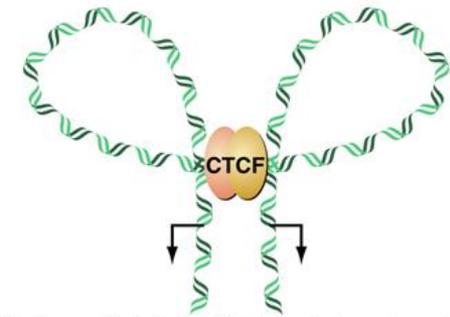
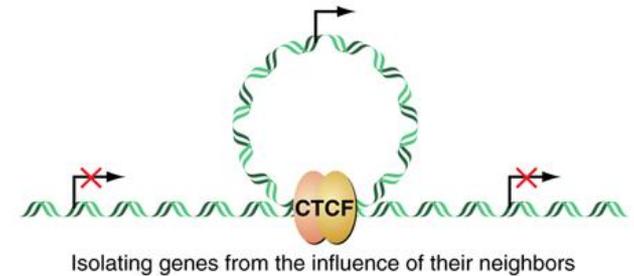
**При включении инсулятора два регуляторных района оказываются в различных доменах, и не способны взаимодействовать**

# Роль транскрипционных факторов CTCF при формировании петель ДНК

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

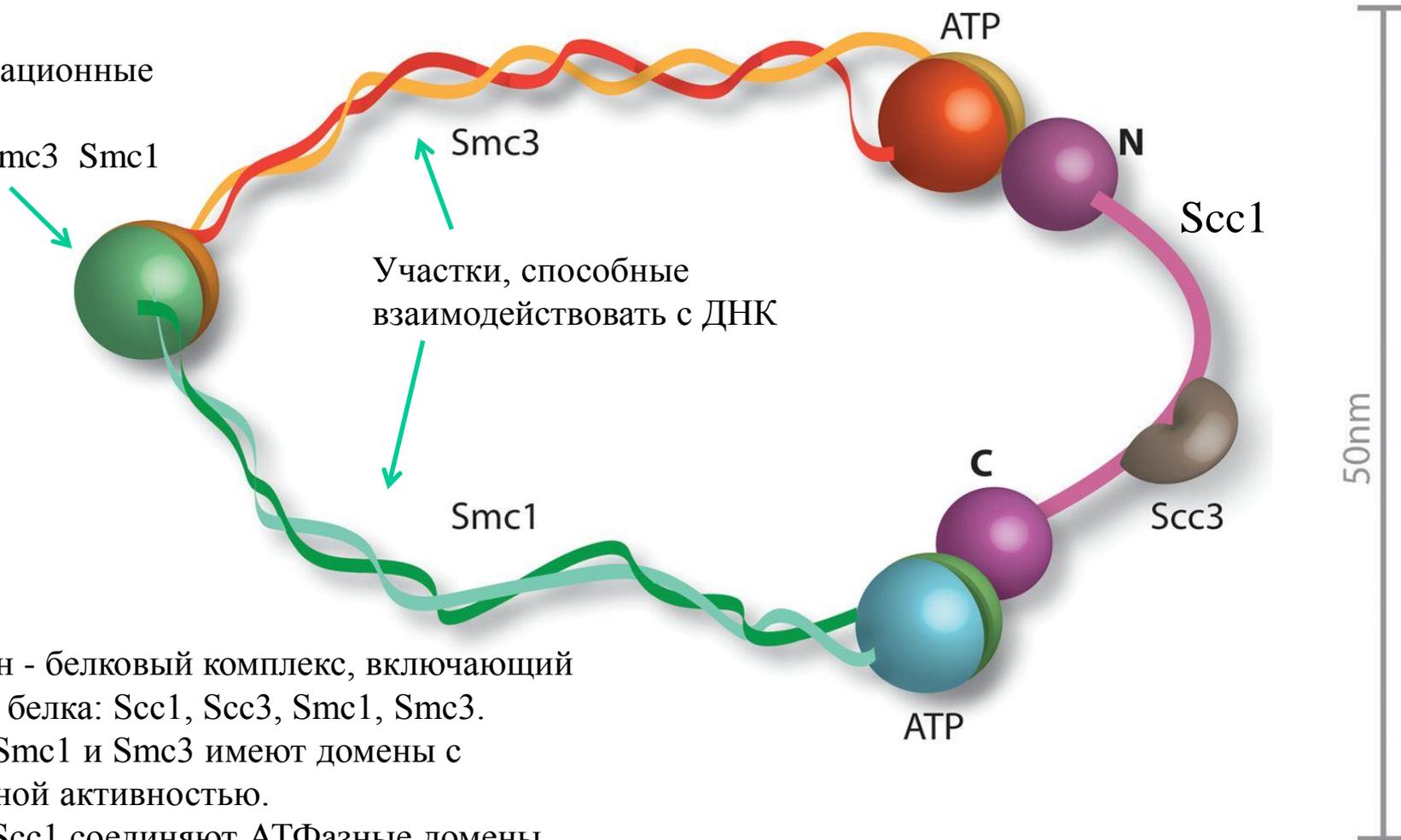
**CTCF** - димерный фактор. Имеет пространственную структуру, обеспечивающую возможность взаимодействовать с различными нитями ДНК, за счет чего в ядре клетки могут формироваться петли ДНК либо межхромосомные контакты.

**CTCF** является маркером участков ДНК, разделяющих активный и репрессированный хроматин



# Когезин – комплекс белков , формирующих кольцеобразную структуру

Димеризационные  
домены  
белков Smc3 Smc1



Участки, способные  
взаимодействовать с ДНК

Smc1

ATP

N

Scc1

C

Scc3

ATP

50nm

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

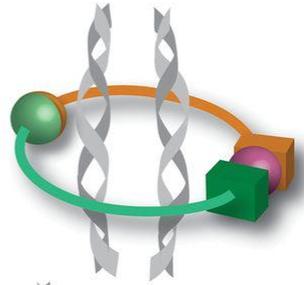
Белки Smc1 и Smc3 имеют домены с  
АТФазной активностью.

