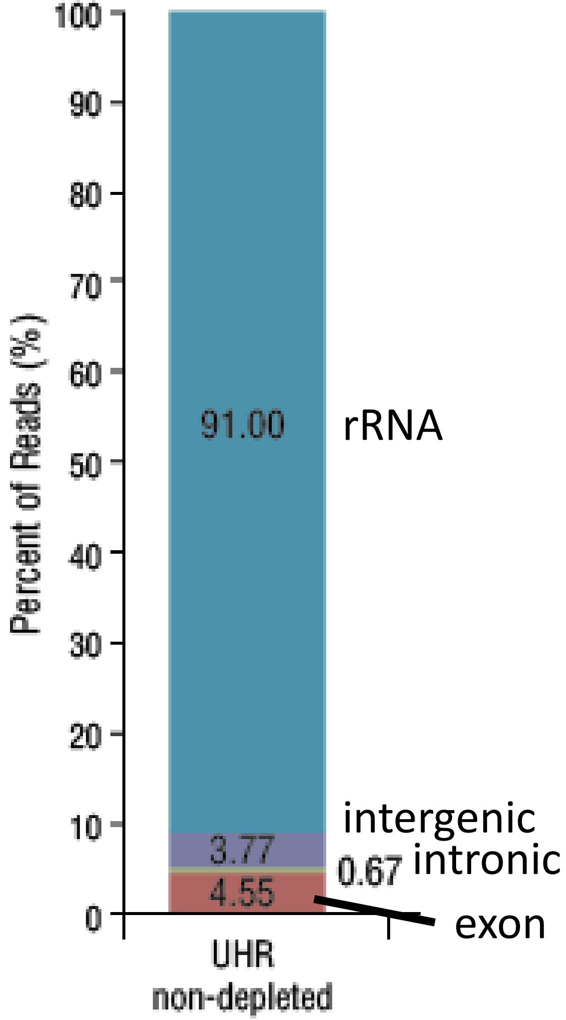
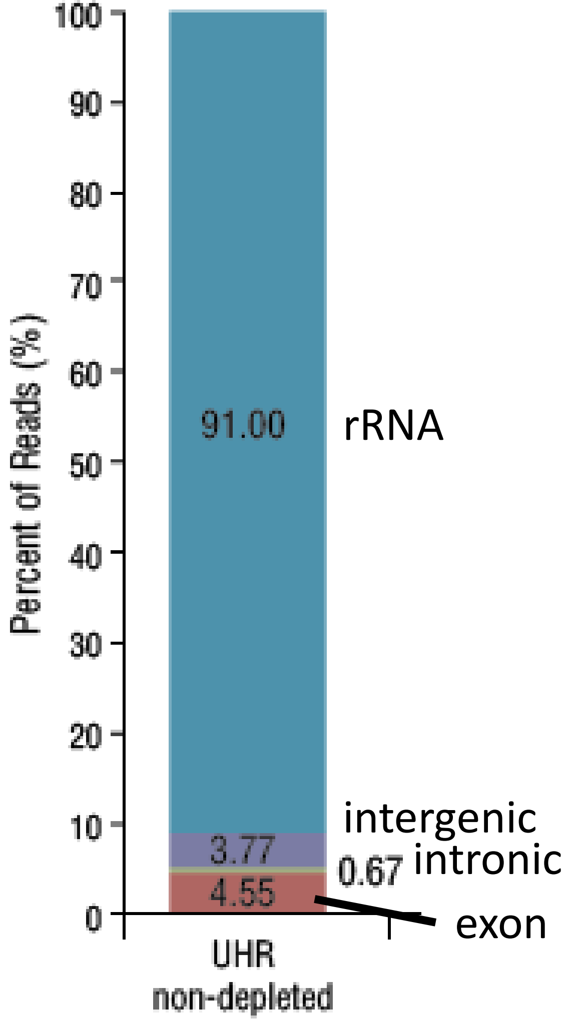
Cross-Site Comparison of Ribosomal Depletion Kits for Illumina RNAseq Library Construction

ABRF
DNA Sequencing Research Group

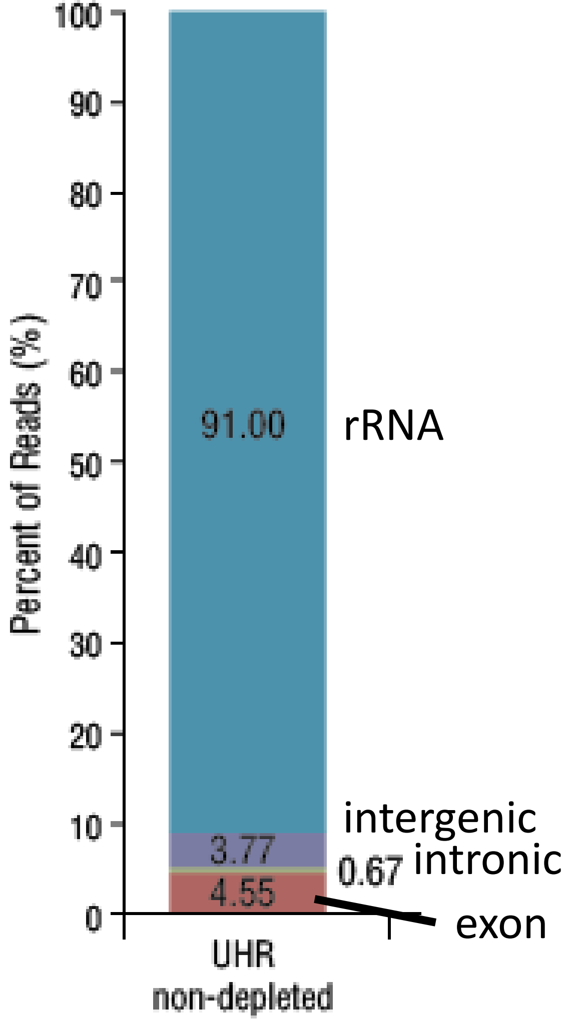
ABRF 2017
ANNUAL MEETING
SAN DIEGO
CA, USA
MARCH 25 – 28, 2017 | TOWN AND COUNTRY RESORT & CONVENTION CENTER
A FORUM FOR ADVANCING TODAY'S CORE TECHNOLOGIES TO ENABLE TOMORROW'S INNOVATIONS



