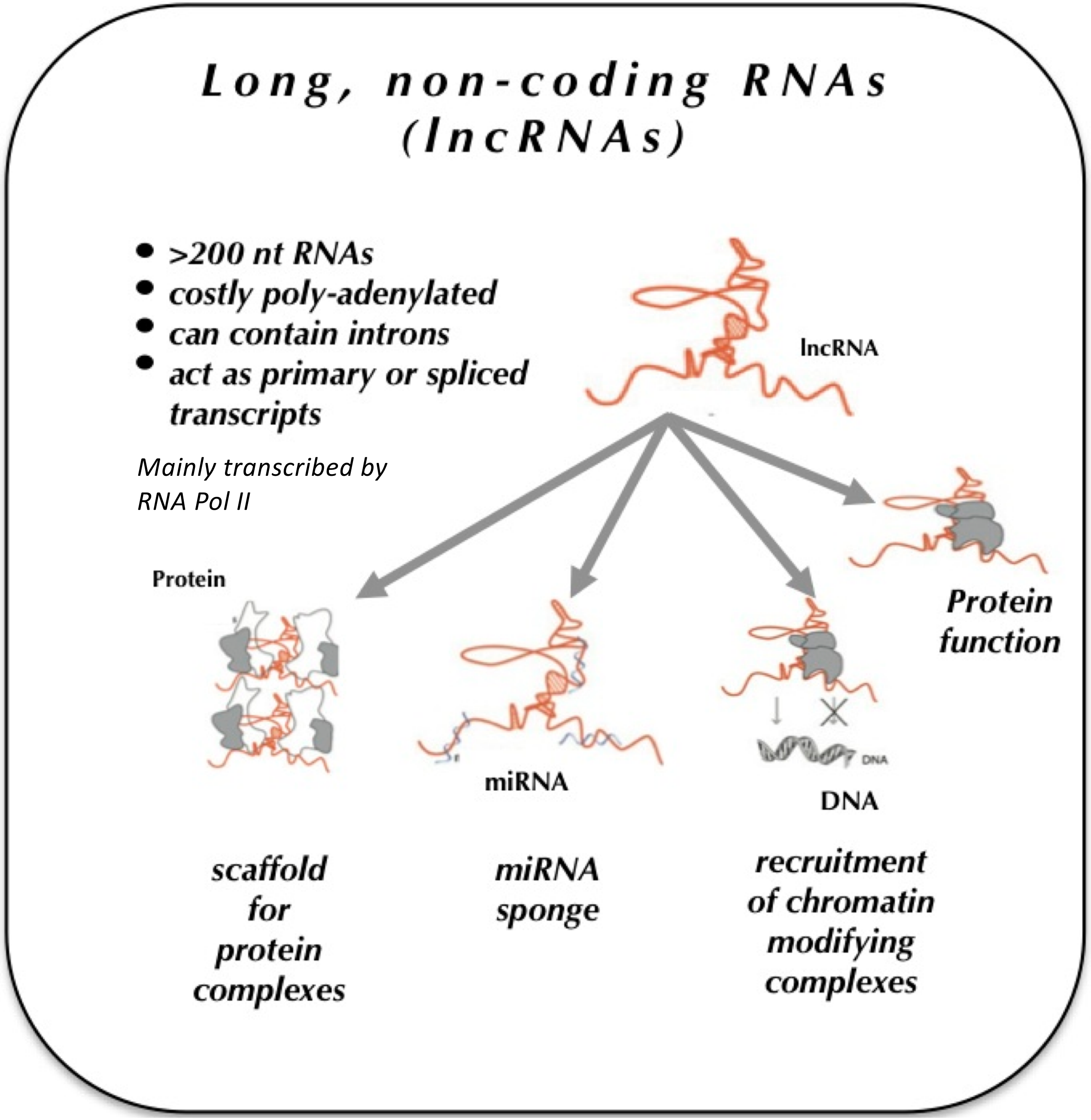
lncRNAs

long, non-coding RNAs

Characteristics of lncRNAs



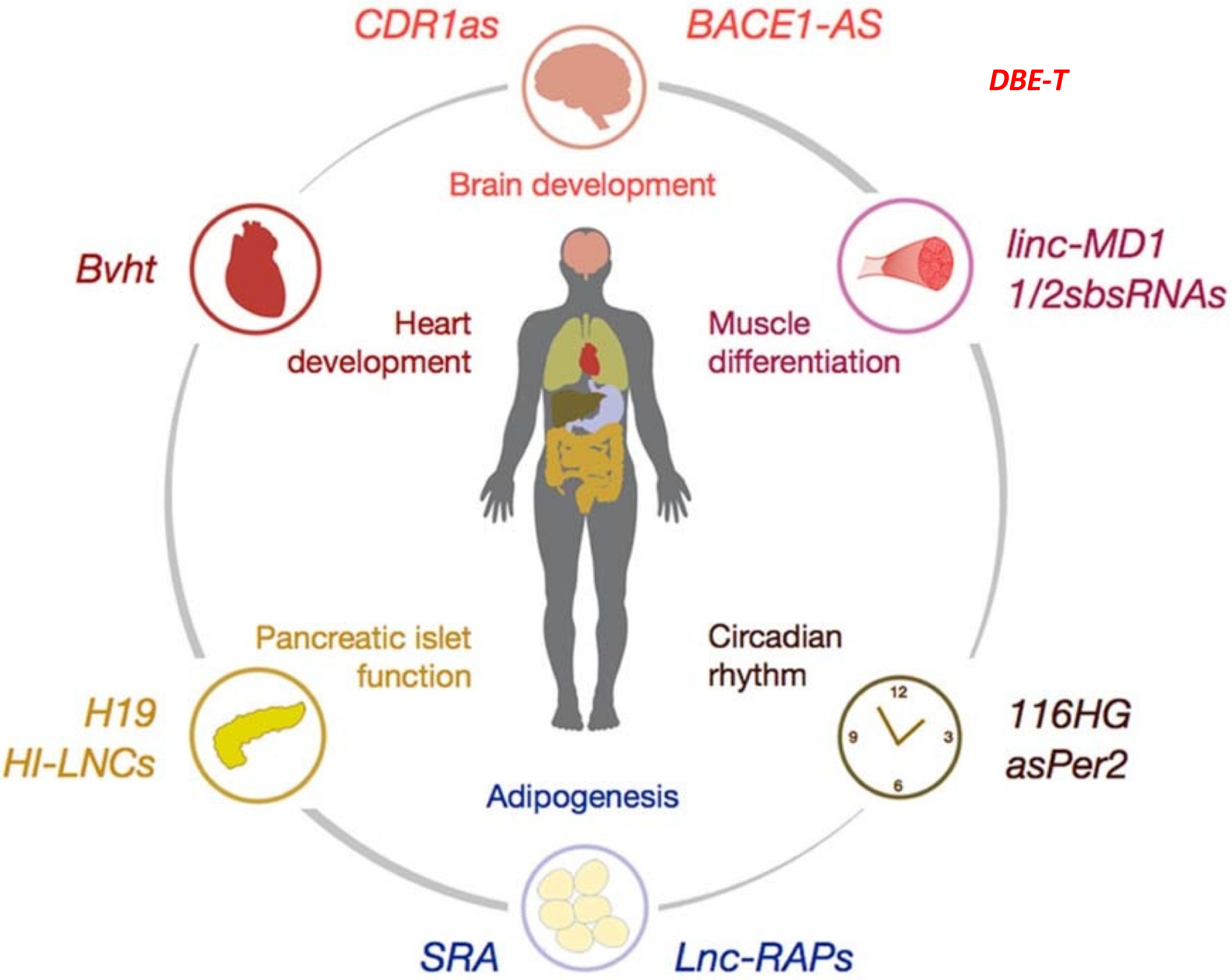
Long, non-coding RNAs (lncRNAs)

- *>200 nt RNAs*
- *costly poly-adenylated*
- *can contain introns*
- *act as primary or spliced transcripts*

