

Supplementary Materials for

The consensus molecular subtypes of esophageal squamous cell carcinoma

Heyang Cui#, Zhongxu Zhu#, Enwei Xu#, Lin Qi#, Yikun Cheng#, Yinghan Zhang#, Ling Zhang, Matthew Yibo Cheng, Bin Yang, Ruifang Sun, Xiaofei Zhuang, Yanfeng Xi, Ting Yan, Caixia Cheng, Ning Ding, Huijuan Liu, Lu Wang, Min Guo, Dinghe Guo, Haoyu Zhang, Meilan Peng, Zhekun An, Yongjia Weng,, Fang Wang, Meng Liu, Ruixin Xiong, Weihua Yin, Bin Song, Weimin Zhang*, Xiaolong Cheng, Zhihua Liu, Qimin Zhan*, Xin Wang*, Yongping Cui*.

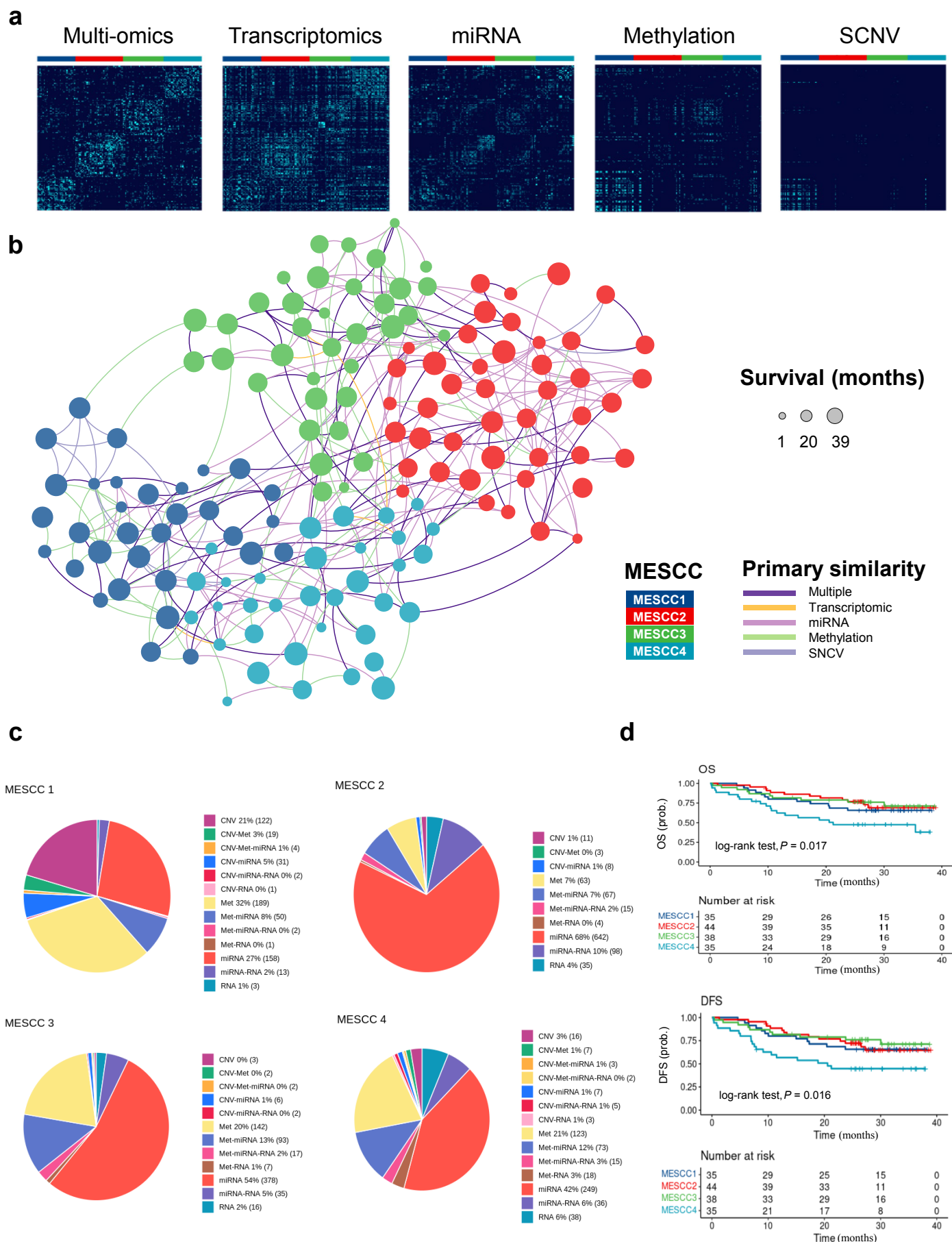
Correspondence to: cuiyp@sphmc.org

This PDF file includes:

