Supplementary data 1. 117 genes with elevated expression in the brain metastasis group of the discovery database (P value<0.05)

IFNA1/13, MR1, CAMP, AHR, CD1A, IL4R, IL3, TMEM173, DPP4, FCAR, C9, IL13RA1, CD59, IFNA2, PTPN22, TNFRSF9, FADD, FCER1A, APP, ICAM1, IFNAR2, TNFRSF10C, RARRES3, DEFB103A, IL7, CD99, IL13, CX3CL1, HFE, PDGFB, PDGFRB, IL26, ARHGDIB, CCL16, XCL1, CD74, ATG10, IL22, CDKN1A, RUNX1, CD276, CSF2, IL23R, IGF2R, HLA-DMB, BATF3, CD46, TNFSF11, ICAM5, FCGR1A/B, CCL22, CX3CR1, C6, SOCS1, IFNGR1, KLRF1, CEACAM6, IL6ST, CCL23, C8B, ENTPD1, NOTCH2, IL17F, CFB, ZEB1, MUC1, THY1, FN1, C8A, CCRL1, TGFBI, PYCARD, CCL18, PLAU, PSMB8, HLA-DMA, STAT3, BST2, IL1R1, IL12B, CD160, CD55, RORC, CD83, CFI, EGR1, ICAM4, KIR_Activating_Subgroup_2, UBE2L3, NT5E, CLEC7A, CCND3, CEACAM1, TLR5, IL19, IL1RL1, CSF1R, MARCO, CFH, NFKBIA, CCBP2, CASP10, CTSS, TBK1, TGFBR1, IL5, FCGRT, CCL15, C4A/B, TNFSF4, NOD1, MCL1, TCF4, CCL11, KLRC4, IFNB1, LY96

Supplementary data 2. 98 genes with elevated expression in the brain metastasis group of the validation database (P value<0.05)

IFIT2, DPP4, MX1, RUNX1, NOD1, PRKCD, ICAM5, ZBTB16, FCGRT, RORC, TLR4, RARRES3, STAT1, FKBP5, CD74, MME, KIR_Activating_Subgroup_1, MSR1, TNFSF12, IFIH1, TYK2, MUC1, HLA-A, STAT2, PSMB9, MCL1, CTSS, IRF5, TRAF6, TAP1, HLA-C, IRF7, IFI35, HLA-B, PPBP, LILRB4, TOLLIP, SELPLG, PSMB8, MYD88, HLA-DRA, CIITA, FCGR2A/C, CD81, TLR5, BST2, CD22, CD163, HLA-DPB1, IL6ST, ICAM1, TGFBR2, STAT5A, CCL8, IRF8, AHR, IL18R1, ICAM4, HLA-DPA1, IFNAR2, STAT3, MRC1, HLA-DMB, PTPN22, HAVCR2, CSF3R, HLA-DRB1, SERPING1, CASP1, KLRC2, PPARG, CD97, C1QB, LAMP3, NFATC1, MAP4K2, HLA-DMA, TNFRSF14, ATG16L1, IL7R, IL1RL1, C1QA, STAT4, LAIR1, CSF2, TNFSF8, EGR1, CCL18, IL6R, PML, KLRK1, PAX5, LILRA4, CISH, HLA-DQB1, TNFSF15, TLR2, CCL23

Supplementary data 3. 28 genes with elevated expression in the brain metastasis group in the discovery and validation database

AHR, DPP4, PTPN22, ICAM1, IFNAR2, RARRES3, CD74, RUNX1, CSF2, HLA-DMB, ICAM5, IL6ST, CCL23, MUC1, CCL18, PSMB8, HLA-DMA, STAT3, BST2, RORC, EGR1, ICAM4, TLR5, IL1RL1, CTSS, FCGRT, NOD1, MCL1

