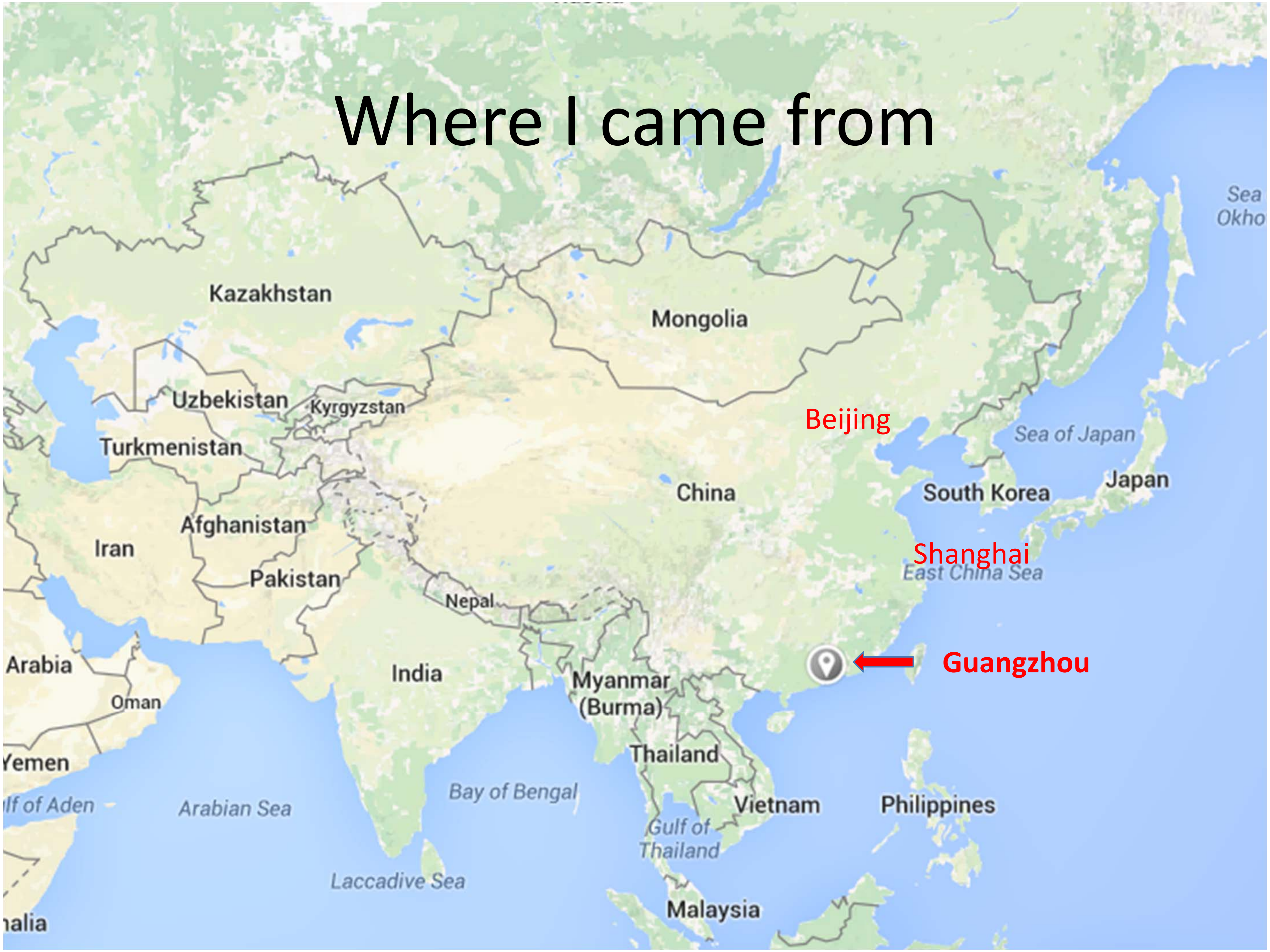
Tissue-based map of the mouse translatome

Zhi Xie
Sun Yat-sun University,
Guangzhou

2017-06-21

Where I came from



Functional diversity of tissues

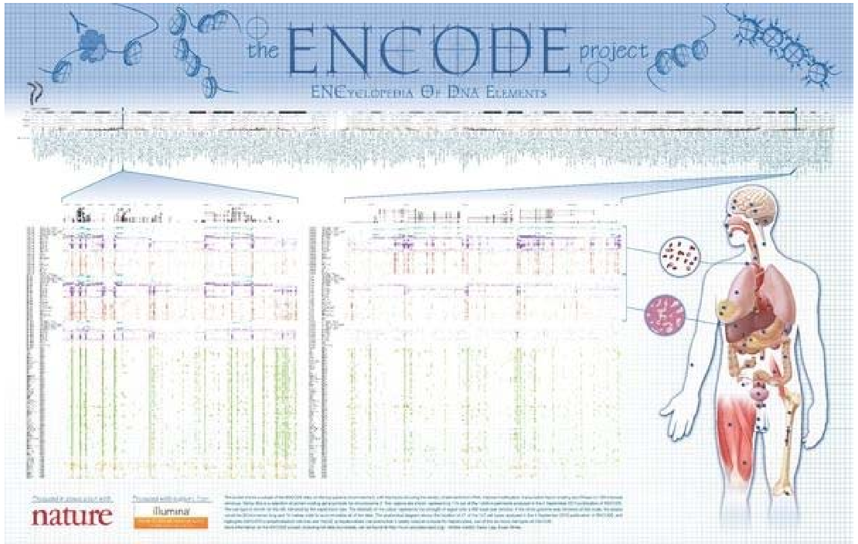
- The physiological functions of mammalian tissues are achieved through the regulation of gene expression in a tissue-restricted manner

- Quantifying gene expression variability across different tissues is important to understand organ morphogenesis and function

Tissue-specific gene expression

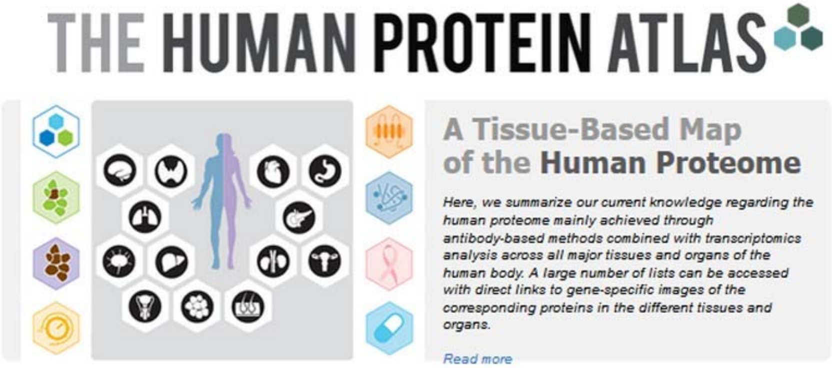
Encyclopedia of DNA Elements (ENCODE)

Genomic, epigenomic and transcriptional levels



Human Protein Atlas (HPA)

Proteomic level

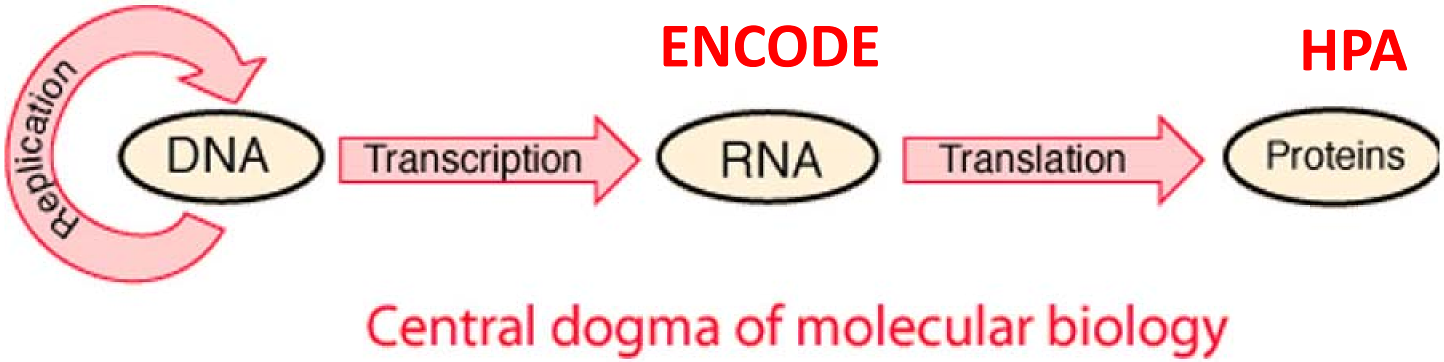

THE HUMAN PROTEIN ATLAS

A Tissue-Based Map of the Human Proteome

Here, we summarize our current knowledge regarding the human proteome mainly achieved through antibody-based methods combined with transcriptomics analysis across all major tissues and organs of the human body. A large number of lists can be accessed with direct links to gene-specific images of the corresponding proteins in the different tissues and organs.

Read more

Control of Gene Expression



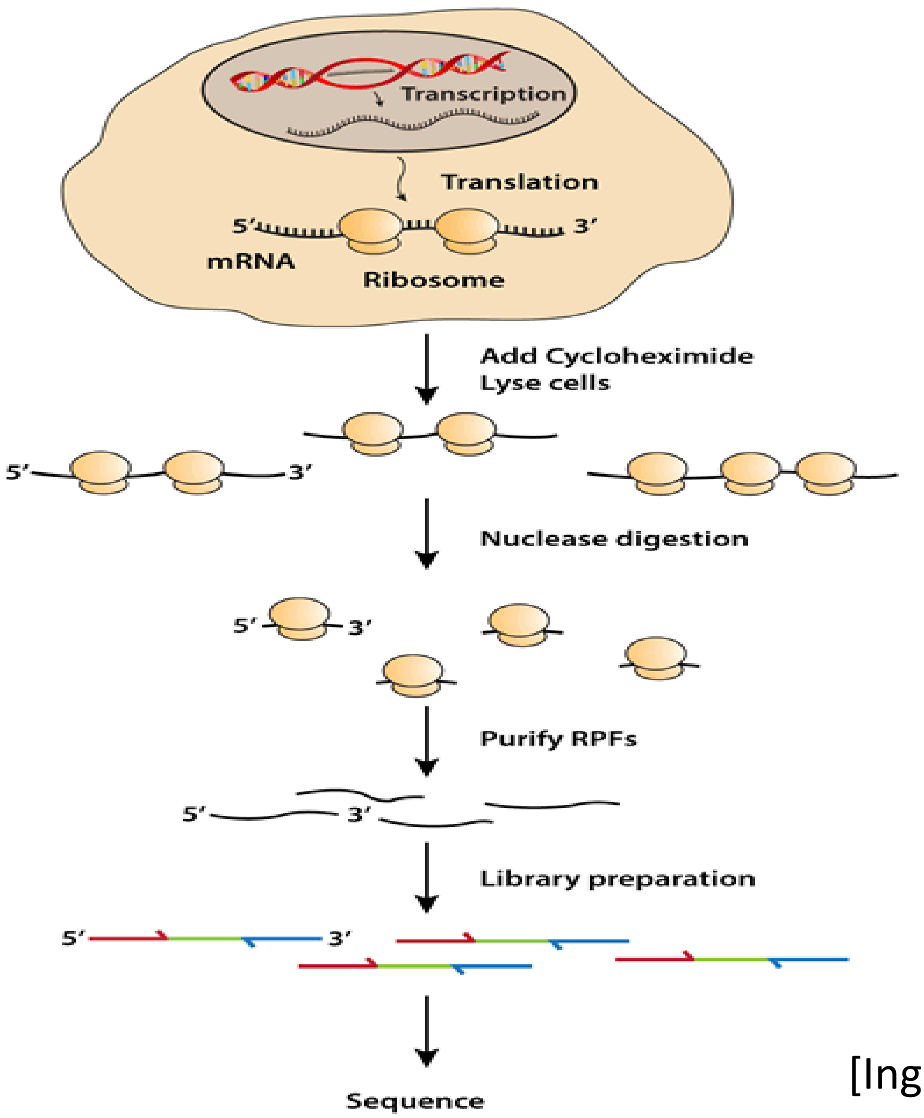
ENCODE

HPA

Central dogma of molecular biology

Translation： **Ribosomes read a messenger RNA and make protein according to its instruction**

Analysis of global translation via ribosome profiling



[Ingolia, Science, 2009]

