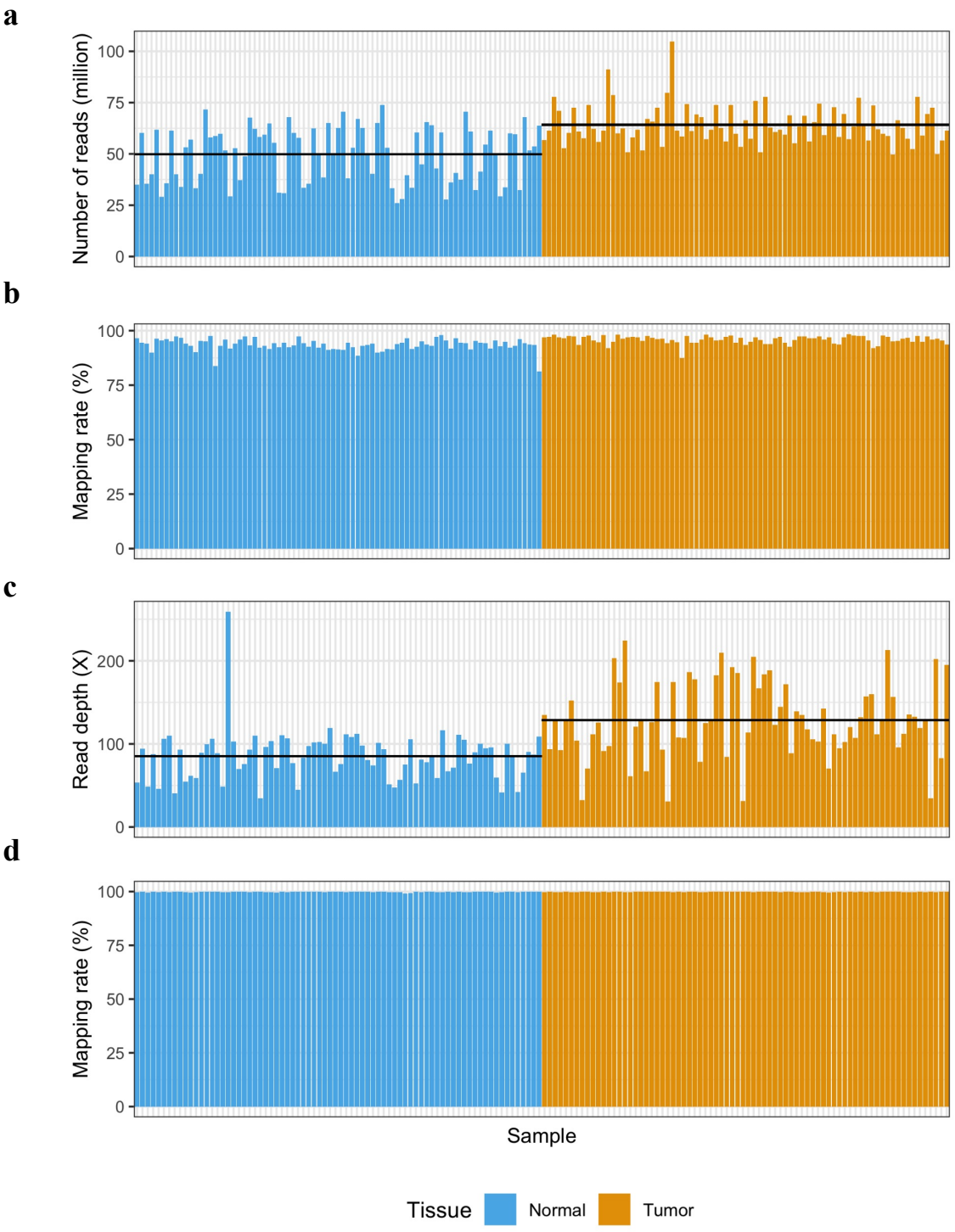
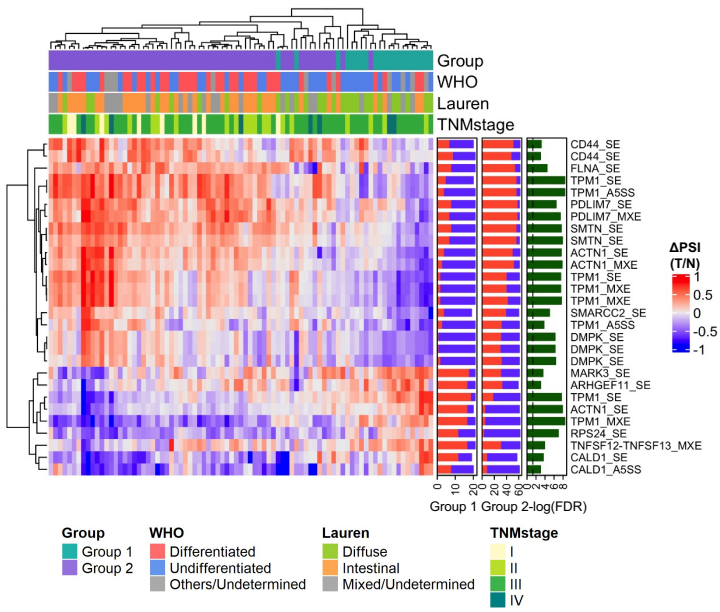


Supplementary Figure 1. Quality control of sequencing data. Number of sequencing reads (a) and mapping rates (b) of RNA-Seq data for normal and tumor samples. Sequencing read depth (c) and mapping rates (d) of exome sequencing data.

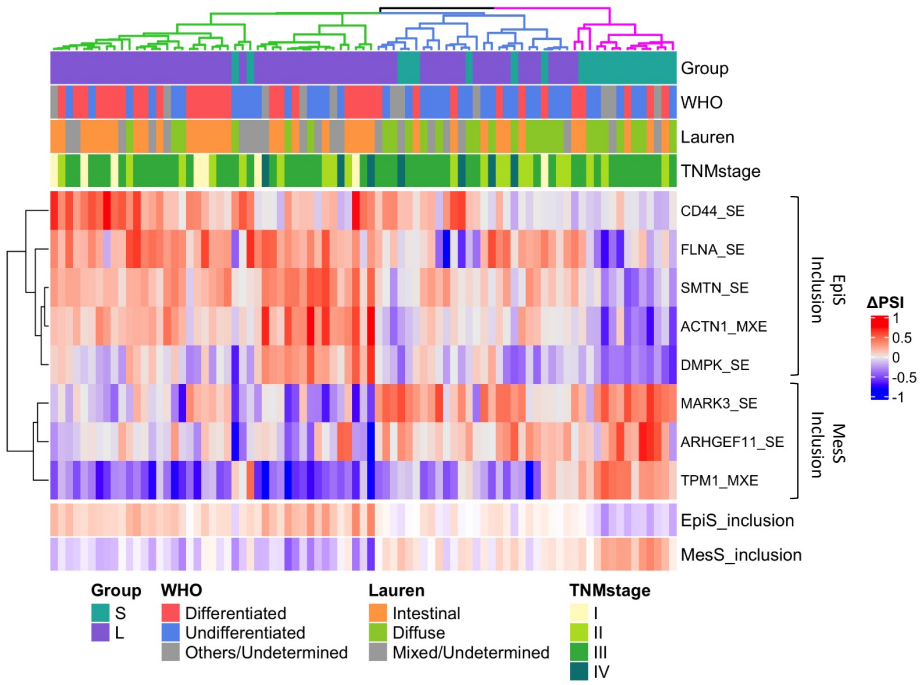
a



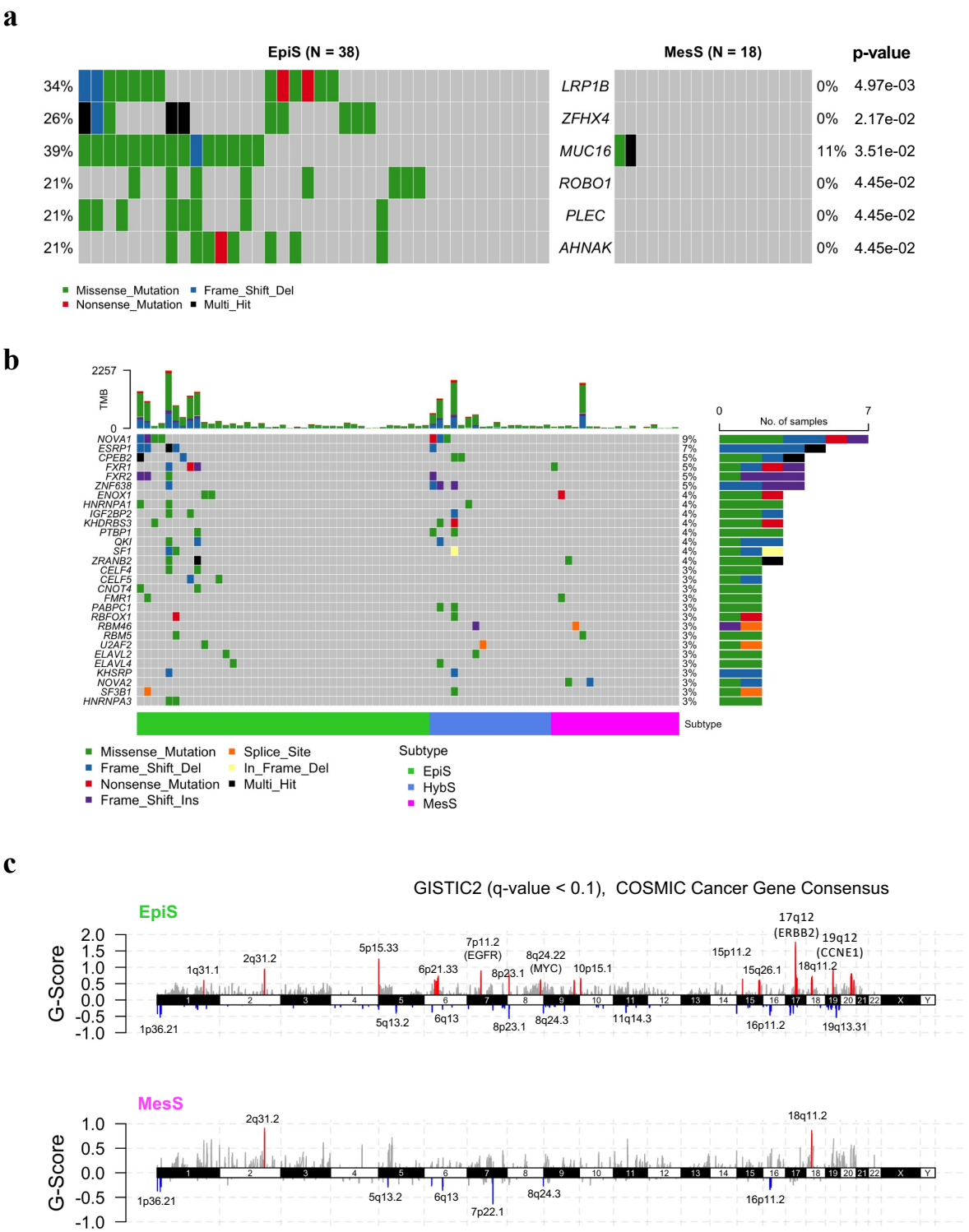
b

No	Gene	# events
1	*ACTN1*	3
2	*ARHGEF11*	1
3	*CALD1*	2
4	*CD44*	2
5	*DMPK*	3
6	*FLNA*	1
7	*MARK3*	1
8	*PDLIM7*	2
9	*RPS24*	1
10	*SMARCC2*	1
11	*SMTN*	2
12	*TNFSF12-TNFSF13*	1
13	*TPM1*	8
		28

