

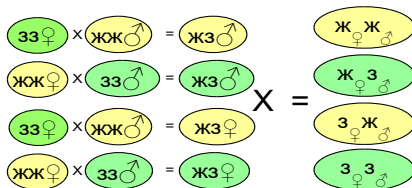
## Эпигенетика – что это такое?

Андрей Миронов

ФББ МГУ, ИППИ РАН

Вороново – 2017

# Немного генетики



Опыты показывают, что иногда наследуется признак только от отца, но при этом генотип может нести оба признака, которые расщепятся в следующем поколении.

## Немного генетики

Проявления признаков, унаследованных только от отца (матери) называется генным импринтингом (ввел Г.Кроуз в 1960 г.)

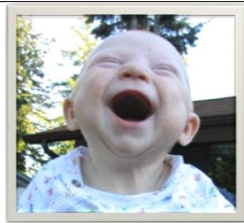
Существует информация, которая идет поверх генов отсюда термин *эпигенетика*, например, метка того, что данный ген пришел от отца.

На самом деле термин возник гораздо раньше в связи с исследованиям развитием организма: при развитии организма клетки наследуют состояние от предков – *соматическое наследование*

# Эпигенетические заболевания

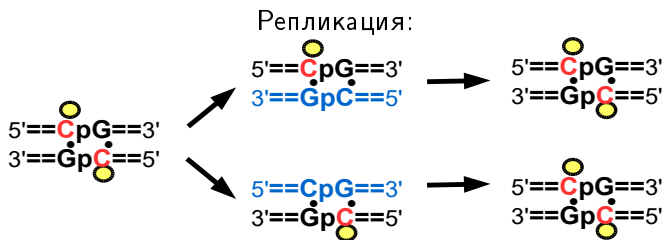
Число импринтированных генов - 200-500 В настоящее время известно 10 наследственных синдромов, связанных с импринтингом (Прадера-Вилли или Ангельмана; Видемана-Беквита; Рассела-Сильвера; синдромы однородительских дисомий)

синдром Ангельмана (счастливой куклы) Приступы неконтролируемого смеха, хлопанье в ладоши, специфическое выражение лица. Частота встречаемости 1/20 тыс. новорожденных.



# Метилирование ДНК

Метилирование ДНК – естественный механизм

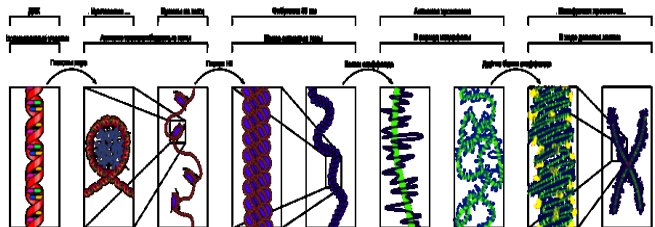


- метилирование аналогично модификациям букв в немецкой и французской письменности:  $A \Rightarrow \ddot{A}$

# Хроматин

ДНК упакована в хромосомы

Уровни организации:

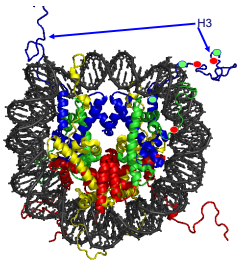


# Нуклеосома

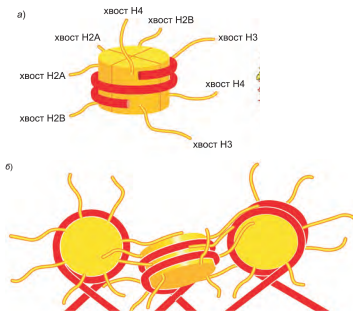
## Нуклеосома:

8 белков с намотанной ДНК Белки – гистоны H2A, H2B, H3, H4  
 ДНК – 1.7 витка вокруг нуклеосомы; 147 нуклеотидов

### Структура нуклеосомы:



### Взаимодействие нуклеосом:



Альбертс, Молекулярная биология клетки

# Модификации хроматина

- Метилирование
- Ацетилирование
- Фосфорилирование
- Замена гистонов (напр. H2A.Z)
- Много чего еще
- Метилирование ДНК по CpG

- *Гетерохроматин* – плотно упакованный хроматин
- *Эухроматин* – более рыхлый хроматин

На самом деле *Эухроматин* и *Гетерохроматин* — граничные состояния хроматина.

