APPENDICES

APPENDIX

Appendix Table 1. Quantification of MUC2 and TFF3 expression in a large panel of CRC cell lines. Cell lines were classified into five different groups based on the quantification of MUC2- and TFF3-positive cells during immunofluorescence analysis. The table describes the average number of counts obtained by each experiment as well as the standard deviation and number of experiments for a given cell line. The comments column marks the cell lines in four distinct categories, with an asterisk (*) when MUC2 and TFF3 staining overlaps, two asterisks (**) when not all MUC2-positive cells stain with TFF3, three asterisks (***) when not all TFF3-positive cells stain with MUC2, and four asterisks (****) when none of TFF3-positive cells stain with MUC2.

Group 1.

Cell line	% MUC2	% TFF3	№ of experiments	Comments
C99	93.34 ± 0.82	100.00	3	***
CL40	93.64 ± 1.47	100.00	3	***
HCA46	100.00	100.00	3	*
HDC9	14.30 ± 0.93	23.84 ± 1.02	6	***
LS411	38.50 ± 1.10	99.39 ± 0.79	3	***
LS513	69.40 ± 1.00	88.90 ± 1.35	3	***
NCIH508	16.09 ± 1.06	16.09 ± 1.06	6	*
NCIH716	98.82 ± 0.73	99.36 ± 0.74	3	***
PCJW	34.33 ± 2.10	34.33 ± 2.10	3	*
RW7213	24.60 ± 1.11	24.60 ± 1.11	3	*

Group 2.

Cell line	% MUC2	% TFF3	№ of experiments	Comments
C125PM	6.02 ± 1.02	6.02 ± 1.02	3	*
C80	6.60 ± 0.07	6.60 ± 0.07	5	*
C84	6.85 ± 0.25	7.60 ± 0.35	3	***

HDC111	7.21 ± 0.36	7.21 ± 0.36	4	*
HDC114	2.48 ± 0.13	1.84 ± 0.11	7	**
HDC143	4.07 ± 0.47	4.07 ± 0.47	3	*
HDC73	3.45 ± 0.36	6.31 ± 0.78	3	***
HT29	1.91 ± 0.07	2.35 ± 0.07	5	***
LS174T	3.28 ± 0.59	3.28 ± 0.59	6	*
LS180	3.42 ± 0.43	3.42 ± 0.43	7	*
T84	4.38 ± 0.01	4.38 ± 0.01	3	*
VACO4A	3.84 ± 0.34	3.84 ± 0.34	3	*
WIDR	3.42 ± 0.20	4.14 ± 0.17	4	***

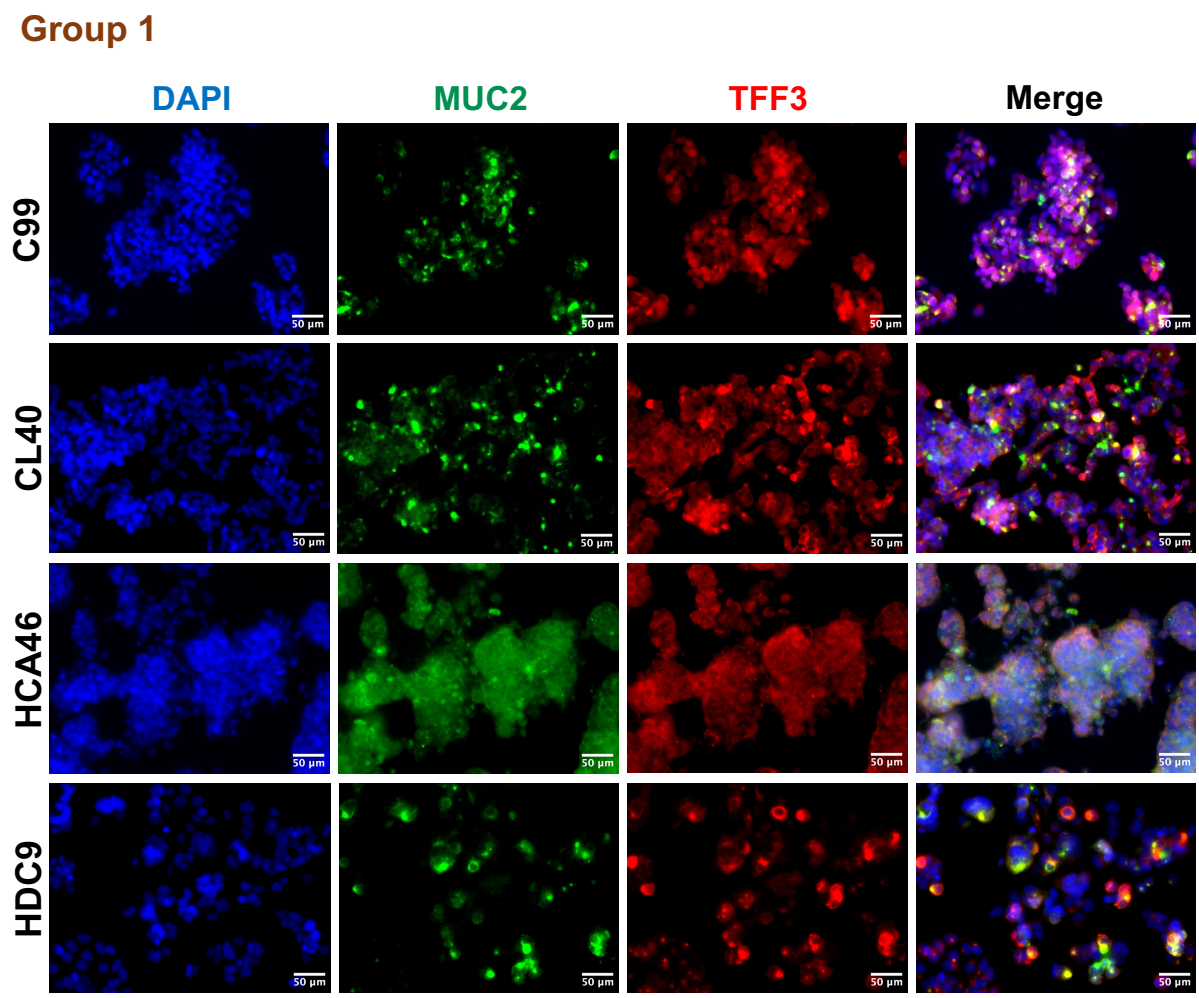
Group 3.

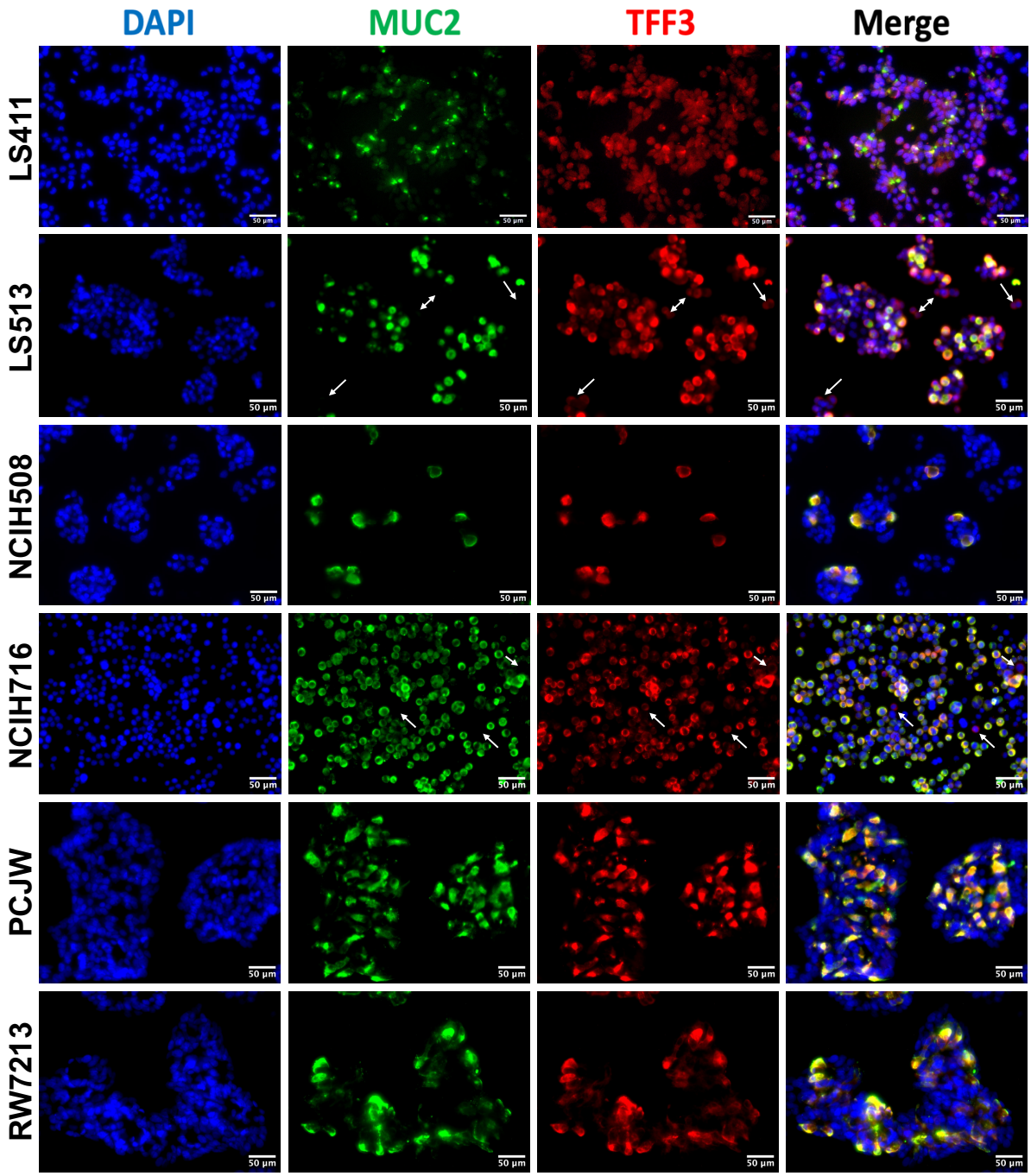
Cell line	% MUC2	% TFF3	№ of experiments	Comments
CCK81	0.76 ± 0.04	0.76 ± 0.04	3	*
CW2	1.25 ± 0.16	1.47 ± 0.15	4	***
CX1	1.01 ± 0.13	1.72 ± 0.18	4	***
HCC56	1.04 ± 0.22	1.04 ± 0.22	3	*
HDC54	0.72 ± 0.06	1.16 ± 0.15	5	***
HDC57	1.19 ± 0.23	1.28 ± 0.30	7	***
JHCOLOYI	0.49 ± 0.11	1.48 ± 0.24	4	***
KM20L2	0.57 ± 0.09	1.17 ± 0.19	3	***
LOVO	0.83 ± 0.10	0.29 ± 0.05	3	**
LS1034	0.89 ± 0.20	1.07 ± 0.23	3	***
SW1222	0.66 ± 0.08	0.96 ± 0.31	4	***
SW1463	0.46 ± 0.06	0.69 ± 0.16	3	***
SW948	1.41 ± 0.11	1.41 ± 0.11	3	*
VACO4S	1.34 ± 0.05	1.34 ± 0.05	3	*

Group 4.

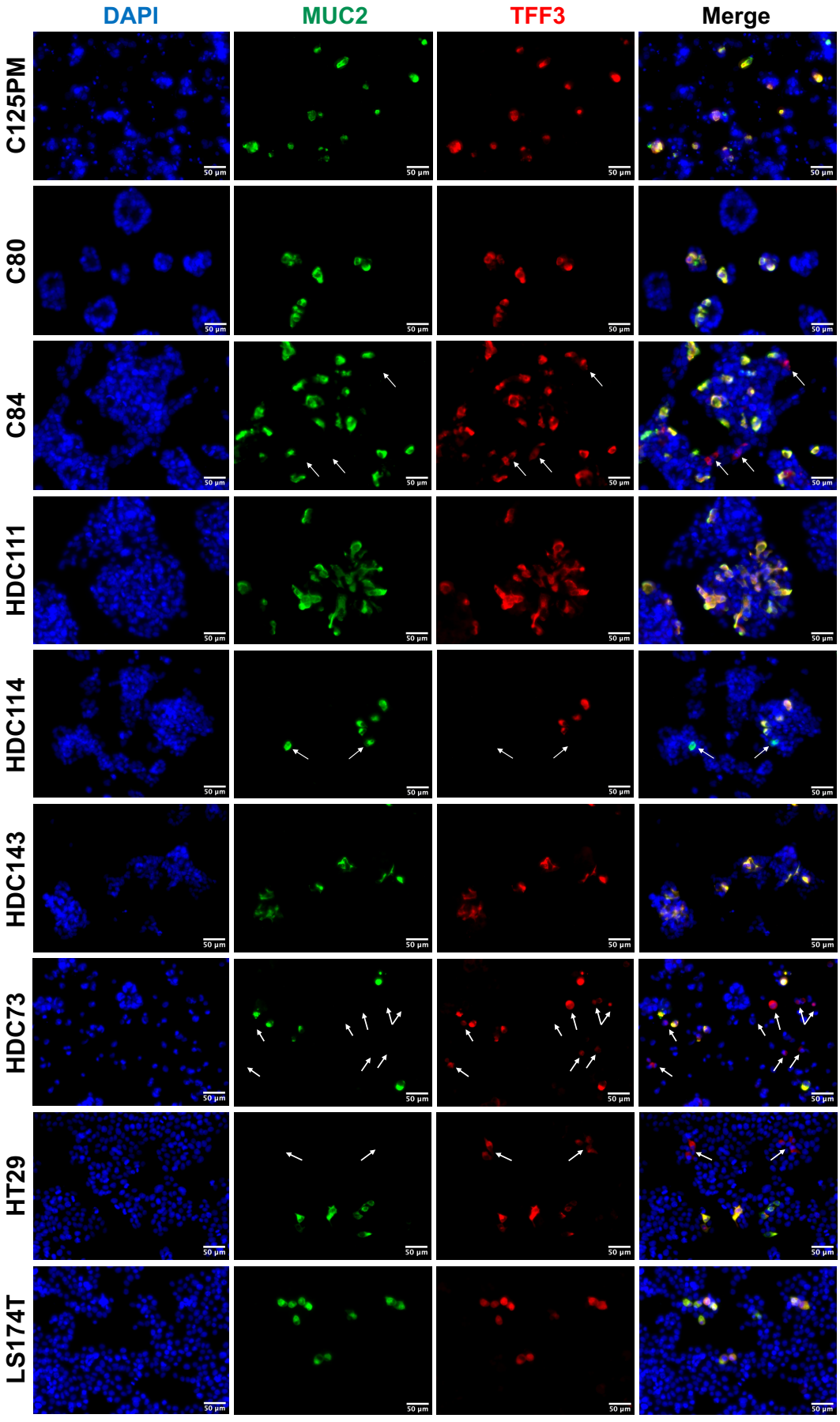
Cell line	% MUC2	% TFF3	№ of experiments	Comments
C106	0	100.00	3	****
C70	0	100.00	3	****
CL14	0	100.00	3	****
COLO201	0	100.00	3	****
COLO205	0	100.00	3	****
COLO206	0	100.00	3	****
GP2D	0	100.00	3	****
GP5D	0	100.00	3	****
HDC142	0	100.00	3	****
NCIH548	0	100.00	3	****
PMFKO14	0	97.59 ± 0.56	3	****
RCM1	0	100.00	3	****
RW2982	0	100.00	3	****
SKCO1	0	100.00	3	****
SNUC1	0	100.00	3	****
SW403	0	100.00	3	****
SW620	0	100.00	3	****
TT1TKB	0	97.87 ± 1.40	3	****
VACO429	0	100.00	3	****
HCT8	0.79 ± 0.12	100.00	3	***
HT55	8.74 ± 1.48	100.00	3	***
LIM1863	0.64 ± 0.09	100.00	4	***
OXCO2	0.36 ± 0.05	100.00	3	***
SNU1235	2.92 ± 0.83	99.03 ± 0.93	4	***
VACO10MS	9.2 ± 1.40	99.00 ± 0.89	3	***
VACO5	2.27 ± 0.30	100.00	4	***
SNUC2B	0	18.73 ± 0.91	3	****

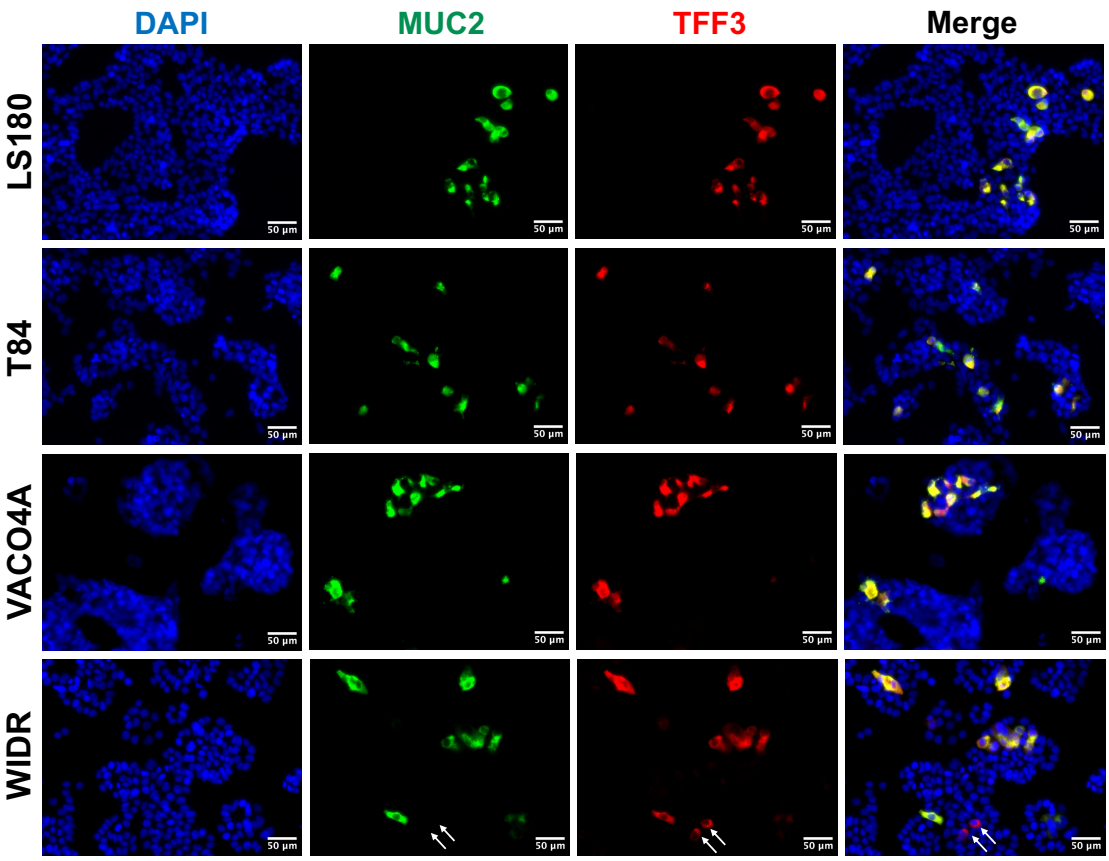
Appendix Figure 1. Representative images of 100 CRC cell lines based on MUC2 and TFF3 staining. The cells were cultured for three days before being fixed and stained with MUC2 and TFF3 antibodies. Magnification: 20X, scale bar: 50 µm.



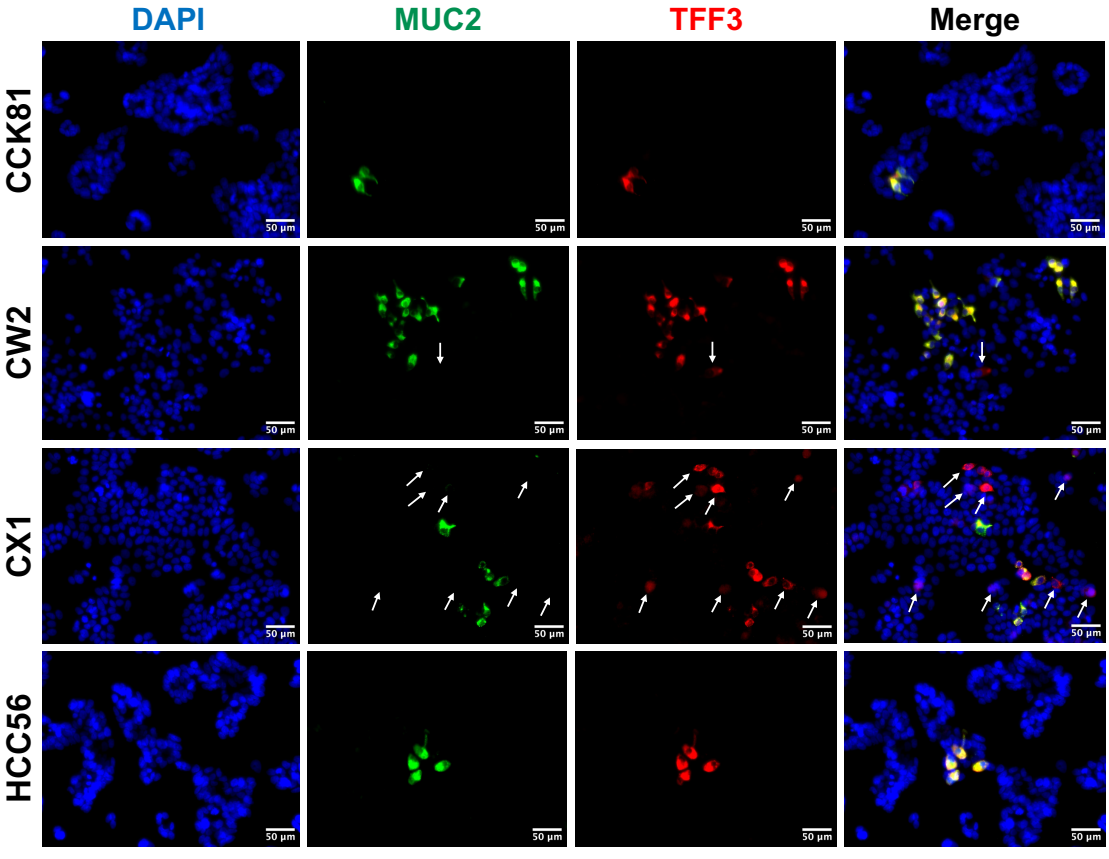

DAPI
MUC2
TFF3
Merge
LS411
LS513
NCIH508
NCIH716
PCJW
RW7213
50 µm

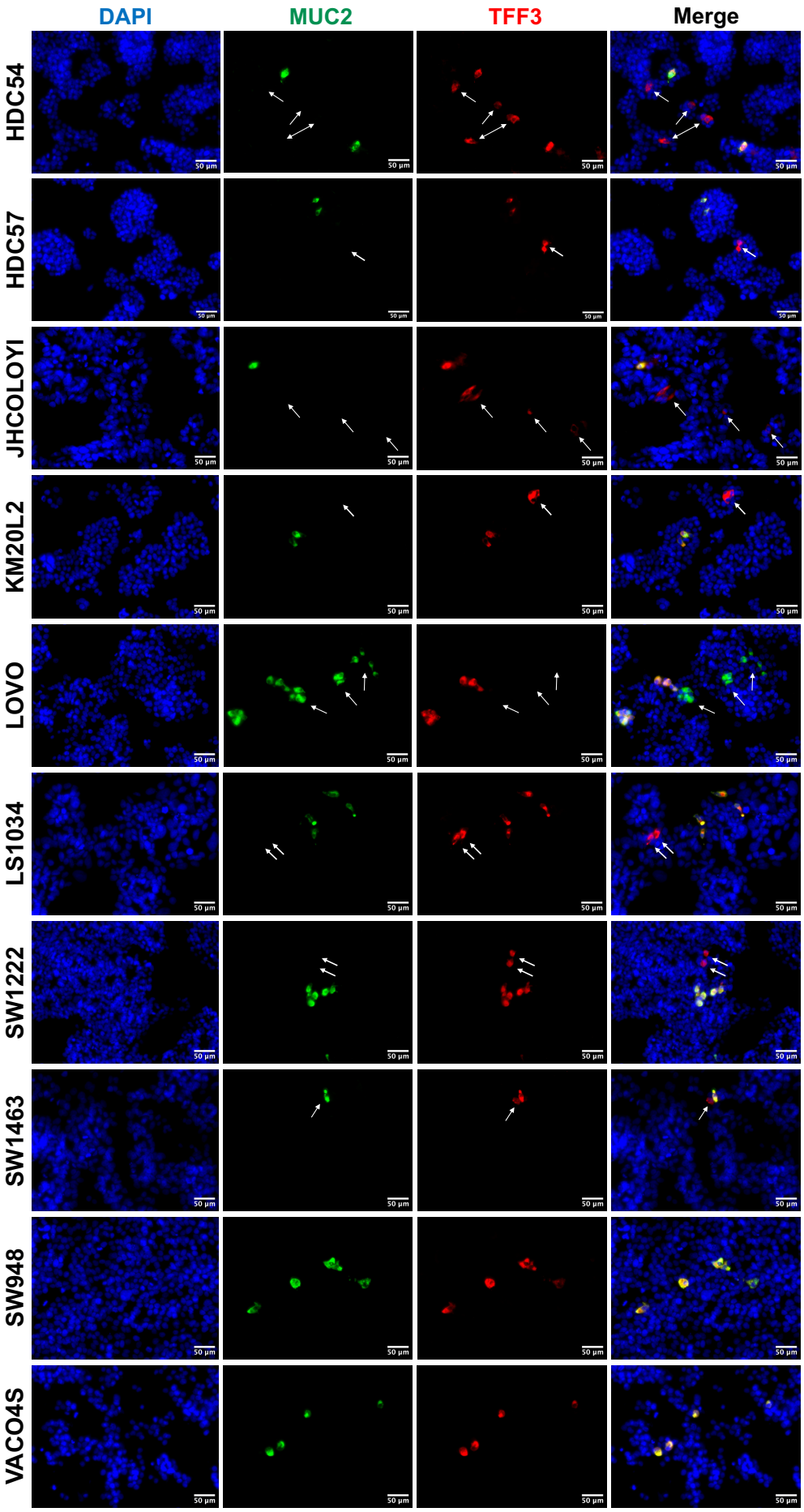
Group 2


DAPI
MUC2
TFF3
Merge
C125PM
C80
C84
HDC111
HDC114
HDC143
HDC73
HT29
LS174T
50 µm

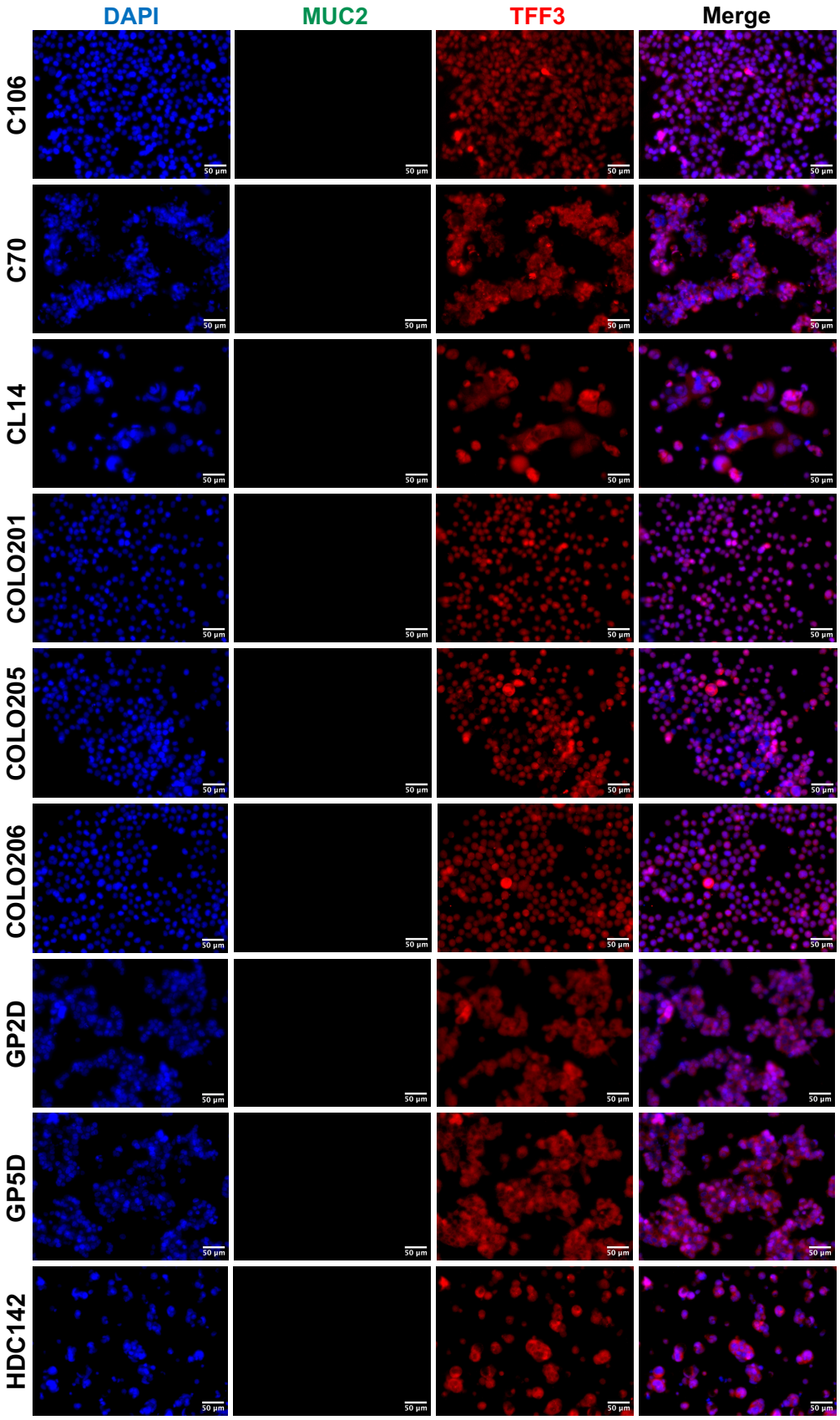

DAPI
MUC2
TFF3
Merge
LS180
T84
VACO4A
WIDR
50 µm

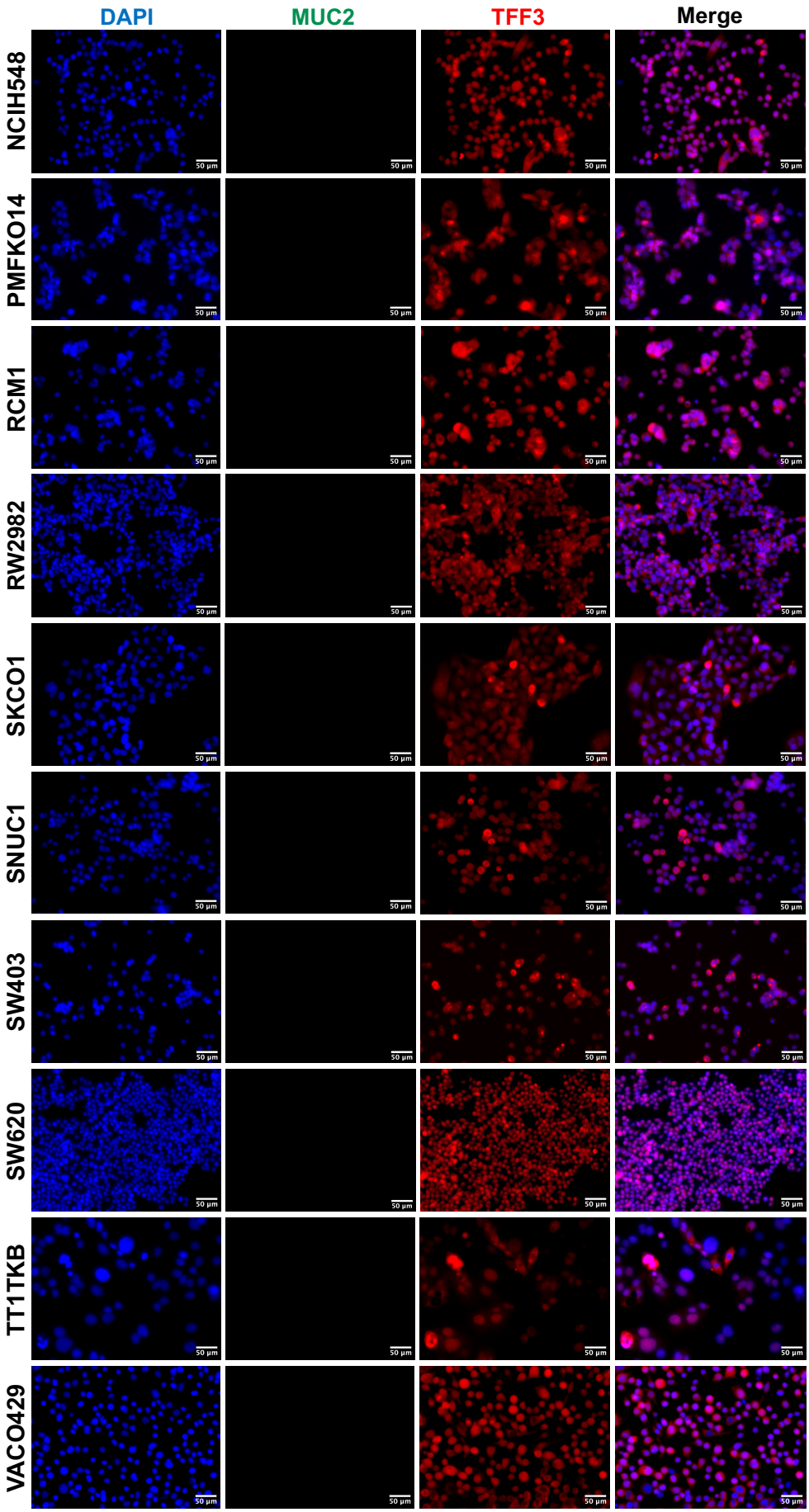
Group 3


DAPI
MUC2
TFF3
Merge
CCK81
CW2
CX1
HCC56
50 µm

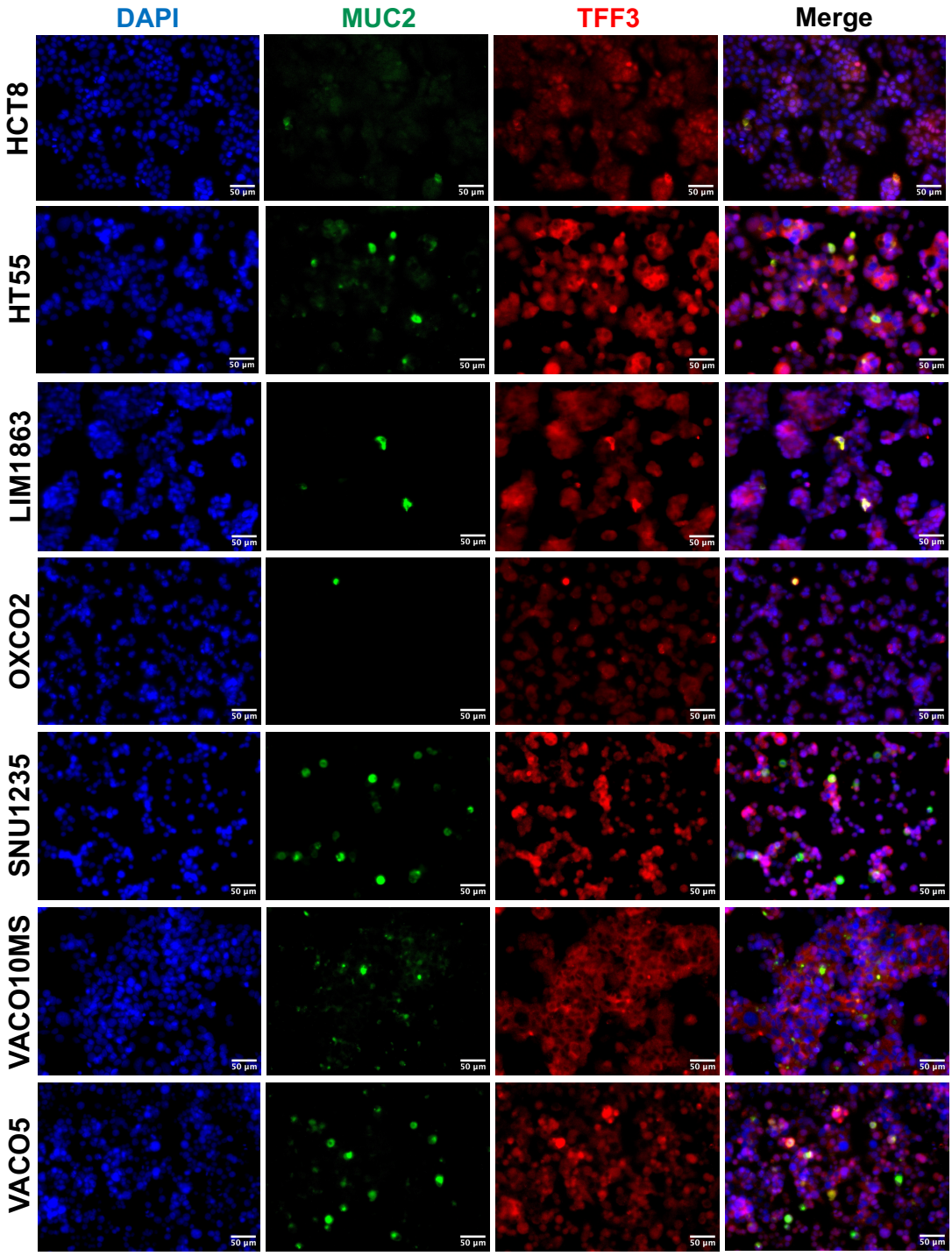

DAPI
MUC2
TFF3
Merge
HDC54
HDC57
JHCOLOYI
KM20L2
LOVO
LS1034
SW1222
SW1463
SW948
VACO4S
50 μm

Group 4

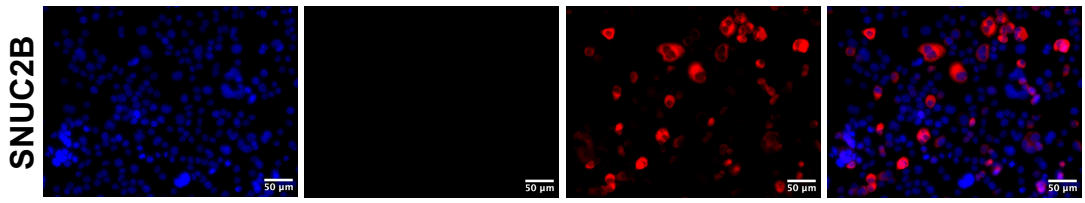

DAPI
MUC2
TFF3
Merge
C106
C70
CL14
COLO201
COLO205
COLO206
GP2D
GP5D
HDC142
50 µm


DAPI
MUC2
TFF3
Merge
NCIH548
PMFKO14
RCM1
RW2982
SKCO1
SNUC1
SW403
SW620
TT1TKB
VACO429
50 µm

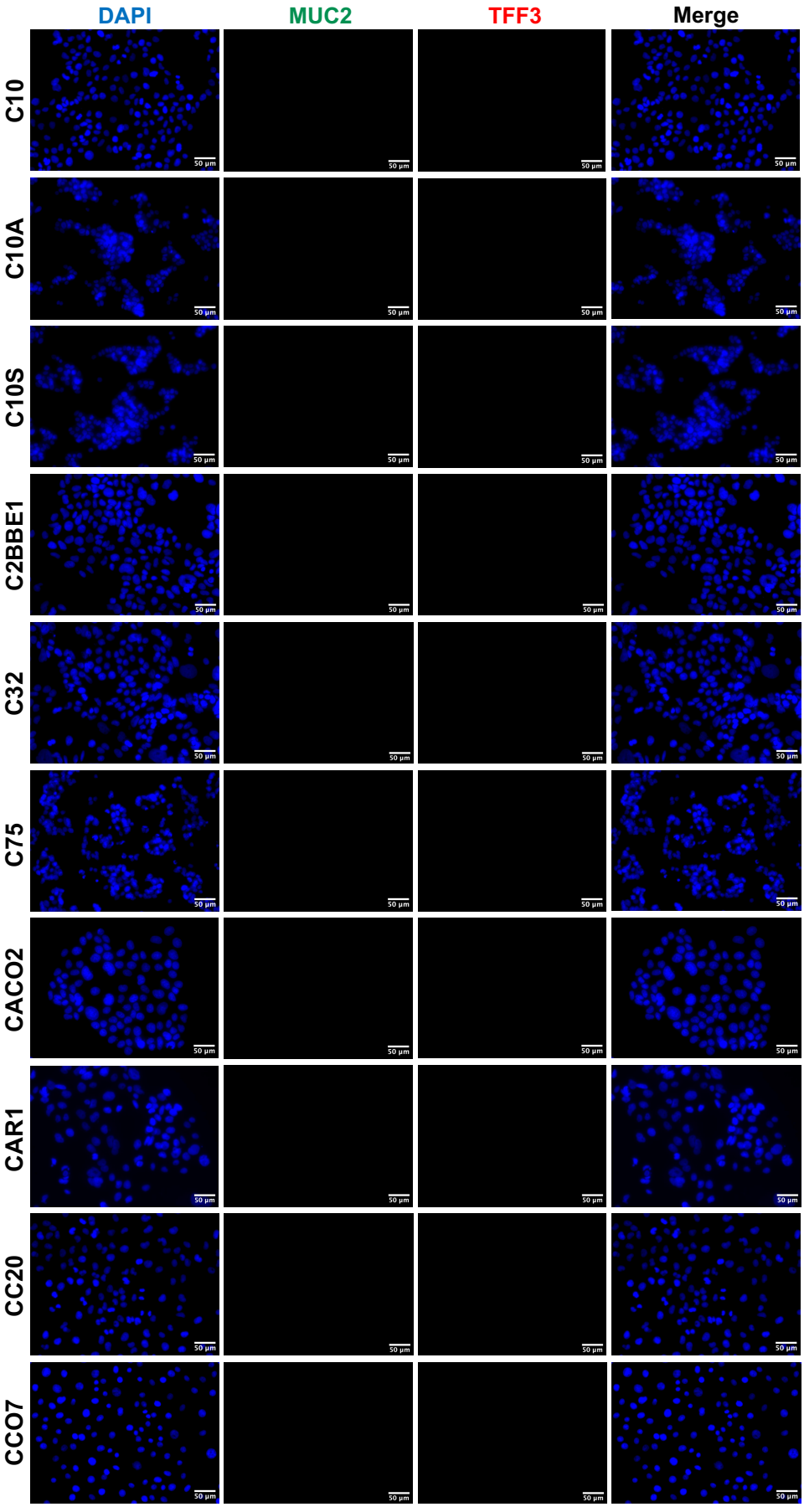
Group 4*

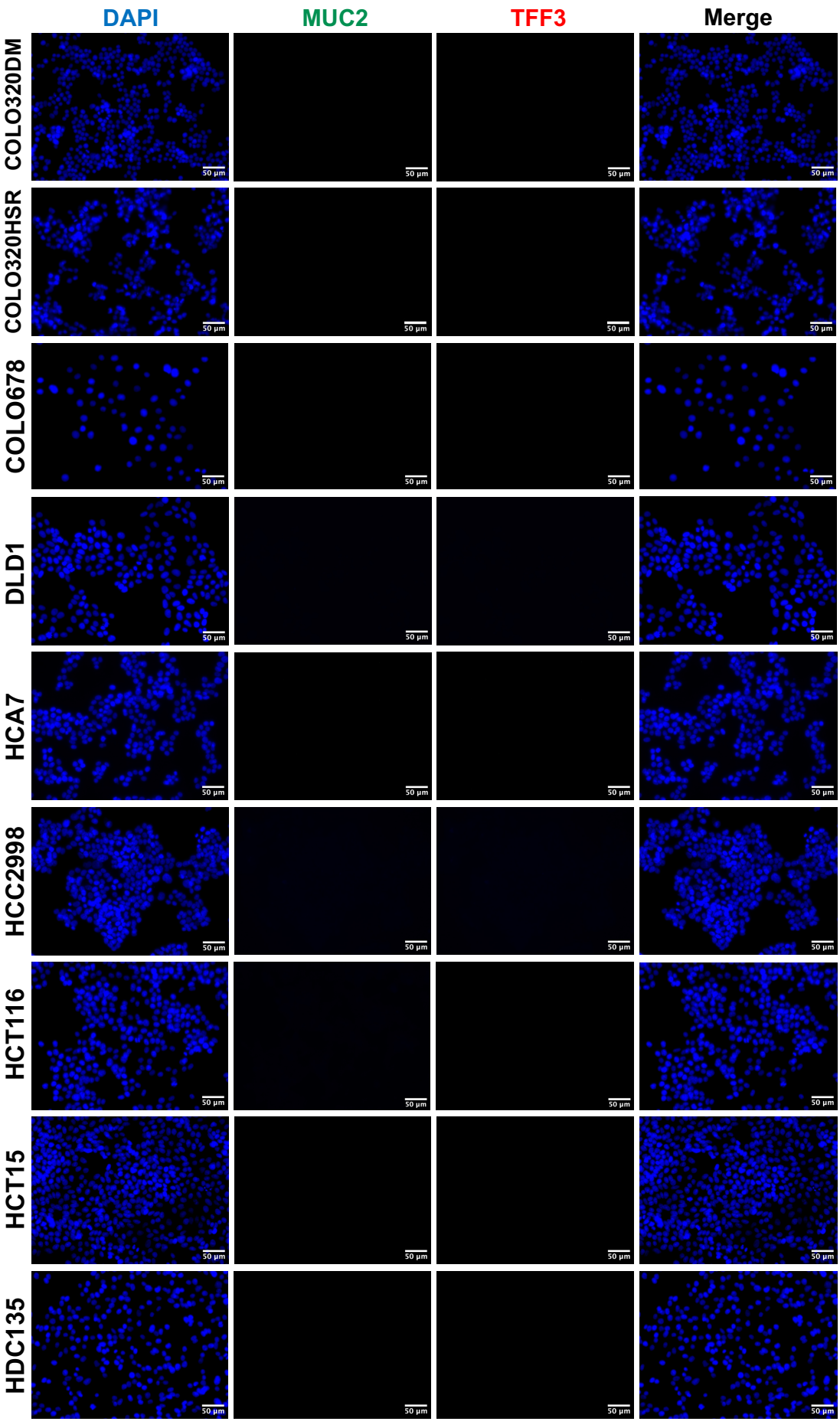


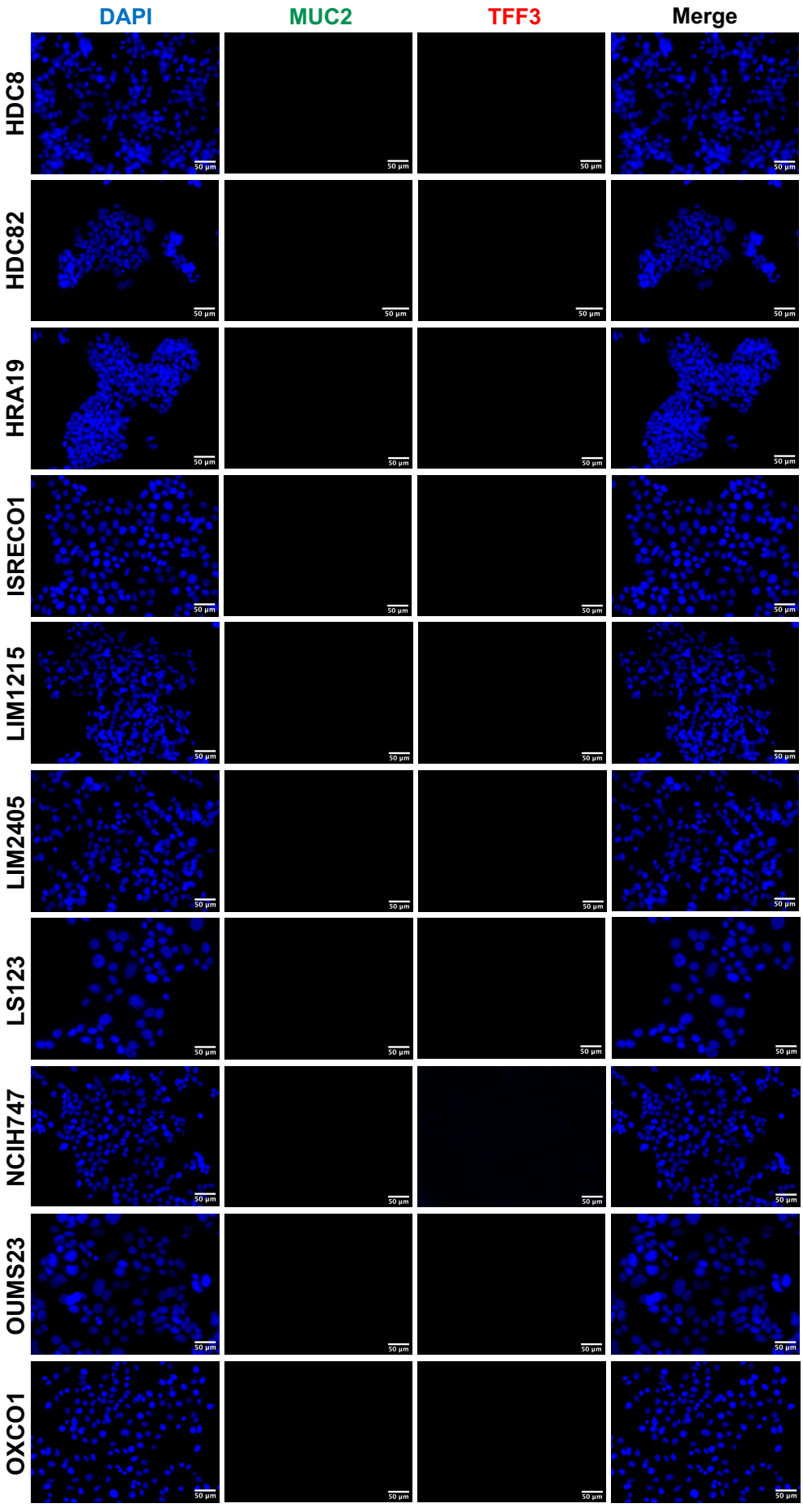
Group 4**

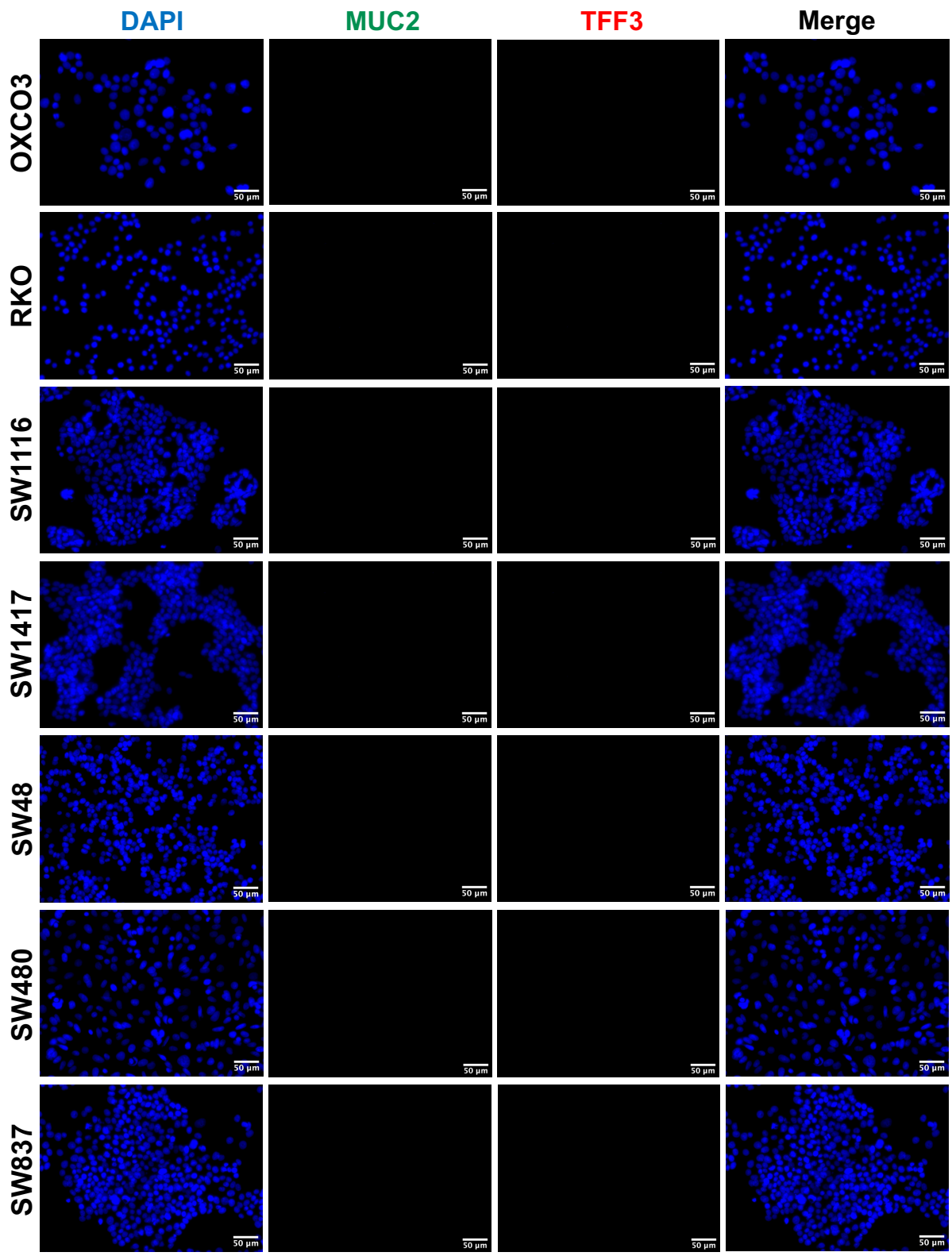


Group 5

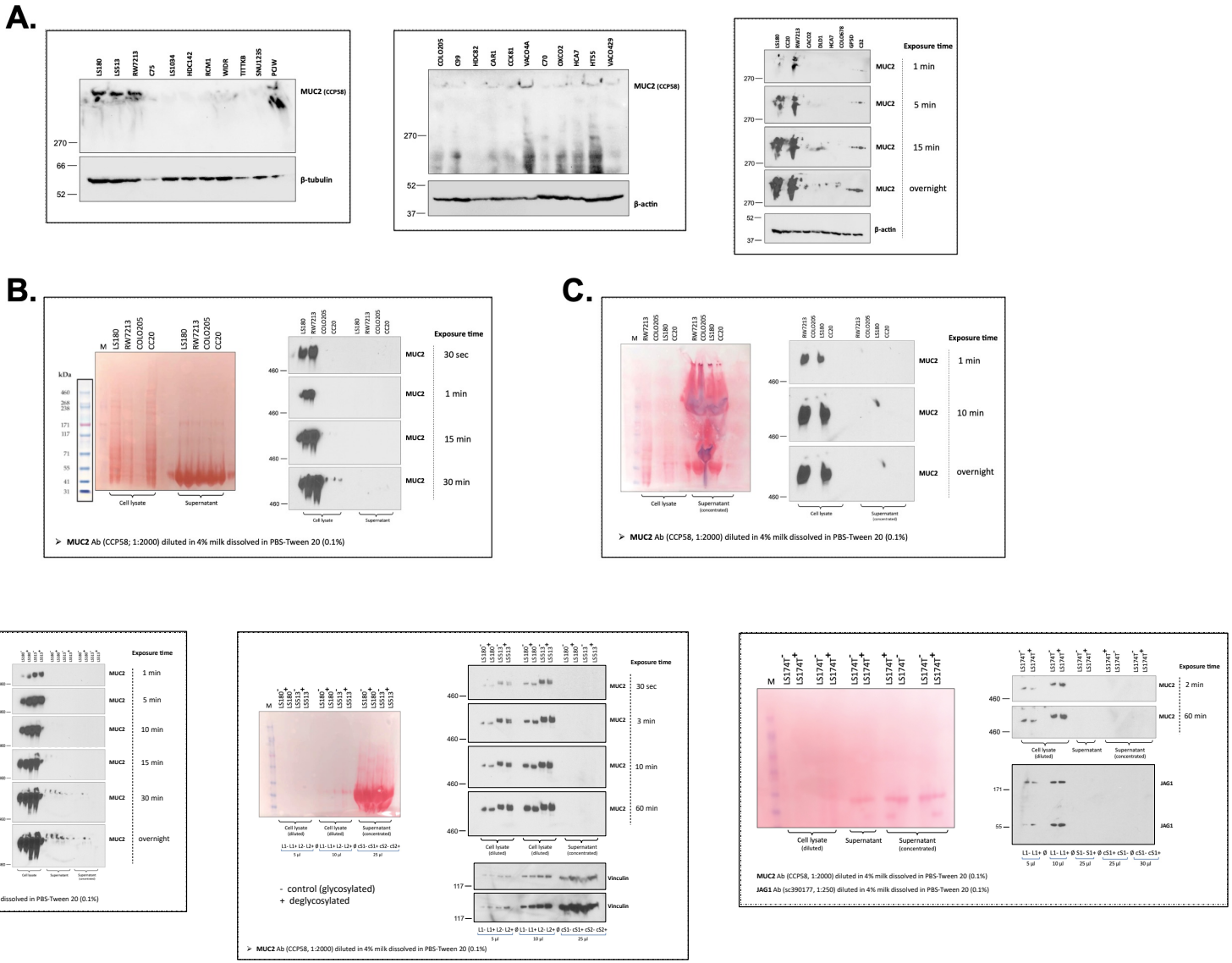

DAPI
MUC2
TFF3
Merge
C10
C10A
C10S
C2BBE1
C32
C75
CACO2
CAR1
CC20
CCO7
50 µm


DAPI
MUC2
TFF3
Merge
COLO320DM
COLO320HSR
COLO678
DLD1
HCA7
HCC2998
HCT116
HCT15
HDC135
50 µm

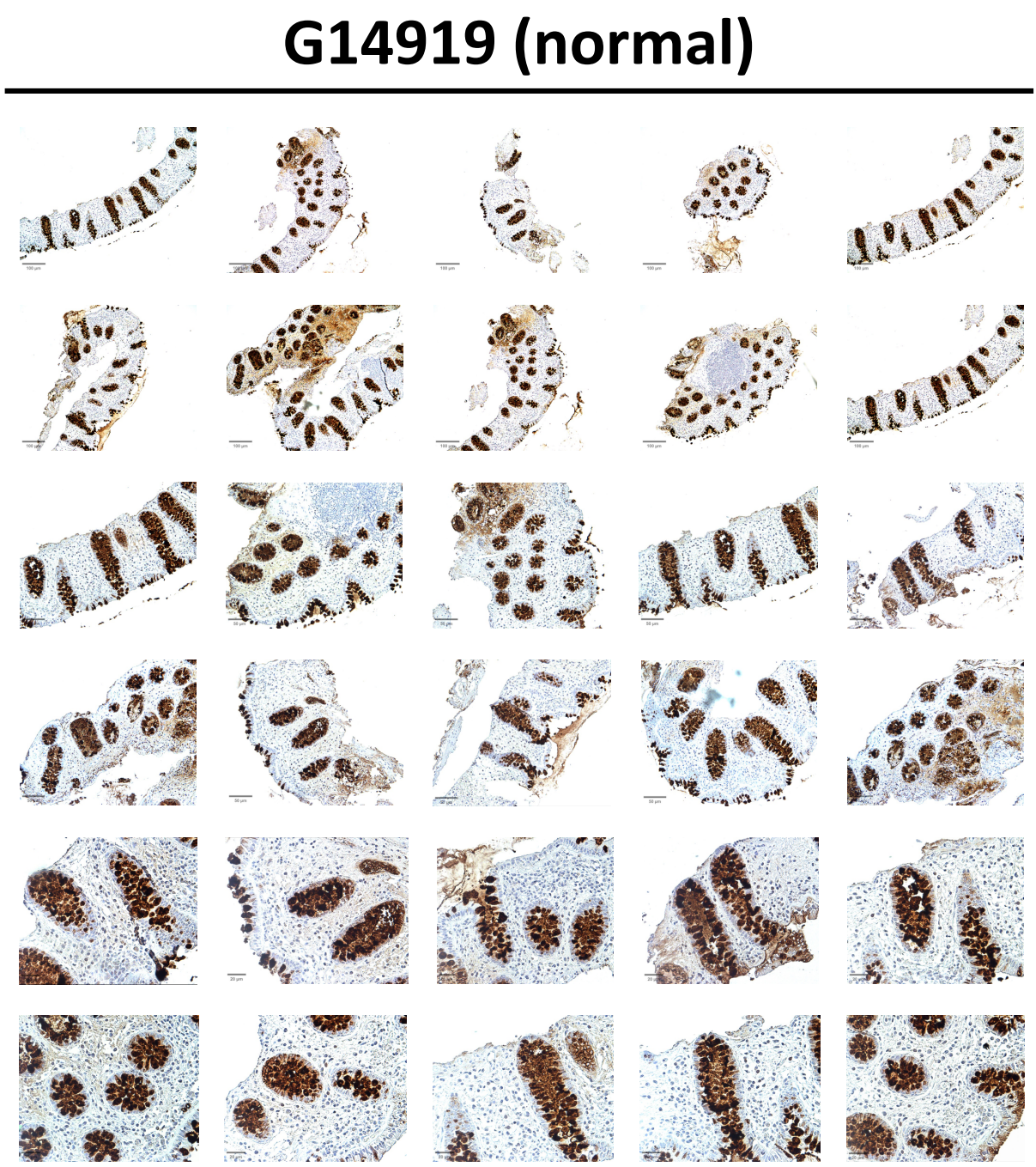

DAPI
MUC2
TFF3
Merge
HDC8
HDC82
HRA19
ISRECO1
LIM1215
LIM2405
LS123
NCIH747
OUMS23
OXCO1
50 µm


DAPI
MUC2
TFF3
Merge
OXCO3
RKO
SW1116
SW1417
SW48
SW480
SW837
50 µm

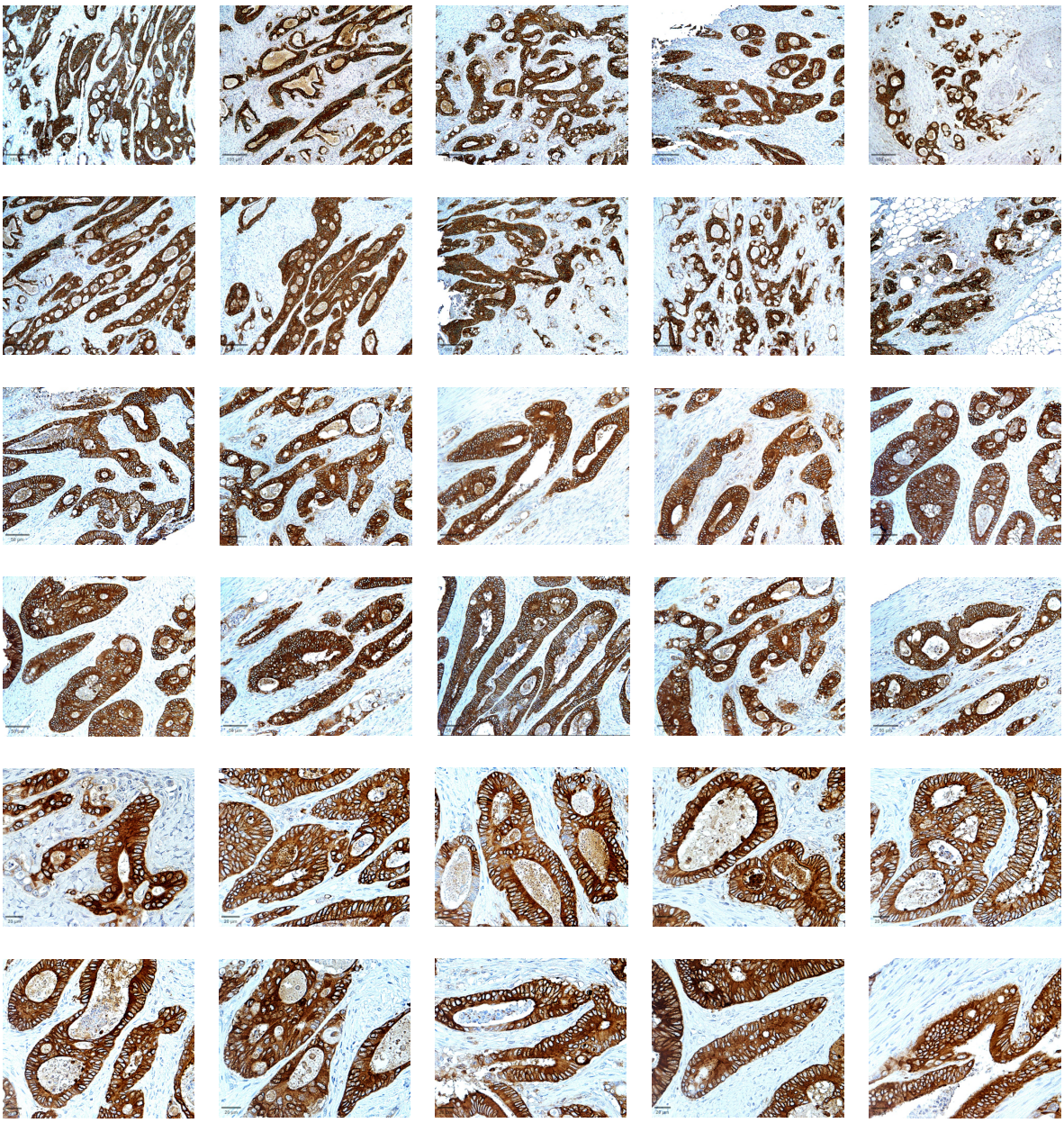
Appendix Figure 2 (following page). Investigation of MUC2 secretion using CCP58 commercial mouse monoclonal antibody. (A) illustrates intracellular MUC2 protein expression across various CRC cell lines. (B) depicts the production of intracellular and extracellular MUC2 in MUC2-positive cell lines, including LS180 and RW7213, as well as in MUC2-negative cell lines, namely COLO205 and CC20. (C) displays the same set of cell lines as shown in (B), but utilising the concentrated supernatant. (D) represents the glycosylated and deglycosylated conditions of lysates and supernatants of LS180, LS174T, and LS513 cell lines.


A.
LS180
LS513
RW7213
C75
LS1034
HDC142
RCM1
WIDR
TTTK8
SNU1235
PCIW
MUC2 (CCP58)
270
66
52
β-tubulin
COLO205
C99
HDC82
CAR1
CCK81
VACO4A
C70
OKCO2
HCA7
HT55
VACO429
MUC2 (CCP58)
270
52
37
β-actin
LS180
CC20
RW7213
CAC02
DLD1
HCA7
COLO678
GP5D
C82
Exposure time
MUC2
1 min
5 min
15 min
overnight
β-actin
B.
M
LS180
RW7213
COLO205
CC20
kDa
460
268
238
171
117
71
55
41
31
Cell lysate
Supernatant
Exposure time
30 sec
1 min
15 min
30 min
MUC2 Ab (CCP58; 1:2000) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)
C.
Cell lysate
Supernatant (concentrated)
Exposure time
1 min
10 min
overnight
MUC2 Ab (CCP58, 1:2000) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)
D.
- control (glycosylated)
+ deglycosylated
Exposure time
1 min
5 min
10 min
15 min
30 min
overnight
MUC2 Ab (CCP58, 1:2000) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)
Cell lysate (diluted)
Supernatant (concentrated)
30 sec
3 min
10 min
60 min
Vinculin
MUC2 Ab (CCP58, 1:2000) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)
LS174T
2 min
60 min
JAG1
MUC2 Ab (CCP58, 1:2000) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)
JAG1 Ab (sc390177, 1:250) diluted in 4% milk dissolved in PBS-Tween 20 (0.1%)

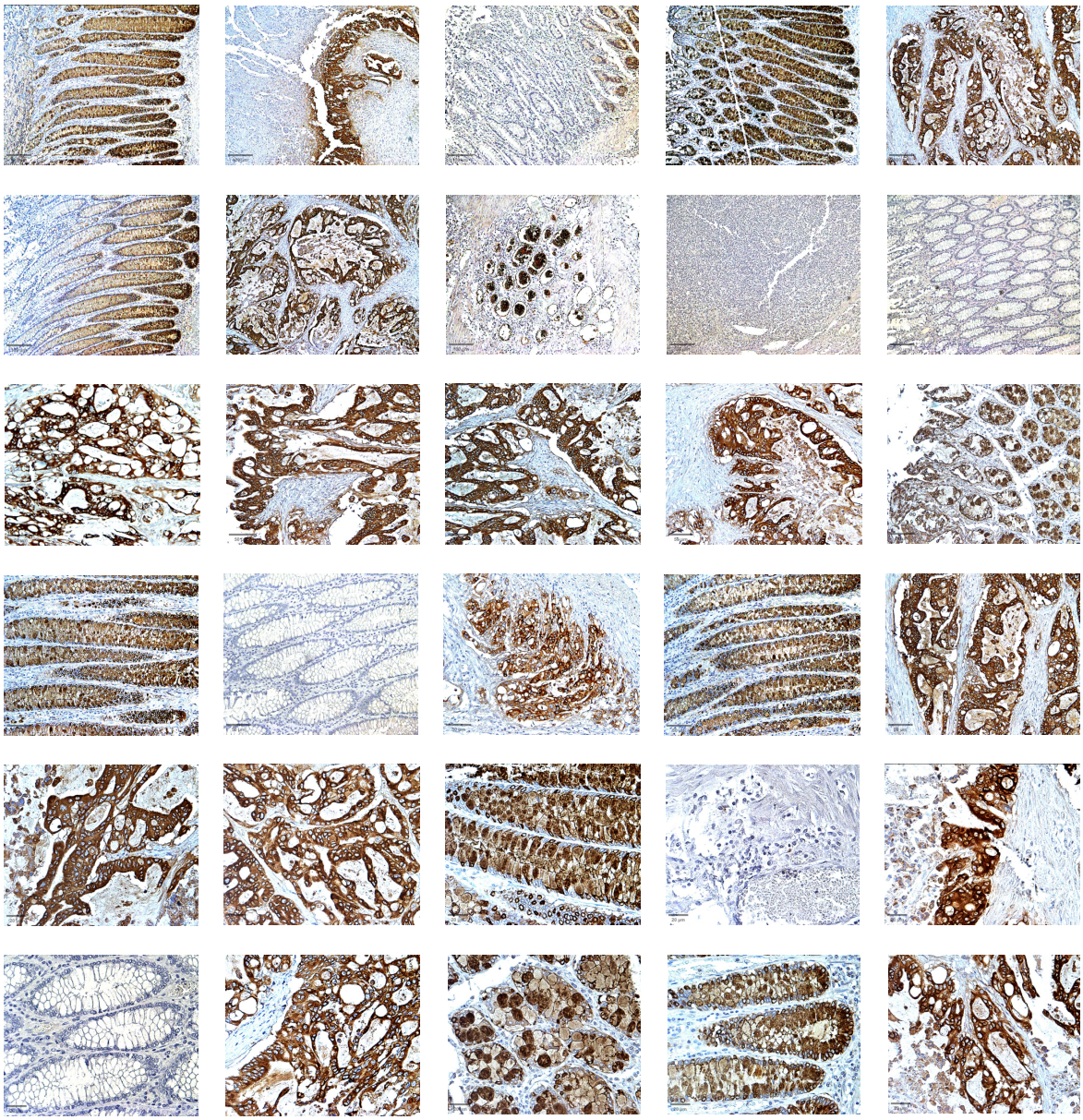
Appendix Figure 3. Representative images of TFF3 protein expression using a monoclonal TFF3 antibody in CRC cell lines, as well as human FFPE tissue sections. Immunofluorescence and immunohistochemistry assays were conducted. Scale bars: 20, 50, and 100 μm.