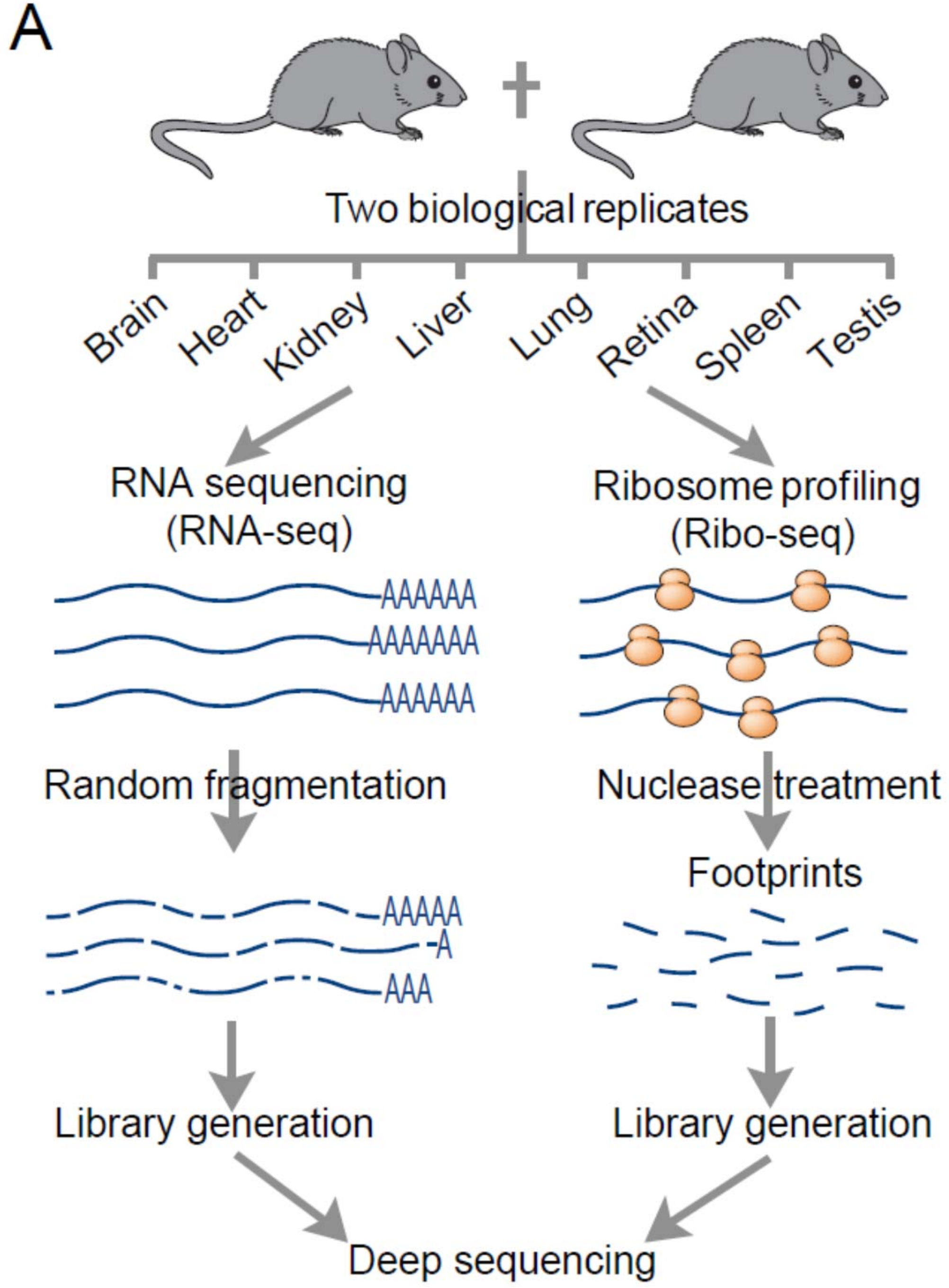
Why Ribosome Profiling

- Ribosome Profiling captures RNA bound by ribosomes giving a snapshot of active translation in the cell
- Tends to provide a better proxy for protein levels in the cell than RNA-seq
- Translational efficiency, non-canonical translational events, micropeptides encoded from lncRNAs

