long non coding RNAs

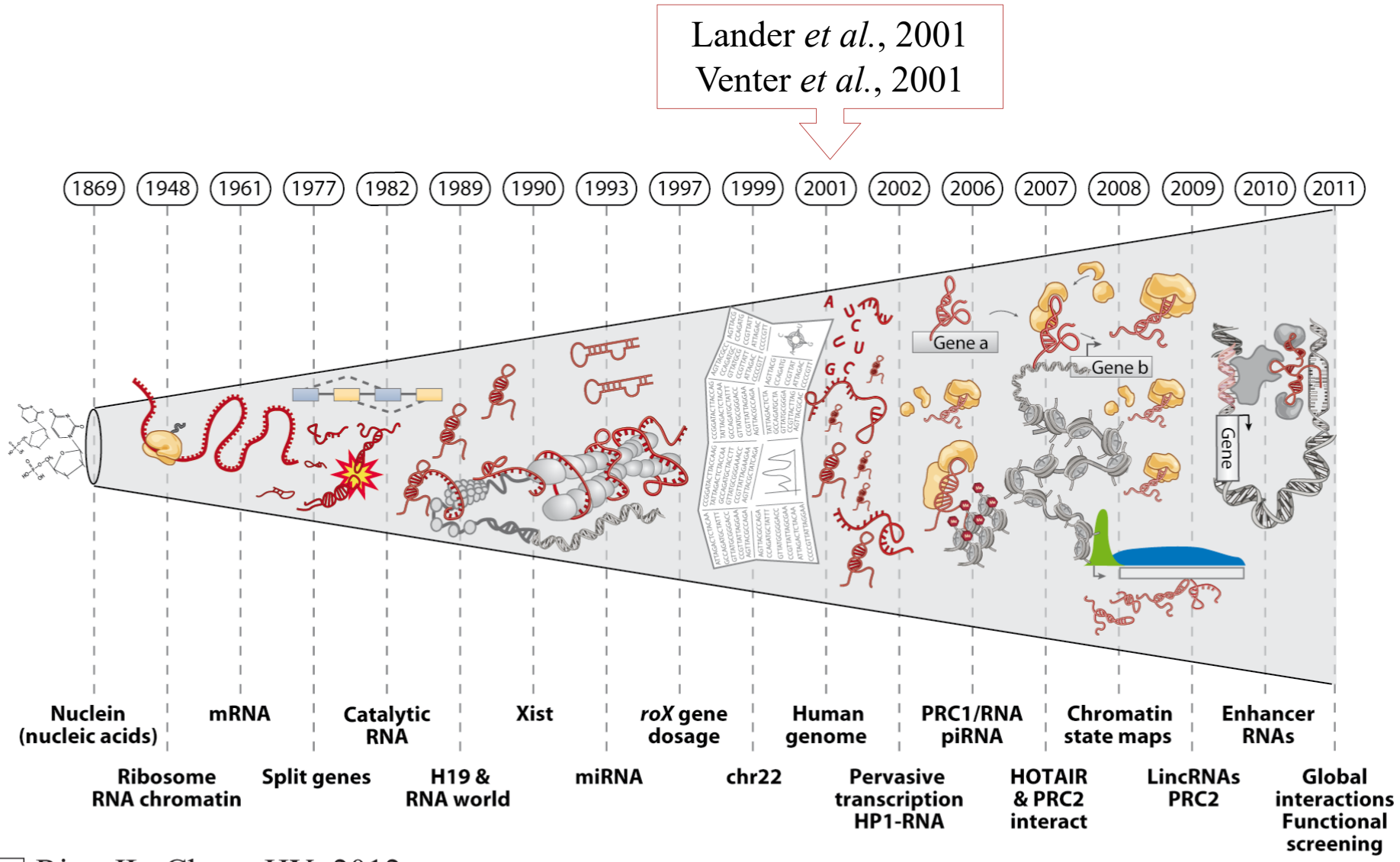
The GENOMIC ERA

– at the beginning of the XXI century, one of the major question was:

how many genes in the human genome?

The huge popular interest in counting the number of genes present in the human genome led even to a public wager named Gene Sweepstake, with an extensive media coverage (nyt Wade 2003)

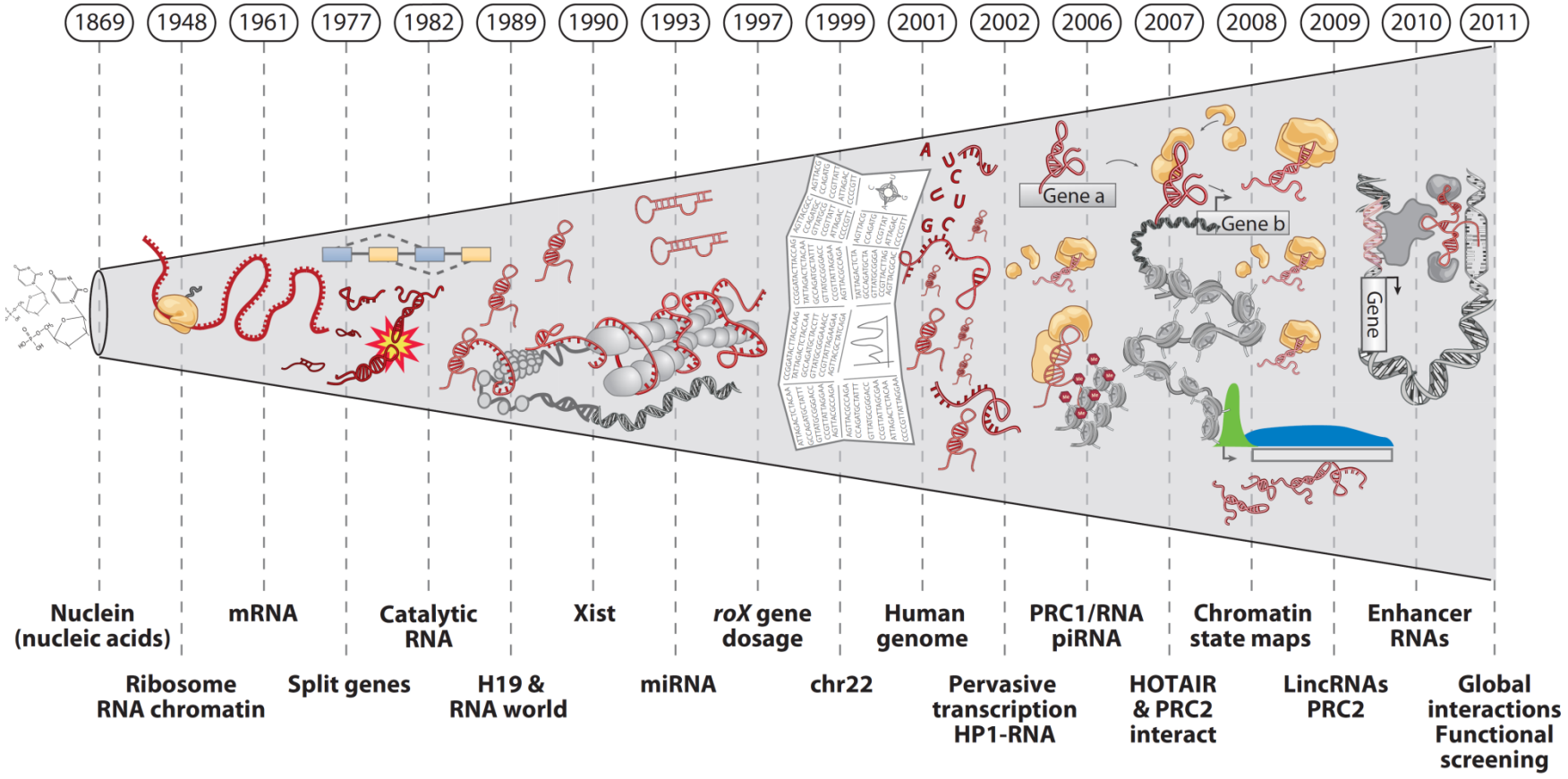
First sequence of the human genome



Rinn JL, Chang HY. 2012.
Annu. Rev. Biochem. 81:145–66

Transcriptome analysis of full length mRNAs

The FANTOM3 Consortium, 2005



Rinn & Chang *Annu Rev Biochem* 2012

Transcriptome analysis

ENCODE

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the **human** genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

FANTOM5

A promoter level mammalian expression atlas
Alistair R.R. Forrest *et al.*, *submitted*

CAGE analysis of the following libraries:

573 human primary cell samples

128 mouse primary cell samples

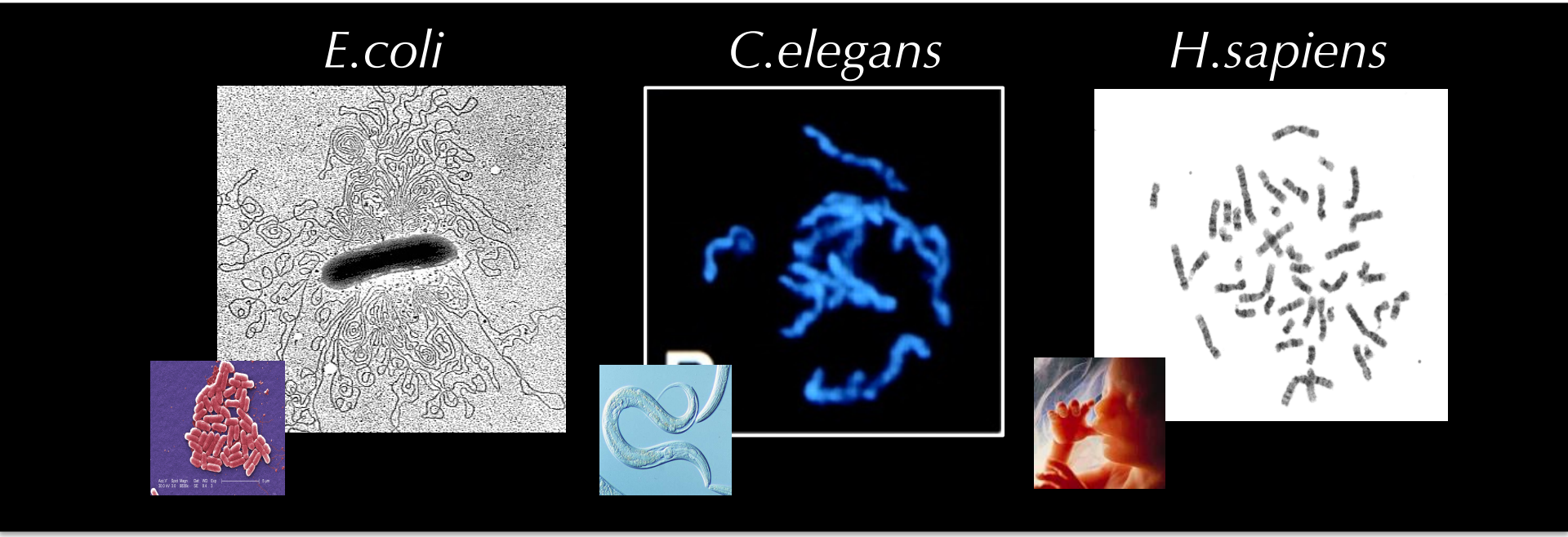
250 different cancer cell lines samples

152 human post-mortem tissues samples

271 mouse developmental tissue samples

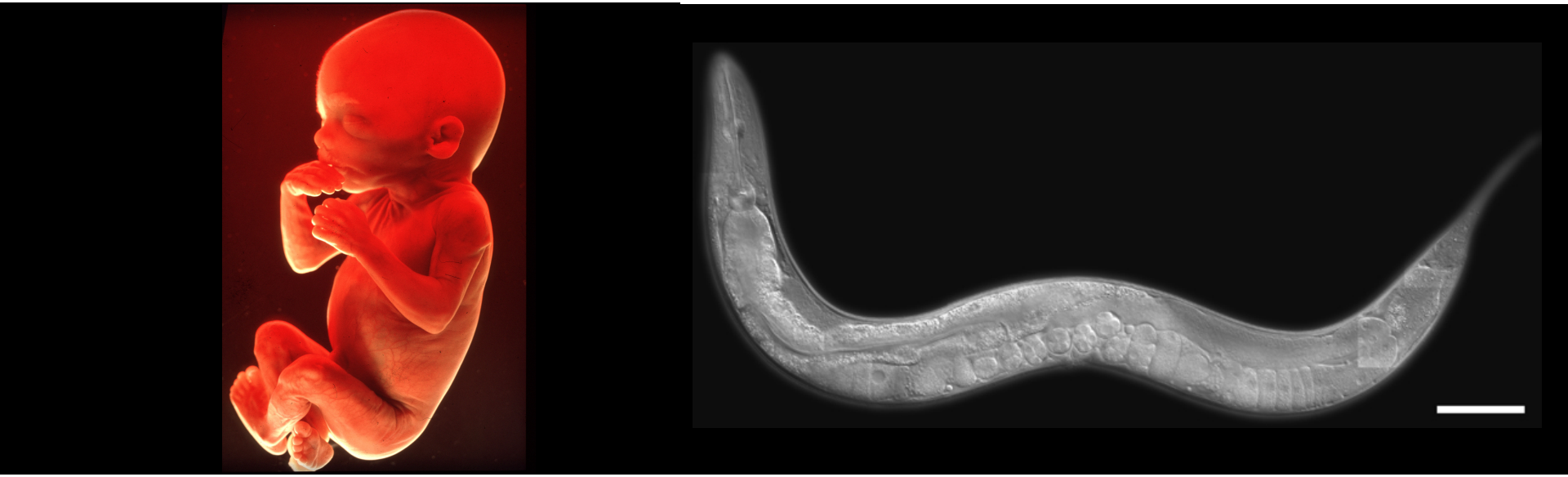
22000 genes encoding for proteins

Genome size and organism complexity



Genome	$5x10^6$ bp	$1x10^8$ bp	$3x10^9$ bp
Chromosomes	1	6	23
Coding genes	6692	20541	21995

The genetic basis of developmental complexity

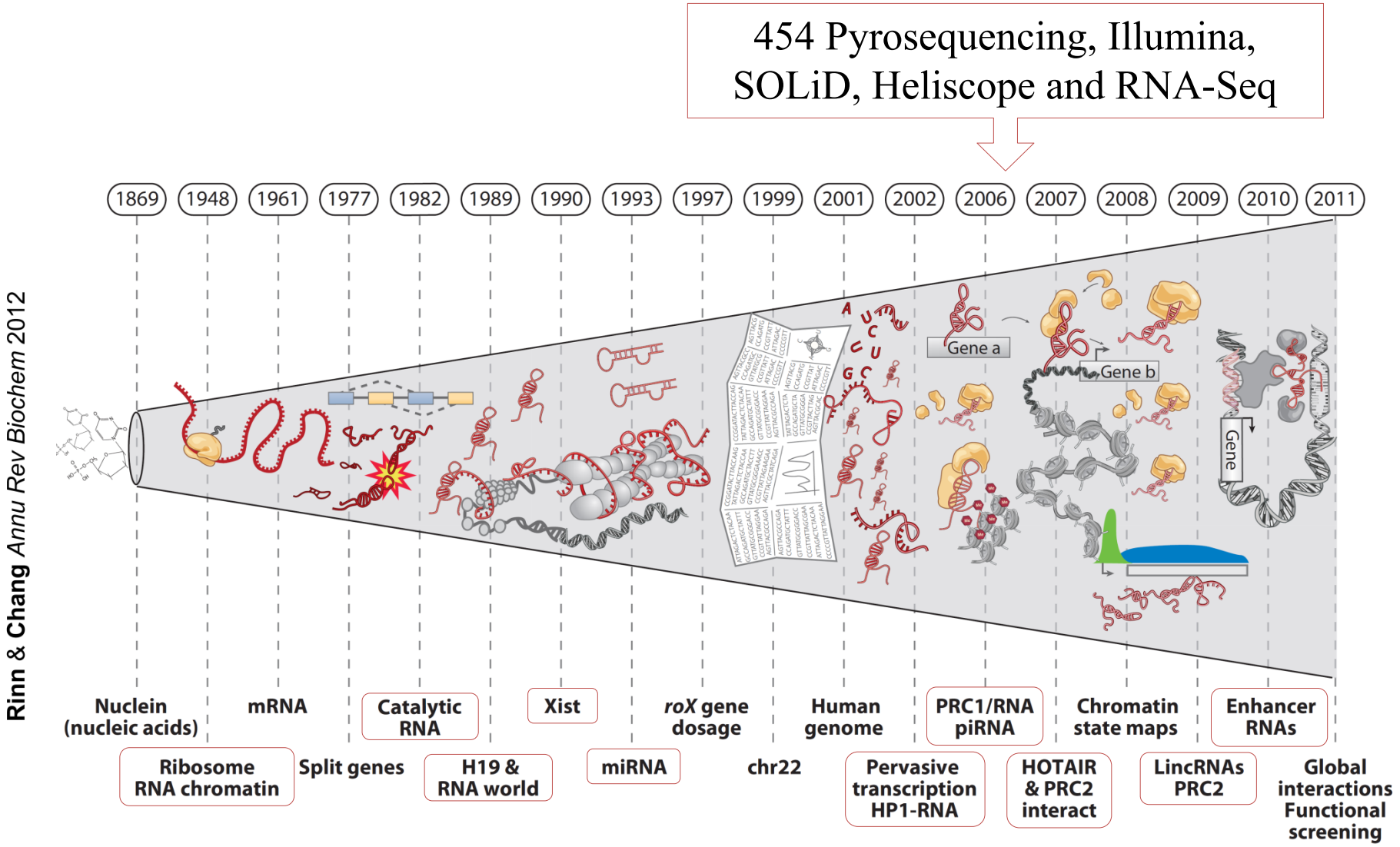


C.elegans - **1000 cells**
H.sapiens - **10^{14} cells - and 10^{11} neurons!!!**

Both have approximately **20.000 proteins**

- Most of the proteins are orthologous and have similar functions from nematodes to humans, and many are common with yeast.
- Where is the information that programs our complexity?

Deep sequencing technologies – identification of low abundance transcripts



Genome Organization

ENCODE

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the **human** genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

FANTOM5

A promoter level mammalian expression atlas
Alistair R.R. Forrest *et al.*, *submitted*

CAGE analysis of the following libraries:

573 human primary cell samples

128 mouse primary cell samples

250 different cancer cell lines samples

152 human post-mortem tissues samples

271 mouse developmental tissue samples

22000 genes that encodes for proteins

>40000 long non-coding RNAs and growing…….
>50% of the genome is functional

Genome Organization

Antisense Transcription in the Mammalian Transcriptome

RIKEN Genome Exploration Research Group and
Genome Science Group (Genome Network Project Core Group)
and the FANTOM Consortium

Antisense transcription (transcription from the opposite strand to a protein-coding or sense strand) has been ascribed roles in gene regulation involving degradation of the corresponding sense transcripts (RNA interference), as well as gene silencing at the chromatin level. Global transcriptome analysis provides evidence that a large proportion of the genome can produce transcripts from both strands, and that antisense transcripts commonly link neighboring "genes" in complex loci into chains of linked transcriptional units. Expression profiling reveals frequent concordant regulation of sense/antisense pairs. We present experimental evidence that perturbation of an antisense RNA can alter the expression of sense messenger RNAs, suggesting that antisense transcription contributes to control of transcriptional outputs in mammals.

>70% of protein encoding genes present antisense transcription

The Antisense Transcriptomes of Human Cells

Yiping He, Bert Vogelstein, Victor E. Velculescu, Nickolas Papadopoulos,* Kenneth W. Kinzler

Transcription in mammalian cells can be assessed at a genome-wide level, but it has been difficult to reliably determine whether individual transcripts are derived from the plus or minus strands of chromosomes. This distinction can be critical for understanding the relationship between known transcripts (sense) and the complementary antisense transcripts that may regulate them. Here, we describe a technique that can be used to (i) identify the DNA strand of origin for any particular RNA transcript, and (ii) quantify the number of sense and antisense transcripts from expressed genes at a global level. We examined five different human cell types and in each case found evidence for antisense transcripts in 2900 to 6400 human genes. The distribution of antisense transcripts was distinct from that of sense transcripts, was nonrandom across the genome, and differed among cell types. Antisense transcripts thus appear to be a pervasive feature of human cells, which suggests that they are a fundamental component of gene regulation.

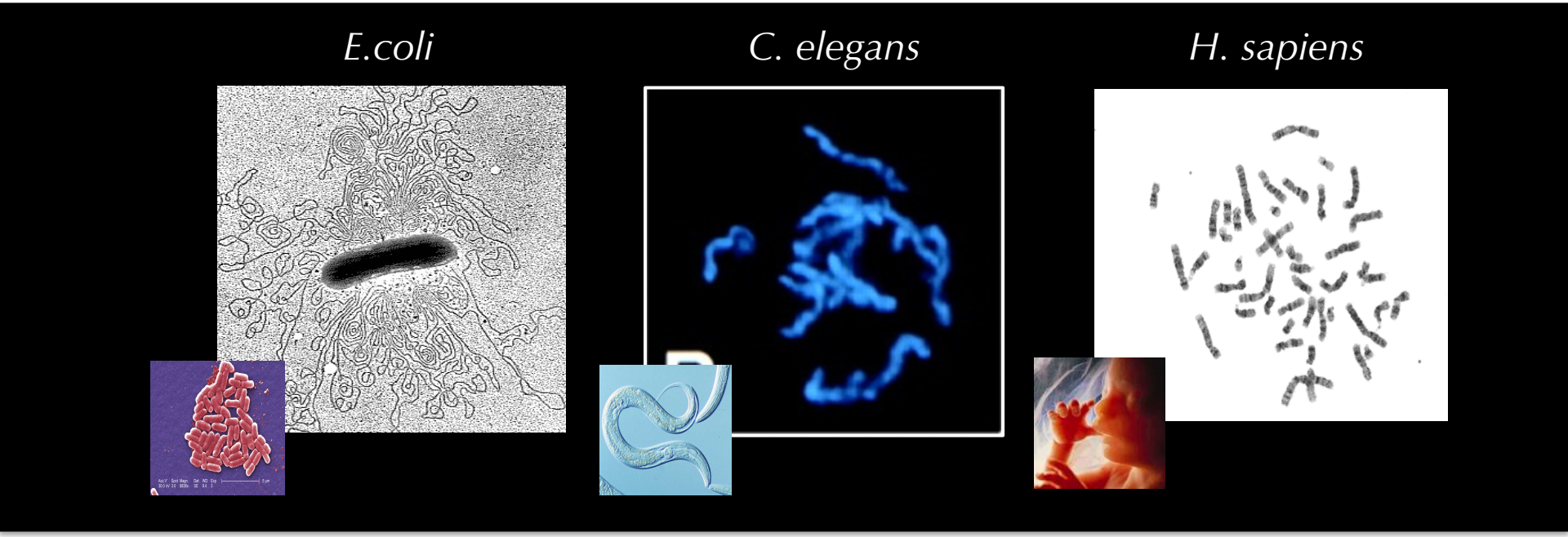
Genome-wide antisense transcription drives mRNA processing in bacteria

Iñigo Lasa[a,1,2], Alejandro Toledo-Arana[a,1], Alexander Dobin[b], Maite Villanueva[a], Igor Ruiz de los Mozos[a], Marta Vergara-Irigaray[a], Víctor Segura[c], Delphine Fagegaltier[b], José R. Penadés[d], Jaione Valle[a], Cristina Solano[a], and Thomas R. Gingeras[b,2]

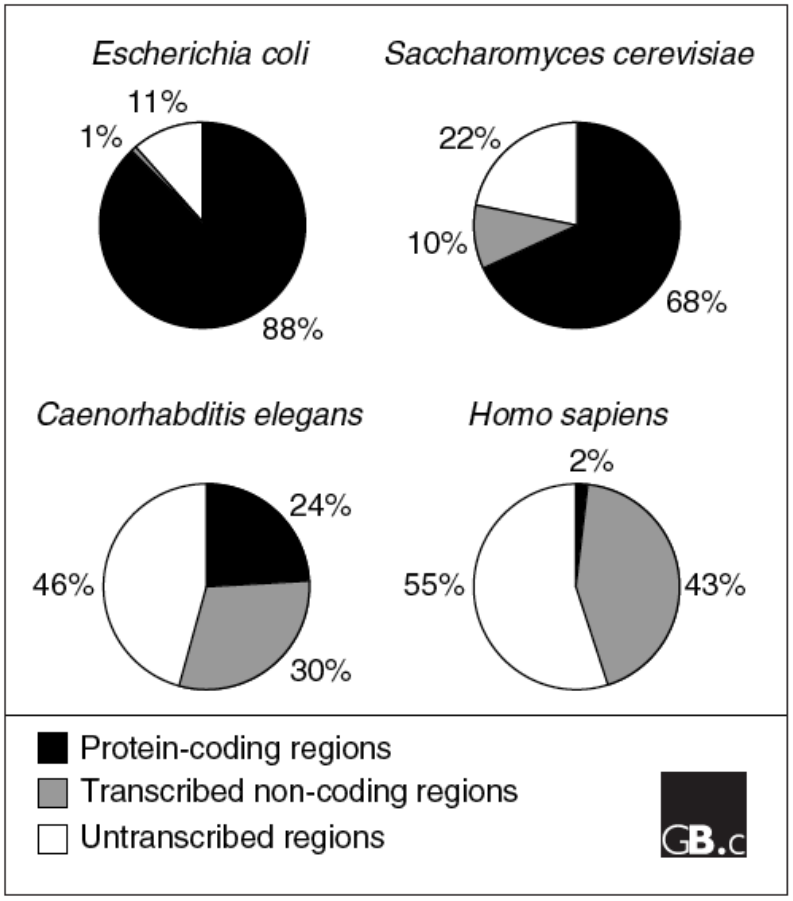
[a]Laboratory of Microbial Biofilms, Instituto de Agrobiotecnología, Consejo Superior de Investigaciones Científicas–Universidad Pública de Navarra–Gobierno de Navarra, 31006 Pamplona, Spain; [b]Laboratory of Functional Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; [c]Genomics, Proteomics and Bioinformatics Unit, Centro de Investigación Médica Aplicada, Universidad de Navarra, 31008 Pamplona, Spain; and [d]Instituto en Ganadería de Montaña-Consejo Superior de Investigaciones Científicas, 24346 León, Spain

Edited by Susan Gottesman, National Cancer Institute, Bethesda, MD, and approved November 8, 2011 (received for review August 19, 2011)

Genome size and organism complexity

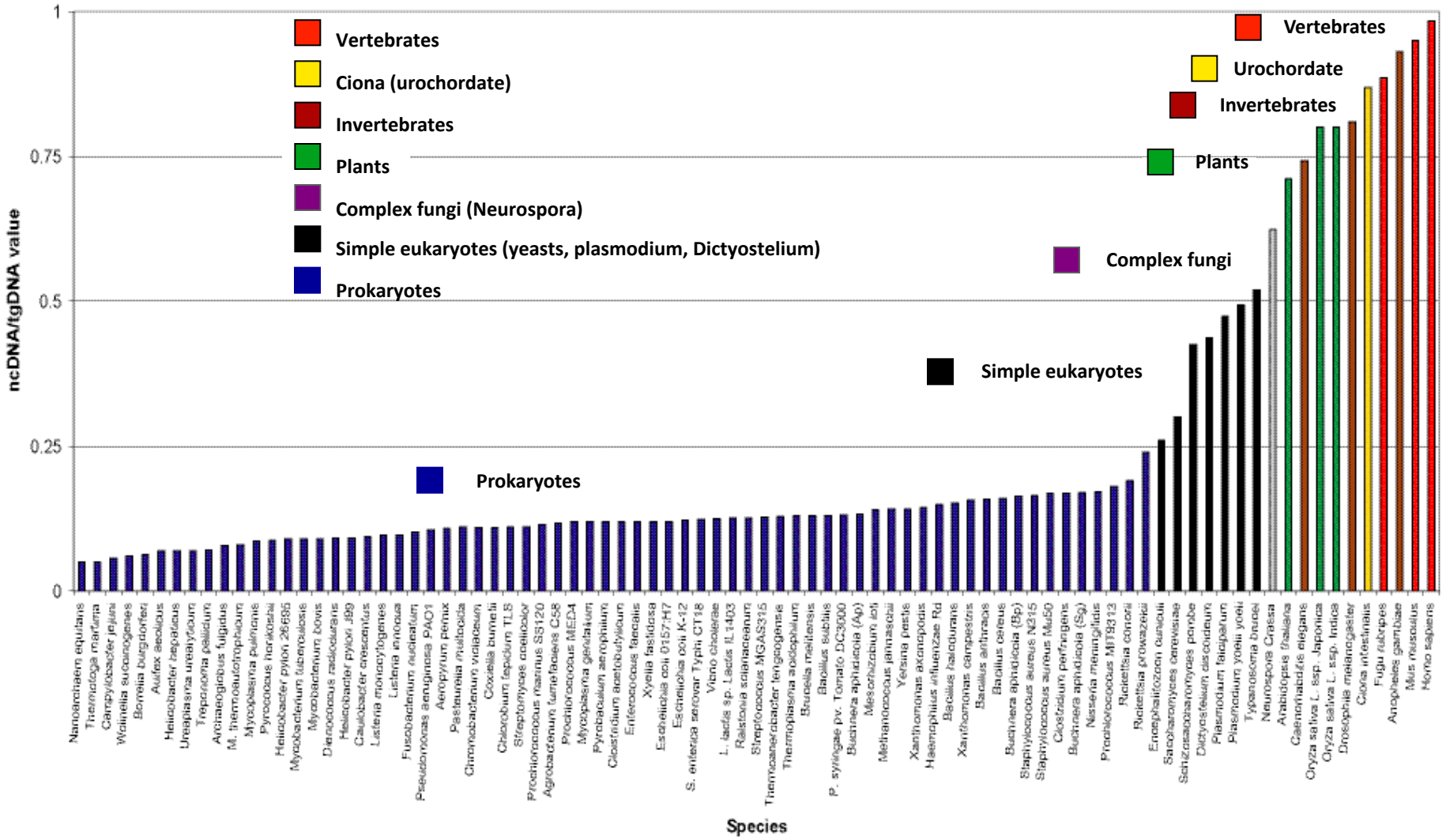


Genome	$5x10^6$ bp	$1x10^8$ bp	$3x10^9$ bp
Chromosomes	1	6	23
Coding genes	6692	20541	21995
ncDNA	5%	60%	98%
non-coding RNA genes	15	23136	ca. 40000
miRNAs	0	224	4274
pseudogenes	21	1522	10616



- Protein-coding genes can't account for all complexity
- ncRNAs represent the larger fraction of the human transcriptome

The proportion of noncoding DNA broadly increases with developmental complexity



J.S. Mattick *Nature Reviews Genetics* 5, 316-323 (2004).
R.J. Taft, M. Pheasant and J.S. Mattick, *Bioessays* 29, 288-299 (2007)

- **The major challenge that evolution had to overcome to evolve developmentally complex organisms was regulatory**

- **the barriers imposed by the rising cost of regulation could be overcome by moving to a hierarchical RNA-based regulatory system**

Scaling of regulatory architecture

- The response to this lack of gene scaling and limited proteomic diversification lies in the power of combinatoric control

- Transcriptional fators, alternative splicing

Dramatic expansion of regulatory complexity

Expansion in the number and complexity of cis-acting sequences recognized by transcriptional factors
The range of regulatory options scales factorially with the numbers of regulatory proteins (>1000 in humans and C.elegans)

The ***central dogma*** of molecular biology states that DNA is transcribed into RNA, which in turn is translated into proteins.

