

Эпигенетика, РНК и все такое

Миронов Андрей Александрович
Факультет биоинженерии и биоинформатики МГУ

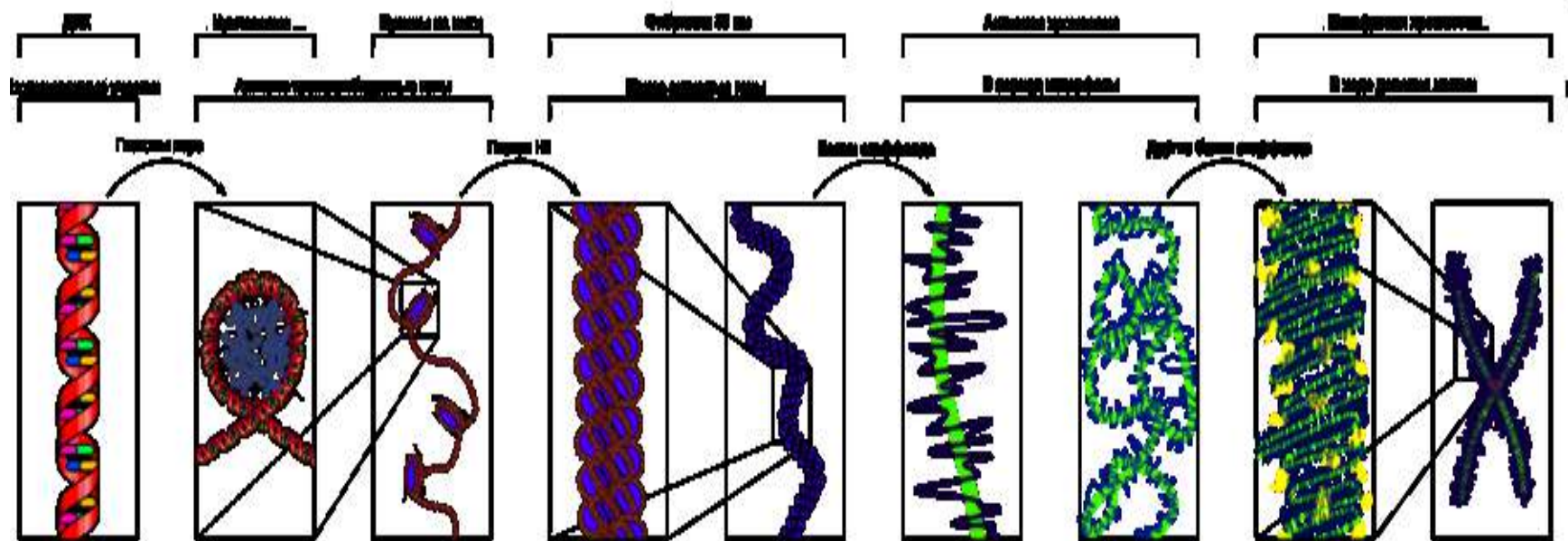
летняя школа по биоинформатике. Рощино 2014.

