long non coding RNAs

Why RNA?

Economist.com

The RNA revolution

Biology's Big Bang

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

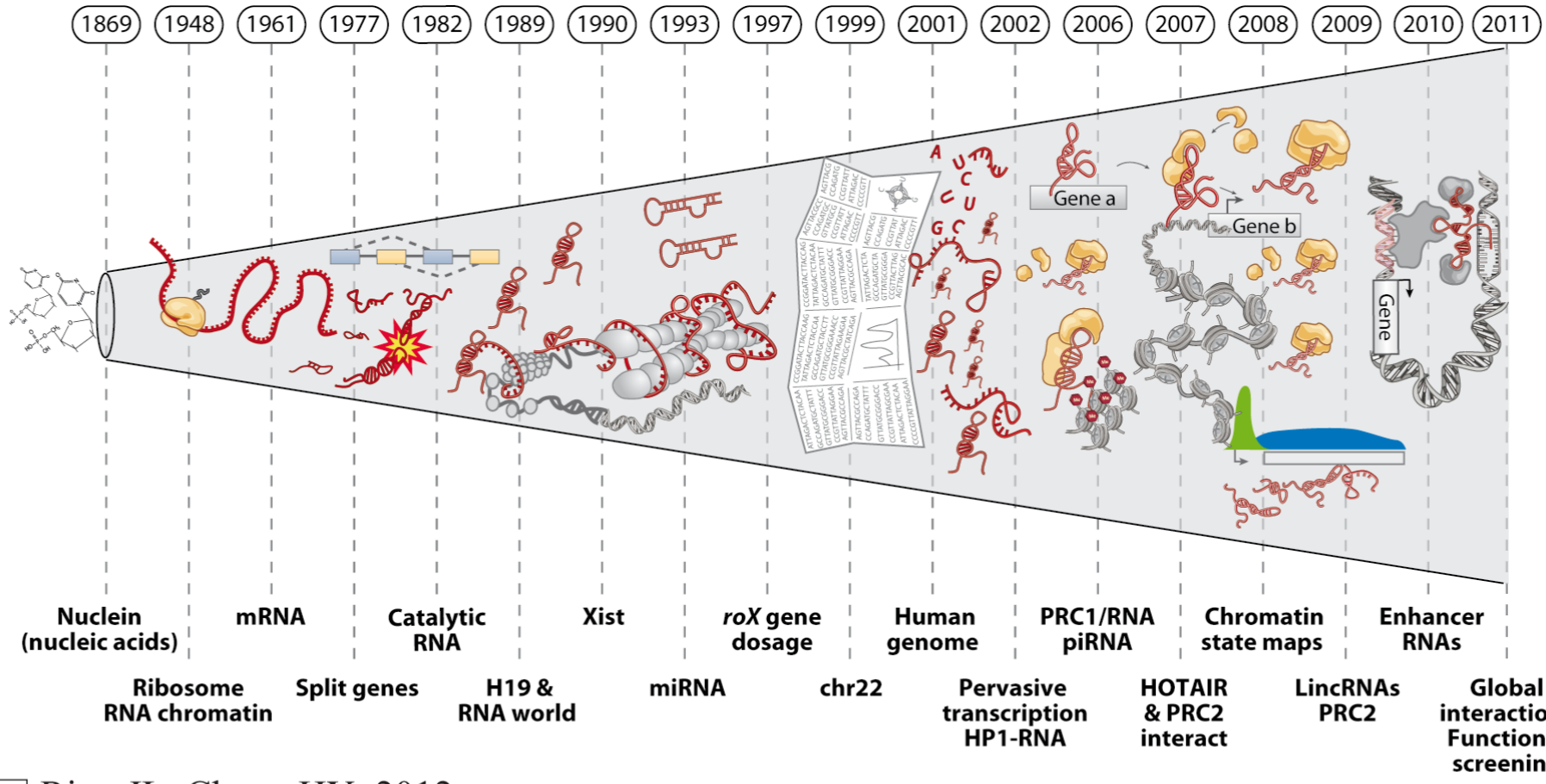
- a whole universe of RNA - predominantly of the noncoding variety - has remained hidden from view, until recently…….many new and unexpected functions

- 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?

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



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)

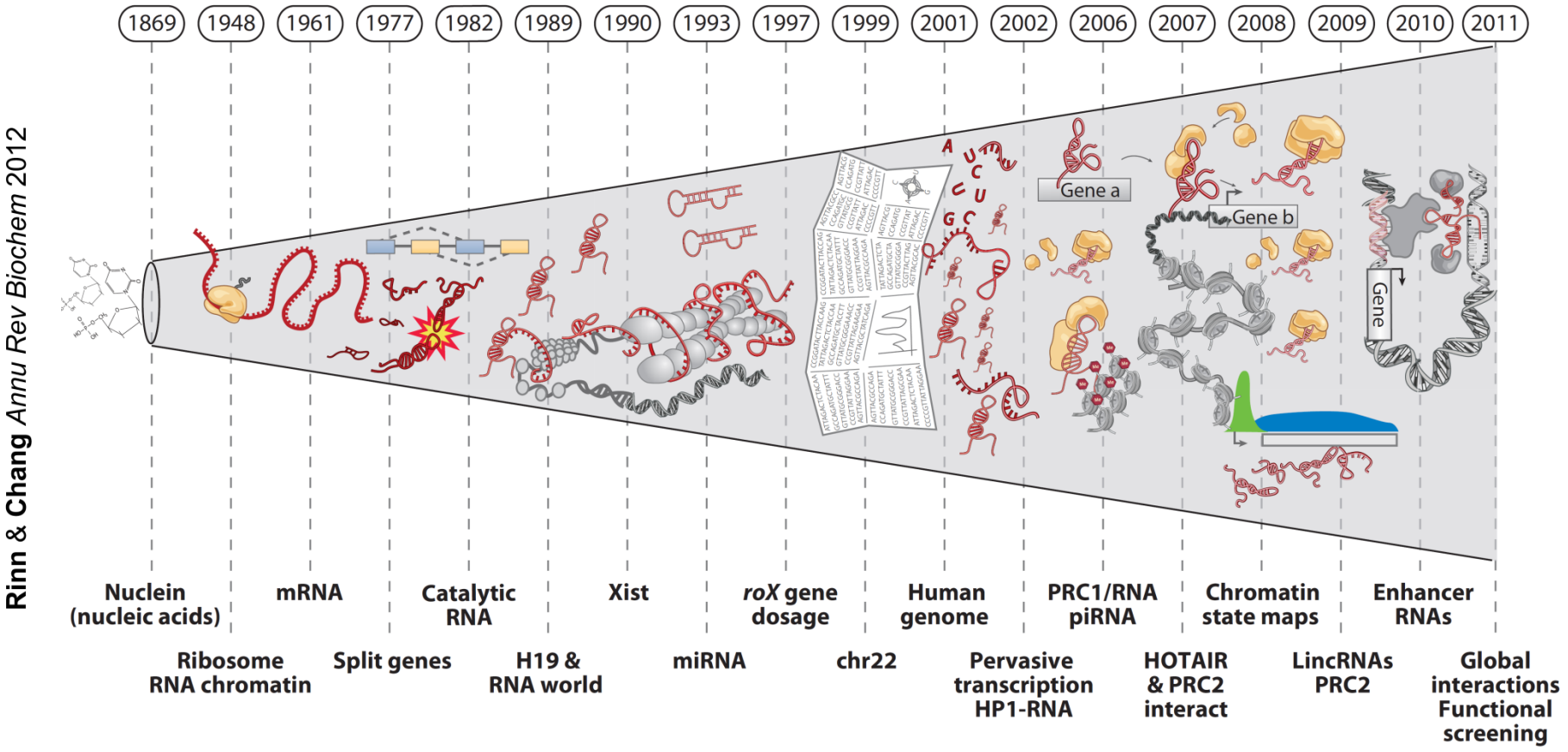
The ***central dogma*** of molecular biology states that DNA is transcribed into RNA, which in turn is translated into proteins.

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.

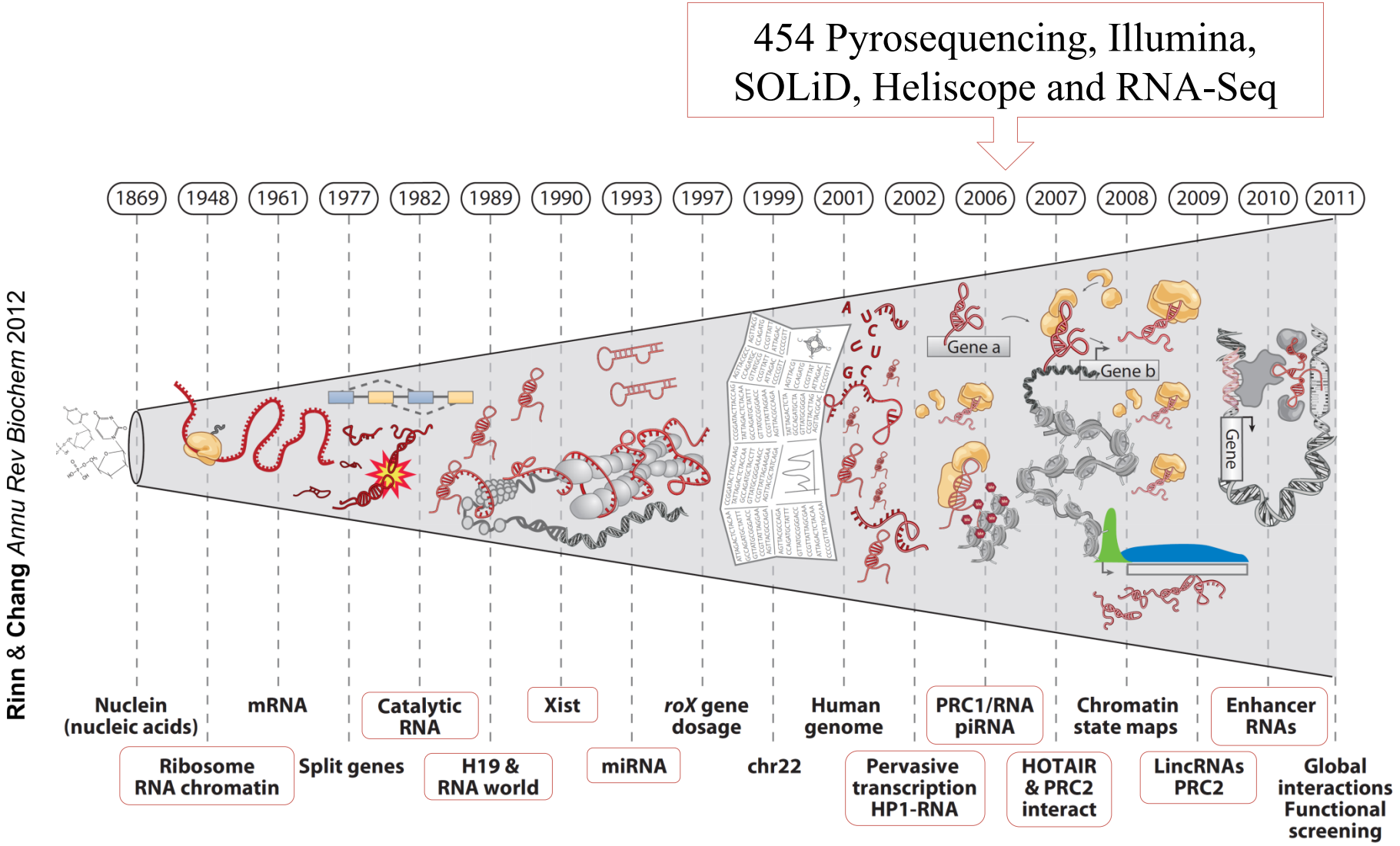
Genome-wide antisense transcription drives mRNA processing in bacteria

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

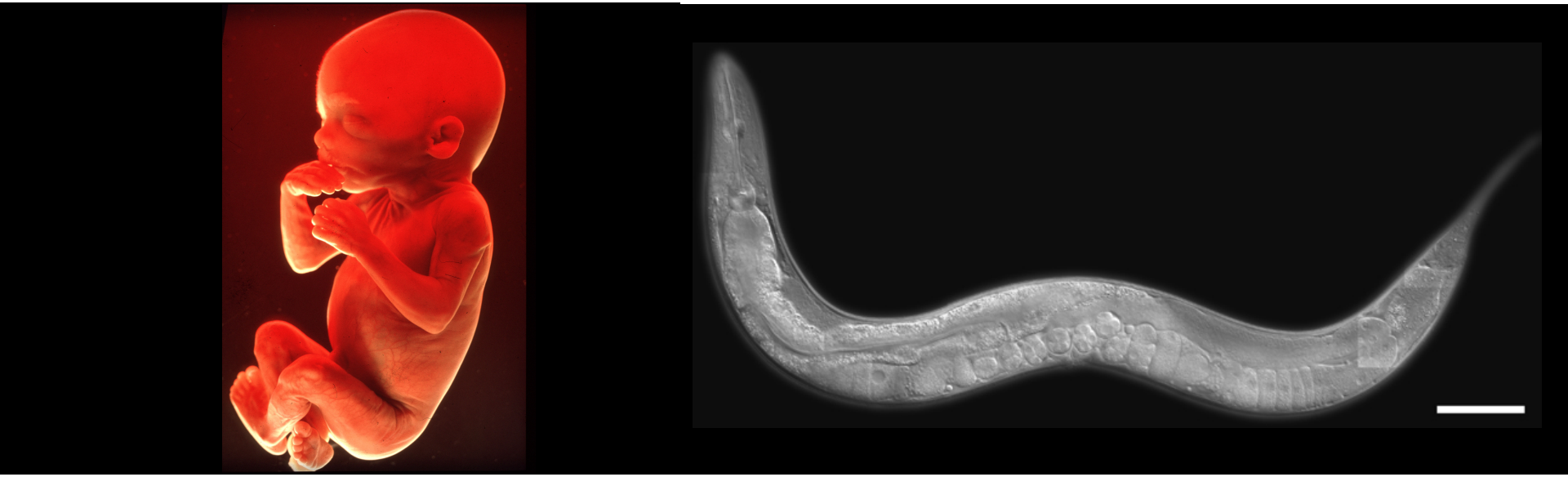
250 different cancer cell lines samples

152 human post-mortem tissues samples

271 mouse developmental tissue samples

22000 genes encoding for proteins

The genetic basis of developmental complexity

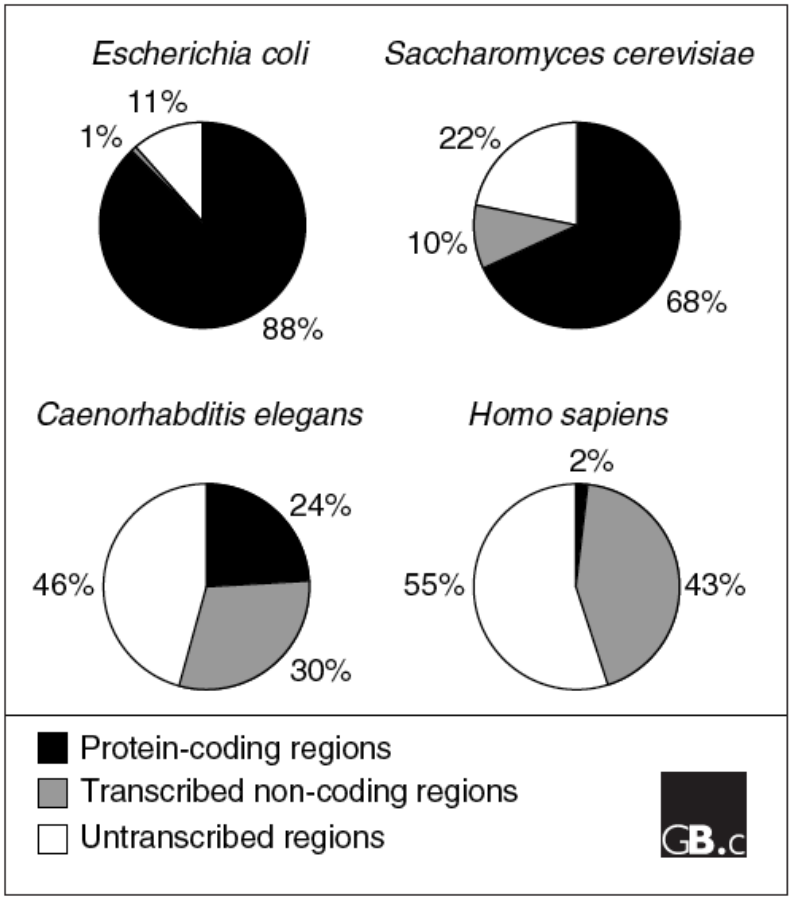


C.elegans - **1000 cells**
H.sapiens - **10^{14} cells - and 10^{11} neurons!!!**