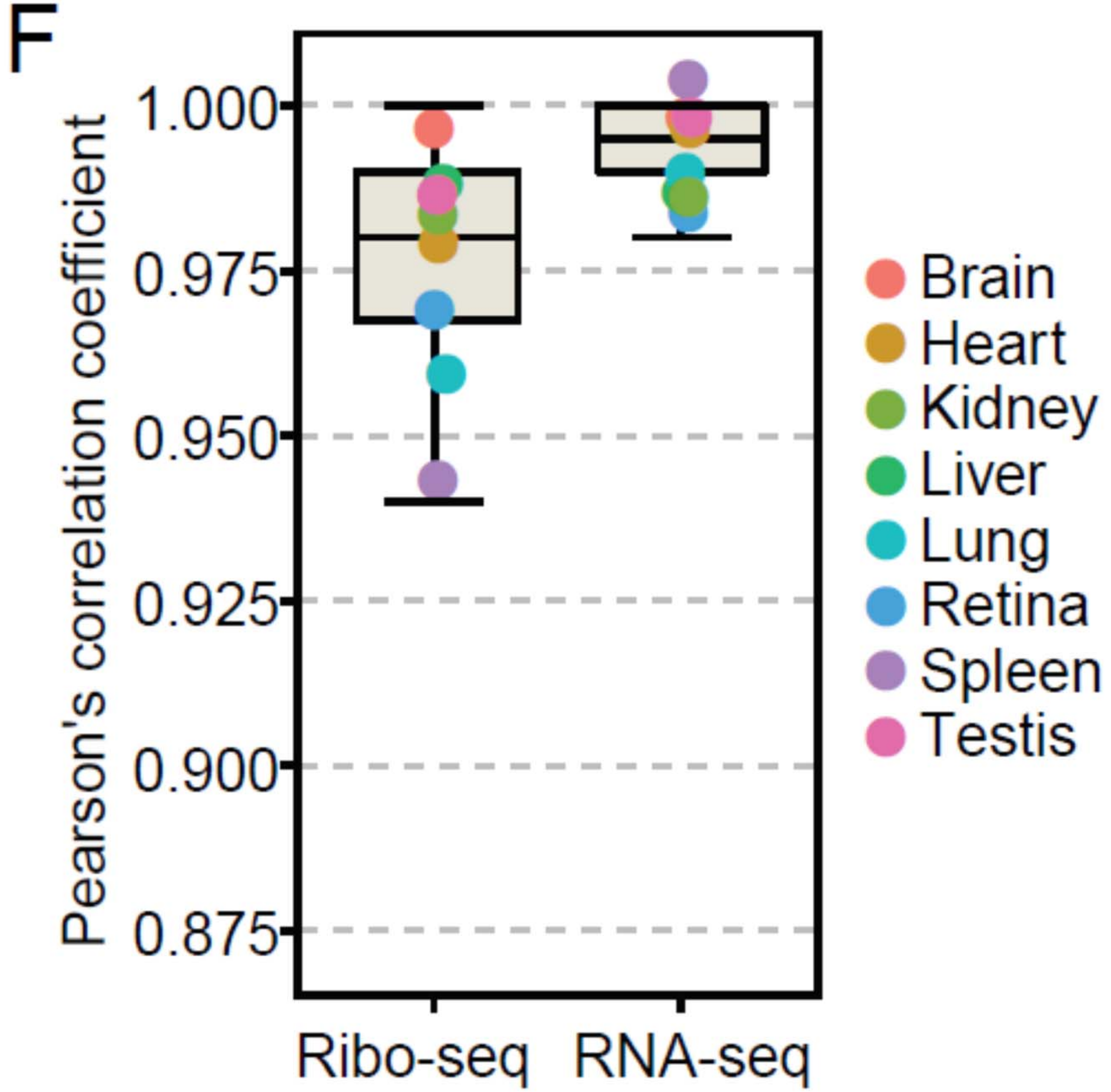
Research Goal

To understand role of RNA translation in maintaining tissue physiology

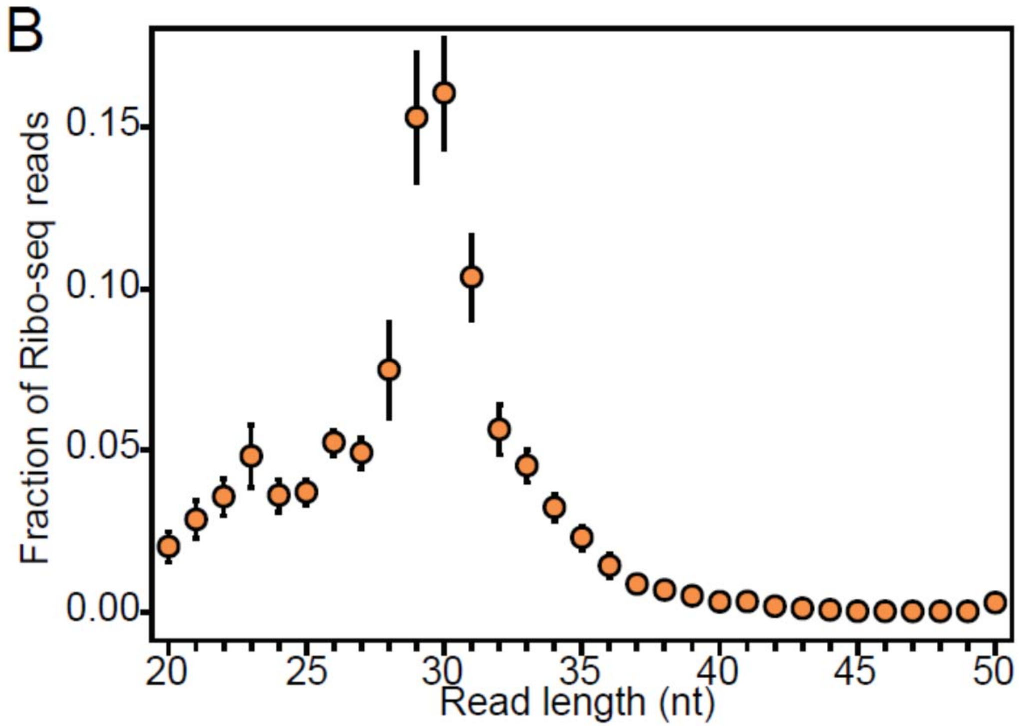
Experimental Design



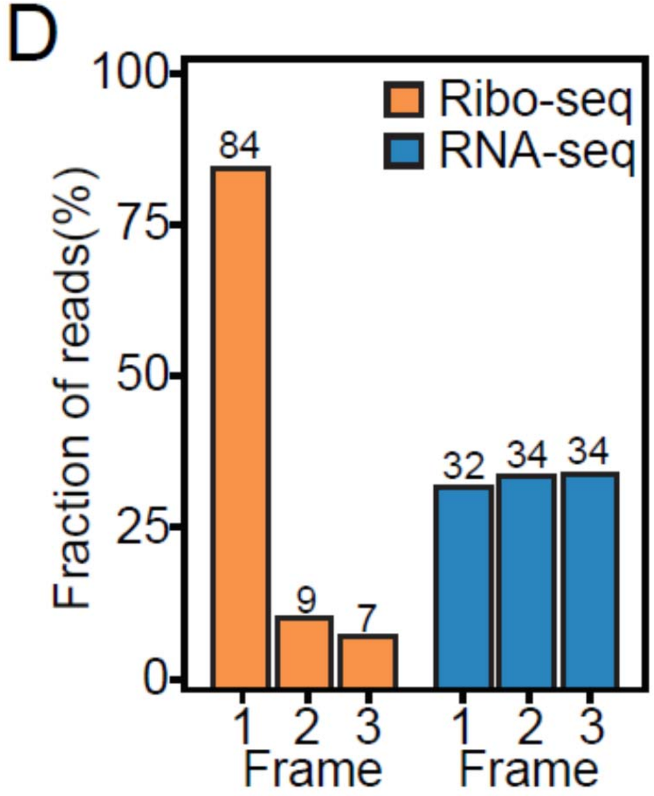
Reproducibility of replicated samples



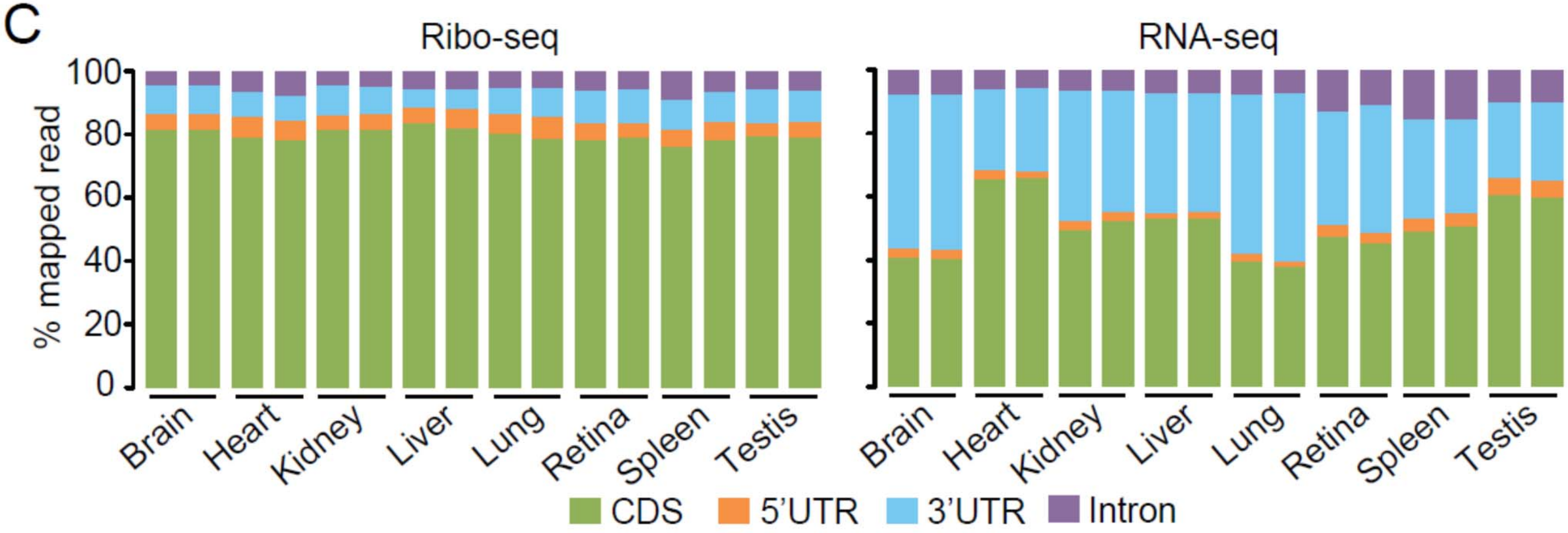
Size distribution of Ribo-seq



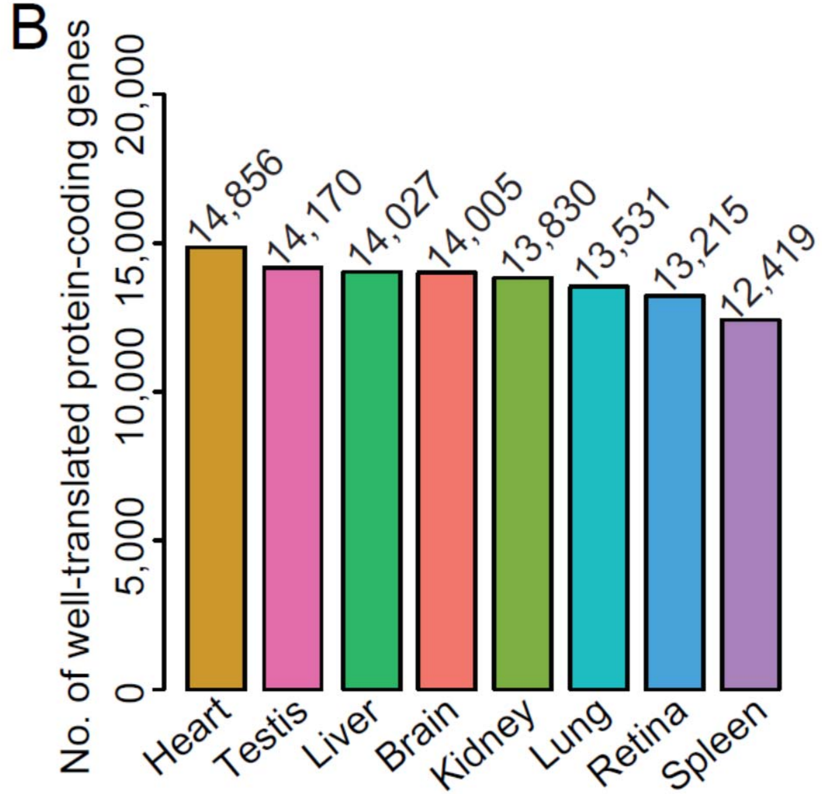
Reading frame analysis



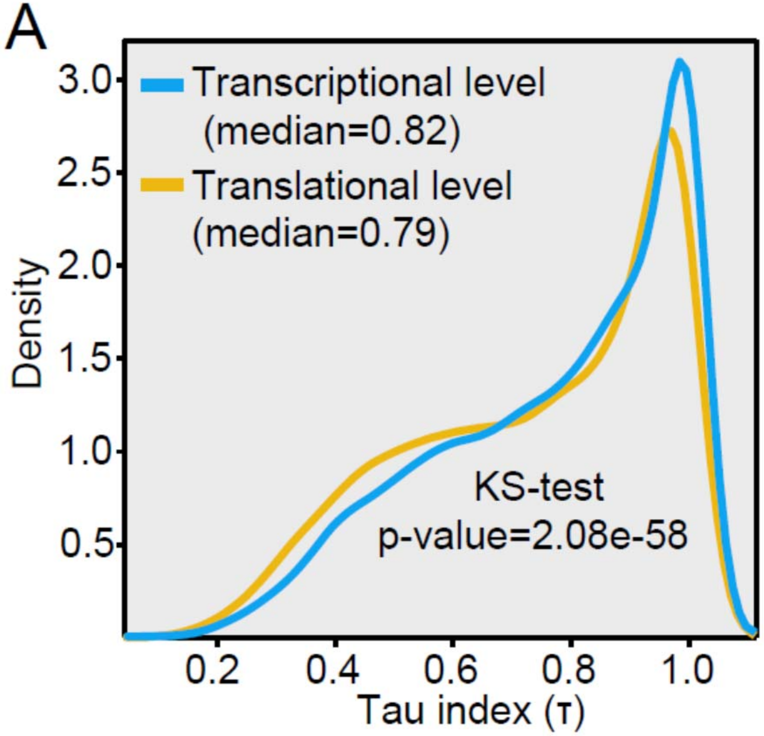
Distribution of reads in genomic regions



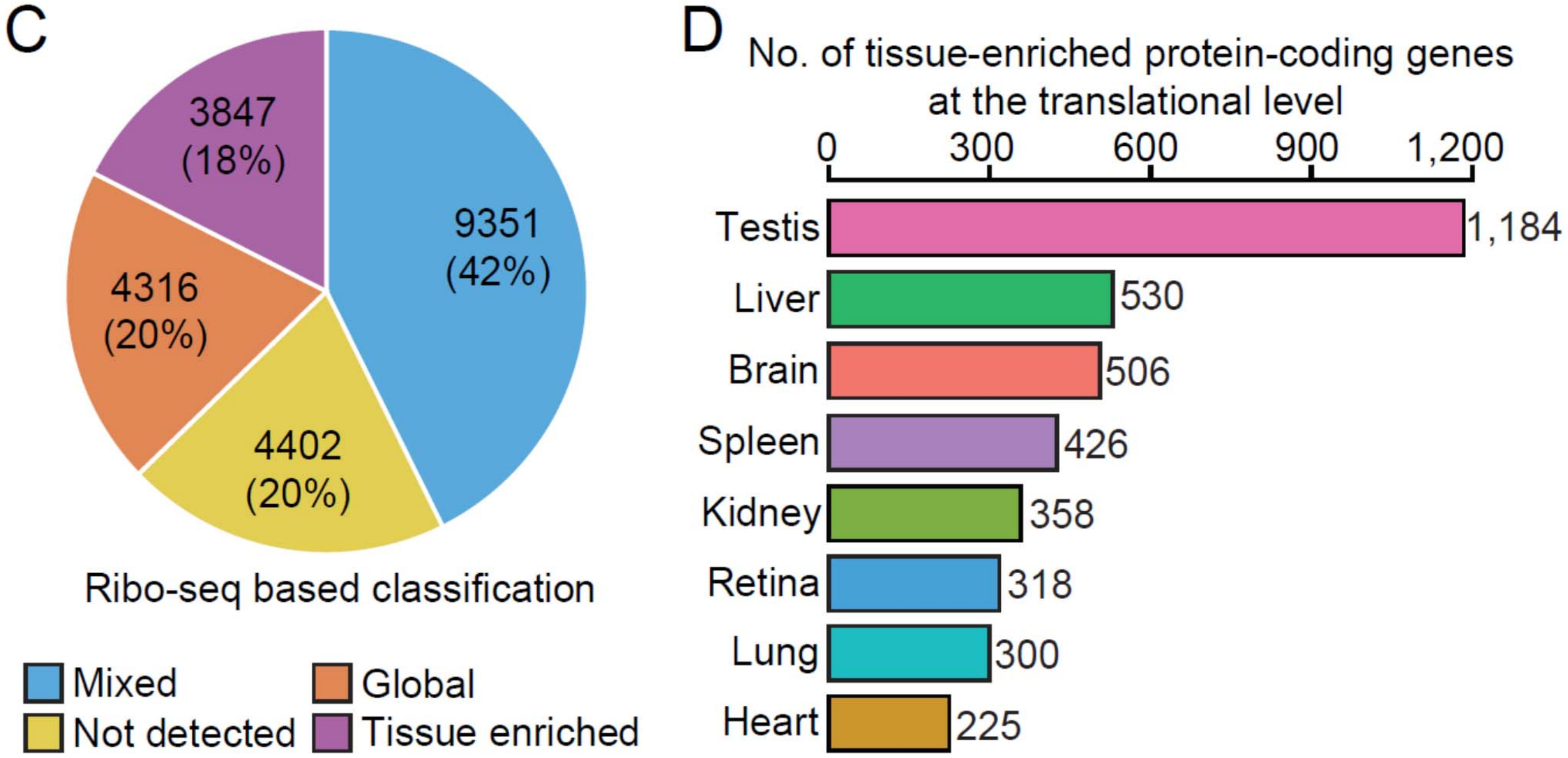
No. of well-translated genes



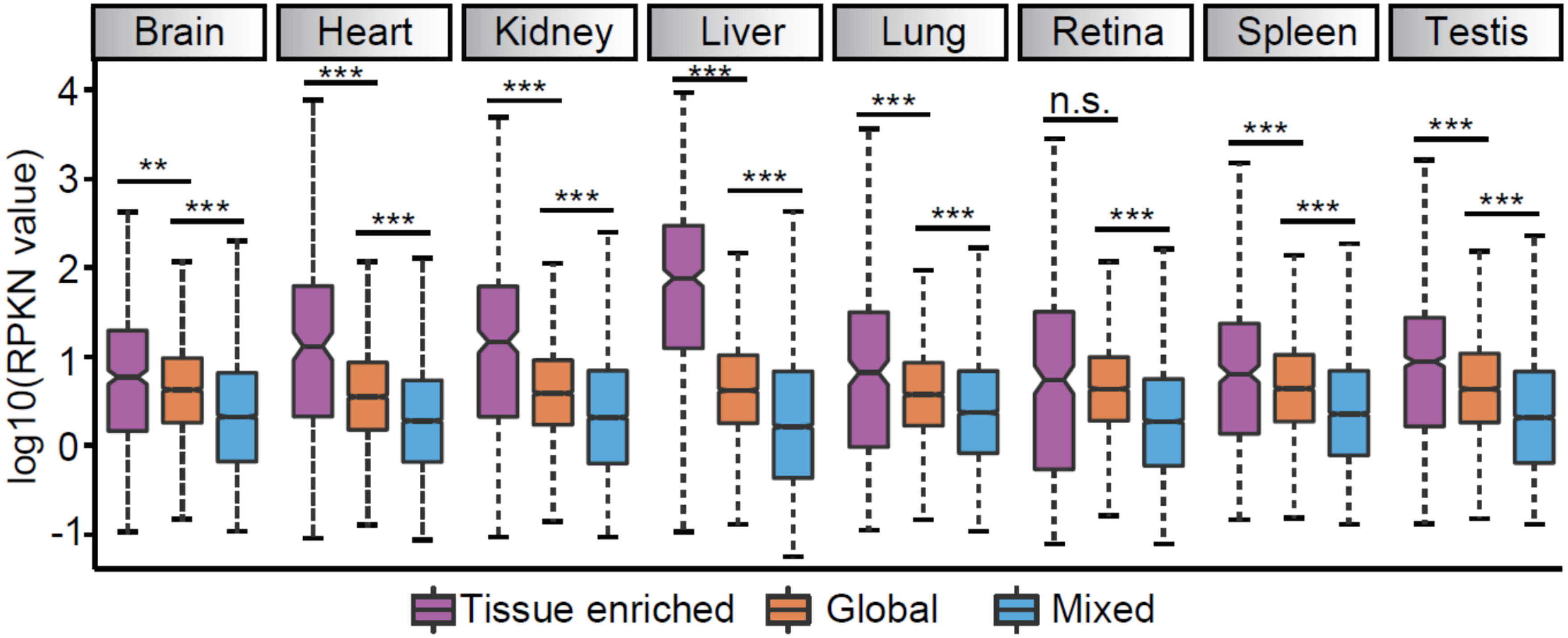
Tissue-specificity



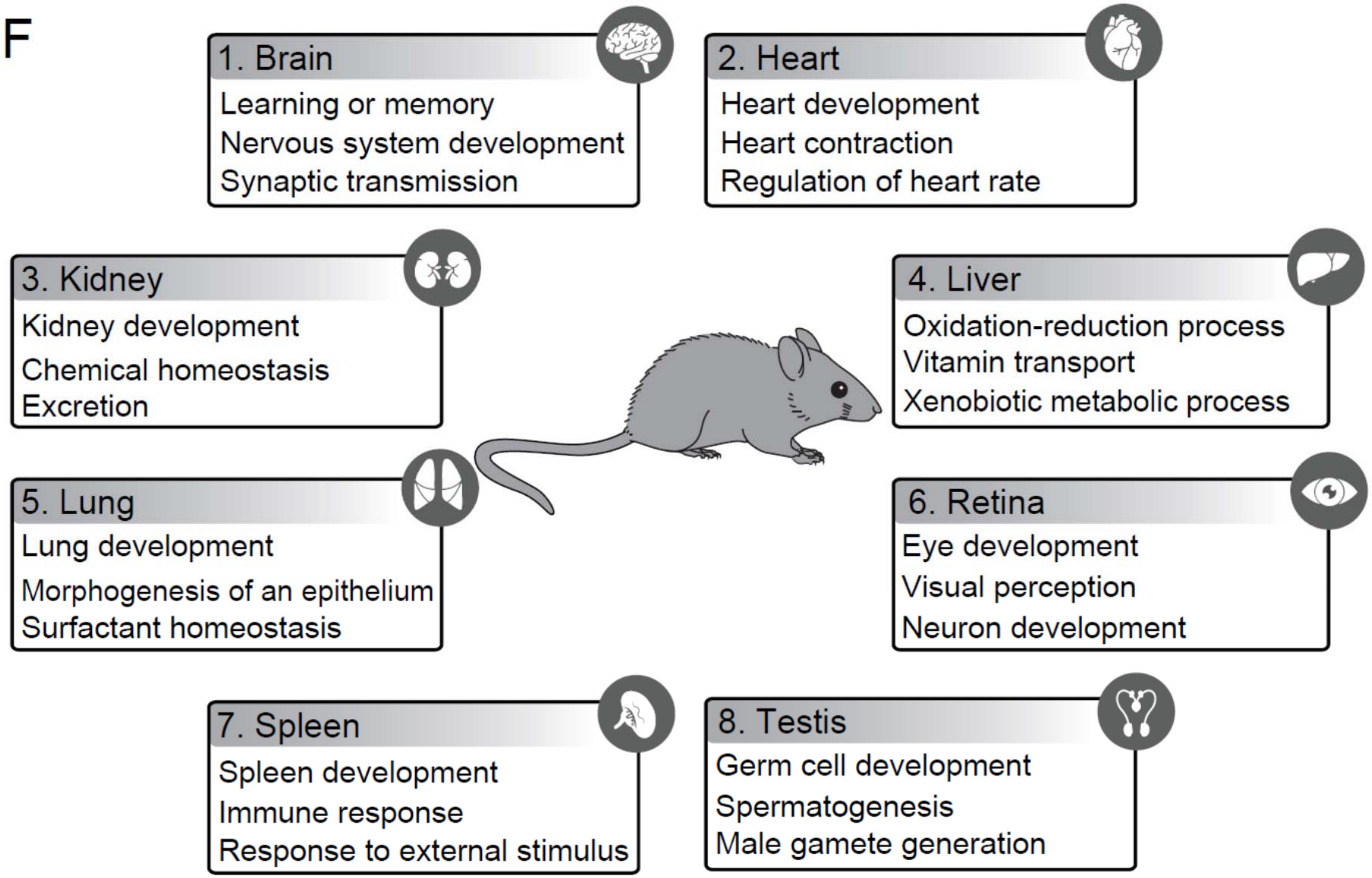
Classification of protein-coding genes based on translation profiles in eight tissues



