

## Supporting information captions

### Supplementary Tables:

**Table S1.** Clinicopathological characteristics of patient cohorts I (CSS or OS)

**Table S2.** Clinicopathological characteristics of patient cohorts II (DFS)

**Table S3.** PCR primers used in this study

**Table S4.** Genes with strongest survival associations by SSAT analysis. A semi-supervised survival tool (SSAT), an R (Bioconductor) based script, was used *in silico* to identify genes whose expression correlated with overall survival in colon cancer. SSAT identified 400 and 269 such genes in two datasets, GSE17536 and GSE17537, respectively, of which 64 were in common to both datasets. Twenty genes, which were most significantly associated with survival based on a weighted rank score are shown in **Table S4**.

**Table S5.** MVA-1. A multivariate cox proportional hazard analysis (backward Wald) identified 3 genes (ULBP2, SEMA5A and PCDH7) that remained in the equation, in addition to stage. These 3 genes and stage were then used for a second MVA (backward Wald), this time using GSE17537 data. ULBP2 and SEMA5A were the only two genes whose upregulation were associated with worse and better overall survival, respectively, independent of stage in this analysis. Downregulation of ULBP2 and upregulation of SEMA5A was related to better overall survival.

**Table S6.** Univariate cox regression analyses of ULBP2 and SEMA5A gene expression with CSS

**Table S7.** Stage 2&3 restricted MVA of CSS with clinicopathological parameters and SU-GIB in GSE17536

**Table S8.** Cox regression analysis with GSE39582 MSS samples only.

**Table S9.** Summary of GIB and Oncotype MVA analyses. GSE17536: Stage, age, gender, grade, GIB (or Oncotype), MSI; GSE39582: Stage, age, gender, GIB (or Oncotype) and MSI

**Table S10.** Summary of GIB and Oncotype MVA analyses - GSE39582

**Table S11.** Drugs that are significantly correlated with both SEMA5A and ULBP2 expression in colon cancer cell lines according to CGP (as tested for colon cancer cell lines)

### Supplementary Figures

**Figure S1. Representative E&H stained tumor sections.** Tumors with well (A), intermediate (B), and poor differentiation (C) are shown. Arrow indicates a site with necrosis.

**Figure S2.** Survival graphs of colon cancer patients stratified based on either ULBP2 or SEMA5A gene expression I (GSE17536). Kaplan-Meier analyses (in silico) comparing “high” and “low” expression groups as defined by SSAT cut-off values for ULBP2 (A) and SEMA5A (B), which were 4 and 6, respectively. Log-rank p values are indicated. Cancer specific survival is in months. C & D: Log-rank test results plotted at all cut-off values: the graphic indicates log-rank values (shown as dots) obtained at every possible cut-off for ULBP2 (C) and SEMA5A (D). A red dot indicates the p value is associated with a HR larger than 1, when the low expression group is the reference, while blue indicates the reverse. Horizontal dotted line p=0.05. Vertical dotted lines: (from left to right) first 25<sup>th</sup> %, median, and 75<sup>th</sup> percentile.

**Figure S3.** Survival graphs of colon cancer patients stratified based on either ULBP2 or SEMA5A gene expression II (GSE17537). Kaplan-Meier analyses (in silico) comparing “high” and “low” expression groups as defined by SSAT cut-off values for ULBP2 (A) and SEMA5A

(B), which were 4 and 6, respectively. Log-rank p values are indicated. Cancer specific survival is in months. C & D: Log-rank test results plotted at all cut-off values: the graphic indicates log-rank values (shown as dots) obtained at every possible cut-off for ULBP2 (C) and SEMA5A (D). A red dot indicates the p value is associated with a HR larger than 1, when the low expression group is the reference, while blue indicates the reverse. Horizontal dotted line  $p=0.05$ . Vertical dotted lines: (from left to right) first 25<sup>th</sup> %, median, and 75<sup>th</sup> percentile.

**Figure S4.** Kaplan-Meier graphs for CSS of GSE17536 stratified based on either ULBP2 or SEMA5A gene expression for stage 2&3 patients. Respective log-rank p values are shown. Survival times are in months. ULBP2 and SEMA5A expression can predict cancer specific survival when restricted to stage 2&3 patients. ULBP2 and SEMA5A cut-off values were 4 and 6, respectively.

**Figure S5.** Survival graphs of colon cancer patients stratified based on either ULBP2 or SEMA5A gene expression III. Kaplan-Meier analyses (*ex vivo*) comparing “high” and “low” expression groups based on the cut-off with the smallest log rank p value within the 25-75<sup>th</sup> % interquartile ranges for ULBP2 (A) and SEMA5A (B) for the Ankara cohort. Log-rank p values are indicated. Overall survival is given in months. C & D: Log-rank test results plotted at all cut-off values: the graphic indicates log-rank values (shown as dots) obtained at every possible cut-off for ULBP2 (C) and SEMA5A (D). A red dot indicates the p value is associated with a HR larger than 1, when the low expression group is the reference, while blue indicates the reverse. Horizontal dotted line  $p=0.05$ . Vertical dotted lines: (from left to right) first 25<sup>th</sup> %, median, and 75<sup>th</sup> percentile.

**Figure S6.** Kaplan-Meier graphs based on the SU-GIB signature for GSE17536 (A), GSE17537 (B) and the Ankara cohort (C); restricted to stage 2&3 patients for *in silico* (A&B) and stage 3 patients for *ex vivo* (C) analysis. Respective log-rank p values are shown. Survival times are in months. SU-GIB signature cut-off values for ULBP2 and SEMA5A were 4 and 6, respectively for *in silico* analyses. For the Ankara cohort, cut-off values correspond to

the median values which were identical to the cut-off with the smallest log rank p value within the 25-75<sup>th</sup> % interquartile range for both genes.

**Figure S7.** Kaplan-Meier graphs for CSS of GSE17536 microsatellite-unstable (MSI) and -stable (MSS) patients as stratified by SU-GIB. ULBP2 and SEMA5A cut-offs were determined previously SSAT (4 and 6, respectively). \* MSI and MSS subgroups were predicted in silico (see Methods). The GSE17536 cohort, separated into MSI and MSS subgroups shows that the G, I and B stratification results in the expected trend in either MSI or MSS patient groups, although only the latter shows significance by the log-rank test.

**Figure S8.** Kaplan-Meier graphs for DFS of GSE17536 and GSE17537 patients as stratified by SU-GIB. G, I and B stratification of patients in both cohorts reveals that this can predict disease-free survival in both cohorts. ULBP2 and SEMA5A cut-offs were determined by SSAT (4 and 6, respectively). Log-rank test p values are shown.

**Figure S9.** TCGA based proteome analysis of colon cancer tumor tissue. RNA seq and proteome data for 132 colon cancer primary tumor tissues downloaded from "cancergenome.nih.gov" via the TCGA data portal were classified according to the SU signature (bad survivors: 40, good survivors: 37, intermediate survivors: 55) revealed increased phosphorylation at EGFR 992 (right) and Shc phosphorylation (left) among patients with better prognosis ( $p < 0.01$  and  $p < 0.001$ , respectively, 1-way Anova).

**Figure S10.** TCGA based RNAseq analysis of colon cancer tumor tissue. TCGA colon tumor samples were divided into GIB groups using medians as cut-offs for ULBP2 and SEMA5A. RPKM values of TNF, TGFB3 and IL1R2 are plotted for 'good', 'intermediate' and 'bad' groups. The median and inter-quartile range for each group is indicated. Ttest p values between 'good' and 'bad' groups, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ . RPKM: reads per kilobase per million mapped reads.

**Figure S11.** SU-GIB stratification is associated with EMT. Scatter plot for tumors in the GSE39582 dataset based on E-cadherin (201131\_S\_AT) and Vimentin (201426\_S\_AT) gene expression. While most of the patients with a good prognosis are also “E-cadherin high, Vimentin low”, patients with bad prognosis show the opposite gene expression pattern (Chi-sq.  $p < 0.00001$ ). Median expression was used as the threshold for both genes. Dotted lines indicate median expression values.

Supplementary Table 1. Clinicopathological characteristics of patient cohorts I (CSS or OS)

| Patient characteristics            | CSS 1:<br>GSE17536 (n = 177) |        |        | CSS 2:<br>GSE17537 (n = 55) |         |        | OS: Ankara cohort<br>(n = 48) |       |       |
|------------------------------------|------------------------------|--------|--------|-----------------------------|---------|--------|-------------------------------|-------|-------|
|                                    | Nr. <sup>#</sup>             | HR*    | P      | Nr.                         | HR      | P      | Nr.                           | HR    | P     |
| <b>Age</b>                         |                              |        |        |                             |         |        |                               |       |       |
| <65 (ref.)                         | 83                           | 0.808  | 0.432  | 33                          | 0.992   | 0.990  | 24                            | 2.173 | 0.104 |
| >65                                | 94                           |        |        | 22                          |         |        | 22                            |       |       |
| Unknown                            | -                            |        |        | -                           |         |        | 2                             |       |       |
| <b>Gender</b>                      |                              |        |        |                             |         |        |                               |       |       |
| Male (ref.)                        | 96                           | 0.918  | 0.534  | 26                          | 2.003   | 0.269  | 24                            | 0.857 | 0.740 |
| Female                             | 81                           |        |        | 29                          |         |        | 24                            |       |       |
| <b>TNM Stage**</b>                 |                              |        |        |                             |         |        |                               |       |       |
| 1                                  | 24                           | 3.623  | <0.001 | 4                           | 14.085  | 0.001  | 3                             | 4.434 | 0.001 |
| 2                                  | 57                           |        |        | 15                          |         |        | 0                             |       |       |
| 3                                  | 57                           |        |        | 19                          |         |        | 38                            |       |       |
| 4                                  | 39                           |        |        | 17                          |         |        | 5                             |       |       |
| Unknown                            | -                            |        |        | -                           |         |        | 2                             |       |       |
| <b>Recurrence</b>                  |                              |        |        |                             |         |        |                               |       |       |
| No (ref.)                          | 109                          | 42.869 | <0.001 | 36                          | 364.058 | <0.001 | NA                            |       |       |
| Yes                                | 36                           |        |        | 19                          |         |        | NA                            |       |       |
| Other/Unknown                      | 32                           |        |        | -                           |         |        | NA                            |       |       |
| <b>Grade</b>                       |                              |        |        |                             |         |        |                               |       |       |
| Well Differentiated                | 16                           | 2.141  | 0.005  | 1                           | 3.771   | 0.119  | 13                            | 1.279 | 0.578 |
| Moderately Differentiated          | 134                          |        |        | 32                          |         |        | 32                            |       |       |
| Poorly Differentiated              | 27                           |        |        | 3                           |         |        | 3                             |       |       |
| Other/Unknown                      | -                            |        |        | 19                          |         |        | -                             |       |       |
| <b>Perineural Invasion</b>         |                              |        |        |                             |         |        |                               |       |       |
| 0 (ref.)                           | NA                           |        |        | NA                          |         |        | 31                            | 4.131 | 0.003 |
| 1                                  | NA                           |        |        | NA                          |         |        | 15                            |       |       |
| Other/Unknown                      | NA                           |        |        | NA                          |         |        | 2                             |       |       |
| <b>Vascular Invasion</b>           |                              |        |        |                             |         |        |                               |       |       |
| 0 (ref.)                           | NA                           |        |        | NA                          |         |        | 23                            | 3.739 | 0.012 |
| 1                                  | NA                           |        |        | NA                          |         |        | 23                            |       |       |
| Other/Unknown                      | NA                           |        |        | NA                          |         |        | 2                             |       |       |
| <b>Microsatellite Instability‡</b> |                              |        |        |                             |         |        |                               |       |       |
| Low or stable (ref.)               | 146                          | 1.319  | 0.415  | NA                          |         |        | NA                            |       |       |
| High                               | 31                           |        |        | NA                          |         |        | NA                            |       |       |

