

MIRNA WARS

THE ISOMIRS MENACE

@lopantano

Lorena Pantano, Ph. D

<https://lpantano.github.io>

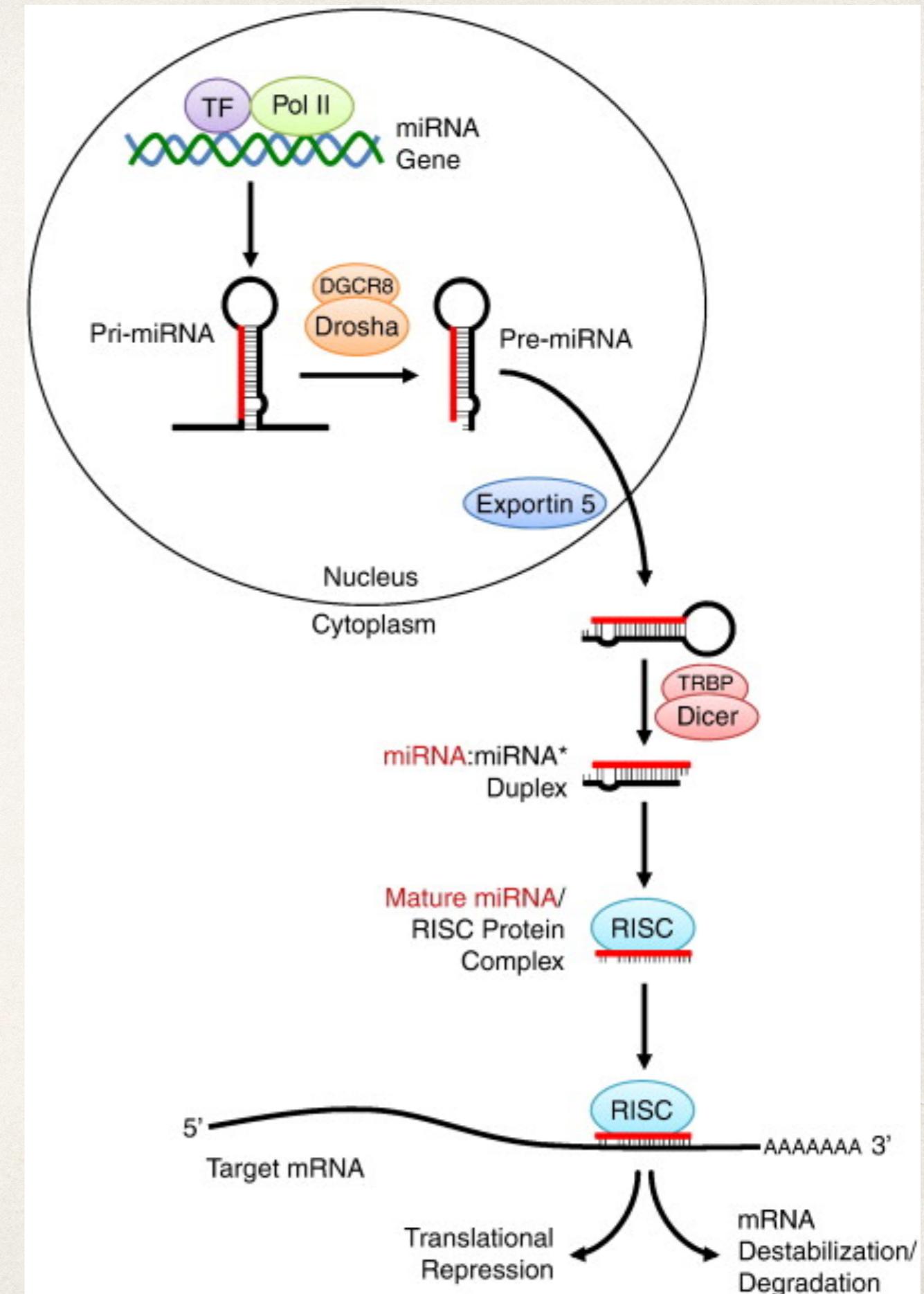
150 human liver samples with small RNAseq data



<https://www.vinylinfo.org/news/population-growth-sustainability-and-our-future>

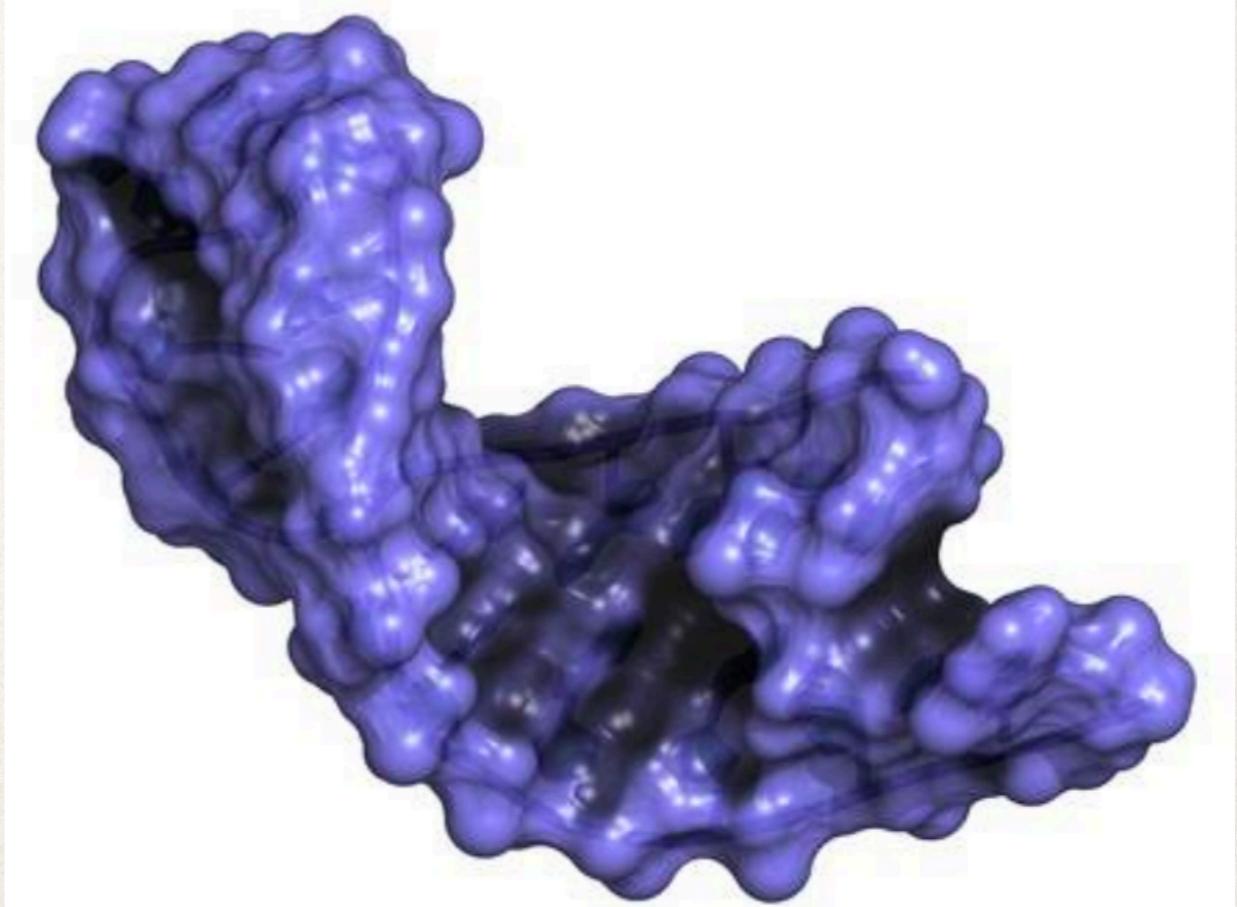
miRNAs

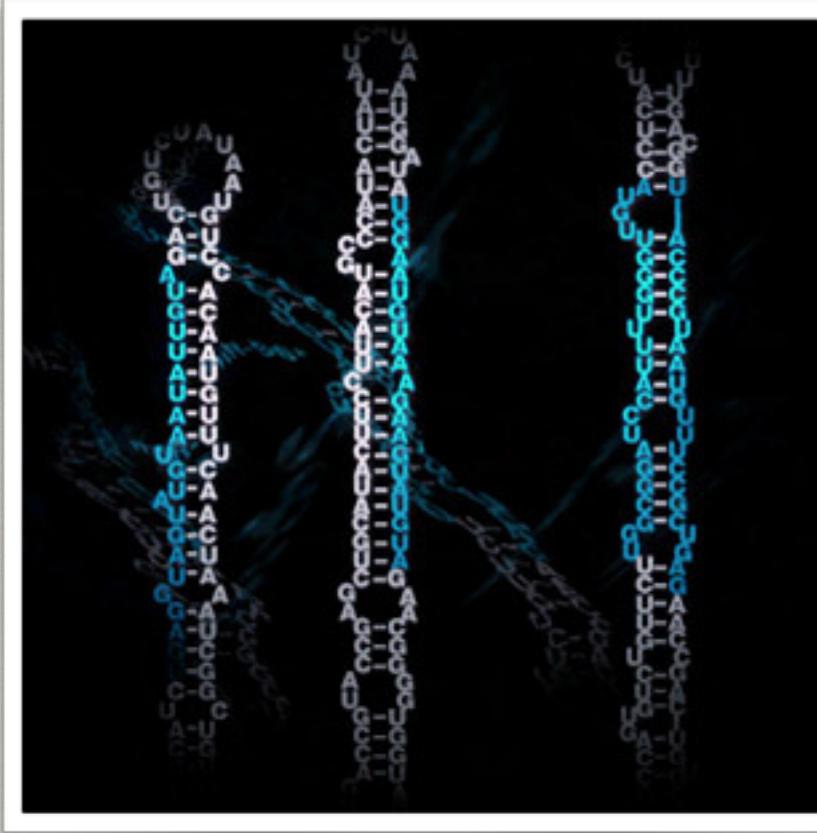
Gene regulation by imperfect complementary between seed region in miRNA and 3'UTR in the targeted RNA molecule.



isomiRs

<https://www.flickr.com/photos/ac1/440441500>





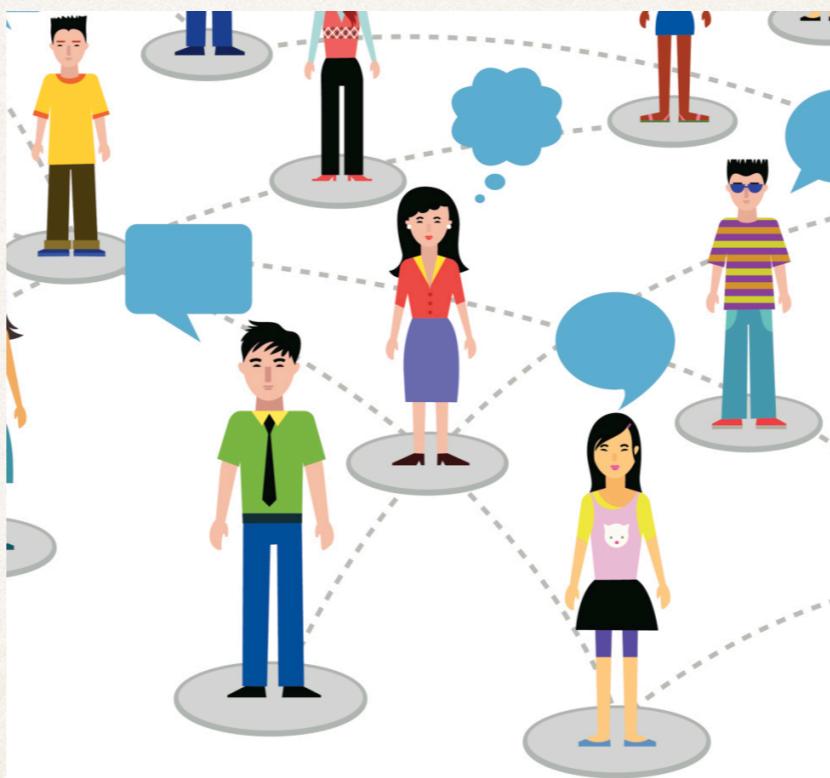
GGG**A**TGAGGTAGTAGGTTGTATAGTT**T**TAGG
TGAGGTAGTAGGTTGTATAGTT
ATGAGGTAGTAGGTTGTATAGTT**T**
TGAGGTAGTAGGTTGTATAGT**T**
TGAGGTAGTAGGTTGTATAGTT**AA**
TGAGGTAGTAGGTTGTATAGT**T****AA**



Software Categories				
» Known miRNA Identification	39	DeAnniso	✓	✗
» isomiRs Identification	15	iMir	✓	✓ (if other species than human, mouse and rat are analysed)
» Novel miRNA/Precursor Analysis	49	isomiR2Function	✓	✓
» Differential Expression Analysis	26	IsomiRage	✓	✓
» Target Prediction	59	isomiReX	✓	✗
» Target Functional Analysis	2			
» miRNA-SNP Analysis	7			
Other Tools				
» Target Prediction Server	10			
» New Prediction				
» My Jobs				
» Sample results				

What is the best way to analyze this data?

