

SUPPLEMENTARY MATERIAL - RLE

1 Model control of reference based RMA normalization

Given the large difference in misclassifications between one-by-one and reference normalized samples, we would like to be able to distinguish between the two in a setting where we are unsure how closely the RMA reference resembles the samples.

1.1 Relative Log Expression

The relative log expression (RLE) is a quality measure for microarrays introduced by Bolstad et al. [1]. If we define $\hat{\theta}_{ij}$ as the estimated log expression for probe i on array j following RMA normalization, the RLE is given by:

$$\text{RLE}(\hat{\theta}_{ij}) = \hat{\theta}_{ij} - \text{median}_j(\hat{\theta}_{ij})$$

i.e. the difference between the expression on a given array and the median expression in the cohort. A non-zero median RLE across probes for an array thus indicates differences in the number of up- and downregulated genes, while a large interquartile range (IQR) indicates that most genes on a given array are differentially expressed [2]. Extreme values of these measures may be used to identify arrays with low-quality data. We propose that it may also be used to evaluate how well a sample resembles a given RMA reference following normalization (how well it has been normalized), by substituting the cohort median with the RMA reference median.

2 Results

For each of the five datasets used in the current study a random subset of 30 samples were extracted and set as a user reference. The remaining samples for each dataset were then RMA normalized as a cohort, against the randomly selected user reference (reference based) or against the four other datasets (one-by-one). RLE values were calculated in all six scenarios and summarized as the median and IQR for each sample. These values are shown in panel A and B in fig. 1 to fig. 5. We find that the RLE values in the user reference scenarios more closely resemble the values calculated from cohort normalization than the values from one-by-one normalization. Furthermore differences between reference-based and on-by-one are more pronounced for the IQR of RLE than for the median. ROC curves showing the ability of the RLE median or IQR to distinguish between a user reference and a one-by-one reference were calculated and plotted using the pROC package version 1.8 [3] in R, as shown in panels C and D in fig. 1 to fig. 5. The area under the curve (AUC) for the ROC curves confirm the superiority of the RLE IQR for distinguishing between a "correct" user reference and a "wrong" reference, i.e. in most cases we observe a higher AUC when using the IQR instead of the median RLE.

Using Youden's index, which maximizes the sum of sensitivity and specificity, we calculated the optimal threshold for each ROC curve based on the RLE IQR as shown in table 1. Calculating the median RLE IQR across all datasets and one-by-one RMA references gives a value of 0.615. By rounding of we set a conservative threshold of 0.6 for reference normalized samples, i.e. if the RLE IQR exceeds this value we advise not to include the sample in subsequent classifications. This does not remove all one-by-one normalized samples, but in most cases it prevents the user from getting poor classifications resulting from biased RMA normalization.