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

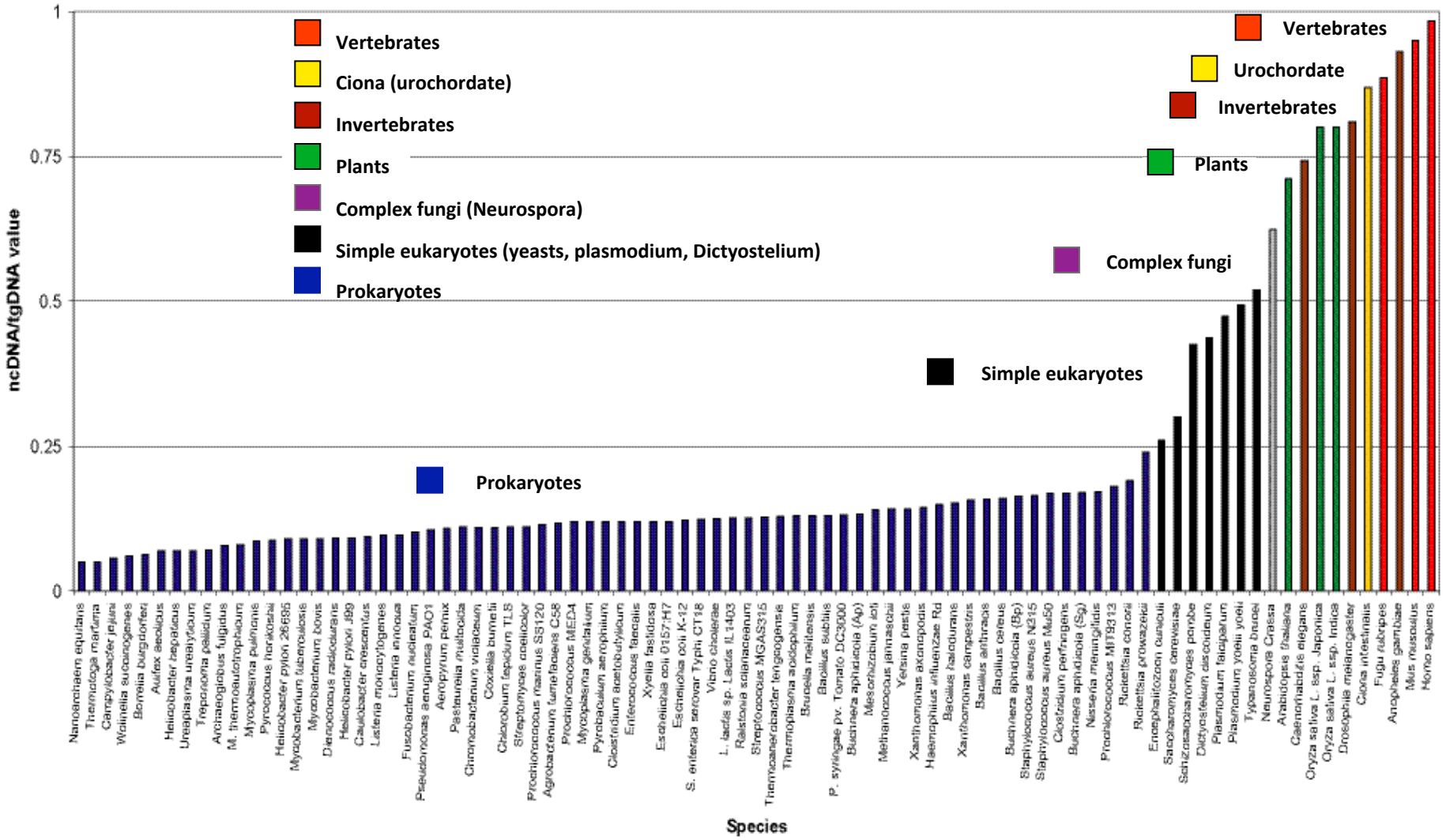
152 human post-mortem tissues samples

271 mouse developmental tissue samples

22000 genes encoding for proteins

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

The proportion of noncoding DNA broadly increases with developmental complexity



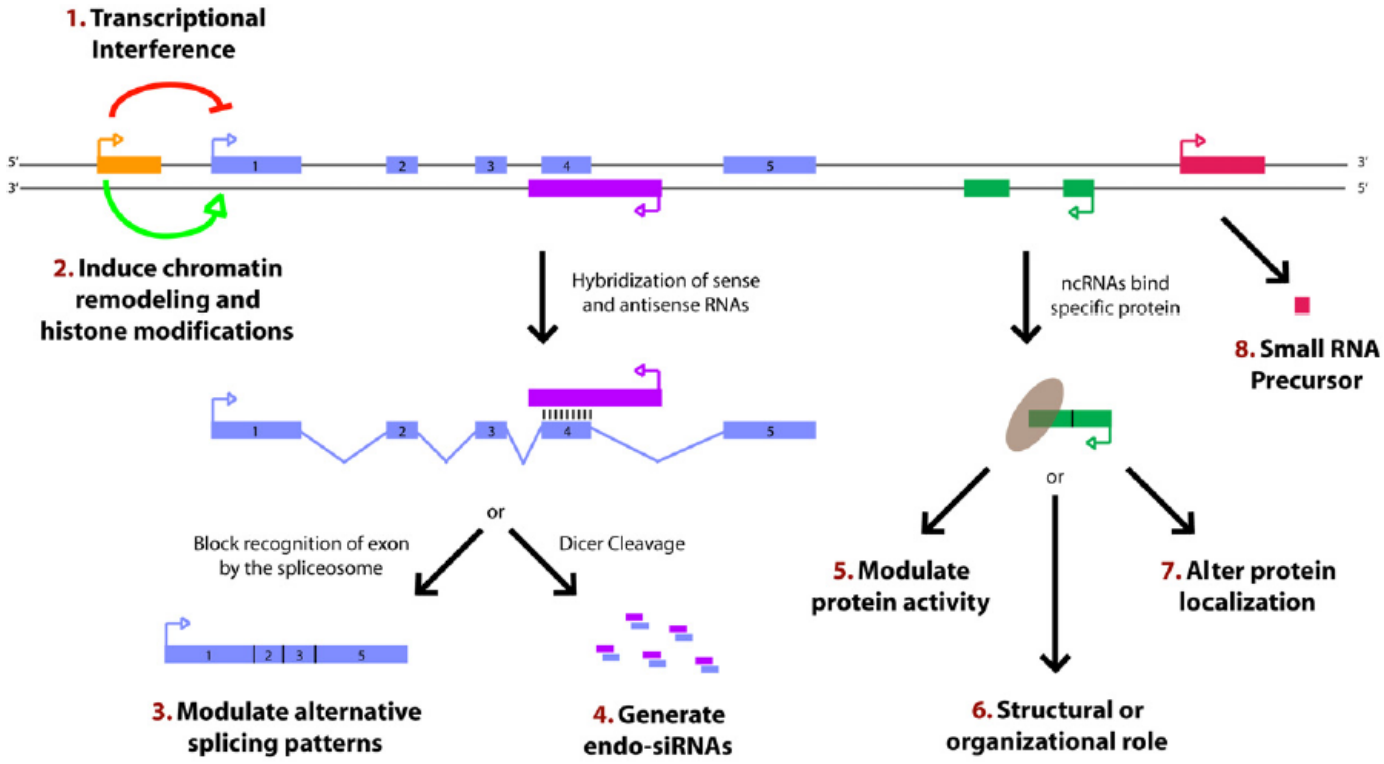
J.S. Mattick *Nature Reviews Genetics* 5, 316-323 (2004).

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

Long non-coding RNAs

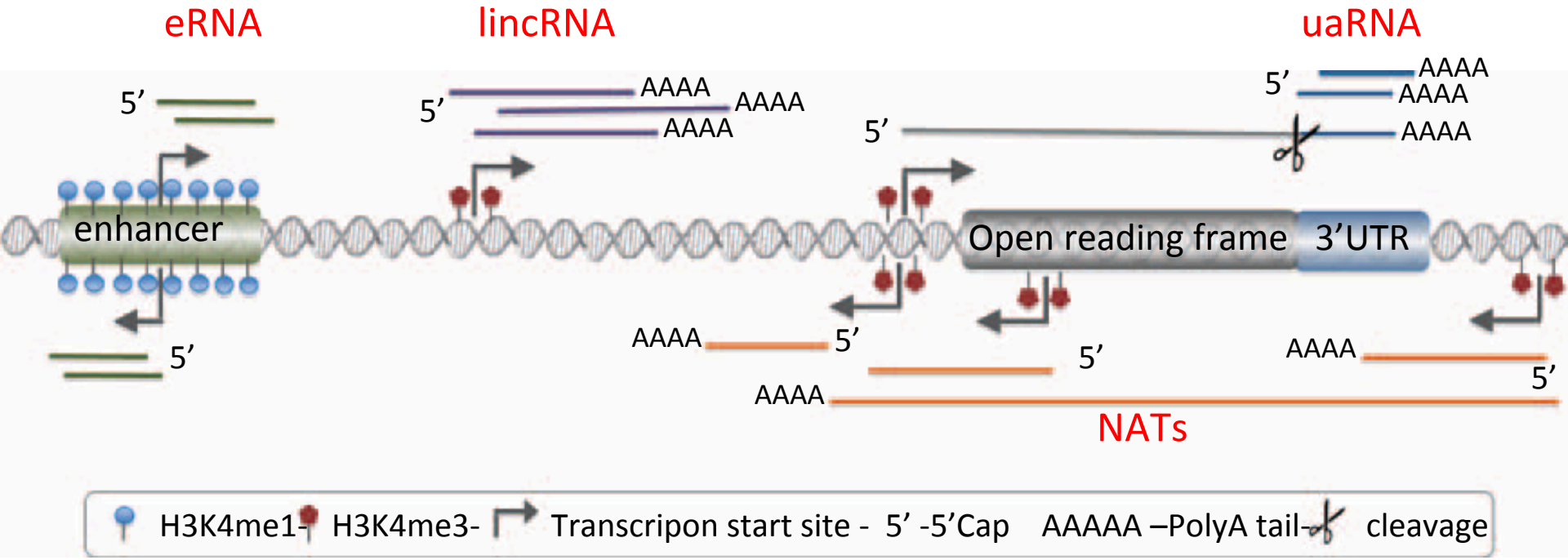
2001 to the Future Break the Dogma

List of long non-coding RNAs

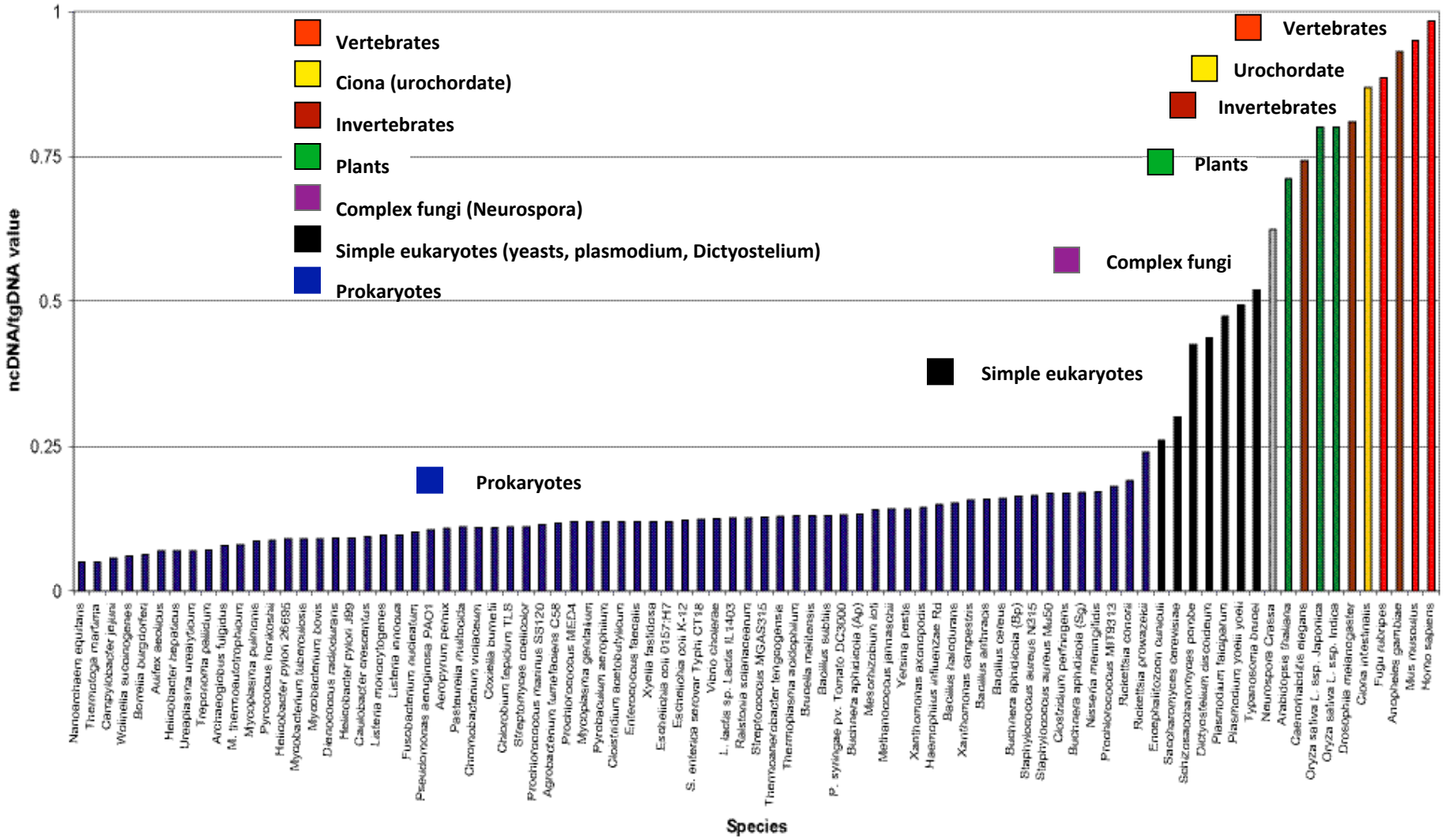


J E Wilusz et al Genes & Development 2009

The current landscape of non coding RNAs



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)

Scaling of regulatory architecture

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

- Transcriptional fators, alternative splicing

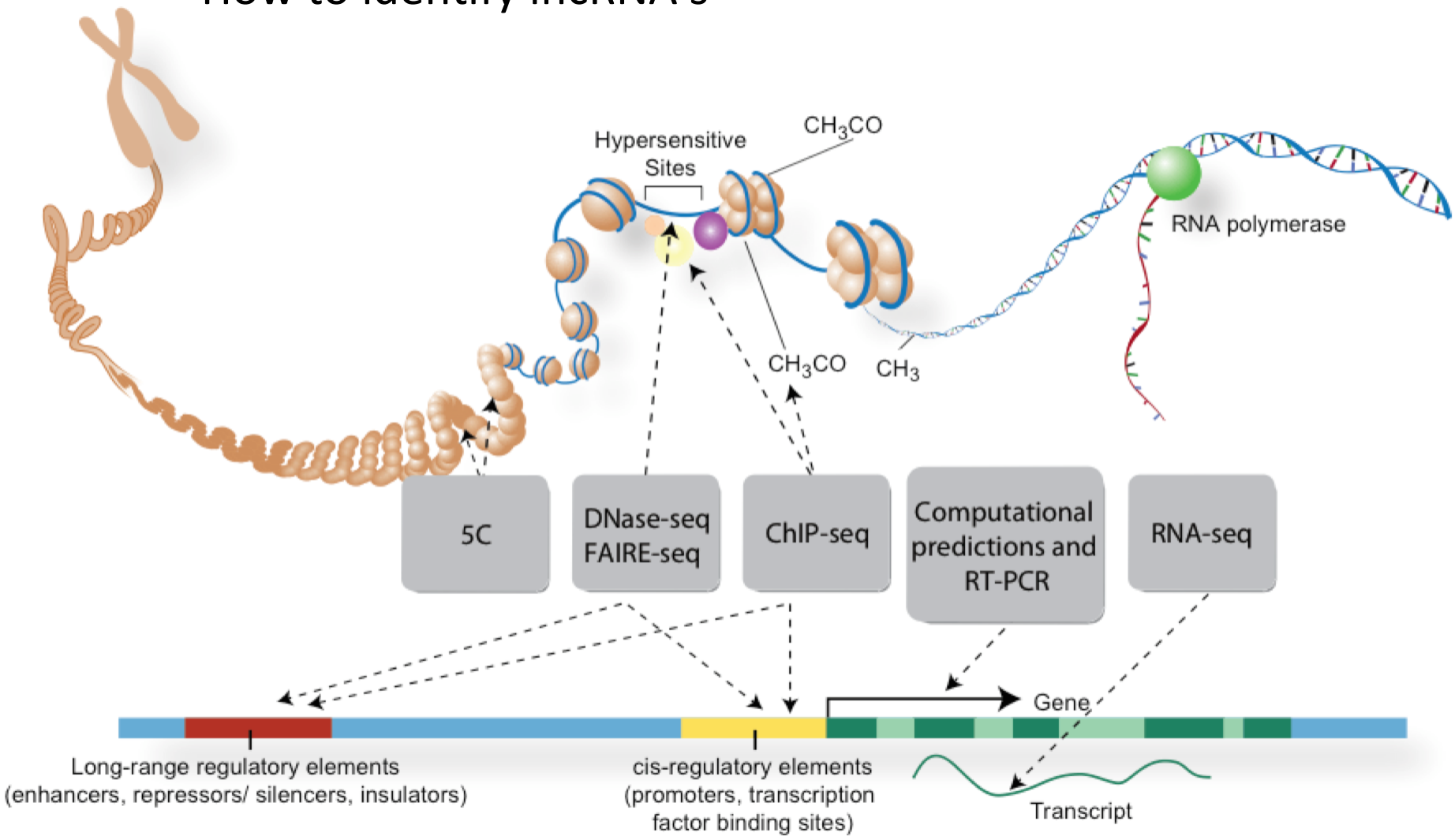
Dramatic expansion of regulatory complexity

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

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

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

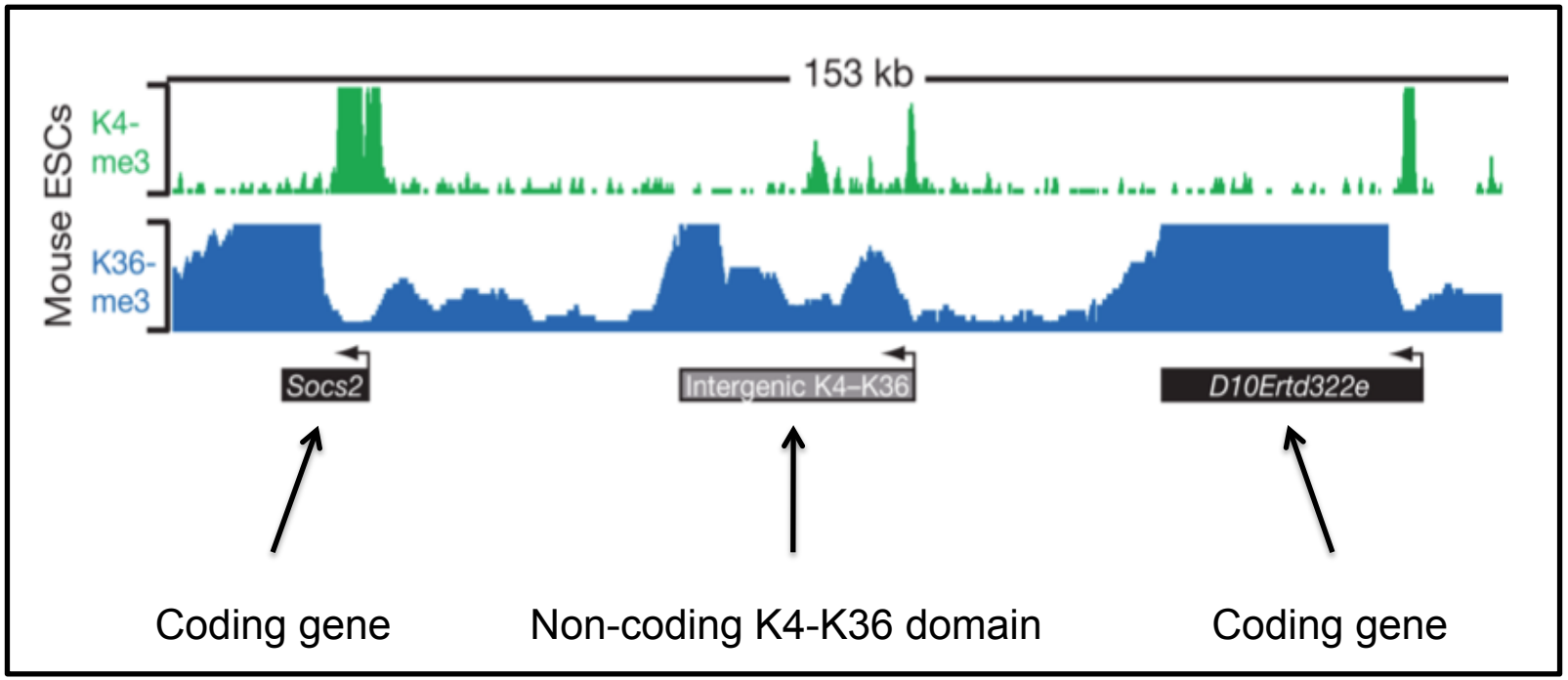
How to identify lncRNA s



lncRNAs are functional transcripts

K4-K36 Domains

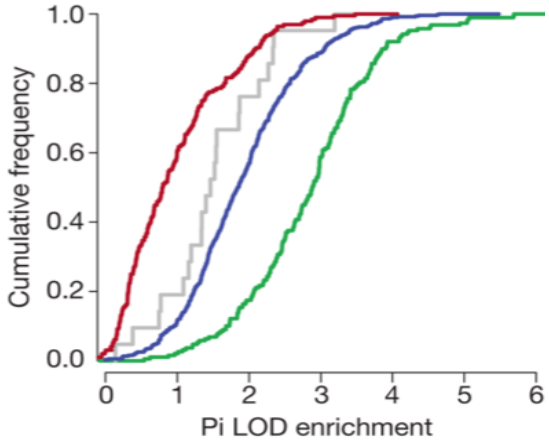
Actively Transcribed Pol II genes are marked by H3 lysine 4 trimethylation (promoter) and H3 lysine 36 trimethylation (whole transcribed region)



