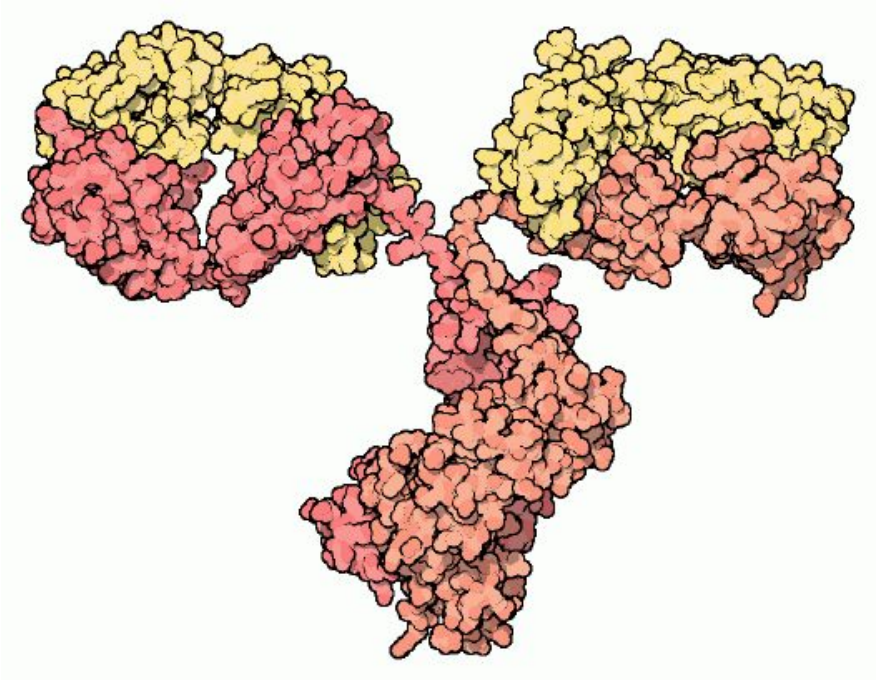


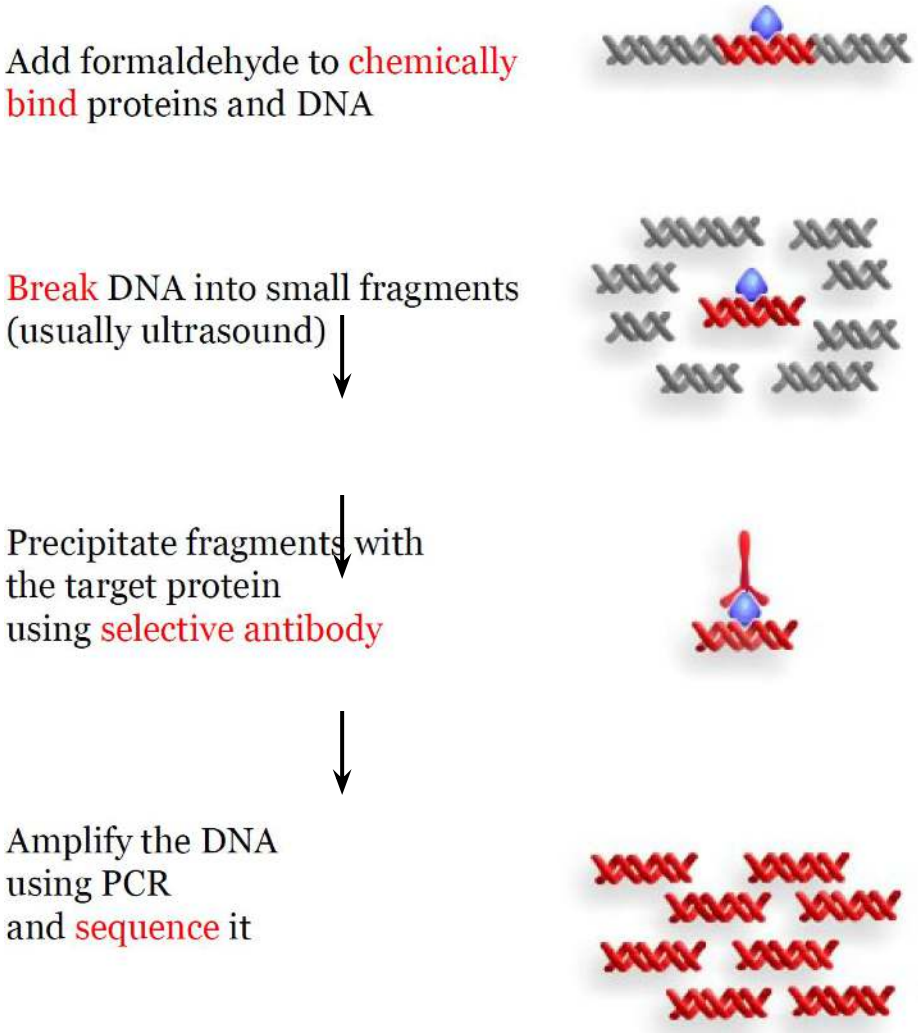
Лекция 3:

Практическое применение ChIP-Seq и родственных методов

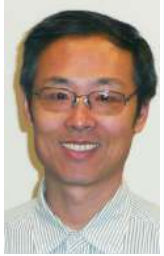
Александр Предеус
Институт Биоинформатики

ChIP-Seq

- Протеины сшиваются с ДНК формальдегидом
- Фрагментация, осаждение при помощи селективных антител
- Расшивание
- ПЦР и секвенирование выделенных фрагментов



Пьедестал забега за ChIP-Seq



Science. 2007 Jun 8;316(5830):1497-502. Epub 2007 May 31.

Genome-wide mapping of in vivo protein-DNA interactions.

Johnson DS[1], Mortazavi A, Myers RM, Wold B.

Author Information

Abstract

In vivo protein-DNA interactions connect each transcription factor with its direct targets to form a gene network scaffold. To map these protein-DNA interactions comprehensively across entire mammalian genomes, we developed a large-scale chromatin immunoprecipitation assay (ChIP-Seq) based on direct ultrahigh-throughput DNA sequencing. This sequence census method was then used to map in vivo binding of the neuron-restrictive silencer factor (NRSF, also known as REST, for repressor element-1 silencing transcription factor) to 1946 locations in the human genome. The data display sharp resolution of binding position (+/-50 base pairs (bp)), which facilitated our finding motifs and allowed us to identify noncanonical NRSF-binding motifs. These ChIP-Seq data also have high sensitivity and specificity (ROC (receiver operator characteristic) area >/= 0.95) and statistical confidence (P <10(-4)), properties that were important for inferring new candidate interactions. These include key transcription factors in the gene network that regulates pancreatic islet cell development.

Cell

Volume 129, Issue 4, 18 May 2007, Pages 823–837

Resource

High-Resolution Profiling of Histone Methylations in the Human Genome

Artem Barski[1, 3], Suresh Cuddapah[1, 3], Kairong Cui[1, 3], Tae-Young Roh[1, 3], Dustin E. Schones[1, 3], Zhibin Wang[1, 3], Gang Wei[1, 3], Iouri Chepelev[2], Keji Zhao[1, *]

nature

ARTICLES

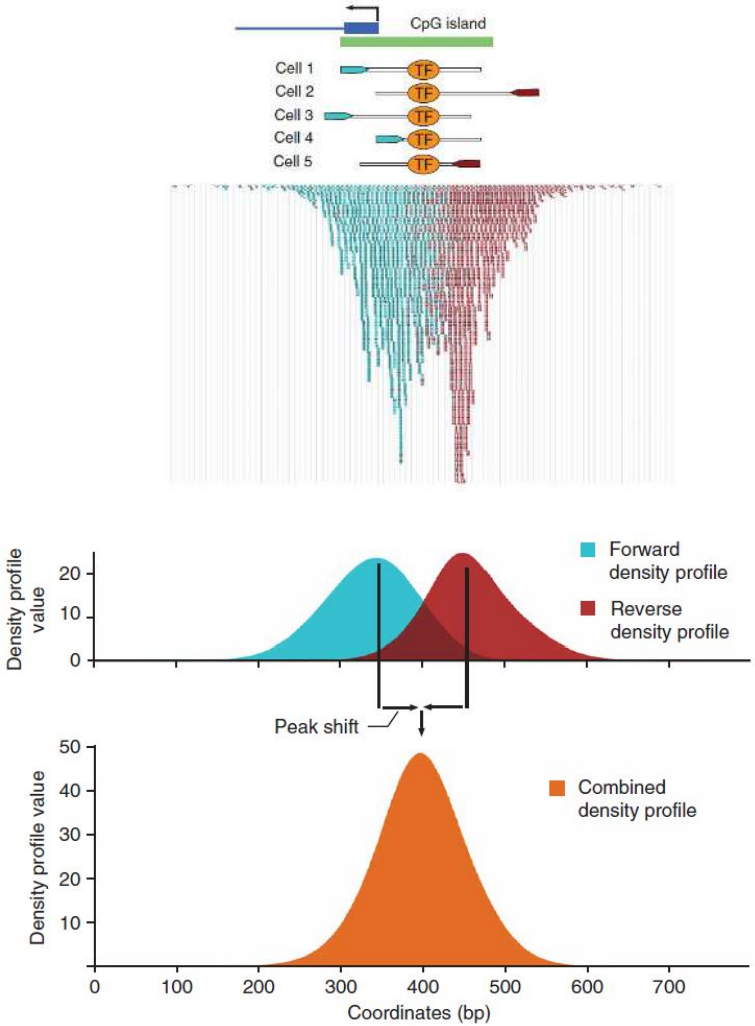
Genome-wide maps of chromatin state in pluripotent and lineage-committed cells

Tarjei S. Mikkelsen[1,2], Manching Ku[1,4], David B. Jaffe[1], Biju Issac[1,4], Erez Lieberman[1,2], Georgia Giannoukos[1], Pablo Alvarez[1], William Brockman[1], Tae-Kyung Kim[2], Richard P. Koche[1,2,4], William Lee[1], Eric Mendenhall[1,4], Aisling O'Donovan[4], Aviva Presser[1], Carsten Russ[1], Xiaohui Xie[1], Alexander Meissner[3], Marius Wernig[3], Rudolf Jaenisch[3], Chad Nusbaum[1], Eric S. Lander[1,5*] & Bradley E. Bernstein[1,4,6*]

http://www.snarkyscientist.com/2013/06/19/the-history-of-the-biggest-technique-of-2009-who-invented-chip-seq/

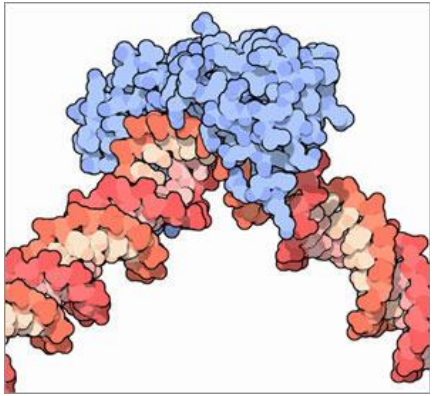
Анализ ChIP-Seq

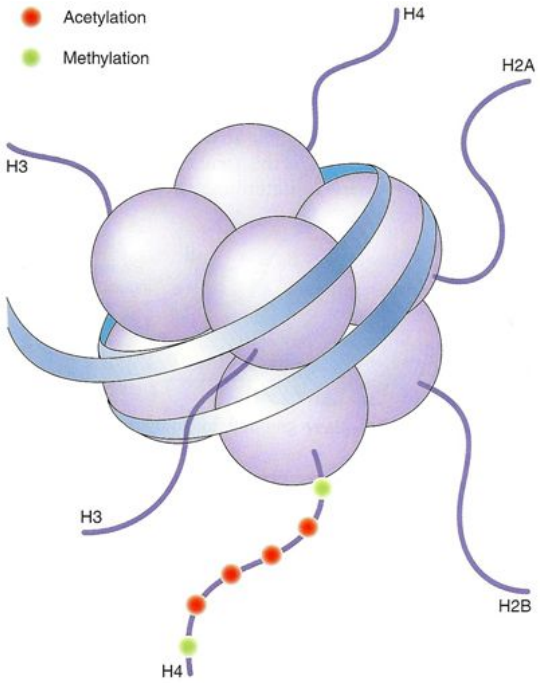
- Полученные данные выравниваются на геном
- Пики идентифицируются при помощи статистических алгоритмов (peak callers)
- Дальнейший анализ сильно зависит от задачи



