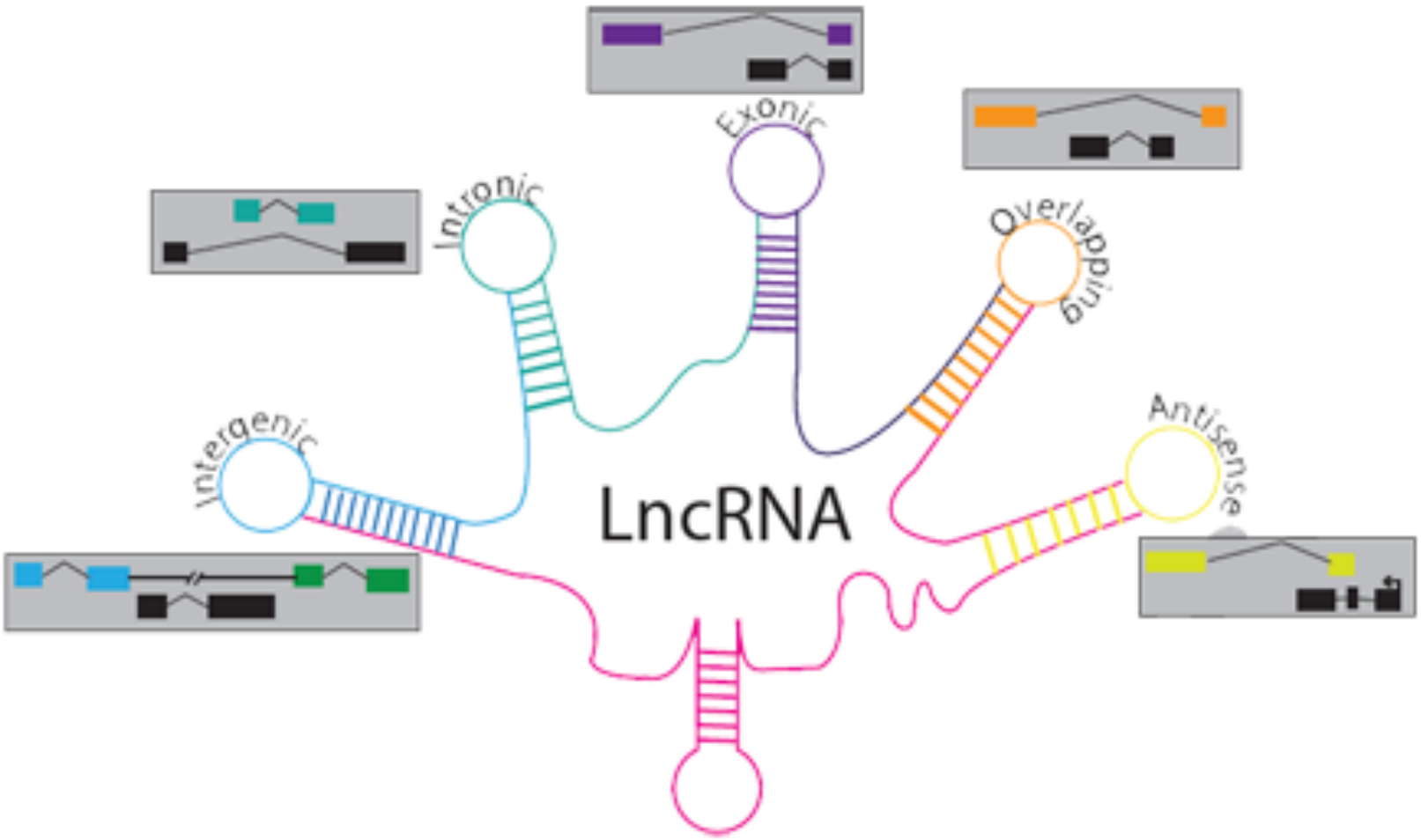
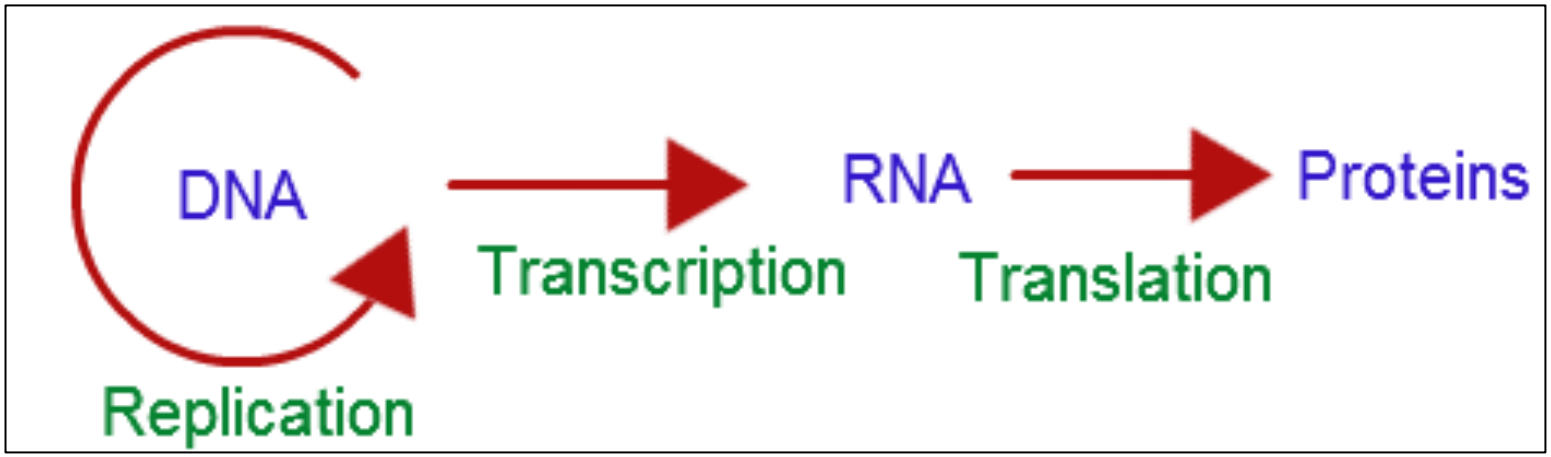


Epigenetic Control of Gene Expression (CEEG)

Ballarino Monica

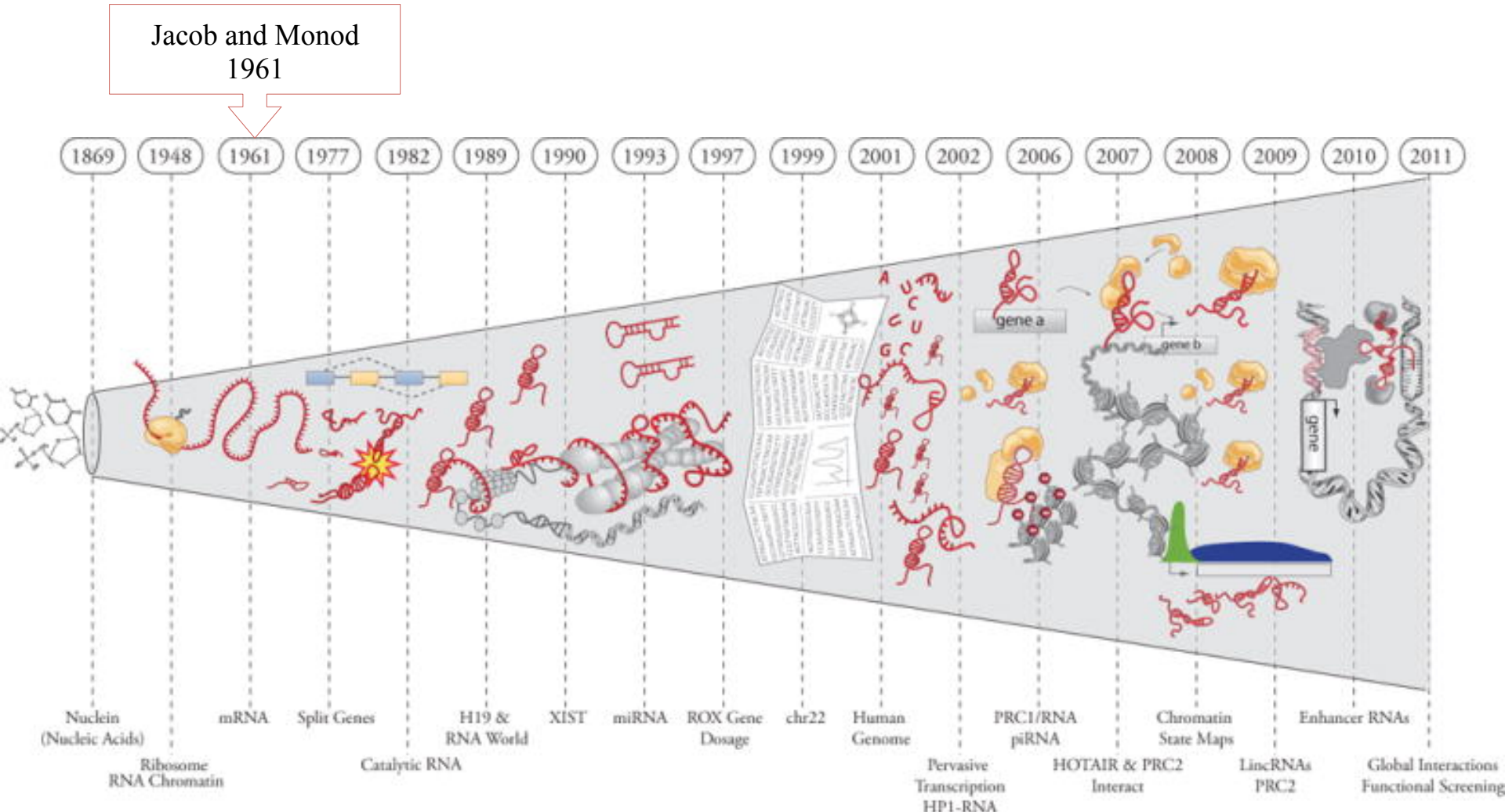
The most fundamental belief in molecular biology is that *genes* are generally *protein-coding*



centrality of RNA in gene regulation

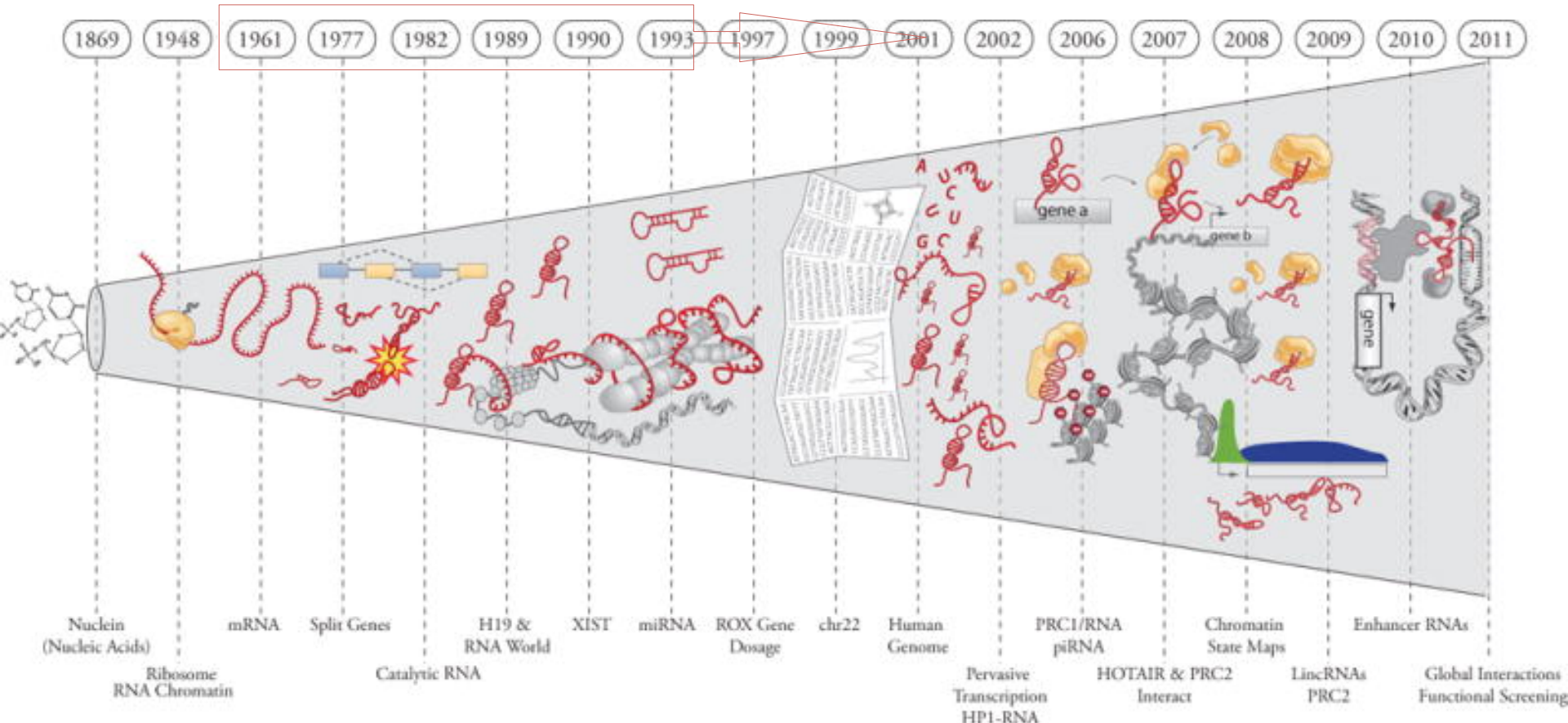
Large-scale sequencing projects have revealed an unexpected ***complexity***: as much as 50% of the transcriptome has *no protein-coding potential* (rather represents an important class of *regulatory* molecules responsible for the fine-tuning of gene expression).

The *centrality* of RNA in the flow of genetic information came to light in Jacob and Monod's 1961 paper "Genetic Regulatory Mechanisms", establishing the concept of **messenger RNAs**. In the 50 years since this landmark paper, numerous regulatory RNAs of all shapes and sizes have been discovered



Studies over the last several decades have pointed to the presence of diverse repertoire of RNAs that was transcribed but *did not encode* proteins i.e. ribosomal RNA, tRNA, RNase P, few lncRNA genes involved in imprinting and other cellular processes (Xist, H19, AIR) and microRNAs.

*These studies only **scratched** the **surface** of functional RNAs in the cell !!!!*



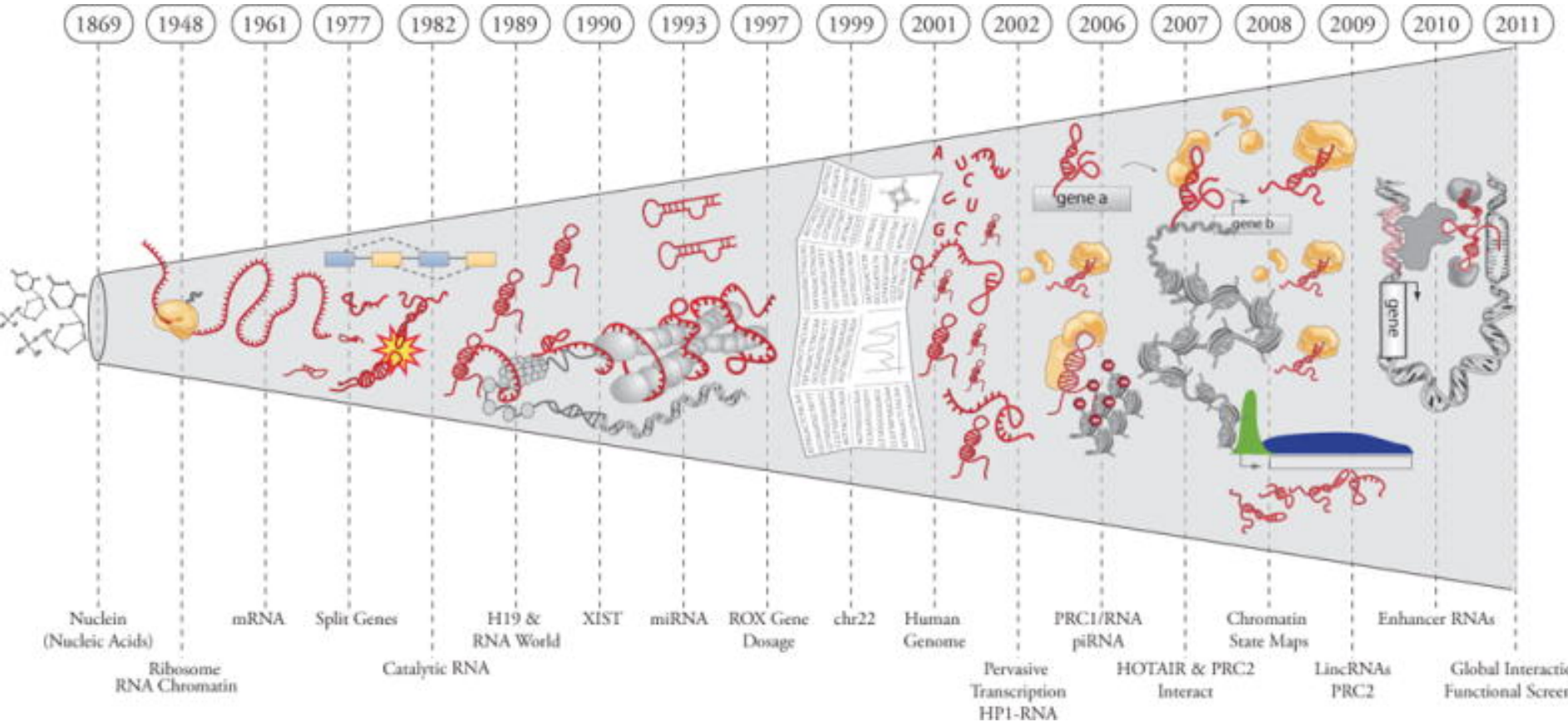
At the turn of the 21st century the scientific community was abuzz with great anticipation of the human genome project.

Burning question:

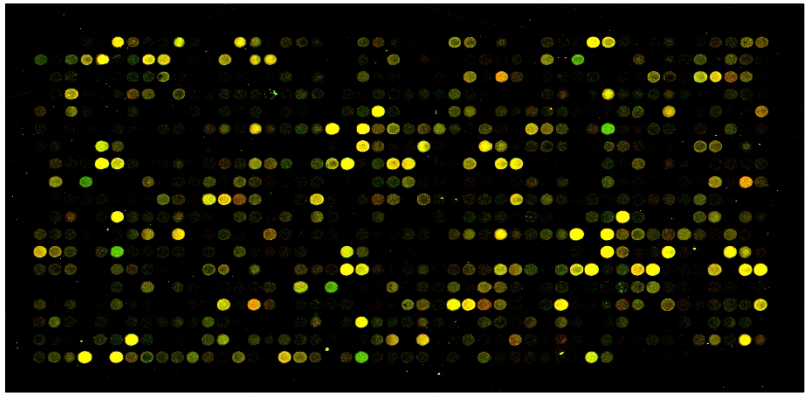
"how many genes are there in the human genome"

Lander *et al.*, 2001
Venter *et al.*, 2001

20.000 transcribed genes (less than 5% of genomic DNA)

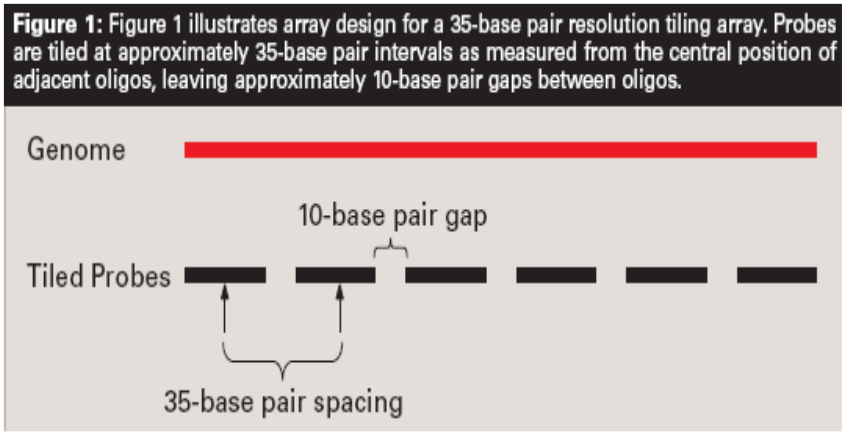


Can the *complexity* of different organisms be explained by the sheer number of classic protein coding genes, their splicing diversity, or perhaps new types of regulation?



cDNA microarrays

Affymetrix Oligonucleotide Arrays



Affymetrix Tiling Arrays

Short Tag High-Throughput Sequencing

NO!

The *biggest* surprise of the genome projects was the discovery that

the number of orthodox (protein-coding) genes *does not scale* with complexity

"You might expect more complex organisms to have progressively larger genomes, but eukaryotic genome size fails to correlate well with apparent complexity. Single-celled amoebae have some of the largest genomes, up to 100-fold larger than the human genome."

C.A. Thomas Jr dubbed it the 'C-value paradox' in 1971.

=

Humans (and other vertebrates) have approximately the *same* number of protein-coding genes (~20,000) as *C. elegans,* and less than those of plants (Arabidopsis ~28,000, rice ~40,000) and protozoa (30,000).

Consortium-wide efforts to define all the transcribed bases in the genome

ENCODE

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

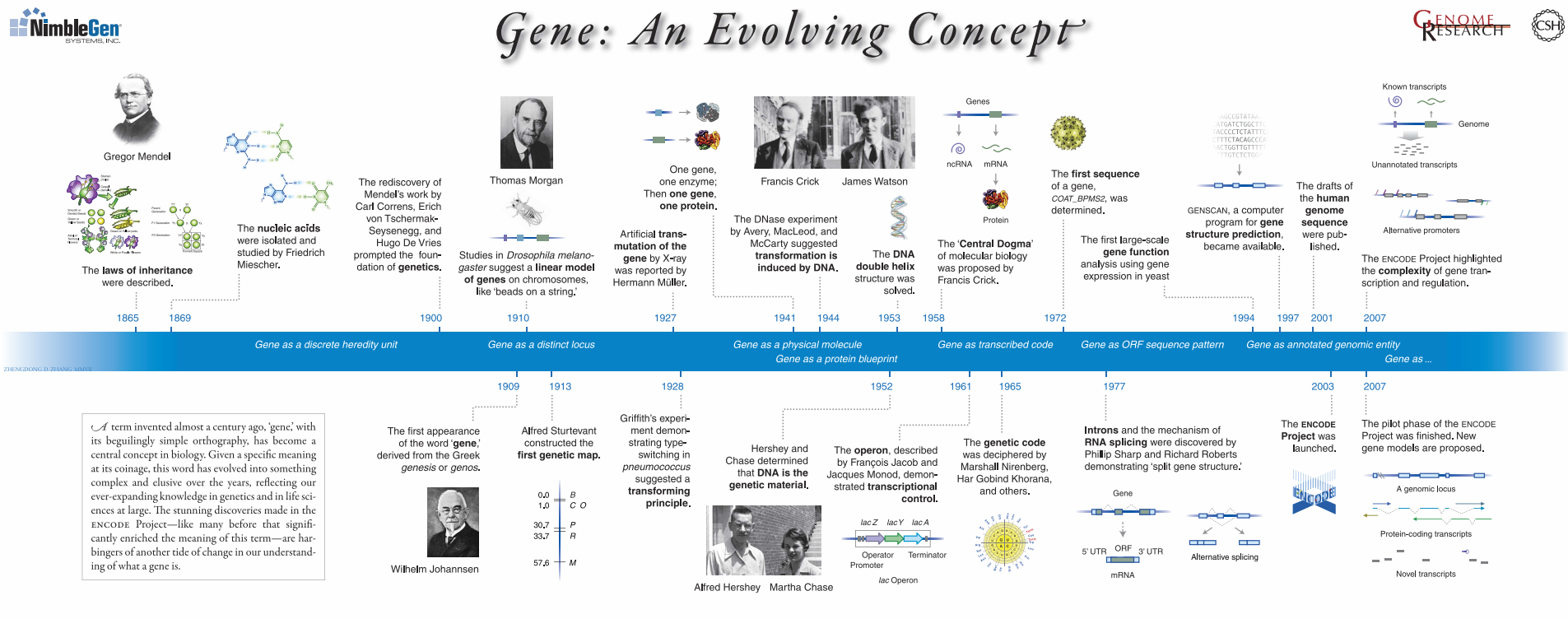
128 mouse primary cell samples

250 different cancer cell lines samples

152 human post-mortem tissues samples

271 mouse developmental tissue samples

>40000 long non-coding RNAs and growing…
80% of the genome is functional

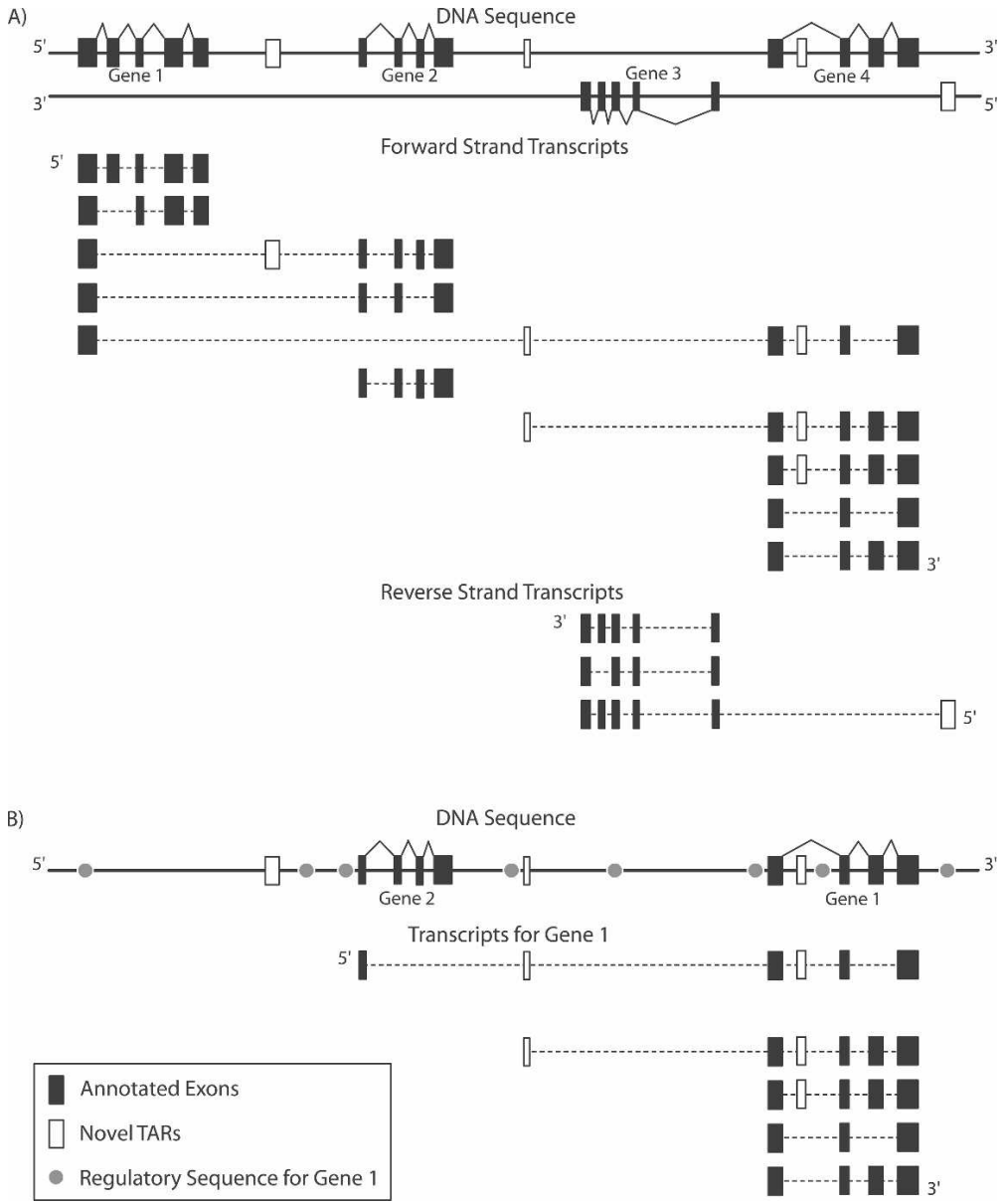


Definition 1860s–1900s: Gene as a discrete unit of heredity
Definition 1910s: Gene as a distinct locus
Definition 1950s: Gene as a physical molecule
Definition 1960s: Gene as transcribed code
Definition 1970s–1980s: Gene as open reading frame (ORF) sequence pattern

What the ENCODE experiments show:

Lattices of long transcripts and dispersed regulation

The ENCODE Project consortium (2007)



1) A vast amount of DNA, not annotated as known genes, is transcribed.

2) A second observation is that there are a large number of unannotated transcription start sites (TSSs) identified by either sequencing of the 5' end of transcribed mRNAs or the mapping of promoter-associated transcription factors via ChIP–chip or ChIP–PET.

3) Thus, some alternative isoforms are transcripts that span multiple gene loci.

4) Many regulatory elements actually reside within the first exon, introns, or the entire body of a gene

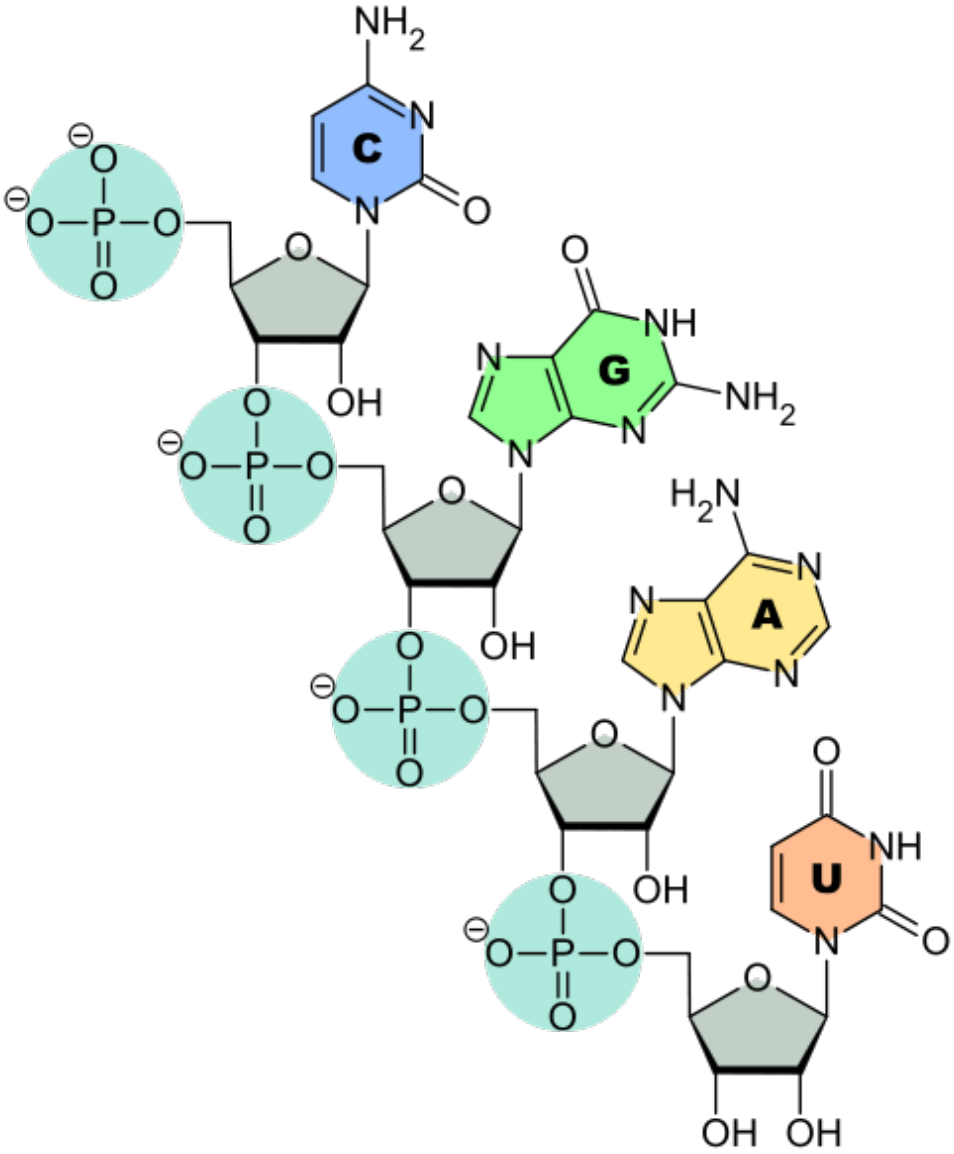
What is a gene, post-ENCODE?

According to traditional definitions genes are unitary regions of DNA sequence separated from each other.

ENCODE reveals that if one attempts to define a gene on the basis of shared overlapping transcripts, then many annotated distinct gene loci coalesce into bigger genomic regions.

Less of a distinction to be made between *genic* and *intergenic* regions.

Where is the information that programs our complexity?



L'evoluzione di organismi complessi è a carico di meccanismi *REGOLATIVI* operati da *sistemi basati sull'RNA*

If RNA-mediated regulation is *real*, then why has this system gone unrecognized for so long?

- **Intellectually *unprepared*** …because of the general assumption that regulatory information is transacted primarily by protein)

- **Biochemically invisible:** RNA is labile, and many, if not most, of the RNA signals are ephemeral and small

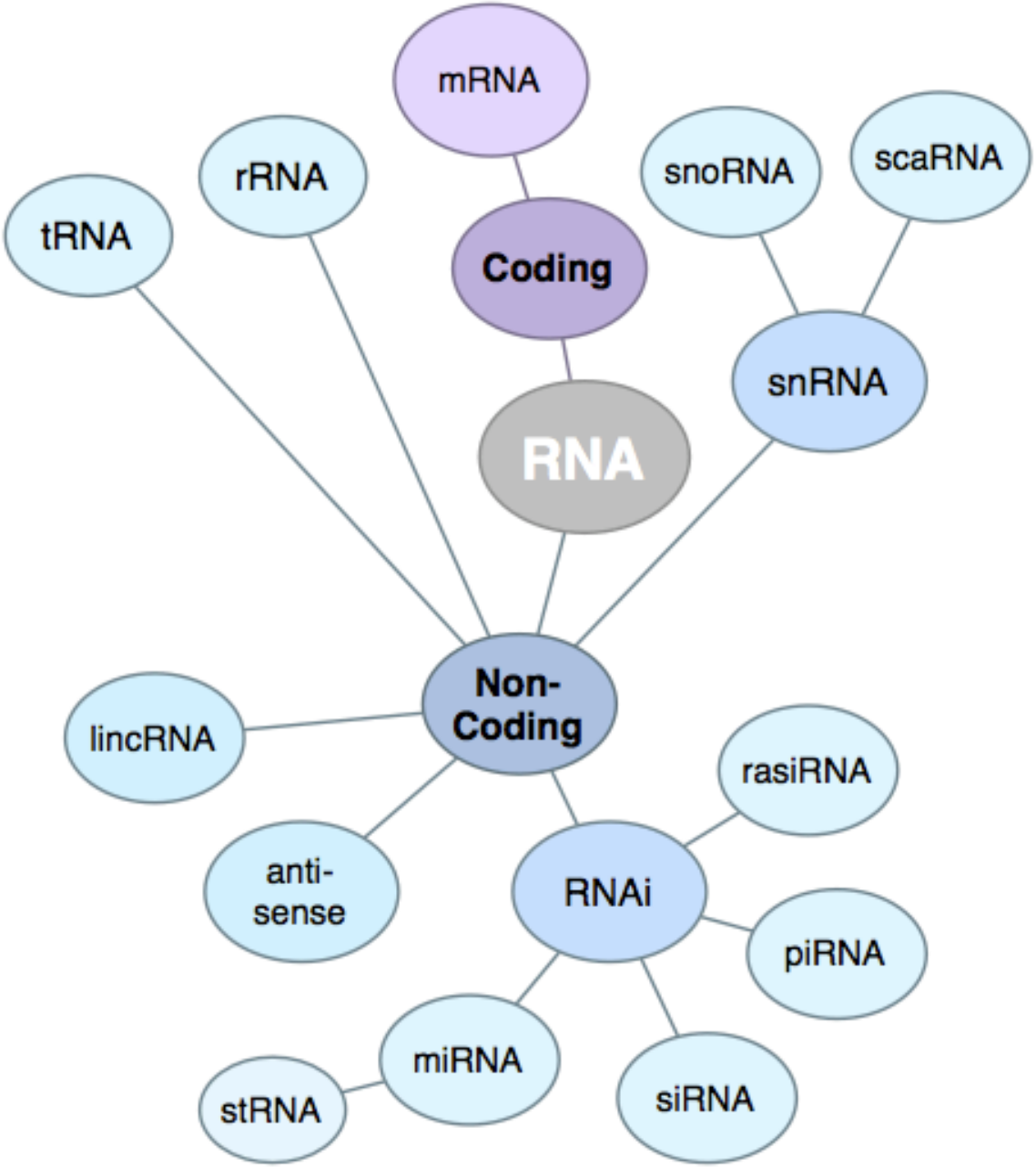
- **Genetically subtle:** mutations that mainly involve single base changes, which in protein-coding sequences can be catastrophic, but in regulatory sequences might have much more subtle consequences.

RNA world

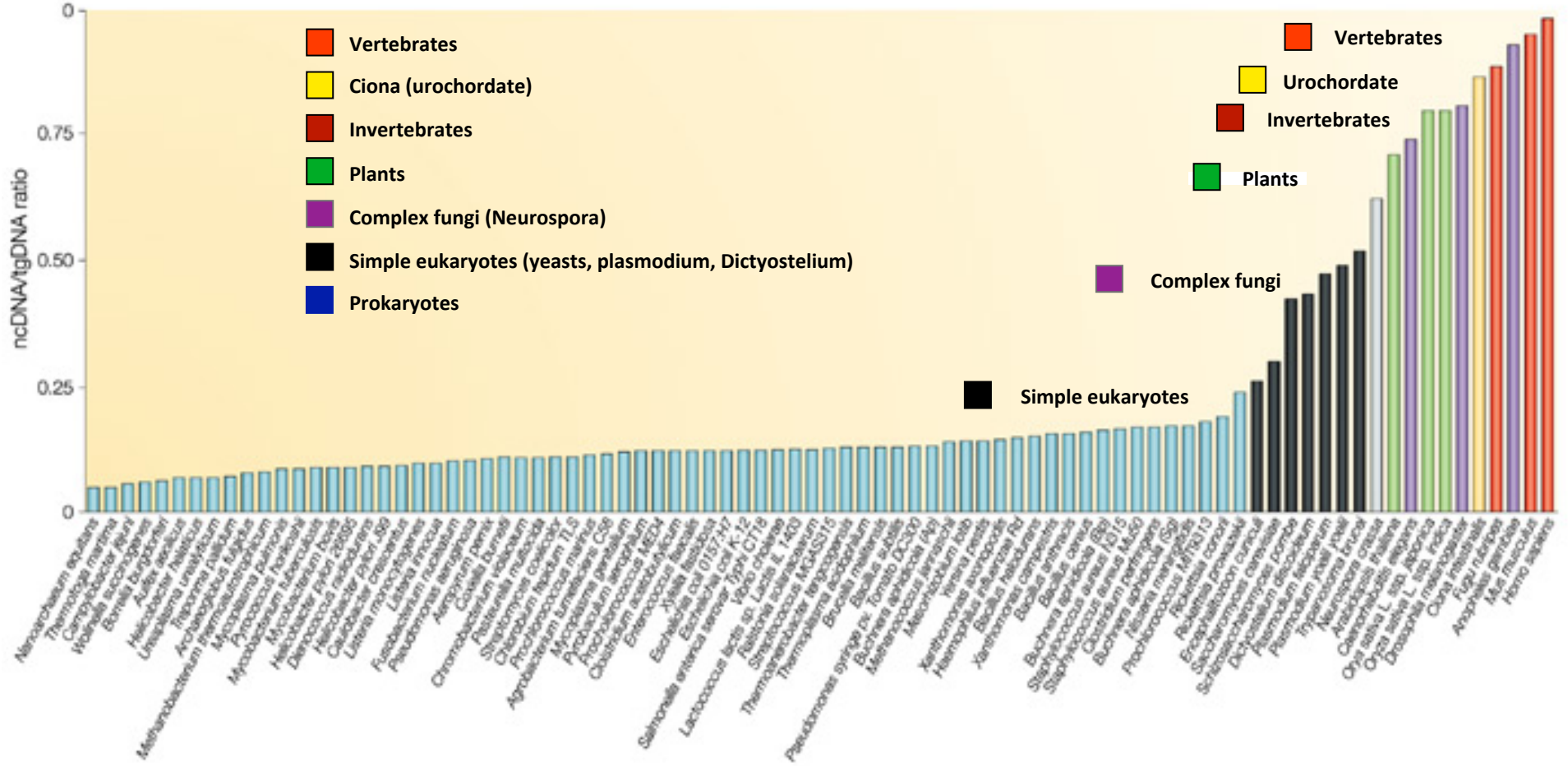
Le cellule contengono una grande varietà di ***RNA non-codificanti***, incluse le componenti *strutturali* del macchinario di espressione genica, come tRNA e rRNA e RNA *regolativi* che influenzano l'espressione di altri geni.

Gli RNA *non codificanti* sono:

- di *differente* natura
- una frazione significativa dei geni di tutti gli organismi *non* codifica per proteine



The proportion of non-coding DNA explains developmental complexity



J.S. Mattick Nature Reviews Genetics 5, 316-323 (2004)

Nature Reviews | Genetics

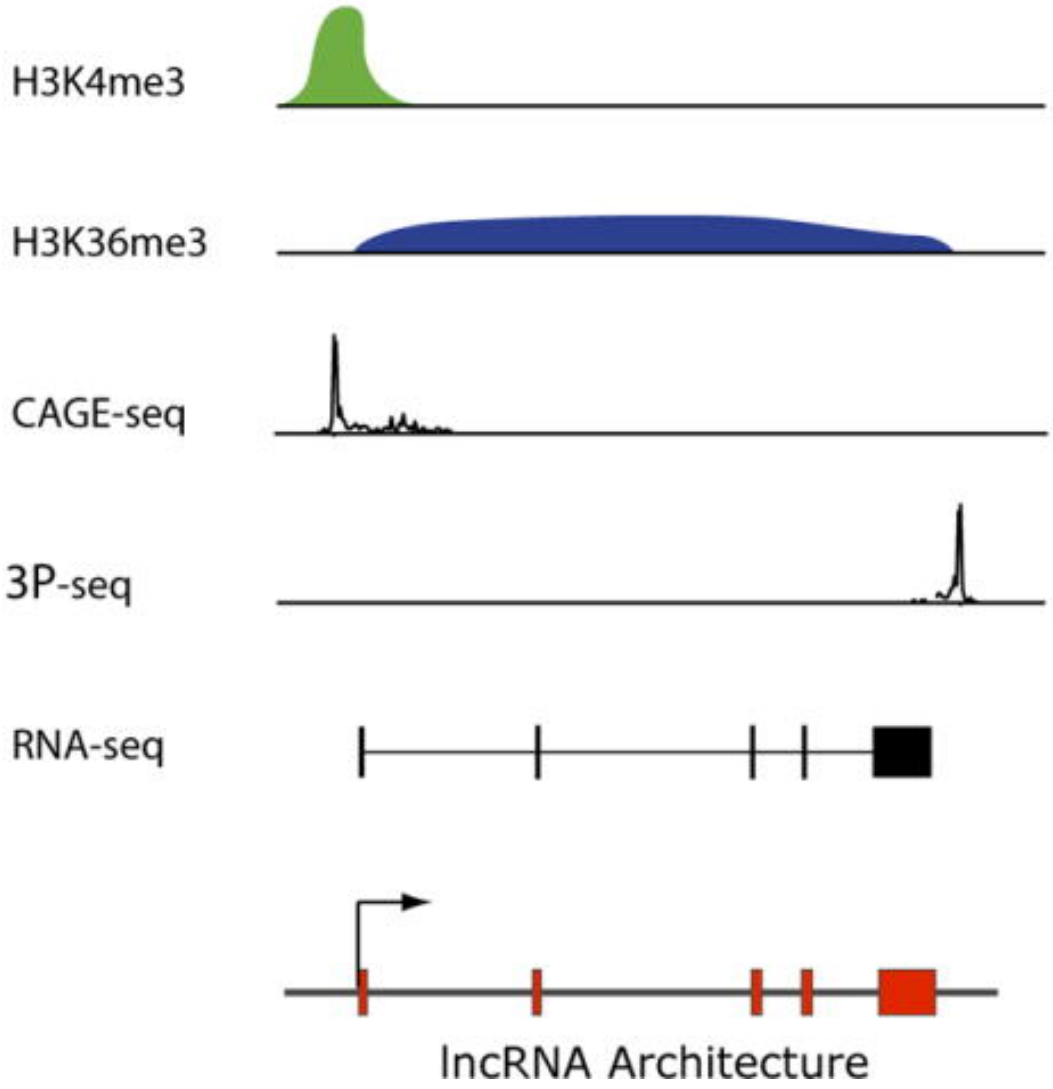
The Transcriptional Landscape of the Mammalian Genome

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

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

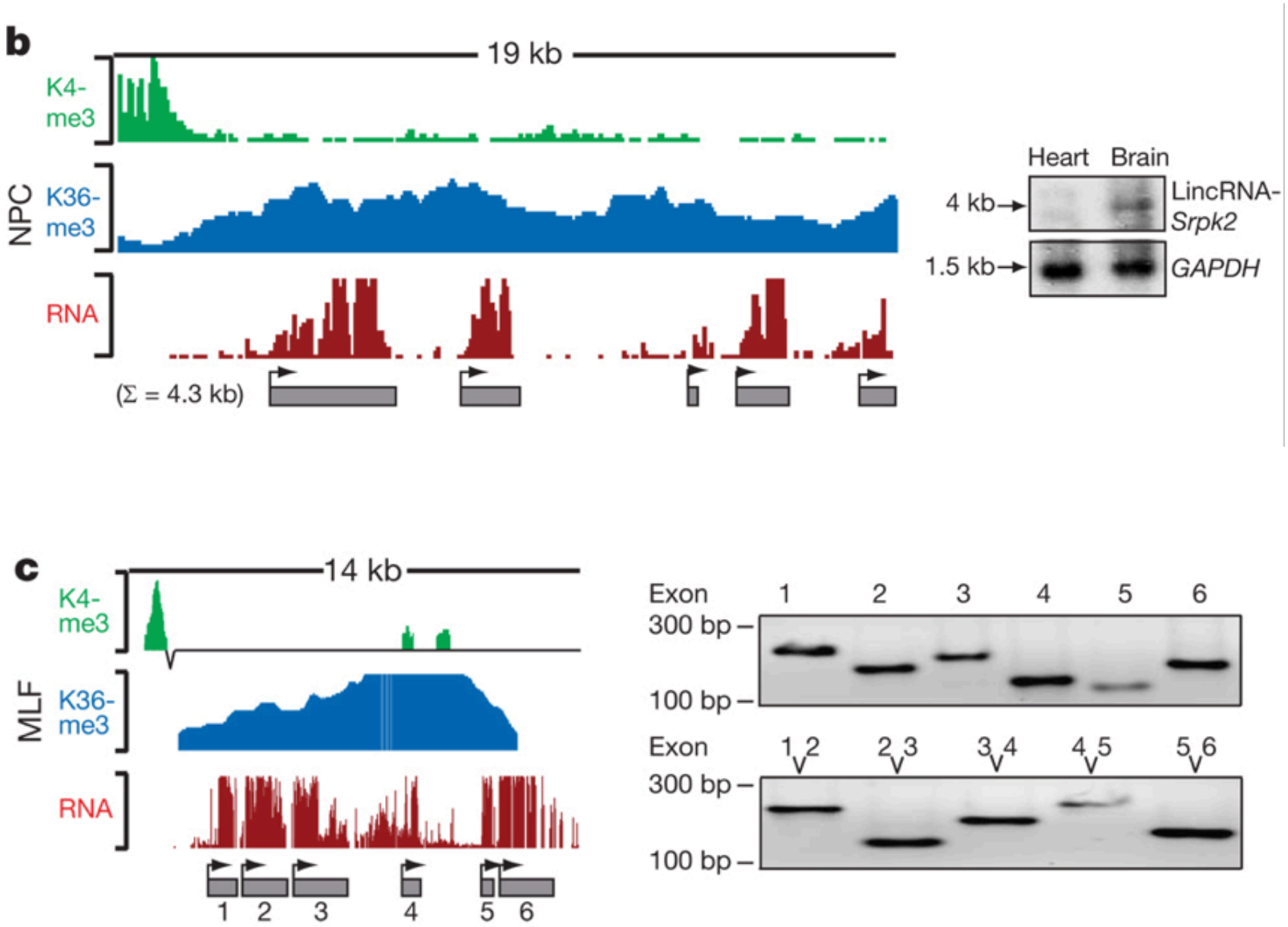
Basal transcriptional ***noise*** or ***functional*** non-coding RNA?

A critical clue for hunting RNA genes came from chromatin



Chromatin marks of transcription initiation (**H3K4me3**) and elongation (**H3K36me3**) define whole transcribed regions of the genome, while sequencing of capped RNA fragments (**CAGE-tag**) or poly-adenylation ends (**3P-seq**) defined the precise beginning and ends of transcripts.