Mainly transcribed by RNA Pol II

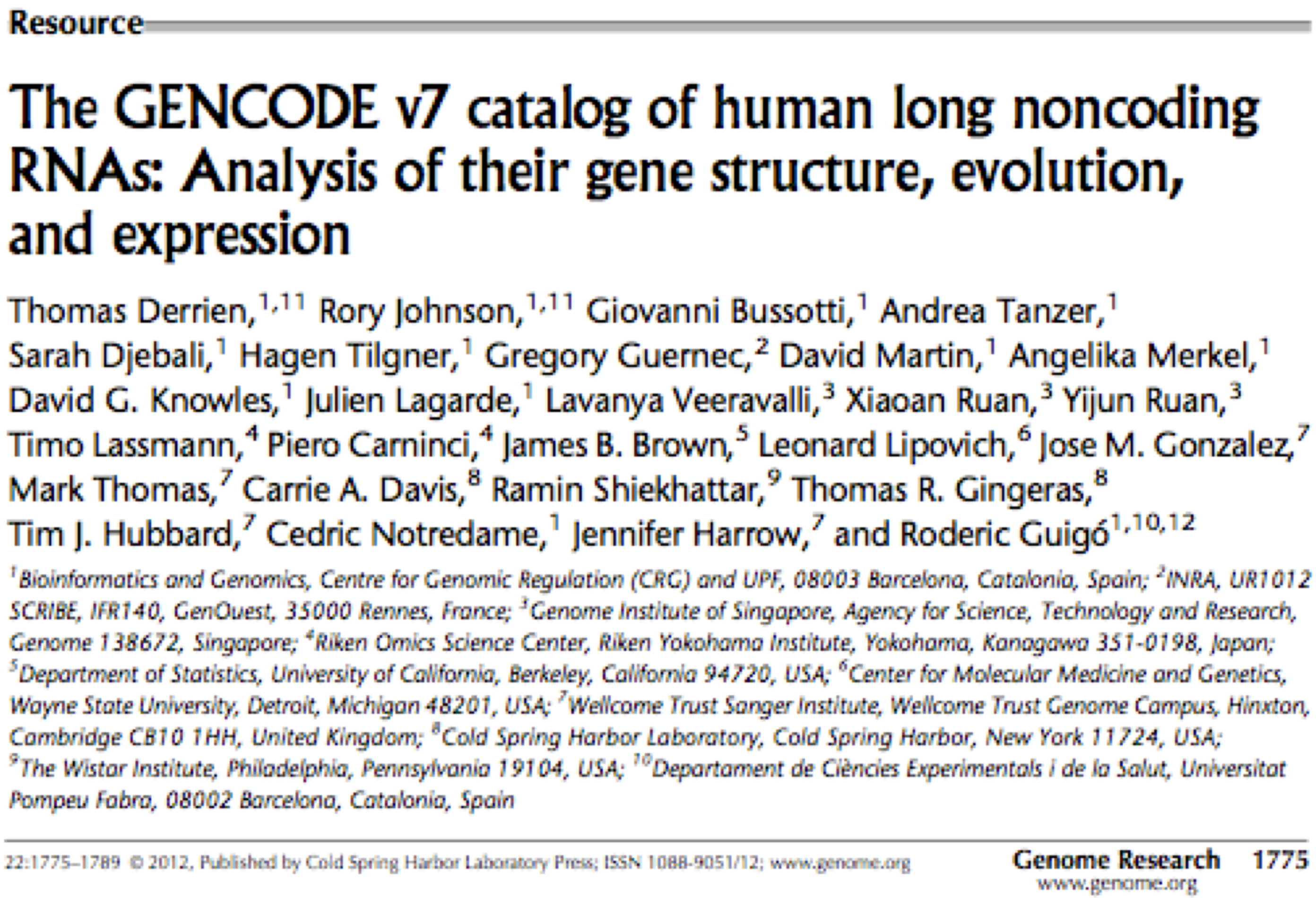
Nuclear and/or cytoplasmatic localiztion

lncRNAs in physiology



DBE-T

COMPARING mRNAs - lncRNAs



Resource

The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression

Thomas Derrien,[1,11] Rory Johnson,[1,11] Giovanni Bussotti,[1] Andrea Tanzer,[1] Sarah Djebali,[1] Hagen Tilgner,[1] Gregory Guernec,[2] David Martin,[1] Angelika Merkel,[1] David G. Knowles,[1] Julien Lagarde,[1] Lavanya Veeravalli,[3] Xiaoan Ruan,[3] Yijun Ruan,[3] Timo Lassmann,[4] Piero Carninci,[4] James B. Brown,[5] Leonard Lipovich,[6] Jose M. Gonzalez,[7] Mark Thomas,[7] Carrie A. Davis,[8] Ramin Shiekhattar,[9] Thomas R. Gingeras,[8] Tim J. Hubbard,[7] Cedric Notredame,[1] Jennifer Harrow,[7] and Roderic Guigó[1,10,12]

[1] Bioinformatics and Genomics, Centre for Genomic Regulation (CRG) and UPF, 08003 Barcelona, Catalonia, Spain; [2] INRA, UR1012 SCRIBE, IFR140, GenOuest, 35000 Rennes, France; [3] Genome Institute of Singapore, Agency for Science, Technology and Research, Genome 138672, Singapore; [4] Riken Omics Science Center, Riken Yokohama Institute, Yokohama, Kanagawa 351-0198, Japan; [5] Department of Statistics, University of California, Berkeley, California 94720, USA; [6] Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan 48201, USA; [7] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom; [8] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; [9] The Wistar Institute, Philadelphia, Pennsylvania 19104, USA; [10] Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08002 Barcelona, Catalonia, Spain

22:1775–1789 © 2012, Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/12; www.genome.org

