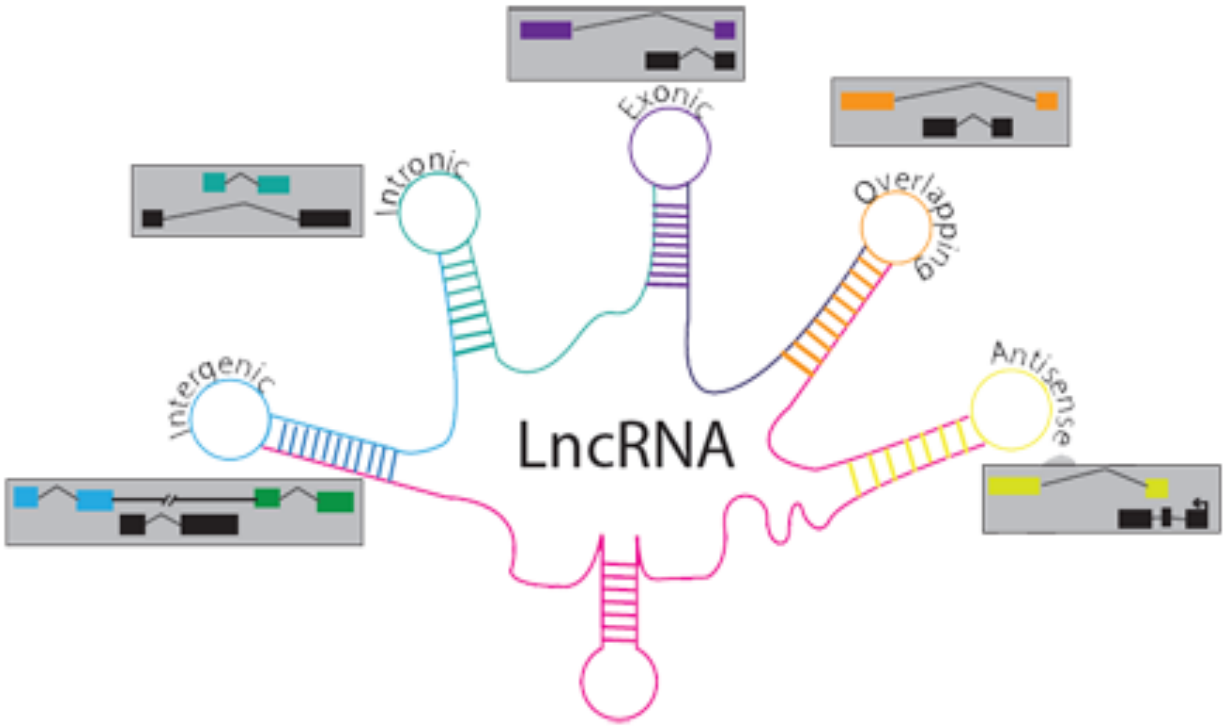
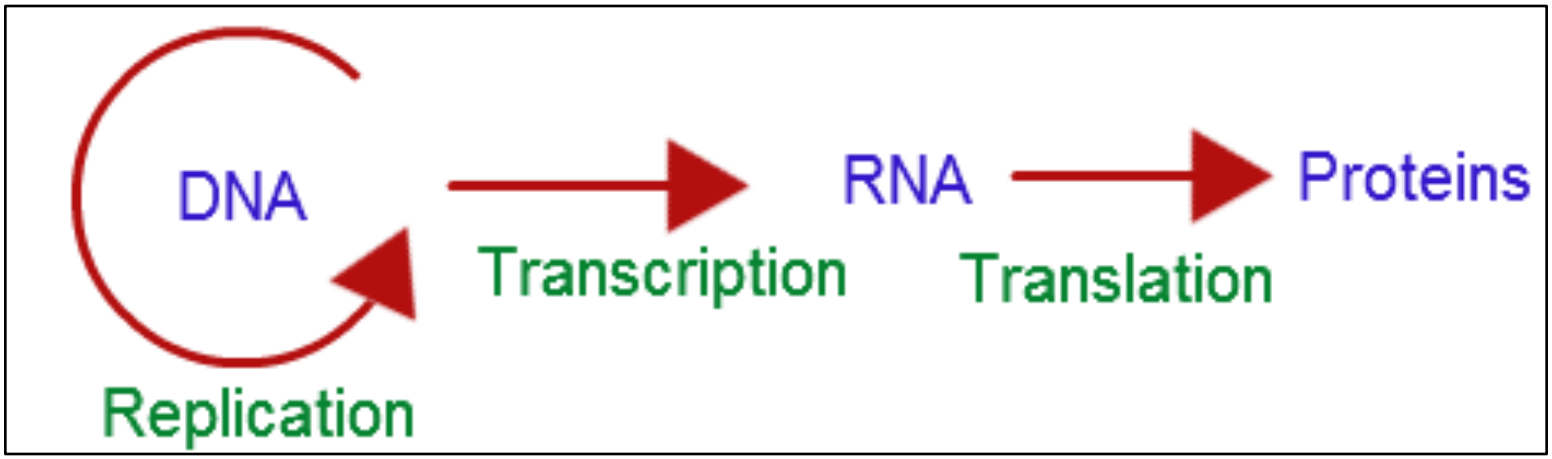
long non coding RNAs



Long Non Coding RNA

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



Why RNA?

Economist.com

The RNA revolution

Biology's Big Bang

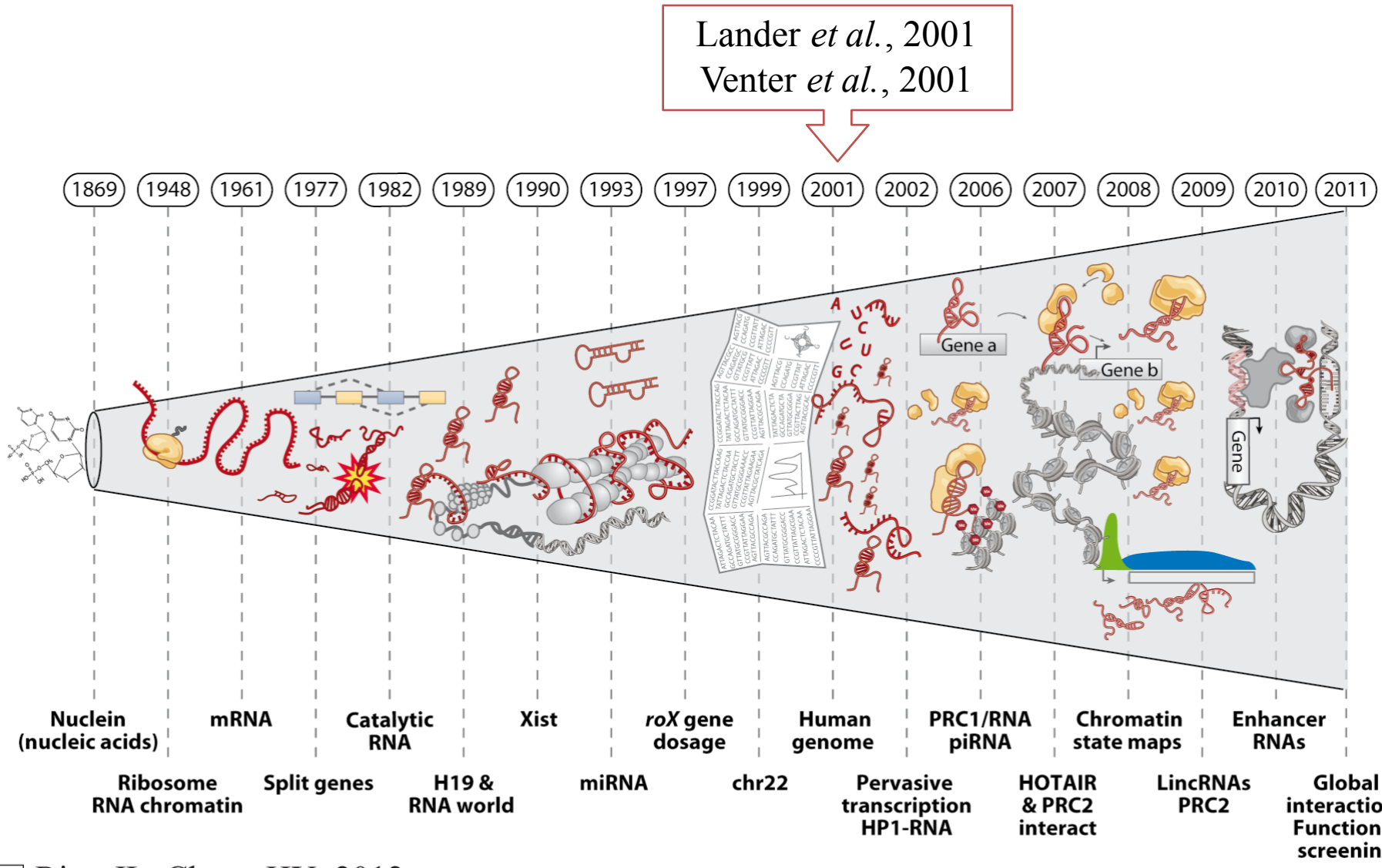
- many processes of gene expression regulation occur at the post-transcriptional level

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

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

The GENOMIC ERA

how many genes in the human genome?



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

The GENOMIC ERA

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

how many genes in the human genome?

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

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

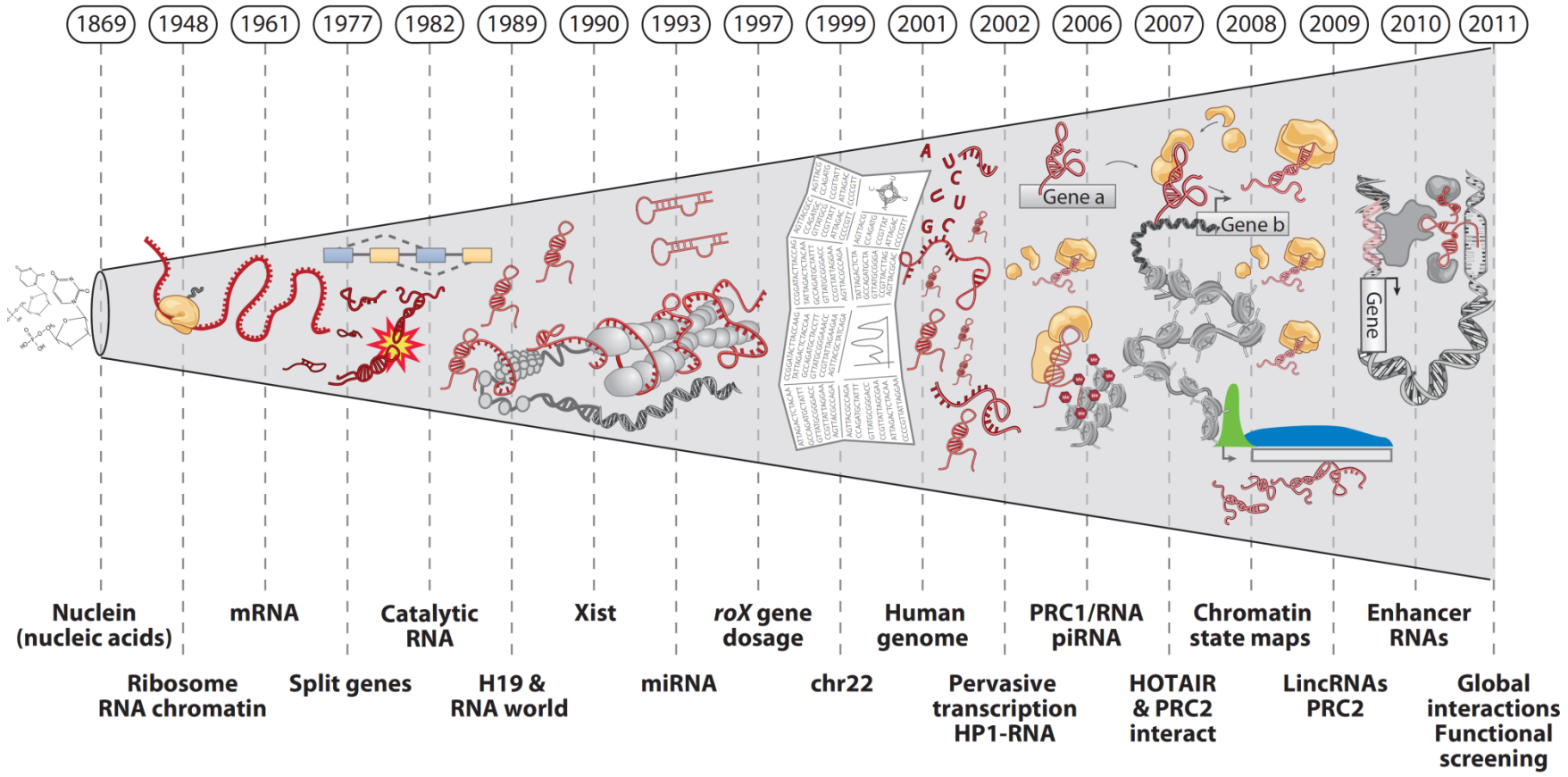
DNA ⟶ RNA ⟶ Protein

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

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

Transcriptome analysis
Characterization of full length transcripts – mapping of 5’ and 3’ ends as well as of alternative splicing events

The FANTOM3 Consortium, 2005



Rinn & Chang *Annu Rev Biochem* 2012

Genome Organization

Antisense Transcription in the Mammalian Transcriptome

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

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

>70% of protein encoding genes present antisense transcription

The Antisense Transcriptomes of Human Cells

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

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

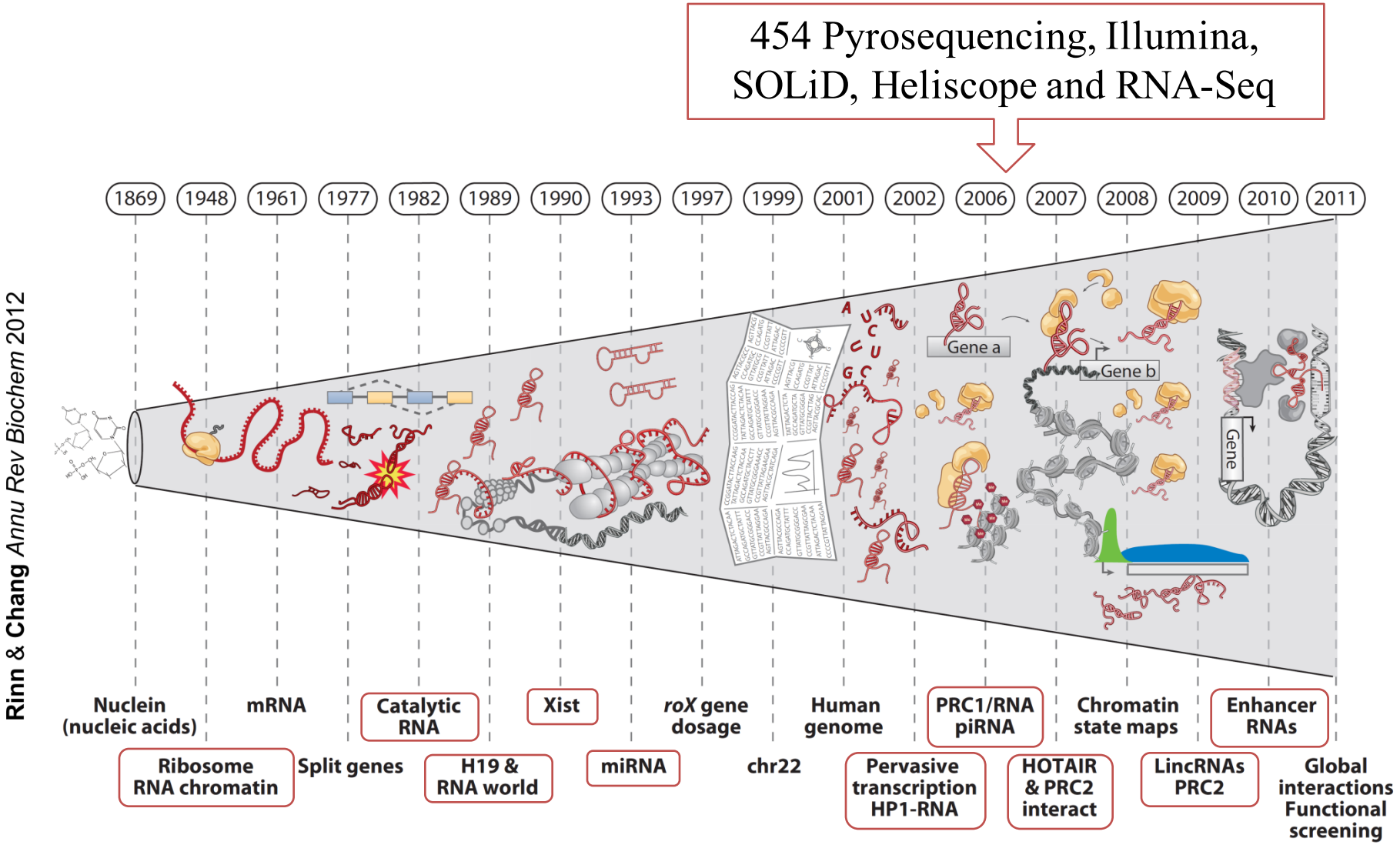
Genome-wide antisense transcription drives mRNA processing in bacteria

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

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

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

Transcriptome analysis – deep sequencing technologies – identification of low abundance transcripts



Transcriptome analysis

ENCODE

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

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

FANTOM 5

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

CAGE analysis of the following libraries:

573 human primary cell samples

128 mouse primary cell samples

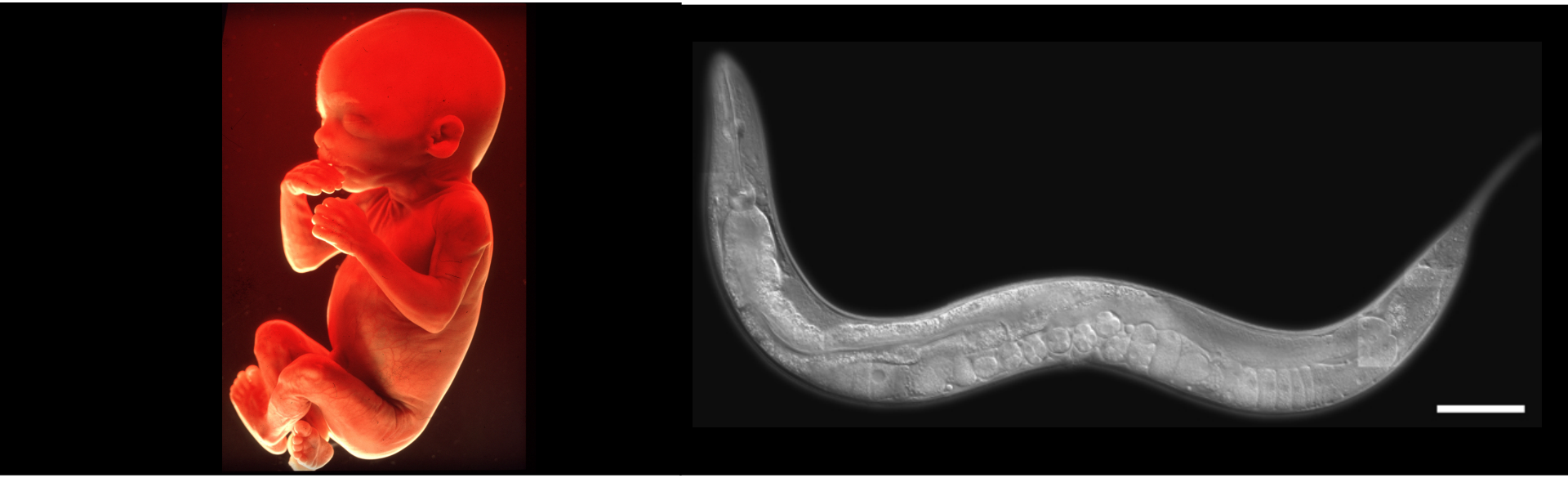
250 different cancer cell lines samples

152 human post-mortem tissues samples

271 mouse developmental tissue samples

22000 genes encoding for proteins

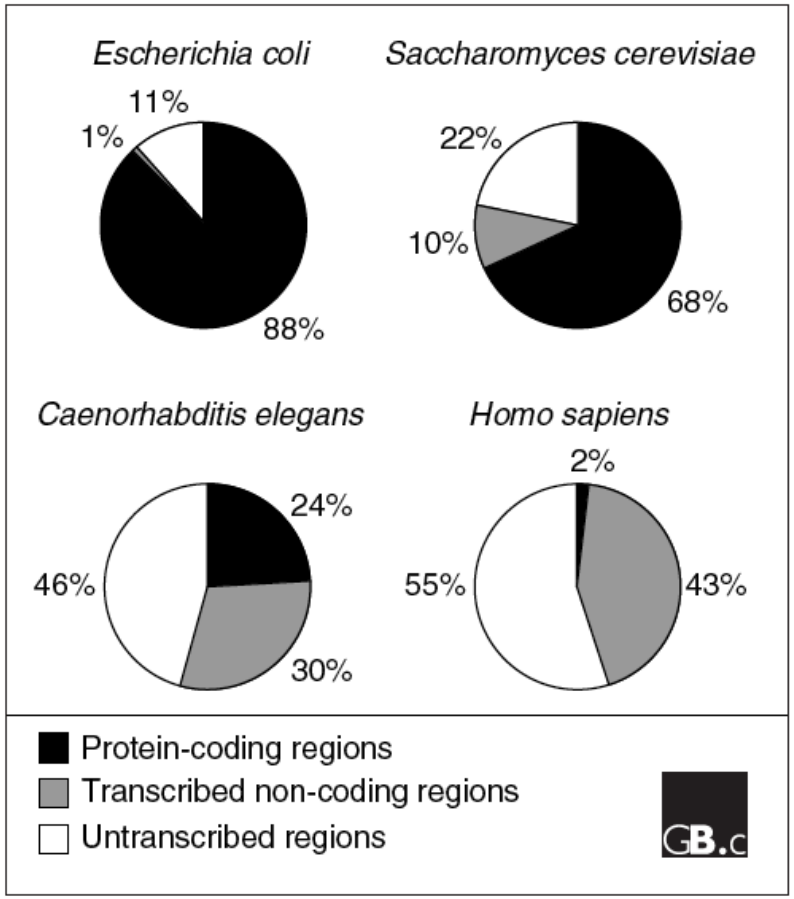
The genetic basis of developmental complexity



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

Both have approximately **20.000 proteins**

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



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

Transcriptome analysis

ENCODE

ARTICLE

doi:10.1038/nature11247

An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

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

FANTOM5

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

CAGE analysis of the following libraries:

