exRNAQC011: RNA purification kit performance (small RNA quantification), part of phase 1

Annelien Morlion - on behalf of exRNAQC Consortium

15/12/2020

Contents

1   Introduction

1.1   Experimental setup

For the evaluation of the different RNA isolation kits in the first phase of exRNAQC, blood was drawn from 1 healthy volunteer. We tested 8 different kits:

- miRNeasy Serum/Plasma Kit (abbreviated to MIR; Qiagen, 217184)
- miRNeasy Serum/Plasma Advanced Kit (abbreviated to MIRA; Qiagen, 217204)
- mirVana PARIS Kit (abbreviated to MIRV and MIRVE; Life Technologies, AM1556)
- NucleoSpin miRNA Plasma Kit (abbreviated to NUC; Macherey-Nagel, 740981.50)
- QIAamp ccfDNA/RNA Kit (abbreviated to CCF; Qiagen, 55184)
- Plasma/Serum Circulating and Exosomal RNA Purification Kit/Slurry Format (abbreviated to CIRC; Norgen Biotek Corp., 42800)
- Maxwell RSC miRNA Plasma and Exosome Kit (Promega, AX5740) in combination with the Maxwell RSC Instrument (abbreviated to MAX; Promega, AS4500)
- MagNA Pure 24 Total NA Isolation Kit (Roche, 07 658 036 001) in combination with the MagNA Pure instrument (abbreviated to MAP; Roche, 07 290 519 001)

Most kits allow a range of plasma input volumes. Therefore, we tested both the minimum and maximum input volume recommended by the supplier. The input volume in ml directly follows the abbreviated name in the plots in this report. This yields 17 unique combinations of kit and input volumes, with 3 technical replicates processed for every combination.

1.2   Metric selection

Seven performance metrics were evaluated. Kits for phase 2 were eventually selected based on transforming metrics for sensitivity and precision to robust z-scores (see Selection for phase 2).

- Sensitivity: absolute number of miRNAs detected (after setting a count cutoff that removes 95% of single positive datapoints between technical replicates).
- Precision: pairwise ALC (area-left-of-curve) calculation between technical replicates.

1.3   RMarkdown set-up

First, basic parameters are set up in this RMarkdown, such as loading dependencies, setting paths and setting up a uniform plot structure.

2 Annotation

Sample annotation with info about kit, used input volume, eluate volume etc.

RNA extraction Control (RC) spike-ins are added to plasma prior to RNA isolation, and Library Prep control (LP) spike-in controls to the RNA eluate prior to library prep. All spike-in RNA molecules were ordered at Integrated DNA Technologies as 5'-monophosphate and 3'OH RNA oligonucleotides, and diluted in 500 nM carrier DNA oligo (TCGAAGTATTC).

3 Sequencing and preprocessing

- Three high-output (HO) runs on NextSeq 500:
    - 20181005_exRNAQC011_02-101750649

    - 20181005_exRNAQC011_04-106020915

    - 20181005_exRNAQC011-99504405
- Original number of single-end reads (min= 2,239,801; median= 6,519,328; max= 10,112,548)

3.1 Filtering

First, adapters were trimmed with cutadapt v1.18, allowing an error rate of 0.15 in the adapter. After trimming, a minimum length of 15 bases was required. Finally, reads where less than 80% of bases have a quality score of 20 or higher were filtered out.

3.2 Downsampling

Randomly downsample each library to the lowest number of reads (at FASTQ level) to make sure the comparison of metrics is fair. E.g. if one sample is sequenced deeper it is likely to yield more miRNAs compared to a sample that was sequenced less deep.

As the lowest number of reads after quality filtering is 1,843,262 (in MIRVE0.1 sample), we downsampled all samples to **1.5M reads**.

3.3 Total number of reads

3.4 Count conversion

RNA spike-in reads are first isolated using a custom Python script. Non-spike reads are then mapped with Bowtie (1.2.2) and further annotated based on miRbase v22 (for miRNAs), UCSC hg38 (for tRNA) & Ensembl v91 (for other small RNAs: snoRNA, snRNA, MT_tRNA, MT_rRNA, rRNA en miscRNA).

3.5 RNA biotypes

- % of mapped reads (spike reads excluded) to different RNA biotypes.
- A very large fraction of miscellaneous RNAs (mainly Y-RNAs), a known issue in platelet-poor human plasma.
- The fraction of reads that map to the human genome but not annotated is much larger for MagnaPure kit samples than for others

Table 1: Sample annotation. PlInputV: plasma input volume; Repl: technical replicate; RNAinput: RNA input volume used for library prep

UniqueID	Biotype	RNAisolation	PlInputV	EluateV	RNAinput	Replicate	Tube	LibraryPrep	LPconc	RCconc	Abbrevation
RNA004060L1	artiPPP	miRNeasySPkit	200	14	5	RNA1	EDTA	SmallRNA	1x	1x	MIR0.2
RNA004063L1	artiPPP	miRNeasySPAkit	200	20	5	RNA1	EDTA	SmallRNA	1x	1x	MIRA0.2
RNA004066L1	artiPPP	miRNeasySPAkit	600	20	5	RNA1	EDTA	SmallRNA	1x	1x	MIRA0.6
RNA004069L1	artiPPP	QIAamp	1000	14	5	RNA1	EDTA	SmallRNA	1x	1x	CCF1
RNA004072L1	artiPPP	QIAamp	4000	14	5	RNA1	EDTA	SmallRNA	1x	1x	CCF4
RNA004075L1	artiPPP	mirVana	100	100	5	RNA1	EDTA	SmallRNA	1x	1x	MIRV0.1
RNA004078L1	artiPPP	mirVana	625	100	5	RNA1	EDTA	SmallRNA	1x	1x	MIRV0.625
RNA004081L1	artiPPP	mirVanaE	100	100	5	RNA1	EDTA	SmallRNA	1x	1x	MIRVE0.1
RNA004084L1	artiPPP	mirVanaE	625	100	5	RNA1	EDTA	SmallRNA	1x	1x	MIRVE0.625
RNA004087L1	artiPPP	NucleoSpin	300	30	5	RNA1	EDTA	SmallRNA	1x	1x	NUC0.3
RNA004090L1	artiPPP	NucleoSpin	900	30	5	RNA1	EDTA	SmallRNA	1x	1x	NUC0.9
RNA004093L1	artiPPP	Norgen	250	100	5	RNA1	EDTA	SmallRNA	1x	1x	CIRC0.25
RNA004096L1	artiPPP	Norgen	5000	100	5	RNA1	EDTA	SmallRNA	1x	1x	CIRC5
RNA004099L1	artiPPP	Maxwell	100	50	5	RNA1	EDTA	SmallRNA	1x	1x	MAX0.1
RNA004102L1	artiPPP	Maxwell	500	50	5	RNA1	EDTA	SmallRNA	1x	1x	MAX0.5
RNA004105L1	artiPPP	MagnaPure	2000	50	5	RNA1	EDTA	SmallRNA	1x	1x	MAP2
RNA004108L1	artiPPP	MagnaPure	4000	50	5	RNA1	EDTA	SmallRNA	1x	1x	MAP4
RNA004061L1	artiPPP	miRNeasySPkit	200	14	5	RNA2	EDTA	SmallRNA	1x	1x	MIR0.2
RNA004064L1	artiPPP	miRNeasySPAkit	200	20	5	RNA2	EDTA	SmallRNA	1x	1x	MIRA0.2
RNA004067L1	artiPPP	miRNeasySPAkit	600	20	5	RNA2	EDTA	SmallRNA	1x	1x	MIRA0.6
RNA004070L1	artiPPP	QIAamp	1000	14	5	RNA2	EDTA	SmallRNA	1x	1x	CCF1
RNA004073L1	artiPPP	QIAamp	4000	14	5	RNA2	EDTA	SmallRNA	1x	1x	CCF4
RNA004076L1	artiPPP	mirVana	100	100	5	RNA2	EDTA	SmallRNA	1x	1x	MIRV0.1
RNA004079L1	artiPPP	mirVana	625	100	5	RNA2	EDTA	SmallRNA	1x	1x	MIRV0.625
RNA004082L1	artiPPP	mirVanaE	100	100	5	RNA2	EDTA	SmallRNA	1x	1x	MIRVE0.1
RNA004085L1	artiPPP	mirVanaE	625	100	5	RNA2	EDTA	SmallRNA	1x	1x	MIRVE0.625
RNA004088L1	artiPPP	NucleoSpin	300	30	5	RNA2	EDTA	SmallRNA	1x	1x	NUC0.3
RNA004091L1	artiPPP	NucleoSpin	900	30	5	RNA2	EDTA	SmallRNA	1x	1x	NUC0.9
RNA004094L1	artiPPP	Norgen	250	100	5	RNA2	EDTA	SmallRNA	1x	1x	CIRC0.25
RNA004097L1	artiPPP	Norgen	5000	100	5	RNA2	EDTA	SmallRNA	1x	1x	CIRC5
RNA004100L1	artiPPP	Maxwell	100	50	5	RNA2	EDTA	SmallRNA	1x	1x	MAX0.1
RNA004103L1	artiPPP	Maxwell	500	50	5	RNA2	EDTA	SmallRNA	1x	1x	MAX0.5
RNA004106L1	artiPPP	MagnaPure	2000	50	5	RNA2	EDTA	SmallRNA	1x	1x	MAP2
RNA004109L1	artiPPP	MagnaPure	4000	50	5	RNA2	EDTA	SmallRNA	1x	1x	MAP4
RNA004062L1	artiPPP	miRNeasySPkit	200	14	5	RNA3	EDTA	SmallRNA	1x	1x	MIR0.2
RNA004065L1	artiPPP	miRNeasySPAkit	200	20	5	RNA3	EDTA	SmallRNA	1x	1x	MIRA0.2
RNA004068L1	artiPPP	miRNeasySPAkit	600	20	5	RNA3	EDTA	SmallRNA	1x	1x	MIRA0.6
RNA004071L1	artiPPP	QIAamp	1000	14	5	RNA3	EDTA	SmallRNA	1x	1x	CCF1
RNA004074L1	artiPPP	QIAamp	4000	14	5	RNA3	EDTA	SmallRNA	1x	1x	CCF4
RNA004077L1	artiPPP	mirVana	100	100	5	RNA3	EDTA	SmallRNA	1x	1x	MIRV0.1
RNA004080L1	artiPPP	mirVana	625	100	5	RNA3	EDTA	SmallRNA	1x	1x	MIRV0.625
RNA004083L1	artiPPP	mirVanaE	100	100	5	RNA3	EDTA	SmallRNA	1x	1x	MIRVE0.1
RNA004086L1	artiPPP	mirVanaE	625	100	5	RNA3	EDTA	SmallRNA	1x	1x	MIRVE0.625
RNA004089L1	artiPPP	NucleoSpin	300	30	5	RNA3	EDTA	SmallRNA	1x	1x	NUC0.3
RNA004092L1	artiPPP	NucleoSpin	900	30	5	RNA3	EDTA	SmallRNA	1x	1x	NUC0.9
RNA004095L1	artiPPP	Norgen	250	100	5	RNA3	EDTA	SmallRNA	1x	1x	CIRC0.25
RNA004098L1	artiPPP	Norgen	5000	100	5	RNA3	EDTA	SmallRNA	1x	1x	CIRC5
RNA004101L1	artiPPP	Maxwell	100	50	5	RNA3	EDTA	SmallRNA	1x	1x	MAX0.1
RNA004104L1	artiPPP	Maxwell	500	50	5	RNA3	EDTA	SmallRNA	1x	1x	MAX0.5
RNA004107L1	artiPPP	MagnaPure	2000	50	5	RNA3	EDTA	SmallRNA	1x	1x	MAP2
RNA004110L1	artiPPP	MagnaPure	4000	50	5	RNA3	EDTA	SmallRNA	1x	1x	MAP4