Intergenic K36-K4 domains produce multiexonic RNAs



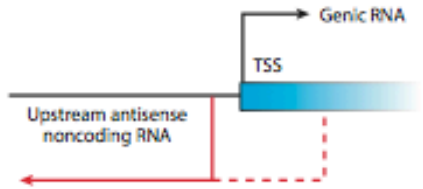
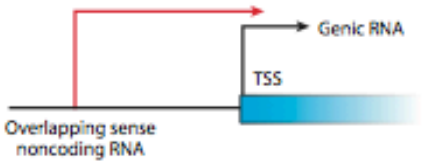
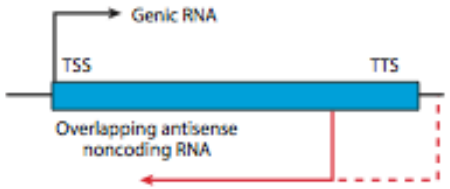
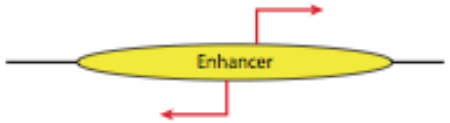
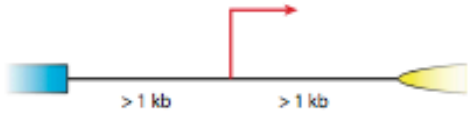
lncRNAs are functional transcripts

- They are *multi-exonic*, *capped*, *spliced* and *polyadenylated* transcripts.

- They can be localised to the *nucleus*, *cytoplasm* (or both).

- Many of them show *tissue specific expression* patterns suggesting potential roles in cell identity

lncRNAs can be defined based on anatomical properties of their gene loci

LncRNA class	Names	Examples	References
Promoter-associated 	pRNAs (uaRNAs, PROMPTs)	Most genes	Sigova et al. 2013 van Bakel et al. 2010
Gene body-associated (sense) 	gsRNAs	*DHFR* locus ecCEBPA *fbp1*$^{+}$ locus	Martianov et al. 2007 Di Ruscio et al. 2013 Hirota et al. 2008
Gene body-associated (antisense) 	gaRNAs (NATs)	Airn Kcnq1ot1 *BDNF* locus	Sleutels et al. 2002 Lee et al. 1999 Modarresi et al. 2012
Enhancer-associated 	eRNAs (ncRNA-a)	All enhancers	Kim et al. 2010 De Santa et al. 2010 Ørom et al. 2010
Intervening 	lRNAs (lincRNAs)	HOTAIR MALAT1	Rinn et al. 2007 Guttman et al. 2009 Cabili et al. 2011

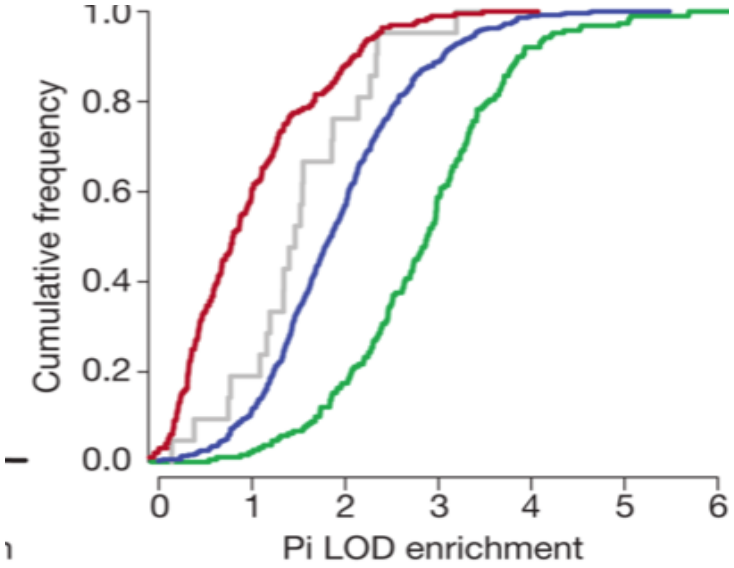
overlapping transcripts

intergenic transcripts

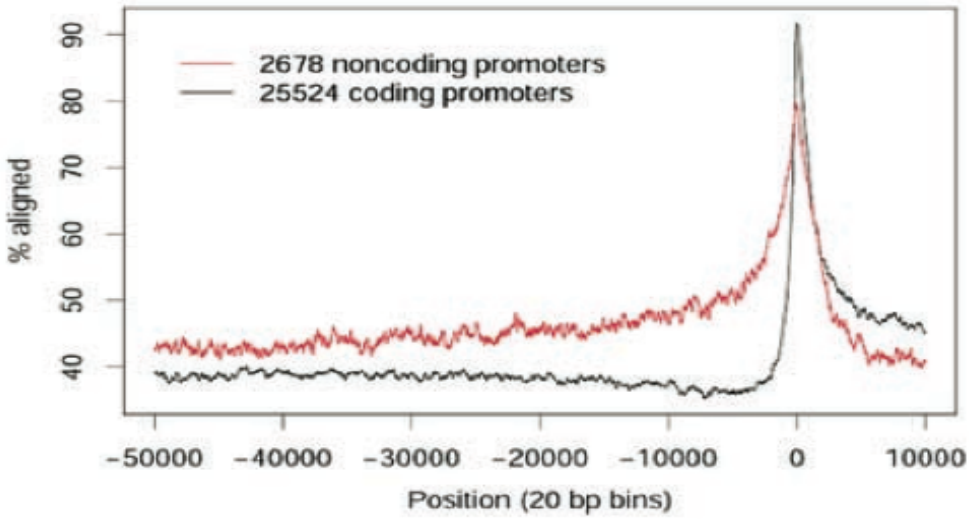
These loci have their own promoter and are marked by the same chromatin modifications found at protein-coding genes.

lncRNAs are functional transcripts

lncRNAs show *low* sequence conservation



lncRNAs show *highly* conserved promoters



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

To suggest functionality….

- Conservation of their promoter, splice junctions, exons, predicted structures, genomic position and expression patterns
- Dynamic expression and alternative splicing during differentiation
- Altered expression or splicing patterns in cancer and other diseases
- Association with particular chromatin signatures that are indicative of actively transcribed genes
- Regulation of their expression by key transcription factors and hormones
- Tissue- and cell-specific expression patterns and subcellular localization

A variety of functions for lncRNAs

NUCLEUS

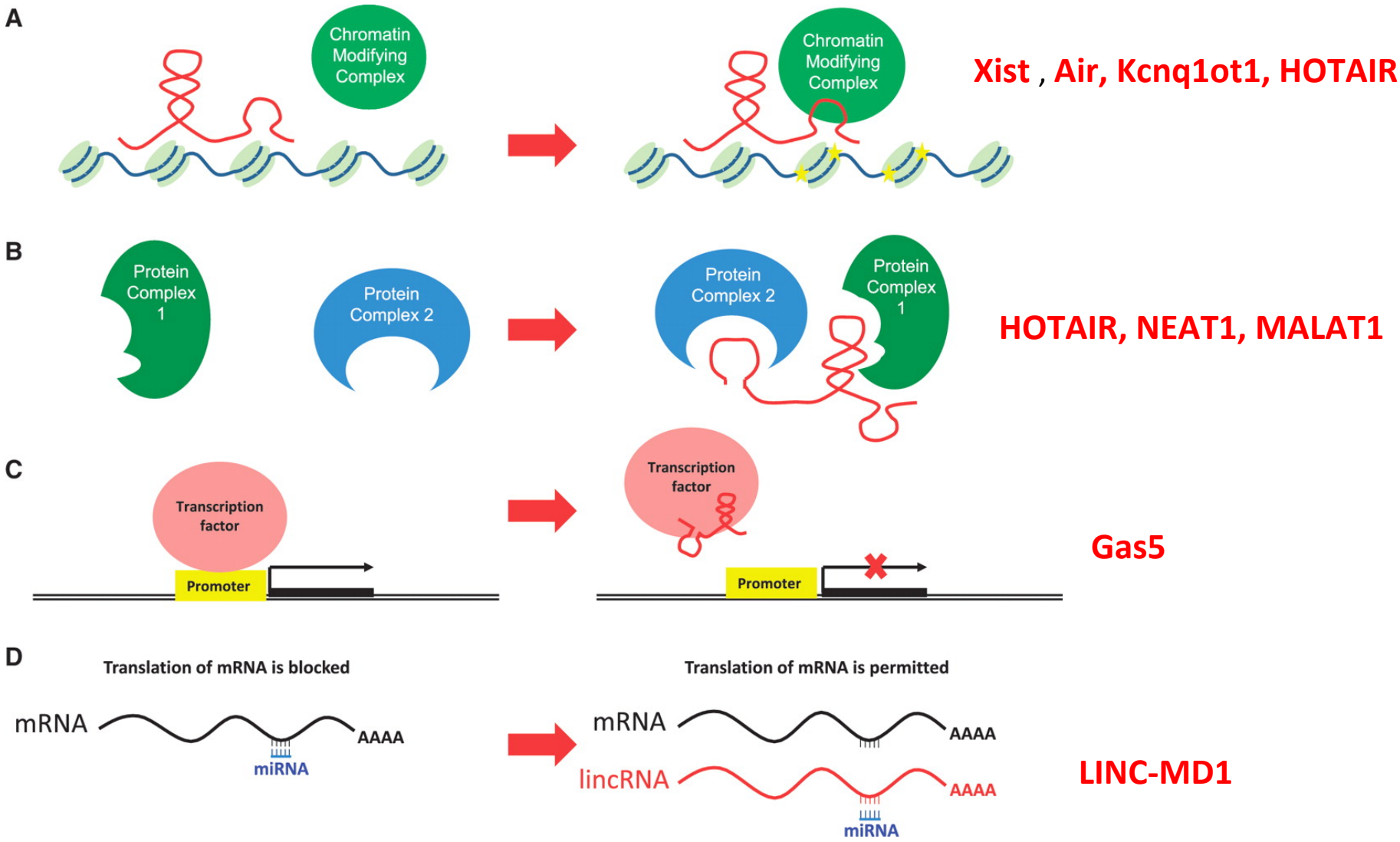
- Chromatin Remodeling → **Xist** – X chromosone inacivation; **Air, Kcnq1ot1, HOTAIR,HOTTIP** – genomic imprinting
- Transcription Regulation → **Gas5** – negative regulator of glucocorticoid receptors
- Nuclear Architecture → **NEAT1** – formation and maintainance of paraspeckles
- Post-transcriptional Regulation → **MALAT1** – phosphorylation of SR proteins

CYTOPLASM

- mRNA decay → **½-sbsRNAs** – STAU1-mediated mRNA decay
- miRNA decoy → **LINC-MD1**– miRNA sponge

etc…

lncRNAs exert their effects by diverse mechanisms



Moran V A et al. Nucl. Acids Res. 2012;40:6391-6400

Nucleic Acids Research

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 stru***cture, which is difficult to predict

LncRNAs controlling the access/dismissal of regulatory proteins to chromatin

TABLE 1

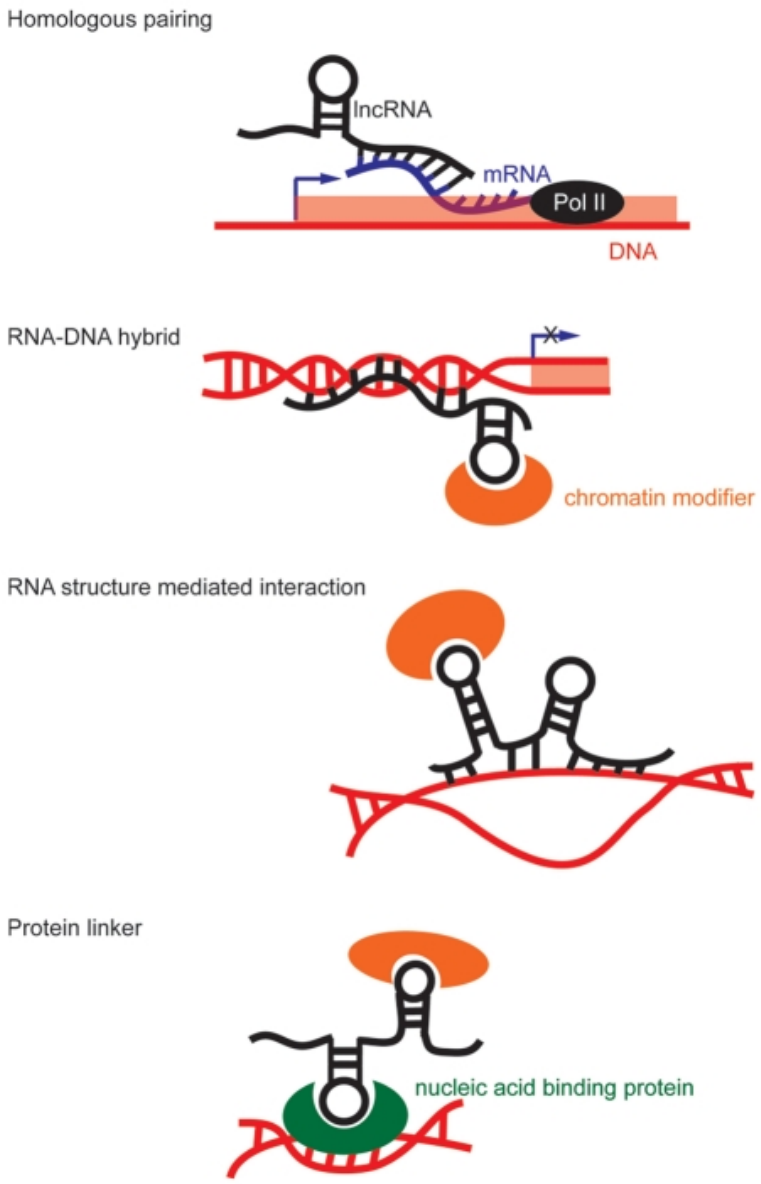
Protein partners of lncRNAs

Many of these lncRNA-protein complexes function at chromatin.

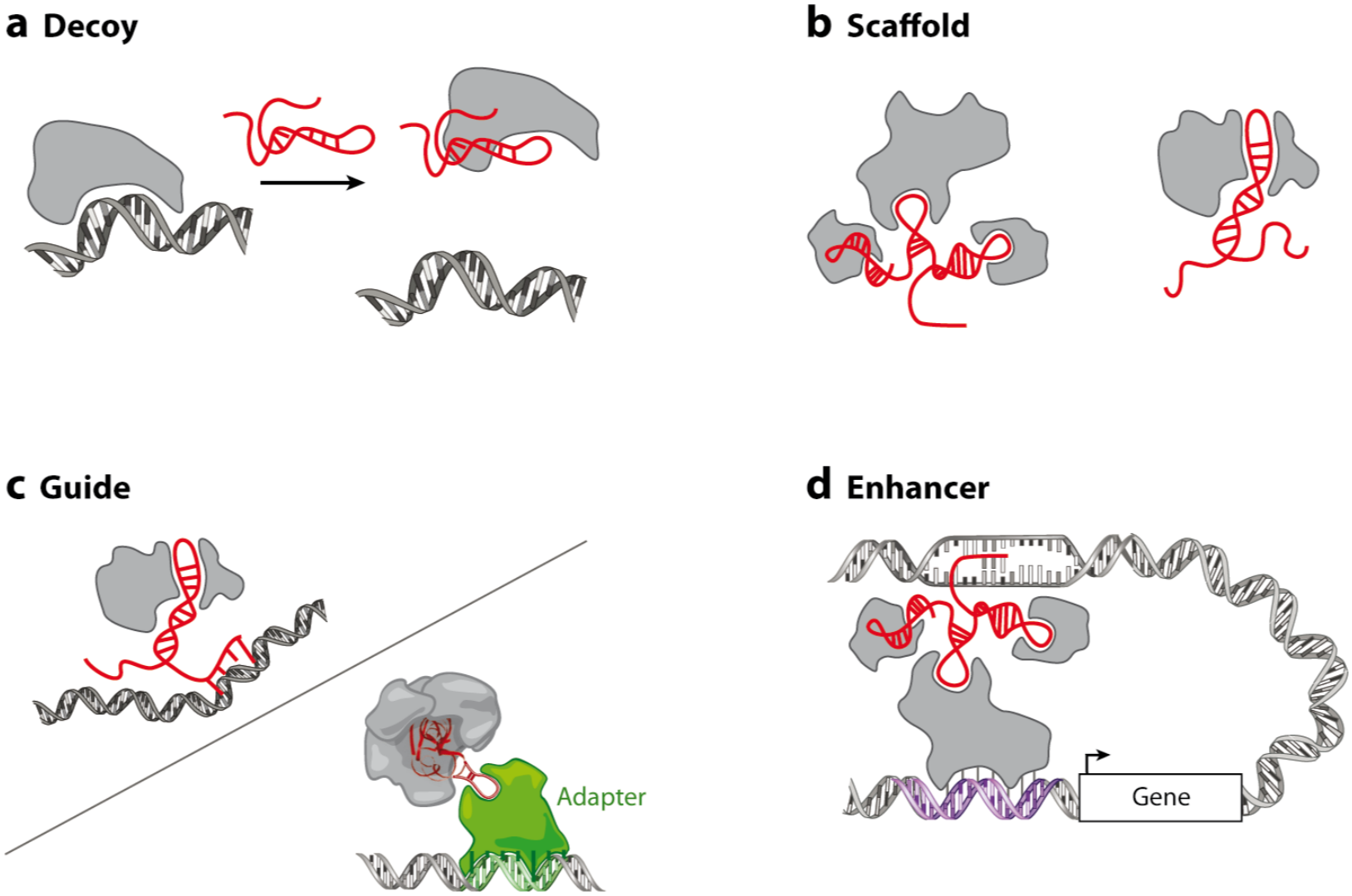
Protein	LncRNA	LncRNA Function	Refs
CTCF	SRA	Enhances insulator function of CTCF.	(92)
DNMT3b	pRNA	Targets DNMT3b in cis to rRNA locus via RNA:DNA:DNA triplex for cytosine methylation and gene silencing.	(90)
G9a	Kcnq1ot1, Air	Targets the H3K9 methylase G9a in cis for imprinting.	(87, 88)
Glucocorticoi d receptor	Gas5	Binds to GR as a decoy and titrates GR away from target genes	(102)
hnRNP-K	Linc-p21	Targets hnRNP-K in trans to mediate p53-dependent gene repression.	(48)
LSD1-coREST	HOTAIR, many others	Targets LSD1 complex to demethylate H3K4me2 to enforce gene silencing.	(49, 84)
MLL-WDR5	HOTTIP, some eRNAs?	Binds to and localizes MLL complex and H3K4me3 via chromosomal looping for gene activation.	(95, 96)
NF-YA	PANDA	p53 inducible and titrates away NF-YA to favor survival over cell death during DNA damage.	(47)
PRC1	ANRIL, Xist	Targets PRC1 in cis for gene silencing. ANRIL influences p16INK4a expression and cell senescence.	(9, 91)
PRC2	Xist, HOTAIR, ANRIL, COLDAIR, Gtl2, Kcnq1ot1, many others	Targets PRC2 either in cis or trans to mediate H3K27 methylation and gene silencing for dosage compensation, imprinting, and developmental gene expression.	(49, 81, 83, 88, 109)
SR splicing factors	MALAT1	Sequesters SR splicing factors to regulate alternative splicing.	(125)
Staufen	½ SBS RNAs	Pairs with mRNAs vis Alu repeast and targets them into non-sense mediated decay pathway.	(126)
Set1 and Hda1/2/3 HDACs	CUTs, XUTs	Antisense RNAs repress sense transcription via control of H3K4me3 and histone deacetylation	(127–130)
TERRA	hnRNP-A	Controls telomerase access to telomeres in cell cycle phase-specific manner.	(131)
TFIIB	DHFR minor	Titrates away TFIIB during cell quiescence to decrease *DHFR* transcription	(99)
TLS	CCND1 promoter ncRNA	Allosterically binds TLS to inhibit CBP and p300 activity, leads to repression of *CCND1* gene.	(132)
YY1	Xist	YY1 binding nucleates Xist on the inactive X chromosome.	(100)

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

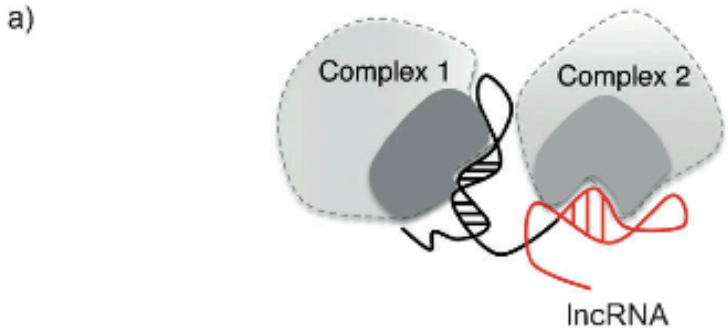


lncRNA mechanisms of action

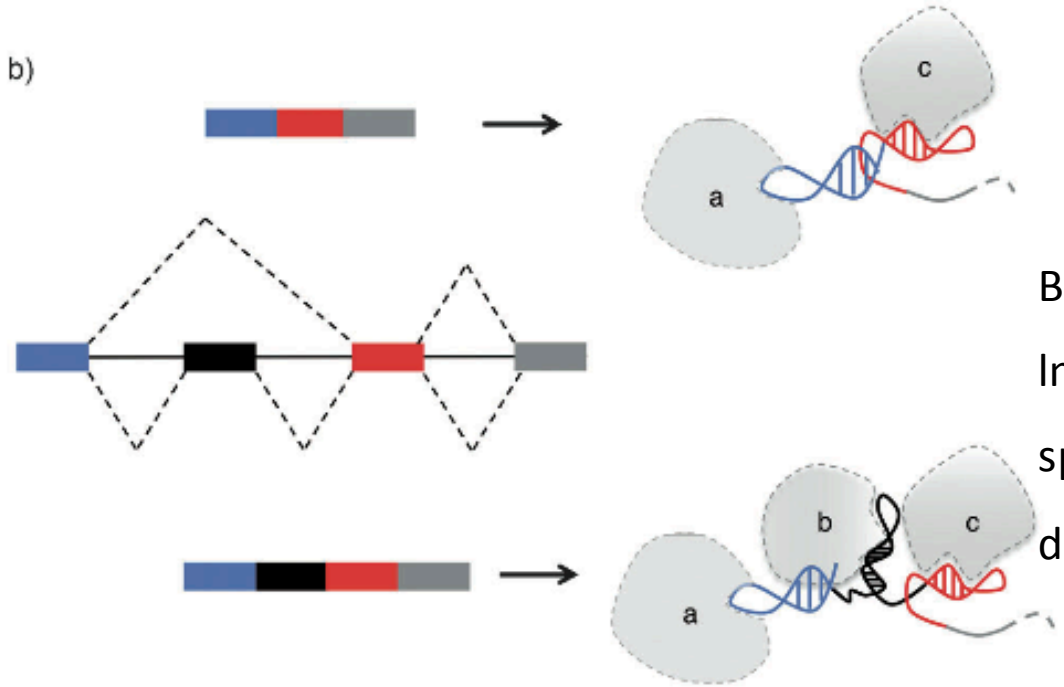


Rinn JL, Chang HY. 2012.
Annu. Rev. Biochem. 81:145–66

Combinations of structural domains create a variety of scaffold lncRNAs



A) lncRNAs can contain independent structural domains that act as scaffolds to bring two protein complexes in spatial proximity.

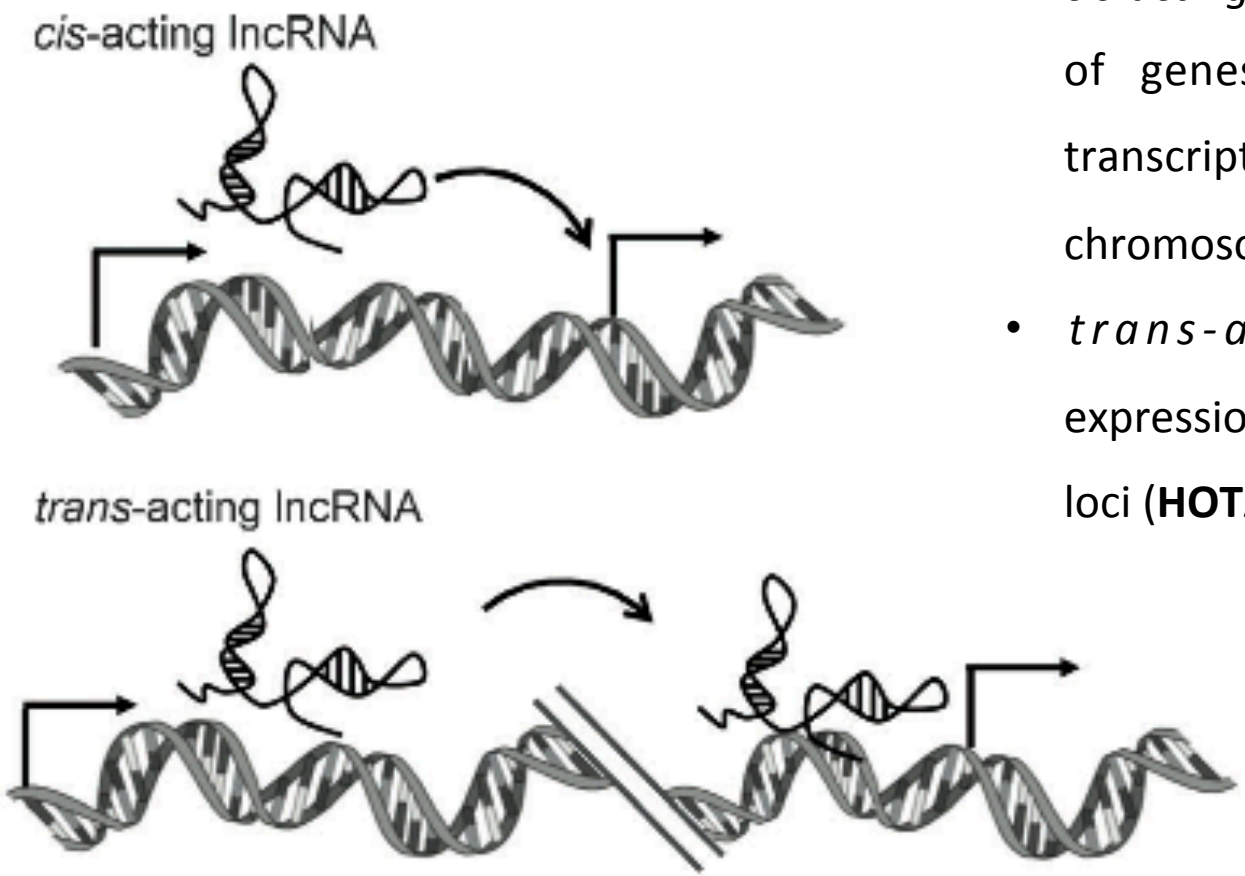


B) Independent structural domains of lncRNAs can be combined by alternative splicing, giving rise to lncRNAs with different scaffold properties.

The Role of Long Noncoding RNAs in the Epigenetic Control of Gene Expression

Mariangela Morlando,[a] Monica Ballarino,[a] Alessandro Fatica,[a] and Irene Bozzoni*[a, b, c]

Nuclear lncRNAs can be classified as:



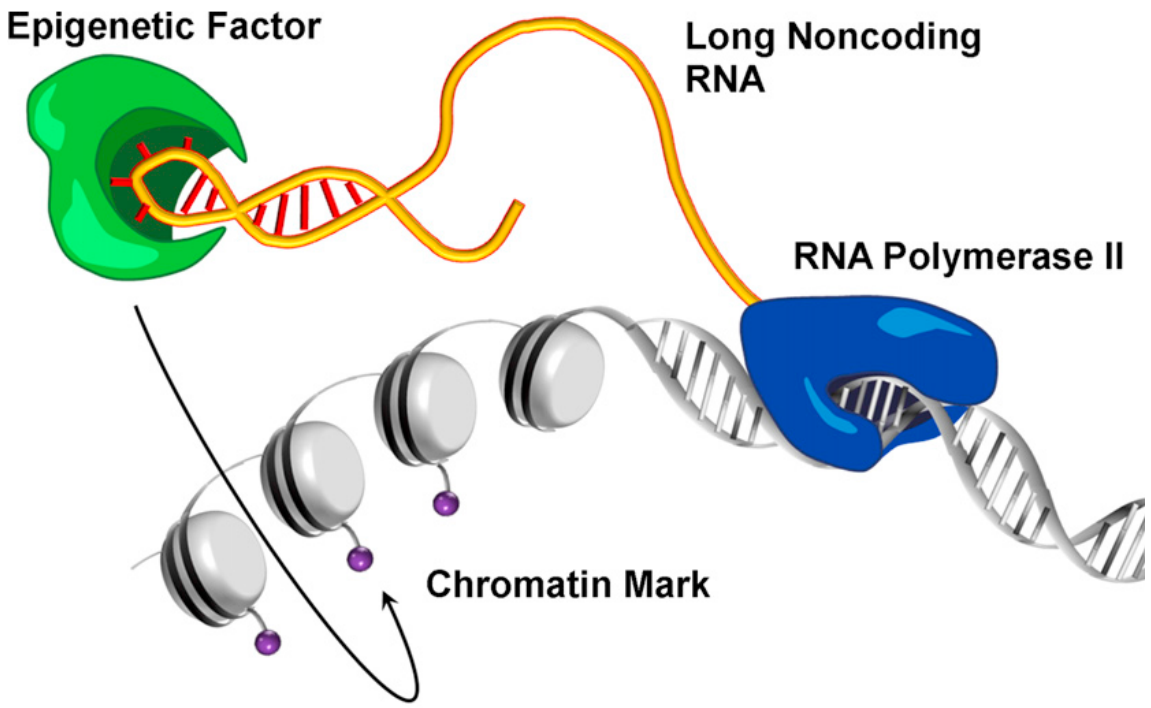
- *cis-acting*, which regulate expression of genes in the vicinity of their transcription site or on the same chromosome (**HOTTIP**, Mistral)
- *trans-acting*, which regulate expression of genes at independent loci (**HOTAIR**).

The Role of Long Noncoding RNAs in the Epigenetic Control of Gene Expression

Mariangela Morlando,[a] Monica Ballarino,[a] Alessandro Fatica,[a] and Irene Bozzoni*[a, b, c]

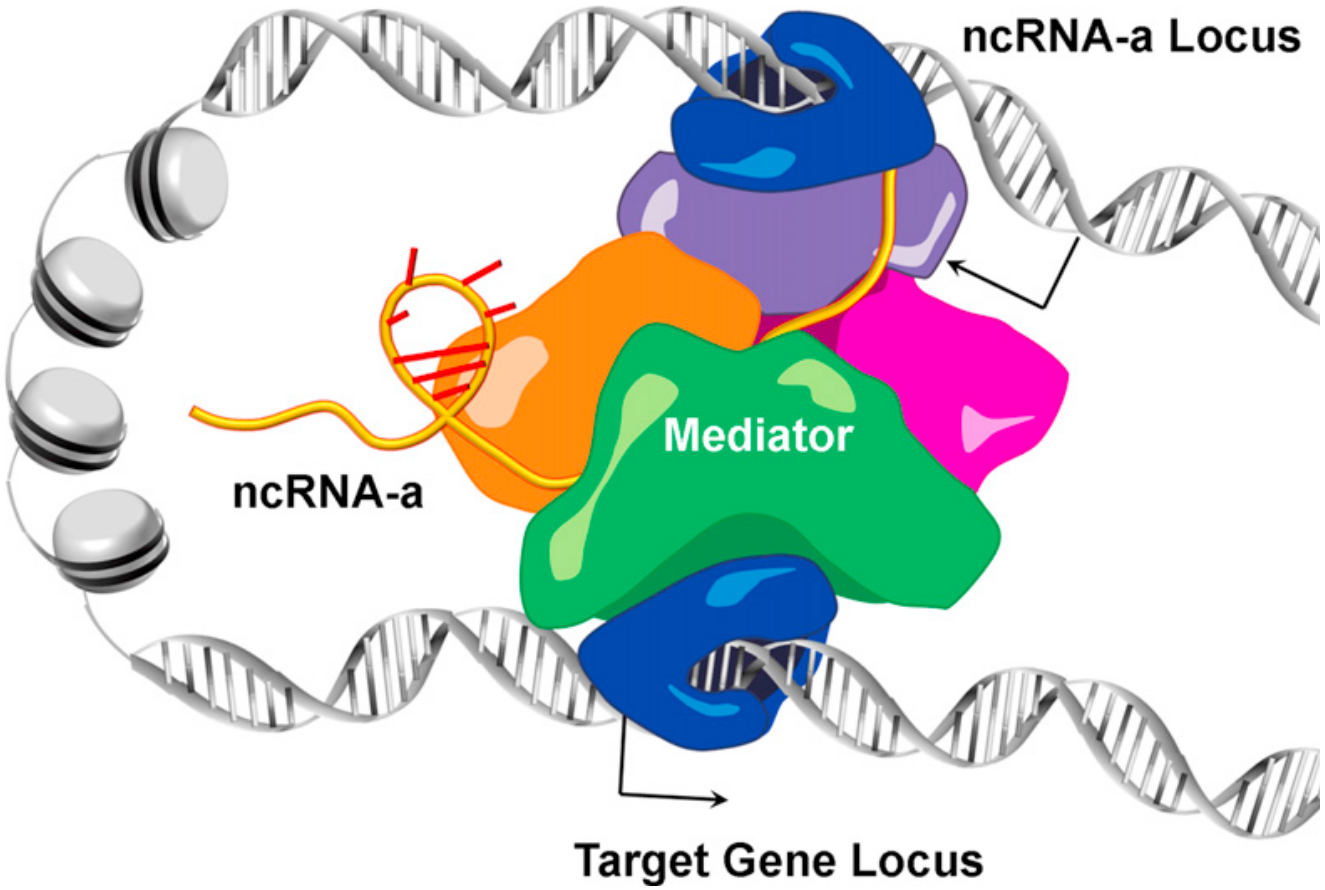
Cis lncRNAs are uniquely suited to act as allele- and locus-specific recruiters by virtue of their:

- *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).



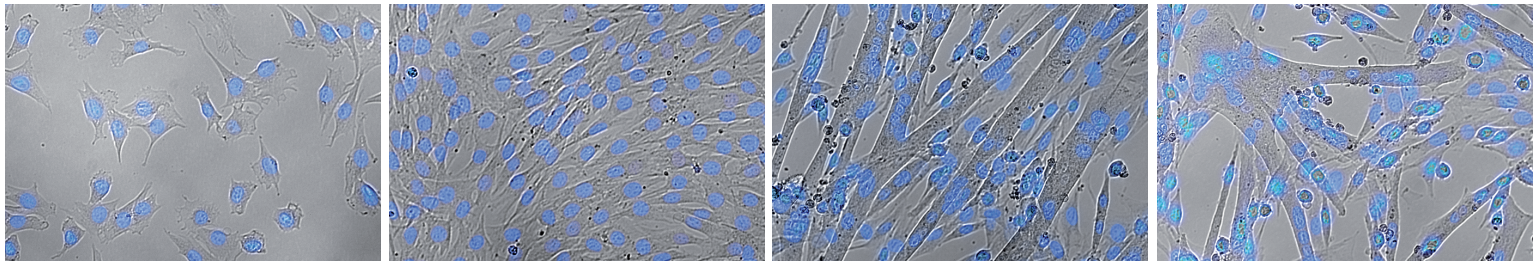
Mediator thus acts as a bridge between transcription factors binding at distant enhancers and the RNA polymerase II (RNAPII) apparatus at target promoters.

How to discover

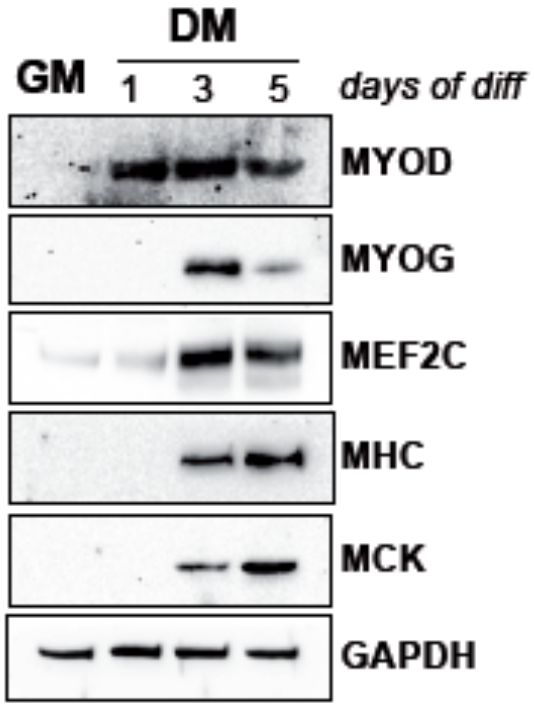
lncRNAs

Investigating the C_2C_{12} transcriptional landscape

GM DM1 DM3 DM5



C_2C_{12} murine cell line



Western blot

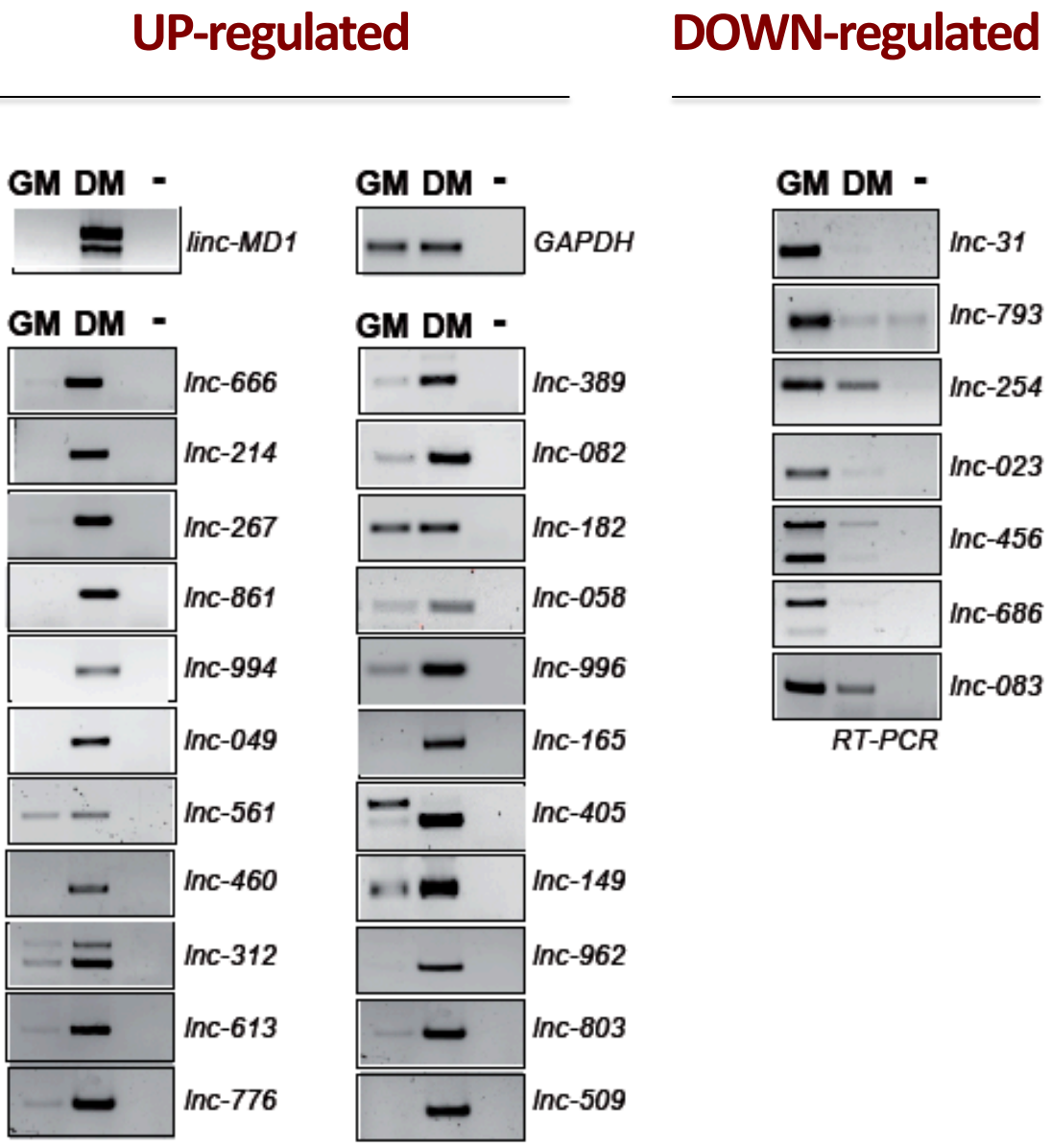
mRNA-seq data

***de-novo* transcript assembly**

Filter to remove non-lncRNAs

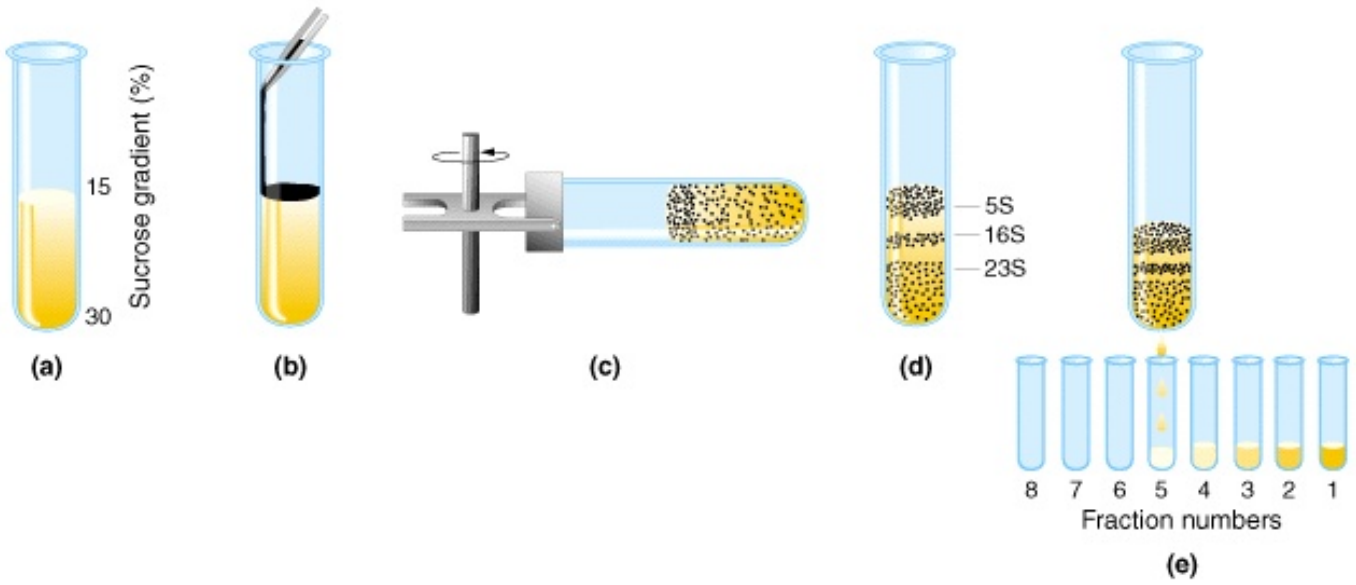
Sorting out the lncRNA mechanisms of action

- stage-specific expression during *in vitro* muscle differentiation
- sub-cellular localization
- tissue specificity and expression in muscle pathology (as dystrophic mouse)



How to discover *nuclear* lncRNA

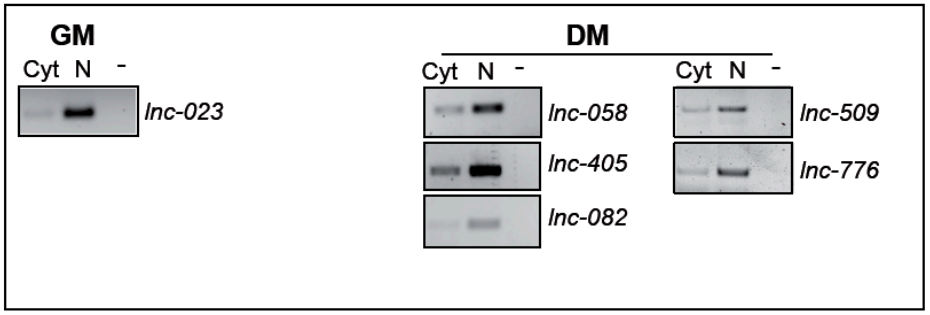
1) SUB-CELLULAR FRACTIONATION

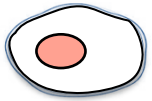


(a,b) A sample of macromolecules is layered over the pre-formed *sucrose* gradient. (c) The tube is placed in an ultracentrifuge and spun at *high* speed for several hours;

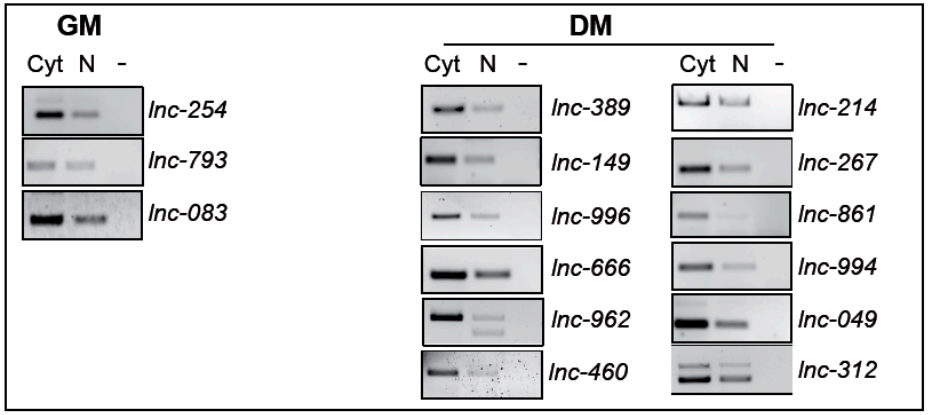
(d) Depending on the sizes & molecular weights of the macromolecular components, they migrate ("*sediment*") through the gradient at different rates: *lighter* molecules will move less quickly than more dense, *compact* molecules. Each molecular type will eventually form a discrete band at its isopycnic point, where its density equals that of the sucrose gradient. (e) Fractions can be recovered by poking a hole in the bottom of the centrifuge tube.

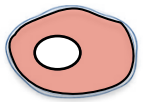
1) SUB-CELLULAR FRACTIONATION



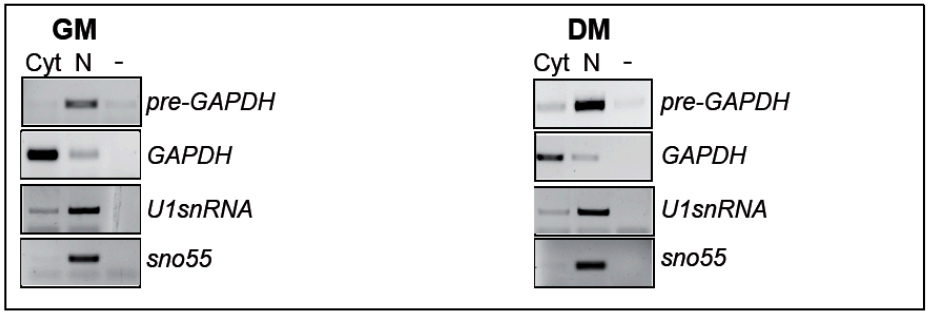


NUCLEAR lncRNAs





CYTOPLASMIC lncRNAs



control transcripts

2) RNA Fluorescent *in situ* hybridization (FISH)

Seeying is believing!

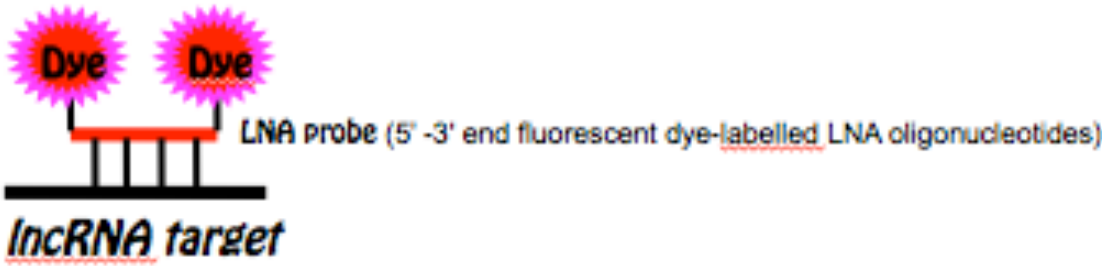
- It is a methodology for *detecting* and *localizing* RNA molecules in fixed cells.

- It utilizes *nucleic acid pro*bes that are complementary to target RNA sequences within the cell which hybridize to their RNA targets *via* standard Watson-Crick base pairing

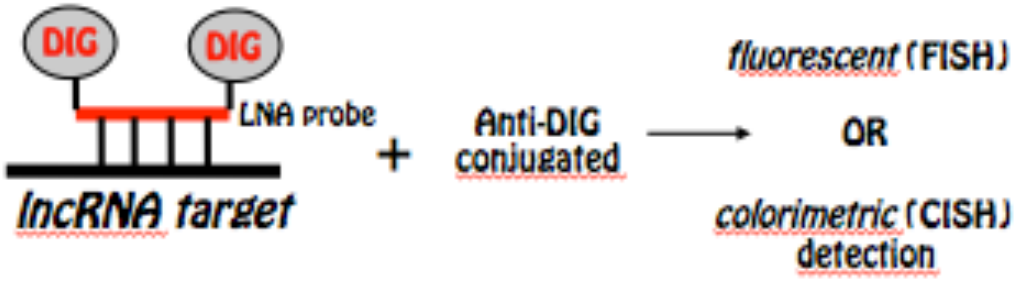
- Detection *via* fluorescence microscopy, either through direct conjugation of fluorescent molecules to the probe, or through fluorescent signal amplification schemes.