\*Cox proportional hazards regression

\*\*Stage: For Ankara cohort - treated as a continuous variable (1: stage 1, 2: stage 3A and 3B, 3: stage 3C and 4). For GSE17536 and GSE17537 - treated as a continuous variable (1, 2, 3, 4)

‡MSI status: Determined in silico (see Methods)

#Nr.: case numbers

**Supplementary Table 2. Clinicopathological characteristics of patient cohorts II (DFS)**

| Patient characteristics            | GSE17536 (n = 177) |       |              | GSE17537 (n = 55) |        |                  | GSE39582 (n=566) |       |                  |
|------------------------------------|--------------------|-------|--------------|-------------------|--------|------------------|------------------|-------|------------------|
|                                    | Nr. <sup>#</sup>   | HR*   | P            | Nr.               | HR     | P                | Nr.              | HR    | P                |
| <b>Age</b>                         |                    |       |              |                   |        |                  |                  |       |                  |
| ≤65 (ref.)                         | 83                 | 0.585 | 0.114        | 33                | 0.677  | 0.429            | 150              | 0.859 | 0.317            |
| >65                                | 94                 |       |              | 22                |        |                  | 415              |       |                  |
| Unknown                            | -                  |       |              | -                 |        |                  | 1                |       |                  |
| <b>Gender</b>                      |                    |       |              |                   |        |                  |                  |       |                  |
| Male (ref.)                        | 96                 | 1.000 | 0.999        | 26                | 2.037  | 0.150            | 310              | 0.790 | 0.124            |
| Female                             | 81                 |       |              | 29                |        |                  | 256              |       |                  |
| <b>TNM Stage**</b>                 |                    |       |              |                   |        |                  |                  |       |                  |
| 0                                  | -                  | 2.047 | <b>0.001</b> | -                 | 13.988 | <b>&lt;0.001</b> | 4                | 2.605 | <b>&lt;0.001</b> |
| 1                                  | 24                 |       |              | 4                 |        |                  | 33               |       |                  |
| 2                                  | 57                 |       |              | 15                |        |                  | 264              |       |                  |
| 3                                  | 57                 |       |              | 19                |        |                  | 205              |       |                  |
| 4                                  | 39                 |       |              | 17                |        |                  | 60               |       |                  |
| <b>Grade</b>                       |                    |       |              |                   |        |                  |                  |       |                  |
| Well Differentiated                | 16                 | 1.848 | 0.076        | 1                 | 1.347  | 0.721            | NA               |       |                  |
| Moderately Differentiated          | 134                |       |              | 32                |        |                  | NA               |       |                  |
| Poorly Differentiated              | 27                 |       |              | 3                 |        |                  | NA               |       |                  |
| Other/Unknown                      | -                  |       |              | 19                |        |                  | NA               |       |                  |
| <b>Microsatellite Instability‡</b> |                    |       |              |                   |        |                  |                  |       |                  |
| Low or stable (ref.)               | 146                | 1.152 | 0.285        | NA                |        |                  | NA               |       |                  |
| High                               | 31                 |       |              | NA                |        |                  | NA               |       |                  |
| <b>Mismatch repair</b>             |                    |       |              |                   |        |                  |                  |       |                  |
| Proficient (ref.)                  | NA                 |       |              | NA                |        |                  | 444              | 0.358 | <b>0.002</b>     |
| Deficient                          | NA                 |       |              | NA                |        |                  | 75               |       |                  |
| Unknown                            | NA                 |       |              | NA                |        |                  | 47               |       |                  |
| <b>KRAS or BRAF mutation</b>       |                    |       |              |                   |        |                  |                  |       |                  |
| Wild Type (ref.)                   | NA                 |       |              | NA                |        |                  | 255              | 1.210 | <b>0.019</b>     |
| Mutant                             | NA                 |       |              | NA                |        |                  | 268              |       |                  |
| Unknown                            | NA                 |       |              | NA                |        |                  | 43               |       |                  |

\*Cox proportional hazards regression

\*\*Stage: For Ankara cohort - treated as a continuous variable (1: stage 1, 2: stage 3A and 3B, 3: stage 3C and 4). For GSE17536 and GSE17537 - treated as a continuous variable (1, 2, 3, 4)

‡MSI status: Determined in silico (see Methods)

#Nr.: case numbers

**Supplementary Table 3. PCR primers used in this study**

|          |                             |
|----------|-----------------------------|
| SEMA5A_F | TCTCTCTCCTTGGCACTTCC        |
| SEMA5A_R | ACTGGATGCTCGGTTCTCTG        |
| ULBP2_F  | GTGCAGGAGCACCCTCG           |
| ULBP2_R  | CATACACCGTAGGTCGTGGG        |
| GAPDH_F* | GGAGCGAGATCCCTCCAAAAT       |
| GAPDH_R* | GGCTGTTGTCATACTTCTCATGG     |
| ACTB_F** | AGAGCTACGAGCTGCCTGAC        |
| ACTB_R** | AGCACTGTGTTGGCGTACAG        |
| 18S_F    | CGTGCATTTATCAGATCAAAACCAACC |
| 18S_R    | ATGGTAGGCACGGCGACTAC        |

\*References 8 and \*\*9



**Supplementary Table 4. Genes with strongest survival associations by SSAT**

| Gene           |   | NCBI gene ID | Rank |
|----------------|---|--------------|------|
| <b>ADAM12</b>  | (ADAM metallopeptidase domain 12)   | 8038         | 1    |
| <b>SFRP2</b>   | (secreted frizzled-related protein 2)   | 6423         | 2    |
| <b>ULBP2</b>   | (UL16 binding protein 2 )   | 80328        | 3    |
| <b>KAL1</b>    | (Kallmann syndrome 1 sequence)  | 3730         | 4    |
| <b>SEMA5A</b>  | (sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A) | 9037         | 5    |
| <b>PXDN</b>    | (peroxidasin homolog (Drosophila))  | 7837         | 6    |
| <b>PCDH7</b>   | (protocadherin 7)   | 5099         | 7    |
| <b>CTGF</b>    | (connective tissue growth factor)   | 1490         | 8    |
| <b>PRKD1</b>   | (protein kinase D1)   | 5587         | 9    |
| <b>COL11A1</b> | (collagen, type XI, alpha 1)  | 1301         | 10   |
| <b>EBF1</b>    | (early B-cell factor 1)   | 1879         | 11   |
| <b>SCFD1</b>   | (sec1 family domain containing 1)   | 23256        | 12   |
| <b>SPP1</b>    | (secreted phosphoprotein 1)   | 6696         | 13   |
| <b>PIM1</b>    | (Pim-1 proto-oncogene, serine/threonine kinase)   | 5292         | 14   |
| <b>HTRA1</b>   | (HtrA serine peptidase 1)   | 5654         | 15   |
| <b>SPON1</b>   | (spondin 1, extracellular matrix protein)   | 10418        | 16   |
| <b>CRYAB</b>   | (crystallin, alpha B)   | 1410         | 17   |
| <b>FABP4</b>   | (fatty acid binding protein 4, adipocyte)   | 2167         | 18   |
| <b>CLDN11</b>  | (claudin 11)  | 5010         | 19   |
| <b>LAMC2</b>   | (laminin, gamma 2)  | 3918         | 20   |

**Supplementary Table 5A. MVA-1 (GSE17536) - CSS**

| <b>Parameters</b>    | <b>Hazard Ratio</b> | <b>95%CI</b> | <b>P*</b> |
|----------------------|---------------------|--------------|-----------|
| TNM Stage**          | 5.648               | 3.48-9.17    | <0.001    |
| ULBP2 (3-8 vs. 1-2)  | 2.682               | 1.41-5.09    | 0.003     |
| SEMA5A (4-8 vs. 1-3) | 0.459               | 0.24-0.88    | 0.019     |
| PCDH7 (4-8 vs. 1-3)  | 2.765               | 1.20-6.35    | 0.017     |

**Supplementary Table 5B. MVA-1 (GSE17537) - CSS**

| <b>Parameters</b>    | <b>Hazard Ratio</b> | <b>95%CI</b> | <b>P*</b> |
|----------------------|---------------------|--------------|-----------|
| TNM Stage**          | 13.829              | 2.78-68.87   | 0.001     |
| ULBP2 (3-8 vs. 1-2)  | 5.808               | 1.46-23.09   | 0.012     |
| SEMA5A (4-8 vs. 1-3) | 0.257               | 0.07-0.95    | 0.042     |

