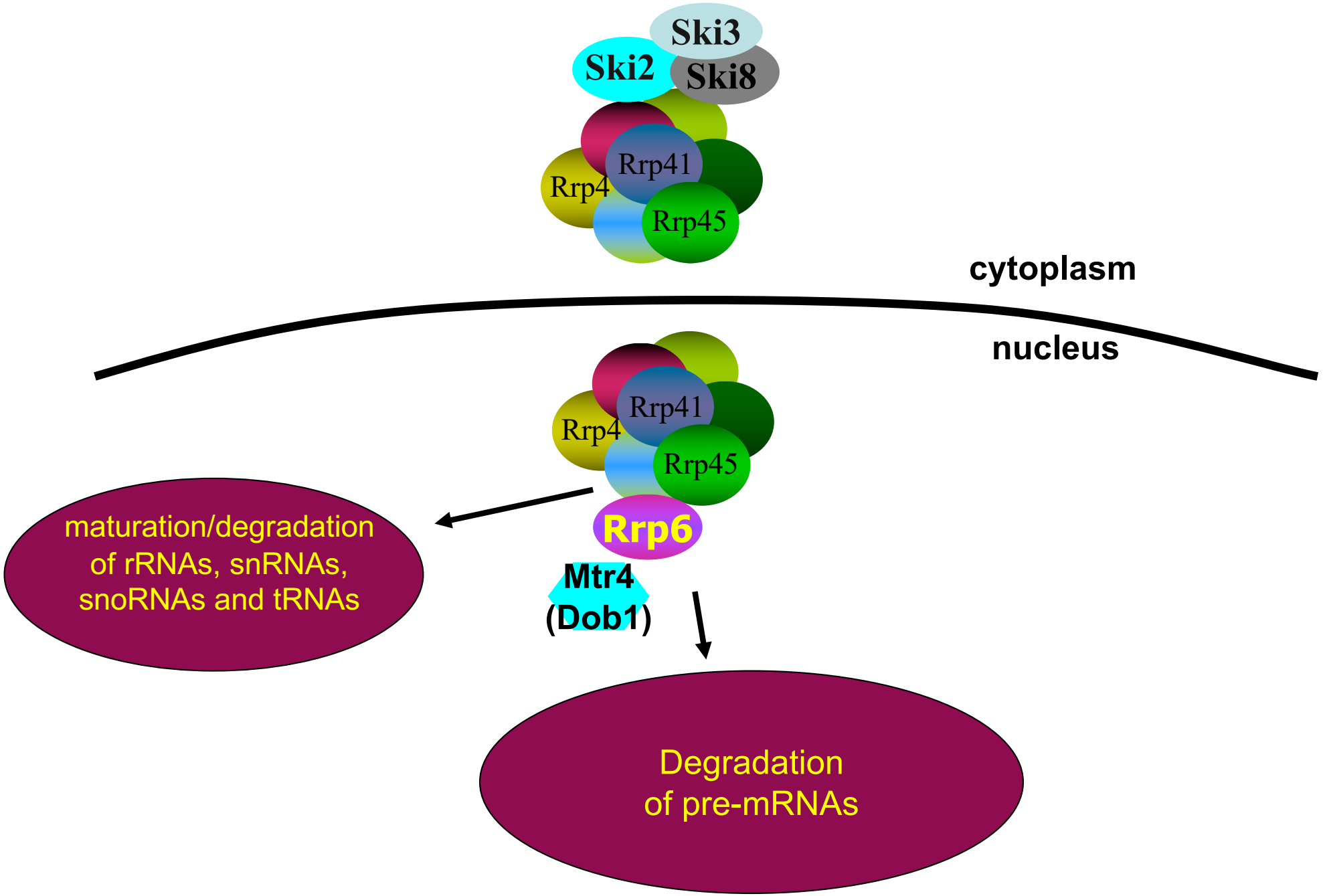
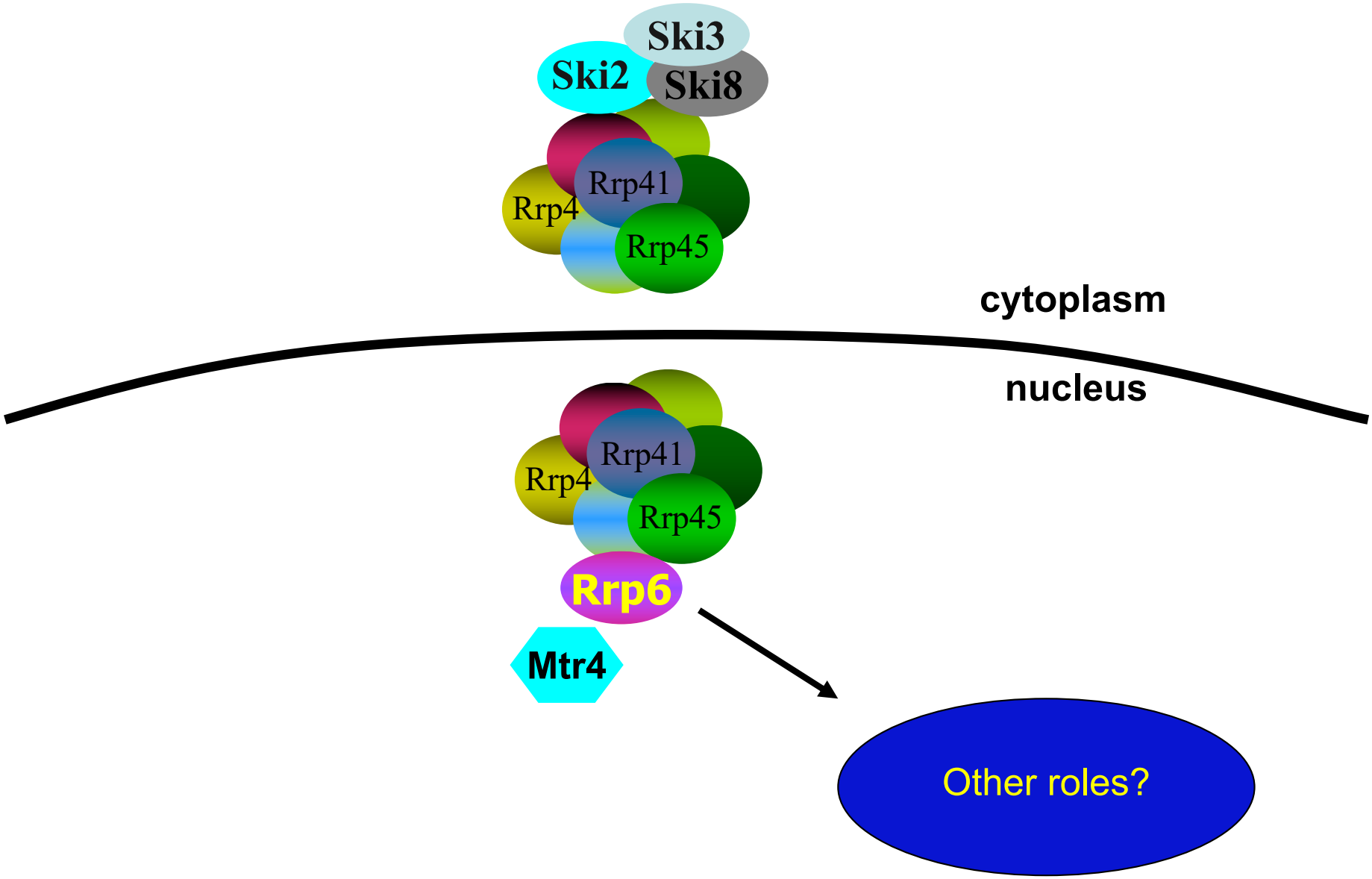
Criptic **U**nstable **T**ranscript**s**

(CUTs)

Looking for substrates of the nuclear exosome



Looking for substrates of the nuclear exosome

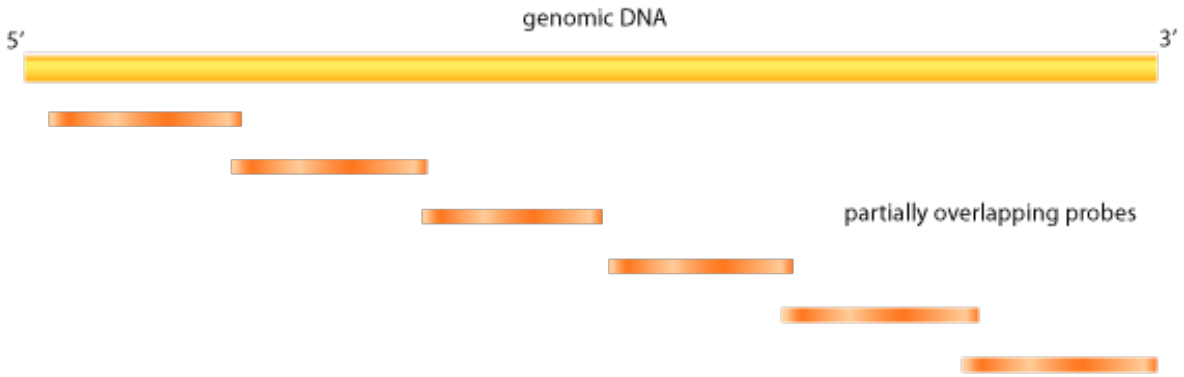


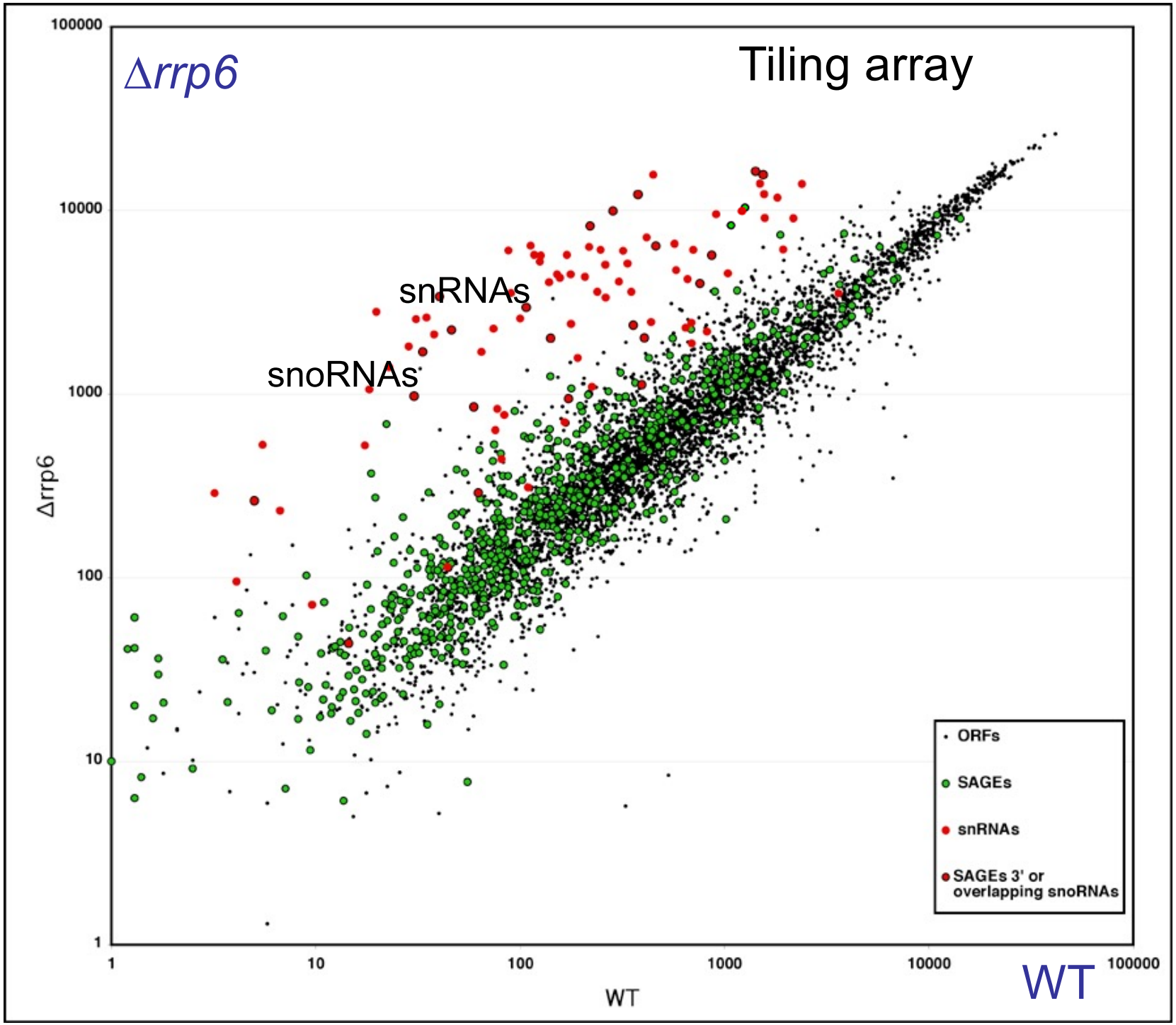
Looking for substrates of the nuclear exosome

WT and Δrrp6 → Total RNA → cDNA (dT oligo) → Tiling array → DE (WT vs Δrrp6)

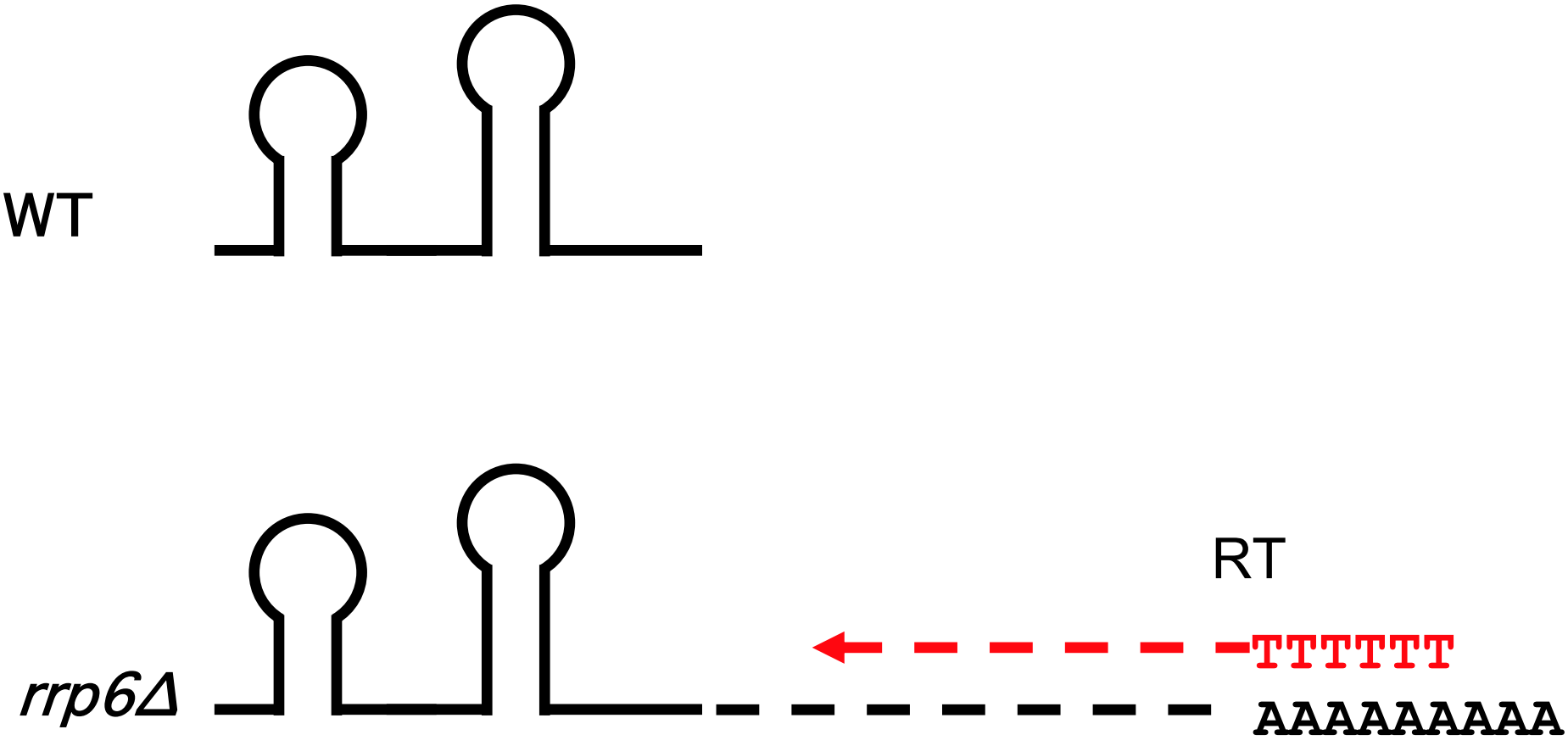
Tiling array:

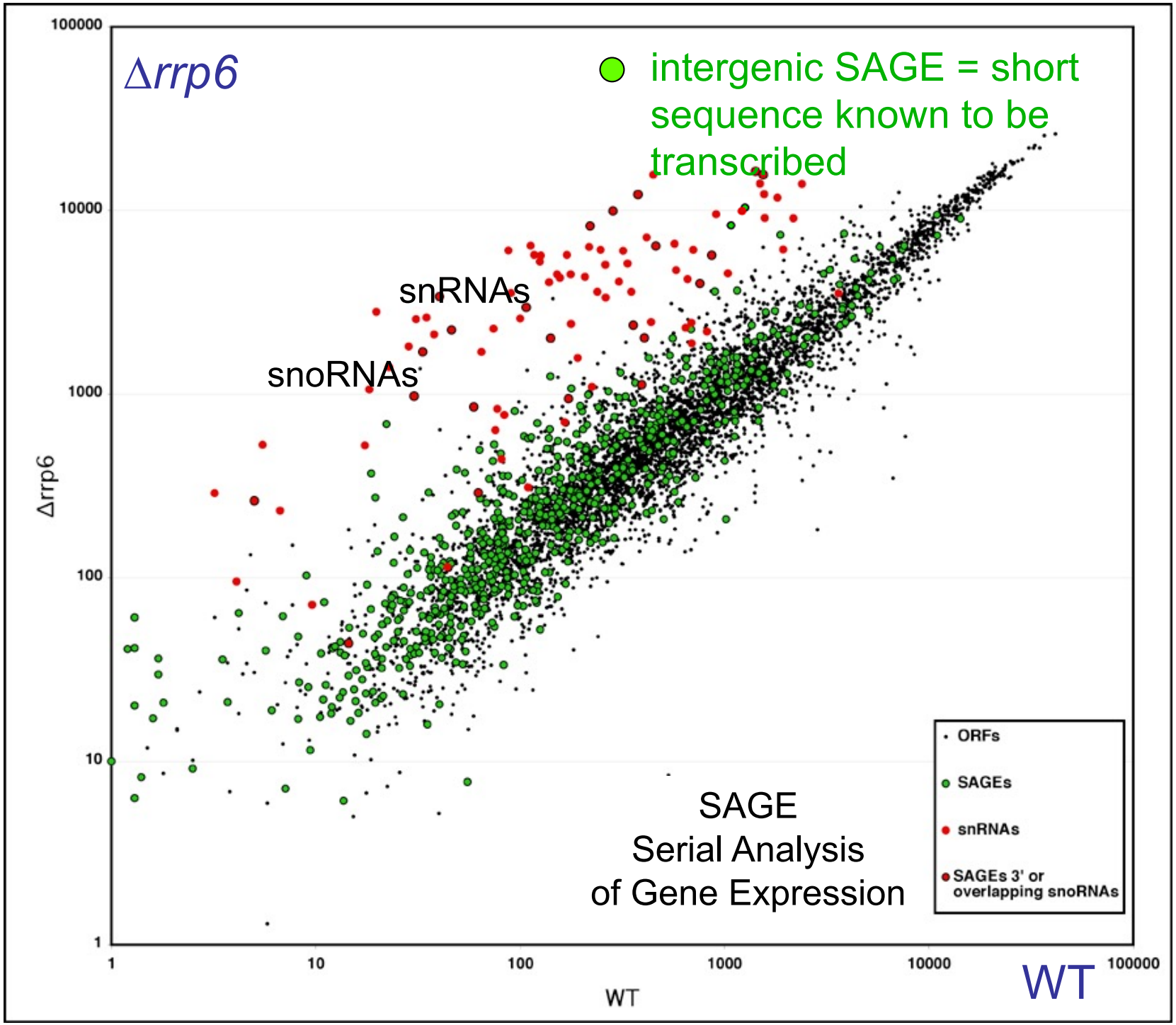
Instead of probing for sequences of known or predicted genes that may be dispersed throughout the genome, **tiling arrays** probe intensively for sequences which are known to exist in a contiguous region.



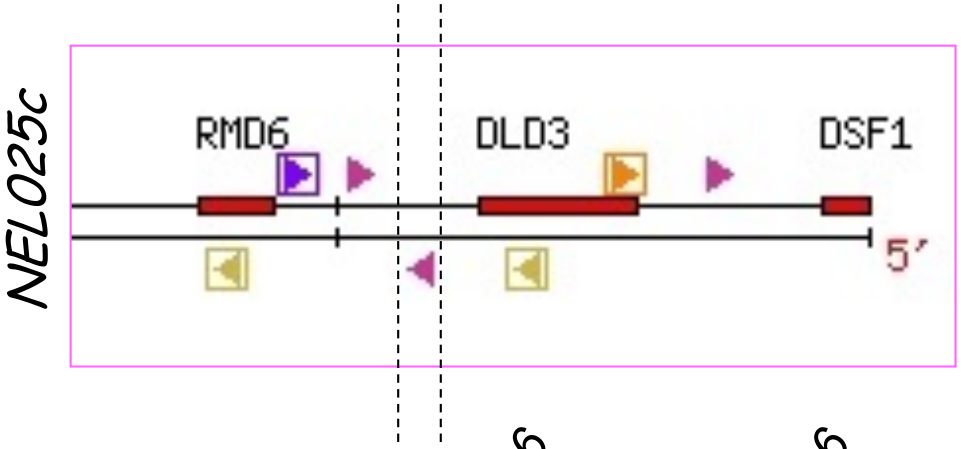

Δrrp6
Tiling array
snRNAs
snoRNAs
ORFs
SAGEs
snRNAs
SAGEs 3' or overlapping snoRNAs
Δrrp6
WT
WT
100000
10000
1000
100
10
1
1
10
100
1000
10000
100000

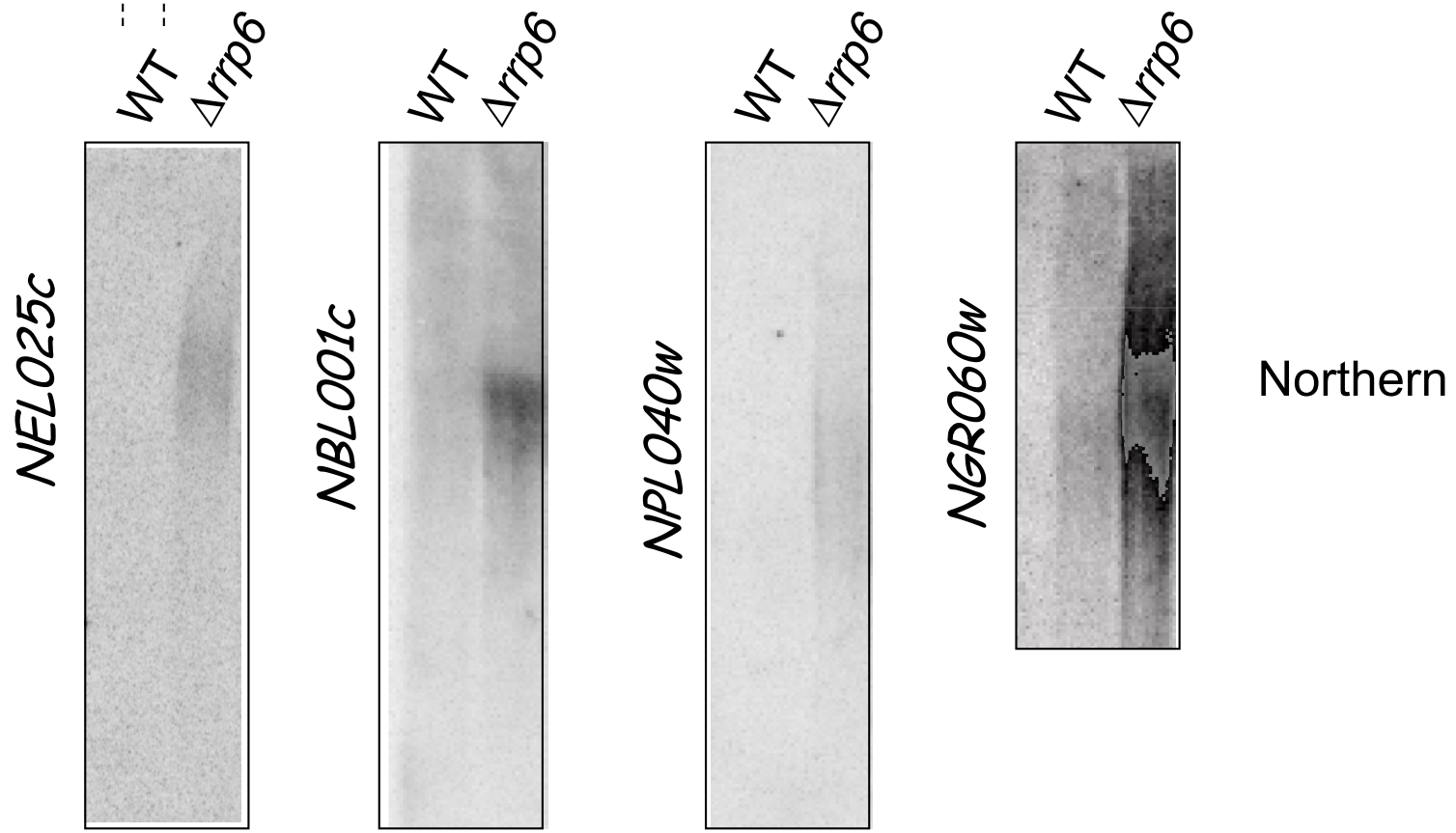
snRNAs, rRNA and snoRNAs get poly(A) tails and become "visible" for the DNA chip