credit: New England Biolabs



credit: New England Biolabs




mRNA

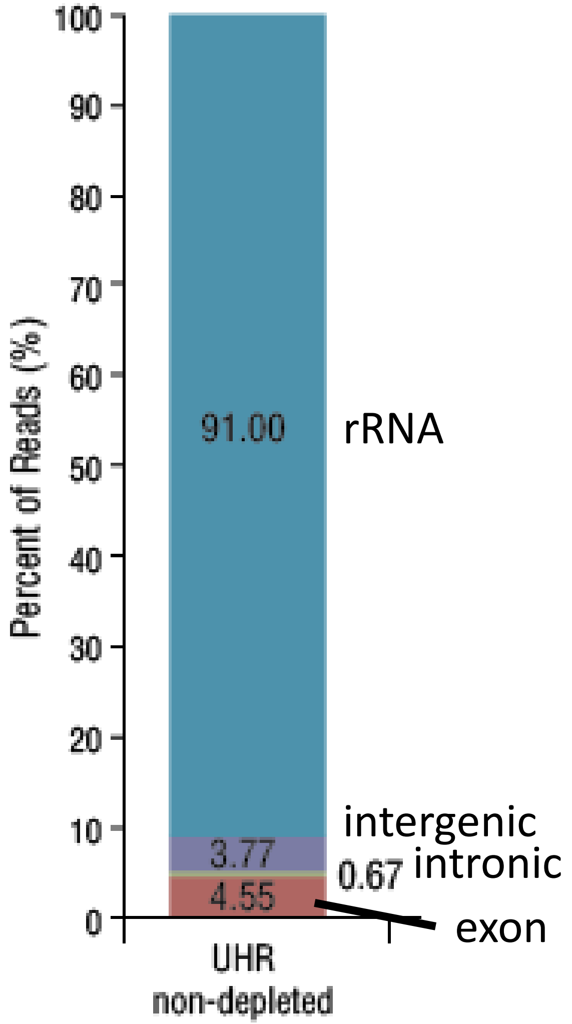
AAAAAA

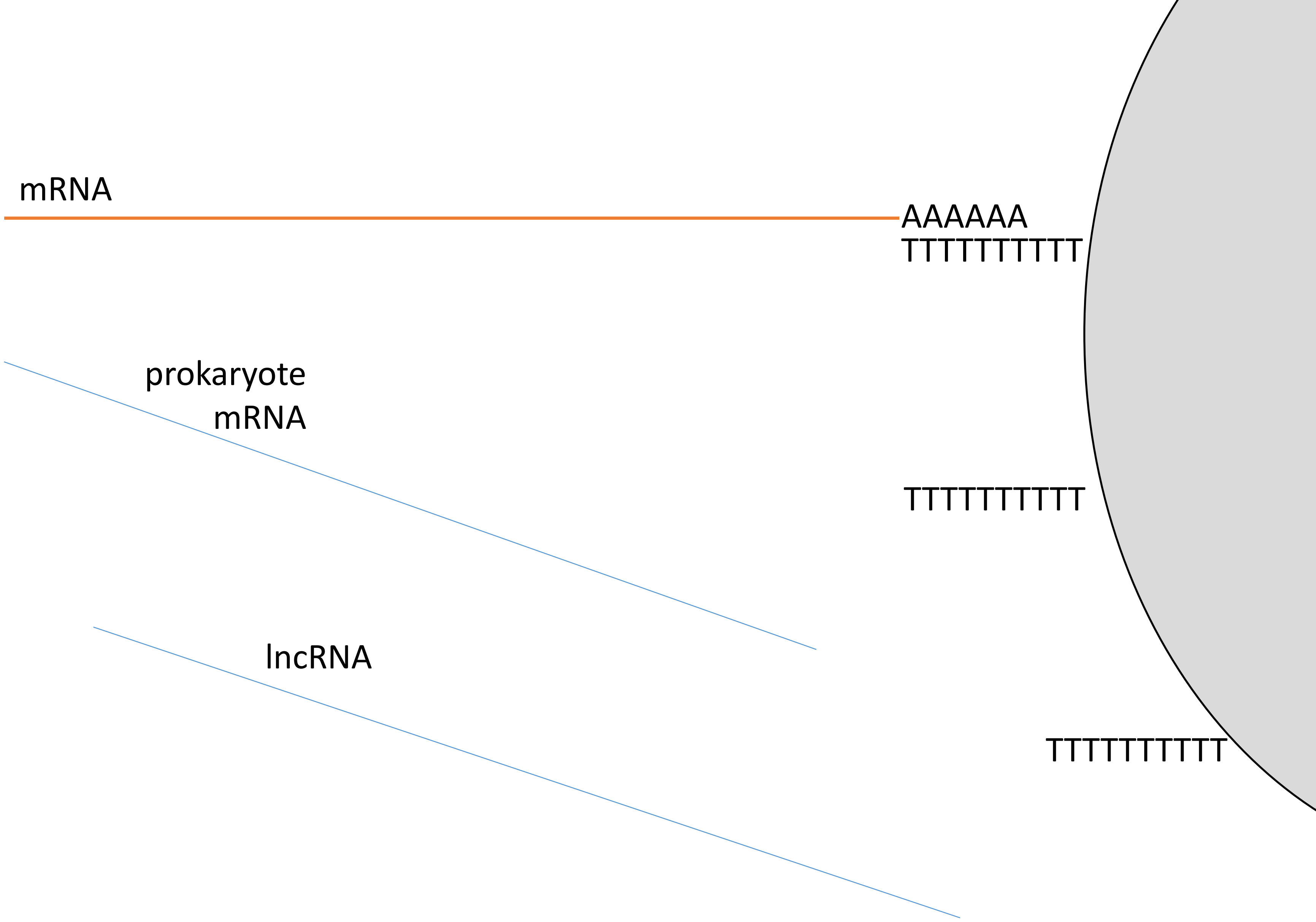
TTTTTTTTTT

TTTTTTTTTT

TTTTTTTTTT

credit: New England Biolabs





mRNA

AAAAAA

TTTTTTTTTT

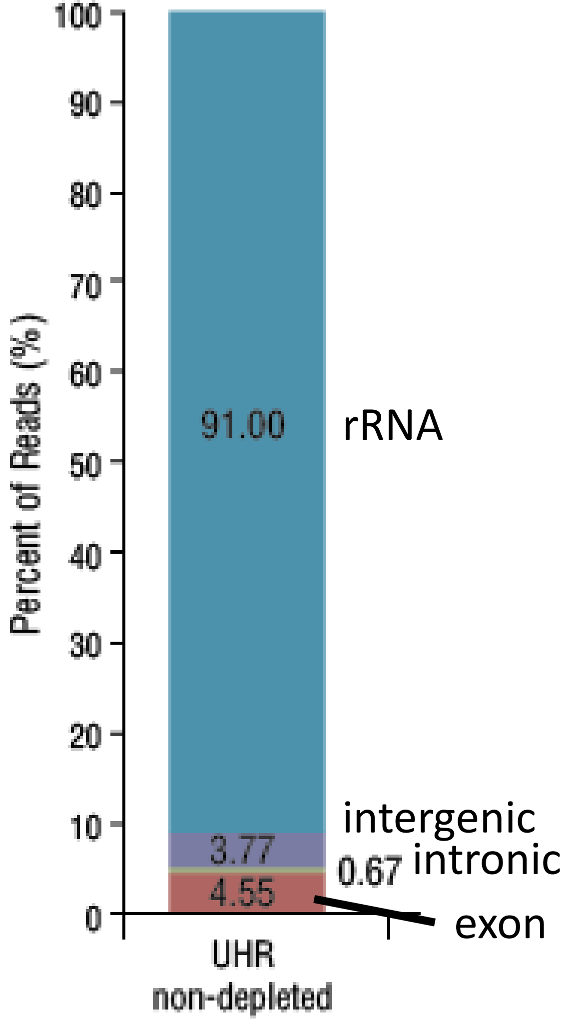
prokaryote
mRNA

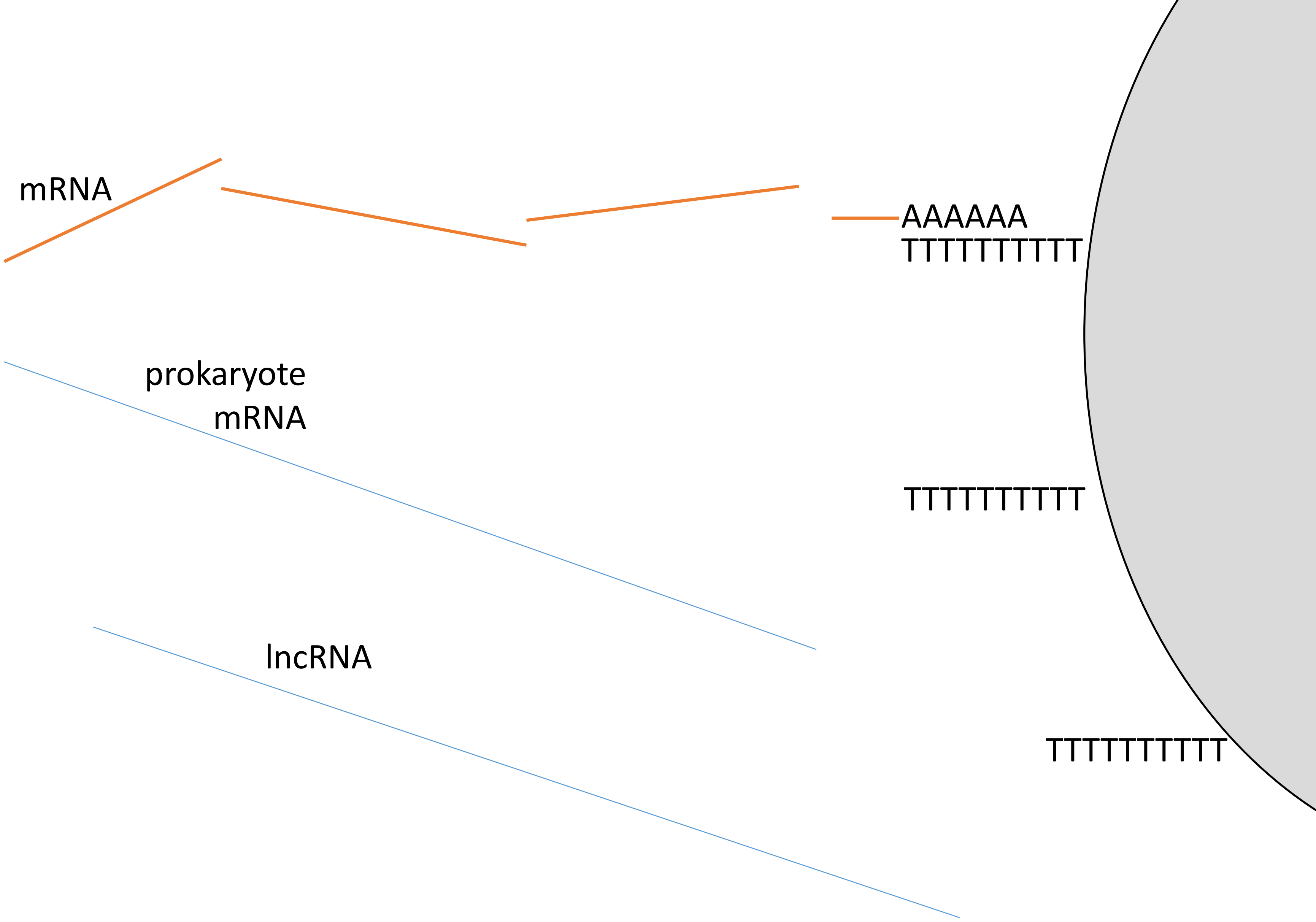
TTTTTTTTTT

lncRNA

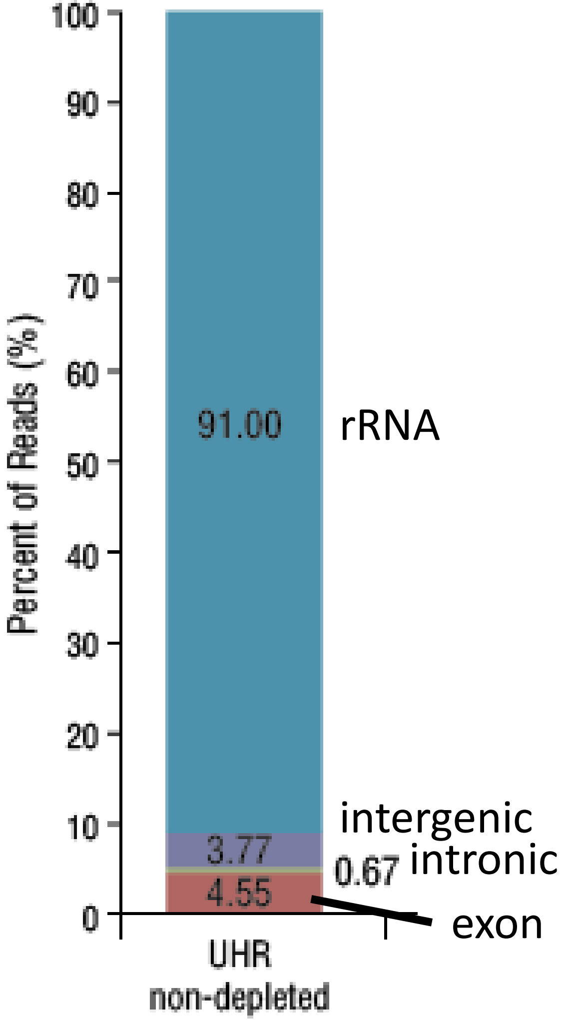
TTTTTTTTTT

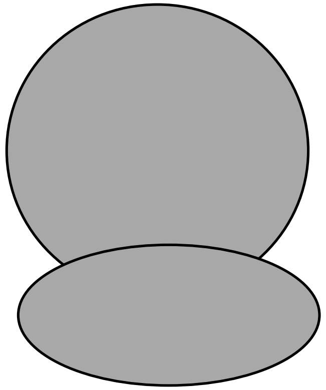
credit: New England Biolabs



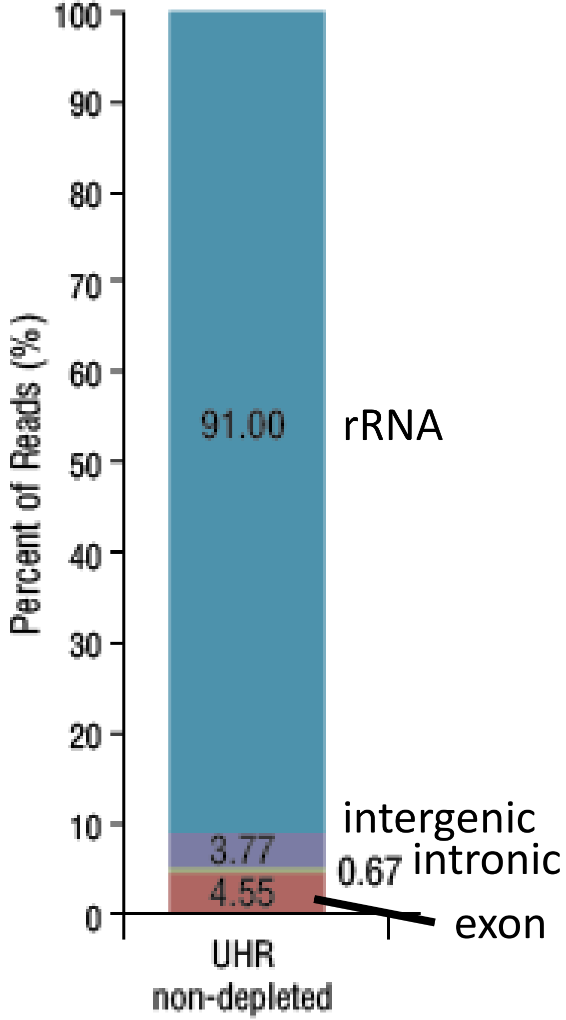


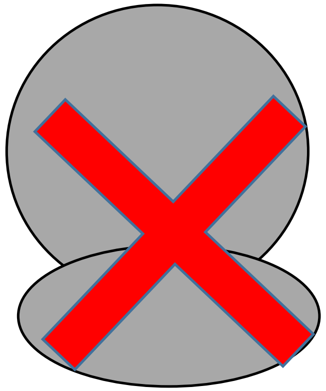
credit: New England Biolabs



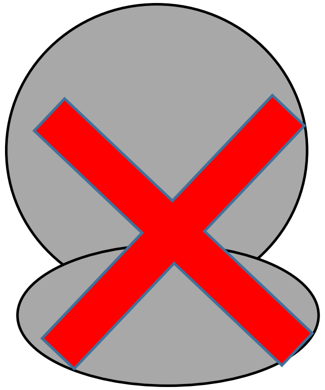


credit: New England Biolabs





credit: New England Biolabs



illumina

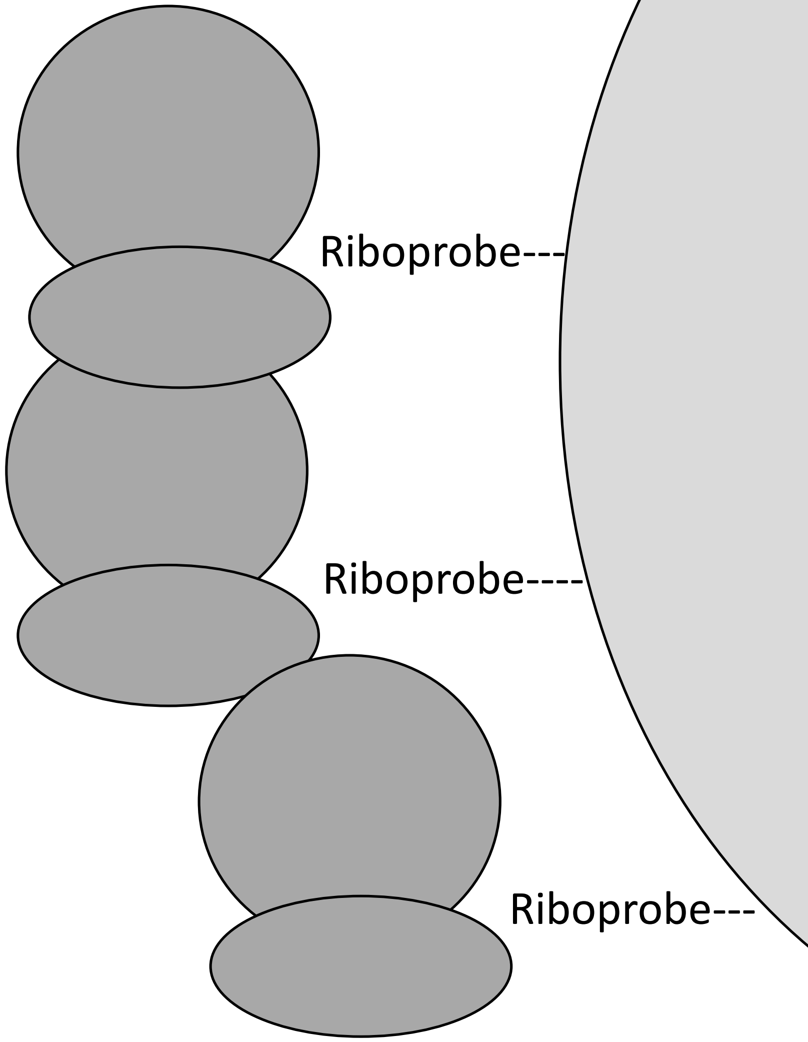
RiboZero

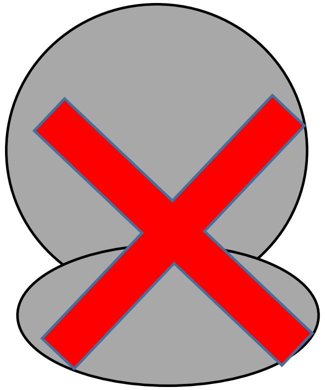
LEXOGEN

RiboCop

QIAGEN

GeneRead rRNA Depletion





illumina
RiboZero

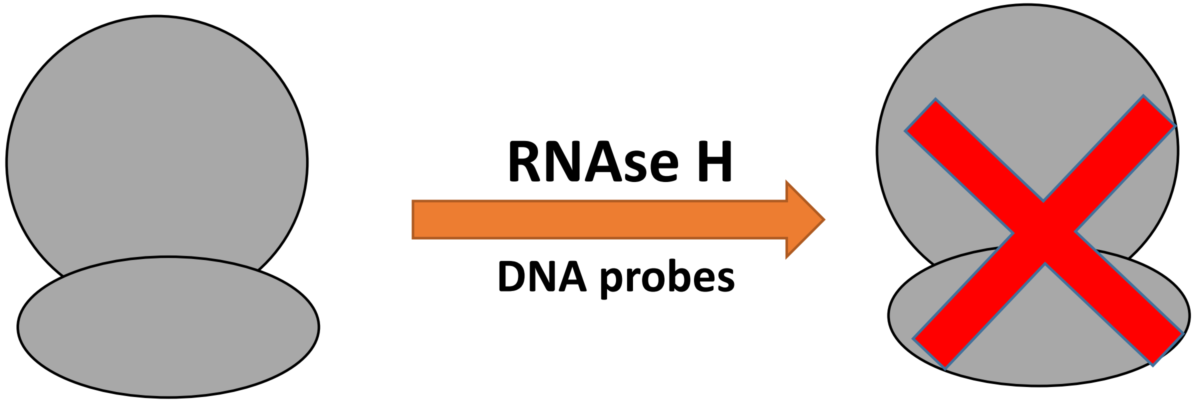
LEXOGEN
RiboCop

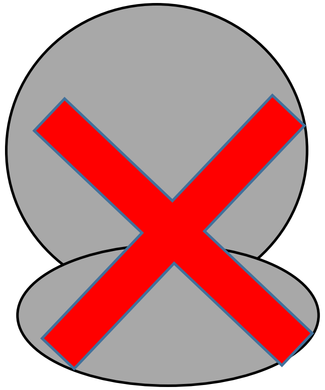
QIAGEN®
GeneRead rRNA Depletion

NEW ENGLAND BioLabs® Inc.
NEBNext rRNA Depletion

KAPA BIOSYSTEMS
RiboErase

TaKaRa
Clontech TaKaRa cellartis
RiboGone





illumina
RiboZero

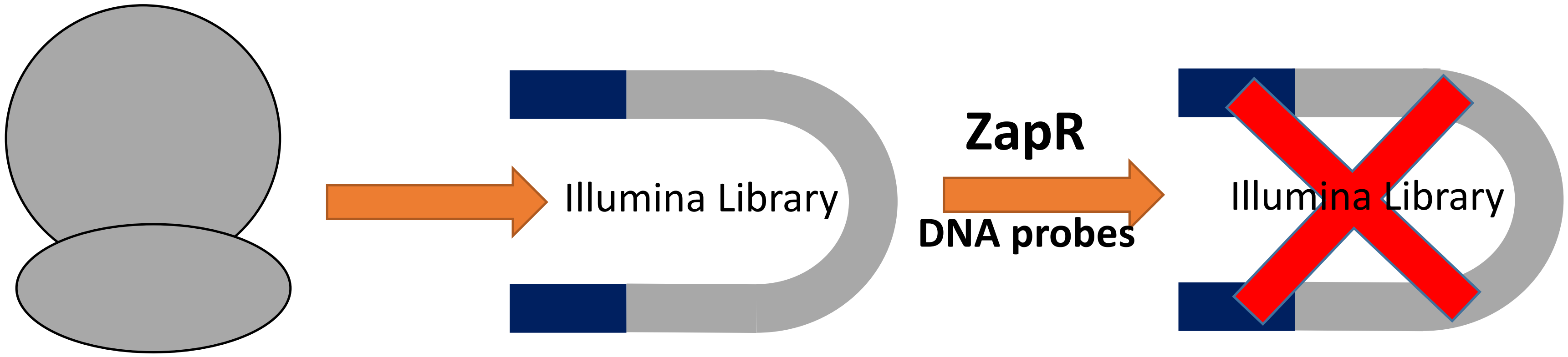
LEXOGEN
RiboCop

QIAGEN
GeneRead rRNA Depletion

NEW ENGLAND BioLabs Inc.
NEBNext rRNA Depletion

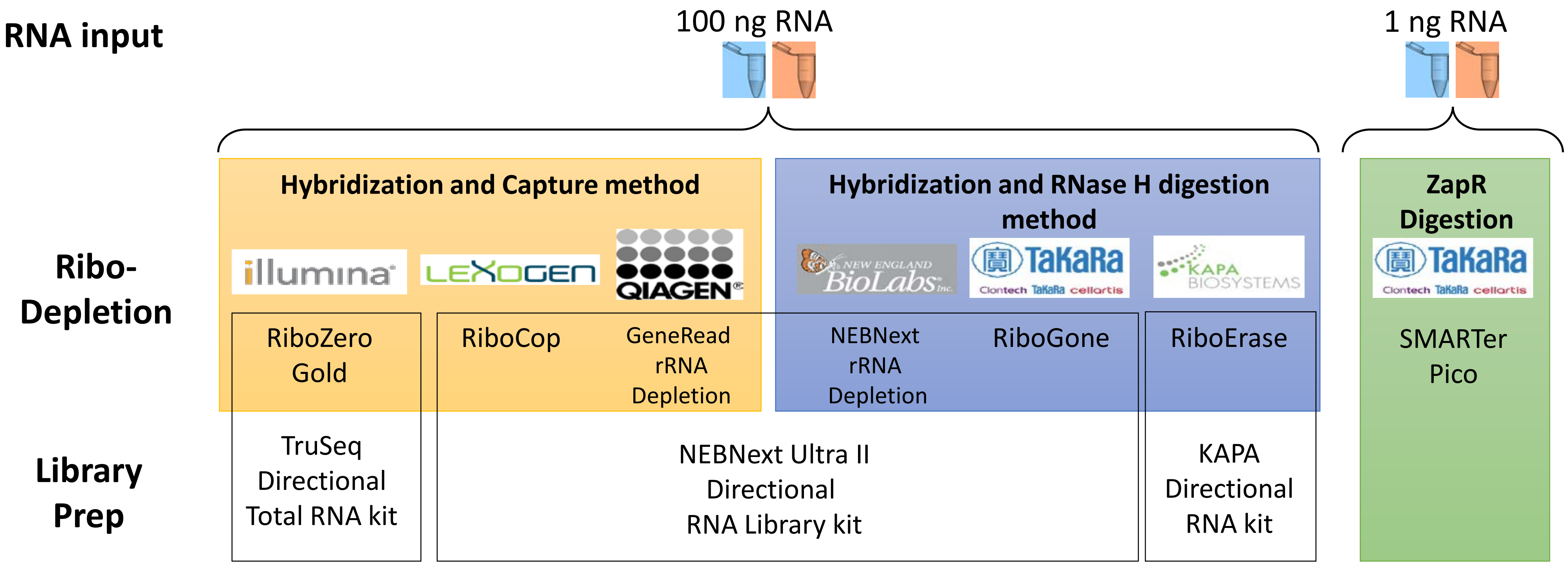
KAPA BIOSYSTEMS
RiboErase

TaKaRa
Clontech TaKaRa cellartis
RiboGone



TaKaRa
Clontech TaKaRa cellartis

SMARTer stranded total RNA - Pico



RNA input	100 ng RNA						1 ng RNA
	Hybridization and Capture method			**Hybridization and RNase H digestion method**			**ZapR Digestion**
Ribo-Depletion	RiboZero Gold	RiboCop	GeneRead rRNA Depletion	NEBNext rRNA Depletion	RiboGone	RiboErase	SMARTer Pico
Library Prep	TruSeq Directional Total RNA kit	NEBNext Ultra II Directional RNA Library kit				KAPA Directional RNA kit	

SAMPLE

UNIVERSAL HUMAN REFERENCE RNA

Agilent Technologies

Mix of RNA from 10 cell lines

SAMPLE

UNIVERSAL HUMAN REFERENCE RNA

Agilent Technologies

Mix of RNA from 10 cell lines

External RNA Controls Consortium SPIKE-IN (A)

Thermo Fisher / Ambion

Mix of 92 synthetic mRNAs across a broad concentration spectrum

SAMPLE

UNIVERSAL HUMAN REFERENCE RNA

Agilent Technologies

Mix of RNA from 10 cell lines

External RNA Controls Consortium SPIKE-IN (A)

Thermo Fisher / Ambion

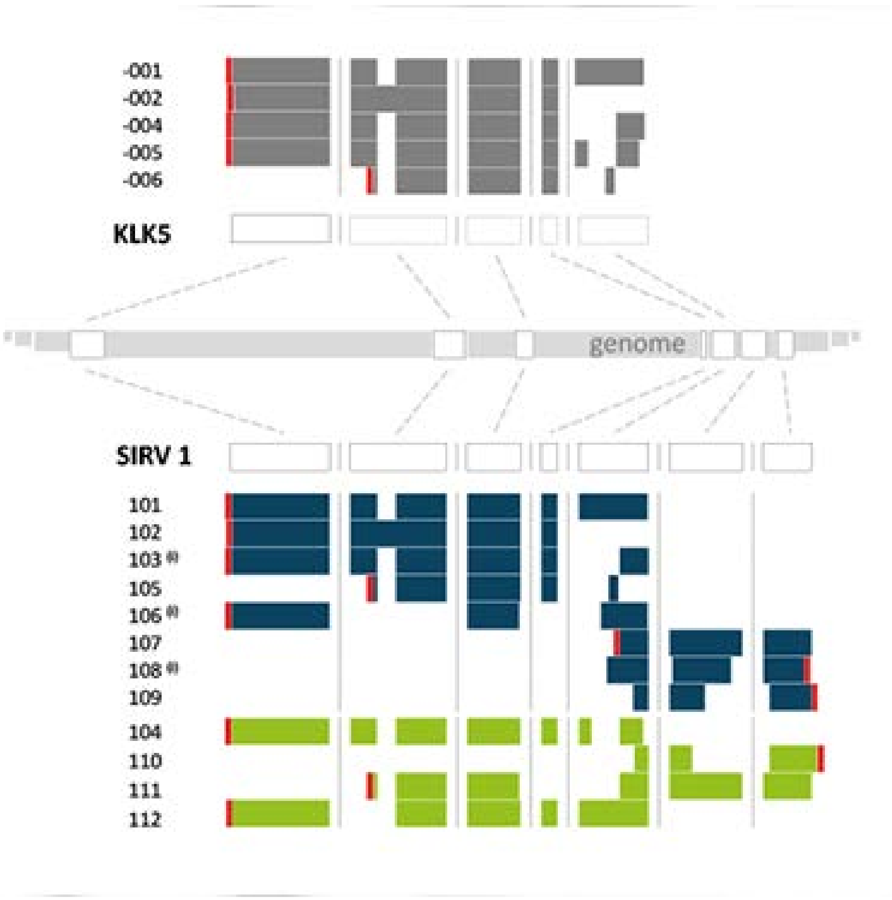
Mix of 92 synthetic mRNAs across a broad concentration spectrum

Spike in RNA Variant Control Mix (Mix E2)

Lexogen

Mix of 70 synthetic mRNAs representing alternative splicing of 10 genes.