Translational levels of protein-coding genes in the different categories



Functions of tissue-enriched genes



F

1. Brain
- Learning or memory
- Nervous system development
- Synaptic transmission

2. Heart
- Heart development
- Heart contraction
- Regulation of heart rate

3. Kidney
- Kidney development
- Chemical homeostasis
- Excretion

4. Liver
- Oxidation-reduction process
- Vitamin transport
- Xenobiotic metabolic process

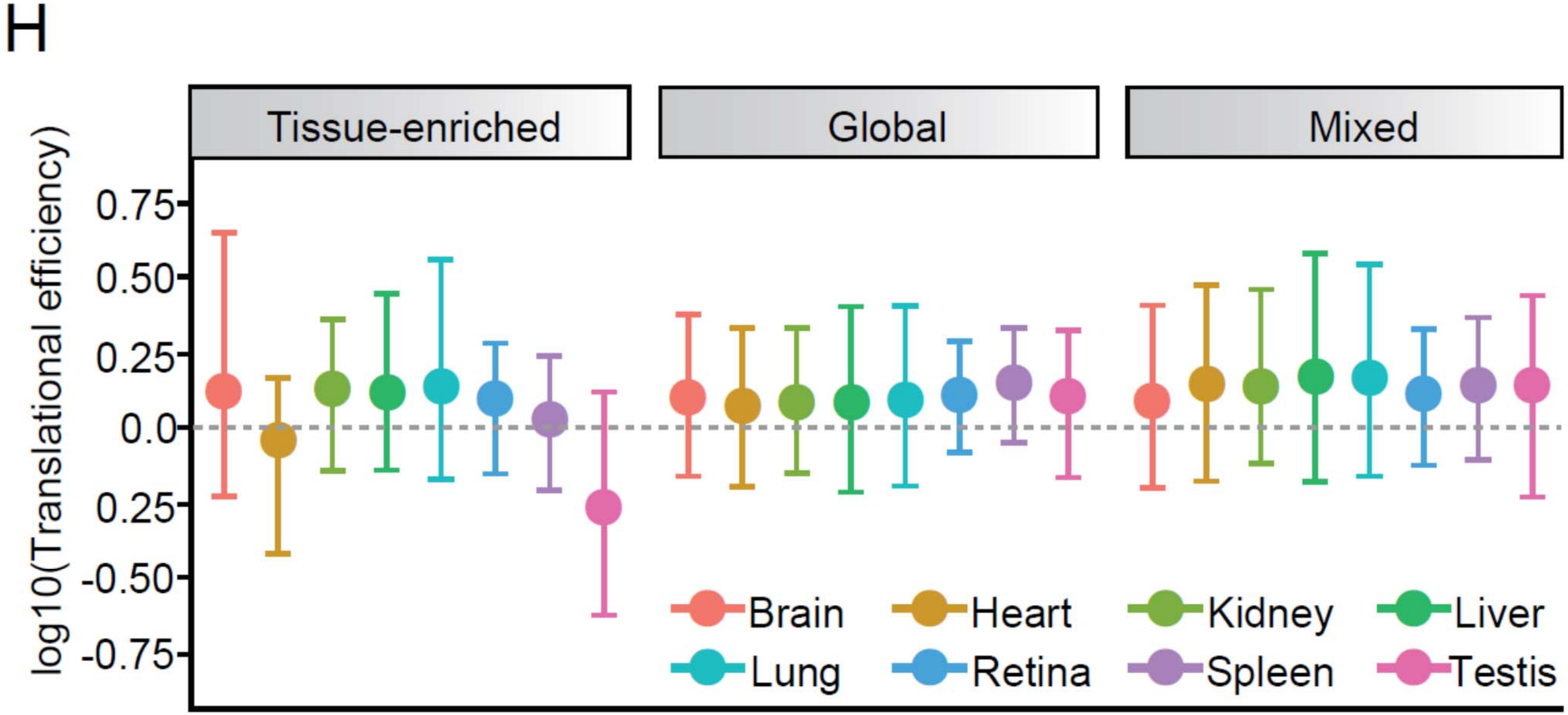
5. Lung
- Lung development
- Morphogenesis of an epithelium
- Surfactant homeostasis

6. Retina
- Eye development
- Visual perception
- Neuron development

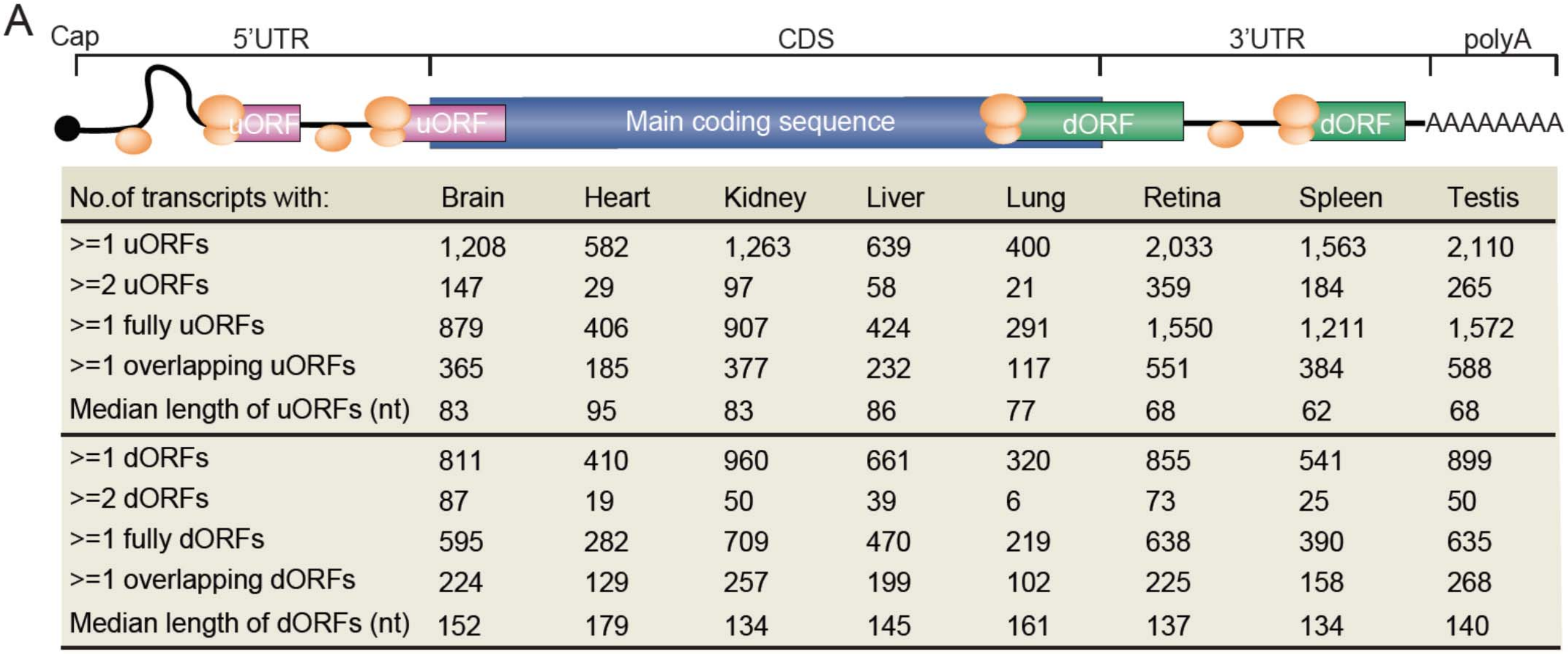
7. Spleen
- Spleen development
- Immune response
- Response to external stimulus

8. Testis
- Germ cell development
- Spermatogenesis
- Male gamete generation

TEs of protein-coding genes in different categories

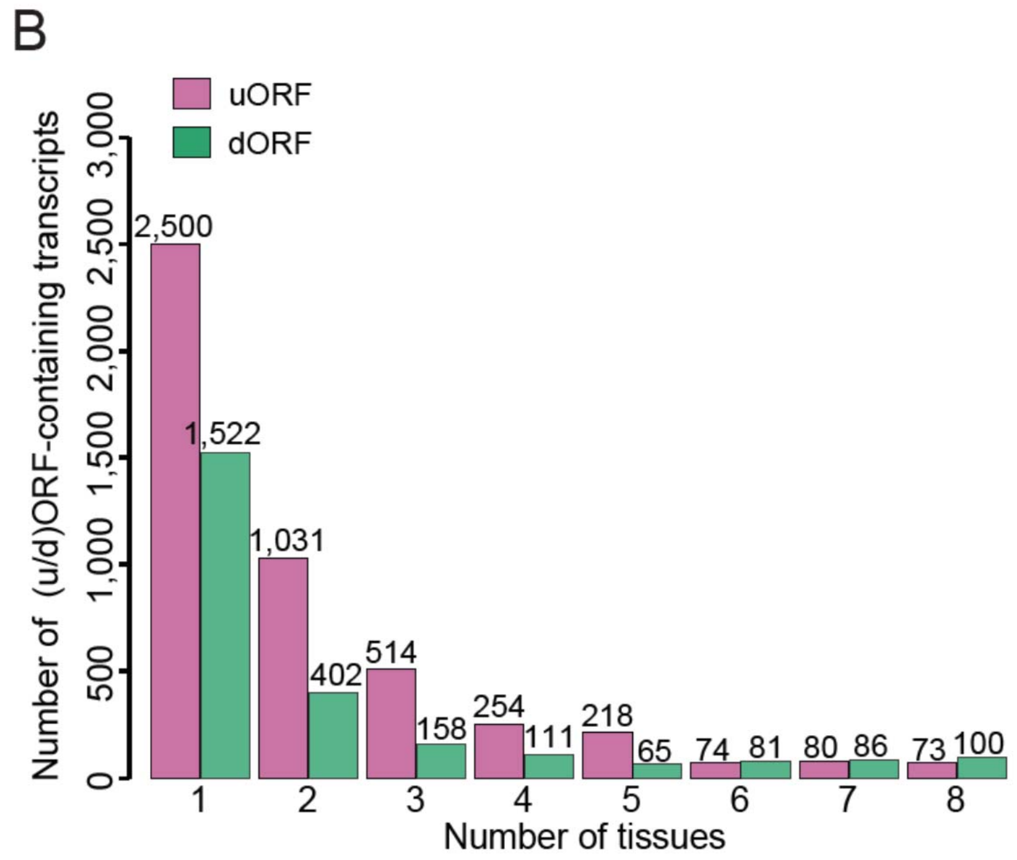


Numbers of uORFs and dORFs in different tissues



No.of transcripts with:	Brain	Heart	Kidney	Liver	Lung	Retina	Spleen	Testis
>=1 uORFs	1,208	582	1,263	639	400	2,033	1,563	2,110
>=2 uORFs	147	29	97	58	21	359	184	265
>=1 fully uORFs	879	406	907	424	291	1,550	1,211	1,572
>=1 overlapping uORFs	365	185	377	232	117	551	384	588
Median length of uORFs (nt)	83	95	83	86	77	68	62	68
>=1 dORFs	811	410	960	661	320	855	541	899
>=2 dORFs	87	19	50	39	6	73	25	50
>=1 fully dORFs	595	282	709	470	219	638	390	635
>=1 overlapping dORFs	224	129	257	199	102	225	158	268
Median length of dORFs (nt)	152	179	134	145	161	137	134	140

Numbers of (u/d)ORF-containing transcripts identified from different numbers of tissues

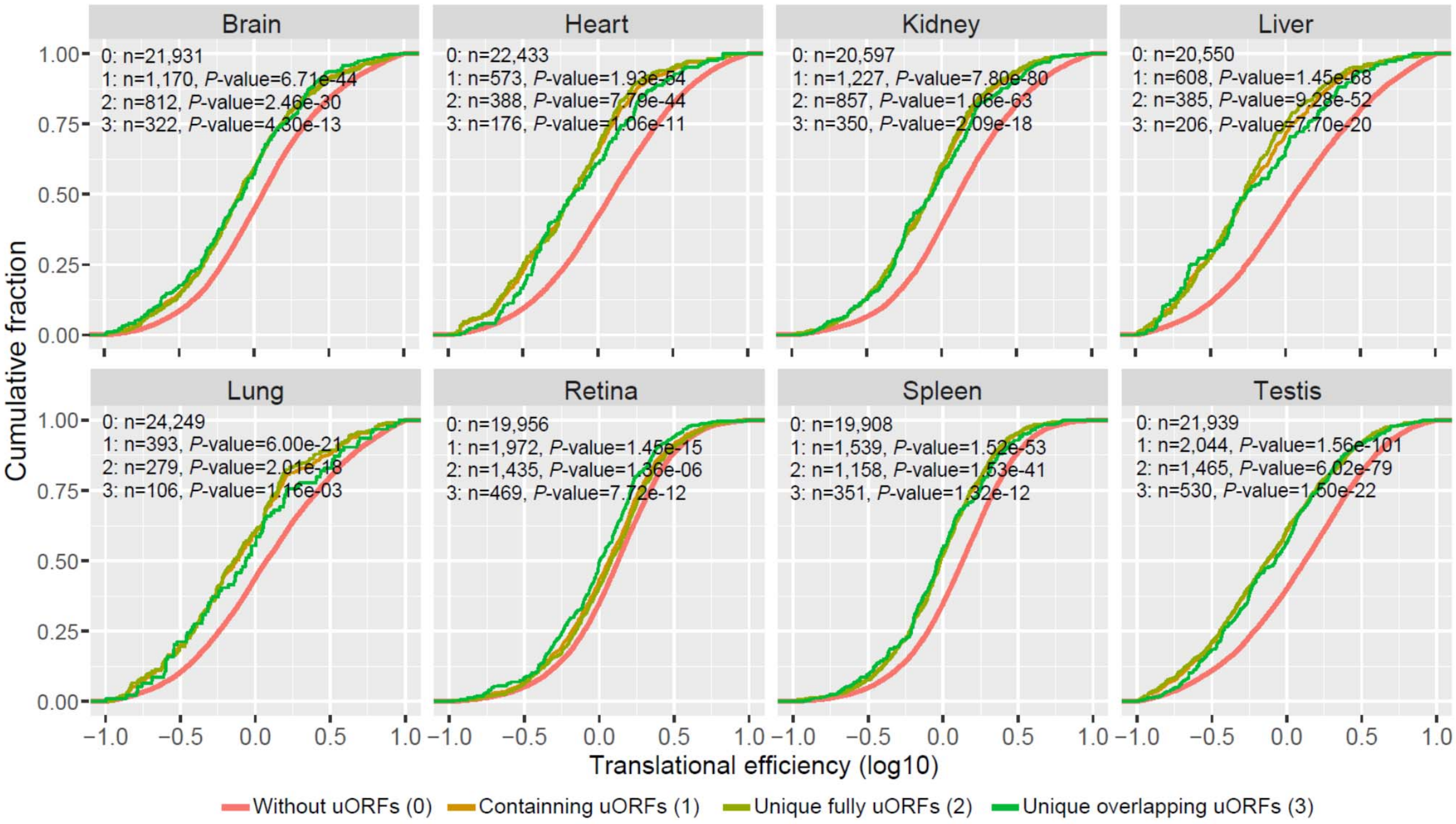


Functional analysis of uORF

- **Most involved in gene regulation, protein modification, and signal transduction**