We now know, however, that as much as 50% of the transcriptome has no protein-coding potential, but rather represents an important class of regulatory molecules responsible for the fine-tuning of gene expression

There are several proposed mechanisms of action for lncRNAs which bring plasticity, adaptability and reactivity to genomic architecture and fine control over gene expression.

Although the central role of RNA in cellular functions and organismal evolution has been advocated periodically during the last 50 years, only recently has RNA received a remarkable level of attention from the scientific community.

Why RNA?

Economist.com

The RNA revolution

Biology's Big Bang

- many processes of gene expression regulation occur at the post-transcriptional level

- a whole universe of RNA - predominantly of the noncoding variety - has remained hidden from view, until recently…….many new and unexpected functions

- RNA molecules can be appropriately modified in order to interfere with gene expression in a sequence-specific way

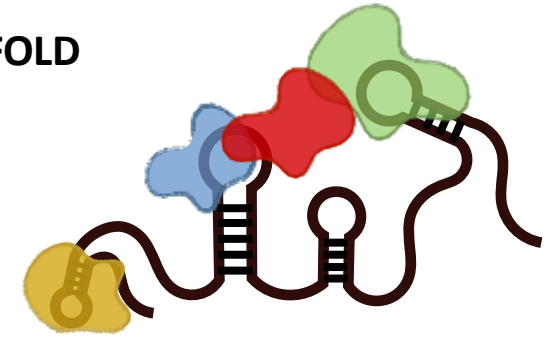
New definition of "GENE" – 2006

- More recently, the Sequence Ontology Consortium called the gene a "locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions" (Pearson 2006).

Models of lncRNA mechanisms of action

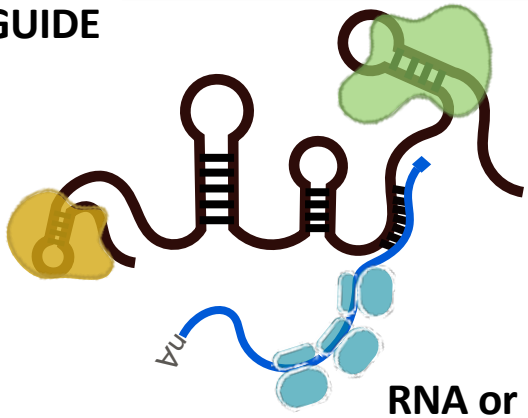
DUAL FUNCTION: lncRNAs can bind proteins as well as RNA/DNA

SCAFFOLD



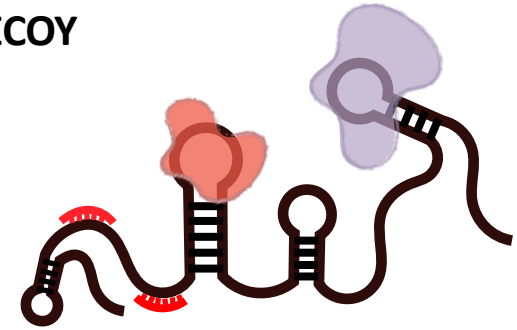
Ability to bring together proteins that normally would not interact each other: ***Increase in the number of protein combinations***

GUIDE



LncRNAs can act as guides to recruit proteins to mRNAs: ***modulation of mRNAs fate***

DECOY

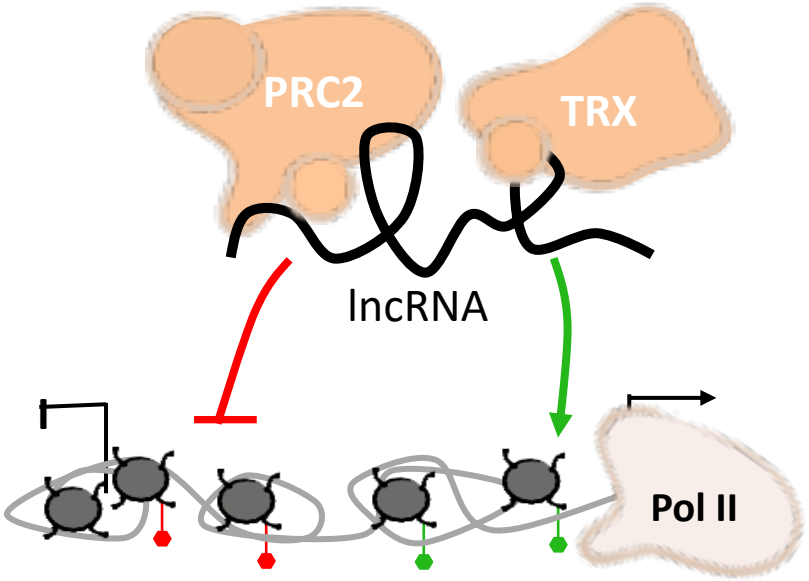


LncRNAs binding to RBPs and miRNAs can reduce the availability of these factors to mRNAs: ***modulation of mRNAs fate***

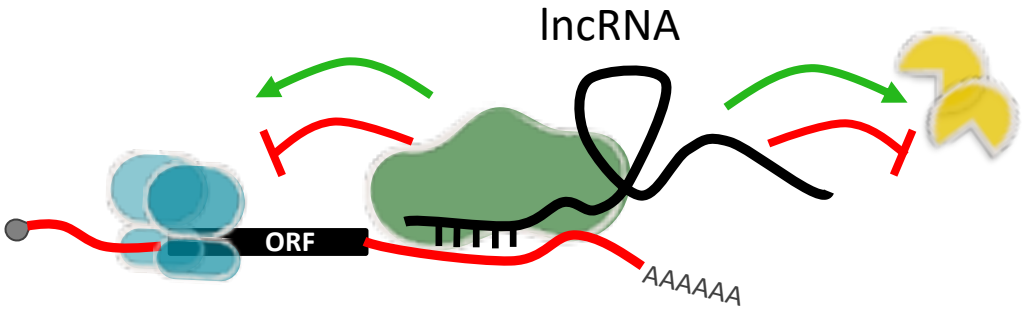
Models of lncRNA mechanisms of action

DUAL FUNCTION: lncRNAs can bind proteins as well as RNA/DNA

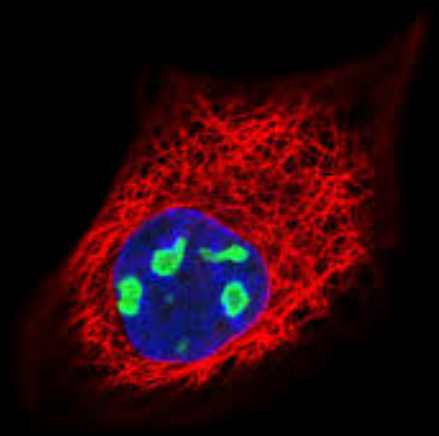
NUCLEAR



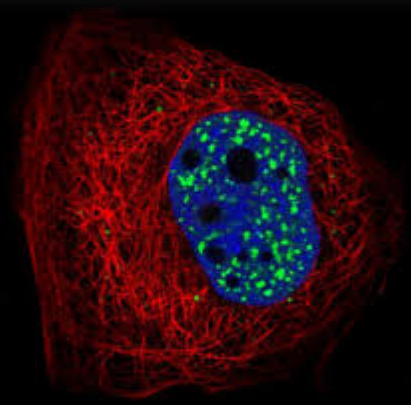
CYTOPLASM



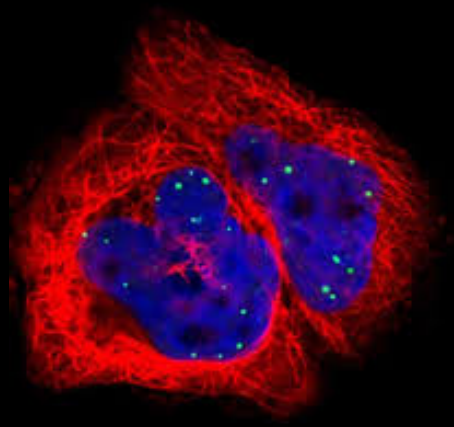
Membrane-less organelles contain RNAs



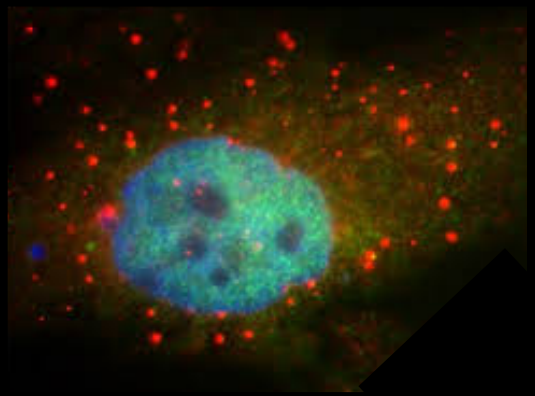
Nucleoli



Nuclear speckles

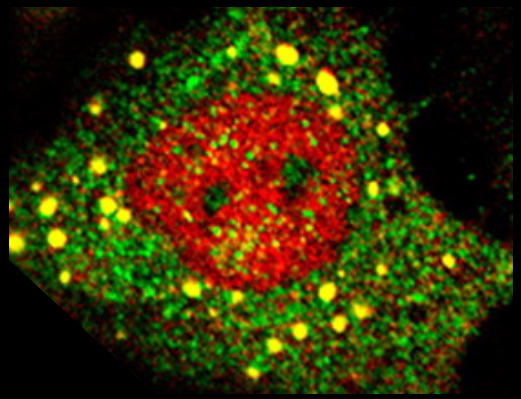


Cajal bodies



P-bodies

These organelles can undergo fission and fusion, and hence their formation has been described as mediated by **liquid-liquid phase separation**

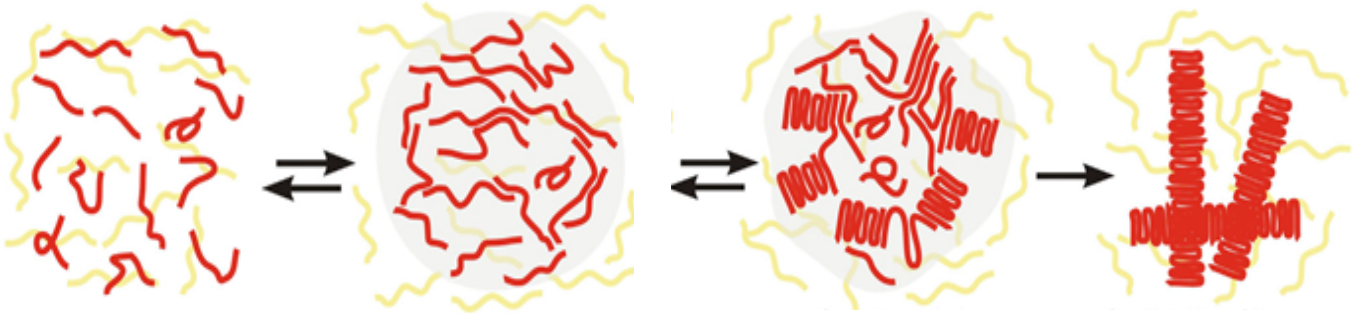


Stress granules

Phase-separated multi-molecular assemblies provide a general regulatory mechanism to compartmentalize biochemical reactions within cells

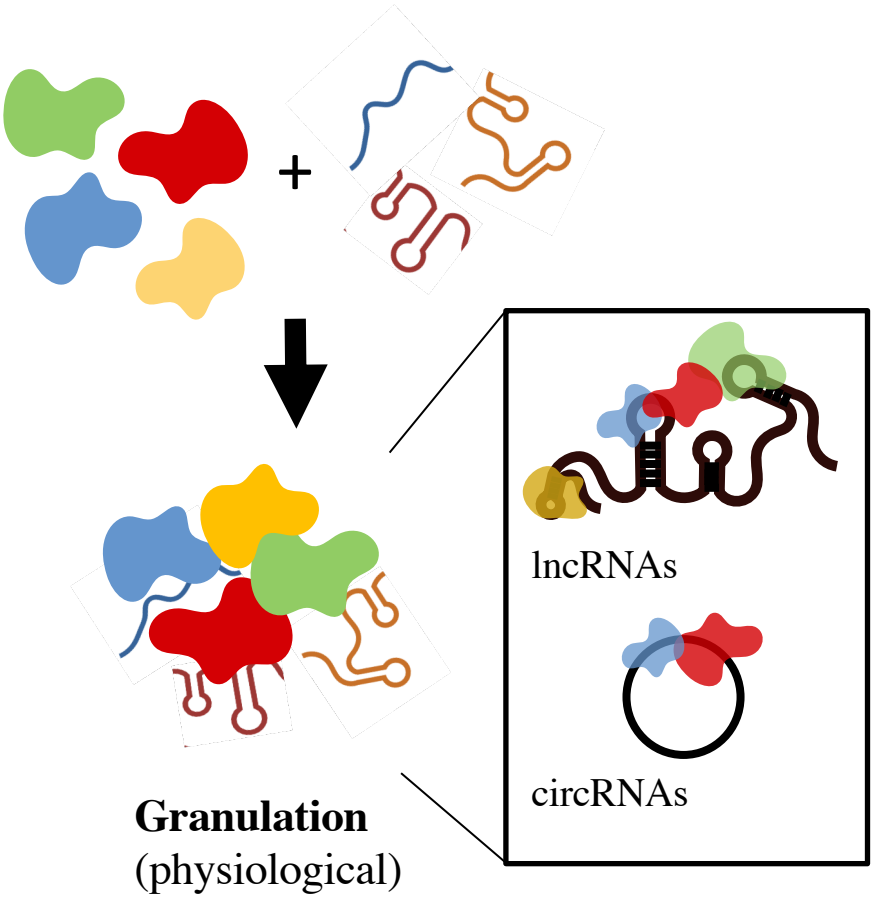
Cellular reactions occur in membrane-less compartments in a ***liquid-like structure***

Increase of compartment viscosity can lead to ***solid-like aggregates***
and alteration of many processes



Alberti & Hyman 2016

RNA contributes to maintain membrane-less compartments in a ***liquid-like structure***



Granulation
(physiological)

Molecular Cell
Volume 60, Issue 2, 15 October 2015, Pages 220-230

RNA controls PolyQ protein phase transitions

Huaiying Zhang[1,2], Shana Elbaum-Garfinkle[2], Erin Langdon[1], Nicole Taylor[2], Patricia Occhipinti[1], Andrew Bridges[1], Clifford P. Brangwynne[2,*], and Amy S. Gladfelter[1,*]

PNAS

RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome

Briana Van Treeck[a], David S. W. Protter[a], Tyler Matheny[a], Anthony Khong[a,b], Christopher D. Link[c], and Roy Parker[a,b,1]

Science
REPORTS

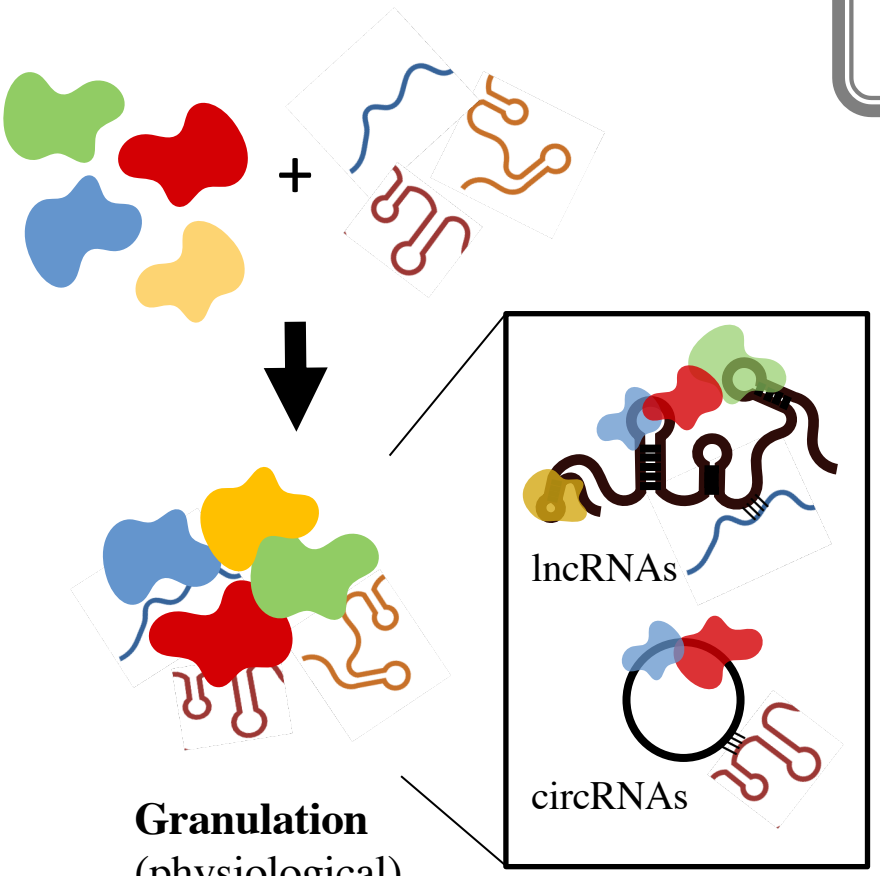
Cite as: S. Maharana *et al.*, *Science* 10.1126/science.aar7366 (2018).

RNA buffers the phase separation behavior of prion-like RNA binding proteins

Shovamayee Maharana,[1] Jie Wang,[1*] Dimitrios K. Papadopoulos,[1,2*] Doris Richter,[1] Andrey Pozniakovsky,[1] Ina Poser,[1] Marc Bickle,[1] Sandra Rizk,[1,3] Jordina Guillén-Boixet,[1] Titus Franzmann,[1] Marcus Jahnel,[1] Lara Marrone,[4] Young-Tae Chang,[5,6] Jared Sterneckert,[4] Pavel Tomancak,[1] Anthony A. Hyman,[1†] Simon Alberti[1†]

RNA contributes to maintain membrane-less compartments in a ***liquid-like structure***

Do ncRNAs control the structure and function of membrane-less compartments?



Granulation
(physiological)

Leading Edge
Review

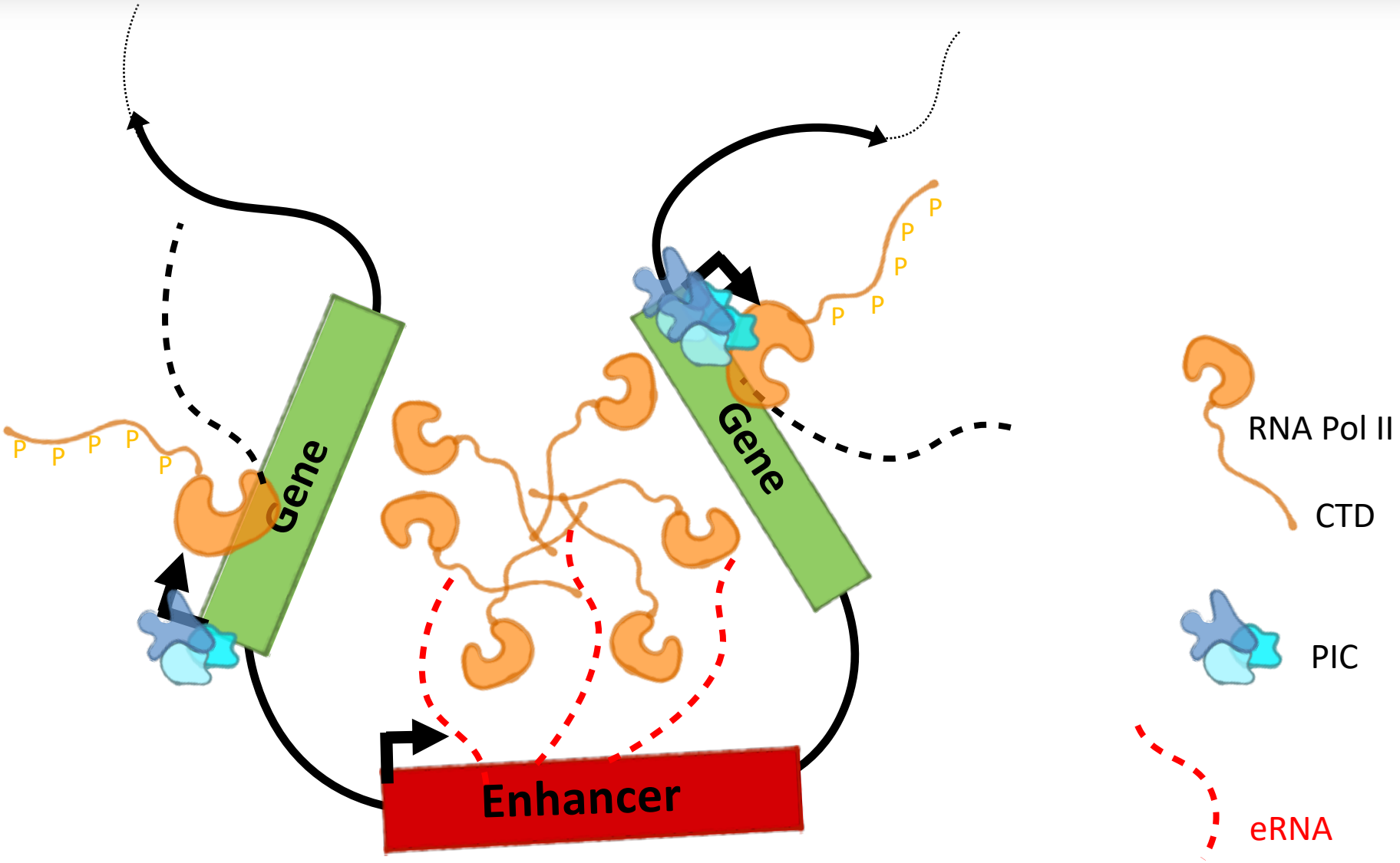
Cell

Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies

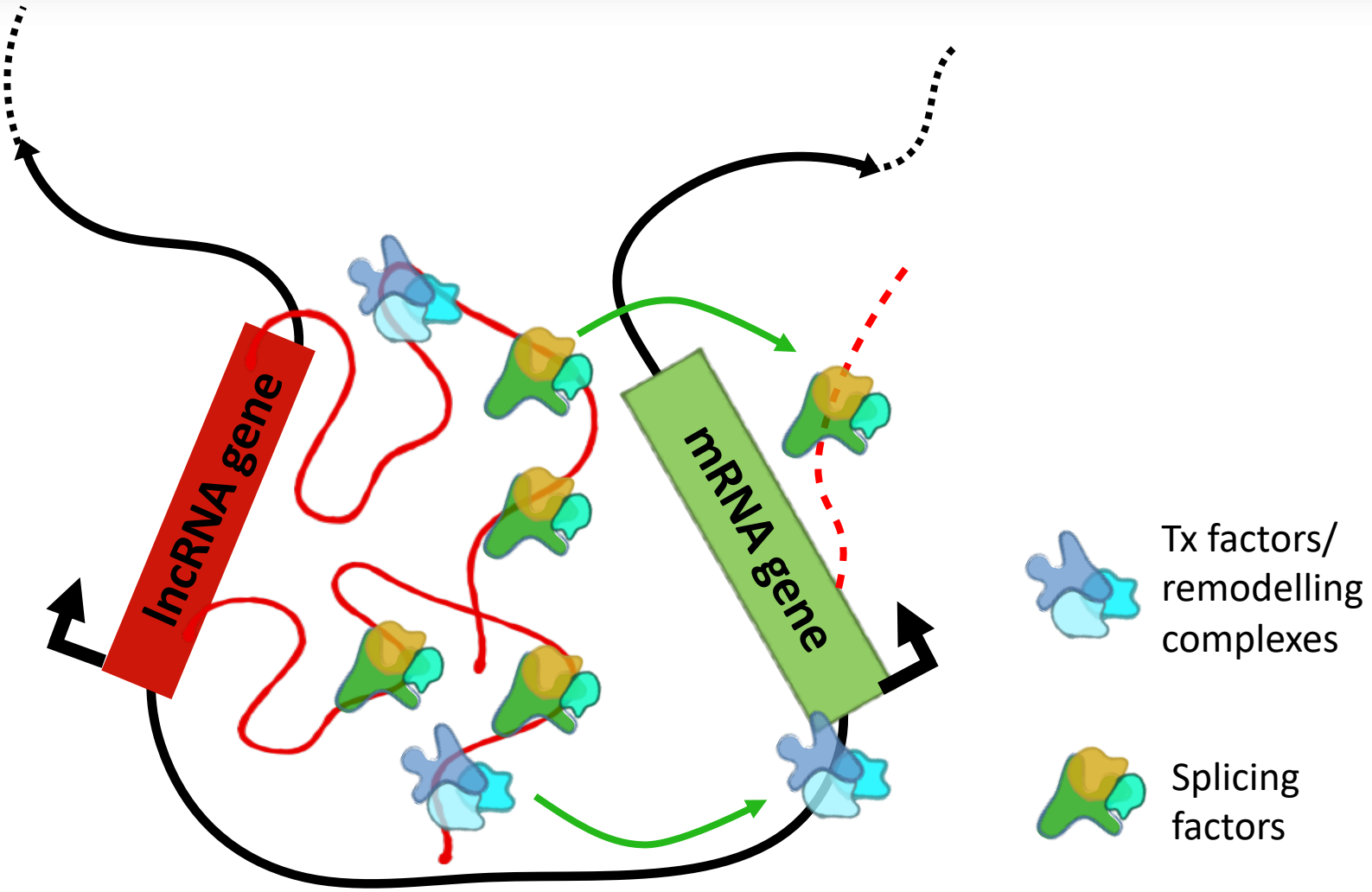
Briana Van Treeck[1] and Roy Parker[1,2,*]
[1]Department of Biochemistry, University of Colorado, Boulder, CO 80309, USA

lncRNA function in the nucleus

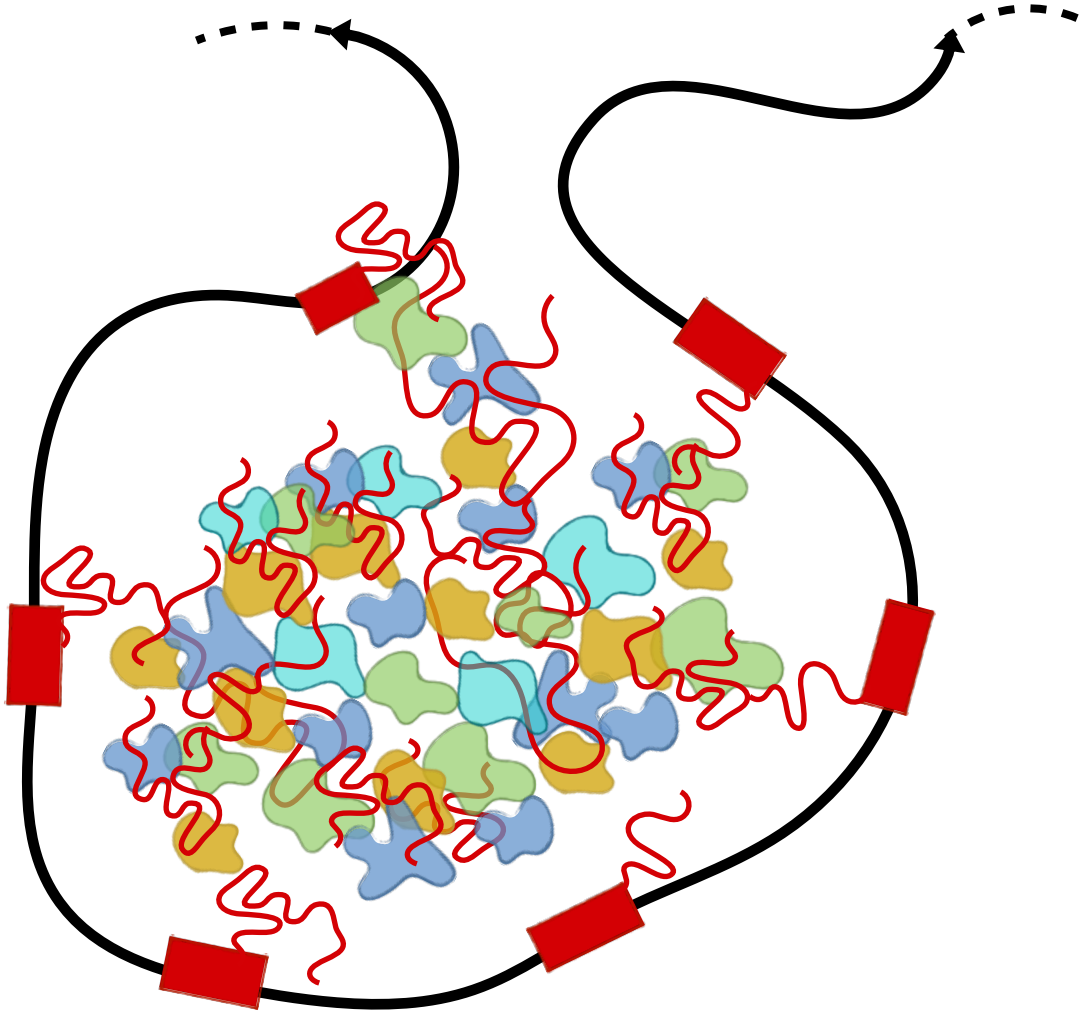
The CTD of pol II contains low complexity domains LCD and forms liquid-phase assemblies at the sites of transcriptional initiation



lncRNA divergent transcription tethers transcriptional factors, chromatin remodelling complexes and splicing factors to the neighbouring gene

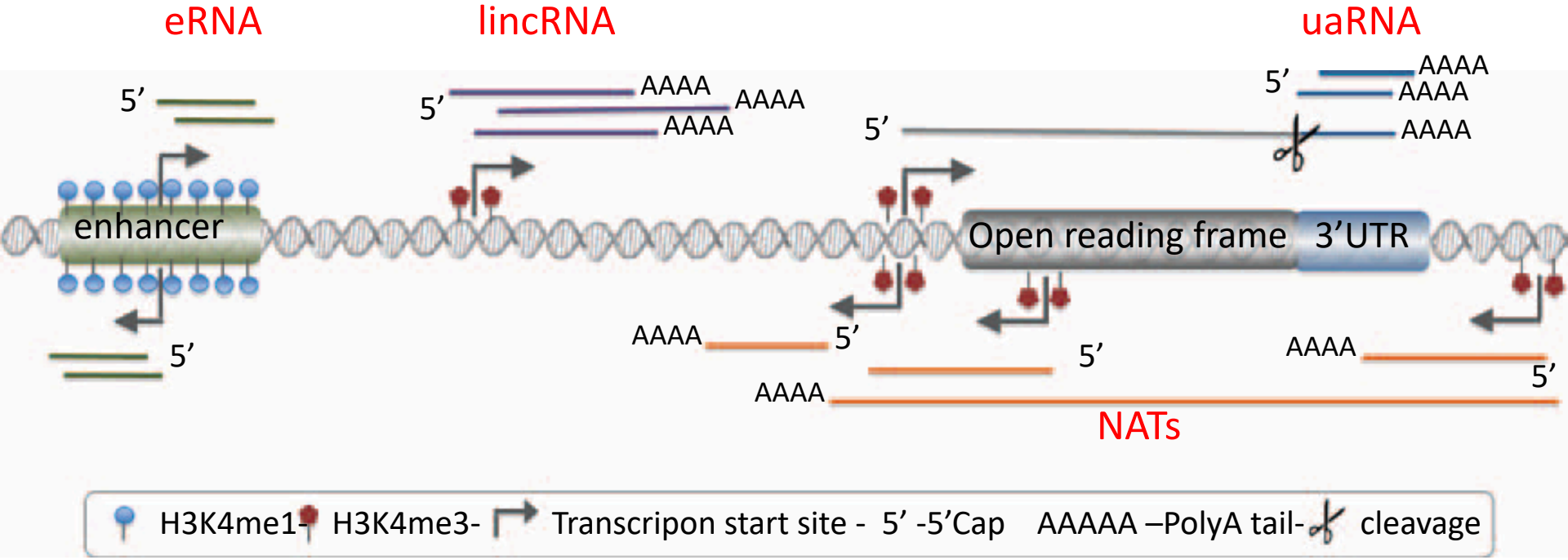


Super enhancer condensates: control genes that have especially prominent roles in cell-type-specific processes

