

miRNAs variants accuracy with sequencing platforms

Lorena Pantano, P.h. D

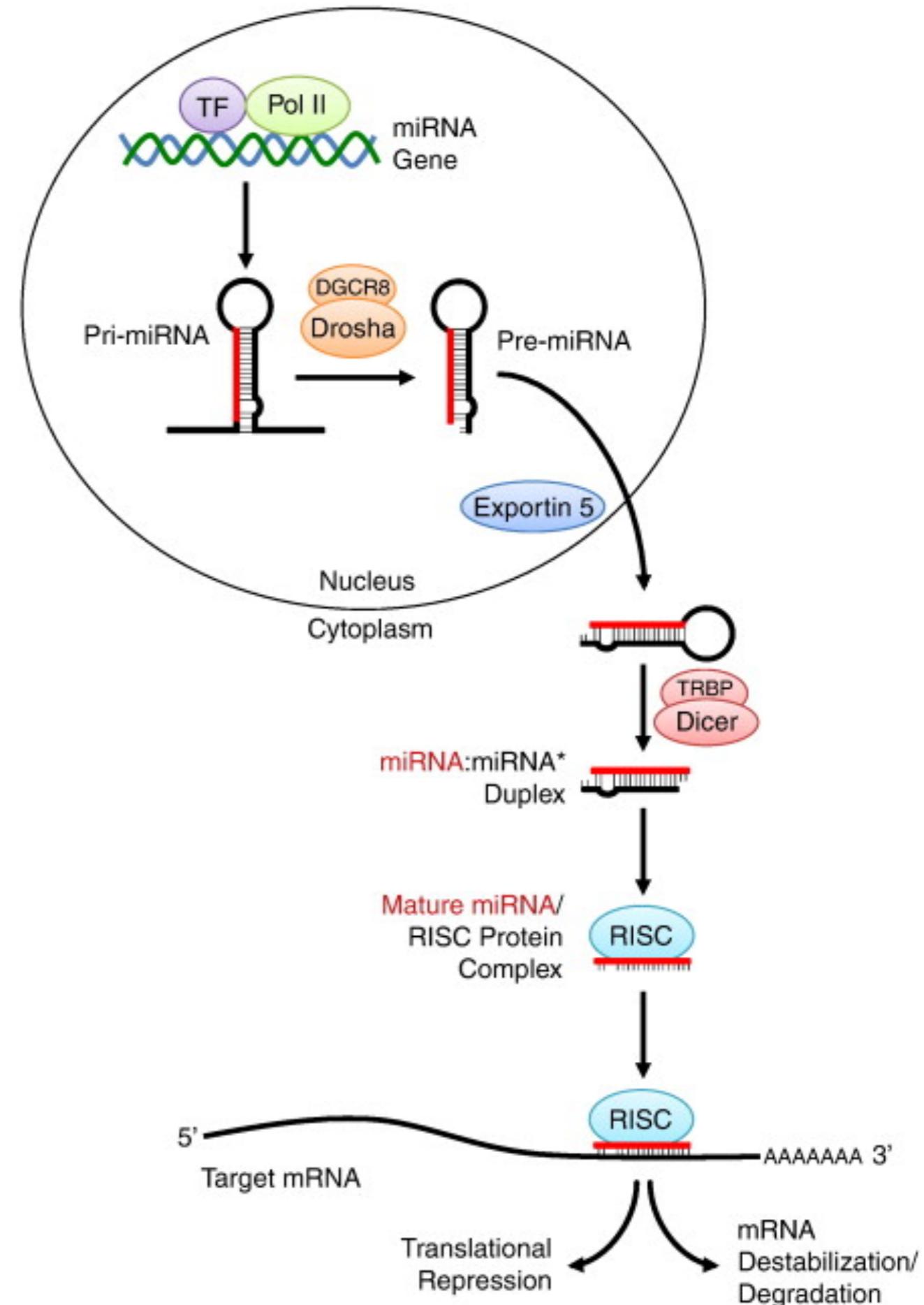
<https://lpantano.github.io>

04-02-2019

@lopantano

miRNAs

Gene regulation by imperfect complementary between seed region in miRNA and 3'UTR in the targeted RNA molecule.





Canonical 7mer-1A site

- Presence of adenine opposite base 1 of miRNA is conservative and improves degree of gene silencing
- 1A is thought to be recognized by RISC



Canonical 7mer-m8 site

- Additional Watson Crick pairing at position 8
- Most abundant type of sites considering only those targeted by highly conserved miRNAs



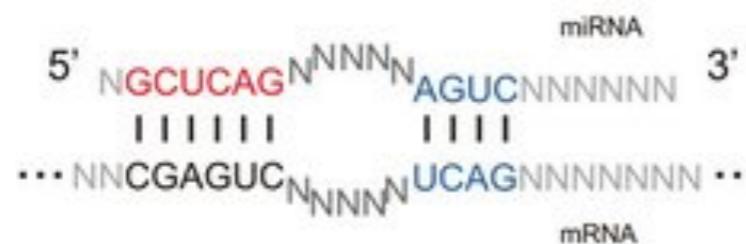
Canonical 8mer site

- Presence of adenine opposite base 1 of miRNA is conservative
- Additional Watson Crick pairing at position 8



6mer site

- Their conservation is only slightly above signal-to-noise ratio
- Generally miRNAs with 6mer sites modestly downregulate targeted mRNAs



3'-supplementary site

- Additional Watson-Crick pairings at positions 13-16
- Only 3 or 4 consecutive paired nucleotides in the 3' region of miRNAs contribute to the efficacy of degree of gene downregulation



3'-compensatory site

- Mismatch in the seed region
- Additional Watson-Crick pairings at positions 13-16 compensate the discontinuity of paring in the seed



miRCarta

hsa-1-52.1

Accession	hsa-1-52.1
similar to following miRBase precursors	hsa-mir-21 (MI0000077)
Organism	Homo sapiens
Genome	GRCh38.p10
Location	17:59,841,273-59,841,332 (+)
miRNA	m-1
miRNA	m-52
Sequence (5' -> 3') (60 nts)	UAGCUUAUCAGACUGAUGUUGACUGUUGAAU CUCAUGG CAACACCAGUCGAUGGGCUGUC
MFE	-24.80 kcal/mol
first miRCarta version	1.0
last miRCarta version	1.1
Clusters (10 kb) (1 precursors)	hsa-1-52.1

isomiRs



GGG**A**TGAGGTAGGTTGTATAGTT**T**TAGG
TGAGGTAGGTTGTATAGTT
ATGAGGTAGGTTGTATAGTT**T**
TGAGGTAGGTTGTATAGT**T**
TGAGGTAGGTTGTATAGTT**AA**
TGAGGTAGGTTGTATAGT**T****AA**

c	GGG A TGAGGTAGTAGGTTGTATAGTT T TAGG	Precursor
	TGAGGTAGTAGGTTGTATAGTT	User defined reference
	A TGAGGTAGTAGGTTGTATAGTT T	iso_5p:-1, iso_3p:+1
	T GAGGTAGTAGGTTGTATAGT T	iso_5p:+1, iso_3p:-1
	TGAGGTAGTAGGTTGTATAGTT AA	iso_add:2
	T GAGGTAGTAGGTTGTATAGT AA	iso_5p:+1, iso_3p:-1, iso_add:2

d

T GAGGTA **G** TAGG TTGTA **TAGTT** *iso_snv*
iso_snv_central_offset

Biomarkers for the early-detection and monitoring of Huntington's Disease

David W. Salzman*, Joli Bregu[^], Nathan S. Ray*, and Richard H. Myers*[^]

*sRNAlytics Inc. AstraZeneca BioHub Incubator, 35 Gatehouse Drive, Waltham MA, 02451

[^]Boston University Medical School, Department of Neurology, 72 East Concord Street, Boston, MA, 02118

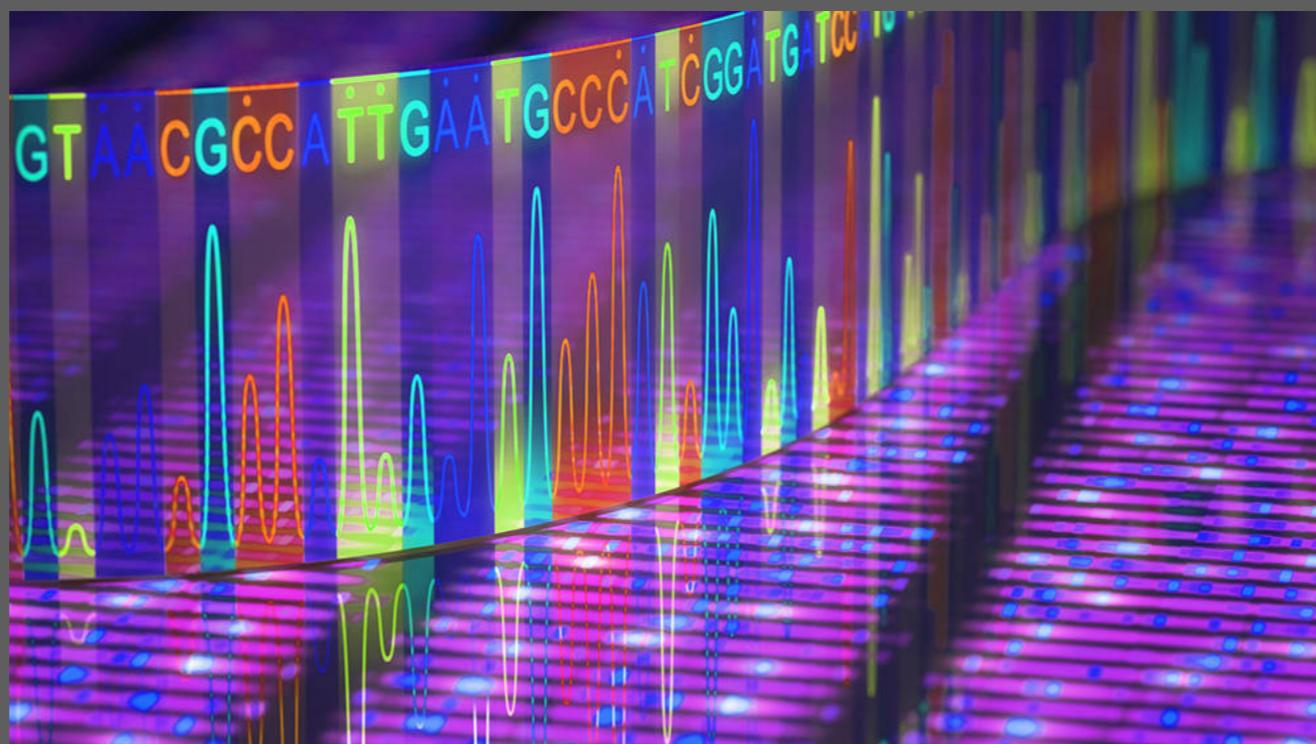
Huntington's Disease (HD) is a genetic disease caused by a CAG trinucleotide repeat in Exon1 of the huntingtin gene. Neurodegeneration results in the loss of cognitive and motor functions, and is caused by aggregation of mutant huntingtin protein in striatal neurons. Volumetric changes in the striatum can be detected decades before the manifestation of clinical phenotypes, indicating that therapeutic intervention would need to occur long before symptomatic presentation. In clinical practice and research settings, the Unified Huntington's Disease Rating Scale (UHDRS) is utilized to evaluate a patients overall physical and neurological health. UHDRS is also the most widely used outcome measure for establishing drug efficacy. However, ...