Table S1 Demographic and clinical characteristics of patients

Variable	Discovery set, number (%)	Validation set, number (%)
Age, median (range) (years)	66 (45–84)	58 (35–86)
Male sex	24 (66.7%)	20 (55.6%)
Smoking history	23 (67.6%)	19 (57.6%)
Pathologic TNM stage		
I	8 (22.2%)	6 (16.7%)
II	10 (27.8%)	10 (27.8%)
III	18 (50%)	20 (55.6%)
Histologic subtype		
Acinar	17 (47.2%)	14 (38.9%)
Papillary	5 (13.9%)	5 (13.9%)
Lepidic	2 (5.6%)	2 (5.6%)
Micropapillary	2 (5.6%)	4 (11.1%)
Solid	10 (27.8%)	11 (30.6%)

Smoking history was collected for 67 patients.

http://dx.doi.org/10.21037/tlcr-20-1056

Table S2 Information on the used antibodies for immunohistochemistry

Name	Clonality	Clone name	Catalog name	Dilution	Manufacturer
DPP4	Rabbit polyclonal		ab231973	1:80	Abcam
ICAM1	Mouse monoclonal	15.2	sc-107	1:50	Santa Cruz
RARRES3	Rabbit polyclonal		NBP1-59395	1:10	Novus Biologicals
CD74	Mouse monoclonal	PIN.1	NB100-1985	1:300	Novus Biologicals
CSF2	Mouse monoclonal	OTI8G5	NBP2-46364	1:50	Novus Biologicals
HLA-DMB	Rabbit polyclonal		21704-1-AP	1:50	Proteintech
ICAM5	Rabbit polyclonal		bs-6686R	1:400	Bioss
MUC1	Mouse monoclonal	E29	M0613	1:25	Dako
CCL18	Rabbit polyclonal		Ab104867	1:200	Abcam
RORC	Rabbit polyclonal		CSB-PA020071LA01HU	1:200	Cusabio
ICAM4	Rabbit polyclonal		CSB-PA440434	1:50	Cusabio
CCL23	Rabbit monoclonal	EPR11363	ab171751	1:50	Abcam
IFNAR2	Rabbit polyclonal		MBS9607261	1:50	Mybiosource
IL6ST	Rabbit polyclonal		NBP2-15776	1:50	Novus Biologicals

Table S3 KEGG pathway analysis in 28 genes

KEGG pathway	P value	Fold enrichment
Inflammatory bowel disease (IBD)	1.89E-05	28.28536184
Antigen processing and presentation	9.39E-04	19.05540166
Rheumatoid arthritis	0.001437	16.4569378
Cytokine-cytokine receptor interaction	0.003143	7.449642625
HTLV-I infection	0.003689	7.127020307
Jak-STAT signaling pathway	0.005934	9.987658802
Staphylococcus aureus infection	0.008539	20.11403509
Viral myocarditis	0.009479	19.05540166
Influenza A	0.009813	8.323049002
Tuberculosis	0.010282	8.181980375
Herpes simplex infection	0.011258	7.913718723
Toxoplasmosis	0.032815	9.874162679
Natural killer cell mediated cytotoxicity	0.039663	8.902933563