Задачи ChIP-Seq

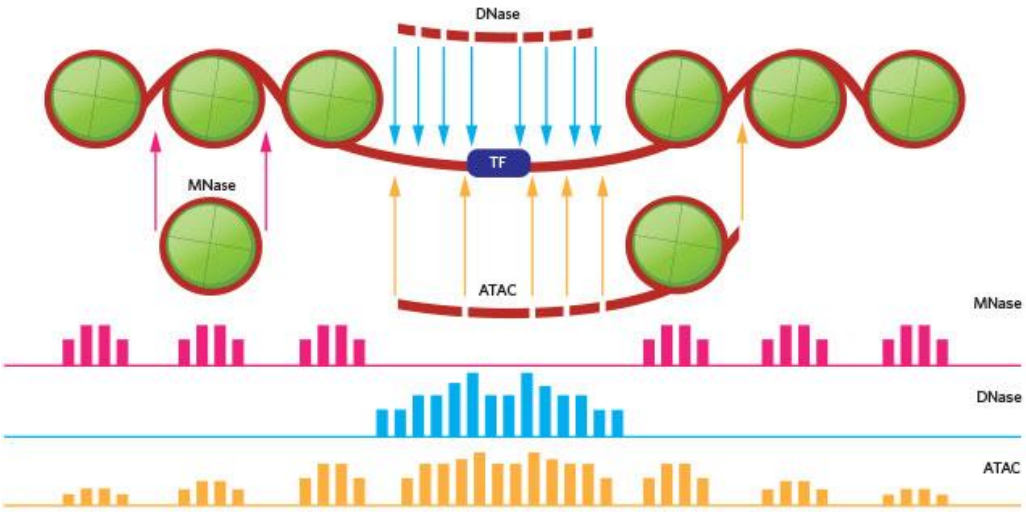
- Два глобальных типа применения ChIP-Seq:
  - транскрипционные факторы
  - гистонные модификации
- Анализ результатов для типов 1 и 2 радикально отличается





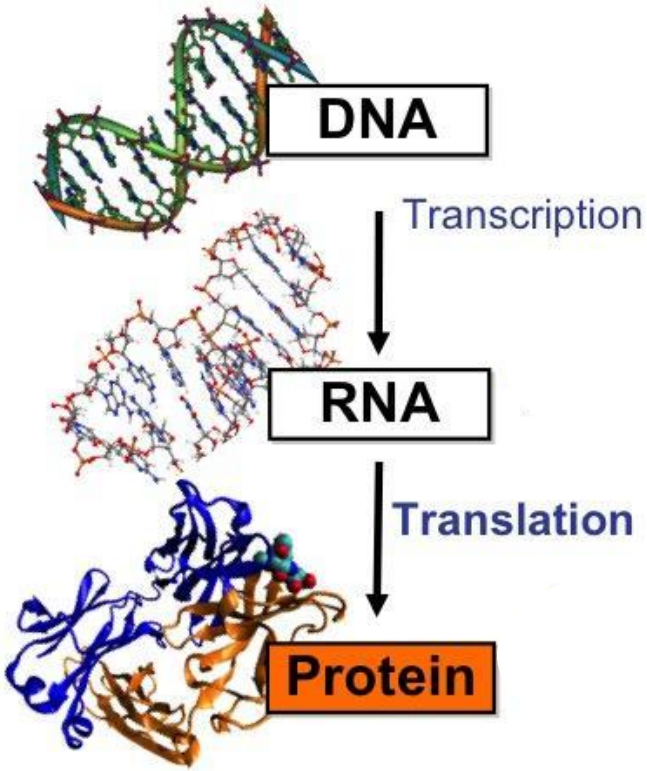
Genome footprinting

- Поймать наиболее развернутые участки хроматина
- DNase-seq, FAIRE-seq, ATAC-seq



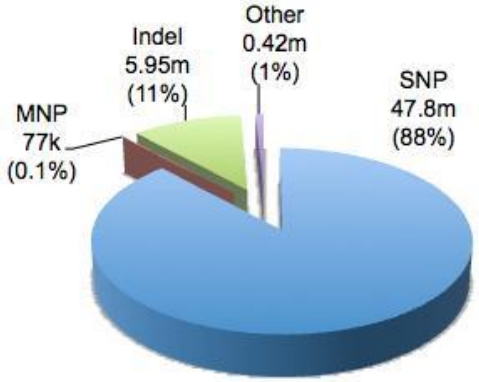
Что изучается ChIP-seq?

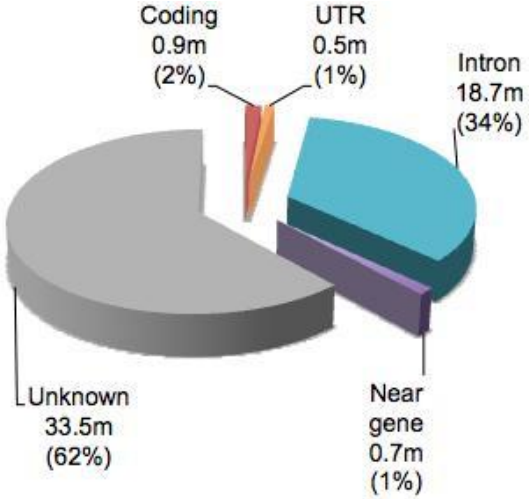
- Методы ChIP-Seq и DNase-Seq помогают понять механизмы регулирования (как правило, экспрессии генов)



А зачем изучать регуляцию?

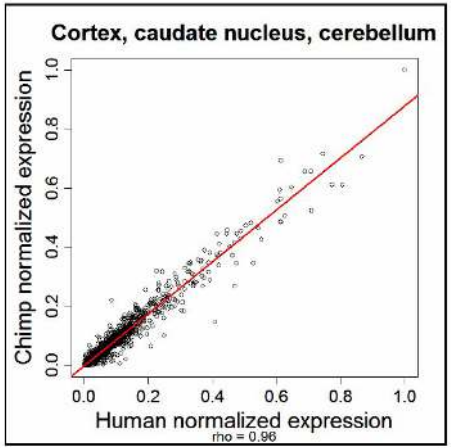
- Практически говоря, регуляцию починить реальнее, чем существующий ДНК
- Огромный процент известных SNP находится в некодирующих последовательностях

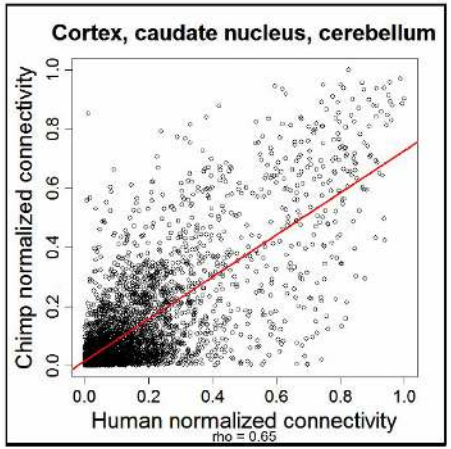


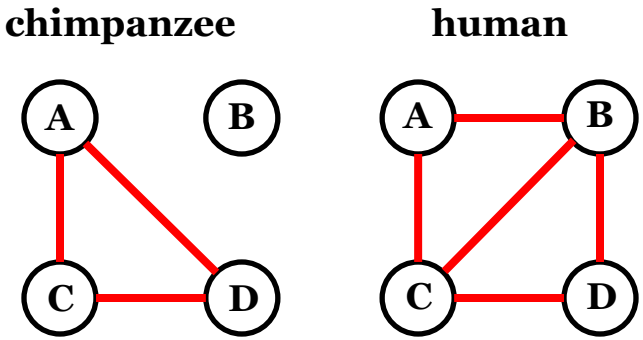


Мозг человека и примата

- структура и уровни экспрессии генов практически совпадают
- регуляция - отличается очень сильно!



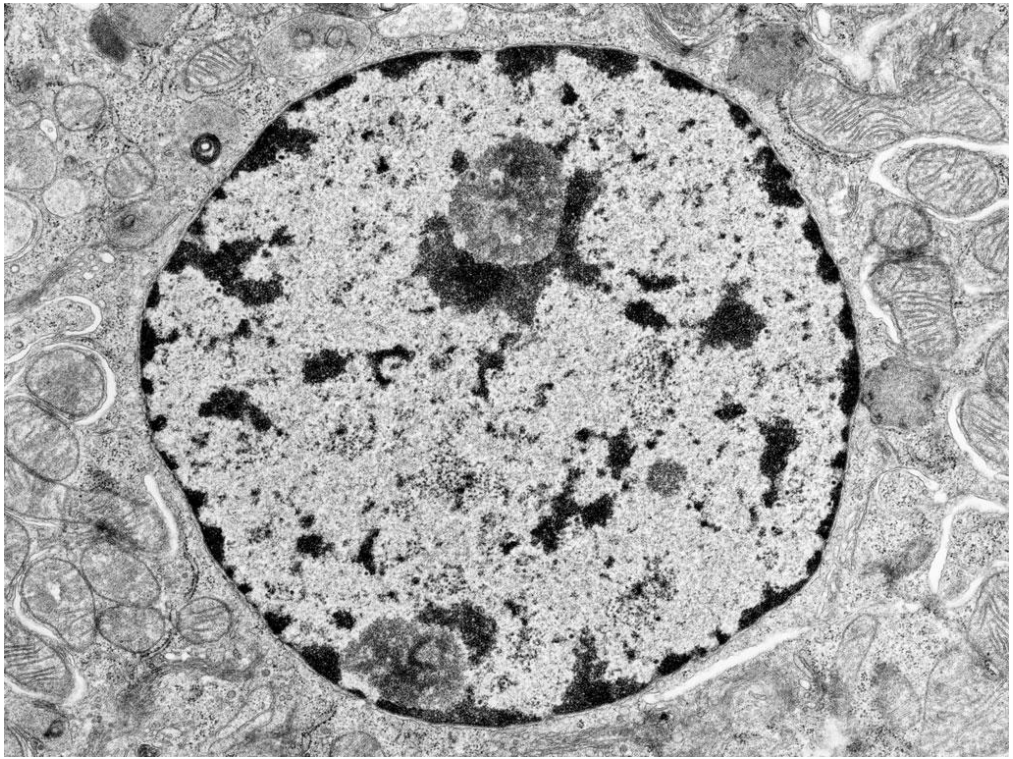




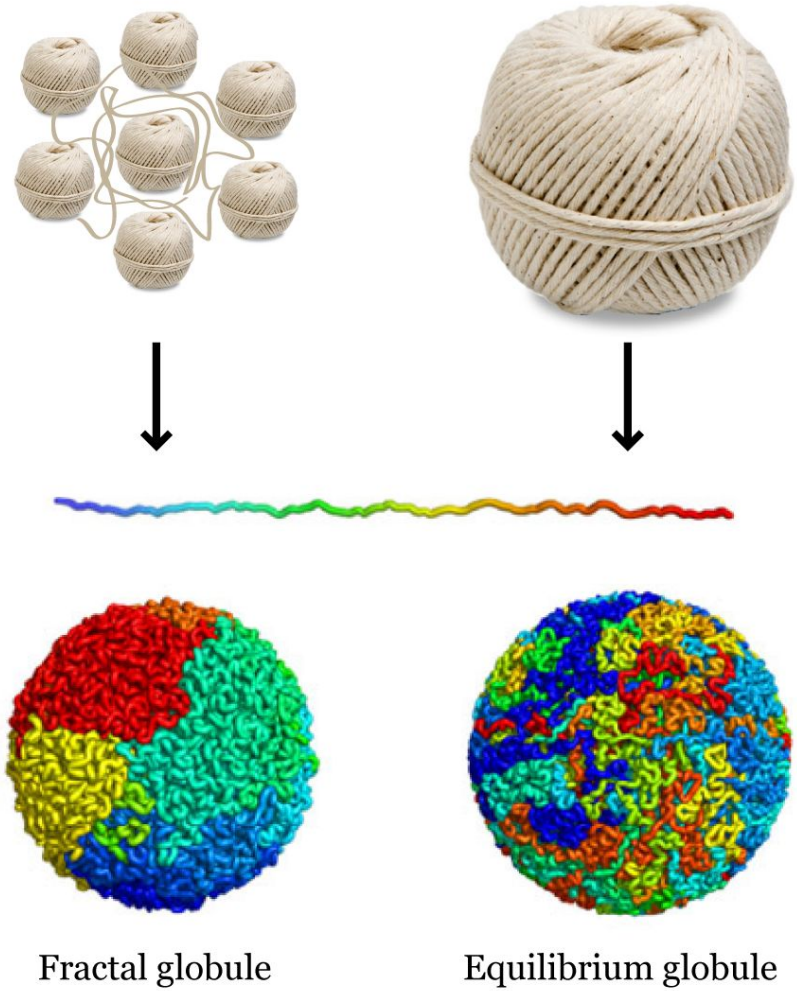
Proc Natl Acad Sci U S A. 2006 Nov 21;103(47):17973-8.

