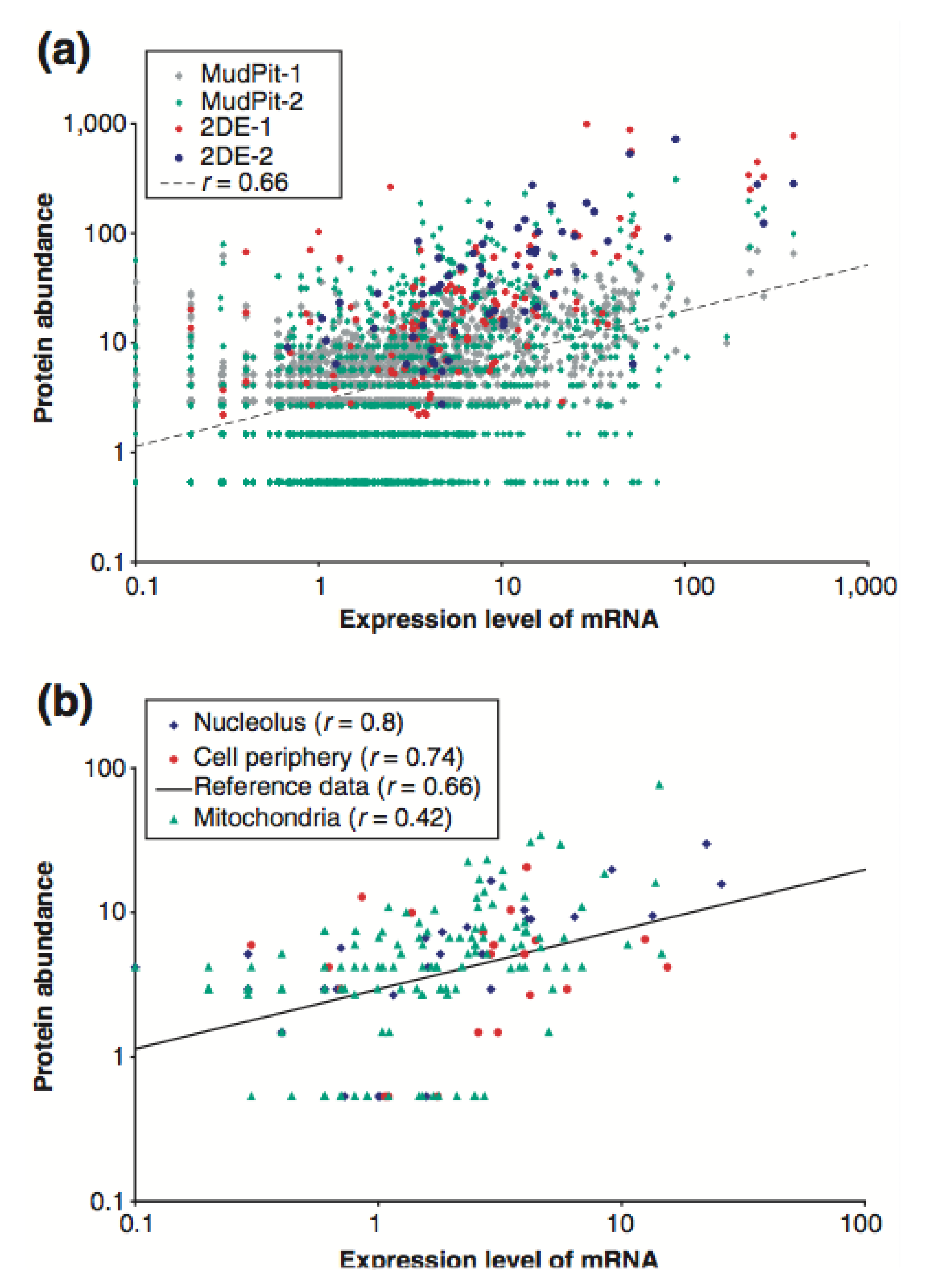
RNA-seq + proteomics

Toward Improved Label-free Quantification of MS/MS Spectra Using Predicted Proteomes Based on 2nd generation RNA-Sequencing

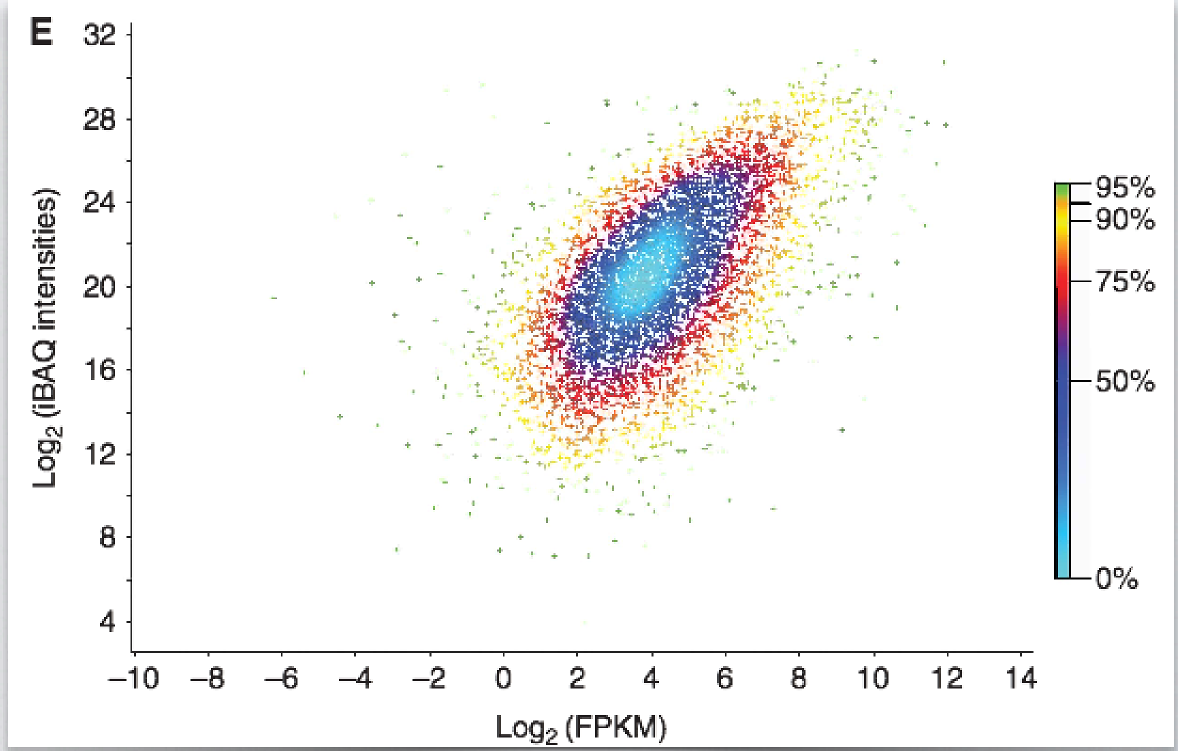
Rob Kitchen
Yale
2013-04-26

mRNA -- protein correlation



Greenbaum et al. 2003 Genome Biol

- ~60% correlation over all genes
- can increase within some functional categories



