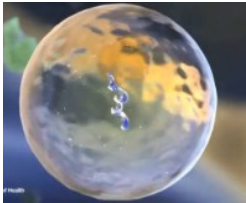
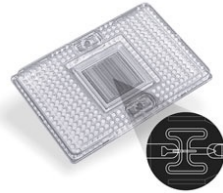


a)

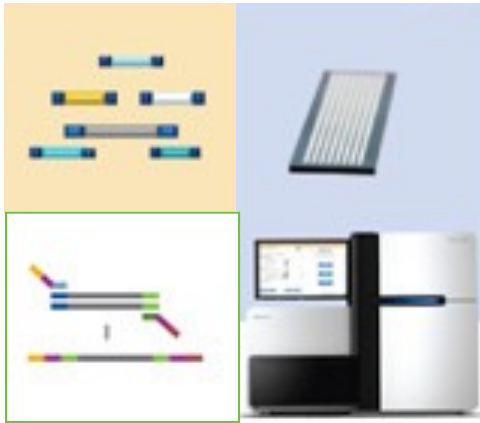
low-input
high-contamination
samples



extracellular
vesicles



single-cell



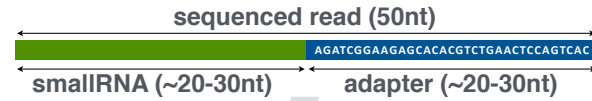
b)

standard formats

.fq .fq.gz .sra

input

remove adapters
+ primers



read-quality filter

fastQC read length analysis

preprocessing

for reads >15nt

map using
bowtie2
(faster, allows
indels)

optional: map to spike-in library

unmapped reads

filtering

map to UniVec contaminants

unmapped reads

QC

map to 45S, 5S, & mitochondrial rRNAs

unmapped reads

bowtie1 alignments



endogenous
alignment

parallel mapping
using bowtie1

hierarchical
quantification
based on user
preferences

engine workflow:

1. assign reads to highest rank library
2. resolve genome vs. library-only alignments
3. overlap mature+hairpin miRNA alignments
4. re-format gencode IDs
5. [count random barcodes]

accepted alignments & [barcode statistics]

- 1 miRNA counts
- 2 tRNA counts
- 3 piRNA counts
- 4 gencode counts
- 5 circRNA counts

(human
&
mouse)