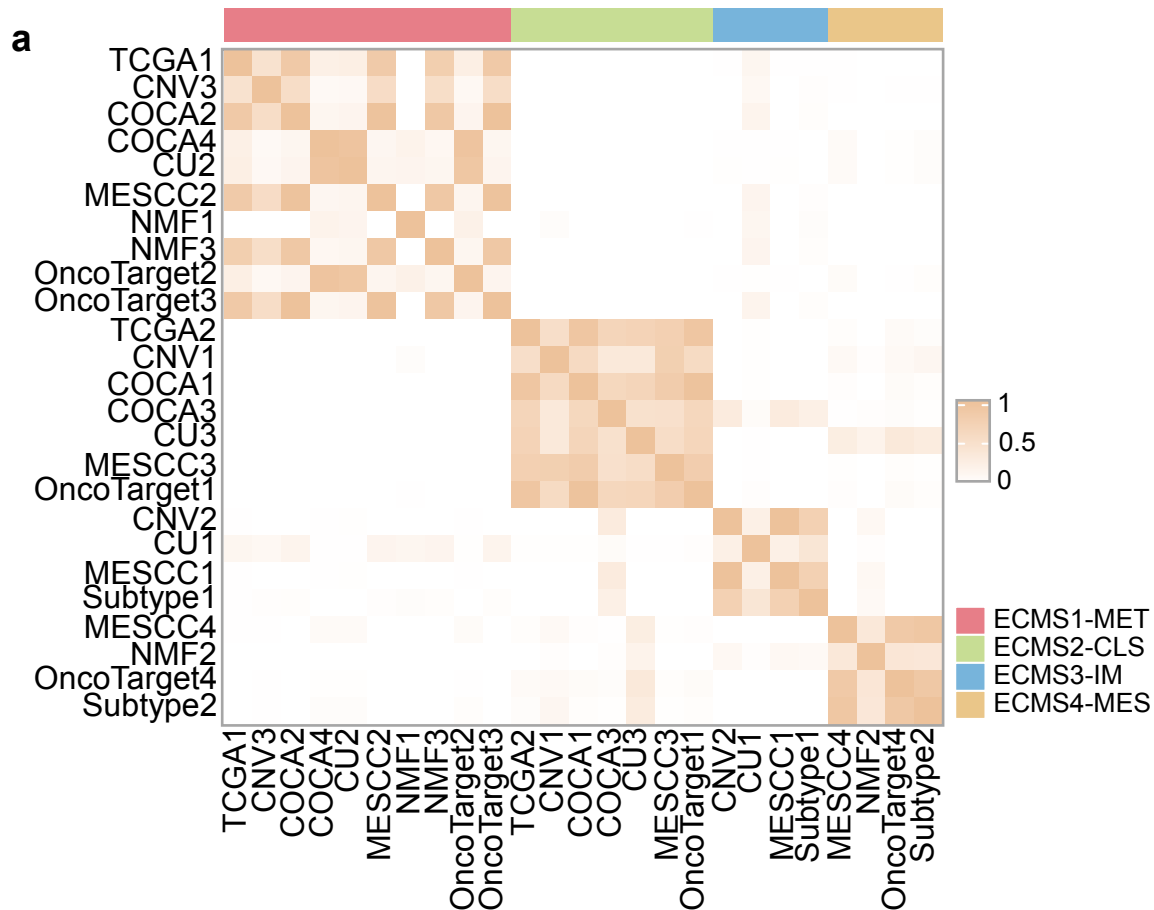
Supplementary Figs. 1-22

Other Supplementary Materials for this manuscript include the following:

Supplementary data 1-15



Supplementary Fig. 1. Multi-omics subtyping of ESCC based on similarity network fusion. (a) Patient-to-patient similarity matrices for 152 ESCC samples based on multi-omics data, mRNA expression, miRNA expression, DNA methylation and SCNv profiles, respectively. (b) Patient fusion network illustrating the associations between patients assigned to the four MESCC subtypes. Each node represents a patient with node color denoting subtype information and node size proportional to the survival time. Edges were colored by the data type showing the strongest similarity between samples. (c) Evidence distribution across four MESCC subtypes. (d) Kaplan-Meier analysis for overall survival and disease-free survival, with patients stratified into the four MESCC subtypes.



b Cui et al.

ECMS1	0.49	1	0.060
ECMS2	0.28	0.87	0.56
ECMS3	0.48	0.34	0.82
ECMS4	0.97	0.0018	0.92

NMF1 NMF2 NMF3

c Du et al.

ECMS1	0.82	0.55	0.16
ECMS2	6.6×10^{-5}	1	0.39
ECMS3	1	3.1×10^{-7}	1
ECMS4	0.46	0.51	1

CNV1 CNV2 CNV3

d Liu et al.

ECMS1	0.99	1.4×10^{-10}	1
ECMS2	0.0049	1	0.030
ECMS3	0.0012	0.98	0.97
ECMS4	1	0.10	0.004

CU1 CU2 CU3

e TCGA

ECMS1	5.3×10^{-10}	1
ECMS2	0.99	0.015
ECMS3	0.99	0.02
ECMS4	0.82	0.31

TCGA1 TCGA2

f Wang et al.

ECMS1	0.53	0.62
ECMS2	0.91	0.14
ECMS3	1.4×10^{-9}	1
ECMS4	1	1.7×10^{-6}

Subtype1 Subtype2

g Xiong et al.

ECMS1	1	6.5×10^{-4}	1.8×10^{-8}	1
ECMS2	3.7×10^{-6}	0.99	0.98	0.82
ECMS3	0.053	0.052	0.99	0.99
ECMS4	1	0.99	1	5.7×10^{-13}

Oncotarget1 Oncotarget2 Oncotarget3 Oncotarget4

h Liu et al.

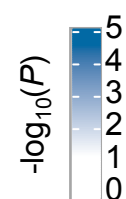
ECMS1	1	9.7×10^{-6}	0.99	0.42
ECMS2	9.2×10^{-4}	0.89	0.44	0.99
ECMS3	0.95	0.98	0.012	32
ECMS4	0.38	0.99	0.88	0.032