Модификации хроматина **наследуются**  
в соматических клетках



# Модификации гистона H3

		Пром.	Энх.	Транс.	Гетеро.	Ploycmb
актив	H3K4me1	+	+++	-	-	-
	H3K4me2	+	+	-	-	-
	H3K4me3	+++	+	-	-	-
	H3K36me3	-	-	++	-	-
	H3K79meX	-	-	++	+	-
репресс	H3K9me1	+	-	-	++	-
	H3K9me2,3	-	-	-	++	-
	H3K27me1	-	-	+	-	-
	H3K27me2,3	-	-	-	+	++

# Промоторы

- Type I: промоторы ткане-специфические:
  - Строгие TSS разваленные нуклеосомы.
  - Помечены только H3K4me3.
  - Нет CpG островов.
- Type II : Большинство генов.
  - Есть CpG острова.
  - Широкие TSS.
- Type III: Промоторы генов развития.
  - Более узкие TSS
  - Репрессируются системой Polycomb
  - Несколько CpG островов.

*N.Harmston, B.Lenhard. NucleicAcidsRes.2013; 41(15) : 7185 – 99.*

# Экспрессия

- Позвоночные: распределение  $H3K4me3$  повторяет распределение CpG островов
- Пром.генов развития  $\Rightarrow H3K27me3$  и с  $H3K4me3$
- Метилирование  $\Rightarrow$  создание длительного репрессированного состояния: молчание гена  $\Rightarrow H3K27me3$   
 $\Rightarrow$  метилирование ДНК
- Энкхансеры обогащены  $H3K4me1$ ,  $H3K4me2$  и  $H3K27ac$ , но не  $H3K4me3$
- CTCF участвует в распространении дальнего Энкхансер-промоторного взаимодействия

# Крупномасштабные изменения хроматина

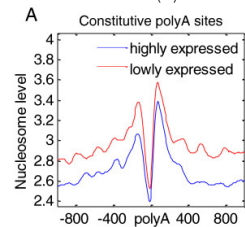
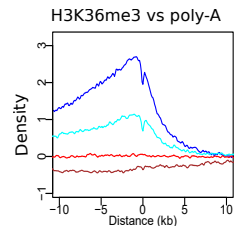
- Уровень PolIII и H3K4me2 сильно меняются на границах LAD (ламина)
- Polycomb и H3K27me3 могут образовывать непрерывные домены (>100 кб), и перекрываются с молчащими генами и межгенными интервалам
- CTCF, Cohesin, малые РНК, играют важную роль в дальних взаимодействиях

# Сплайсинг

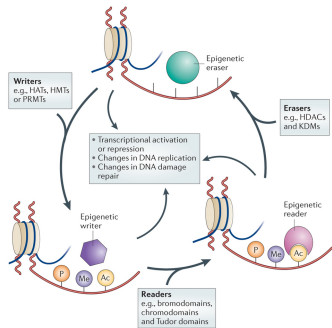
- Si РНК модулирует модификации гистонов и имеет влияние на скорость Pol-II (кинетическая модель). (*Allo, 2009*)
- Модификации гистонов H3K36me3 взаимодействует с РТВ (поли-Ру-связывающий белок) (*Lucio 2010*)
- Метилированные гены чаще имеют альт. сплайсинг (*Flores 2012*)
- Хроматин может влиять на сплайсинг непосредственно привлекая факторы сплайсинга (*Brown 2012*)
- Репрессивная метка H3K9me3 и HP1 $\gamma$  усиливают включение альтернативного экзона