Nucleic Acids Res. 2017 Apr 7;45(6):2973-2985. doi: 10.1093/nar/gkx082.

Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types.

Telonis AG¹, Magee R¹, Loher P¹, Chervoneva I², Londin E¹, Rigoutsos I¹.

 Author information



High-throughput sequencing

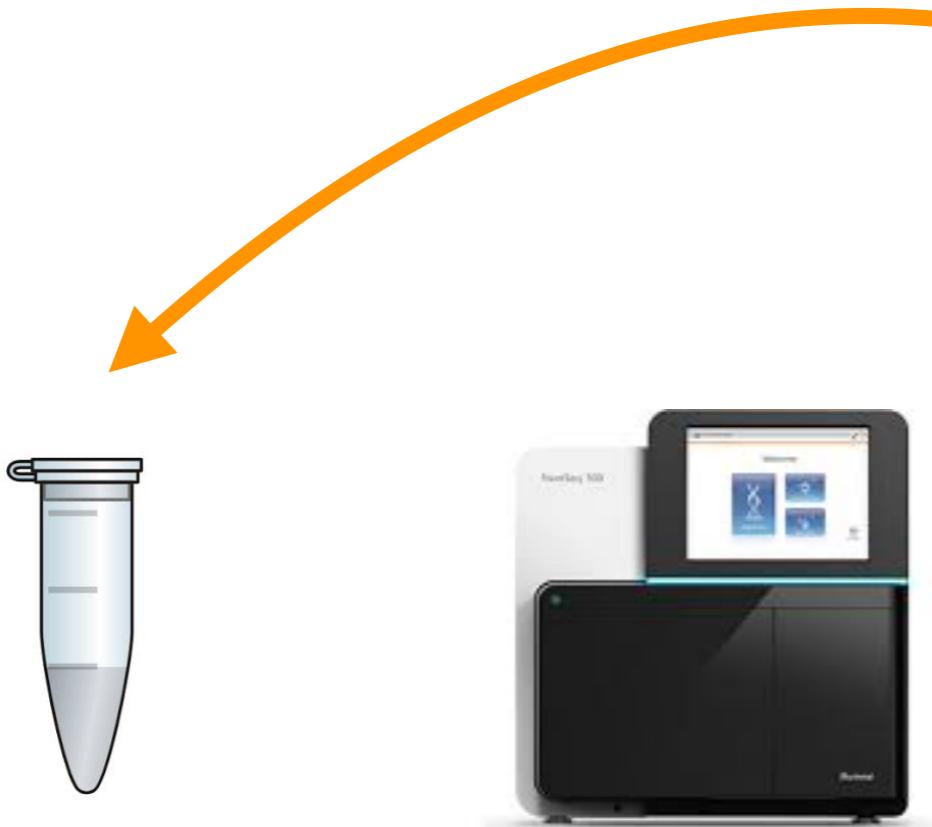
Source of Variation

Library
preparation

Sequencing

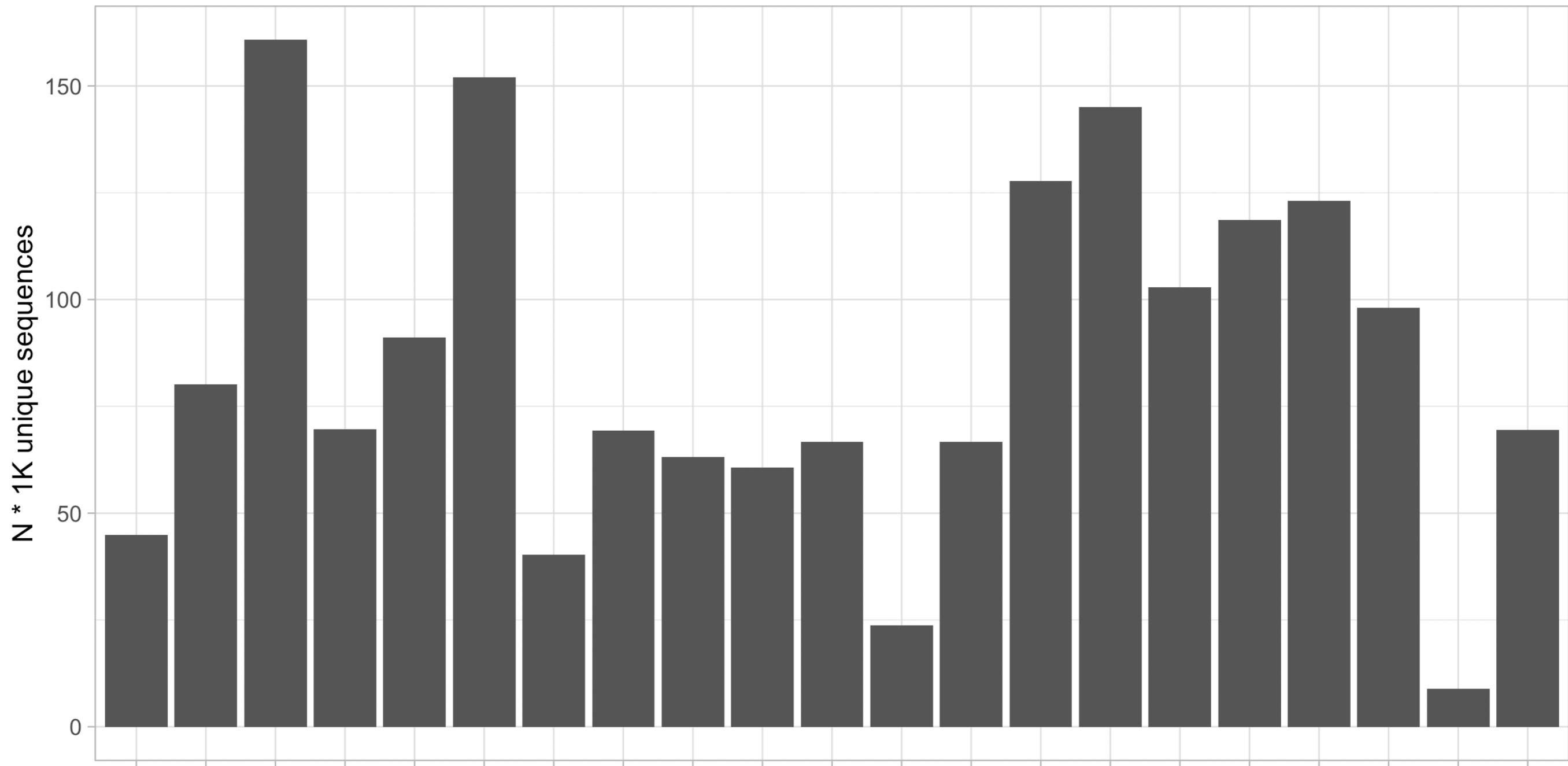
Analysis

Synthetic sample



N synthetic miRNAs