COCA1 COCA2 COCA3 COCA4

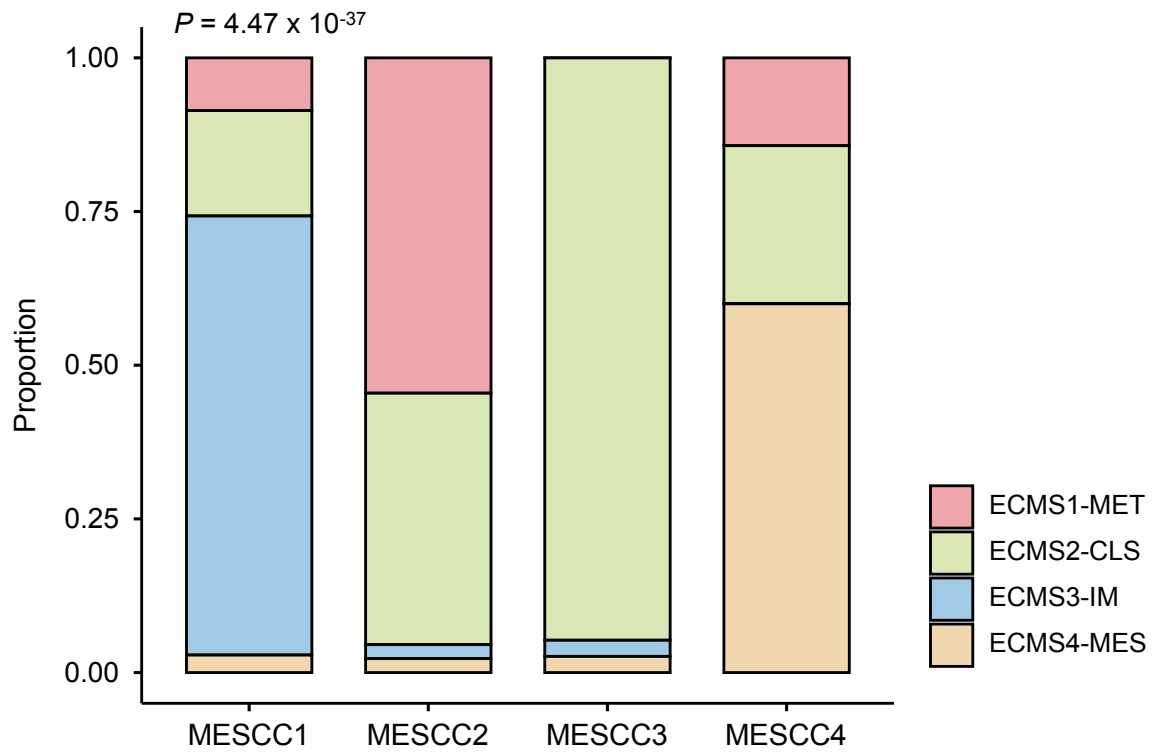
i MESCC

ECMS1	0.99	8.8×10^{-15}	1	0.91
ECMS2	1	0.81	1.8×10^{-13}	0.99
ECMS3	1.9×10^{-18}	1	1	1
ECMS4	0.99	1	0.99	1.1×10^{-13}

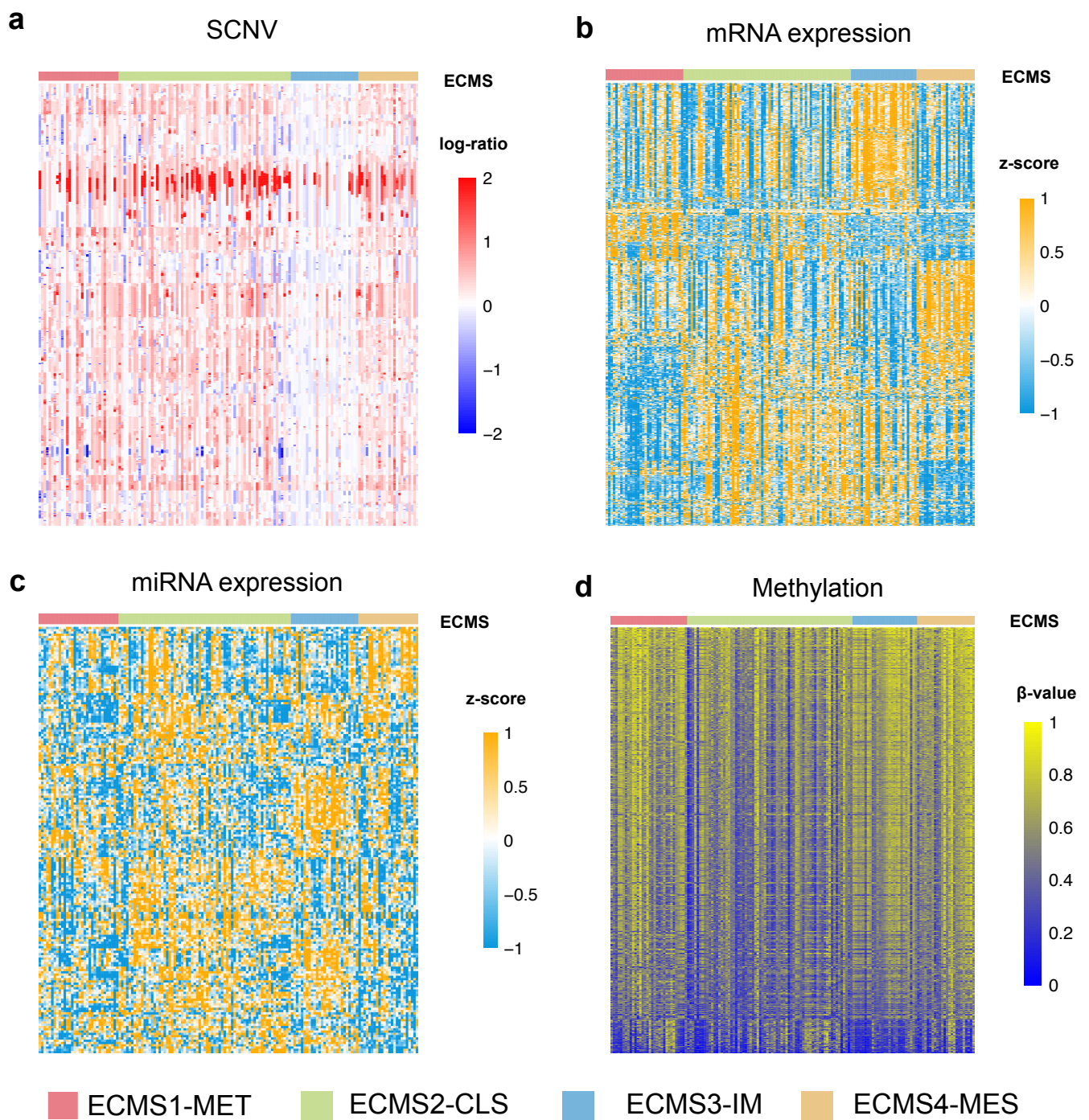
MESCC1 MESCC2 MESCC3 MESCC4



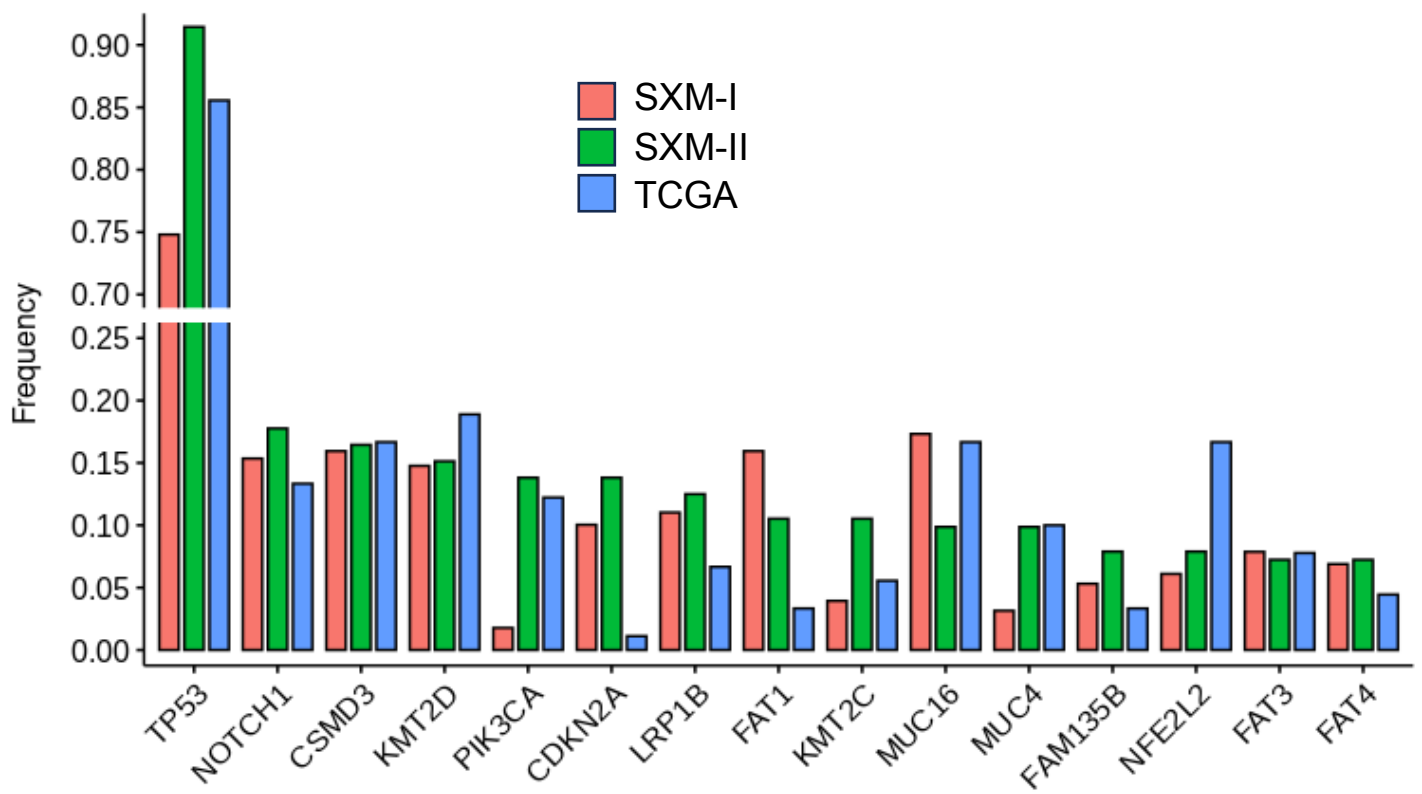
Supplementary Fig. 2. Statistical associations between different ESCC subtyping systems and the consensus matrix