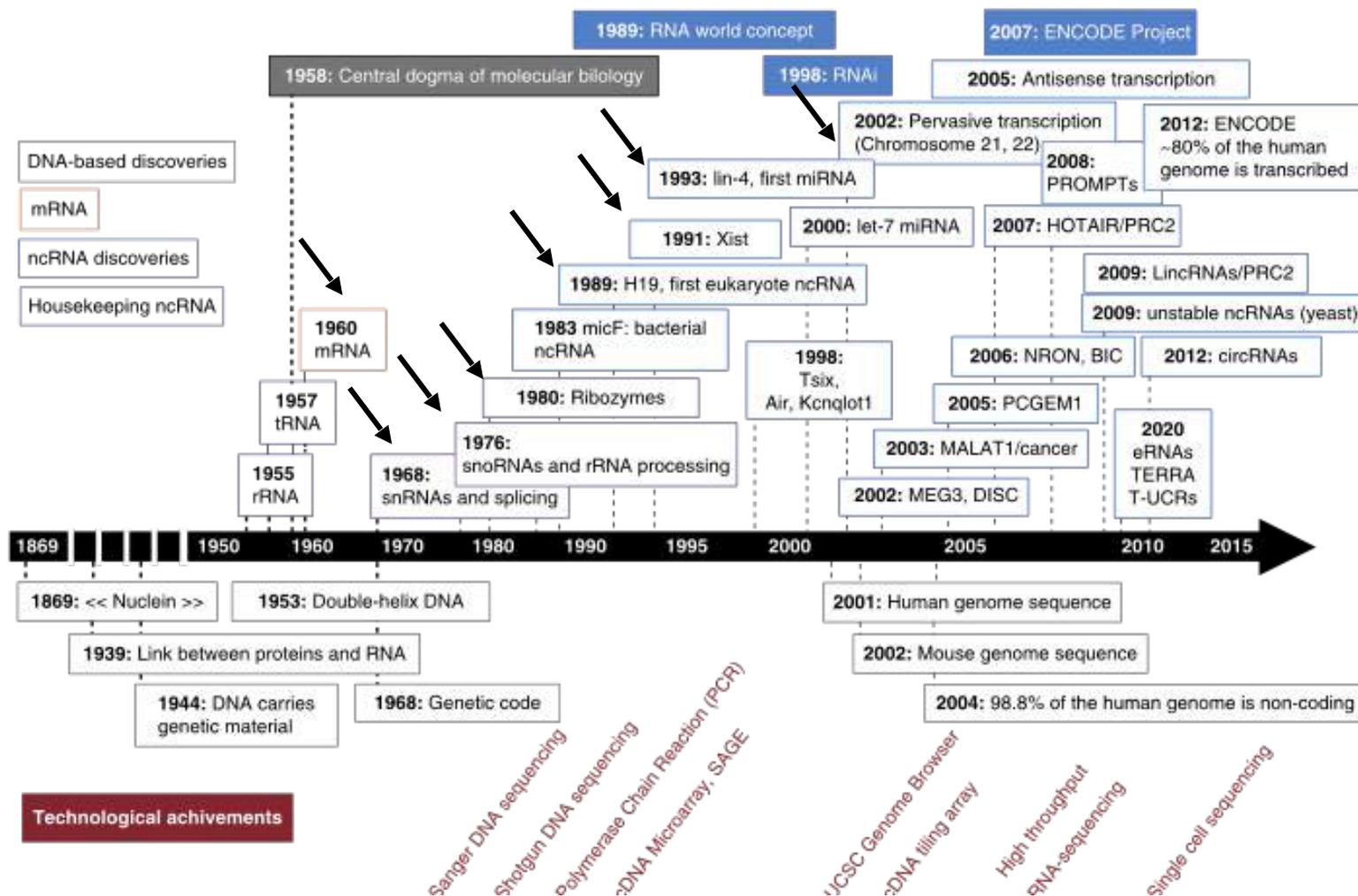


Χρονοδιάγραμμα των σημαντικότερων ανακαλύψεων της βιολογίας των ριβονουκλεϊκών οξέων, στους ευκαρυώτες



1 History, Discovery, and Classification of lncRNAs

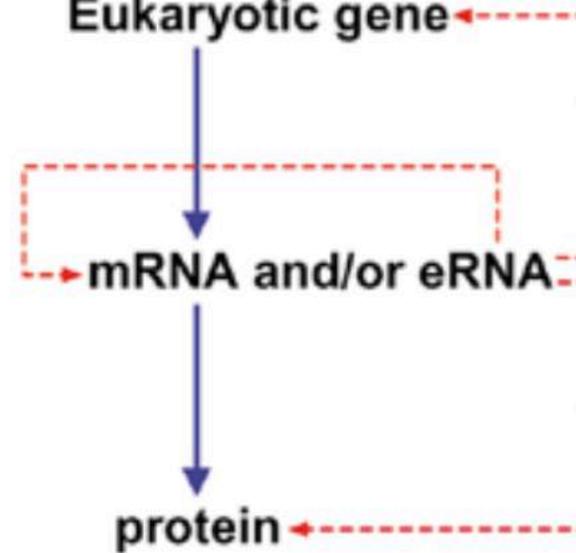
*Single output
Simple system*

Prokaryotic gene



*Multiplex output
Parallel processing*

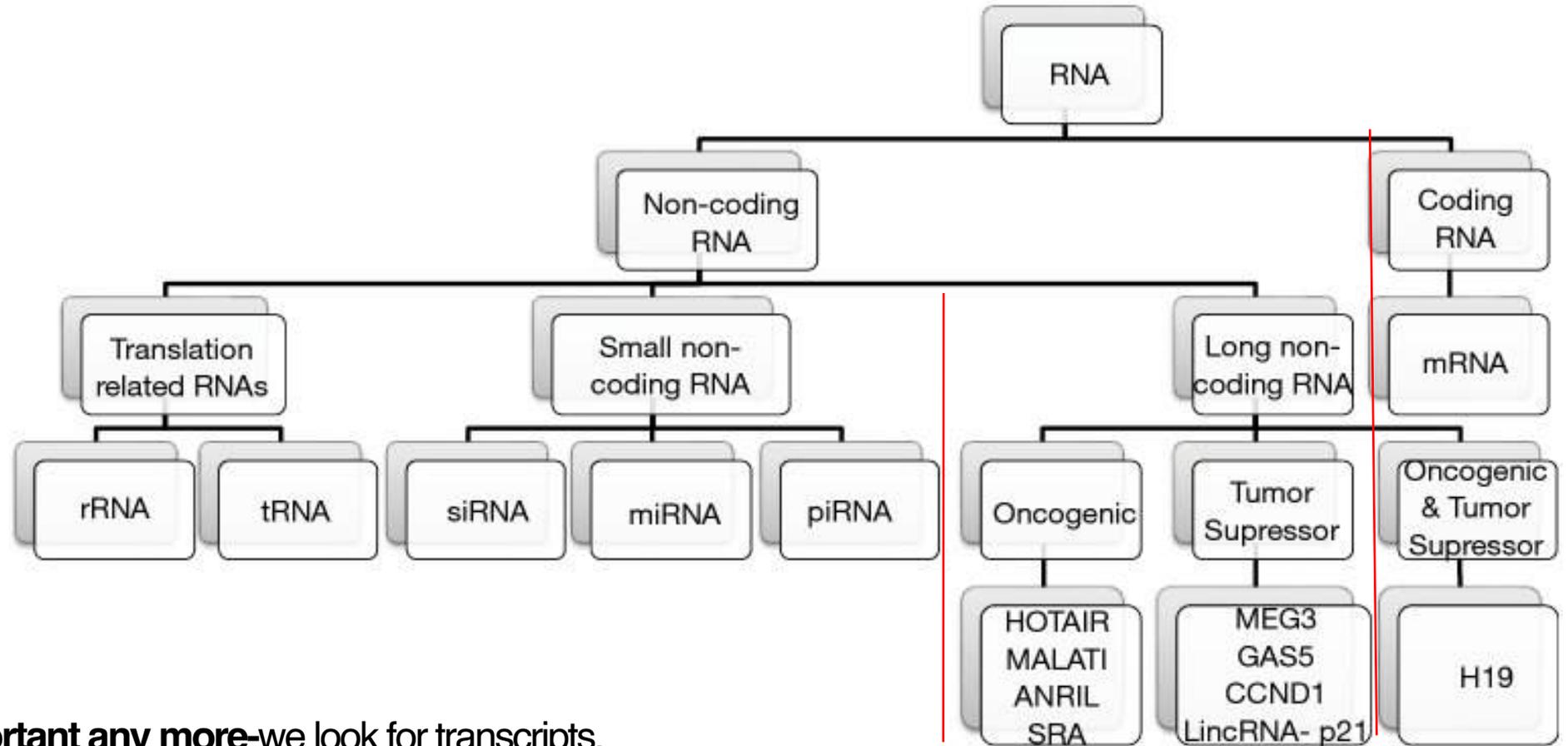
Eukaryotic gene



Replace eRNA
with any
regulatory
RNA

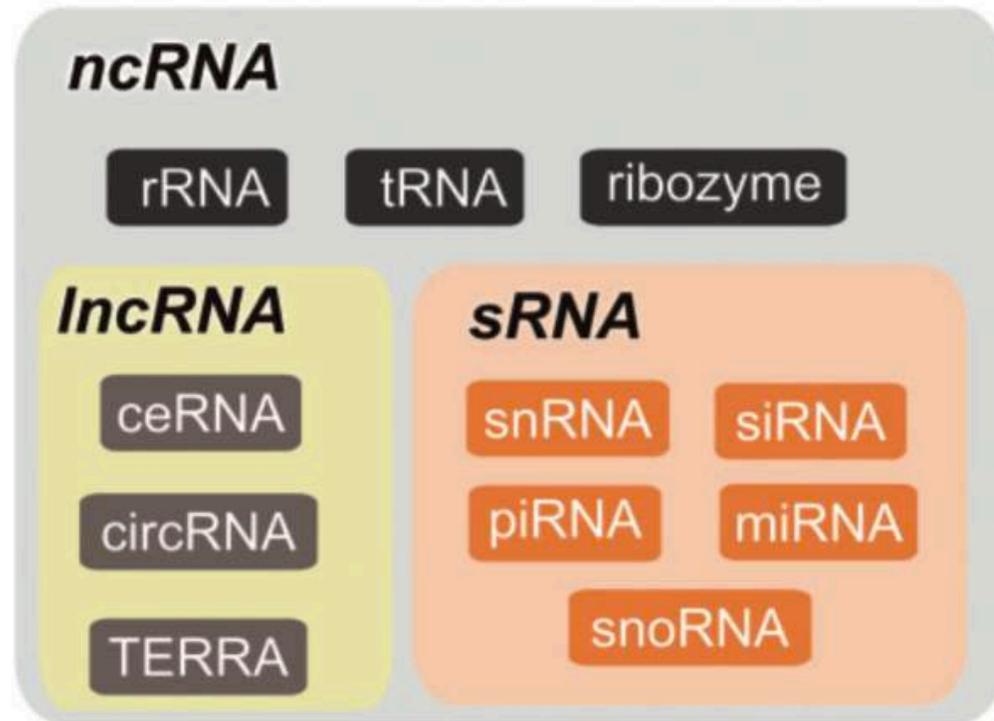
- Comparison of the prokaryotic and proposed eukaryotic genetic operating systems. The left panel shows the central dogma in which genes code, via mRNA, for proteins, which carry out the catalytic, structural, signal transduction and regulatory functions of the cell. The right panel shows the proposed operating system in eukaryotes wherein genes may express two levels of information: mRNA for proteins, and eRNAs that carry out concomitant networking and other functions within the organism. Thus there are three types of genes in eukaryotes: those that encode only protein (which are rare), those that encode only eRNA, and those that encode both.

The brave new world of RNAs



- **Genes are not so important any more**-we look for transcripts.
- The rise in RNA-mediated regulation could be concomitant with the evolution of genome complexity.
- **Stress/inflammation and cancer promotes RNA accumulation**
- **The importance of being RNA in stress**, when lowering the energetic cost of induction of stress response is highly desirable.
- The resulting stressed-induced senescence and stress-associated secretory phenotype (pro-inflammatory, pro-cancerous state) besides accumulating RNA and misfolded proteins, **secretes extracellular vesicles to promote or contain systemic stress, that contain a multitude of ncRNAs.**

ncRNAs



Διακυτταρική επικοινωνία

Αναδιοργάνωση του πυρήνα

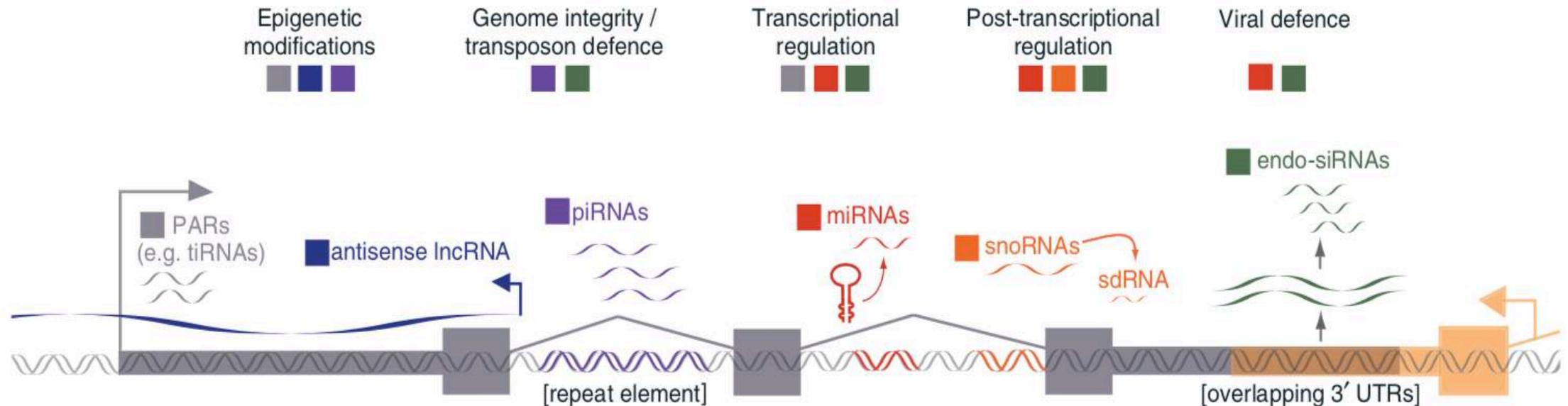
Φύλαξη του γονιδιώματος και μετα-μεταγραφική καταστολή του γονιδιώματος

Επιγενετική, στρες

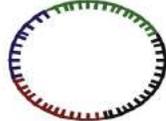
Σχηματισμός ογανιδίων χωρίς μεμβράνη

Απεικόνιση των ρόλων των ρυθμιστικών ncRNAs

- PARs, promoter-associated RNAs; lncRNAs, long non-coding RNAs; miRNAs, microRNAs; snoRNAs, small nucleolar RNAs; sdRNAs, sno-derived RNAs; endo-siRNAs, endogenous siRNAs; piRNAs, PIWI-interacting RNAs; tiRNAs, transcription initiation RNAs



Οι κύριες κατηγορίες των ρυθμιστικών ncRNAs

(A) miRNA	(B) lncRNA	(C) circRNA
 <ul style="list-style-type: none">- Produced as longer transcripts and processed to 22 bp- mature miRNA are contained in the RNA-induced silencing complex (RISC)- Interact with mRNA resulting in post-transcriptional gene regulation	 <ul style="list-style-type: none">- Nonprotein coding transcripts- Longer than 200 bp- Form complex secondary structures- Can be post-translationally processed, (e.g., through splicing or 5' capping)- Interact with RNA, DNA, and proteins	 <ul style="list-style-type: none">- Formed by back-splicing of pre-mRNA- Can contain multiple exons- Stable due to exonuclease resistance- Interact with mRNA, miRNA, proteins, and DNA

Μακρά μη-κωδικοποιά RNAs (long ncRNAs, lncRNAs)

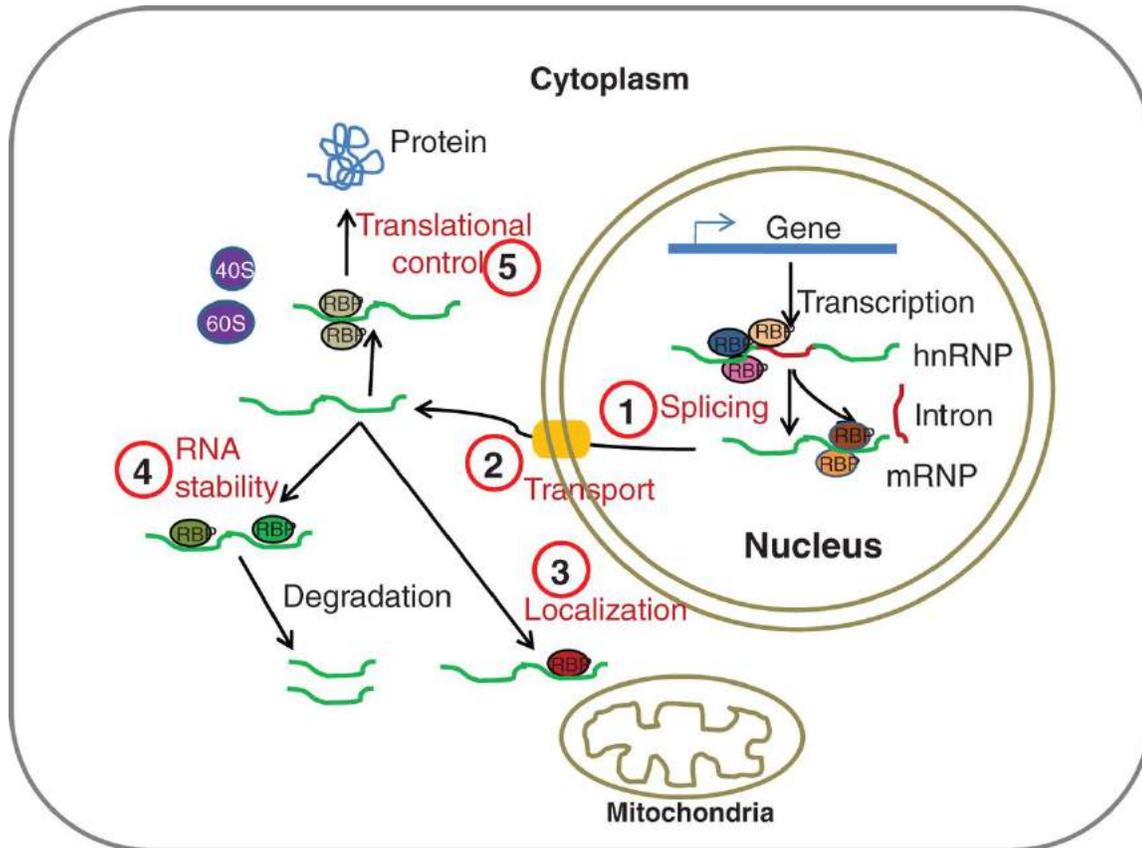
- Τα μακρά μη-κωδικοποιά RNAs (lncRNAs) αντιπροσωπεύουν την κυρίαρχη και λειτουργικά πιο ποικιλόμορφη κατηγορία μη-κωδικοποιών, (>200nts σε μήκος)
- Κάποια από τα τα lncRNAs είναι ψευδογονίδια τα οποία εξαιτίας μεταλλάξεων έχουν μετατραπεί σε μη-κωδικοποιά. Όπως το PTENP1
- Εκτός από πολύ-λειτουργικά, επηρεάζουν πολλαπλά επίπεδα της γονιδιακής έκφρασης τα lncRNAs είναι και πολύ-επίπεδα
- Αλληλοεπιδρούν πέρα από RNA/DNA, με πρωτεΐνες, λιπίδια και πιθανώς και με άλλα βιοπολυμερή (όπως σάκχαρα).
- Είναι τα πιο πρόσφατα εξελιγμένα γονίδια (παρουσιάζουν το λιγότερο βαθμό συντήρησης) και έχουν και αυτά (όπως τα miRNAs) ιστο-ειδικό ή σταδιο-ειδικό προφίλ έκφρασης.
- Αν και στο παρελθόν θεωρούνταν μεταγραφικός «θόρυβος» ή τεχνήματα από την αναδιαμόρφωση της χρωματίνης, πλέον θεωρούμε ότι είναι γονίδια όπως τα πρωτεϊνικά γονίδια.

Long non-coding RNAs (LncRNAs)

Newly evolved low-copy mRNA-like transcripts (>200nts in size) that cover as a class a very large fraction of the human genome.