План

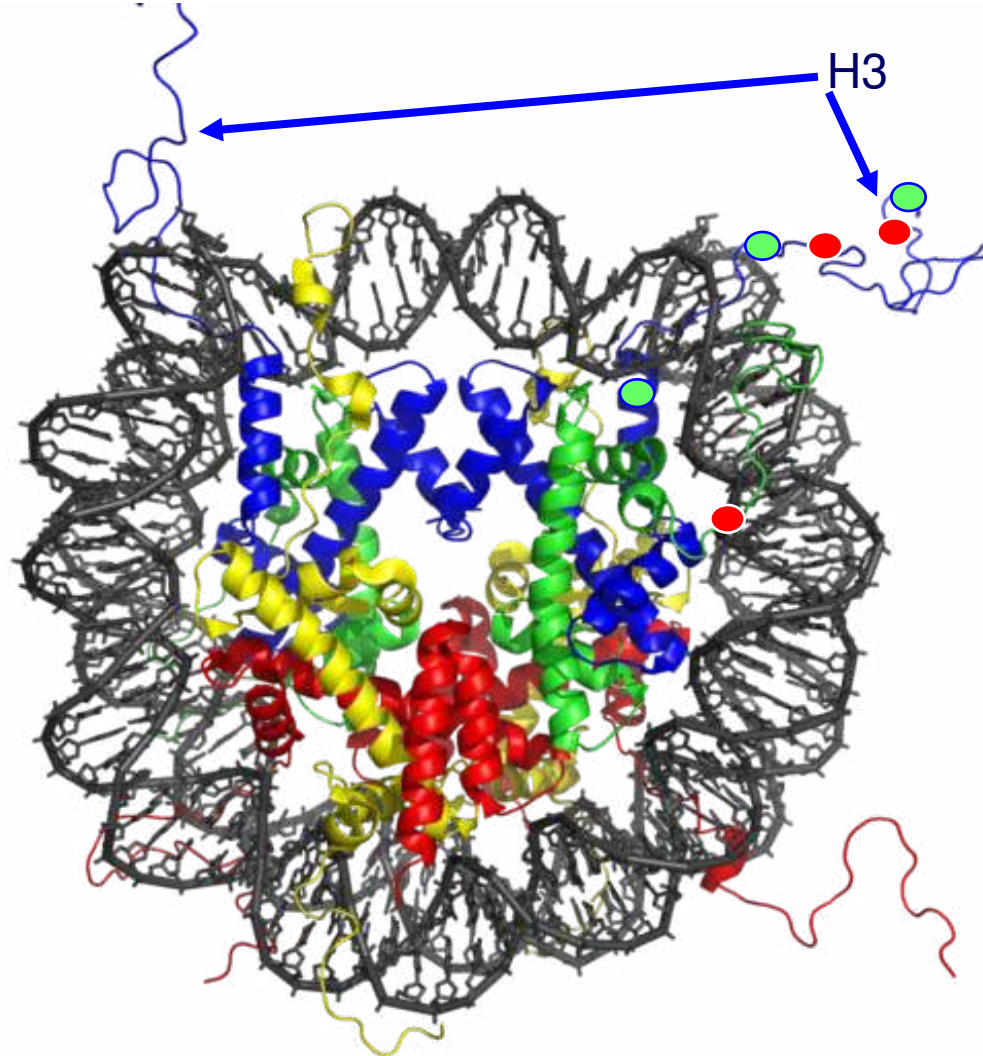
Эпигеномные модификации

- Источники данных
- Роль в экспрессии
- Роль в сплайсинге
- Роль в импринтинге и т. п.
- Модели и анализ данных
- РНК и хроматин

Уровни структурной организации хроматина



Структура нуклеосомы



Эпигенетика

- Модификации
ГИСТОНОВ
 - Метитлирование
 - Ацетилирование
 - Фрсфорилирование
 - Много чего еще

- Метилирование
ДНК по CpG

Регуляция экспрессии генов

Транскрипция

Инсуляторы

Сплайсинг

3D — структура хроматина

Источники данных

- **Модификации гистонов:**
 - ChIP-seq
- **Метилирование ДНК:**
 - Рестриктазы, специфичные к CpG:
 - HpaII vs MspI (C^m-чувствительная и устойчивая)
 - Бисульфитная конверсия
 - MeDIP
- **Доступность хроматина: DNaseI; FAIRE-seq**
- **Pol-II ChIP-seq**
- **CTCF**
- **Когезины: ChIP-seq на Smc3, Rad21**

Роль в экспрессии

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

На самом деле Эухроматин и гетерохроматин — граничные состояния хроматина.

Основные модификации H3 me

	Пром.	Энх.	Транс.	CpG-me	Гетеро.	Ploucmb
<u>H3K4me1</u>	+	+++				
<u>H3K4me2</u>	+	+				
<u>H3K4me3</u>	+++	+				
<u>H3K9me1</u>	+				++	
<u>H3K9me2,3</u>				+	++	
<u>H3K27me1</u>			+			
<u>H3K27me2,3</u>				++	+	++
<u>H3K36me3</u>			++			
<u>H3K79meX</u>			++		+	

Другие модификации

Ацетилирование H2A, H2B, H3, H4	Активация транскрипции, позиционирование нуклеосом, репарация ДНК
H4R3me	Активация транскрипции.
H4K20me1	Подавление транскрипции
H4K20me3	Активация транскрипции. Check point
H4K59me	Подавление транскрипции

Прочие модификации:

Фосфорилирование, Биотинилирование,
Убиквитинирование,...