Table 2: Different RNA extraction Control (RC) spike-ins are used. For each RC spike-in, the ID and sequence is shown

SpikeID	sequence
RC1-01	ACUCAUCUACGUACGCAUCUAGUCU
RC1-03	UGCUAUCAUAUCACAGUACGCGAGC
RC1-04	UAGAUGAGAUACUGAUAGCGACGUA
RC1-06	AUCGUCUCGUCAUCUCAUAUCUACA
RC1-07	UAUGCAUAUGAUCACGAGACUCAGU
RC1-09	GCUCUACACUCUACUCGUCAGCUGU
RC1-10	CGAUGCUAUAGACUCUCACGUGAUG
RC1-11	CGCAUCAGUCGUCAUCUAGAUACAG
RC1-12	CUGAUGAUAGAUACGCGCACACAGU
RC2-25	AUGCUGAUGAUAGACGCUACUGACU
RC2-28	CGUAUCGCUGCUCUGAGUCACUAUCUAC
RC2-34	AGAUAGUACUGAUCUGCUGCGACGAGUGACUGUC

Table 3: Different Library Prep Control (LP) spike-ins are used. For each LP spike in, the ID and sequence is shown

SpikeID	sequence
LP1-01	GUCCCACUCCGUAGAUCUGUUC
LP1-02	GAUGUAACGAGUUGGAAUGCAA
LP1-03	UAGCAUAUCGAGCCUGAGAACA
LP1-04	CAUCGGUCGAACUUAUGUGAAA
LP1-06	UCUUAACCCGGACCAGAAACUA
LP1-07	AGGUUCCGGAUAAGUAAGAGCC
LP1-10	UGAUACGGAUGUUAUACGCAGC
LP1-11	CCUGGAACUUAGGACGUGAAUC
LP1-12	UCAUGAGUCCGUACCUUGAUUG
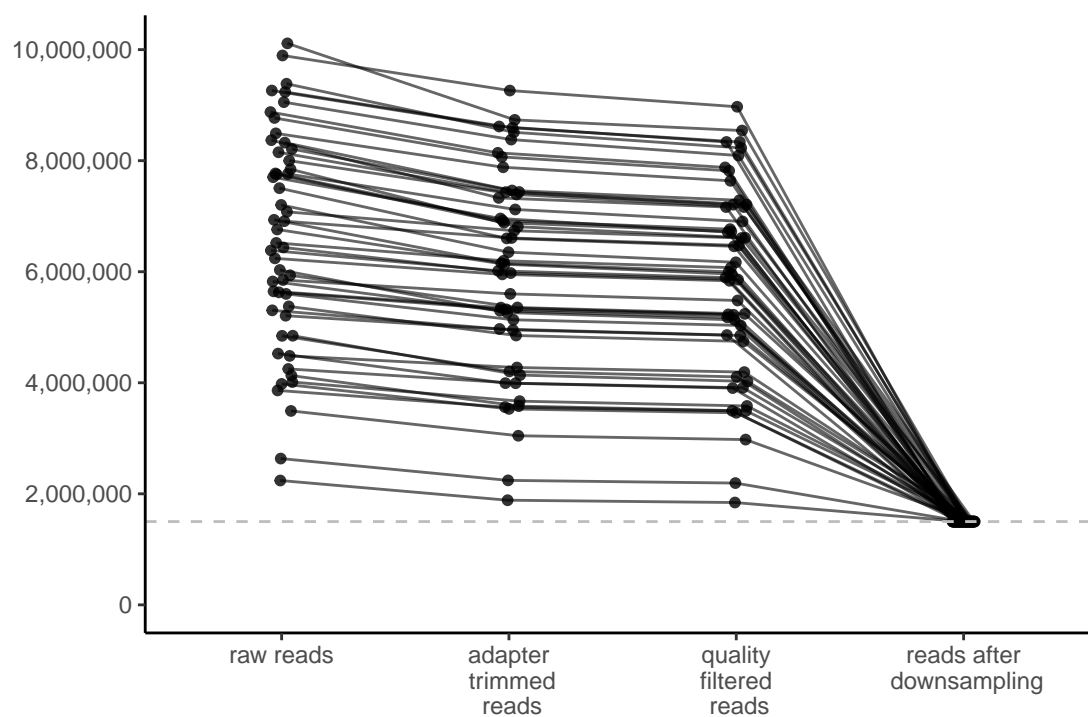
Figure 1: Number of reads at different stages in the data processing workflow. Raw reads: total number of reads in raw FASTQ files at start; adapter trimmed reads: number of reads after adapter trimming and requiring a minimum length of 15; quality filtered reads: number of adapter trimmed reads where at least 80% of bases have a phred score of 20 or higher; reads after downsampling: all samples were downsampled to 1.5M.
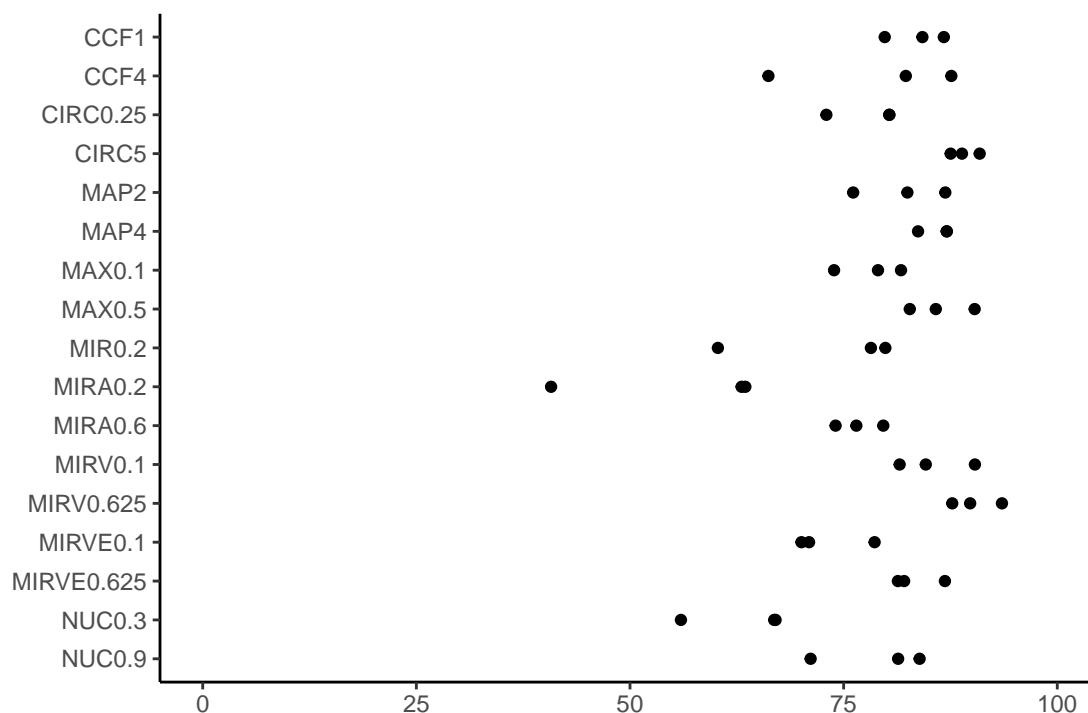
Figure 2: Differences in mapping percentages. How many % of the 1.5M reads collectively map to spikes, miRbase v22 (for miRNAs), UCSC hg38 (for tRNA) & Ensembl v91 (for other small RNAs: snoRNA, snRNA, MT_tRNA, MT_rRNA, rRNA en miscelaneous RNA). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAP2: MagNA Pure 24 total NA isolation kit, 2 ml input; MAP4: MagNA Pure 24 total NA isolation kit, 4 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

- Together with the strandedness issue at mRNA level for this kit (potential DNA contamination), we decided to leave this kit out of further analyses.

- Another observation is that the RNA2 replicate samples (from the second library size selection) show a consistently different pattern in terms of number of miRNAs than RNA1 and RNA3. There may have been an issue here with size selection. Therefore, for comparison of detected miRNAs, we apply a random downsampling (without replacement) to the minimum at the level of the miRNA counts (see [miRNA counts]).

Figure 3: Fraction of mapped reads for each RNA biotype. Spike-in RNAs are not included. (miRNA: microRNA; misc_RNA: miscellaneous RNA - mainly Y-RNAs; multiple_piRNA_tRNA: multimapping to piRNA and tRNA; multiple_snRNA_misc_RNA: multimapping to small nucle(ol)ar RNA and miscellaneous RNA; not_annotated: mapping to reference genome but not annotated as small RNA; piRNA: piwi-interacting RNA; rRNA: ribosomal RNA; snoRNA: small nucleolar RNA; snRNA: small nuclear RNA; tRNA: transfer RNA)

3.6 Spikes

RC (RNA extraction Control) spikes were added before RNA purification (volume adapted according to plasma input volume: (1 microL RC/ 100 microL plasma)). If there are more RC spikes detected, there is less endogenous RNA in plasma (of note, we used a large plasma pool from only one donor)

LP (Library Preparation) spikes were added to the RNA eluate after RNA purification (2 microL LP/ 12 microL eluate), prior to library preparation. The more LP spikes detected, the less endogenous RNA present in the eluate. Given that we always used the same eluate volume for library prep, more LP spikes indicate a lower endogenous RNA concentration in the eluate.

3.6.1 Spike percentage in mapped reads

Note that MIRVE0.1 has a large fraction of LP spikes, pointing to a low RNA concentration in eluate, see miRNA concentration.