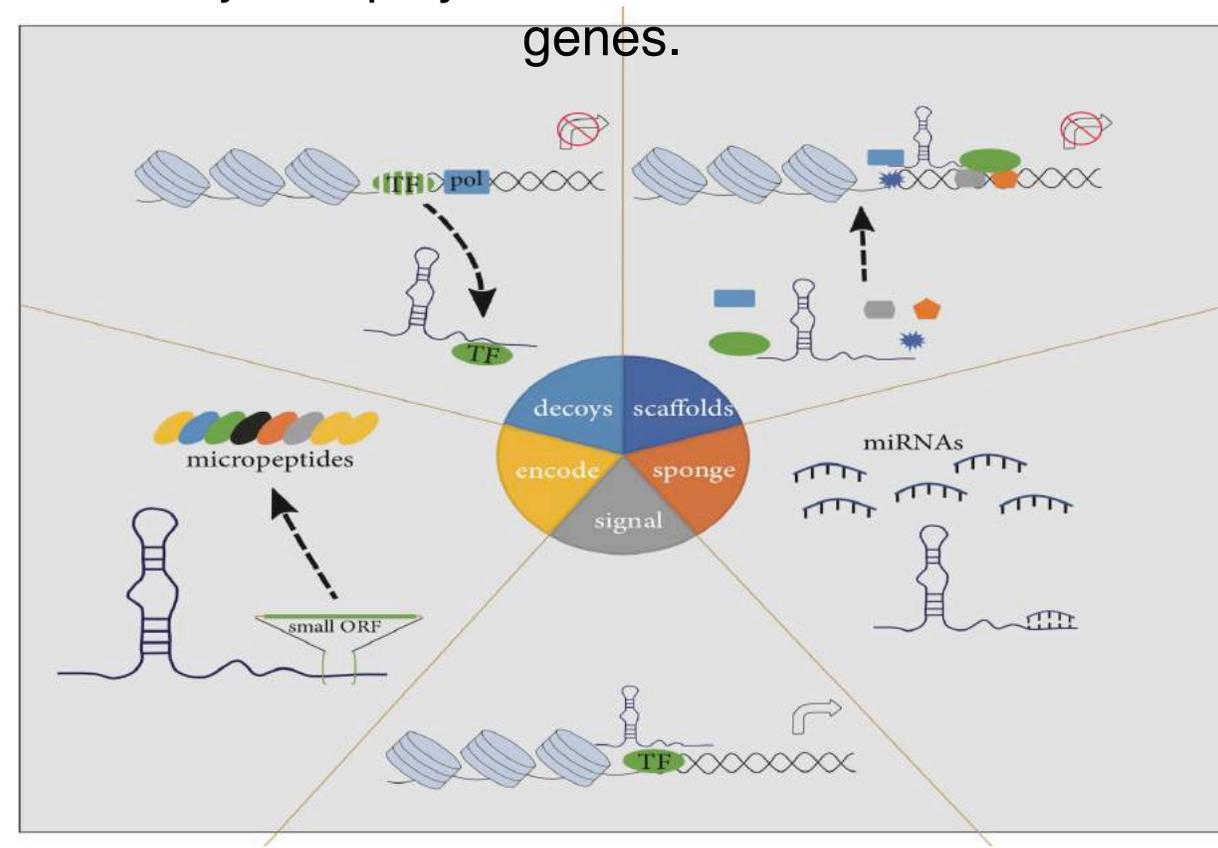
Multi-functional (like miRs)

Fine-tuners of all steps of gene expression



Multi-dimensional

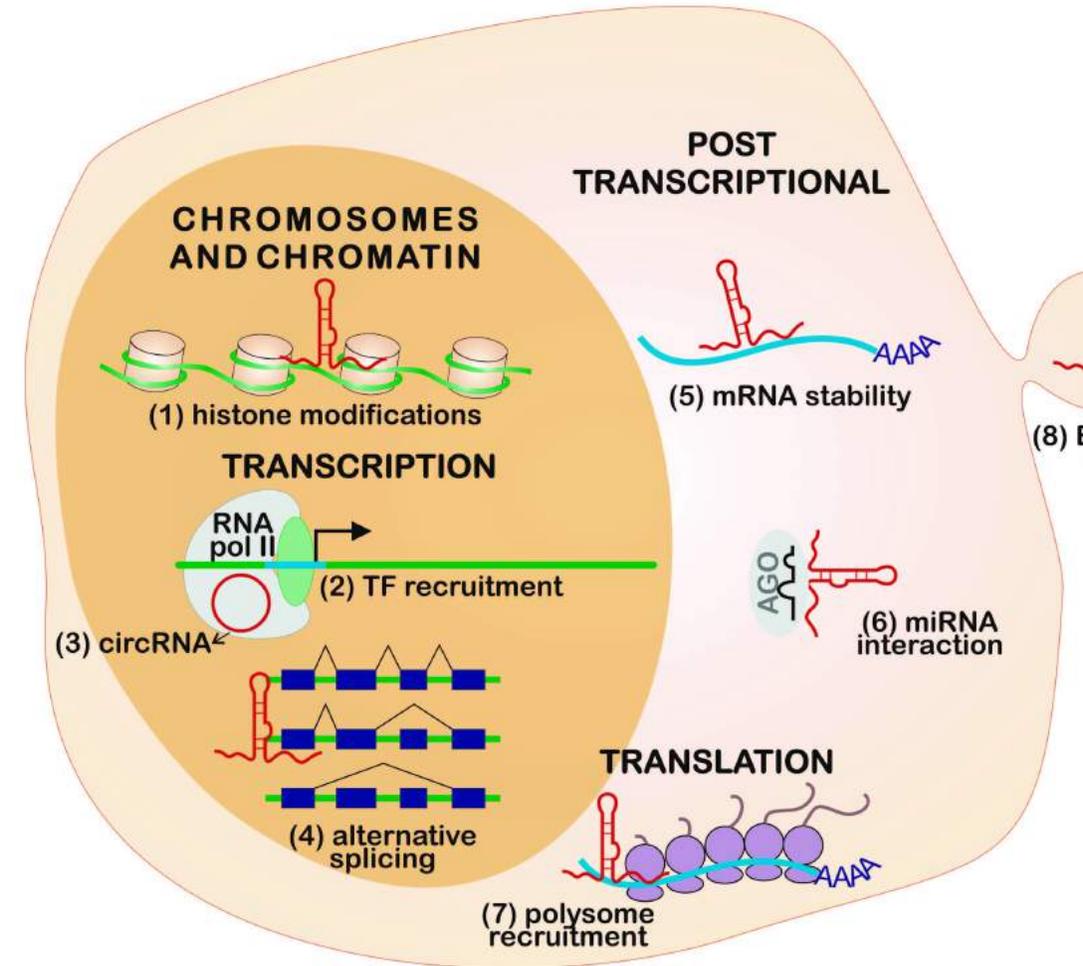
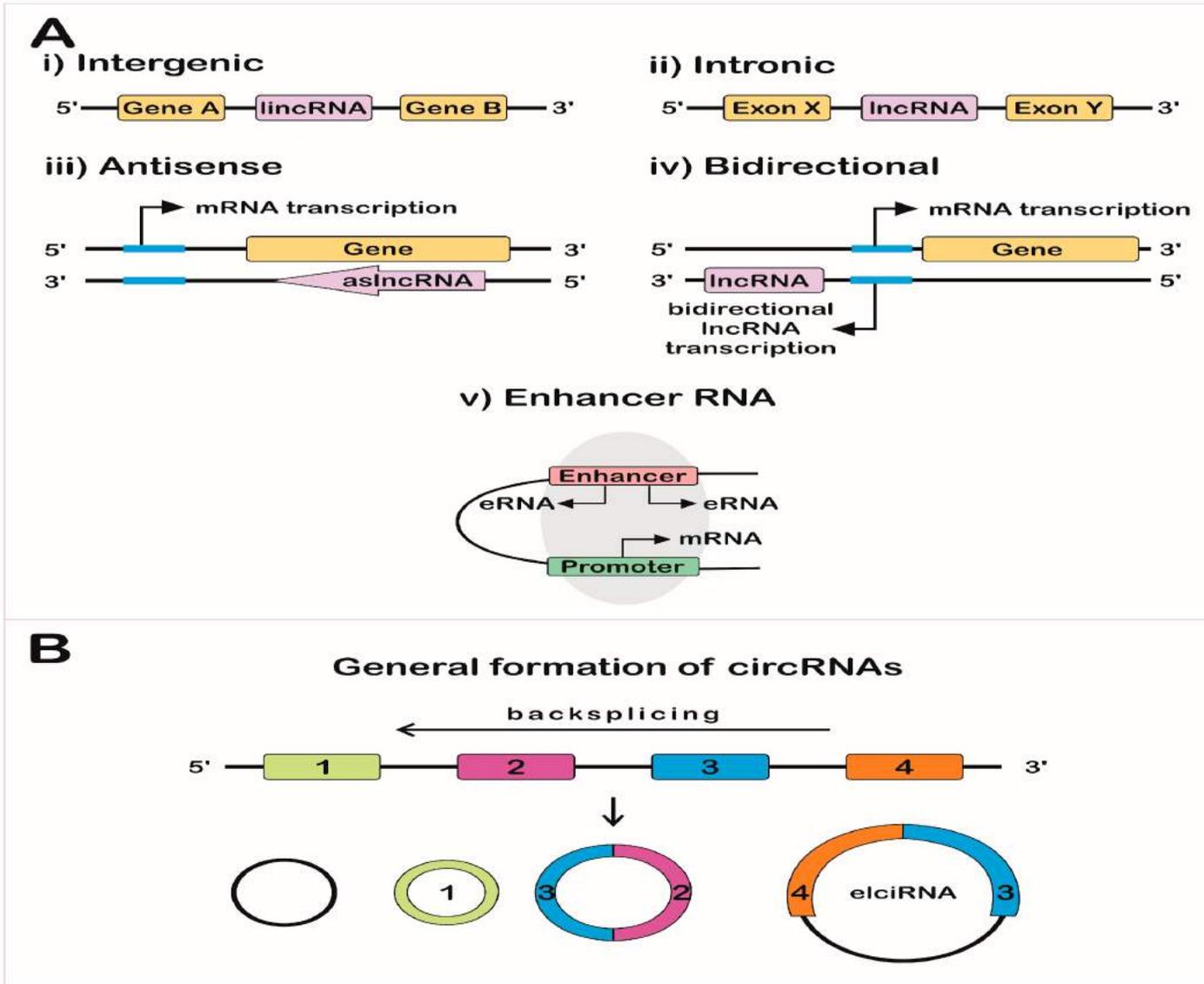
Regulators, modifiers, sensors, scaffolds of all major biopolymers and source of de-novo genes.



Λειτουργικοί ρόλοι των lncRNAs

- Υπολογίζεται ότι τα lncRNAs ανέρχονται από 20,000 έως και 60,000–100,000 γονίδια στον άνθρωπο. Παρόλα αυτά, μόνο περίπου 200 από αυτά είναι καλά μελετημένα.
- Η έκφραση τους είναι εξαιρετικά ιστο-ειδική ή σταδιο-ειδική έκφραση (περισσότερο από τα μικροRNAs) και αυτό καταδεικνύει πολύ σημαντικό λειτουργικό ρόλο στην ομοιόσταση, στην κυτταρική απάντηση του στρες και στις ασθένειες του ανθρώπου, όπως τον καρκίνο, τα νευροεκφυλιστικά νοσήματα, αυτό-άνοσες ασθένειες και τα μεταβολικά νοσήματα.
- Τα lncRNA είναι εκτός από πολύ-λειτουργικά και πολύ-επίπεδα στη φύση τους εργαλεία ρύθμισης κυτταρικών διεργασιών.
 - Επιγενετική επίδραση μέσα από πρόσδεση με σύμπλοκα τροποποίησης της χρωματίνης ή πρόσδεση με ρυθμιστικά στοιχεία του DNA για το έλεγχο της έναρξης της μεταγραφής.
 - Ρύθμιση της γονιδιωματικής ακεραιότητας
 - Μετα-μεταγραφική ρύθμιση της γονιδιακής έκφρασης.
 - Ρύθμιση της πρωτεϊνοσύνθεσης
 - Ρύθμιση μεταβολισμού
- Τα μακρά μη-κωδικοποιά RNAs μπορούν να προσλαμβάνουν δευτεροταγείς δομές και επομένως να προσδένουν ή να αλληλοεπιδρούν με ειδικό τρόπο με ένα εύρος βιοπολυμερών (DNA, RNA, πρωτεΐνες, λιπίδια).
- Τα επίπεδα έκφρασης των lncRNAs μπορεί να είναι χαμηλά αλλά μπορούν να μεταβληθούν πολύ αποτελεσματικά για τη ρύθμιση της γονιδιακής έκφρασης.

Λειτουργικοί ρόλοι των lncRNAs

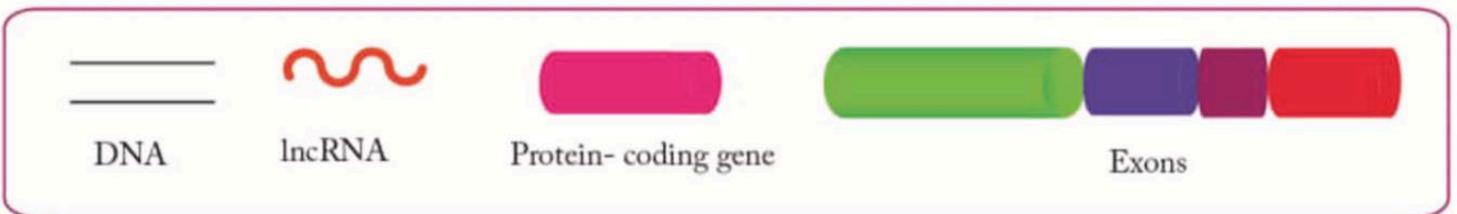
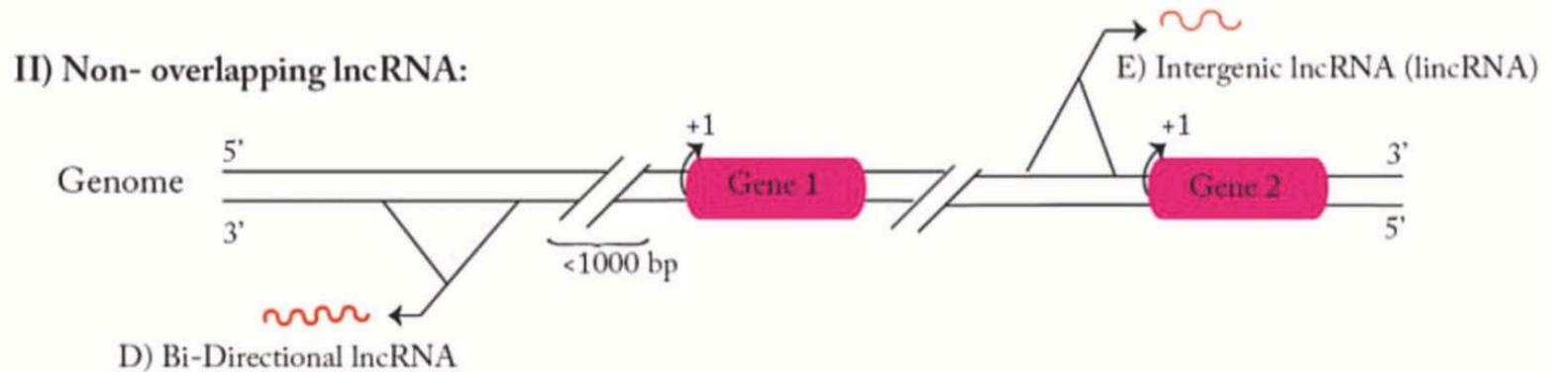
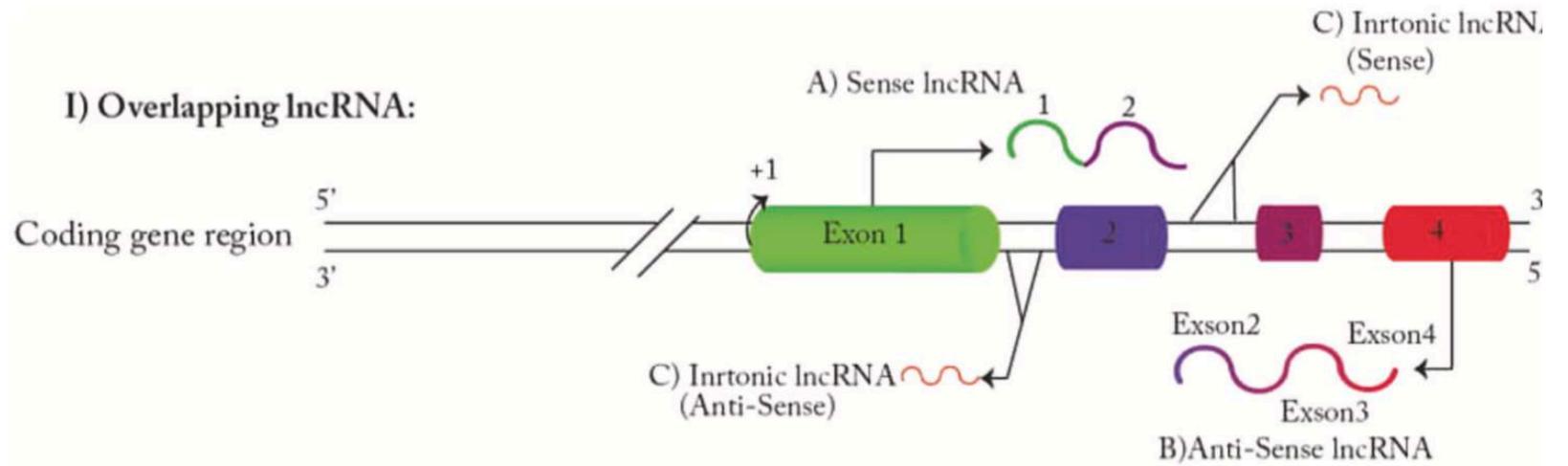


Κατηγοριοποίηση των lncRNAs

- Τα lncRNAs μπορούν να κατηγοριοποιηθούν σε διάφορες ομάδες σύμφωνα με το γονιδιωματικό τόπο και γονιδιακή δομή.
- Πολλά lncRNA αλληλεπικαλύπτουν, εν μέρη ή πλήρως, πρωτεϊνικά γονίδια. Υπολογίζεται γύρω στο 20% του μεταγραφώματος στον άνθρωπο σχηματίζοντας νοσηματικά-αντινοσηματικά ζεύγη (sense-antisense pairs).
- Τα περισσότερα όμως lncRNAs εντοπίζονται έξω από τα σώματα των κωδικοποιών γονιδίων.

Αλληλεπικαλυπτόμενα vs. Μη-αλληλεπικαλυπτόμενα

- Τα lncRNAs διαχωρίζονται λειτουργικά σε cis-δραστικά και trans-δραστικά lncRNAs. Τα cis-δραστικά lncRNAs λειτουργούν στο ίδιο χρωμόσωμα από το οποίο παράγονται και αν ρυθμίσουν τη έκφραση ενός αριθμού γονιδίου τοπικά. Τα trans-δραστικά lncRNAs ρυθμίζουν γονίδια που βρίσκονται σε διαφορετικά χρωμοσώματα.



Αλληλεπικαλυπτόμενα (overlapping) lncRNA

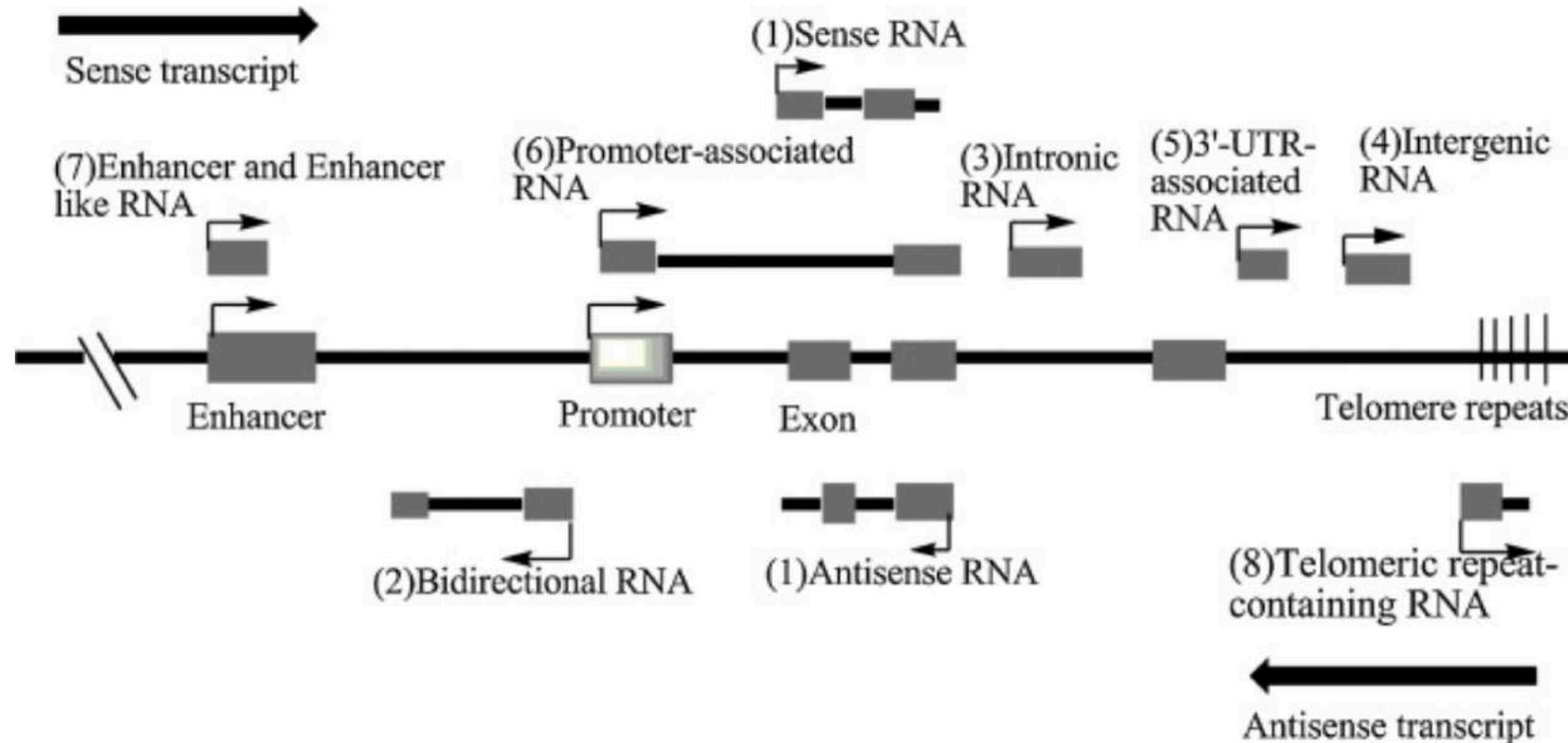
Τα αλληλεπικαλυπτόμενα lncRNAs ομαδοποιούνται σε τρεις κατηγορίες:

- Τα νοηματικά lncRNAs (sense lncRNAs). Αυτά τα lncRNA γονίδια αλληλεπικαλύπτουν ένα ή περισσότερα εξώνια πρωτεϊνικών γονιδίων και μεταγράφονται με την ίδια κατεύθυνση όπως τα πρωτεϊνικά γονίδια. Παραδείγματα είναι τα Gas5 και το MALAT1.
- Τα αντινοηματικά lncRNAs (antisense lncRNAs) αλληλεπικαλύπτουν ένα ή περισσότερα εξώνια κωδικοποιών γονιδίων με αντίθετη μεταγραφική. Κλασσικό παράδειγμα αποτελεί το HOTAIR.
- Τα ιντρονικά lncRNAs (intronic lncRNAs) εντοπίζονται σε ιντρόνια πρωτεϊνικών γονιδίων (όπως το COLDAIR). Ειδική κατηγορία των ιντρονικών lncRNAs είναι τα κυκλικά lncRNAs (circular lncRNAs) τα οποία λόγω δομής είναι εξαιρετικά ανθεκτικά εναντίον εξωνουκλεασών.

Μη-αλληλεπικαλυπτόμενα (non-overlapping) lncRNAs

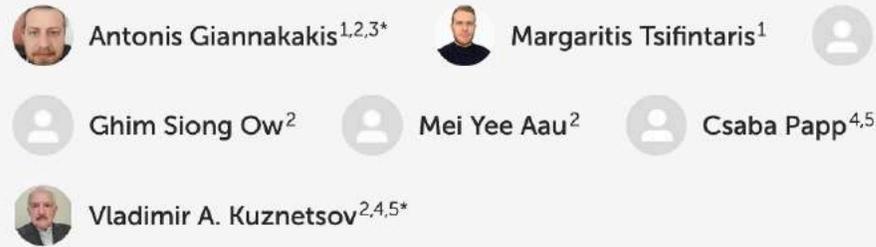
- Τα αμφίδρομα (bidirectional) lncRNAs που εντοπίζονται σε κοντινή απόσταση από τα κωδικοποιά γονίδια αλλά στην αντινοσηματική φορά. Συνήθως εντοπίζονται σε ενισχυτές (σχετιζόμενα με ενισχυτές, enhancer-associated lncRNAs, eRNAs) και υποκινητές κωδικοποιών γονιδίων.
- Τα σχετιζόμενα με υποκινητές lncRNAs (promoter-associated lncRNAs) και σχετιζόμενα με το μεταγραφικό τέλος των κωδικοποιών γονιδίων lncRNAs (terminal-associated lncRNAs) τα οποία μπορούν να μεταγράφονται είτε στη νοσηματική είτε στην αντινοσηματική κατεύθυνση.
- Τα διαγονιδιακά lncRNAs (long intergenic ncRNAs, lincRNAs), που το κάθε άκρο του βρίσκεται σε μακρινές αποστάσεις (>5 Kbp) από κωδικοποιά γονίδιο. Κλασικά παραδείγματα είναι τα H19, Xist και lincRNA-p21.

