A novel mesenchymal-associated transcriptomic signature for risk-stratification and therapeutic response prediction in colorectal cancer

Takatoshi Matsuyama, **Raju Kandimalla**, Toshiaki Ishikawa, Naoki Takahashi, Yasuhide Yamada, Masamichi Yasuno, Yusuke Kinugasa, Torben Frostrup Hansen, Marwan Fakih, Hiroyuki Uetake, Balázs Győrffy, and Ajay Goel

Supplementary Online Content

Supplementary Table S1:

Clinicopathological characteristics of in-house training and validation cohorts

Supplementary Table S2:

Primer sequence IDs

Supplementary Table S3:

The coefficients of seven genes MATS classifier in the cox regression model

Supplementary Table S4:

The list of candidate genes, which have positive correlation with VIM in our laser capture micro dissected transcriptomic data

Supplementary Table S5:

The list of candidate EMT genes which were upregulated in cancer vs. normal and these were used for lasso cox regression model

Supplementary Table S6:

The AUC of each candidate gene for identifying CMS4 subtype in six publicly available datasets

Supplementary Figure S1:

Study Design

Supplementary Figure S2:

The association between each MATS gene and CMS status in GSE39582 dataset.

Supplementary Figure S3:

The association between each MATS gene and CMS status in GSE17536 dataset.

Supplementary Figure S4:

The association between each MATS gene and CMS status in GSE33113 dataset.

Supplementary Figure S5:

The association between each MATS gene and CMS status in TCGA RNA seq dataset.

Supplementary Figure S6:

The association between each MATS gene and CMS status in TCGA microarray dataset.

Supplementary Figure S7:

The association between each MATS gene and CMS status in GSE104645 dataset.

Supplementary Figure S8:

Association of MATS with TP53, BRAF and KRAS gene mutations status.

Supplementary Figure S9:

Survival analyses in stage II, and stage III patients in the validation cohort.

Supplementary Table S1: Clinicopathological characteristics of in-house training and validation cohorts

		Training cohort N (%) (N=142)	Validation cohort N (%) (N=286)
Gender			
	Male	84 (59)	162 (57)
	Female	58 (41)	124 (43)
Age			
	Mean±SD	59 ±10	68 ±11
Location			
	Rectum	71 (50)	102 (36)
	Left colon	47 (33)	89 (31)
	Right colon	24 (17)	95 (33)
Size			
	Mean±SD (mm)	49 ±17	53 ±37
Tumor grade			
	Low	132 (93)	261 (91)
	High	10 (7)	25 (9)
T stage			
	T4	41 (29)	88 (31)
	T1-3	101 (71)	198 (69)
Lymphovascular invasion			
	Absent	39 (27)	22 (8)
	Present	103 (73)	263 (92)
	unknown	0	1 (0)
No. of lymphnode examined			
	<12	13 (9)	50 (17)
	12 or more	129 (91)	234 (82)
	unavailable	0	2 (1)
TNM stage			
	II	54 (38)	161 (56)
	III	88 (62)	125 (44)
Adjuvant therapy			
	No	73 (51)	187 (65)
	Yes	69 (49)	99 (35)
MSI status			
	MSI-H	-	25 (9)
	MSS, MSH-L	-	251 (88)
	unavailable	-	10 (3)
Median follow up period (Month)			
		71	47

Supplementary Table S2: Primer IDs from Thermo Fisher Scientific

Gene Name	Assay ID	Amplicon Length
FN1	Hs01549972_m1	63bp
COL3A1	Hs00943809_m1	65bp
PRR16	Hs01083087_m1	67bp
POSTN	Hs01566750_m1	67bp
BCAT1	Hs00398962_m1	74bp
COL1A2	Hs01028956_m1	71bp
DKK3	Hs00951304_m1	64bp
FSTL1	Hs00907496_m1	64bp

Supplementary Table S3: The coefficients of seven genes MATS classifier in the cox regression model

COL1A2	**COL3A1**	**FN1**	**POSTN**	**FSTL1**	**BCAT1**	**DKK3**
0.339183	0.419288	0.066279	0.566933	0.378621	-0.15683	-0.19965

Supplementary Table S4: The list of candidate genes, which have positive correlation with VIM in our laser capture micro dissected transcriptomic data