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

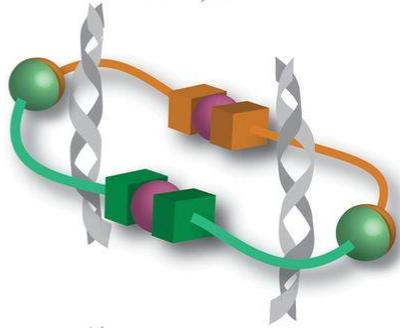
Белок Scc3 связывается с С-терминальным  
участком белка Scc1

# Модели 3-Д структур хроматина с участием когезина

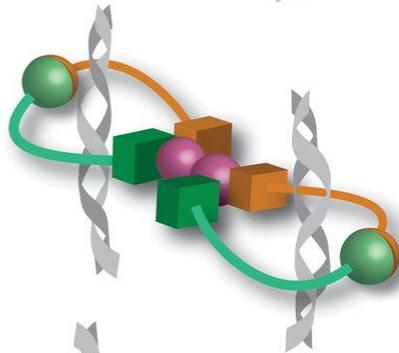
A) *Ring model:*



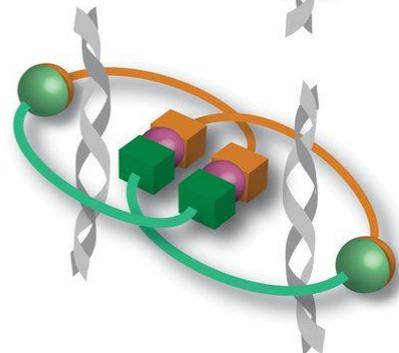
B) *Handcuff model (1):*



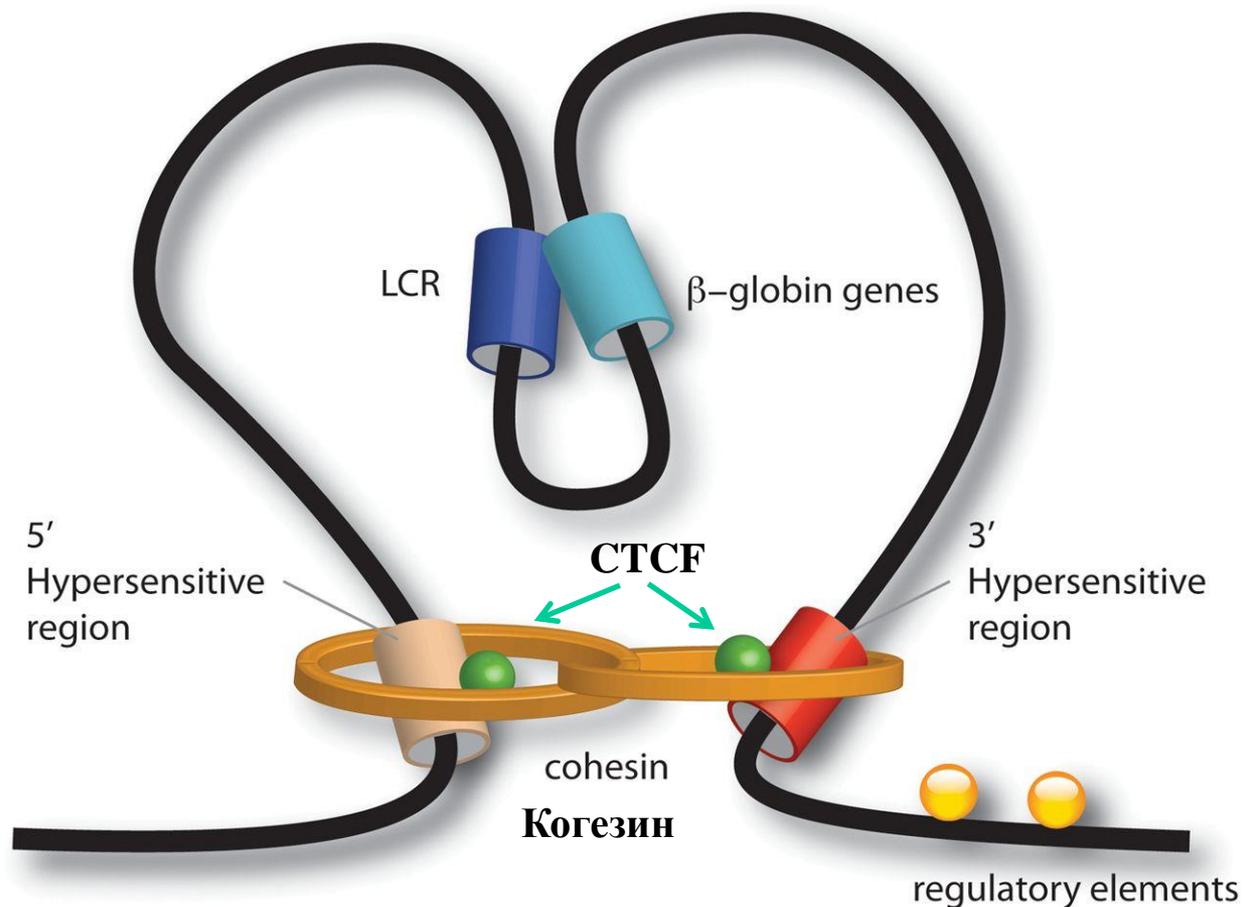
*Handcuff model (2):*



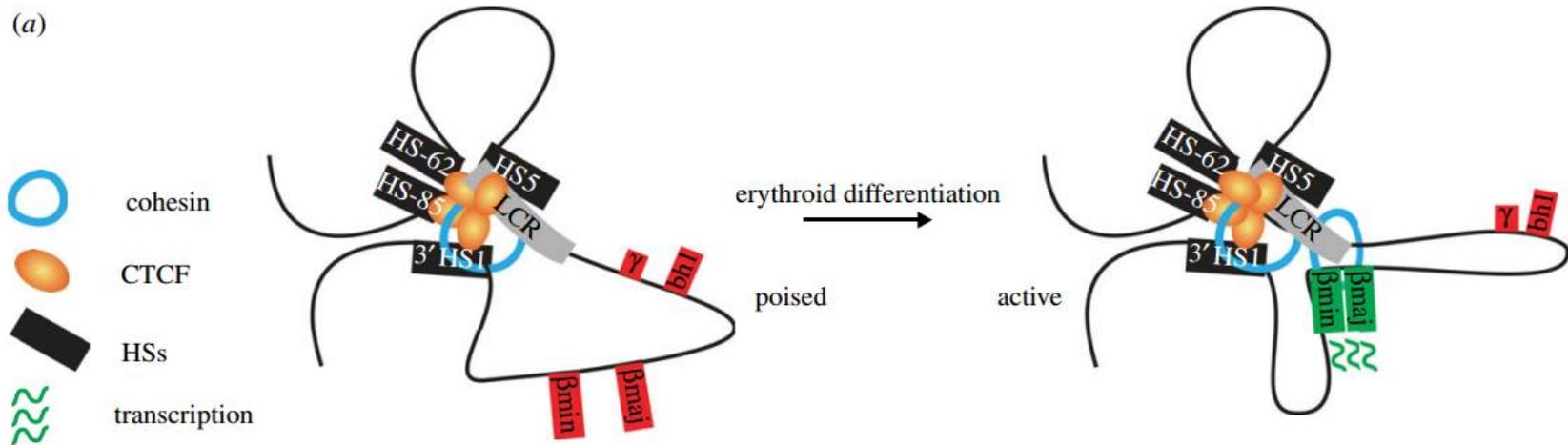
*Handcuff model (3):*



# Кластер генов $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -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

