

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*

Protein



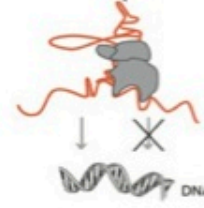
*scaffold
for
protein
complexes*

miRNA



*miRNA
sponge*

DNA



*recruitment
of chromatin
modifying
complexes*

lncRNA

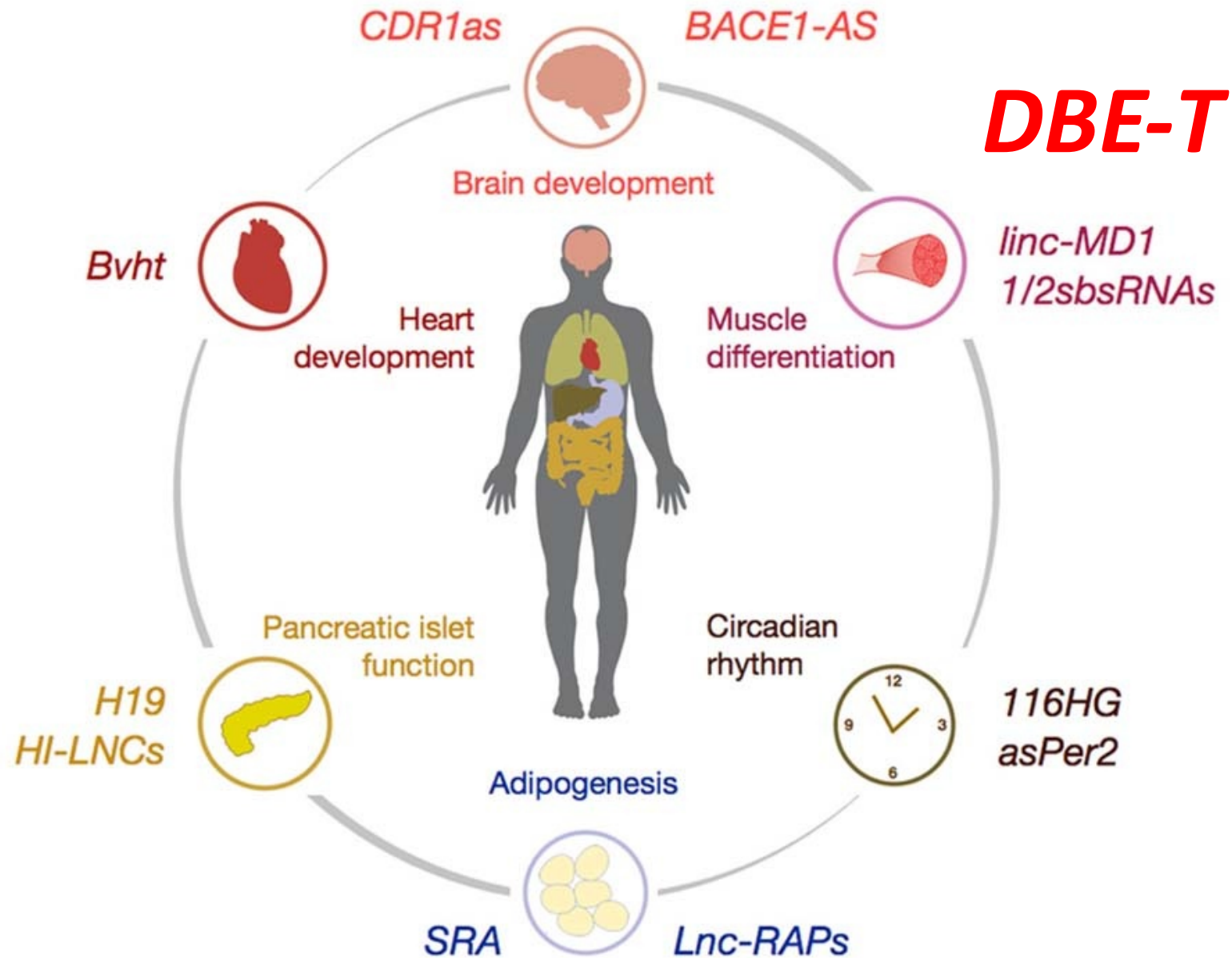


**Protein
function**



Nuclear
and/or
cytoplasmic
localization

lncRNAs in physiology



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}

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

COMPARING mRNAs - lncRNAs

lncRNAs can be

1. Intergenic (lincRNA):

Do not intersect with protein coding gene

2. Genic:

Intersect a protein coding gene

- Exonic
- Intronic
- overlapping

Intergenic lncRNA

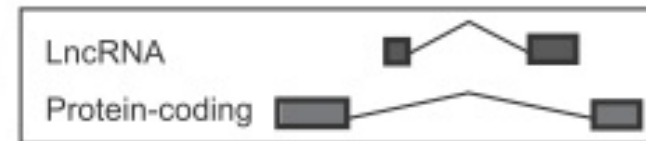


Genic lncRNA

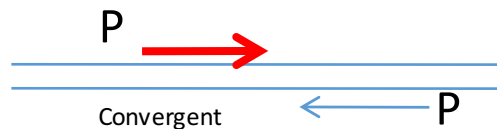
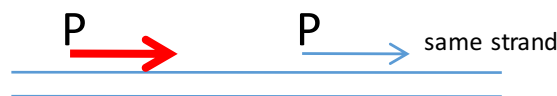
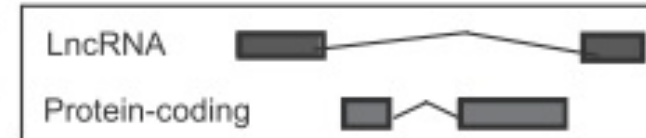
Exonic



Intronic



Overlapping



| 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 | |

P: promoter

Why? Why? Why?