Symbol	Symbol	Symbol
COL3A1	ADAMTS2	CD14
COL1A2	FERMT2	CHN1
IGFBP7	NA	MYLK
SPARC	CDH11	GUCY1A3
DCN	GPNMB	TAGLN
COL5A2	WISP1	LOC100132891
ACTA2	PMP22	A2M
CTSK	MSN	TUBB6
TIMP2	NID2	CD163
FSTL1	SPARCL1	TNC
COL12A1	COL5A1	FPR3
COL4A1	SRGN	BCAT1
LGALS1	CRISPLD2	SERPINF1
COL6A3	ADAMTS12	CTSL
AXL	GUCY1B3	LOX
RFTN1	FRMD6	TMEM204
SDC2	CYBRD1	CTGF
FBN1	MEIS1	ADAM12
DKK3	LUM	ATP8B2
DDR2	THY1	PRRX1
C1S	INHBA	PRR16
POSTN	CTHRC1	LHFP
ANTXR1	GLT8D2	CTSC
MSRB3	COL15A1	SERPING1
VCAN	LAPTM5	HEG1
RAB31	PDLIM3	FN1
CALD1	MS4A4A	GPR176
NRP1	COL8A1	
SNAI2	FILIP1L	
SULF1	LPHN2	

Supplementary Table S5: The list of candidate EMT genes which were upregulated in cancer vs. normal and these were used for lasso cox regression model

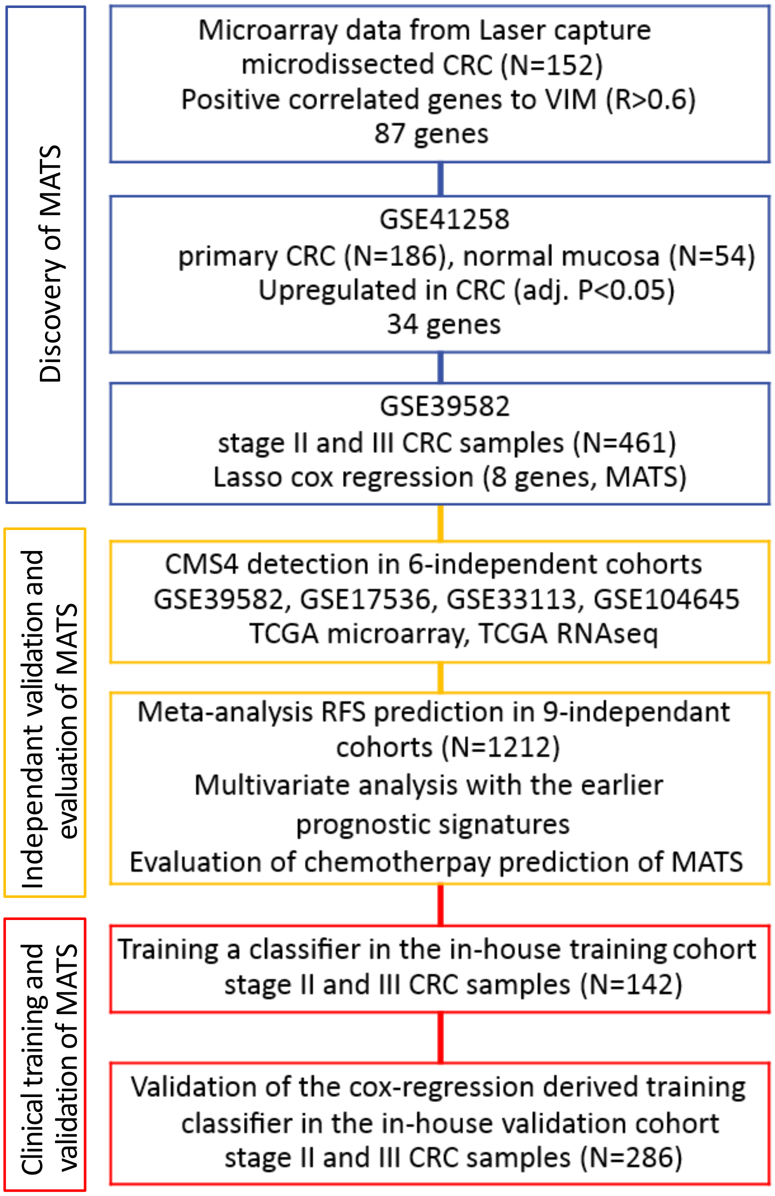
Symbol
COL3A1
COL1A2
IGFBP7
SPARC
COL5A2
CTSK
FSTL1
COL4A1
LGALS1
COL6A3
FBN1
DKK3
POSTN
ANTXR1
VCAN
RAB31
NRP1
SULF1
ADAMTS2
CDH11
WISP1
NID2
COL5A1
LUM
THY1
INHBA
COL8A1
CHN1
BCAT1
LOX
ADAM12
PRRX1
PRR16
FN1

Supplementary Table S6: The AUC of each candidate gene for identifying CMS4 subtype in six publicly available datasets