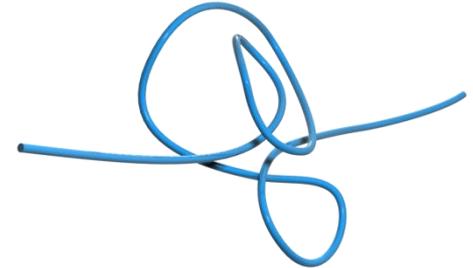
Исследованы контакты в клеточных культурах человека MCF7 и K562



Промоторная модель

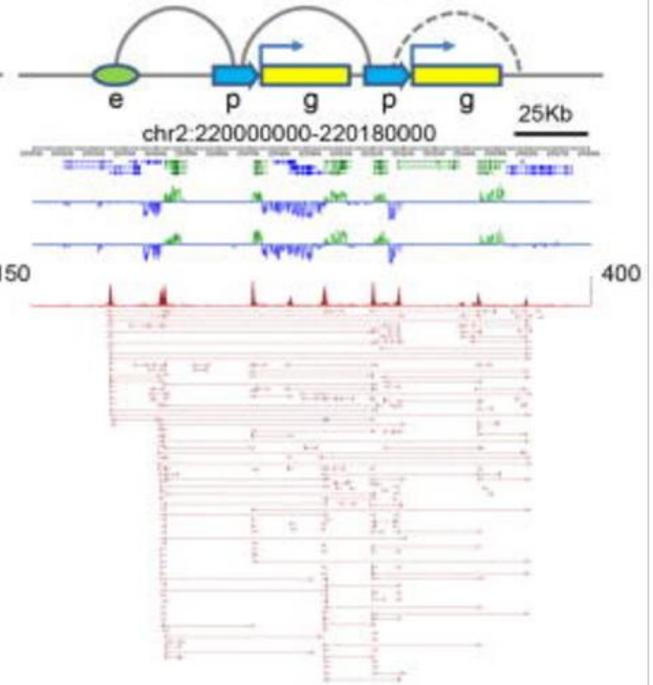
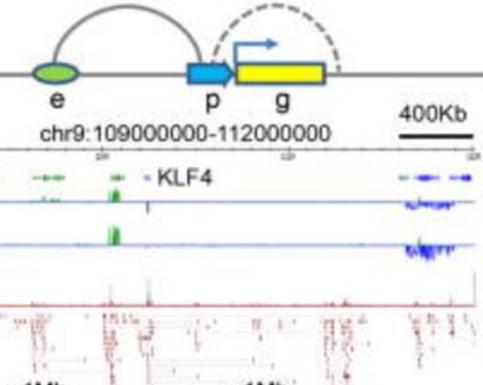
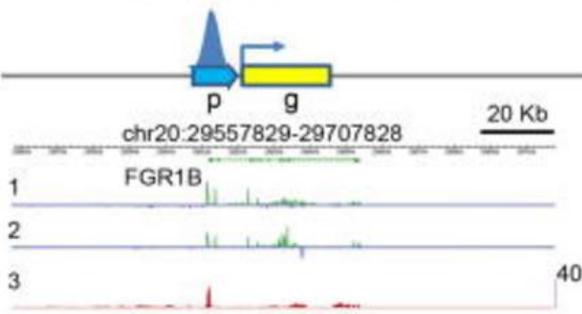