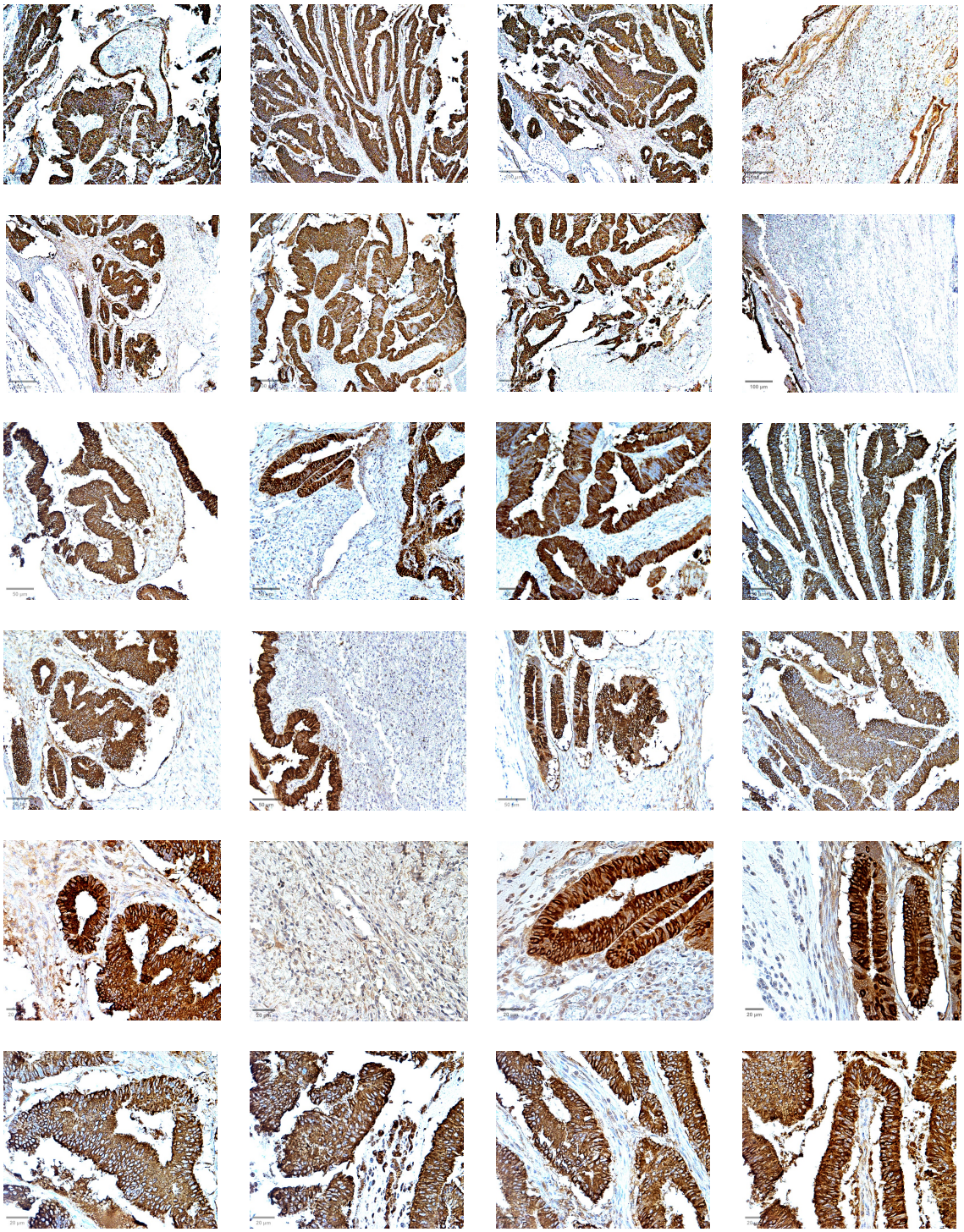
VT11189 (colorectal cancer)



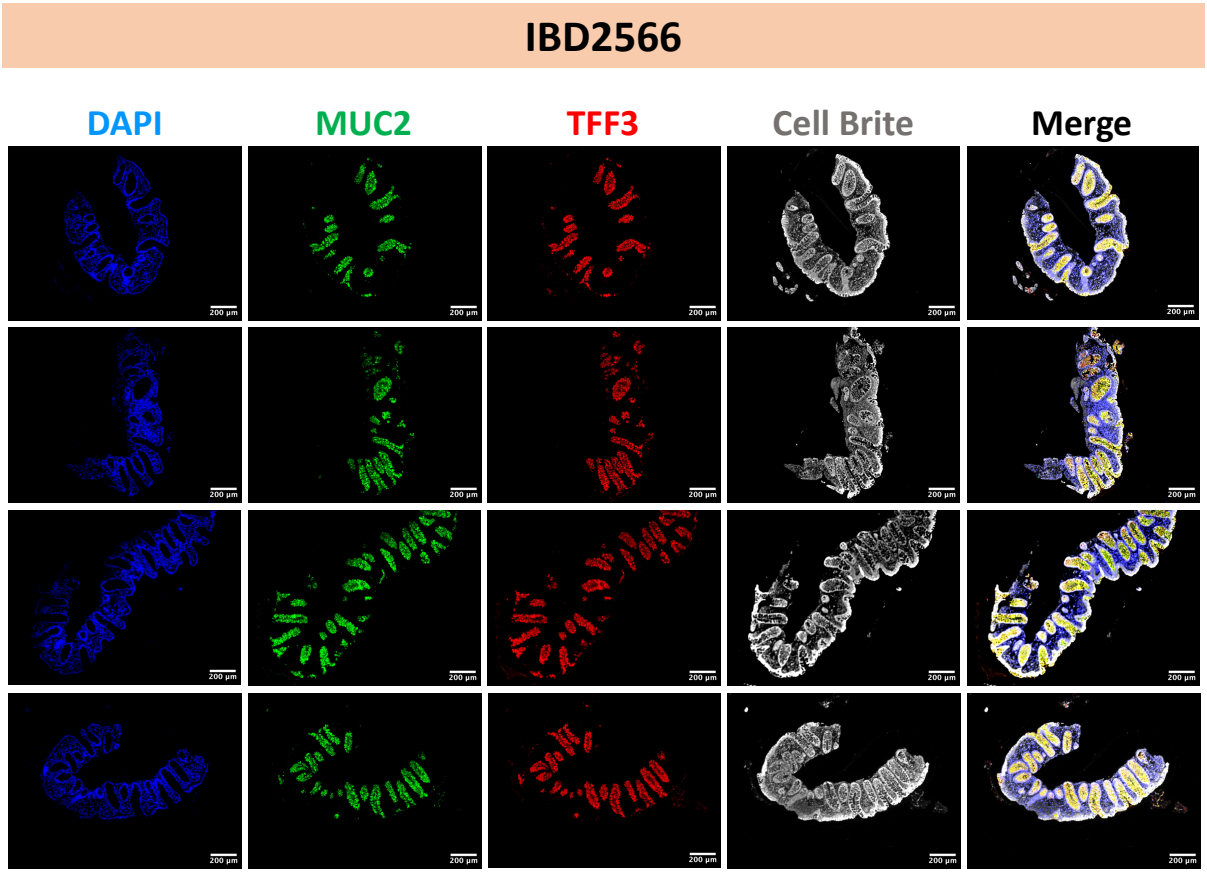
VT11385 (colorectal cancer)

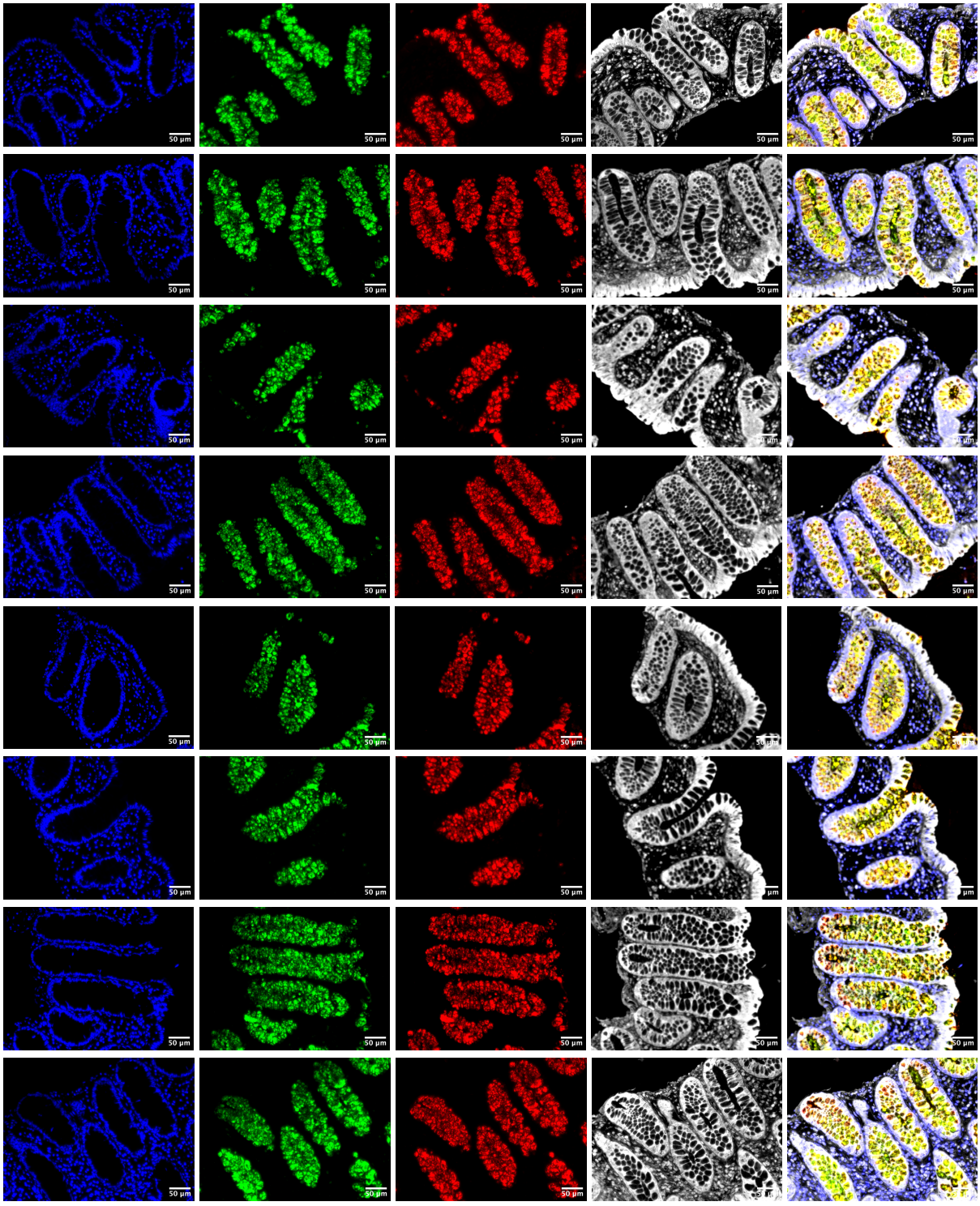


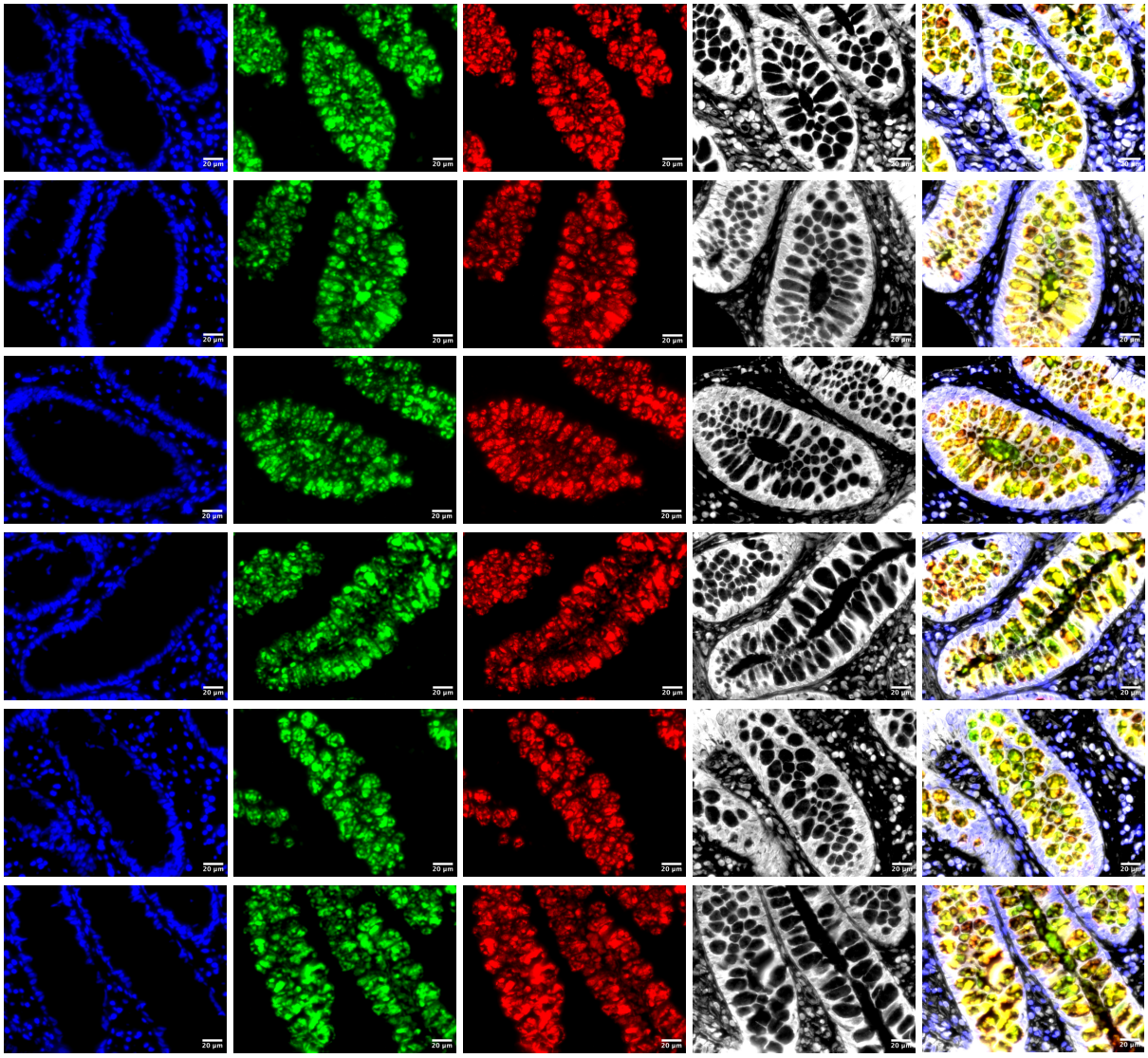
VT11401 (colorectal cancer)

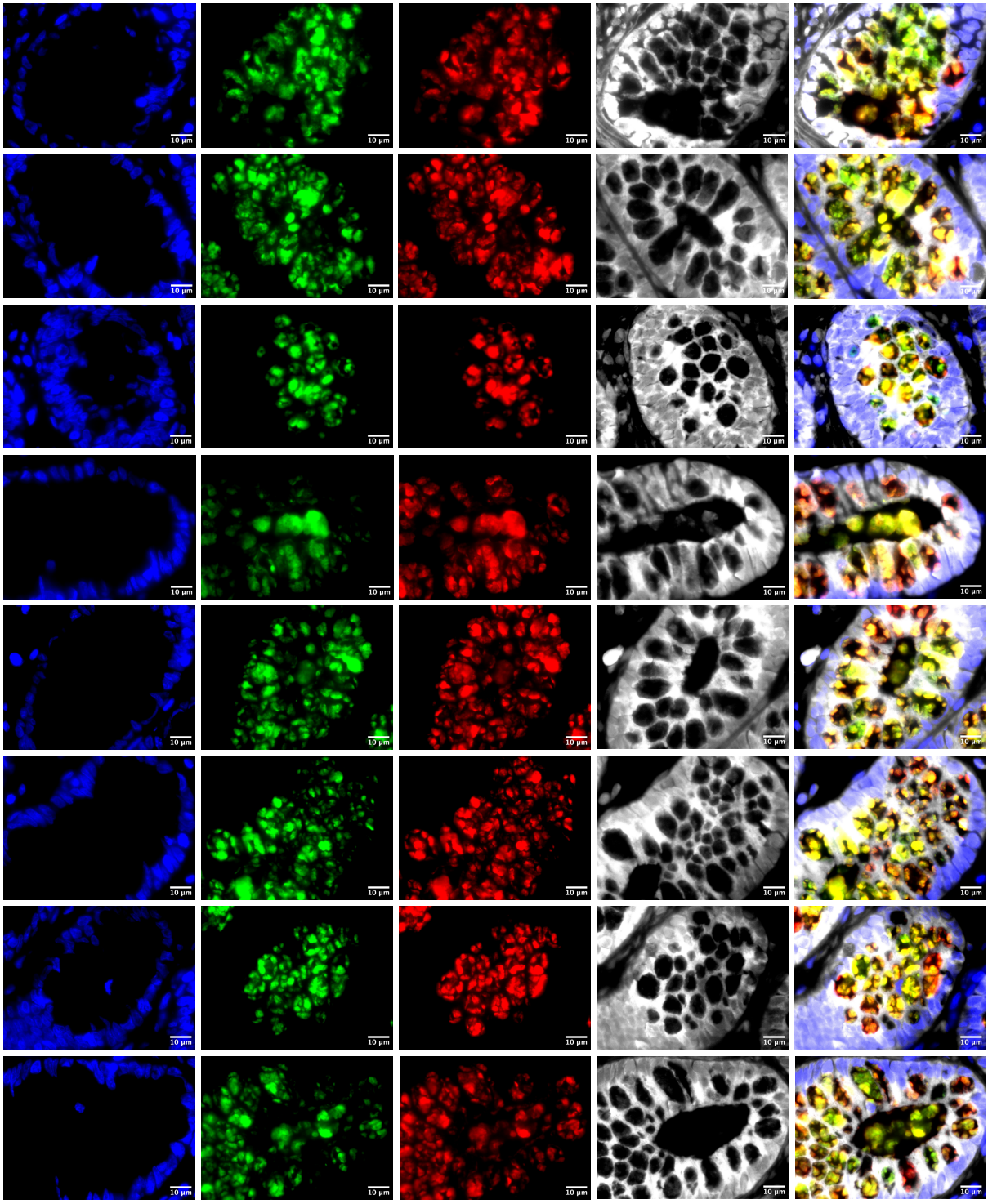


Appendix Figure 4. Staining of MUC2, TFF3, and CellBrite in human tissue section, IBD2566. Immunofluorescence analysis was conducted using MUC2 (PR5D5, 1:250), TFF3 (1:1000), and CellBrite (1:2000). Scale bars: 10 μm, 20 μm, and 50 μm.

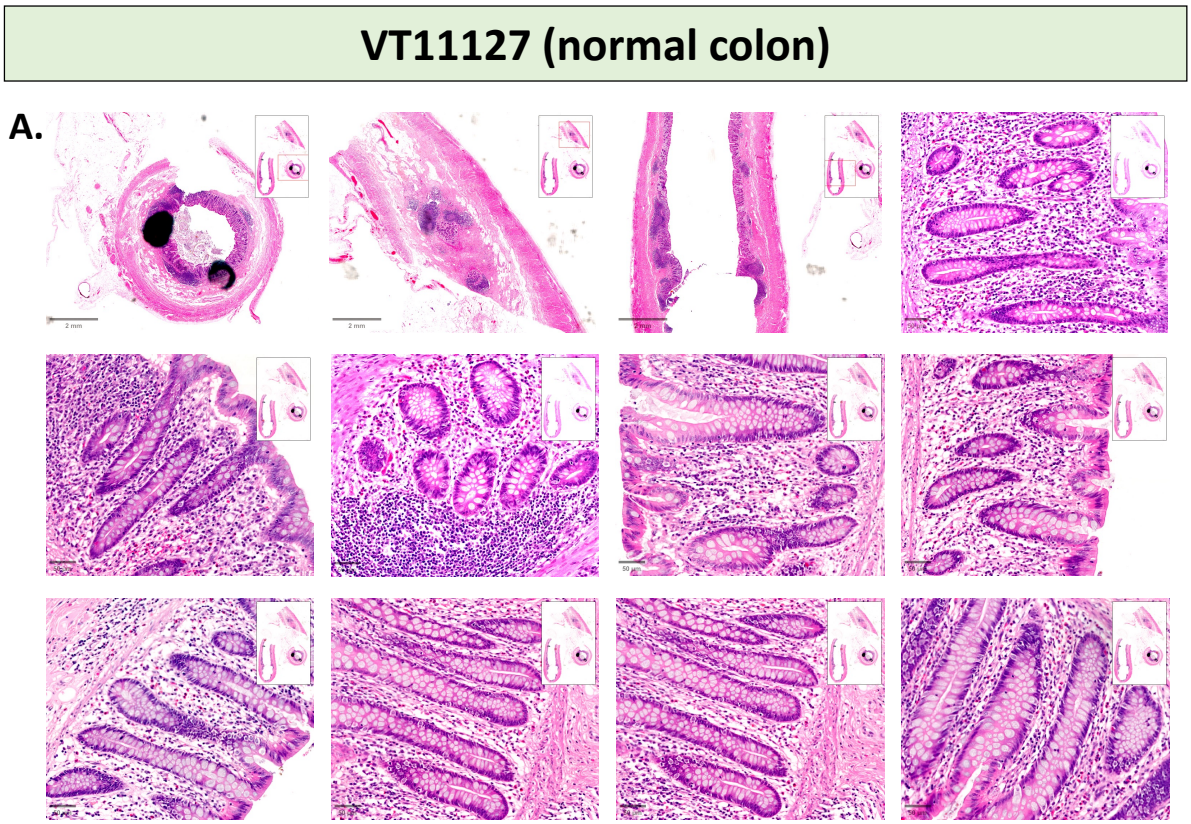



50 µm


20 µm

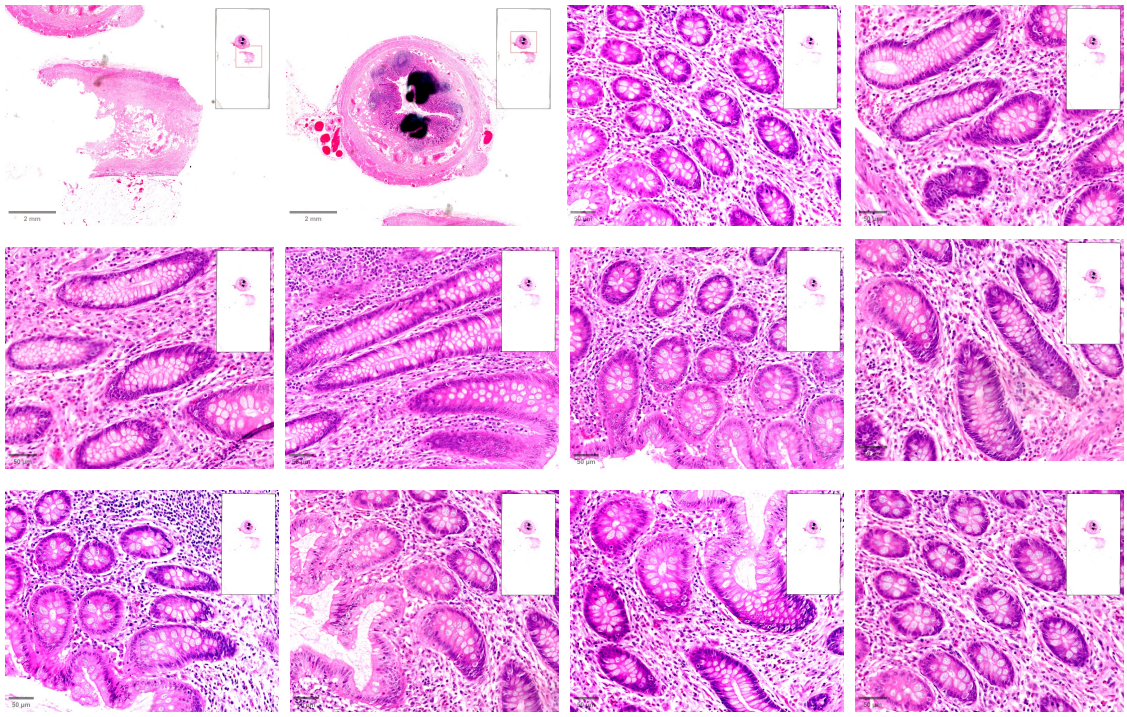

10 µm

Appendix Figure 5. H&E and immunofluorescence staining of the four paired normal colon tissues. (A) shows immunofluorescence staining of MUC2, TFF3, and CellBrite in two healthy colon tissues, VT11146 and VT11184. (B) provides a visual representation of H&E staining in four normal colon FFPE sections, including VT11127, VT11140, VT11146, and VT11184. Scale bars: 10 μm, 20 μm, 50 μm and 2 mm.



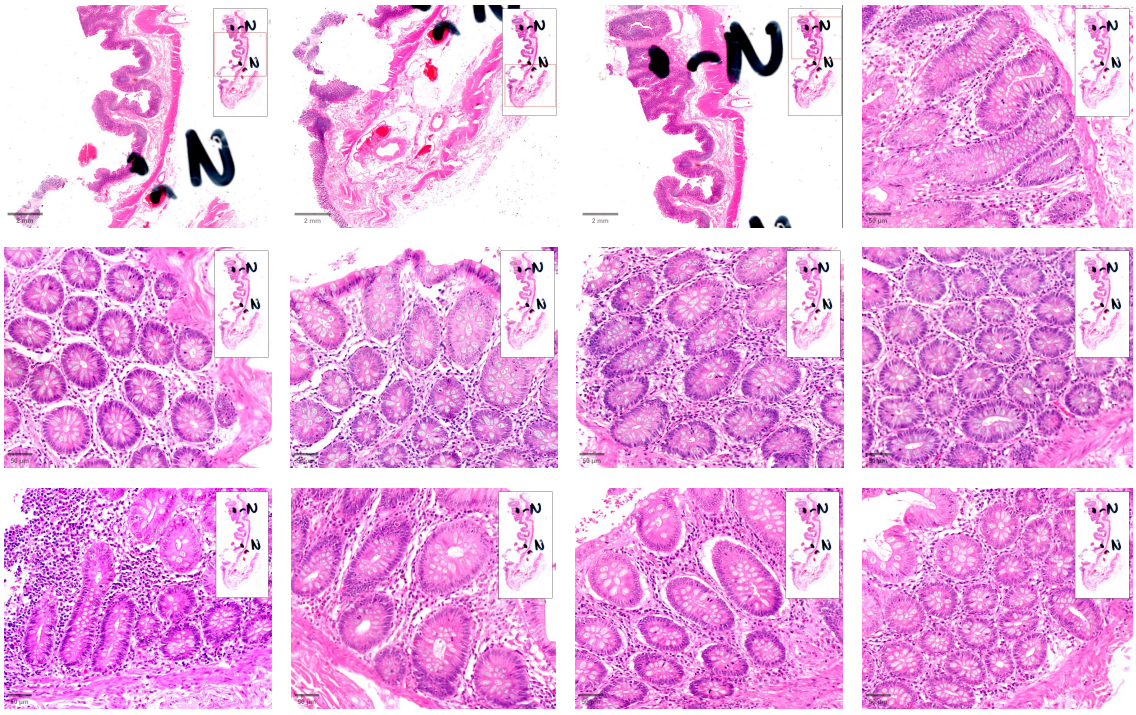
VT11140 (normal colon)

A.

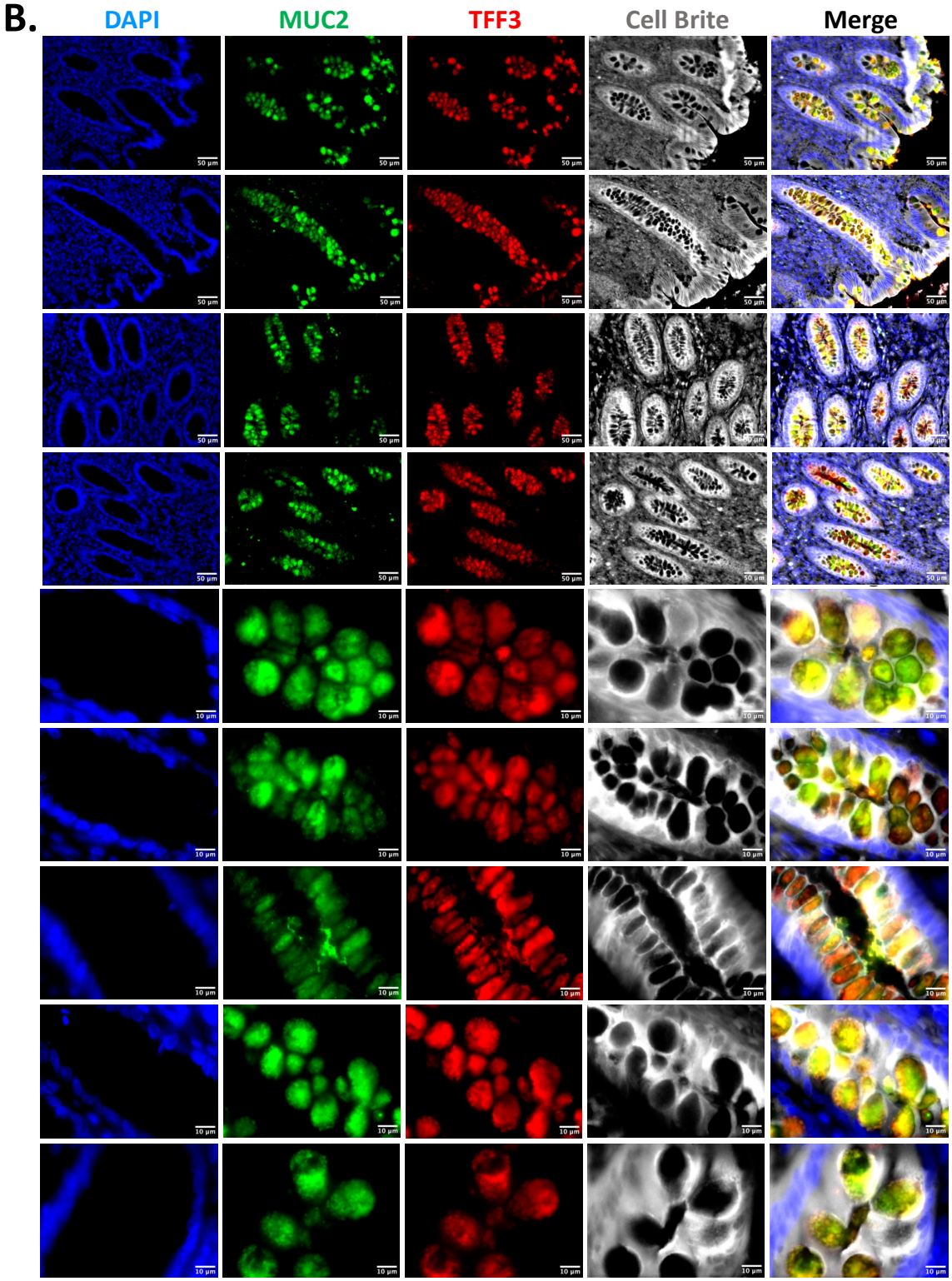


VT11146 (normal colon)

A.

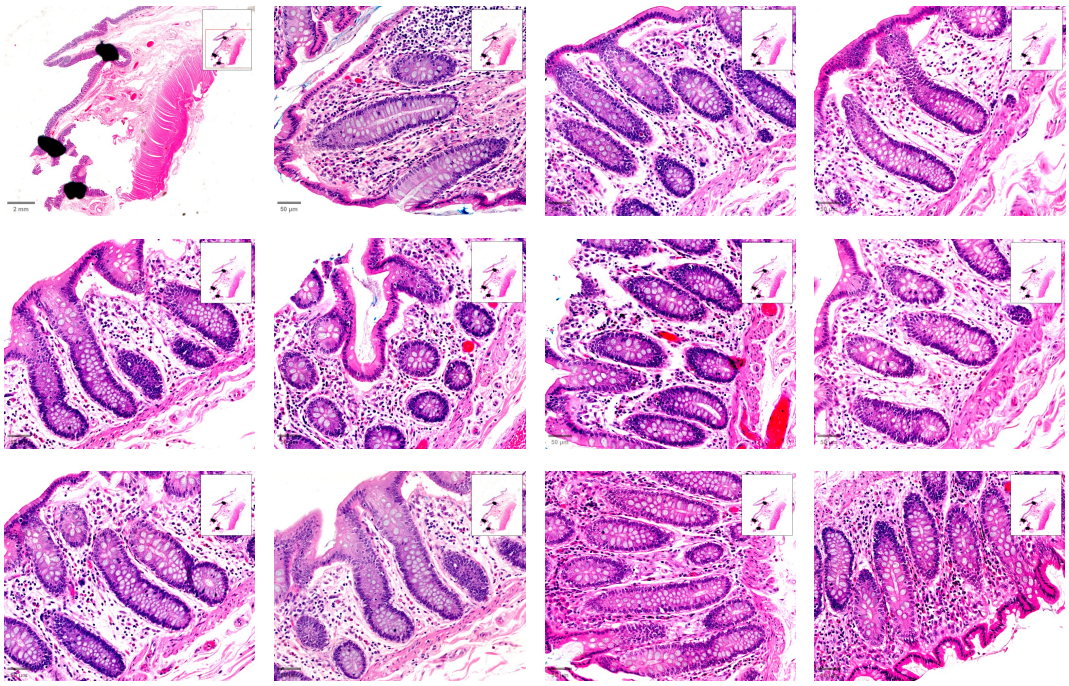


VT11146 (normal colon)



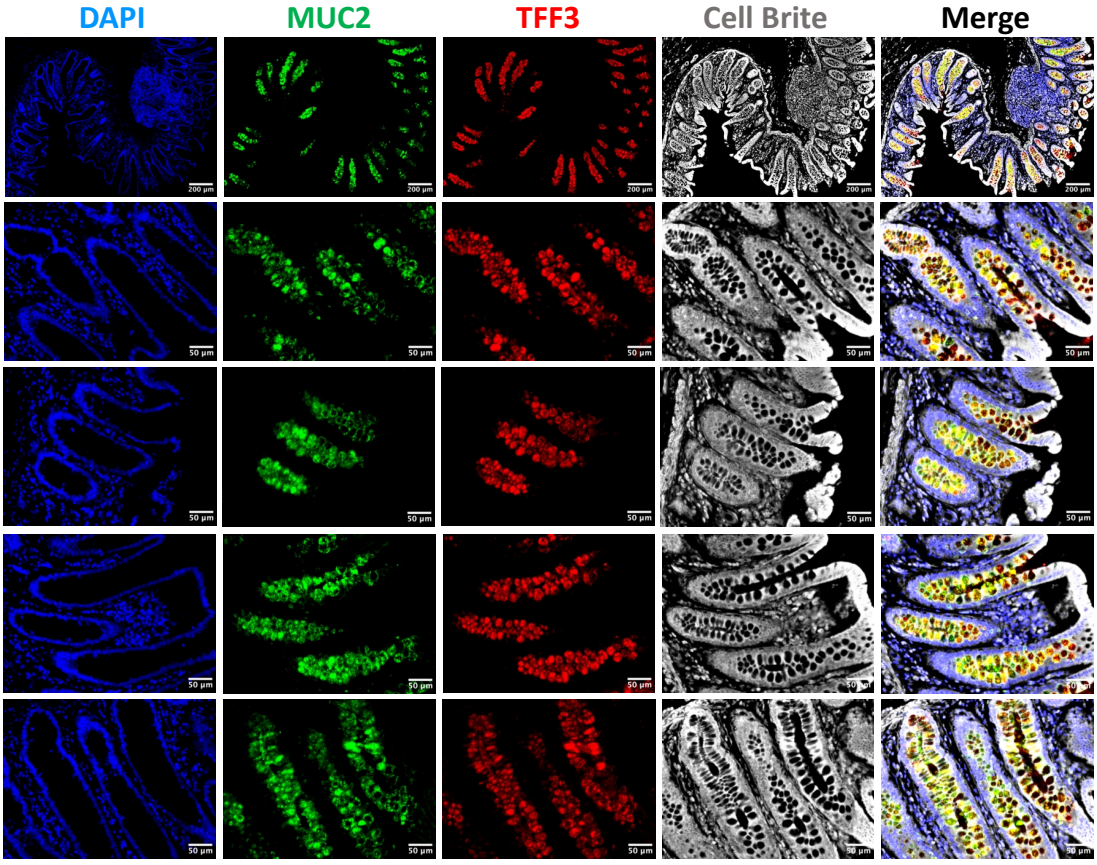
VT11184 (normal colon)

A.

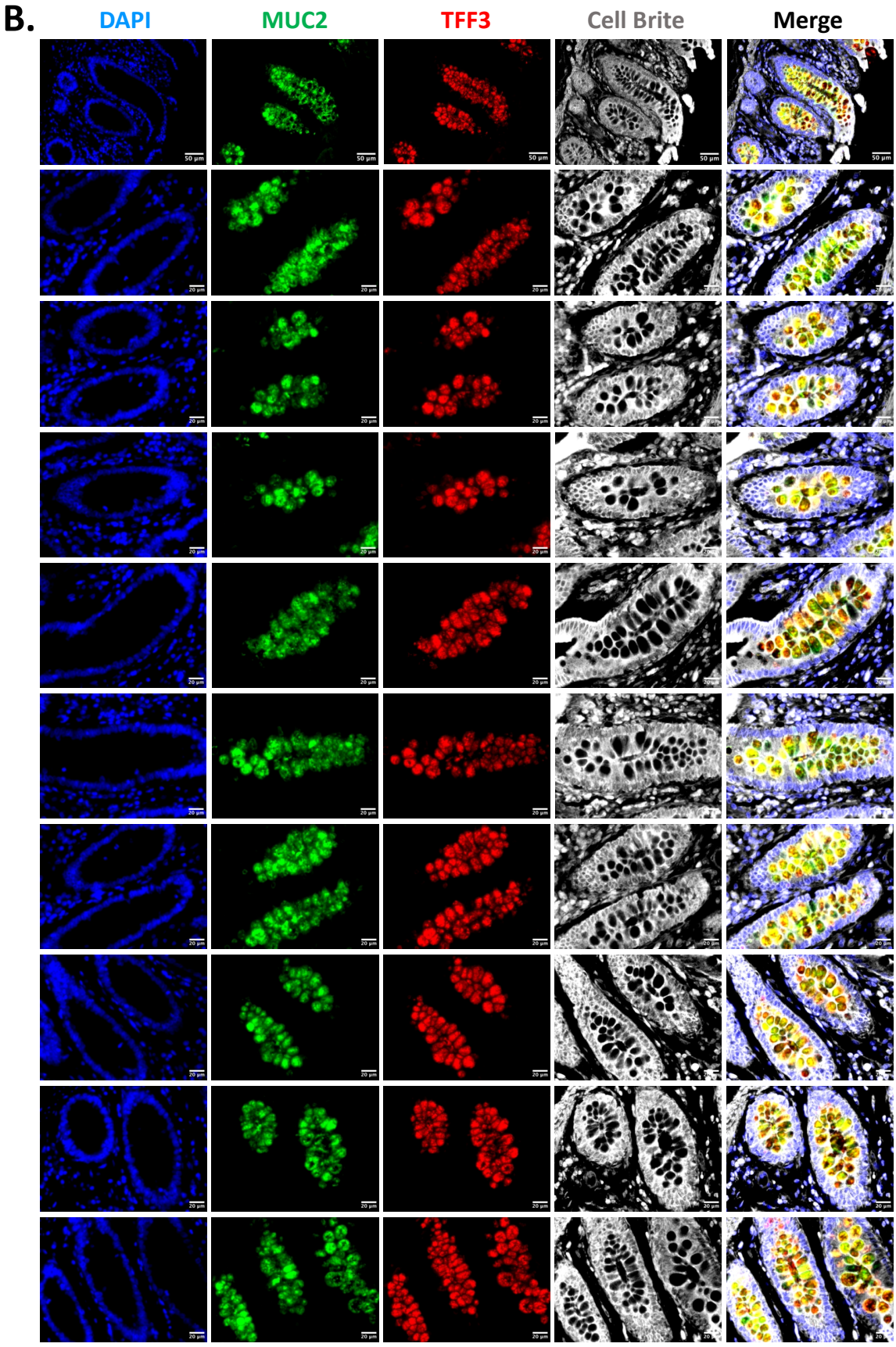


VT11184 (normal colon)

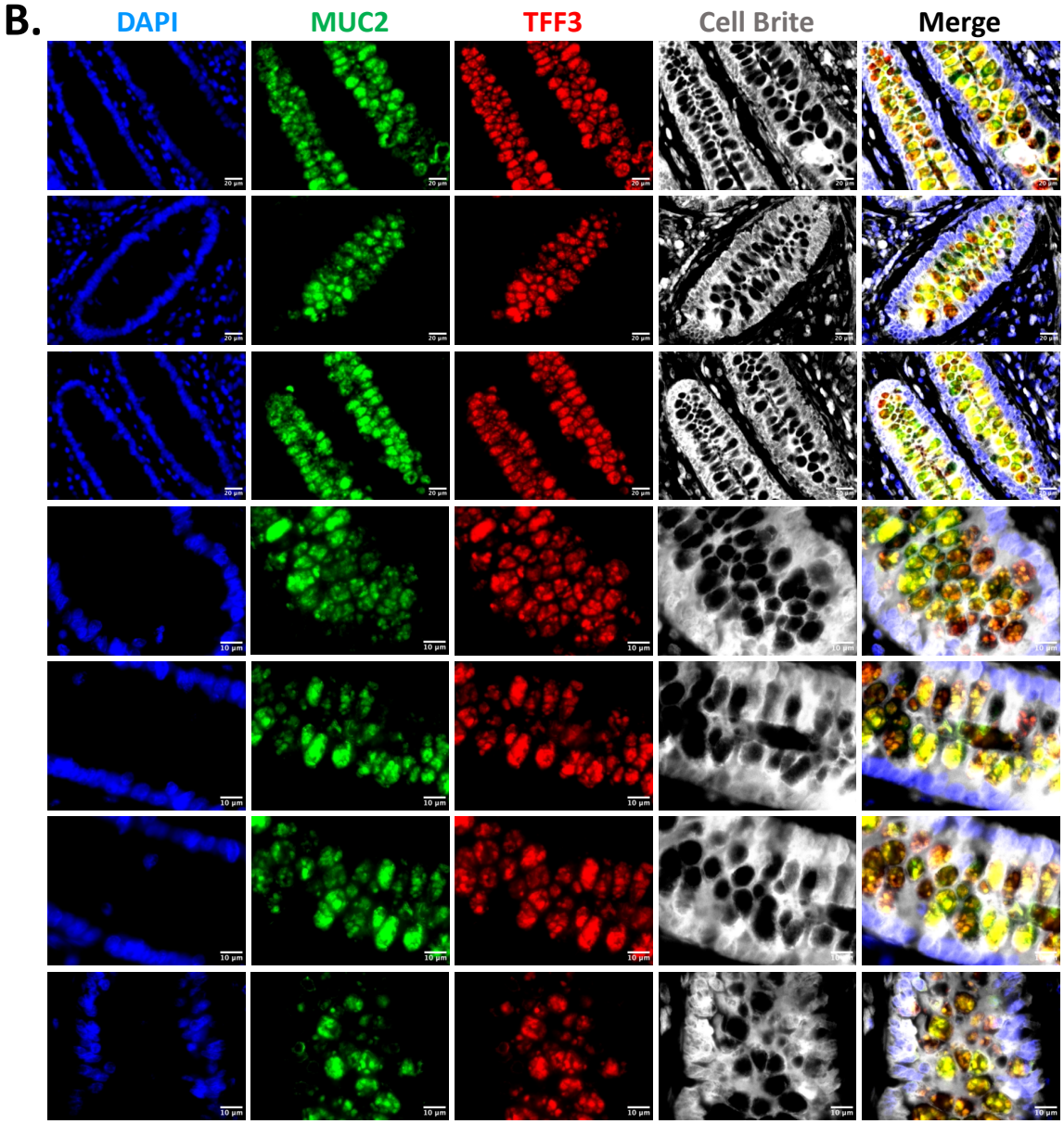
B.



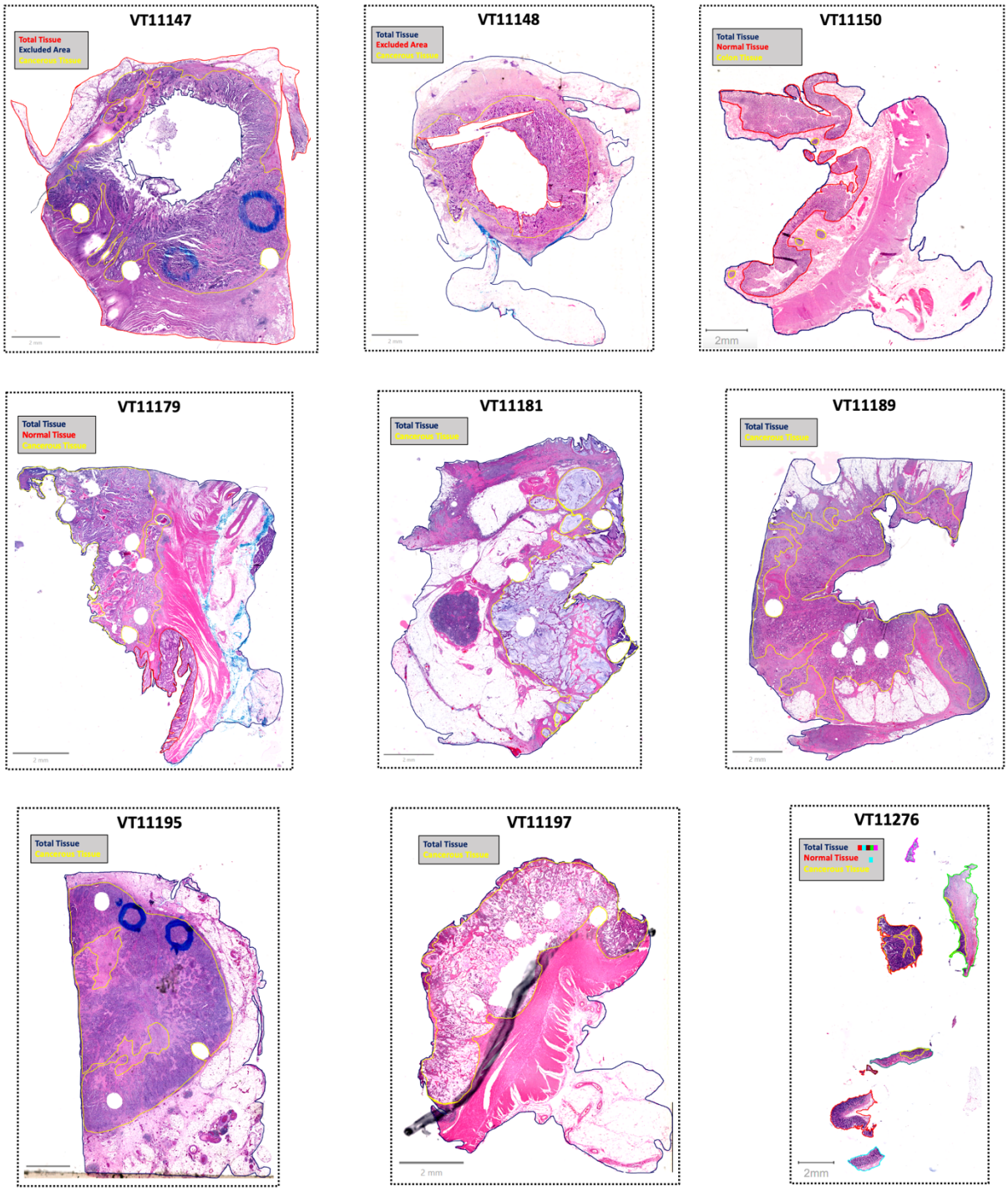
VT11184 (normal colon)

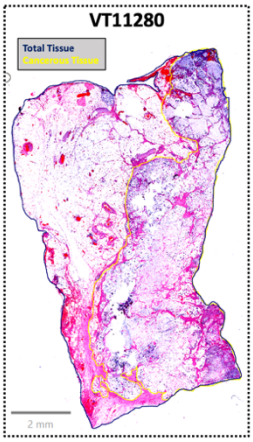


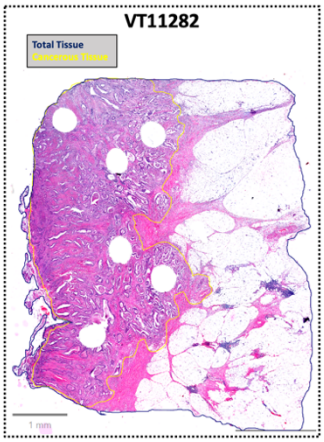
VT11184 (normal colon)



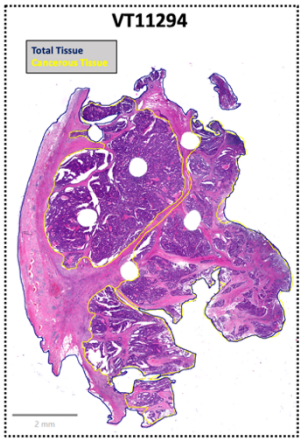
Appendix Figure 6. Histological assessment of colon cancer tissue sections. The measurement was conducted based on the surface area of normal colon, cancerous, and total tissues. The surface area of the tissue was annotated using QuPath software and matched with the colour of the corresponding tissue in the figure legend. Scale bar: 2 mm.

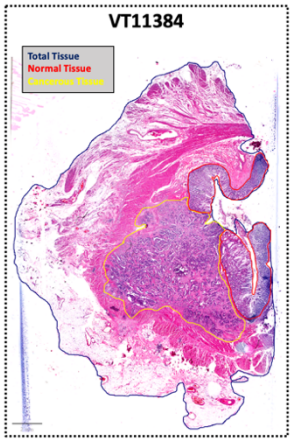


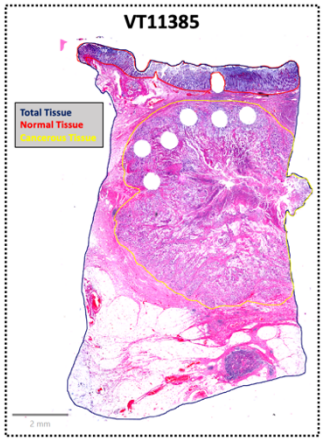

VT11280
Total Tissue
Cancerous Tissue
2 mm

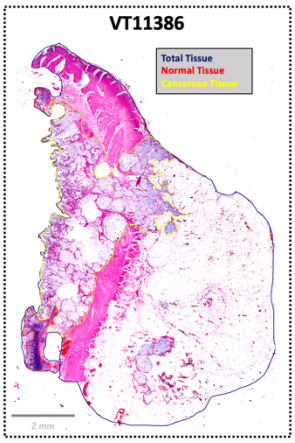

VT11282
Total Tissue
Cancerous Tissue
1 mm

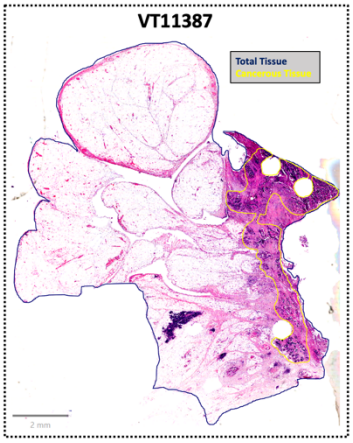

VT11290
Total Tissue
Cancerous Tissue

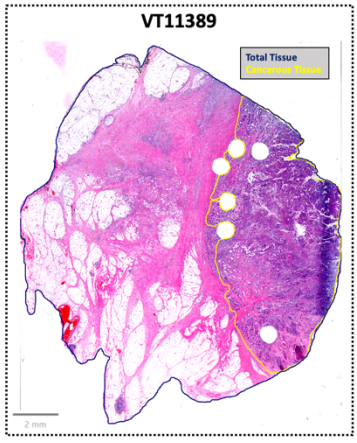

VT11294
Total Tissue
Cancerous Tissue
2 mm

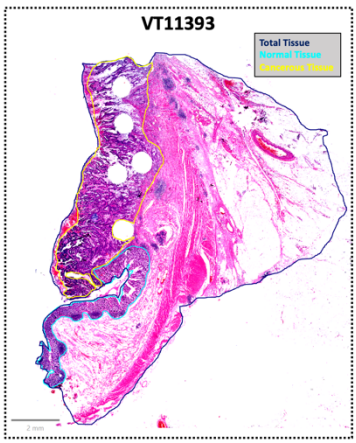

VT11384
Total Tissue
Normal Tissue
Cancerous Tissue

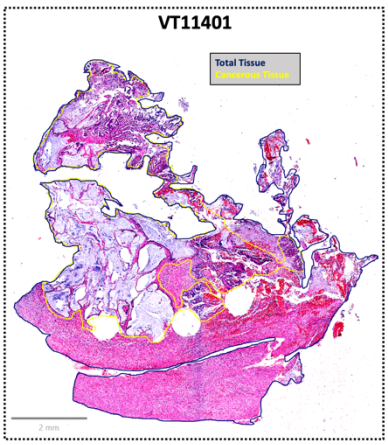

VT11385
Total Tissue
Normal Tissue
Cancerous Tissue
2 mm


VT11386
Total Tissue
Normal Tissue
Cancerous Tissue
2 mm


VT11387
Total Tissue
Cancerous Tissue
2 mm

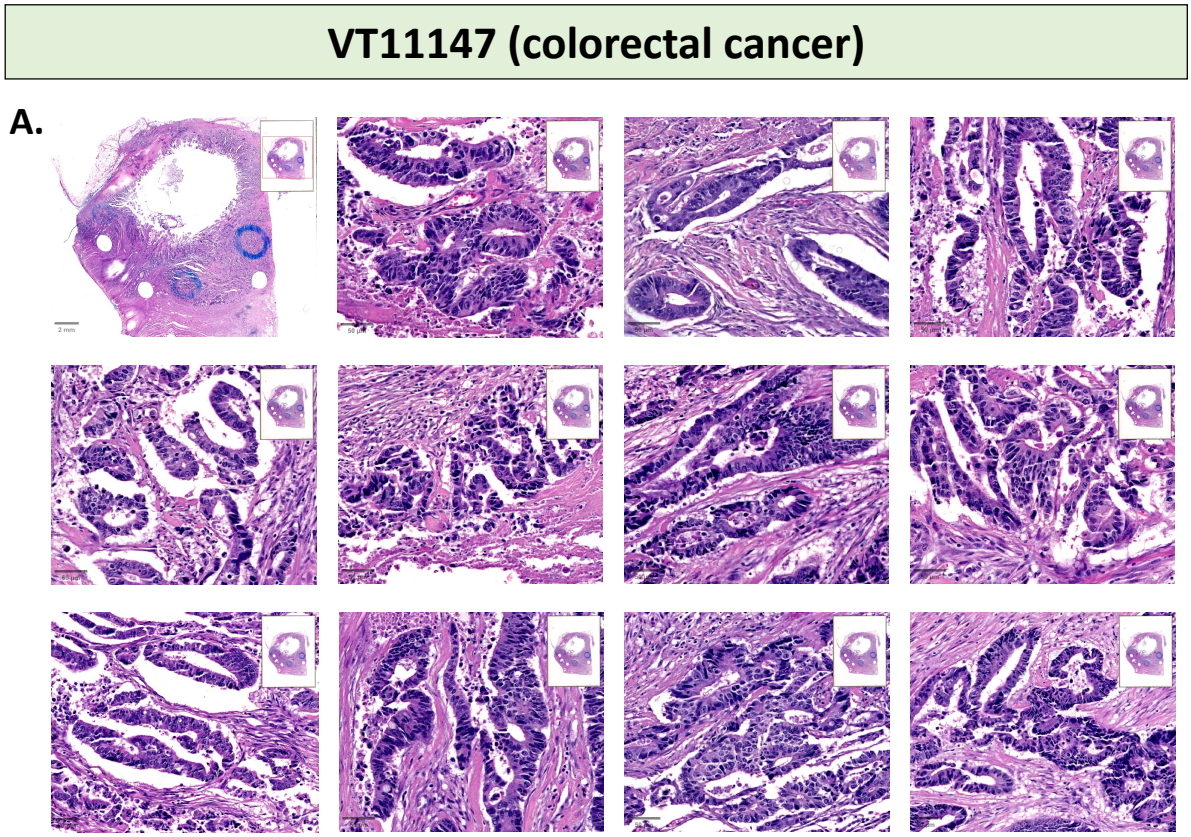

VT11389
Total Tissue
Cancerous Tissue
2 mm


VT11393
Total Tissue
Normal Tissue
Cancerous Tissue


VT11401
Total Tissue
Cancerous Tissue
2 mm

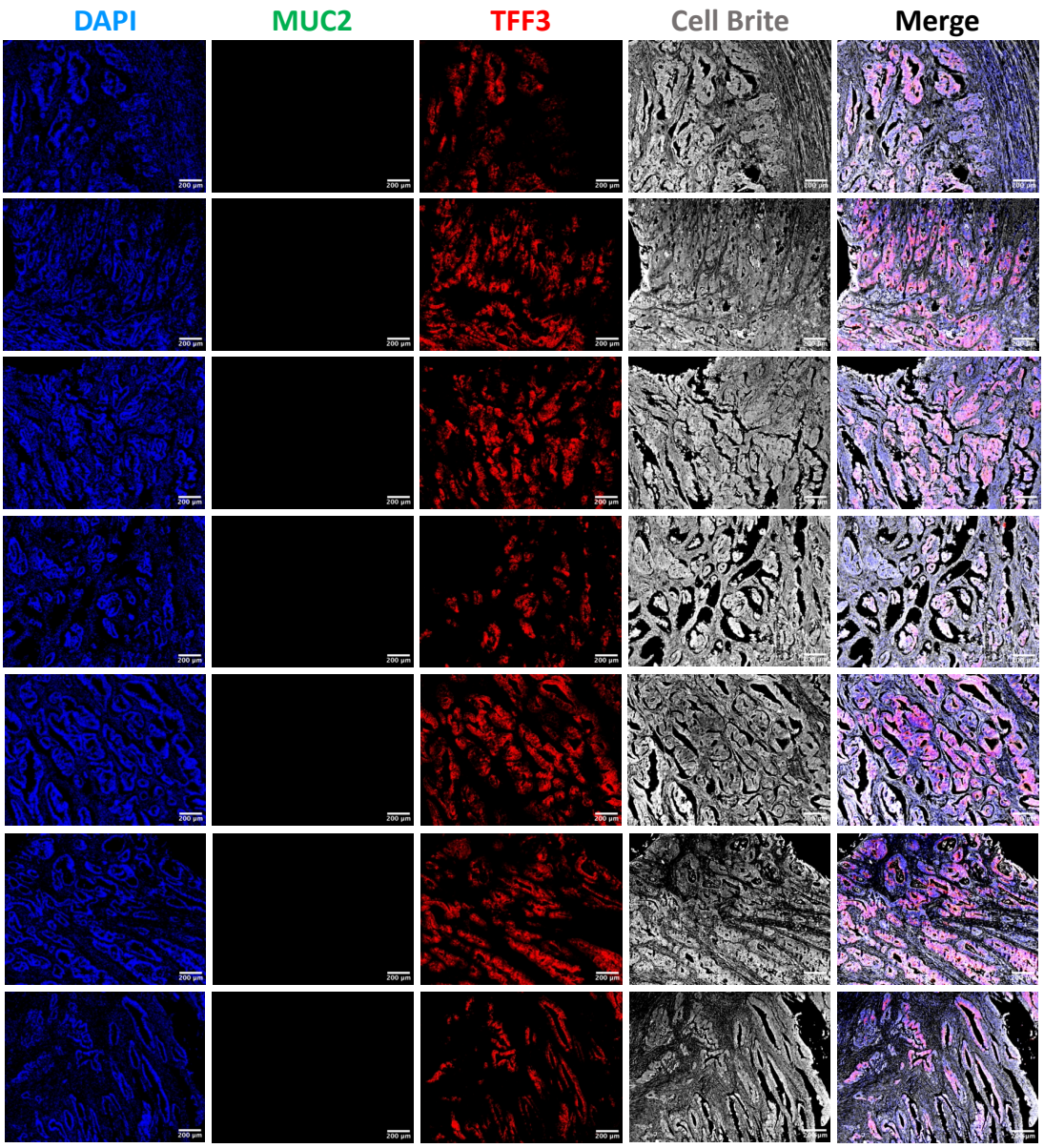

VT11429
Total Tissue
Normal Tissue
Cancerous Tissue
2 mm

Appendix Figure 7. H&E and immunofluorescence images of selected fields of view from all colon cancer tissue slides obtained from the VICTOR trial. (A) shows H&E staining of colon cancer tissues, whereas (B) describes immunofluorescence images of FFPE sections stained with MUC2 (PR5D5, 1:250), TFF3 (1:1000), DAPI (1:10000) and CellBrite (1:2000). VT11146 and VT11184. Representative images of each category are described in Figure 4.5. Scale bars: 10 µm, 20 µm, 50 µm and 2 mm.



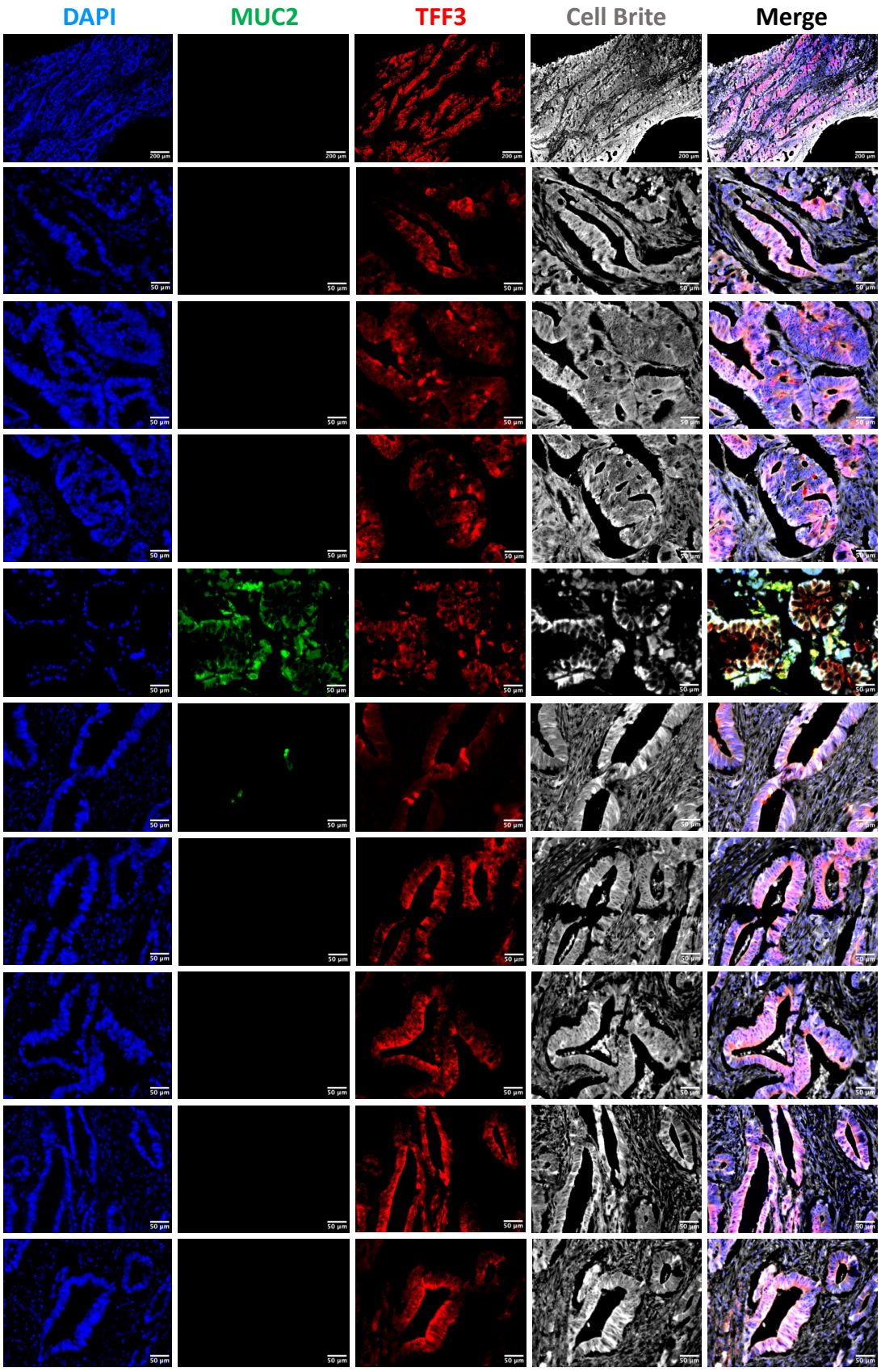
VT11147 (colorectal cancer)

B.

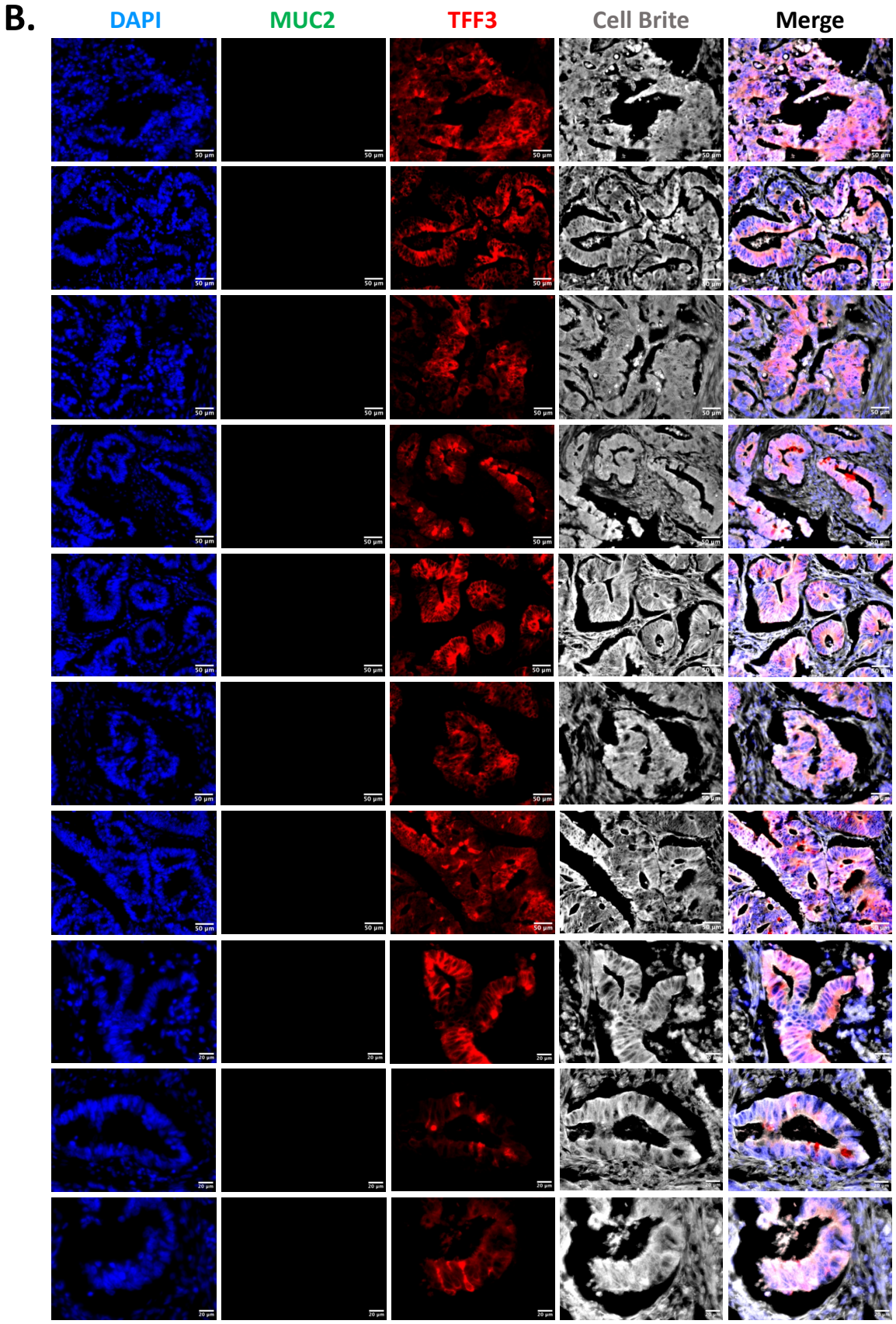


VT11147 (colorectal colon)

B.

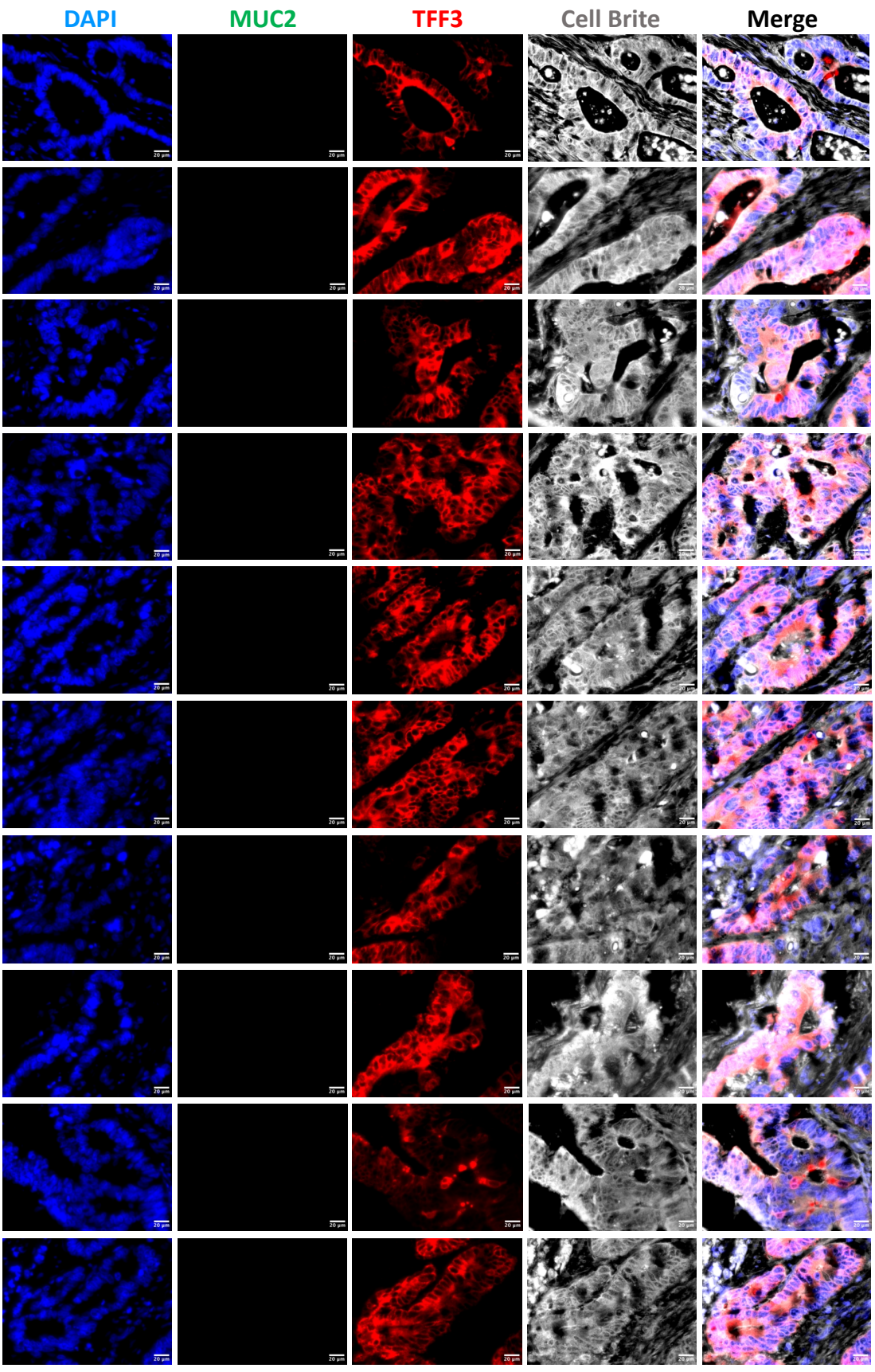


VT11147 (colorectal colon)



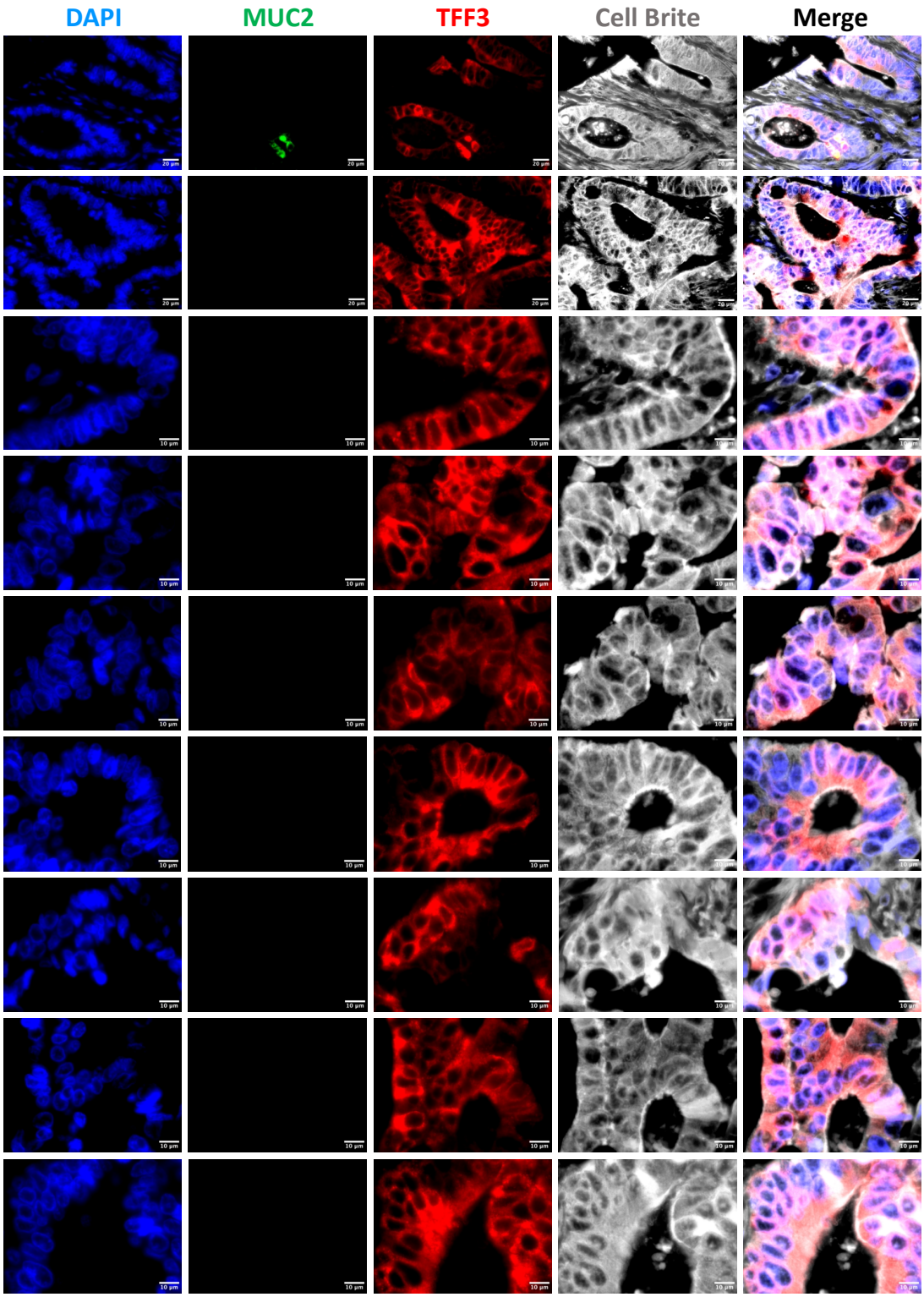
VT11147 (colorectal colon)

B.



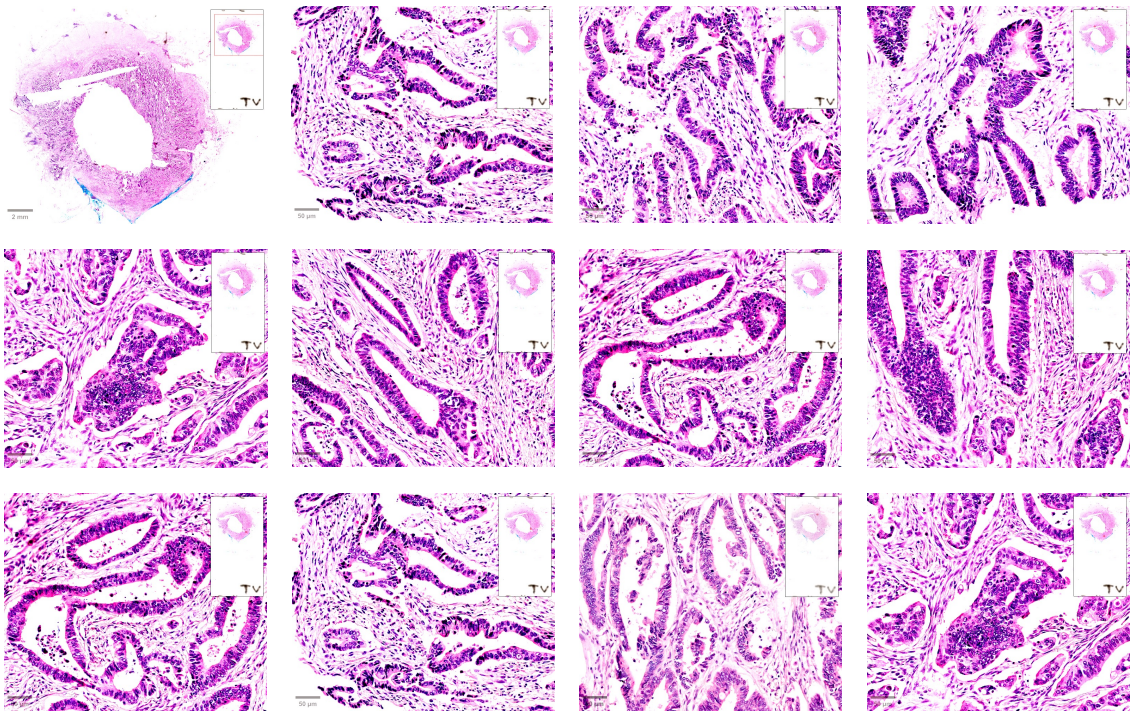
VT11147 (colorectal colon)

B.



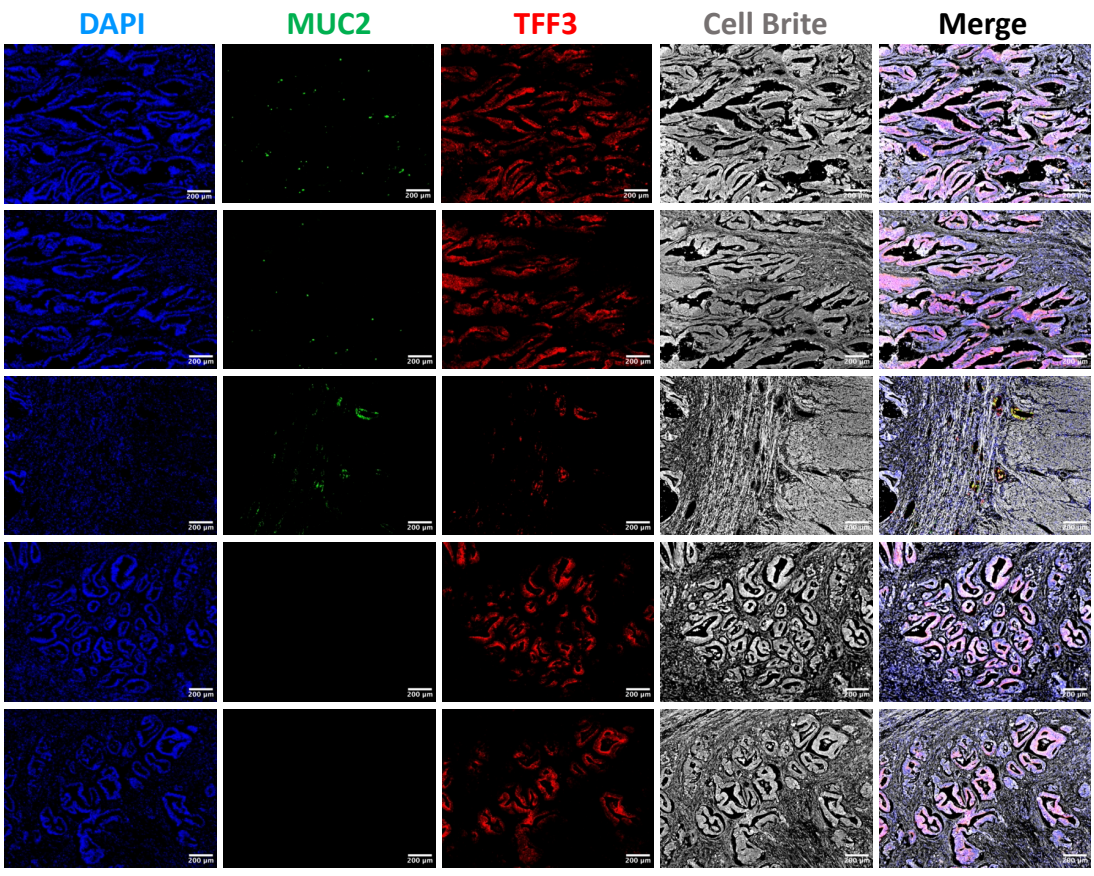
VT11148 (colorectal cancer)

A.

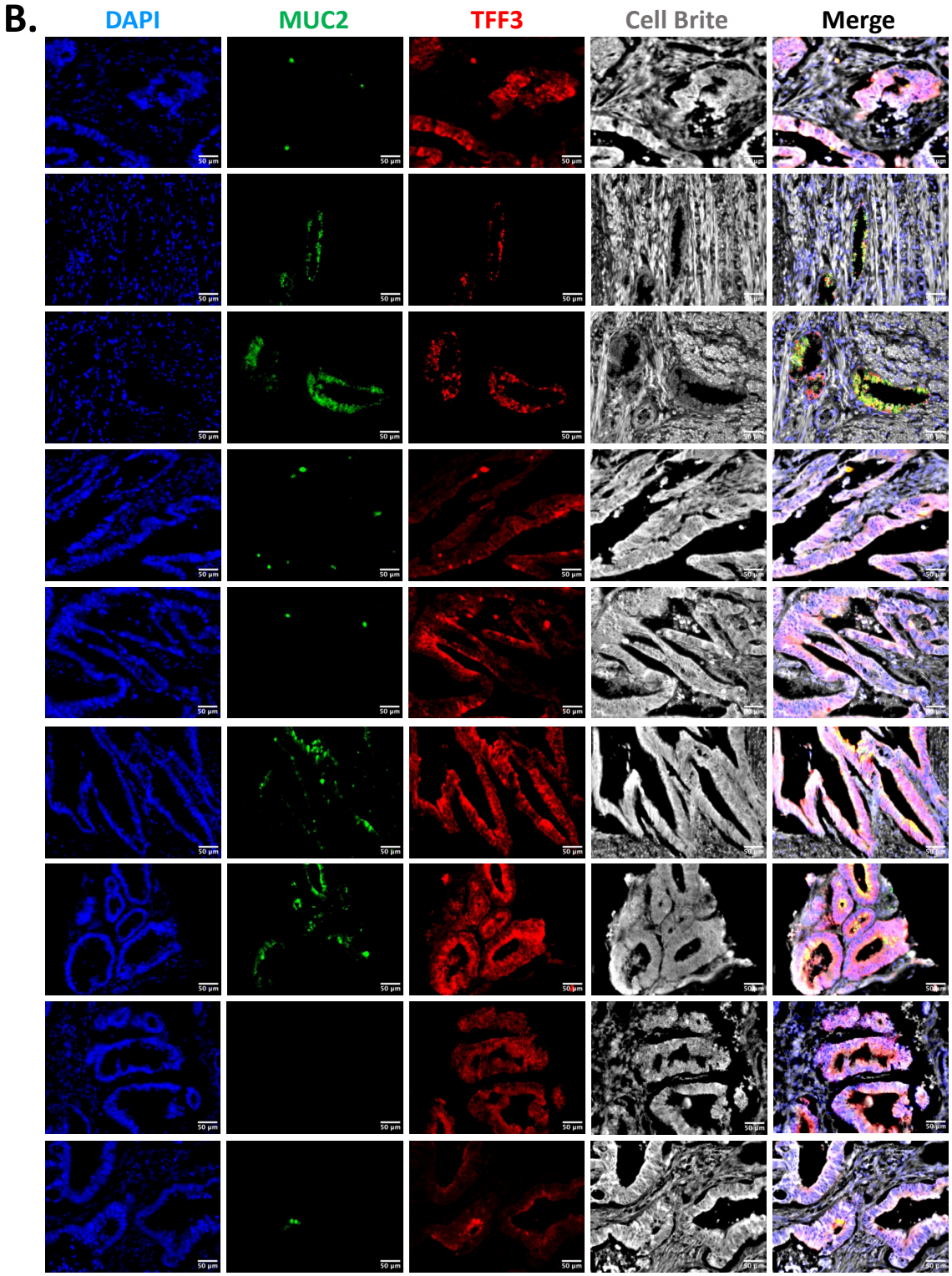


VT11148 (colorectal cancer)

B.