Промоторы

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

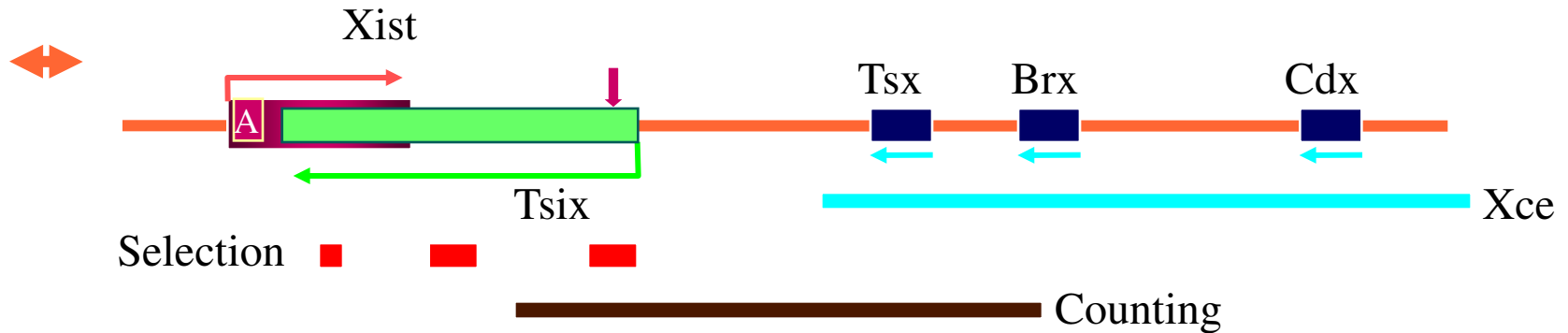
Экспрессия

- Для большинства позвоночных распределение H3K4me3 повторяет распределение CpG островов
- Промоторы генов развития ассоциированы с репрессивной меткой H3K27me3 и с активной меткой H3K4me3
- метилирование отвечает за создание длительного репрессированного состояния:
молчание гена → H3K27me3 → метилирование
- Эnhансеры обогащены H3K4me1, H3K4me2 и H3K27ac, но не H3K4me3
- CTCF участвует в распространении дальнего Эnhансер-промоторного взаимодействия

Крупномасштабные изменения хроматина

- Уровень Pol-II 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
- Дальние взаимодействия мало изучены. По-видимому, CTCF, Cohesin, малые РНК, играют важную роль в этом процессе.

Xist: X-inactivation specific transcript



Xist – ген ncRNA с интронами, быть может с альт. сплайсингом. Содержит много моб. элементов и повторов. Повтор **A** важен для инициации инактивации.

Tsix – Anti-sense Xist.

DXPas34 – мини-сателлитный маркер. Имеет разный уровень метилирования

Tsx, Brx, Cdx – белок-кодирующие гены (Чел. : Tsx – псевдоген)

Xce – локус, который влияет на выбор (генетические данные — разные аллели имеют разное предпочтение)

Выбор – локусы, влияющие на выбор

Counting – локус, который подсчитывает число активных X-хромосом.

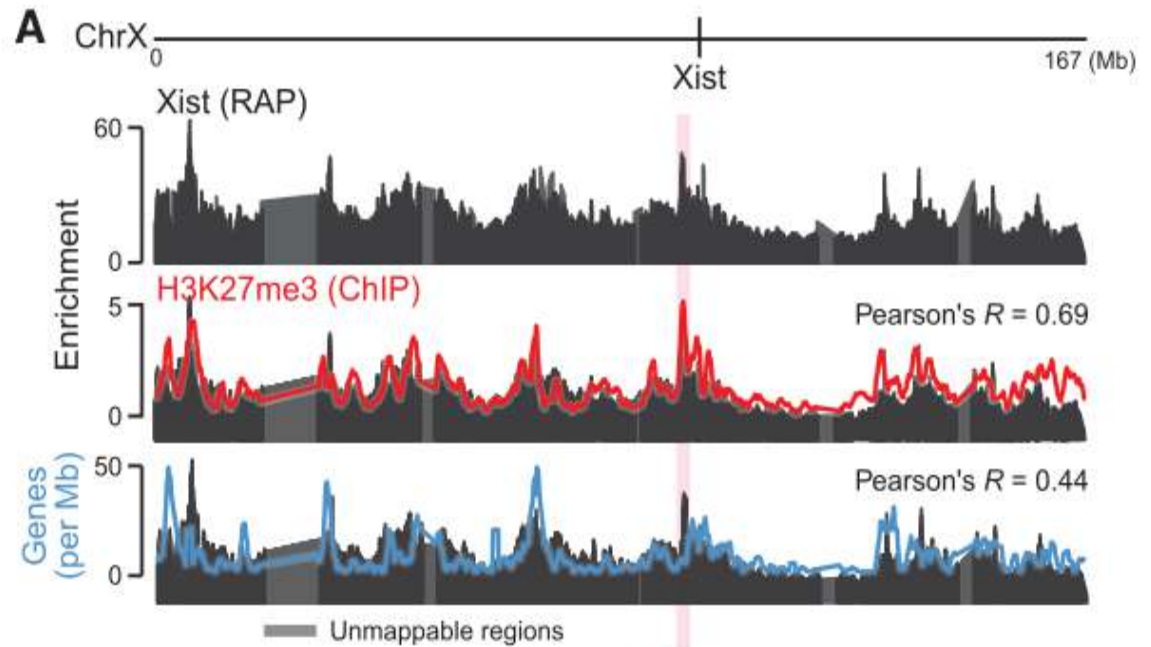
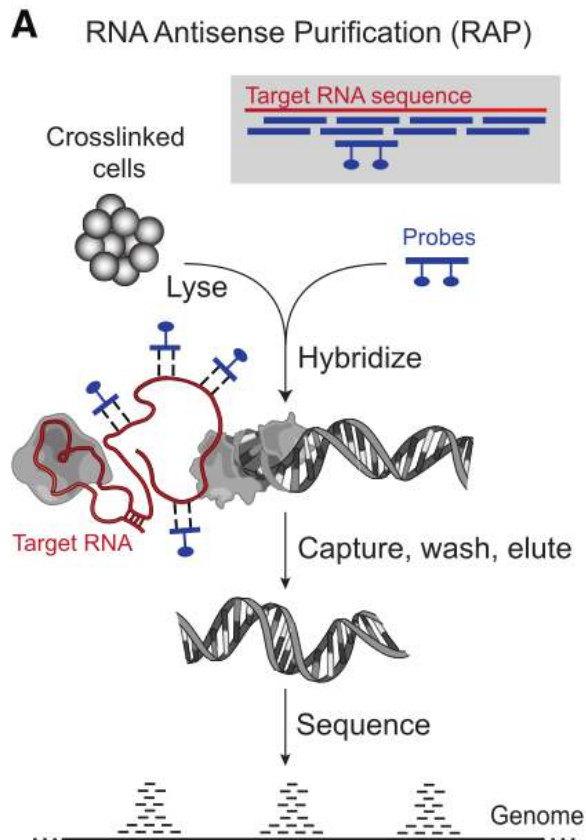
Инактивация X-chromosome с помощью *Xist*