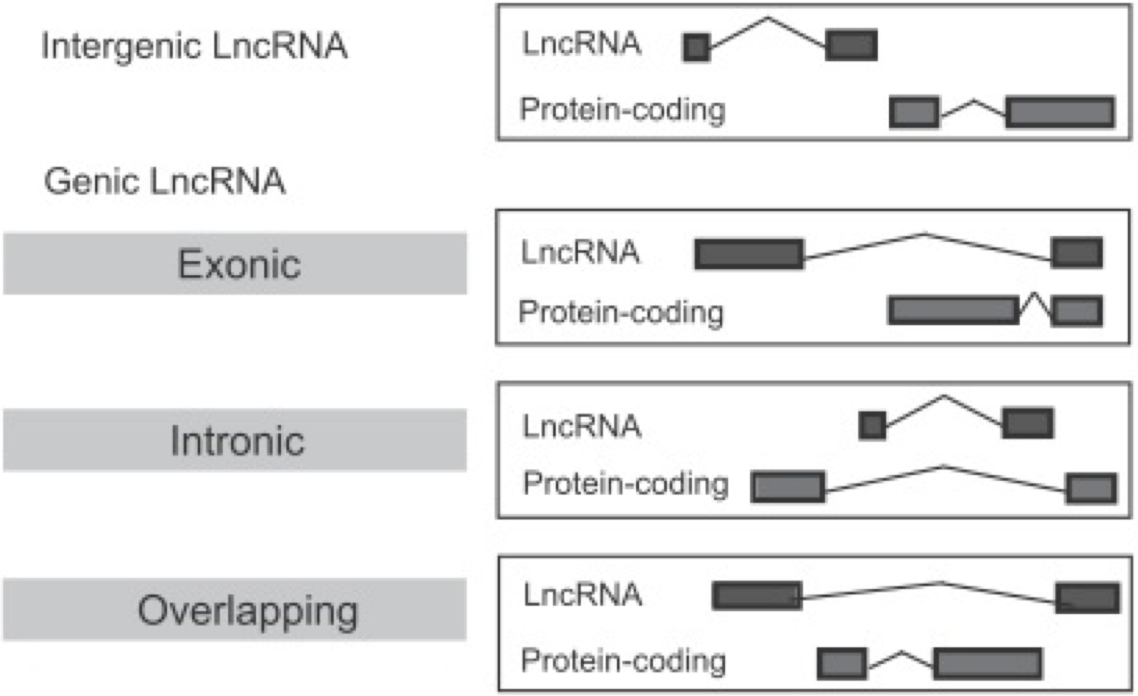
Genome Research 1775
www.genome.org

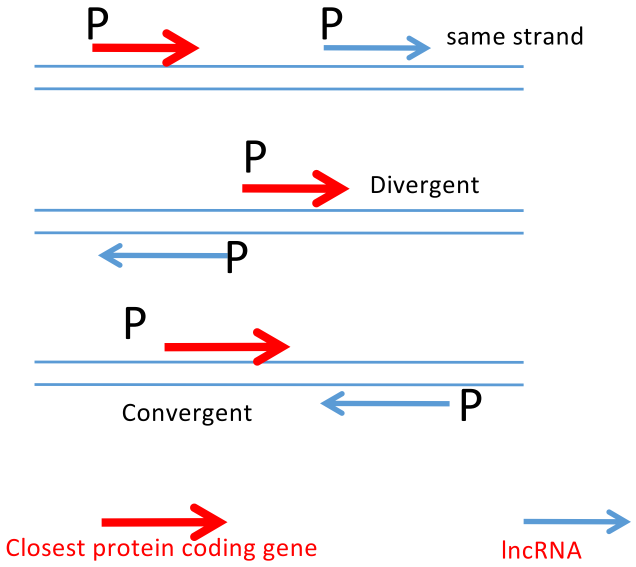
COMPARING mRNAs - lncRNAs

lncRNAs can be

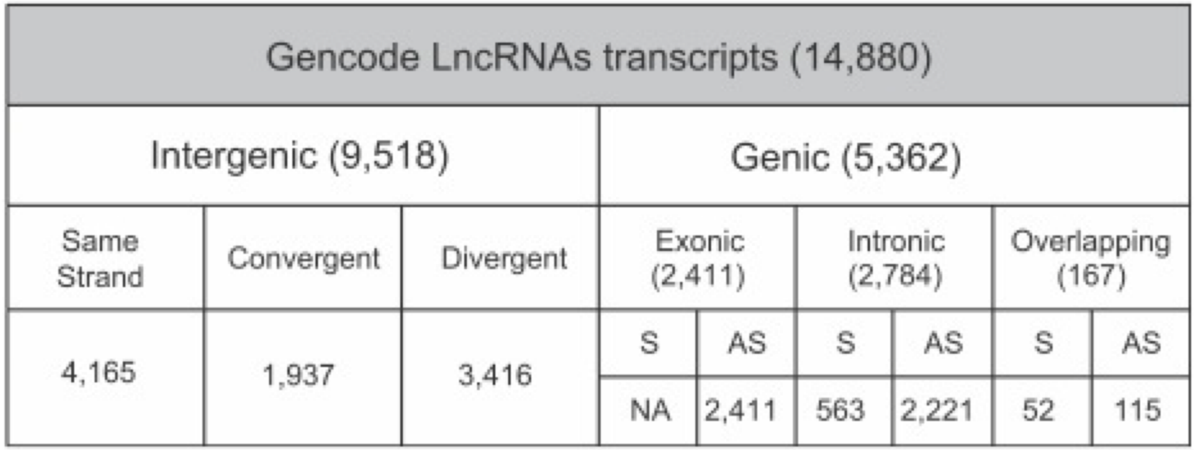
1. **Intergenic lncRNAs (lincRNA)**:
Do not intersect with protein coding gene

2. Genic lncRNAs:
Intersect a protein coding gene
- Exonic
- Intronic
- overlapping





P: promoter



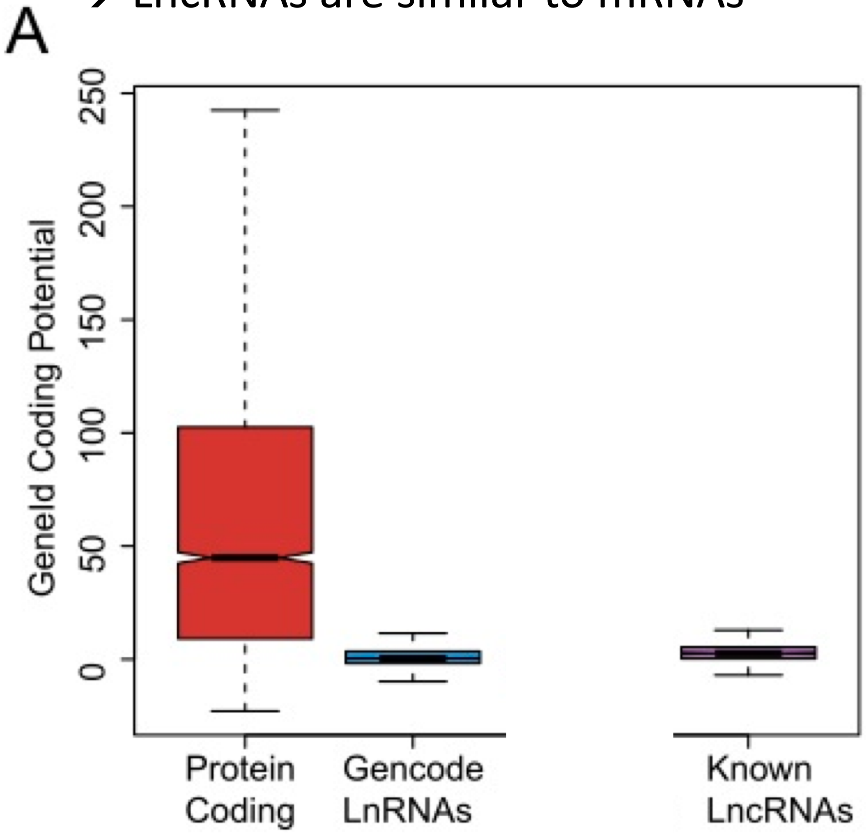
Gencode LncRNAs transcripts (14,880)								
Intergenic (9,518)			Genic (5,362)					
Same Strand	Convergent	Divergent	Exonic (2,411)		Intronic (2,784)		Overlapping (167)	
			S	AS	S	AS	S	AS
4,165	1,937	3,416	NA	2,411	563	2,221	52	115

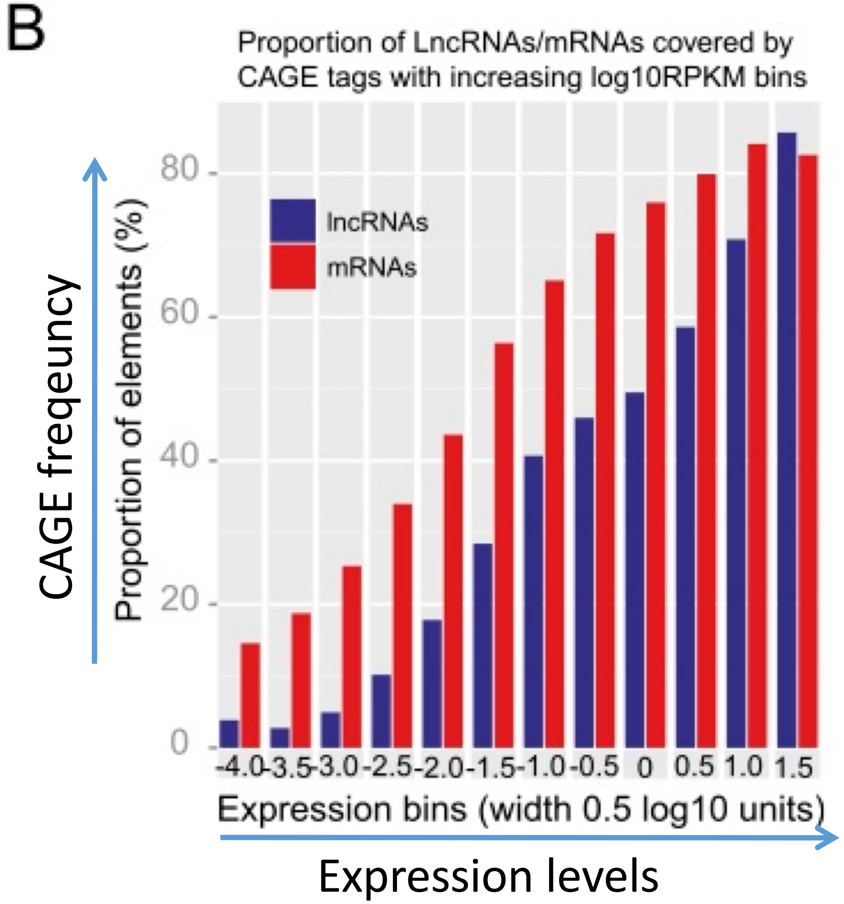
Why? Why? Why?

COMPARING mRNAs - lncRNAs

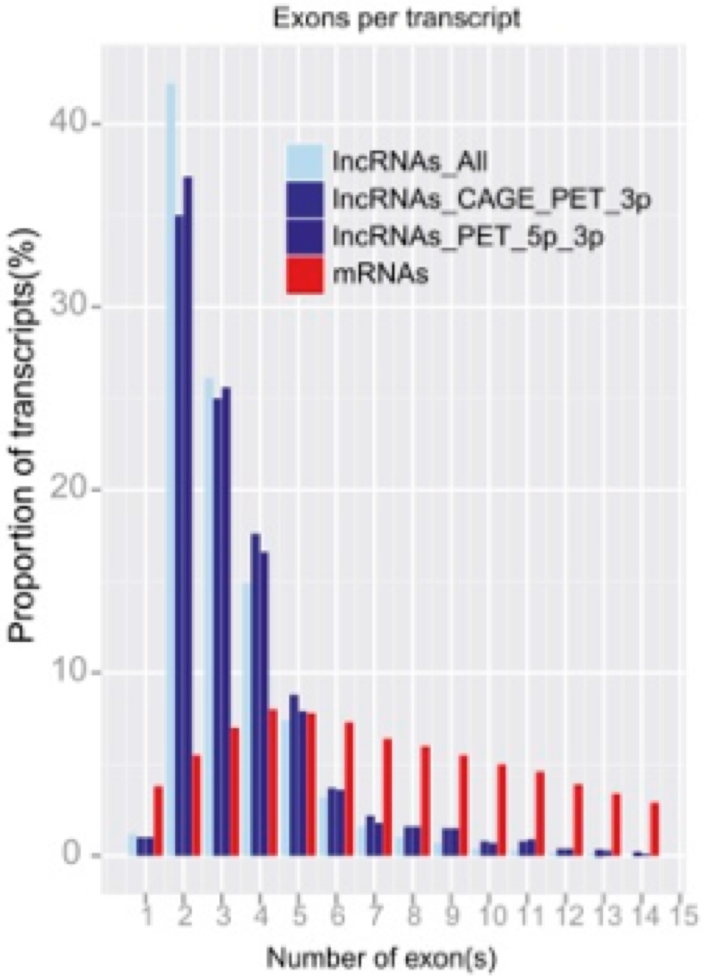
A. LncRNAs do not have coding potential:
→ Longest possible ORF was searched in mRNA/lncRNAs