	MATS	**COL1A2**	**COL3A1**	**FN1**	**POSTN**	**FSTL1**	**BCAT1**	**DKK3**	**PRR16**
GSE17536	0.92	0.87	0.86	0.85	0.85	0.87	0.85	0.88	0.80
GSE33113	0.99	0.86	0.87	0.82	0.90	0.95	0.89	0.90	0.65
GSE39582	0.94	0.88	0.90	0.85	0.85	0.91	0.84	0.86	0.78
TCGA RNA seq	0.97	0.93	0.94	0.89	0.91	0.97	0.92	0.92	0.89
TCGA microarray	0.95	0.92	0.83	0.81	0.86	0.93	0.82	0.91	0.85
GSE104645	0.92	0.86	0.82	0.87	0.77	0.76	0.48	0.75	0.62

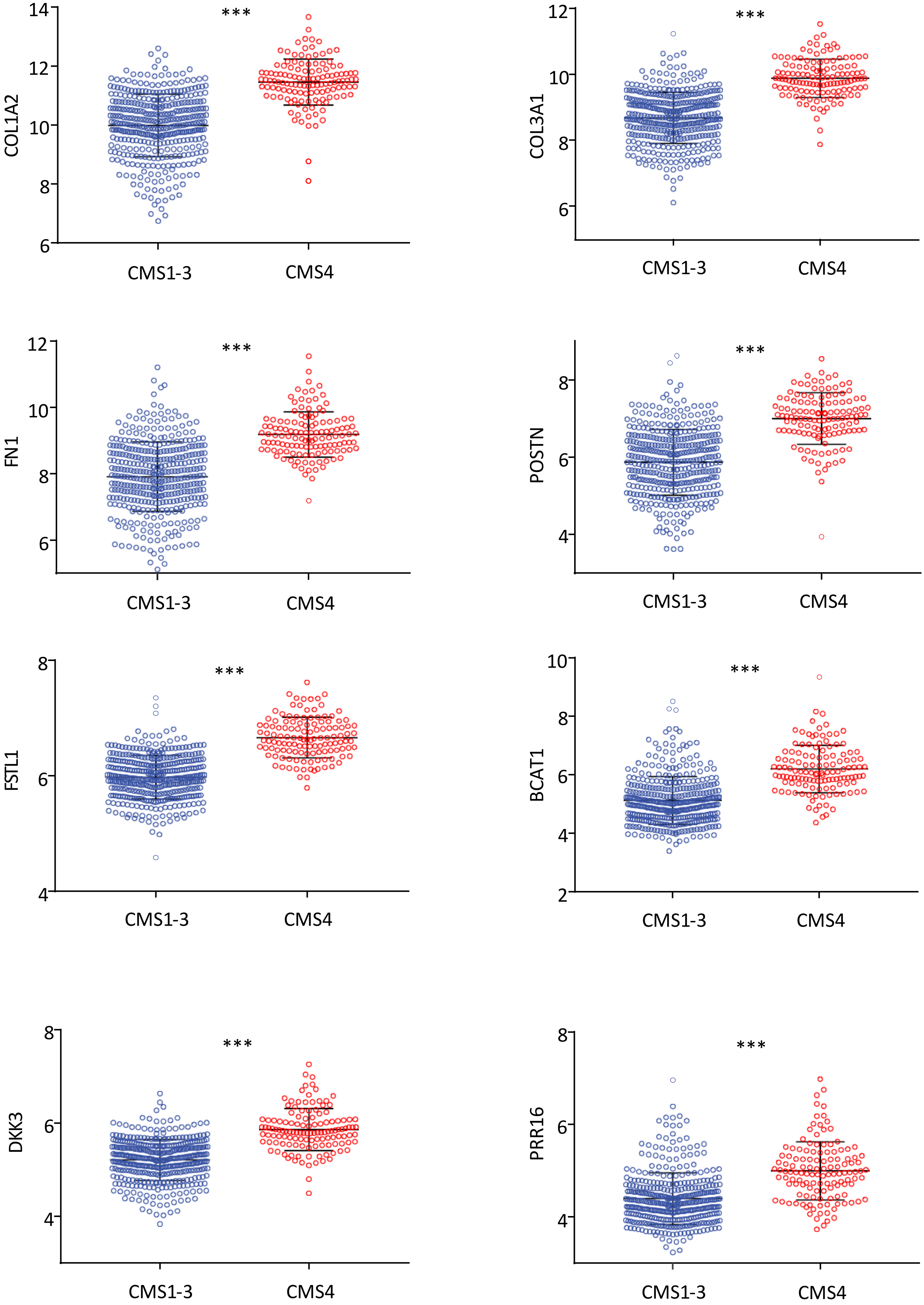
Supplementary Figure S1

Study Design



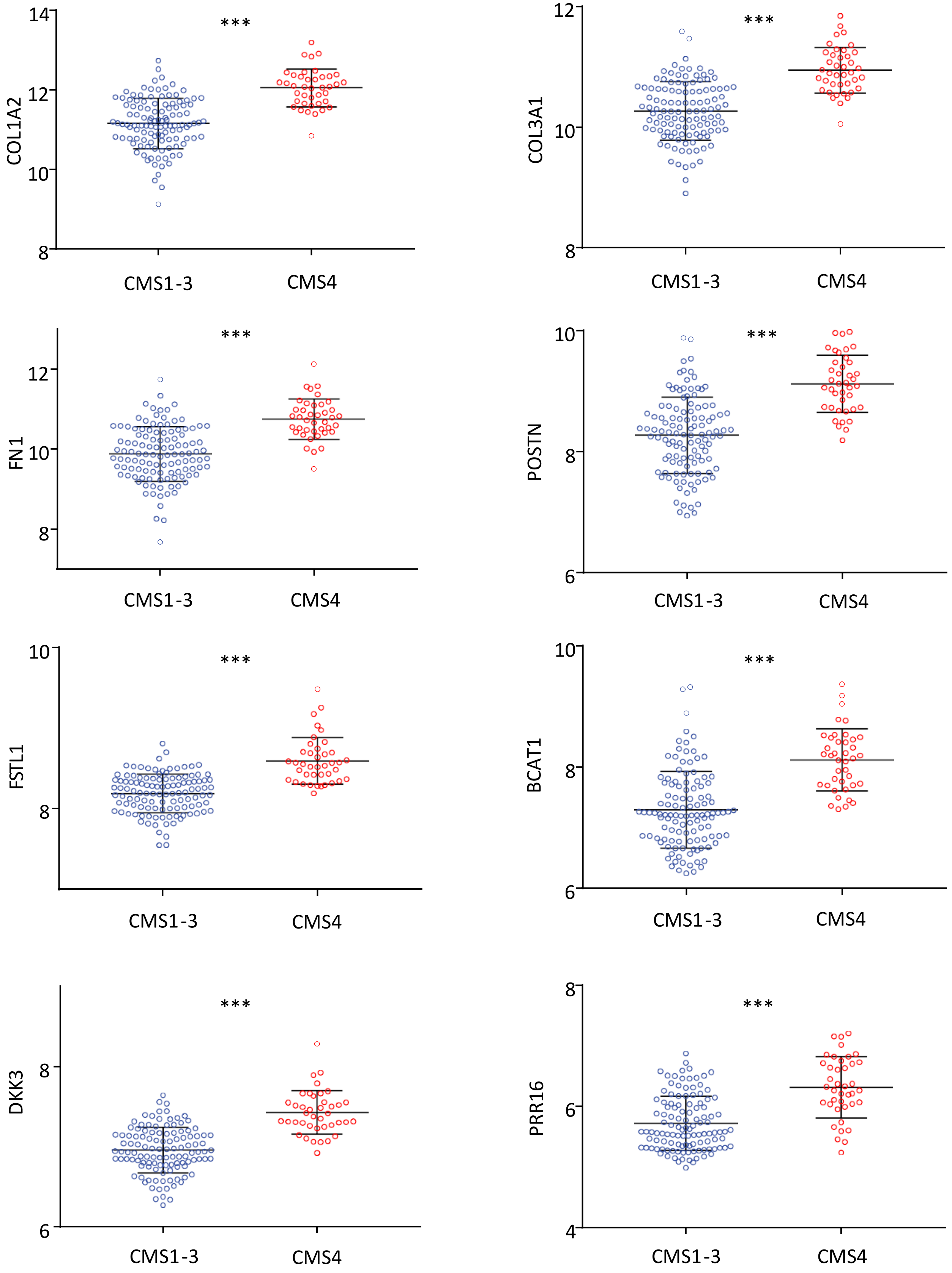
Supplementary Figure S2:

The association between each MATS gene and CMS status in GSE39582 dataset.