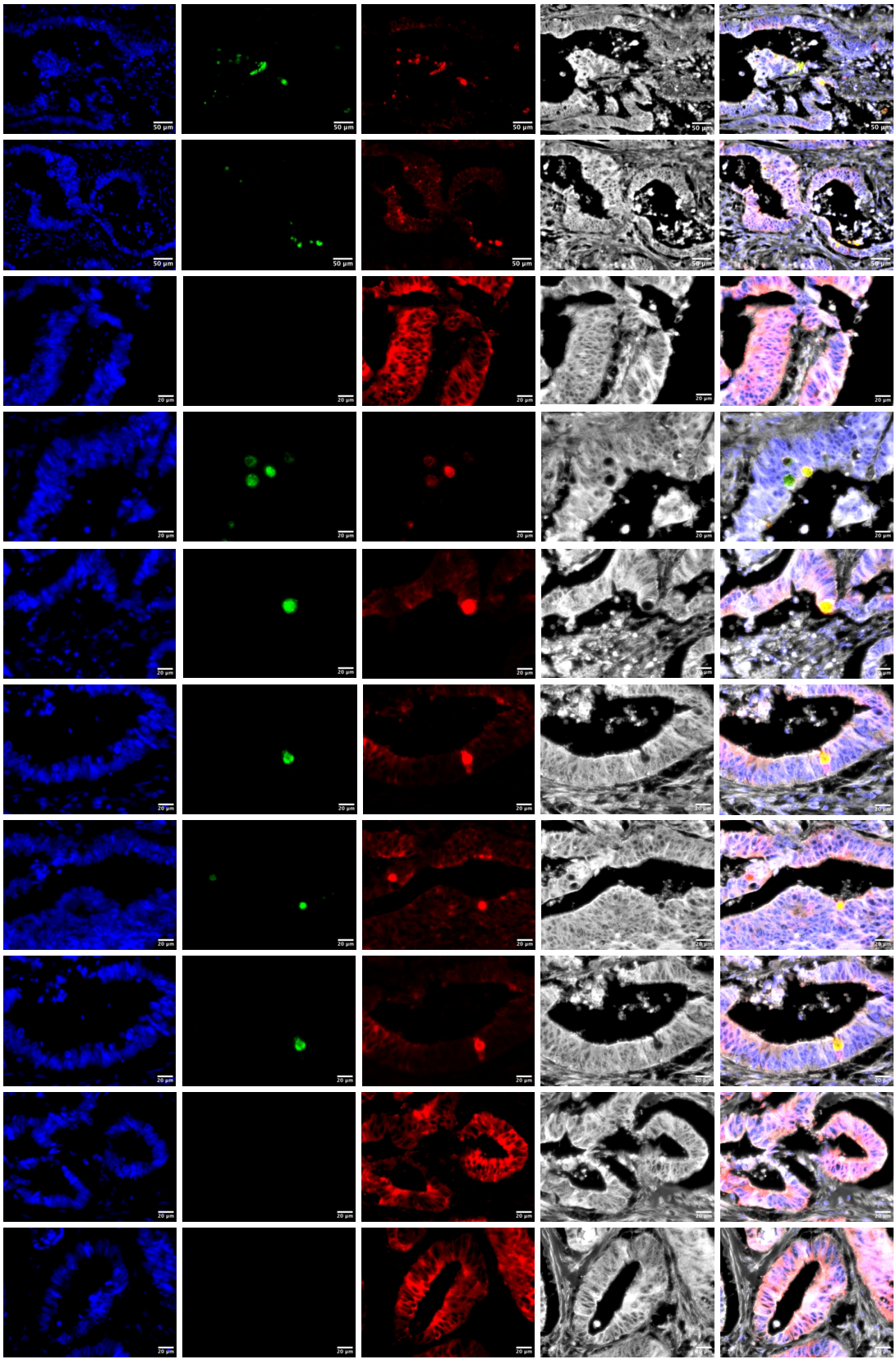


VT11148 (colorectal colon)



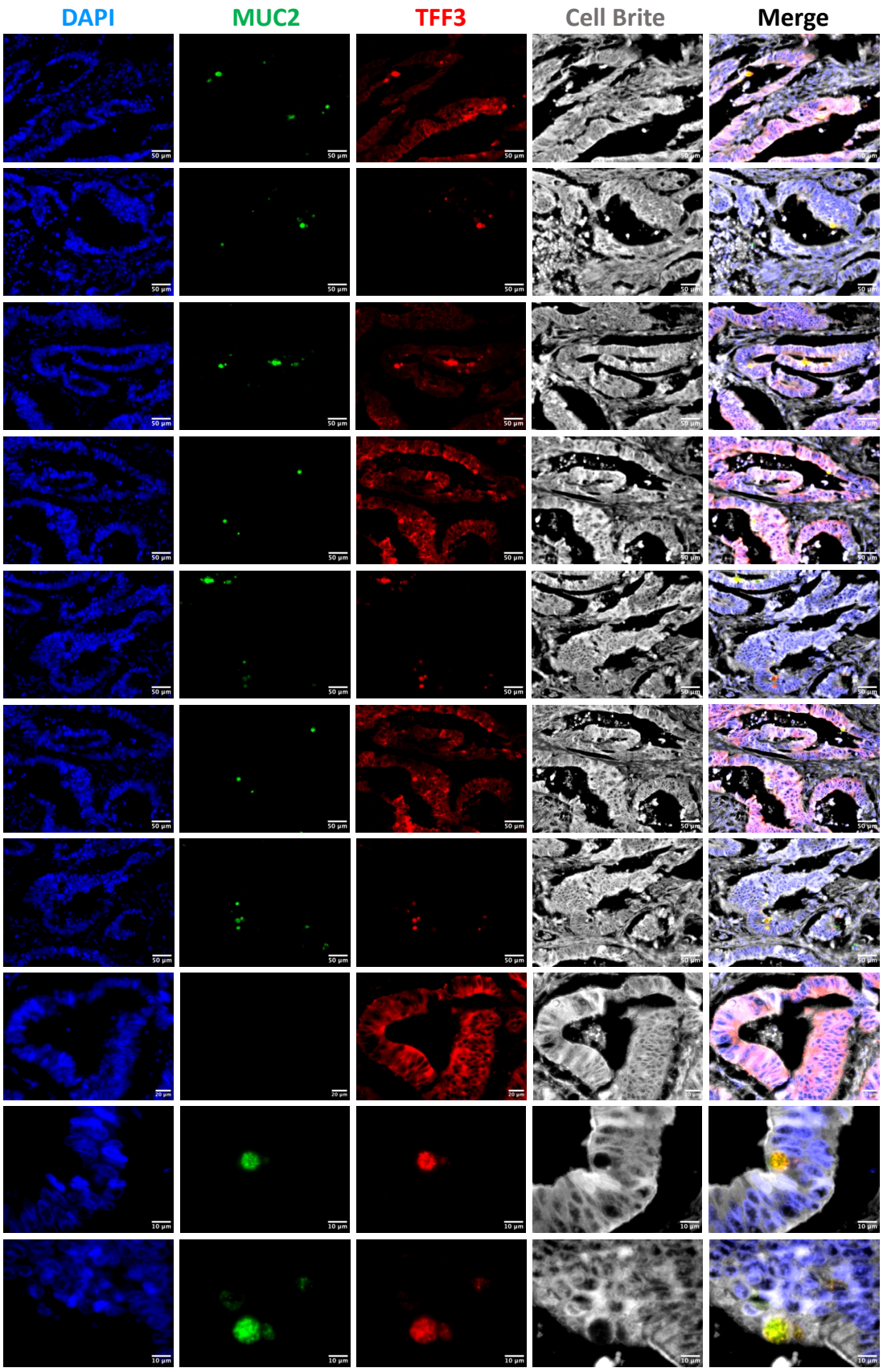
VT11148 (colorectal colon)

B.



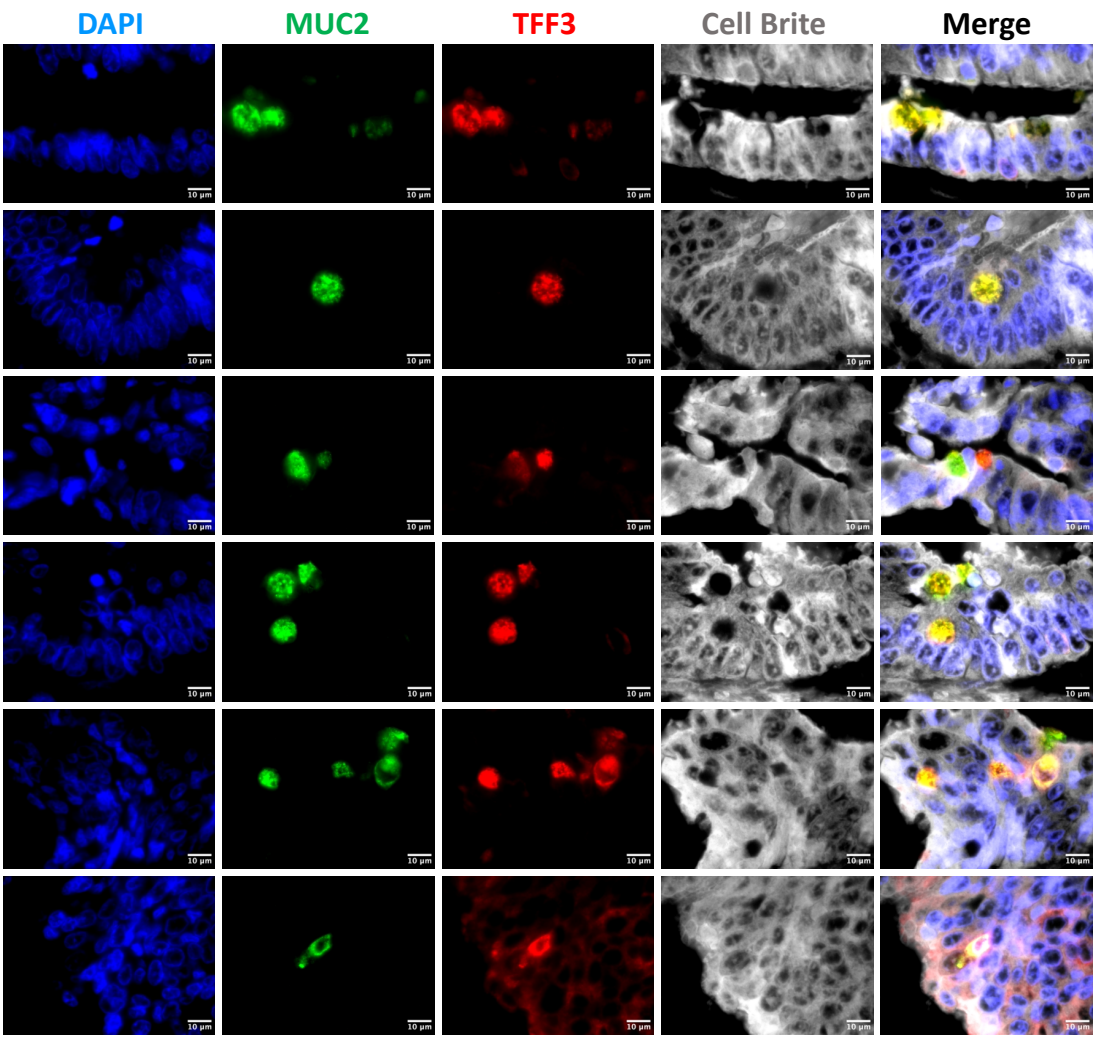
VT11148 (colorectal colon)

B.



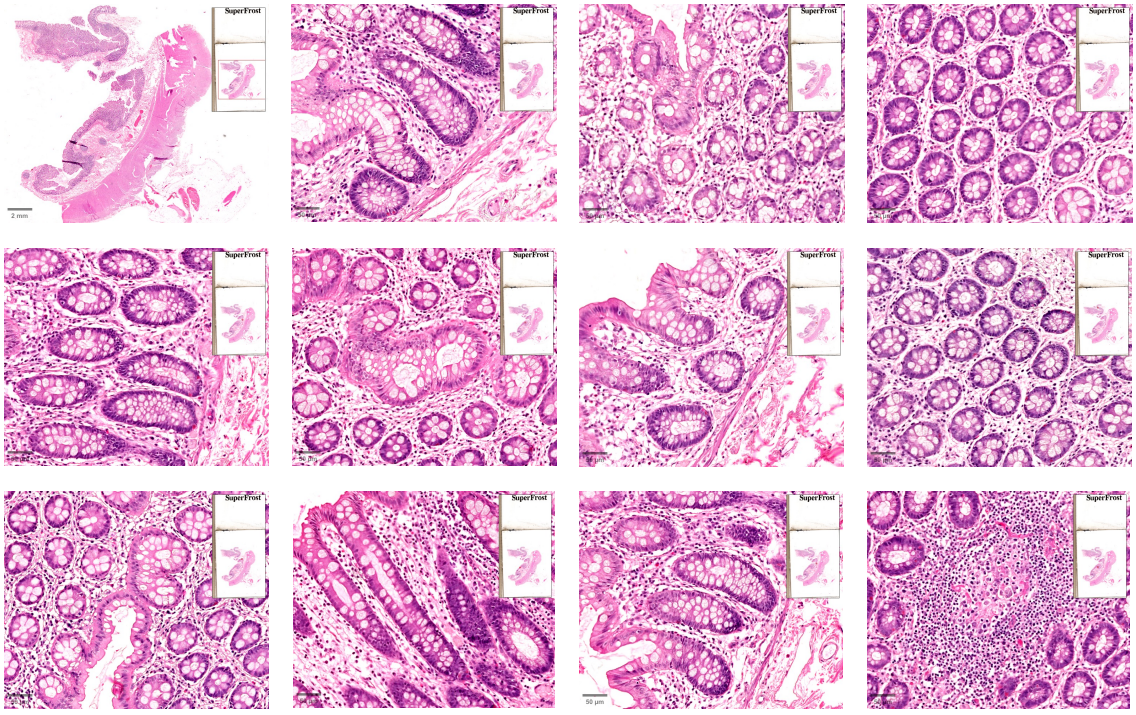
VT11148 (colorectal colon)

B.



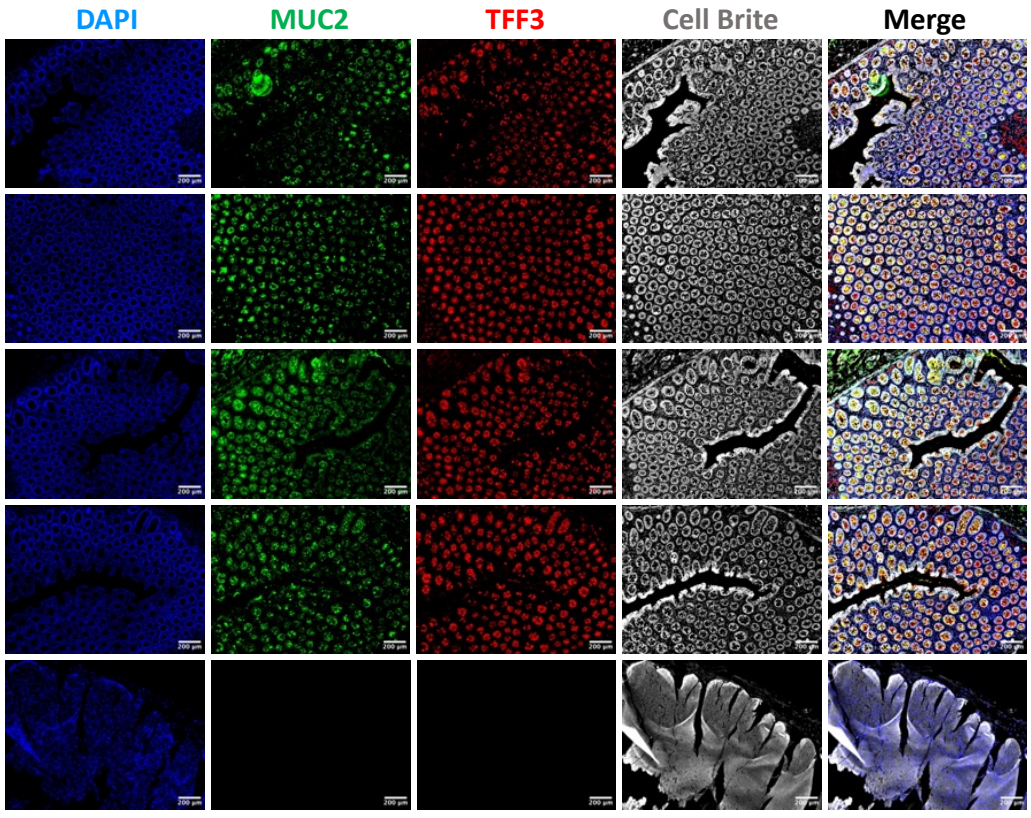
VT11150 (colorectal cancer)

A.

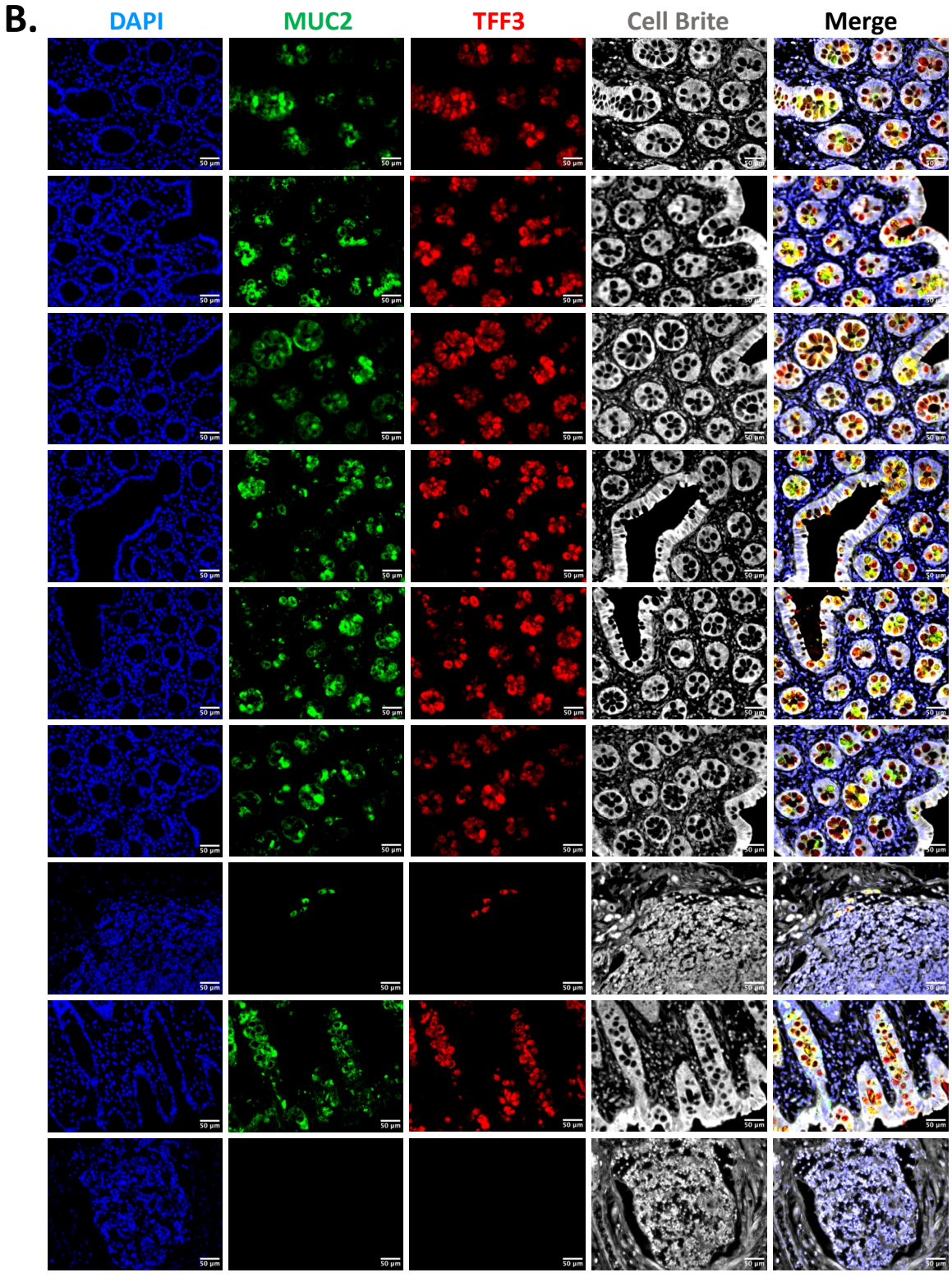


VT11150 (colorectal cancer)

B.

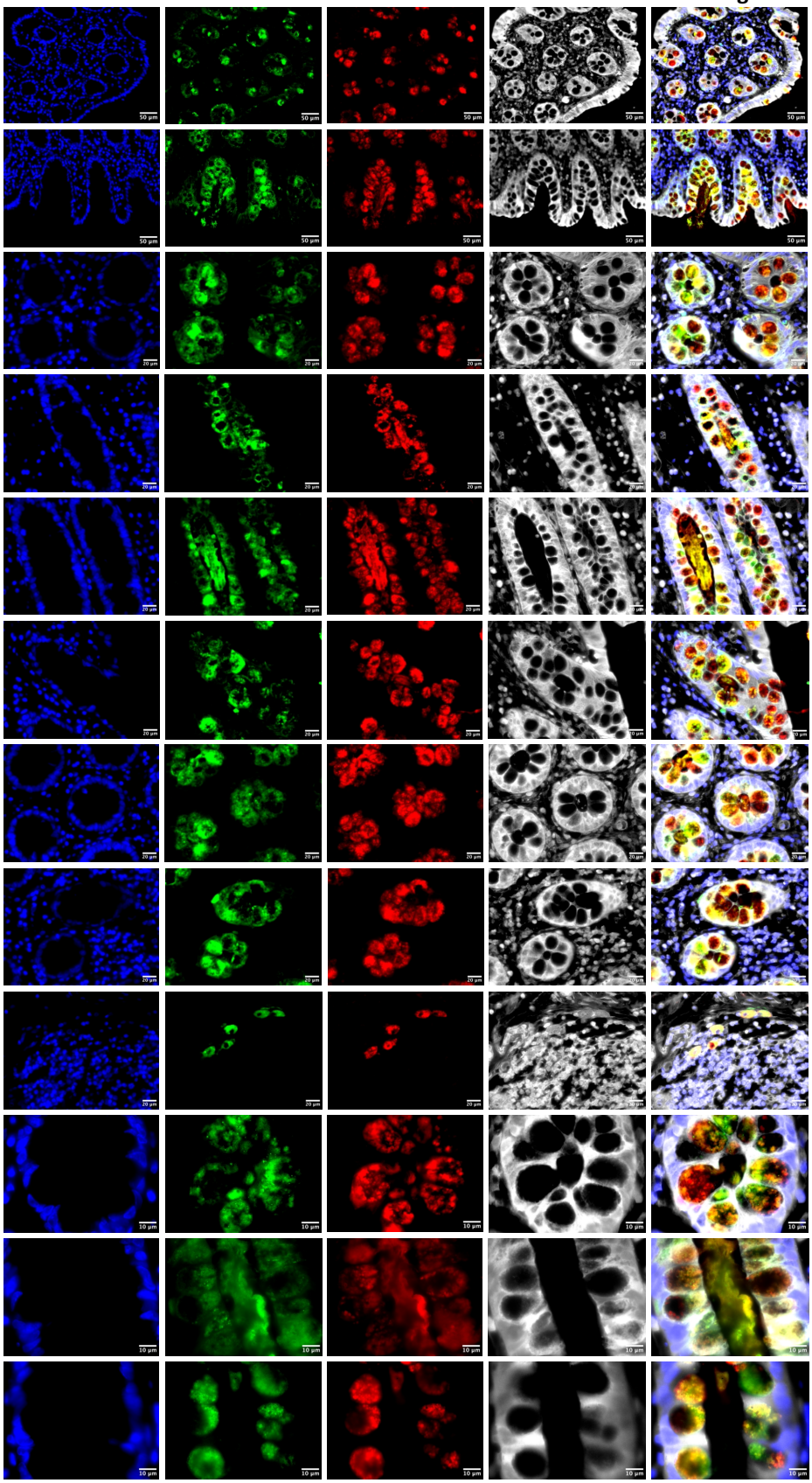


VT11150 (colorectal colon)



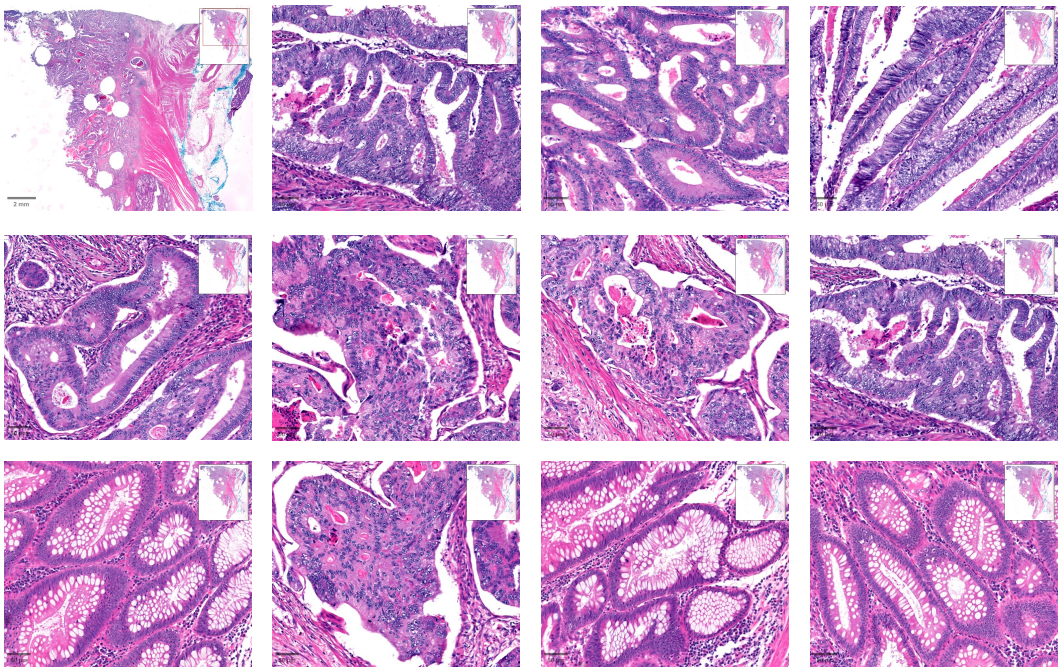
VT11150 (colorectal colon)

B.



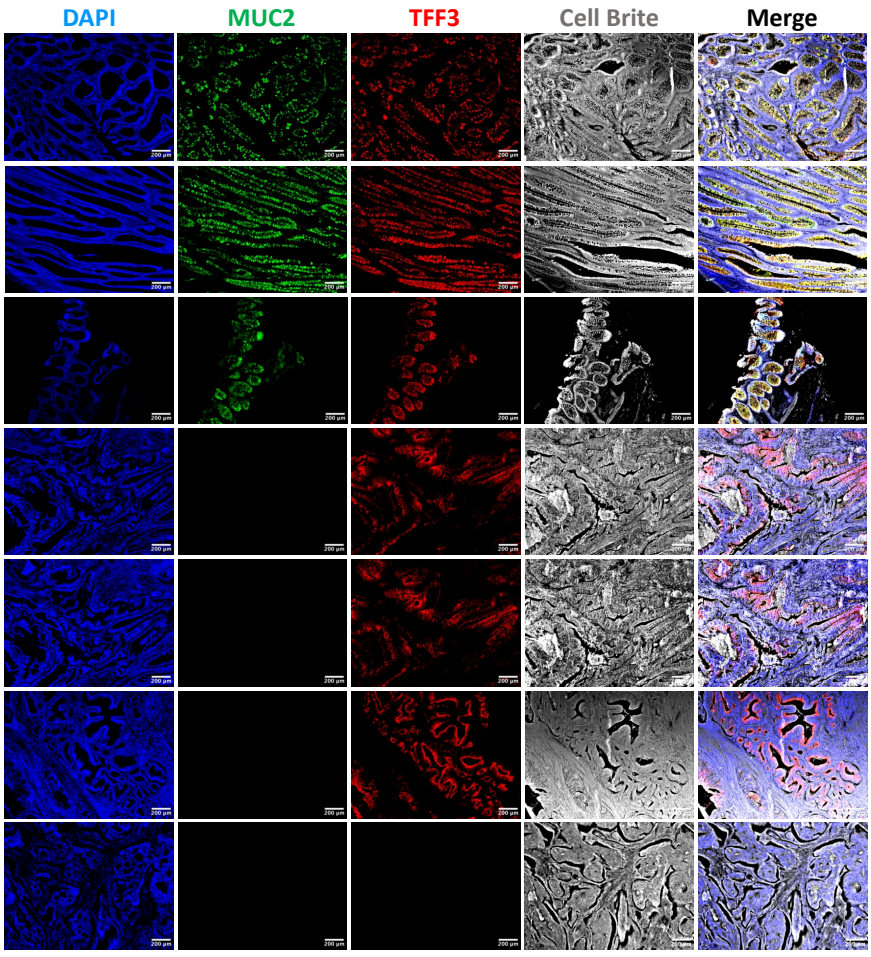
VT11179 (colorectal cancer)

A.



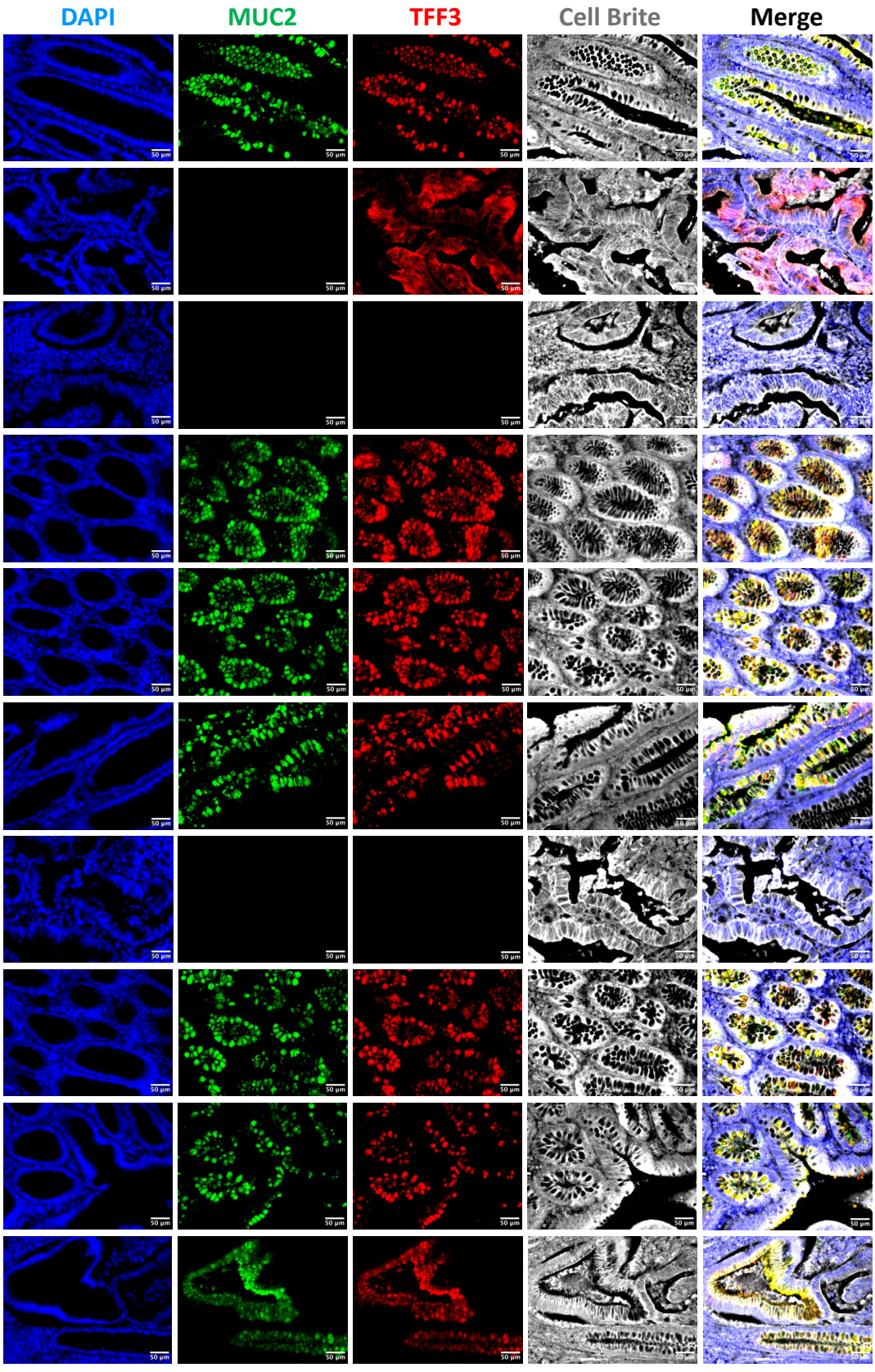
VT11179 (colorectal cancer)

B.



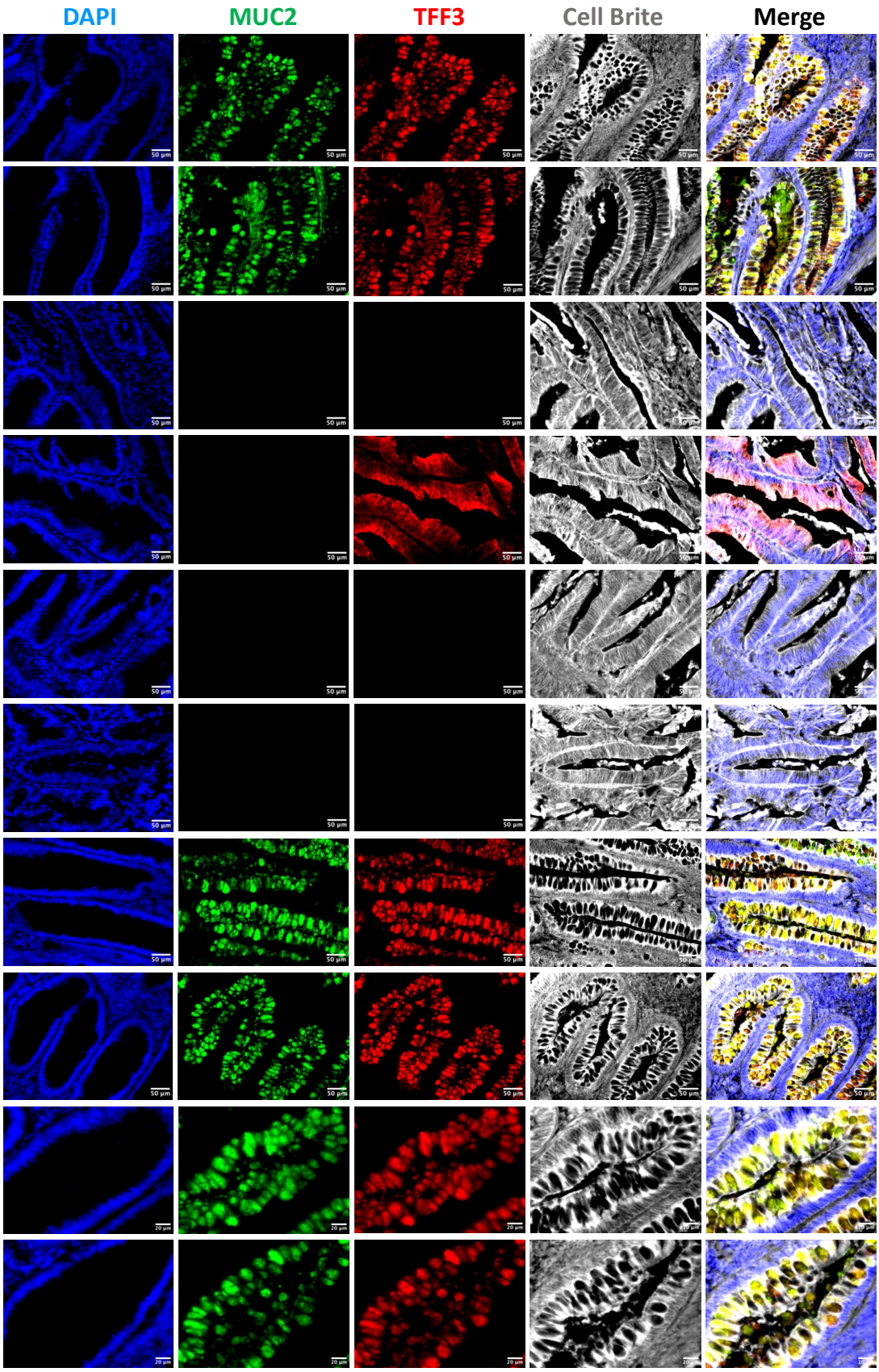
VT11179 (colorectal colon)

B.



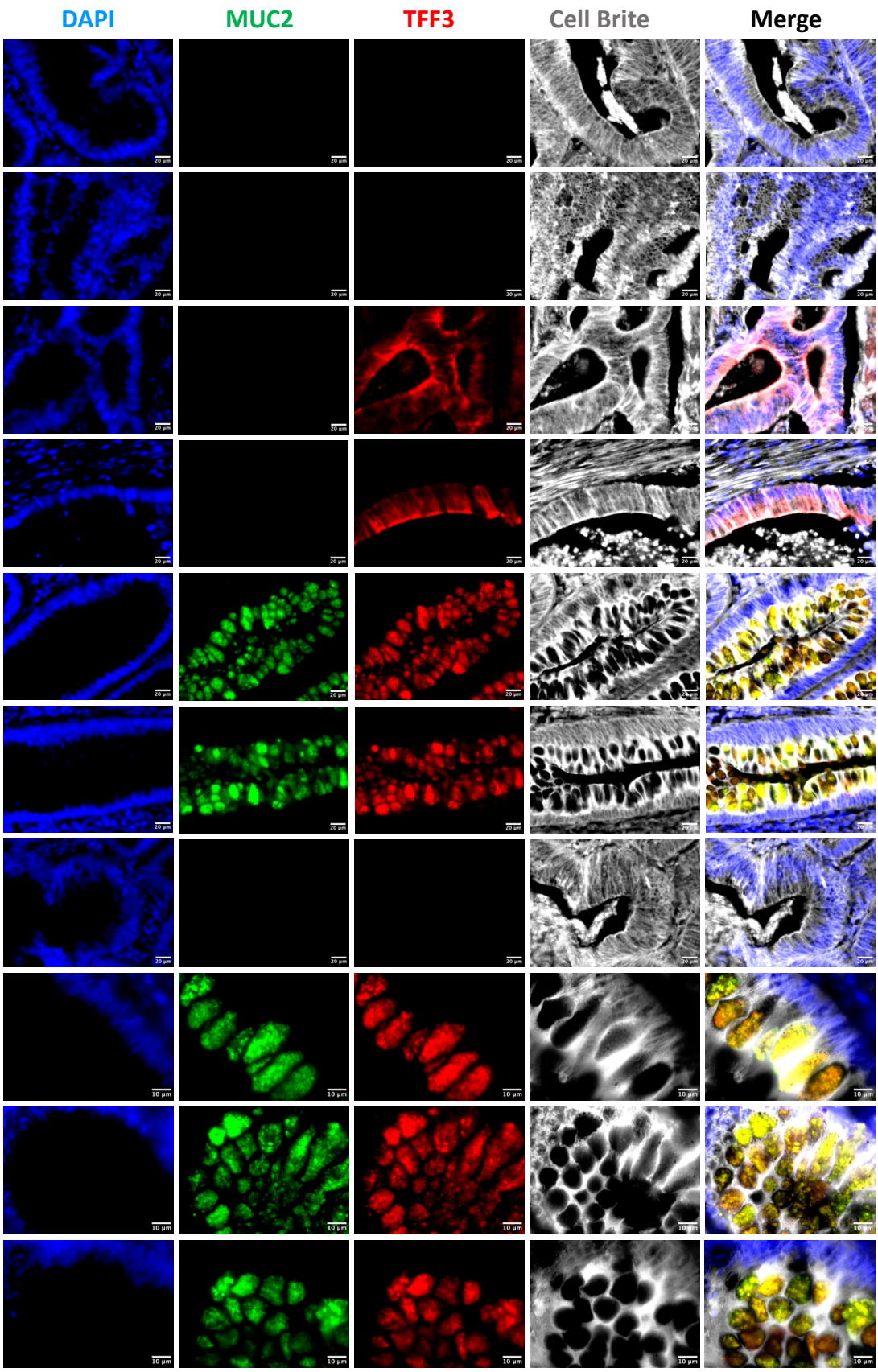
VT11179 (colorectal colon)

B.



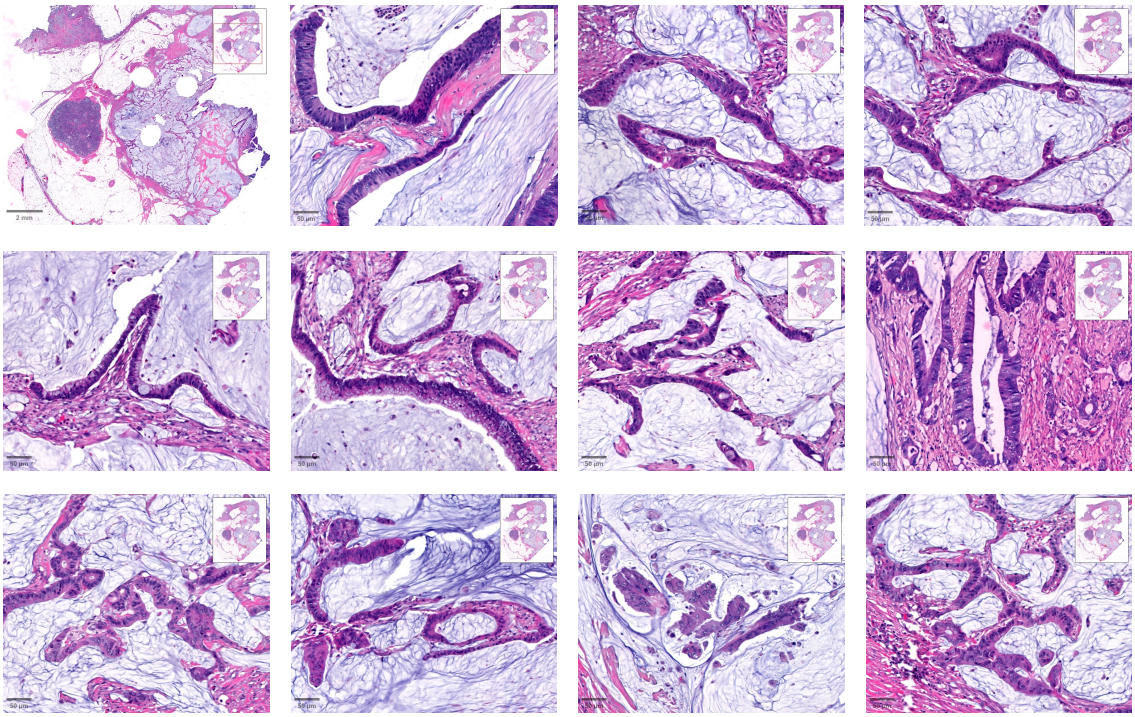
VT11179 (colorectal colon)

B.



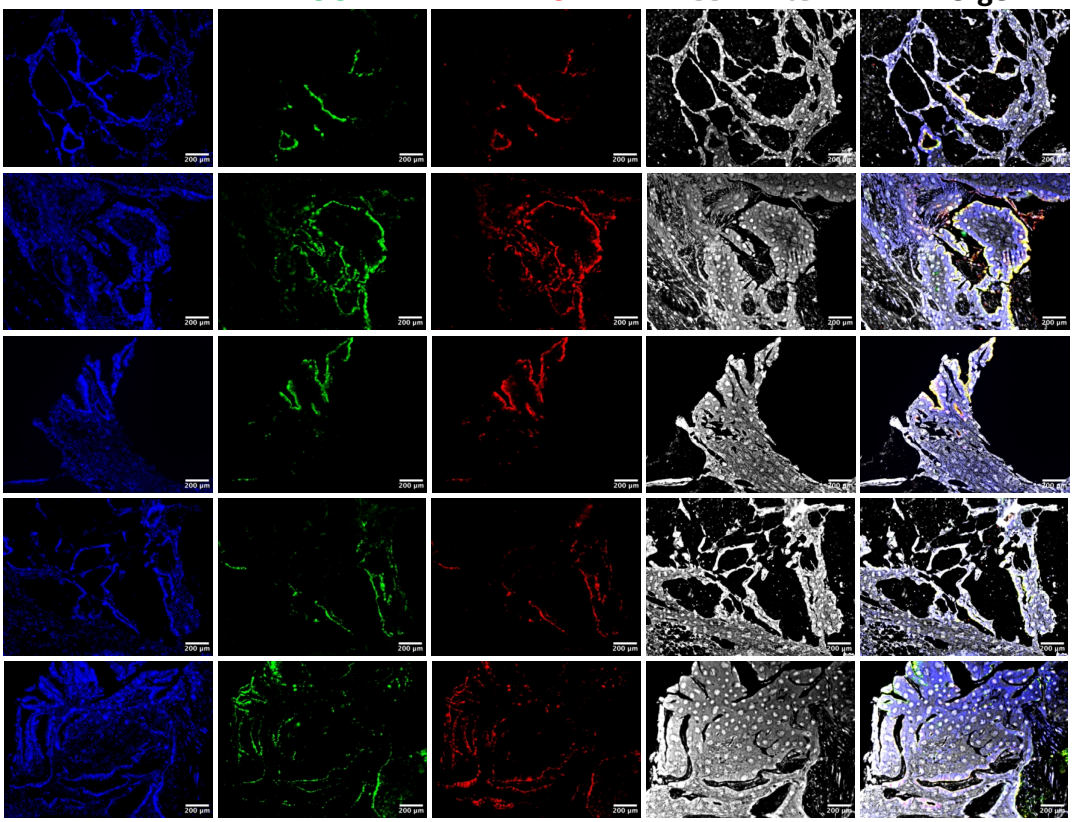
VT11181 (colorectal cancer)

A.



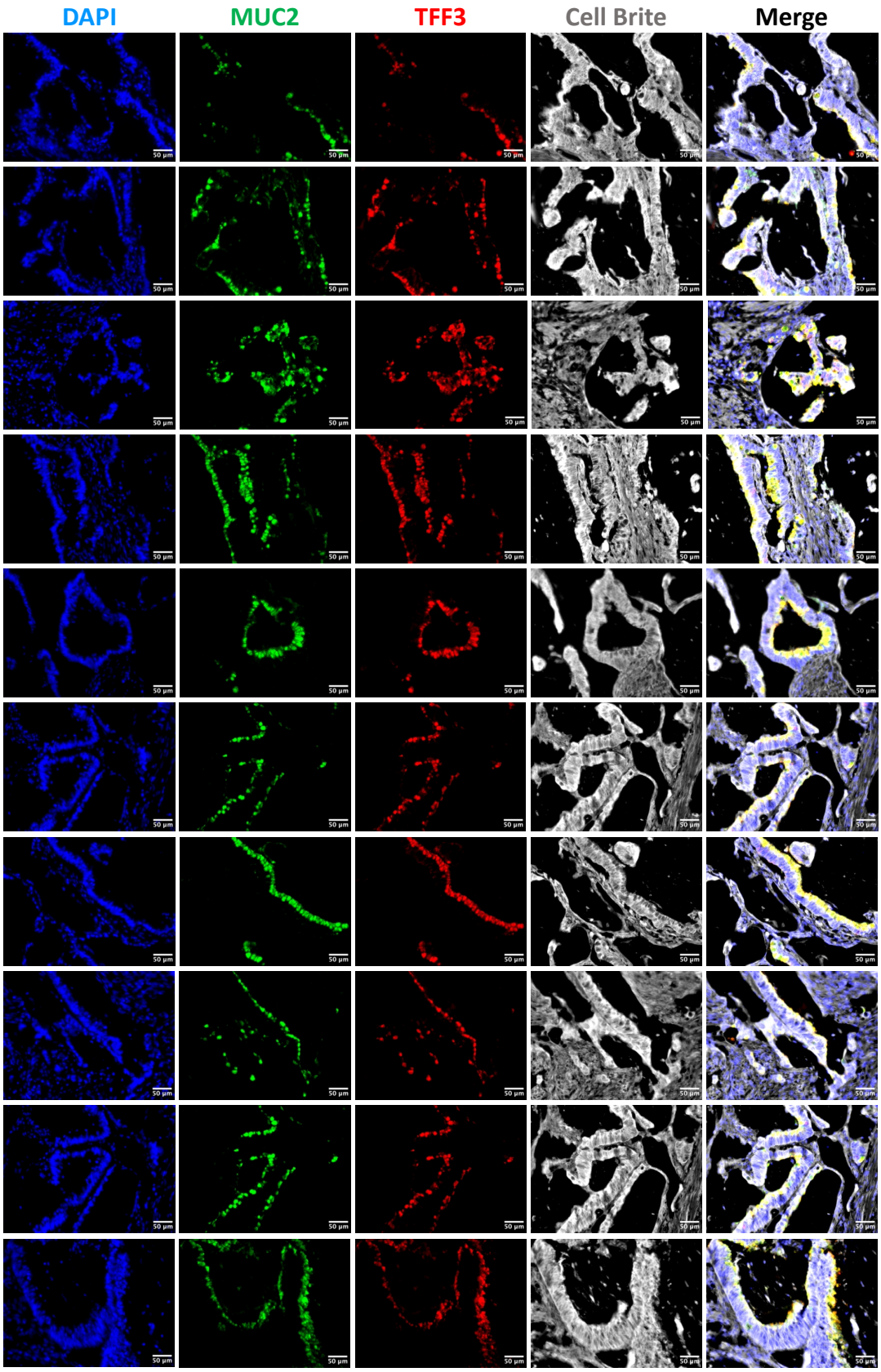
VT11181 (colorectal colon)

B.



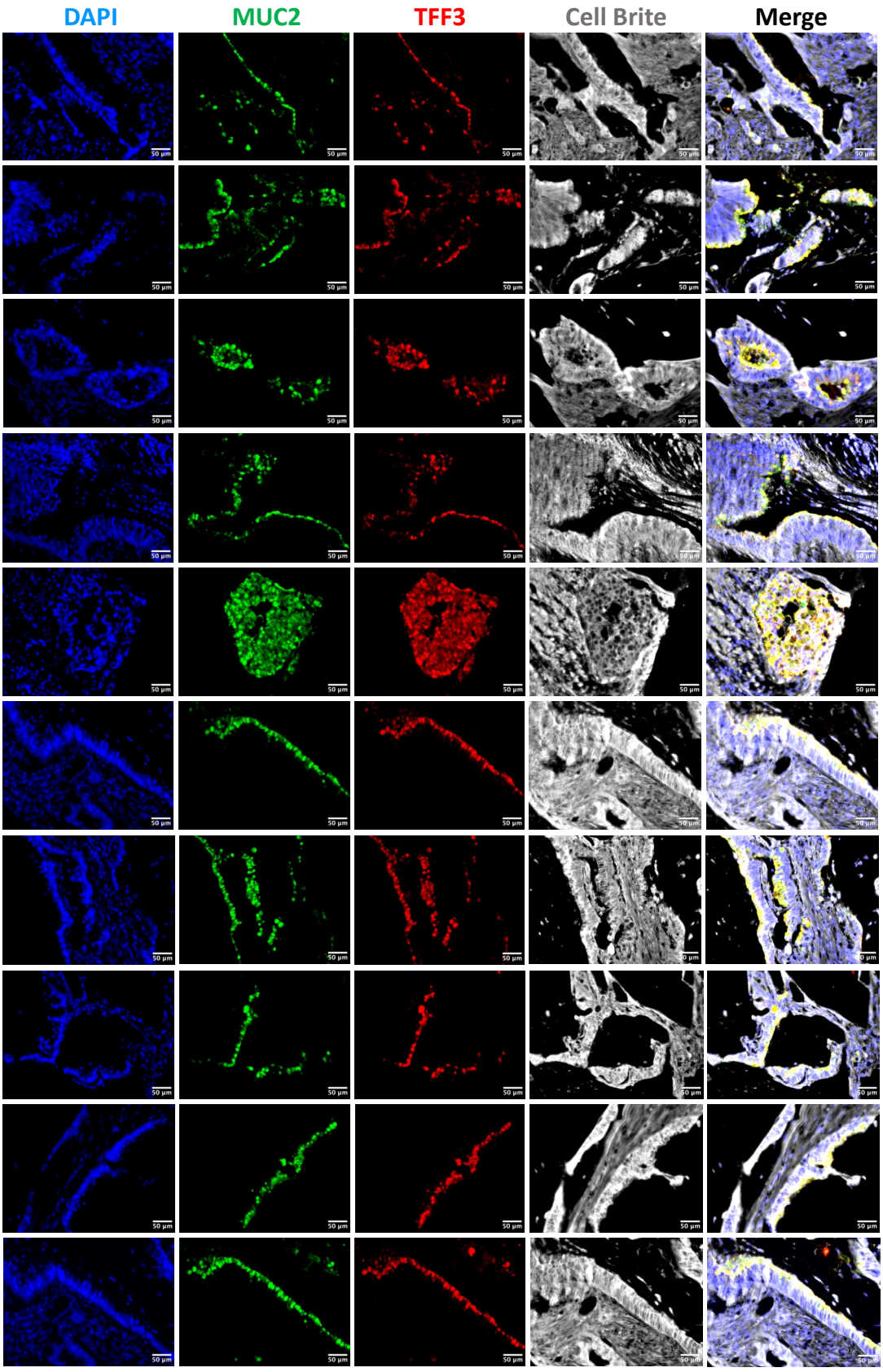
VT11181 (colorectal colon)

B.



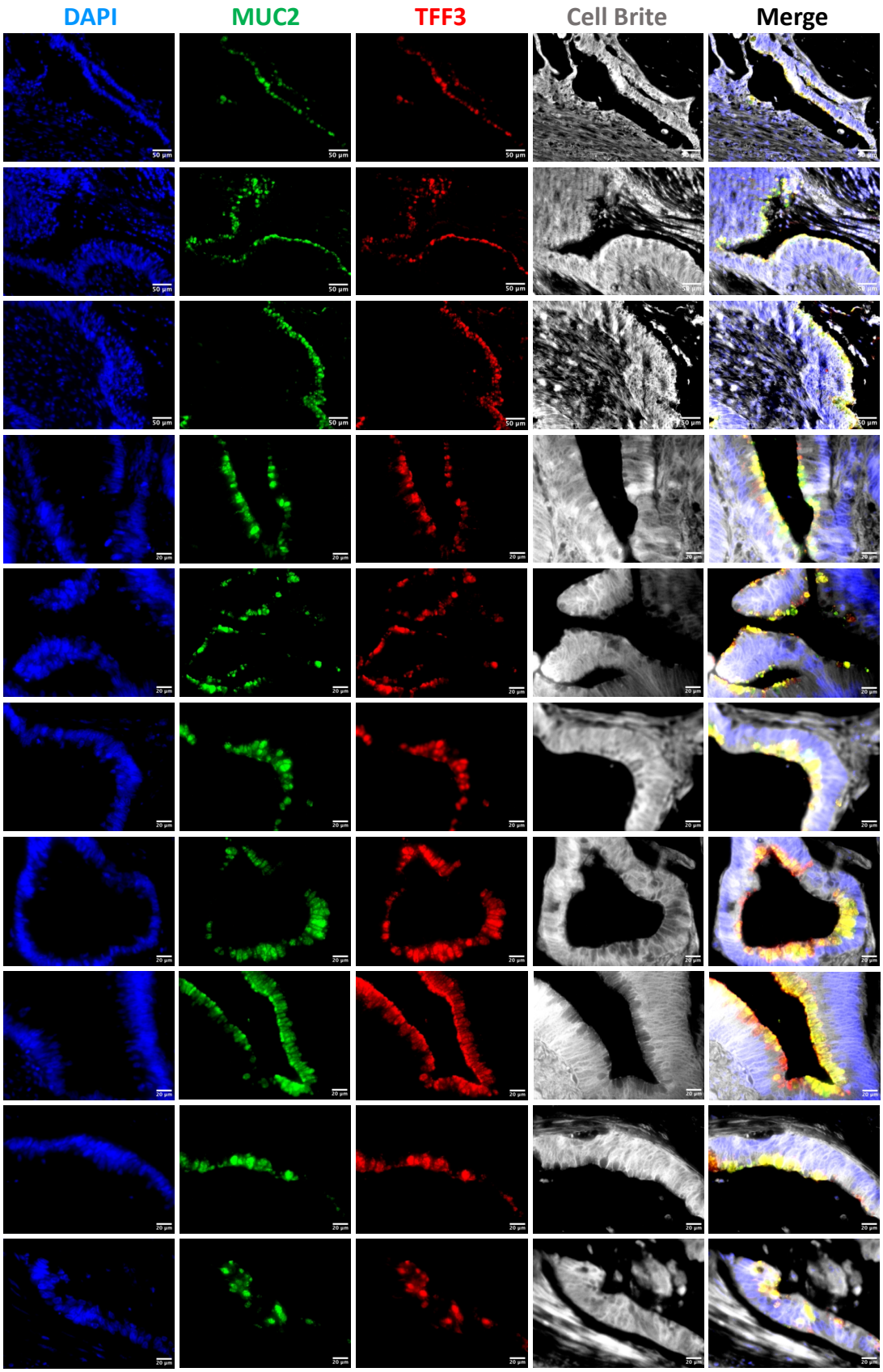
VT11181 (colorectal colon)

B.



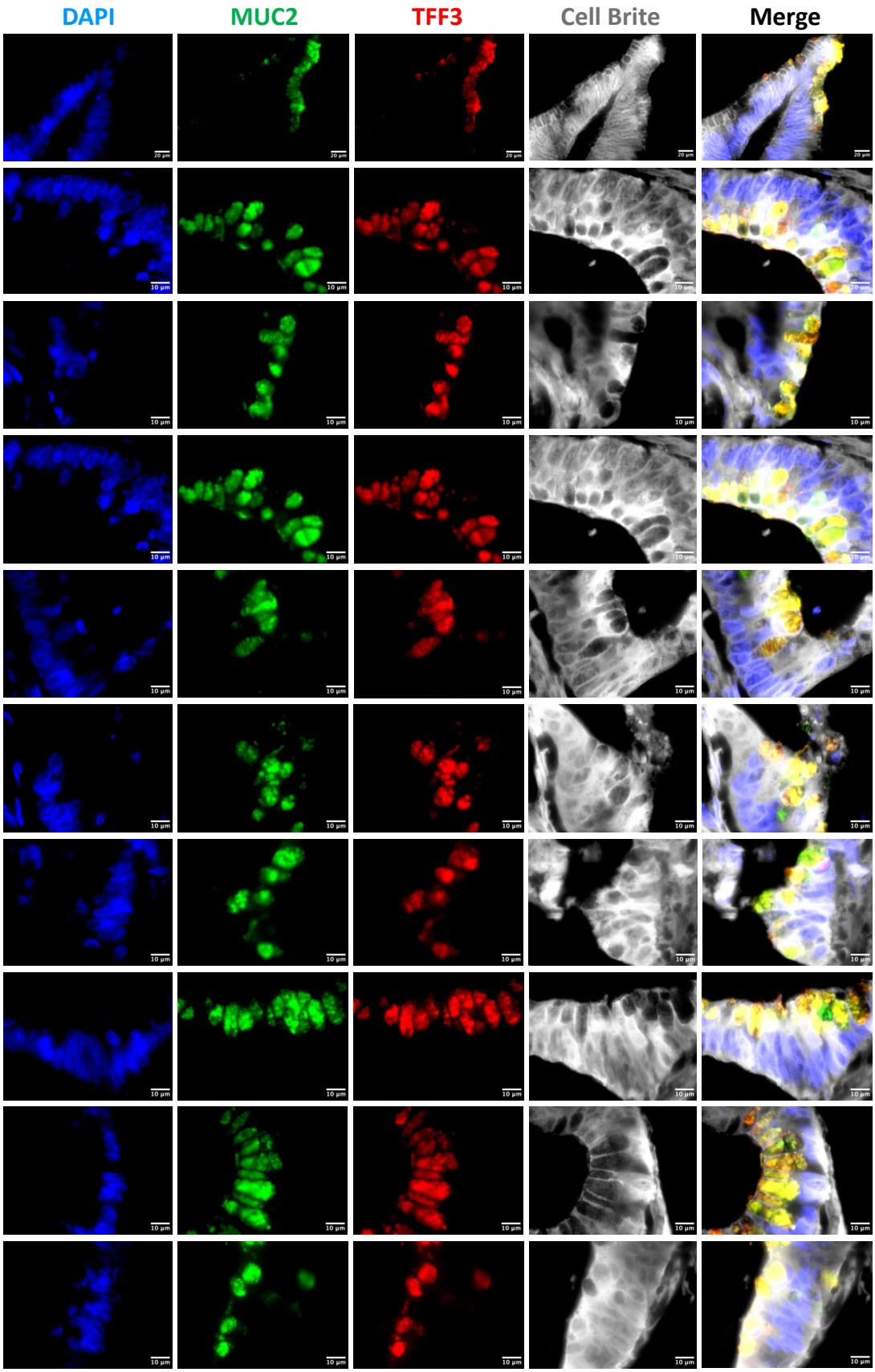
VT11181 (colorectal colon)

B.



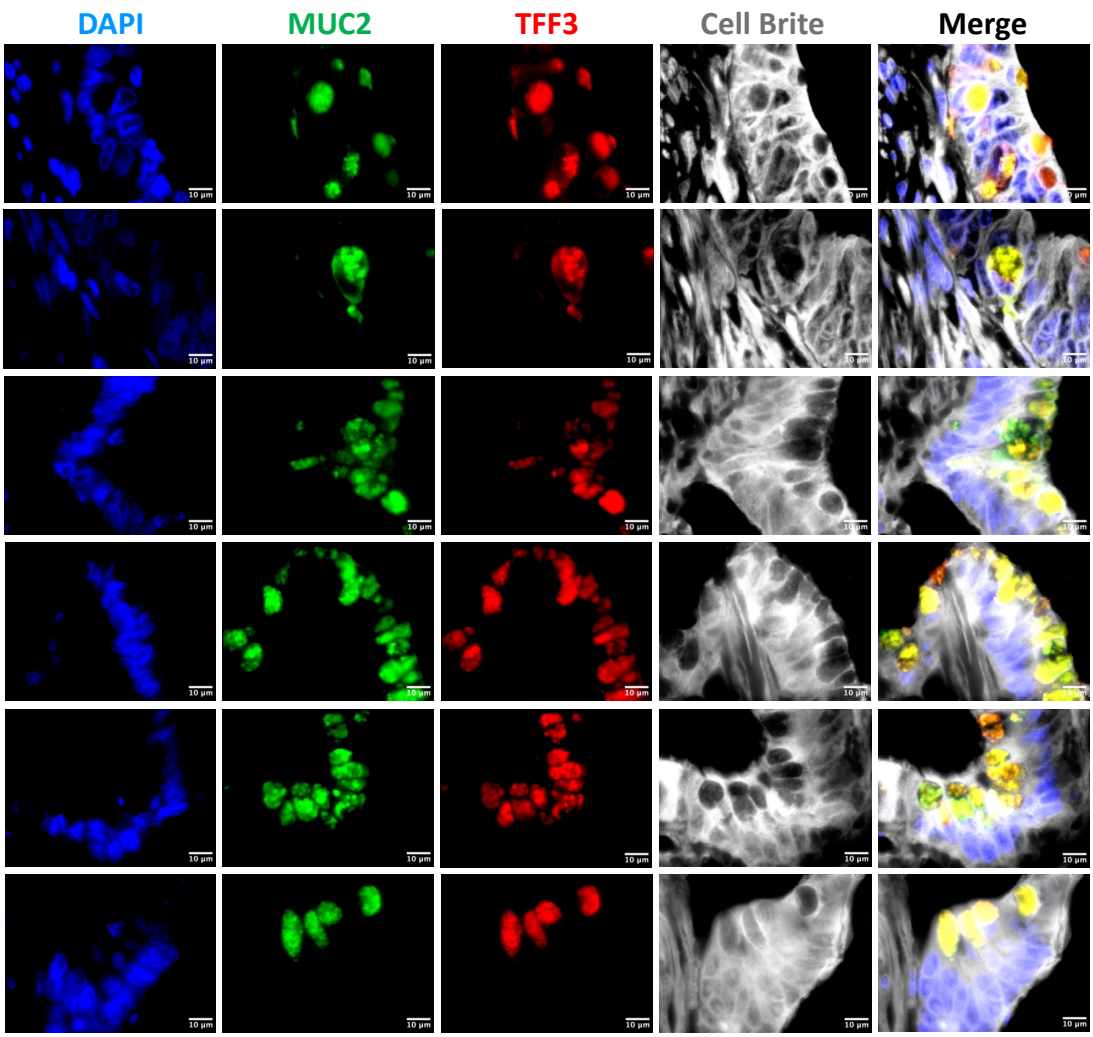
VT11181 (colorectal colon)

B.



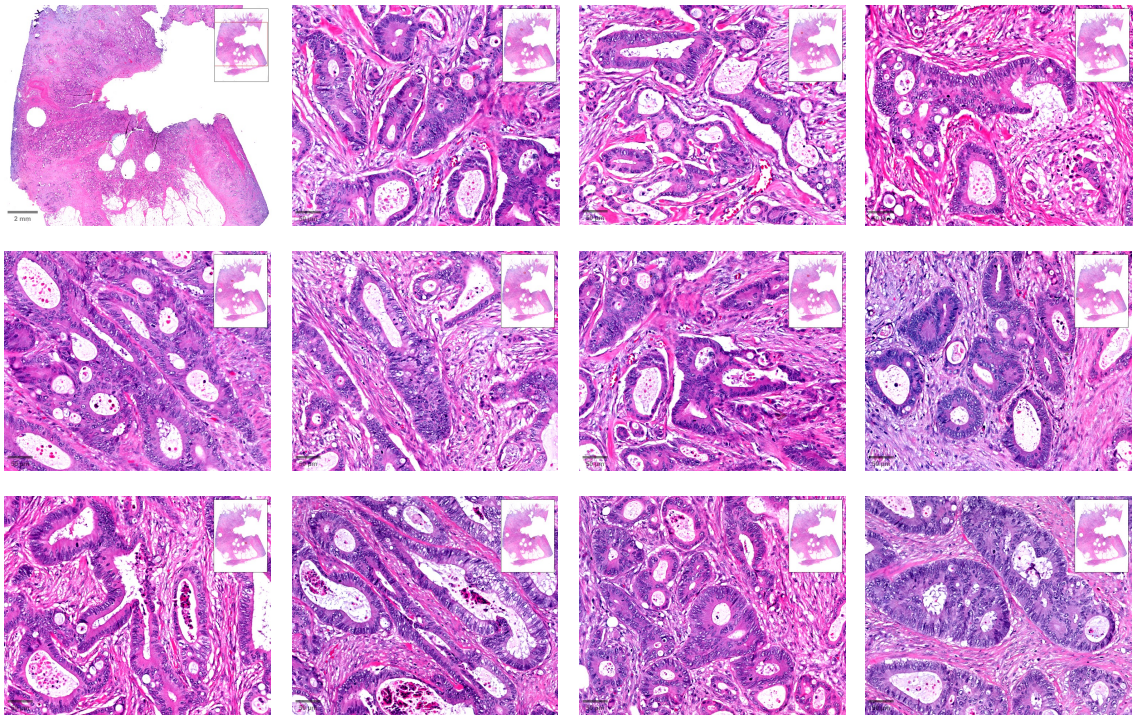
VT11181 (colorectal colon)

B.



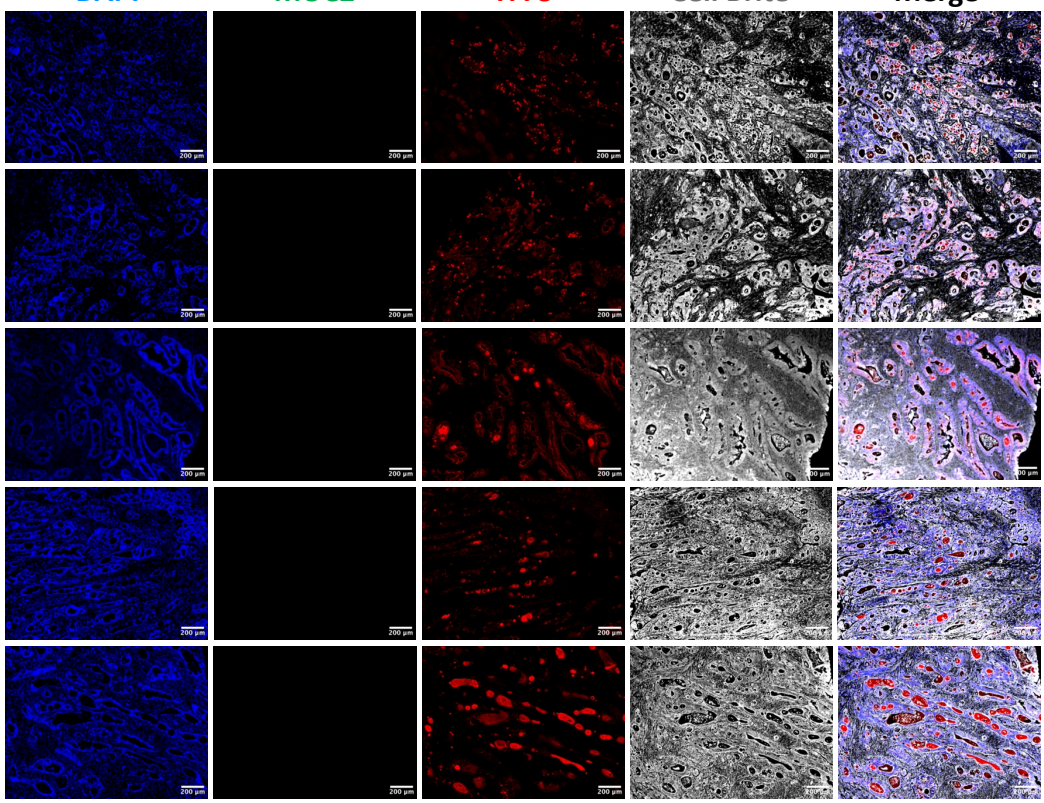
VT11189 (colorectal cancer)

A.



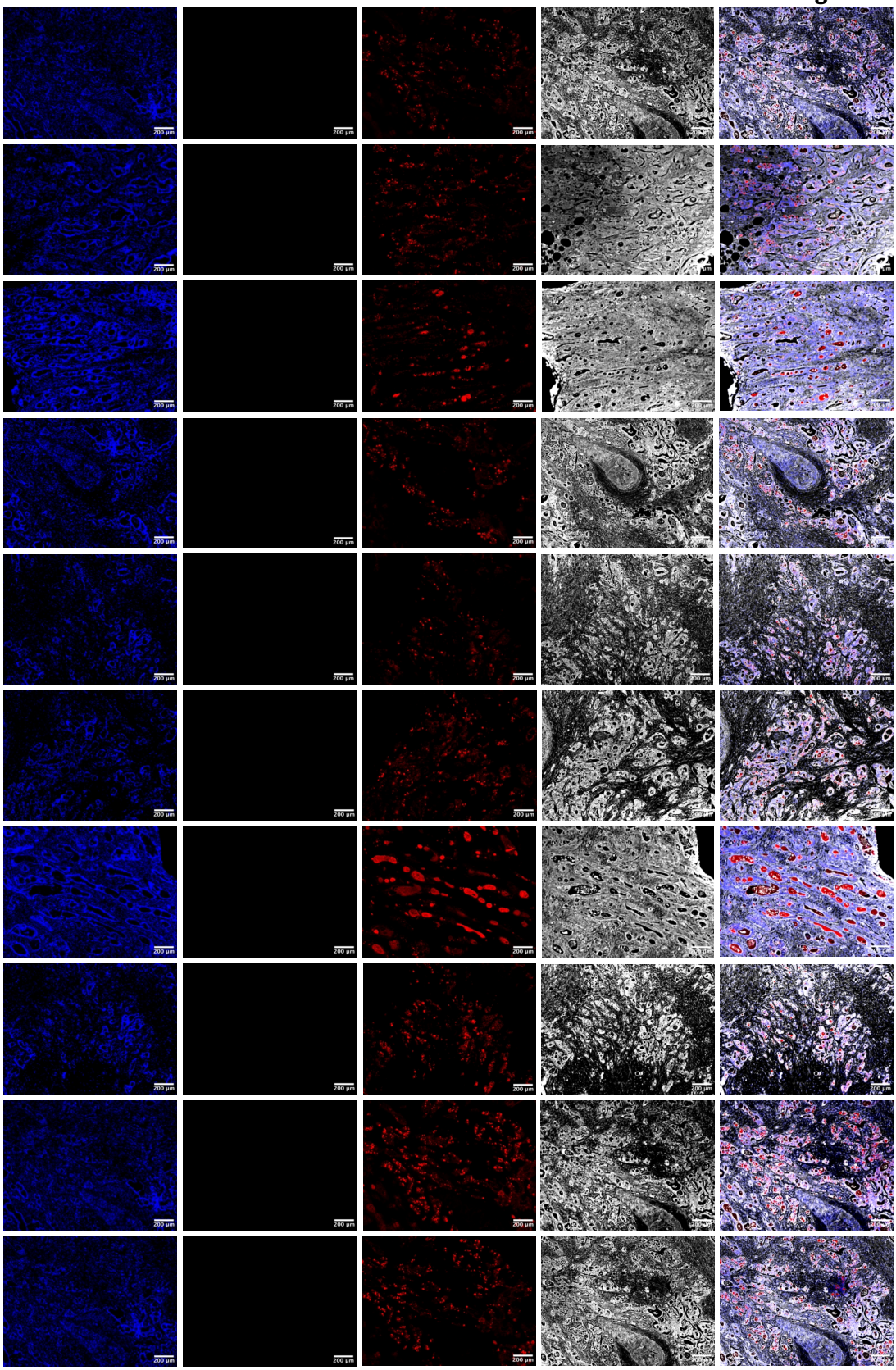
VT11189 (colorectal cancer)

B.



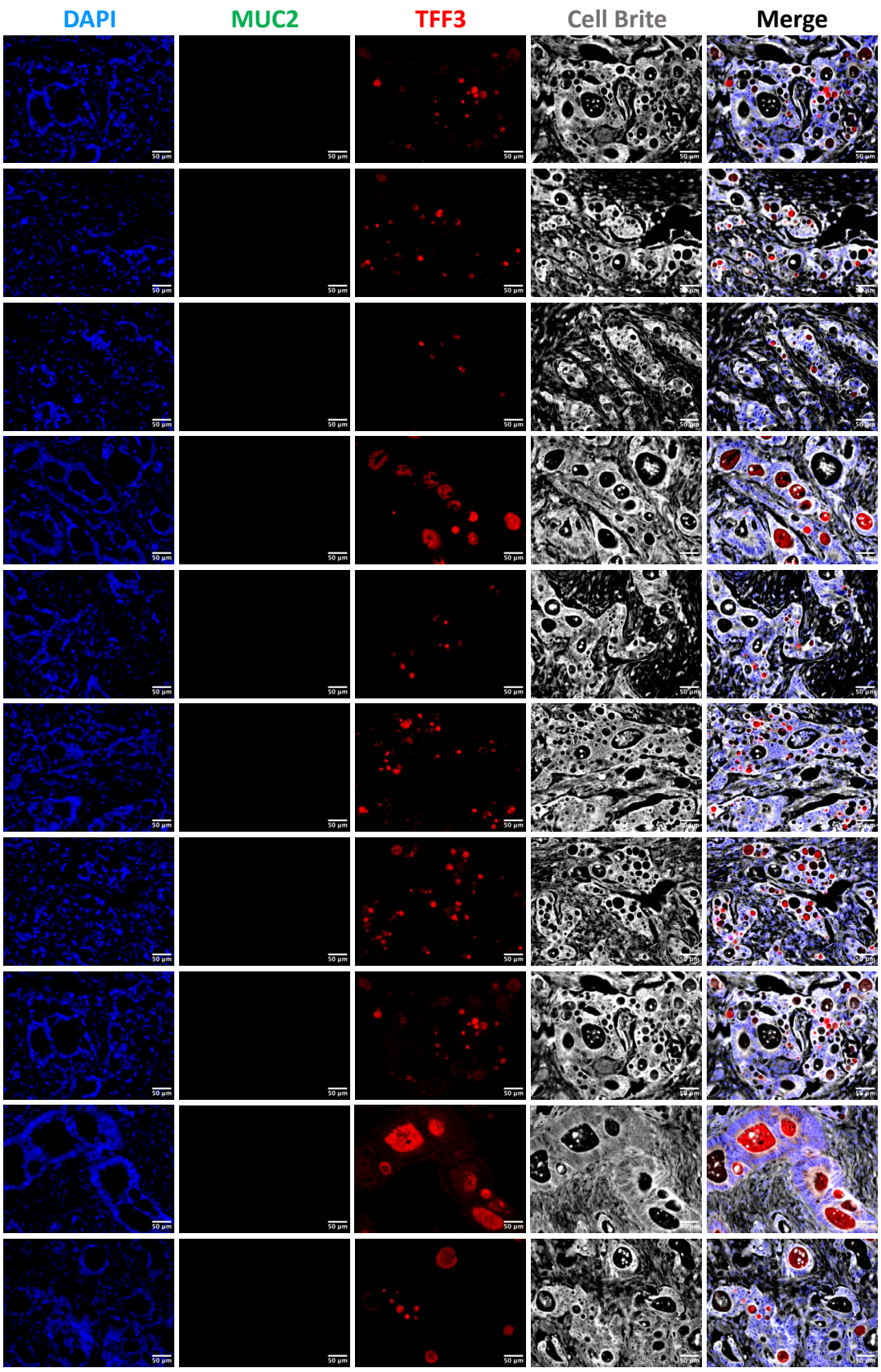
VT11189 (colorectal colon)

B.