- Выбор – конкуренция *Xist*–*Tsix*. Аллельное состояние локуса *Xce* влияет на выбор (*Xce* не содержит *Xist* и *Tsix*).
- Инициация:
 - Замена H2A →macroH2A
 - Метилирование H3K27→H3K27me
 - Метилирование промоторов
 - Компактизация хроматина
 - Некоторые гены продолжают транскрибироваться с не активной X-хромосомы
 - Важен повтор A. Участвует система ***Polycomb***
 - По-видимому важно взаимодействие *Xist* с матриксом
- Инактивация начинается в раннем эмбрионе и затем поддерживается (как?)



Patchy paint cats determined by X chromosome

Xist



Engreitz J.M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. (2013) Science; v.341

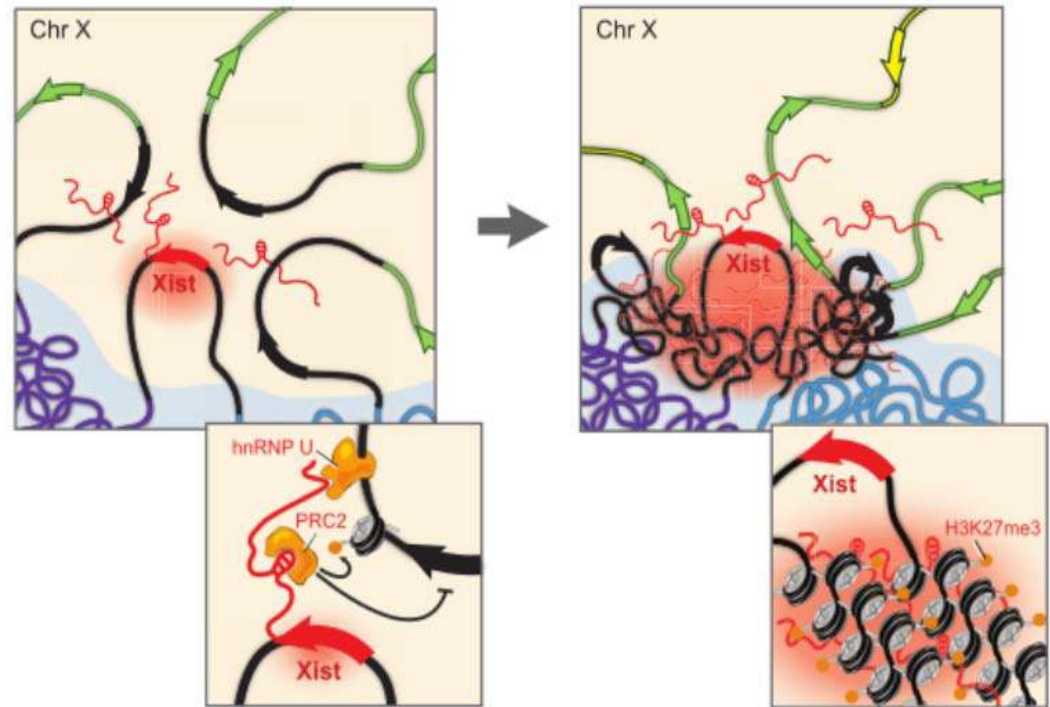
Xist

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.