unmapped reads

map using
bowtie2

map to endogenous genome (again!)
and repetitive elements

filtering

unmapped reads

map using
bowtie & STAR

map to exogenous
miRNAs and rRNAs

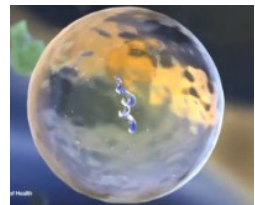
exogenous
alignment

unmapped reads

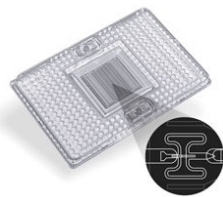
map using STAR

map to all genomes in ensembl & NCBI

a) low-input high-contamination samples



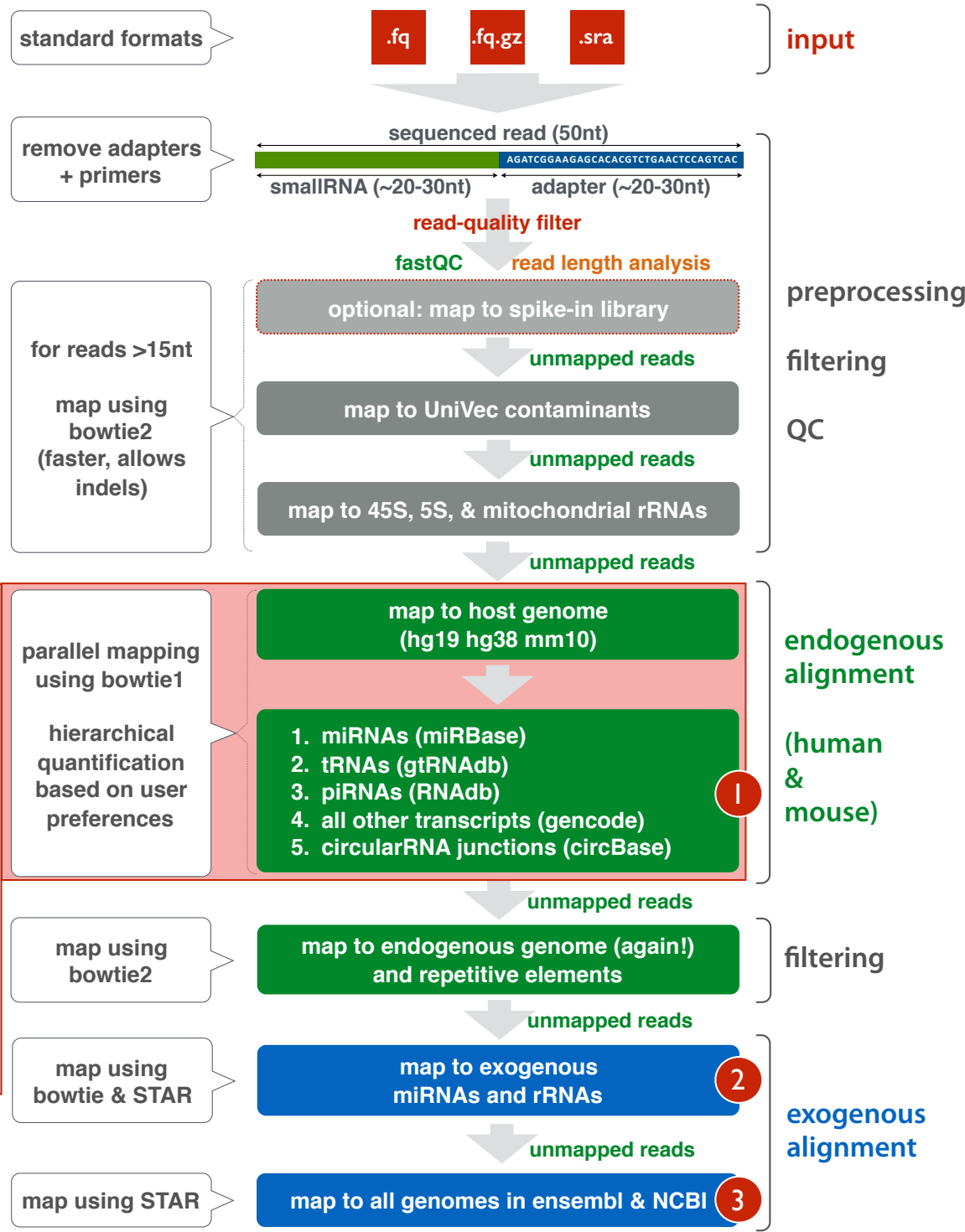
extracellular vesicles



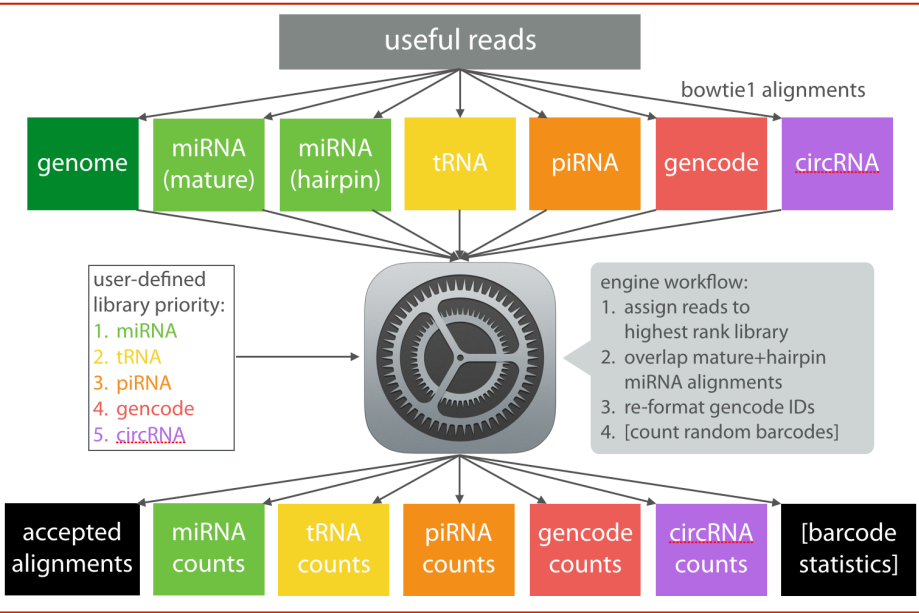
single-cells

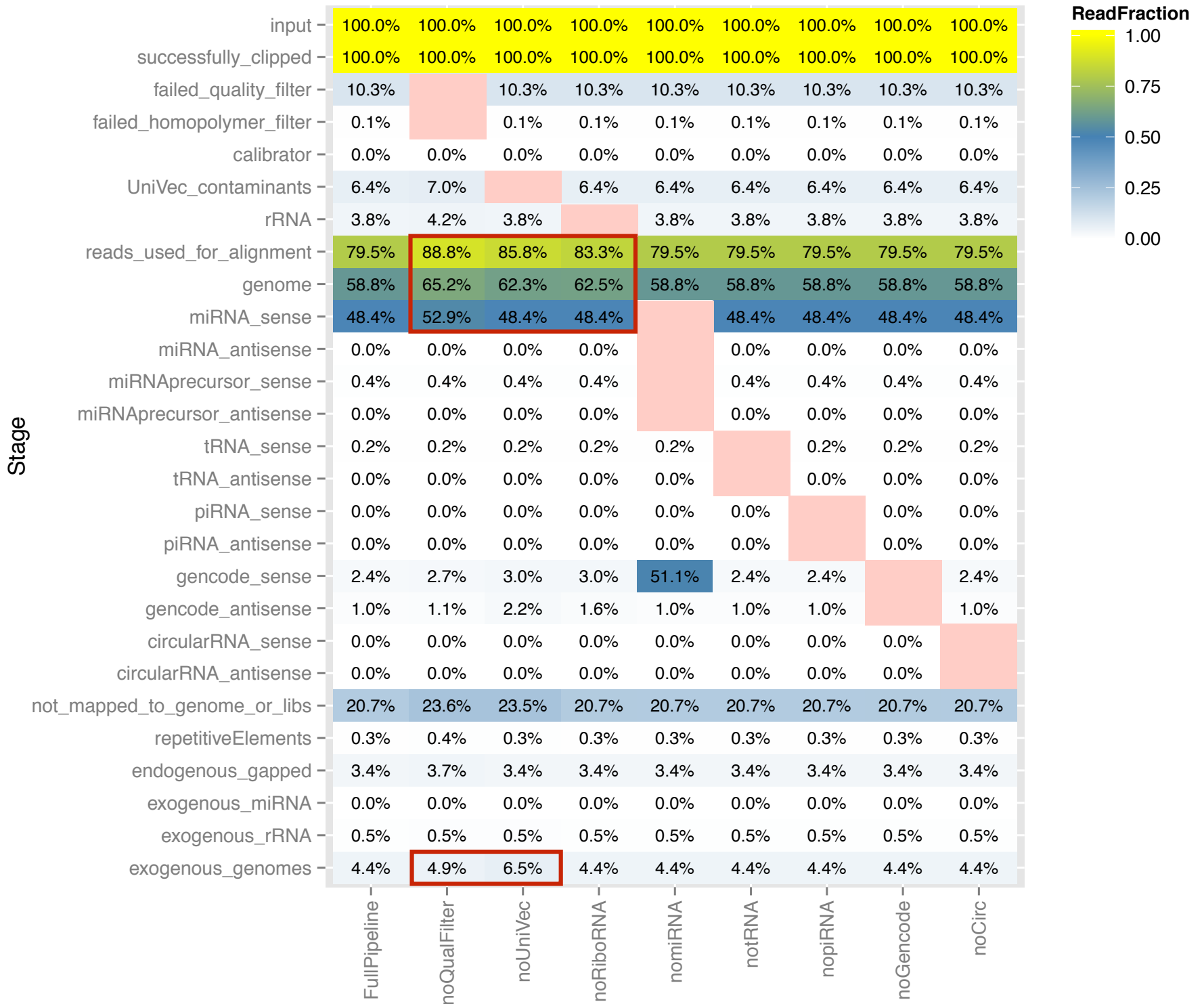
b)