3.6.2 Spike-count linearity

Spike counts are proportional to their relative concentration in the mix
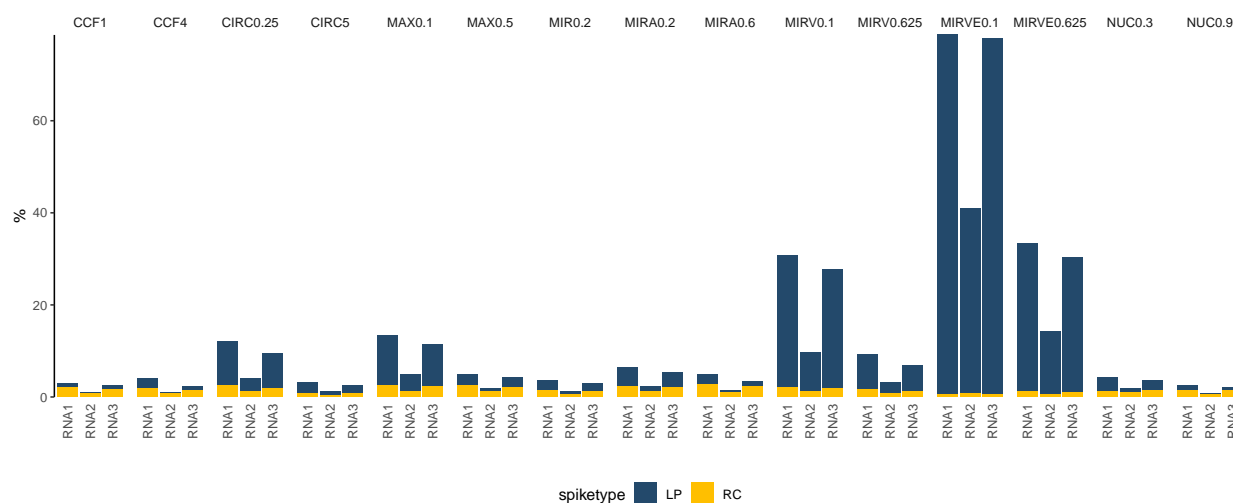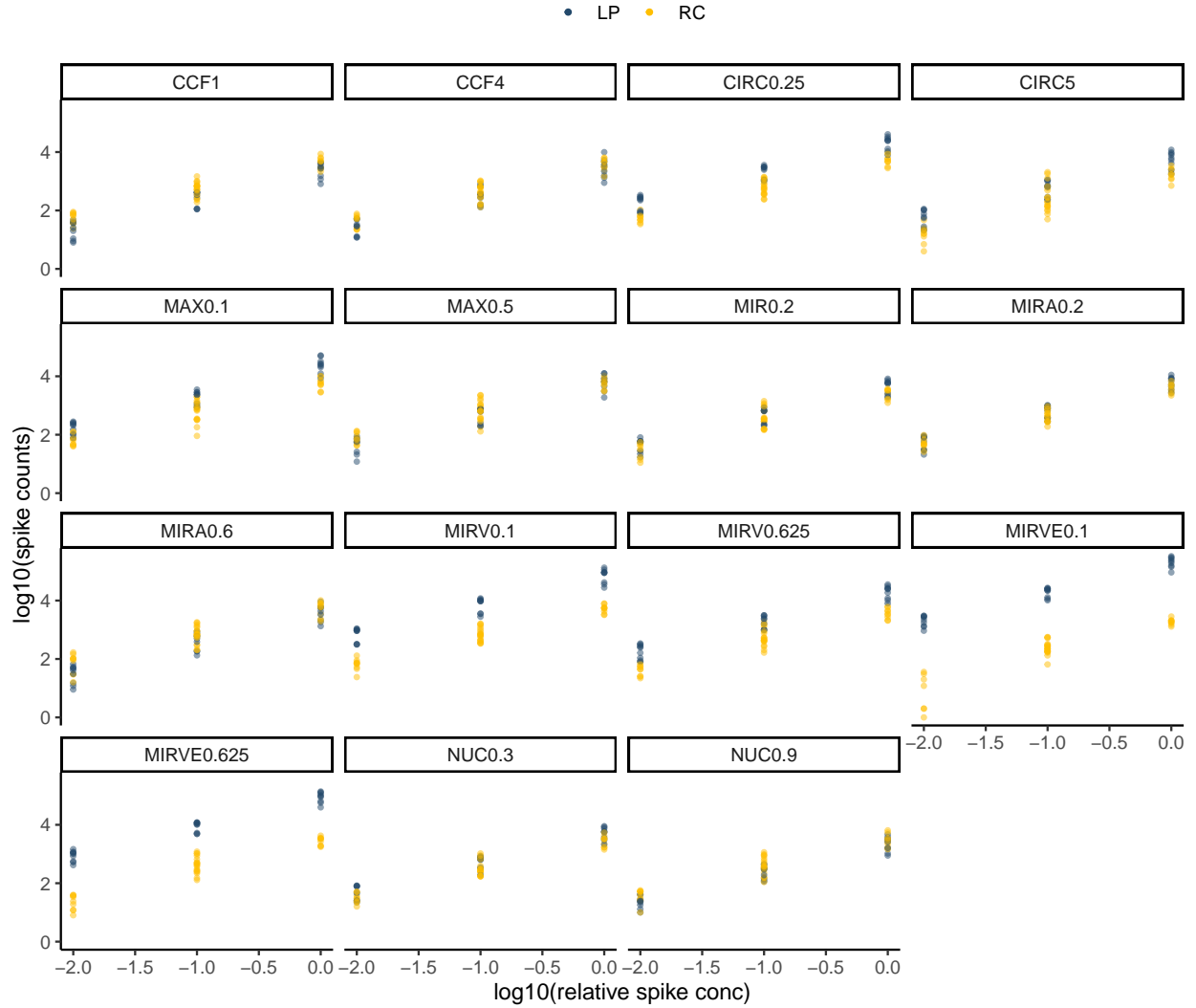
Figure 4: Percentage of mapped reads that goes to spike-in RNAs. LP (Library Preparation) spikes were added prior to library preparation, after RNA purification (blue). RC (RNA extraction Control) spikes were added prior to RNA purification (yellow). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

Figure 5: Spike counts are proportional to their relative concentration in the mix. LP (Library Preparation) spikes were added prior to library preparation, after RNA purification (blue). RC (RNA extraction control) spikes were added prior to RNA purification (blue). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

4 Performance metrics

Seven performance metrics are calculated.

In order to compare different kits in a uniform way across the different performance metrics, we transformed these into **z-scores**. We made sure that a **higher value always corresponds to better performance**. To account for the small sample size, we calculated **robust z-scores**.

4.1 miRNA concentration

The ratio of endogenous miRNA to LP count fractions reflects the relative concentration of endogenous miRNA in the eluate: if there is more endogenous miRNA in the input for the library prep, there will be less LPs and therefore the ratio endogenous miRNA/LP is higher. Remember that some kits have a much larger eluate volume after RNA isolation. A larger total eluate volume results in more diluted endogenous RNA (lower concentration) and therefore less endogenous RNA in library prep (given constant input volume for all library preparations).

Scoring measure: the higher the concentration, the better

4.2 miRNA yield

For small RNA sequencing purposes, we are most interested in the concentration of the eluate as we can only use a limited volume for library prep. However, by multiplying the relative miRNA concentrations above with the total eluate volume, we obtain the relative miRNA yield in the eluate after RNA isolation. In case the total eluate volume is larger, the RNA will be more diluted (this is for example the case for MIRV: 100 microL eluate compared to only 12 microL for CCF).

If the miRNA yield is high, but the eluate volume is large, further concentrating the total eluate before library prep may give better results . Of note, we did not evaluate this in our study. **Scoring: the higher the yield, the better**

4.3 Efficiency of kits

Based on the previous plot with miRNA yield, we observe differences in RNA isolation efficiencies among kits. By **correcting the yield for the plasma input volume**, we obtain a better picture:

- **with more input, you expect to have more yield in eluate. To correct for this: divide yield by input volume**
- (endogenous/LP)*EluateV / PlasmaInputV
- compare CCF4 to MAX0.5: CCF4 has 8x more plasma (and therefore also more RNA) in input than MAX0.5, but this does not translate in 8x more yield in the eluate -> MAX0.5 seems to extract the lower volume more efficiently than CCF4
- Although the yield is higher within a given kit when using the maximum input volume, this sometimes seems to be associated with a lower efficiency than the minimal input volume

Remark: while we could also look at RC/LP ratio corrected for input and eluate volume (should give similar results), we decided to look at endogenous RNA as a more representative metric as this is the biomaterial of interest.

There is a clear difference in kit efficiency, with a difference of factor 10 or more.