Δrrp6
intergenic SAGE = short sequence known to be transcribed
snRNAs
snoRNAs
SAGE
Serial Analysis
of Gene Expression
ORFs
SAGEs
snRNAs
SAGEs 3' or overlapping snoRNAs
Δrrp6
WT
WT
100000
10000
1000
100
10
1
1
10
100
1000
10000
100000

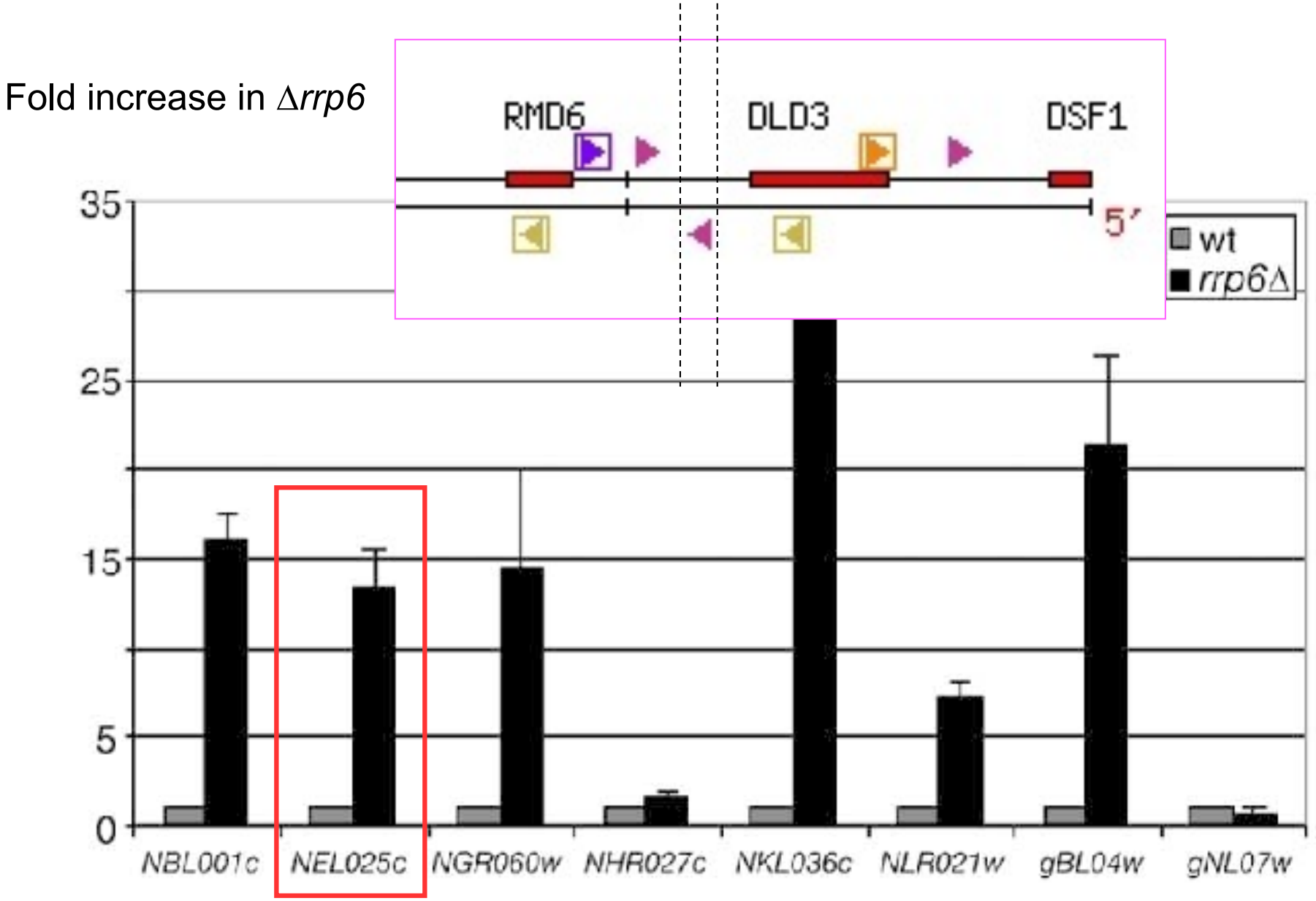
Intergenic regions are expressed in Δ*rrp6*



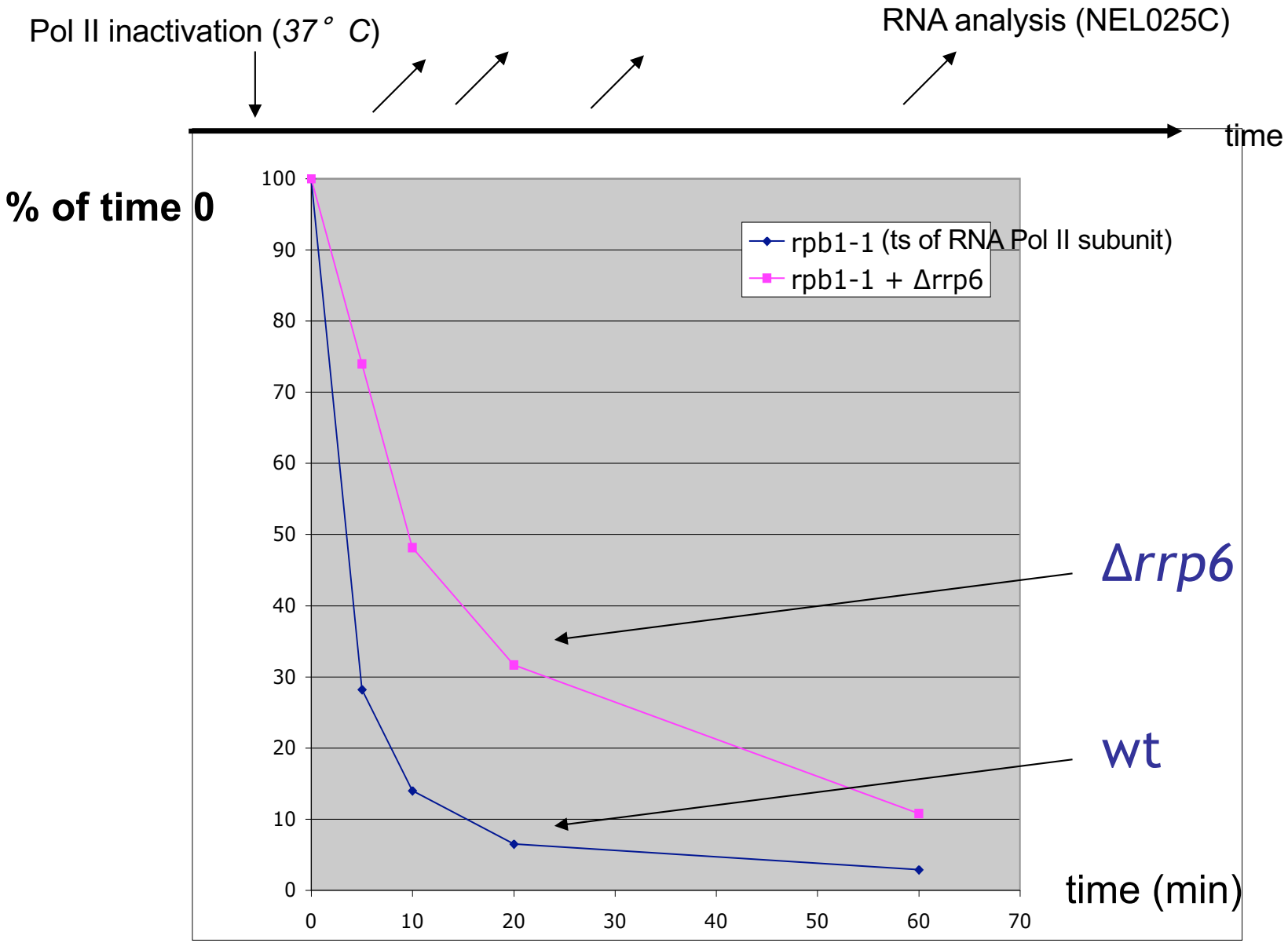


Northern

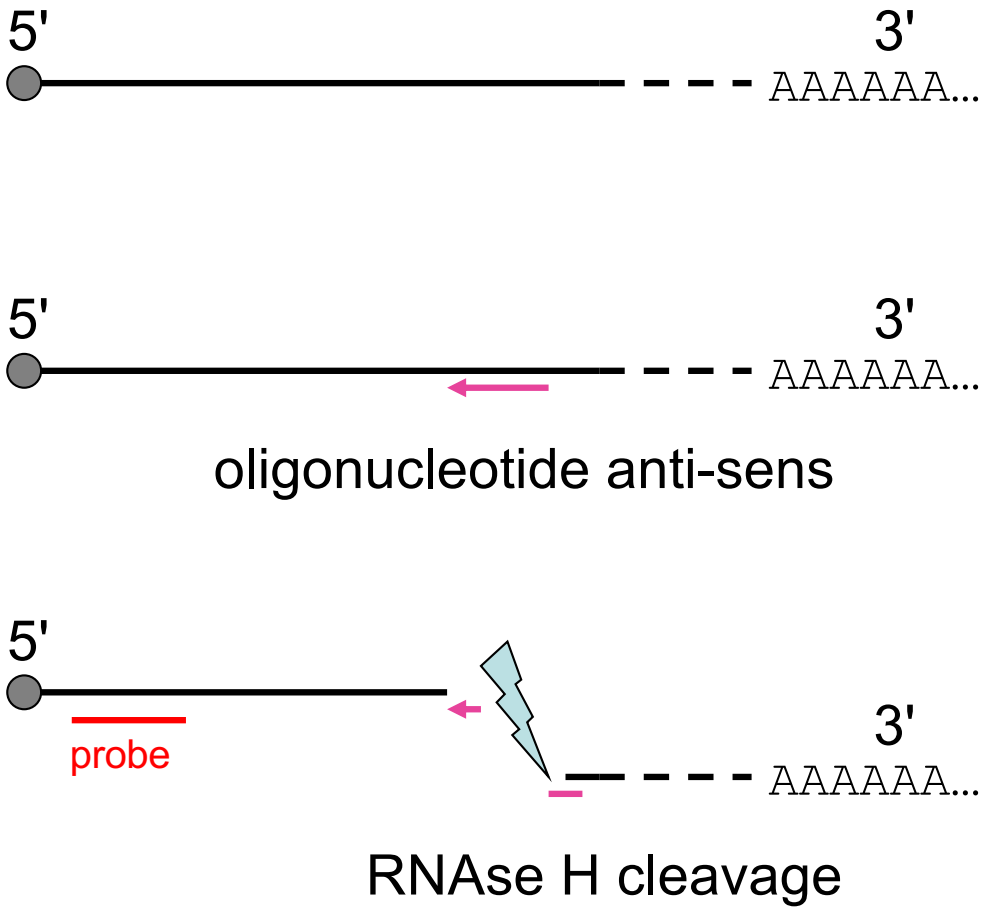
Quantitative real time PCR confirms transcription from intergenic regions

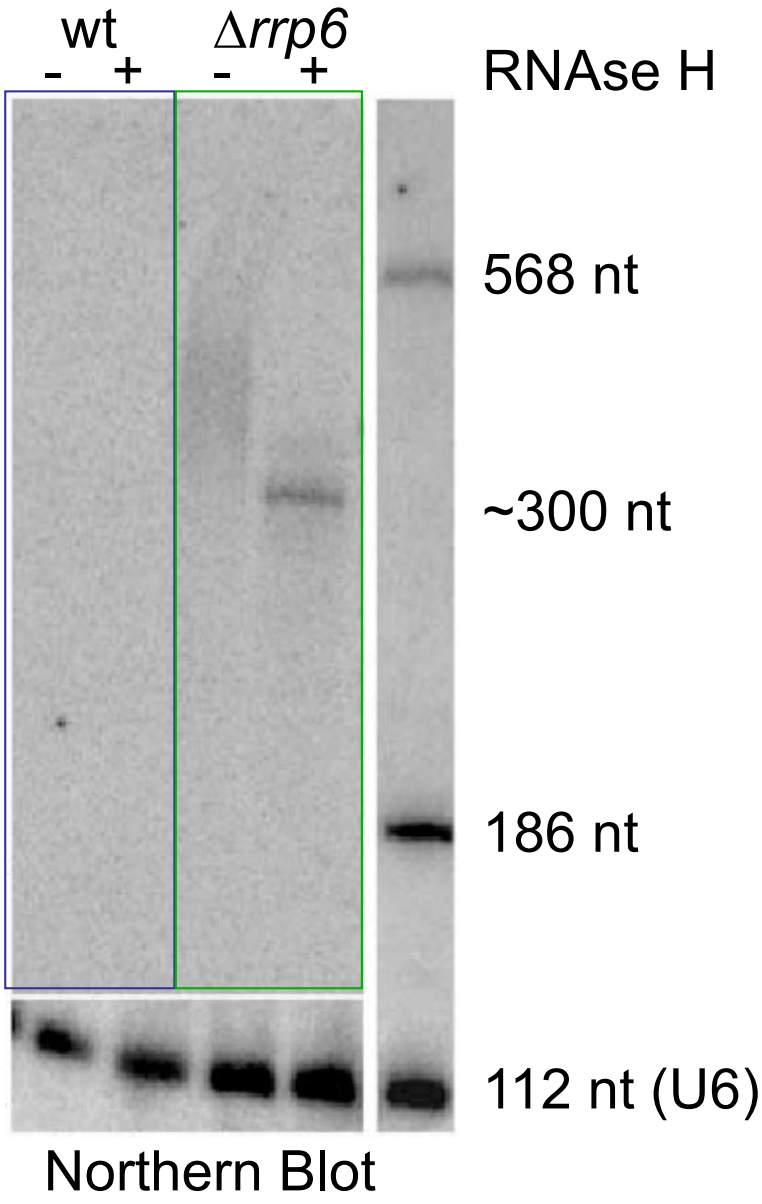


NEL025 is transcribed by RNA polymerase II and its turnover is slower in Δ*rrp6*

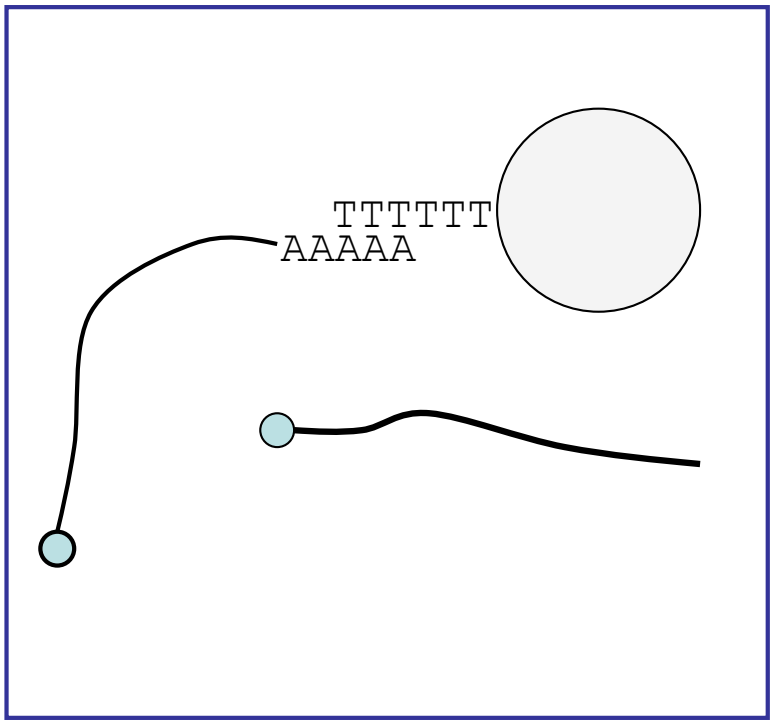


NEL025c transcripts have a defined capped 5′-end

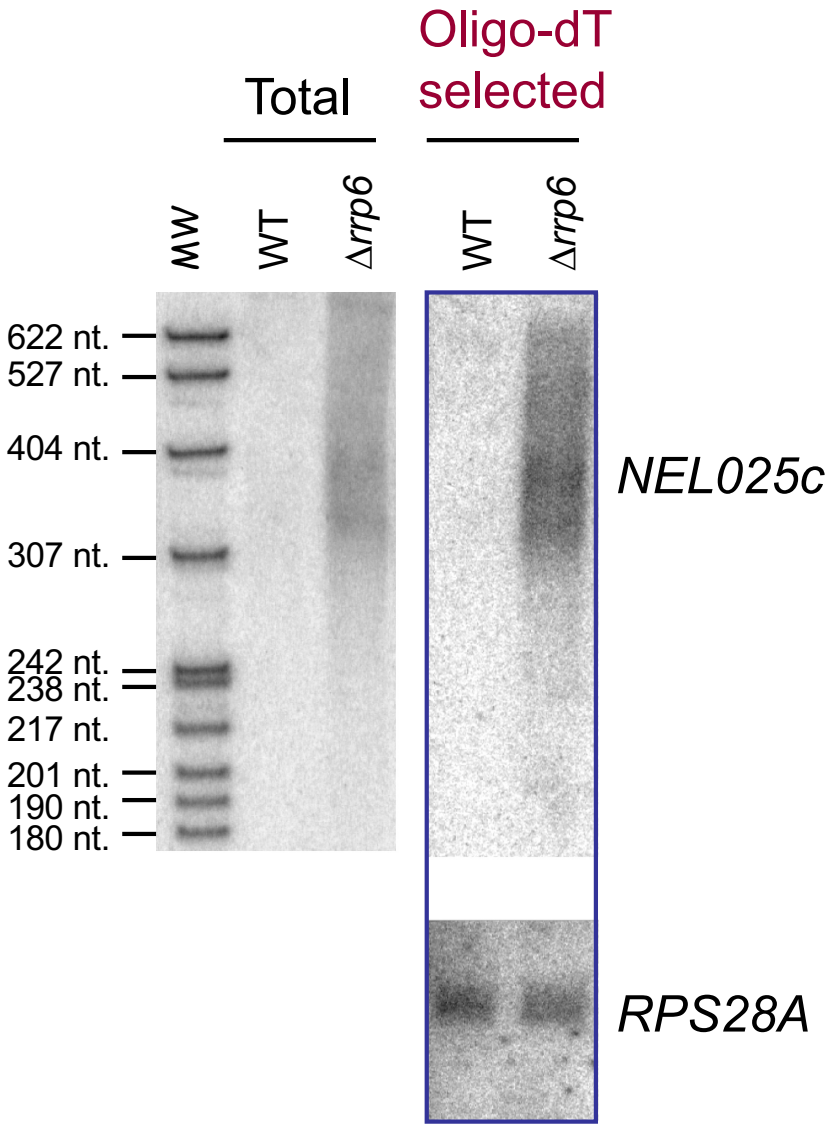




NEL025c transcripts are adenylated in Δ*rrp6*

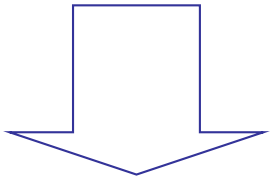


Oligo-dT chromatography to select poly(A) RNAs



Intergenic transcripts are RNA polymerase II products

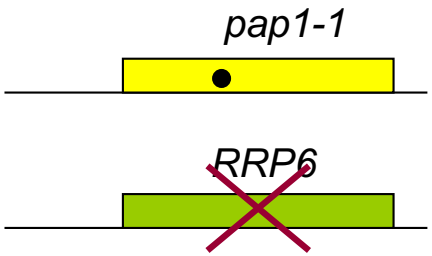
- Without defined promoters
- Capped and polyadenylated
- Unstable in a wild type strain (but detectable)
- Degraded by the nuclear exosome



CUTs: Criptic Unstable Transcripts

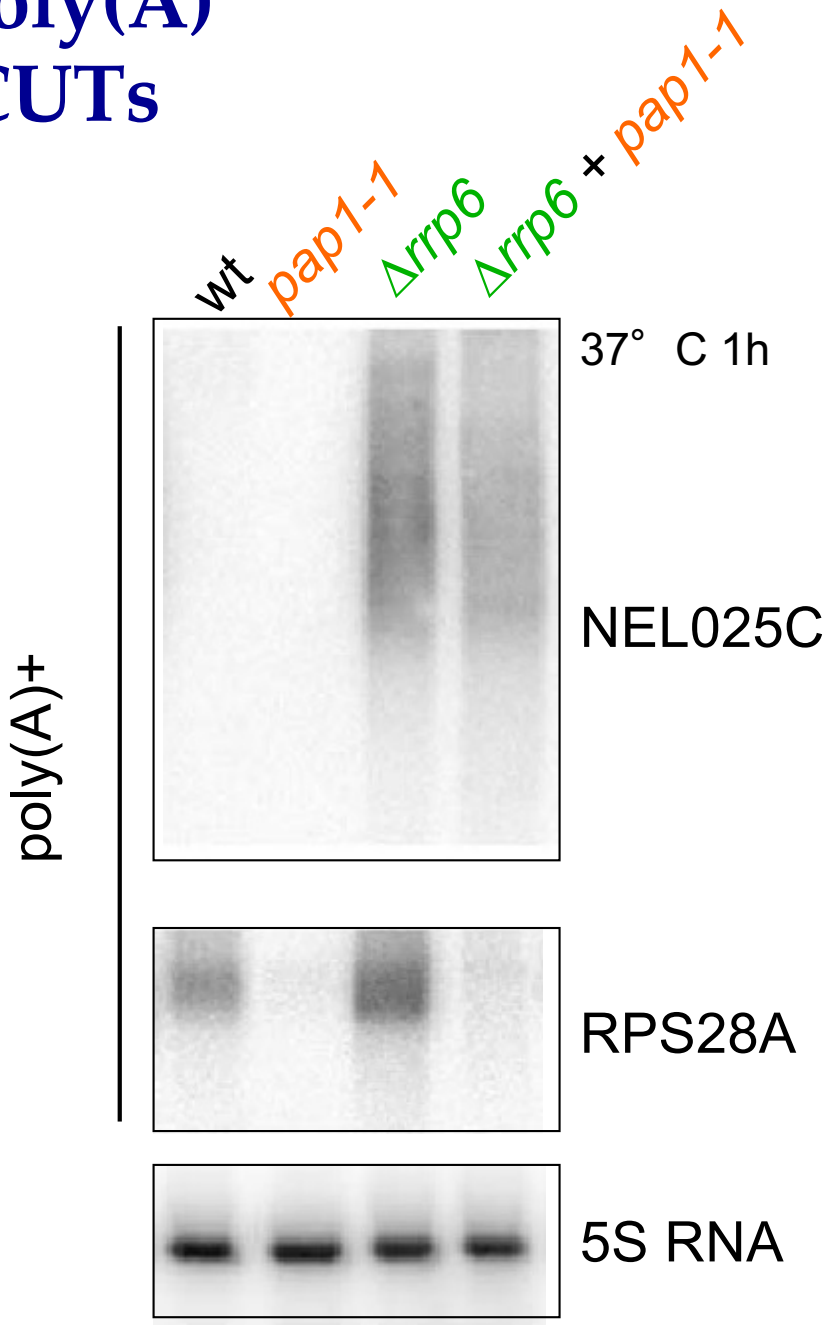
Who's responsible for CUTs polyadenylation?
What targets a CUT to degradation?