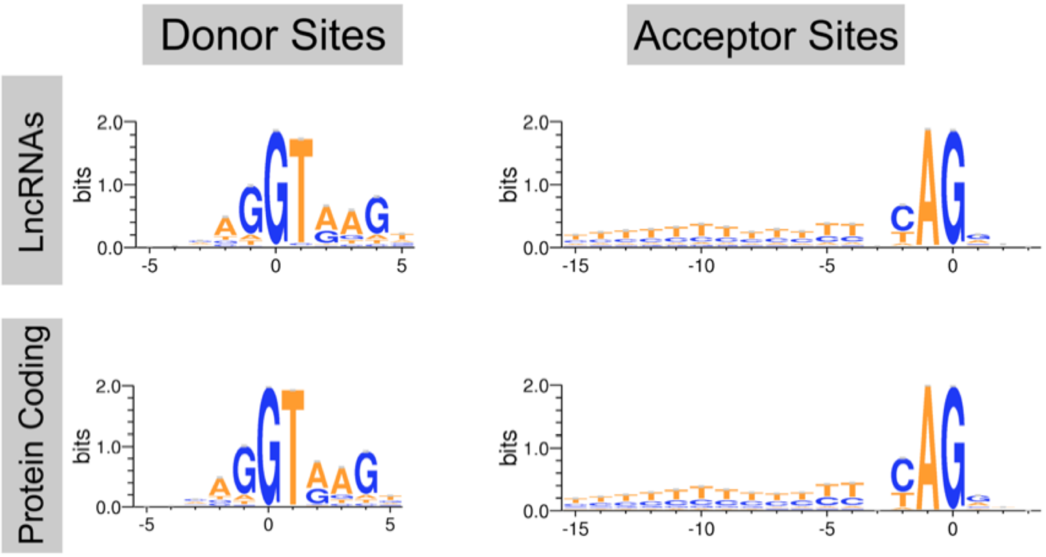
B. CAGE tags can be found in lncRNAs → defined transcriptional start site
CAGE tag freqeuncy increases with increased lncRNA expression levels.
mRNAs are characterized by more CAGE tags
→ LncRNAs are similar to mRNAs





COMPARING mRNAs - lncRNAs





SPLICING:

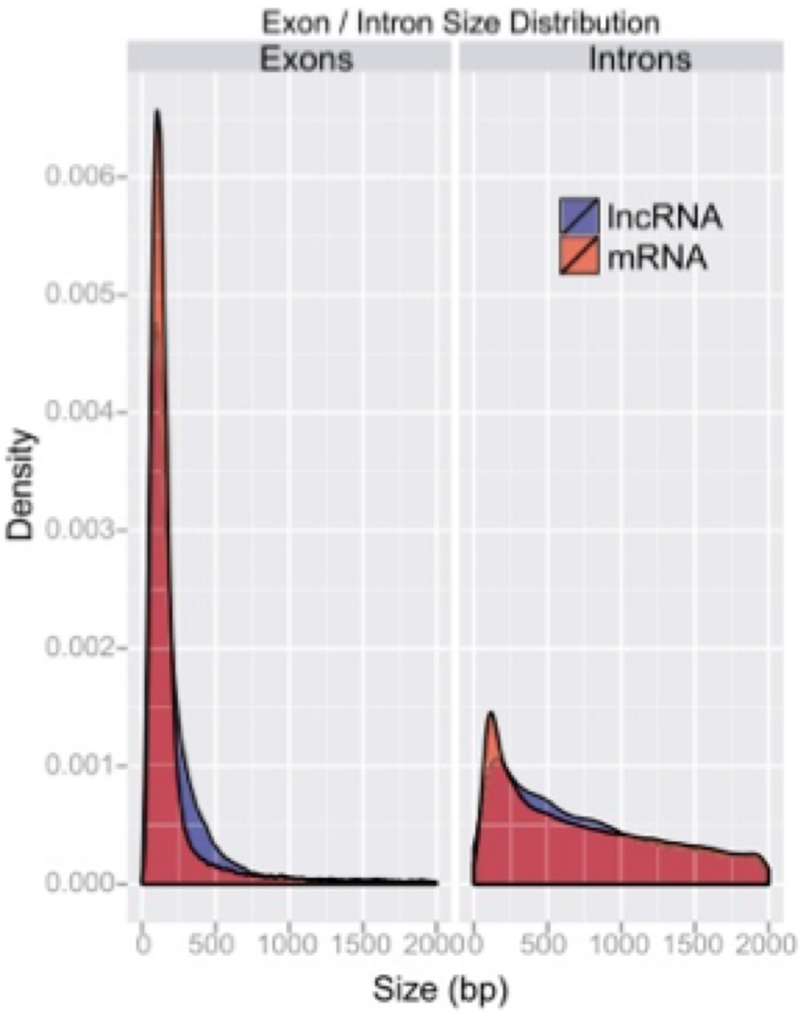
- 98% of lncRNAs are spliced,

- **Relevant Splice-site prerequisites at splice donor/acceptor are conserved**
- **Remarkable tendency of lncRNAs to have only 2 exons (42%, mRNAs: 6%)**

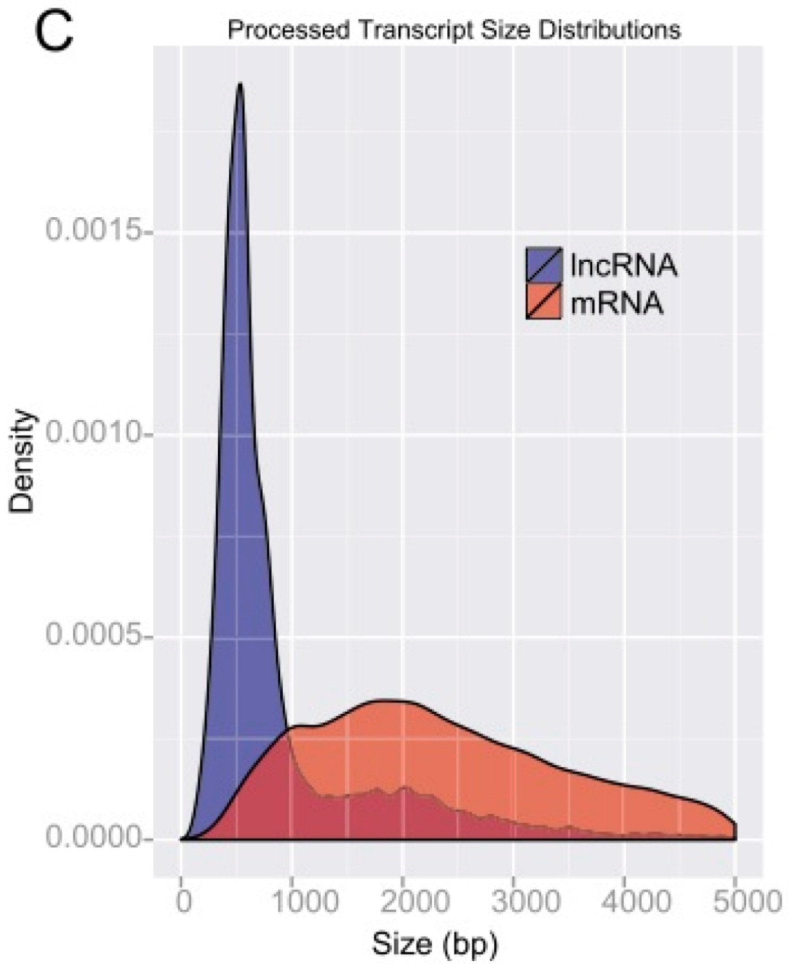
BUT: LncRNAs contain fewer INTRONs!! most lncRNA have only 1 intron!!!; mRNAs 3-6