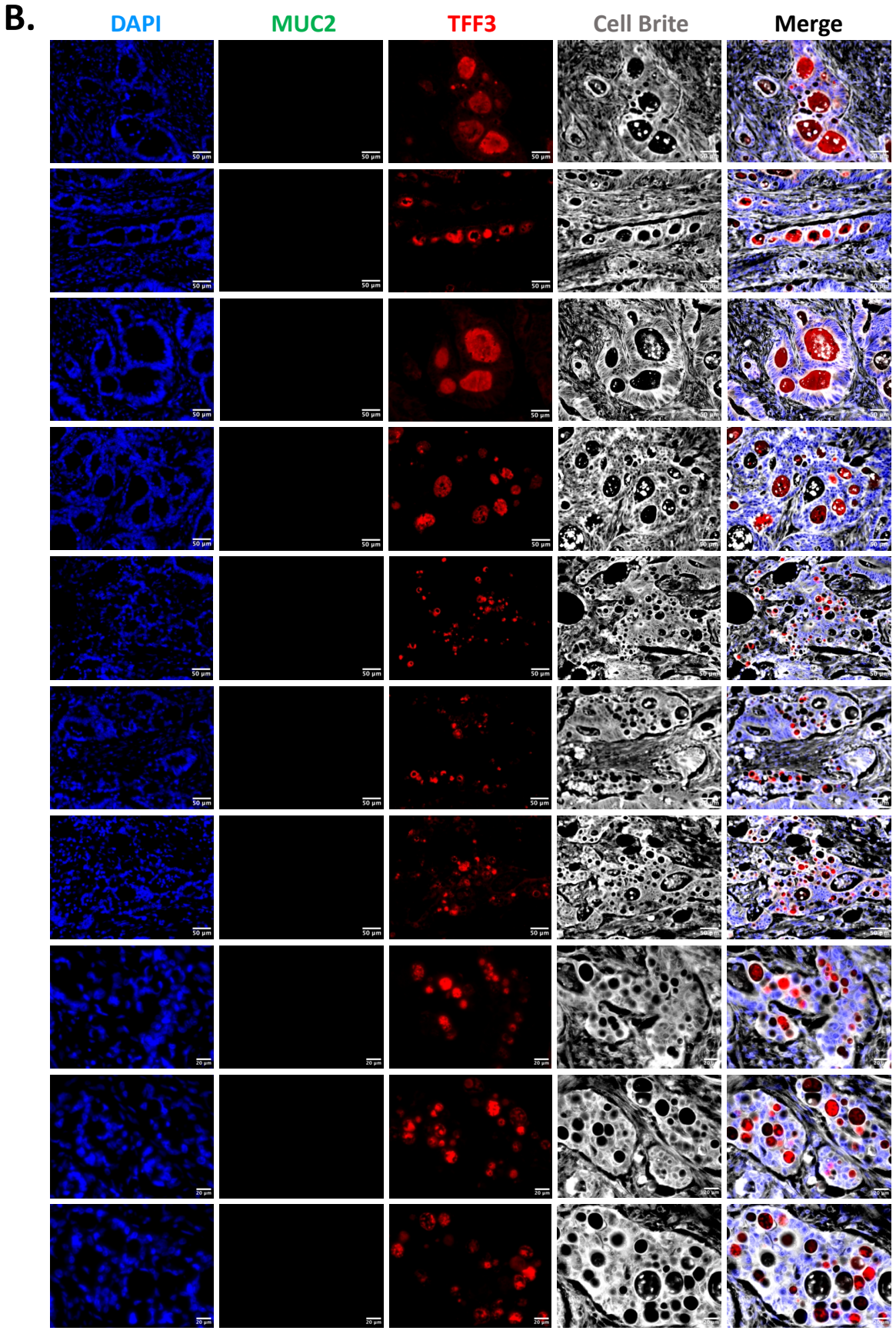


VT11189 (colorectal colon)

B.

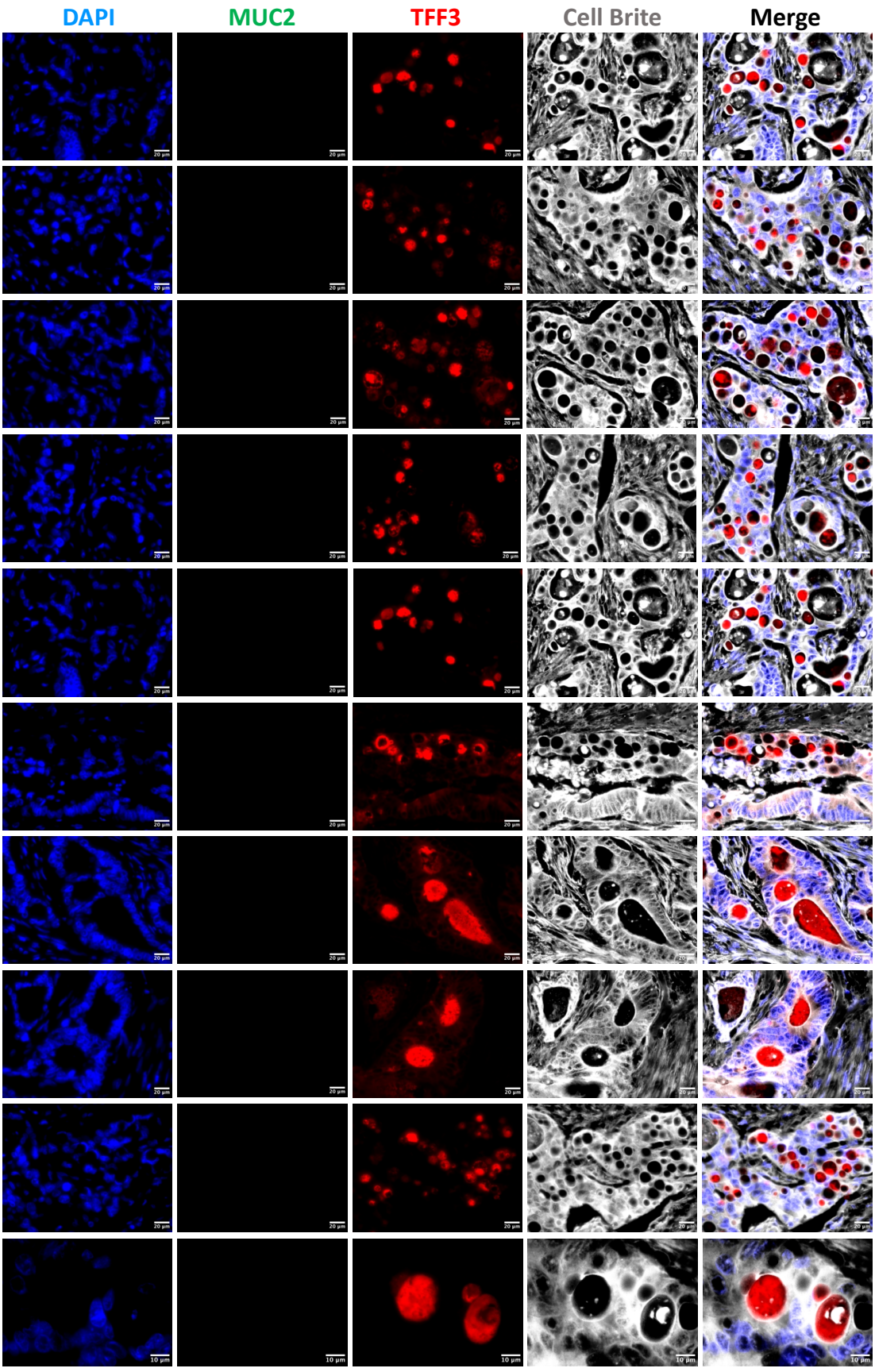


VT11189 (colorectal colon)



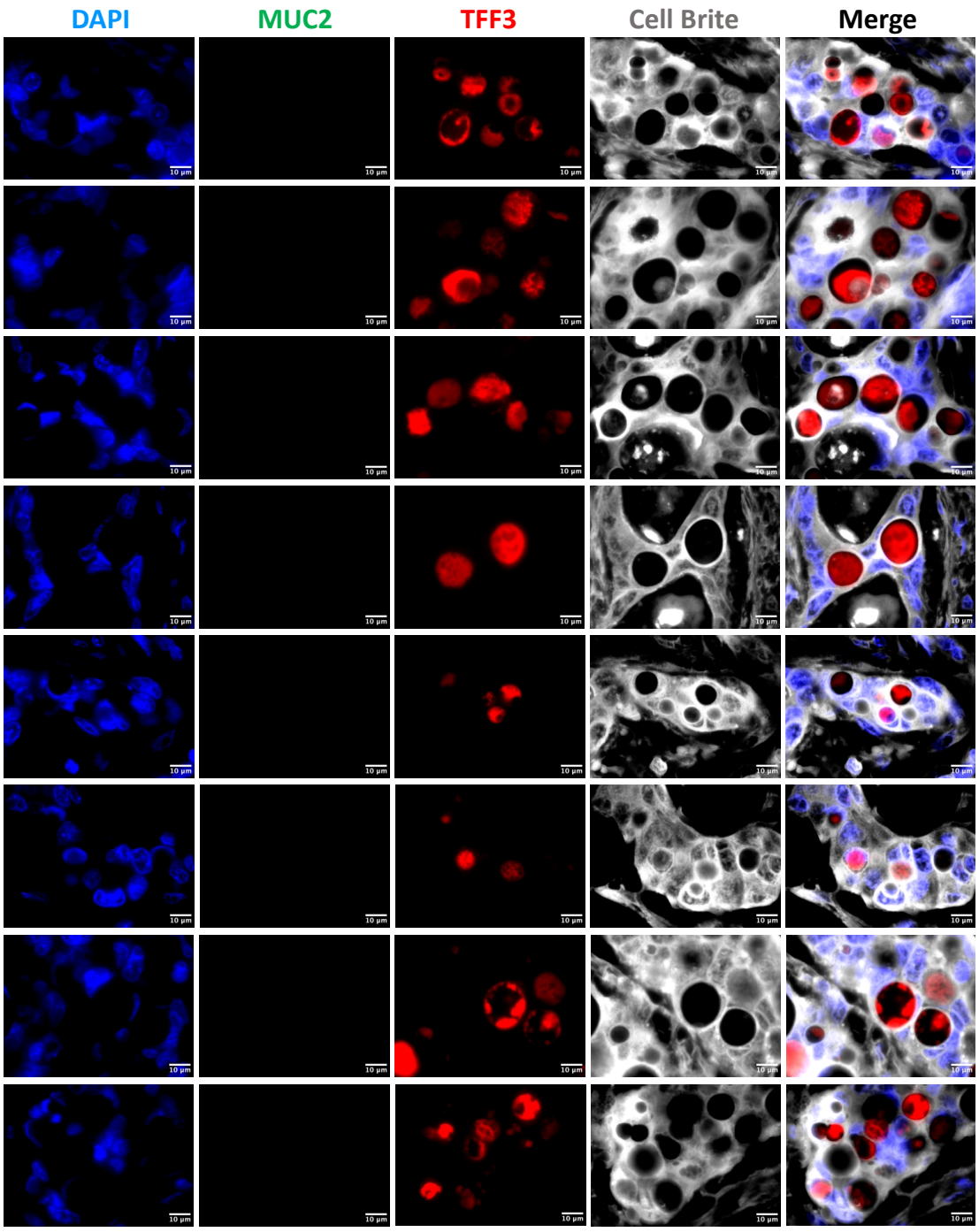
VT11189 (colorectal colon)

B.



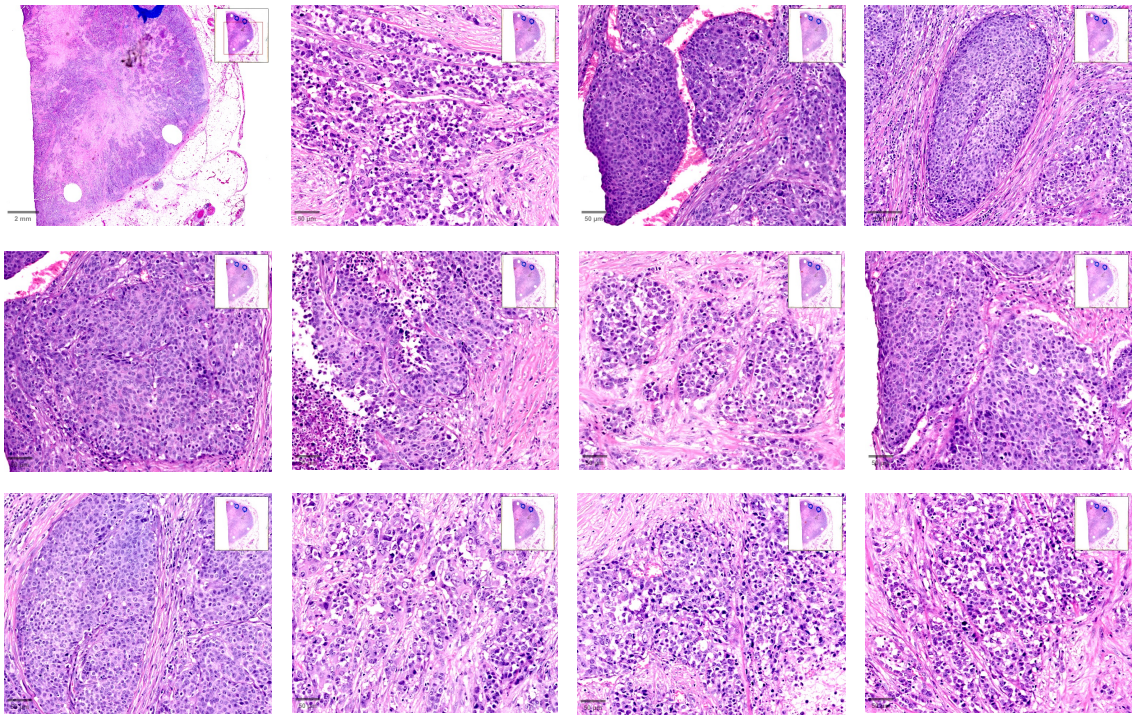
VT11189 (colorectal colon)

B.



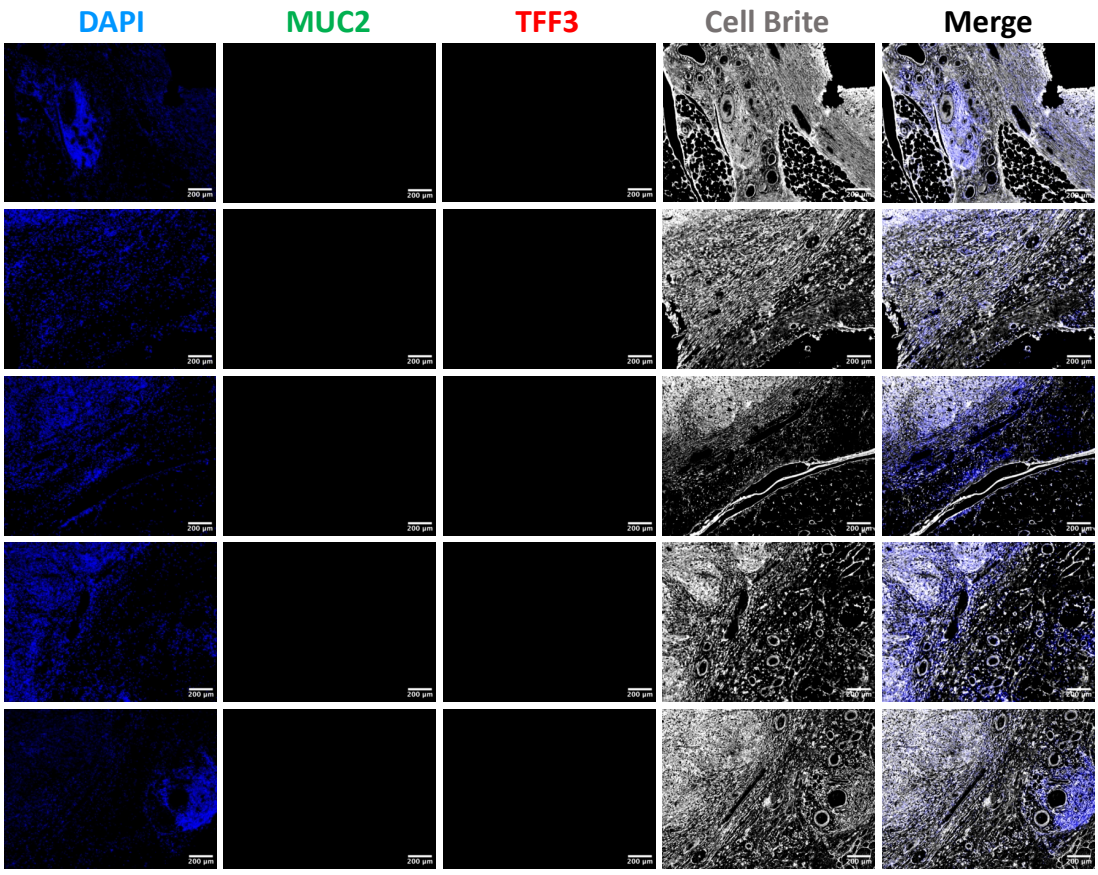
VT11195 (colorectal cancer)

A.



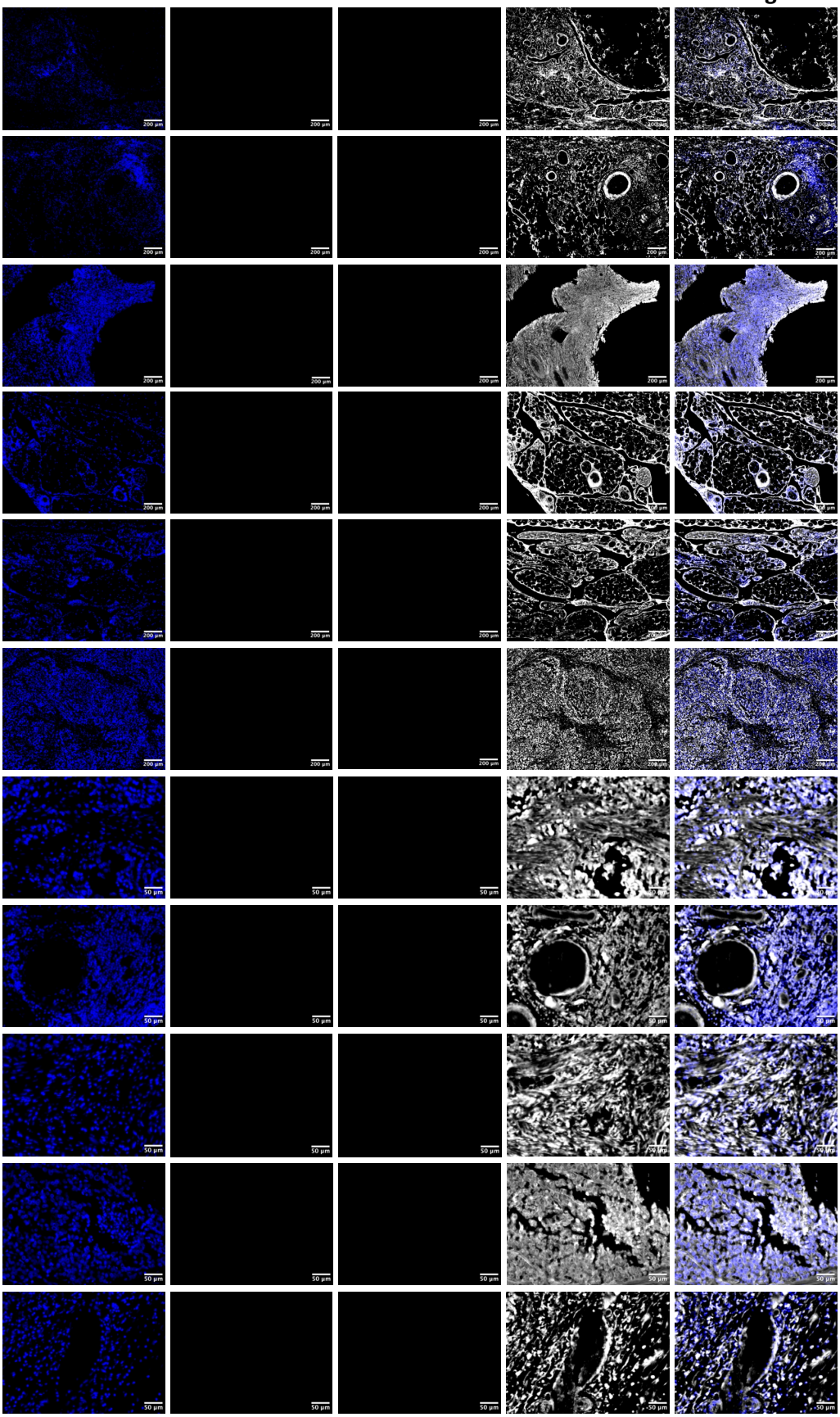
VT11195 (colorectal cancer)

B.



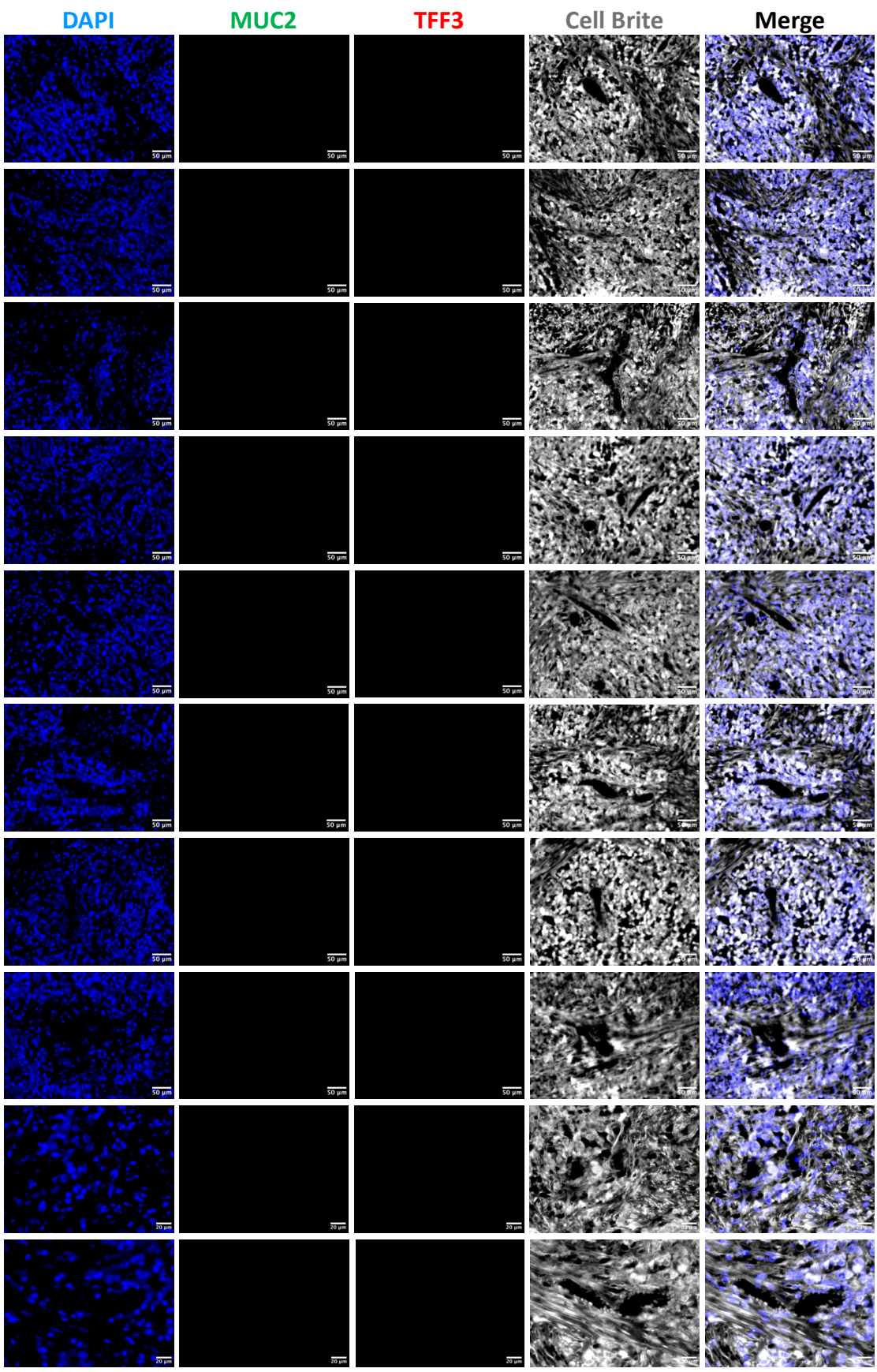
VT11195 (colorectal colon)

B.



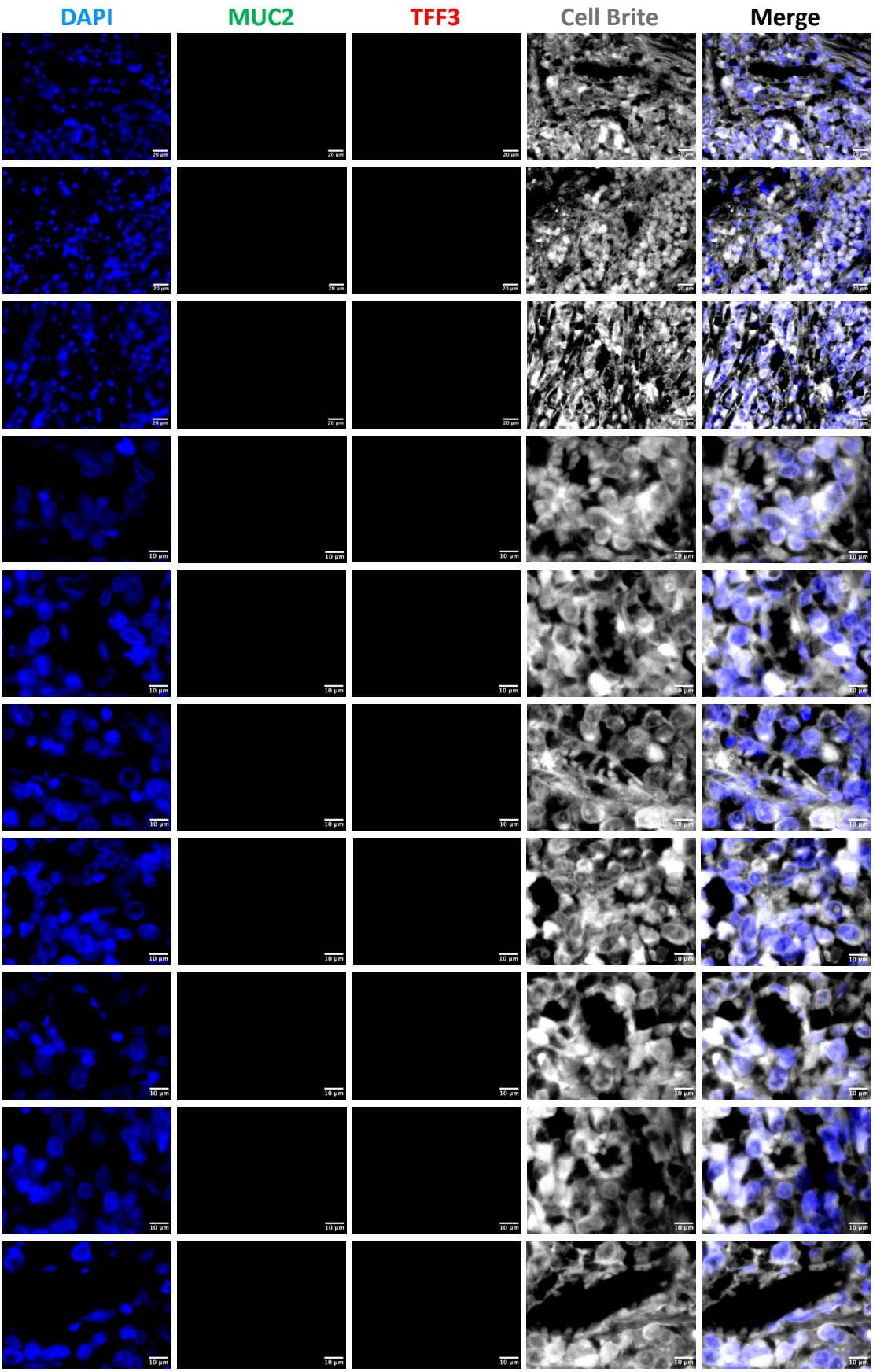
VT11195 (colorectal colon)

B.



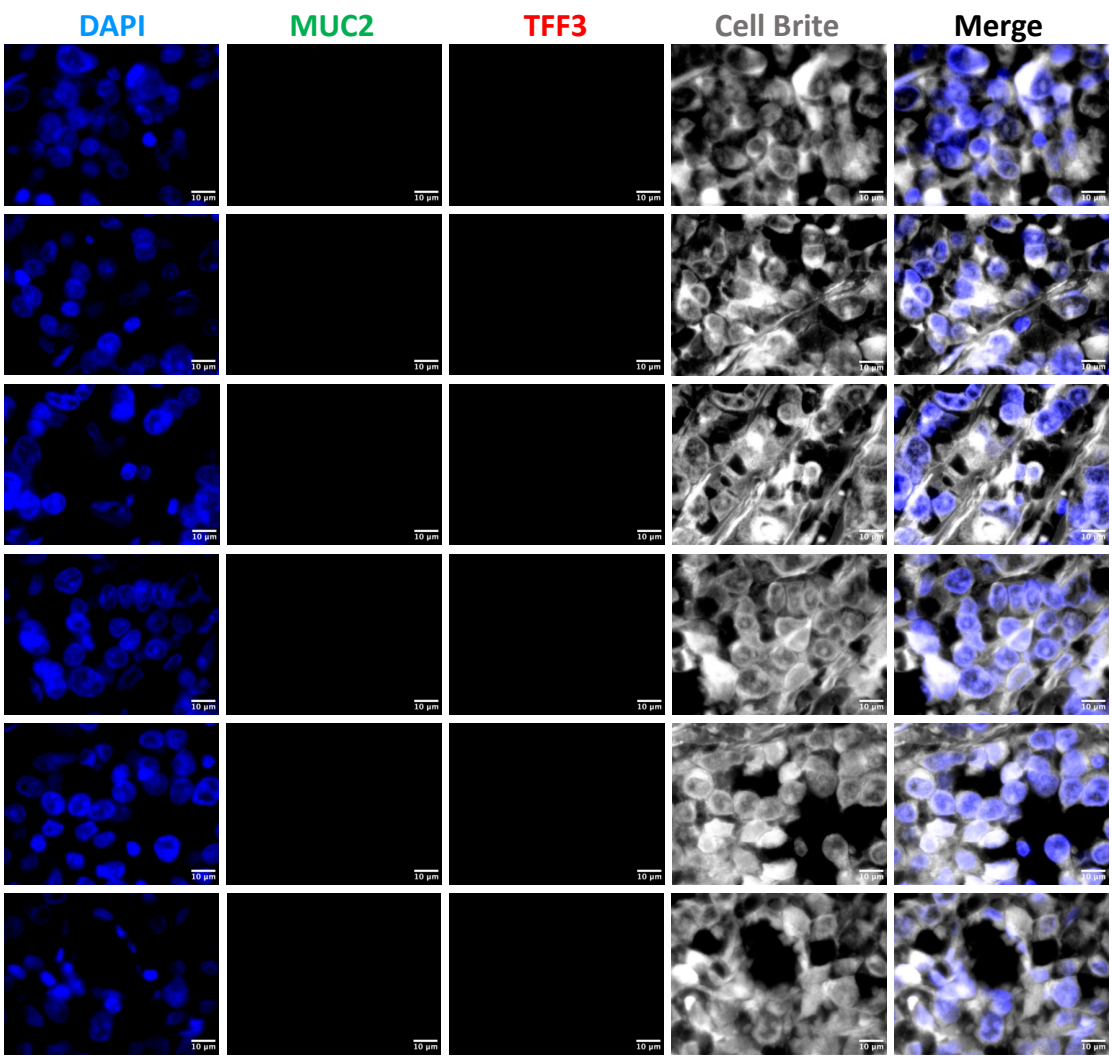
VT11195 (colorectal colon)

B.



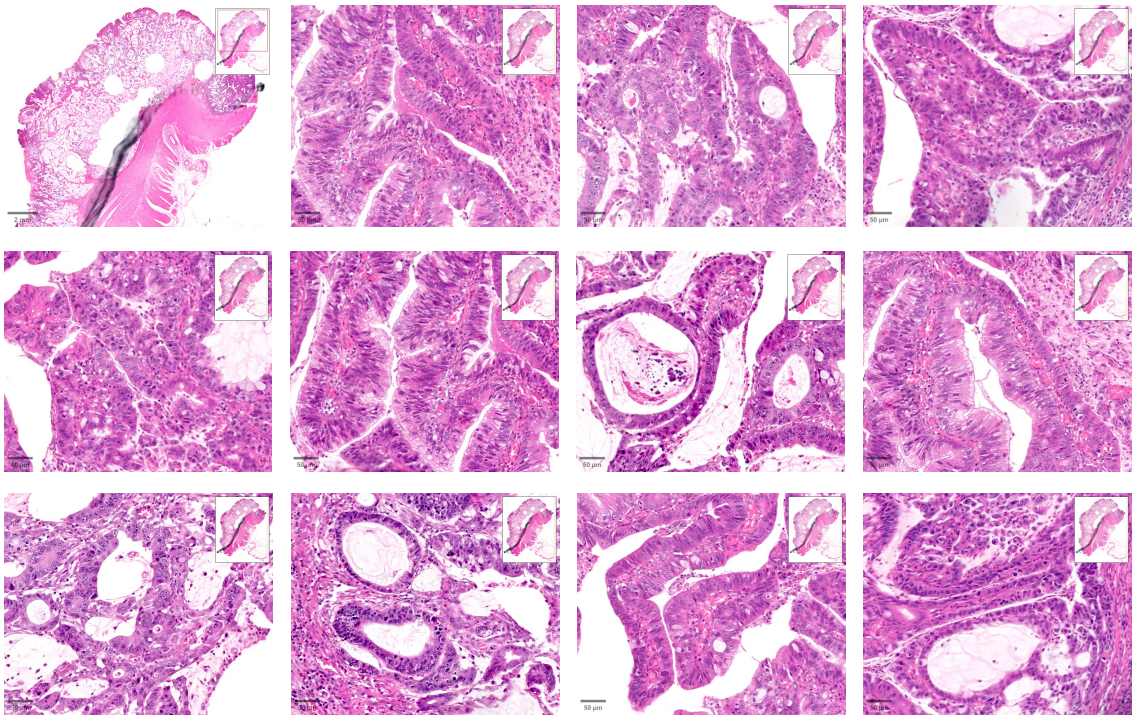
VT11195 (colorectal colon)

B.



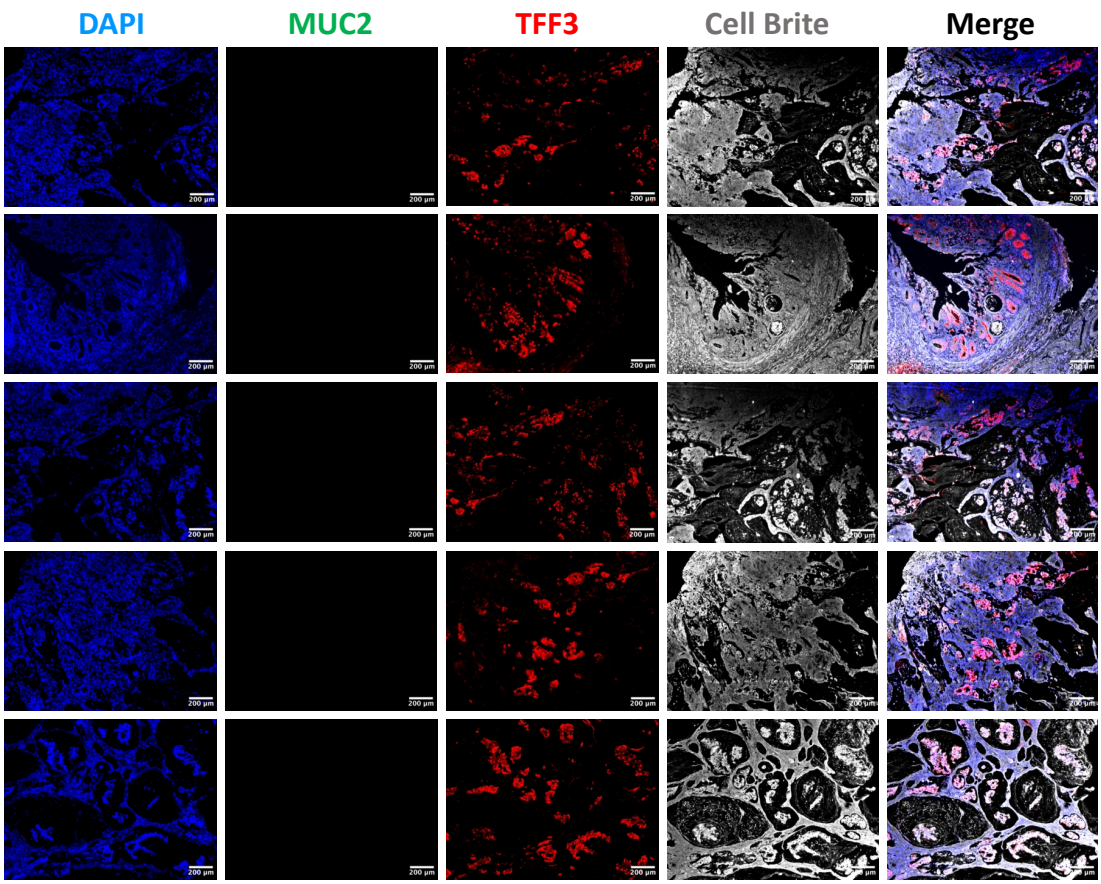
VT11197 (colorectal cancer)

A.



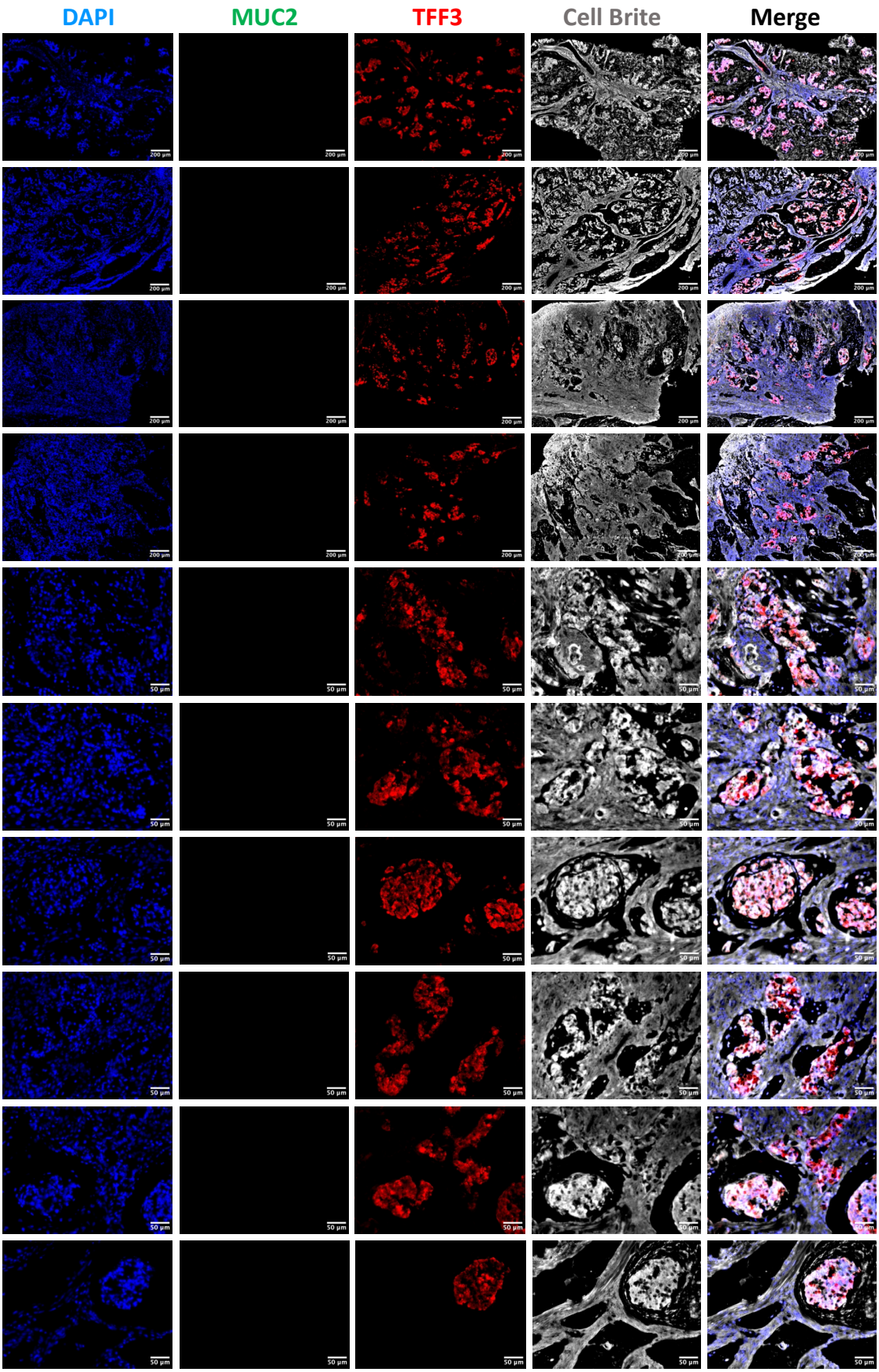
VT11197 (colorectal cancer)

B.



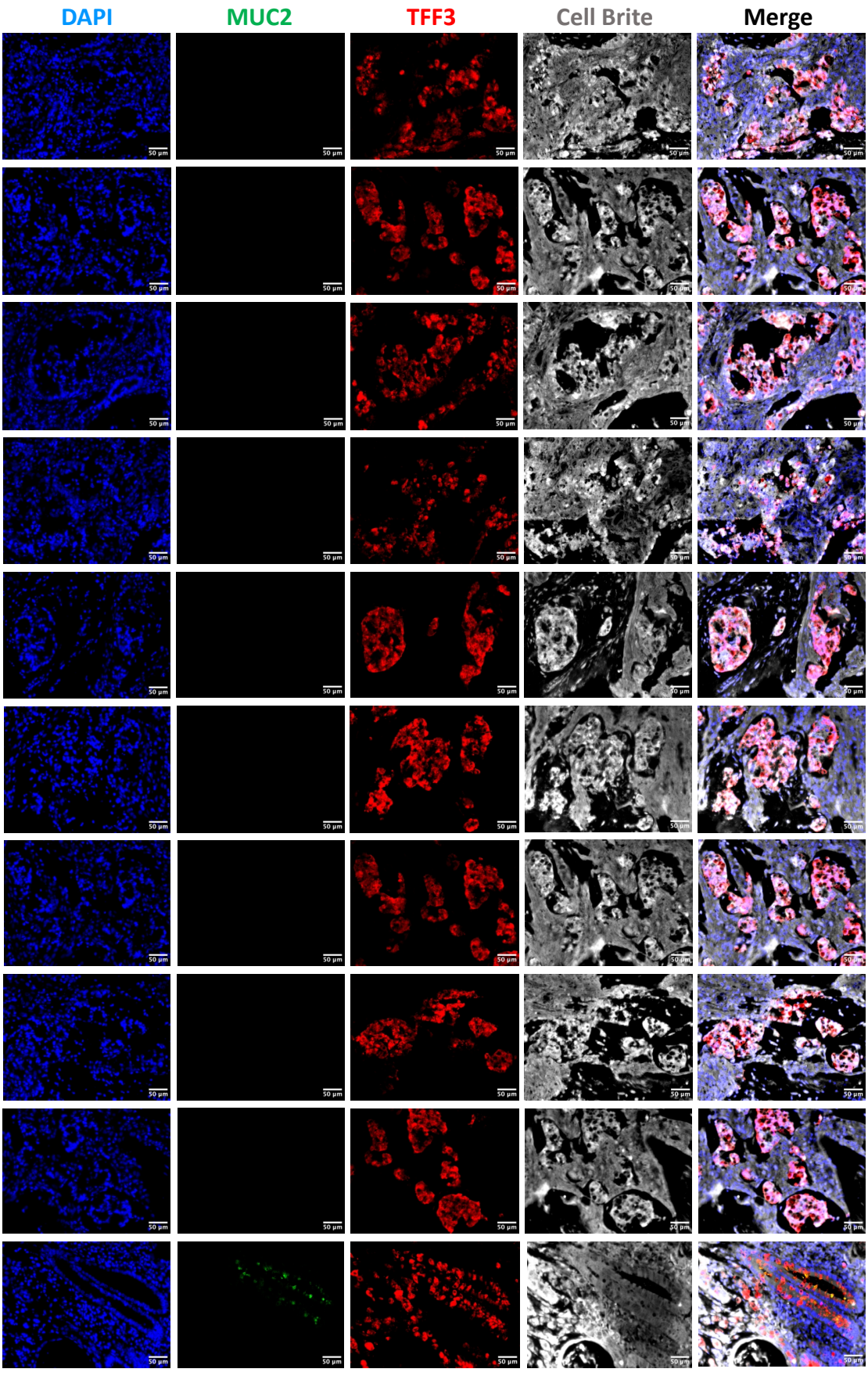
VT11197 (colorectal colon)

B.



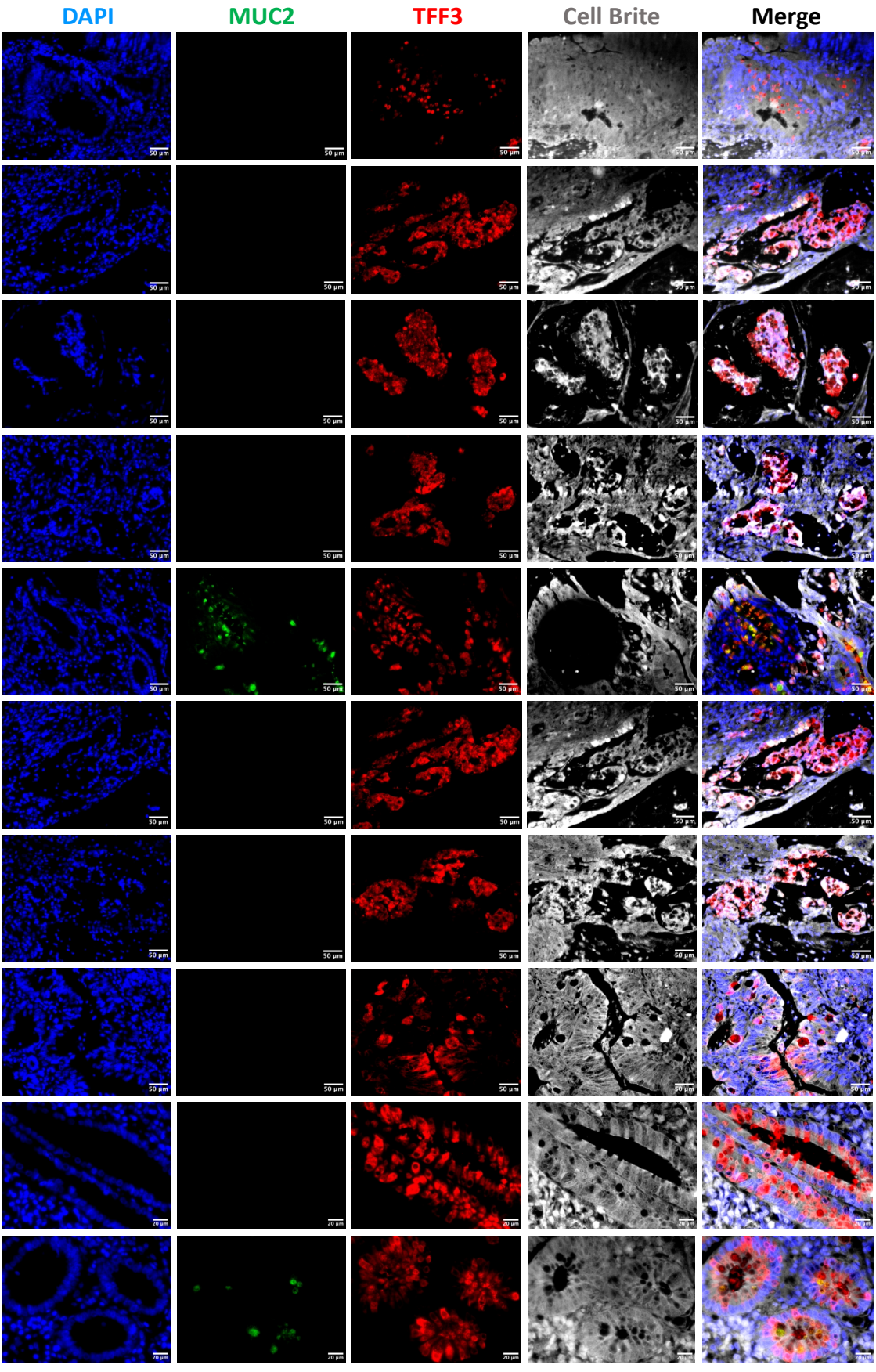
VT11197 (colorectal colon)

B.



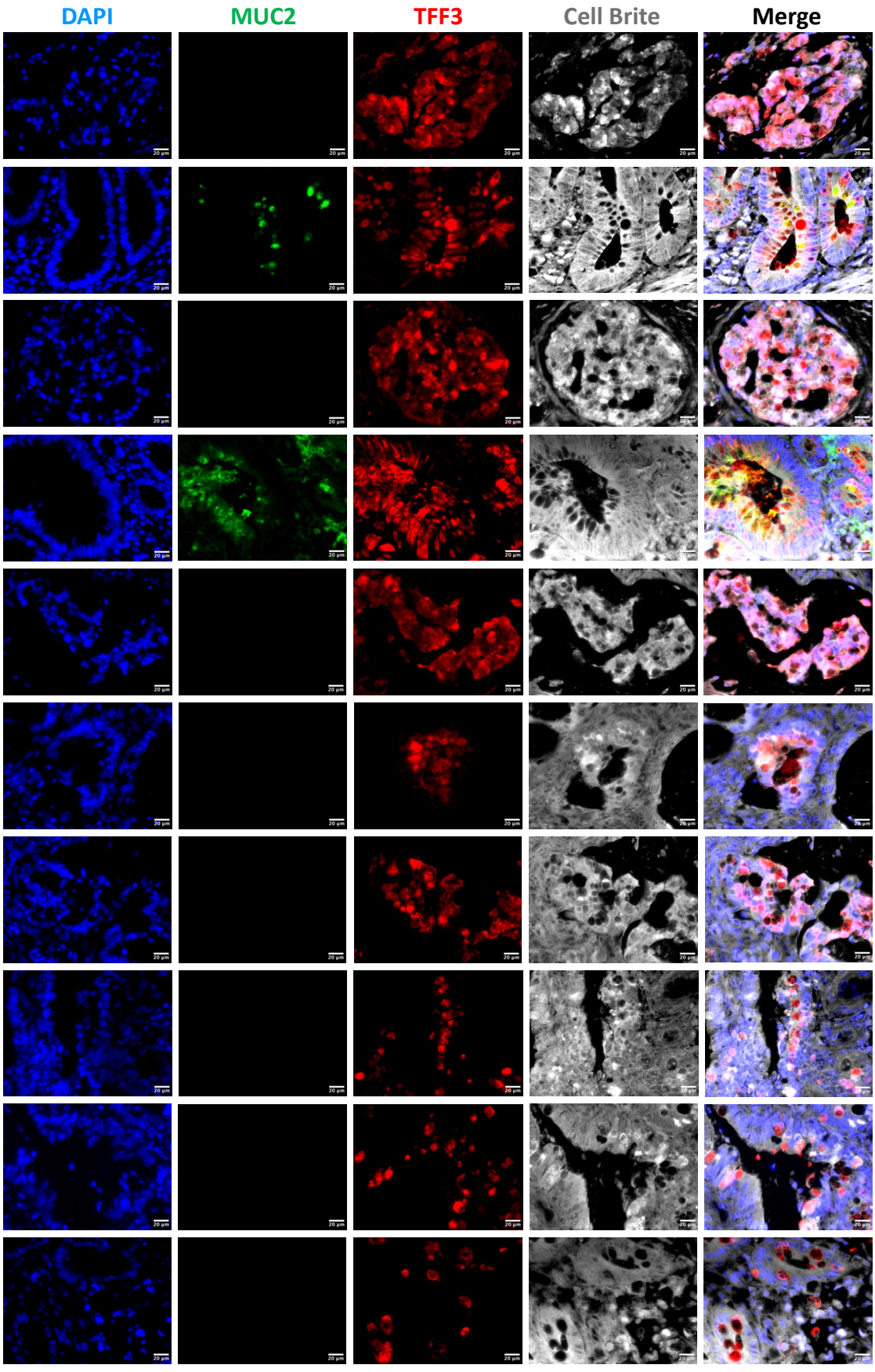
VT11197 (colorectal colon)

B.



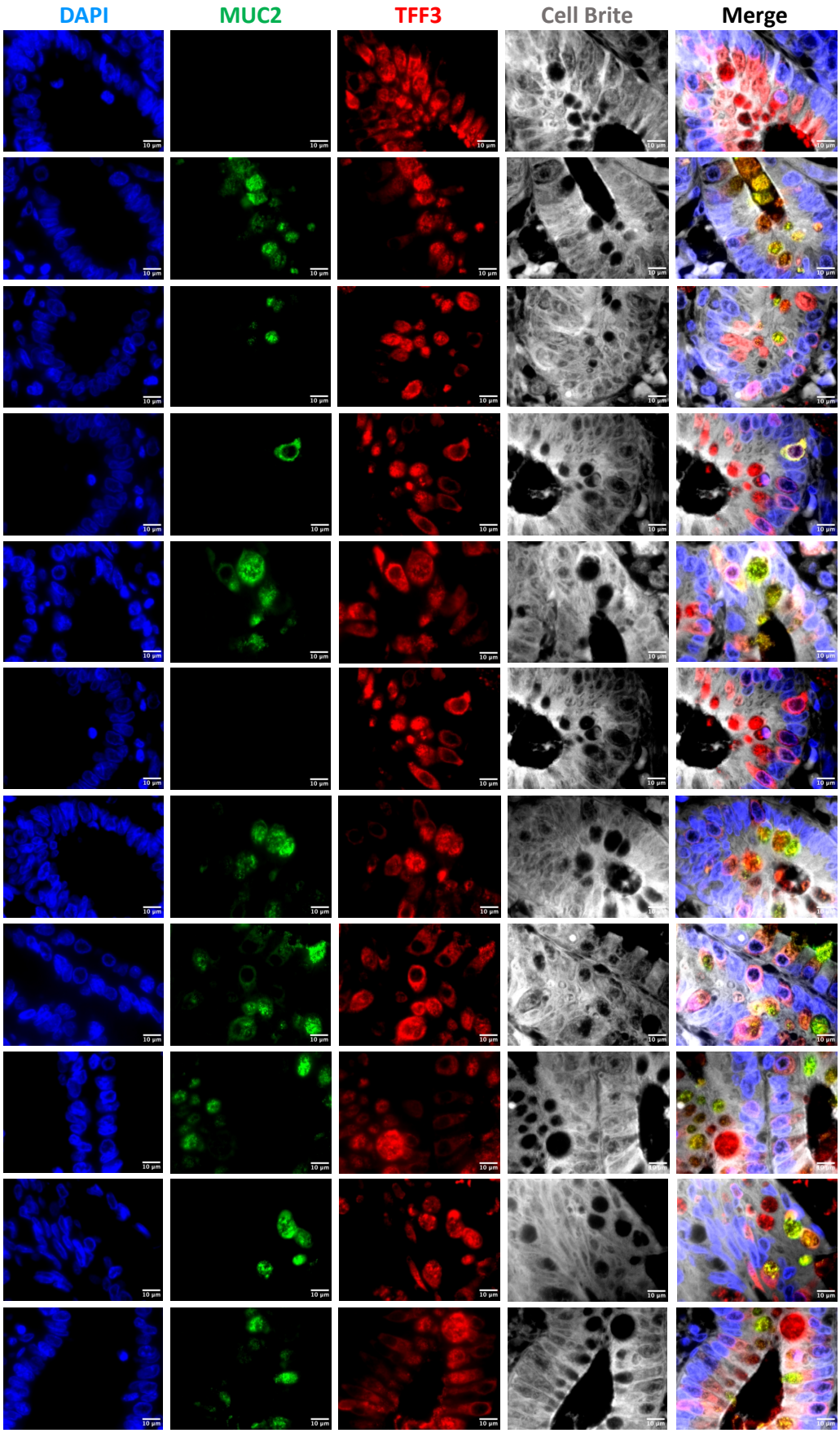
VT11197 (colorectal colon)

B.



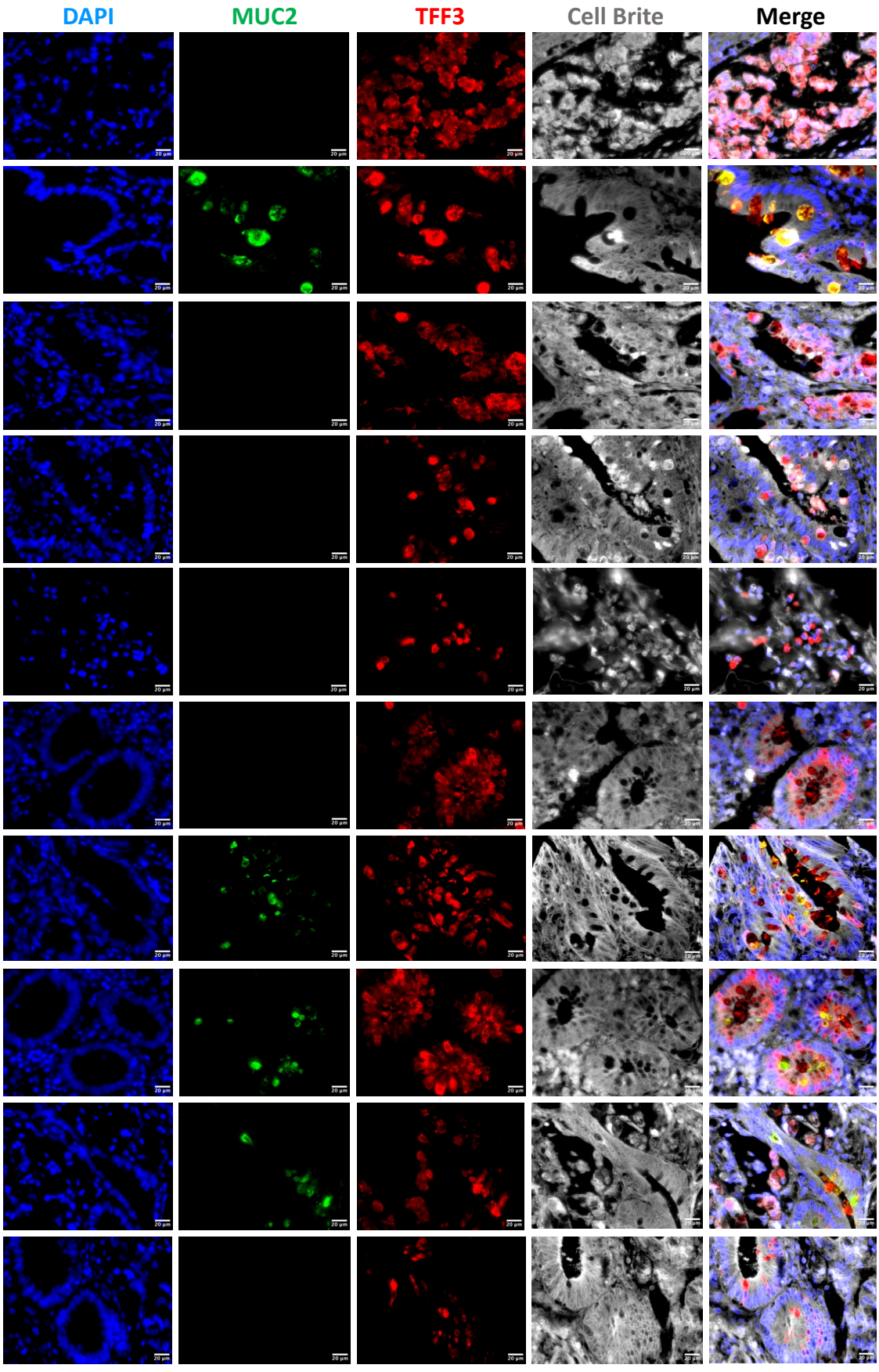
VT11197 (colorectal colon)

B.



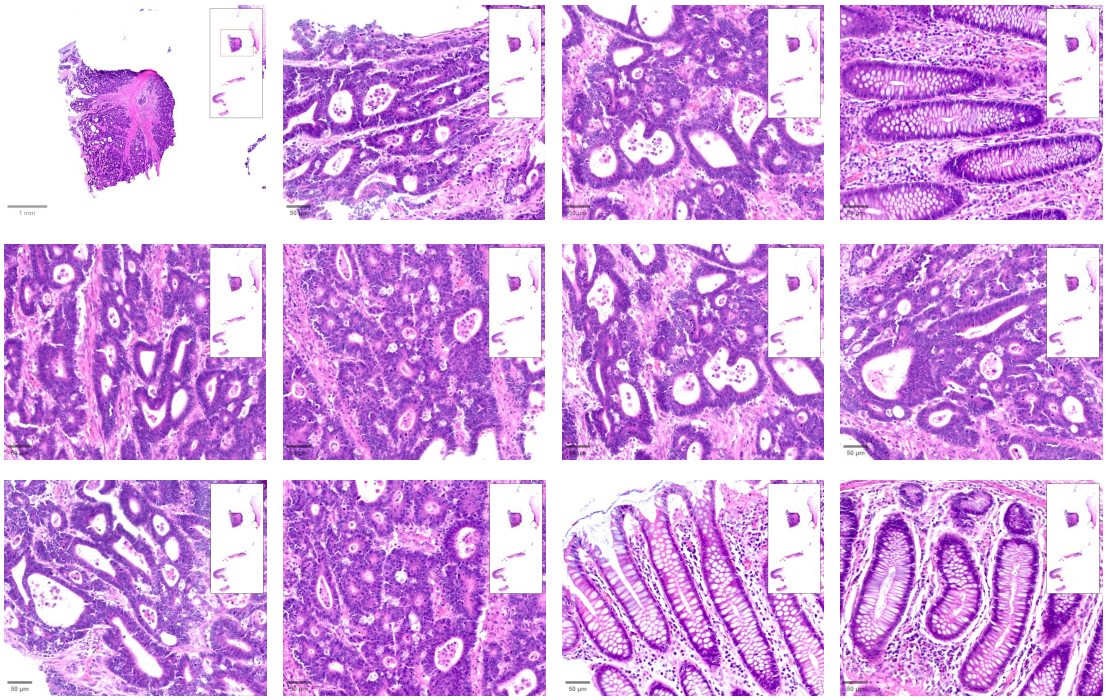
VT11197 (colorectal colon)

B.



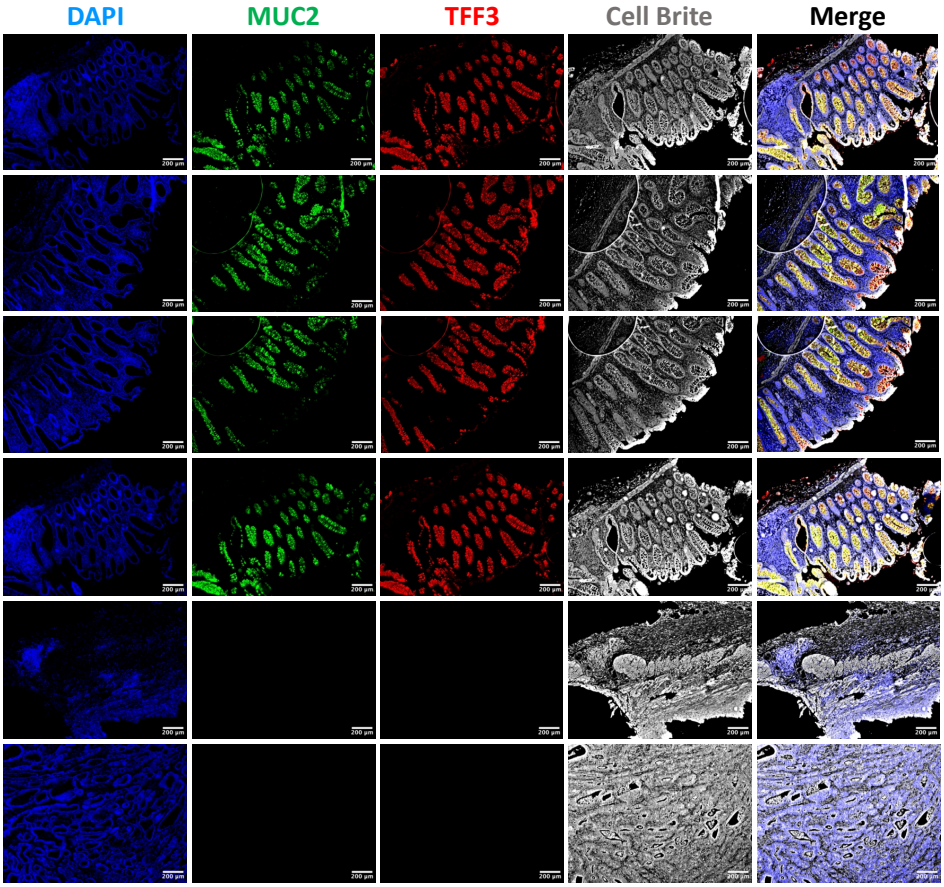
VT11276 (colorectal cancer)

A.



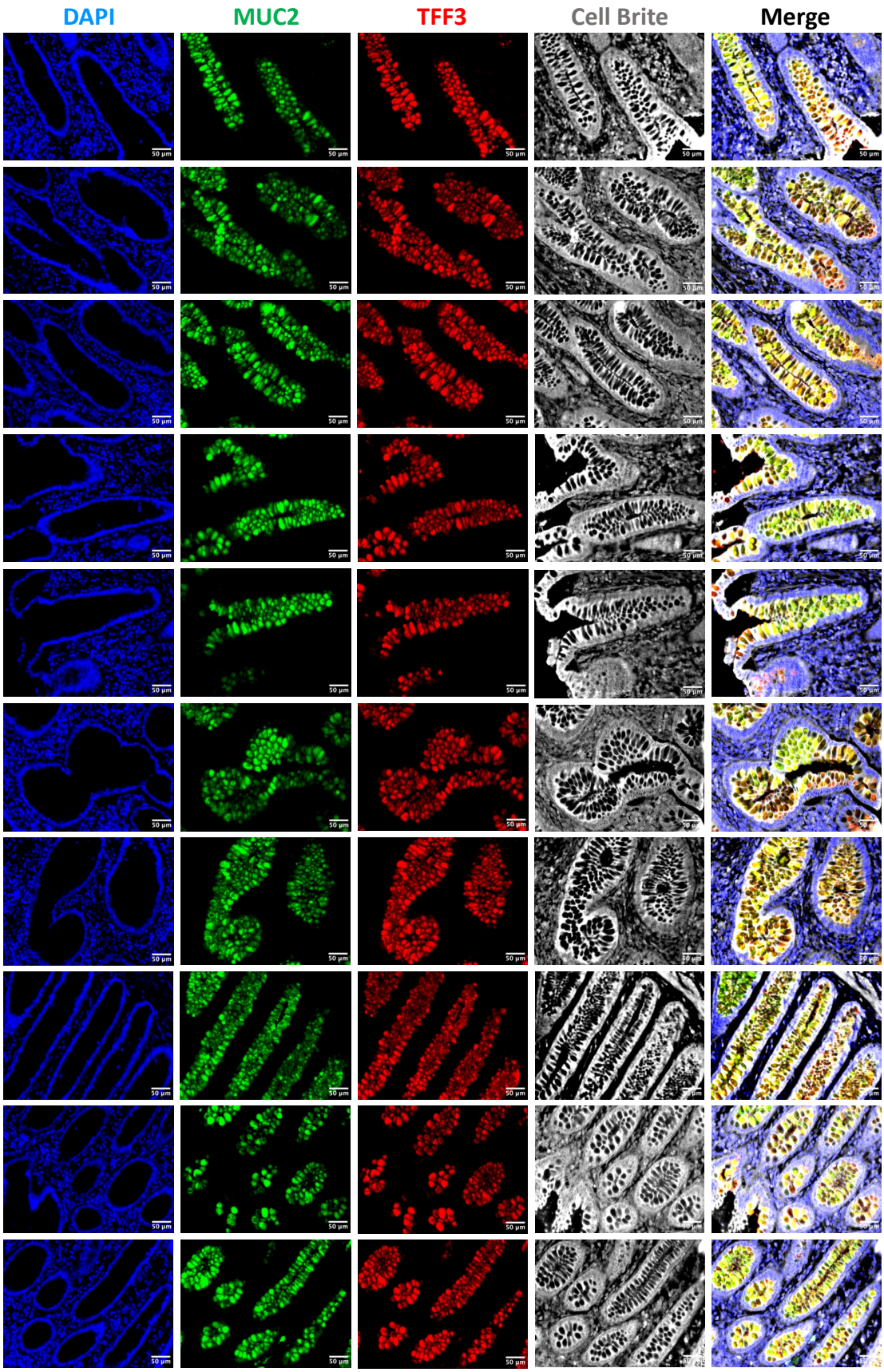
VT11276 (colorectal cancer)

B.



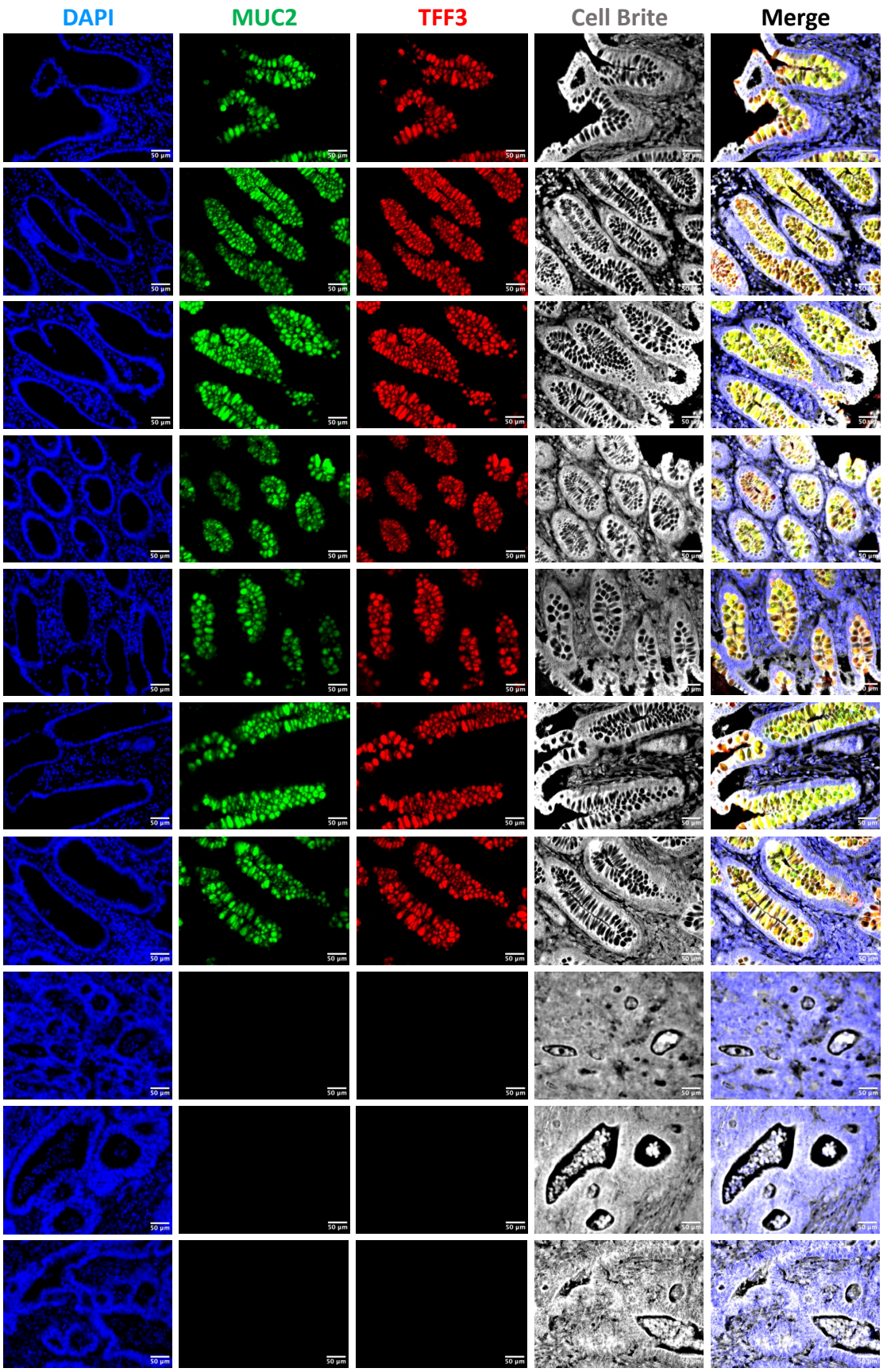
VT11276 (colorectal colon)

B.



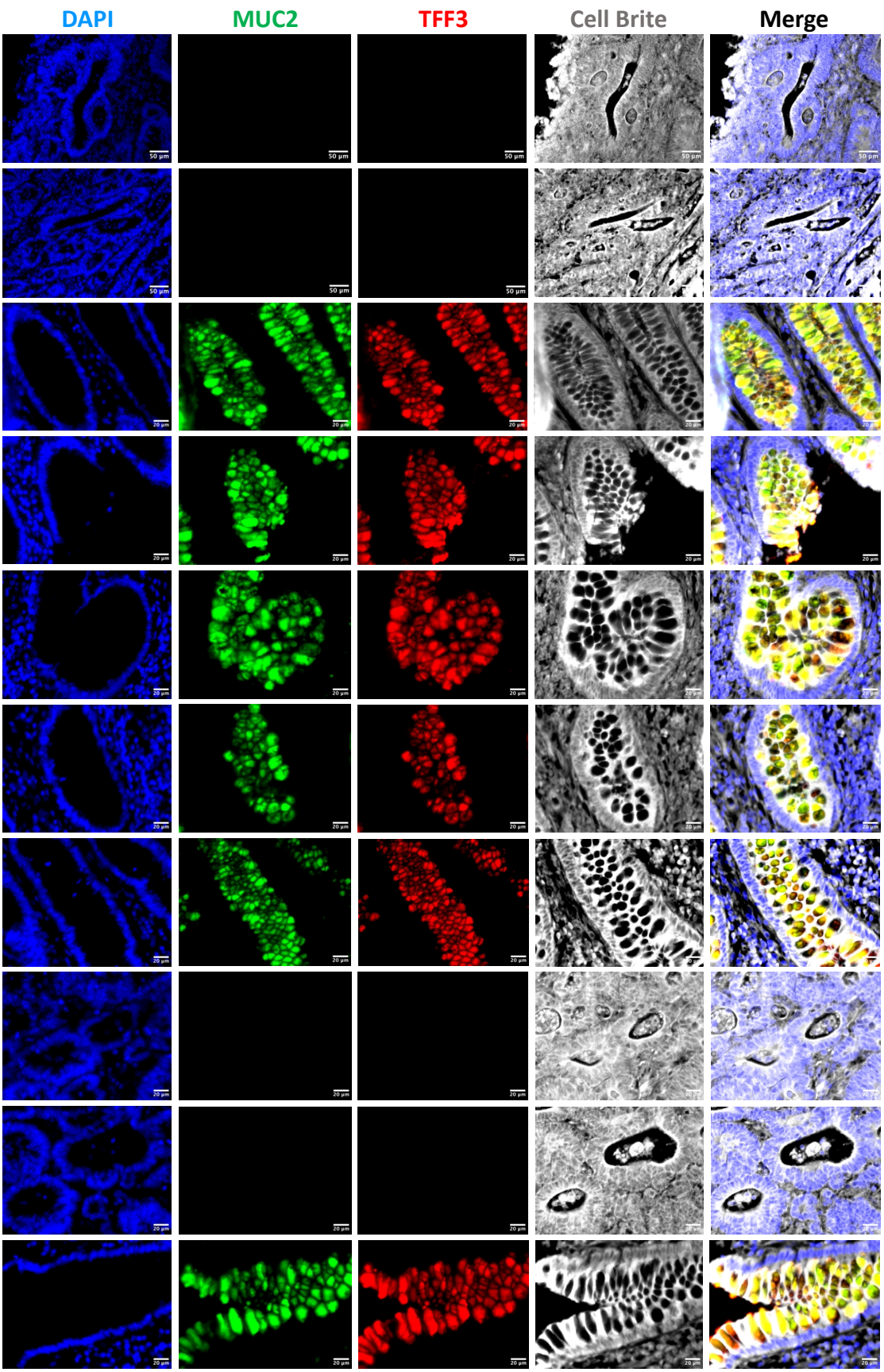
VT11276 (colorectal colon)

B.