If some adjustments would be made to kits with low input volume, but high efficiency (i.e. increase allowed plasma input V and keep eluate V as small as possible), **the overall performance may further improve**. Of note, we did not evaluate this in our study.
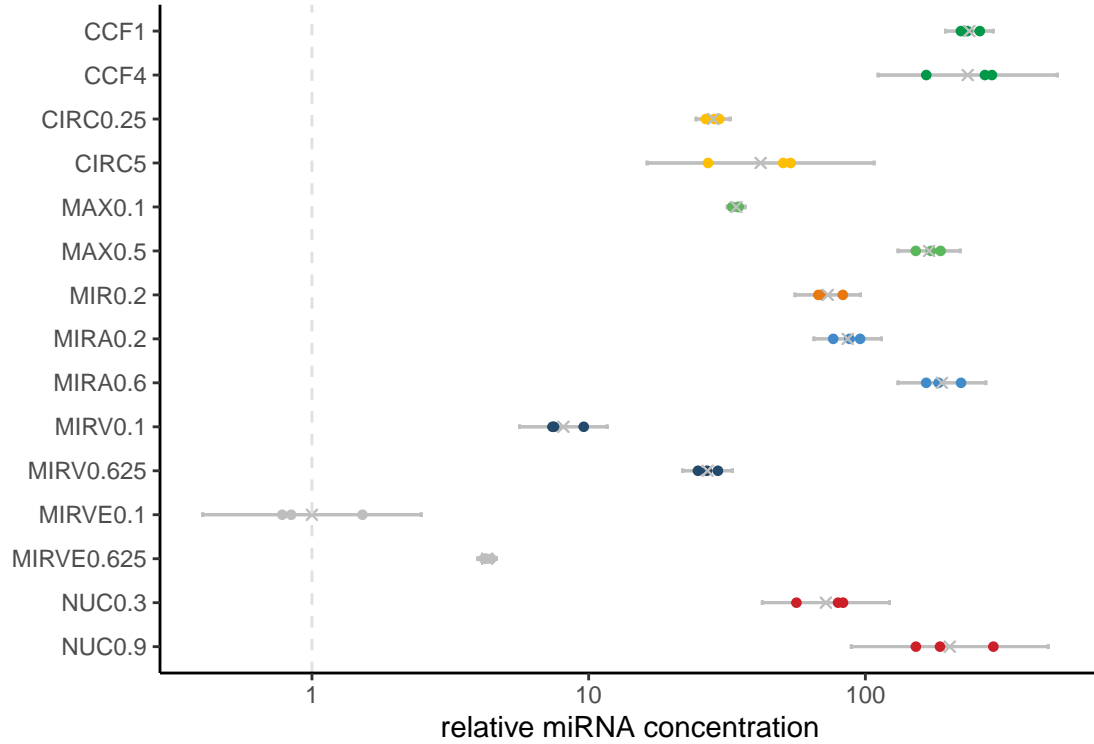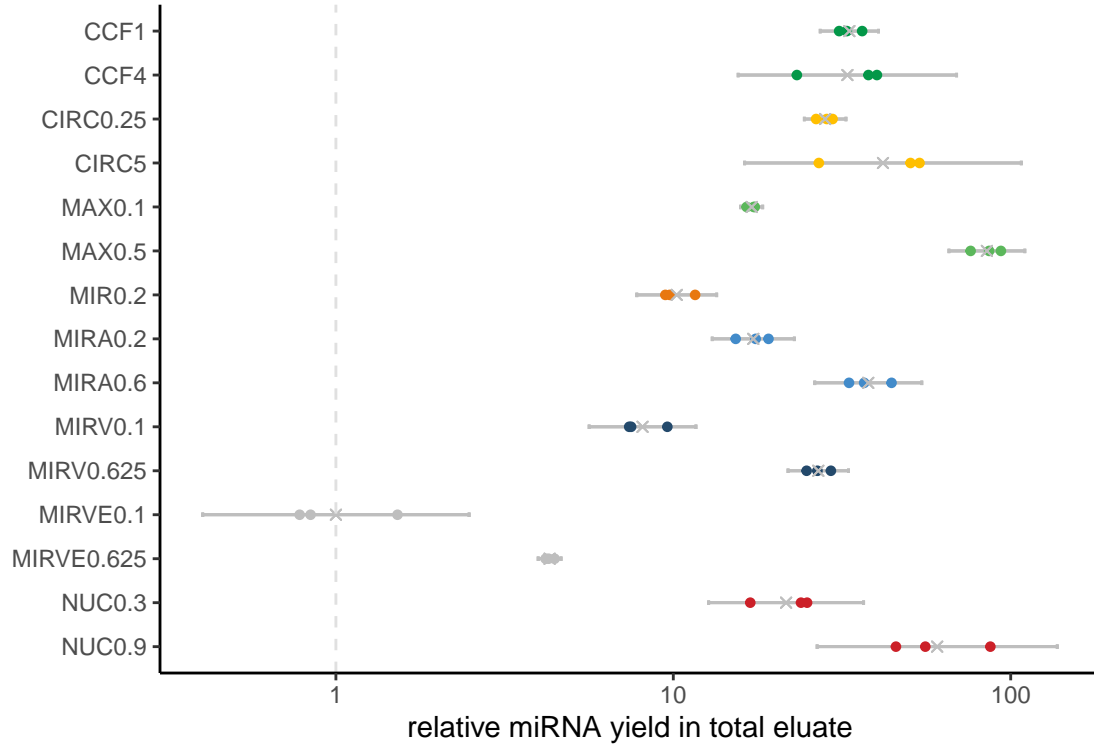
Scoring: the higher the efficiency, the better

Figure 6: Relative miRNA concentration in eluate after RNA purification. Concentration: ratio of endogenous miRNA to LP spikes. Values were first log transformed and rescaled to the average of MIRVE0.1, then transformed back to linear scale. Mean per kit (cross) and 95% confidence intervals shown (grey lines). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)
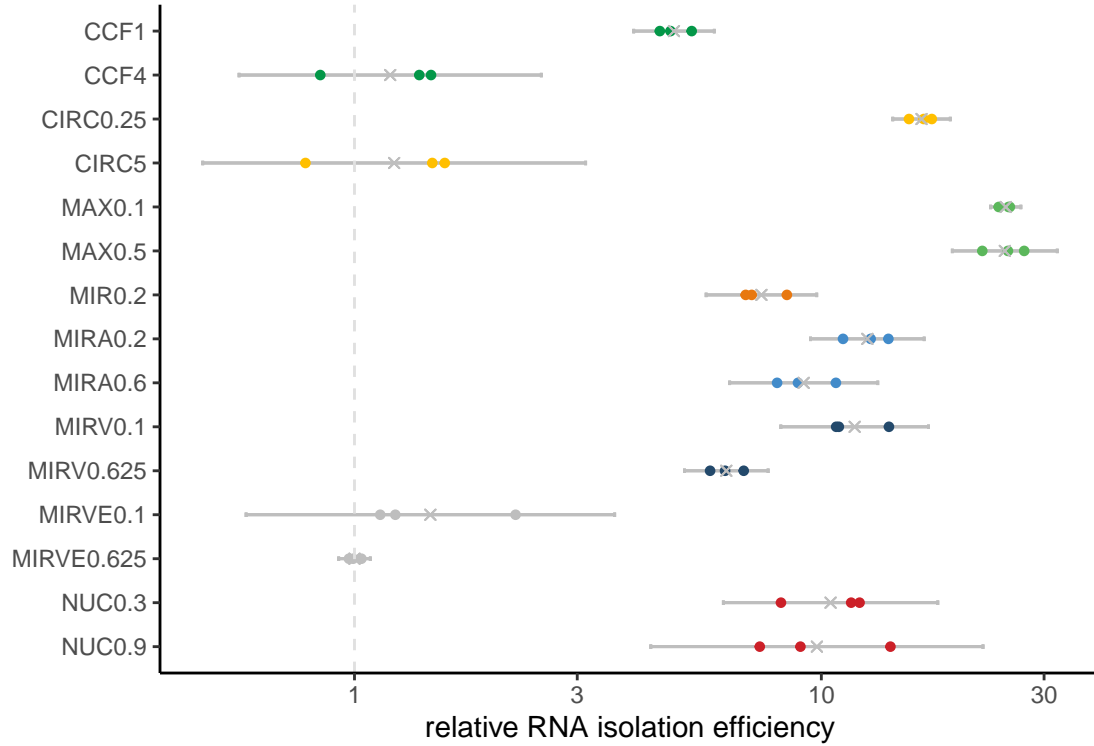
Figure 7: Relative miRNA yield of kits. Yield: eluate volume corrected miRNA concentration. Values were first log transformed and rescaled to the average of MIRVE0.1, then transformed back to linear scale. Mean per kit (cross) and 95% confidence intervals shown (grey lines). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

13

Figure 8: Relative efficiency of kits. Efficiency: plasma input volume corrected miRNA yield. Values were first log transformed and rescaled to the average of MIRVE0.625, then transformed back to linear scale. Mean per kit (cross) and 95% confidence intervals shown. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

4.4 Filter threshold

- We want to have a filter threshold that eliminates 95% of single positive genes (SP) between technical replicates, i.e. genes detected in only one technical replicate (cf. miRQC study of Mestdagh et al., 2014, Nature Methods):
  - **Only miRNA counts are considered here**
  - All pseudocounts < 1 are first rounded down to 0
  - Pairwise comparison of technical replicates (3 pairs per kit-volume combination)
  - Determine threshold at which at least 95% of the single positives are removed
  - Take median cutoff per combination of kit and volume
- **95% SP elimination cutoff (which is specific for each kit-volume combination) will be used for filtering throughout ALL analyses**
  - This is our proposed strategy to make data comparable, we do not claim that this is the only way to do this

4.4.1 cutoff examples

Two examples of pairwise kit-volume comparisons together with their cutoff and r-squared value (based on linear model (y=x) of log counts). Histograms show the relative number of RNAs with counts in that bin. For an overview of the cutoffs for each kit-volume combination, see next tab 95% SP elimination cutoffs.

4.4.2 95% SP elimination cutoffs

- If all counts smaller than or equal to cutoff are eliminated, at least 95% of single positives are removed, resulting in data that is highly reproducible

- **Cutoff is always higher for lower input volume within same kit** (with lower input volume, there is more variation in which genes are detected in each replicate)

- **Cutoffs are close to each other** except in the mirVana kits (MIRV & especially MIRVE). Note that 1 count difference can already have a major impact on the number of miRNAs filtered out

- We use the median cutoff per kit-volume combination for filtering in further analyses (see table below)

- **Scoring: take the negative of the cutoff values (so that higher = better precision)**

4.4.3 Impact of filtering

- After applying the repeatability cutoff, remaining miRNA counts per sample: min=27,962; mean=112,015; max=293,373
- **Scoring: data retention: more % of counts remaining = better precision**

  - Another scoring measure could be: the higher the number of total reads remaining, the better
  - Is related to the cutoff & initial number of reads

- As the cutoffs are close together, the % filtered out is similar among kits (except for MIRVE0.1)
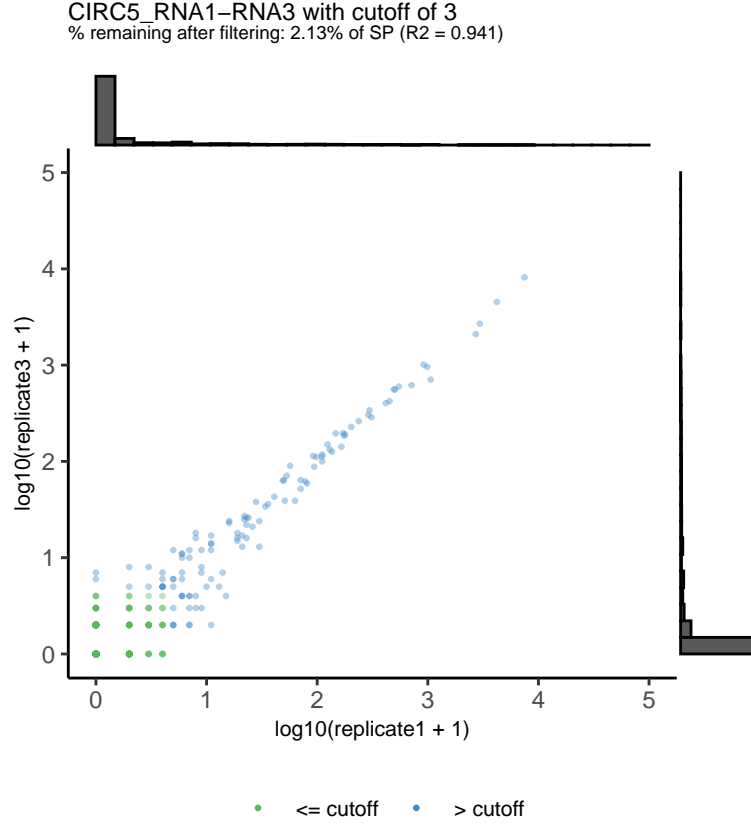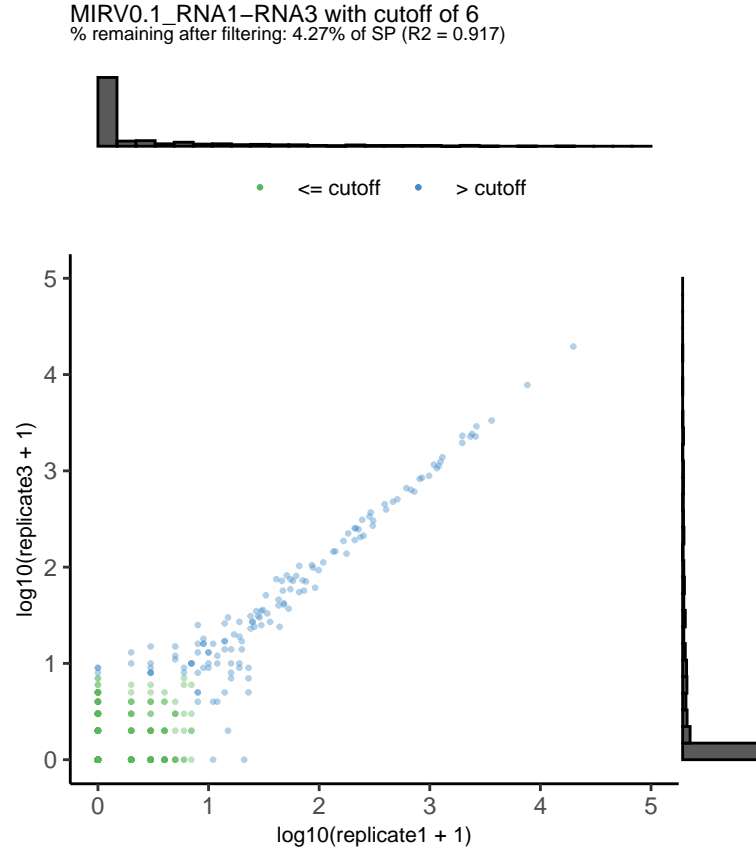
Figure 9: Pairwise miRNA count comparison of first and third replicate of the CIRC5 kit. The coefficient of determination is 0.941 (linear model that fits log10 values). Single positives are 0 in one replicate and > 0 in other. The cutoff of 3 removes 97.87% of single positives. Green dots show data points that are filtered out with this cutoff. (CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input)