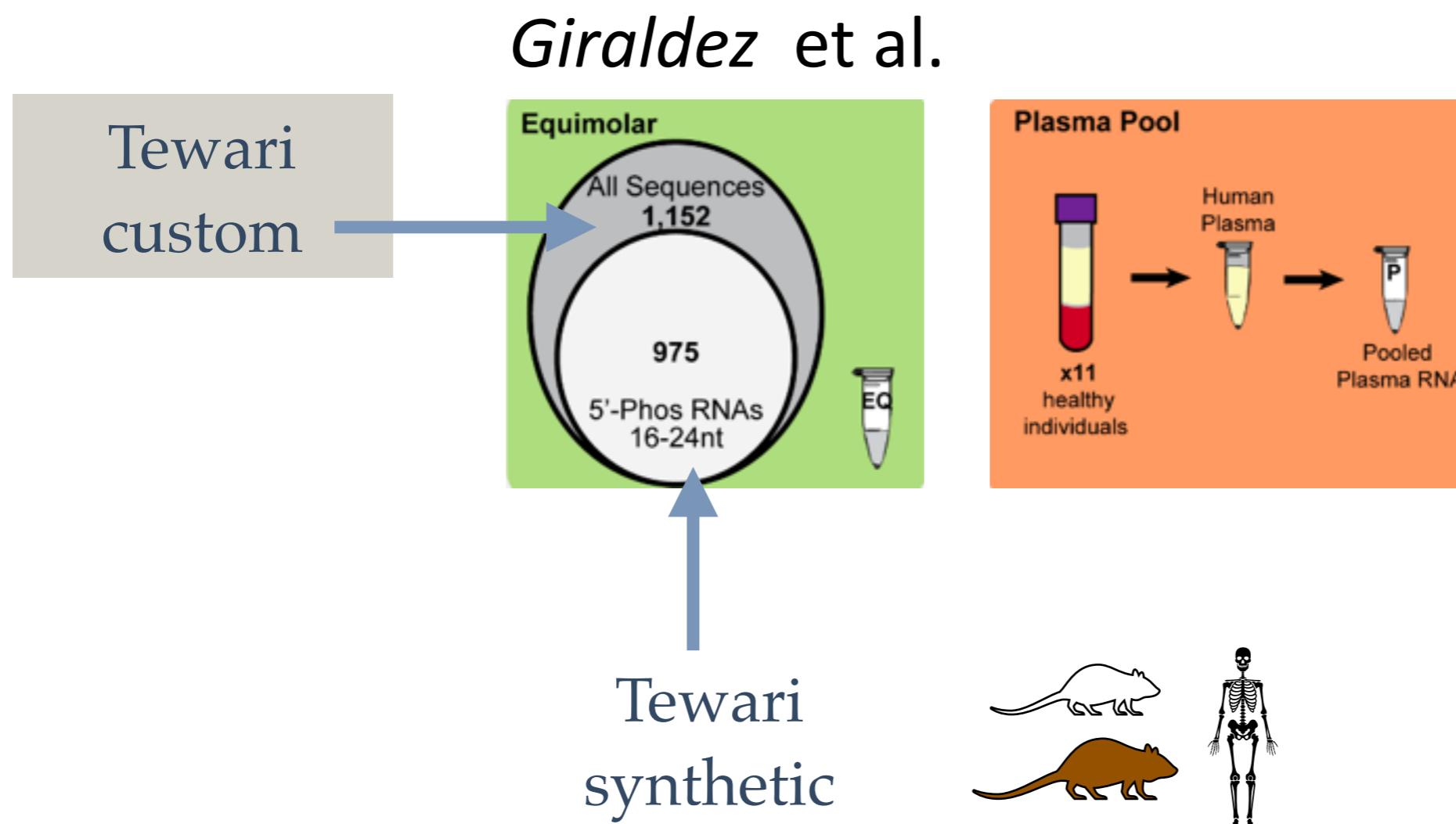
Sample with 450 synthetic miRNAs across 21 replicates



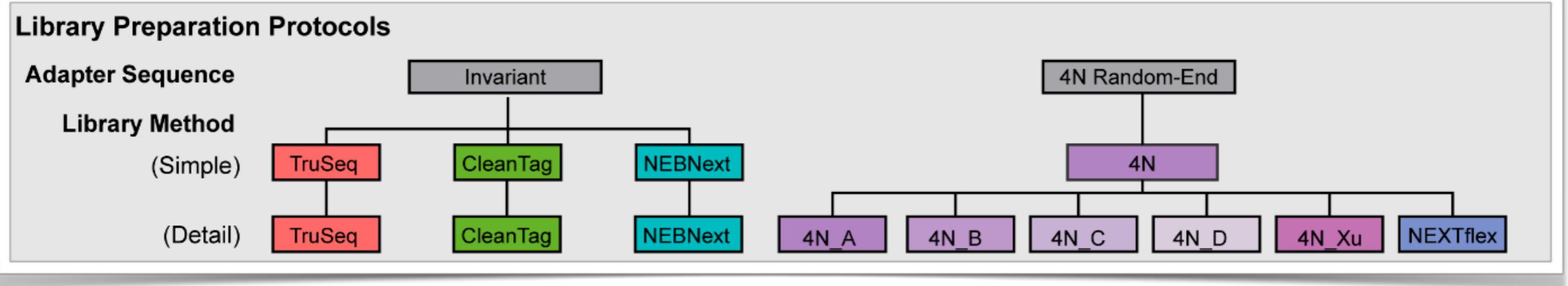
“How well NGS detect isomiRs?”

– *Scientists*

Samples suitable for benchmarking



Different protocols

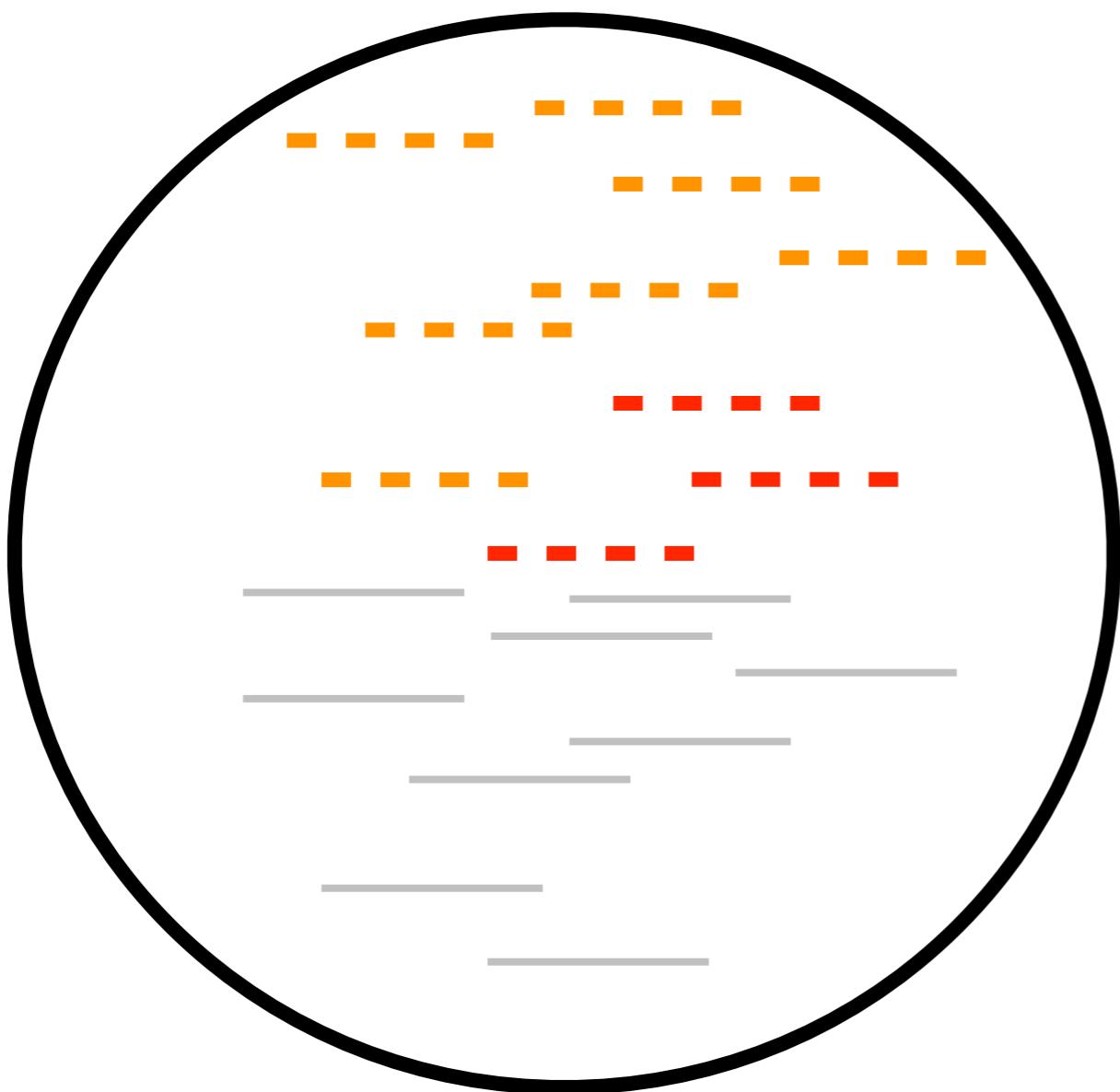


Variations

	3 adapter	5 adapter	PEG
TrueSeq			
NEBNext			Yes
NextFlex	4N	4N	5-15%
AQSeq	4N	4N	20%
CleanTag	Methylphosphonate	2'0me	
CATS	poly-T	TSO with rX	
SMARTer	poly-T	TSO	

