Supplementary Methods

Removal of hematoxylin contamination from RNA

Total RNA from 4 CRC cases was extracted from FFPE samples by the PureLink FFPE Total RNA Isolation Kit. Despite performing the washing step 3 times in the extraction process, the extracted RNA samples were blue instead of clear, probably due to hematoxylin contamination (**Supplementary Fig. S1**). Because the washing buffer contains alcohol, it is possible that hematoxylin, which is not dissolved in alcohol, could not be removed by washing. Following this, the concentration and purity of 4 RNA samples were quantified. Although the OD 260/280 ratio of RNA samples was greater than 2.00 (ranging from 2.01 to 2.08), the RNA samples had a low OD 260/230 ratio (ranging from 0.74 to 1.18), probably due to hematoxylin contamination. Therefore, we added a centrifugation step after proteinase K digestion to remove hematoxylin, and extracted total RNA from the same samples (adjacent FFPE sections) again. By this method, RNA samples were clear but not blue, and OD 260/230 ratios were improved (ranging from 1.47 to 1.92; **Supplementary Fig. S1**).

Because hematoxylin contamination in RNA samples resulted in low OD260/230 ratios, we investigated the influence of hematoxylin contamination on qRT-PCR results. We first compared *miR-21* expression levels in samples with and without centrifugation using 4 of the Japanese cohort RNA samples. Although RNA was extracted from almost the same FFPE sections (adjacent FFPE sections), *miR-21* expression levels varied. Before normalization by *RNU48*, the expression level of *miR-21* without centrifugation was less than half of that after centrifugation in 2 of the 4 samples (**Supplementary Fig. S2A-C**). Even when expression levels of *miR-21* were normalized by *RNU48*, the expression level of *miR-21* without centrifugation was still less than half the value with centrifugation in the same 2 samples (**Supplementary Fig.**

S2A-C). These results suggest that qRT-PCR results from RNA samples with hematoxylin contamination do not accurately reflect those without hematoxylin contamination. Therefore, we further investigated the direct effect of hematoxylin contamination on qRT-PCR results. RNA from the HCT116 CRC cell line was extracted (OD 260/230 ratio, 2.27), mixed with serially diluted hematoxylin, and the concentration and purity of RNA samples were quantified (**Supplementary Fig. S2D-F**). The concentration of RNA samples with added hematoxylin was shown to not match those without hematoxylin, and a high amount of hematoxylin was correlated with a low OD 260/230 ratio. Therefore, the low OD260/230 ratios of RNA extracted from FFPE samples are due, at least in part, to hematoxylin contamination. Next, we performed qRT-PCR analysis on these samples. In general, both *miR-21* and *RNU48* expression levels in samples with hematoxylin were lower than those without hematoxylin. Before normalization by *RNU48*, expression levels of *miR-21* from RNA samples with more than 1/1000 hematoxylin (OD 260/230 ratio, 1.33) were less than half those of RNA samples without hematoxylin. However, after normalization with *RNU48*, expression levels of *miR-21* from RNA samples with more than 1/400 hematoxylin (OD 260/230 ratio, 1.02) had less than half the expression level of RNA samples without hematoxylin. Therefore, we conclude that RNA samples with OD260/230 ratios of 1.00 or less are not suitable for qRT-PCR analysis.

Supplementary Figure S1.

RNA extraction method in the Japanese cohort. When RNA was extracted without centrifugation after proteinase K digestion, the extracted RNA sample was blue instead of clear. In contrast, when RNA was extracted with centrifugation after proteinase K digestion, the extracted RNA sample was clear.

Supplementary Figure S2.