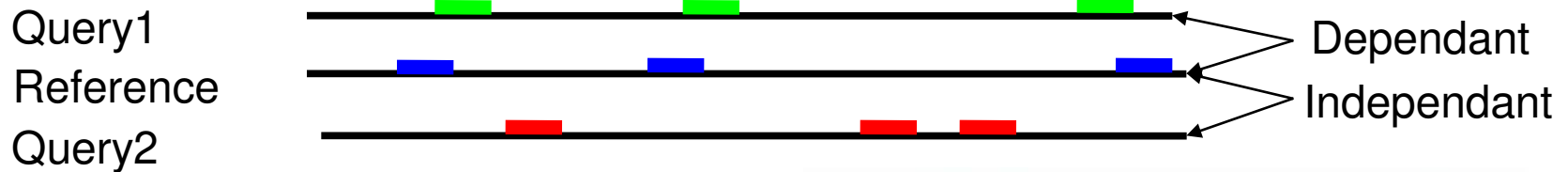
Engreitz J.M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. (2013) Science; v.341

Консервативность

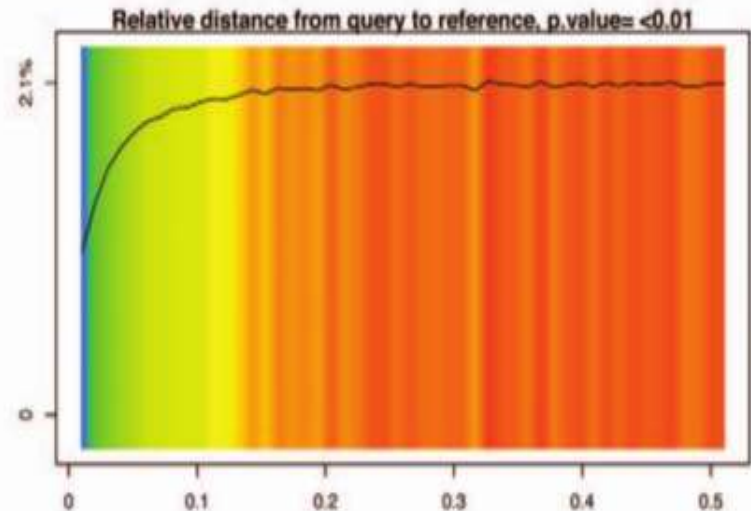
Сравнение эпигеномов человек-мышь-свинья

- лишь 5 % из OCT4 -и NANOG сайтов связывания есть в гомологичных последовательностях (ES)
- 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

Корреляции (Genometricorr)



- Рассмотрены два набора интервалов. Определяется статистическая значимость их зависимости.



Chikina M.D., Troyanskaya O.G.; An effective statistical evaluation of ChIPseq dataset similarity. (2012) Bioinformatics;

Favorov A., et al. Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput Biol (2012), <http://genometricorr.sourceforge.net/>

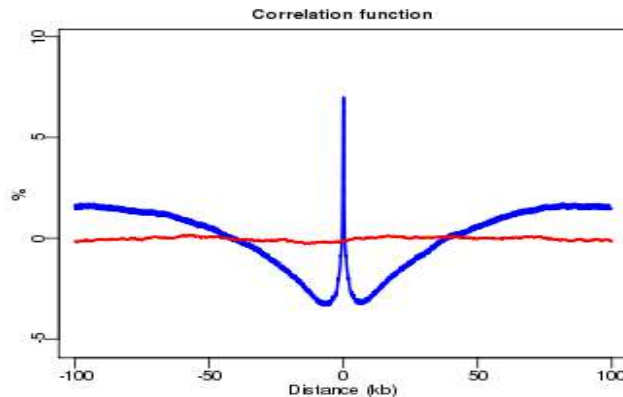
Корреляции (StereoGene)

Пусть есть две разметки (профиля) $f(x)$ и $g(x)$
Тогда для их сравнения можно использовать меру

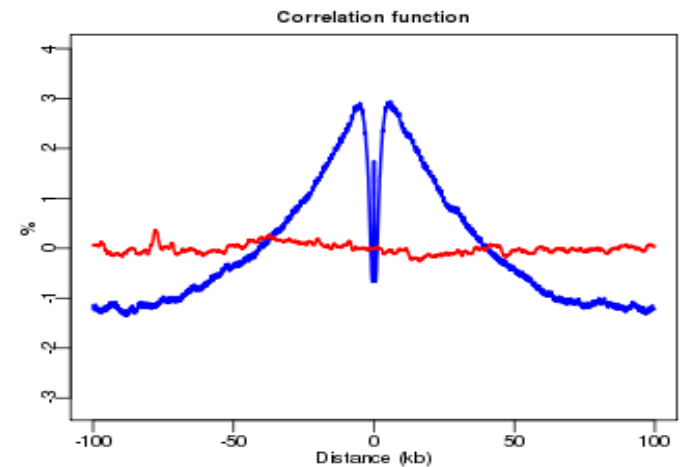
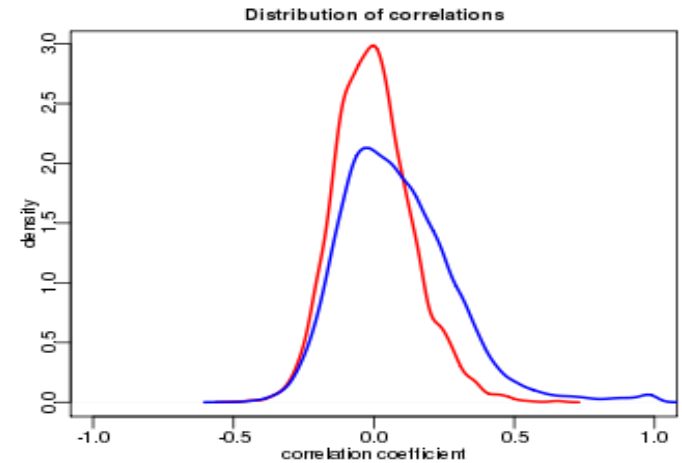
$$Q(f, g) = \int_0^L \int_0^L f(x)g(y)\rho(x - y)dx dy$$