Metrics

Sample

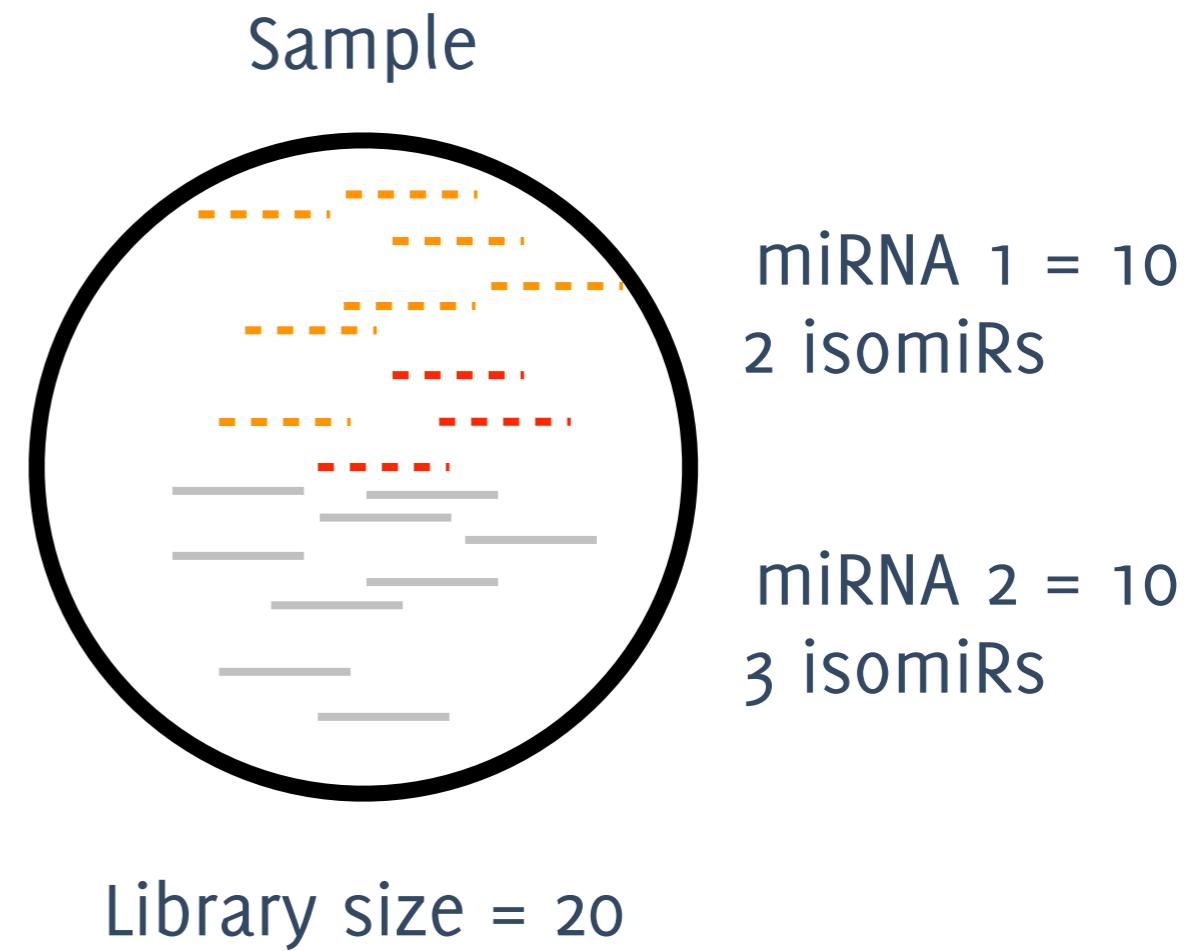
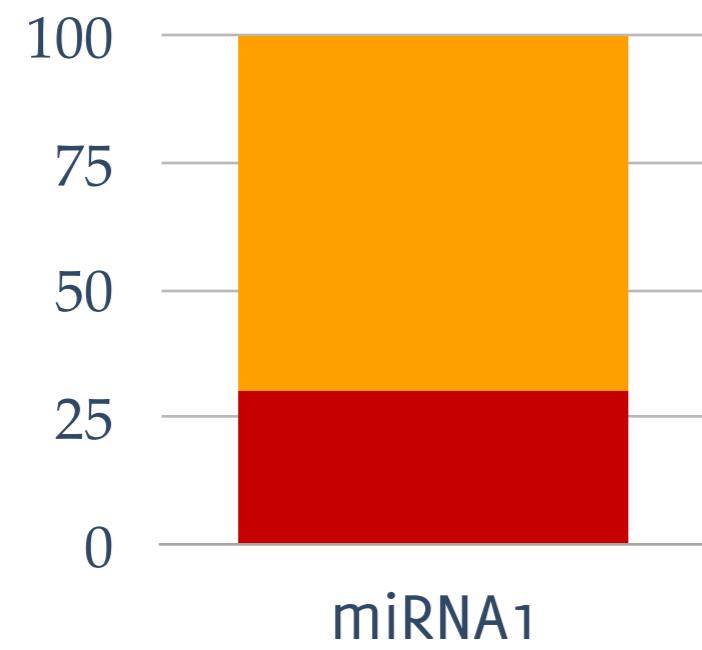


miRNA 1 = 10
2 isomiRs

miRNA 2 = 10
3 isomiRs

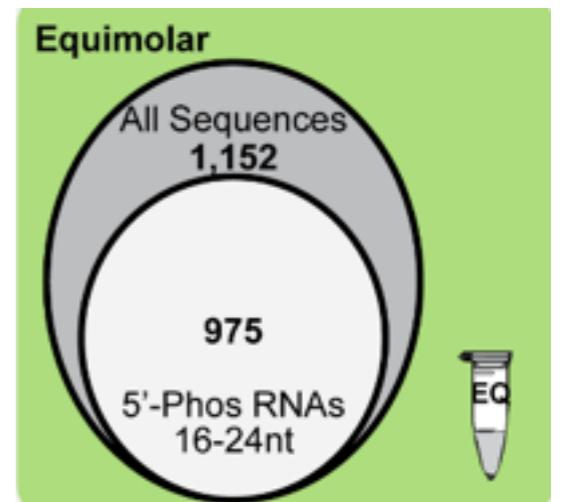
Library size = 20

IMPORTANCE = isomiR 1.1 is $7/10=70\%$ of the miRNA1 reads

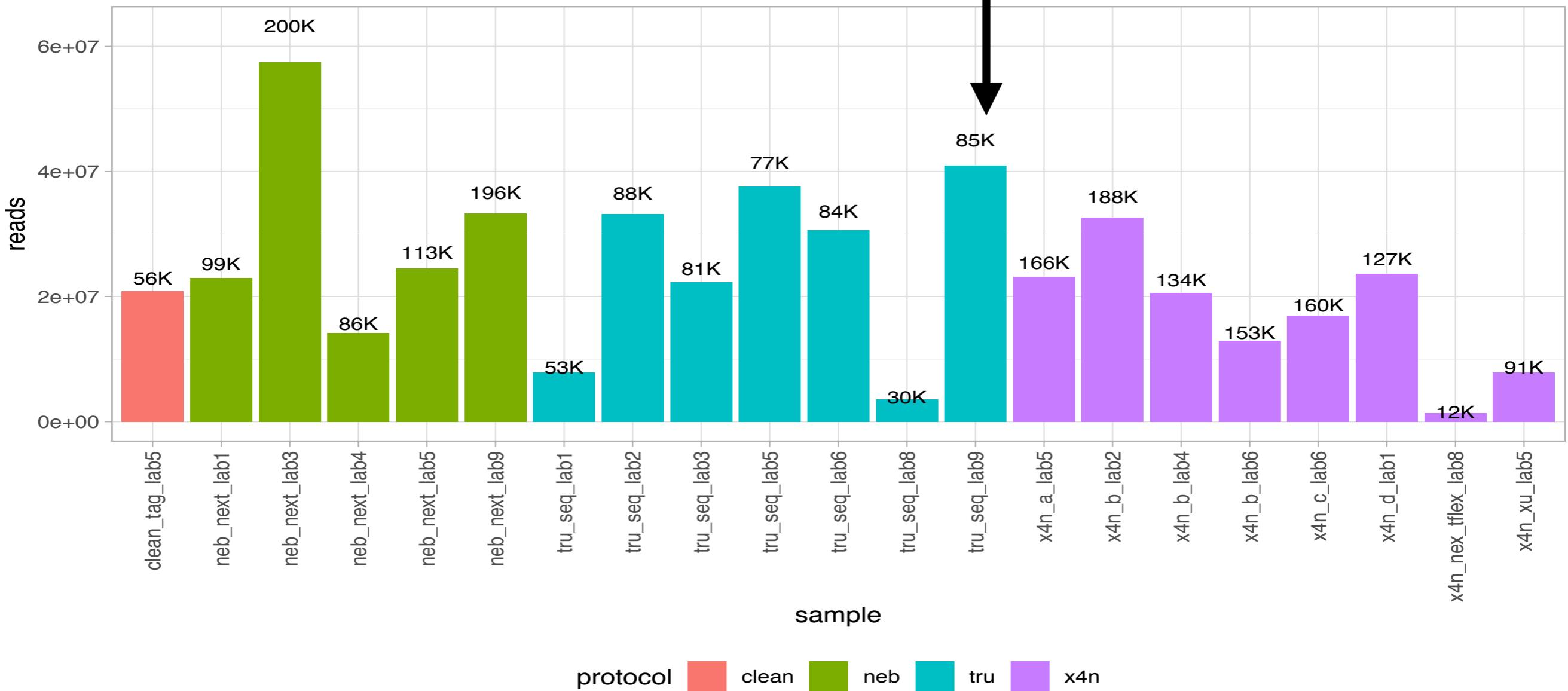


synthetic: library size

Giraldez et al.