Одногенная модель  
(Промотор-Энхансер)  
Single-gene interaction model



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

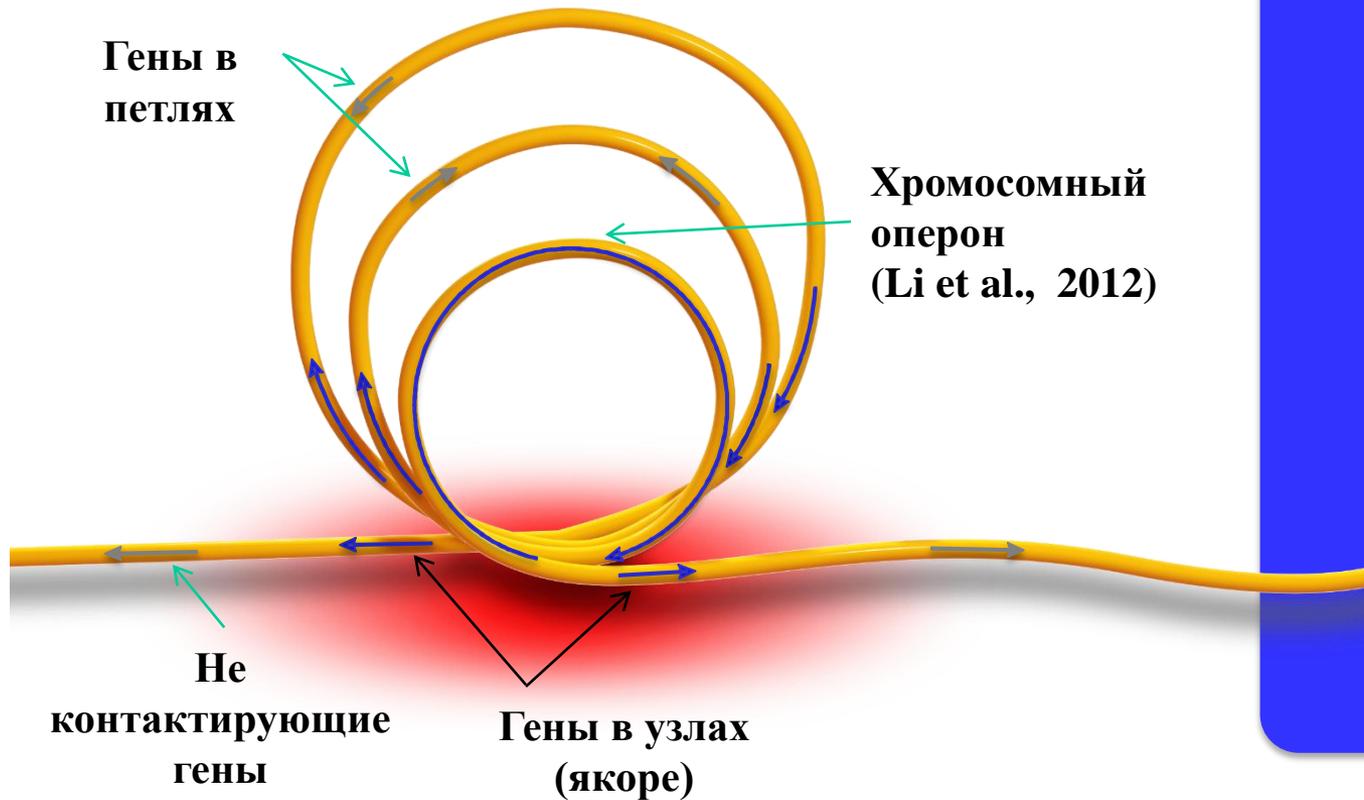
Basal promoter model



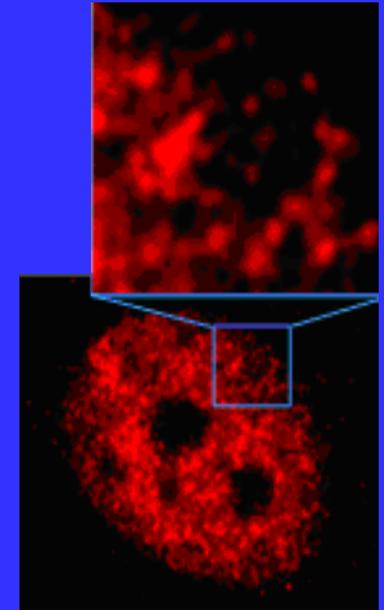
Li G. Et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012 Jan 20;148(1-2):84-98.

# Хромосомные опероны и транскрипционные фабрики

Структуры, найденные с помощью ChIA-PET



Ядро клетки – участки транскрипции под микроскопом



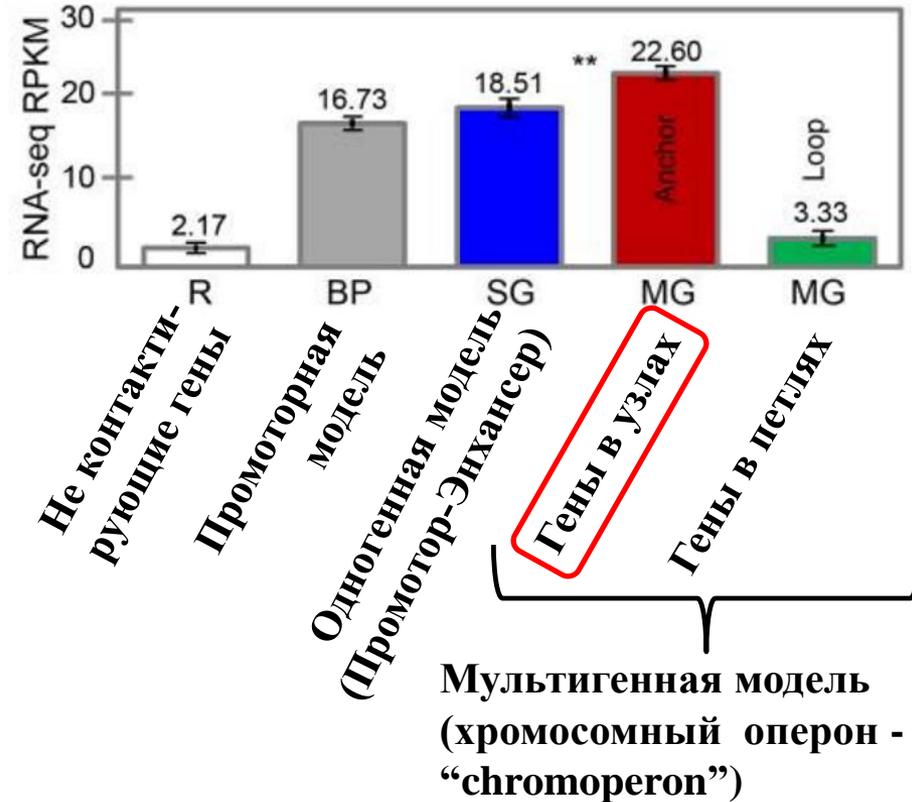
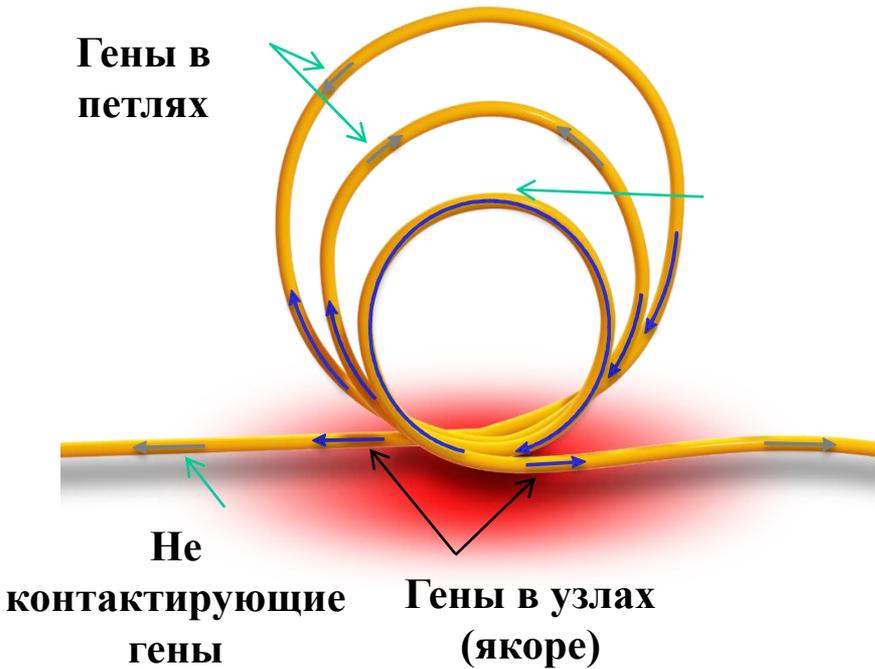
FISH