COMPARING mRNAs - lncRNAs

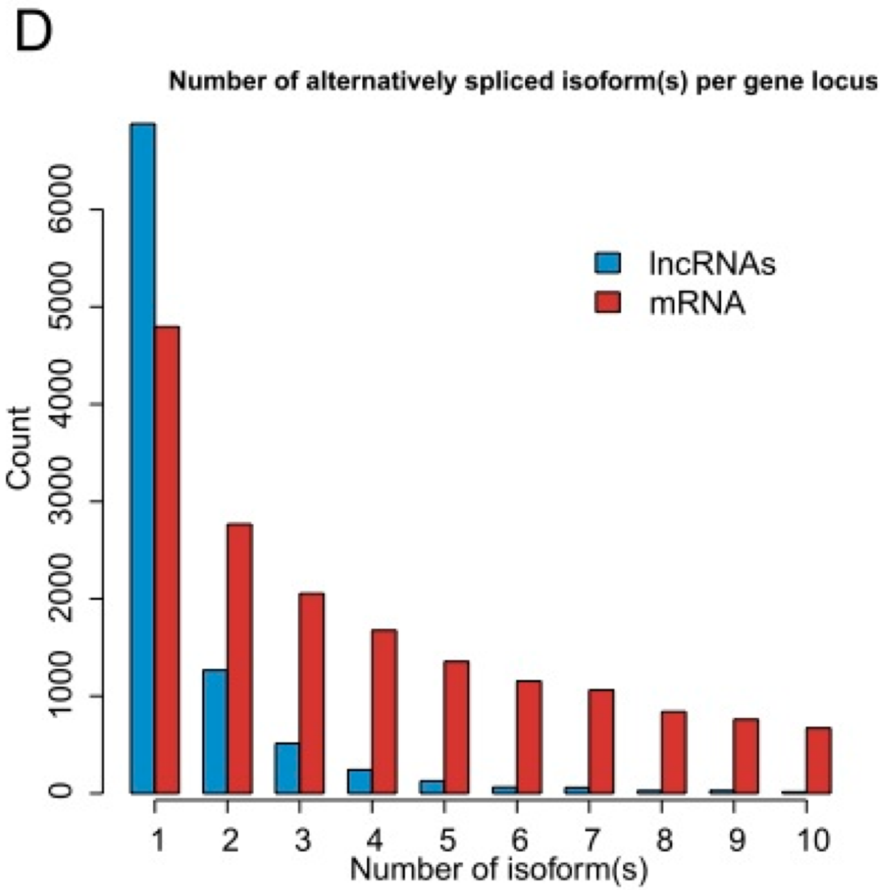
Introns/Exons from lncRNAs are slightly longer



COMPARING mRNAs - lncRNAs



- lncRNAs are on average much shorter: ca. 500nt
- mRNAs are longer and have wider size distribution

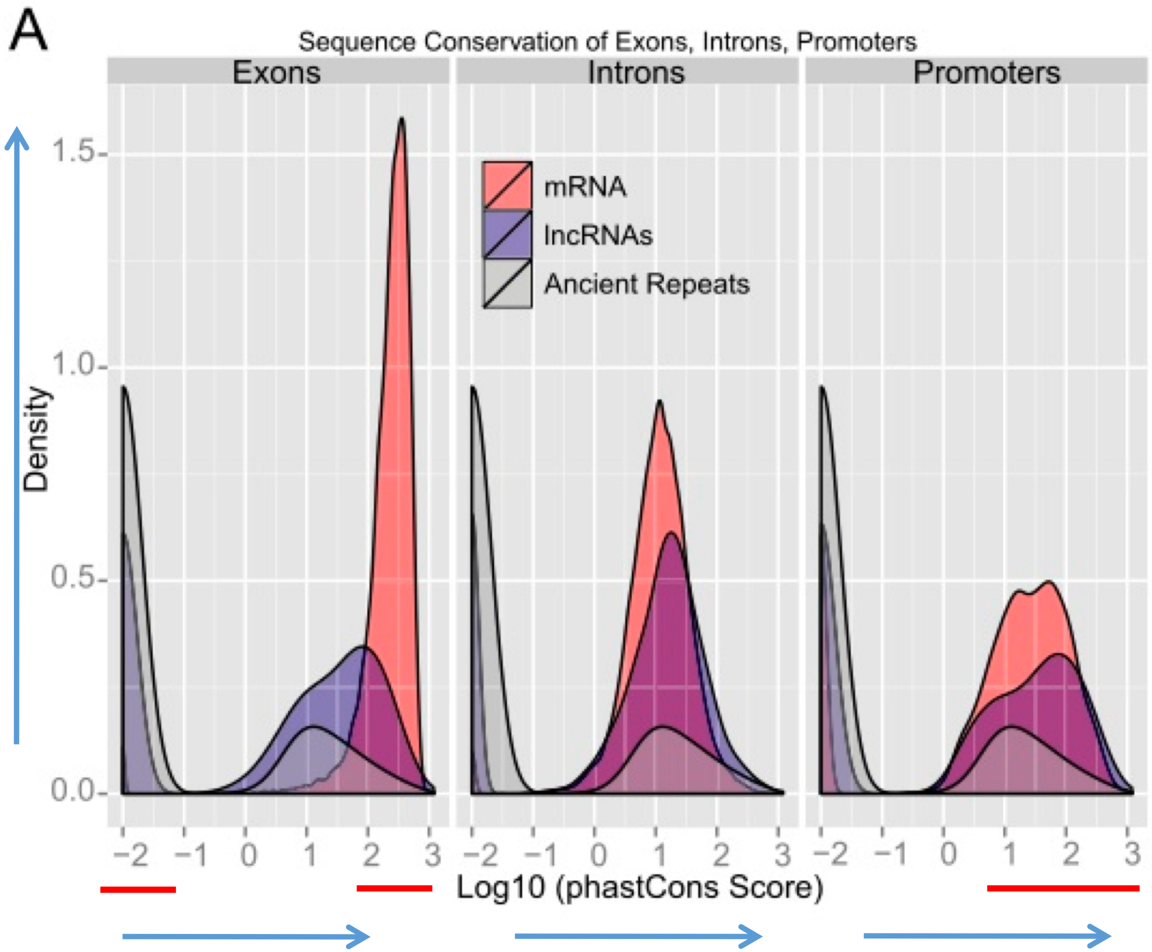


- lncRNAs are uniform → little alternative splicing
- mRNAs: large variety of alternative splicing

COMPARING mRNAs - lncRNAs

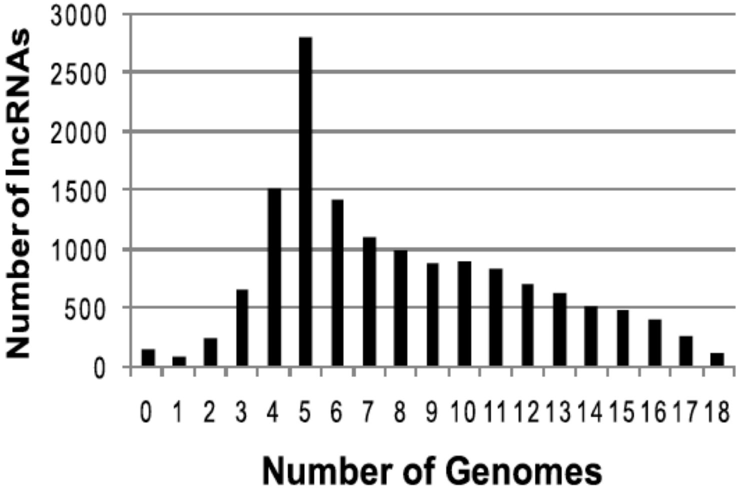
EXONS:
mRNA: high conservation
lncRNA: reduced conservation
But: conservation is higher than mRNA intron conservation