http://dx.doi.org/10.21037/tlcr-20-1056

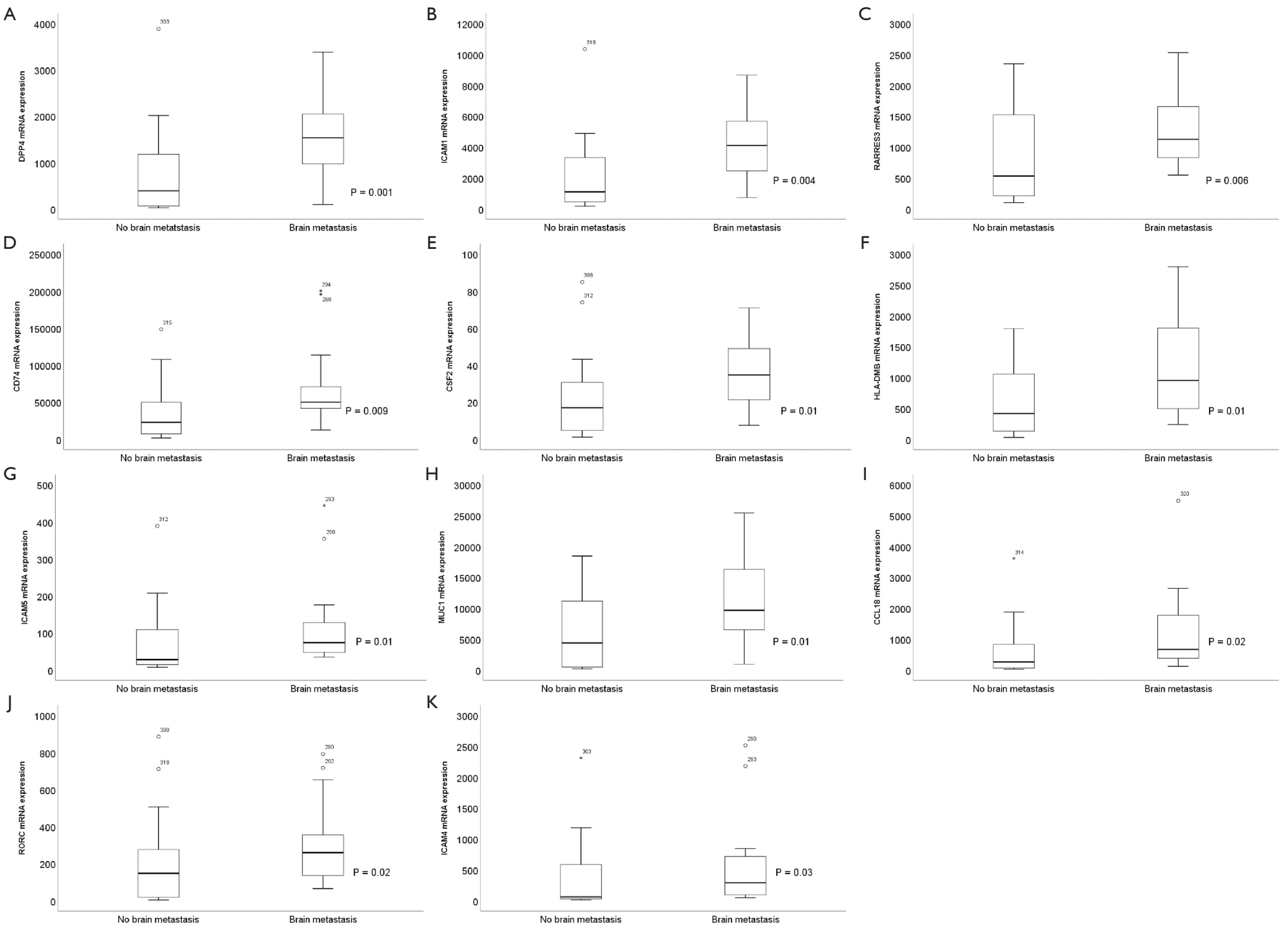

Figure S1 Box plots for 11 genes in discovery set. (A) *DPP4*, (B) *ICAM1*, (C) *RARRES3*, (D) *CD74*, (E) *CSF2*, (F) *HLA-DMB*, (G) *ICAM5*, (H) *MUC1*, (I) *CCL18*, (J) *RORC*, (K) *ICAM4*.