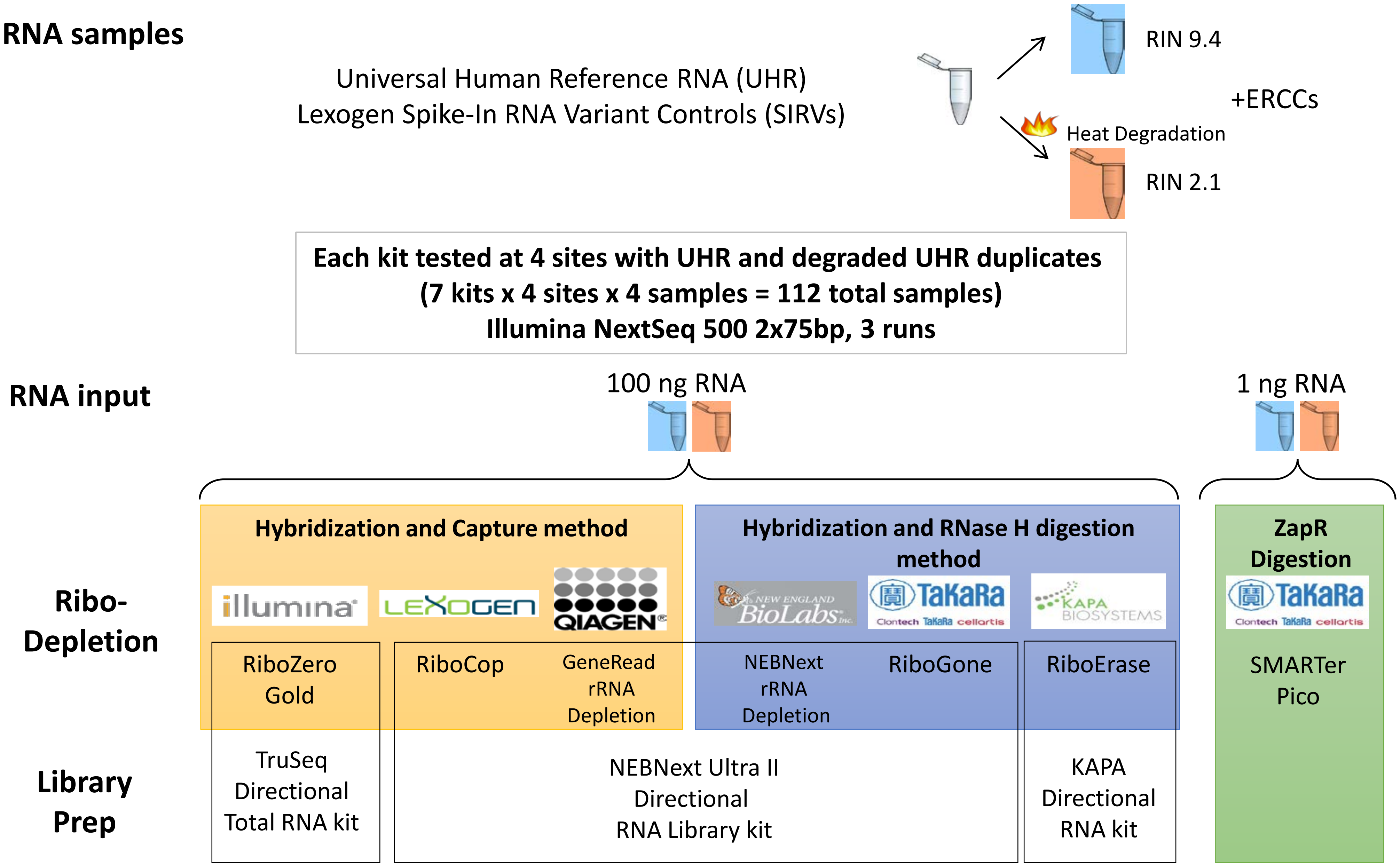
SAMPLE



Spike in RNA Variant Control Mix (Mix E2)

Lexogen

Mix of 70 synthetic mRNAs representing alternative splicing of 10 genes.



RNA samples

Universal Human Reference RNA (UHR)
Lexogen Spike-In RNA Variant Controls (SIRVs)

RIN 9.4

+ERCCs

Heat Degradation

RIN 2.1

Each kit tested at 4 sites with UHR and degraded UHR duplicates
(7 kits x 4 sites x 4 samples = 112 total samples)
Illumina NextSeq 500 2x75bp, 3 runs

RNA input

100 ng RNA

1 ng RNA

Ribo-Depletion

Hybridization and Capture method

RiboZero Gold

RiboCop

GeneRead rRNA Depletion

Hybridization and RNase H digestion method

NEBNext rRNA Depletion

RiboGone

RiboErase

ZapR Digestion

SMARTer Pico

Library Prep

TruSeq Directional Total RNA kit

NEBNext Ultra II Directional RNA Library kit

KAPA Directional RNA kit

BOSTON UNIVERSITY

BYU

CU

MARSHALL

DANA-FARBER
CANCER INSTITUTE

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

NH
University of
New Hampshire

TEXAS
The University of Texas at Austin

OU

VAN ANDEL
RESEARCH INSTITUTE

UNIVERSITY OF CALIFORNIA
SAN DIEGO
LET THERE BE LIGHT

BOSTON UNIVERSITY

BYU

CU

MARSHALL

DANA-FARBER CANCER INSTITUTE

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

NH
University of New Hampshire

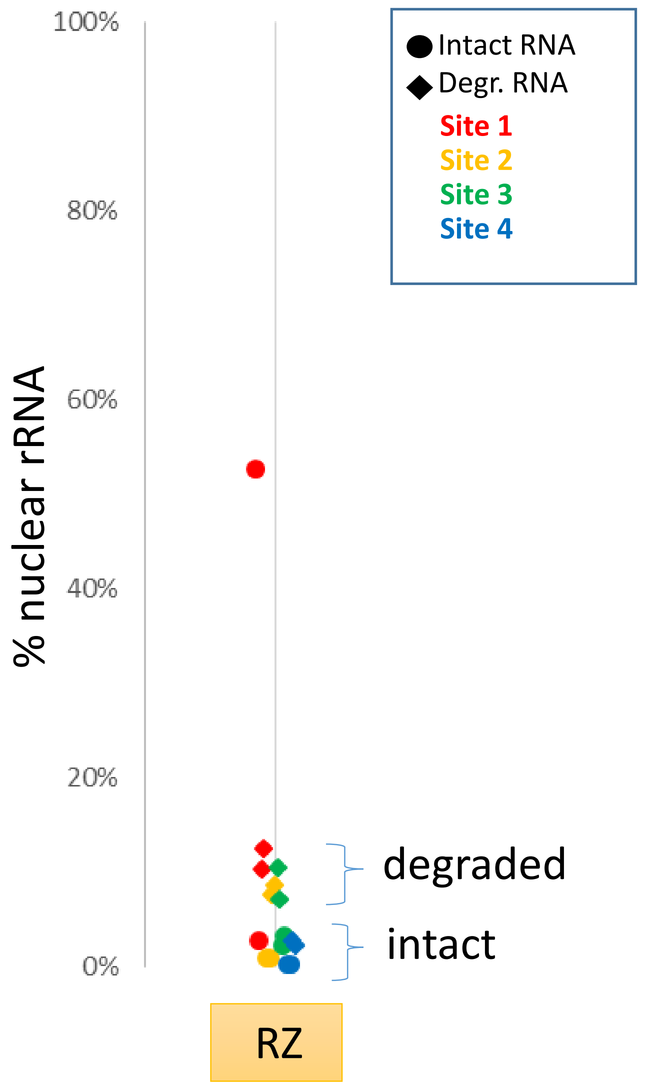
TEXAS
The University of Texas at Austin

OU

VAN ANDEL RESEARCH INSTITUTE

UNIVERSITY OF CALIFORNIA · SAN DIEGO ·

CONFERENCE CALL BETWEEN THE SITES AND VENDOR BEFORE EXPERIMENTS BEGAN



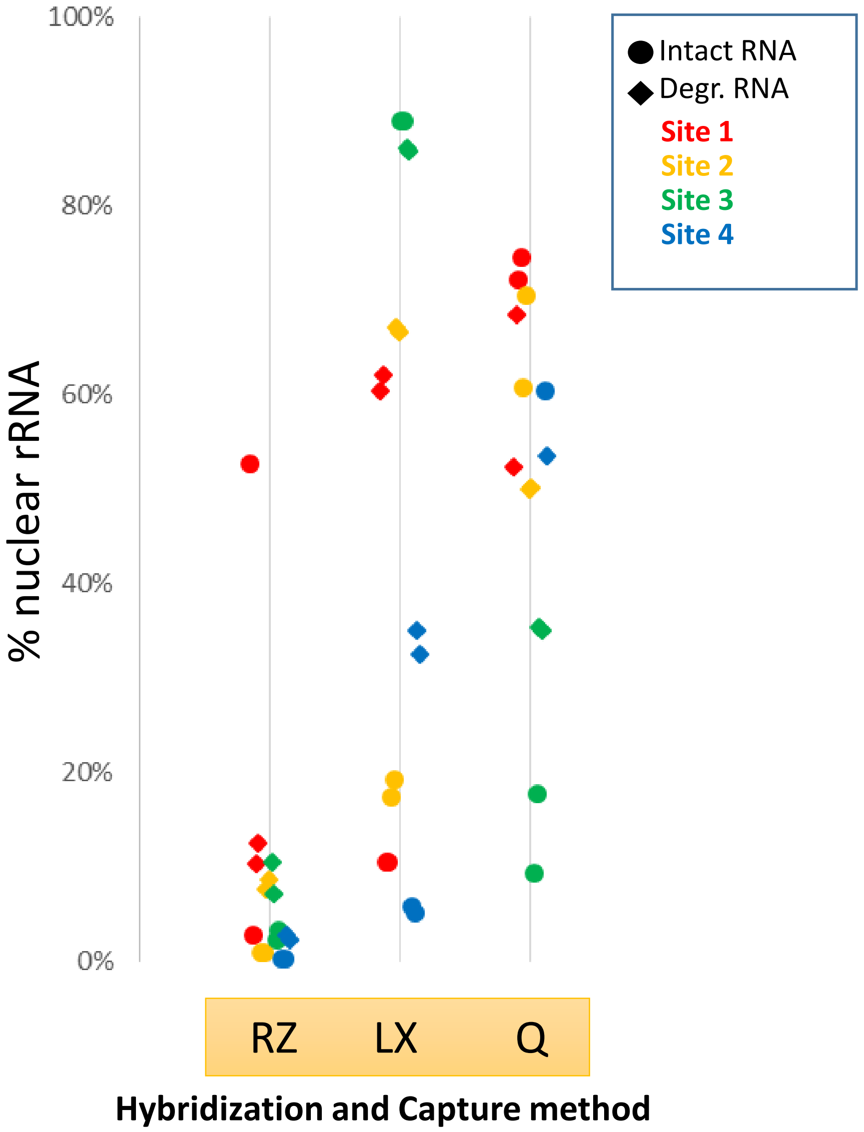
Quality metrics for the rRNA libraries.

BWA was used to map 1 million reads per sample to rRNA sequences from hg19.

- SITE# is unique per kit. (Site 1 for RZ is not necessarily site 1 for any other kit.

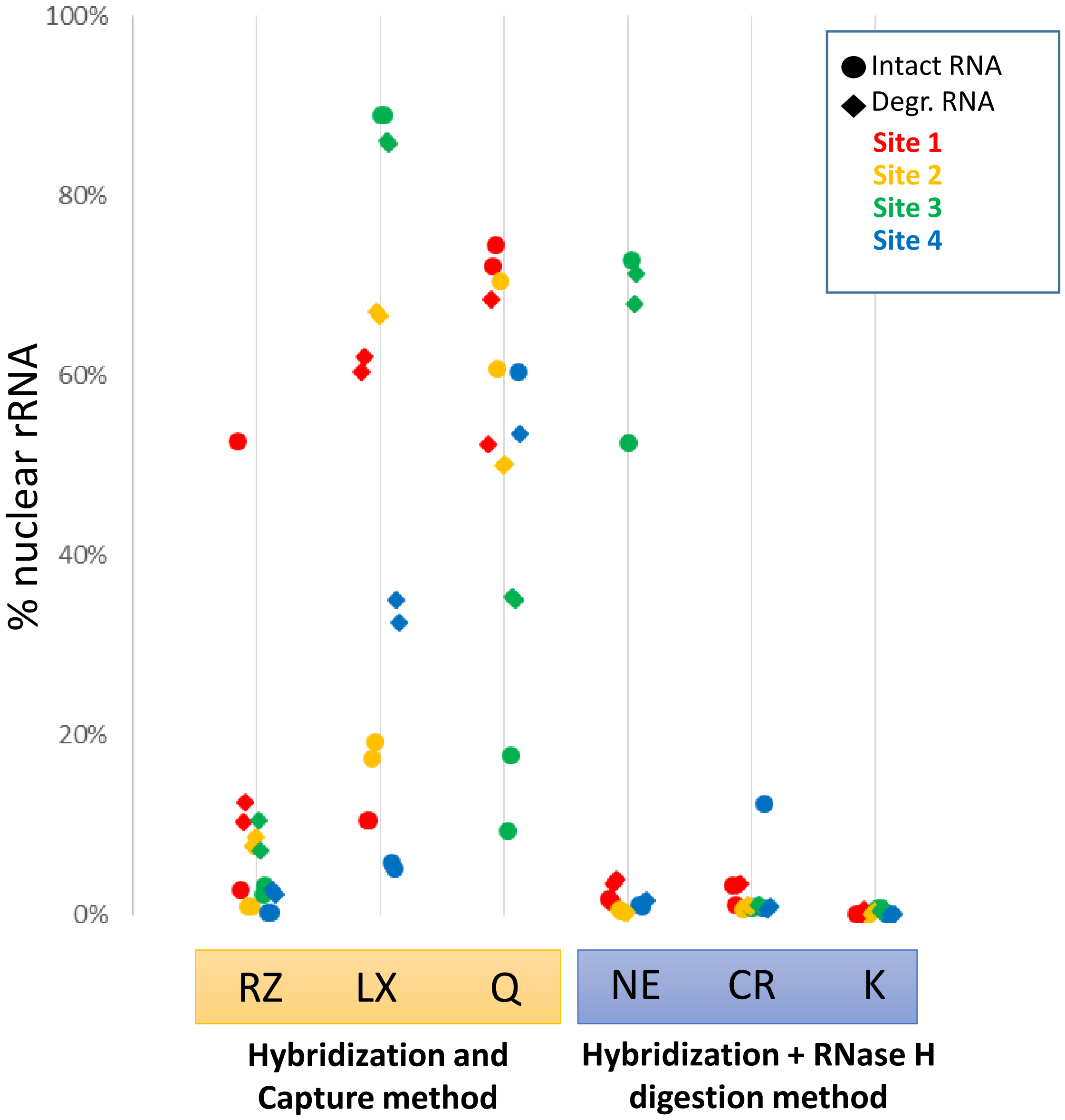
The Association of Biomolecular Resource Facilities

Research • Technology • Communication • Education

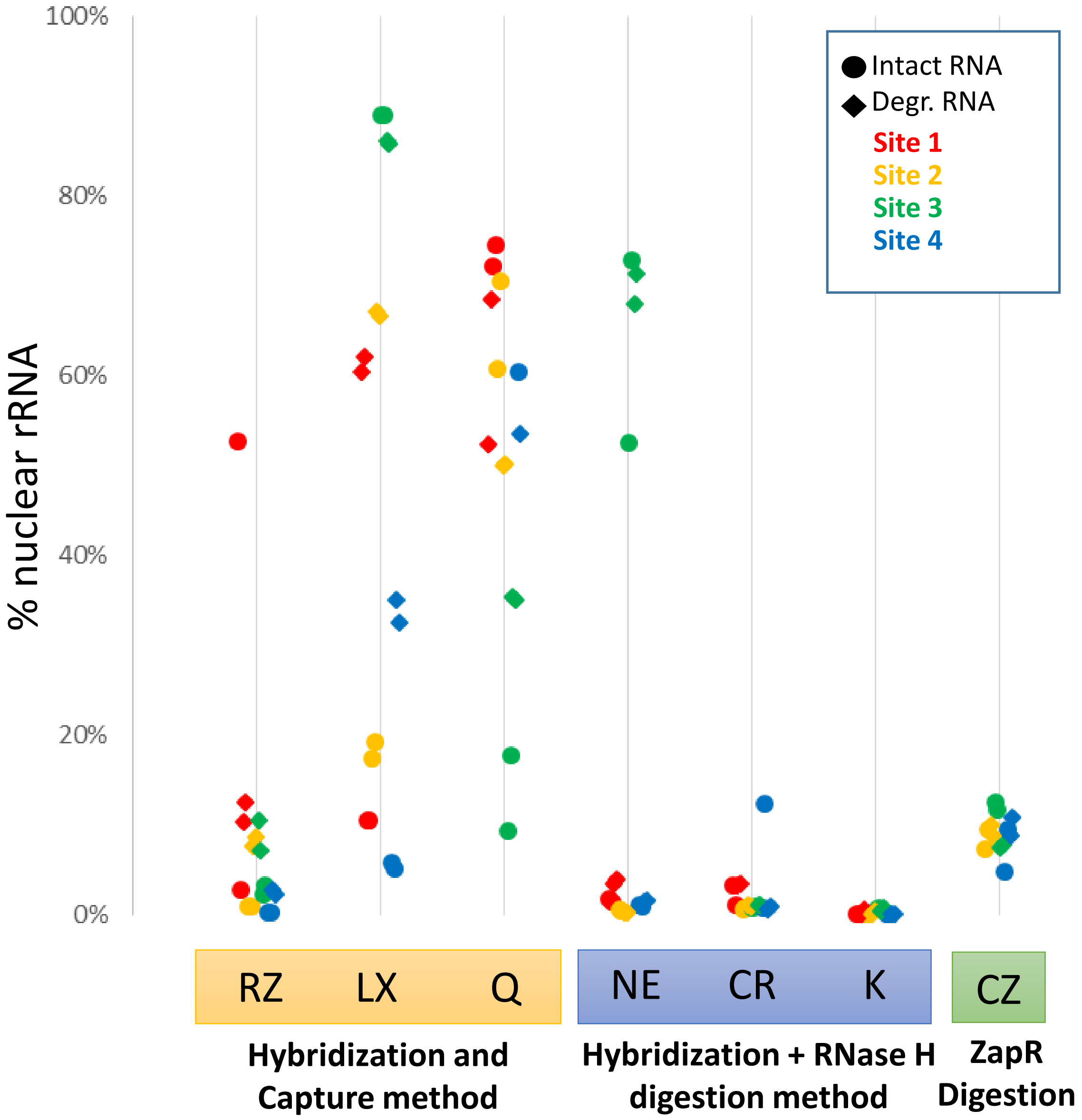


- ➢ Newer capture methods are challenging
- ➢ Capture methods often show differences between intact and degraded samples