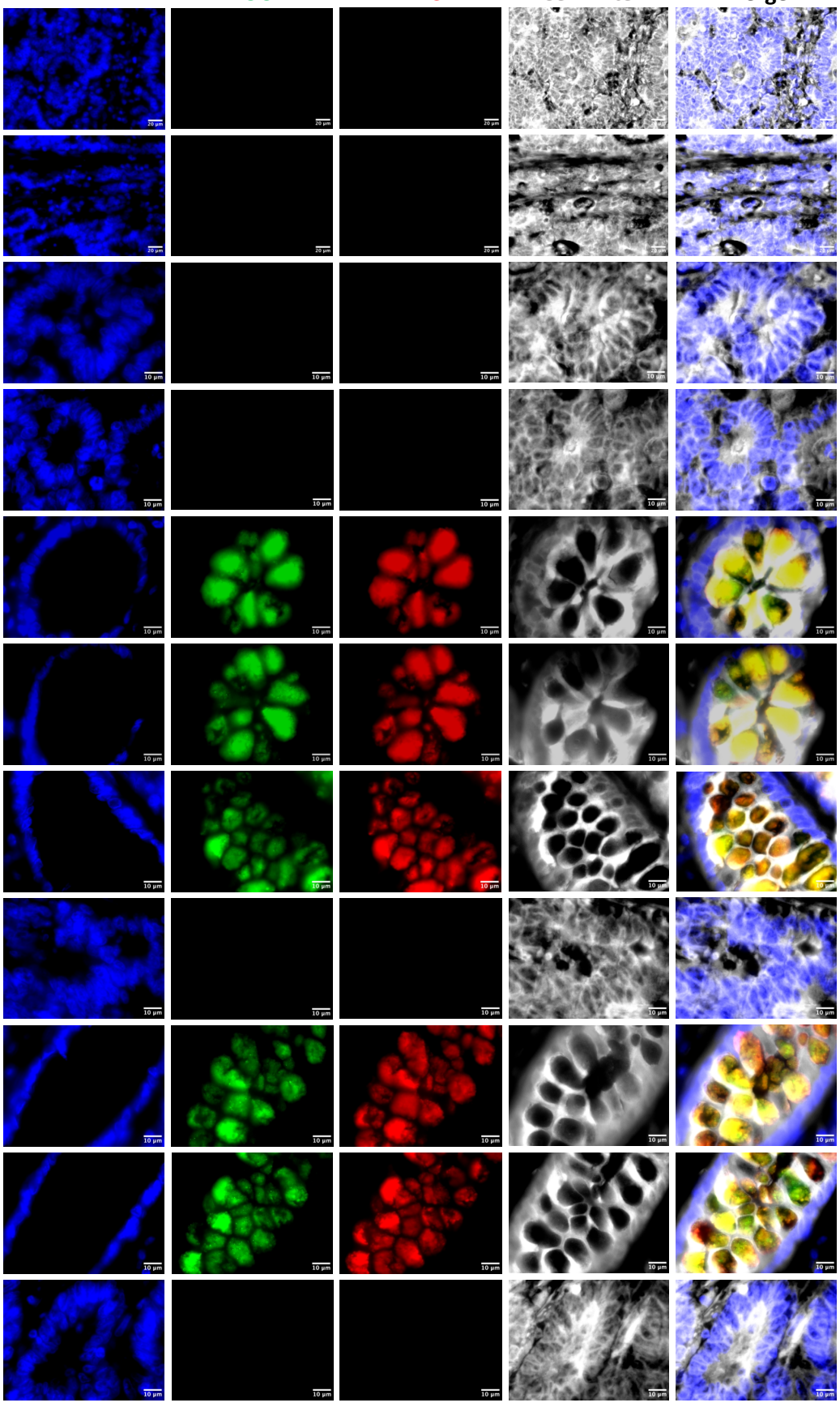
VT11276 (colorectal colon)

B.



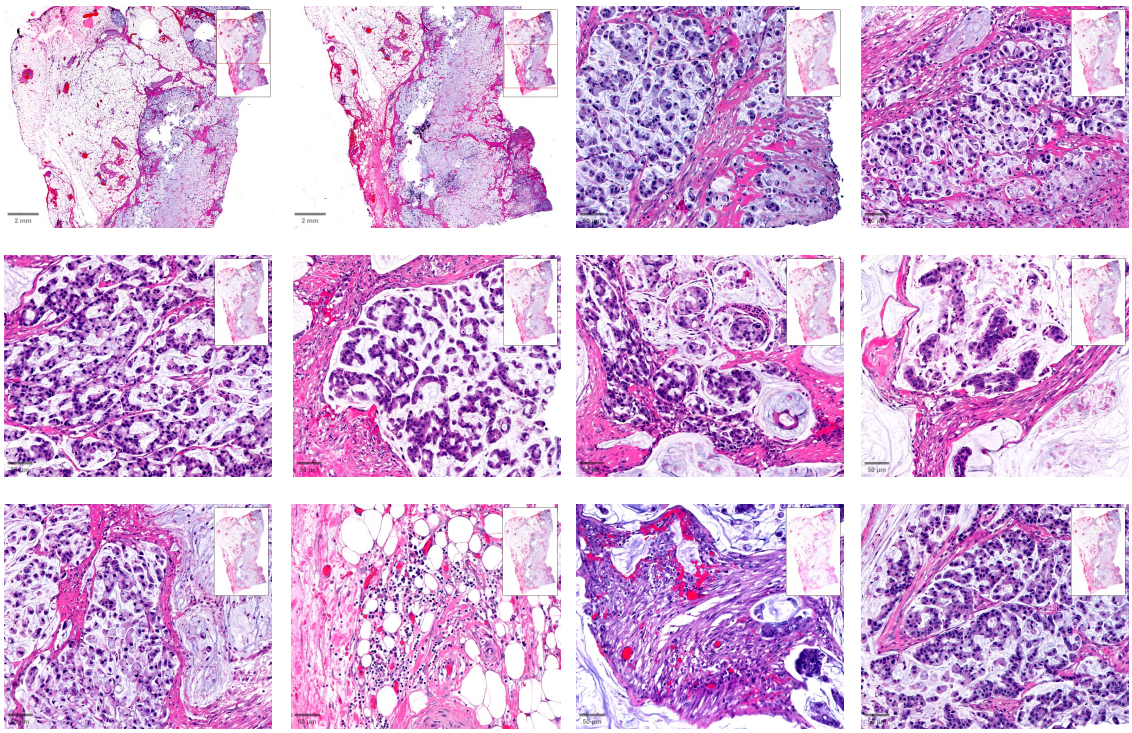
VT11276 (colorectal colon)

B.



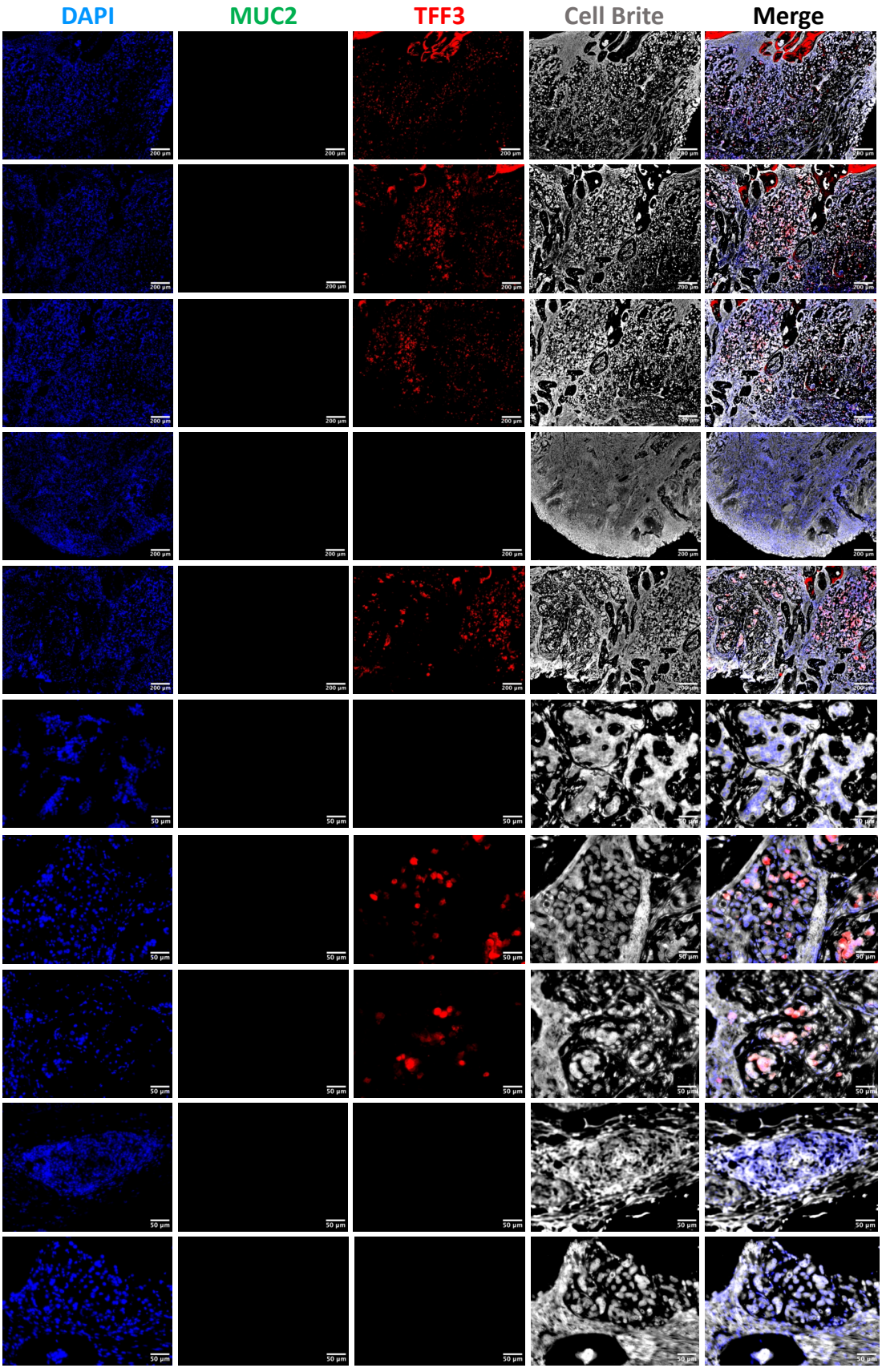
VT11280 (colorectal cancer)

A.



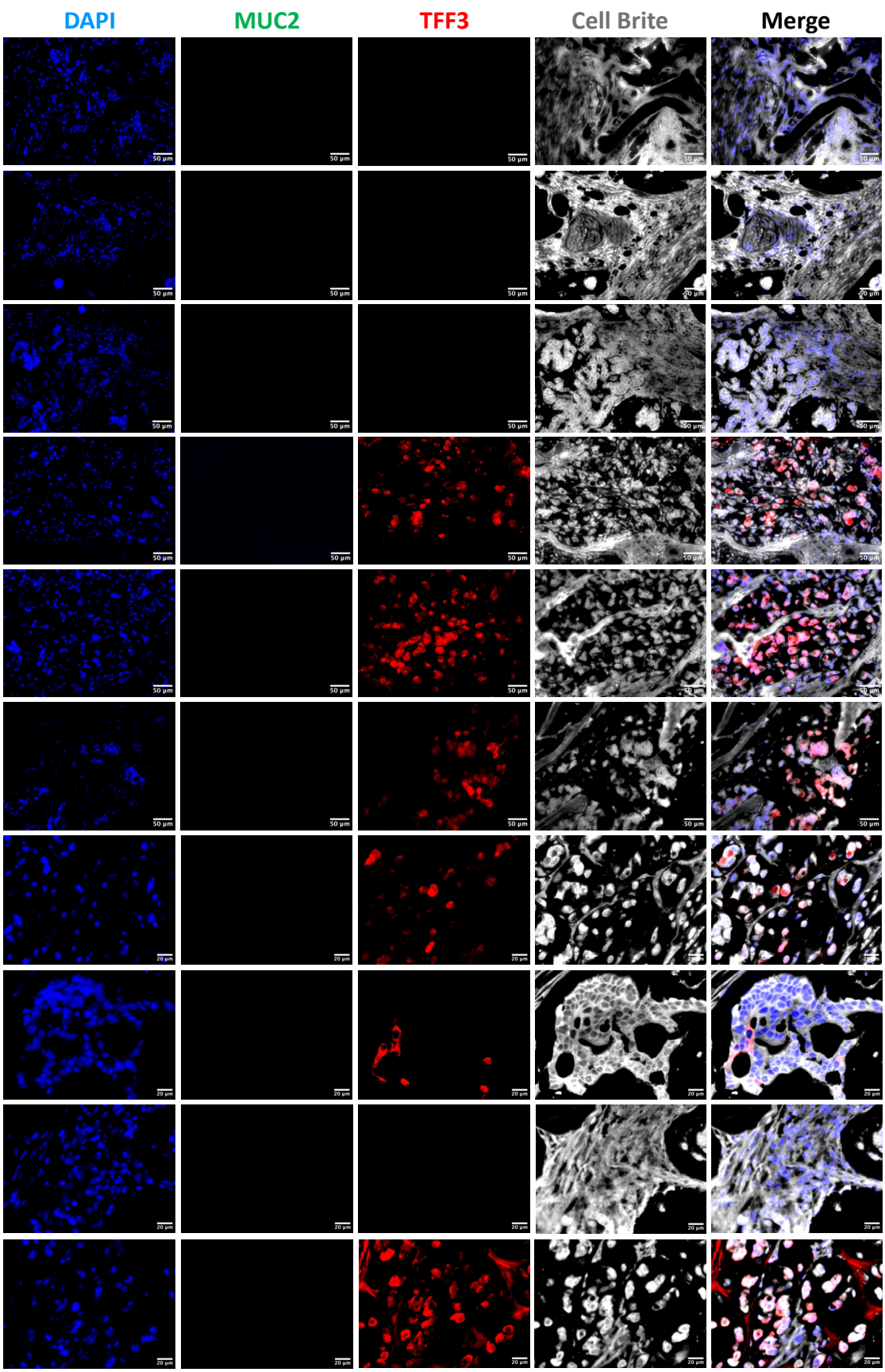
VT11280 (colorectal cancer)

B.



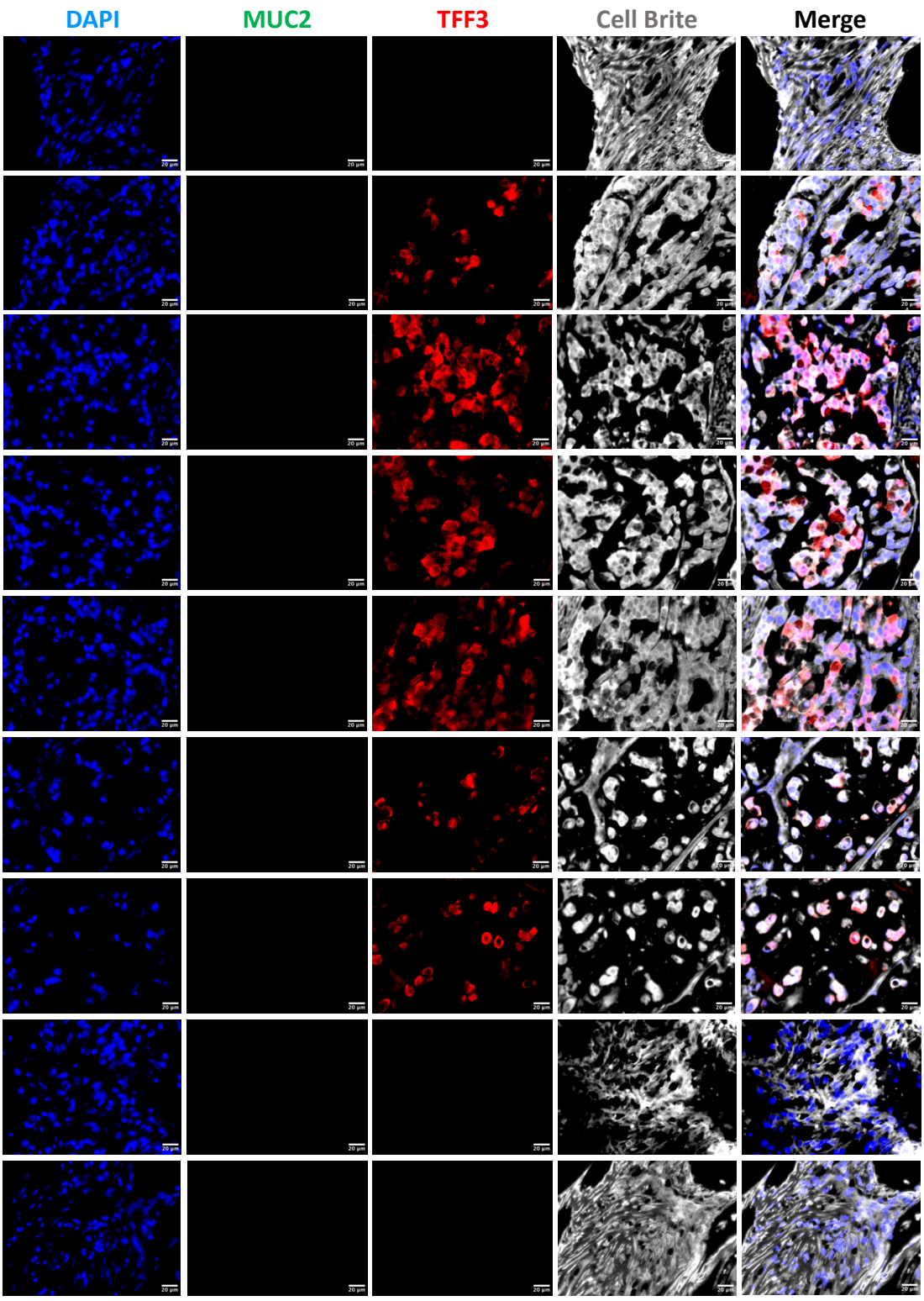
VT11280 (colorectal colon)

B.



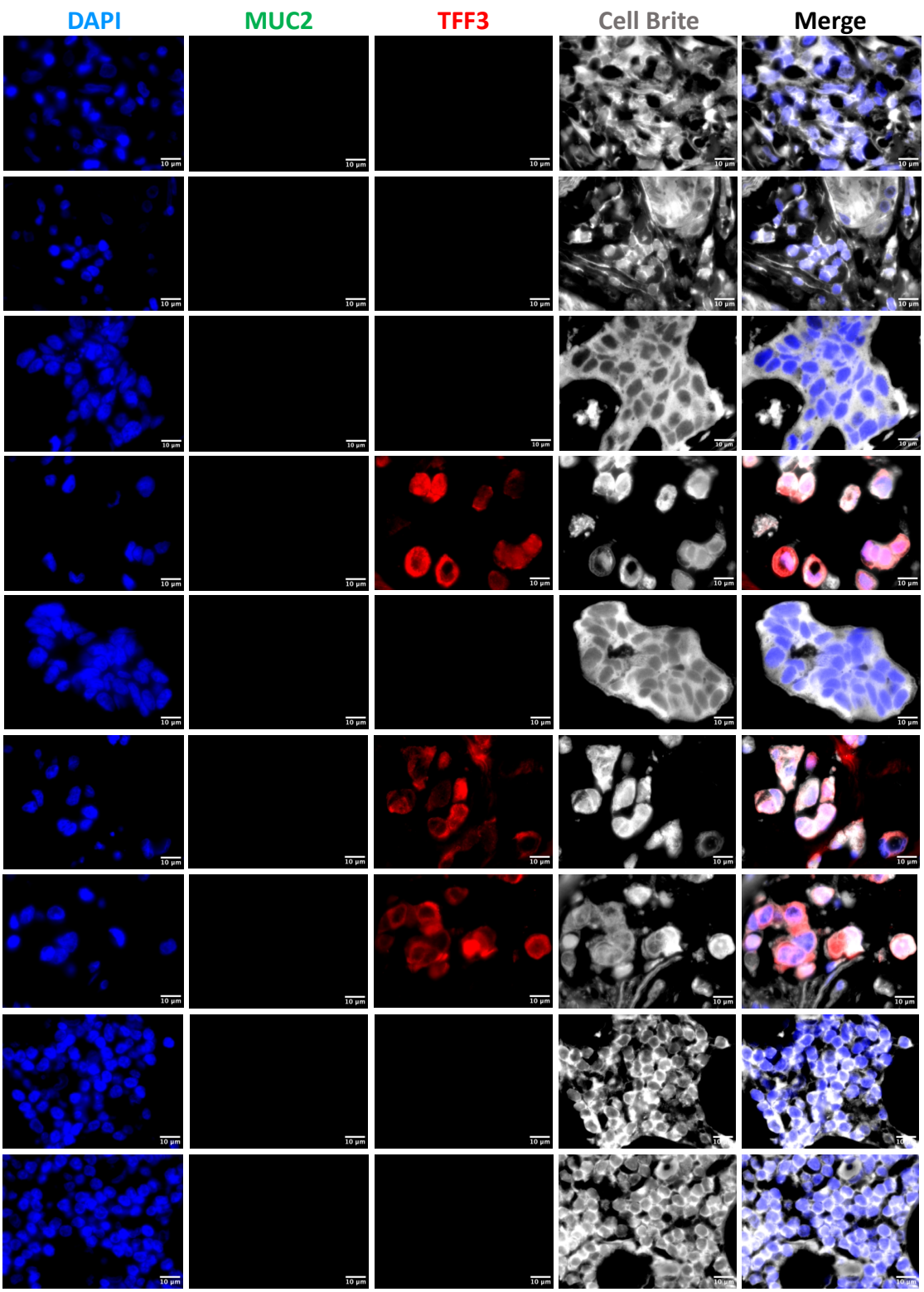
VT11280 (colorectal colon)

B.



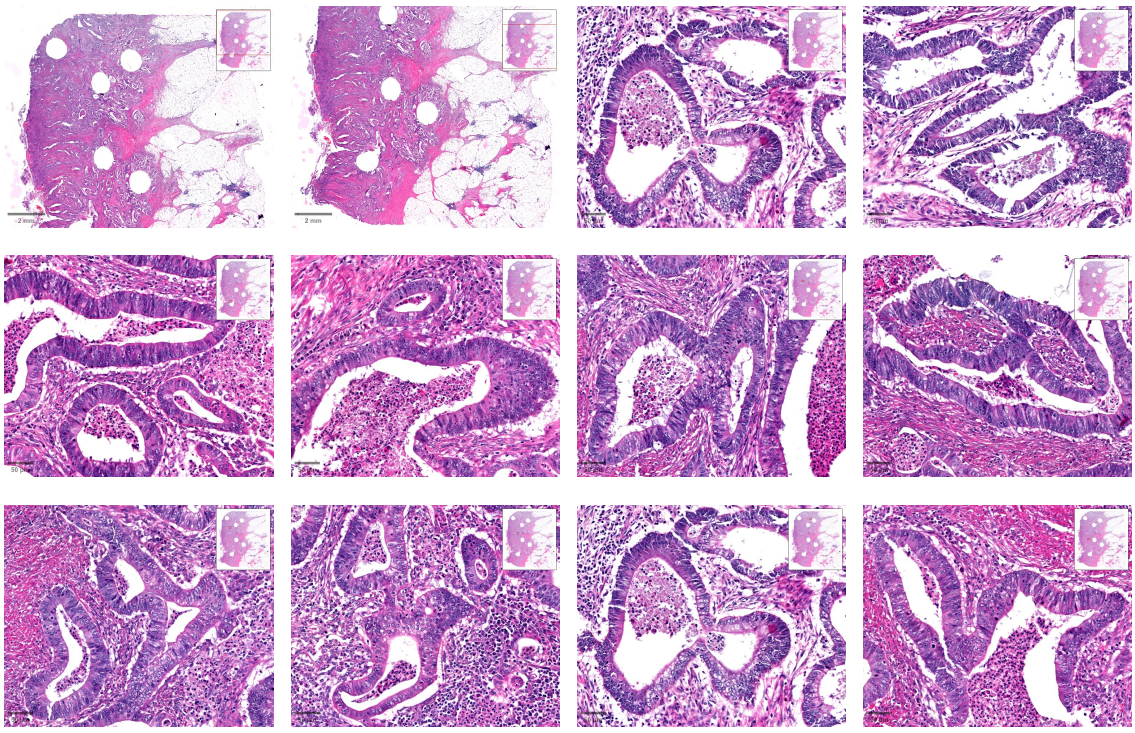
VT11280 (colorectal colon)

B.



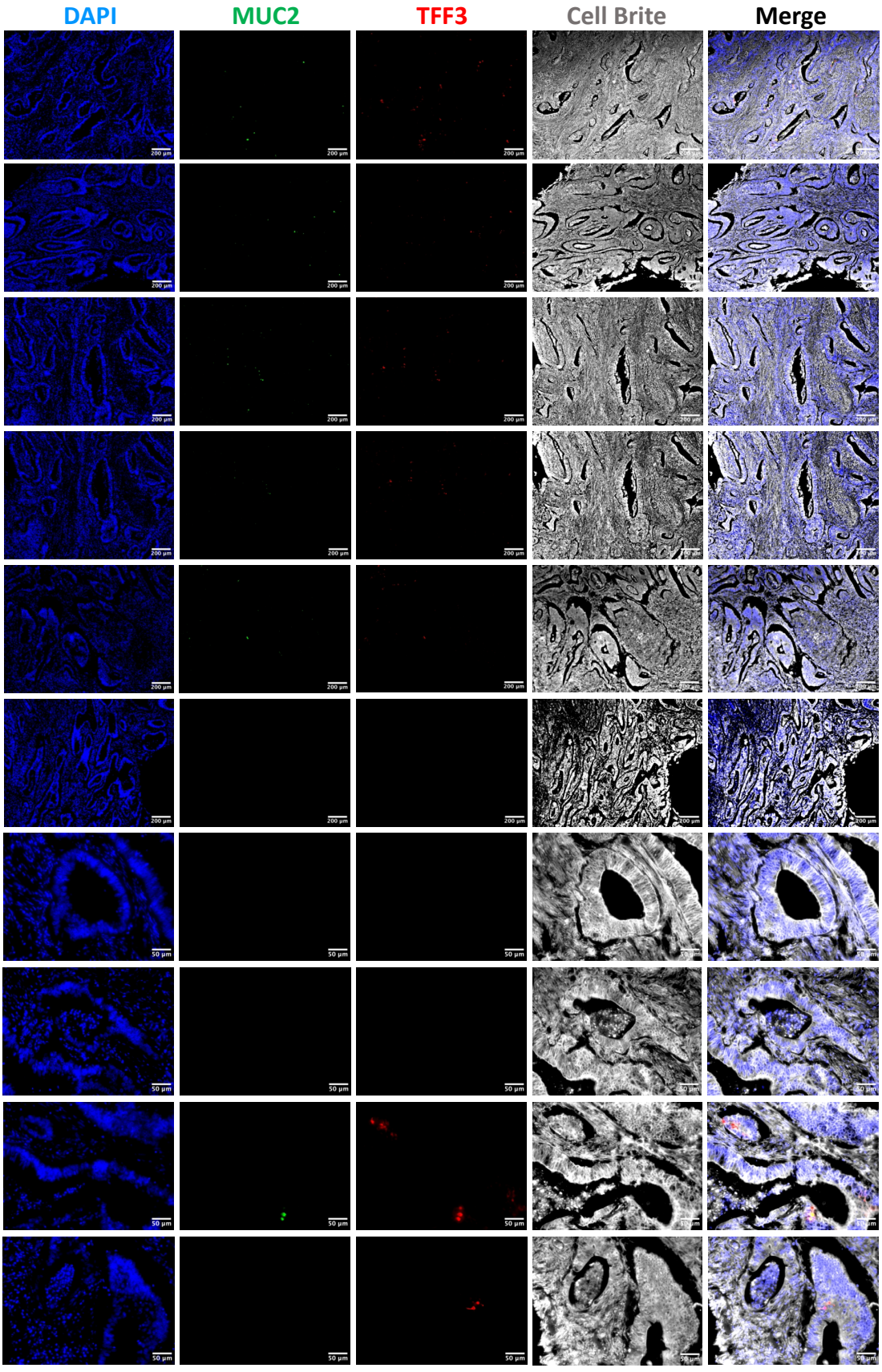
VT11282 (colorectal cancer)

A.



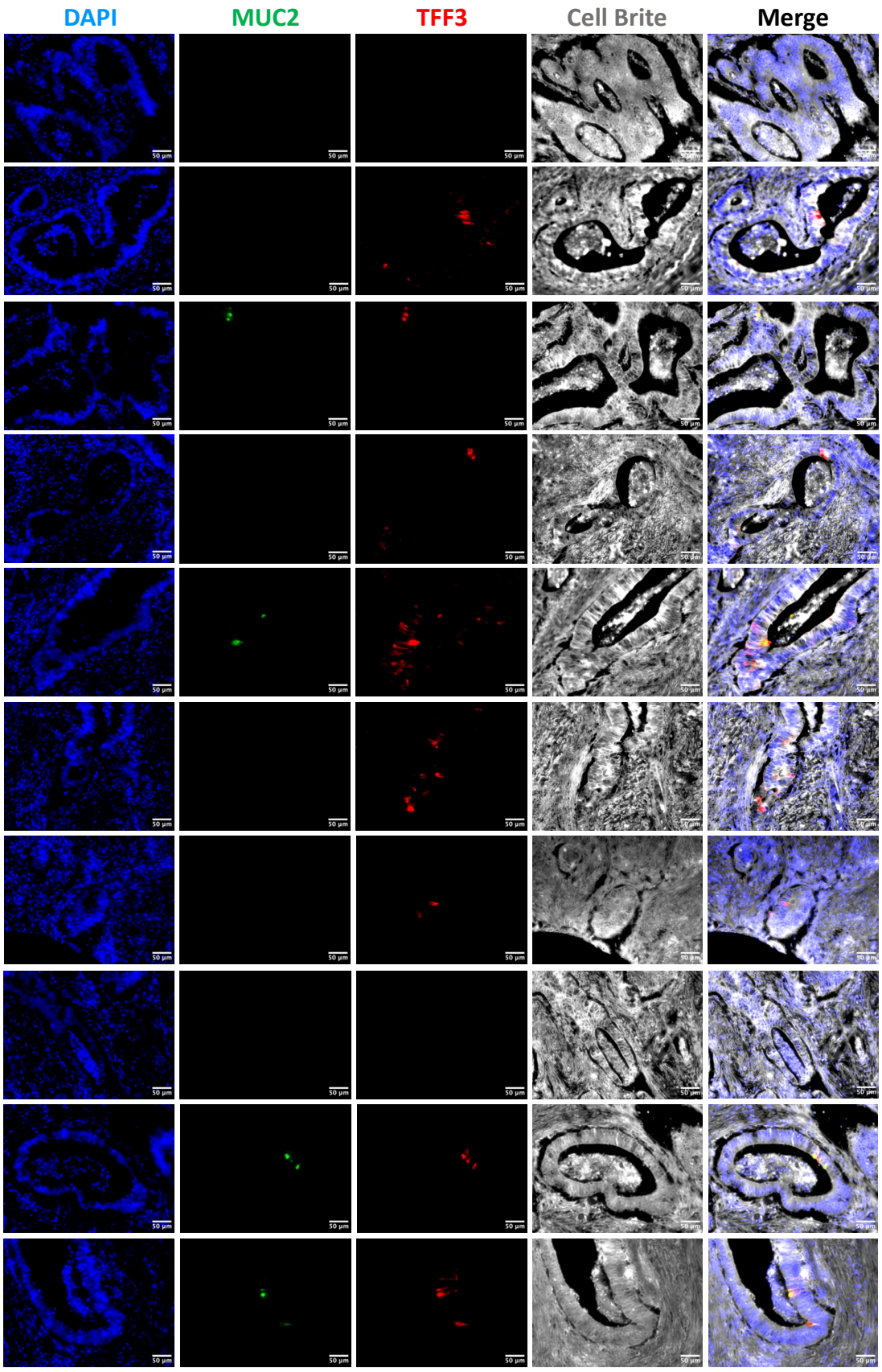
VT11282 (colorectal cancer)

B.



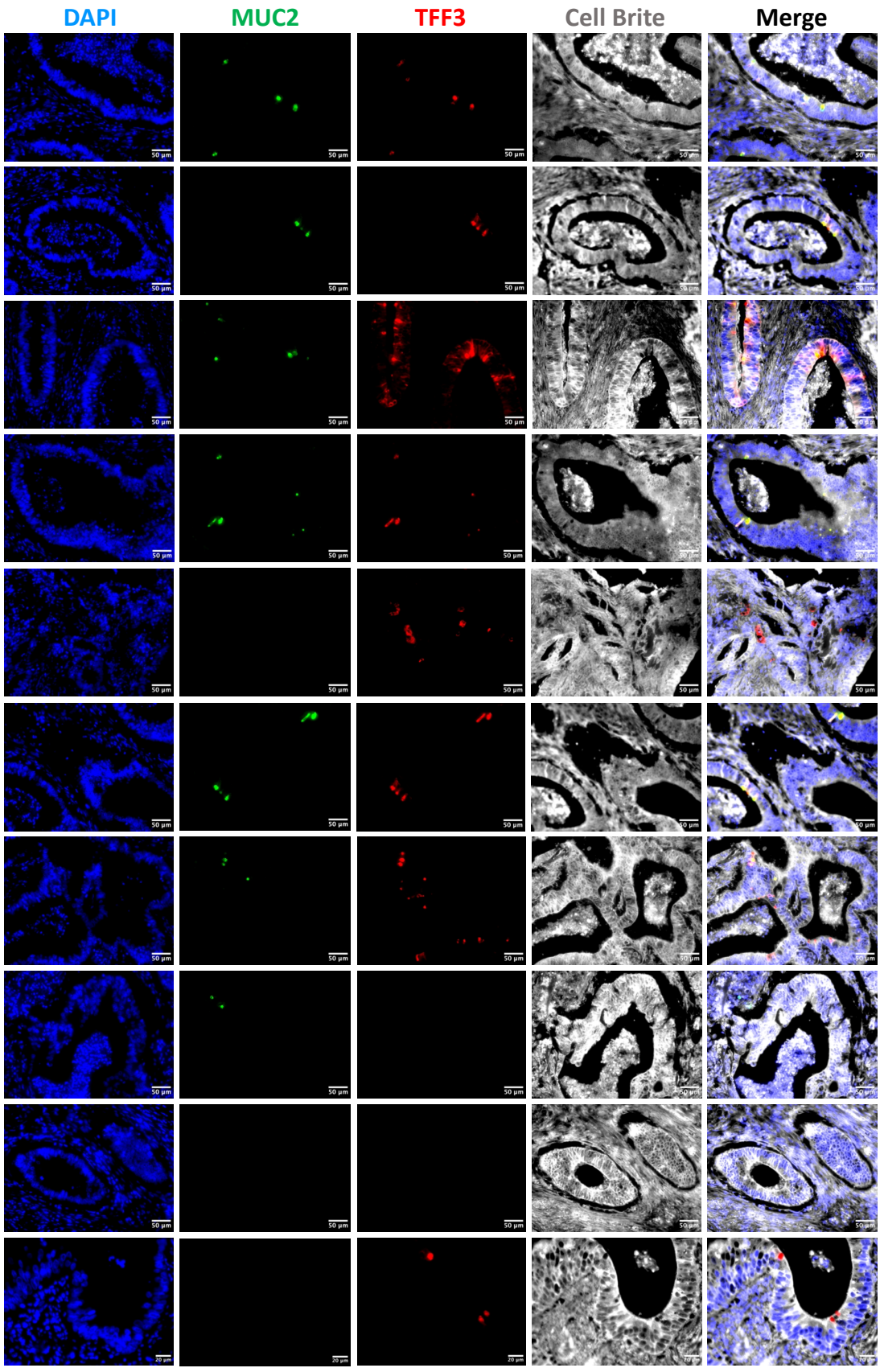
VT11282 (colorectal cancer)

B.



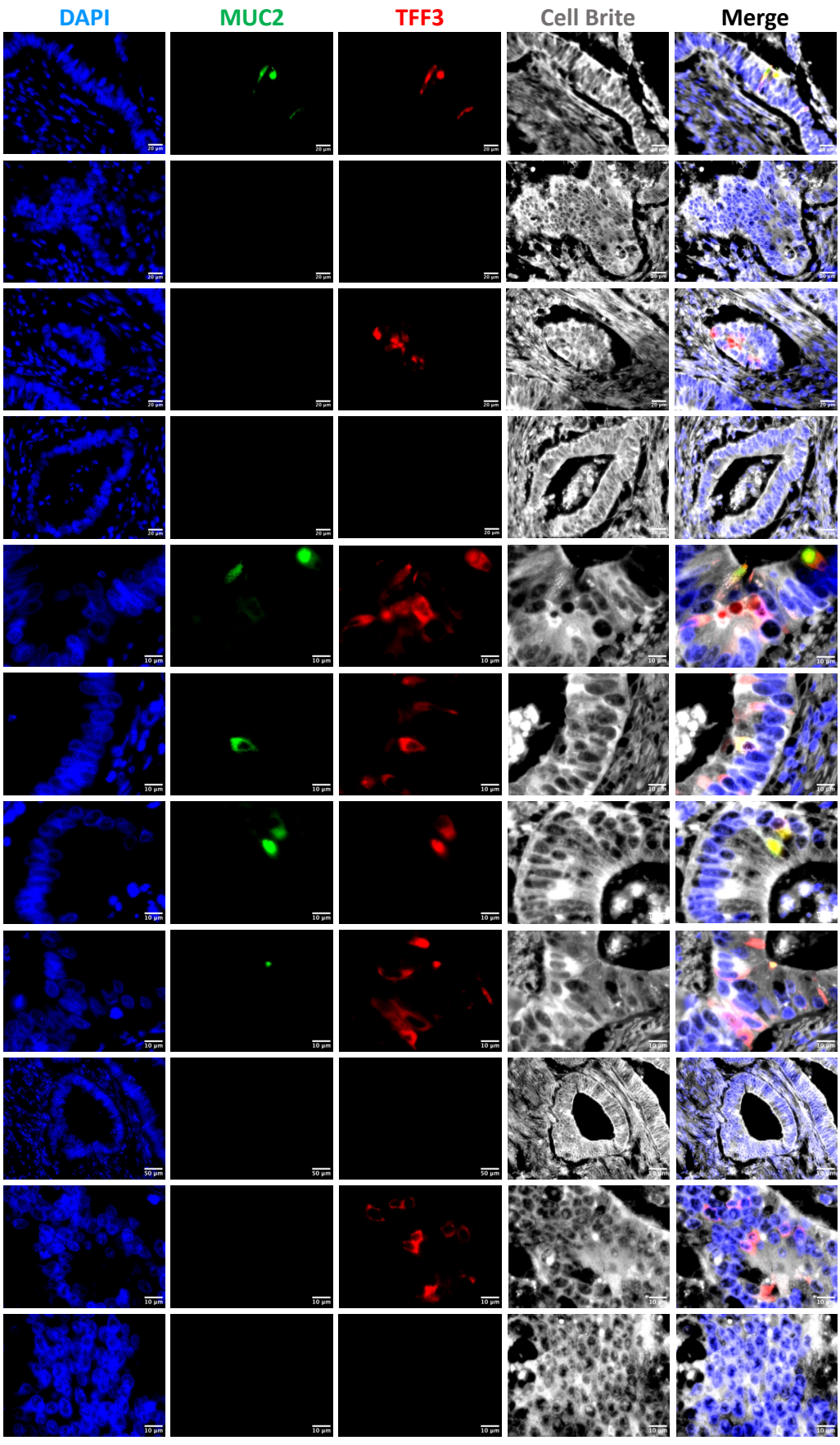
VT11282 (colorectal cancer)

B.



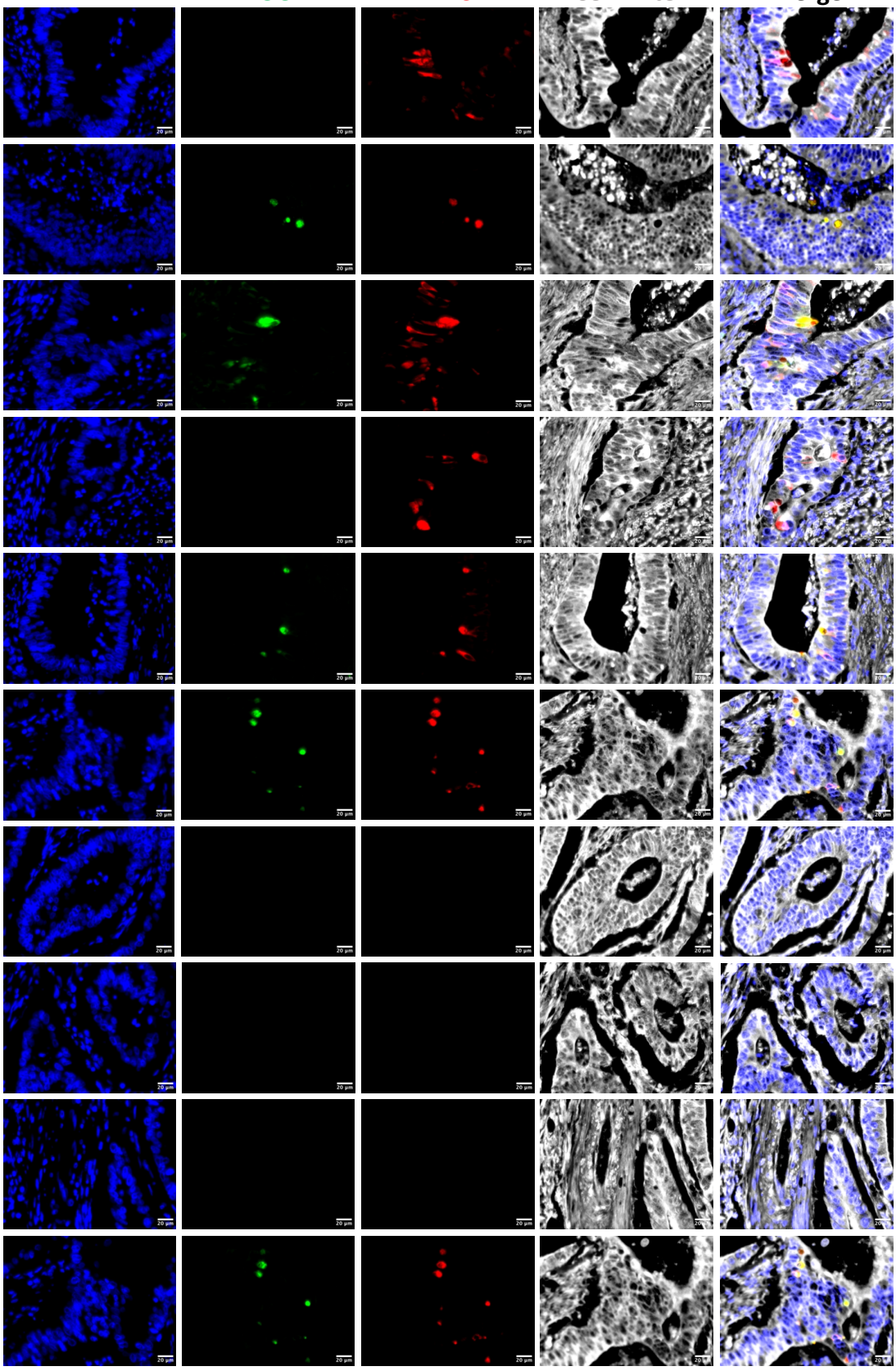
VT11282 (colorectal cancer)

B.



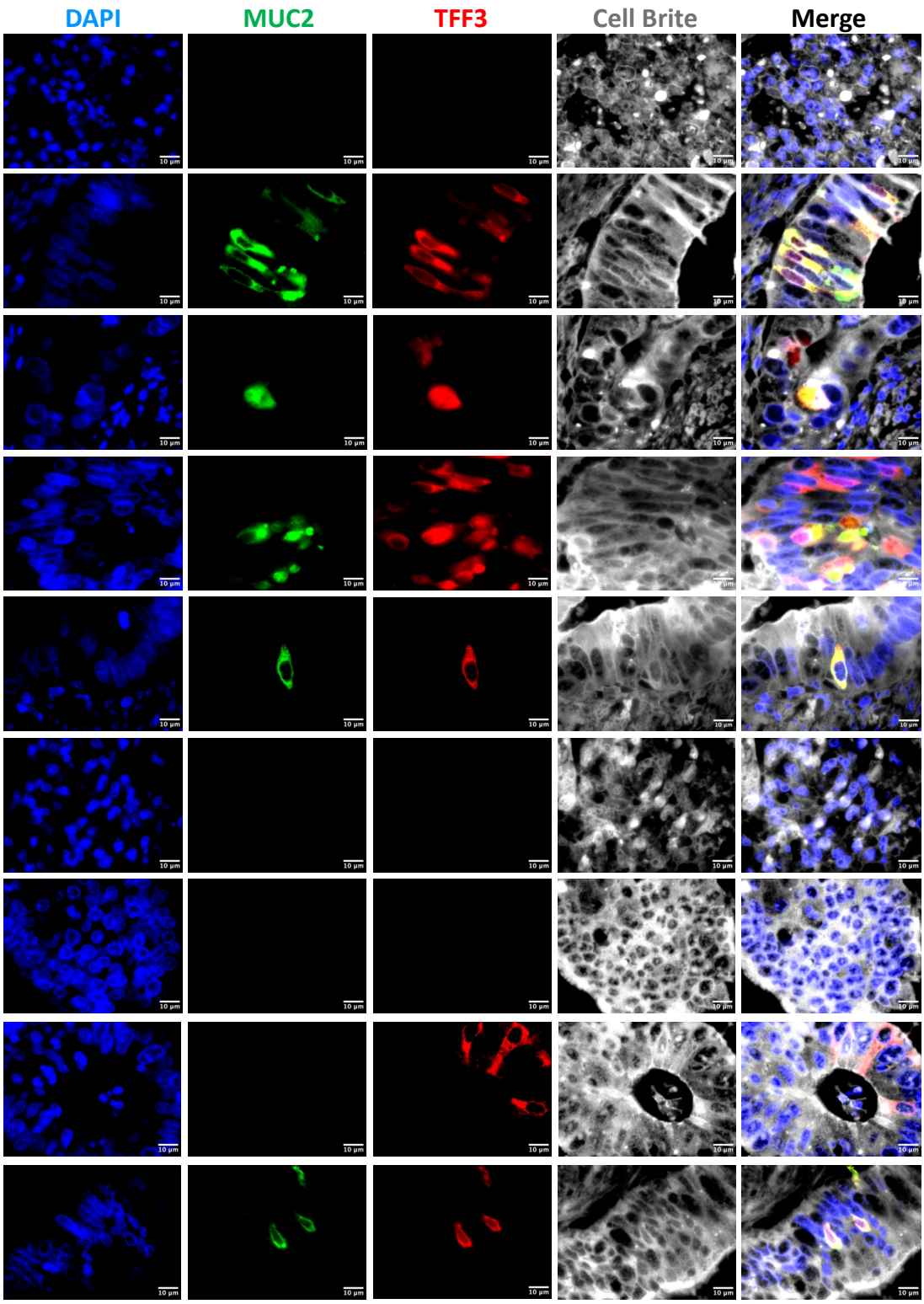
VT11282 (colorectal cancer)

B.



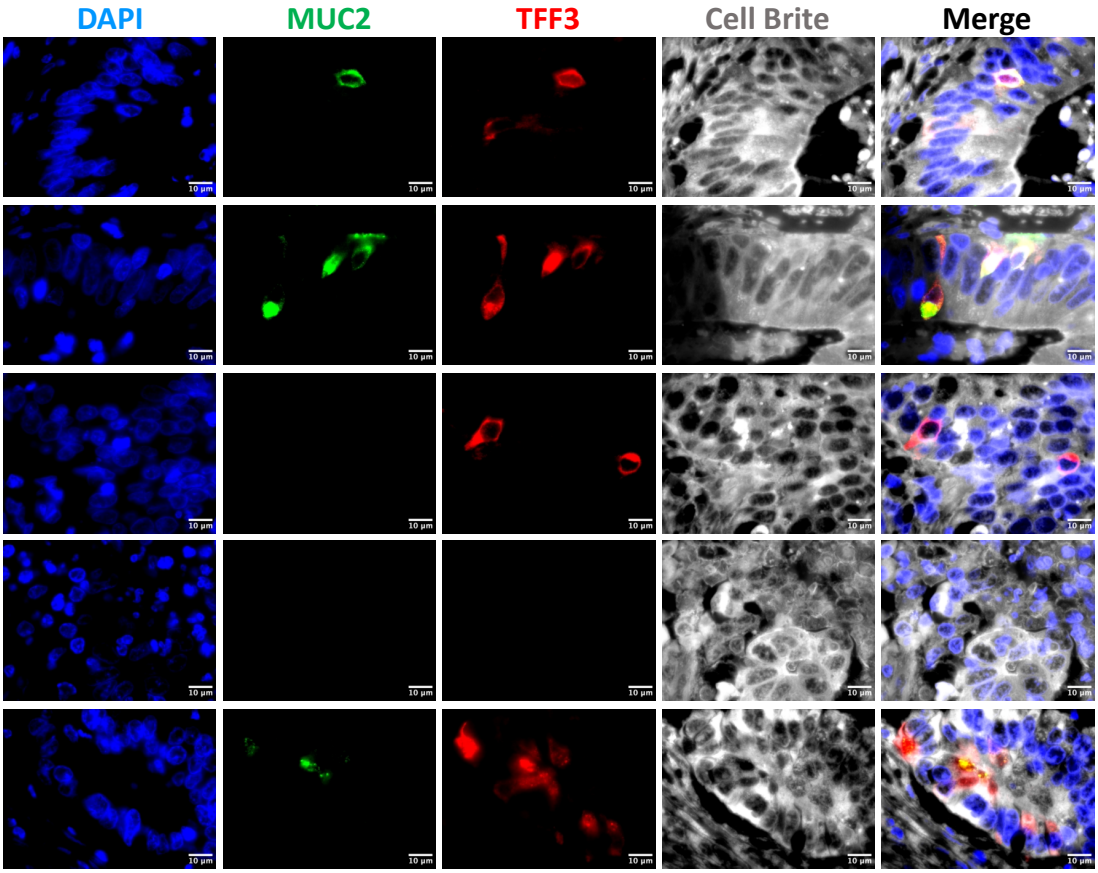
VT11282 (colorectal cancer)

B.



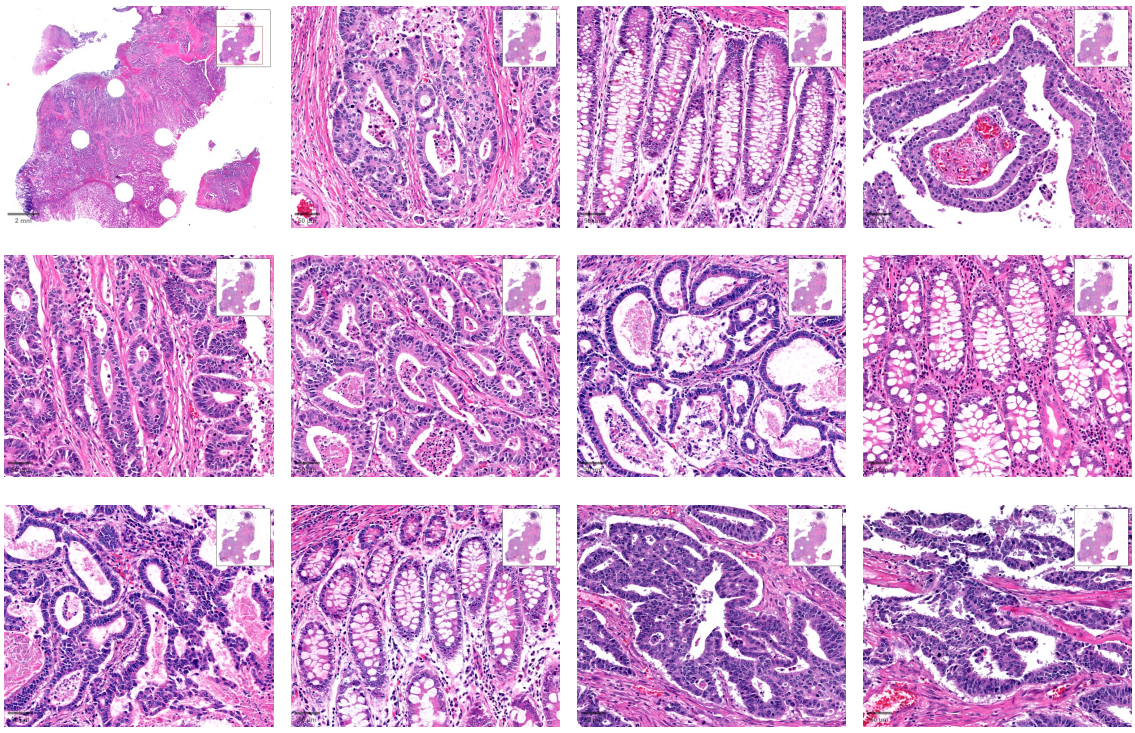
VT11282 (colorectal cancer)

B.



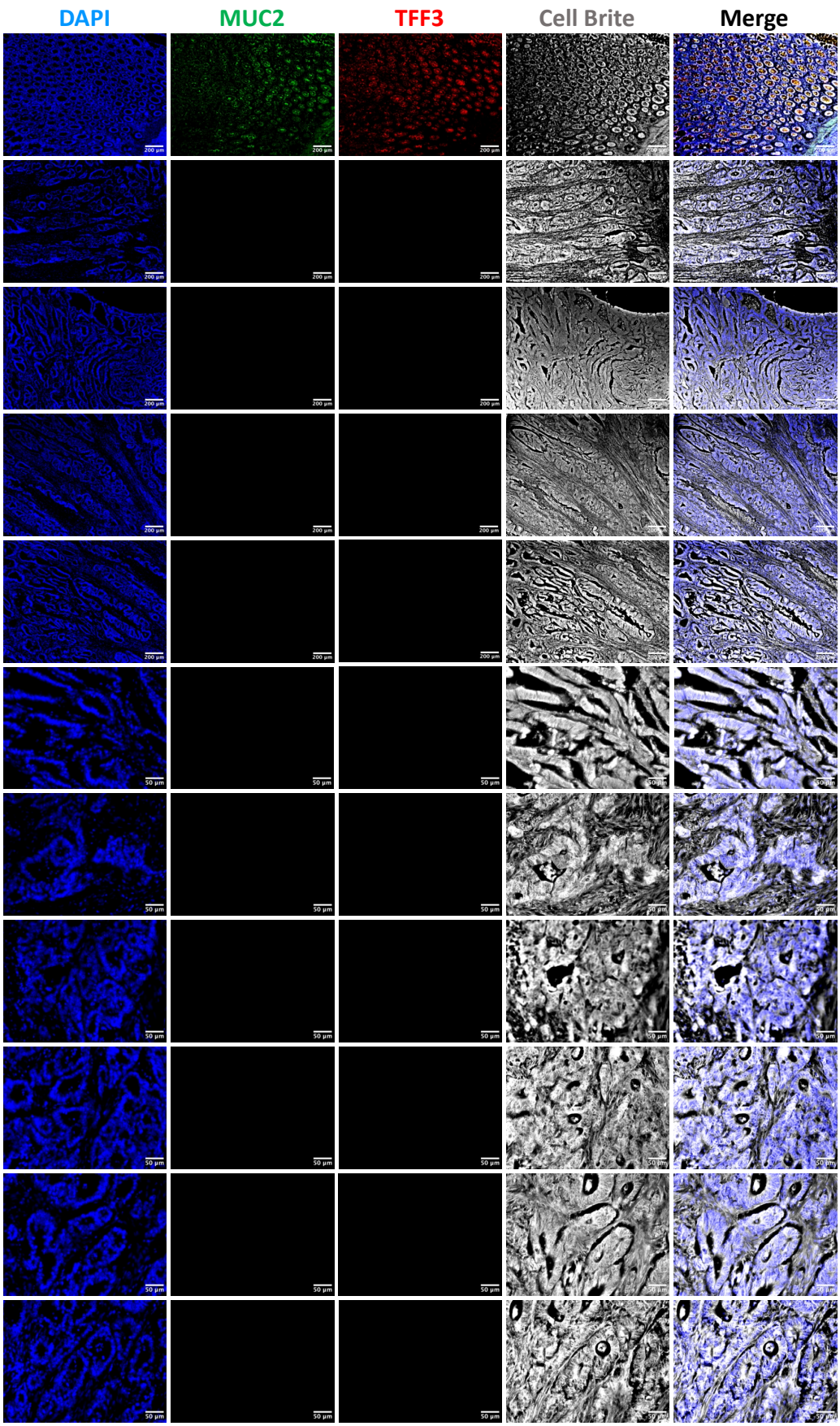
VT11290 (colorectal cancer)

A.



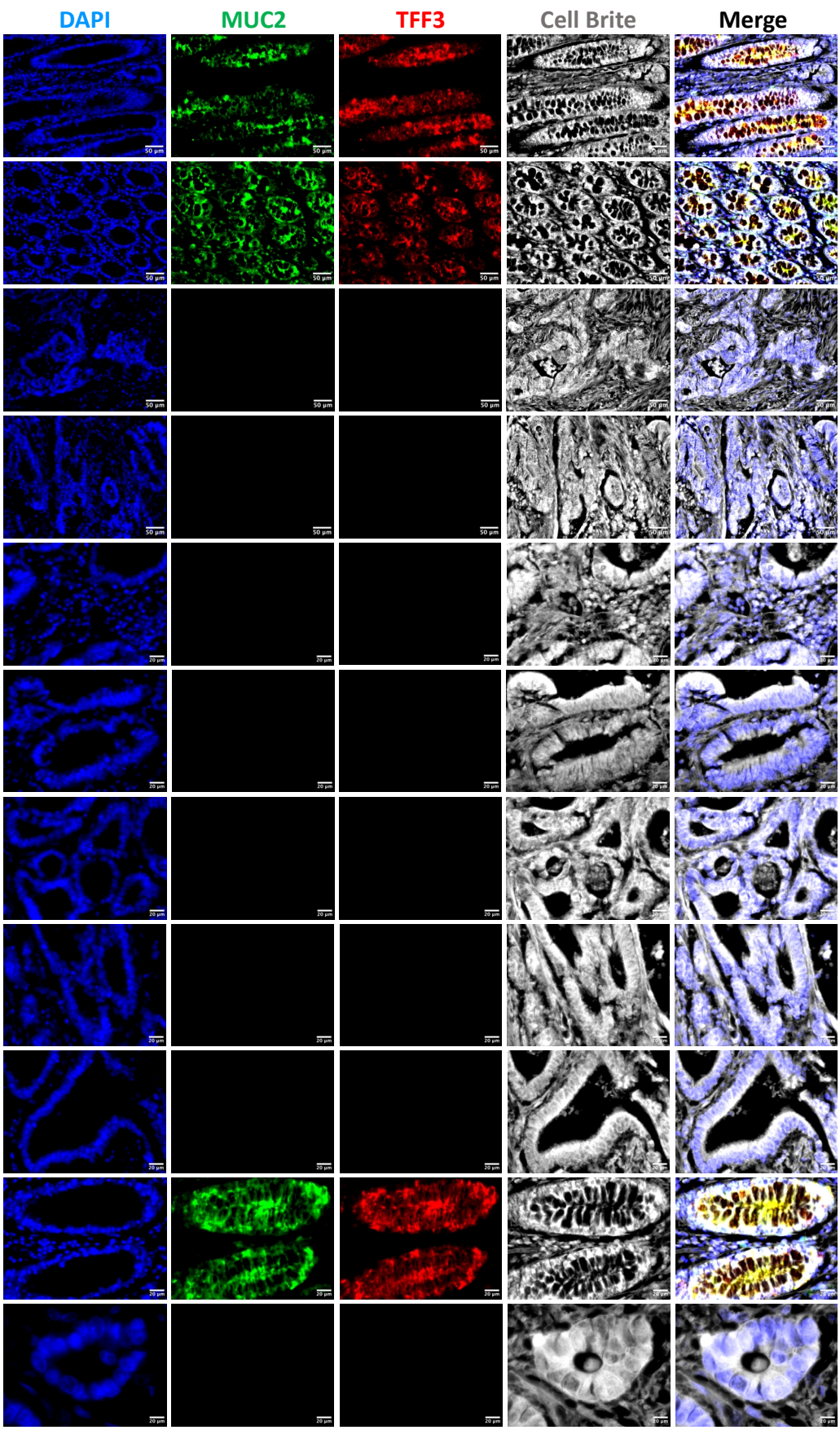
VT11290 (colorectal cancer)

B.



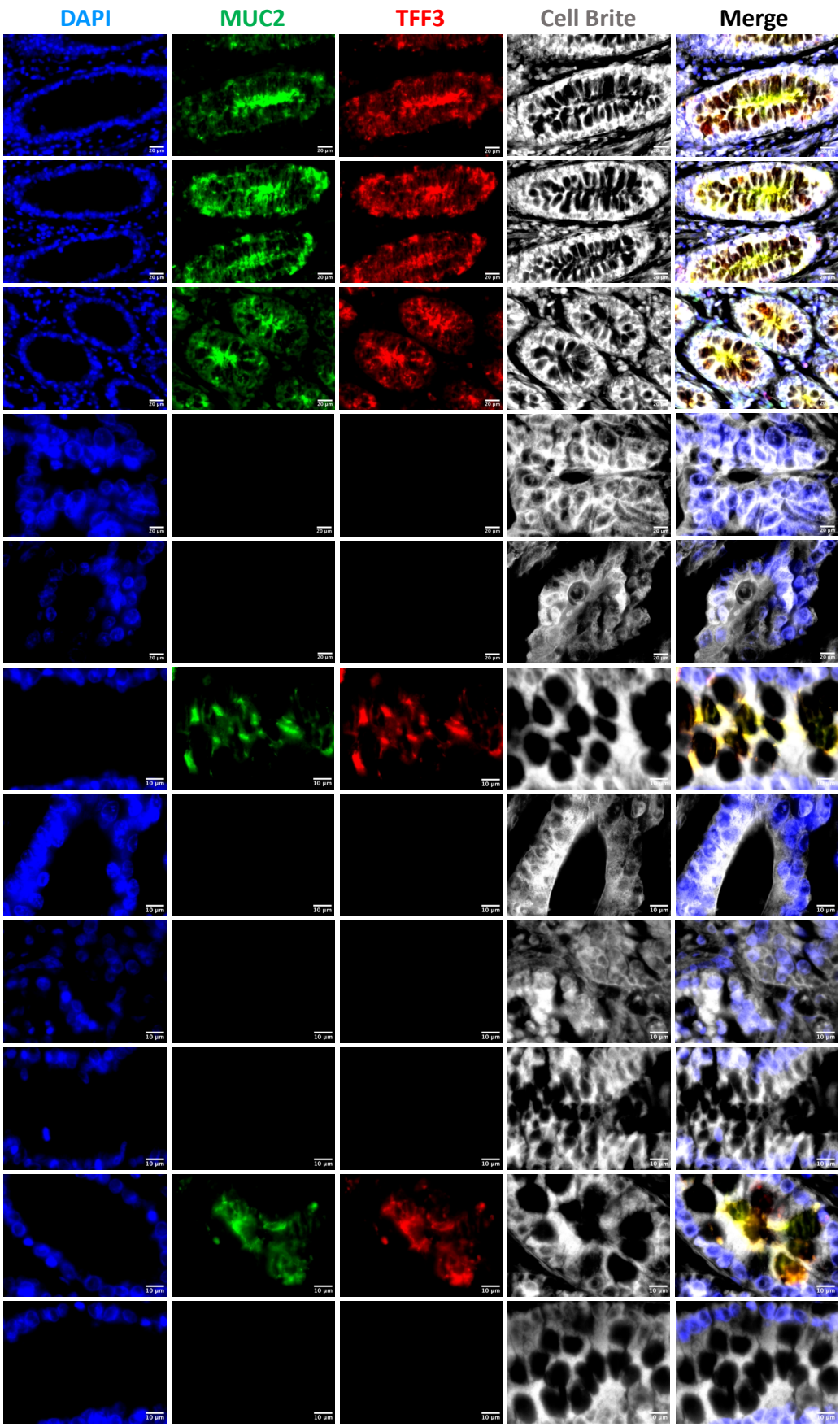
VT11290 (colorectal cancer)

B.



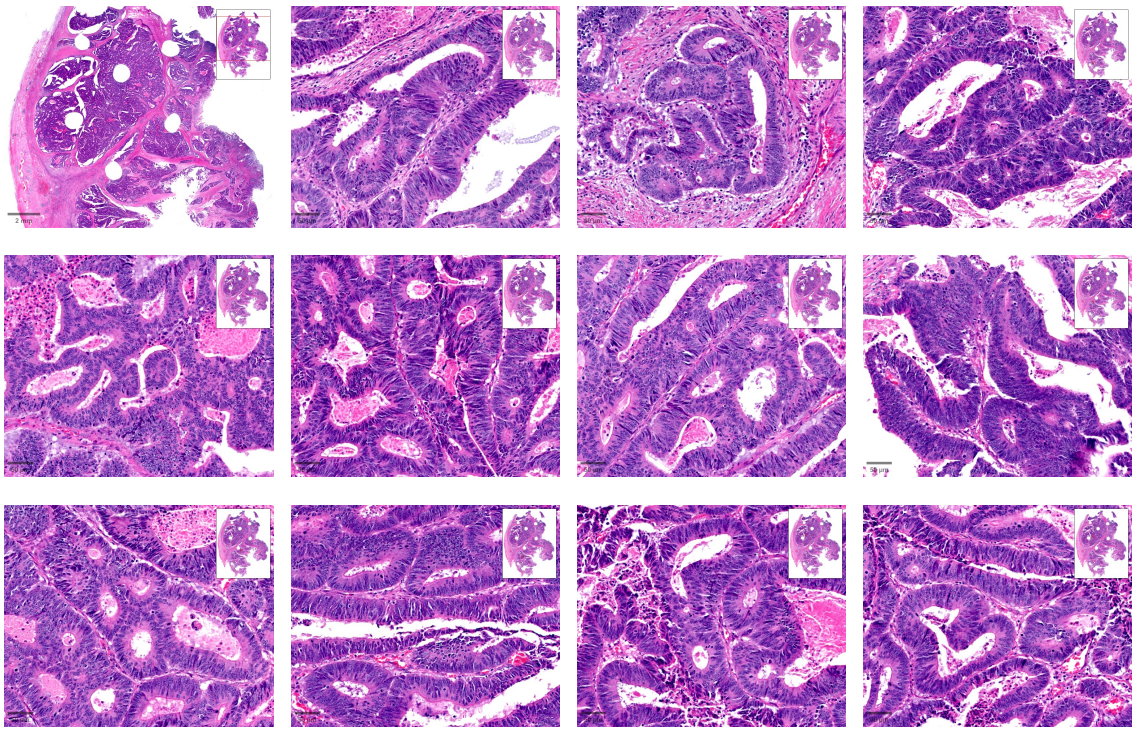
VT11290 (colorectal cancer)

B.



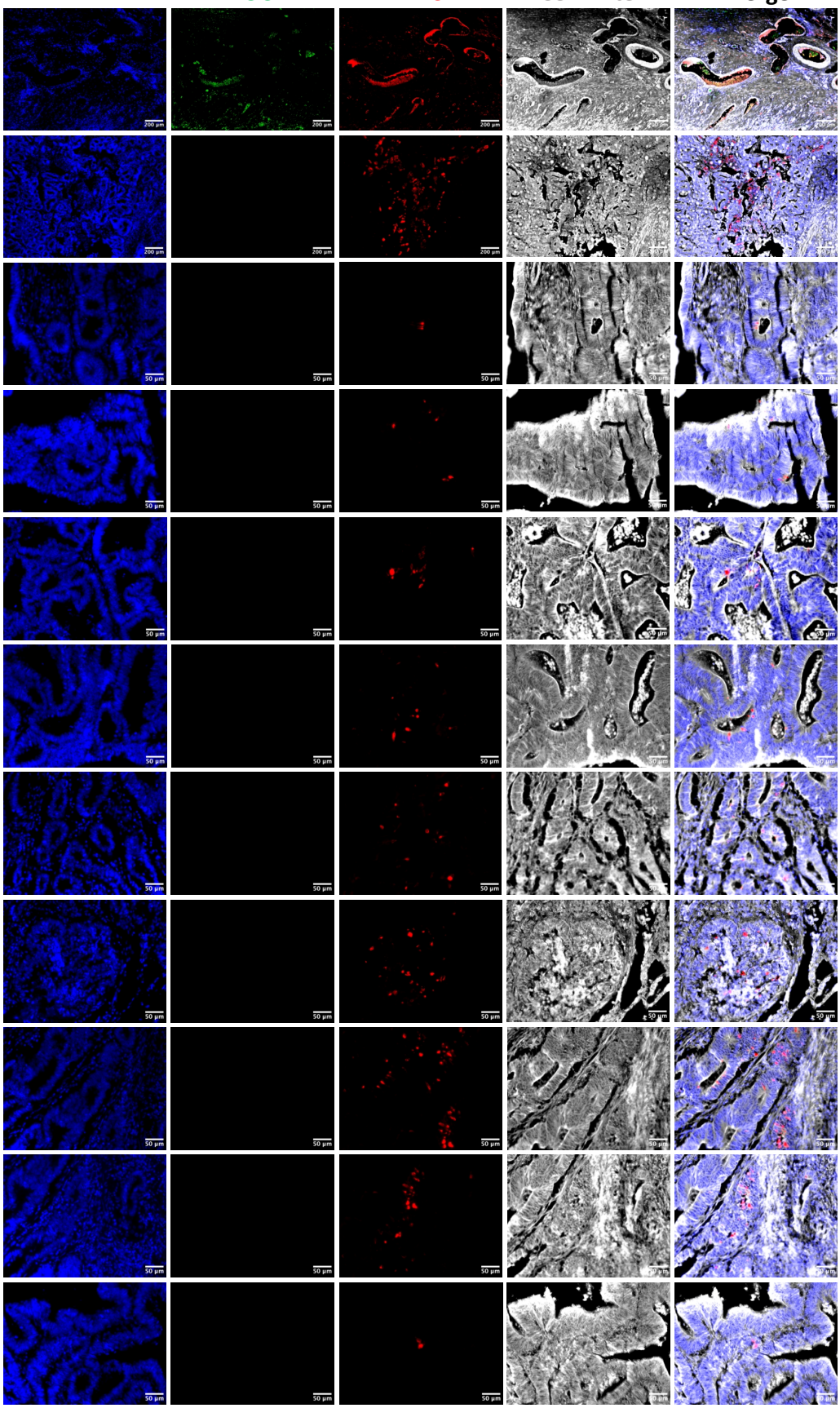
VT11294 (colorectal cancer)

A.



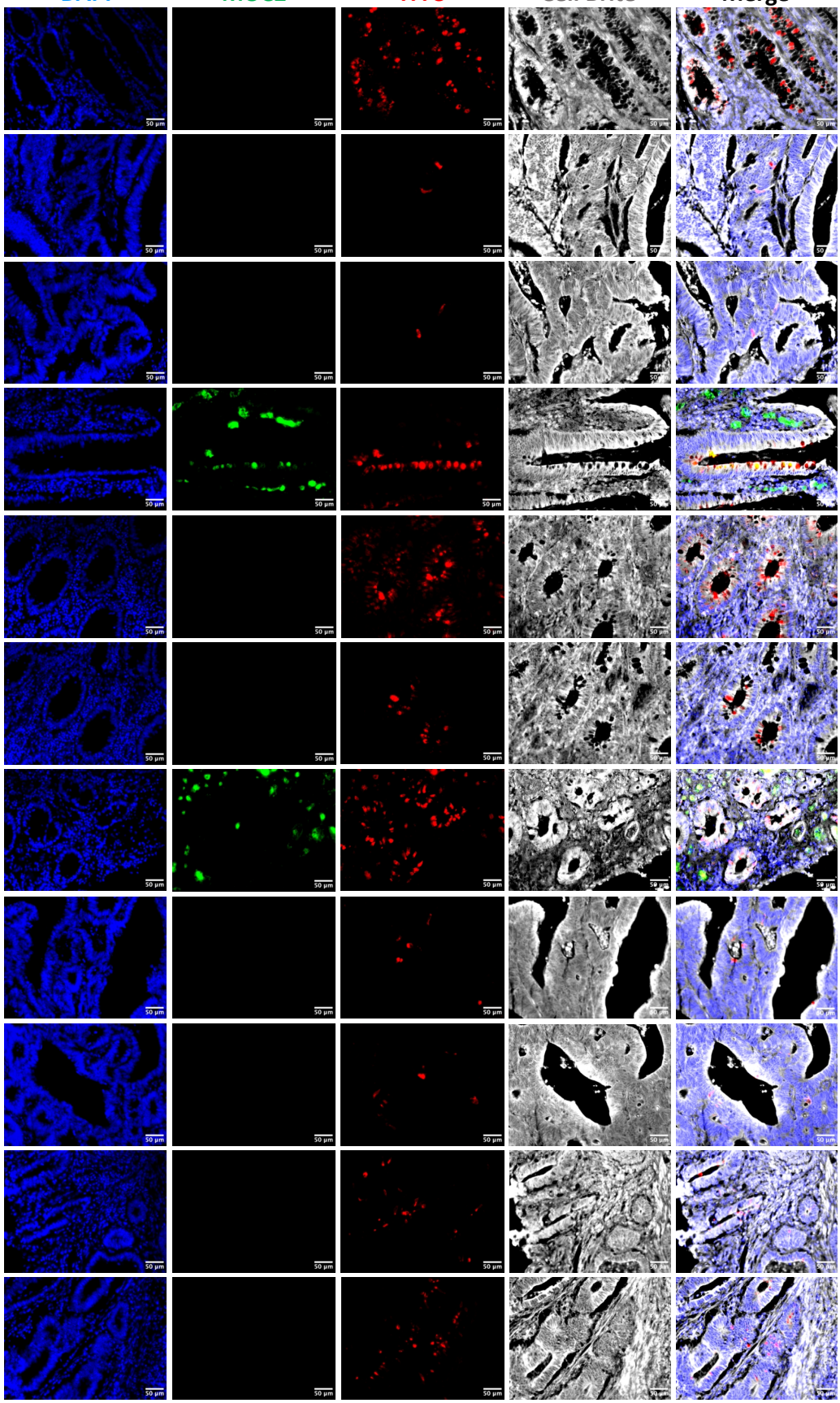
VT11294 (colorectal cancer)

B.



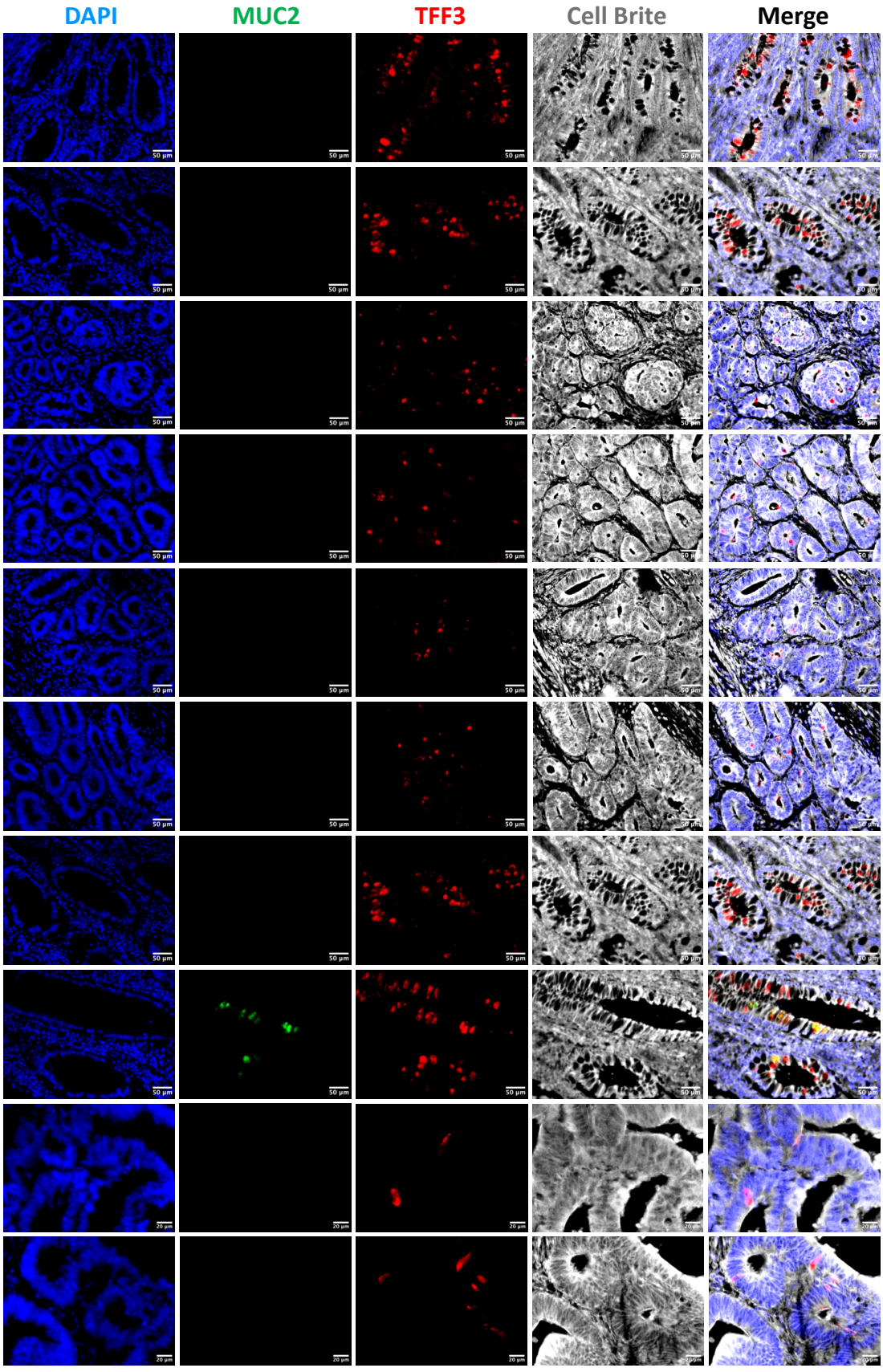
VT11294 (colorectal cancer)

B.