Nagaraj et al. 2011 Mol Sys Biol

peptide detection

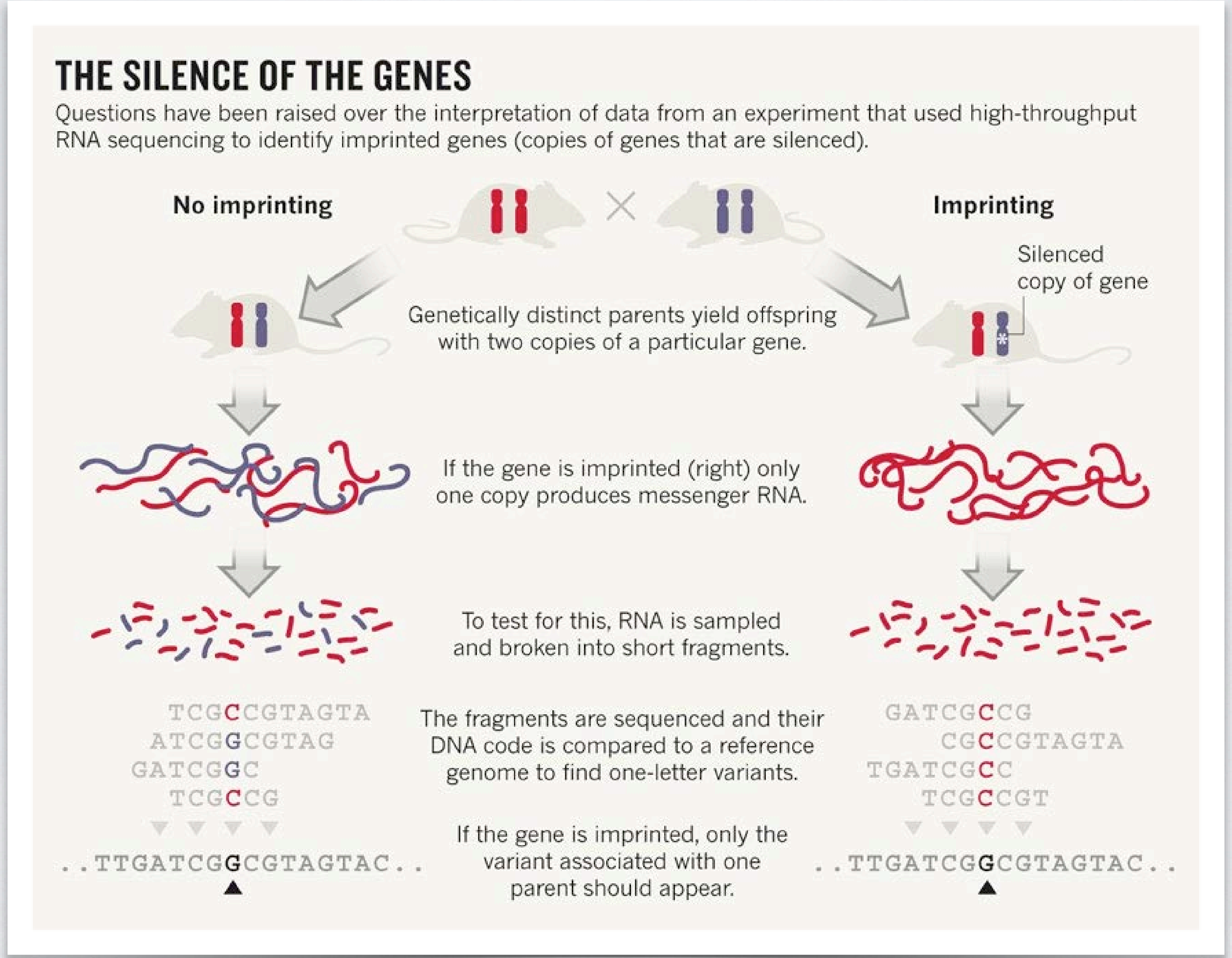
- 20,248 reviewed human proteins in UniProtKB/Swiss-Prot DB

- 62,901 predicted human proteins in UniProtKB/TrEMBL DB

- 20,843 observed human proteins in YPED (243,749 peptides)

Instrument	N spectra	N unique peptide assignments	proteins with unique peptide assignments
5600TT	51,747	5,349	1,424
Orbi	40,195	7,565	1,916
Orbi - fractionated	256,702	33,388	6,177

imprinting / allele-specific exp.



THE SILENCE OF THE GENES

Questions have been raised over the interpretation of data from an experiment that used high-throughput RNA sequencing to identify imprinted genes (copies of genes that are silenced).

Genetically distinct parents yield offspring with two copies of a particular gene.

If the gene is imprinted (right) only one copy produces messenger RNA.

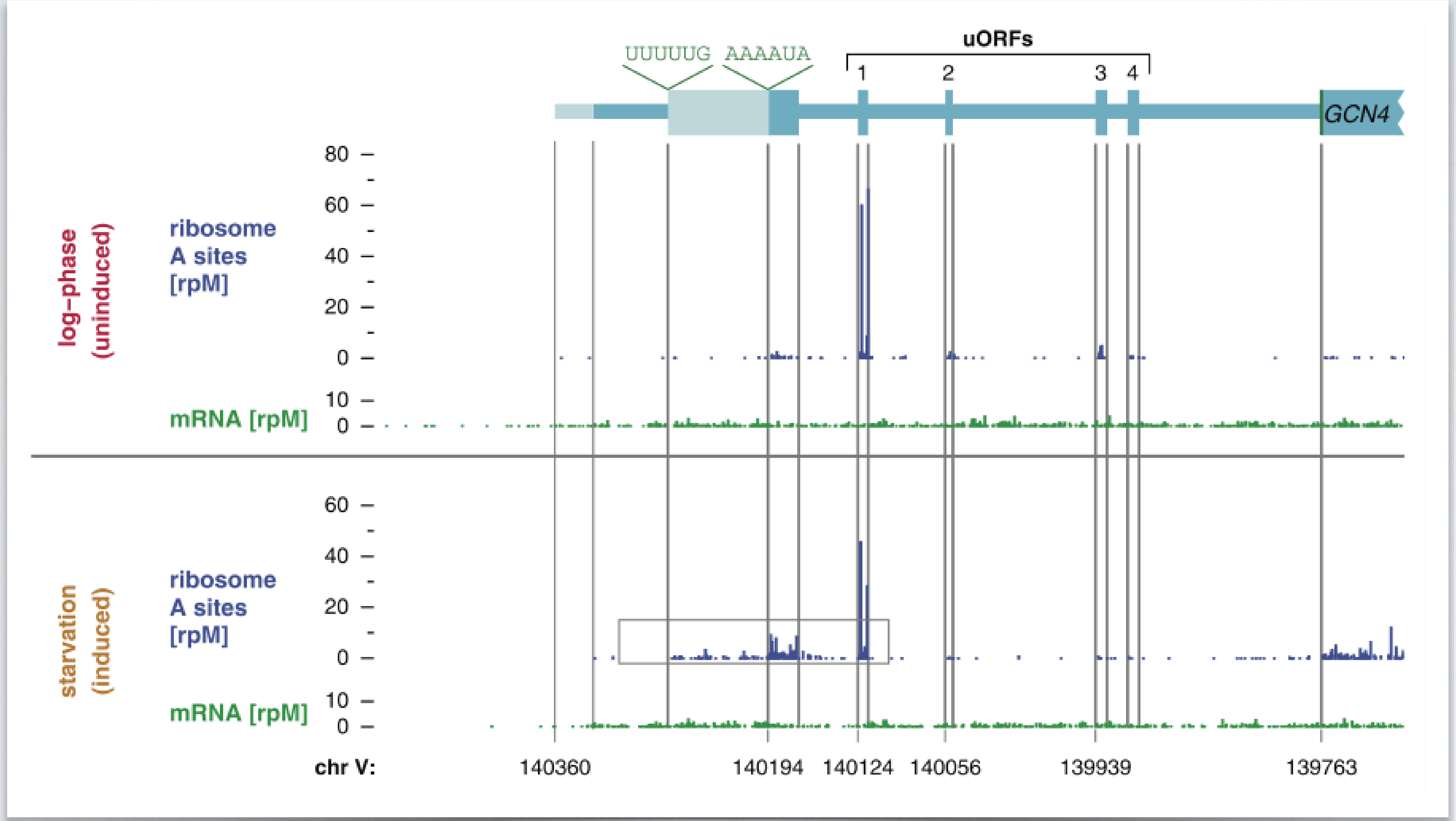
To test for this, RNA is sampled and broken into short fragments.

The fragments are sequenced and their DNA code is compared to a reference genome to find one-letter variants.

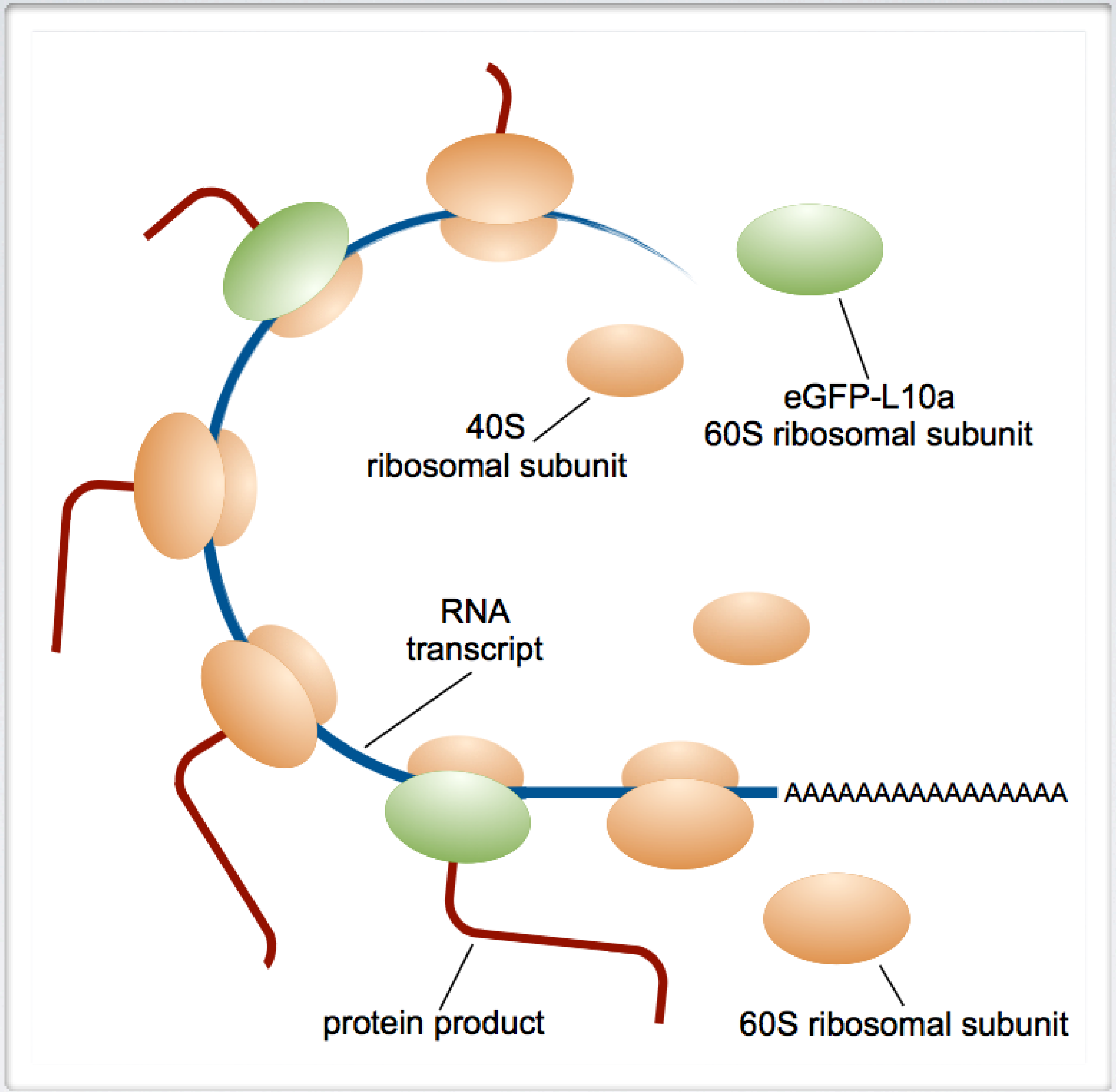
If the gene is imprinted, only the variant associated with one parent should appear.