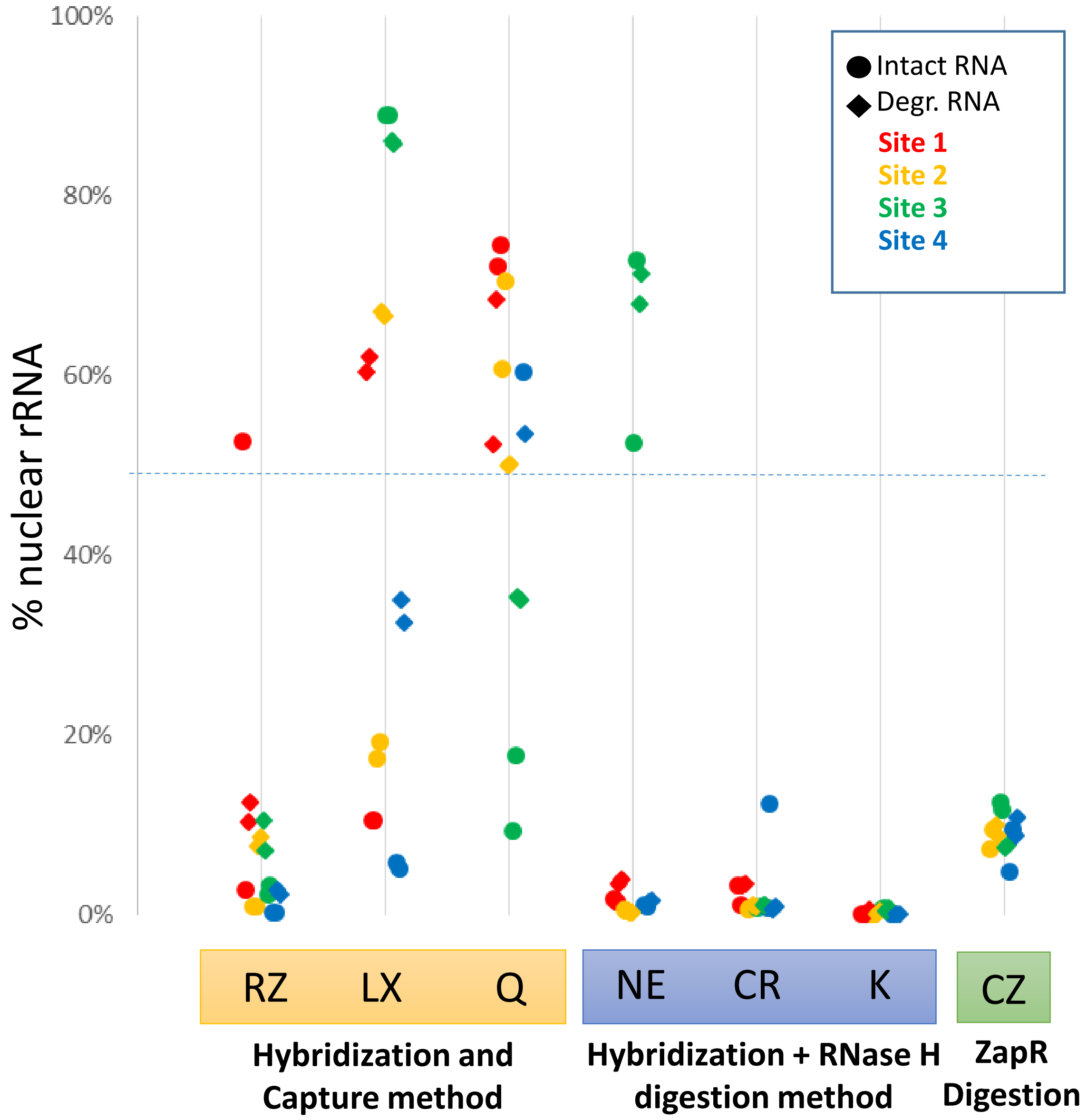
The Association of Biomolecular Resource Facilities
Research • Technology • Communication • Education



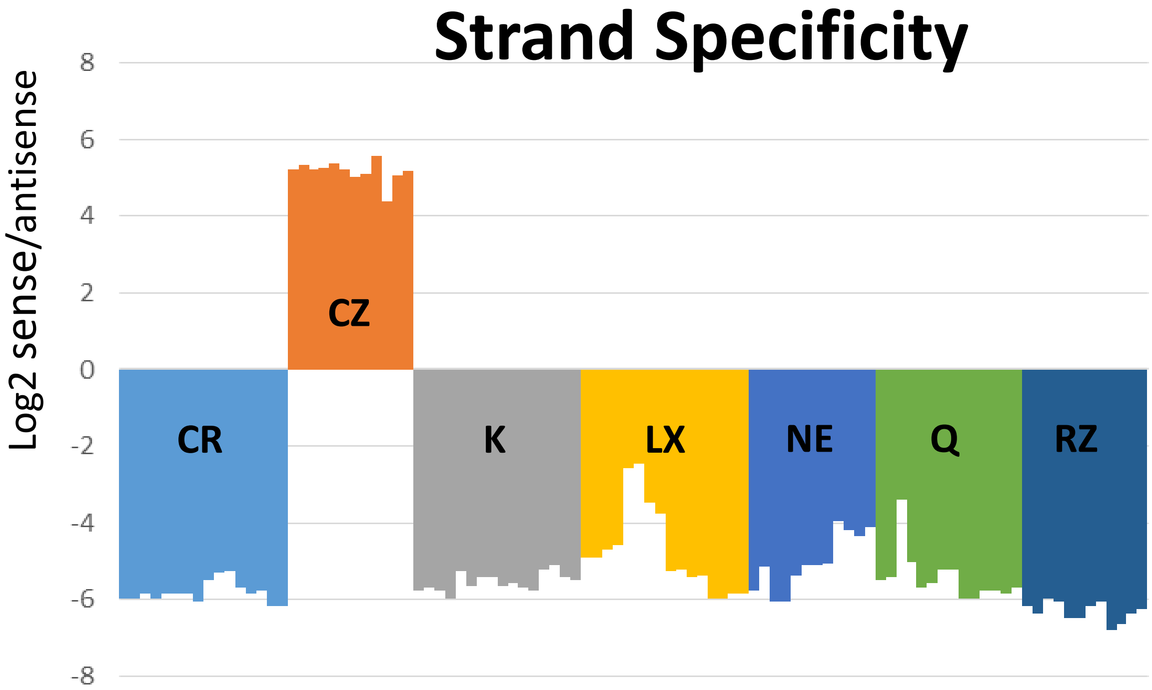
- Newer capture methods are challenging
- Capture methods often show differences between intact and degraded samples
- RNaseH methods were generally easier to implement and preformed very well.



- Newer capture methods are challenging
- Capture methods often show differences between intact and degraded samples
- RNaseH methods were generally easier to implement and preformed very well.
- ZapR method worked well when it produced library.



- Newer capture methods are challenging
- Capture methods often show differences between intact and degraded samples
- RNaseH methods were generally easier to implement and preformed very well.
- ZapR method worked well when it produced library.
- All kits can work well at removing rRNA.
- SAMPLES >50% rRNA REMOVED FROM FURTHER ANALYSIS

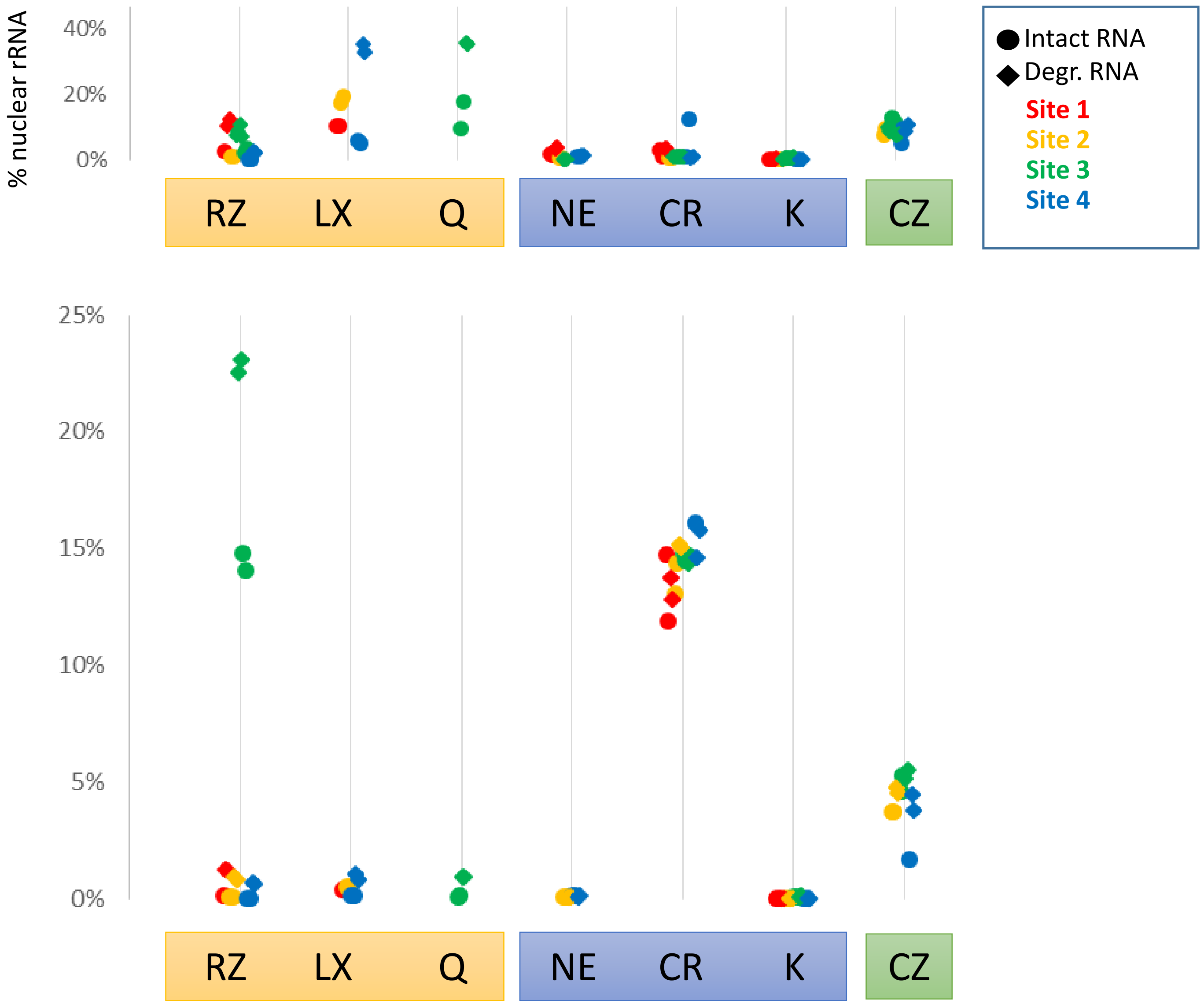


Quality metrics for the rRNA libraries.

BWA was used to map 1 million reads per replicate to hg19 build and scored based on overlapping exons annotated in UCSC knowngene track.

- Stranding of all kits is robust
- ZapR kit is on Watson.

Quality metrics for the rRNA libraries. mtRNA fraction determined using RSEM mapping to ENSEMBL assembly and summing reads mapping to mtRNA transcripts.



% mt rRNA

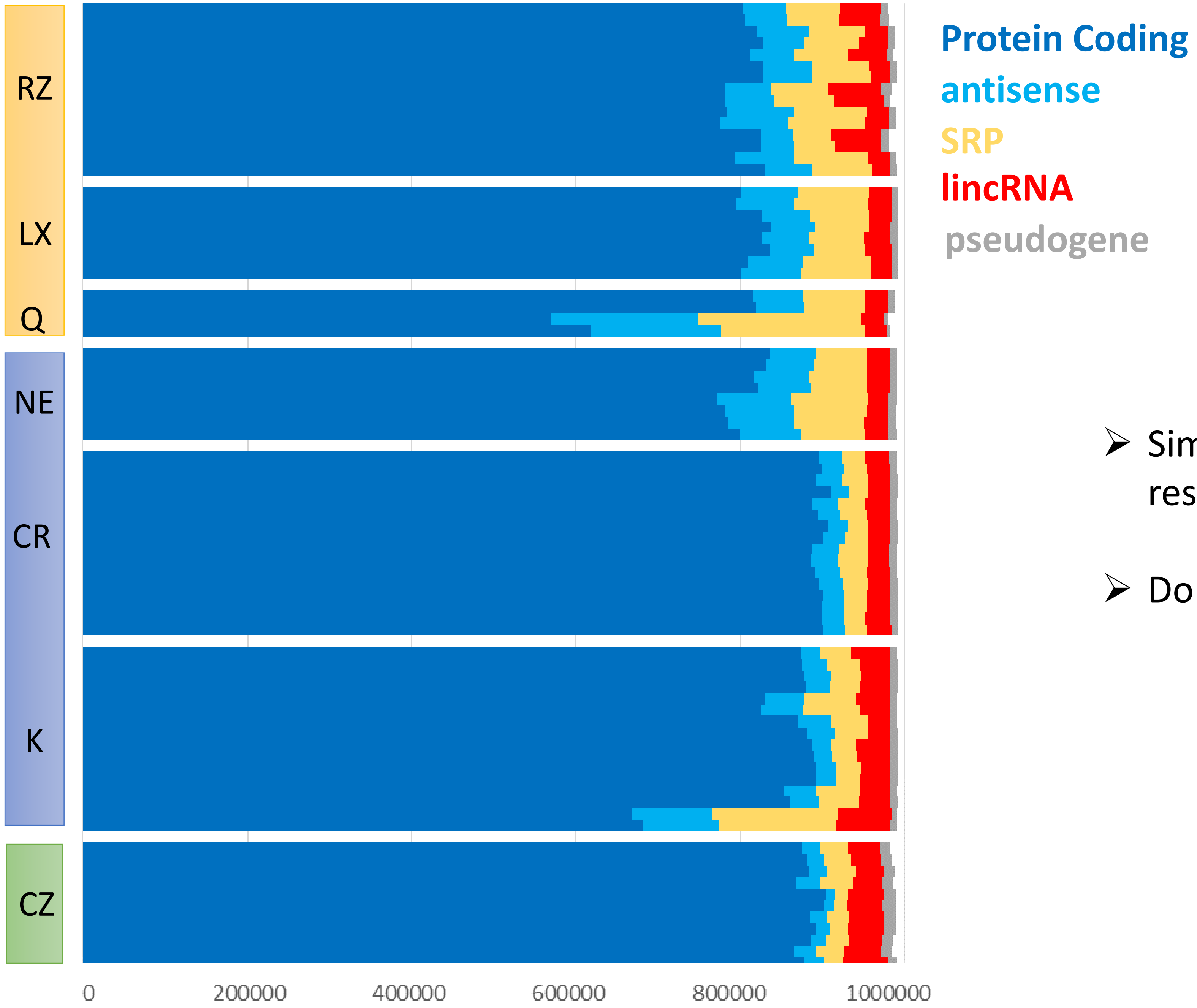
- Ribogone kit does not remove mtRNAs.
- Nor does RiboZero unless you buy the ‘Gold’ version.
- mtRNA can take up a significant fraction of reads.

DNA Sequencing Research Group

The Association of Biomolecular Resource Facilities

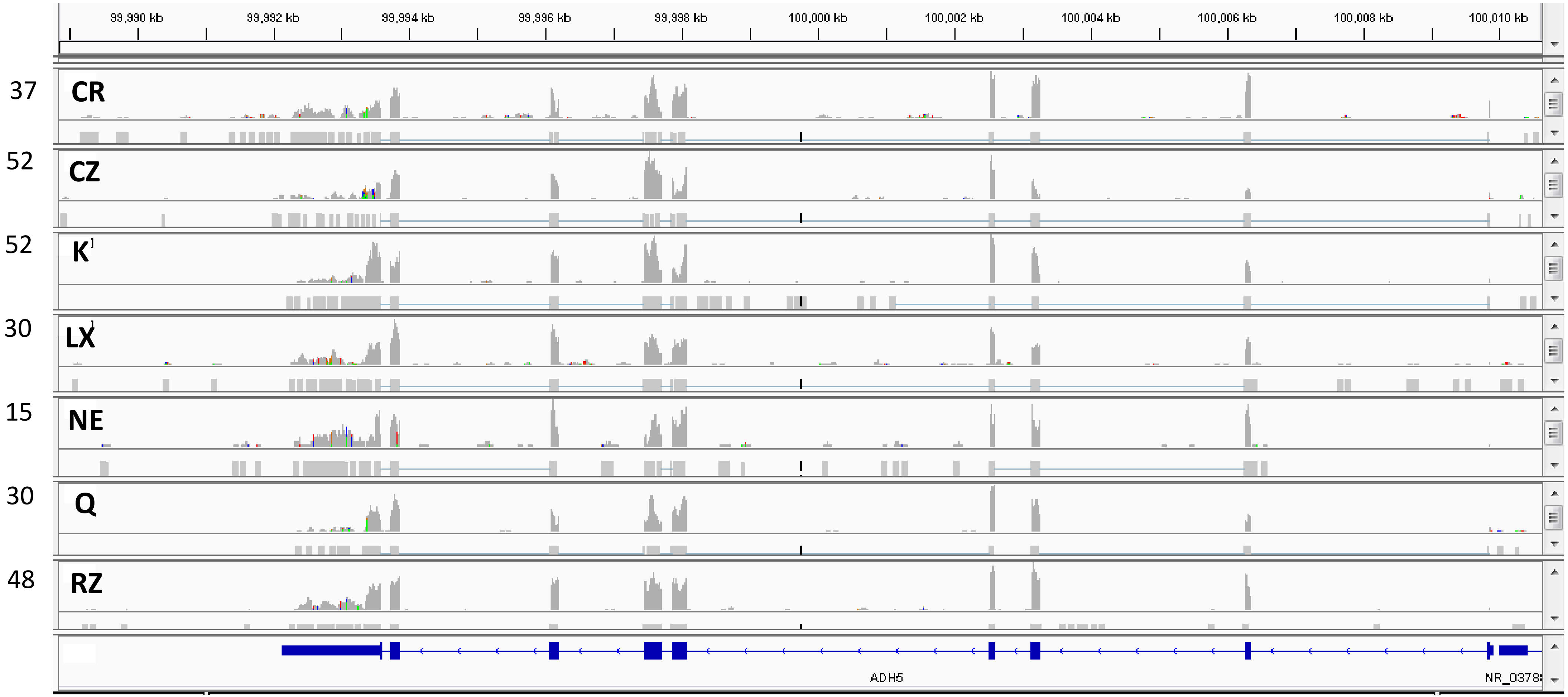
Research • Technology • Communication • Education