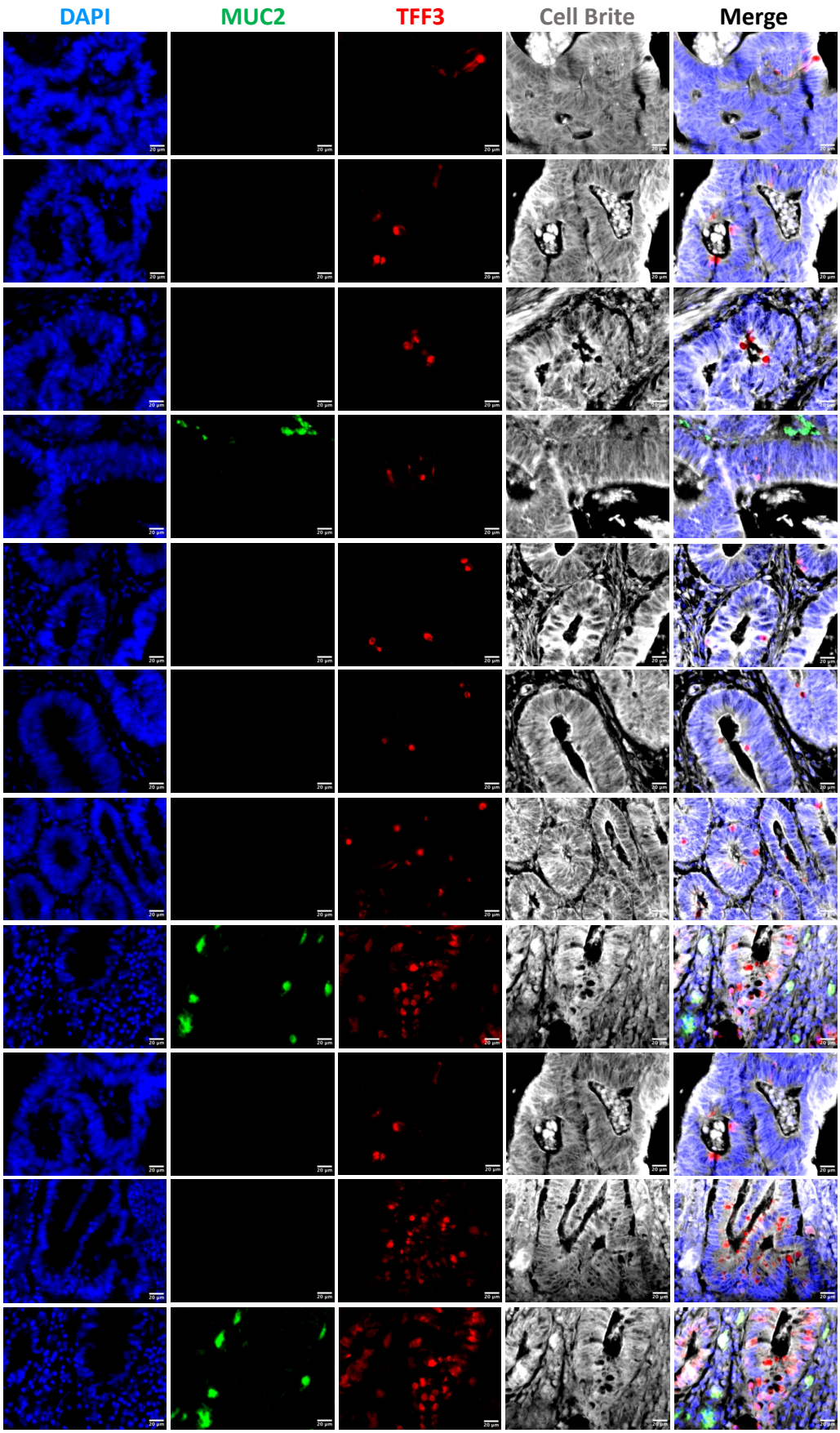
VT11294 (colorectal cancer)

B.



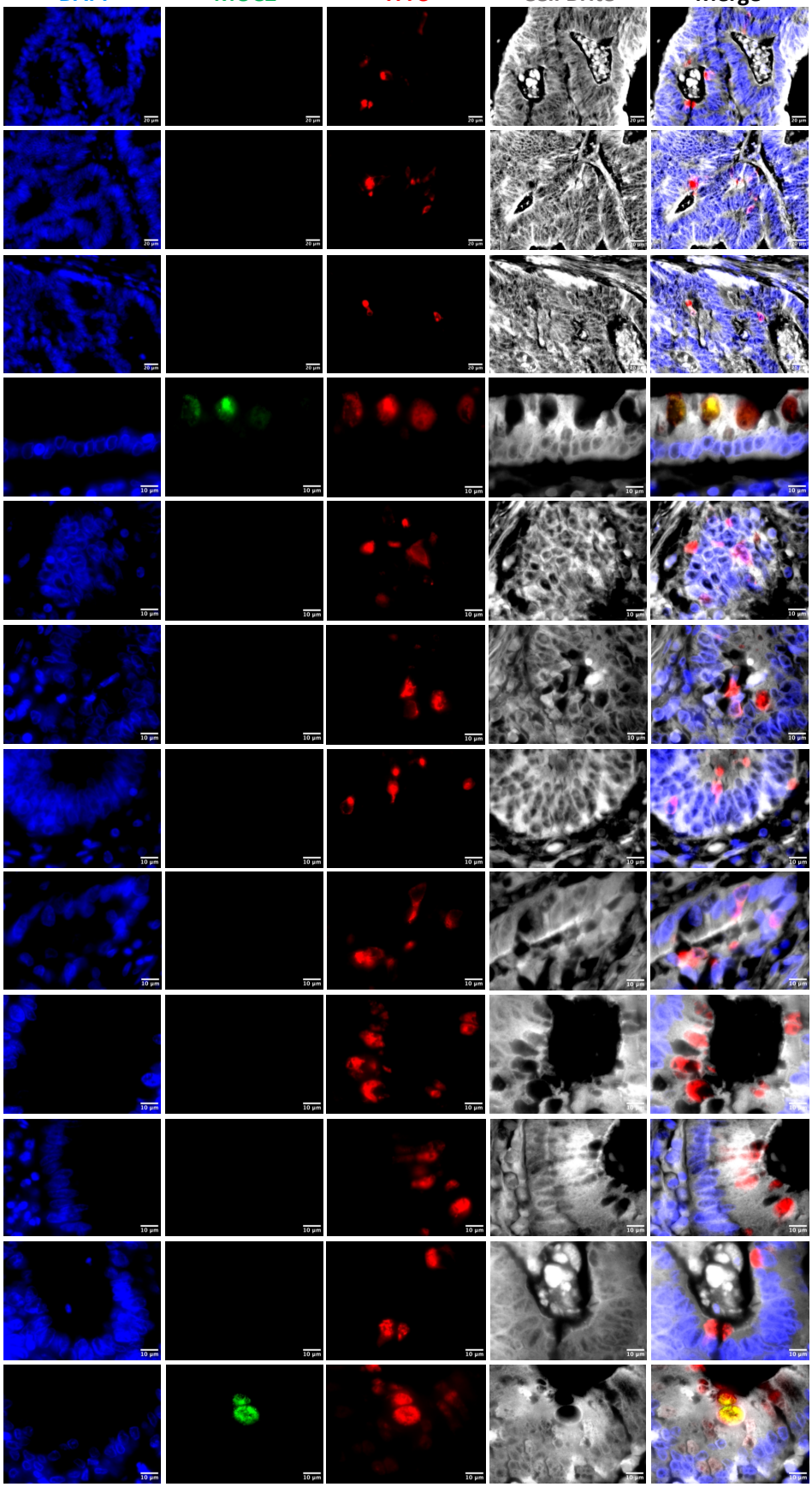
VT11294 (colorectal cancer)

B.



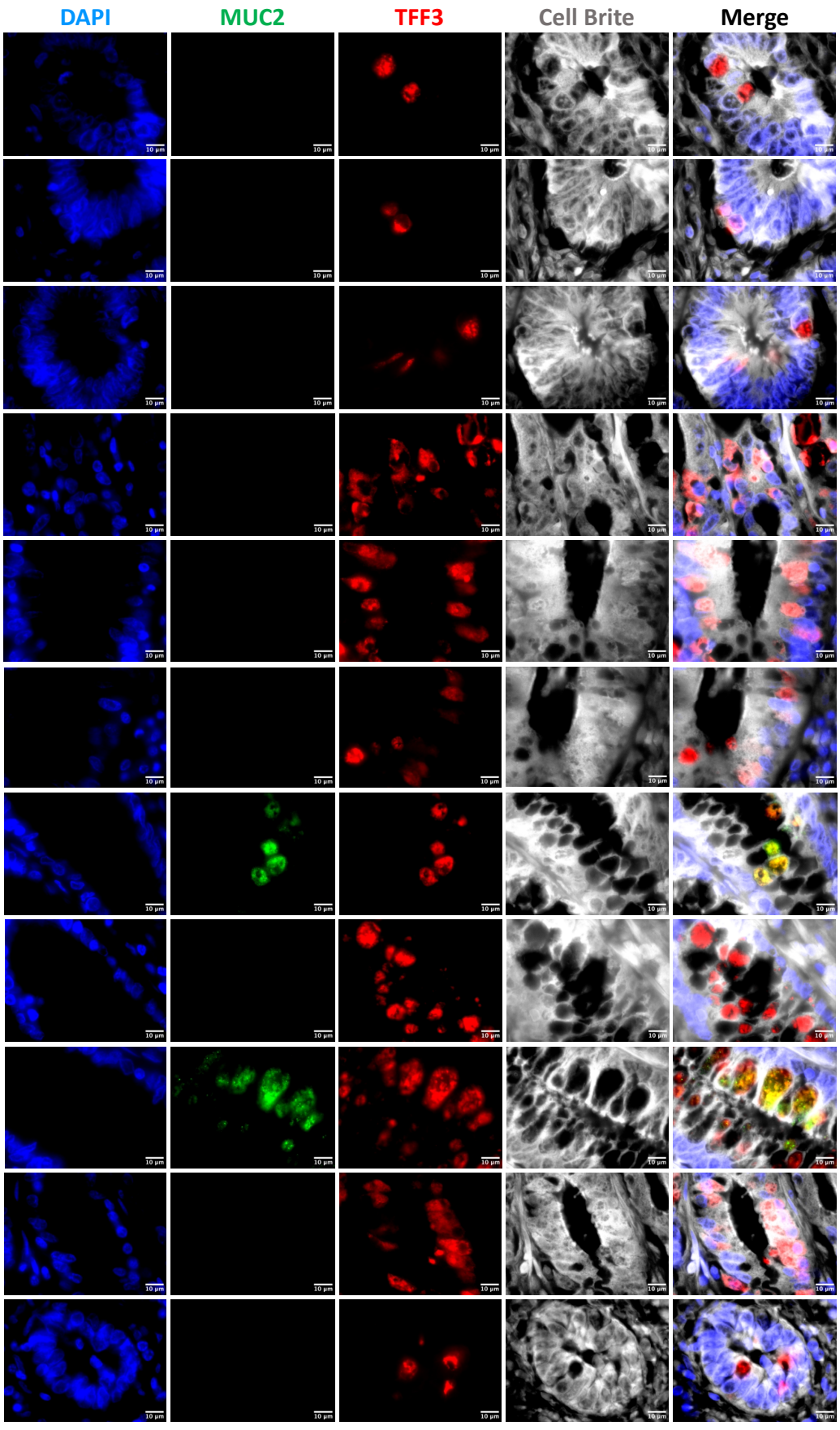
VT11294 (colorectal cancer)

B.



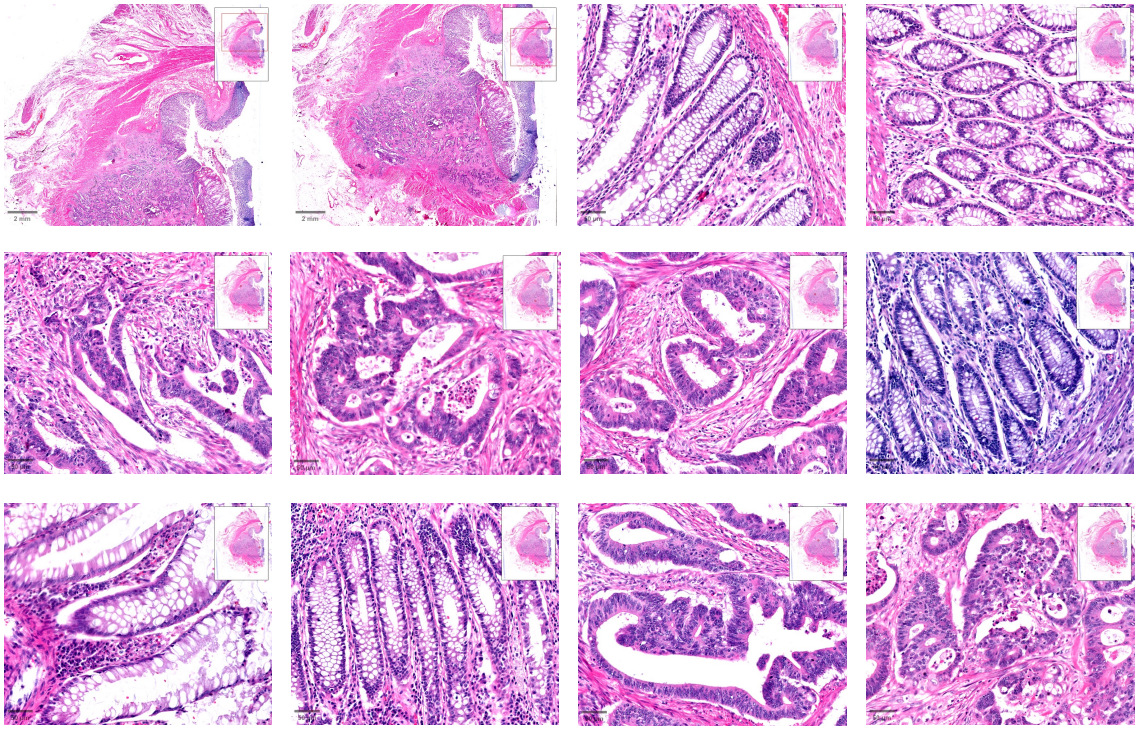
VT11294 (colorectal cancer)

B.



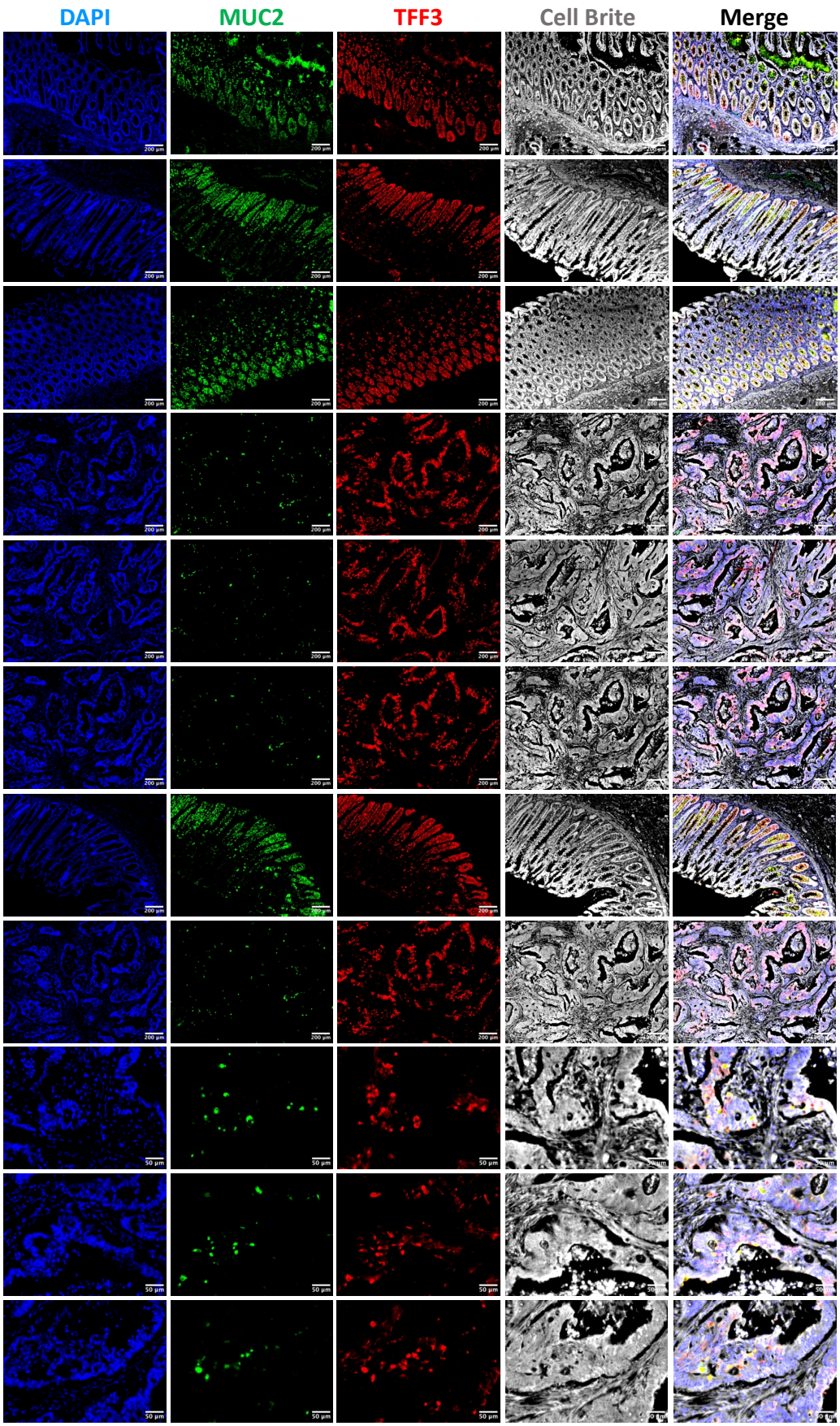
VT11384 (colorectal cancer)

A.



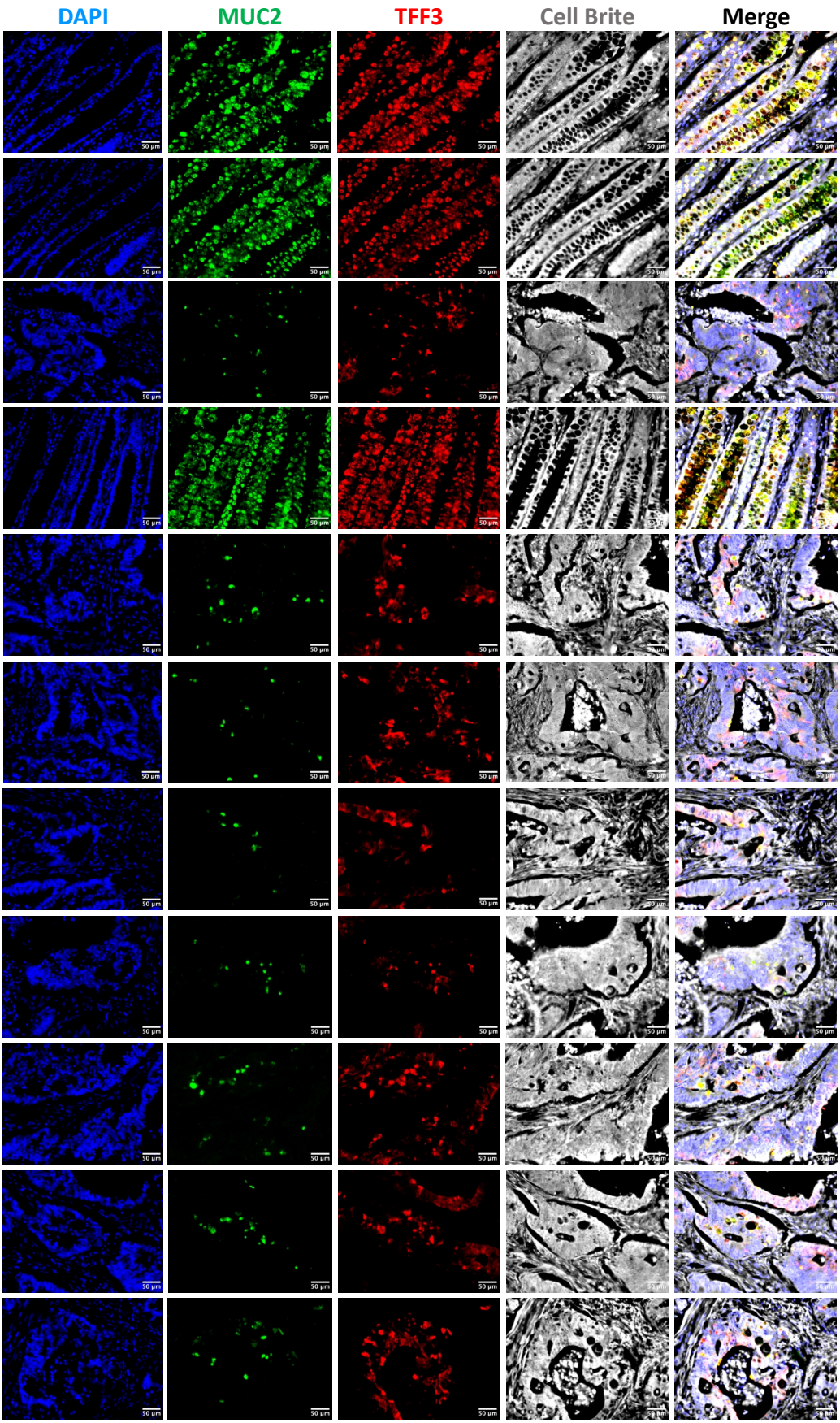
VT11384 (colorectal cancer)

B.



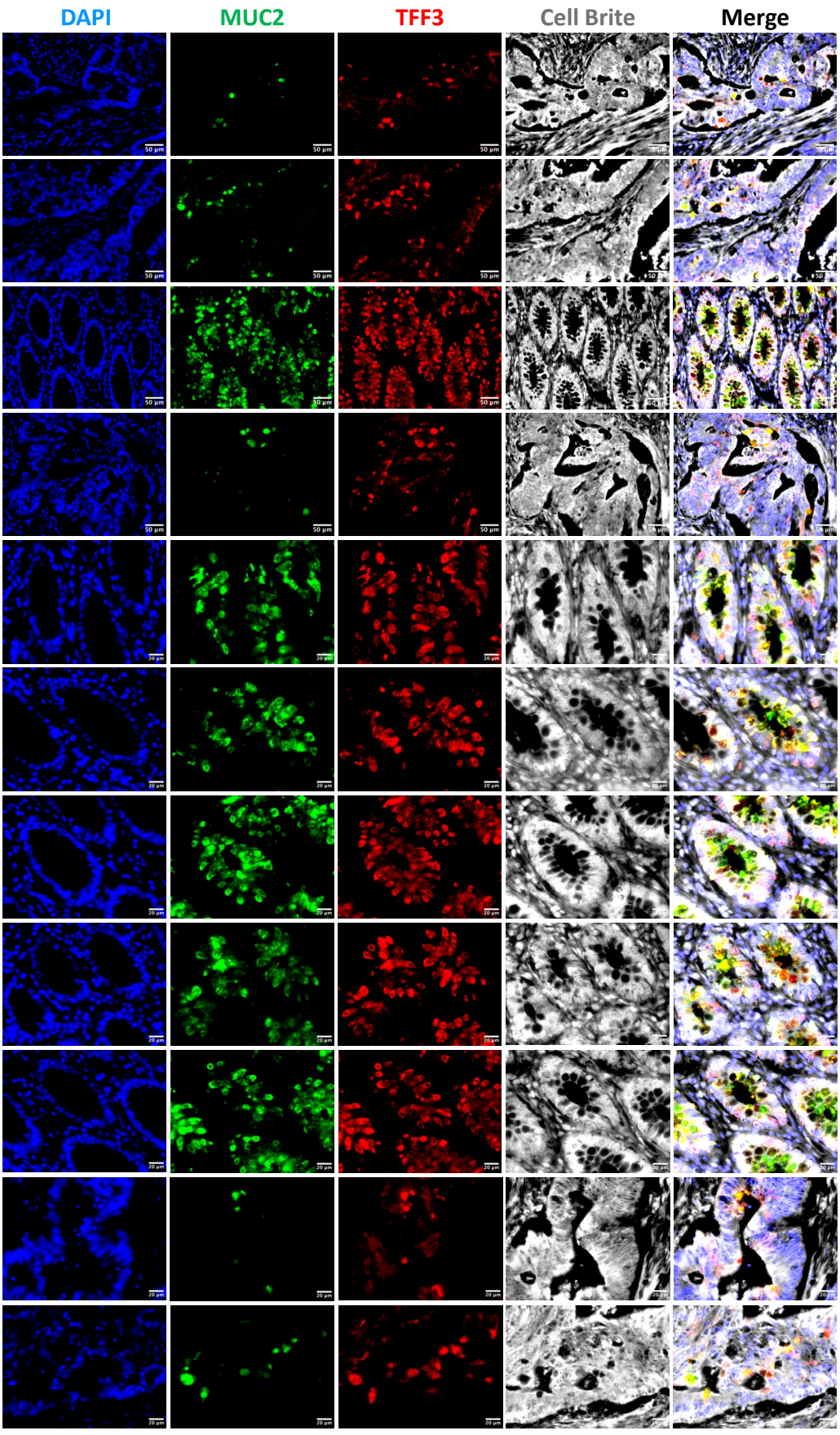
VT11384 (colorectal cancer)

B.



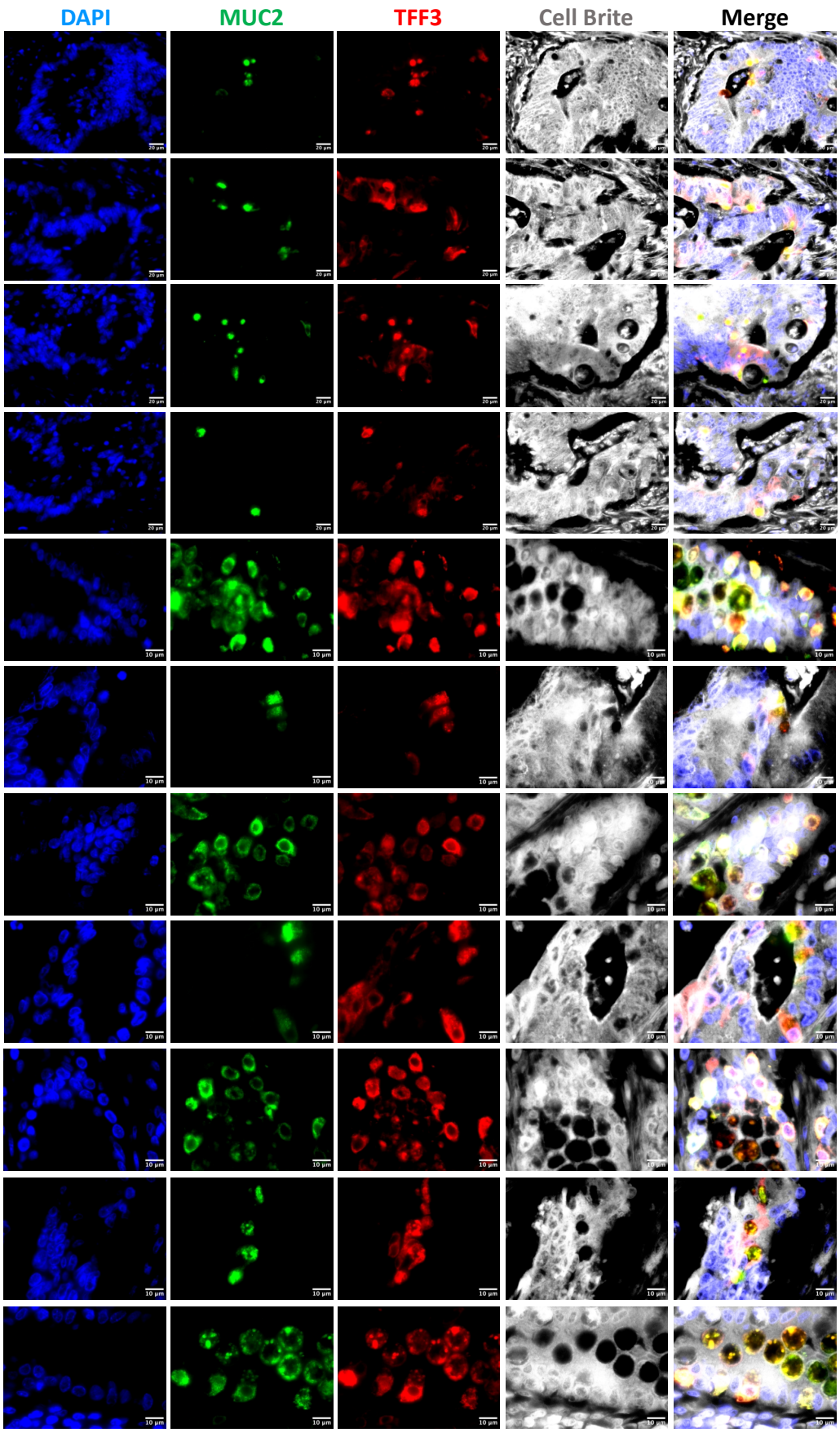
VT11384 (colorectal cancer)

B.



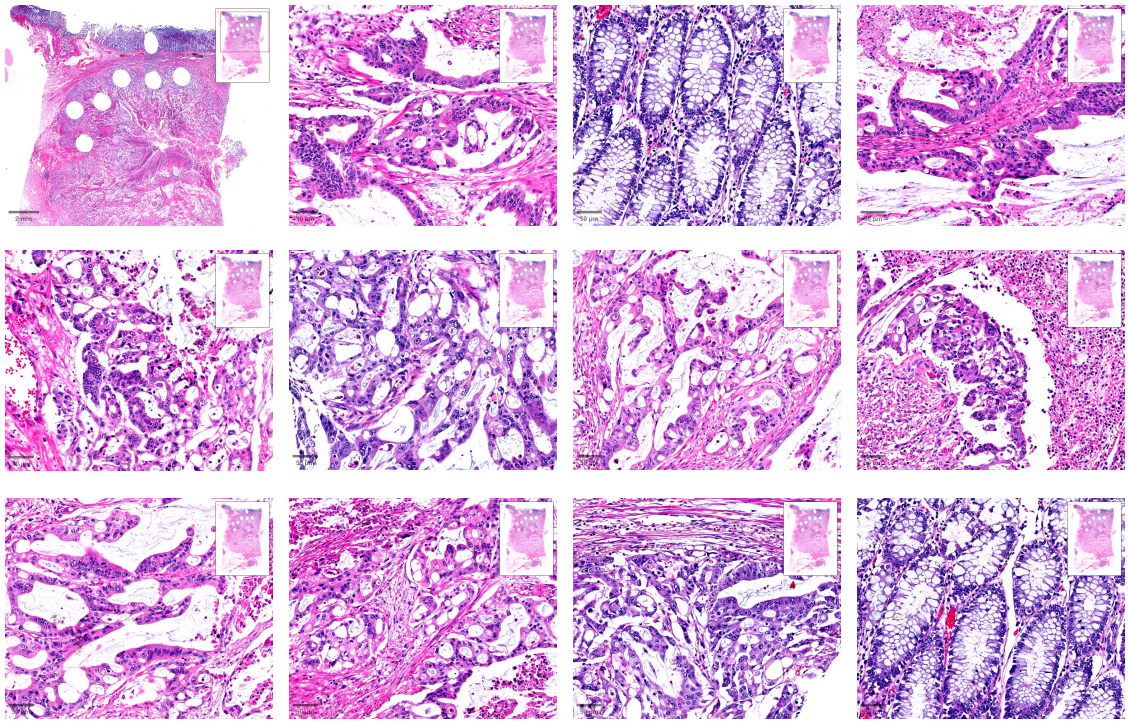
VT11384 (colorectal cancer)

B.



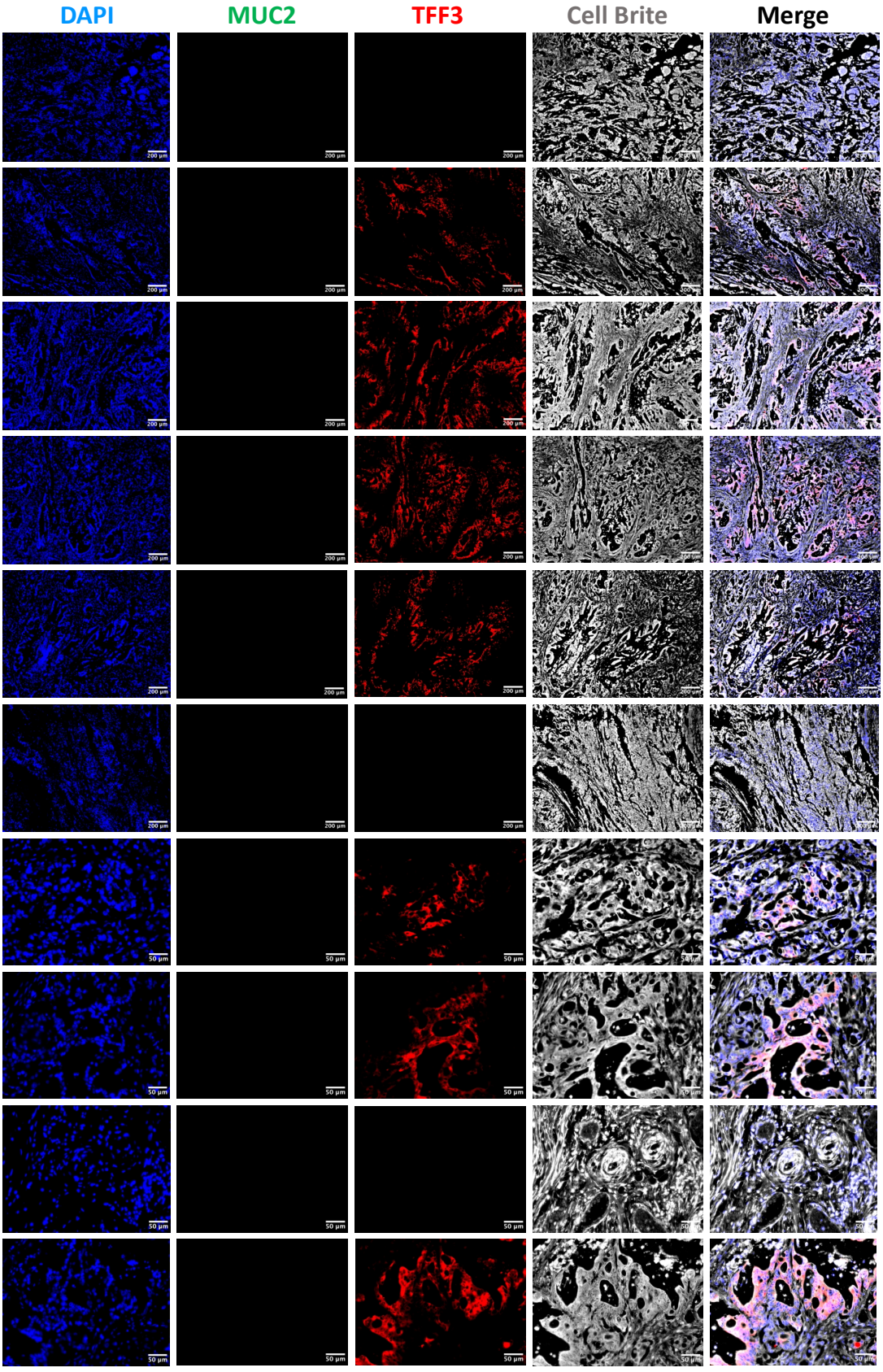
VT11385 (colorectal cancer)

A.



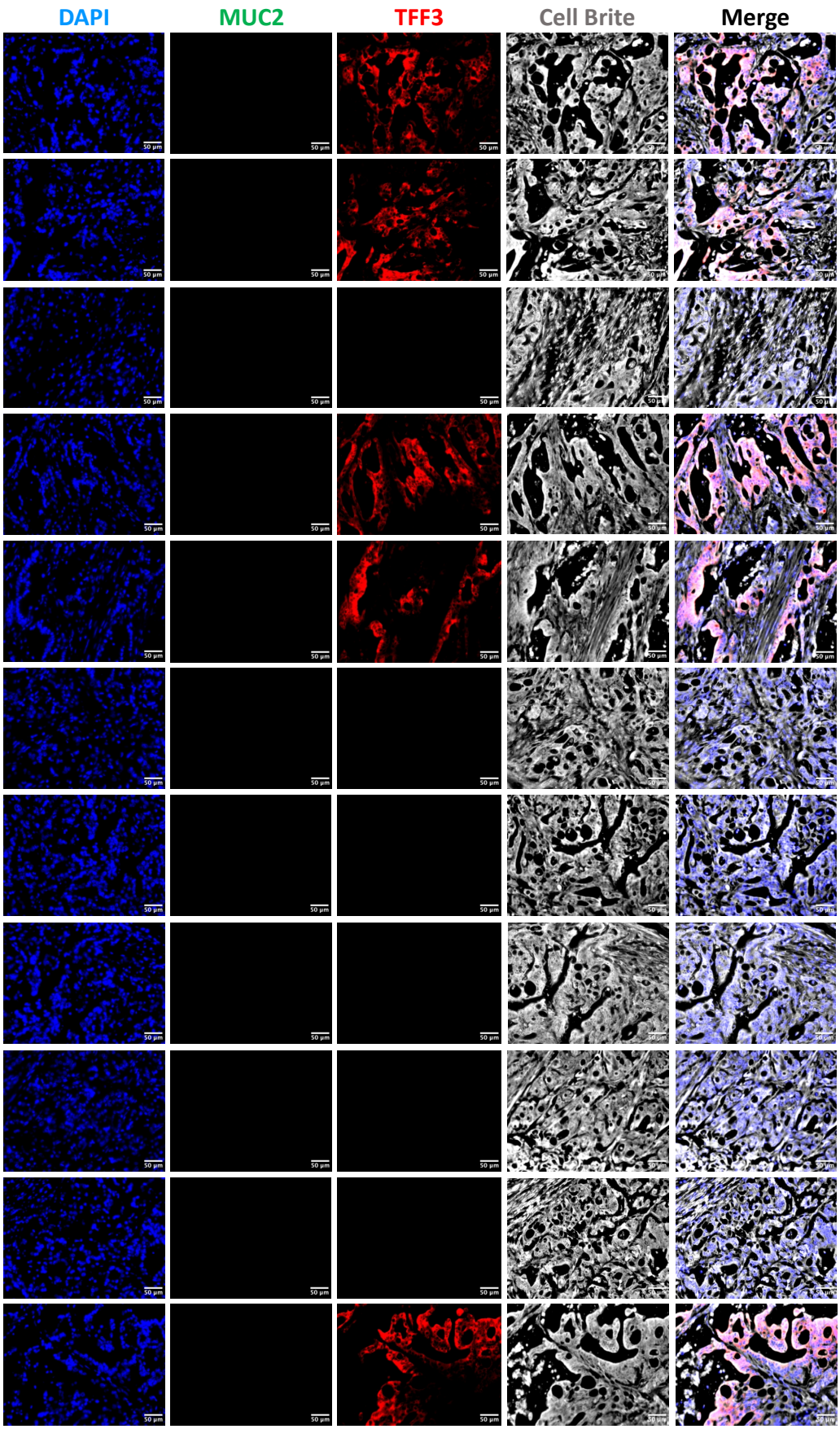
VT11385 (colorectal cancer)

B.



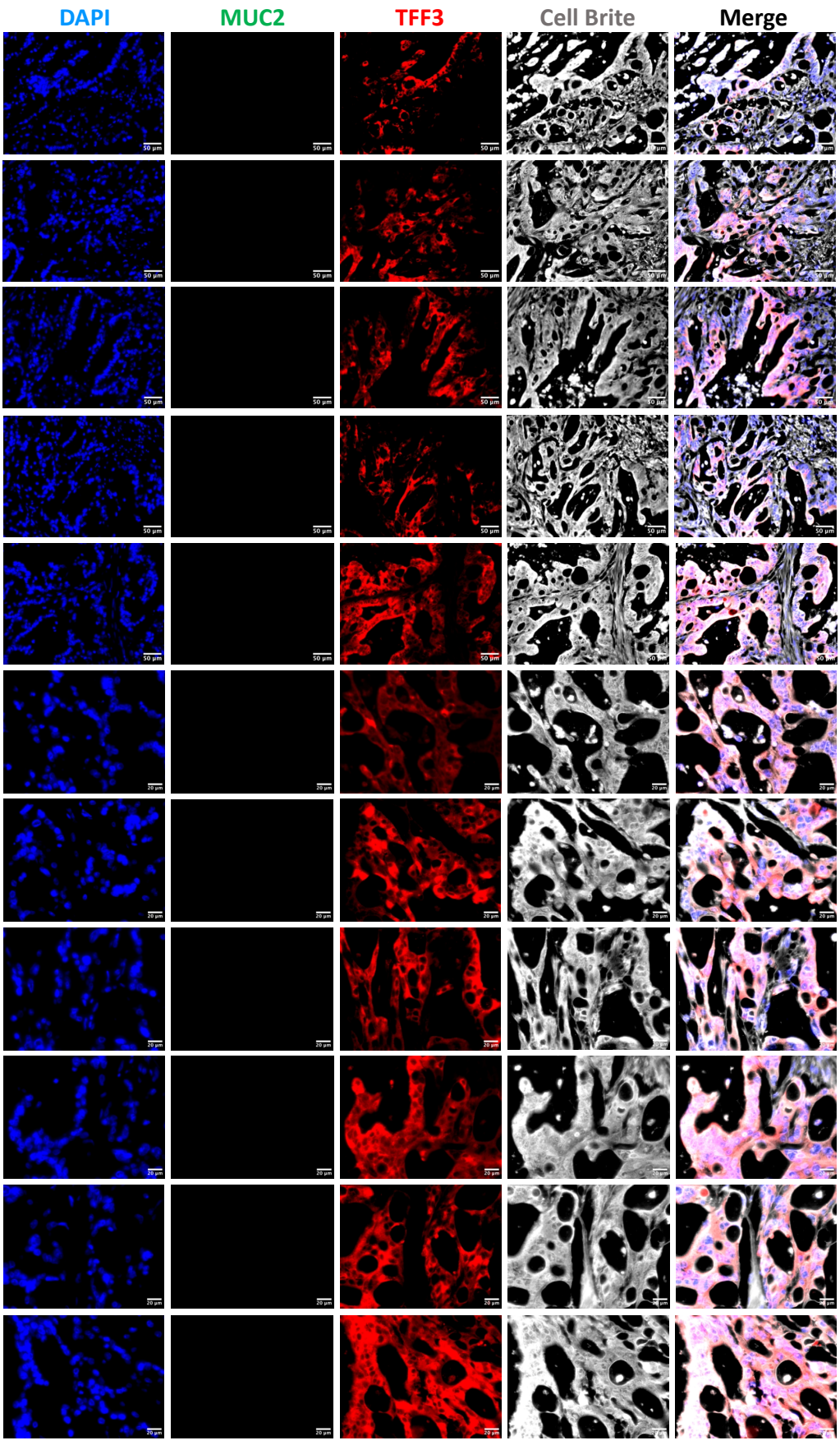
VT11385 (colorectal cancer)

B.



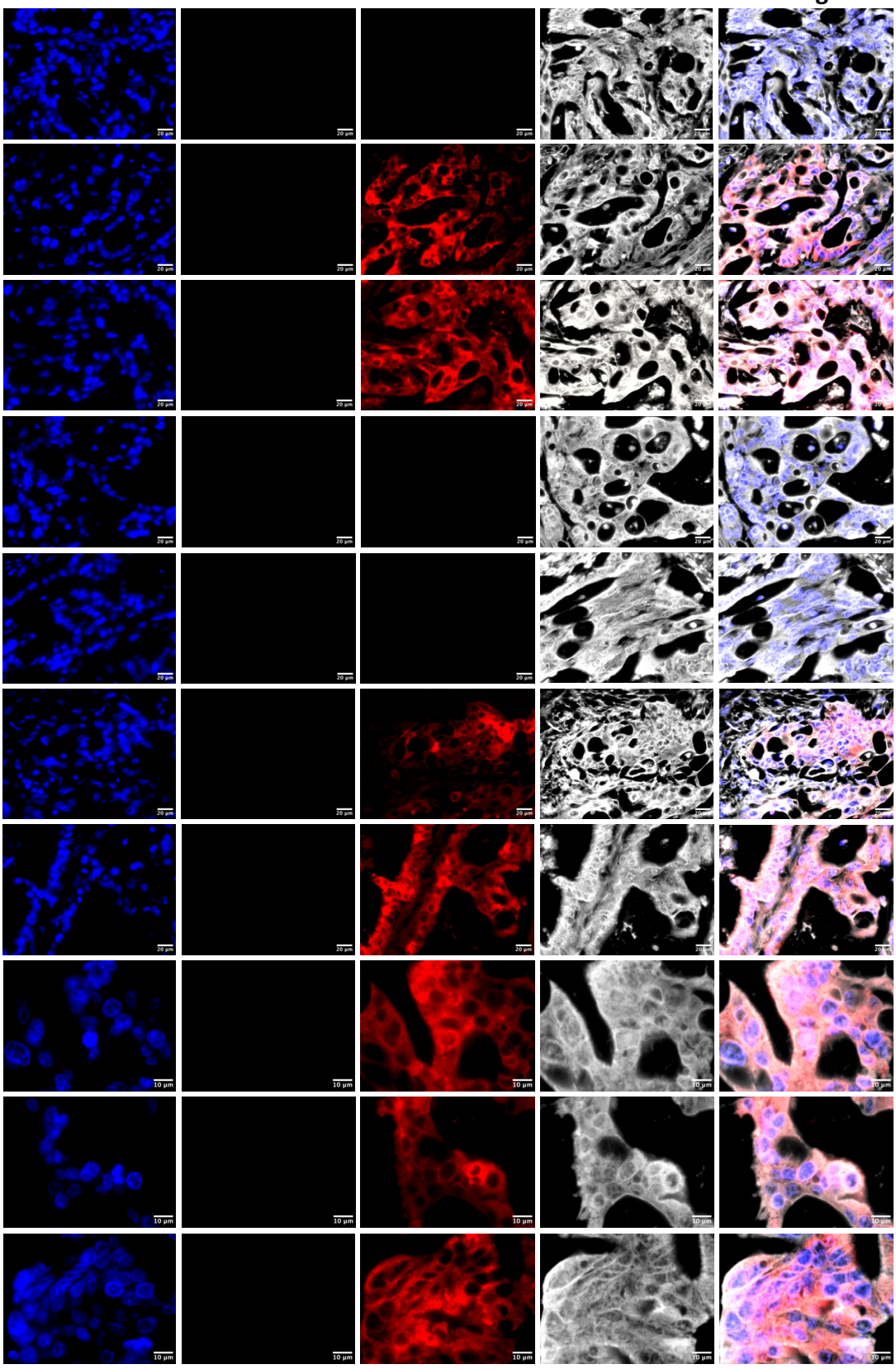
VT11385 (colorectal cancer)

B.



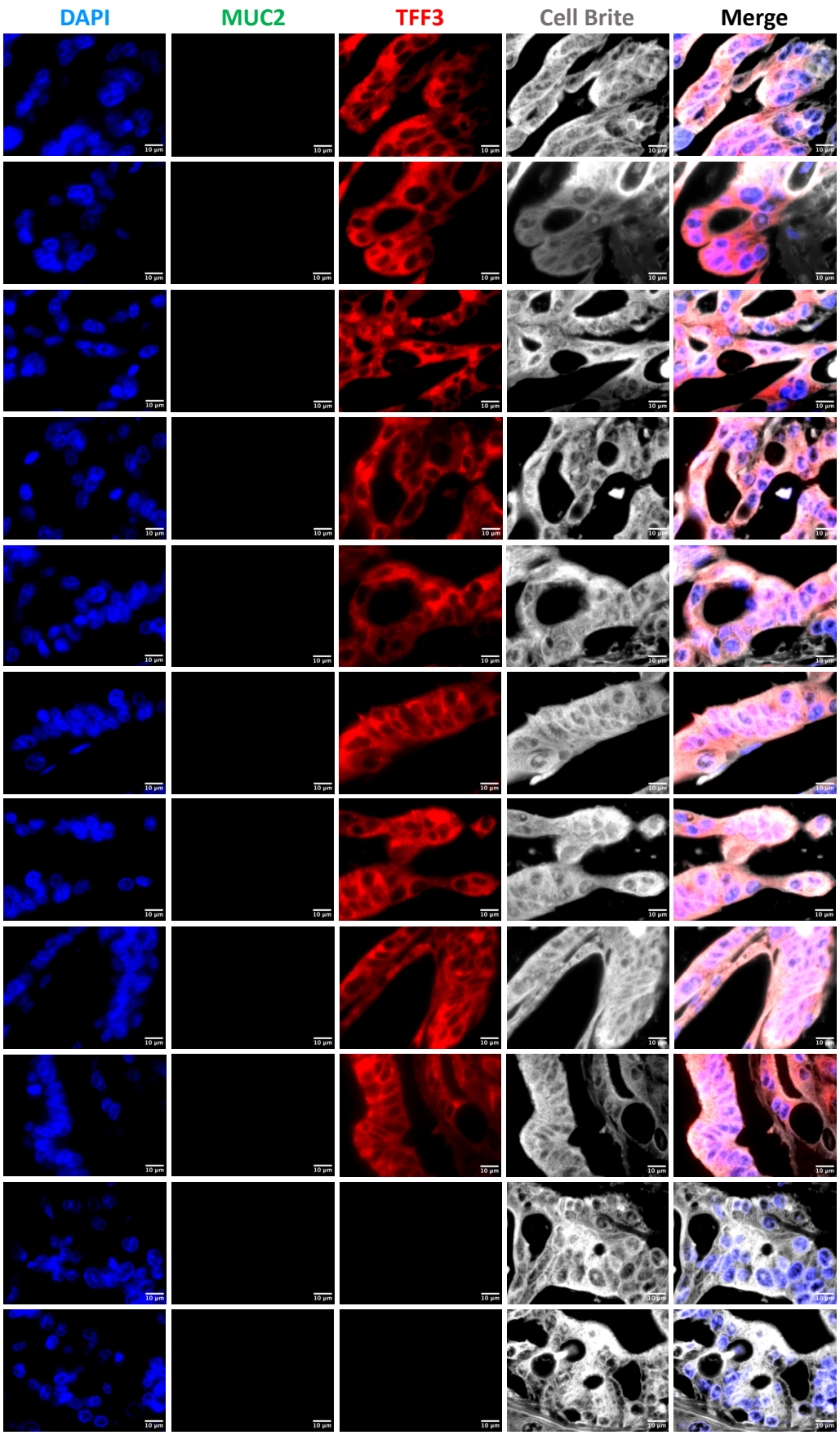
VT11385 (colorectal cancer)

B.



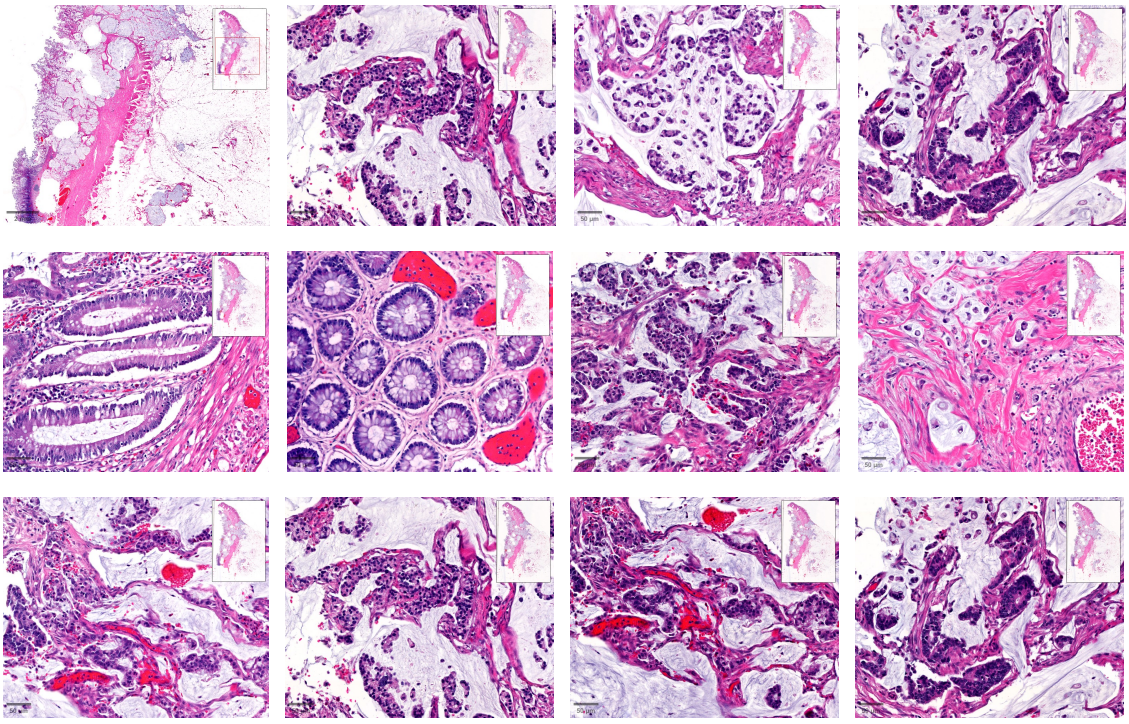
VT11385 (colorectal cancer)

B.



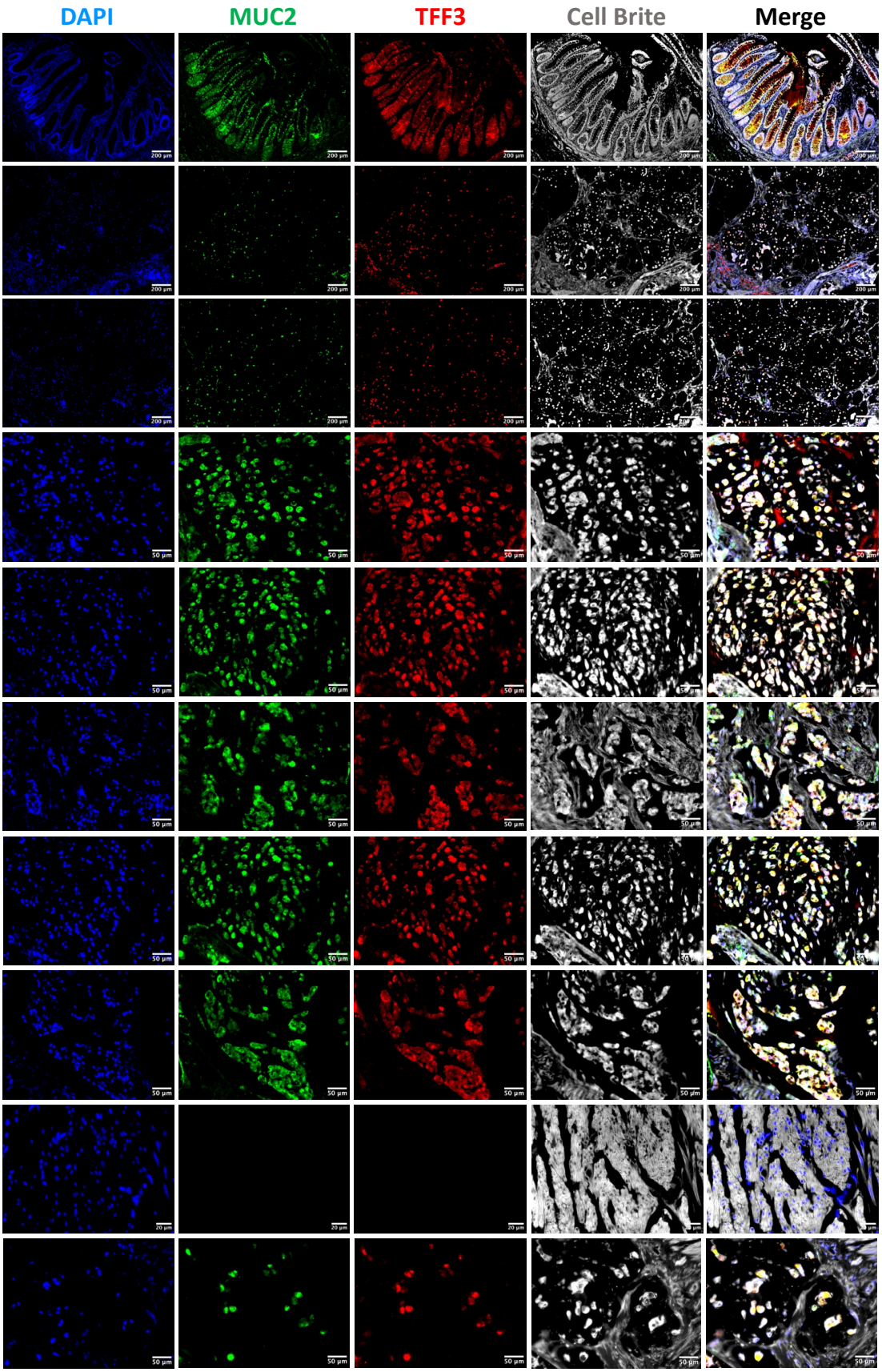
VT11386 (colorectal cancer)

A.



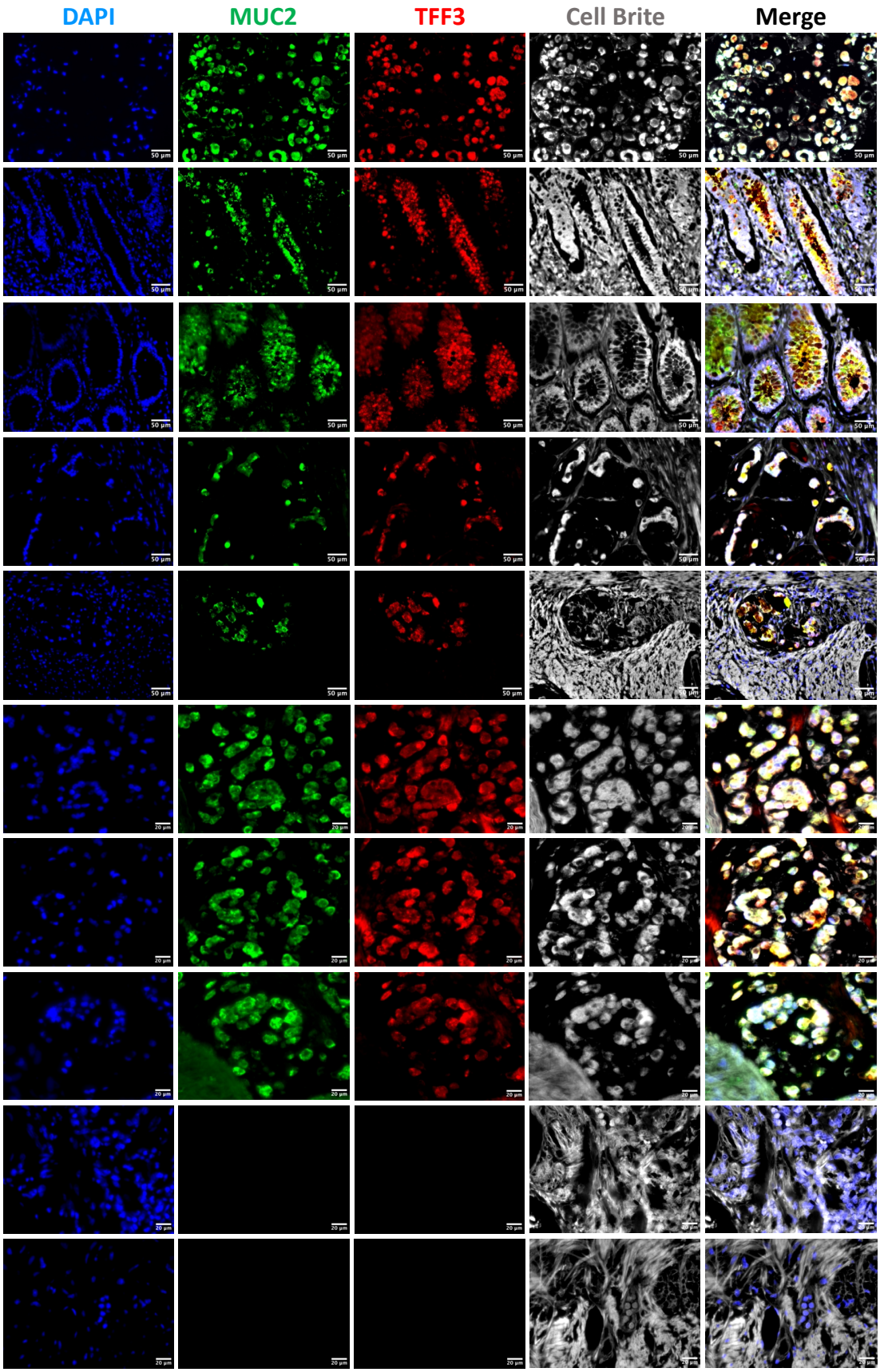
VT11386 (colorectal cancer)

B.



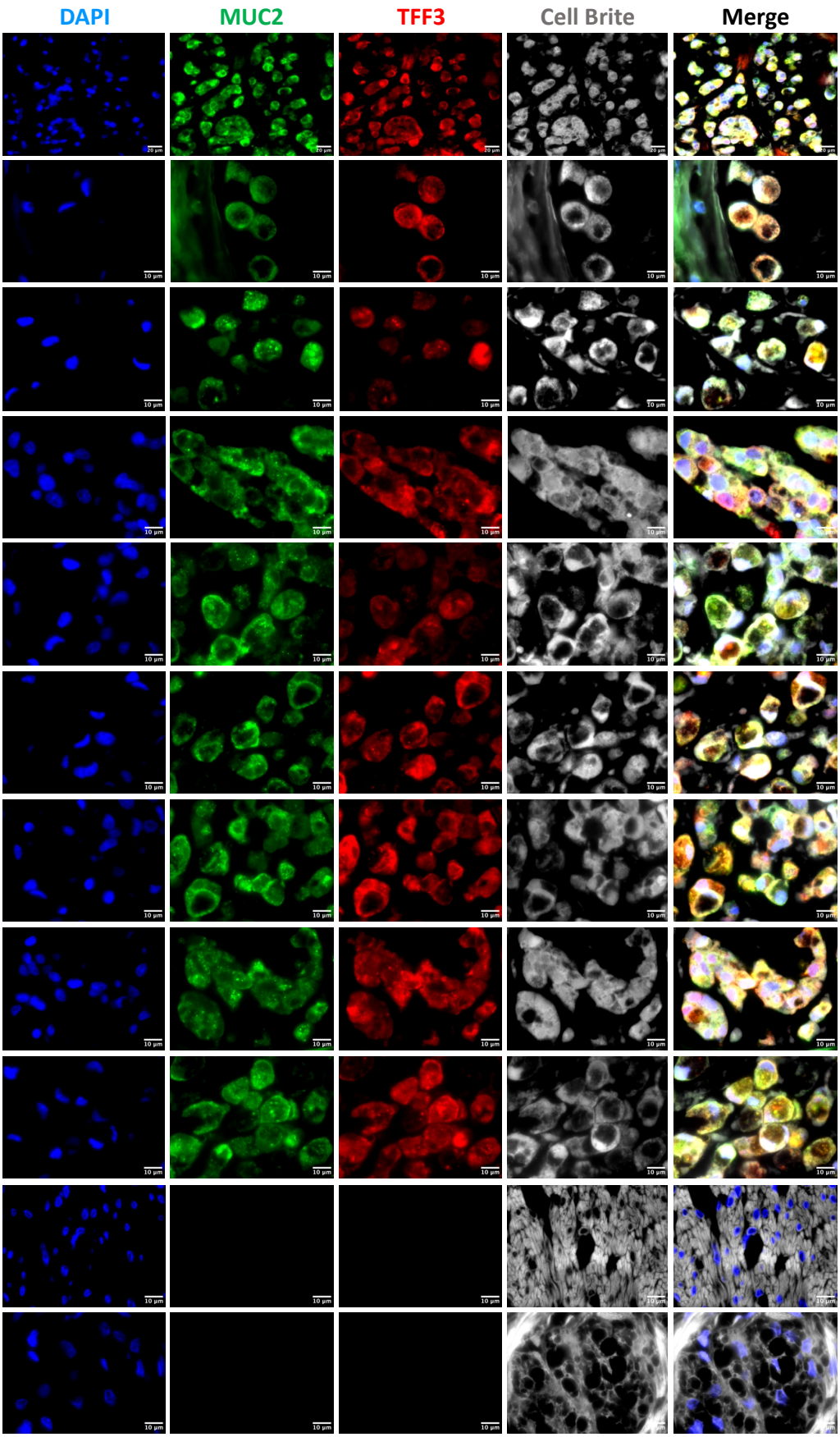
VT11386 (colorectal cancer)

B.



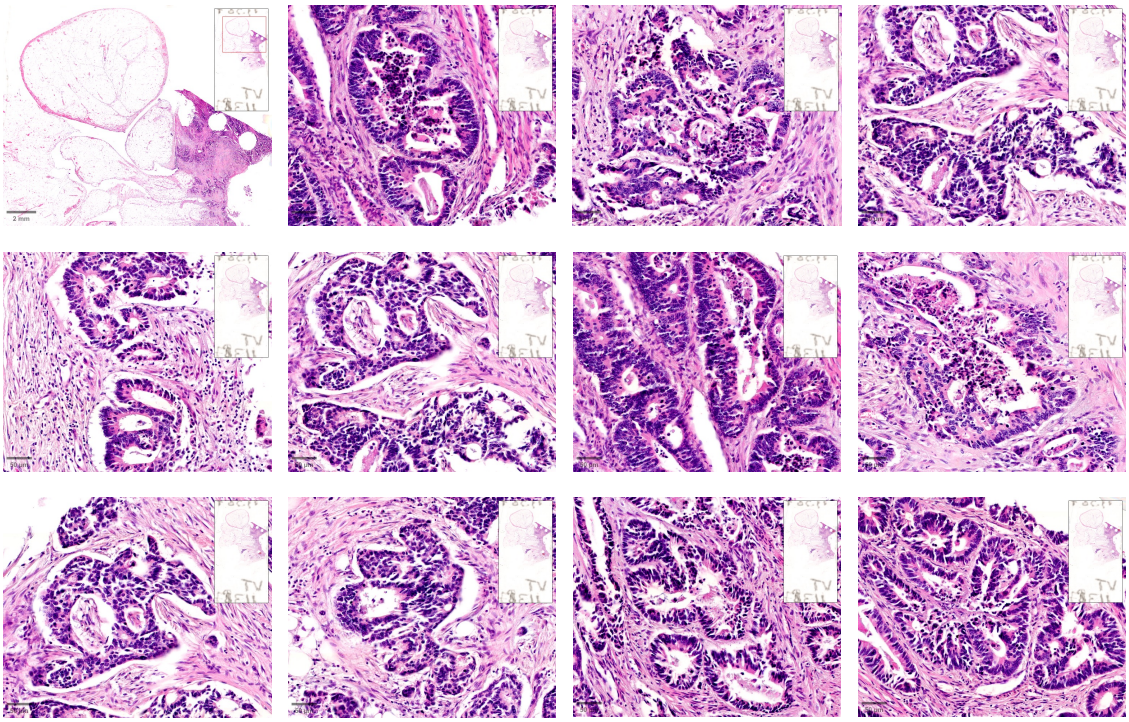
VT11386 (colorectal cancer)

B.



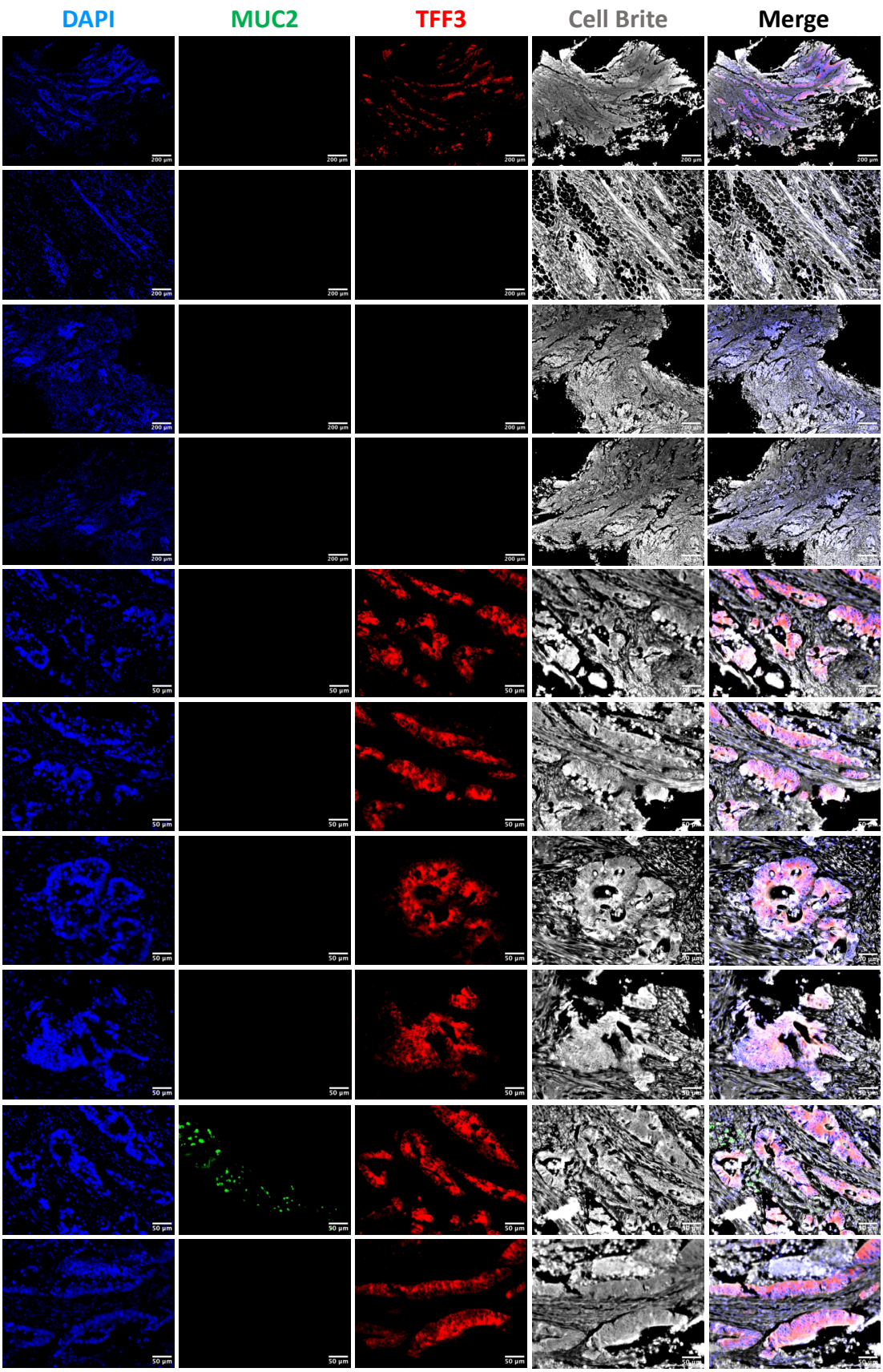
VT11387 (colorectal cancer)

A.



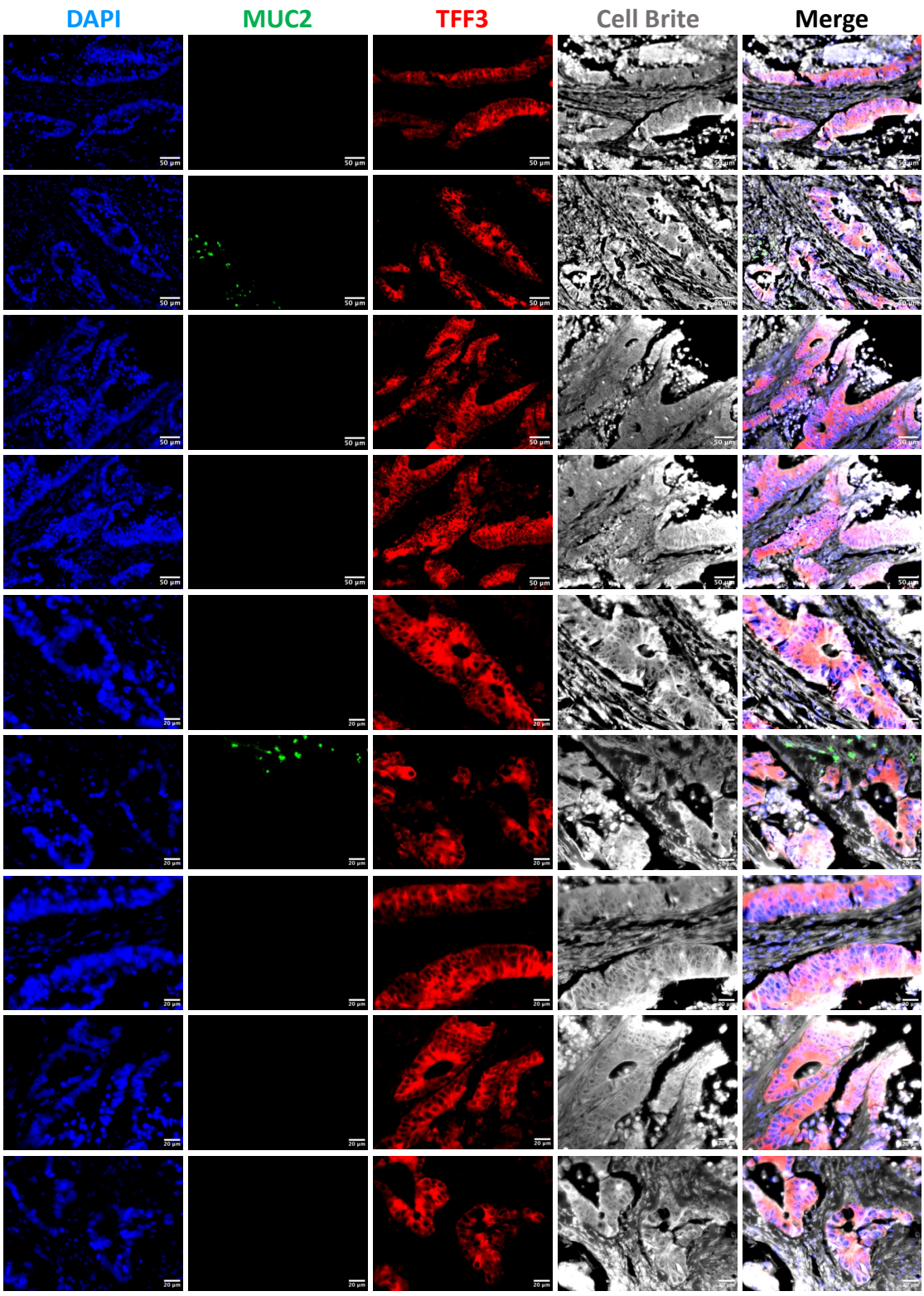
VT11387 (colorectal cancer)

B.



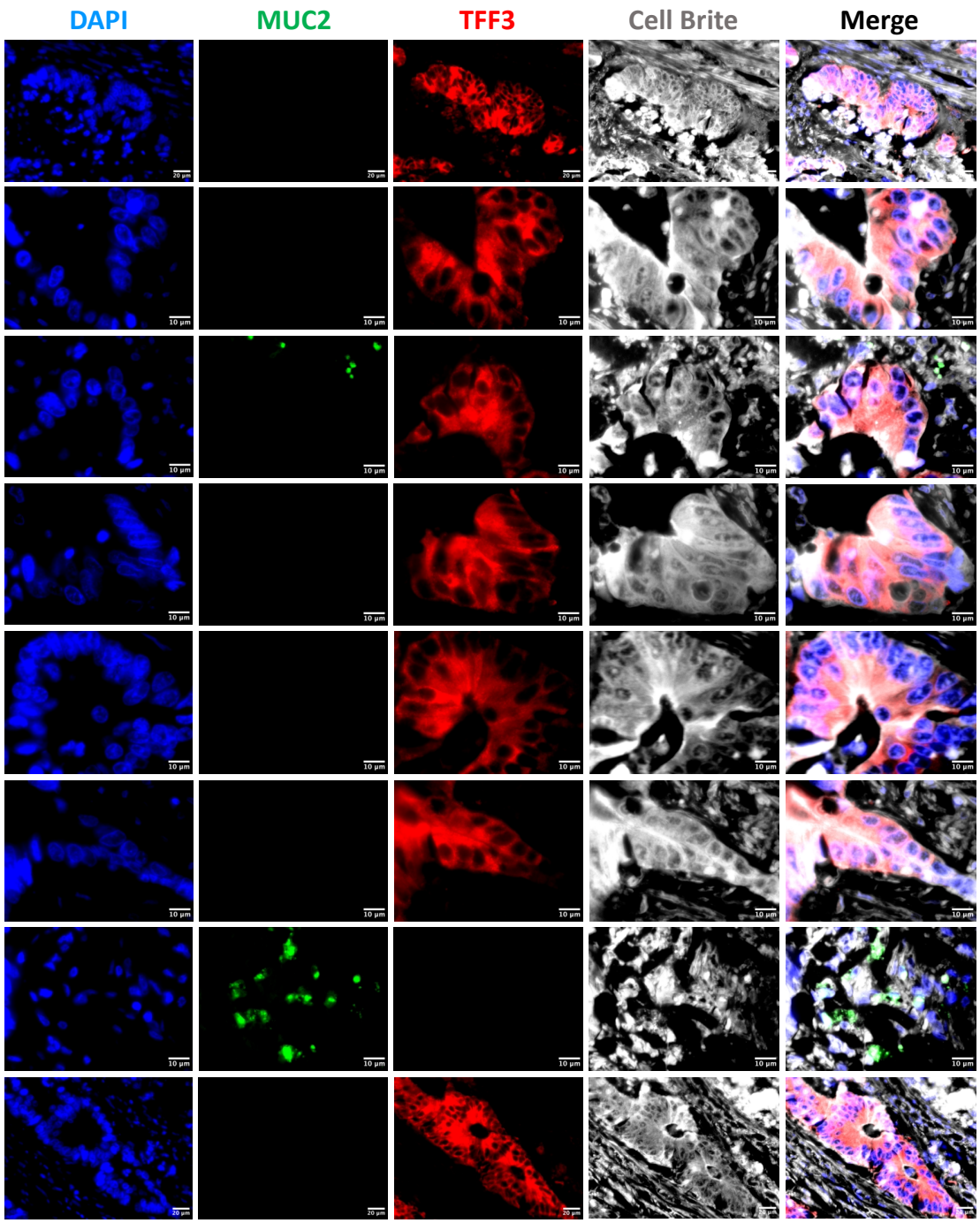
VT11387 (colorectal cancer)

B.