# Полиаденилирование

- Низкая доступность хроматина вокруг поли (A)-сайтов, особенно для дистального поли (A)
- H3K36me3 образует пик в активно транскрибируемых генах особенно в экзонах
- Альт. сайты поли-А, как правило, имеют более высокий уровень H3K36me3 чем уникальные поли (A) сайтов



# Писатели, Читатели, Стиратели, Корректоры



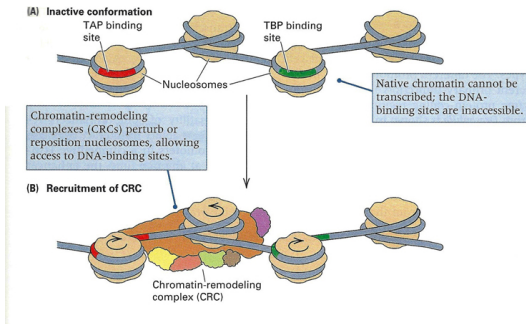
Nature Reviews | Drug Discovery

- Писатели Модифицируют хроматин
- Читатели Считывают хроматиновую метку
- Стиратели Стирают хроматиновую метку
- Корректоры Перемещают нуклеосомы

*K.J.Falkenberg and R.W. Johnstone. Nature Reviews Drug Discovery 13, 673~691(2014)*

# Писатели, Читатели, Стриратели, Корректоры

## Action of chromatin-remodeling complex



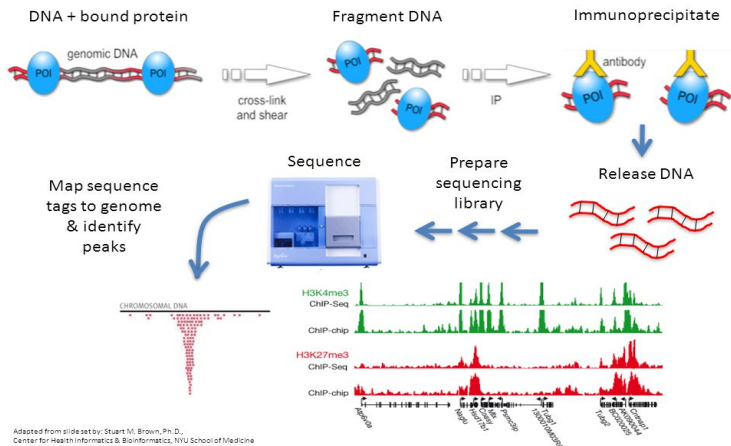
<http://www.discoveryaninnovation.com/BIOL202/notes/lecture19.html>



## Данные

## Модификации гистонов: ChIP-seq

## ChIP-seq overview



Adapted from slide set by: Stuart M. Brown, Ph.D.,  
Center for Health Informatics & Bioinformatics, NYU School of Medicine

## re-ChIP

На одном месте видим две модификации. Что это значит?

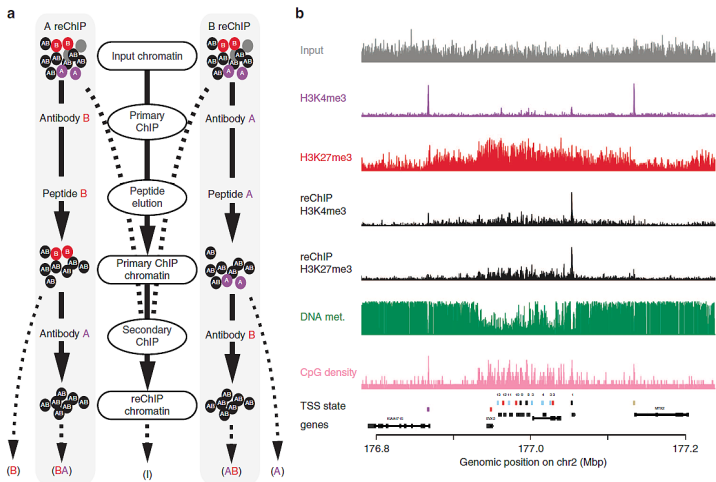
- В клетке есть две копии хромосом.
- На каждой нуклеосоме два гистона одного типа.
- В эксперименте участвуют  $> 100\ 000$  клеток

В разных клетках разные модификации?