Γονιδιωματική κατηγοριοποίηση των lncRNAs

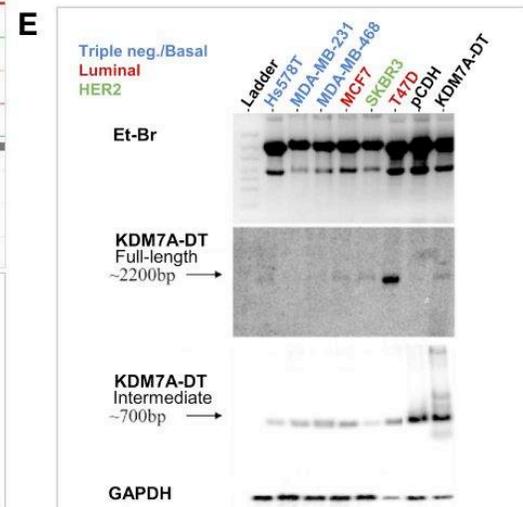
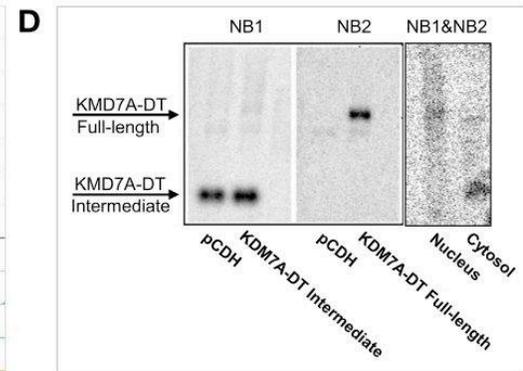
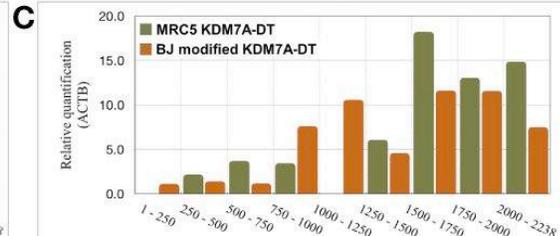
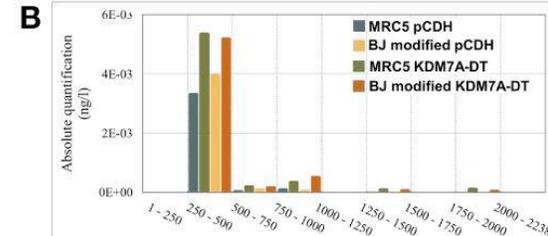


Η κατηγοριοποίηση με βάση το περιεχόμενο δεν παρέχει απαραίτητα πληροφορία για τη λειτουργία των lncRNAs πέρα από ενδεχόμενο να λειτουργούν *in cis* or *in trans*

KDM7A-DT induces genotoxic stress, tumorigenesis, and progression of p53 missense mutation-associated invasive breast cancer

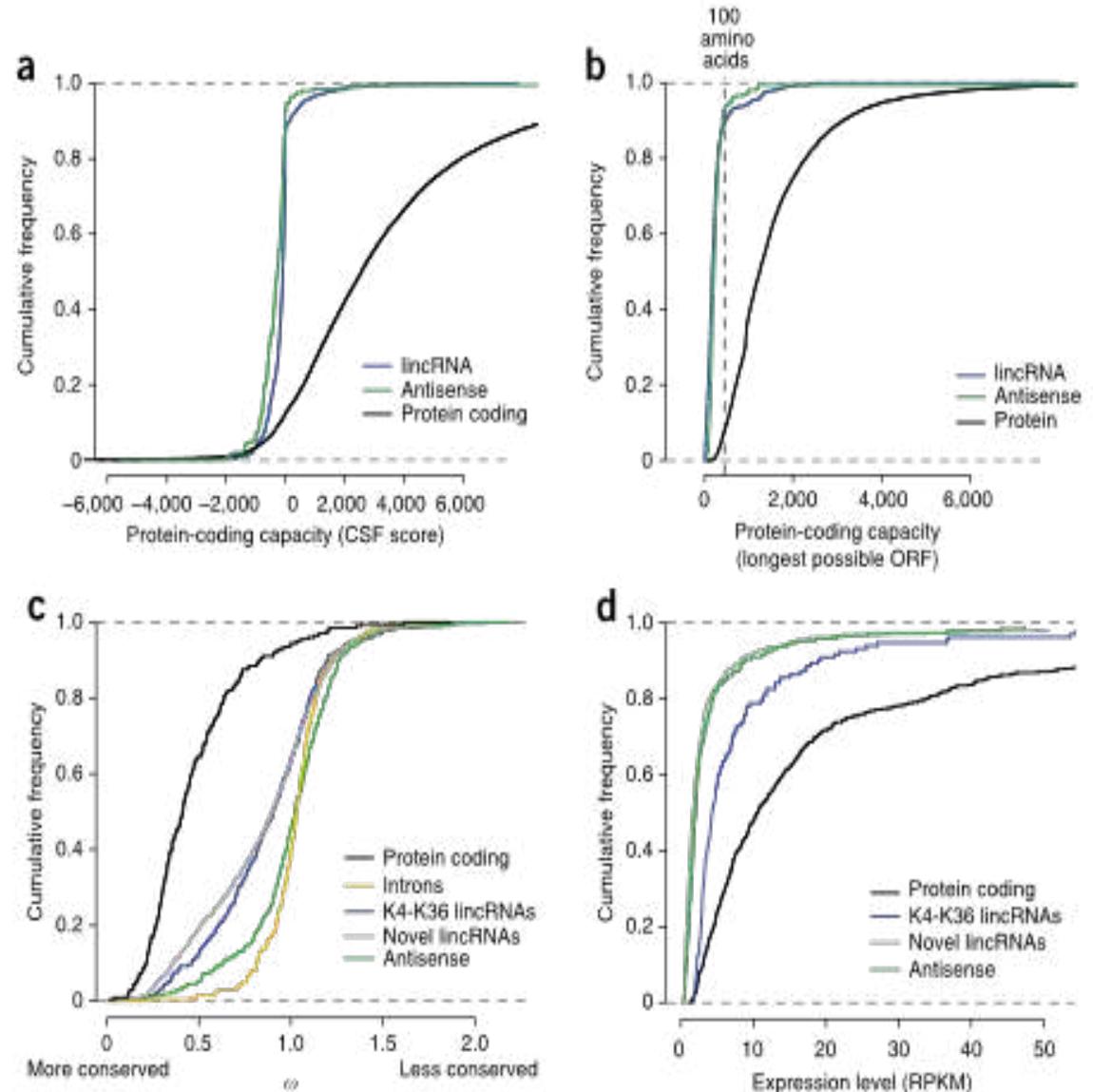


¹ Department of Molecular Biology and Genetics, Democritus University of Thrace,
² Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR),
³ University Research Institute for the Study of Genetic & Malignant Disorders in Children, University of Athens, Athens, Greece
⁴ Department of Urology, The State University of New York (SUNY) Upstate Medical
⁵ Department of Biochemistry and Molecular Biology, The State University of New York, Syracuse, NY, United States



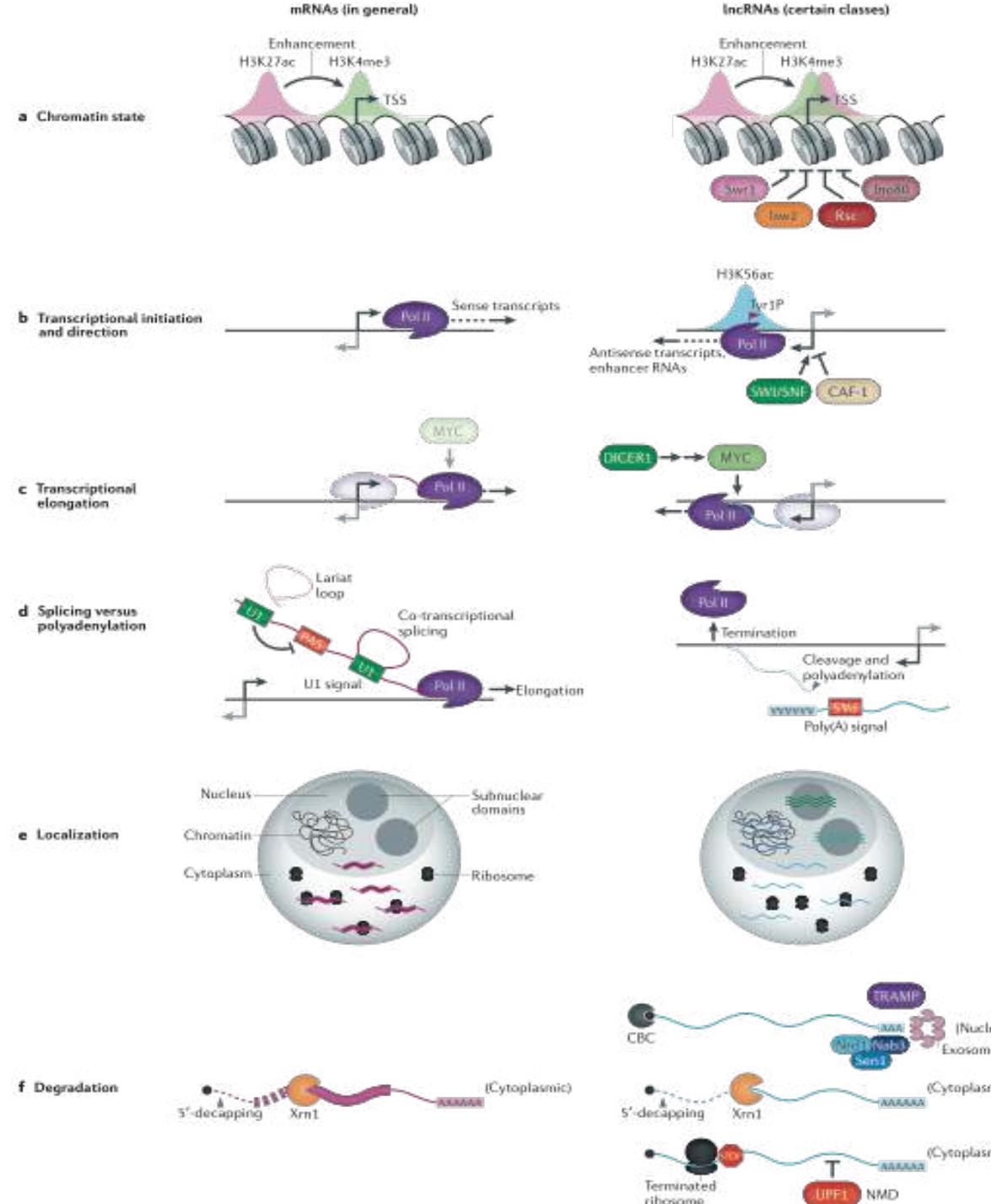
Χαρακτηριστικά μακρών μη-κωδικοποιών RNAs

- Σύντομα (100-300nt) ή ανύπαρκτα ανοιχτά πλαίσια ανάγνωσης
- Μη συντηρημένα ή συντηρημένα αλλιώς (δευτεροταγής και τριτοταγής δομή, π.χ. tRNA).
- Κατά μέσο όρο 10 φορές λιγότερα επίπεδα έκφρασης σε σχέση με τα κωδικοποιά.
- Εξαιρέσεις lincRNAs που είναι συντηρημένα (NORAD) ή και εκφράζονται σε αφθονία (MALAT1, NEAT1)



Ομοιότητες/Διαφορές mRNAs και lncRNAs

- Like mRNAs, many lncRNAs are transcribed by RNA polymerase II (Pol II) from genomic loci with similar chromatin states to mRNAs
- They are often 5' capped, spliced and polyadenylated;
- They expressed at relatively low levels and exhibit poorer primary sequence conservation
- Located in all sub-cellular compartments
- In most instances, they lack any biochemical distinction from mRNAs besides the absence of a translated ORF.

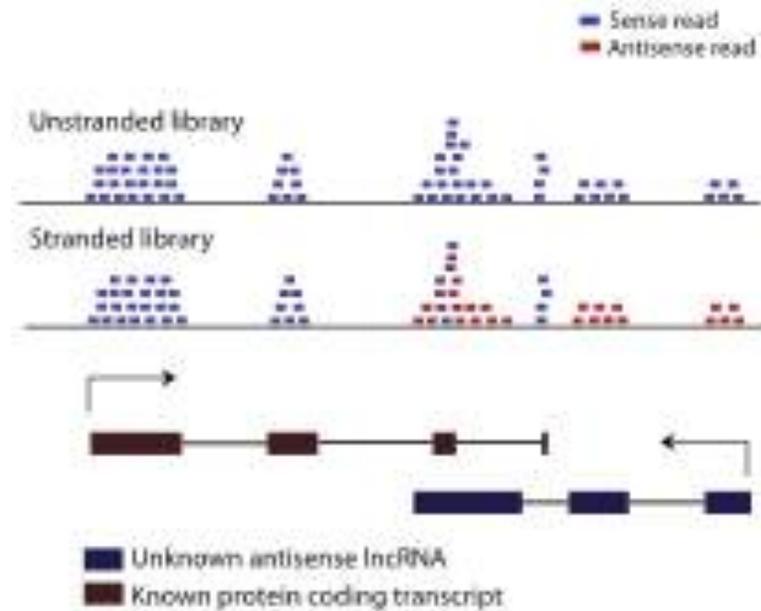
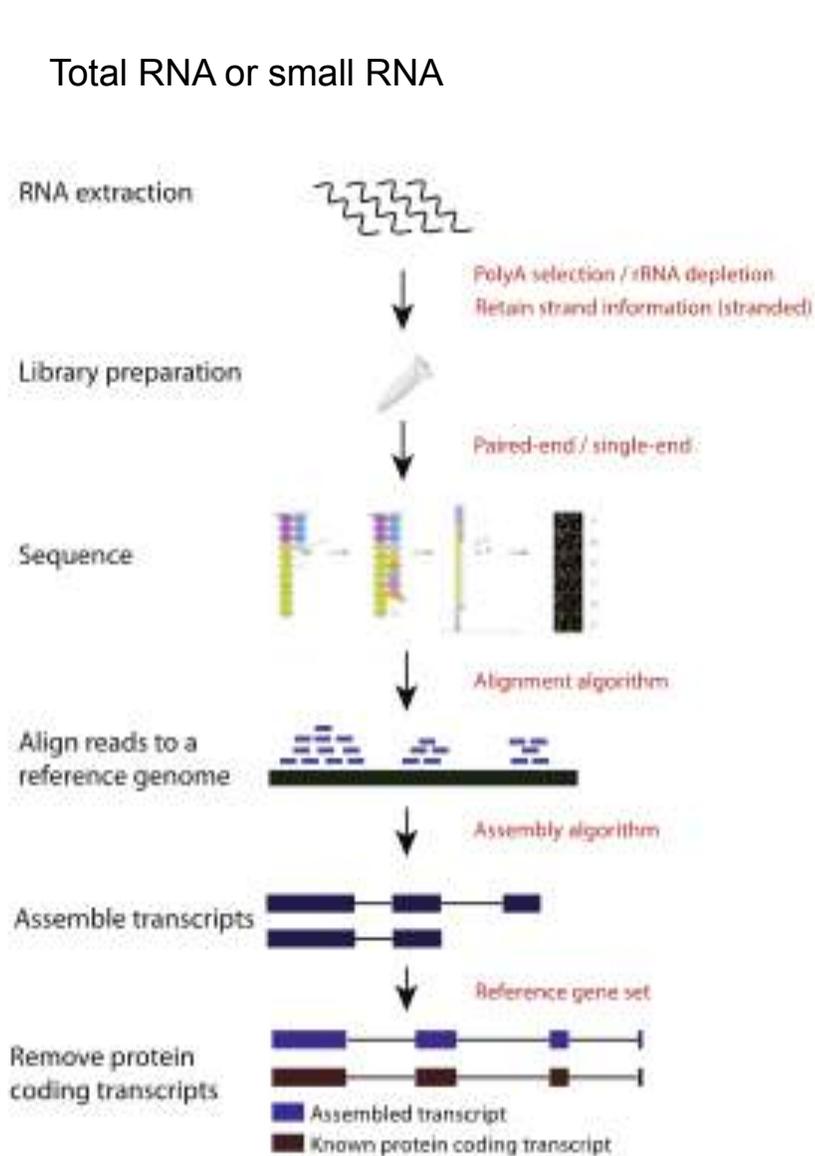


Υπολογιστικοί μέθοδοι για τη διάκριση μεταξύ mRNAs και lncRNAs γονιδίων

- Το μήκος του ανοιχτού πλαισίου ανάγνωσης (>300 νουκλεοτίδια ώστε να μην θεωρείται τυχαίο)
- Η σύσταση νουκλεοτιδίων, κωδικονίων (σε σχέση με τα mRNAs του κάθε οργανισμού).
- Η συχνότητα και η τοποθεσία σημειακών μεταλλάξεων (σύσταση αμινοξέων σε σχέση με τα mRNAs του κάθε οργανισμού).
- Η παρουσία αλληλουχιών που κωδικοποιούν λειτουργικές δομικές περιοχές πρωτεϊνών (Pfam)
- Ομολογία με αλληλουχίες γνωστών πρωτεϊνών (GenBank, Ensembl) - συνήθως δεν αρκεί καθώς πρέπει να υπάρχει και το απαραίτητο ανοιχτό πλαίσιο ανάγνωσης.
- Άλλα κριτήρια - Βαθμός συντήρησης (τα lncRNAs εξελίσσονται πιο γρήγορα -> λιγότερο συντηρημένα από τα mRNAs – ωστόσο υπάρχουν αρκετές εξαιρέσεις)

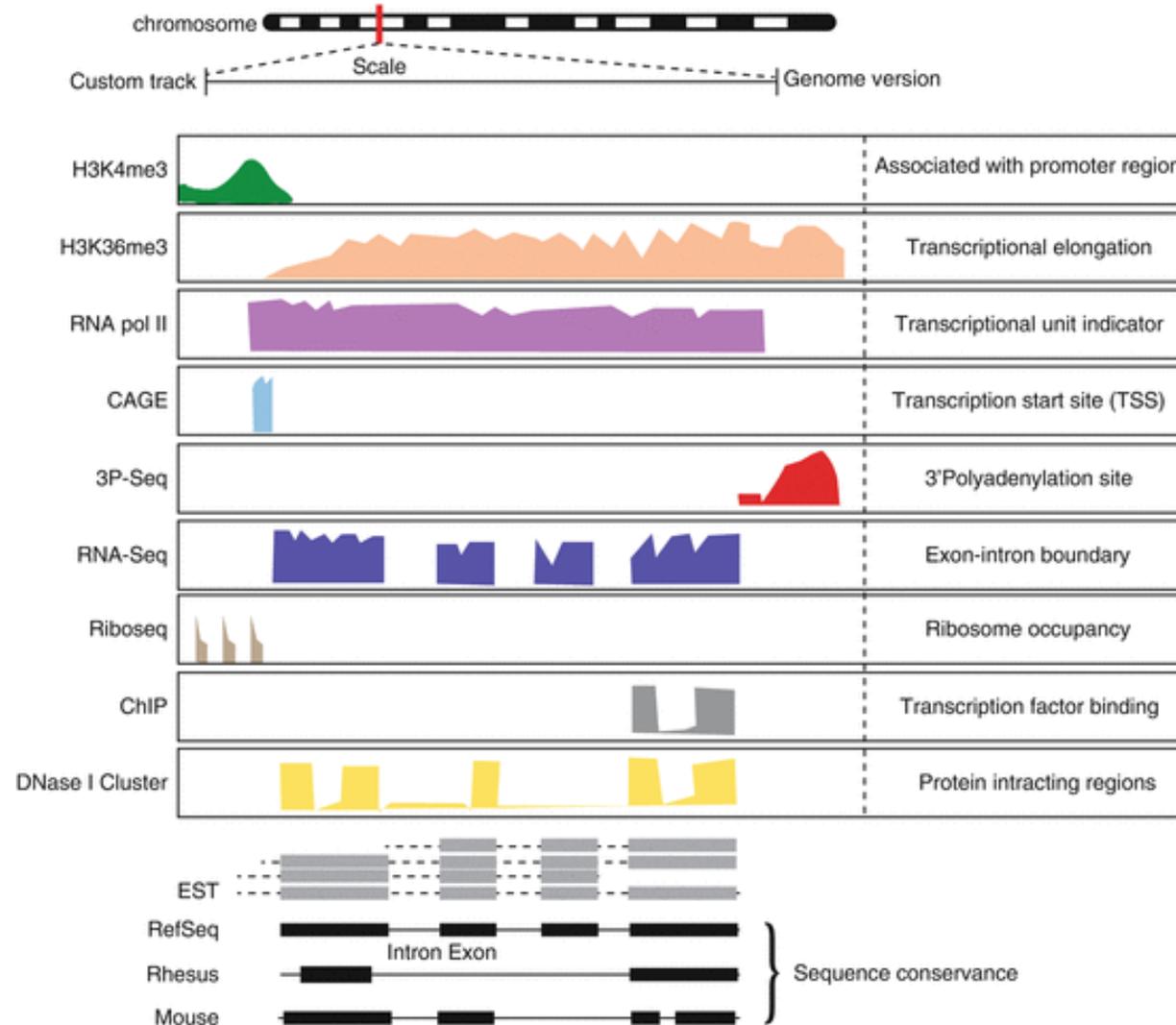
Αλληλουχοποίηση RNA επόμενης γενιάς

Total RNA or small RNA



- RNA library preparation by size selection of RNA fragments of interest (small ncRNAs <200nts or lncRNAs > 200nts)
- Ribosomal RNA depletion is necessary
- Paired-end sequencing is preferred as ncRNAs are enriched for repetitive elements.
- Directionality is important to discover ncRNAs species around coding genes
- Depth of sequencing is critical for ncRNA identification and characterization
- Cell-specific or state-specific sequencing is also critical as many lncRNAs are only expressed in specific types or states of cells

Διαδεδομένες τεχνικές για την πρόβλεψη και σχολιασμό γονιδιακών τόπων lncRNA με NGS