\*Cox proportional hazards regression

\*\*TNM Stage: Treated as a continuous variable (1, 2, 3, 4)

**Supplementary Table 6. Univariate cox regression analysis of ULBP2 and SEMA5A gene expression - CSS**

| <b>Dataset</b>  | <b>Parameters</b> | <b>Hazard Ratio</b> | <b>95%CI</b>  | <b>P*</b> |
|-----------------|-------------------|---------------------|---------------|-----------|
| <b>GSE17536</b> | ULBP2**           | 1.335               | 1.173 - 1.519 | <0.001    |
|                 | SEMA5A**          | 0.862               | 0.747 - 0.996 | 0.044     |
| <b>GSE17537</b> | ULBP2**           | 1.670               | 1.119 - 2.492 | 0.012     |
|                 | SEMA5A**          | 0.719               | 0.541 - 0.956 | 0.023     |

\*Cox proportional hazards regression

\*\*Treated as a continuous variable

**Supplementary Table 7. Stage 2&3 restricted MVA of clinicopathological parameters and SU-GIB in GSE17536 - CSS**

| <b>GSE17536</b>      |                      | <b>Hazard ratio</b> | <b>95% CI</b>  | <b><i>P</i>*</b> |
|----------------------|----------------------|---------------------|----------------|------------------|
| Stage (stage 3 vs 2) |                      | 2.176               | 0.936 - 5.058  | 0.071            |
| SU-GIB               | Baseline             | 1                   |                | 0.001            |
|                      | Intermediate vs good | 3.501               | 1.374 - 8.922  | 0.009            |
|                      | Bad vs good          | 7.176               | 2.546 - 20.225 | <0.001           |

\*Cox proportional hazards regression

**Supplementary Table 8. Cox regression analysis with GSE39582 MSS samples - DFS**

|  |                      | <b>Hazard Ratio</b> | <b>95%CI</b>    | <b>P*</b> |
|--|----------------------|---------------------|-----------------|-----------|
| KRAS or BRAF mutation (Mutant vs. Wild type)** |                      | 1.135               | 0.797 - 1.616   | 0.482     |
| Stage  | Baseline             |                     |                 | <0.001    |
|  | Stage 2 vs 1         | 6.139               | 0.844 - 44.637  | 0.073     |
|  | Stage 3 vs 1         | 10.297              | 1.423 - 74.501  | 0.021     |
|  | Stage 4 vs 1         | 36.488              | 4.973 - 267.726 | <0.001    |
| SU-GIB   | Baseline             |                     |                 | 0.016     |
|  | Intermediate vs good | 1.472               | 0.959 - 2.259   | 0.077     |
|  | Bad vs good          | 1.978               | 1.241 - 3.154   | 0.004     |

\*Cox proportional hazards regression

\*\*KRAS or BRAF mutation: Samples are considered as wild type if only both genes are wild type, other situations are considered mutated.

**Supplementary Table 9. MVA analyses for stage 2 & 3 patients**

| <b>GSE17536 - DFS</b>              | <b>Hazard ratio</b> | <b>95% CI</b> | <b>P*</b> |
|------------------------------------|---------------------|---------------|-----------|
| Age (>65 vs ≤65)                   | 0.422               | 0.187 - 0.956 | 0.039     |
| Gender (male vs female)            | 1.005               | 0.688 - 1.467 | 0.98      |
| Stage (stage 3 vs. 2)              | 1.732               | 0.802 - 3.742 | 0.162     |
| Grade***                           | 0.91                | 0.424 - 1.954 | 0.809     |
| MSI status (Instable vs. Stable)** | 2.157               | 0.856 - 5.437 | 0.13      |
| Oncotype (High vs. Low risk)****   | 2.471               | 1.179 - 5.181 | 0.017     |
|                                    |                     |               |           |
| <b>GSE17536 - DFS</b>              | <b>Hazard ratio</b> | <b>95% CI</b> | <b>P*</b> |
| Age (>65 vs ≤65)                   | 0.459               | 0.209 - 1.008 | 0.052     |
| Gender (male vs female)            | 0.988               | 0.681 - 1.435 | 0.952     |
| Stage (stage 3 vs 2)               | 1.872               | 0.864 - 4.057 | 0.112     |
| Grade***                           | 0.925               | 0.468 - 1.830 | 0.824     |
| MSI status (Instable vs stable)**  | 1.135               | 0.415 - 3.107 | 0.805     |
| SU-GIB*****                        | 2.004               | 1.158 - 3.469 | 0.013     |

| <b>GSE39582 - DFS</b>           | <b>Hazard ratio</b>  | <b>95% CI</b> | <b>P*</b>     |        |
|---------------------------------|----------------------|---------------|---------------|--------|
| Age (>65 vs ≤65)                | 0.921                | 0.647 - 1.312 | 0.65          |        |
| Gender (female vs male)         | 0.795                | 0.557 - 1.135 | 0.206         |        |
| Stage (stage 3 vs 2)            | 1.734                | 1.214 - 2.475 | 0.002         |        |
| MSI status (Instable vs Stable) | 2.008                | 1.049 - 3.842 | 0.035         |        |
| Oncotype (High vs low risk)**** | 1.207                | 1.012 - 1.439 | 0.037         |        |
|                                 |                      |               |               |        |
| <b>GSE39582 - DFS</b>           | <b>Hazard ratio</b>  | <b>95% CI</b> | <b>P*</b>     |        |
| Age (>65 vs ≤65)                | 1.016                | 0.712 - 1.450 | 0.929         |        |
| Gender (female vs male)         | 0.708                | 0.488 - 1.026 | 0.068         |        |
| Stage (stage 3 vs 2)            | 1.806                | 1.263 - 2.583 | 0.001         |        |
| MSI status (Instable vs stable) | 0.367                | 0.186 - 0.726 | 0.004         |        |
| SU-GIB                          | Baseline             |               | 0.002         |        |
|                                 | Intermediate vs good | 1.731         | 1.104 - 2.713 | 0.017  |
|                                 | Bad vs good          | 2.47          | 1.494 - 4.083 | <0.001 |

\*Cox proportional hazards regression

\*\*MSI status: Determined in silico (see Methods)

\*\*\*Grade: Treated as a continuous variable; poorly differentiated (1), moderately differentiated (2), well differentiated (3); not available for GSE39582.

\*\*\*\*Oncotype: Determined in silico (see Methods)

\*\*\*\*\*SU-GIB: Treated as a continuous variable (1: Good, 2: Intermediate, 3: Bad)

**Supplementary Table 10. MVA analyses for stage 2 & 3 patients including both Oncotype and SU-GIB**

| <b>GSE39582 – DFS - stage 2 &amp; 3</b> |                      | <b>Hazard ratio</b> | <b>95% CI</b> | <b>P*</b> |
|---|----------------------|---------------------|---------------|-----------|
| Age (>65 vs ≤65)                        |                      | 1.022               | 0.717 - 1.458 | 0.903     |
| Gender (female vs male)                 |                      | 0.72                | 0.497 - 1.043 | 0.082     |
| Stage (stage 3 vs 2)                    |                      | 1.73                | 1.203 - 2.488 | 0.003     |
| MSI status (Instable vs stable)         |                      | 0.37                | 0.187 - 0.732 | 0.004     |
| Oncotype (High vs low risk)**           |                      | 1.131               | 0.944 - 1.356 | 0.183     |
| SU-GIB                                  | Baseline             |                     |               | 0.006     |
|   | Intermediate vs good | 1.665               | 1.059 - 2.619 | 0.027     |
|   | Bad vs good          | 2.296               | 1.374 - 3.837 | 0.002     |

\*Cox proportional hazards regression

\*\*Oncotype: Determined in silico (see Methods)

**Supplementary Table 11. Drugs that are significantly correlated with both SEMA5A and ULBP2 expression in colon cancer cell lines according to CGP**

| Drugs               | Target                               | Ranksum | SEMA5A      |        |      | ULBP2       |       |      |
|---------------------|--------------------------------------|---------|-------------|--------|------|-------------|-------|------|
|                     |                                      |         | Pearson's r | p      | rank | Pearson's r | p     | rank |
| <b>NVP-BEZ235</b>   | PI3K/MTOR                            | 5       | 0.60        | <0.001 | 1    | -0.47       | 0.006 | 4    |
| <b>Bortezomib</b>   | PARP                                 | 8       | 0.43        | 0.013  | 6    | -0.51       | 0.002 | 2    |
| <b>AZ628</b>        | Farnesyltransferase                  | 8       | 0.47        | 0.005  | 3    | -0.47       | 0.006 | 5    |
| <b>Sorafenib</b>    | PDGFRA, PDGFRB, KDR, KIT, FLT3       | 12      | 0.38        | 0.029  | 9    | -0.52       | 0.002 | 3    |
| <b>Bleomycin</b>    | DNA Damage                           | 12      | 0.46        | 0.007  | 5    | -0.43       | 0.013 | 7    |
| <b>Etoposide</b>    | Topoisomerase 2                      | 12      | 0.49        | 0.004  | 2    | -0.37       | 0.032 | 10   |
| <b>AG-014699</b>    | PARP1/2                              | 13      | 0.37        | 0.034  | 12   | -0.54       | 0.001 | 1    |
| <b>BI-D1870</b>     | HSP90                                | 16      | 0.47        | 0.006  | 4    | -0.35       | 0.048 | 12   |
| <b>Thapsigargin</b> | Ca <sup>++</sup> transporting ATPase | 19      | 0.35        | 0.043  | 13   | -0.47       | 0.006 | 6    |
| <b>AZD7762</b>      | Chk 1/2                              | 19      | 0.38        | 0.031  | 11   | -0.39       | 0.023 | 8    |
| <b>CEP-701</b>      | FLT3, JAK2, NTRK1, RET               | 19      | 0.38        | 0.029  | 10   | -0.38       | 0.029 | 9    |
| <b>17-AAG</b>       | HSP90                                | 19      | 0.38        | 0.028  | 8    | -0.37       | 0.033 | 11   |
| <b>Tipifarnib</b>   | Farnesyl-transferase                 | 20      | 0.41        | 0.017  | 7    | -0.35       | 0.049 | 13   |