of ESCC. (a) Consensus matrix, defined by the frequency with each pair of subtypes partitioned into the same cluster. **(b-i)** Heatmaps illustrating the associations between ECMS and other subtyping systems by hypergeometric tests.



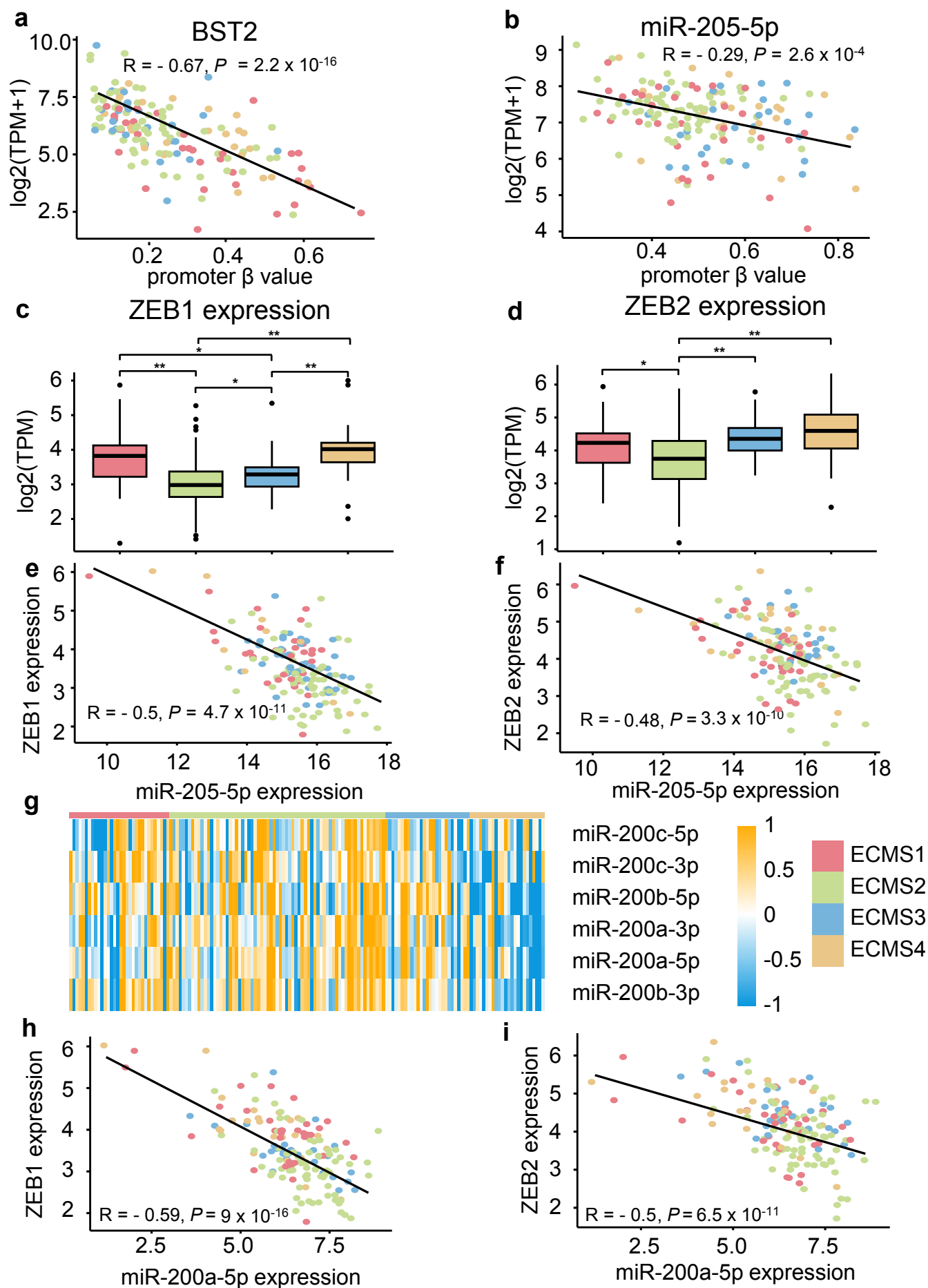
Supplementary Fig. 3. Association between MESCC and ECMS. Barplot showing the distribution of ECMS subtypes within each MESCC subtype.