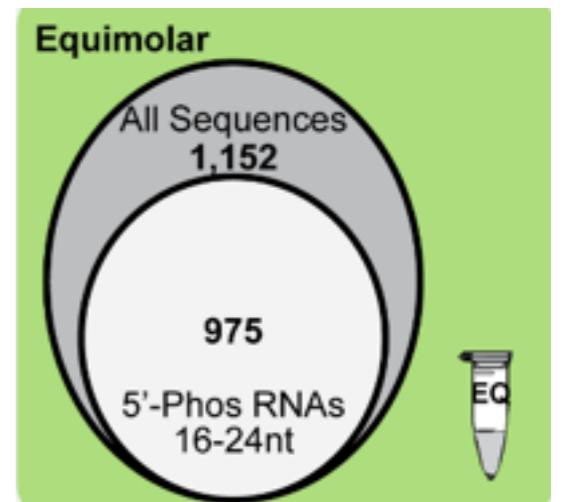


40 million reads and 85K different sequences
mapping to ~450 synthetic miRNAs

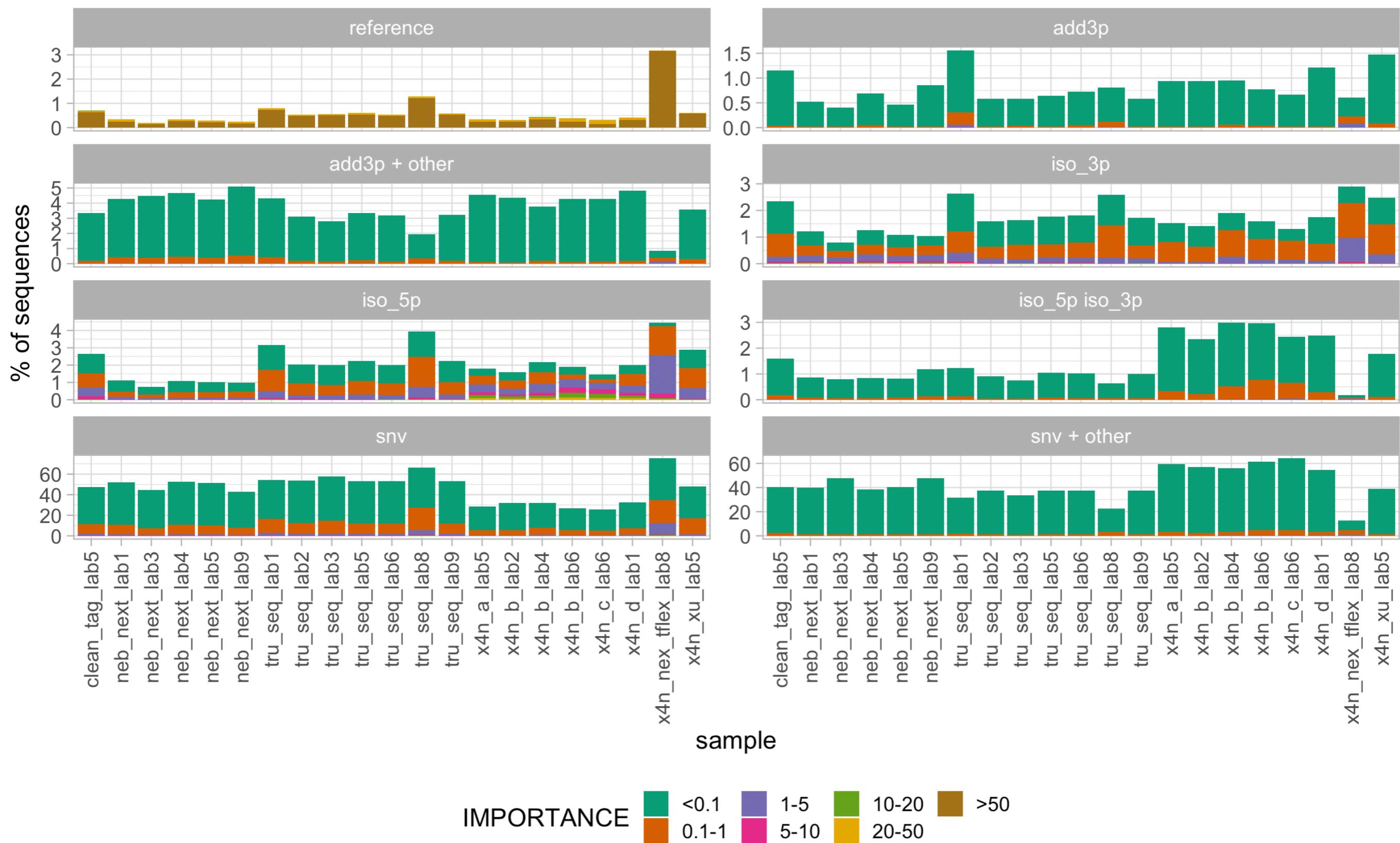


synthetic: Importance of the ‘isomiRs’ detected

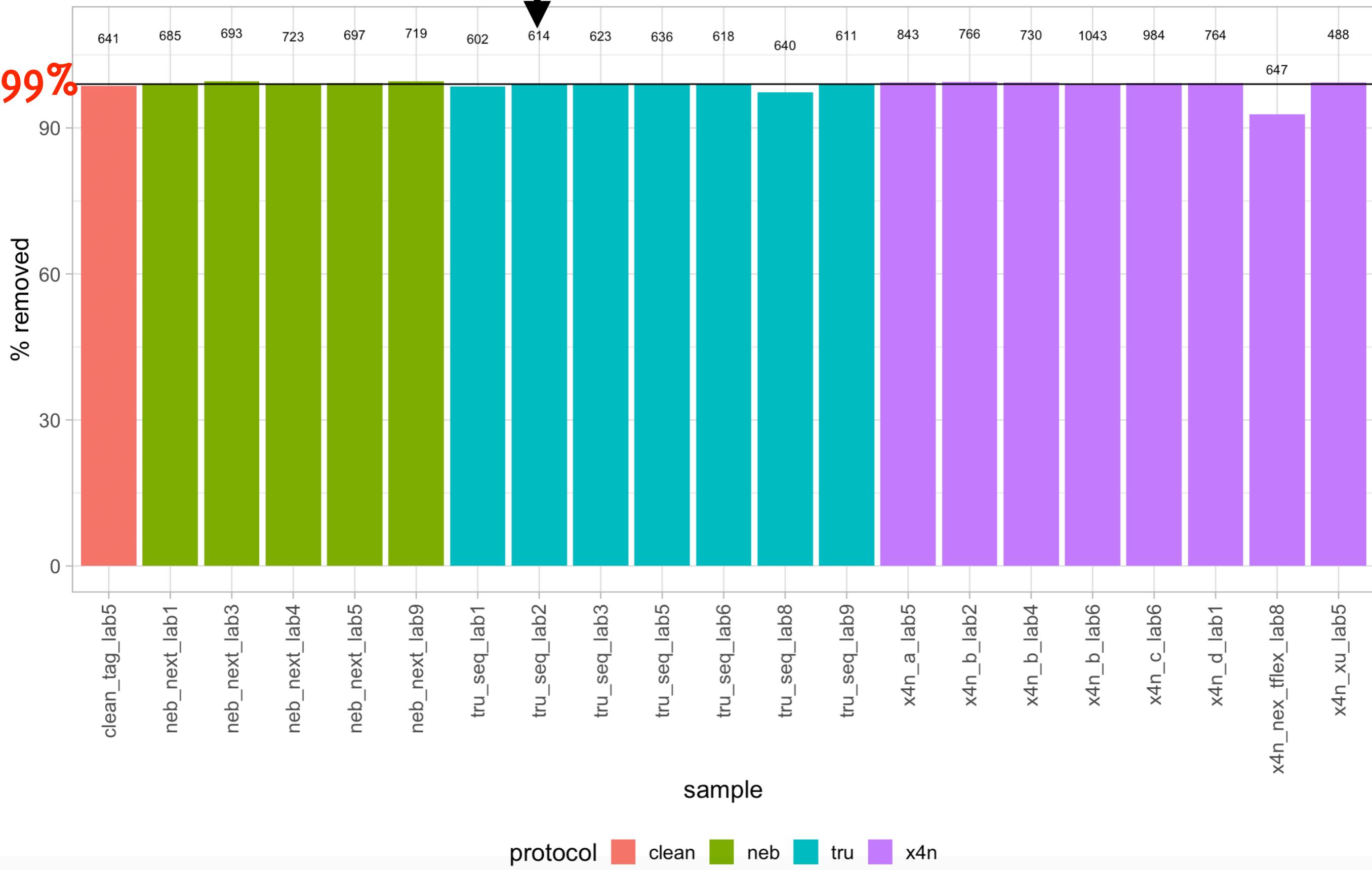
Giraldez et al.



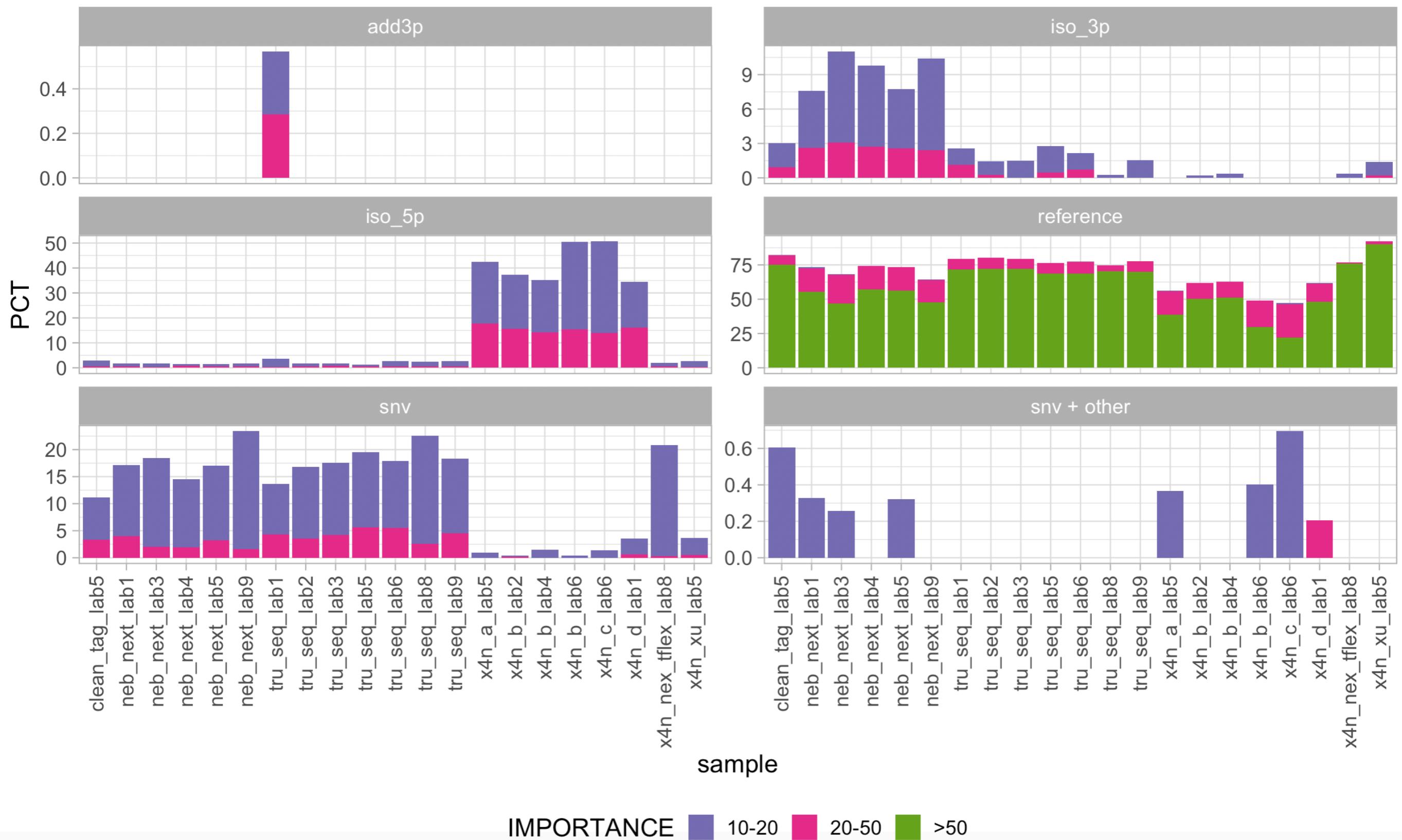
isomiRs importance by type (miRNAs > 1000 counts)