Хромосомы и хроматин

- Хроматин = ДНК + протеины
- общая длина ДНК - около 2 м, нужна упаковка!
- Типы хроматина - эухроматин (открытый, активный) и гетерохроматин (плотный, как правило неактивный)

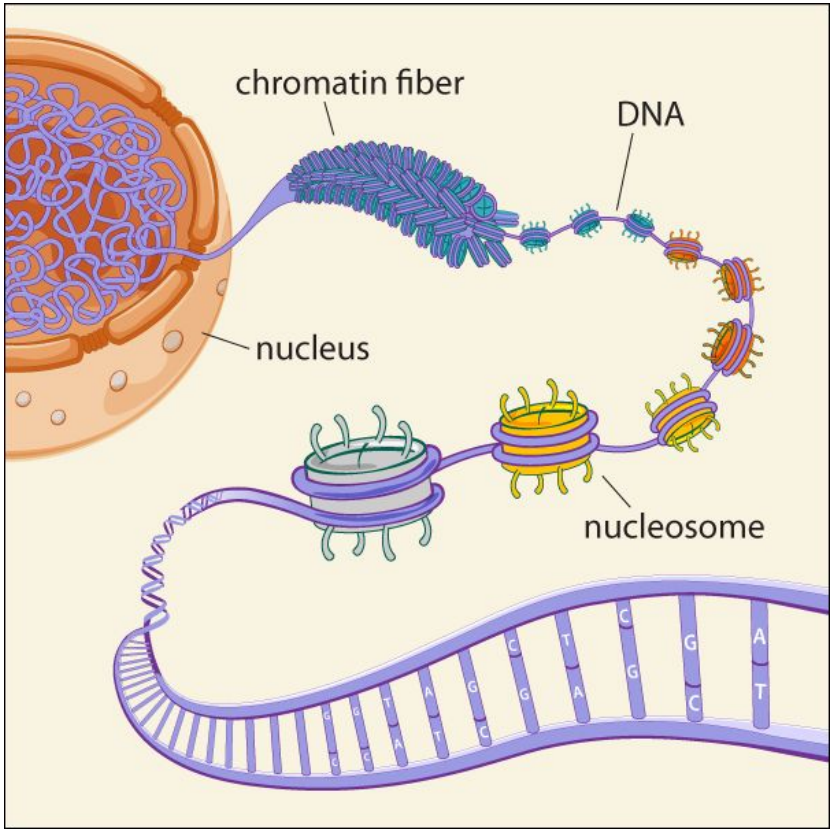


Модель фрактальной глобулы

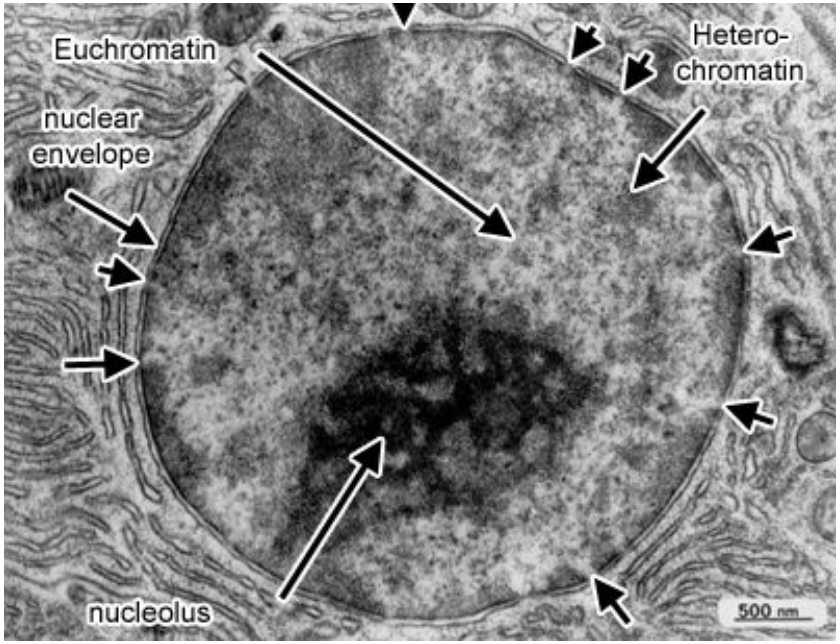


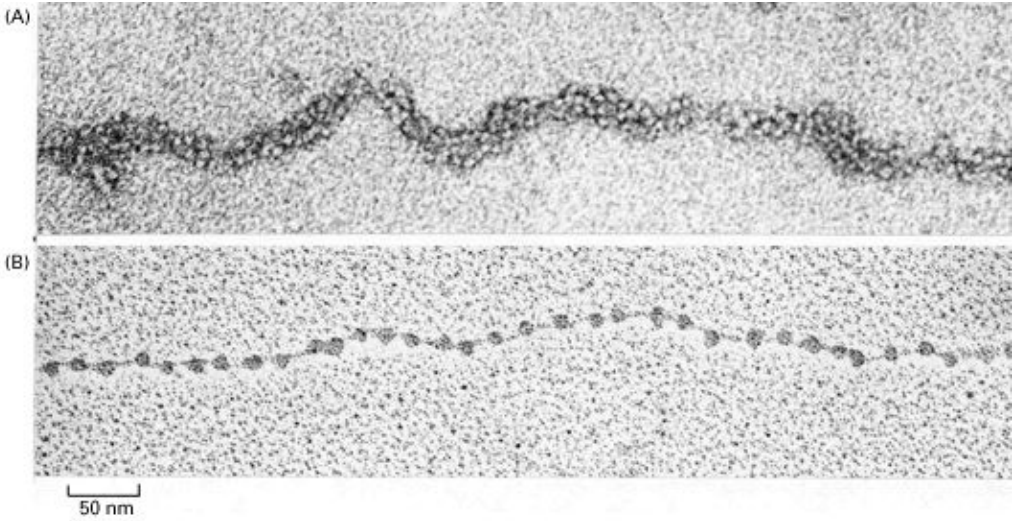
Уровни организации хроматина

- Три главных уровня организации хроматина:
  - «бусы на нитке» - нуклеосомы
  - 30нм фибрилы
  - плотно упакованный хроматин (гетерохроматин)
- Нуклеосомы включают в себя гистонные протеины



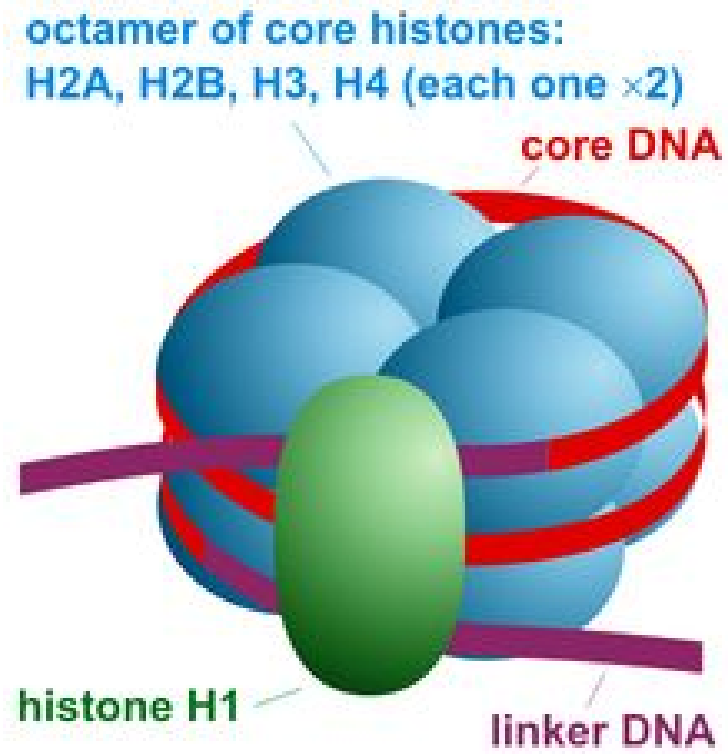
Хроматин под микроскопом





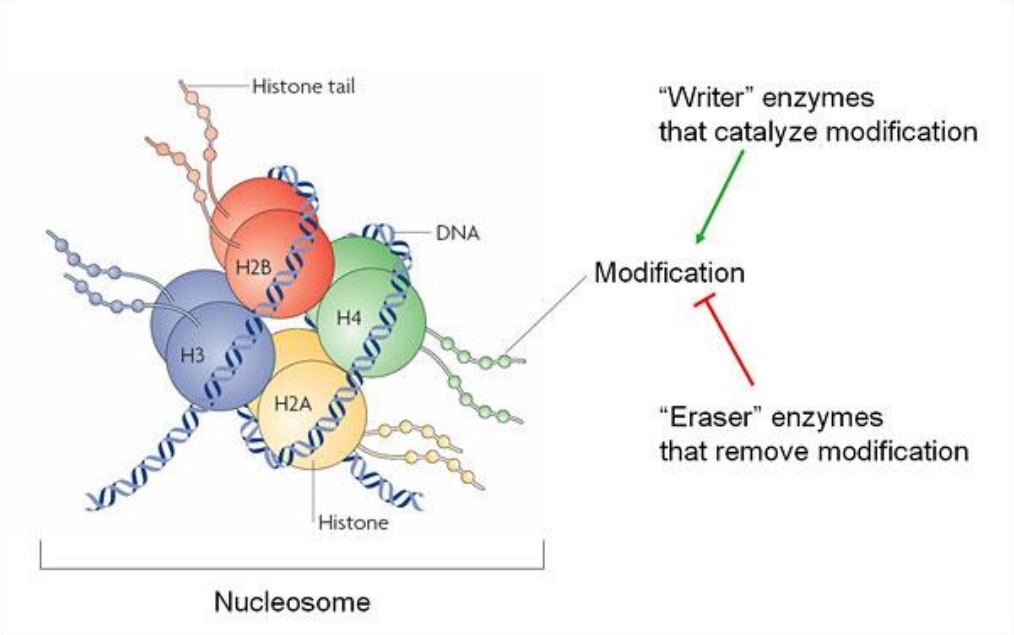
Строение нуклеосомы

- Нуклеосома включает в себя 8 протеинов, по 2 каждого из 4х типов (H2A, H2B, H3, H4)
- Гистон H1 участвует в построении фибрил
- ДНК почти никогда не существует свободно от гистонов, но бывает ими обеднена



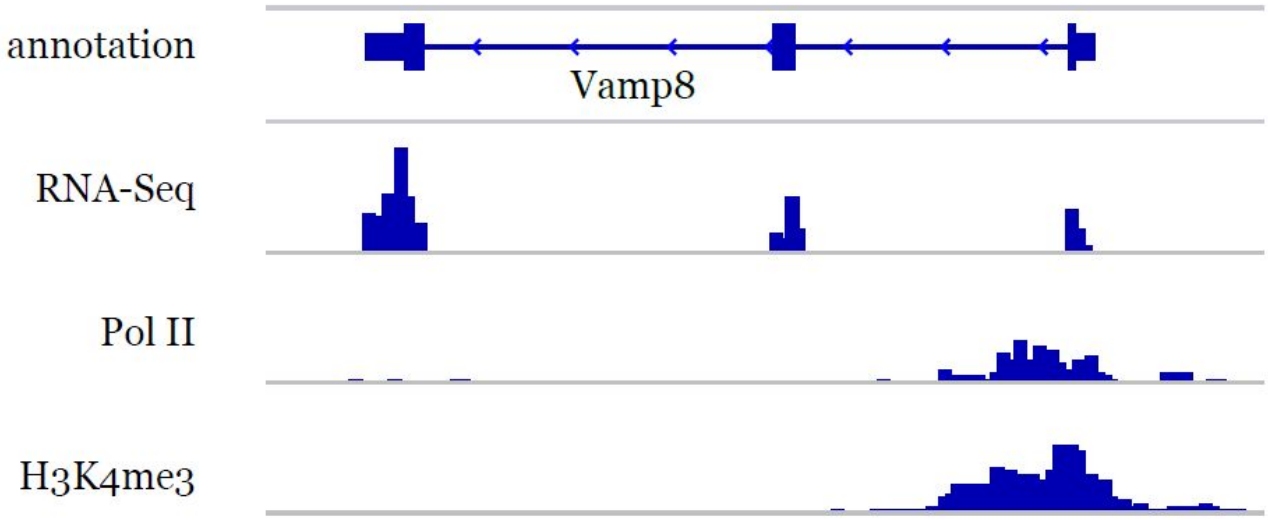
Модификации гистонов

- у гистонов есть хвосты!
- хвосты часто оказываются модифицированными химически (ПТМ)
- модификации меняют лабильность нуклеосом и распознаются протеинами