Reconstructed transcript Computational
e.g. Cufflinks

Active transcription Chip-Seq

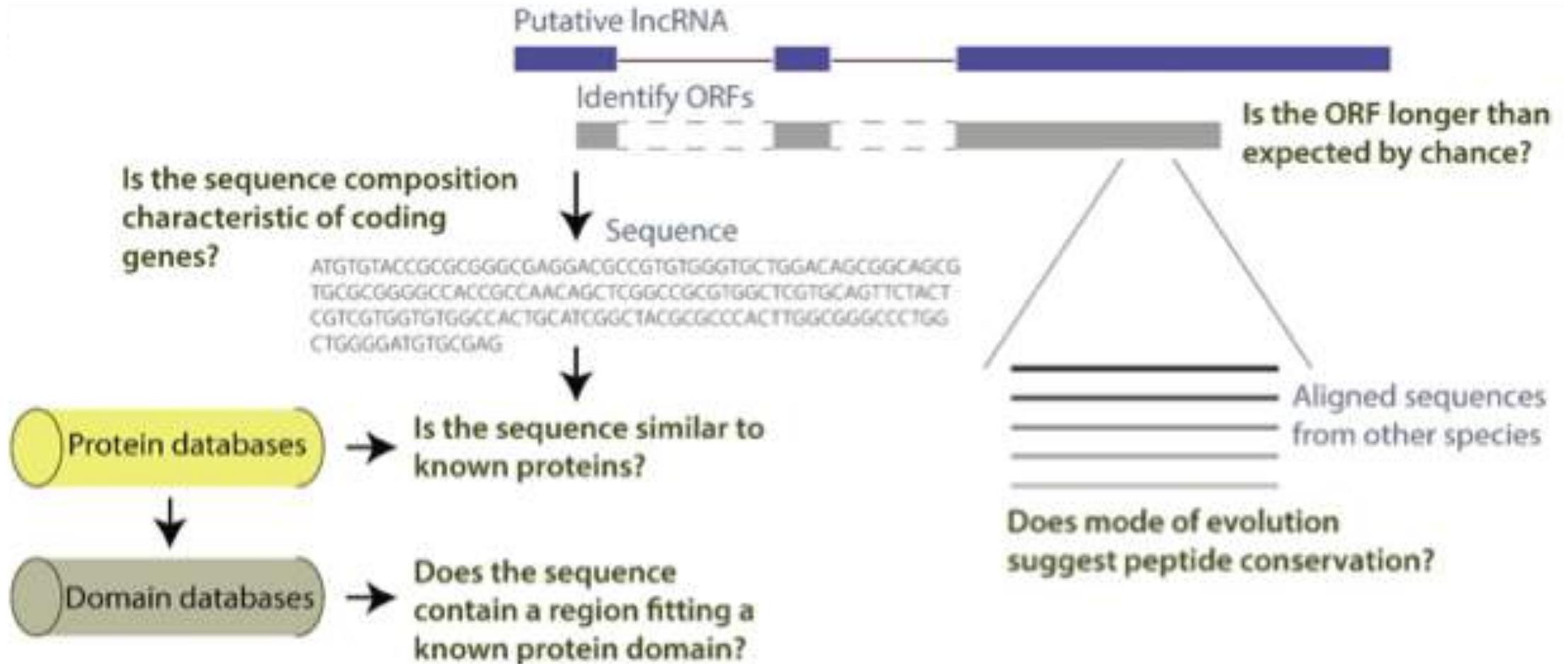
Promoter of gene Chip-Seq

Body of gene Chip-Seq

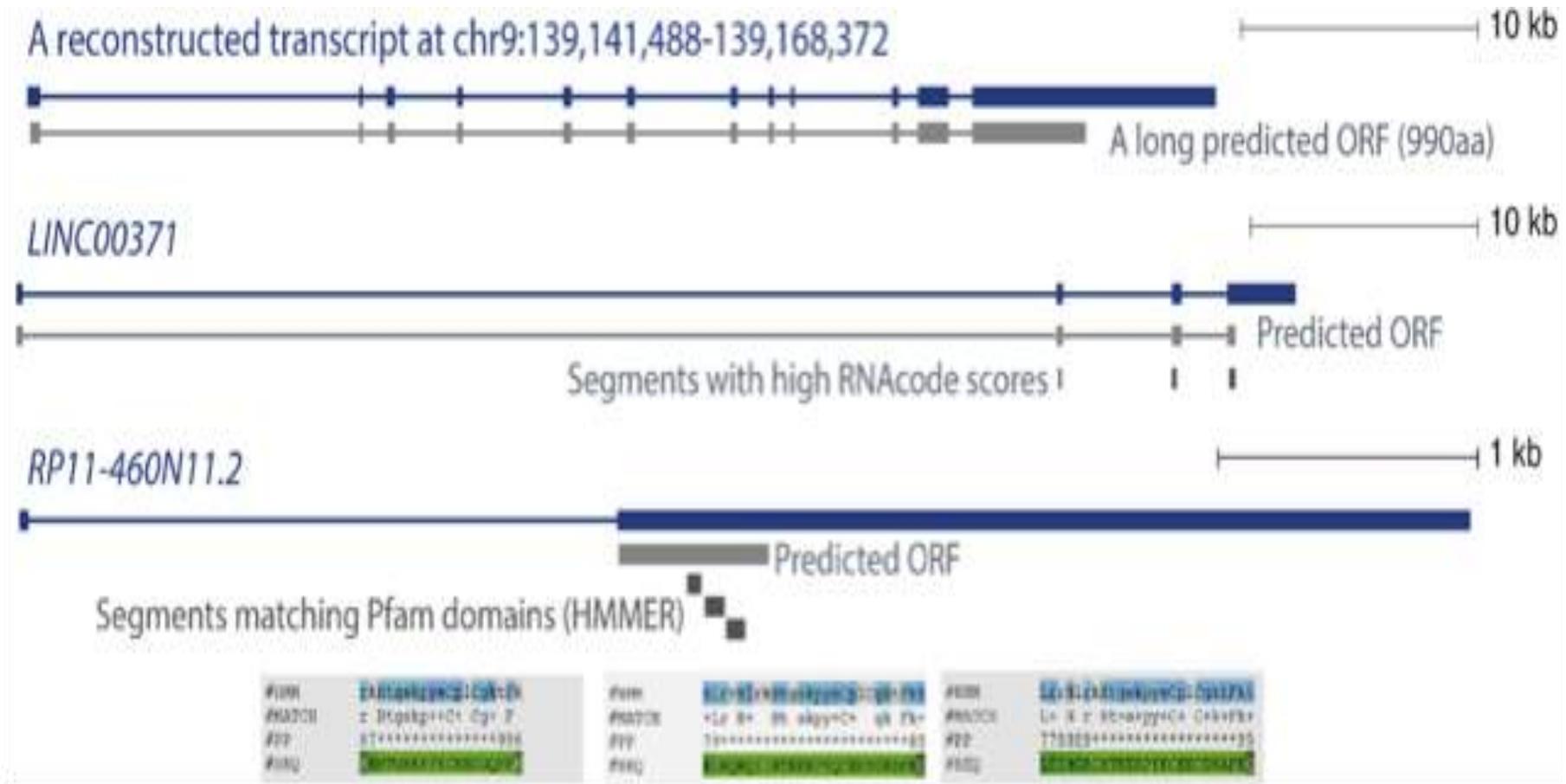
Distribution of reads RNA-Seq

Reference database e.g. ENSEMBL

Υπολογιστικοί μέθοδοι για τη διάκριση μεταξύ mRNAs και lncRNAs γονιδίων



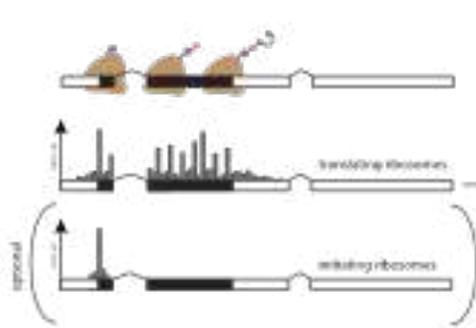
Παραδείγματα γονιδίων που δεν περνάνε ένα από τα κριτήρια ώστε να χαρακτηριστούν ως lncRNAs



Πειραματικές μεθοδολογίες σχολιασμού μακρών μη-κωδικοποιών RNAs

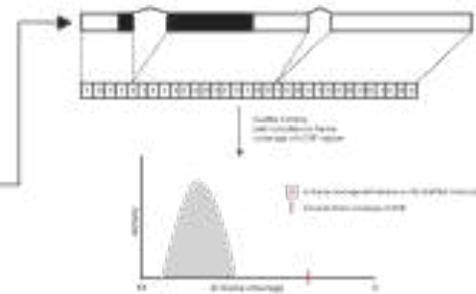
Ribosome-Sequencing

TIS-calling



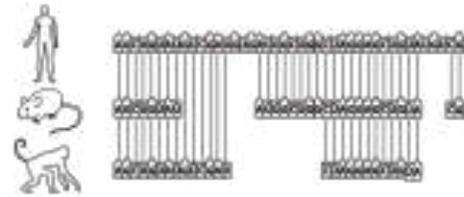
Computational filtering based on statistical characteristic of PCGs

noise filtering



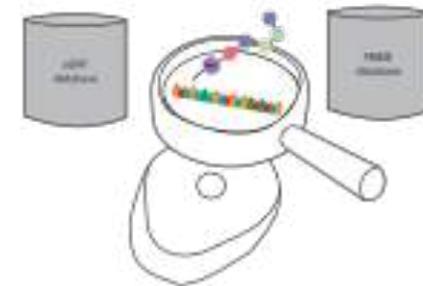
Computational annotation (alignment across species)

inner-BLAST

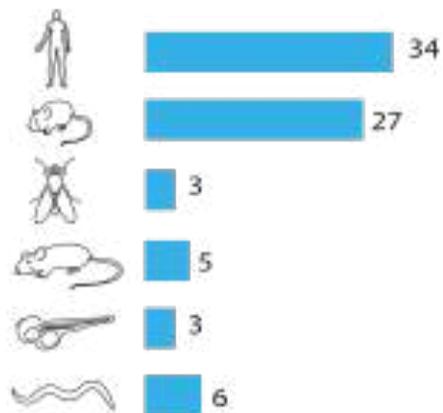


Peptides from Mass Spec data

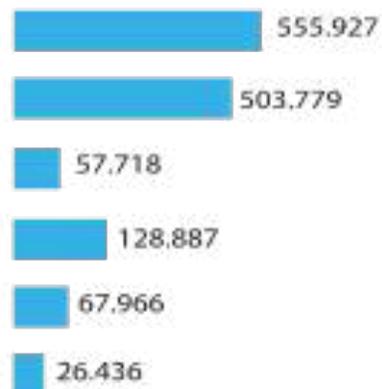
PRIDE-ReSpin



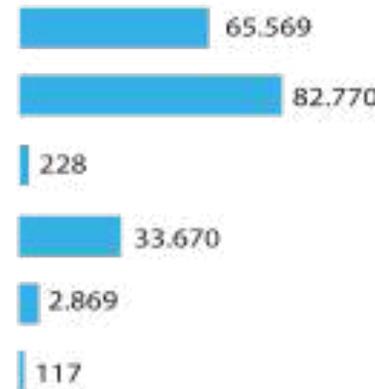
datasets



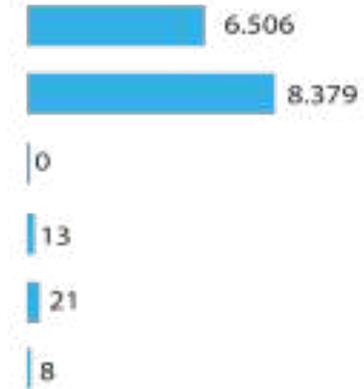
unique sORFs



sORFs with sequence similarity

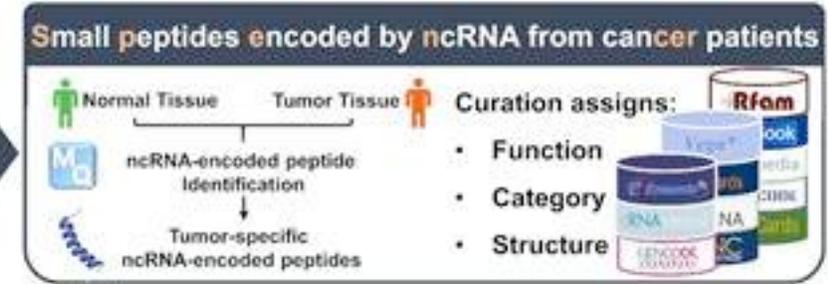
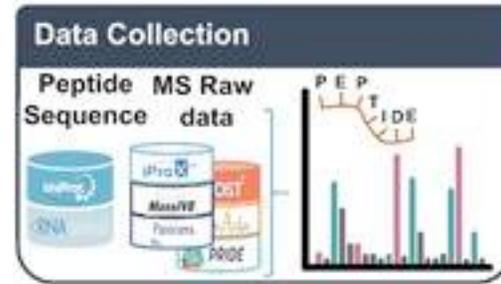
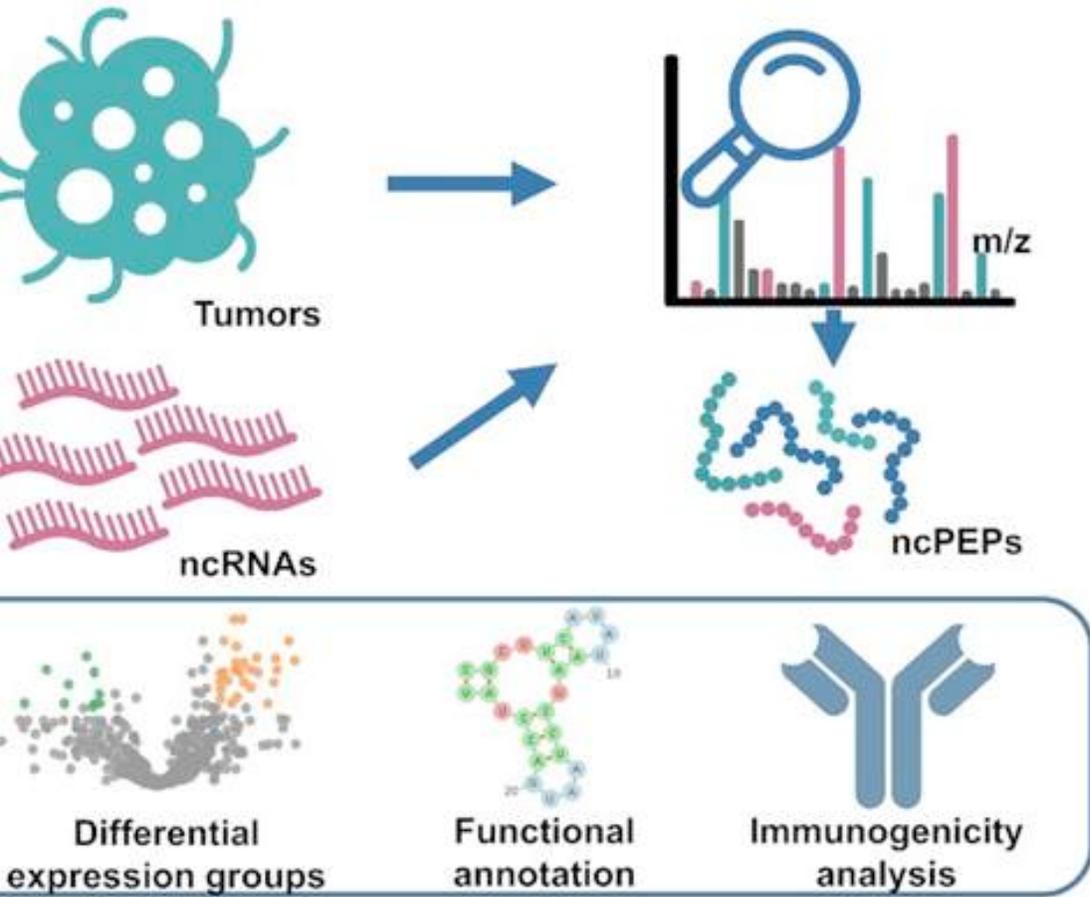


PRIDE-ReSpin sORFs



SPENCER

A comprehensive database for small peptides encoded by ncRNA in cancer patient (spencer.renlab.org)



SPENCER Database

CATEGORY

- Cancer associated ncRNA-encoded small peptides

Function

- Cancer functional study
- Potential neoantigen

Immunogenicity analysis
A circular diagram shows the cycle between **Tumor**, **ncRNAs**, **MHC-I**, and **Immunotherapy**.

Peptides
A bar chart shows the **Number of peptides** across various cancer types: Colon cancer, Gastric cancer, Liver cancer, Skin cancer, Breast cancer, Leukemia, Ovarian cancer, Pancreatic cancer, Prostate cancer, Bladder cancer, Thyroid cancer, Lung cancer, Kidney cancer.

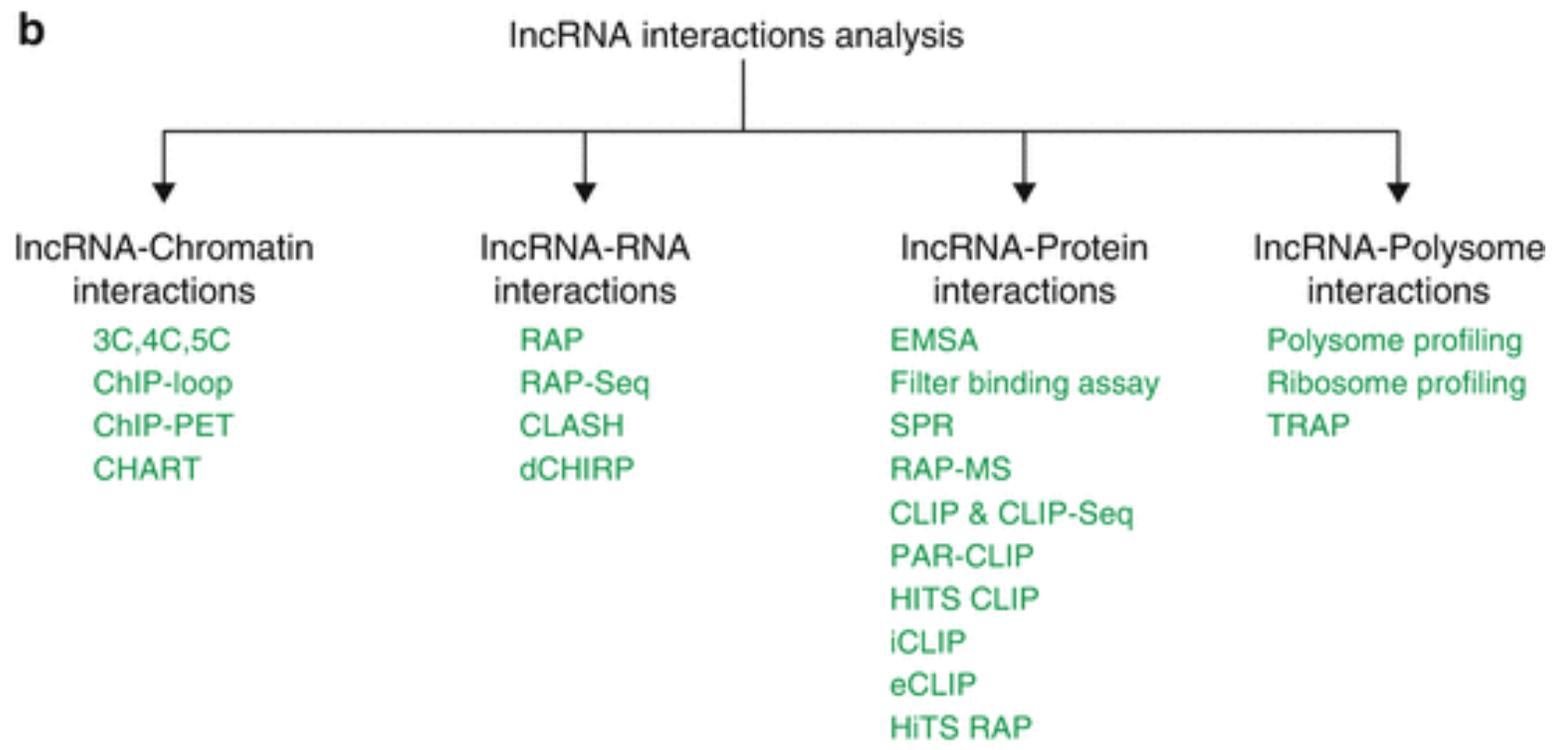
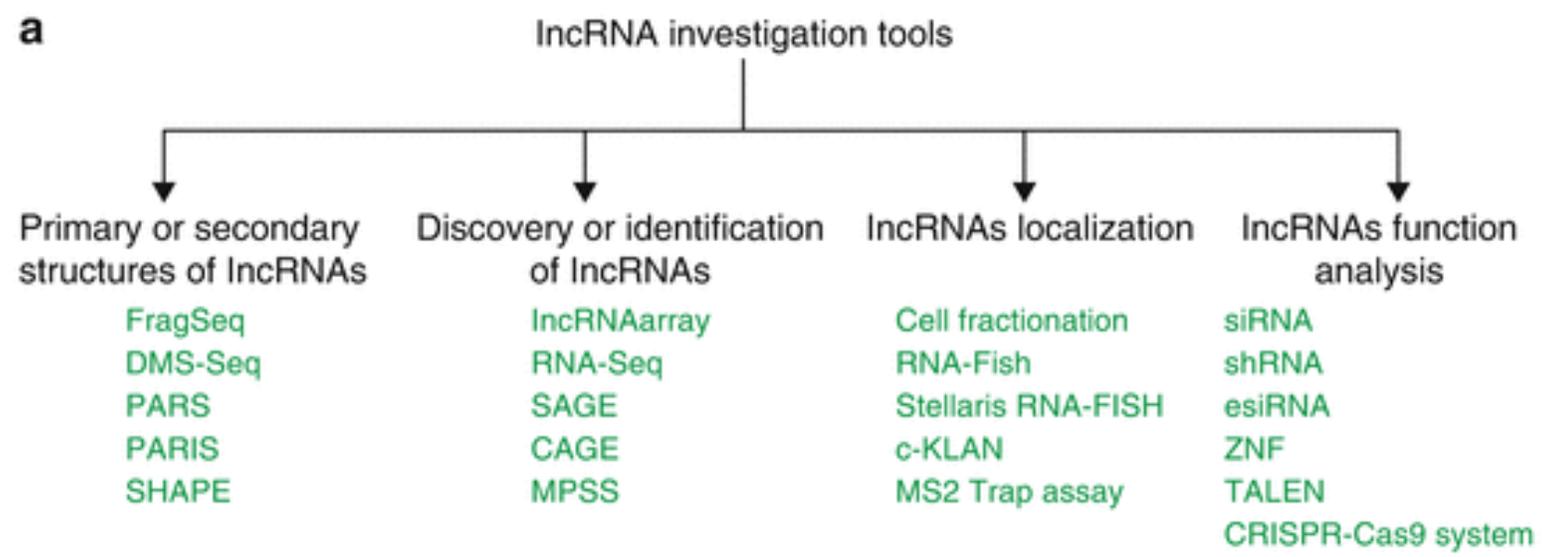
Expression in cancer patients

LINC00488	0.0171	-0.8962	1.9958	-0.6909	-0.5731
lnc-LSM2-1	-0.4259	-1.5640	1.9487	-0.2450	2.6622
lnc-STX1A-1	-0.1438	-0.3883	-0.6997	0.7149	-0.6894
LINC00174	-0.3551	-0.5731	2.3186	-0.4252	-1.3576
SEC24B-AS1	-0.1269	2.8822	-0.4602	0.7548	-1.1202

Immunogenicity analysis
A diagram shows an **MHC-I** molecule binding to a **Candidate** peptide.

Visualization
A sequence alignment shows **AEIAEISK MLDQTLIDLNEM** with positions 981, 997, and 1114 marked. Below it, a genomic track shows **ORF1**, **ORF2**, and **ORF3** regions.