exceRpt pipeline



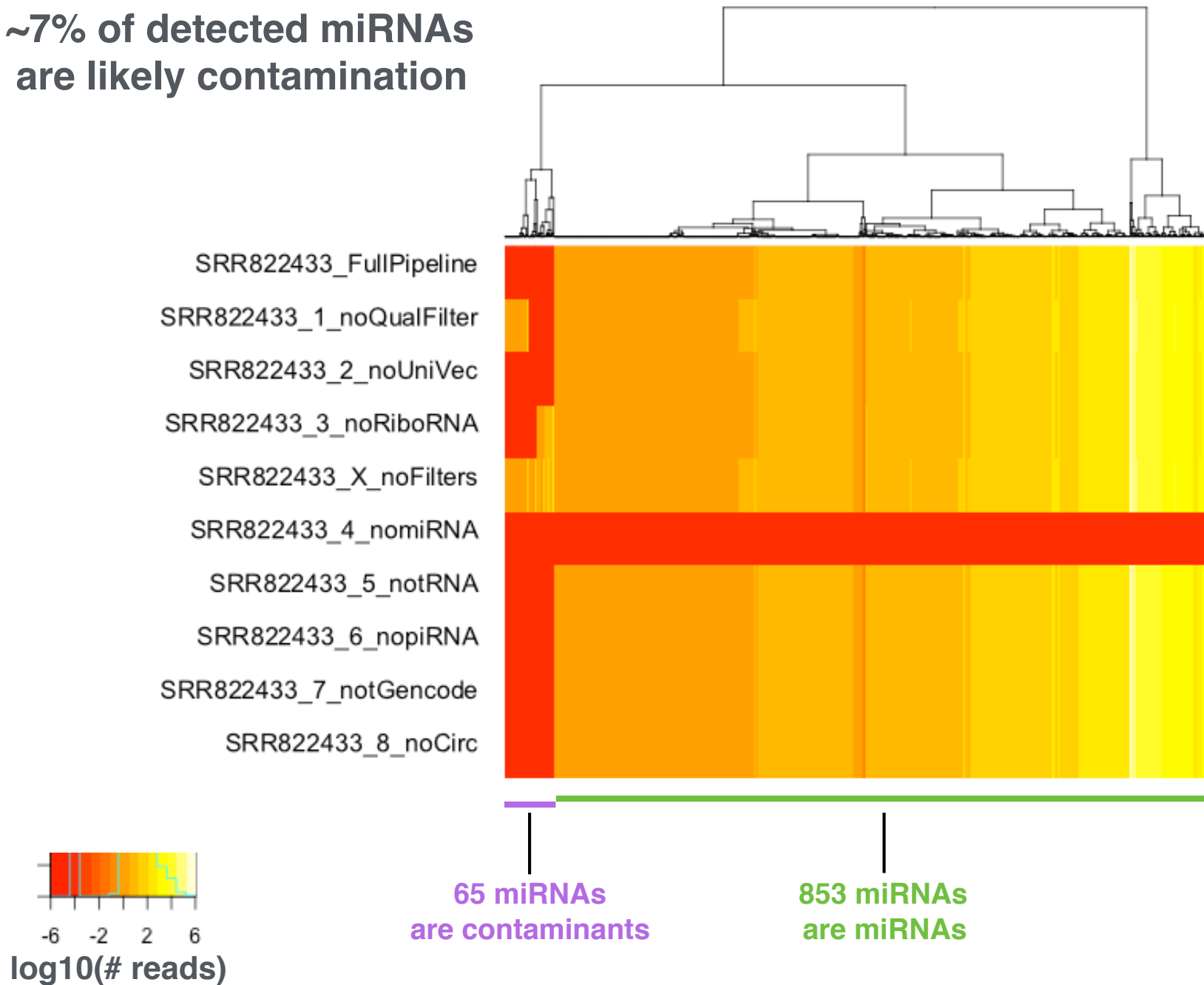
c)