**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

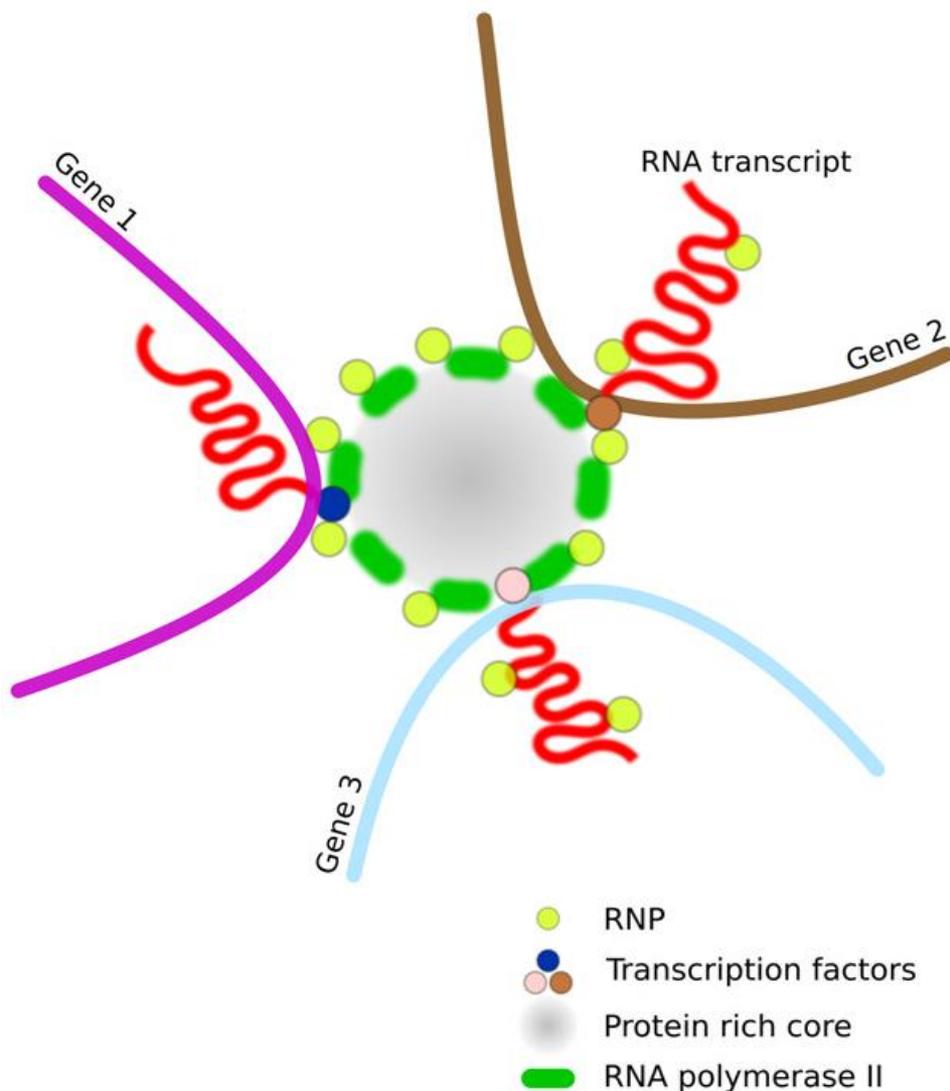
Структуры, найденные с помощью методики ChIA-PET, в MCF7 клетках человека

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)

KCl

Нейрональные клетки мыши

Деполяризация мембраны

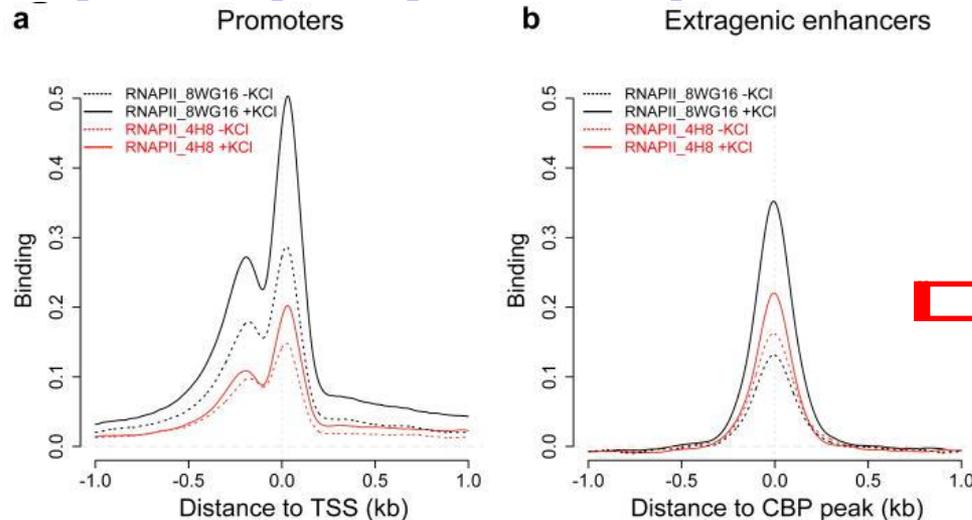
Активация коактиваторного белка P300/CBP

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)

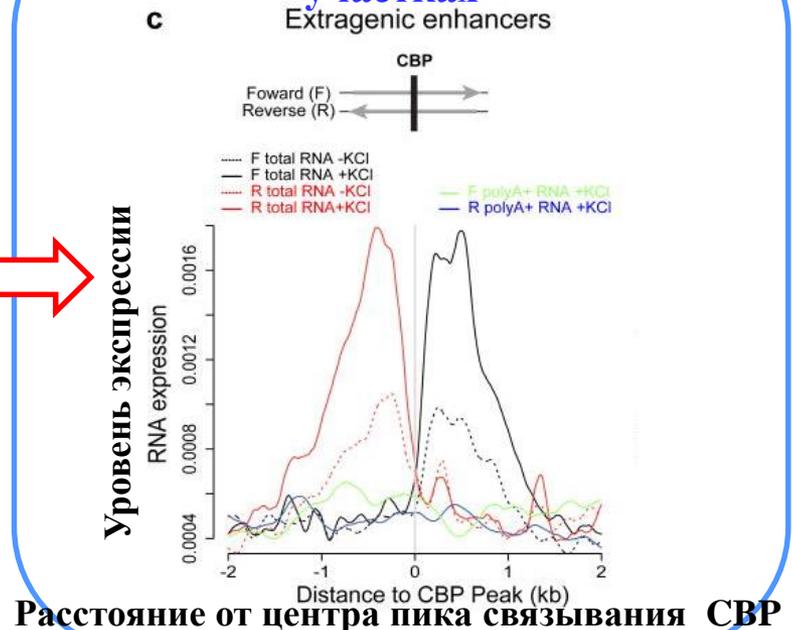
## Связывание РНК полимеразы II (RNAPII) в районах промоторов и энхансеров



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)



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).