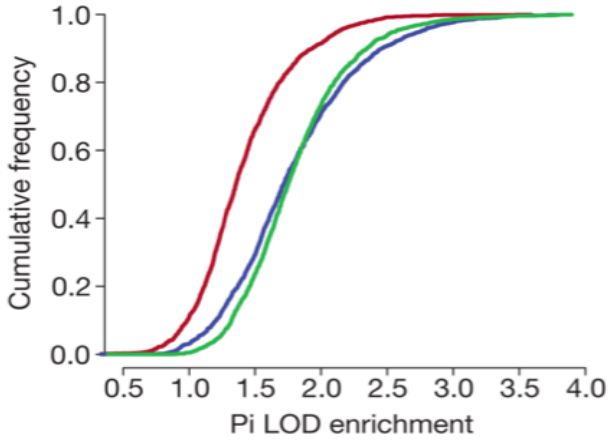
Some intergenic regions and active genes share a common chromatin signature

lncRNAs are functional transcripts

Sequence conservation

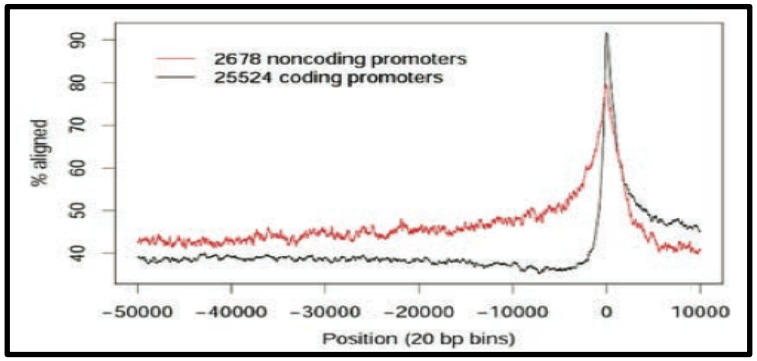


Coding sequences are strongly conserved among mammals (green) whereas lncRNAs show lower scores (blue), but still more than introns and UTRs.



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

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



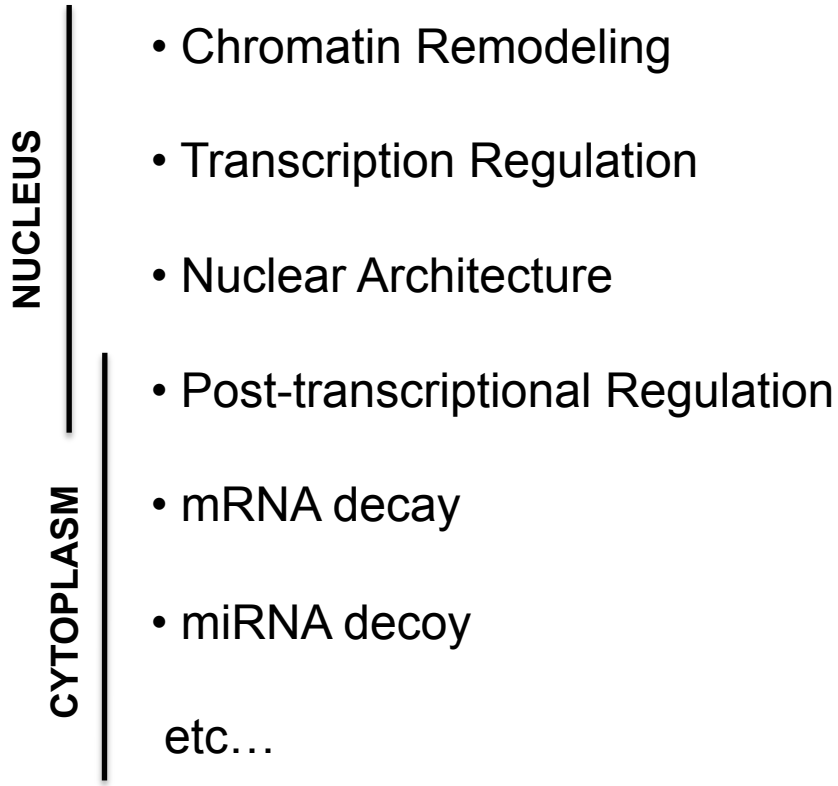
Functions of lncRNAs

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

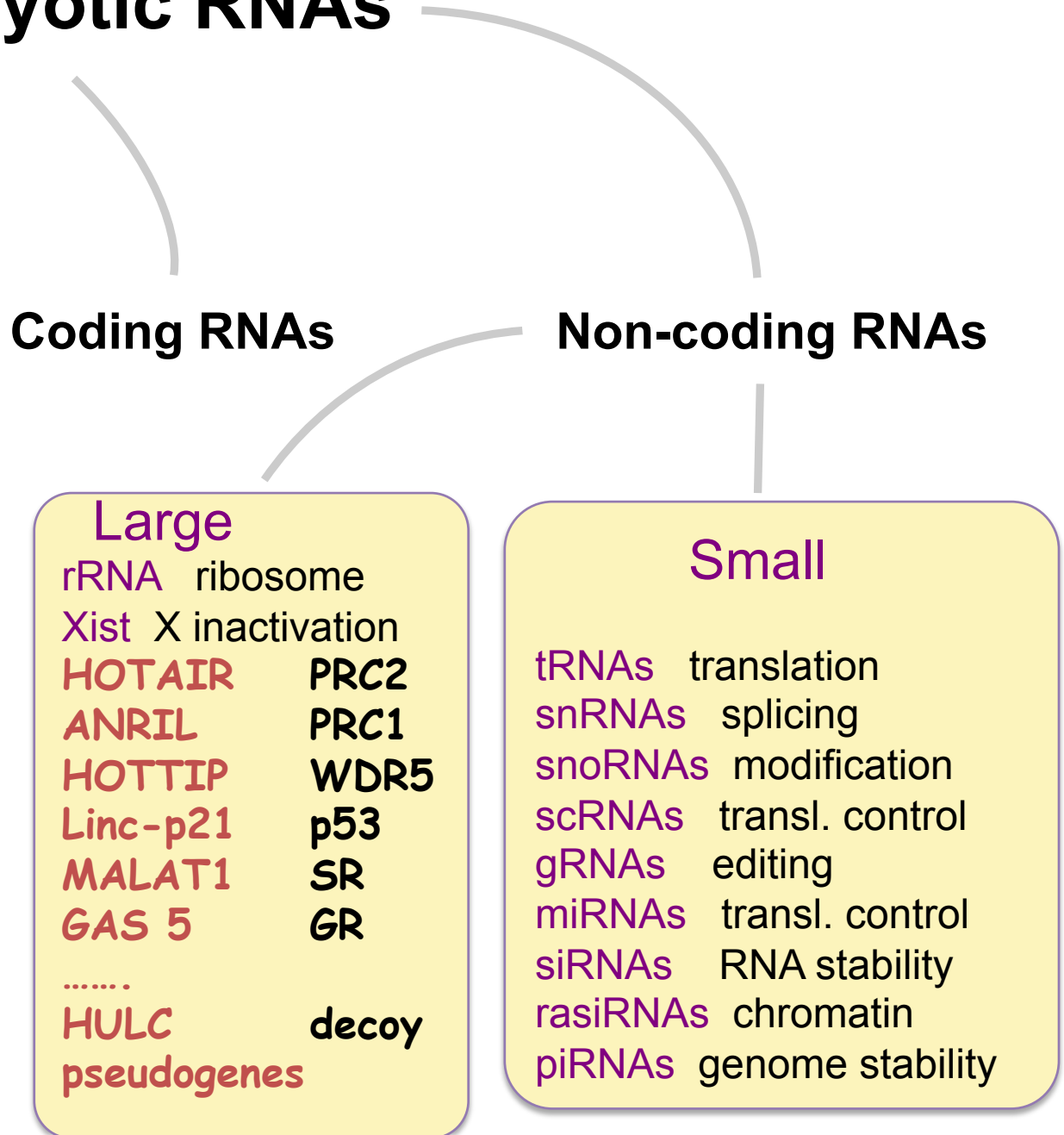
- Sequence conservation average is lower than in coding gene but there are peaks in specific functional regions
- Sequence-dependent functions with respect to structure are easier to discover (sequence analysis) and to study (mutation analysis)
- Many lncRNAs can have functions depending on their 3D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

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

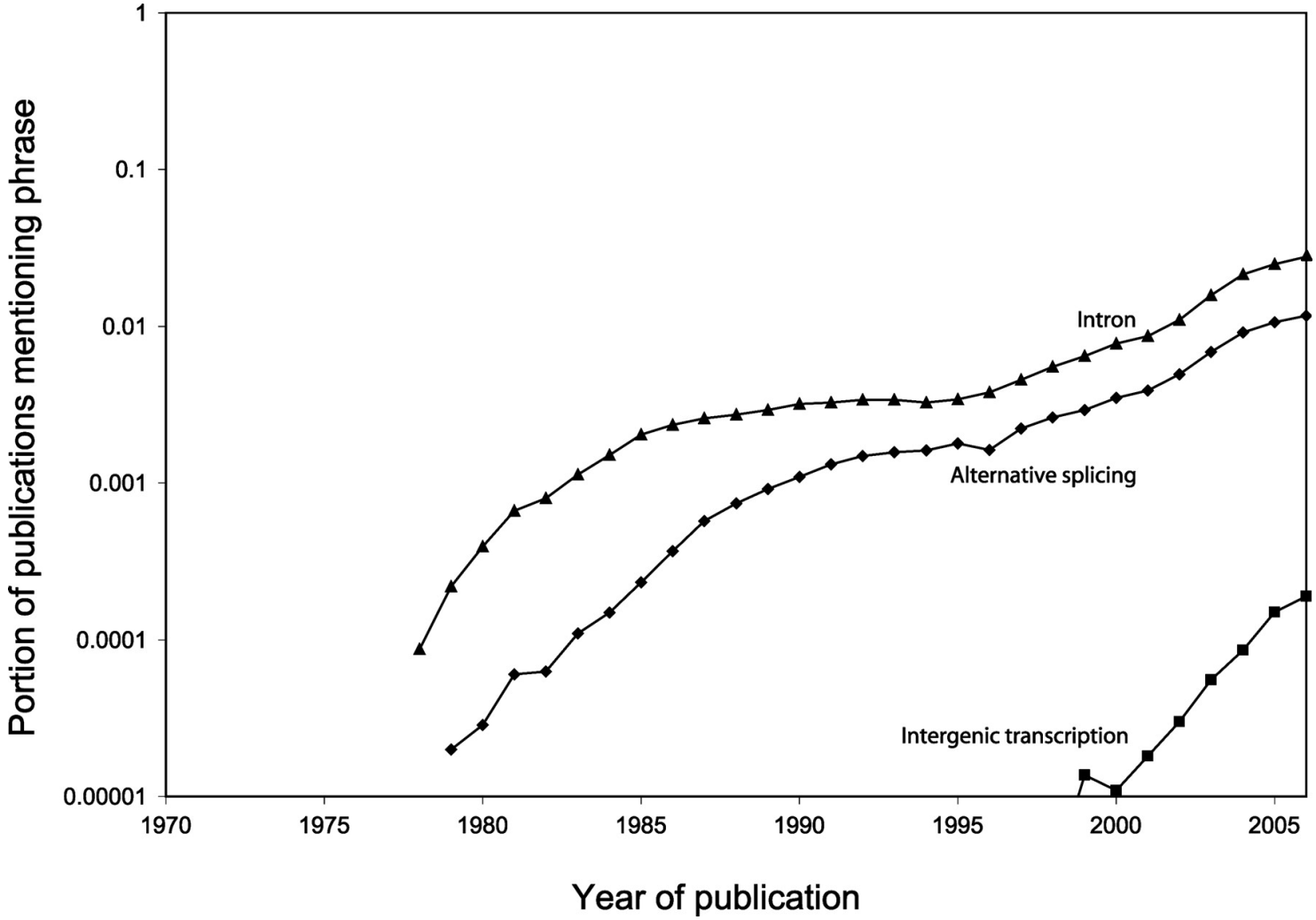
Functions of lncRNAs



Eukaryotic RNAs



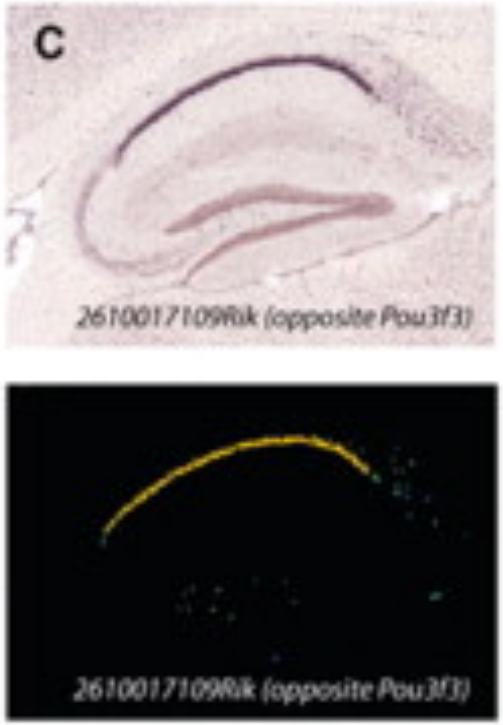
Keyword analysis and complexity of genes.



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

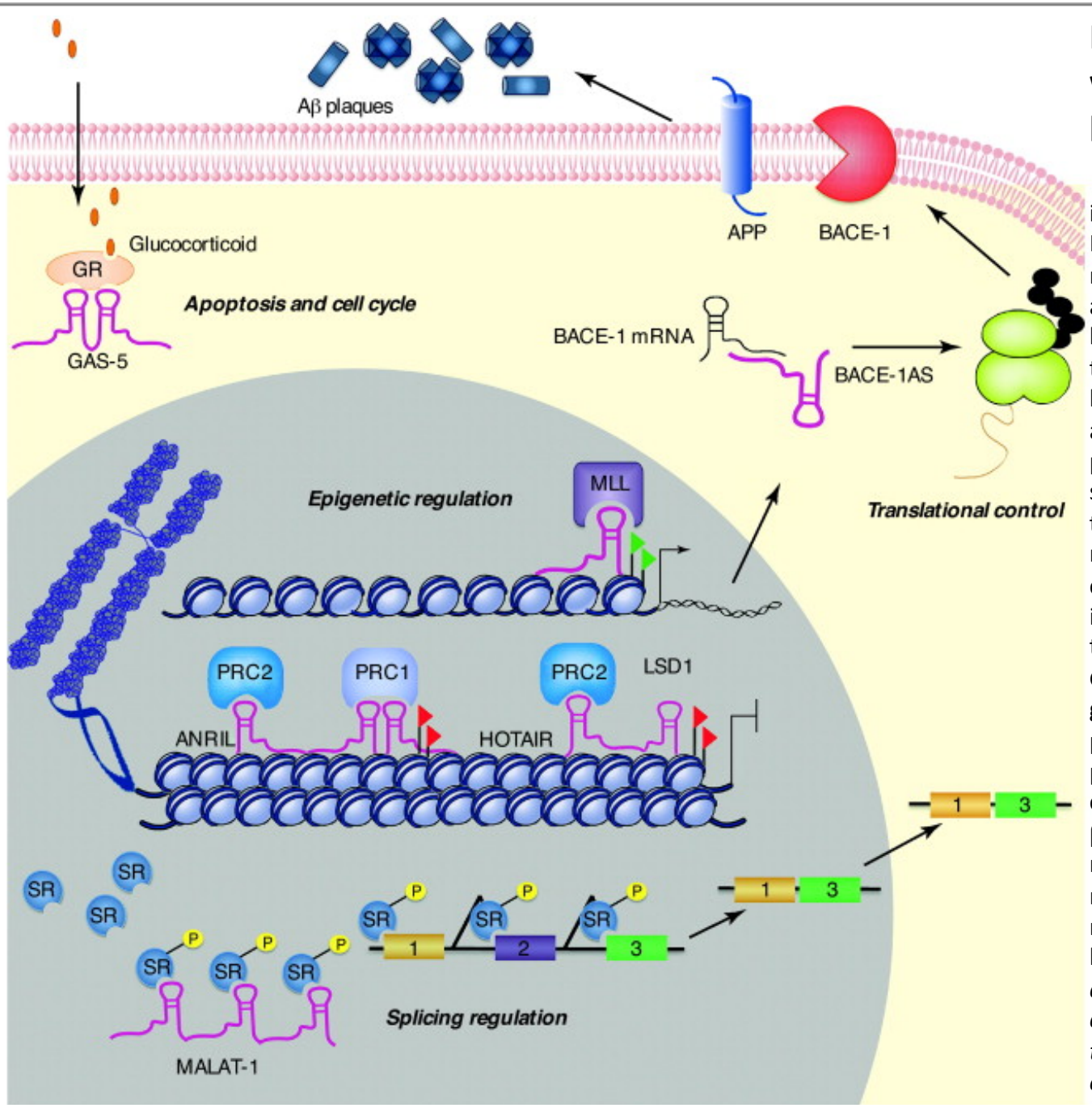
Why has this system gone unnoticed?