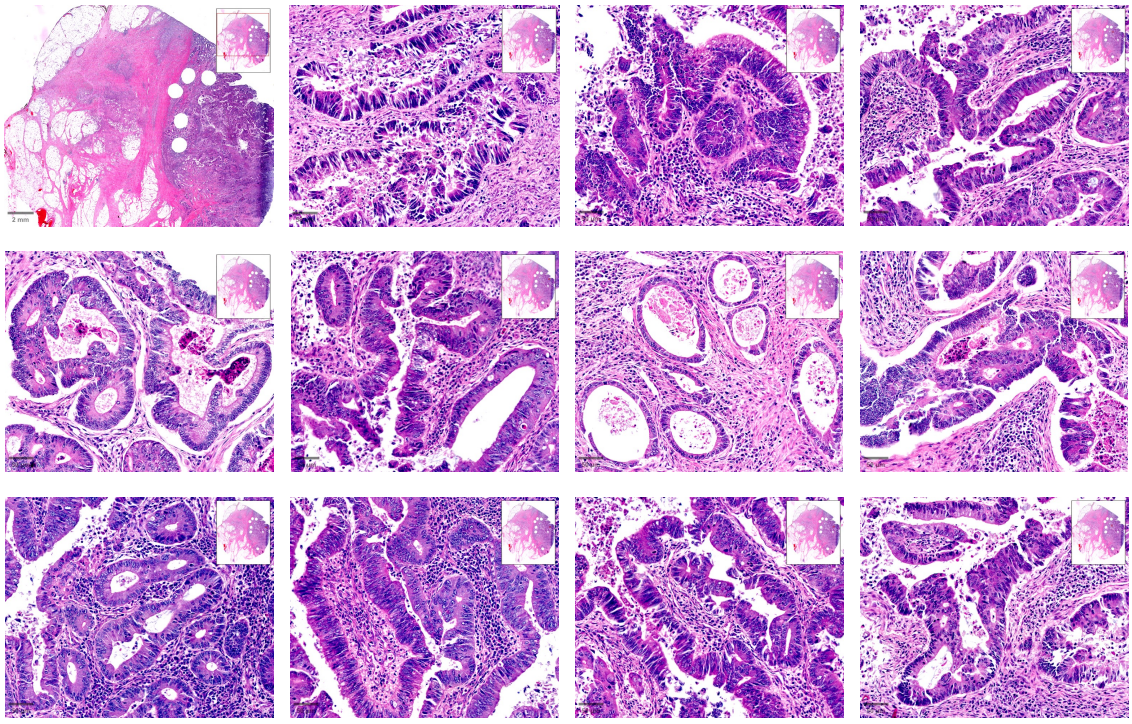
VT11387 (colorectal cancer)

B.



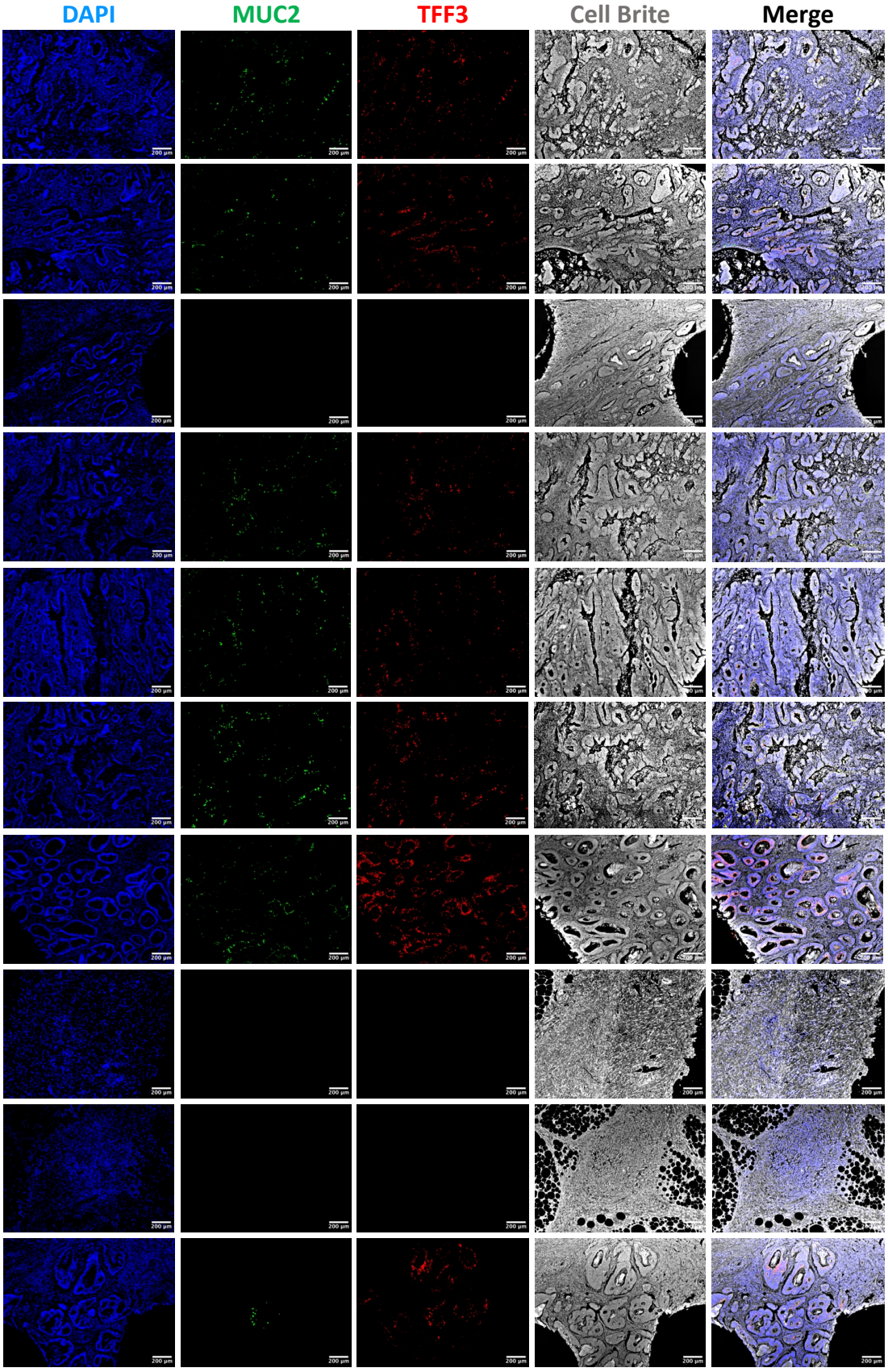
VT11389 (colorectal cancer)

A.



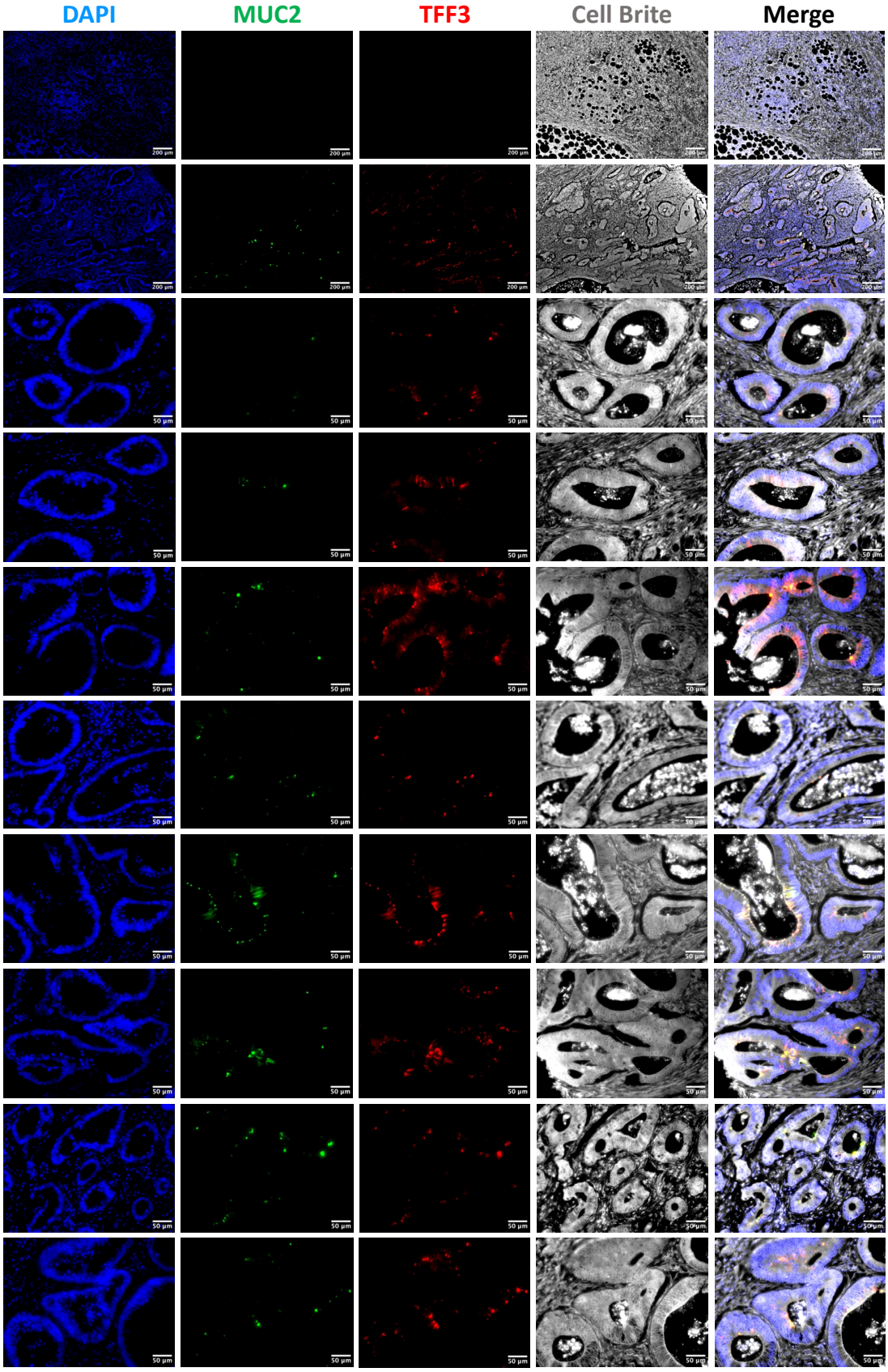
VT11389 (colorectal cancer)

B.



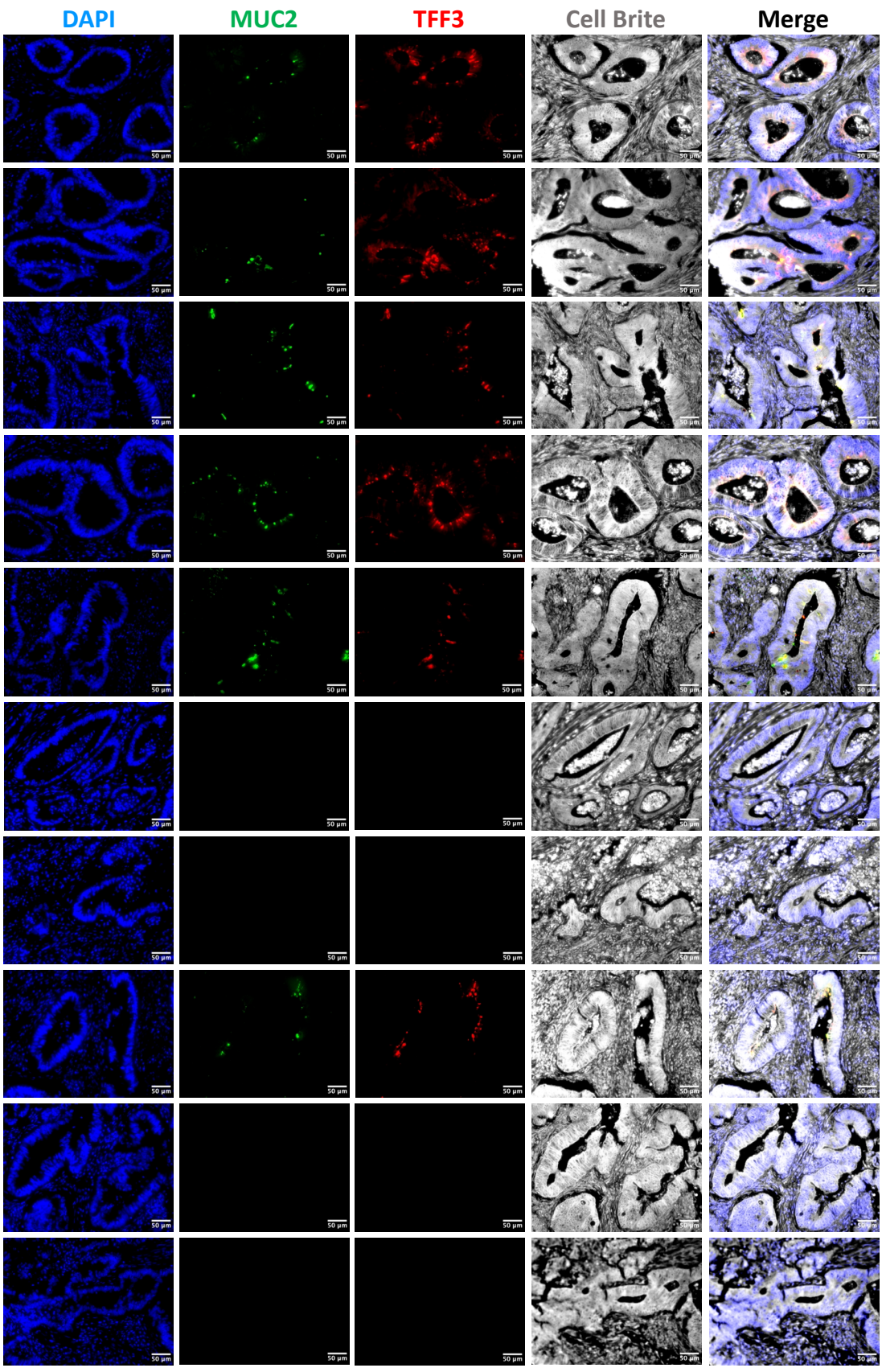
VT11389 (colorectal cancer)

B.



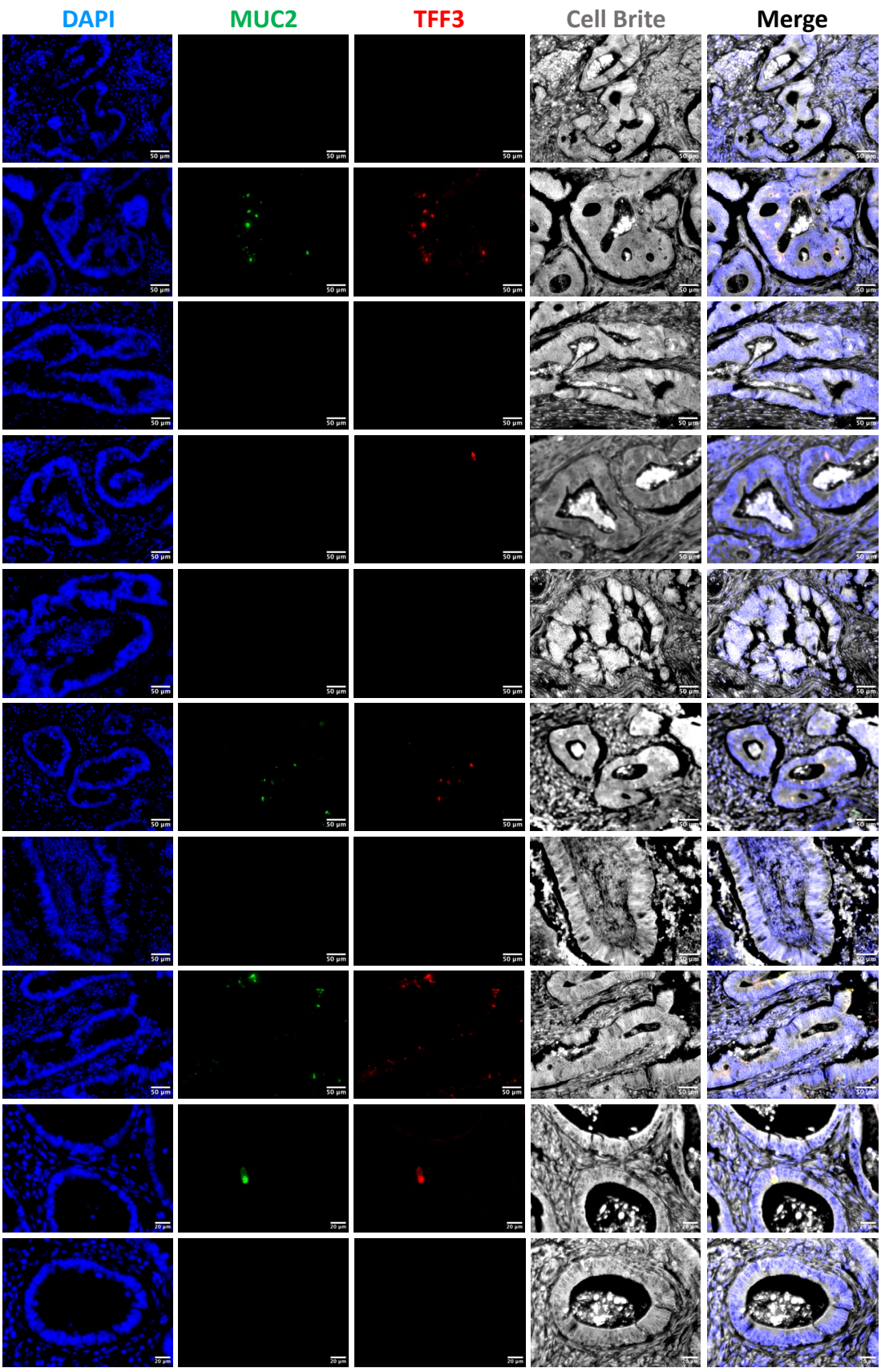
VT11389 (colorectal cancer)

B.



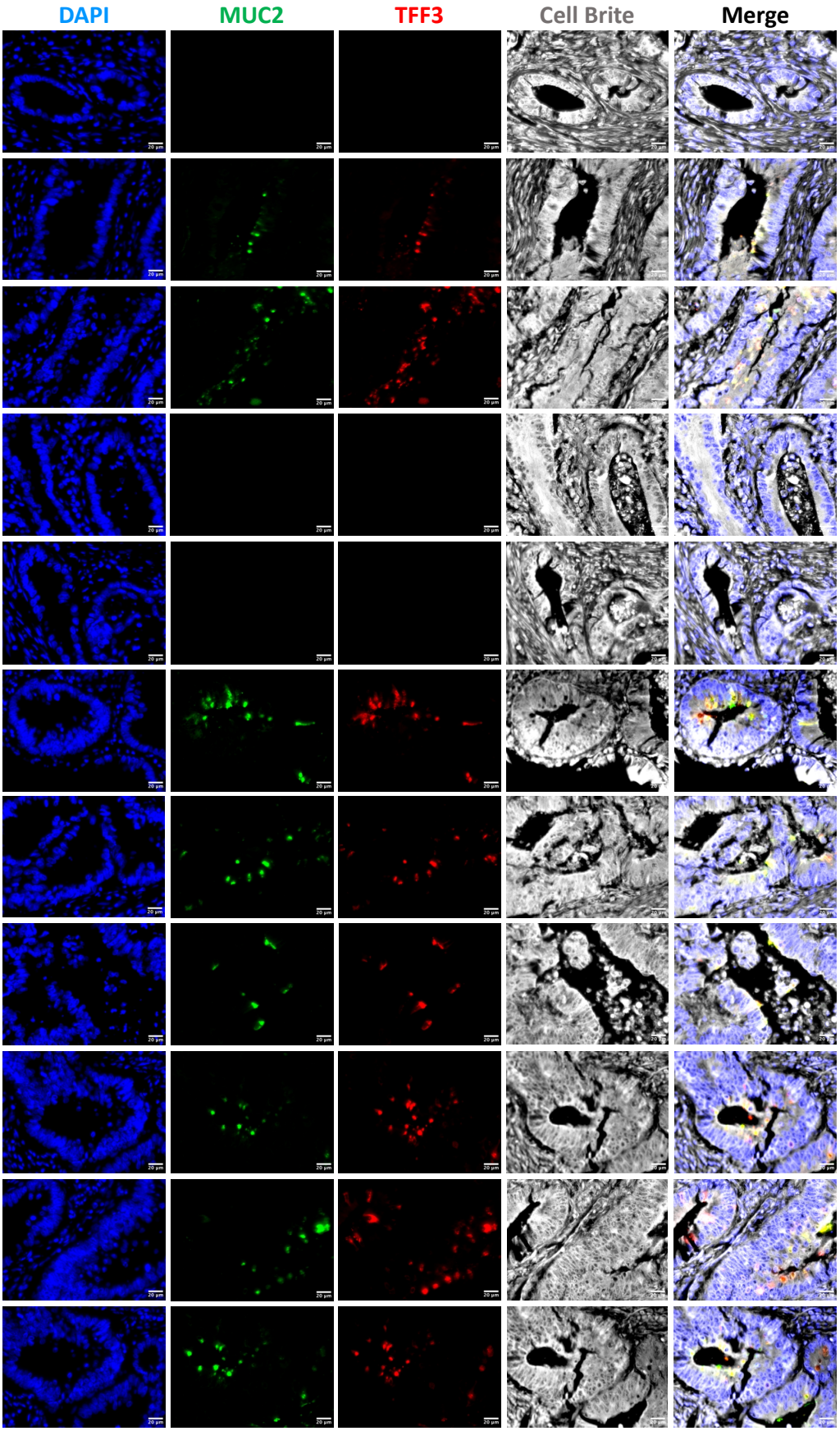
VT11389 (colorectal cancer)

B.



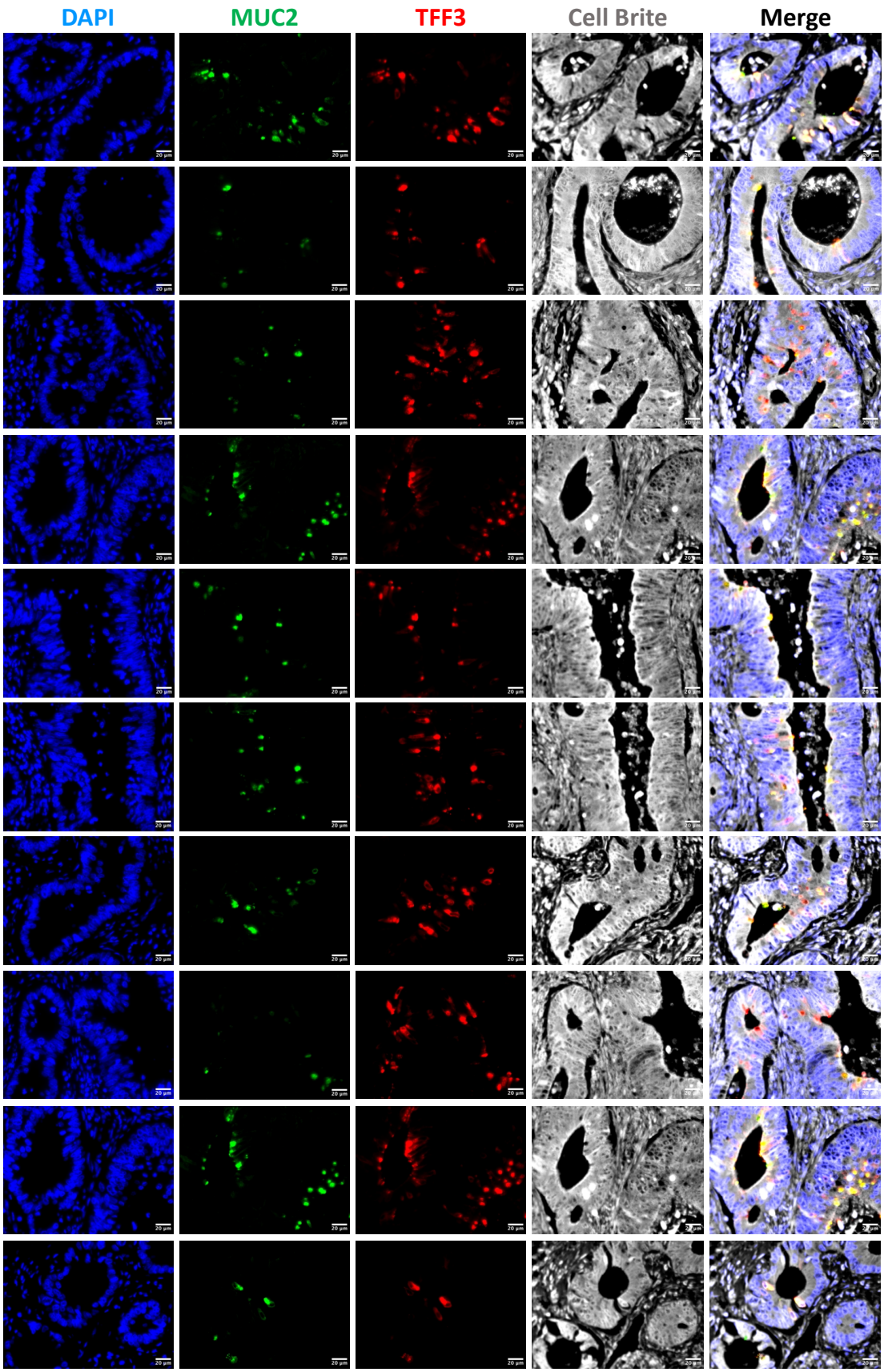
VT11389 (colorectal cancer)

B.



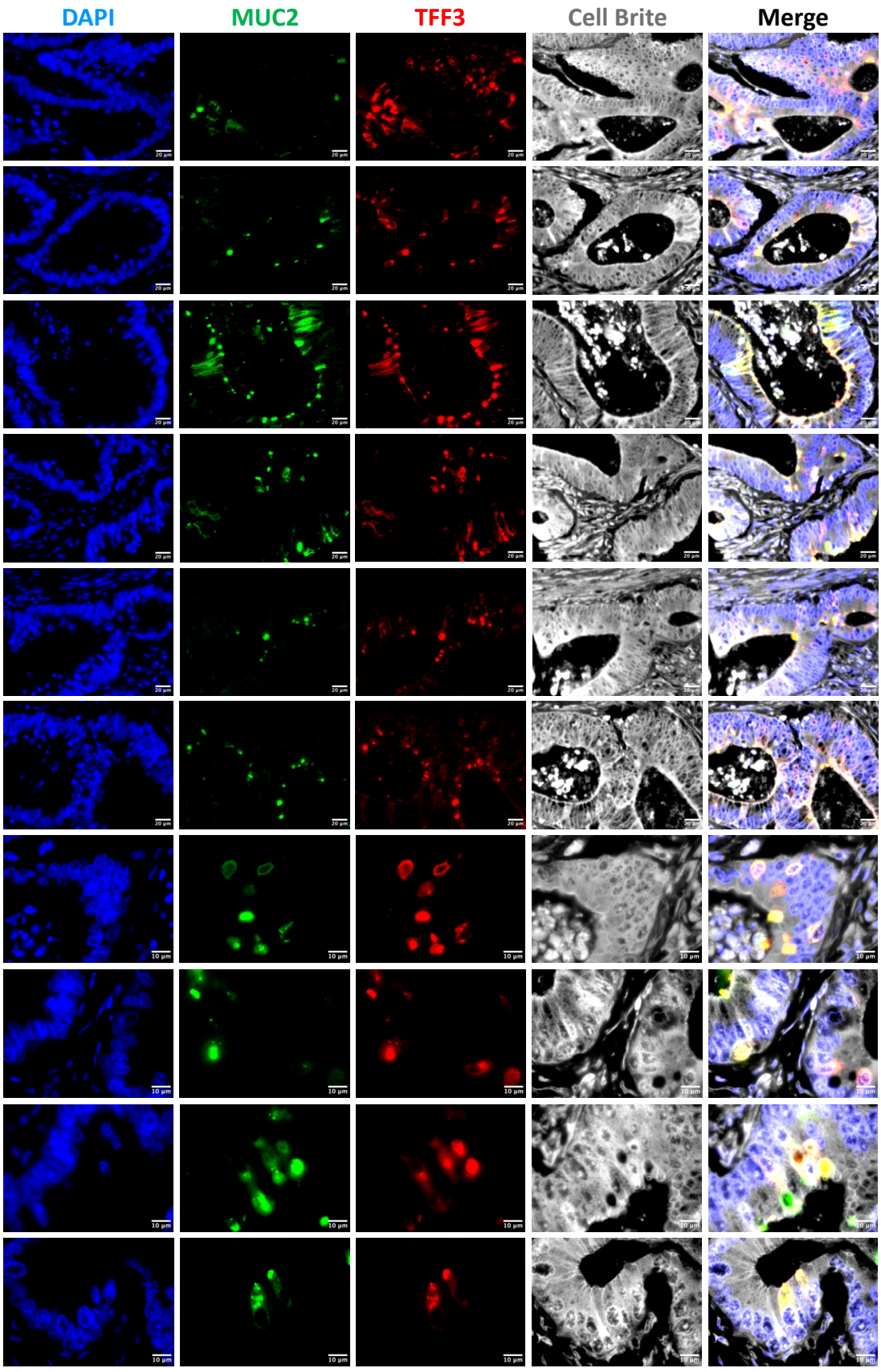
VT11389 (colorectal cancer)

B.



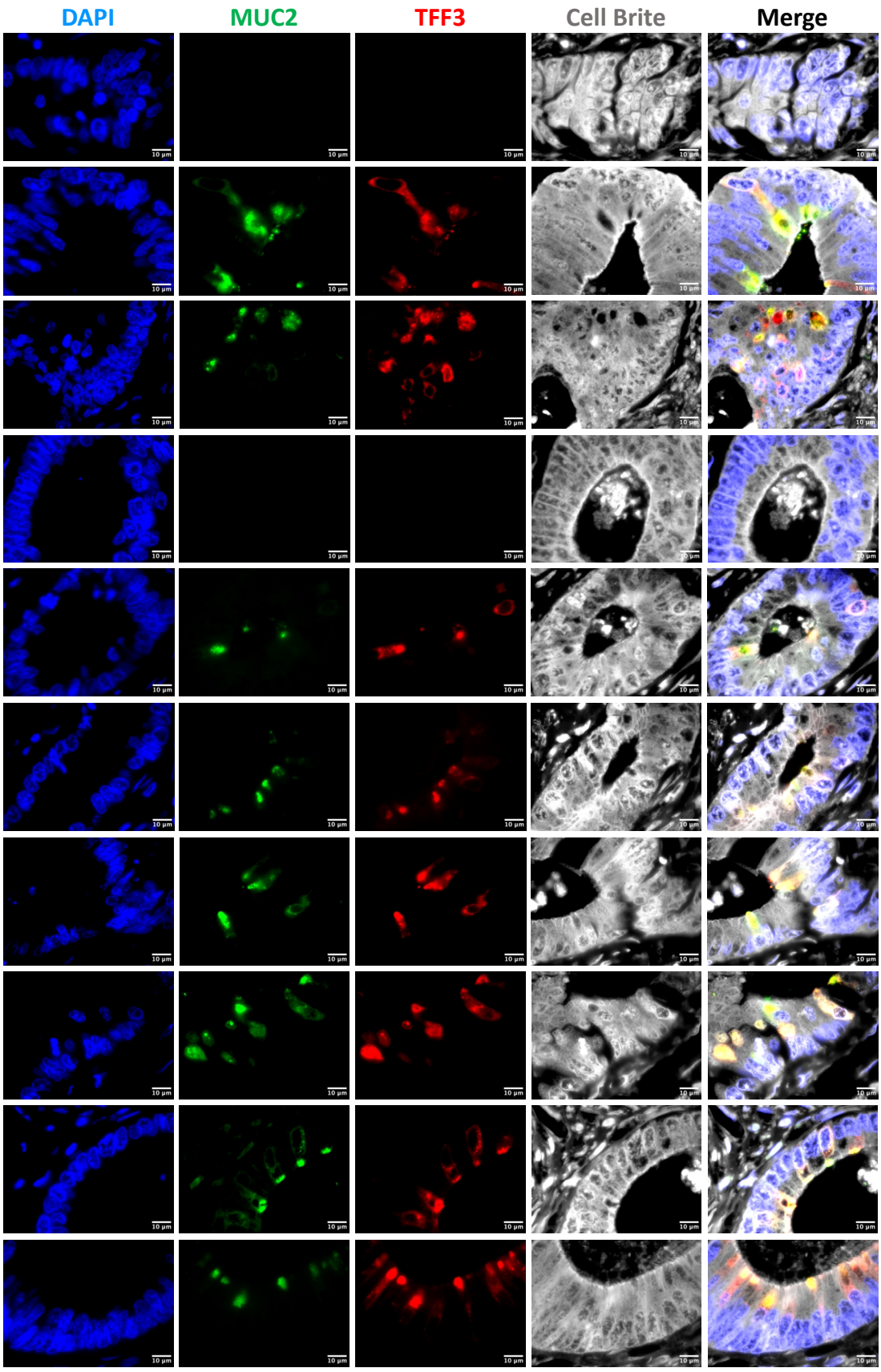
VT11389 (colorectal cancer)

B.



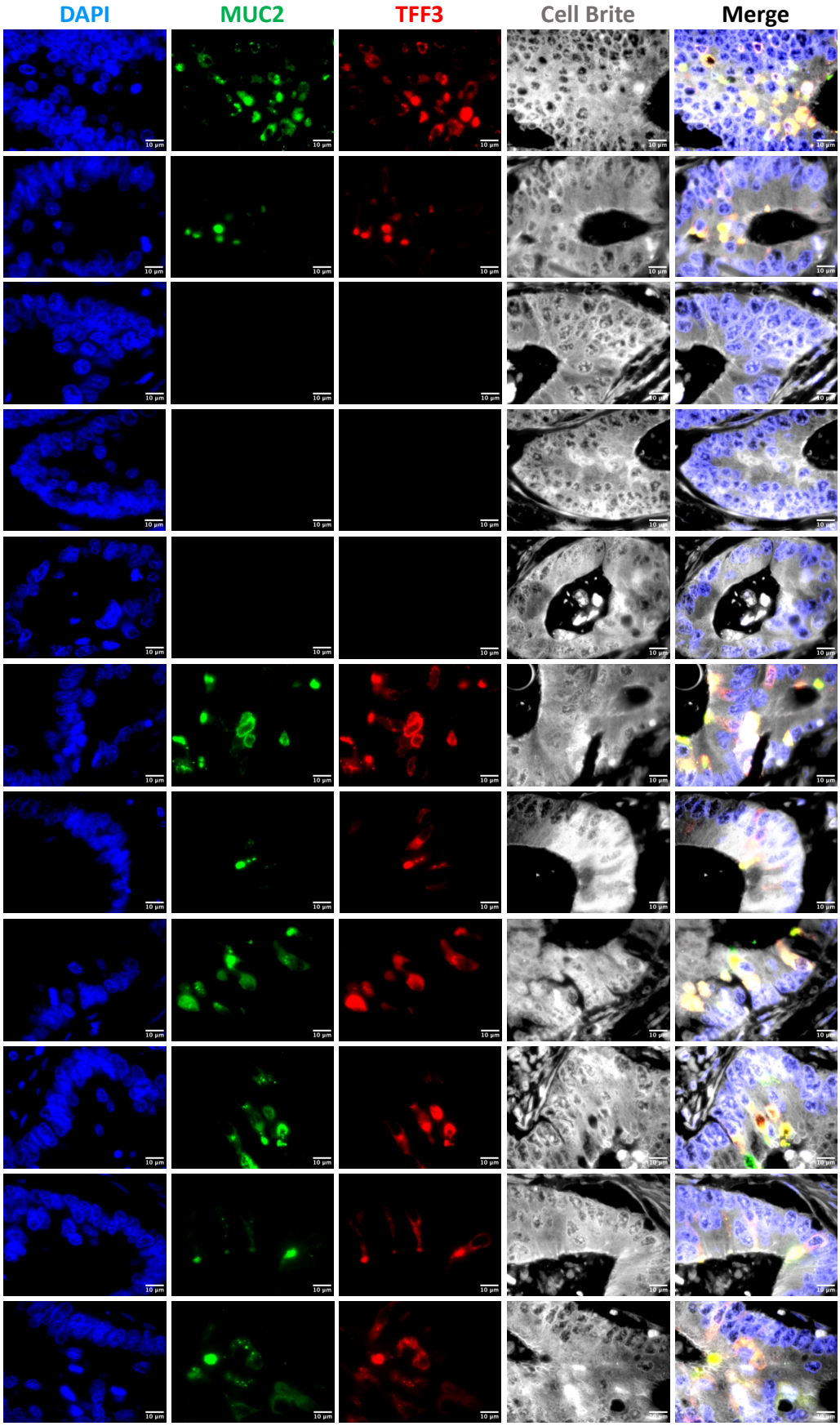
VT11389 (colorectal cancer)

B.



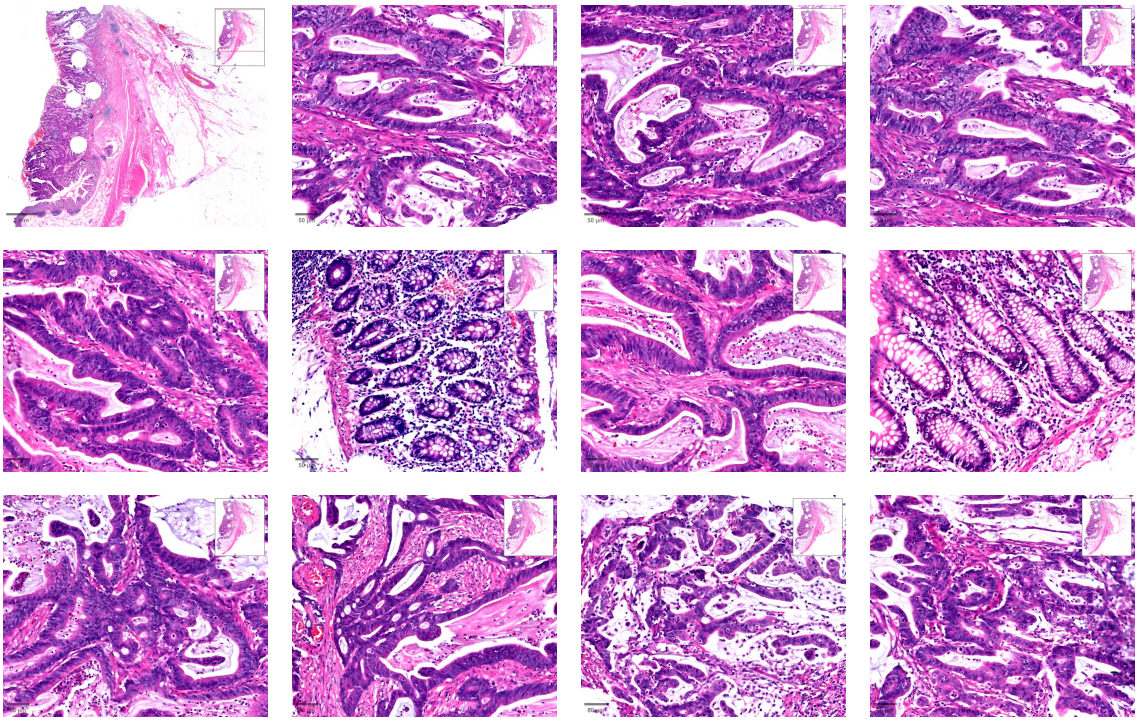
VT11389 (colorectal cancer)

B.



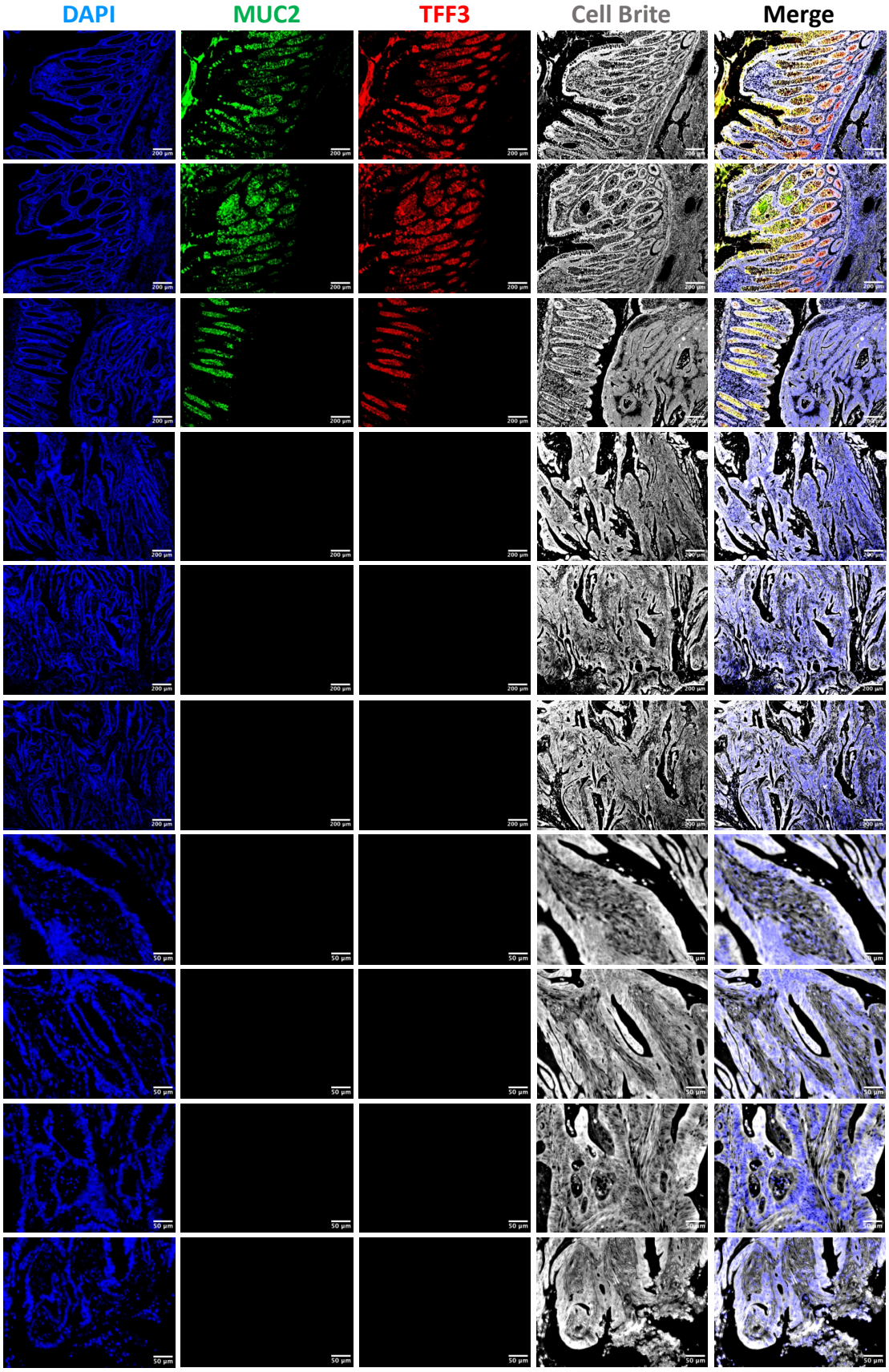
VT11393 (colorectal cancer)

A.



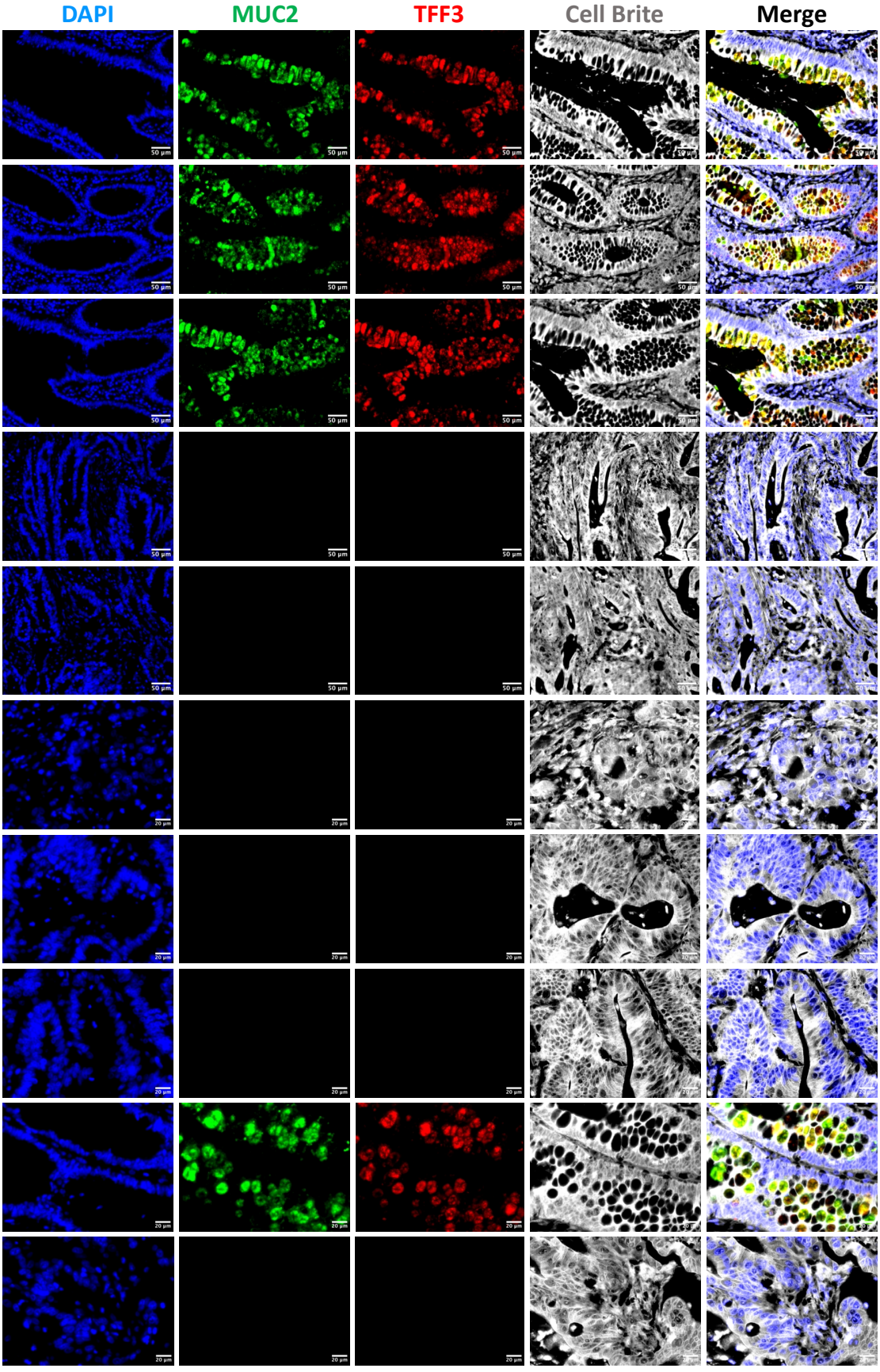
VT11393 (colorectal cancer)

B.



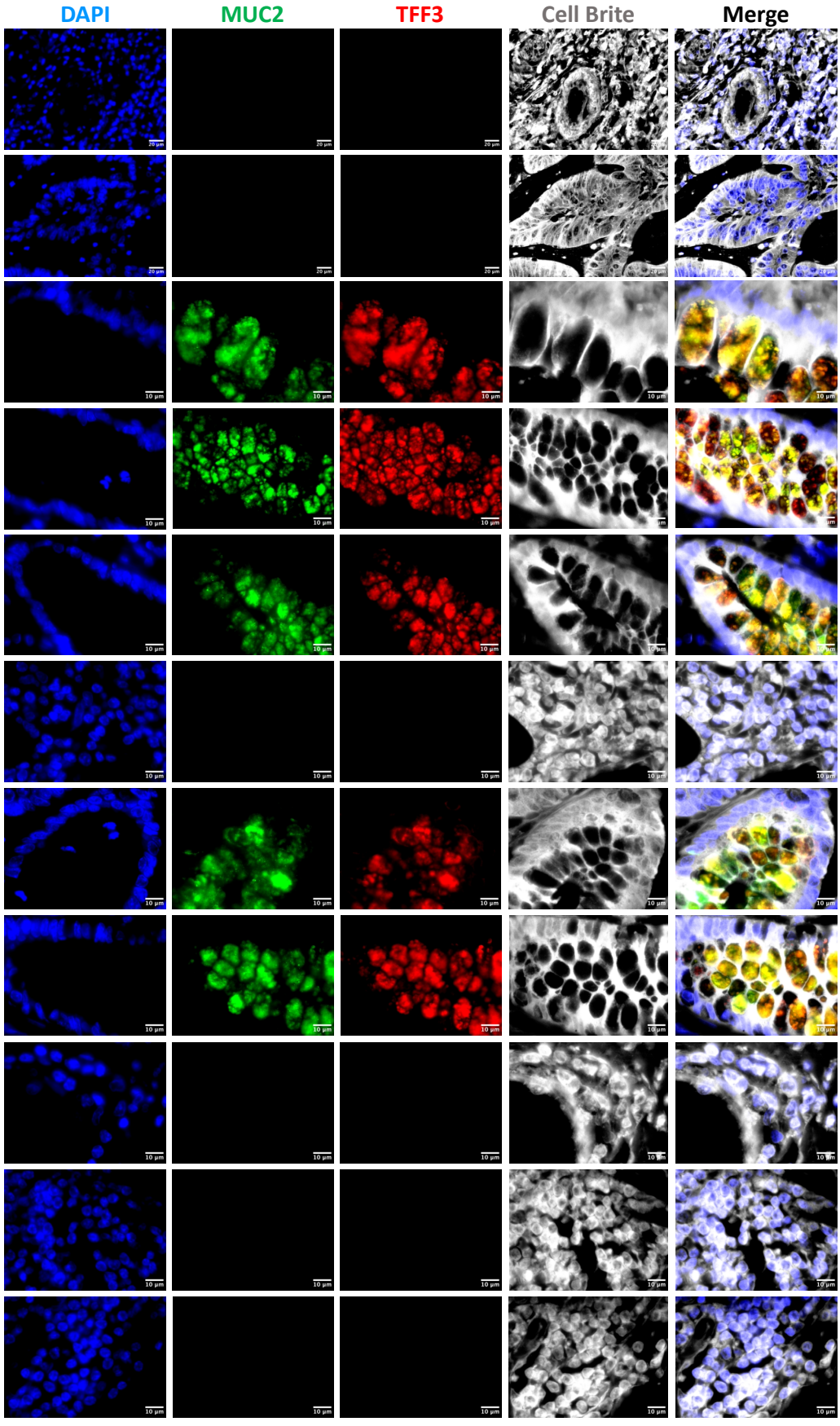
VT11393 (colorectal cancer)

B.



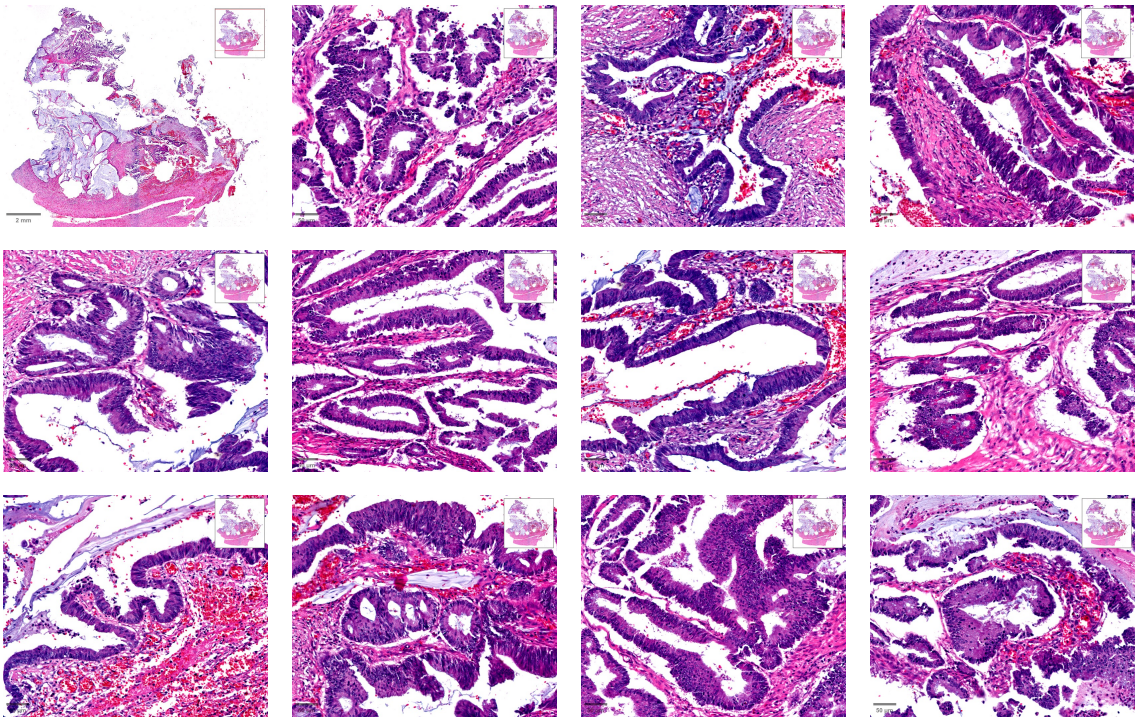
VT11393 (colorectal cancer)

B.



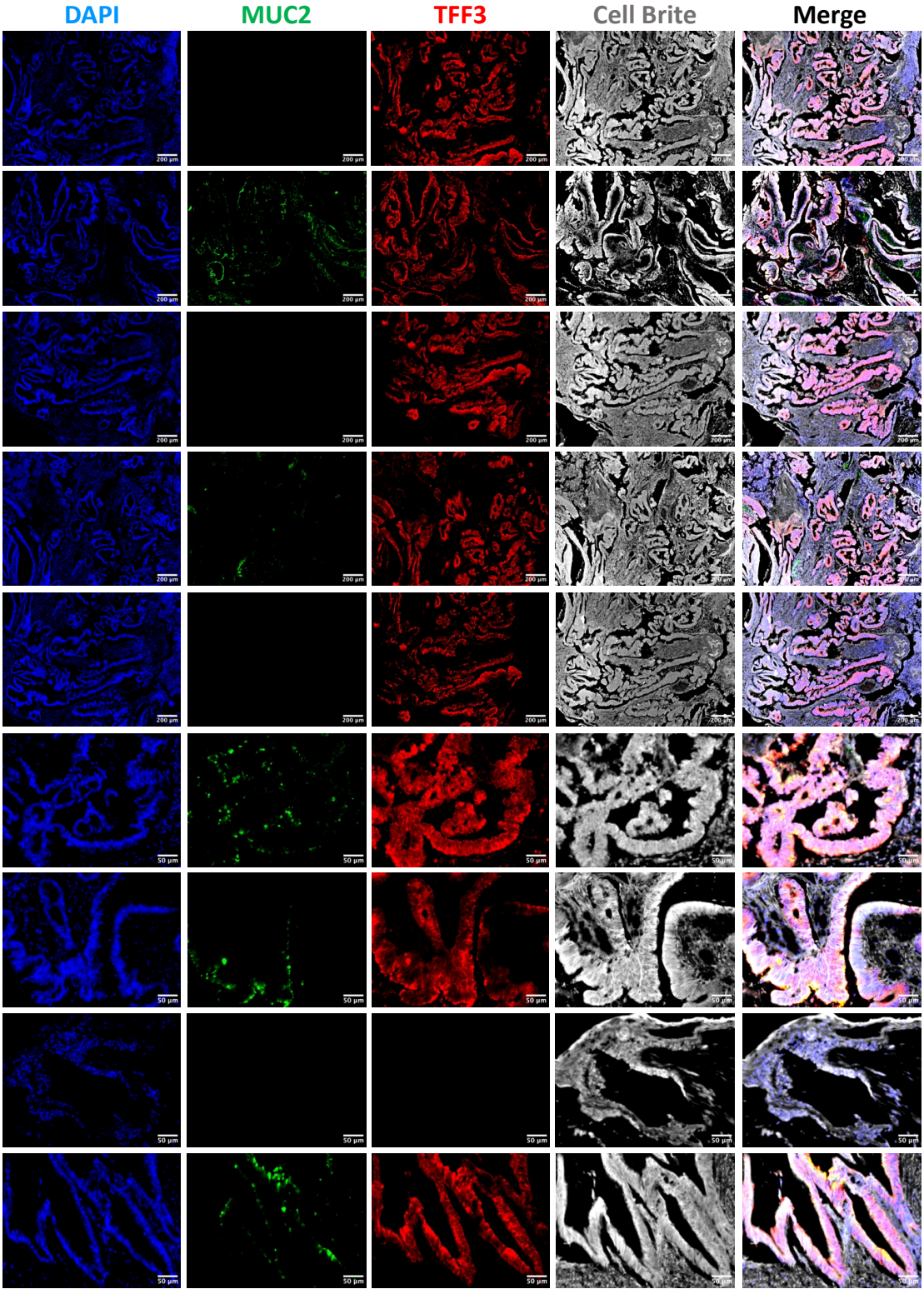
VT11401 (colorectal cancer)

A.



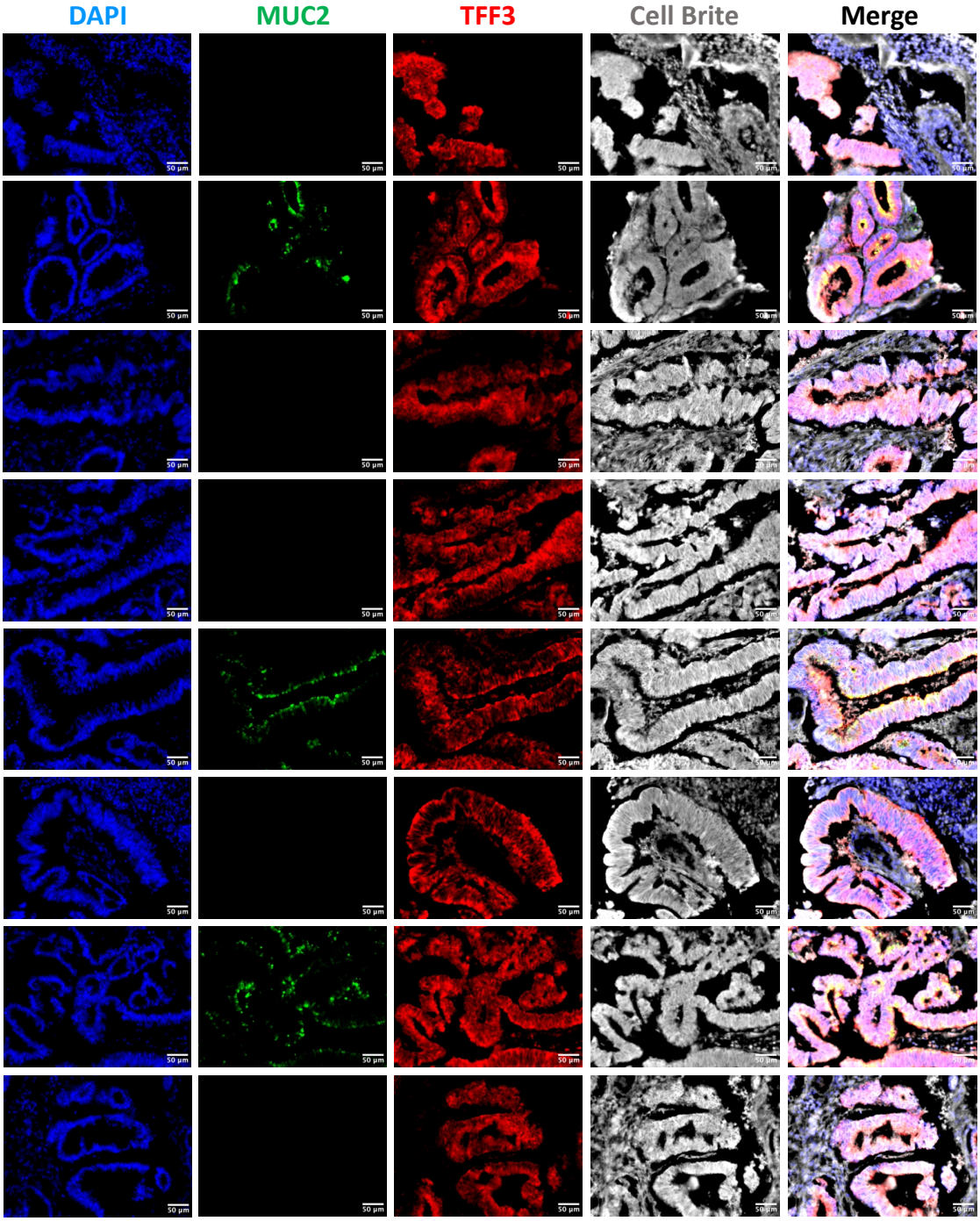
VT11401 (colorectal cancer)

B.



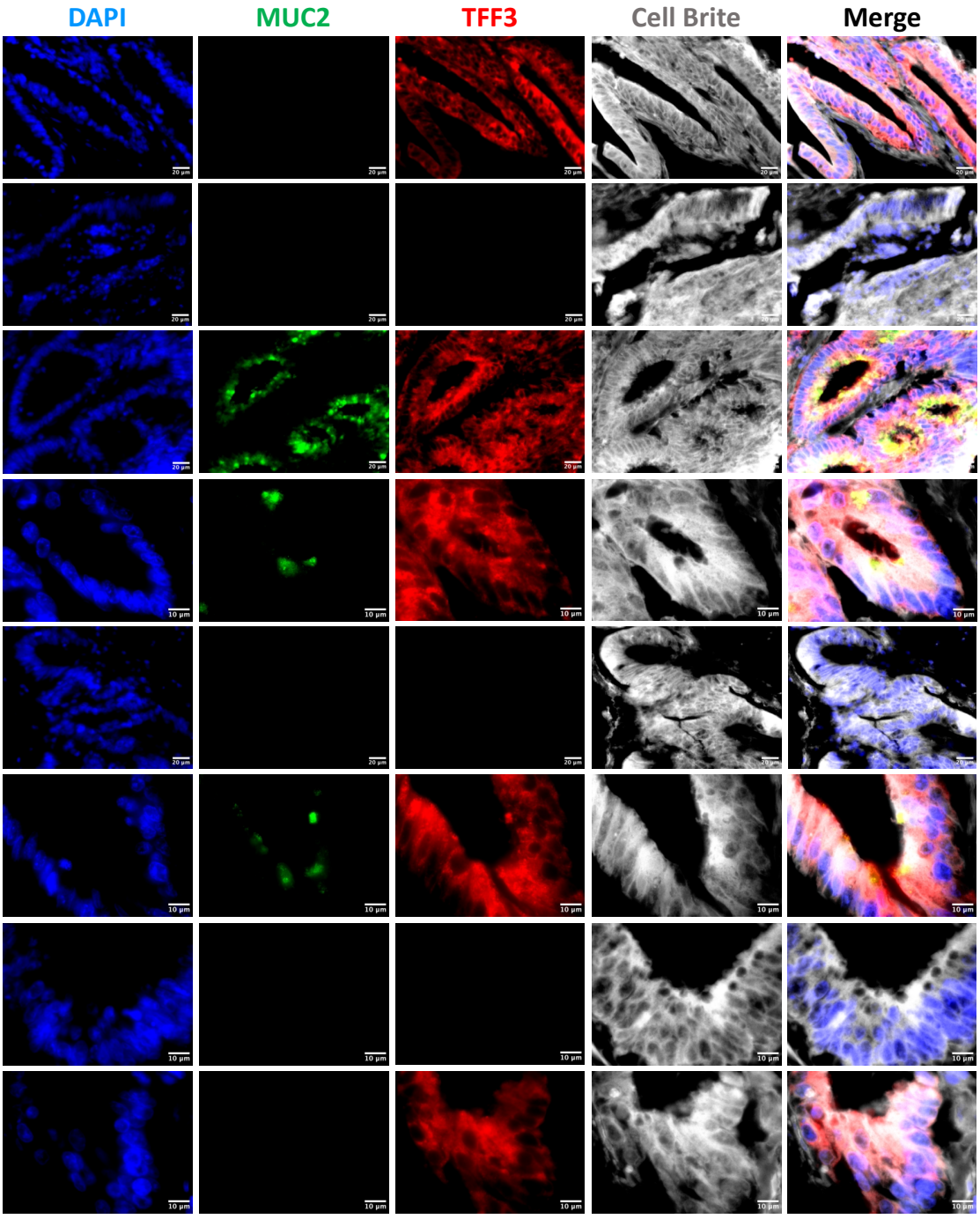
VT11401 (colorectal cancer)

B.



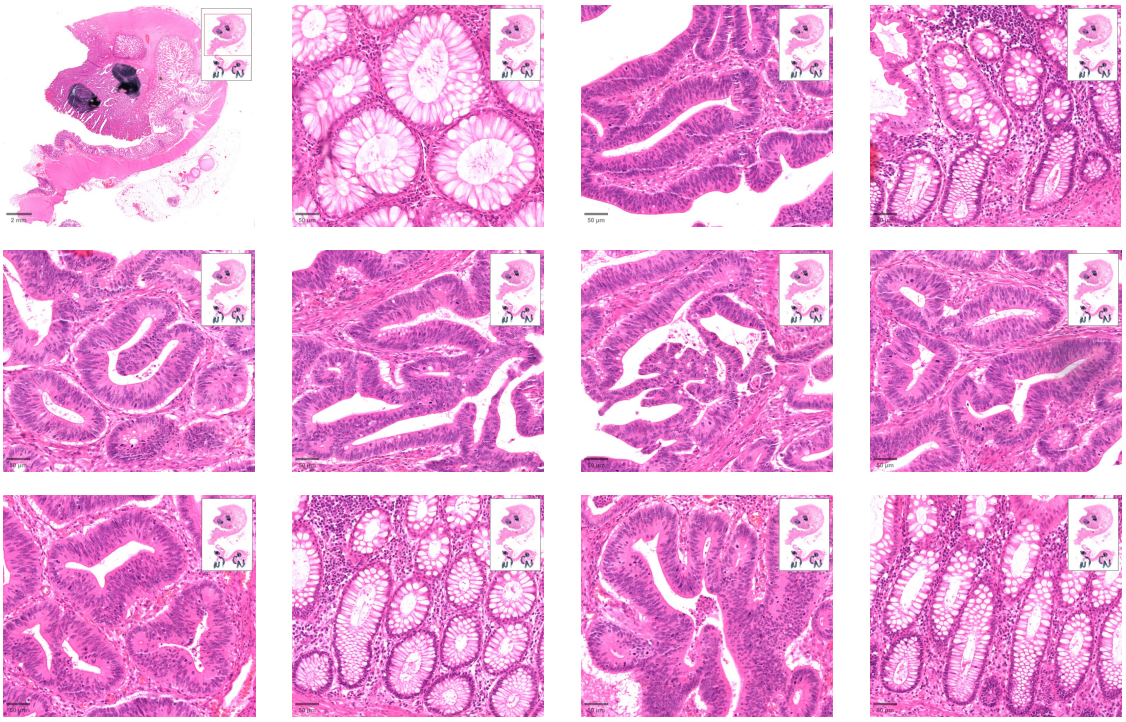
VT11401 (colorectal cancer)

B.



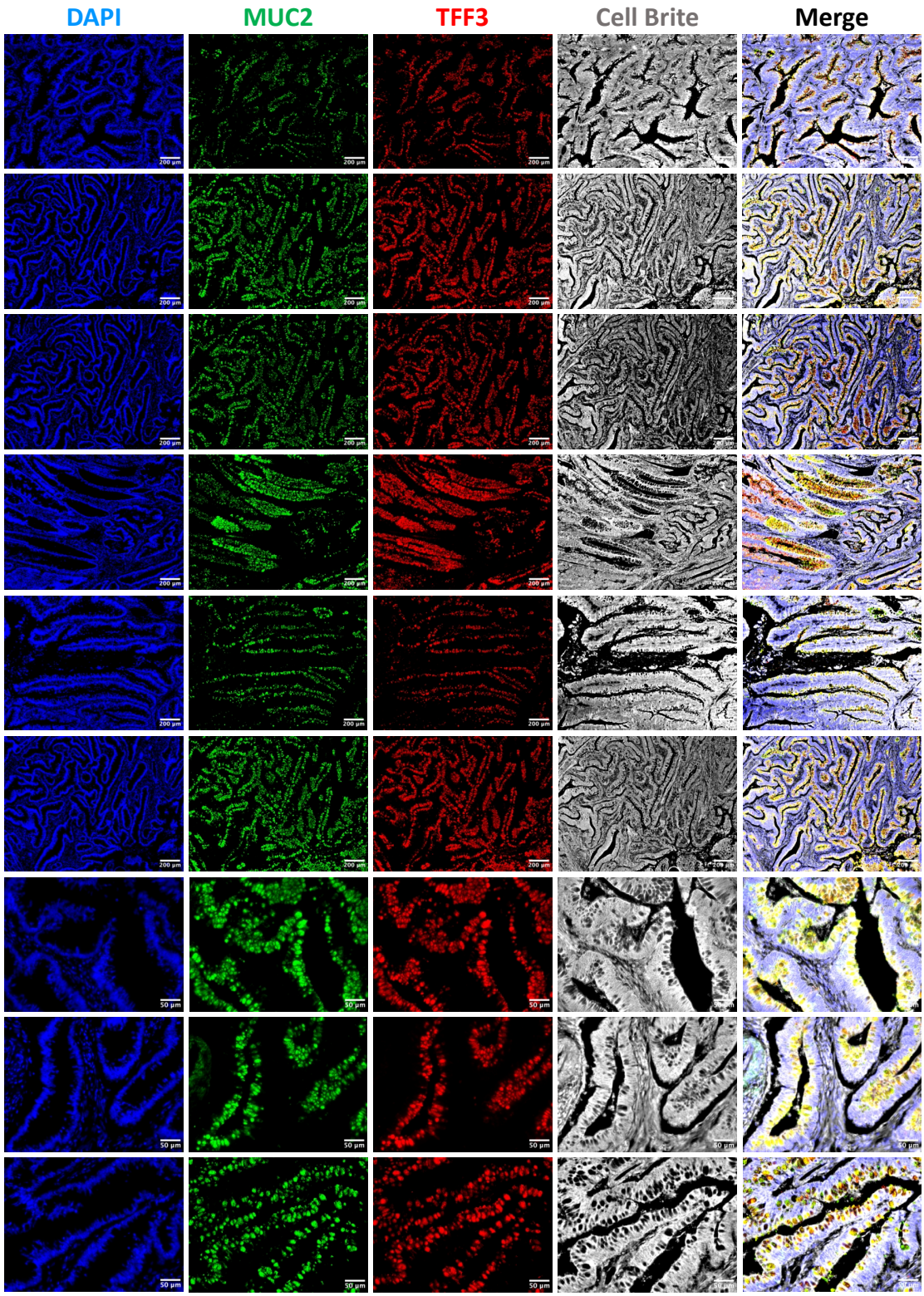
VT11429 (colorectal cancer)

A.



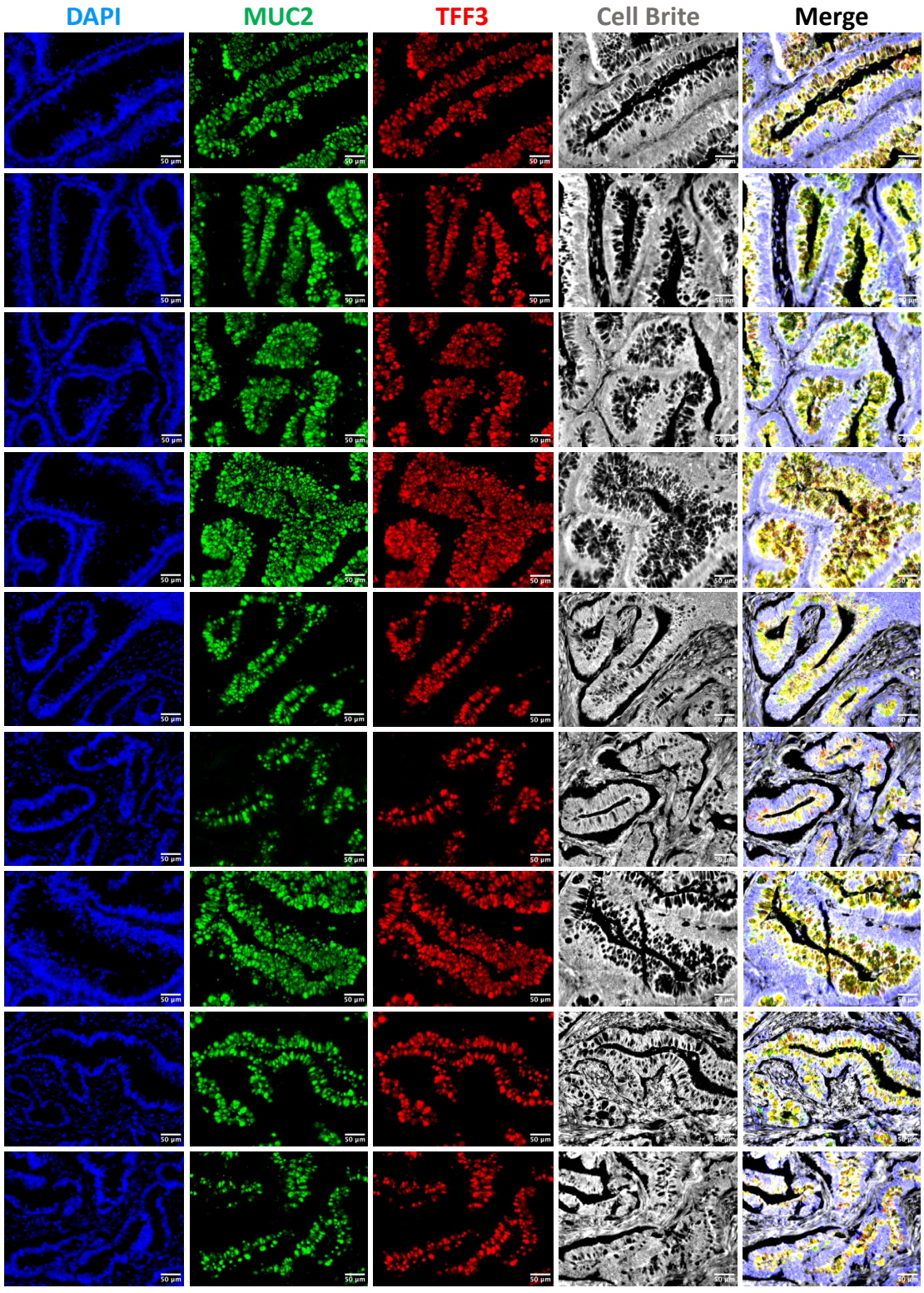
VT11429 (colorectal cancer)

B.



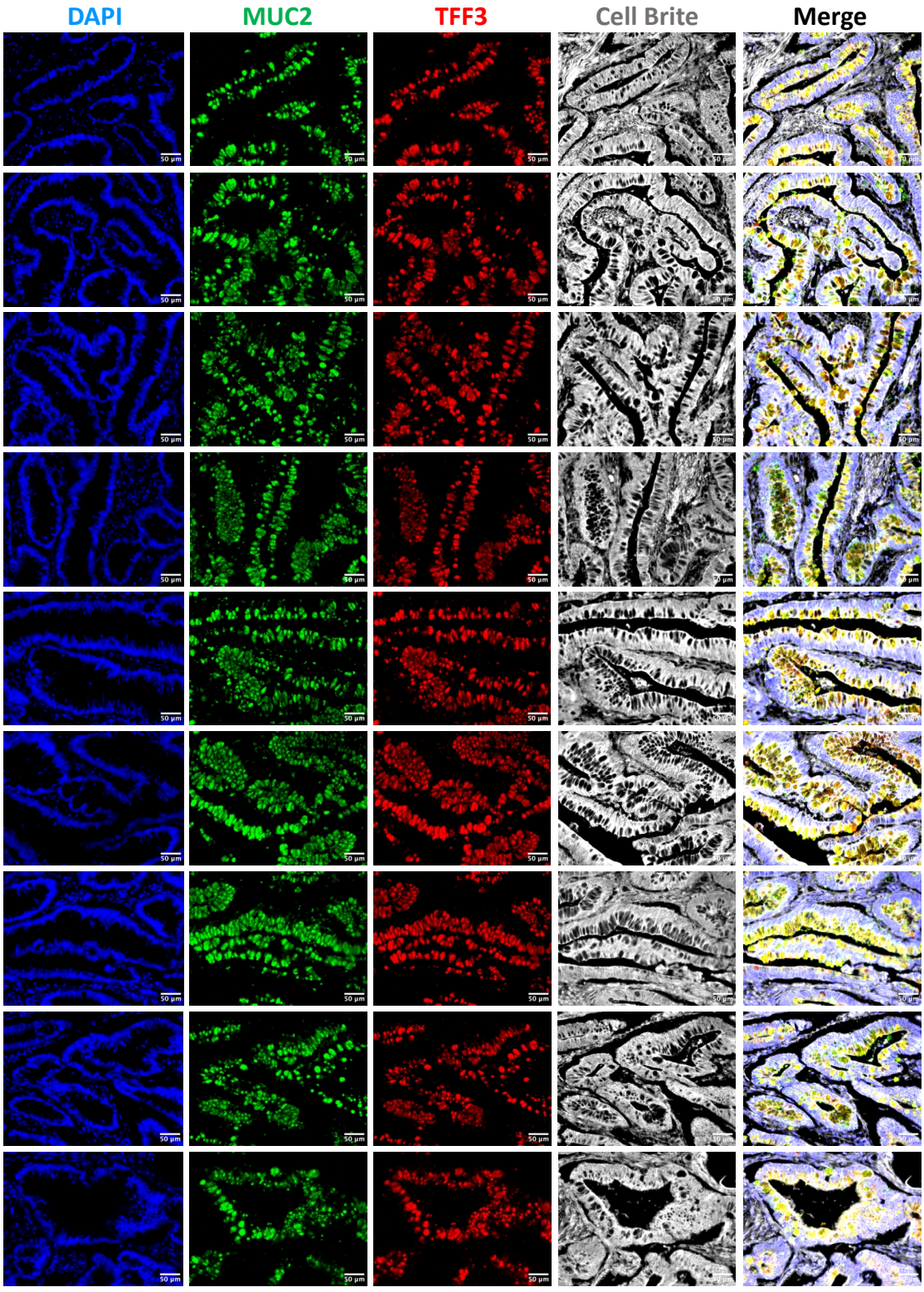
VT11429 (colorectal cancer)

B.