2) RNA Fluorescent *in situ* hybridization (FISH)

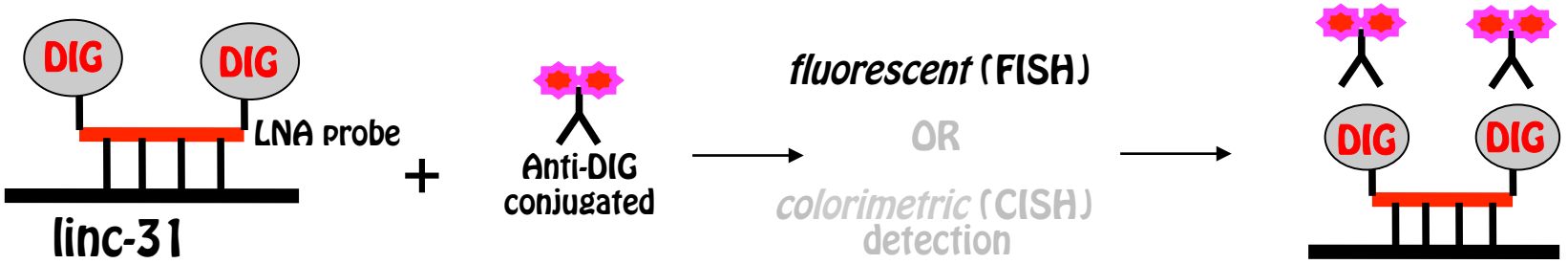
- DIRECT method (abundant lncRNAs)

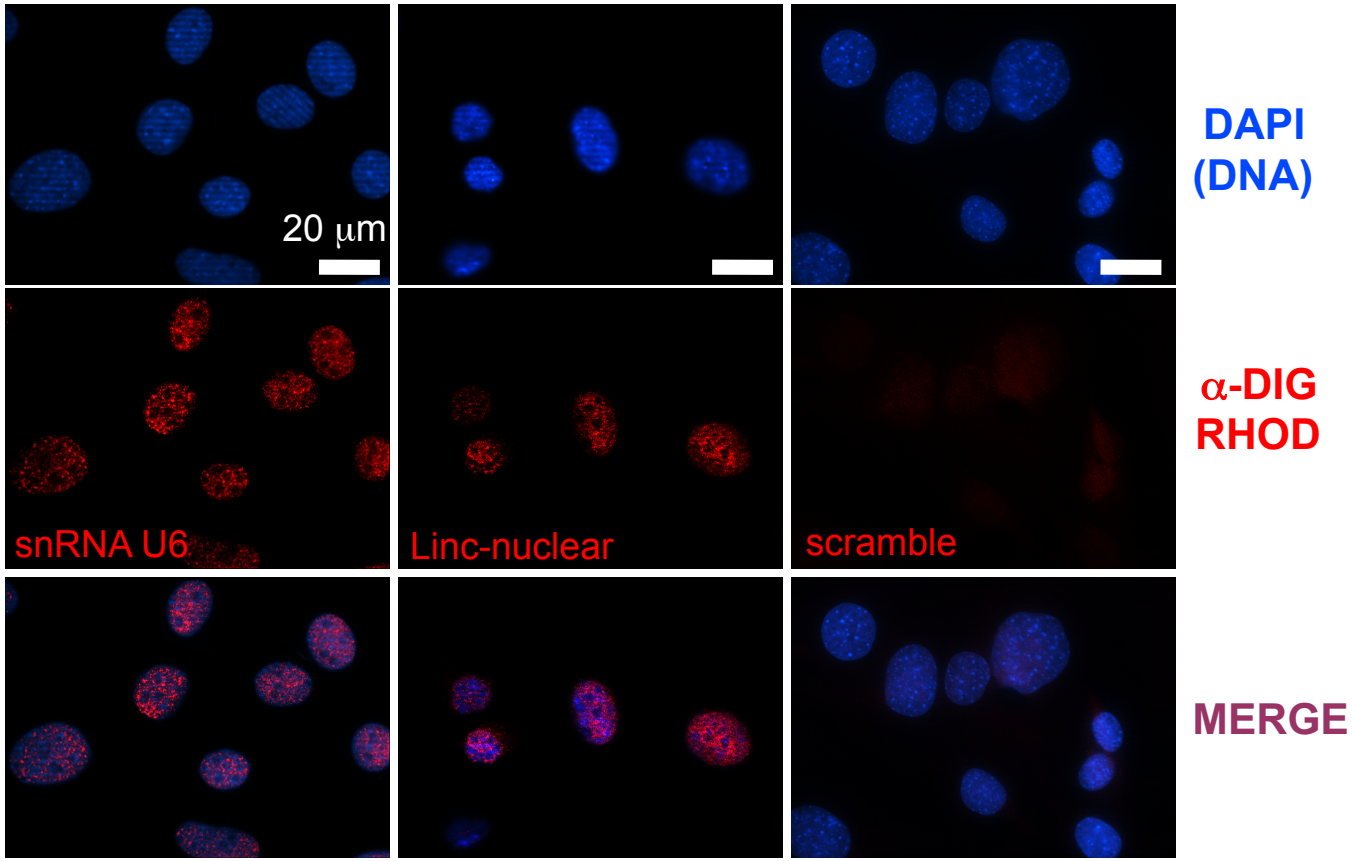


- INDIRECT method (low-expressed lncRNAs)

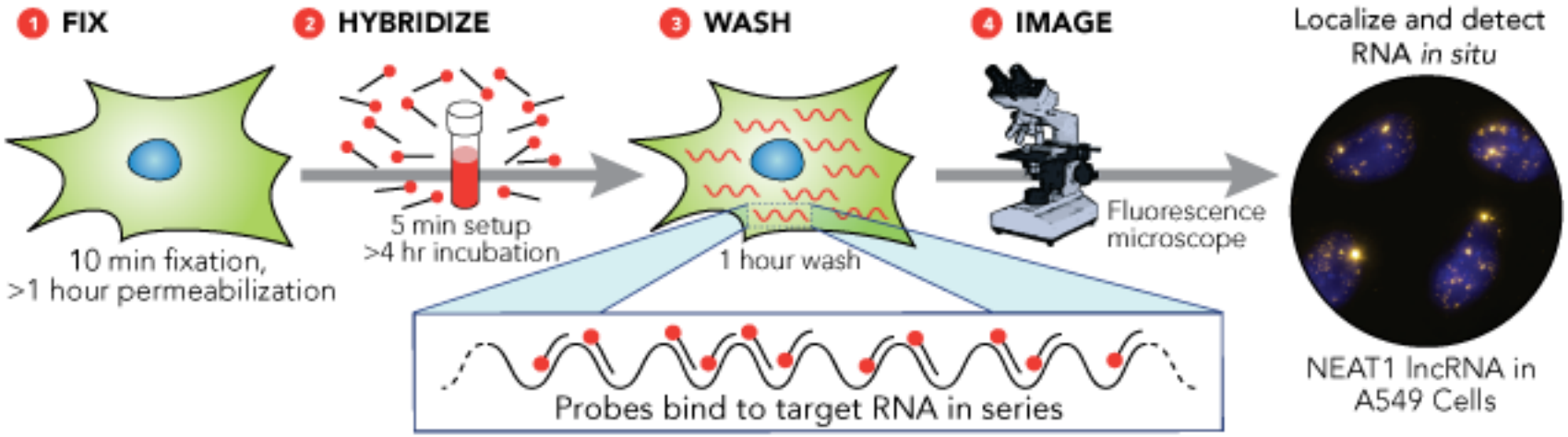


2) RNA Fluorescent *in situ* hybridization (FISH): in*direct* method

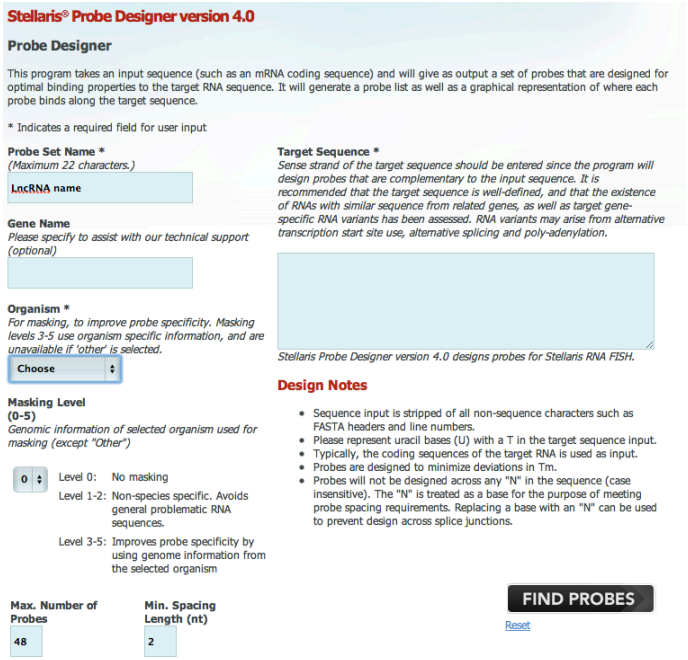




2) RNA Fluorescent *in situ* hybridization (FISH): *direct* method



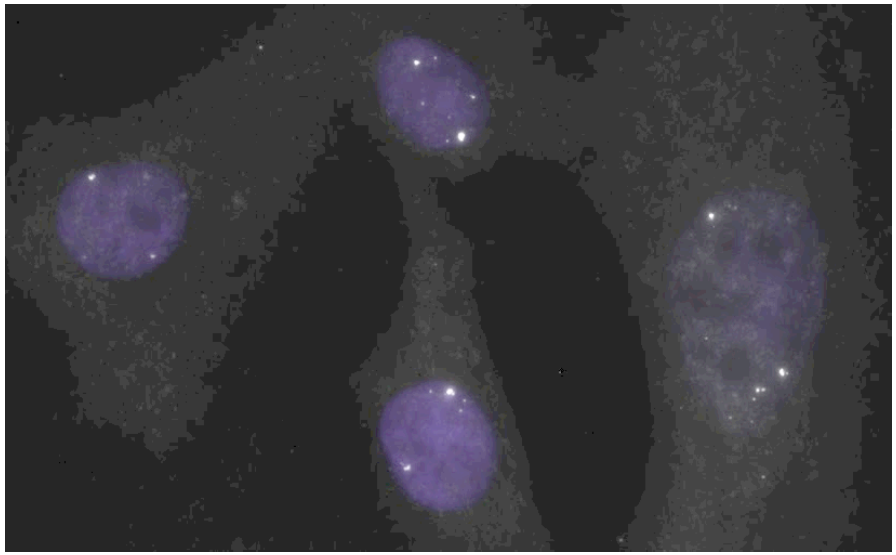
PROBES DESIGN tool
Stellaris website



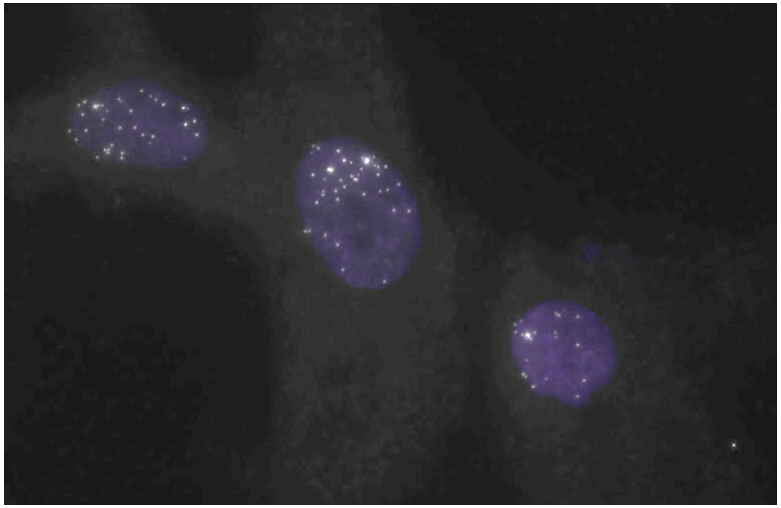
2) RNA Fluorescent *in situ* hybridization (FISH): *direct* method

…examples

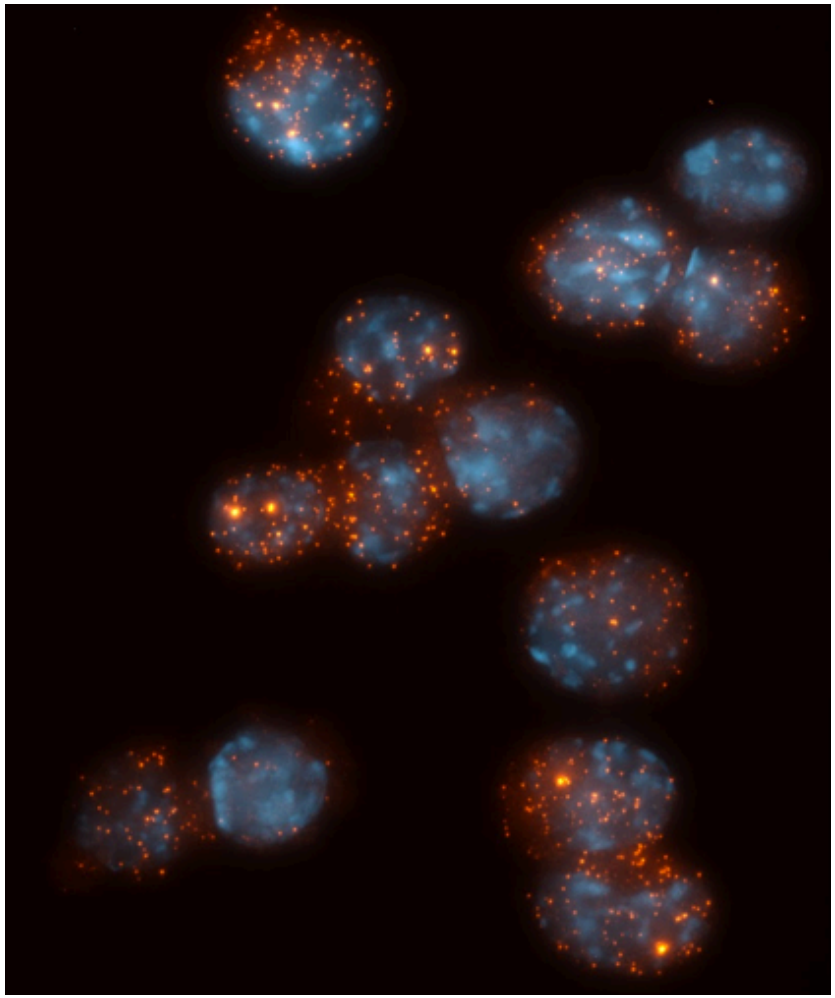
linc-SFPQ



linc-MKLN1A

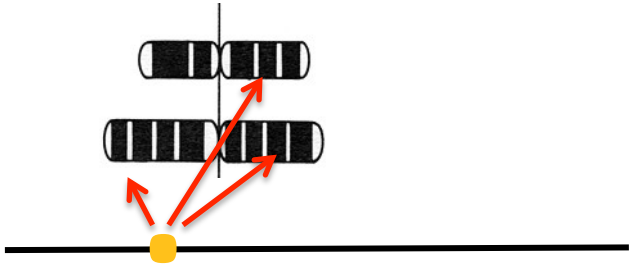


linc-HOXA1



***Raj Lab* website**

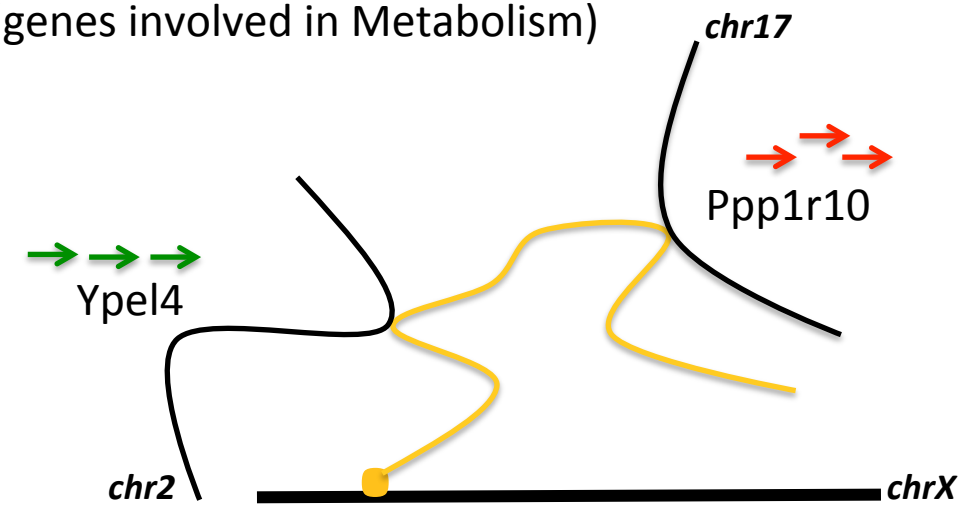
Discovering lncRNA trans-genomic regulation



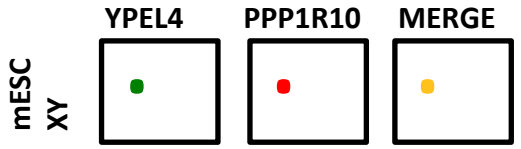
Firre lncRNA *locus* bring together different chromosomal domains

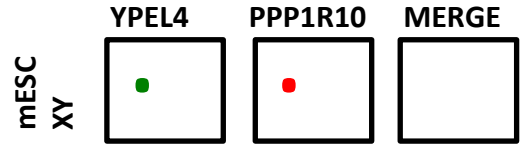
CHART: Ppp1r10 and **Ypel4** (genes involved in Metabolism)

Co-FISH:



Yellow=colocalization





ΔlincFIRRE

Cellular identity in an organismis determined by epigenetic factors that modulate specific gene expression programs

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

transcriptionally ***active*** genes correlate with

- H3K4me2/3 (di- and trimethyl),
- H3K36me2/3 (di- and trimethyl),
- H3K79me2 (dimethyl)

transcriptionally ***silent*** **regions** contain

- H3K9me3 (trimethyl),
- H3K27me2/3 (di- and trimethyl),
- H4K20me1 (monomethyl)

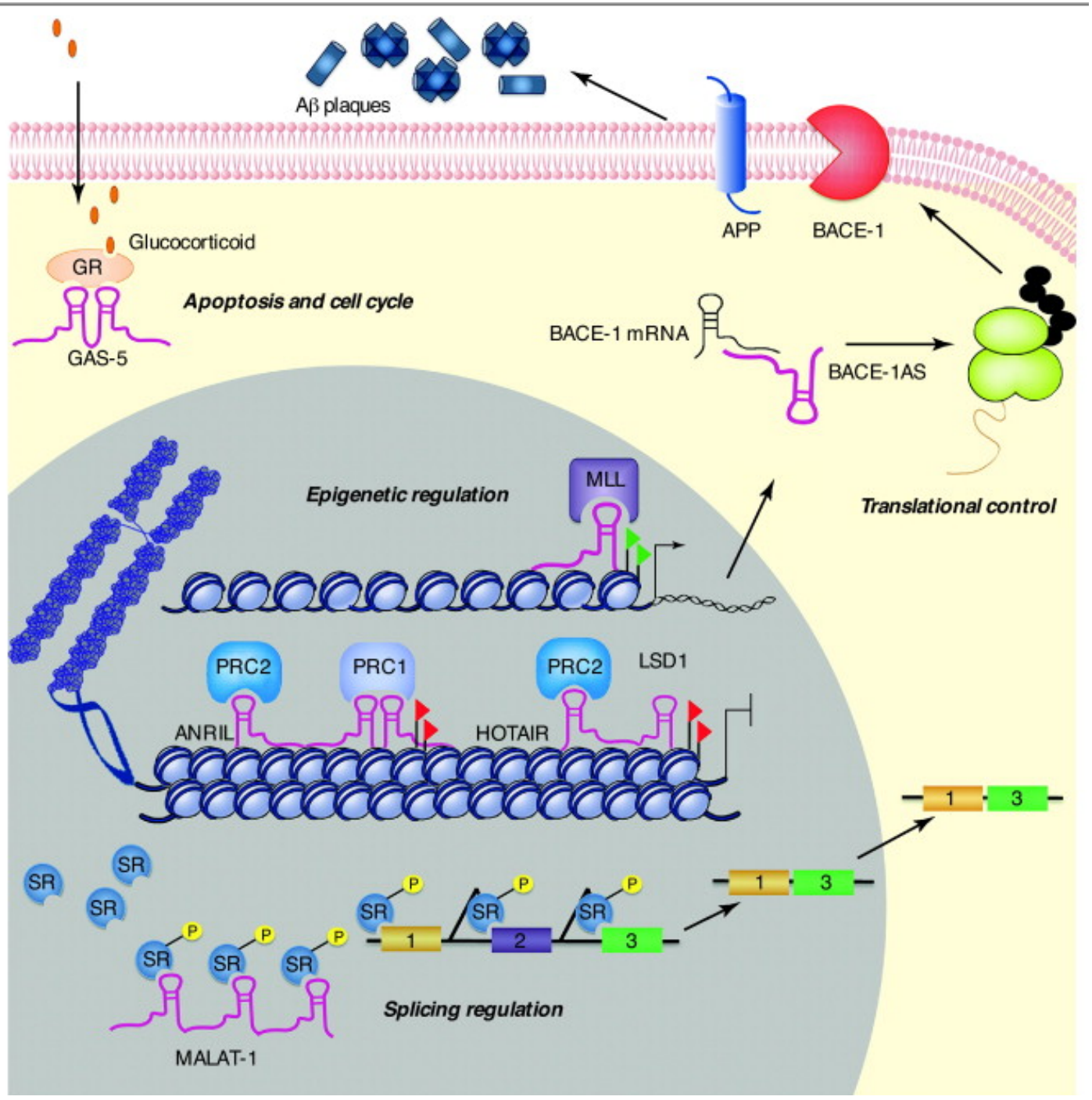
The histone code

Several combinations of histone modifications establish a "**histone code**" able to *demarcate* distinct regions within *enhancers*, core *promoters* and *ORFs* in a way that is critical for the regulation of chromatin-related processes.

The different types of modifications are deposited by a variety of well-characterized enzymes, which include two main systems of chromatin-modifying activities:

Polycomb (PcG) and Trithorax (TrxG) groups of proteins

How do these ubiquitous enzymes, which lack DNA binding capacity, recognise their target genes in the various cell types ?



Recent examples of mutated lncRNAs implicated in disease include **ANRIL** and **HOTAIR** that bind to chromatin-remodeling complexes PRC1 and PRC2 to alter chromatin and transcription. **GAS5** lncRNA acts as a decoy for the GR transcription factor and prevents GR from binding to DNA and transcriptional activation. **MALAT1** RNA binds to SR proteins to regulate mRNA alternative splicing, whereas BACE-1AS RNA binds to the complementary BACE-1 mRNA to regulate BACE-1 translation.

Red chromatin marks denote transcriptional inhibition. Green chromatin marks denote transcriptional activation. Abbreviations: GR, glucocorticoid receptor; GAS-5, growth arrest-specific 5 ncRNA; Aβ plaques, amyloid-β plaques; APP, amyloid precursor protein; BACE-1, β-site APP-cleaving enzyme; βAPP BACE-1AS mRNA, βAPPBACE-1 antisense ncRNA; MLL, mixed-lineage leukemia; PRC1, polycomb repressive complex 1; PRC2, polycomb repressive complex 2; ANRIL; antisense lncRNA of the *INK4 locus;* HOTAIR, HOX antisense ncRNA; MALAT-1, metastasis associated in lung adenocarcinoma transcript; SR, serine/arginine-rich family of nuclear phosphoproteins; P, phosphorylation.

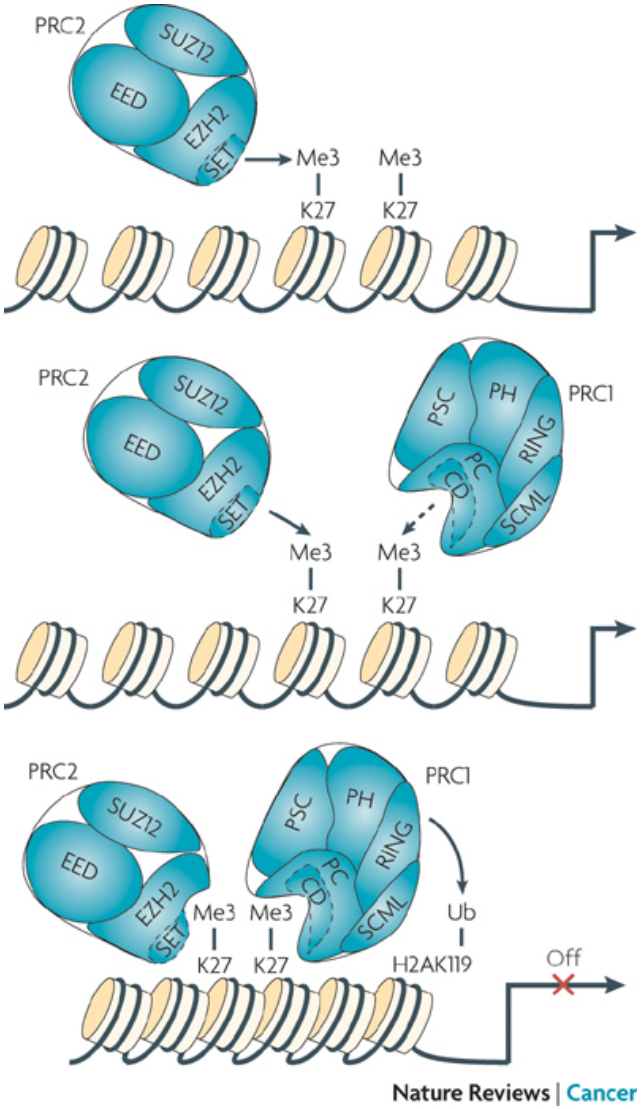
Polycomb group proteins (PcG)

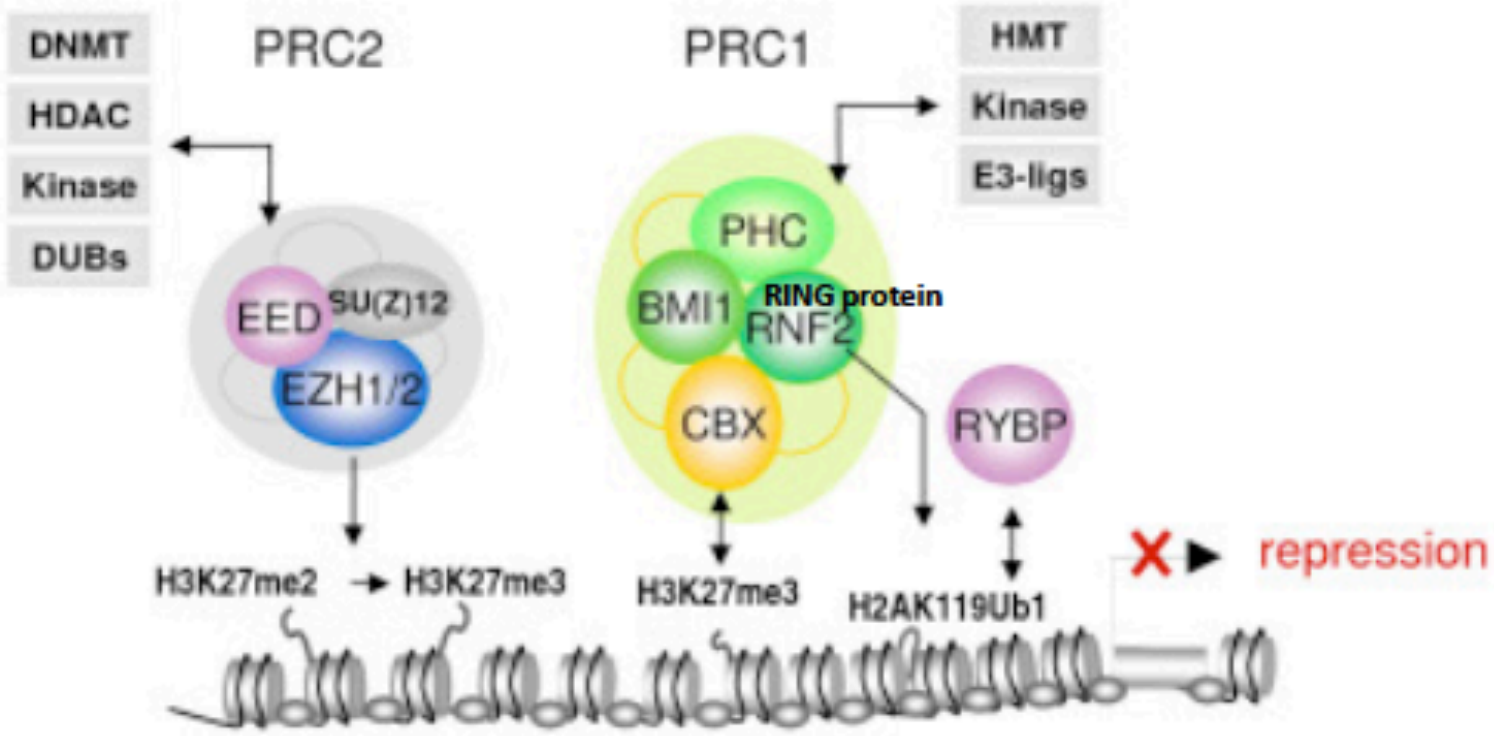
The PcG proteins form multiprotein repressive complexes, called Polycomb repressive complexes (PRCs), which repress transcription by a mechanism that involves the modification of chromatin

Two major Polycomb repressive complexes (PRCs) have been described:

•The **PRC2** contains the histone methyltransferase **EZH2**, which together with **EED** and **SUZ12** catalyses the **trimethylation of histone H3 at lysine K27 (H3K27me3).**

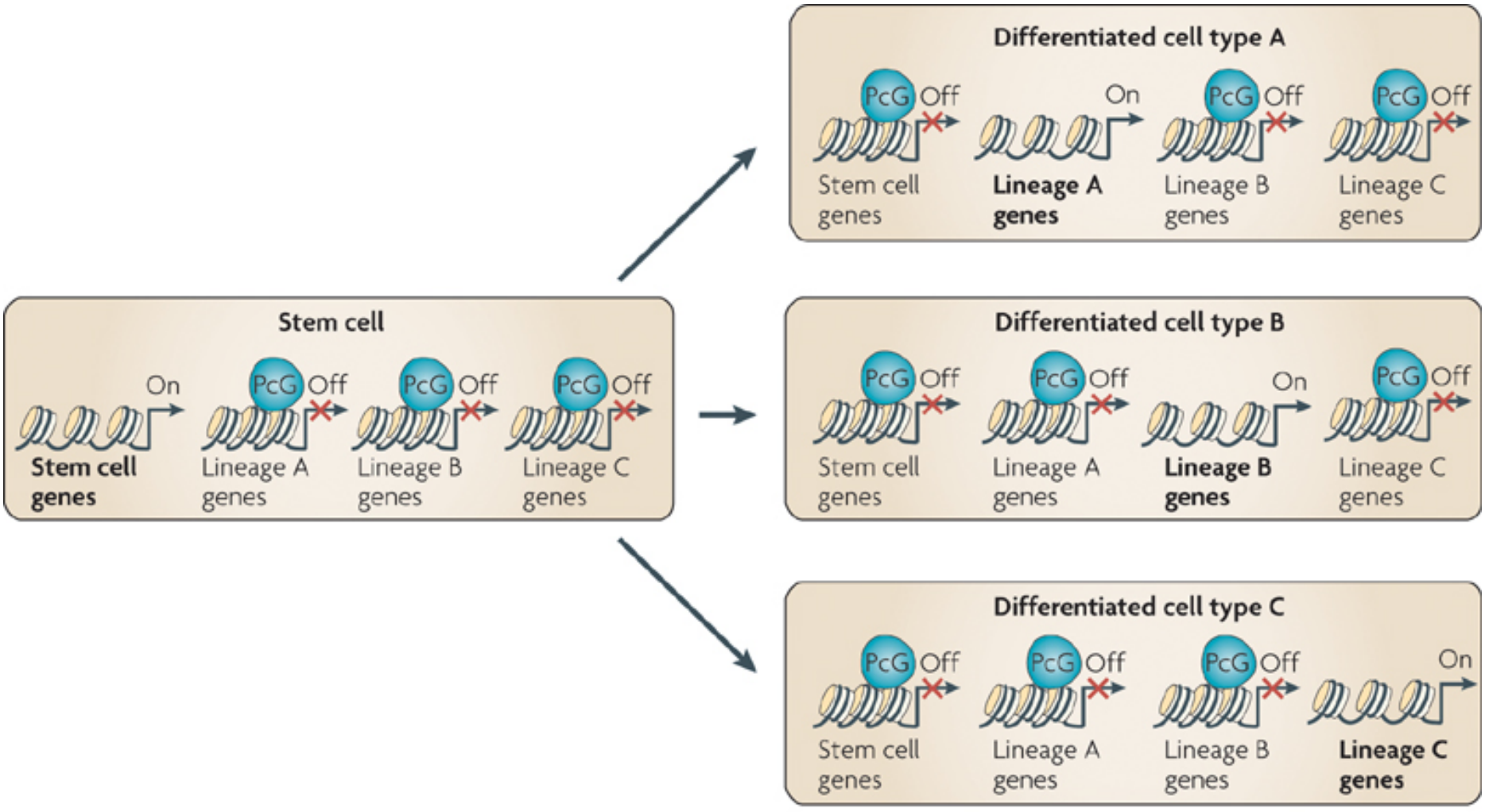
•The **PRC1** complexes are recruited by the affinity of chromodomains in chromobox (Cbx) proteins to the H3K27me3 mark. PRC1 recruitment results in the RNF1 and RNF2-mediated **ubiquitylation of histone H2A on lysine 119**, which is thought to be important for transcriptional repression.




DNMT
HDAC
Kinase
DUBs
PRC2
EED
SU(Z)12
EZH1/2
H3K27me2
H3K27me3
PRC1
HMT
Kinase
E3-ligs
PHC
BMI1
RING protein
RNF2
CBX
RYBP
H3K27me3
H2AK119Ub1
repression

PcGs and differentiation

PcG proteins are required for maintaining the correct identities of stem, progenitor and differentiated cells

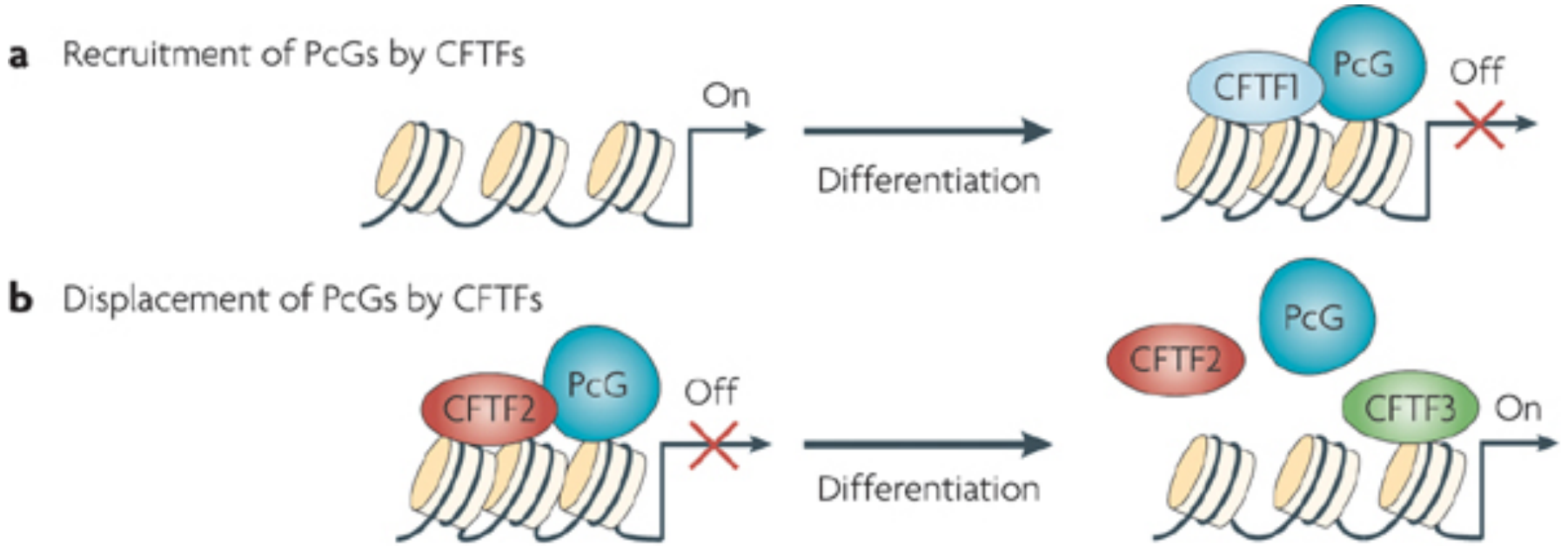


Nature Reviews | Cancer

PcG recruitment to target genes

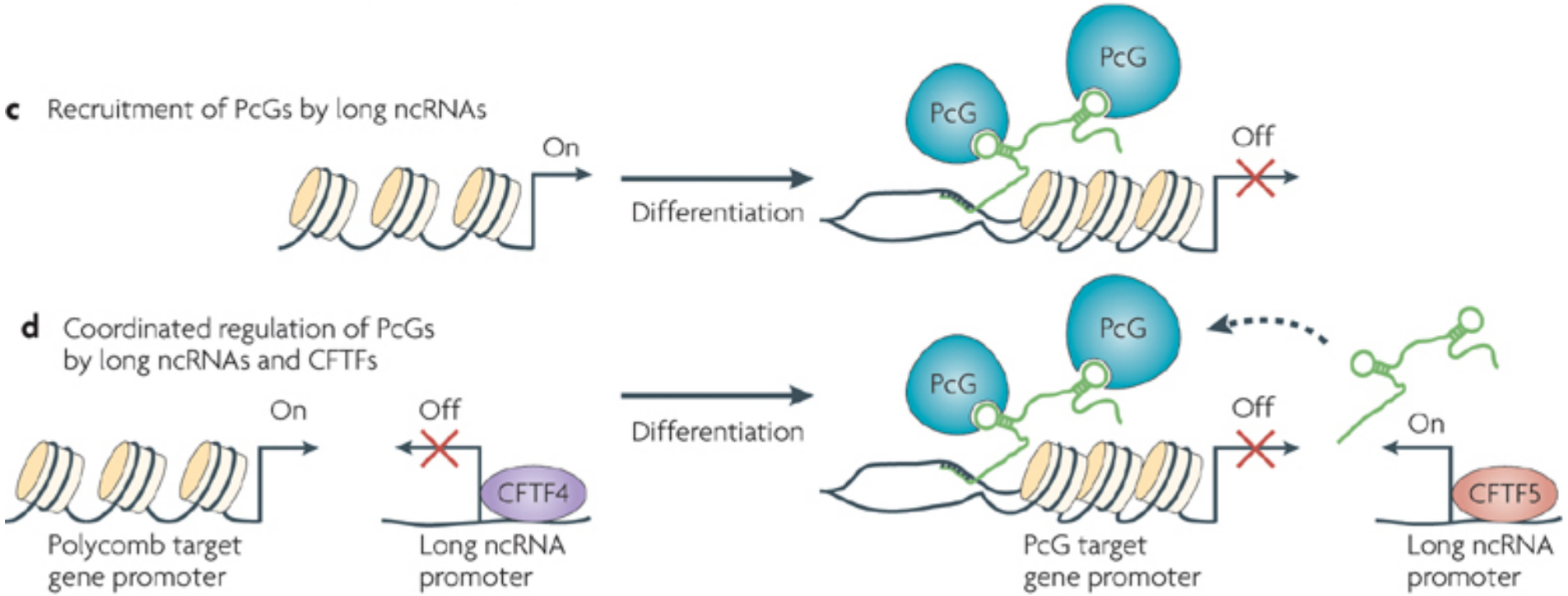
- PcG proteins do not have the ability to bind specific DNA motifs (exception in *Drosophila* - PRE)

- cell fate transcription factors **CFTFs** may regulate PRC binding to specific promoters

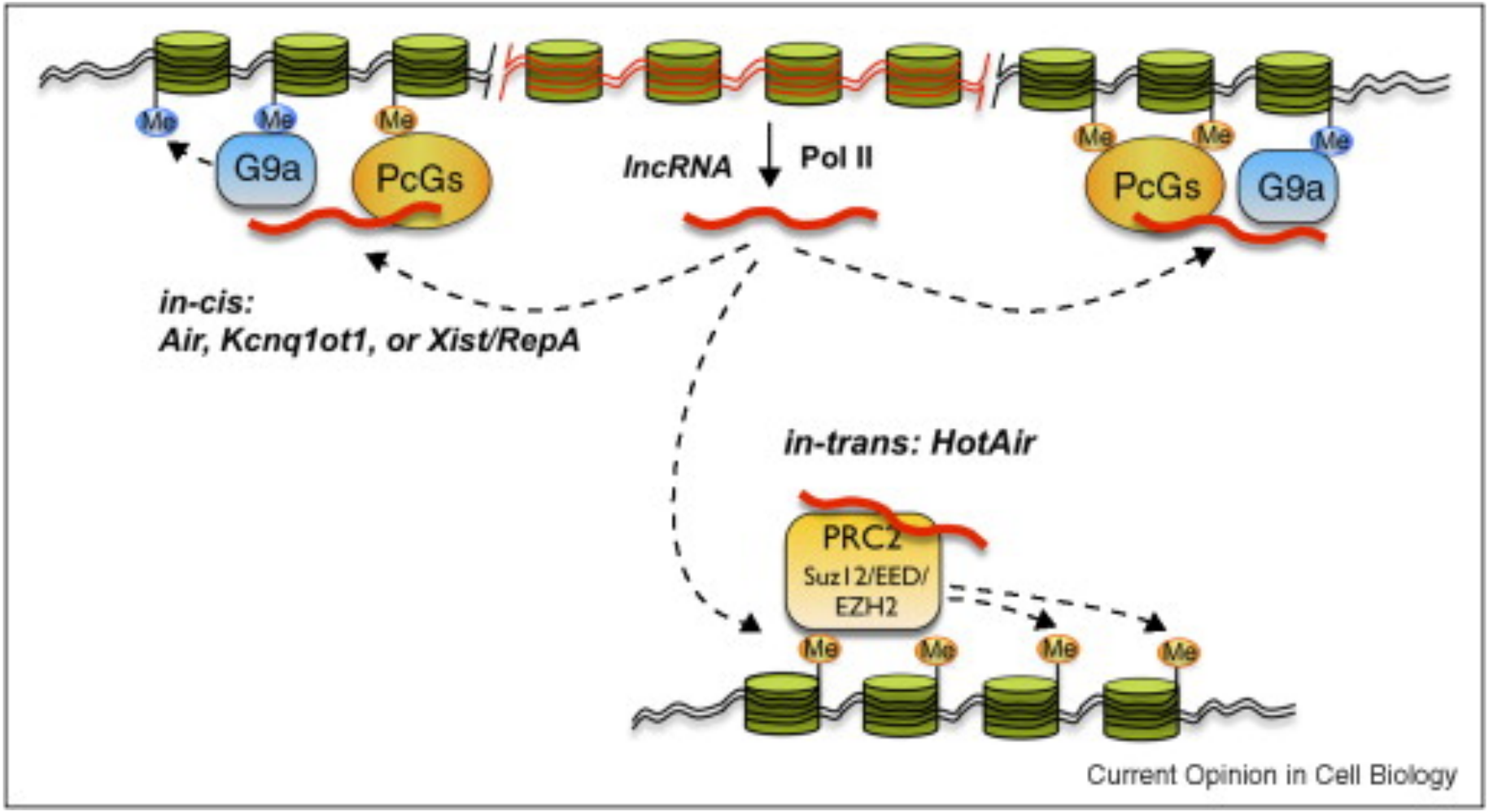


PcG recruitment to target genes

Long non coding RNAs may also regulate PRC binding to specific promoters



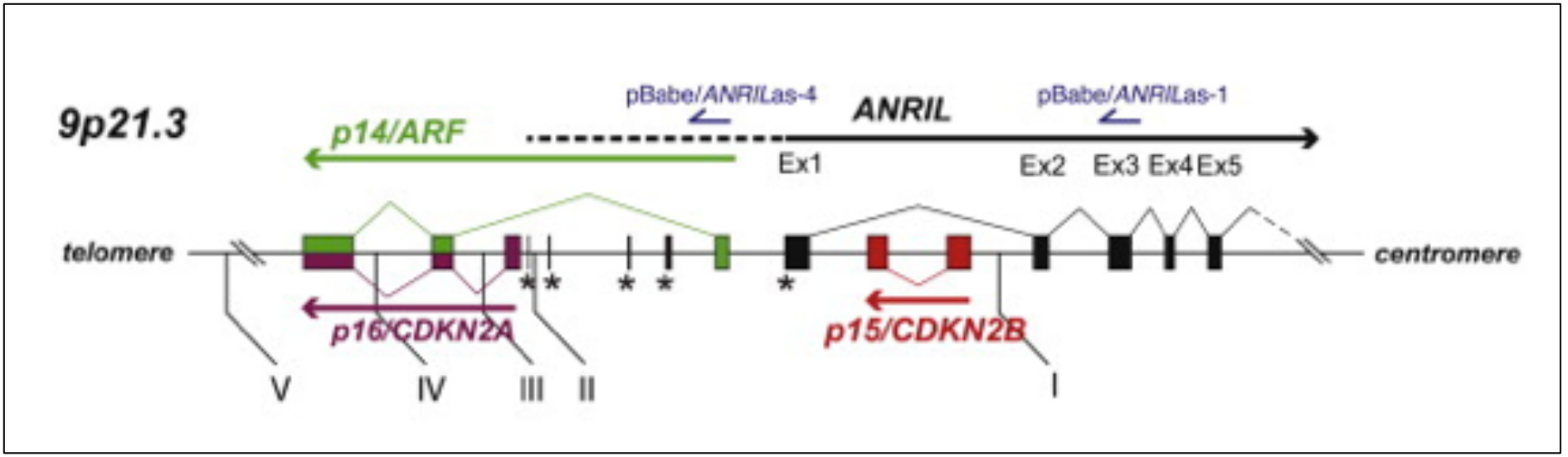
LncRNAs are key component of epigenetic regulatory networks (e.g. XIST, HOTAIR, AIR and KCNQ1QT1). Around 20% of them associate with chromatin repressive complexes (e.g. PRC2) and their depletion affects the ability of PRC2 to regulate specific genes.



ANRIL is a signature transcript for docking PcG proteins

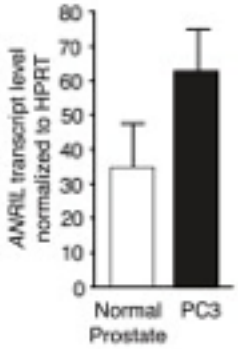
ANRIL is a long noncoding antisense RNA transcript overlapping the INK4b/ARF/INK4a locus (Pasmant et al., 2007) with estimates of the transcript spanning over 30–40 kb in length. This transcript promotes and maintains the epigenetic state of the CDKN2B (INK4b) gene and is an antisense transcript to INK4b.