Erika Hayden 2012 Nature (news)

novel translation products

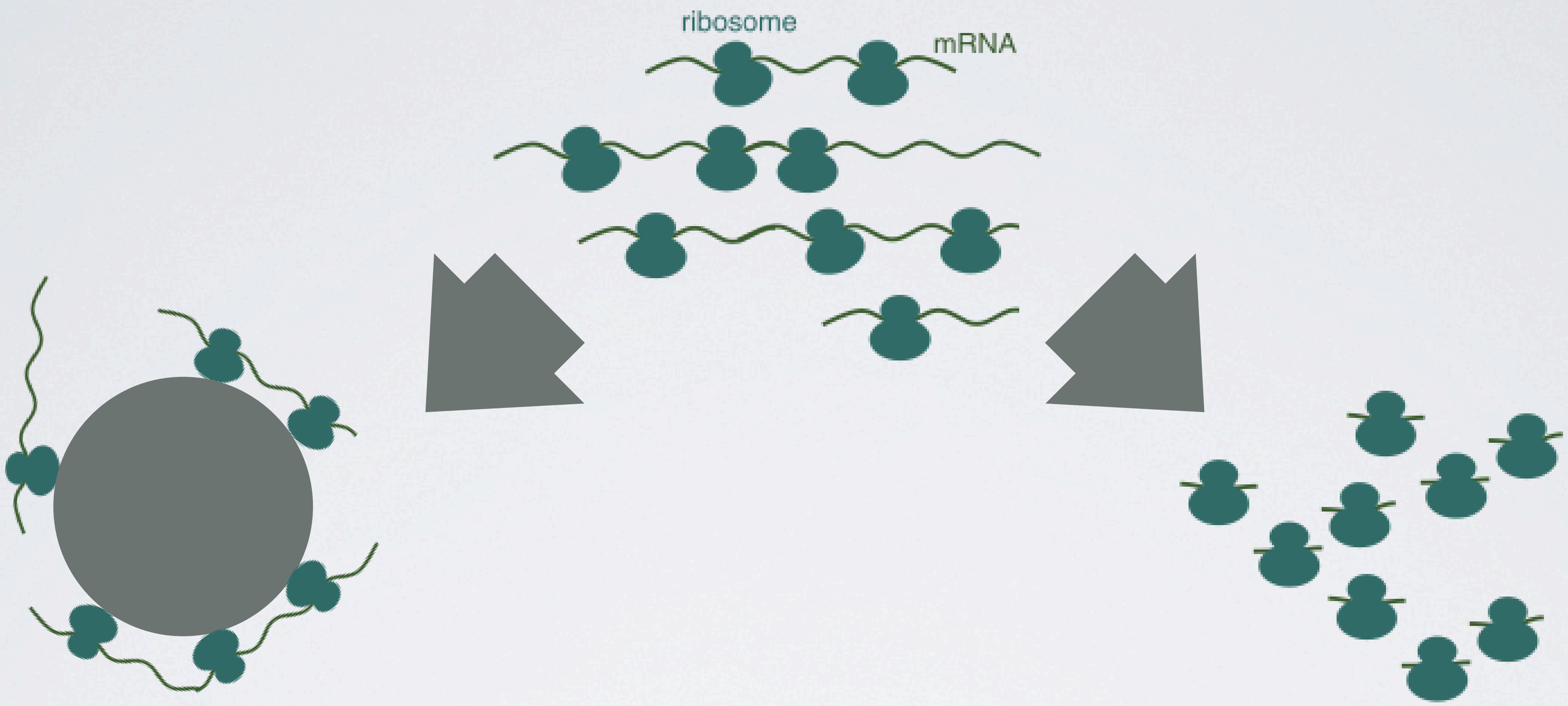


Ingolia et al. 2009
Science



profiling translation

state of the art



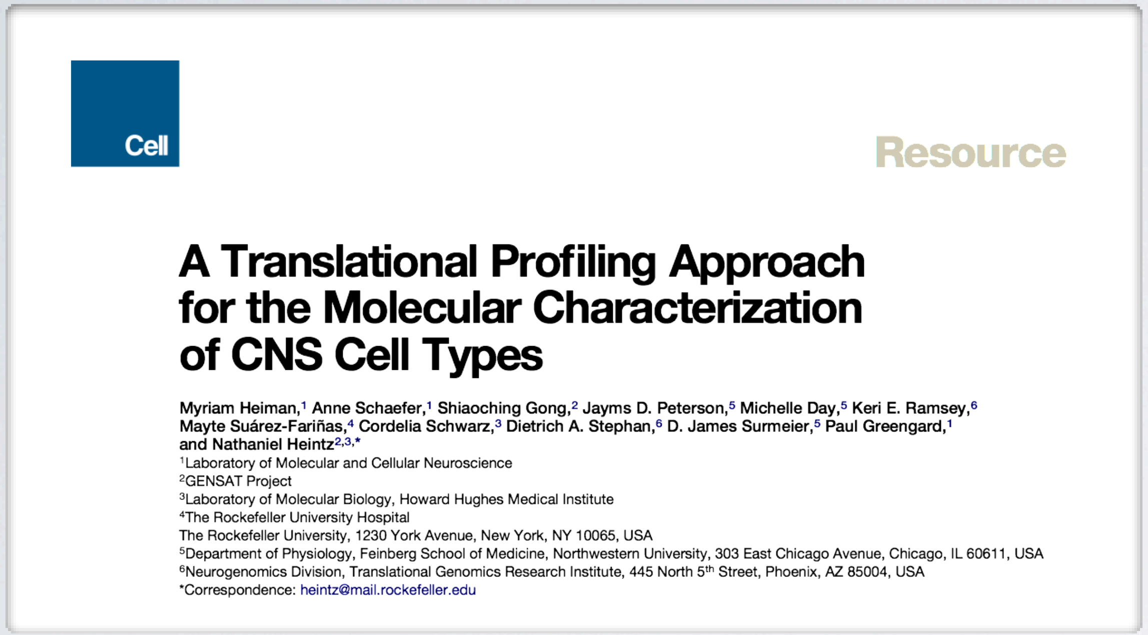
IP

- use antibodies against either exogenous or endogenous ribosome components

footprinting

- nuclease-treatment degrades RNA except ribosome-protected fragments

state of the art

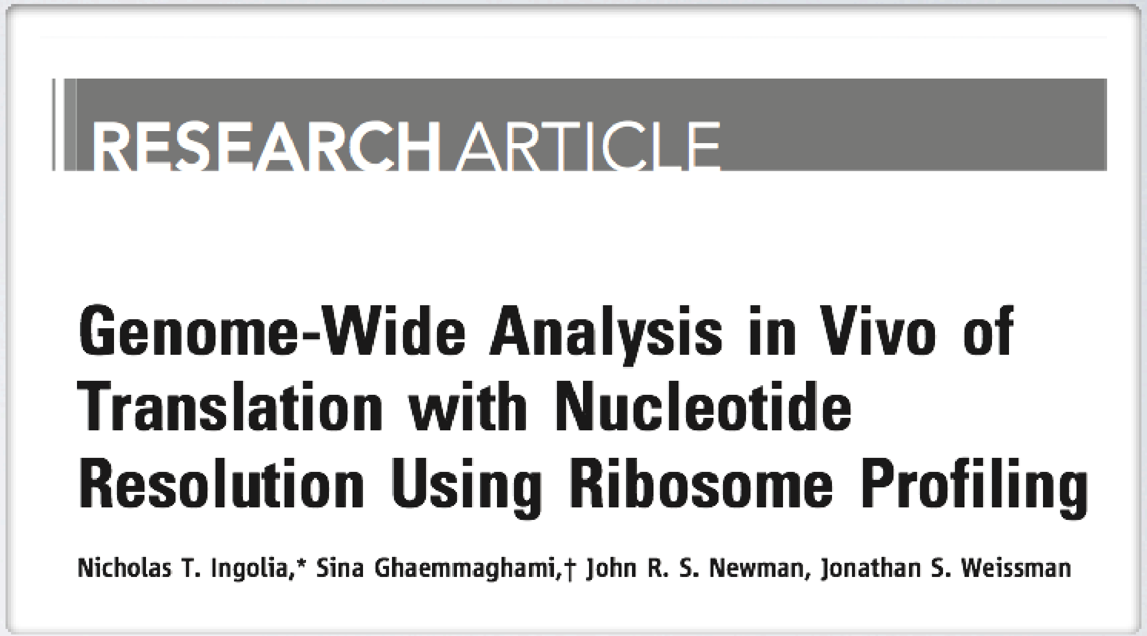


Cell

Resource

A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types

Myriam Heiman,[1] Anne Schaefer,[1] Shiaoching Gong,[2] Jayms D. Peterson,[5] Michelle Day,[5] Keri E. Ramsey,[6]
Mayte Suárez-Fariñas,[4] Cordelia Schwarz,[3] Dietrich A. Stephan,[6] D. James Surmeier,[5] Paul Greengard,[1]
and Nathaniel Heintz[2,3,*]
[1]Laboratory of Molecular and Cellular Neuroscience
[2]GENSAT Project