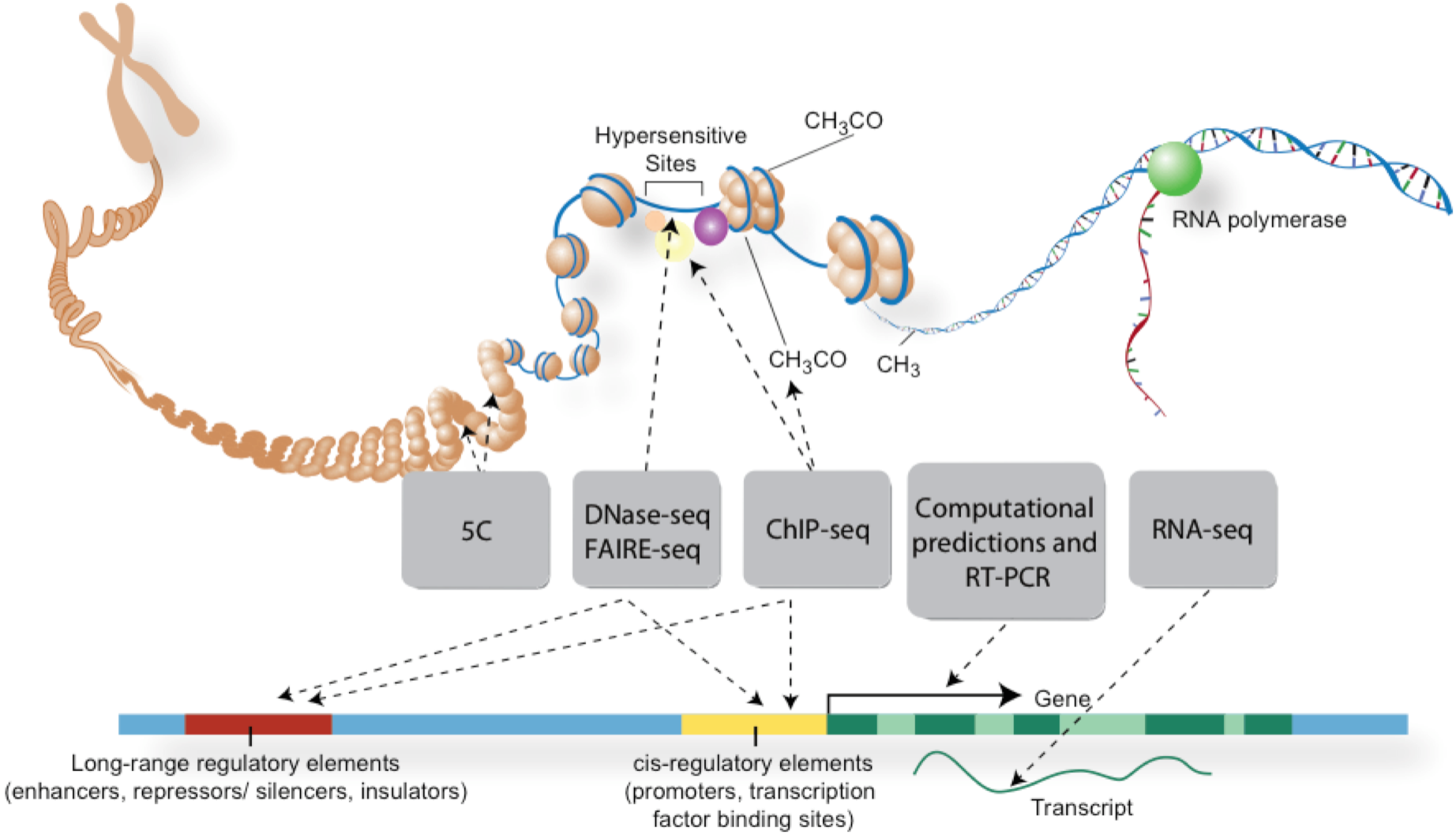


super-enhancers can be considered to be co-operative assemblies of high densities of transcription factors, transcriptional co-factors, chromatin regulators, non-coding RNAs and RNA Polymerase II.

The current landscape of non coding RNAs



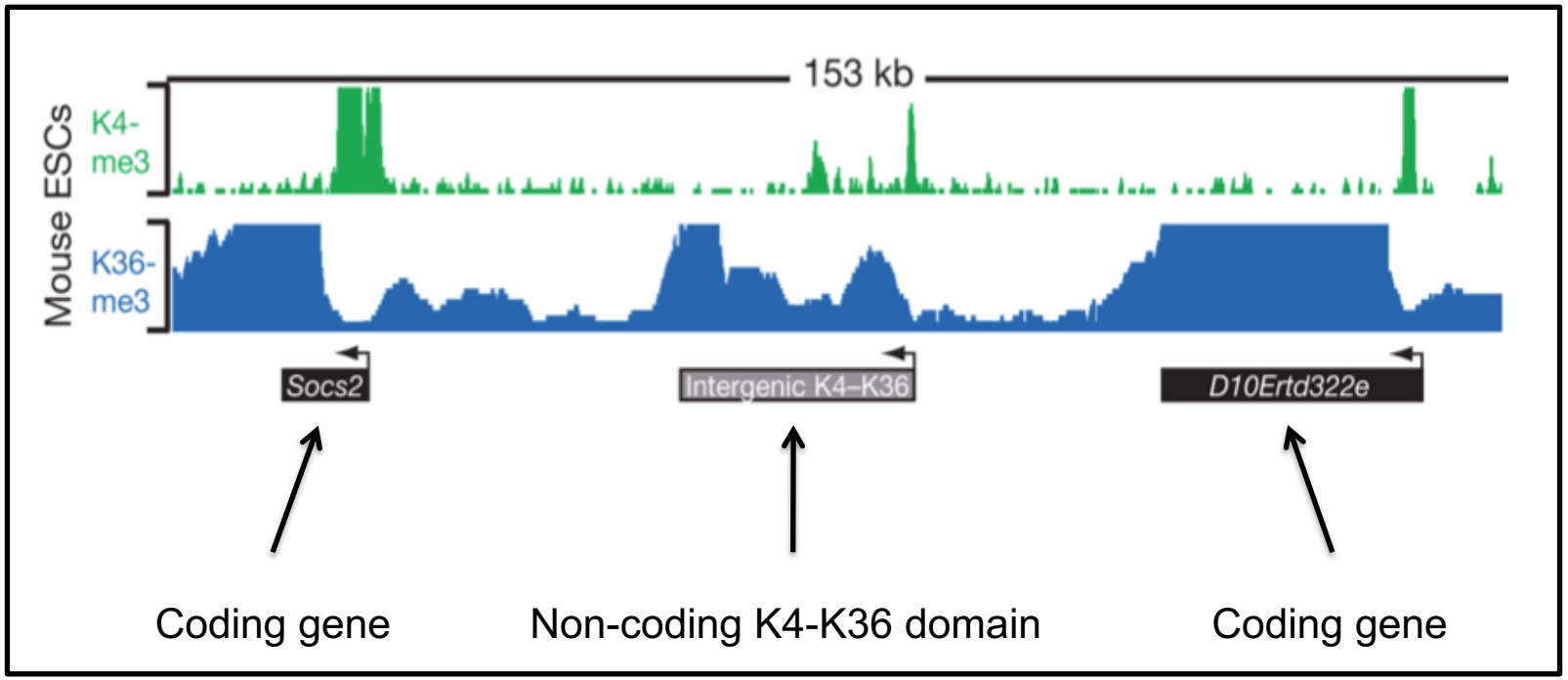
How to identify a lncRNA coding region



lncRNAs are functional transcripts

K4-K36 Domains

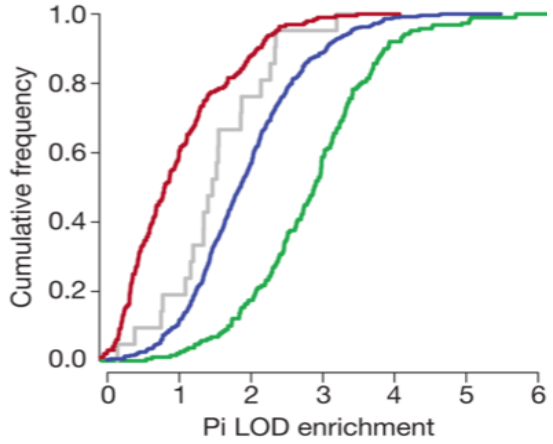
Actively Transcribed Pol II genes are marked by H3 lysine 4 trimethylation (promoter) and H3 lysine 36 trimethylation (whole transcribed region)



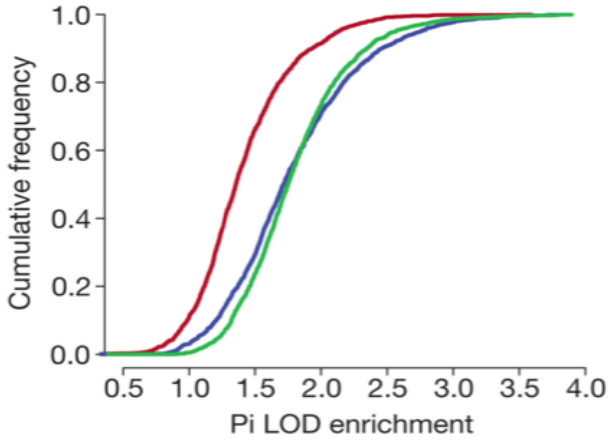
Some intergenic regions and active genes share a common chromatin signature

lncRNAs are functional transcripts

Sequence conservation

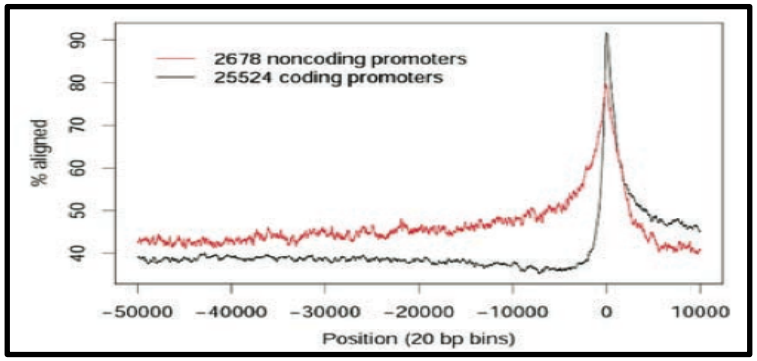


Coding sequences are strongly conserved among mammals (green) whereas lncRNAs show lower scores (blue), but still more than introns and UTRs.



In which regions should a functional transcript undergo a selective pressure?
mRNAs ➔ coding sequence
lncRNAs ➔ structure/functional seq.
both of them ➔ regulatory elements

In their promoters, lncRNAs show conservation scores even higher than coding genes!



Functions of lncRNAs

NUCLEUS

- Chromatin Remodeling
- Transcription Regulation
- Nuclear Architecture

CYTOPLASM

- Post-transcriptional Regulation
- mRNA decay/translation
- miRNA decoy
- organizers of membrane-less compartments

etc…

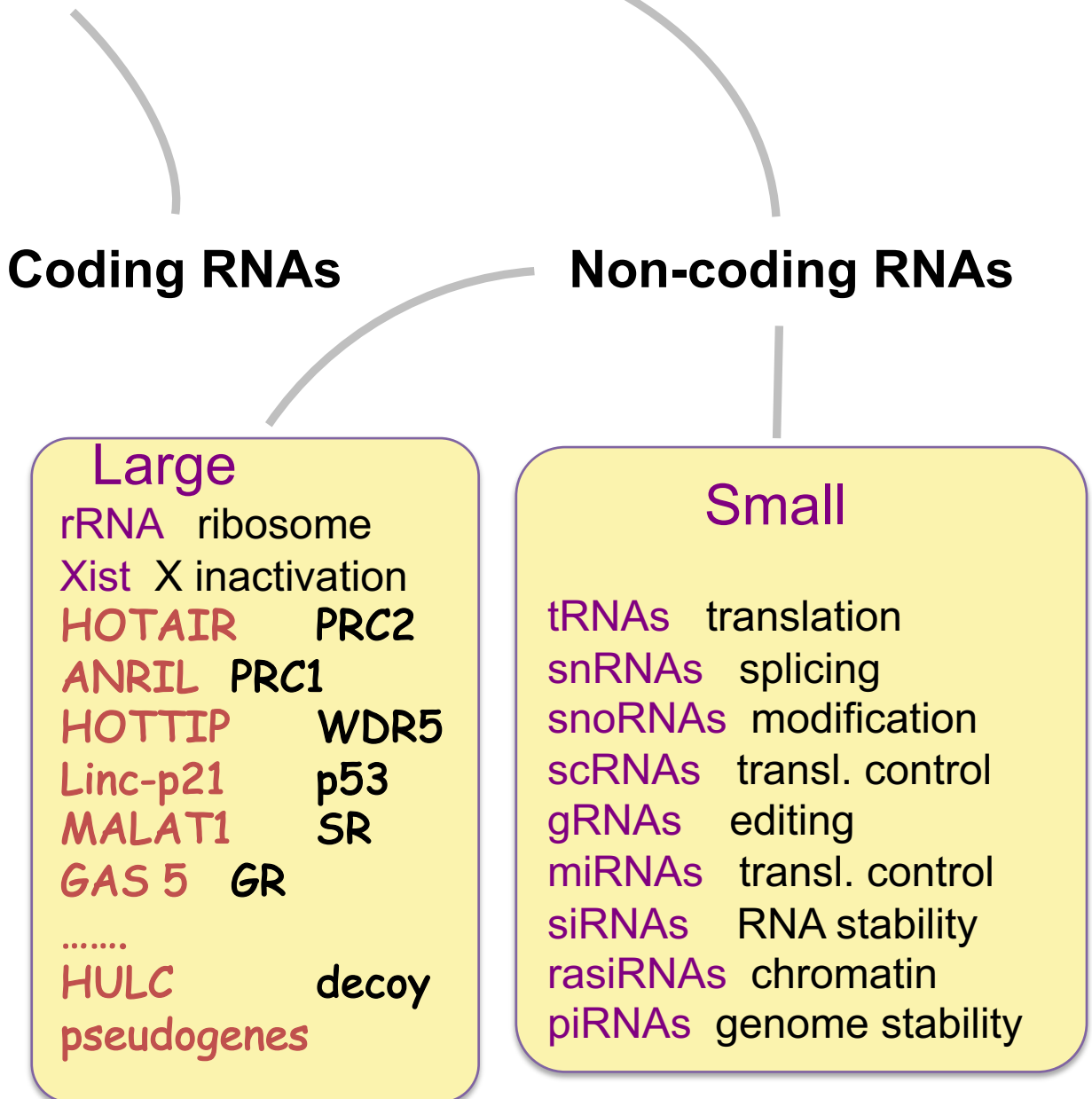
Functions of lncRNAs

All these functions are accomplished in a sequence-dependent manner… so why there is such a little sequence conservation among lncRNAs?

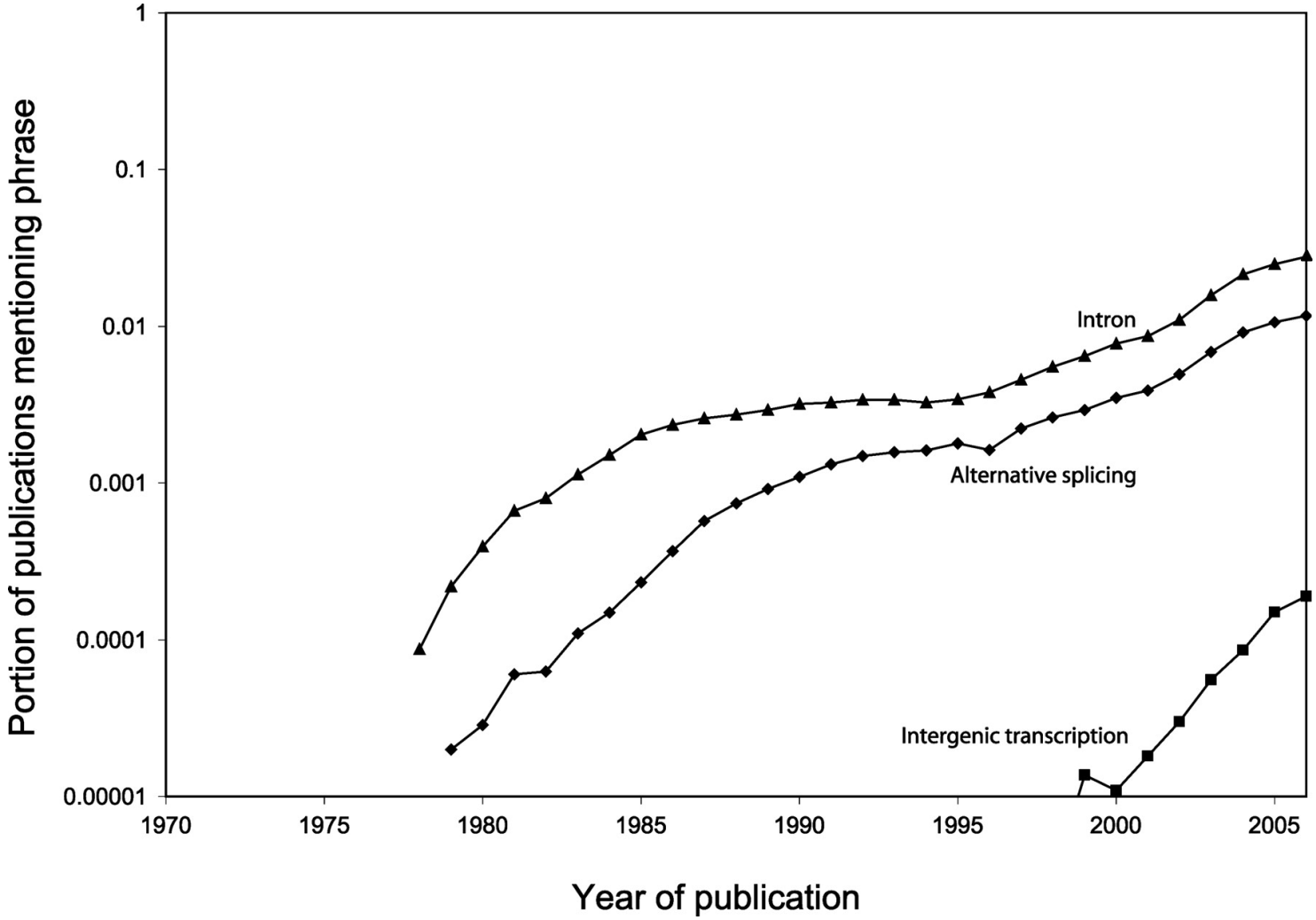
- Sequence conservation average is lower than in coding gene but there are peaks in specific functional regions
- Sequence-dependent functions with respect to structure are easier to discover (sequence analysis) and to study (mutation analysis)
- Many lncRNAs can have functions depending on their 3D structure, which is difficult to predict

So the well-characterized non-coding RNAs are not a reliable sample of the whole non-coding transcriptome!!!

Eukaryotic RNAs



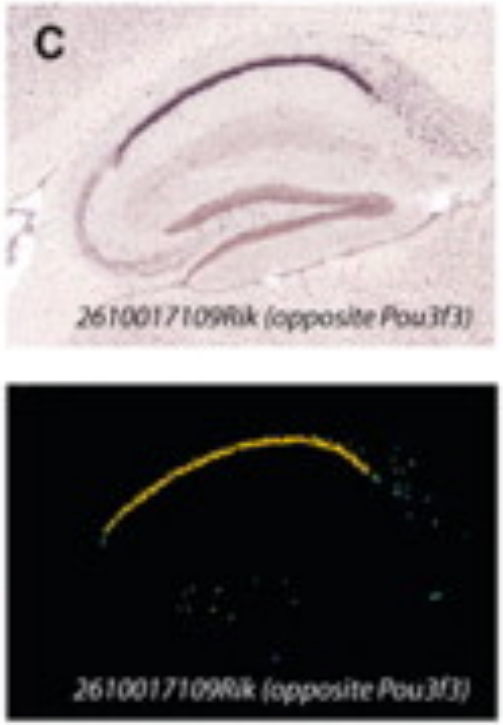
Keyword analysis and complexity of genes.



Gerstein M B et al. Genome Res. 2007;17:669-681

Why has this system gone unnoticed?

- **Intellectually unprepared**
- **Genetically subtle**
- **Biochemically invisible**

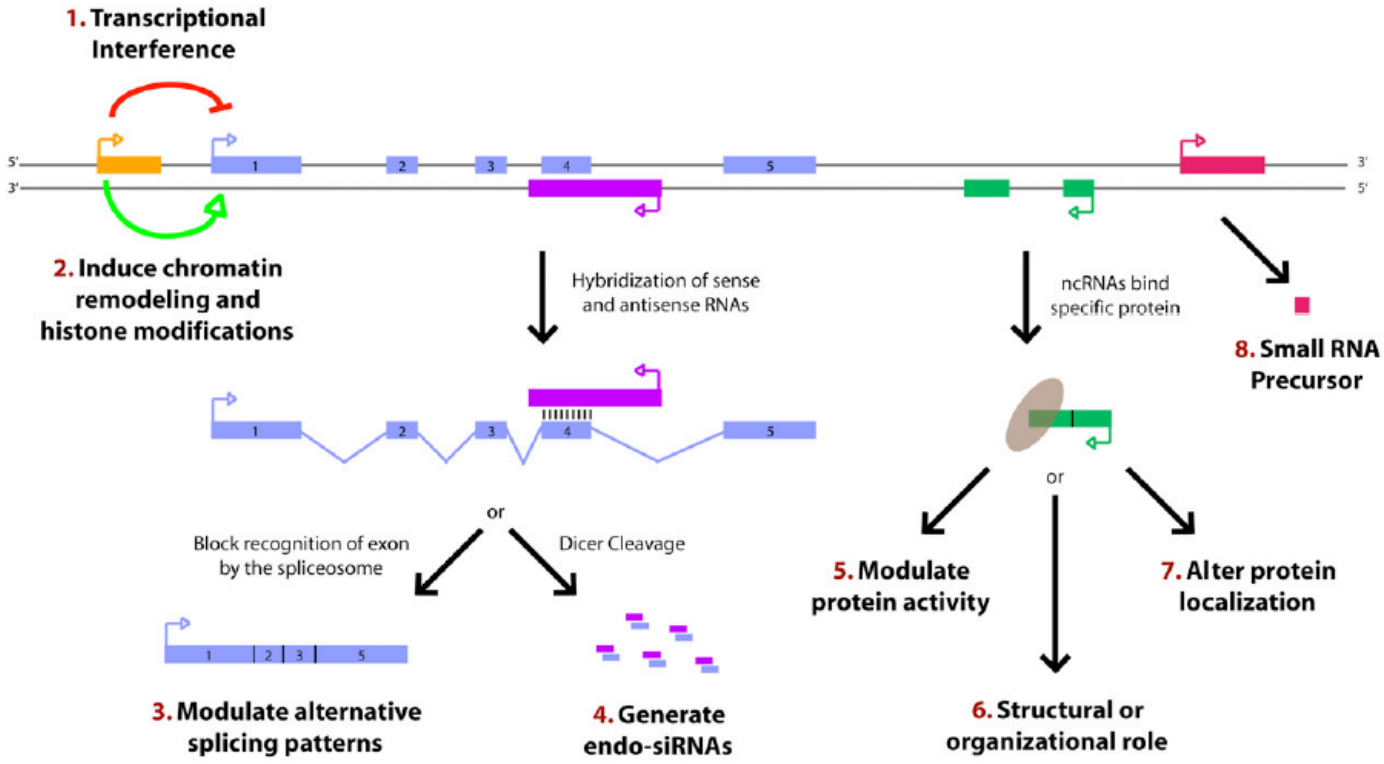


A survey of the expression of over 1300 lncRNAs in mouse brain showed that over 600 were expressed in highly specific locations, such as different regions of the hippocampus, different layers of the cortex, or different parts of the cerebellum

Long non-coding RNAs

2001 to the Future Break the Dogma

List of long non-coding RNAs



J E Wilusz et al Genes & Development 2009

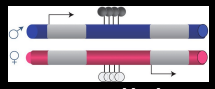
Why to study lncRNAs

AGING-CANCER

Telomere function

hTR lncRNA

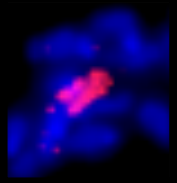
DEVELOPMENTAL DEFECTS



mono-allelic Gene expression

HOTAIR lncRNA

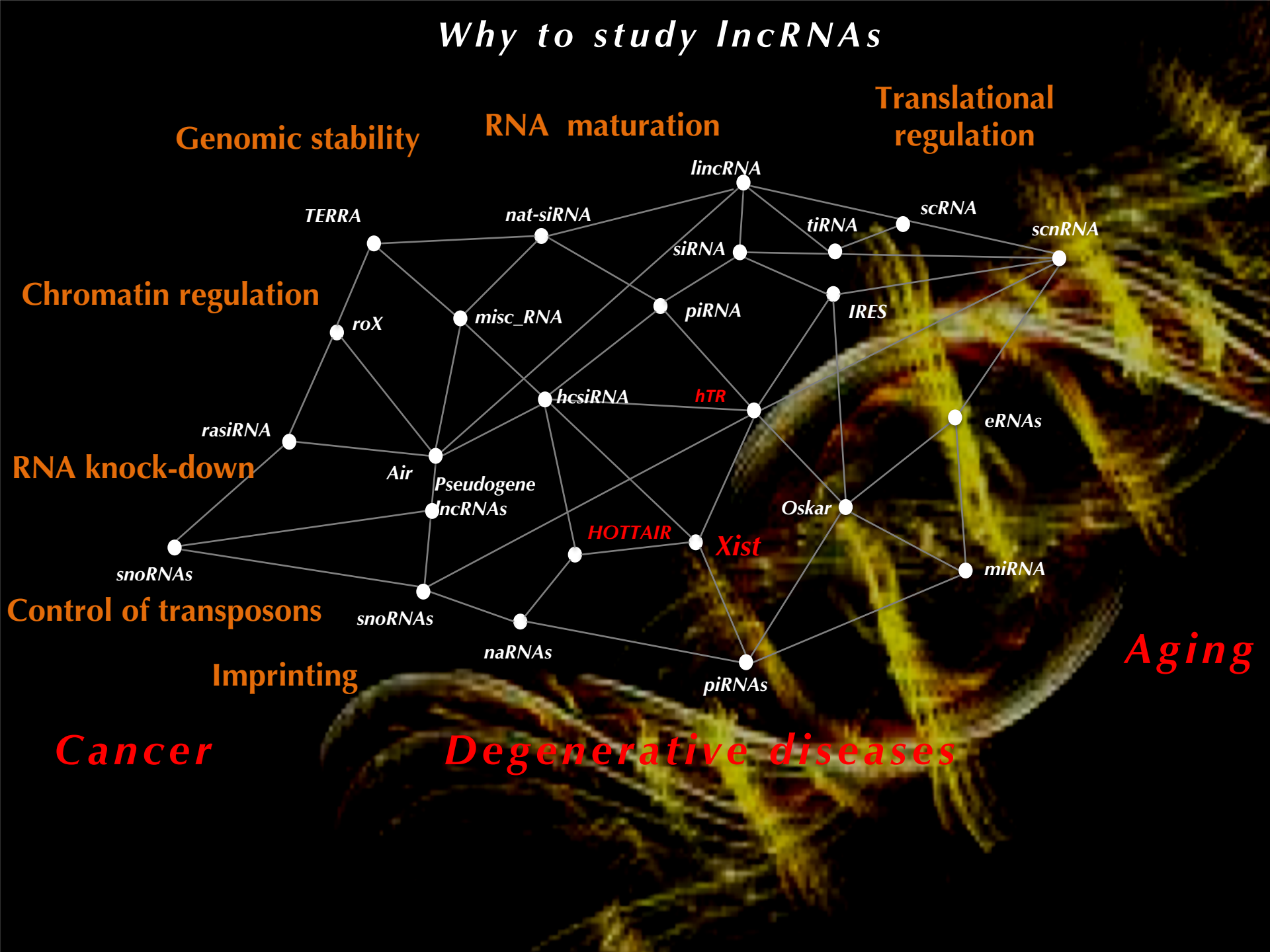
Xist lncRNA



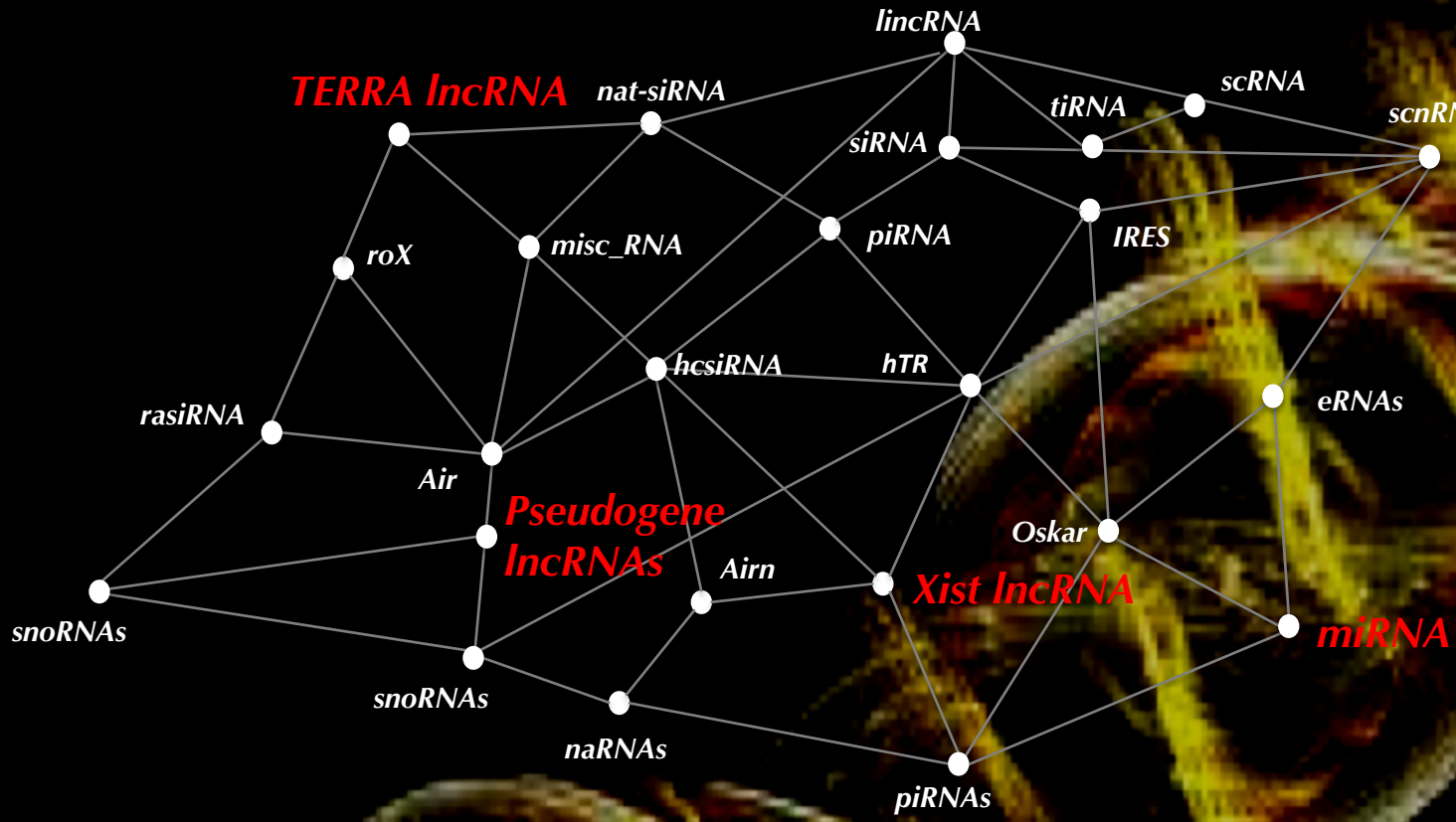
X inactivation Silencing 1000 genes on the Xi