а для пространственной корреляции

использовать $c(x) = \int \tilde{f}(t)\tilde{g}(t - x)dt$



H3K36me3 vs H3K27me3



H3K36me3 vs H3K27ac

Сплайсинг

- Si-РНК был использован для регуляции альтернативного сплайсинга. 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) (Liso 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 γ усиливают включение альтернативного экзона

Кинетическая модель

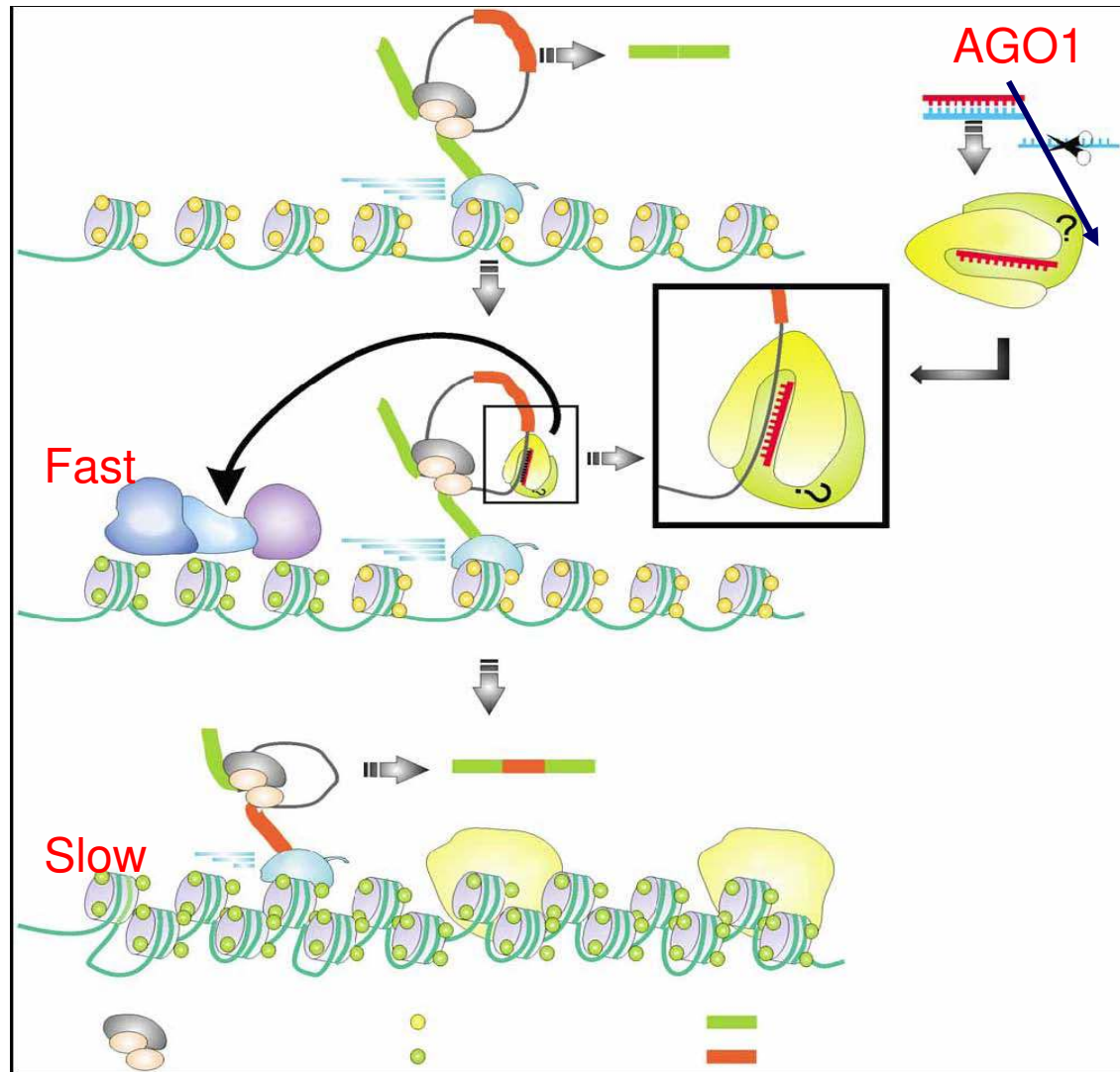
Добавление si-РНК:

Взаимодействие с мРНК

Изменение статуса метилирования гистонов

Модификация хроматина

Белок HP1- α конденсированного хроматина



Полиаденилирование

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

Модели

Состояния

Активный промотор

Слабый промотор

Неактивный промотор

Сильный энхансер (1,2)

Слабый энхансер (1,2)

Инсулятор

Транскрипция (1,2)

Слабая транскрипция

Polysomb

Гетерохроматин

Повторы

Скрытая Марковская Модель.

Chromatin states	State	Chromatin mark observation frequency (%)									Candidate state annotation	Luciferase relative light units
		CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac		
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	

РНК зависимая регуляция

- **PTGR – пост-транскрипционная регуляция**
 - siRNA – деградация mRNA с помощью RISC
 - miRNA – подавление инициации трнсляции и элонгации, стабильность
- **TGR – Регуляция транскрипции**
 - Хроматин, РНК-зависимое метилирование ДНК
 - Инициация транскрипции
 - Элонгация транскрипции
- **RITS** (RNA-induced initiation of transcriptional gene silencing)

Note: RNA-dependent RNA polymerase (RdRp):
The enzyme found in plants, *C.elegance*, *drosophila*
Mammals: RdRp activity
dsRNA amplification. Necessary for many processes

Типы регуляторных РНК

- **miRNA**
- **piRNA – PIWI** – связывающая РНК
- **C/D РНК** (80..300 nucl.) – содержит взаимно-комплементарные блоки C & D вблизи концов, ассоциированы с фибронектином (гликопротеиновым комплексом):
 - sno – модификация рРНК, некоторые другие функции
 - miRNA и piRNA.
 - Некоторые вовлечены в импринтинг
 - Некоторые вовлечены в редактирование РНК(A→I)
- **Длинные некодирующие РНК**
 - Xist/Tsix – инактивируют вторую копию X-хромосомы
 - Kcnq1/Kcnq1ot1, Air – вовлечены в импринтинг некоторых генов
 - HOTAIR – вовлечено в регуляцию гомеобоксных генов
 - lincRNA – длинные некодирующие РНК (about 4000 transcripts).

Импринтинг

- Селективная экспрессия исключительно отцовских либо материнских генов
- **Человек:** 80 генов под импринтингом
- ICE (ICR) - **I**mprinting **C**ontrol **E**lement (**R**egion)
- DMR - **D**ifferentially **M**ethylated **R**egions
- Каждый кластер генов под импринтингом экспрессирует одну или более длинных нкРНК (часто анти-сенс)
- Подавление транскрипции этих РНК восстанавливает экспрессию репрессированных аллелей