Pap1 is the major nuclear poly(A) polymerase - but not for CUTs



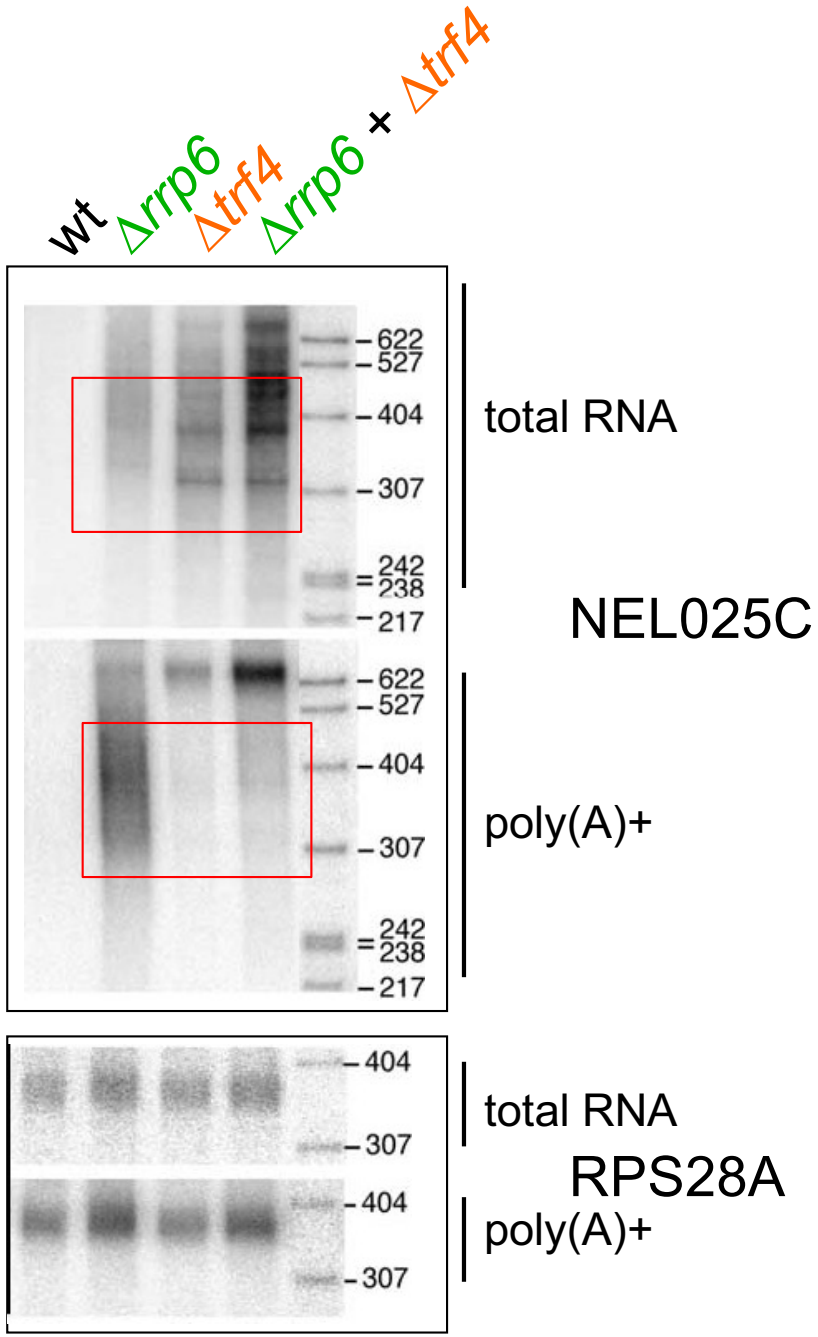
Oligo-dT chromatography to select poly(A) RNAs

Northern



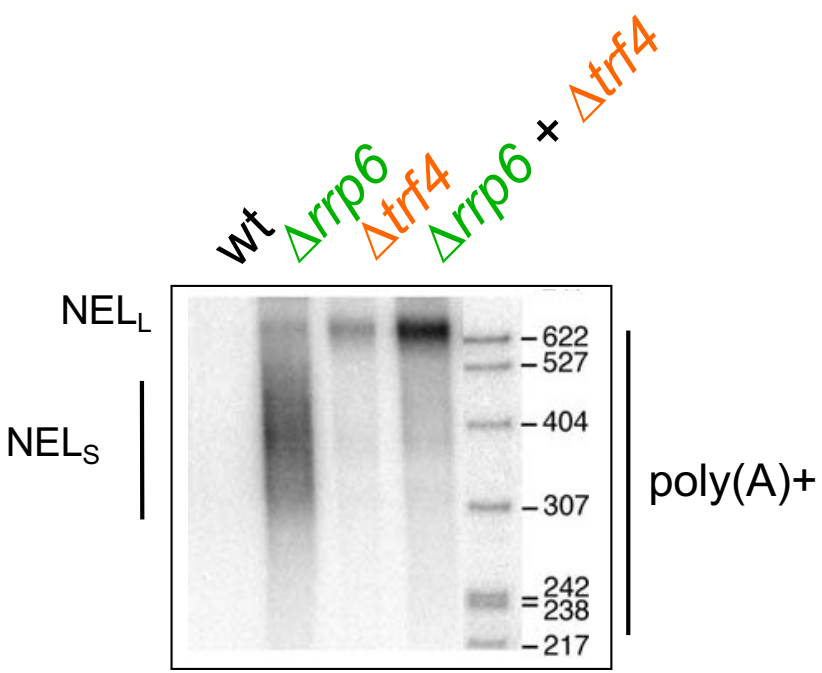
Trf4 is essential for CUTs polyadenylation and degradation

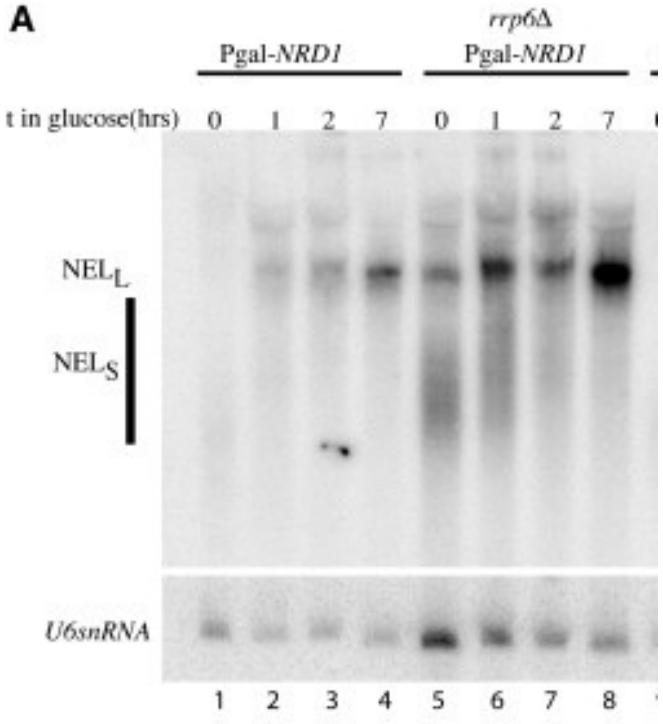
The absence of Trf4 leads to accumulation of unpolyadenylated NEL025C transcripts



Termination of CUTs is Nrd1 dependent

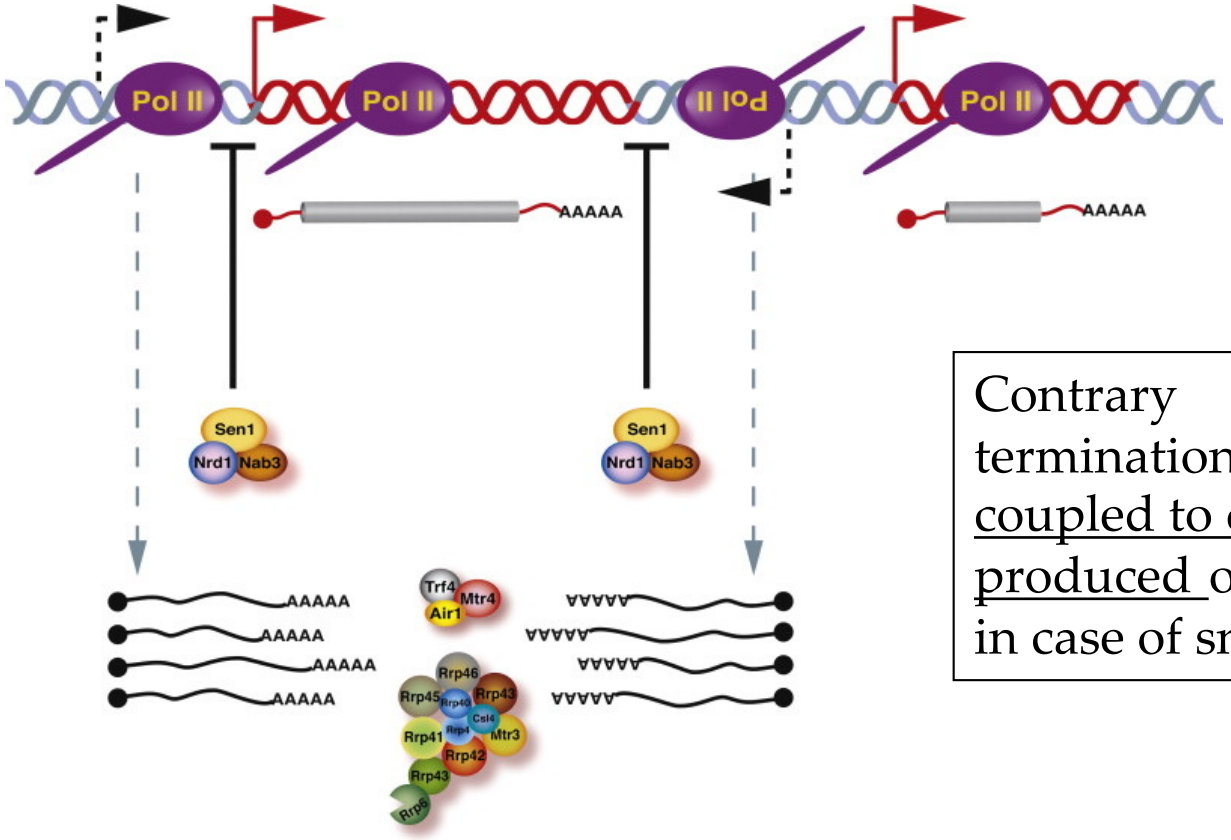
NEL is a CUT expressed under two forms: (1) a series of short transcripts (hereafter called **NEL_S**) with heterogeneous 3′ ends and (2) a longer transcript of homogeneous size (**NEL_L**). Both forms can only be observed in mutants of the exosome/TRAMP, with the short forms representing the vast majority of NEL025c transcripts. In a trf4Δ strain, only the NEL_L form is polyadenylated. NEL_L is polyadenylated by Pap1p and is processed by the canonical 3′ end mRNA termination/polyadenylation pathway.





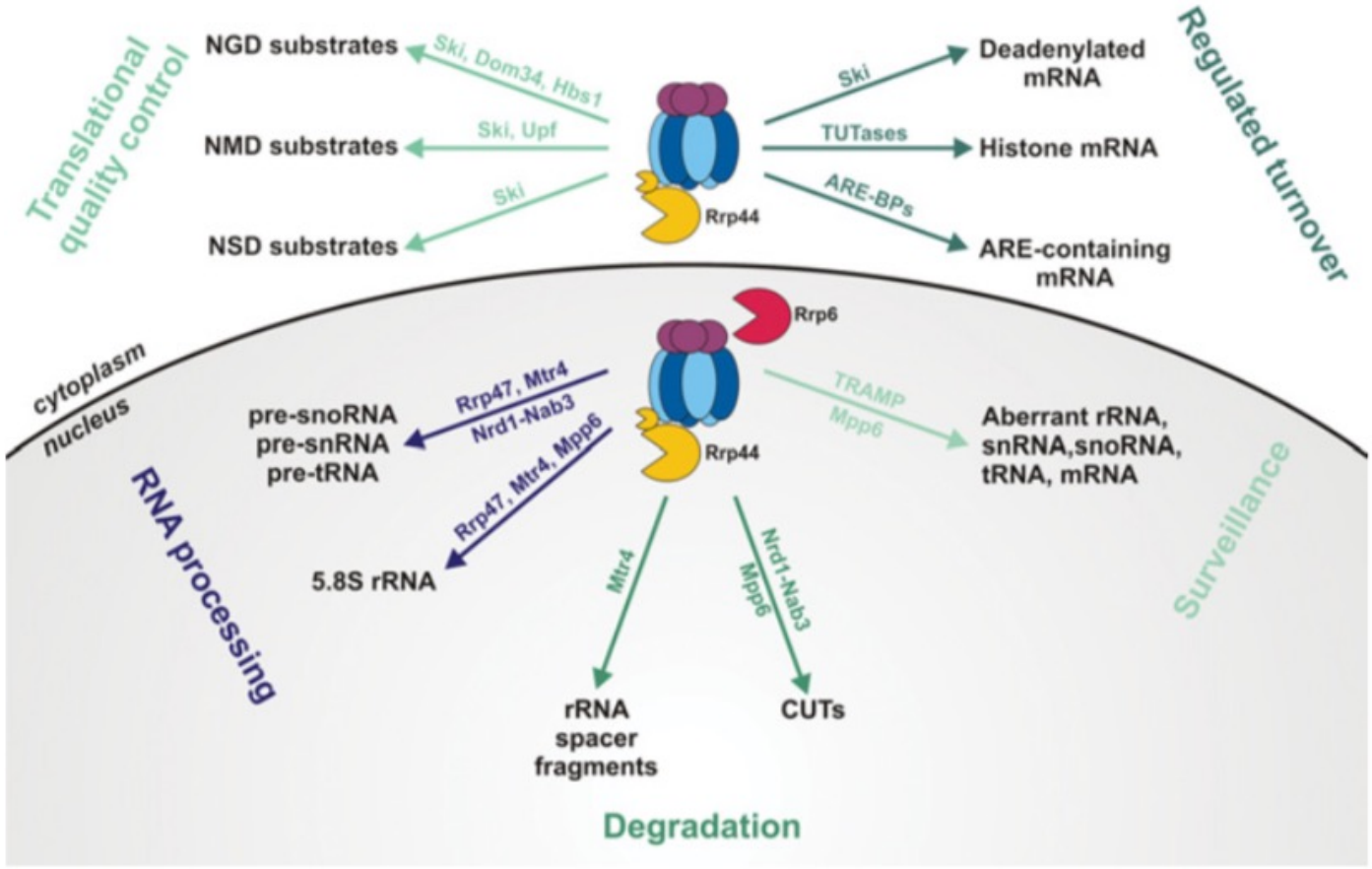
Quality control of transcription by the Nrd1p-dependent termination pathway

Transcription of CUT from regions upstream of or antisense to mRNA-coding genes (red DNA) is terminated by the Nrd1p complex and CUTs are rapidly degraded by the exosome/TRAMP. Quality control by the Nrd1p complex operates to avoid overlapping transcription and to promote degradation of potentially toxic RNAs by the exosome/TRAMP.



Contrary to the CPF pathway, termination by the Nrd1 pathway is coupled to degradation of the transcript produced or trimming of the precursor in case of snRNAs and snoRNAs

Characterized substrates and cofactors for the exosome



CUTs

•Due to their rapid turnover in wild-type cells, these ncRNAs-which are 200-800 nt long, are transcribed by RNAPII, capped and polyadenylated-were named **cryptic unstable transcripts (CUTs)**. The termination of CUT transcription is dependent on the Nrd1–Sen1 pathway.

•CUTs account for 13% of total yeast transcripts

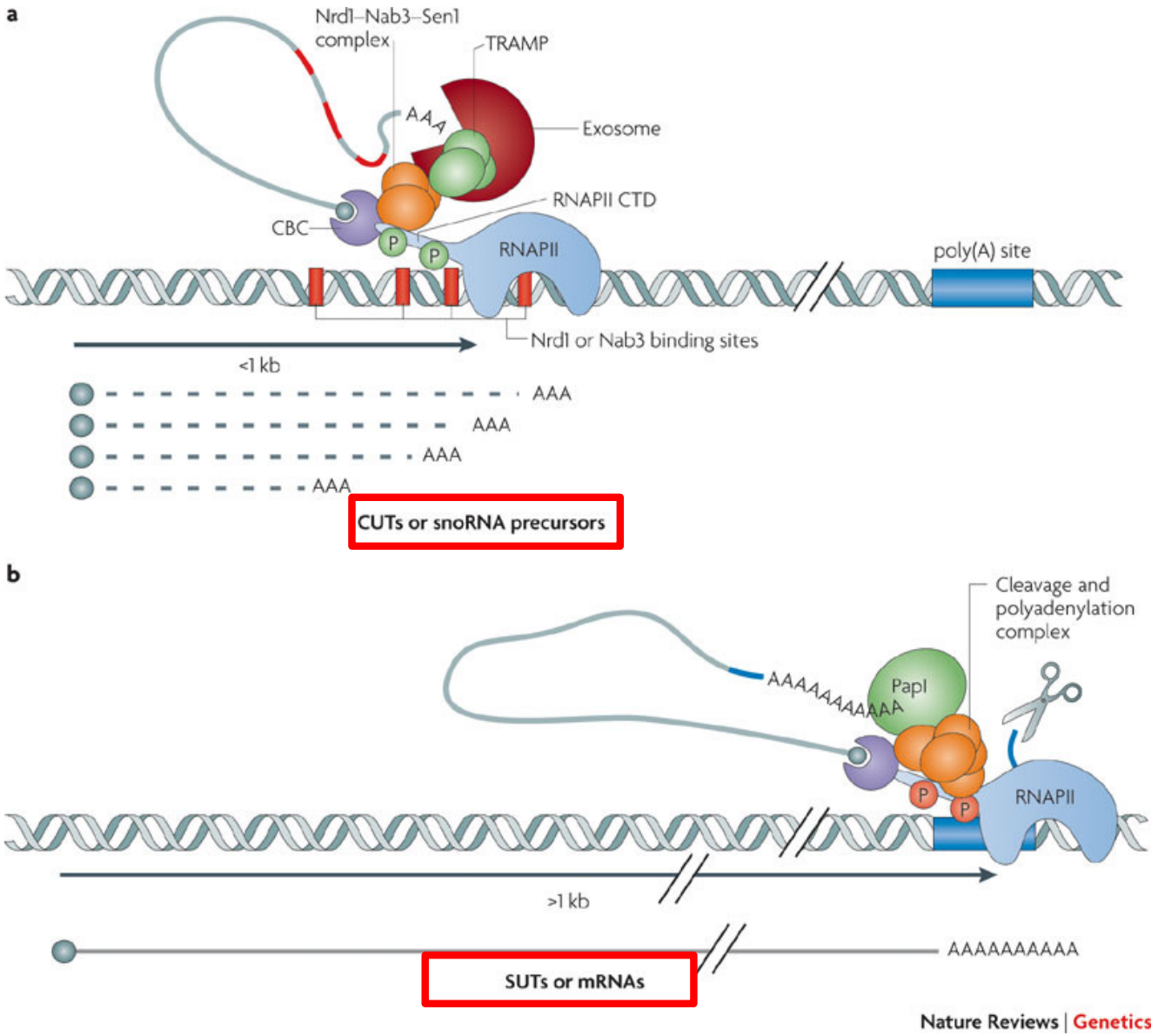
•Nrd1 recruit the TRAMP polyadenylation complex to the CUT. The subsequent TRAMP-dependent polyadenylation of the CUT is believed to facilitate its degradation by targeting the exosome to the transcript

SUTs (stable unannotated transcripts)

•**SUTs** are detectable in wild-type yeast strains, which were appropriately named **stable unannotated transcripts (SUTs)**. These SUTs, of unknown function at present, account for 12% of the transcripts identified by the tiling microarray

•The termination of SUT transcription is dependent on the classical Poly-A pathway (see NEL_L)

The instability of CUTs is linked to their mode of transcription termination



Pervasive transcription in eukaryotic genomes

- up to **85% of the genome is transcribed** in *S. cerevisiae*
- a large number of the transcripts arise from intronic and intergenic regions
- another important pool of non-coding RNAs (ncRNAs) aligns with known ORFs, either in the same orientation (sense) or in opposite orientation (antisense) to the coding transcript
- 11% of the human transcriptome derives from this latter category

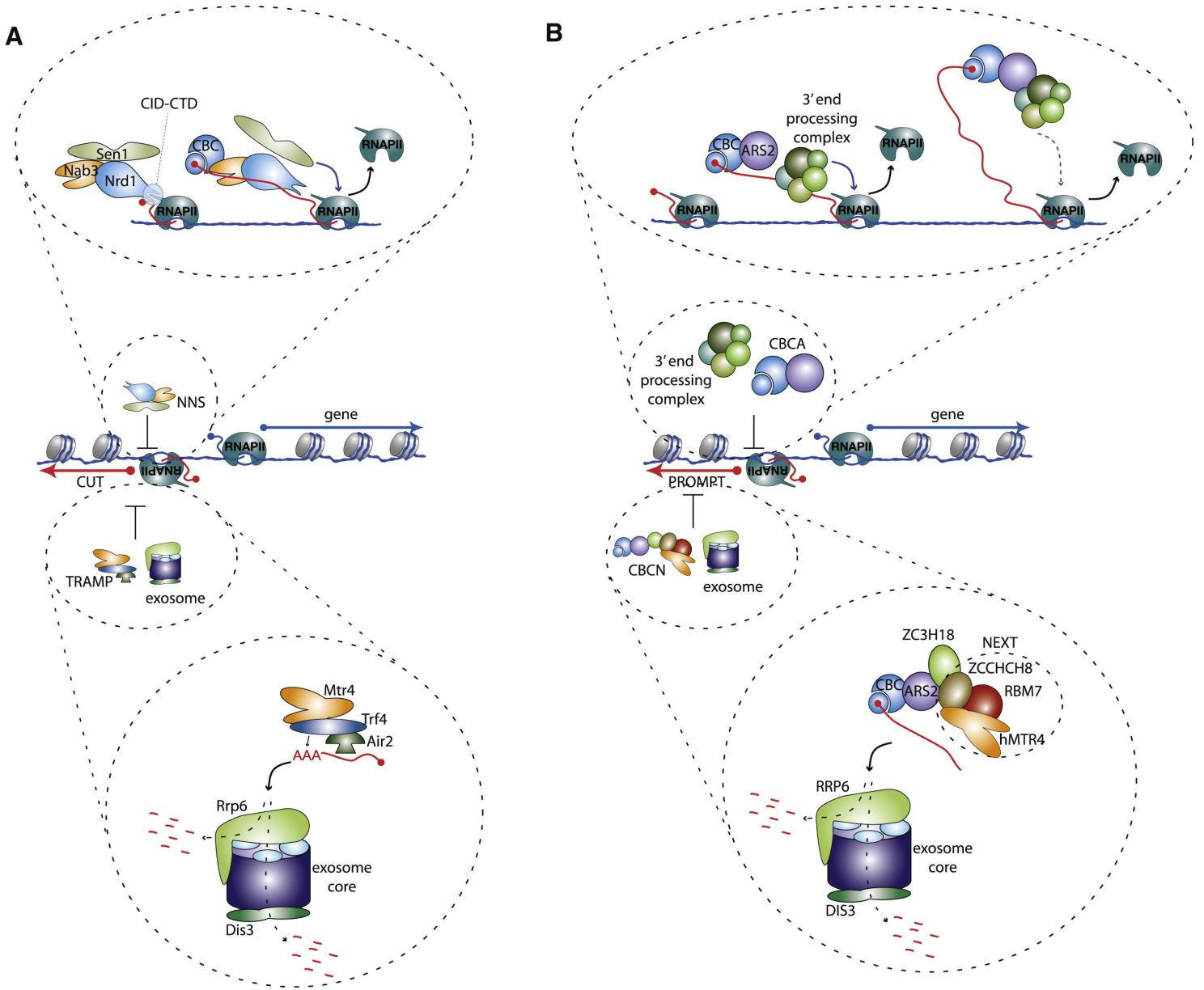
Function for the CUTs and SUTS

•Expression of some genes (not only in yeast) is regulated by overlapping or antisense non coding RNAs. Some CUTs/SUTs have acquired a function in gene expression regulation !

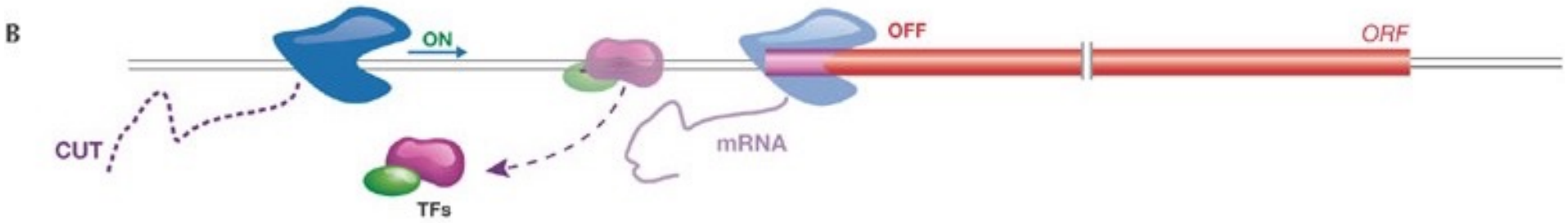
•Hidden transcripts may represent genetic material in evolution towards a functional role (turn noise in sound !!)