[3]Laboratory of Molecular Biology, Howard Hughes Medical Institute
[4]The Rockefeller University Hospital
The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA
[5]Department of Physiology, Feinberg School of Medicine, Northwestern University, 303 East Chicago Avenue, Chicago, IL 60611, USA
[6]Neurogenomics Division, Translational Genomics Research Institute, 445 North 5th Street, Phoenix, AZ 85004, USA
*Correspondence: heintz@mail.rockefeller.edu



RESEARCH ARTICLE

Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling

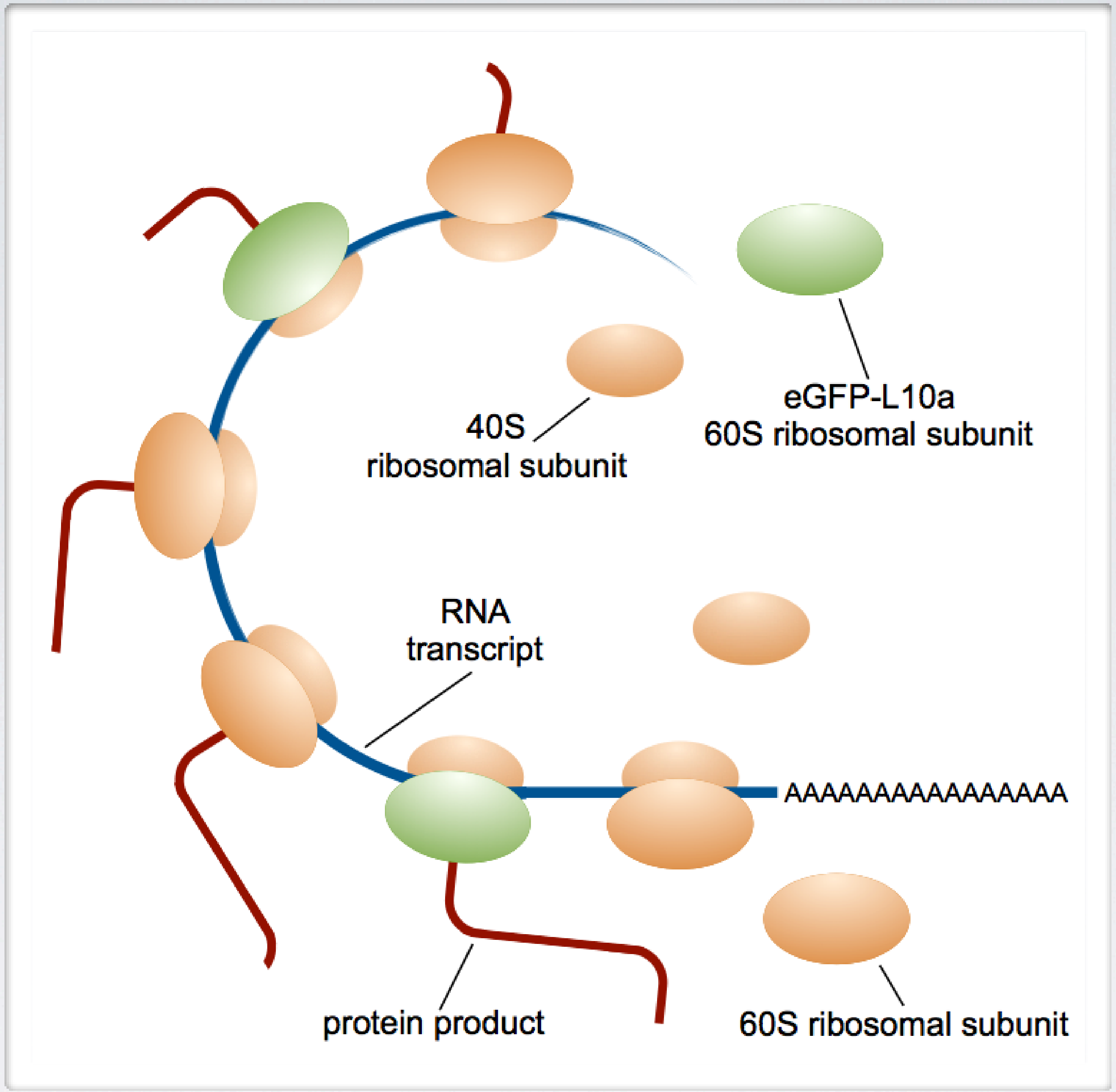
Nicholas T. Ingolia,* Sina Ghaemmaghami,† John R. S. Newman, Jonathan S. Weissman

immunoprecipitation

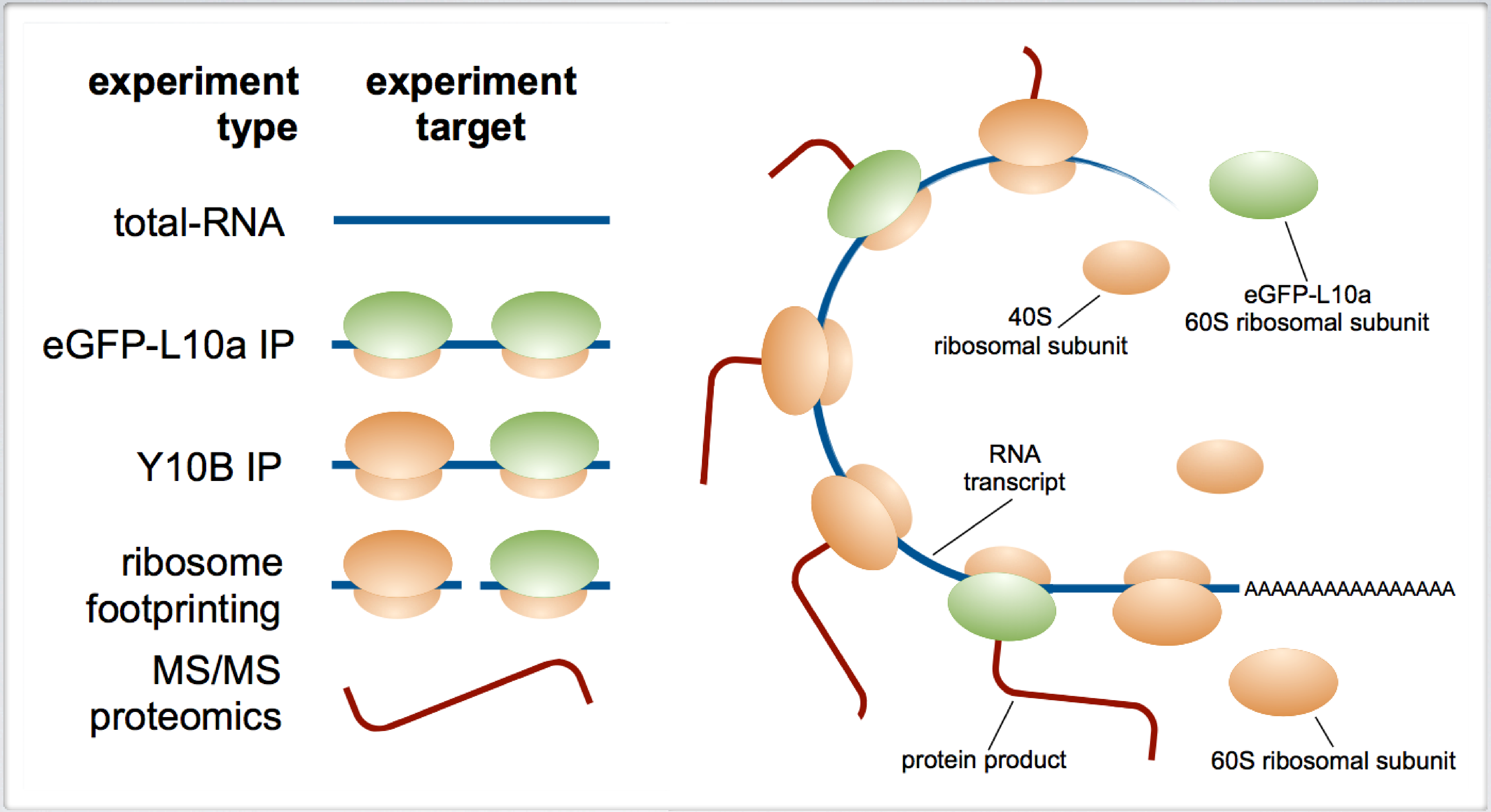
- ✔ can assess splicing
- ✔ can be cell-type specific
- ✔ accurate expression est.
- ✘ antibody specificity?
- ✘ RNA cross-reactivity?