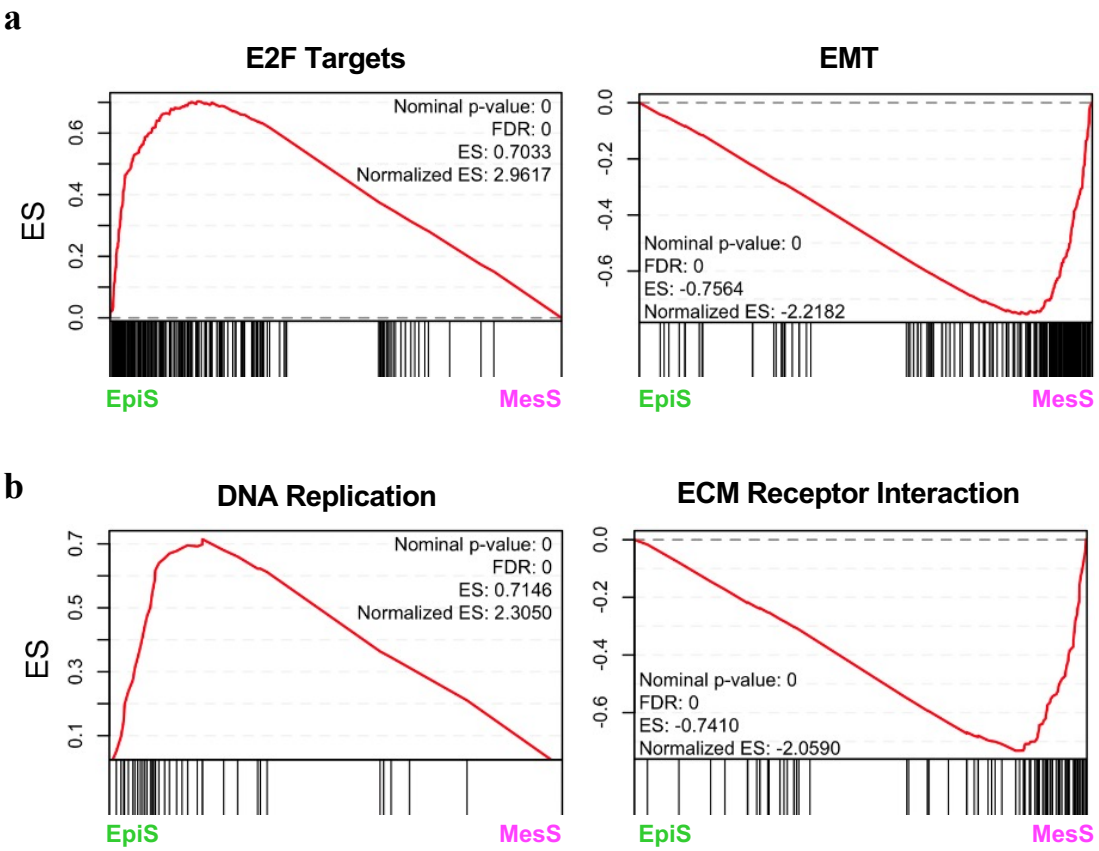
Supplementary Figure 2. Genome-wide survey of variable AS events. (**a**) Hierarchical clustering of gastric cancer patients using 28 variable AS events, based on ΔPSI values between Group 1 and Group 2, shown for each patient. AS events are named using gene name and event type (SE- skipped exon; A5/3SS- alternative 5'/3' splice site; MXE- mutually exclusive exons; RI- retained intron). WHO and Lauren classification, as well as tumor stage are shown. Relative ratio of patients with positive and negative ΔPSI values group 1 and 2 with the FDR values is shown as bar plots on the right. (**b**) Number of variable AS events detected for each spliced gene.



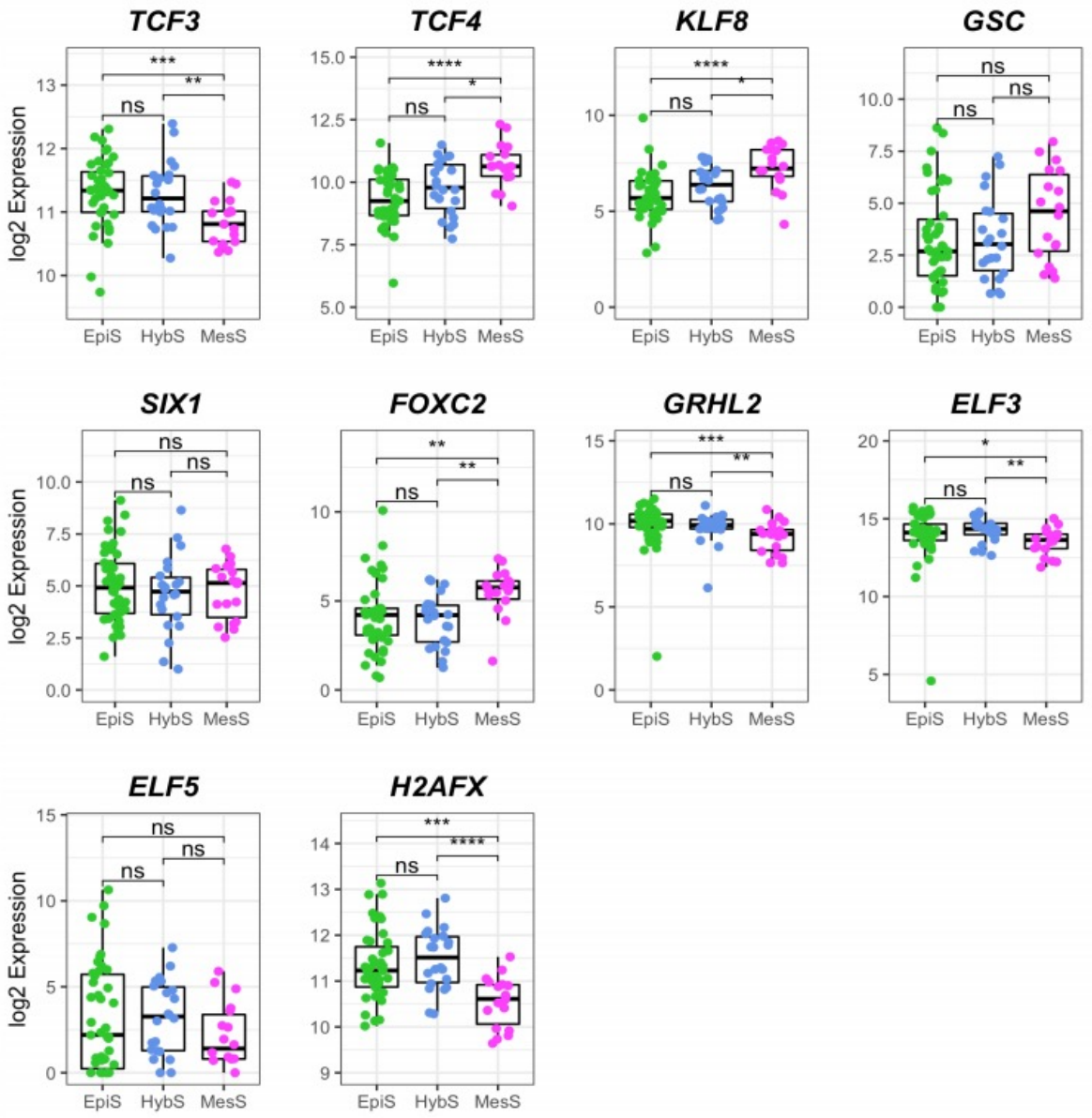
Supplementary Figure 3. Hierarchical clustering of patients and 8 variable AS events. Heatmap based on the ΔPSI values.



