Регуляция и Эпигеномика

Андрей Миронов

ФББ МГУ

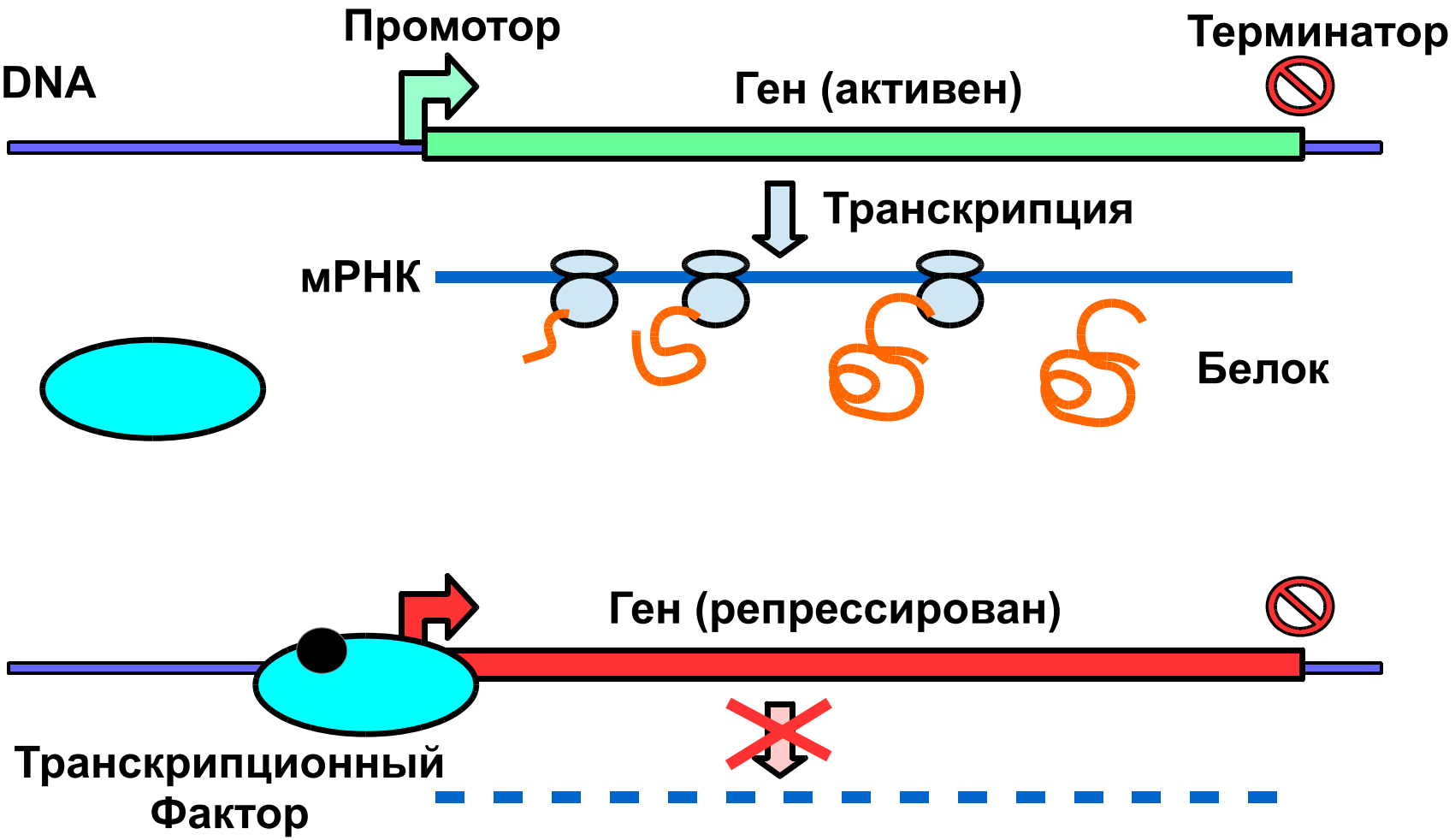
МФК – 2017

Регуляция Экспрессии

- В геноме фага λ около 40 генов
- В геноме бактерии *Mycoplasma* около 400 генов
- В геноме бактерии *E.coli* около 4000 генов
- В геноме плодовой мушки *Drosophila* 14 тыс генов
- В геноме человека около 25 тыс. генов
- ...

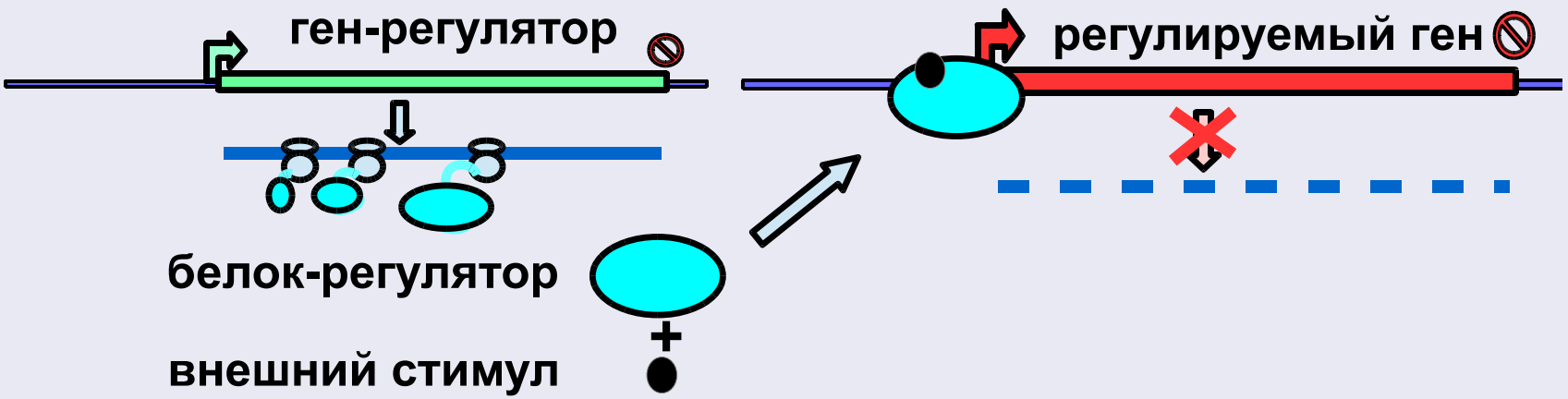
Все ли гены производят продукт (белок) всегда с одинаковой интенсивностью?
Существует регуляция экспрессии генов – экспрессия генов происходит с разной интенсивностью в зависимости от условий