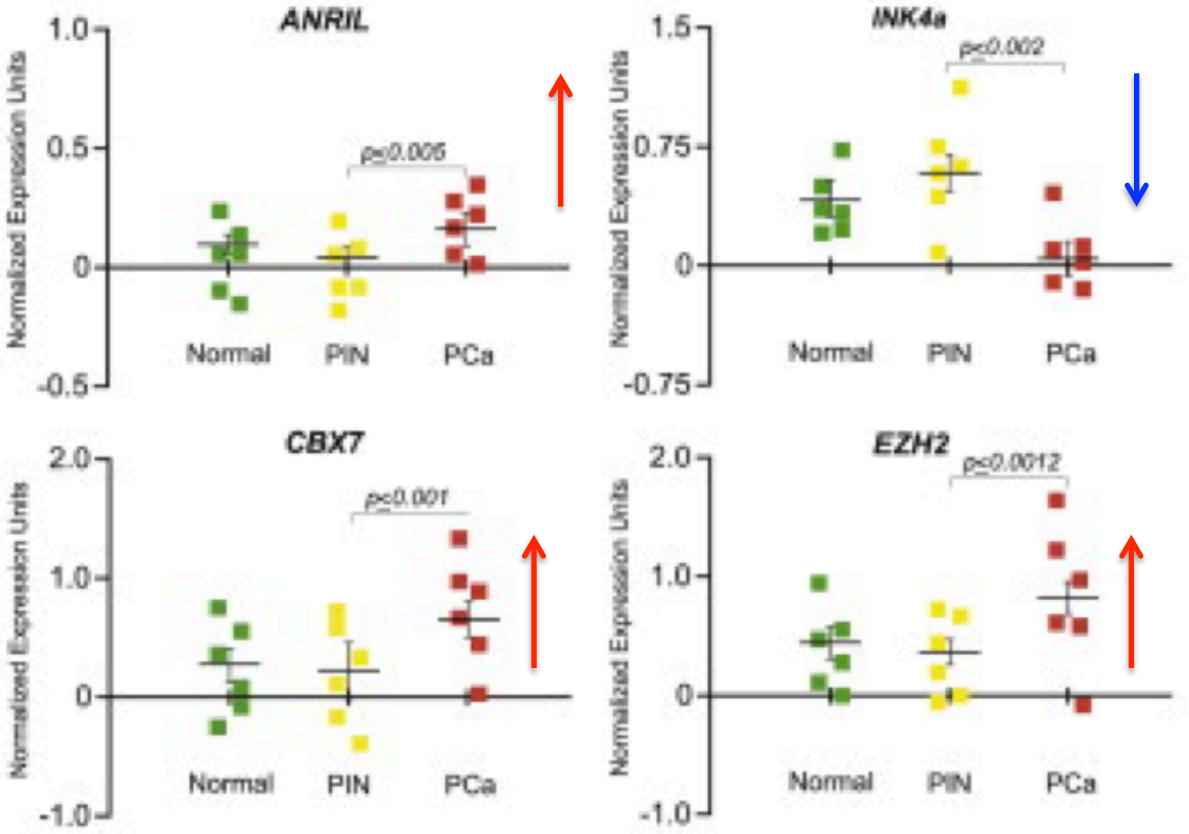
Genomic organization of the INK4a/ARF gene cluster



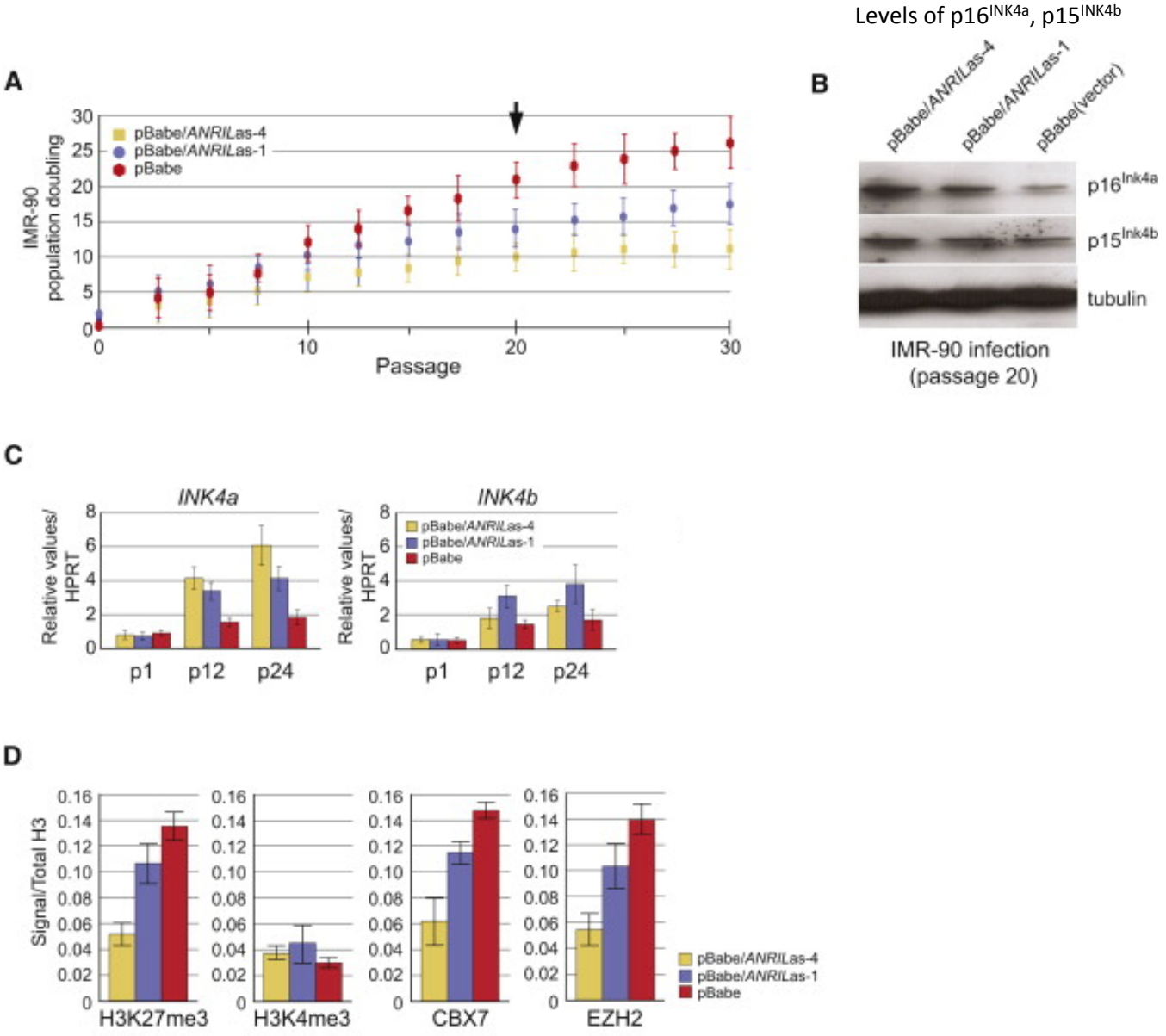
CBX7 and its interacting ncRNA ANRIL are upregulated in prostate cancer tissues

Normalized expression levels of **ANRIL**, **p16**, **CBX7**, and **EZH2** in *normal* prostate epithelium cells, and *preneoplastic* PIN and PCa prostate *cancer* cells.



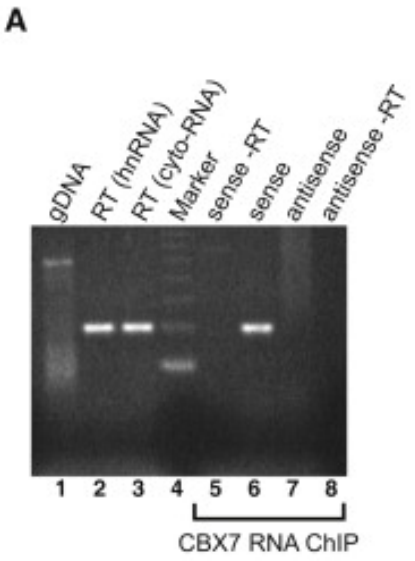


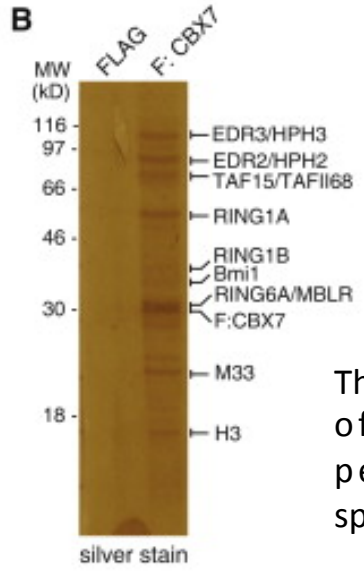
Inhibition of ANRIL Transcript Interactions Results in Loss of INK4a Repression and Limitation of Cellular Life Span

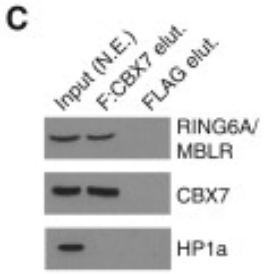


ANRIL RNA Is Associated with Polycomb Protein Recruitment to the INK4a/ARF Locus

Recovery of ANRIL transcript by CLIP analysis of CBX7 in chromatin

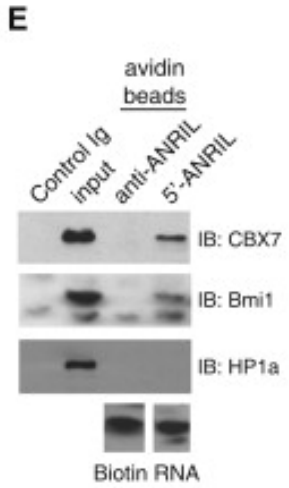


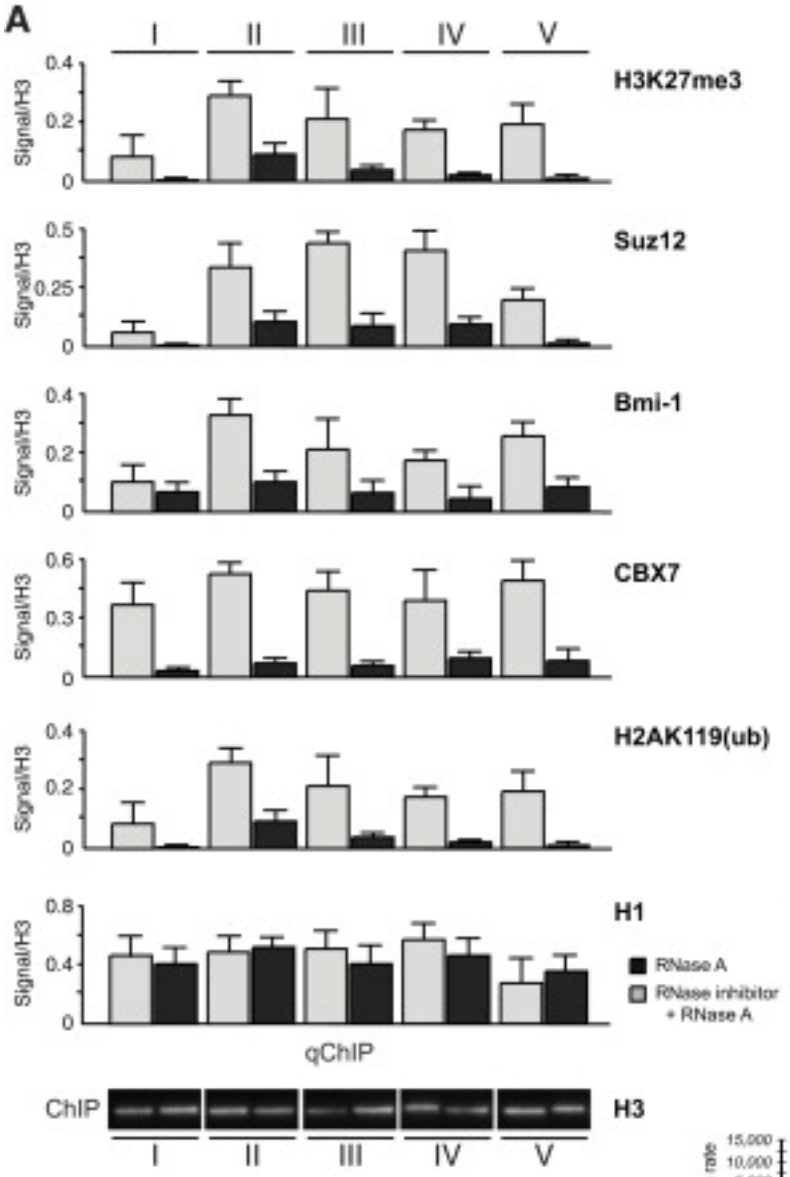




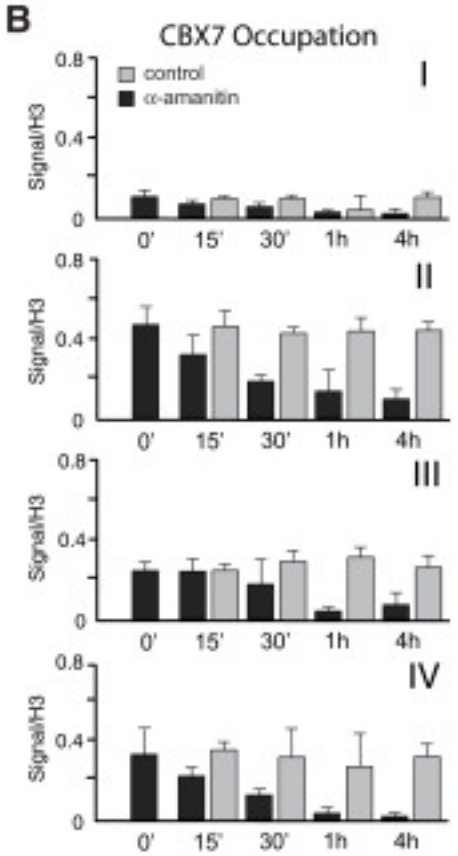
The human **CBX7 complex**: identification of CBX7/PRC1 components was performed by MALDI-TOF mass spectrometry

Nuclear extract from PC3 cells was used to evaluate by western blot analysis the *association* of **PRC1** proteins CBX7, Bmi1, and HP1α with the biotinylated **ANRIL** transcript

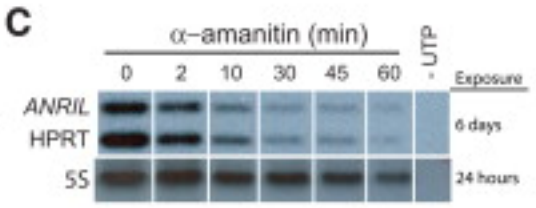




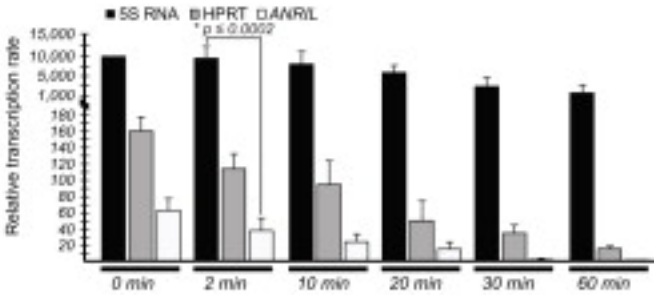
Recruitment of Polycomb Proteins to the INK4a/ARF Locus: RNAse treatment reduced the recruitment



Depleted levels of CBX7 upon whole-cell treatment of α-amanitin: the RNA associating with CBX7 is a Pol II *nascent* transcript



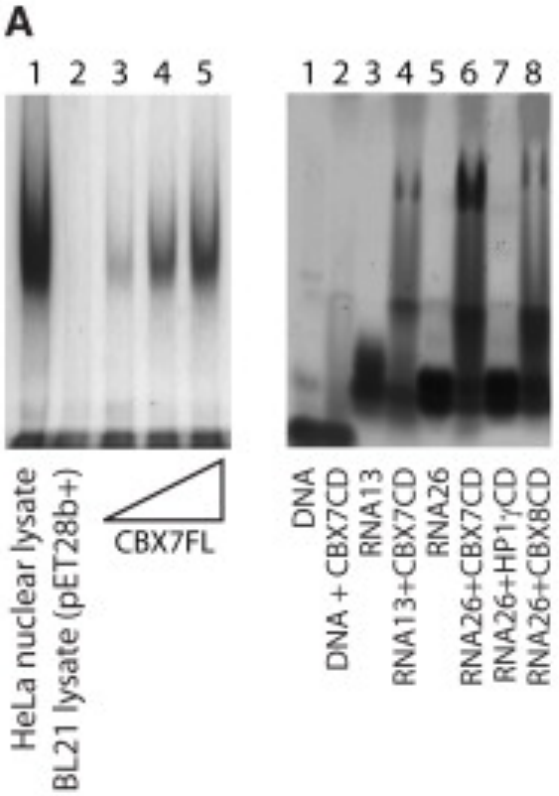
In vitro transcription: ANRIL is a nascent transcript generated by Pol II



CBX7 chromodomain binds H3K27me3 or ANRIL with distinc domains

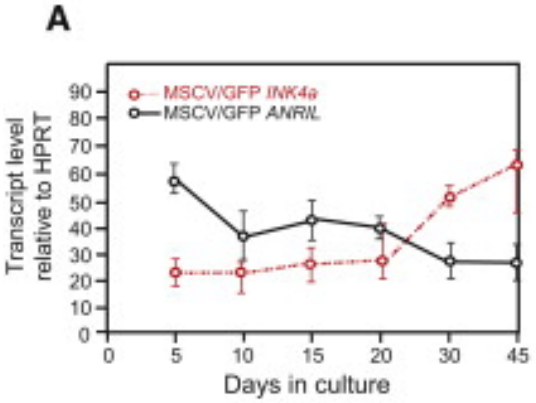
RNA-EMSA showing CBX7 full-length (CBX7FL, *left*) and chromodomain (CBX7CD, *right*) directly associate to RNA.

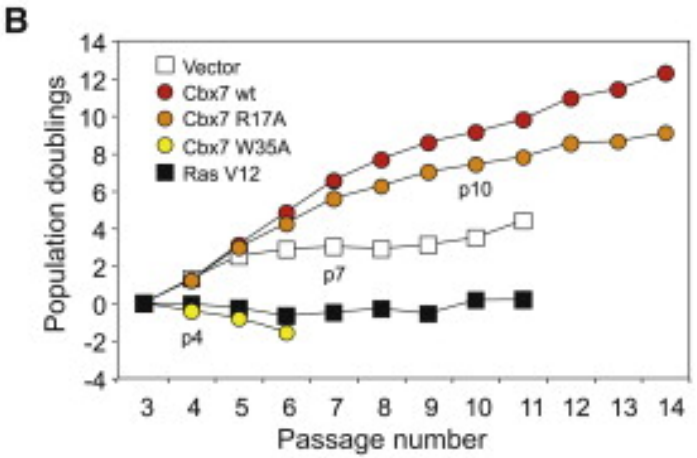
13 nt ssDNA, 13 nt RNA (RNA13), or 26 nt RNA (RNA26)



- The three components (*CBX7CD*, *H3K27me3* peptide, and *ANRIL* loop RNA) are able to form a ternary complex (NMR analysis)
- CBX7 Chromodomain Uses Distinct Regions and Residues for Binding H3K27me or RNA (NMR analysis)

Relative Contributions of CBX7 Interactions with H3K27me and RNA to Senescence Regulation

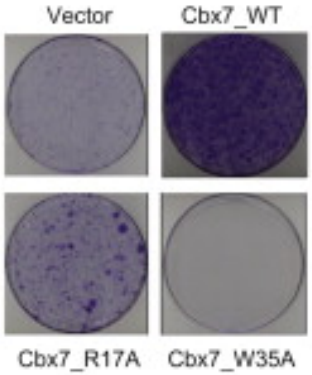




Overespression of WT of mutants

- Cbx7 wt→ extended lifespan, increased cell growth
- Cbx7 R17A mutant=RNA defective→ slight decrease cell growth
- Cbx7 W35A mutant=H3K27me defective→ growth arrest

C colony formation assay

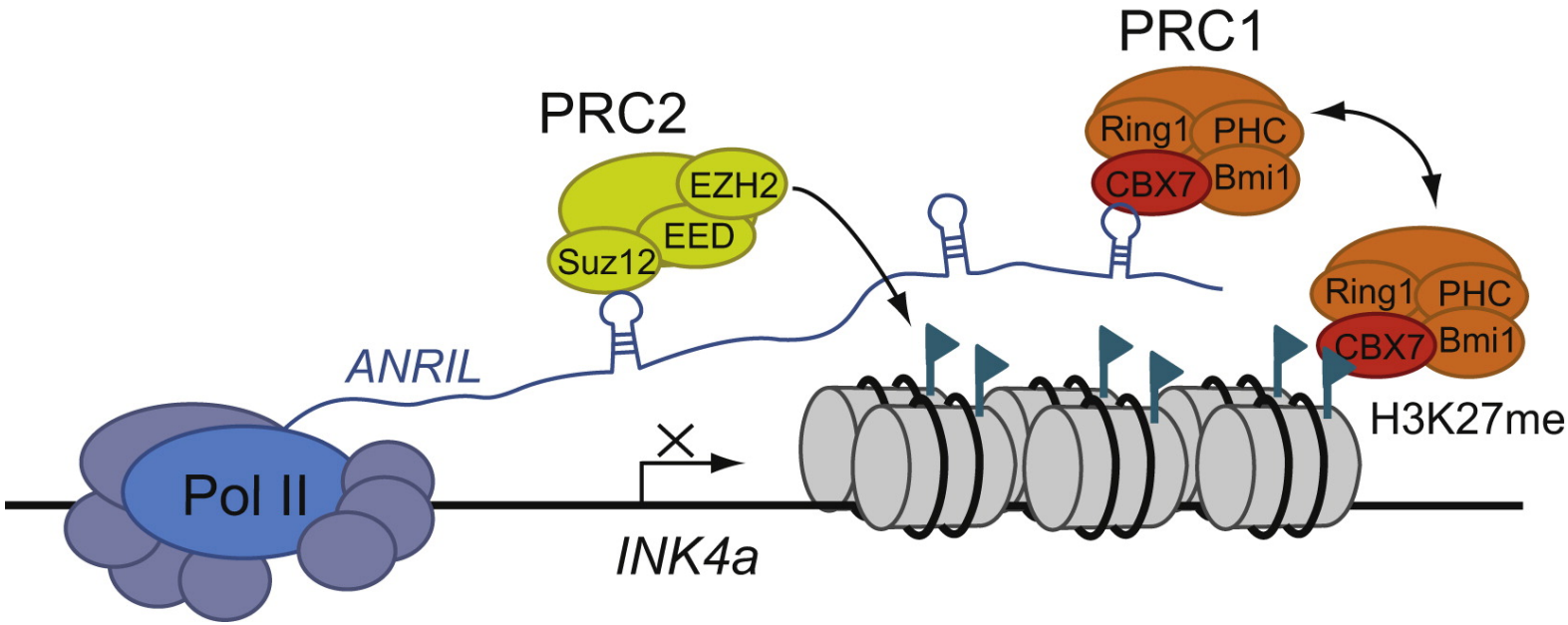


Both interactions are required for the INK4b/ARF/INK4a *locus* repression by CBX7

Ablation of either interaction results in premature growth arrest

Model of ncRNA-mediated gene repression of the INK4b/ARF/INK4a locus by Polycomb group complexes.

- molecular interplay between methylated H3K27 and ncRNA transcripts ANRIL act for silencing of the INK4b/ARF/INK4a locus by Polycomb CBX7.
- Both PRC1 and PRC2 complexes are retained at a repression site of a target gene through their association with the nascent ANRIL transcripts of Pol II.
- INK4b/ARF/INK4a is a tumor suppressor locus: oncogenic function!!!



Maintenance of HOX expression patterns is under complex epigenetic regulation

Anatomic specific expression: **HOTTIP** and **HOTAIR**

Nature. 2011 Apr 7;472(7341):120-4. doi: 10.1038/nature09819. Epub 2011 Mar 20.

A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.

Wang KC1, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, Wysocka J, Lei M, Dekker J, Helms JA, Chang HY.

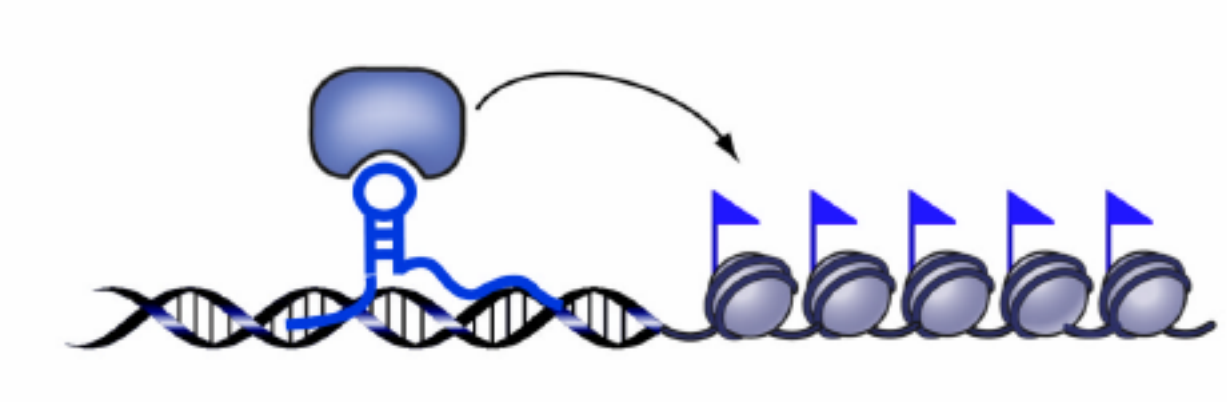
Abstract

The genome is extensively transcribed into long intergenic noncoding RNAs (lincRNAs), many of which are implicated in gene silencing. Potential roles of lincRNAs in gene activation are much less understood. Development and homeostasis require coordinate regulation of neighbouring genes through a process termed locus control. Some locus control elements and enhancers transcribe lincRNAs, hinting at possible roles in long-range control. In vertebrates, 39 Hox genes, encoding homeodomain transcription factors critical for positional identity, are clustered in four chromosomal loci; the Hox genes are expressed in nested anterior-posterior and proximal-distal patterns colinear with their genomic position from 3' to 5'of the cluster. Here we identify HOTTIP, a lincRNA transcribed from the 5' tip of the HOXA locus that coordinates the activation of several 5' HOXA genes in vivo. Chromosomal looping brings **HOTTIP** into close proximity to its target genes. HOTTIP **RNA binds the adaptor protein WDR5 directly and targets WDR5/MLL complexes across HOXA, driving histone H3 lysine 4 trimethylation and gene transcription**. Induced proximity is necessary and sufficient for HOTTIP RNA activation of its target genes. Thus, by serving as key intermediates that transmit information from higher order chromosomal looping into chromatin modifications, lincRNAs may organize chromatin domains to coordinate long-range gene activation.

An additional facet of lncRNA functional mechanism is revealed by studies employing **chromosome conformation capture (3C)**. The cis-acting HOTTIP RNA has been found to recruit the gene activating Trithorax group complex MLL to coordinately regulate loci in the HOXA cluster, as far apart as 40 kb, that have been brought into close proximity in 3D through long-range chromosomal interactions (Wang and Chang, 2011).

Chromosome conformation capture (3C) technologies and variants

HOTTIP as a paradigm of “guide” in *cis*



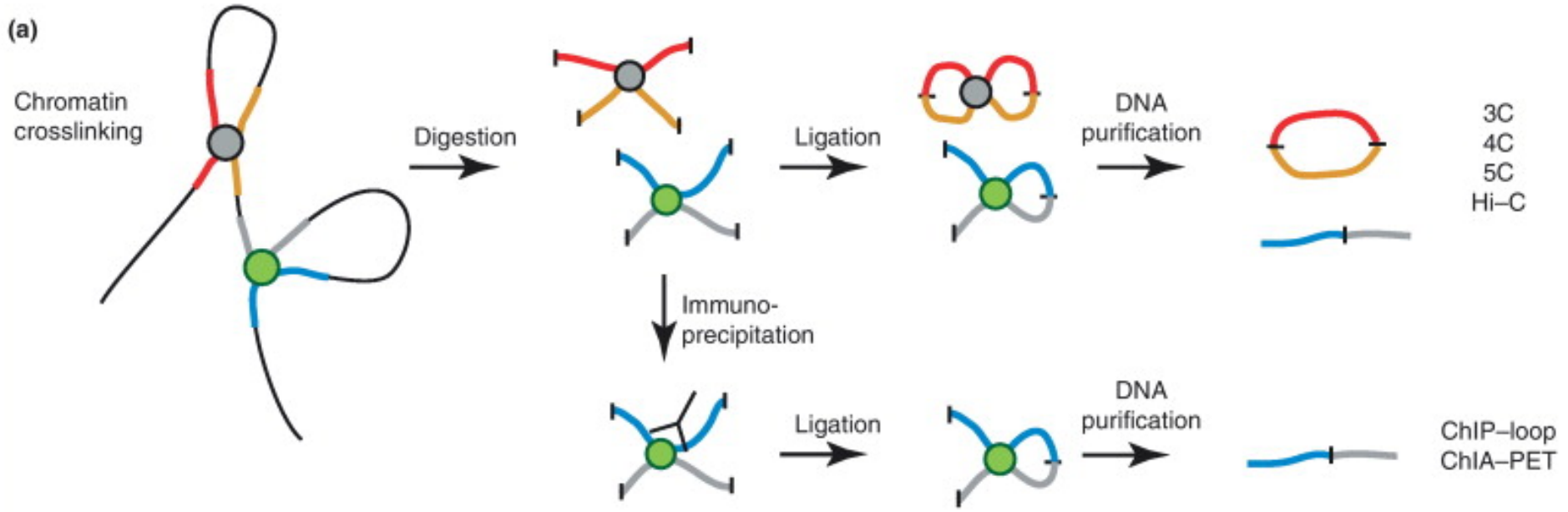
- The “HOXA transcript at the distal tip” HOTTIP is a lincRNA transcribed from the 5’ tip of the HOXA locus that coordinates the activation of multiple 5’HOXA genes i*n vivo*.

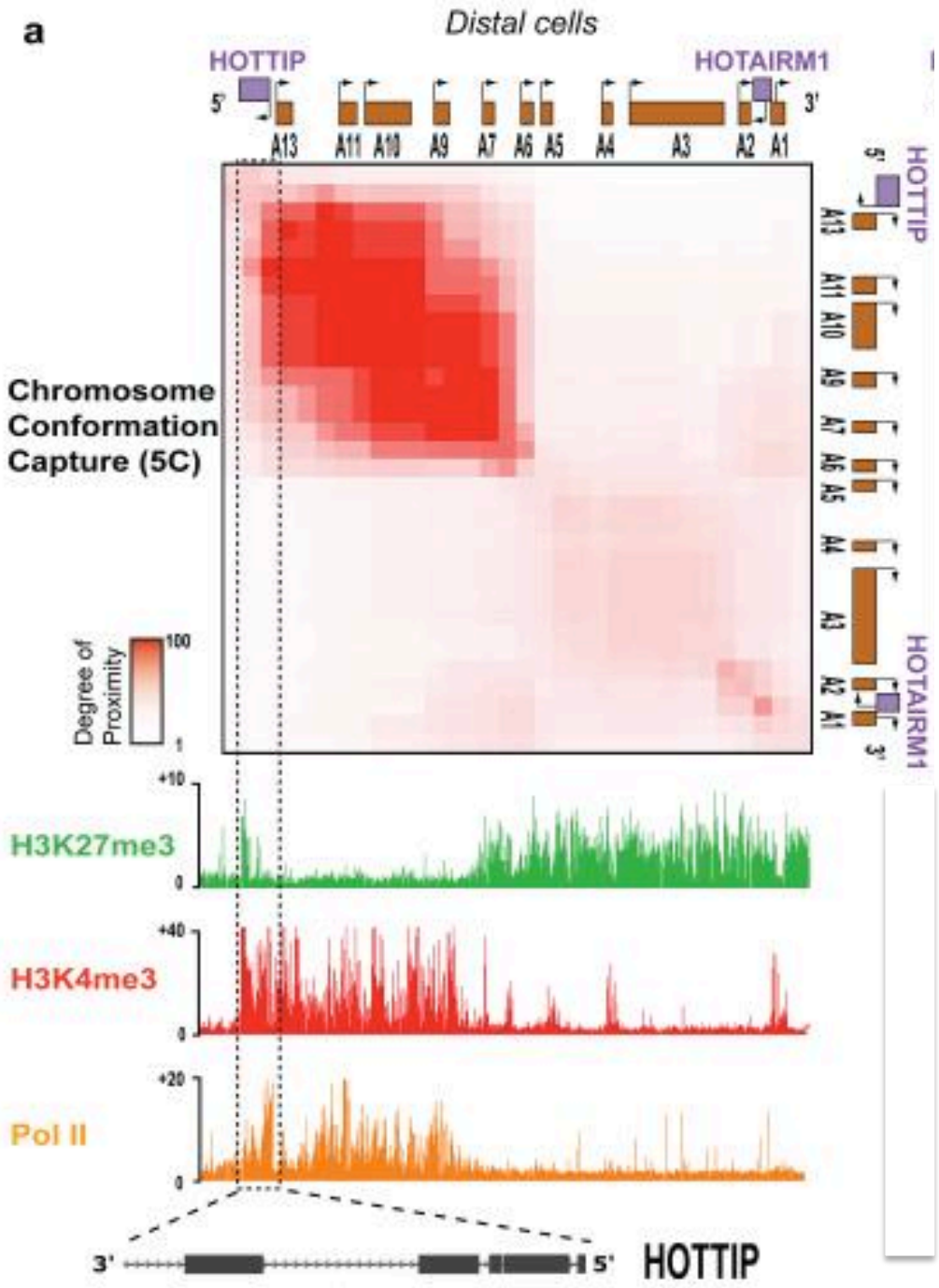
- Chromosomal looping brings HOTTIP in close proximity to the 5’HOXA genes.

- HOTTIP transcription acts as a *switch* to produce HOTTIP lincRNA, which binds to and targets WDR5-MLL complexes to the 5’HOXA locus, yielding a broad domain of H3K4me3 and transcription activation

Chromosome conformation capture (3C) technologies

The library represents the *sum* of DNA–DNA interactions over the cell population





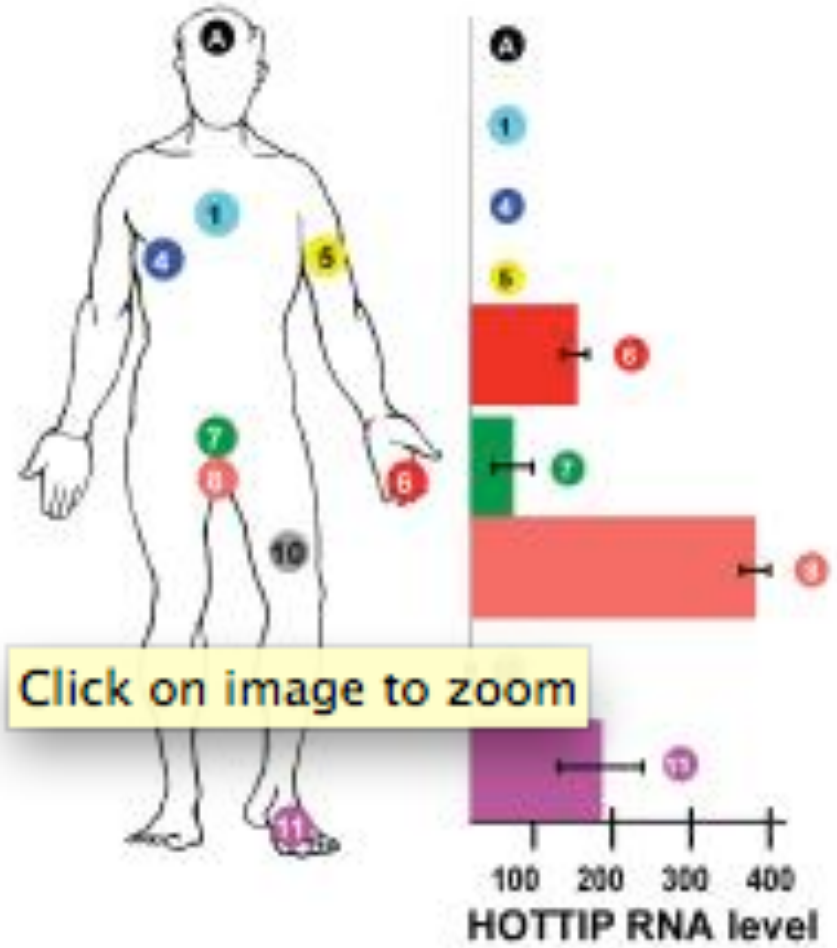
In *distal* cells (e.g. foreskin and foot fibroblasts), abundant chromatin interactions within the transcriptionally active 5' *HOXA* locus.

Proximal cells (e.g. lung fibroblasts) have the diametrically opposite pattern.

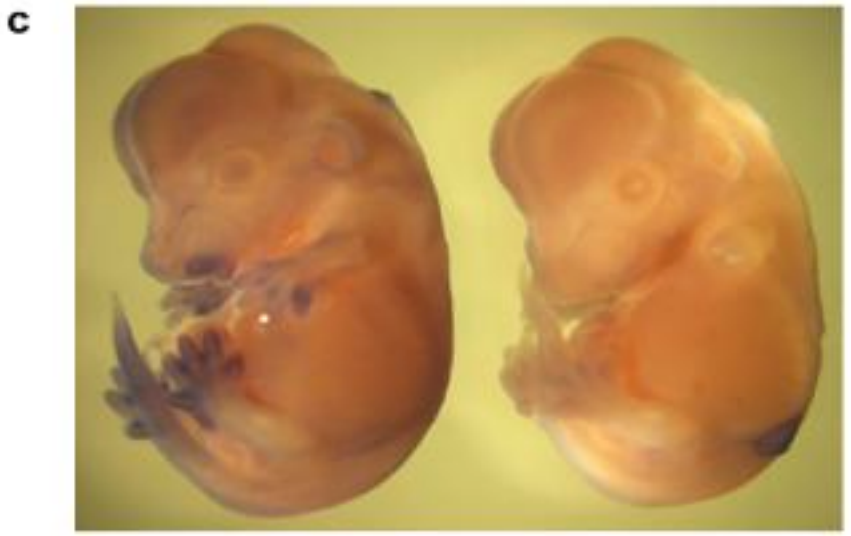
Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation
Nature. Apr 7, 2011; 472(7341): 120–124.

HOTTIP is a lincRNA expressed in distal anatomic sites

Genes near the 5’ end of each HOX cluster tend to be expressed in more posterior and/or distal anatomical locations. Consistent with its genomic location 5’ to HOXA13 , HOTTIP is expressed in distal and/or posterior anatomic sites.



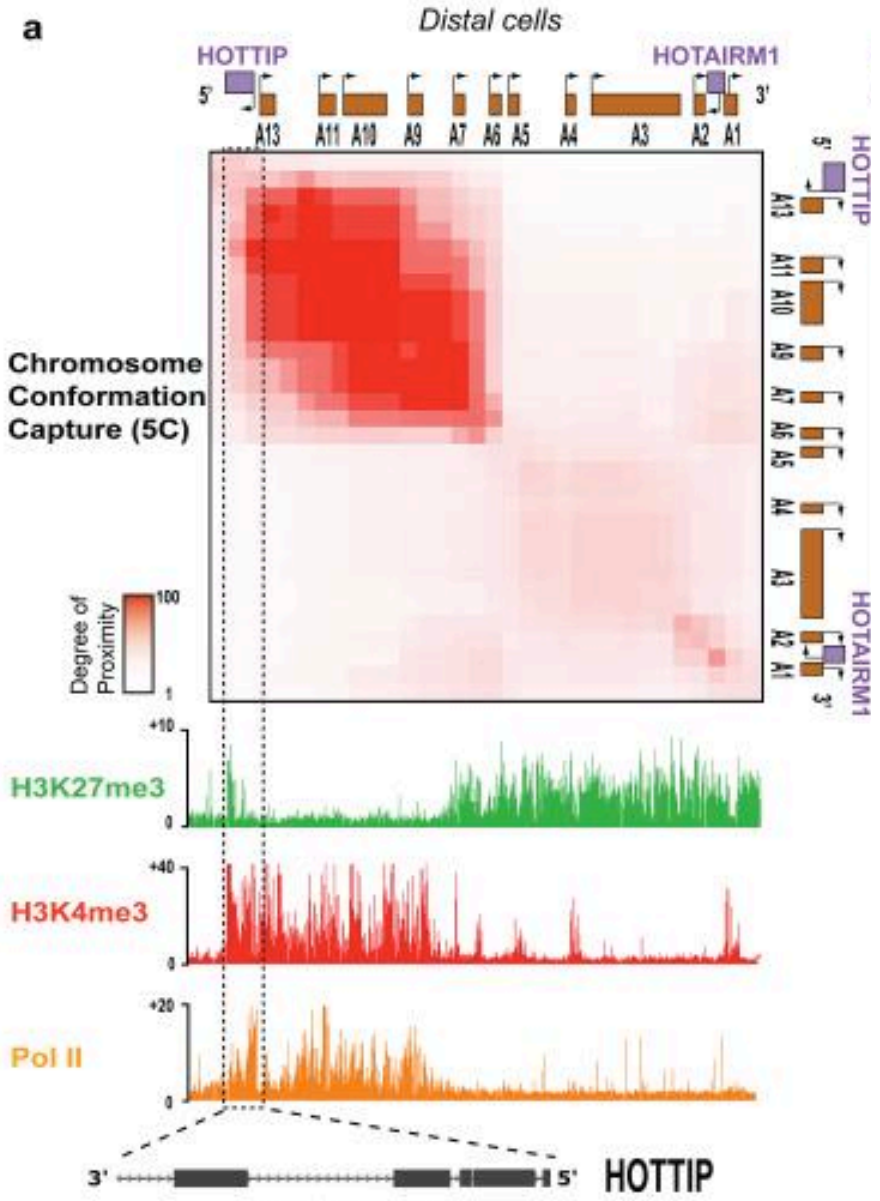
HOTTIP expression in primary human fibroblast from 11 anatomic sites



In situ hybridization
in E13.5 mouse embryo

Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation
Nature. Apr 7, 2011; 472(7341): 120–124.

On the very 5’ and 3’ edges of the two respective interaction clusters are two lincRNA loci that exhibit distinct chromatin modifications.

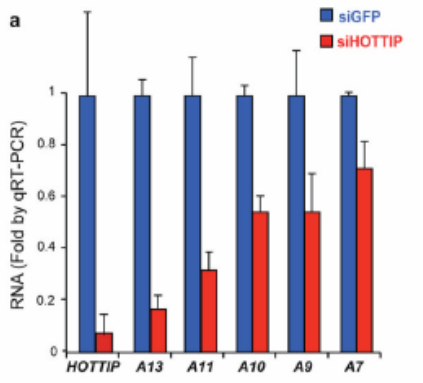


HOTTIP exhibits *bivalent* H3K4me3 and H3K27me3, a histone modification pattern associated with ***poised*** regulatory sequences

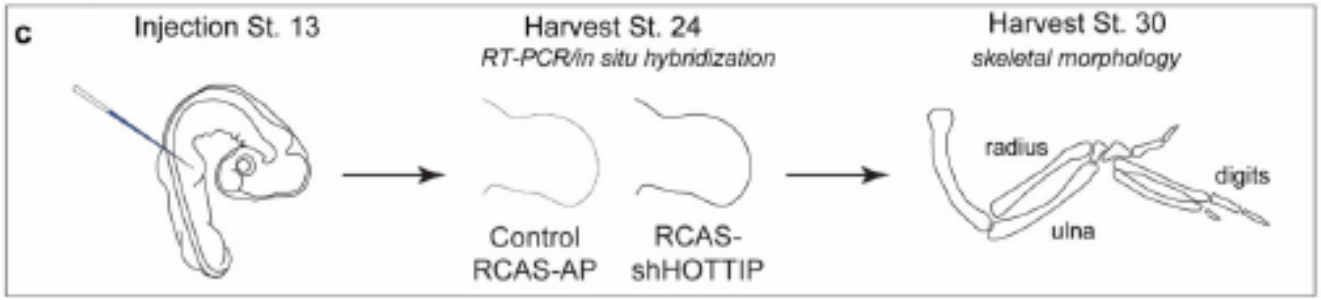
ChIP analysis confirmed that *HOTTIP* is occupied by both PRC2 and MLL complexes, consistent with the bivalent histone marks

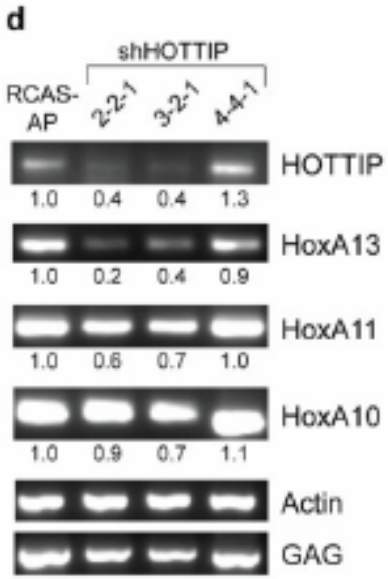
Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation
Nature. Apr 7, 2011; 472(7341): 120–124.

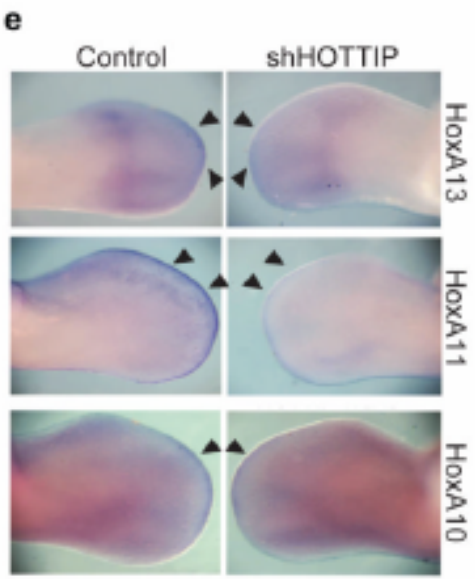
HOTTIP knockdown abrogates expression of distal HOXA genes

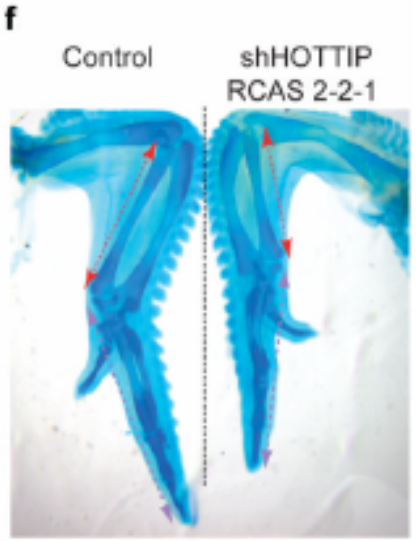


…across 40 kilobases with a trend dependent on the distance to *HOTTIP*.





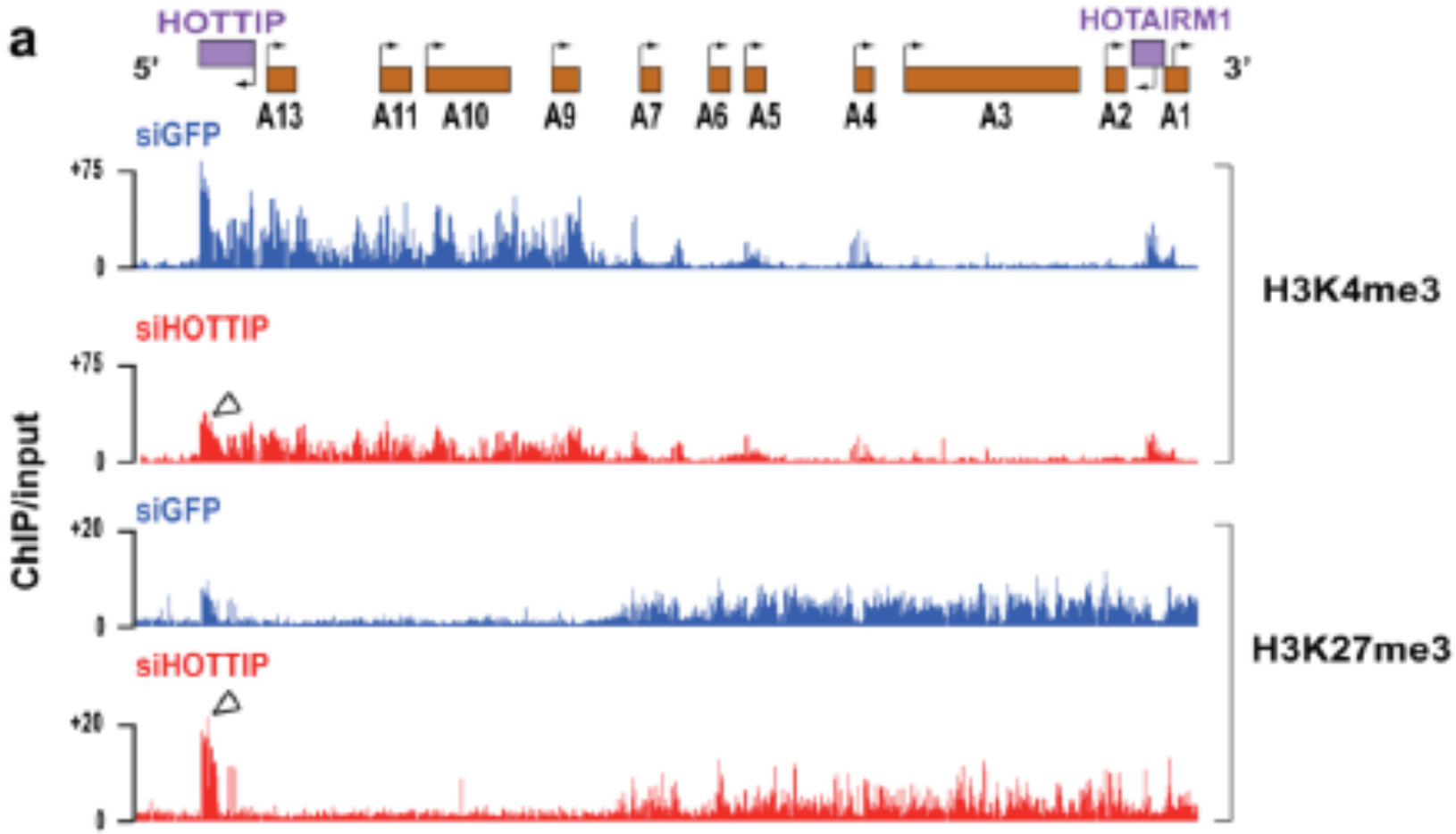




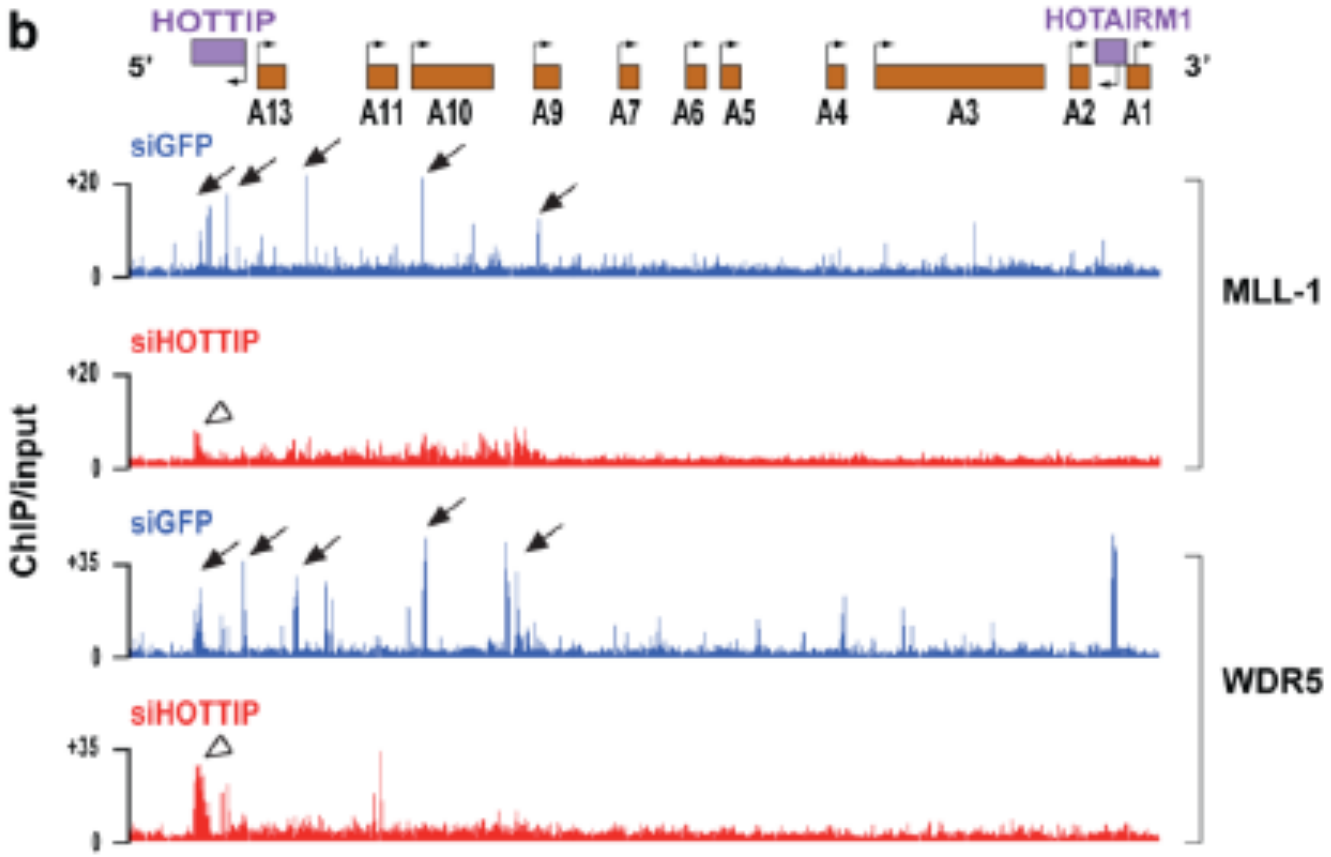
Chick limb

HOTTIP is required for the active chromatin state of 5’ HOXA cluster

Knockdown of HOTTIP broadly decreases H3K4me3 across 5’HOXA locus but focally affects H3K27me3 at HOTTIP itself.