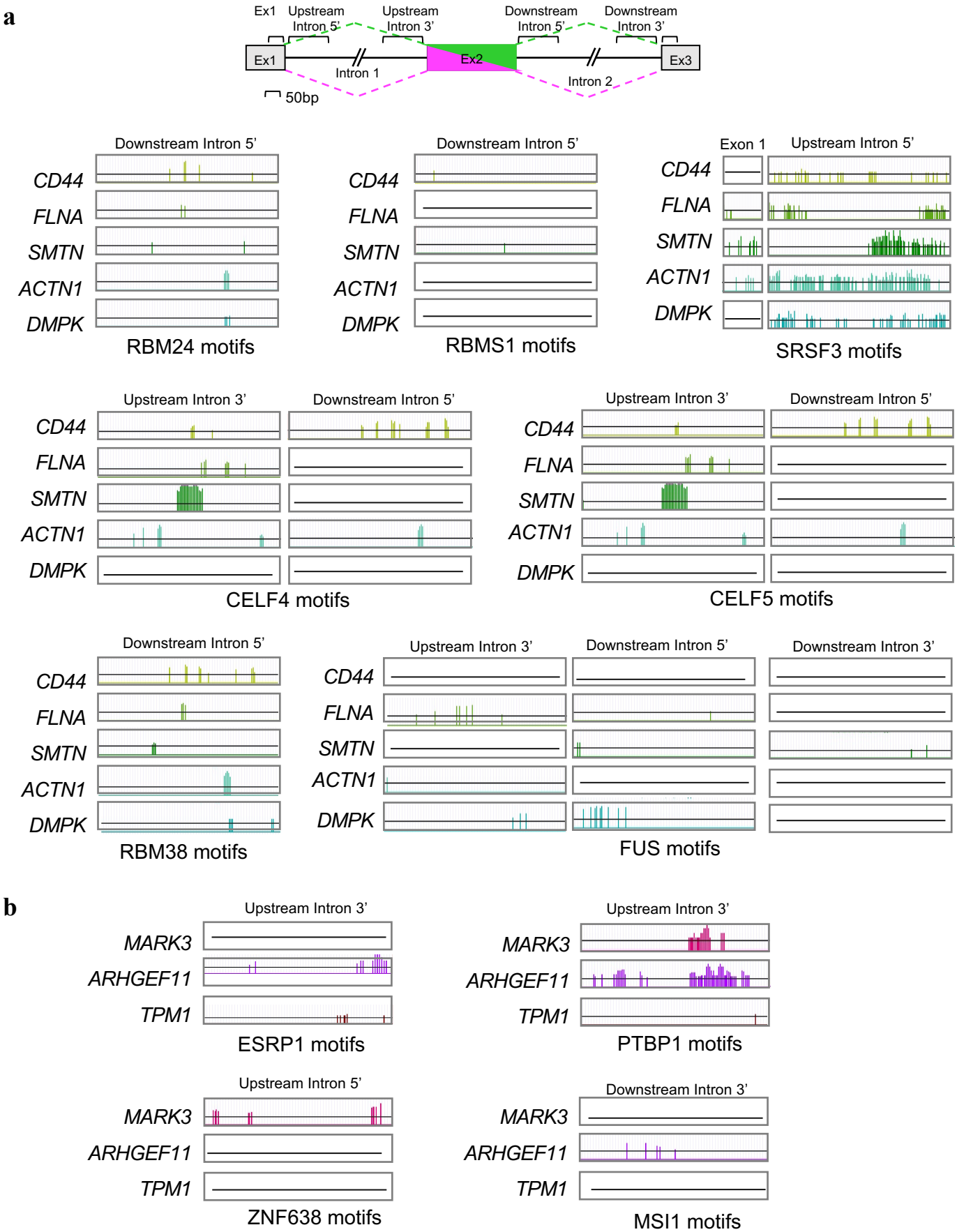
Supplementary Figure 4. Comparison of somatic mutations and copy number alterations between EpiS and MesS subtypes. (**a**) Co-mutation plot of somatic mutations that were significantly different between the EpiS (n=38) and MesS (n=18) subtypes. Mutation types are color-coded according to the functional significance. (**b**) Mutation plot for splicing factors and RBP genes (29 recurrently mutated genes). **(c)** Significantly altered somatic copy number regions in in the EpiS and MesS subtypes. GISTIC2 was used to identify regions of recurrent amplifications (red) or deletions (blue).



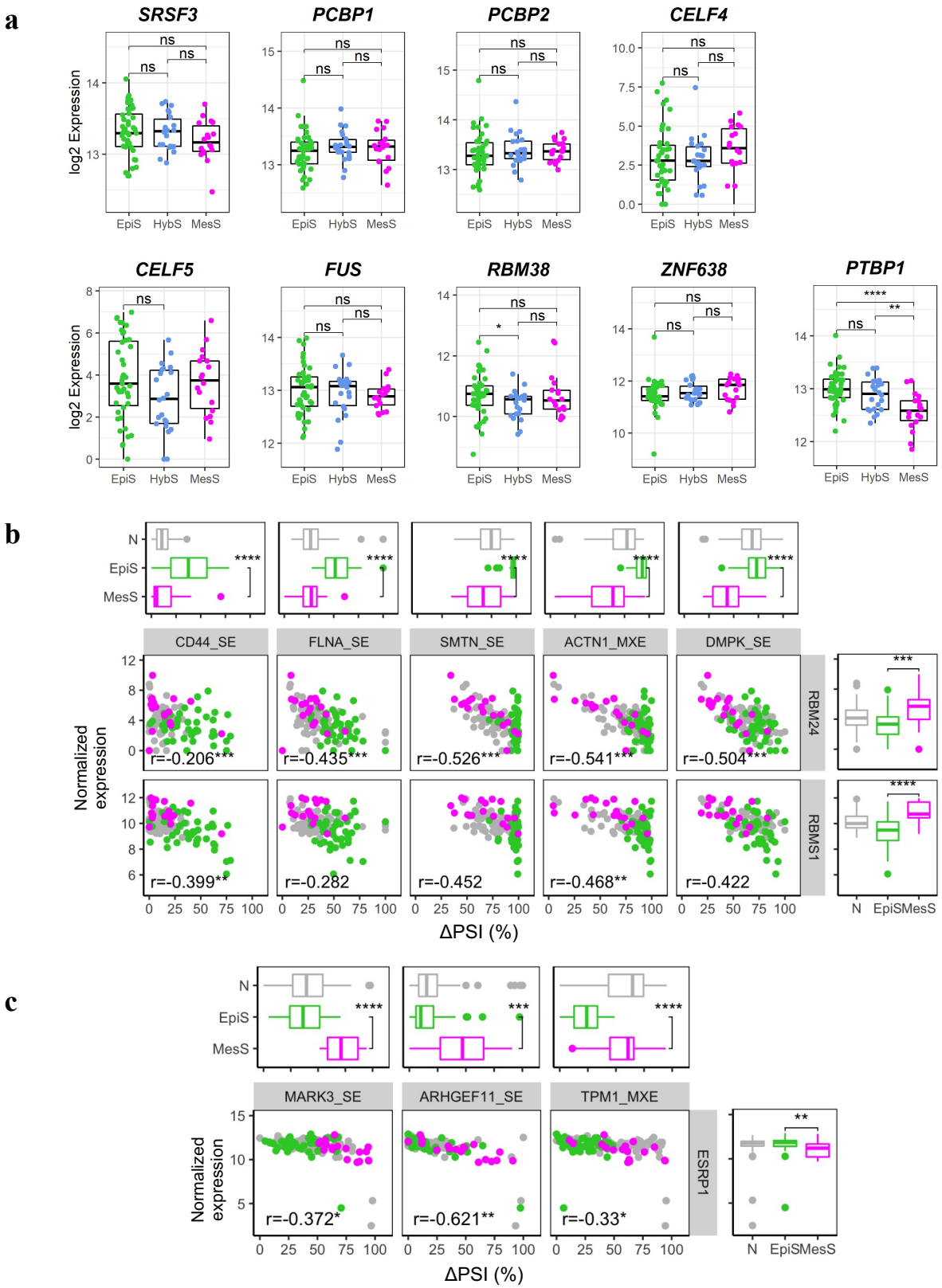
Supplementary Figure 5. Enrichment plots from Gene Set Enrichment Analysis (GSEA). (a) Statistically enriched gene sets in EpiS and MesS subtypes identified from GSEA analysis using the 50 Hallmark gene sets of MSigDB. E2F targets and EMT gene sets showed the highest enrichment score in EpiS and MesS subtypes, respectively. **(b)** Statistically enriched gene sets in EpiS and MesS subtypes identified from GSEA analysis using the KEGG pathways from MSigDB. DNA replication and ECM receptor interactions were significantly enriched in EpiS and MesS subtypes, respectively.



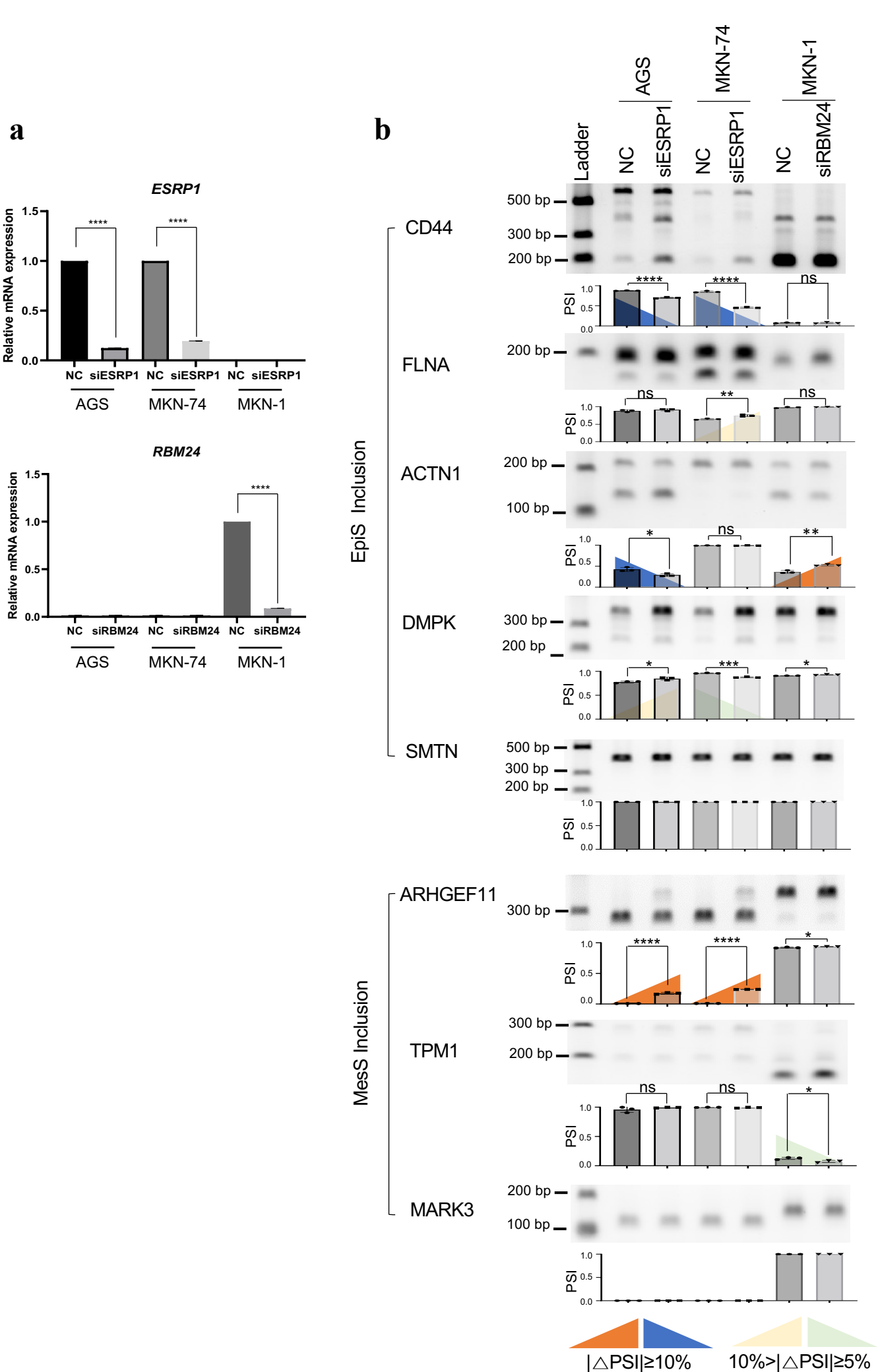
Supplementary Figure 6. Expression of EMT regulators in each AS-event based gastric tumor subtype. Gene expression of TFs *TCF3*, *TFC4*, *KLF8*, *GSC*, *SIX1*, and *FOXC2* known to be upregulated in the EMT process, or TFs *GRHL2*, *ELF3*, and *ELF5* known to promote the epithelial cell state in EpiS, HybS, and MesS tumors (Wilcoxon rank sum test, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001). All TFs behave as expected except *TCF3* (the opposite trend) and *ELF5* (no statistical significance). *H2AFX* (*H2AX*) is a marker for DNA damage (See discussion).