footprinting

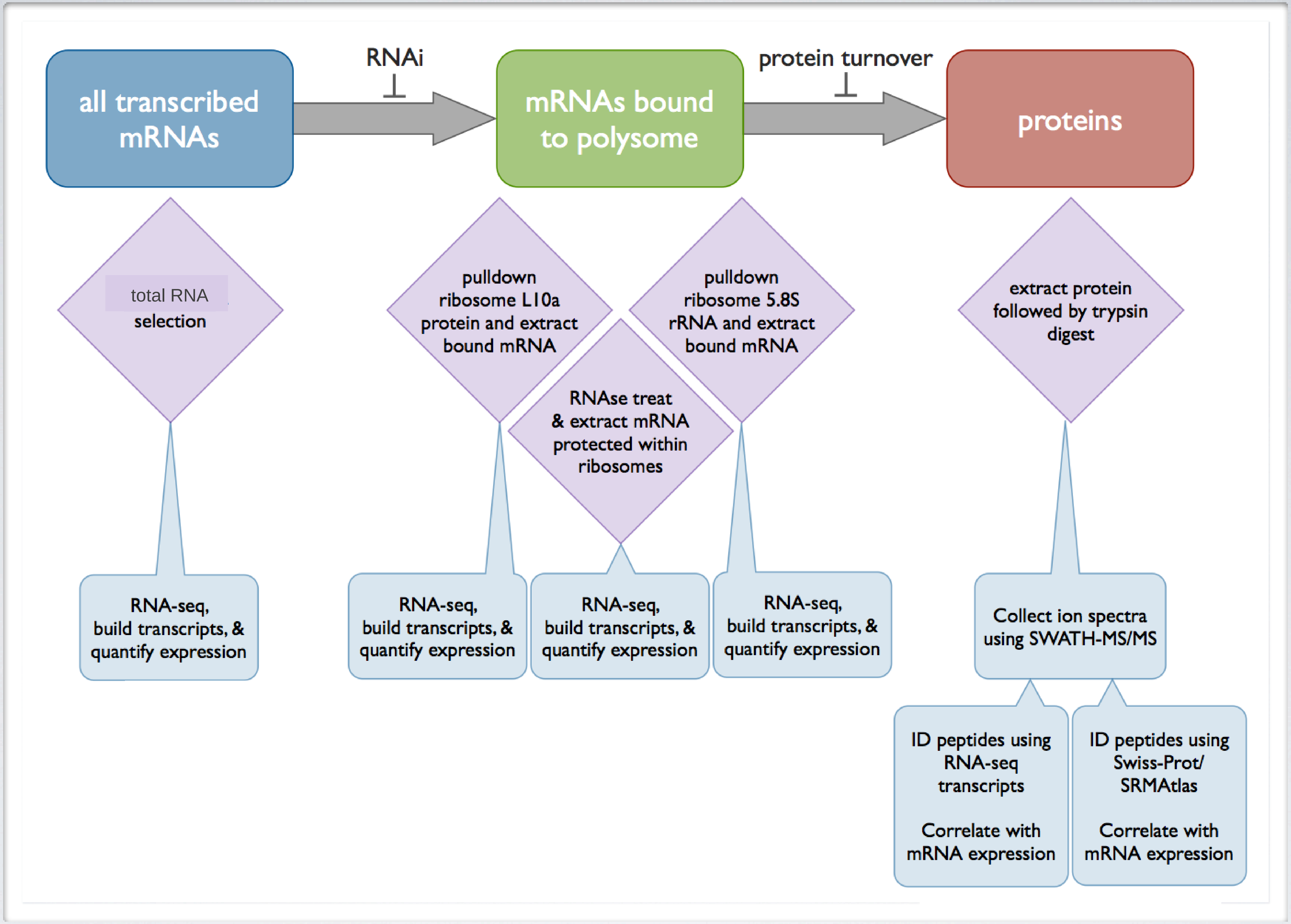
- ✔ precisely defines CDS
- ✔ measure translation rate
- ✘ short RNA fragments