http://dx.doi.org/10.21037/tlcr-20-1056

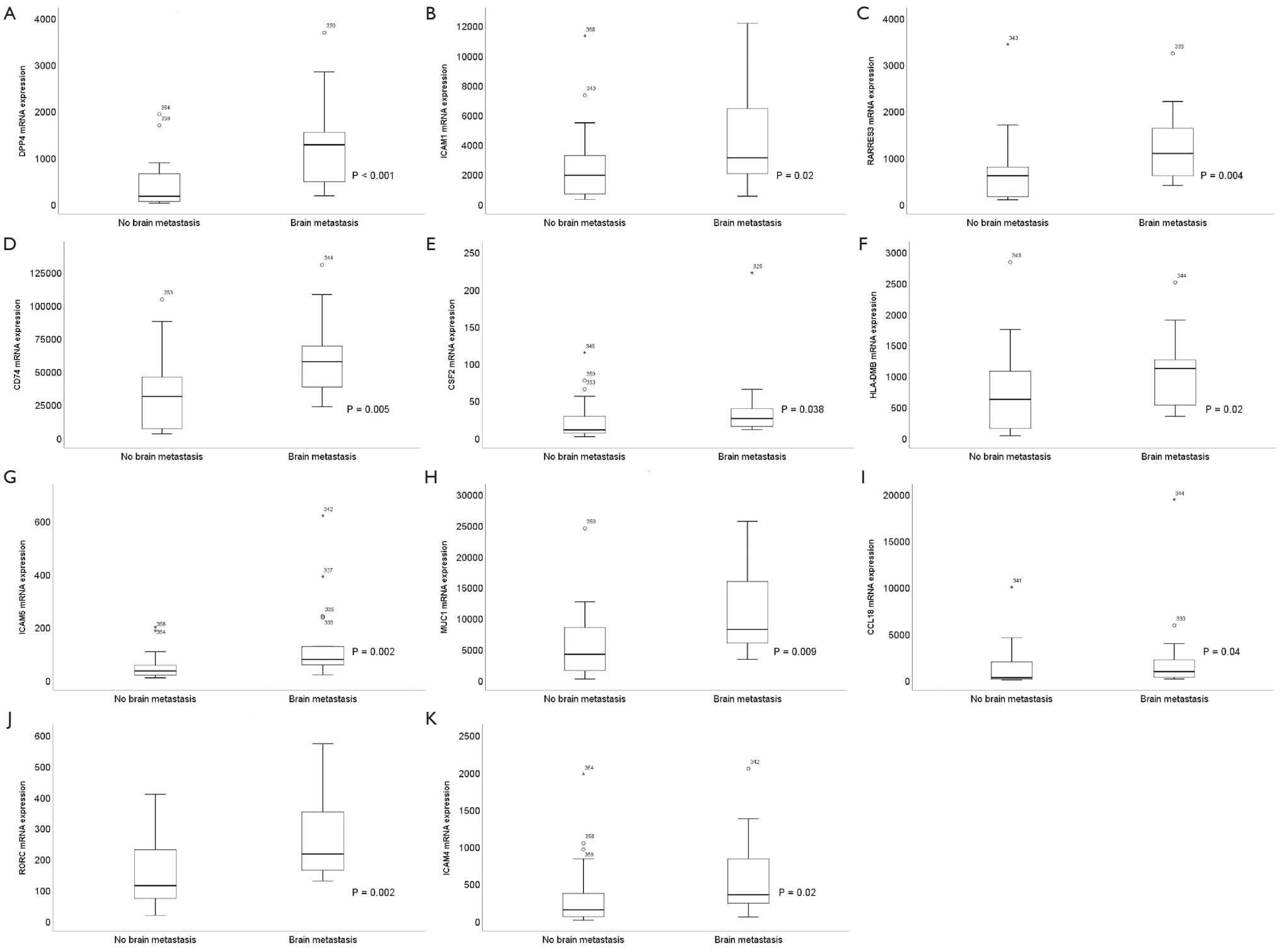

Figure S2 Box plots for 11 genes in validation set. (A) *DPP4*, (B) *ICAM1*, (C) *RARRES3*, (D) *CD74*, (E) *CSF2*, (F) *HLA-DMB*, (G) *ICAM5*, (H) *MUC1*, (I) *CCL18*, (J) *RORC*, (K) *ICAM4*.