COMPARING mRNAs - lncRNAs

A. lncRNAs do not have coding potential:

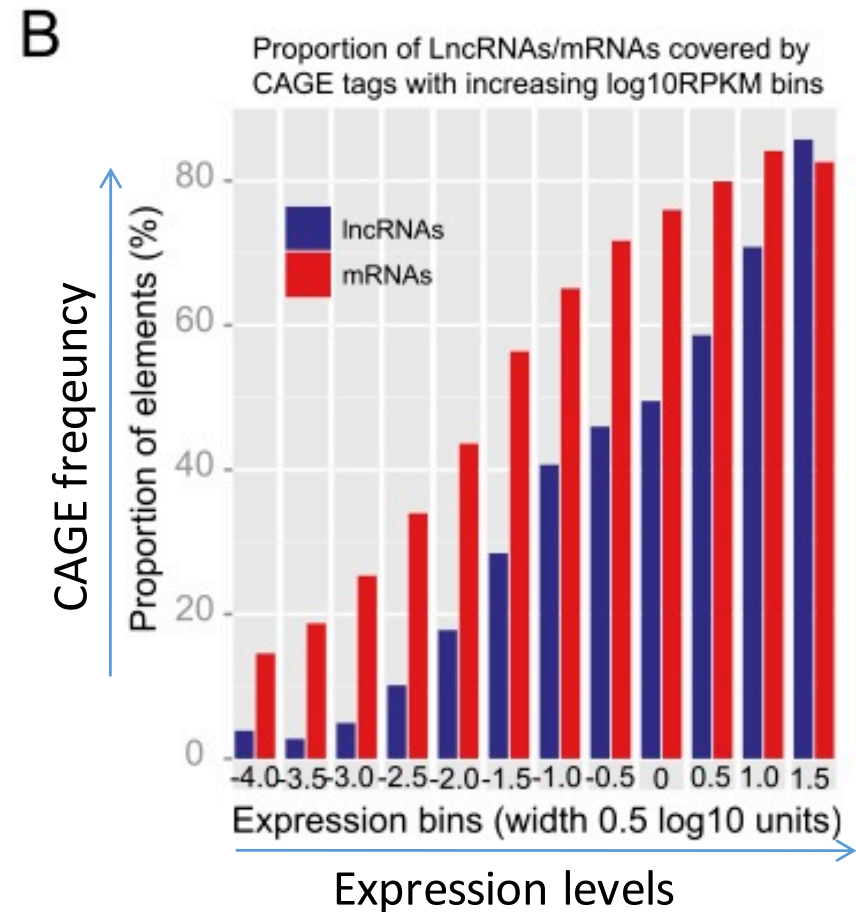
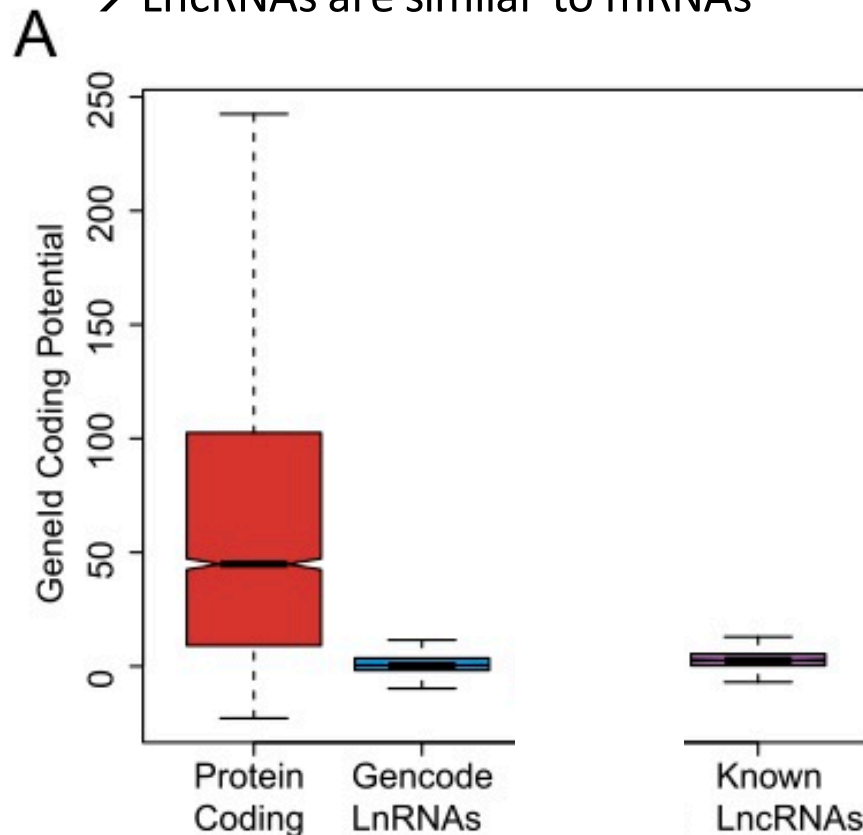
→ Longest possible ORF was searched in mRNA/lncRNAs

B. CAGE tags can be found in lncRNAs → defined transcriptional start site

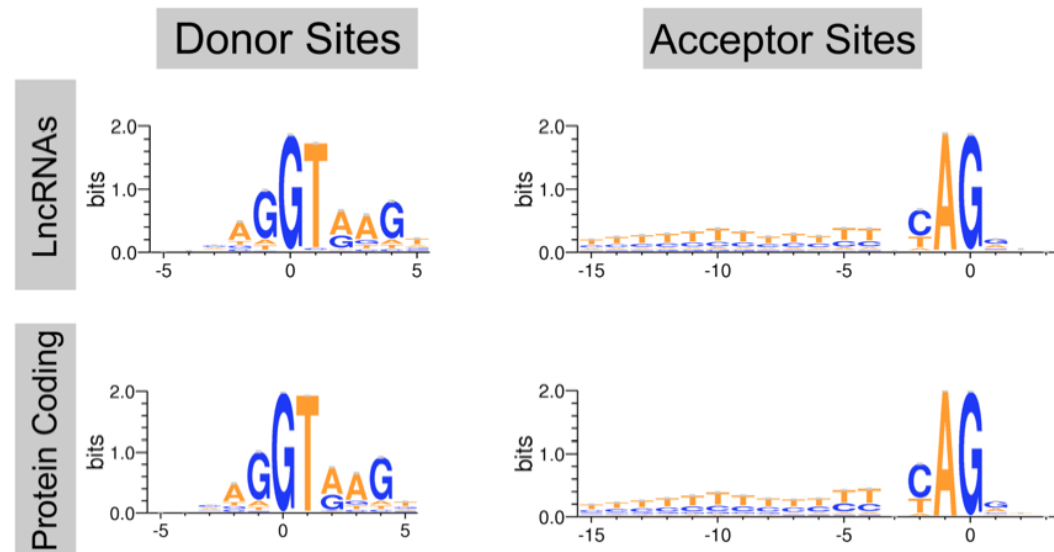
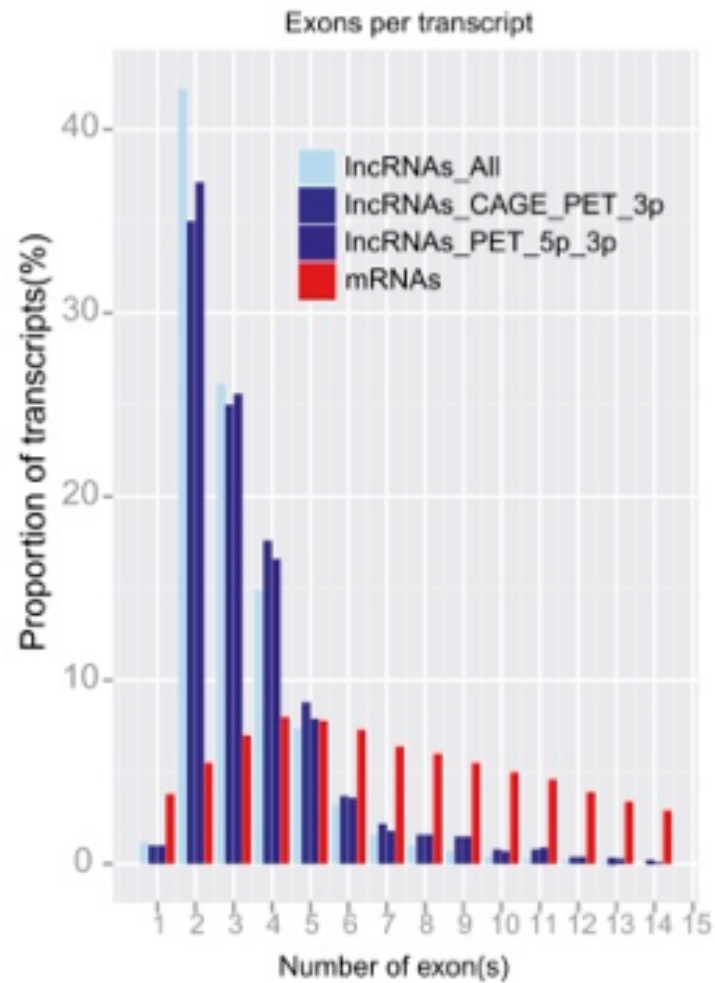
CAGE tag frequency increases with increased lncRNA expression levels.