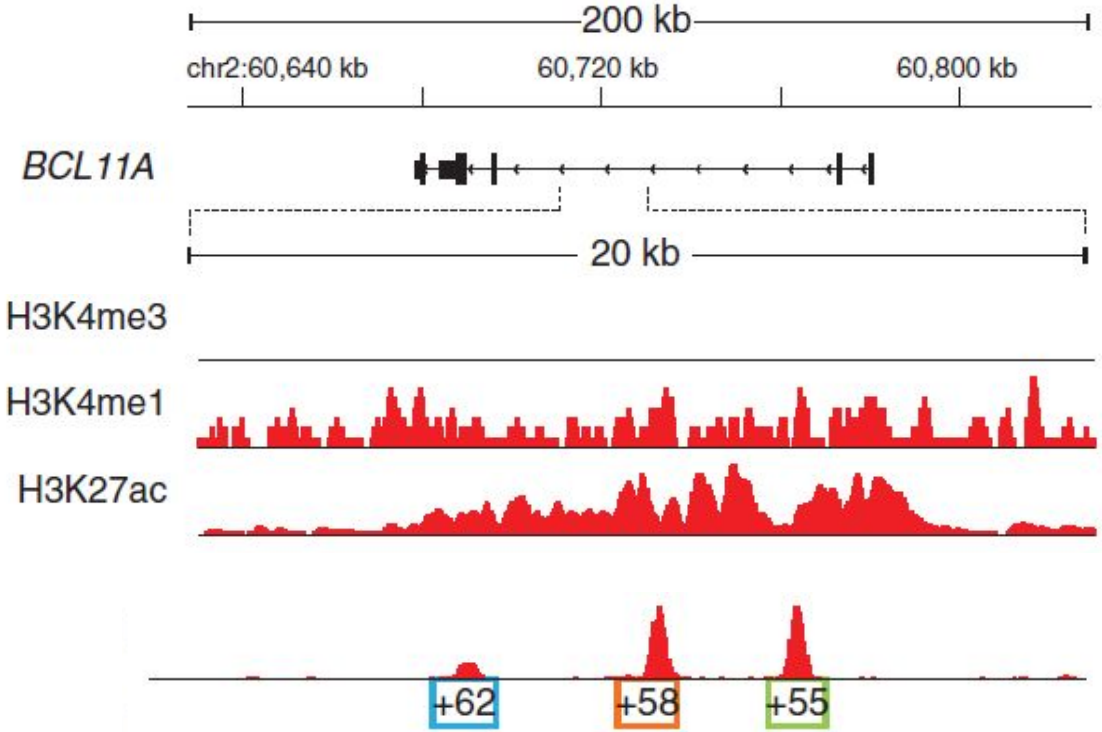
Марки промотеров: H3K4me3

- четкий, сильно выраженный пик



Модификации энхансеров

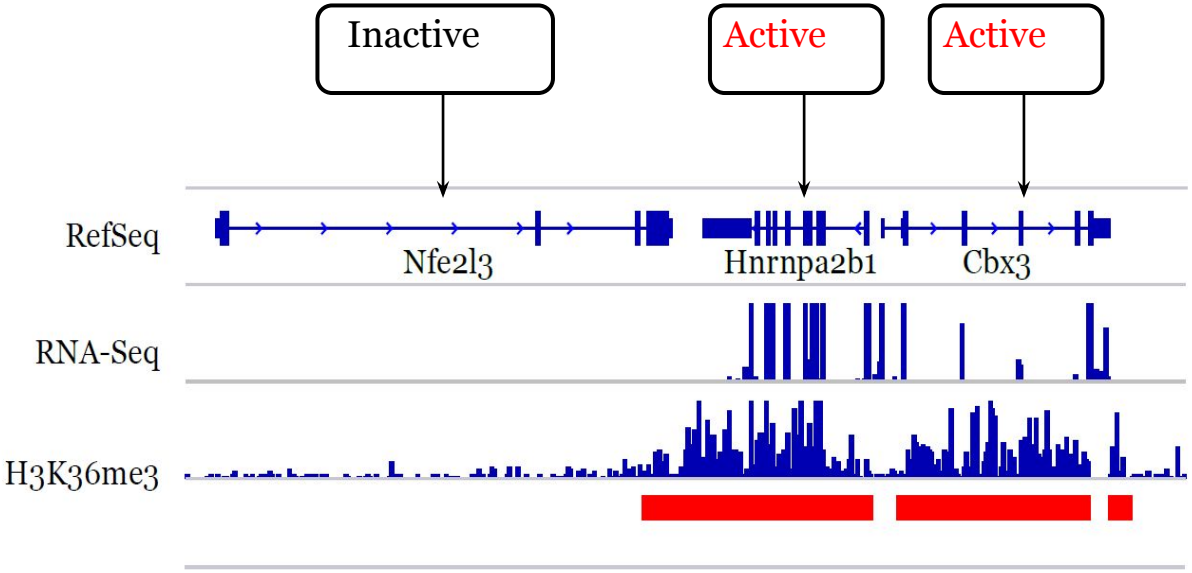
- Энхансеры ассоциированы с H3K4me1 и H3K27ac
- считается что H3K27ac - метка активно используемого энхансера



Метки транскрипции

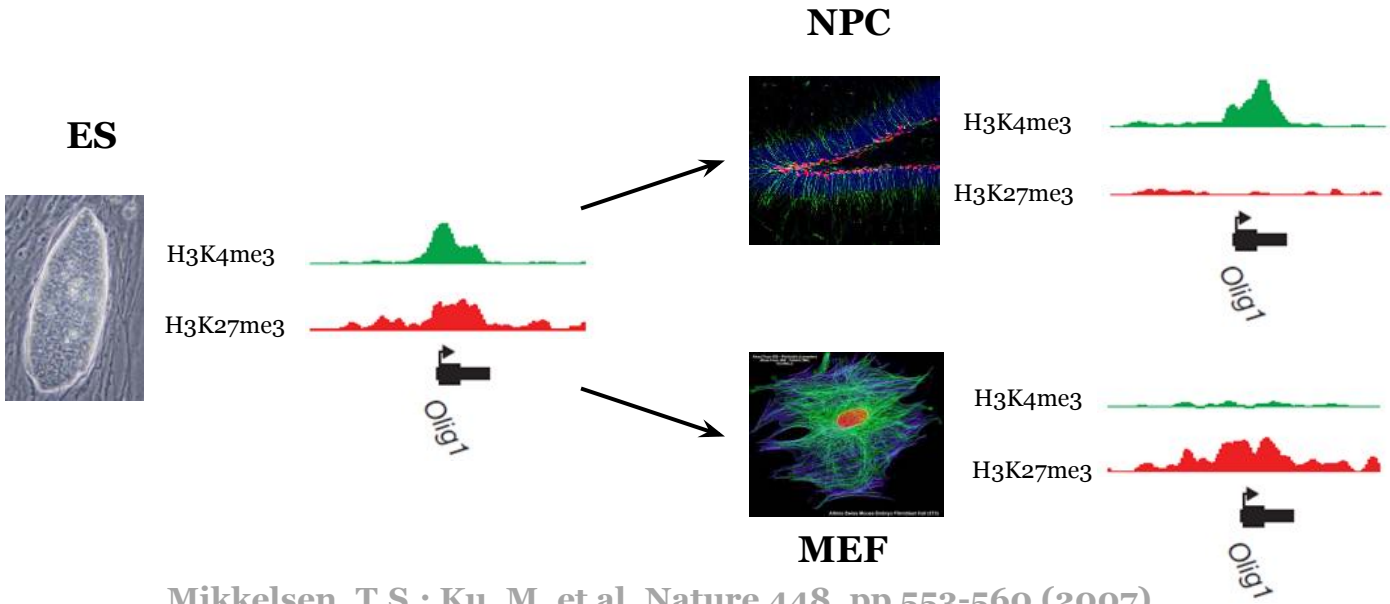
- Элонгация - H3K36me3 and H3K79me2/3

Mouse proB cells



Обратимая репрессия хроматина

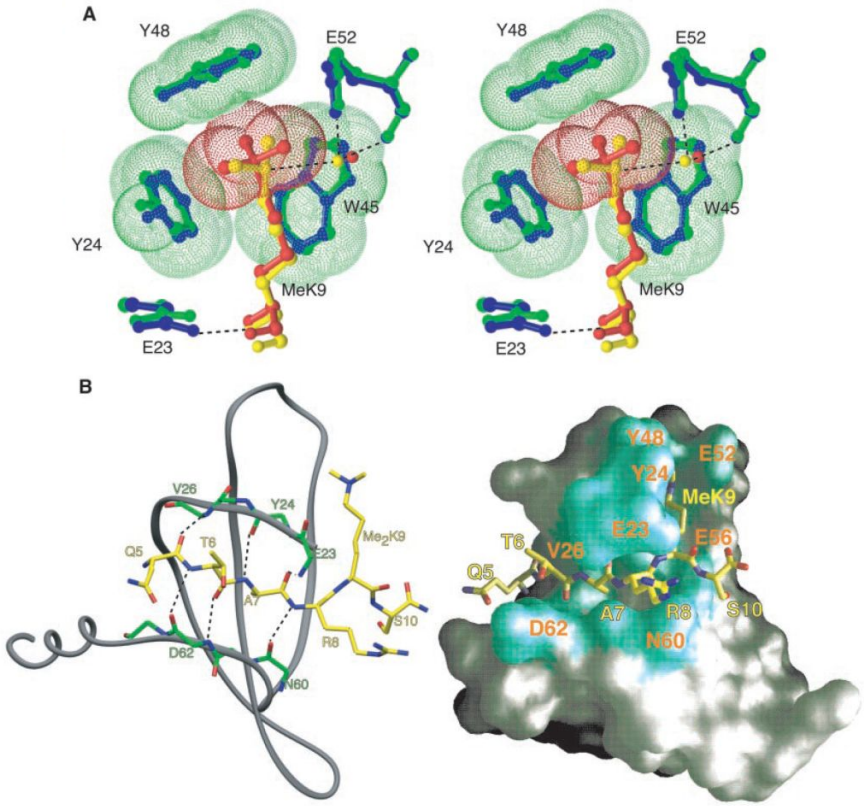
- H3K27me3 - марка неактивных генов, которая считается относительно обратимой



Mikkelsen, T.S.; Ku, M. et al, Nature 448, pp 553-560 (2007)
Vastenhouw N.L.; Zhang Y. et al, Nature 464, pp 922-6 (2010)

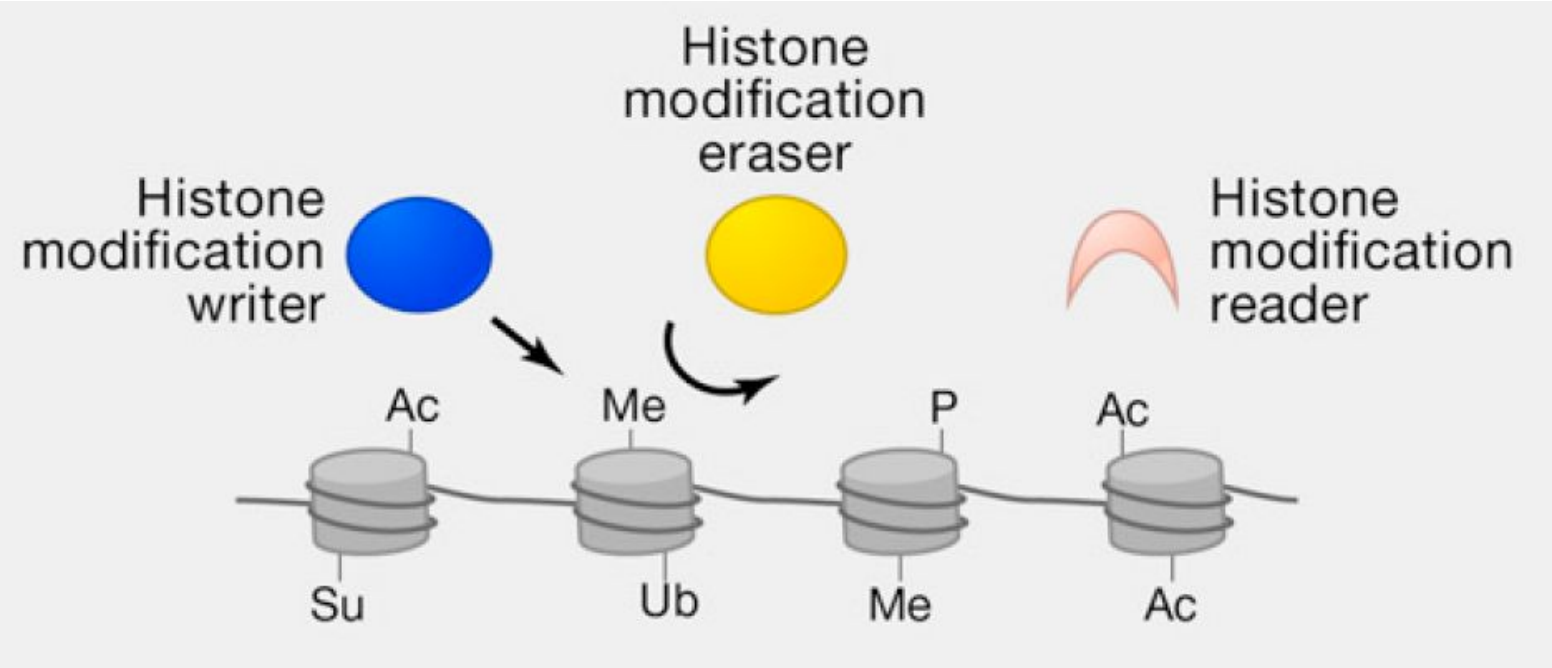
Марки гетерохроматина

- Марки H3K9me2\3 ассоциированы с перманентным гетерохроматином
- Связываются с протеином HP1



Jacobs, S.A.; Khorasanizadeh, S., Science, 295(5562), pp 2080-3 (2002)

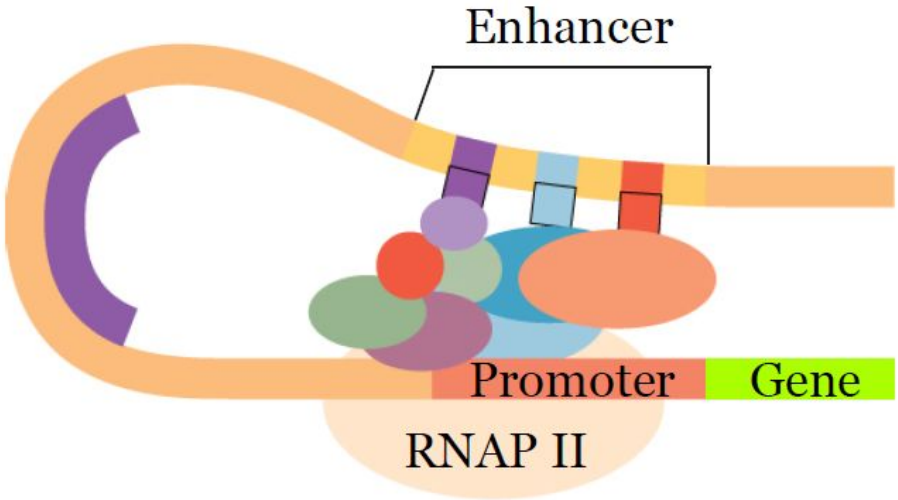
Общая роль модификаций



Lee, T.I.; Young, R.A., Cell, 152(6), pp 1237-51 (2013)