In exploring the evolutionary potential of RNA molecules the RNA quality control would maintain a sub-threshold level of fidelity in gene expression

Control of Pervasive Transcription by Termination-Coupled Degradation

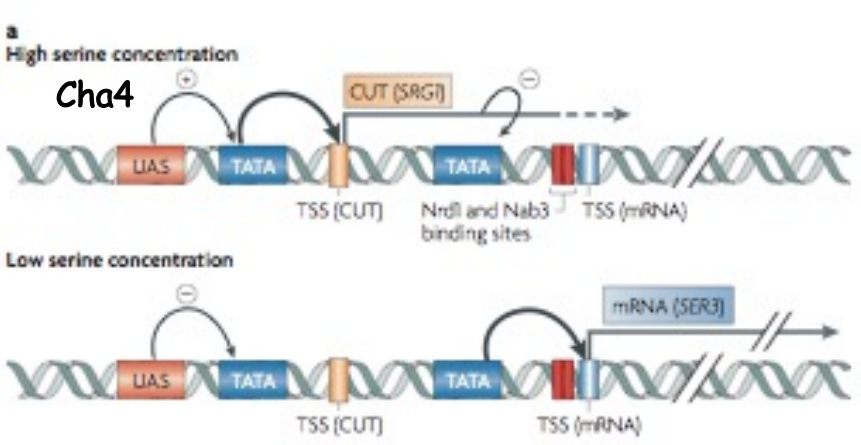


Possible mechanisms for the regulation of genome expression by non-coding transcription



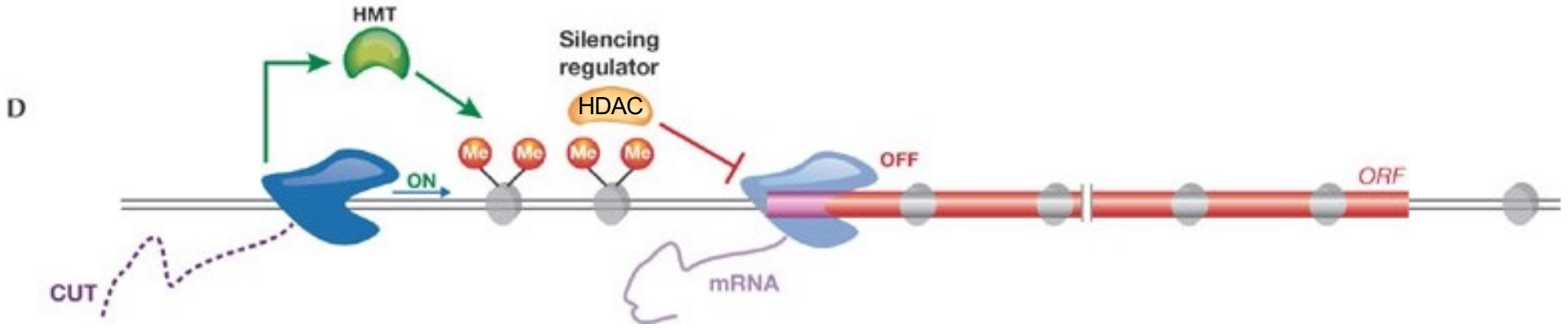
A) The **<u>transcriptional interference mechanism</u>**, in which transcription factors (TFs) are displaced from the mRNA promoter by the upstream cryptic transcription, is shown.

Ex: The *SRG1 cryptic non-coding RNA (ncRNA) interferes with the promoter of the downstream SER3 gene through this mechanism.*



At a high serine concentration, the serine-dependent transcription activator Cha4 binds an upstream activating sequence (UAS) to promote CUT transcription

Possible mechanisms for the regulation of genome expression by non-coding transcription.



B) <u>Transcription-induced chromatin modifications</u>, in which cryptic transcription modifies promoter proximal chromatin to attenuate gene expression.

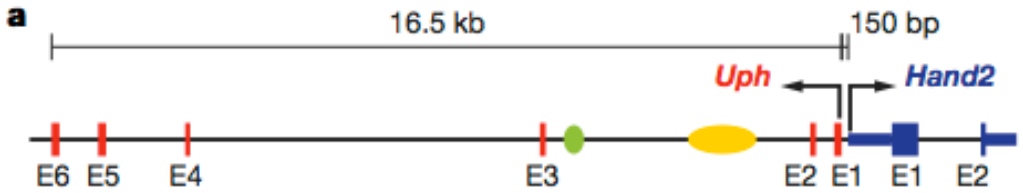
Ex: The *GAL10–GAL1 locus is regulated through this mechanism; cryptic transcription that originates upstream from the GAL10–GAL1 promoter induces the methylation of H3K36 and tethers the Rpd3S histone deacetylase complex (HDAC) to attenuate gene expression of the GAL locus*.

Transcription is sufficient for lncRNA function

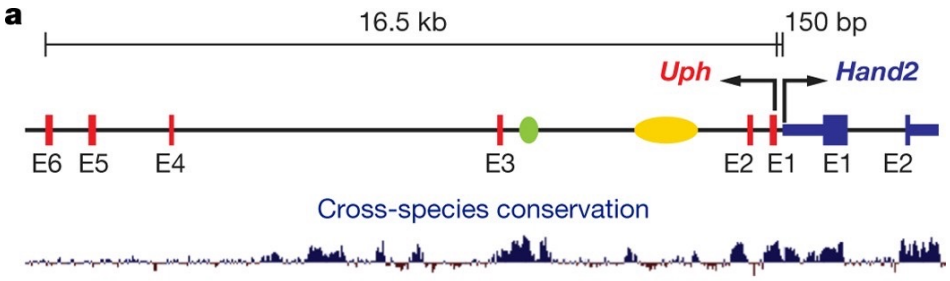
nature 26 October 2016

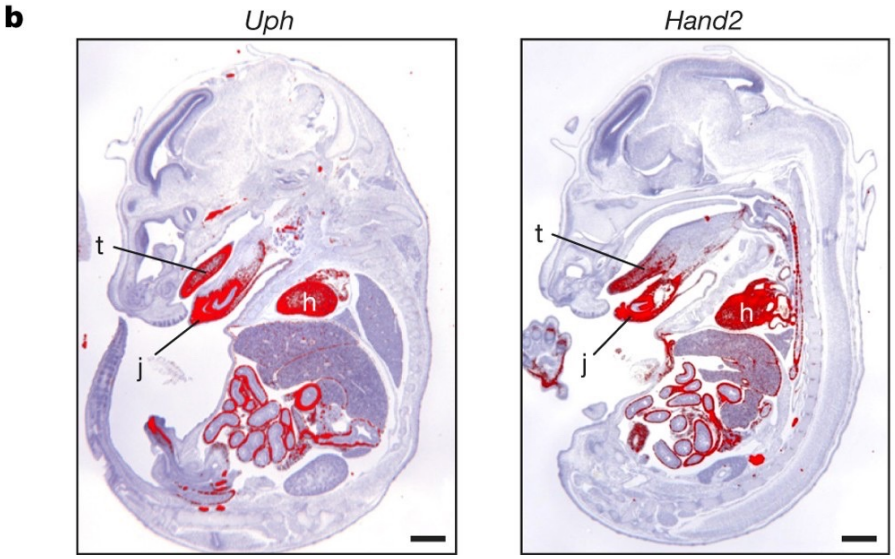
Transcription of the non-coding RNA upperhand controls *Hand2* expression and heart development

Kelly M. Anderson[1,2], Douglas M. Anderson[1,2], John R. McAnally[1,2], John M. Shelton[3], Rhonda Bassel-Duby[1,2] & Eric N. Olson[1,2]

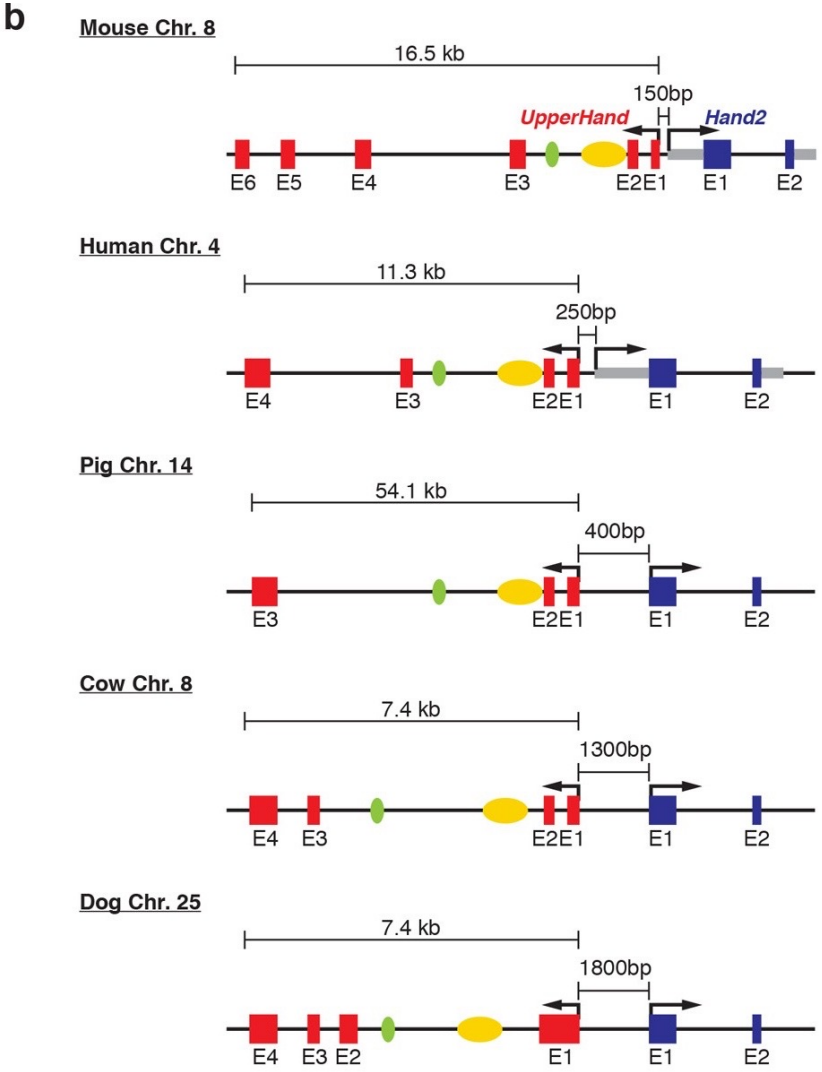


The lncRNA *Upperhand (Uph)* is transcribed upstream of *Hand2* and contains two *Hand2* enhancers (in green and yellow) within one of its introns. Blockade of Uph transcription, but not knockdown of the mature transcript *in vivo* reduced enhancer activity, resulting in RNAPII pausing and loss of Hand2 expression in the heart.



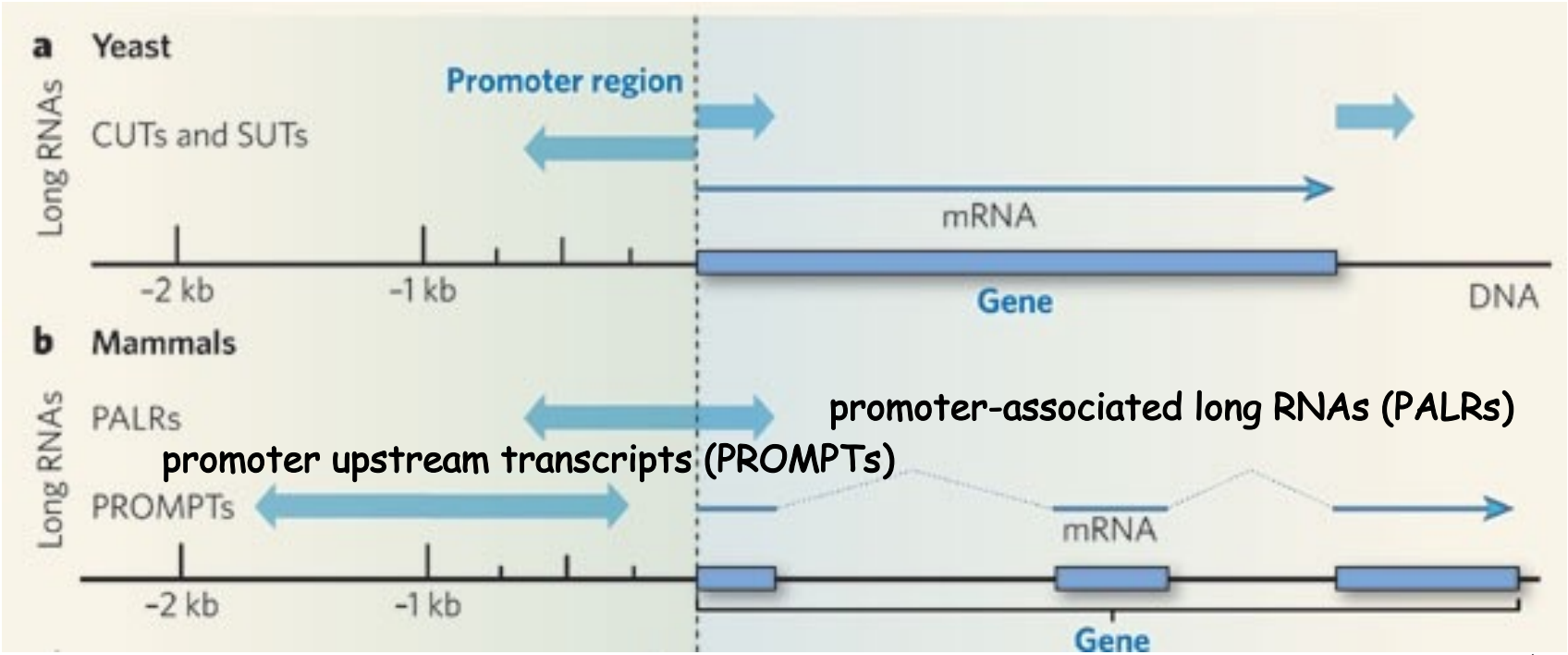


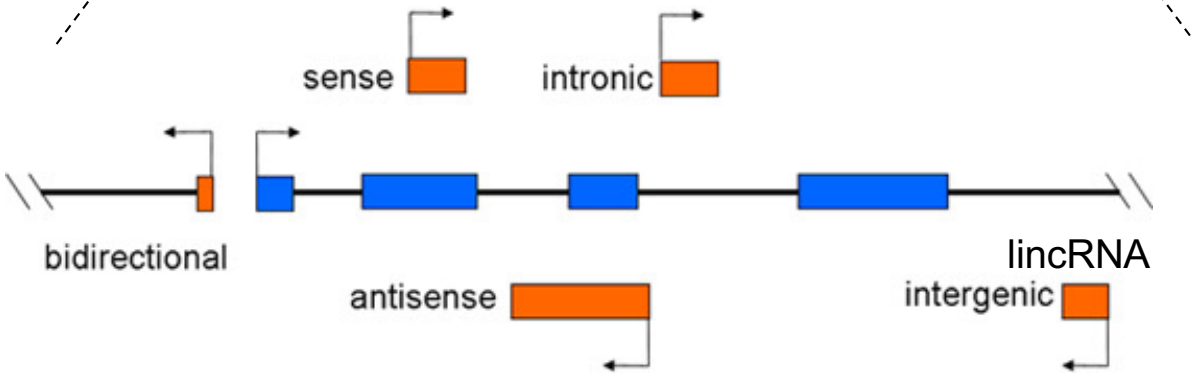
whole-mount *in situ* hybridization for *Uph* in mouse embryos revealed expression in the heart, distal branchial arches and the developing limb bud at embryonic day (E)10.5, overlapping with *Hand2* expression



Similar to most characterized lncRNAs, the Uph nucleotide sequence is not well-conserved, with only ~56% homology between mouse and human, and little conservation across other species. However, Uph orthologues share a promoter and contain the conserved Hand2-associated cardiac and brachial arch enhancers within their second introns.

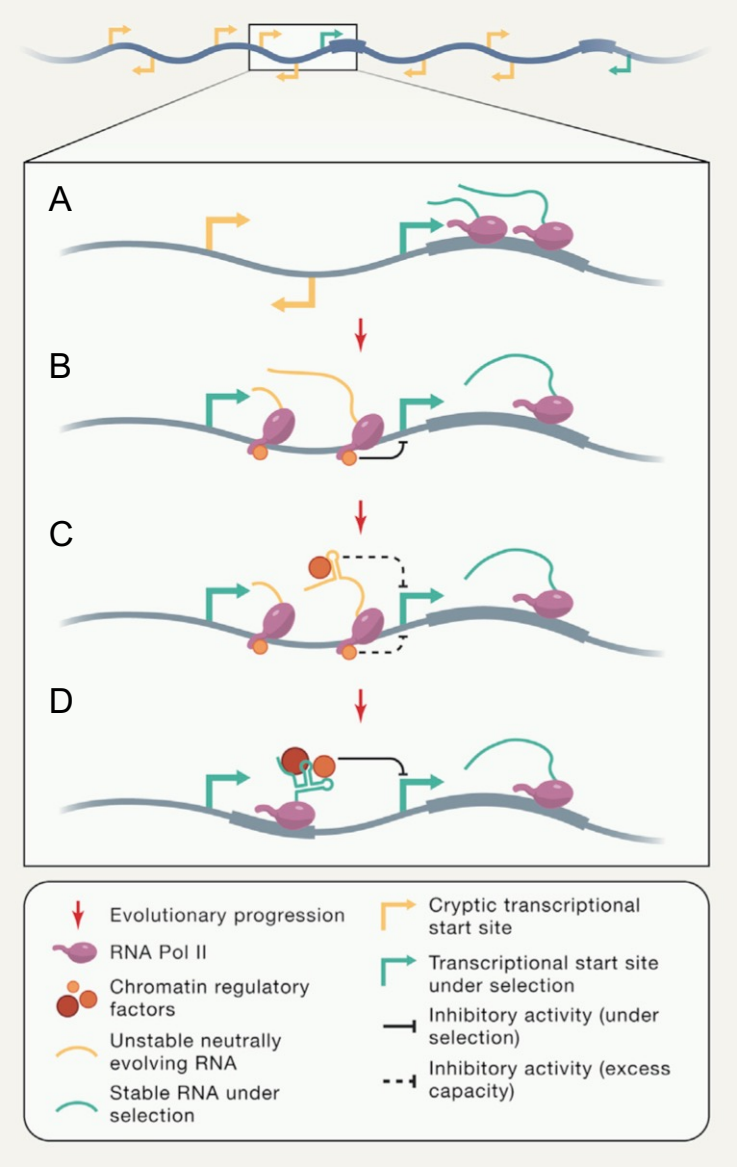
Pervasive transcription in eukaryotic genomes





Evolution of lncRNAs from intergenic transcription

(**A**) Weak, cryptic intergenic transcriptional start sites evolve through neutral mutations.
(**B**) An intergenic transcriptional start site promotes RNA polymerase II- dependent transcription that overlaps a downstream promoter of a functional gene. During transcription, the CTD of RNA polymerase II recruits chromatin- regulatory complexes that silence the promoter. If this regulation becomes beneficial, the intergenic transcriptional start site becomes subject to purifying selection, whereas its transcript evolves neutrally.
(**C**) The junk transcript recruits chromatin regulatory factors either independently or in conjunction with the CTD.
(**D**) The excess capacity of the junk transcript relieves the requirement for overlapping transcription; the transcript is now under purifying selection and evolves into a functional lncRNA that independently inhibits the nearby promoter.



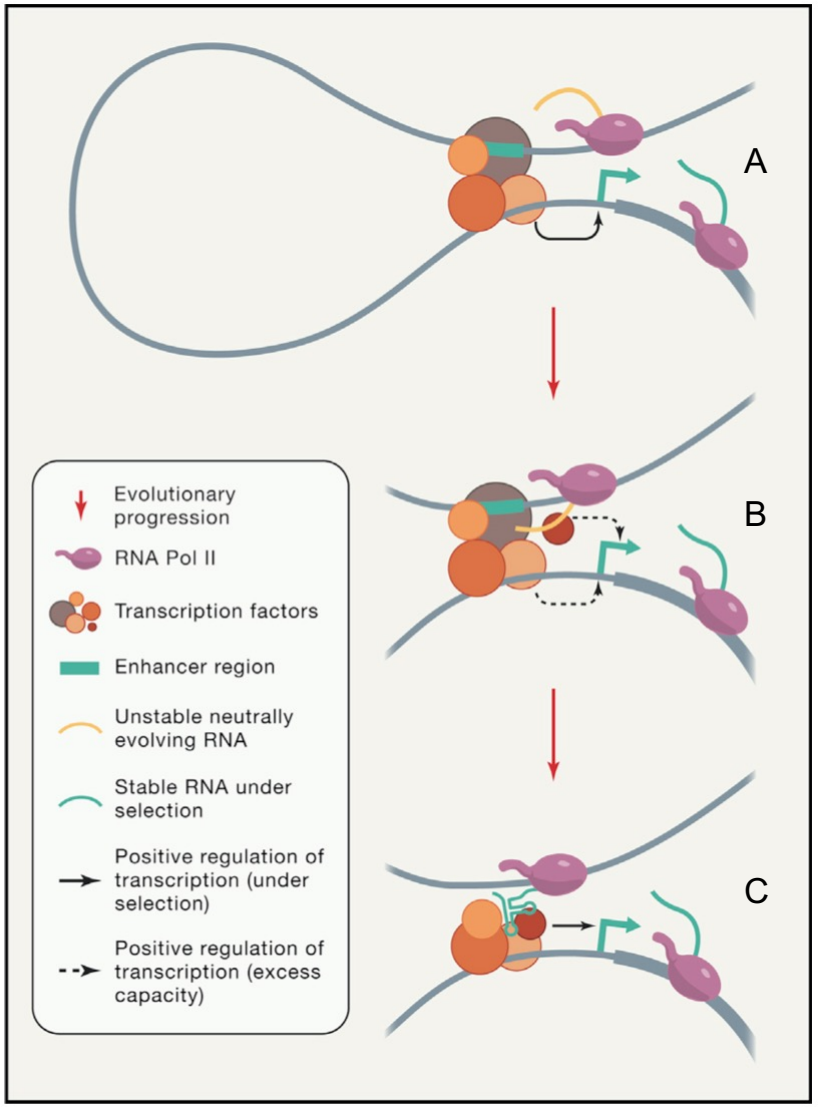
Palazzo and Konin. Cell 2020

Evolution of lncRNAs from eRNAs

(A) The transcriptional start site is regulated by an enhancer region and its associated transcription factors. This region also generates unstable eRNAs that do not contribute to gene regulation.

(B) The eRNA recruits transcription factors that contribute to either DNA looping and/or gene regulation. This excess capacity relieves the requirement for the enhancer region to be maintained.

(C) The transcript is now under purifying selection and evolves into a functional lncRNA that upregulates the transcription of the adjacent gene.



Palazzo and Konin. Cell 2020

Human genome contains many lncRNAs

Version 27 (January 2017 freeze, GRCh38) - Ensembl 90

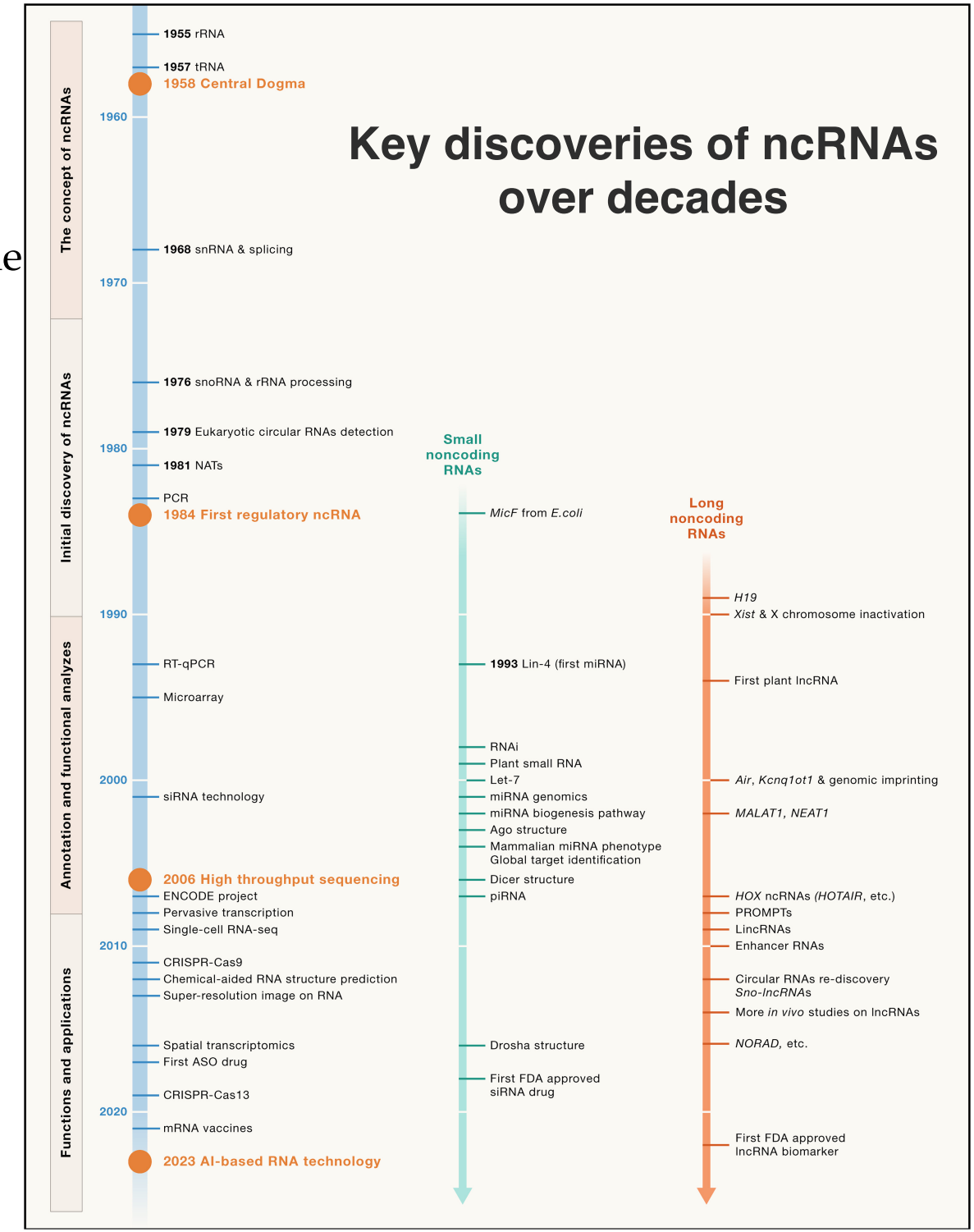
General stats

Total No of Genes	58288
Protein-coding genes	19836
Long non-coding RNA genes	15778
Small non-coding RNA genes	7569
Pseudogenes	14694
- processed pseudogenes:	10704
- unprocessed pseudogenes:	3469
- unitary pseudogenes:	206
- polymorphic pseudogenes:	63
- pseudogenes:	18
Immunoglobulin/T-cell receptor gene segments	
- protein coding segments:	410
- pseudogenes:	234

Total No of Transcripts	200401
Protein-coding transcripts	80930
- full length protein-coding:	55406
- partial length protein-coding:	25524
Nonsense mediated decay transcripts	14208
Long non-coding RNA loci transcripts	27908
Total No of distinct translations	60297
Genes that have more than one distinct translations	13580

Long non coding RNA > 200 nt
Short non coding RNA < 200 nt

Large-scale analyses, like those of the human ENCODE project (Djebali et al., 2012), showed that a combined total of 75% of the genome is transcribed within 15 human cell lines analyzed. Primary transcripts from protein-coding genes explain just 25% of genomic output (1.5% of this being exonic RNA) but are an order of magnitude more abundant than ncRNA. Such high-genome activity implies significant overlap among transcripts (albeit not necessarily in the same tissue).