Figure S1

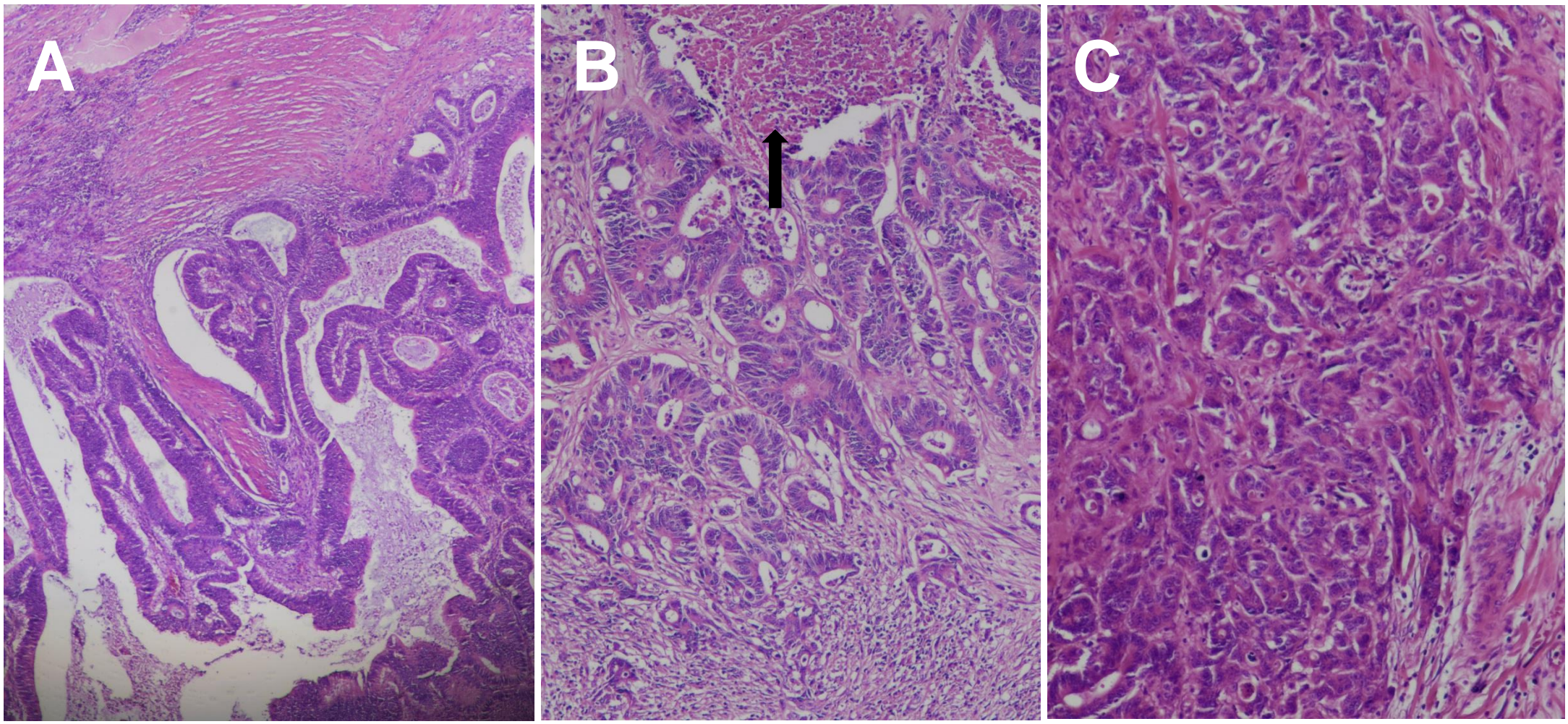


Figure S2

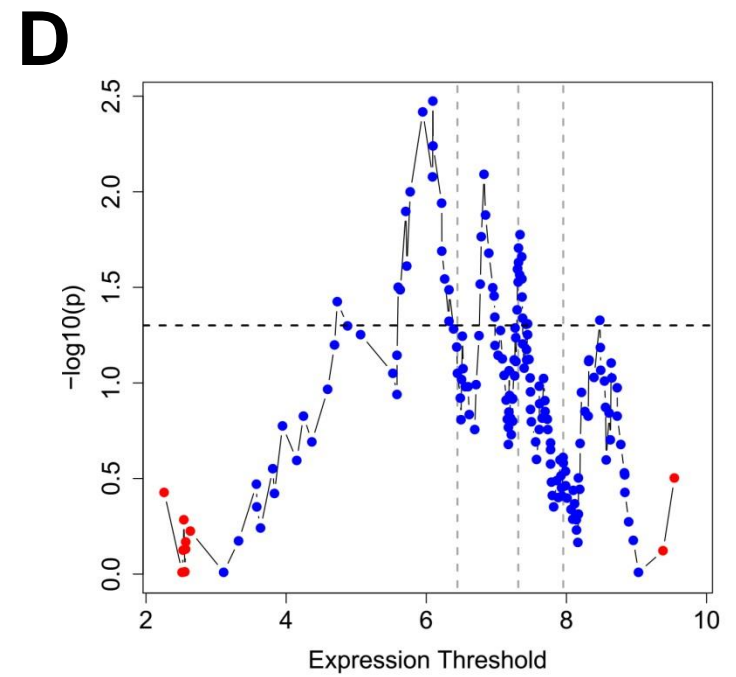
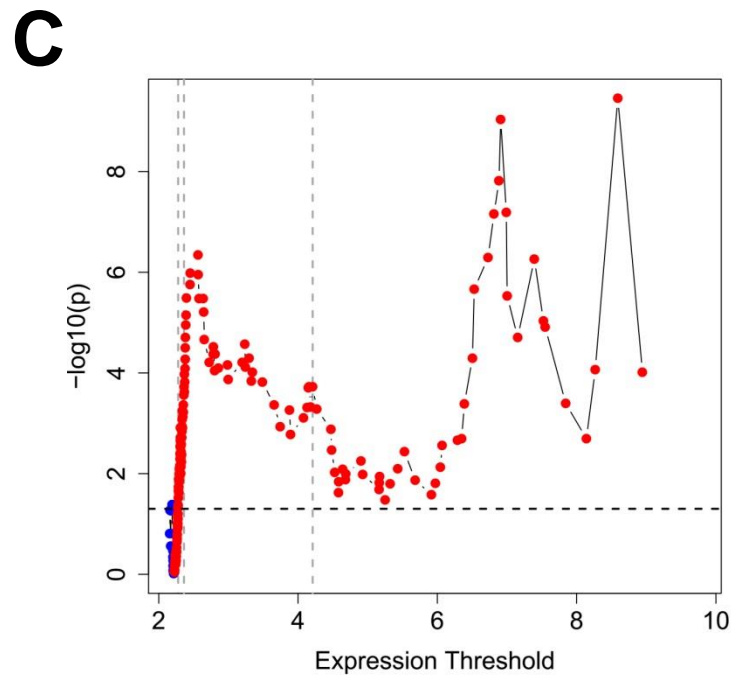
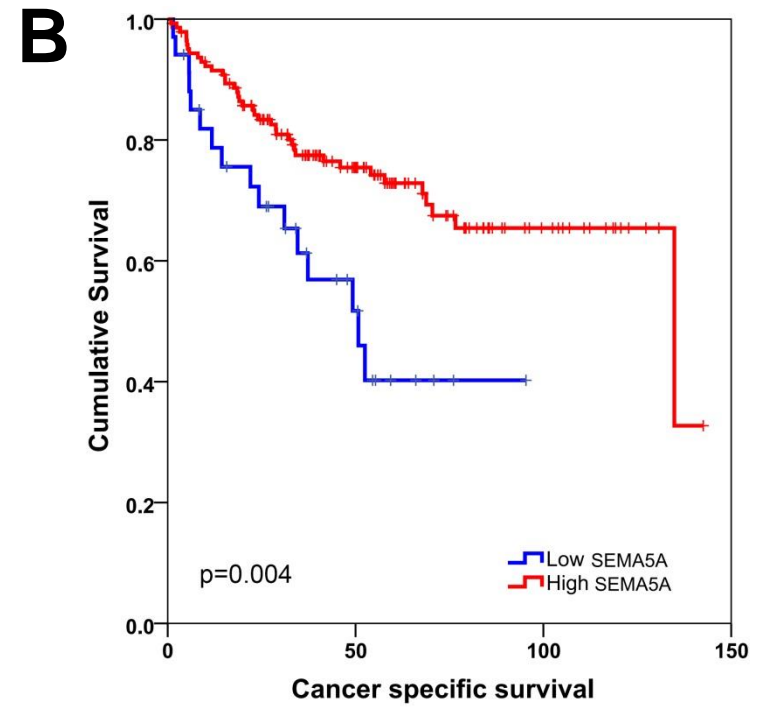
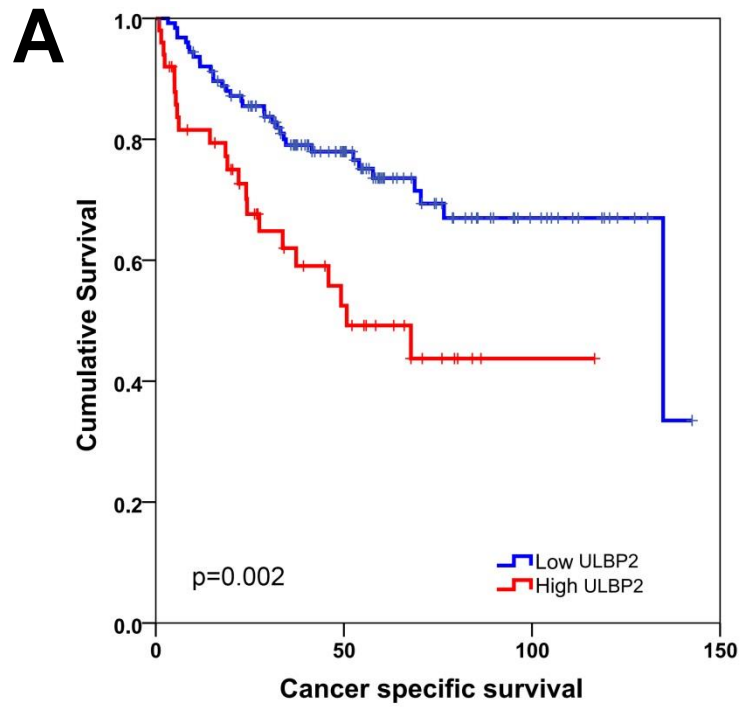