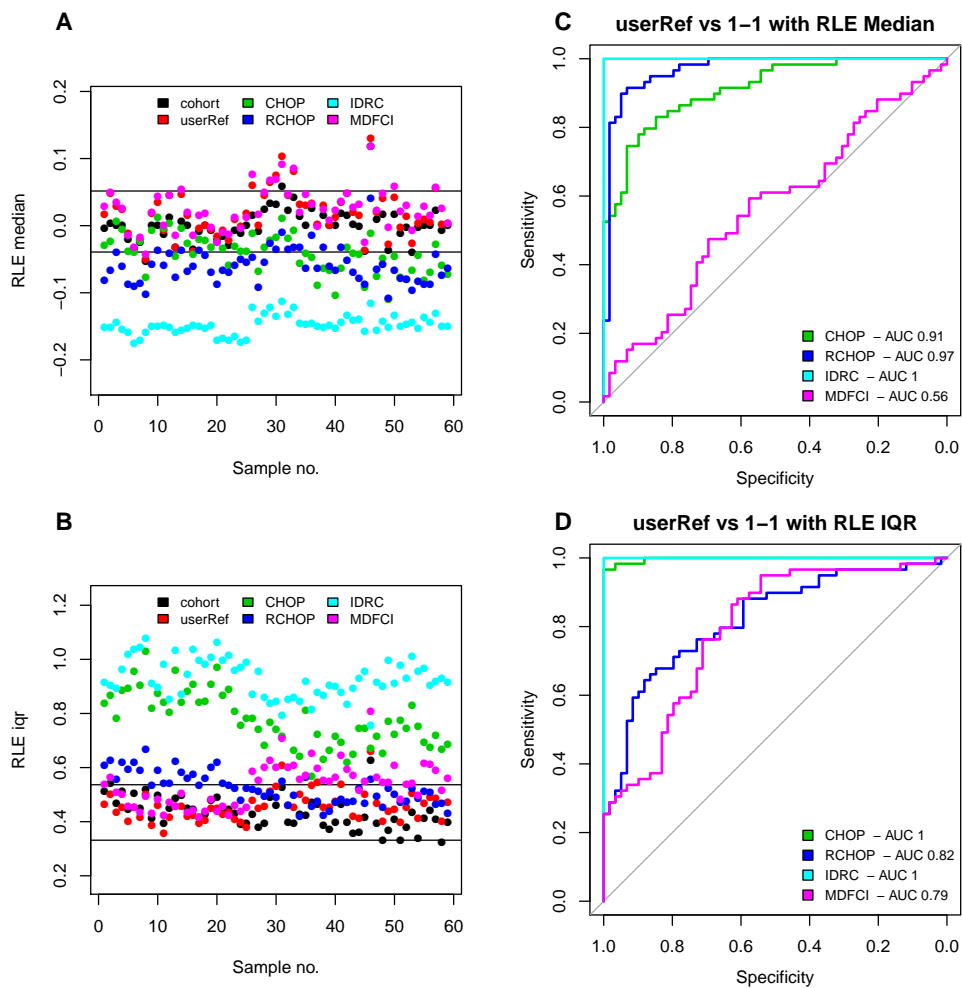


Figure 1: Median and IQR RLE values for different RMA normalizations of the CHEPRETRO dataset and ROC curves