Non coding RNAs

Non coding transcripts have very diverse properties and mechanisms of action and can be classified in:

- ***Antisense RNAs,*** which overlap with exons or introns or protein-coding genes.

- ***Long intergenic noncoding RNAs*** (**lincRNAs**), which do not overlap exons of either protein-coding or other non-lincRNA types of genes. Some of them can encode for small functional peptide.

- ***Enhancer RNAs*** **(eRNAs)**, which are produce from coding gene promoters and may contribute to enhancer function during transcription.

- ***Circular RNAs*** (**circRNAs**), circular molecules that can arise from exons (exonic circRNA) or introns (intronic circRNA). They can be also coding.

- **Antisense RNAs**

Antisense RNAs

Antisense transcripts were initially discovered in bacteria more than 30 years ago; soon after this, examples were found in all eukaryotes.

Notably, more than 30% of annotated transcripts in humans have antisense transcription. However, antisense transcripts are more than 10-fold lower in abundance than sense expression.

Antisense transcripts preferentially accumulate in the nucleus. However, some antisense transcripts have been found to be associated with a range of distinct locations, including the mitochondria and the cytoplasm

The expression of some antisense transcripts is linked to the activity of neighbouring genes, whereas many others have distinct expression patterns during different processes, such as cellular differentiation and cancer progression.

Reviewed in Pelechano and Steinmetz, 2013 Nat. Rev. Genetics

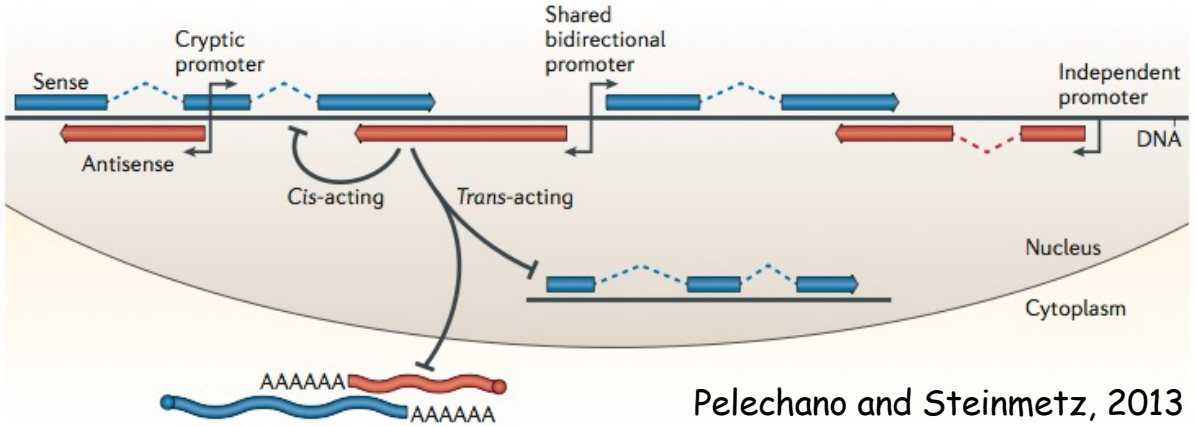
Expression of antisense RNAs

They can arise from ***independent promoters***, ***bidirectional promoters*** of divergent transcription units, or ***cryptic promoters***.

In addition, they can also originate from ***RNA-dependent RNA polymerase*** (RdRP) activity.

The expression of antisense transcripts is subject to regulation at the level of RNA stability; for example, many antisense are targeted for early degradation by the nuclear exosome.

Antisense transcripts can recognize their reverse-complementary sequence in RNA or DNA and can also carry protein-binding domains to modulate gene expression.



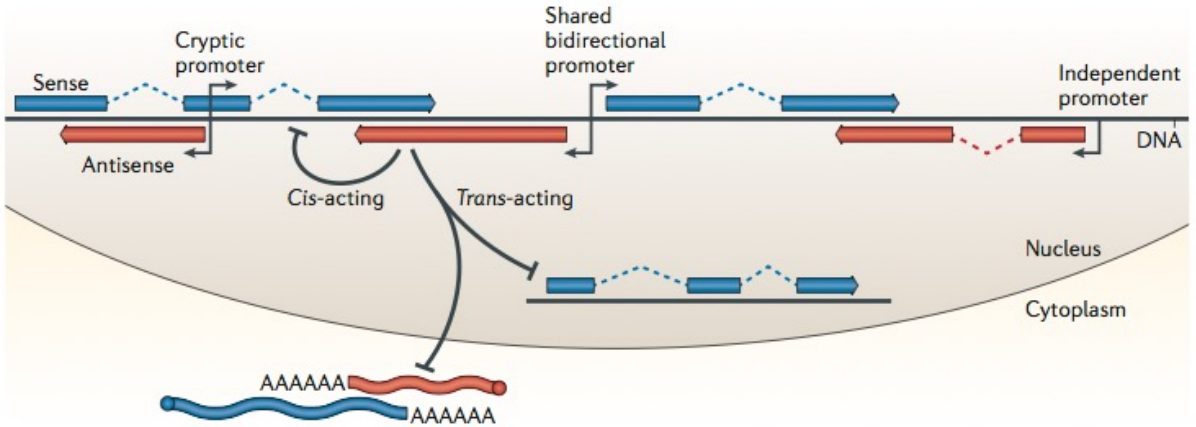
Pelechano and Steinmetz, 2013 Nat. Rev. Genetics

Function of antisense RNAs

The function of an antisense transcript can be mediated by either the **transcript itself** or the **act of its transcription**

The fact that both antisense and sense transcripts are transcribed from the same region suggests that antisense transcripts function more frequently *in cis*. However, regions of antisense transcription can interact with other loci through the three-dimensional organization of chromatin, which can mediate *trans* effects.

The fact that most antisense transcripts function in *cis* also makes it difficult to experimentally identify putative functions of these transcripts.



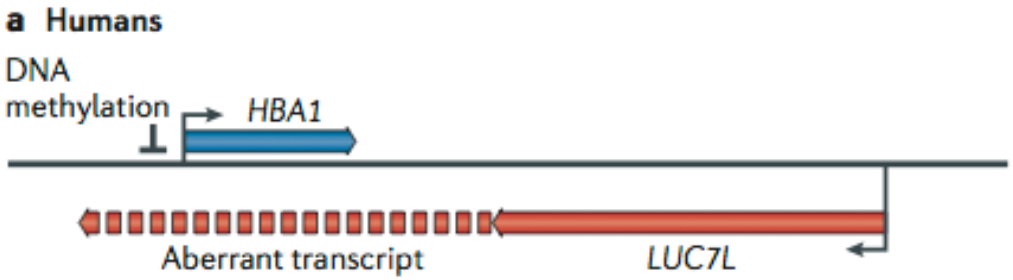
Functional antisense RNAs

	Mechanism of action	Antisense locus	Effects	Species
1.	DNA methylation	*LUC7L*	Methylates *HBA1* promoter CpG island, which represses its expression	Humans
2.	Chromatin modifications	*XIST* and *TSIX*	Inactivates X chromosome gene expression	Mammals
		ANRIL	Represses the tumour suppressor locus *CDKN2B–CDKN2A* by both histone H3 lysine 27 (H3K27) methylation and DNA methylation	Humans
		BDNF-AS	Represses *BDNF* by histone modification	Mammals
		HOTAIR	Silences the *HOXD* locus in *trans* by the recruitment of Polycomb proteins	Humans
		COOLAIR	Represses *FLC* sense gene by H3K4 demethylation and recruits Polycomb proteins, which increase H3K27me3 levels	Plants
		COLDAIR	Antisense to *COOLAIR*; represses *FLC* sense gene by the recruitment of Polycomb proteins	Plants
		AS to *PHO84*	Represses *PHO84* by histone deacetylation both in *cis* and in *trans*	*S. cerevisiae*
		RTL	Silences transcription of the *Ty1* retrotransposon in *trans* through chromatin modification and post-transcriptionally controls its retrotransposition	*S. cerevisiae*
3.	Transcriptional interference	*RME2*	Represses *IME4* by transcriptional interference in *cis* and functions after transcription initiation of *IME4*	*S. cerevisiae*
		Airn	Regulates *Igf2r* imprinting	Mice
4.	Isoform variation	*ZEB2-AS*	Induces exon skipping in *ZEB2*, which produces an alternative isoform that has increased translation efficiency	Humans
5.	Translation efficiency	AS to *Uchl1*	Increases translation efficiency of *Uchl1* using a SINEB2 domain	Mice
		SymR	Decreases translation efficiency of *SymE* by competing with binding of the 30S ribosome	Enterobacteria
6.	RNA stability	*BACE1-AS*	Increases stability of *BACE1* by masking an microRNA-binding site	Humans
		WDR83 and *DHPS*	Increase their mutual stability by forming a duplex within their 3′ untranslated regions	Humans

Functional antisense RNAs

DNA methylation:

An inherited form of *anemia* (alpha-thalassemia) is due to a mutation that results in a termination defect of the widely expressed gene (LUC7L) becoming juxtaposed to a structurally normal alpha-globin gene (HBA2). Aberrant transcriptional extension of the LUC7L (putative RNA-binding protein Luc7-like) produces a DNA methylation of a CpG island in the *HBA1* promoter and represses its expression.

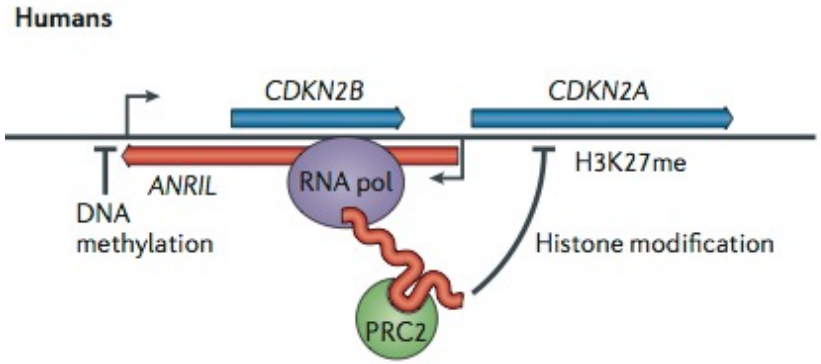


Tufarelli et al., 2003. Nat. Genetics

Functional antisense RNAs

Chromatin modification:

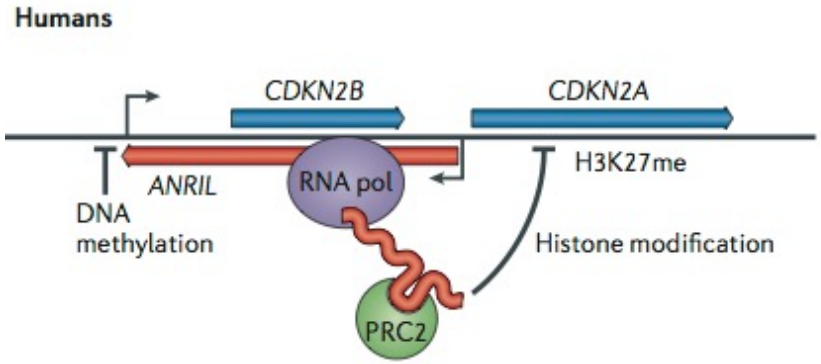
ANRIL (antisense ncRNA in the *INK4* locus*),* the expression of which is increased in prostate cancer, mediates the specific repression of the tumour suppressor locus *CDKN2B–CDKN2A* by recruiting PRC2 and inducing repressive H3K27me3 *in cis* (Yap et al. 2010 Mol Cell). This also leads to long-term promoter DNA methylation at this locus (Yu et al., 2008 Nature).



Functional antisense RNAs

Chromatin modification:

ANRIL (antisense ncRNA in the *INK4* locus), the expression of which is increased in prostate cancer, mediates the specific repression of the tumour suppressor locus *CDKN2B–CDKN2A* by recruiting PRC2 and inducing repressive H3K27me3 *in cis* (Yap et al. 2010 Mol Cell). This also leads to long-term promoter DNA methylation at this locus (Yu et al., 2008 Nature).

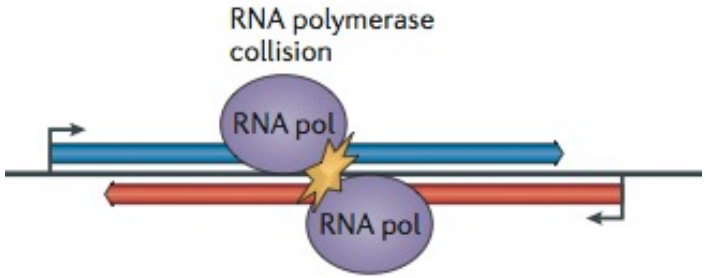


When antisense transcription affects sense expression through chromatin-mediated mechanisms, it is not necessary for the antisense and sense transcripts to be present in the same cell at the same time.

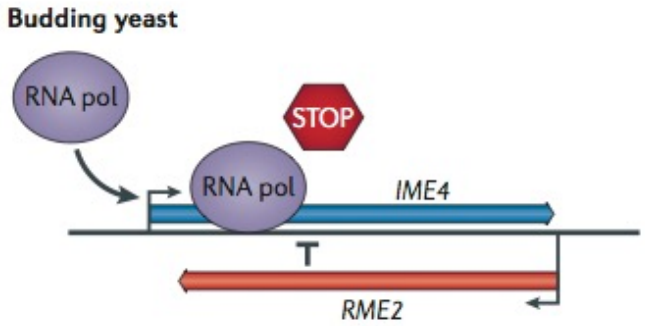
Functional antisense RNAs

Transcriptional interference:

Antisense expression can regulate gene expression after transcription initiation by transcriptional interference that occurs co-transcriptionally. Head-to-head transcription can lead to RNA polymerase collision (Hobson et al. 2012 Mol Cell)



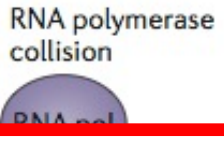
An example of transcriptional interference that functions after transcription initiation is the repression of the *IME4* locus (which encodes a key regulator of meiosis) in budding yeast by its antisense transcript regulator of meiosis 2 (*RME2*). (Hongay et al., 2006 Cell).



Functional antisense RNAs

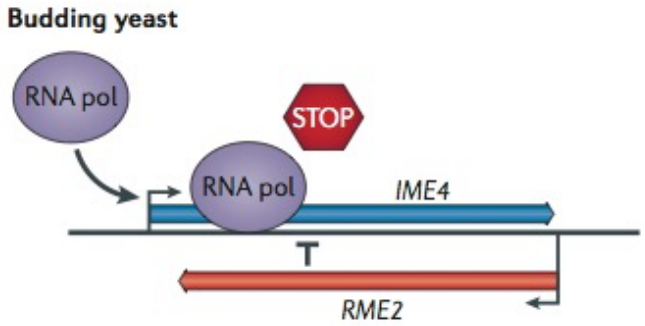
Transcriptional interference:

Antisense expression can regulate gene expression after transcription initiation by transcriptional interference that occurs co-transcriptionally. Head-to-head transcription can lead to RNA polymerase collision (Hobson et al. 2012 Mol Cell)



When antisense transcription affects sense expression through transcriptional interference, it is not necessary for the antisense transcripts to be stably accumulated in the cell.

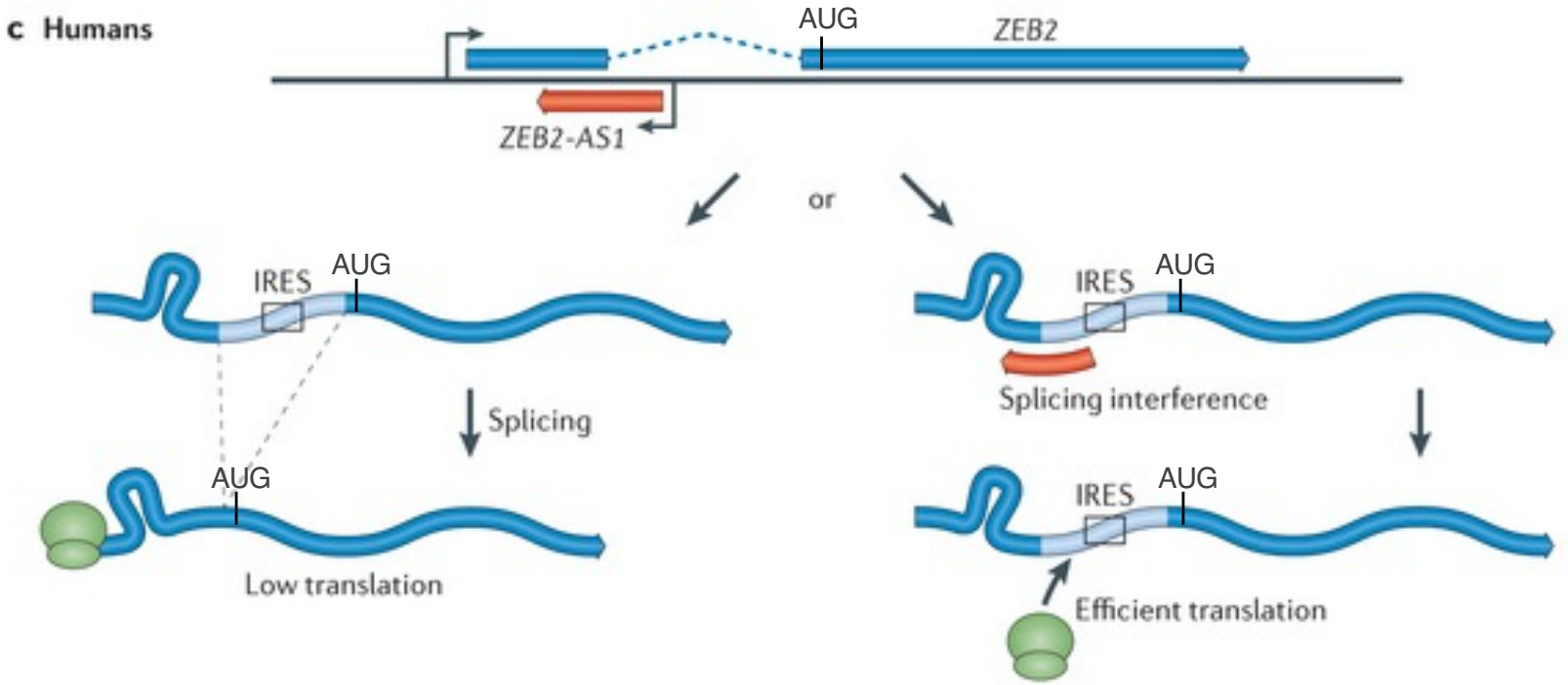
A n is
th ling
ye rnal
region of the *IME4* gene is necessary for antisense-mediated repression, which suggests that antisense-mediated transcriptional interference blocks the elongation, but not the initiation, of the *IME4* transcript (Hongay et al., 2006 Cell).



Functional antisense RNAs

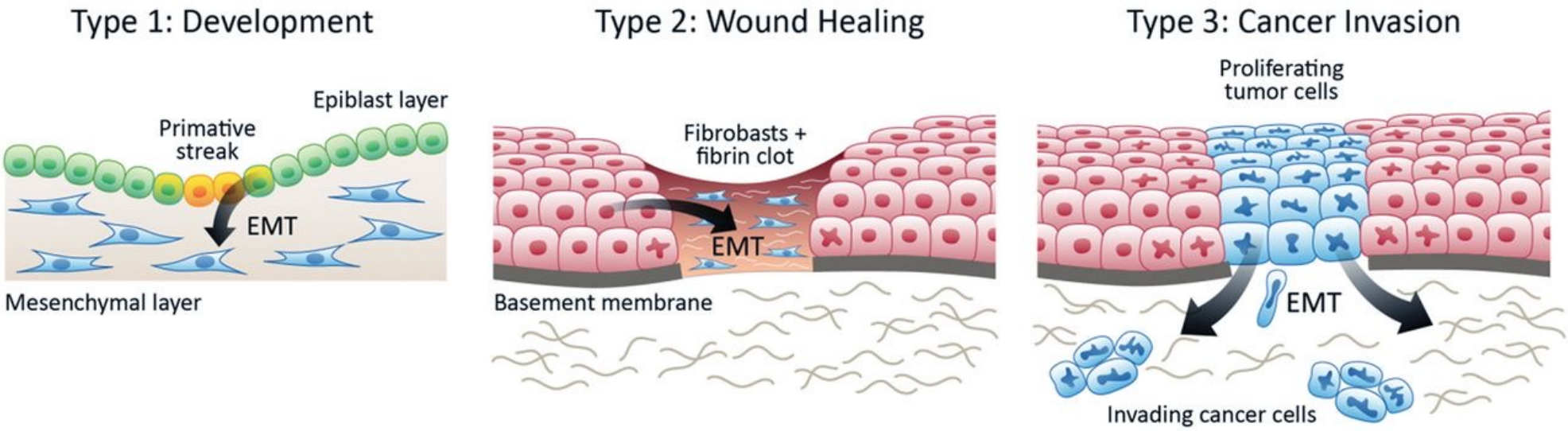
Splicing regulation:

Upon epithelial-to-mesenchymal transition, a **transcript antisense to the *ZEB2* gene (ZEB2-AS1)** is produced, and sequence complementarity between the sense-antisense pairs precludes splicing of a long intron in *ZEB2* pre-mRNA, which encodes a transcriptional repressor of E-cadherin (it plays an important role in cell adhesion). The retained intron includes an IRES close to the start of translation. A sequence in the 5'-UTR region of the *ZEB2* mRNA limits ribosome scanning, such that only the presence of this IRES in the final product allows efficient *ZEB2* translation, allowing ZEB2 protein synthesis in mesenchymal cells (Beltran et al., 2008 Genes Dev).



Epithelial-to-Mesenchymal Transition (EMT)

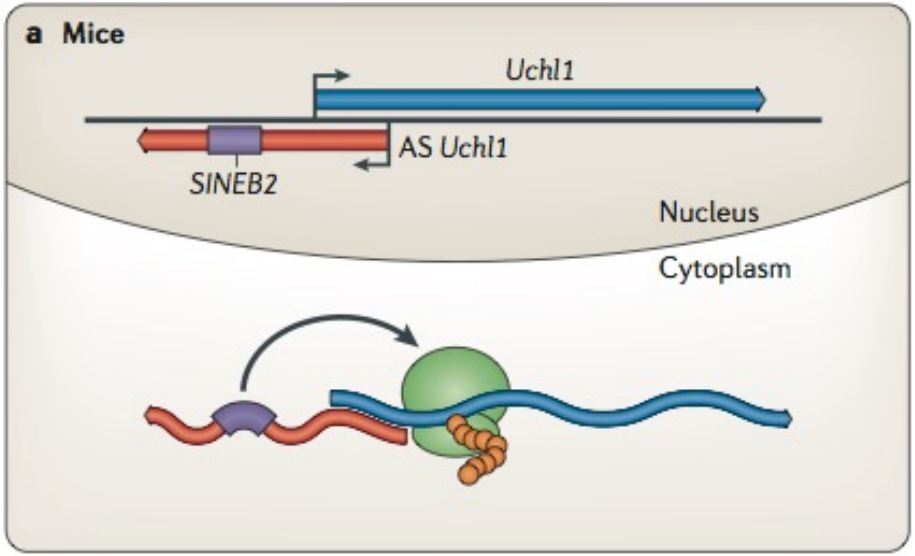
The epithelial–mesenchymal transition (EMT) is a process by which epithelial cells lose their cell polarity and cell-cell adhesion, and gain migratory and invasive properties to become mesenchymal stem cells; these are multipotent stromal cells that can differentiate into a variety of cell types. EMT is essential for numerous developmental processes including mesoderm formation and neural tube formation (1). EMT has also been shown to occur in wound healing (2) and in the initiation of metastasis for cancer progression (3).