Supplementary Figure 7. RBP binding motifs in EpiS and MesS AS events. Positions of RBP binding motifs (See Supplementary Table 5) in the EpiS (**a**) or MesS (**b**) AS events. A schematic representation of intronic and exonic regions in which the enriched motifs are mapped is shown at the top.

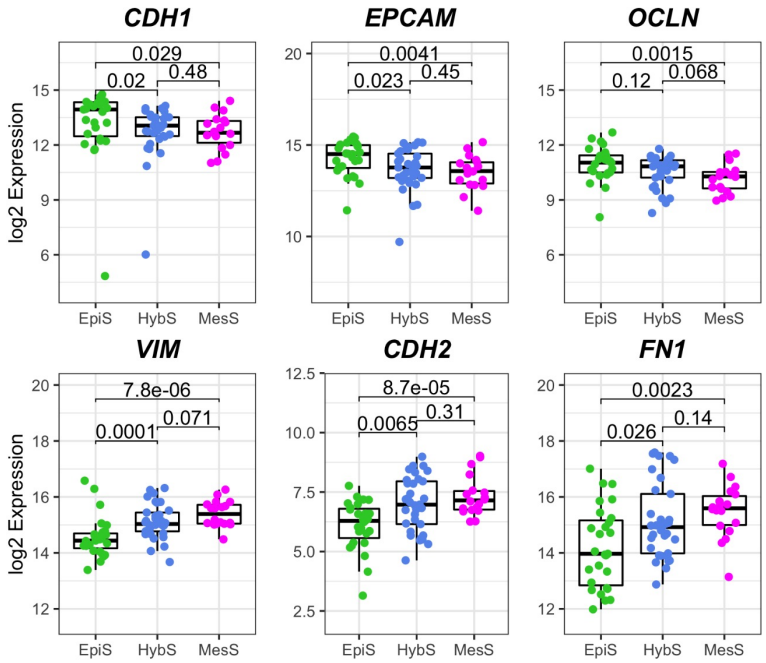


Supplementary Figure 8. Correlation between RBP expression and AS events. (a) Positions of RBP motifs in EpiS and MesS AS events. (**a**) Normalized RBP gene expression in each patient subtypes (Wilcoxon rank sum test, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001). (**b,c**) Correlation between RBP expression and splicing of AS events for each RBP-AS pair, calculated from RNA-Seq data, for EpiS (b) and MesS (c) inclusion events, colored by patient tissues. Box plots indicate the subtype difference in AS event (top) and RBP expression (right) (Wilcoxon rank sum test, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*< 0.0001).

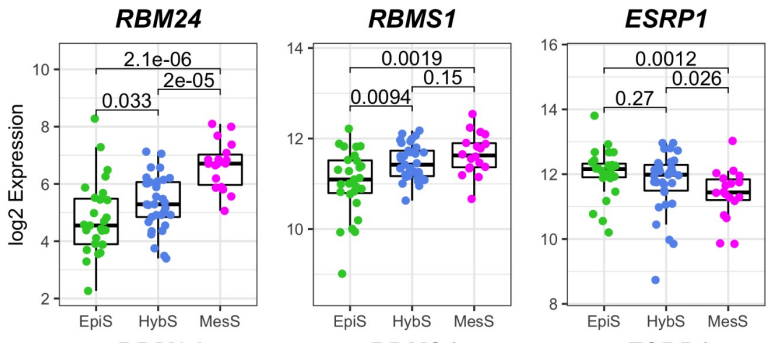

a
ESRP1
Relative mRNA expression
1.5
1.0
0.5
0.0
****
****
NC siESRP1 NC siESRP1 NC siESRP1
AGS
MKN-74
MKN-1
RBM24
Relative mRNA expression
****
NC siRBM24 NC siRBM24 NC siRBM24
AGS
MKN-74
MKN-1
b
Ladder
AGS
MKN-74
MKN-1
NC
siESRP1
siRBM24
EpiS Inclusion
CD44
500 bp
300 bp
200 bp
PSI
FLNA
200 bp
ACTN1
200 bp
100 bp
DMPK
300 bp
200 bp
SMTN
500 bp
300 bp
200 bp
MesS Inclusion
ARHGEF11
300 bp
TPM1
300 bp
200 bp
MARK3
200 bp
100 bp
ns
*
**
***
****
|△PSI|≥10%
10%>|△PSI|≥5%

Supplementary Figure 9. Alternative splicing patterns in gastric cancer cell lines with ESRP1 or RBM24 knock down. (a) mRNA expression of *ESRP1* or *RBM24* in AGS, MKN-74, and MKN-1 cell lines transfected with siRNAs targeting ESRP1 or RBM24 (siESRP1 or siRBM24) or negative control (NC). **(b)** RT-PCR validation of AS events in gastric cancer cell lines with siRNAs targeting ESRP1 or RBM24 (siESRP1 or siRBM24) or negative control (NC). For each AS event, a representative gel is shown along with PSI values calculated as % included/ (%included+skipped); significant ΔPSI (siRNA-NC) values are colored (n=3; **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001. ns- not significant).