READ DISTRIBUTION



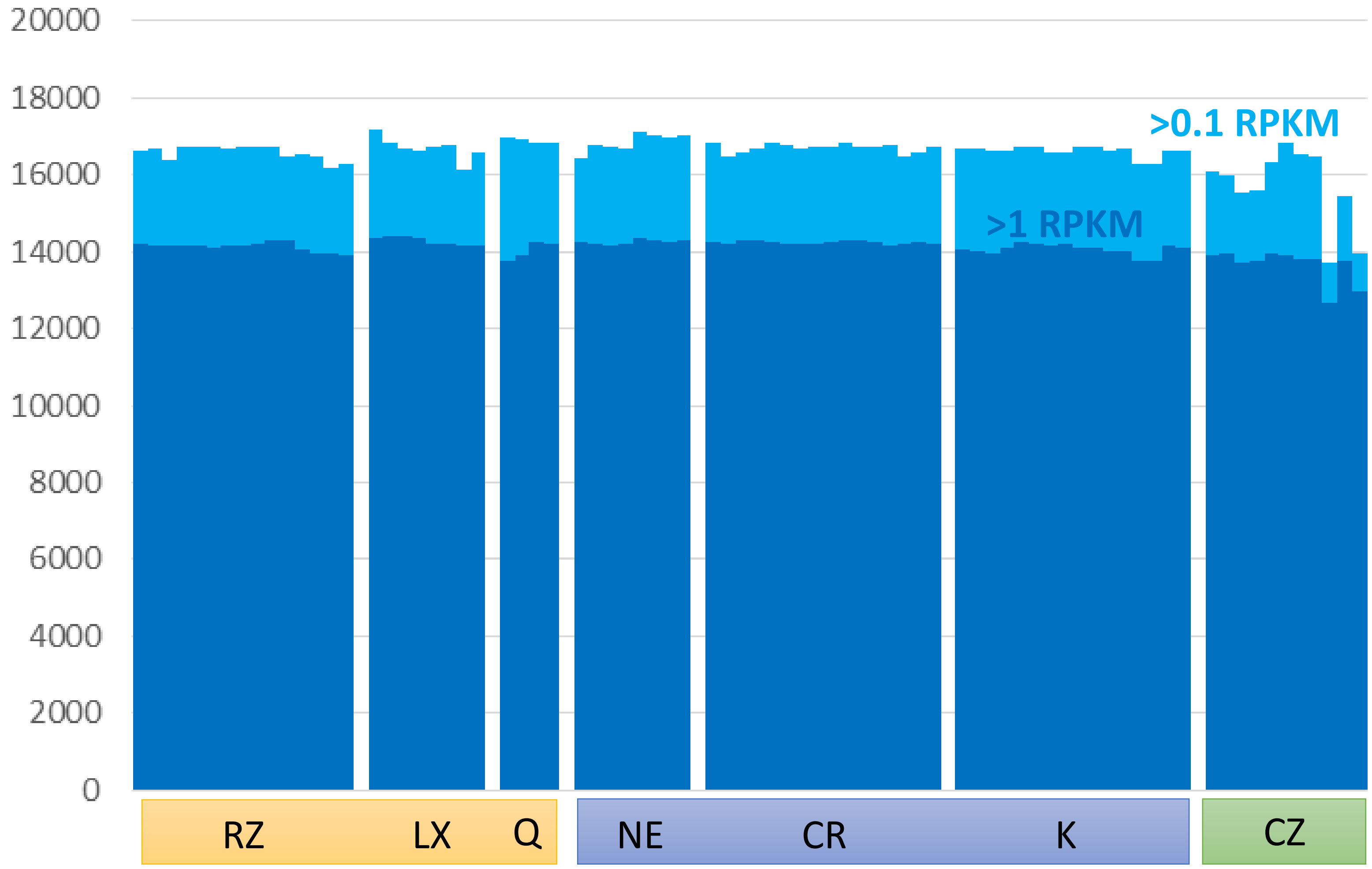
- Similar profiles from all kits after removing residual rRNA and mtRNAs
- Dominated by protein coding RNAs.

PROTEIN CODING GENES



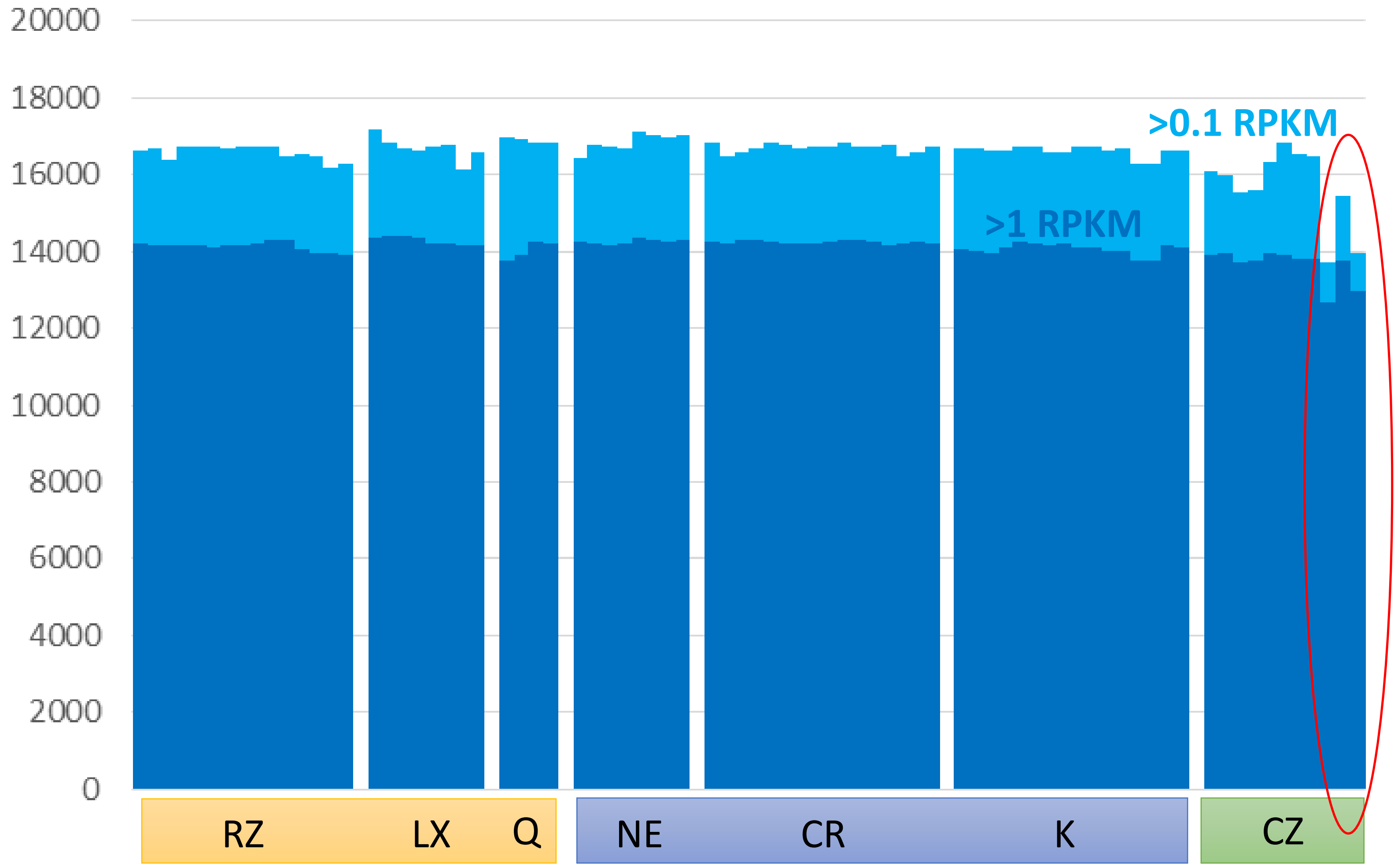
PROTEIN CODING GENES

- All kits detect ~14k genes at >1RPKM and ~16k genes at >0.1RPKM



PROTEIN CODING GENES

- All kits detect ~14k genes at >1RPKM and ~16k genes at >0.1RPKM

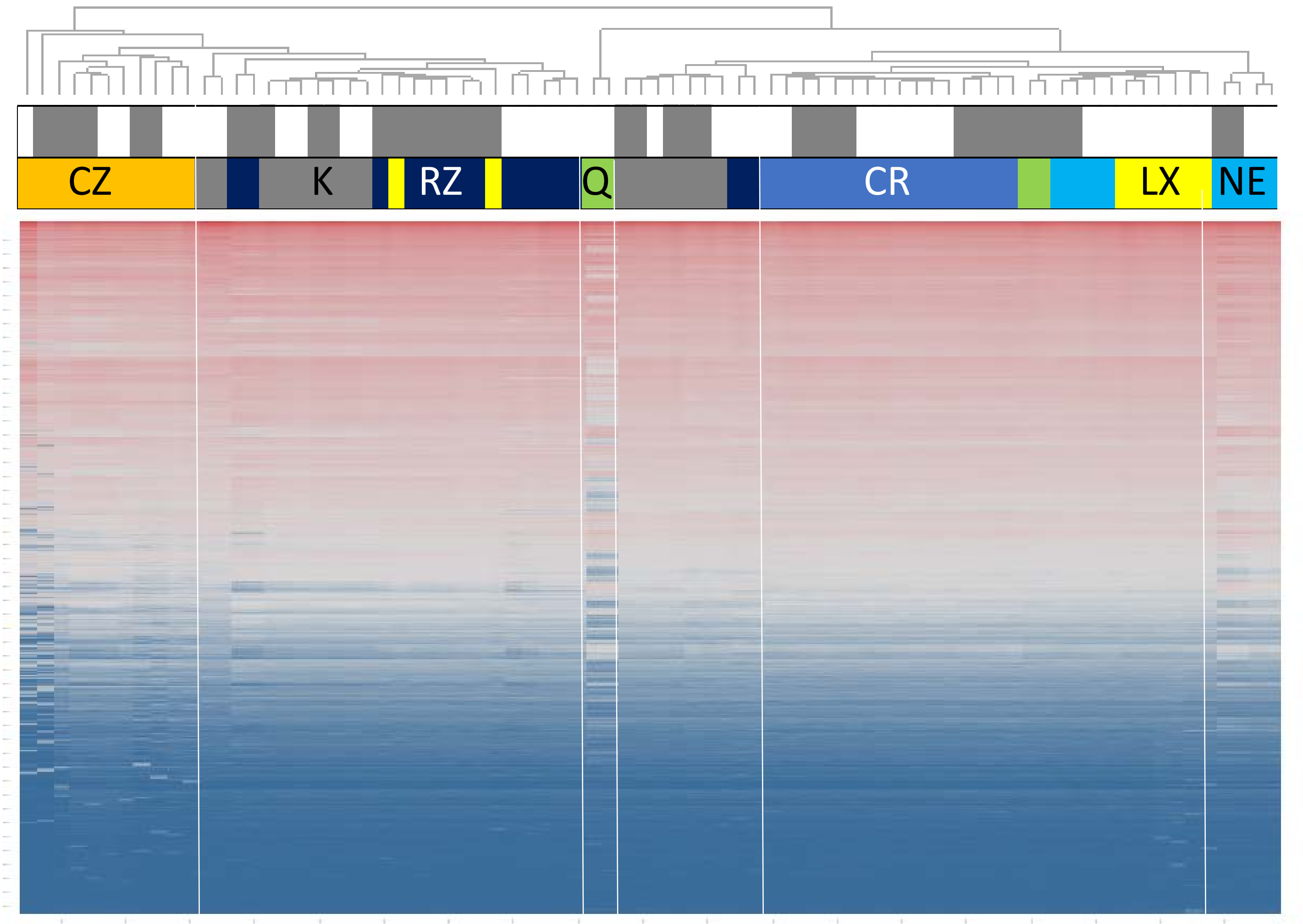


DNA Sequencing Research Group

The Association of Biomolecular Resource Facilities

Research • Technology • Communication • Education

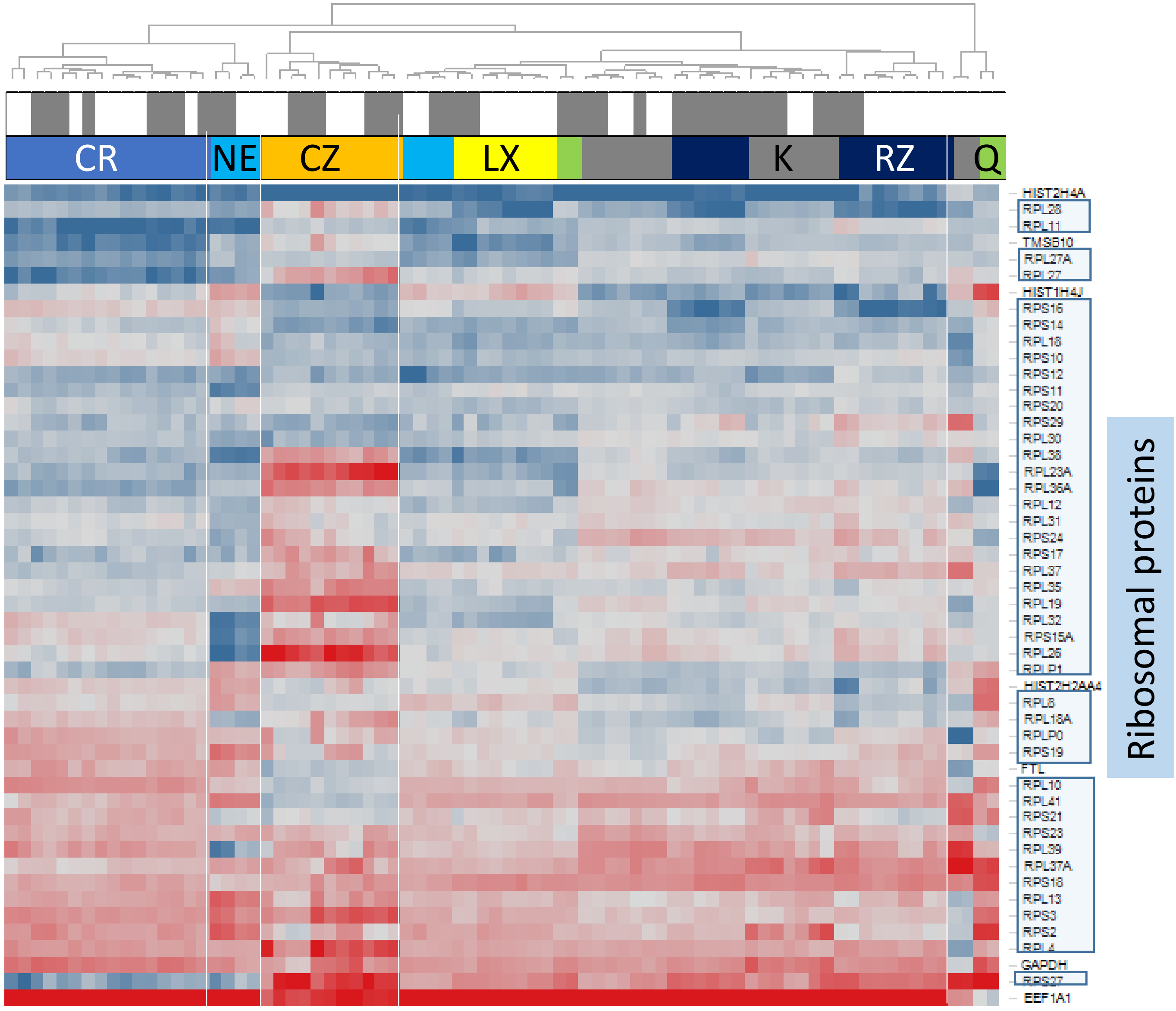
PROTEIN CODING GENES – GLOBAL CLUSTER



N=20238 protein-coding genes (ENSEMBL)

- Looking for clades separating kits.
- NEB RNAprep cluster together – globally.
- Clear clades for ZapR kit only

HIGHEST VARIABILITY GENES

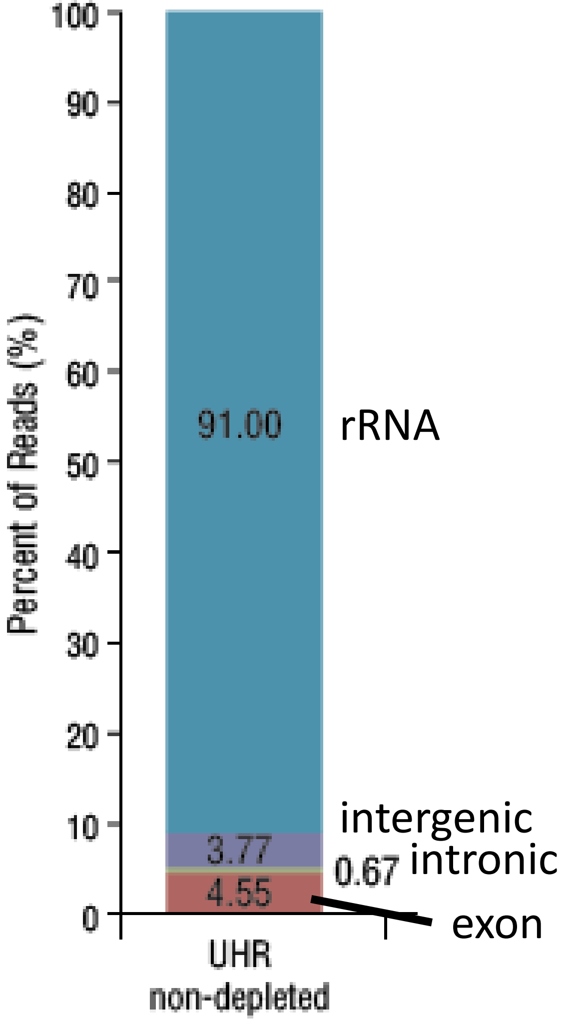


- ➢ More clear clades
  - ZapR & RiboGone

- ➢ Most proteins are ribosomal

Overall Success Rate:

Kit	SUCCESS:TOTAL	FAIL MODE	COMMENTS
Illumina RiboZero Gold	**4:4**		
Qiagen GeneRead rRNA Depletion Nano	**1:4**	Lack of rRNA depletion	
Lexogen RiboCop	**3:4 – 0:4 degraded**	No depletion, poor on degraded	
KAPA RiboErase	**4:4**		
NEBNext rRNA Depletion	**3:4**	Lack of rRNA depletion	
Clonetech RiboGone	**4:4**		Possible differences in transcript levels.
Clonetech SMARTer Pico	**2:4**	No library / poor complexity	Possible differences in transcript levels.