LETHAL

Why to study lncRNAs



Why to study lncRNAs



Long non-coding RNAs

The Transcriptional Landscape of the Mammalian Genome

The FANTOM Consortium* and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group)*

	Total	Average per TU cluster	Average per TK cluster
Total number of transcripts	158,807	7.59	7.30
RIKEN full-length	102,801		
Public (non-RIKEN) mRNAs	56,006		
GFs	25,027	1.20	1.15
Framework clusters	31,992	1.53	1.47
TUs	44,147	2.11	2.03
With proteins	20,929	1.00	0.96
Without proteins	23,218	1.11	1.07
TK	45,142	2.16	2.07
With proteins	21,757	1.04	1.00
Without proteins	23,385	1.12	1.07
Splicing patterns	78,393	3.75	3.60

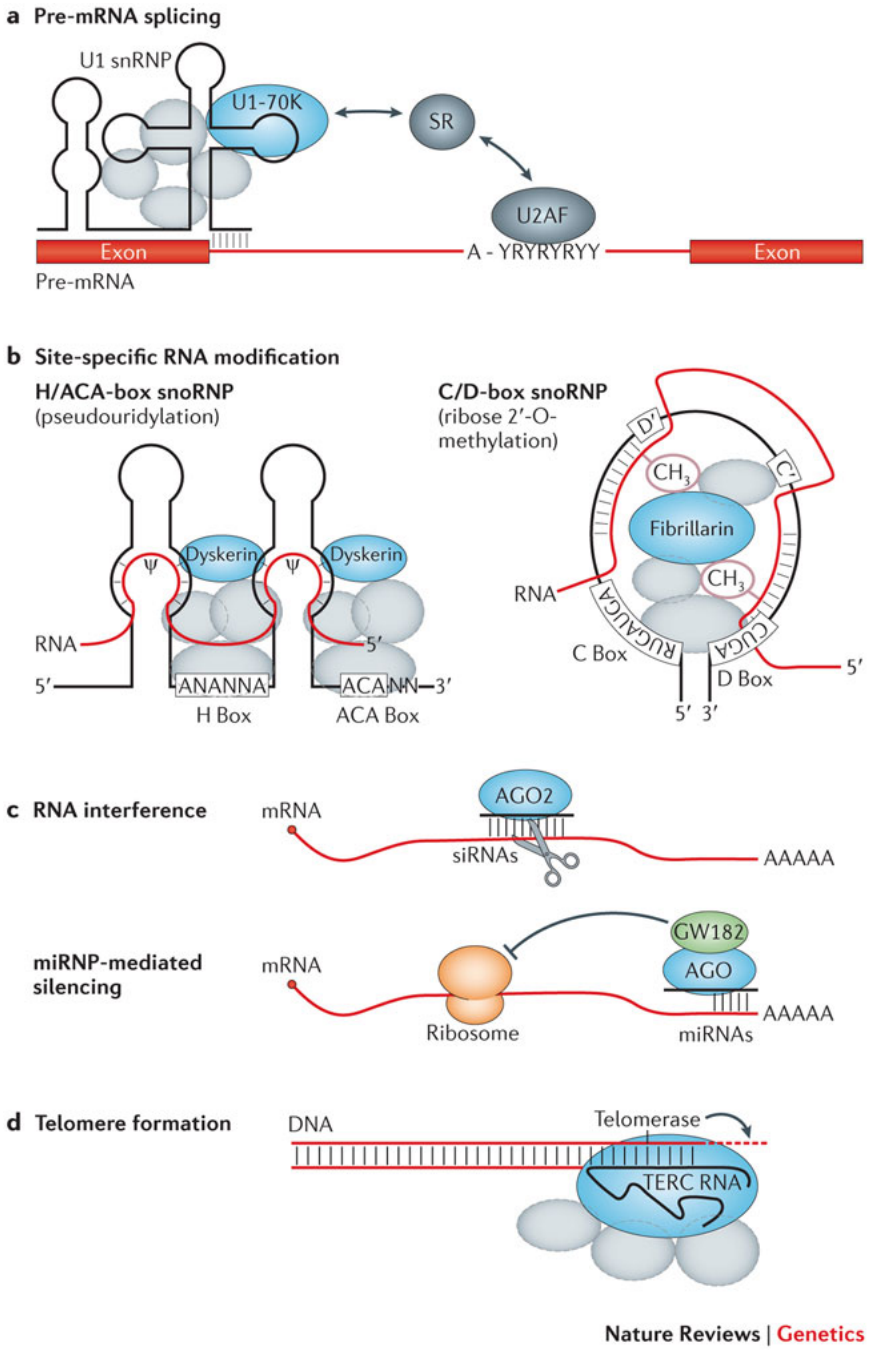
Out of the 181000 transcripts detected in the mammalian transcriptome, about a half are non-coding RNAs

Basal trancriptional noise or functional non-coding RNA?

Long non-coding RNAs: new players in cell differentiation and development

Alessandro Fatica & Irene Bozzoni

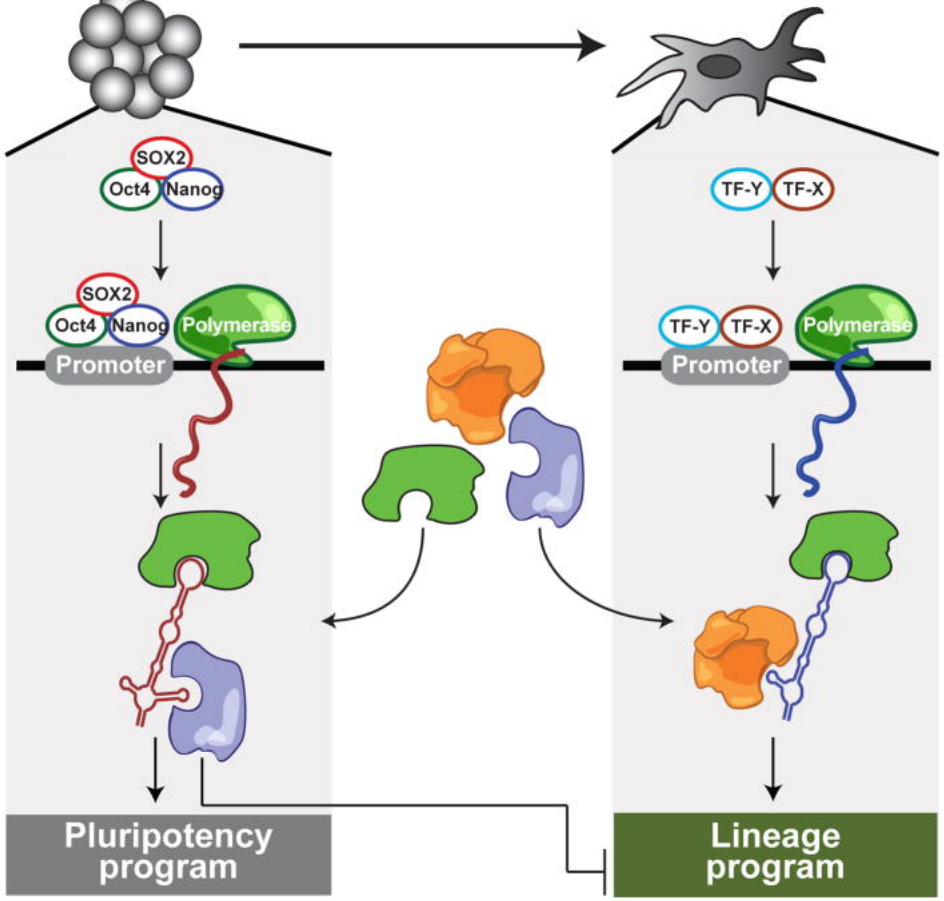
Nature Reviews Genetics (2013)



The modular scaffold model

lincRNAs act in the circuitry controlling pluripotency and differentiation

Mitchell Guttman……..& Eric S. Lander Nature 477, 295–300

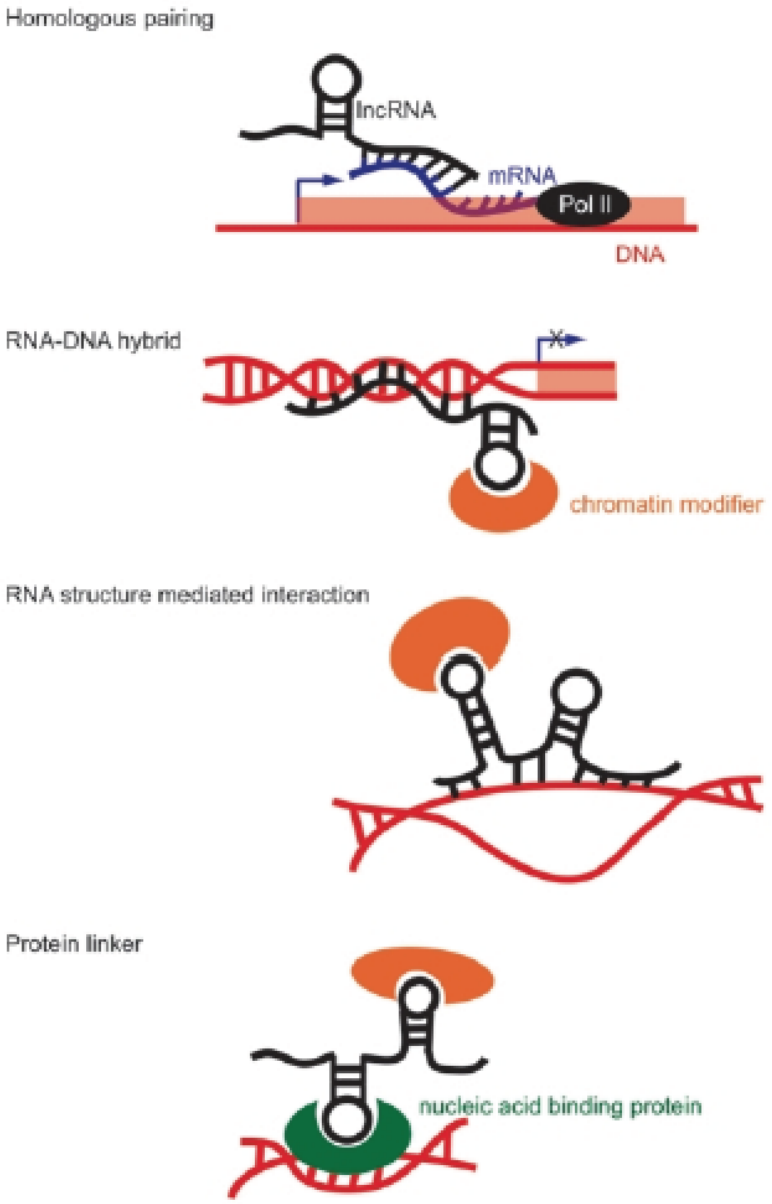


A model for lincRNA integration into the molecular circuitry of the cell

Left: ESC-specific transcription factors (such as Oct4, Sox2, and Nanog) bind to the promoter of a lincRNA gene and drive its transcription. The lincRNA binds to ubiquitous regulatory proteins, giving rise to cell-type specific RNA–protein complexes. Through different combinations of protein interactions, **the lincRNA–protein complex can give rise to unique transcriptional programs**. Right: A similar process may also work in other cell types with specific transcription factors regulating lincRNAs, creating cell-type–specific RNA–protein complexes and regulating cell-type–specific expression programs.

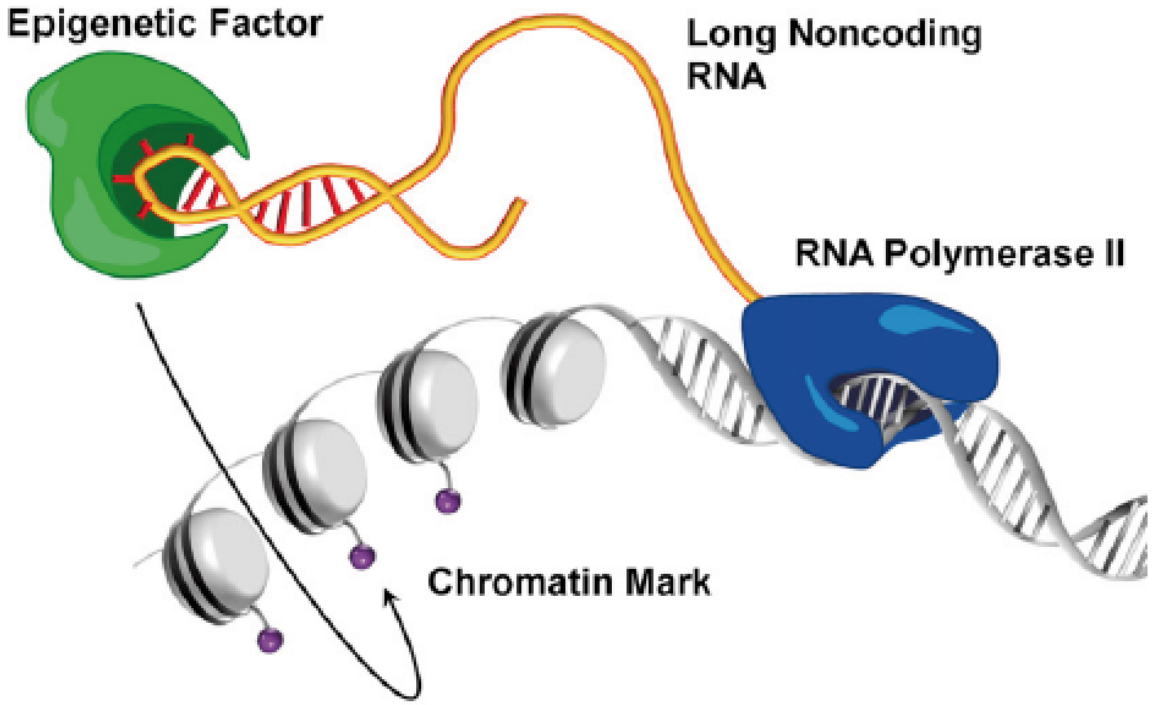
How does a lncRNA interface with selective regions of the genome?

- RNA:RNA hybrid of lncRNA with a nascent transcript
- formation of a RNA: DNA:DNA triplex
- RNA: DNA hybrid that displaces a single-strand of DNA (so called R-loop)
- RNA binding to a sequence-specific DNA binding protein



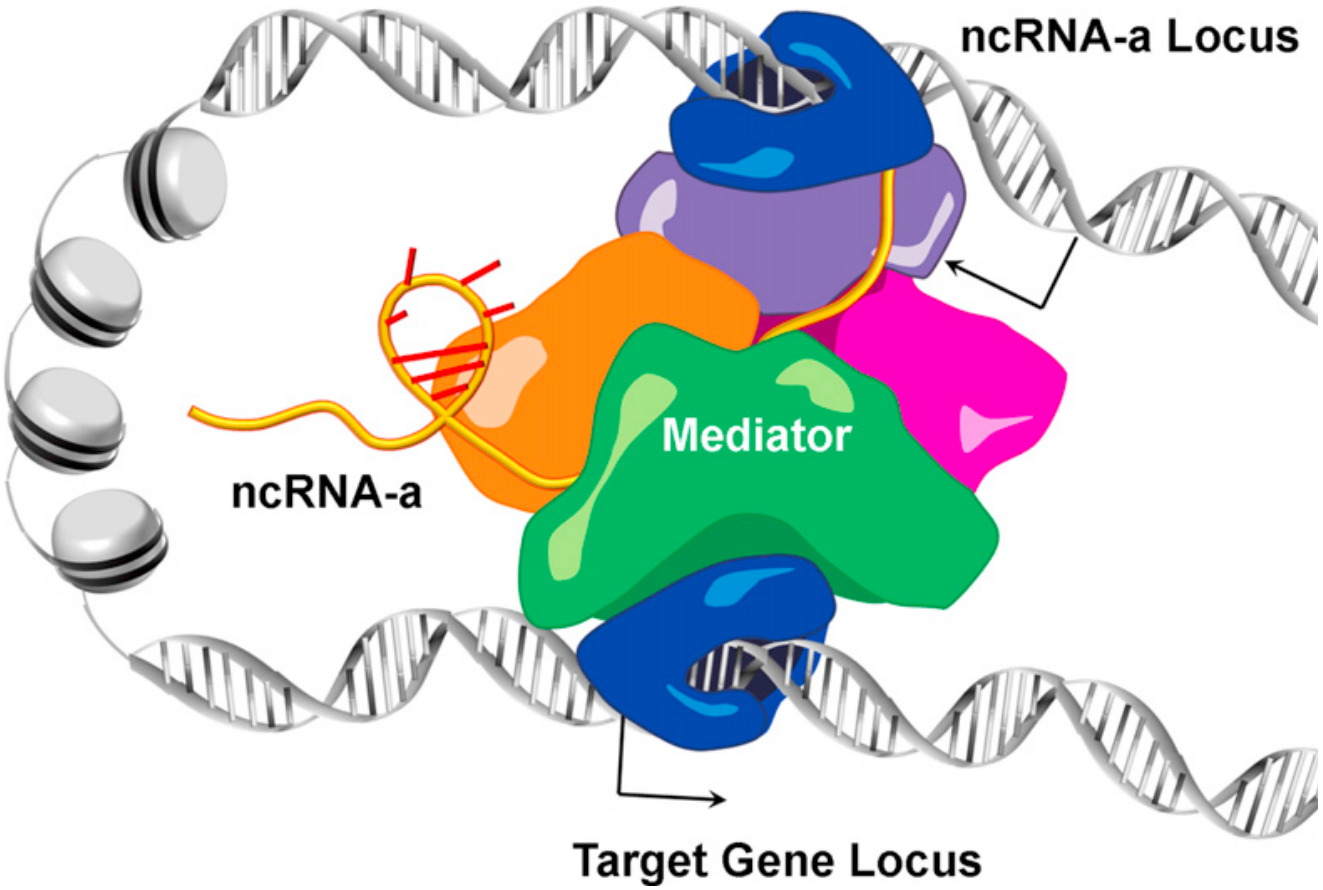
Cis lncRNAs are uniquely suited to act as allele- and locus-specific recruiters by virtue of their:

- *length* (allowing them to reach out and capture protein factors while tethered to chromatin)
- *specificity* (since most lncRNAs emanate from single loci)
- possibility to *hybridize* to chromatin through DNA:RNA heteroduplexes during transcription



Enhancer noncoding RNAs (eRNAs)

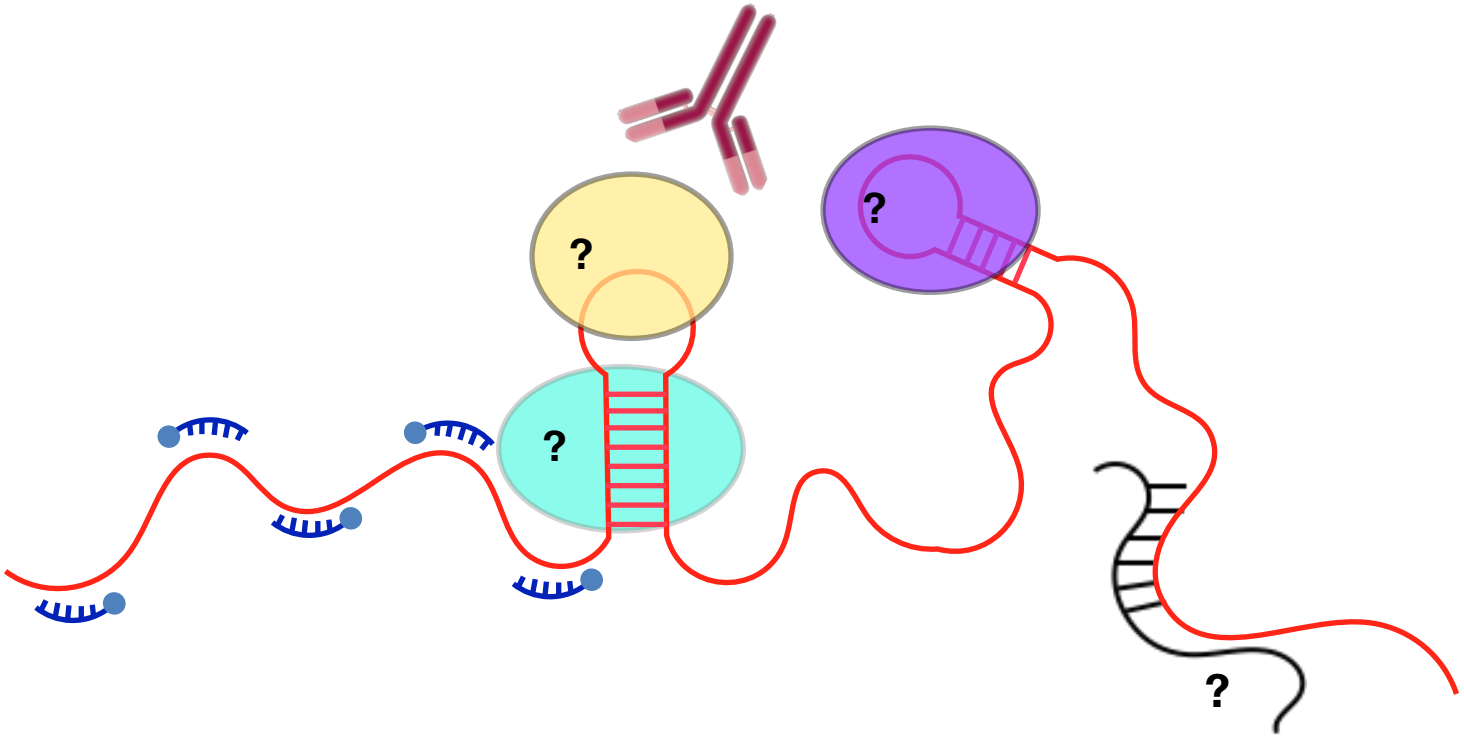
Example of *cis-acting* lncRNA: non-coding transcription from neuronal enhancers (Bond et al., 2009; Onodera et al., 2012) produces a class of activating lncRNAs called ‘‘ncRNA-a’’ (Ørom et al., 2010).



Mediator thus acts as a bridge between transcription factors binding at distant enhancers and the RNA polymerase II (RNAPII) apparatus at target promoters.

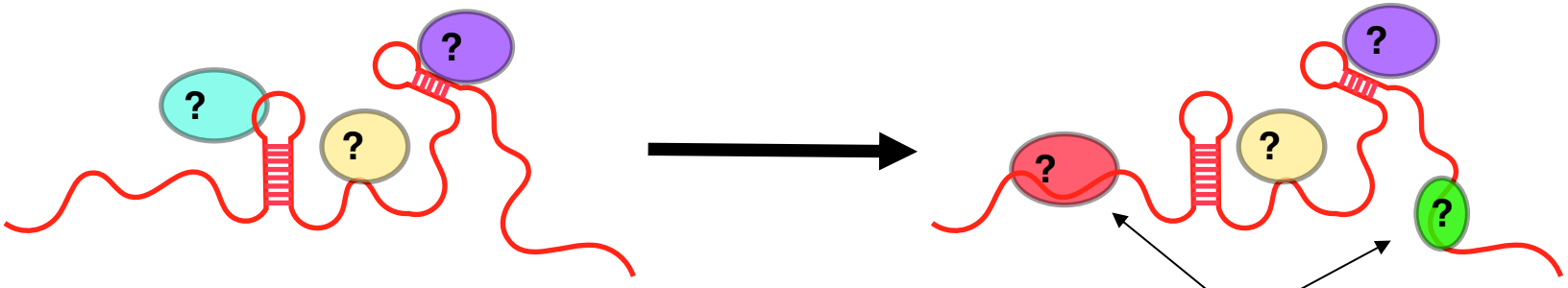
RNA pull down versus IP

Pull-down of a known RNA with biotinylated oligos or IP with Ab for a known protein



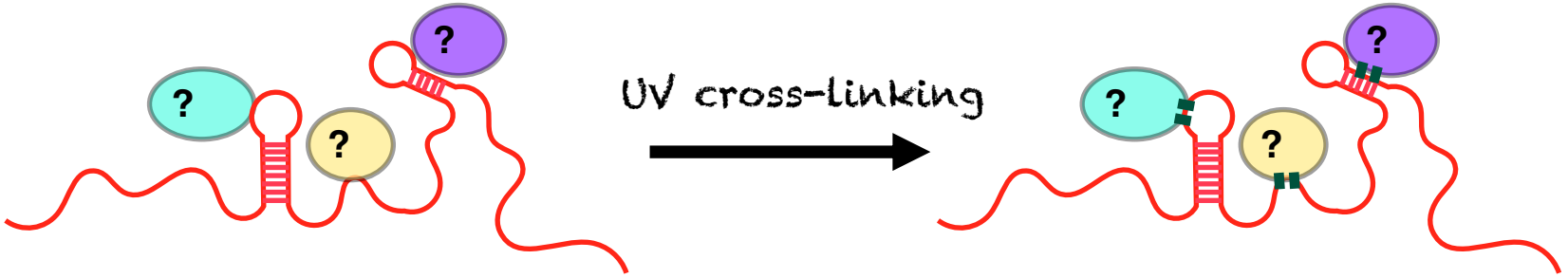
RNA-protein interaction: native or cross-linking?

native conditions can led to non-specific RNA-protein interactions in solution



Tris–HCl 50 mM, **NaCl 150 mM**, $MgCl_2$ 1 mM, NP-40 0,05%

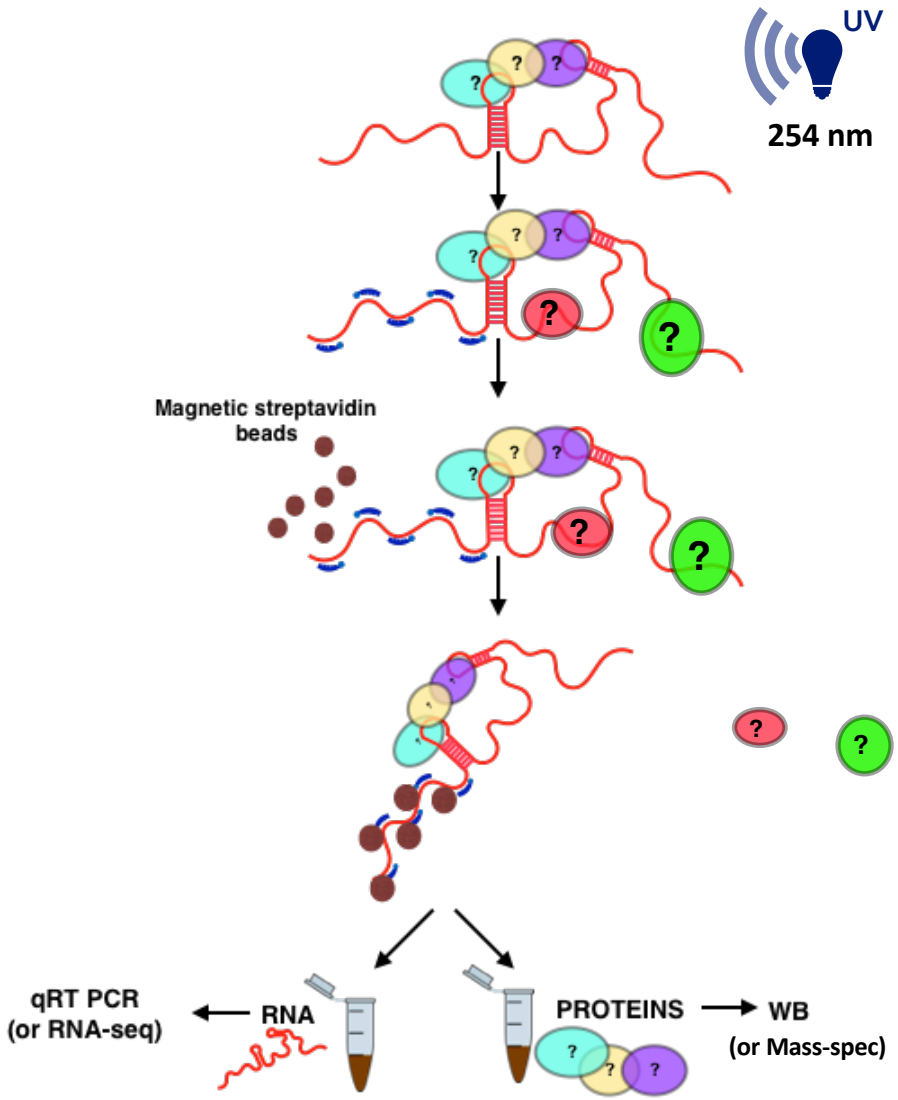
False Positives



Tris–HCl 50 Mm, **NaCl 1 M** (or LiCl), EDTA 1 Mm, NP-40 1 %, SDS 0,1% Na deoxyc. 0.5 %

CLIP (UV-RIP)
PAR-CLIP
i-CLIP (CLIP-seq)

CROSS-LINKED RNA PULLDONW (RAP)



UV cross-linking
(254 nm)

