UNIVERSITÀ
DEGLI STUDI DI TRIESTE

DIPARTIMENTO DI
SCIENZE DELLA VITA

Distinctive patterns of transcription and RNA processing for human lincRNAs

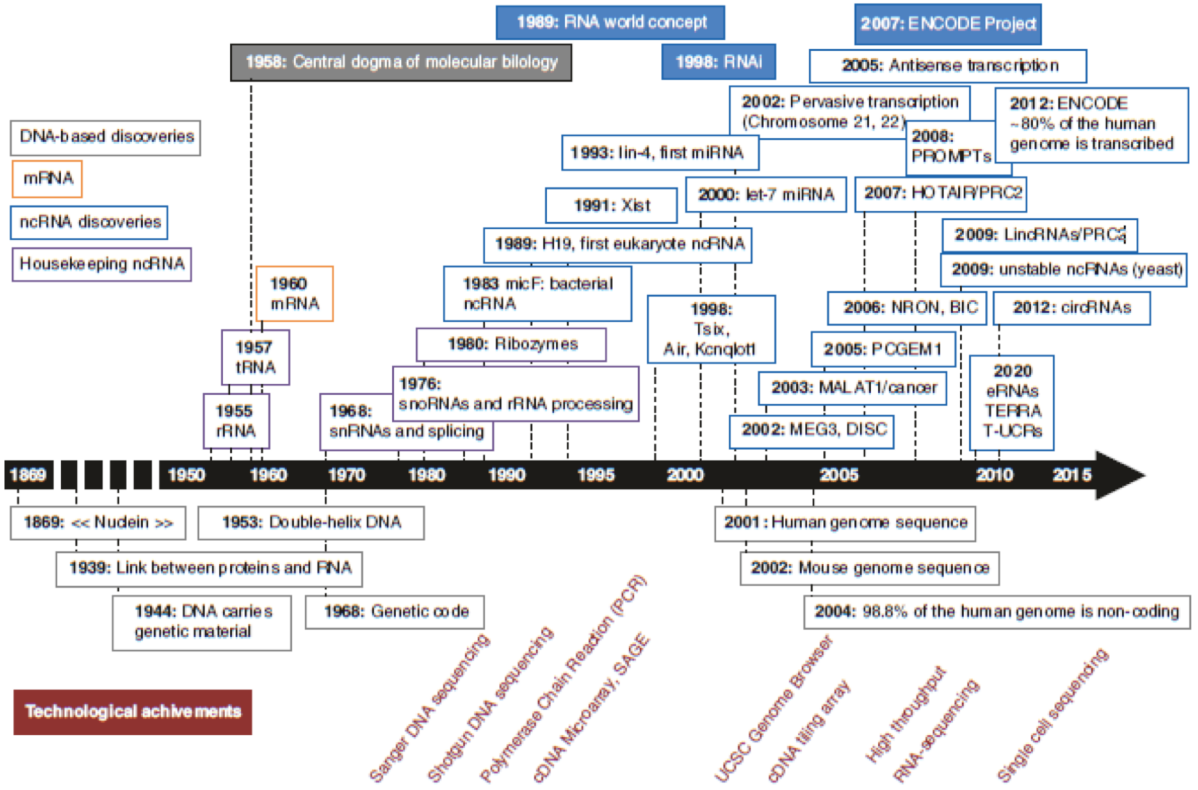
Gabriele Muscia

Matteo Furchì

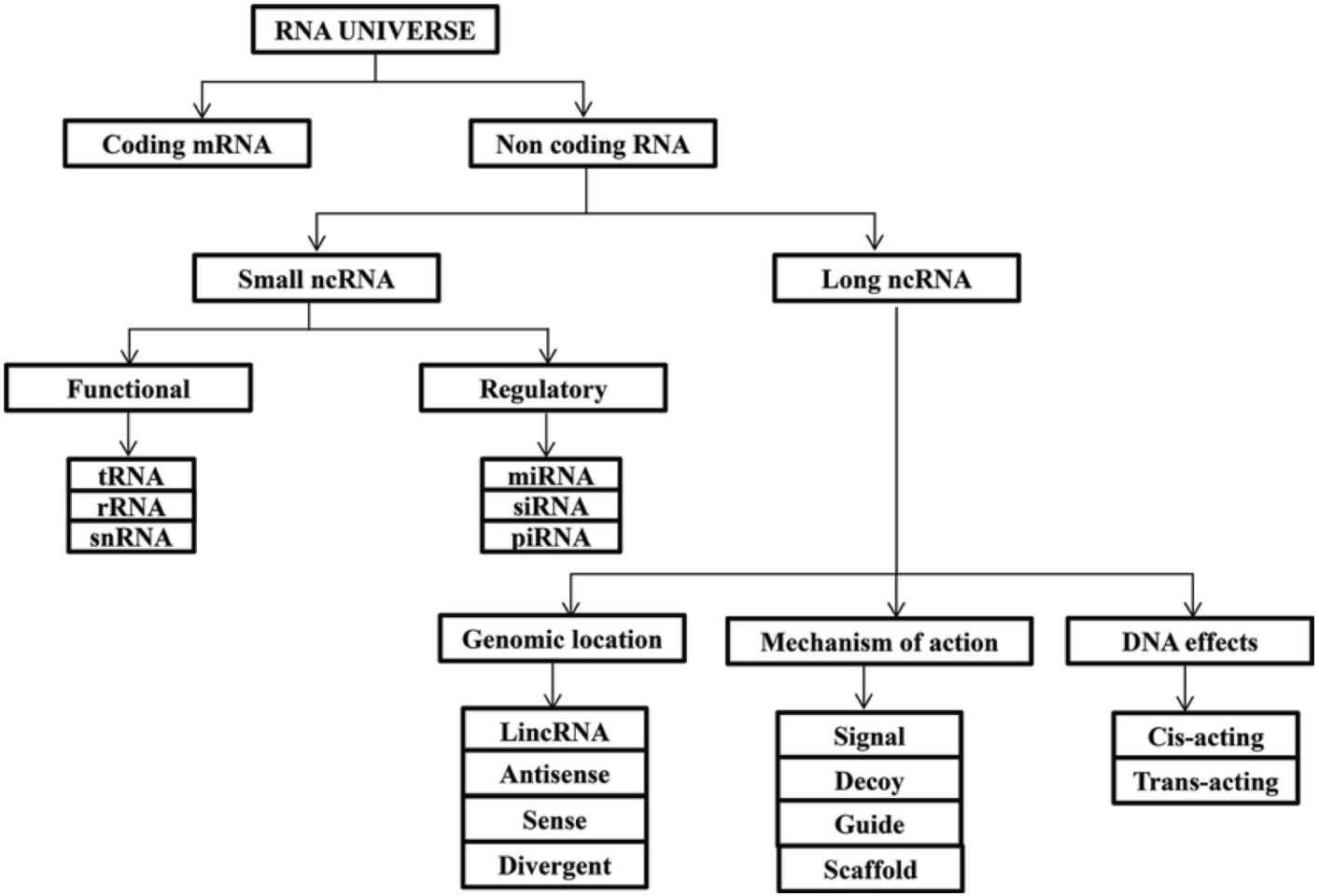
Francesco Radin

History

- 1955: first ncRNA housekeeping: rRNA
- 1983: first sncRNA: micF
- 1989: first lncRNA: H19 + RNA world concept
- 1991: XIST
- 2001: HGP completed
- 2004: Only 1,2% of the human genome codes for proteins



non-coding RNA



lincRNAs functions

Chromatin topology: gene transcription regulations

Scaffolding and modulating the activity of proteins and RNAs

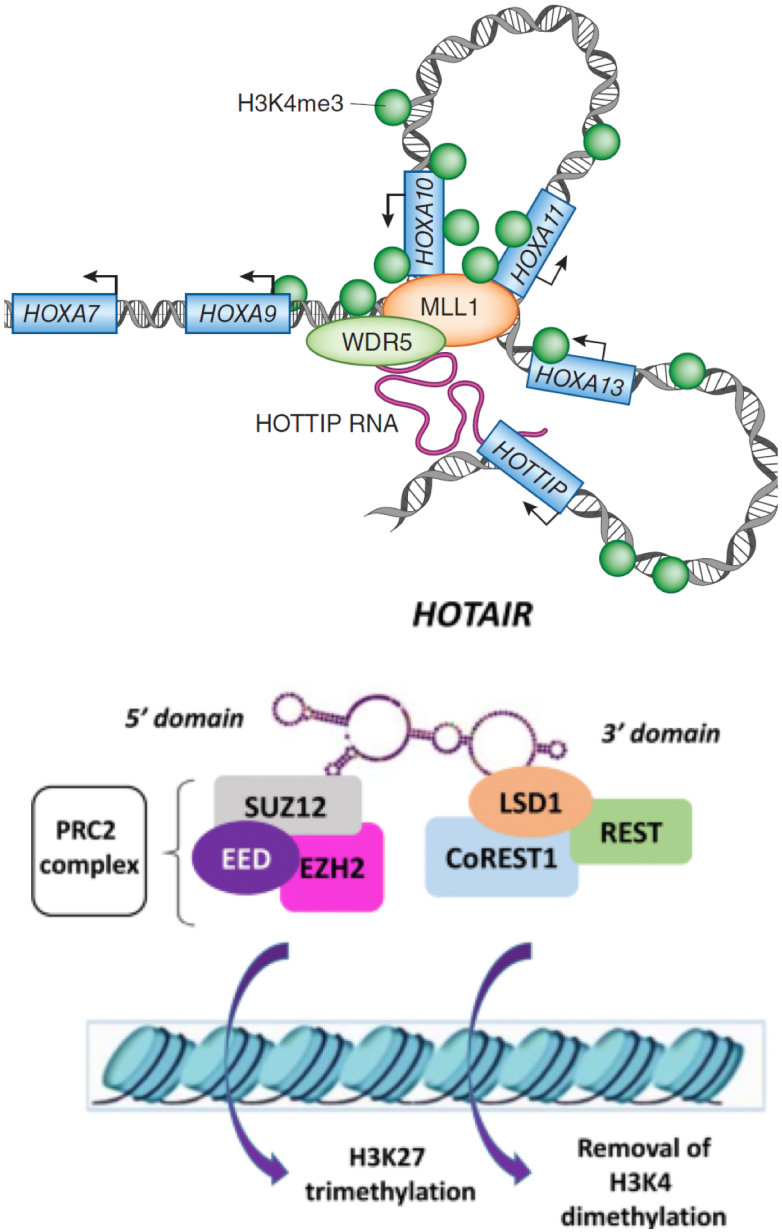
Protein and RNA decoy

Encoding functional micropeptides

lincRNAs functions

Chromatin topology: gene transcription regulations

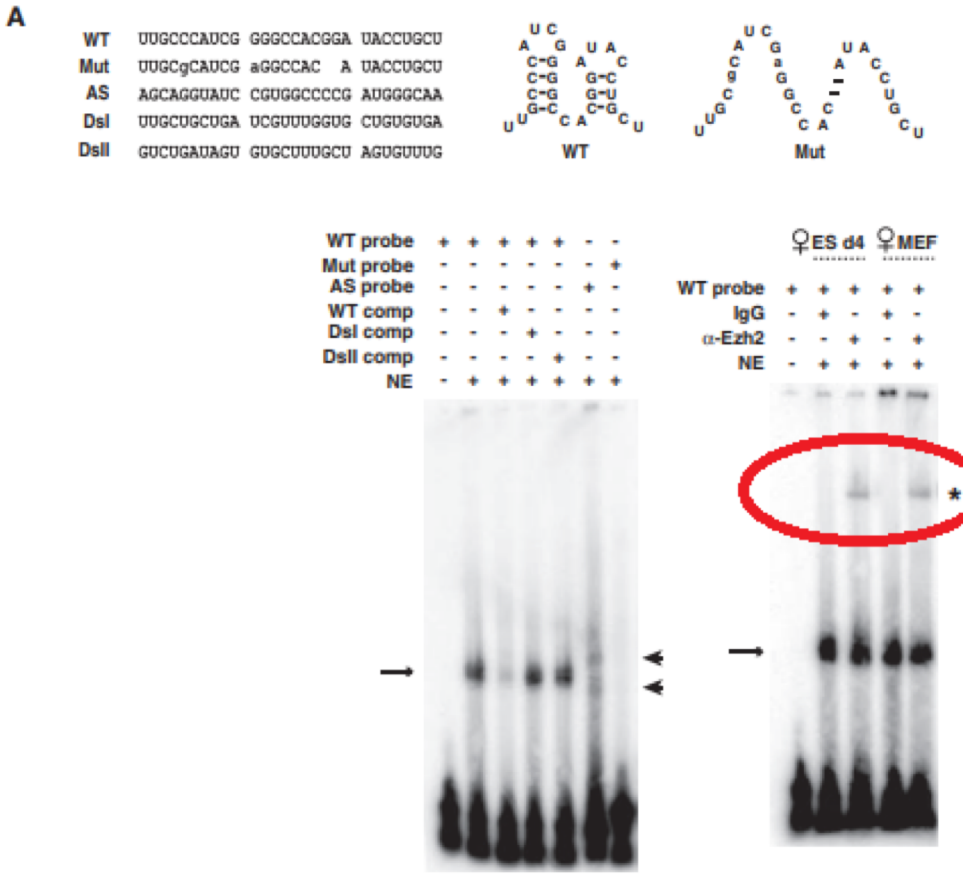
Epigenetic markers derive from the recruitment of methyl-transferase binded to lincRNA such HOTTIP pr HOTAIR



lincRNAs functions

Scaffolding and modulating the activity of proteins and RNAs (RepA-EZH2)