profiling translation



profiling translation



all transcribed mRNAs

RNAi

mRNAs bound to polysome

protein turnover

proteins

total RNA selection

pulldown ribosome L10a protein and extract bound mRNA

pulldown ribosome 5.8S rRNA and extract bound mRNA

RNAse treat & extract mRNA protected within ribosomes

extract protein followed by trypsin digest

RNA-seq, build transcripts, & quantify expression

RNA-seq, build transcripts, & quantify expression

RNA-seq, build transcripts, & quantify expression

RNA-seq, build transcripts, & quantify expression

Collect ion spectra using SWATH-MS/MS

ID peptides using RNA-seq transcripts

Correlate with mRNA expression

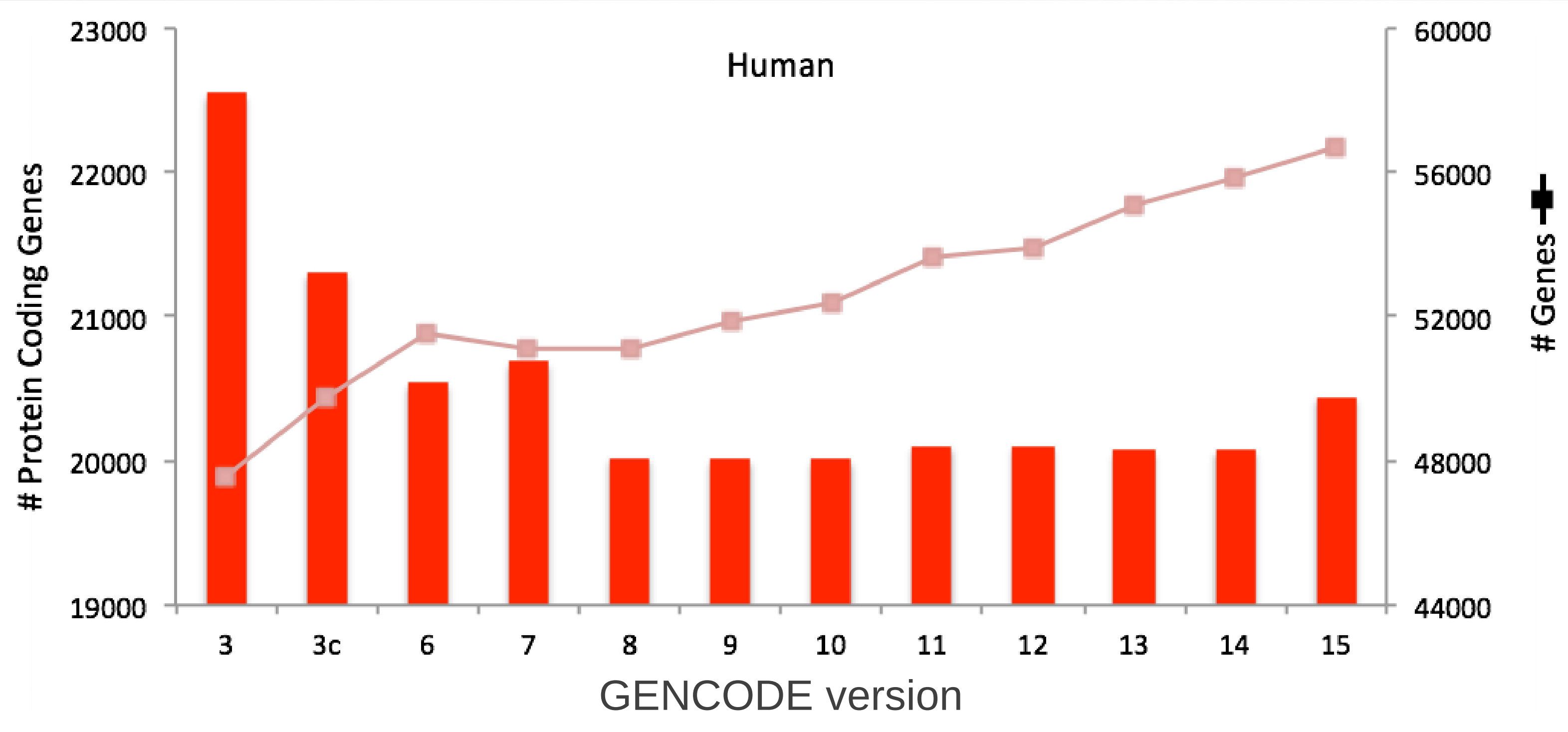
ID peptides using Swiss-Prot/ SRMAtlas

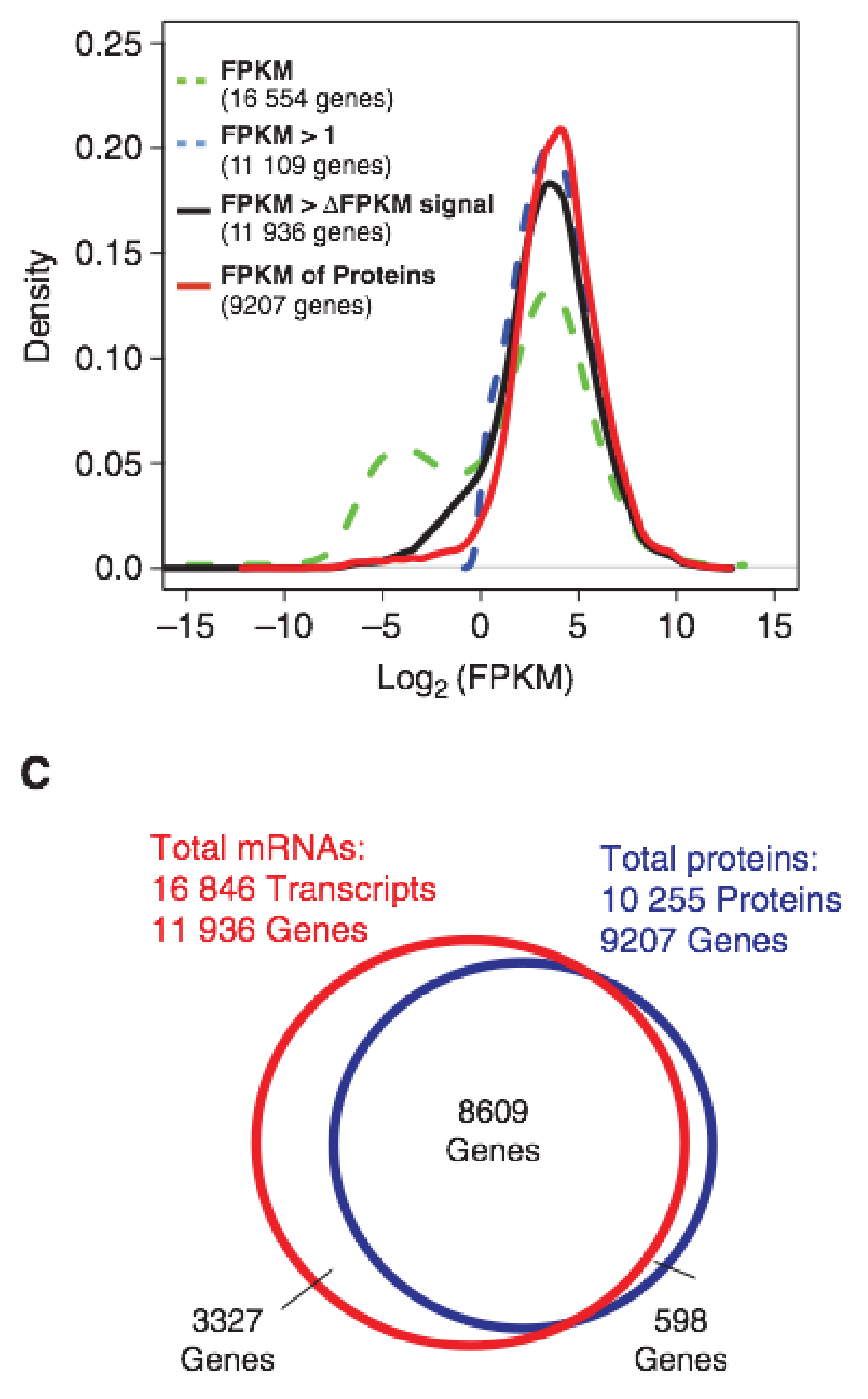
Correlate with mRNA expression

experiment design

choice of comparison set...

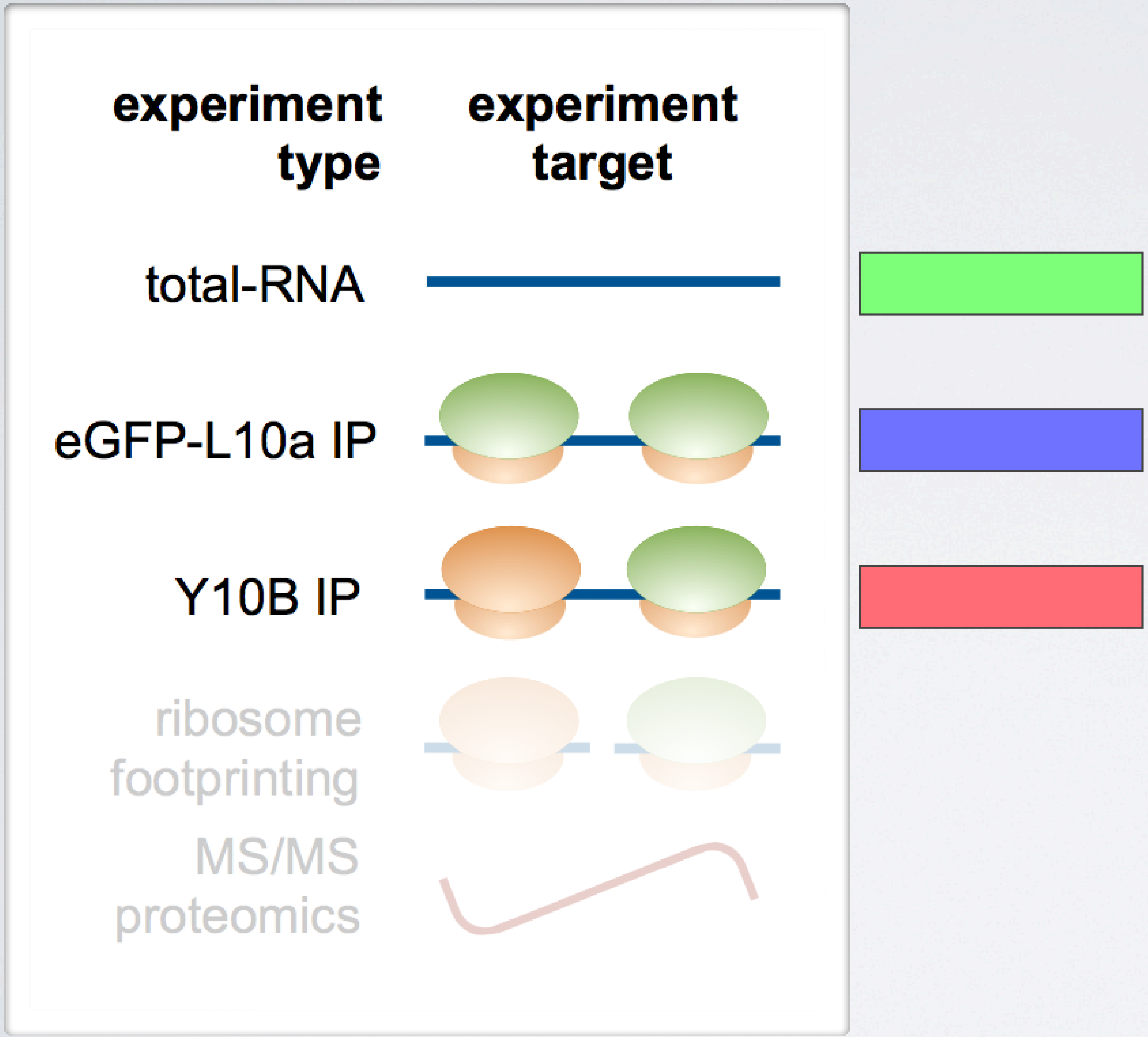
- ... is important!
- no reason to limit analysis to mRNA