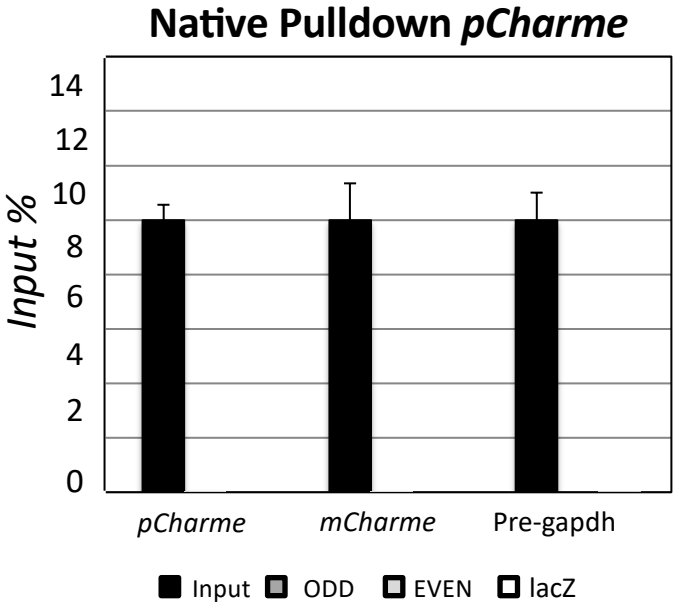
90nt long biotinylated
probes binding with extract

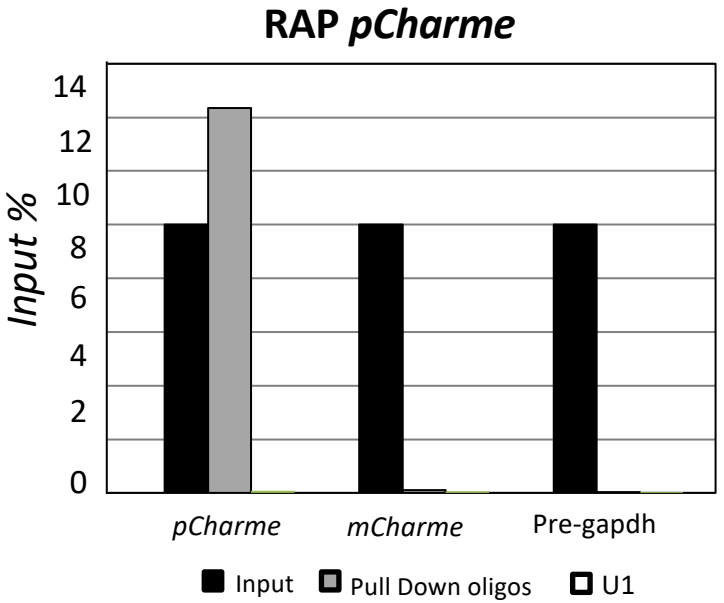
RNA/probes binding to
streptavidin magnetic beads

Capture RNA/probes
complexes from lysate
and washes in high salt buffers (1M LiCl)

RNA and
PROTEIN elution

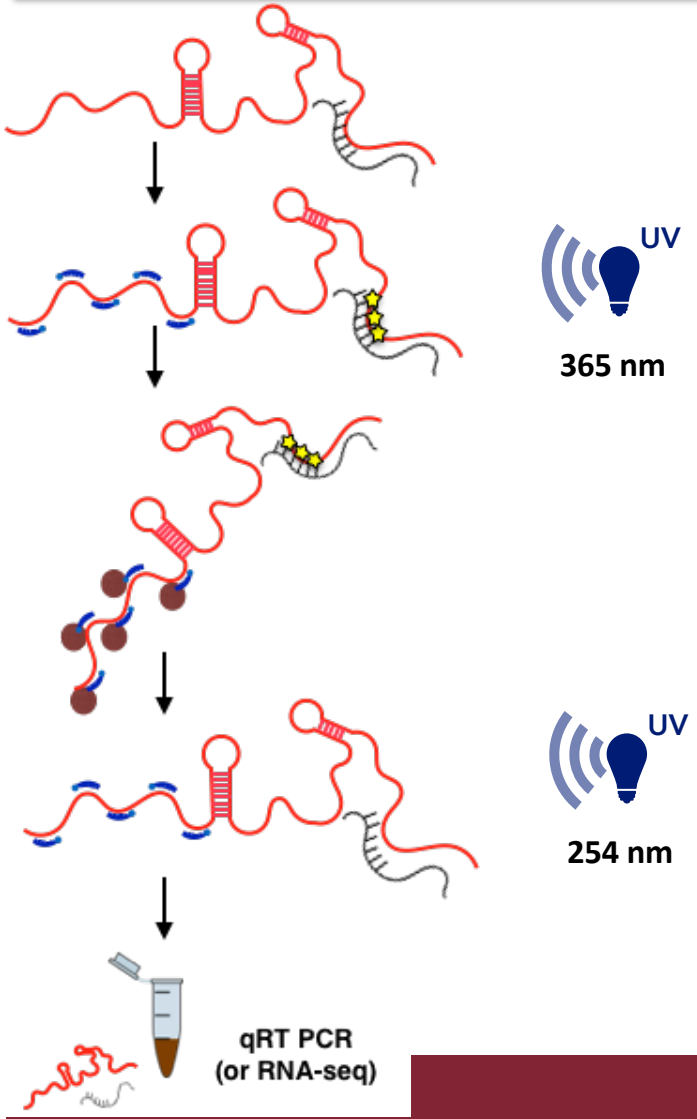
Native RNA pull-down vs RAP





- RNA Antisense Purification (RAP) allows a **higher percentage of RNA/protein complexes recovery** if compared to a native pulldown.
- Moreover thanks to the **UV crosslinking**, and to the **higher stringent washing conditions** it allows to detect only **directly-bound protein interactors** and to strongly **reduce the rate of false positives**.

RNA-RNA interaction: native or cross-linking?



PSORALEN-CROSSLINKED RNA PULLDOWN

- Psoralen (200 ug/ml)
- UV-crosslink at 365 nm for 1 hour

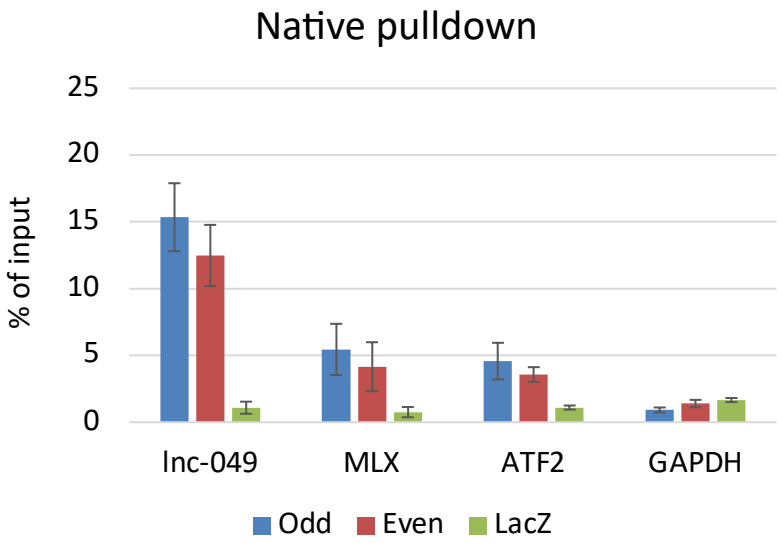
Lyse cells with guanidinium hydrocloride and extract total RNA

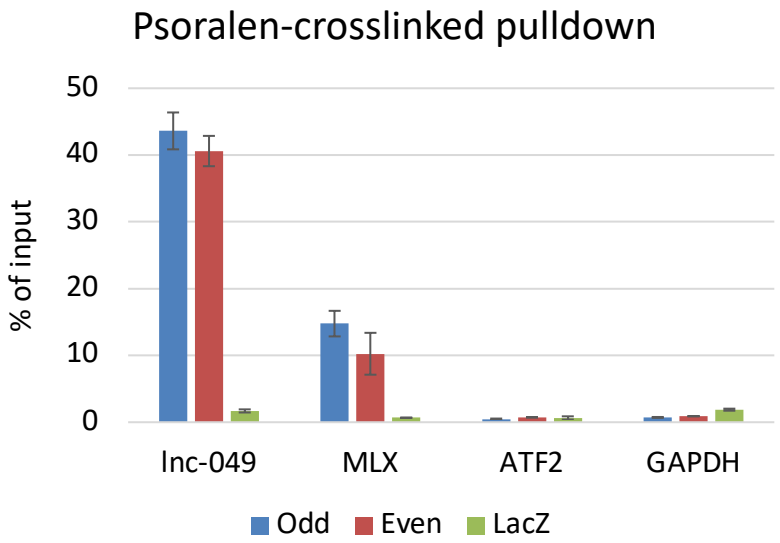
95 °C for 3 minutes - add the biotinylated probes and heat at 65 °C for 5 minutes

Recover the RNA complex by addition of streptavidin-conjugated magnetic beads (ON)

Wash the beads and elute RNA

- Reverse crosslinking by UV irradiation of the RNA sample at 254 nm for 10 minutes
- Analyze RNA





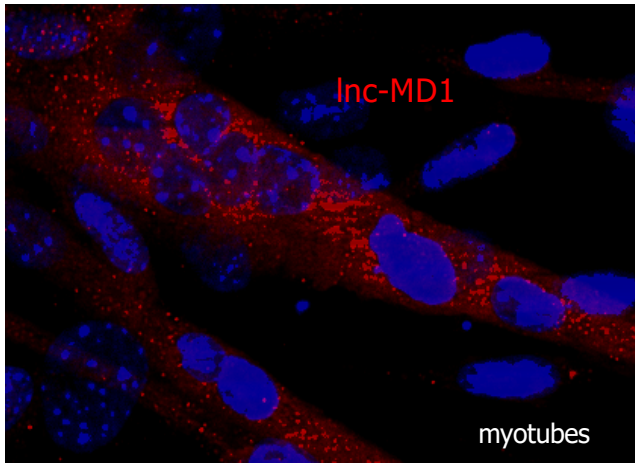
Psoralen-crossliked RNA pulldown has a higher percentage of recovery of the target RNA if compared to a native pulldown performed with the same probes

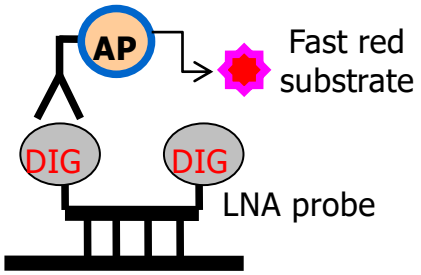
Thanks to the reversible crosslinking, it allows to detect only directly-bound interactors and to reduce the rate of false positives

In situ analysis of lncRNAs and miRNAs

Enzimatic Indirect method for citoplasmic targets

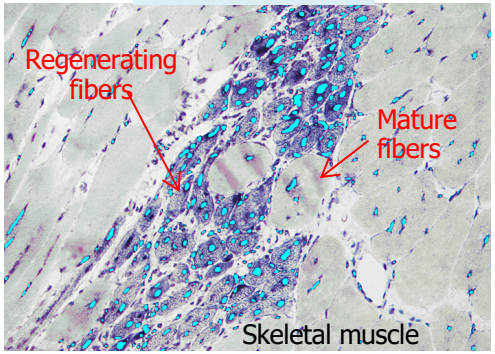
Fluorescent

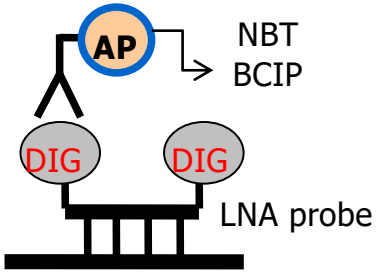




Primary anti-DIG antibodies directly conjugated with enzyme reporter proteins (Alkaline Phosphatase) that uses fluorescent substrates (Fast Red) or chromogenic chemical compounds that gives a very stable dark blue dye, visible with brightfield microscopy.

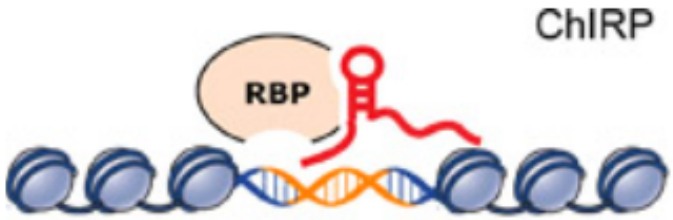
chromogenic





ChIRP
(Chromatin Isolation by Rna Purification)

AIM: Identification of the genomic binding sites of long noncoding RNAs.



- Chromatin associated lncRNAs
- Discrimination between ***cis*** and ***trans*** action

J Vis Exp. 2012 Mar 25;(61)

Chromatin Isolation by RNA Purification (ChIRP).

Chu C, Quinn J, Chang HY.

Source

Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine.

Abstract

Long noncoding RNAs are key regulators of chromatin states for important biological processes such as dosage compensation, imprinting, and developmental gene expression (1,2,3,4,5,6,7). The recent discovery of thousands of lncRNAs in association with specific chromatin modification complexes, such as Polycomb Repressive Complex 2 (PRC2) that mediates histone H3 lysine 27 trimethylation (H3K27me3), suggests broad roles for numerous lncRNAs in managing chromatin states in a gene-specific fashion (8,9). While some lncRNAs are thought to work in cis on neighboring genes, other lncRNAs work in trans to regulate distantly located genes. For instance, Drosophila lncRNAs roX1 and roX2 bind numerous regions on the X chromosome of male cells, and are critical for dosage compensation (10,11). However, the exact locations of their binding sites are not known at high resolution. Similarly, human lncRNA HOTAIR can affect PRC2 occupancy on hundreds of genes genome-wide(3,12,13), but how specificity is achieved is unclear. LncRNAs can also serve as modular scaffolds to recruit the assembly of multiple protein complexes. The classic trans-acting RNA scaffold is the TERC RNA that serves as the template and scaffold for the telomerase complex (14); HOTAIR can also serve as a scaffold for PRC2 and a H3K4 demethylase complex (13). Prior studies mapping RNA occupancy at chromatin have revealed substantial insights (15,16), but only at a single gene locus at a time. The occupancy sites of most lncRNAs are not known, and the roles of lncRNAs in chromatin regulation have been mostly inferred from the indirect effects of lncRNA perturbation. Just as chromatin immunoprecipitation followed by microarray or deep sequencing (ChIP-chip or ChIP-seq, respectively) has greatly improved our understanding of protein-DNA interactions on a genomic scale, here we illustrate a recently published strategy to map long RNA occupancy genome-wide at high resolution (17). This method, Chromatin Isolation by RNA Purification (ChIRP) (Figure 1), is based on affinity capture of target lncRNA:chromatin complex by tiling antisense-oligos, which then generates a map of genomic binding sites at a resolution of several hundred bases with high sensitivity and low background. ChIRP is applicable to many lncRNAs because the design of affinity-probes is straightforward given the RNA sequence and requires no knowledge of the RNA's structure or functional domains.

ChIRP
(Chromatin Isolation by Rna Purification)

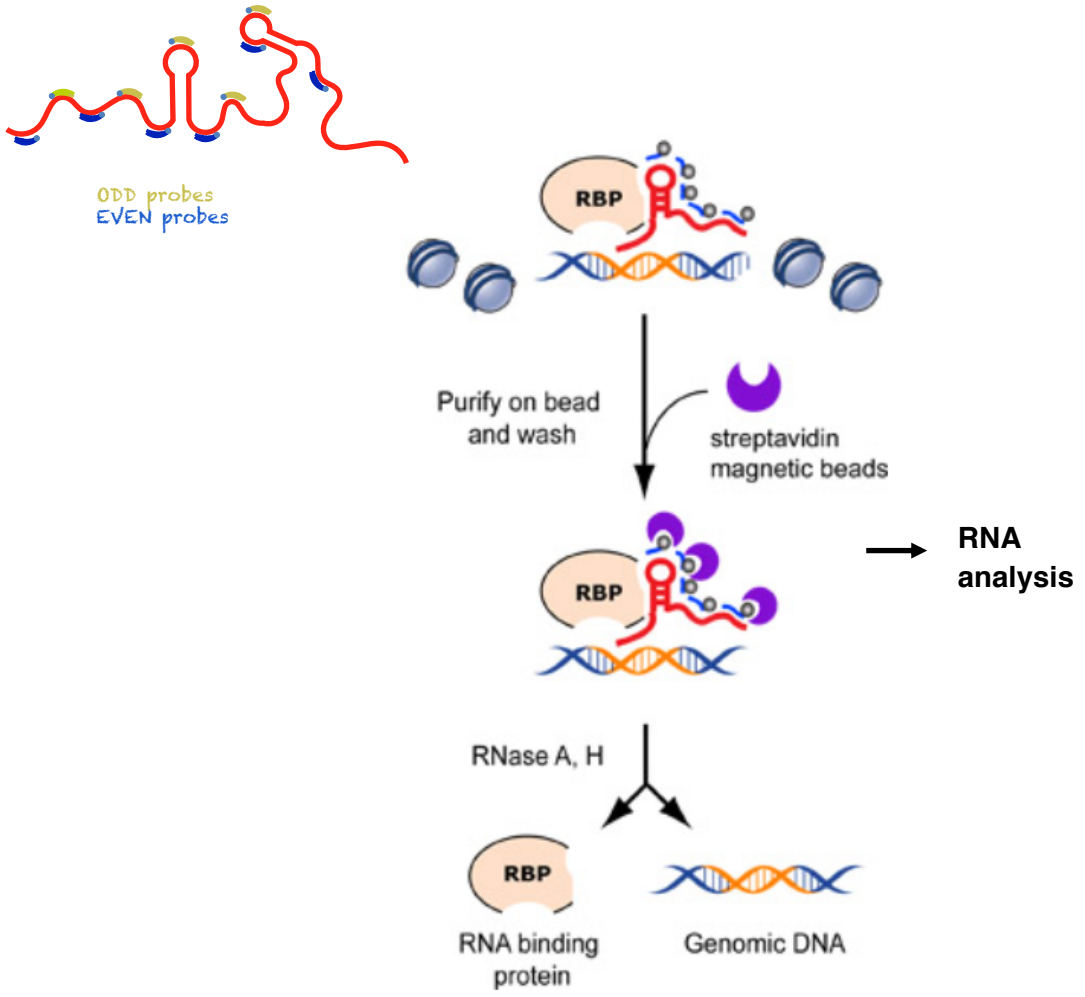
1. Probe Design

2. Cross-linking

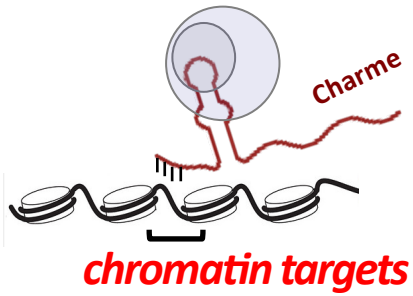
3. Chromatin Sonication

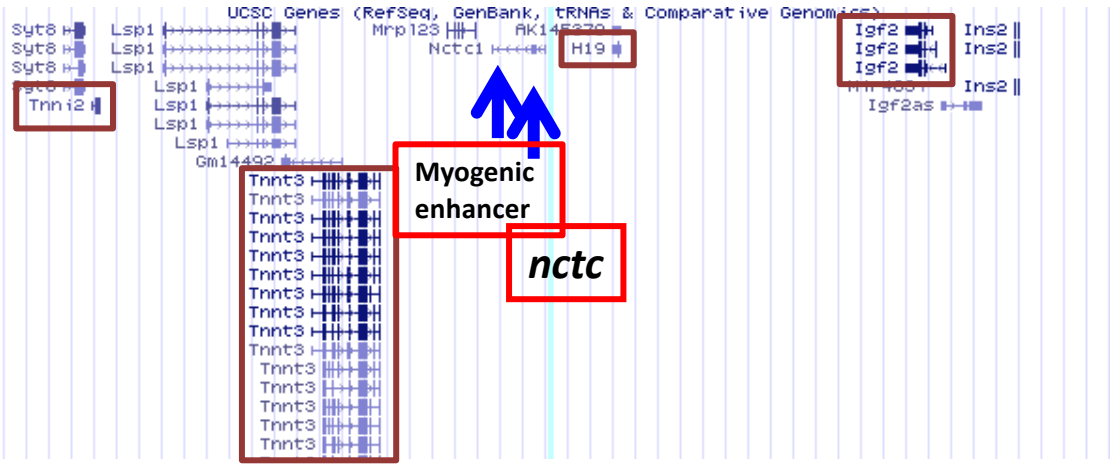
4. Probes Hybridization and precipitation

5. DNA and Protein Isolation



ChIRP: select the genomic regions revealed with both odd and even oligo sets

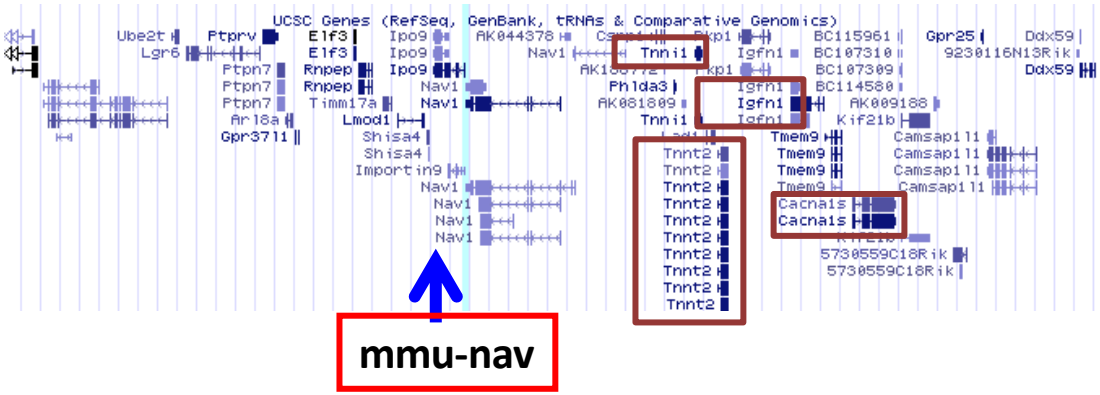




Top ranked – Charme locus

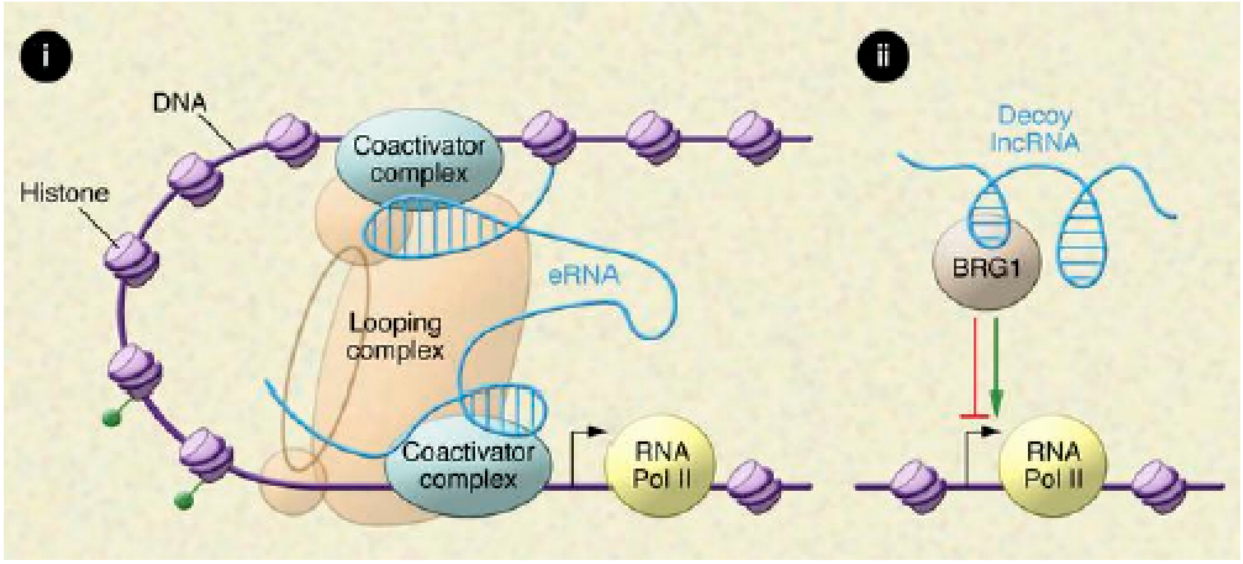
- the other two top regions contain the genes most affected by *Charme* depletion

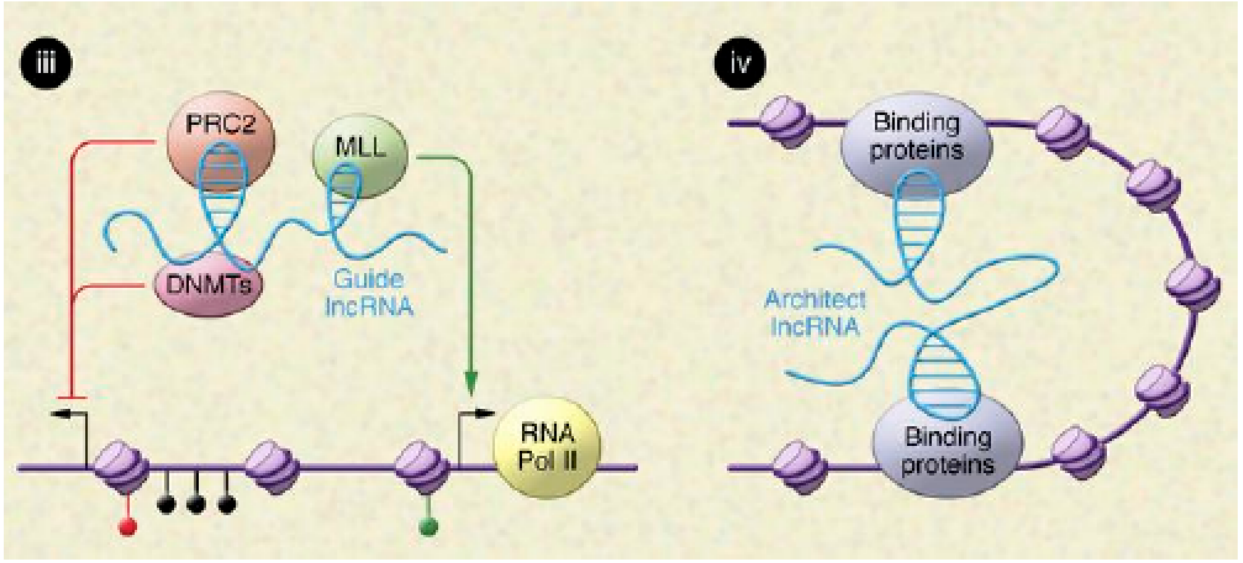
- one of them contains a crucial myogenic enhancer



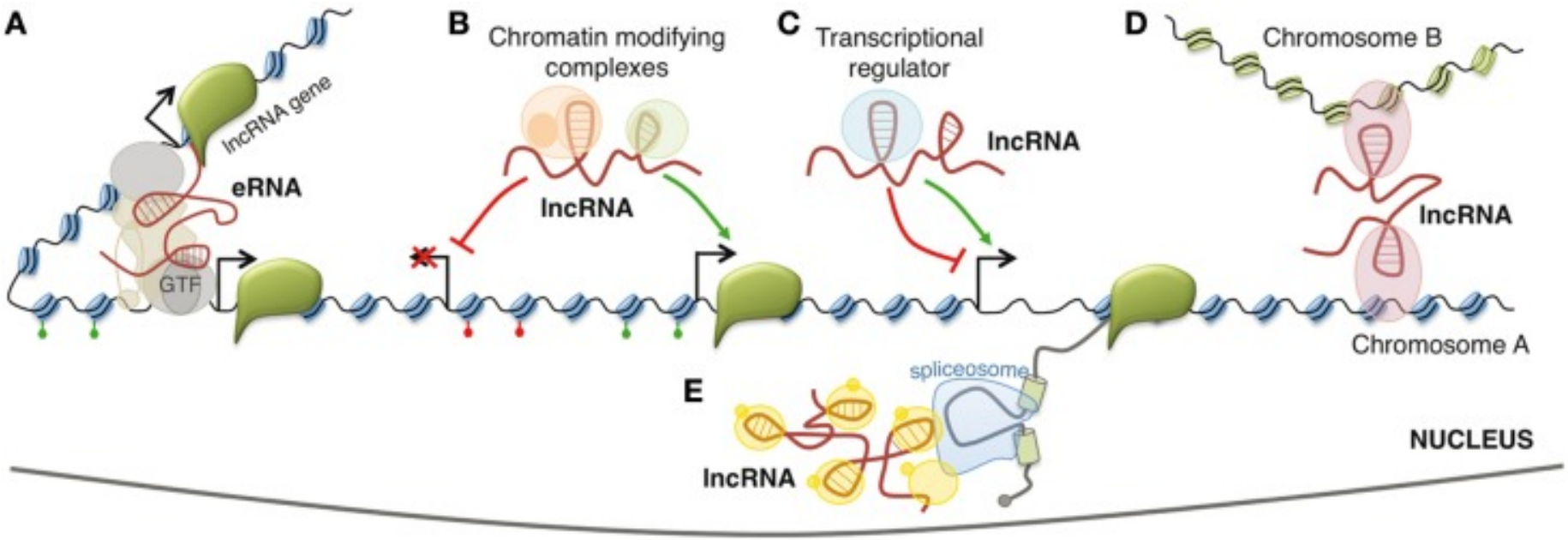
Nuclear long non coding RNAs

Models of *nuclear* lncRNA functions



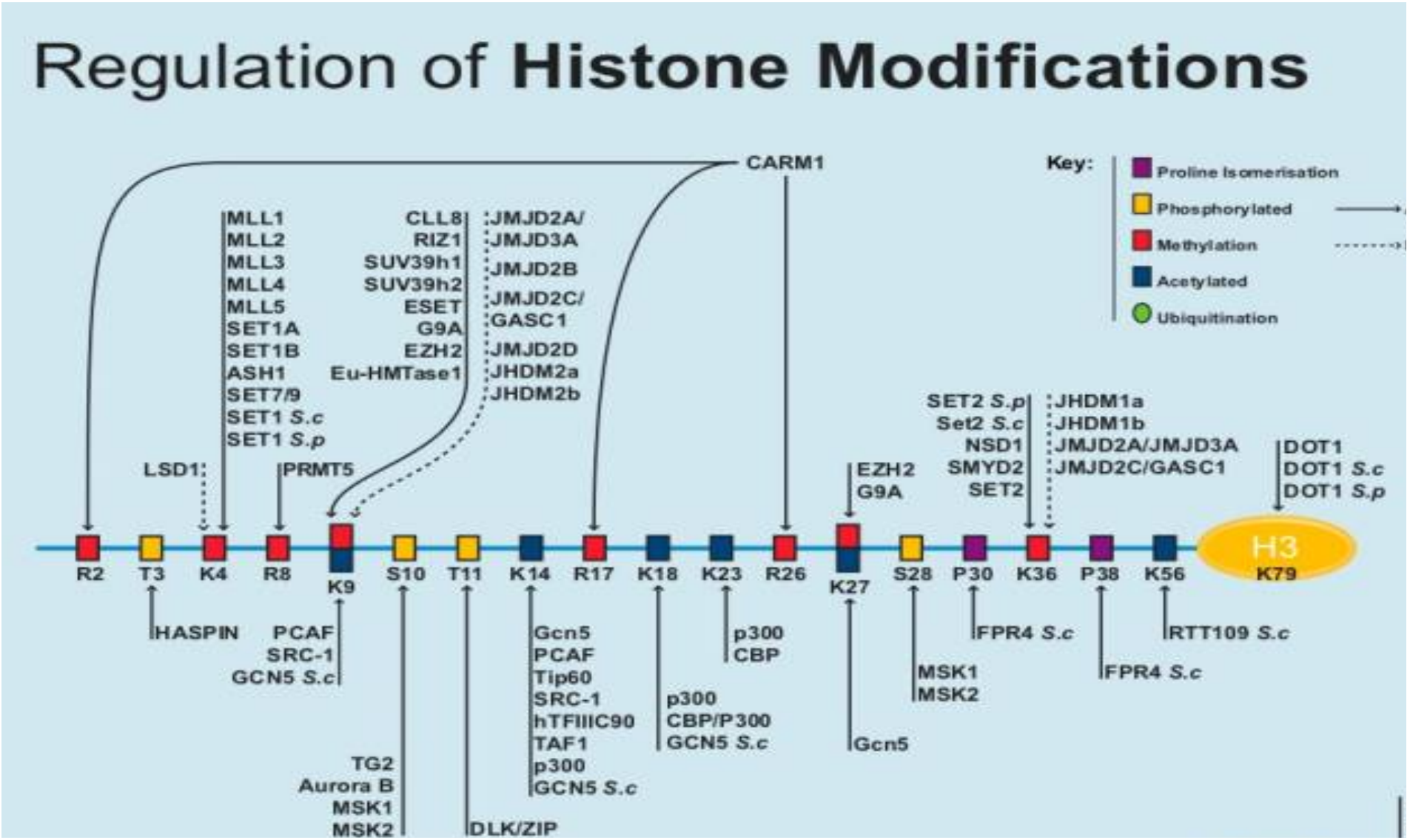


Nuclear lncRNAs may act as:
(i) **eRNAs**, which regulate transcription through enhancer-like functions (such as core enhancer RNA and DRRRNA (***MUNC***) in myogenesis);
(ii) **decoy** lncRNAs, which act by sequestering chromatin or transcriptional regulators (such as MyHeart, which inhibits the chromatin remodeling factor BRG1);
(iii) **guide** lncRNAs, which act by recruiting epigenetic regulators onto specific chromosomal loci (such as ***Bvht***, ***Fendrr***, ***DBE-T***, ***Dum***, ***Meg3***);
(iv) **architect** lncRNAs, which act by modifying the three-dimensional conformation of chromatin (such as ***Kcnq1ot1*** lncRNA). Activating (green) or repressing (red) histone modifications together with the sites of DNA methylation (black) are indicated.