Knockdown of HOTTIP abrogates peaks of MLL1 and WDR5 occupancy near TSSs of 5'HOXA genes and leads to accumulation of these proteins at HOTTIP itself.

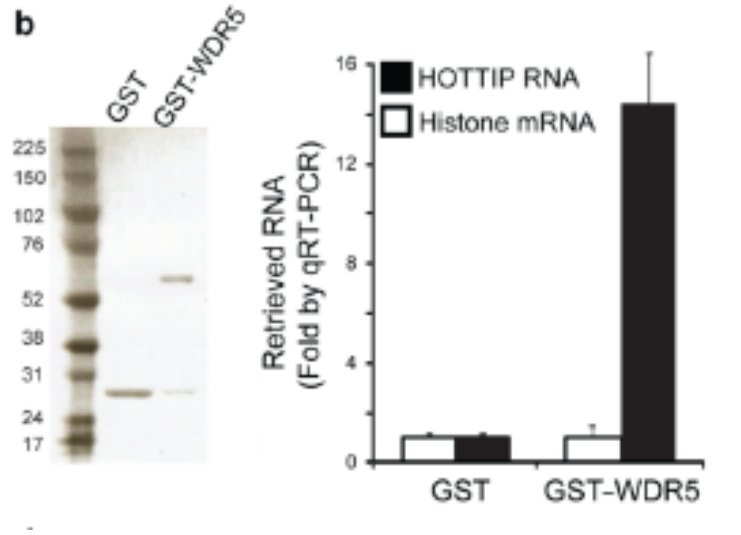


Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation
Nature. Apr 7, 2011; 472(7341): 120–124.

RNA-protein interaction studies

The indicated recombinant protein were purified and used to retrieve purified HOTTIP or control histone RNA in vitro. Only GST-WDR5 specifically retrieved HOTTIP.

	Specific binding to HOTTIP
GST	-
GST-WDR5	+
GST-MLL (C-term)	-
GST-RBBP5/Ash2L	-
GST-TRF1	-



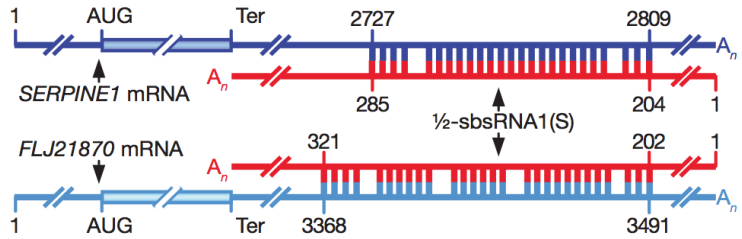
Long noncoding RNA programs active chromatin domain to coordinate homeotic gene activation
Nature. Apr 7, 2011; 472(7341): 120–124.

HOTTIP is critical for maintaining a specific pattern of MLL complex occupancy (*via* WDR5) across the HOXA locus to facilitate H3K4me3 and active transcription.

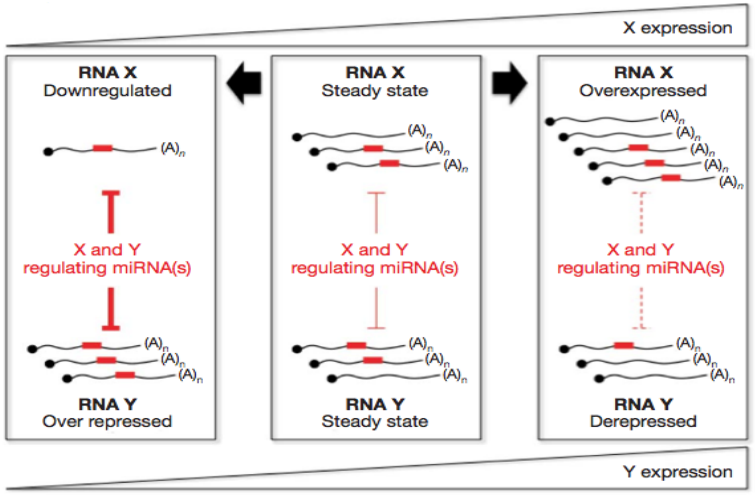
Functions of lncRNAs

Cytoplasmic lncRNAs

lncRNAs can regulate mRNA expression in a sequence-dependent manner



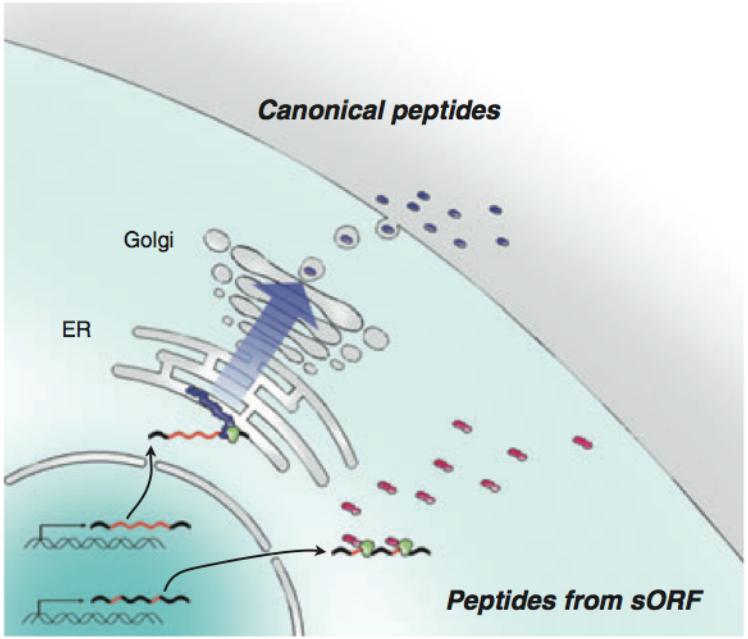
lncRNAs carrying Alu sequences induce mRNA decay by duplexing with their partially complementary 3'UTR and recruiting Staufen-1 on their targets.



Pseudogene non-coding transcripts can sequester microRNAs that target their coding counterparts.
Also mRNAs can act as "sponges" for miRNA level regulation.

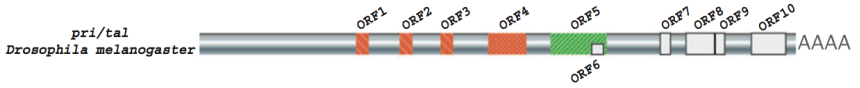
Tumor suppressor PTEN mRNA and its pseudogene-derived transcript PTENP1 are both targeted by miR-19b and miR-20a. PTENP1 is lost in many types of cancer.

Apparently non-coding RNAs are shown to be translated in functional small peptides.

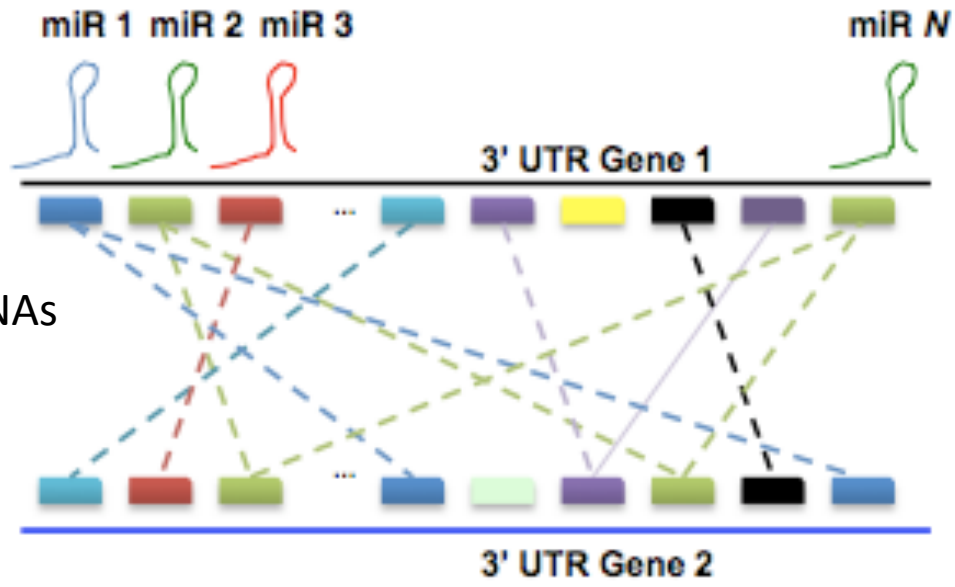


Canonical peptides are produced by processing a long precursor in the ER and are released into the extracellular space *via* trans-Golgi.

sORF (small open reading frame: 11-100aa) encoded small peptides are direclty produced in the cytoplasm in a non-canonical translation process.



Protein coding RNA transcripts can cross talk by competing for common miRNAs



competing endogenous RNAs

ceRNAs

Karreth et al., *Cell* 147, October 14, 2011

Tay et al., *Cell* 147, October 14, 2011

Sumazin et al., *Cell* 147, October 14, 2011

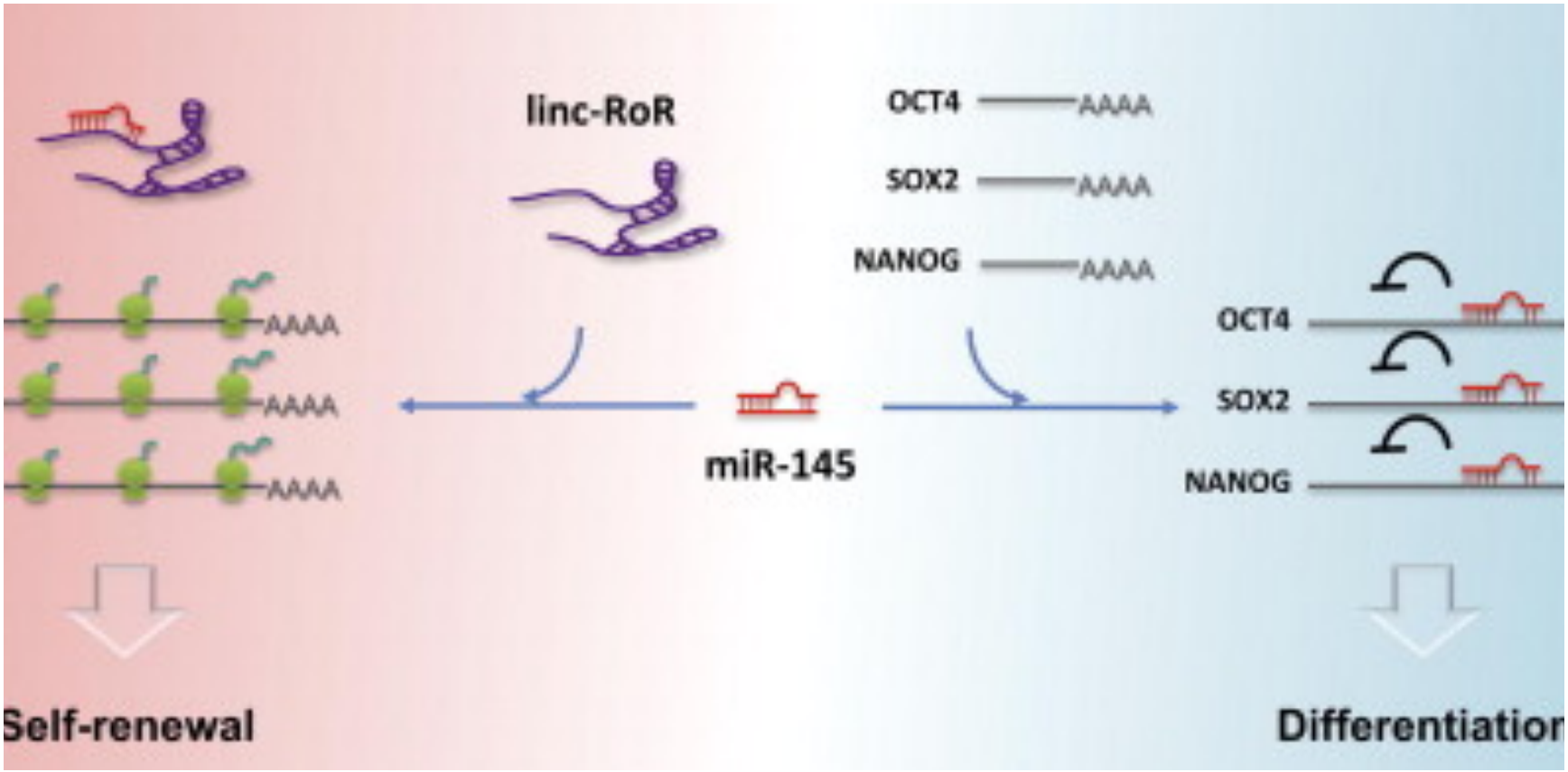
long non coding RNAs can act as ceRNA competing for miRNA binding, thereby modulating the expression of miRNA targets

The muscle-specific long non-coding RNA, linc-MD1, governs the time of muscle differentiation by acting as a ceRNA in mouse and human myoblasts

Endogenous miRNA Sponge lincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal.

Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, Wu M, Xiong J, Guo X, Liu H.

Dev Cell. 2013 25:69-80.

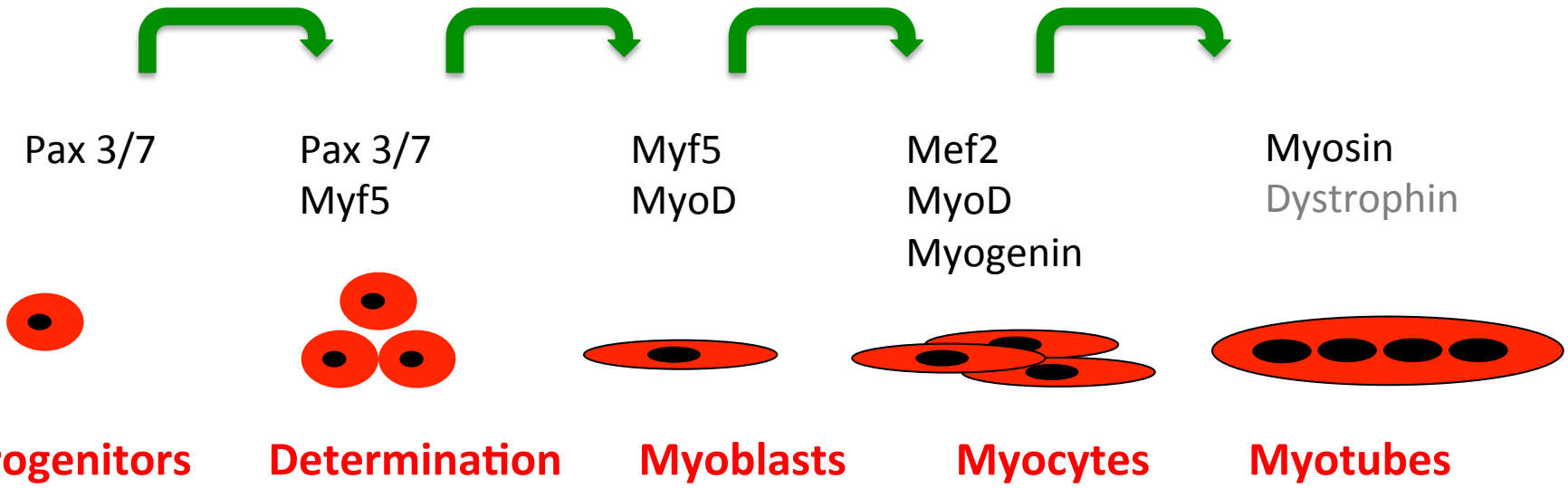


A Competition for miR-145 between linc-RoR and mRNAs Encoding the Core TFs. The presence of linc-RoR in hESCs traps miR-145, preventing it from repressing the translation of the core pluripotency factors and ensuring the stem cell fate. The disappearance of linc-RoR in differentiating hESCs releases miR-145, allowing it to repress the translation of core pluripotency factors.

Function of non coding RNA in normal versus pathological conditions

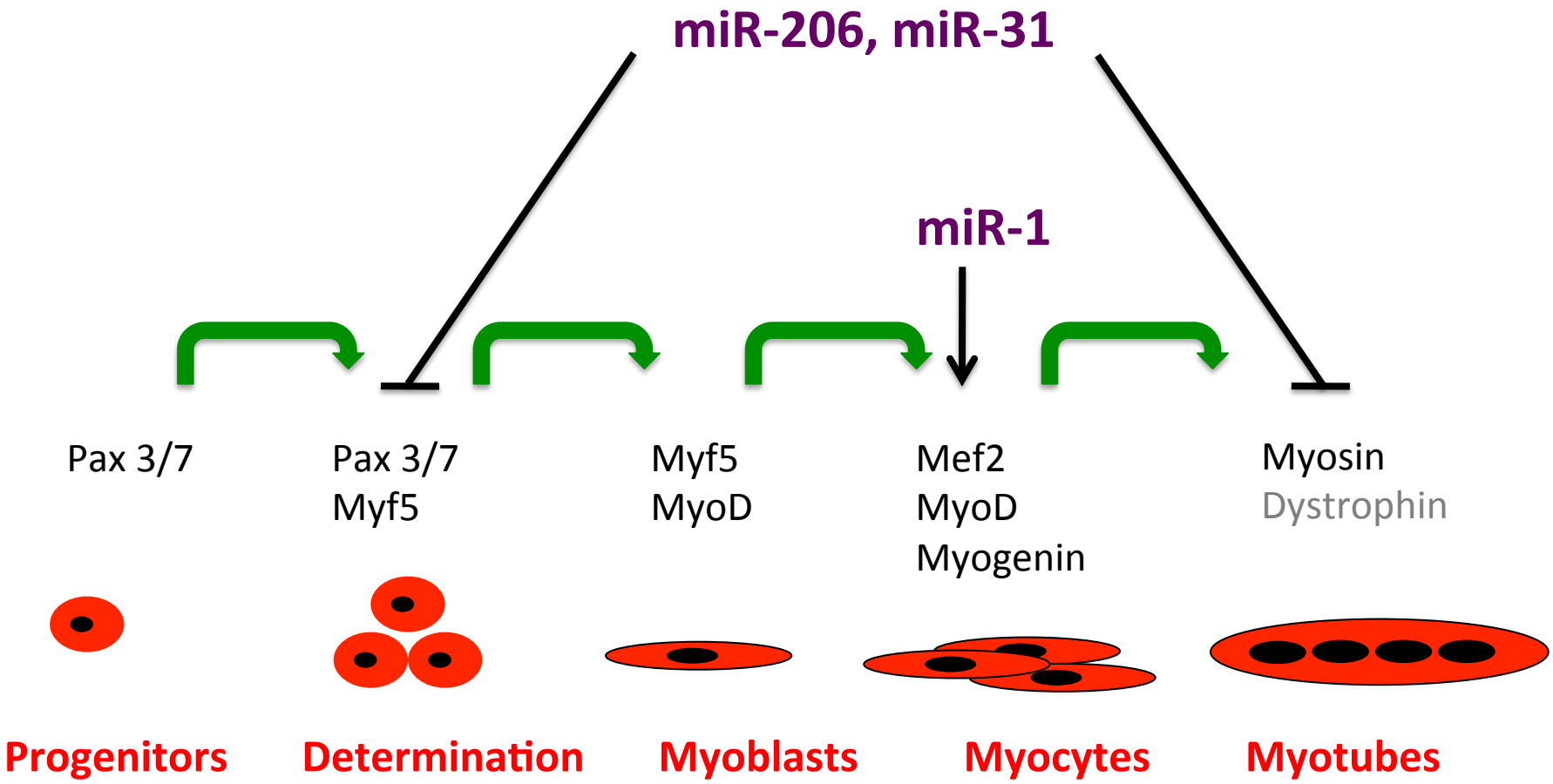
……long non coding RNAs in differentiation

Muscle differentiation



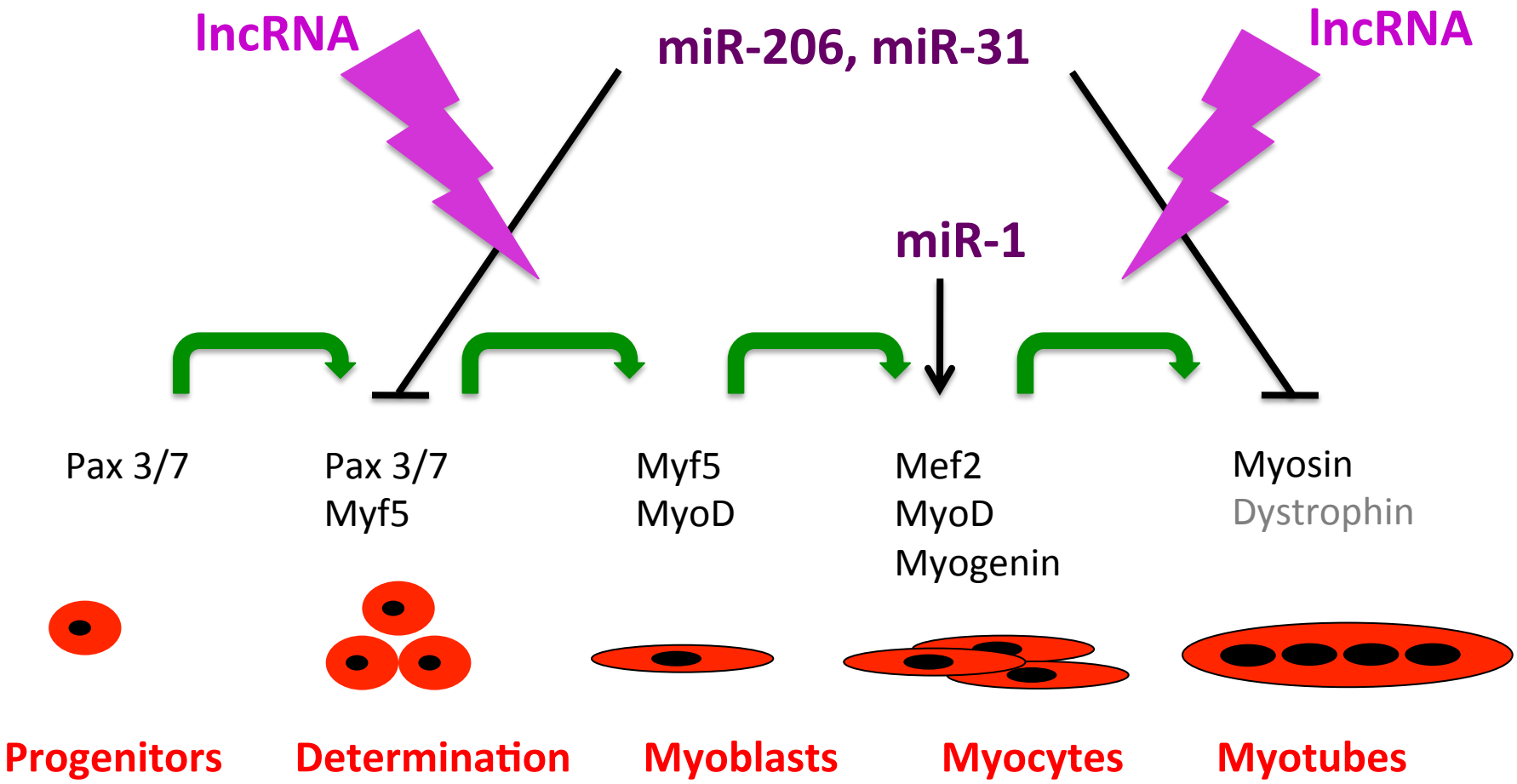
Transcriptional factors as master controllers of cell differentiation

Muscle differentiation



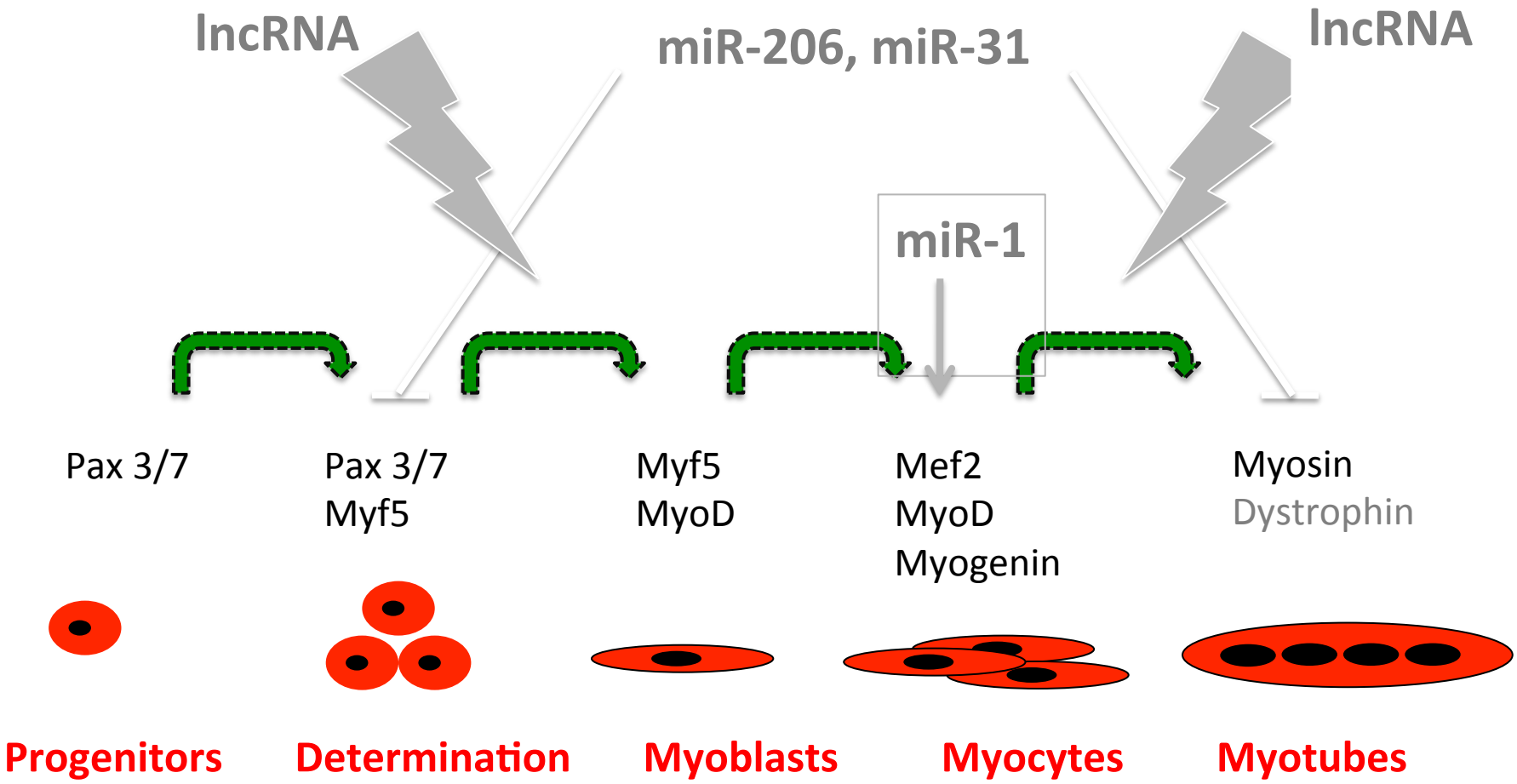
miRNAs as master controllers of cell differentiation

Muscle differentiation



long non-coding RNAs as master controllers of cell differentiation

Muscle differentiation



Deregulation of ncRNAs can lead to pathology

Cell. Oct 14, 2011; 147(2): 358–369.
doi: 10.1016/j.cell.2011.09.028
PMCID: PMC3234495

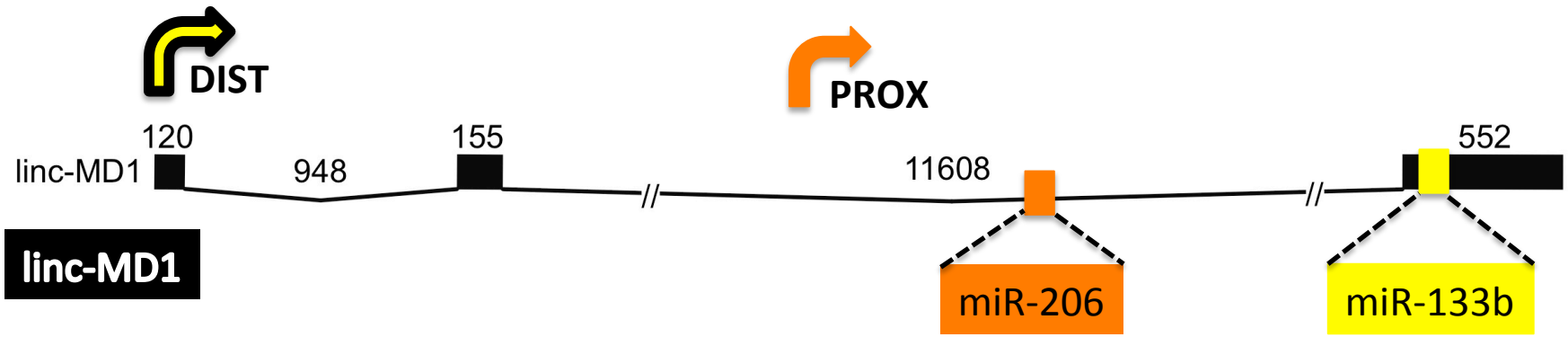
A Long Noncoding RNA Controls Muscle Differentiation by Functioning as a Competing Endogenous RNA

Marcella Cesana,[1,6] Davide Cacchiarelli,[1,6] Ivano Legnini,[1] Tiziana Santini,[1] Olga Sthandier,[1] Mauro Chinappi,[2] Anna Tramontano,[2,3,4] and Irene Bozzoni[1,3,4,5,*]

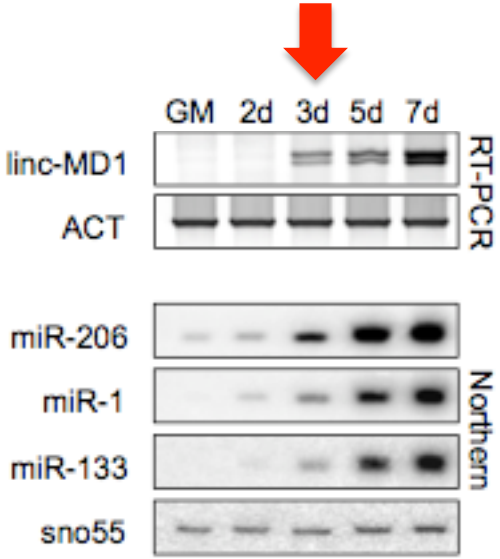
Author information ► Article notes ► Copyright and License information ►

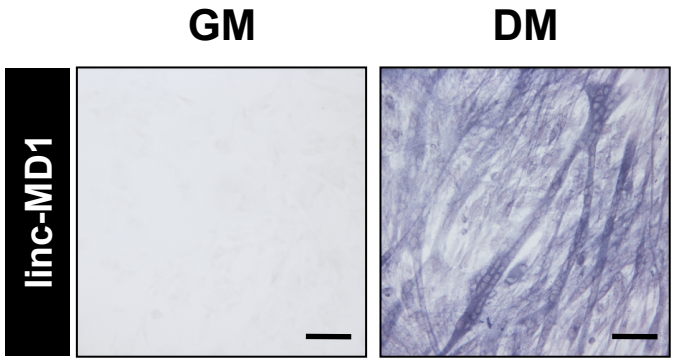
Recently, a new regulatory circuitry has been identified in which RNAs can crosstalk with each other by competing for shared microRNAs. Such competing endogenous RNAs (ceRNAs) regulate the distribution of miRNA molecules on their targets and thereby impose an additional level of post-transcriptional regulation. Here we identify a muscle-specific long noncoding RNA, linc-MD1, which governs the time of muscle differentiation by acting as a ceRNA in mouse and human myoblasts. Downregulation or overexpression of linc-MD1 correlate with retardation or anticipation of the muscle differentiation program, respectively. We show that linc-MD1 “sponges” miR-133 and miR-135 to regulate the expression of MAML1 and MEF2C, transcription factors that activate muscle-specific gene expression. Finally, we demonstrate that linc-MD1 exerts the same control over differentiation timing in human myoblasts, and that its levels are strongly reduced in Duchenne muscle cells. We conclude that the ceRNA network plays an important role in muscle differentiation.

linc-MD1 is a cytoplasmic polyA+ long non-coding RNA



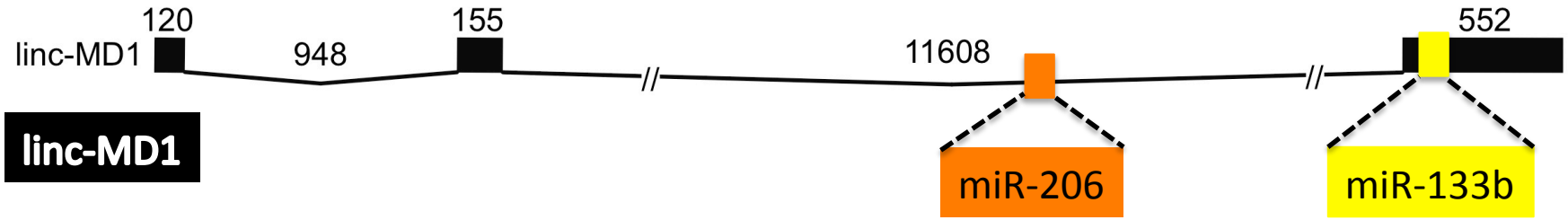
linc-MD1 is expressed in early phases of differentiation from the DIST promoter

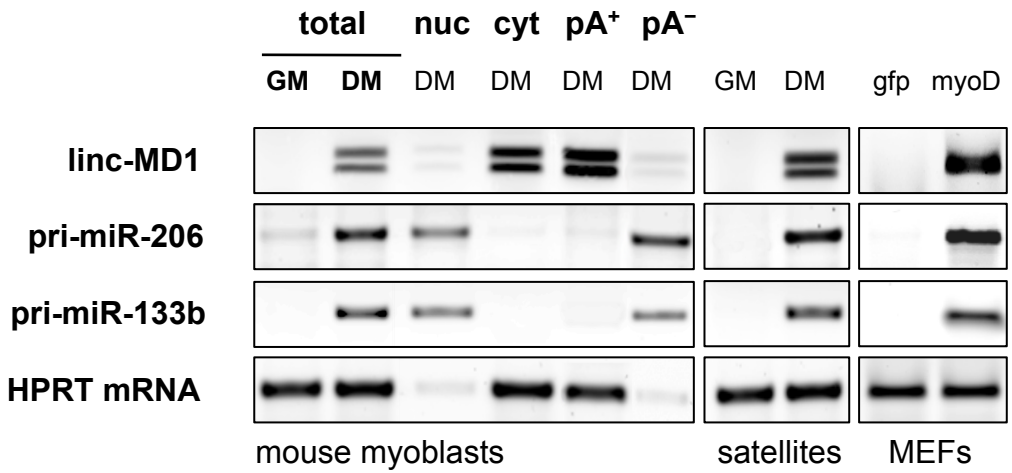




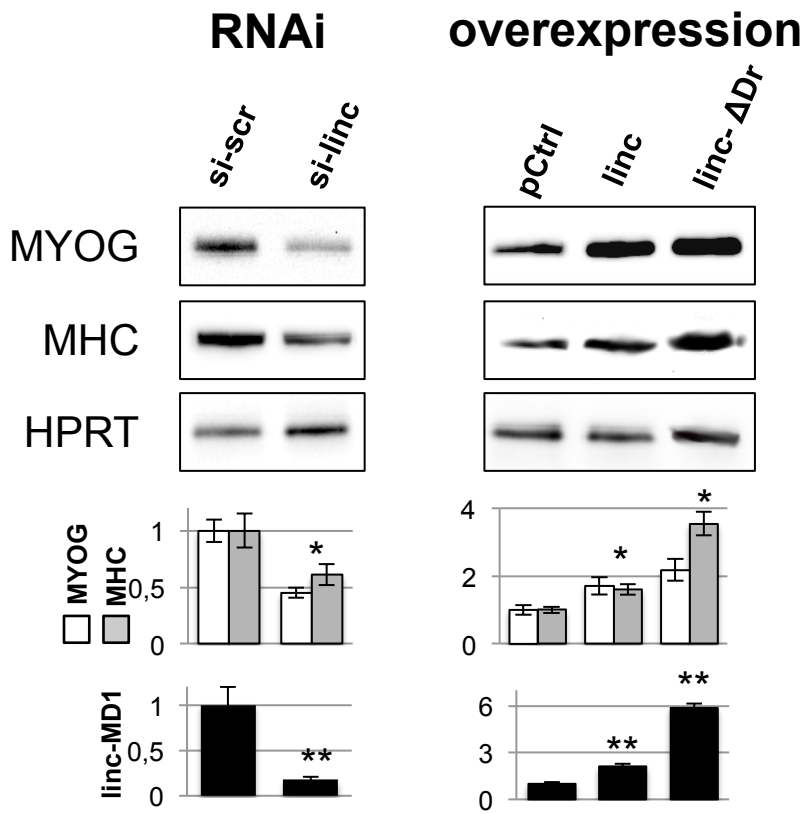
C2 myoblasts

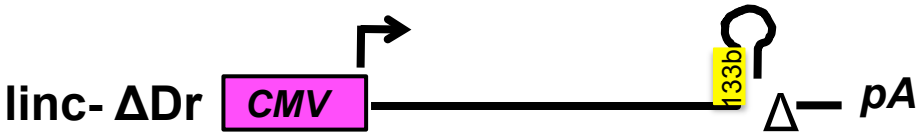
linc-MD1 is a cytoplasmic polyA+ long non-coding RNA



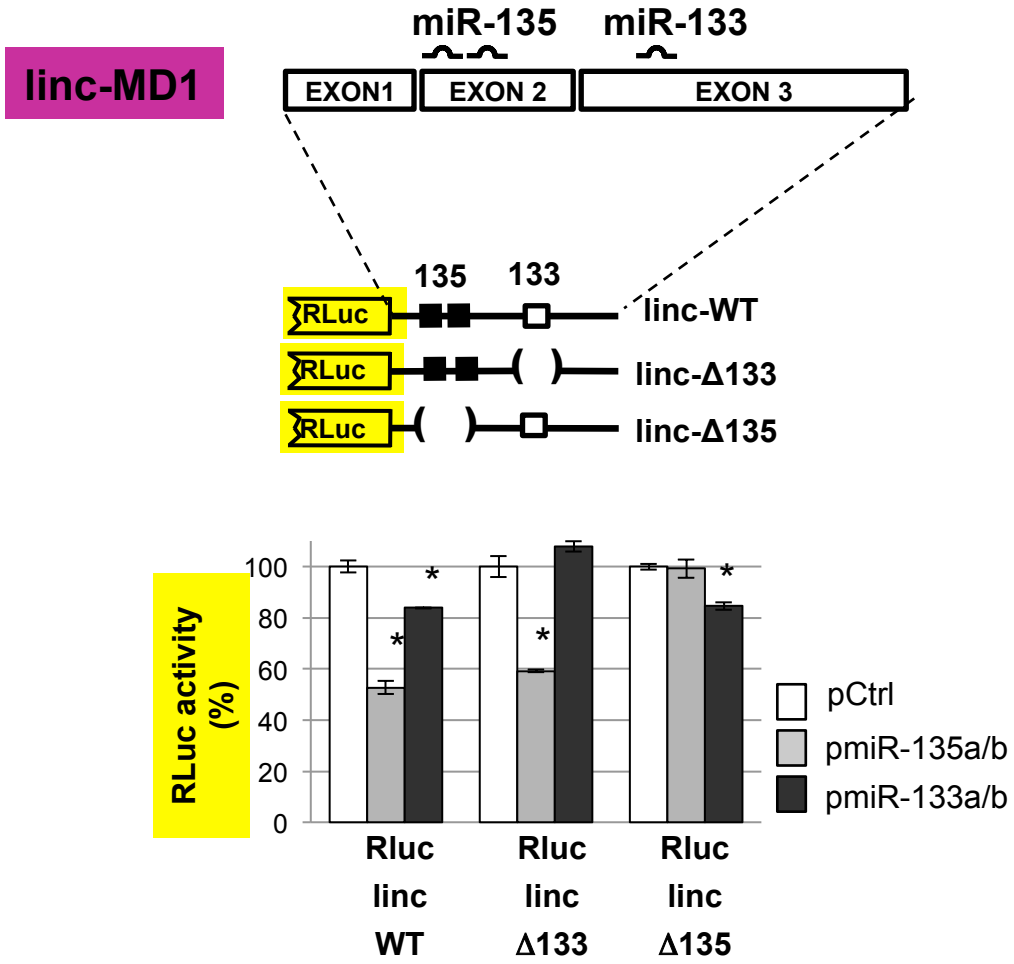


linc-MD1 levels affect myoblasts differentiation

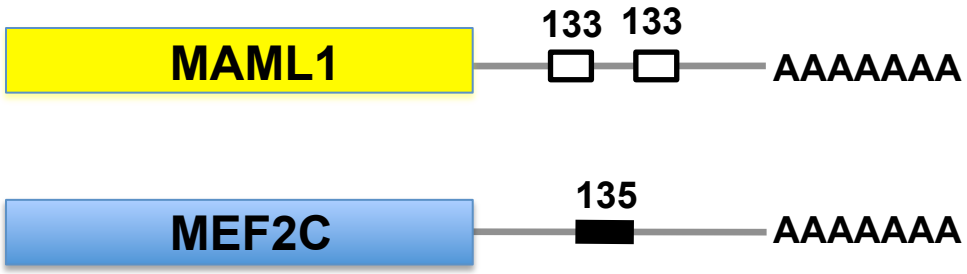




linc-MD1 is target of miR-135 and miR-133



The 3’ UTRs of the myogenic factors MAML1 and MEF2C contain miR-135 and miR-133 binding sites



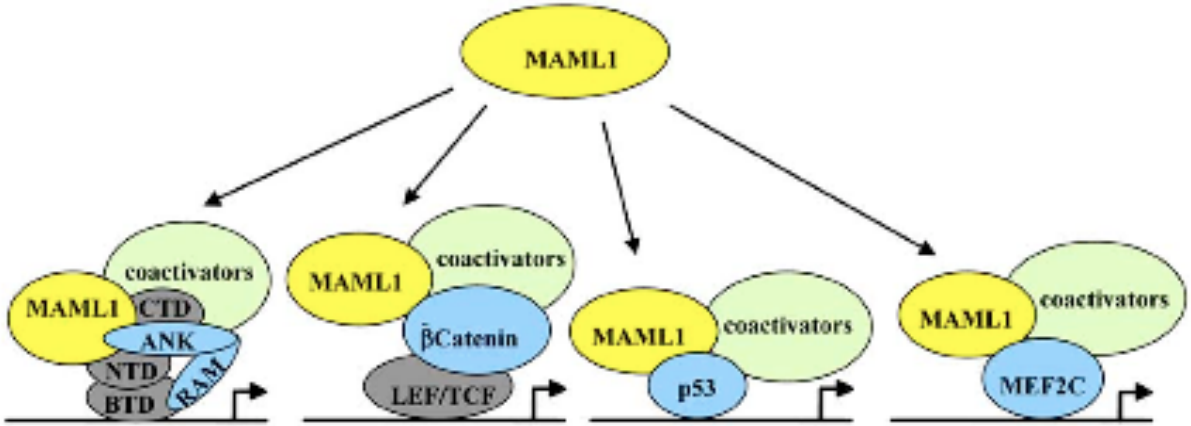
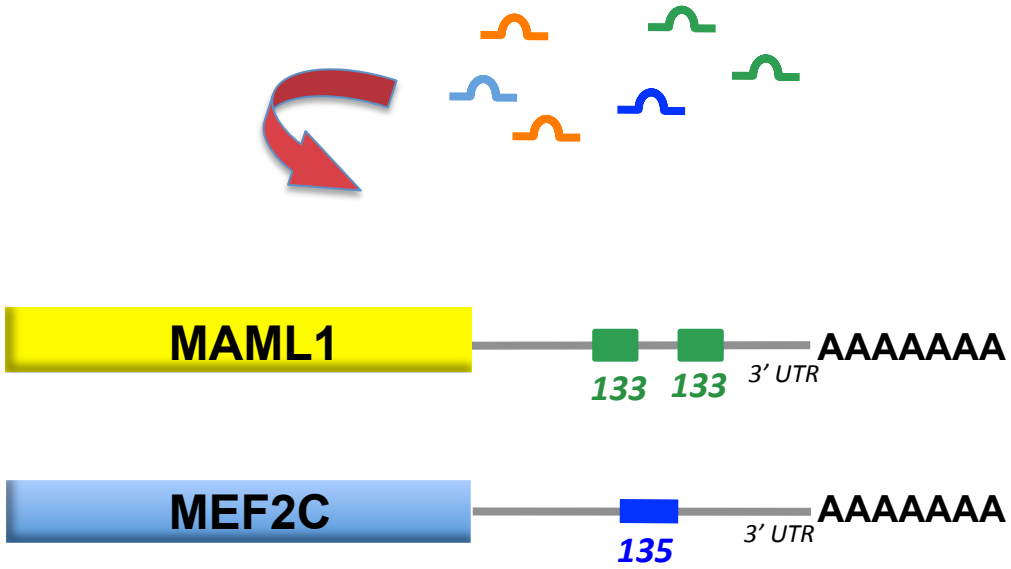


Fig. (1). The MAML1 protein acts as a coactivator for diverse transcription factors. MAML1 is recruited by the ANK (ankyrin repeats) domain of Notch and forms a complex with the DNA-binding protein CSL (composed of NTD (N-terminal domain), CTD (C-terminal domain), and BTD (β-trefoil domain)). The RAM (RBP-Jk associated molecule) domain of Notch interacts with the BTD of CSL. In addition, MAML1 is recruited by β-catenin, p53 and MEF2C to regulate various signalling pathways. Most likely, additional coactivators are working cooperatively with MAML1 in gene regulation.