Figure S3

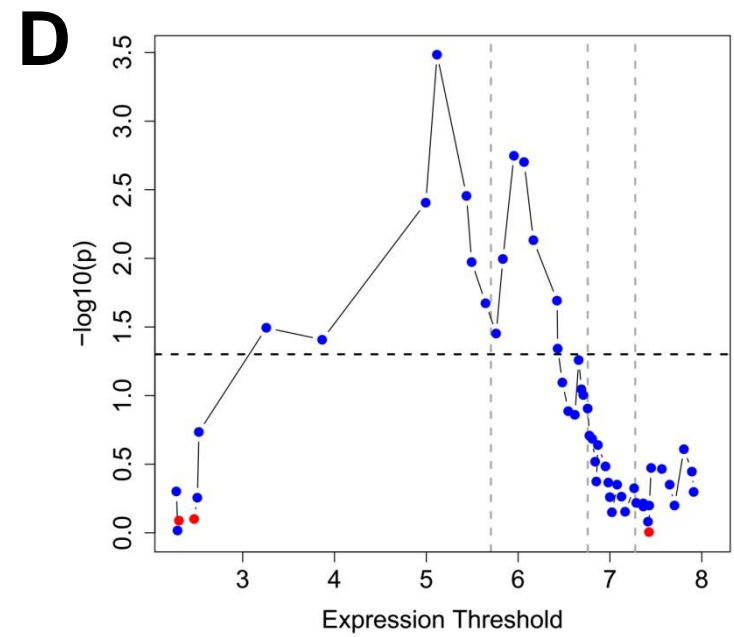
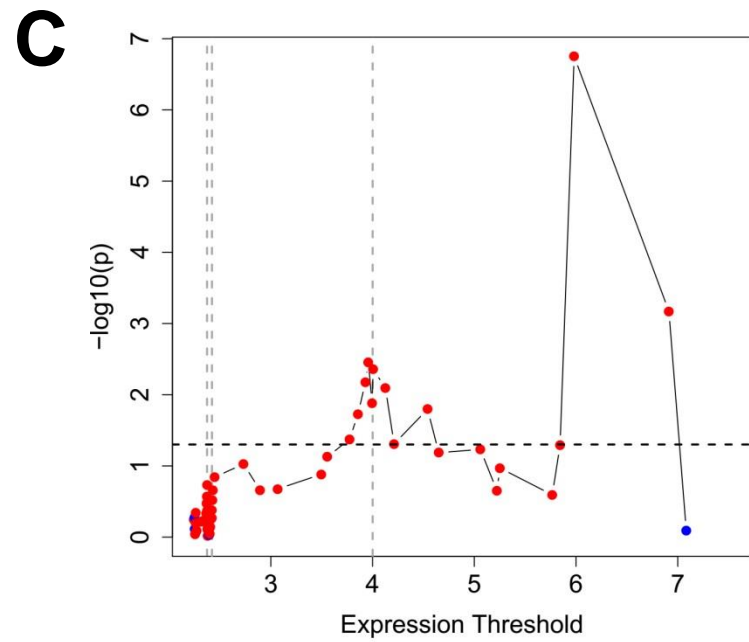
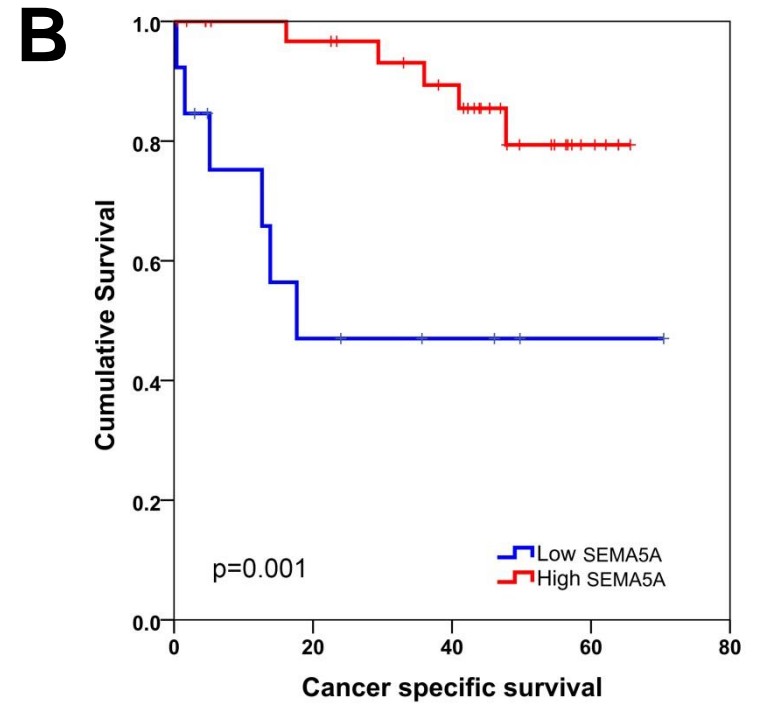
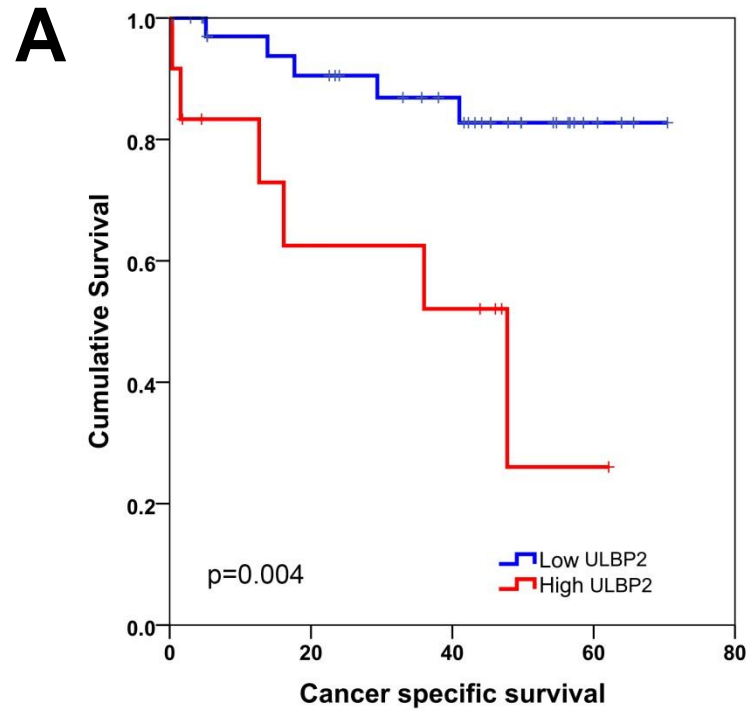
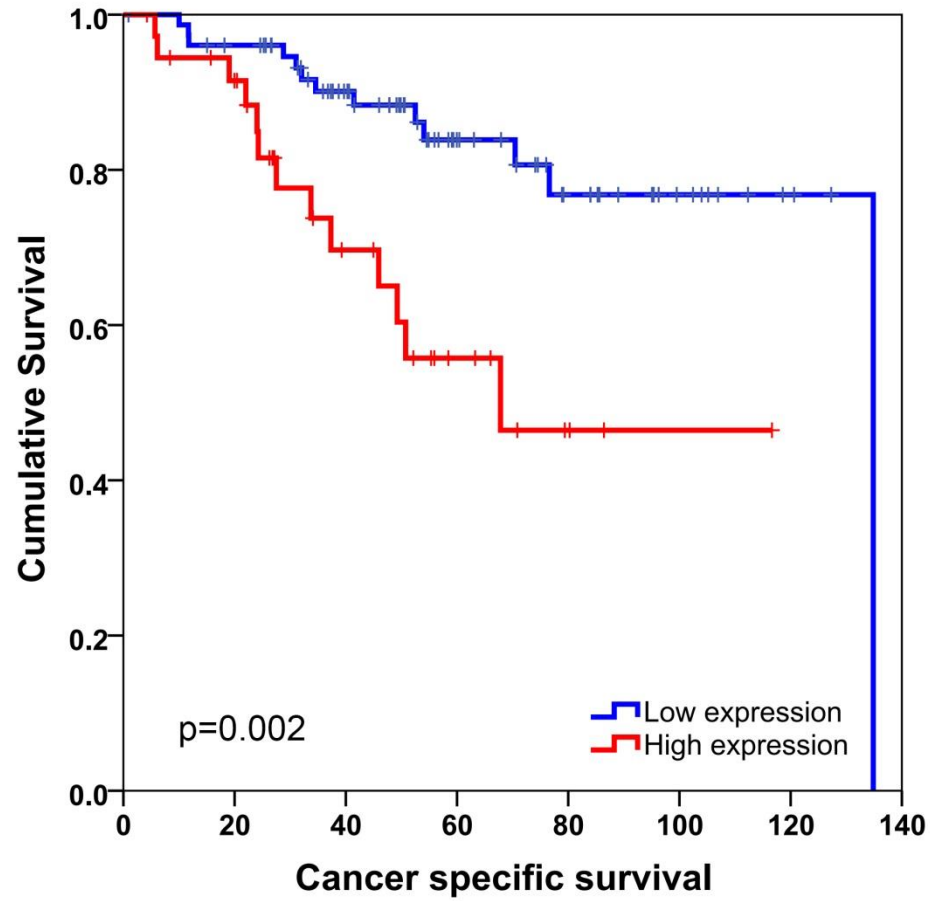