The influence of hematoxylin contamination on qRT-PCR results. A-C, Expression levels of miR-21 and RNU48 in 4 Japanese-cohort RNA samples extracted with or without centrifugation. A, Expression levels of miR-21 before normalization by RNU48. B, Expression levels of RNU48. C, Expression levels of miR-21 after normalization by RNU48. The units are arbitrary, and expression levels of miR-21 and RNU48 were standardized to 40ng total RNA from HCT116, defined as 1.0. D-F, qRT-PCR results of RNA extracted from HCT116 with or without hematoxylin. HCT116 RNA was extracted, mixed with serially diluted hematoxylin, and the concentration and purity of RNA samples was quantified. D, Expression levels of miR-21 before normalization with RNU48. E, Expression levels of RNU48. F, Expression levels of miR-21 after normalization with RNU48. The units are arbitrary, and expression levels of miR-21 and RNU48 were standardized to 40ng total RNA from HCT116 without hematoxylin, defined as 1.0.

Supplementary Figure S3.

Expression of miR-21 in the Japanese and German cohort stratified by clinicopathological features. Dot plots represent miR-21 relative threshold cycle values from the qRT-PCR.

Horizontal red bars indicate the median expression value. Differences in miR-21 expression levels between groups were tested by the Mann–Whitney *U* test.

Abbreviations: NS, not significant.

Supplementary Figure S4.

Kaplan–Meier analysis of miR-21 levels in the Japanese cohort (stage I-IV). Kaplan–Meier plot of the overall survival in the Japanese cohort.

Supplementary Figure S5.

Kaplan–Meier analysis of miR-21 stratified by MSI status. Kaplan-Meier analysis of MSI status in the Japanese (A) and German (C) cohorts. Kaplan–Meier analysis of miR-21 stratified by MSI status in the Japanese (B) and German (D) cohorts.