A large black circle highlights the "Target Prediction Server" section of the sidebar. A large black arrow points from the bottom left towards the "Target Prediction Server" button.







HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH



Harvard Medical School
Research Computing



JOHNS HOPKINS
SCHOOL *of* MEDICINE

SciLifeLab



National Institutes of Health

O UNIVERSITY OF
OREGON



UNIVERSIDAD
DE GRANADA

U **HEALTH**
UNIVERSITY OF UTAH

BROAD
INSTITUTE



BC
CAN
CER

Provincial Health Services Authority

CRG[®]
Centre
for Genomic
Regulation



Jefferson

HOME OF SIDNEY KIMMEL MEDICAL COLLEGE

BRIGHAM HEALTH



**BRIGHAM AND
WOMEN'S HOSPITAL**





mirTOP

miRNA transcriptome open project

<http://mirtop.github.io>

Repositories 5 People 20 Teams 2 Projects 0 Settings

Find a repository... Type: All Language: All Customize pinned repositories [New](#)

incubator

Where all ideas and discussions happen to lead to new repositories

● R ★ 3 ⚡ 4 Updated a minute ago



miRTOP.github.io

project for small RNA standard annotations

● HTML ★ 2 ⚡ 1 MIT 1 issue needs help Updated 4 days ago



mirtop

command lines tool to annotate miRNAs with a standard mirna/isomir naming

formatter mirna gff isomirs smallrna-seq

● Python ★ 6 ⚡ 13 MIT 17 issues need help Updated 5 days ago



simulator

first ideas and brainstorming for small RNA simulator

● C++ Updated 25 days ago

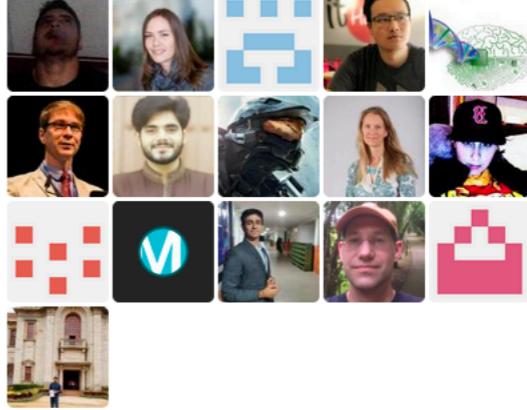


Top languages

● C++ ● Python ● R ● HTML

People

20 >



Invite someone

Main projects:

- a format
 - a way to compare
-



Format derived from GFF3: mirGFF3

a

```
## mirGFF3. VERSION 1.1
## source-ontology: miRBasev21 doi:10.25504/fairsharing.hmgte8
## COLDATA: sample1
```

e

b

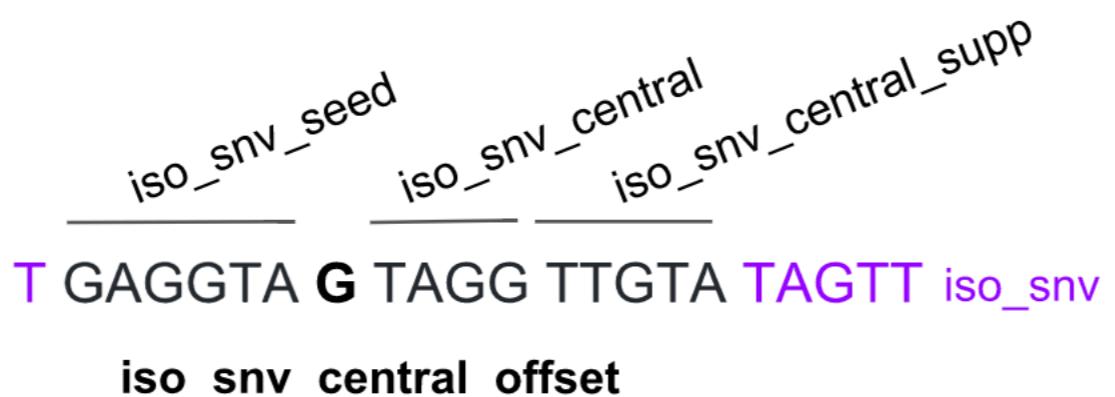
Read=GATGAGGTAGTAGGTTGTATAGTT -> UID=**iso-24-5URPV39QFE**

Read=ATGAGGTAGTAGGTTGTATAGTT -> UID=**iso-23-I0S31NSL0E**

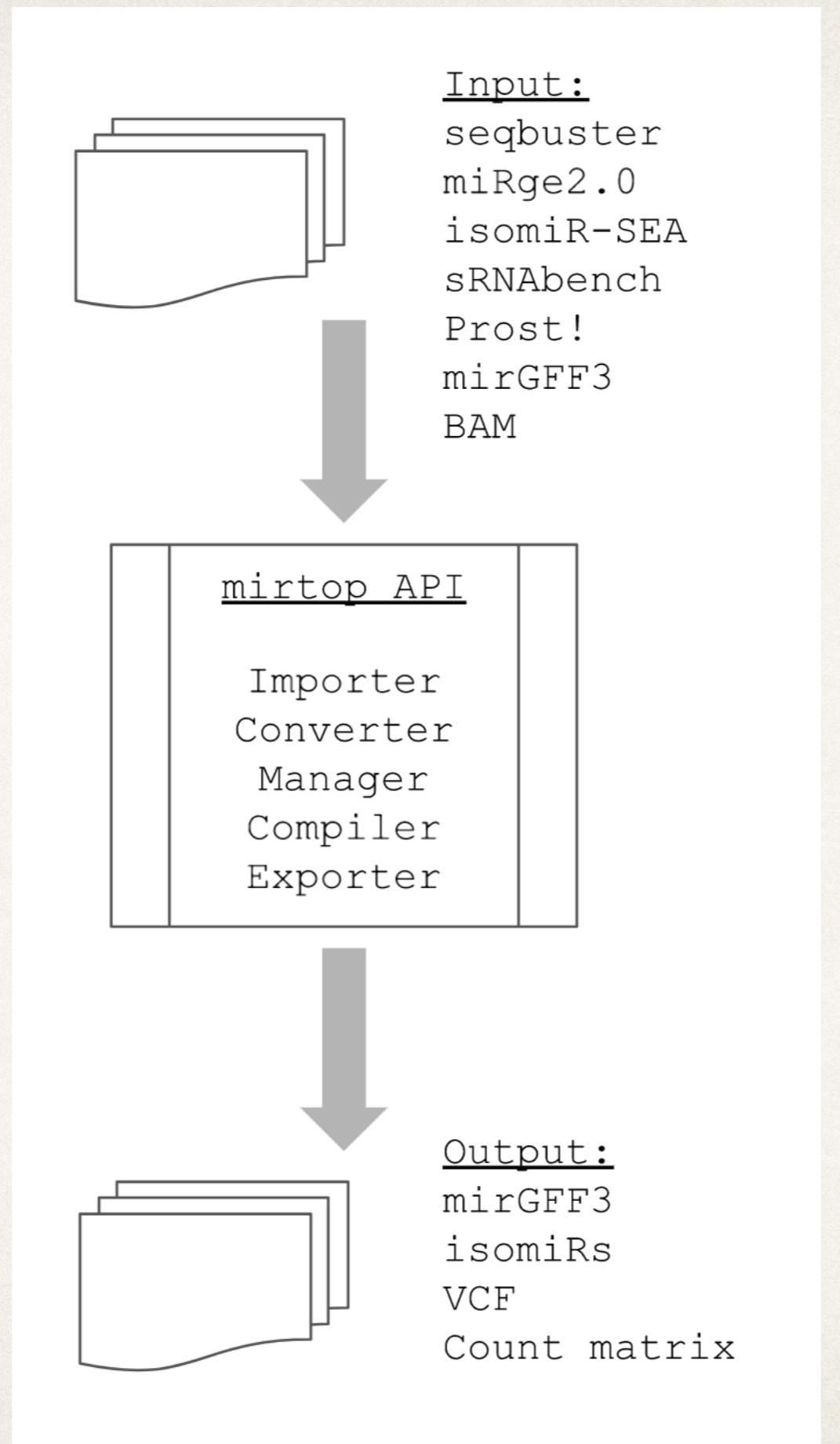
c

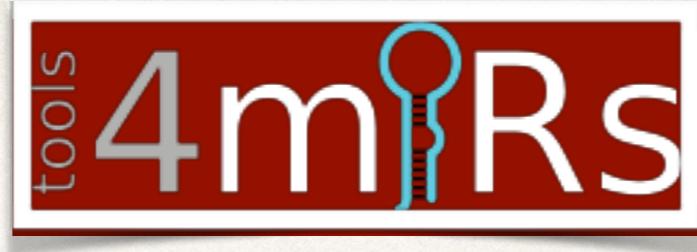
GGG A TGAGGTAGTAGGTTGTATAGTT T TAGG	Precursor
TGAGGTAGTAGGTTGTATAGTT	User defined reference
A TGAGGTAGTAGGTTGTATAGTT T	iso_5p:-1, iso_3p:+1
T GAGGTAGTAGGTTGTATAGT T	iso_5p:+1, iso_3p:-1
TGAGGTAGTAGGTTGTATAGTT AA	iso_add:2
T GAGGTAGTAGGTTGTATAGT T AA	iso_5p:+1, iso_3p:-1, iso_add:2

d



Python API: mirtop



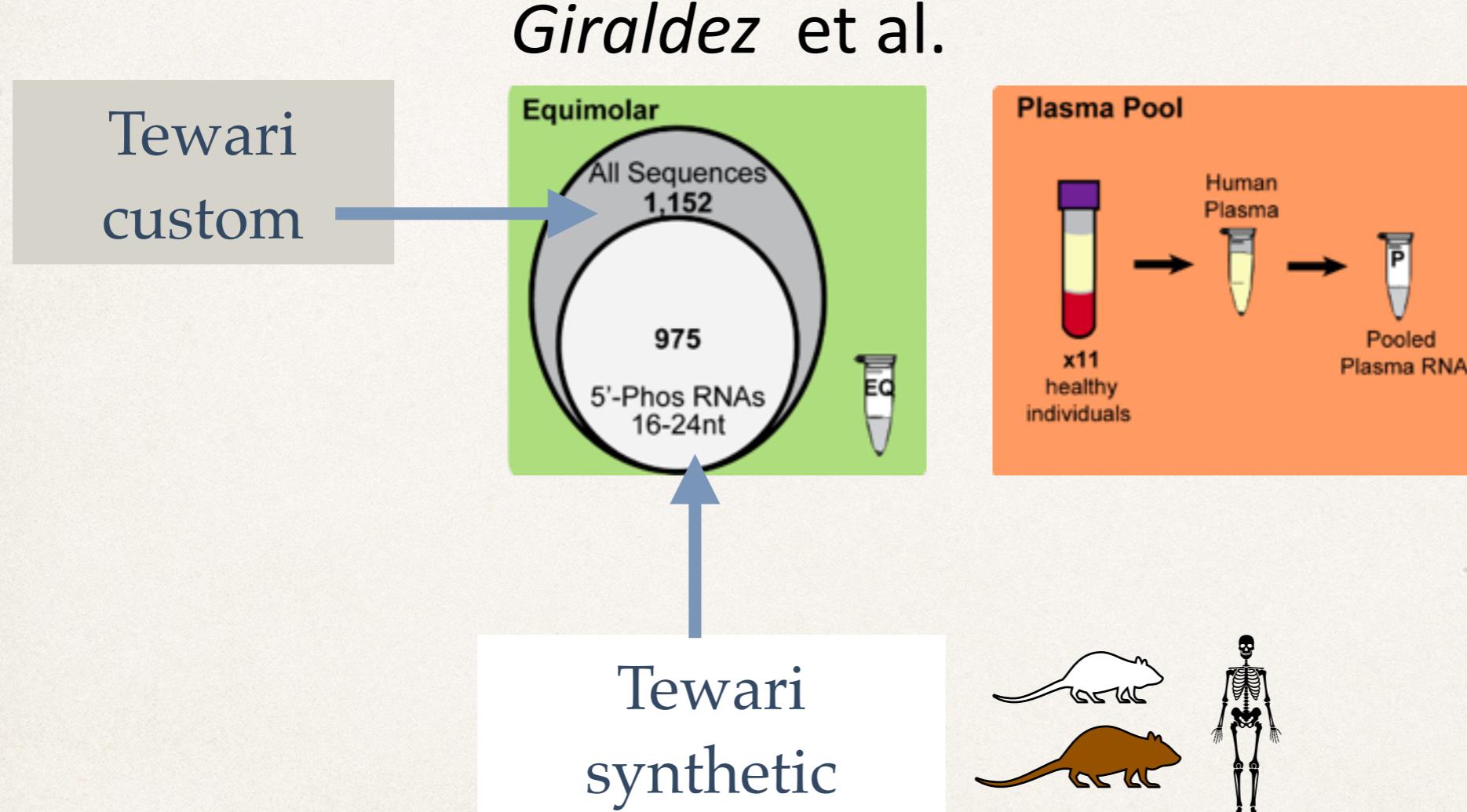


Software Categories				
» Known miRNA Identification	39	DeAnniso	✓	✗
» isomiRs Identification	15	iMir	✓	✓ (if other species than human, mouse and rat is analysed)
» Novel miRNA/Precursor Analysis	49	isomiR2Function	✓	✓
» Differential Expression Analysis	26	IsomiRage	✓	✓
» Target Prediction	59	isomiReX	✓	✗
» Target Functional Analysis	2			
» miRNA-SNP Analysis	7			
Other Tools				
» Target Prediction Server	10			
» New Prediction				
» My Jobs				
» Sample results				