Figure S4

## ULBP2



## SEMA5A

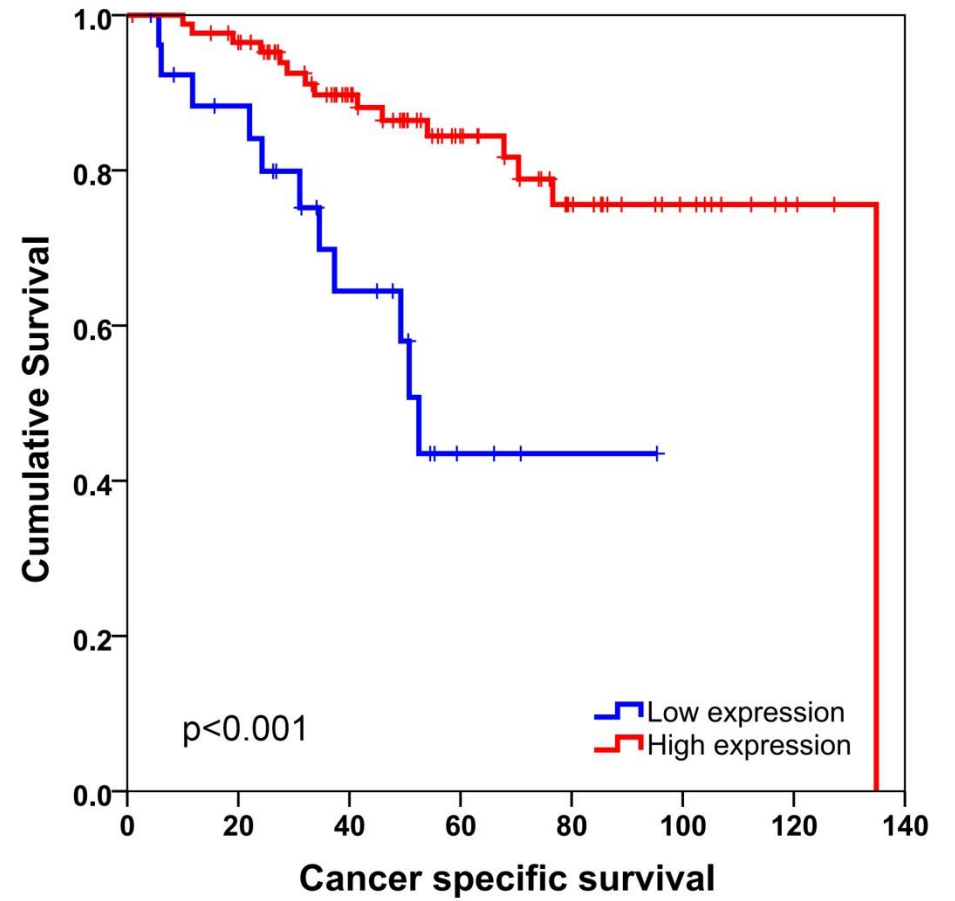


Figure S5

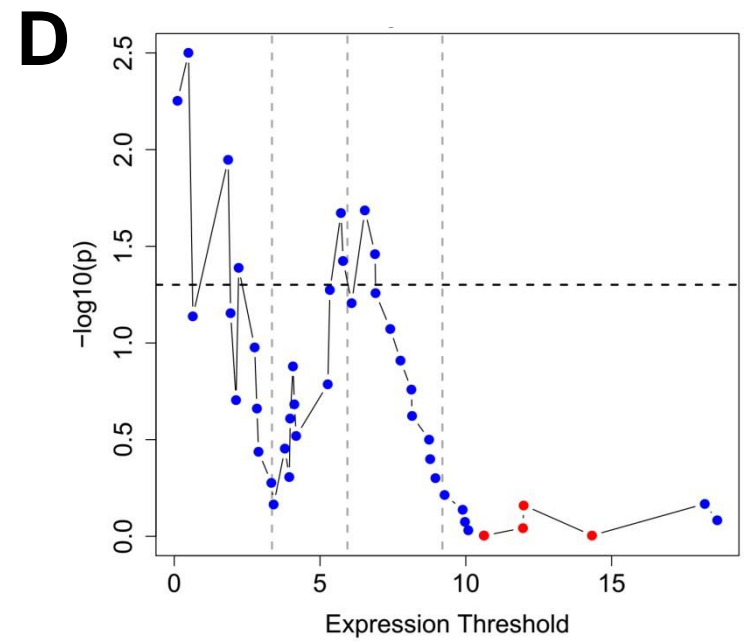
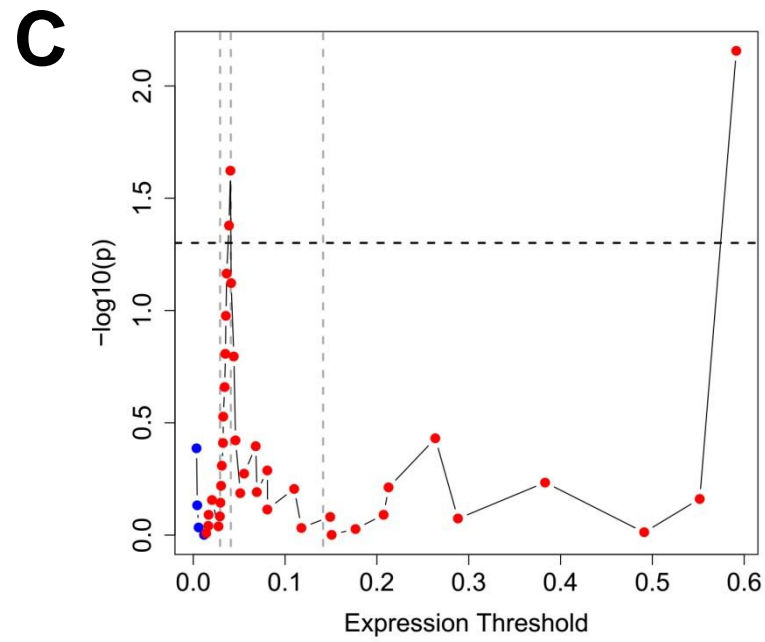
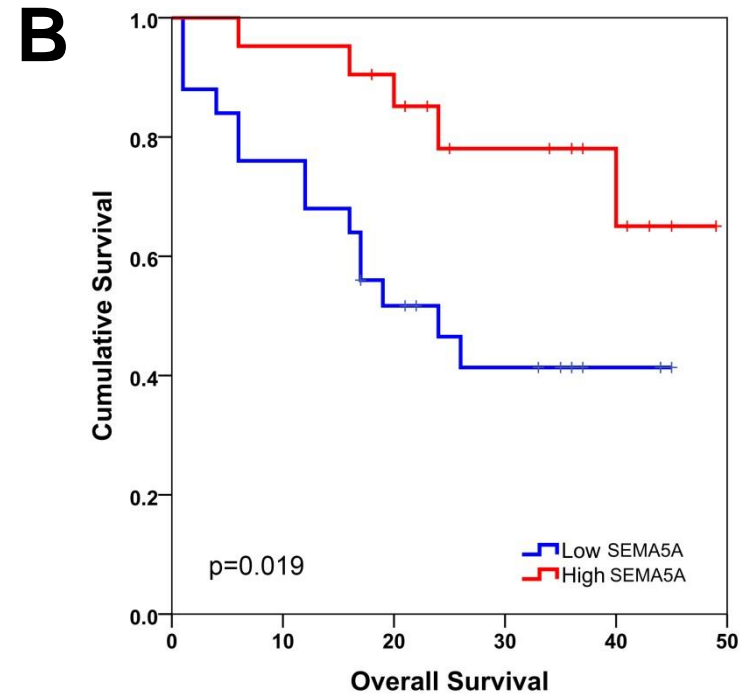
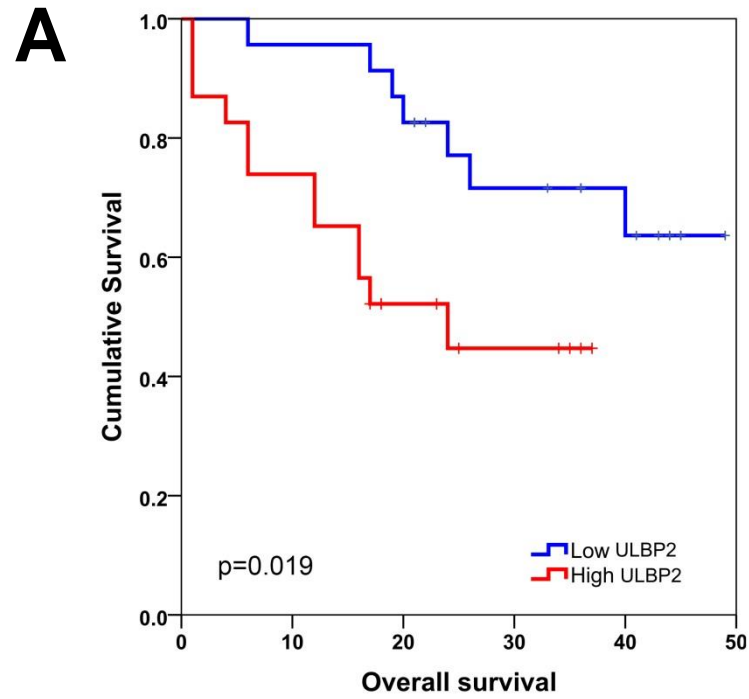