573 human primary cell samples

128 mouse primary cell samples

250 different cancer cell lines samples

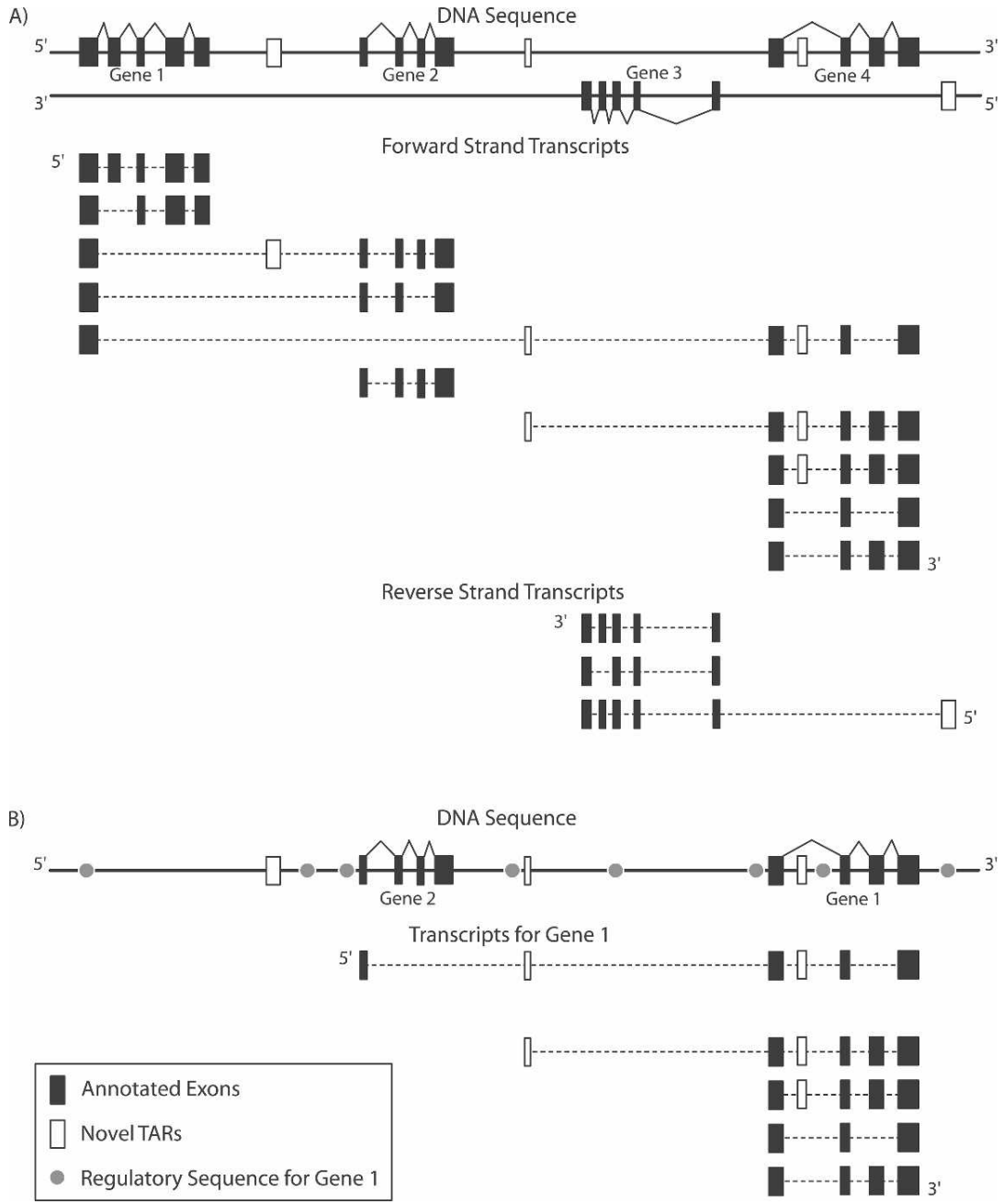
152 human post-mortem tissues samples

271 mouse developmental tissue samples

22000 genes encoding for proteins

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

The ENCODE Project consortium (2007)



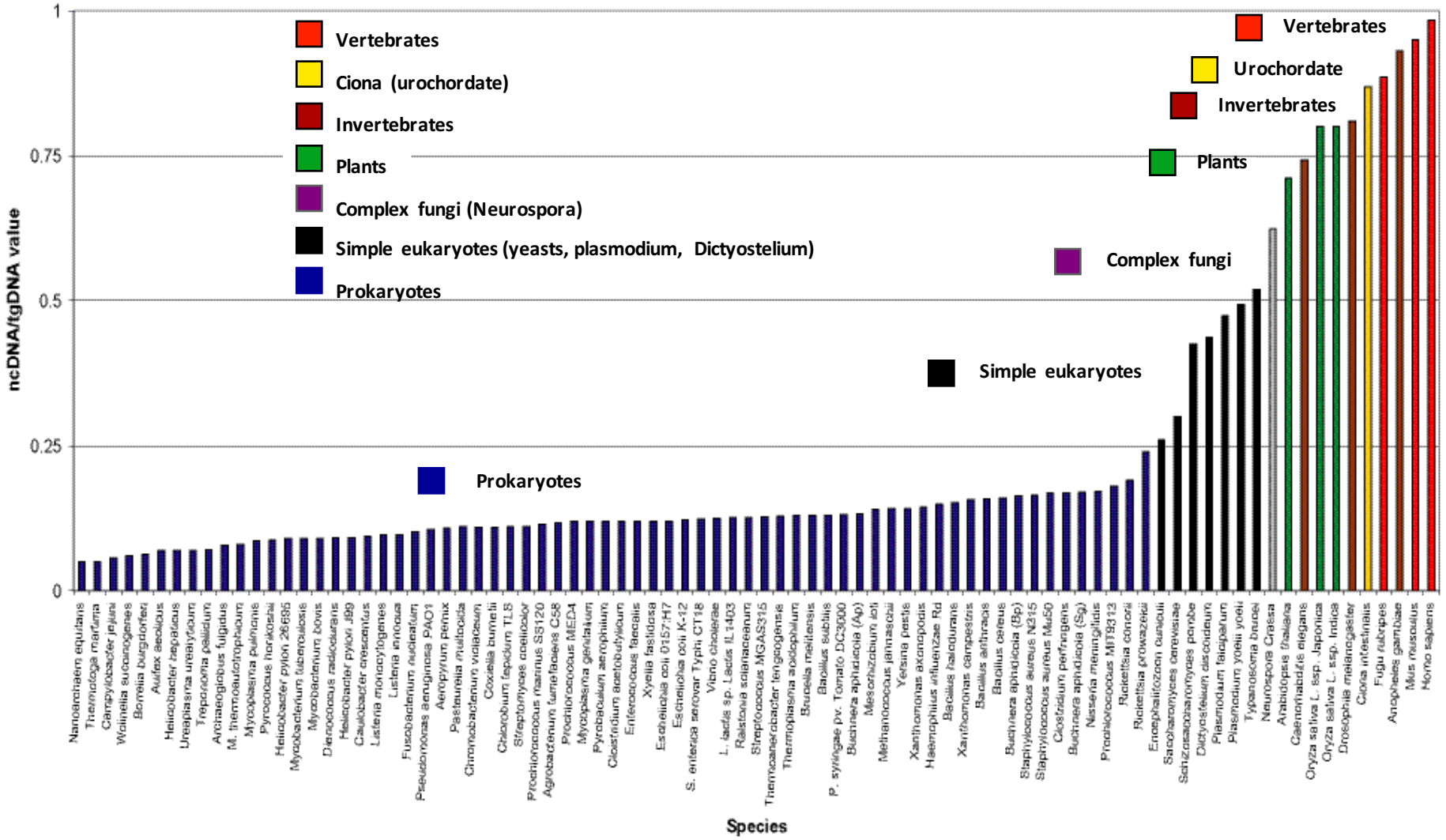
1) A vast amount of DNA, not annotated as known genes, is transcribed.

2) there are a large number of **unannotated <u>transcription start sites</u> (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**.

Lattices of long transcripts and dispersed regulation

The proportion of noncoding DNA broadly increases with developmental complexity

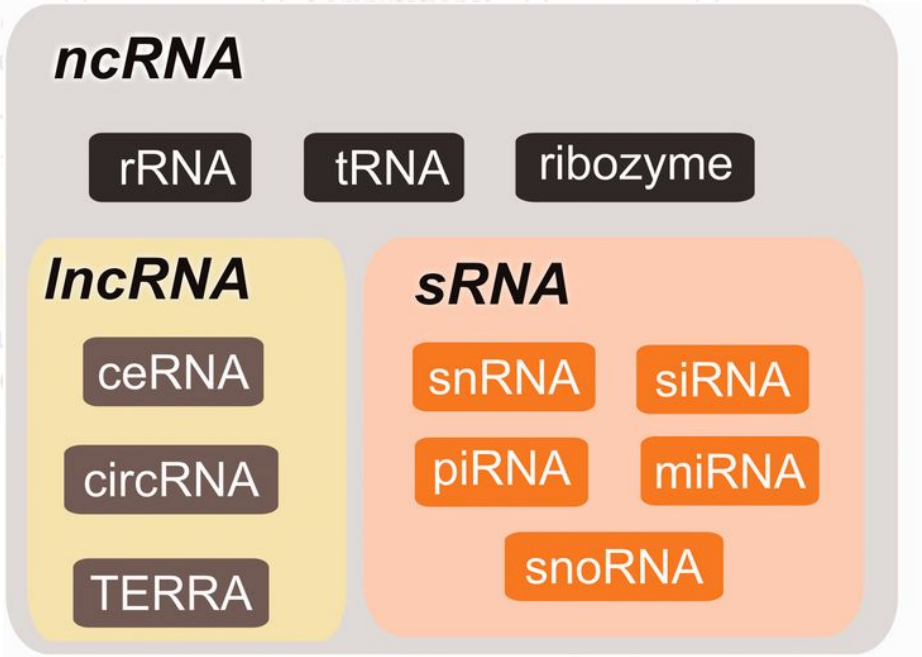


J.S. Mattick *Nature Reviews Genetics* 5, 316-323 (2004).

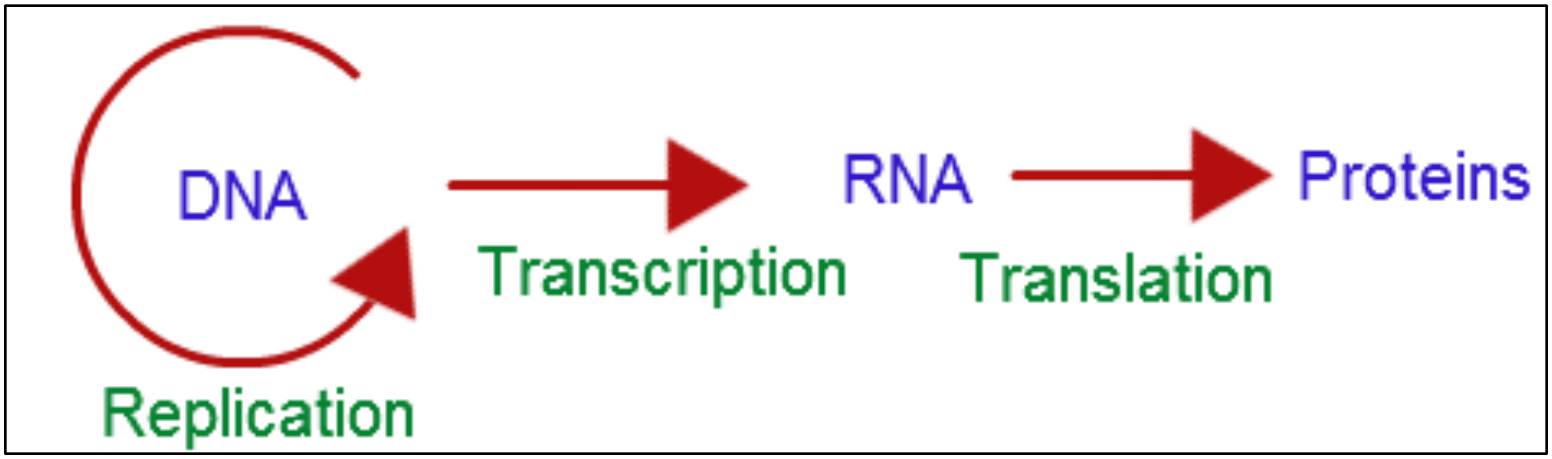
R.J. Taft, M. Pheasant and J.S. Mattick, *Bioessays* 29, 288-299 (2007)

Major highlights:

- Human genome is pervasively transcribed.
- A large fraction of the non coding portion is **functional**



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



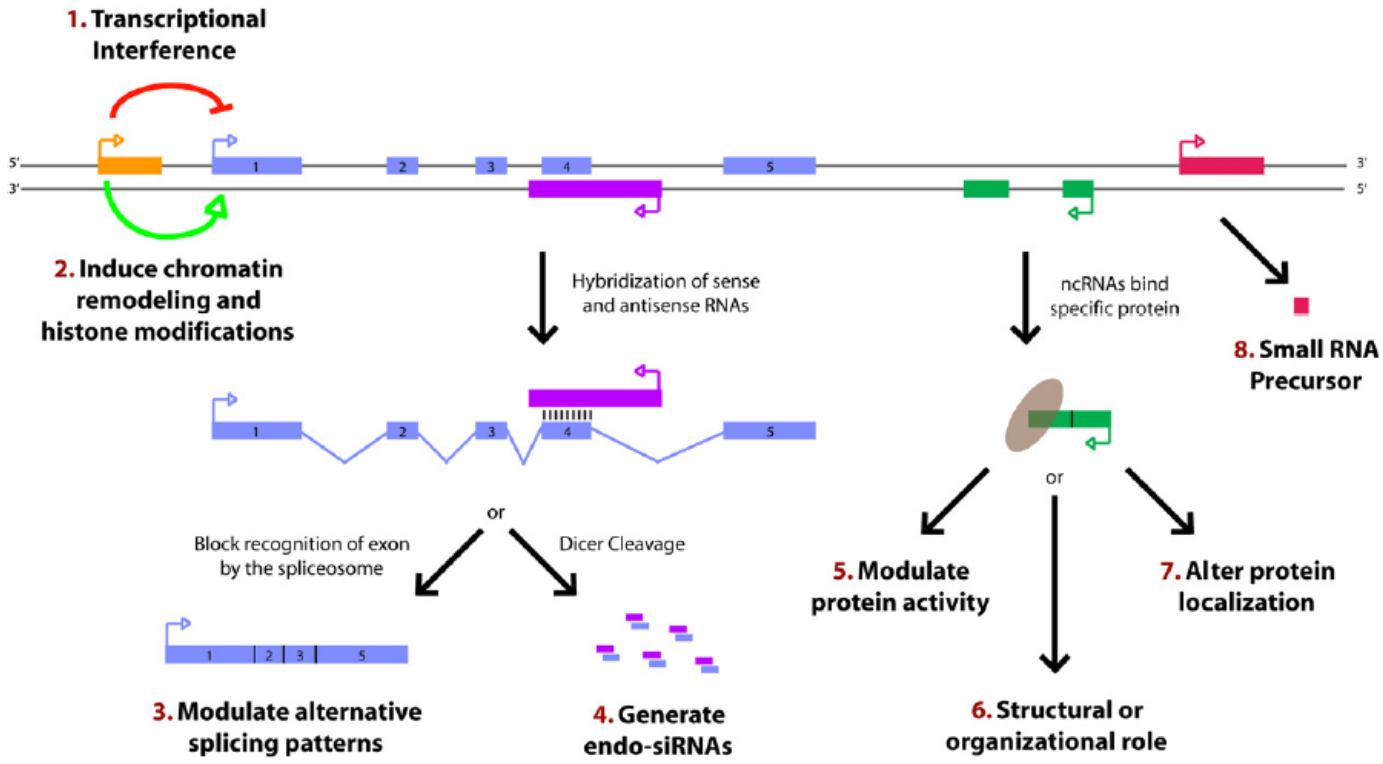
centrality of RNA in gene regulation

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).

Long non-coding RNAs

2001 to the Future Break the Dogma

List of long non-coding RNAs



J E Wilusz et al Genes & Development 2009

"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.

NO!

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

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

=

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

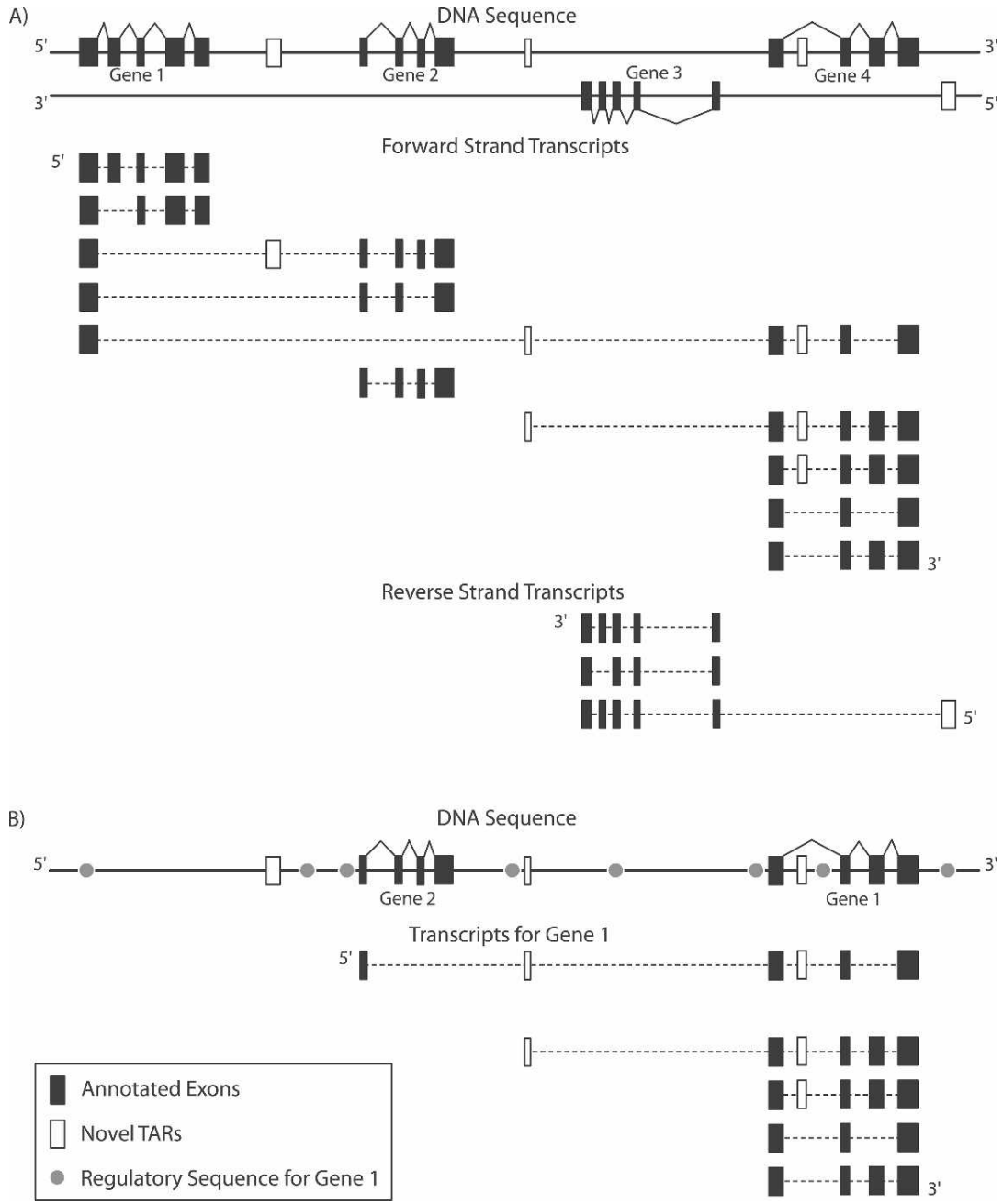
152 human post-mortem tissues samples

271 mouse developmental tissue samples

80% of the genome is functional

>40000 long non-coding RNAs and growing…

The ENCODE Project consortium (2007)



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

2) there are a large number of **unannotated <u>transcription start sites</u> (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**.

Lattices of long transcripts and dispersed regulation

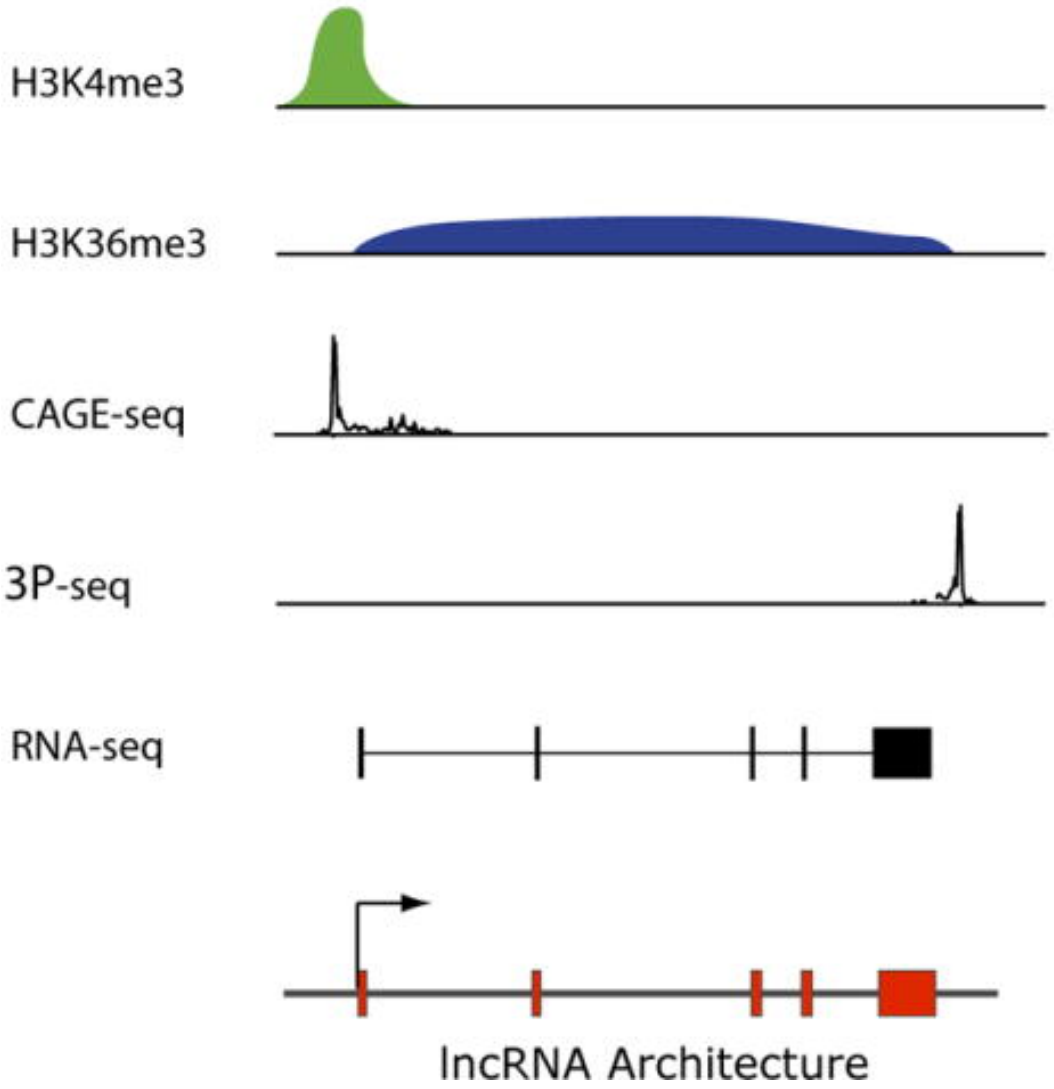
What is a gene, post-ENCODE?

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.

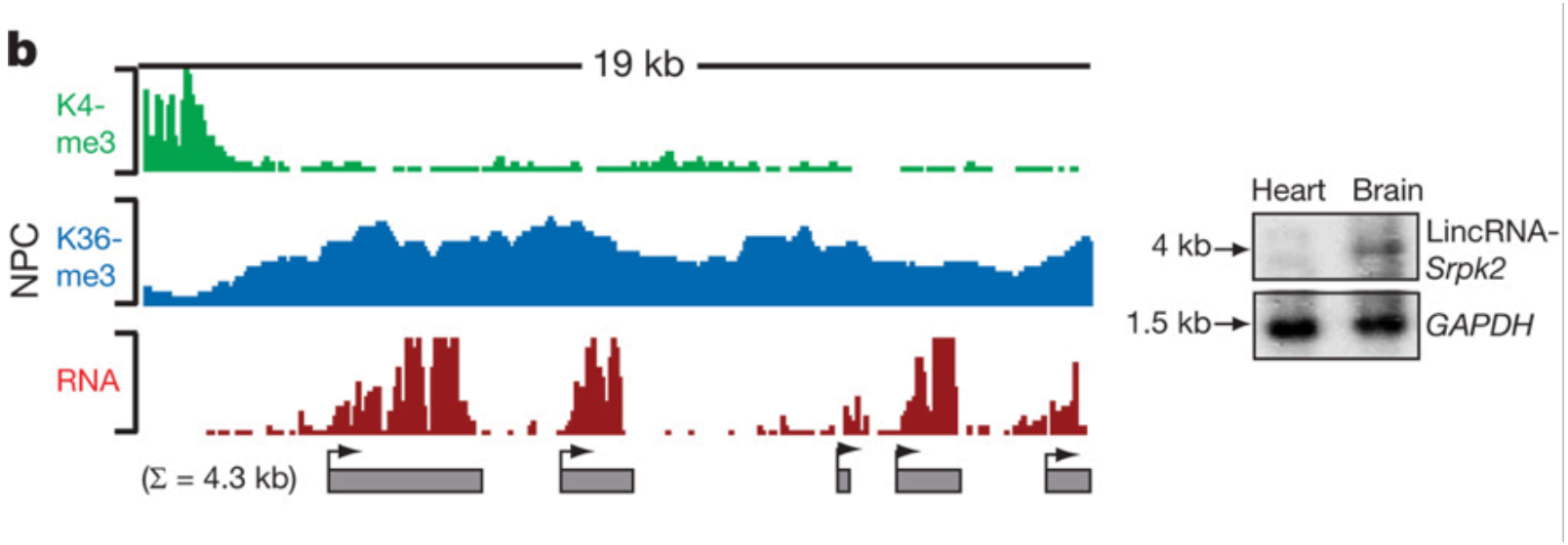
Less of a distinction to be made between *genic* and *intergenic* regions.