Table 4: Median cutoff per kit

Kit	cutoff
CCF1	3.00
CCF4	3.00
CIRC0.25	3.00
CIRC5	3.00
MAX0.1	5.00
MAX0.5	3.00
MIR0.2	3.00
MIRA0.2	3.00
MIRA0.6	2.00
MIRV0.1	6.00
MIRV0.625	4.00
MIRVE0.1	15.36
MIRVE0.625	7.00
NUC0.3	3.00
NUC0.9	3.00

Figure 10: Pairwise miRNA count comparison of first and third replicate with MIRV0.1 kit. Coefficient of determination is 0.917 (based on linear model that fits log10 values). Single positives are 0 in one replicate and > 0 in other. The cutoff of 6 removes 95.61% of single positives. Green dots show data points that are filtered out with this cutoff. (MIRV0.1: mirVana PARIS kit, 0.1 ml input)
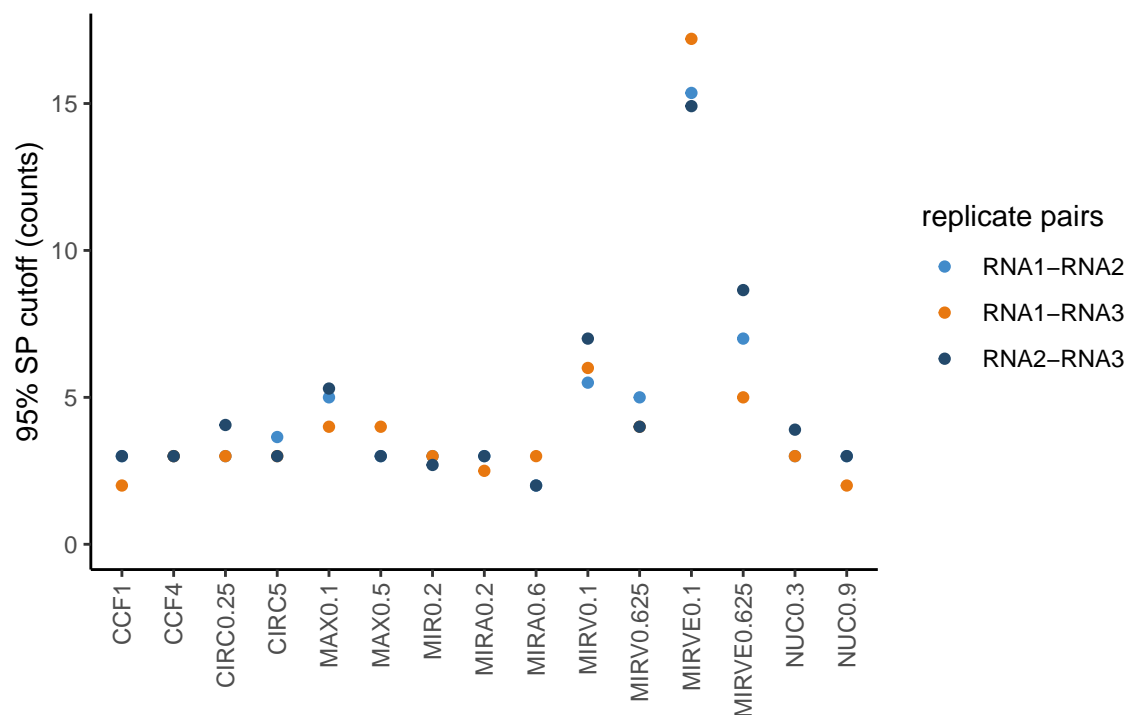
Figure 11: Count threshold that removes 95% of single positives for each pairwise comparison of replicates. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

Figure 12: Impact of filtering (filter removes 95% of single positives). Left: % of total counts that are kept after applying filter; right: sum of counts that are not filtered out. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)
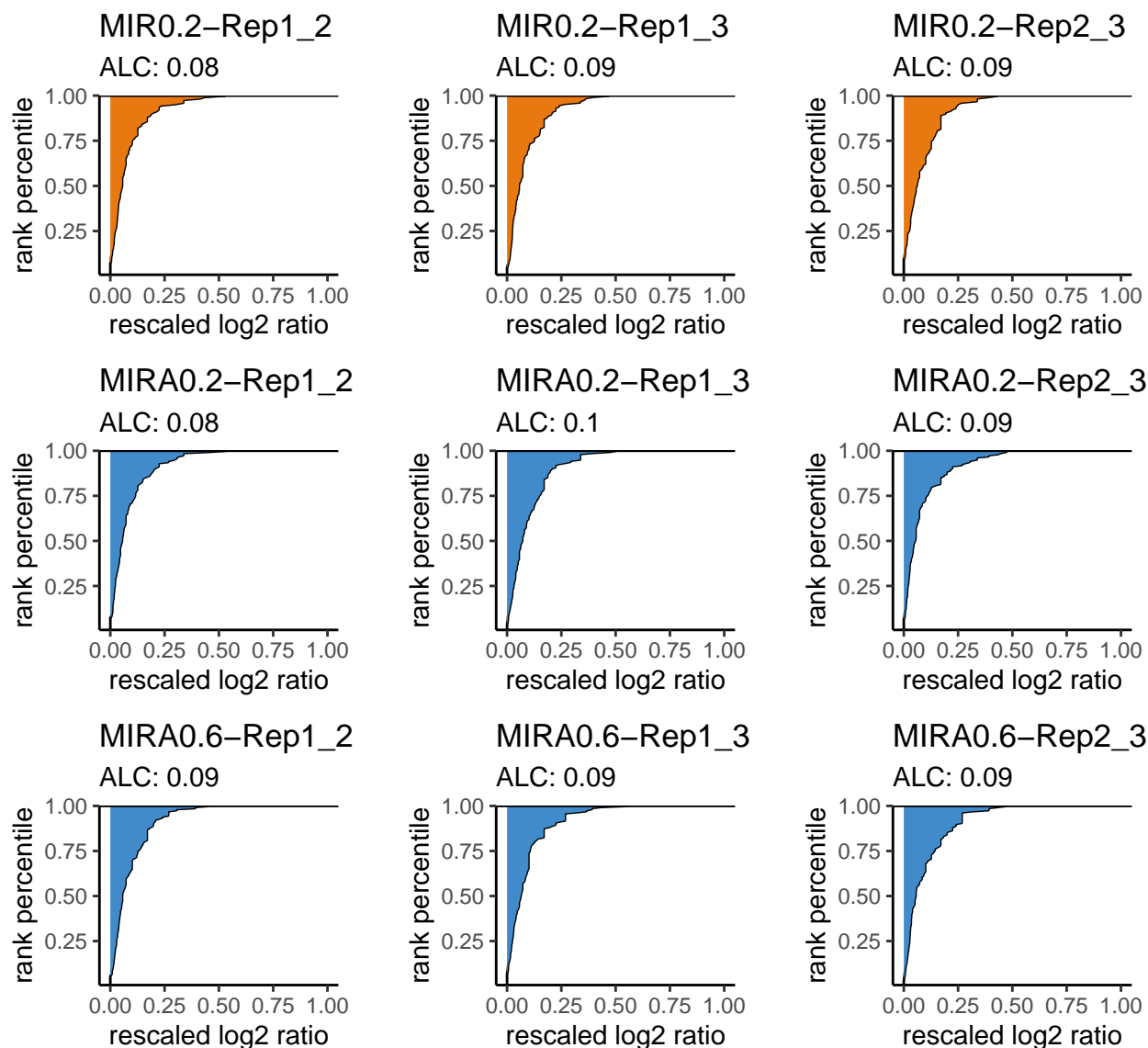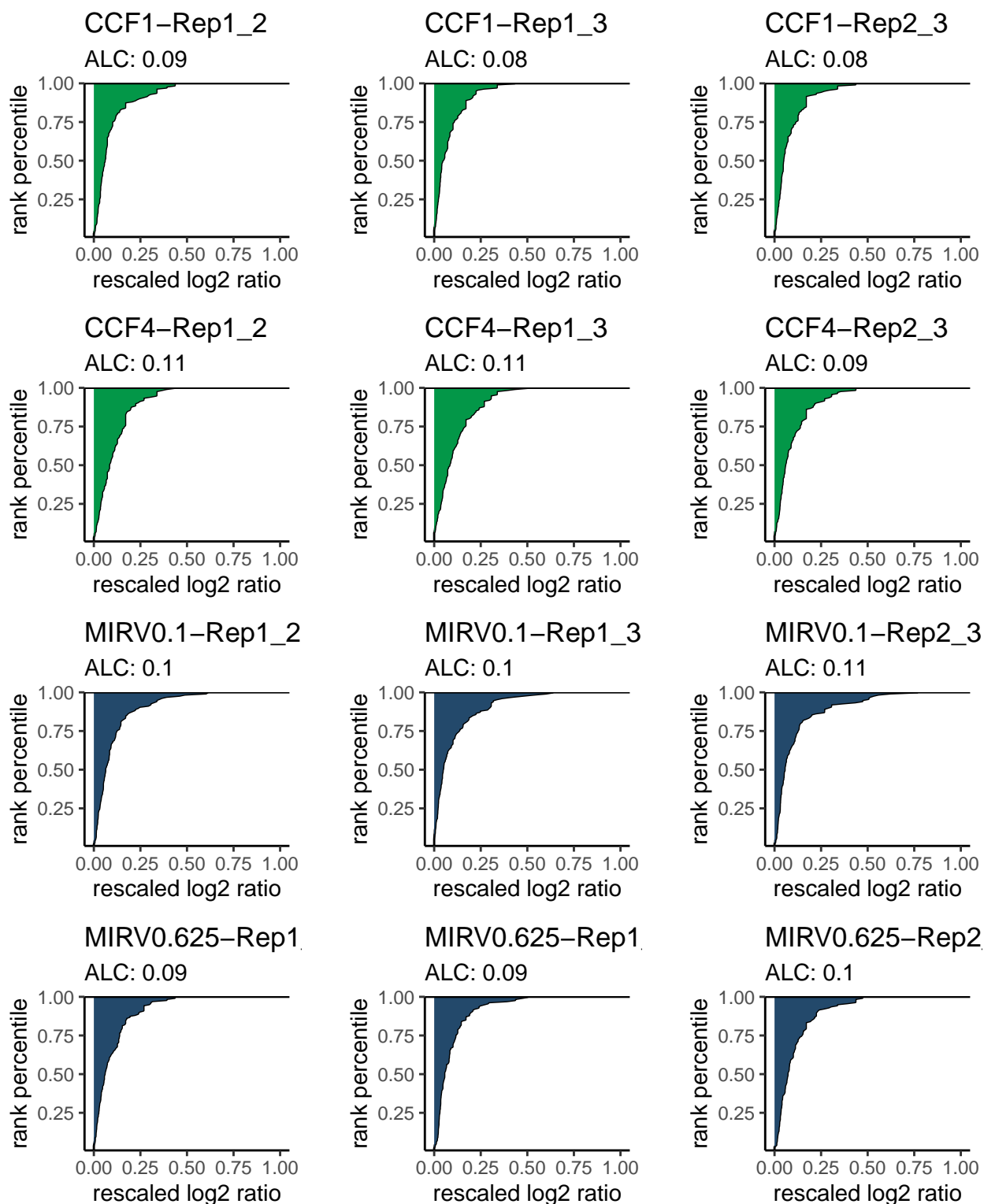
4.5   Number of miRNAs