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



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

LncRNAs participate in a wide repertoire of biological process.



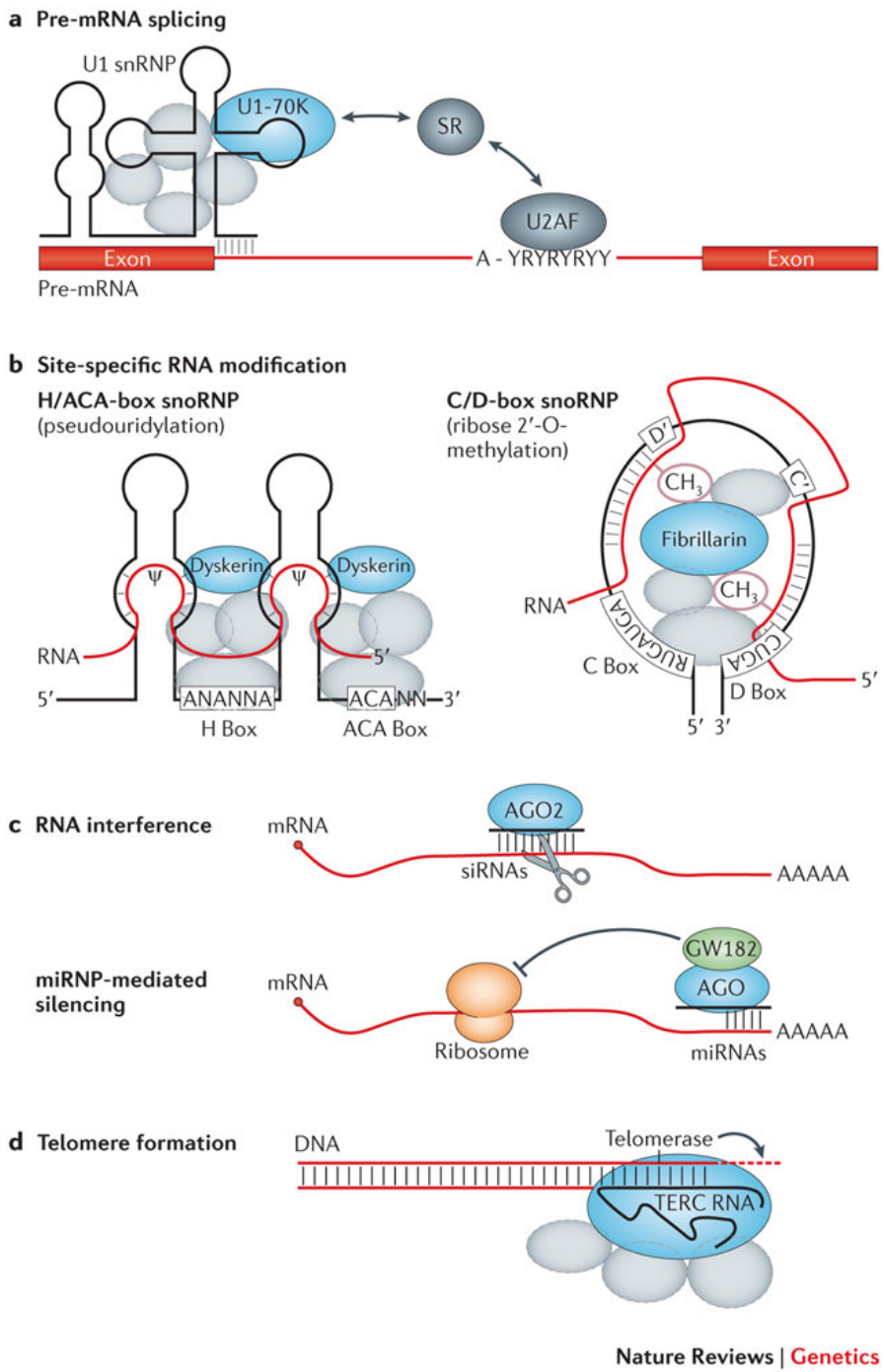
Recent examples of mutated lncRNAs implicated in disease include ANRIL and HOTAIR that bind to chromatin-remodeling complexes PRC1 and PRC2 to alter chromatin and transcription. GAS5 lncRNA acts as a decoy for the GR transcription factor and prevents GR from binding to DNA and transcriptional activation. 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.*

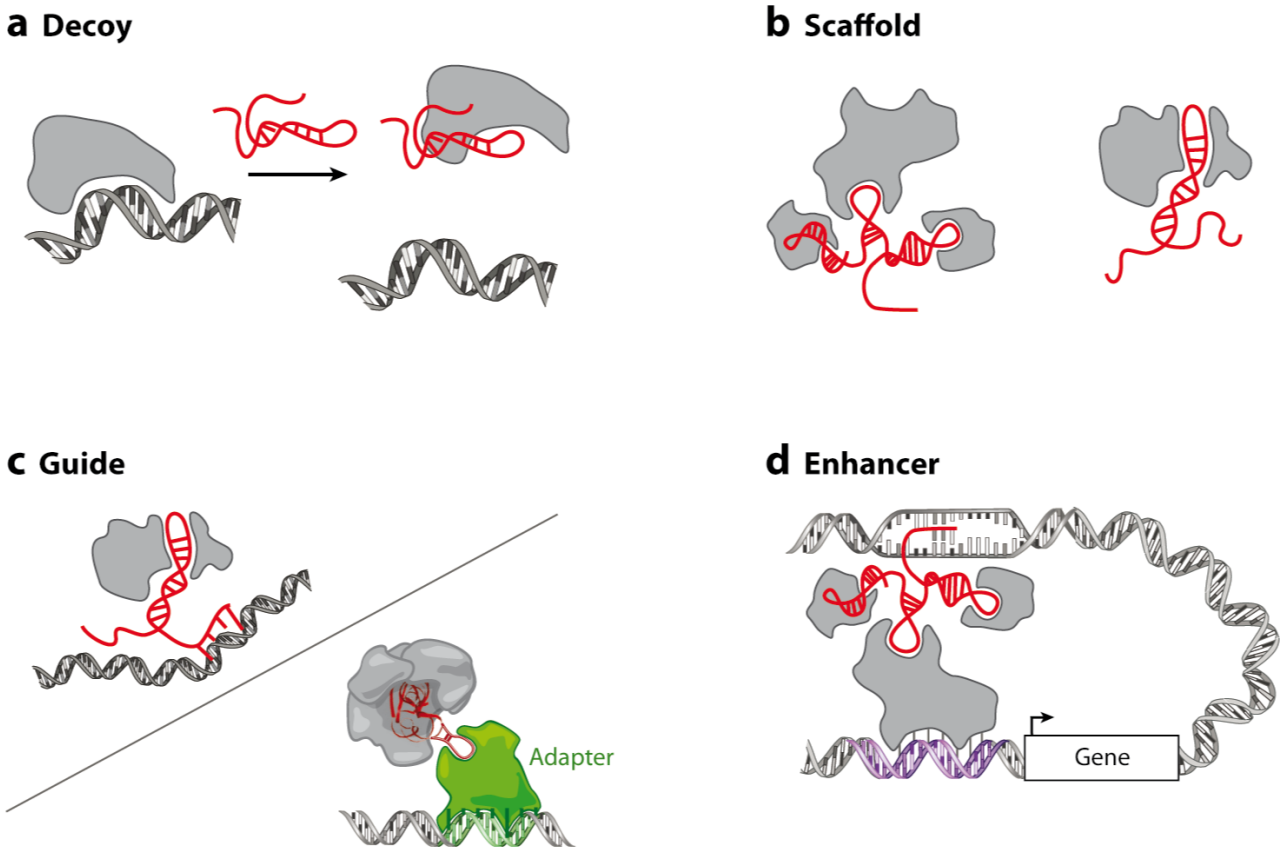
Long non-coding RNAs: new players in cell differentiation and development

Alessandro Fatica & Irene Bozzoni
Nature Reviews Genetics (2013)

RNA has the ability to bind to other nucleic acids as well as to proteins

RNPs





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

Models of long noncoding RNA (lncRNA) mechanisms of action. (*a*) The lncRNAs can act as decoys that titrate away DNA-binding proteins, such as transcription factors. (*b*) These lncRNAs may act as scaffolds to bring two or more proteins into a complex or spatial proximity and (*c*) may also act as guides to recruit proteins, such as chromatin modification enzymes, to DNA; this may occur through RNA-DNA interactions or through RNA interaction with a DNA-binding protein. (*d*) Such lncRNA guidance can also be exerted through chromosome looping in an enhancer-like model, where looping defines the *cis* nature and spread of the lncRNA effect.

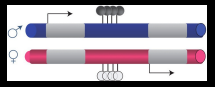
long non coding RNAs

AGING-CANCER

Telomere function

hTR lncRNA

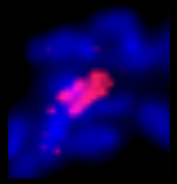
DEVELOPMENTAL DEFECTS



HOX gene control

HOTAIR lncRNA

Xist lncRNA



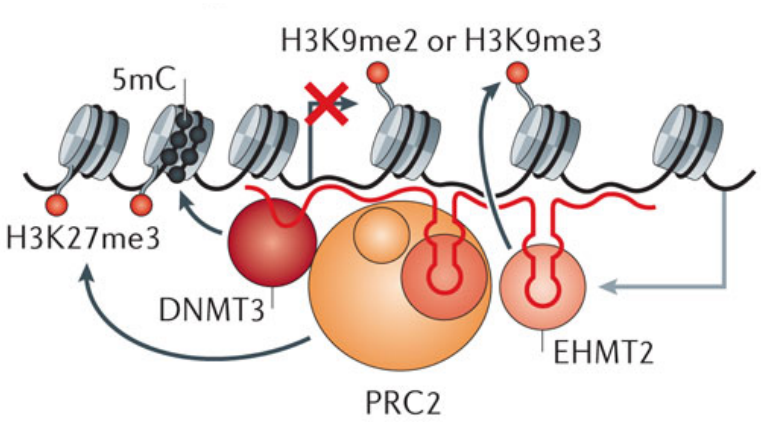
X inactivation Silencing 1000 genes on the Xi

LETHAL

Nuclear lncRNAs

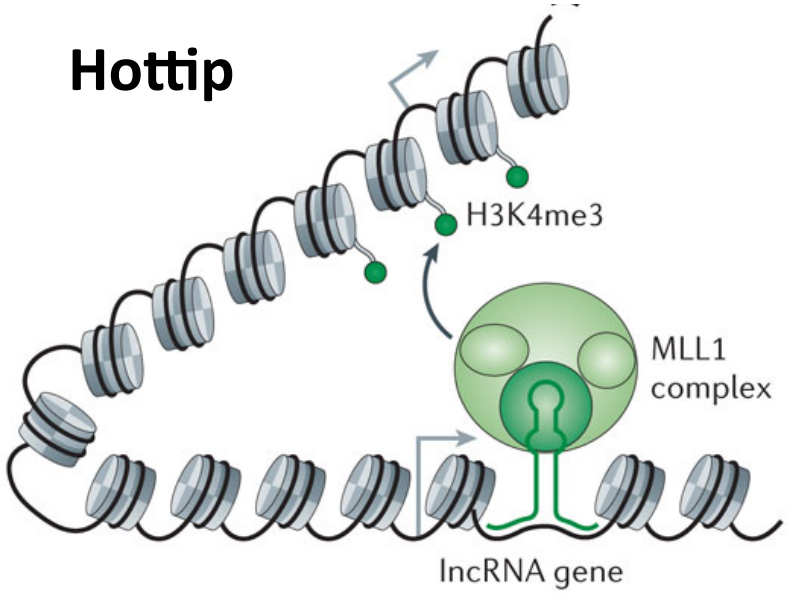
Nuclear lncRNAs

Xist



repressors

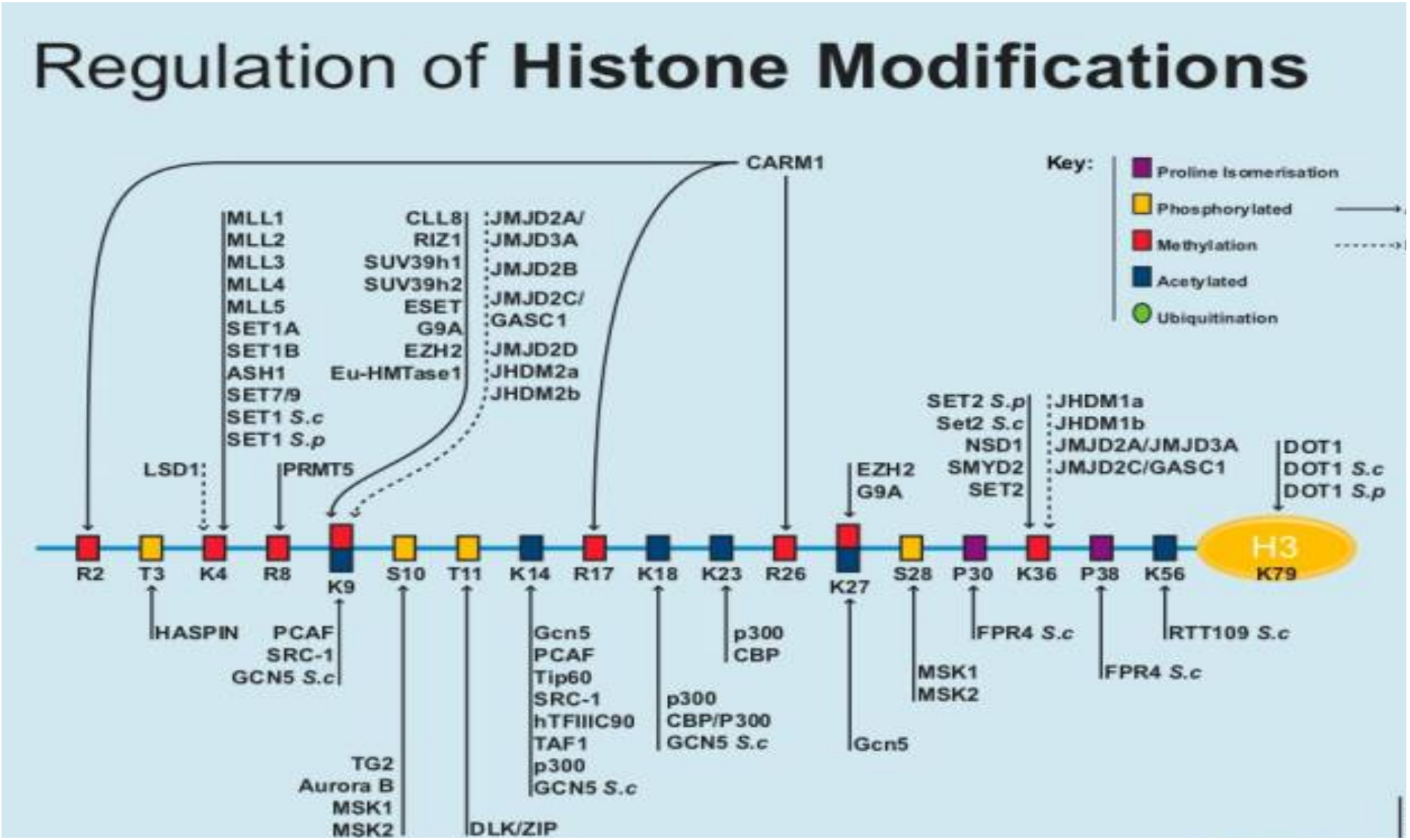
Hottip



activators

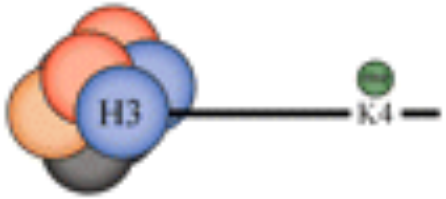
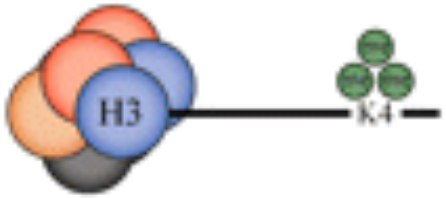
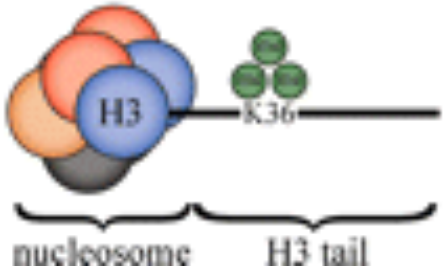
lncRNAs tether chromatin remodelling complexes to specific genomic loci

Nature Reviews Genetics, 2014
Fatica A. and Bozzoni I.