Functional antisense RNAs

Translational regulation:

An example of antisense RNA with a direct effect on translation is the mouse ubiquitin carboxy-terminal hydrolase L1 (*Uchl1*) antisense transcript (AS Uchl1), which increases the translation of *Uchl1* (Carrieri et al., 2012 Nature). Specifically, the antisense transcript binds to the 5′-region of the sense transcript, and the SINEB2 domain on the antisense molecule then increases *Uchl1*-translation efficiency by acting as an IRES. In addition, nuclear–cytoplasmic shuttling of the antisense transcript regulates the efficiency of *Uchl1* translation.

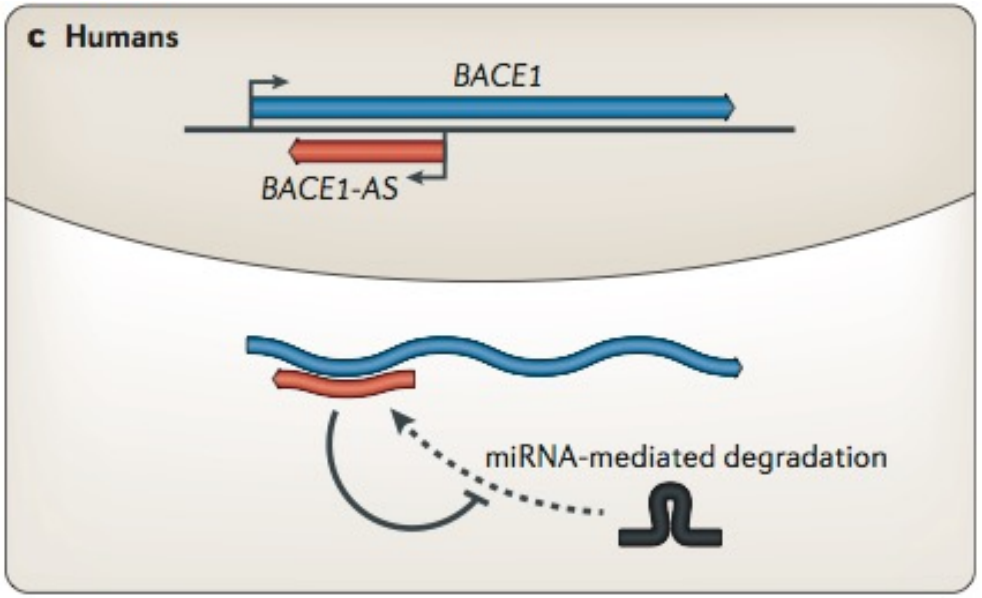


Functional antisense RNAs

mRNA stability:

Antisense expression can also affect the stability of target mRNAs. Antisense transcripts have been shown to increase the stability of their target sense mRNAs by masking specific sites that would otherwise lead to mRNA degradation. One example, is that of the antisense transcript to the **β-site APP-cleaving enzyme 1 gene (*BACE1*)**, which encodes β-secretase 1 — an enzyme that has a central role in the progression of Alzheimer's disease. The antisense transcript forms an RNA duplex with the sense mRNA, and this duplex masks a binding site for the miRNA miR-485-5p, which consequently suppresses miRNA-induced decay and translational repression of *BACE1* (Faghihi et al., 2010 Genome Biol.)

The BACE1-antisense transcript is markedly up-regulated in brain samples from Alzheimer's disease patients



- **LincRNAs**

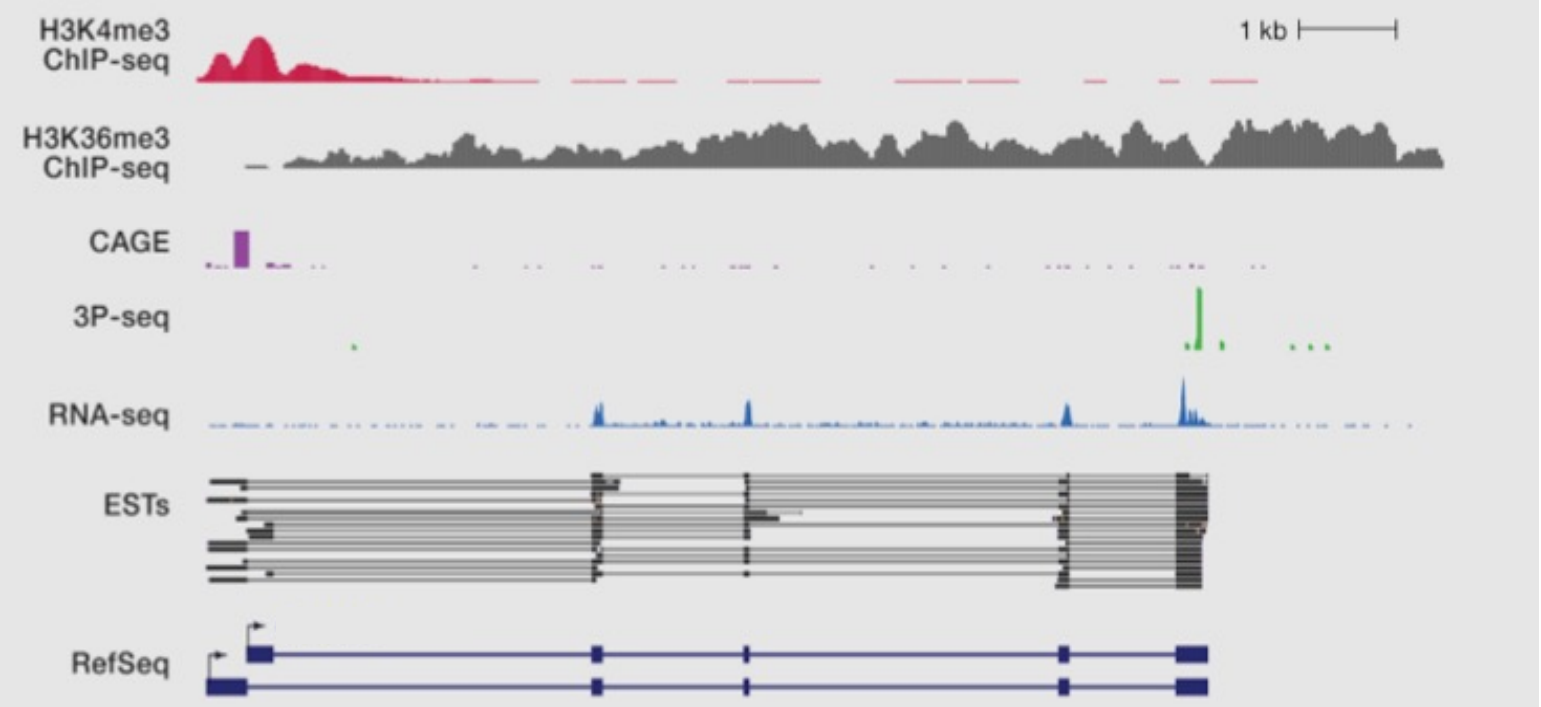
LincRNAs (long intergenic or long intervening non-coding RNAs)

- *not* overlapping transcripts of other types of genes
- *not* coding for proteins (but some of them contain small ORF and can produce small peptides)
- being RNA Pol II products (capped and polyadenylated)
- being longer than 200 nt
- transcriptional regulation, chromatin- modification patterns, and splicing signals of lincRNAs are similar to those of protein-coding genes
- generally less express than coding mRNAs, the median lincRNA level is only about a tenth that of the median mRNA level
- can be localised in both nucleus and cytoplasm

Reviewed in Ulitsky and Bartel, 2013 Cell

lincRNA identification

High-throughput sequencing of short RNA fragments (**RNA-seq**) is widely used for lincRNA identification. RNA-seq has yielded billions of strand-specific paired-end reads of ~ 100 nt each, and those can be sufficient for reconstruction of even very lowly expressed transcripts, such us lincRNAs. Additional data sets that can improve transcript models include chromatin maps (**ChIP-seq**) and data from methods used to identify transcript start (CAGE) and polyadenylation sites (**3P-seq**)



Ulitsky and Bartel, 2013 Cell

lincRNA evolution

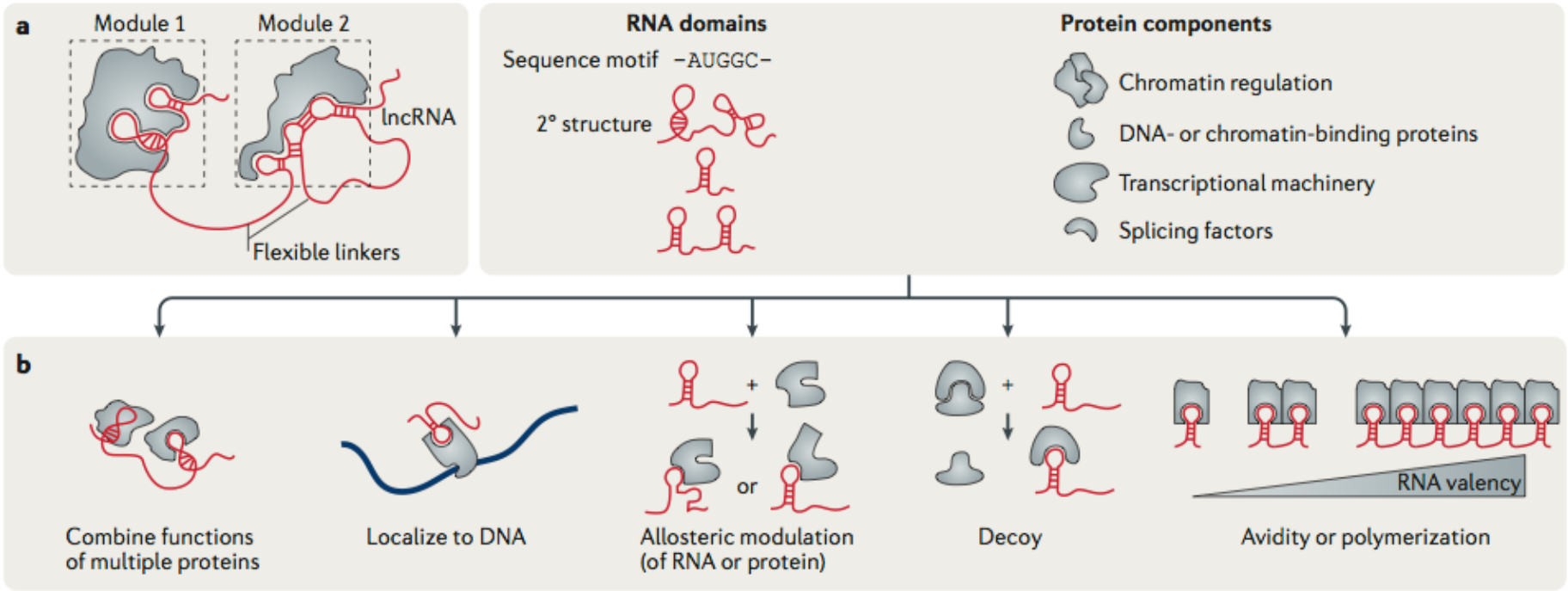
In stark contrast to mRNAs and many classes of noncoding RNAs, mammalian lincRNAs lack known orthologs in species outside of vertebrates - one exception is the Telomeric repeat-containing RNA (Terra), which is conserved between human and yeast.

Only ~12% of human and mouse lincRNAs appear to be conserved in the other species but more lincRNAs are at conserved genomic locations.

LincRNAs might be under pressure to conserve structure but not sequence, and thus homologs would be missed with methods that focus on primary-sequence homology, or the fraction of lincRNAs that are non functional is large.

Nuclear lincRNA

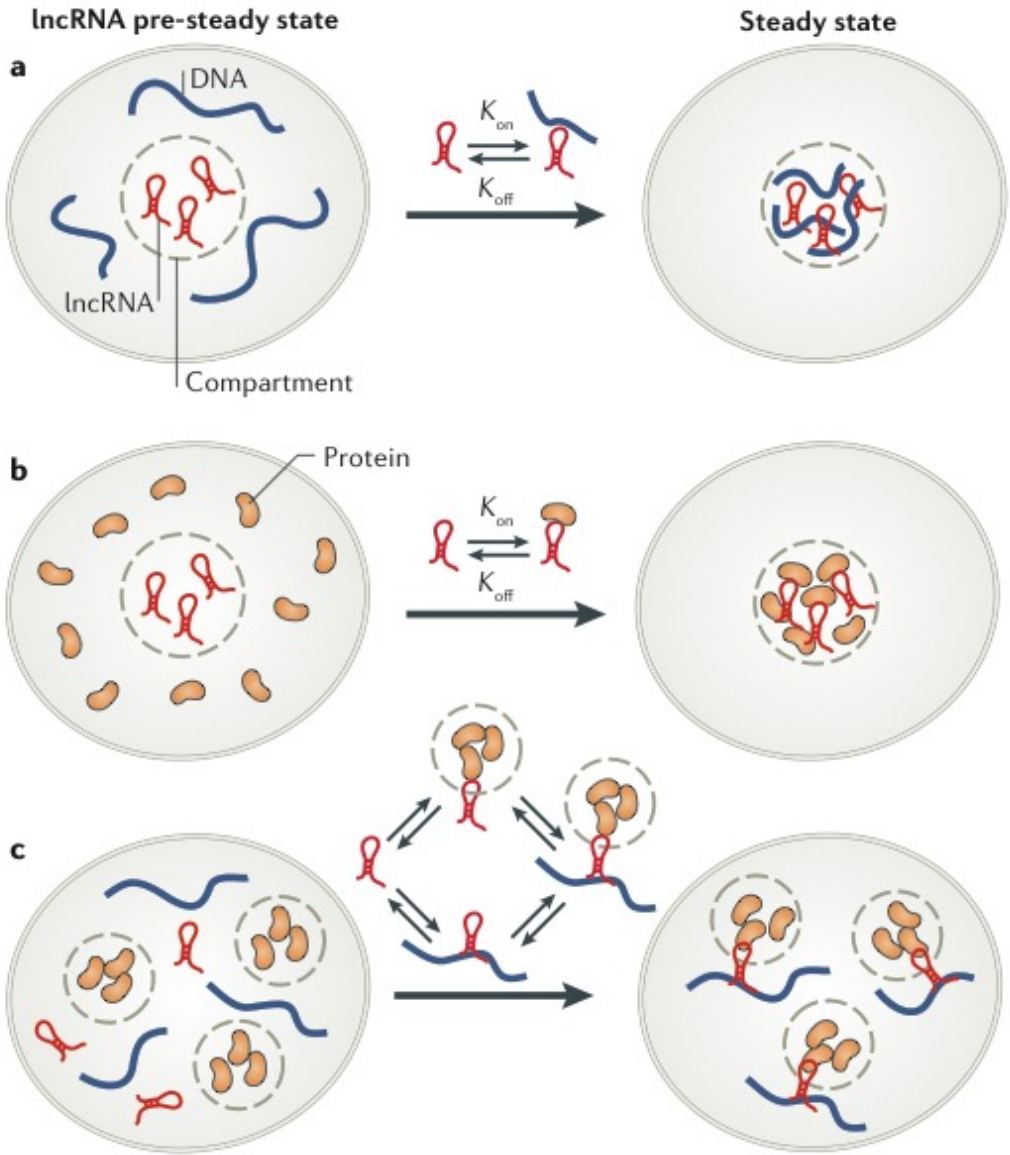
Various regions of the long non-coding RNA (lncRNA) molecule can interact with distinct protein complexes in a modular manner. LncRNA–protein interactions can have various functions, including: **combining the functions of multiple proteins**; **localizing lncRNAs to genomic DNA**; **modifying the structure of lncRNAs or proteins**; **competitively inhibiting protein function** (as decoys); and **providing a multivalent platform**, for example, to increase the avidity of protein interactions or to promote RNA-protein complex polymerization.



Engreitz et al., 2018. Nature Rev. Mol. Cell Biol.

LincRNAs can form spatial compartments in the nucleus

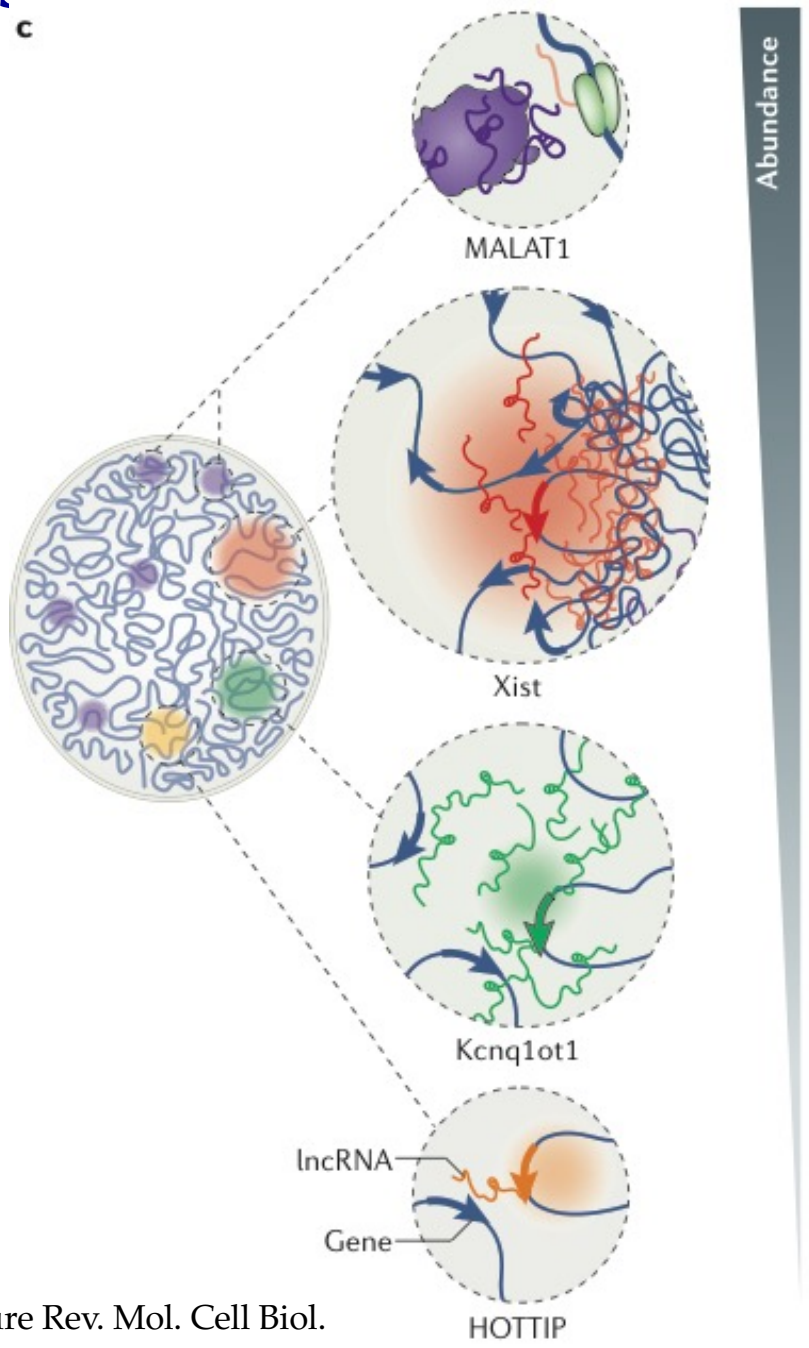
Interactions with proteins and DNA enable lncRNAs to form dynamic spatial compartments in the nucleus. Assembly may be controlled by passive diffusion and interactions that are characterized by specific association and dissociation constants (K_{on} and K_{off}).



Engreitz et al., 2018. Nature Rev. Mol. Cell Biol.

LincRNAs an spatially amplify regulatory signals in the nucleus

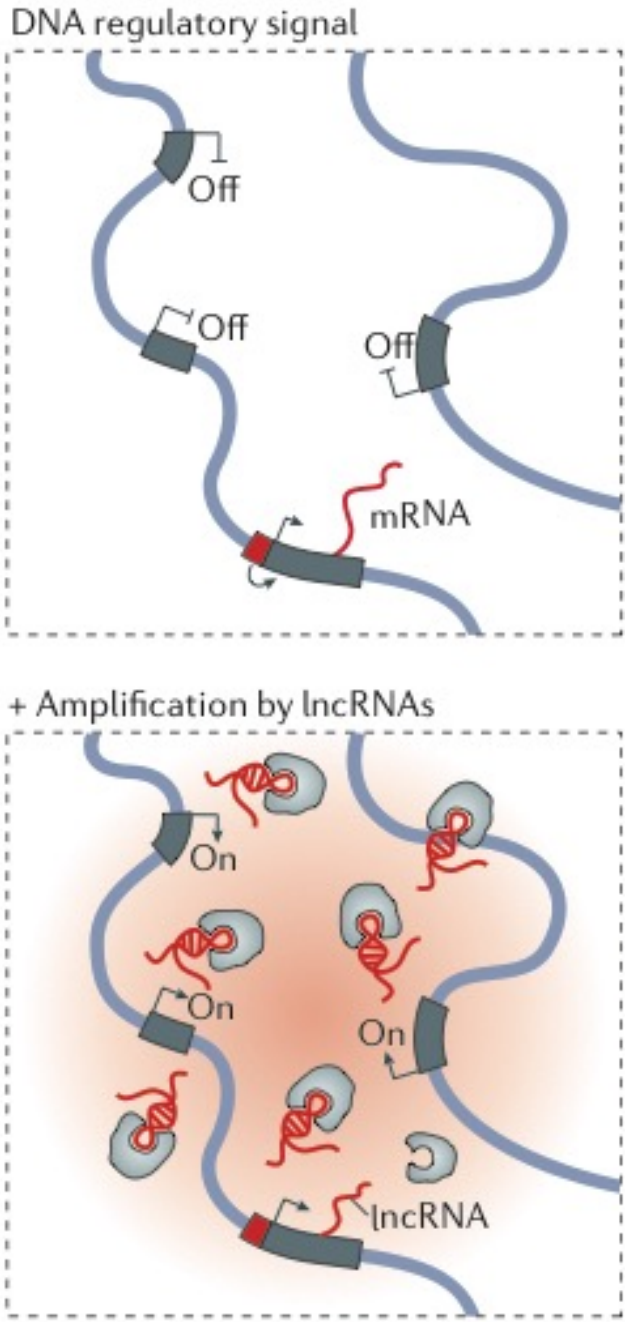
LncRNAs are unique in their ability to spatially amplify regulatory information encoded by DNA. Unlike proteins, lncRNAs can act in close proximity to their site of transcription; and unlike DNA regulatory elements, lncRNAs can amplify DNA-encoded regulatory signals to different extents according to their expression levels.



Engreitz et al., 2018. Nature Rev. Mol. Cell Biol.

LincRNAs an spatially amplify regulatory signals in the nucleus

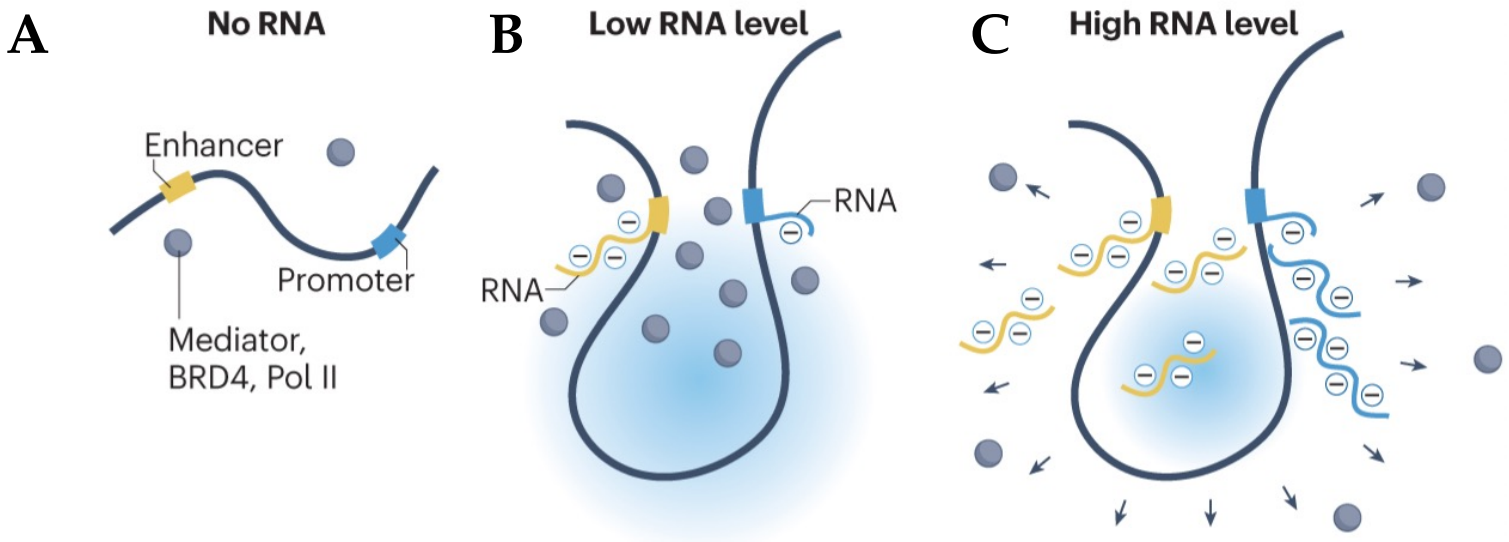
LncRNAs are unique in their ability to spatially amplify regulatory information encoded by DNA. Unlike proteins, lncRNAs can act in close proximity to their site of transcription; and unlike DNA regulatory elements, lncRNAs can amplify DNA-encoded regulatory signals to different extents according to their expression levels.