Fig. S1

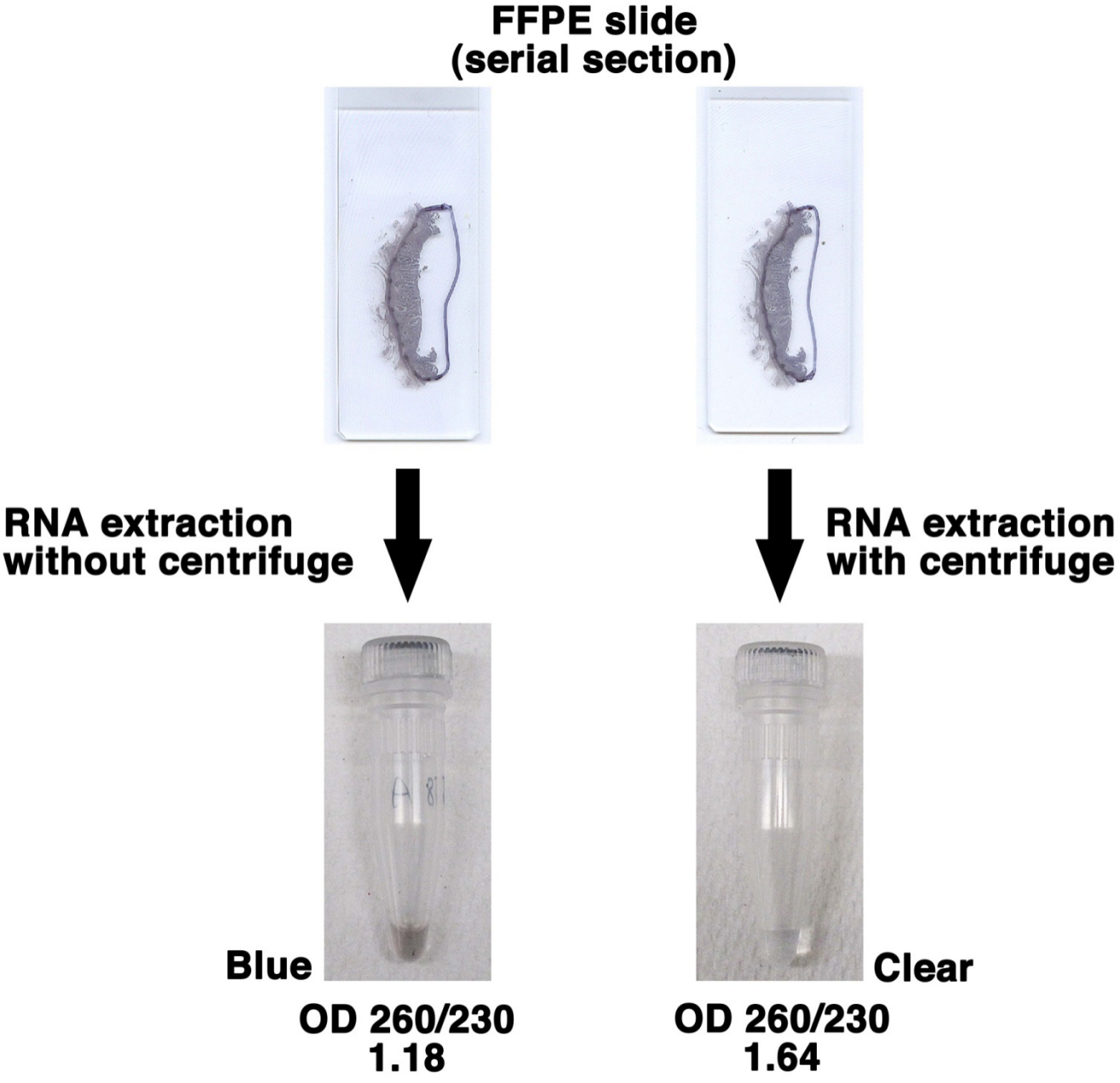

FFPE slide
(serial section)
RNA extraction
without centrifuge
RNA extraction
with centrifuge
Blue
Clear
OD 260/230
1.18
OD 260/230
1.64