Interaction between RepA and PRC2 demonstrated by an Electophoretic Mobility Supershift Assay



lincRNAs functions

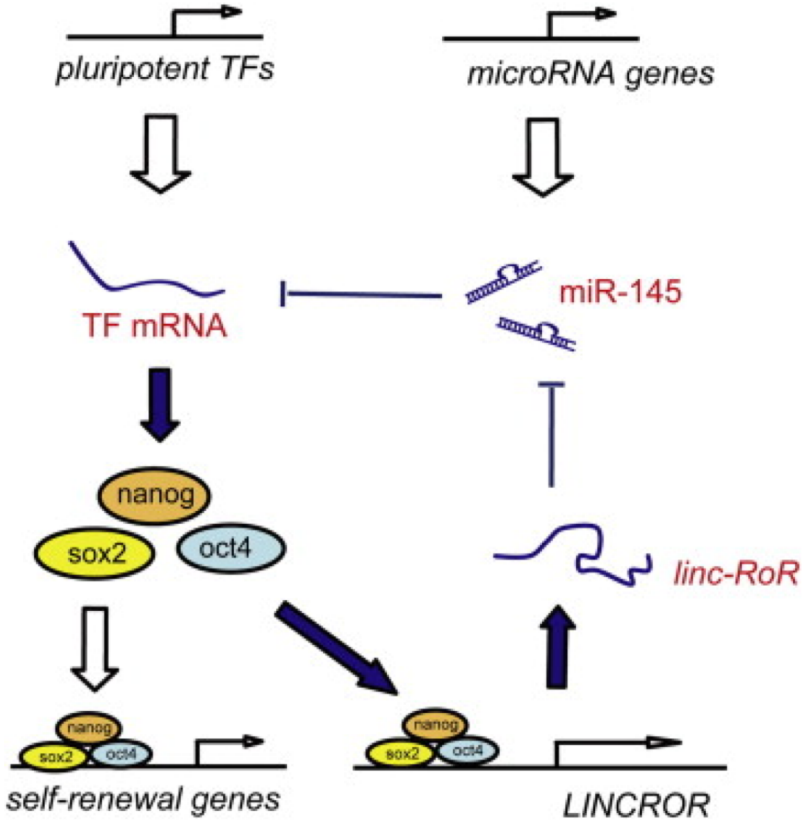
Protein and RNA decoy

Some lncRNA can act as ceRNA

linc- RoR maintains stem cell pluripotency.

In pluripotent stem cells, linc-RoR sequesters miR-145, thereby promoting the accumulation of OCT4, the transcription factor SOX2 and the homeobox protein Nanog, which are miR-145 targets.

The levels of linc-RoR decrease during differentiation, and miR-145 is released and promotes the degradation of SOX2, Nanog and OCT4 mRNAs



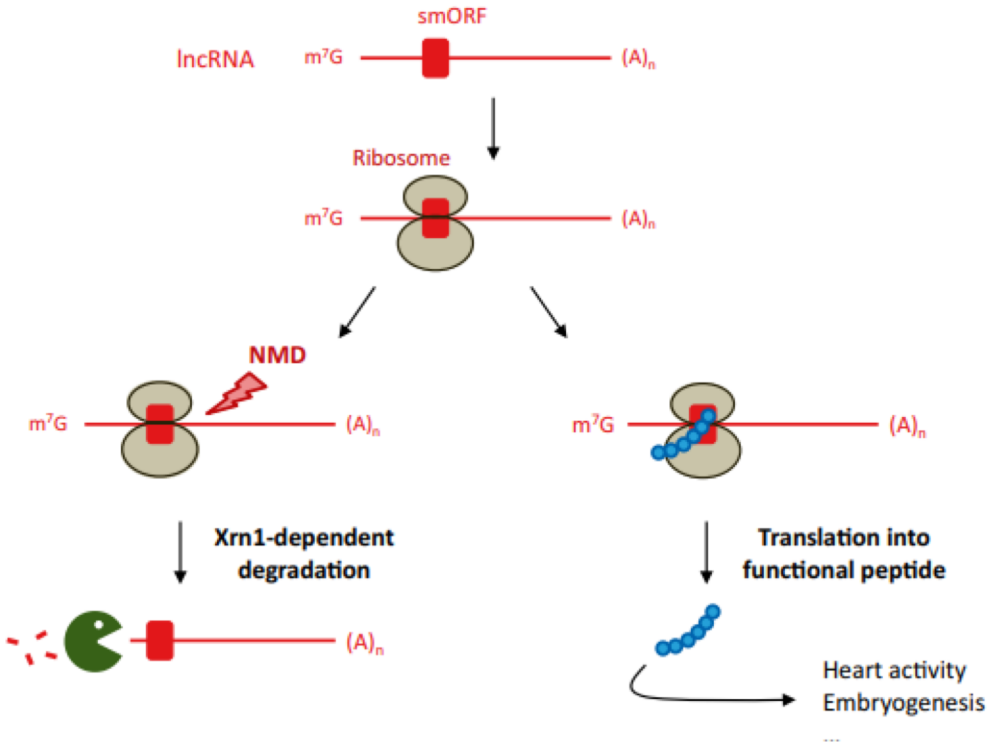
lincRNAs functions

Encoding functional micropeptides

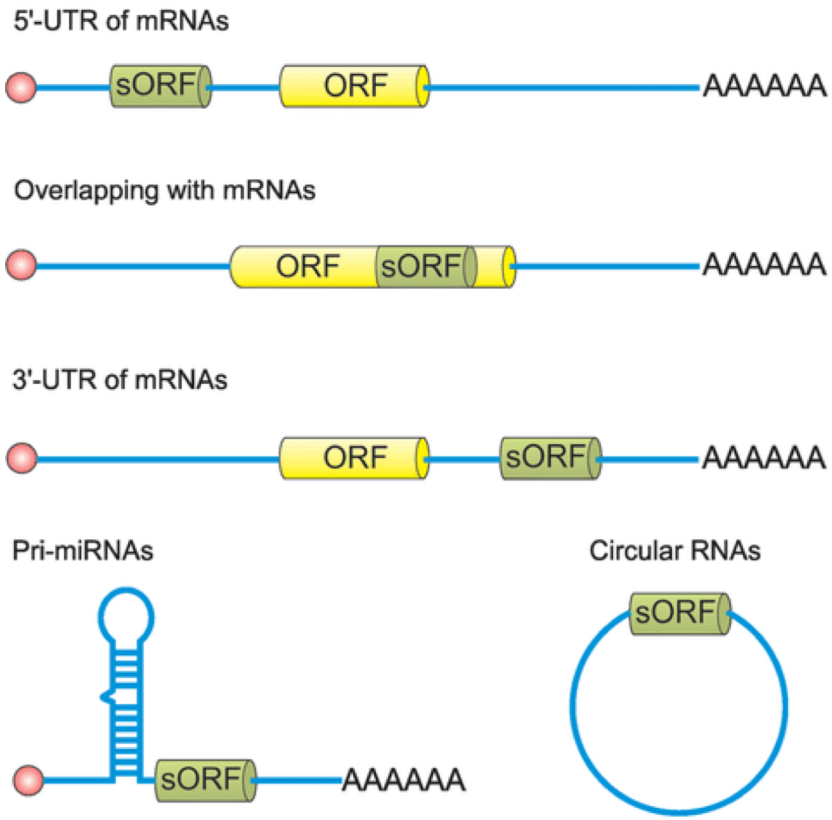
lincRNAs can contain little ORFs in their sequence. They are called smORFs.

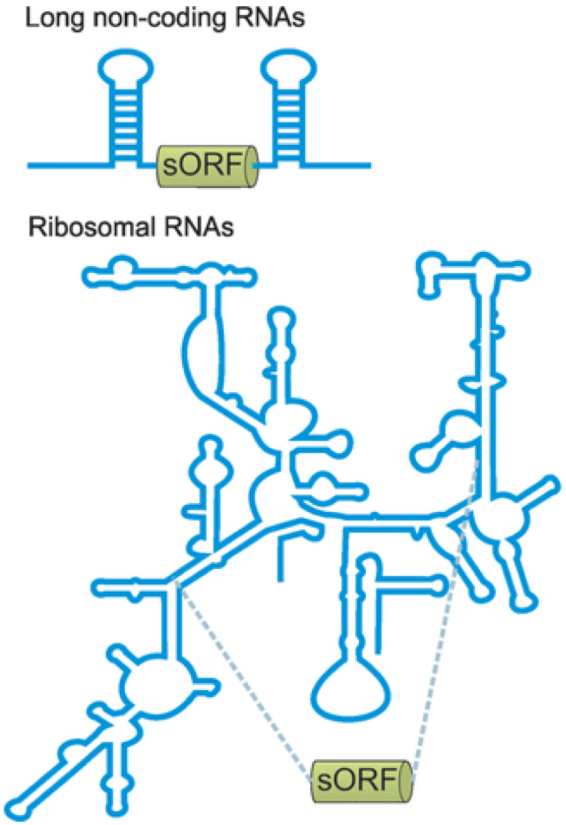
smORFs are translated in micropeptides that can be functional for another structure (SERCA).

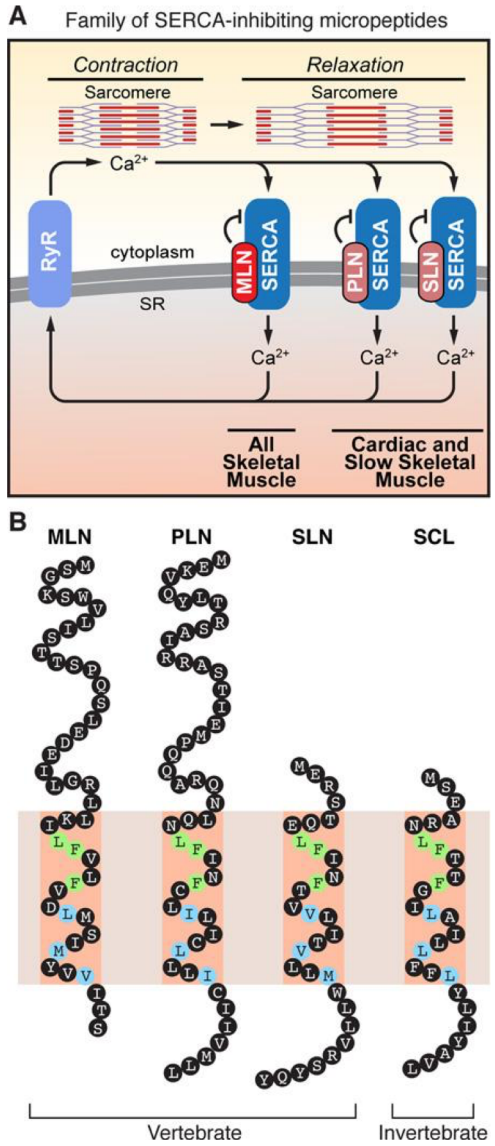
Another meaning for the existence of these smORFs is the presence of an early stop codon that activate the Nonsense-Mediated Decay



lincRNAs functions







Characterization of lincRNAs

- >200 nt
- 5’ capping (CAGE)
- 3’ polyadenylation (3P-seq)
- Different splicing from mRNA
- Epigenetic markers similar to mRNA
- Inefficiently polyadenylated

- lack of primary structure conservation despite protein-coding gene
- average of 40 Kb compared to other genes
- *NOT* coding for proteins
- *NOT* overlapping with other transcripts
- Degraded by exosomes in the nucleus