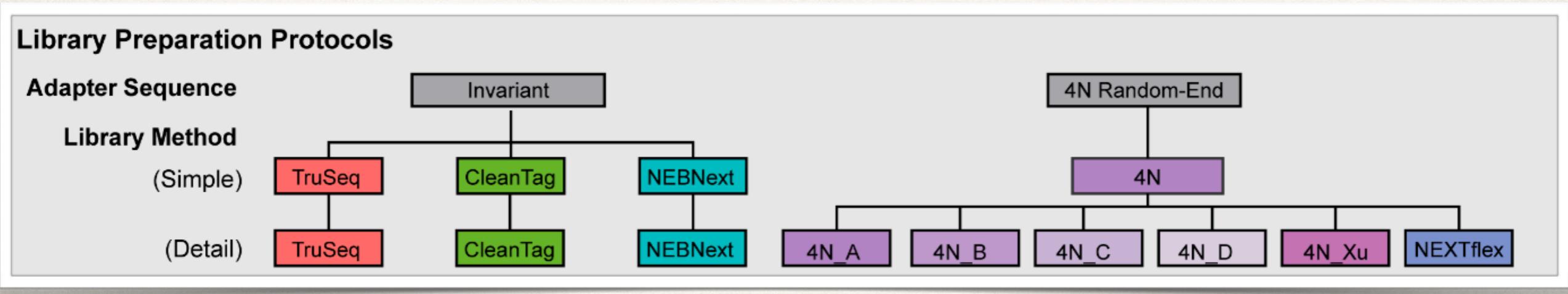
What is the best way to analyze this data?

A large black circle highlights the "Target Prediction Server" section of the sidebar. A large black arrow points from this highlighted area towards the "isomiReX" row in the main table.

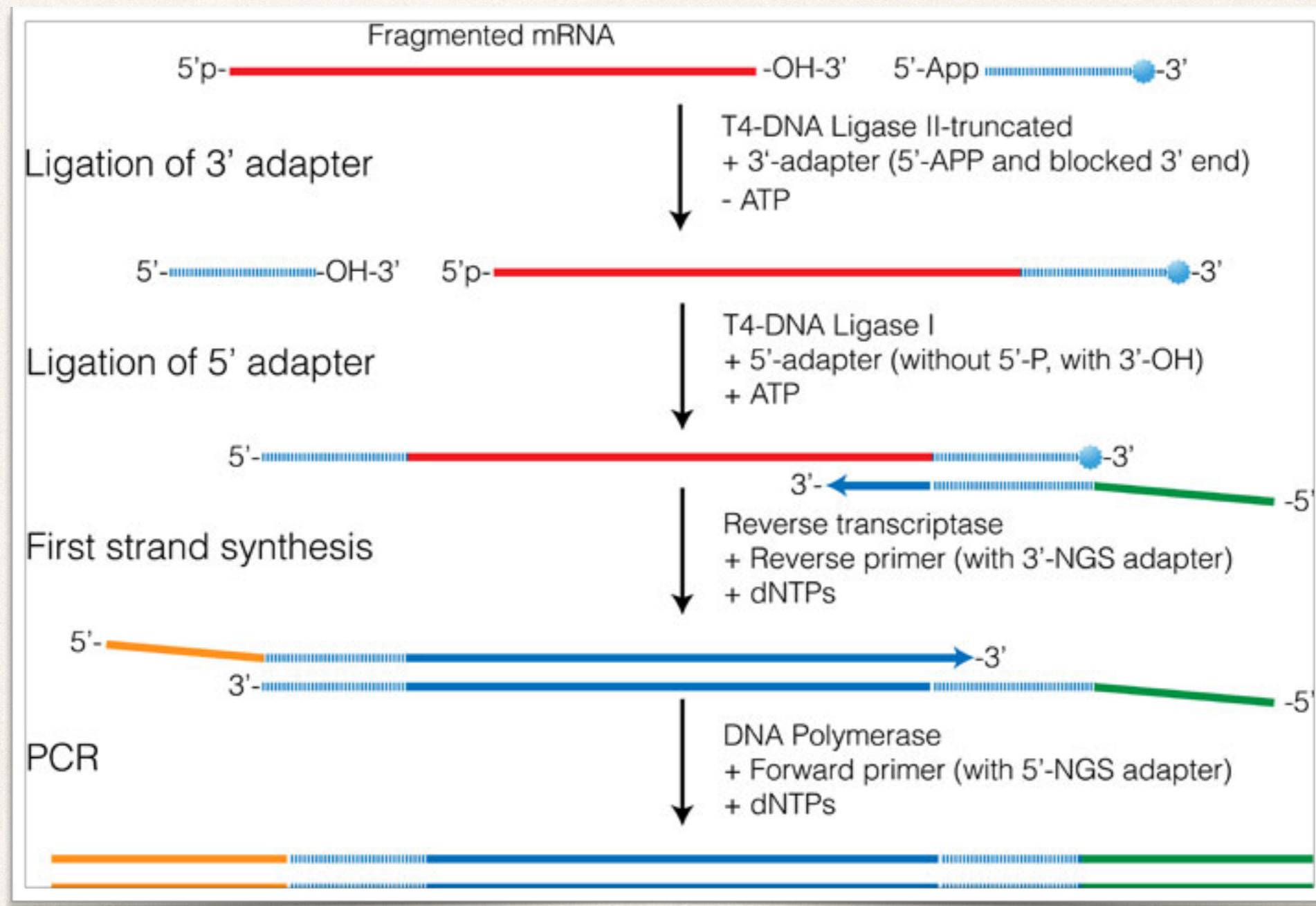
Samples suitable for benchmarking



Different protocols



Protocols

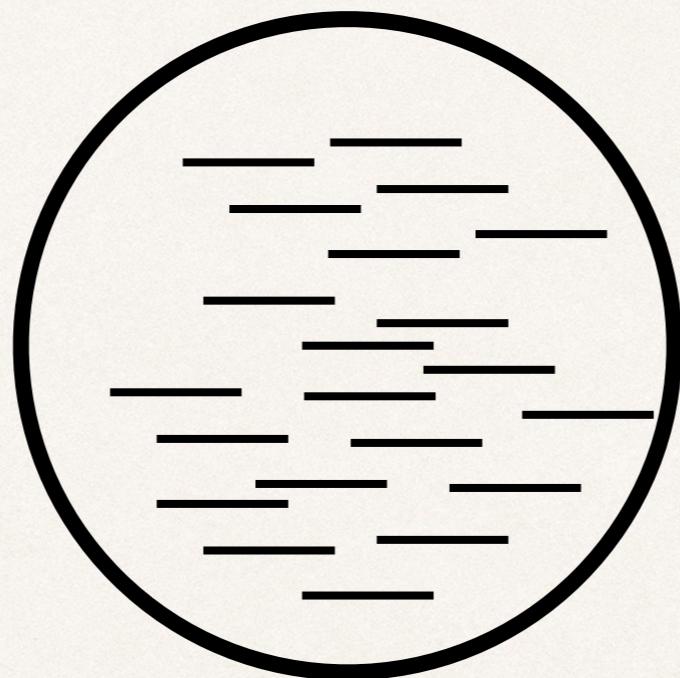


Variations

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

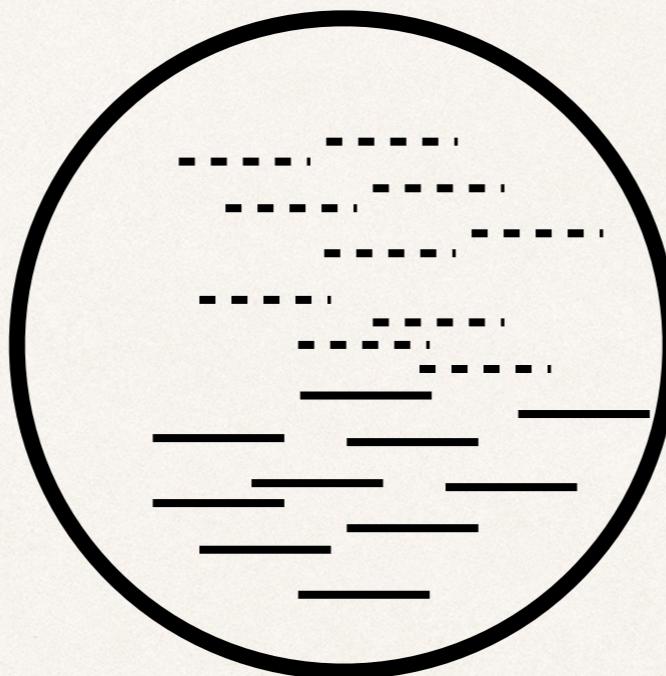
Methods and Metrics

Sample



Library size = 20

Sample

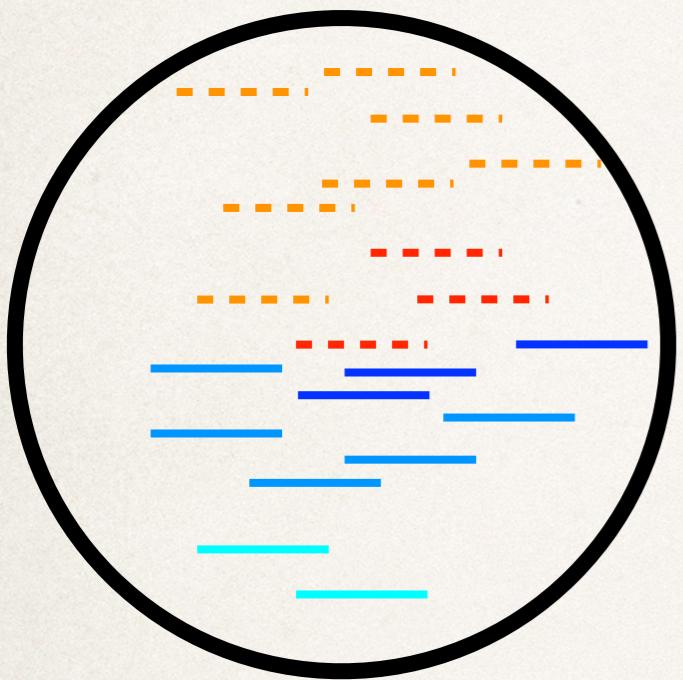


miRNA 1 = 10

miRNA 2 = 10

Library size = 20

Sample

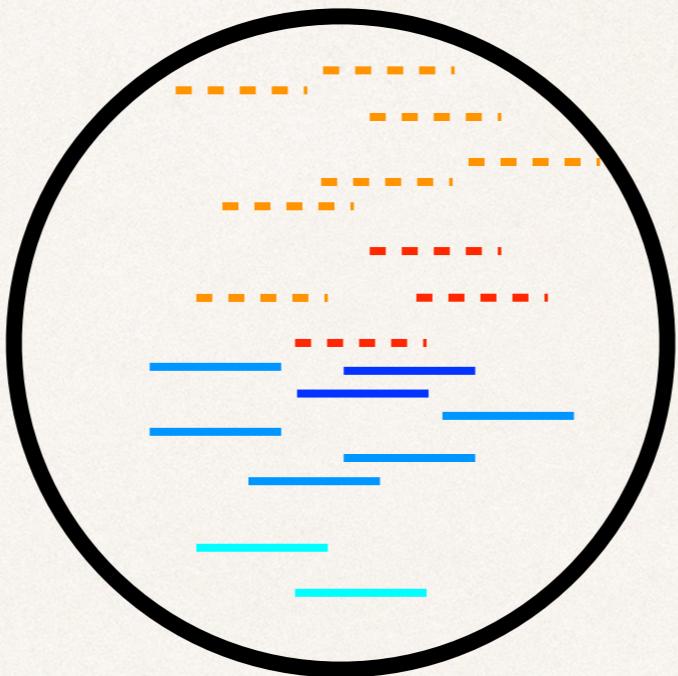


Library size = 20

miRNA 1 = X10
isomiR 1.1 = X7
(match perfectly spike-in)
isomiR 1.2 = X3
(NOT match perfectly spike-in)