Обе модификации на одной нуклеосоме или на разных?

Обе модификации на одной молекуле гистона или на разных?

## re-ChIP



*S. Kinkley et al, (2016) Nat Commun; vol.7; pp.12514 – 12514*

## re-ChIP

## Истино бивалентные



## Псевдо бивалентные

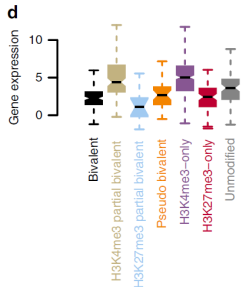
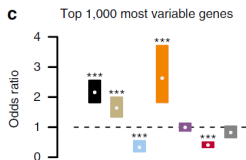
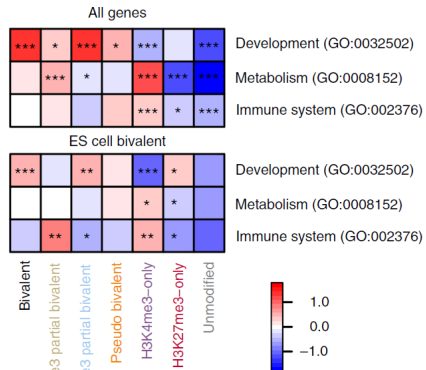


## Частично бивалентные



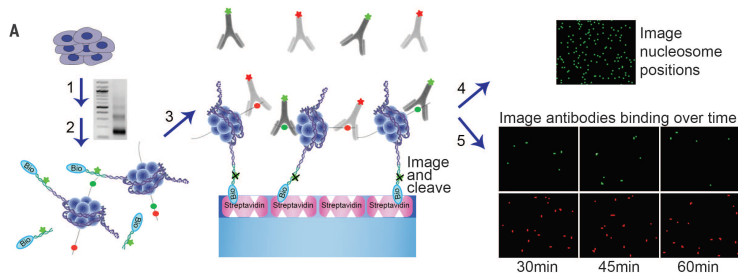
## re-ChIP

## 5 Доноров:

**b**

*S.Kinkley et al, (2016) NatCommun; vol.7; pp.12514 – 12514*

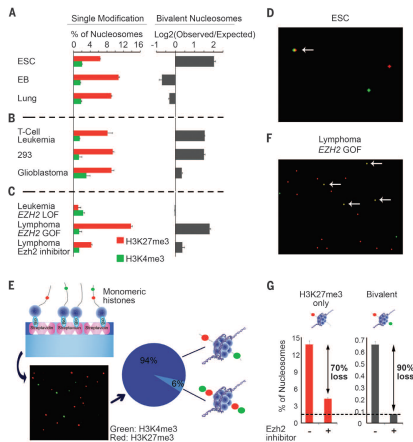
# Одно-нуклеосомный метод



- Фиксируем формальдегидом
- Режем на куски нуклеазой
- Шьем Bio - линкер
- Наливам на SA – слайд
- Обрабатываем фл. антителами
- Смотрим в микроскоп

*E. Shema et al; (2016) Science; vol.352(6286)*

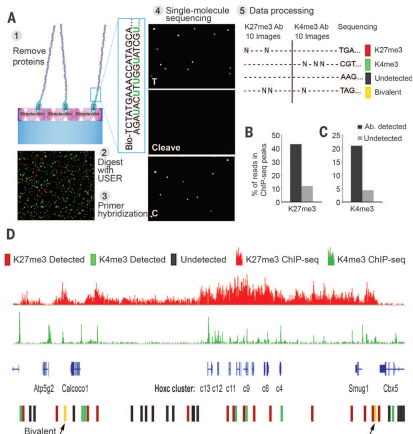
# Одно-нуклеосомный метод



- Делали на разных клетках:
  - ESC – эмбриональные стволовые в слое
  - EB – плюрипотентные стволовые клетки в виде "эмбрионального тела"
- При разных мутациях гистоновых метилаз
- Могли различить модификации на разных гистонах и двойные модификации