mRNAs are characterized by more CAGE tags

→ lncRNAs are similar to mRNAs



COMPARING mRNAs - lncRNAs



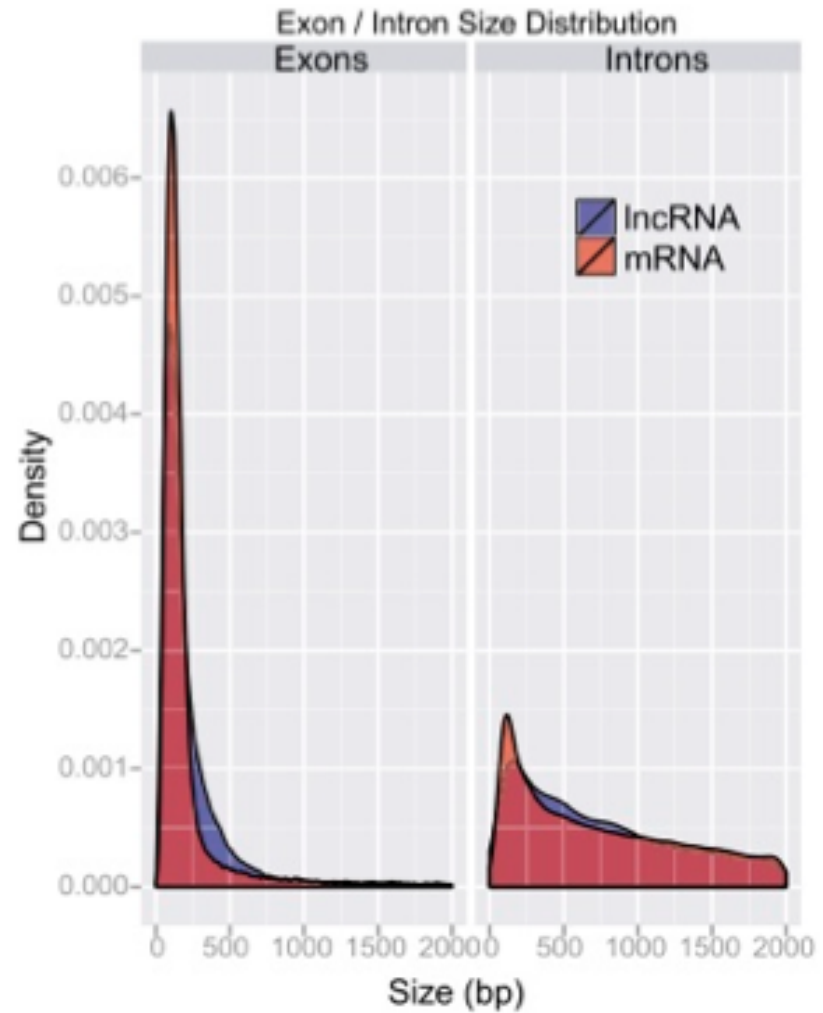
SPLICING:

- **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%)

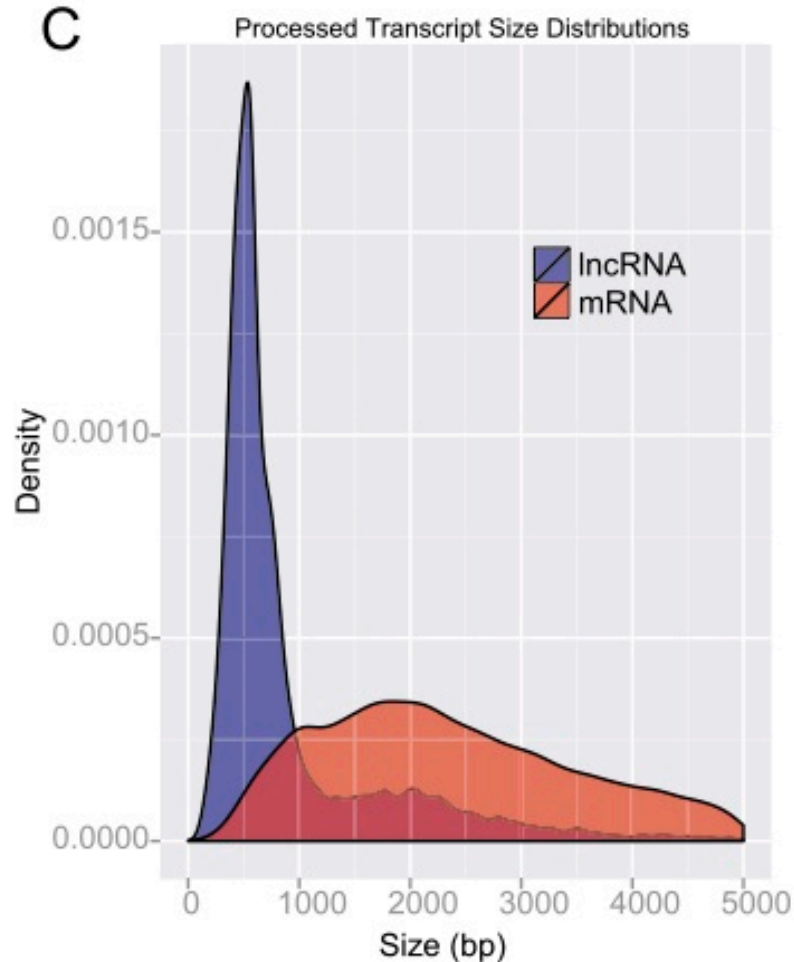
BUT: lncRNAs contain fewer INTRONS!! most lncRNA have only 1 intron!!!; mRNAs 3-6