Fig. S2

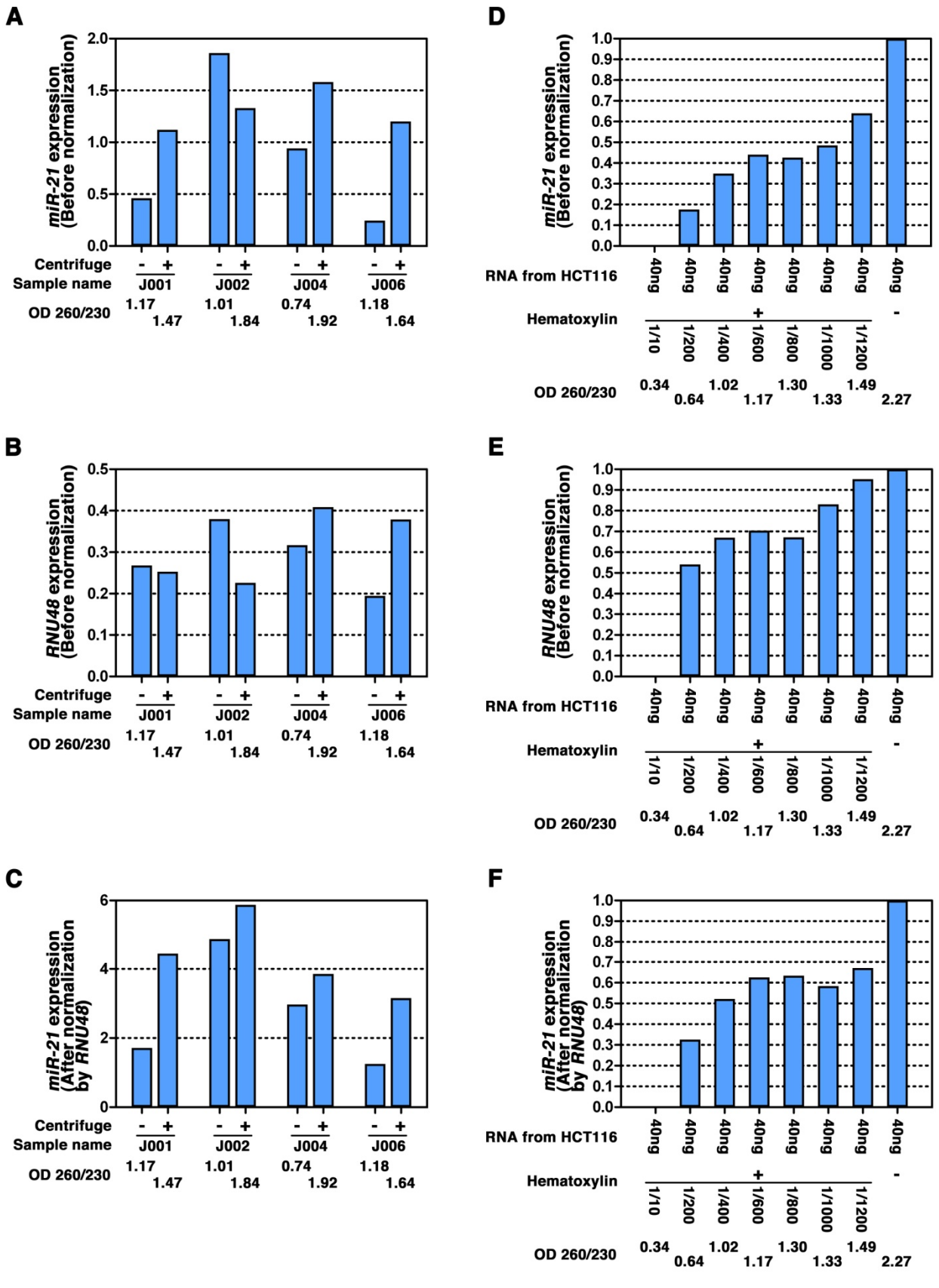
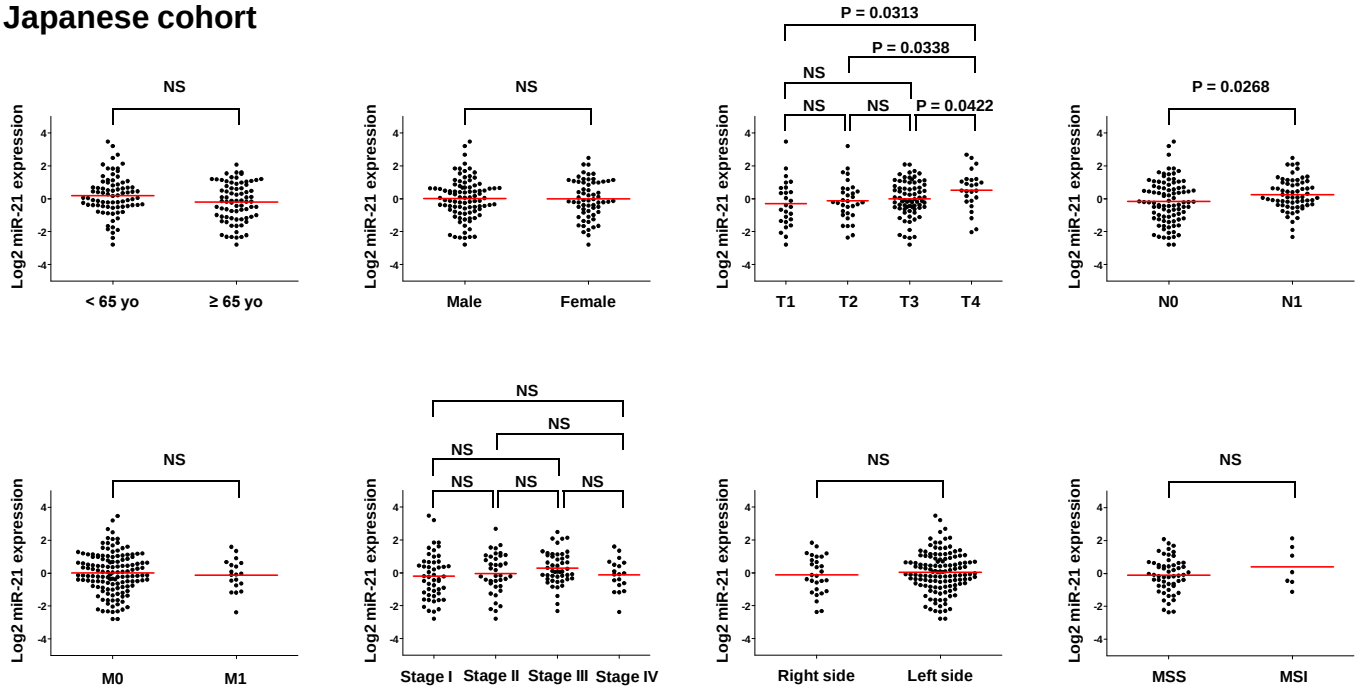