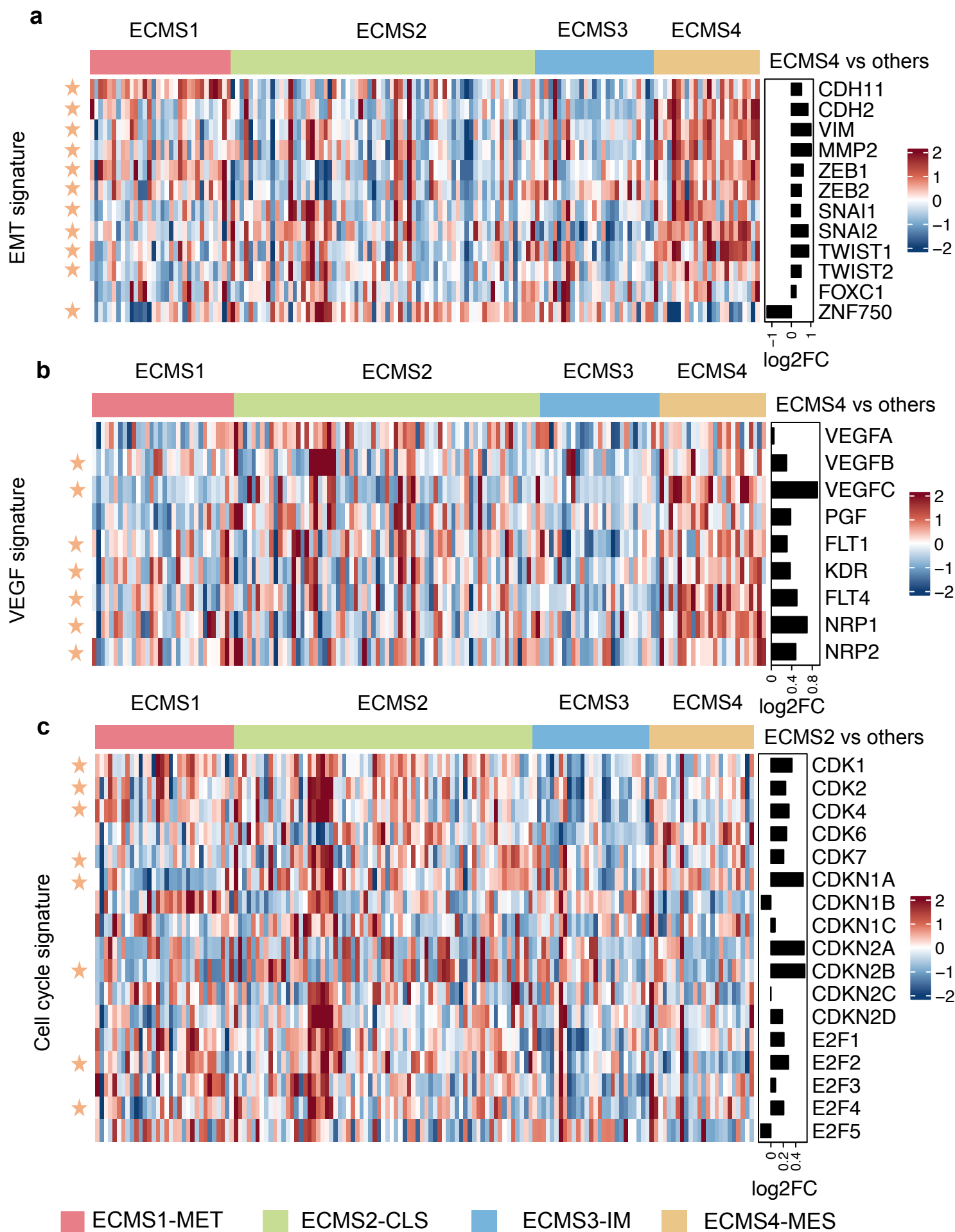
Supplementary Fig. 4. Multi-omics signatures for ECMS subtypes. Heatmaps showing the patterns of (a) copy number changes (shades of red for gains and blue for losses), (b) differentially expressed mRNAs, (c) differentially expressed miRNAs (shades of blue for downregulation and orange for upregulation), (d) DNA methylation β -values in the four ECMSs.



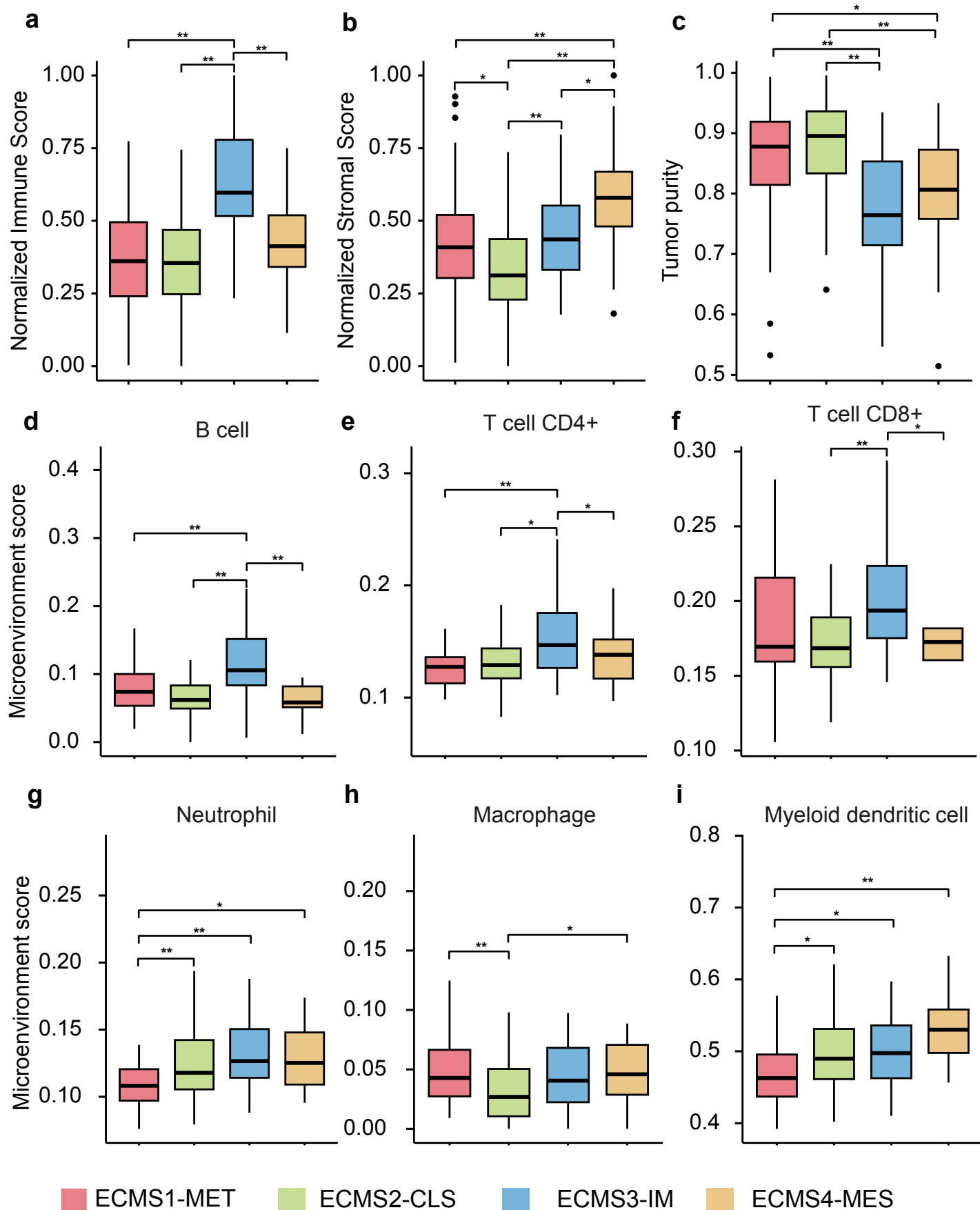
Supplementary Fig. 5. A comparison of mutation frequencies in TCGA-ESCC cohort and two Chinese ESCC cohorts