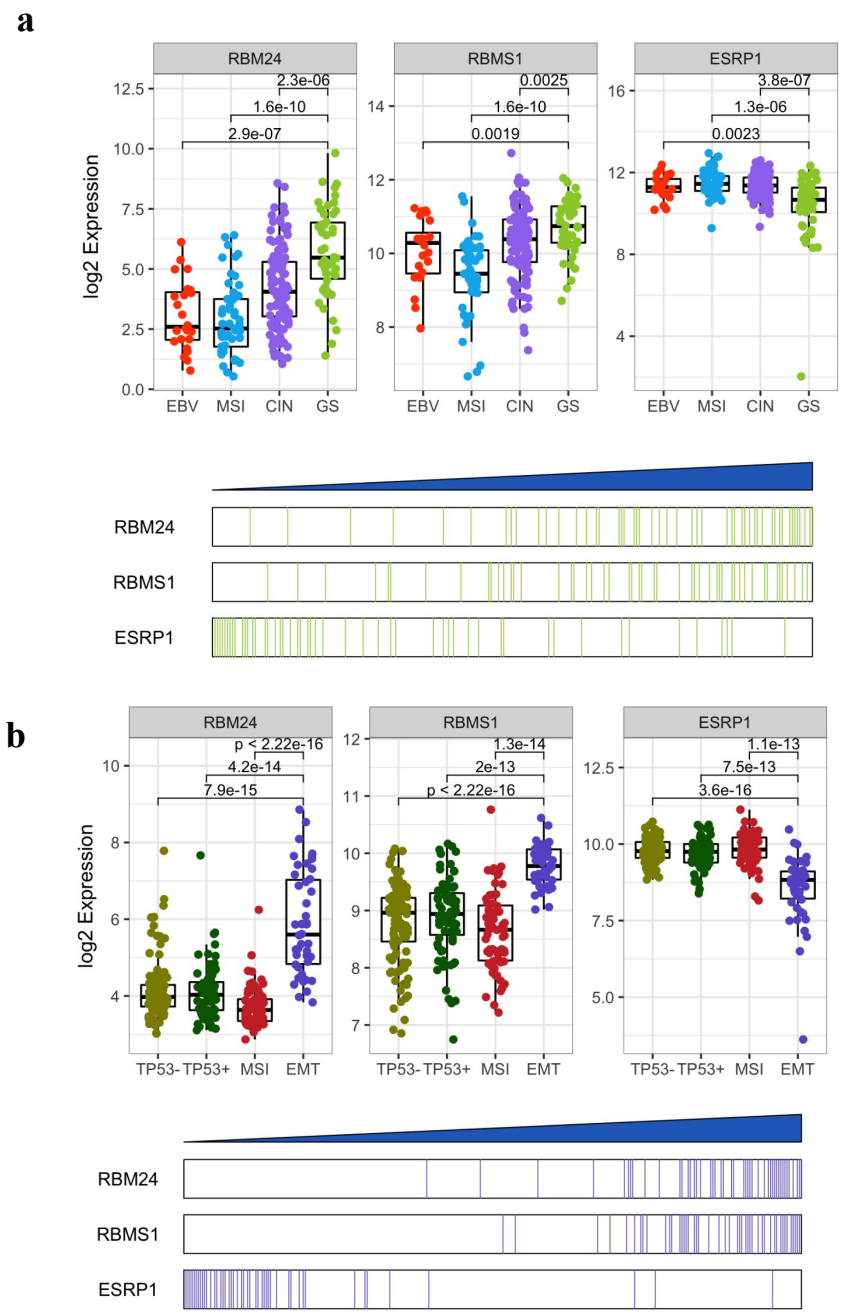
a



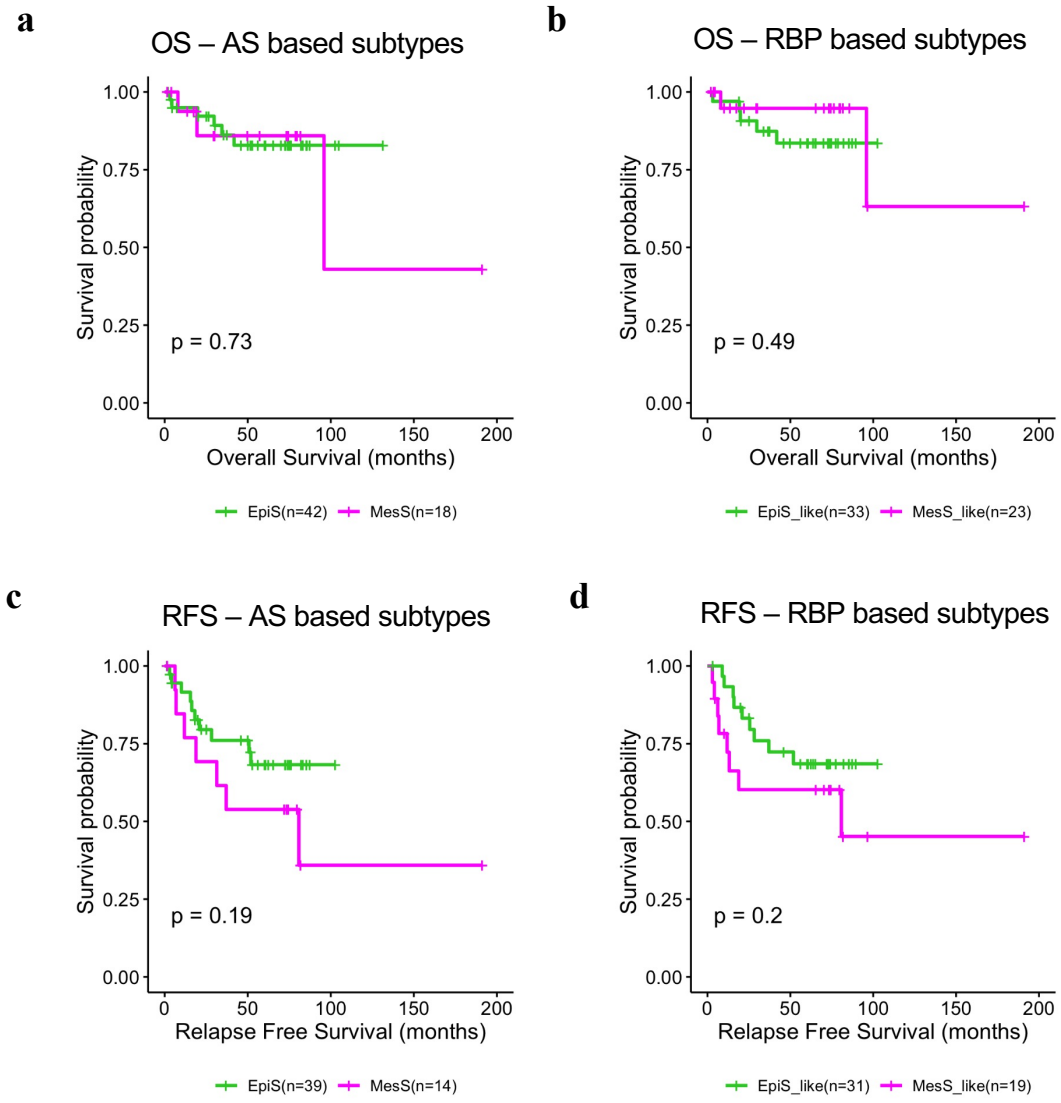
b



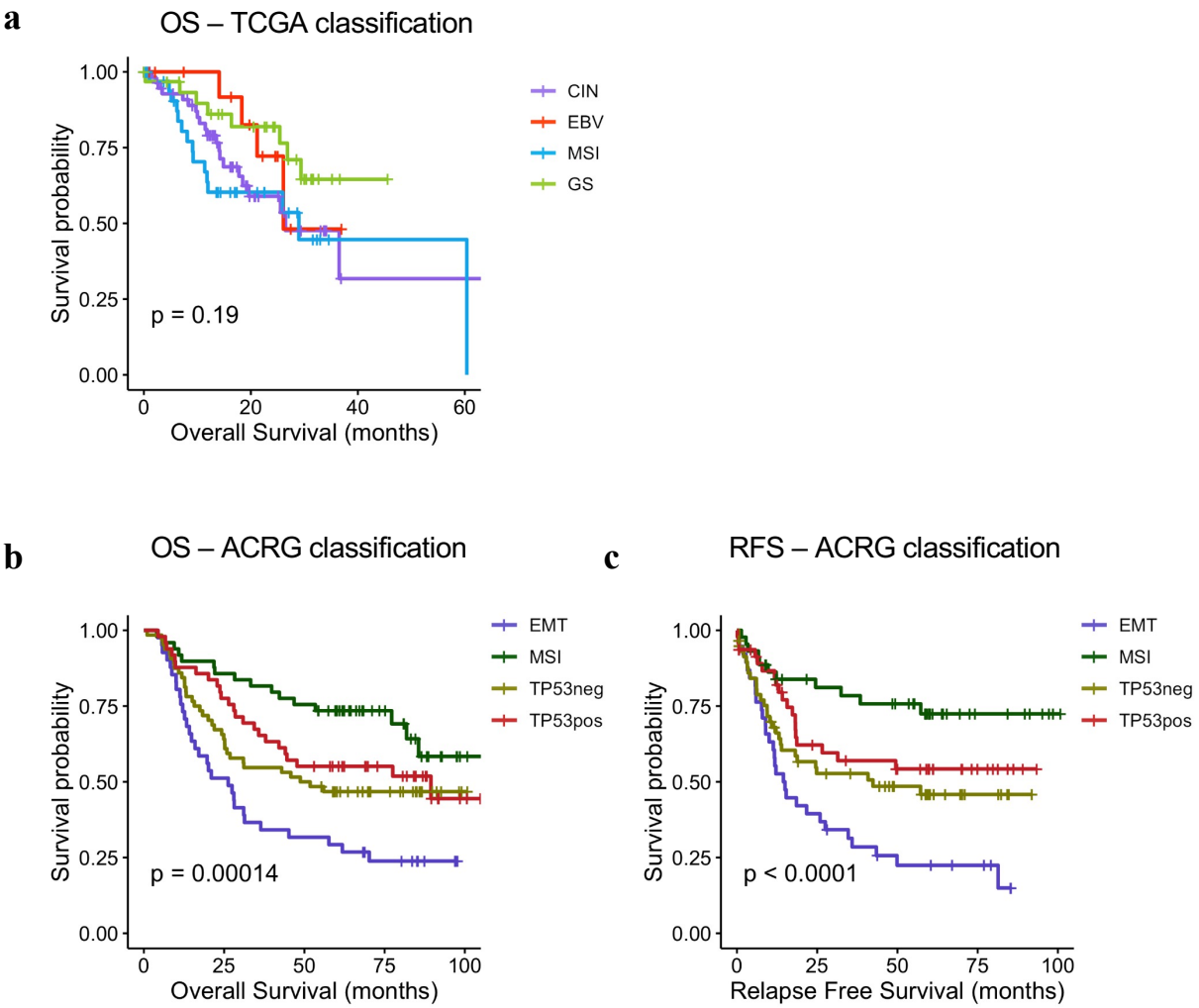
Supplementary Figure 10. Functional characterization of AS-event based subtypes in the Hwang's proteogenomic cohort. (**a**) Gene expression of *CDH1, EPCAM,* and *OCLN* epithelial and *VIM, CDH2* and *FN1* mesenchymal markers in each patient subtype (Wilcoxon rank sum test, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001). (**b**) Gene expression of regulatory RBPs in each patient subtype (Wilcoxon rank sum test, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001).



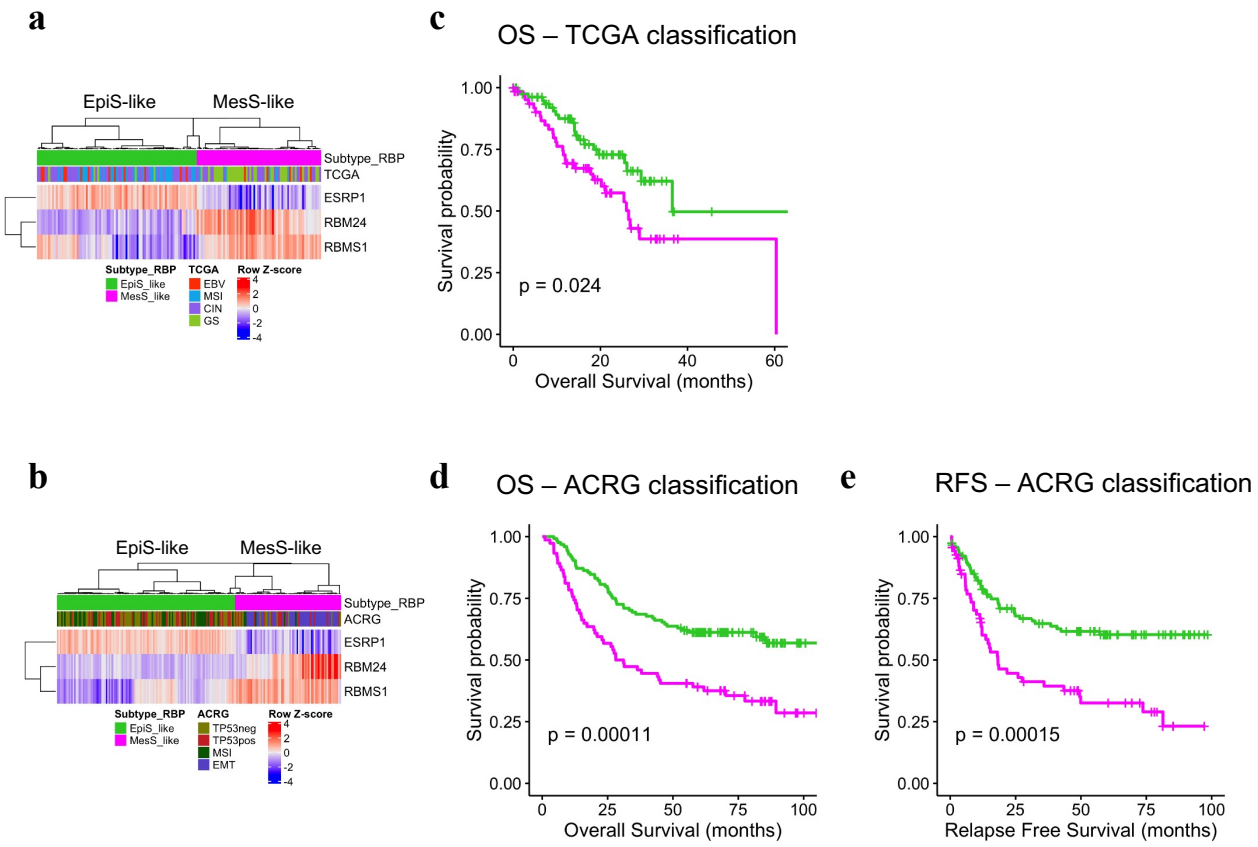
Supplementary Figure 11. Gene expression of splicing regulatory RBPs in the TCGA and ACRG cohorts. (**a,b**) RBP gene expression in TCGA (a) and ACRG (b) patient cohorts. Box plots indicate the gene expression for each subtype (Wilcoxon rank sum test P-values). The enrichment plots are shown in the bottom, where the position of the patients are indicated as tick marks (green in the GS subtype and violet in the EMT subtype) after positioning patients by each RBP expression in the increasing order. Note that RBM24 has slightly better P-value and enrichment than RBMS1 in both cohorts.