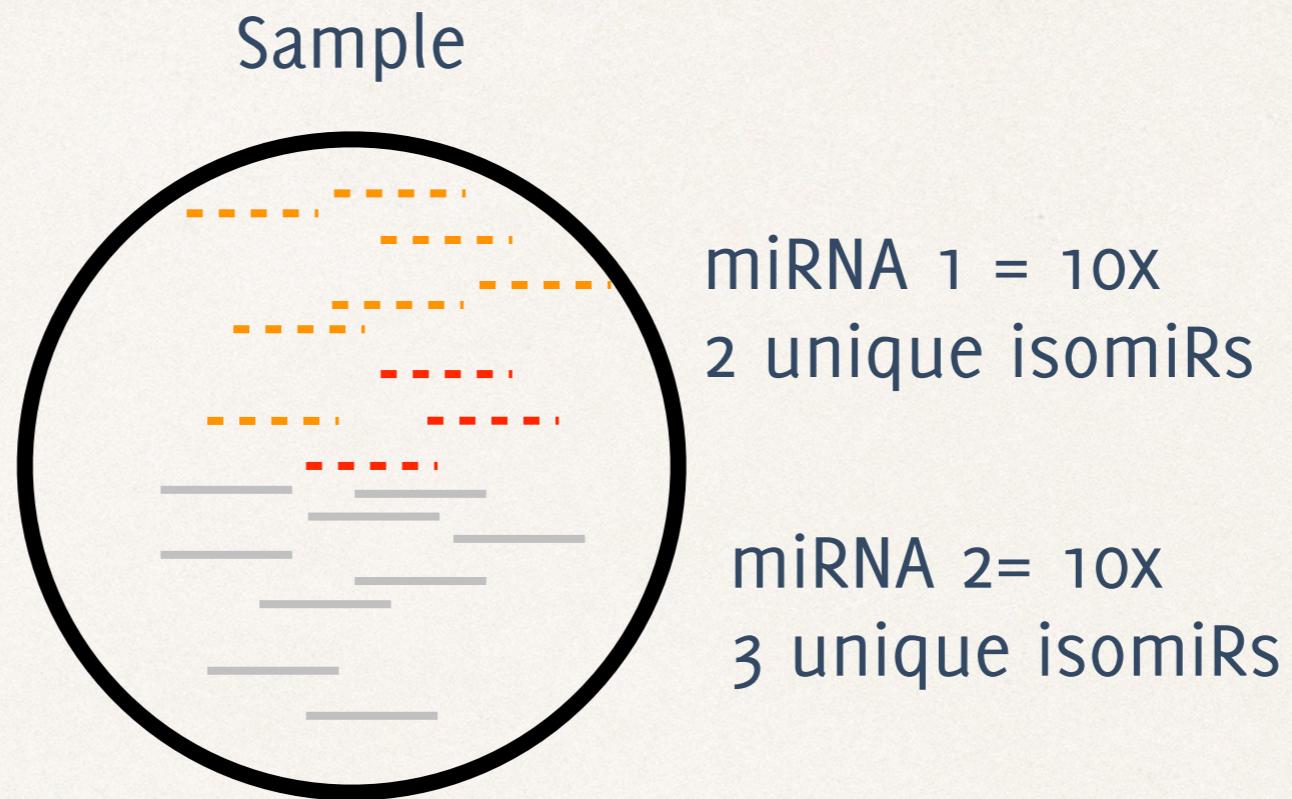
miRNA 2 = X10
isomiR 2.1 = X3
(match perfectly spike-in)
isomiR 2.2 = X5
(NOT match perfectly spike-in)
isomiR 2.2 = X2
(NOT match perfectly spike-in)

Sample



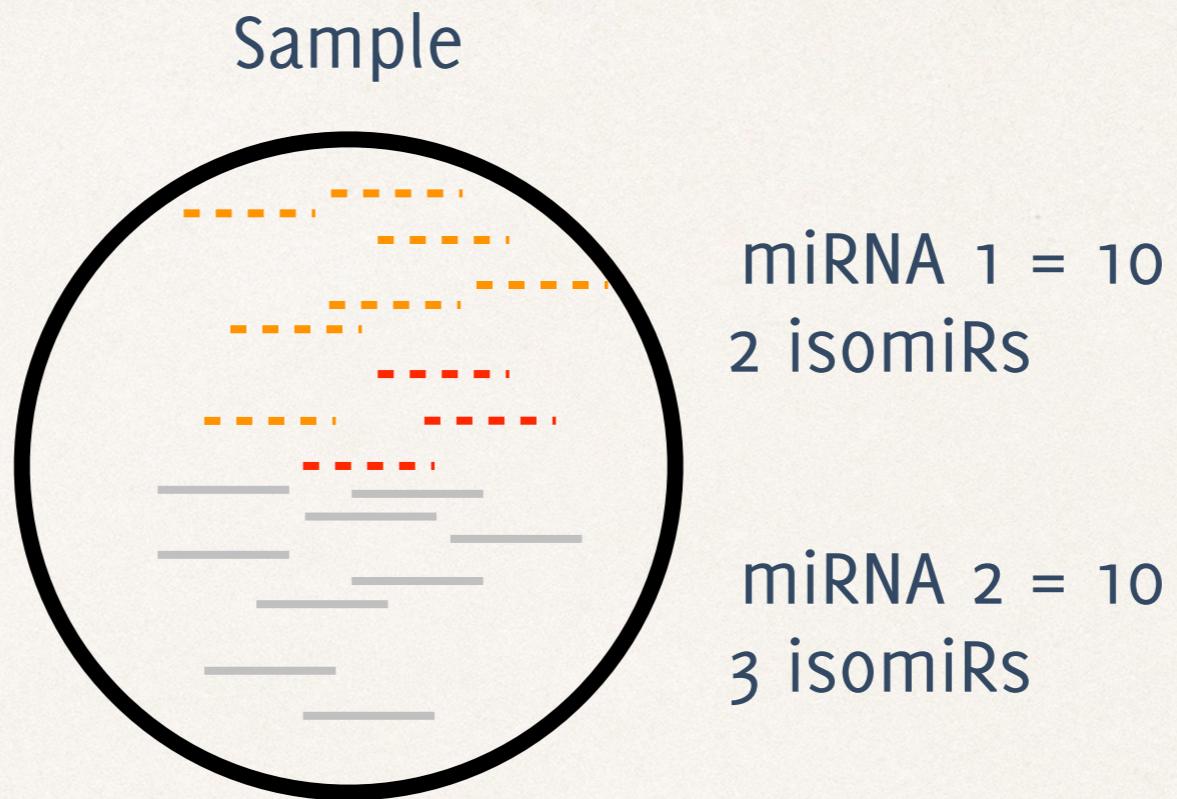
Library size = 20

There are 20 reads in total (lines), but 5 unique sequences (colors)



Library size = 20

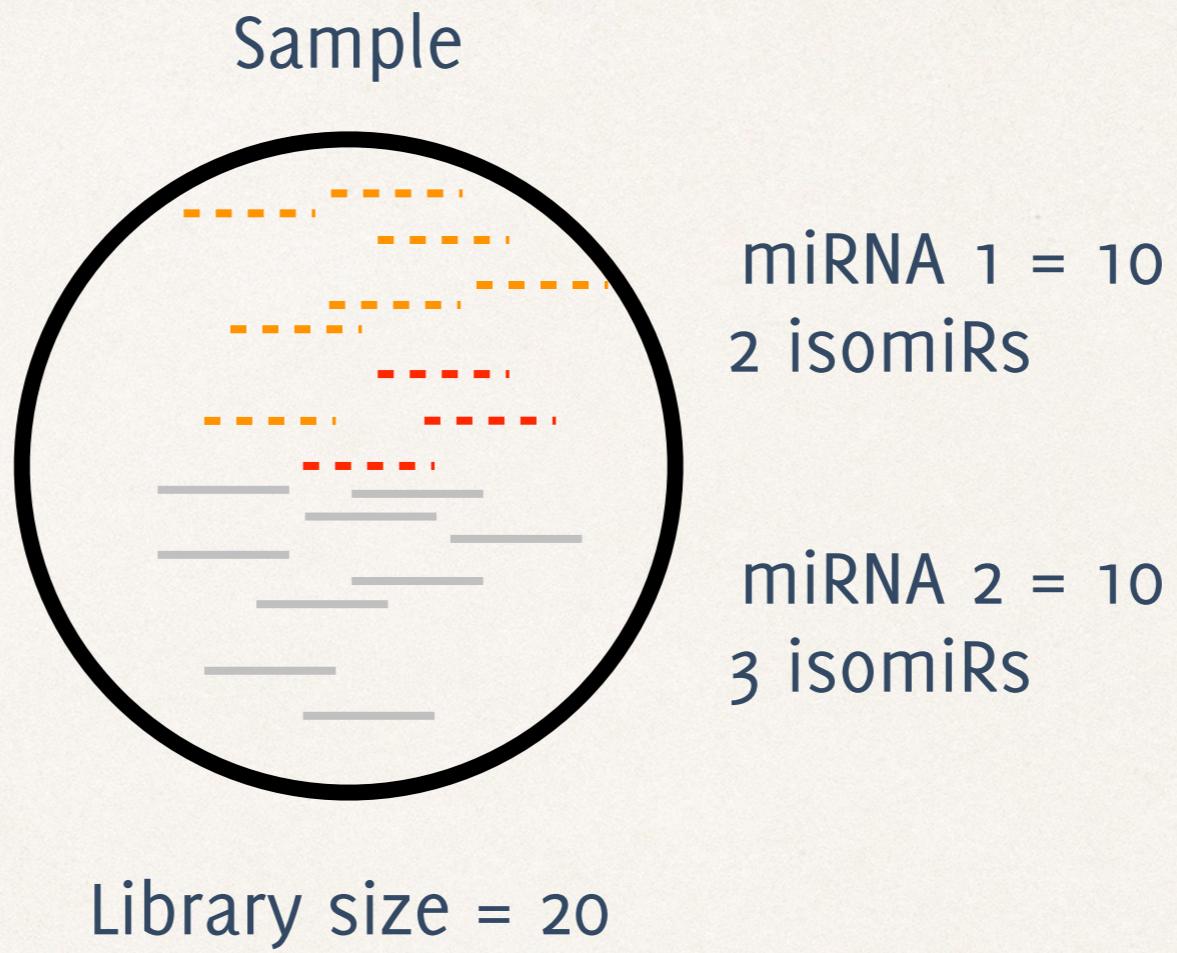
isomiR 1.1 is $7/20=35\%$ of the reads and $1/5=20\%$ of the sequences