http://dx.doi.org/10.21037/tlcr-20-1056

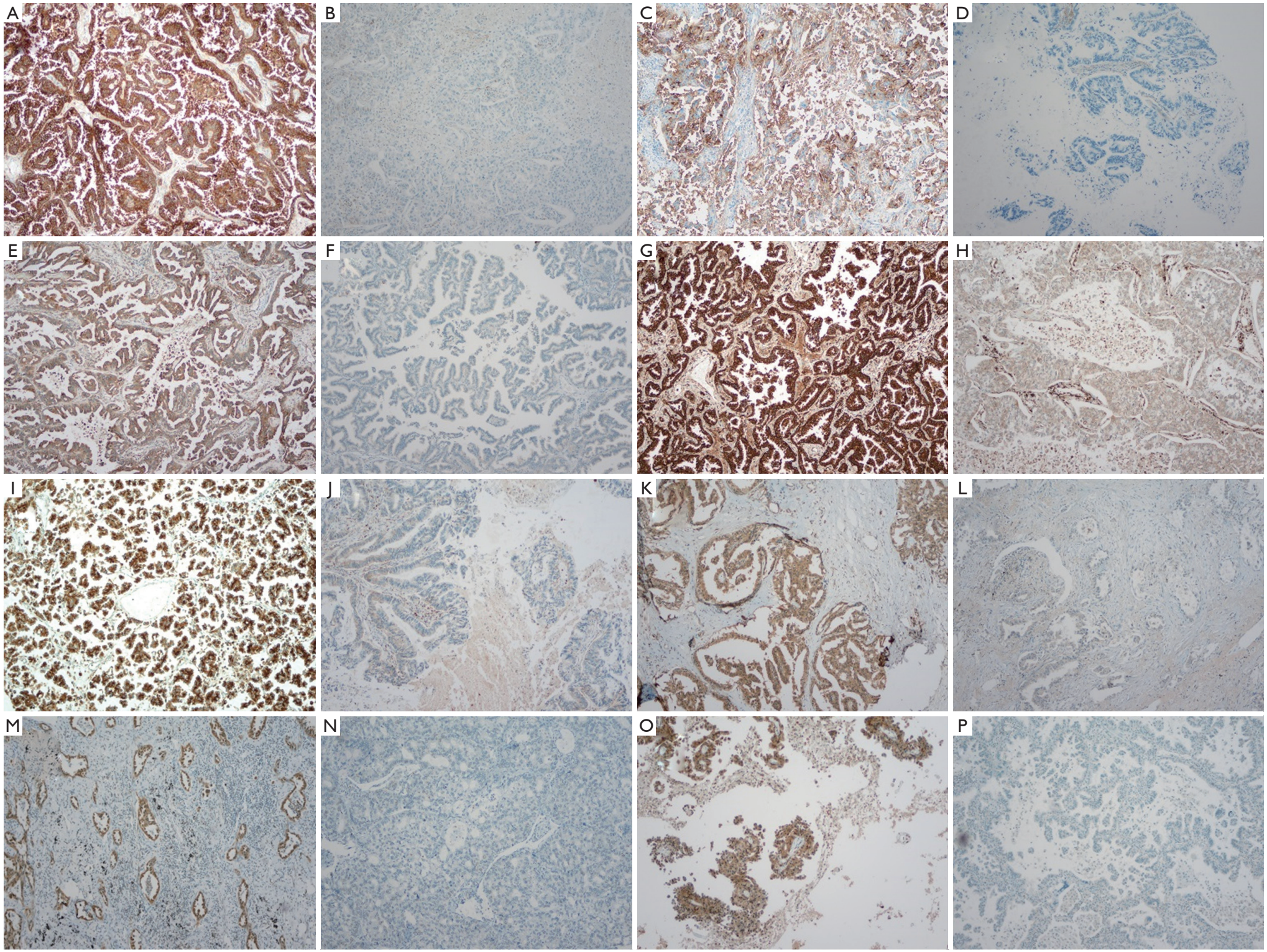

Figure S3 Low-magnification immunohistochemical expression in the group with or without brain metastases (×100). High *DPP4* (A), *ICAM1* (C), *RARRES3* (E), *CD74* (G), *HLA-DMB* (I), *ICAM5* (K), *CCL18* (M), *ICAM4* (O) expression in the group with brain metastases. Low *DPP4* (B), *ICAM1* (D), *RARRES3* (F), *CD74* (H), *HLA-DMB* (J), *ICAM5* (L), *CCL18* (N), *ICAM4* (P) expression in the group without brain metastases.