→ Adds modifications

⇢ Removes modifications

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

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

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

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

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

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

Two opposing groups of histone modifying complexes

Trithorax group (TrxG) of H3K4 HMTase

and

Polycomb complex (PcG) H3K27 HMTase,

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

Polycomb repressive complex (PRC)

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

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

PRC2 possesses histone H3K27 methyltransferase activity

PRC1 components ubiquitinate H2A following the H3K27 trimethylation

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

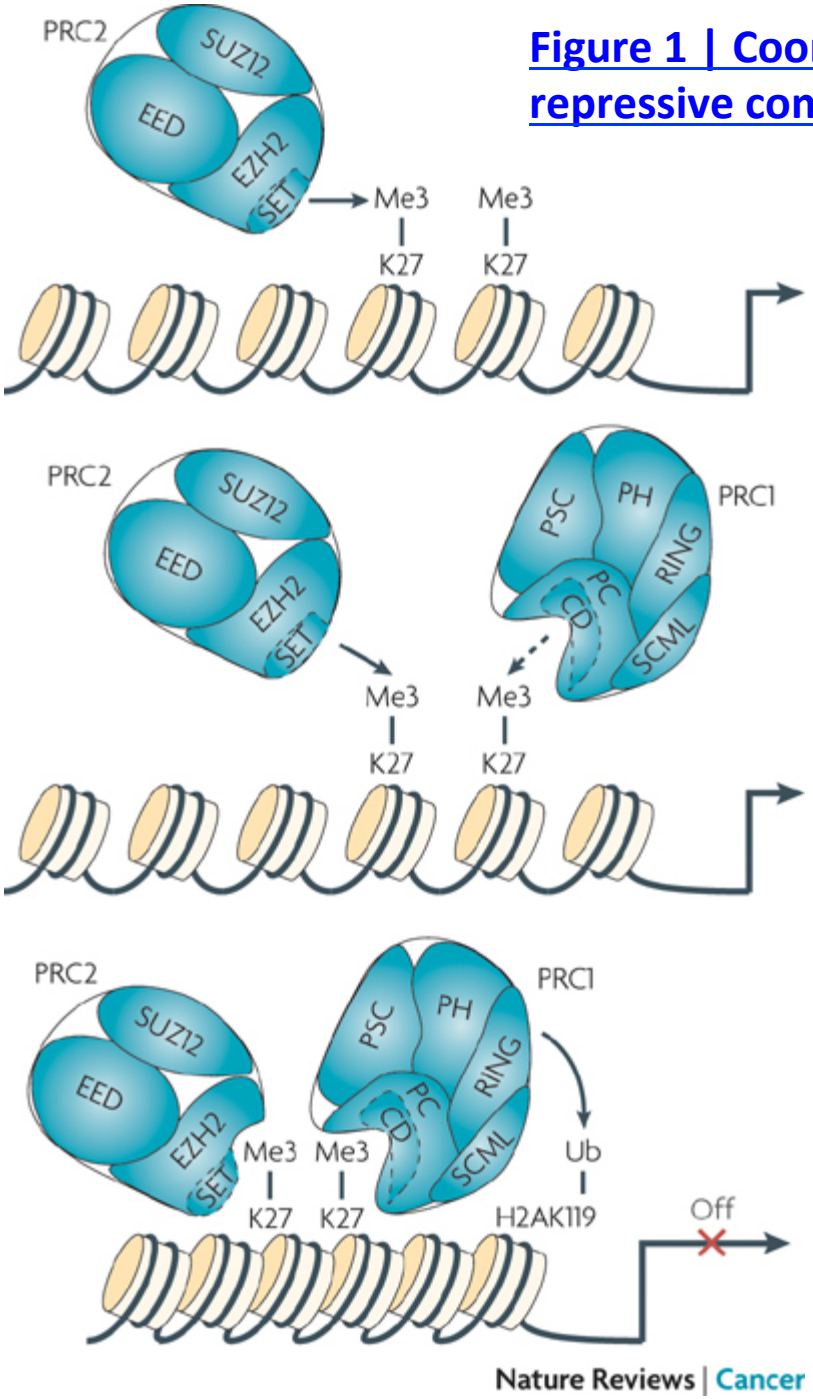
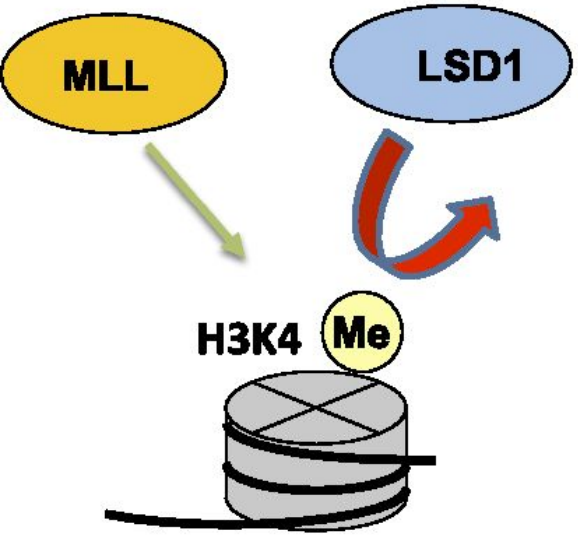
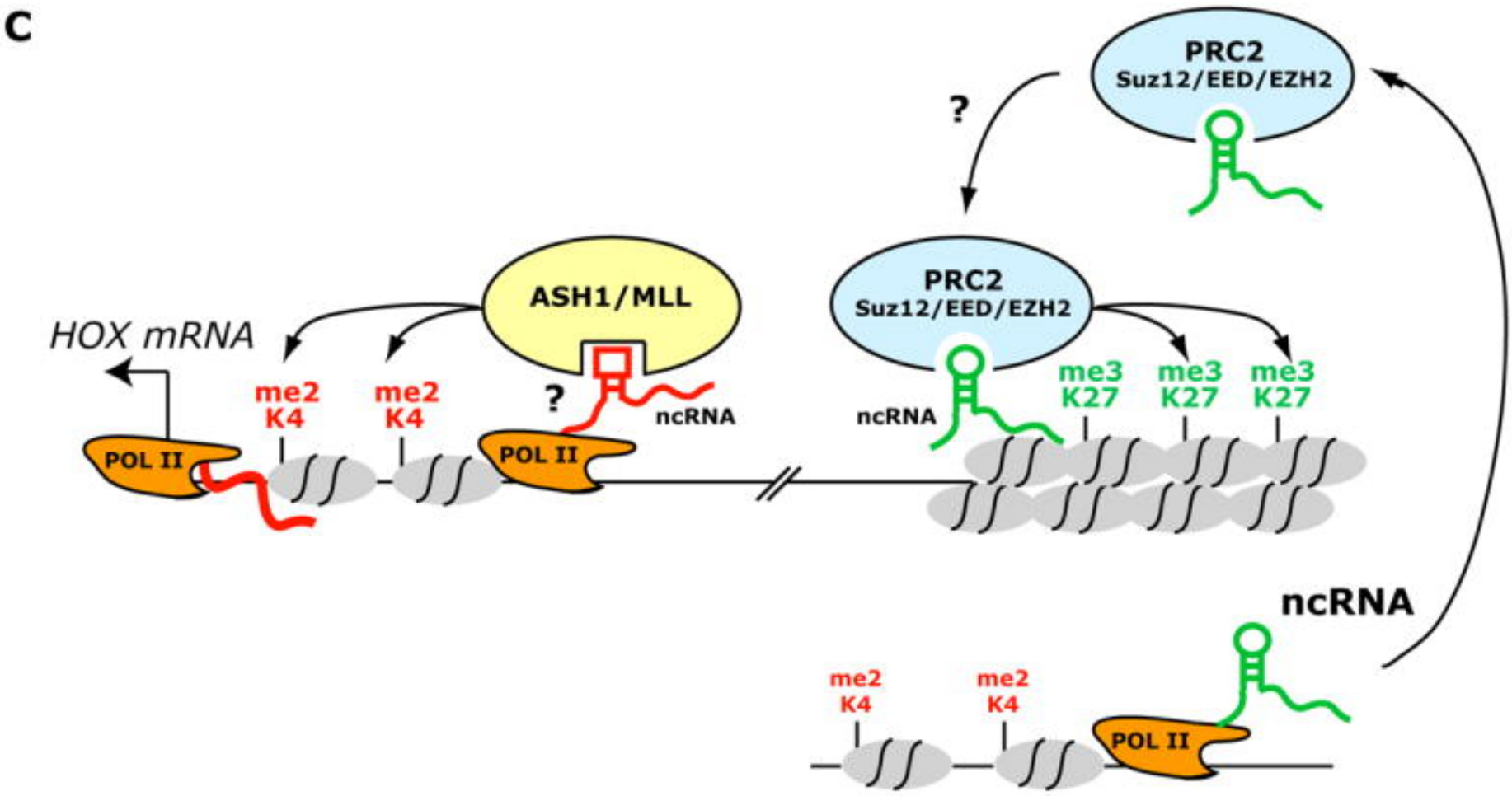


Figure 1 | Coordinated action of Polycomb repressive complexes.

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

How do chromatin remodelling complexes find their sites of action on the chromatin?



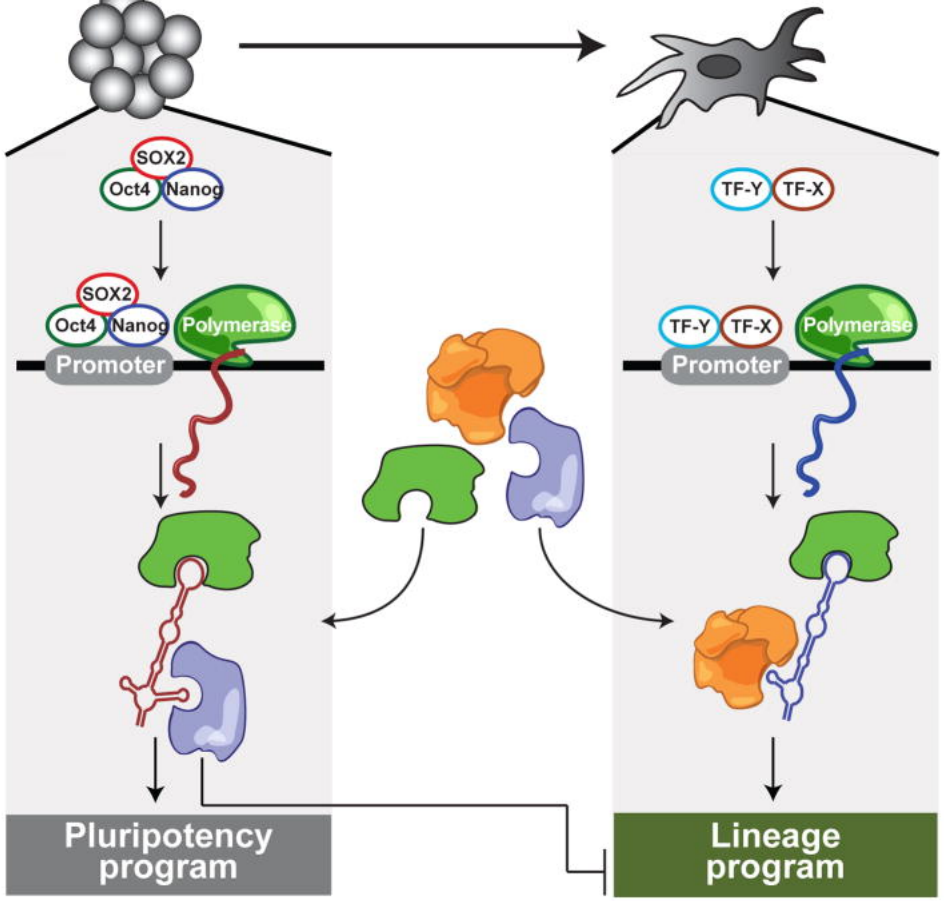


Model of long ncRNA regulation of chromatin domains via histone modification enzymes. Transcription of ncRNAs *in cis may increase the accessibility of TrxG proteins such as ASH1 or MLL or directly recruit them, leading to H3K4 methylation and* ***transcriptional activation*** *of the downstream HOX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

The modular scaffold model

lincRNAs act in the circuitry controlling pluripotency and differentiation

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



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

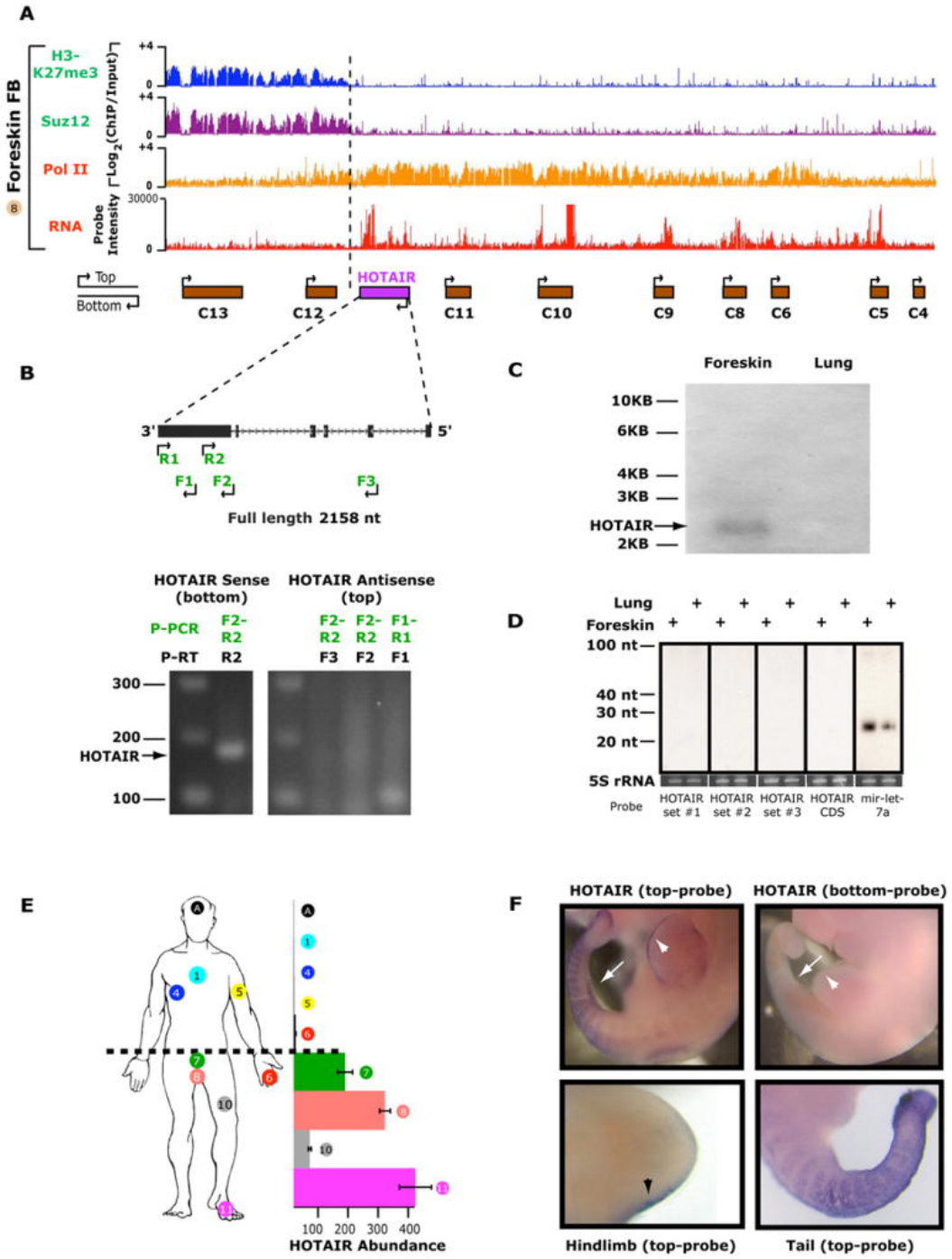
Left: ESC-specific transcription factors (such as Oct4, Sox2, and Nanog) bind to the promoter of a lincRNA gene and drive its transcription. The lincRNA binds to ubiquitous regulatory proteins, giving rise to cell-type specific RNA–protein complexes. Through different combinations of protein interactions,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

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

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

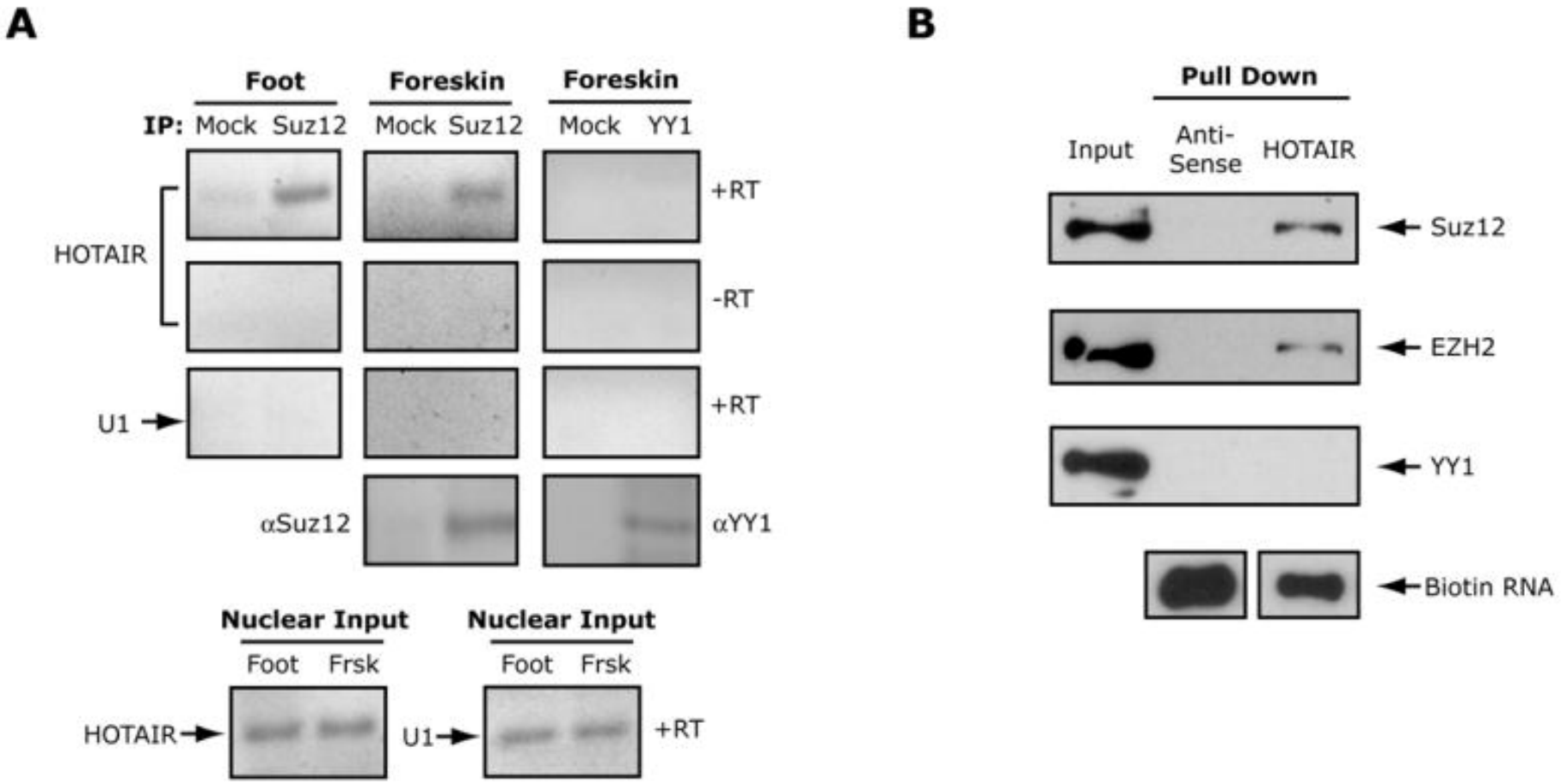
Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY.

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.



HOTAIR, an antisense intergenic long ncRNA of the HOXC locus

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



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

J Vis Exp. 2012 Mar 25;(61)

Chromatin Isolation by RNA Purification (ChIRP).

Chu C, Quinn J, Chang HY.

Source

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

Abstract

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

Chromatin Isolation by RNA Purification (ChIRP).