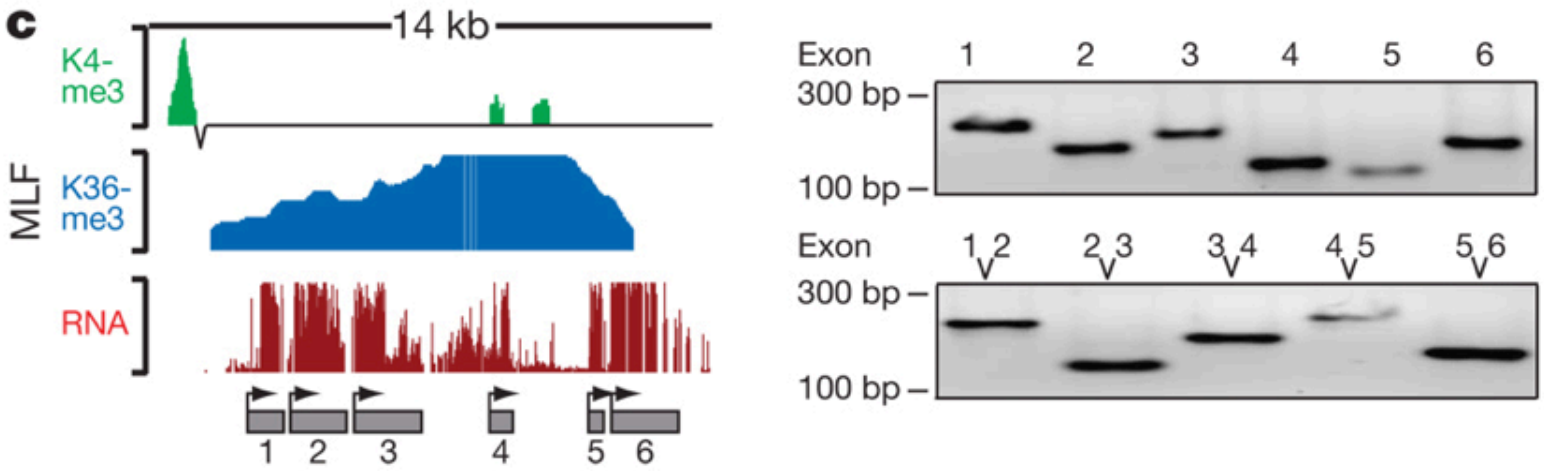
A critical clue for hunting RNA genes came from chromatin



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



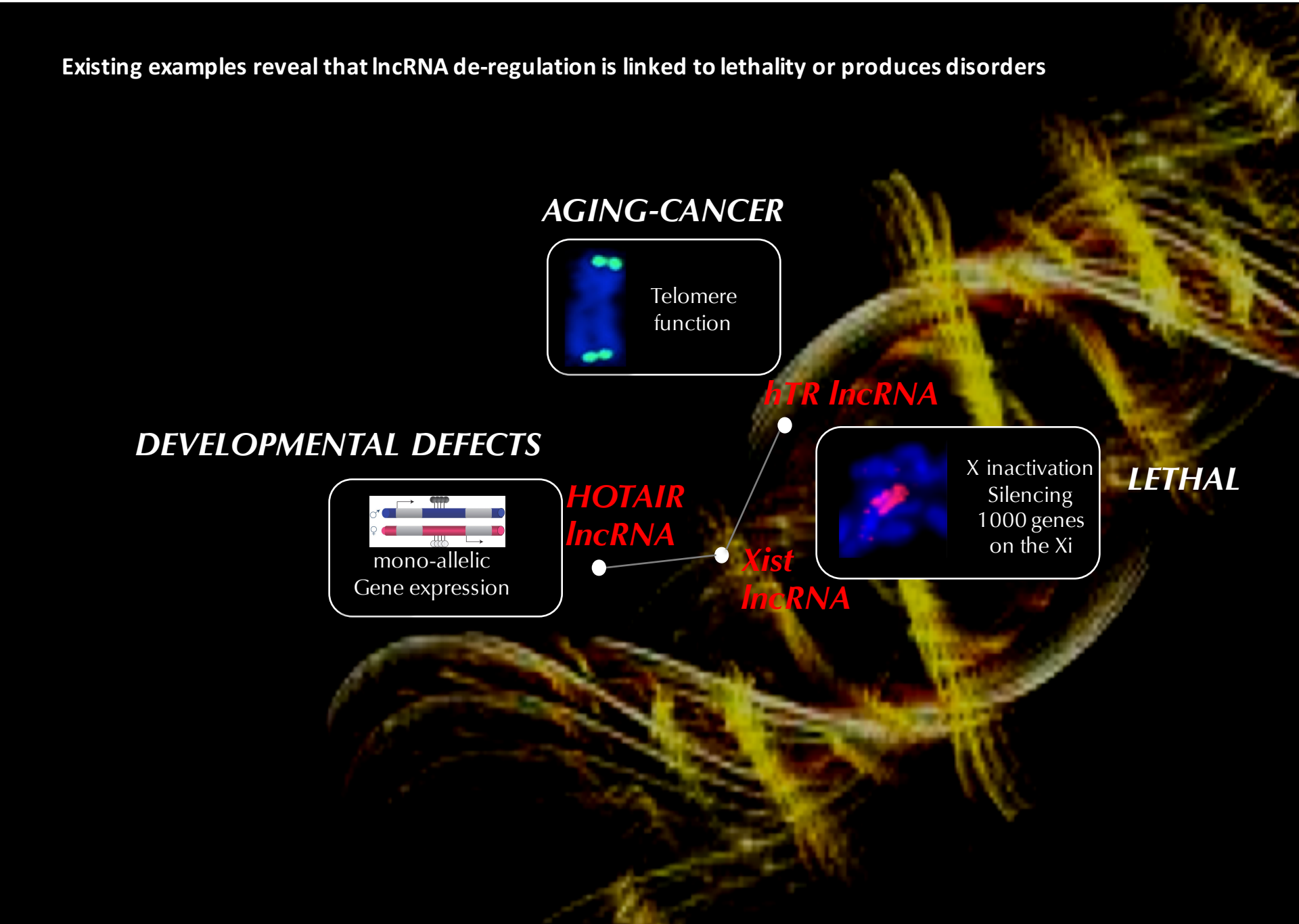


LncRNAs are functional transcripts

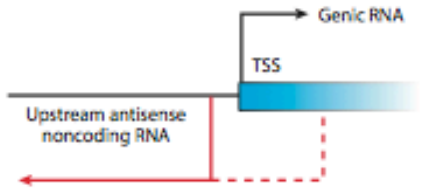
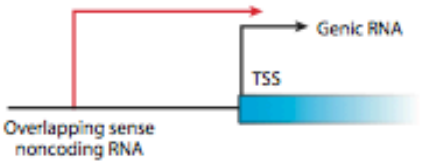
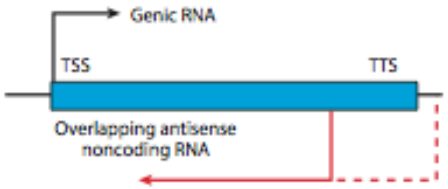
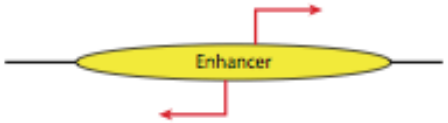
- They are *multi-exonic*, *capped*, *spliced* and *polyadenylated* transcripts
- They can be localised to the *nucleus*, *cytoplasm* (or both)

Why to study lncRNAs?

Existing examples reveal that lncRNA de-regulation is linked to lethality or produces disorders



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.

A variety of functions for lncRNAs

NUCLEUS

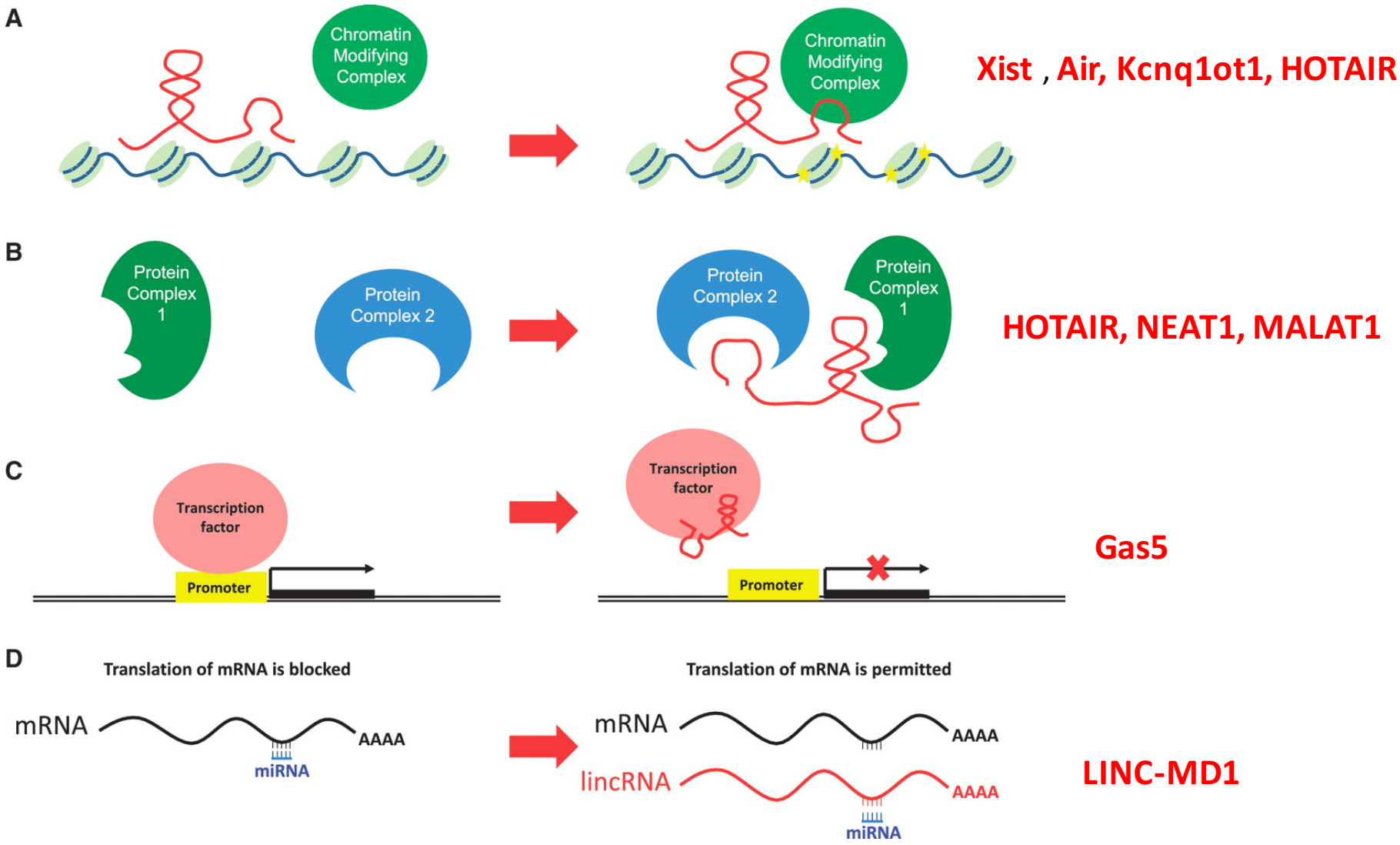
- Chromatin Remodeling → **Xist** – X chromosone inacivation; **Air, Kcnq1ot1, HOTAIR,HOTTIP** – genomic imprinting
- Transcription Regulation → **Gas5** – negative regulator of glucocorticoid receptors
- Nuclear Architecture → **NEAT1** – formation and maintainance of paraspeckles
- Post-transcriptional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

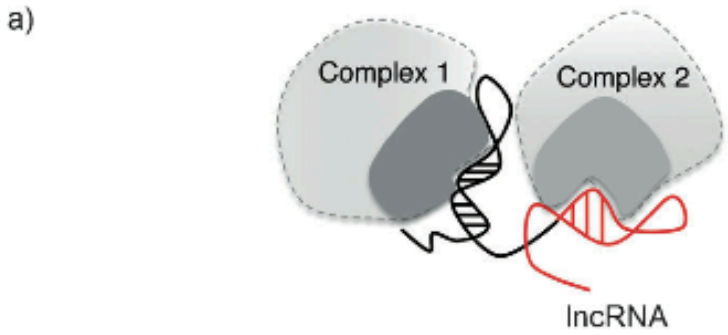
…***why*** there is such a *little* sequence conservation among lncRNAs?

- Sequence conservation average is *lower* than in coding gene but there are *peaks* in specific functional regions

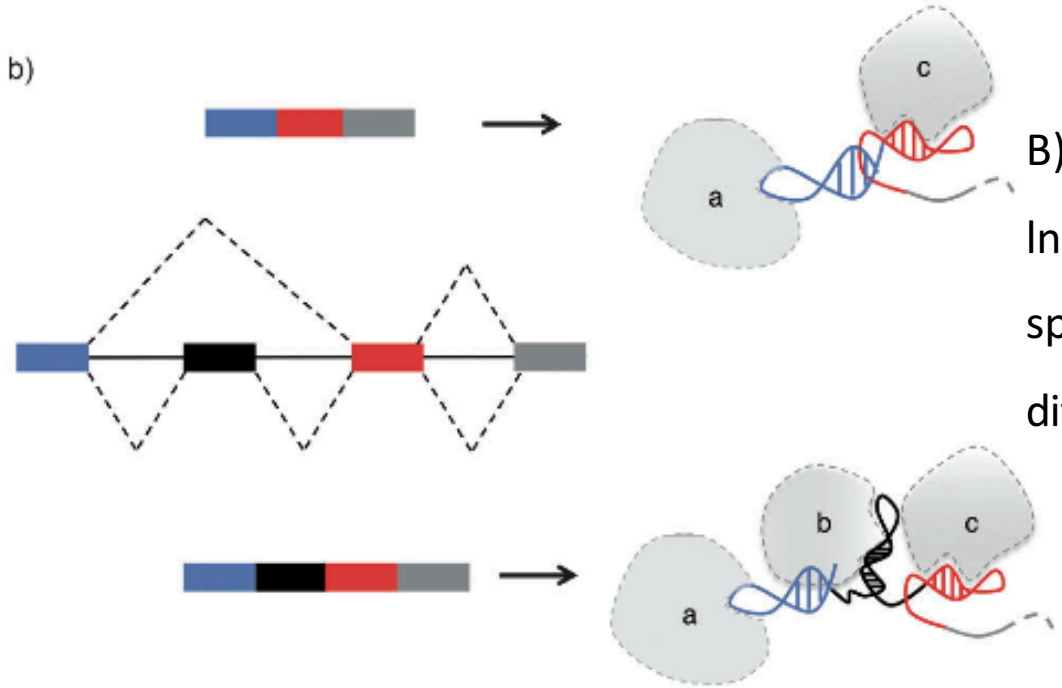
- Many lncRNAs can have functions depending on ***their 3D structure***, which is difficult to predict

- Sequence-dependent functions with respect to structure are easier to discover (sequence analysis) and to study (mutation analysis)

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.

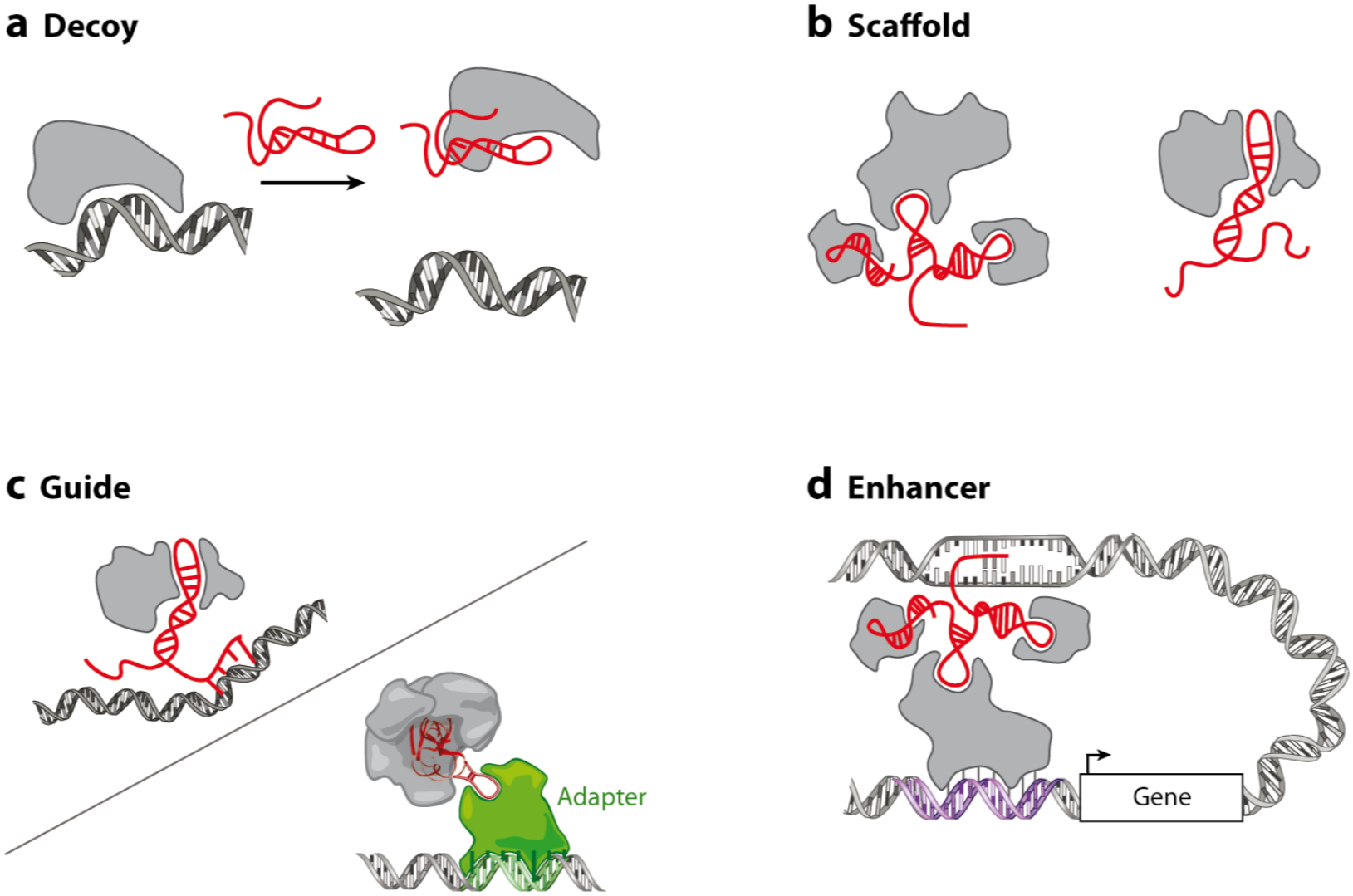


B) Independent structural domains of lncRNAs can be combined by alternative splicing, giving rise to lncRNAs with different scaffold properties.

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]

lncRNA mechanisms of action



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

LncRNAs controlling the access/dismissal of regulatory proteins to chromatin

TABLE 1

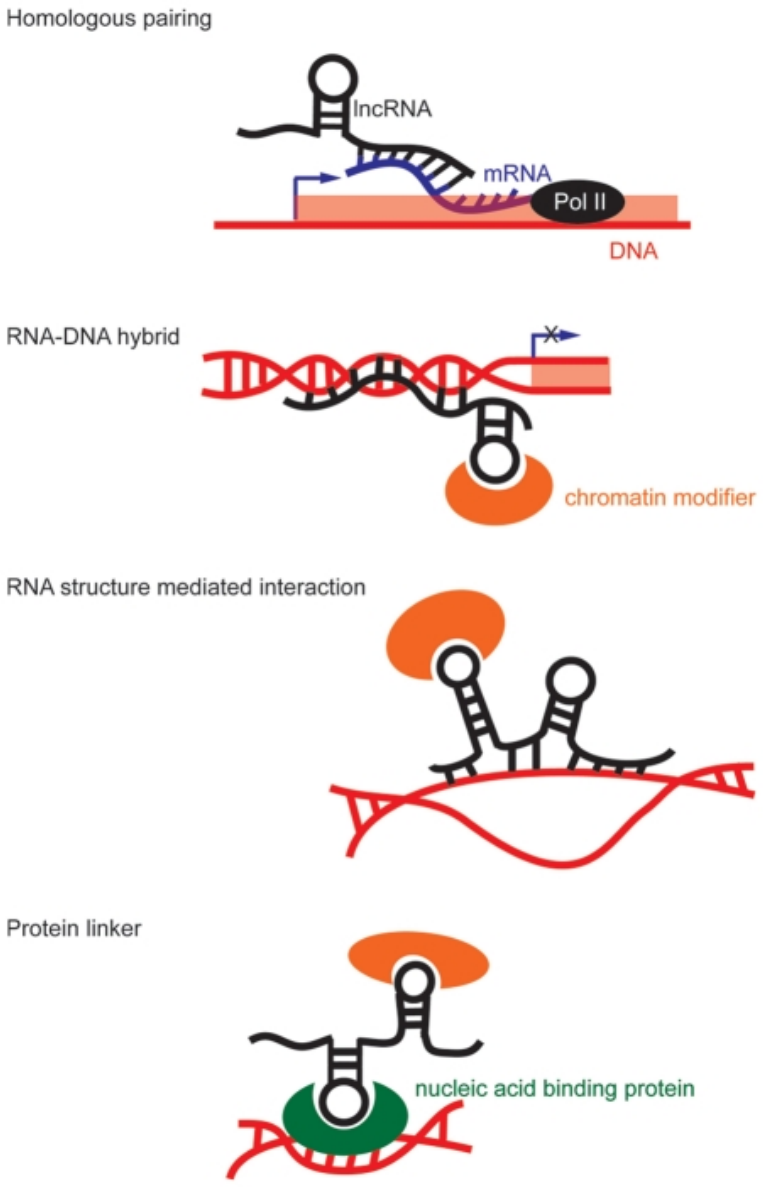
Protein partners of lncRNAs

Many of these lncRNA-protein complexes function at chromatin.