COMPARING mRNAs - lncRNAs

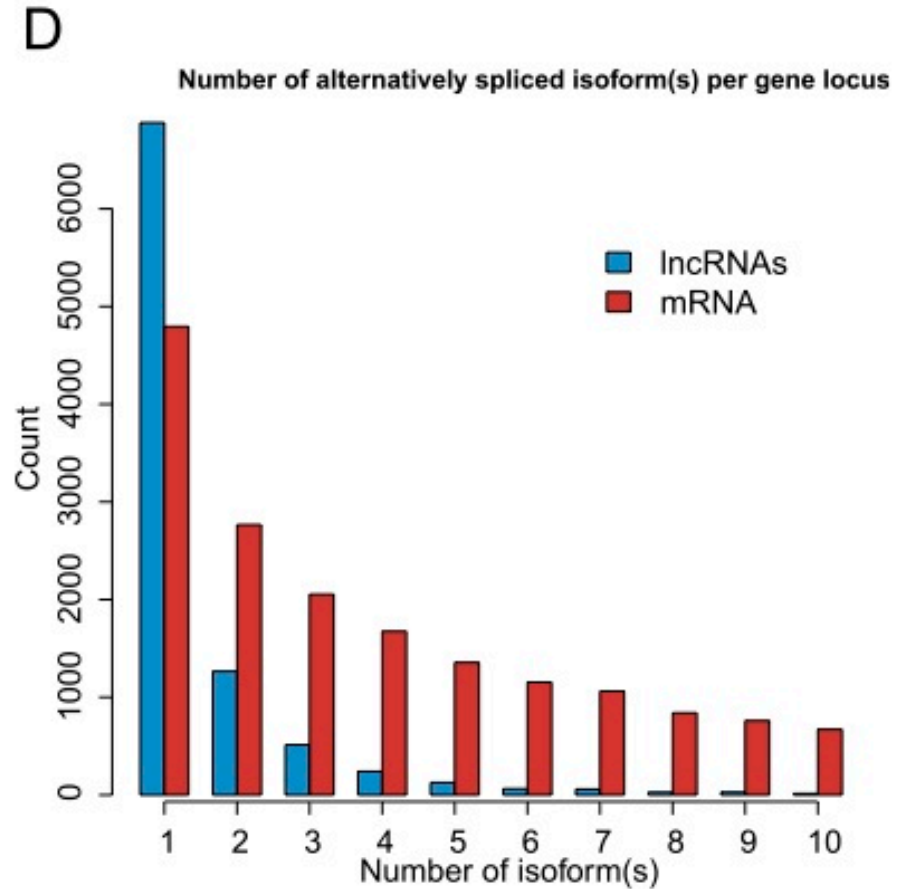
Introns/Exons from lncRNAs are slightly longer



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:

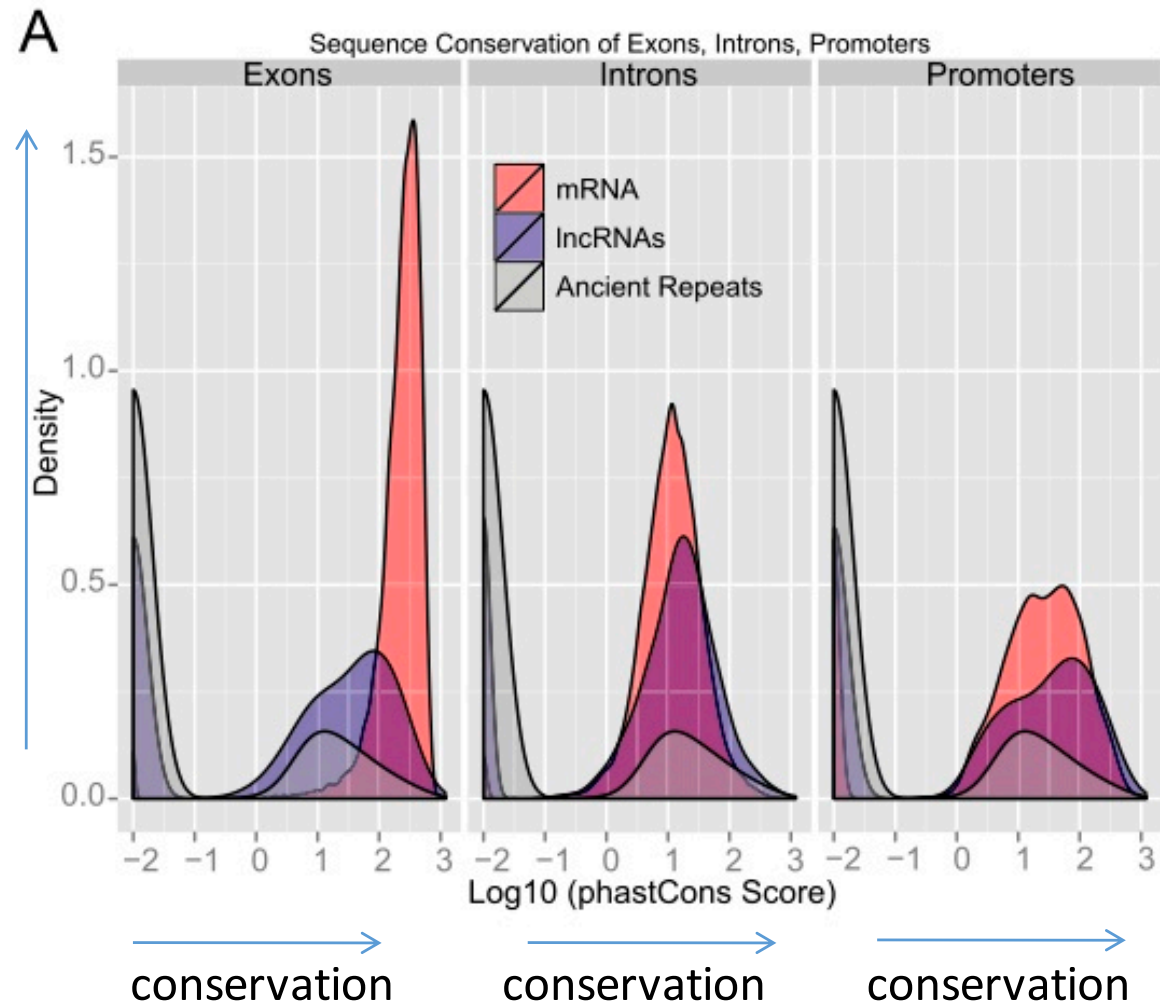
mRNA: high conservation

lncRNA: reduced conservation

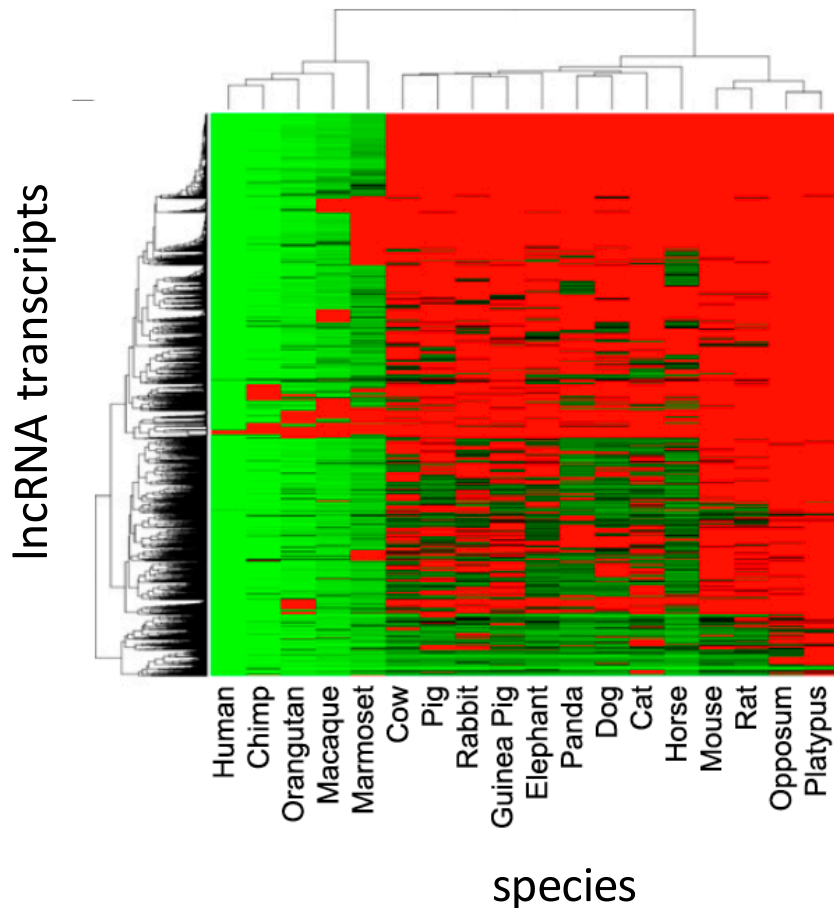
But: conservation is higher than mRNA intron conservation

INTRONS:

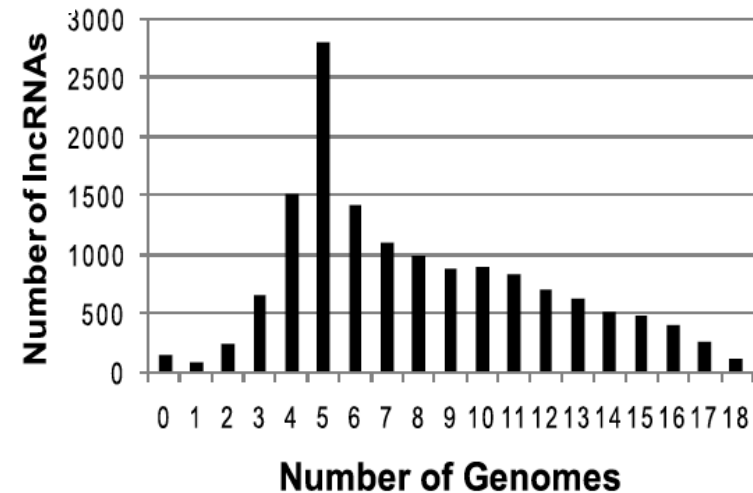
mRNA: higher conservation than lncRNAs



Evolutional conservation of lncRNAs



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



Most lncRNA transcripts are conserved in 4-6 species

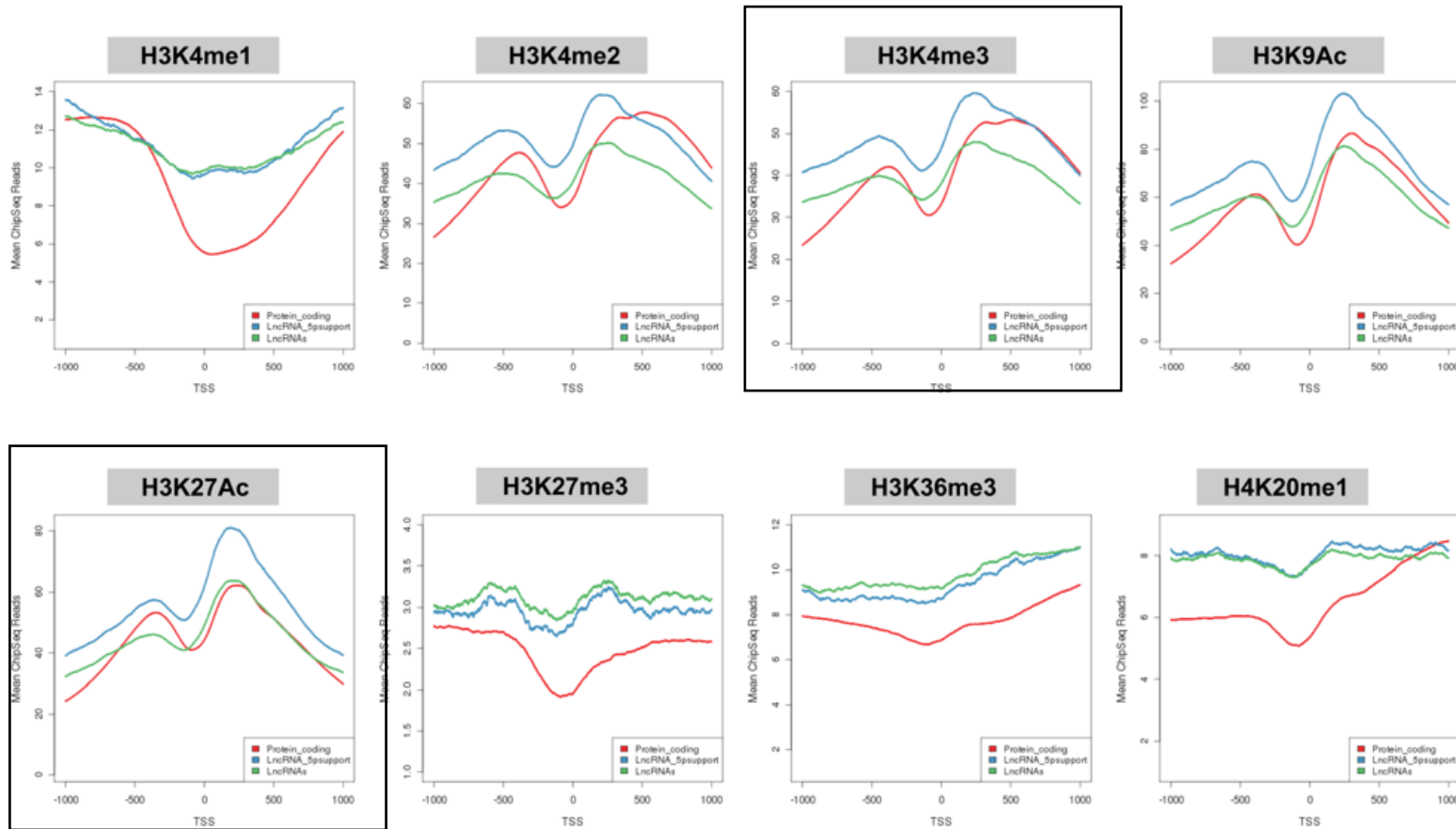
A total of **0.7%** (101) of transcripts appear to be **specific to the human lineage.**