Library size = 20

isomiR 1.1 is $1/2=50\%$ of the miRNA1 sequences

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

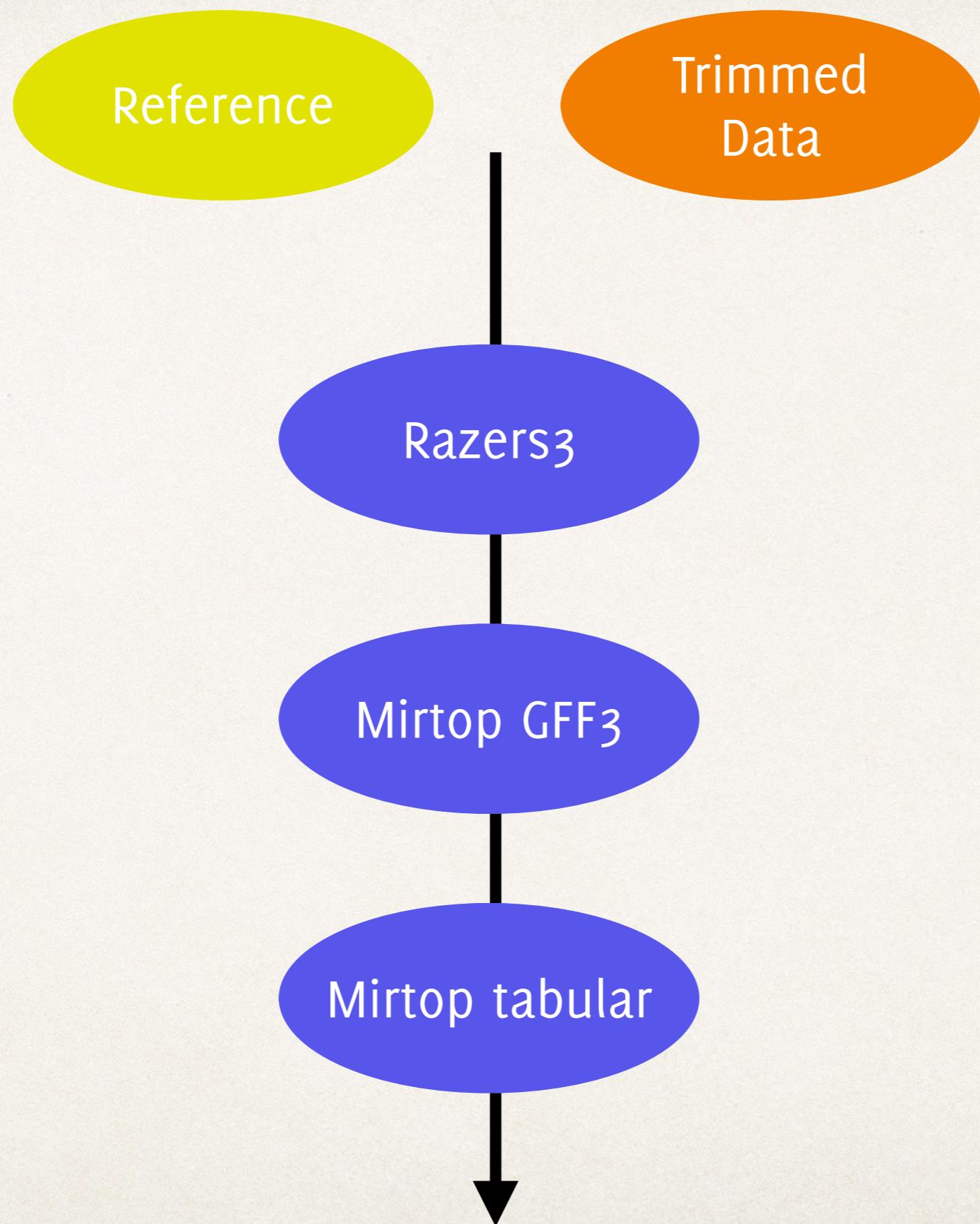


PCT = isomiR 1.1 is $1/5=20\%$ of the sequences

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

Pipeline razers3 + mirtop

snakemake

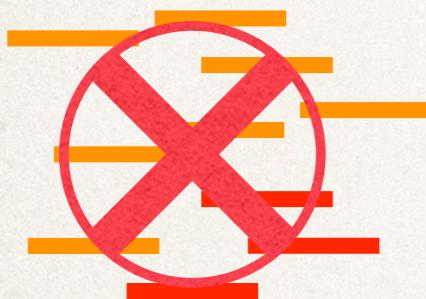


Data analysis - Filters

All sequences in a miRNA



NO

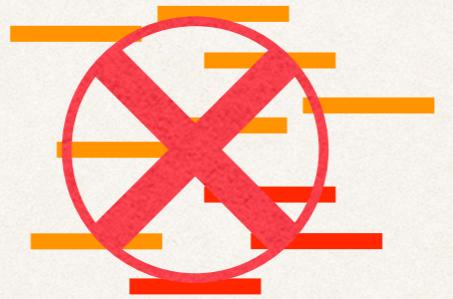


Human?

YES



NO

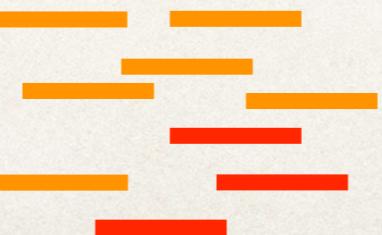


Is the spike-in
detected?

YES

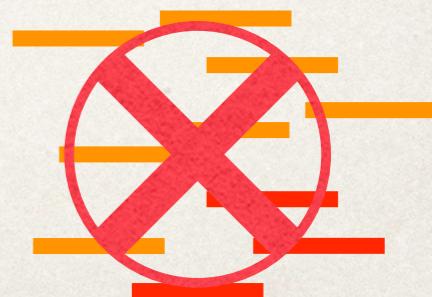


NO



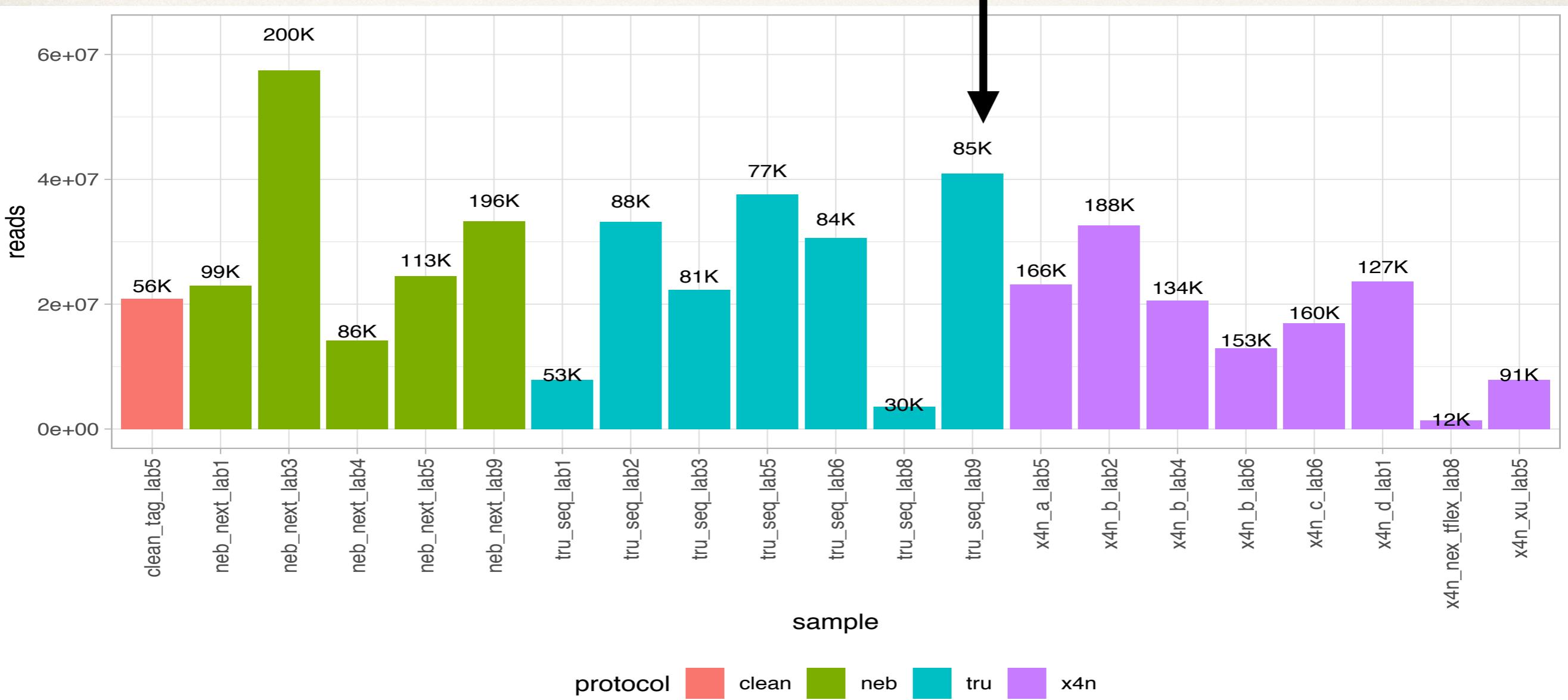
Any sequence in
group mapped to
other miRNA?

YES



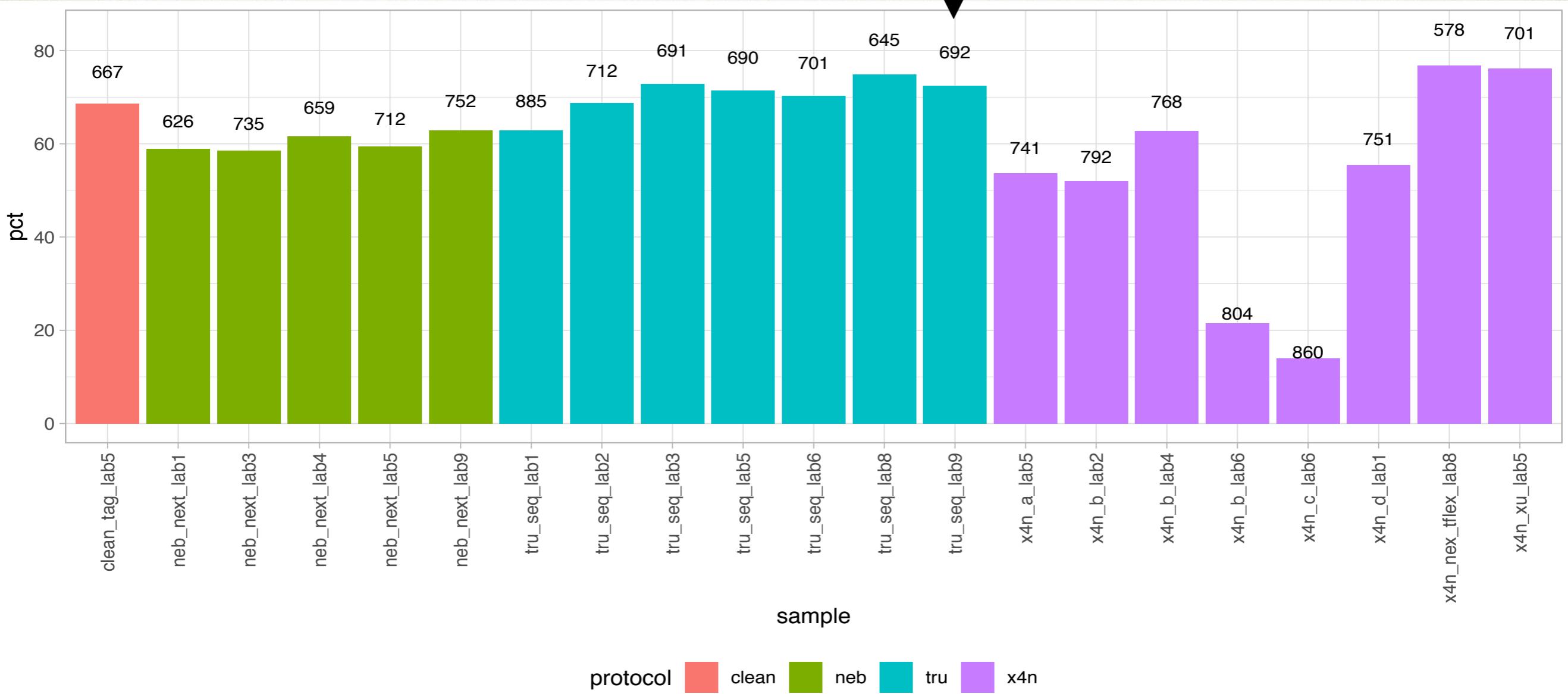
Tewari-synthetic: library size

40 million reads and 85K different sequences.

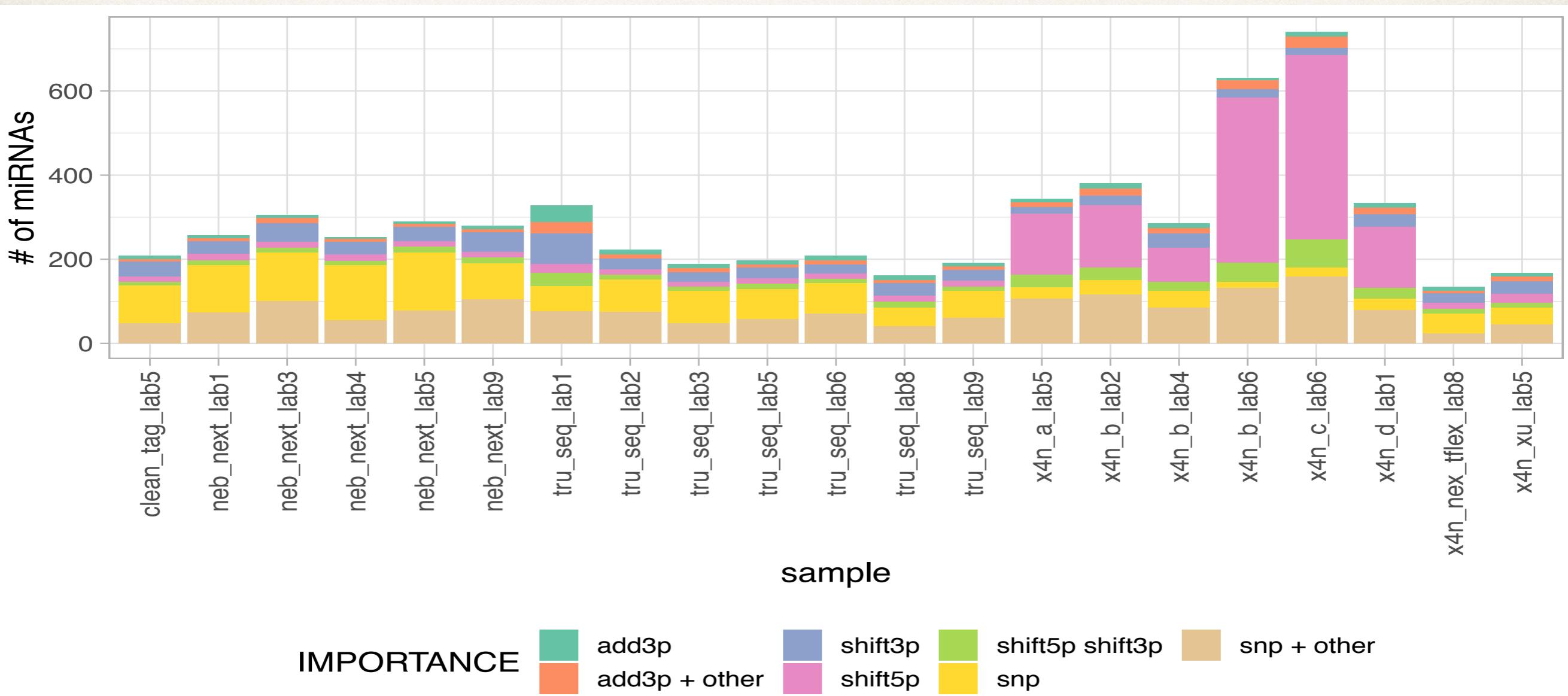


Tewari-synthetic: spike-ins are the top sequence?