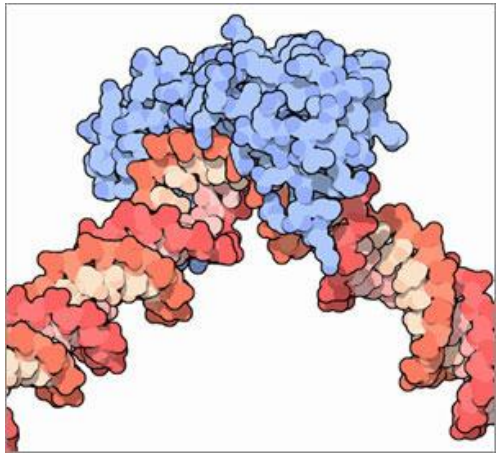
Регуляция транскрипции

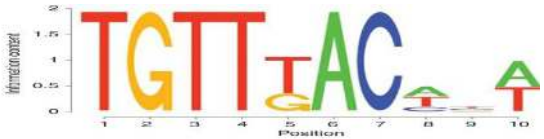
- Базальная транскрипция - GTF + Pol II
- Регуляция - энхансеры и хроматин
- Промотеры - очень близко к началу гена, энхансеры могут быть далеко



Транскрипционные факторы

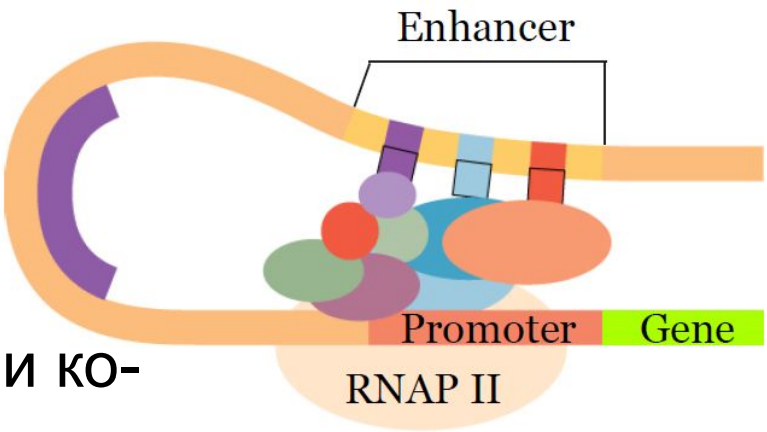
- Около 1,500-2000 ТФ в геноме человека
- Мотив определяет связывание данного ТФ с ДНК
- Для данного типа клеток часто можно определить т.н. master regulators, но они не всегда известны





ТФ и модуляция экспрессии

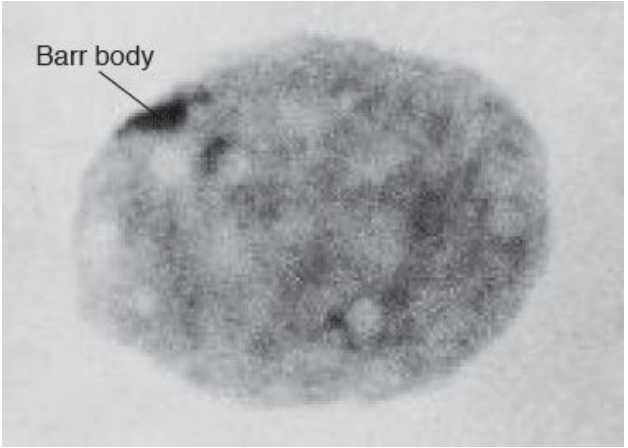
- Механизмы модуляции экспрессии ТФ:
  - стабилизация или блокирование взаимодействия Pol II с ДНК
  - ацетилирование или де-ацетилирование гистонов
  - связывание с ко-репрессором или ко-активатором



Если не ты, то кот?

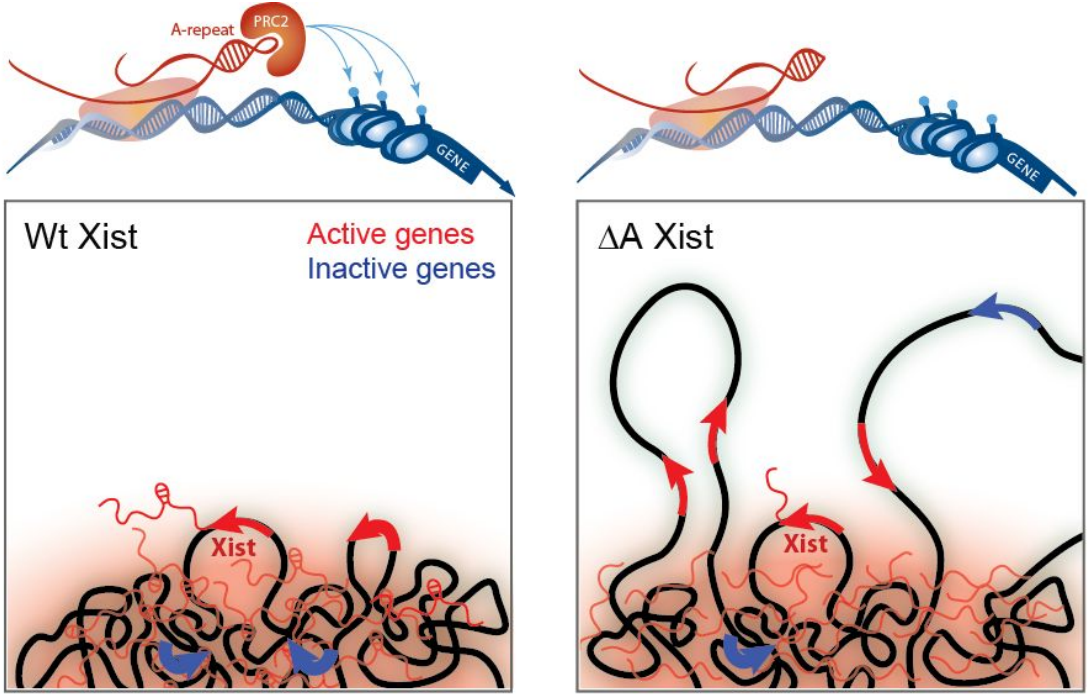
lincRNAs и X-инактивация

- у самок млекопитающих одна из X хромосом инактивируется
- неактивная X-хромосома известна как тельце Барра
- в мышах инактивируется X отца, во многих других видах - случайно



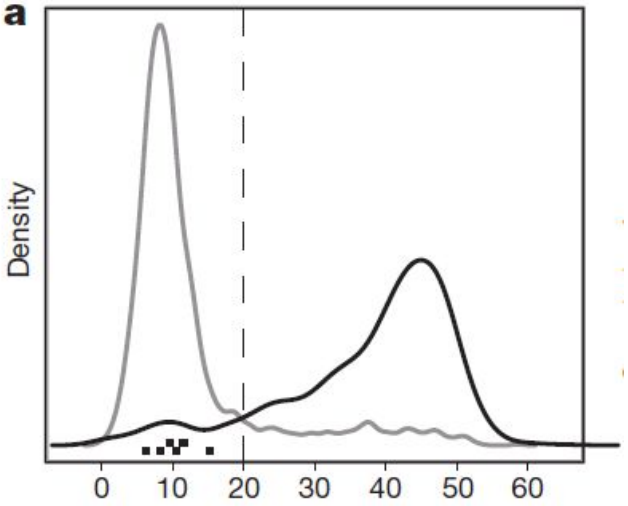
XIST

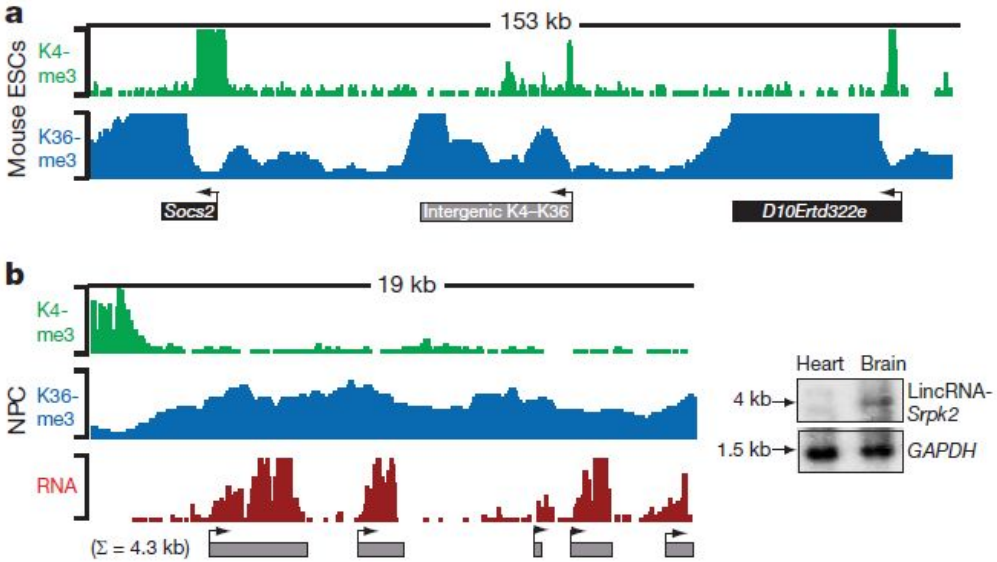
- Xist известен тем, что нужен для X-инактивации



Сколько же всего lincRNA?

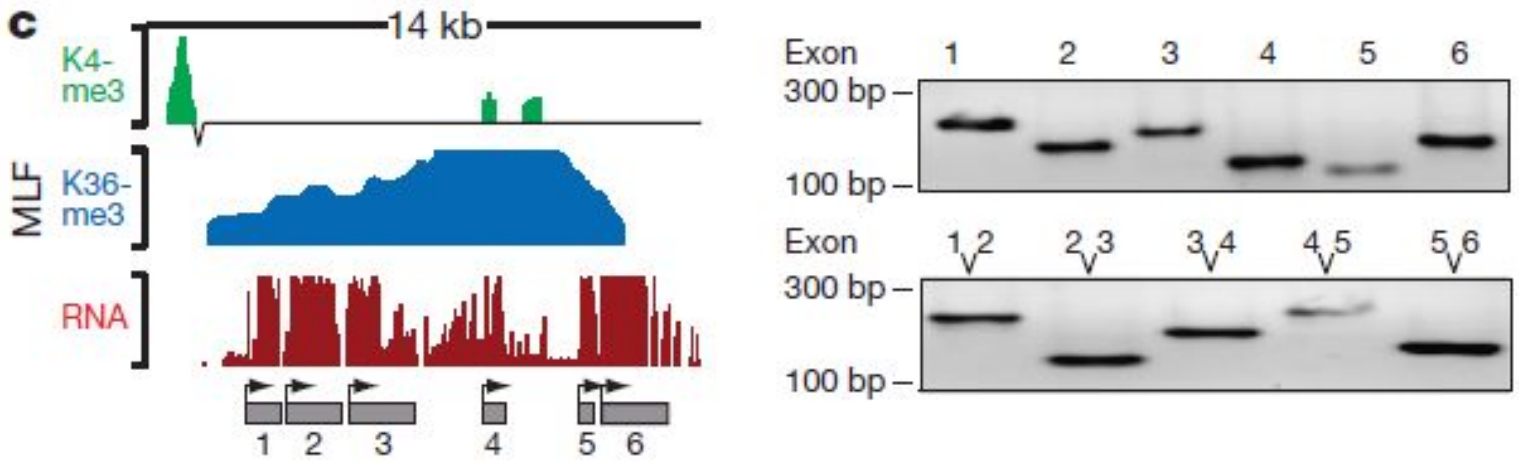
- в 2008-м году их считали артефактами
- .. а потом сделали ChIP-Seq на пару полезных гистонных модификаций





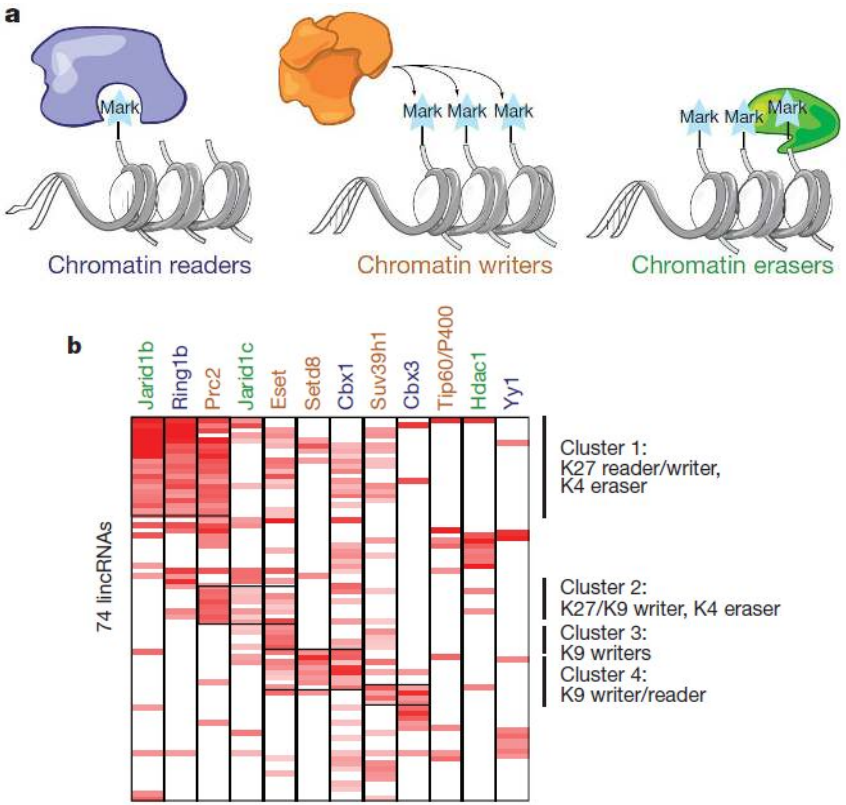
Подтверждение наличия lincRNAs

- RT-PCR подтвердили наличие более 10,000 нетранслируемых транскриптов!!!