INTRONS:
mRNA: higher conservation than lncRNAs



conservation conservation conservation

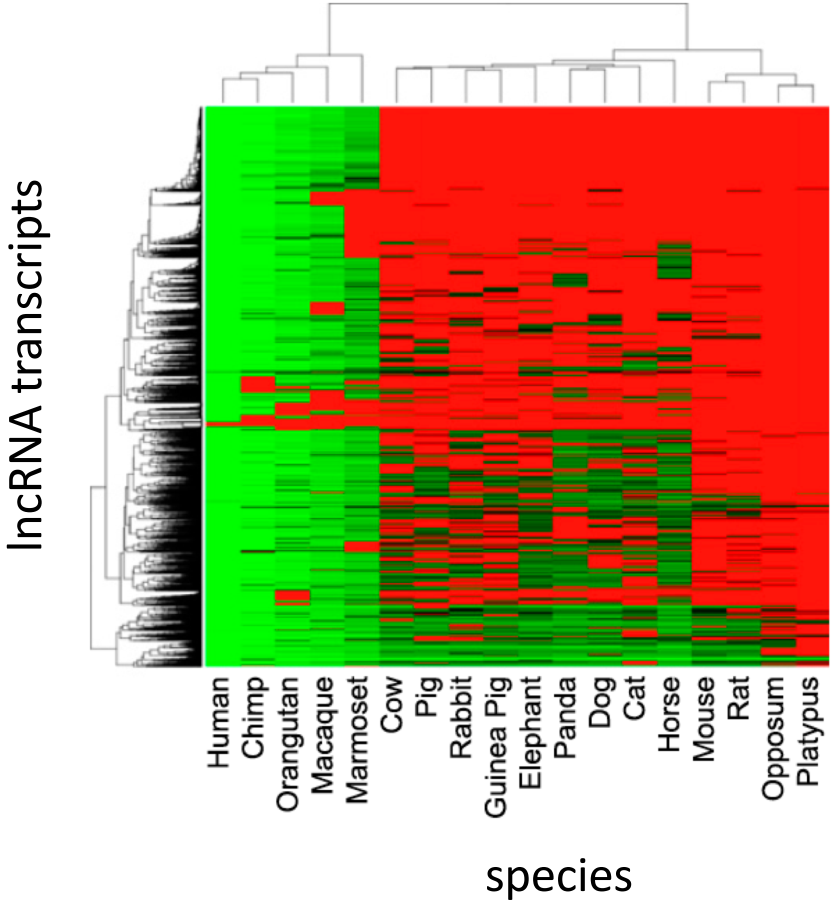
Evolutional conservation of lncRNAs



Most lncRNA transcripts are conserved in 4-6 species

A total of **0.7%** of transcripts appear to be **specific to the human** lineage.
A similar number (**1.0%**) is **found in all of the 18 species** analyzed here.

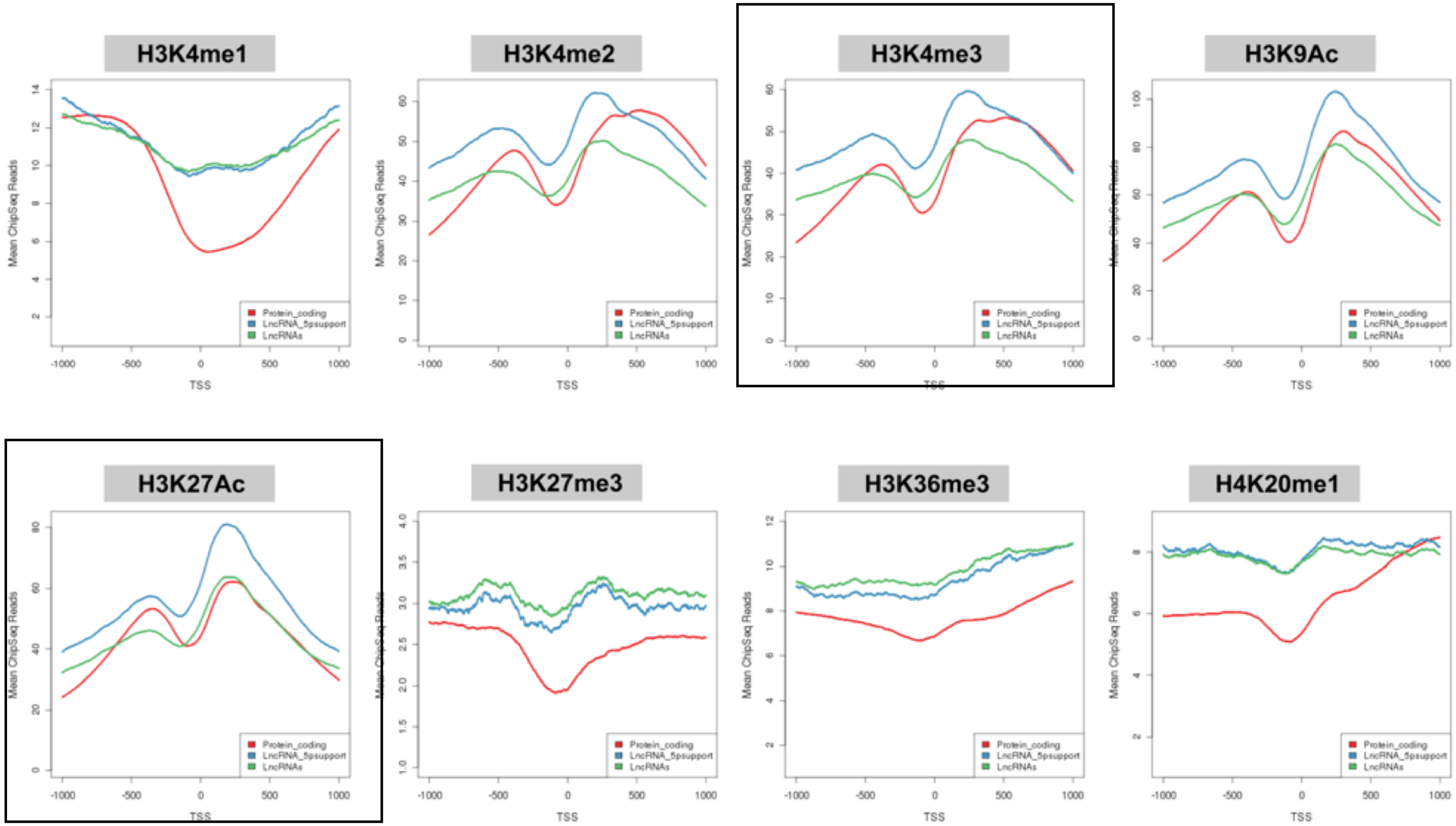
lncRNAs show relevant species specificity



BLAST of human lncRNAs (green) against all available **mammalian** genomes and identify homologous genes in other species:
30% of human lncRNA transcripts (n = 4546) appear to be primate specific.

COMPARING mRNAs - lncRNAs

Chromatin signatures at transcriptional start sites are conserved between mRNAs and lncRNAs: MOST lncRNAs ARE TRANSCRIBED BY RNA Pol II



Chromatin signatures around TSS of protein-coding and lncRNA transcripts expressed in the same cell lines where the signatures were monitored by ChipSeq. Shown on the y-axis is the average density of reads covering the TSS of various gene sets, with position plotted on the x-axis (bp relative to positive strand TSS). Protein coding genes are plotted in red, Gencode v7 lncRNAs in green, and lncRNAs with 5′ experimental support (n=2,793) in blue. N.B. A more extensive analysis of histone modifications in multiple cell types is available at http://big.crg.cat/bioinformatics_and_genomics/lncrna_data.