Figure S6

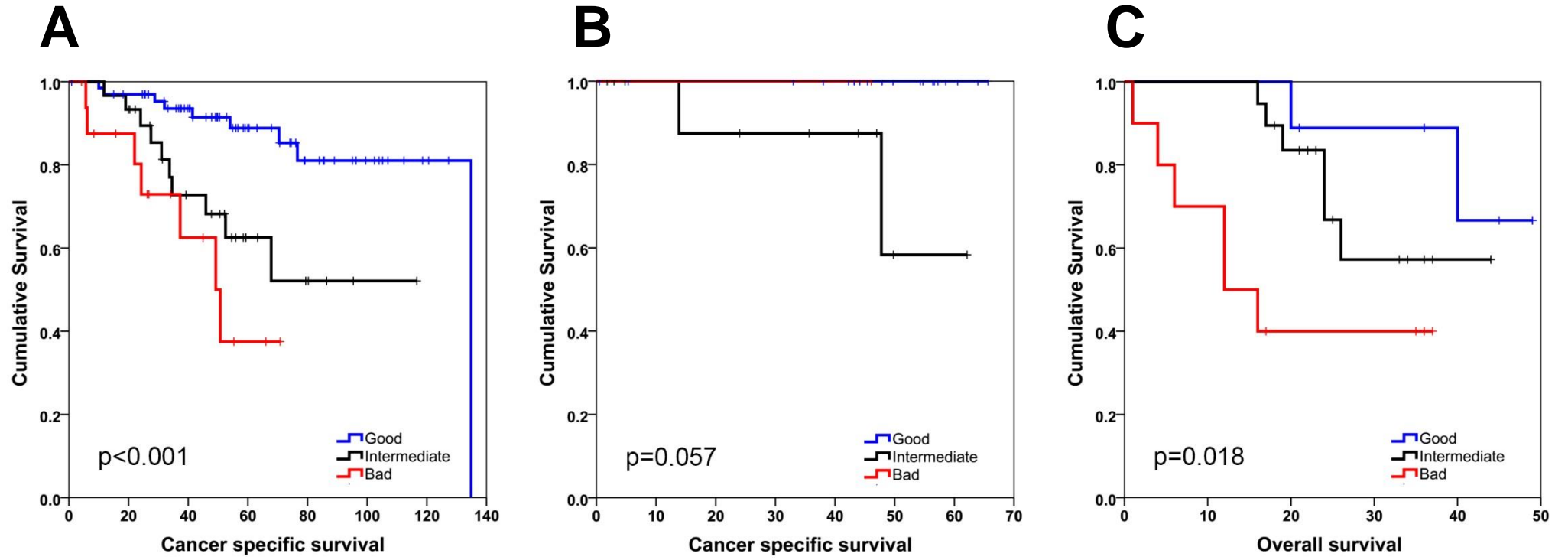
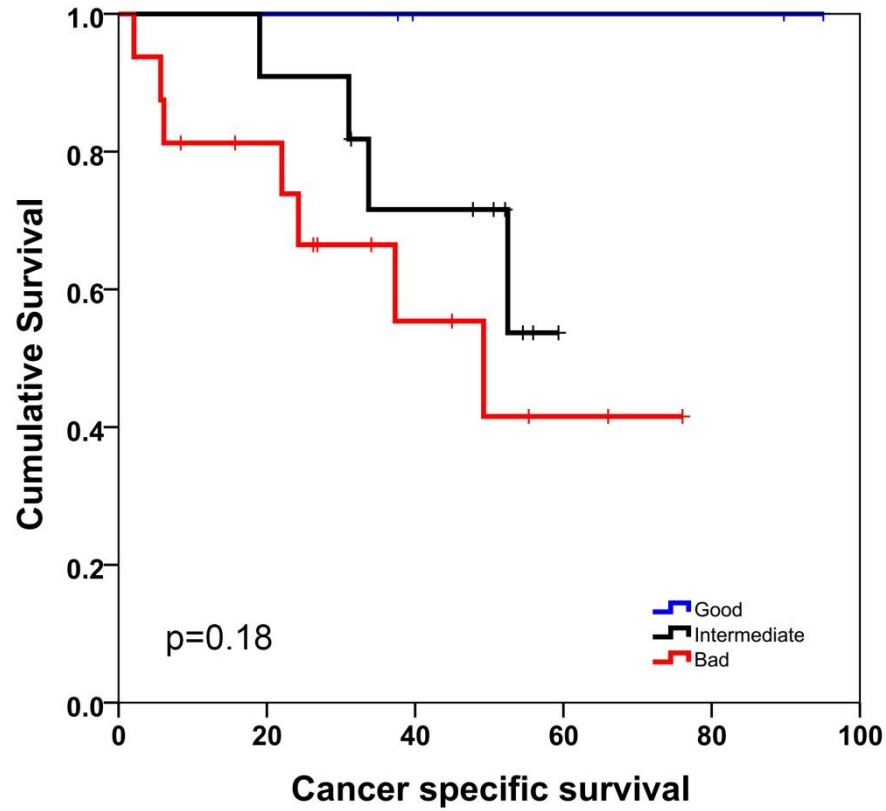


Figure S7

### MSI\*



### MSS

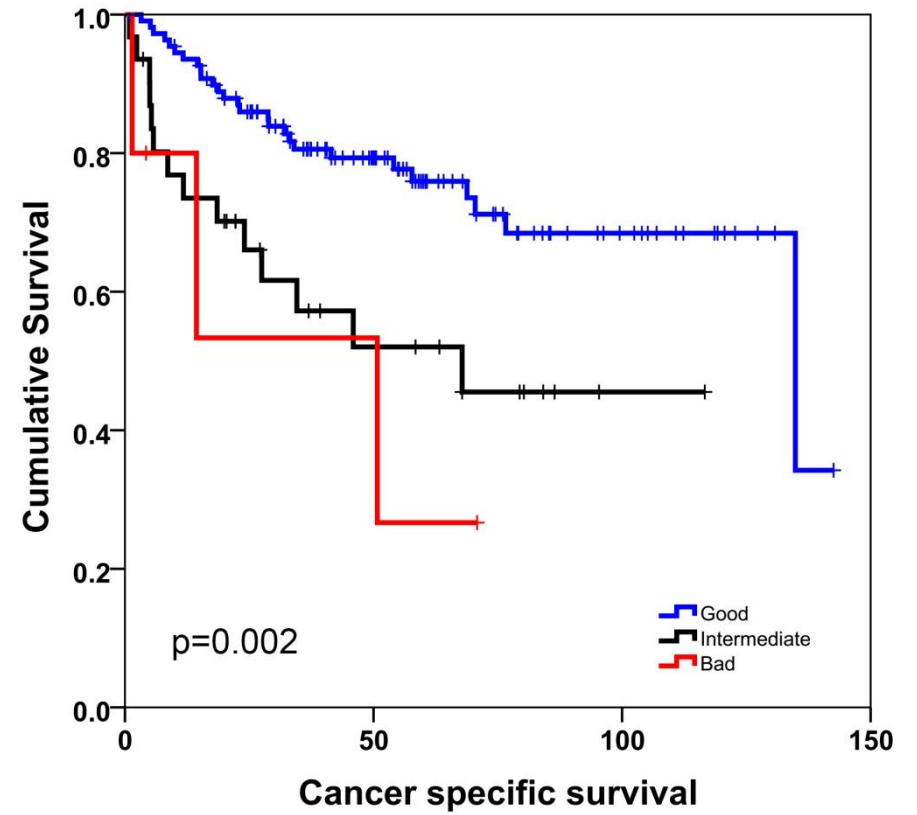


Figure S8

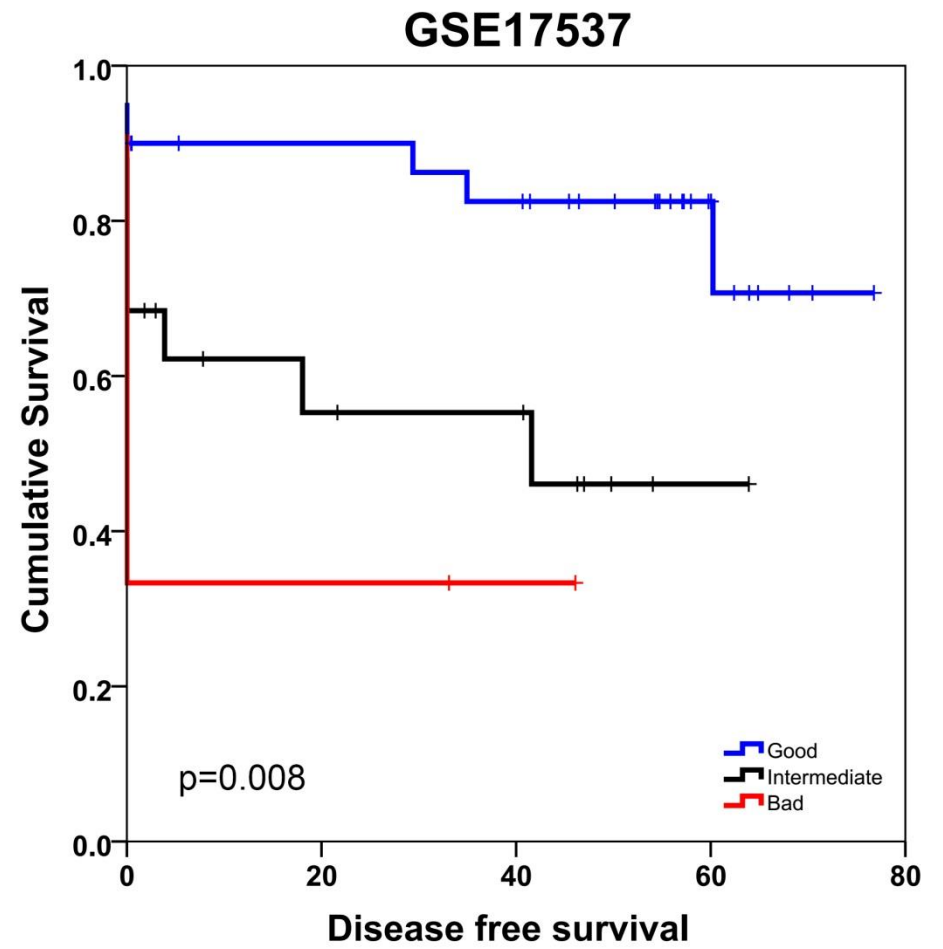
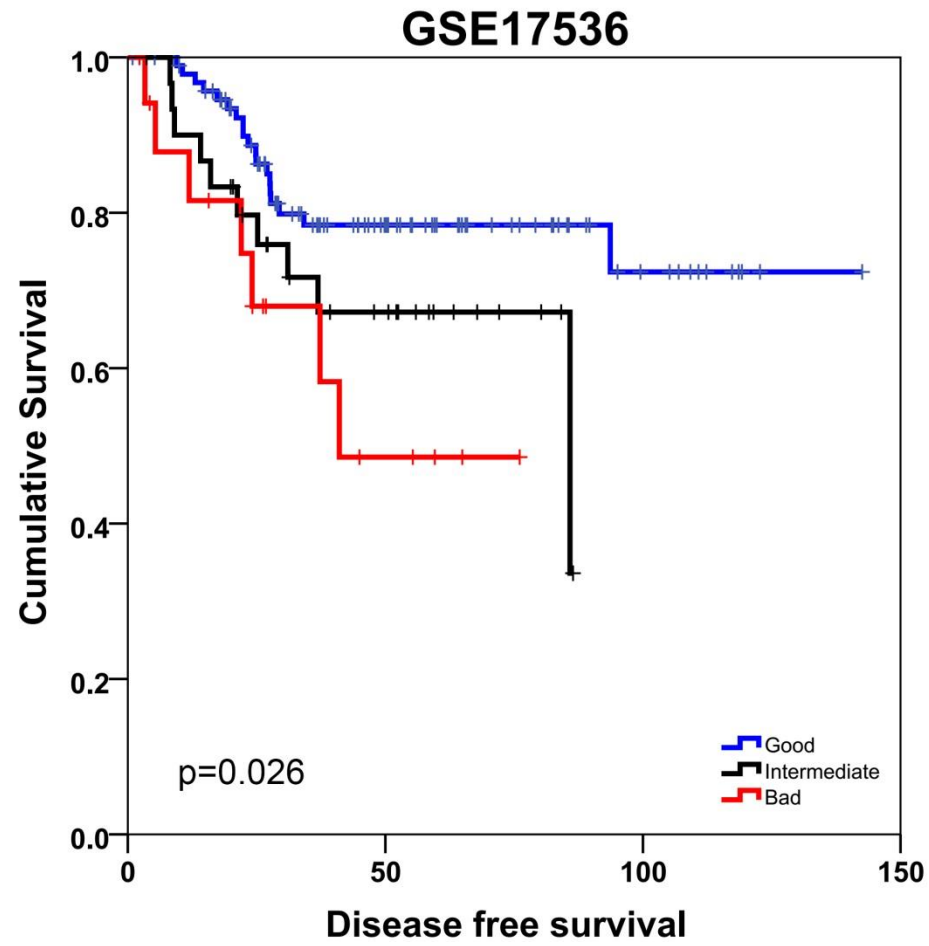