Регуляция транскрипции

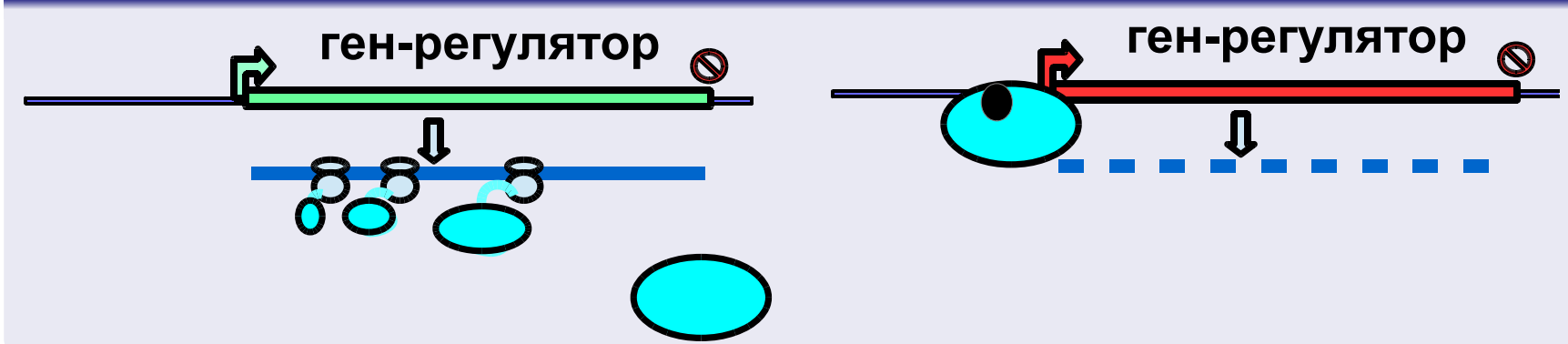


Регуляторы – тоже гены

Регуляторы – тоже гены

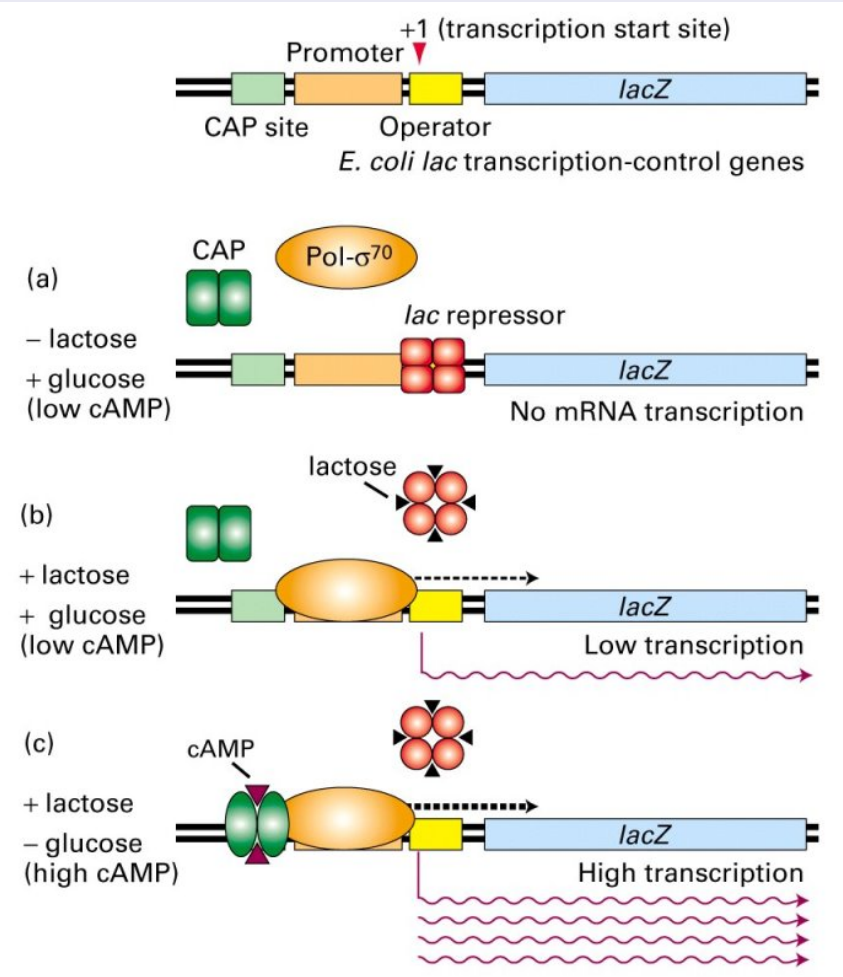


Авторегуляция



Пример 1

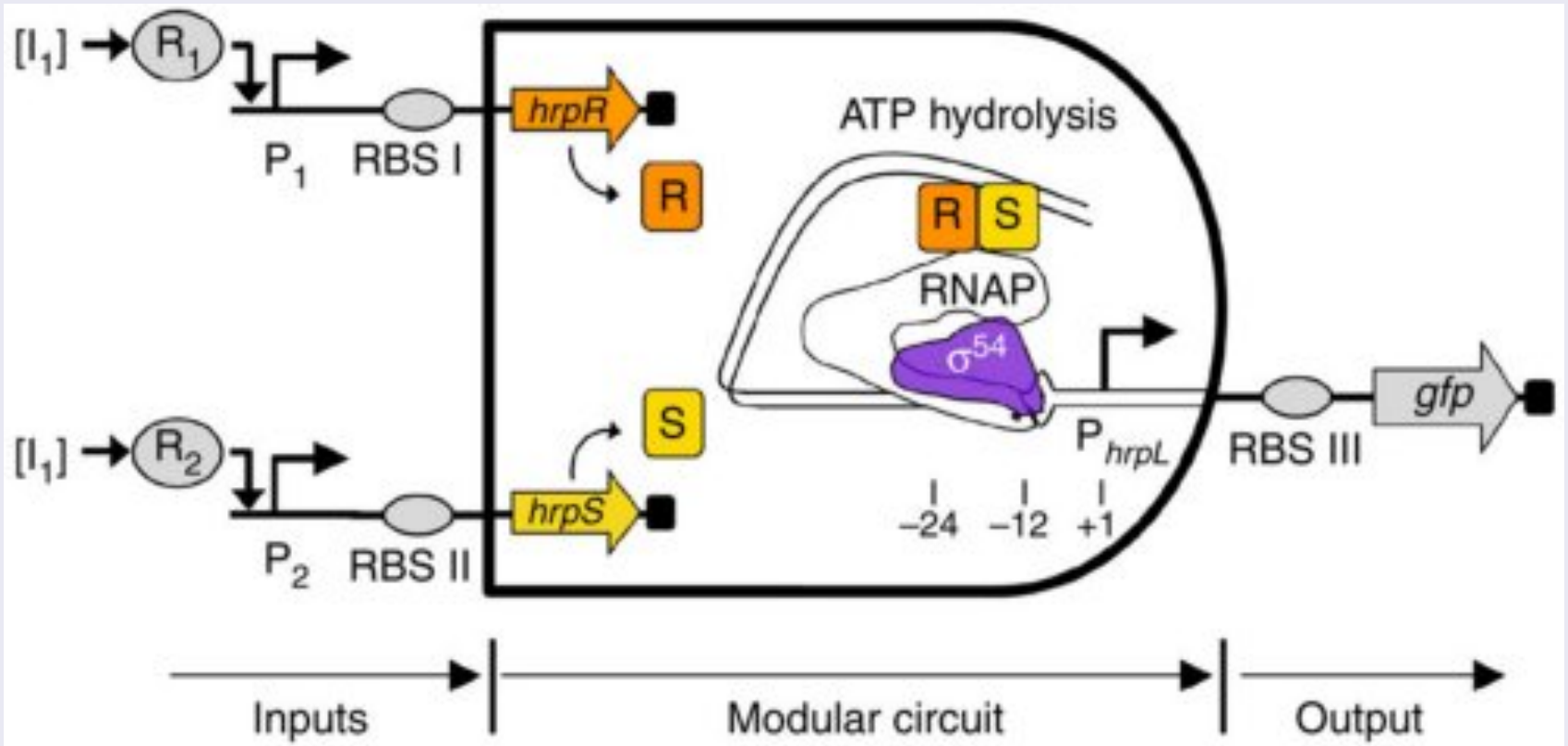
Lactose



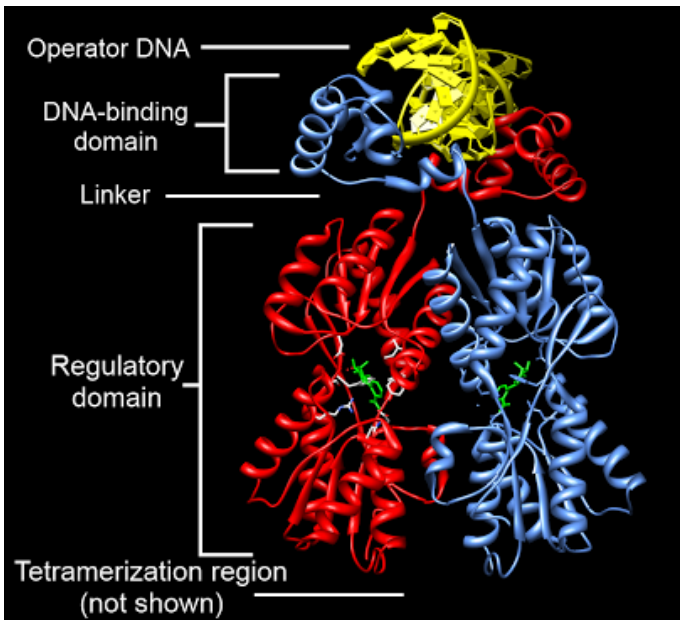
- Есть глюкоза $\Rightarrow$ лактоза не нужна
- Глюкоза кончилась $\Rightarrow$ голод, надо что-то делать $\Rightarrow$ cAMP
- Включаем потребление лактозы

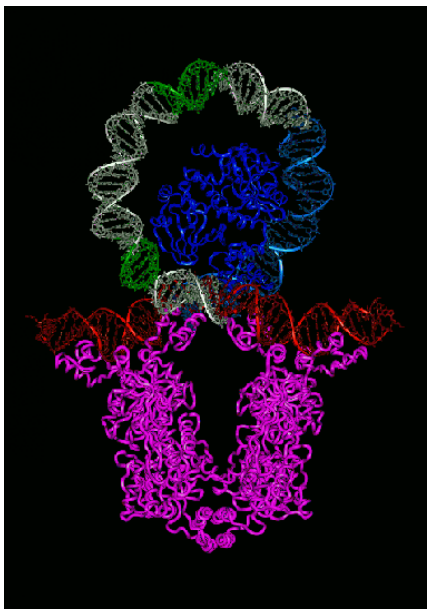
Пример 2

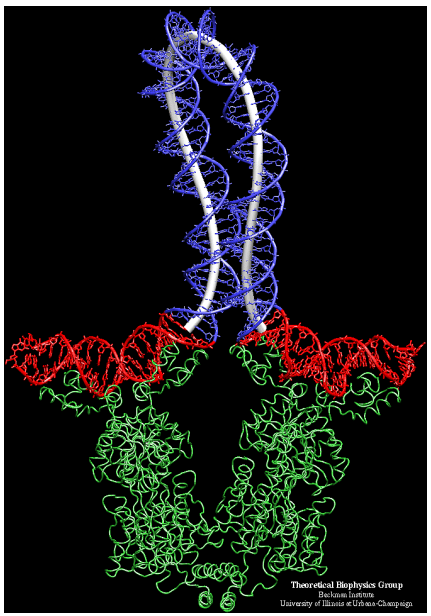
Генетические Вычисления



Связывание с ДНК







Связывание с ДНК

сайты связывания purR

C A C G C A A A C G T T T T C G T T
C A C G C A A A C G G T T T C G T C
T A C G C A A A C G T T T T C T T T