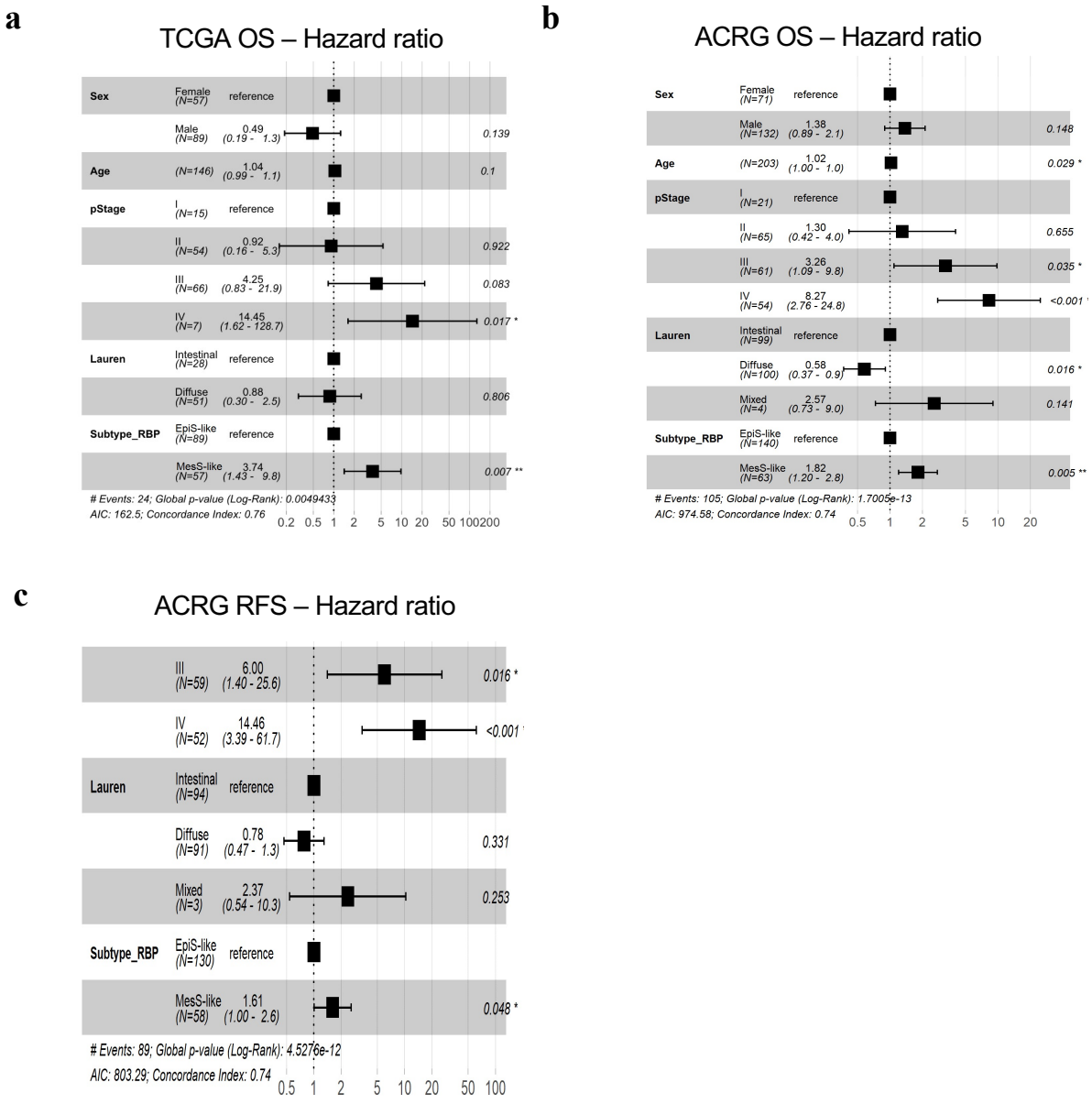
Supplementary Figure 12. Survival of gastric cancer patients classified by splicing-derived subtypes. Overall survival (a,b) and relapse free survival (c,d) analysis according to the AS-based (a,c) and RBP-based subtypes (b,d) (log-rank test *P*-values).



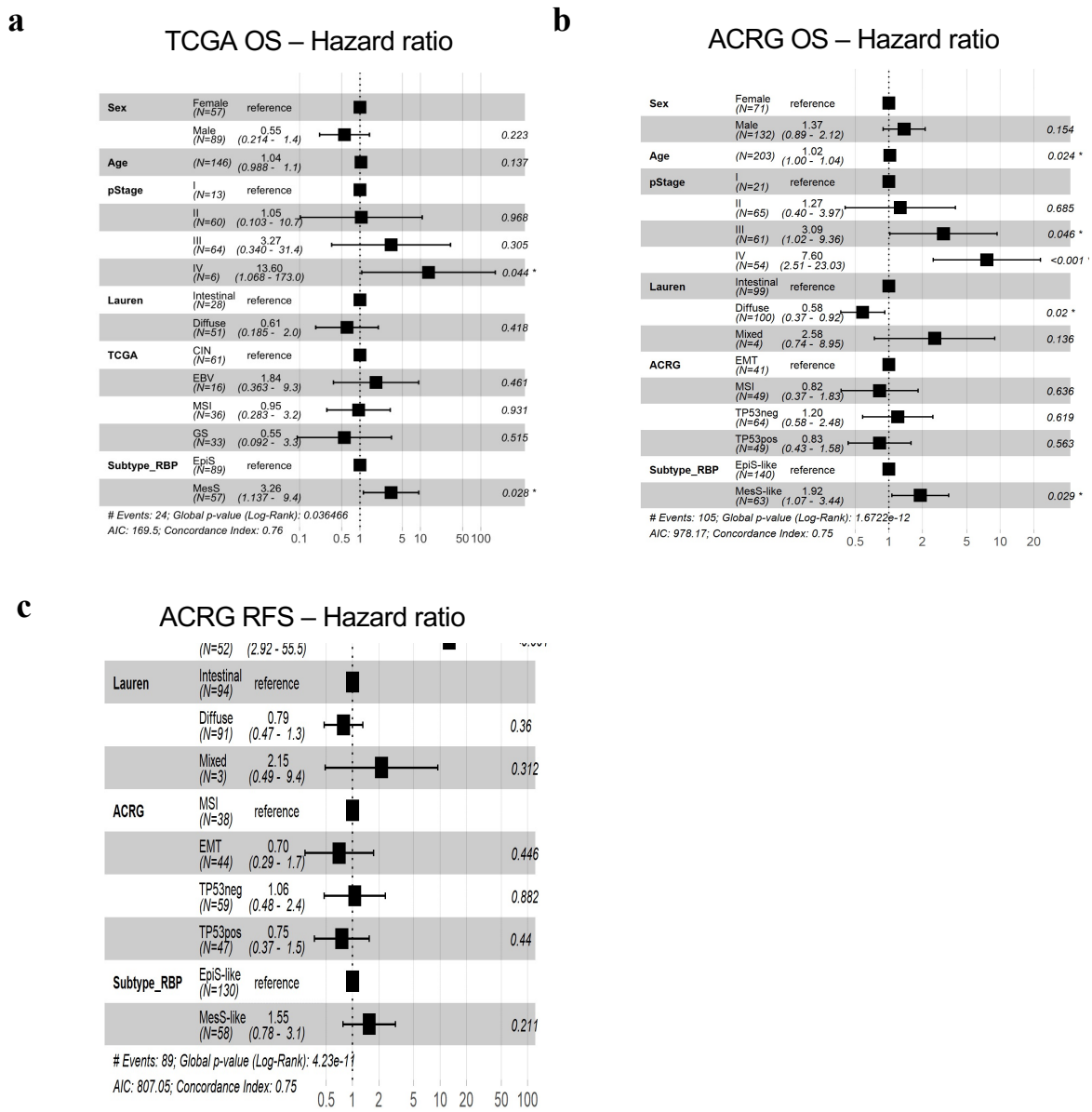
Supplementary Figure 13. Survival analysis of TCGA and ACRG gastric cancer patient cohorts. (a) Overall survival of the TCGA cohort according to the TCGA subtypes. **(b, c)** Overall (b) and relapse-free (c) survival in the ACRG cohort according to the ACRG subtypes.