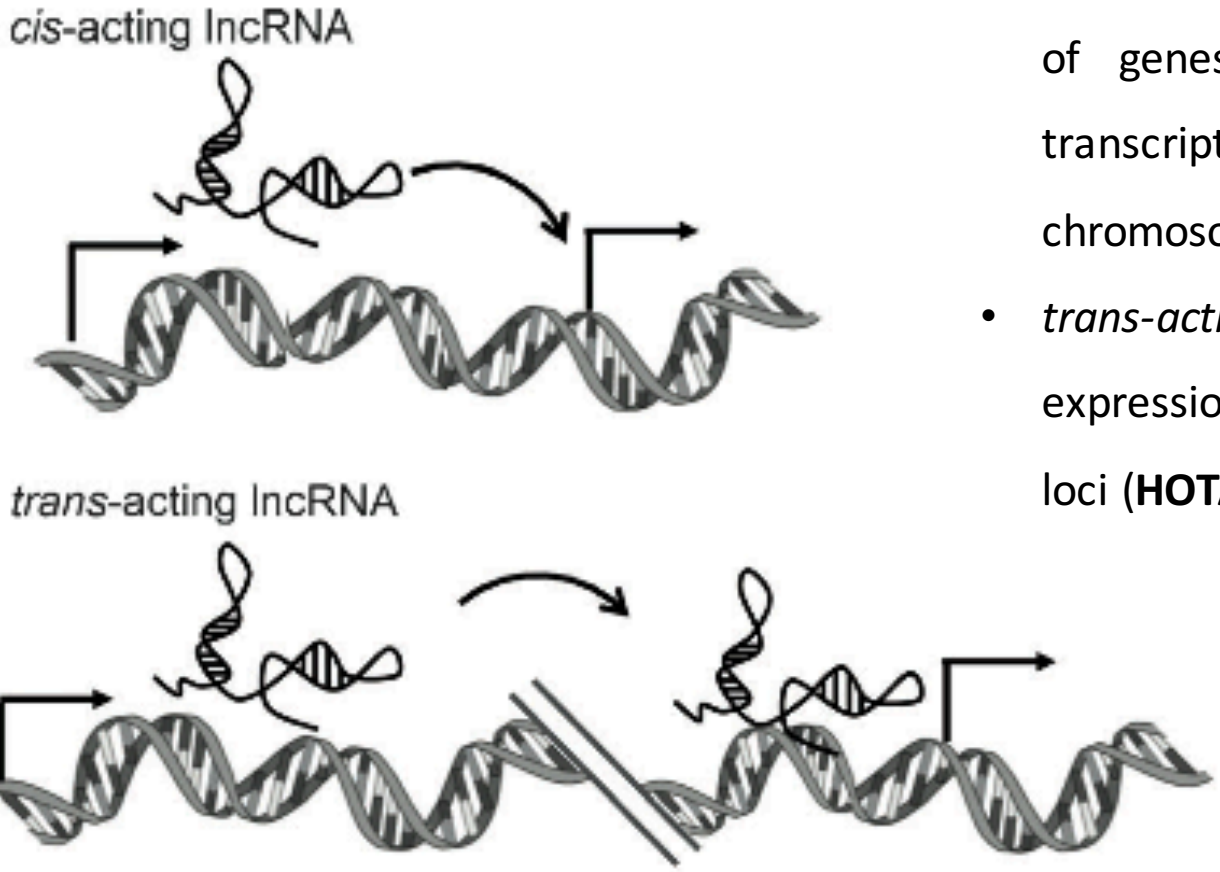
Protein	LncRNA	LncRNA Function	Refs
CTCF	SRA	Enhances insulator function of CTCF.	(92)
DNMT3b	pRNA	Targets DNMT3b in cis to rRNA locus via RNA:DNA:DNA triplex for cytosine methylation and gene silencing.	(90)
G9a	Kcnq1ot1, Air	Targets the H3K9 methylase G9a in cis for imprinting.	(87, 88)
Glucocorticoi 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



Nuclear lncRNAs can be classified as:



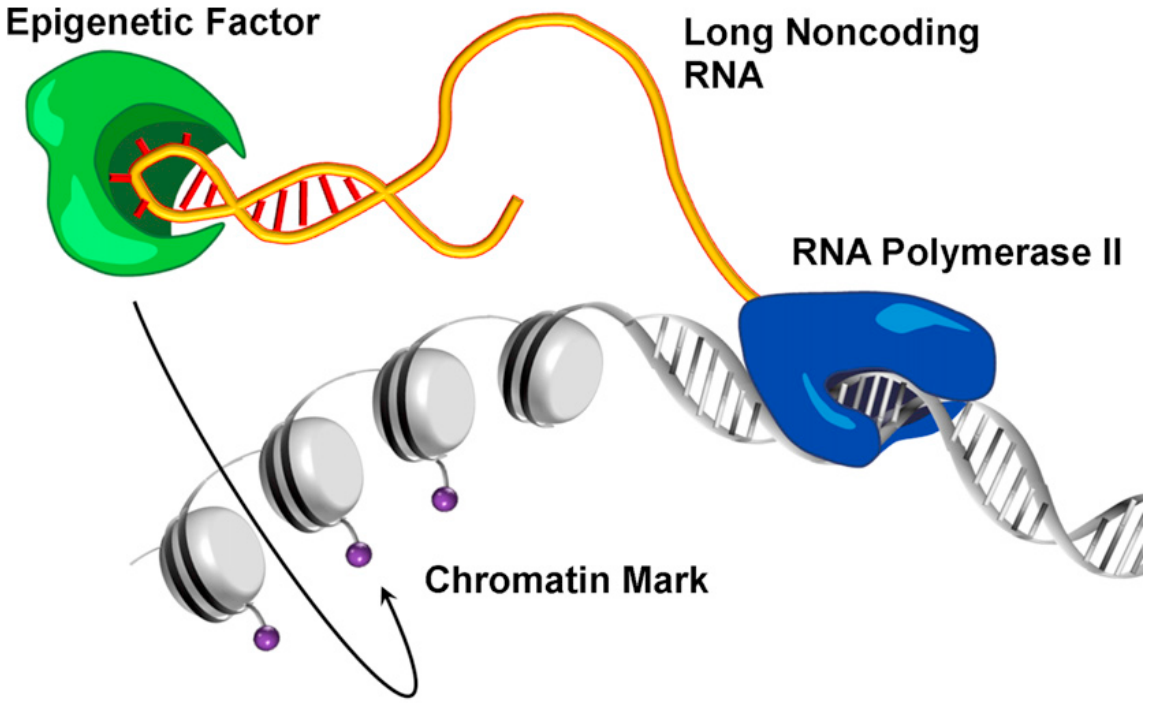
- *cis-acting*, which regulate expression of genes in the vicinity of their transcription site or on the same chromosome (**HOTTIP**, Mistral)
- *trans-acting*, which regulate expression of genes at independent loci (**HOTAIR**).

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]

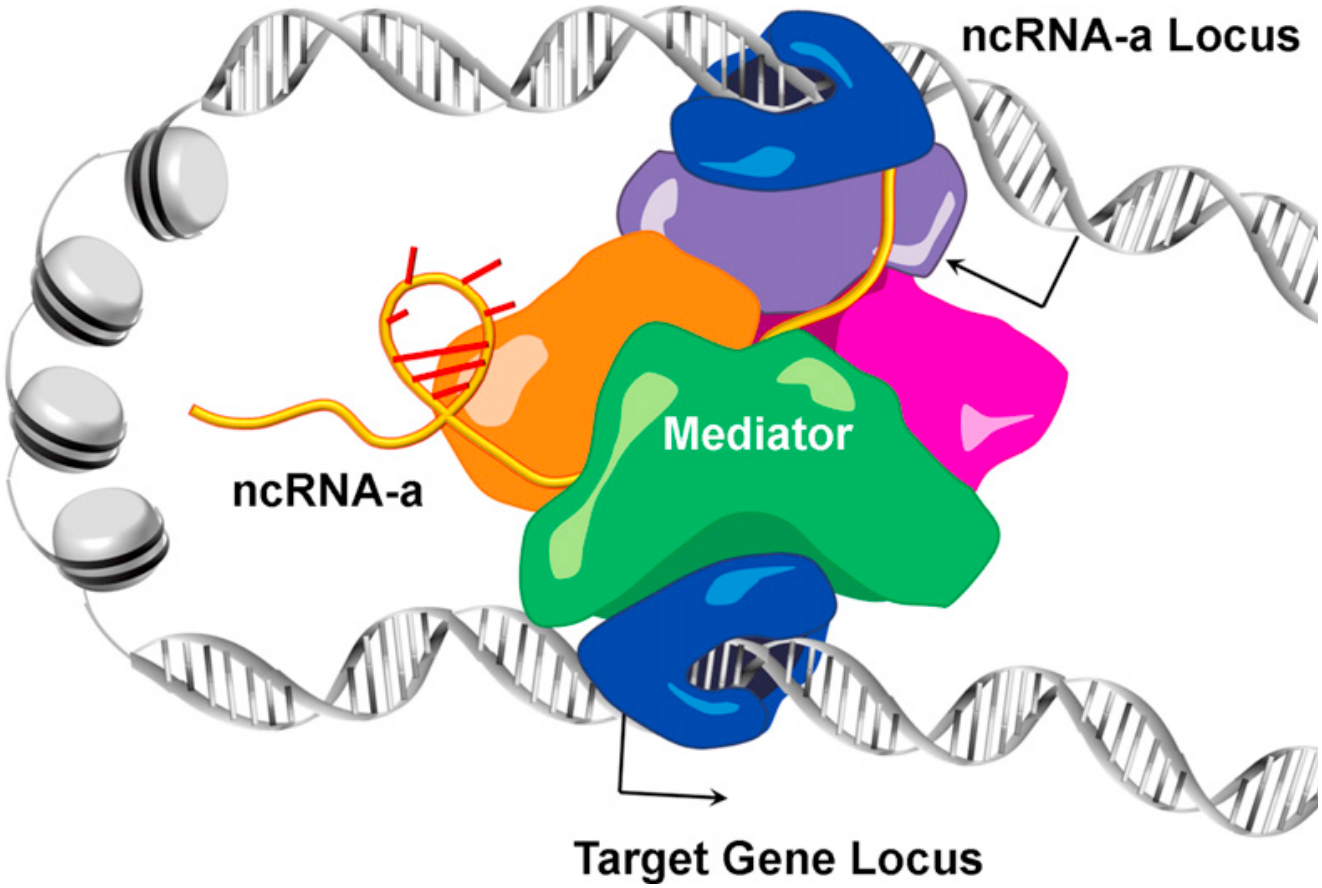
Cis lncRNAs are uniquely suited to act as allele- and locus-specific recruiters by virtue of their:

- *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).

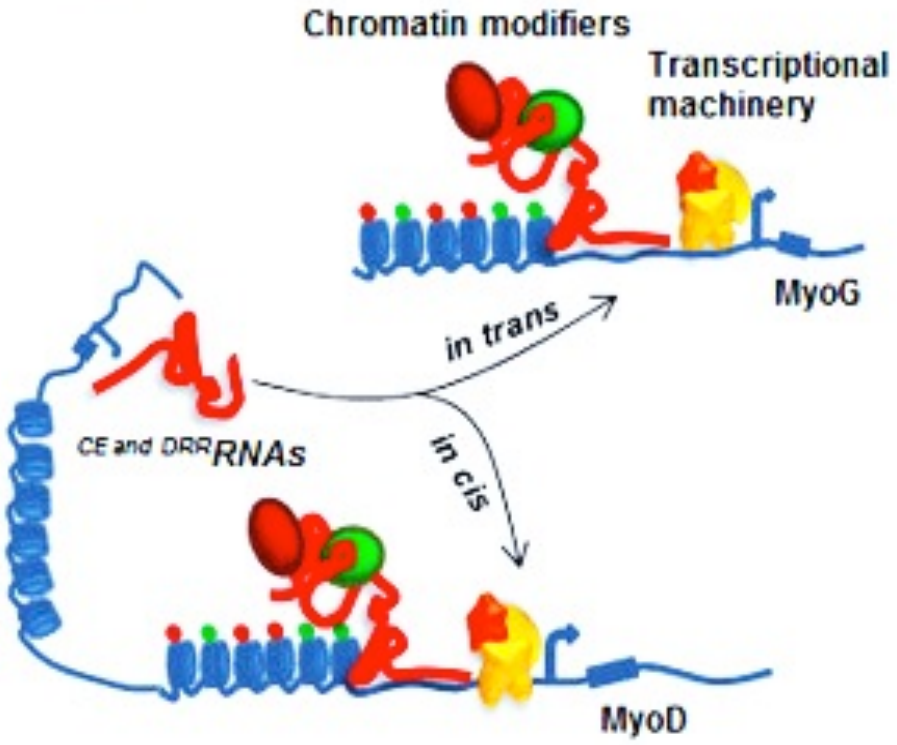


Mediator thus acts as a bridge between transcription factors binding at distant enhancers and the RNA polymerase II (RNAPII) apparatus at target promoters.

A) eRNA

Enhancer lncRNAs (eRNAs)

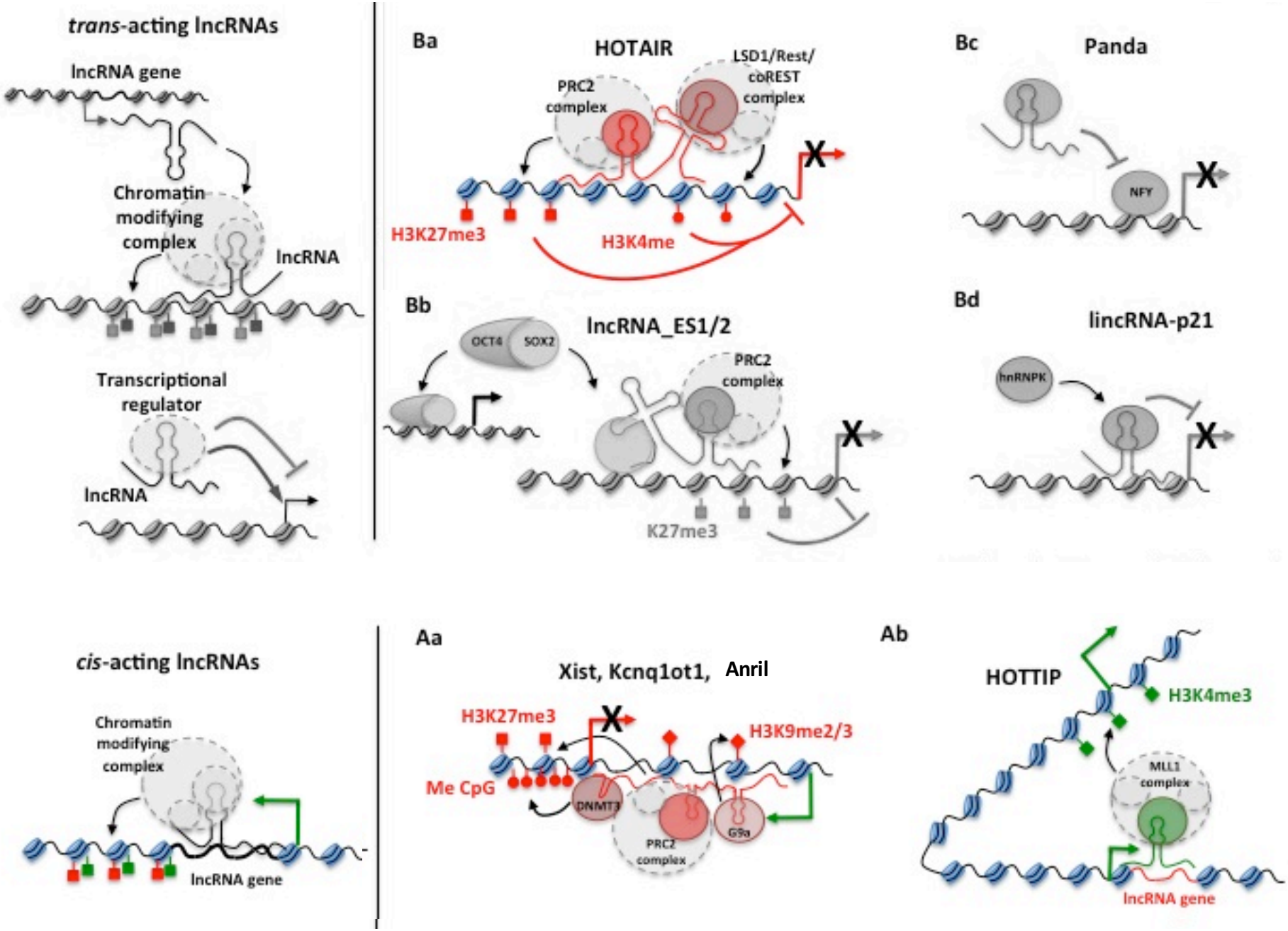
In muscle differentiation:



- Extensive MyoD and MyoG occupancy in the extragenic regions
- Two eRNAs are generated by upstream regulatory regions of MyoD (**CE** and **DRR**) regulate the expression of MyoD and MyoG: distal regulatory regions (DRR) and core enhancer (CE) lncRNAs
- They differ in their mode of action: while the **CERNA** functions in *cis* to activate expression of MyoD, **DRRRNA** works in trans to promote MyoG transcription and muscle differentiation.

At their site of action, both eRNAs mediate increased chromatin accessibility and recruitment of RNAPII.

eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci.
Mousavi K, Zare H, Dell'orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager GL, **Sartorelli** V.
Mol Cell. 2013 Sep 12;51(5):606-17.


trans-acting lncRNAs
lncRNA gene
Chromatin modifying complex
lncRNA
Transcriptional regulator
lncRNA
Ba
HOTAIR
PRC2 complex
LSD1/Rest/ coREST complex
X
H3K27me3
H3K4me
Bb
lncRNA_ES1/2
OCT4
SOX2
PRC2 complex
X
K27me3
Bc
Panda
NFY
X
Bd
lincRNA-p21
hnRNPK
X
cis-acting lncRNAs
Chromatin modifying complex
lncRNA gene
Aa
Xist, Kcnq1ot1,
Anril
H3K27me3
X
H3K9me2/3
Me CpG
DNMT3
PRC2 complex
G9a
Ab
HOTTIP
H3K4me3
MLL1 complex
lncRNA gene

Histone post-transcriptional modifications and readout

1. Type of modification
   - Which amino-acid
   - Number of modifications (me)
2. Position in genome
   - Promoter: H3K36me, H3K9me are repressive
   - Coding region: H3K36me, H3K9me are activating and prevent cryptic initiation of transcription in ORF
3. Other histone modifications
   - combinatorial (occur together)
   - H3K4me + H3K9me: transcriptional activation
   - H4K20me + H3K9me: heterochromatin formation
   - H3K27me + H3K4me: "bivalent" mark in stem cells
4. Size of histone modification domain
   - large: heritable (can be copied more easily)
     - H3K27me can recruit PRC2 has H3K27me3 activity
     - H3K4me recruits WDR5 (MLL thrithorax): H3K4me
5. Cycles of modifications
   - H2Bub → H2B required for transcriptional elongation

Histone code

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 enzymes

which *lack* DNA binding capacity

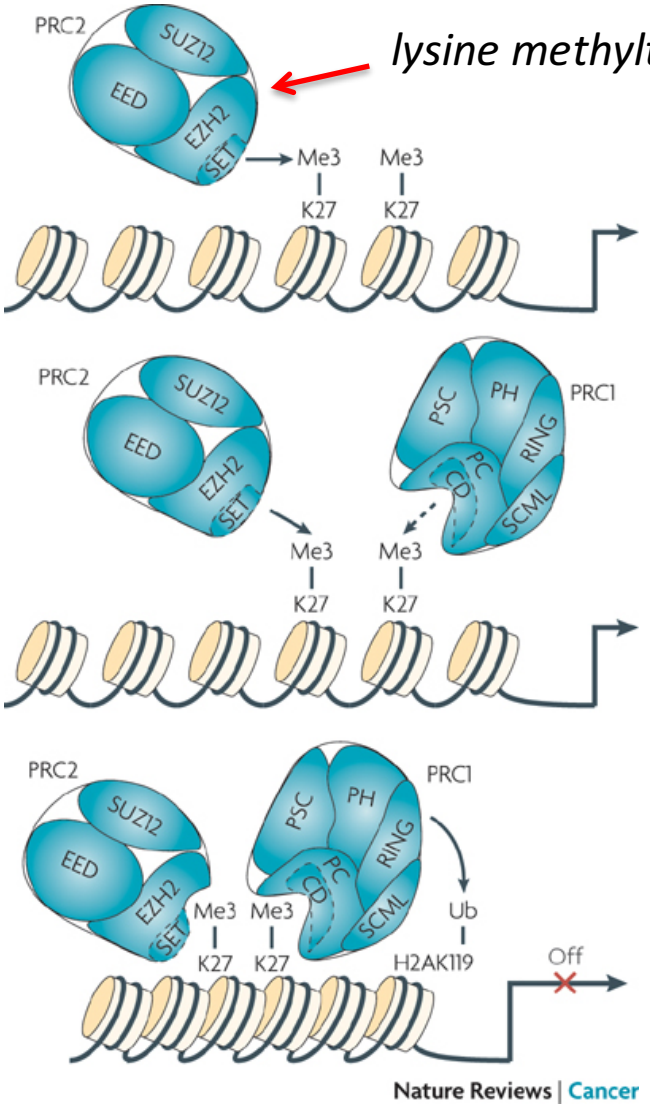
recognise their target genes in the various cell types

Polycomb group proteins (PcG)

Polycomb repressive complexes (PRCs), repress transcription by a mechanism that involves chromatin *modification*. *Two* major Polycomb repressive complexes (PRCs) have been described:

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

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



Maintenance of HOX gene expression patterns is under epigenetic regulation

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

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

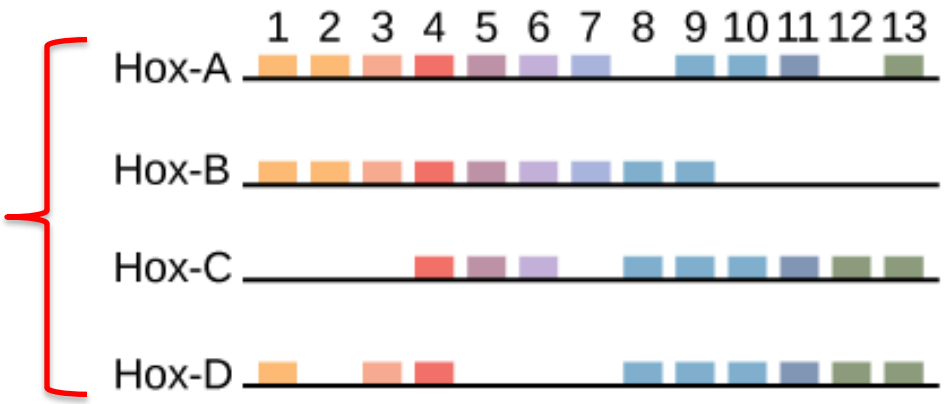
Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.

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

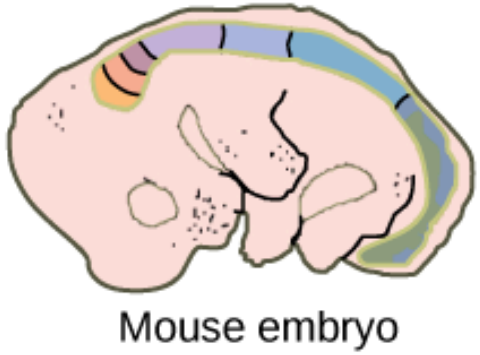
Noncoding RNAs (ncRNA) participate in epigenetic regulation but are poorly understood. Here we characterize the transcriptional landscape of the four human HOX loci at five base pair resolution in 11 anatomic sites and identify 231 HOX ncRNAs that extend known transcribed regions by more than 30 kilobases. 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.

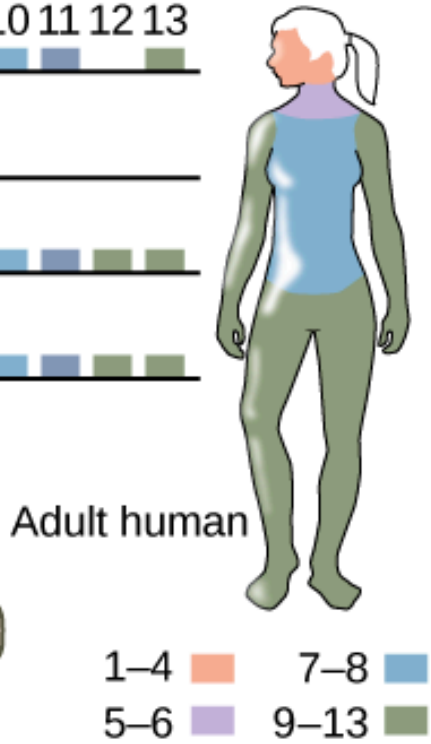
Demarcation of *active* and *silent* chromatin domains in HOX loci by ncRNAs

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).