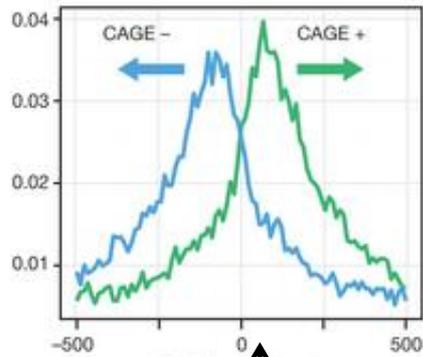
Были выявлены транскрипты (короткие РНК), которые считывались как с прямой (+), так и обратной (-) цепи ДНК в районе энхансера. Расстояние между (+) и (-) пиками соответствовало длине участка ДНК, упакованного в нуклеосому (180 п.о.).

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

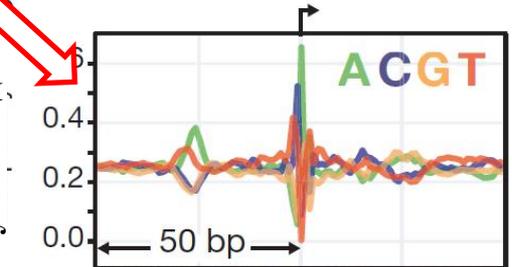


Позиция относительно центра  
(пик связывания белка Р300)

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

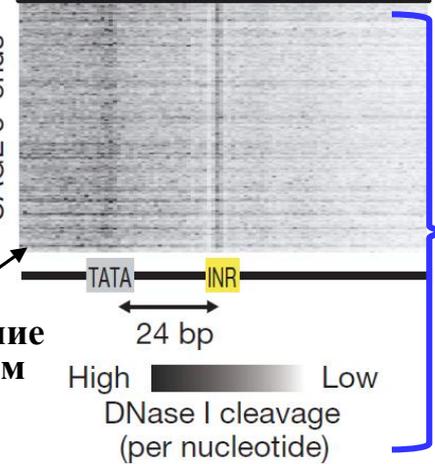
Inr – инициаторный элемент, можно описать консенсусным мотивом YYANWYY, где Y = C либо T

Частота встречаемости нуклеотида



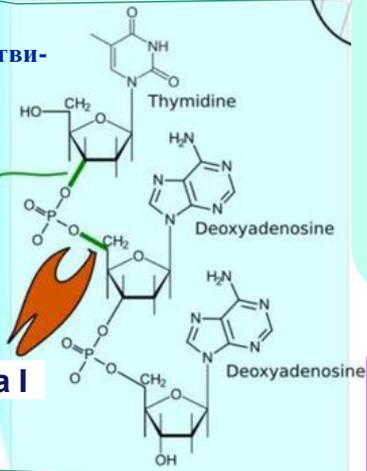
Enhancer  
CAGE 5' ends

Расщепление ферментом DNase I



Сайт гиперчувствительности к DNase I

Фосфодиэфирная связь



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

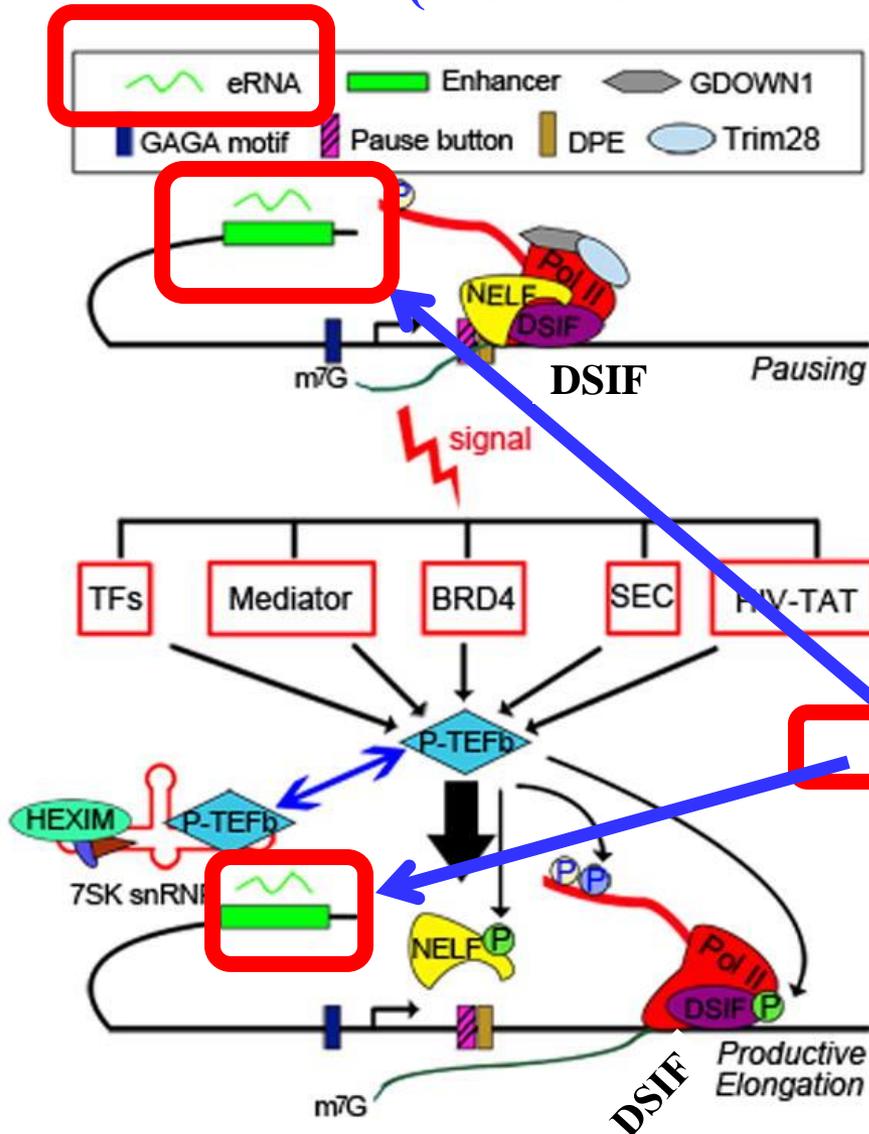
Позиции TATA боксов и INR элементов характеризовались наибольшей чувствительностью к DNase I (менее плотная нуклеосомная упаковка, наибольшая доступность для контакта с белками), что подтверждает их функциональную значимость.