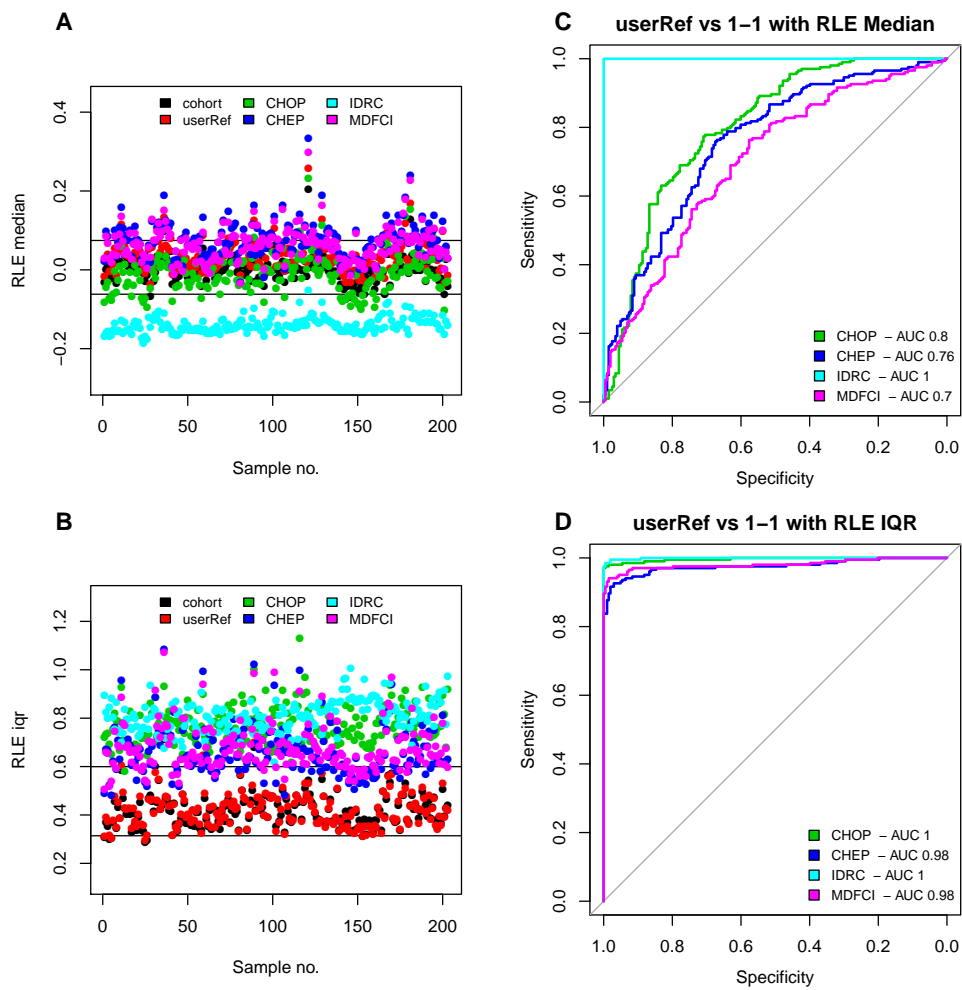


Figure 2: Median and IQR RLE values for different RMA normalizations of the RCHOP dataset and ROC curves

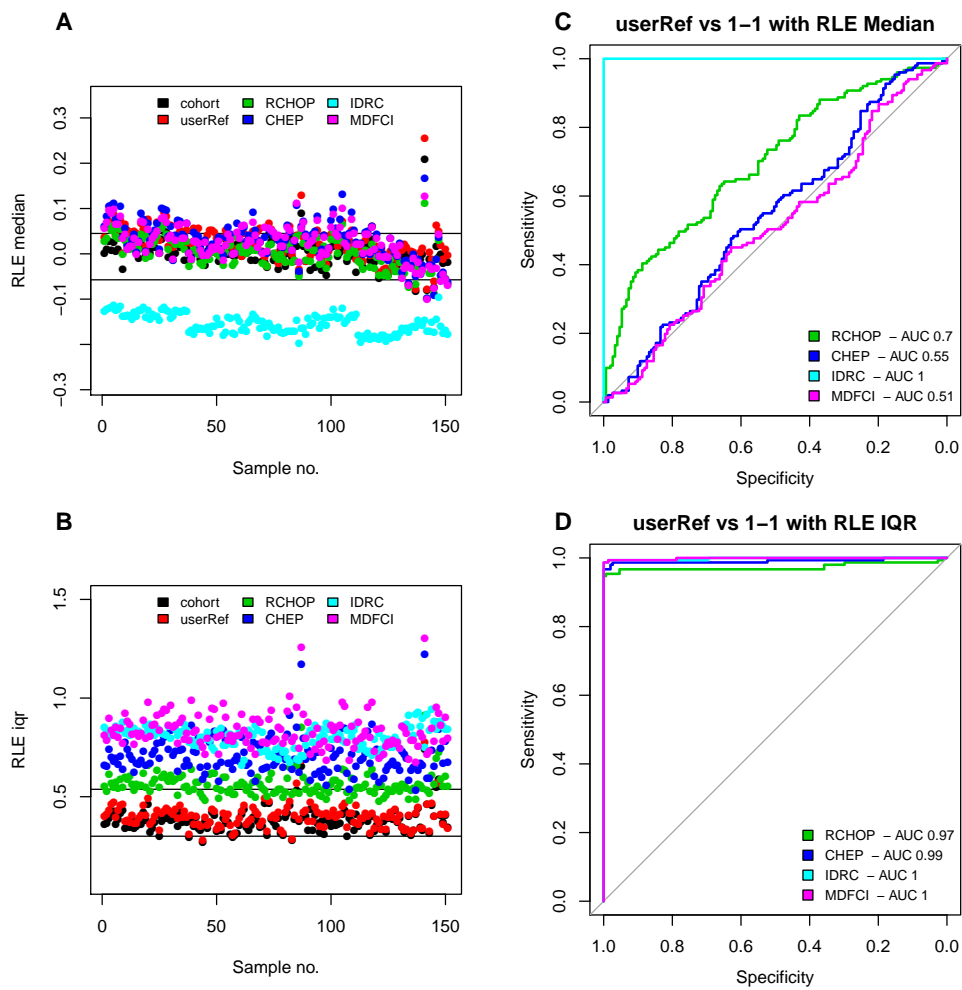


Figure 3: Median and IQR RLE values for different RMA normalizations of the CHOP dataset and ROC curves