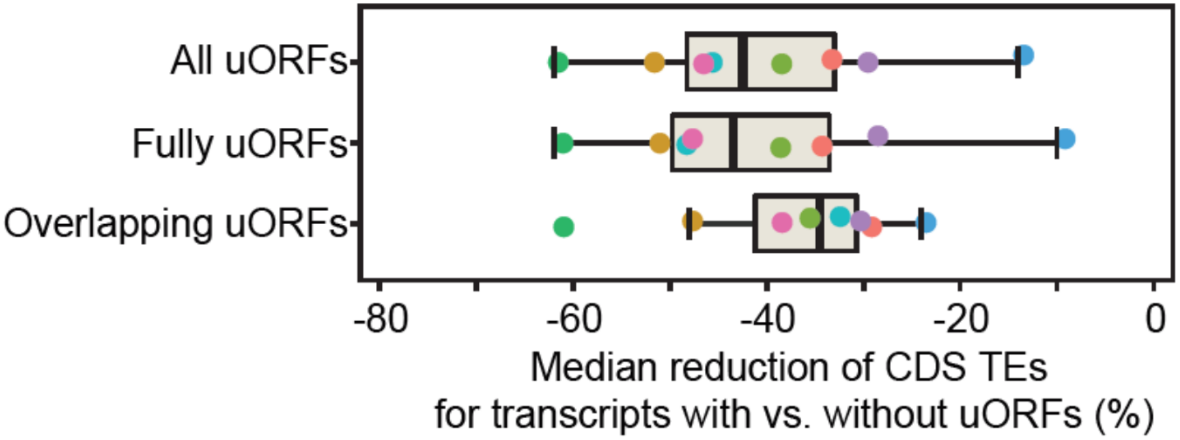
- **Some uORF-containing transcripts had a direct link with physiological functions of each tissue**
  - **Retina: photoreceptor cell maintenance and visual perception**
  - **Testis: spermatogenesis**