70% (692 being total detected miRNAs) of miRNAs
with top sequence to be the expected one.

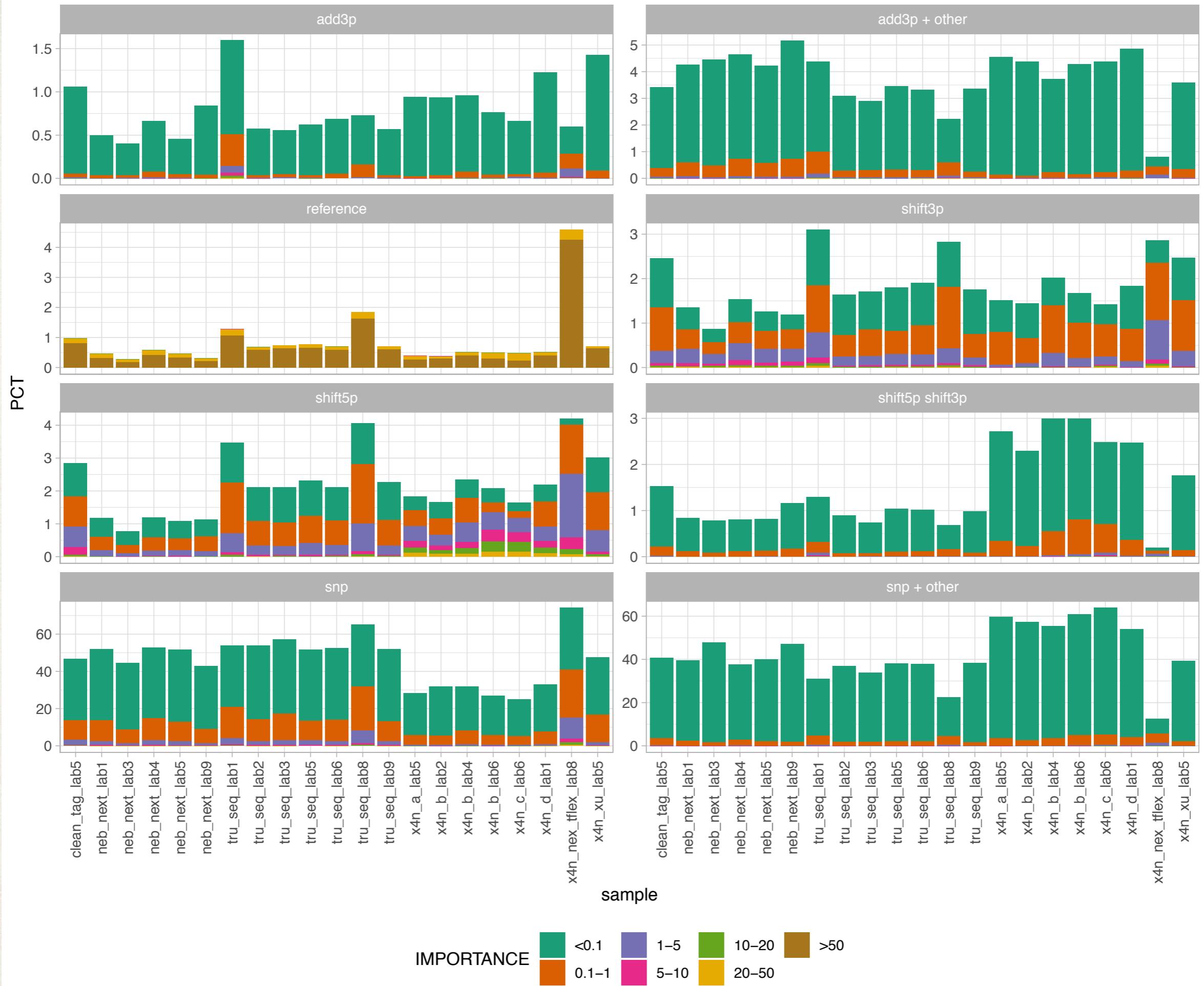


Most abundant type of isomiR for each miRNA without the spike-in as the most abundant