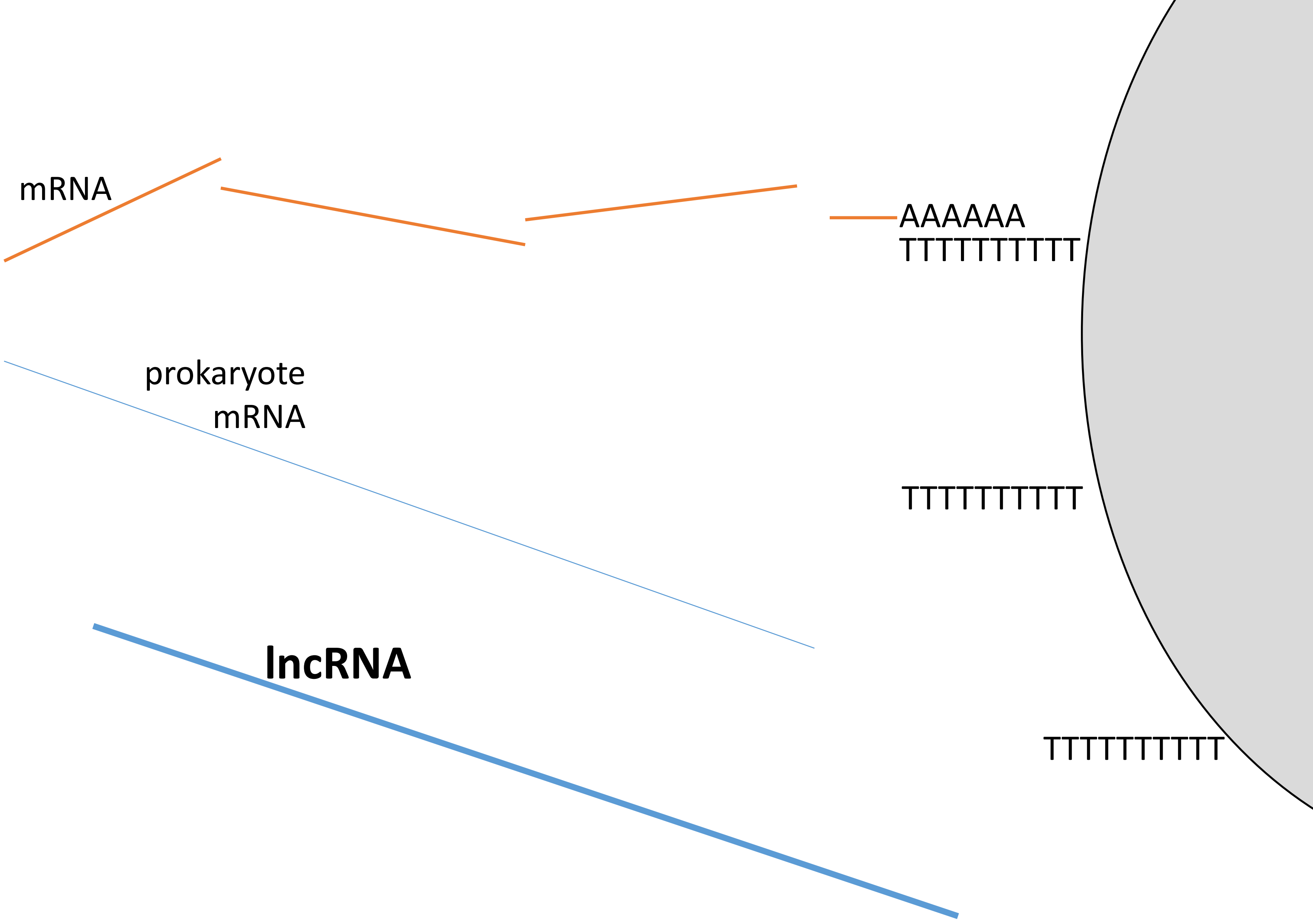


rRNA

intergenic

intronic

exon



mRNA

AAAAAA

TTTTTTTTTT

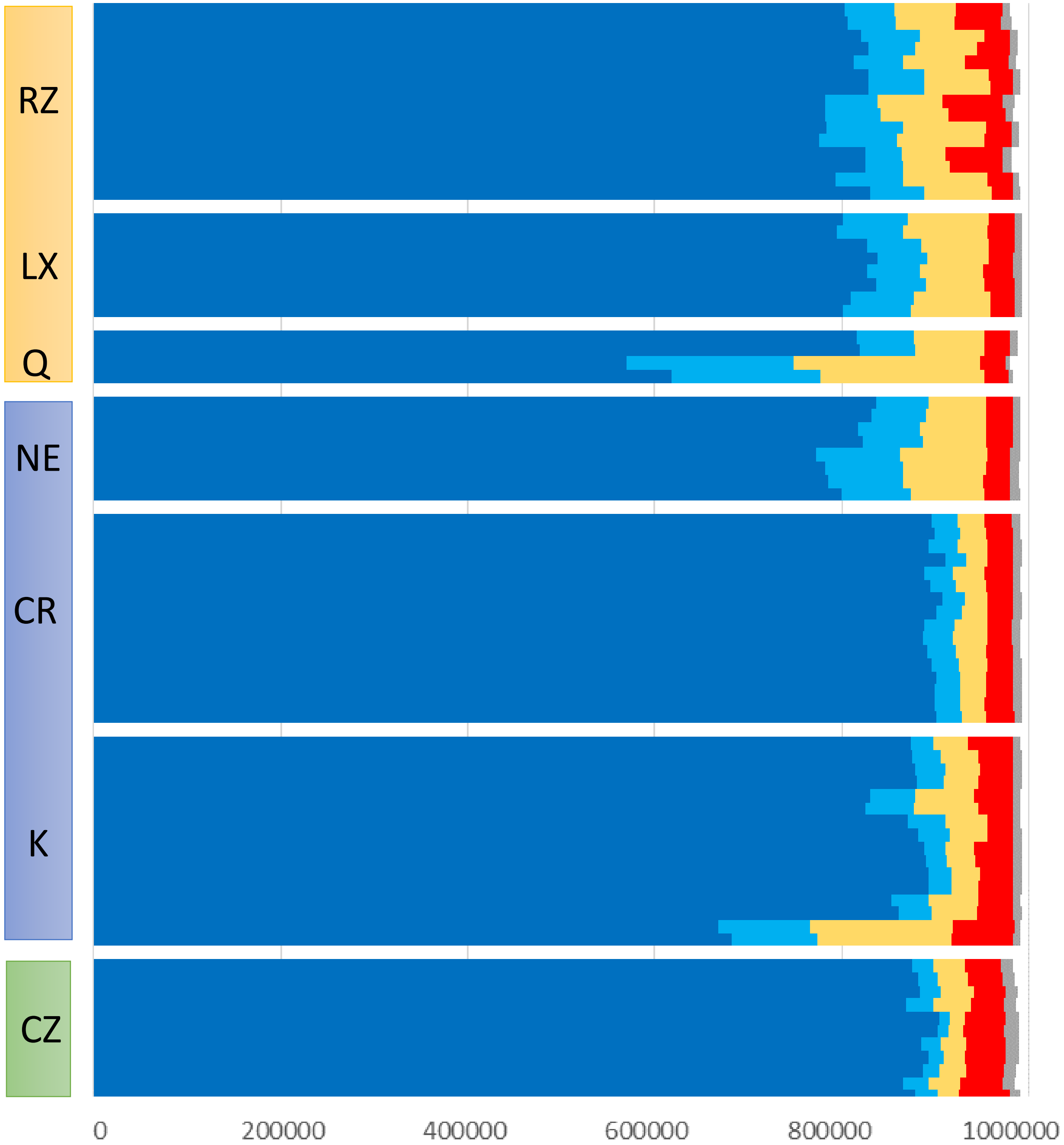
prokaryote
mRNA

TTTTTTTTTT

lncRNA

TTTTTTTTTT

READ DISTRIBUTION



Protein Coding

antisense

SRP

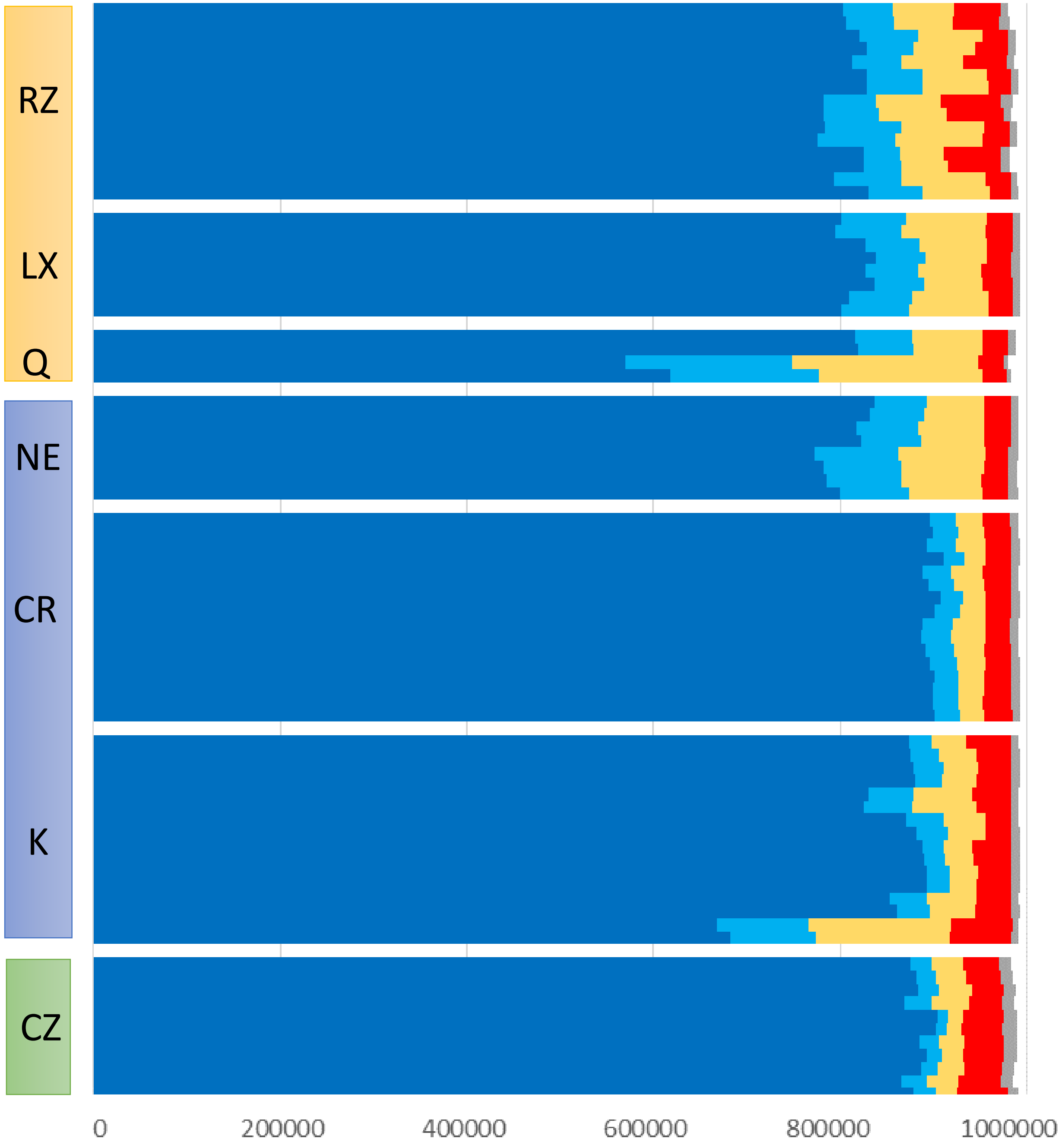
lincRNA

pseudogene

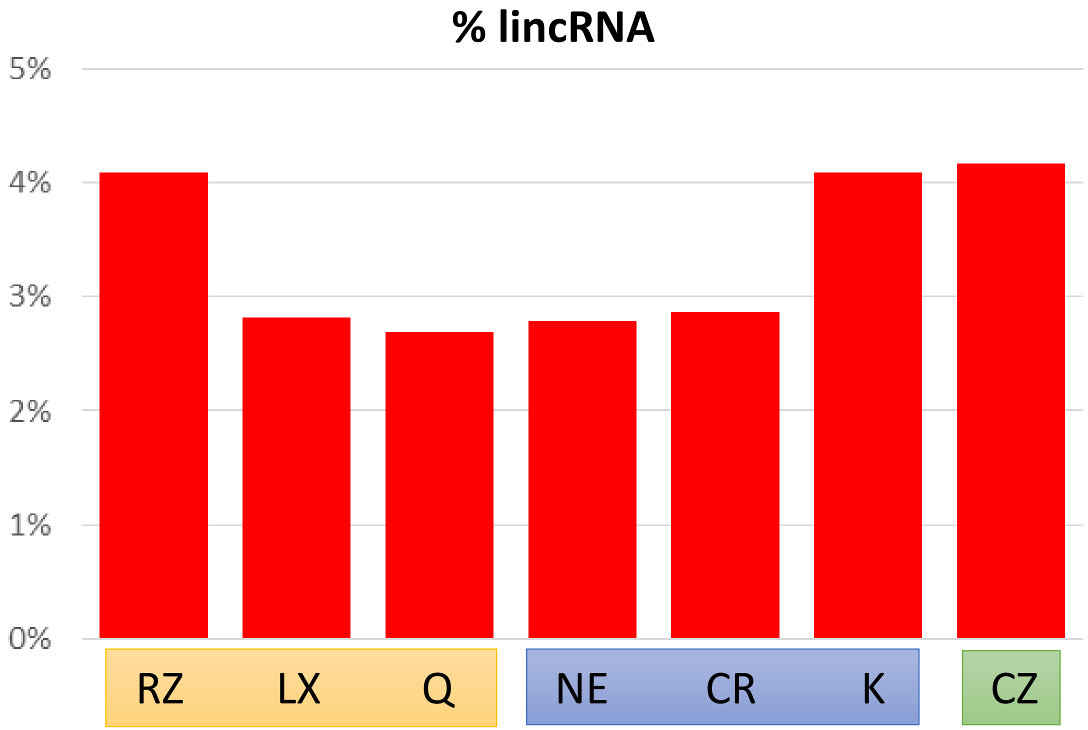
The Association of Biomolecular Resource Facilities