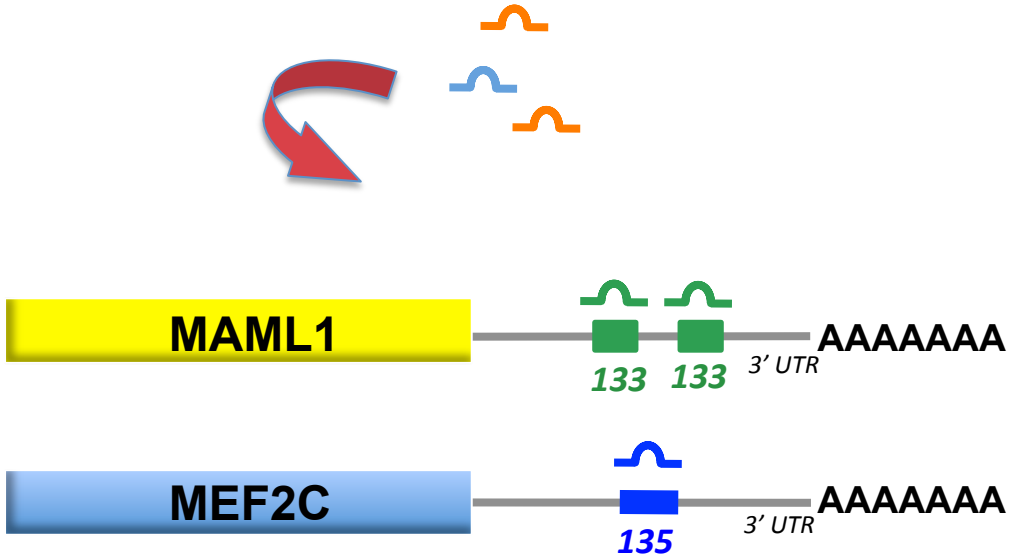
Competition for miRNA binding



MAML1 and MEF2C are myogenic factors controlling muscle differentiation

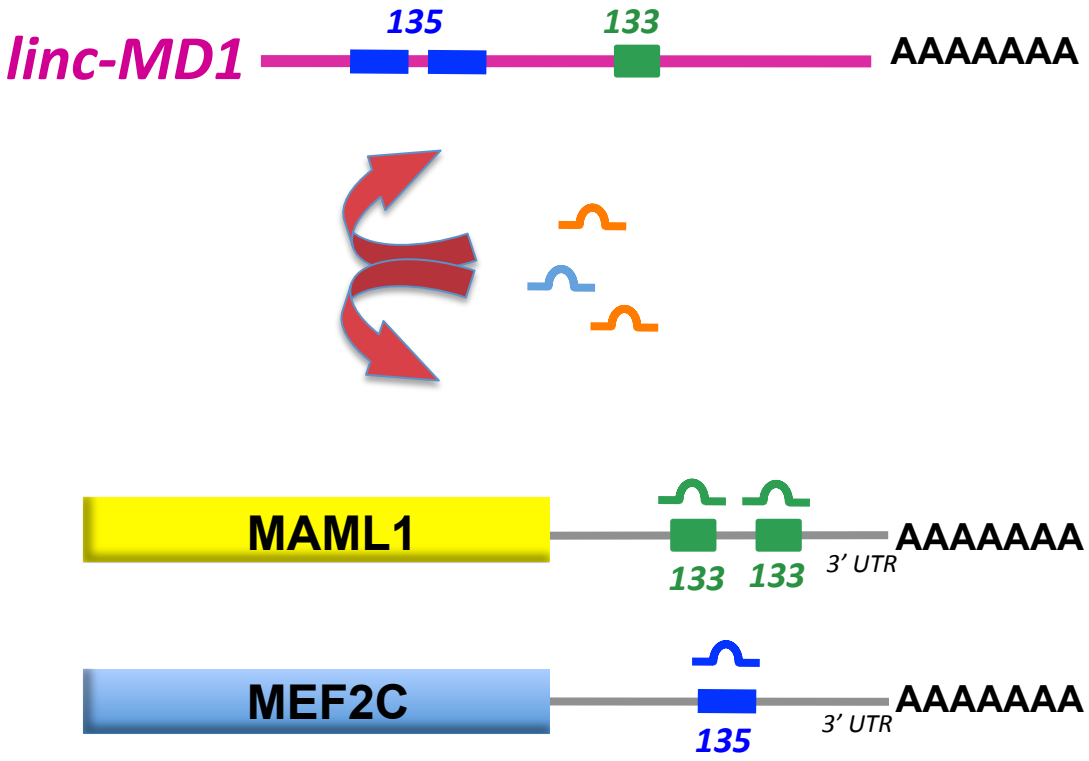
Competition for miRNA binding

- Maml1 and Mef2C are target of miR-135 and miR-133



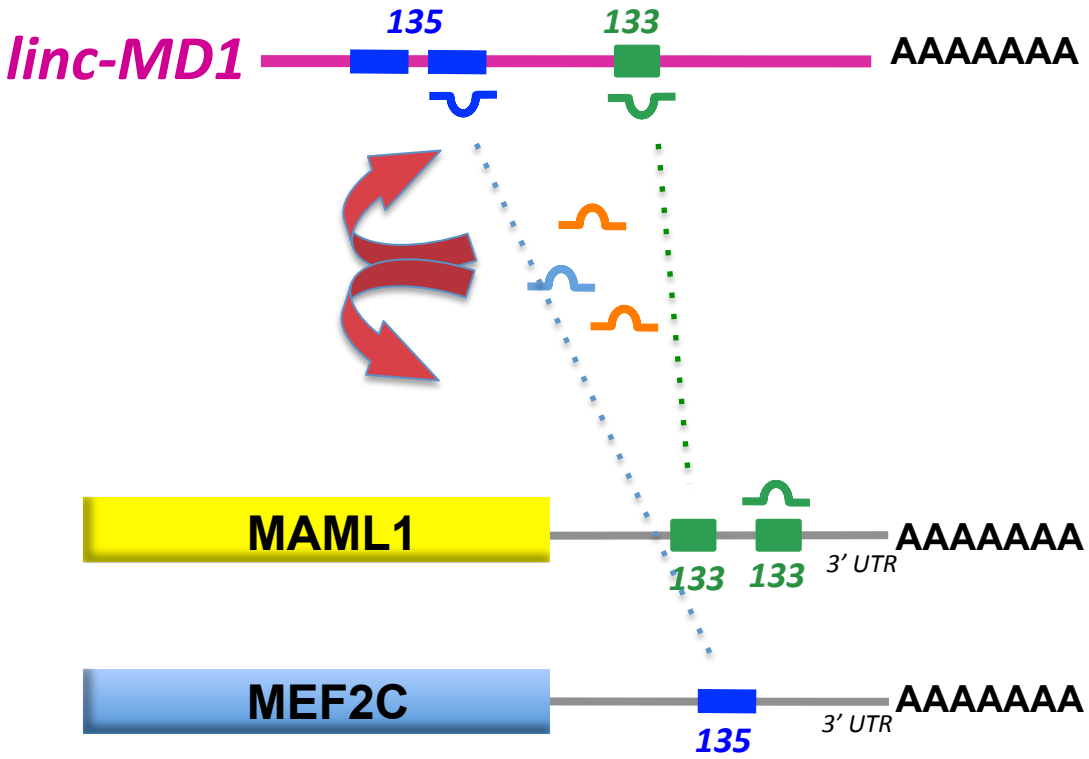
miR-133 and miR-135 repress MAML1 and MEF2C

Competition for miRNA binding



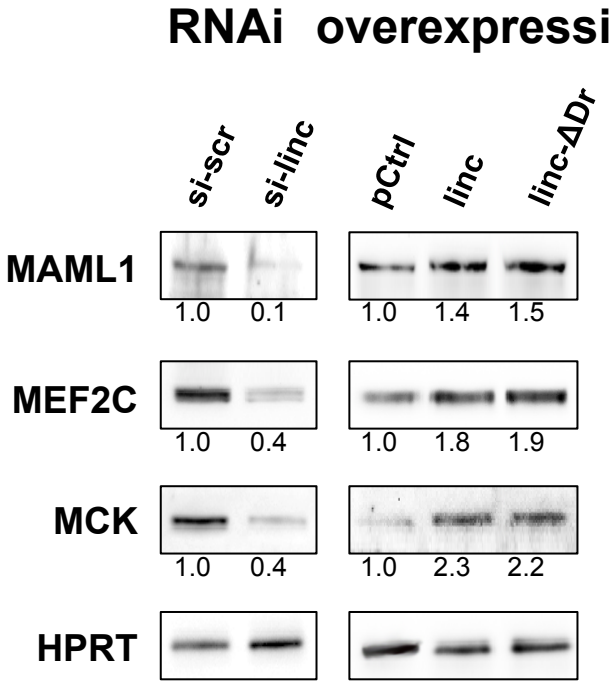
when linc-MD1 is expressed……

Competition for miRNA binding

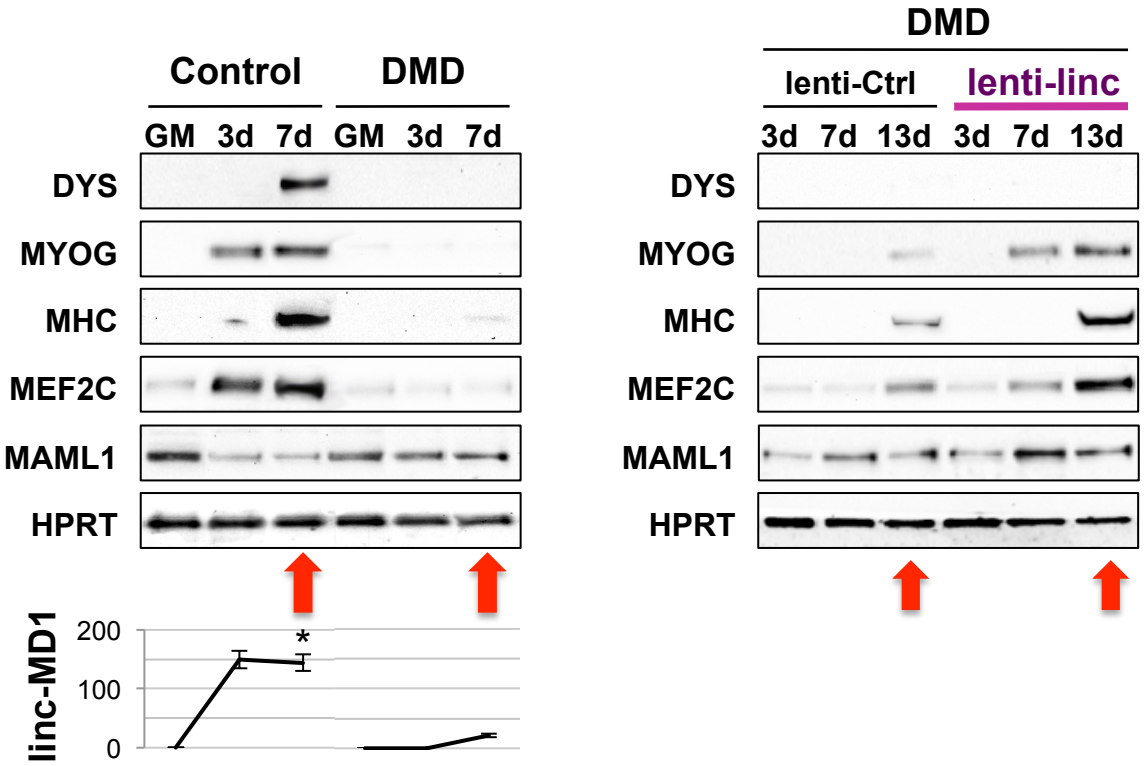


…it allows the derepression of Maml1 and Mef2C synthesis

linc-MD1 controls the levels of MAML1 and MEF2C

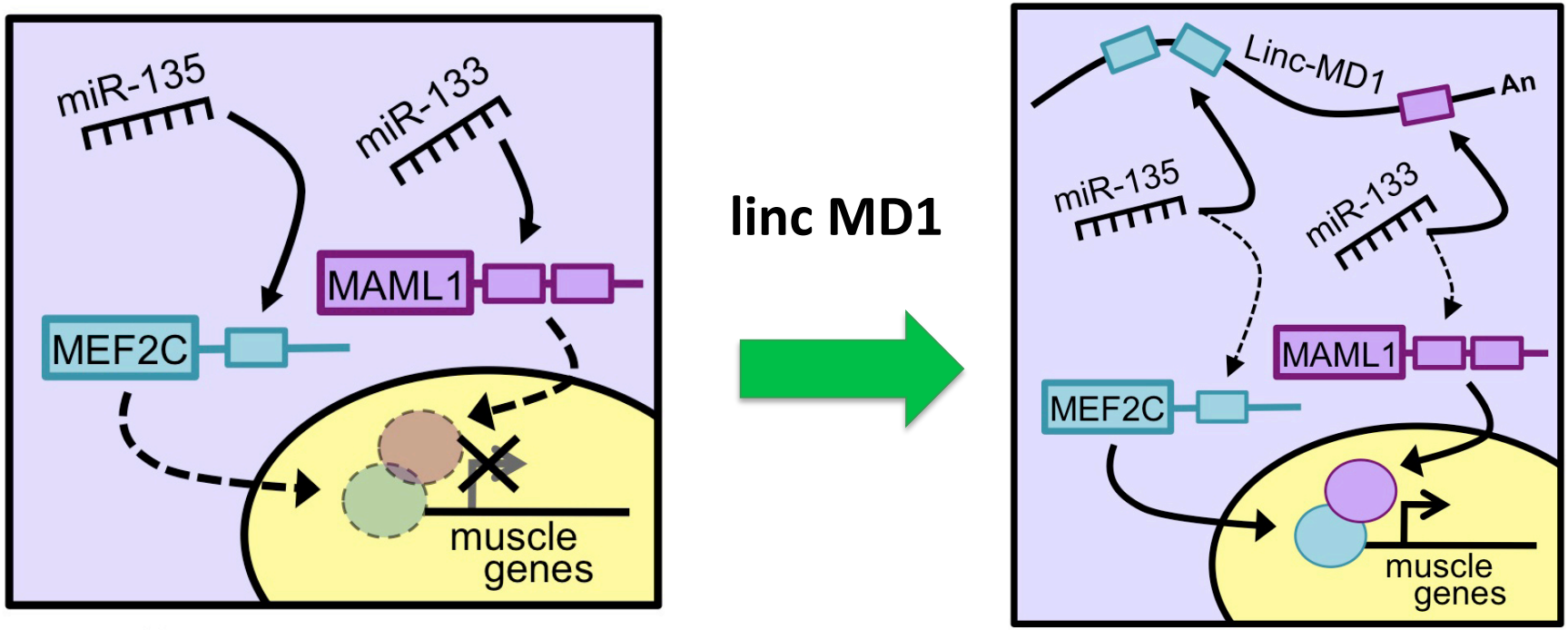


- linc-MD1 controls differentiation of human myoblasts
- linc-MD1 is down-regulated in Duchenne myoblasts



- low levels of linc-MD1 account for the delay in differentiation of DMD myoblasts
- rescue of linc-MD1 restores an almost wt differentiation timing

linc-MD1 acts as a natural decoy (sponge) for specific miRNAs



differentiation

Cesana et al., *Cell* 147, *358-369*, 2011

A Long ncRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy

Daphne S. Cabianca,[1] Valentina Casa,[1,2] Beatrice Bodega,[3,5] Alexandros Xynos,[1] Enrico Ginelli,[3] Yujiro Tanaka,[4] and Davide Gabellini[1,*]

[1]Dulbecco Telethon Institute at San Raffaele Scientific Institute, Division of Regenerative Medicine, Stem Cells, and Gene Therapy, 20132 Milan, Italy

[2]Università Vita-Salute San Raffaele, 20132 Milan, Italy

[3]Department of Biology and Genetics for Medical Sciences, University of Milan, 20133 Milan, Italy

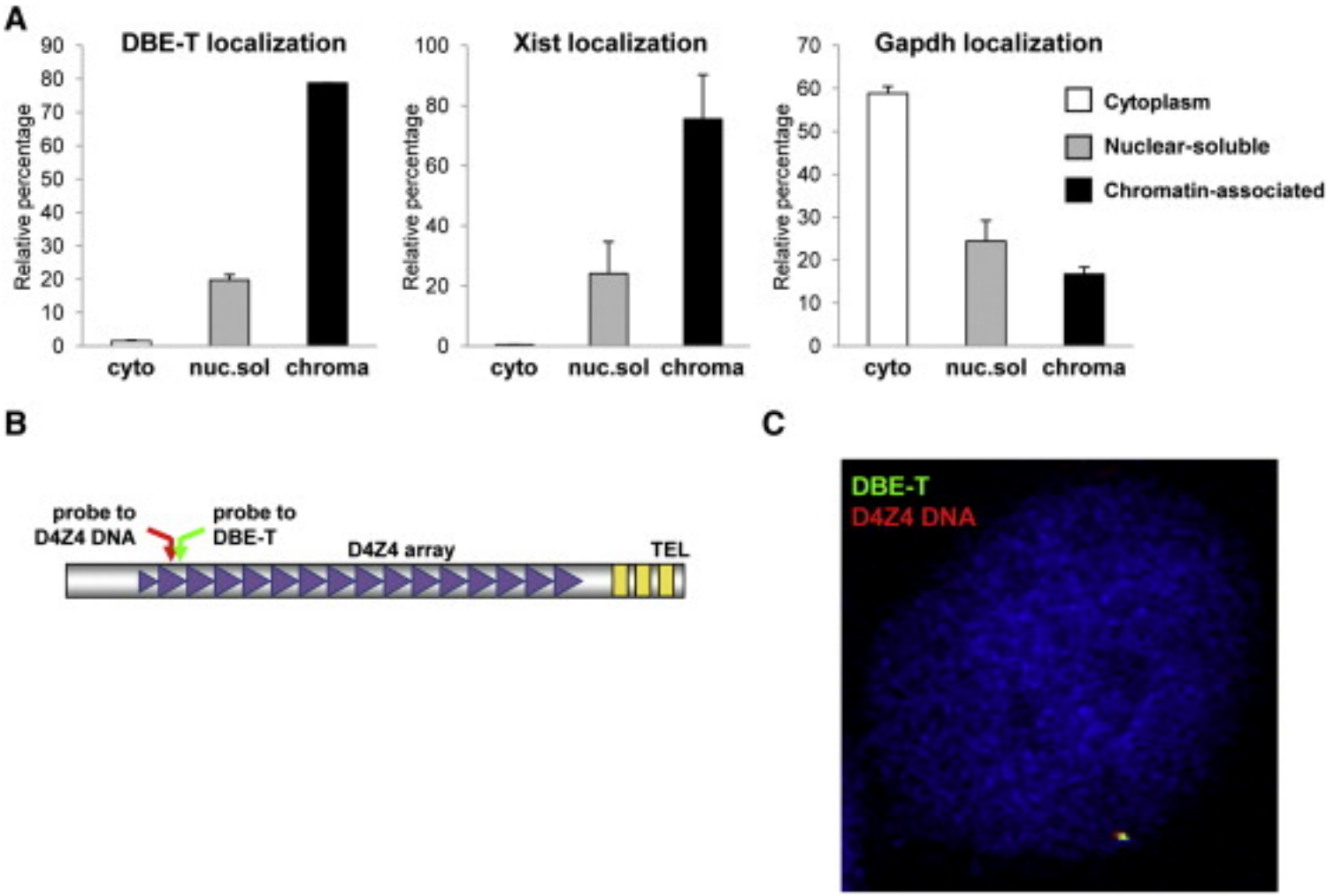
[4]Genome Structure and Regulation, School of Biomedical Science and Biochemical Genetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

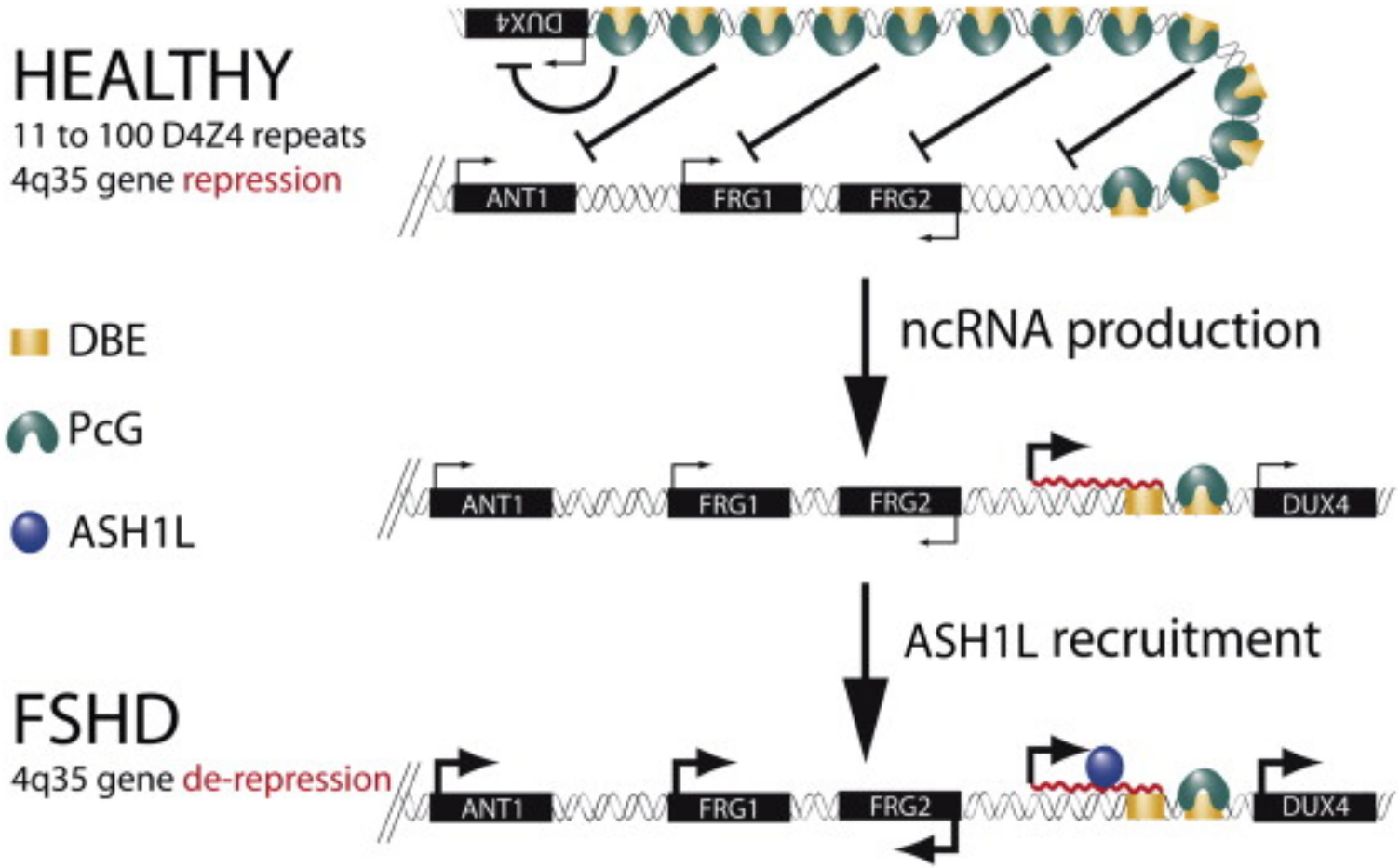
[5]Present address: Dulbecco Telethon Institute at Fondazione Santa Lucia, 00143 Rome, Italy

*Correspondence: gabellini.davide@hsr.it

DOI 10.1016/j.cell.2012.03.035

Repetitive sequences account for more than 50% of the human genome. Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal-dominant disease associated with reduction in the copy number of the D4Z4 repeat mapping to 4q35. By an unknown mechanism, D4Z4 deletion causes an epigenetic switch leading to de-repression of 4q35 genes. Here we show that the Polycomb group of epigenetic repressors targets D4Z4 in healthy subjects and that D4Z4 deletion is associated with reduced Polycomb silencing in FSHD patients. We identify DBE-T, a chromatin-associated noncoding RNA produced selectively in FSHD patients that coordinates de-repression of 4q35 genes. DBE-T recruits the Trithorax group protein Ash1L to the FSHD locus, driving histone H3 lysine 36 dimethylation, chromatin remodeling, and 4q35 gene transcription. This study provides insights into the biological function of repetitive sequences in regulating gene expression and shows how mutations of such elements can influence the progression of a human genetic disease.


A
DBE-T localization
Relative percentage
90
80
70
60
50
40
30
20
10
0
cyto
nuc.sol
chroma
Xist localization
Relative percentage
100
80
60
40
20
0
cyto
nuc.sol
chroma
Gapdh localization
Relative percentage
70
60
50
40
30
20
10
0
cyto
nuc.sol
chroma
Cytoplasm
Nuclear-soluble
Chromatin-associated
B
probe to
D4Z4 DNA
probe to
DBE-T
D4Z4 array
TEL
C
DBE-T
D4Z4 DNA



Model for DBE-T-Mediated 4q35 Gene De-Repression in FSHD

Normal individuals carry multiple D4Z4 copies that are extensively bound by PcG proteins, promoting the maintenance of repressed chromatin at 4q35.

In FSHD patients, D4Z4 deletion leads to insufficient binding of PcG, causing the production of DBE-T. DBE-T recruits the TrxG protein ASH1L and promotes a topological reorganization of the FSHD locus leading to de-repression of 4q35 genes.

The various mechanisms through which lncRNAs function during disease

Transcript	Target	Mechanism	Disease Relevance	Reference
GAS5	*Glucocorticoid receptor (GR)*	Decoy	Cancer, Autoimmune disease	28,29
APOA1-AS	*APOA1*	Scaffold, Guide	Cardiovascular disease	83
BDNF-AS	*BDNF*	Scaffold, Guide	Neurodegenerative disorders	38
DLEU1, DLEU2	*13q14.3 gene locus*	Scaffold, Guide	Leukemia	73
H19	*E-cadherin*	Scaffold, Guide	Cancers	75–77
HOTAIR	*HOXD 8-11, 13*	Scaffold, Guide	Cancers	56–68
KCNQ1OT1	*KCNQ1*	Scaffold, Guide	Beckwith-Wiedemann Syndrome, adrenal neuroblastoma	43,44
ANRIL	*Ink4b/ARF/INK4a*	Scaffold, Guide	Cancers	69,70
PCAT-1	*BRCA2, CENPF, CENPE*	Scaffold, Guide	Prostate cancer	71
PINT	p53 pathway genes	Scaffold, Guide	Cancers	78
*PTENpg1-*α isoform	*PTEN*	Scaffold, Guide	Cancers	74
TUG1	*HOXB7*	Scaffold, Guide	Cholangiocarcinoma	79

LncRNAs are involved in the epigenetic regulation of genes and have been implicated in a number of diseases.

Long ncRNAs and Cancer

In cancer, several lncRNAs are dysregulated and act to *repress* important tumor suppressor genes.

Examples:

- *HOTAIR* overexpression promotes tumor metastasis by re-targeting PRC2 to metastasis suppressor genes. Furthermore, knockdown of *HOTAIR* inhibits cancer invasiveness, highlighting its potential as a target for cancer therapeutics.

- *ANRIL* was first identified as an antisense transcript overlapping the INK4b/ARF/INK4a tumor suppressor locus that emerges from the 403 kb germ-line deletion observed in patients with melanoma- neural system tumor syndrome.

Strategy for gene therapy: De-Repressing (by lncRNA KD) Tumor Suppressor Genes

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

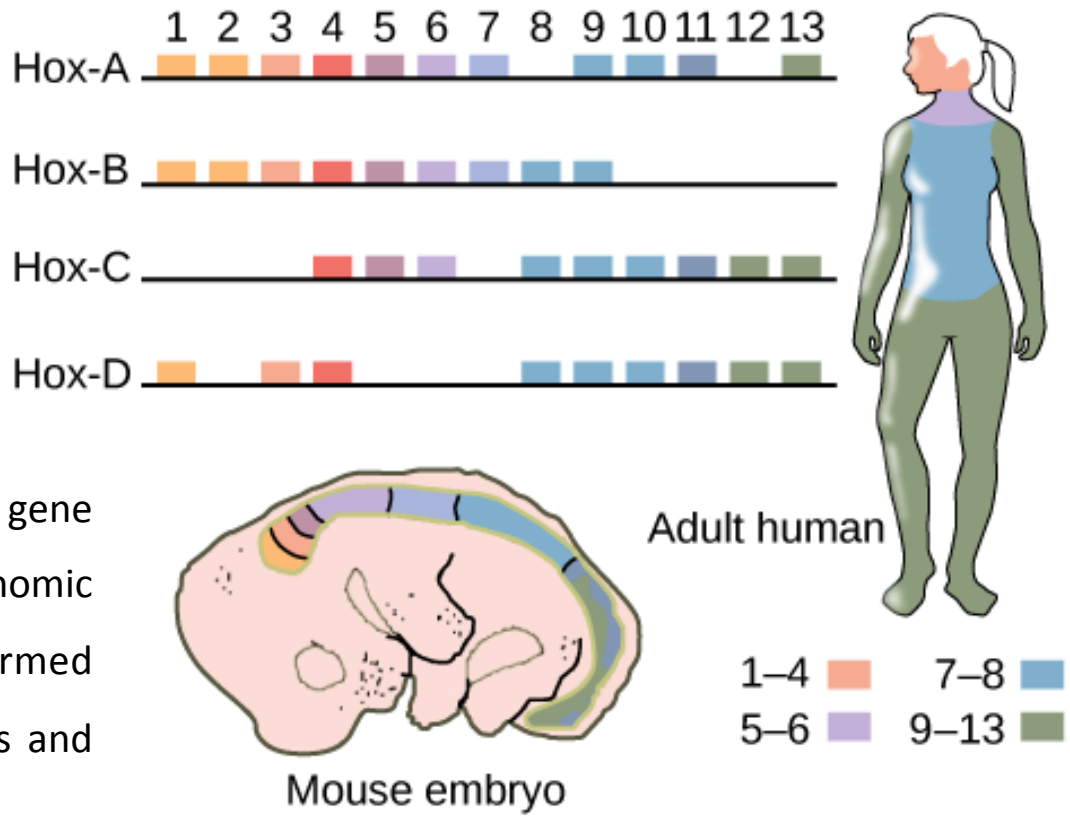
Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.

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.

Demarcation of *active* and *silent* chromatin domains in HOX loci by ncRNAs

In mammals, 39 homeobox transcription factors (HOX) clustered into 4 chromosomal clusters (HOXA through HOXD) are essential for specifying the positional identities of cells.



The temporal and spatial pattern of HOX gene expression is often correlated to their genomic location within each loci, a property termed *colinearity* (Kmita and Duboule, 2003; Lemons and McGinnis, 2006).

gene position= ***spatial*** position
along the anterior-posterior *anatomic axis*

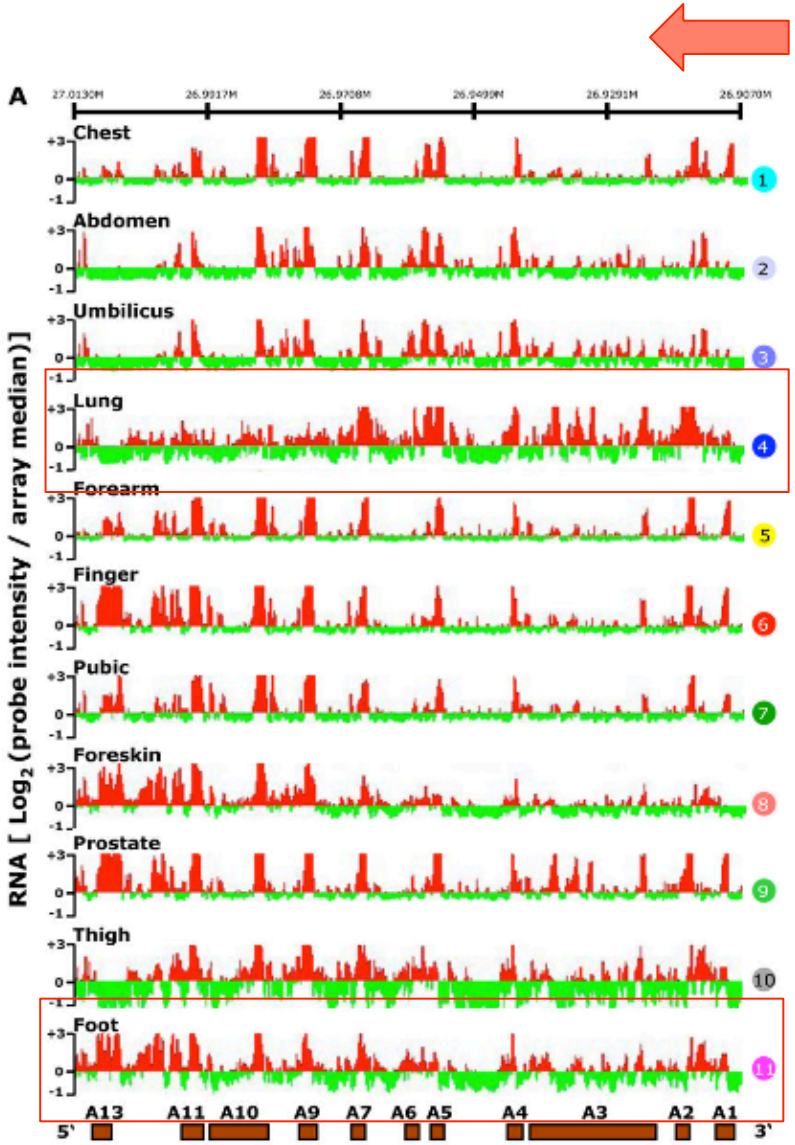
The human HOX transcriptome

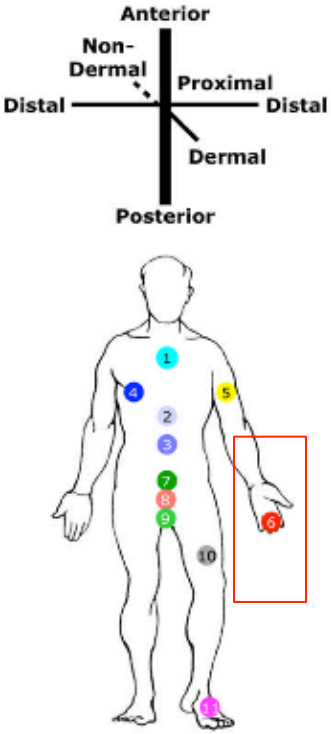
- DNA microarray for all 4 human HOX loci at five base pair (bp) resolution

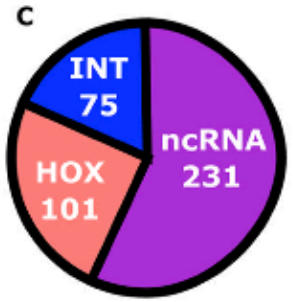
Prossimale

- Site-specific transcription of the HOXA locus.

Distale





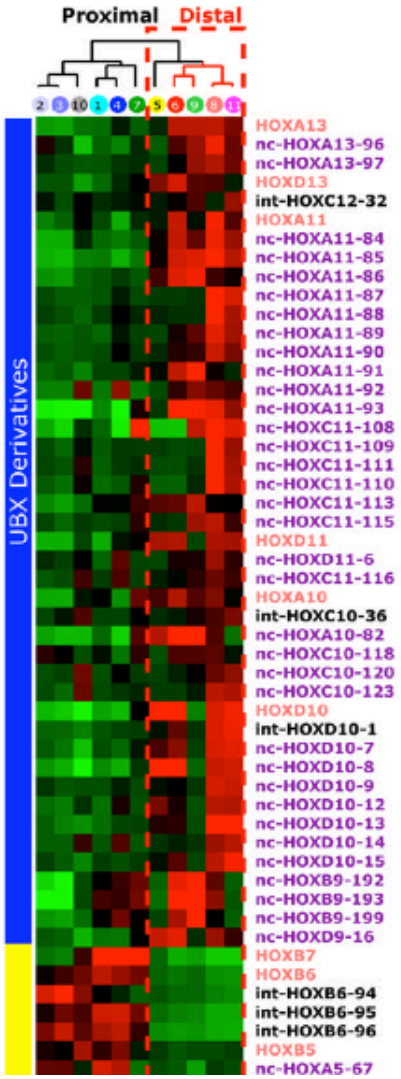


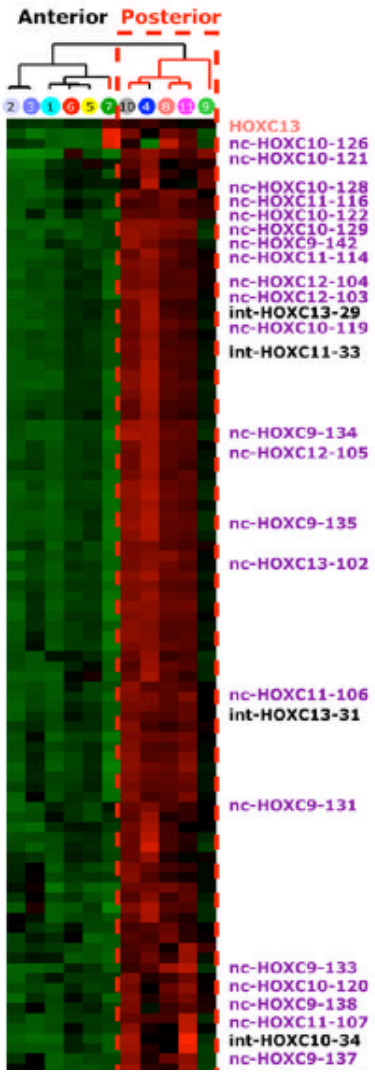
11 samples

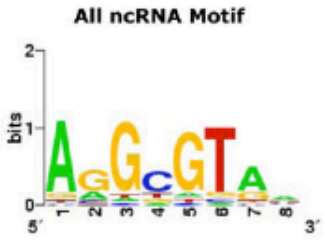
- Numerous long noncoding RNAs were found to be transcribed within the human HOX clusters (*Rinn et al., 2007*)

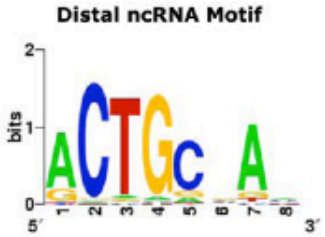
- The lncRNAs were found to be also *co-linear* with the overall anatomic expression pattern of the HOX loci

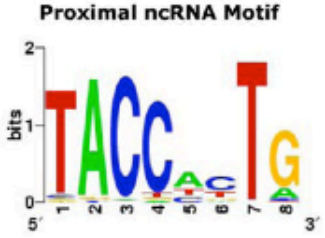
- in addition to their distinctive expression patterns, we found that the ncRNAs also possess specific sequence motifs.



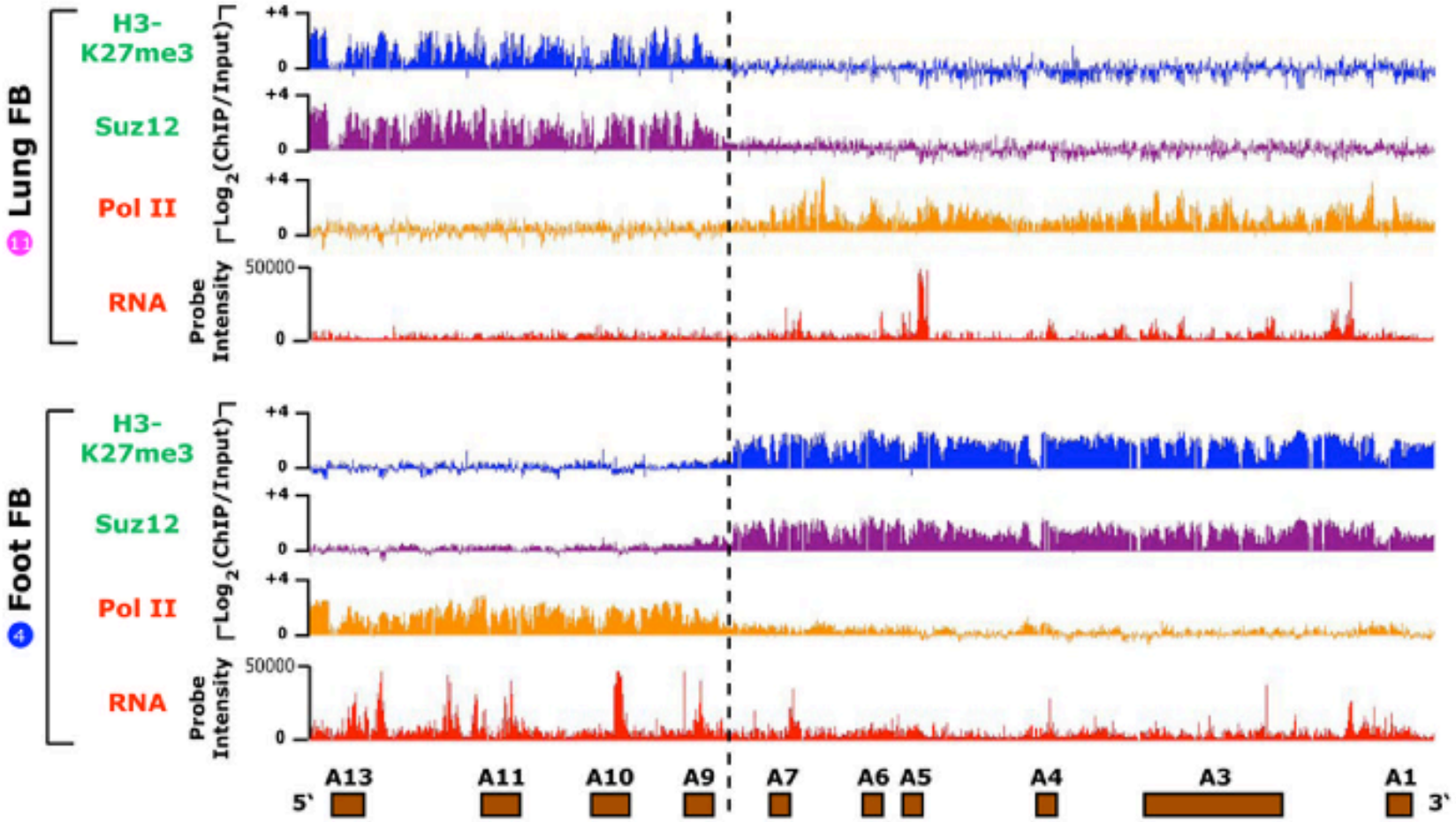






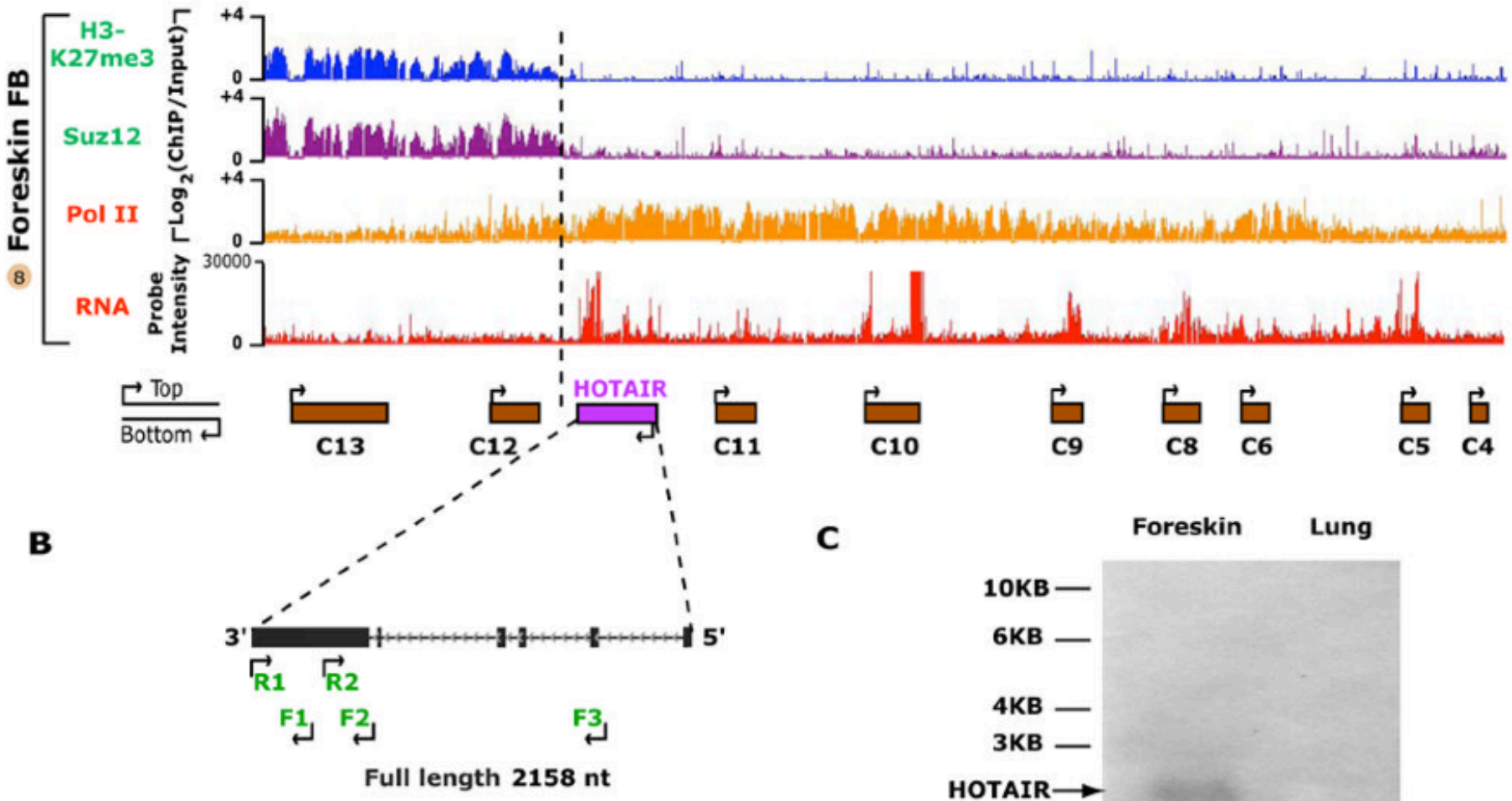


Diametrically opposed chromatin modifications and transcriptional accessibility in the HOXA locus



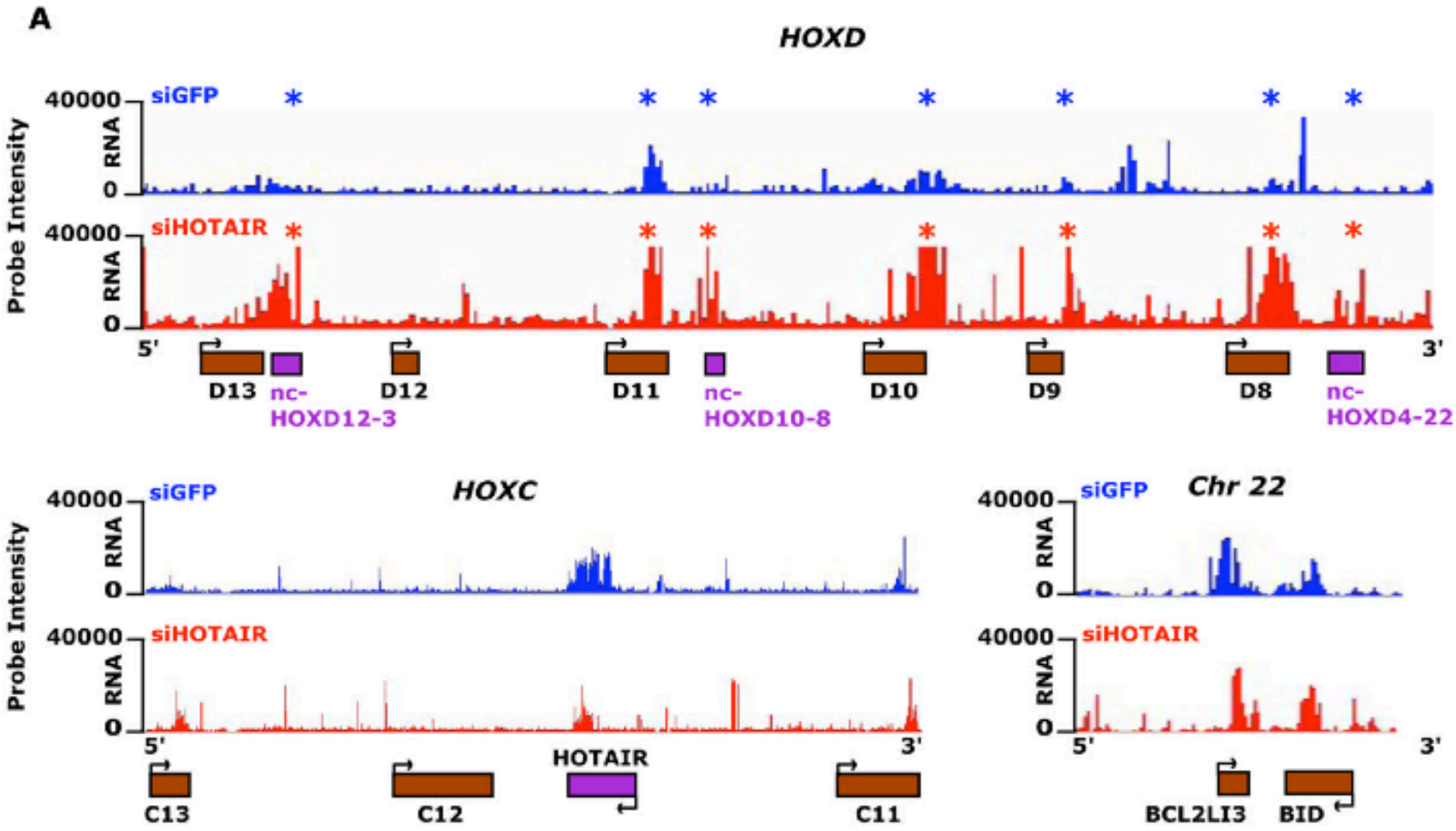
ChIP data

Genomic location of HOTAIR at the boundary of two chromatin domains.