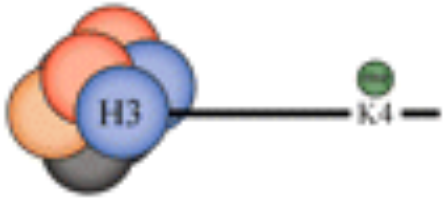
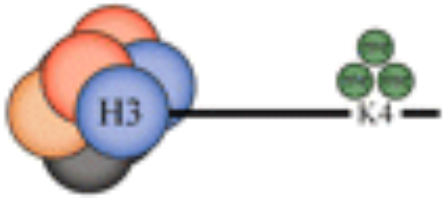
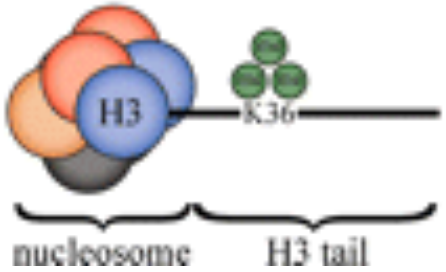
Munc (**A**, **C**)
Anril (**B**) *cis*
HOTTIP (**B**) *cis*
Xist (**B**) *cis* and *trans*
HOTAIR (**B**) *trans*
FIRRE, *Charme* (**D**)

Regulate gene expression



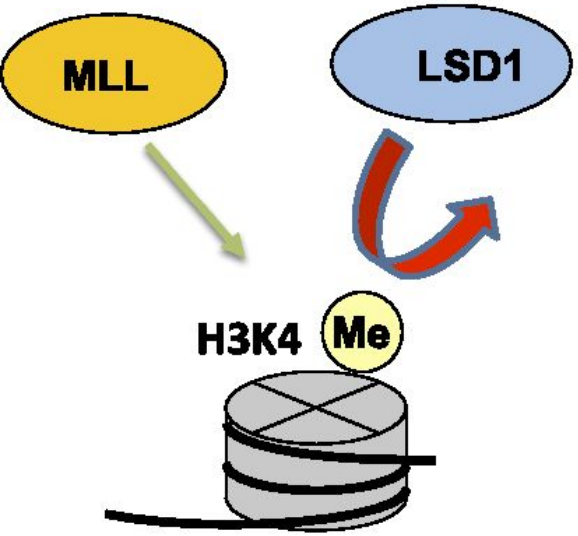
Adds modifications

Removes modifications

Modification	Localization	Writer	Writer complex
 H3K4me1	enhancers	?	?
 H3K4me3	promoters	Set1 (yeast) hSET1A (human) hSET1B (human) MLL1-4 (human)	COMPASS (yeast) MLL1-4 complexes (human)
 H3K36me3	transcribed regions	Set2 (yeast) HYPB/Setd2 (human)	Set2 (yeast) HYPB/Setd2 (human)

Modifications of histone H3. Lysine residues on histone H3 can be mono-, di- or tri-methylated. Shown are modifications H3K4me1, H3K4me3 and H3K36me3, which mark active/poised enhancers, active/poised promoters and actively transcribed regions, respectively. me, methylation.

At least six mammalian homologs of COMPASS exist, including MLL1–4 complexes, hSET1A and hSET1B, and their recruitment to active promoters can result in H3K4me3


MLL
LSD1
H3K4
Me

Two opposing groups of histone modifying complexes

Trithorax group (TrxG) of H3K4 HMTase

and

Polycomb complex (PcG) H3K27 HMTase,

maintain open and closed chromatin domains in the HOX loci, respectively, over successive cell division (Ringrose and Paro, 2007).

Generally, **transcriptionally silent regions** contain
H3K9me3 (trimethyl),
H3K27me2/3 (di- and trimethyl),
H4K20me1 (monomethyl),

whereas **active genes correlate** with
H3K4me2/3 (di- and trimethyl),
H3K36me2/3 (di- and trimethyl),
H3K79me2 (dimethyl)

(Sims et al. 2003; Margueron et al. 2005; Martin and Zhang 2005).

Polycomb repressive complex (PRC)

Polycomb repressive complex (PRC), a chromatin remodeling complex, is an epigenetic gene silencer and crucial regulator of genomic programming and differentiation. This complex cooperates in transcriptional repression of target genes by altering chromatin structure

Polycomb group proteins form large multimeric complexes of 2 general types:

PRC2 possesses histone H3K27 methyltransferase activity

PRC1 components ubiquitinate H2A following the H3K27 trimethylation

Enhancer of Zeste homolog (EZH2) histone methyltransferase is an enzymatic subunit of PRC2 and methylates H3K27 to mediate gene silencing. The Bmi-1 (RING finger protein 51) is a PRC1 protein that binds to the H3K27 trimethylation and catalyzes the ubiquitination of histone H2A

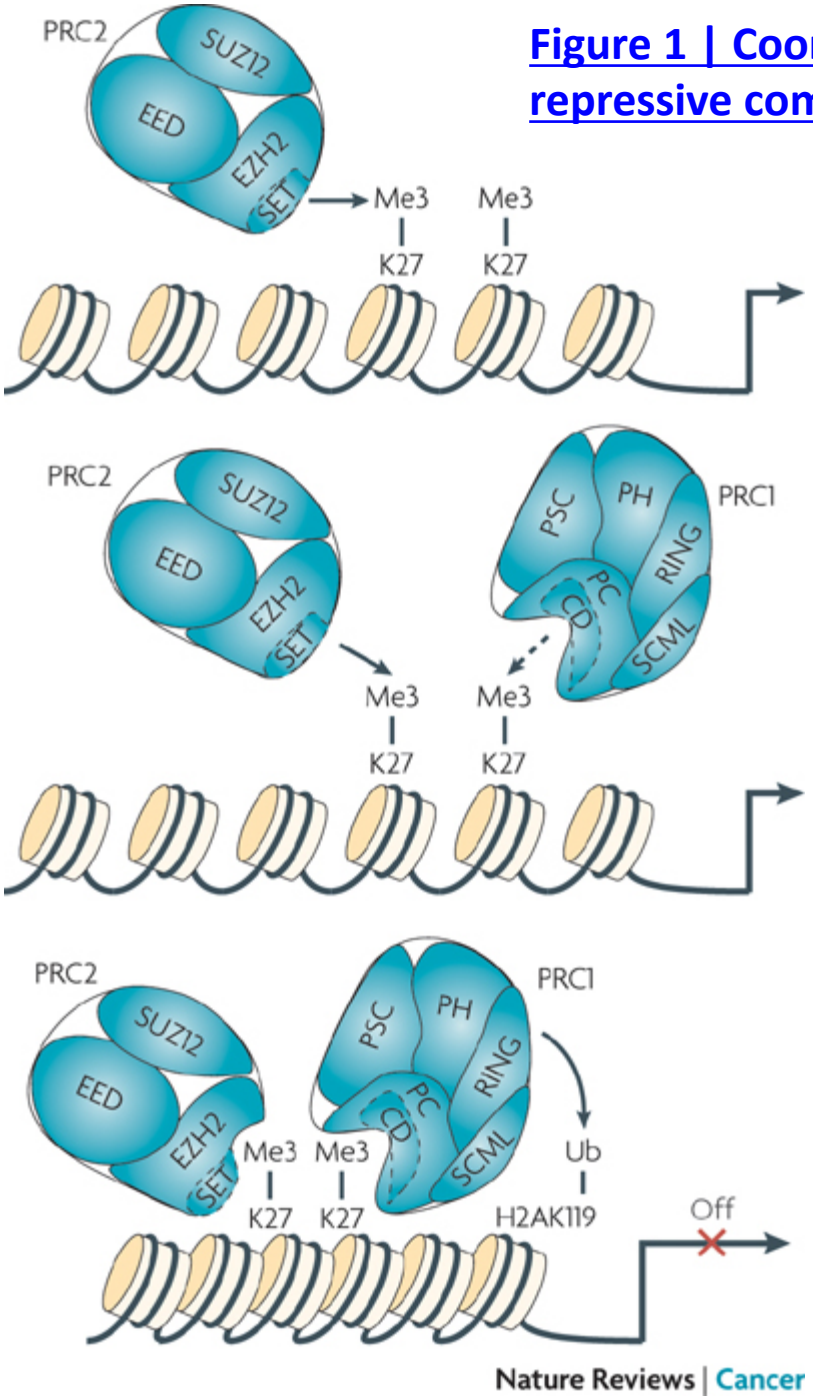
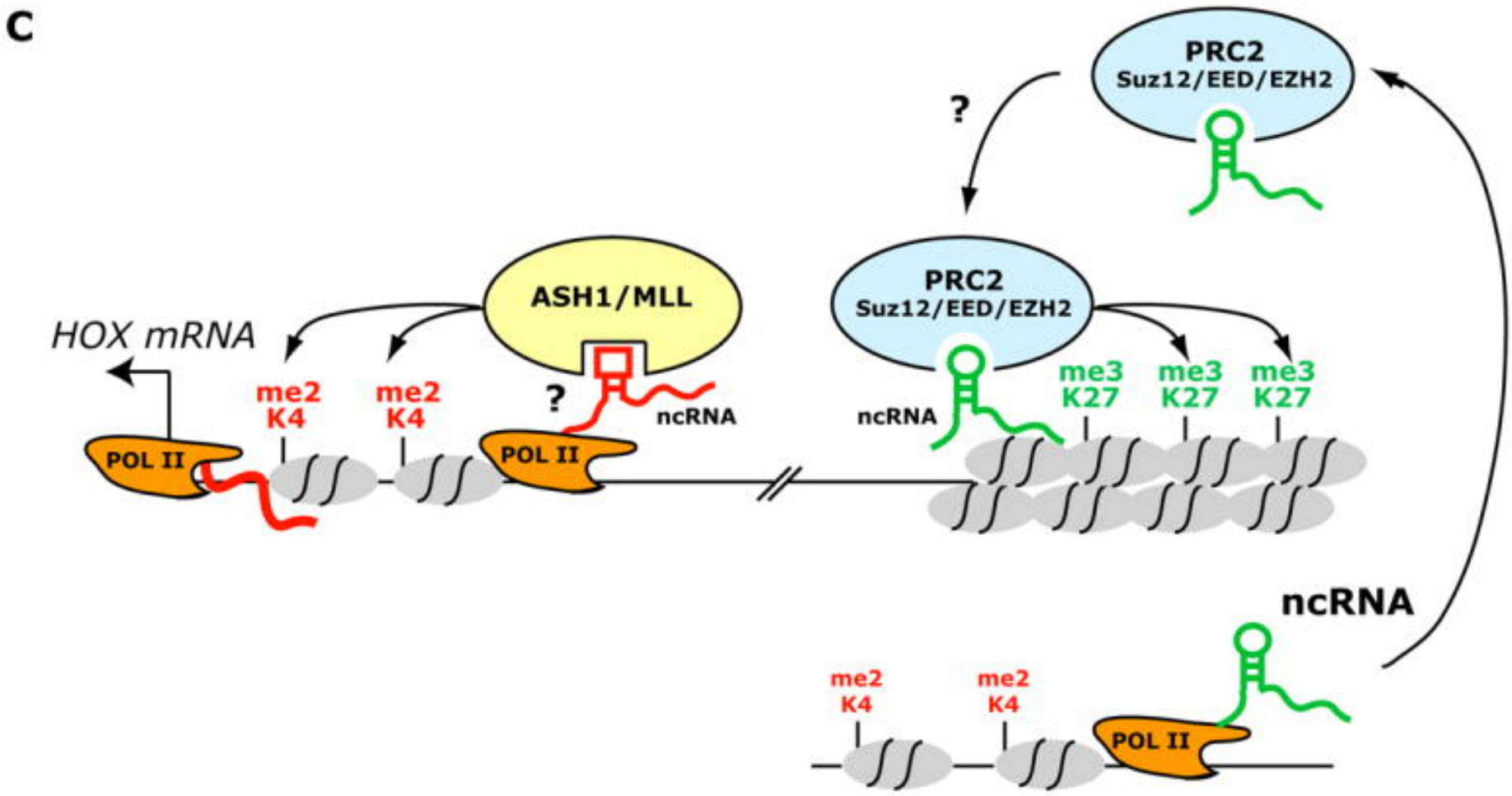
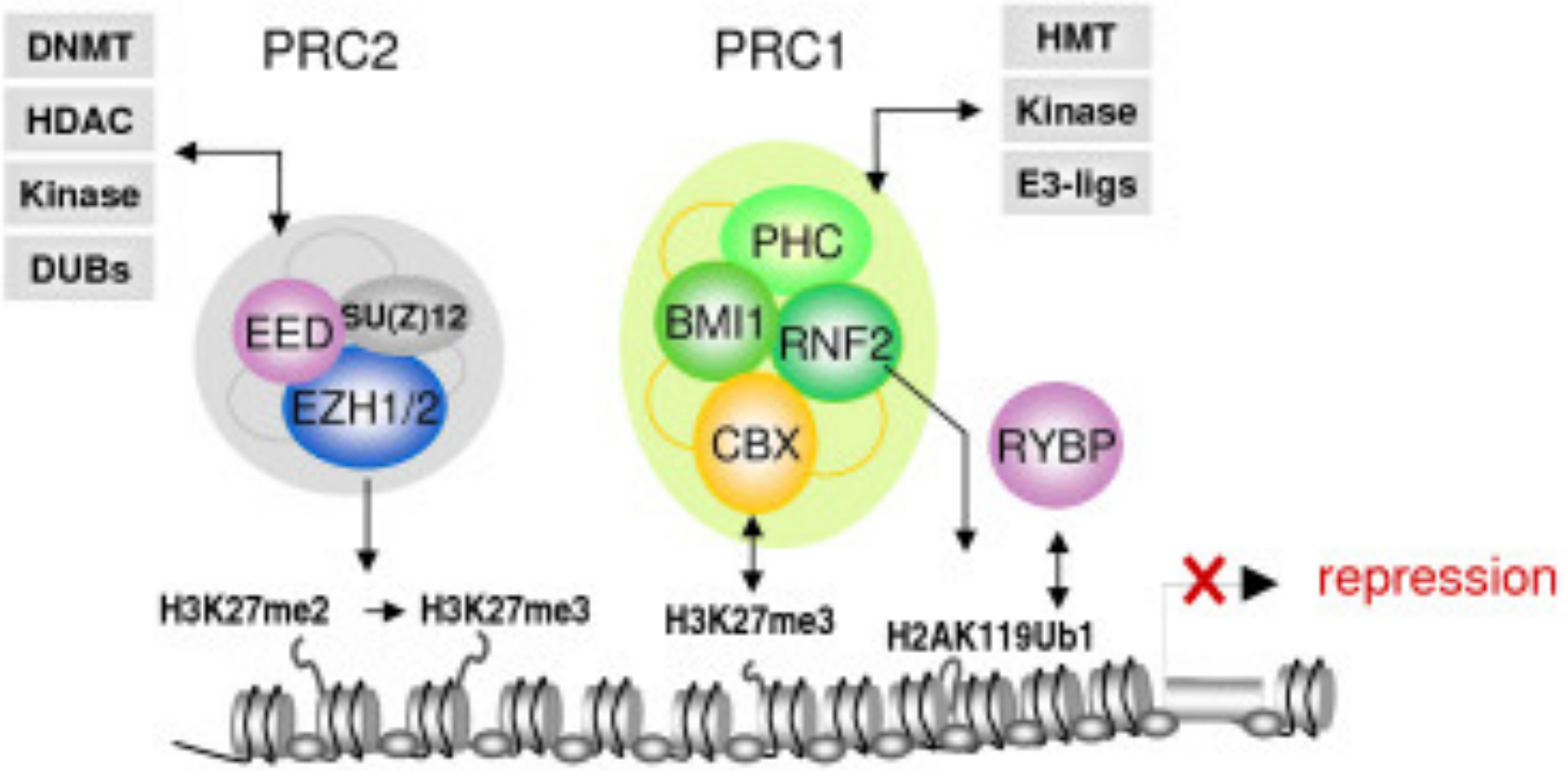


Figure 1 | Coordinated action of Polycomb repressive complexes.

Two major Polycomb repressive complexes (PRCs) have been described. The PRC2 complex contains the histone methyltransferase enhancer of zeste homologue 2 (EZH2), which together with embryonic ectoderm development (EED) and suppressor of zeste 12 homolog (SUZ12) catalyses the trimethylation of histone H3 at lysine K27 (H3K27me3). The EZH2 SET domain confers this activity. Multiple forms of the PRC1 complex exist and these contain combinations of at least four PC proteins (CBX2, CBX4, CBX7 and CBX8), six PSC proteins (BMI1, MEL18, MBLR, NSPC1, RNF159 and RNF3), two RING proteins (RNF1 and RNF2), three PH proteins (HPH1, HPH2 and HPH3) and two SCML proteins (SCML1 1 and SCML2). Some results have suggested that PRC1 complexes are recruited by the affinity of chromodomains in chromobox (Cbx) proteins to the H3K27me3 mark**. PRC1 recruitment results in the RNF1 and RNF2-mediated ubiquitylation of histone H2A on lysine 119, which is thought to be important for transcriptional repression**. PC, Polycomb; PSC, Posterior sex combs ; SCML, Sex combs on midleg .



Model of long ncRNA regulation of chromatin domains via histone modification enzymes. Transcription of ncRNAs *in cis may increase the accessibility of TrxG proteins such as ASH1 or MLL or directly recruit them, leading to H3K4 methylation and* ***transcriptional activation*** *of the downstream HOX gene(s). In contrast, recruitment of PRC2 is programmed by ncRNAs produced in trans, which targets PRC2 activity by yet incompletely defined mechanisms to target loci. PRC2 recruitment leads to H3K27 methylation and* ***transcriptional silencing*** *of neighboring HOX genes.*



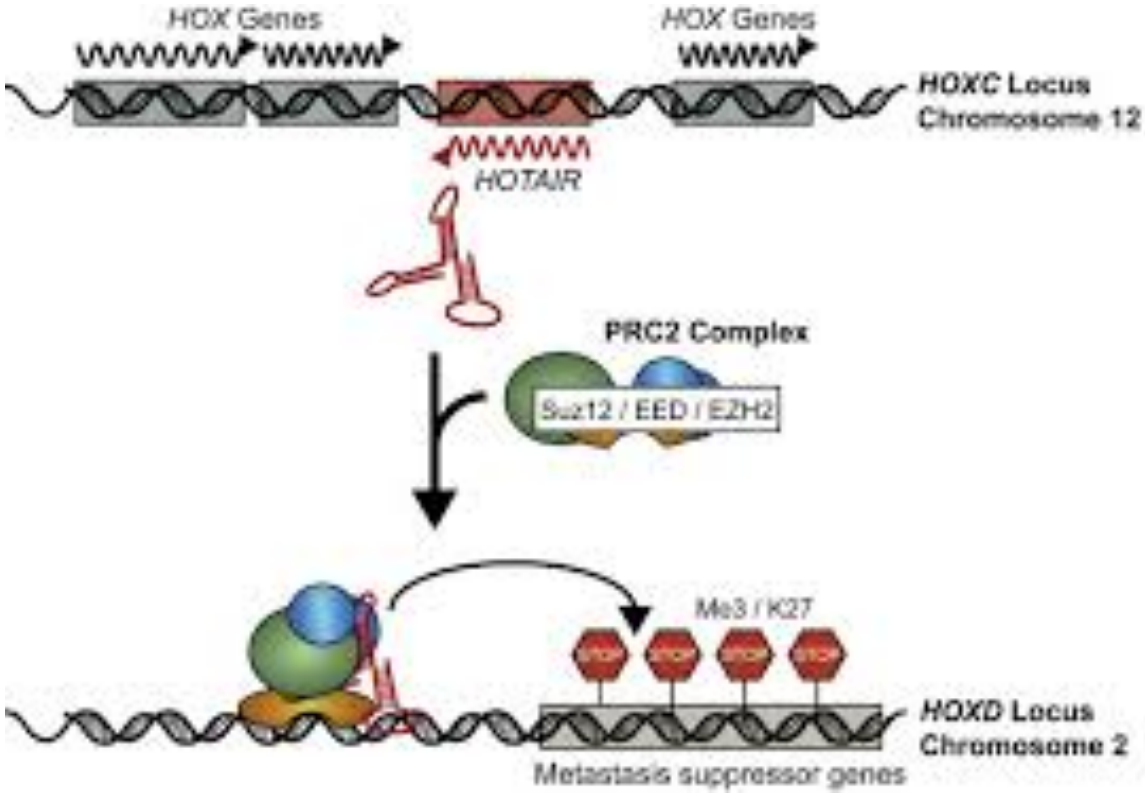
A model of how Polycomb complexes work. The initial methylation of histones by the ESC-E(Z) complex sets the stage for PRC1 binding, thereby maintaining the repressed states of the corresponding regions.

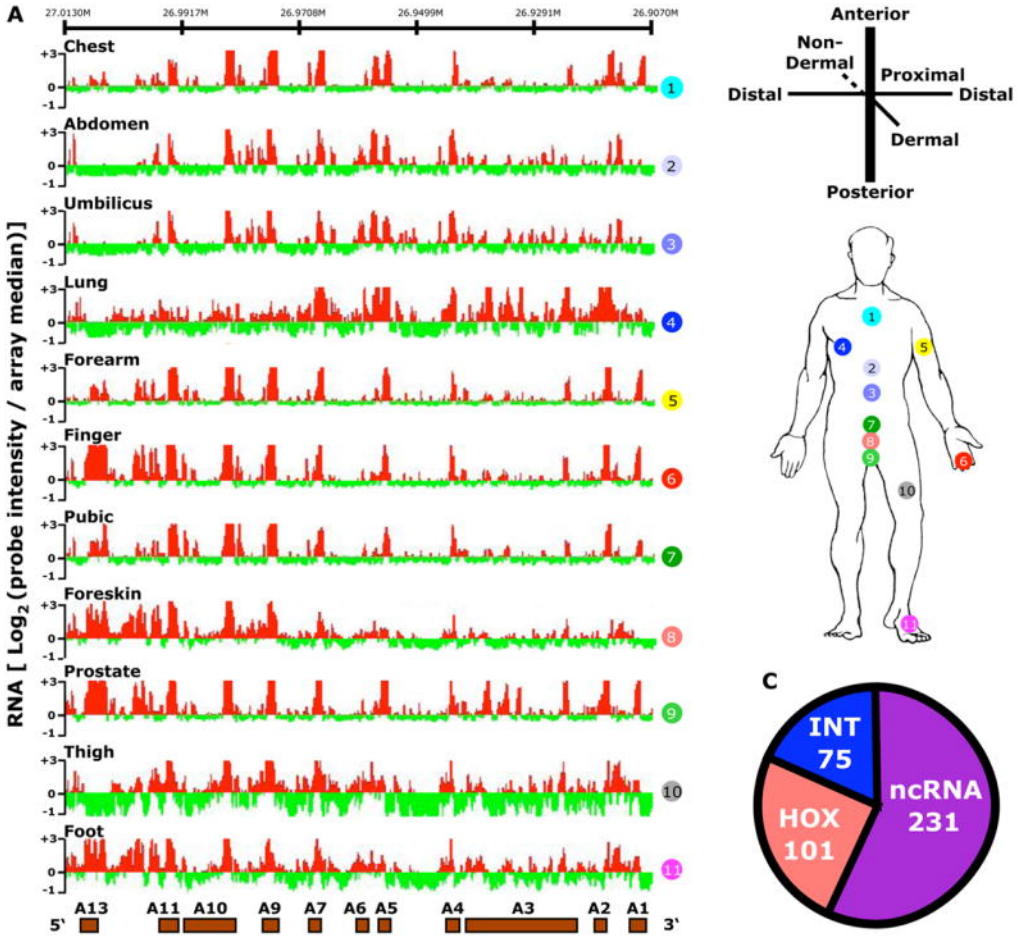
Cell. 2007 Jun 29;129(7):1311-23. **HOTAIR discovery**

Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY.

Noncoding RNAs (ncRNA) participate in epigenetic regulation but are poorly understood. Here we characterize the transcriptional landscape of the four human HOX loci at five base pair resolution in 11 anatomic sites and identify 231 HOX ncRNAs that extend known transcribed regions by more than 30 kilobases. HOX ncRNAs are spatially expressed along developmental axes and possess unique sequence motifs, and their expression demarcates broad chromosomal domains of differential histone methylation and RNA polymerase accessibility. We identified a 2.2 kilobase ncRNA residing in the HOXC locus, termed **HOTAIR**, **which represses transcription in trans across 40 kilobases of the HOXD locus**. HOTAIR interacts with Polycomb Repressive Complex 2 (PRC2) and is required for PRC2 occupancy and histone H3 lysine-27 trimethylation of HOXD locus. Thus, transcription of ncRNA may demarcate chromosomal domains of gene silencing at a distance; these results have broad implications for gene regulation in development and disease states.


HOX Genes
HOX Genes
HOXC Locus
Chromosome 12
HOTAIR
PRC2 Complex
Suz12 / EED / EZH2
Me3 / K27
HOXD Locus
Chromosome 2
Metastasis suppressor genes



The human HOX transcriptome

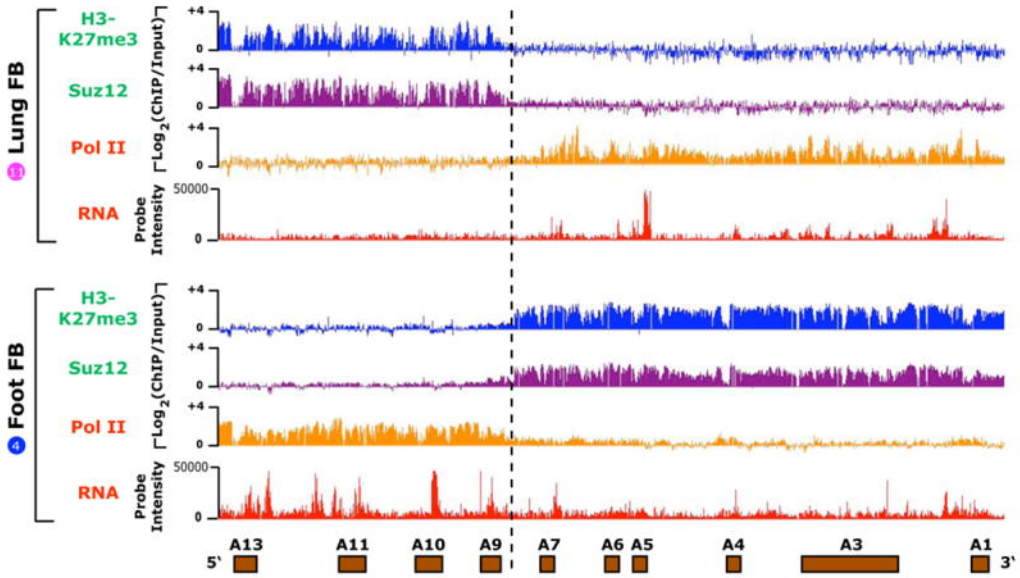
(A) Site-specific transcription of the HOXA locus.

red and green bars indicate expression above or below the array mean, respectively

Transcribed regions were identified by contiguous signals on tiling array, then compared with Refseq sequence to define genic [exonic (pink color) and intronic (blue)] and intergenic transcribed regions (purple).

The human HOX transcriptome

(A) Site-specific transcription of the **HOXA locus**. Left: The hybridization intensity of 50,532 probes that tile the human HOXA locus for each of the 11 samples (numbered in circles). The intensity of each probe is displayed as the $\log_2$ of the ratio of the individual probe intensity divided by the average intensity of all 301,027 probes on the array. The $\log_2$ ratio of each probe was averaged over a 100 bp window; red and green bars indicate expression above or below the array mean, respectively. Genomic locations of protein-coding HOX genes are displayed as brown boxes. Right: Anatomic origins of the 11 fibroblast samples with respect to the developmental axes. (B) **Transcribed regions were identified by contiguous signals on tiling array, then compared with Refseq sequence to define genic [exonic (pink color) and intronic (blue)] and intergenic transcribed regions (purple)**. Each predicted HOX exon or intron was named HOXn or int-HOXn, respectively. Intergenic transcribed regions were named as nc-HOXn where n is the HOX paralog located 3′ to the ncRNA on the HOX coding strand. (C) Summary of transcribed regions in all four HOX loci defining the number of HOX genic, intronic, and ncRNA transcribed regions.