Research • Technology • Communication • Education

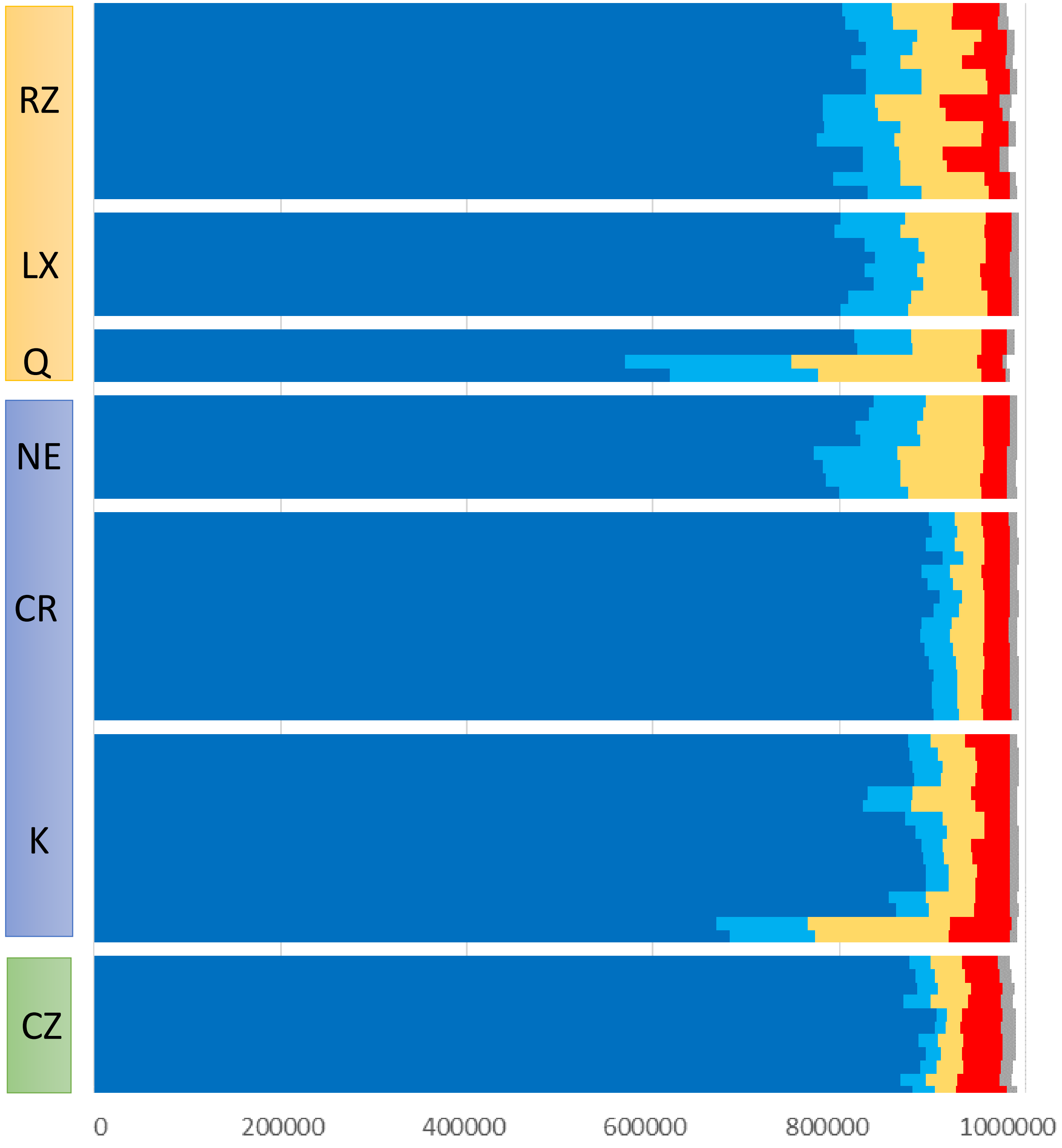
READ DISTRIBUTION

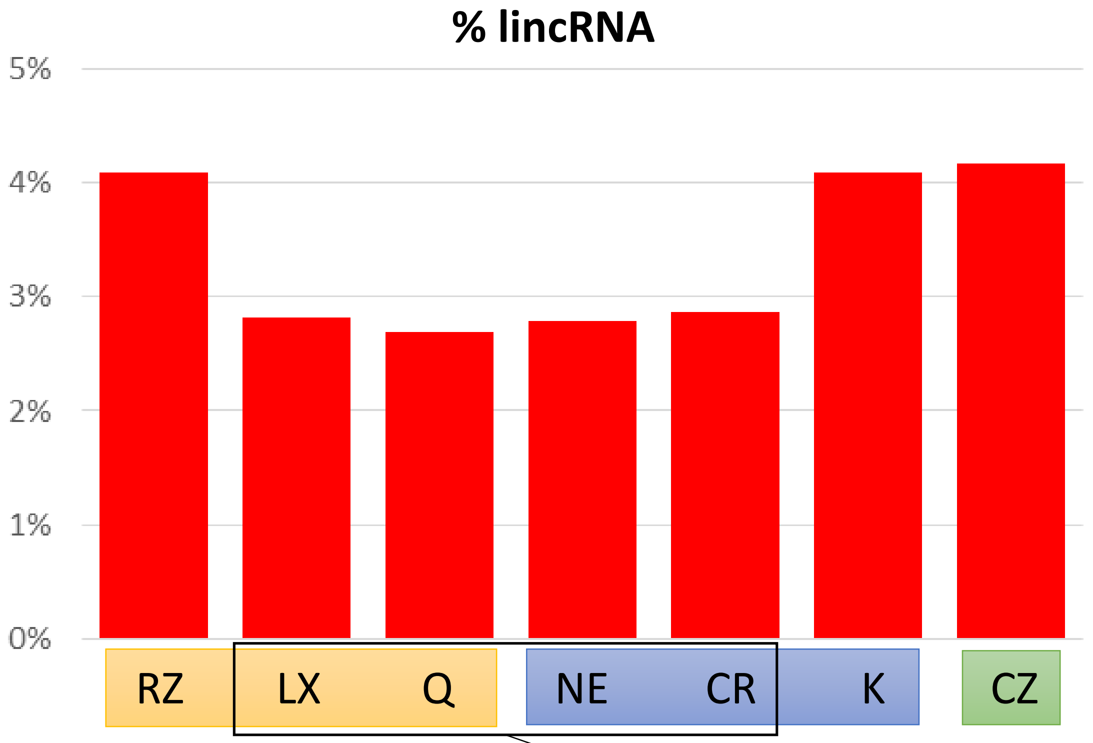


Protein Coding
antisense
SRP
lincRNA
pseudogene

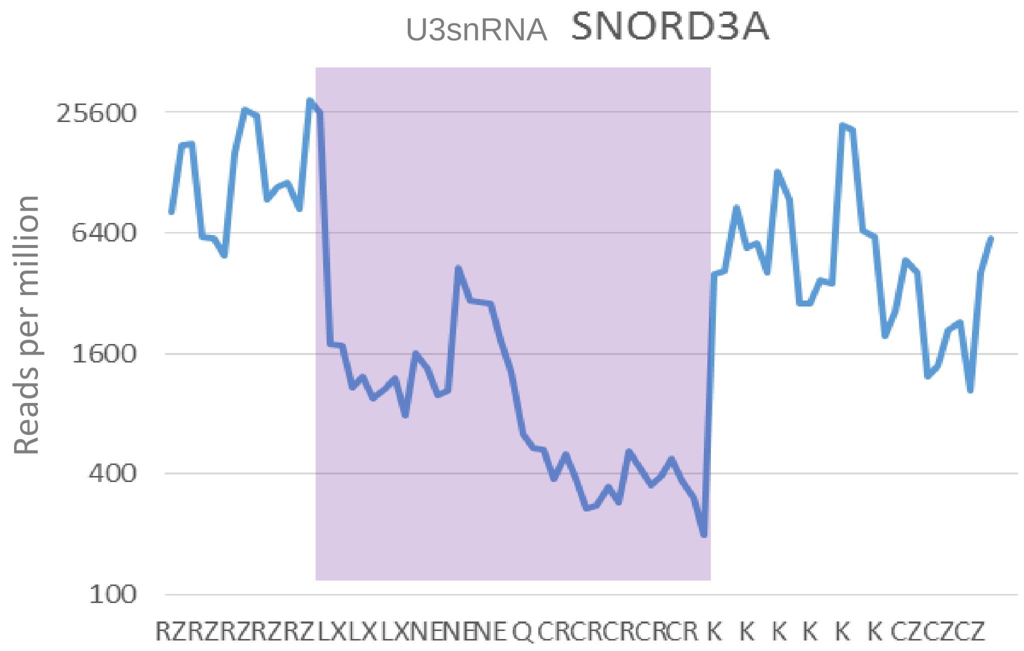


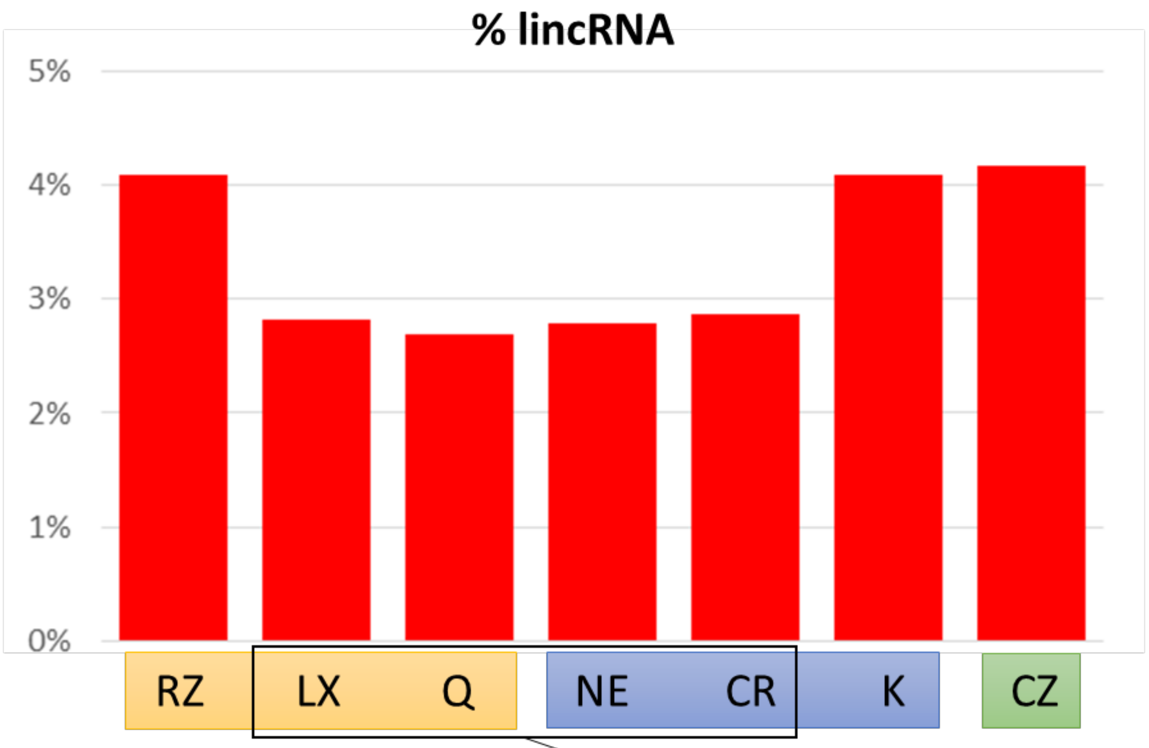
READ DISTRIBUTION





NEB Ultra II Directional RNAseq prep.





NEB Ultra II Directional RNAseq prep.

LOOKING FOR GROUND TRUTH.

UNIVERSAL HUMAN REFERENCE RNA

Agilent Technologies

Mix of RNA from 10 cell lines

External RNA Controls Consortium SPIKE-IN

Thermo Fisher / Ambion

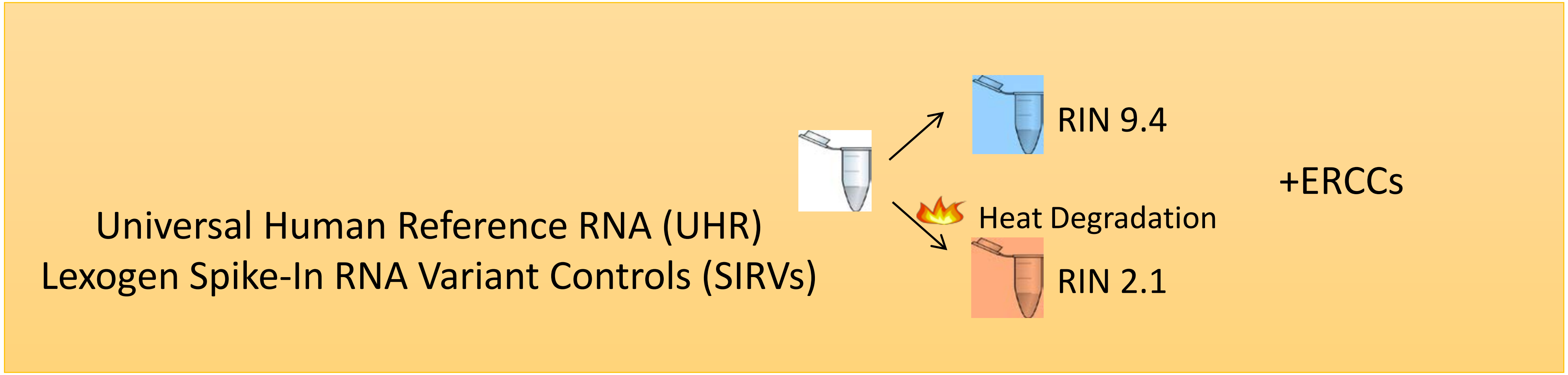
Mix of 92 synthetic mRNAs across a broad concentration spectrum

Spike in RNA Variant Control Mix (Mix E2)

Lexogen

Mix of 70 synthetic mRNAs representing alternative splicing of 10 genes.

LOOKING FOR GROUND TRUTH.



Universal Human Reference RNA (UHR)
Lexogen Spike-In RNA Variant Controls (SIRVs)

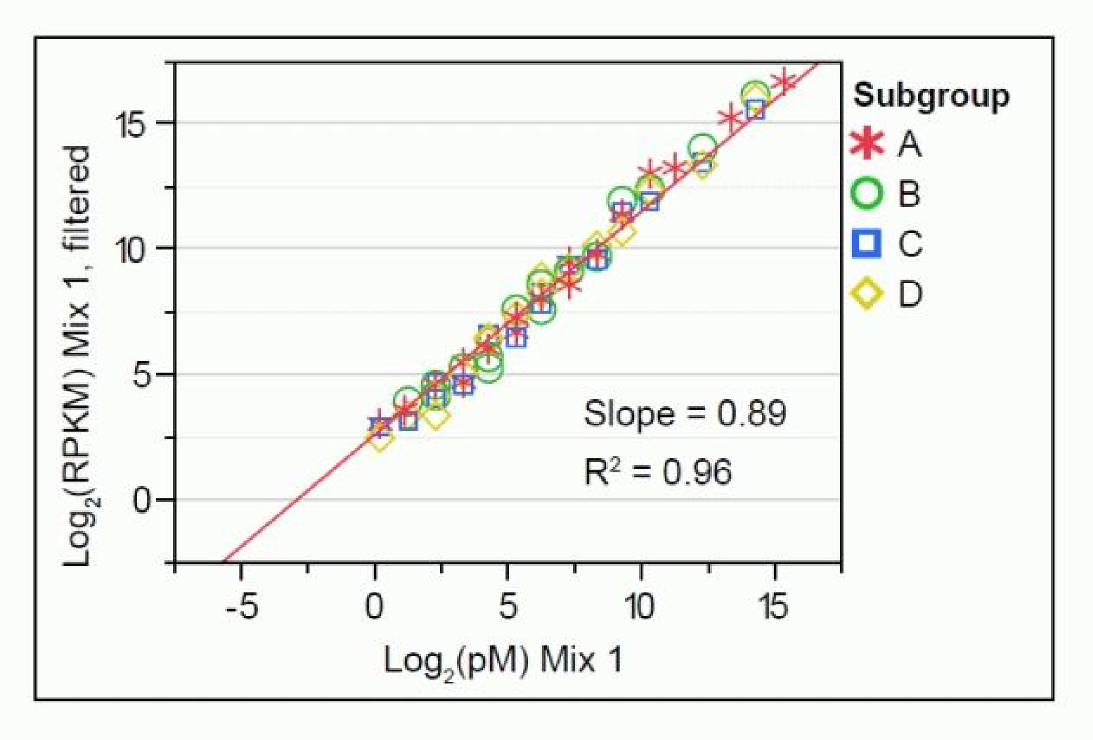
RIN 9.4

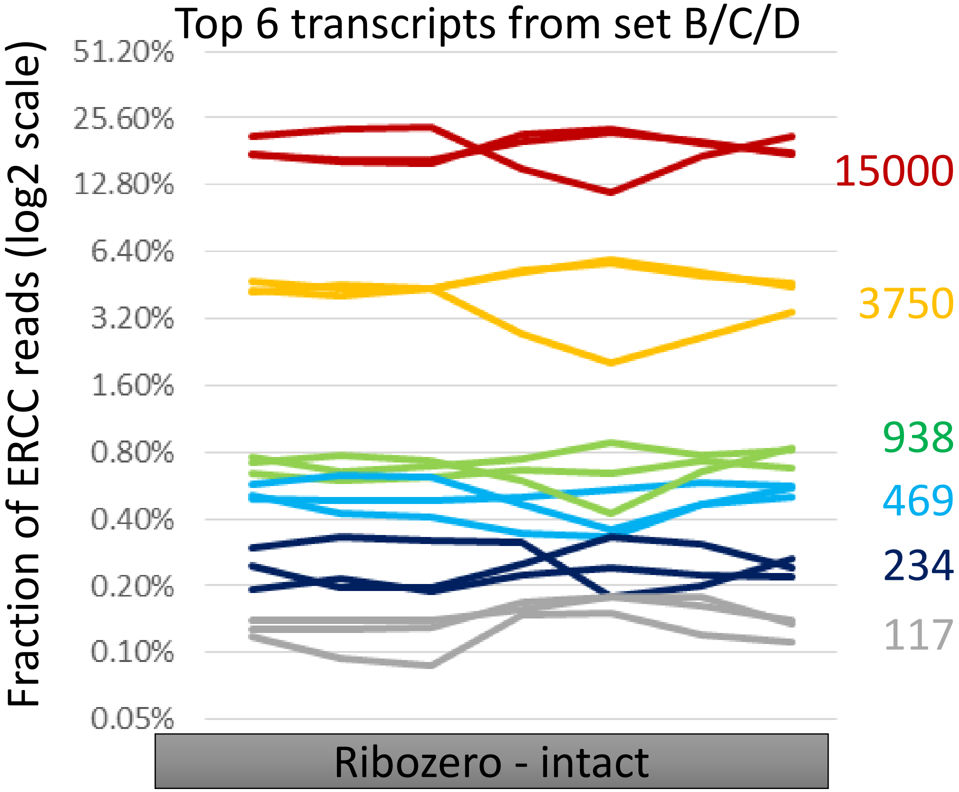
Heat Degradation

RIN 2.1

+ERCCs

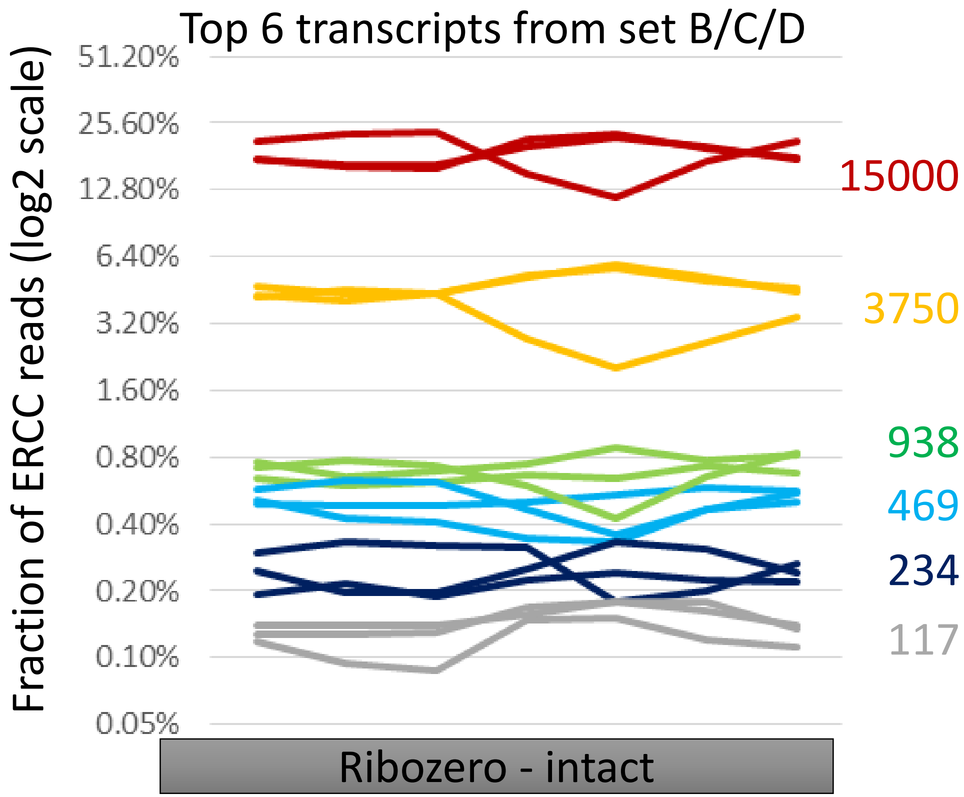
ERCC


Subgroup
A
B
C
D
Slope = 0.89
R^2 = 0.96
Log_2(RPKM) Mix 1, filtered
Log_2(pM) Mix 1
15
10
5
0
-5
0
5
10
15


Top 6 transcripts from set B/C/D
Fraction of ERCC reads (log2 scale)
51.20%
25.60%
12.80%
6.40%
3.20%
1.60%
0.80%
0.40%
0.20%
0.10%
0.05%
15000
3750
938
469
234
117
Ribozero - intact

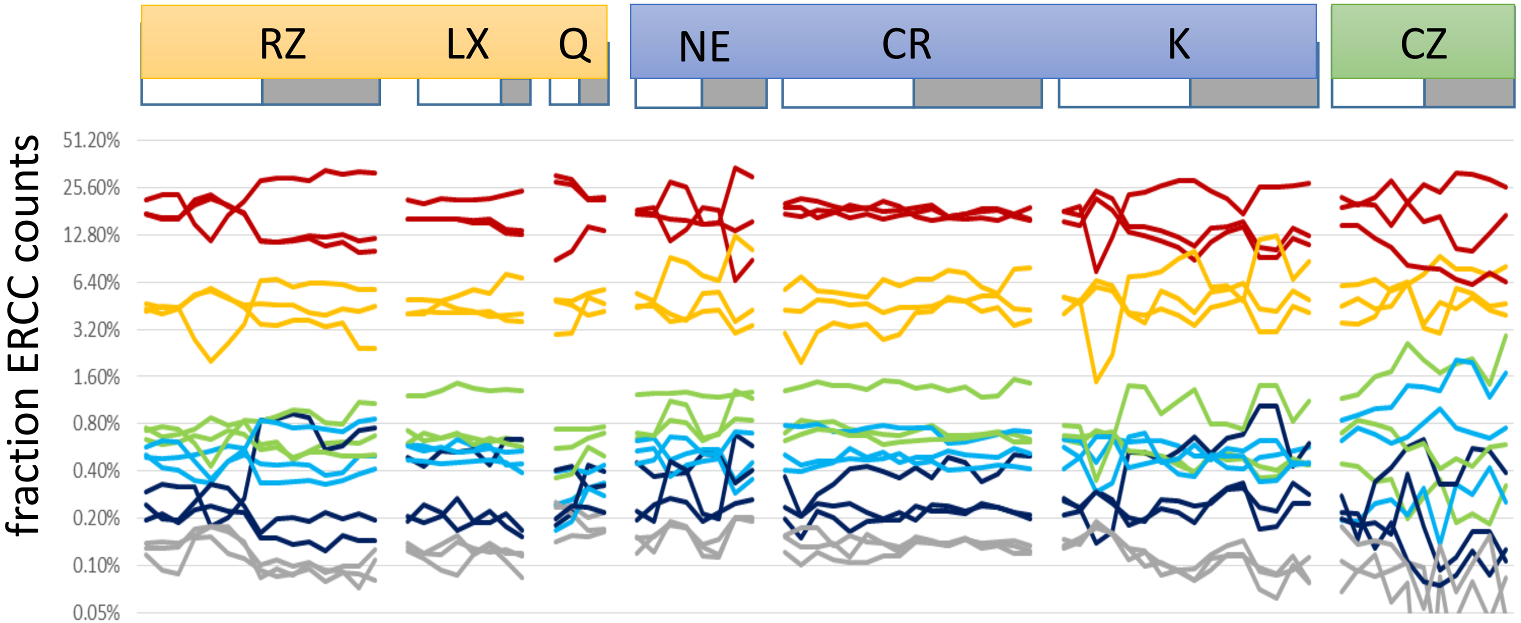
ERCC

- Observing quite large drift in ERCC samples from site to site – even within one kit.