Figure S9

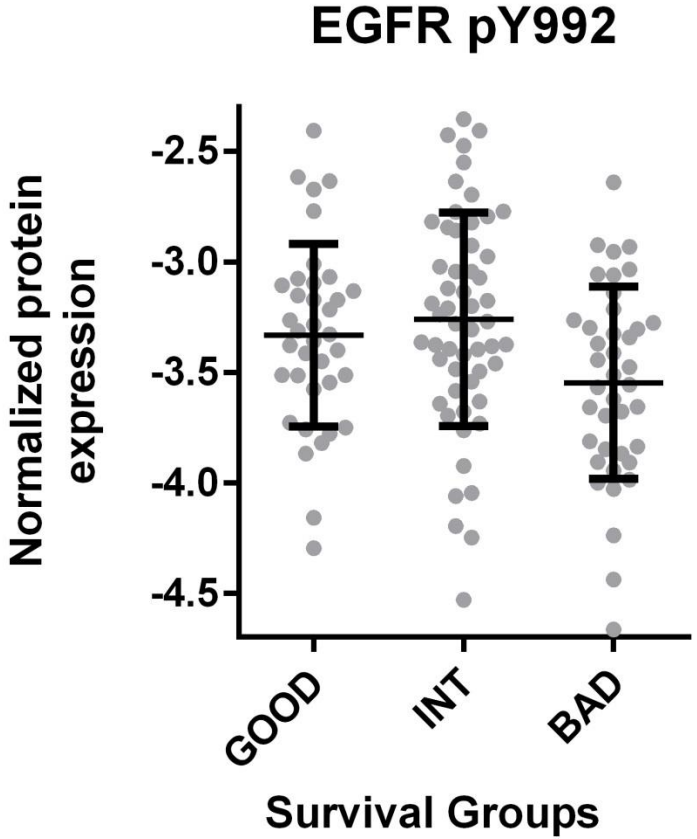
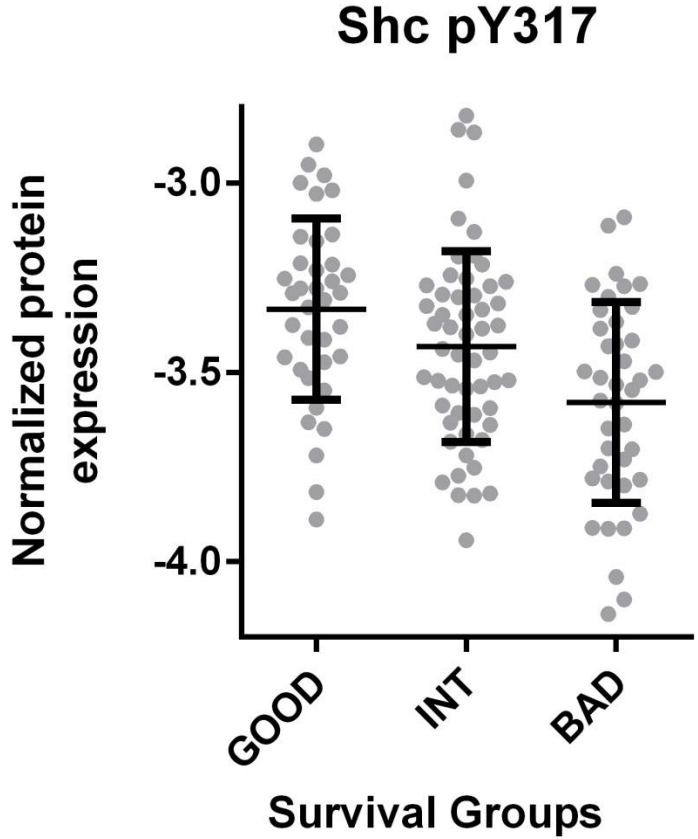


Figure S10

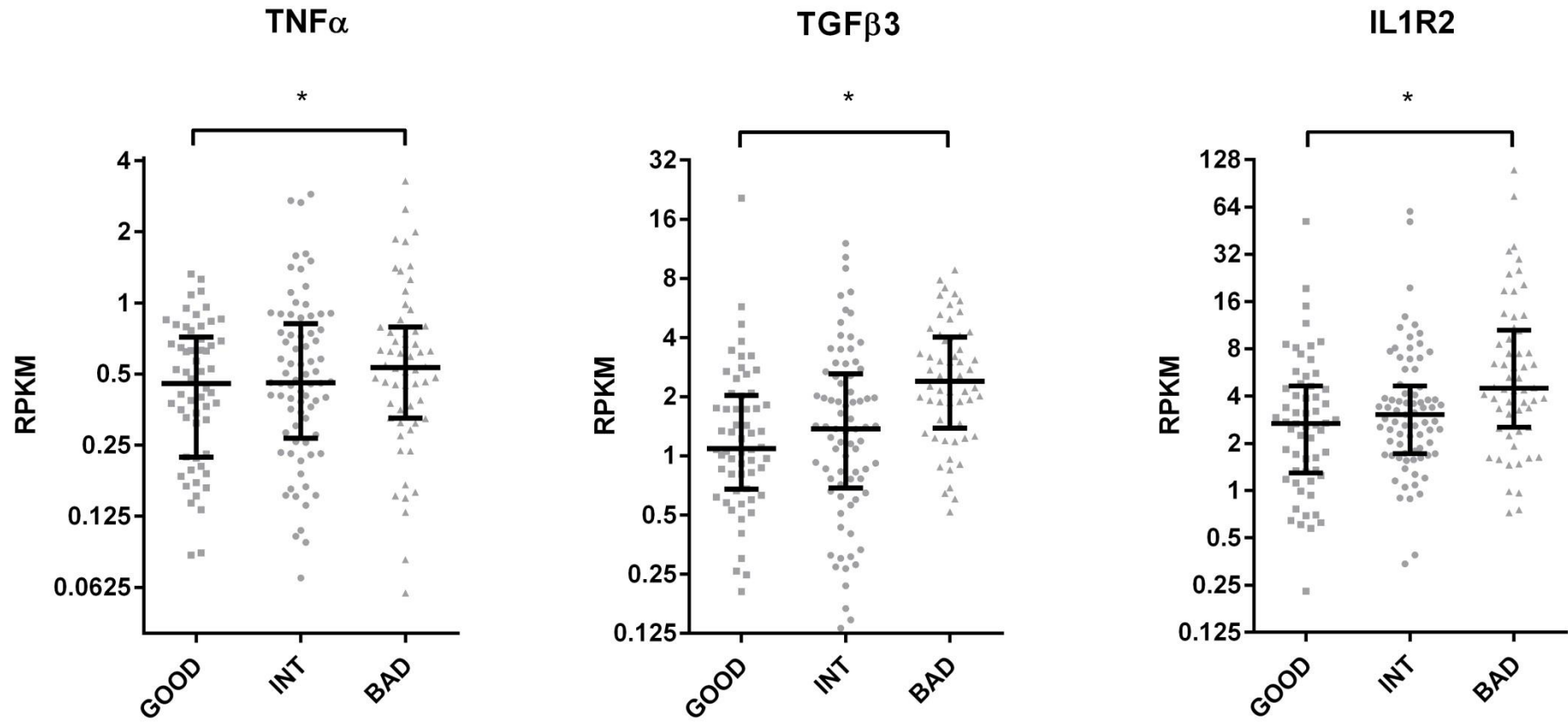


Figure S11