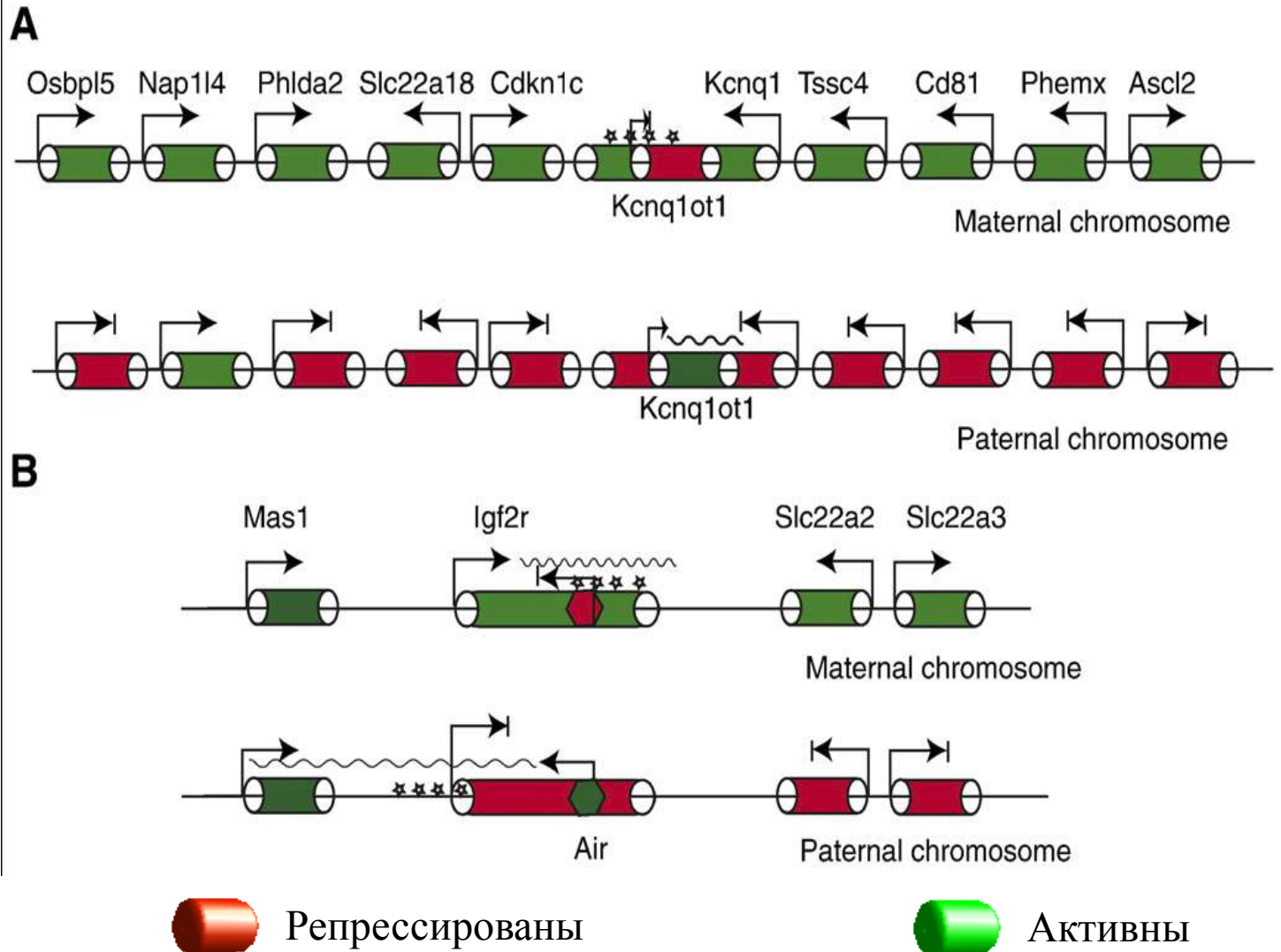
Импринтинг.

Цис-регуляция: Kcnq1 & Air RNA

Kcnq1ot1:
L=60Kb

Air:
L=108 Kb – не
сплайсировано
в ядре

L=0.5-1.5 Kb –
сплайсировано
в цитоплазме



Система Polycomb / Trithorax

- Polycomb group = PcG; Trithorax group=TrxG
- Система ассоциирована с развитием и гомеобоксами (Hox). Поддерживает эпигенетическую память.
- PcG - репрессор, TrxG - активатор.
- PRE / TRE - Polycomb / Trithorax Response Elements
- Гипотеза: поддержание эпигенетического статуса в митозе обусловлено локальными концентрациями.
- Наблюдение: PRE / TRE – репрессия/активация зависит от нкРНК

РНК-зависимая репарация ДНК

Модель - *Saccharomyces cerevisiae*

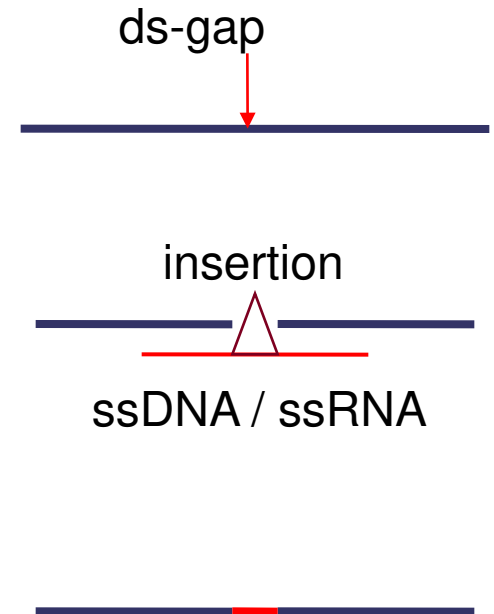
Репарация двунитевых разрывов

Обычный механизм –

рекомбинация с использованием
второй хромосомы

Сделано:

- Двунитевые разрывы в гене LEU2
- Добавлены однонитевые ДНК или РНК
=> наблюдали восстановление (в том числе и со вставкой кодона).



РНК в перетройках генома

Инфузории: Микронуклеус и Макронуклеус

Тетрахимена: малые РНК (piwi - like systems)
ответственны за нарезку.

Oxytricha trifallax: малые РНК также
обеспечивают перестановки и инверсии
(показано с помощью микроинъекций).

Роль РНК в метилировании ДНК в растениях

Сравнили два экотипа. Смотрели на высоко метилированные сегменты.

Показано:

Разные пулы малых РНК, комплементарных сайтам метилирования.