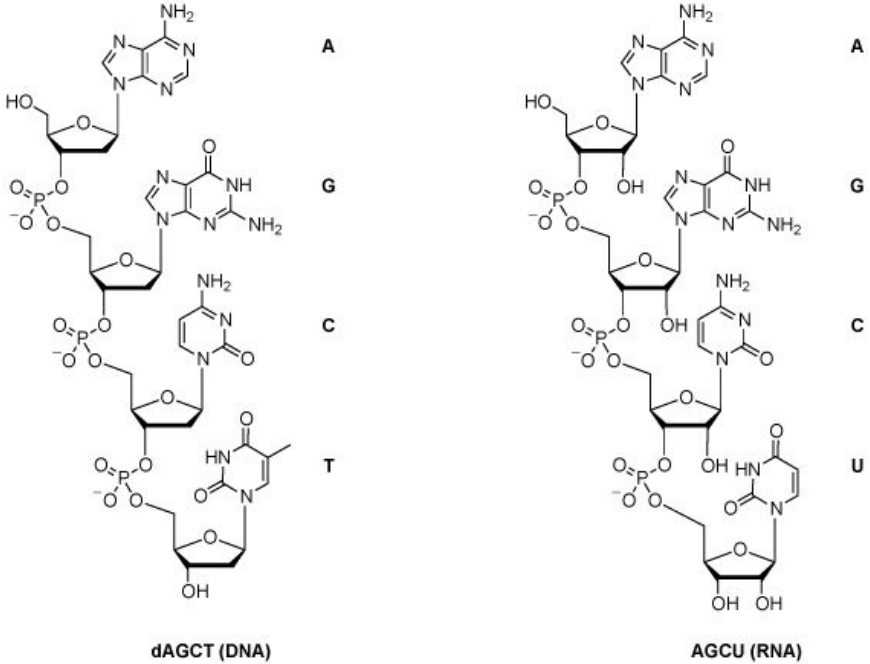
Есть ли функция у lincRNA?

- LincRNA практически неконсервативны
- Однако заведомо функциональные lincRNAs (Air, Xist) тоже неконсервативны!



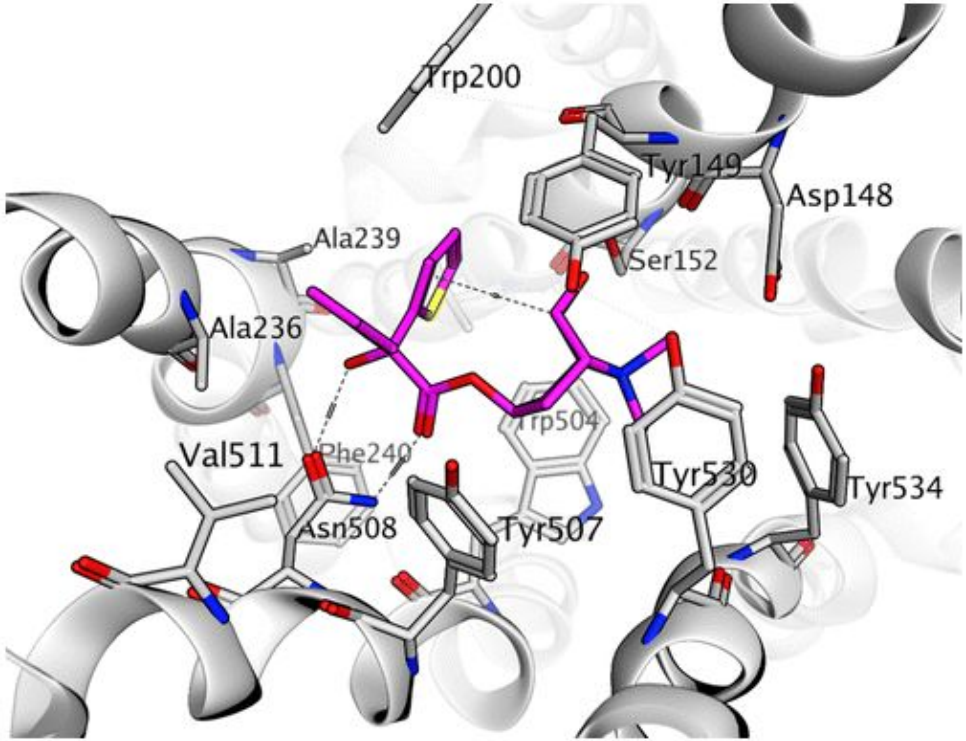
Медицина и транскрипция

- на транскрипционные факторы почти невозможно влиять лекарствами



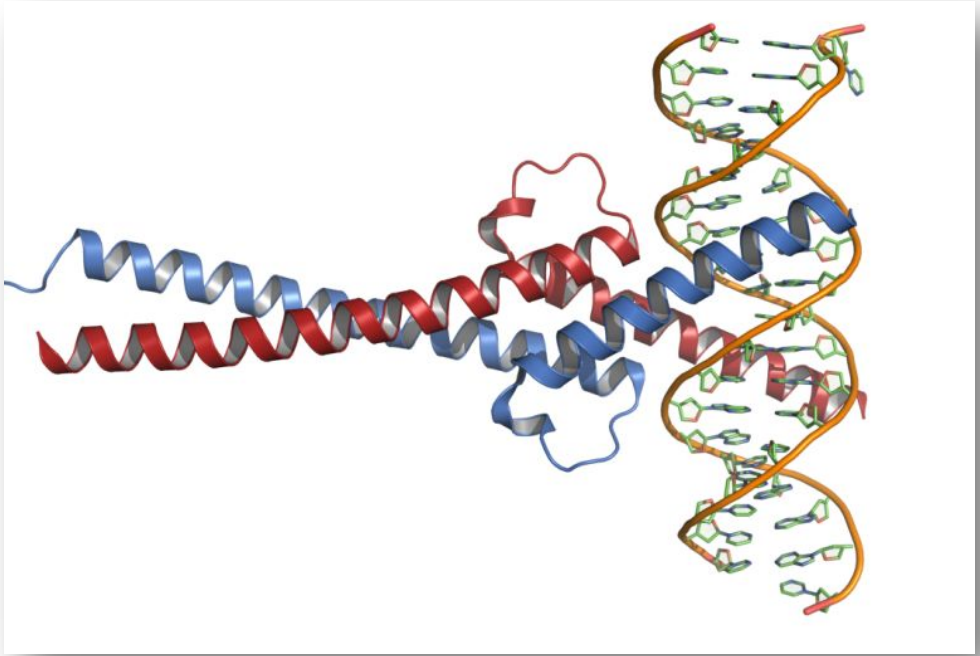
Существующие лекарства

- Большинство существующих лекарств связывается с небольшими полостями на поверхности энзимов и других протеинов



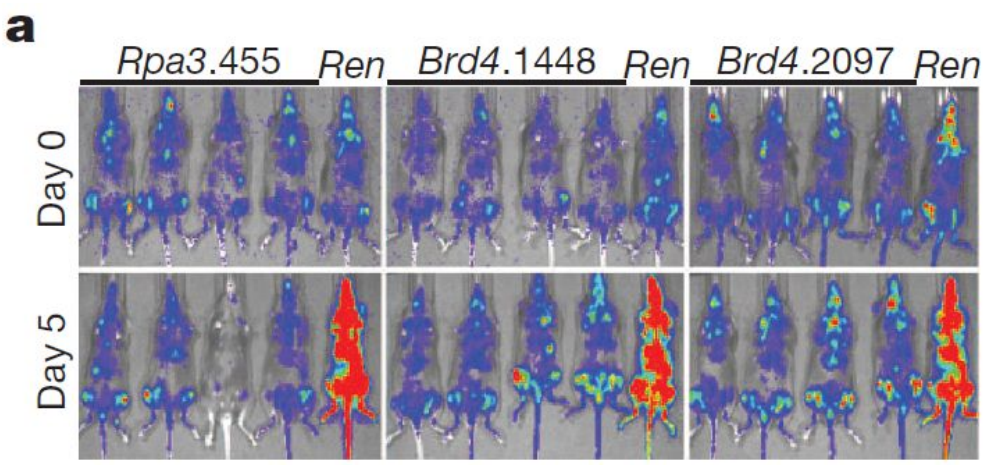
Пример: Myc

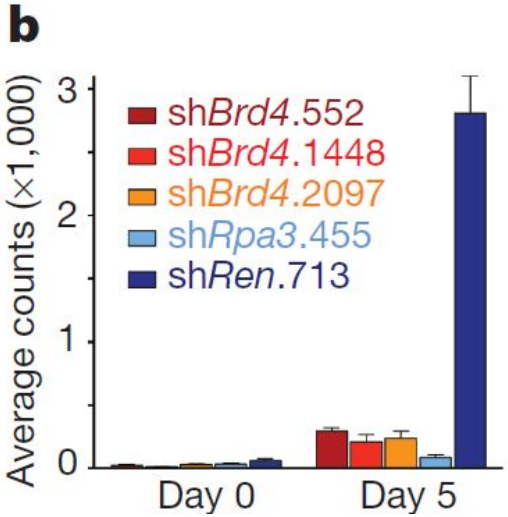
- MYC регулирует до 15% человеческих генов
- MYC называли «святым Граалем» раковой терапии в 1985-м году
- с тех пор терапия по MYC «не стала ближе ни на шаг»



Идентификация BRD4

- BRD4 был идентифицирован как критический для рака (AML) в мышах и shRNA-культурах



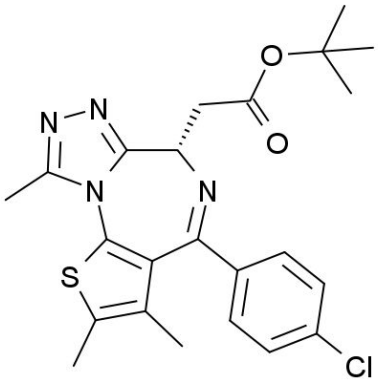


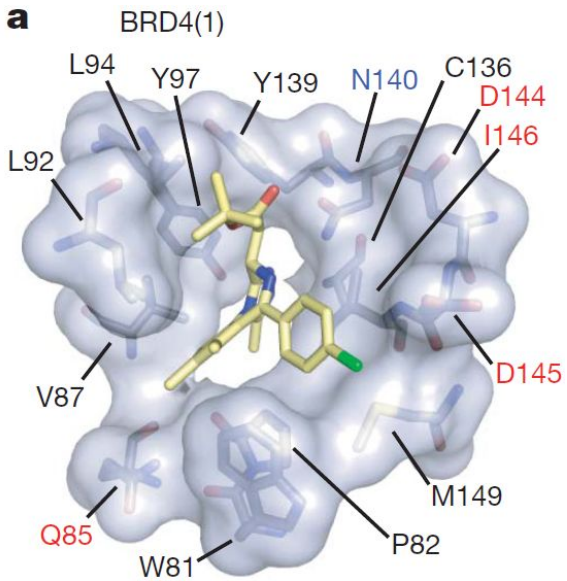
JQ1

- JQ1 был разработан как мужской контрацептив (?!)
- Одно из первых проявлений “веяния”, в котором фарма делится кандидатами в лекарства с учеными

История JQ1

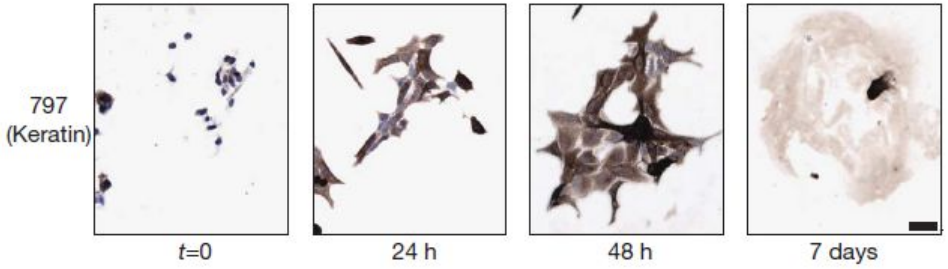
- JQ1 связывается с бромодоменом
- Бромодомен распознает ацетилированные хвосты гистонов
- JQ1 ингибирует BRD4