*E. Shema, et al; (2016) Science; vol.352(6286)*

# А где же последовательности?



- Одно-молекулярное секвенирование синтезом на тех же слайдах
- Получили ок 300 000 прочтений картировалось 80%
- 26 000 позитивных по H3K27me3 картировались на ChIP-seq пики

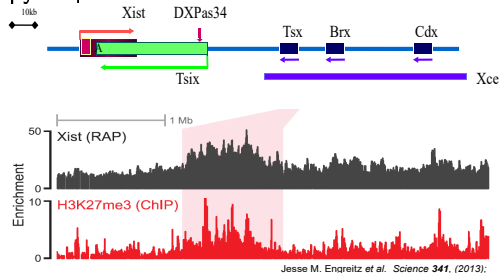
*E. Shema, et al; (2016) Science; vol.352(6286)*



# Инактивация X-хромосомы

Позвоночные:  $\boxed{\sigma = XY \quad \text{♀} = XX}$   $\Rightarrow$  проблема "дозы гена"

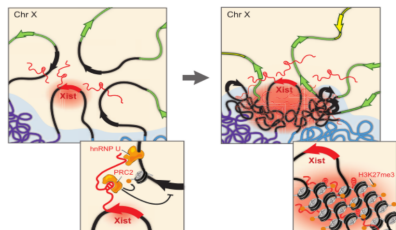
Инактивация происходит с помощью системы двух некодирующих РНК: XIST и TSIX.



# инактивация X-хромосом

## Модель:

- При индукции Xist распространяется на пространственно близкие локусы к активным генам.
- Неактивные регионы медленнее инактивируются.
- Взаимодействие с хроматином идет через hnRNP U.
- Xist использует Polycomb систему для модификации хроматина.



# Консервативность

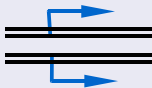
## Сравнение эпигеномов человек-мышь-свинья

- Межвидовые вариации эпигенома больше, чем внутри-видовые.
- Распределение H3K36me3 является более консервативными, чем H3K4me3
- Бивалентные домены ( H3K27me3 + H3K4me2 / 3 ) более консервативны, чем отдельные модификации
- Уровень консервативности выше в областях с ускоренным изменением последовательности

*S. Xiao, et al. Cell. 2012*

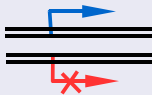
# Моноаллельная экспрессия

## Биаллельная экспрессия



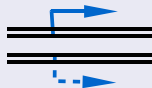
Большинство генов

## Моноаллельная экспрессия



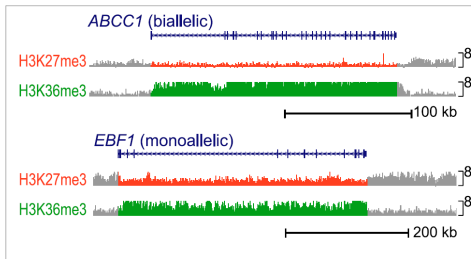
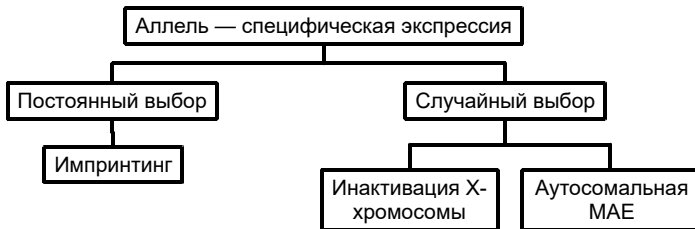
X - хромосома;  
импринтинг;  
имунная система;  
рецепторы запаха