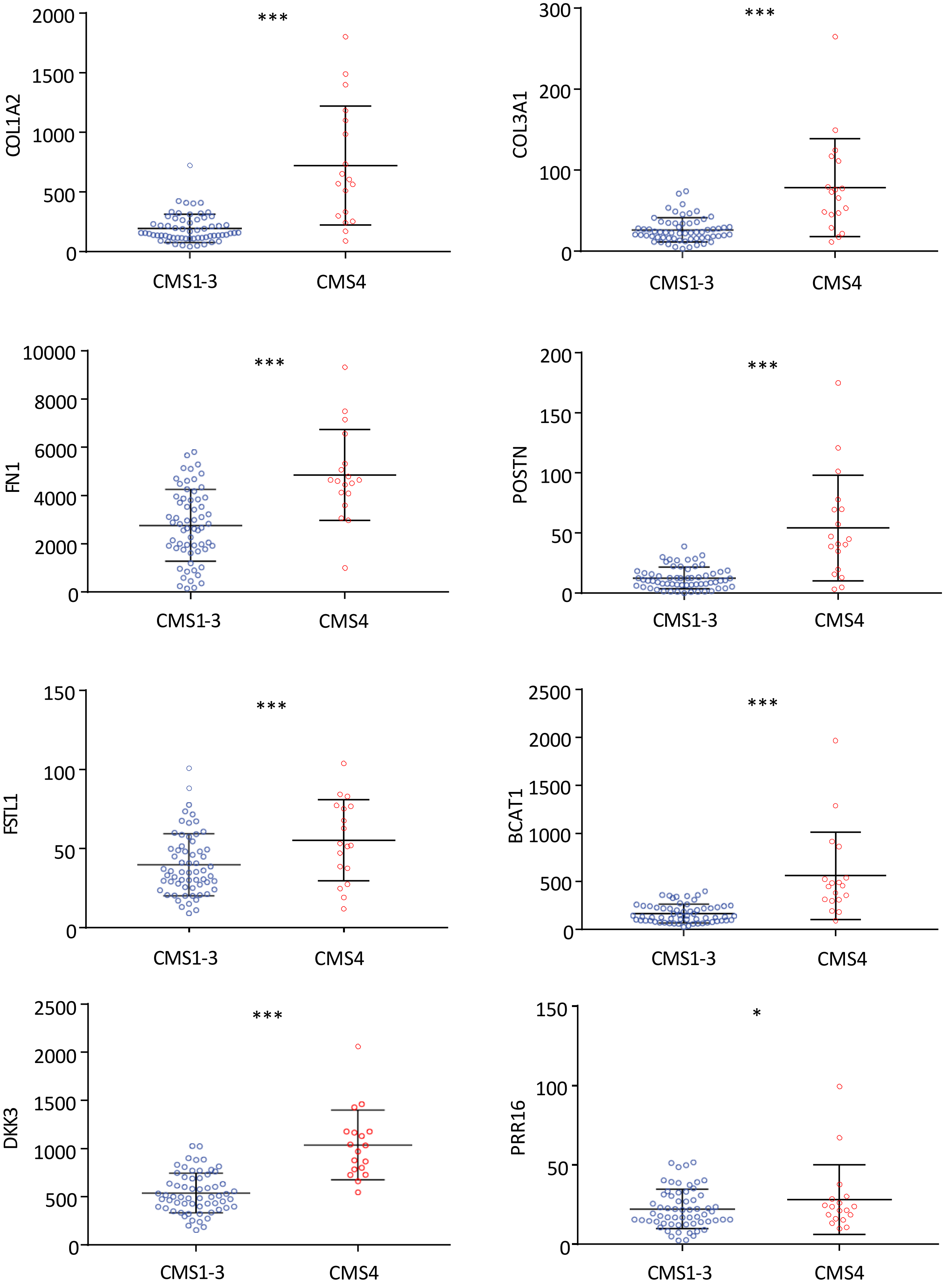
Supplementary Figure S3:

The association between each MATS gene and CMS status in GSE17536 dataset.