Πειραματικά εργαλεία διερεύνησης της λειτουργίας lncRNAs



Μεθοδολογίες ανίχνευσης νέων γονιδίων

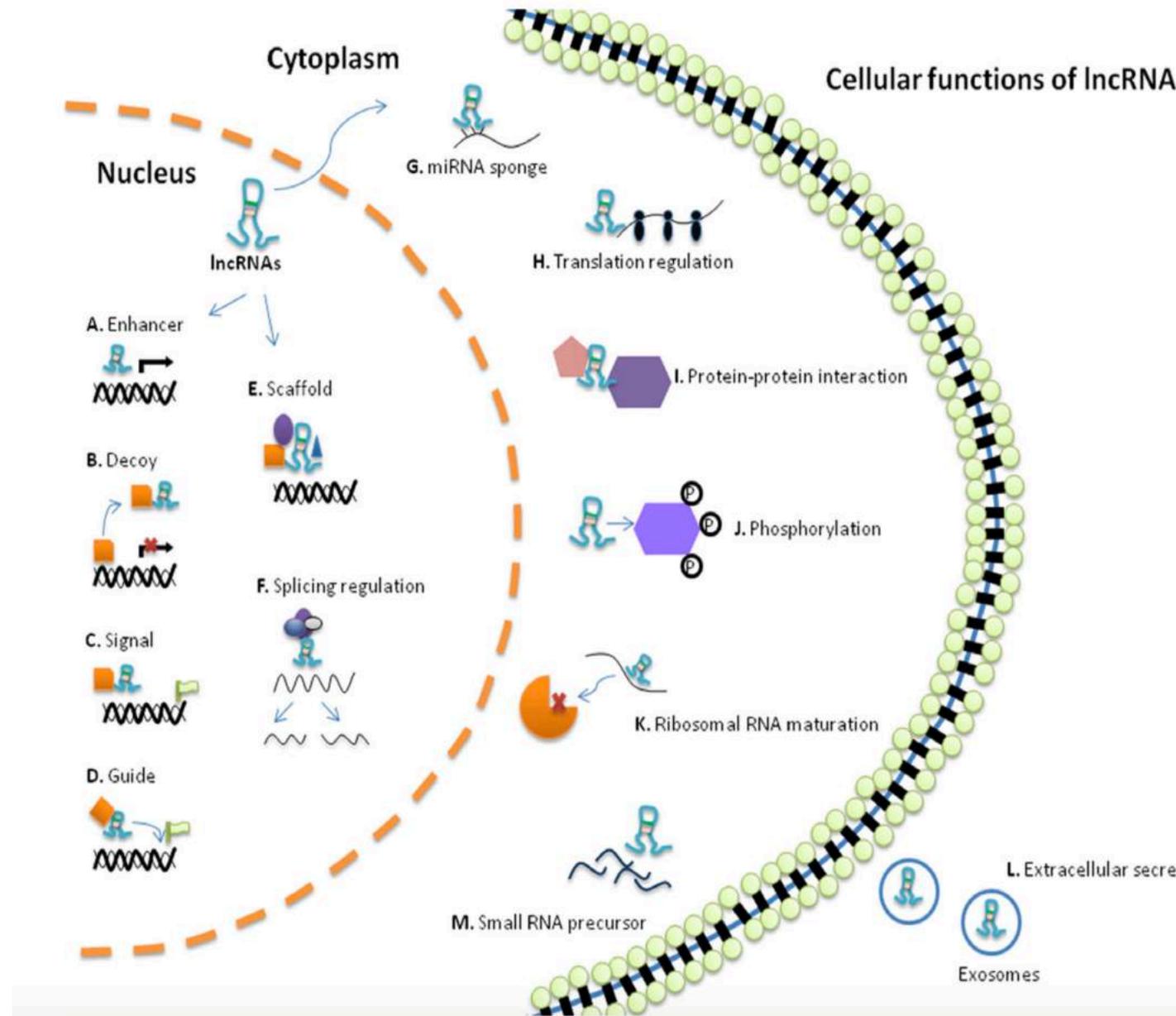
Table 1 | High-throughput technology used for transcriptome analyses

Technique	Principle	Aim	Organism	Main references
SAGE	Reverse transcription to cDNA of the transcriptome Short sequence tag production from unique poly(A)+ transcripts Cloning and sequencing of sequence tag arrays Quantitation of tag recovery	Quantitative gene expression data without prerequisite of sequence knowledge	<i>S. cerevisiae</i>	Velculescu <i>et al</i> , 1997
3'-long SAGE	Similar to SAGE Introduction of a recognition site for a IIS restriction enzyme at the 3' end of cDNA, recovery of 3' end tags High-throughput sequencing	Identification of 3' ends Distinction between overlapping transcripts	<i>S. cerevisiae</i>	Neil <i>et al</i> , 2009
CAGE	Similar to SAGE but from CAP-selected RNAs	Identification of transcription start sites	Mouse Human	Carninci <i>et al</i> , 2006
ASSAGE	Similar to SAGE RNA treatment with bisulphite (changes cytosine residues to uracil) before reverse transcription	Unambiguous strand specificity	Human	He <i>et al</i> , 2008
RNA-Seq	Reverse transcription to cDNA of poly(A)+ RNA High-throughput sequencing	Quantitative transcriptome data High sensitivity, low background, high resolution	<i>S. cerevisiae</i> <i>S. pombe</i>	Nagalakshimi <i>et al</i> , 2008; Wilhelm <i>et al</i> , 2008
GRO-seq	Extension of nascent RNA associated with engaged RNAPII (nuclear run-on) Isolation and purification of the RNAs Reverse transcription to cDNA High-throughput sequencing	Genome-wide positional mapping, determination of transcript amount, orientation of engaged RNAPII	Human	Core <i>et al</i> , 2008
High-resolution tiling arrays	Reverse transcription of total or selected RNA followed by hybridization on DNA chip Probes overlap and cover the entire genome	Quantitative transcriptome data from sequenced genomes Strand specific	<i>S. cerevisiae</i> <i>S. pombe</i> Rice Human	David <i>et al</i> , 2006; Xu <i>et al</i> , 2009; Dutrow <i>et al</i> , 2008; Wilhelm <i>et al</i> , 2008; L. Li <i>et al</i> , 2006; Kapranov <i>et al</i> , 2007a; Preker <i>et al</i> , 2008
ChIP-chip	Chromatin immunoprecipitation of RNAPII and TFs Hybridization on DNA array	Mapping of chromatin-bound RNAPII and TFs, genome-wide	<i>S. cerevisiae</i> Human	Steinmetz <i>et al</i> , 2006; Kim <i>et al</i> , 2005
ChIP-Seq	Chromatin immunoprecipitation of RNAPII High-throughput sequencing	Mapping of chromatin-bound RNAPII, also at low complexity sequences (such as telomeres and transposons)	Mouse	Seila <i>et al</i> , 2008

The techniques used in the ENCODE and FANTOM projects are not detailed. RNAPII, RNA polymerase II; *S. cerevisiae*, *Saccharomyces cerevisiae*; *S. pombe*, *Schizosaccharomyces pombe*; TF, transcription factor.

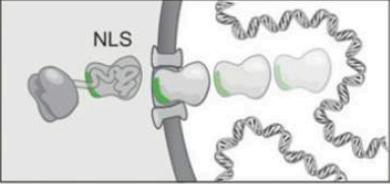
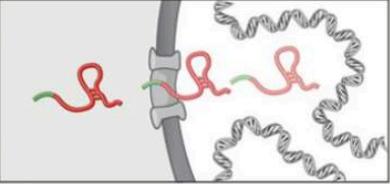
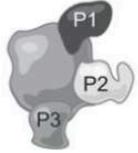
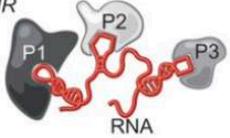
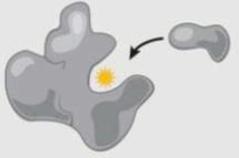
Λειτουργίες lncRNAs

- Γονιδιακοί τόποι lncRNAs σε σχέση με τους ρυθμιστικούς τους ρόλους στον πυρήνα, κυτταρόπλασμα και τον εξωκυττάριο χώρο. Τα lncRNAs που εντοπίζονται στον πυρήνα δρουν «εν δυνάμει» ως (A) ενισχυτές για την επαγωγή της μεταγραφής *in cis* ή *in trans*; (B) δολώματα σε μεταγραφικούς παράγοντες και ρυθμιστές της δομής της χρωματίνης, (C) μοριακά σήματα ενεργοποίησης ή αποσιώπησης της γονιδιακής έκφρασης; (D) οδηγοί πρωτεϊνών σε ειδικές θέσεις του DNA για τη ρύθμιση της μεταγραφής ή της δομής της χρωματίνης; (E) ικριώματα σε πρωτεϊνικά σύμπλοκα για την στρατολόγηση/ενεργοποίηση τους; (F) μπορούν να ρυθμίζουν το εναλλακτικό μάτισμα των pre-mRNAs. Στο κυτταρόπλασμα, τα lncRNAs δρουν «εν δυνάμει» ως (G) σφουγγάρια/δολώματα των miRNAs ώστε να αναστείλουν την δραστηριότητά τους (H) σε μεταφραστικές διεργασίες (αλληλεπίδραση RNA-RNA ή πρωτεΐνης-RNA); (I) στην αλληλεπίδραση πρωτεϊνών (J) στη φωσφορυλίωση πρωτεϊνών και την ενεργοποίηση σηματοδοτικών μονοπατιών; (K) στην ωρίμανση ριβοσωμικών RNA. Τέλος κάποια lncRNAs (L) μπορούν να φορτωθούν σε εξωσώματα και να μεταφερθούν σε άλλα κύτταρα και (M) κάποια λειτουργούν απλώς ως φορείς miRNAs και άλλων ρυθμιστικών μικρών RNA



Functionally equivalent RNA and DNA domains

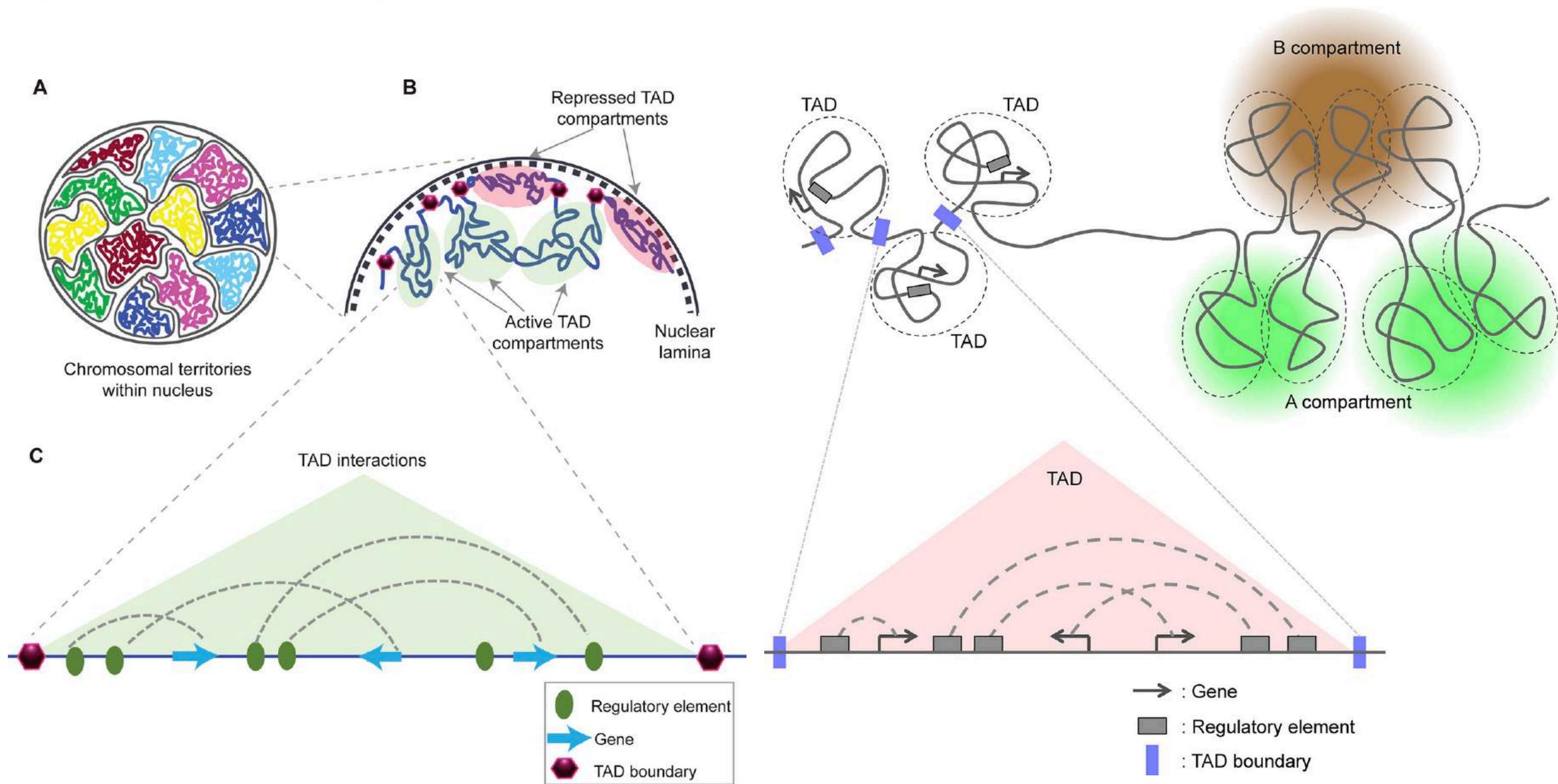
- Despite distinct mechanisms by which they achieve their equivalence, RNA domains are capable of affecting many biological activities that have traditionally been ascribed exclusively to proteins. Much work remains, however, in identifying common RNA primary, secondary, and tertiary structures that comprise orthologous functional domains and can mediate these and other subcellular activities.

	Protein Domains	RNA equivalent	References
Localization	Signal peptide (SRP) NLS 	<i>BORG</i> <i>FIRRE</i> 	<ul style="list-style-type: none"> Zhang, et al. <i>Mol. Cell. Biol.</i>, 2014 Hacisuleyman, et al. <i>Nat. Struct. Mol. Biol.</i>, 2014
Scaffolding	A-kinase anchor proteins (AKAPs) <i>KSR</i> in MAPK signaling <i>HOMER</i> (calcium signaling) 	Ribosomal RNA <i>XIST</i> <i>HOTAIR</i> 	<ul style="list-style-type: none"> Brimacombe & Stiege, <i>Biochem. J.</i>, 1985 Chaumeil, et al., <i>Genes & Development</i>, 2006 Yoon, et al., <i>Nat. Commun</i>, 2013 Zappulla & Cech, <i>CSH Symp Quant Biol</i> 2006
DNA binding	Leucine zipper bHLH 	Direct Hybridization Hoogsteen binding (triplex) 	<ul style="list-style-type: none"> Grote, et al., <i>Developmental Cell</i>, 2013 Schmitz, et al., <i>Genes & Development</i>, 2010 O'Leary, et al., <i>Cell Reports</i>, 2015
RNA binding	RNA recognition motif hnRNPs splicing regulation/polyA stabilization 	miRNA/siRNA targeting mRNA hybridization 	<ul style="list-style-type: none"> Engreitz, et al., <i>Cell</i>, 2014 Kretz, et al., <i>Nature</i>, 2013
Enzymatic	Protein kinase domain 	Group I introns RNase P <i>glmS</i> 	<ul style="list-style-type: none"> Cech, <i>Annu. Rev. Biochem.</i>, 1990 Guerrier-Takada, et al., <i>Cell</i>, 1983 Winkler, et al., <i>Nature</i>, 2004

Τοπική ρύθμιση της δομής της χρωματίνης ή της γονιδιακής έκφρασης

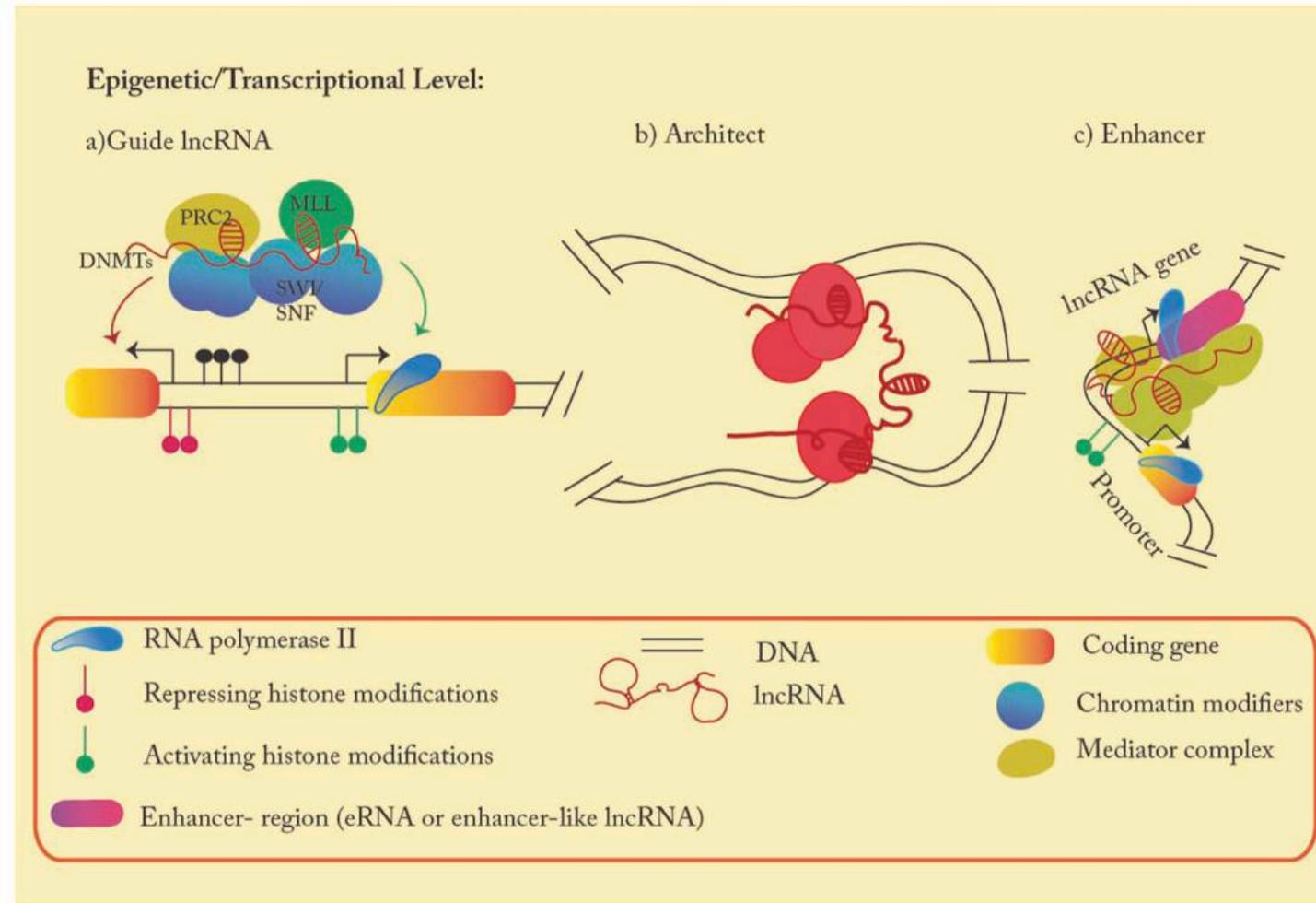
- Τα cis-δραστικά lncRNAs χρησιμοποιούν τρεις μηχανισμούς: Πρώτον, τα lncRNAs στρατολογούν άμεσα ρυθμιστικά σύμπλοκα σε γειτονικούς γονιδιακούς τόπους; Δεύτερον, ρυθμίζουν άμεσα τη μεταγραφή ή το εναλλακτικό μάτισμα ως RNA ρυθμιστικά στοιχεία και τέλος έμμεσα προσδένοντας σε DNA ρυθμιστικά στοιχεία υποκινητών, ενισχυτών γονιδίων.
- Το lncRNA μετάγραφο XIST (X-inactive specific transcript), είναι το πιο γνωστό παράδειγμα cis-δραστικού στοιχείου που όμως ενέχεται στην απενεργοποίηση ολόκληρου του χρωμοσώματος X.

Nuclear architecture

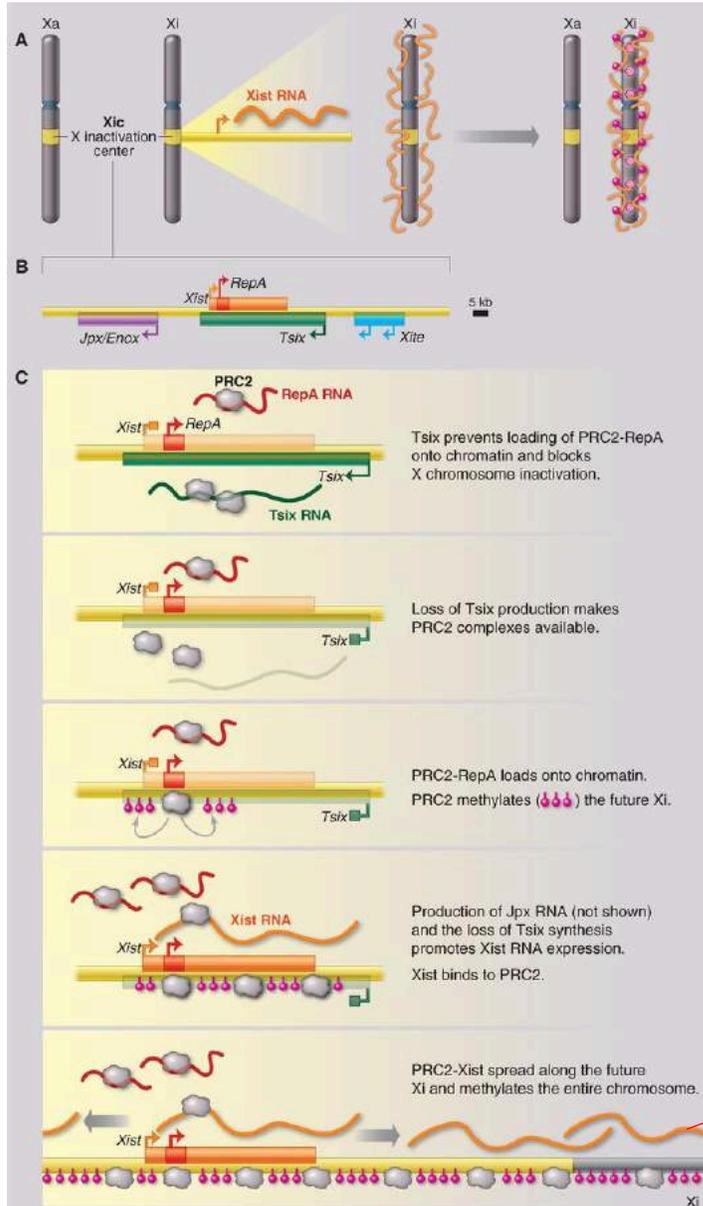


Τα lncRNAs, είναι πολυδιάστατα βιοχημικά πολυμερή; Επιγενετικό/μεταγραφικό επίπεδο

- a. Τα lncRNAs-οδηγοί (το μοντέλο δεμένου ικρίωματος; tethered scaffold model): ρύθμιση της γονιδιακής έκφρασης μέσω στρατολόγησης ή αποσυναρμολόγησης ρυθμιστών της δομής της χρωματίνης; Με αυτόν τον τρόπο ένα lncRNA αναγνωρίζει και προσδένει ειδικά μοτίβα στο DNA και στη συνέχεια στρατολογεί ειδικά σύμπλοκα όπως το PRC2.
- b. Τα lncRNAs-αρχιτέκτονες (το μοντέλο ικρίωμα γέφυρας; the bridging scaffold model): με αυτό τον τρόπο τα lncRNAs λειτουργούν αναδιαμορφώνοντας την τρισδιάστατη δομή της χρωματίνης.
- c. Τα lncRNAs που συσχετίζονται με περιοχές ενισχυτών ή δρουν ως ενισχυτές (eRNAs, enhancer-like lncRNA): με αυτό τον τρόπο ένα lncRNA που εκφράζεται από ενισχυτές ή ανοδικά του γονιδίου, στρατολογεί μεταγραφικούς παράγοντες σε ρυθμιστικές περιοχές (ενισχυτές ή υποκινητές) πρωτεϊνικών γονιδίων.