vulnerability of smallRNA quantification to contamination

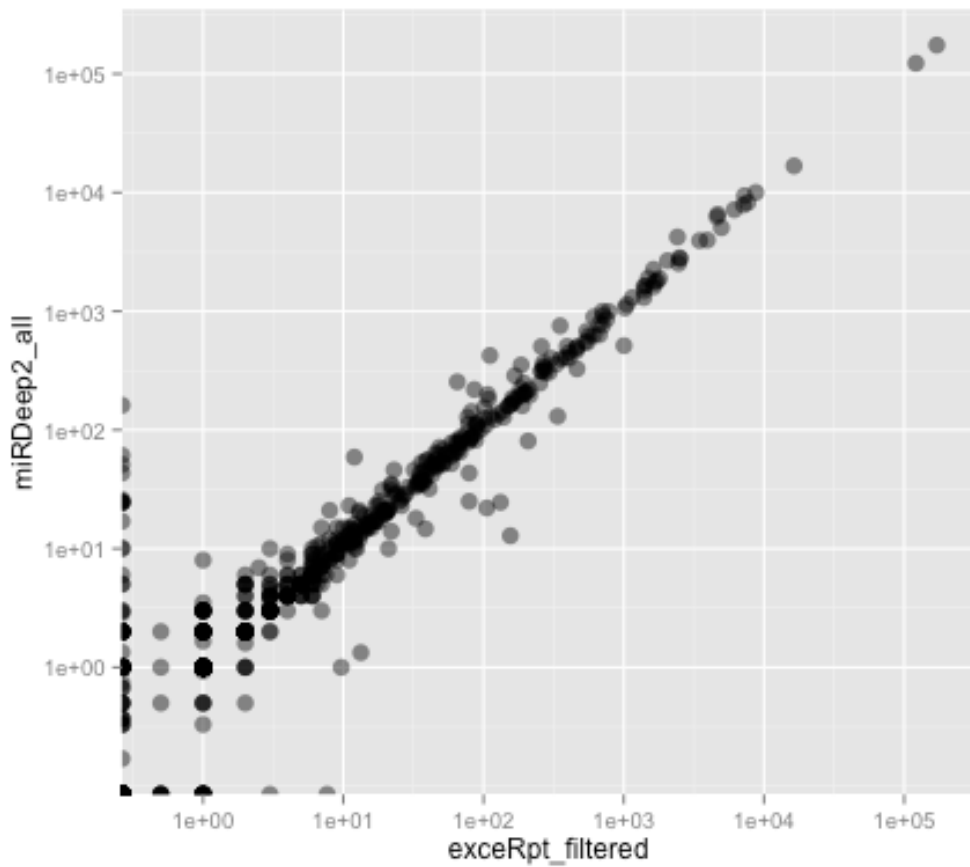
~7% of detected miRNAs are likely contamination