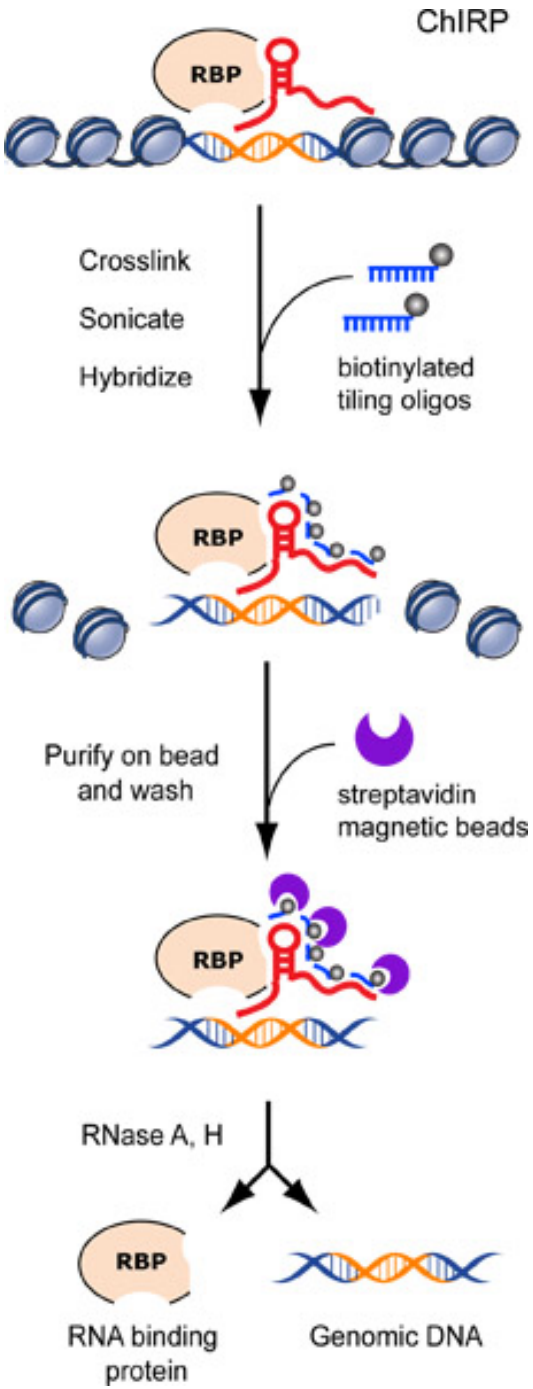
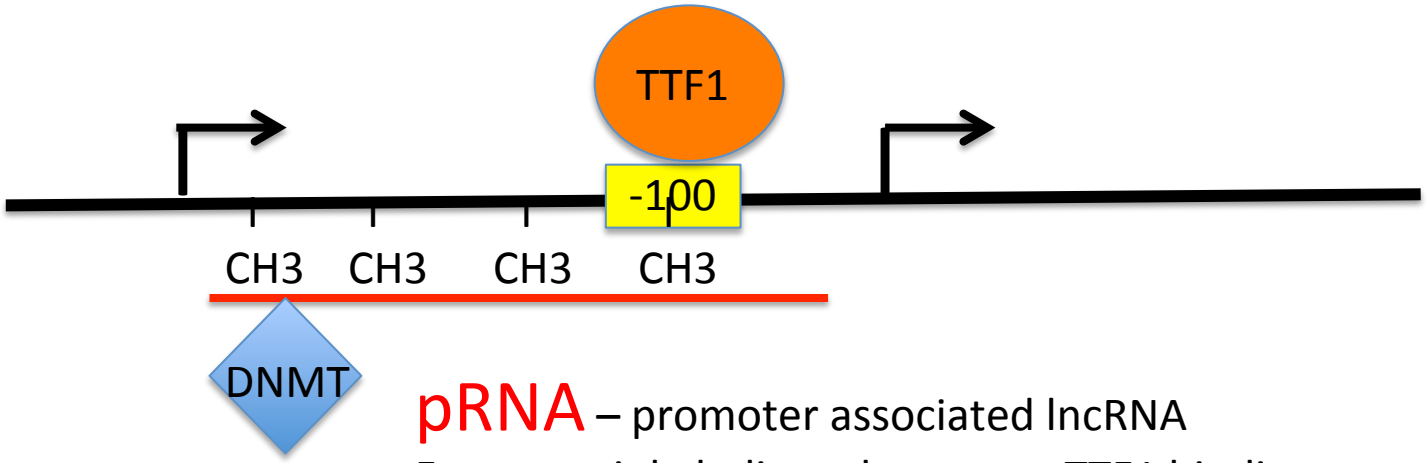


Figure 1. Flow chart of the ChIRP procedure. Chromatin is crosslinked to lncRNA:protein adducts *in vivo*. Biotinylated tiling probes are hybridized to target lncRNA, and chromatin complexes are purified using magnetic streptavidin beads, followed by stringent washes. We elute lncRNA bound DNA or proteins with a cocktail of Rnase A and H. A putative lncRNA binding sequence is schematized in orange. Previously published in Chu *et al*. 2011.17

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

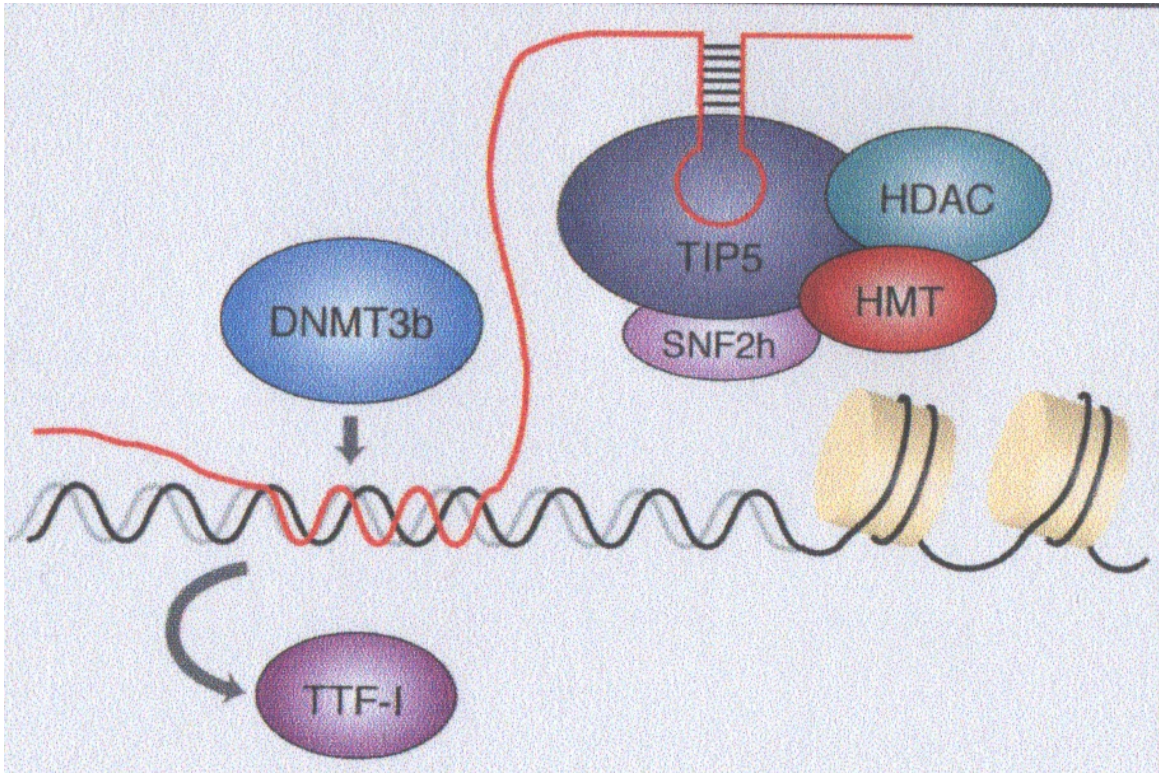
While recruiting DNMT3b

Ingrid Grummt – Heidelberg

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

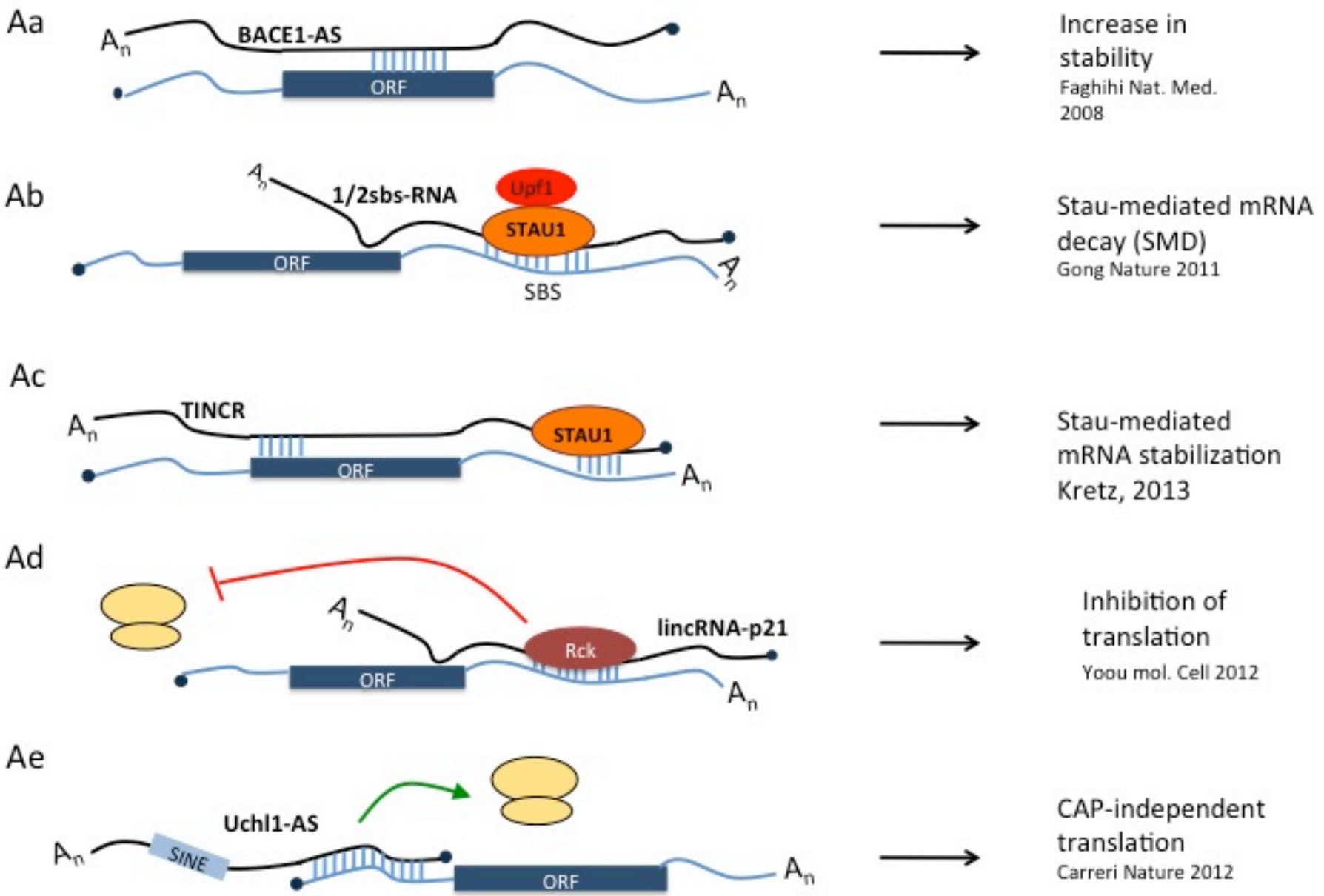
Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes.

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.



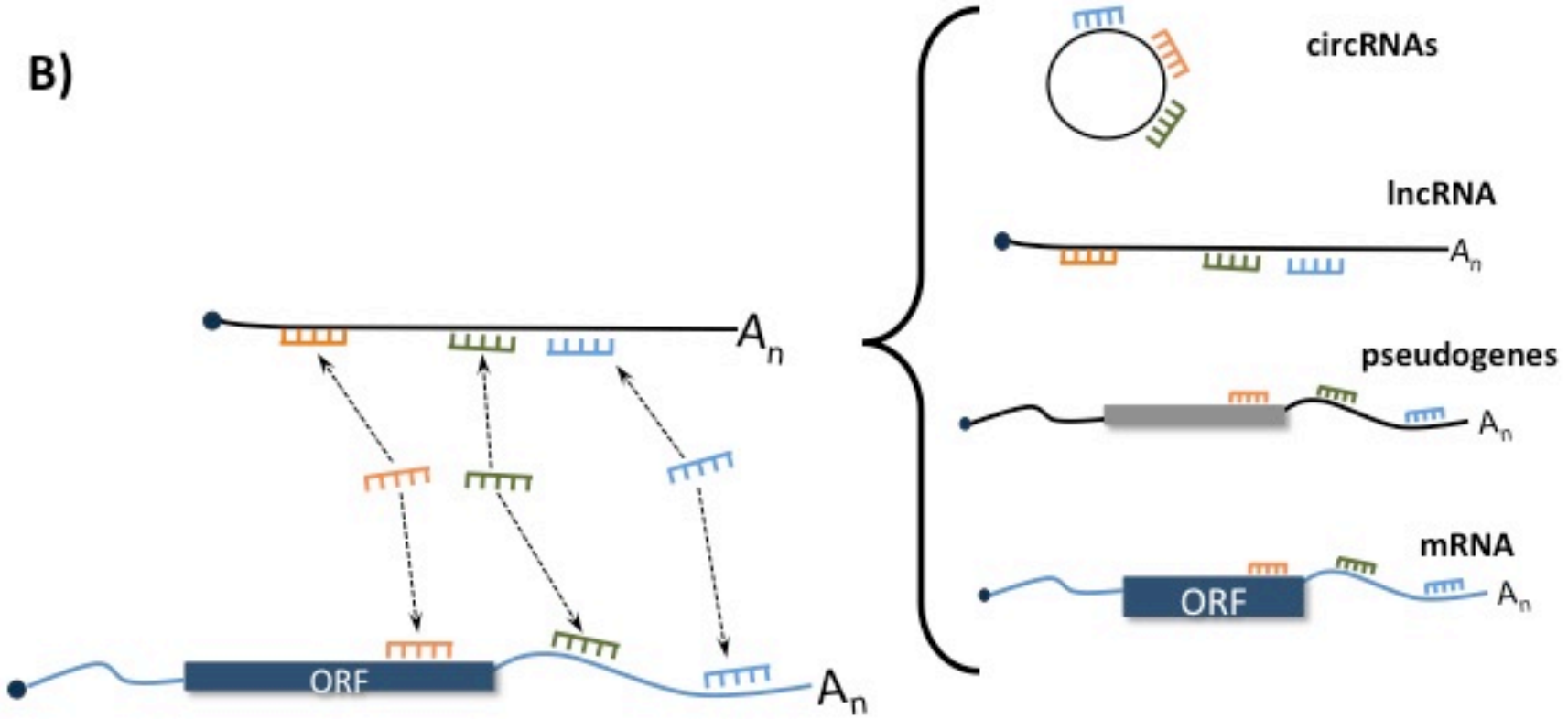
In addition, pRNA directly interacts with DNA, forming a DNA:DNA:RNA triple helix with the bind ing site of the transcription factor TTF-i, leading to displacement of TTF-i. The triple helical structure is recognized by the DNA methyltransferase DNMT3b, which methylates the rDNA promoter, leading to

Cytoplasmic lncRNA



Fatica & Bozzoni, Figure 2A

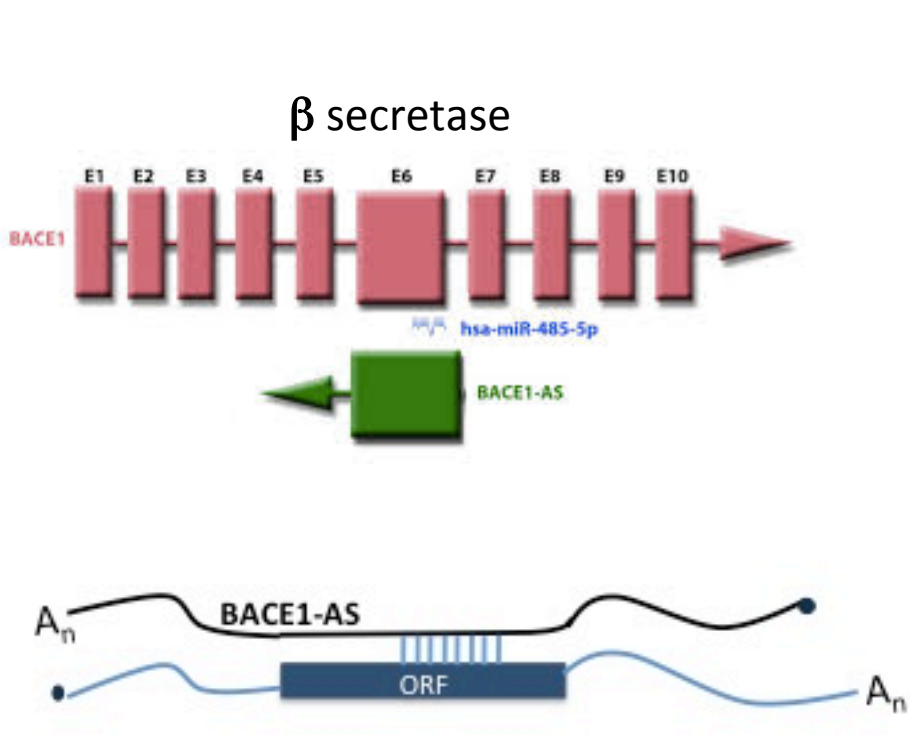
Competing endogenous RNAs for miRNA binding

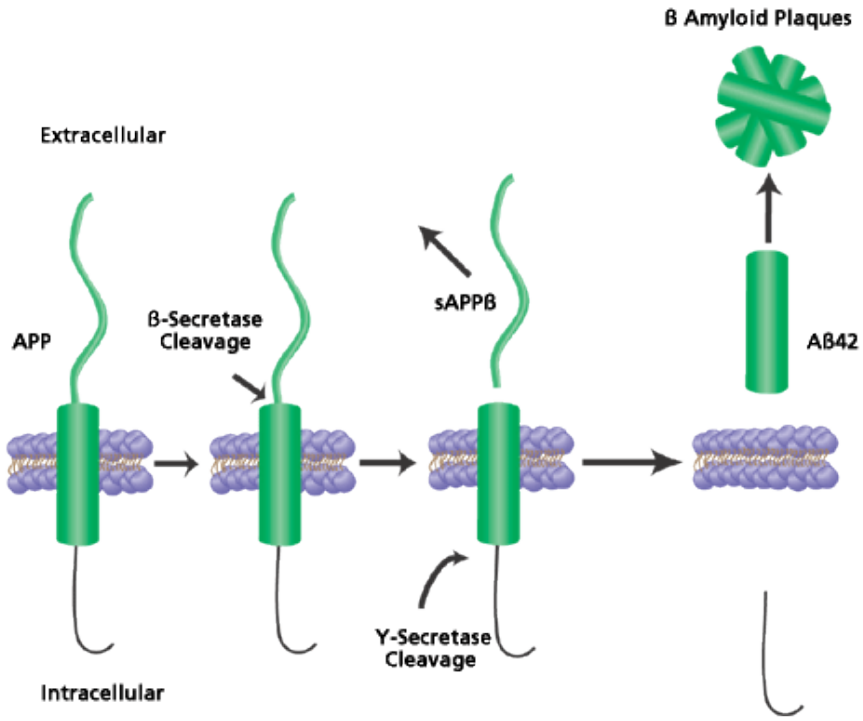


lncRNAs in brain disorders

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

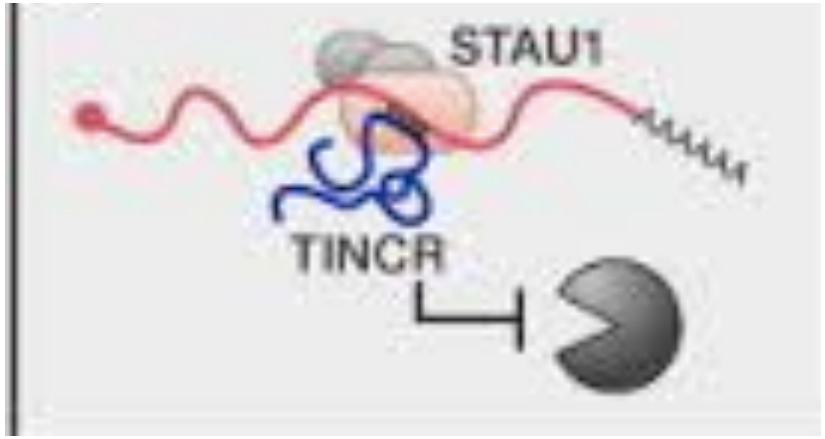
Nature. 2013 Jan 10;493(7431):231-5.