http://dx.doi.org/10.21037/tlcr-20-1056

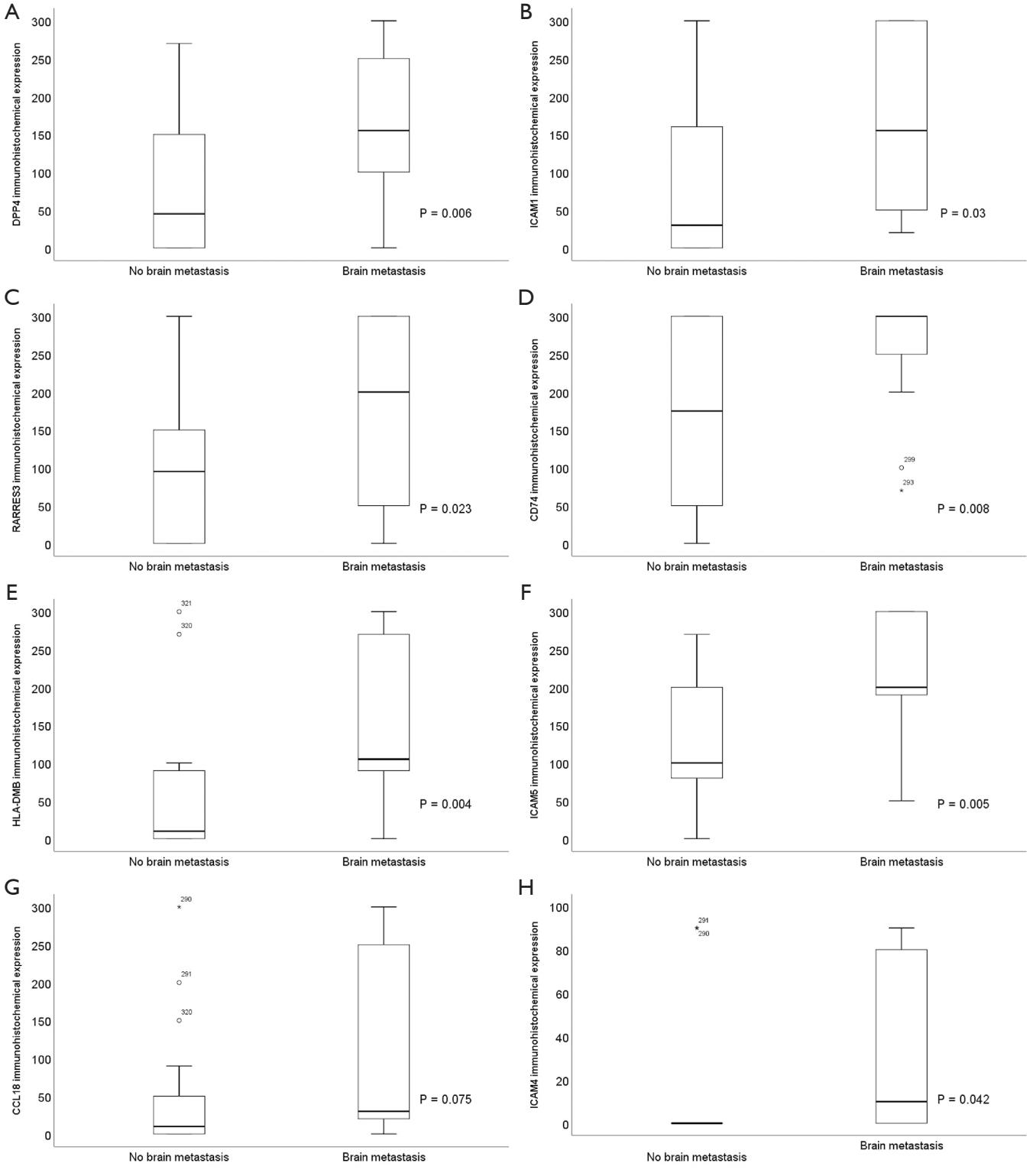

Figure S4 Box plots for 8 gene's immunohistochemical expression in discovery set. (A) *DPP4*, (B) *ICAM1*, (C) *RARRES3*, (D) *CD74*, (E) *HLA-DMB*, (F) *ICAM5*, (G) *CCL18*, (H) *ICAM4*.