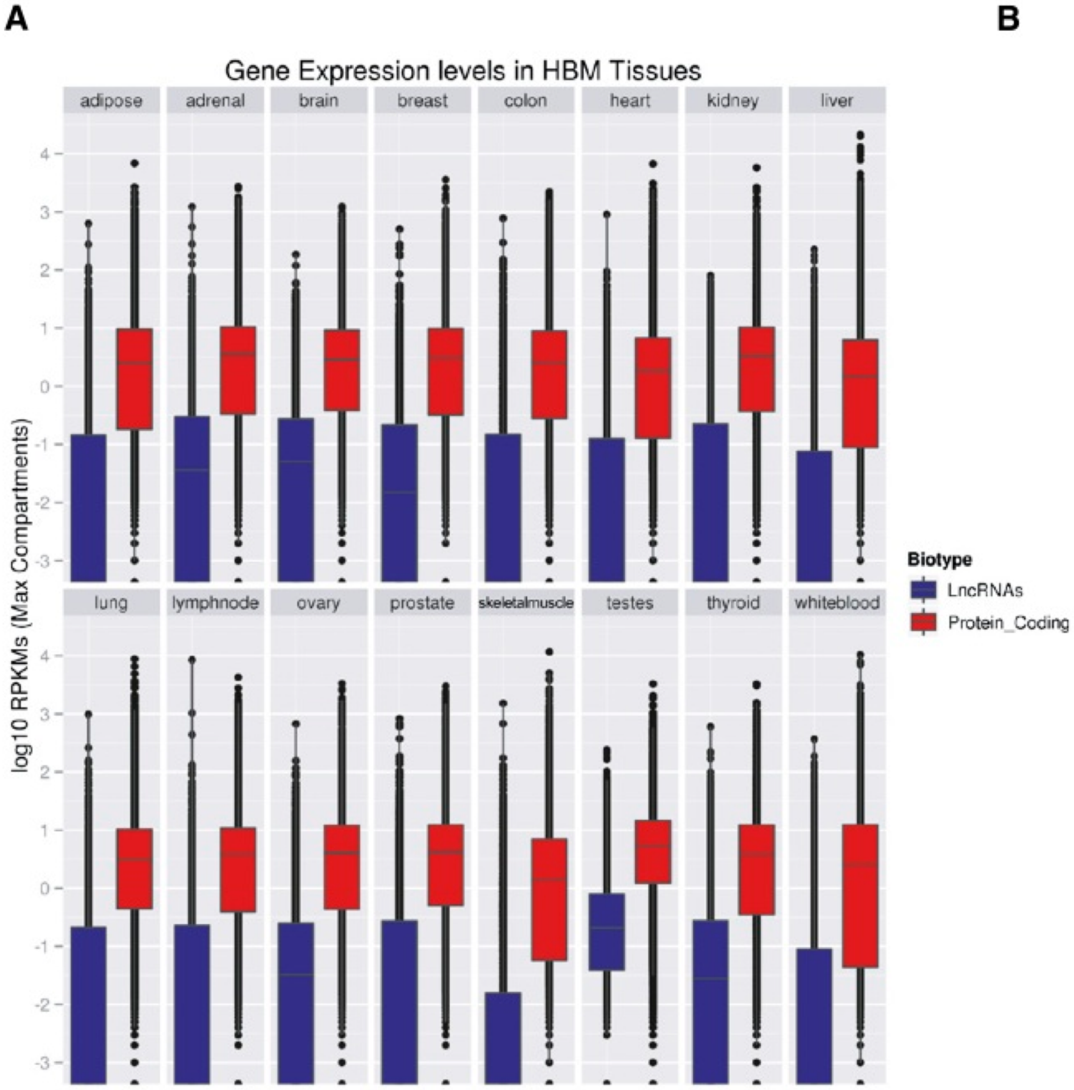
mRNA
lncRNA 5psupport
lncRNA

K562

Not shown:
RNA Pol II peaks at transcription start site of most lncRNAs

COMPARING mRNAs - lncRNAs

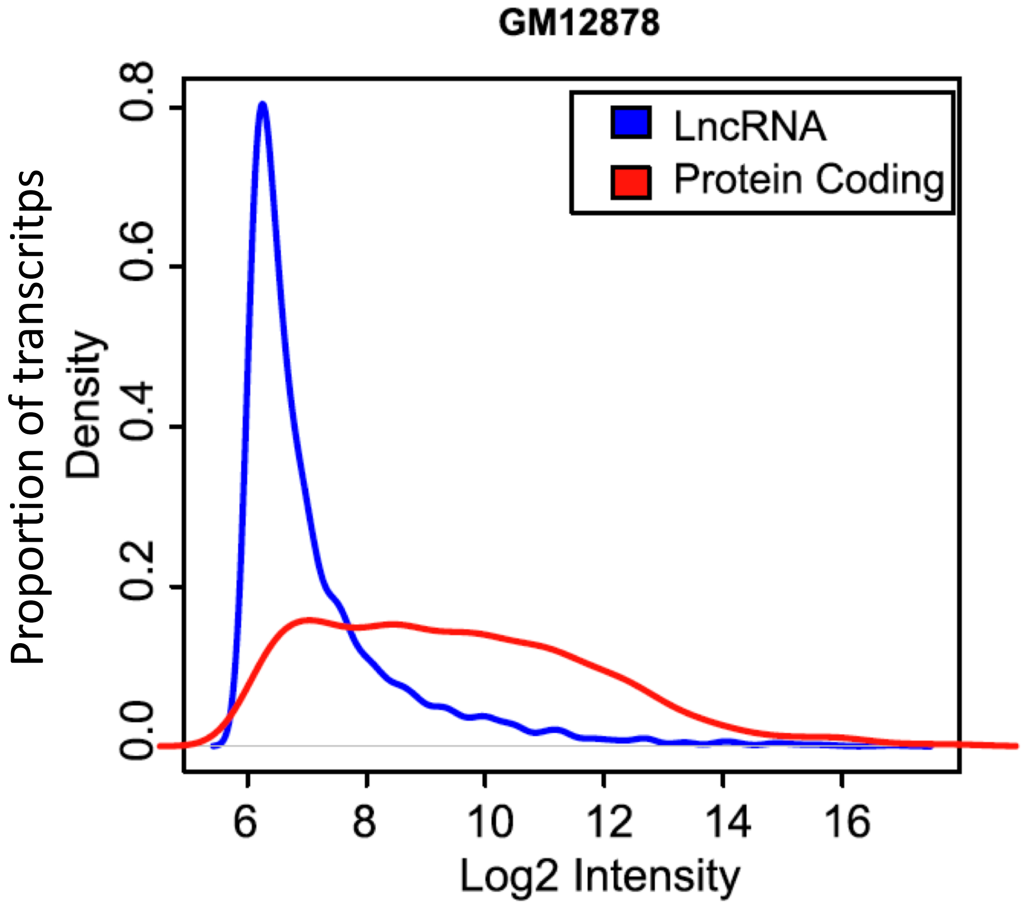
Absolute expression
of lncRNAs is much lower than
mRNA expression