XIST

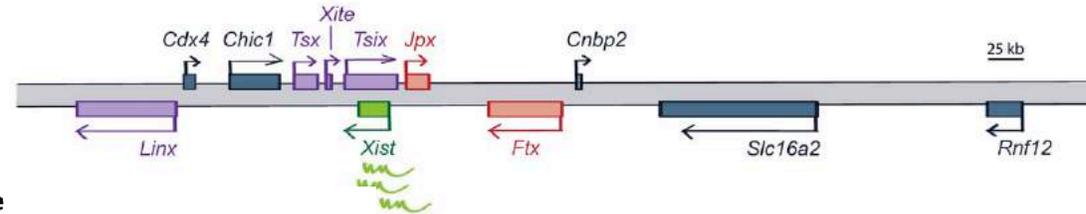


A

Mouse X Chromosome



Xic
~800 Kb
(Nora et al. 2012)

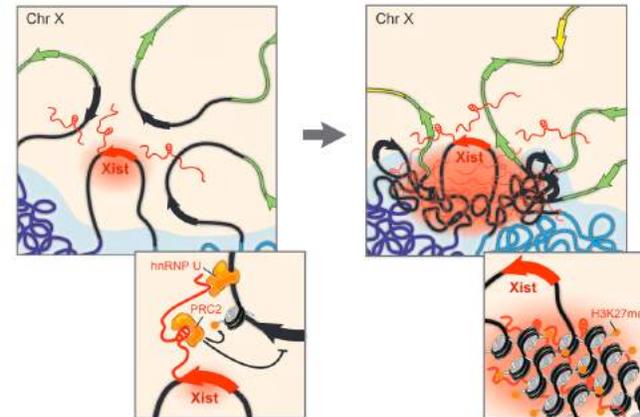
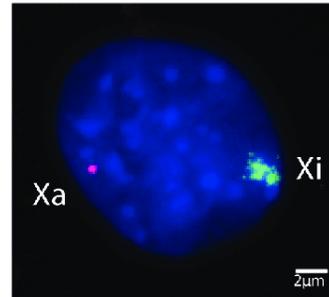


A Xist RNA structure



B Xist RNA expression pattern

Xist RNA
AtrX RNA
DNA

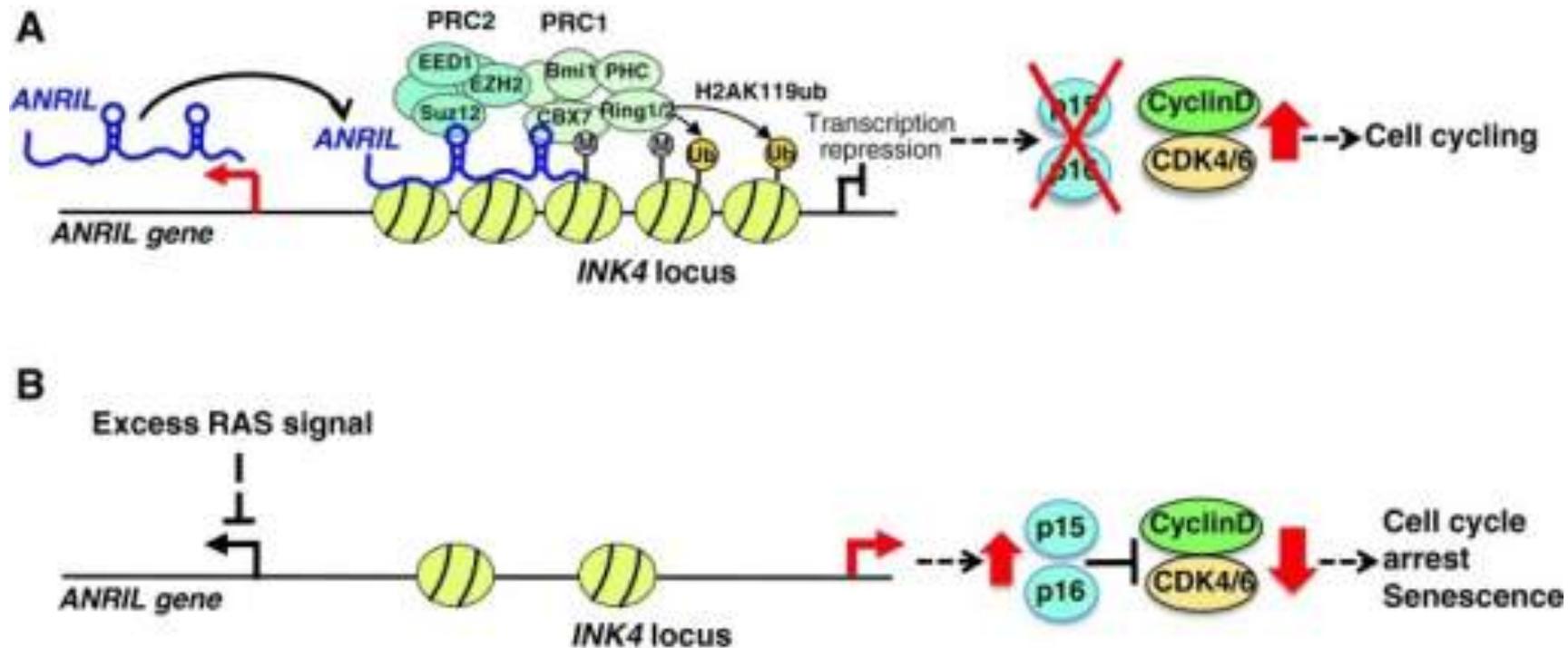


Xist is a ~17kb lncRNA essential for mammalian X chromosome inactivation. Xist RNA spreads along the inactive X chromosome, and this is followed by induction of a series of PRC2-mediated repressive chromatin marks. Interestingly, the most conserved regions of Xist correspond to low-copy repetitive elements, where the repetitive element A (repA) is the most highly conserved. RepA localizes at the 5' end of Xist, a region which has been found to be essential for X chromosome inactivation. RepA binds, and recruits the chromatin remodeling polycomb repressor complex 2 (PRC2), consisting of EZH2, SUZ12 and EED, which initiate the X inactivation by chromatin remodeling

The Xist gene is the tool to deactivate one of the chromosomes. Xist gene has very specific function which is to produce an unusually large cis-acting RNA transcript and never translated into a protein. It acts as RNA transcript to paint the whole chromosome to be deactivated. The deactivated chromosomes is called Barr body after Murray Barr, the cytologist who discovered it.

Cis-acting scaffold regulation of transcription

- ANRIL acts as scaffold for PRC2-mediated repression of the INK4 locus. Excess RAS suppresses ANRIL expression, promoting cell cycle arrest.



Cis acting antisense ncRNAs

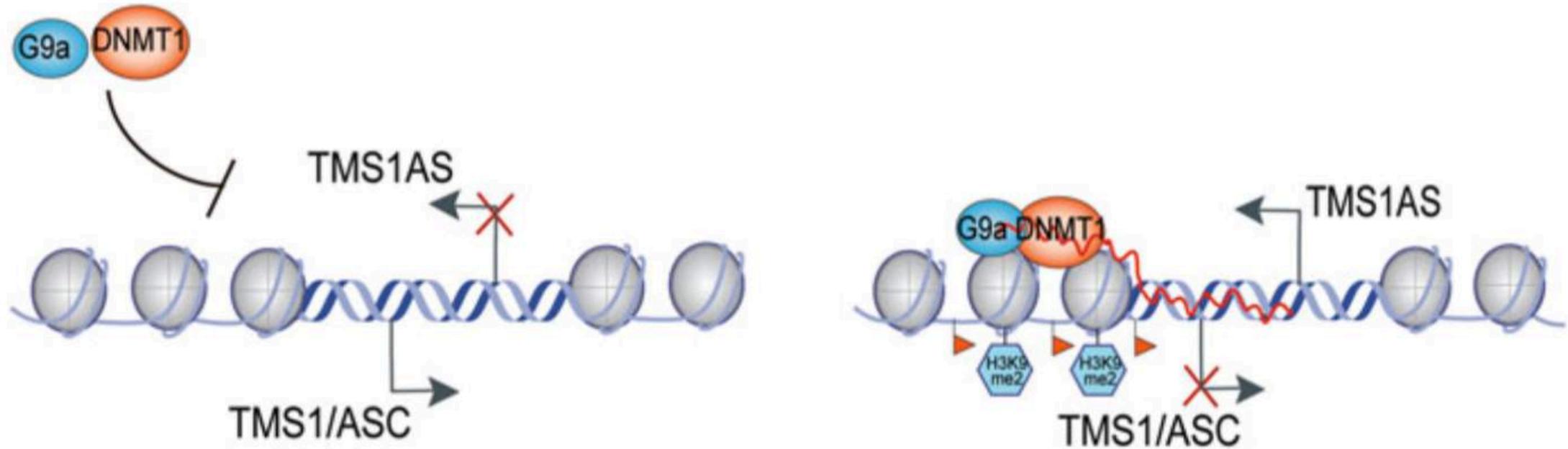


Fig. 2.1 TMS1AS regulates TMS1/ASC at the epigenetic level. Upon being transcribed, TMS1AS can act in cis to recruit the chromatin repressor proteins DNMT1 and G9a to TMS1/ASC promoter

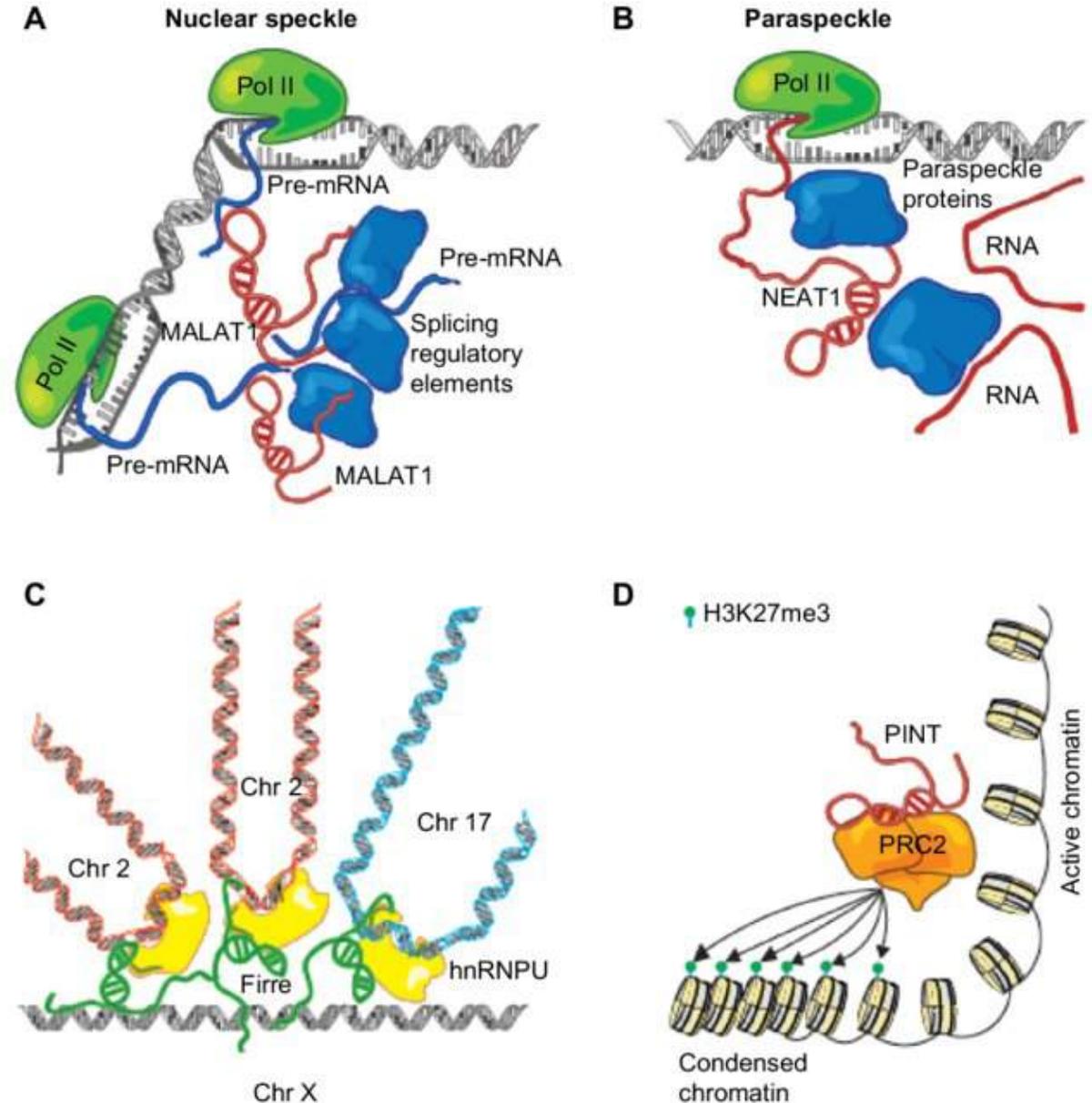
Ρύθμιση των κυτταρικών λειτουργιών με ένα *trans*-δραστικό τρόπο.

Πολλά lncRNAs ρυθμίζουν την κυτταρική λειτουργία με λειτουργίες οι οποίες είναι ξεχωριστές από τον γενετικό τόπο που μεταγράφονται. Αυτές χωρίζονται σε τρεις κατηγορίες

1. Τα lncRNAs ρυθμίζουν τη γονιδιακή έκφραση μέσω αλληλοεπίδρασης με σύμπλοκα χρωματίνης ή άμεσης πρόσδεσης στο DNA. Χαρακτηριστικό παράδειγμα είναι το lncRNAs HOTAIR
2. Τα lncRNAs οργανώνουν την αρχιτεκτονική του πυρήνα λειτουργώντας ως ικριώματα.
3. Τέλος, τα lncRNAs ελέγχουν τη ενδοκυτταρική μεταφορά, διακυτταρική επικοινωνία, κυτταρική σηματοδότηση και την ενεργότητα των πρωτεϊνών ή άλλων RNA πολυμερών.

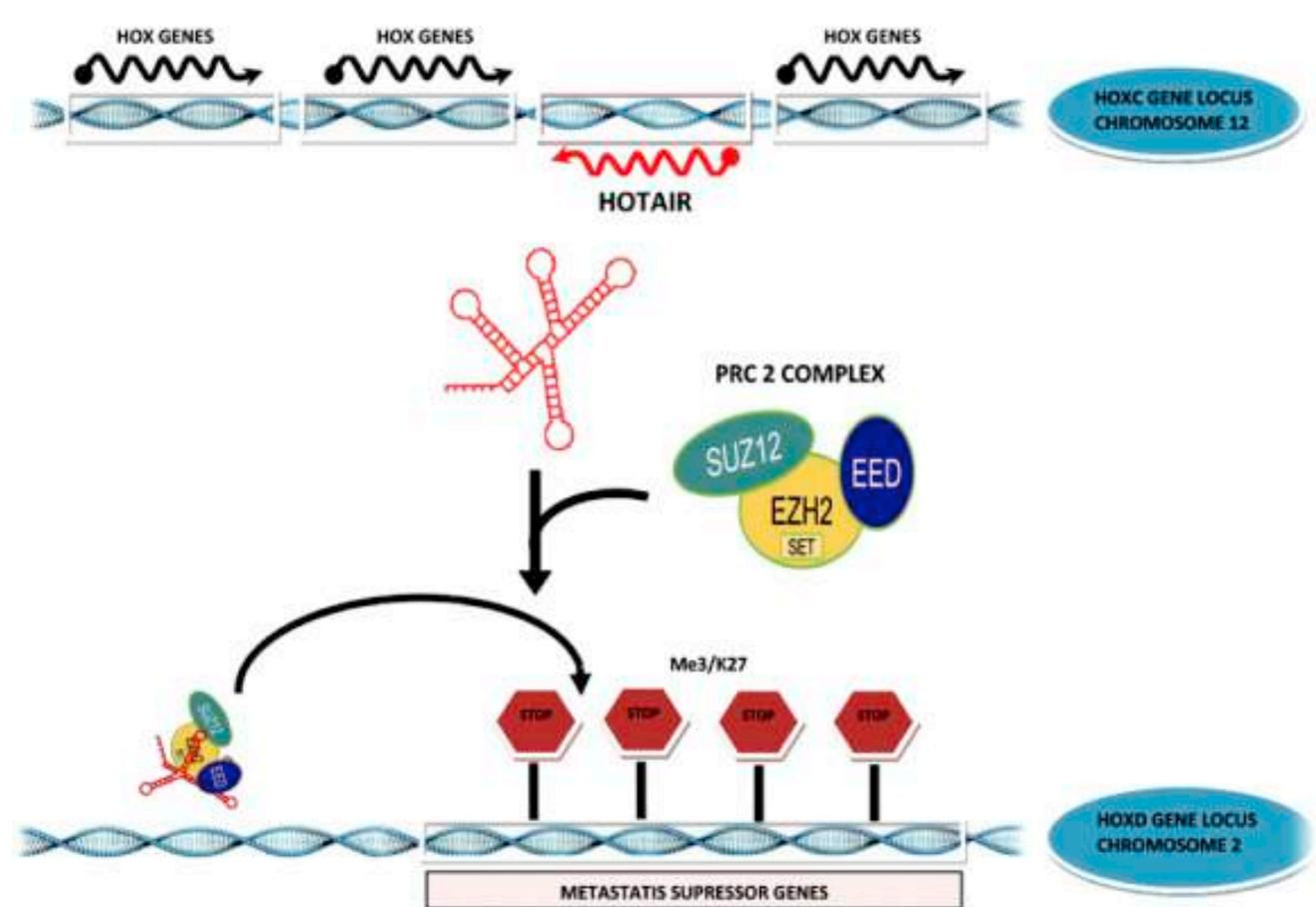
Μηχανισμοί trans-δραστικών lncRNAs

- (A) Το NEAT1 εντοπίζεται σε πυρηνικά paraspeckles όπου βρίσκονται πολλοί παράγοντες του εναλλακτικού ματίσματος για τη ρύθμιση πρόδρομων μεταγράφων mRNAs (pre-mRNAs). (B) Το MALAT1 είναι συστατικό πυρηνικού speckle component, και αλληλοεπιδρά με πρόδρομα RNAs και πρωτεΐνες εναλλακτικού ματίσματος. (C) Το Firre μεταγράφεται από το χρωμόσωμα X και δεσμεύει την πρωτεΐνη hnRNPU ώστε να αλληλοεπιδράσει με άλλους γονδιωματικούς τόπους άλλων χρωμοσωμάτων. (D) Το Pint αλληλοεπιδρά με την PRC2 για την διαμεσολάβηση της κατάθεσης H3K27me3 και σήμανσης της αποσιώπησης γονιδίων του μονοπατιού p53.

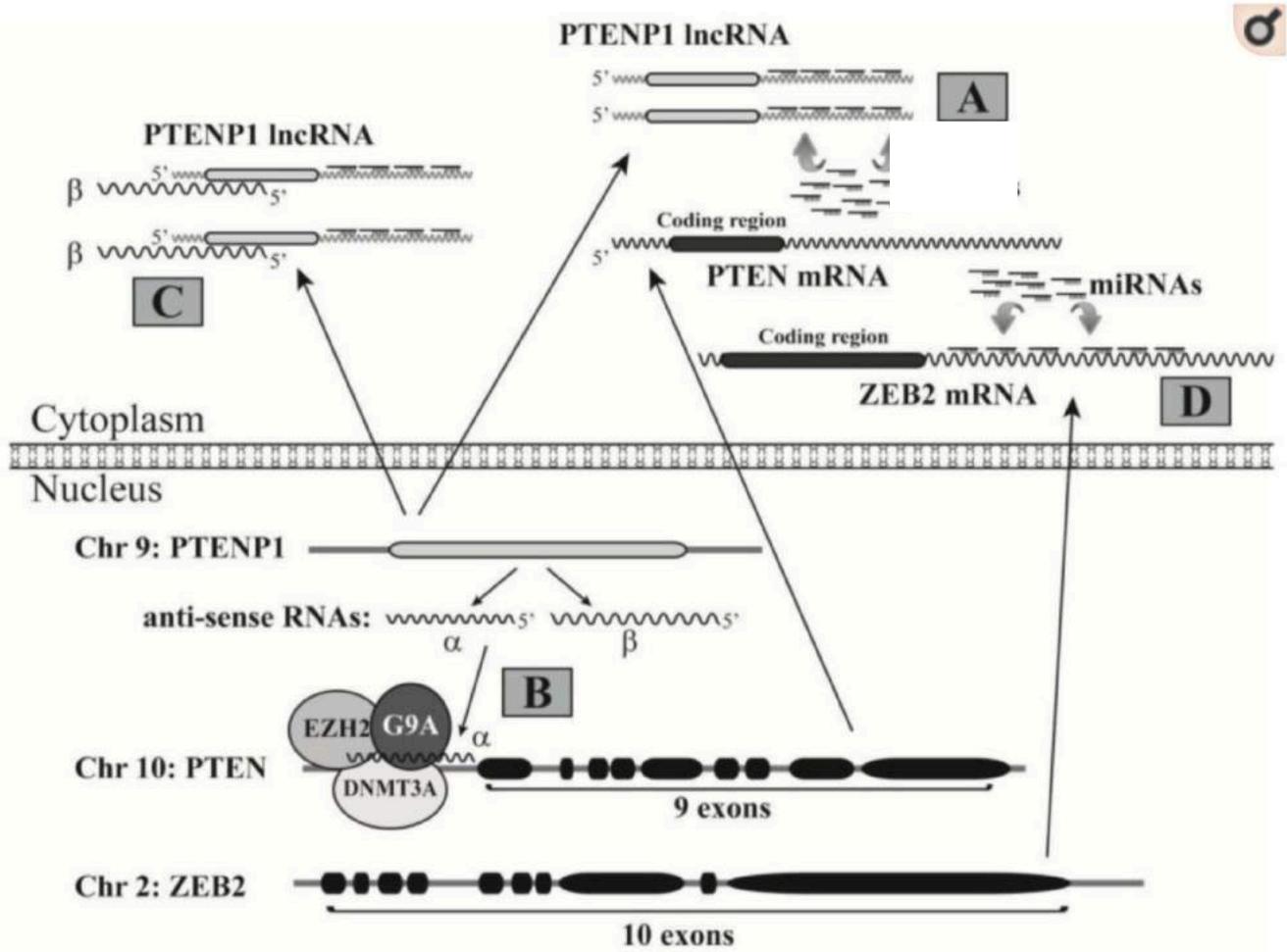


HOTAIR

- HOTAIR is transcribed from the antisense strand of the HOXC gene, which is located between HOXC11 and HOXC12 on chromosome 12.
- HOTAIR has been shown to be abnormally expressed in a variety of cancers and is associated with poor prognosis.
- With the emergence of tumor resistance, HOTAIR has also been confirmed to be involved in chemotherapy resistance of tumors.

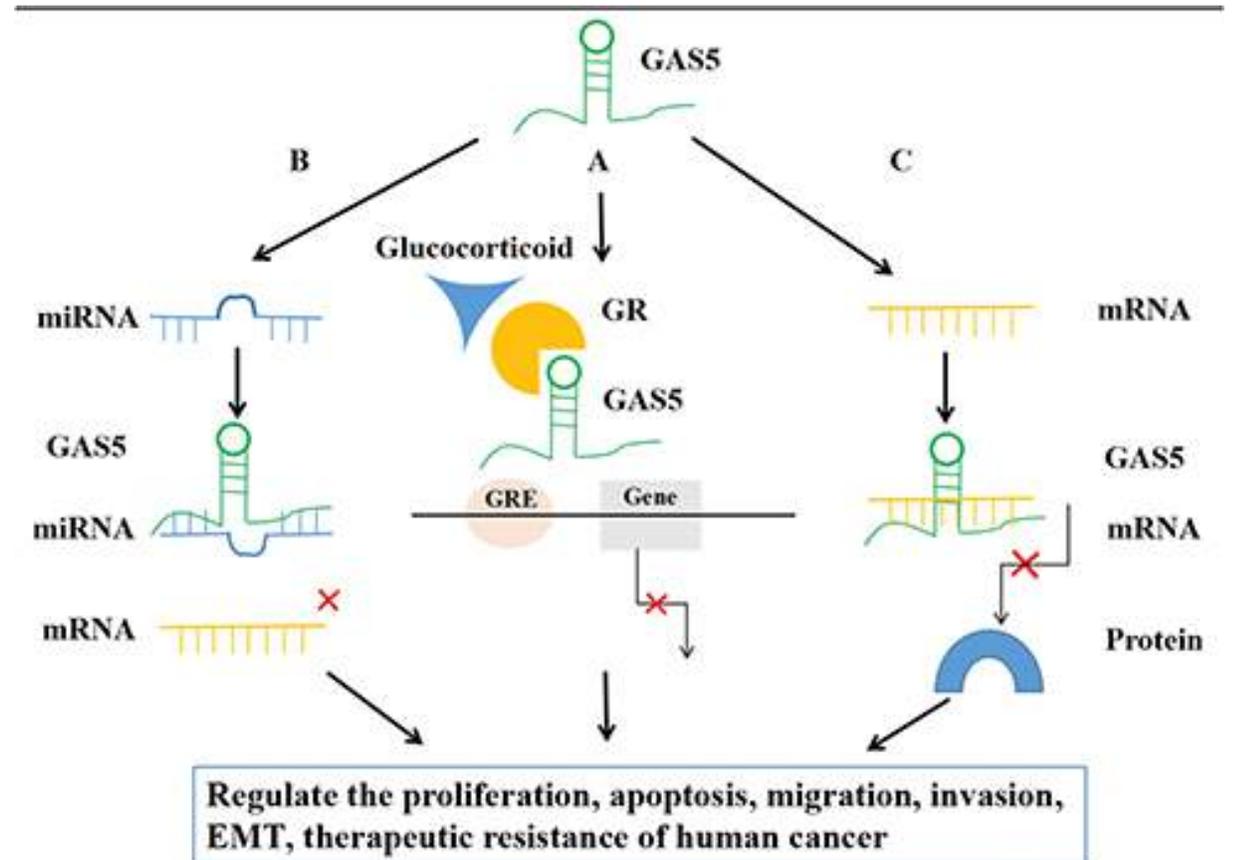
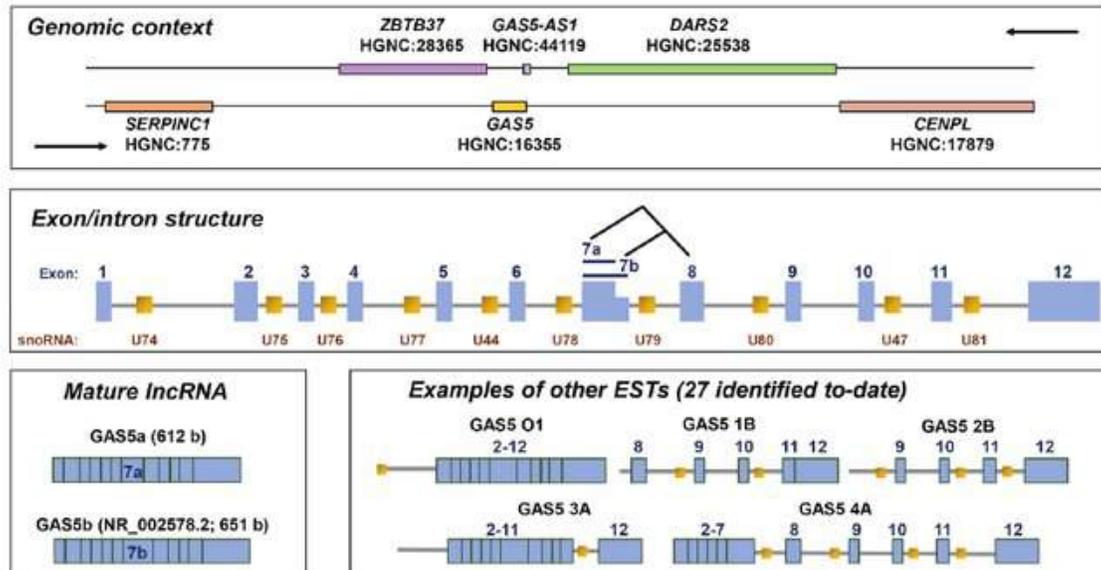