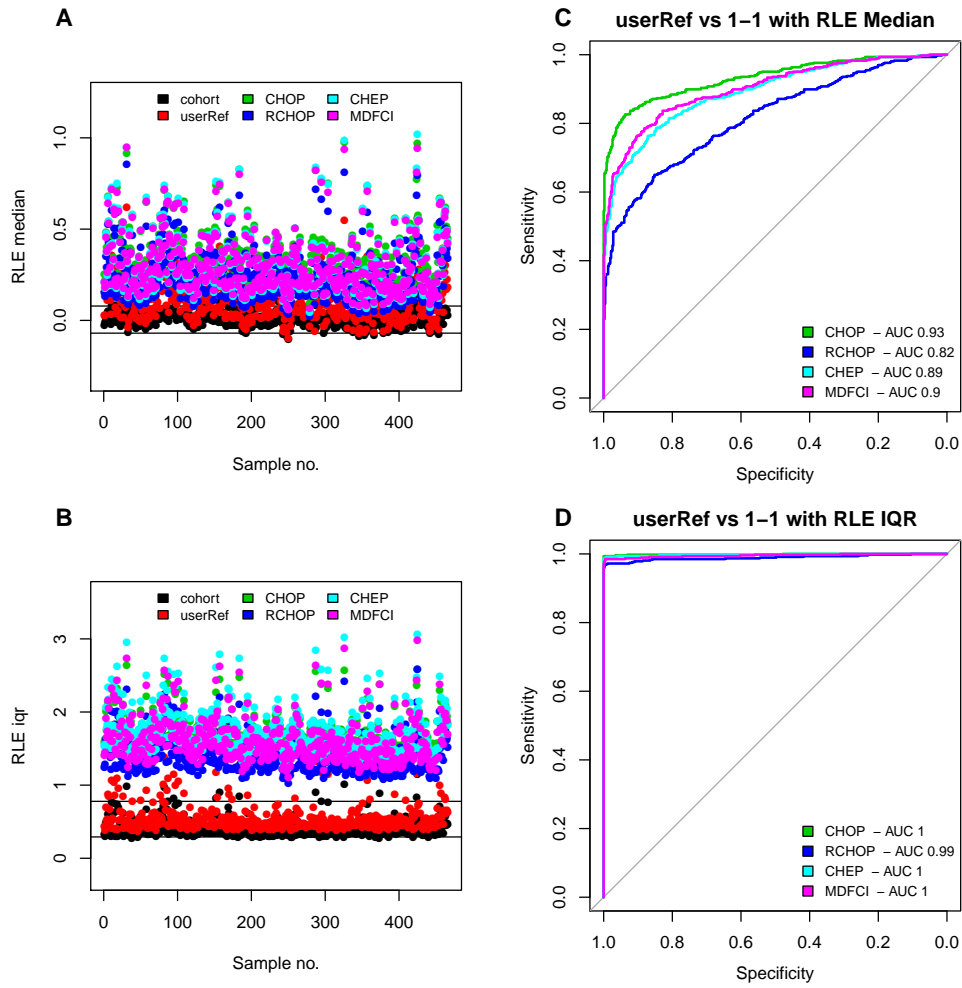


Figure 4: Median and IQR RLE values for different RMA normalizations of the IDRC dataset and ROC curves