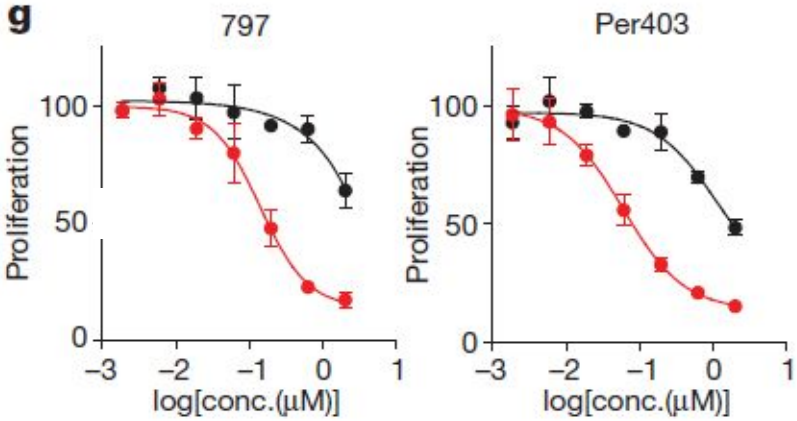




JQ1 против рака

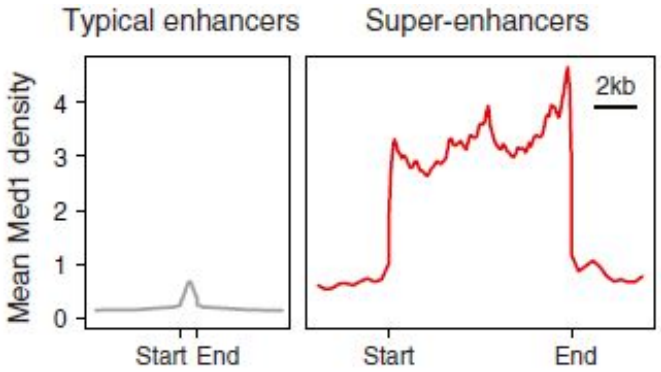
- JQ1 оказался эффективным в подавлении роста целого ряда раковых клеток





Суперэнхансеры

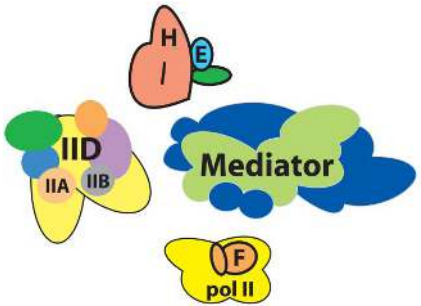
- В геноме человека около 400,000 энхансеров
- В отдельном типе клеток активны от 5,000 до 100,000
- Можно выделить отдельный тип энхансеров длиной около 10кпо
- Обогащены Med1, H3K27ac, H3K4me1, и master regulators



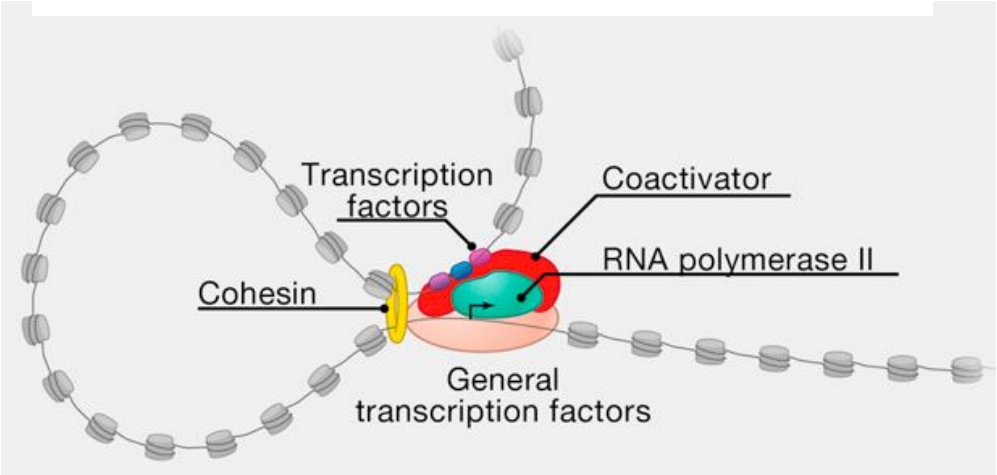
	Typical enhancers		Super-enhancers	
Number:	8563		231	
Median size:	703 bp		8667 bp	
	Total signal	Density at constituents	Total signal	Density at constituents
Med1:	1x	1x	28x	8.1x
H3K27ac:	1x	1x	26x	4.8x
H3K4me1:	1x	1x	10x	1.3x
DNaseI:	1x	1x	8x	2.2x

Med1 как метка супер-энхансера

- Компонент медиаторного комплекса, который вовлечен в работу многих (возможно почти всех) энхансеров

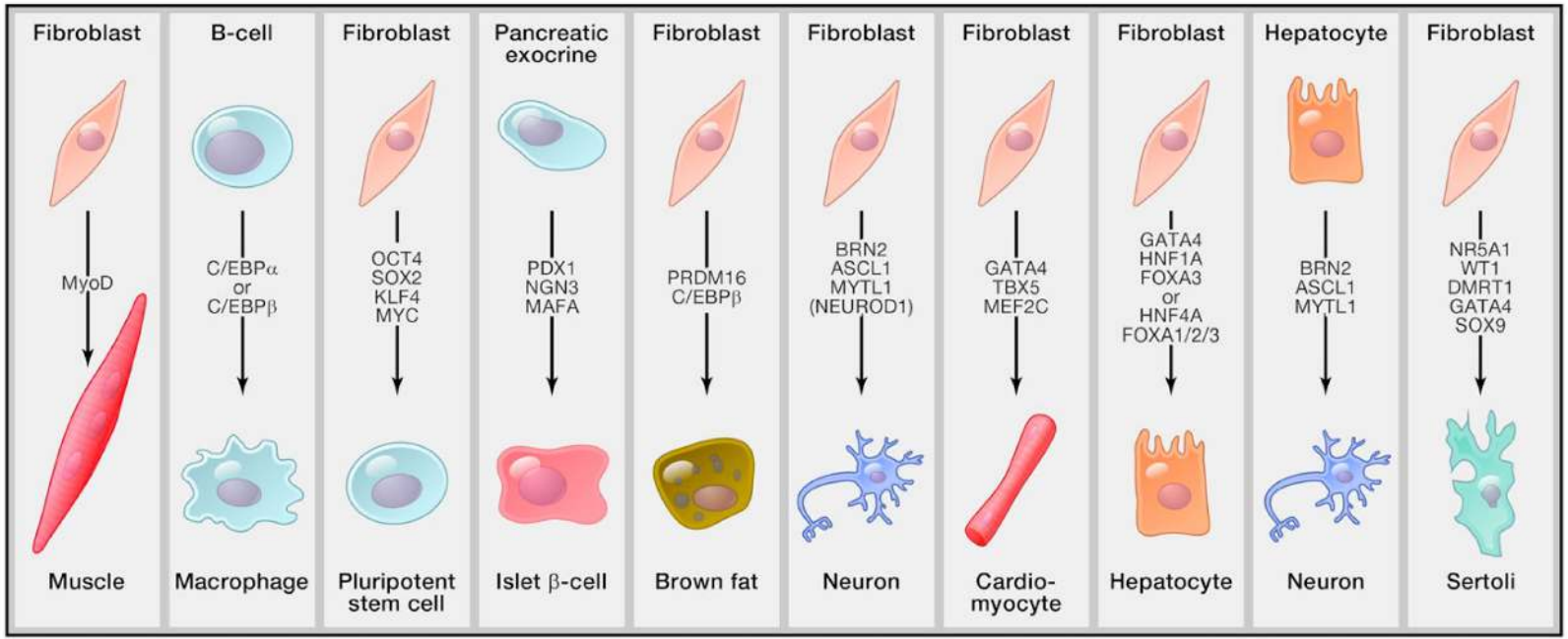


*TFIIA: 2 subunits**
TFIIB: 1 protein, 33 kDa
*TFIID: 16 subunits (TBP + 15 TAFs)***
TFIIE: 2 subunits
TFIIF: 2 subunits
TFIIH: 10 subunits
pol II: 12 subunits
Mediator: 26 subunits



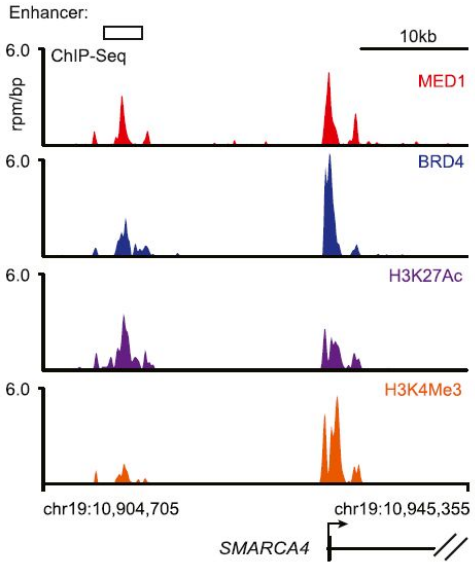
Master regulators

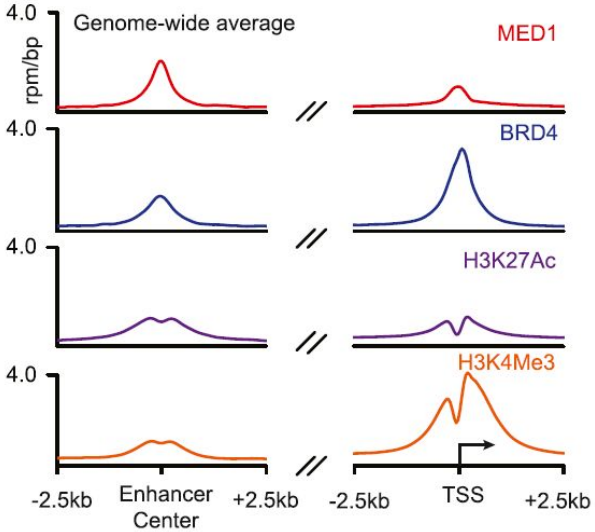
- Определяют судьбу данной ткани или типа клеток
- Мощная экспрессия - 10-100к копий на клетку



Рак и суперэнхансеры

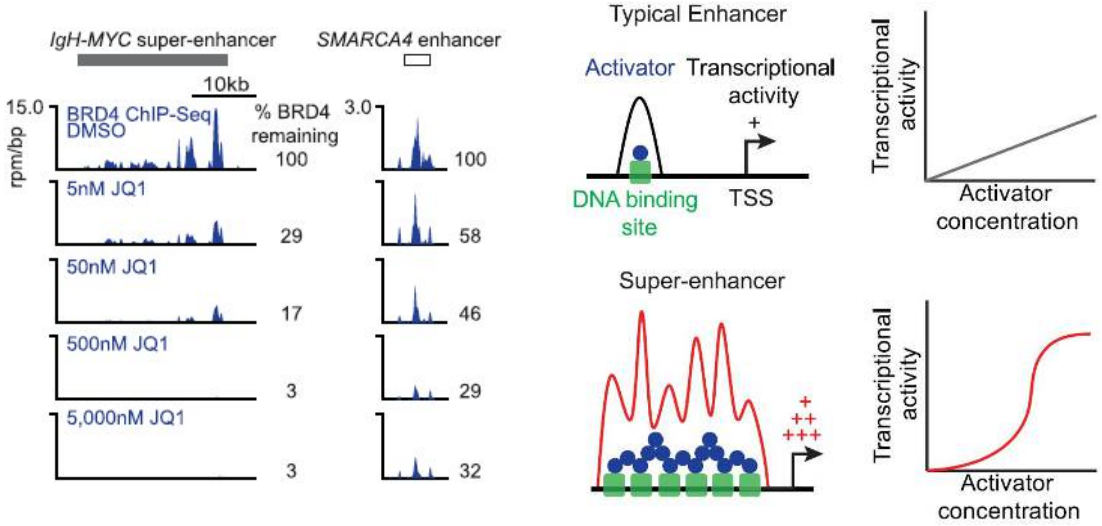
- Почему же JQ1 работает?
- BRD4 и MED1 работают на энхансерах вместе!





ТФ-аддикция

- Суперэнхансеры зависят от кооперативных взаимодействий больше чем обычные
- поэтому они сильнее зависят от концентрации компонентов!