Cumulative distribution of TEs of transcripts grouped by the types of uORFs



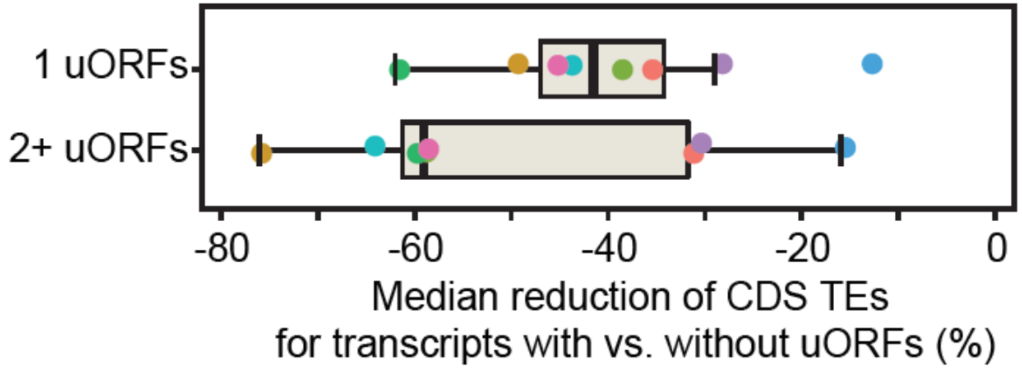
The presence of uORFs

E

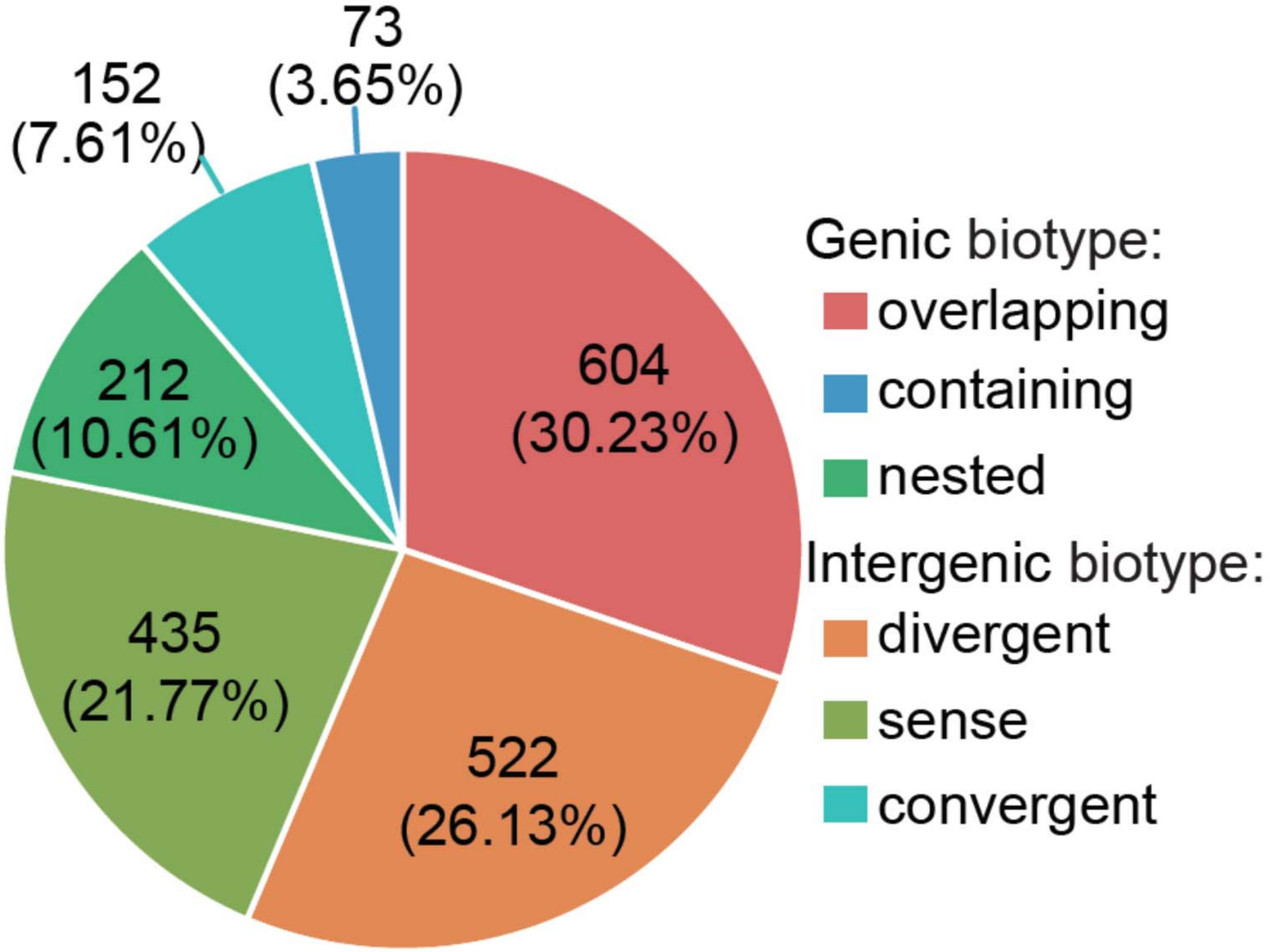


The number of uORFs

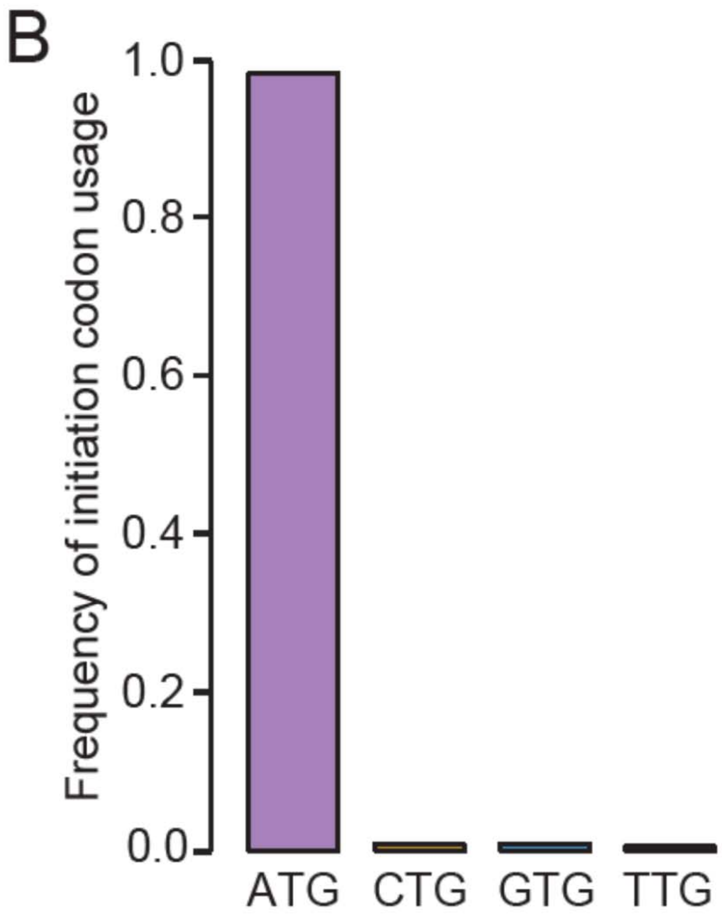
F



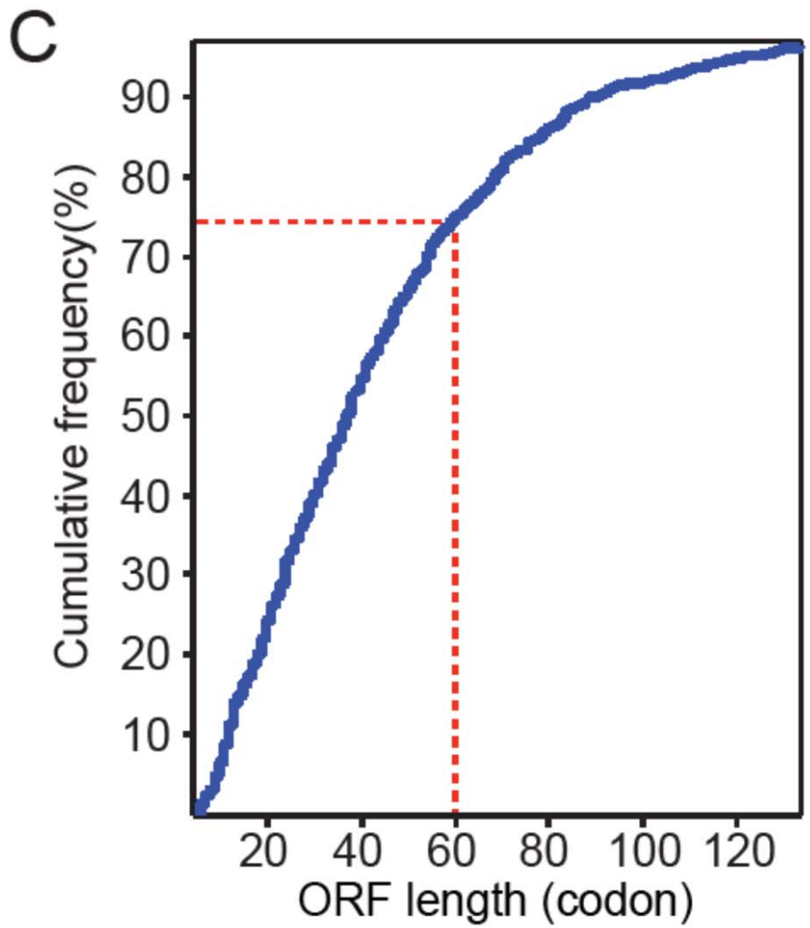
Identification of sORFs



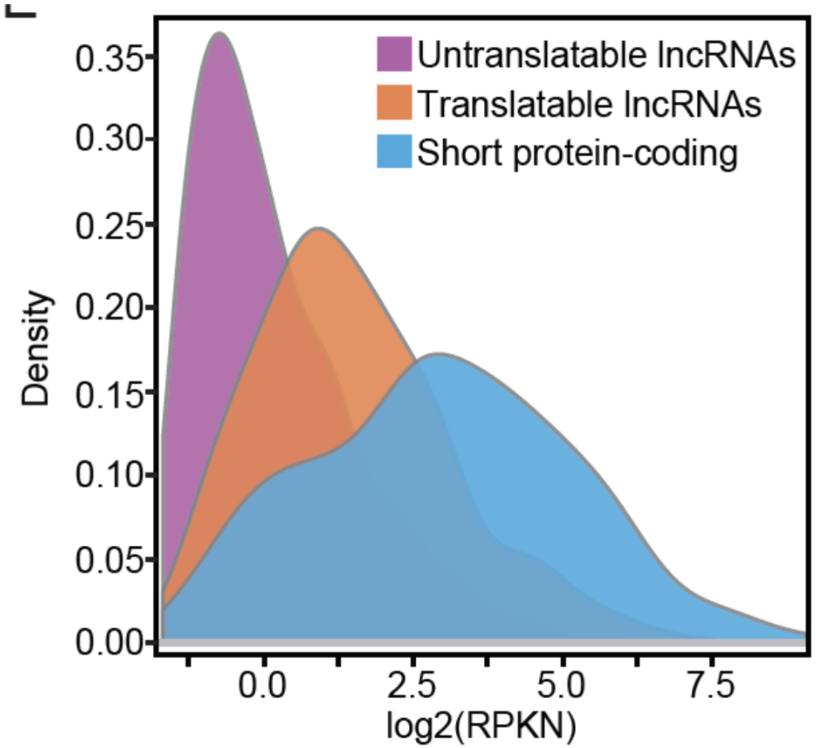
Translation initiation codon usage

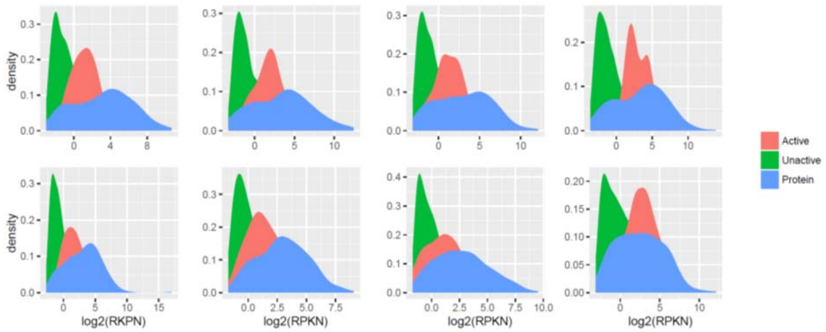


Cumulative distribution of actively translated sORF lengths



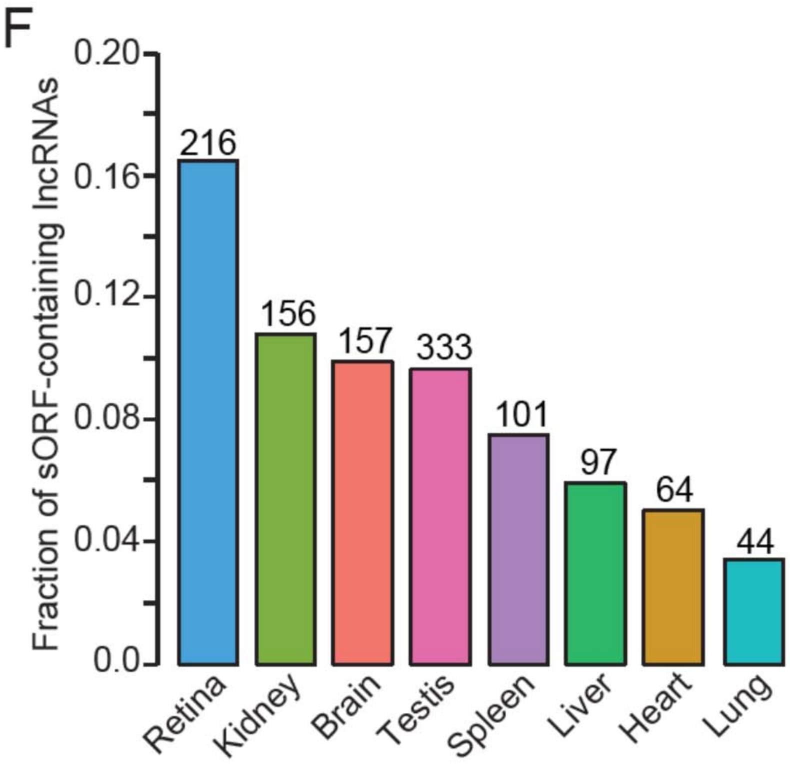
Comparison of transcriptional abundance for translatable and untranslatable lncRNAs

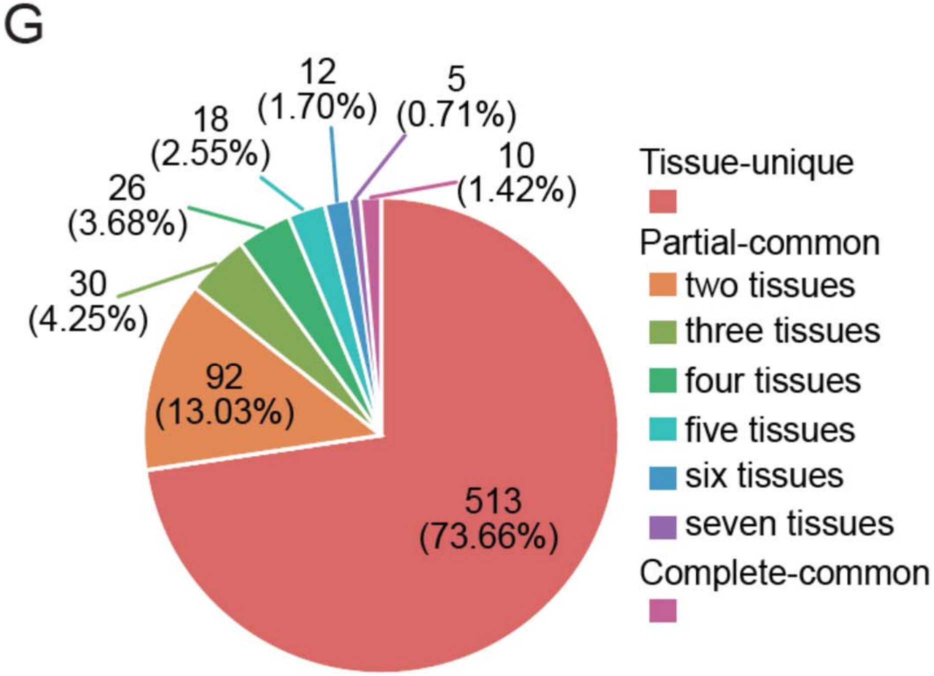




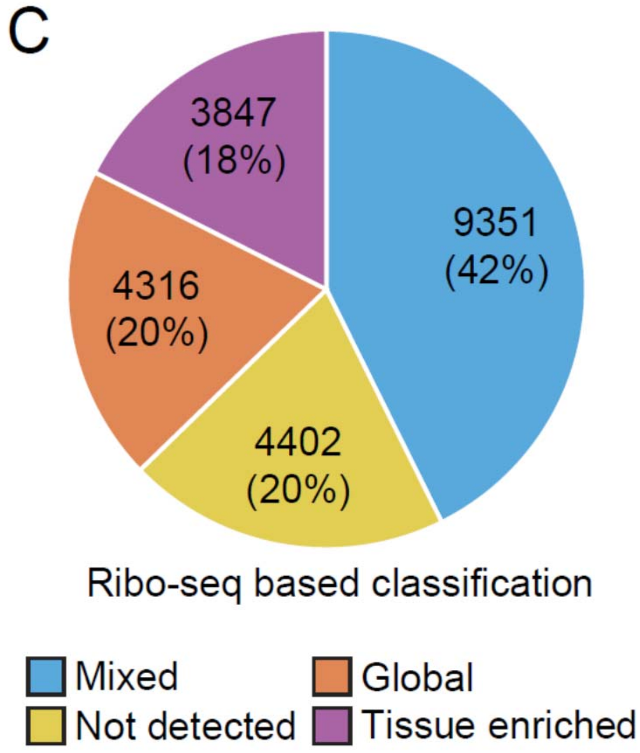
Percentage of sORF-containing lncRNAs in each tissue

Numbers of translatable lncRNAs in different tissues

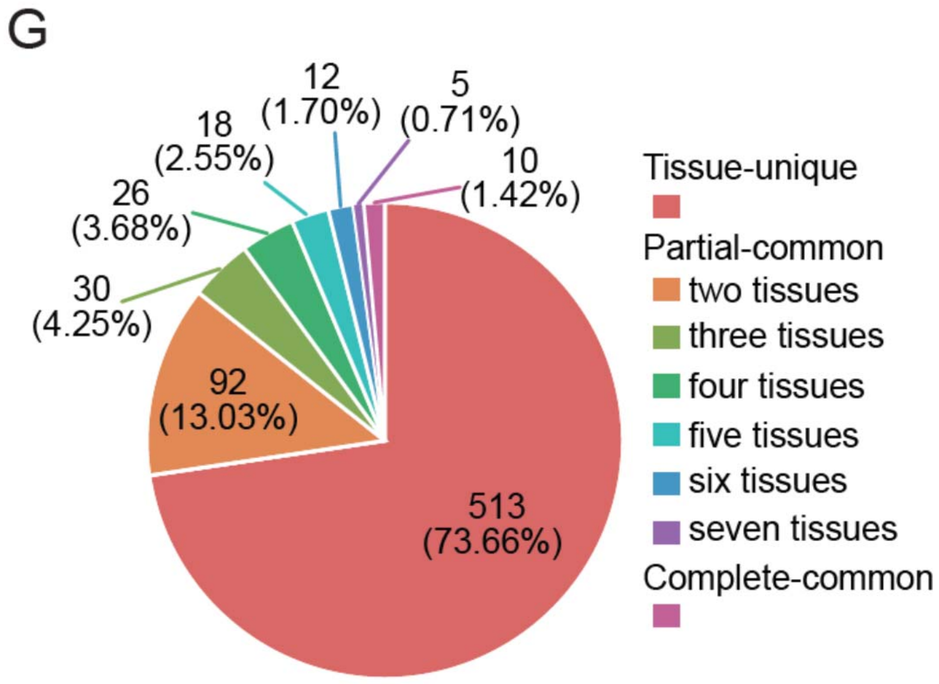




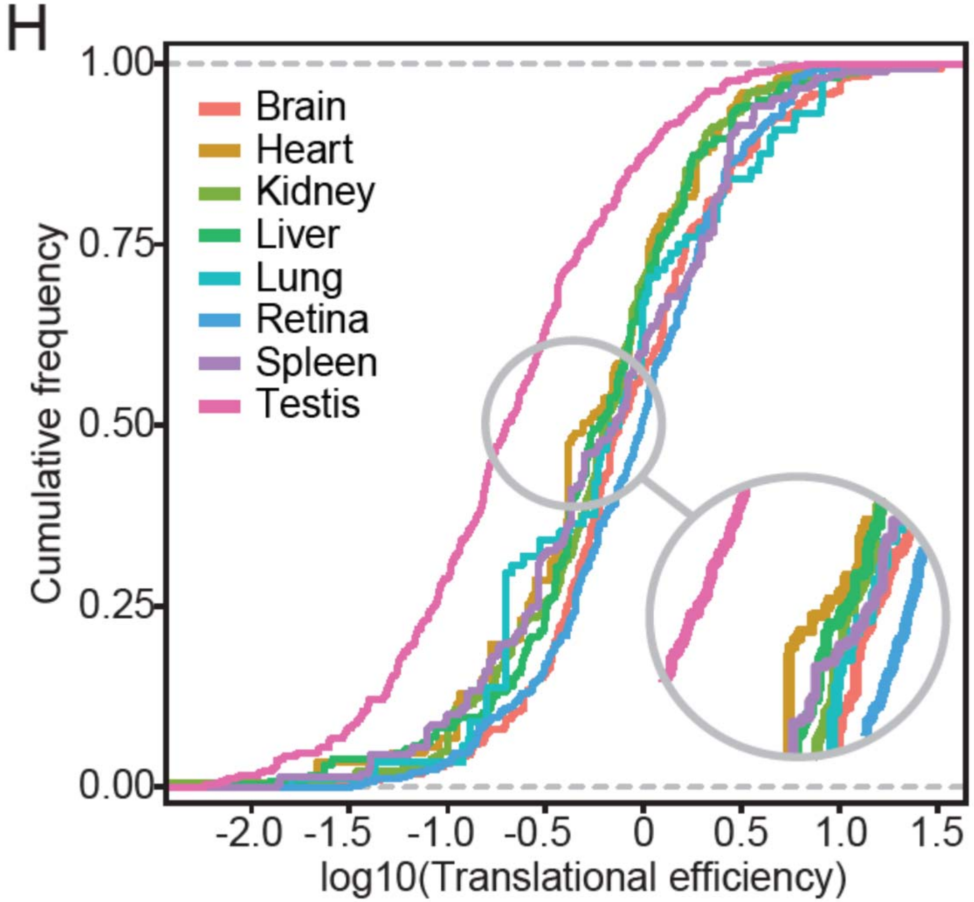
Protein-coding genes



sORFs from lncRNAs

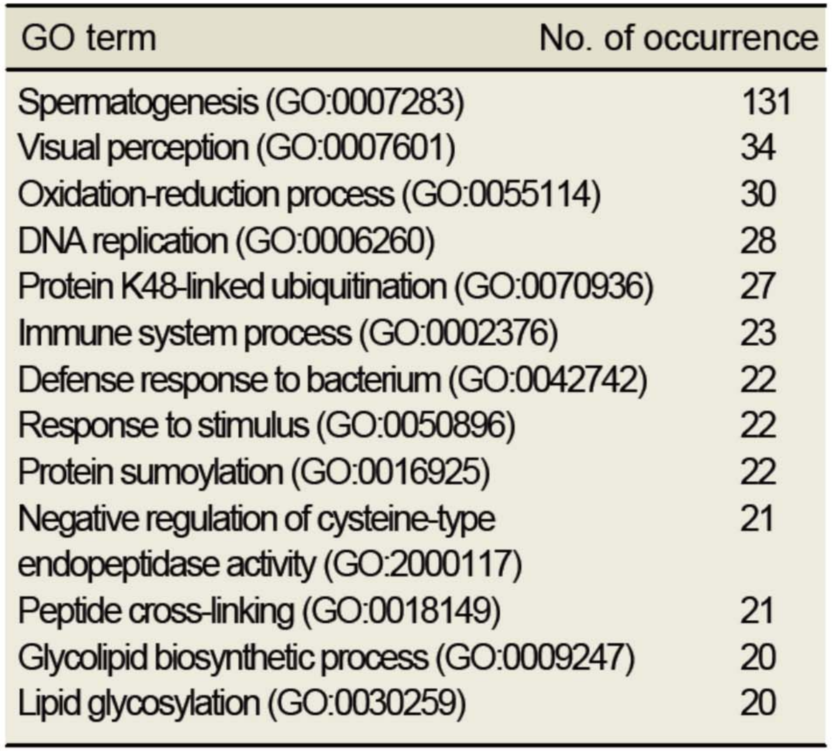


TEs of translatable lncRNAs in eight tissues



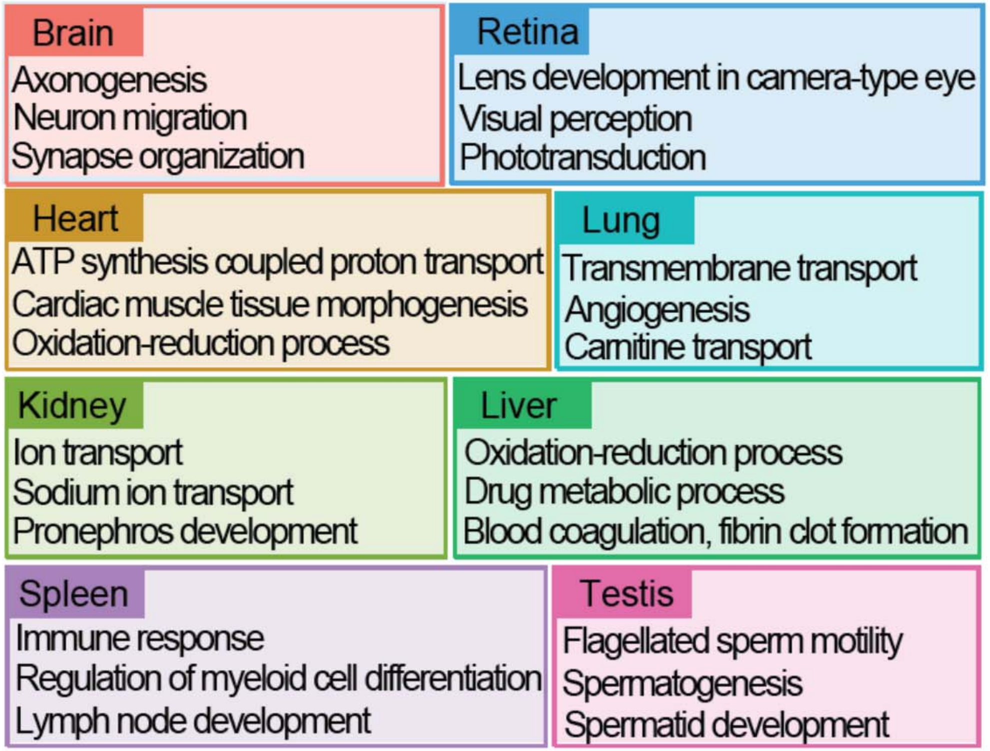
Major GO terms associated with translatable lncRNAs