- Filter: only keep **miRNAs** that reach the median 95% SP cutoff per kit in terms of counts
- Observations before and after filtering:
  - Less variabillily among kits than at mRNA level (see exRNAQC004)
  - A higher input volume does not necessarily result in a higher number of detected genes

- **Scoring: more miRNAs that reach precision threshold = better detection sensitivity**
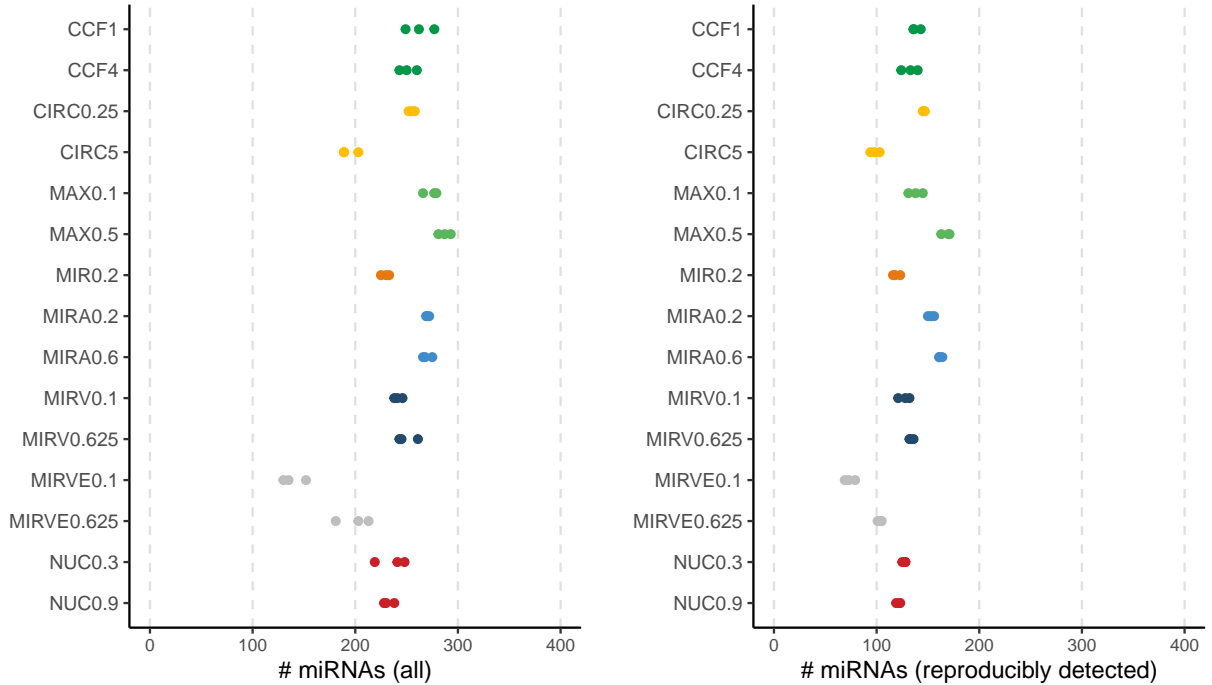
Figure 13: Number of miRNAs that are detected per 1.5 M reads. Left: all miRNAs that are detected with at least one count; right: miRNAs that are reproducibly detected (precision threshold that eliminates 95% of single positives). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

4.6   Area-left-of-curve

ALC (area-left-of-curve) **as repeatability metric** (see Mestdagh et al., 2014, Nature Methods). Compare two technical replicates at a time, only considering miRNAs that reach the precision threshold (which eliminates 95% of single positives) in at least one of the samples and 1 count in the other sample. Replace all zero counts by NA. Determine the absolute value of the log2 ratios of counts for each miRNA. Plotted
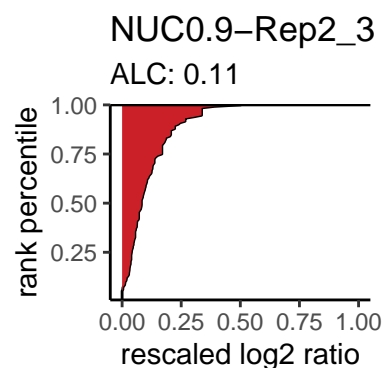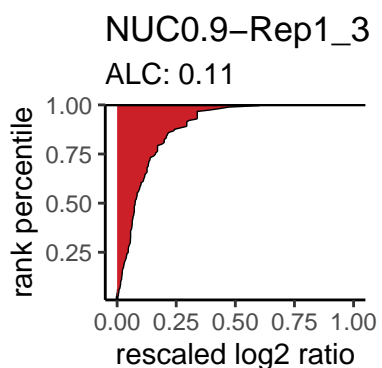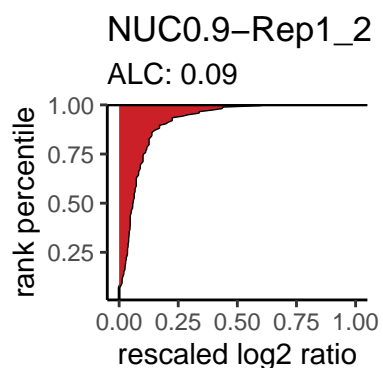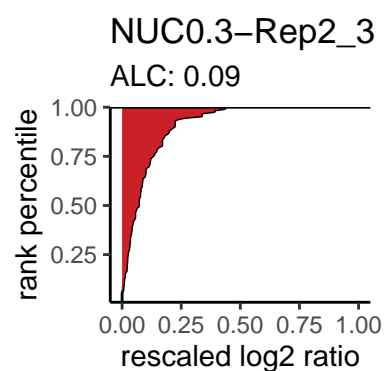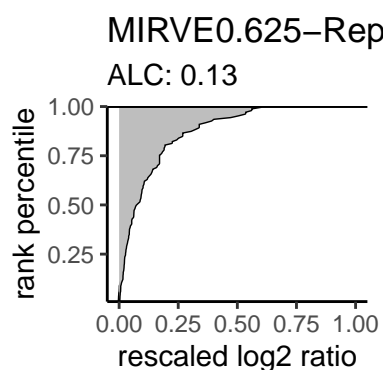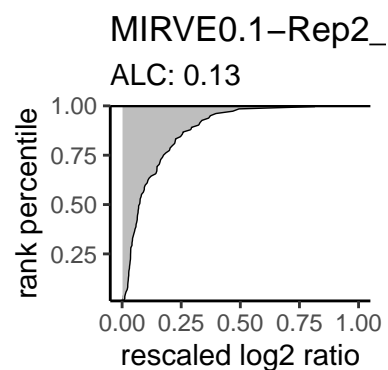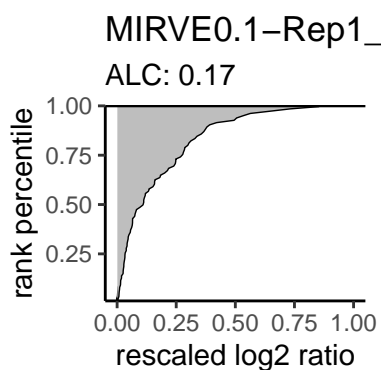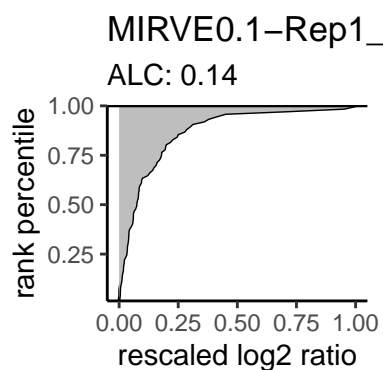
are the (percentage) ranks of these absolute log2 ratios. The faster the curve reaches 100% (the smaller the ALC), the better. A small ALC indicates that the replicates give similar counts for each miRNA.
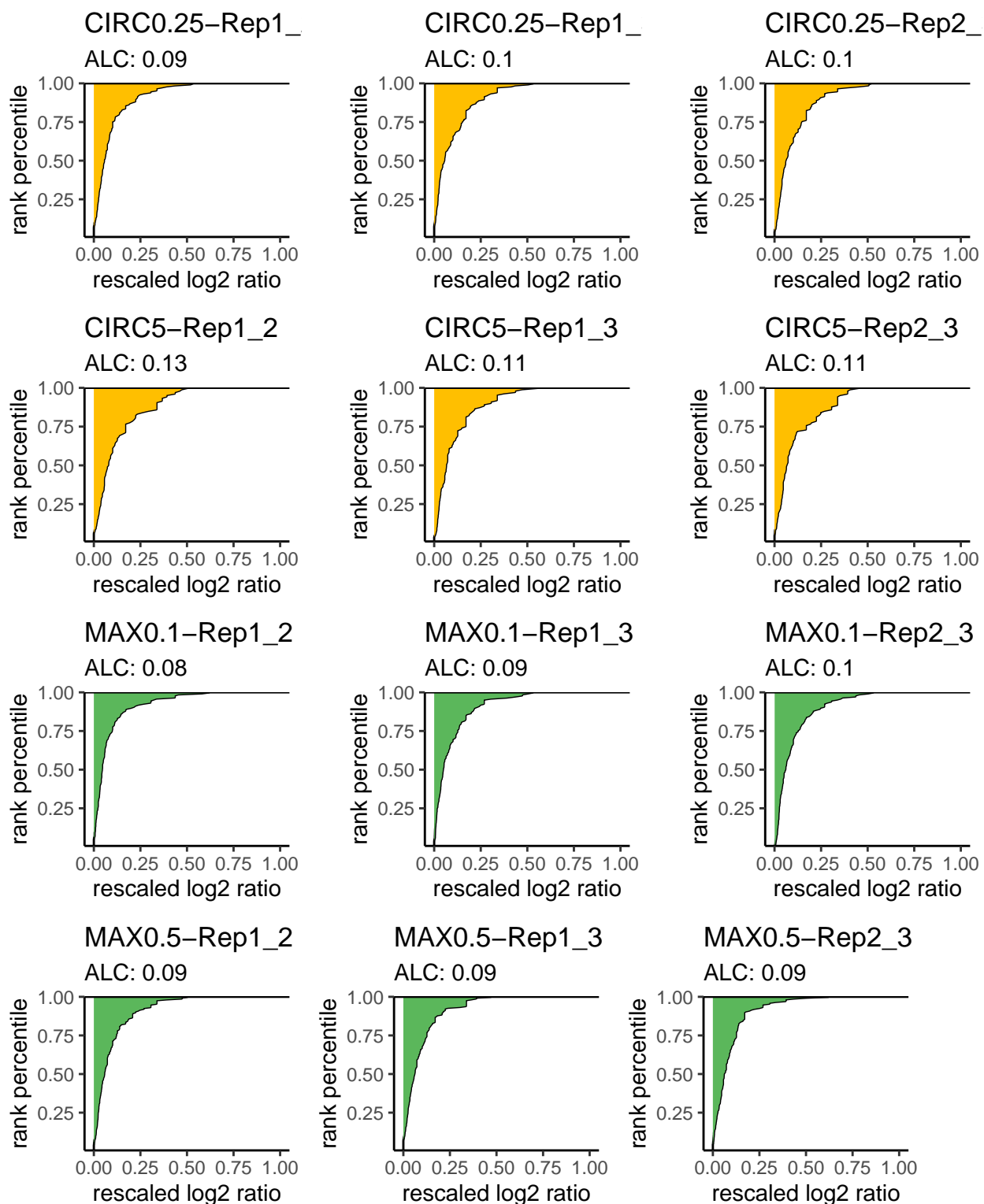
Scoring: lower ALC = better precision