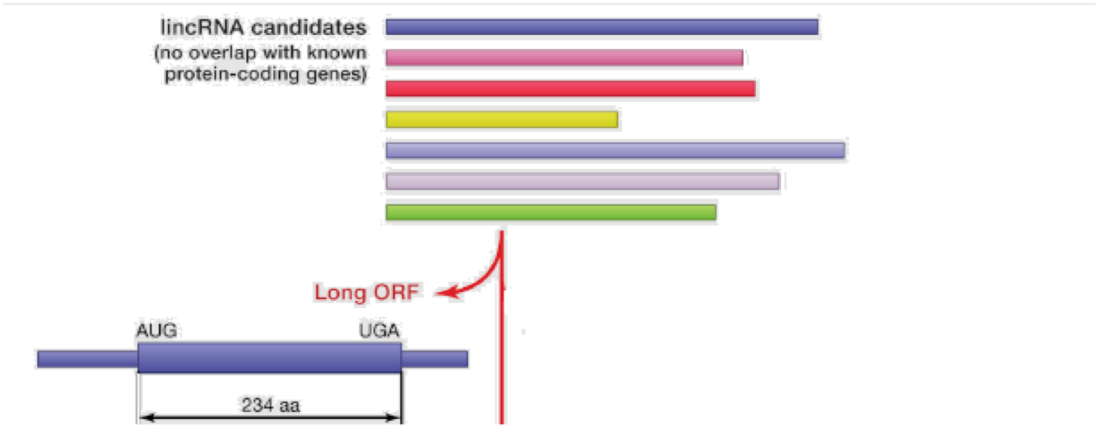
Human

Khalil et al., 2009	Chromatin marks, tiling arrays	Collection of approximate exonic regions, chromatin domain ≥ 5 kb	CSF	3,289 loci
Jia et al., 2010	cDNAs	Overlap with mRNAs allowed		5,446 transcripts
Ørom et al., 2010	cDNAs	Restricted to loci >1 kb away from known protein-coding genes, ≥200 nt mature length	Manual curation based on length, conservation and other characteristics of the ORFs	3,019 transcripts from 2,286 loci
Cabili et al., 2011	RNA-seq	Multi-exon only, ≥200 nt mature length	PhyloCSF, Pfam	8,195 transcripts (4,662 in the stringent set)
Derrien et al., 2012	cDNAs	Overlap with mRNAs allowed (intergenic transcripts reported separately), ≥200 nt mature length	Manual curation based on length, conservation and other characteristics of the ORFs	14,880 transcripts from 9,277 loci, including 9,518 intergenic transcripts
Sigova et al., 2013	RNA-seq, cDNAs, chromatin marks,	Antisense overlap with mRNA introns allowed, ≥100 nt mature length	CPC	3,548 loci from embryonic stem cells, and 3,986 loci from endodermal cells

lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

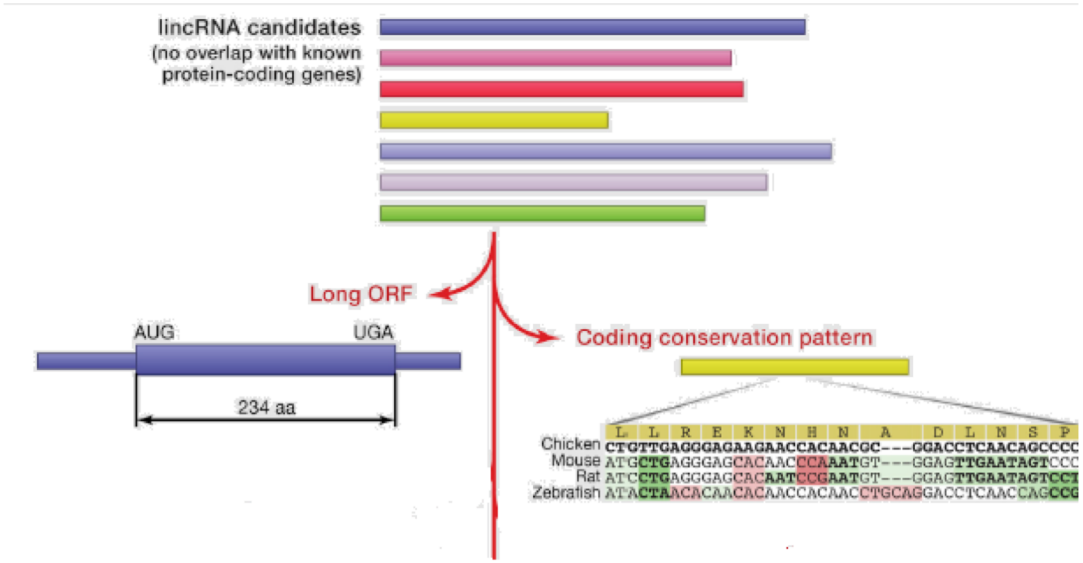
- Coding regions tend to be much longer than expected by chance



lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

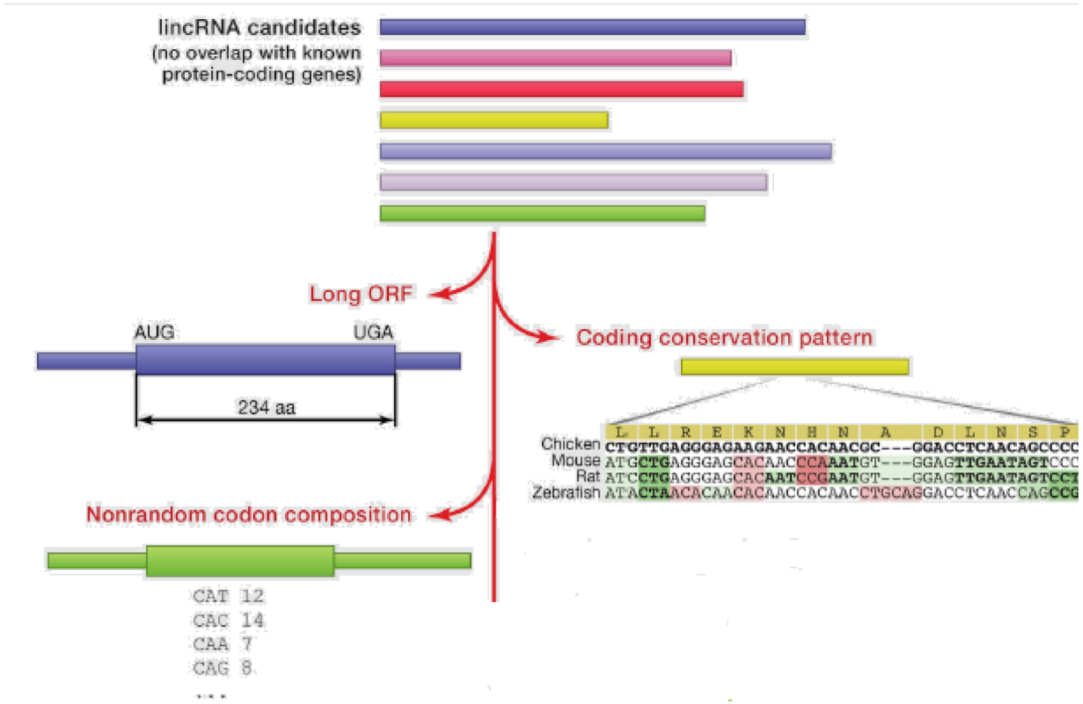
- Coding regions tend to be much longer than expected by chance
- During evolution selective pressures bias nucleotide substitutions in coding sequences



lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

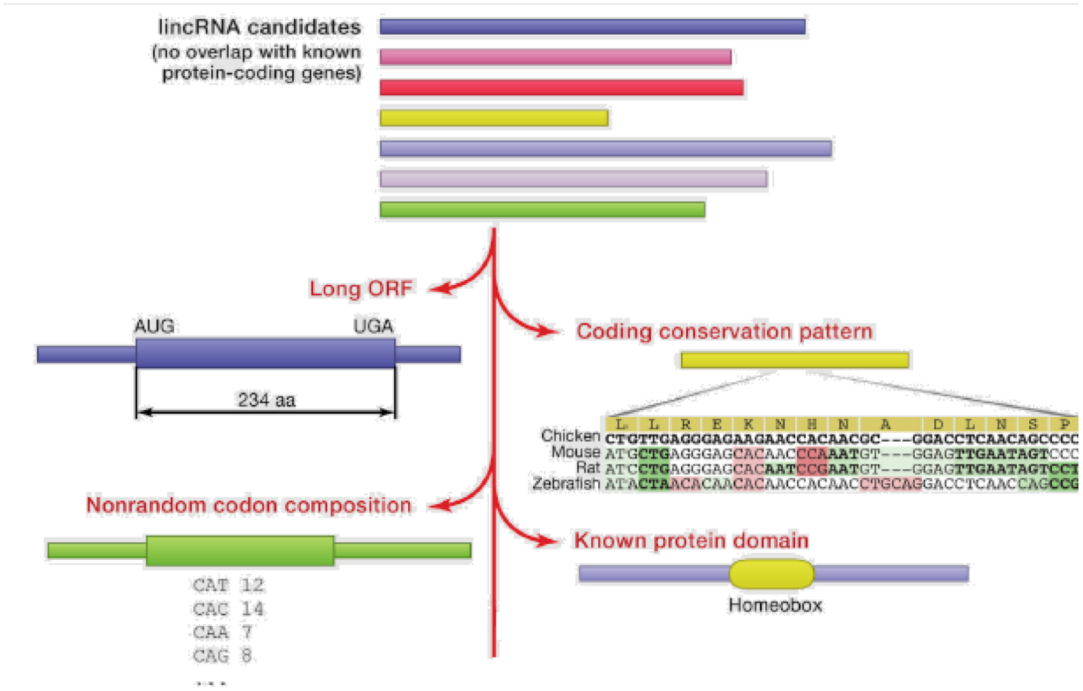
- Coding regions tend to be much longer than expected by chance
- During evolution selective pressures bias nucleotide substitutions in coding sequences
- nucleotide frequencies of functional ORFs are dictated by nonrandom codon usage



lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

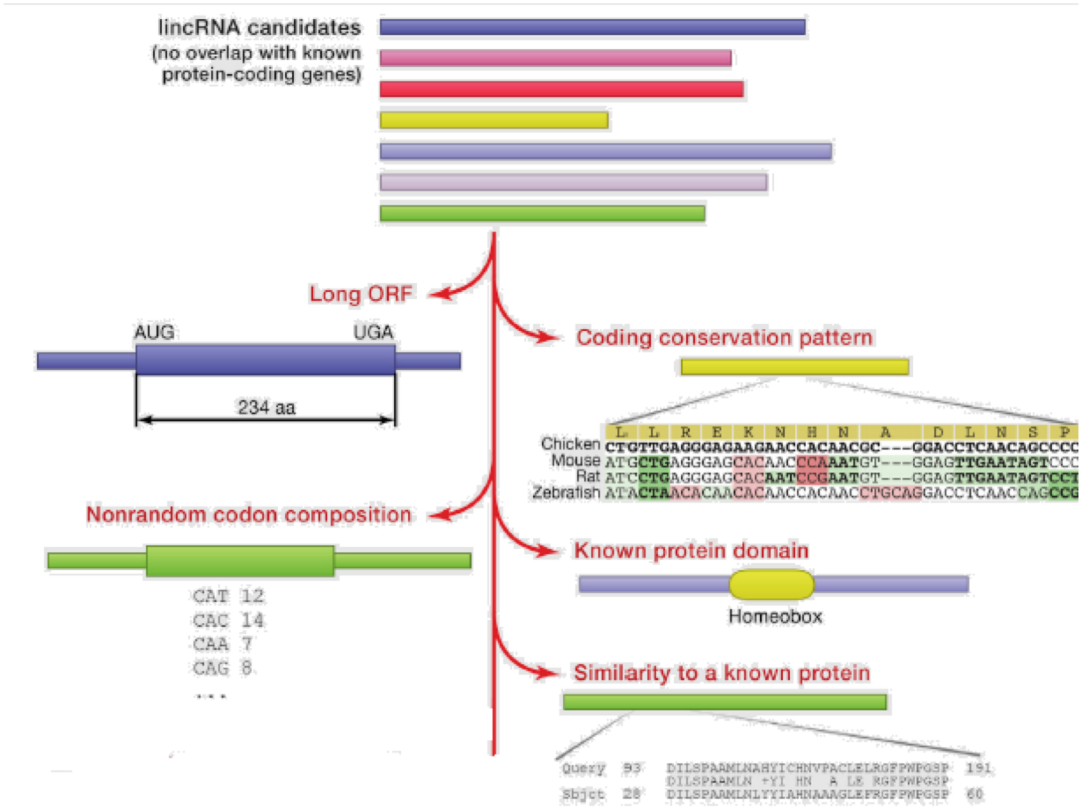
- Coding regions tend to be much longer than expected by chance
- During evolution selective pressures bias nucleotide substitutions in coding sequences
- nucleotide frequencies of functional ORFs are dictated by nonrandom codon usage
- Protein coding genes typically contain known protein domains



lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

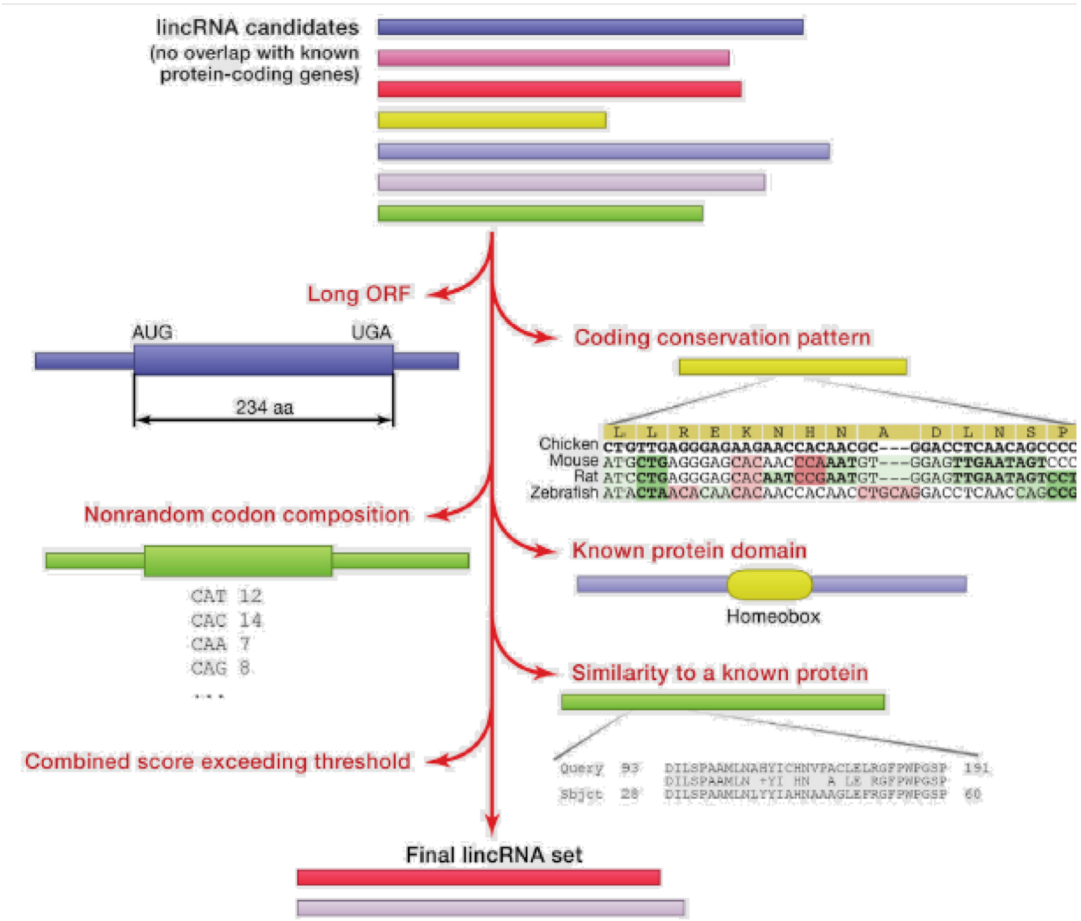
- Coding regions tend to be much longer than expected by chance
- During evolution selective pressures bias nucleotide substitutions in coding sequences
- nucleotide frequencies of functional ORFs are dictated by nonrandom codon usage
- Protein coding genes typically contain known protein domains
- Coding regions are likely to bear sequence similarities to entries in protein databases



lincRNAs (mis)identification?

ncRNA can have ORFs ⟹ imperfect criteria:

- Coding regions tend to be much longer than expected by chance
- During evolution selective pressures bias nucleotide substitutions in coding sequences
- nucleotide frequencies of functional ORFs are dictated by nonrandom codon usage
- Protein coding genes typically contain known protein domains
- Coding regions are likely to bear sequence similarities to entries in protein databases



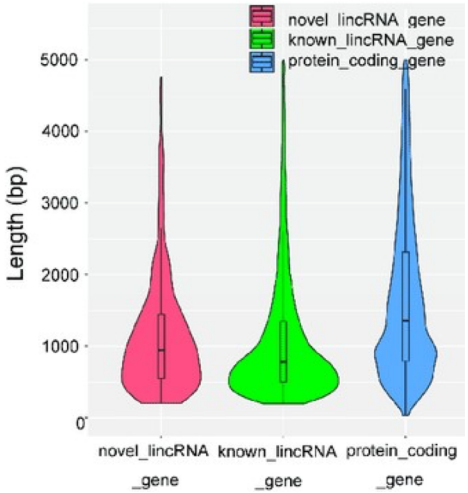
Comparison between mRNA and lincRNA: genes

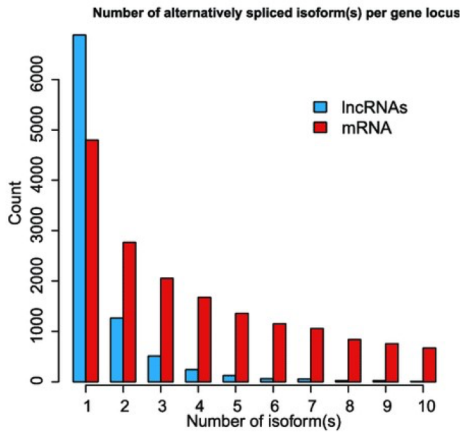
mRNA

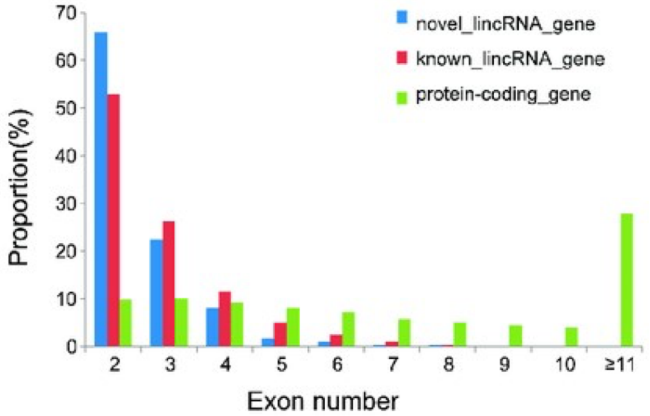
About 20.000 genes. Average of 11 exons of 3 Kb each. Higher gene density than lincRNA

lincRNA

About 13.000 genes. Average of 3 exons of 1 Kb each.







Comparison between mRNA and lincRNA: localization

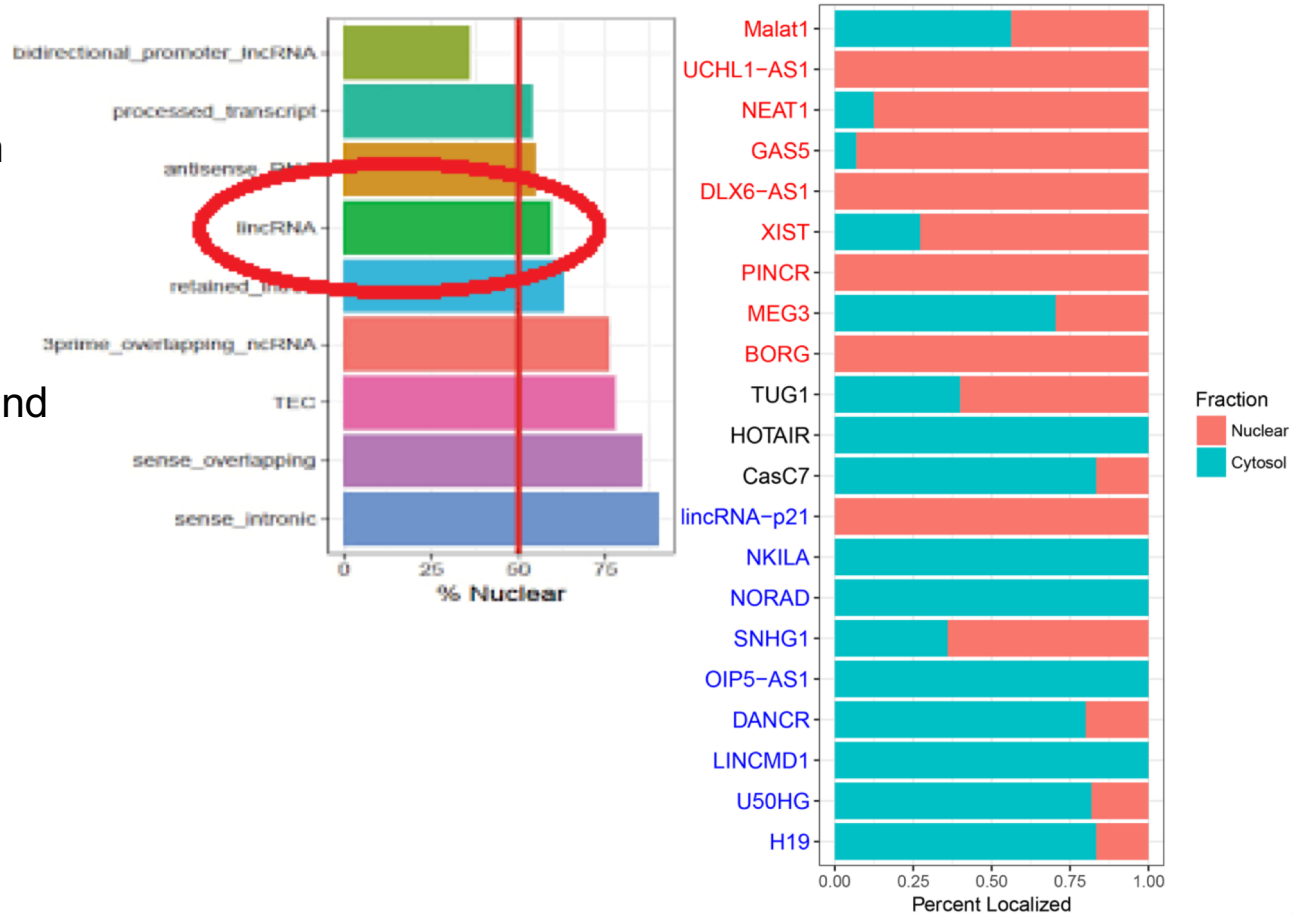
mRNA

Synthesized in nucleus and carried out in cytoplasm to be translated

lincRNA

Perform their functions both in nucleus and cytoplasm

They are degraded with exosomes in nucleoplasm

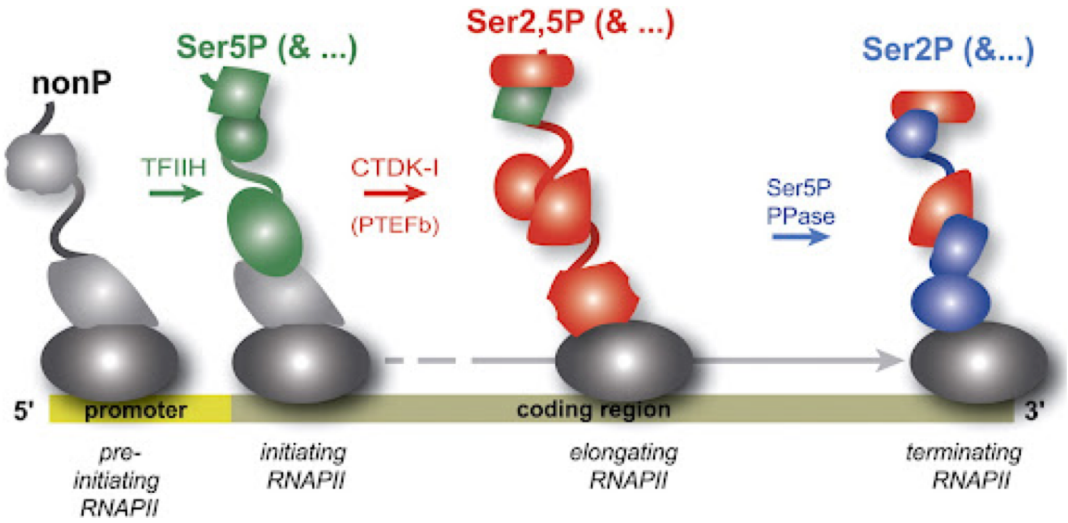


Comparison between mRNA and lincRNA: phosphorylation of RNA Pol II CTD and RNA maturation

mRNA

They have a peculiar pattern of phosphorylation of RNA Polymerase II's CTD:

Ser5 phosphorylated during early elongation
Ser2 phosphorylated during later elongation



And what about the phosphorylation of RNA Polymerase II's CTD in lincRNA?