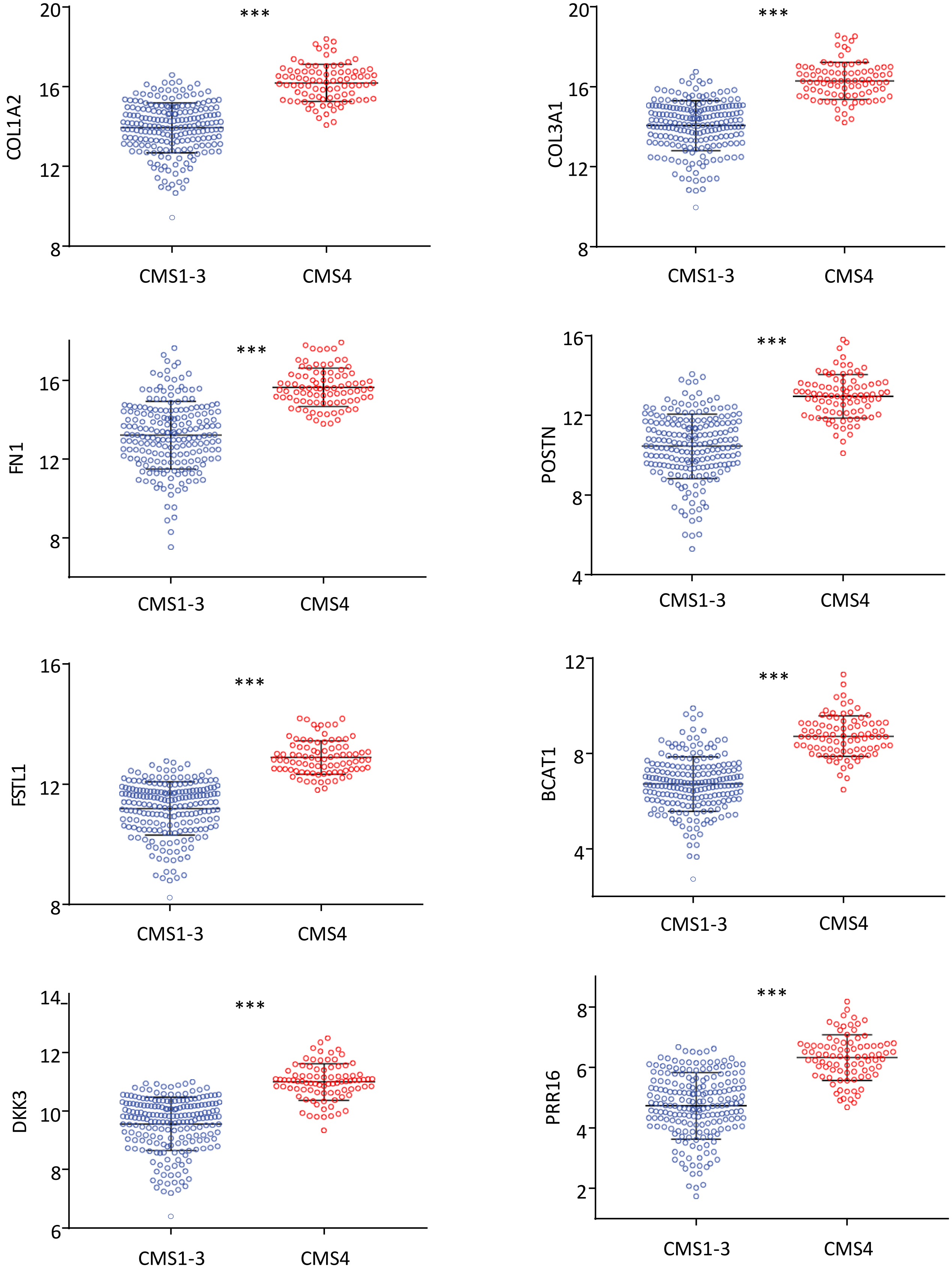
Supplementary Figure S4:

The association between each MATS gene and CMS status in GSE33113 dataset.



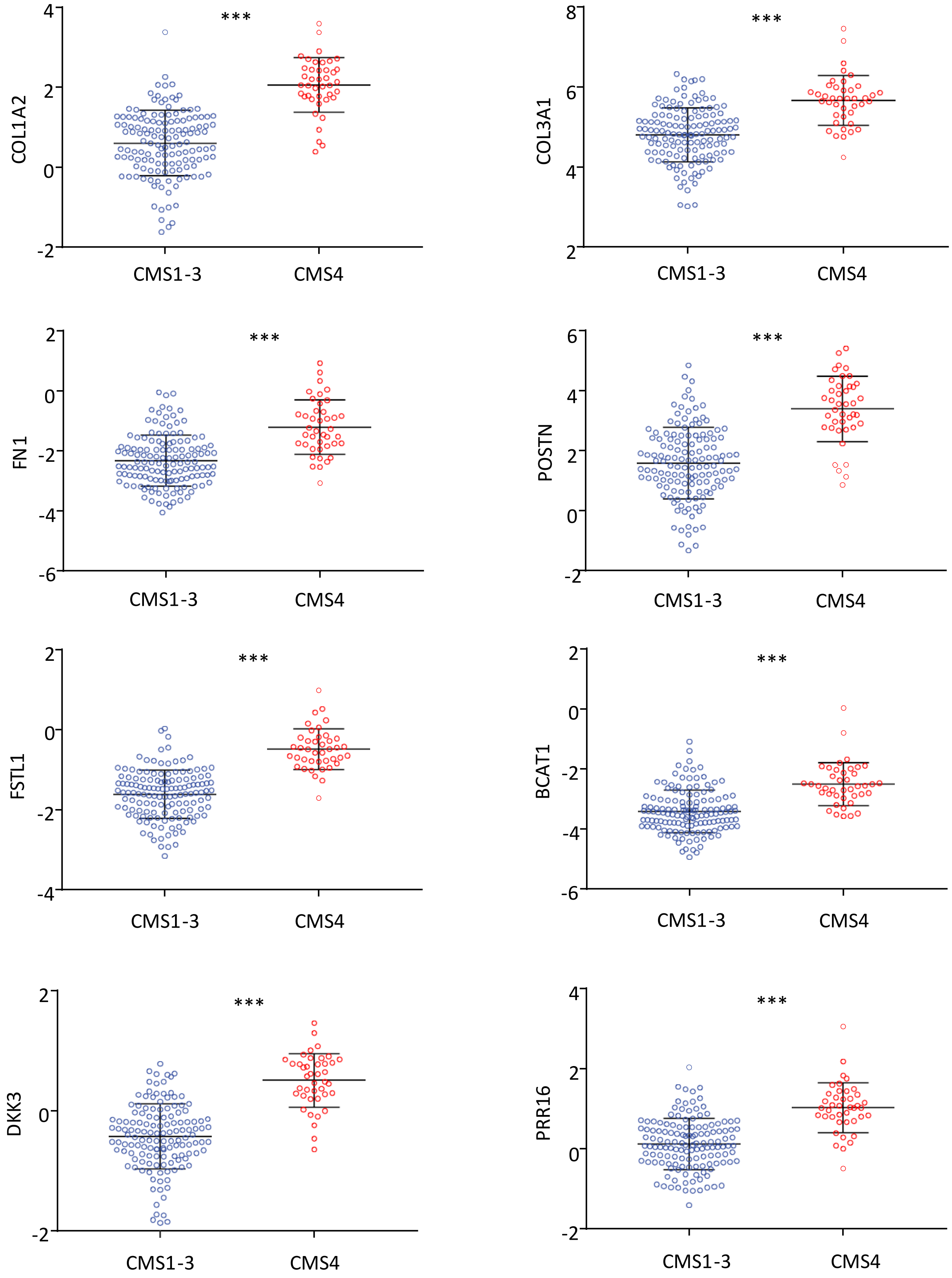
Supplementary Figure S5:

The association between each MATS gene and CMS status in TCGA RNA seq dataset.



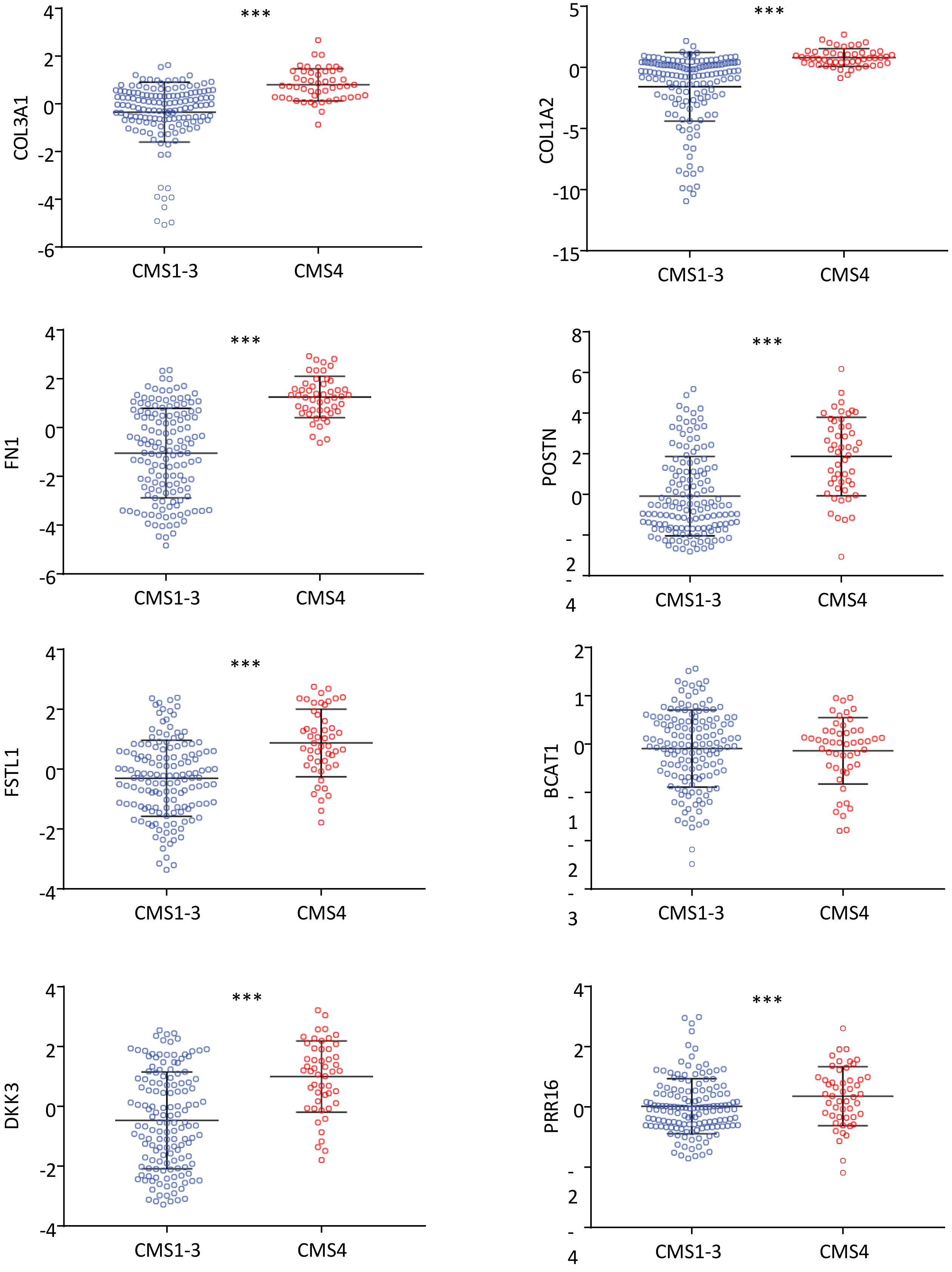
Supplementary Figure S6:

The association between each MATS gene and CMS status in TCGA microarray dataset.



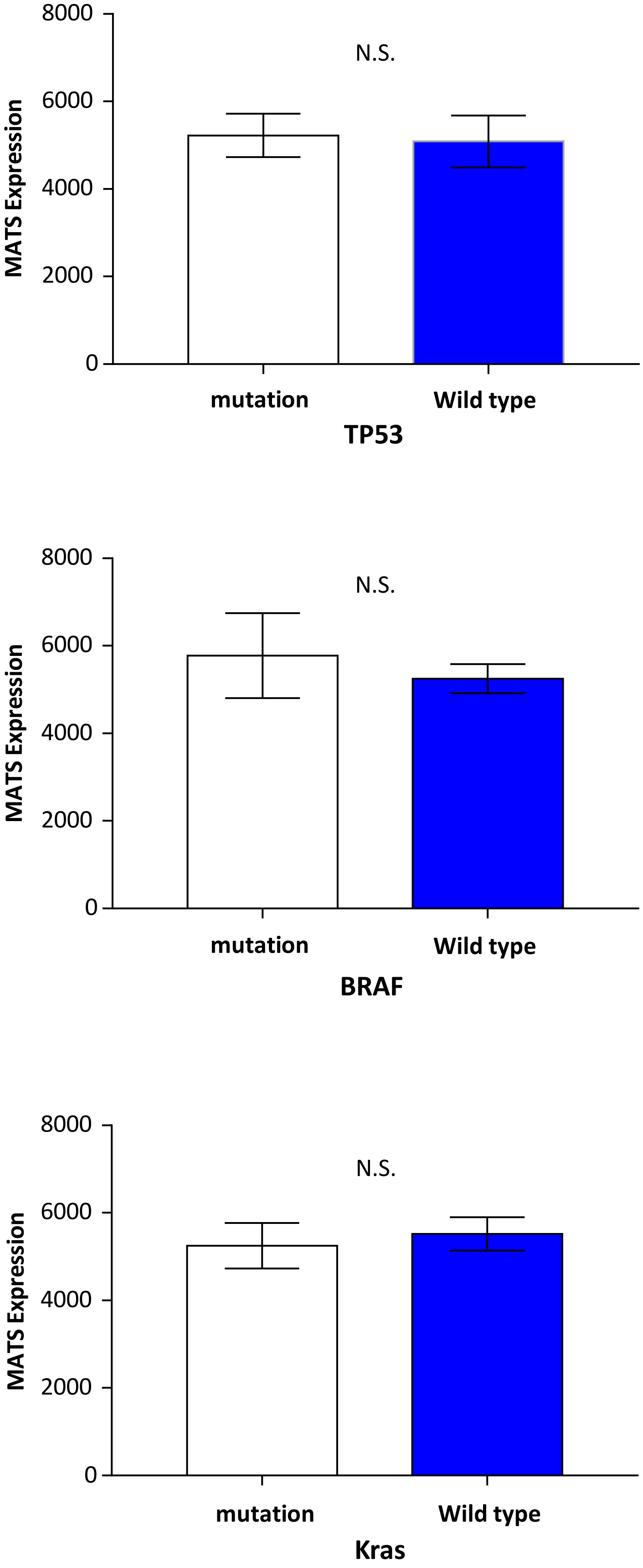
Supplementary Figure S7:

The association between each MATS gene and CMS status in GSE104645 dataset.



Supplementary Figure S8:

Association of MATS with TP53, BRAF and KRAS gene mutations status.



Supplementary Figure S9:

Kaplan-Meier survival analyses revealed MATS high risk group showed poorer RFS than those with low risk in stage II (A), and stage III (B) patients respectively in the validation cohort.