4.6.1 Individual ALC plots

CCF1−Rep1_2
ALC: 0.09

CCF1−Rep1_3
ALC: 0.08

CCF1−Rep2_3
ALC: 0.08

CCF4−Rep1_2
ALC: 0.11

CCF4−Rep1_3
ALC: 0.11

CCF4−Rep2_3
ALC: 0.09

MIRV0.1−Rep1_2
ALC: 0.1

MIRV0.1−Rep1_3
ALC: 0.1

MIRV0.1−Rep2_3
ALC: 0.11

MIRV0.625−Rep1
ALC: 0.09

MIRV0.625−Rep1
ALC: 0.09

MIRV0.625−Rep2
ALC: 0.1

CIRC0.25−Rep1_
ALC: 0.09

CIRC0.25−Rep1_
ALC: 0.1

CIRC0.25−Rep2_
ALC: 0.1

CIRC5−Rep1_2
ALC: 0.13

CIRC5−Rep1_3
ALC: 0.11

CIRC5−Rep2_3
ALC: 0.11

MAX0.1−Rep1_2
ALC: 0.08

MAX0.1−Rep1_3
ALC: 0.09

MAX0.1−Rep2_3
ALC: 0.1

MAX0.5−Rep1_2
ALC: 0.09

MAX0.5−Rep1_3
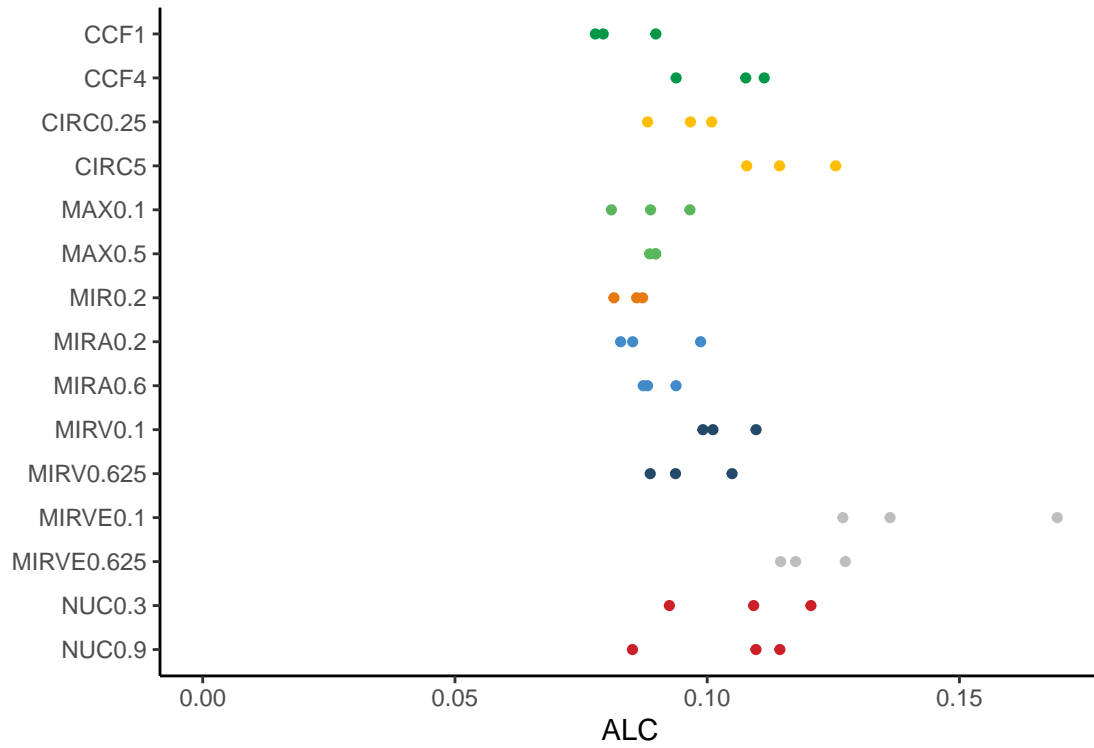ALC: 0.09

MAX0.5−Rep2_3
ALC: 0.09

Figure 14: Precision based on ALCs (area-left-of-curve). The lower the ALC, the better (less difference between replicates). (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)

4.6.2 Overview ALC

4.7 Overview

Based on robust z-scores (for each performance metric: higher robust z-score is better)

4.7.1 Correlation among metrics

- Differences between kits are less pronounced for miRNAs than for mRNAs (see exRNAQC004). This also means that if the ranking of performance of kits slightly deviates in different metrics (as a result of small differences) - i.e. one kit scores better on first method while other kit scores slightly better for second metric - the correlation (r) between metrics will be low.
- Some remarks:
  - Yield and efficiency are theoretical metrics: how well would the kit perform regardless of input and/or eluate volume restrictions (see miRNA yield and Efficiency of kits)
  - The two precision metrics (ALC and filter threshold) highlight a different aspect of precision: ALC shows how similar miRNA counts between replicates are (see Area-left-of-curve), while the threshold gives an idea of the amount of noise - from which count threshold onward can you "trust" count values (see Filter threshold)

4.7.2 Comparison of kits

5 Selection for phase 2

The RNA isolation for phase 2 is based on robust z-score transformed metric for sensitivity (# detected genes, see Number of miRNAs) and for precision (ALC, see Area-left-of-curve). Higher z-score = better

We looked at both metrics but in close calls, the sensitivity was given a higher weight. We also wanted to include at least one kit which is able to purify miRNA in case you have less than 1 ml of plasma as it is not always possible to collect or use multiple ml.