## 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
- ....



# 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)

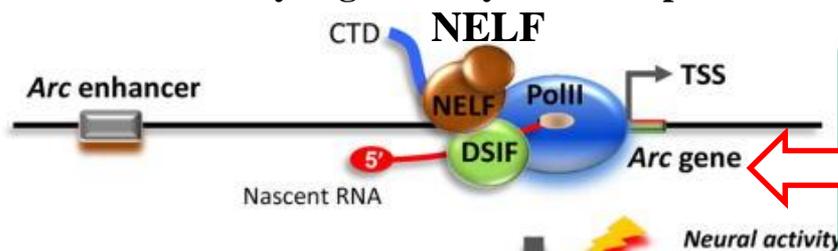
- Дифференцировке стволовых клеток (SEC, ТФ)

- 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**

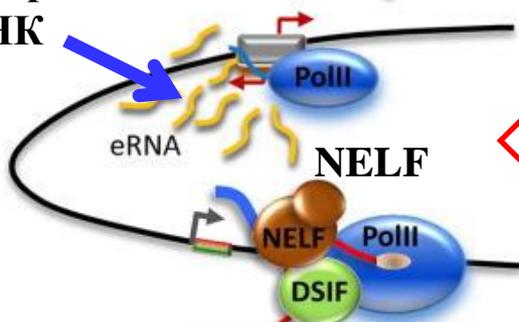
Ген Arc = Activity-regulated cytoskeletal protein



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

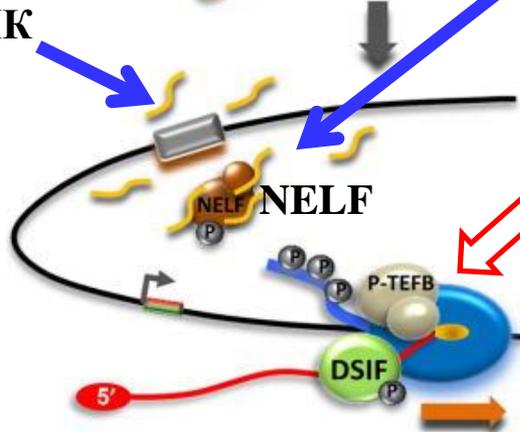
**НЕГАТИВНЫЕ РЕГУЛЯТОРЫ ЭЛОНГАЦИИ:**  
NELF = negative elongation factor  
DSIF = DRB Sensitivity Inducing Factor

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»

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

**Nucleic Acids Research**  
Now Available: the 2017 Web Server Issue

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

### Nucleic Acids Research

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

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

**Explore the 2017 NAR Database Issue today**  
54 new database descriptions

**Current Issue**  
Volume 45, Issue 17  
29 September 2017

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

Advance articles Submit Purchase About All Nucleic Acids Res Advanced Search

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

### Database issue

#### NUCLEIC ACID SEQUENCE, STRUCTURE, AND REGULATION

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

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

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

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

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

Database Resources of the National Center for Biotechnology Information

Ссылка на «online Database Collection»

# Журнал NAR: online Database Collection

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



OXFORD ACADEMIC Journals

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

### NAR Database Summary Paper Alphabetic List

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

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

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

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

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

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

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

Полный список баз данных,  
сгруппированных по категориям .

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



OXFORD ACADEMIC Journals

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

### NAR Database Summary Paper Category List

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

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

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

The screenshot shows the Nucleic Acids Research website interface. At the top, the journal title "Nucleic Acids Research" is displayed in white on a maroon background. Below this, there are navigation links: "ABOUT THIS JOURNAL", "CONTACT THIS JOURNAL", "SUBSCRIPTIONS", and "CUI". The main content area features a breadcrumb trail: "Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper". The title "2015 NAR Database Summary Paper" is prominently displayed. Underneath, there are several blue hyperlinks: "Nucleotide Sequence Databases", "International Nucleotide Sequence Database Collaboration", "Coding and non-coding DNA", "Gene structure, introns and exons, splice sites", and "Transcriptional regulator sites and transcription factors". A list of 80 database names is shown in red text, starting with "3D-Footprint" and ending with "FlvTF". A blue bracket on the right side of the list indicates that it contains 80 entries.

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

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

**EPD**

Eukaryotic Promoter Database

Geneva, Switzerland

# EPD = Eukaryotic Promoter Database

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

(объединенный ресурс, включающий две базы)



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

The screenshot shows the EPD website interface. At the top, there is a search bar with a dropdown menu set to 'All databases' and a 'SEARCH' button. Below the search bar, there is a news section with a date '10-10-2016' and a title 'New liftOver option added to "Select / Download" page'. The main content area is divided into two sections: 'EPDnew' and 'EPD'. The 'EPDnew' section lists various organisms and their corresponding promoter counts, along with a list of animals, plants, and fungi. The 'EPD' section provides a brief description of the database and its contents.

**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**

**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.

**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 interface. At the top, there is a logo for SIB (Russian Science Center) and the EPD logo, which includes a rabbit and a DNA helix. The text 'EPD EUKARYOTIC PROMOTER DATABASE' is displayed. Below the logo, there is a navigation bar with 'Computational Cancer Genomics | ExPASy | EPFL'. A search bar is present with a text input field, 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 also features a section titled 'EPD database' with a descriptive paragraph, a note that the current version is based on 'EMBL 124', a 'Collection accessibility' section, and a 'Reference' section with a link to a detailed description of the database's principles.

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

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	<a href="#">Homo sapiens</a> (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	<a href="#">NT_032977.8</a> [3474400, -70361425] <a href="#">HapMap</a>
CLEANEX	<a href="#">HS_AK2</a>
DNA References	AL020995.14 [46401, -104596] <a href="#">ENA</a> <a href="#">GenBank</a> <a href="#">DDBJ</a>
SWISSPROT	<a href="#">P54819</a> KAD2_HUMAN
REFSEQ	<a href="#">NM_001625</a> [ <a href="#">DBTSS</a> ]
	<a href="#">NM_013411</a> [ <a href="#">DBTSS</a> ]
MIM	<a href="#">103020</a>
References	
[1]	MEDLINE= <a href="#">10521335</a> Strausberg RL., Feingold EA., Klausner RD., Collins FS. The mammalian gene collection Science 286:455-457(1999).
Promoter-specific information	
Sequence	tgtgagcggcagtgaggacgtgctggtgctgcttgacctgggaaGCACTGGACCT
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



# 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

The screenshot shows the EPDnew database website. The left sidebar lists species: H. sapiens, M. musculus, D. melanogaster, A. mellifera, D. rerio, C. elegans, A. thaliana, Z. mays, S. cerevisiae, and S. pombe. The main content area is titled 'EPDnew databases' and describes the project as a series of species-specific databases of experimentally validated promoters. It lists 10 supported species: 6 animals (H. sapiens, M. musculus, D. melanogaster, A. mellifera, C. elegans, and D. rerio), 2 plants (A. thaliana and Z. mays), and 2 fungus (S. cerevisiae and S. pombe). Below the text are icons for each species' database. The 'Collection accessibility' section explains four ways to access the data: (1) using input forms, (2) searching by gene symbol or IDs, (3) downloading promoters by genomic context, and (4) bulk download via FTP. A reference is provided at the bottom.