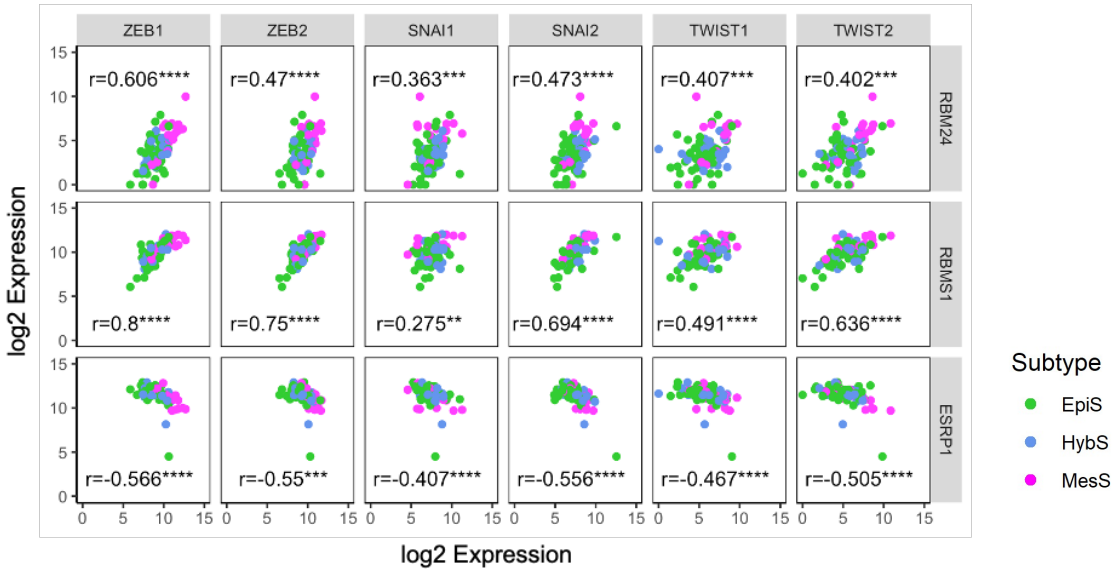
Supplementary Figure 14. Patient subgrouping of the TCGA and ACRG cohorts based on expression of three RBPs. (**a,b**) Three RBP-based classification of TCGA (a) and ACRG (b) patients according to *RBM24, RBMS1,* and *ESRP1* expression levels. (**c,d**) Kaplan-Meier plot for overall survival between the EpiS-like and MesS-like RBP-subtypes in the TCGA (c) and ACRG (d) cohort (log-rank test *P*-values). (**e**) Kaplan-Meier plot for relapse-free survival between EpiS-like and MesS-like RBP-subtypes in the ACRG cohort (log-rank test *P*-values).



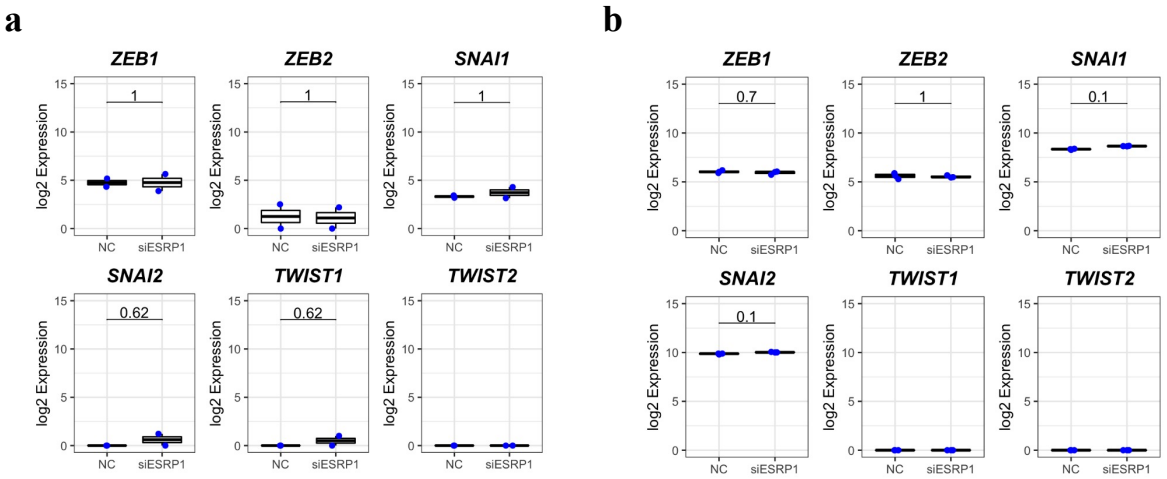
Supplementary Figure 15. Multivariate Cox regression analysis in the TCGA and ACRG cohorts. (**a**) Forest plot of hazard ratios (HR) with 95% confidence intervals (CI) for overall survival in the TCGA cohort. (**b, c**) Forest plot of HR with 95% CI for overall survival (b) and relapse-free survival (c) in the ACRG cohort. Continuous variables HR larger than 1 represent factors associated with additional risk of death, whereas those with HR from 0 to 1 represent decreasing risk (Cox hazard , **P*<0.05, ***P*<0.01).



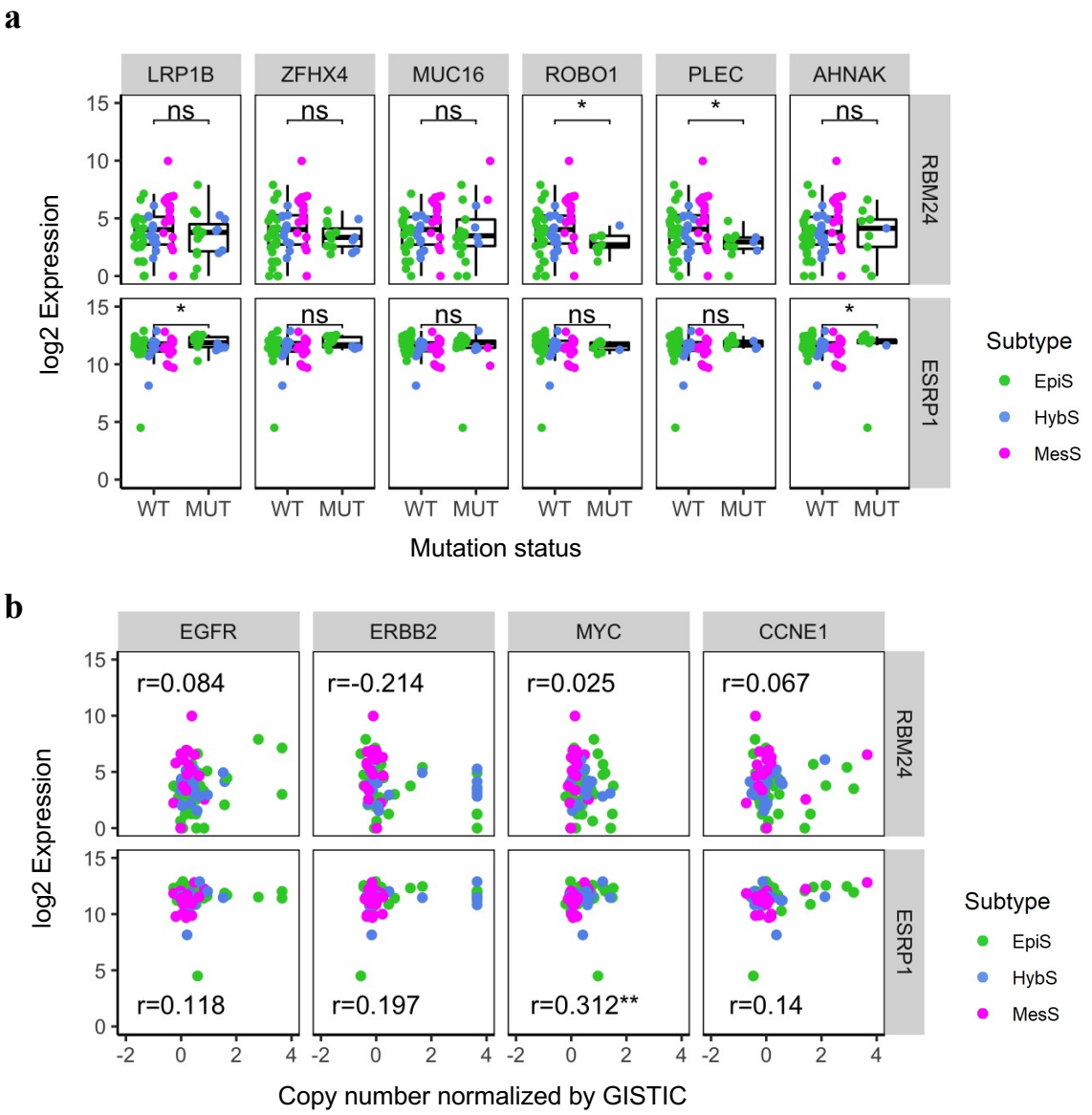
Supplementary Figure 16. Multivariate Cox regression analysis in the TCGA and ACRG cohorts including RBP-based subtypes with either the TCGA or ACRG subtypes. (a) Hazard ratios (HR) with 95% confidence intervals (CI) for overall survival in the TCGA cohort. **(b, c)** HR with 95% CI for overall survival (b) and relapse-free survival (c) in the ACRG cohort. Continuous variables HR larger than 1 represent factors associated with additional risk of death, whereas those with HR from 0 to 1 represent decreasing risk (Cox hazard, **P*<0.05, ** *P*<0.01).



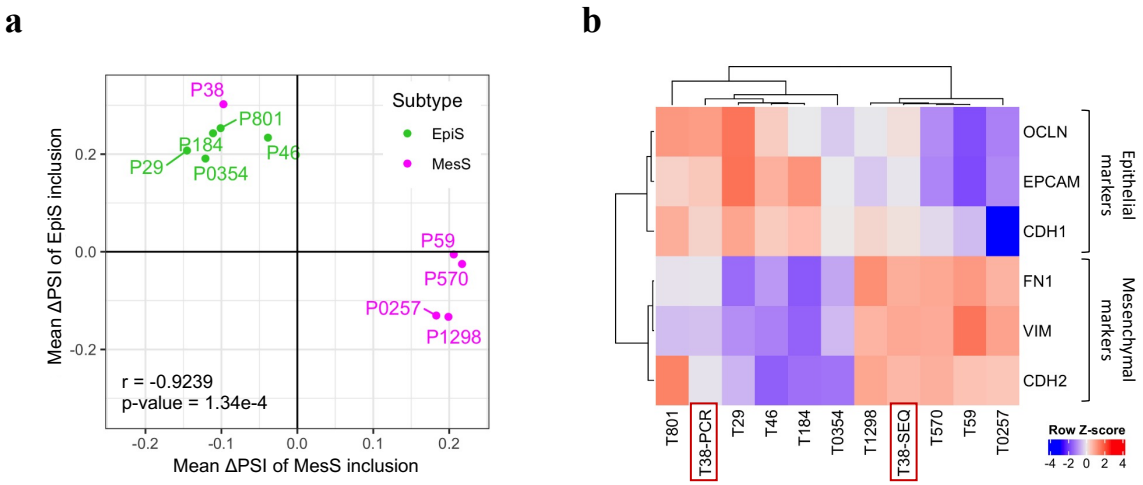
Supplementary Figure 17. Correlation between EMT-TFs and RBPs expression in gastric tumors. Each dot represents patients colored by AS event-based subtypes (Spearman correlation coefficient and *P*-value, **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001). Tumors are color coded by splicing subtype.