http://dx.doi.org/10.21037/tlcr-20-1056

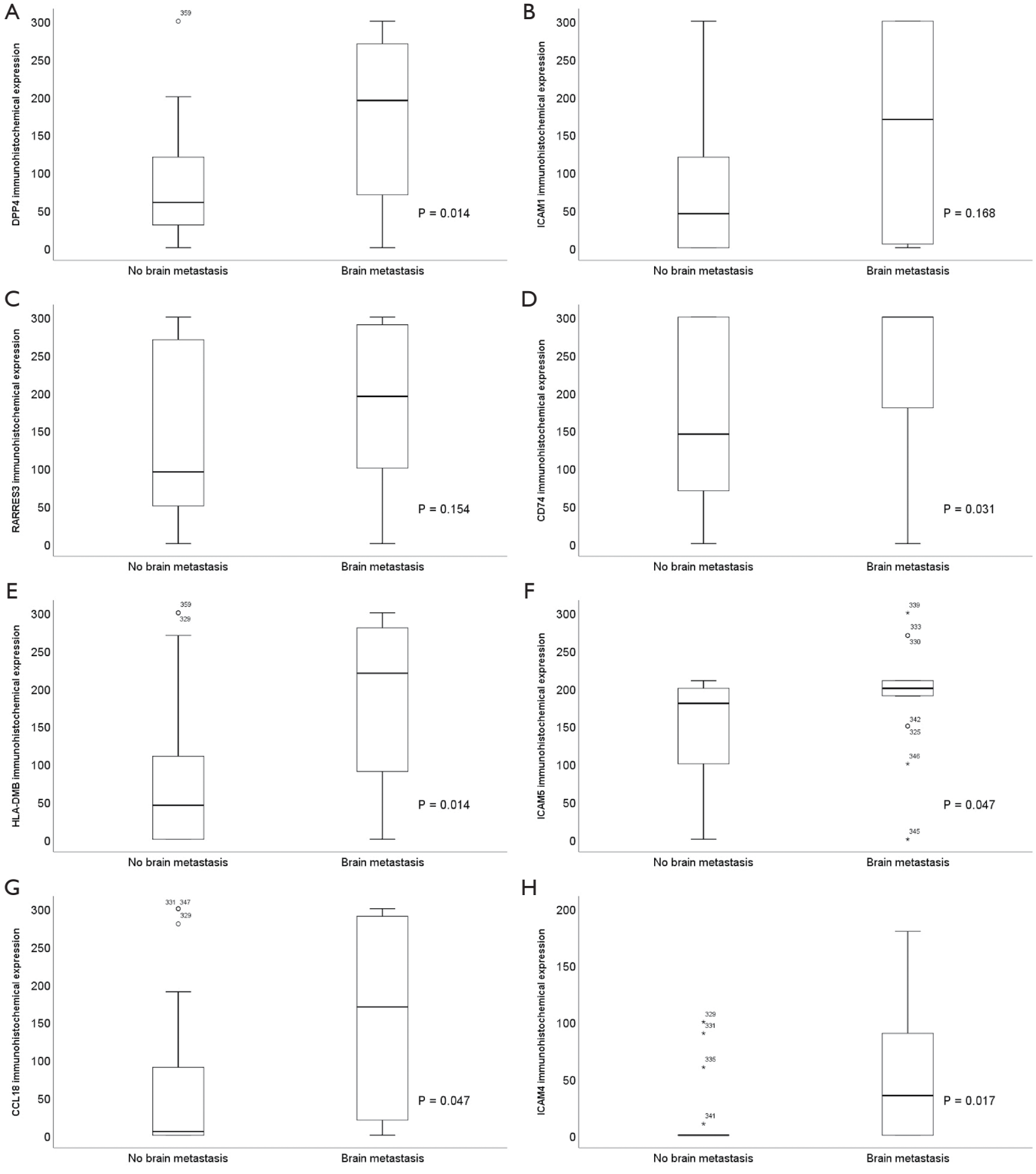

Figure S5 Box plots for 8 gene's immunohistochemical expression in validation set. (A) *DPP4*, (B) *ICAM1*, (C) *RARRES3*, (D) *CD74*, (E) *HLA-DMB*, (F) *ICAM5*, (G) *CCL18*, (H) *ICAM4*.

http://dx.doi.org/10.21037/tlcr-20-1056