A similar number (134; **1.0%**) is **found in all of the 18 species** analyzed here.

lncRNAs show relevant species specificity

COMPARING mRNAs - lncRNAs

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



mRNA
lncRNA 5psupport
lncRNA

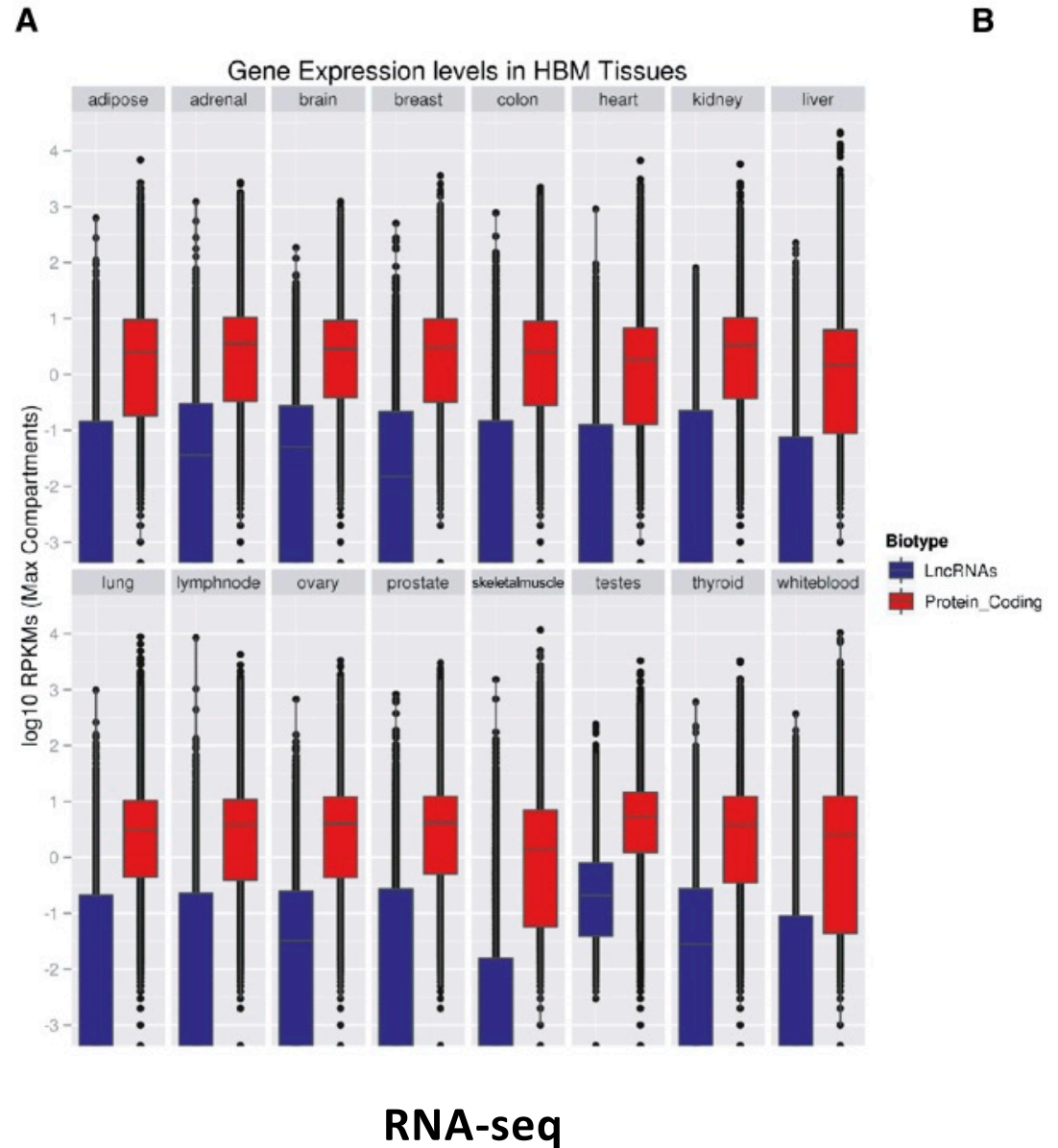
K562

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

Chromatin signatures around TSS of protein-coding and lncRNA transcripts expressed in the same cell lines where the signatures were monitored by ChIP-seq. 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.