The paper tries to answer the following questions:

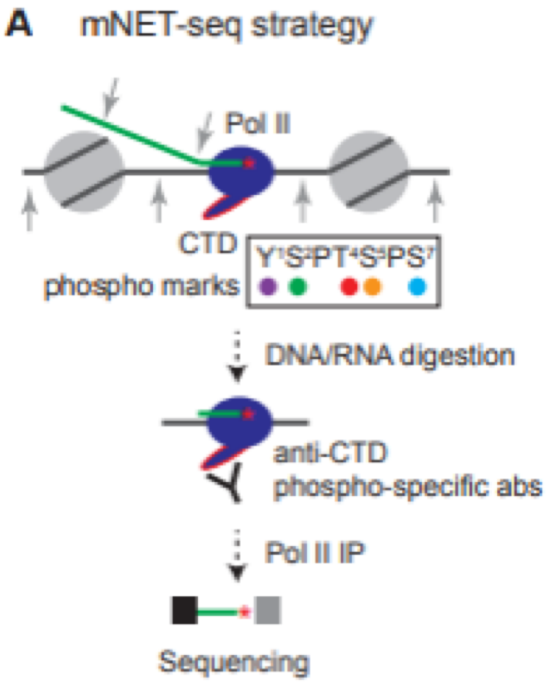
- How does Pol II CTD phosphorylation differ between protein-coding and lincRNA genes?
- Are there any differences between splicing of protein coding and splicing of lincRNA?
- Are lincRNA and protein-coding genes differentially polyadenylated?
- Why are lincRNA levels substantially reduced in the nucleoplasm?
- Are lincRNAs co-transcriptionally cleaved?
- Could lincRNA endonucleolytic cleavage be mediated by the microprocessor?

How does Pol II CTD phosphorylation differ between protein-coding and lincRNA genes?

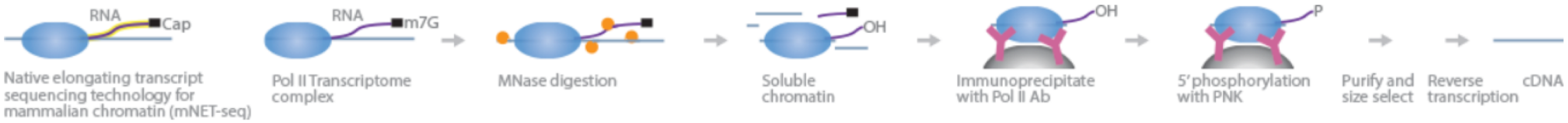
Specific **Pol II CTD phosphorylation states** are associated with **different stages of transcription**:

- **Ser5P** > early elongation, 5’ capping and active splicing
- **Ser2P** > later elongation and 3’ end processing
- Y1P, T4P, S7P, unph > additional phosphorylation states

mNET-seq can be used to sequence nascent RNA by employing Pol II antibodies against specific CTD phosphorylation states in order to isolate RNA from immunoprecipitated Pol II



mNET-seq (Mammalian Native Elongating Transcript sequencing)



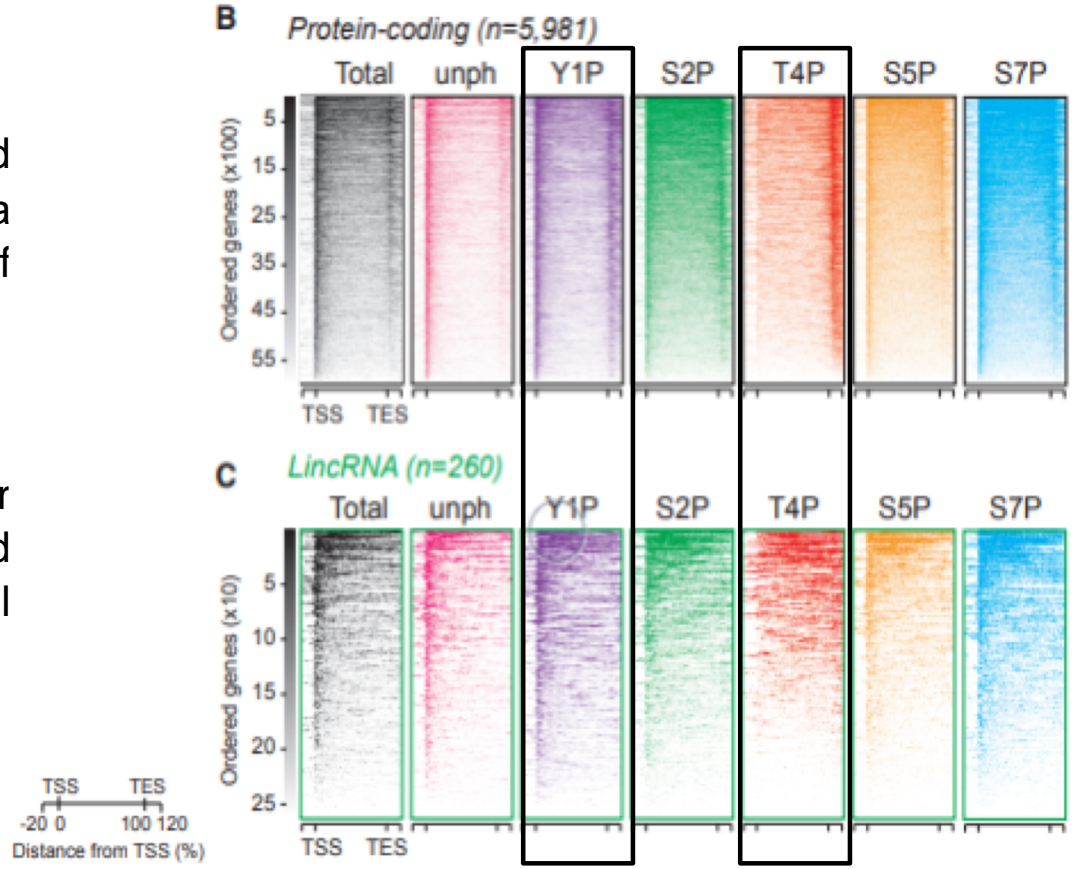
Key steps:

- RNAPII complexes are isolated through chromatin fractionation
- MNase is used to digest all exposed DNA while leaving RNA strands protected by RNAPII or spliceosomes intact
- RNAPII complexes are immunoprecipitated using RNAPII antibodies and 5' phosphorylated by T4 PNK
- 3' linkers are ligated to the 3' hydroxyl end of the RNA strand
- Nascent RNAs are isolated, size-selected for 35–100 nt, processed into cDNA sequencing libraries, and sequenced

mNET-seq analysis:

- *lincRNA genes* show less pronounced **unph** and **Y1P TSS** peaks and a generally **more even distribution** of mNET-seq reads across the gene body

- *protein-coding genes* show a higher **T4P** signal in the **TES** region compared to lincRNA genes, where the **T4P** signal is more evenly distributed

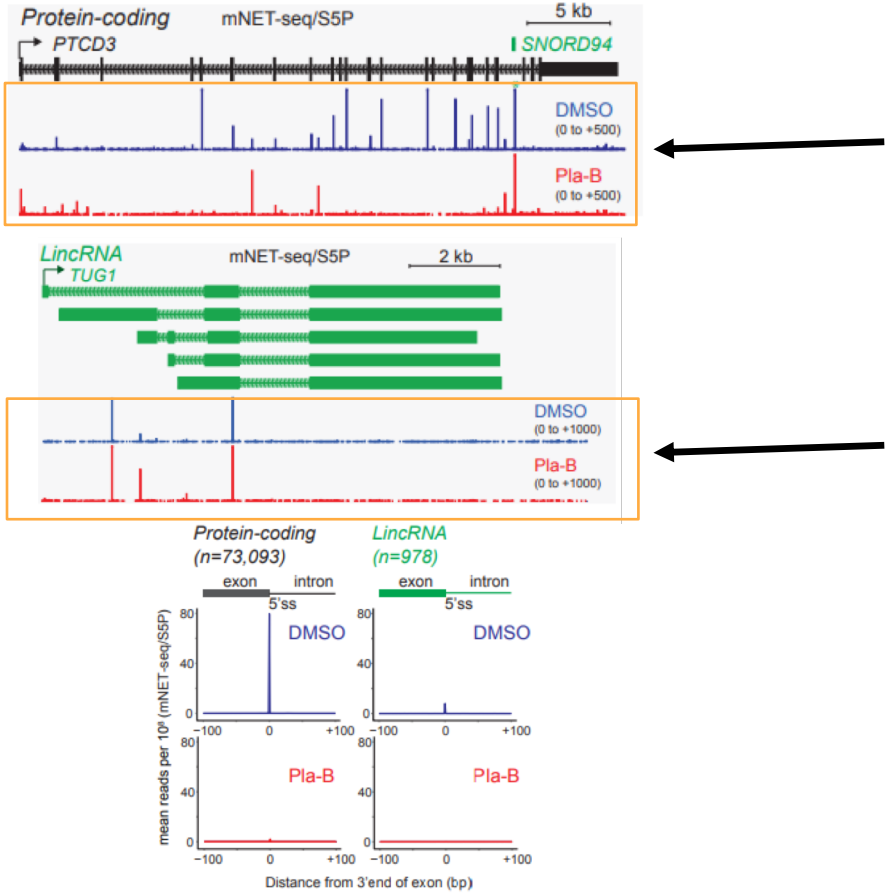


Conclusion > Pol II termination probably occurs at multiple positions across lincRNA genes

Splicing differences between lincRNA and protein-coding genes

Analysis of specific lincRNAs using splicing specific mNET-seq/S5P profiles:

- When HeLa cells are treated with Pla-B, most **S5P CTD-specific 5'ss peaks** on *protein-coding genes*, such as PTCD3, **are lost**
- *lincRNA genes* are **less sensitive** to Pla-B treatment
- 55-70% of protein-coding introns are associated with 5'ss peaks. In contrast, only 20-30% of lincRNA exons show 5'ss peaks

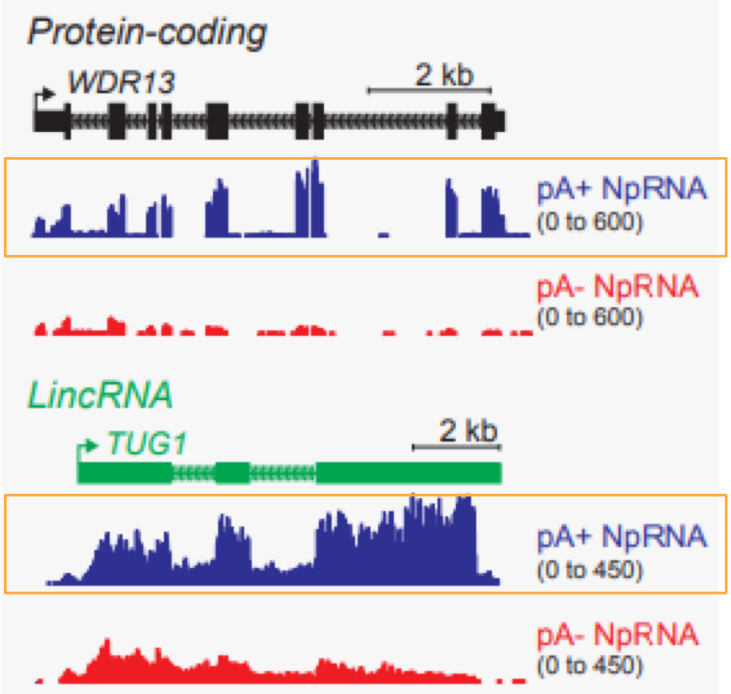


Conclusion > lincRNAs are inefficiently spliced compared to protein-coding genes

Duplicate HeLa cell transcript libraries from either **pA+** or **pA-** nuclear RNA were prepared to measure splicing efficiency directly:

- **pA+ reads** across the *protein-coding gene* **WDR13** are **exon restricted** (> efficient co-transcriptional splicing), with little signal detected in the pA- NpRNA-seq profile

- the *lincRNA* **TUG-1 pA+** profile shows **significant levels of intron reads** over its annotated intron regions, whereas the pA- profile revealed a higher level of intron signal

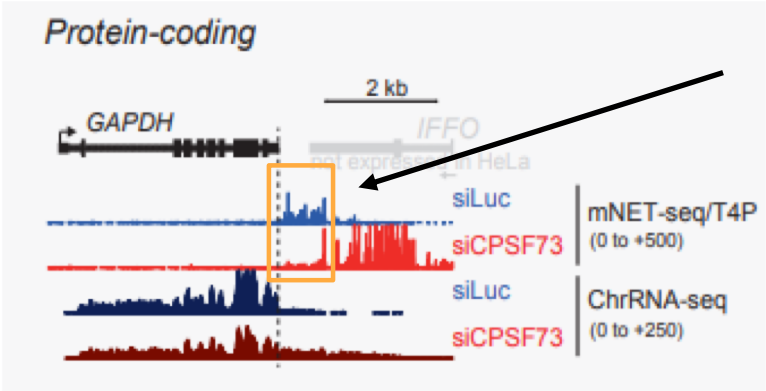


Comparison of splicing events between these two transcript classes shows a consistently lower splicing rate for lincRNAs

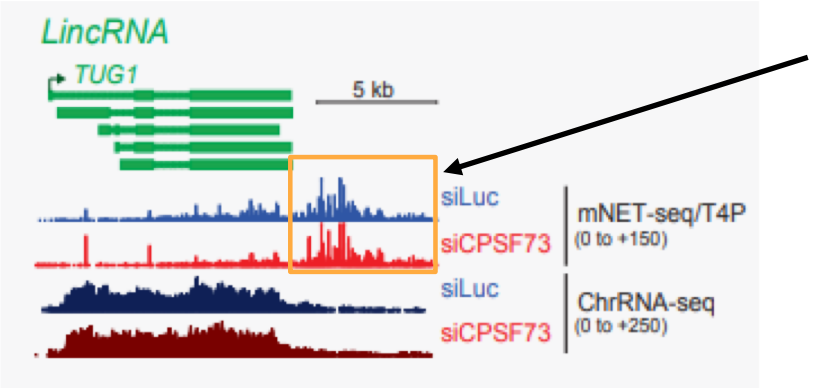
Are lincRNA and protein-coding genes differentially polyadenylated?