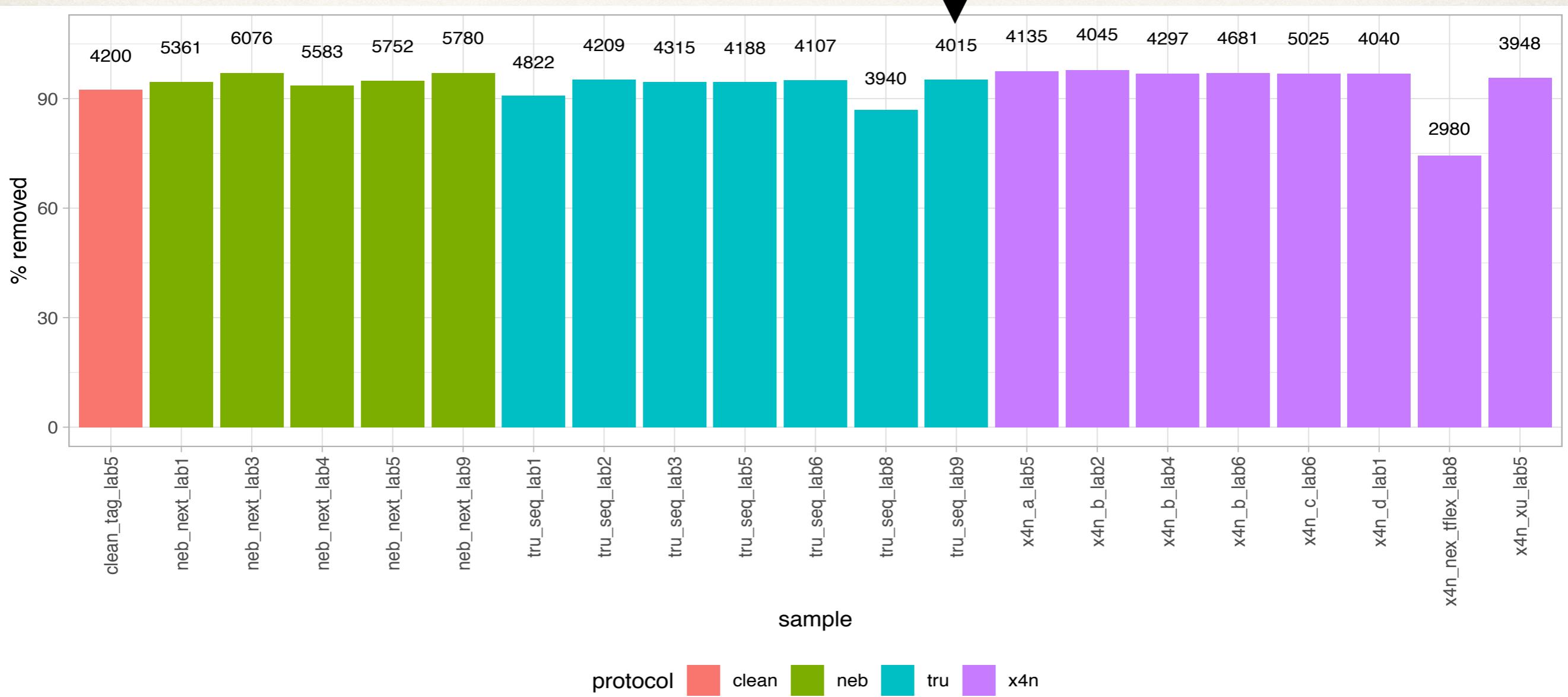


Tewari-synthetic: Importance of the 'isomiRs' detected

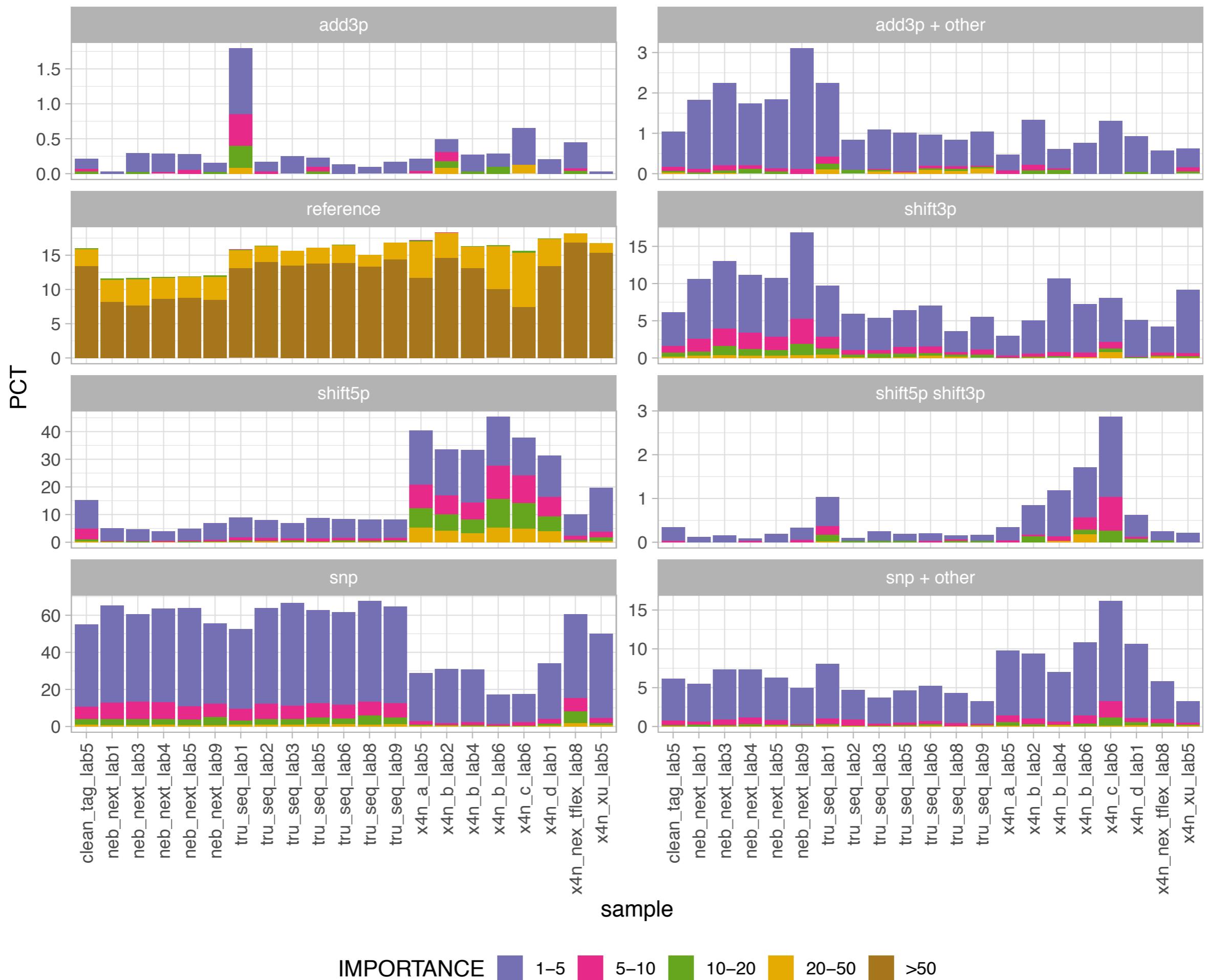
isomiRs importance by type

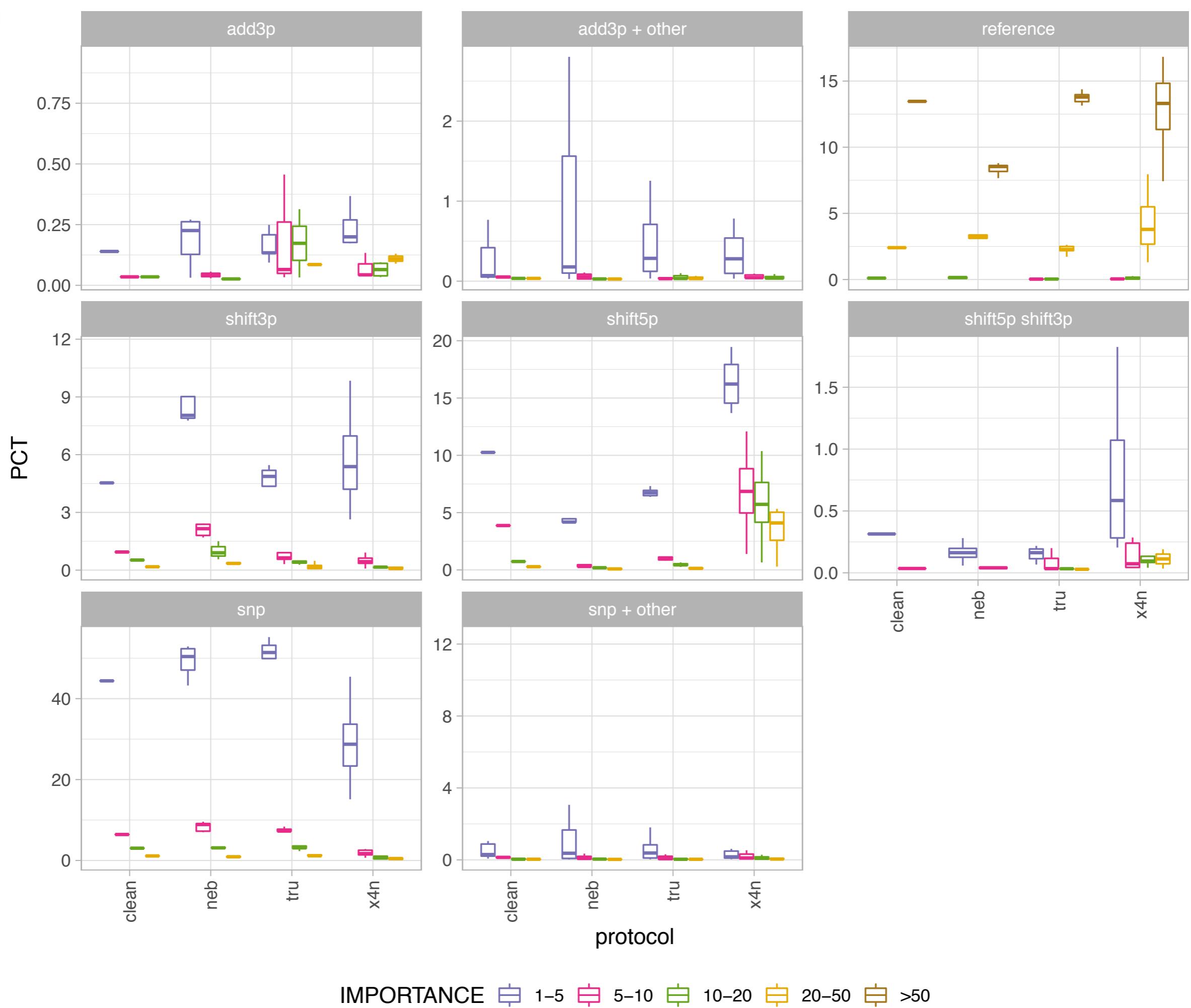


From 85K to 4K unique sequences.
90% of sequences are removed.