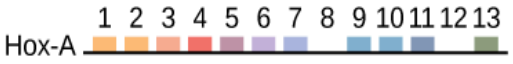
gene position= ***spatial*** position
along the anterior-posterior *anatomic axis*

The human HOX transcriptome

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

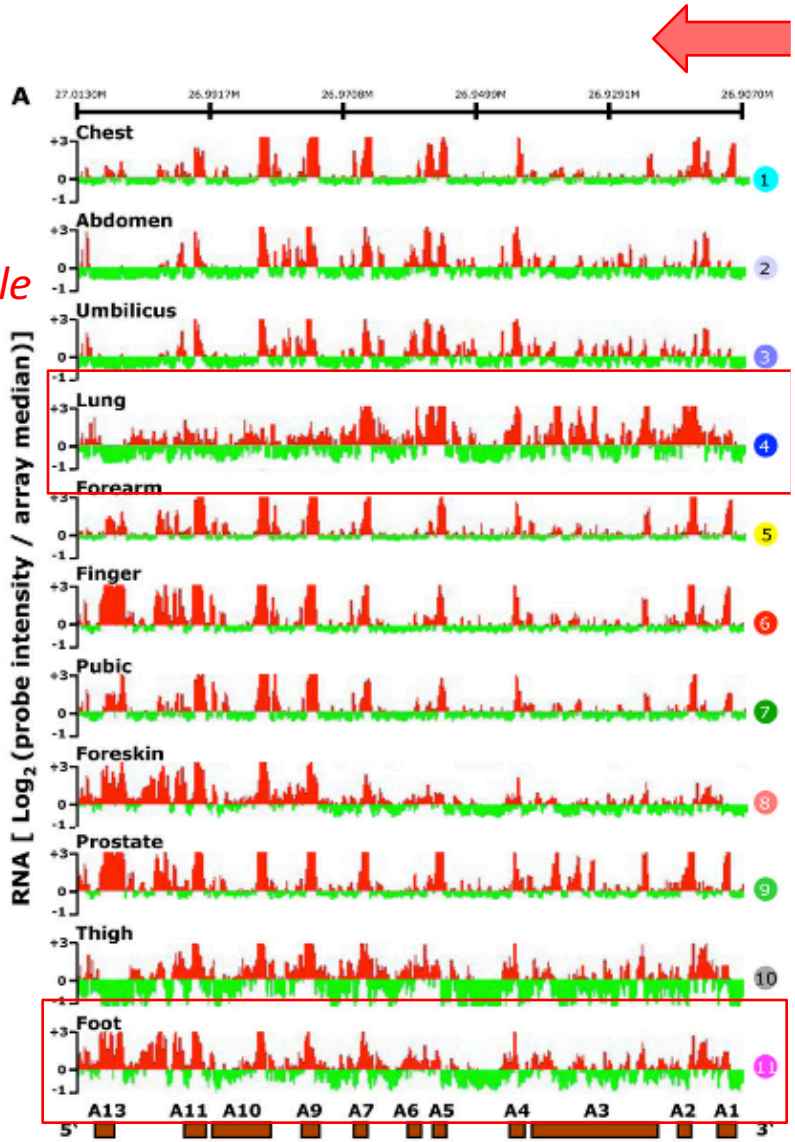
Prossimale

Site-specific transcription of the HOXA locus !



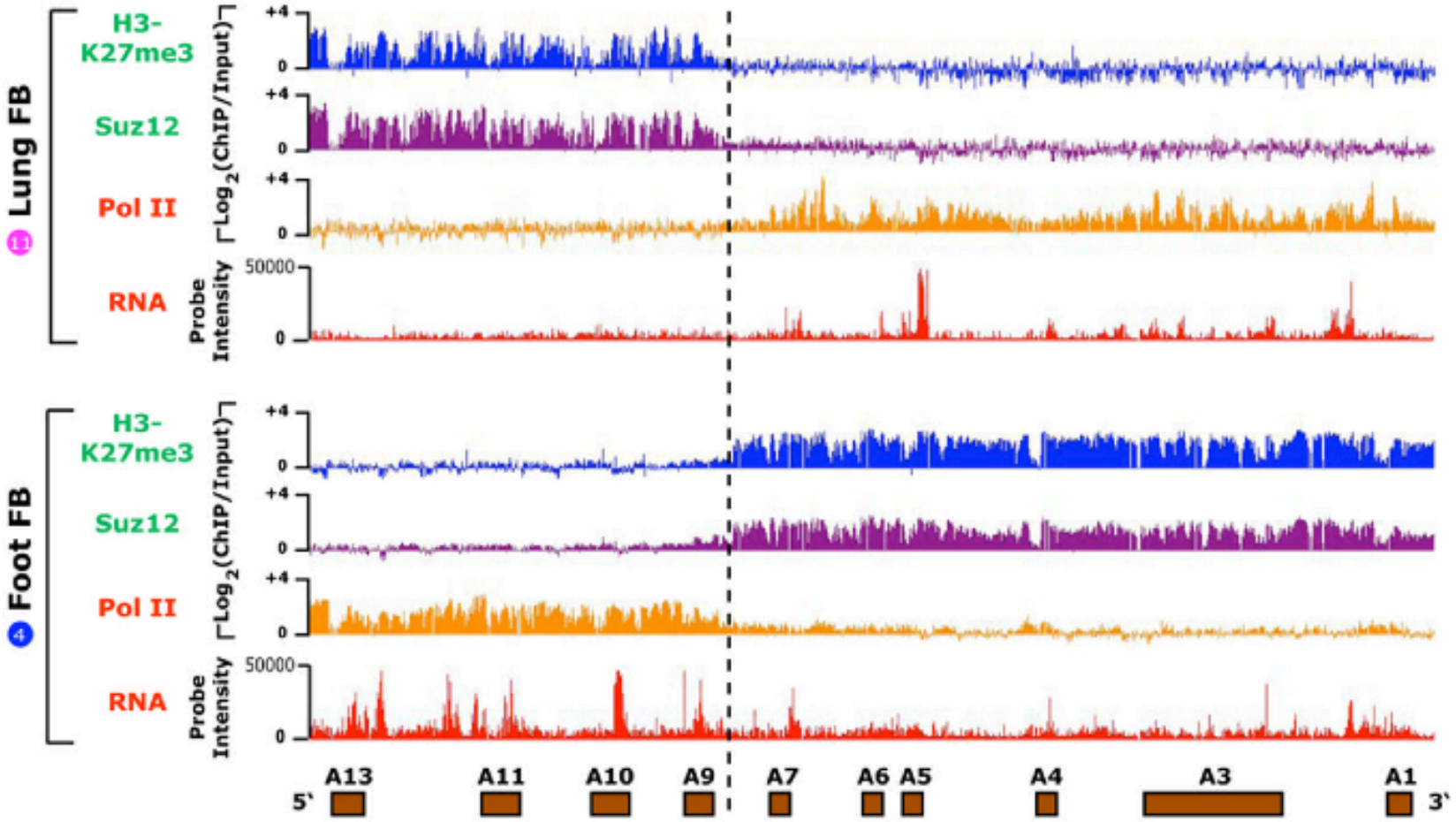
Distale

1–4 7–8
5–6 9–13



11 samples

Diametrically opposed chromatin modifications and transcriptional accessibility in the HOXA locus

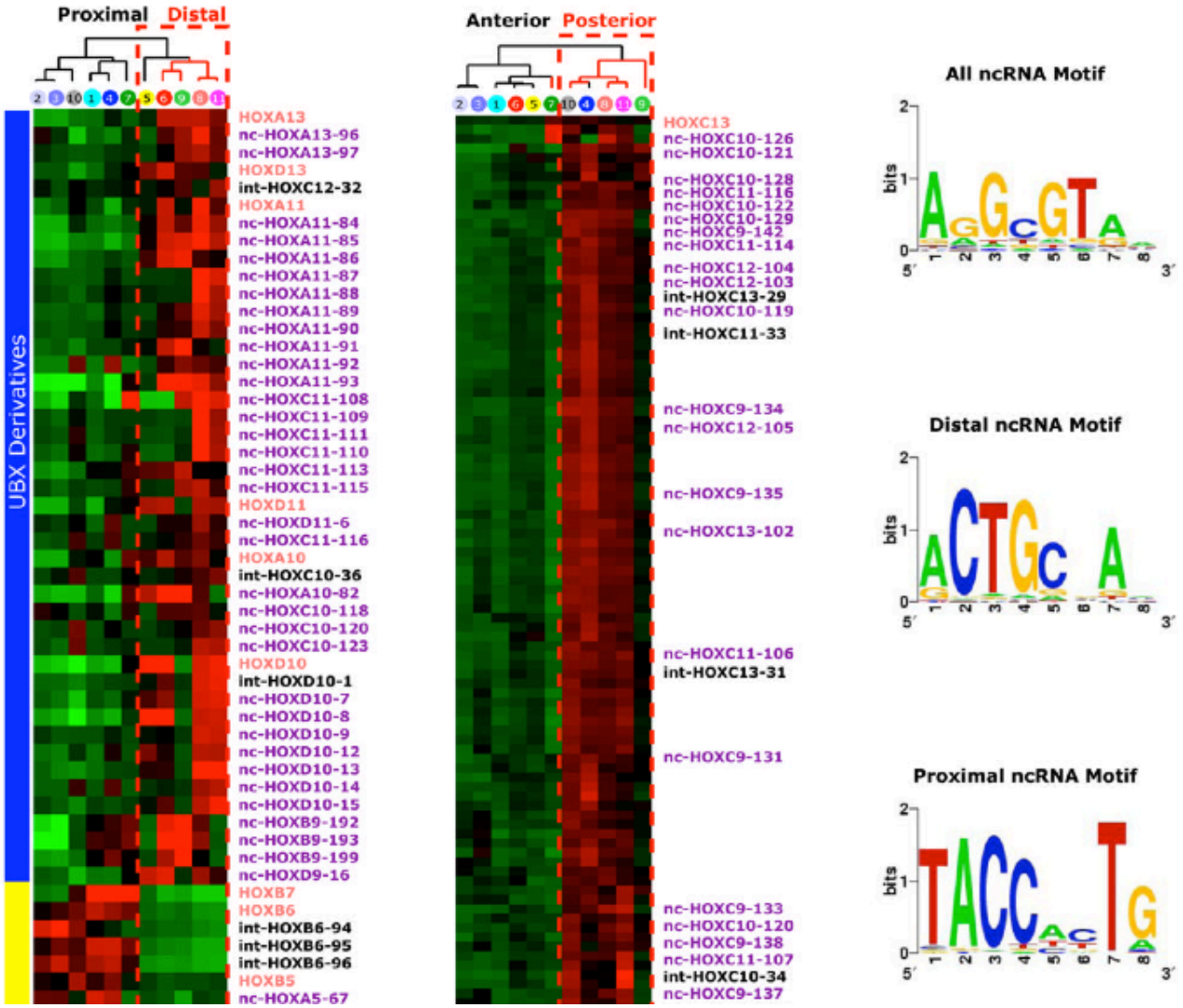


ChIP data

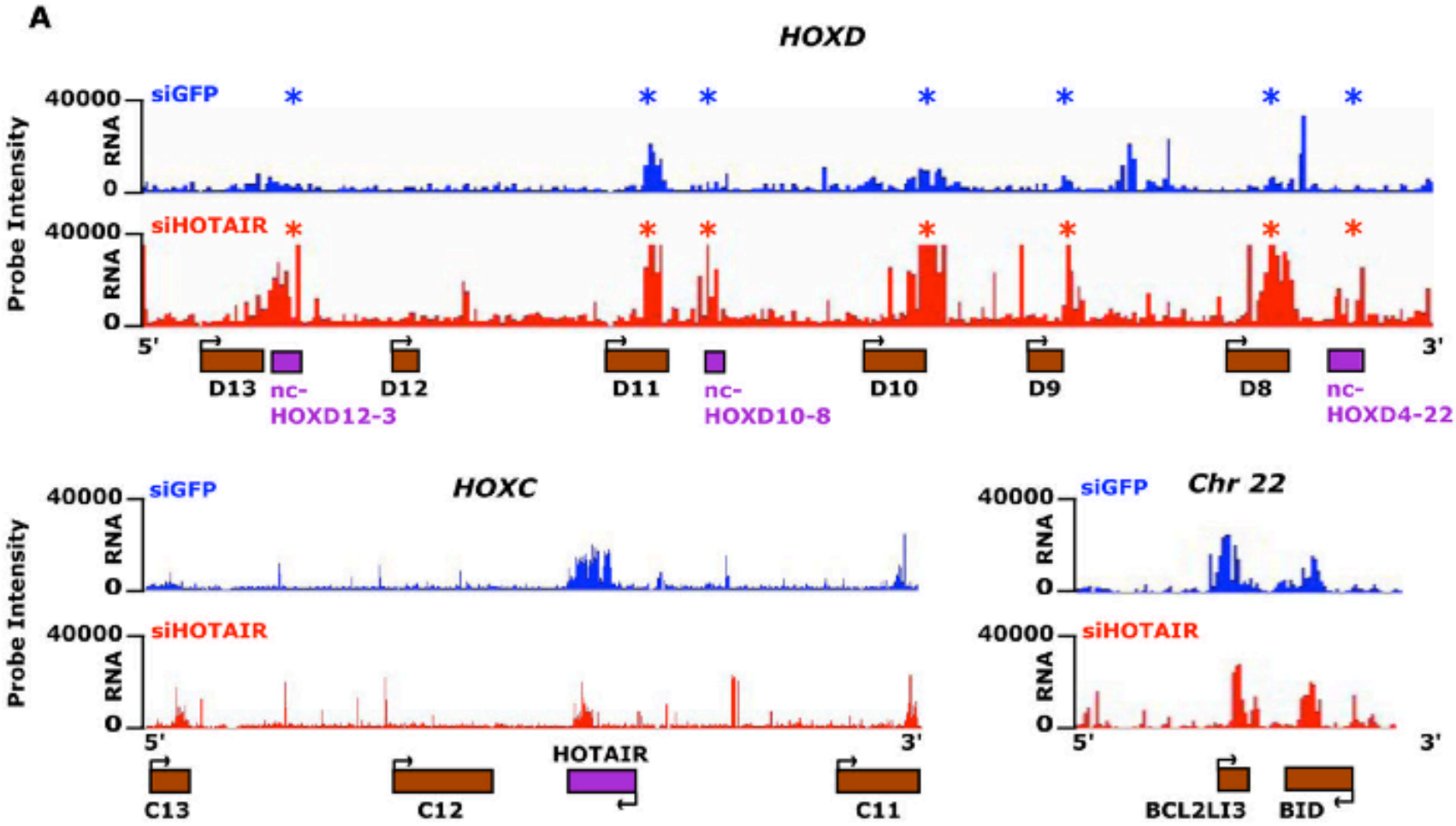
- *Numerous* long noncoding RNAs were found to be transcribed within the human HOX clusters (Rinn et al., 2007)

- The lncRNAs were found to be **also** *colinear* 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.



Loss of HOTAIR results in transcriptional induction of HOXD locus

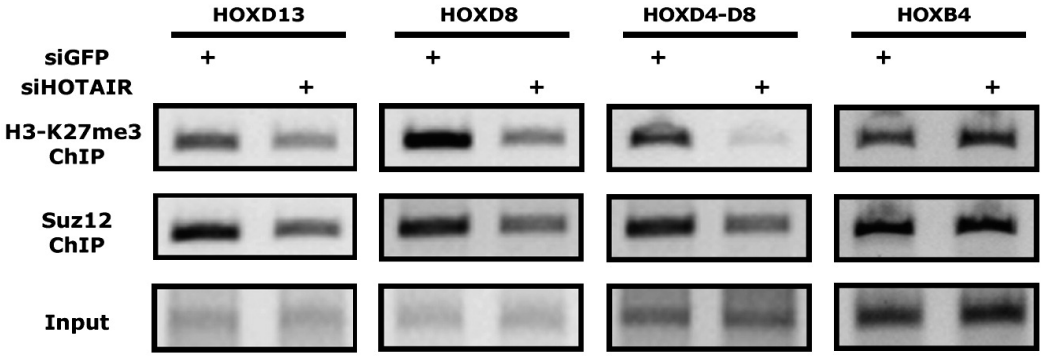


RNAi against HOTAIR in primary fibroblast led to *dramatic* transcriptional activation of the HOXD locus on chromosome 2 spanning over 40 kb, including HOXD8, HOXD9, HOXD10, HOXD11 and multiple ncRNAs

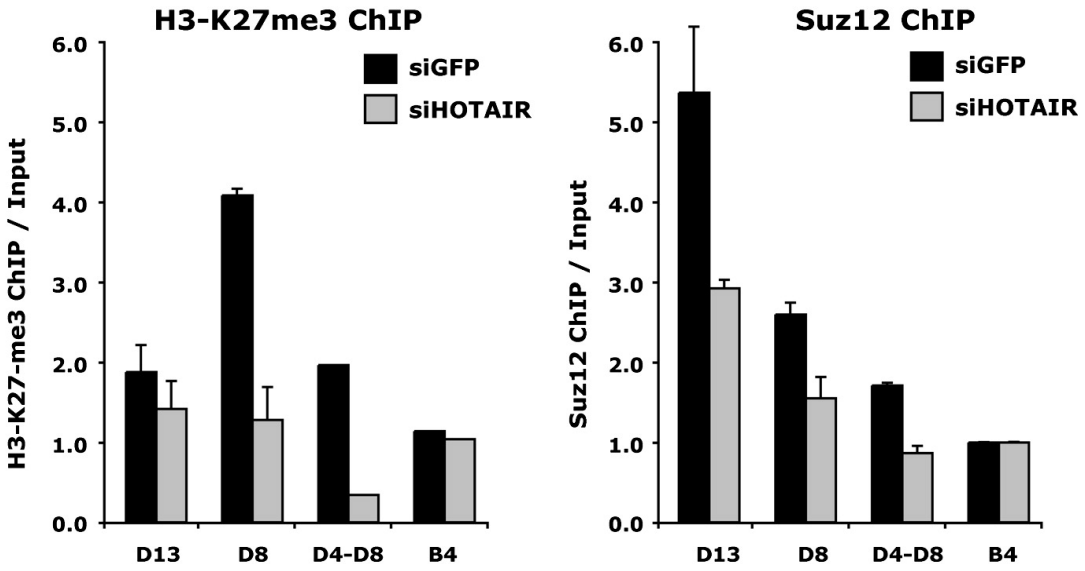
HOTAIR Enhances PRC2 Activity at the HOXD Locus

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).

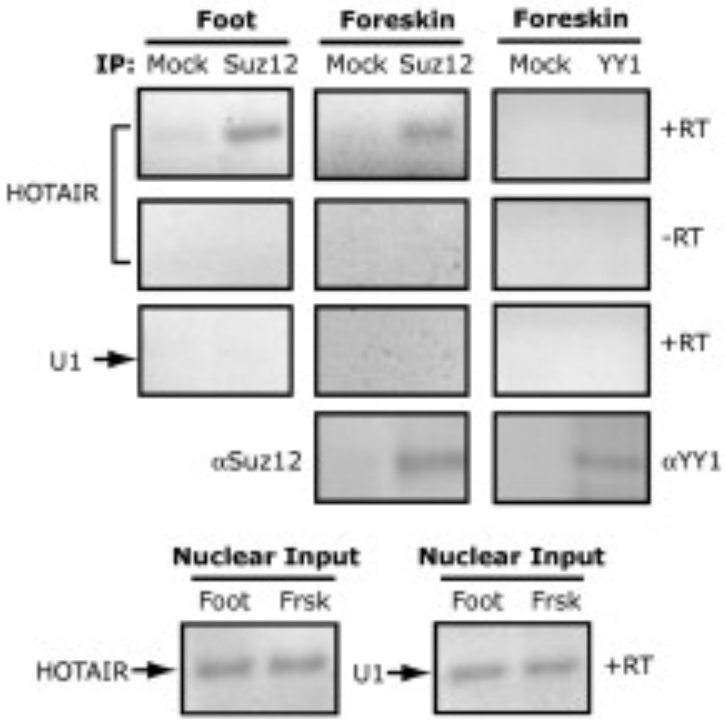


Cell. 2007 June 29; 129(7): 1311–1323.

HOTAIR ncRNA Binds PRC2

Immunoprecipitation

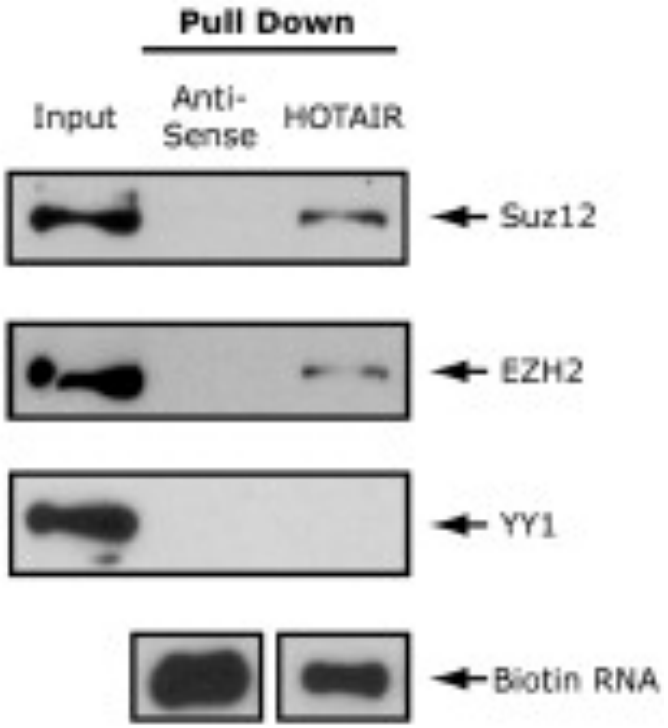
IP of Suz12 retrieves endogenous HOTAIR



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.

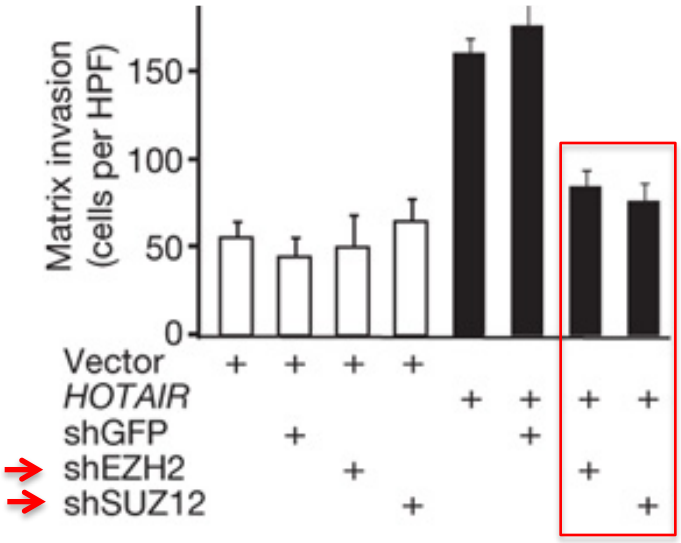
RNA Pull Down

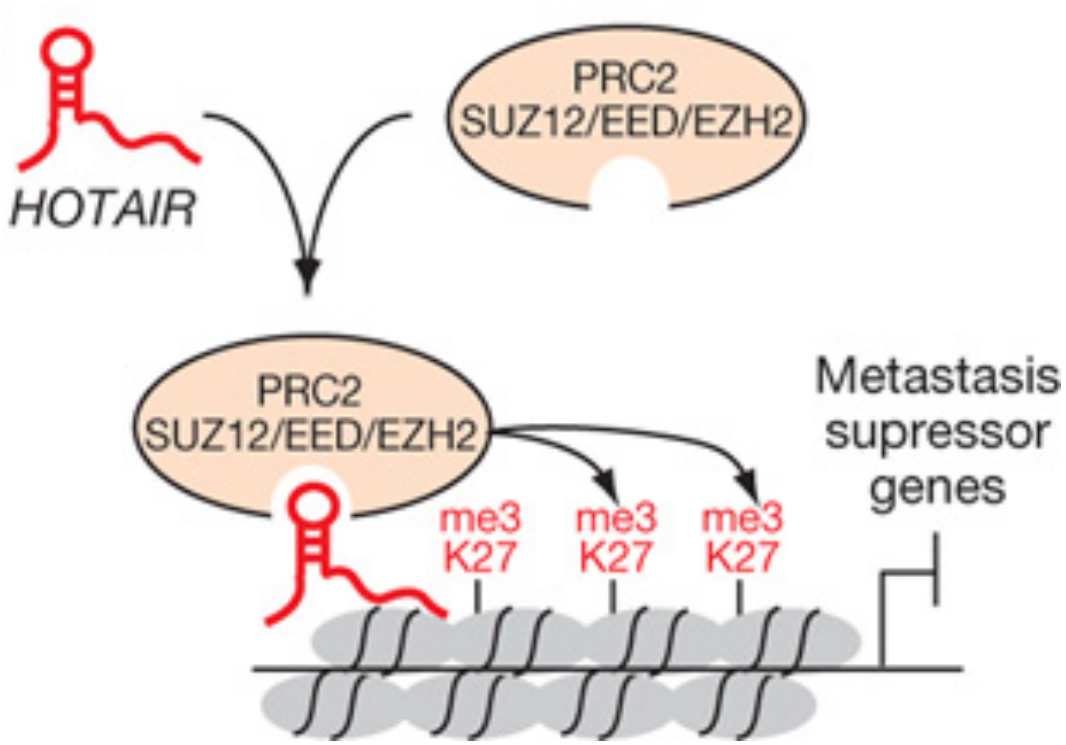
In vitro-transcribed HOTAIR retrieves PRC2 subunits



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

HOTAIR requires PRC2 for function





Promuove le metastasi tramite PRC2

Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes

Miao-Chih Tsai,[1] Ohad Manor,[2] Yue Wan,[1] Nima Mosammaparast,[3] Jordon K. Wang,[1] Fei Lan,[3,4] Yang Shi,[3] Eran Segal,[2] and Howard Y. Chang[1,*]

Author information ► Copyright and License information ►

The publisher's final edited version of this article is available at Science

See other articles in PMC that cite the published article.