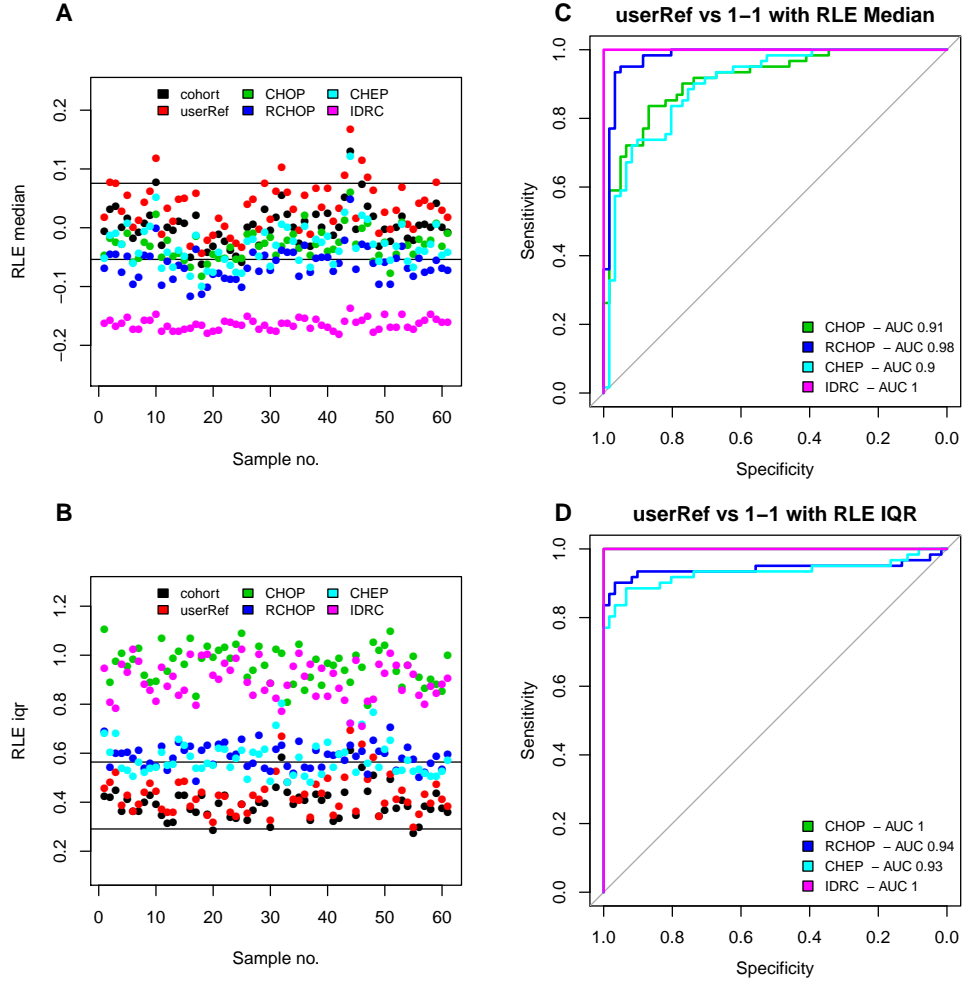


Figure 5: Median and IQR RLE values for different RMA normalizations of the MDFCI dataset and ROC curves

Dataset	RMA reference	Threshold	Sensitivity	Specificity
CHEPRETRO	CHOP	0.56	0.97	1.00
CHEPRETRO	RCHOP	0.47	0.68	0.85
CHEPRETRO	IDRC	0.71	1.00	1.00
CHEPRETRO	MDFCI	0.54	0.95	0.54
RCHOP	CHOP	0.61	0.98	1.00
RCHOP	CHEPRETRO	0.52	0.93	0.97
RCHOP	IDRC	0.68	0.99	1.00
RCHOP	MDFCI	0.53	0.94	0.99
CHOP	RCHOP	0.48	0.95	0.99
CHOP	CHEPRETRO	0.51	0.97	1.00
CHOP	IDRC	0.62	0.99	1.00
CHOP	MDFCI	0.62	0.99	1.00
IDRC	CHOP	1.39	0.99	1.00
IDRC	RCHOP	1.08	0.97	1.00
IDRC	CHEPRETRO	1.32	0.99	1.00
IDRC	MDFCI	1.19	0.98	1.00
MDFCI	CHOP	0.74	1.00	1.00
MDFCI	RCHOP	0.51	0.90	0.97
MDFCI	CHEPRETRO	0.50	0.89	0.93
MDFCI	IDRC	0.70	1.00	1.00

Table 1: Optimal thresholds for RLE IQR

References

- [1] B.M. Bolstad, F. Collin, K.M. Simpson, R.A. Irizarry, and T.P. Speed. Experimental Design and Low-Level Analysis of Microarray Data. *International Review of Neurobiology*, 60:25–58, 2004.
- [2] Matthew N McCall, Peter N Murakami, Margus Lukk, Wolfgang Huber, and Rafael a Irizarry. Assessing affymetrix GeneChip microarray quality. *BMC bioinformatics*, 12(1):137, 2011.
- [3] Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frdrique Lisacek, Jean-Charles Sanchez, and Markus Mller. proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*, 12:77, 2011.