RNA-seq

COMPARING mRNAs - lncRNAs

Absolute expression of lncRNAs is much lower than mRNA expression

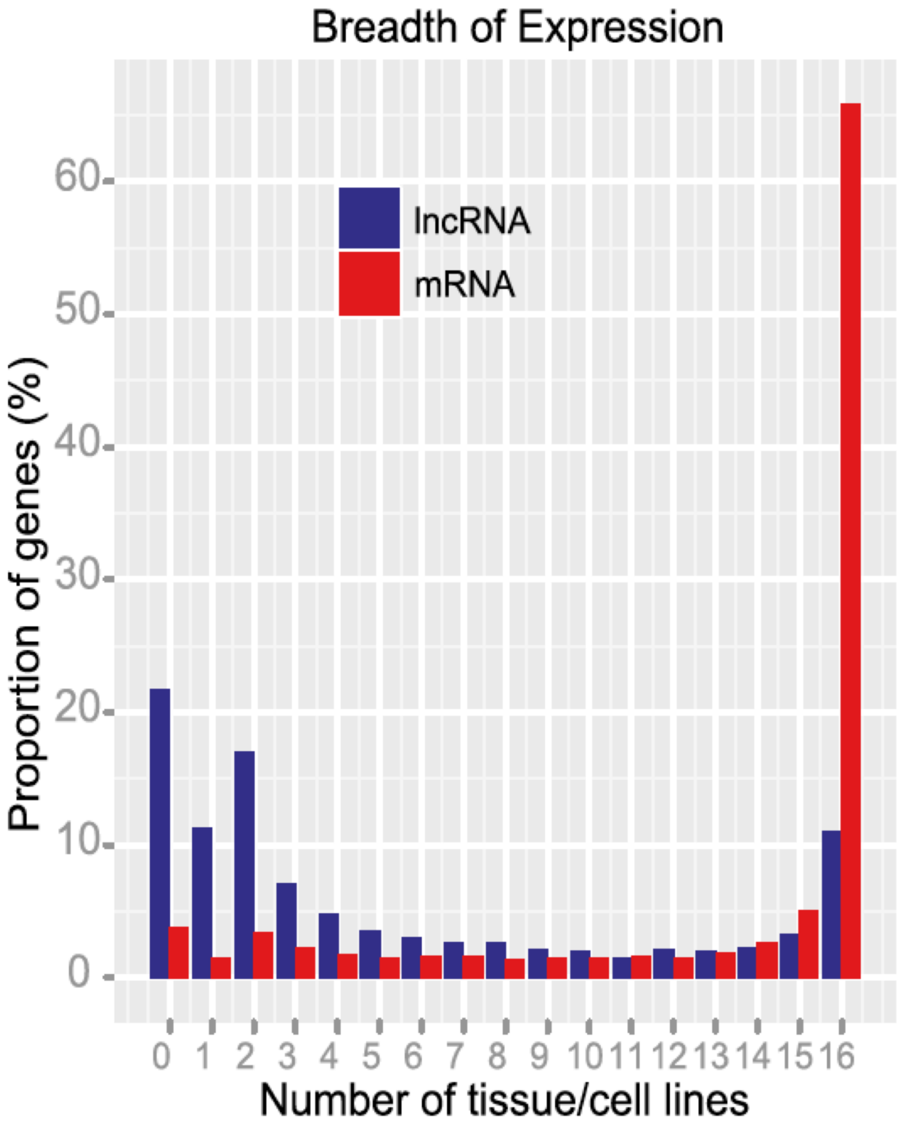


Microarray

COMPARING mRNAs - lncRNAs

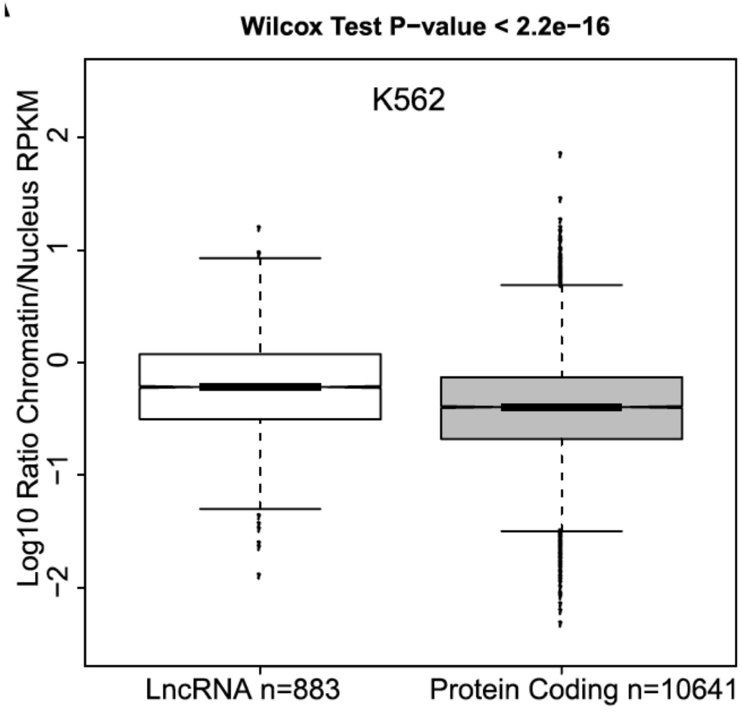
mRNAs show broad expression in different tissues – 65% of mRNAs are expressed in 16 tissues

lncRNAs show tissue/cell soecifc expression – 50% of lncRNAs expressed in 4-5 tissues.

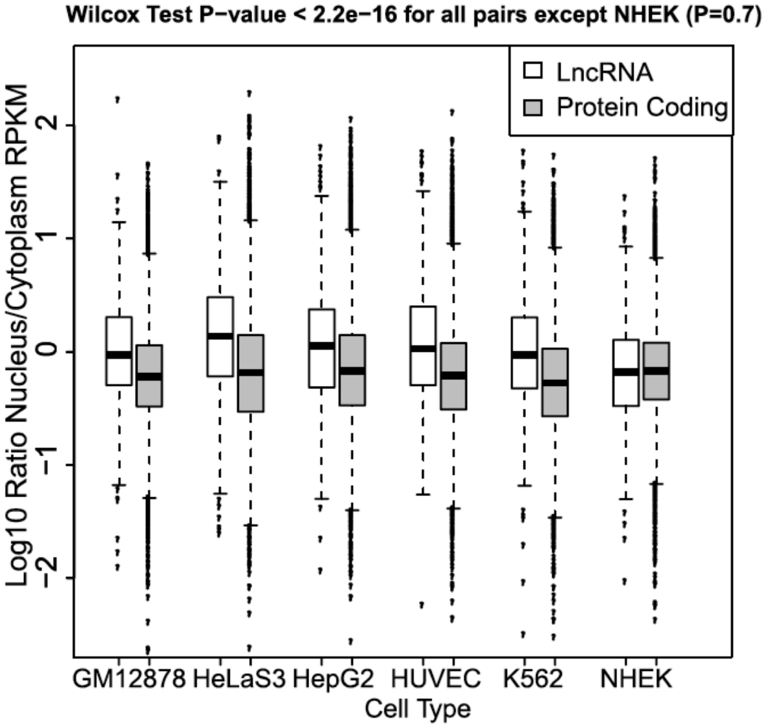


COMPARING mRNAs - lncRNAs

Determination of transcript abundance
Chromatin/nucleus



Determination of transcript abundance
Nucleus/cytoplasma



lncRNAs locate more frequently to chromatin than mRNAs

lncRNAs locate more frequently to the nucleus than mRNAs

Data: RNA seq of RNAs located in the nucleoplasma and sticking to chromatin

Data: RNAseq of RNAs located in the nucleoplasma and cytoplasma

COMPARING mRNAs - lncRNAs

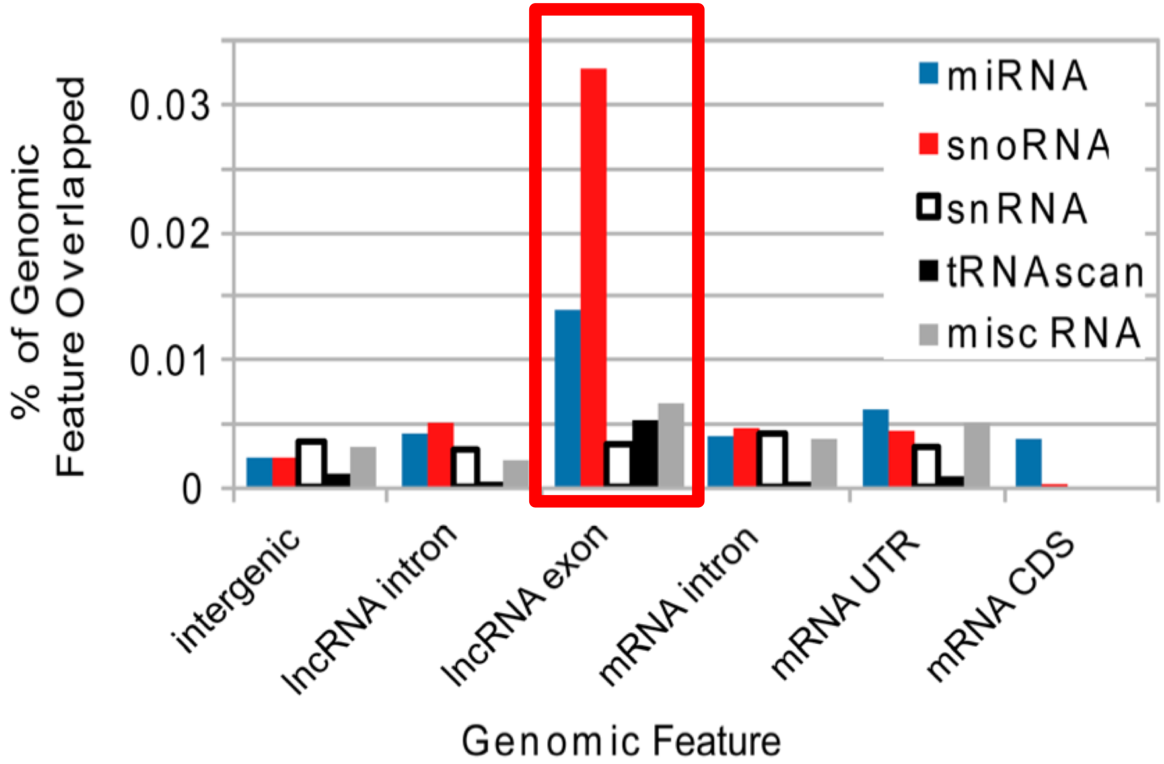


Figure S7: LncRNAs as a source of small RNAs. Shown is the proportion of nucleotides in exons/introns/(UTRS) from protein-coding genes and lncRNAs that overlap different classes of annotated small RNAs. Note that all values refer to cases where the small RNA is on the same strand as the indicated lncRNA/mRNA.

lncRNA exons
Have a remarkable
Potential to encode
functional small RNAs
(**sRNAs = RNAs <200nt**)
HIGHER THAN EXONS OF
mRNAS!!!!

snoRNA:Small nucleolar RNAs (snoRNAs) are a class of small RNA molecules that primarily guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs.
snRNA: Small nuclear ribonucleic acid (snRNA), also commonly referred to as U-RNA, is a class of small RNA molecules that are found within the splicing speckles and Cajal bodies of the cell nucleus in eukaryotic cells.
tRNA: translation
miscRNAs: MiscRNA is short for miscellaneous RNA, a general term for a series of miscellaneous small RNA. It serves a variety of functions, including some enzyme-like catalysis and processing RNA after it is formed. Besides, some of these small RNAs may serve as switches.

COMPARING mRNAs - lncRNAs

lncRNAs represent a big class of functional elements that

- show controlled gene expression
- are processed
- lack protein coding potential
- defined localization
- frequently encode sRNAs
- low overall conservation
- 35% of lncRNAs are primate specific
- expression is rather low – but tightly controlled!
- mostly transcribed by RNA Pol II

FUNCTION: FOR THE VAST MAJORITY OF lncRNAs THE BIOLOGICAL FUNCTION IS UNKNOWN!