From 85K to 614 unique sequences.
99% of sequences are removed.



isomiRs importance by type (miRNAs > 1000 counts)



Source of Variation

Synthesis

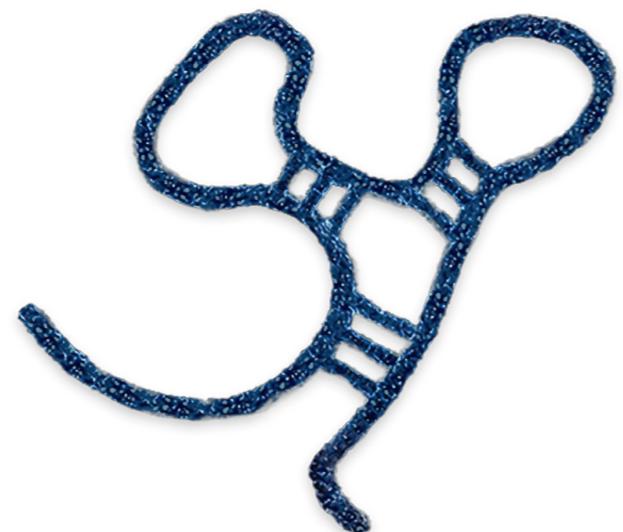
Shipment

Library
preparation

Sequencing

Analysis

We assume real molecules will be detected across replicates



Sample 1

Sample 2

Sample 3

Sample 4

Synthetic miRNAs

All

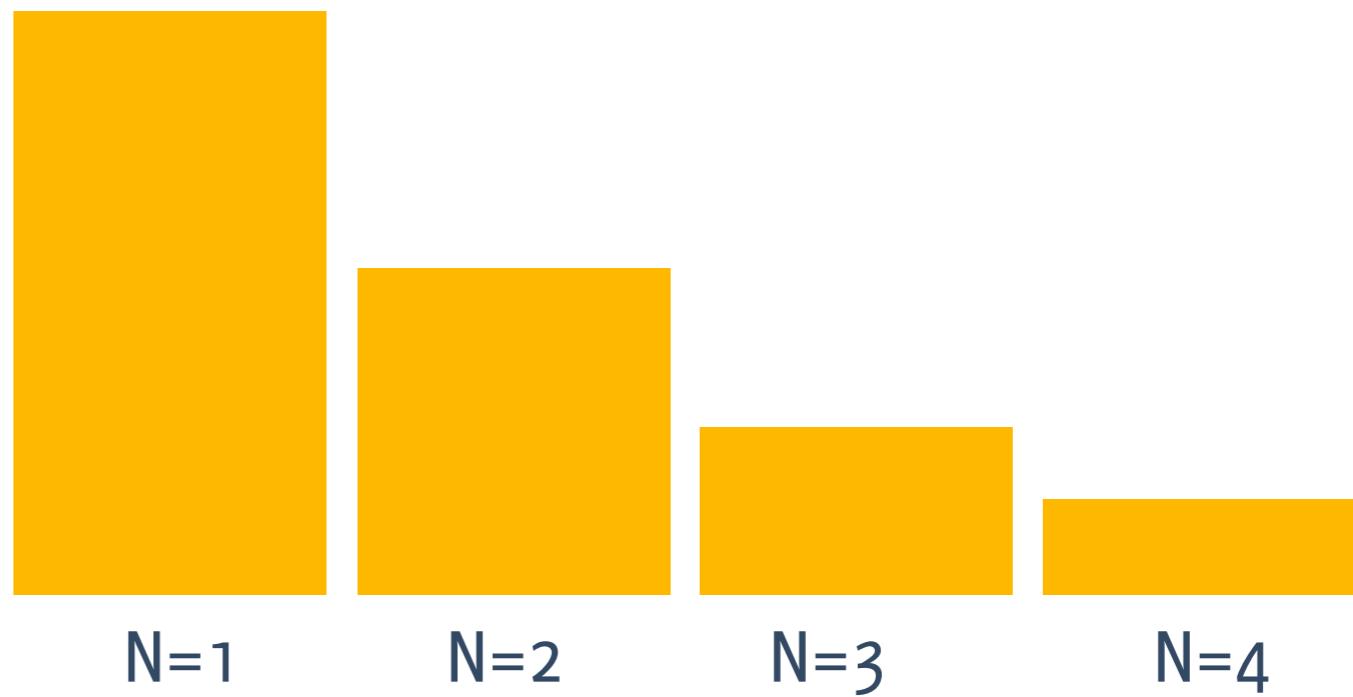
N=1

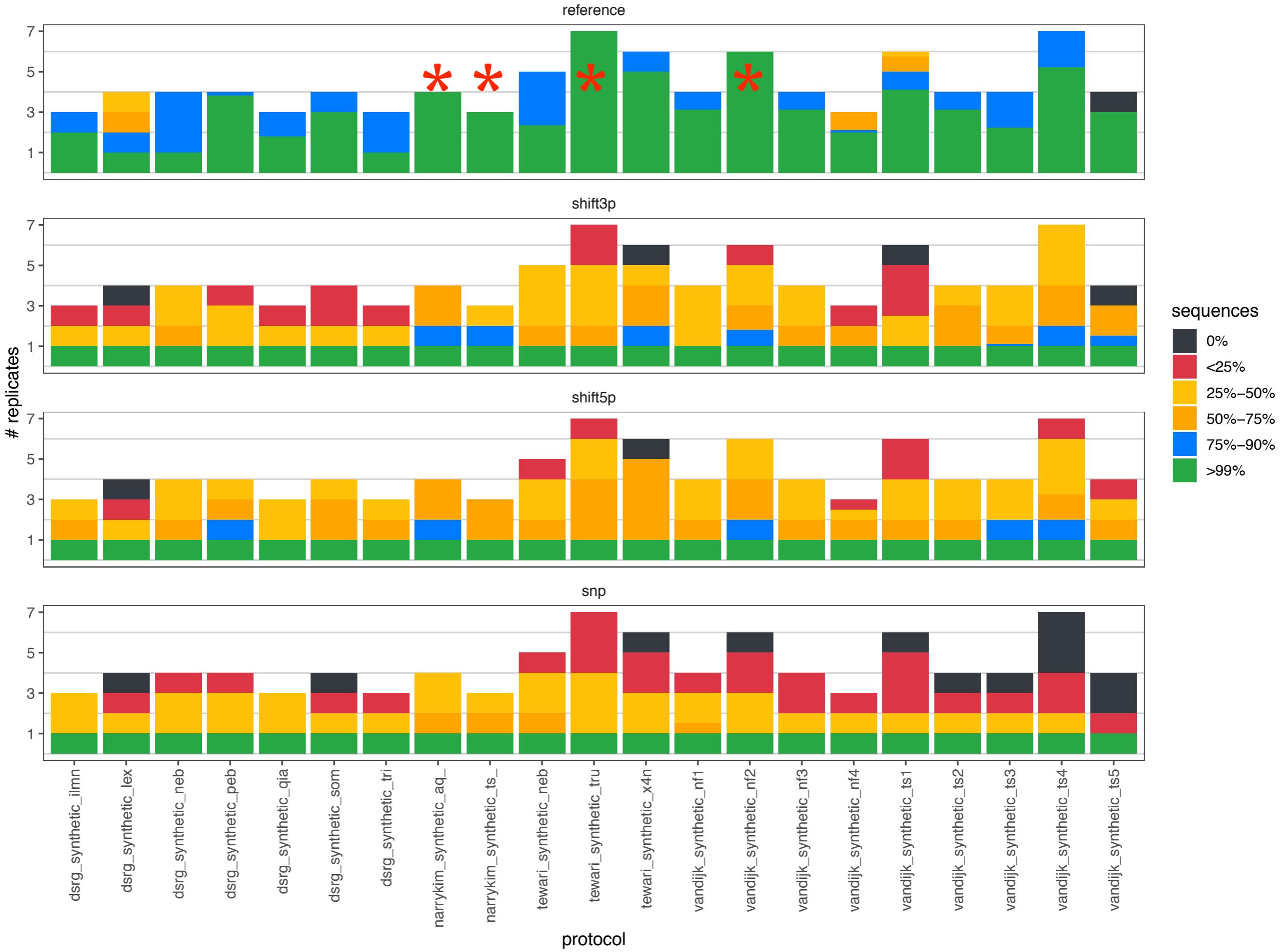
N=2

N=3

N=4

Noise





Narry Kim Data

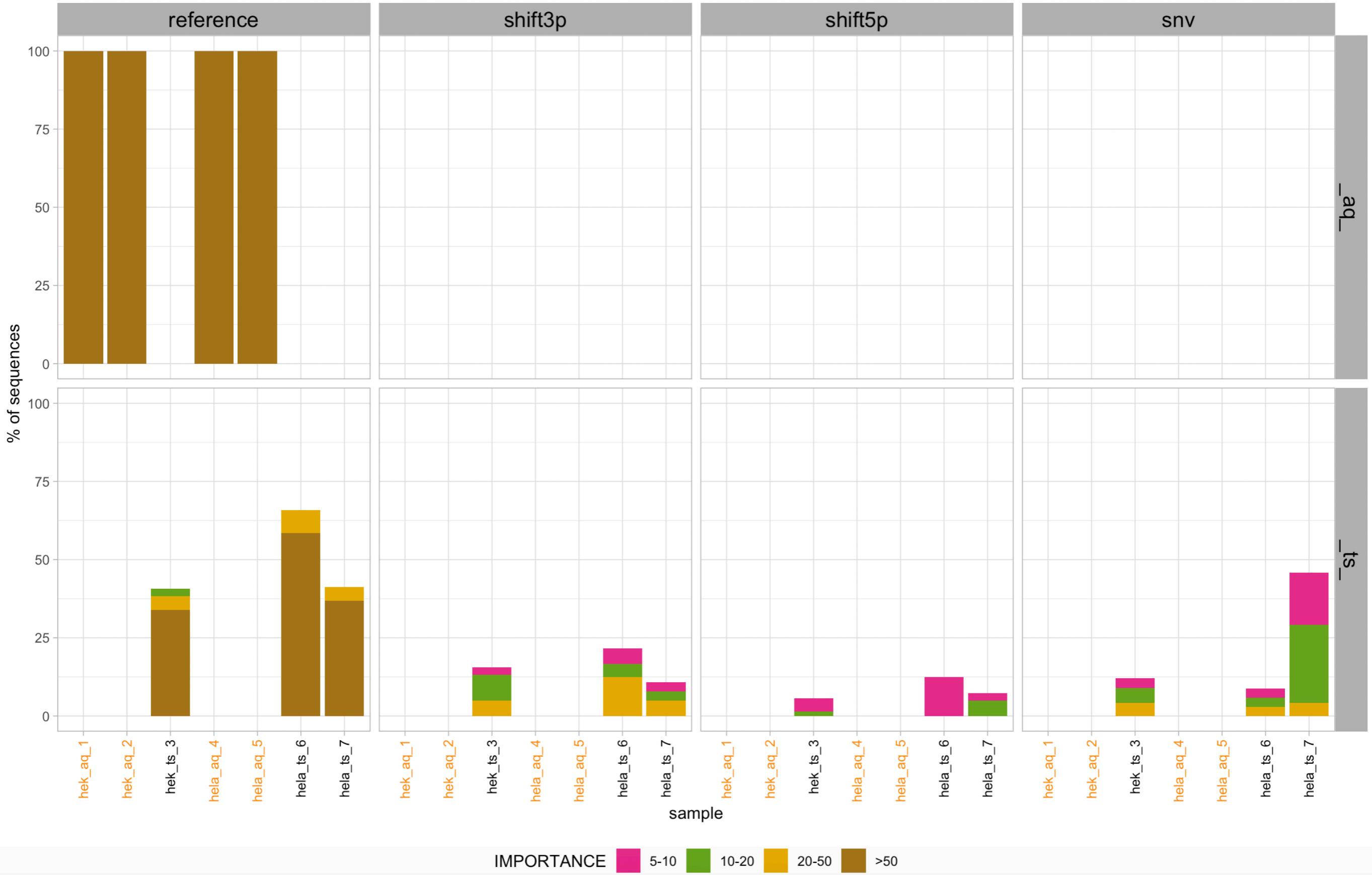
Protocol 1

30 spikeins x 3 replicates

Protocol 2

30 spikeins x 4 replicates

Keep isomiRs > 5%



Summary

- ✓ Each miRNA generates a diversity of isomiRs:
 - ✓ 99% contribute to < 10% of the miRNA abundance
 - ✓ Independent of pipelines and data sets