miRNA quantification: exceRpt vs. miRDeep2

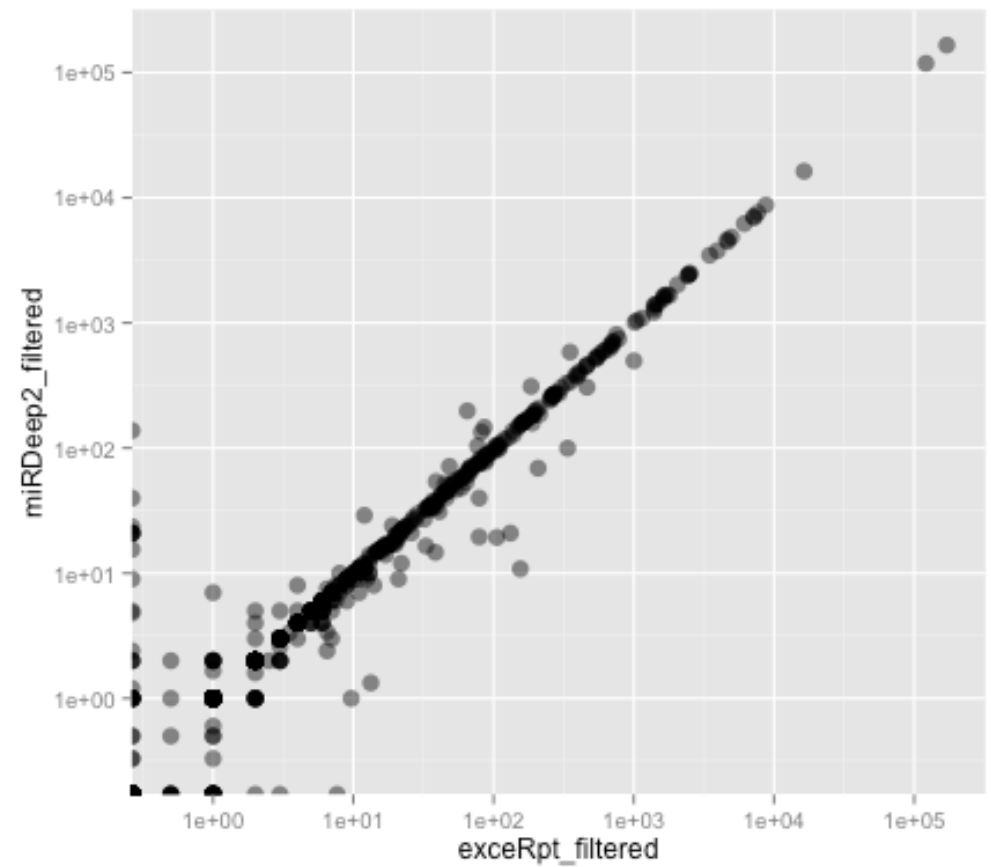
miRDeep2 (all reads)
vs exceRpt

cor = 0.999804



miRDeep2 (exceRpt filtered reads)
vs exceRpt

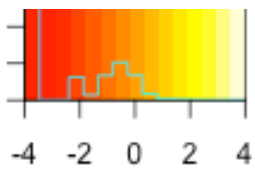
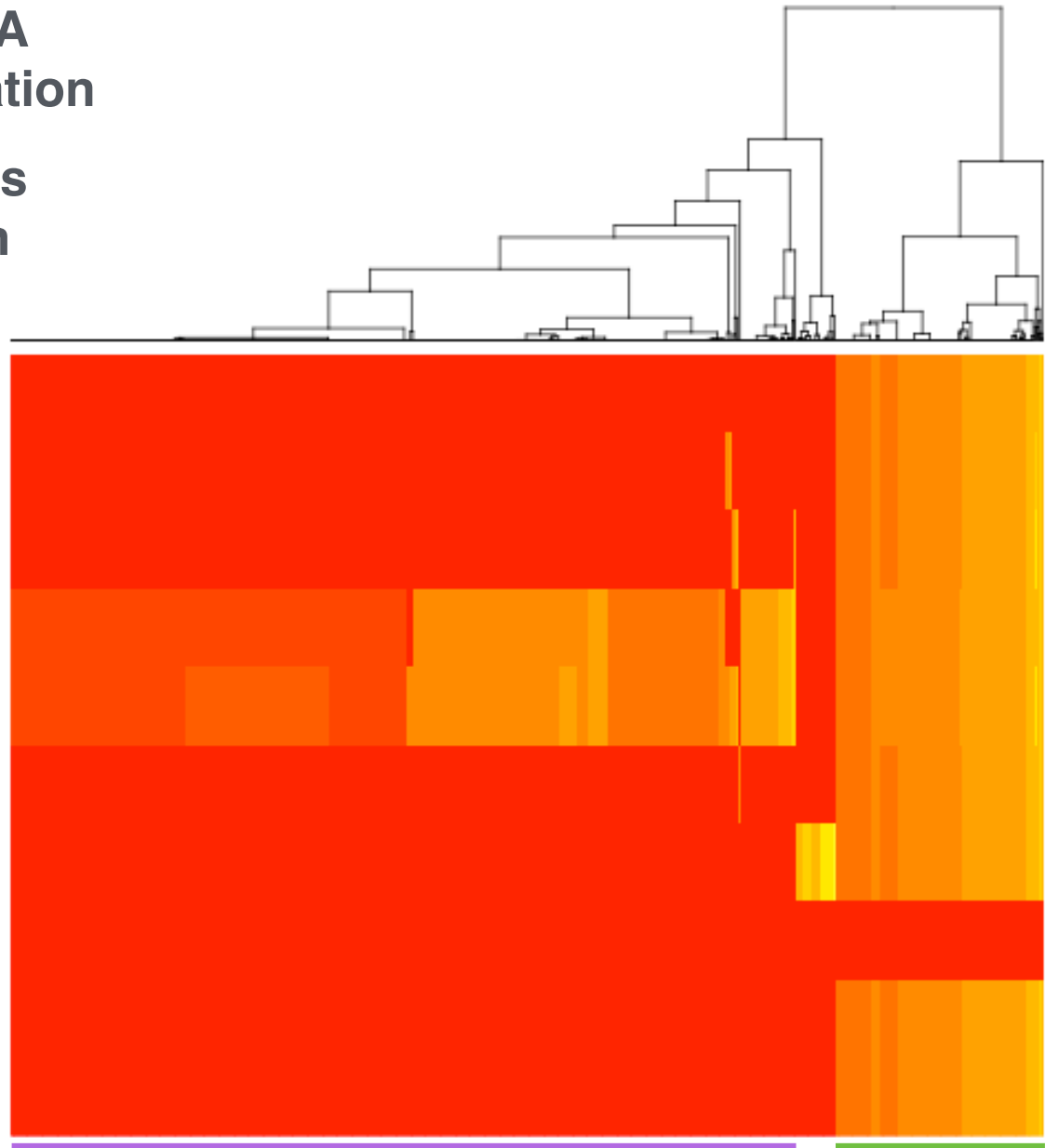
cor = 0.999985



vulnerability of smallRNA quantification to contamination

~80% of detected piRNAs are likely contamination

SRR822433_FullPipeline
SRR822433_1_noQualFilter
SRR822433_2_noUniVec
SRR822433_3_noRiboRNA
SRR822433_X_noFilters
SRR822433_4_nomiRNA
SRR822433_5_notRNA
SRR822433_6_nopiRNA
SRR822433_7_notGencode
SRR822433_8_noCirc



log₁₀(# reads)

763 piRNAs
are contaminants

39 piRNAs
are tRNAs

202 piRNAs
are piRNAs

to-do

- compare to cellular prep
- compare to single cell with gencode RNAs?