Engreitz et al., 2018. Nature Rev. Mol. Cell Biol.

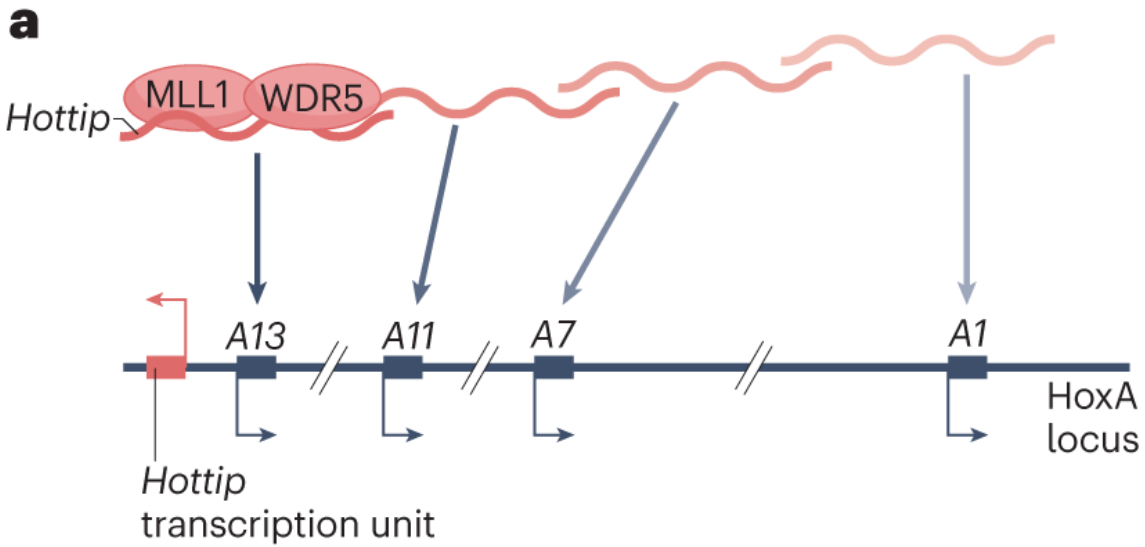
Local RNA abundance provides feedback on transcription initiation

(A) The Mediator complex, the histone-acetylation reader bromodomain-containing protein 4 (BRD4) and RNA polymerase II (Pol II) are present in low abundance at transcriptionally inactive promoter and enhancer elements. (B) Upon transcription initiation, nascent RNAs produced from promoter and enhancer regions nucleate the formation of a condensate, which increases the local concentration of transcription regulators, thereby causing a burst in transcription. (C) As transcription proceeds, the increase in local RNA abundance beyond a certain threshold generates electrostatic repulsive forces that disperse the transcriptional condensates, thereby ending the transcription burst.



LncRNAs as scaffolds

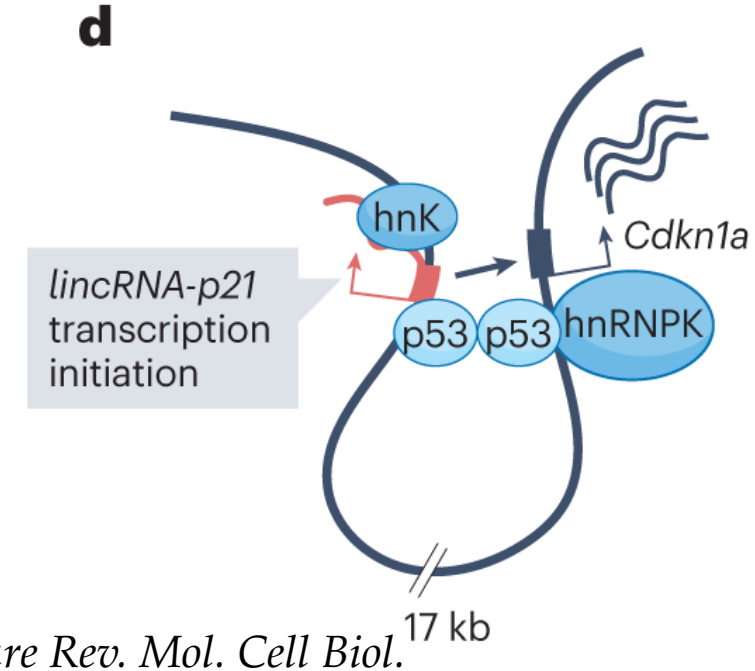
lncRNA HOTTIP, has been proposed to form a local concentration gradient that provides a scaffold for locus-specific recruitment of regulatory complexes, which in turn regulate the transcription of neighbouring genes The lncRNA Hottip is expressed from the HoxA locus and serves as a scaffold for the local recruitment of the histone methyltransferase complex comprising MLL1 (also known as KMT2A) and WDR5 to HoxA gene (A1– A13) sites of transcription. Consistent with local activity, RNAi-mediated depletion of Hottip preferentially affects Hottip proximal, compared with distal, HoxA genes (fading colour gradient)



Ferrer and Dimiytrova, 2024. Nature Rev. Mol. Cell Biol.

Cis-activation by transcription initiation of lncRNAs

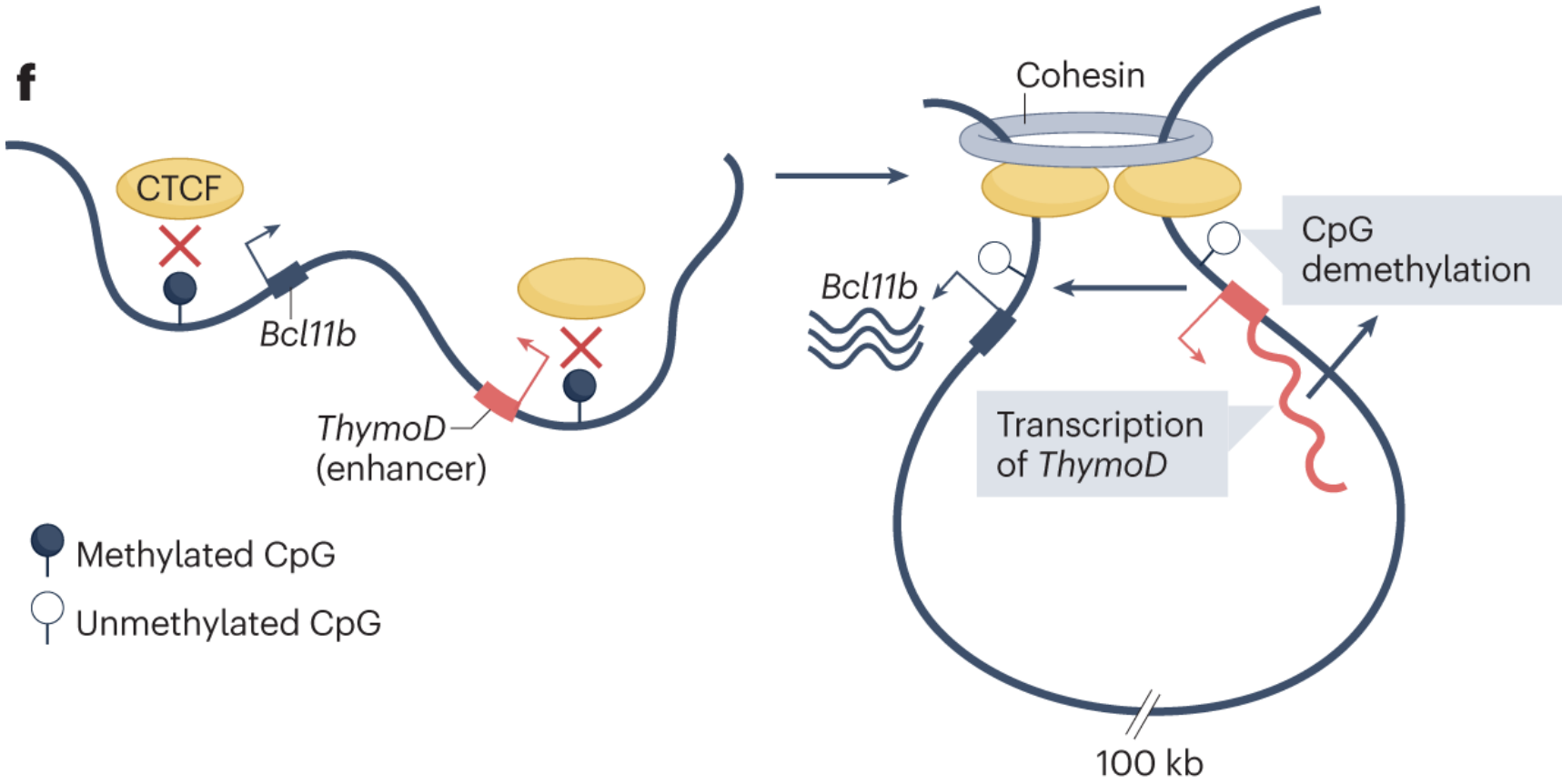
lincRNA-p21 illustrates the functional importance of transcription initiation.
lincRNA-p21 and its cis-activated target, the neighbouring gene Cdkn1a (encoding p21), are in 3D proximity, and are co-regulated by the TF p53. Transcription initiation of lincRNA-p21 is sufficient to enhance the expression of Cdkn1a by creating a scaffold for the recruitment of the Cdkn1a transcriptional co-activator heterogeneous nuclear ribonucleoprotein K (hnRNPK).
Interference with *lincRNA-p21* transcription initiation significantly reduced *Cdkn1a* expression, indicating that the earliest steps of production of nascent *lincRNA-p21* are required for *Cdkn1a cis*-activation. By contrast, premature transcription termination, ribozyme-mediated degradation of *lincRNA-p21* had no effect on *Cdkn1a*, indicating that the mature transcript is dispensable for p21 upregulation



Ferrer and Dimiytrova, 2024. Nature Rev. Mol. Cell Biol.

LncRNAs as chromatin organizers

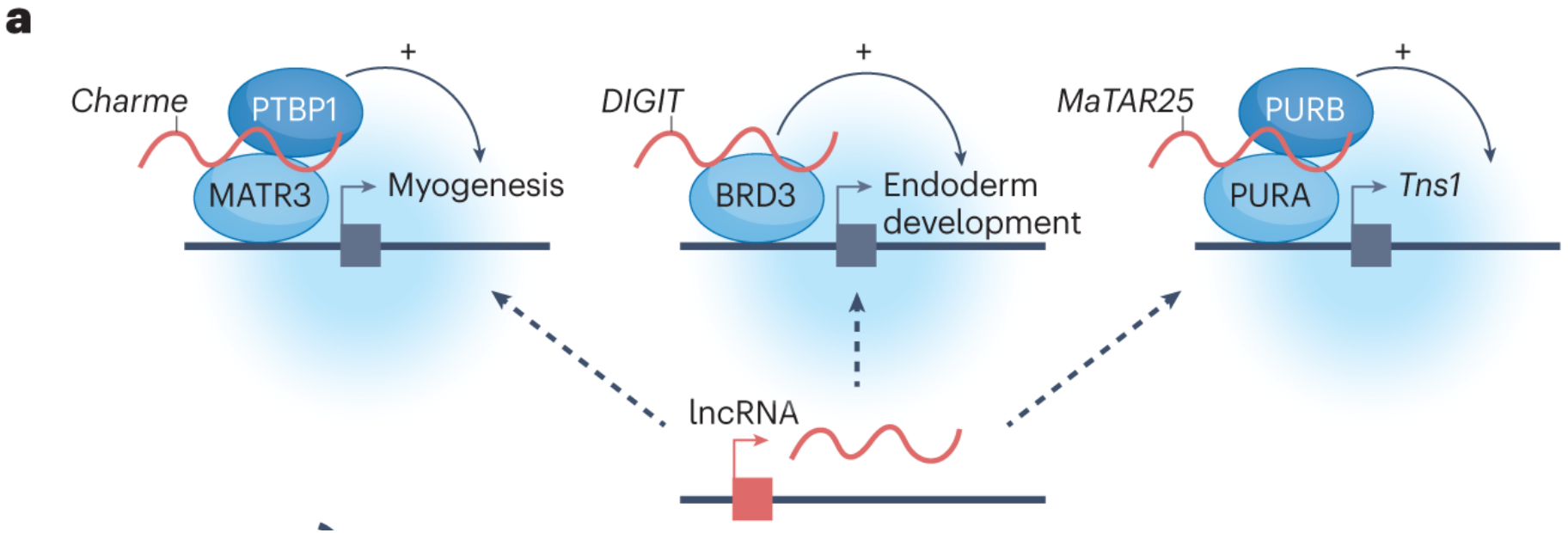
Transcription of a lncRNA in the *Bcl11b* locus named thymocyte differentiation factor (*ThymoD*) promotes the demethylation of CTCF-binding sites and, therefore, CTCF recruitment and chromatin reorganization. This process brings the *ThymoD*-associated enhancer region in proximity with their target, the promoter *of Bcl11b*, resulting in transcription activation



Ferrer and Dimiytrova, 2024. Nature Rev. Mol. Cell Biol.

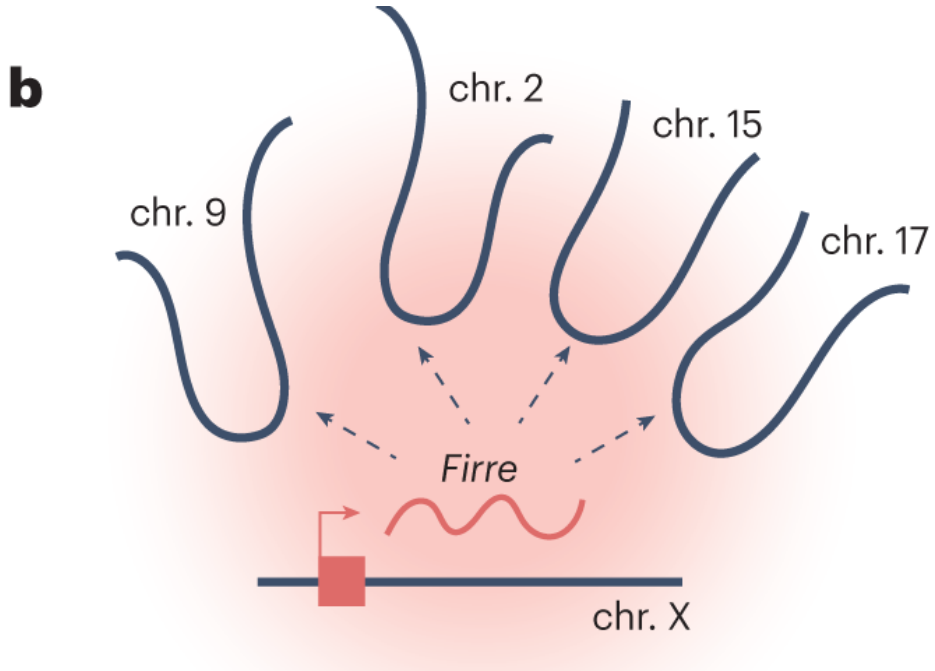
Mechanisms of transcription regulation by *trans*- acting lncRNAs

Association of the lncRNAs *Charme*, *DIGIT* and mammary tumour- associated RNA 25 (*MaTAR25*) with polypyrimidine tract-binding protein 1 (PTBP1)–matrin 3 (MATR3), bromodomain-containing protein 3 (BRD3) and purine-rich element-binding protein A (PURA)–PURB, respectively, drives condensate formation (blue background) and localization at target genes. This localization promotes the activation of broad developmental or disease- associated transcription programmes



Mechanisms of transcription regulation by *trans*- acting lncRNAs

The lncRNA *Firre* promotes inter-chromosomal contacts, which facilitates the co-regulation of a genes with shared functions in energy metabolism.

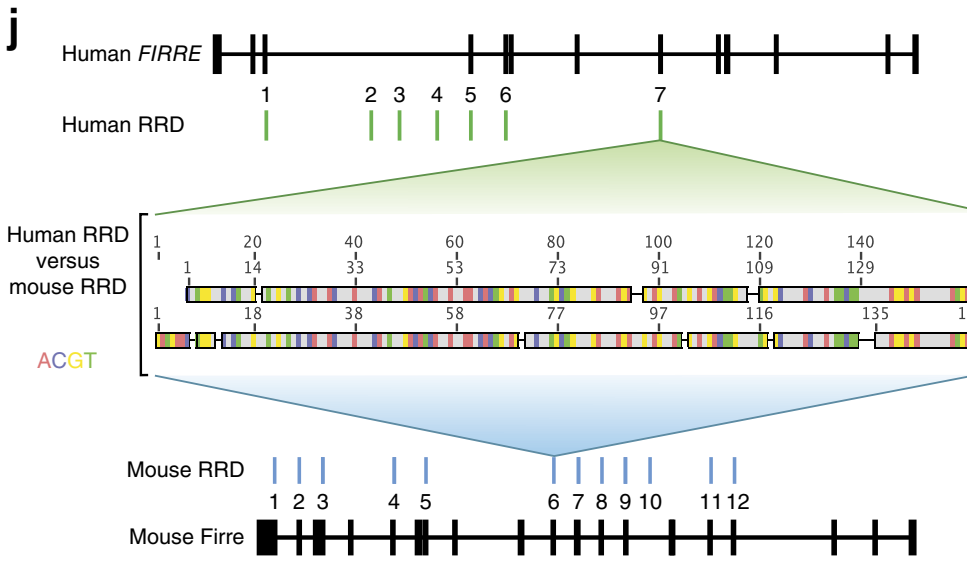


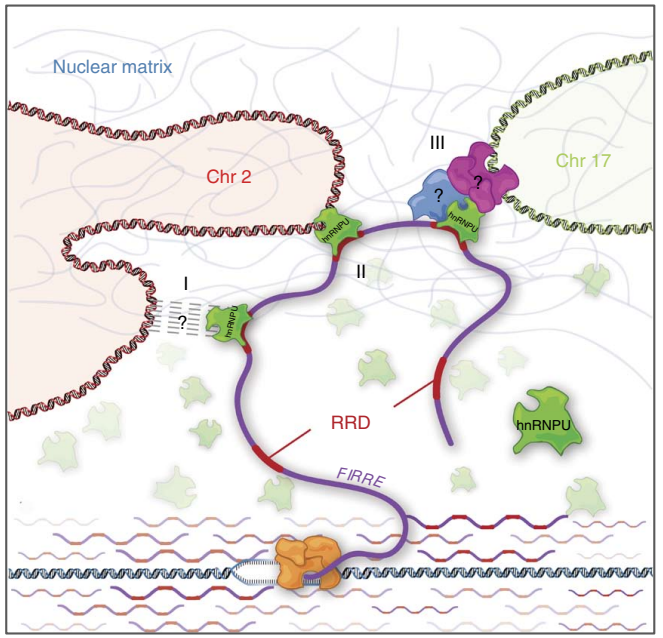
The lncRNA Firre

Firre is expressed in the mouse neural crest tissue.

It has a human ortholog located on the X chromosome.

It contains a 156-bp repeat domain (RRD) that is repeated 16 times in mice and 8 times in humans; this domain shows 96% conservation within species and 68% conservation between species.
By interacting with nuclear matrix proteins, including HNRNPU, it is able to influence the topological organization of multiple chromosomes. It brings together genes involved in the same biological processes—such as energy metabolism and adipogenesis—thereby enabling their spatial and temporal co-regulation.

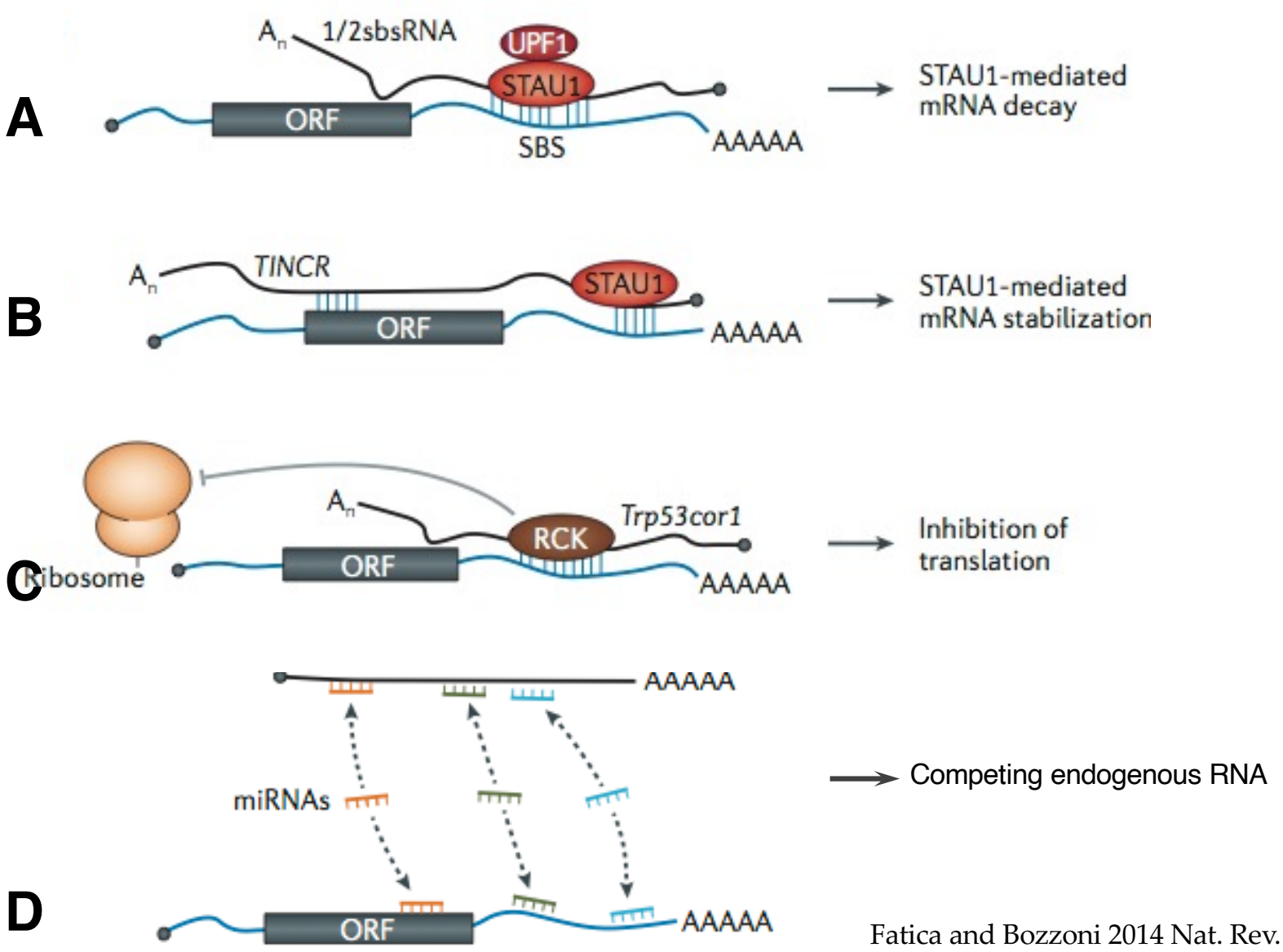




Hacisuleyman et al., NSMB 2014

Cytoplasmic lincRNA

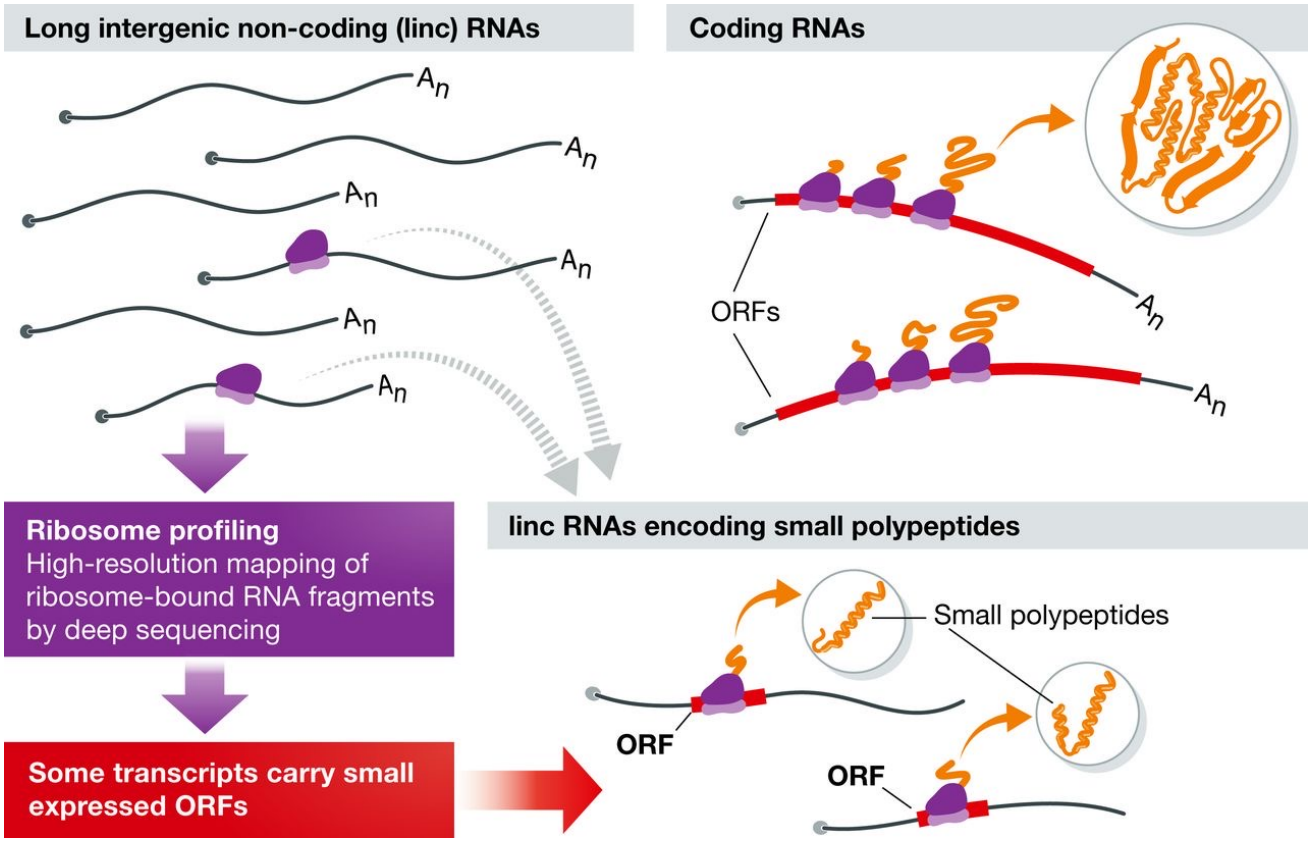
Many lincRNA-mediated mechanisms of gene regulation have been identified in the cytoplasm. These lincRNAs often show sequence complementarity with regulated transcripts.