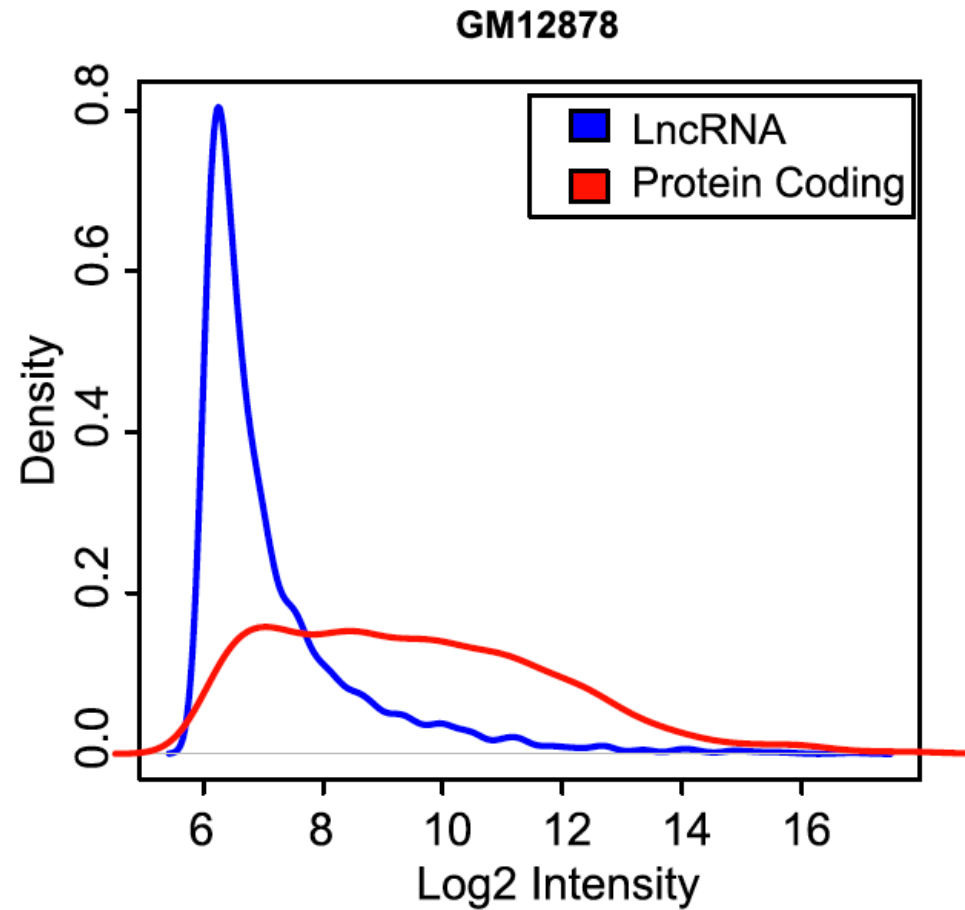
COMPARING mRNAs - lncRNAs

Absolute expression of lncRNAs is much lower than mRNA expression



COMPARING mRNAs - lncRNAs

Absolute expression of lncRNAs is much lower than mRNA expression

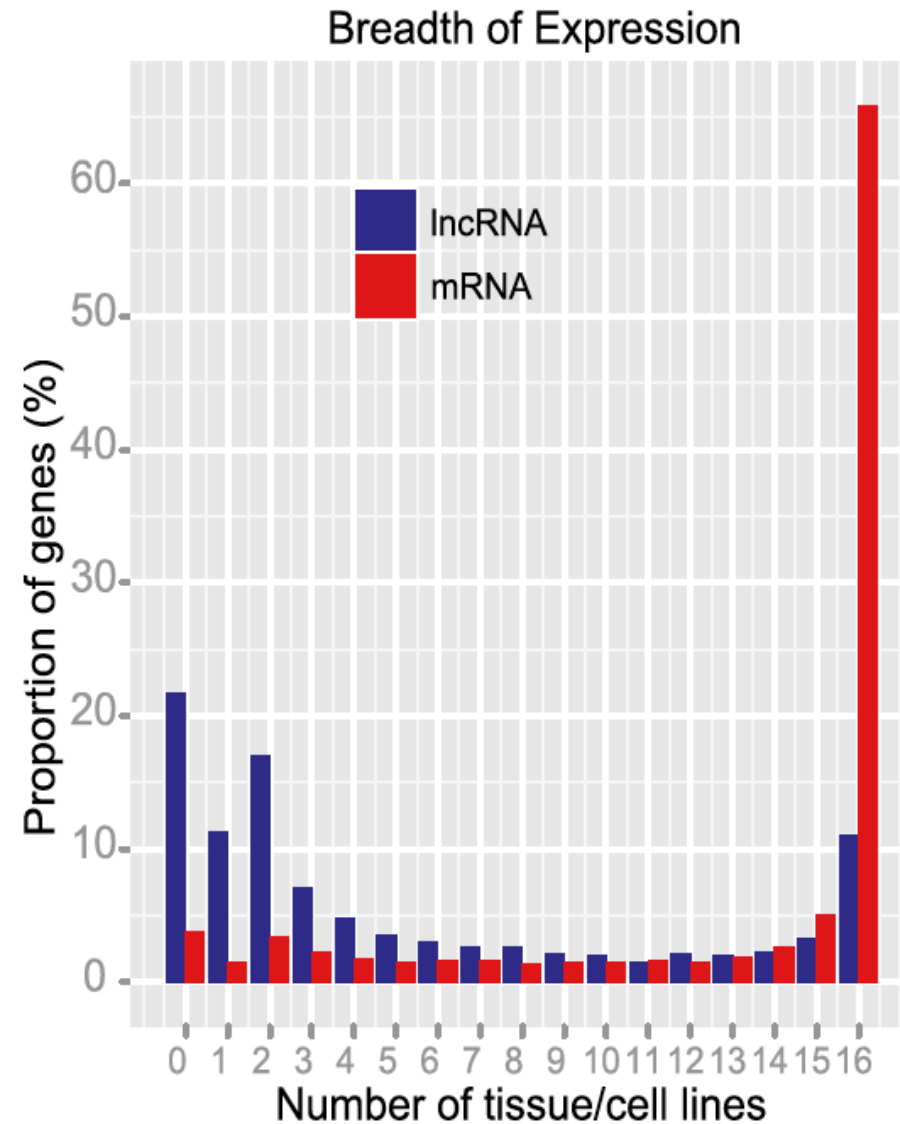


Microarray

COMPARING mRNAs - lncRNAs

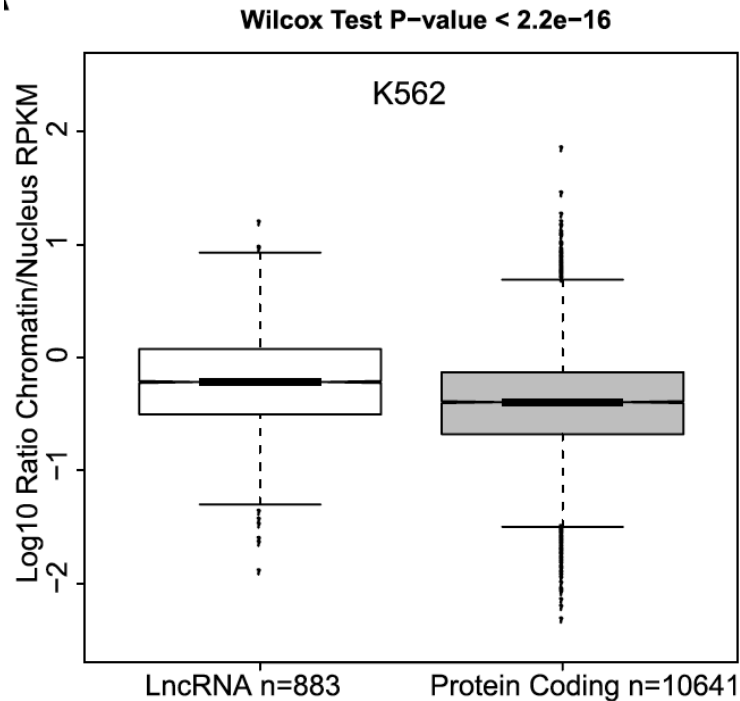
mRNAs show broad expression in different tissues – 65% of mRNAs are expressed in 16 tissues

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



COMPARING mRNAs - lncRNAs

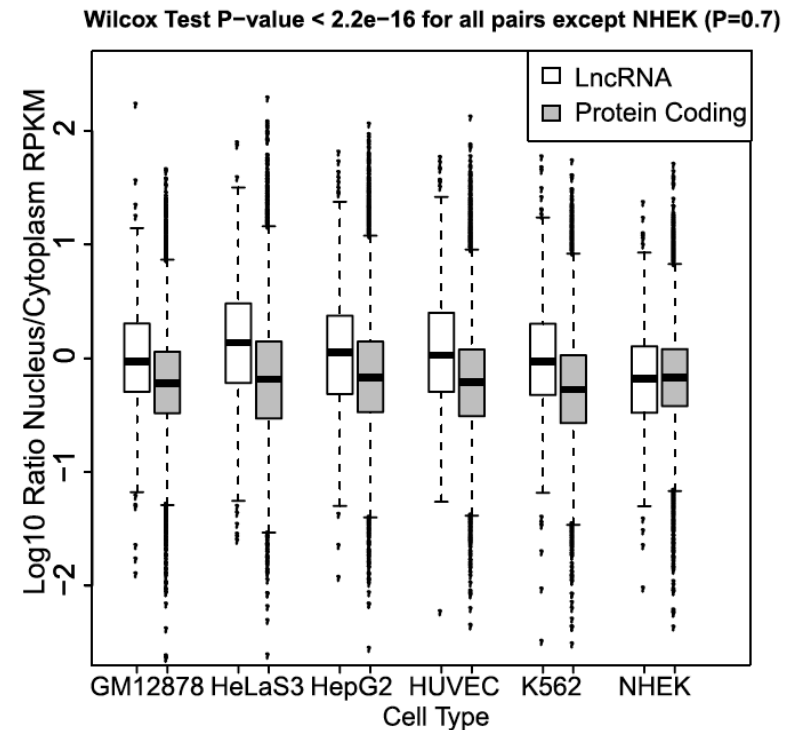
Determination of transcript abundance Chromatin/nucleus



lncRNAs locate more frequently to chromatin than mRNAs

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