T G C G C A A A C G T T T T C G T T
C A C G C A A C C G T T T T C C T T
C T C G C A A A C G T T T G C T T T
T A C G C A A A C G T G T G C G T C
C G A T G A A C C G G T T G C C G C

Site Logo

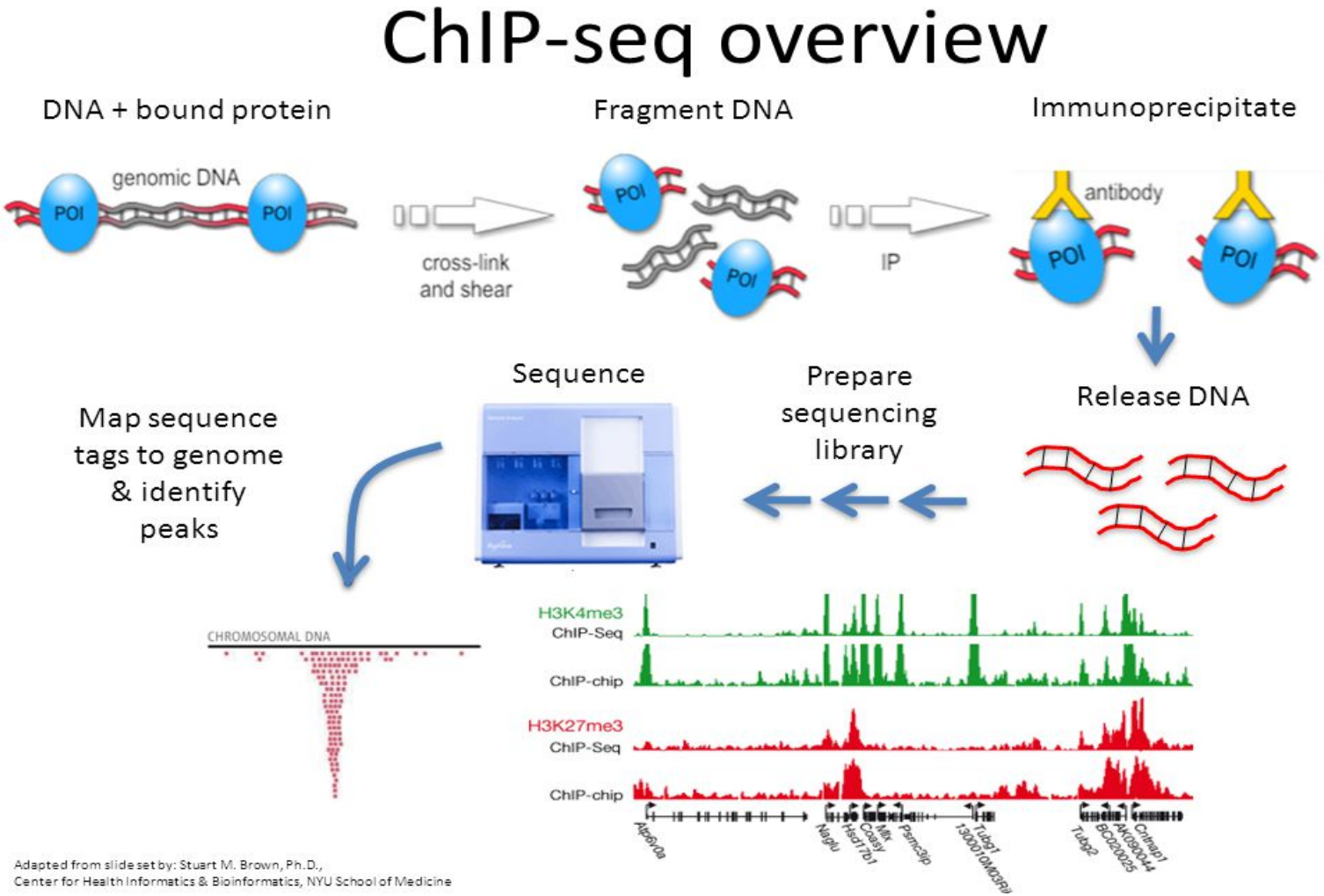
Описание мотива

	C	A	C	G	C	A	A	A	C	G	T	T	T	T	C	G	T	T
	C	A	C	G	C	A	A	A	C	G	G	T	T	T	C	G	T	C
	T	A	C	G	C	A	A	A	C	G	T	T	T	T	C	T	T	T
	T	G	C	G	C	A	A	A	C	G	T	T	T	T	C	G	T	T
	C	A	C	G	C	A	A	C	C	G	T	T	T	T	C	C	T	T
	C	T	C	G	C	A	A	A	C	G	T	T	T	G	C	T	T	T
	T	A	C	G	C	A	A	A	C	G	T	G	T	G	C	G	T	C
	C	G	A	T	G	A	A	C	C	G	G	T	T	G	C	C	G	C
A	0	5	1	0	0	8	8	6	0	0	0	0	0	0	0	0	0	0
C	5	0	7	0	7	0	0	2	8	0	0	0	0	0	8	2	0	3
G	0	2	0	7	1	0	0	0	0	8	2	1	0	3	0	4	1	0
T	3	1	0	1	0	0	0	0	0	0	6	7	8	5	0	2	7	5