Analysis of mNET-seq/T4P datasets shows a close correlation between the CTD **T4P mark** and **protein-coding gene termination**, whereas lincRNAs show a more widespread T4P mNET-profile across the whole transcription unit (TU)

- **Depletion of CPSF73** (cleavage and polyadenylation factor) causes a **substantial decrease in T4P mNET-seq reads over the termination region** of the protein-coding gene **GAPDH**

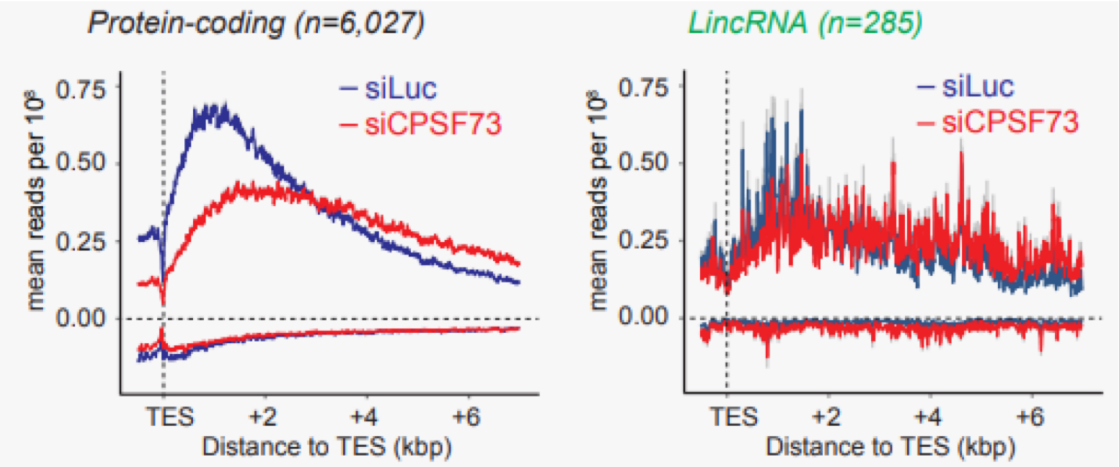


- The lincRNA **TUG1** mNET-seq/T4P profile is not affected by CPSF73 depletion > **TUG1 termination is CPSF-independent**



Are lincRNA and protein-coding genes differentially polyadenylated?

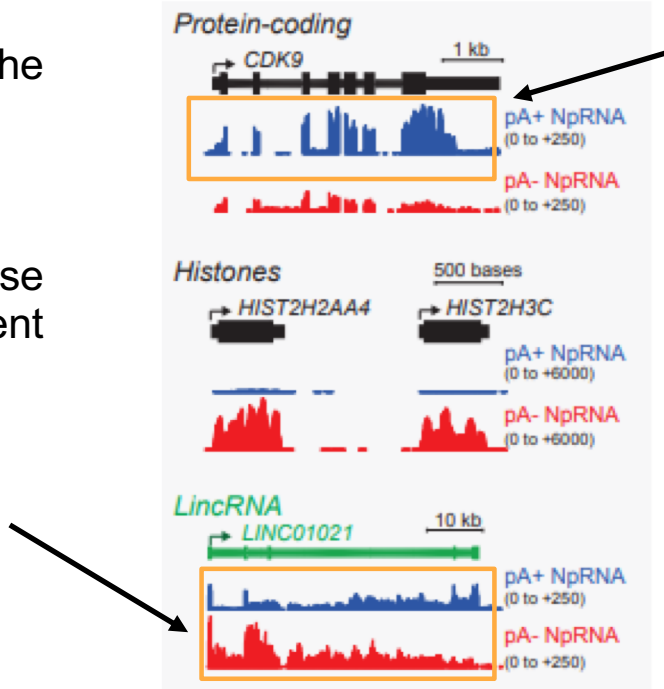
Meta-analysis of the termination region associated with mNET-seq/T4P profiles shows that protein coding, but not lincRNA gene termination, is strongly affected by CPSF73 depletion



Are lincRNA and protein-coding genes differentially polyadenylated?

pA+ and pA- NpRNA-seq libraries were employed to examine the degree of 3' polyadenylation in lincRNAs:

- *protein-coding transcripts* are predominantly **pA+**
- histone RNAs are exclusively in the pA- fraction because histone mRNA is maturated by a PAS-independent mechanism
- *lincRNAs*, such as LINC01021, are **more pA- than pA+**



Conclusion > lincRNAs are inefficiently polyadenylated compared to protein-coding transcripts

Why are lincRNA levels substantially reduced in the nucleoplasm?

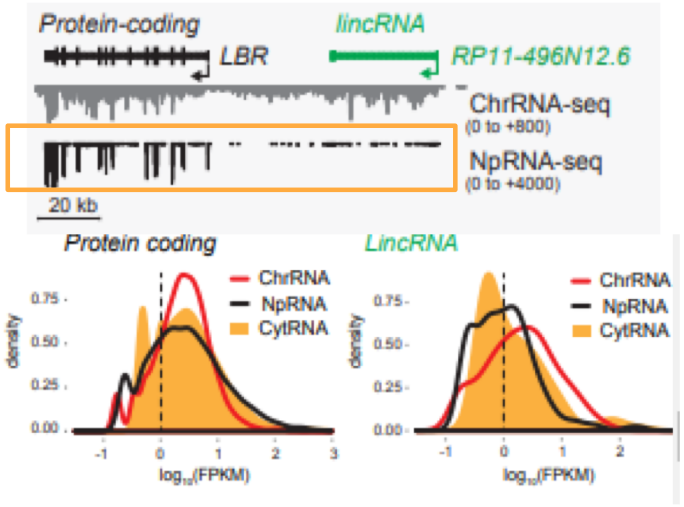
lincRNA and protein-coding gene transcripts are often similar in abundance in the chromatin fraction

	coding (n=6027)	*lincRNA (n=285)*	*Antisense (n=510)*
FPKM, TSS+500bp, ChrRNA	**23.3** (31.5) [25.4]	**21.2** (37.7) [43.1]	**20.1** (25.5) [20.5]
FPKM, TSS+500bp, NpRNA	**25.3** (44.5) [51.3]	**7.0** (11.5) [13.2]	**4.9** (9.1) [10.2]
Maximum number of different exons	**9** (11.1) [6.8]	**3** (3.2) [1.8]	**2** (2.3) [1.2]
Gene length, bp	**32151** (49420.5) [47784.0]	**9077** (25981.5) [38625.1]	**2529.5** (7659.2) [11012.9]

median (Mean 90%, excluding top 5% and bottom 5%) [stddev 90%]

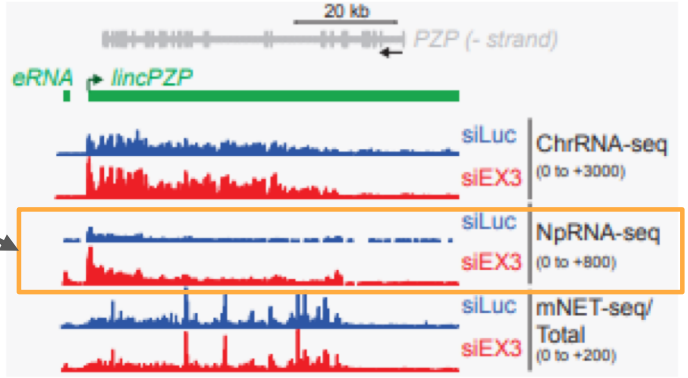
Transcription profiles for a tandem lincRNA and protein-coding gene **LBR** show lower levels of lincRNA in the nucleoplasm compared to chromatin-associated lincRNA

RNA-seq data were analyzed for lincRNA expression in the cytoplasm to exclude the possibility of rapid nuclear export > **less cytoplasmic lincRNA is detected compared to chromatin-associated lincRNA**



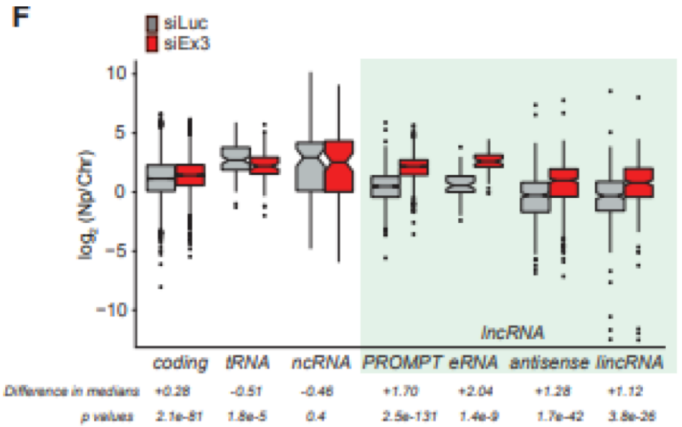
lincRNAs are substrates of the RNA exosome (shown in mESCs)

ChrRNA-seq and NpRNA-seq following depletion of the RNA exosome component **EXOSC3** > **lincRNAs were all significantly increased in the nucleoplasm**



Comparison of the ratio of chromatin to nucleoplasm RNA levels between protein-coding and definable classes of lncRNAs following exosome depletion:

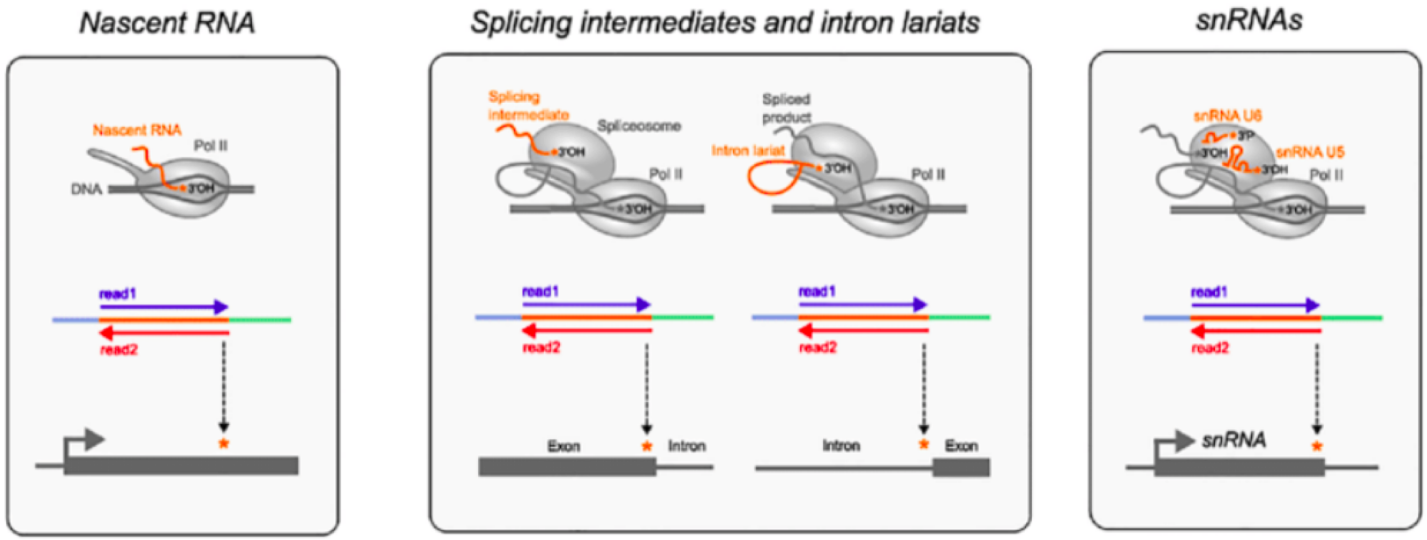
- protein-coding RNA levels are slightly stabilized
- **all categories of lncRNAs show significant nucleoplasmic stabilization**
- tRNAs, structural ncRNAs and small nuclear RNAs were significantly destabilized > known role of the exosome in tRNA and snRNA maturation