A
miR-21 expression (Before normalization)
Centrifuge
Sample name
J001
J002
J004
J006
OD 260/230
1.17
1.47
1.01
1.84
0.74
1.92
1.18
1.64
B
RNU48 expression (Before normalization)
C
miR-21 expression (After normalization by RNU48)
D
RNA from HCT116
40ng
Hematoxylin
1/10
1/200
1/400
1/600
1/800
1/1000
1/1200
0.34
0.64
1.02
1.17
1.30
1.33
1.49
2.27
E
F

Fig. S3

Japanese cohort



German cohort

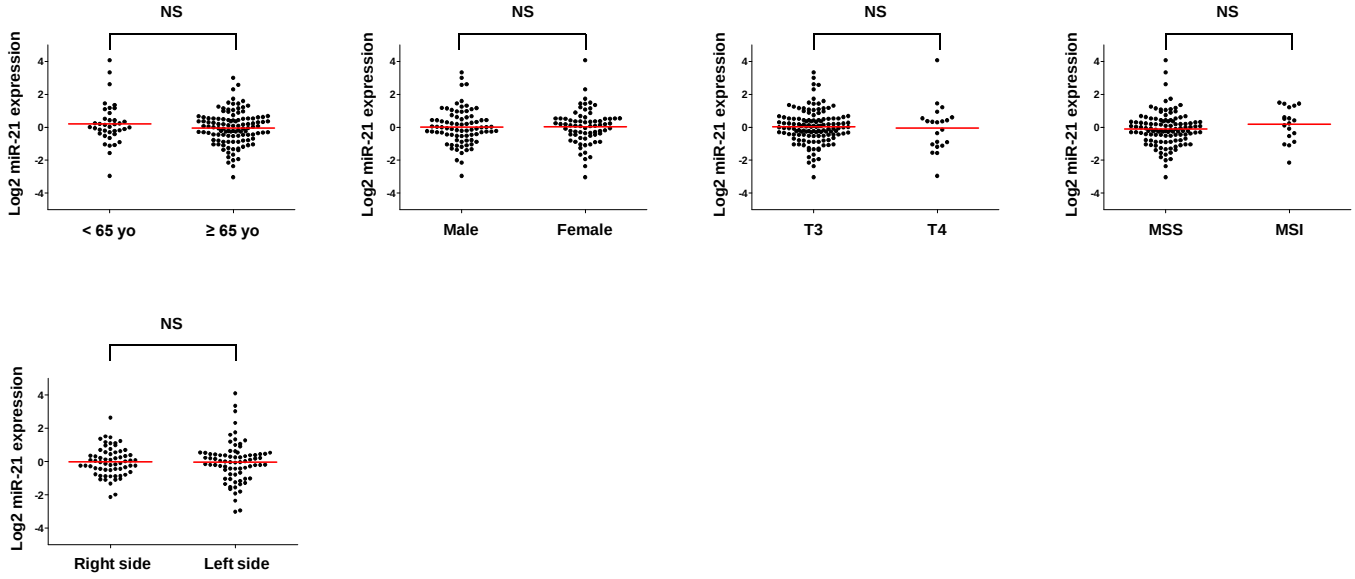
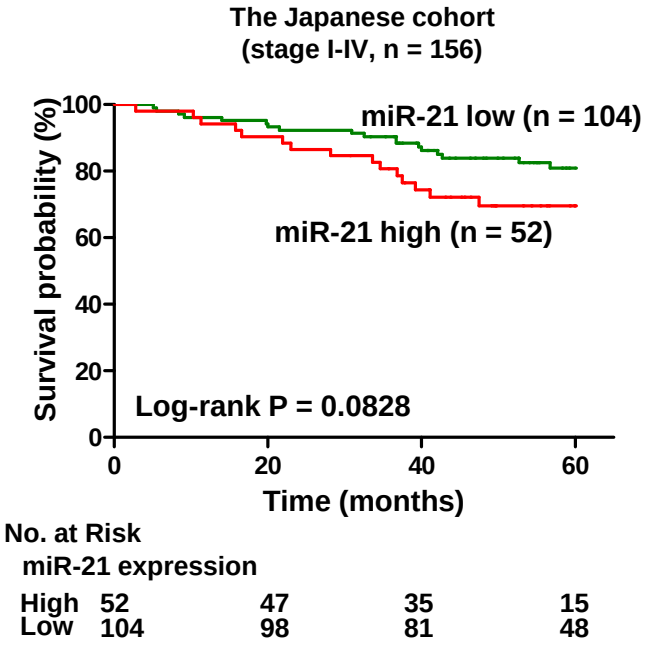