Модель мотива: Позиционная Весовая Матрица

$$W(pos, \alpha) = \log \frac{n_\alpha(pos) + \psi_\alpha}{N + \sum \psi_\alpha}; \quad I(pos) = -\sum_\alpha f^\alpha_{pos} \log \frac{f^\alpha_{pos}}{p^\alpha}$$

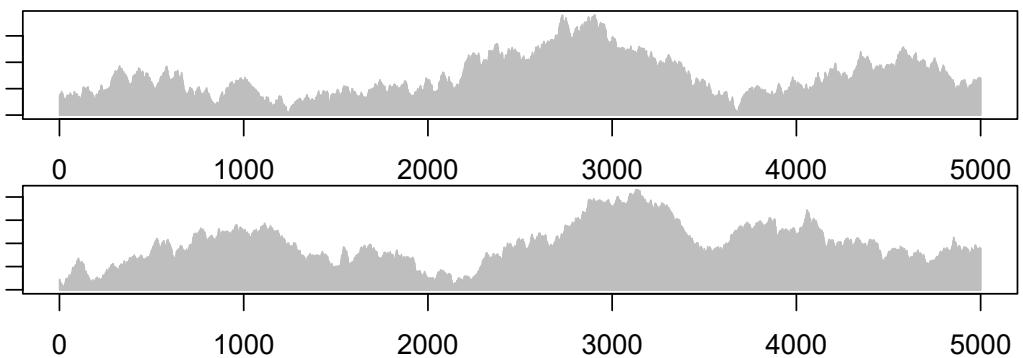
Данные


ChIP-seq overview
DNA + bound protein
genomic DNA
POI
POI
cross-link
and shear
Fragment DNA
POI
POI
IP
Immunoprecipitate
antibody
POI
POI
Release DNA
Prepare
sequencing
library
Sequence
Map sequence
tags to genome
& identify
peaks
CHROMOSOMAL DNA
H3K4me3
ChIP-Seq
ChIP-chip
H3K27me3
ChIP-Seq
ChIP-chip
Adapted from slide set by: Stuart M. Brown, Ph.D.,
Center for Health Informatics & Bioinformatics, NYU School of Medicine

Биоинформатика: Поиск сайтов

Исходные данные:

- ChIP – seq
- Гомолгия регуляторных областей
- SELEX
- Данные о ко-регуляции генов