Conclusion > lincRNAs are post-transcriptionally degraded by the nuclear exosome

Are lincRNAs co-transcriptionally cleaved? (shown in HeLa)

- The mNET-seq technique involves the ligation of a linker oligonucleotide onto any RNA 3′ end protected from micrococcal nuclease digestion
- RNA 3′ ends principally derive from the Pol II active site, reflecting nascent transcription
- Co-precipitated RNA processing complexes can generate RNA 3′ ends (detected by mNET-seq): e.g. splicing intermediates or microRNA precursors

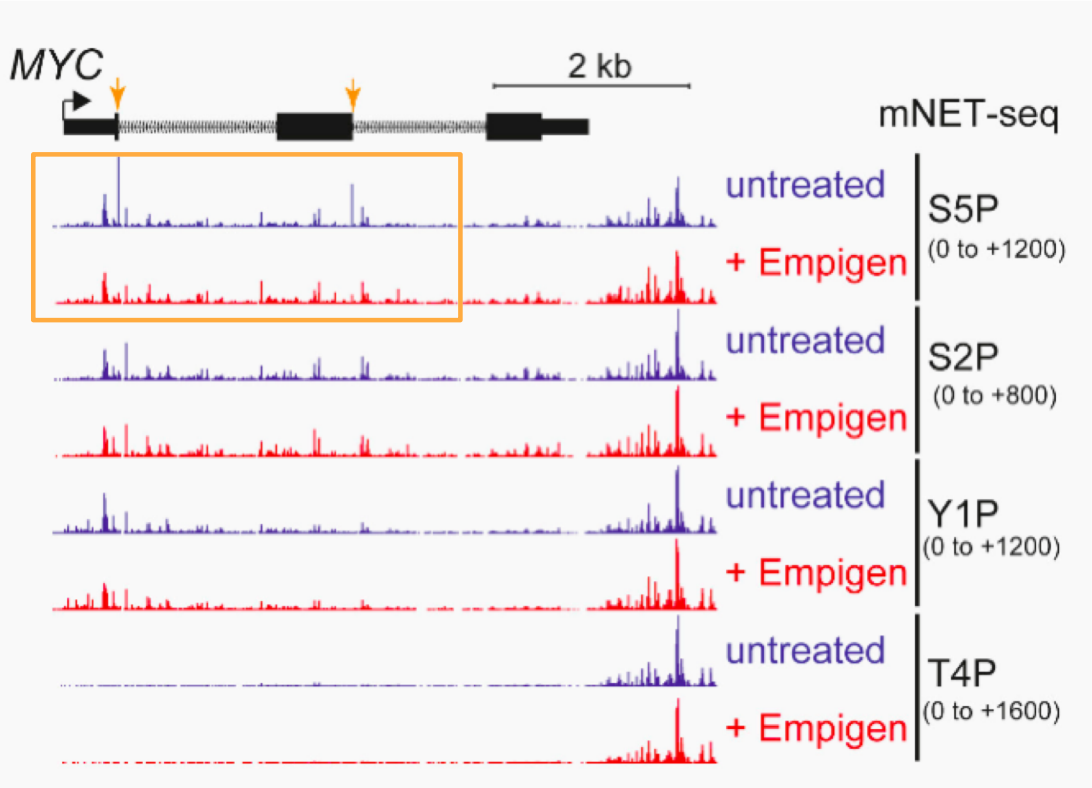


Are lincRNAs co-transcriptionally cleaved? (shown in HeLa)

Empigen is employed to separate mNET-seq reads derived from Pol II active site RNA 3′ ends and those derived from co-precipitated RNA processing complexes

mNET-seq after Empigen treatment:

- *MYC* gene: S5P-specific 5′splicing sites peaks are specifically lost

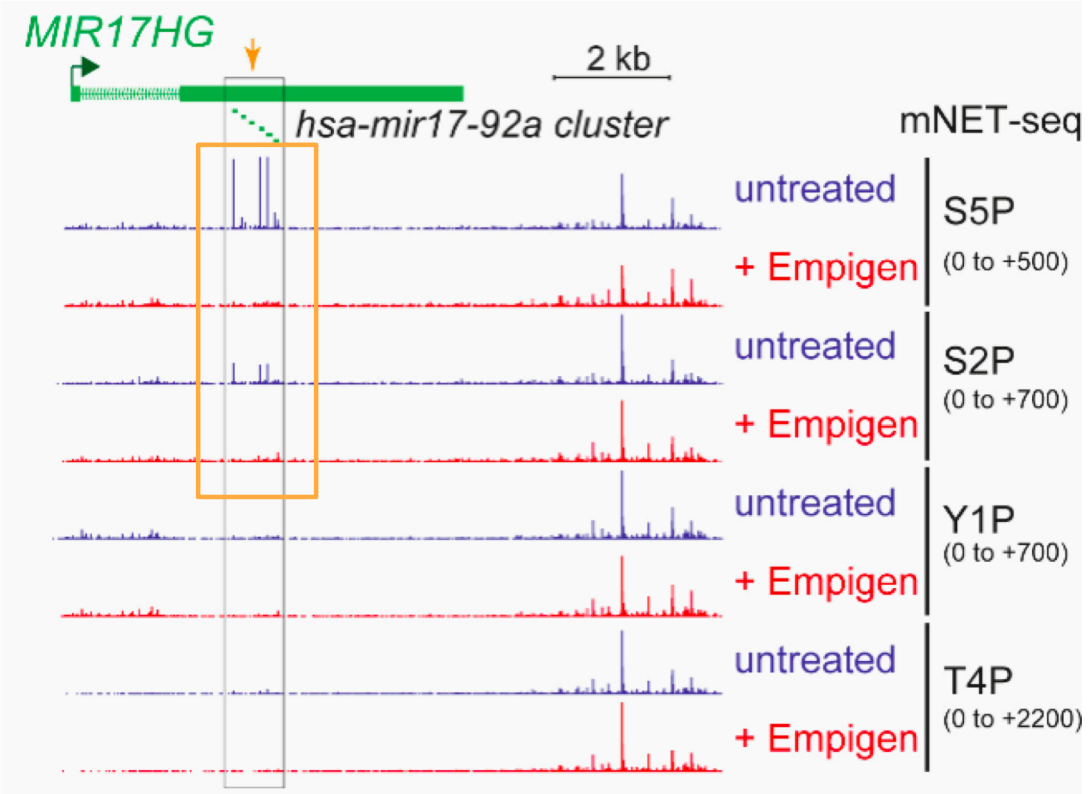


Are lincRNAs co-transcriptionally cleaved? (shown in HeLa)

Empigen is employed to separate mNET-seq reads derived from Pol II active site RNA 3′ ends and those derived from co-precipitated RNA processing complexes

mNET-seq after Empigen treatment:

- *MYC* gene: S5P-specific 5′splicing sites peaks are specifically lost
- lincRNA *MIR17HG:* S5P-/S2P-specific microprocessor-mediated RNA cleavage intermediate is lost

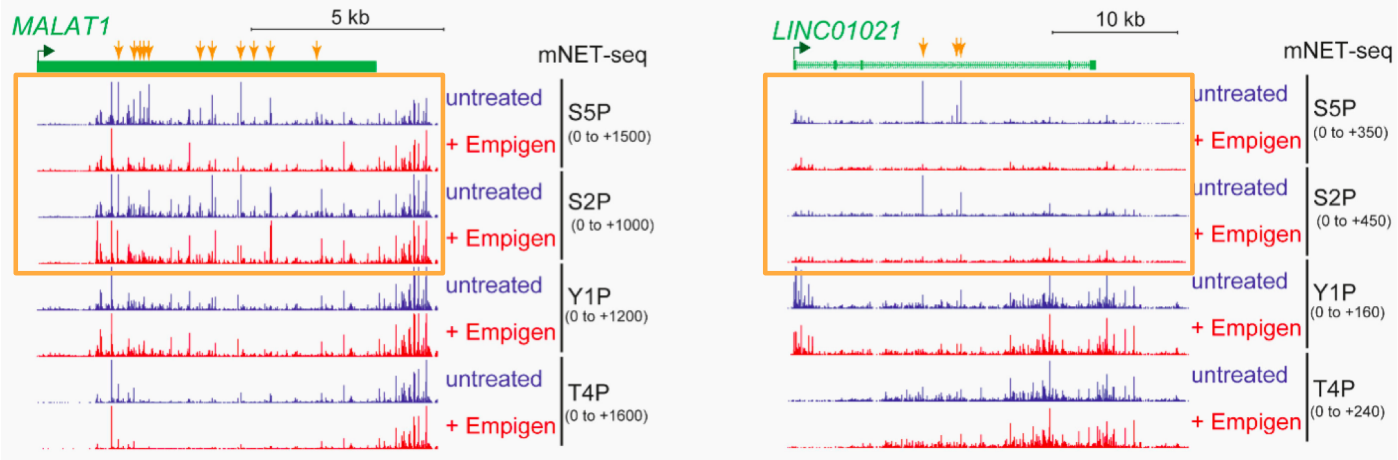


Are lincRNAs co-transcriptionally cleaved? (shown in HeLa)

Empigen is employed to separate mNET-seq reads derived from Pol II active site RNA 3′ ends and those derived from co-precipitated RNA processing complexes

mNET-seq after Empigen treatment:

- *MYC* gene: S5P-specific 5′splicing sites peaks are specifically lost
- lincRNA *MIR17HG:* S5P-/S2P-specific microprocessor-mediated RNA cleavage intermediate is lost
- *MALAT1* and *LINC01021*: lots of S5P and S2P peaks are reduced



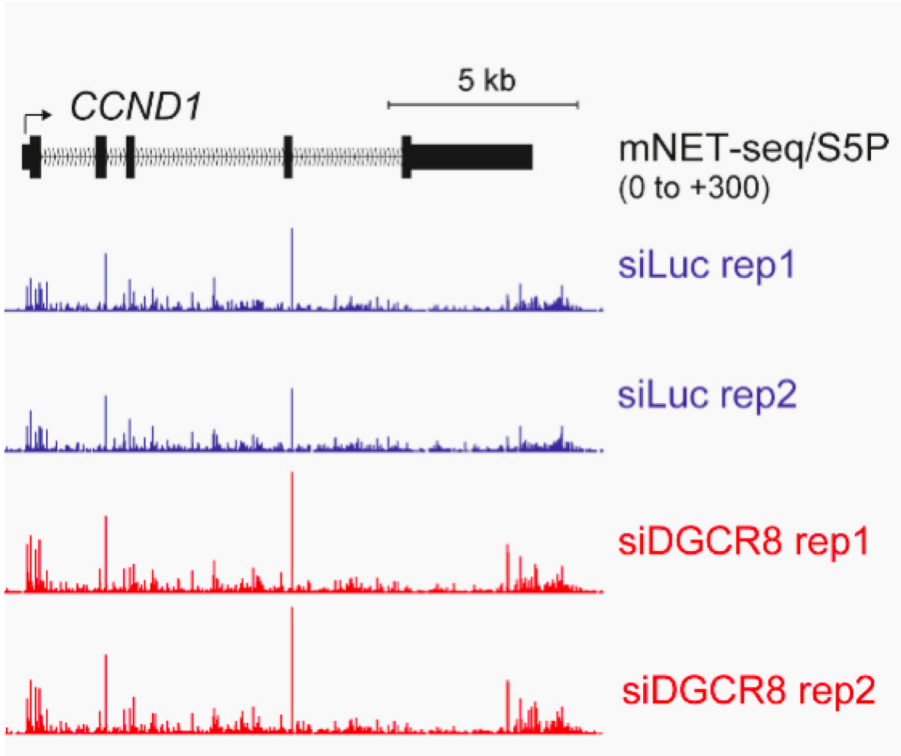
Conclusion > lincRNAs are co-transcriptionally cleaved at multiple positions across their TUs and most Empigen-sensitive lincRNA peaks are insensitive to Pla-B treatment, indicating that they are distinct from splicing intermediates

Could lincRNA endonucleolytic cleavage be mediated by the microprocessor?

(shown in HeLa)

mNET/S5P datasets using chromatin from HeLa cells depleted for either DGCR8 (a double-stranded RNA binding protein) or Dicer. DGCR8 depletion also inactivates Drosha as an integral part of the microprocessor.