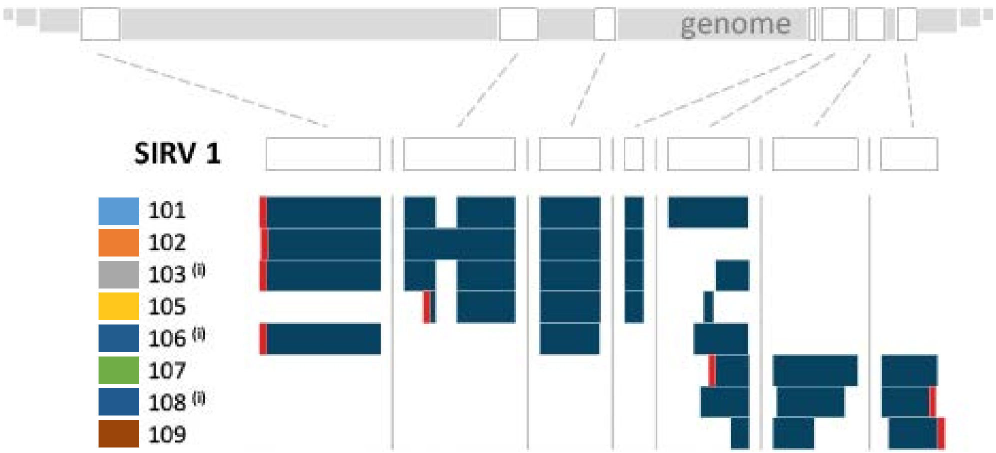


ERCC measurements are highly variable

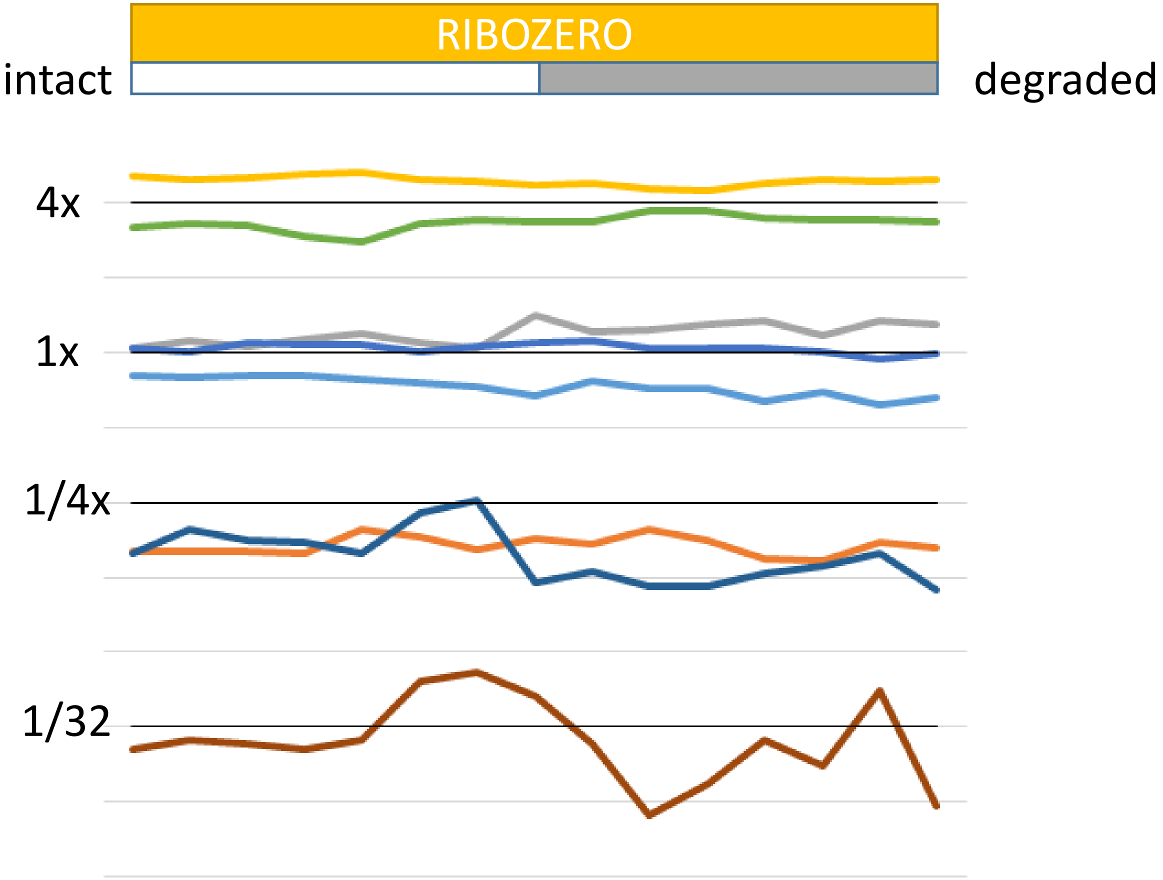
- Observing quite large drift in ERCC samples from site to site – even within one kit.
- Variability is inconsistent. Lexogen and Clontech Ribogone have most even coverage of ERCC spike-ins.
- Definite changes in ratios observed in the degraded samples *even though ERCCS were not degraded*



SIRV1

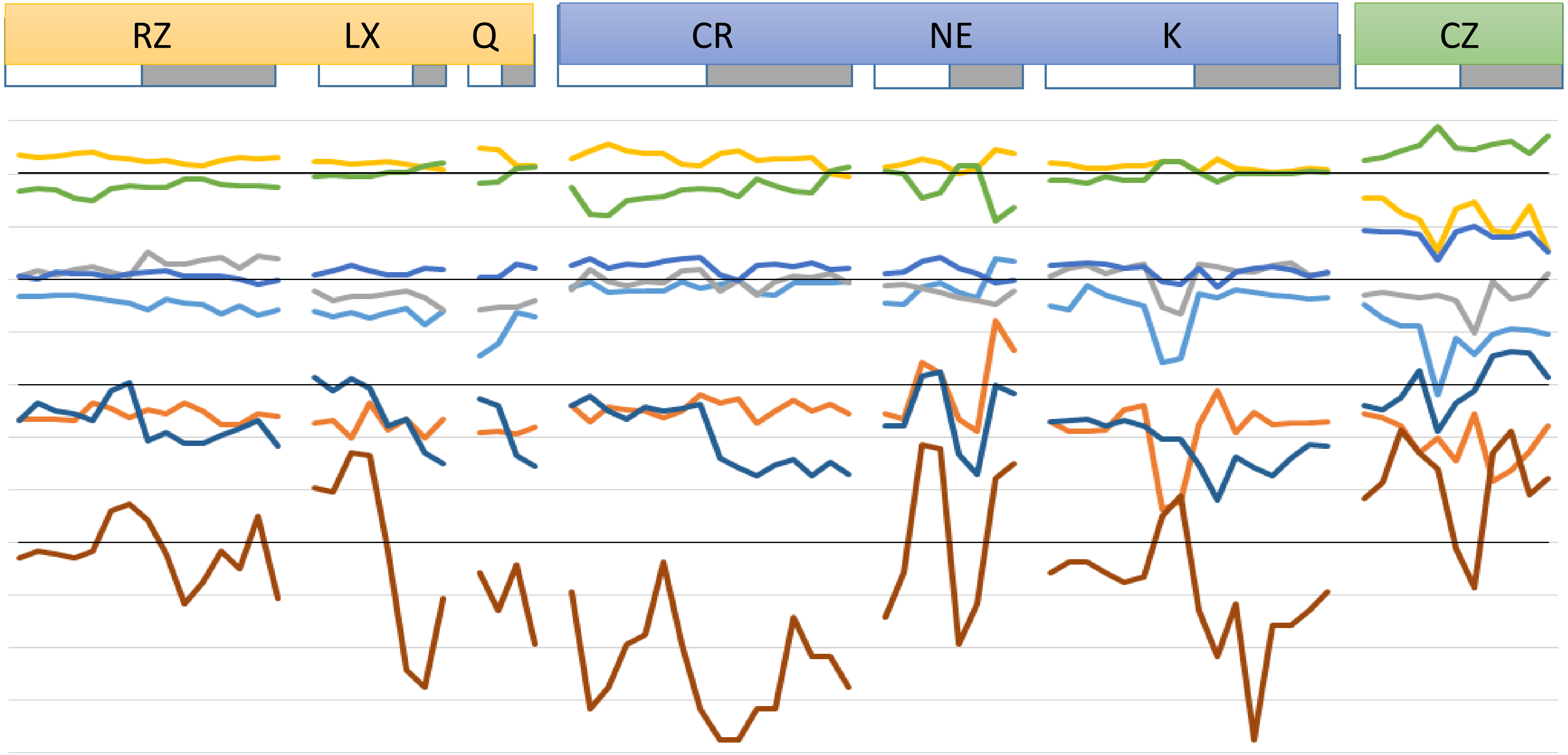


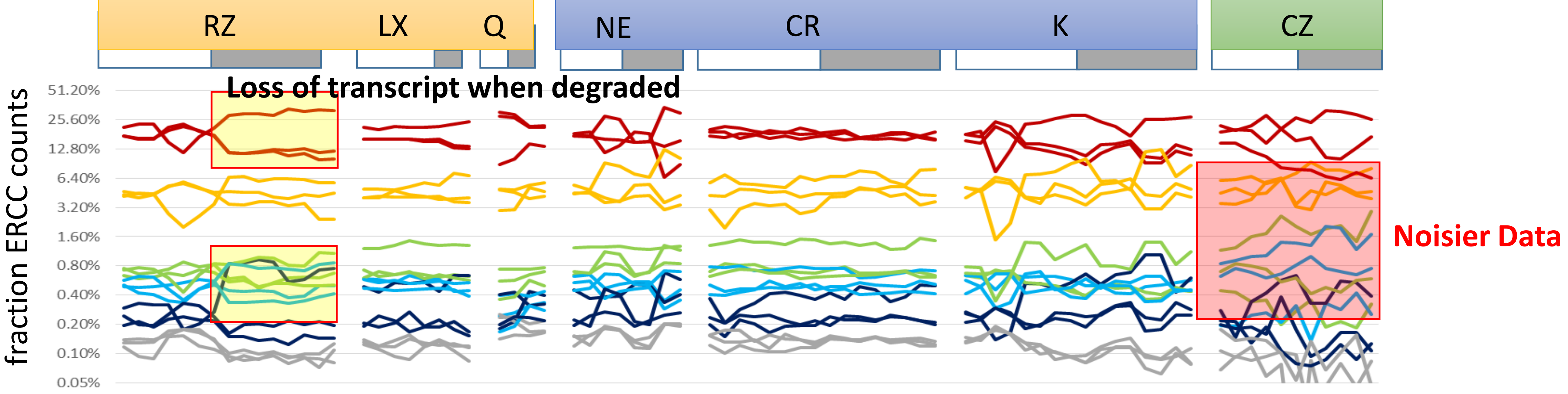
SIRV1



- SIRVs appear to behave better than the ERCCs

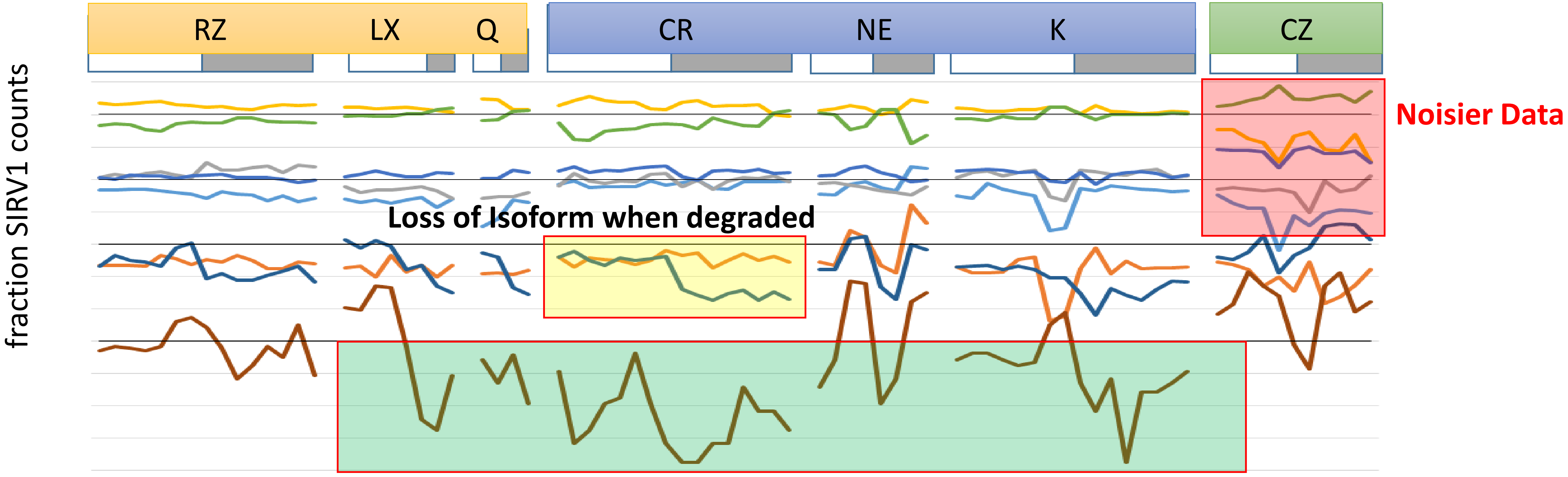
SIRV1





Loss of transcript when degraded

Noisier Data



Noisier Data

Loss of Isoform when degraded

Different Minimum Detected

DNA Sequencing Research Group

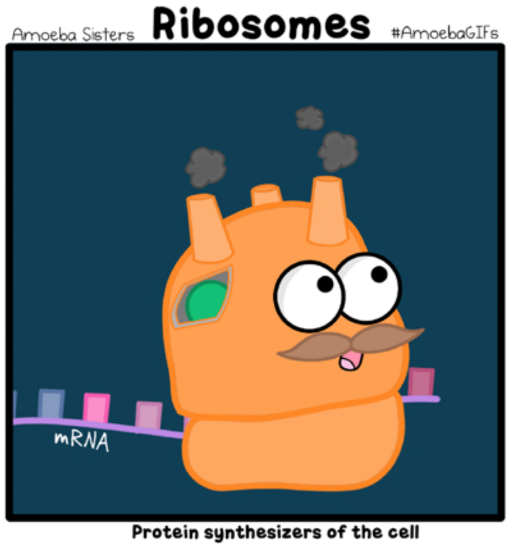
The Association of Biomolecular Resource Facilities

Research • Technology • Communication • Education

Cross-Site Comparison of Ribosomal Depletion Kits for Illumina RNAseq Library Construction

- All kits were capable of reducing rRNAs and produce similar output based on gene count.
- Ribosomal depletion based kits were more challenging to implement and struggled more with degraded samples
- Clontech Pico kit showed stronger biases from low input sampling but worked well at removing rRNA.
- Some evidence Clontech Ribogone works less well on short RNAs, but had some of the best performance on spike-ins

- ERCC spike-ins were less consistent. Lexogen SIRVs show promise as an alternative.
- NEB RNAseq prep kit may have some bias against lincRNAs – still looking at this.



DNA Sequencing Research Group

The Association of Biomolecular Resource Facilities
Research • Technology • Communication • Education

PHASE II

- Analysis here was done using STAR aligner followed by RSEM.
- Annotations based on ENSEMBL.

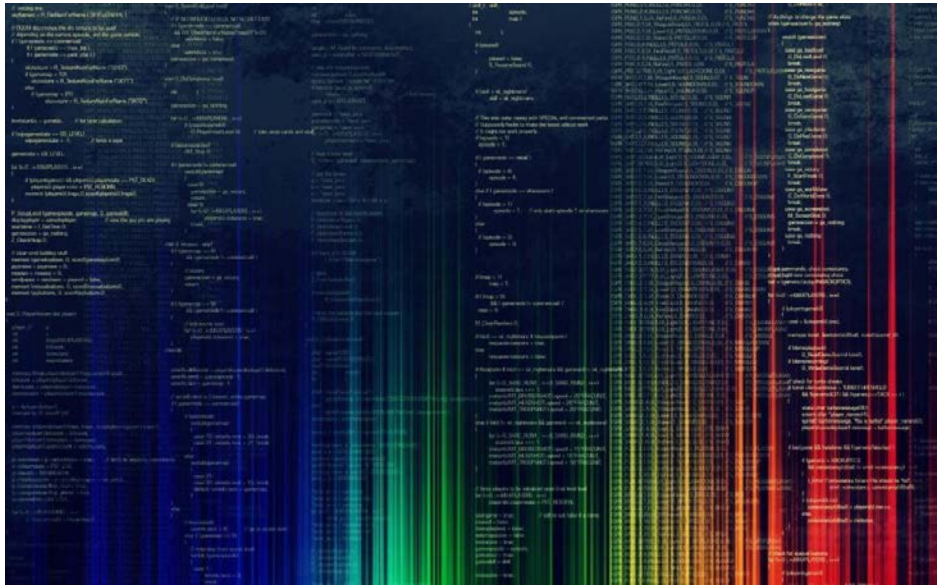


Image from the Price lab

- Work closely with **GBIRG** to use spike-in data on a comparative analysis of RNAseq algorithms and packages.
- Data set provides a unique real world experiment with different chemistries on well understood samples as well both intact and degraded samples.

ACKNOWLEDGEMENTS

ABRF
DNA Sequencing Research Group

- **Marie Adams**, Van Andel Inst. DSRG Co-chair
- **Zach Herbert**, DFCI
- **Jamie Kershner**, U.Colorado
- **Jyothi Thimmapuram**, Purdue U.
- **Jun Fan**, Marshall U.
- **Yuriy Alekseyev**, Boston U.
- **Jessica Podnar**, UT Austin

- **Savita Shanker**, U. Florida
- **Charlie Nicolet**, USC, Ad Hoc
- **Jeanne Geskes**, Illumina
- **Joanna Hamilton**, Roche / Kapa

Additional Collaborators

- **Krystalynne Morris**, UNH
- **Edward Wilcox**, BYU
- **Allison Gillaspy**, U. Oklahoma
  Jenny Gipson, U. Oklahoma
- **Kristen Jepsen**, UCSD
- **Maura Berkeley,** DFCI
- **Leslie Grimmett,** DFCI
- **Gary Sommerville,** DFCI

- **Vincent Butty**, MIT

The other 36 sites that volunteered to be a part of this study!

Corporate Partners

illumina
LEXOGEN
QIAGEN®
TaKaRa
Clontech TaKaRa cellartis
NEW ENGLAND BioLabs® Inc.
KAPA BIOSYSTEMS

ABRF 2017

ANNUAL MEETING

SAN DIEGO
CA, USA

MARCH 25 – 28, 2017 | TOWN AND COUNTRY RESORT & CONVENTION CENTER

A FORUM FOR ADVANCING TODAY'S CORE TECHNOLOGIES TO ENABLE TOMORROW'S INNOVATIONS