Selection: Maxwell (**MAX0.5**) & miRNeasy serum/plasma advanced (**MIRA0.6**)
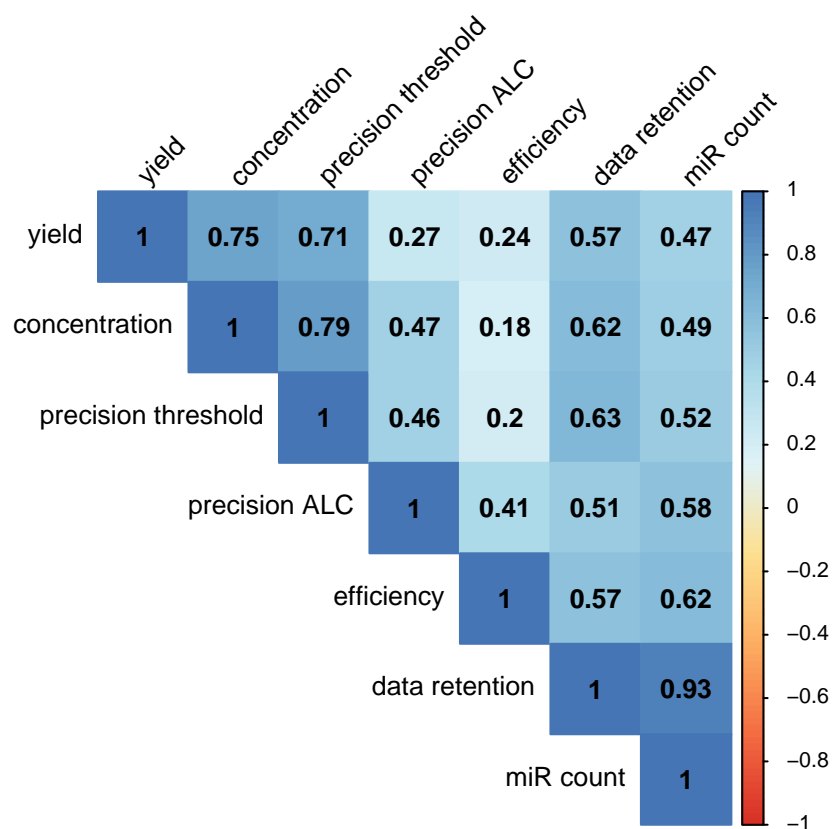
Figure 15: Correlation among metrics based on robust z-scores. Spearman correlation with complete hierarchical clustering. (concentration: relative endogenous miRNA concentration based LP spikes; data retention: % of counts left after applying precision threshold; efficiency: yield corrected for plasma input volume ; miR count: number of miRNAs after applying precision threshold; precision ALC: precision based on area-left-of-curve; precision threshold: count threshold to filter out 95% of single positives between replicates; yield: concentration corrected for eluate volume)
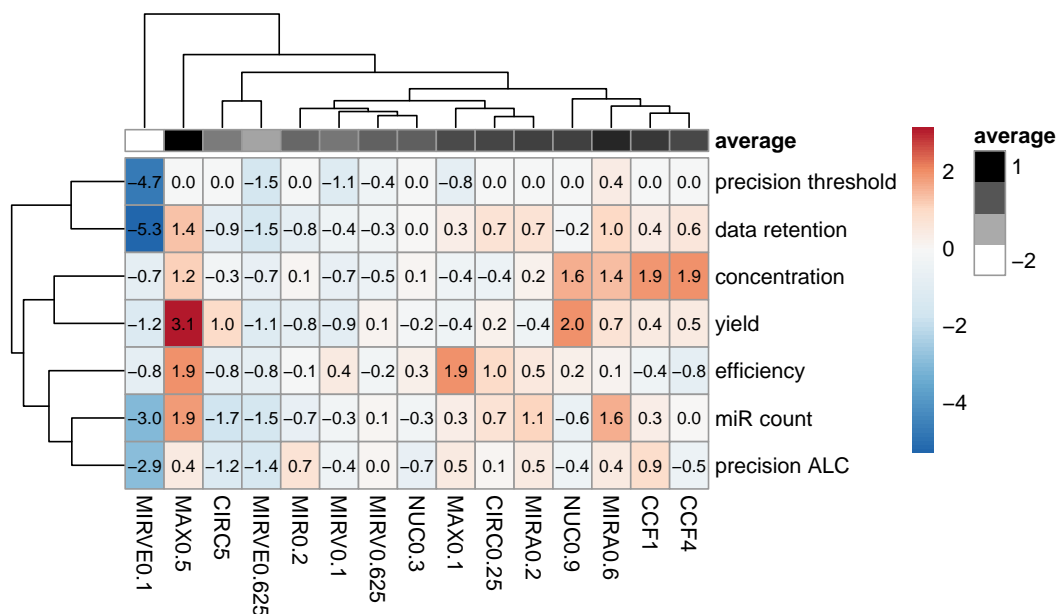
Figure 16: Comparison of kit performance based on robust z-scores. Higher means a better performance. Complete hierarchical clustering based on Euclidean distance. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input; concentration: relative endogenous miRNA concentration based LP spikes; data retention: % of counts left after applying precision threshold; efficiency: yield corrected for plasma input volume ; miR count: number of miRNAs after applying precision threshold; precision ALC: precision based on area-left-of-curve; precision threshold: count threshold to filter out 95% of single positives between replicates; yield: concentration corrected for eluate volume; average: average z score over all metrics)
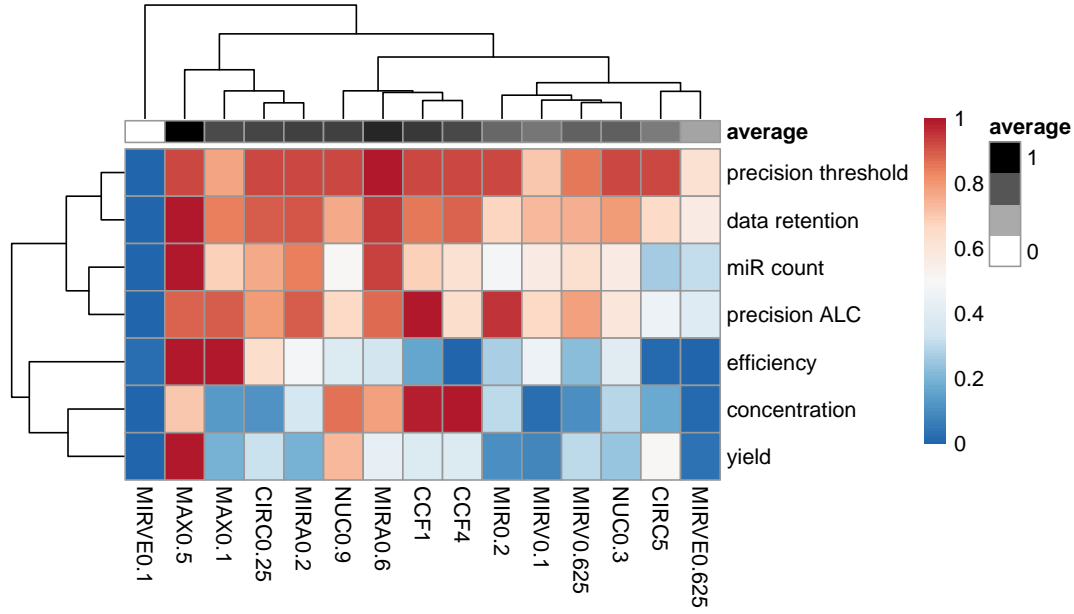
Figure 17: Comparison of kit performance based on rescaled robust z-scores. Robust z-scores were rescaled to range of [0,1] to emphasize differences within a metric. Closer to 1 means a better performance. Complete hierarchical clustering based on Euclidean distance. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input; concentration: relative endogenous miRNA concentration based LP spikes; data retention: % of counts left after applying precision threshold; efficiency: yield corrected for plasma input volume ; miR count: number of miRNAs after applying precision threshold; precision ALC: precision based on area-left-of-curve; precision threshold: count threshold to filter out 95% of single positives between replicates; yield: concentration corrected for eluate volume; average: average z score over all metrics)
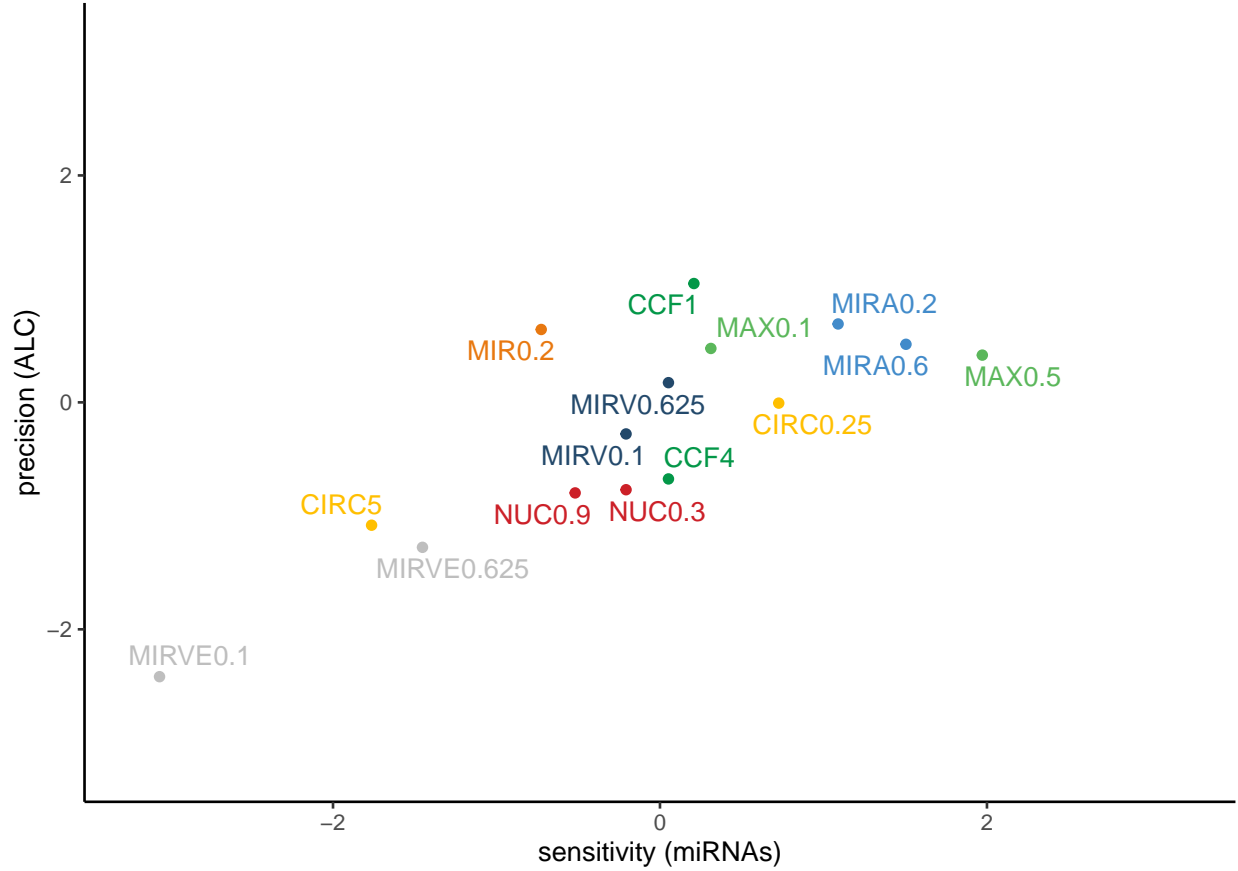
Figure 18: Robust z-scores (median per RNA isolation kit) for sensitivity (number of miRNAs) (x) and precision (ALC, area-left-of-curve) (y). MIRA0.6 and MAX0.5 kits are selected for phase II. (CCF1: QIAamp ccfDNA/RNA kit, 1 ml input; CCF4: QIAamp ccfDNA/RNA kit, 4 ml input; CIRC0.25: plasma/serum circulating and exosomal RNA purification kit, 0.25 ml input; CIRC5: plasma/serum circulating and exosomal RNA purification kit, 5 ml input; MAX0.1: Maxwell RSC miRNA plasma and exosome kit, 0.1 ml input; MAX0.5: Maxwell RSC miRNA plasma and exosome kit, 0.5 ml input; MIR0.2: miRNeasy serum/plasma kit, 0.2 ml input; MIRA0.2: miRNeasy serum/plasma advanced kit, 0.2 ml input; MIRA0.6: miRNeasy serum/plasma advanced kit, 0.6 ml input; MIRV0.1: mirVana PARIS kit, 0.1 ml input; MIRV0.625: mirVana PARIS kit, 0.625 ml input; MIRVE0.1: mirVana PARIS kit with enrichment, 0.1 ml input; MIRVE0.625: mirVana PARIS kit with enrichment, 0.625 ml input; NUC0.3: NucleoSpin miRNA plasma kit, 0.3 ml; NUC0.9: NucleoSpin miRNA plasma kit, 0.9 ml input)