Abstract

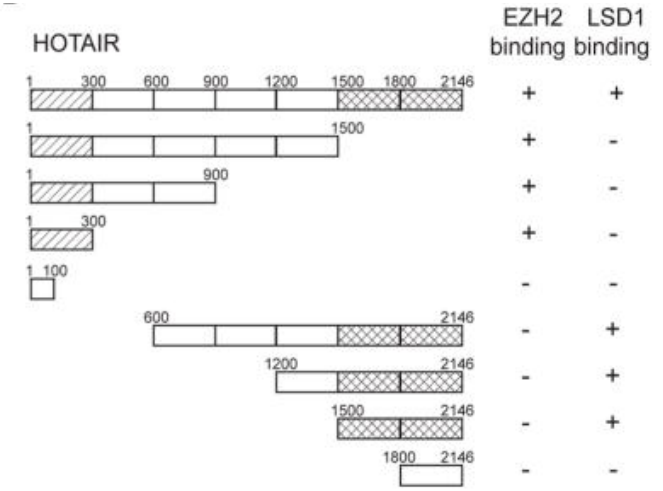
Go to:

Long intergenic noncoding RNAs (lincRNAs) regulate chromatin states and epigenetic inheritance. Here we show that the lincRNA HOTAIR serves as a scaffold for at least two distinct histone modification complexes. A 5′ domain of HOTAIR binds Polycomb Repressive Complex 2 (PRC2) while a 3′ domain of HOTAIR binds the LSD1/CoREST/REST complex. The ability to tether two distinct complexes enables RNA-mediated assembly of PRC2 and LSD1, and coordinates targeting of PRC2 and LSD1 to chromatin for coupled histone H3 lysine 27 methylation and lysine 4 demethylation. Our results suggest that lincRNAs may serve as scaffolds by providing binding surfaces to assemble select histone modification enzymes, and thereby specify the pattern of histone modifications on target genes.

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.

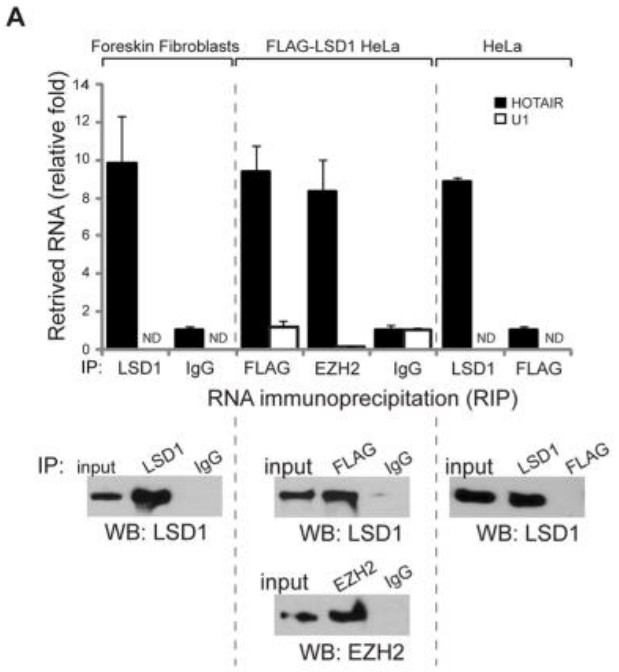


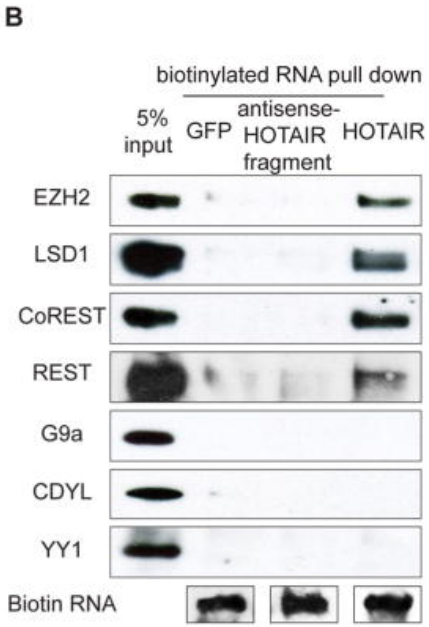
Science. Aug 6, 2010; 329(5992): 689–693.

HOTAIR can link a histone *methylase* and a *demethylase*

by acting as a modular scaffold

RIP

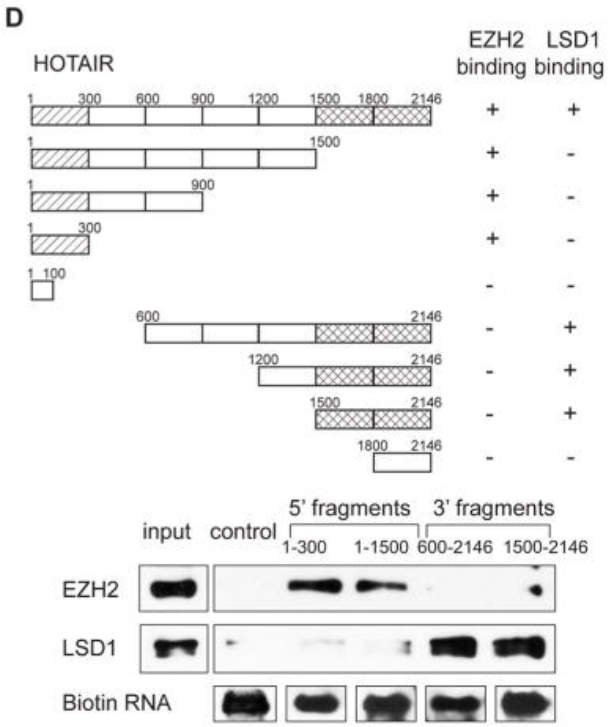




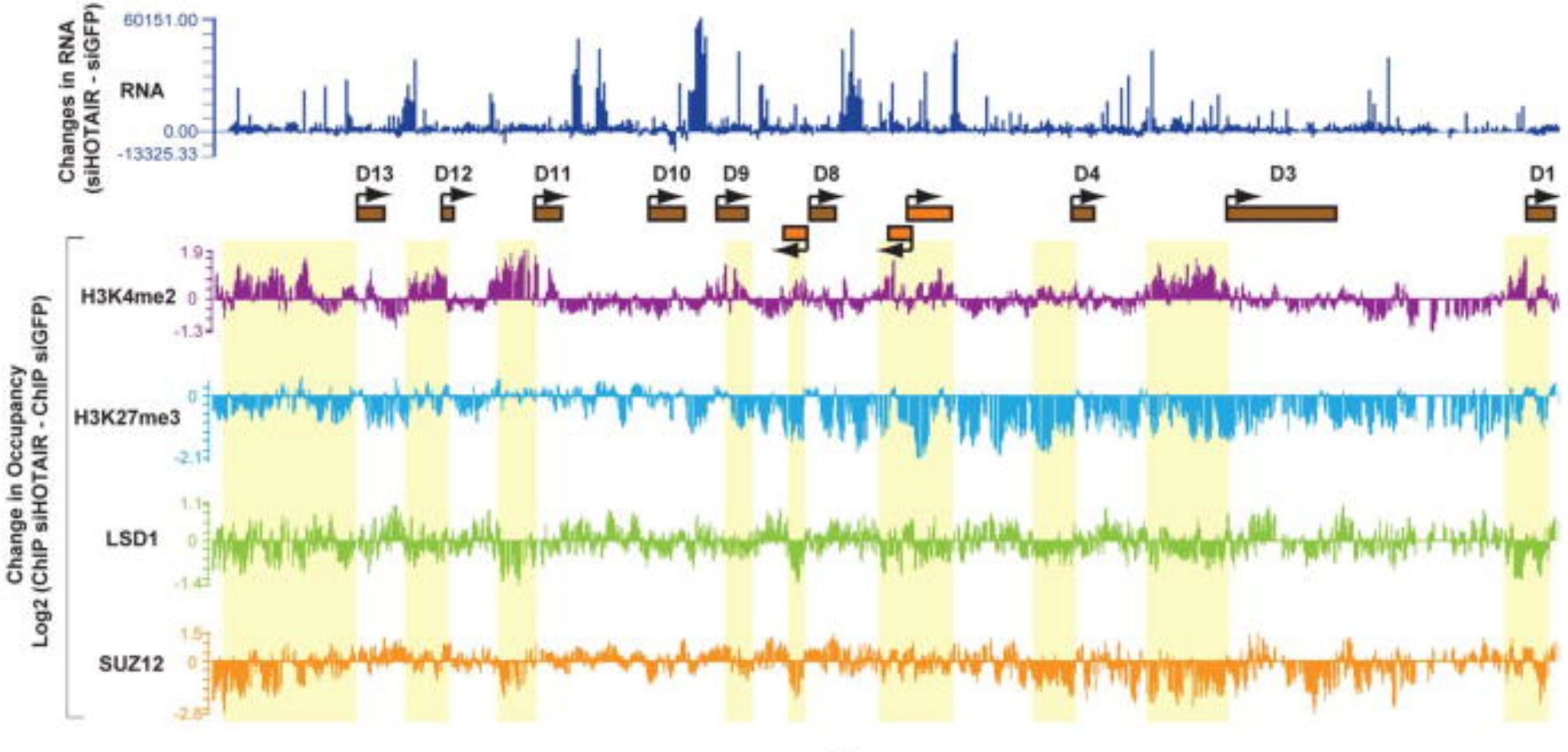
(**A**) LSD1 IP specifically retrieves HOTAIR RNA.

RNA PULL-DOWN

The presence of *independent* binding sites for PCR2 and LSD1 on HOTAIR suggests that HOTAIR may bridge PRC2 and LSD1 complexes.



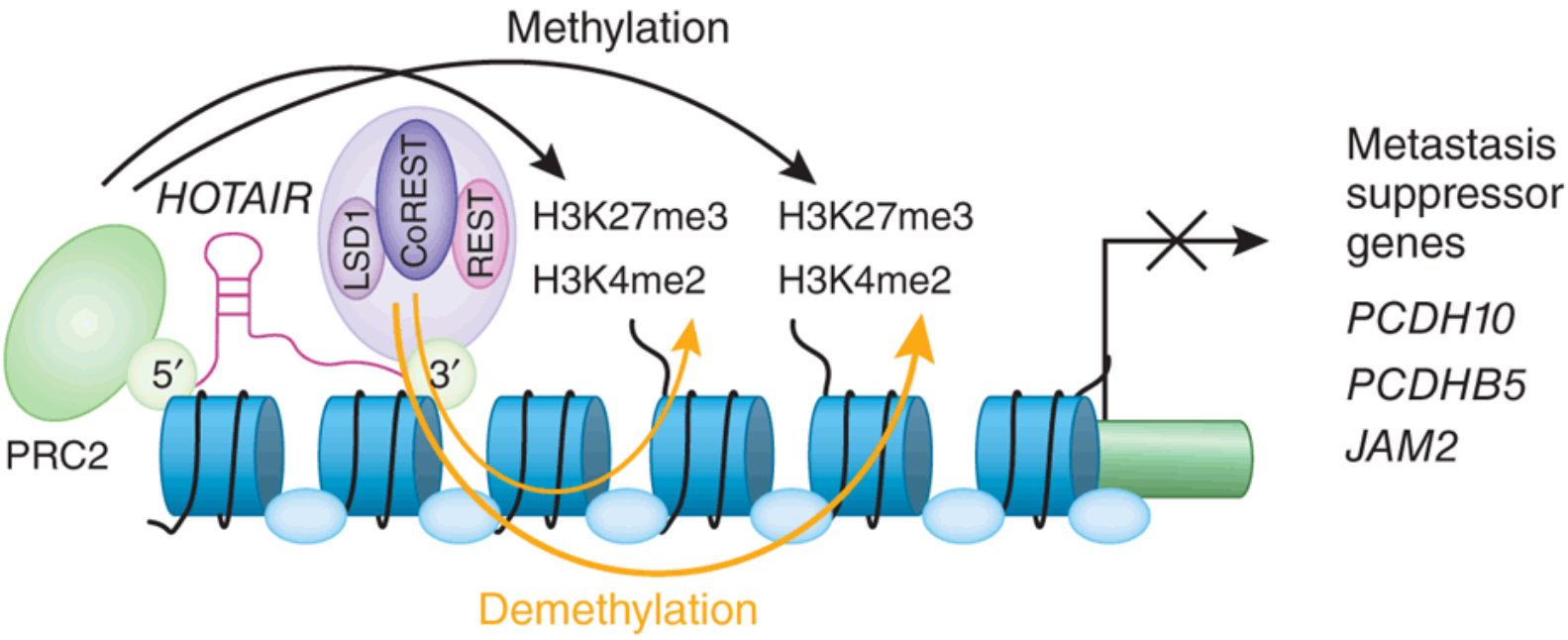
HOTAIR-mediated bridging of PRC2 and LSD1 complexes enables their coordinate binding to target genes on chromatin.



- Changes in mRNA and occupancy of H3K4me2, H3K27me3, LSD1, and SUZ12 across *HOXD* locus after RNAi of HOTAIR in foreskin fibroblasts.
- Coordinate loss of SUZ12 and LSD1 occupancy caused by HOTAIR knockdown were concentrated in proximal promoters of *HOXD* genes

HOTAIR can link a histone methylase and a demethylase by acting as a modular scaffold

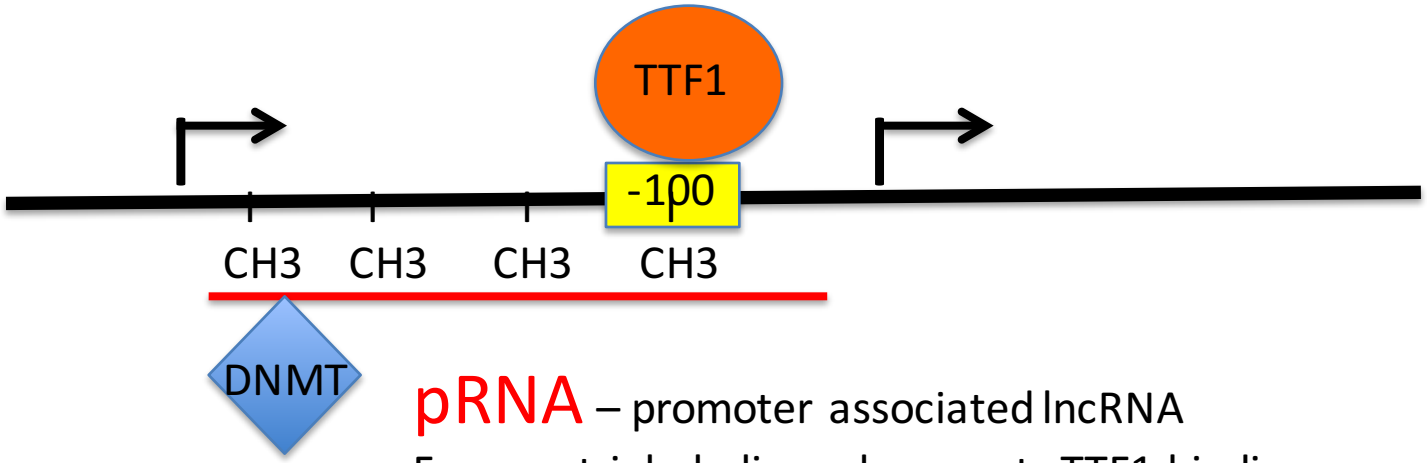
The resulting molecular complex is bound to the promoter of genes encoding metastasis suppressors (such as *PCDH10*, *PCDHB5* and *JAM2*) to coordinately regulate the histone modifications H3K27me3 trimethylation and H3K4me2 demethylation (that removes an active chromatin mark), which in turn, silence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



Ingrid Grummt – Heidelberg

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

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



pRNA – promoter associated lncRNA

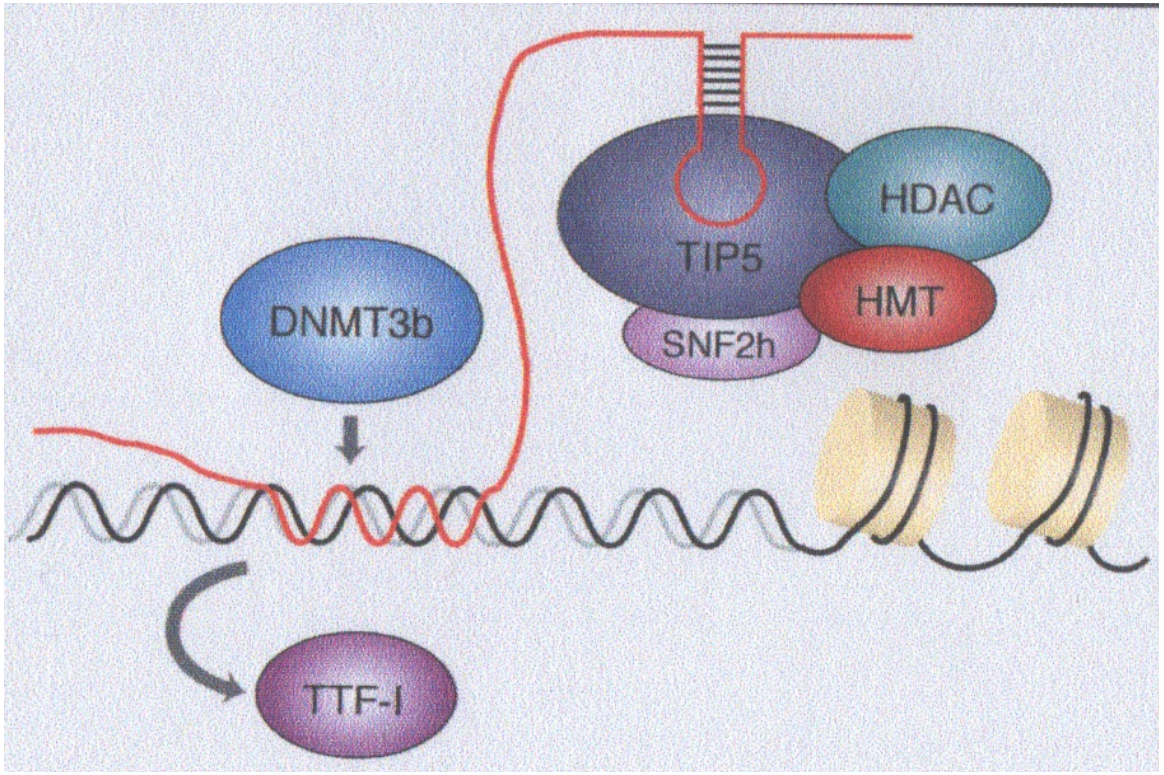
Forms a triple helix and prevents TTF1 binding

While recruiting DNMT3b

Ingrid Grummt – Heidelberg

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

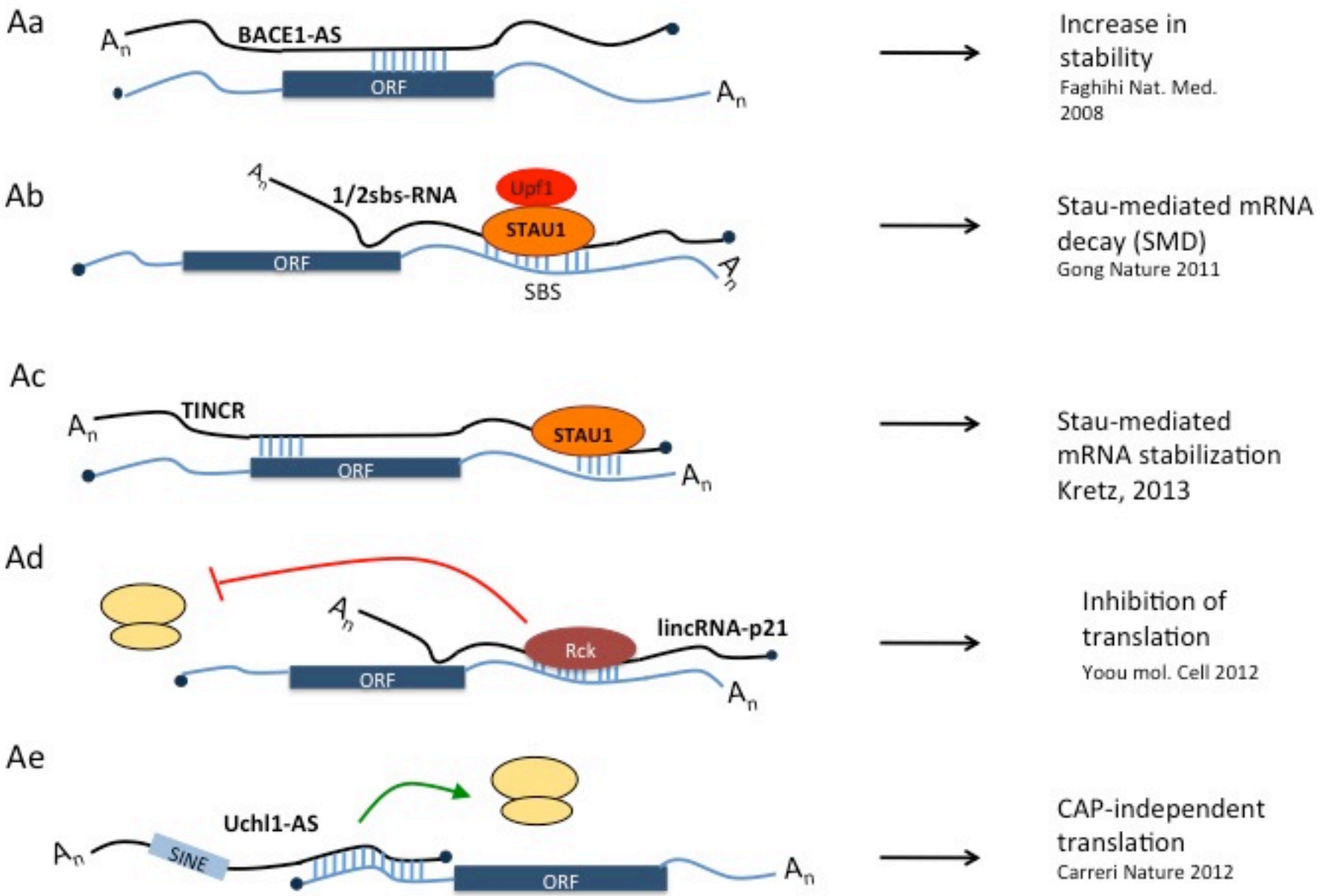
Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.