Το παράδειγμα ρύθμισης PTEN από μικρά και μακρά μη-κωδικοποιά RNAs (Trans-acting antisense RNA)



Ρύθμιση της μεταγραφής από lncRNAs

GAS5

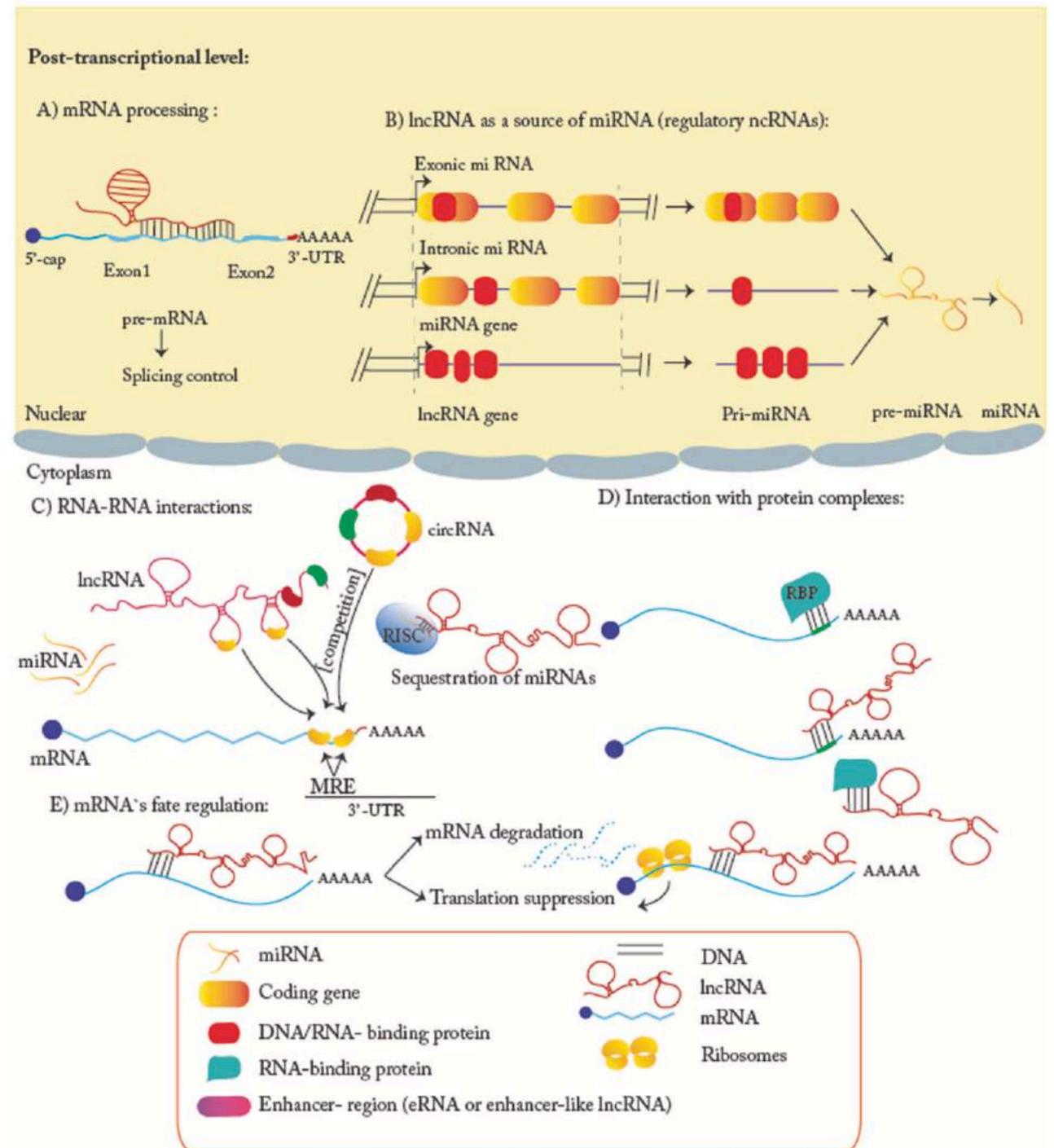


Ρύθμιση της μεταγραφής από lncRNAs

- Παροδική έκφραση lncRNAs σε περιοχές ενεργών ενισχυτών. Κάποια lncRNAs αποτελούν τα RNA ενισχυτές (enhancer RNAs, eRNAs), τα οποία μεταγράφονται αμφίροπα σε περιοχές ενισχυτών και η έκφραση αυτών των RNAs συσχετίζεται με αλλαγές έκφρασης γειτονικών γονιδίων. Έχει προταθεί ότι συμμετέχουν στην αναδιαμόρφωση της χρωματίνης – δηλαδή της σταθερότητας το μεταγραφικού βρόγχου DNA σε κοντινές αποστάσεις γονιδίων στόχων.
- Ανεξάρτητα lncRNAs που δρουν ως ενισχυτές. Αυτά τα lncRNAs ονομάζονται ενεργοποιητικά ncRNAs (activating ncRNAs) και μεταγράφονται σε περιοχές 1 kb ανοδικά σε κωδικοποιά γονίδια.

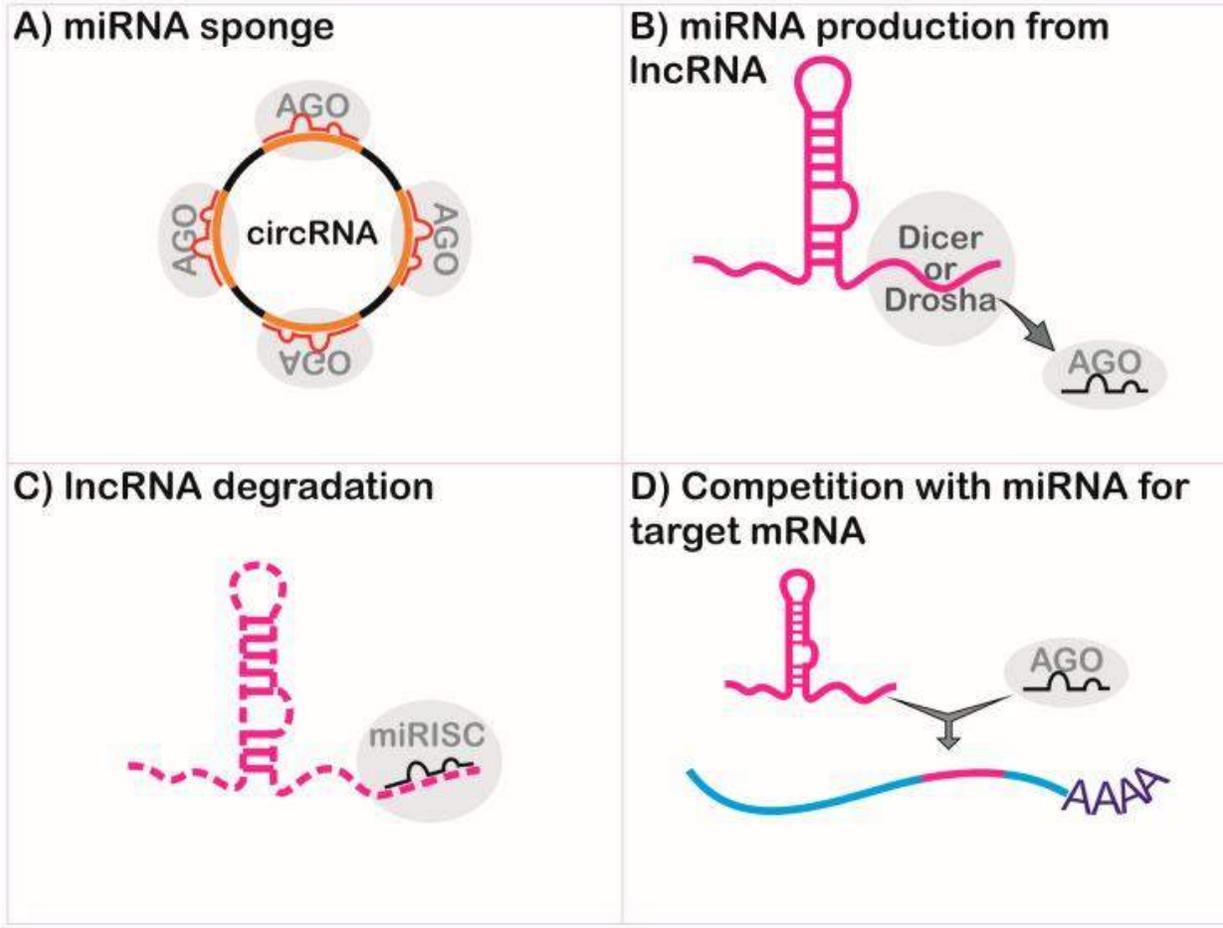
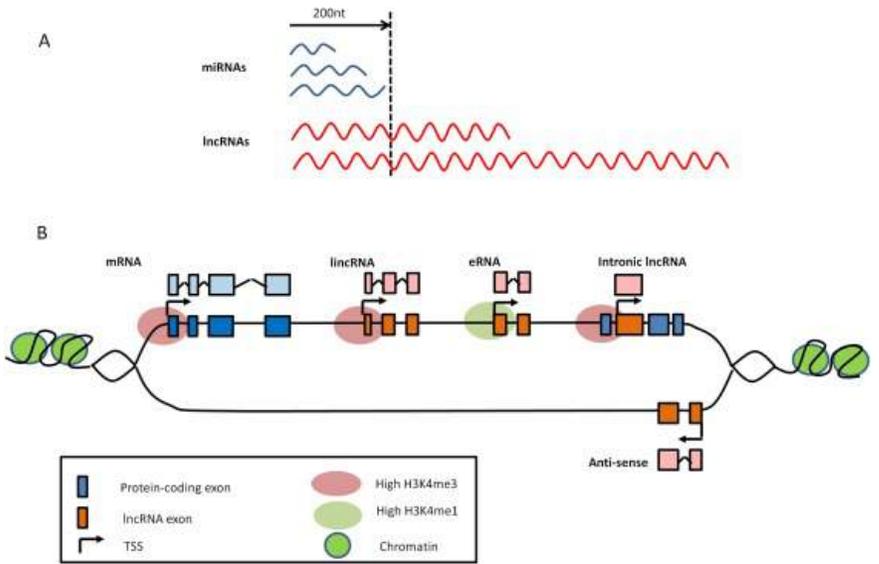
Τα lncRNAs, είναι πολυδιάστατα βιοχημικά πολυμερή: Μετά-μεταγραφικό επίπεδο

- Τα αντινοσηματικά lncRNAs μπορούν να ρυθμίσουν είτε το εναλλακτικό μάτισμα είτε τη διαδικασία επεξεργασίας/τροποποίησης RNA των πρόδρομων mRNA-στόχων τους.
- Ρυθμιστικά ncRNAs: Πολλά γονίδια lncRNA περιέχουν αλληλουχίες miRNA (κόκκινα κουτιά) είτε στα εξώνια τους (κίτρινα κουτιά) είτε στα ιντρόνια τους (γραμμές). Επιπρόσθετα, miRNAs μπορούν να δημιουργηθούν μέσω ανεξάρτητων μεταγραφικών μονάδων στο γονιδίωμα. Αυτά τα τρία μονοπάτια οδηγούν σε διαφορετικά πρωταρχικά μετάγραφα miRNAs (pri-miRNA) αλλά όλα έχουν ως συνέπεια τη βιογένεση ώριμων miRNA.
- Αλληλεπιδράσεις RNA-RNA : Η υπόθεση ενδογενώς ανταγωνιστικού RNA (competing endogenous RNA, ceRNA): Τα miRNAs συνιστούν κυρίως αναστολείς της γονιδιακής έκφρασης. Προσδέχονται στο mRNA στόχο τους μέσω στοιχείων DNA με το όνομα αλληλουχίες MRE (miRNA response element, οβάλ σχήματα) οι οποίες λαμβάνουν χώρα κυρίως στην περιοχή 3'UTR (untranslated region) του mRNA. Παρόλα αυτά MREs υπάρχουν επίσης και σε άλλους τύπους RNA όπως lncRNAs, circRNAs, ψευδογονίδια και ανεξάρτητα/εναλλακτικά μετάγραφομενα 3'UTR mRNAs . Όλοι αυτοί οι τύπου των ncRNAs που περιέχουν MRE είναι εδνογενείς ανταγωνιστές των mRNAs ως προς την πρόσδεση και κατακράτηση miRNAs (called ceRNA) και επομένως ρυθμίζουν εν δυνάμει θετικά τη γονιδιακή έκφραση.
- Αλληλεπίδραση με πρωτεϊνικά σύμπλοκα. Τα lncRNAs μπορούν να καθορίσουν τη μοίρα των mRNA, ρυθμίζοντας την δέσμευση τους από πρωτεΐνες πρόσδεσης του RNA (RNA-binding protein, RBPs) και επιδρώντας έτσι στη σταθερότητά τους, μεταφορά και τοποθέτησή τους μέσα στο.
- Τα αντινοσηματικά lncRNAs μπορούν να επηρεάσουν τη μετάφραση των mRNA στόχων και τη σταθερότητα στο κυτταρόπλασμα.



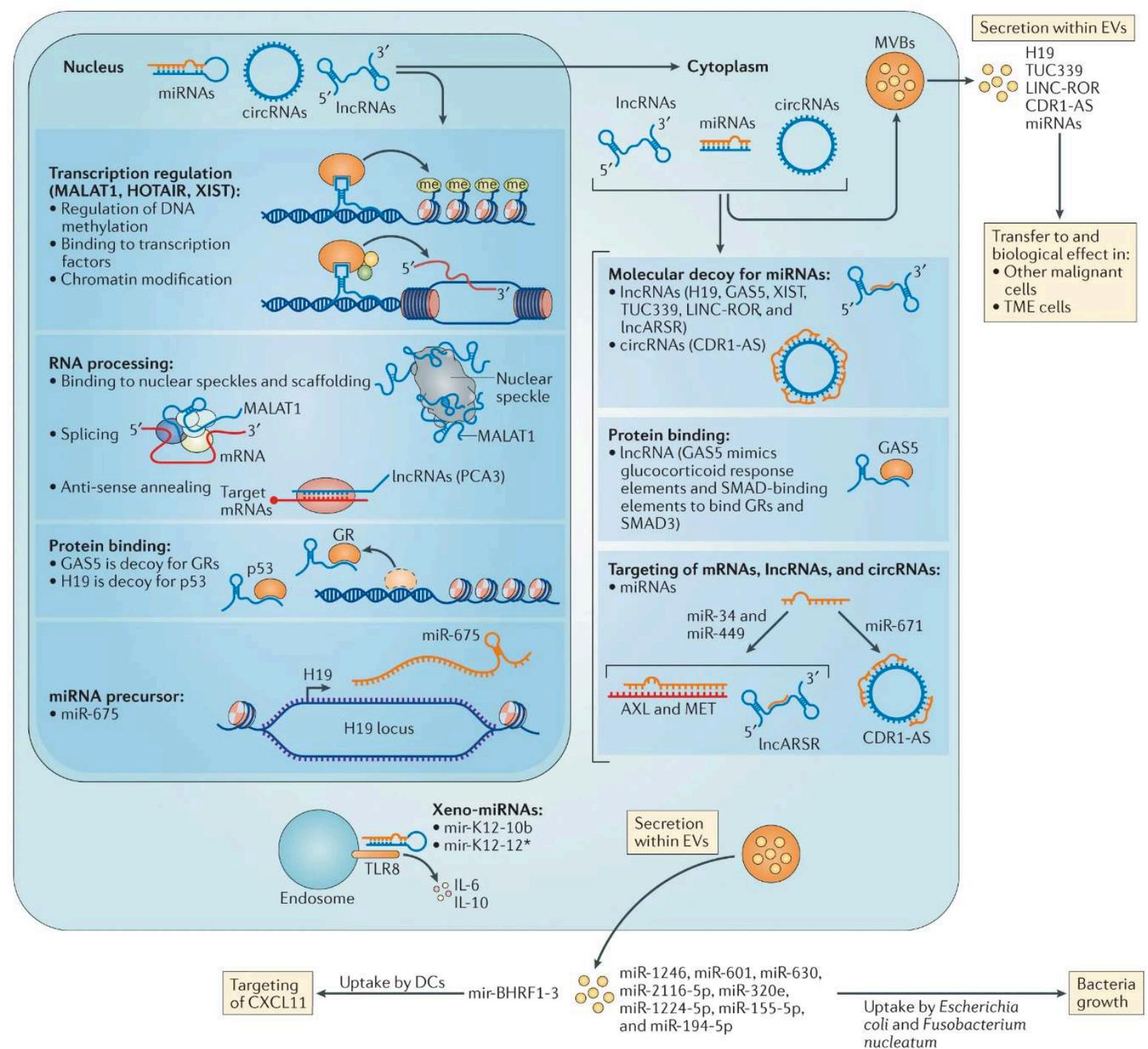
Μετα-μεταγραφική ρύθμιση από lncRNAs

- Τα lncRNAs ως ρυθμιστές της επεξεργασίας των mRNA
- Αλληλοεπιδράσεις lncRNA και microRNA
- lncRNAs ως πηγή miRNAs
- lncRNAs as δολώματα miRNA (ceRNAs)
- lncRNAs as miRNA-independent regulators of mRNA
- Circular RNAs



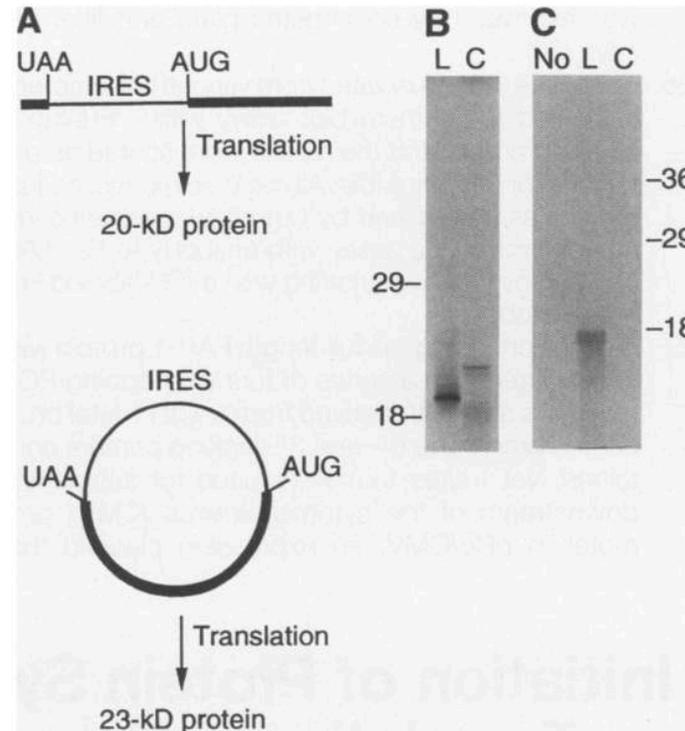
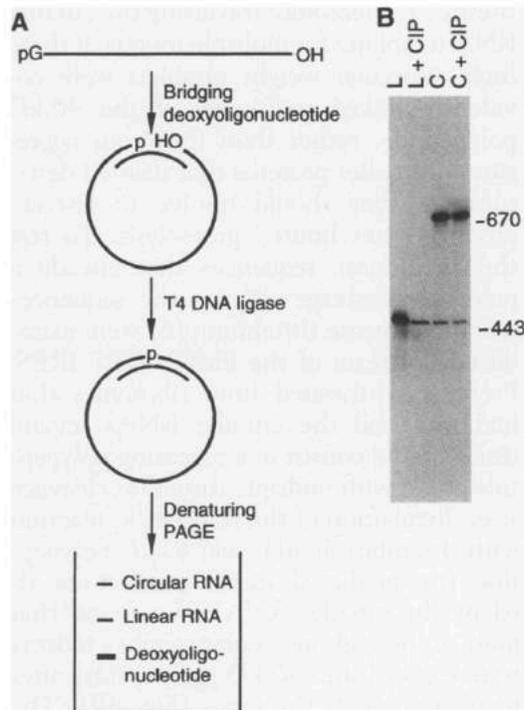
Περίληπτικά...

- Biological functions of lncRNAs
- Control of chromatin architecture
- Enhancer action
- Formation of biomolecular condensate



Circular RNAs (CircRNAs)

mRNA maturation is a seamlessly integrated process of transcription, splicing, capping, polyadenylation, export and surveillance. Back-splicing of exons for circRNA formation is an unusual type of alternative splicing. About 25 years ago, circular RNAs (circRNAs) engineered with an internal ribosome entry site (IRES) to make possible cap-independent translation were shown to undergo translation into proteins.



Biogenesis of circRNAs

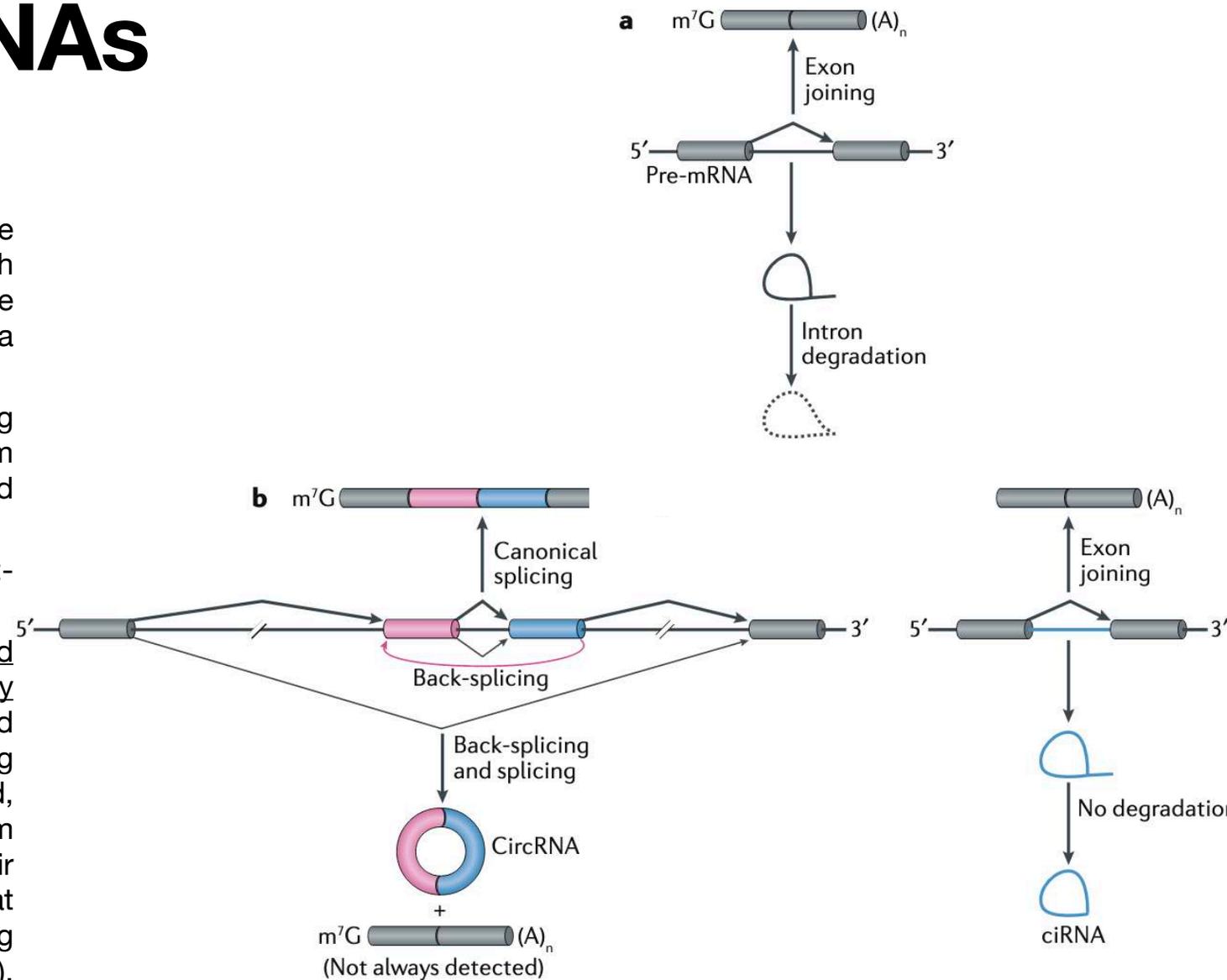
Back-splicing

Many protein-coding genes in higher eukaryotes can produce circular RNAs (circRNAs) through back-splicing of exons, in which a downstream 5' splice site is joined to an upstream 3' splice site in reverse order across an exon or exons, to form circRNAs with a 3',5'- phosphodiester bond between the back-spliced exons.