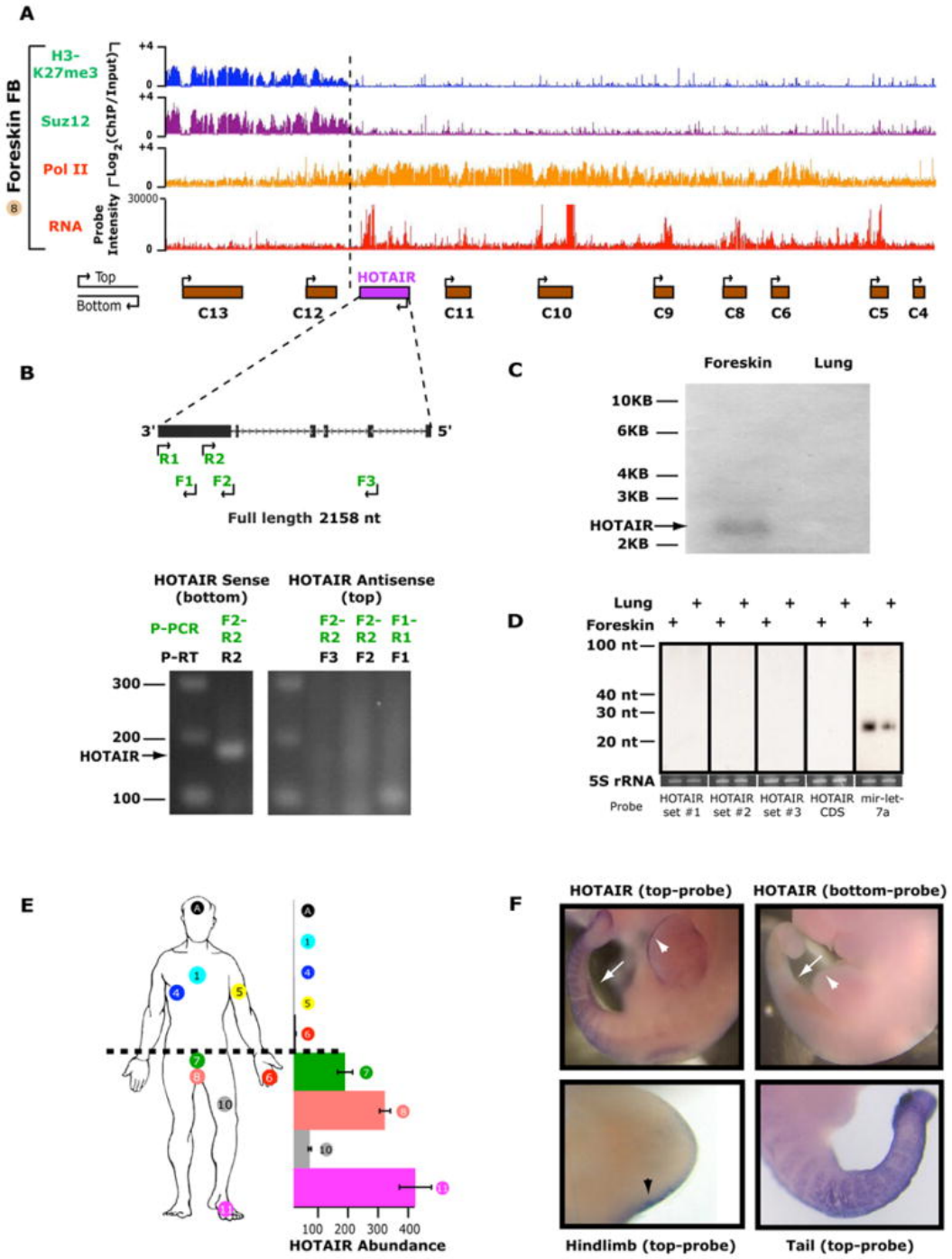


Simmetrical opposed chromatin modifications and transcriptional accessibility in the HOXA locus

Occupancy of Suz12, H3K27me3, and pol II versus transcriptional activity over ~100 kb of the HOXA locus for primary lung (top) or foot (bottom) fibroblasts (Fb). For chIP data, the $\log_2$ ratio of ChIP/Input is plotted on the Y-axis. For RNA data, the hybridization intensity on a linear scale is shown. Dashed line highlights the boundary of opposite configurations of chromatin modifications and intergenic transcription.

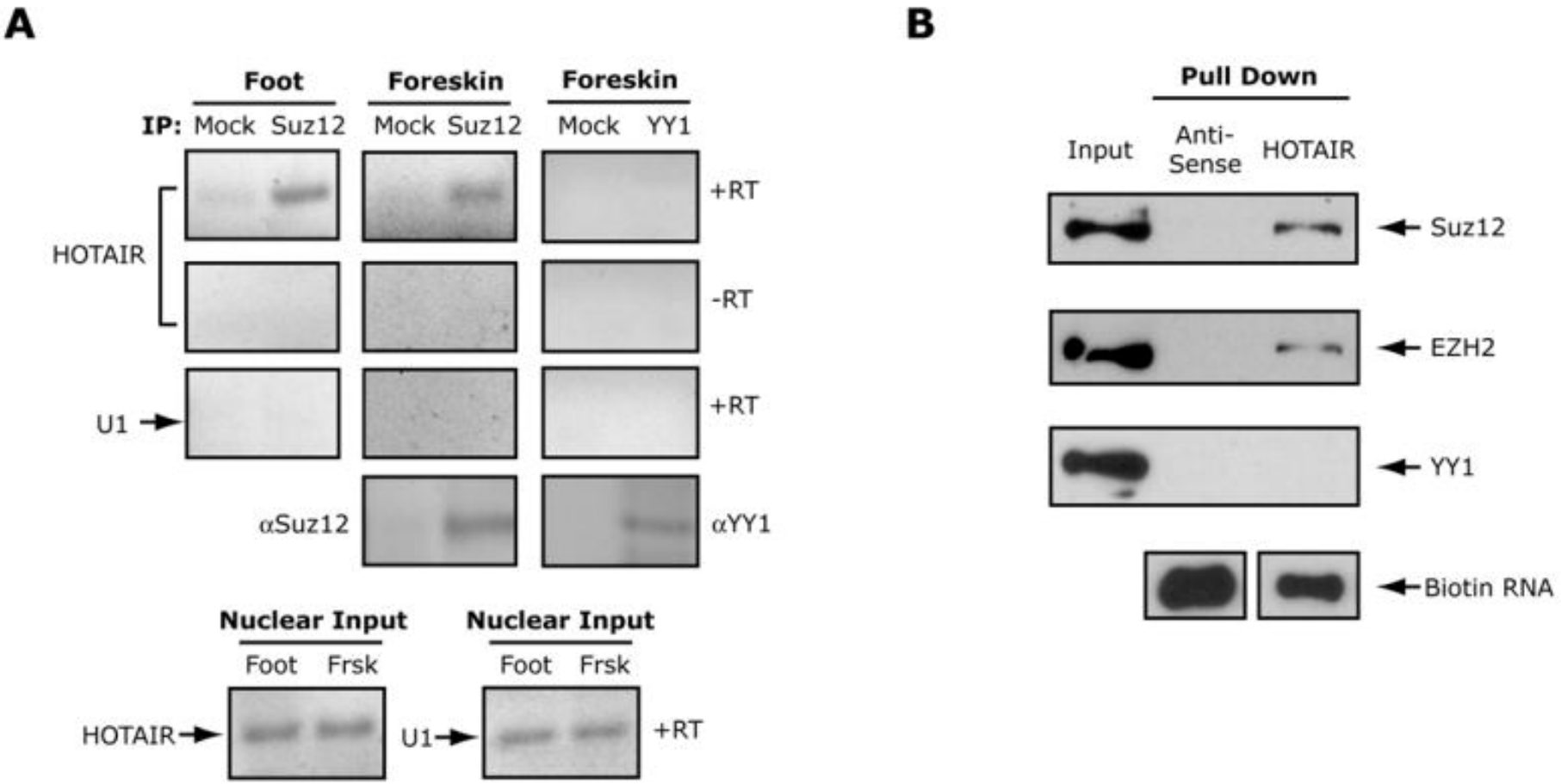
Previously, several reports indicated the link between small RNAs, derived from the RNAi pathway, and transcriptional gene silencing in plants and lower eukaryotes. However, for long non-coding RNAs there were only indirect evidences of possible interconnection with chromatin; namely Xist RNA in the X-chromosome inactivation and ncRNAs in imprinted loci. Other reports instead described the interaction of several PcG proteins with RNA, and that their asssociation with chromatin partly depended on the presence of RNA. Nevertheless, the functional and mechanistic insights as to how RNA could influence either active or repressed chromatin and how these epigenetic states are established and maintained during development or cell differentiation remained major unanswered questions.

It is truly surprising how detailed the transcriptional analysis performed is! One year before the burst of RNA-seq methodologies profiling these landscapes must have been very challenging indeed. While conventional gene expression microarrays were only designed to detect the expression of protein-coding mRNAs, unbiased RNA detection methods had to be conceived. John Rinn had already established tiling arrays at high resolution in 2003, covering all of the unique sequences of human Chromosome 22. In this paper, **the tiling array has been adapted to all four human HOX loci,** allowing the authors to distinguish amo



HOTAIR, an antisense intergenic long ncRNA of the HOXC locus

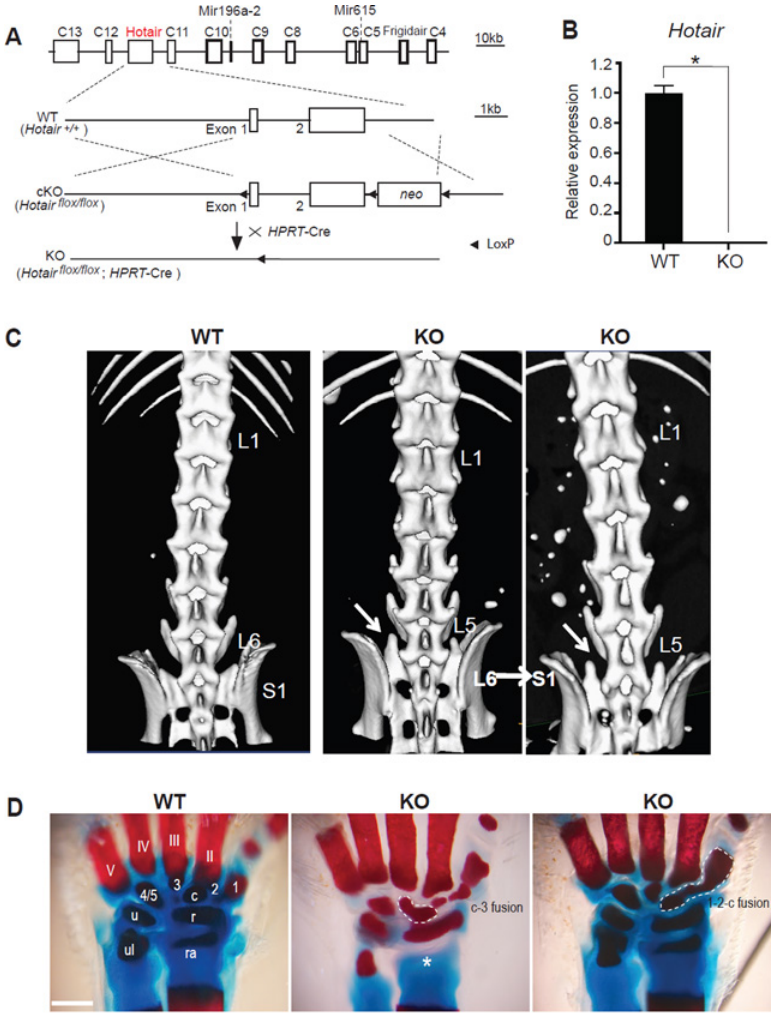
(A) Genomic location of HOTAIR at the boundary of two chromatin domains. ChIP-chip and RNA expression on tiling array are as shown in Fig. 3.
(B) Strand specific RT-PCR shows exclusive expression of HOTAIR from the strand opposite to HOXC genes (bottom). Primers for reverse transcription (P-RT) and PCR (P-PCR) were designed to specifically target either the top (primers F1–F3) or bottom strand (primer R1) of HOTAIR.
(C) Northern blot analysis of HOTAIR in lung and foreskin fibroblast RNA.
(D) Size-fractionated small RNA was probed with pools of oligonucleotides spanning HOTAIR (sets #1–3), full length antisense HOTAIR (CDS), or a probe against miRNA let7a.
(E) Posterior and distal expression of HOTAIR in human fibroblasts as measured by qRT-PCR. The site of origin of each fibroblast sample is indicated by the sample number on the anatomic cartoon. “A” is derived from the scalp. The relative abundance of HOTAIR in each position, relative to scalp (most anterior) is shown on the X-axis.
(F) Whole mount in situ hybridization using HOTAIR sense (bottom strand) or antisense (top strand) probes in embryonic day 10.5 whole mount embryos. (top panels) and the hind limb and tail (bottom left and right panels, respectively). Expression of HOTAIR in posterior hindlimb (arrowhead) and tail (arrow) are highlighted.



HOTAIR ncRNA binds Polycomb Repressive Complex 2 (A) Immunoprecipitation of Suz12 retrieves endogenous HOTAIR. Nuclear extracts of foot or foreskin fibroblasts were immunoprecipiated by IgG (lanes 1, 3, 5), anti-Suz12 (lanes 2, 4), or anti-YY1 (lane 6). Co-precipitated RNAs were detected by RT-PCR using primers for HOTAIR (rows 1 and 2) or U1 small nuclear RNA (row 3). To demonstrate that the HOTAIR band was not due to DNA contamination, each RT-PCR was repeated without reverse transcriptase (-RT, row 2). Immunoprecipitation of Suz12 and YY1 were successful as demonstrated by IP-western using the cognate antibodies (row 4). RT-PCR of nuclear extracts demonstrated equal input RNAs (row 5).(B) *In vitro transcribed HOTAIR retrieves PRC2 subunits. Immunoblot analysis of the indicated proteins is shown; five percent of input extract (5 µg) was loaded as input control.*

Cell Rep. 2013 Oct 17;5(1):3-12.

Targeted disruption of Hotair leads to homeotic transformation and gene derepression.



targeted deletion of mouse Hotair lncRNA leads to derepression of hundreds of genes, resulting in homeotic transformation of the spine and malformation of metacarpal-carpal bones

E

Skeletal abnormalities in *Hotair* KO mice:

Phenotype	Penetrance of skeletal abnormalities		
	KO	WT	*p* Value
Lumbar-sacral transition (L6→S1)	58% (n=31)	6% (n=18)	0.0002
Wrist defects	56% (n=18)	9% (n=11)	0.019

Mol Cell Biol. 2015 Feb;35(3):498-513. doi: 10.1128/MCB.01079-14. Epub 2014 Nov 17.

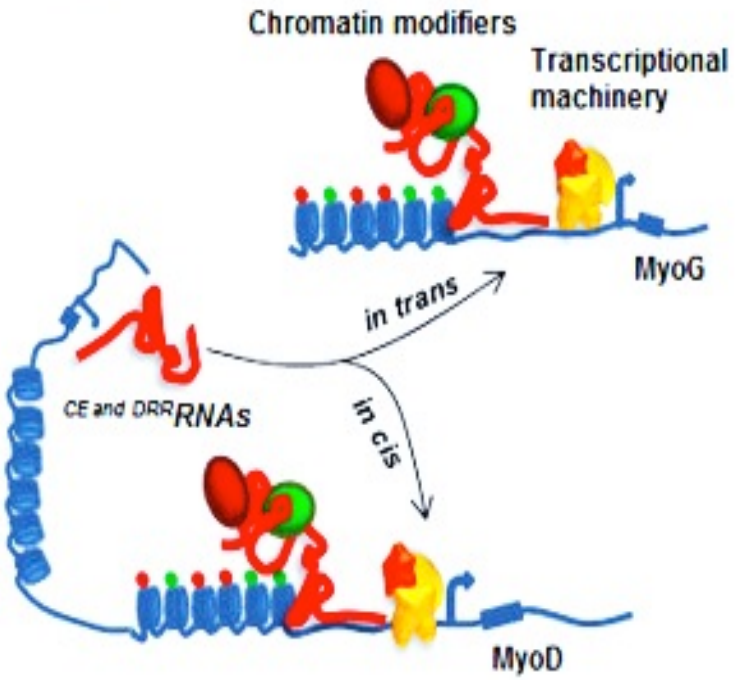
MUNC, a long noncoding RNA that facilitates the function of MyoD in skeletal myogenesis.

Mueller AC[1], Cichewicz MA[1], Dey BK[1], Layer R[1], Reon BJ[1], Gagan JR[1], Dutta A[2].

Author information

Abstract

An in silico screen for myogenic long noncoding RNAs (lncRNAs) revealed nine lncRNAs that are upregulated more than 10-fold in myotubes versus levels in myoblasts. One of these lncRNAs, MyoD upstream noncoding (MUNC, also known as DRR(eRNA)), is encoded 5 kb upstream of the transcription start site of MyoD, a myogenic transcription factor gene. MUNC is specifically expressed in skeletal muscle and exists as in unspliced and spliced isoforms, and its 5' end overlaps with the cis-acting distal regulatory region (DRR) of MyoD. Small interfering RNA (siRNA) of MUNC reduced myoblast differentiation and specifically reduced the association of MyoD to the DRR enhancer and myogenin promoter but not to another MyoD-dependent enhancer. Stable overexpression of MUNC from a heterologous promoter increased endogenous MyoD, Myogenin, and Myh3 (myosin heavy chain, [MHC] gene) mRNAs but not the cognate proteins, suggesting that MUNC can act in trans to promote gene expression but that this activity does not require an induction of MyoD protein. MUNC also stimulates the transcription of other genes that are not recognized as MyoD-inducible genes. Knockdown of MUNC in vivo impaired murine muscle regeneration, implicating MUNC in primary satellite cell differentiation in the animal. We also discovered a human MUNC that is induced during differentiation of myoblasts and whose knockdown decreases differentiation, suggesting an evolutionarily conserved role of MUNC lncRNA in myogenesis. Although MUNC overlaps with the DRR enhancer, our results suggest that MUNC is not a classic cis-acting enhancer RNA (e-RNA) acting exclusively by stimulating the neighboring MyoD gene but more like a promyogenic lncRNA that acts directly or indirectly on multiple promoters to increase myogenic gene expression.



MUNC facilitates **MyoD binding** to the **DRR**

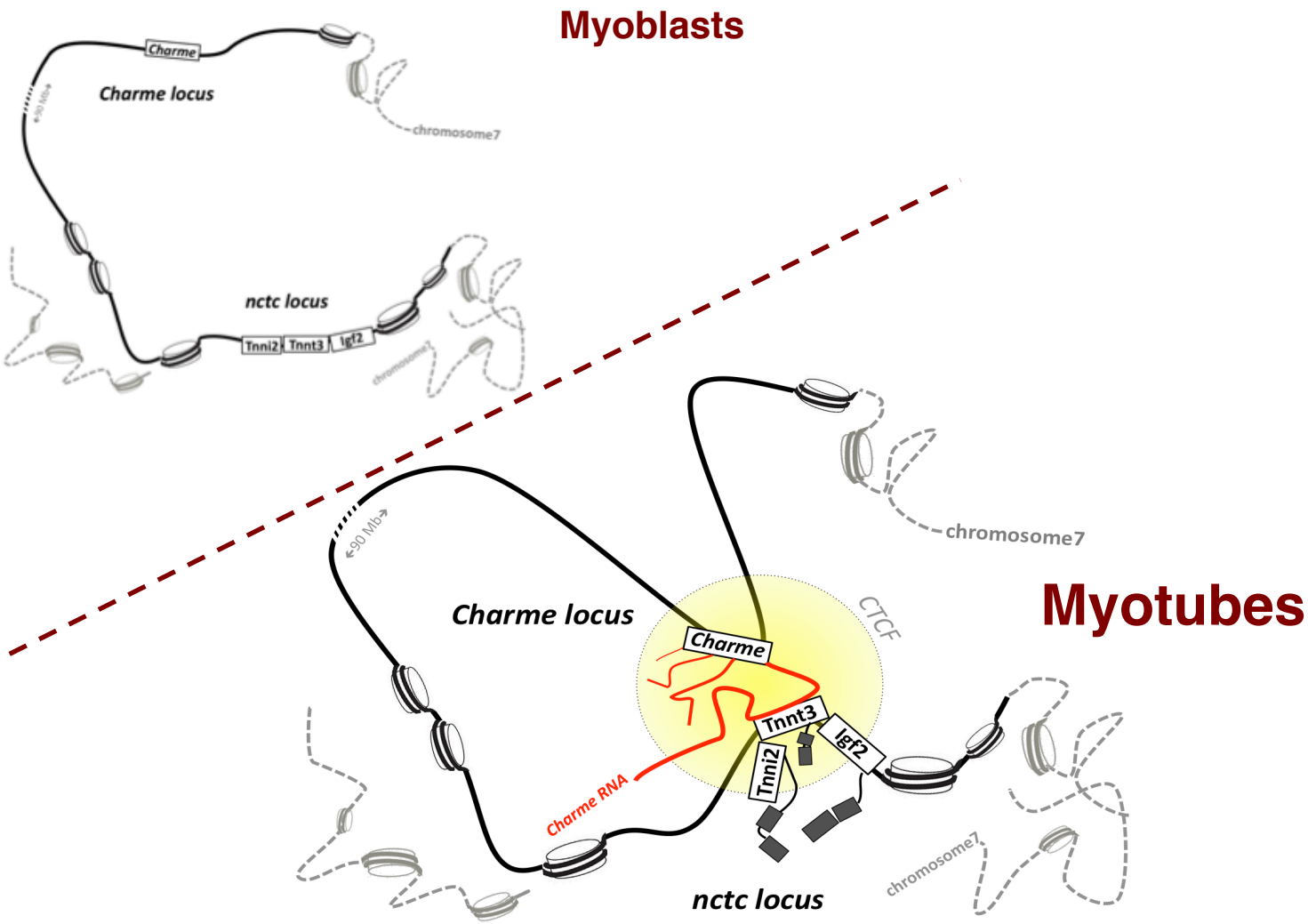
and to a lesser extent to the ***Myogenin*** **promoter**

thereby

promoting the expression of MyoD targets (*Myogenin*, *MyoD* and MUNC)

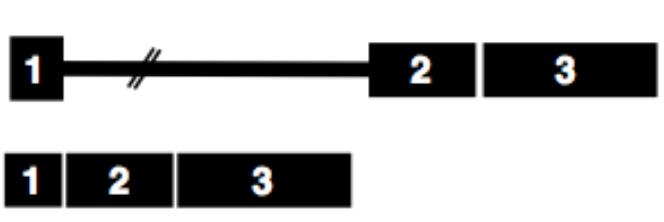
Charme

Chromatin ***ar***chitect of ***m***uscle ***e***xpression



lncRNA_*Charme*

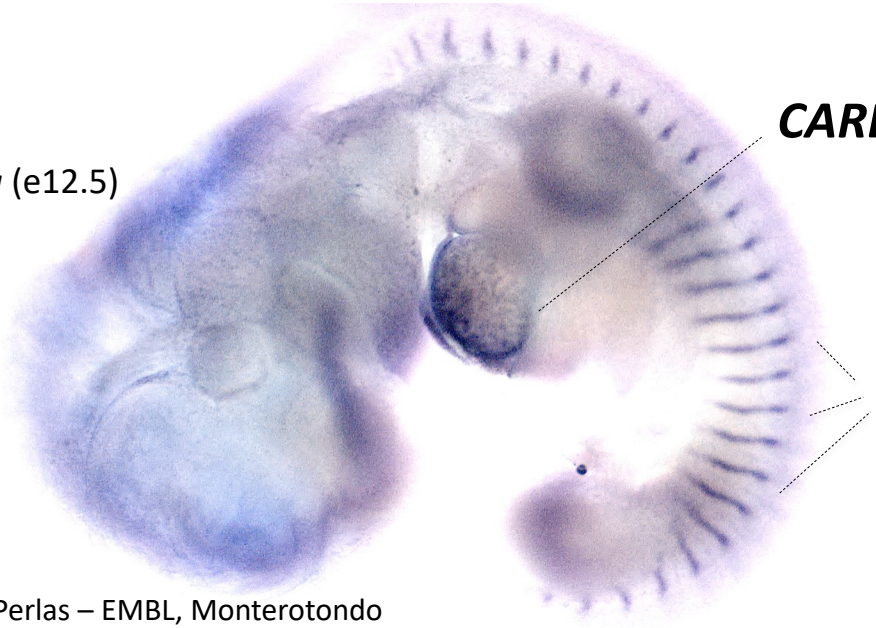
hsCharme



Charme isoforms

=

in situ (e12.5)



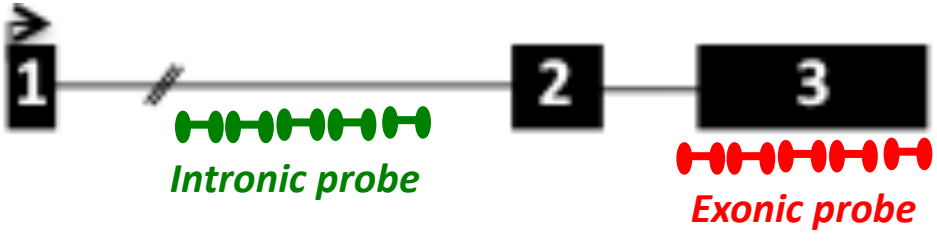
CARDIAC muscle

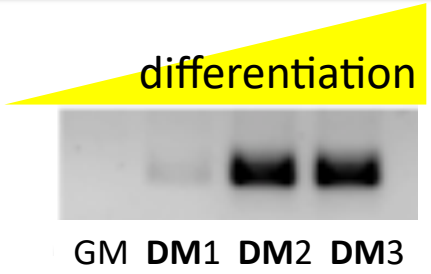
SKELETAL muscle

E.Perlas – EMBL, Monterotondo

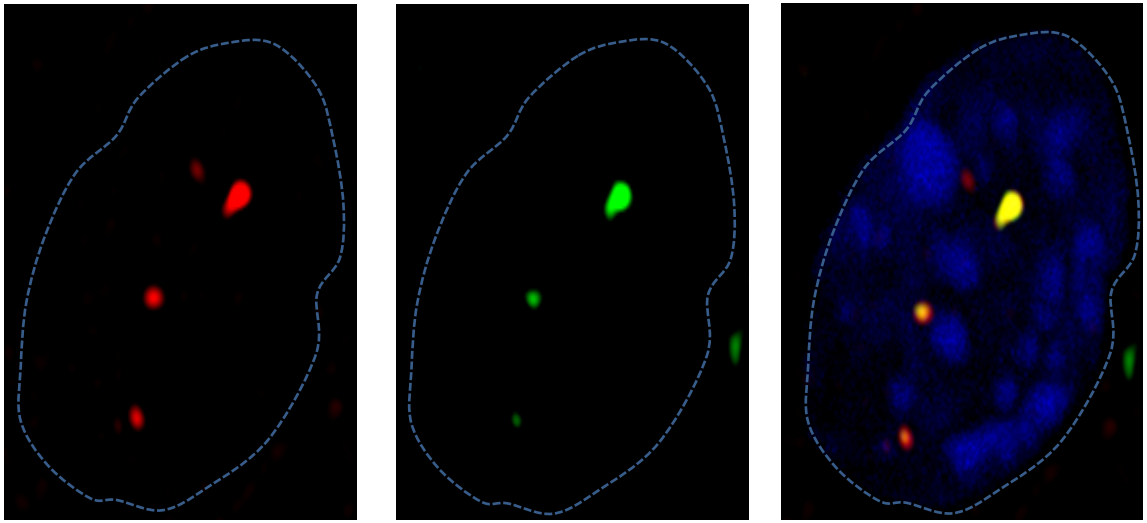
Ballarino M., Cipriano et al, 2018 EMBO J.