Fig. S4



No. at Risk

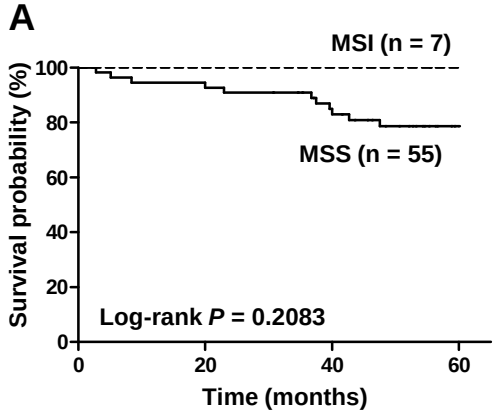
miR-21 expression				
High	52	47	35	15
Low	104	98	81	48

Fig. S5

The Japanese cohort (stage I-IV, n = 62)

MSI status

A

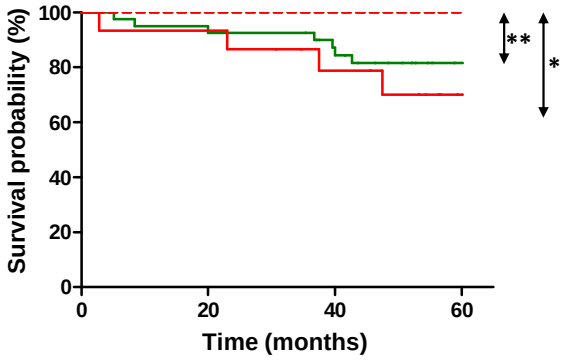


No. at Risk

MSI status				
MSI	7	7	7	4
MSS	55	51	41	23

***miR-21* expression and MSI status**

B



No. at Risk

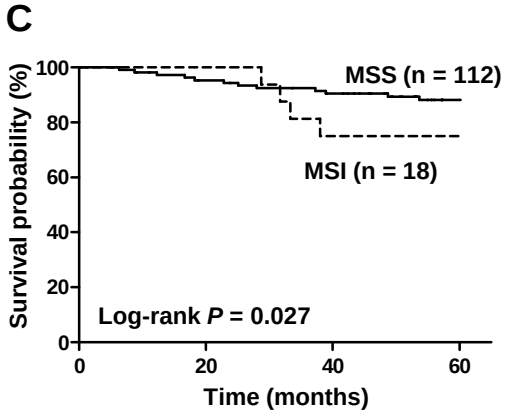
MSI group by *miR-21* expression				
High	3	3	3	1
Low	4	4	4	3
MSS group by *miR-21* expression				
High	15	14	10	3
Low	40	37	31	20

* High-*miR-21* expression group
MSI group vs MSS group, Log-rank *P* = 0.34

** Low-*miR-21* expression group
MSI group vs MSS group, Log-rank *P* = 0.37

The German cohort (stage II, n = 130)