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

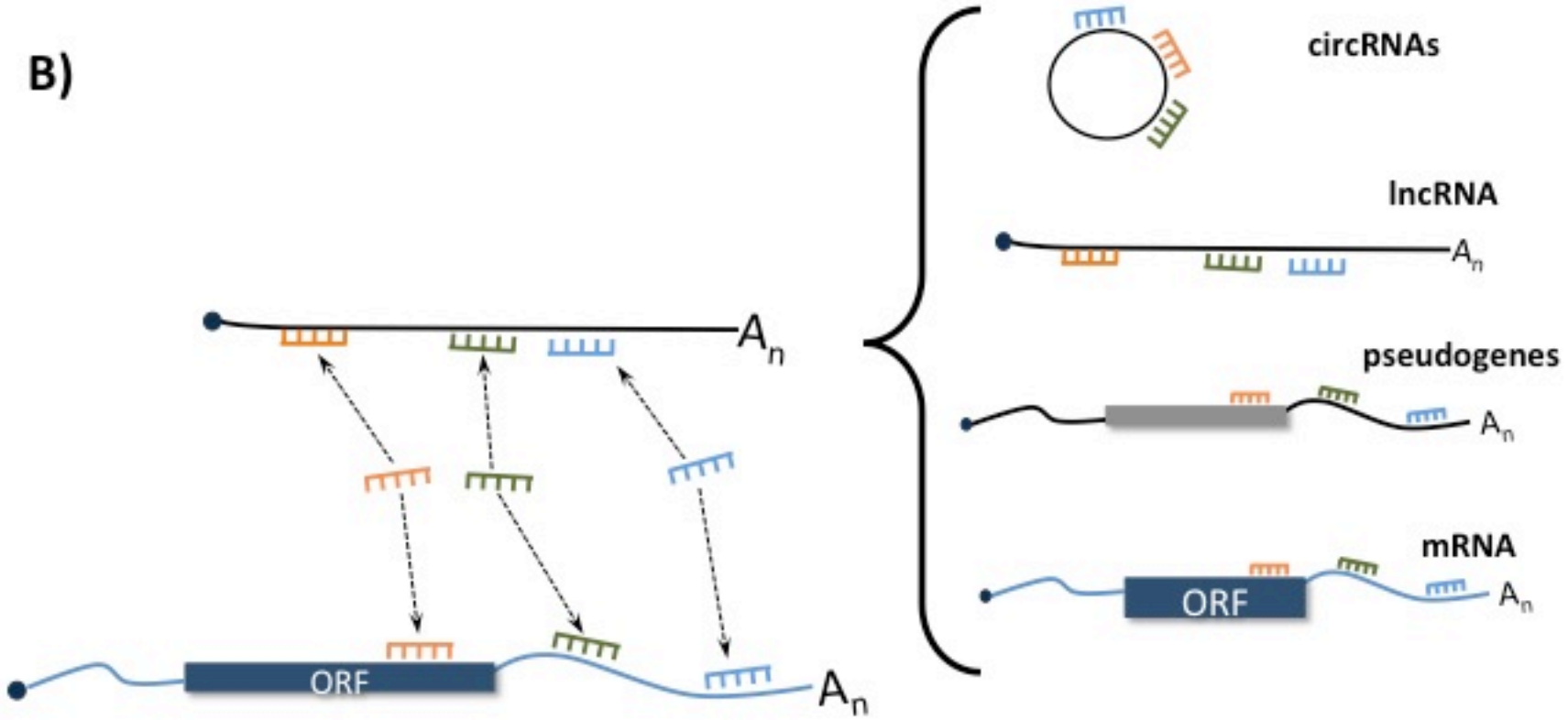
In addition, pRNA directly interacts with DNA, forming a DNA:DNA:RNA triple helix with the bind ing site of the transcription factor TTF-i, leading to displacement of TTF-i. The triple helical structure is recognized by the DNA methyltransferase DNMT3b, which methylates the rDNA promoter, leading to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

Cytoplasmic lncRNA



Fatica & Bozzoni, Figure 2A

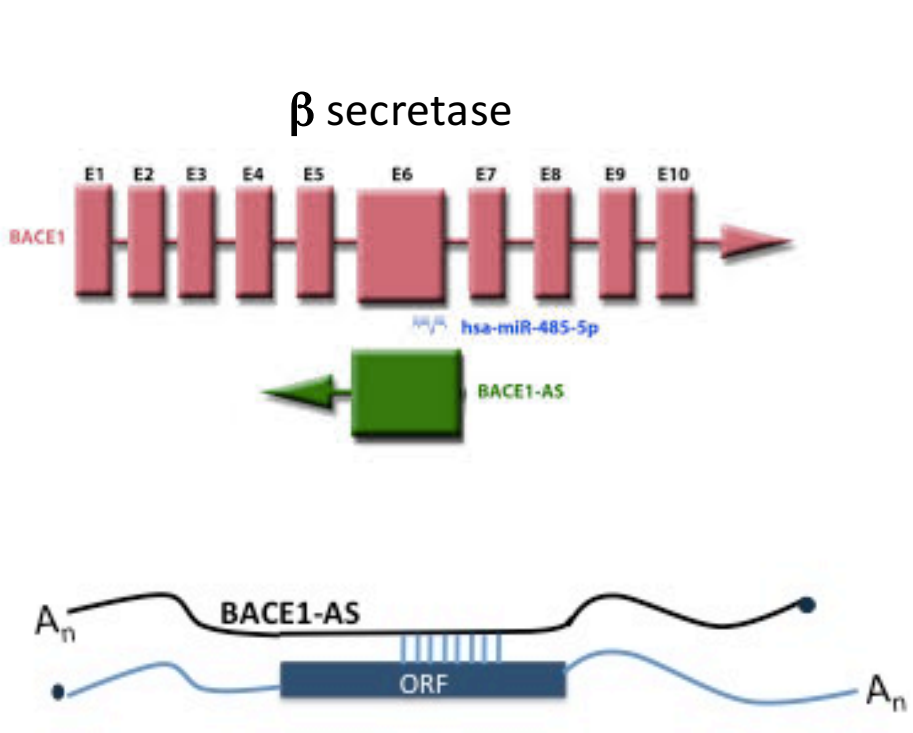
Competing endogenous RNAs for miRNA binding

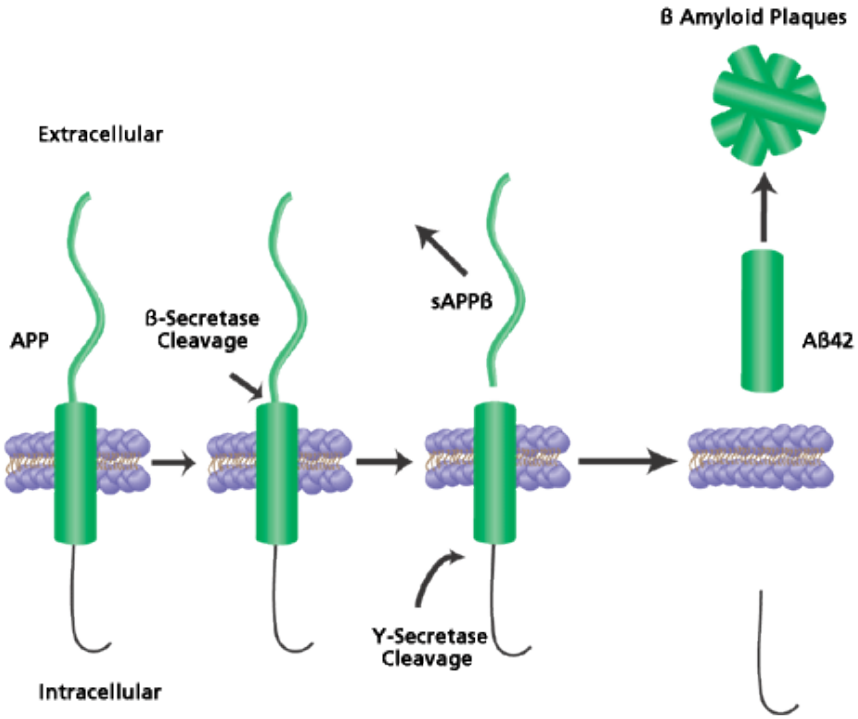


lncRNAs in brain disorders

Expression of the non coding **BACE1-AS is elevated in Alzheimer's disease** and drives regulation of beta-secretase.

The BACE1-antisense transcript (BACE1-AS) regulates BACE1 mRNA and subsequently BACE1 protein expression in vitro and in vivo**. Upon exposure to various cell stressors, expression of BACE1-AS becomes elevated, increasing BACE1 mRNA stability and generating additional Abeta** 1-42 through a post-transcriptional feed-forward mechanism. **concentrations were elevated in subjects with Alzheimer's disease and in amyloid precursor protein transgenic mice.**





Nat Med. 2008, 14:723-30

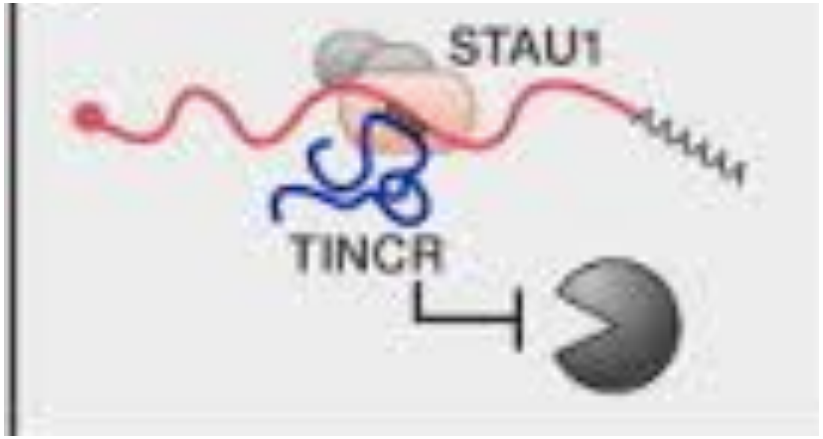
Nature. 2013 Jan 10;493(7431):231-5.

Control of somatic tissue differentiation by the long non-coding RNA TINCR.

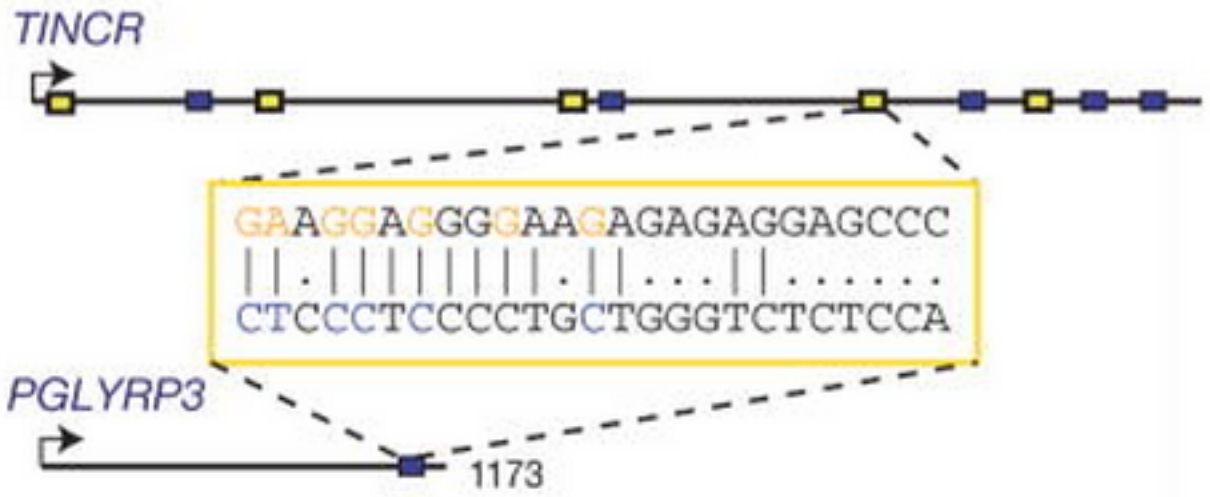
Kretz et al.

Abstract

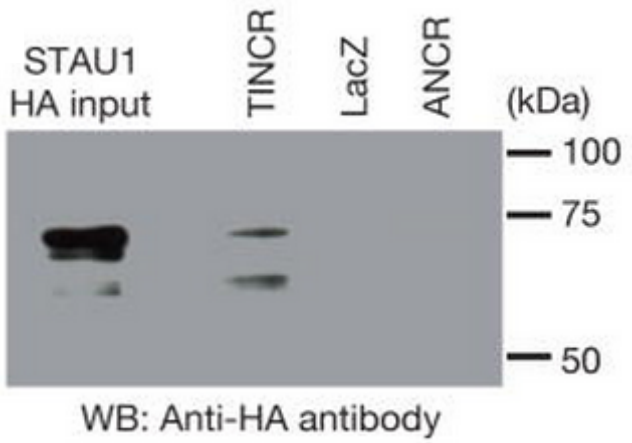
Several of the thousands of human long non-coding RNAs (lncRNAs) have been functionally characterized; however, potential roles for lncRNAs in somatic tissue differentiation remain poorly understood. Here we show that a **3.7-kilobase lncRNA**, terminal differentiation-induced ncRNA (TINCR), controls human **epidermal differentiation** by a post-transcriptional mechanism. TINCR is required for high messenger RNA abundance of key differentiation genes, many of which are mutated in human skin diseases, including FLG, LOR, ALOXE3, ALOX12B, ABCA12, CASP14 and ELOVL3. TINCR-deficient epidermis lacked terminal differentiation ultrastructure, including keratohyalin granules and intact lamellar bodies. Genome-scale RNA interactome analysis revealed that **TINCR interacts with a range of differentiation mRNAs**. TINCR-mRNA interaction occurs through a **25-nucleotide 'TINCR box'** motif that is strongly enriched in interacting mRNAs and required for TINCR binding. A high-throughput screen to analyse TINCR binding capacity to approximately 9,400 human recombinant proteins revealed direct binding of **TINCR RNA to the staufen1 (STAU1)** protein. STAU1-deficient tissue recapitulated the impaired differentiation seen with TINCR depletion. Loss of UPF1 and UPF2, both of which are required for STAU1-mediated RNA decay, however, did not have differentiation effects. Instead, the TINCR-STAU1 complex seems to mediate stabilization of differentiation mRNAs, such as KRT80. These data identify TINCR as a key lncRNA required for somatic tissue differentiation, which occurs through lncRNA binding to differentiation mRNAs to ensure their expression.



Given its cytoplasmic TINCR control of epidermal barrier genes may occur at the post-transcriptional level through direct association with target mRNAs. To test this, we developed RNA interactome analysis, followed by deep sequencing (RIA-Seq). Thirty-eight biotinylated DNA probes were designed in even- and odd-numbered pools. These two pools were used separately in a multiplex fashion **for pull-down of endogenous TINCR and associated RNAs** in differentiated keratinocytes



25-nucleotide motif that was strongly enriched in TINCR-interacting mRNAs



Cell

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

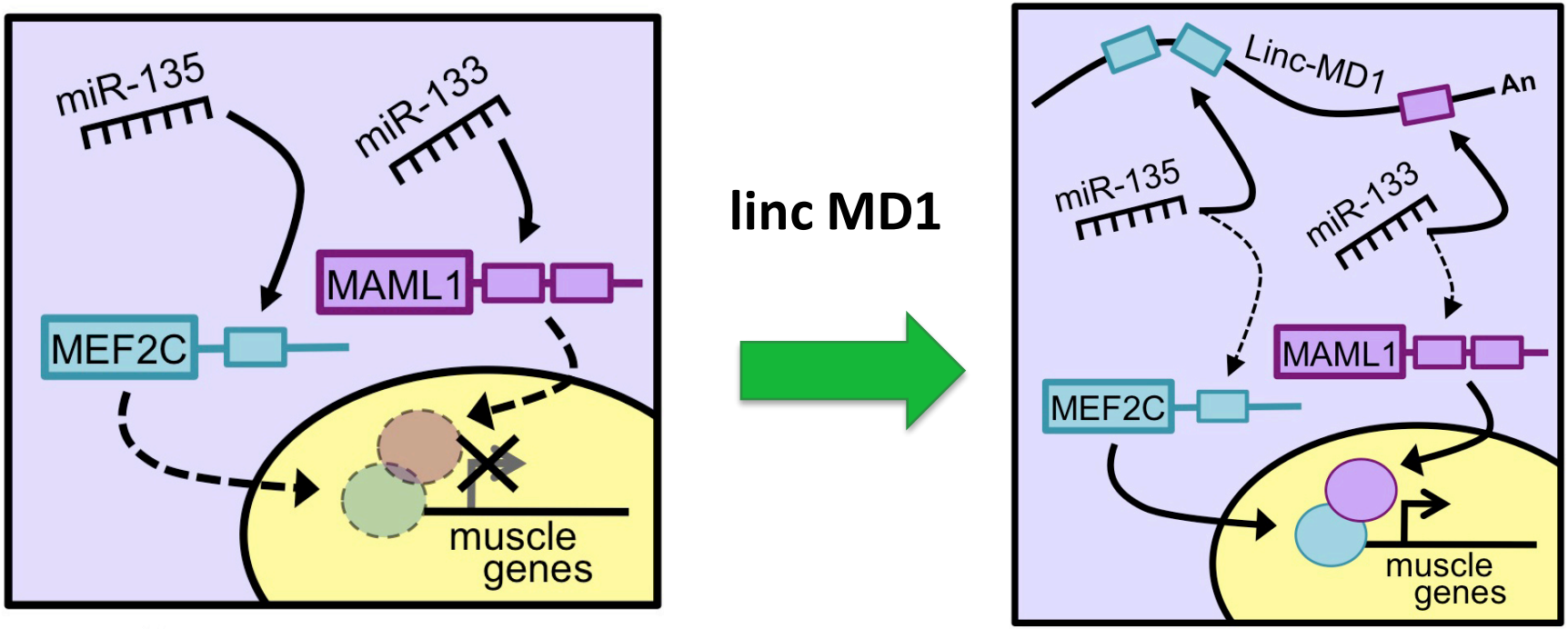
Marcella Cesana,[1,6] Davide Cacchiarelli,[1,6] Ivano Legnini,[1] Tiziana Santini,[1] Olga Sthandier,[1] Mauro Chinappi,[2] Anna Tramontano,[2,3,4] and Irene Bozzoni[1,3,4,5,*]

[1]Department of Biology and Biotechnology "Charles Darwin"

SUMMARY

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.

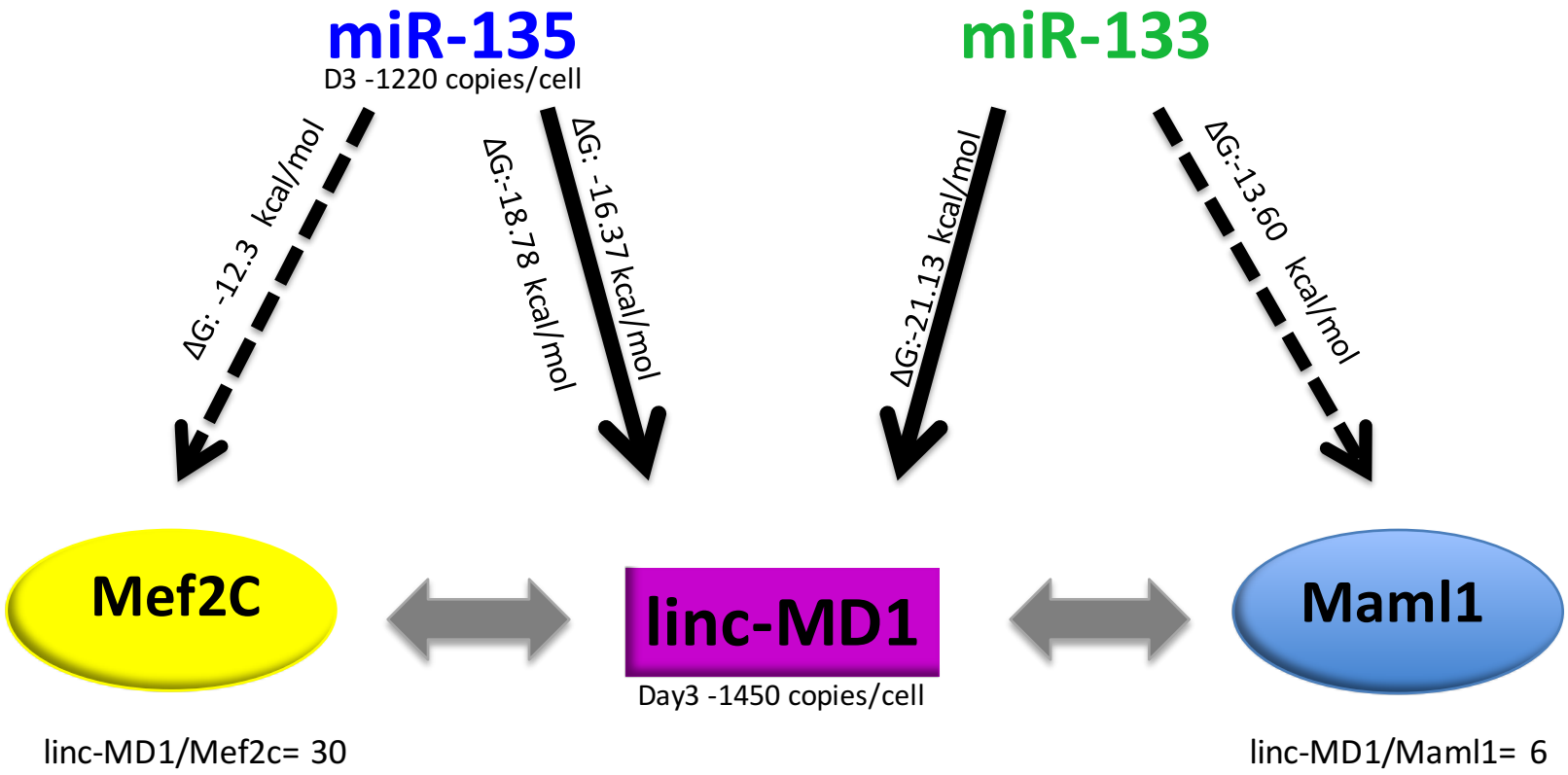
linc-MD1 acts as a sponge for specific miRNAs



differentiation

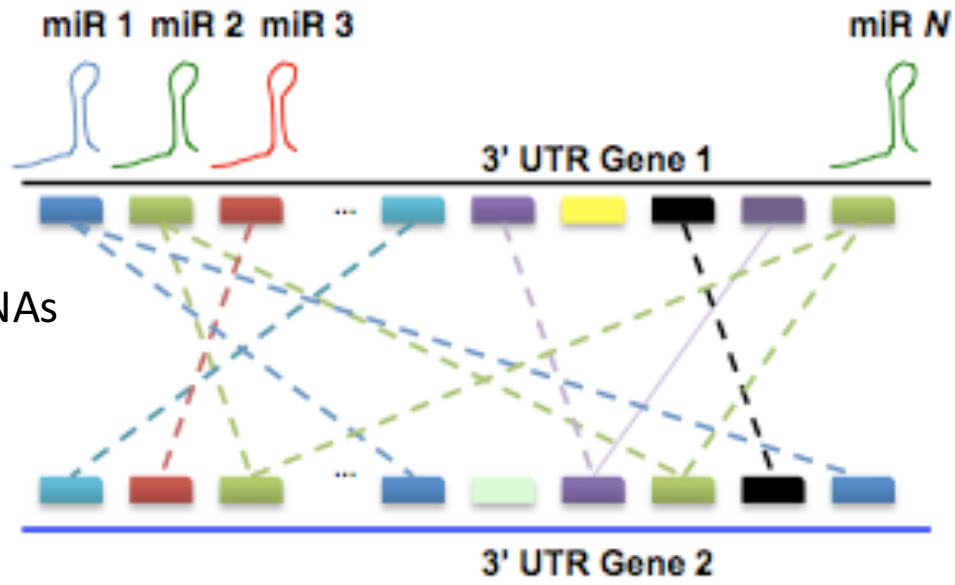
Cesana et al., *Cell* 147, *358-369*, 2011

Crosstalk between coding and non coding RNAs



ΔG values were obtained from miRanda (Enright et al., 2003)

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



competing endogenous RNAs

ceRNAs

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

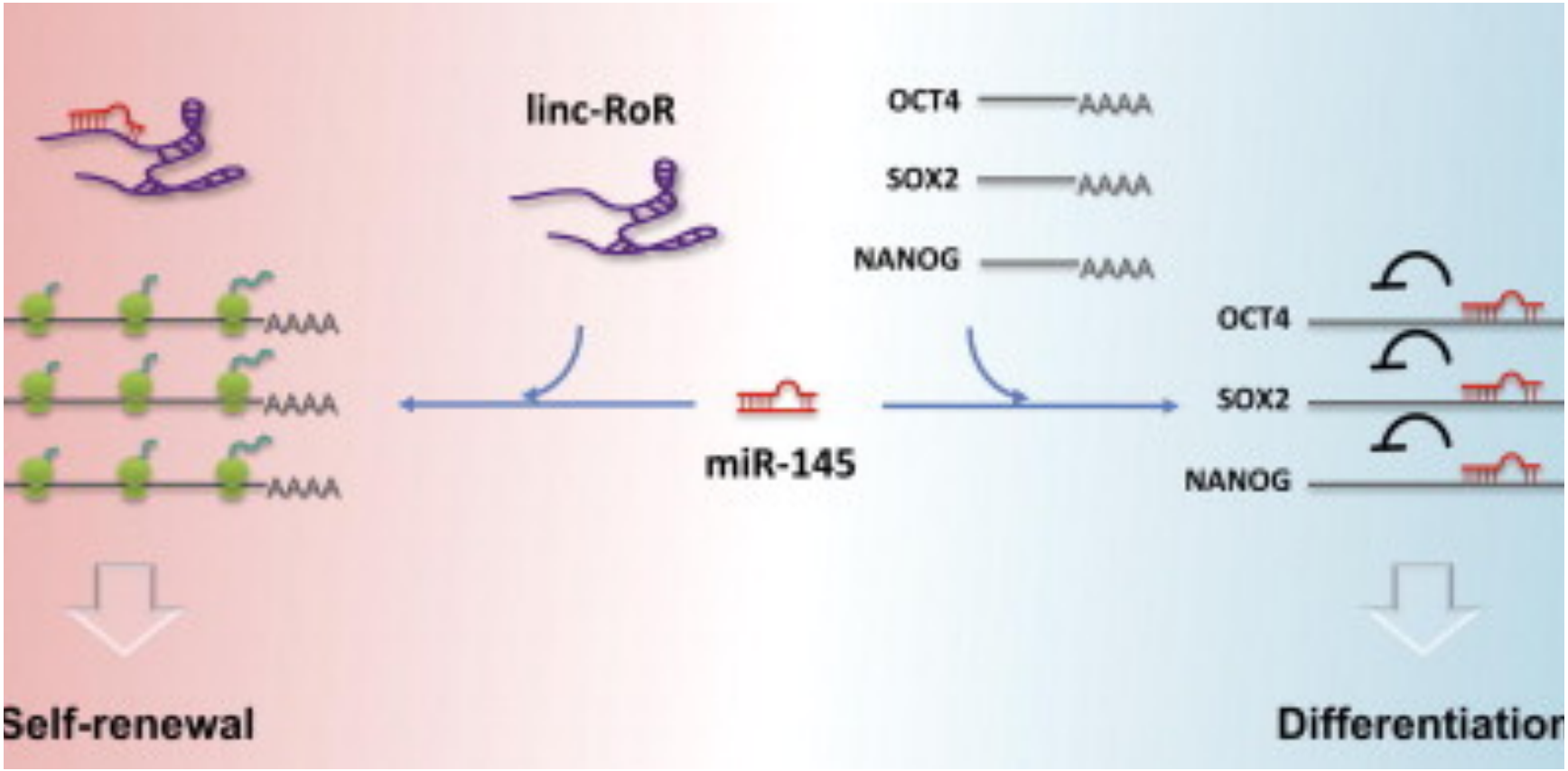
Tay et al., *Cell* 147, October 14, 2011

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

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.