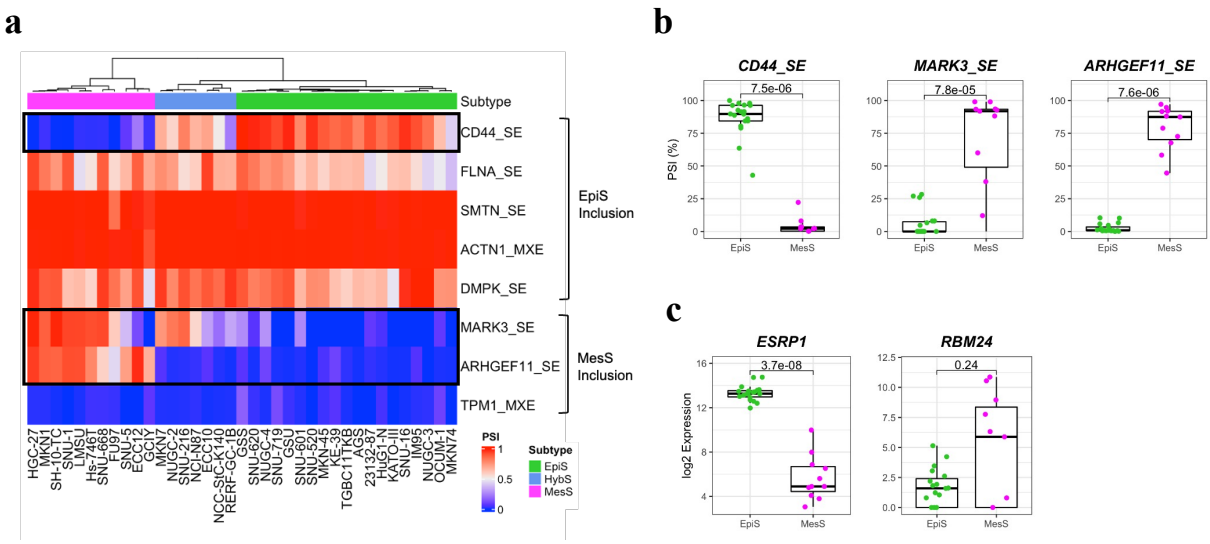
Supplementary Figure 18. Expression EMT-associated TFs upon ESRP1 knockdown in epithelial cells. TFs levels were compared between control (NC) and ESRP1 knockdown (siESRP1) in LCC2 and LCC9 breast (SRP180303) **(a)** and H358 lung (SRP066789) **(b)** cancer cell lines (n=3; median ± interquartile range; Wilcoxon rank-sum test *P*-values).



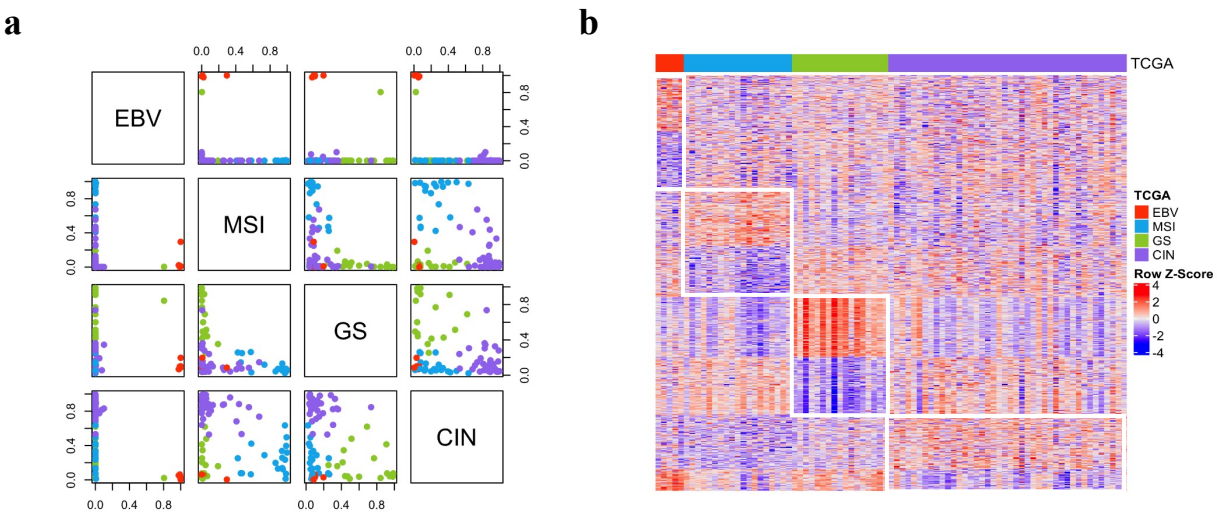
Supplementary Figure 19. Association analysis of RBP expression. (**a**) *RBM24* or *ESRP1* expression in tumors wild-type (WT) or mutant (MUT) for indicated genes (Wilcoxon rank sum test *P*-values). (**b**) Correlation plots of *RBM24* or *ESRP1* expression and copy number variations in tumors for indicated genes. Tumors are color coded by splicing subtype.



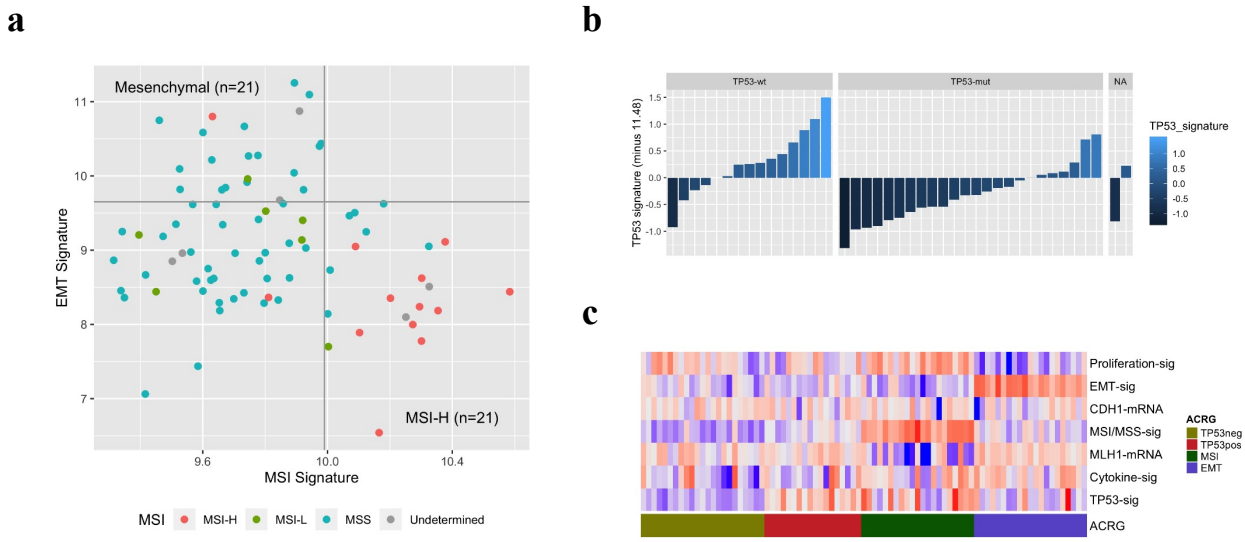
Supplementary Figure 20. Tumor heterogeneity in P38 patient. (**a**) Scatter plot of two AS-event types for the 10 patients used in experimental RT-PCR validation. Note that P38 was positioned in the EpiS Subtype region even though it was classified as a MesS subtype in the original RNA-Seq data. (**b**) Expression EMT marker genes from RNA-Seq data. T38-PCR and T38-SEQ represent the original RNA sample used for RNA-Seq and a second RNA sample from the same patient used for RT-PCR validation and then sequenced also by RNA-Seq, respectively. Note that the EMT markers of T38-PCR and T38-SEQ behave oppositely even though they were from a single person (confirmed by NGSCheckMate).



Supplementary Figure 21. Alternative splicing of representative AS events in gastric cancer cell lines. (a) Hierarchical clustering using PSI values for eight AS events classifies 37 gastric cancer cell lines into three AS-event based subtypes. Three AS events indicated by boxes account for most of the subtype differences. **(b)** PSI values distribution for AS events in *CD44*, *MARK3*, and *ARHGEF11* in EpiS and MesS gastric cancer cell lines (Wilcoxon rank-sum test *P*-values). **(c)** *ESRP1* and *RBM24* expression in EpiS and MesS gastric cancer cell lines (Wilcoxon rank-sum test *P*-values).



Supplementary Figure 22. Classification of our patients according to the TCGA method. (**a**) Scatter plot matrix of Bayesian probability for each gastric cancer subtype predictor in the TCGA cohort. Each patient was assigned with four probability scores used in the plot. Note that the probability scores are high for the type of patient subtype almost exclusively. (**b**) Expression heatmap of signature genes as provided by the TCGA. Note that the signature genes from the TCGA were also up- and down-regulated specifically in the corresponding subtype in our data. Thus, we can conclude that our clustering results using the TCGA classification scheme was quite successful.



Supplementary Figure 23. Classification of our patients according to the ACRG method. (**a**) Plot of EMT signature *vs.* MSI signature. The signature scores were obtained as the average expression value of signature genes as provided by the ACRG consortium. Patients in the second and fourth quadrant were assigned as the EMT subtype and the MSI subtype, respectively. The MSI status was determined independently from fragment analysis. Note that MSI subtype was highly enriched with our MSI-H patients (red dots). (**b**) Plot of *TP53* signature scores for the remaining patients. We verified the *TP53*mutation status for our patients. Majority of the *TP53* wildtype and mutant patients showed positive and negative signature scores, respectively. (**c**) Heatmap of signatures and marker genes for ACRG classification. Each subtype showed highly activated signatures of its own, as expected.