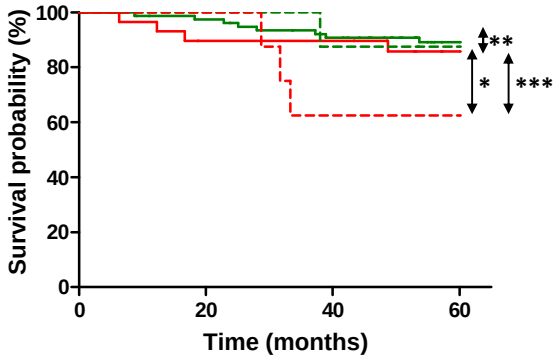
C



No. at Risk

MSI status				
MSI	18	18	12	11
MSS	112	100	92	67

D



No. at Risk

MSI group by *miR-21* expression				
High	9	9	6	5
Low	9	9	7	6
MSS group by *miR-21* expression				
High	30	25	25	18
Low	82	75	67	49

* MSI group
high vs low *miR-21* expression, Log-rank *P* = 0.22

** MSS group
high vs low *miR-21* expression, Log-rank *P* = 0.58

*** High-*miR-21* expression group
MSI group vs MSS group, Log-rank *P* = 0.19

MSS group by miR-21 expression	MSI group by miR-21 expression
High	High
Low	Low

Supplementary Table S1. Univariate and multivariate Cox regression analysis of miR-21 expression levels and overall survival in the Japanese cohort (stage I-IV, n = 156)

	Univariate analysis		Multivariate analysis[a]	
Characteristic	HR (95% CI)	*P* value	HR (95% CI)	*P* value
miR-21 expression				
Low	1 (Ref.)	0.087	1 (Ref.)	0.094
High	1.81 (0.92-3.61)		1.80 (0.91-3.58)	
TNM stage				
Stage I/II	1 (Ref.)	< 0.001	1 (Ref.)	<0.001
Stage III/IV	6.09 (2.64-14.07)		6.1 (2.65-14.12)	
Age				
< 65	1 (Ref.)	0.113	1 (Ref.)	0.078
65 and > 65	1.76 (0.88-3.54)		1.87 (0.93-3.77)	
Tumor Location				
Proximal	1 (Ref.)	0.923		
Distal	0.95 (0.40-2.32)			
Sex				
Male	1 (Ref.)	0.84		
Female	0.81 (0.40-1.64)			

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a]Multivariable model was selected based on stepwise removal of variables with a significance threshold for removal set at p<0.1.

Supplemantary Table S2. Univariate and multivariate Cox regression analysis of miR-21 expression and MSI status with overall survival in the German cohort (n = 145)

	Univariate analysis		Multivariate analysis (n=125)[a]	
Characteristic	HR (95% CI)	*P* value	HR (95% CI)	*P* value
miR-21 expression				
Low	1 (Ref.)	0.055	1 (Ref.)	0.239
High	2.42 (0.98-5.95)		1.95 (0.64-5.96)	
T classification				
T3	1 (Ref.)	0.06	1 (Ref.)	0.024
T4	2.67 (0.96-7.41)		4.25 (1.21-14.90)	
Tumor Location (n=139)				
Distal	1 (Ref.)	0.118	1 (Ref.)	0.052
Proximal	2.26 (0.81-6.27)		0.27 (0.07-1.01)	
MSI status (n=130)				
MSS	1 (Ref.)	0.193	1 (Ref.)	0.347
MSI	2.12 (0.68-6.58)		2.00 (0.474-8.41)	
Adjuvant chemotherapy[b]				
Did not received	1 (Ref.)	0.879		
Received	0.91 (0.26-3.12)			
Age				
< 65	1 (Ref.)	0.16		
65 and > 65	2.87 (0.66-12.46)			
Sex				
Male	1 (Ref.)	0.648		
Female	1.23 (0.50-3.04)			

Abbreviations: HR, hazard ratio; CI, confidence interval; MSI, microsatellite instability; MSS, microsatellite stable.

[a]The multivariate model included miR-21, T classification, tumor location and MSI status. Only the 125 cases for which MSI status and Tumor location was available was used for the multivariate analysis. [b]Chemotherapy regimens were primarily 5-fluorouracil based regimens.