Alternatively, intron lariats excised during conventional splicing can sometimes escape debranching and retain a circular form with a 2',5'-phosphodiester bond between the splice donor and the branch point.

Back-splicing occurs both co-transcriptionally and post-transcriptionally.

Back-splicing is carried out by the spliceosome and is regulated by cis-complementary sequences in flanking introns and by specific proteins. The efficiency of back-splicing is low compared with that of canonical splicing, but the kinetics of back-splicing can be altered in some conditions (e.g. in stress). Once produced, circRNAs are generally stable, largely exported to the cytoplasm and tend to have different structural conformations from their cognate linear RNAs. Ongoing investigations have revealed that circRNAs can regulate gene expression by modulating transcription and splicing, titrating micro- RNAs (miRNAs), interacting with proteins and acting as templates for the synthesis of polypeptides



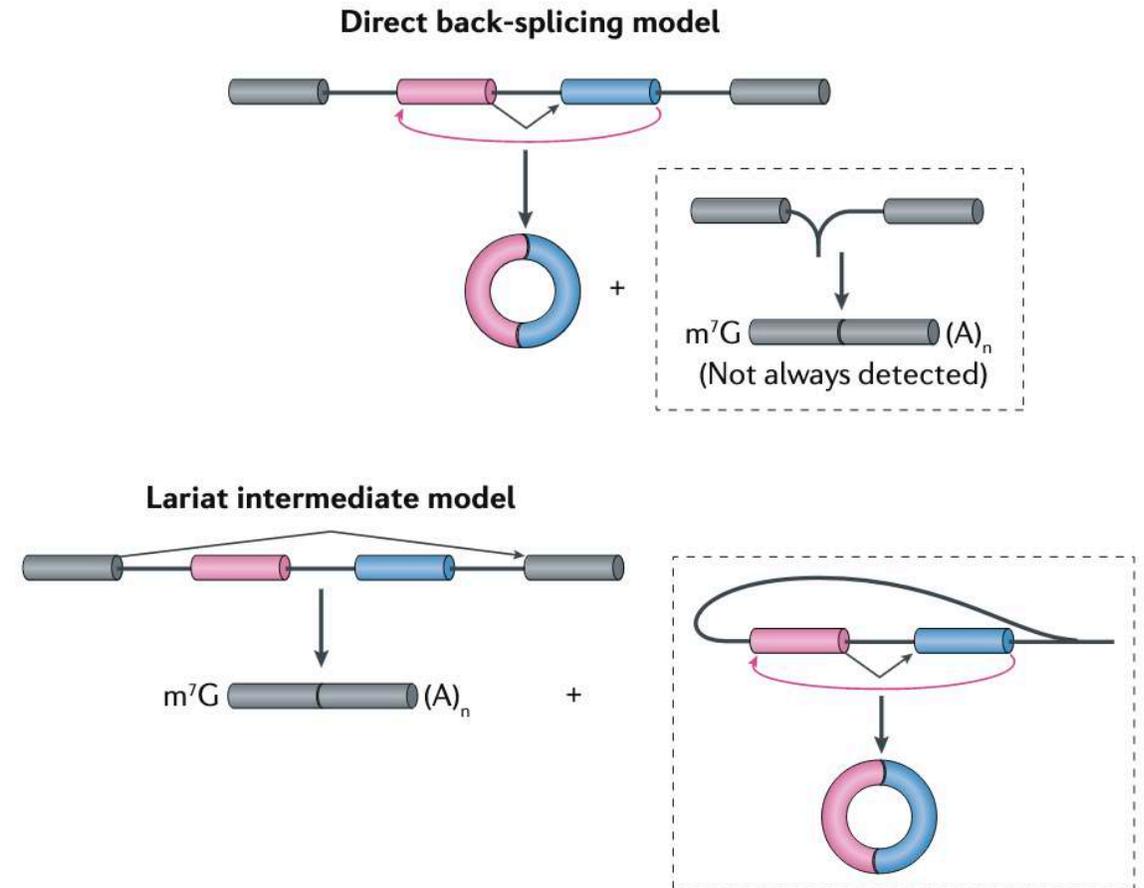
Back-splicing vs canonical splicing

Two proposed models of how back-splicing is coupled with canonical splicing.

-In the direct back-splicing model (left), back-splicing occurs first and generates a circRNA, followed by the production of linear RNA from an exon–(multiple) intron–exon intermediate (box).

-In the lariat intermediate model (right), canonical splicing occurs first and produces a linear RNA lacking introns, and a long intron lariat containing the skipped exons that subsequently undergoes back-splicing

Back-splicing often occurs in middle exons of annotated genes and theoretically should affect the splicing of linear RNAs. However, the efficiency of back-splicing is much lower than that of canonical splicing (less than 1%) in most endogenous human loci, as revealed by measurement of the levels of nascent circRNAs. Nevertheless, it has been suggested that the use of 5' and 3' splice sites for back-splicing can compete with linear RNA splicing, thereby increasing the proportion of linear RNAs lacking the circularized exons. Although not all alternatively skipped exons can produce circRNAs, it has been observed that the more back-splicing an exon undergoes, the less it is included in the fully processed mRNA⁵¹; for example, in the rat cytochrome P450 2C24 gene



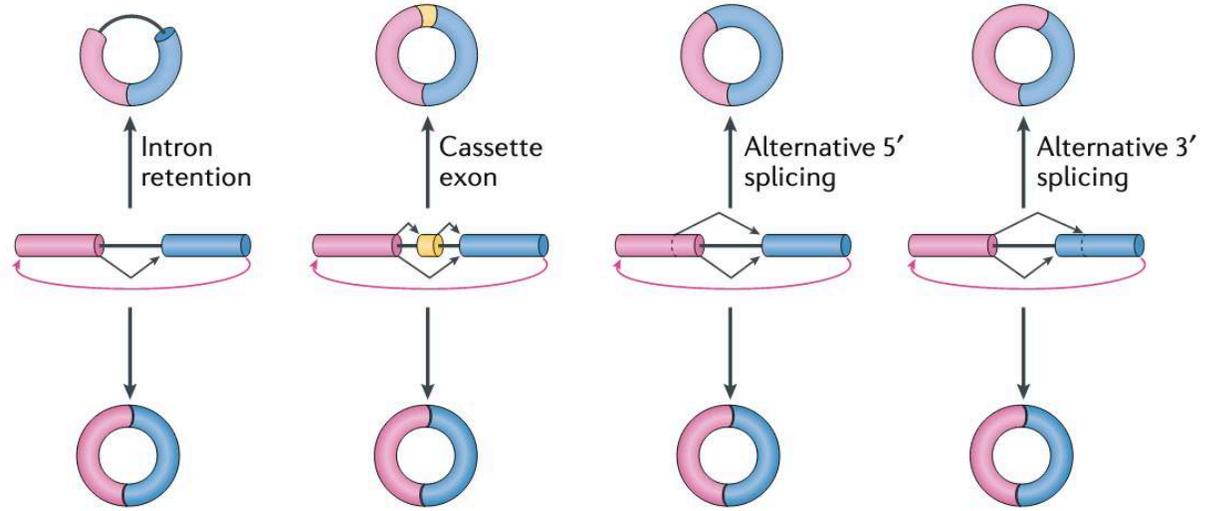
Alternative back-splicing and alternative splicing (with the same back-splice site) of circRNAs

Alternative back-splice site selection produces different single-exon circRNAs, in parallel with the common two-exon circRNAs produced from the same pre-mRNA.

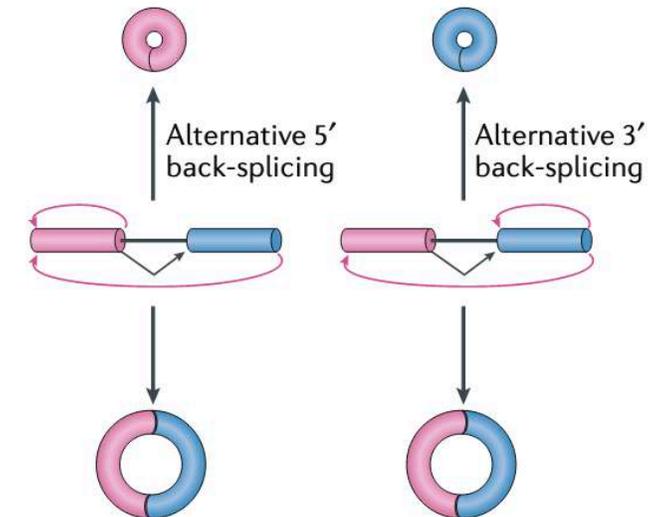
Four basic types of alternative splicing in circRNAs that contain multiple exons.

There are at least three unique aspects of circRNA-specific alternative splice site selection compared with linear RNA splicing. First, production of circRNAs can somehow bypass the canonical splice sites used in the production of their cognate mRNAs; for example, leading to the production of two circRNA isoforms that either include or do not include a retained intron in the CAMSAP1 gene. Second, thousands of previously unannotated cassette exons are included in circRNAs by alternative (back-) splice site selection; these exons are frequently undetectable in the cognate mRNAs. Many such circRNA-specific exons are less conserved and have not been annotated. Third, circRNA-specific cassette exons appear to be highly dynamically regulated. For example, inclusion of some cassette exons is significantly increased in circRNAs, such as those produced from the XPO1 gene, during the differentiation of human embryonic stem cells (ES cells) to neuronal lineages, but are absent in linear RNAs.

f Alternative splicing within circRNAs (with the same back-splice site)



e Alternative back-splicing



Abundance of circRNAs

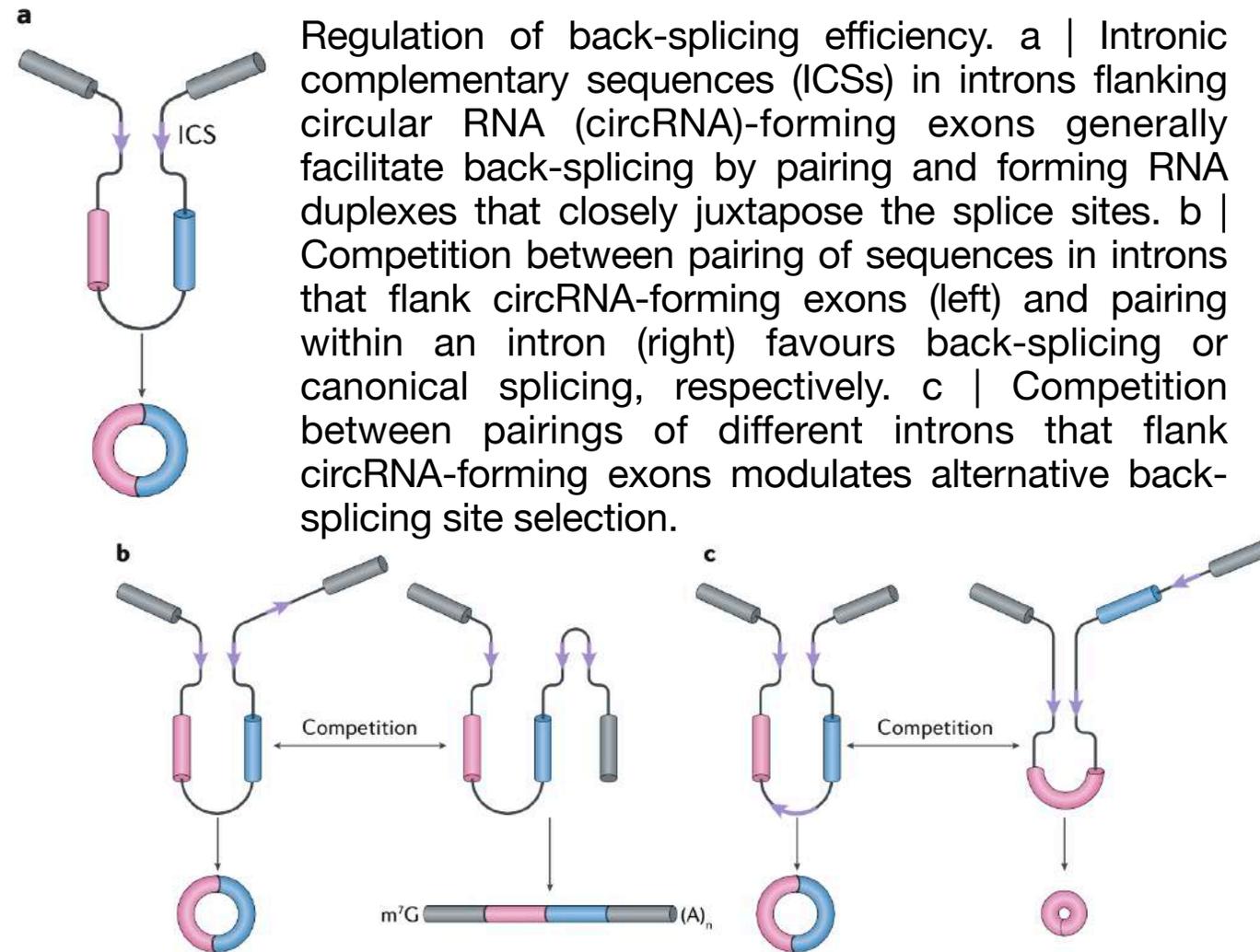
Thousands of novel regulators and polypeptides

Owing to the lack of 3' polyadenylated tails, most circRNAs were not detected in the classical RNA sequencing (RNA-seq) datasets, which mainly contain polyadenylated RNAs. More recently, analyses of non-polyadenylated transcriptomes and RNase R-treated transcriptomes have uncovered how wide-spread the expression of circRNAs is in metazoans, from nematodes, zebrafish and fruitflies to mice, pigs, monkeys and humans as well as in protists, fungi and plants.

The most recent analyses using computational pipelines to identify back-splice junction sites from RNA-seq reads or reconstructing full-length circRNAs from long RNA-seq reads have identified more than 183,000 circRNAs from human transcriptomes, more than 96,000 circRNAs from macaque transcriptomes and more than 82,000 circRNAs from mouse transcriptomes. In these studies, ~1,334 Gb from 70 human samples, ~443 Gb from 26 macaque samples and ~512 Gb from 28 mouse samples were analysed.

Intronic complementary elements (ICSs) facilitate back-splicing

Although, in general, ICSs boost circRNA formation, their presence is neither sufficient nor necessary for it. First, RNA pairing in the same intron competes with RNA pairing between flanking introns, thereby facilitating splicing of linear RNA and reducing circRNA production from the same locus. Second, multiple RNA pairings can presumably form between different sets of introns in the same gene, and this competition can lead to alternative back-splicing and the formation of different circRNAs. Third, repetitive elements are enriched in mammalian genomes, but in other organisms ICSs appear to have become less crucial for circRNA formation

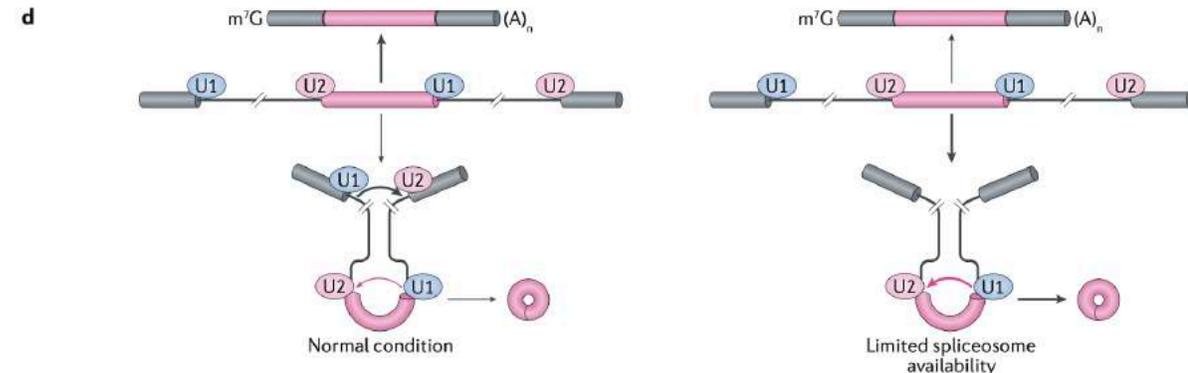
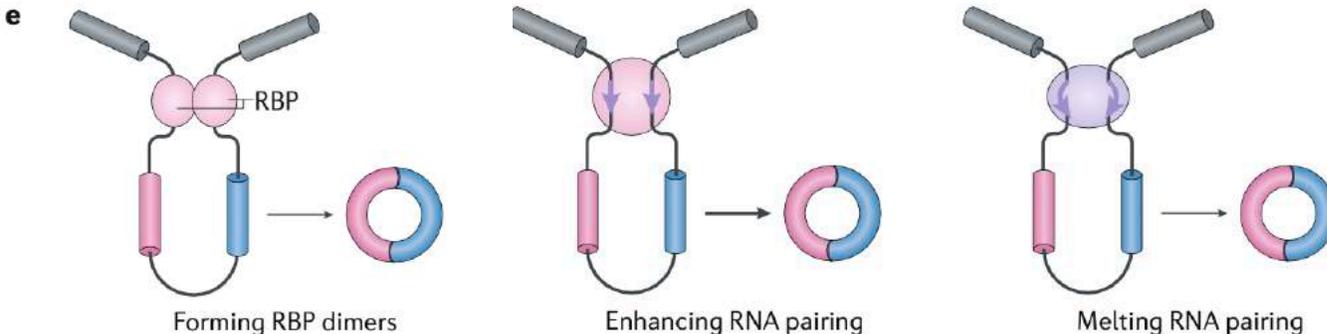


Intronic complementary elements (ICSs) facilitate back-splicing

Regulation of back-splicing in trans. Splicing and alternative splicing are regulated also by proteins involved in spliceosome function and by RNA-binding proteins (RBPs) that bind to cis elements in flanking introns. In addition to the role of ICSs in modulating back-splicing, a number of RBPs control circRNA biogenesis.

RNA-binding proteins (RBPs) can directly bridge distal splice sites to promote back-splicing. f | RBPs bind to ICSs to enhance or suppress back-splicing by enhancing (left) or reducing (right), respectively, the pairing capacity of ICSs.

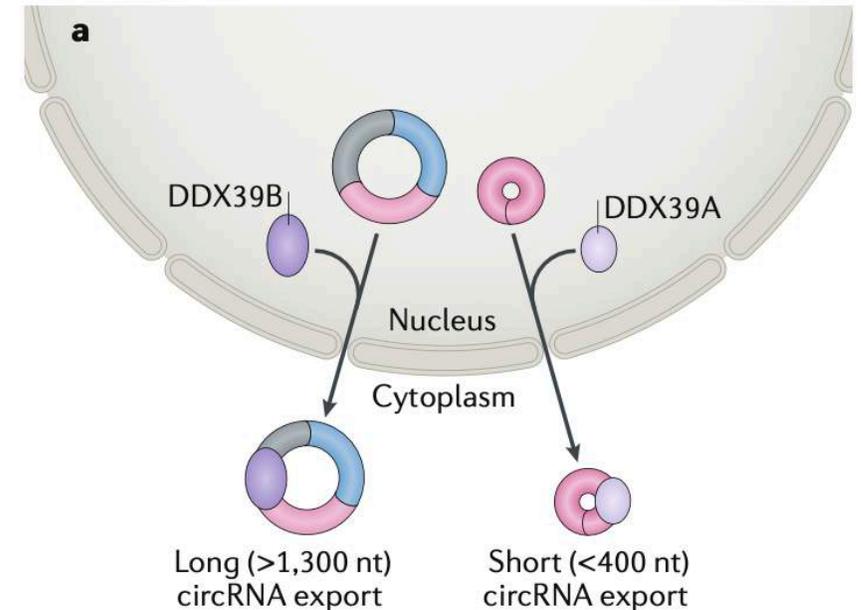
d | Conditions that limit spliceosome availability cause a shift from canonical splicing (left), to back-splicing that produces single-exon circRNAs (right), presumably by reducing the formation of cross-intron exon definition complexes (not shown)



Nuclear export of circular RNAs.

Similarly to many linear mRNAs⁶⁷, circRNAs containing retained introns are frequently sequestered in the nucleus^{23,68,69}. However, the majority of annotated back-splicing circRNAs are primarily localized in the cytoplasm. How nuclear export or retention of circRNAs is regulated remained unknown until the recent publication of a study that aimed to identify key factors controlling circRNA export using RNAi screening. This work nicely showed that depletion of the ATP-dependent RNA helicase Hel25E (also known as WM6) in *D. melanogaster* cells led to specific accumulation of long (more than 800 nucleotides) circRNAs in the nucleus. Likewise, depletion of DDX39B or DDX39A (the two human homologues of Hel25E) led to the accumulation of long (more than 1,300 nucleotides) or short (fewer than 400 nucleotides) circRNAs in the nucleus, respectively.

These results suggest that the lengths of circRNAs are somehow involved in the mode of their export



The turnover of circular RNAs.

Back-splicing is inefficient, but circRNAs can accumulate to high levels owing to their circular structure, which makes them resistant to degradation by RNA decay machineries. CircRNAs are more stable than their cognate linear transcripts, with the examined circRNAs having a median half-life ranging from 18.8 to 23.7 hours, compared with 4.0–7.4 hours for their cognate linear RNAs.

Regardless of the innate stability of circRNAs, recent studies have shown they undergo degradation in normal conditions and in stress conditions.

(a) CircRNA binding by miRNAs can initiate circRNA decay, as exemplified by the circRNA CDR1as

(b) The cytoplasmic endonuclease RNase L, which is activated during viral infection by the presence of pathogenic dsRNA, is known to cleave linear RNAs after UN dinucleotides and can globally degrade circRNAs

(c) A subset of m⁶A-modified circRNAs can be degraded by the ribonuclease complex RNase P/MRP; the degradation is mediated by the m⁶A reader protein YTHDF2 and HRSP12.

(d) Highly structured circRNAs can be targeted and degraded by the ATP-dependent RNA helicase upstream frameshift 1 (UPF1) and its associated endonuclease G3BP1 (ReF.83) (Fig. 3e). UPF1 is a key factor of nonsense-mediated mRNA decay and of other, functionally diverse mRNA decay pathways.