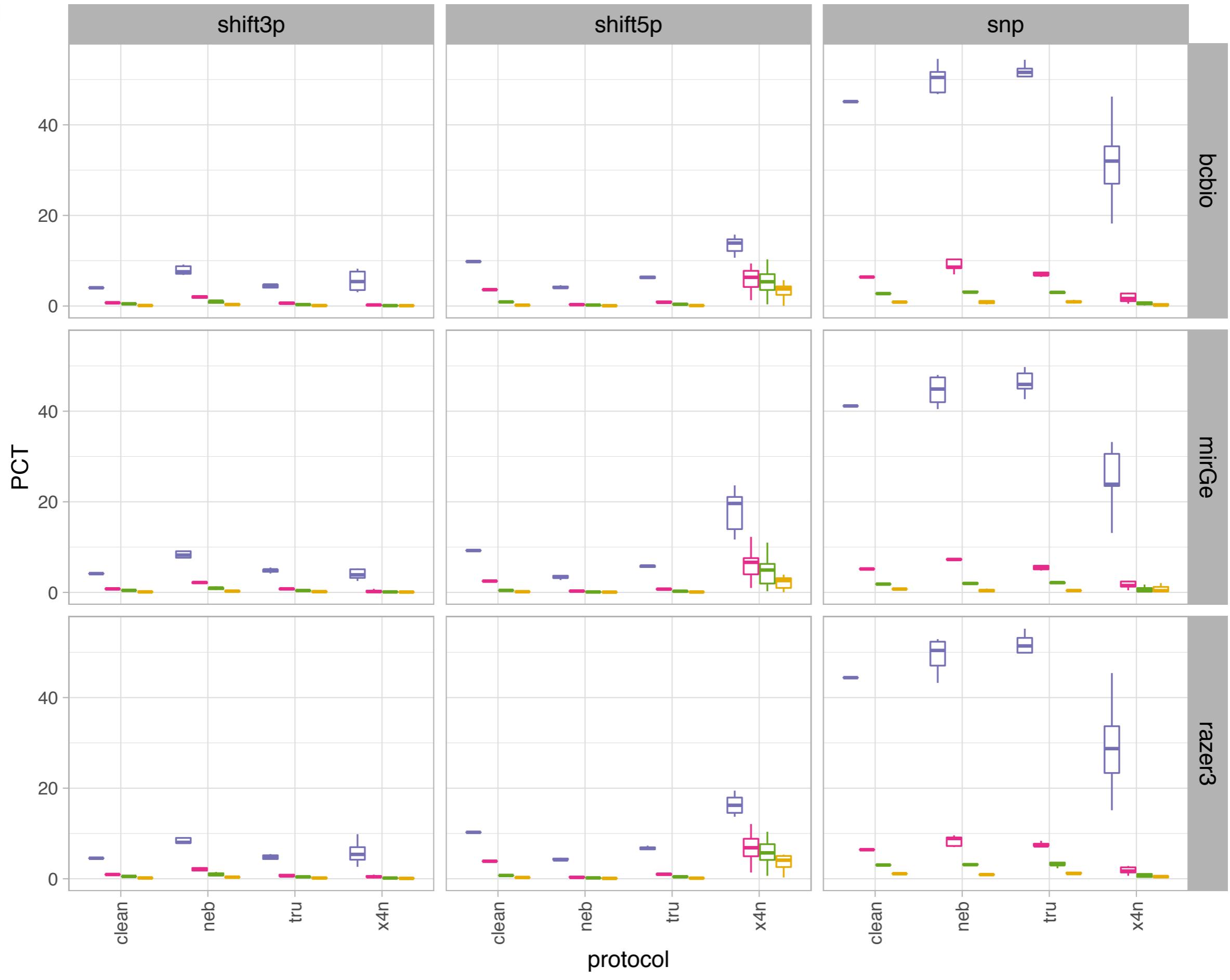


isomiRs importance by type with pct > 1

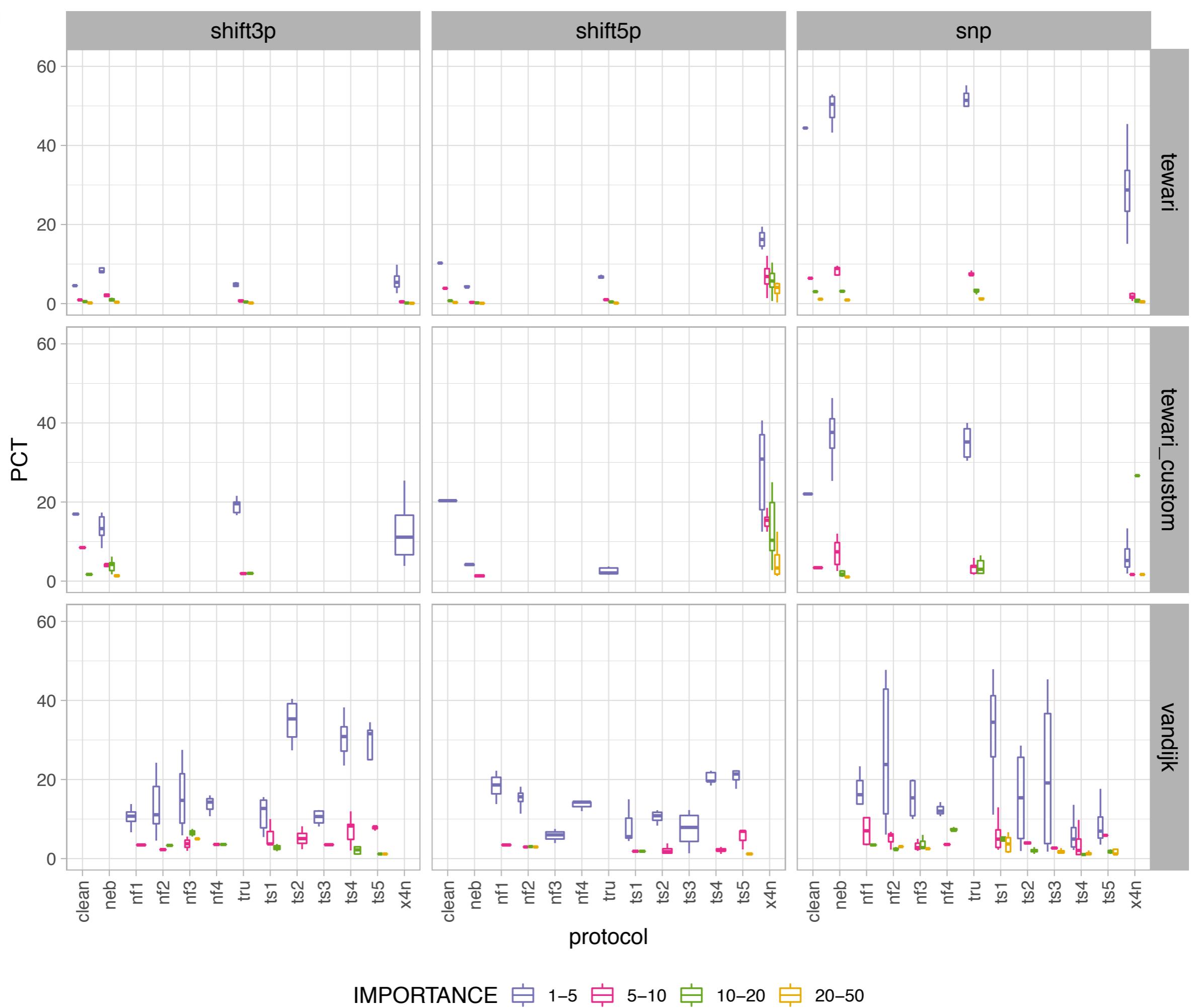




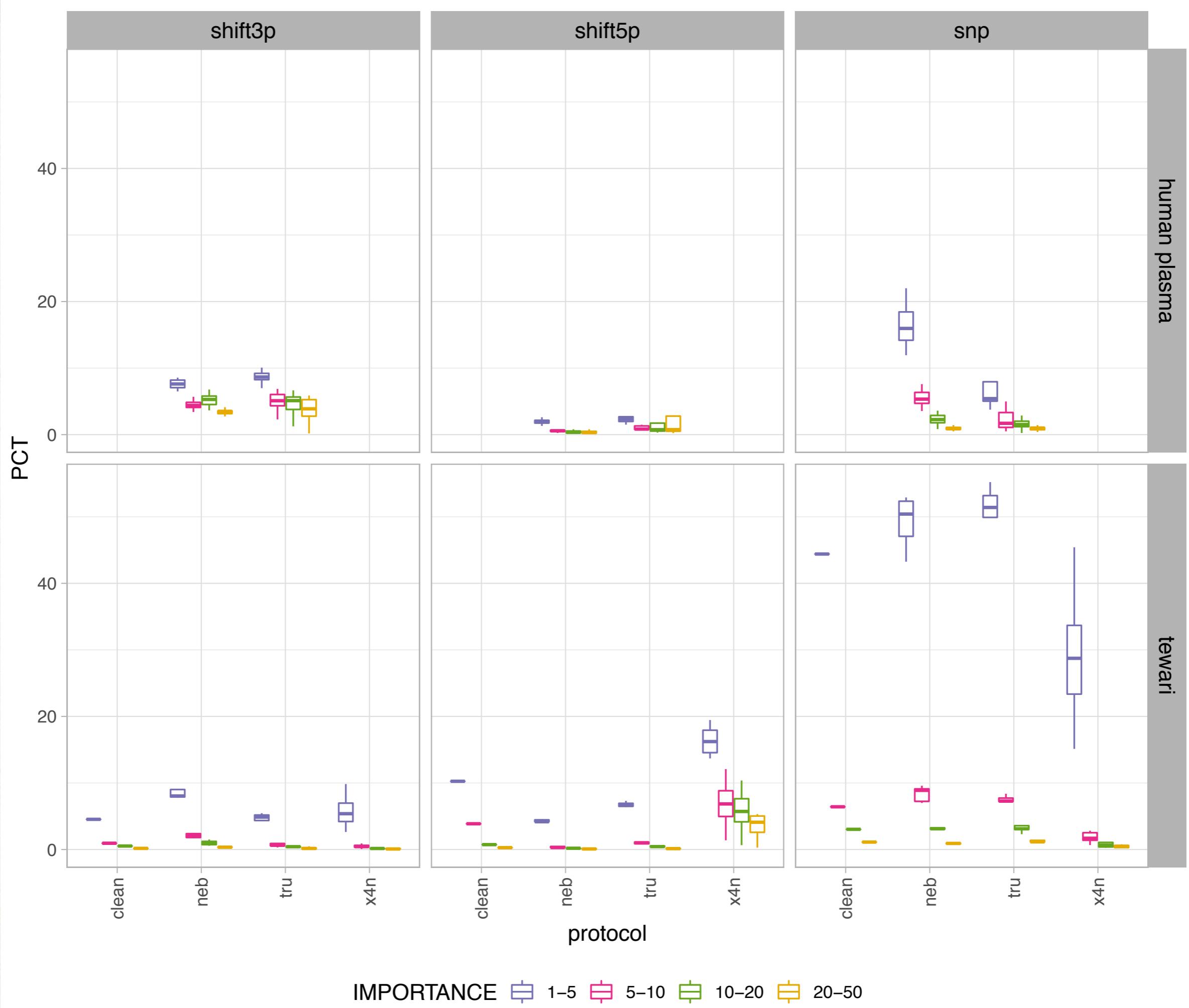
Tewari-synthetic: Other tools



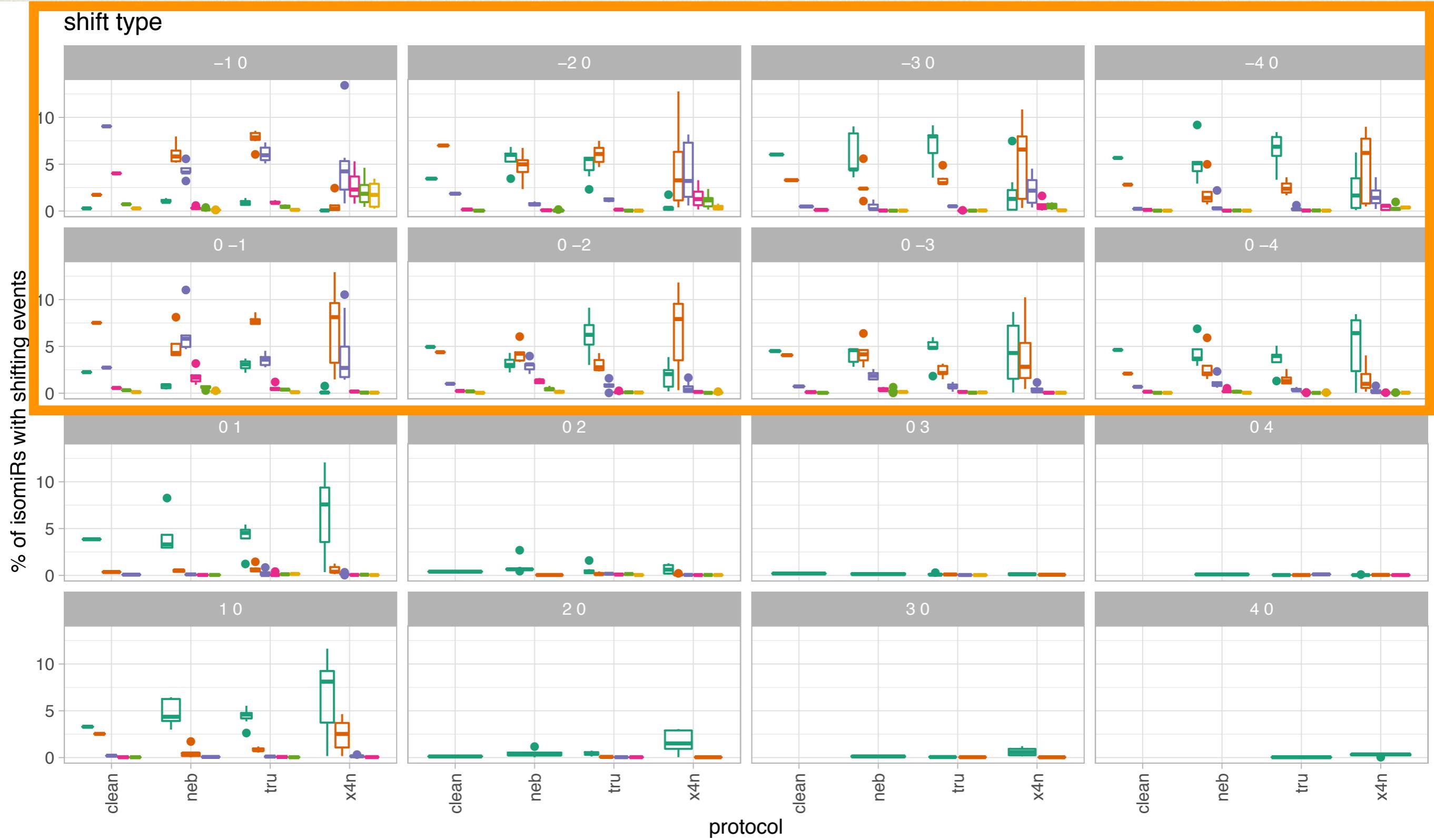
Tewari-synthetic: Other data



Tewari: synthetic vs human plasma

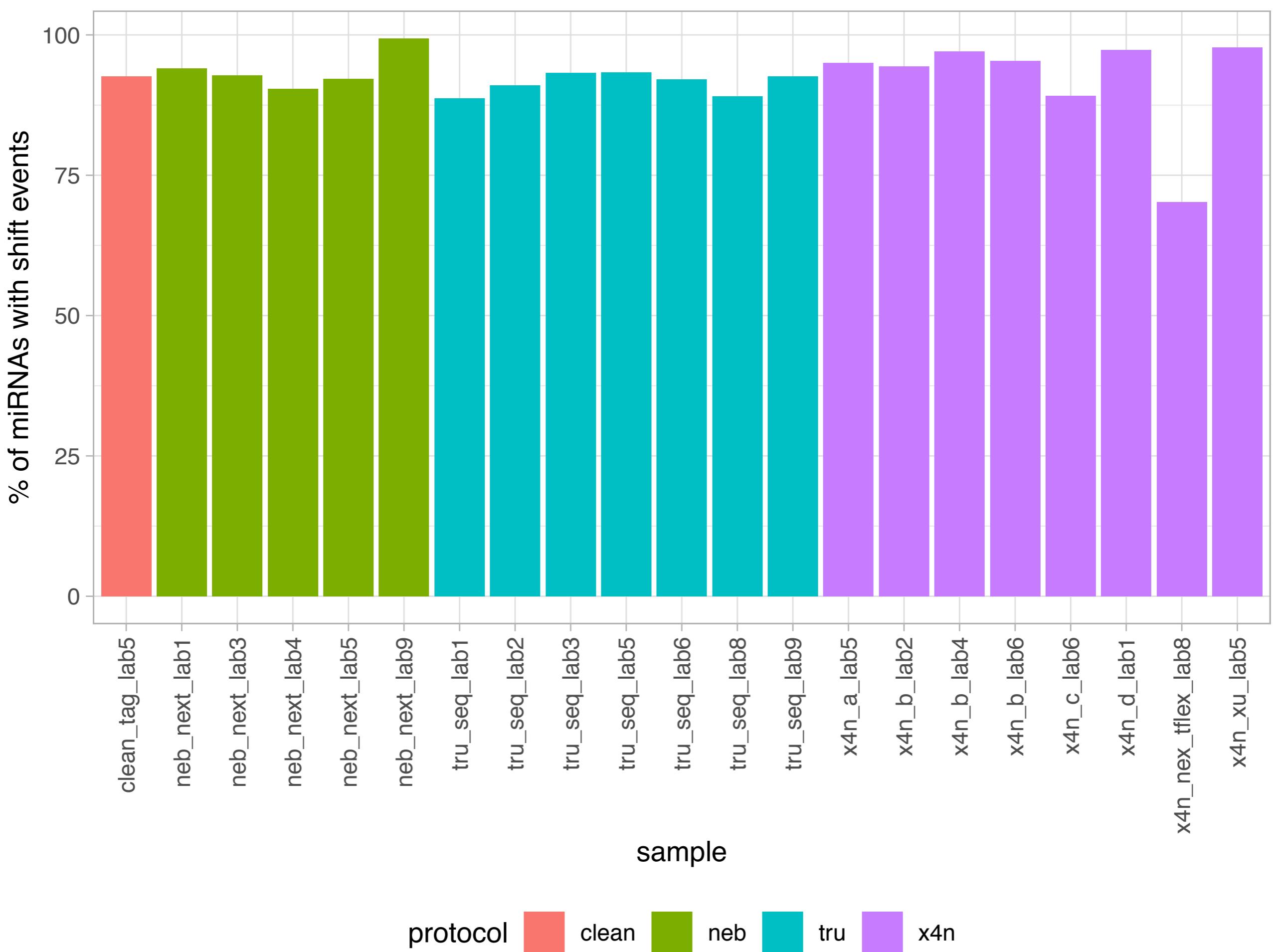


Tewari-synthetic: Shift type



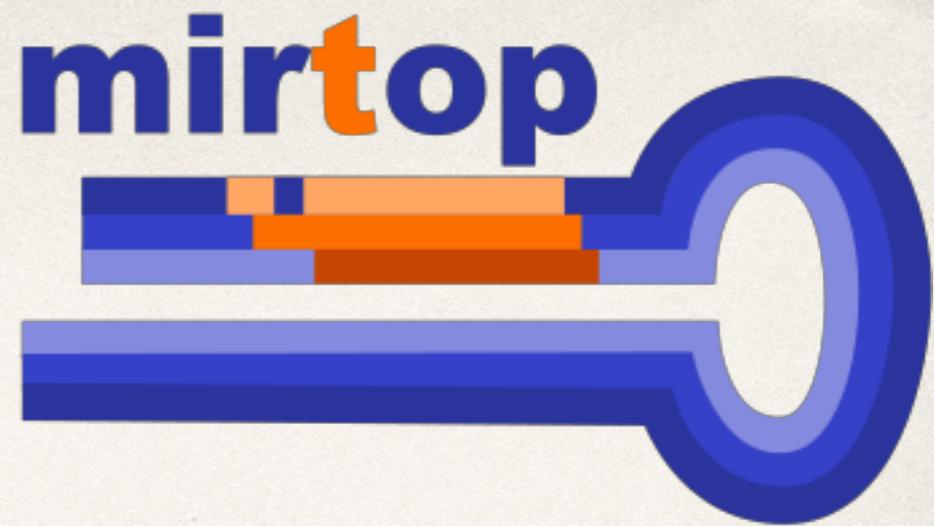
pct_cat

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



Summary

- ✓ Each miRNA generates a diversity of isomiRs:
 - ✓ 90% contribute to < 1% of the miRNA abundance
 - ✓ 10% enrichment on truncations at both sides
 - ✓ Independent of pipelines and data sets
- ✓ 90% of miRNAs affected
- ✓ Custom 4N protocols perform worst



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

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

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



HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH



Harvard Medical School
Research Computing

JOHNS HOPKINS
SCHOOL of MEDICINE

SciLifeLab

NIH
National Institutes of Health

UNIVERSITY OF
OREGON

Thomas Desvignes
Phillipe Loher
Karen Ellbeck
Bastian Fromm
Gianvito Urgese
Isidore Rigoutsos
Michael Hackenberg
Ioannis S. Vlachos
Marc K. Halushka

BROAD
INSTITUTE

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

CRG
Centre for Genomic
Regulation



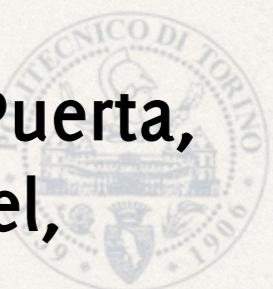
Jefferson

HOME OF SIDNEY KIMMEL MEDICAL COLLEGE



BRIGHAM HEALTH

BRIGHAM AND
WOMEN'S HOSPITAL



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

BC

Provincial Health Services Authority

MIRNA WARS: THE NEW HOPE

Thank you!