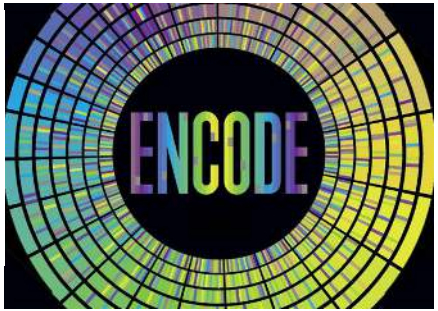


проект ENCODE

- ENCODE = ENCyclopedia Of DNA Elements
- наследник Human Genome Project
- цена пилота (2007): $55M, на данный момент ~$300M
- RNA-Seq, ChIP-seq основных (?) ТФ и гистонных модификаций, а также метилирование ДНК
- публикации - осень 2012 (6 Nature + 24 статьи в других журналах)

National Human Genome Research Institute

ENCODE



ENCODE By the Numbers

147 cell types studied
80% functional portion of human genome
20,687 protein-coding genes
18,400 RNA genes
1640 data sets
30 papers published this week
442 researchers
$288 million funding for pilot, technology, model organism, and current project

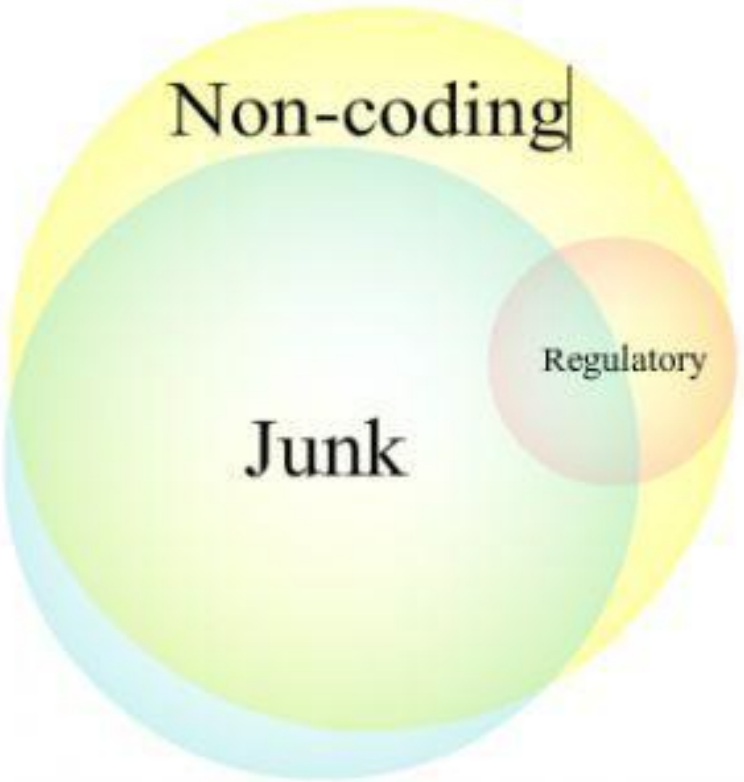
http://www.sciencemag.org/content/337/6099/1159/F2.expansion.html

Результаты ENCODE

- 400,000 энхансеров и 70,000 промотеров
- “*More than 80% of human genome is functionally active*”

Критика ENCODE

- 80% ДНК не может быть функциональны, т.к. только 5-15% консервативны



Скандалы, интриги, расследования!

- Общий научный консенсус - ENCODE сделал отличные данные, но многократно нарушил научную этику в анализе и публицистических обзорах результатов (см. погоня за сенсациями)

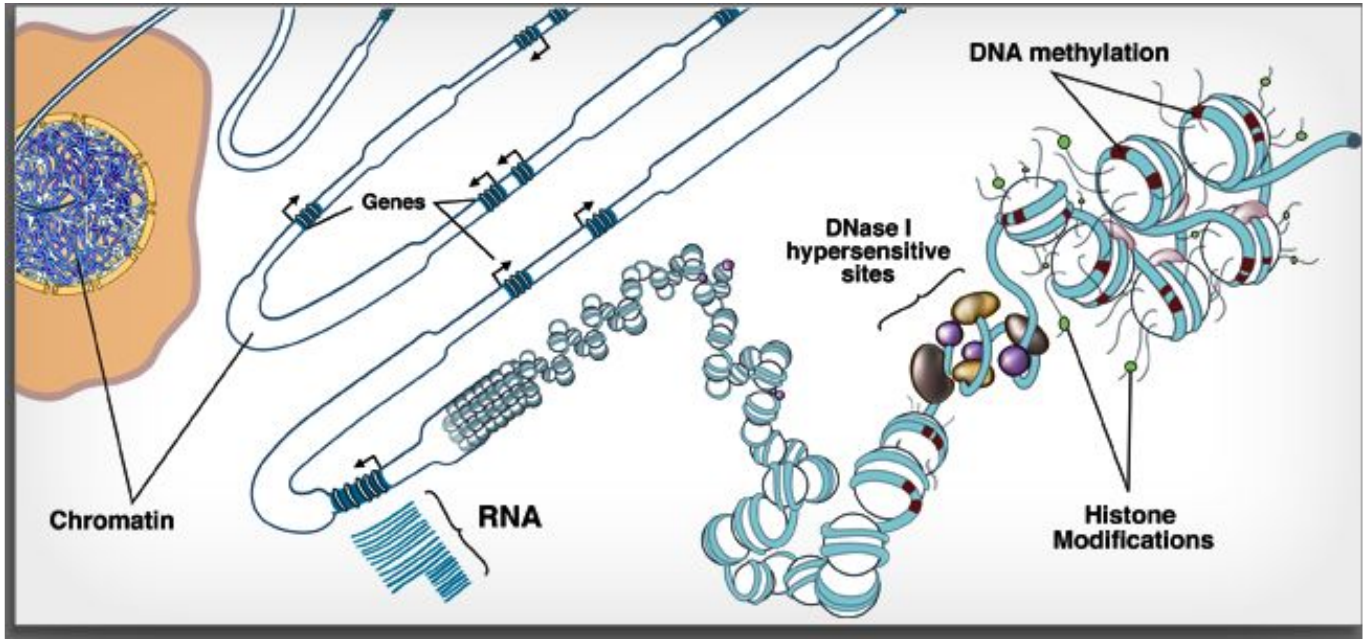
http://blogs.scientificamerican.com/guest-blog/2012/09/17/junk-dna-junky-pr/

Скандалы, интриги, расследования!

- многие уважаемые люди были огорчены. и поделом.

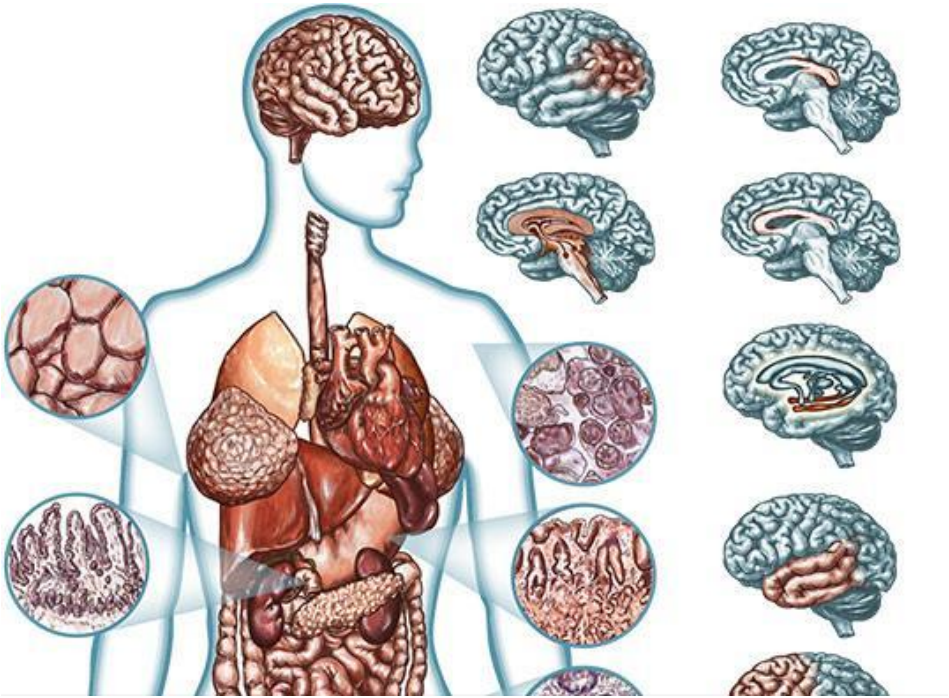
http://blogs.scientificamerican.com/guest-blog/2012/09/17/junk-dna-junky-pr/

Roadmap Epigenomics Project



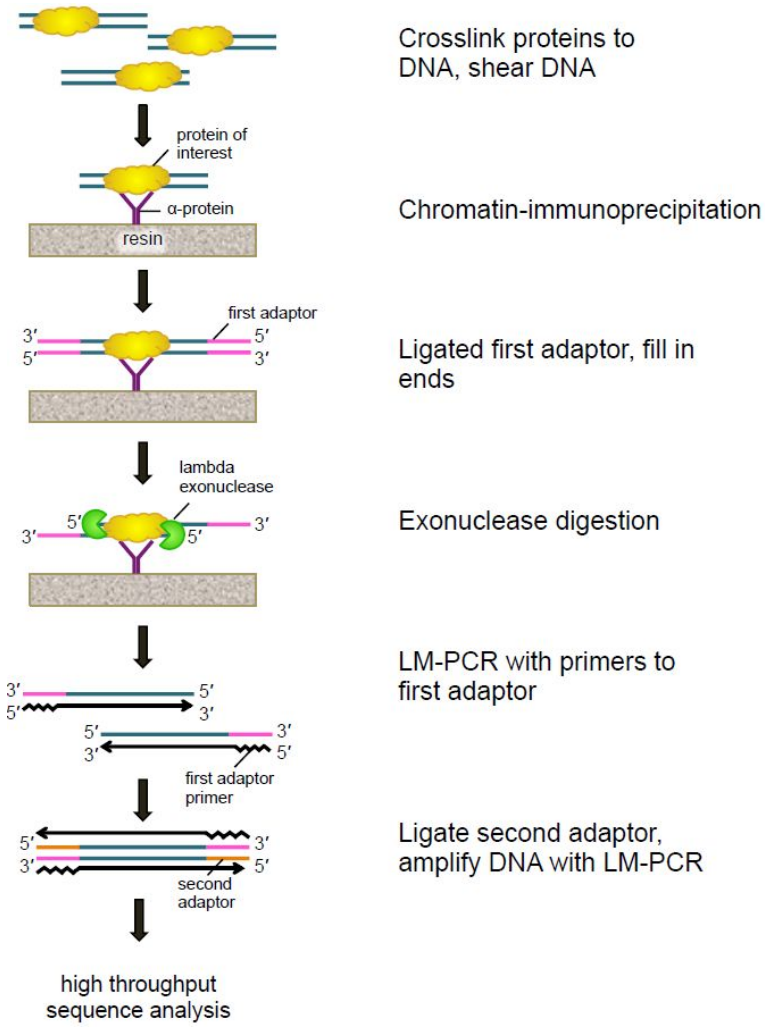
Roadmap Epigenomics Project

- Более 100 первичных тканей (а не клеточных линий)
- *«2,804 genome-wide datasets, including 1,821 histone modification datasets, 360 DNase datasets, 277 DNA methylation datasets, and 166 RNA-Seq datasets»*



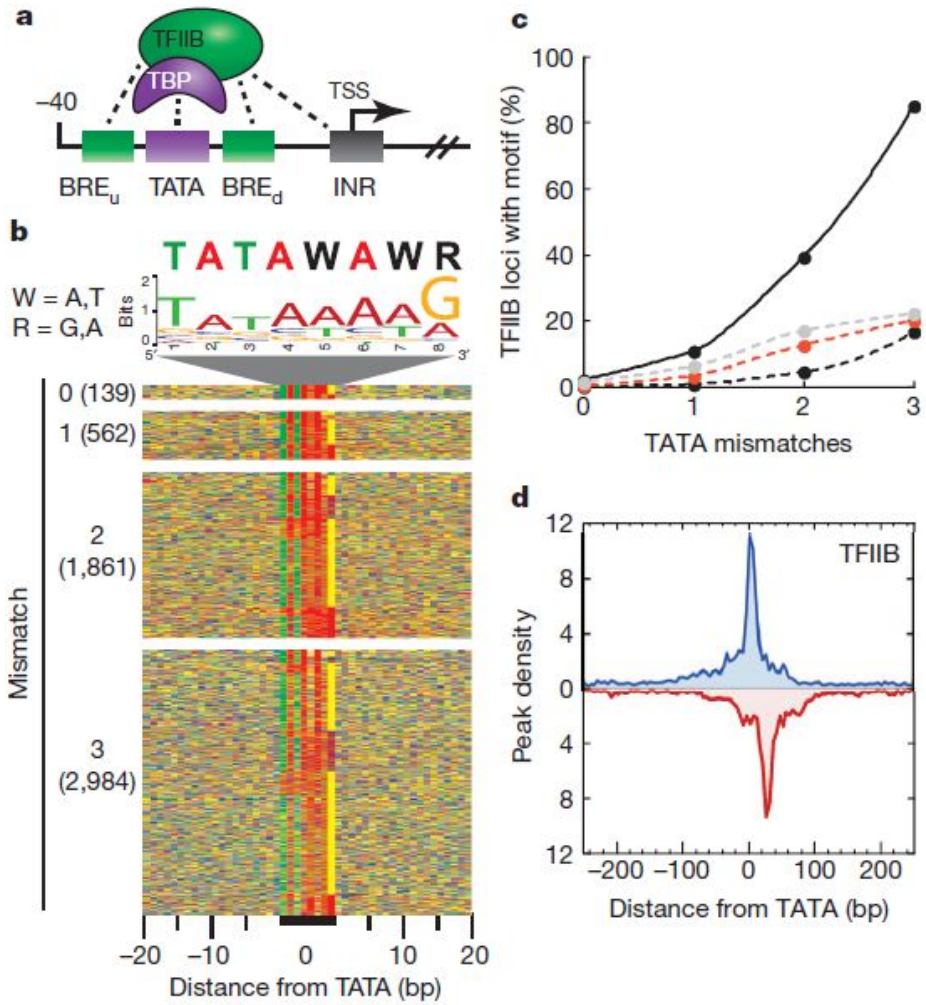
Прогресс в методах: ChIP-Exo

- ChIP-Exo позволяет существеннейшим образом увеличить точность локализации ТФ на геноме



ChIP-Exo в деле

- Фундаментальная поправка к существующей парадигме: TATA box есть почти везде!



Как видите, все довольно просто.

Спасибо за внимание.