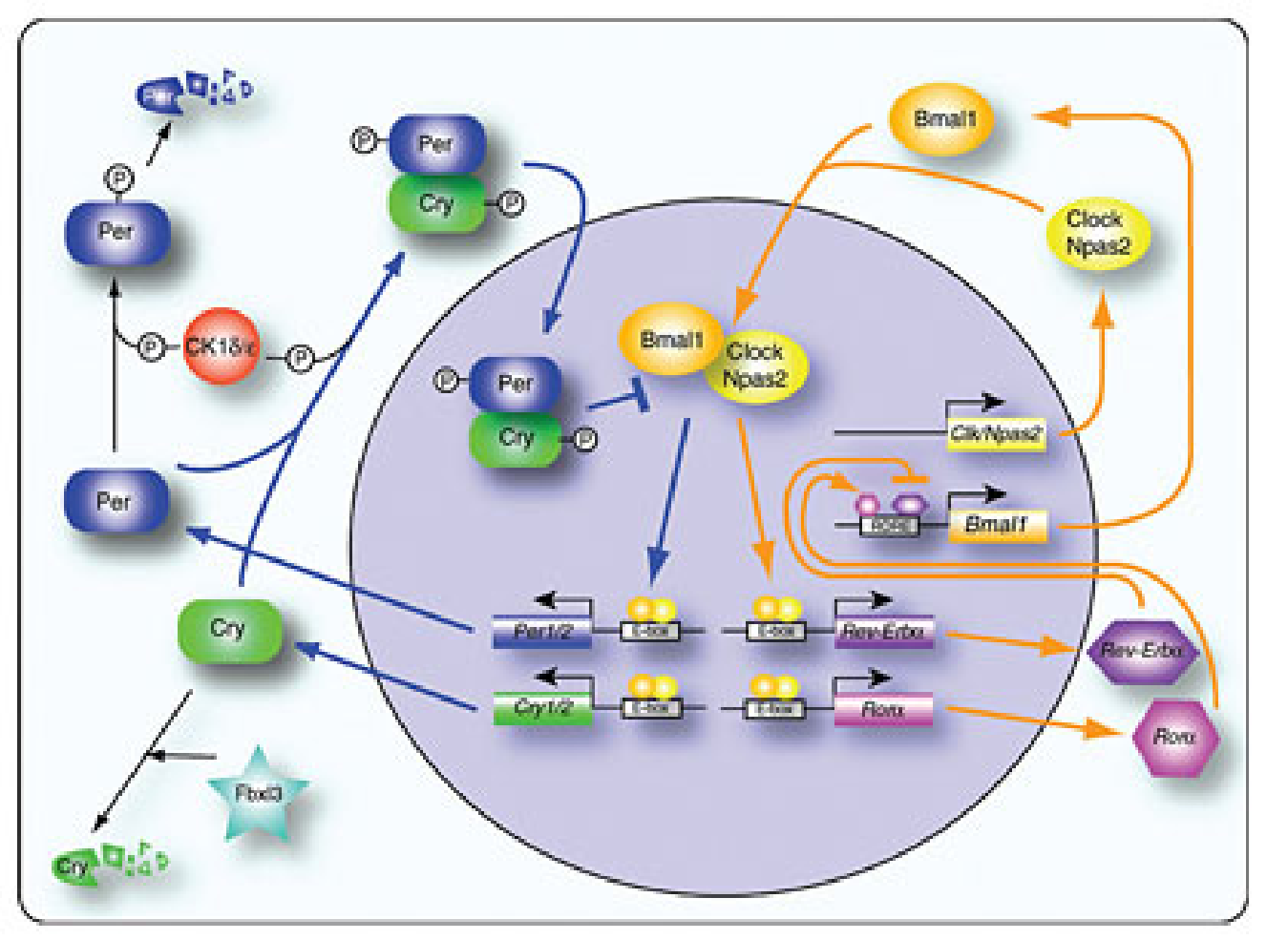
задача

Дано: Множество фрагментов последовательностей

Найти: Мотив

Регуляторная сеть

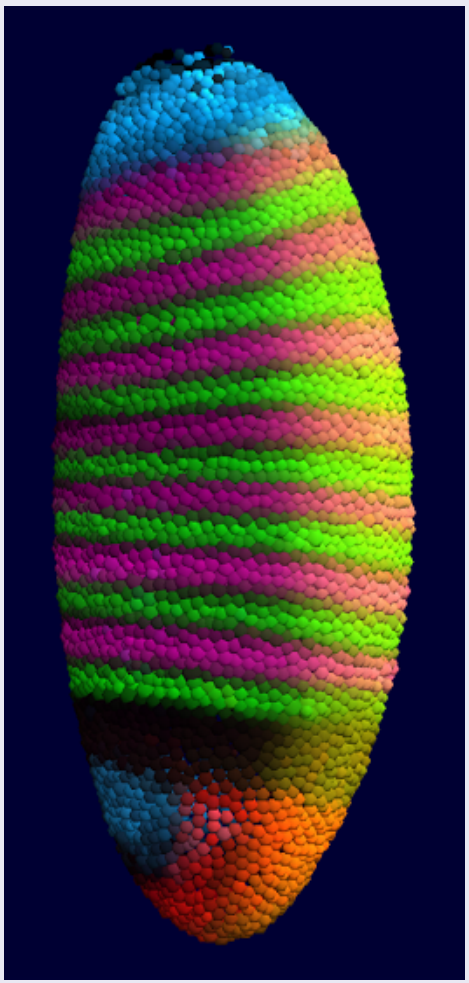
Регуляторная сеть

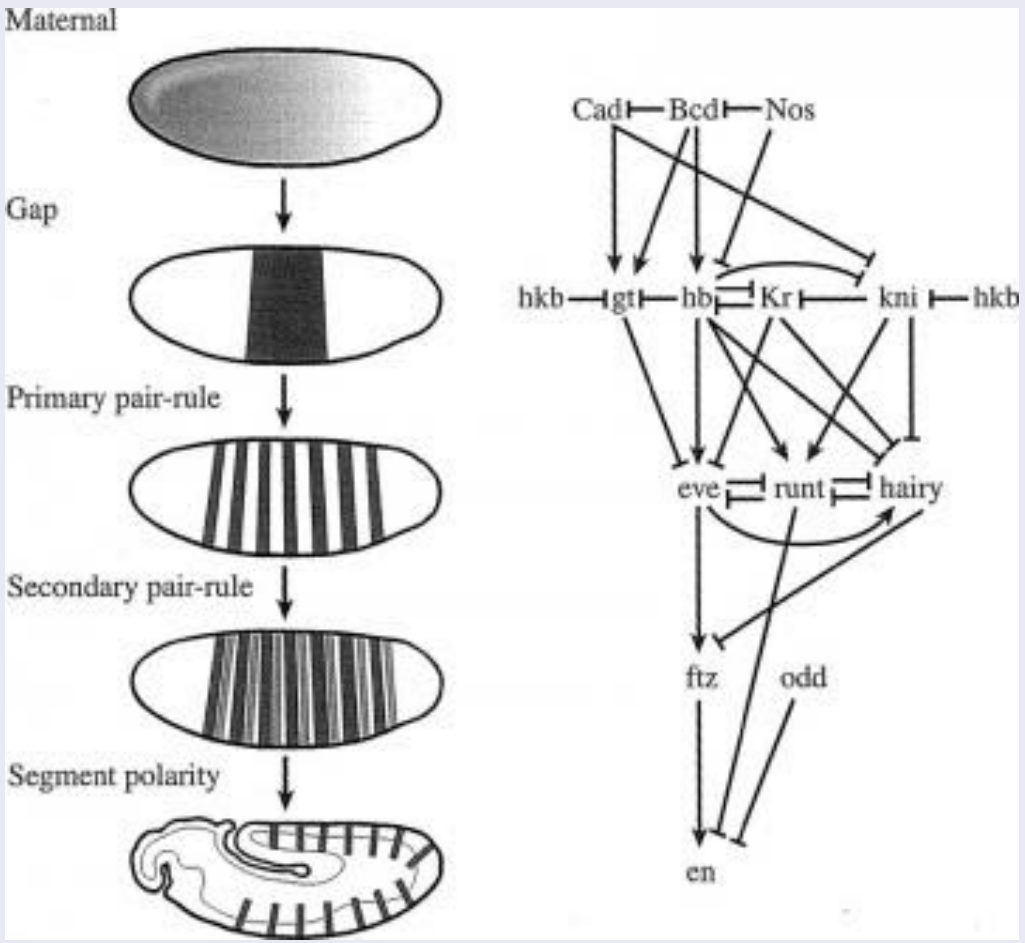


Elementy.ru

Сегментация дрозофилы

Экспресия разных регуляторов в раннем эмбрионе



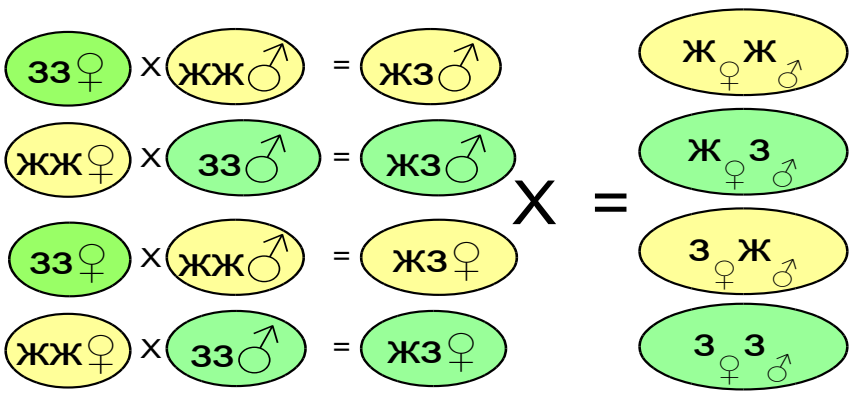


Уровни регуляции

Уровни регуляции

- Транскрипция
- Трансляция
- Стабильность РНК
- Стабильность белка
- Модификации белков

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*

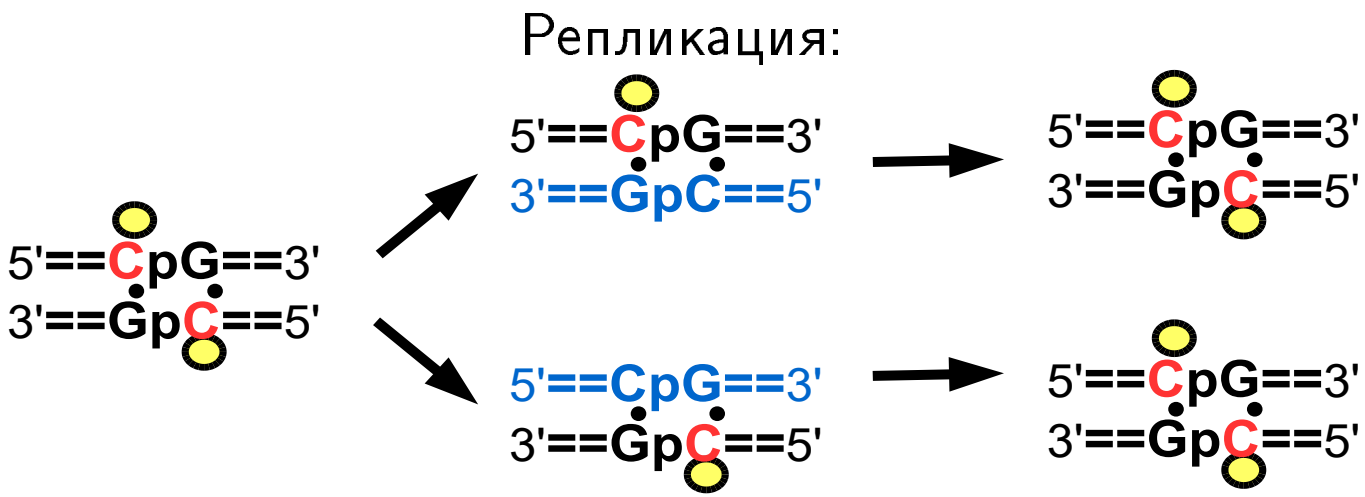
Эпигенетические заболевания

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 всасываемости 1/20 тыс. новорожденных.