VT11429 (colorectal cancer)

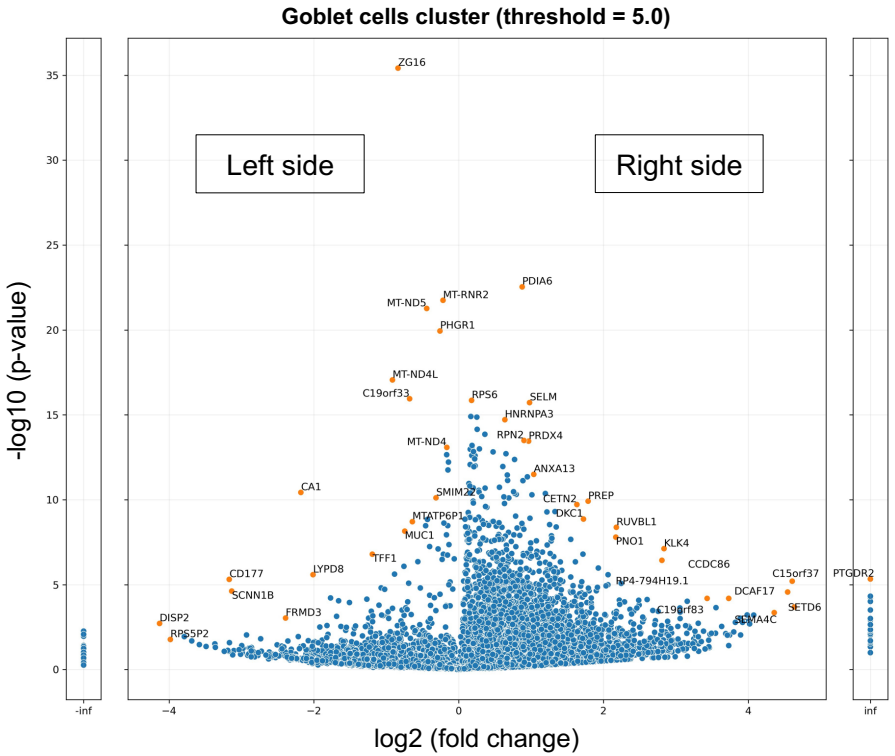
B.

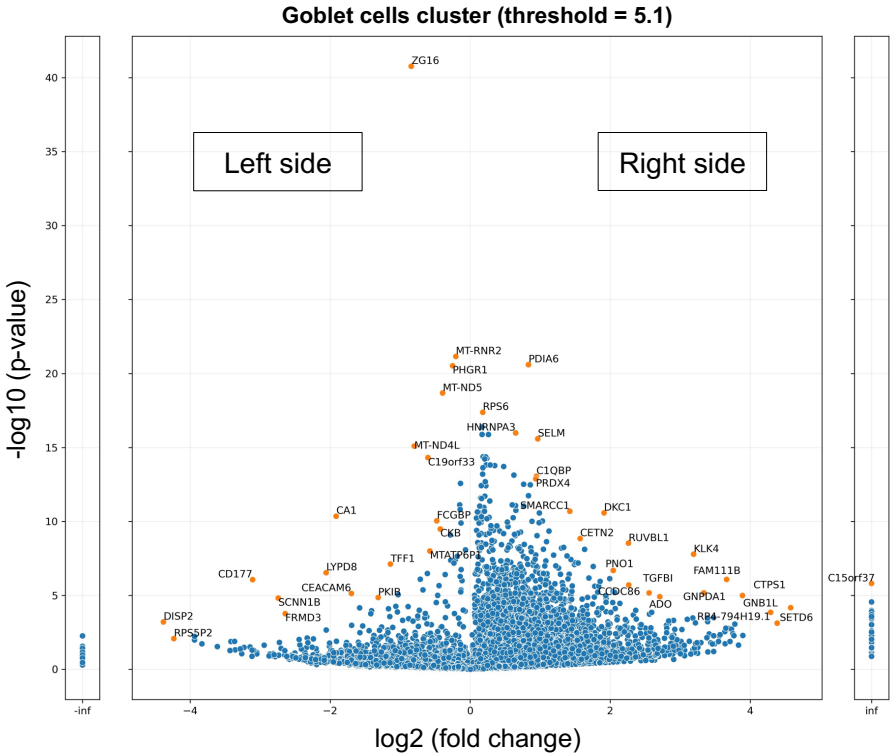


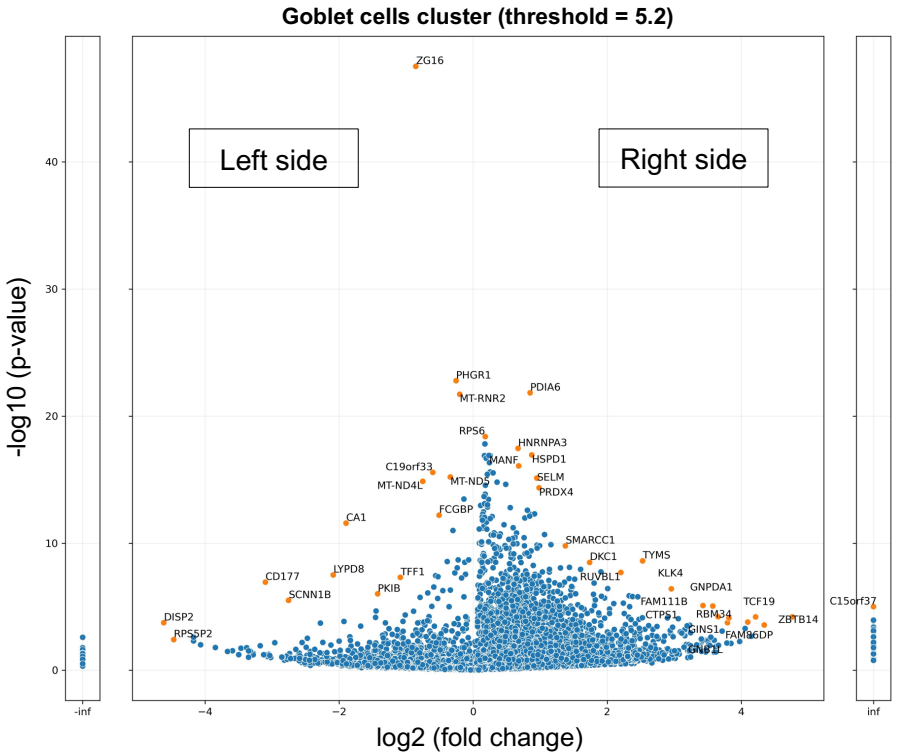
Appendix Table 2. List of differentially expressed genes in the goblet cell cluster compared to other clusters excluding the immune cells in scRNA-seq data of normal colon. The genes were chosen based on the volcano plot (Figure 5.4) and arranged by fold change values in descending order.

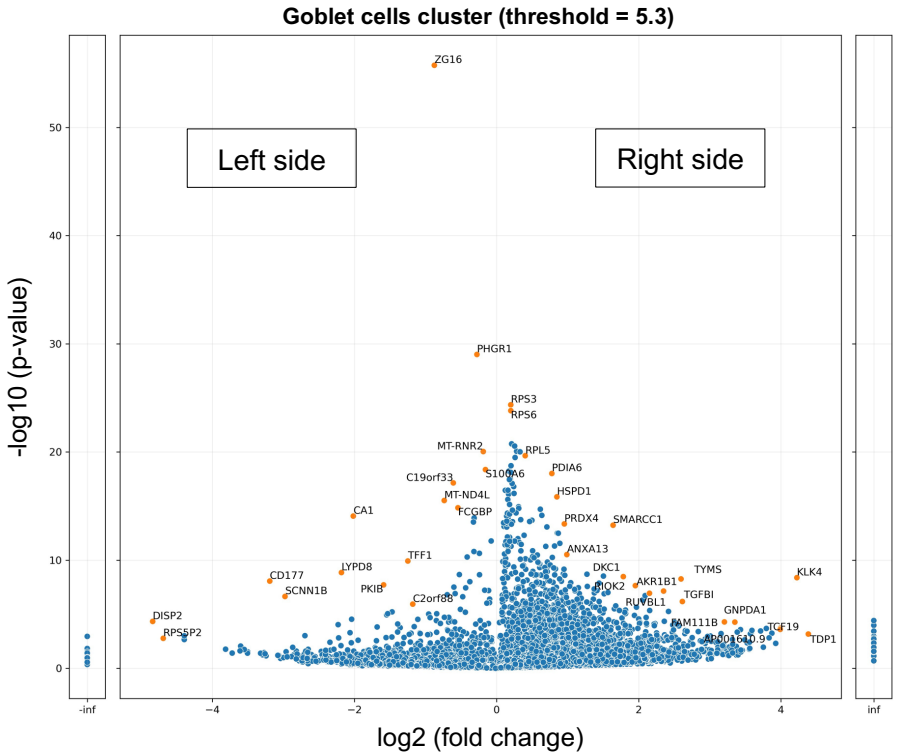
№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	CLCA2	inf	1.22E-37	29.	FCGBP	10.82	3.82E-225
2.	CHRNA3	207.05	6.72E-21	30.	NUCB2	10.01	2.67E-258
3.	ROBO2	137.41	5.57E-13	31.	DNAJC10	9.73	2.74E-174
4.	SPTSSB	108.06	7.95E-26	32.	SPDEF	9.42	2.05E-249
5.	FRZB	74.80	8.45E-229	33.	CREB3L1	7.33	8.84E-188
6.	AC110619.2	74.34	2.53E-59	34.	NPDC1	7.06	4.82E-318
7.	MB	54.56	1.10E-229	35.	FOXA3	6.73	2.49E-196
8.	SLC16A7	53.69	3.57E-99	36.	TSPAN13	5.76	1.02E-256
9.	FAM177B	49.34	3.48E-104	37.	MANF	4.75	1.44E-205
10.	ATOH1	42.93	3.12E-251	38.	HERPUD1	4.11	4.13E-291
11.	REG4	39.47	2.32E-290	39.	XBP1	3.95	7.95E-242
12.	SELM	36.21	1.00E-323	40.	PDIA3	3.76	5.67E-244
13.	RETNLB	36.05	1.18E-196	41.	TPM1	2.84	5.66E-211
14.	FMOD	32.87	1.18E-194	42.	BSG	0.26	2.65E-80
15.	RAP1GAP	29.87	3.49E-261	43.	CES2	0.16	9.90E-71
16.	ATP6V0E2	29.11	5.99E-149	44.	SELENBP1	0.12	1.75E-63
17.	KCNMA1	28.32	1.08E-167	45.	CA2	0.12	9.36E-74
18.	DNAJC12	26.03	3.09E-173	46.	ADIRF	0.09	1.56E-57
19.	B3GNT6	25.20	5.79E-118	47.	SLC26A2	0.07	1.64E-49
20.	AQP3	23.39	1.06E-136	48.	FABP1	0.07	1.07E-101
21.	CAPN9	23.17	6.13E-184	49.	PRSS3	0.07	2.41E-100
22.	TMEM61	20.49	3.49E-183	50.	SLC51B	0.05	6.74E-51
23.	KDELR3	17.39	9.64E-156	51.	MALL	0.04	5.11E-53
24.	HEPACAM2	16.84	7.93E-254	52.	SLC26A3	0.03	1.19E-48
25.	CREB3L4	15.05	2.98E-158	53.	CEACAM7	0.02	2.76E-43
26.	FKBP11	13.74	9.41E-283	54.	TMIGD1	0.01	8.97E-27
27.	RP11-234B24.2	11.61	6.07E-172	55.	BMP3	0.01	1.94E-12
28.	MLPH	11.28	3.16E-232	56.	CHGA	0.01	2.37E-01

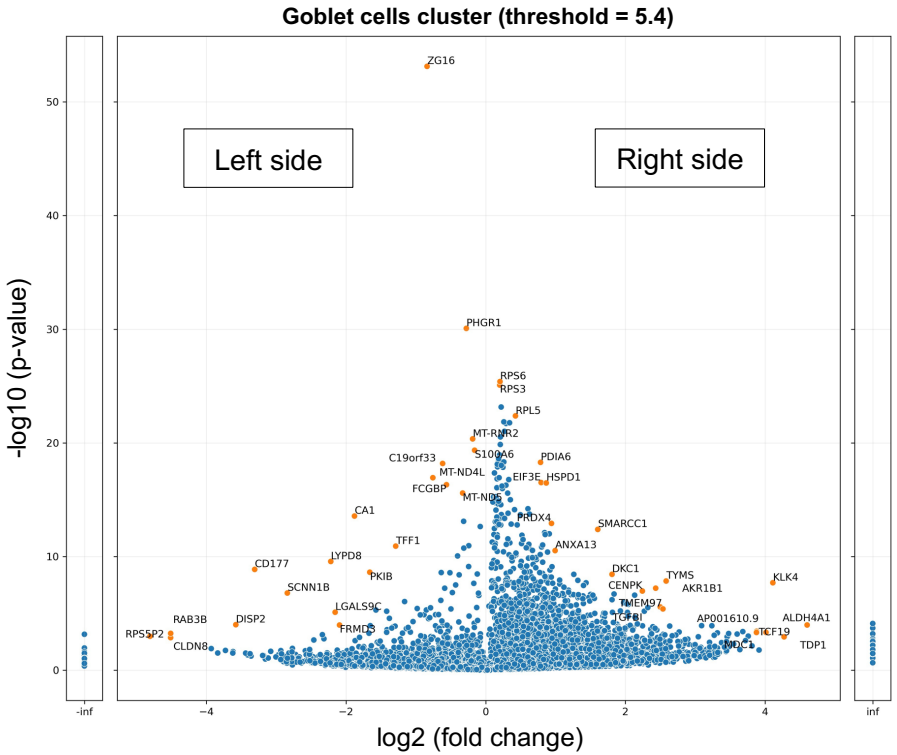
Appendix Figure 8 (following page). Variations between the right and left side of the goblet cell cluster. Volcano plots compare the goblet cell cluster using a range threshold values from 5.0 to 6.0.

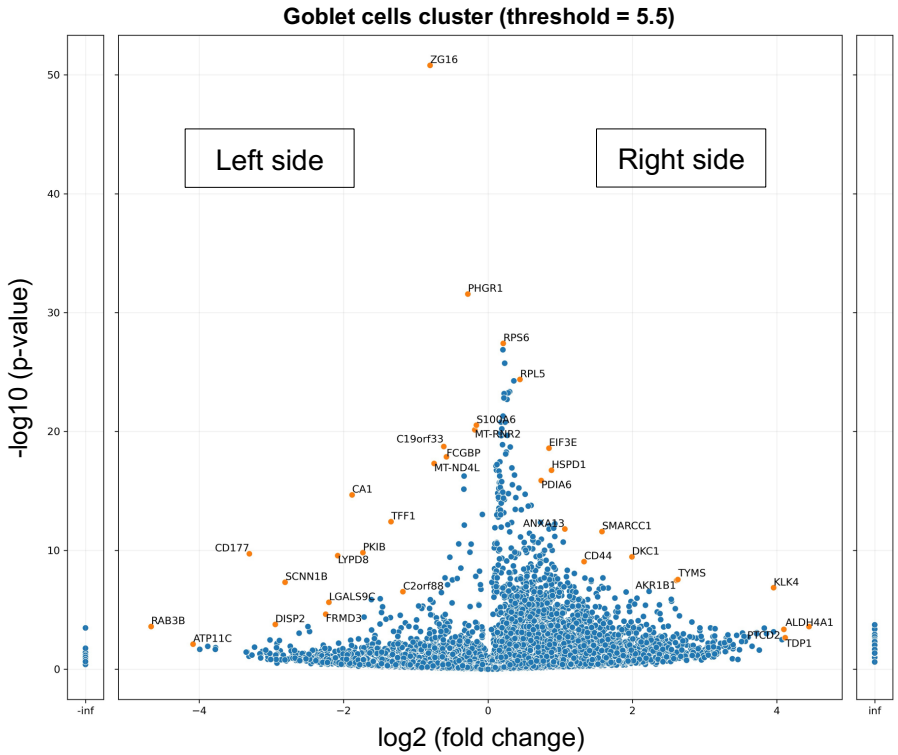

Goblet cells cluster (threshold = 5.0)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PDIA6
MT-ND5
MT-RNR2
PHGR1
MT-ND4L
C19orf33
RPS6
SELM
HNRNPA3
MT-ND4
RPN2
PRDX4
ANXA13
CA1
SMIM22
CETN2
PREP
DKC1
MTATP6P1
RUVBL1
MUC1
PNO1
KLK4
TFF1
CCDC86
C15orf37
PTGDR2
CD177
LYPD8
RP4-794H19.1
SCNN1B
DCAF17
C19orf83
SETD6
FRMD3
ATMA4C
DISP2
RPS5P2

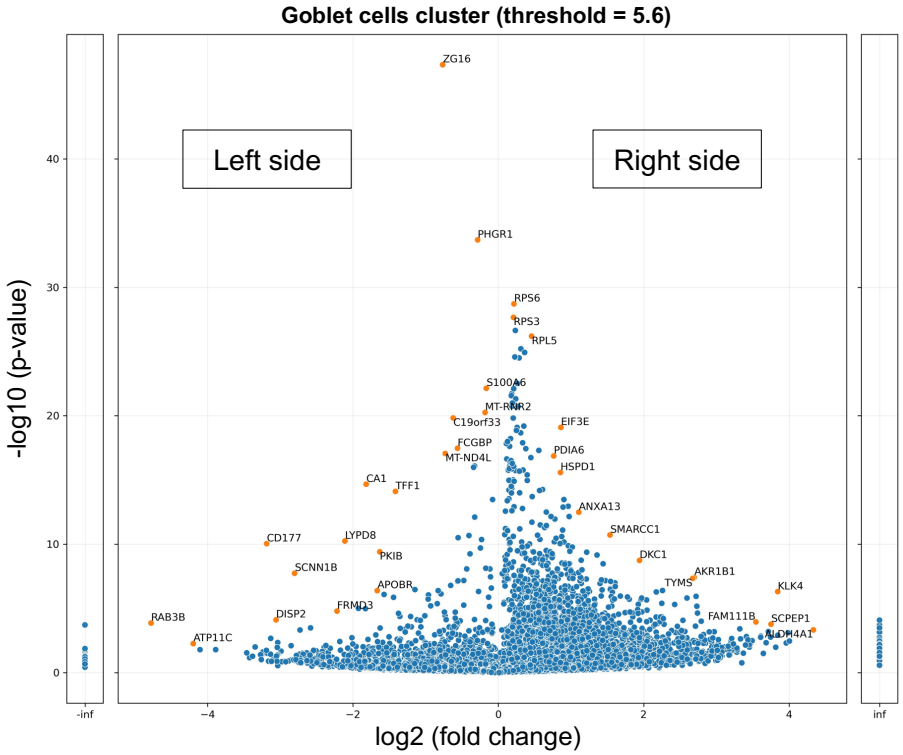

Goblet cells cluster (threshold = 5.1)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
MT-RNR2
PDIA6
PHGR1
MT-ND5
RPS6
HNRNPA3
SELM
MT-ND4L
C19orf33
C1QBP
PRDX4
SMARCC1
DKC1
CA1
FCGBP
CKB
CETN2
RUVBL1
KLK4
PNO1
MTATP6P1
TFF1
FAM111B
CD177
LYPD8
TGFBI
CTPS1
CEACAM6
PKIB
CCDC86
C15orf37
SCNN1B
ADO
GNPDA1
GNB1L
FRMD3
RP4-794H19.1
SETD6
DISP2
RPS5P2

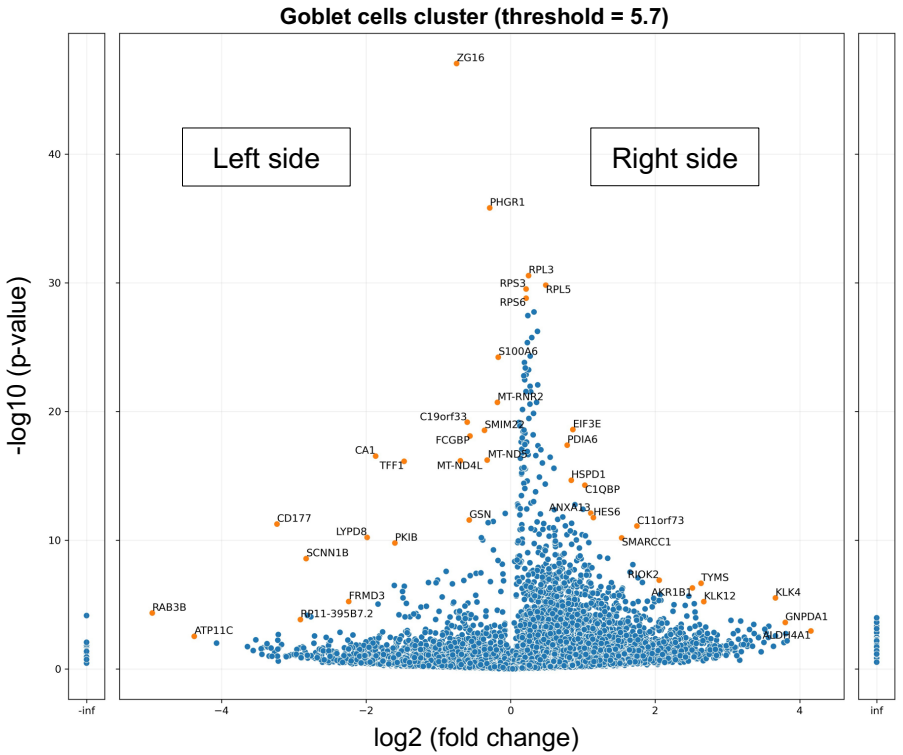

Goblet cells cluster (threshold = 5.2)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
MT-RNR2
PDIA6
RPS6
HNRNPA3
MANF
HSPD1
C19orf33
MT-ND4L
MT-ND5
SELM
PRDX4
FCGBP
CA1
SMARCC1
DKC1
TYMS
CD177
LYPD8
TFF1
RUVBL1
KLK4
PKIB
GNPDA1
SCNN1B
FAM111B
TCF19
C15orf37
DISP2
CTPS1
RBM34
ZBTB14
RPS5P2
GINS1
FAM86DP
GNB1L
-inf
inf

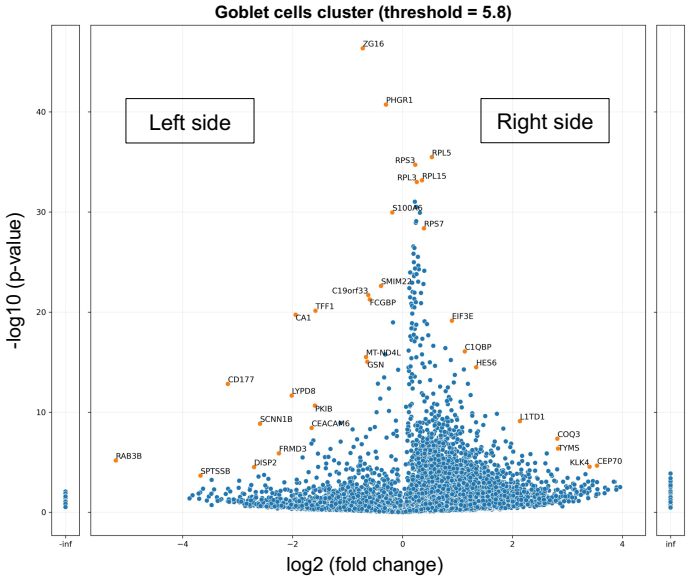

Goblet cells cluster (threshold = 5.3)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
RPS3
RPS6
MT-RNR2
RPL5
PDIA6
C19orf33
S100A6
MT-ND4L
HSPD1
CA1
FCGBP
PRDX4
SMARCC1
ANXA13
TFF1
DKC1
TYMS
KLK4
CD177
LYPD8
AKR1B1
SCNN1B
PKIB
RIOK2
TGFBI
C2orf88
RUVBL1
GNPDA1
DISP2
FAM111B
TCF19
RPS5P2
TDP1
-inf
inf

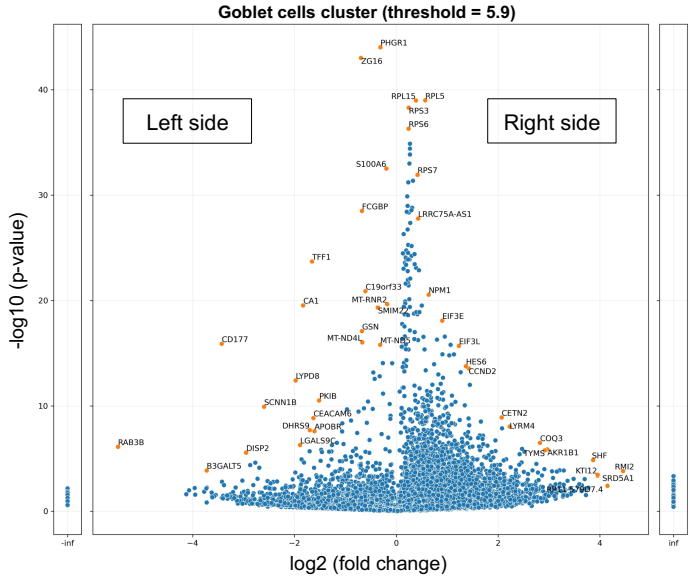

Goblet cells cluster (threshold = 5.4)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
RPS6
RPS3
RPL5
MT-RNR2
S100A6
C19orf33
PDIA6
MT-ND4L
EIF3E
HSPD1
FCGBP
MT-ND5
CA1
PRDX4
SMARCC1
TFF1
ANXA13
CD177
LYPD8
PKIB
DKC1
TYMS
CENPK
AKR1B1
KLK4
SCNN1B
TMEM97
LGALS9C
RAB3B
DISP2
FRMD3
AP001610.9
ALDH4A1
RPS5P2
TCF19
CLDN8
TMDC1
TDP1

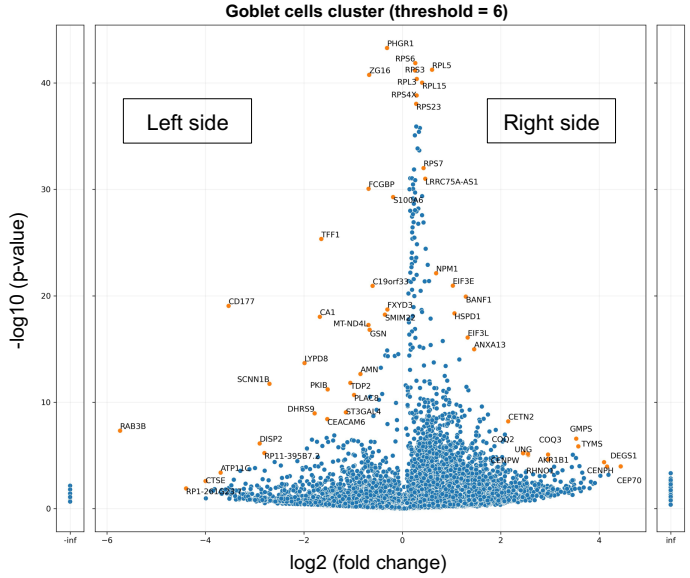

Goblet cells cluster (threshold = 5.5)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
RPS6
RPL5
S100A6
MT-RNR2
C19orf33
EIF3E
FCGBP
HSPD1
MT-ND4L
PDIA6
CA1
TFF1
ANXA13
SMARCC1
CD177
PKIB
LYPD8
CD44
DKC1
SCNN1B
C2orf88
TYMS
AKR1B1
KLK4
LGALS9C
RAB3B
DISP2
FRMD3
ALDH4A1
ATP11C
PTGDS
TDP1


Goblet cells cluster (threshold = 5.6)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
RPS6
RPS3
RPL5
S100A6
MT-RNR2
C19orf33
EIF3E
FCGBP
PDIA6
MT-ND4L
HSPD1
CA1
TFF1
ANXA13
CD177
LYPD8
SMARCC1
PKIB
DKC1
SCNN1B
AKR1B1
APOBR
TYMS
KLK4
RAB3B
DISP2
FRMD3
FAM111B
SCPEP1
ATP11C
ALDH4A1


Goblet cells cluster (threshold = 5.7)
Left side
Right side
-log10 (p-value)
log2 (fold change)
ZG16
PHGR1
RPL3
RPS3
RPL5
RPS6
S100A6
MT-RNR2
C19orf33
SMIM22
EIF3E
FCGBP
PDIA6
CA1
TFF1
MT-ND4L
HSPD1
C1QBP
CD177
GSN
ANXA13
HES6
C11orf73
LYPD8
PKIB
SMARCC1
SCNN1B
PIOK2
TYMS
AKR1B1
KLK12
KLK4
FRMD3
RAB3B
RP11-395B7.2
GNPDA1
ATP11C
ALDH4A1


Goblet cells cluster (threshold = 5.8)
Left side
Right side
-log10 (p-value)
log2 (fold change)


Goblet cells cluster (threshold = 5.9)
Left side
Right side
-log10 (p-value)
log2 (fold change)


Goblet cells cluster (threshold = 6)
Left side
Right side
-log10 (p-value)
log2 (fold change)

Appendix Table 3. List of significantly expressed genes within the goblet cell cluster based on a range of threshold values from 5.0 to 6.0. The genes were highlighted in the volcano plots (Figure 5.5B) and sorted by fold change values from the highest to the lowest.

Goblet cell cluster (threshold = 5.0)							
№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	PTGDR2	inf	4.65E-06	21.	RPS6	1.13	1.41E-16
2.	SETD6	24.88	2.08E-04	22.	MT-ND4	0.89	8.27E-14
3.	C15orf37	24.37	6.25E-06	23.	MT-RNR2	0.86	1.79E-22
4.	RP4-794H19.1	23.32	2.72E-05	24.	PHGR1	0.84	1.15E-20
5.	SEMA4C	20.51	4.43E-04	25.	SMIM22	0.80	7.71E-11
6.	DCAF17	13.27	6.44E-05	26.	MT-ND5	0.74	5.43E-22
7.	C19orf83	10.79	6.44E-05	27.	MTATP6P1	0.64	1.96E-09
8.	KLK4	7.14	7.58E-08	28.	C19orf33	0.63	1.13E-16
9.	CCDC86	7.00	3.69E-07	29.	MUC1	0.60	7.07E-09
10.	RUVBL1	4.52	4.19E-09	30.	ZG16	0.56	3.83E-36
11.	PNO1	4.50	1.59E-08	31.	MT-ND4L	0.53	8.79E-18
12.	PREP	3.45	1.21E-10	32.	TFF1	0.44	1.61E-07
13.	DKC1	3.30	1.37E-09	33.	LYPD8	0.25	2.59E-06
14.	CETN2	3.10	1.92E-10	34.	CA1	0.22	3.71E-11
15.	ANXA13	2.05	3.20E-12	35.	FRMD3	0.19	9.25E-04
16.	SELM	1.97	1.91E-16	36.	SCNN1B	0.11	2.40E-05
17.	PRDX4	1.95	3.53E-14	37.	CD177	0.11	4.89E-06
18.	RPN2	1.87	3.24E-14	38.	RPS5P2	0.06	1.69E-02
19.	PDIA6	1.84	2.92E-23	39.	DISP2	0.06	1.92E-03
20.	HNRNPA3	1.56	1.93E-15				

Goblet cell cluster (threshold = 5.1)							
№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	C15orf37	inf	1.56E-06	22.	RPS6	1.13	4.24E-18
2.	GNB1L	23.90	6.80E-05	23.	MT-RNR2	0.87	7.09E-22
3.	SETD6	20.92	7.56E-04	24.	PHGR1	0.84	3.04E-21
4.	RP4-794H19.1	19.60	1.41E-04	25.	MT-ND5	0.76	2.09E-19

5.	CTPS1	14.83	1.02E-05	26.	CKB	0.74	3.27E-10
6.	FAM111B	12.68	8.35E-07	27.	FCGBP	0.72	9.16E-11
7.	GNPDA1	10.11	6.81E-06	28.	MTATP6P1	0.67	1.01E-08
8.	KLK4	9.14	1.65E-08	29.	C19orf33	0.66	4.87E-15
9.	ADO	6.54	1.22E-05	30.	MT-ND4L	0.58	8.16E-16
10.	CCDC86	5.88	6.83E-06	31.	ZG16	0.56	1.73E-41
11.	TGFBI	4.81	1.99E-06	32.	TFF1	0.45	7.65E-08
12.	RUVBL1	4.80	2.95E-09	33.	PKIB	0.40	1.39E-05
13.	PNO1	4.13	2.07E-07	34.	CEACAM6	0.31	7.46E-06
14.	DKC1	3.77	2.65E-11	35.	CA1	0.27	4.49E-11
15.	CETN2	2.97	1.44E-09	36.	LYPD8	0.24	2.93E-07
16.	SMARCC1	2.68	2.07E-11	37.	FRMD3	0.16	1.73E-04
17.	SELM	1.95	2.62E-16	38.	SCNN1B	0.15	1.52E-05
18.	C1QBP	1.93	8.93E-14	39.	CD177	0.12	8.67E-07
19.	PRDX4	1.92	1.33E-13	40.	RPS5P2	0.05	8.38E-03
20.	PDIA6	1.78	2.56E-21	41.	DISP2	0.05	6.25E-04
21.	HNRNPA3	1.57	1.05E-16				

Goblet cell cluster (threshold = 5.2)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	C15orf37	inf	9.90E-06	21.	HNRNPA3	1.59	3.47E-18
2.	ZBTB14	27.17	6.32E-05	22.	RPS6	1.13	4.09E-19
3.	GNB1L	20.31	2.75E-04	23.	MT-RNR2	0.87	1.93E-22
4.	TCF19	18.58	6.35E-05	24.	PHGR1	0.84	1.64E-23
5.	FAM86DP	17.12	1.62E-04	25.	MT-ND5	0.79	6.21E-16
6.	RBM34	14.09	7.64E-05	26.	FCGBP	0.70	6.28E-13
7.	GINS1	13.88	1.82E-04	27.	C19orf33	0.66	2.68E-16
8.	CTPS1	12.61	6.12E-05	28.	MT-ND4L	0.59	1.36E-15
9.	GNPDA1	11.93	8.87E-06	29.	ZG16	0.55	3.00E-48
10.	FAM111B	10.78	8.13E-06	30.	TFF1	0.47	4.89E-08
11.	KLK4	7.77	3.93E-07	31.	PKIB	0.37	9.68E-07
12.	TYMS	5.77	2.47E-09	32.	CA1	0.27	2.65E-12
13.	RUVBL1	4.61	2.03E-08	33.	LYPD8	0.24	3.17E-08
14.	DKC1	3.34	3.28E-09	34.	SCNN1B	0.15	3.16E-06
15.	SMARCC1	2.60	1.63E-10	35.	CD177	0.12	1.17E-07
16.	PRDX4	1.98	4.41E-15	36.	RPS5P2	0.05	3.90E-03
17.	SELM	1.93	7.53E-16	37.	DISP2	0.04	1.81E-04
18.	HSPD1	1.83	1.15E-17				

19.	PDIA6	1.80	1.47E-22				
20.	MANF	1.60	8.31E-17				

Goblet cell cluster (threshold = 5.3)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	TDP1	20.93	6.75E-04	19.	RPS6	1.15	1.52E-24
2.	KLK4	18.72	4.13E-09	20.	RPS3	1.15	4.34E-25
3.	AP001610.9	15.93	1.94E-04	21.	S100A6	0.90	4.23E-19
4.	TCF19	15.91	2.52E-04	22.	MT-RNR2	0.88	9.09E-21
5.	GNPDA1	10.22	5.32E-05	23.	PHGR1	0.83	9.62E-30
6.	FAM111B	9.23	5.12E-05	24.	FCGBP	0.69	1.44E-15
7.	TGFBI	6.13	6.64E-07	25.	C19orf33	0.66	7.19E-18
8.	TYMS	6.05	5.45E-09	26.	MT-ND4L	0.60	3.04E-16
9.	AKR1B1	5.10	7.18E-08	27.	ZG16	0.55	1.79E-56
10.	RUVBL1	4.44	1.15E-07	28.	C2orf88	0.44	1.16E-06
11.	RIOK2	3.87	2.25E-08	29.	TFF1	0.42	1.19E-10
12.	DKC1	3.44	3.31E-09	30.	PKIB	0.33	1.92E-08
13.	SMARCC1	3.12	5.88E-14	31.	CA1	0.25	8.40E-15
14.	ANXA13	1.98	3.05E-11	32.	LYPD8	0.22	1.39E-09
15.	PRDX4	1.94	4.44E-14	33.	SCNN1B	0.13	2.27E-07
16.	HSPD1	1.80	1.40E-16	34.	CD177	0.11	8.57E-09
17.	PDIA6	1.72	9.67E-19	35.	RPS5P2	0.04	1.68E-03
18.	RPL5	1.32	2.20E-20	36.	DISP2	0.03	4.59E-05

Goblet cell cluster (threshold = 5.4)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	ALDH4A1	24.28	1.04E-04	22.	S100A6	0.90	4.45E-20
2.	TDP1	19.32	1.15E-03	23.	MT-RNR2	0.88	4.47E-21
3.	KLK4	17.28	2.01E-08	24.	PHGR1	0.83	8.45E-31
4.	MDC1	16.15	4.70E-04	25.	MT-ND5	0.80	2.60E-16
5.	AP001610.9	14.71	3.74E-04	26.	FCGBP	0.68	4.78E-17
6.	TCF19	14.69	4.74E-04	27.	C19orf33	0.65	6.47E-19
7.	TYMS	5.99	1.42E-08	28.	MT-ND4L	0.59	1.16E-17
8.	TMEM97	5.81	4.02E-06	29.	ZG16	0.56	7.60E-54
9.	TGFBI	5.65	2.71E-06	30.	TFF1	0.41	1.21E-11

10.	AKR1B1	5.40	6.05E-08	31.	PKIB	0.32	2.39E-09
11.	CENPK	4.73	1.08E-07	32.	CA1	0.27	2.73E-14
12.	DKC1	3.50	3.64E-09	33.	FRMD3	0.23	1.06E-04
13.	SMARCC1	3.04	4.09E-13	34.	LGALS9C	0.22	7.87E-06
14.	ANXA13	1.99	2.98E-11	35.	LYPD8	0.21	2.70E-10
15.	PRDX4	1.92	1.20E-13	36.	SCNN1B	0.14	1.61E-07
16.	HSPD1	1.82	3.36E-17	37.	CD177	0.10	1.35E-09
17.	EIF3E	1.73	3.08E-17	38.	DISP2	0.08	9.83E-05
18.	PDIA6	1.72	5.18E-19	39.	RAB3B	0.04	5.71E-04
19.	RPL5	1.34	4.23E-23	40.	CLDN8	0.04	1.35E-03
20.	RPS6	1.15	4.04E-26	41.	RPS5P2	0.04	1.03E-03
21.	RPS3	1.15	8.18E-26				

Goblet cell cluster (threshold = 5.5)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	ALDH4A1	21.80	2.58E-04	18.	PHGR1	0.82	2.77E-32
2.	TDP1	17.34	2.23E-03	19.	FCGBP	0.67	1.36E-18
3.	PTCD2	17.12	4.45E-04	20.	C19orf33	0.65	1.86E-19
4.	KLK4	15.51	1.39E-07	21.	MT-ND4L	0.60	4.97E-18
5.	TYMS	6.19	2.87E-08	22.	ZG16	0.57	1.61E-51
6.	AKR1B1	6.11	3.37E-08	23.	C2orf88	0.44	2.99E-07
7.	DKC1	3.98	3.53E-10	24.	TFF1	0.39	3.87E-13
8.	SMARCC1	2.99	2.64E-12	25.	PKIB	0.30	1.55E-10
9.	CD44	2.51	8.90E-10	26.	CA1	0.27	2.17E-15
10.	ANXA13	2.09	1.60E-12	27.	LYPD8	0.24	2.80E-10
11.	HSPD1	1.84	1.84E-17	28.	LGALS9C	0.22	2.34E-06
12.	EIF3E	1.80	2.57E-19	29.	FRMD3	0.21	2.36E-05
13.	PDIA6	1.67	1.34E-16	30.	SCNN1B	0.14	4.85E-08
14.	RPL5	1.36	4.22E-25	31.	DISP2	0.13	1.70E-04
15.	RPS6	1.16	3.93E-28	32.	CD177	0.10	1.95E-10
16.	S100A6	0.89	2.98E-21	33.	ATP11C	0.06	7.81E-03
17.	MT-RNR2	0.88	7.32E-21	34.	RAB3B	0.04	2.56E-04

Goblet cell cluster (threshold = 5.6)							
№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	ALDH4A1	20.17	4.66E-04	18.	PHGR1	0.82	2.06E-34
2.	KLK4	14.35	4.86E-07	19.	FCGBP	0.68	3.43E-18
3.	SCPEP1	13.47	1.68E-04	20.	C19orf33	0.65	1.48E-20
4.	FAM111B	11.66	1.12E-04	21.	MT-ND4L	0.60	8.71E-18
5.	AKR1B1	6.47	3.88E-08	22.	ZG16	0.59	4.80E-48
6.	TYMS	6.38	4.55E-08	23.	TFF1	0.37	7.72E-15
7.	DKC1	3.84	1.82E-09	24.	PKIB	0.32	3.91E-10
8.	SMARCC1	2.90	1.91E-11	25.	APOBR	0.32	4.08E-07
9.	ANXA13	2.15	3.21E-13	26.	CA1	0.28	2.13E-15
10.	EIF3E	1.82	8.11E-20	27.	LYPD8	0.23	5.60E-11
11.	HSPD1	1.81	2.60E-16	28.	FRMD3	0.21	1.60E-05
12.	PDIA6	1.70	1.34E-17	29.	SCNN1B	0.14	1.83E-08
13.	RPL5	1.37	6.39E-27	30.	DISP2	0.12	7.65E-05
14.	RPS6	1.16	1.97E-29	31.	CD177	0.11	9.05E-11
15.	RPS3	1.16	2.21E-28	32.	ATP11C	0.05	5.45E-03
16.	S100A6	0.89	7.26E-23	33.	RAB3B	0.04	1.36E-04
17.	MT-RNR2	0.88	5.59E-21				

Goblet cell cluster (threshold = 5.7)							
№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	ALDH4A1	17.81	1.10E-03	21.	MT-RNR2	0.88	1.94E-21
2.	GNPDA1	13.93	2.41E-04	22.	PHGR1	0.82	1.52E-36
3.	KLK4	12.67	2.99E-06	23.	MT-ND5	0.80	5.90E-17
4.	KLK12	6.38	5.81E-06	24.	SMIM22	0.78	2.88E-19
5.	TYMS	6.21	2.19E-07	25.	FCGBP	0.68	7.96E-19
6.	AKR1B1	5.72	5.04E-07	26.	GSN	0.67	2.67E-12
7.	RIOK2	4.16	1.26E-07	27.	C19orf33	0.66	6.68E-20
8.	C11orf73	3.36	7.74E-12	28.	MT-ND4L	0.62	6.72E-17
9.	SMARCC1	2.90	6.55E-11	29.	ZG16	0.60	9.25E-48
10.	HES6	2.21	1.73E-12	30.	TFF1	0.36	7.33E-17
11.	ANXA13	2.16	7.86E-13	31.	PKIB	0.33	1.62E-10
12.	C1QBP	2.04	5.27E-15	32.	CA1	0.27	2.89E-17
13.	EIF3E	1.82	2.50E-19	33.	LYPD8	0.25	6.01E-11
14.	HSPD1	1.79	2.18E-15	34.	FRMD3	0.21	5.70E-06

15.	PDIA6	1.72	4.09E-18	35.	SCNN1B	0.14	2.65E-09
16.	RPL5	1.40	1.51E-30	36.	RP11-395B7.2	0.13	1.44E-04
17.	RPL3	1.19	2.77E-31	37.	CD177	0.11	5.49E-12
18.	RPS6	1.16	1.58E-29	38.	ATP11C	0.05	2.89E-03
19.	RPS3	1.16	3.03E-30	39.	RAB3B	0.03	4.44E-05
20.	S100A6	0.89	5.99E-25				

Goblet cell cluster (threshold = 5.8)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	CEP70	11.62	2.21E-05	17.	FCGBP	0.66	5.52E-22
2.	KLK4	10.58	2.85E-05	18.	C19orf33	0.65	2.00E-22
3.	TYMS	7.11	4.33E-07	19.	GSN	0.64	9.25E-16
4.	COQ3	7.05	4.44E-08	20.	MT-ND4L	0.63	3.14E-16
5.	L1TD1	4.39	7.86E-10	21.	ZG16	0.61	4.49E-47
6.	HES6	2.53	3.25E-15	22.	TFF1	0.33	7.46E-21
7.	C1QBP	2.19	8.56E-17	23.	PKIB	0.33	2.23E-11
8.	EIF3E	1.86	7.36E-20	24.	CEACAM6	0.32	3.76E-09
9.	RPL5	1.45	3.33E-36	25.	CA1	0.26	1.85E-20
10.	RPS7	1.31	4.31E-29	26.	LYPD8	0.25	2.22E-12
11.	RPL15	1.28	6.87E-34	27.	FRMD3	0.21	1.31E-06
12.	RPL3	1.20	1.01E-33	28.	SCNN1B	0.17	1.43E-09
13.	RPS3	1.17	1.93E-35	29.	DISP2	0.15	3.06E-05
14.	S100A6	0.88	1.13E-30	30.	CD177	0.11	1.52E-13
15.	PHGR1	0.81	1.88E-41	31.	SPTSSB	0.08	2.19E-04
16.	SMIM22	0.76	2.46E-23	32.	RAB3B	0.03	6.69E-06

Goblet cell cluster (threshold = 5.9)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	RMI2	21.94	1.56E-04	23.	S100A6	0.87	3.08E-33
2.	RP11-579D7.4	17.75	3.86E-03	24.	PHGR1	0.81	9.59E-45
3.	SRD5A1	15.53	3.52E-04	25.	MT-ND5	0.80	1.62E-16
4.	KTI12	15.49	4.41E-04	26.	SMIM22	0.78	4.72E-20
5.	SHF	14.59	1.41E-05	27.	C19orf33	0.66	1.33E-21
6.	AKR1B1	7.83	1.35E-06	28.	MT-ND4L	0.63	9.56E-17

7.	TYMS	7.56	1.66E-06	29.	FCGBP	0.63	3.24E-29
8.	COQ3	7.07	3.25E-07	30.	GSN	0.63	8.14E-18
9.	LYRM4	4.69	9.26E-09	31.	ZG16	0.62	9.82E-44
10.	CETN2	4.20	1.28E-09	32.	PKIB	0.35	3.10E-11
11.	CCND2	2.68	2.74E-14	33.	APOBR	0.33	2.54E-08
12.	HES6	2.58	1.75E-14	34.	CEACAM6	0.32	1.45E-09
13.	EIF3L	2.35	2.06E-16	35.	TFF1	0.32	2.01E-24
14.	EIF3E	1.87	8.29E-19	36.	DHRS9	0.31	2.03E-08
15.	NPM1	1.56	2.86E-21	37.	CA1	0.28	2.96E-20
16.	RPL5	1.48	1.03E-39	38.	LGALS9C	0.27	5.25E-07
17.	LRRC75A-AS1	1.35	1.71E-28	39.	LYPD8	0.25	3.91E-13
18.	RPS7	1.34	1.22E-32	40.	SCNN1B	0.17	1.23E-10
19.	RPL15	1.31	1.06E-39	41.	DISP2	0.13	2.77E-06
20.	RPS3	1.18	5.22E-39	42.	CD177	0.09	1.29E-16
21.	RPS6	1.18	5.10E-37	43.	B3GALT5	0.08	1.37E-04
22.	MT-RNR2	0.88	2.24E-20	44.	RAB3B	0.02	7.74E-07

Goblet cell cluster (threshold = 6)

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	CEP70	21.66	1.07E-04	29.	FXYD3	0.81	1.93E-19
2.	CENPH	17.86	1.11E-04	30.	PHGR1	0.81	5.18E-44
3.	DEGS1	17.13	4.28E-05	31.	SMIM22	0.78	5.88E-19
4.	TYMS	11.92	1.40E-06	32.	C19orf33	0.66	1.16E-21
5.	GMPS	11.57	2.66E-07	33.	GSN	0.63	1.55E-17
6.	RHNO1	7.82	2.37E-05	34.	ZG16	0.63	1.74E-41
7.	AKR1B1	7.79	8.29E-06	35.	FCGBP	0.62	8.79E-31
8.	CENPW	5.89	9.10E-06	36.	MT-ND4L	0.62	5.50E-18
9.	COQ3	5.87	7.37E-06	37.	AMN	0.55	2.21E-13
10.	COQ2	5.62	5.87E-06	38.	PLAC8	0.51	2.08E-11
11.	UNG	5.48	6.23E-06	39.	TDP2	0.48	1.52E-12
12.	CETN2	4.44	6.23E-09	40.	ST3GAL4	0.45	8.83E-10
13.	ANXA13	2.75	1.05E-15	41.	PKIB	0.35	6.23E-12
14.	EIF3L	2.51	8.30E-17	42.	CEACAM6	0.35	3.84E-09
15.	BANF1	2.44	1.19E-20	43.	TFF1	0.32	4.55E-26
16.	HSPD1	2.08	4.45E-19	44.	CA1	0.31	9.38E-19
17.	EIF3E	2.04	1.11E-21	45.	DHRS9	0.29	1.11E-09
18.	NPM1	1.61	7.55E-23	46.	LYPD8	0.25	2.08E-14
19.	RPL5	1.52	5.72E-42	47.	SCNN1B	0.15	1.86E-12

20.	LRRC75A-AS1	1.38	1.00E-31	48.	RP11-395B7.2	0.14	5.95E-06
21.	RPS7	1.35	9.96E-33	49.	DISP2	0.13	7.47E-07
22.	RPL15	1.32	9.44E-41	50.	CD177	0.09	8.84E-20
23.	RPL3	1.23	4.19E-41	51.	ATP11C	0.08	4.12E-04
24.	RPS4X	1.22	1.49E-39	52.	CTSE	0.06	2.54E-03
25.	RPS23	1.22	9.19E-39	53.	RP1-261G23.7	0.05	1.25E-02
26.	RPS6	1.20	1.36E-42	54.	RAB3B	0.02	4.72E-08
27.	RPS3	1.20	6.78E-42				
28.	S100A6	0.88	5.26E-30				

Appendix Table 4. Summary of up- and down-regulated genes between MUC2 or TFF3-positive and MUC2 or TFF3-negative cells in scRNA-seq data of LS180 cell line. (A) The genes were highlighted in the volcano plot (Figure 5.10A) for MUC2 and (B) the volcano plot (Figure 5.10B) for TFF3 were aligned in descending order based on their fold change values.

A.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	AC004067.1	inf	1.53E-08	13.	KLK12	36.29	5.97E-09
2.	AC027601.4	inf	1.53E-08	14.	DLL4	30.61	9.47E-08
3.	CHST11	inf	7.27E-16	15.	H3F3A	24.09	3.94E-10
4.	GAST	inf	8.11E-16	16.	UGT2B10	21.03	3.88E-09
5.	MUC2	inf	4.03E-70	17.	NCALD	15.81	8.76E-10
6.	FCGBP	326.79	6.07E-31	18.	GUCY2C	10.48	1.03E-09
7.	U2AF1L5	130.22	4.83E-08	19.	HMGCS2	9.60	1.17E-09
8.	ACE	57.95	2.38E-09	20.	REG4	6.41	1.72E-11
9.	KRT20	53.08	6.87E-09	21.	AGR2	2.97	9.14E-11
10.	MUC5AC	52.32	4.95E-12	22.	SMIM19	0.07	6.55E-02
11.	HYPK	50.19	4.72E-16	23.	IMPDH1	0.06	3.44E-02
12.	TFF3	42.73	1.09E-38	24.	PDAP1	0.03	5.46E-03

B.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	AC090772.3	inf	6.72E-07	16.	TFF2	8.20	2.39E-06
2.	TFF3	inf	2.37E-33	17.	RASSF10	7.66	1.03E-04
3.	FCGBP	91.32	7.31E-09	18.	HMBOX1	6.70	1.54E-05
4.	AC009996.1	66.44	1.28E-04	19.	REG4	4.23	7.95E-10
5.	AD000671.3	58.51	6.78E-07	20.	CEACAM5	3.98	1.50E-09
6.	AL139289.1	54.80	3.30E-04	21.	CEACAM6	3.59	6.05E-07
7.	KRT20	46.35	5.21E-05	22.	AC016586.1	2.97	8.27E-06
8.	MUC2	41.48	2.55E-11	23.	ANO9	2.90	5.19E-07
9.	DLL4	38.47	3.21E-06	24.	TSPAN8	1.99	9.56E-06
10.	AC243964.2	29.73	4.48E-05	25.	PSMB5	1.97	8.92E-06
11.	ARAP3	25.11	3.18E-05	26.	S100A16	1.83	2.58E-06
12.	AC024941.2	19.90	5.22E-05	27.	KRT18	1.49	4.35E-06
13.	AC104113.1	15.89	1.15E-04	28.	MALSU1	0.16	5.05E-03
14.	TFF1	13.04	6.35E-09	29.	MMP14	0.06	3.15E-02
15.	AC069503.1	11.76	7.49E-05				

Appendix Table 5. List of up- and downregulated genes in MUC2- and TFF3-positive cell lines. (A) describes a summary of 40 genes up- and down-regulated between MUC2-positive and MUC2-negative cell lines based on our microarray data. The genes were selected from the volcano plot (Figure 6.1A) and arranged by their fold change values in descending order. (B) shows a list of 47 genes up- and down-regulated between TFF3-positive and TFF3-negative cell lines based on our microarray data. The genes were chosen based on the volcano plot (Figure 6.1B) and sorted by fold change.

A.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	APOA2--2	10.77	5.22E-02	21.	ABCC13--4	1.50	1.38E-06
2.	APOA2--1	9.44	5.17E-02	22.	ZNF575	1.45	2.11E-06
3.	LOC100510557 /// MUC6	8.41	3.24E-02	23.	IL4--2	1.22	1.43E-06
4.	LOC100505875	8.18	5.34E-03	24.	RC3H2--2	0.66	1.22E-05
5.	MUC2	7.94	3.20E-06	25.	PFDN4--1	0.62	1.49E-04
6.	FCGBP	7.26	6.58E-09	26.	CAMSAP1L1--2	0.55	5.80E-04
7.	EN1	5.63	1.01E-02	27.	MAP7D1	0.53	6.27E-04
8.	NCRNA00238	5.54	1.71E-02	28.	FLNA--2	0.44	1.96E-03
9.	OR51E1	5.41	8.31E-03	29.	KATNAL1--2	0.32	2.70E-03
10.	RETNLB--1	4.29	5.89E-04	30.	IQCJ-SCHIP1 /// SCHIP1	0.32	4.75E-03
11.	CLDN2	3.78	1.33E-03	31.	GPX8--1	0.26	6.94E-03
12.	ST6GALNAC1	3.19	3.30E-05	32.	NXN	0.21	2.82E-03
13.	FAM3D	3.10	4.47E-05	33.	RDX--3	0.20	6.31E-03
14.	CA12--4	2.95	8.88E-05	34.	RDX--2	0.16	3.85E-03
15.	RETNLB--2	2.91	5.21E-04	35.	KLHL5--1	0.14	2.01E-02
16.	NR1I2--1	2.61	7.99E-05	36.	FN1--4	0.08	2.74E-02
17.	CDC42EP5	2.27	6.09E-06	37.	FN1--3	0.08	3.11E-02
18.	C2orf89	2.24	1.19E-05	38.	SLC2A3--1	0.07	7.91E-02
19.	AMN--2	2.08	9.30E-06	39.	SLC2A3--3	0.05	9.32E-02
20.	ATP1B1--2	1.93	7.78E-06	40.	IGF2 /// INS-IGF2--1	0.04	1.92E-01

B.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	GCG	22.92	3.16E-01	26.	ALDH2	2.08	6.37E-09
2.	TFF3	17.55	2.30E-06	27.	MGST2	1.86	1.68E-09
3.	SMOC2--1	14.27	3.34E-02	28.	SPIN1--1	0.66	6.16E-10
4.	COL14A1--1	13.83	3.89E-01	29.	PRPF31--1	0.63	3.38E-09
5.	NTS	13.44	3.24E-01	30.	FXR1--3	0.50	2.46E-11
6.	REG4--2	12.47	3.44E-02	31.	BCAS4--1	0.45	5.83E-10
7.	TFF1	11.47	5.70E-02	32.	WWC2--3	0.41	3.44E-08
8.	REG4--1	11.28	3.82E-02	33.	SAMD4A--1	0.36	2.13E-08
9.	KIAA1324--2	8.80	1.29E-03	34.	TSPAN4--2	0.28	9.33E-08
10.	SPINK1	8.14	3.14E-04	35.	RDX--3	0.23	2.85E-07
11.	LRP4	7.14	9.10E-04	36.	MYO5A--3	0.23	1.95E-07
12.	PTPRD--3	6.88	1.16E-04	37.	RDX--1	0.22	1.65E-07
13.	DPEP1	5.38	1.07E-05	38.	ARHGAP29--4	0.21	1.36E-06
14.	AKR1C3	4.49	7.37E-06	39.	KLHL5--1	0.16	3.03E-06
15.	TDGF1 /// TDGF3	4.39	8.06E-06	40.	CAV1--2	0.11	8.04E-06
16.	C9orf152	4.19	6.15E-08	41.	FRMD6--2	0.11	7.06E-06
17.	PPP1R1B	3.88	9.14E-08	42.	CYR61--1	0.11	2.57E-06
18.	CLRN3	3.85	4.02E-07	43.	FERMT2--2	0.10	7.92E-06
19.	VIL1--4	3.36	4.64E-09	44.	WNT5A--1	0.07	2.72E-04
20.	VIL1--3	3.33	4.38E-09	45.	AXL--2	0.07	7.23E-04
21.	FOXA2--1	3.23	4.21E-08	46.	TFPI2--2	0.04	2.24E-02
22.	STARD10--1	3.00	2.61E-13	47.	IGFBP7--2	0.03	5.95E-02
23.	STARD10--2	2.86	4.70E-13				
24.	PRR15L	2.68	3.72E-09				
25.	TST	2.42	3.59E-10				

Appendix Table 6. List of selected genes up- and down-regulated in different groups (groups 1-5) derived from protein expression levels of MUC2 and TFF3. The genes were labelled in volcano plots (Figure 6A-D) and sorted by their corresponding R values. (A) compares group 1 cell lines versus the cell lines belonging to groups 2 and 3. (B) shows the differences between groups 1 and 4. (C) displays highly expressed genes in MUC2-positive groups (groups 1-3) versus group 4 cell lines. (D) represents differentially expressed genes in both MUC2- and TFF3-positive cell lines versus group 5.

A.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	GCG	96.47	9.20E-02	28.	MR1	1.94	3.11E-04
2.	COL14A1	60.38	1.50E-01	29.	N4BP2L2--3	1.91	2.91E-04
3.	CXCL5--3	29.03	8.96E-02	30.	EGFR--2	1.83	7.06E-05
4.	HAPLN1	26.56	3.50E-02	31.	MID1IP1	1.80	2.86E-04
5.	CXCL5--2	20.54	6.59E-02	32.	GRAPL	1.66	1.17E-04
6.	APOA2--2	18.08	1.55E-01	33.	PTGER3	1.32	4.07E-06
7.	HAPLN1--3	17.91	4.34E-02	34.	MBD3L1	1.28	6.13E-05
8.	SST	17.03	1.52E-01	35.	GTF3C5	0.78	1.37E-04
9.	TUBA1A	9.12	4.67E-02	36.	CDC25C	0.62	1.49E-04
10.	DNAJC12	8.52	5.64E-03	37.	CDCA3	0.57	1.15E-04
11.	LOC100289026	7.37	1.30E-02	38.	CDC25C--3	0.54	1.76E-05
12.	DNAJC12--2	5.35	5.88E-03	39.	CASP6	0.47	3.83E-04
13.	SMAD9--2	4.85	3.18E-02	40.	ADA	0.43	1.61E-02
14.	HAPLN1--2	4.35	3.29E-02	41.	MTHFD2L--5	0.42	2.38E-02
15.	GABRB3--4	4.26	3.23E-02	42.	KIF13A--5	0.42	1.94E-02
16.	MUC2	4.17	2.44E-02	43.	SAMD5	0.38	2.02E-02
17.	ZMAT3--2	2.99	6.71E-03	44.	C6ORF150	0.37	2.42E-02
18.	MR1--6	2.93	3.28E-04	45.	CNTNAP2--2	0.36	1.72E-02
19.	GLT25D2	2.71	5.71E-03	46.	SLC26A2--2	0.29	3.73E-02
20.	GNE	2.68	3.94E-03	47.	SLC26A2	0.24	5.67E-02
21.	ZMAT3	2.61	3.36E-03	48.	CNTNAP2--3	0.24	5.79E-02
22.	CAPN8	2.59	1.43E-02	49.	UGT1A6	0.22	8.65E-03
23.	PLCB4--3	2.45	1.12E-02	50.	SPON1--4	0.16	1.01E-01
24.	N4BP2L2--4	2.19	3.00E-03	51.	BCAT1--4	0.12	2.99E-01
25.	N4BP2L2--5	2.16	1.55E-03	52.	FAM3B	0.12	1.35E-01
26.	NAV1--3	2.14	2.46E-03	53.	BCAT1--3	0.06	3.16E-01
27.	SLC16A6--2	2.03	5.55E-04	54.			

B.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	GCG	96.51	1.10E-01	26.	GLP1R--3	1.18	8.56E-05
2.	COL14A1	58.80	1.72E-01	27.	LOC100131510	1.17	7.08E-05
3.	CXCL5--3	50.11	1.02E-01	28.	MKI67IP	0.76	2.53E-04
4.	CXCL5--2	46.75	7.25E-02	29.	CDK12--5	0.73	8.98E-04
5.	HAPLN1	27.60	4.50E-02	30.	POLR2G	0.70	5.84E-04
6.	HEPACAM2	20.93	1.10E-01	31.	C1ORF124	0.70	5.99E-04
7.	APOA2--2	19.80	1.75E-01	32.	GPAM--2	0.58	1.10E-04
8.	HAPLN1--3	16.65	5.62E-02	33.	NDRG3--3	0.58	2.19E-03
9.	MUC2	9.15	1.26E-02	34.	RBL1--4	0.56	4.64E-03
10.	LOC100289026	9.10	1.39E-02	35.	PCYOX1	0.53	1.90E-03
11.	FCGBP	7.60	1.42E-06	36.	MOSC2--3	0.50	1.35E-02
12.	DNAJC12	7.14	1.02E-02	37.	BMP7	0.44	2.44E-02
13.	RETNLB	6.79	1.27E-02	38.	BMP7--3	0.44	2.65E-02
14.	MATN2	6.44	5.79E-02	39.	BMP7--2	0.34	4.31E-02
15.	LRRC19	5.11	1.26E-02	40.	WWTR1--2	0.16	5.44E-02
16.	DNAJC12--2	4.86	1.07E-02	41.	FN1--3	0.14	1.41E-01
17.	METTL7A	4.70	1.78E-03	42.	FAM3B	0.11	1.59E-01
18.	CELF2--4	3.84	7.04E-03	43.	FZD10	0.10	1.38E-01
19.	CELF2--5	3.27	7.08E-03	44.	IGF2 /// INS-IGF2--3	0.10	1.75E-01
20.	SEMA6A--4	2.83	6.73E-03	45.	IGF2 /// INS-IGF2--2	0.07	1.77E-01
21.	TMEM173	2.67	2.89E-03	46.	IGF2 /// INS-IGF2	0.02	1.36E-01
22.	HES1--2	1.83	1.02E-03				
23.	GABBR2--4	1.31	4.88E-05				
24.	LOC339822	1.28	1.73E-05				
25.	RBM11	1.27	1.03E-04				

C.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	GCG	31.81	3.69E-01	34.	FARS2	1.18	6.10E-04
2.	COL14A1	19.63	4.40E-01	35.	OR7E24	0.87	6.98E-04
3.	CXCL5--3	17.33	3.44E-01	36.	SPOP--3	0.82	2.98E-04
4.	CXCL5--2	16.62	2.88E-01	37.	IDH3A--3	0.78	1.08E-04
5.	REG1A	13.13	7.97E-02	38.	YAP1--3	0.71	8.18E-04
6.	LEFTY1	9.73	2.17E-01	39.	YAP1--2	0.70	7.53E-04
7.	HAPLN1	9.61	2.66E-01	40.	PTPN14	0.64	4.66E-04
8.	HEPACAM2	9.01	2.62E-01	41.	KRT6B	0.60	1.55E-03
9.	APOA2--2	7.13	4.37E-01	42.	CDKN1C--3	0.59	1.86E-03
10.	OLFM4	6.57	2.77E-02	43.	NOTUM	0.59	9.25E-04
11.	SST	6.30	4.39E-01	44.	WWTR1--3	0.58	2.31E-03
12.	APOA2	6.05	4.38E-01	45.	LYPD3	0.54	3.97E-03
13.	HAPLN1--3	6.00	2.94E-01	46.	WWTR1	0.53	1.40E-03
14.	FCGBP	5.92	6.45E-03	47.	GRHL3	0.52	4.40E-03
15.	COL3A1	5.65	4.28E-01	48.	FYN--3	0.46	7.01E-03
16.	REG3A	5.37	1.74E-01	49.	FOXQ1	0.43	5.11E-04
17.	MUC2	4.44	7.49E-02	50.	MSX2--3	0.43	9.46E-03
18.	C15ORF48	4.37	3.50E-03	51.	KLK6	0.33	8.99E-03
19.	MATN2	4.22	7.10E-02	52.	CTGF	0.28	1.04E-02
20.	ABCB1 /// ABCB4	4.15	1.02E-02	53.	AKAP12--2	0.23	1.02E-02
21.	ABCB1	4.07	5.89E-03	54.	S100A4	0.21	4.65E-03
22.	ALOX5--2	3.89	4.55E-02	55.	WWTR1--2	0.20	1.50E-03
23.	RETNLB	3.72	6.81E-02	56.	KRT6A	0.17	2.25E-02
24.	METTL7A	3.46	1.38E-03	57.	IGF2 /// INS-IGF2--3	0.17	2.90E-02
25.	CELF2--4	3.02	5.68E-03	58.	KRT6B--2	0.14	8.11E-02
26.	ABCB1--2	2.95	6.37E-03	59.	SLC1A2--4	0.13	2.03E-01
27.	CELF2--5	2.93	4.41E-03	60.	IGF2 /// INS-IGF2--2	0.13	3.00E-02
28.	CCND2	2.85	8.28E-03	61.	FZD10	0.12	1.08E-02
29.	IQGAP2	2.57	6.72E-03	62.	IGF2 /// INS-IGF2	0.12	2.47E-02
30.	CCND2--3	2.28	1.62E-03	63.	ITM2A	0.12	1.09E-01
31.	MAP3K13--2	1.35	7.73E-04	64.	NTS	0.11	2.30E-01
32.	MAP3K13--3	1.33	3.87E-04	65.	FGF19	0.11	1.87E-01
33.	BCL2L14--4	1.27	8.57E-04	66.			

D.

№	Genes	Fold change	P-value	№	Genes	Fold change	P-value
1.	OLFM4	116.54	3.69E-01	34.	MGST2	2.03	1.24E-10
2.	REG4	36.65	4.40E-01	35.	SLC35D1--2	1.83	3.81E-09
3.	REG4--2	31.50	3.44E-01	36.	SPIN1	0.65	4.04E-11
4.	PIGR--2	23.13	2.88E-01	37.	BICD2--2	0.56	2.51E-10
5.	SMOC2	20.59	7.97E-02	38.	CAMSAP1L1--2	0.52	1.83E-10
6.	GCG	20.32	2.17E-01	39.	FXR1--3	0.51	2.49E-11
7.	TFF3	16.18	2.66E-01	40.	BCAS4	0.44	4.20E-11
8.	COL14A1	12.72	2.62E-01	41.	GADD45B	0.33	2.49E-09
9.	NTS	12.07	4.37E-01	42.	NXN	0.26	1.35E-08
10.	TFF1	10.73	2.77E-02	43.	MYBL1	0.25	9.67E-09
11.	PLA2G2A	9.84	4.39E-01	44.	TIMP2--3	0.22	2.28E-08
12.	SPINK1	8.98	4.38E-01	45.	RDX--2	0.22	8.17E-09
13.	KIAA1324--2	8.49	2.94E-01	46.	RDX--3	0.22	7.47E-09
14.	LGR5--2	7.97	6.45E-03	47.	RDX	0.21	3.93E-09
15.	IL33	7.72	4.28E-01	48.	KLHL5	0.15	1.50E-07
16.	TOX3--3	7.42	1.74E-01	49.	FERMT2	0.15	6.70E-07
17.	TOX3	7.21	7.49E-02	50.	CYR61--2	0.13	9.10E-07
18.	TOX3--2	6.76	3.50E-03	51.	CAV1--2	0.12	1.75E-06
19.	C9ORF152	6.43	7.10E-02	52.	GJC1--3	0.12	7.01E-06
20.	DPEP1	6.32	1.02E-02	53.	CAV1	0.11	4.98E-06
21.	PPP1R1B	5.98	5.89E-03	54.	FRMD6--2	0.10	3.61E-07
22.	LGALS4	4.64	4.55E-02	55.	SLC2A3	0.10	6.89E-05
23.	AGR2	4.47	6.81E-02	56.	CYR61	0.10	6.25E-08
24.	POF1B--3	4.02	1.38E-03	57.	FRMD6	0.10	1.42E-06
25.	VIL1--4	3.82	5.68E-03	58.	FERMT2--2	0.09	3.67E-07
26.	FOXA2	3.78	6.37E-03	59.	WNT5A	0.09	1.06E-04
27.	VIL1--3	3.60	4.41E-03	60.	TUSC3--3	0.08	2.79E-04
28.	MYB	3.52	8.28E-03	61.	AKAP12--2	0.07	2.68E-04
29.	STARD10	2.89	6.72E-03	62.	AXL--2	0.06	8.95E-05
30.	STARD10--2	2.71	9.06E-10	63.	GNG11	0.04	3.75E-03
31.	TST	2.56	1.32E-09	64.	TFPI2--2	0.04	8.83E-03
32.	ALDH2	2.26	2.52E-09	65.	IGFBP7--2	0.03	3.11E-02
33.	FAM108C1	2.23	9.64E-10	66.			

Code Appendix 1: Single-cell data analysis in scRNA-seq data of normal colon

```
## Recreate the work in the SingleCell jupyter notebook on Alison Simmons data
from utils_plotting import plot_embedding, plot_embedding_binary,
plot_embedding_continuous, plot_embedding_binary_multigene, calc_volcano_data
import matplotlib.pylab as plt
import seaborn as sns
import numpy as np
from sklearn.cluster import KMeans
import pandas as pd
import umap

DATA_PATH = "rsc/data/Simmons/Healthy"
SEED = 42

## Load in the processed data
all_data = pd.read_feather(f"DATA_PATH/Alison_Simmons_Healthy.feather")
all_data = all_data.set_index("index")
all_data.index.name = None
log2_data = np.log2(all_data + 1)

## Run UMAP
embedding = umap.UMAP(random_state=SEED, n_components=2,
verbose=True).fit_transform(all_data)

## Plot the overall distribution
embedding_df = pd.DataFrame(embedding, columns=["UMAP1", "UMAP2"])
embedding_df.index = all_data.index
plot_embedding_binary(embedding_df, log2_data, "MUC2", "RdBu",
legend_title=”MUC2 log2(expression level)”)
plot_embedding_embedding_continuous(embedding_df, log2_data, "MUC2", "RdBu",
legend_title=”MUC2 log2(expression level)”)
plot_embedding_binary_multigene(embedding_df, all_data, gene_set = ["MUC2",
"TFF3", "MUC2/TFF3"], set_colours = "None": 'lightgray', "MUC2": 'red', "TFF3": 'blue',
"MUC2/TFF3": 'green')

## Run clustering on it
kmeans = KMeans(n_clusters=6, n_init=10, random_state=SEED).fit(embedding)
plot_embedding(embedding_df, hue=kmeans.labels_, palette=”Dark2”)
```

```
## Separate out the different groups

labelled_data = all_data.copy()
labelled_data.loc[:, “cluster”] = kmeans.labels
labelled_data.loc[:, [“UMAP1”, “UMAP2”] = embedding_df

## Volcano plot data between goblet and all not immune

Goblet_Data = labelled_data.query(‘cluster == 1’).drop(columns=[“cluster”, “UMAP1”, “UMAP2”])
Background_data = labelled_data.query(‘cluster !=1’).query(‘cluster !=3’).drop(columns=[“cluster”, “UMAP1”, “UMAP2”])
cluster_volcano_df = calc_volcano_data(goblet_data, background_data)
cluster_volcano_df.to.csv(“out/goblet_cluster_comparison.csv”)

## Volcano plot data of left and right section of goblet cells

goblet_indexes = goblet_data.index
goblet_embedding = embedding_df.loc[goblet_indexes, :]
log2_goblet_data = np.log2(goblet_data + 1)
for gene in [‘AGR2’, ‘FCGBP’, ‘ZG16’]:
        plot_embedding_continuous(goblet_embedding, log2_goblet_data, gene, ‘RdBu’)

left_upreg, right_upreg = [], []
for thresh in np.array(list(range(50, 61, 1)))/10:
left_indexes = goblet_embedding.query(f”UMAP < (thresh)”).index
right_indexes = goblet_embedding.query(f”UMAP > (thresh)”).index
cluster_volc_df = calc_volcano_data(
        log2_goblet_data.loc[left, indexes, :],
        log2_goblet_data.loc[right_indexes, :],
)
cluster_volcano_df.to.csv(f”out/goblet_cells_volcano_(thresh).csv”)

## Special thankful to James Ashford
```

Code Appendix 2: Single-cell data analysis in LS180-COLO678 cell lines

```
## Recreate the work in the SingleCell jupyter notebook on LS180-COLO678 data
from utils_plotting import plot_embedding, plot_embedding_binary,
plot_embedding_binary_multigene, plot_embedding_continuous,
gene_expression_volcano, calc_volcano_data, multi_label
from utils_twobytwo import twobytwo_comparison
from sklearn.cluster import KMeans
import pandas as pd
import numpy as np
import umap
DATA_PATH = "rsc/data/LS180_COLO678"
SEED = 42

## Load in the processed data
all_data = pd.read_feather(f"(DATA_PATH)/filtered_data.feather")
all_data = all_data.set_index("index")
all_data.index.name = None

## Run UMAP
embedding = umap.UMAP(random_state=SEED, n_components=2,
verbose=True).fit_transform(all_data)

## Plot the overall distribution
embedding_df = pd.DataFrame(embedding, columns=["UMAP1", "UMAP2"])
plot_embedding_binary(embedding_df, all_data, "MUC2", "Oranges")
plot_embedding_continuous(embedding_df, all_data, "MUC2", "Oranges")
plot_embedding_binary_multigene(embedding_df, all_data, gene_set = ["MUC2",
"TFF3", "MUC2/TFF3"], set_colours = ("None": 'lightgray', "MUC2": 'red', "TFF3":
'blue', "MUC2/TFF3": 'green'))

## Run clustering on it
kmeans = KMeans(n_clusters=2, n_init=10, random_state=SEED).fit(embedding)
plot_embedding(embedding_df, hue=np.array(["COLO678",
"LS180"])[kmeans.labels_])

## Compare the expressions between the two clusters
c0_data = all_data.iloc[kmeans.labels_ == 0, :]
c1_data = all_data.iloc[kmeans.labels_ == 1, :]
cluster_volc_df = calc_volcano_data(c0_data, c1_data)
cluster_volc_df.to_csv("out/cluster_comparison.csv")
```

```
## Extract occurrence counts of sets of gene combinations in each cluster
_, _, counts = multi_label(c1_data, gene_set = ["MUC2", "TFF3", "MUC2\TFF3"])
print(counts)

## Compare gene expression between MUC2 positive barcodes and all cells
gene_expression_volcano(c1_data, "MUC2")
gene_expression_volcano(c1_data, "TFF3")
gene_expression_volcano(c0_data, "ALPP")

## 2x2 tables
categorise = lambda x: pd.Categorical(pd.Series(x > 0, dtype=int), categories=[0, 1])
c0_categorised_data = c0_data.apply(categorise)
c1_categorised_data = c1_data.apply(categorise)
twobytwo_comparison(c1_categorised_data, "MUC2")
twobytwo_comparison(c1_categorised_data, "TFF3")
twobytwo_comparison(c0_categorised_data, "ALPP")

## Special thankful to James Ashford
```