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

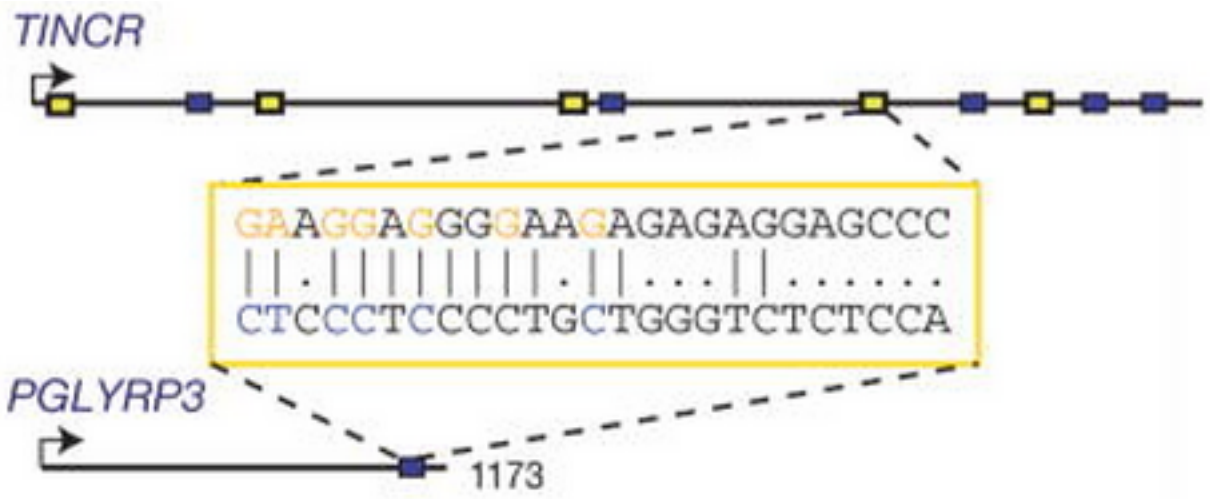
Kretz et al.

Abstract

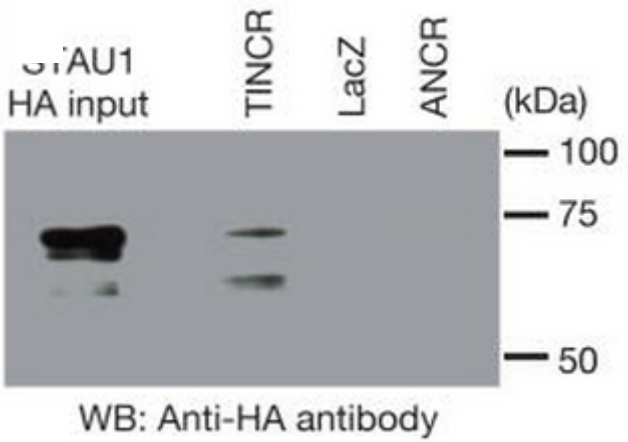
Several of the thousands of human long non-coding RNAs (lncRNAs) have been functionally characterized; however, potential roles for lncRNAs in somatic tissue differentiation remain poorly understood. Here we show that a **3.7-kilobase lncRNA**, terminal differentiation-induced ncRNA (TINCR), controls 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

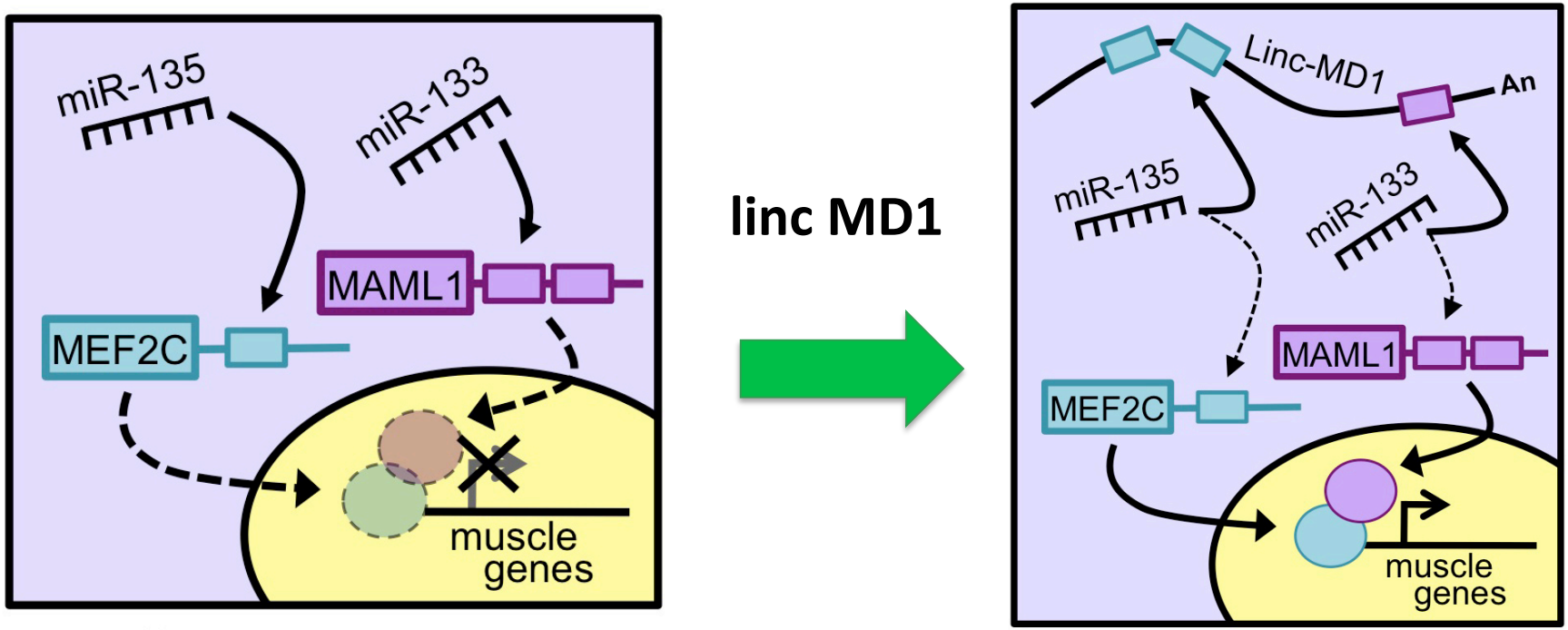
Marcella Cesana,[1,6] Davide Cacchiarelli,[1,6] Ivano Legnini,[1] Tiziana Santini,[1] Olga Sthandier,[1] Mauro Chinappi,[2] Anna Tramontano,[2,3,4] and Irene Bozzoni[1,3,4,5,*]

[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.

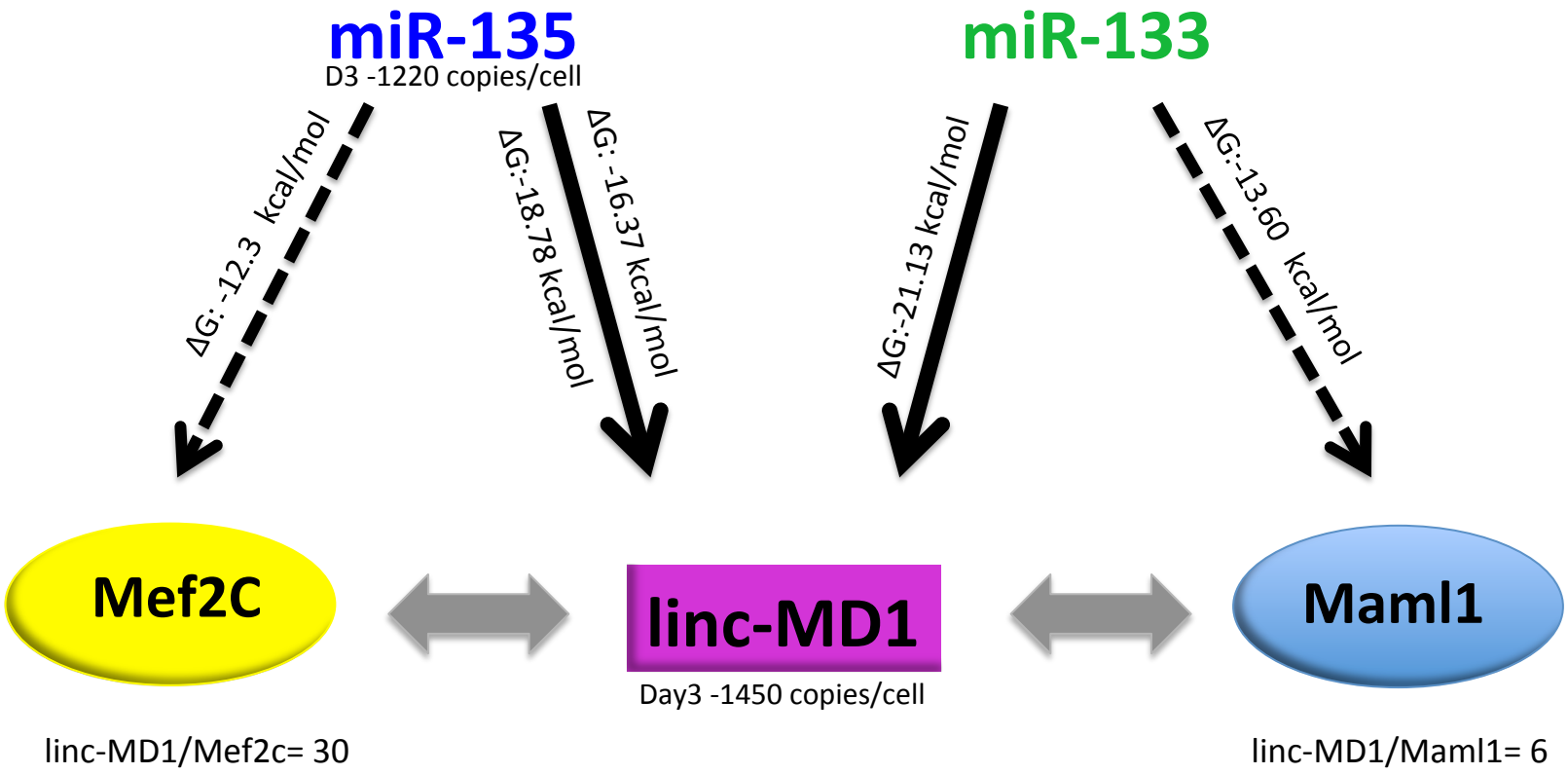
linc-MD1 acts as a sponge for specific miRNAs



differentiation

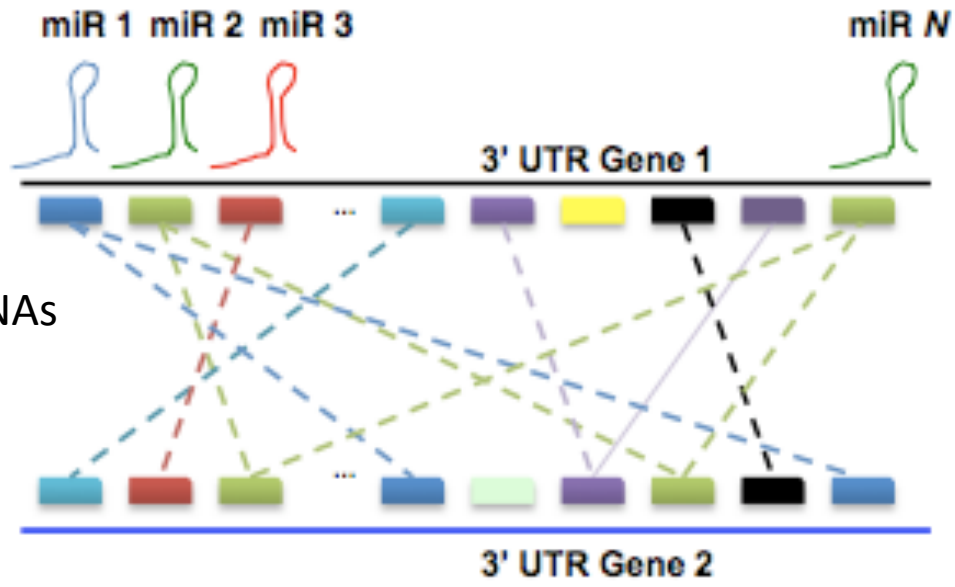
Cesana et al., *Cell* 147, *358-369*, 2011

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

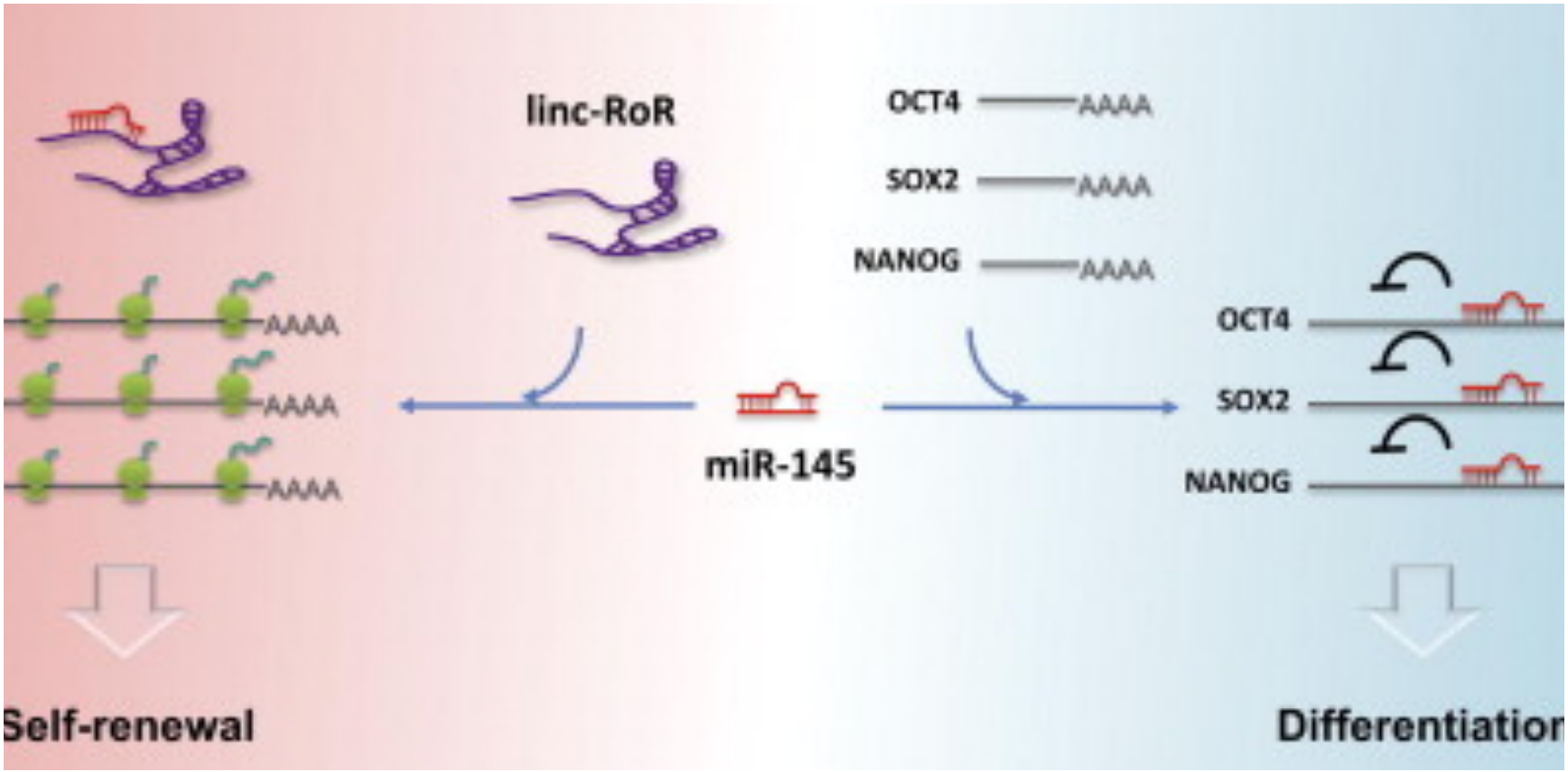
Tay et al., *Cell* 147, October 14, 2011

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.