Метилирование ДНК

Метилирование ДНК – естественный механизм

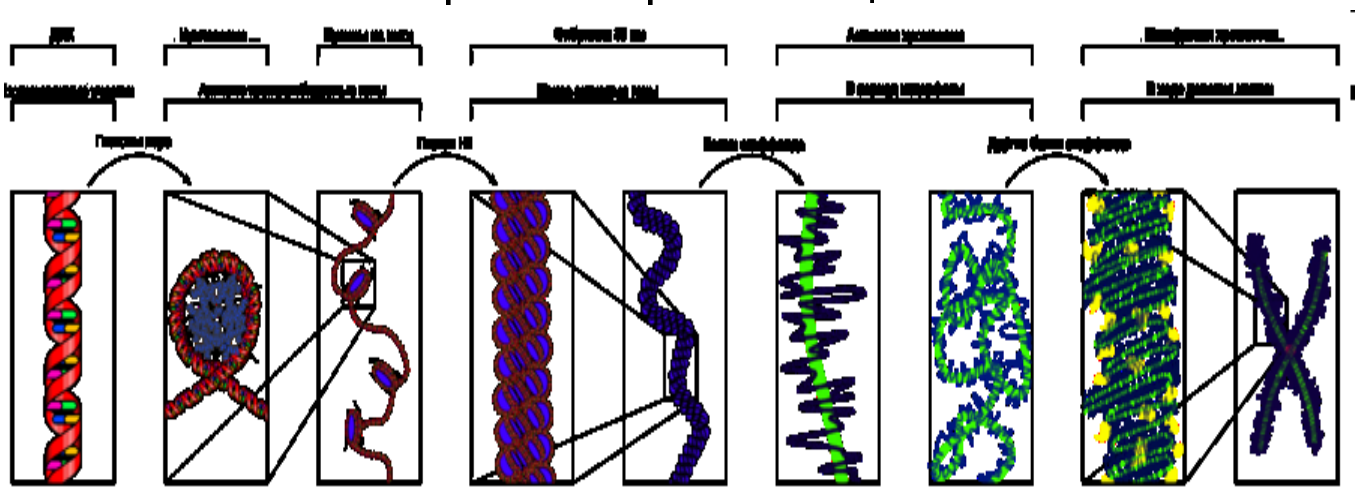


- метилирование аналогично модификациям букв в немецкой и французской письменности: **A** $\Rightarrow$**Ä**

Хроматин

ДНК упакована в хромосомы

Уровни организации:

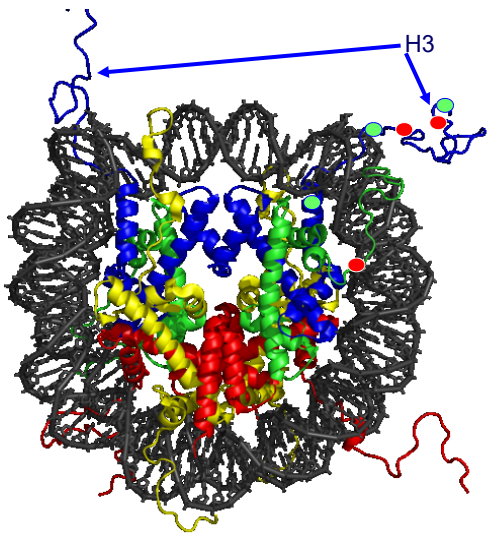


Нуклеосома

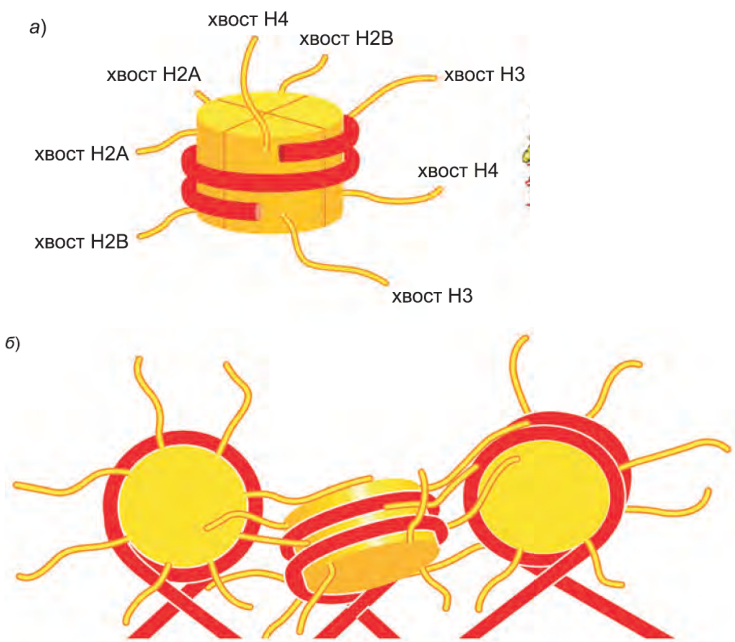
Нуклеосома:

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
- Асетилирование
- Фрсфорилирование
- 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

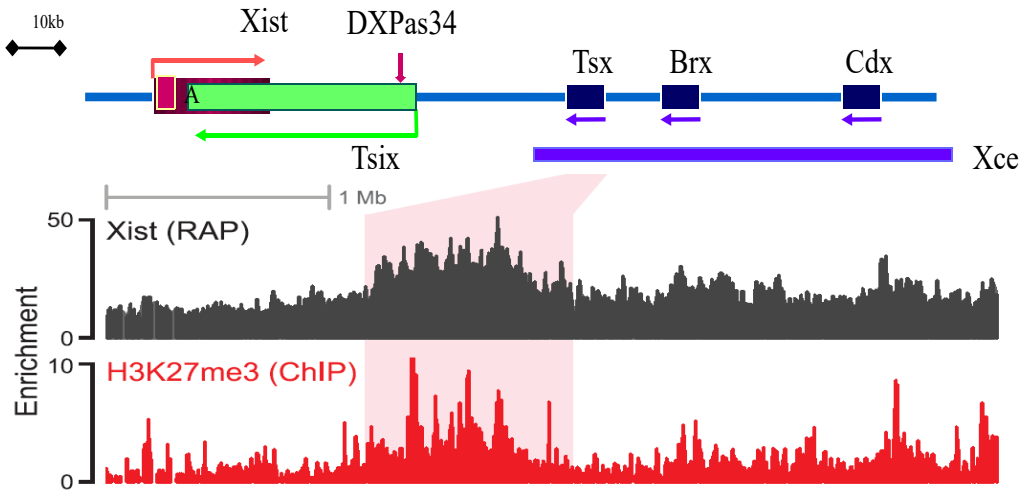
Модификации гистона H3

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

инактивация X-хромосомы

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

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



Jesse M. Engreitz *et al.* *Science* ***341***, (2013);