- ✓ >90% of miRNAs affected
- ✓ Custom 4N protocols perform worst, BUT NEXTFLEX shows good performance, and AQSeq looks the best

Guidelines

- ✓ NEXFlex/TrueSeq shows less unexpected sequences
- ✓ Removing isomiRs <10% of importance may clean your data
- ✓ Increase your replicates
- ✓ Don't filter by sample individually



https://github.com/miRTop/incubator/tree/master/projects/tewari_equimolar

<https://github.com/miRTop/mirGFF3>

<https://github.com/miRTop/mirtop>



Jeffery Ma, Jason Sydes, Yin Lu, Ernesto Aparicio-Puerta,
Shruthi Bandyadka, Victor Barrera, Peter Batzel,
Rafa Allis, Roderic Espin



Kieran O'Neill, Eric Londin, Aristeidis G. Telonis,
Elisa Ficarra, John H. Postlethwait,

Provincial Health Services Authority



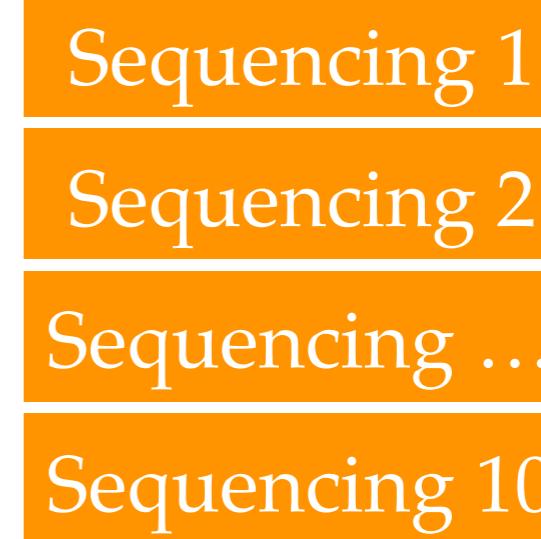
Thank you!