# 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 <sup>(12)</sup>	148	3841
<i>M. musculus</i>	CAGE from FANTOM5 ( <a href="http://fantom.gsc.riken.jp/5/">http://fantom.gsc.riken.jp/5/</a> )	339	6236
<i>D. melanogaster</i>	CAGE from modENCODE ( <a href="ftp://data.modencode.org/">ftp://data.modencode.org/</a> ) TSS-seq from Machibase <sup>(13)</sup>	57	646
<i>D. rerio</i>	CAGE from Nepal et al. <sup>(14)</sup> , downloaded from SRA <sup>(8)</sup> , ID SRA055273	12	65
<i>C. elegans</i>	GRO-cap from Kruesi et al. <sup>(15)</sup>	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 Dreos<sup>1</sup>, Giovanna Ambrosini<sup>1,2</sup>, Rouayda Cavin Périer<sup>2</sup> and Philipp Bucher<sup>1,2,\*</sup>

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:

**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**

**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**

**Text view** | **Genome view**

**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: cctctcggcgcgccccctcctcgcgccgccccgcccgcctcAGTCTGGCAGG  
Position in the genome: Chromosome [NC\_000022.10]; Strand [-]; Position [22221935]  
Ensembl: [ENSG00000100030](#)  
RefSeq: [NM\\_002745](#)  
NCBI Gene: [MAPK1](#)  
GeneCards: [MAPK1](#)

**External resources:**

EPD view at UCSC (hg19): [chr22:22219935-22223935](#)  
EPD view at UCSC (hg18): [chr22:20549935-20553935](#)  
SwissRegulon: [chr22:22219935-22223935](#)  
FANTOM4: [chr22:22219935-22223935](#)  
MPromDB: [chr22:22219935-22223935](#)

**Promoter image:**

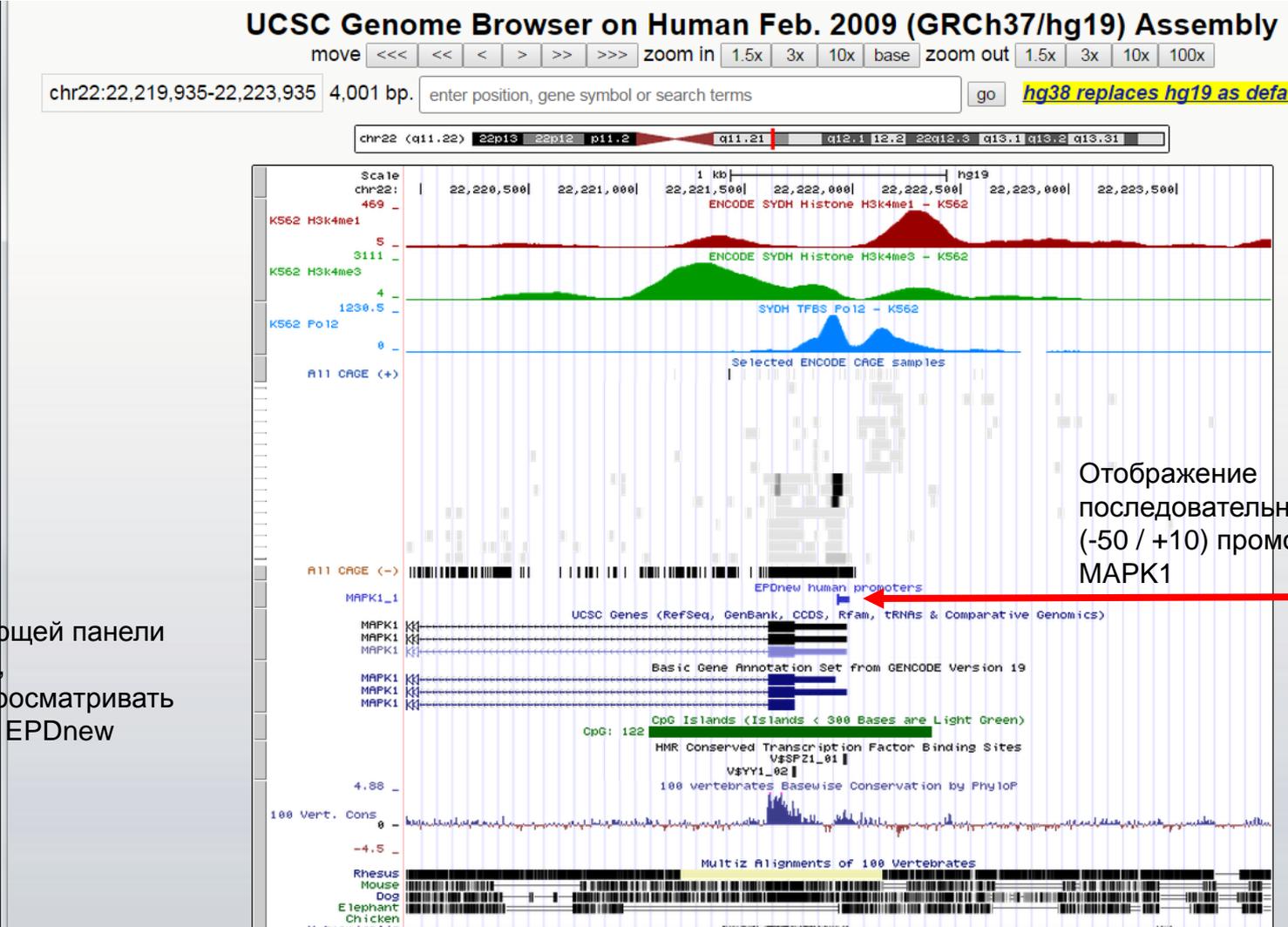
Scale chr22: 22,220,500 | 22,221,000 | 1 kb | 22,222,000 | 22,222,500 | hg19 | 22,223,000 | 22,223,500

K562 H3K4me1  
K562 H3K4me3  
K562 Po12  
Selected ENCODE ChIP samples

A11 CAGE (+)  
A11 CAGE (-)

MAPK1\_1  
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)  
CpG Islands (Islands < 300 Bases are Light Green)  
HMR Conserved Transcription Factor Binding Sites  
V\$P21\_01  
V\$Y1\_02  
100 vertebrates Basewise Conservation by PhyloP

# 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

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) задукомментированным 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 as 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

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

Submit

### Sequence Extraction Tool (FASTA format)

From:

To:

Submit

### Downstream Analysis

Motif Enrichment

?

Motif Discovery

?

Chromatin analysis

?

Конец 4-ой лекции