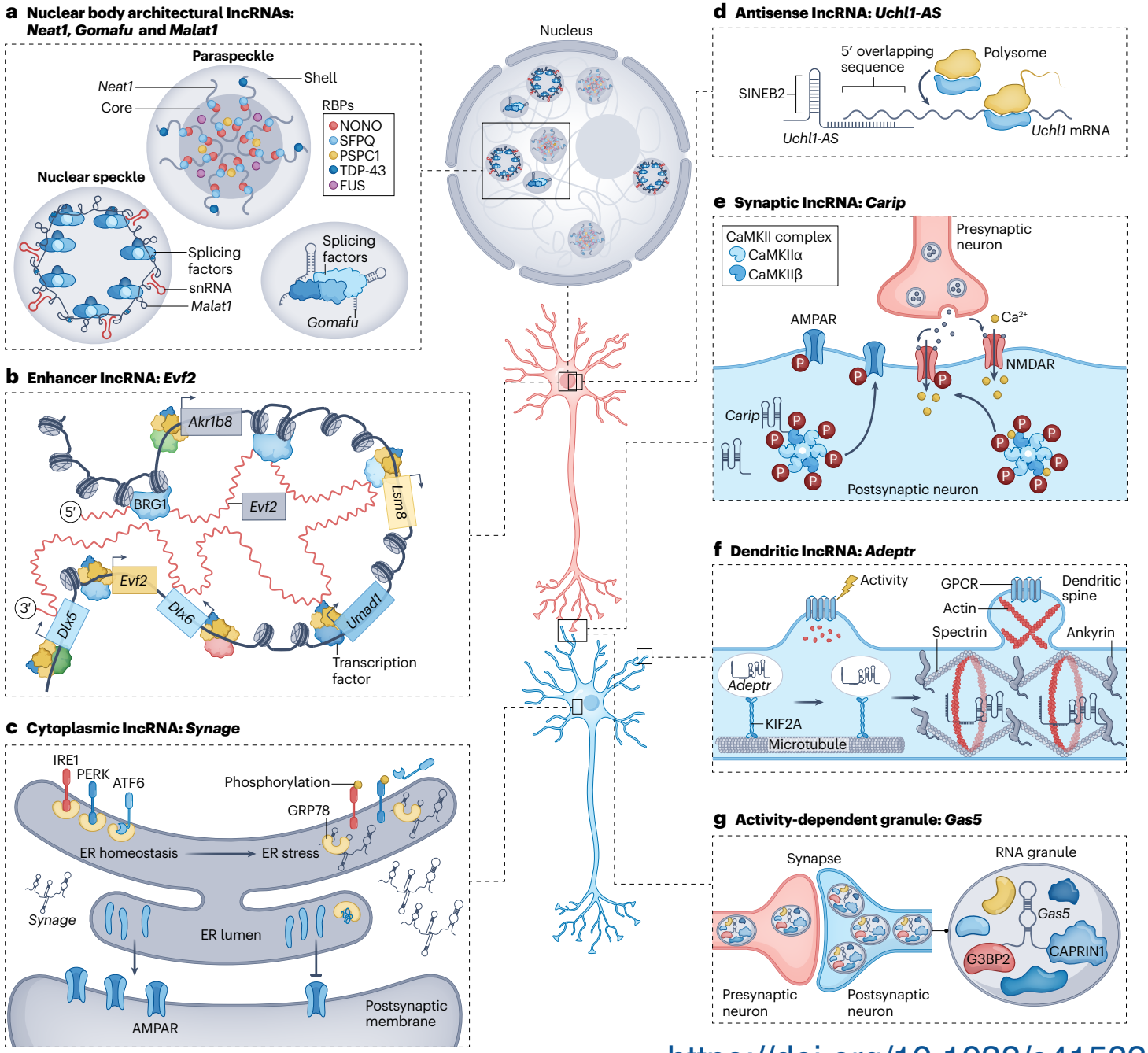
Fatica and Bozzoni 2014 Nat. Rev. Genetics

Several lincRNAs encode for small polypeptide

Recent papers identify lincRNAs that work in a more conventional way—**encoding protein**—in each case a small polypeptide with a biological activity (Magny et al, 2013 Science; Pauli et al, 2014 Science; Bazzini et al, 2014 EMBO J; Anderson et al., 2015 Cell).



lncRNA function in neurons



https://doi.org/10.1038/s41583-025-00960-z

Epiloque

Like all biochemical processes, the transcription machinery is not perfect and can produce spurious RNAs that have no purpose. A sensible hypothesis is that most of the currently annotated long noncoding RNAs are not functional.

Despite general agreement that some long noncoding RNAs are functional and others are not, opinions vary widely as to the fraction that is functional.

However, even a scenario in which only 10% are functional implies the existence of more than a thousand human loci generating noncoding transcripts with biological roles.

And regardless of function, long noncoding RNAs might have diagnostic applications, with changes in their expression already associated with cancer and several disorders.

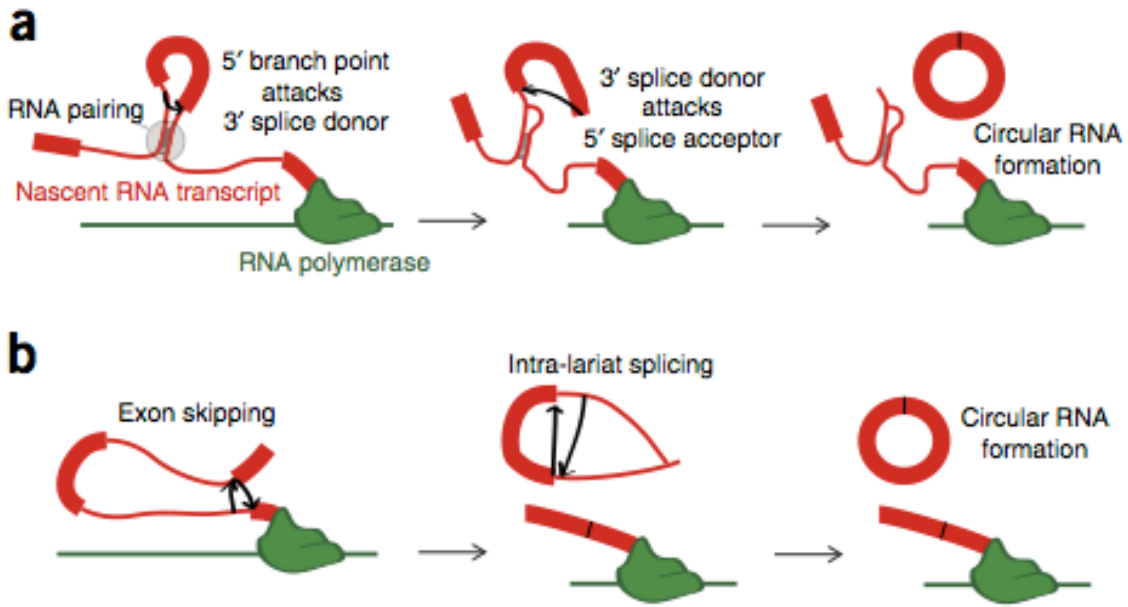
- **Circular RNAs**

Exonic circRNAs

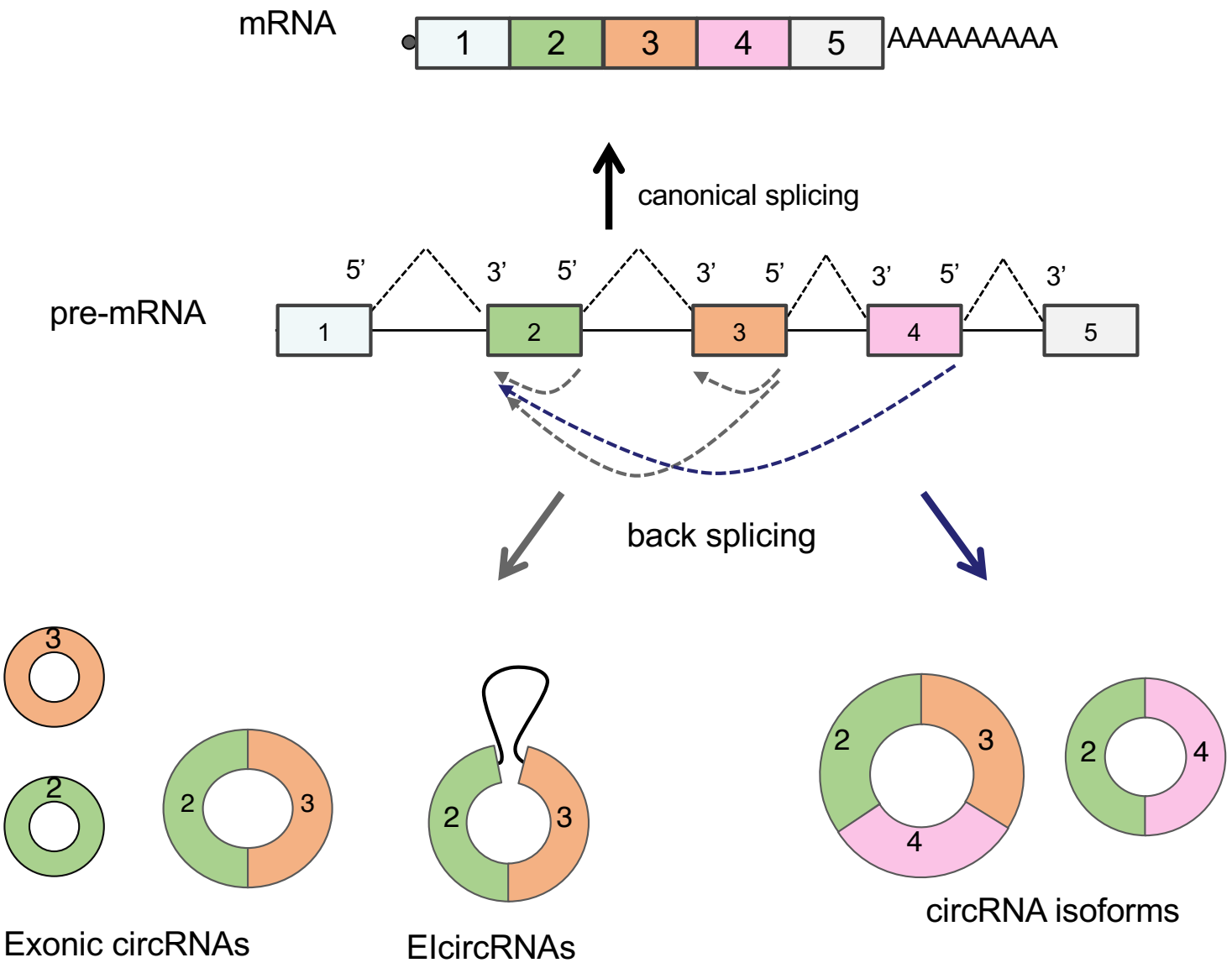
Two mechanisms have been proposed for mammalian exonic circRNA formation:

a.In the first mechanism—'**direct backsplicing**'—a downstream splice donor pairs with an unspliced upstream splice acceptor and the intervening RNA is circularized

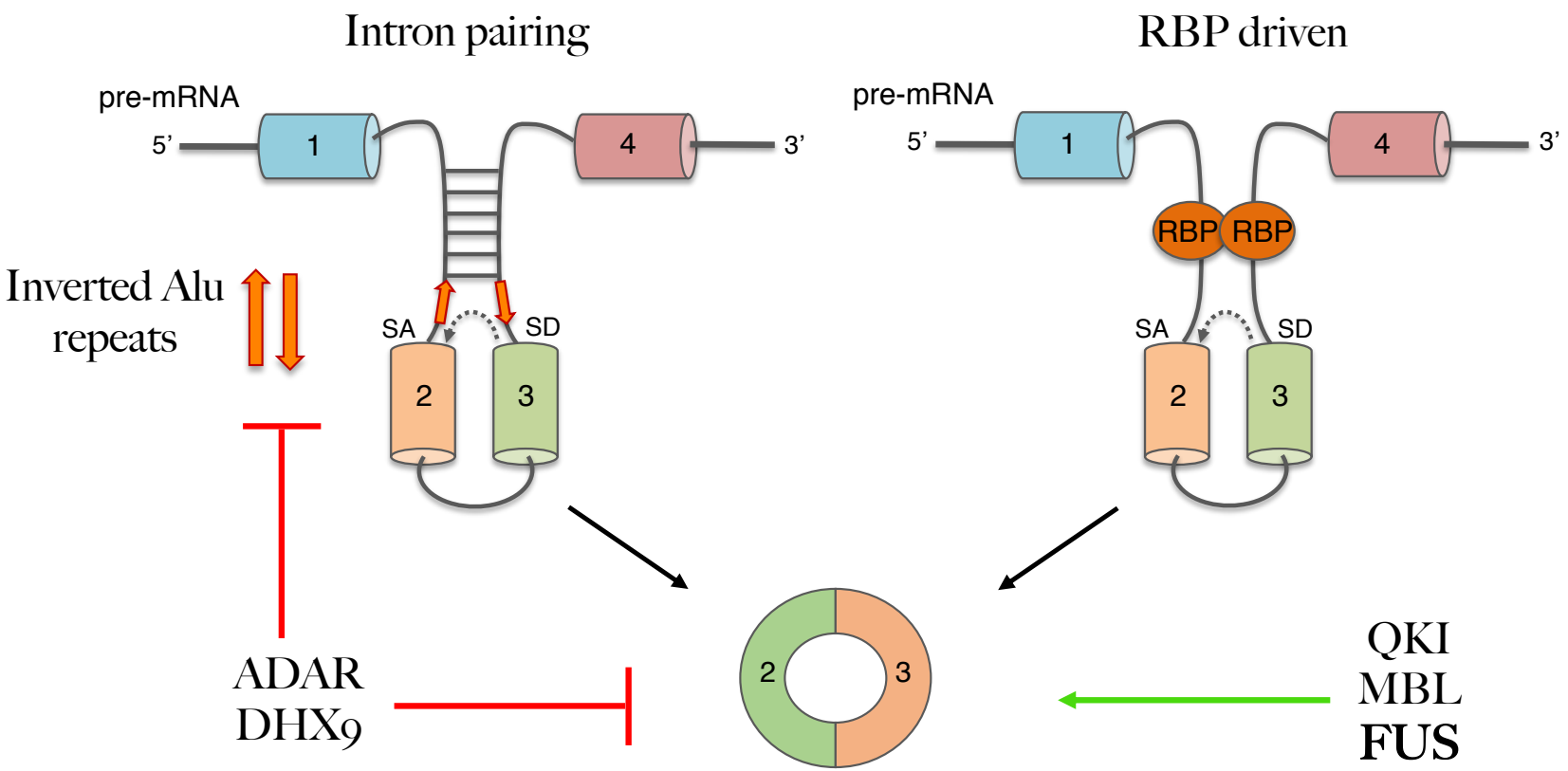
b.The second mechanism—known as the '**lariat intermediate**' or '**exon skipping' mechanism**—involves splicing within lariats produced from exon skipping



circRNAs originate through a back-splicing event



The direct back-splicing mechanisms



Adapted from D'Ambra et al., 2019

Circular RNA are a recent addition to the growing list of types of non-coding RNAs

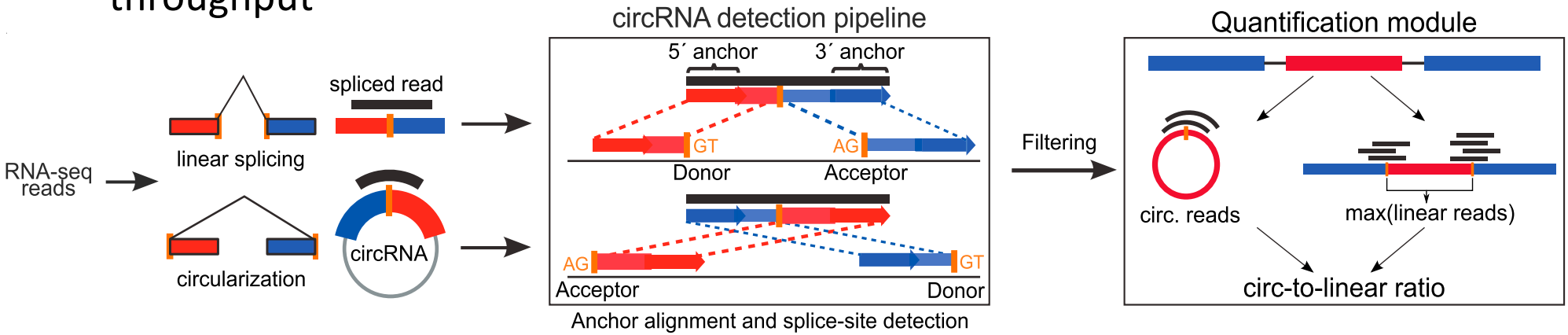
CircRNAs have eluded identification until recently for several reasons:

- Circular RNAs, unlike miRNAs and other small RNAs, are not easily separated from other RNA species by size or electrophoretic mobility.
- Commonly used molecular techniques that require amplification and/or fragmentation destroy circularity
- circRNAs have no free 3’ or 5’ end, they cannot be found by molecular techniques that rely on a polyadenylated free RNA end
- circRNAs have an out-of-order arrangement of exons, known as a ‘backsplice’,and early RNA-seq mapping algorithms filtered out such sequences

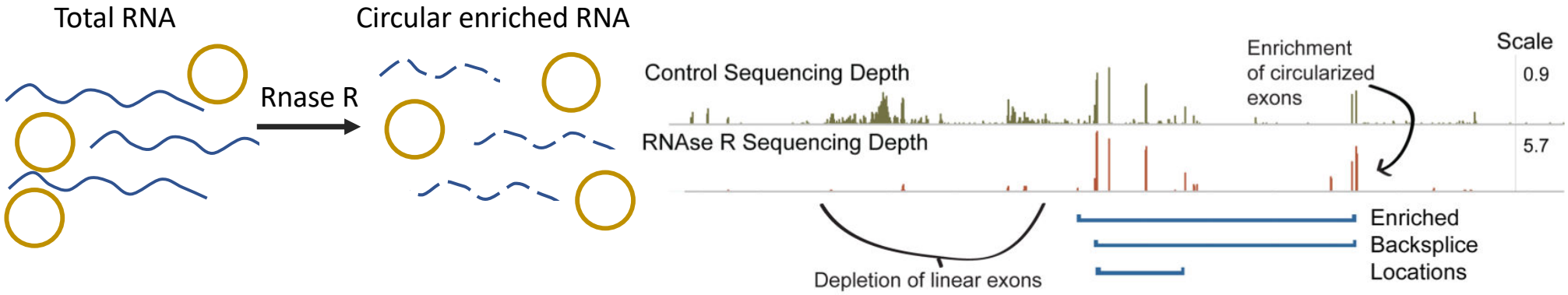
The problems of circRNA detection have recently been addressed

- Sequencing of ribosomal RNA (rRNA)-depleted RNA libraries (rather than polyA-enriched libraries).

- Novel bioinformatic tools, sequencing with longer reads and higher throughput



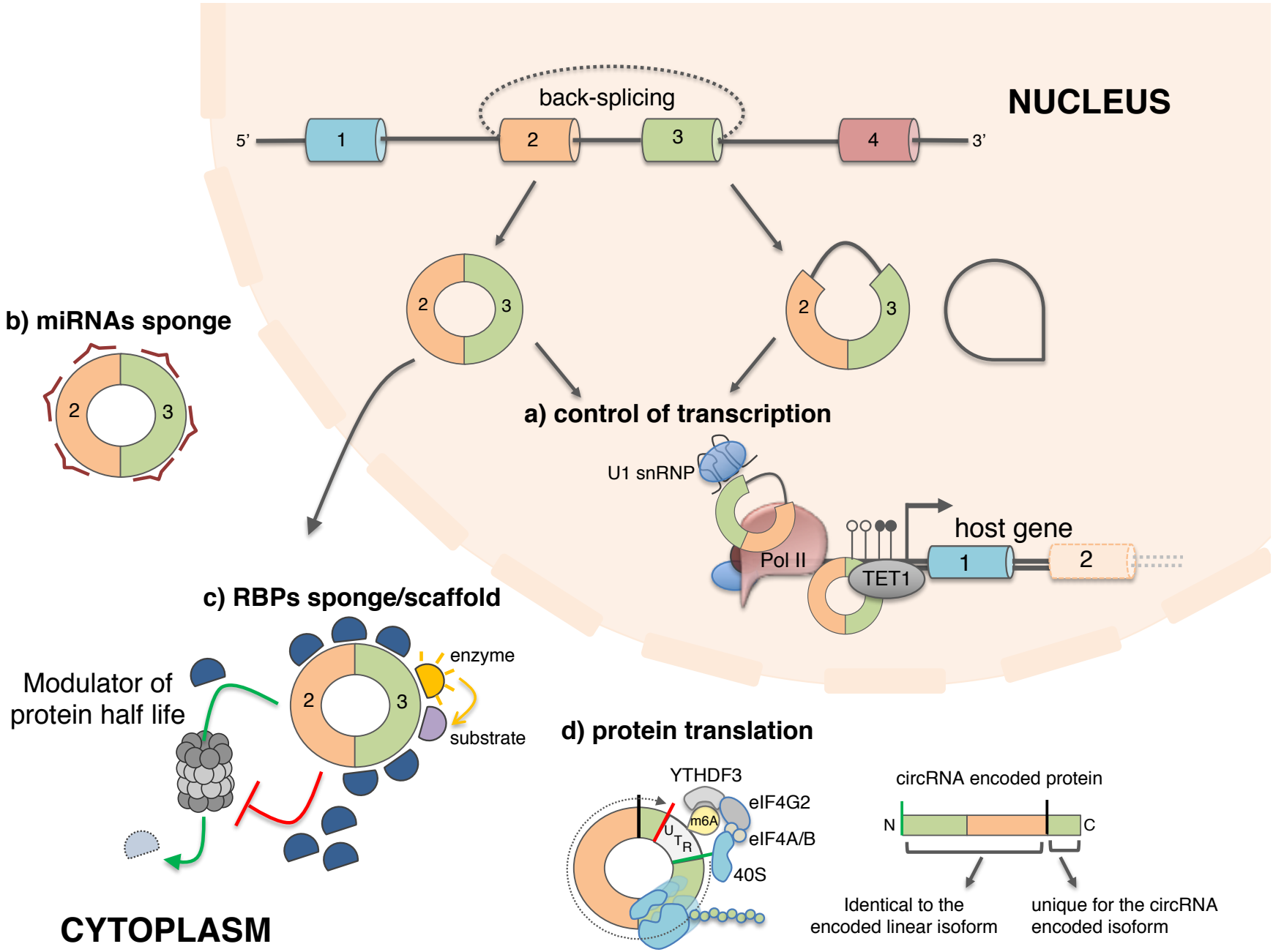
- Exonuclease-based enrichment approaches



These studies revealed that:

- over **10% of expressed genes** in examined cells and tissues can produce circRNAs
- circRNAs are **evolutionary conserved** and expressed in a time-, cell type- and gene-specific manner.
- most circRNAs are lowly abundant, but some that are ubiquitously expressed are present at higher copy numbers (>10-fold) as compared to their linear transcripts
- circRNAs are detectable in blood and other peripheral tissues in association with endosomes and microvesicles, thus representing excellent candidates as disease biomarkers

Functions of circRNAs



D'Ambra et al., 2019