## Смещенная экспрессия



разные гены  
(до 10 %)

# Моноаллельная экспрессия



Характерно:

транскрипционная метка  
**H3K36me3** и  
 репрессивная метка  
**H3K27me3**

# Источники данных

## Проекты

- *ENCODE* и *modENCODE*
  - > 200 образцов и клеточных линий
  - > 30 модификаций хроматина
  - транскрипционные факторы
- Консорциум *NIH Roadmap Epigenomics Mapping Consortium*
- Множество отдельных экспериментов, не входящих в эти проекты

## Genome Browser



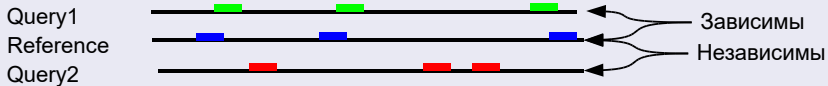
Гены

мРНК

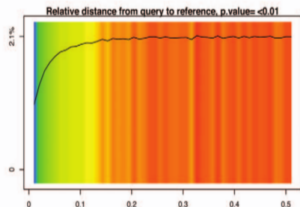
H3K27ac

# Анализ корреляций

## сравнение наборов интервалов



- Разные подходы определения статистической значимости пространственной ко-локализации двух наборов интервалов.
- Всегда набор интервалов.
- Данные преобразуются в набор интервалов с помощью *порога*



Chikina M.D., Troyanskaya O.G.; Bioinformatics (2012)  
Favorov A., et al. PLoS Comput Biol (2012)  
Kravatsky, Y., et al. DNA Res (2015)

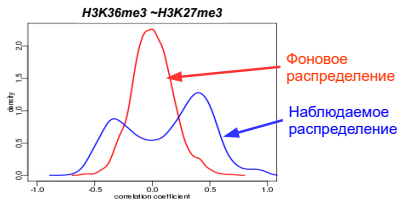


# Анализ корреляций: Stereogene

Корреляция с ядром  $\rho(t)$

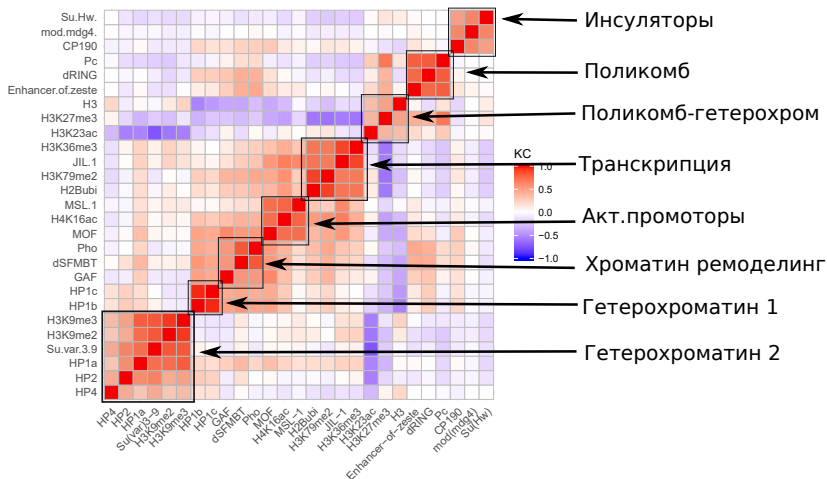
$$C_\rho = \frac{1}{|G|} \int_{G \otimes G} \frac{(f(x) - \bar{f}) \cdot (g(y) - \bar{g}) \rho(x - y)}{\sigma_f^\rho \cdot \sigma_g^\rho} dx dy$$

- Позволяет чувствовать корреляции на некотором расстоянии.
- Двумерный интеграл вычисляется с помощью преобразования Фурье



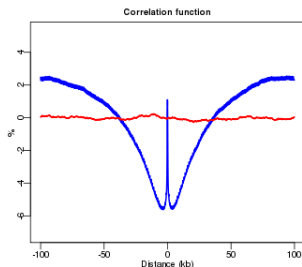
# Анализ корреляций: Stereogene

## Эмриональная клеточная линия Дрозофилы S2-DRSC



## Шило в ...

Мы наблюдаем перепредставленность положительных корреляций и часто странный пик в кросс-корреляции.