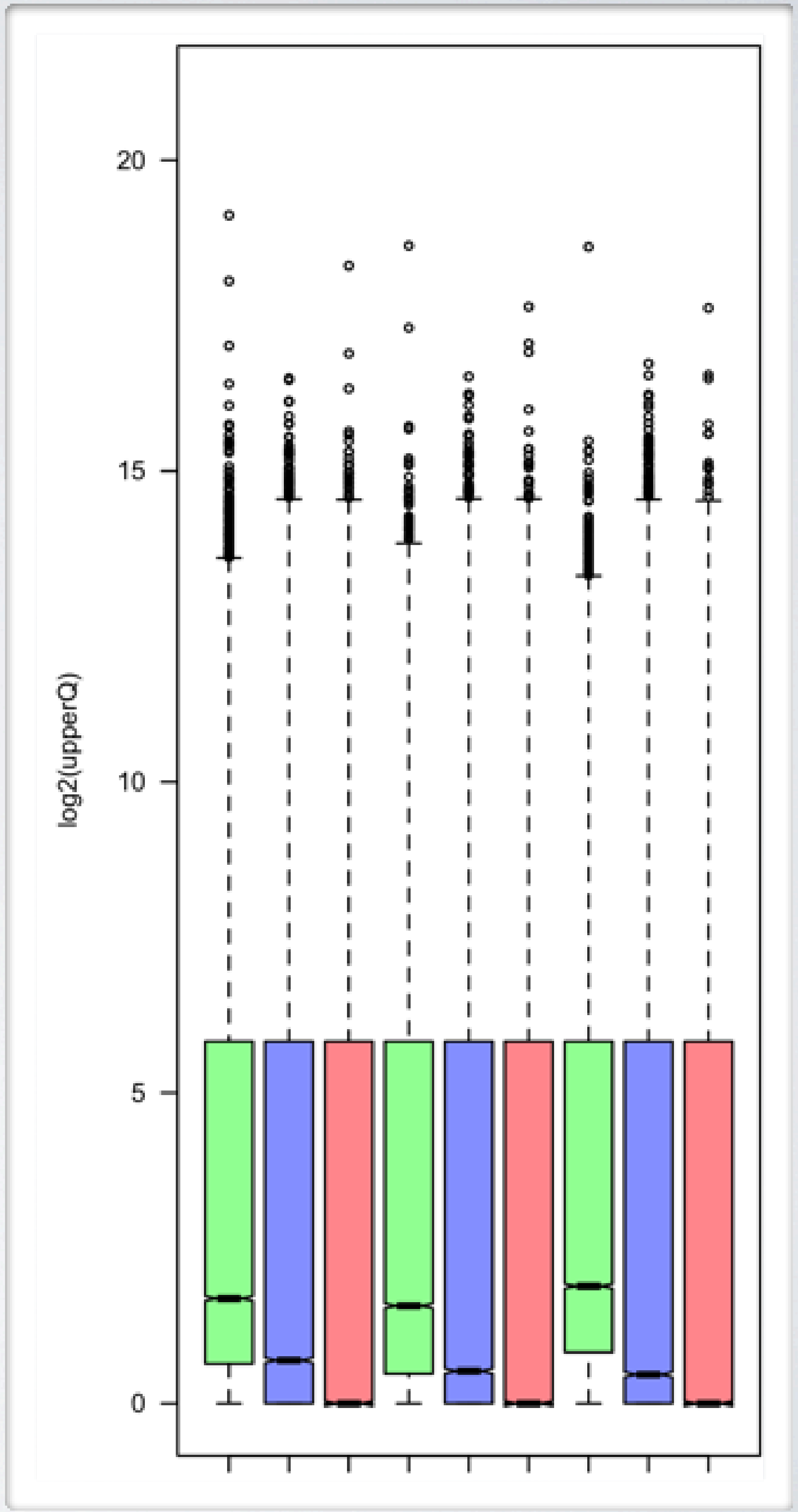


Nagaraj et al. 2011 Mol Sys Biol

expression distributions

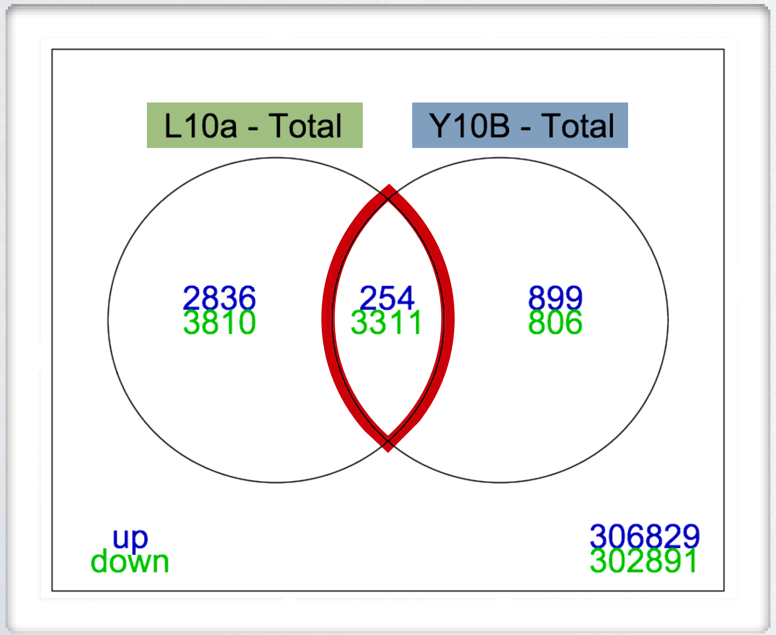


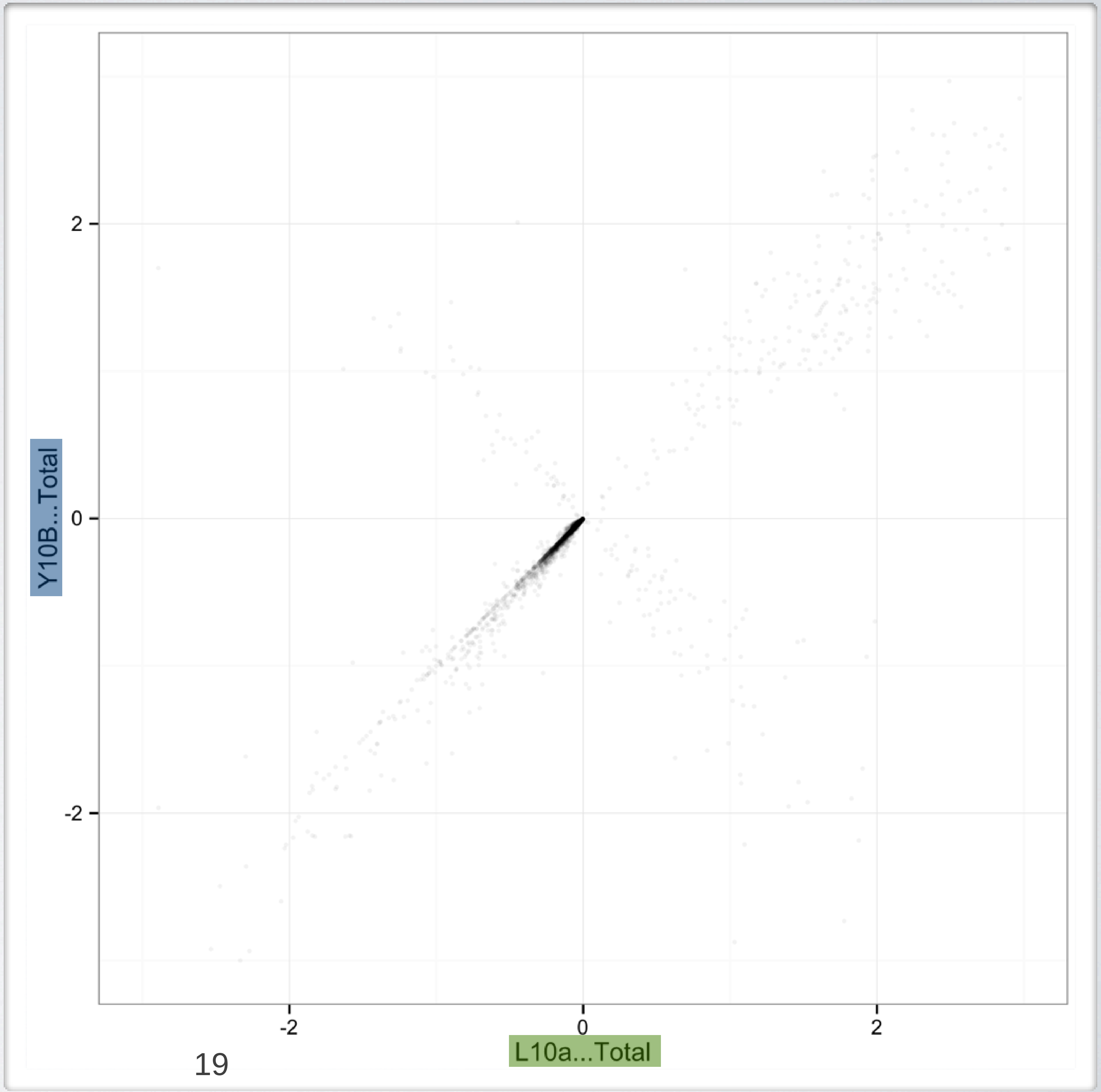
experiment type	experiment target	
total-RNA		🟩
eGFP-L10a IP		🟦
Y10B IP		🟥
ribosome footprinting		
MS/MS proteomics		



differential expression

- intersection of transcripts called differentially expressed between totalRNA and the two IPs:

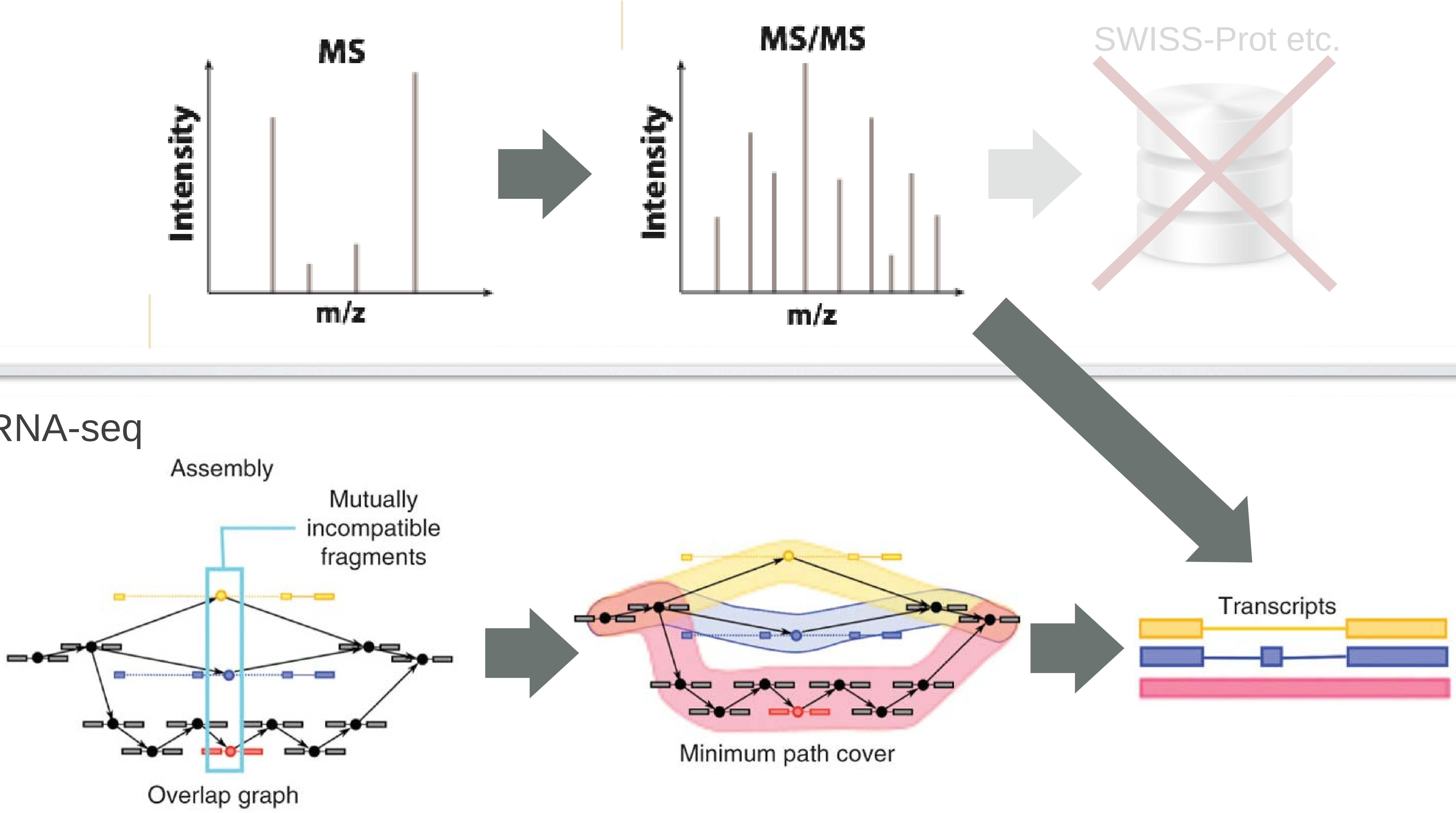




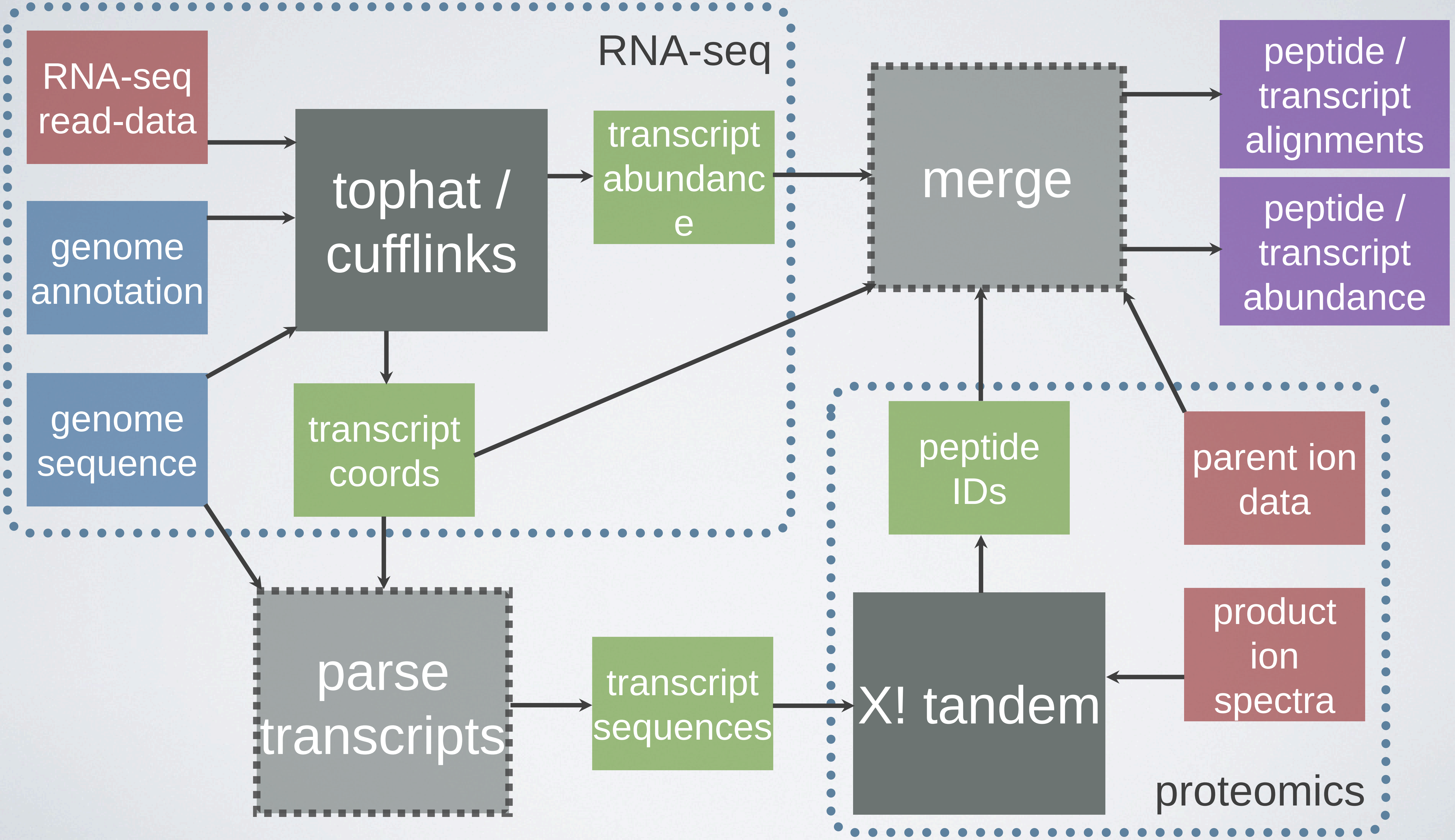
integrated analysis

label-free MS/MS

RNA-seq



integrated analysis -- detail



RNA-seq

RNA-seq read-data

genome annotation

genome sequence

tophat / cufflinks

transcript abundance

transcript coords

parse transcripts

transcript sequences

merge

peptide / transcript alignments

peptide / transcript abundance

peptide IDs

parent ion data

X! tandem

product ion spectra

proteomics

neuroproteomics application

human

6 subjects

chimp

5 subjects

macaque

5 subjects

sequenced mRNA and miRNA from 16 regions of each brain:

Frontal lobe
DFC Dorsolateral prefrontal cortex
M1C Primary motor cortex
MFC Medial prefrontal cortex
OFC Orbital prefrontal cortex
VFC Ventrolateral prefrontal cortex

Temporal lobe
A1C Primary auditory cortex

Parietal lobe
IPC Posterior inferior parietal cortex
S1C Primary sensory cortex

Temporal lobe
A1C Primary auditory cortex
ITC Inferior temporal cortex
STC Posterior superior temporal cortex

Sub-cortex
AMY Amygdala
CBC Cerebellum
HIP Hippocampus
MD Thalamus
STR Striatum

acknowledgements

Nairn lab, Yale
Angus Nairn
Becky Carlyle

Gerstein lab, Yale
Mark Gerstein
Joel Rozowsky

Proteomics, Yale
Ken Williams
Tu Lam
Chris Colangelo
Kathy Stone
Tom Abbott

Šestan lab, Yale
Nenad Šestan
Yuka Kawasawa
Mingfeng Li

Steitz lab, Yale
Joan Steitz
Demian Cazalla

YCGA, Yale
Shrikant Mane
John Overton

Allen Brain Institute
Ed Lein