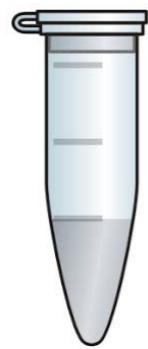
UHBR
+ 30 spikes



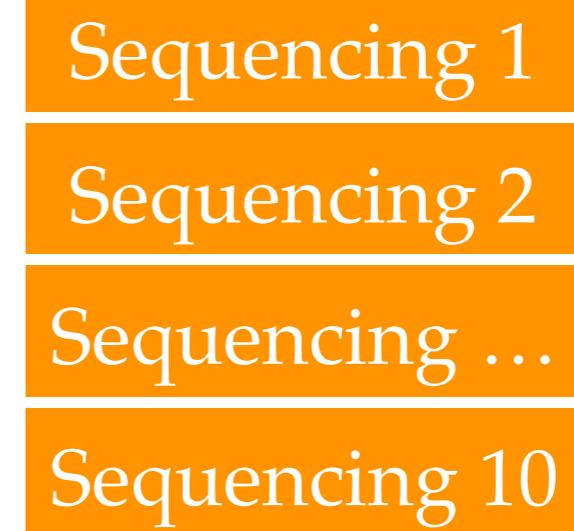
UHLR
+ 30 spikes

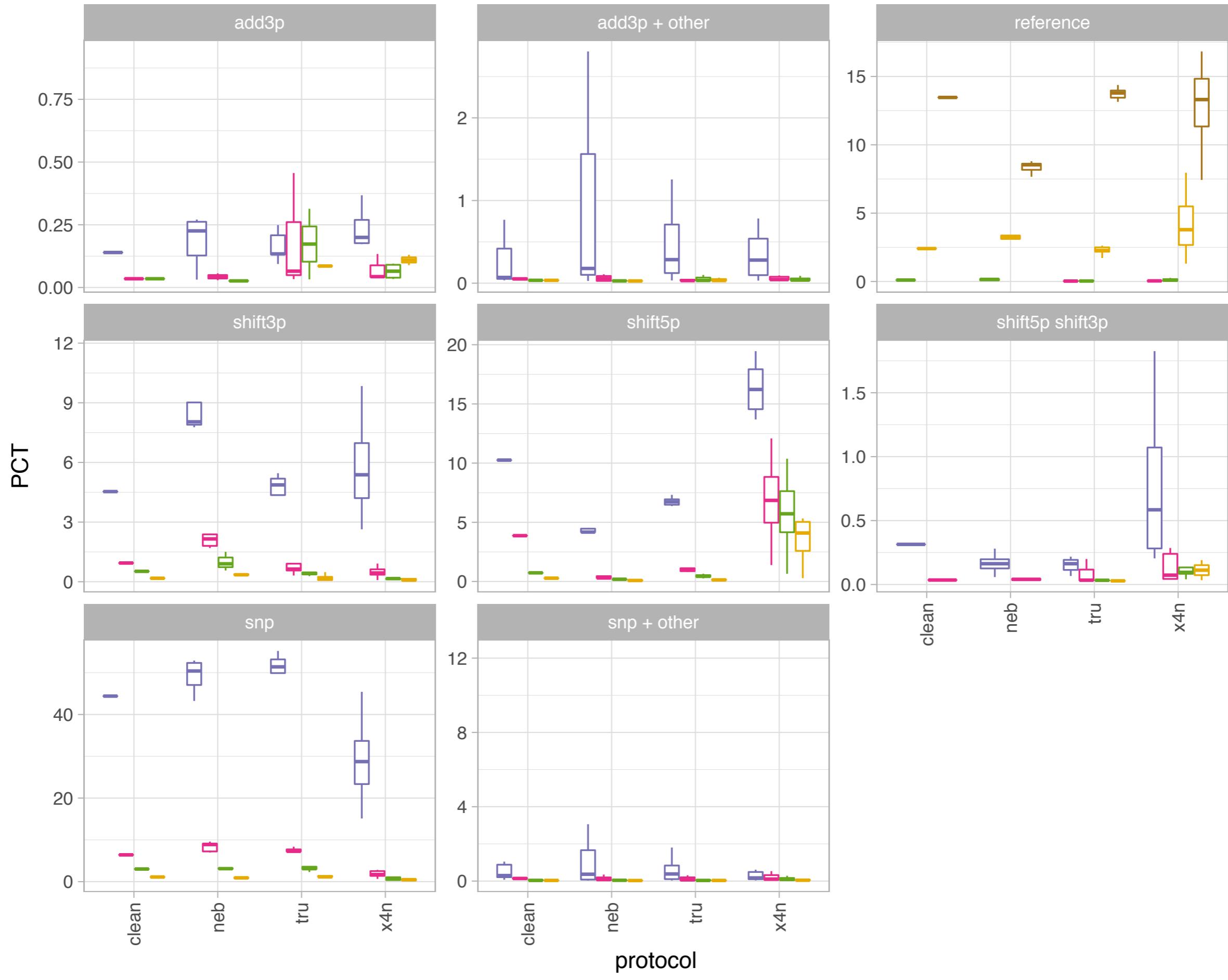


UHBR
+ 30 spikes



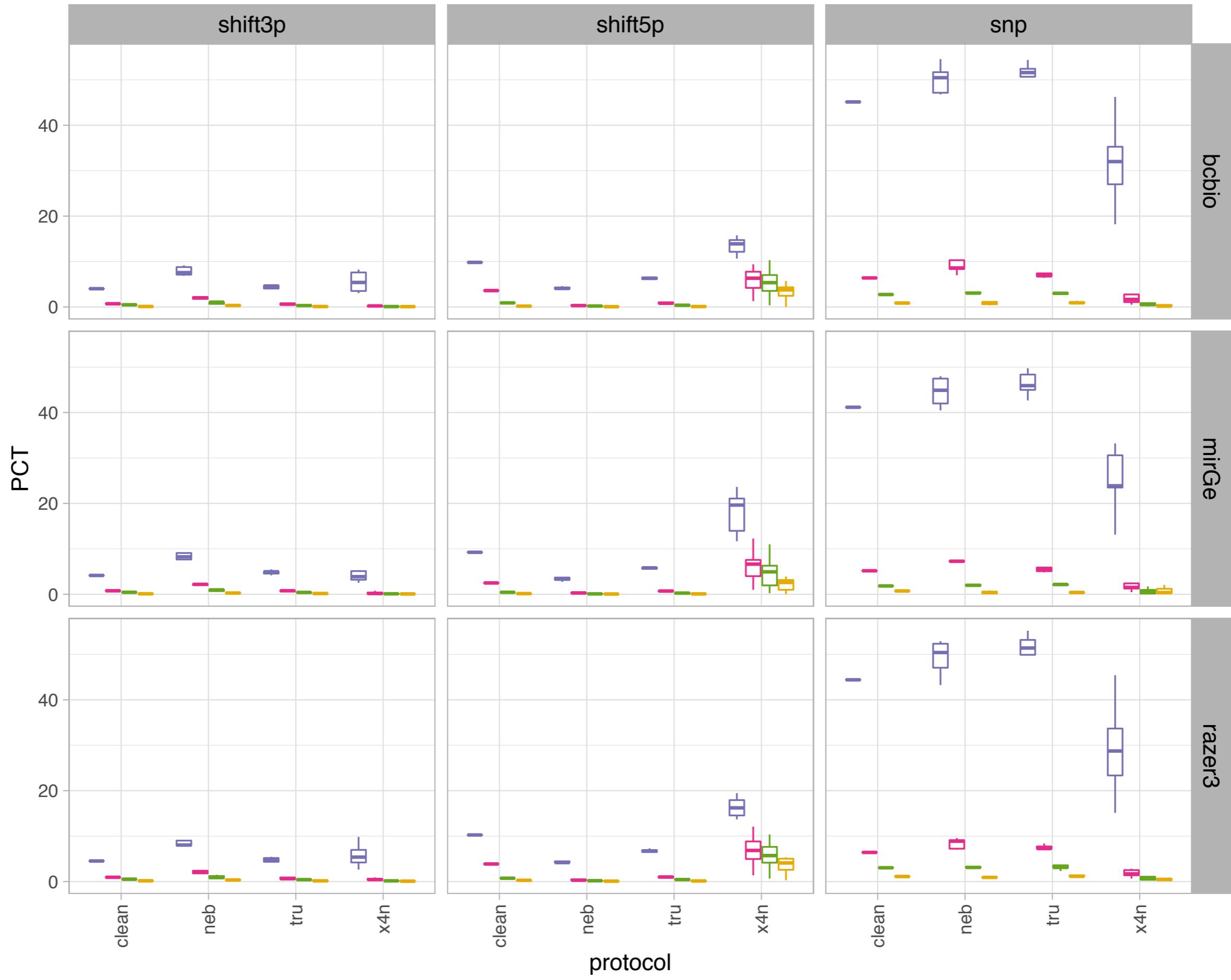
Library 1



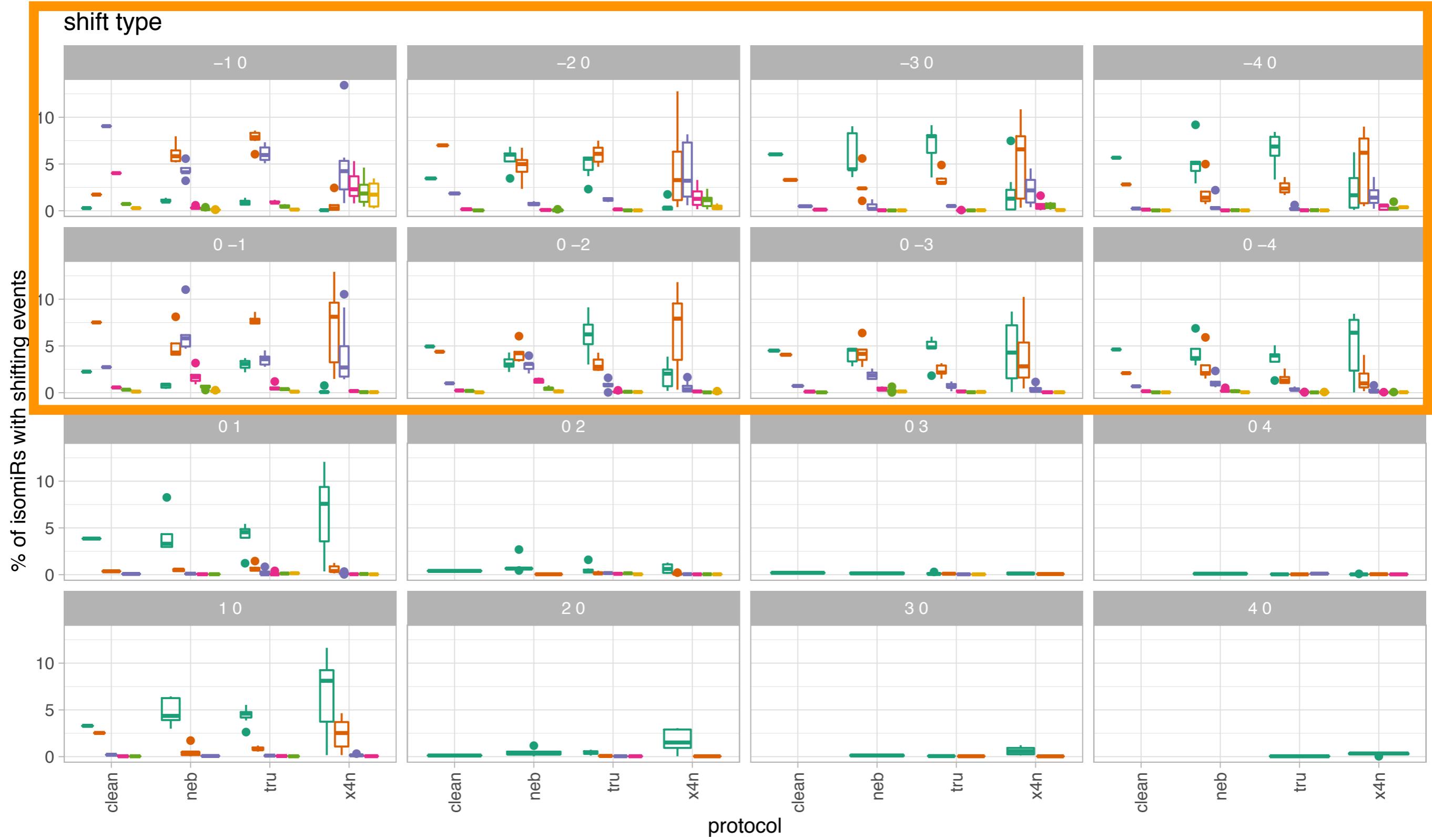


isomiRs importance by type with pct > 5





IMPORTANCE 1-5 5-10 10-20 20-50



pct_cat

- <0.1
- 0.1-1
- 1-5
- 5-10
- 10-20
- 20-50

