Supporting Information

Doc S1. Algorithm: combinatorial optimization for multicandidate miRNAs

Notation:

Nm: Number of miRNAs.

mR(i): The i th miRNA where i = 1 TO Nm.

Ns: Number of miRNA sets with better performance to limit at each step.

Mc: Maximum number of miRNA combinations that will be examined.

Sc: Score (same as Accuracy) = (TP + TN) / (TP + FP + FN + TN), where the meaning of each variable is as follows:

	Cancer	No cancer
Prediction positive	TP: number of true positives	FP: number of false positives
Prediction negative	FN: number of false negatives	TN: number of true negatives

LDA-LOO(): Function of Linear Discriminant Algorithm (LDA) with Leave-One-Out Cross-Validation (LOO-CV). We used LDA function included as lda() in R language MASS package version 7.3-45. The score (Sc) by LOO-CV is calculated as the accumulated number, counting the results (hit or not) of evaluating each leave-out sample, divided by number of trials.

Processing steps:

Step 1: Division of the data to be analyzed into halves for training data and test data.

Step 2: Setting of Ns = 20 and Mc = 10.

(For training data)

Step 3: Combinatorial optimization

```
FOR I = 1 TO Nm
        Sc[i] = LDA-LOO( mR(i) )
miRNA_index_of_sorted[1:Nm] = ORDER( Sc[1:Nm], descending order )
Pivot[1:Ns] = miRNA_index_of_sorted[1:Ns]      # get indexes of top Ns as Pivot
```

```
i = 1
FOR j = 1 TO Mc
        Sc = 0
        FOR k = 1 TO Nm
                IF k is not included in Pivot_set[j] {Sc[i] = LDA-LOO( mR[Pivot_set[j]], mR[k] )}
                i = i + 1
        miRNA_set[Ns * (Nm-i)] = ORDER( Sc[Ns * (Nm-i)], descend order )
        Pivot_set[1:Ns] = miRNA_set[1:Ns]      # get indexes of top Ns as Pivot
```

Step 4: Calculation of coefficients of Ns set of linear discriminant models.

Step 5: Calculation of INDEX: An offset is added to the discrimination model such that the cutoff value becomes 0.

Step 6: Calculation of performances for Ns discriminant models with sensitivity, specificity, accuracy, positive predictive value, negative predictive value, and area under the ROC curve on the training data. (For test data)

Step 7: Calculation of the same performances (as above) using the Ns discriminant models for test data.

Table S1. Patient characteristics

	N	Age	LDH level
Leiomyosarcoma	6	59.7 ± 16.5	219.0 ± 83.2
Adenosarcoma	2	46.0 ± 5.7	186.5 ± 65.8
ESS	2	43.0 ± 7.1	197.5 ± 37.5
STUMP	1	65	281
Leiomyoma	18	46.8 ± 11.1	286.0 ± 46.3

ESS, endometrial stromal sarcoma; STUMP, smooth muscle tumor of uncertain malignant potential.

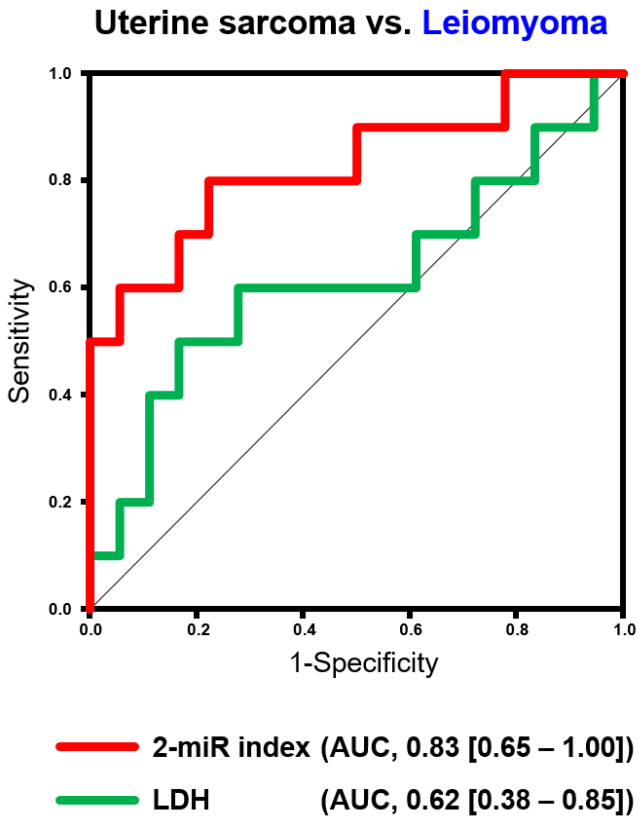

Figure S1. Performance of the diagnostic model for uterine sarcomas. ROC curves for distinguishing US patients from ULM controls using the two-miRNA strategy developed for diagnostic model 1. N = ULMS 6; ULM, 18; ESS, 2; adenosarcoma, 2; STUMP, 1.