Strand specific RT-PCR shows exclusive expression of HOTAIR from the strand opposite to HOXC genes

Loss of HOTAIR results in transcriptional induction of HOXD locus

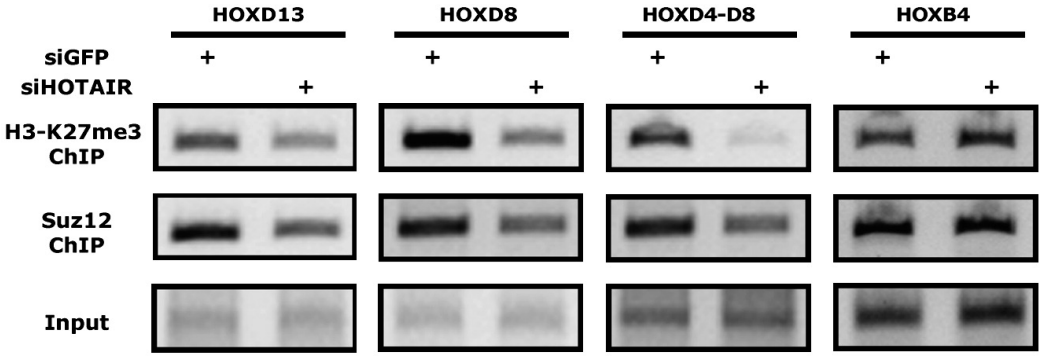


RNAi against HOTAIR in primary fibroblast led to *dramatic* transcriptional activation of the HOXD locus on chromosome 2 spanning over 40 kb, including HOXD8, HOXD9, HOXD10, HOXD11 and multiple ncRNAs

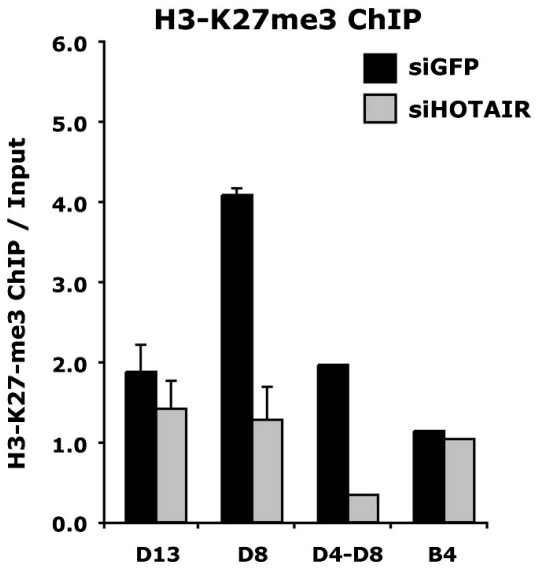
HOTAIR Enhances PRC2 Activity at the HOXD Locus

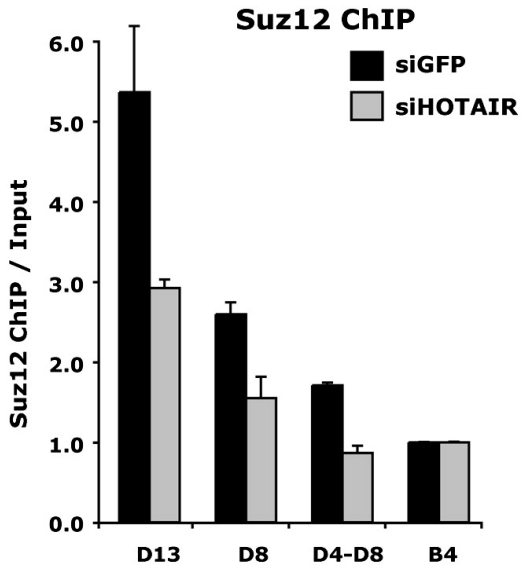
HOTAIR is required **for H3K27 trimethylation** and **Suz12** occupancy of the *HOXD locus*

ChIP of **H3K27me3** and Suz12 of select promoters across the HOXD locus after siRNA treatment targeting GFP or HOTAIR.



Bottom: quantitation of ChIP assays (mean ± standard error).



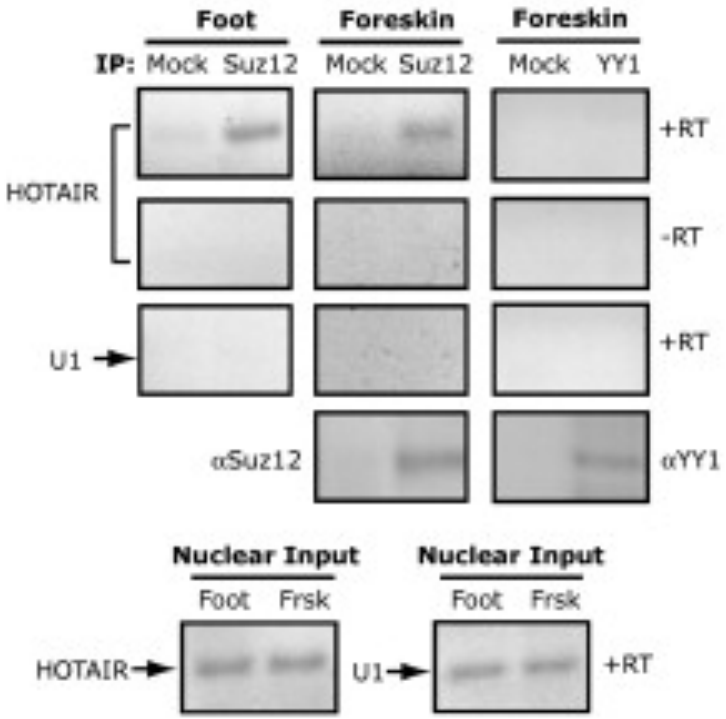


Cell. 2007 June 29; 129(7): 1311–1323.

HOTAIR ncRNA Binds PRC2

Immunoprecipitation

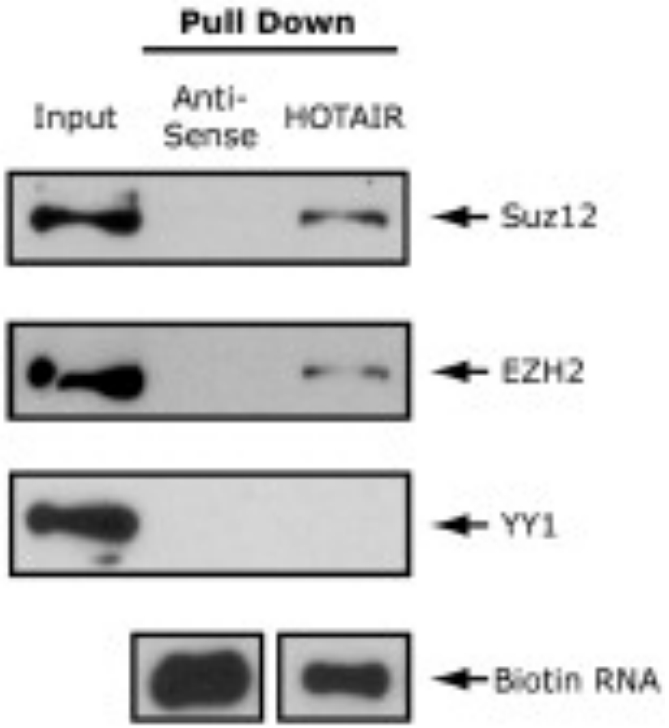
IP of Suz12 retrieves endogenous HOTAIR



Nuclear extracts of fibroblasts were immunoprecipiated by IgG (Mock), anti-Suz12, or anti-YY1. Co-precipitated RNAs were detected by RT-PCR using primers for HOTAIR or U1 small nuclear RNA.

RNA Pull Down

In vitro-transcribed HOTAIR retrieves PRC2 subunits

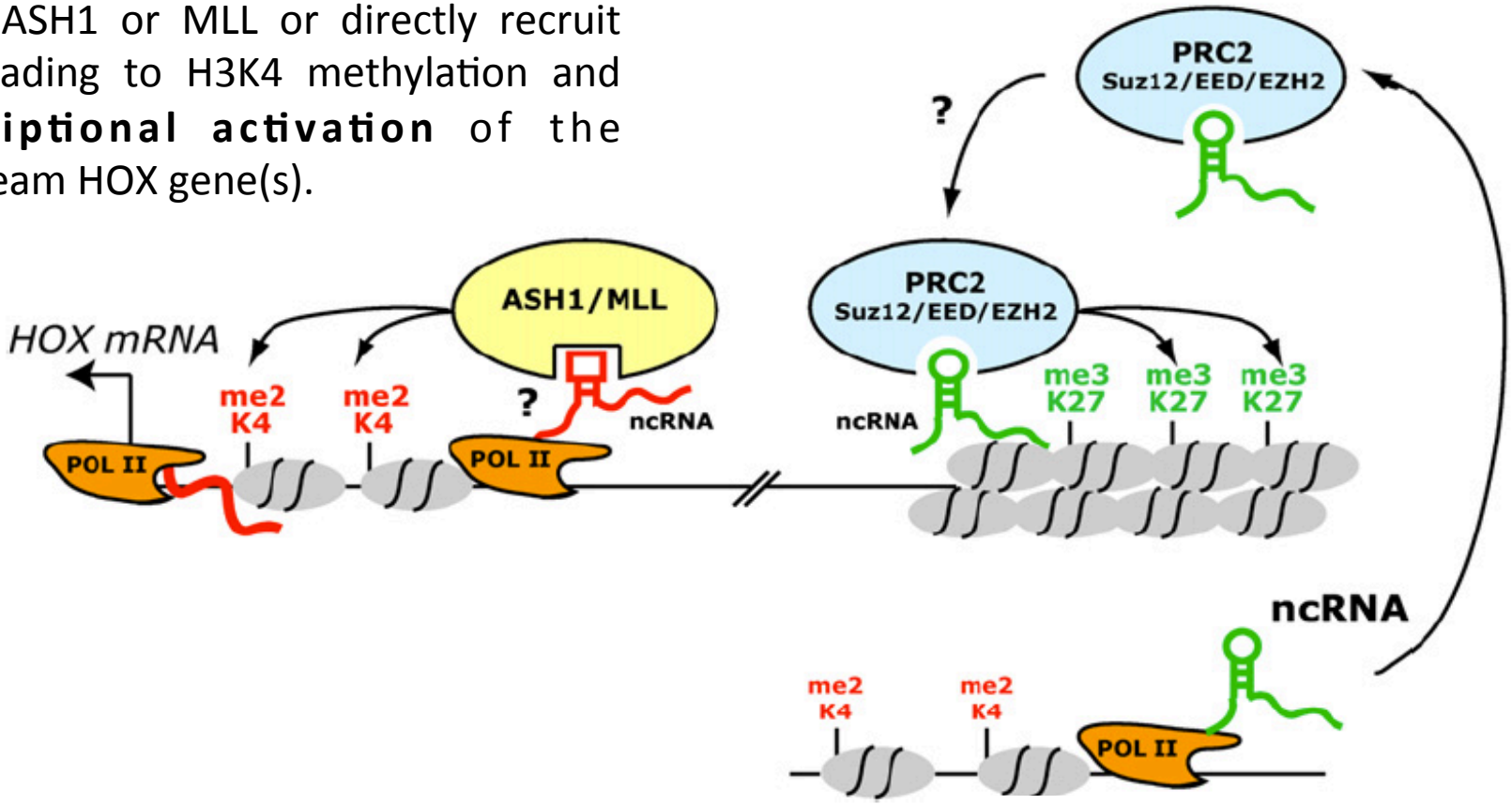


Cell. 2007 June 29; 129(7): 1311–1323.

cis and *trans* lncRNA regulation of chromatin domains

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).

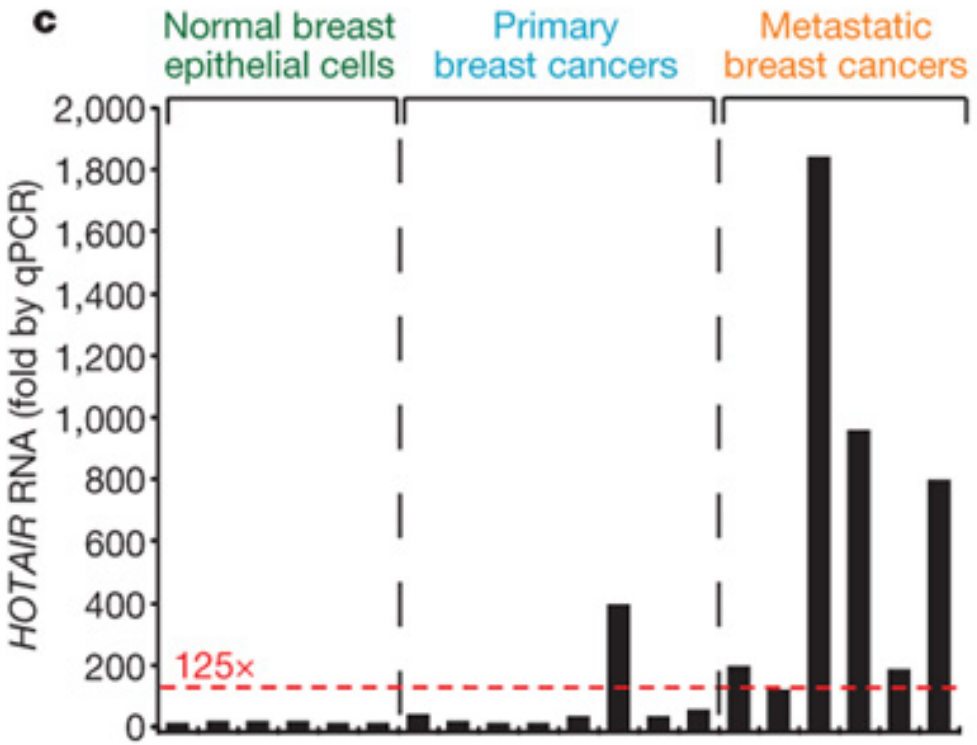
In contrast, recruitment of PRC2 is programmed by ncRNAs produced *in trans*, which targets PRC2 activity by yet incompletely defined mechanisms to target loci. PRC2 recruitment leads to H3K27 methylation and **transcriptional silencing** of neighboring HOX genes.



Cell. 2007 June 29; 129(7): 1311–1323.

HOX lincRNAs are systematically dysregulated in breast carcinoma and have prognostic value for metastasis and survival

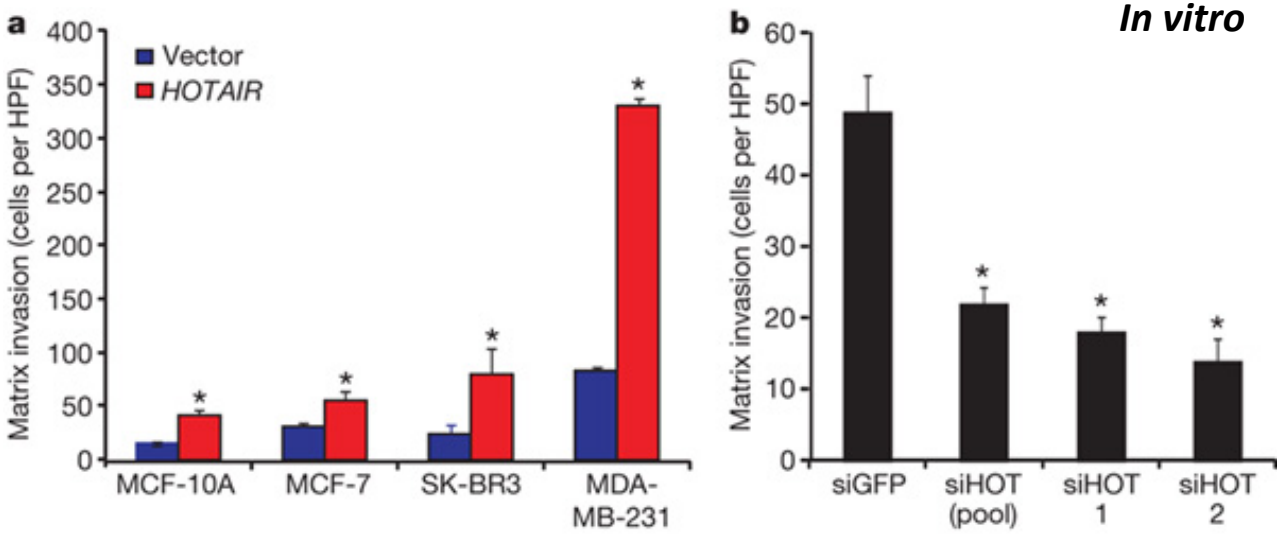
Metastatic tumours had a minimum of 125-fold higher levels of *HOTAIR* than normal breast epithelia.

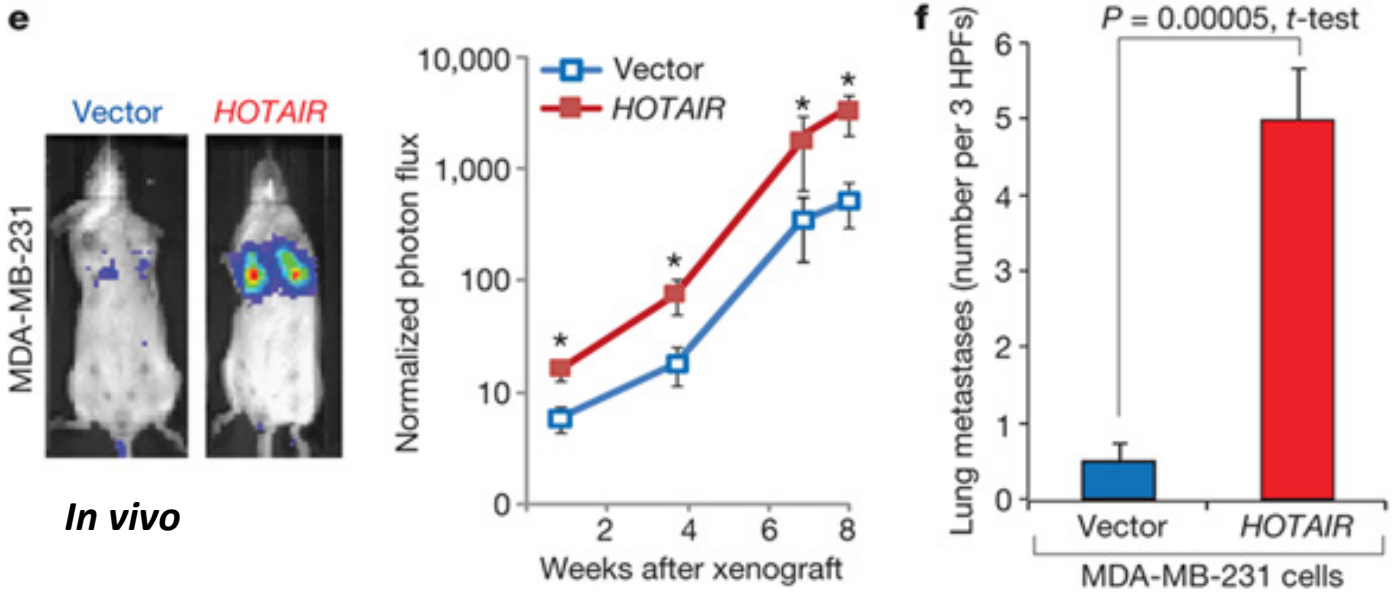


HOTAIR promotes invasion of breast carcinoma cells

a *Increase* in matrix invasion in four breast carcinoma cell lines after enforced *HOTAIR* expression. **b** Matrix invasion in the MCF-7 breast carcinoma cell line transfected with individual or pooled siRNAs targeting *HOTAIR*

In vitro



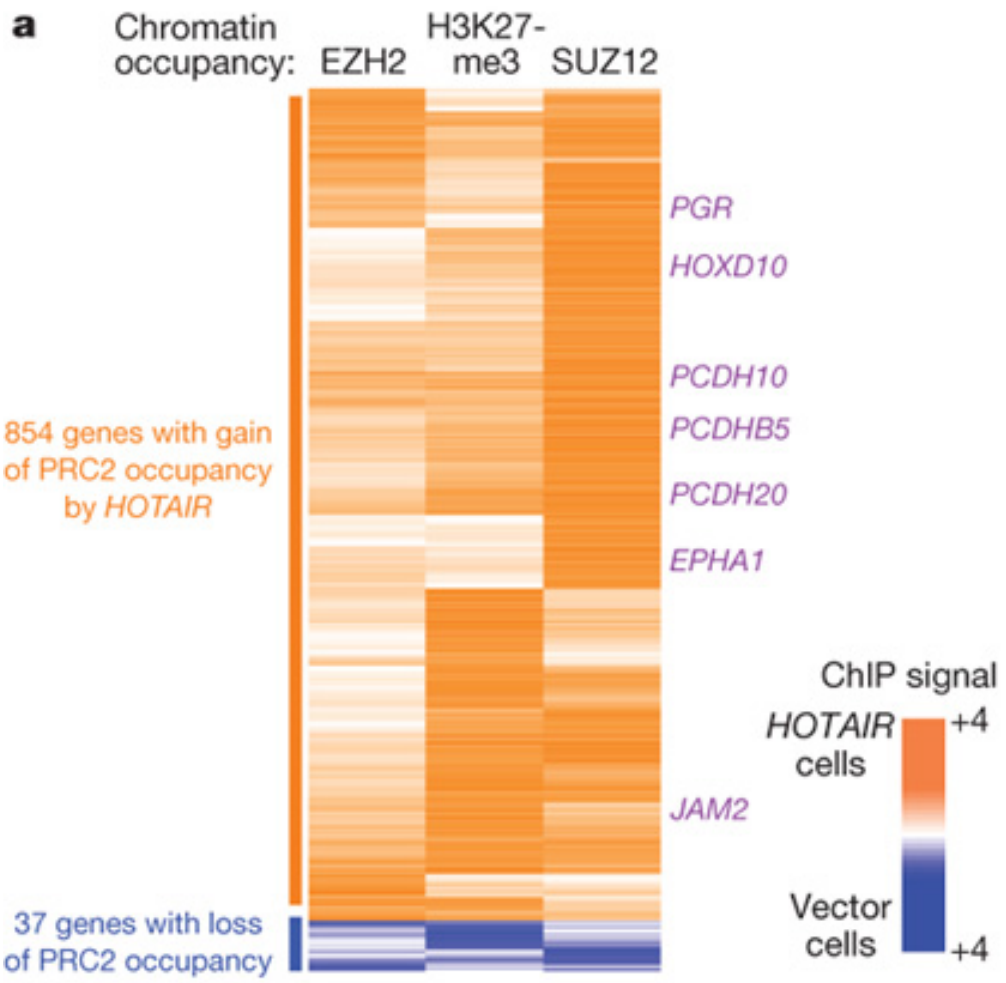


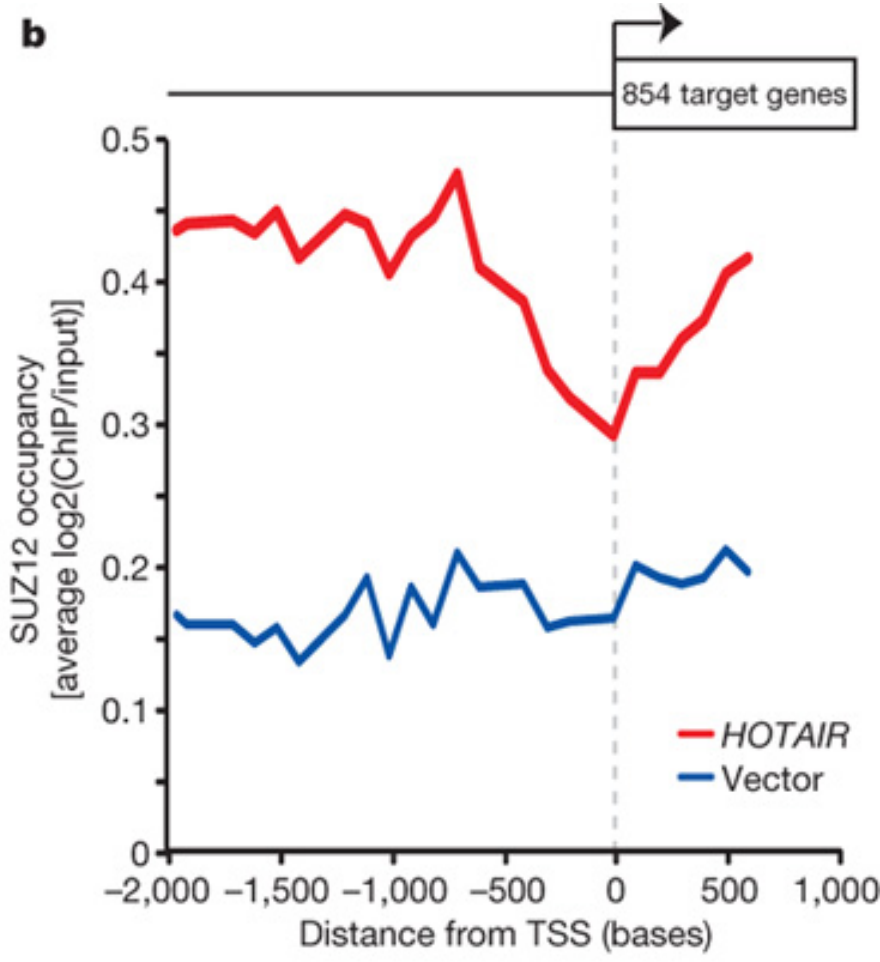
In vivo

Histological analysis of the number of *lung metastasis* in vector- or *HOTAIR*-expressing MDA-MB-231 cells 8–9 weeks after tail vein xenograft

HOTAIR promotes genome-wide re-targeting of PRC2 and H3K27me3

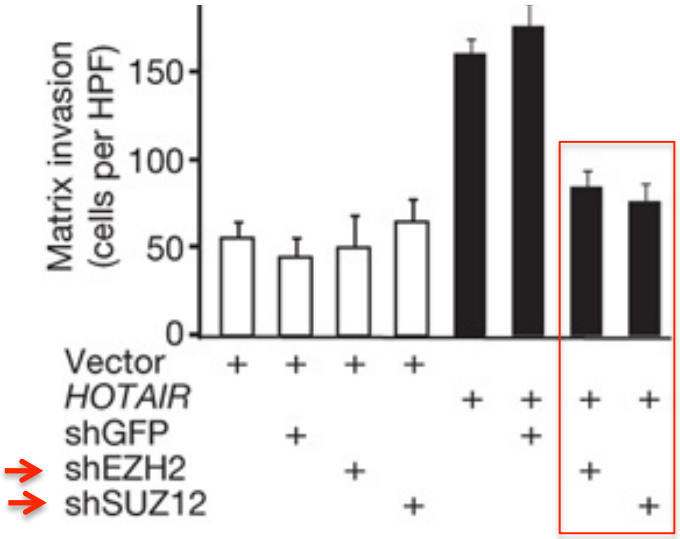
ChIP-chip (compared to vector expressing cells)

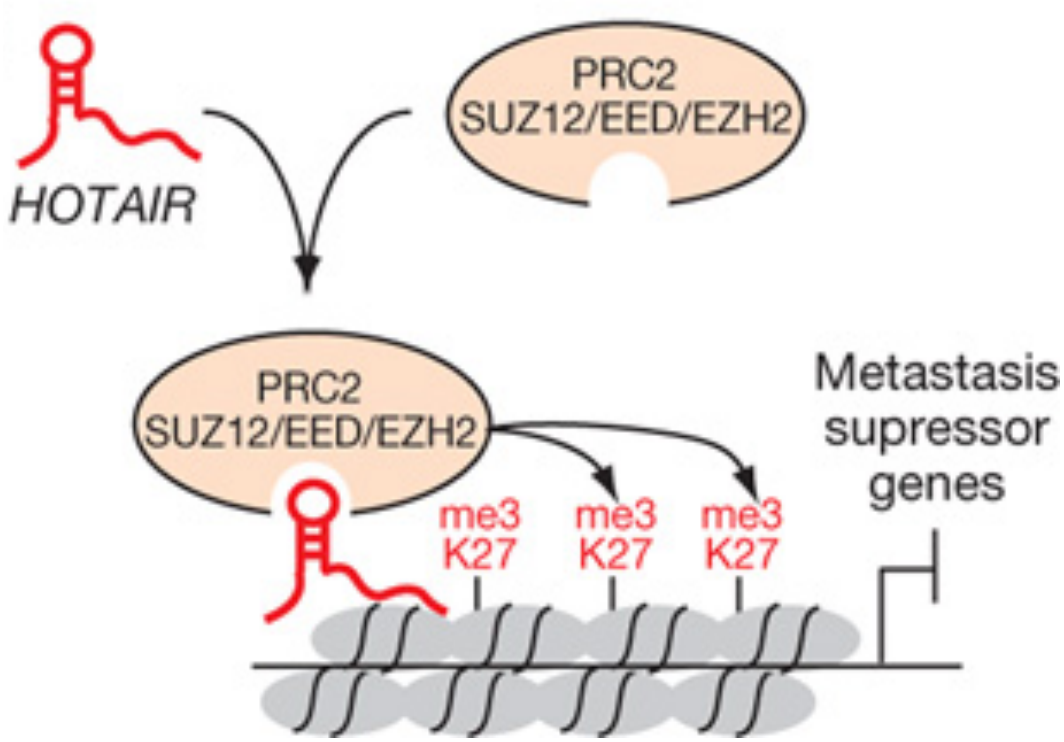




Control or HOTAIR cells were subjected to ChIP using anti-EZH2, H3K27me3 and SUZ12 antibodies followed by interrogation on a genome-wide promoter array

HOTAIR requires PRC2 for function



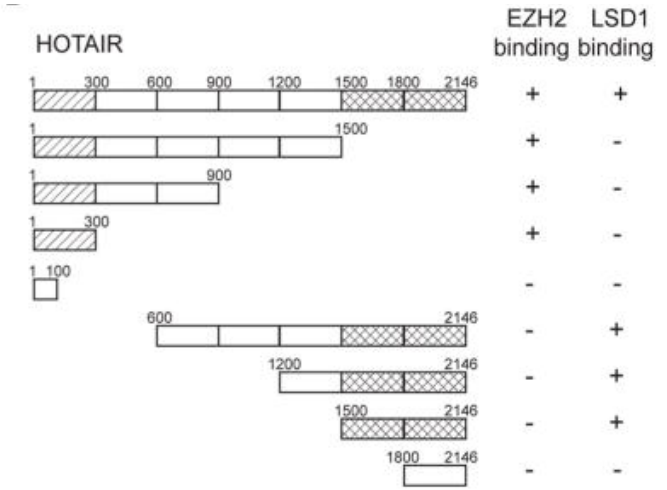


Sopprime le metastasi
quando è overespresso
in combinazione con PRC2

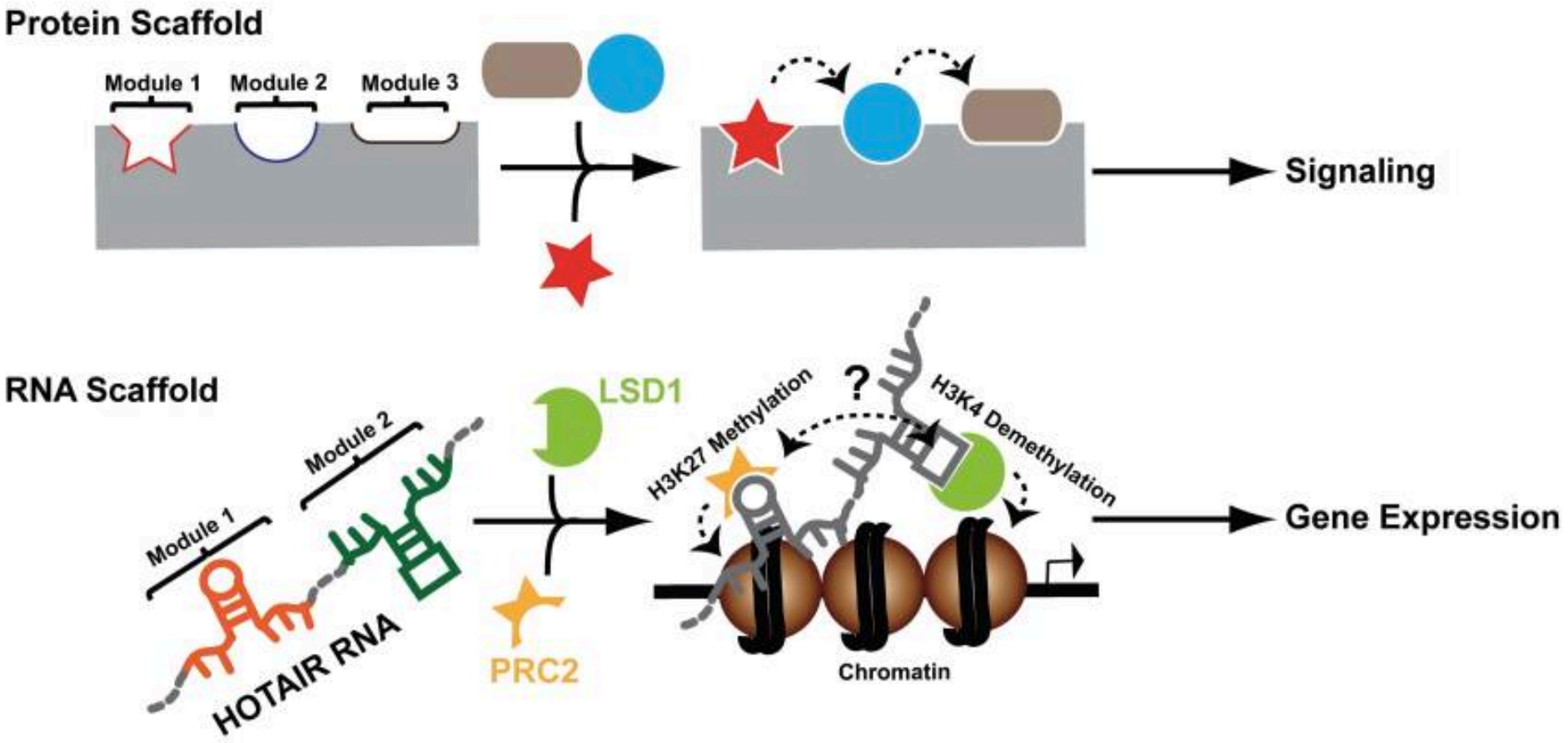
HOTAIR as a paradigm of “molecular scaffold”

lncRNAs can serve as adaptors to bring two or more proteins into discrete complexes

HOTAIR coordinately interacts with both PRC2 and LSD1. A 5' domain of HOTAIR binds polycomb repressive complex 2 (PRC2), whereas a 3' domain of HOTAIR binds the LSD1/CoREST/REST complex. HOTAIR can link a histone *methylase* and a *demethylase* by acting as a modular scaffold

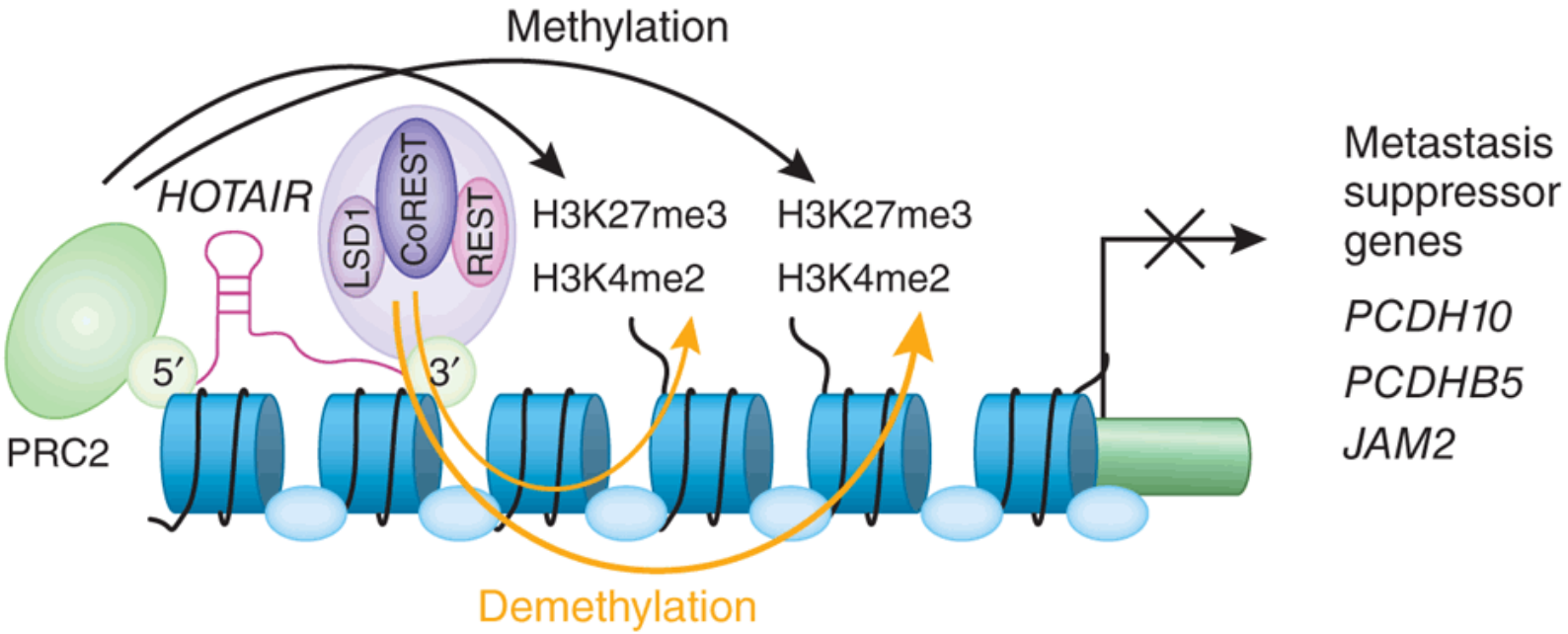


Science. Aug 6, 2010; 329(5992): 689–693.



Parallels between protein and RNA scaffolds. Protein scaffolds provide a meeting place for proteins to recognize their substrate or control a downstream molecular signal. *Similar logic* can be employed by RNA to control cellular signaling events, such as chromatin modification.

The resulting molecular complex is bound to the promoter of genes encoding metastasis suppressors (such as *PCDH10*, *PCDHB5* and *JAM2*) to coordinately regulate the histone modifications H3K27me3 trimethylation and H3K4me2 demethylation (that removes an active chromatin mark), which in turn, silence expression of the target genes.



In cancer, several lncRNAs are dysregulated and act to *repress* important tumor suppressor genes.

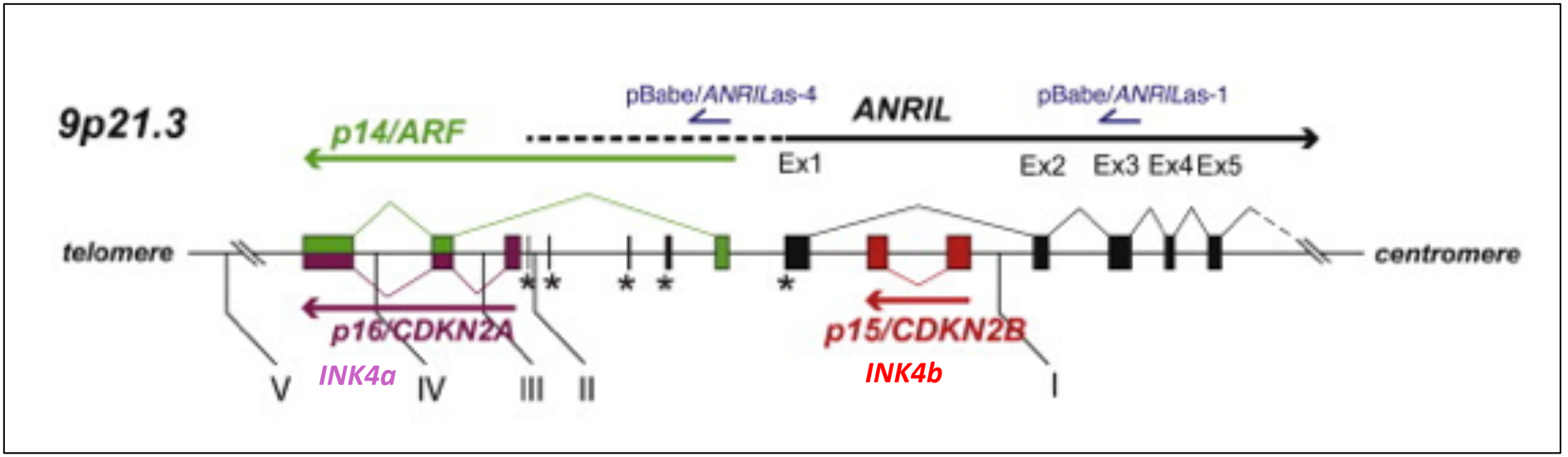
Example: HOTAIR overexpression promotes tumor metastasis by re-targeting PRC2 to metastasis suppressor genes. Furthermore, knockdown of *HOTAIR* inhibits cancer invasiveness, highlighting its potential as a target for cancer therapeutics.

ANRIL is a signature transcript for docking PcG proteins

ANRIL is a long noncoding antisense RNA transcript overlapping the INK4b/ARF/INK4a *tumor suppressor* locus (Pasmant et al., 2007) with estimates of the transcript spanning over 30–40 kb in length. This transcript promotes and maintains the epigenetic state of the CDKN2B (INK4b) gene.

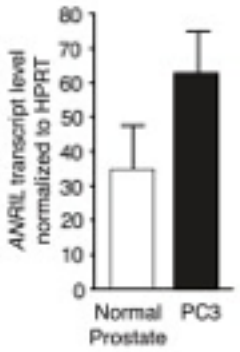
Expression of the INK4b/ARF/INK4a tumor suppressor locus in normal and cancerous cell growth is controlled by H3K27me as directed by the Polycomb group proteins.

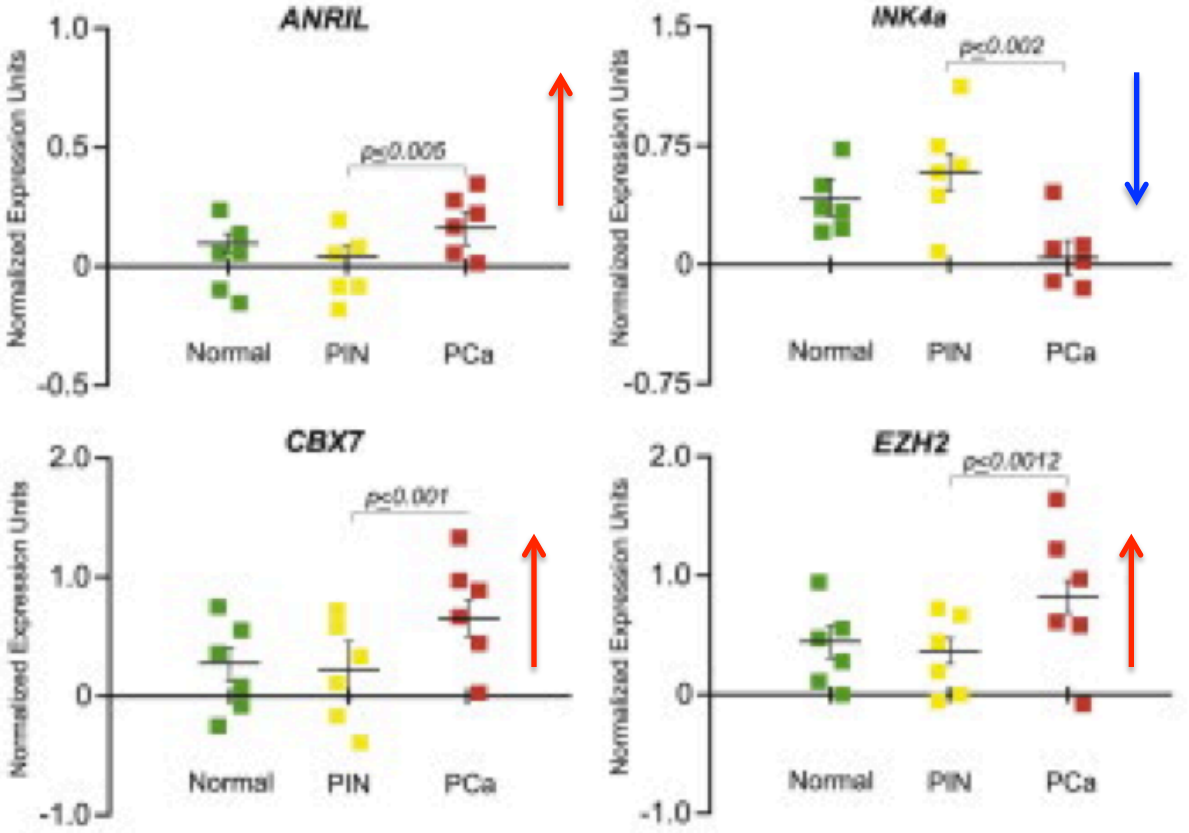
Genomic organization of the INK4a/ARF gene cluster



CBX7 and its interacting ncRNA ANRIL are upregulated in prostate cancer tissues

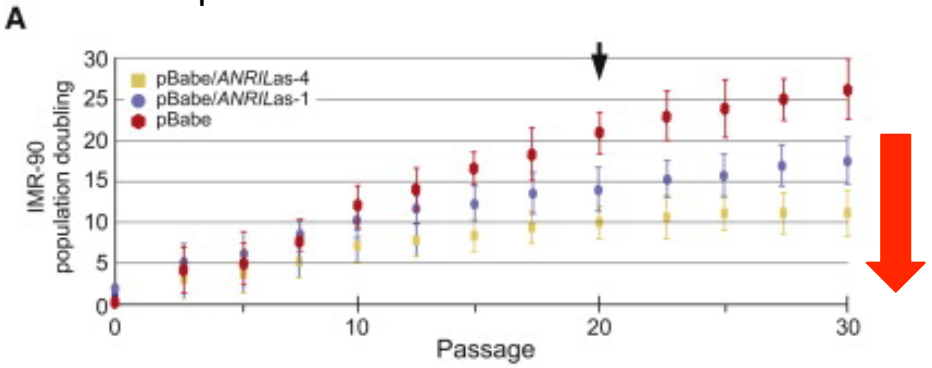
Expression levels of **ANRIL**, **INK4a (p16)**, **CBX7**, and **EZH2** in *normal* prostate epithelium cells, and *preneoplastic* PIN and PCa prostate *cancer* cells.