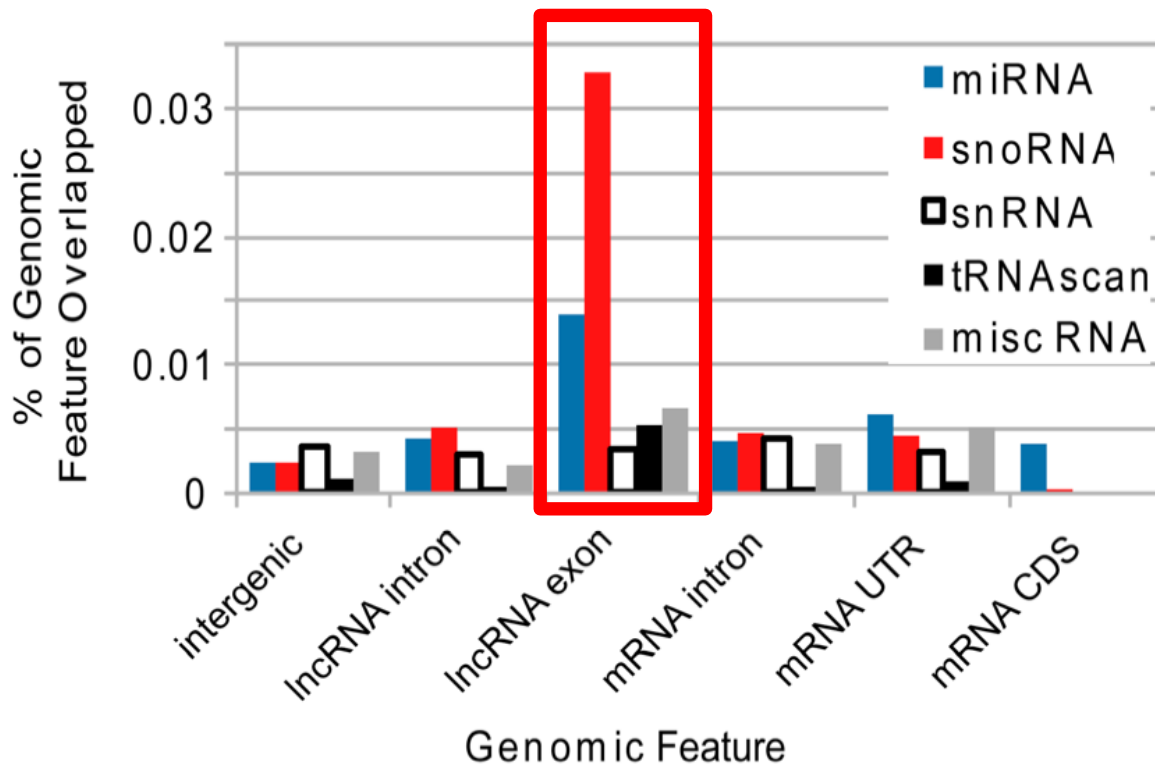
Determination of transcript abundance Nucleus/cytoplasm



lncRNAs locate more frequently to the nucleus than mRNAs

Data: RNAseq of RNAs located in the nucleoplasm and cytoplasm

COMPARING mRNAs - lncRNAs



lncRNA exons
 Have a remarkable
 Potential to encode
 functional small RNAs
 (sRNAs)
**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.

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.

COMPARING mRNAs - lncRNAs

lncRNAs represent a big class of functional elements that

- 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 controlled!
- mostly transcribed by RNA Pol II

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