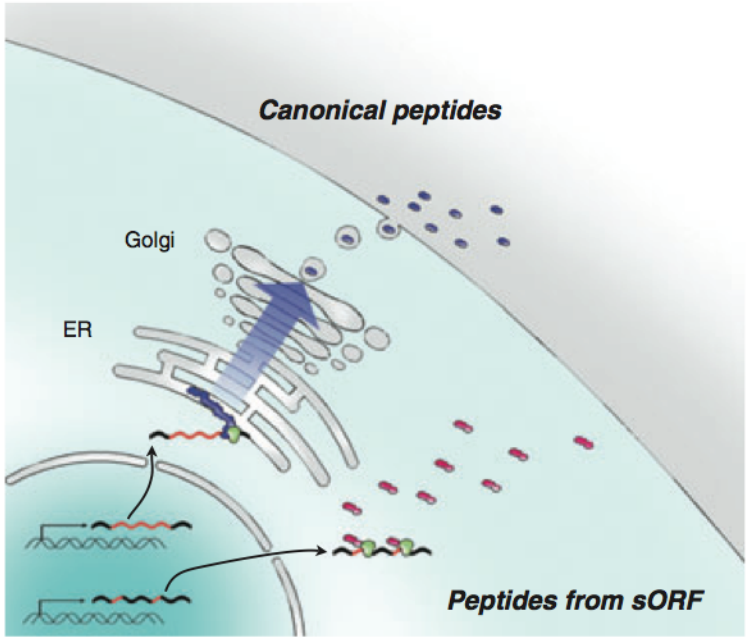


A Competition for miR-145 between linc-RoR and mRNAs Encoding the Core TFs. The presence of linc-RoR in hESCs traps miR-145, preventing it from repressing the translation of the core pluripotency factors and ensuring the stem cell fate. The disappearance of linc-RoR in differentiating hESCs releases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

Functions of lncRNAs

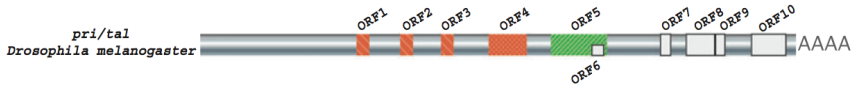
Cytoplasmic lncRNAs

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.



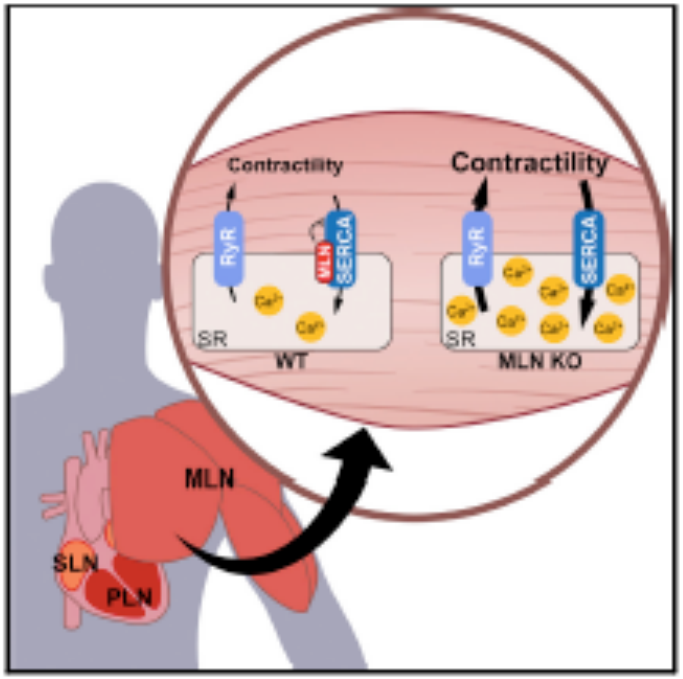
lncRNAs can encode for short peptides

Article

Cell

A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance

Graphical Abstract



Authors

Douglas M. Anderson,
Kelly M. Anderson, ...,
Rhonda Bassel-Duby, Eric N. Olson

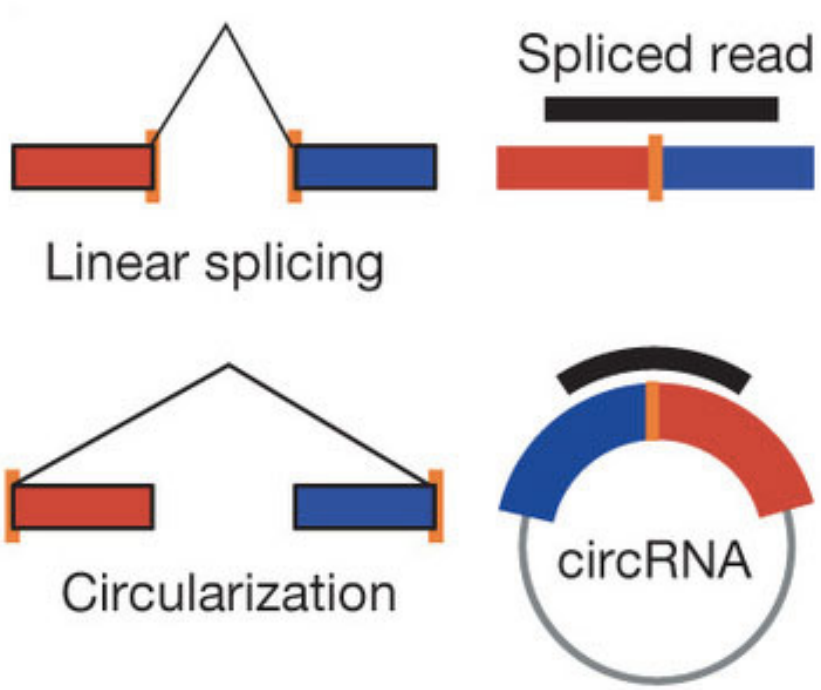
Correspondence

eric.olson@utsouthwestern.edu

In Brief

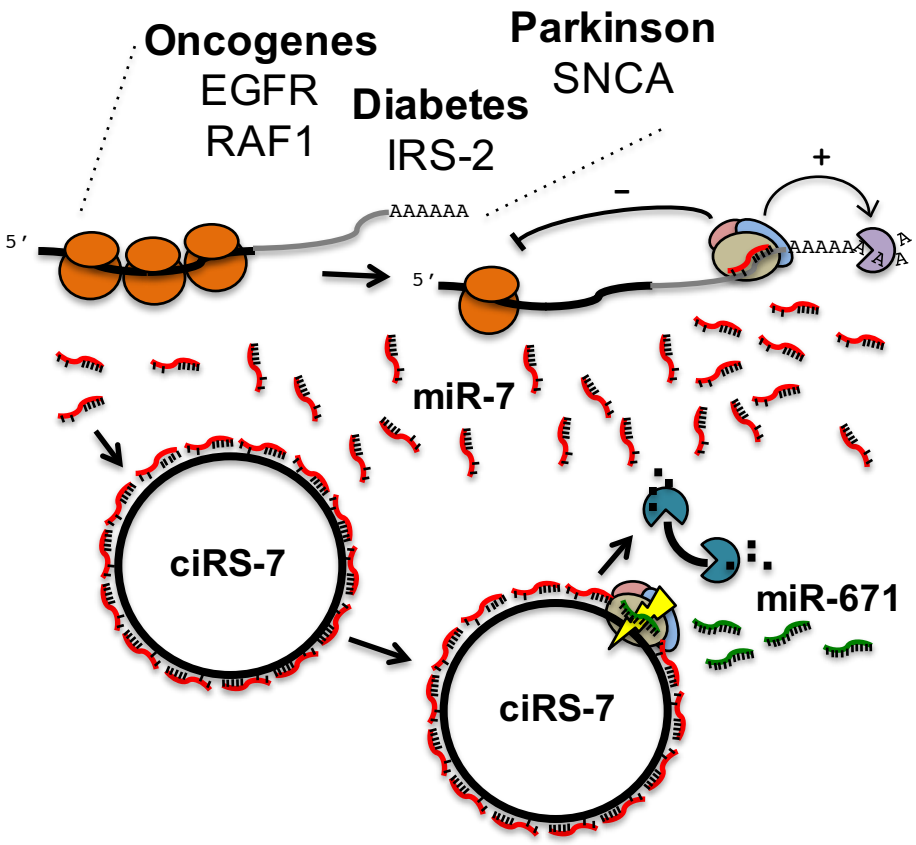
Myoregulin is a skeletal muscle-specific micropeptide that regulates muscle performance by modulating intracellular calcium handling.

…..more non coding RNAs …….**circular RNAs**



Biogenesis of circular RNAs. A gene can be transcribed and spliced into linear and circular RNAs. Note the unique 'head-to-tail' splice junctions formed by an acceptor splice site at the 5' end of an exon and a donor site at the 3' end of a downstream exon.

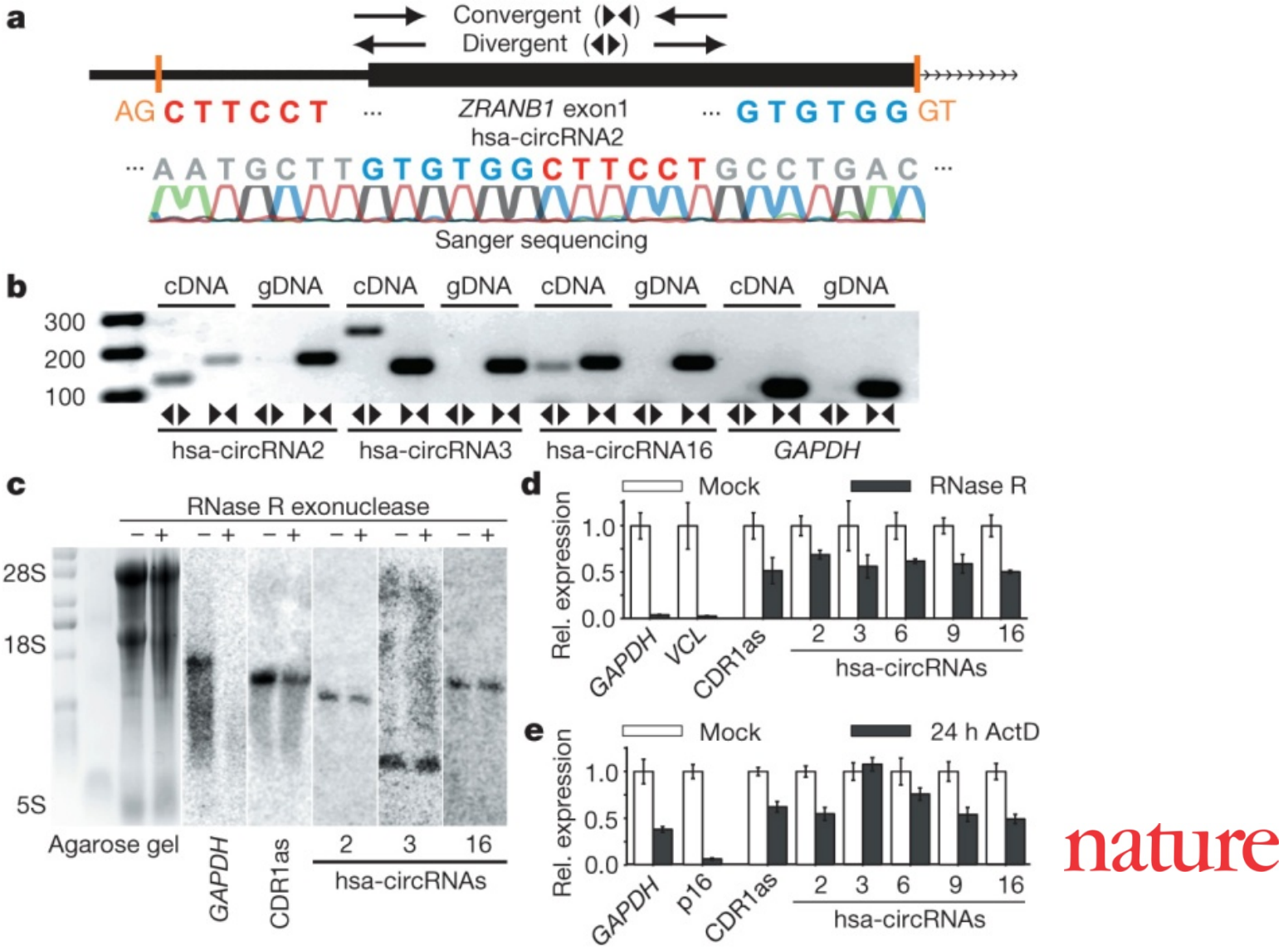
Circular RNA (circRNA) Sponge



Nature 495, 333–338 (21 March 2013)

Circular RNAs are a large class of animal RNAs with regulatory potency

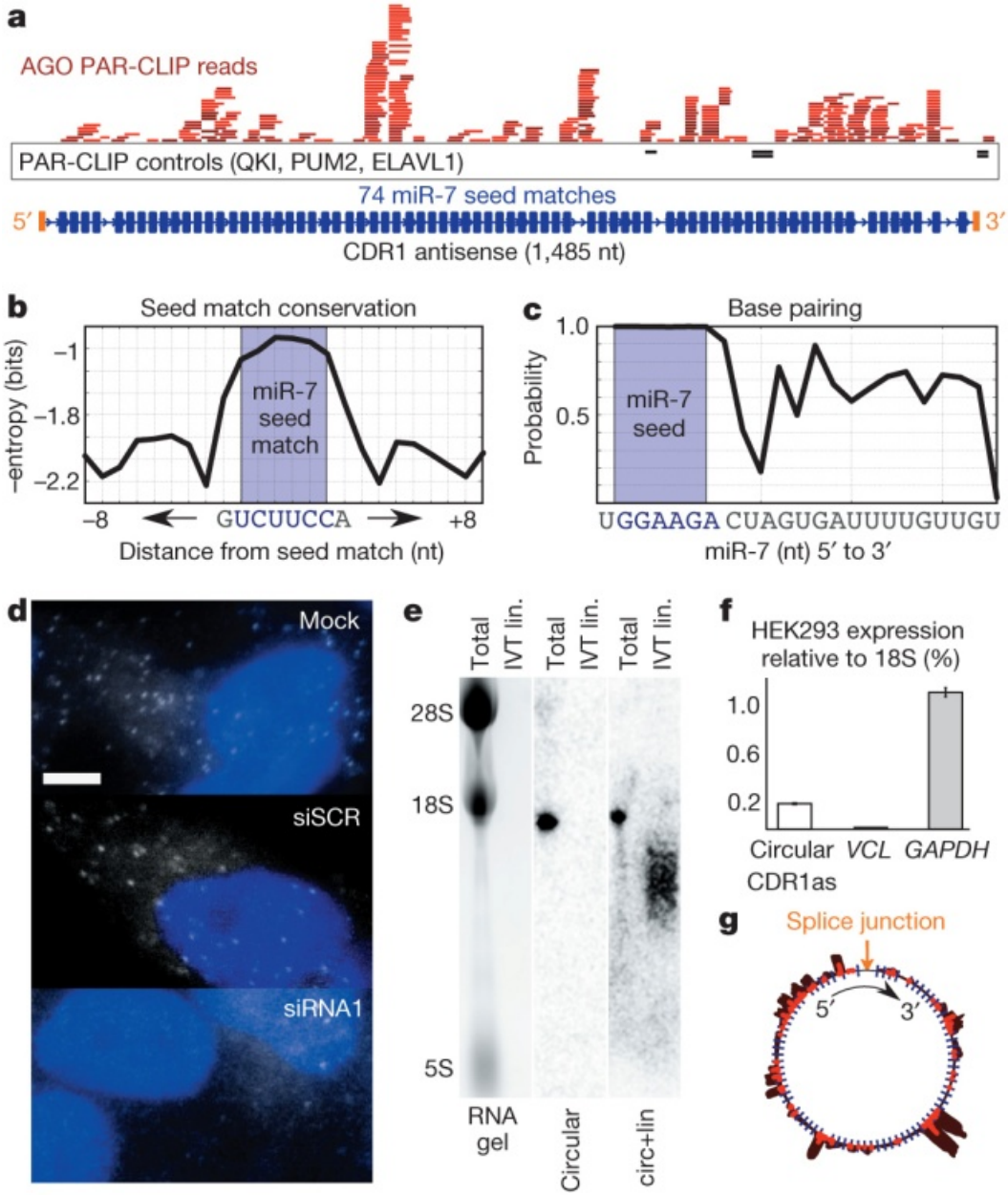
Sebastian Memczak ……. & Nikolaus Rajewsky



nature

Figure 2 | CircRNAs are stable transcripts with robust expression.
a, Human (hsa) ZRANB1 circRNA exemplifies the validation strategy. Convergent (divergent) primers detect total (circular) RNAs. Sanger sequencing confirms head-to-tail splicing. b, Divergent primers amplify circRNAs in cDNA but not genomic DNA (gDNA). GAPDH, linear control, size marker in base pairs. c, Northern blots of mock (2) and RNase R (1) treated HEK293 total RNA with head-to-tail specific probes for circRNAs. GAPDH, linear control. d, e, circRNAs are at least 10-fold more RNase R resistant than GAPDH mRNA (d) and stable after 24 h transcription block

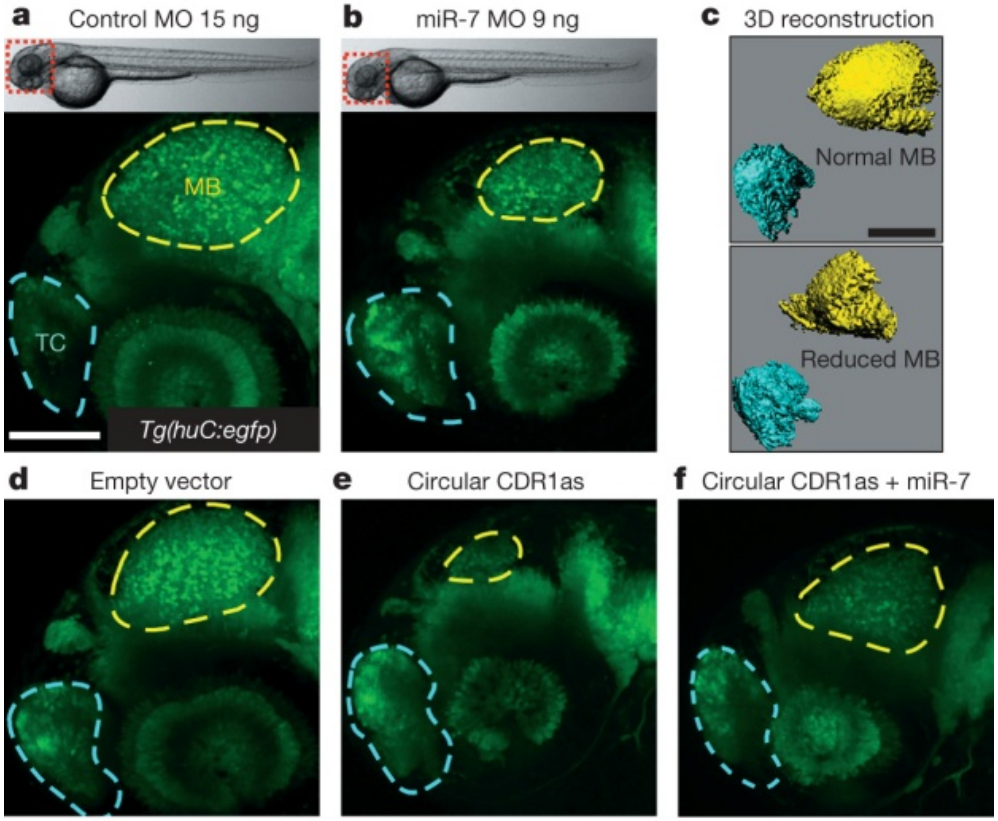
The circRNA CDR1as is bound by the miRNA effector protein AGO, and is cytoplasmic.



a, CDR1as is densely bound by AGO (red) but not by unrelated proteins (black). Blue boxes indicate miR-7 seed matches. nt, nucleotides. b, c, miR-7 sites display reduced nucleotide **variability across 32 vertebrate genomes** (b) and high base-pairing probability within seed matches (c). d, CDR1as RNA is cytoplasmic and disperse (white spots; single-molecule RNA FISH; maximum intensity merges of Z-stacks). siSCR, positive; siRNA1, negative control. Blue, nuclei (DAPI); scale bar, 5 µm (see also Supplementary Fig. 10 for uncropped images). e, Northern blotting detects circular but not linear CDR1as in HEK293 RNA. Total, HEK293 RNA; circular, head-to-tail probe; circ+lin, probe within splice sites; IVT lin., in vitro transcribed, linear CDR1as RNA. f, Circular CDR1as is highly expressed (qPCR, error bars indicate standard deviation). g, CDR1as. Blue, seed matches; dark red, AGO PAR-CLIP reads; bright red, crosslinked nucleotide conversions.

nature

In zebrafish, knockdown of miR-7 or expression of CDR1as causes midbrain defects.



nature

zebrafish has lost the cdr1 locus, whereas miR-7 is conserved and highly expressed in the embryonic brain

a, b, Neuronal reporter (Tg(huC:egfp)) embryos (top, light microscopy) 48 h post fertilization (bottom, representative confocal z-stack projections; blue dashed line, telencephalon (TC) (control); yellow dashed line, midbrain (MB)). Embryos after injection of 9 ng miR-7 morpholino (MO) (b) display a reduction in midbrain size. Panel a shows a representative embryo injected with 15 ng control morpholino. c, Three-dimensional volumetric reconstructions. d, Empty vector control. e, Expression vector encoding human circular CDR1as. f, Rescue experiment with miR-7 precursor.