I



GO term	No. of occurrence
Spermatogenesis (GO:0007283)	131
Visual perception (GO:0007601)	34
Oxidation-reduction process (GO:0055114)	30
DNA replication (GO:0006260)	28
Protein K48-linked ubiquitination (GO:0070936)	27
Immune system process (GO:0002376)	23
Defense response to bacterium (GO:0042742)	22
Response to stimulus (GO:0050896)	22
Protein sumoylation (GO:0016925)	22
Negative regulation of cysteine-type endopeptidase activity (GO:2000117)	21
Peptide cross-linking (GO:0018149)	21
Glycolipid biosynthetic process (GO:0009247)	20
Lipid glycosylation (GO:0030259)	20

J



Brain
Axonogenesis
Neuron migration
Synapse organization

Retina
Lens development in camera-type eye
Visual perception
Phototransduction

Heart
ATP synthesis coupled proton transport
Cardiac muscle tissue morphogenesis
Oxidation-reduction process

Lung
Transmembrane transport
Angiogenesis
Carnitine transport

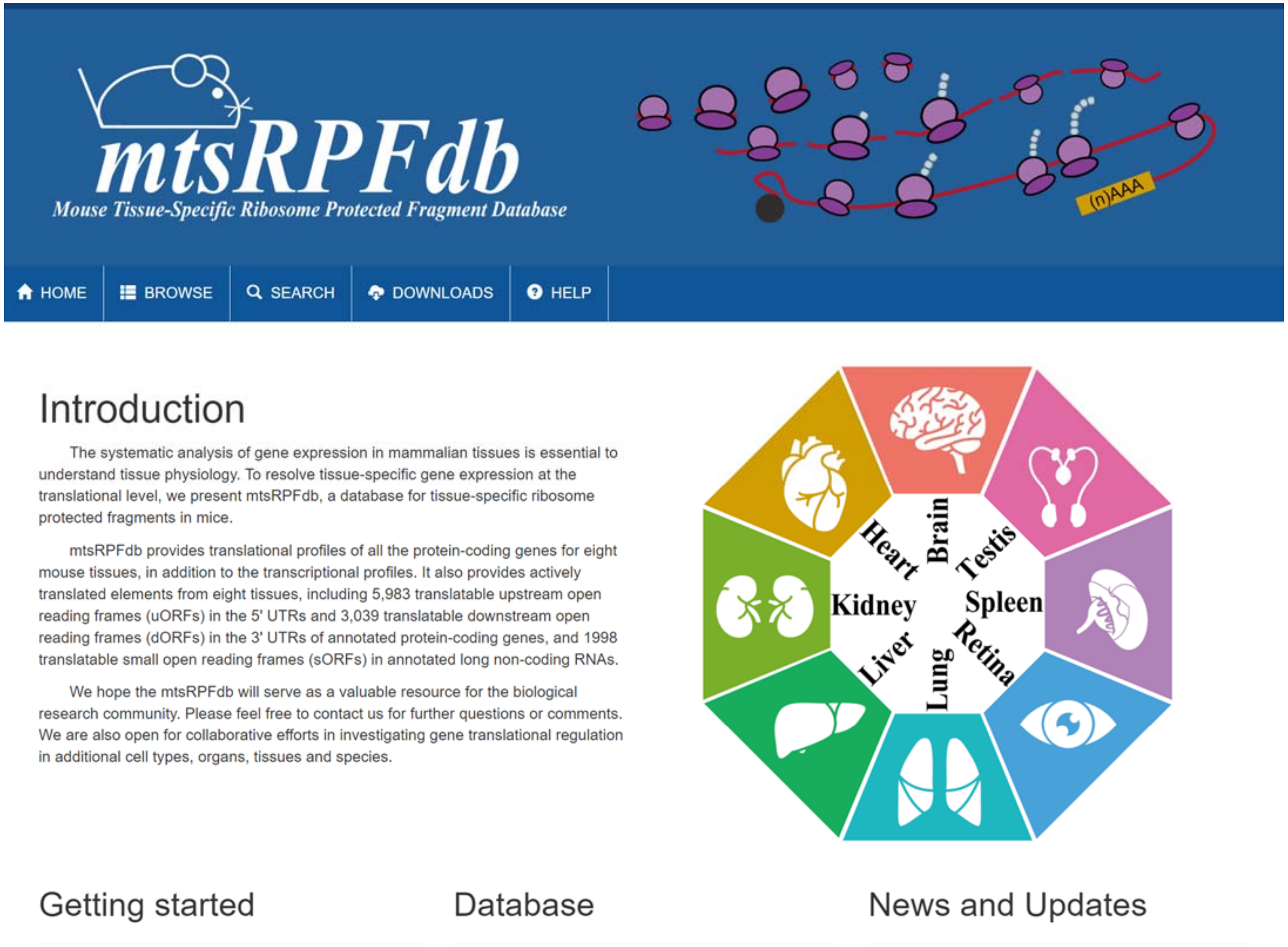
Kidney
Ion transport
Sodium ion transport
Pronephros development

Liver
Oxidation-reduction process
Drug metabolic process
Blood coagulation, fibrin clot formation

Spleen
Immune response
Regulation of myeloid cell differentiation
Lymph node development

Testis
Flagellated sperm motility
Spermatogenesis
Spermatid development

http://sysbio.sysu.edu.cn/mtsRPFdb



Introduction

The systematic analysis of gene expression in mammalian tissues is essential to understand tissue physiology. To resolve tissue-specific gene expression at the translational level, we present mtsRPFdb, a database for tissue-specific ribosome protected fragments in mice.

mtsRPFdb provides translational profiles of all the protein-coding genes for eight mouse tissues, in addition to the transcriptional profiles. It also provides actively translated elements from eight tissues, including 5,983 translatable upstream open reading frames (uORFs) in the 5' UTRs and 3,039 translatable downstream open reading frames (dORFs) in the 3' UTRs of annotated protein-coding genes, and 1998 translatable small open reading frames (sORFs) in annotated long non-coding RNAs.

We hope the mtsRPFdb will serve as a valuable resource for the biological research community. Please feel free to contact us for further questions or comments. We are also open for collaborative efforts in investigating gene translational regulation in additional cell types, organs, tissues and species.

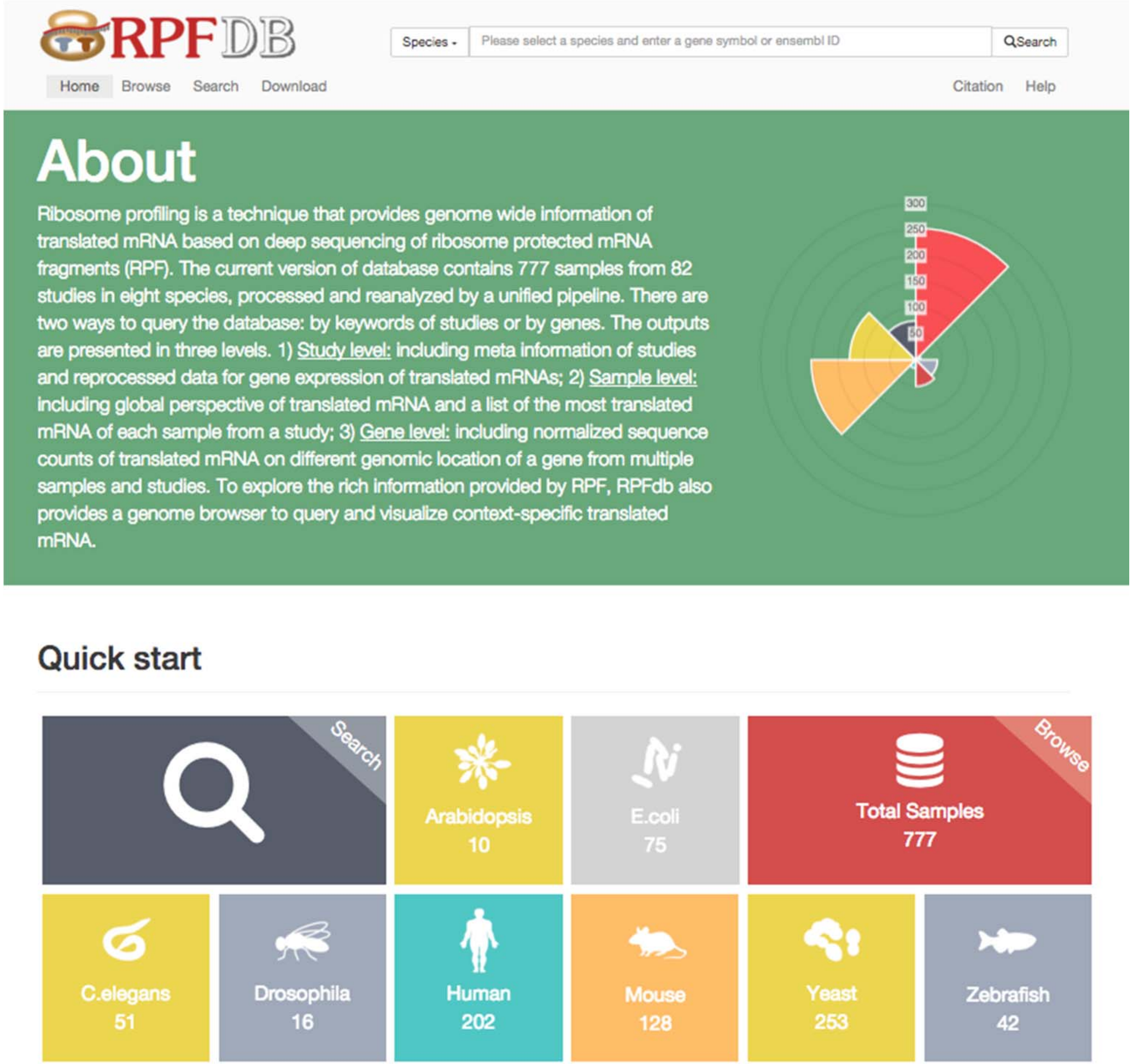
Getting started

Database

News and Updates

Conclusion

- First survey of RNA translation of mammalian tissues
- Translational changes of protein-coding genes
- Pervasive actively translated regions
  - uORFS, dORFS and sORFs from lncRNAs
- Translation events related to physiological functions of tissues



About

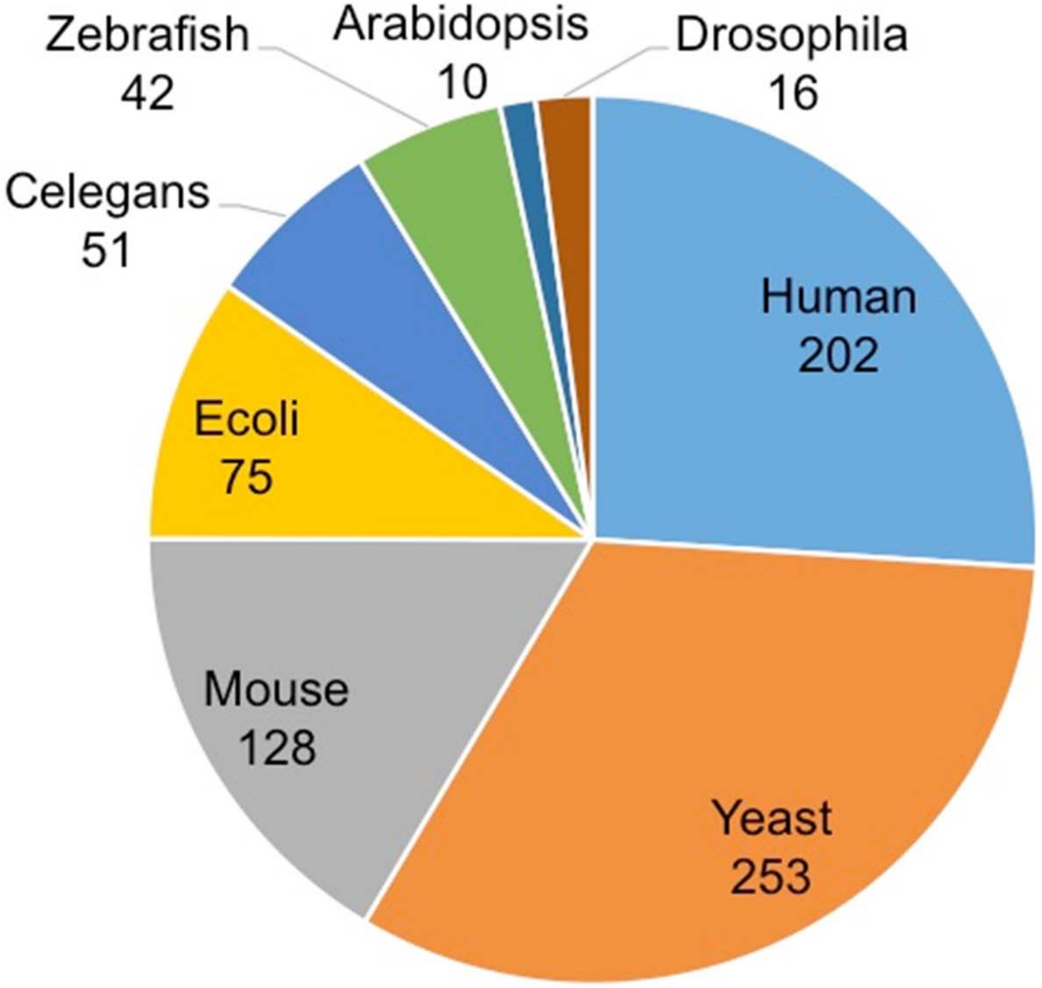
Ribosome profiling is a technique that provides genome wide information of translated mRNA based on deep sequencing of ribosome protected mRNA fragments (RPF). The current version of database contains 777 samples from 82 studies in eight species, processed and reanalyzed by a unified pipeline. There are two ways to query the database: by keywords of studies or by genes. The outputs are presented in three levels. 1) Study level: including meta information of studies and reprocessed data for gene expression of translated mRNAs; 2) Sample level: including global perspective of translated mRNA and a list of the most translated mRNA of each sample from a study; 3) Gene level: including normalized sequence counts of translated mRNA on different genomic location of a gene from multiple samples and studies. To explore the rich information provided by RPF, RPFdb also provides a genome browser to query and visualize context-specific translated mRNA.