- protein-coding gene *CCND1*: neither DGCR8 nor Dicer depletion affected mNET-seq/S5P profiles

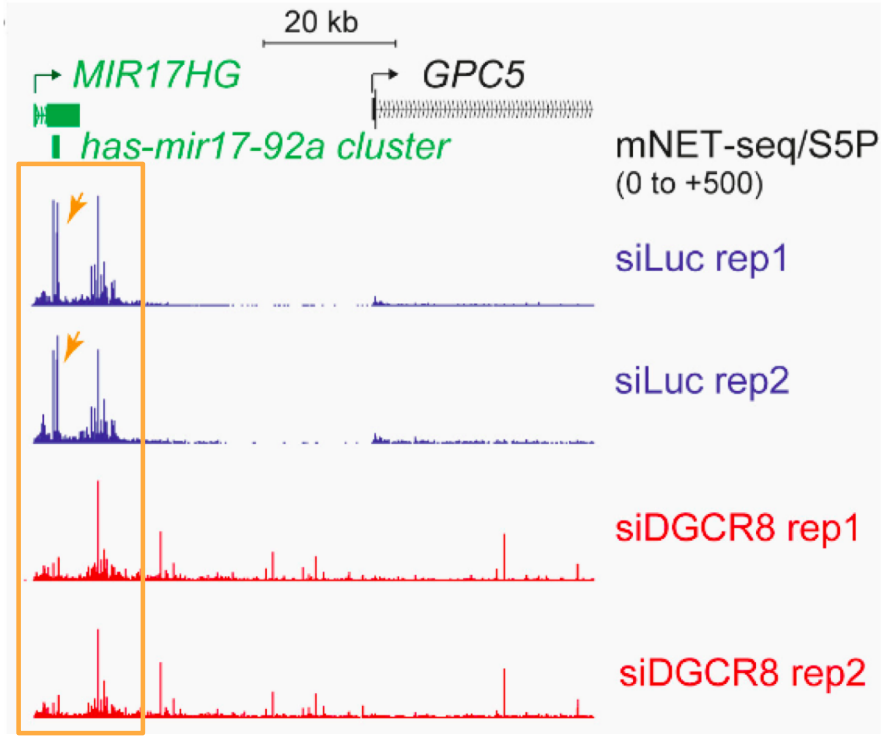


Could lincRNA endonucleolytic cleavage be mediated by the microprocessor?

(shown in HeLa)

mNET/S5P datasets using chromatin from HeLa cells depleted for either DGCR8 (a double-stranded RNA binding protein) or Dicer. DGCR8 depletion also inactivates Drosha as an integral part of the microprocessor.

- protein-coding gene *CCND1*: neither DGCR8 nor Dicer depletion affected mNET-seq/S5P profiles
- *MIR17HG*, which encodes the miR17-92a cluster: DGCR8 depletion affected mNET-seq peaks corresponding to release of these pre-miRNAs.

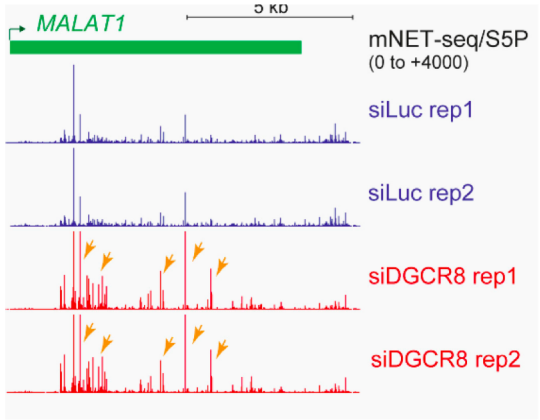


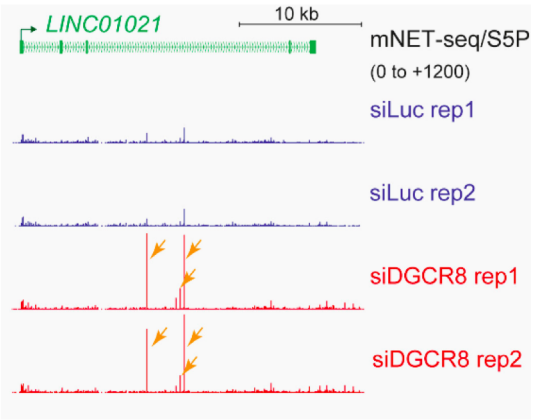
Could lincRNA endonucleolytic cleavage be mediated by the microprocessor?

(shown in HeLa)

mNET/S5P datasets using chromatin from HeLa cells depleted for either DGCR8 (a double-stranded RNA binding protein) or Dicer. DGCR8 depletion also inactivates Drosha as an integral part of the microprocessor.

- protein-coding gene *CCND1*: neither DGCR8 nor Dicer depletion affected mNET-seq/S5P profiles
- *MIR17HG*, which encodes the miR17-92a cluster: DGCR8 depletion affected mNET-seq peaks corresponding to release of these pre-miRNAs.
- lincRNA: neither loss of DGCR8 nor Dicer caused a general loss of mNET-seq/S5P peaks



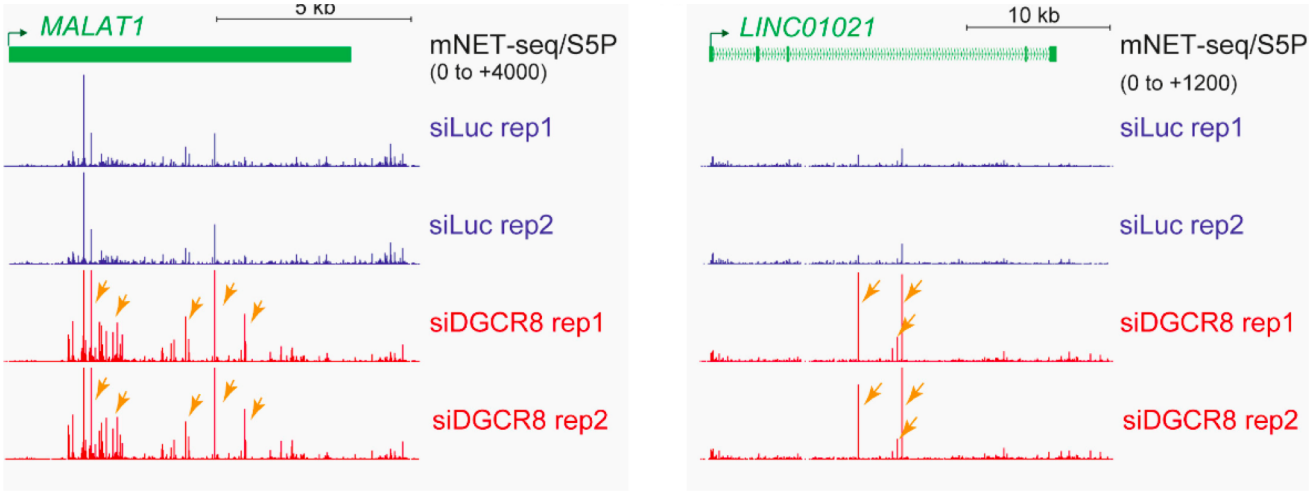


Seems that these endonucleases have not a role in lincRNA cleavage

Could lincRNA endonucleolytic cleavage be mediated by the microprocessor?

(shown in HeLa)

- DGCR8 interacts with nuclear RNA exosome components, independently of the endonuclease Drosha, facilitating exosome recruitment to degrade abundant lncRNAs.
- DGCR8, but not Dicer, depletion acted to selectively increase Empigen-sensitive mNET-seq/S5P peaks on lincRNA genes (*MALAT1* and *LINC01021*).



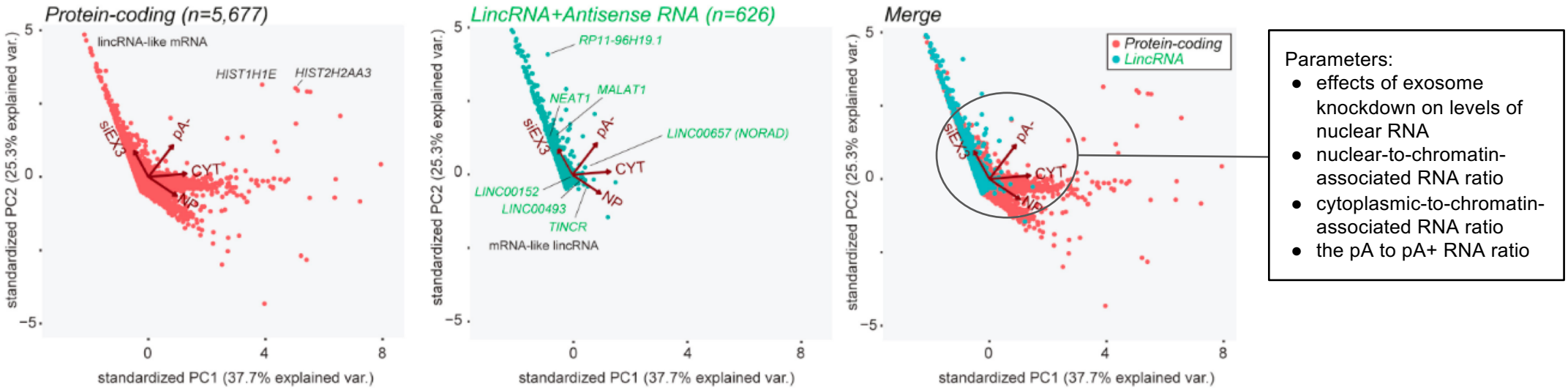
Conclusion > DGCR8 also acts to recruit the exosome to co-transcriptionally cleaved lincRNA, independently of miRNA

PCA reveals lincRNAs are generally distinct from protein-coding genes

Principal-component analysis (PCA) compare protein-coding versus lincRNA TUs based on multiple parameters.

Main features:

- lincRNA TUs → upregulation upon exosome knockdown and general lack of polyA
- protein-coding TUs → stability within the nucleoplasm and cytoplasm

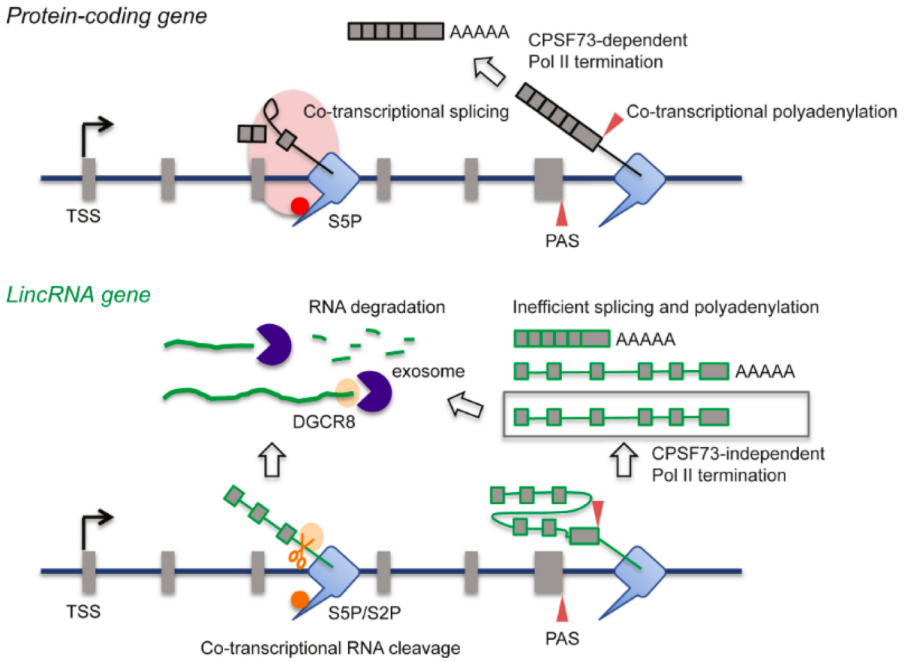


CONCLUSION:

	lincRNAs	**mRNAs**
Pol II phospho-CTD isoforms	CTD profiles appear less selective, T4P signal is more evenly distributed	show higher selectivity for specific CTD modifications
Trancription termination	mainly cleavage and polyadenylation factor (CPA)-independent manner	cleavage and polyadenylation factor (CPA)-dependent manner
Polyadenylation	mainly non-polyadenylated	polyadenylated
Splicing	rarely spliced	spliced
exosome degradation	are post-transcriptionally degraded by the nuclear exosome	low-level turnover by the exosome
co-transcriptional cleavage	are co-transcriptionally cleaved at multiple positions across their TUs.	

STILL TO DISCUSS:

LincRNAs appear unlikely to possess sequence-specific functions. Possibly, the act of transcription rather than the nature of the transcript underlies their biological purpose. However, it remains an attractive possibility that tissue-specific RNA-binding proteins (possibly absent in HeLa cells) may selectively restrict lincRNA turnover and so allow their sufficient accumulation to promote functional roles at least for some of these RNAs.




Thanks for your attention!