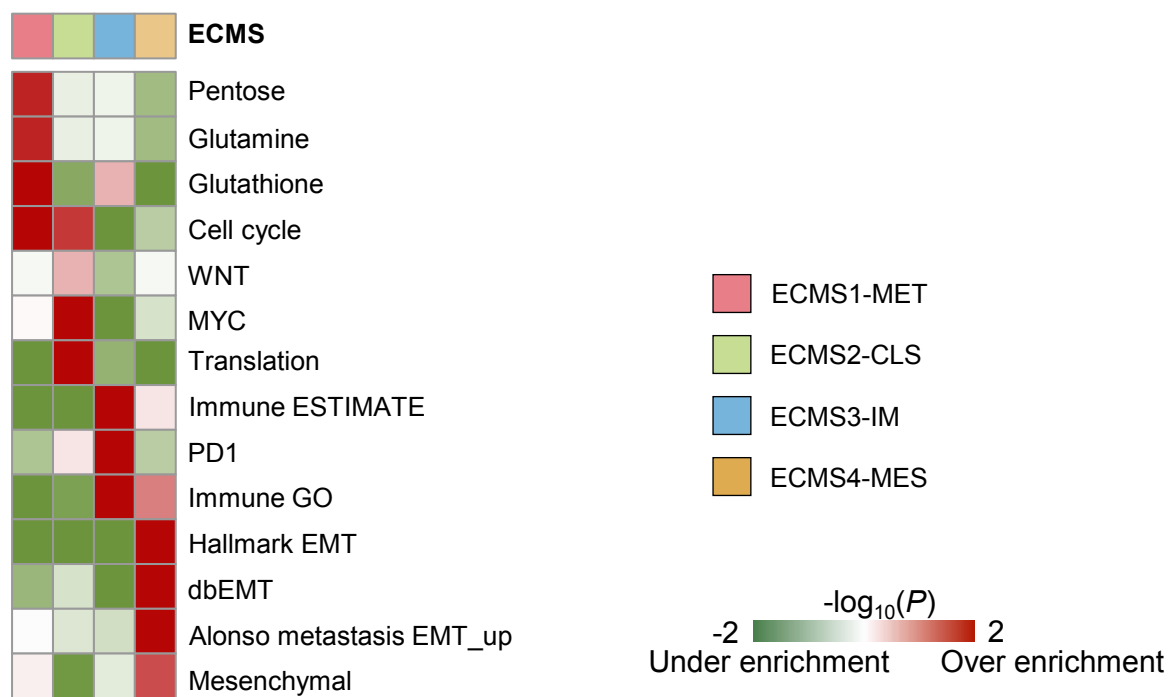
Supplementary Fig. 6. Subtype-specific gene expression regulations. (a) Scatterplot showing the correlation between the promoter methylation and expression of *BST2*. (b) Scatterplot showing the correlation between the promoter methylation and expression of miR-205-5p. (c) Boxplot showing the distribution of *ZEB1* expression across ECMSs. (d) Boxplot showing the distribution of *ZEB2* expression across ECMSs. (e-f) Scatterplot showing the correlation between the expression levels of miR-205-5p and (e) *ZEB1* and (f) *ZEB2* expression levels. (g) Heatmap showing the expression profiles of miR-200 family across different ECMSs. (h-i) Scatterplot showing the correlation between miR-200a-5p and (h) *ZEB1* and (i) *ZEB2* expression levels. (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon signed-rank tests)