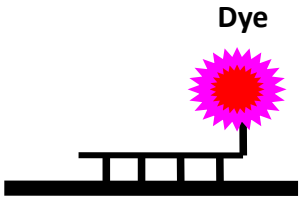
Charme is nuclear – most remains at the sites of its own transcription
(C2C12 cells are triploid for chromosome 11)





myofibers



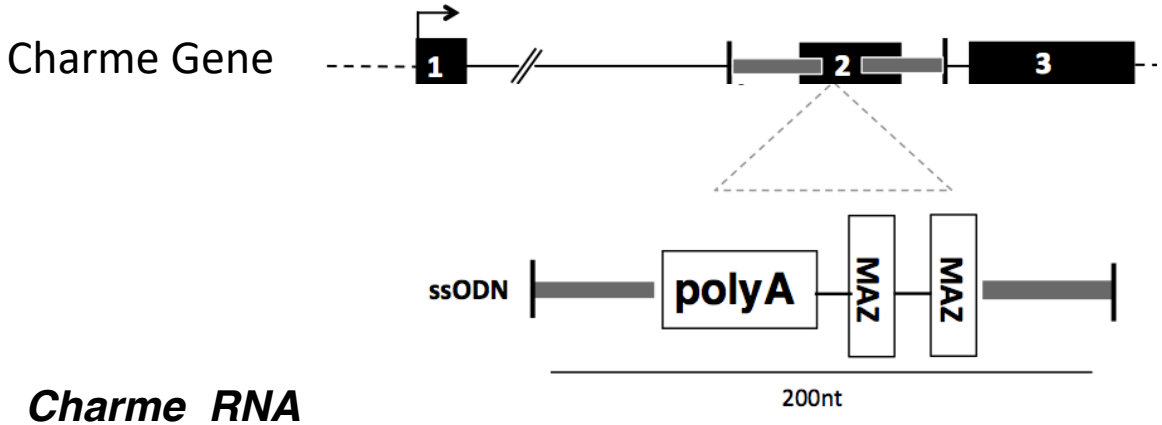


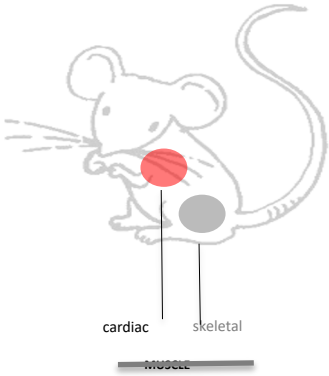
RNAseq of *Charme*-depleted cells

several down-regulated genes are associated with different types of cardiomyopathies

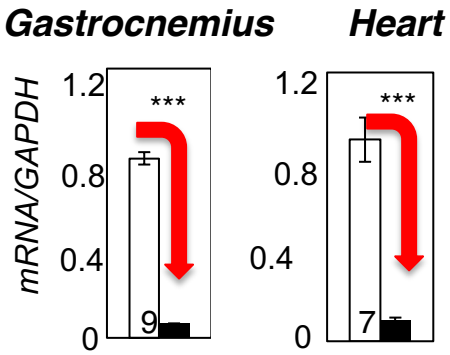
Gene	Simbolo	Prevalenza (%)
• Proteine sarcomeriche		
Catena pesante della β-miosina	MYH7	25-30
Proteina C legante la miosina	MYBPC3	25-30
Troponina T	TNNT2	5
Troponina I	TNNI3	~5
α-tropomiosina	TPM1	<5
Catena leggera della miosina		
Essenziale	MYL3	<1
Regolatrice	MYL2	<1
Catena pesante dell'α-miosina	MYH6	rara
Troponina C	TNNC	rara
Actina	ACTC	rara

Production of *Charme*$^{-/-}$ mice through CRISPR/Cas9 gene editing





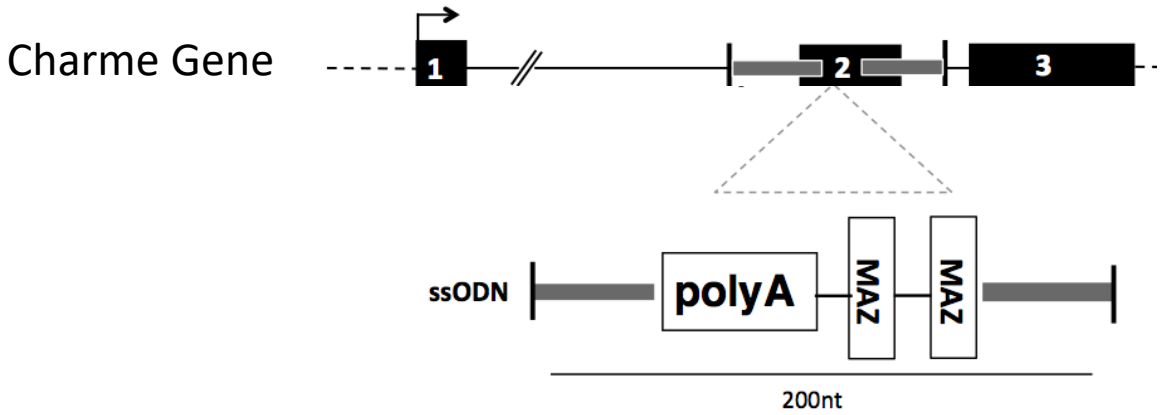
Charme RNA

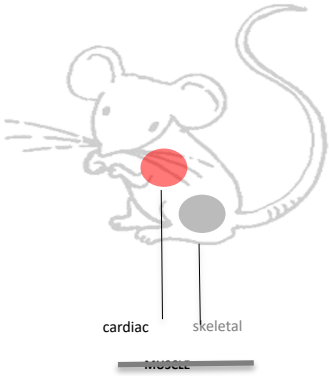


Charme +/+

Charme -/-

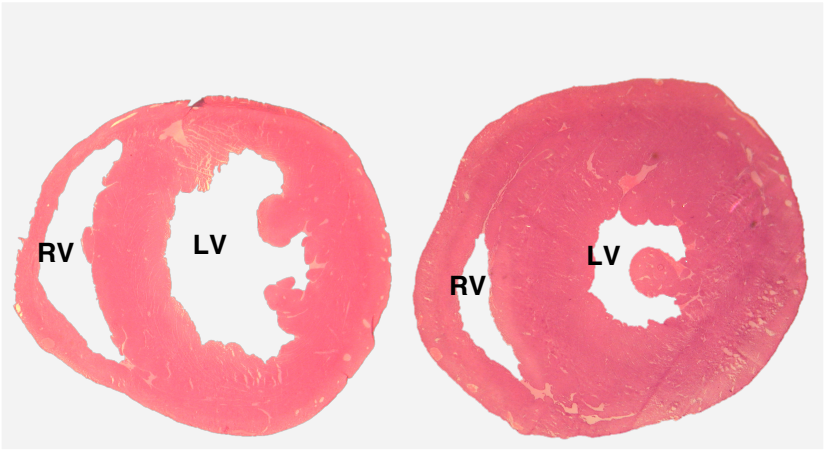
Production of *Charme*$^{-/-}$ mice through CRISPR/Cas9 gene editing





1 month-old

+/+ **-/-**

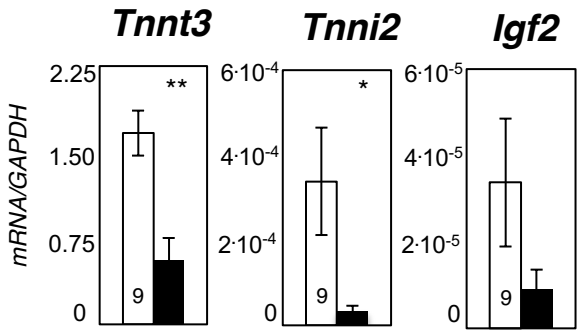


Charme$^{-/-}$ mice showed a **reduced lifespan** that never extended beyond 1 year of age.

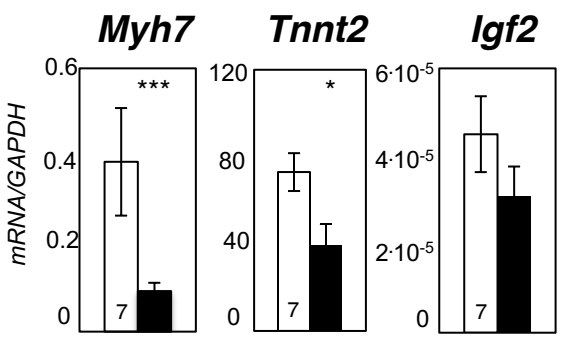
Histological analysis showed a pronounced decrease of ventricle volumes and increase in ventricular septum.

Charme depletion in vivo affects the same muscles genes identified in vitro

Gastrocnemius



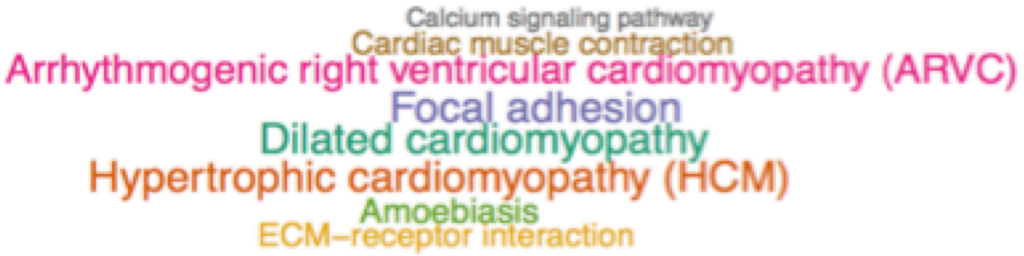
Heart



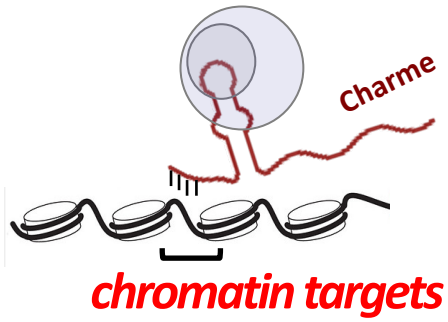
Charme +/+

Charme -/-

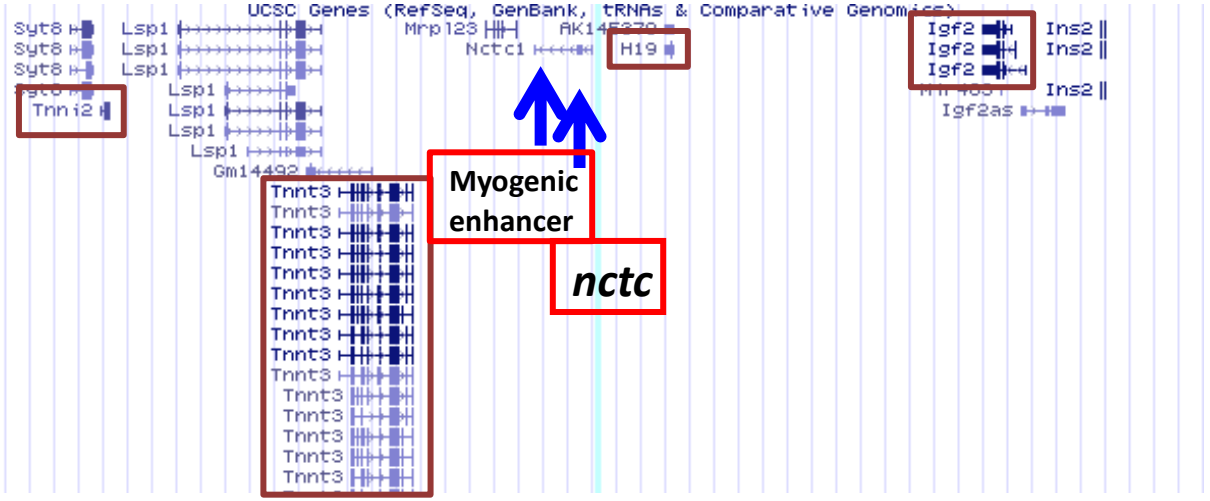
down-regulated genes



ChIRP: one top region contains the genes most affected by *Charme* depletion



Top ranked: the Charme locus

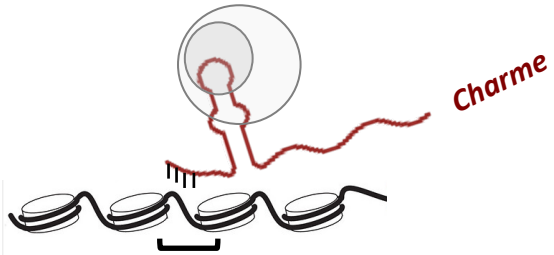


Combining RNA and DNA FISH analysis: *Charme* RNA localizes at the *nctc* locus

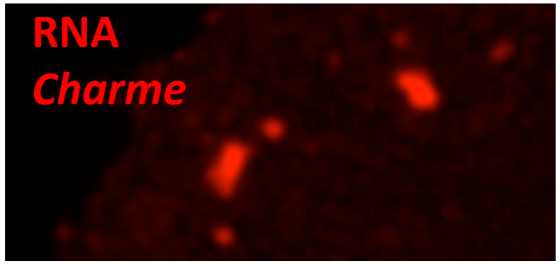
***Charme* RNA**

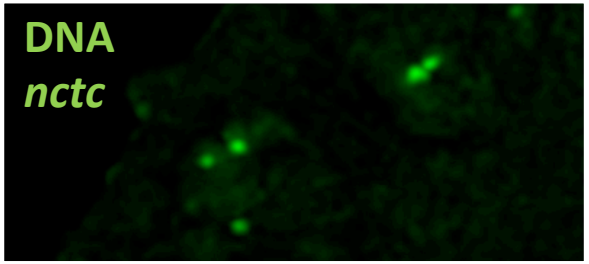
***nctc* DNA**

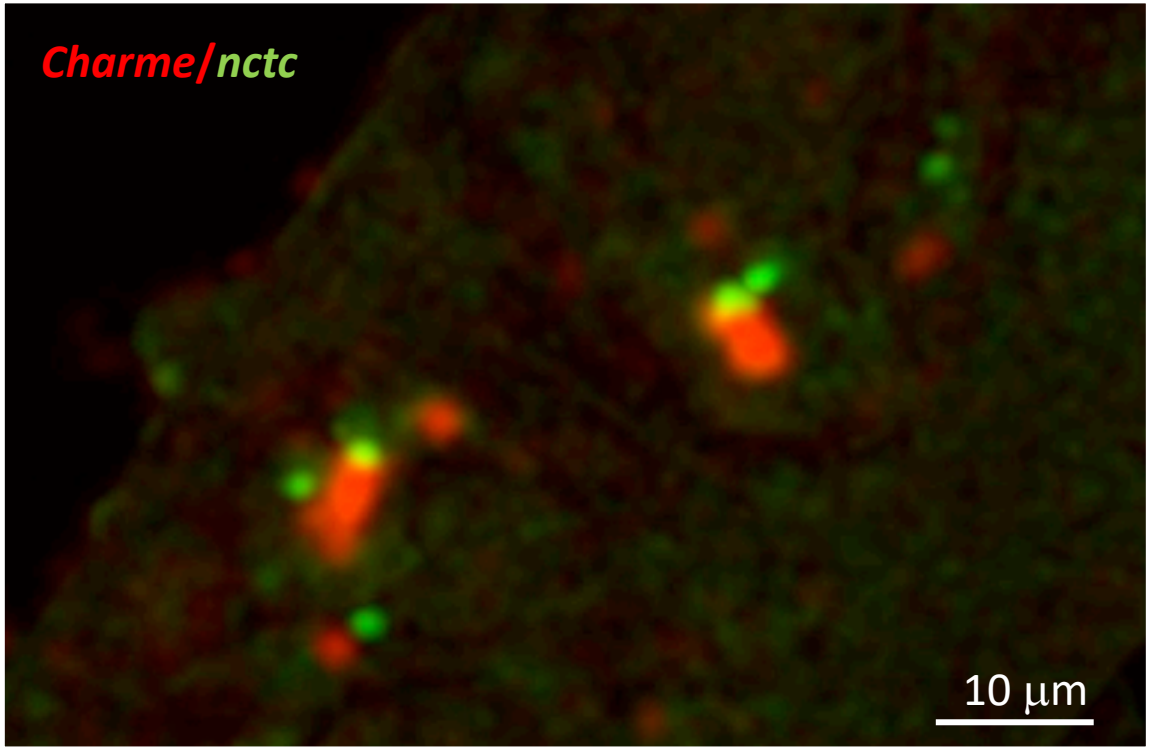
(TNNT3, TNNi2, Igf2)



***nctc* locus**

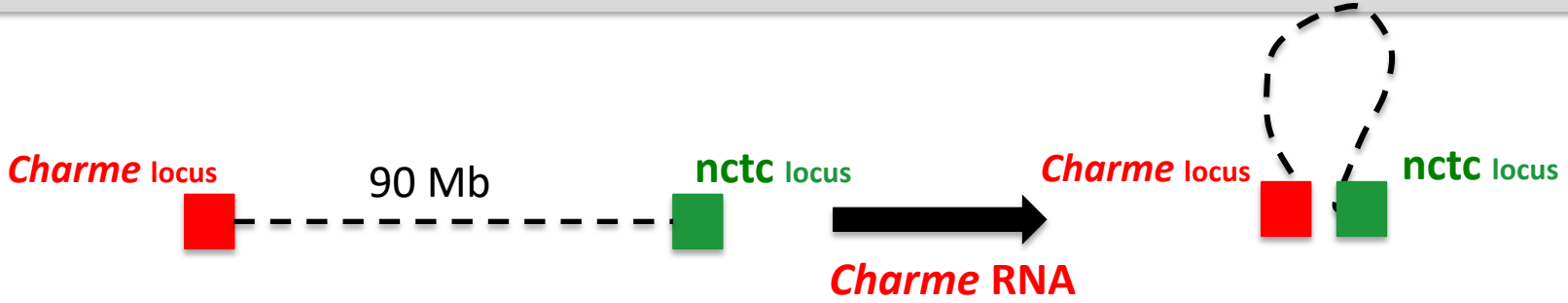


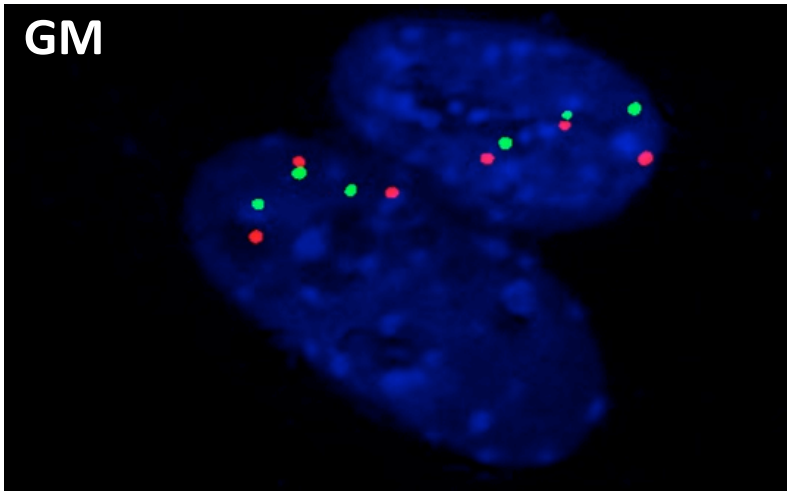


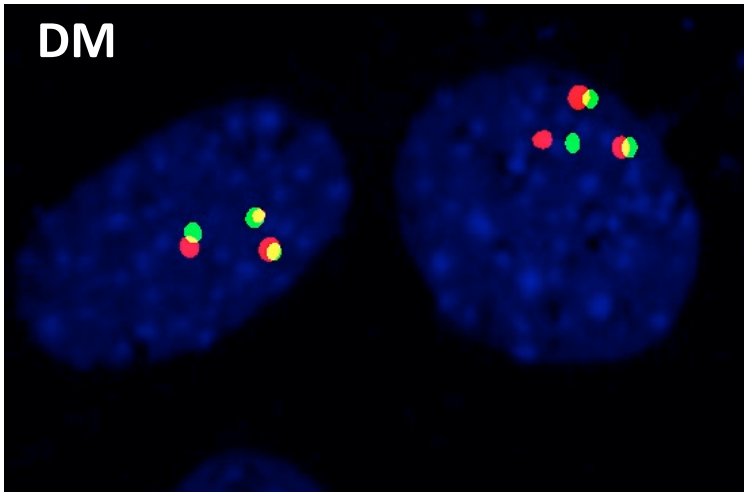


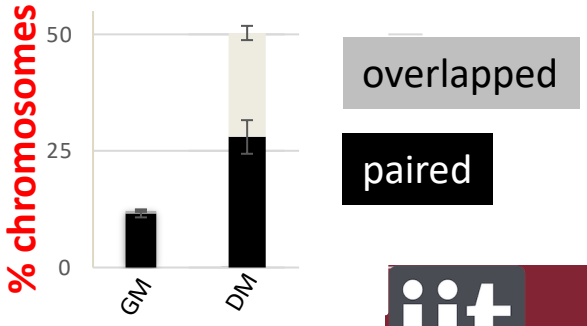
DM conditions

Double DNA FISH analysis: *upon differentiation (**DM**), the Charme coding locus is found in close proximity to the nctc region (TNNT3)*



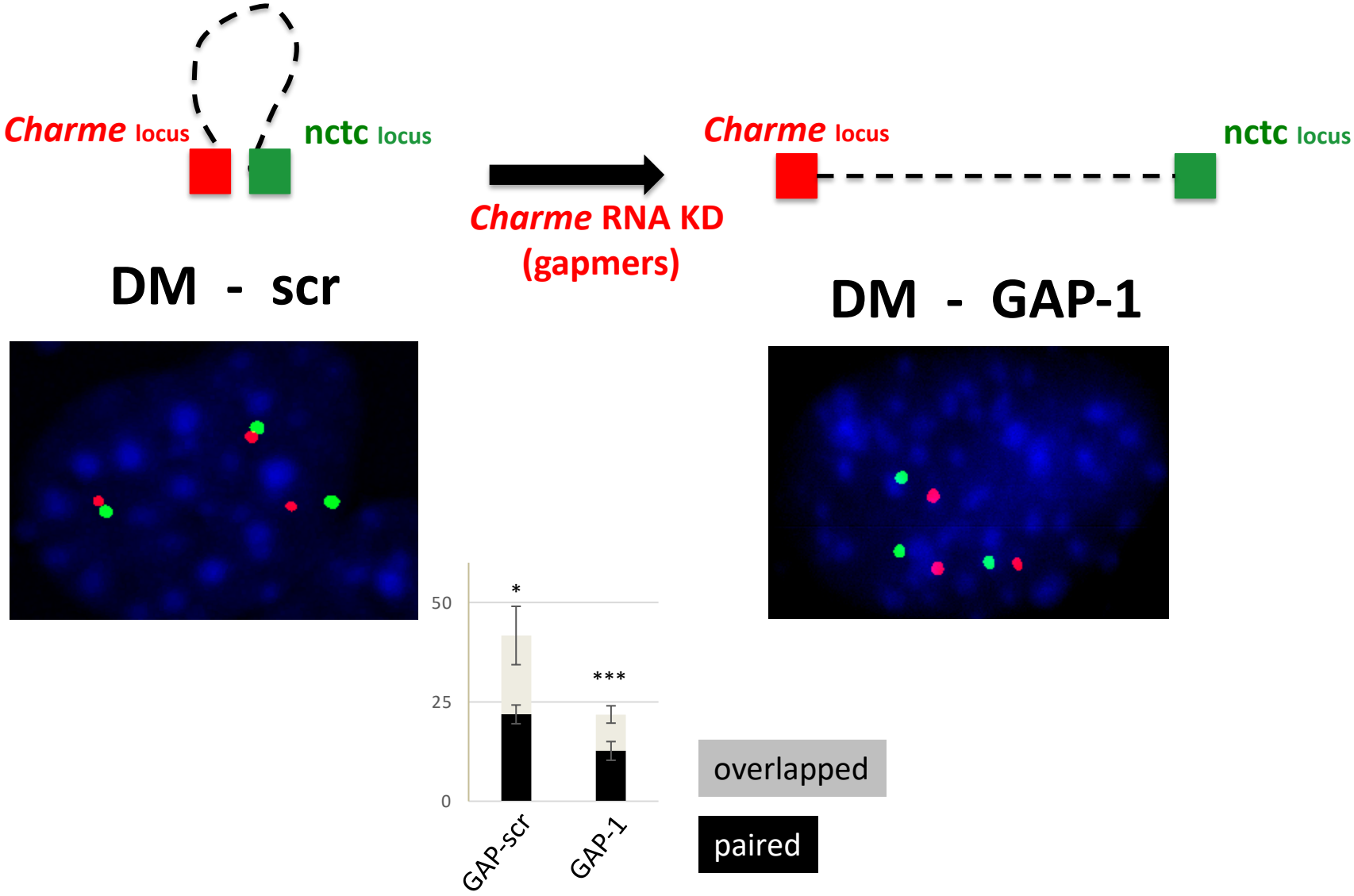


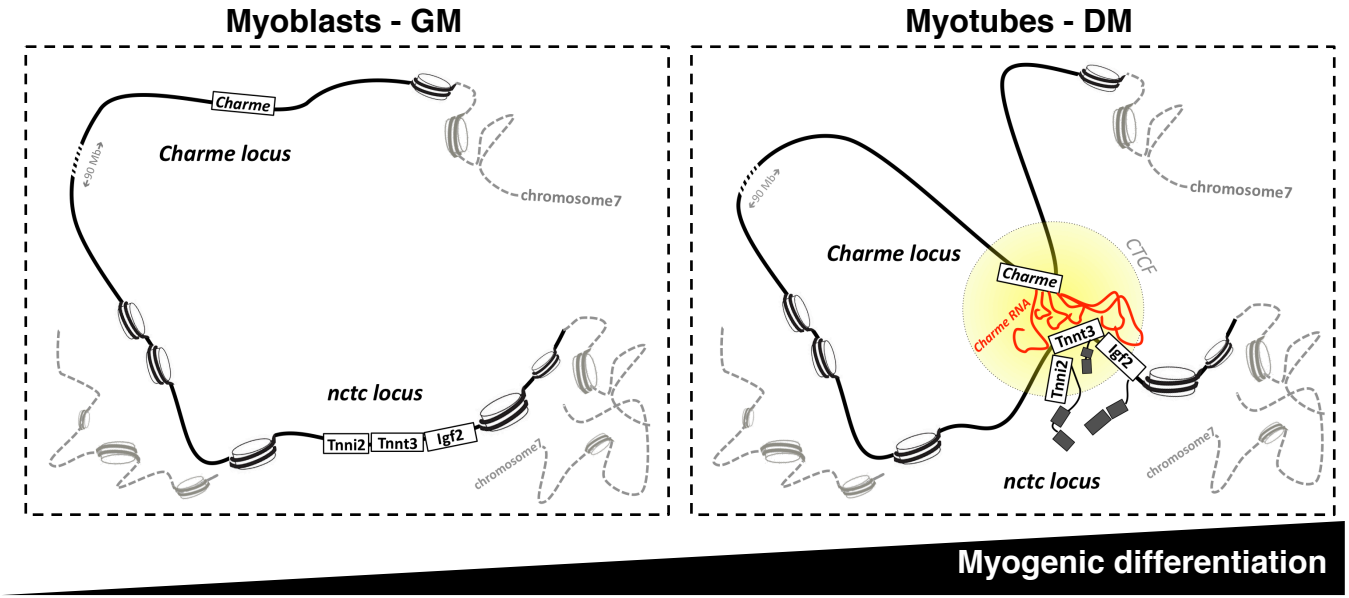


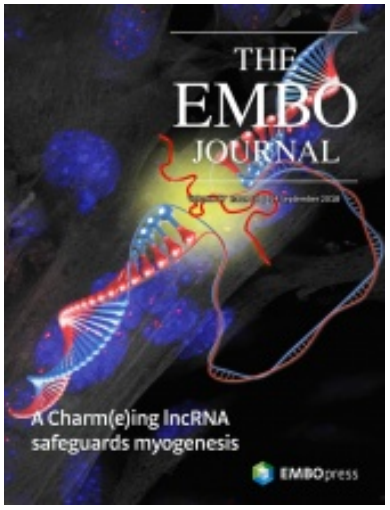


Double DNA FISH analysis:

Charme is required for maintaining the long-range interaction





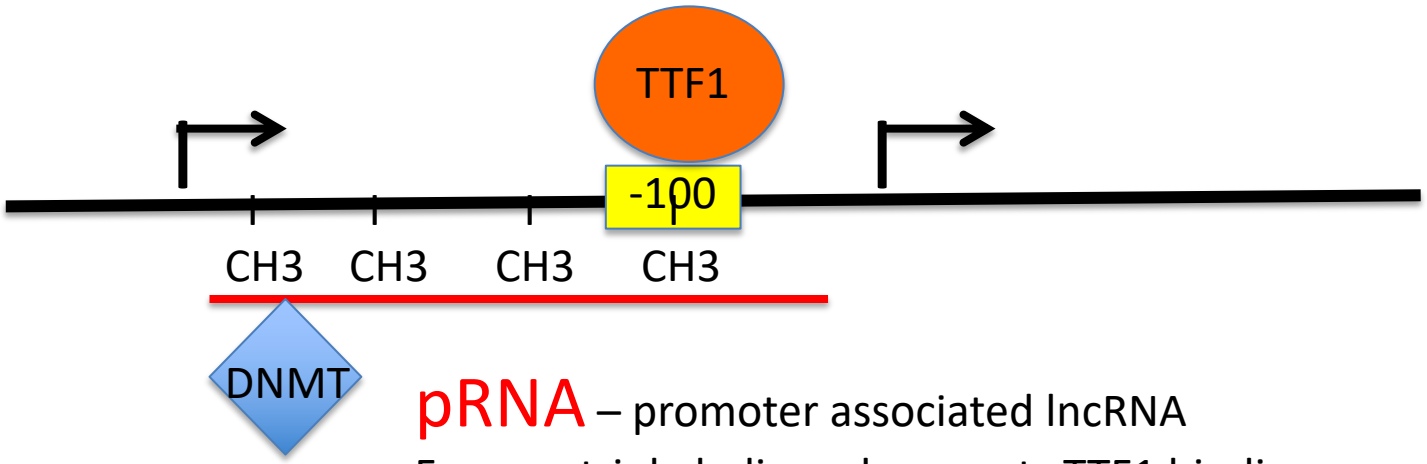


Ballarino M., Cipriano et al, 2018 EMBO J.

Ingrid Grummt – Heidelberg

Genes Dev. 2010 Oct 15;24(20):2264-9.

Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.



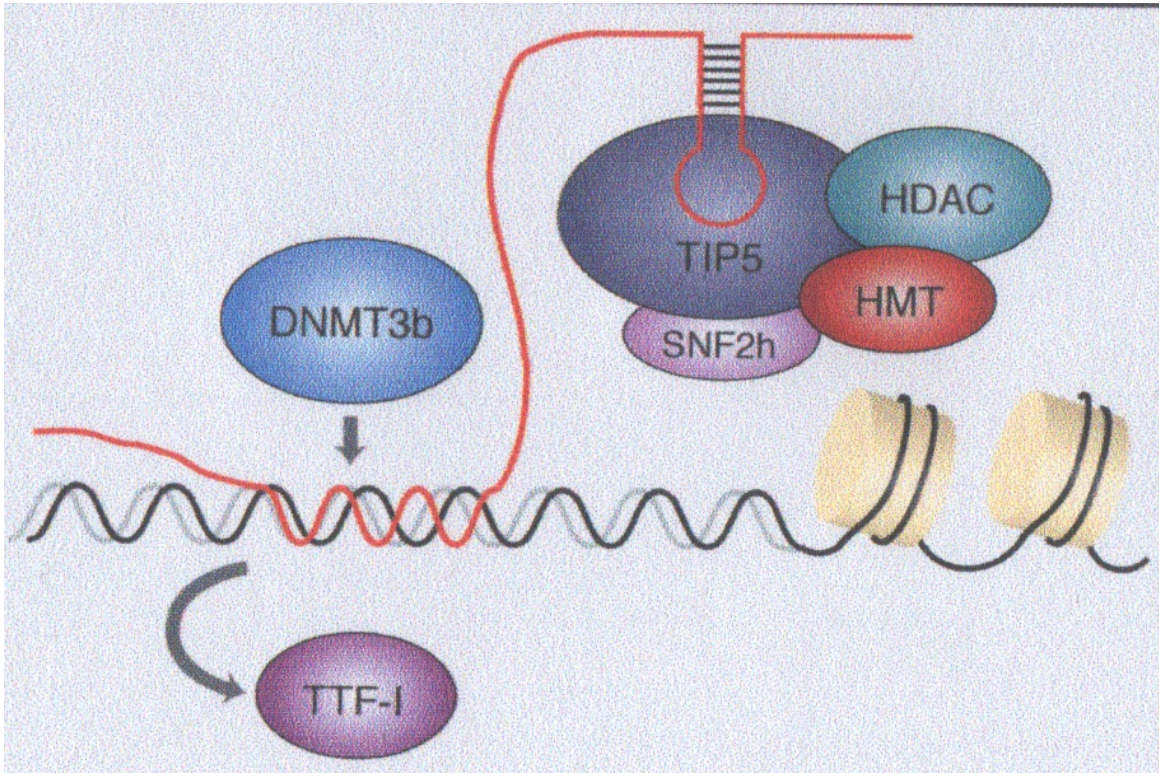
pRNA – promoter associated lncRNA

Forms a triple helix and prevents TTF1 binding while recruiting DNMT3b

Ingrid Grummt – Heidelberg

Genes Dev. 2010 Oct 15;24(20):2264-9.

Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.



Model illustrating the role of pRNA in recruiting chromatin modifying enzymes to rDNA. Transcripts that match the rDNA promoter, dubbed pRNA (promoter-associated RNA), form a specific secondary structure that is recognized by TiP5, the large subunit of the chromatin remodeling complex NoRC. NoRC is associated with histone deacetylases (HDACs) and histone methyltransferases (HMTs) that establish hetero-chromatic features at the rDNA promoter transcriptional silencing.

In addition, pRNA directly interacts with DNA, forming a DNA:DNA:RNA triple helix with the bind ing site of the transcription factor TTF-i, leading to displacement of TTF-i. The triple helical structure is recognized by the DNA methyltransferase DNMT3b, which methylates the rDNA promoter, leading to transcritpional silencing

Ingrid Grummt – Heidelberg

If pRNA↓ CH3↓
pRNA recruits DNMT3b and prevents TTF1 binding

pRNA forms triple helix with the DNA

in vivo – biotinylated oligo > psoralen > UV
DNMT3b binds triple helix

EMSA with 20 nt pRNA + rDNA
Resistant to RNAse H and A