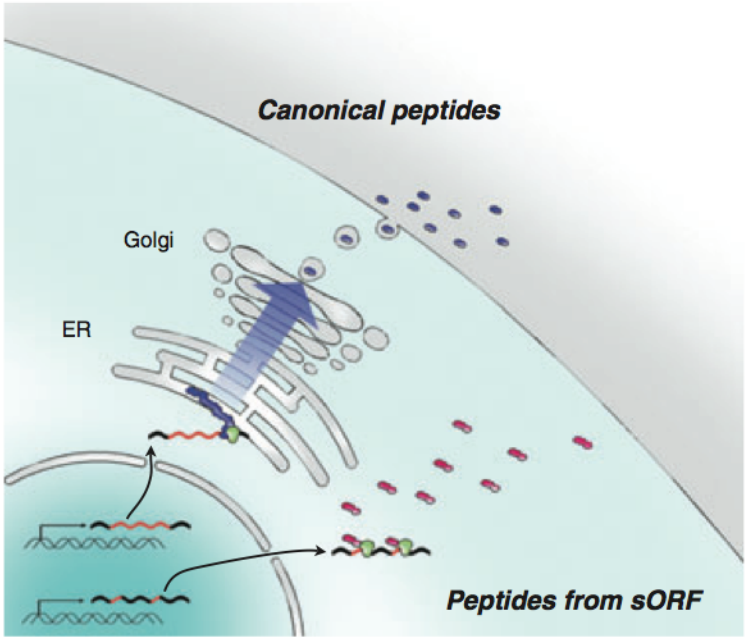


A Competition for miR-145 between linc-RoR and mRNAs Encoding the Core TFs. The presence of linc-RoR in hESCs traps miR-145, preventing it from repressing the translation of the core pluripotency factors and ensuring the stem cell fate. The disappearance of linc-RoR in differentiating hESCs releases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

Functions of lncRNAs

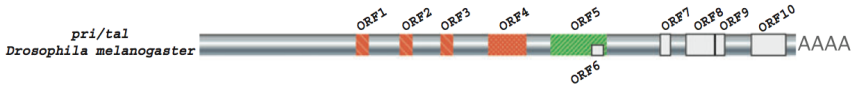
Cytoplasmic lncRNAs

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.



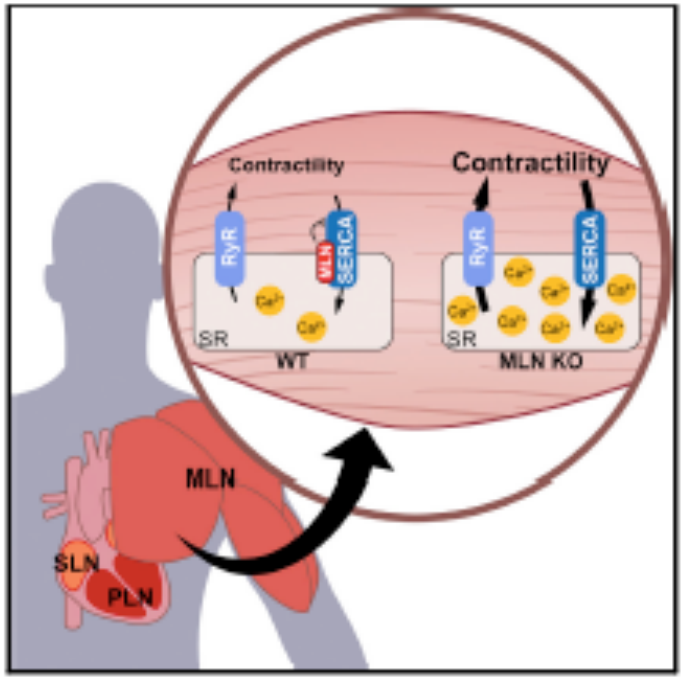
lncRNAs can encode for short peptides

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

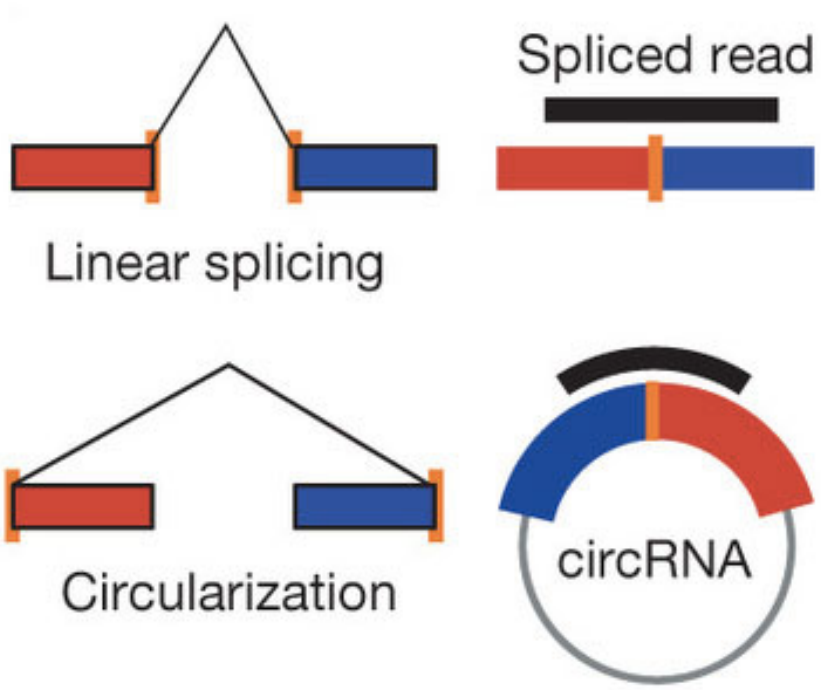
Correspondence

eric.olson@utsouthwestern.edu

In Brief

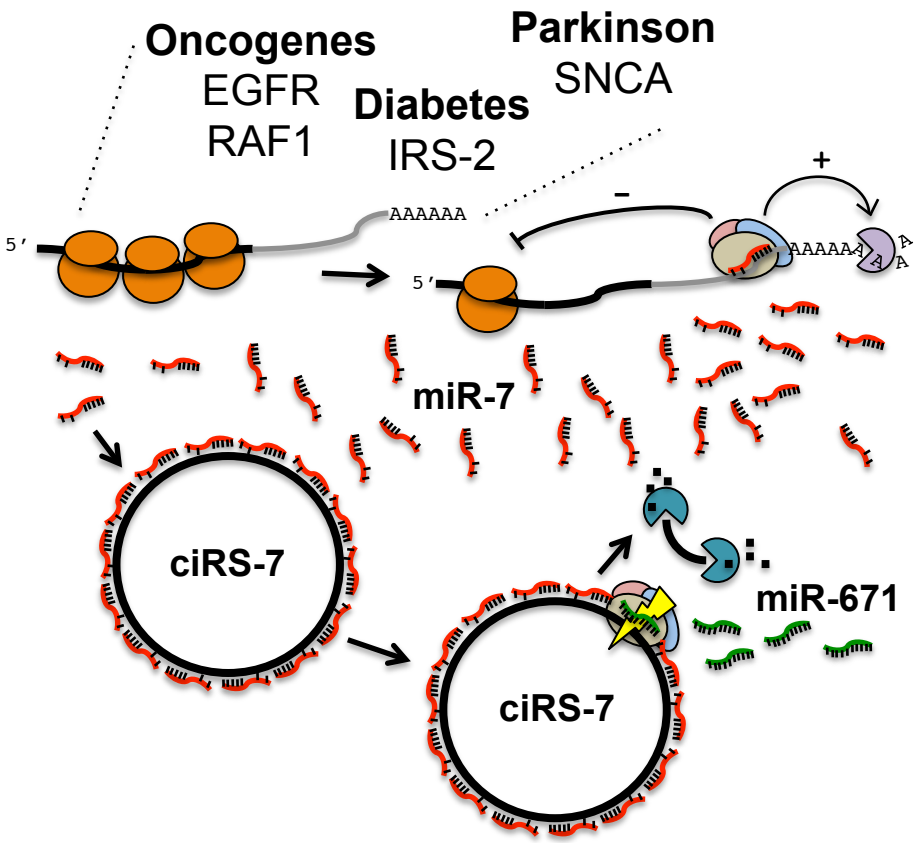
Myoregulin is a skeletal muscle-specific micropeptide that regulates muscle performance by modulating intracellular calcium handling.

…..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.

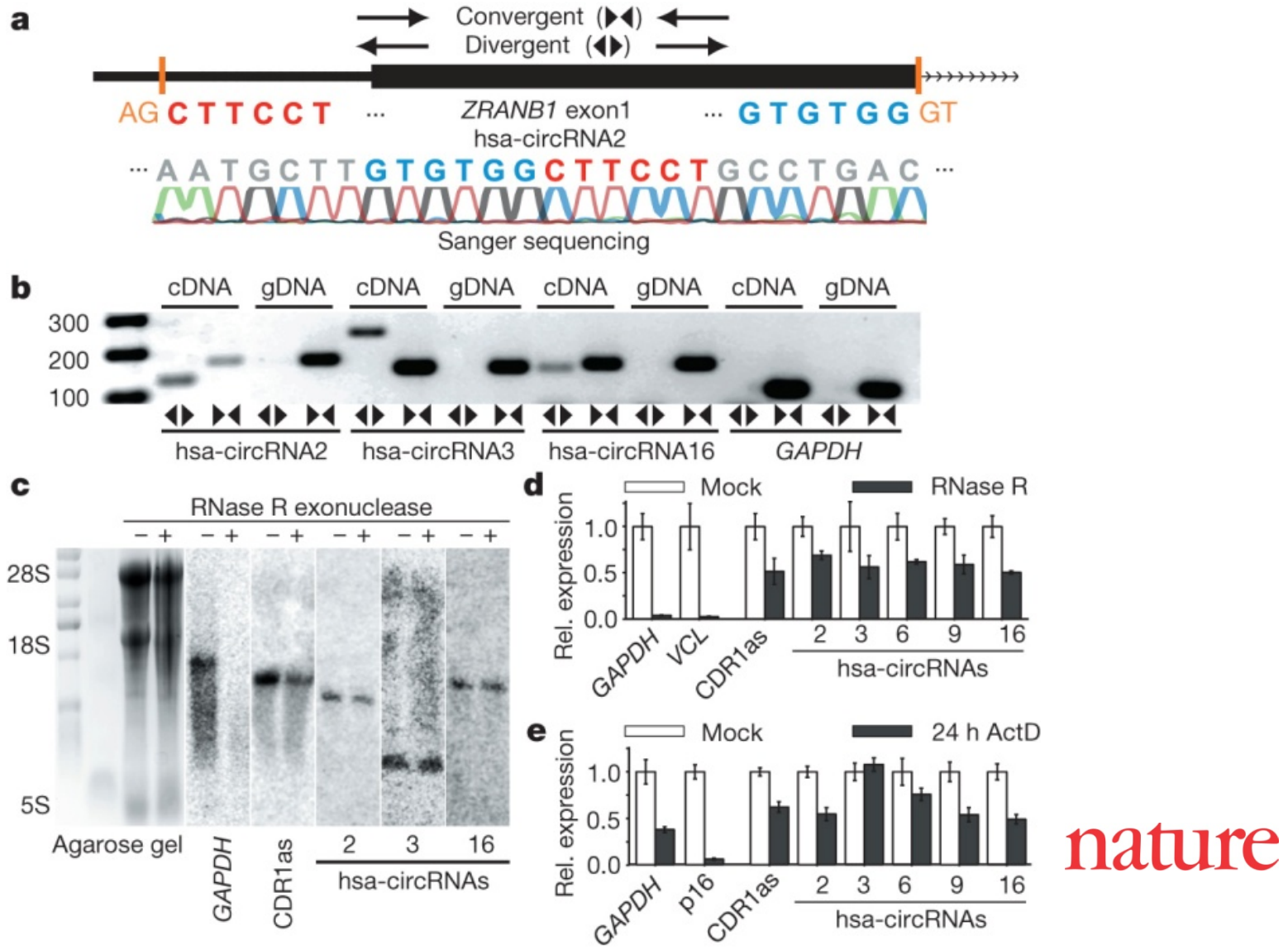
Circular RNA (circRNA) Sponge



Nature 495, 333–338 (21 March 2013)

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

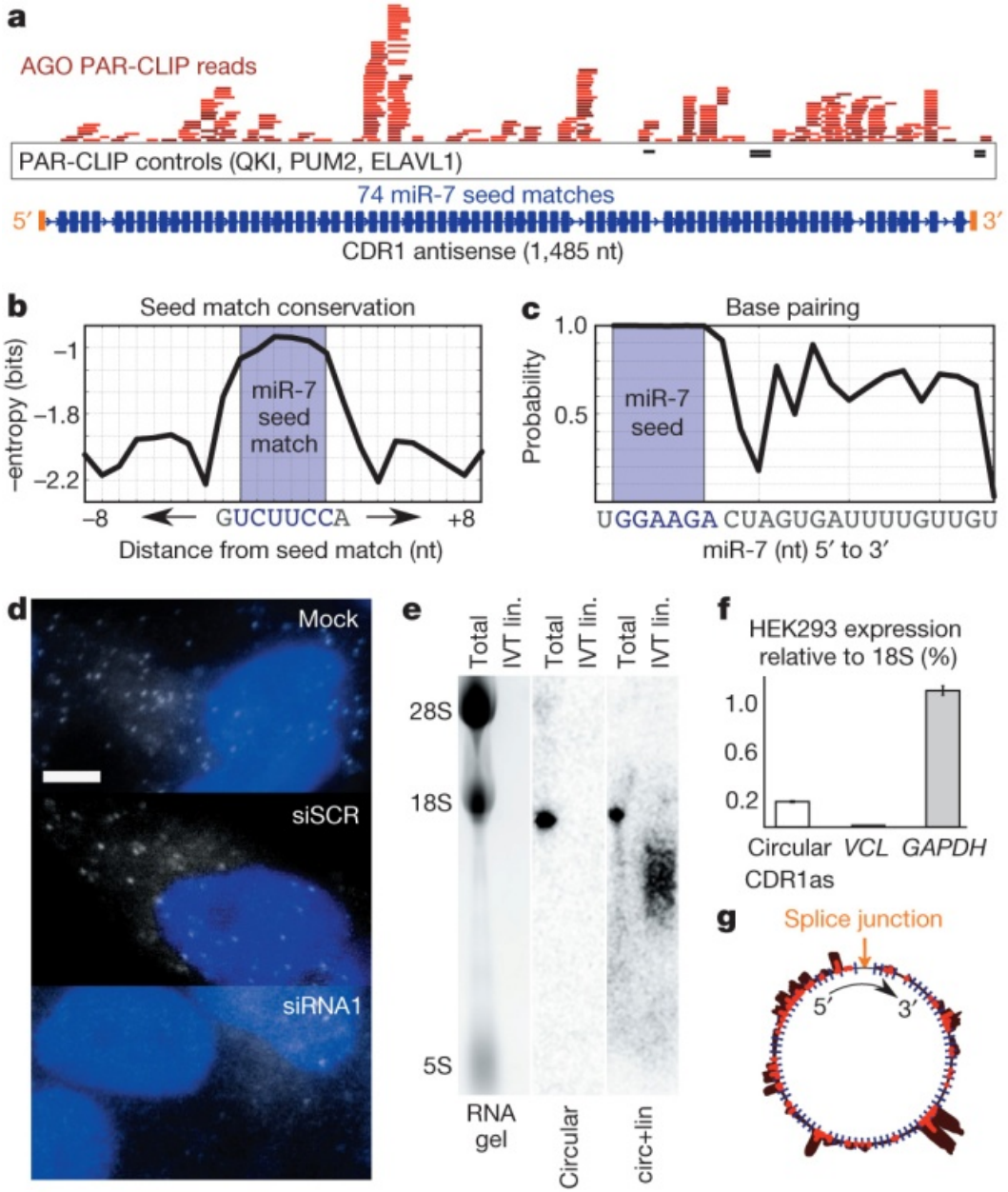
Sebastian Memczak ……. & Nikolaus Rajewsky



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

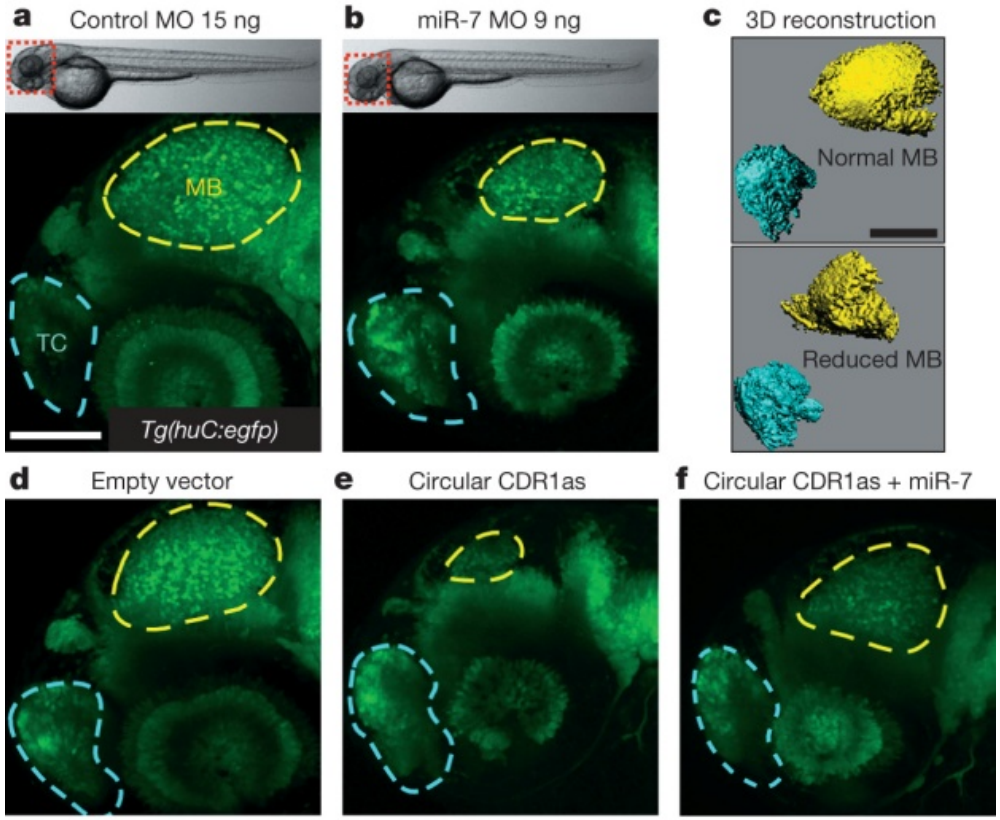
The circRNA CDR1as is bound by the miRNA effector protein AGO, and is cytoplasmic.



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.