Quick start

Arabidopsis	10
E.coli	75
Total Samples	777
C.elegans	51
Drosophila	16
Human	202
Mouse	128
Yeast	253
Zebrafish	42

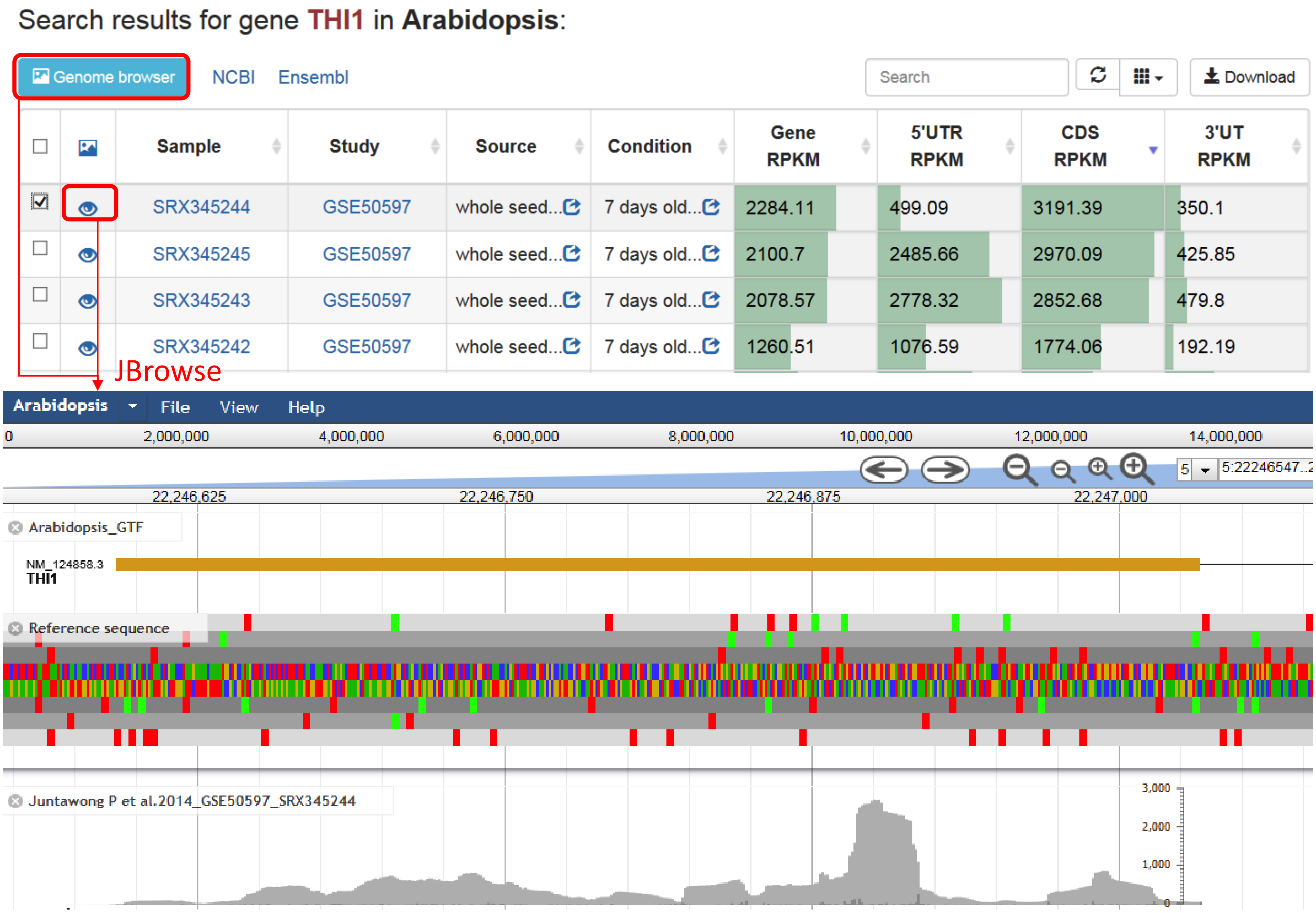
www.rpfdb.org

NAR, 2016



777 datasets from 88 studies

www.rpfdb.org



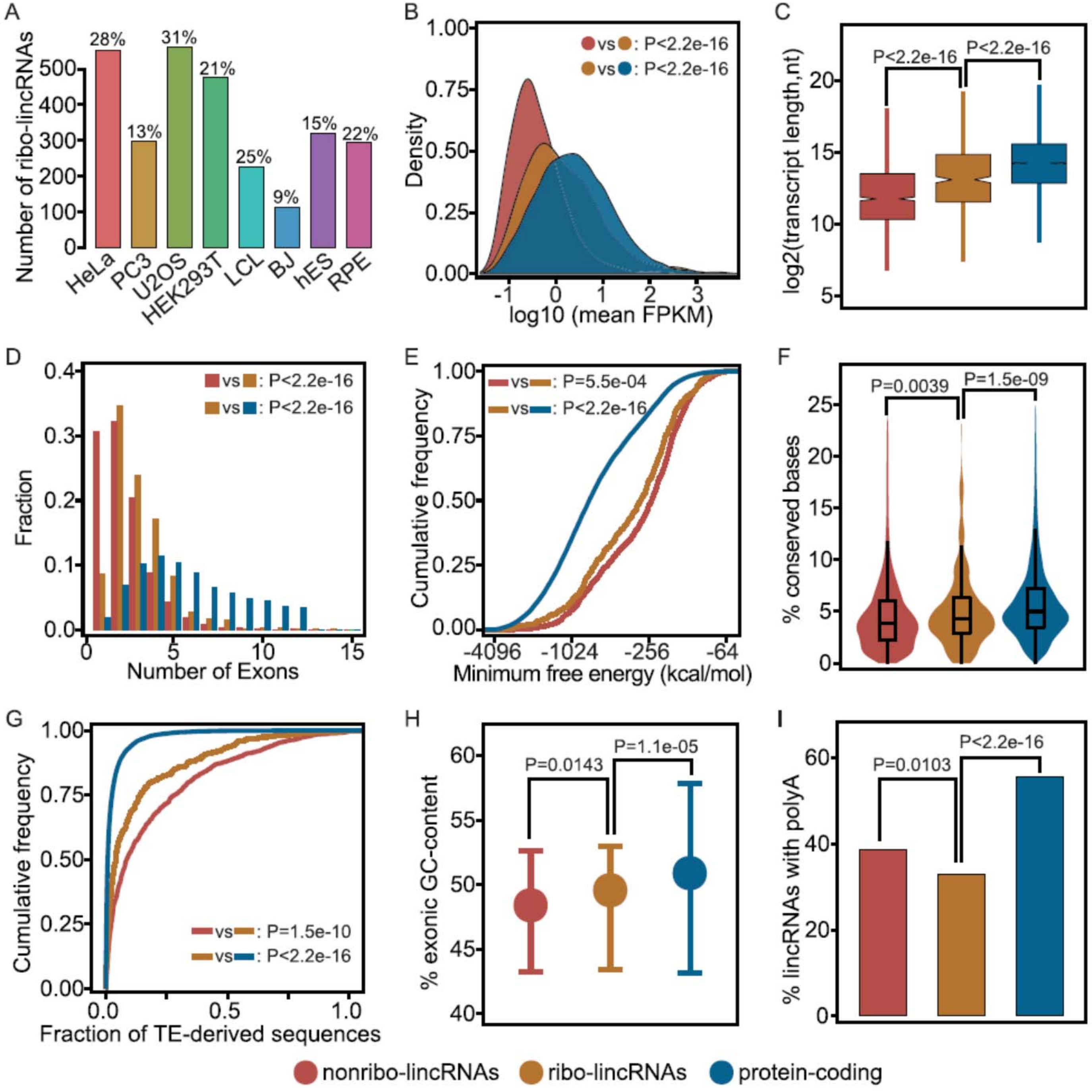
Search results for gene **THI1** in **Arabidopsis**:

☐		Sample	Study	Source	Condition	Gene RPKM	5'UTR RPKM	CDS RPKM	3'UT RPKM
☑		SRX345244	GSE50597	whole seed...	7 days old...	2284.11	499.09	3191.39	350.1
☐		SRX345245	GSE50597	whole seed...	7 days old...	2100.7	2485.66	2970.09	425.85
☐		SRX345243	GSE50597	whole seed...	7 days old...	2078.57	2778.32	2852.68	479.8
☐		SRX345242	GSE50597	whole seed...	7 days old...	1260.51	1076.59	1774.06	192.19

Acknowledgment

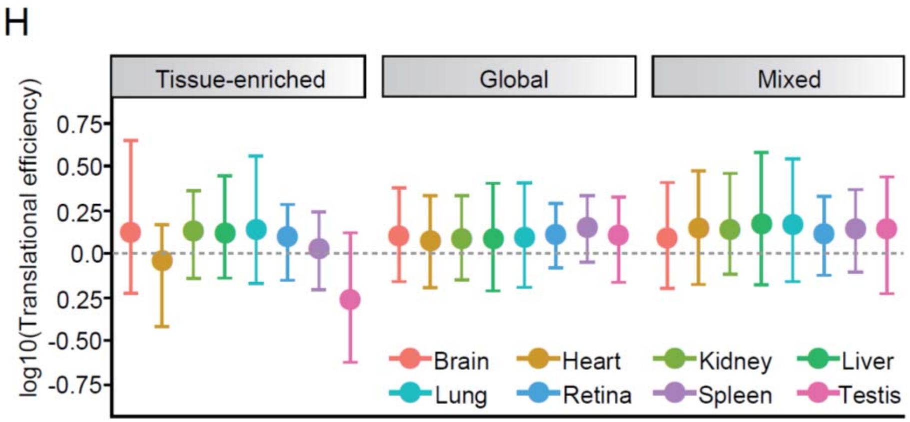
- Data analysis
  - Hongwei Wang
  - Yan Wang
- Experiments
  - Jiaqi Yang
- Website
  - Pohao Ye

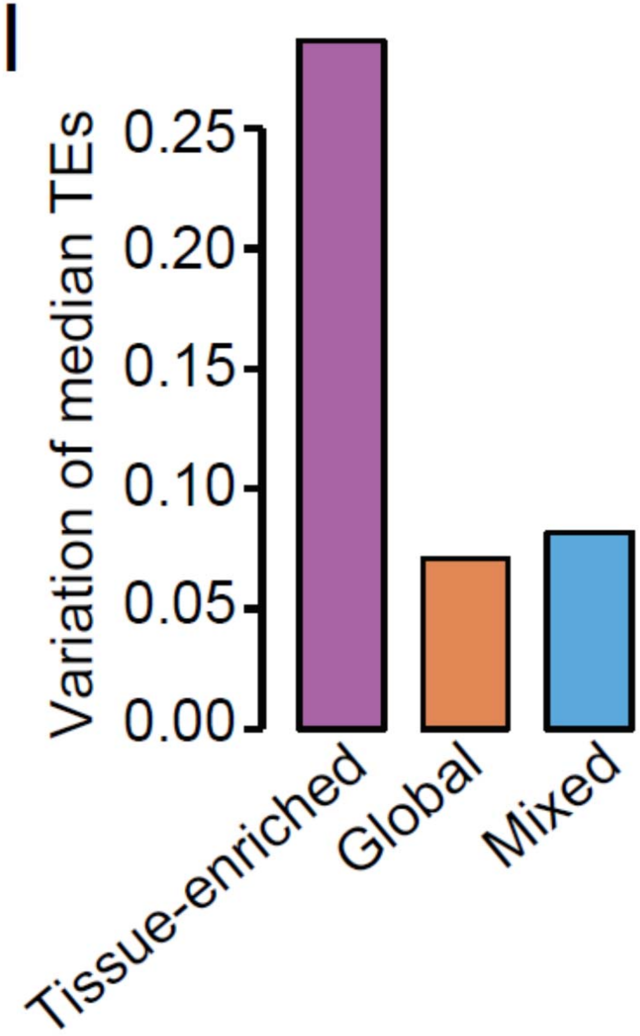
Thanks!


A
Number of ribo-lincRNAs
0
100
200
300
400
500
28%
13%
31%
21%
25%
9%
15%
22%
HeLa
PC3
U2OS
HEK293T
LCL
BJ
hES
RPE
B
Density
0.00
0.25
0.50
0.75
1.00
● vs ●: P<2.2e-16
● vs ●: P<2.2e-16
-1
0
1
2
3
log10 (mean FPKM)
C
log2(transcript length,nt)
5
10
15
20
P<2.2e-16
P<2.2e-16
D
Fraction
0.0
0.1
0.2
0.3
0.4
■ vs ■: P<2.2e-16
■ vs ■: P<2.2e-16
0
5
10
15
Number of Exons
E
Cumulative frequency
0.00
0.25
0.50
0.75
1.00
▬ vs ▬: P=5.5e-04
▬ vs ▬: P<2.2e-16
-4096
-1024
-256
-64
Minimum free energy (kcal/mol)
F
% conserved bases
0
5
10
15
20
25
P=0.0039
P=1.5e-09
G
Cumulative frequency
0.00
0.25
0.50
0.75
1.00
▬ vs ▬: P=1.5e-10
▬ vs ▬: P<2.2e-16
0.0
0.5
1.0
Fraction of TE-derived sequences
H
% exonic GC-content
40
45
50
55
60
P=0.0143
P=1.1e-05
I
% lincRNAs with polyA
0
20
40
60
P=0.0103
P<2.2e-16
● nonribo-lincRNAs
● ribo-lincRNAs
● protein-coding

NAR, 2016

Variations of median TEs for different categories





Ribosome footprints of the eIF4G1 transcript