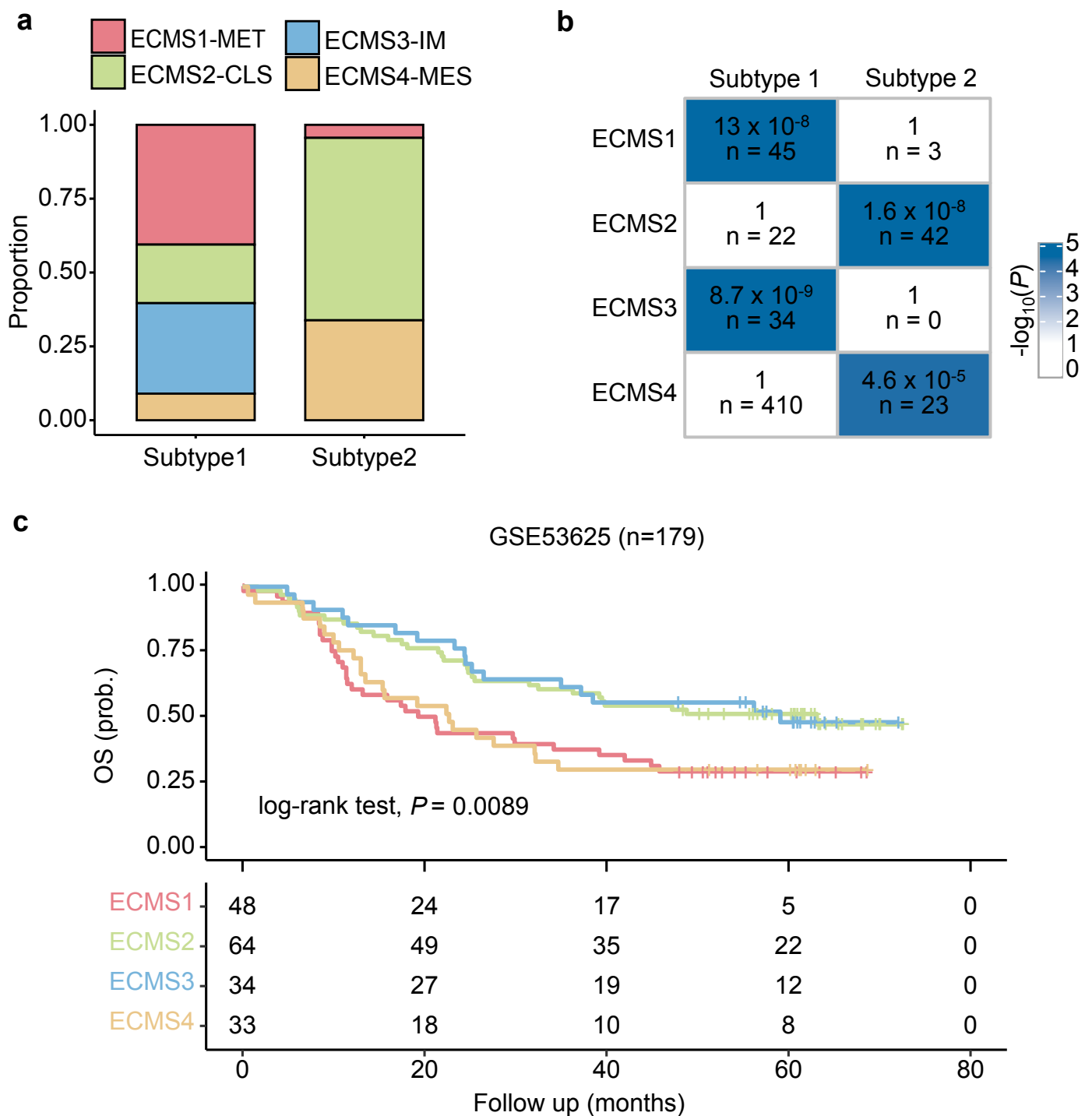
Supplementary Fig. 7. Heatmap illustrating the expression patterns of EMT, VEGF and cell cycle signature genes between the four ECMSs. Heatmaps comparing the expression of (a) EMT, (b) VEGF and (c) cell cycle signature genes between the four ECMSs. Star denotes a significant dysregulation.



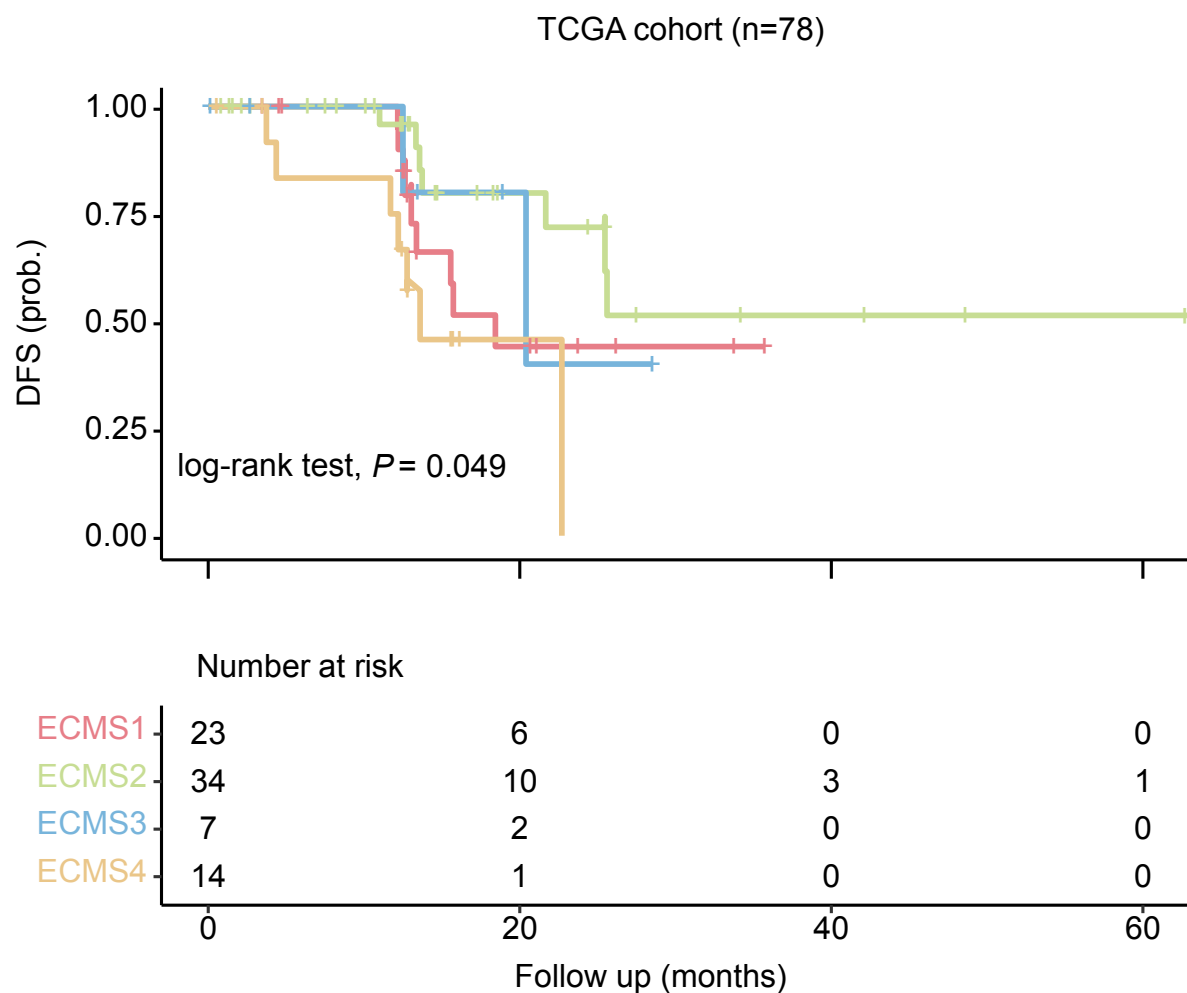
Supplementary Fig. 8. Tumor microenvironment characterizations. (a-c) Boxplots comparing (a) Immune score, (b) stromal score and (c) tumor purity calculated by ESTIMATE between the four ECMSs. The score was scaled between 0 and 1. (d-i) Boxplots comparing the microenvironment scores calculated by TIMER, including (d) B cell, (e) CD4+ T cell, (f) CD8+ T cell, (g) neutrophil, (h) macrophage and (i) myeloid dendritic cell. (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon signed-rank tests)