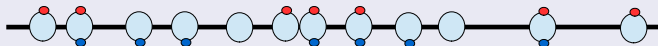


## Гипотеза

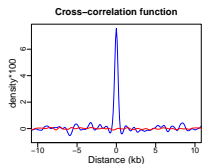
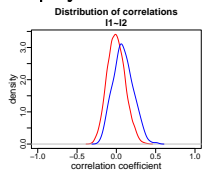
Перепредставленность положительных корреляций и Пик отражает то, что нуклеосомы точно позиционированы

# Эксперимент

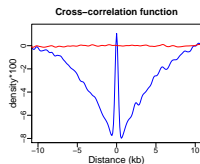
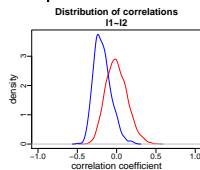
Эксперимент: случайные нуклеосомы, разбросаны две метки



## Бернуллиевские метки



## Марковские метки



# Эпигенетические подписи

## Процедура

- Дискретизируем разметки с помощью порога. Получаем разметки вдоль генома типа  $[0,1]$ .
- Строим Обобщенную Скрытую Марковскую модель с заданным числом состояний. Эмиссия – вектор дискретных разметок.
- Оптимизируем параметры методом Баума-Вельча.
- Повторяем для разного количества состояний и получаем

## Модель

*Ernst J., et al. Nature (2011)*

# Эпигенетические подписи

- (1) Активный промотор
- (2) Слабый промотор
- (3) Неактивный промотор
- (4-7) Эхансер
- (8) Инсулятор
- (9-11) Транскрипция
- (12) Polycomb
- (13) Гетерохроматин
- (14-15) Повторы

**b**

State	CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac	WCE	Candidate state annotation
1	16	2	2	6	17	93	99	96	98	2	Active promoter
2	12	2	6	9	53	94	95	14	44	1	Weak promoter
3	13	72	0	9	48	78	49	1	10	1	Inactive/poised promoter
4	11	1	15	11	96	99	75	97	86	4	Strong enhancer
5	5	0	10	3	88	57	5	84	25	1	Strong enhancer
6	7	1	1	3	58	75	8	6	5	1	Weak/poised enhancer
7	2	1	2	1	56	3	0	6	2	1	Weak/poised enhancer
8	92	2	1	3	6	3	0	0	1	1	Insulator
9	5	0	43	43	37	11	2	9	4	1	Transcriptional transition
10	1	0	47	3	0	0	0	0	0	1	Transcriptional elongation
11	0	0	3	2	0	0	0	0	0	0	Weak transcribed
12	1	27	0	2	0	0	0	0	0	0	Polycomb repressed
13	0	0	0	0	0	0	0	0	0	0	Heterochrom; low signal
14	22	28	19	41	6	5	26	5	13	37	Repetitive/CNV
15	85	85	91	88	76	77	91	73	85	78	Repetitive/CNV

Chromatin mark observation frequency (%)

**c**

**d**

Chromatin states

Luciferase relative light units

Ernst J., et al. *Nature* (2011)

# Задачи

- 1 Построение эпигеномного ландшафта по непрерывным эпигеномным данным
- 2 Исключение конфаундера – точного позиционирования нуклеосом
- 3 Поиск особенностей последовательностей, обеспечивающих точное позиционирование
- 4 Межвидовое сравнение треков и ландшафтов
- 5 Совместные ландшафты эпигеномики и транскрипционных факторов
- 6 Совмещение с хроматиновыми контактами

ФСЁ!!!