Консервативность

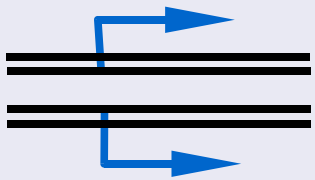
Сравнение эпигеномов человек-мышь-свинья

- 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

Xiao S, et al. Comparative epigenomic annotation of regulatory DNA. Cell. 2012

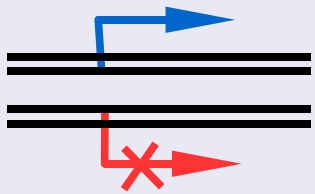
Моноаллельная экспрессия

Биаллельная экспрессия



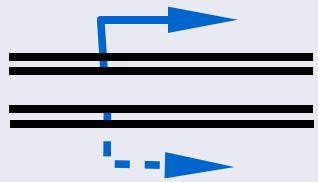
Большинство генов

Моноалельная экспрессия



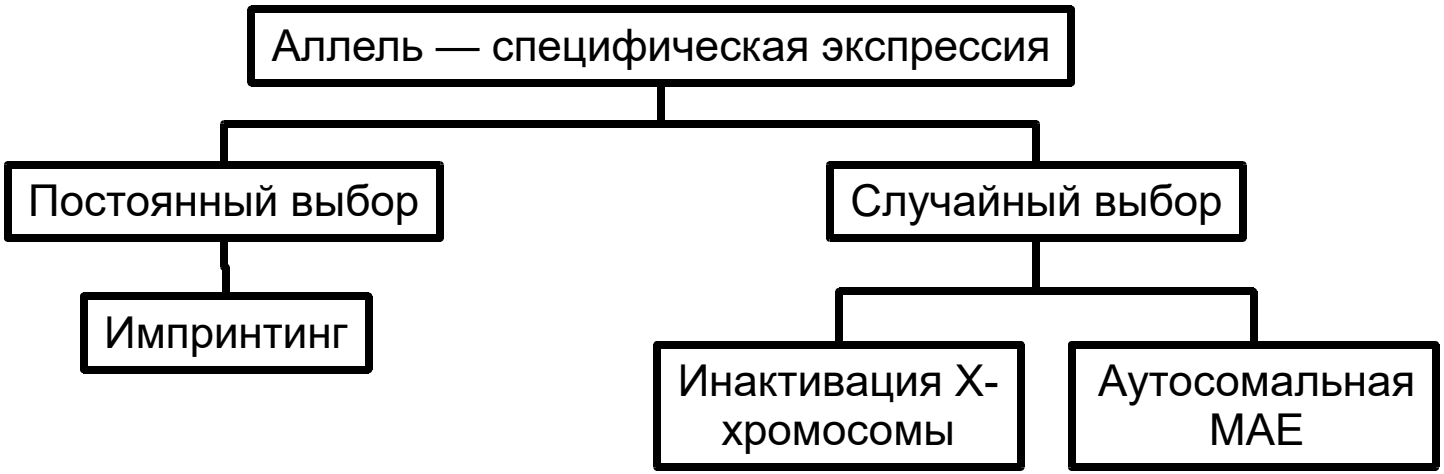
X - хромосома;
импринтинг;
имунная система;
рецепторы запаха

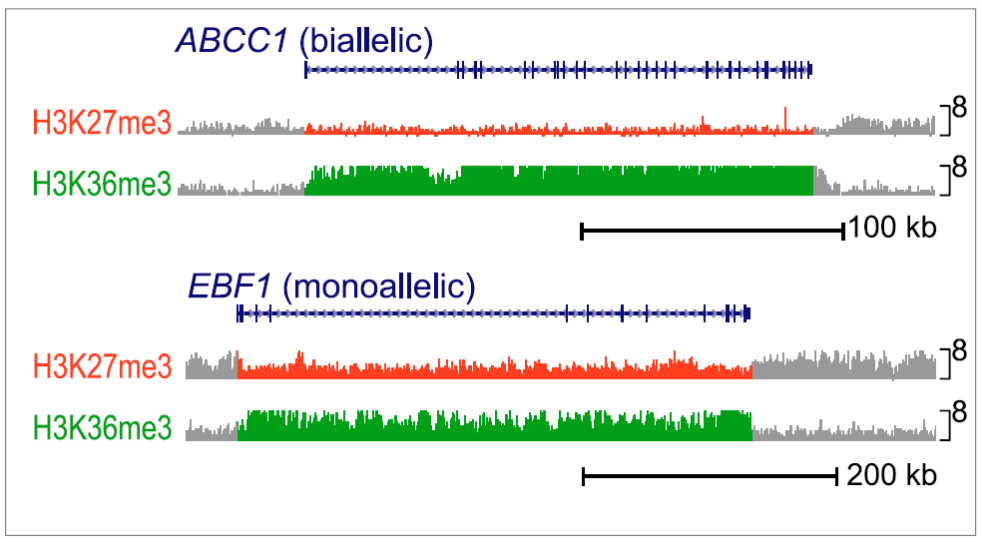
Смещенная экспрессия



разные гены
(до 10 %)

Моноаллельная экспрессия





Характерно:

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

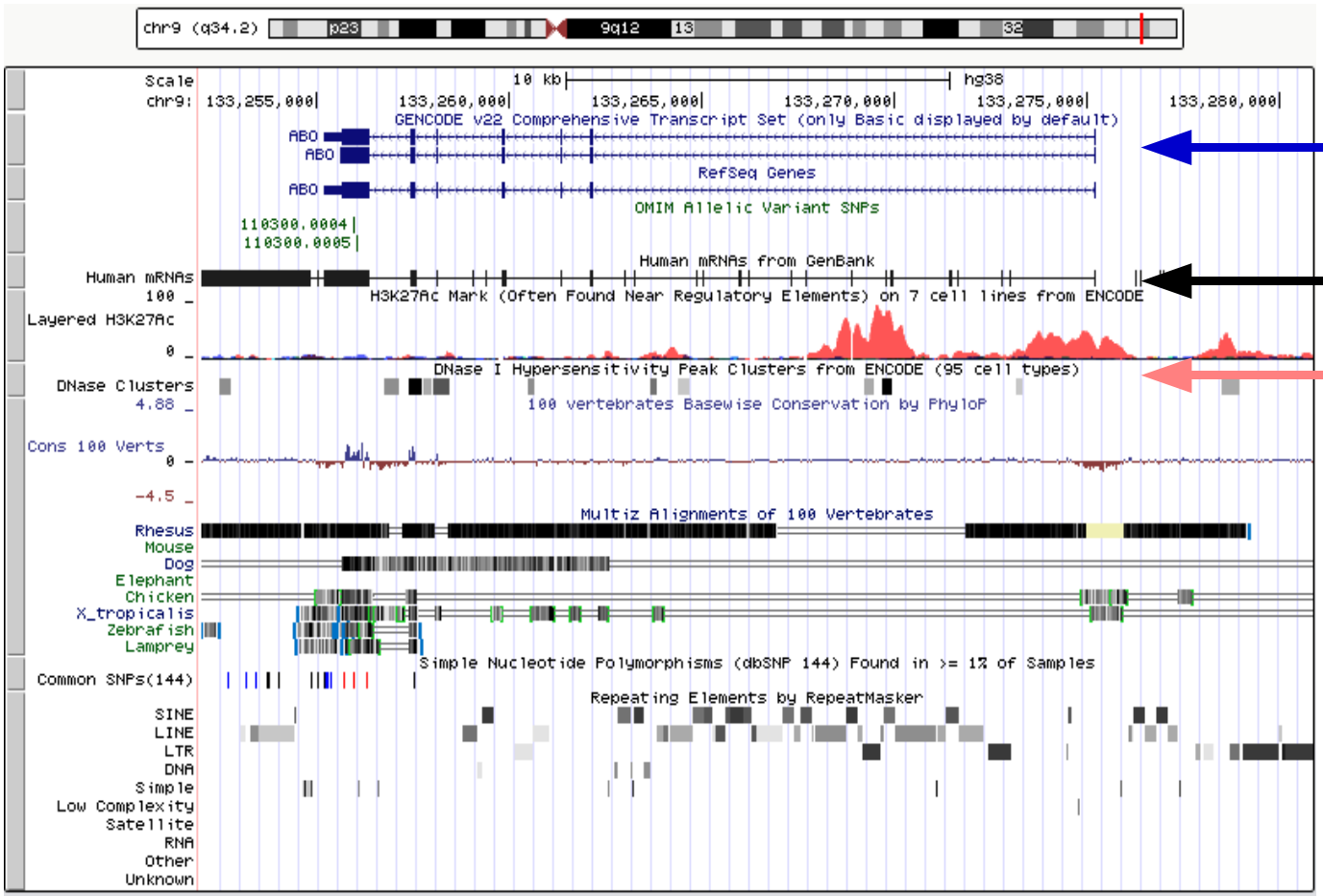
A.Nag *et al.* eLife 2013

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



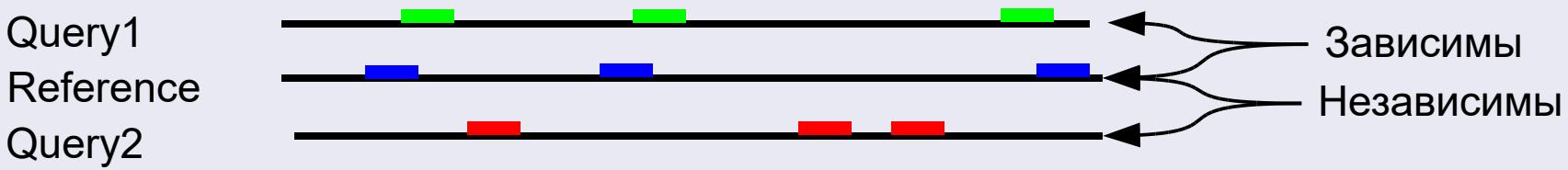
Гены

мРНК

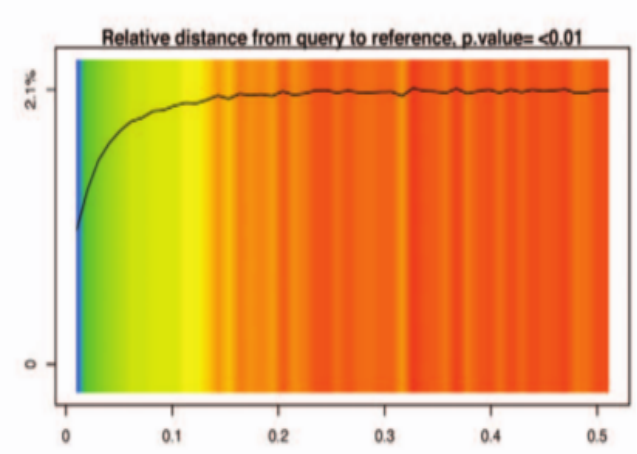
H3K27ac

Анализ корреляций:**Genometricorr**

сравнение наборов интервалов



Определяется статистическая значимость пространственной колокализации двух наборов интервалов.



Chikina M.D., Troyanskaya O.G.; Bioinformatics (2012)

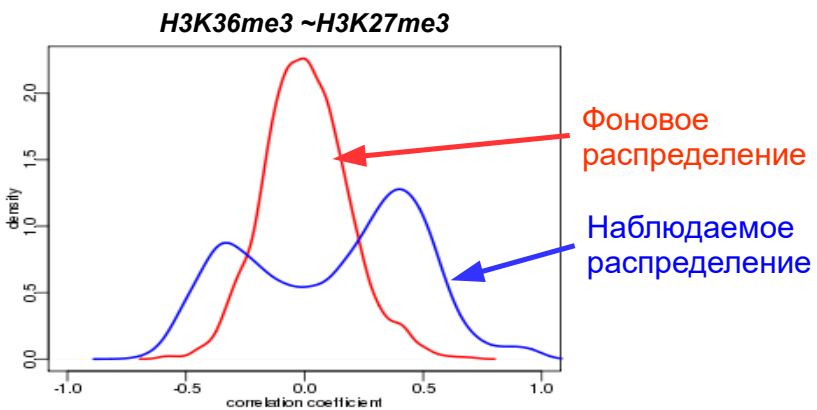
Favorov A., et al. PLoS Comput Biol (2012)

Анализ корреляций: **Stereogene**

Кореляция с ядром $\rho(t)$

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

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



Эпигенетические подписи

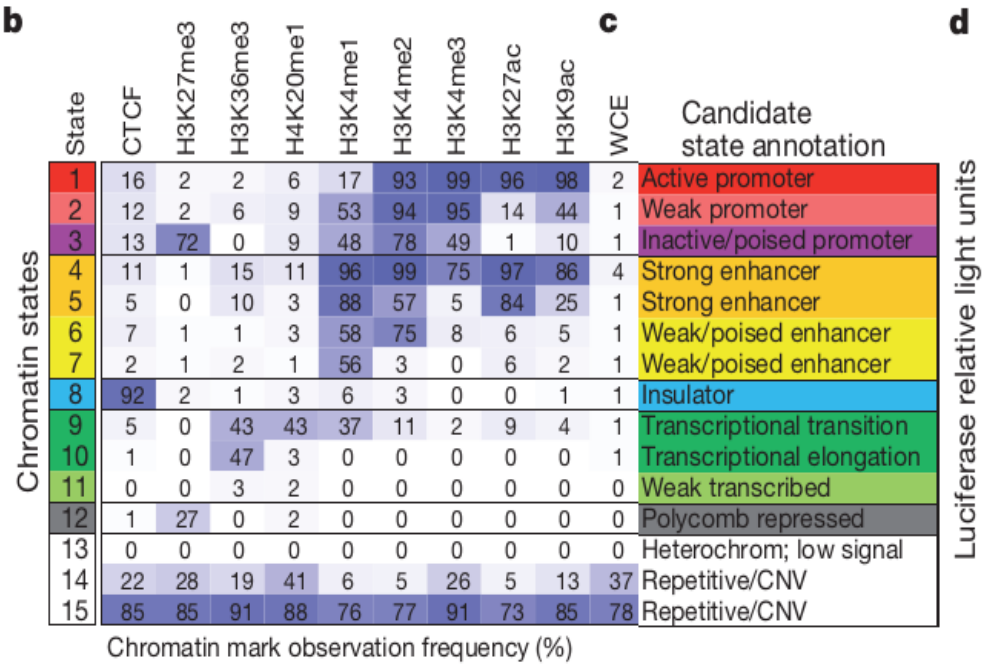
Процедура

- Дискретизуем 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) Энхансер
- (8) Инсулятор
- (9-11) Транскрипция
- (12) Polycomb
- (13) Гетерохроматин
- (14-15) Повторы



Ernst J., et al. Nature (2011)

Протезирование данных (imputation)

линейный предиктор

- Пусть у нас есть треки про n хромосомных меток.
- Построим линейную модель $f_n(x) = \sum a_i f_i(x)$
- Обучим эту модель на некотором наборе данных

Этот предиктор можно использовать для восстановления неизвестных треков и для уточнения "плохих"треков

Ernst J., Kellis M.; Nat Biotechnol (2015)