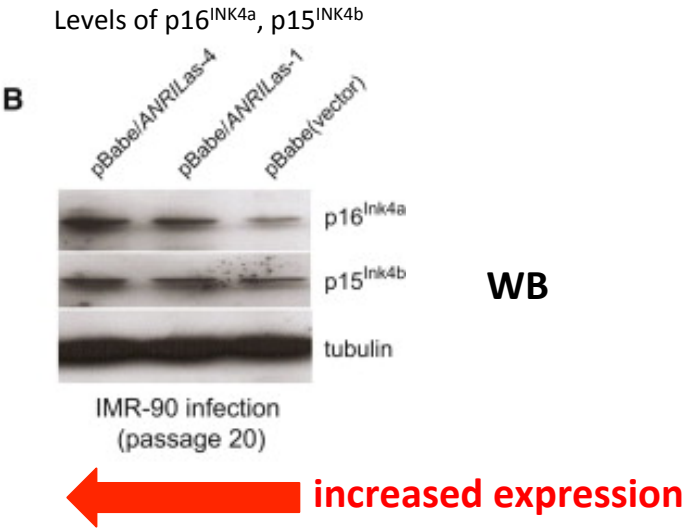


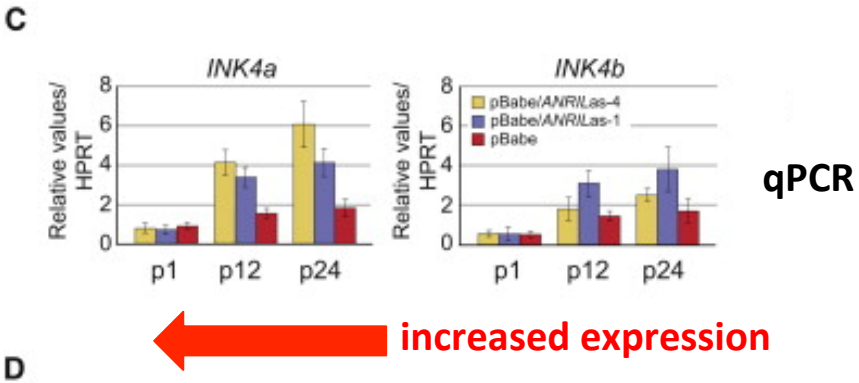


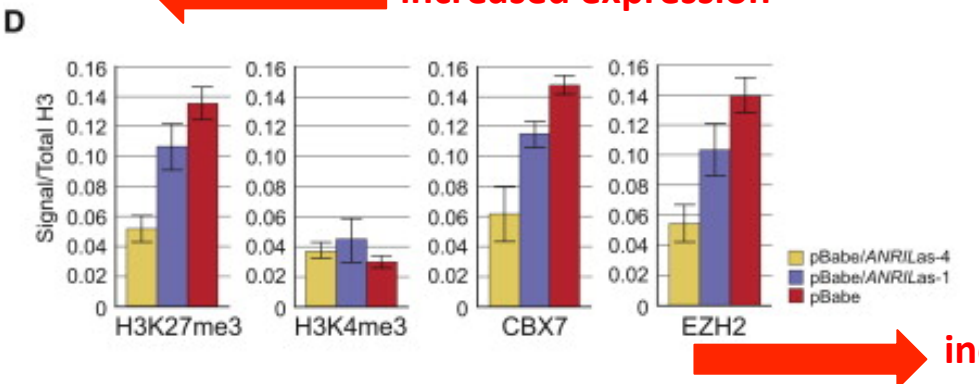
Inhibition of ANRIL results in limitation of life span and in loss of INK4a repression

effects of transcripts antisense to ANRIL
on life span of IMR-90 diploid fibroblasts



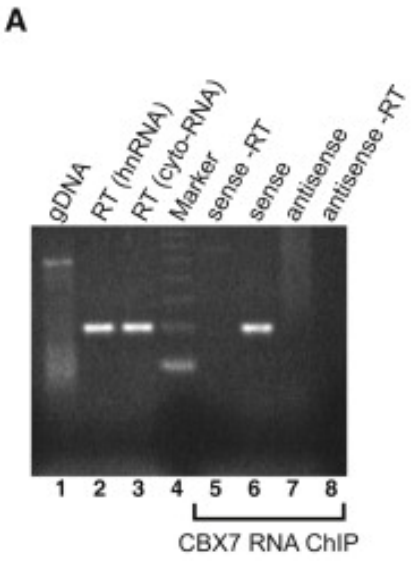


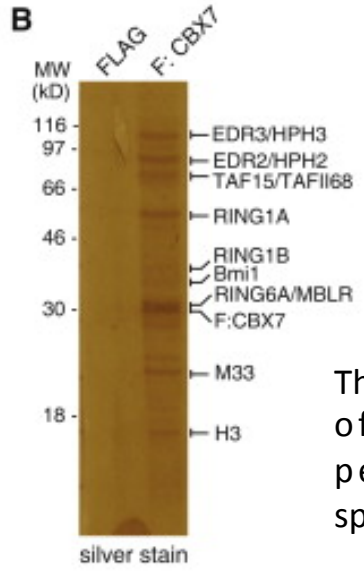


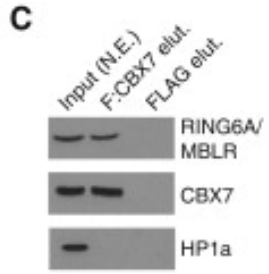


ANRIL RNA Is Associated with Polycomb Protein Recruitment to the INK4a/ARF Locus

Recovery of ANRIL transcript by CLIP analysis of **CBX7** in chromatin

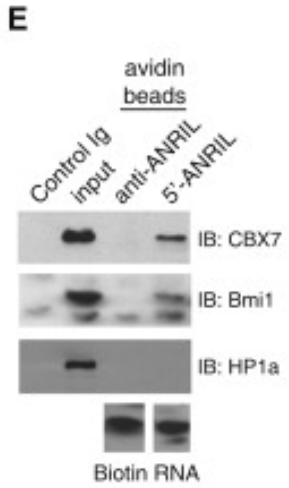




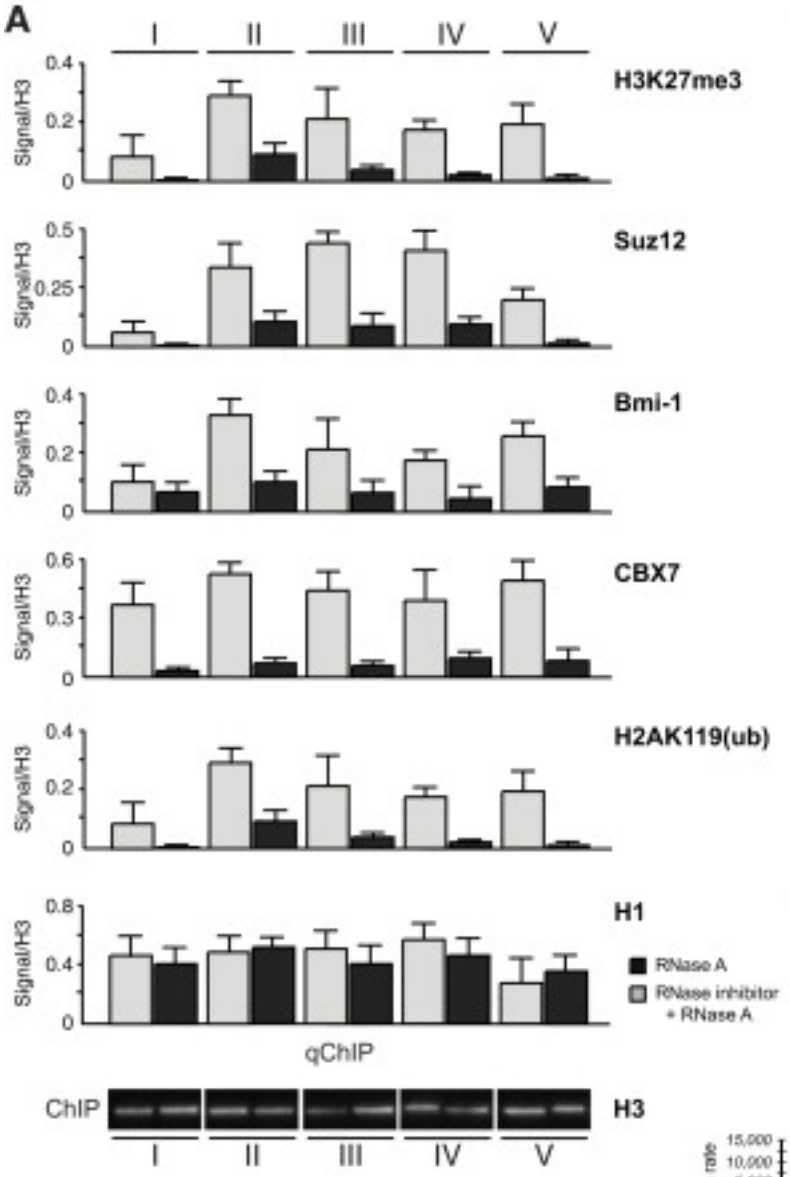


The human **CBX7 complex**: identification of CBX7/PRC1 components was performed by MALDI-TOF mass spectrometry

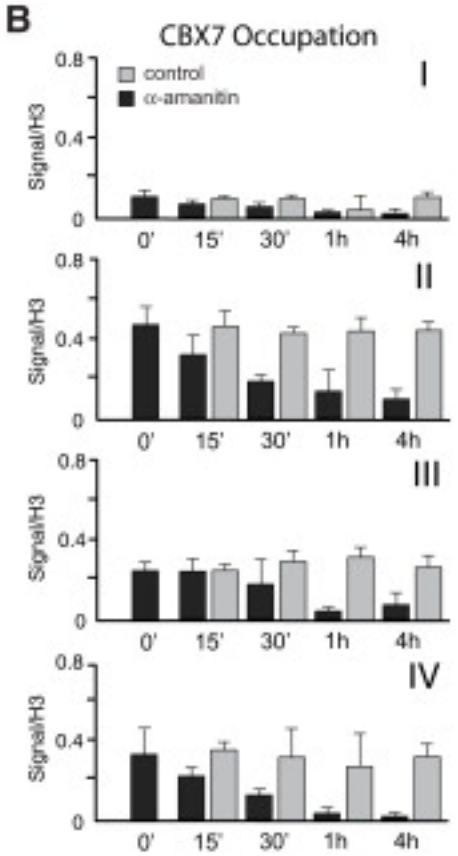
Nuclear extract from PC3 cells was used to evaluate by western blot analysis the *association* of **PRC1** proteins CBX7, Bmi1, and HP1α with the biotinylated **ANRIL** transcript



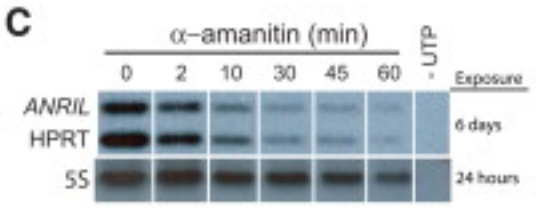
Direct role of ANRIL in repressing the INK4b/ARF/INK4a locus.



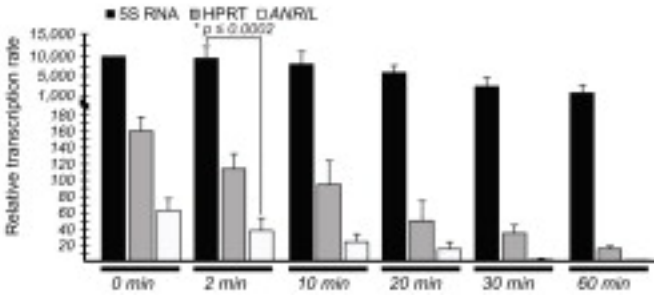
Recruitment of Polycomb Proteins to the INK4a/ARF Locus: RNAse treatment reduced the recruitment



Depleted levels of CBX7 upon whole-cell treatment of α-amanitin: the RNA associating with CBX7 is a Pol II *nascent* transcript



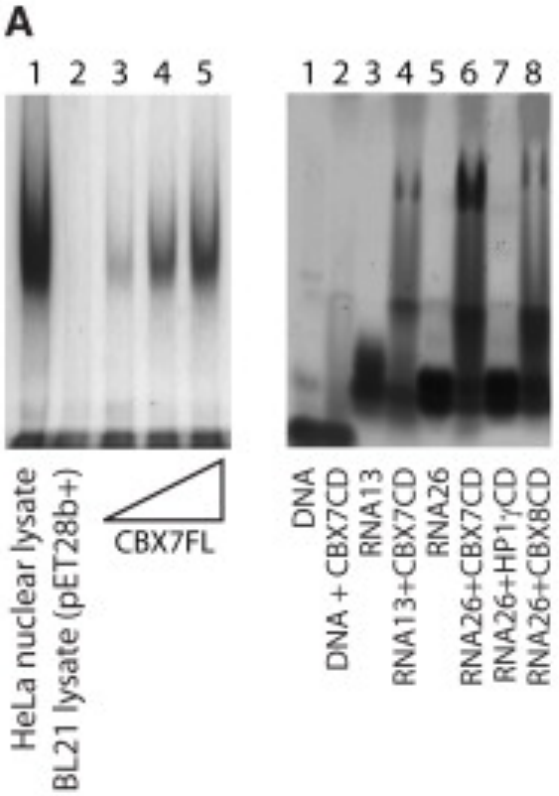
In vitro transcription: ANRIL is a nascent transcript generated by Pol II



CBX7 chromodomain binds H3K27me3 or ANRIL with distinct domains

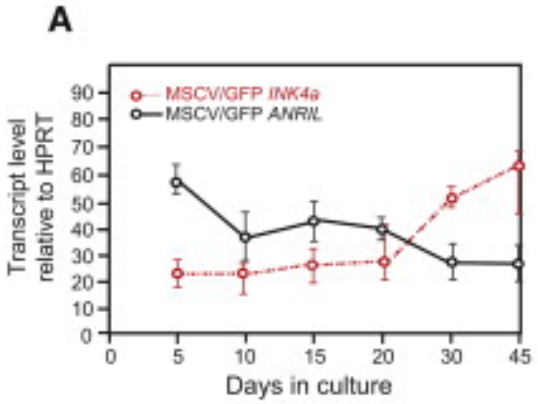
RNA-EMSA showing CBX7 full-length (CBX7FL, *left*) and chromodomain (CBX7CD, *right*) directly associate to RNA.

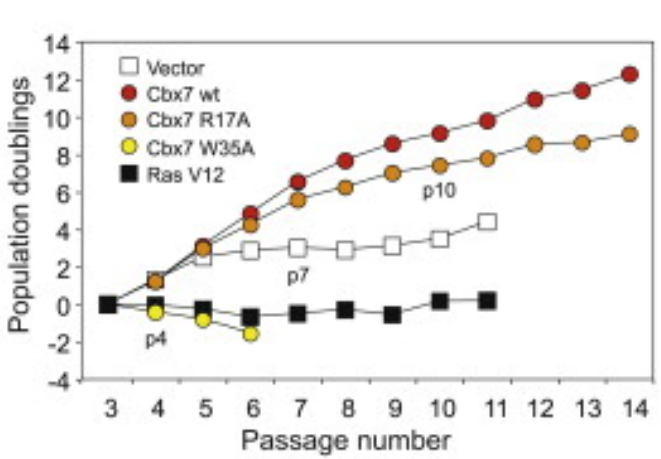
13 nt ssDNA, 13 nt RNA (RNA13), or 26 nt RNA (RNA26)



- The three components (*CBX7CD*, *H3K27me3* peptide, and *ANRIL* loop RNA) are able to form a ternary complex (NMR analysis)
- CBX7 Chromodomain Uses Distinct Regions and Residues for Binding H3K27me or RNA (NMR analysis)

Relative Contributions of CBX7 Interactions with H3K27me and RNA to Senescence Regulation

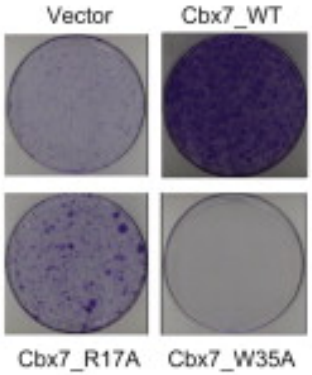




Overespression of Cbx WT or mutants

- Cbx7 wt→ extended lifespan, increased cell growth
- Cbx7 R17A mutant=RNA defective→ slight decrease cell growth
- Cbx7 W35A mutant=H3K27me defective→ growth arrest

C colony formation assay

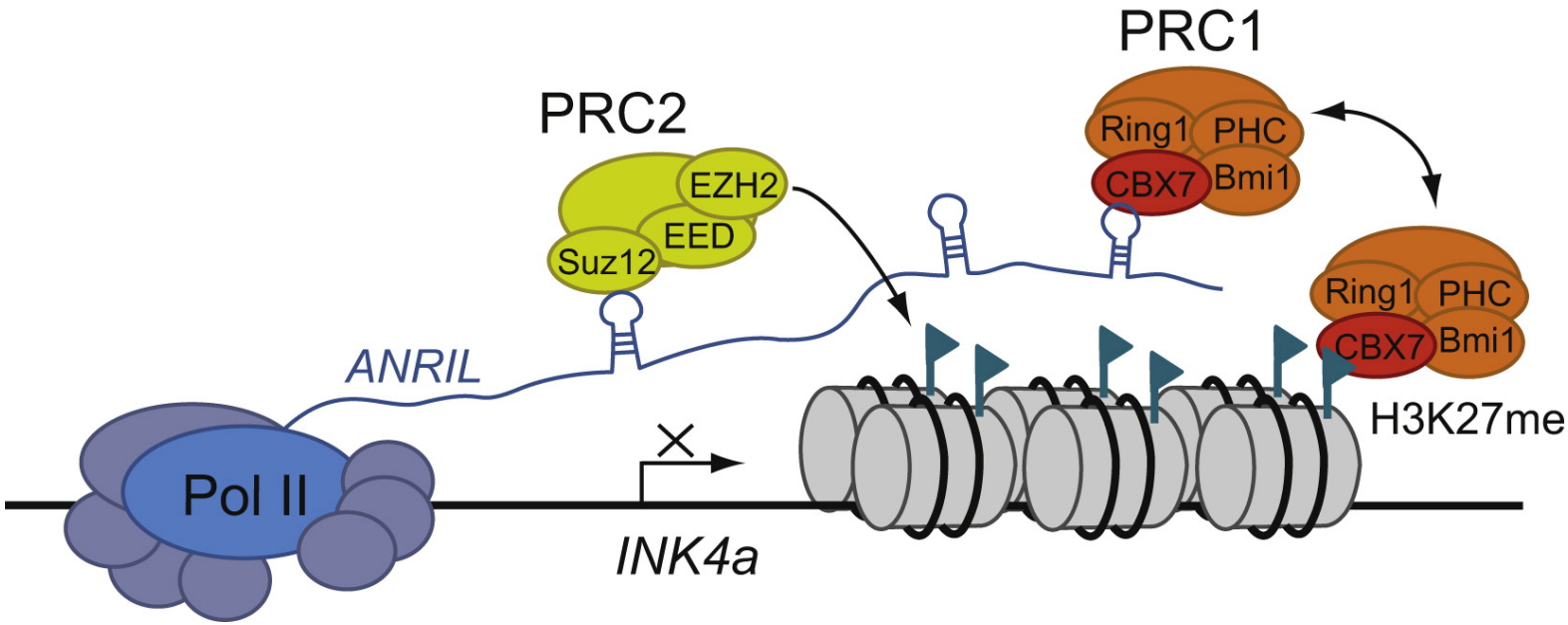


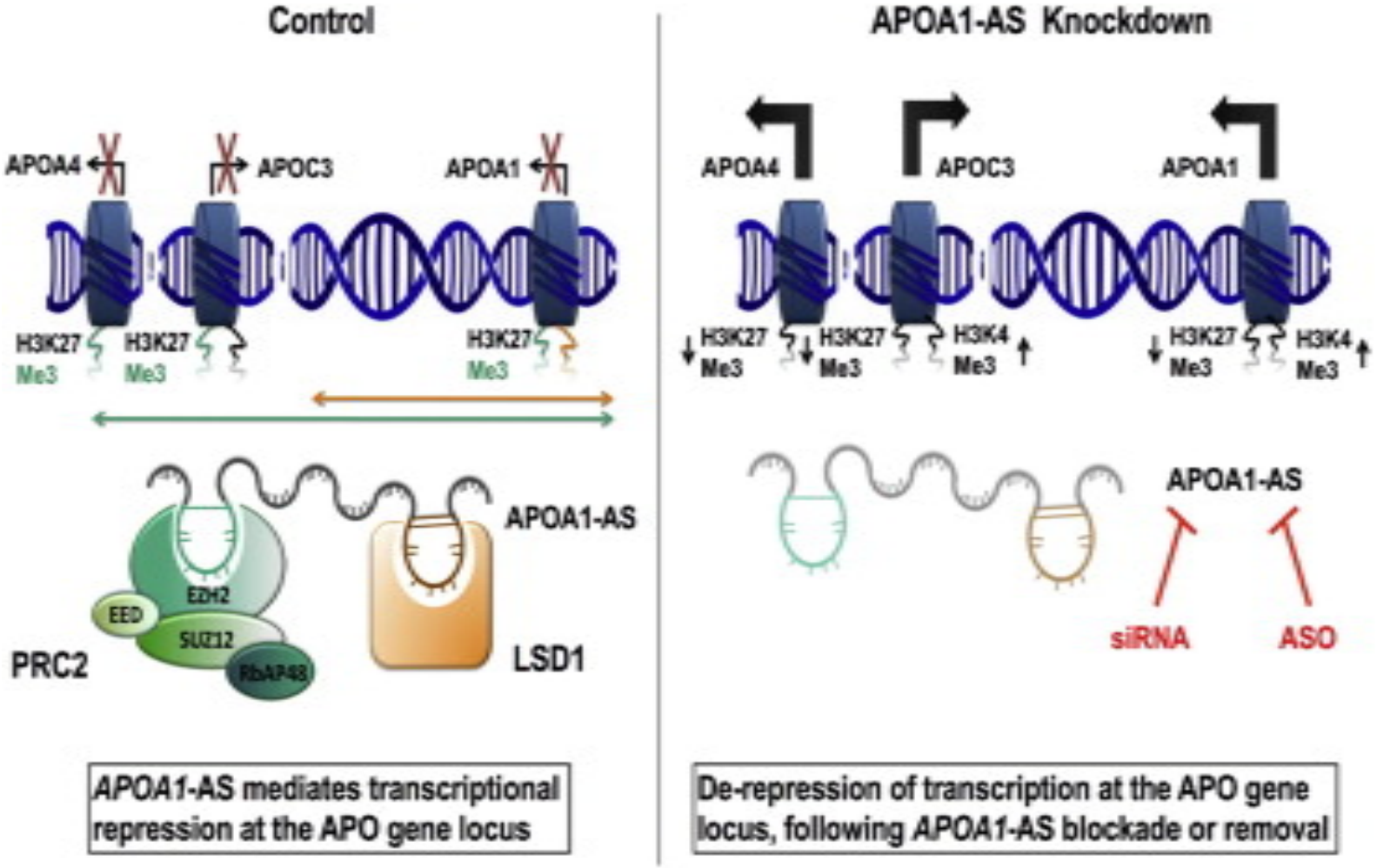
Both interactions are required for the INK4b/ARF/INK4a *locus* repression by CBX7

Ablation of either interaction results in premature growth arrest

ncRNA-mediated gene repression of the INK4b/ARF/INK4a locus by Polycomb group complexes.

- molecular interplay between methylated H3K27 and ncRNA transcripts ANRIL act for *silencing* of the INK4b/ARF/INK4a locus by Polycomb CBX7.
- Both PRC1 and PRC2 complexes are retained at a repression site of a target gene through their association with the nascent ANRIL transcripts of Pol II.
- INK4b/ARF/INK4a is a tumor suppressor locus: oncogenic function!!!





APOA1-AS mediates transcriptional repression at the APO gene locus

De-repression of transcription at the APO gene locus, following *APOA1*-AS blockade or removal

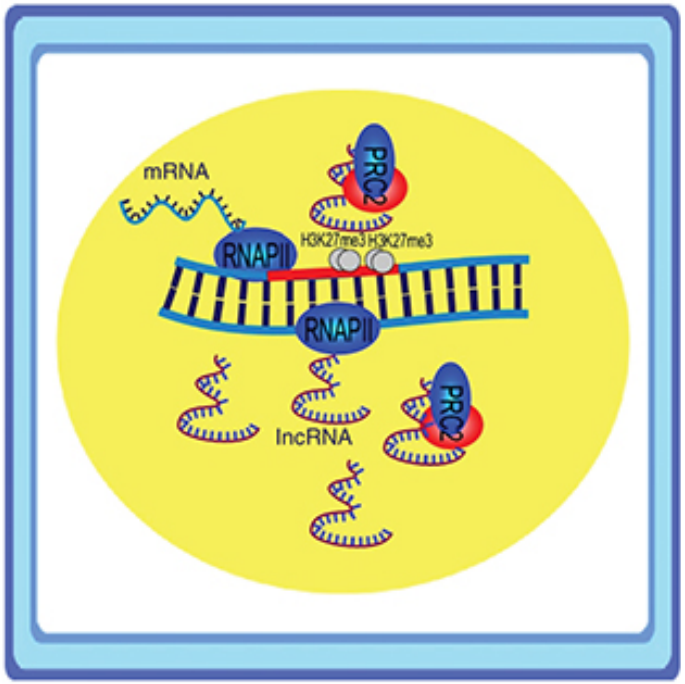
Designing LncRNA-Directed Therapeutics

LncRNAs represent attractive therapeutic targets due:

- to their presence in a development and cell-type specific manner
- to their ability to target a specific subset of genes

Designing LncRNA-Directed Therapeutics

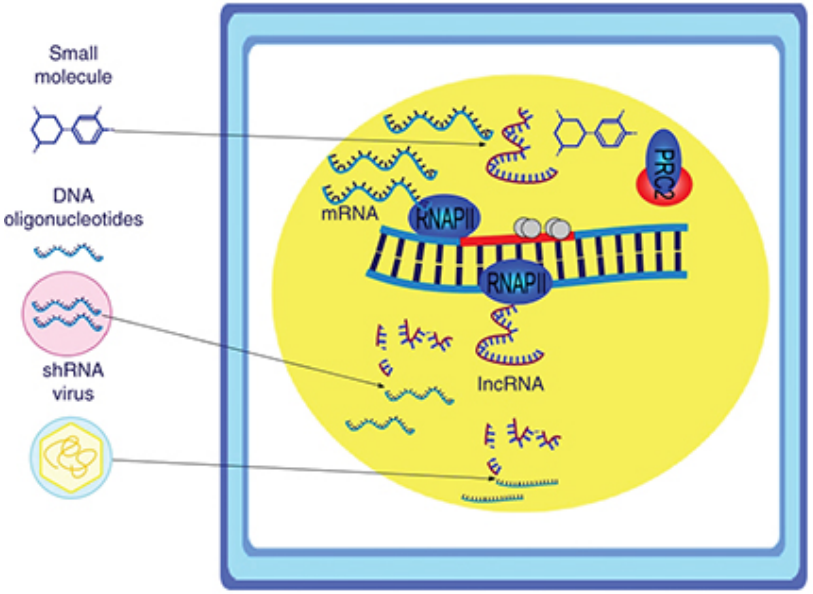
a) LncRNAs bind to the PRC2 complex and guide the complex to target gene promoters (red) to induce H3K27met3) and repress transcription.



b) Potential approaches to inhibit the repressive action of lncRNAs:

- *small molecule* drugs prevent the formation of a functional PRC2-lncRNA complex.
- Naked or liposome-encapsulated DNA *oligonucleotides* can target lncRNAs to form DNA-lncRNA duplexes that recruit RNase H. RNase H results in lncRNA cleavage and decay.
- Viral particles containing lncRNA-targeting shRNAs can deliver shRNAs that generate siRNAs (targeting the RISC complex to the lncRNA) in the cell. The RISC complex cleaves and degrades lncRNAs.

These lncRNA-targeting approaches result in the dissociation of the PRC2 complex from the target promoter, de-repressing genomic loci to upregulate the expression of therapeutic genes.

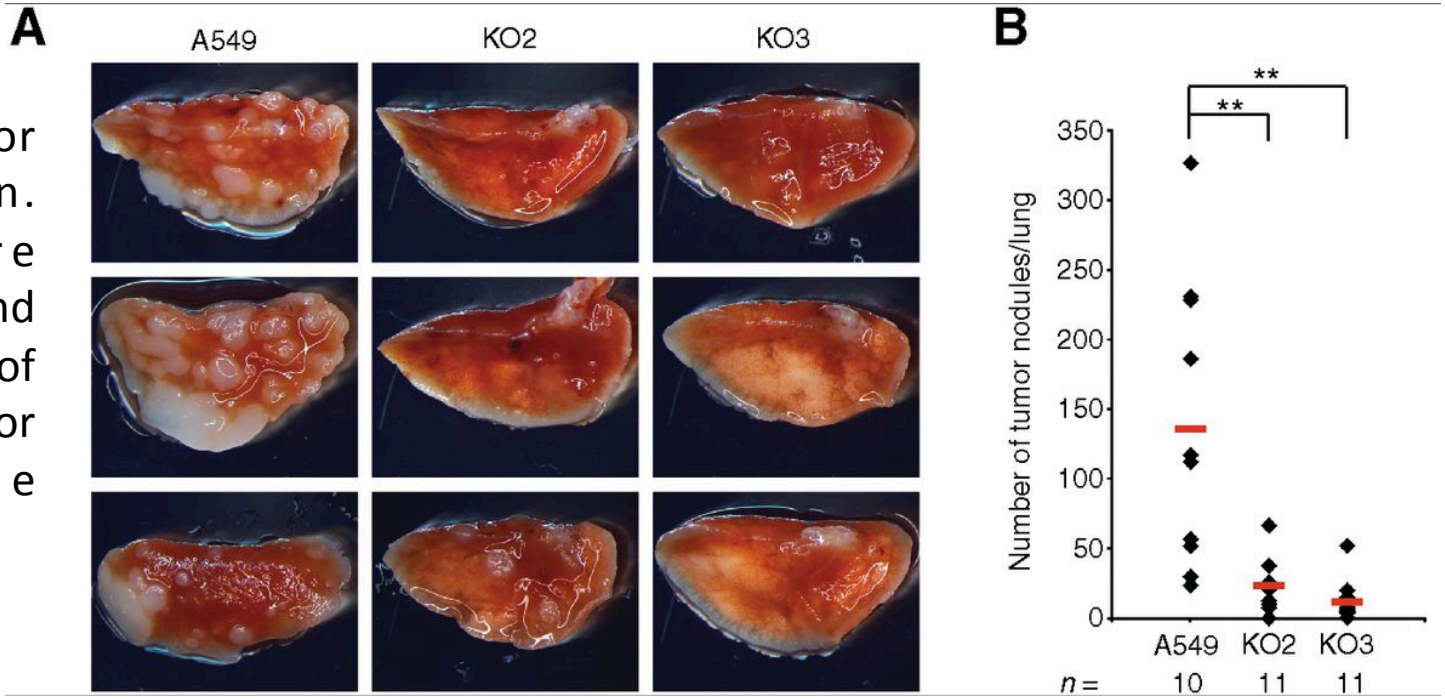


One of the first cancer-associated lncRNAs discovered was MALAT1 (metastasis-associated lung adenocarcinoma transcript 1)

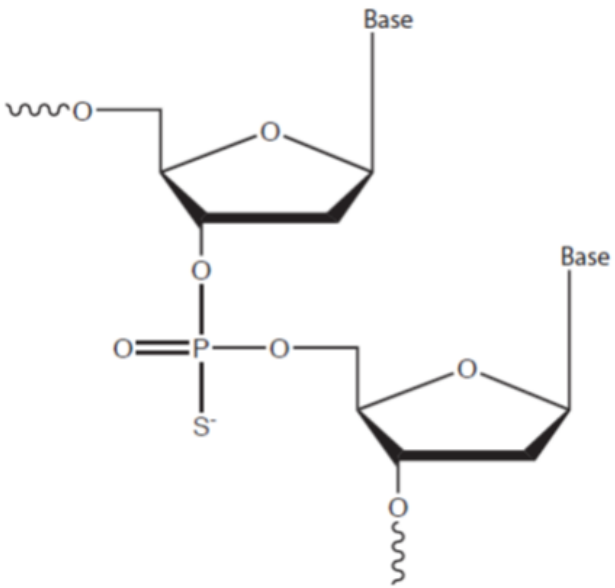
Subcutaneous administration of ASO antisense oligonucleotides targeting *MALAT1* effectively inhibited human lung cancer cell proliferation in a mouse xenograft model.

mouse *xenograft assay* by injecting either A549 *MALAT1* WT cells or the two knockout cell lines with the lowest *MALAT1* expression (KO2 and KO3) into the tail vein of nude mice and analyzed the formation of lung tumor nodules after two months.

Lung cancer tumor nodule formation. The lungs were resected, fixed, and the total number of tumors and the tumor area were determined.

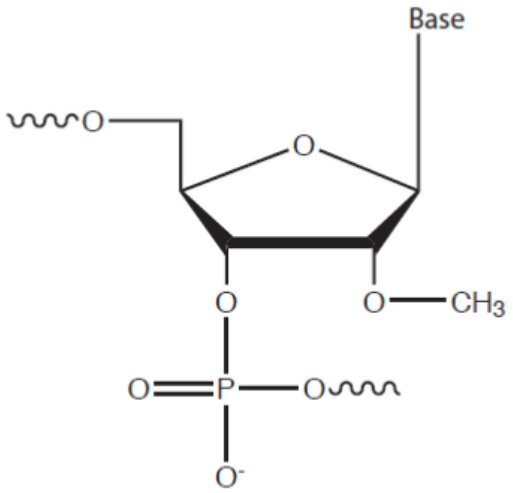


Antisense oligonucleotides (ASOs)



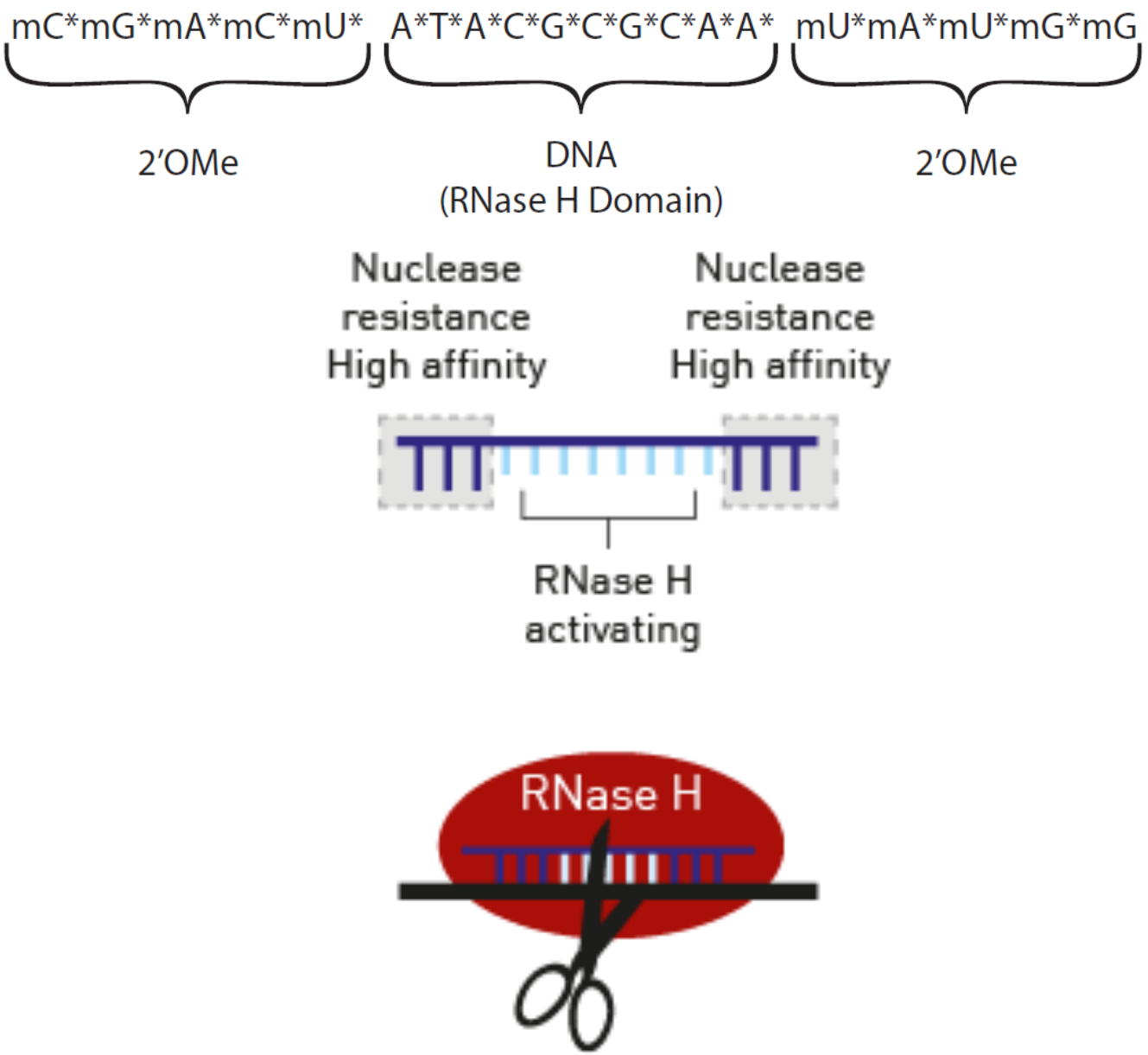
ASO are short, synthetic 14–22 nt oligonucleotides that localize to the nucleus. They include phosphorothioate (PS) linkages that confer nuclease resistance, thus enhancing intracellular stability. A PS bond is created by replacing one of the non-bridging oxygen atoms in the phosphate backbone of the oligo with a sulfur atom.

Inclusion of PS bonds increases oligonucleotide half-life in human serum by up to 10 hours compared to approximately 1 hour for an unmodified oligonucleotide with the same nucleotide sequence.



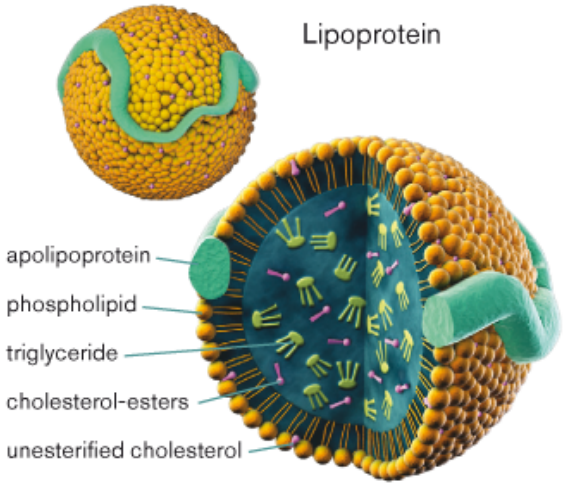
ASOs containing these 2'OMe modifications display enhanced nu-clease resistance in addition to lower toxicity and increased hybridization affinities.

RNase H, an endogenous enzyme found in all cells, specifically cleaves RNA linkages in a double-stranded RNA:DNA heteroduplex



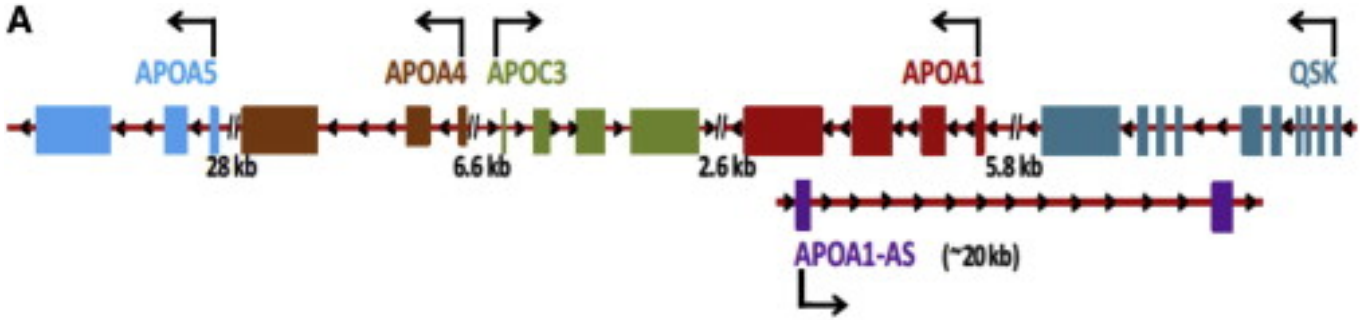
Long ncRNAs and Cardiovascular Disease

Examples: **Apolipoprotein A-1** (APOA-1) is the major protein component of high-density lipoprotein (HDL) in plasma where it plays an important role in *cholesterol efflux*. Elevated HDL levels have been shown to protect against coronary heart disease making APOA1 a strong candidate for pharmacological intervention for patients at risk for heart disease.

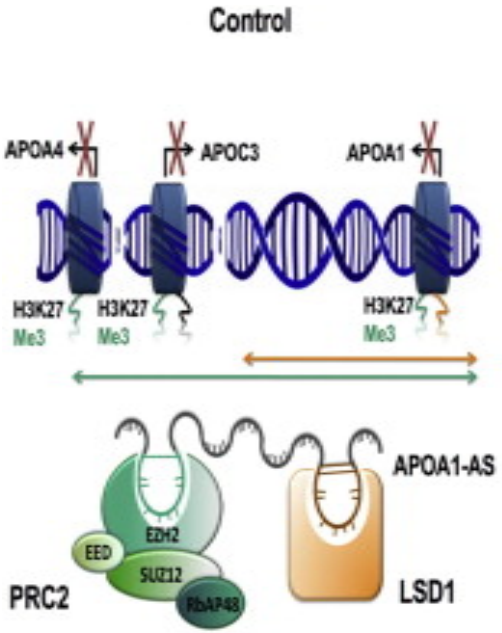


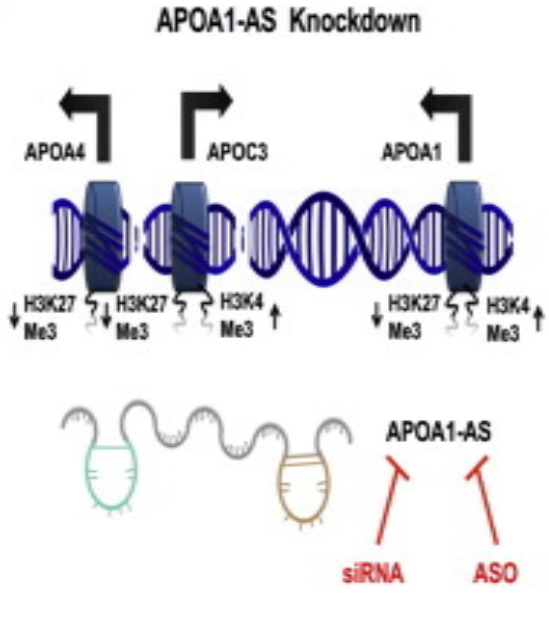
Strategy for gene therapy: Upregulating (by lncRNA KD) Therapeutic Genes

Analysis of the ***APOA1*** gene locus resulted in the identification of a long noncoding antisense transcript, *APOA1-AS*. The *APOA1*-AS transcript regulates *APOA1 in cis* through epigenetic mechanisms. *APOA1*-AS KD upregulates *APOA1*, increases activating H3K4m3 marks and reduces suppressive H3K27.

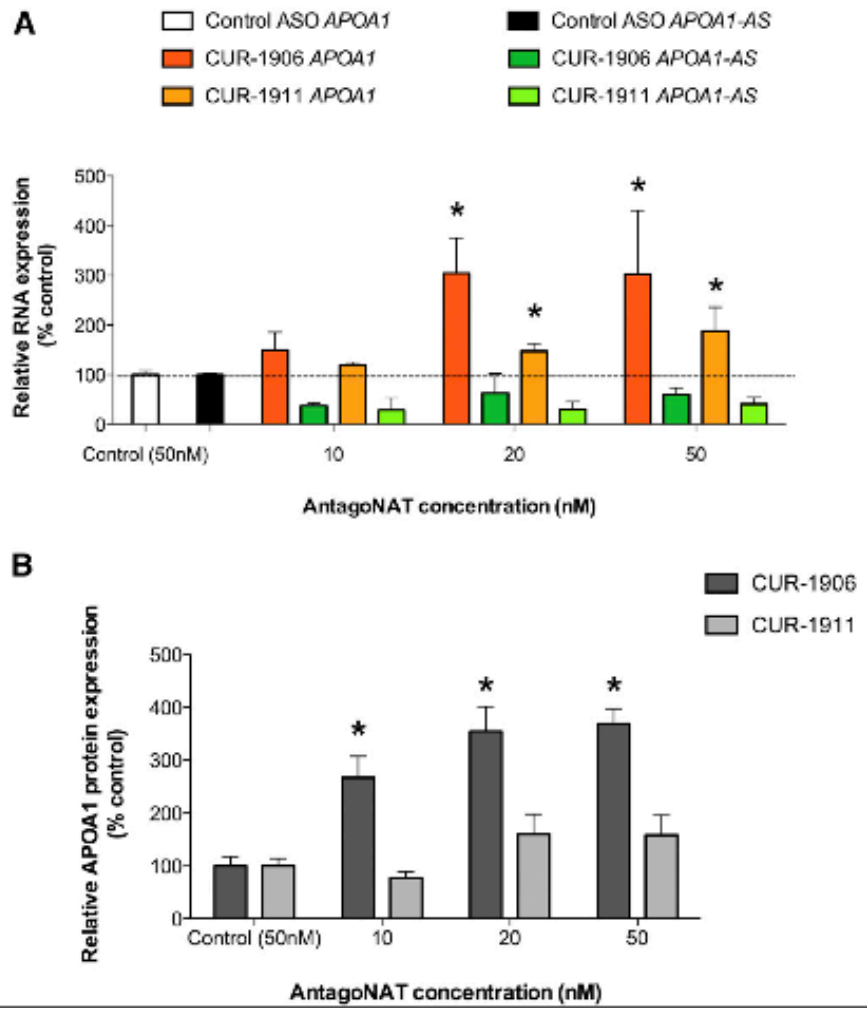


Therefore, *APOA1*-AS can be used to directly modulate APOA1 protein, making it a potential therapeutic target for cardiovascular disease.

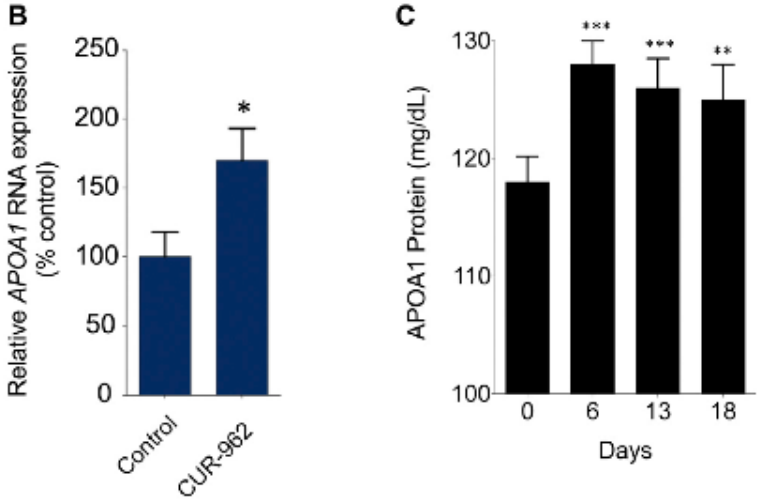




HepG2 cells were transiently transfected with AntagoNATs against APOA1- AS.



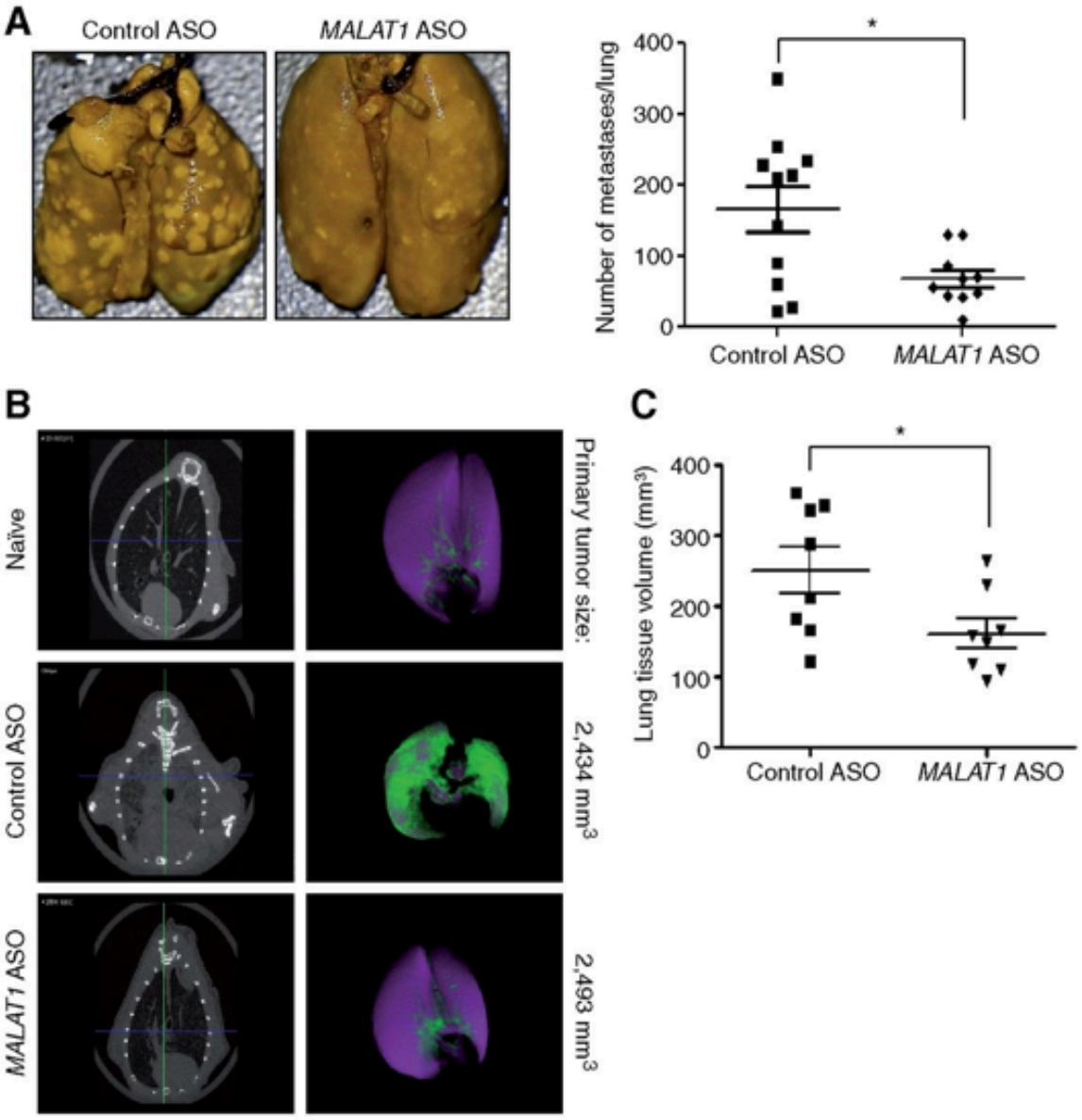
Primary African monkey hepatocytes were transiently transfected with AntagoNATs against APOA1-AS



Data show the effectiveness of a relatively low dose of CUR-962 in increasing APOA1 mRNA and protein concentration *in vivo*. Potential applicability of Antago-NAT strategies for inducing therapeutic gene upregulation in a clinical setting.

MALAT1 might be a potential therapeutic target

Development of an approach to target *MALAT1 in vivo* by the administration of free-uptake ASOs



A, animals treated with *MALAT1* ASO had significantly fewer tumor nodules in the lung compared with control ASO-treated. **B** and **C**, a significant decrease in tumor volume in the *MALAT1* ASO-treated group compared with the control ASO group. Purple indicates airway volumes and green represents lung tissue volumes including tumor mass.

Long ncRNAs and Neurological Disorders

Examples:

Brain-derived neurotrophic factor (*BDNF*), one of the most abund
factors in the brain, is a protein that is critical for the developme
and maintenance of neurons in the nervous system. Furthermor
therapeutic target for several neurodegenerative and neurodev
disorders in which BDNF is severely downregulated.[86,87,88] The
gives rise to a 2.2 kb antisense transcript, called *BDNF*-AS, that c
regulates *BDNF* mRNA and protein levels *in cis*.[38,89] Blocking
down *BDNF*-AS correlates with a reduction in H3K27me3 at
promoter region and a reduction of enhancer of zeste 2 (EZH
subunit of the PRC2 complex) occupancy, suggesting that *BDNF*-A
repressive effect on the *BDNF* locus by interacting with PRC2
BDNF expression *in cis*. Therefore, *BDNF*-AS or its site of inter
PRC2 could be a viable therapeutic target to upregulate BD
expression.

Strategy for gene therapy: Upregulating (by lncRNA KD) Neuroprotective Proteins

Small molecules modulators of lncRNA-Protein Interactions

The use of small molecules is an alternative therapeutic approach and can be pursued by designing measurable pharmacological assays to quantitate lncRNA-protein interactions. Although lncRNAs lack sequence conservation, they are structurally conserved, making them more stable and predictable physical targets,

Ex: targeting lncRNAs at their site of interaction with chromatin modifying enzymes would:

- increase the specificity of compounds,
- reduce off-target effects
- provide reversible inhibition of chromatin modifying enzymes at non-catalytic domains of the protein.

However, in order to target this specific interaction, it is crucial to understand the biochemistry of how RNAs and proteins interact (CLIP, RIP, PAR-CLIP…). There is a need for the development of target-based pharmacological assays that can quantify lncRNA-protein interactions for the purpose of high throughput compound screening.