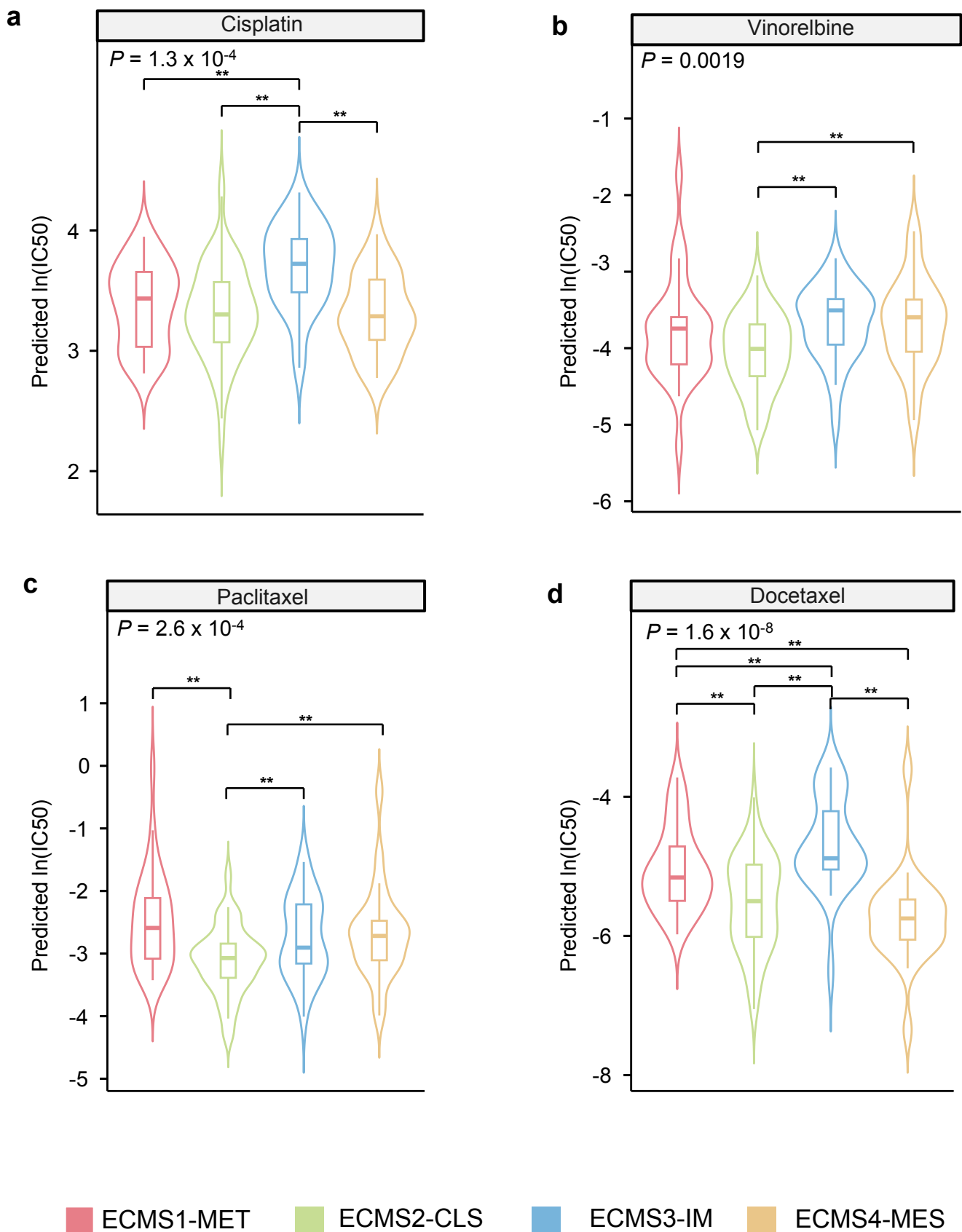
Supplementary Fig. 9. Molecular characterizations of ECMS subtypes. Gene set enrichment analysis based on proteomics profiles in the SXM-I cohort.



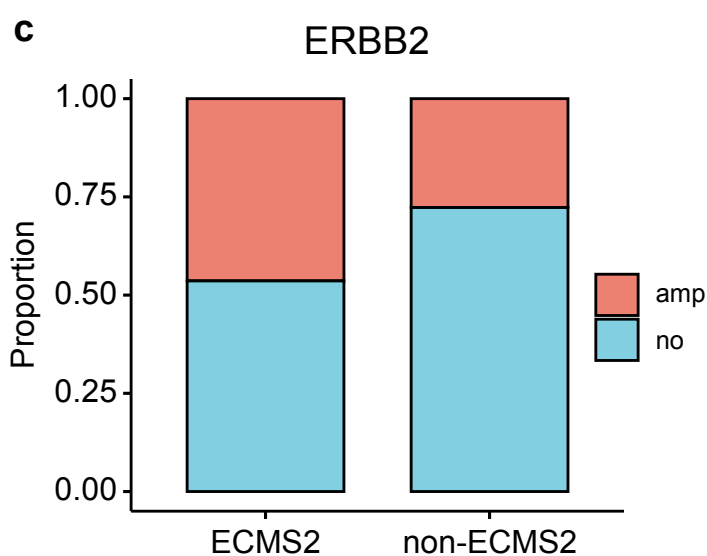
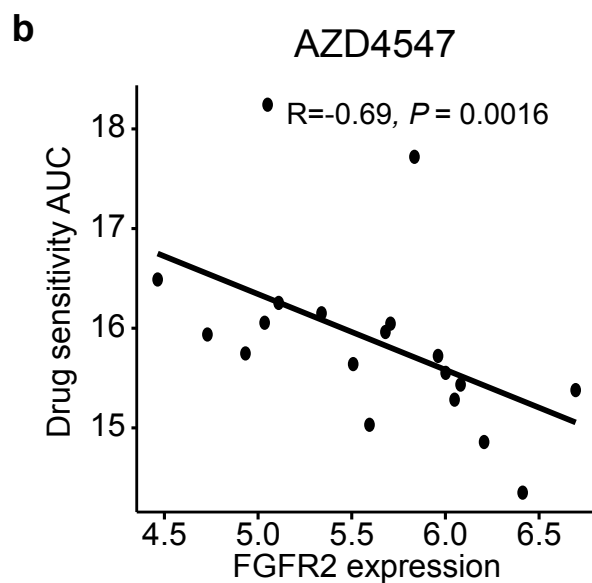
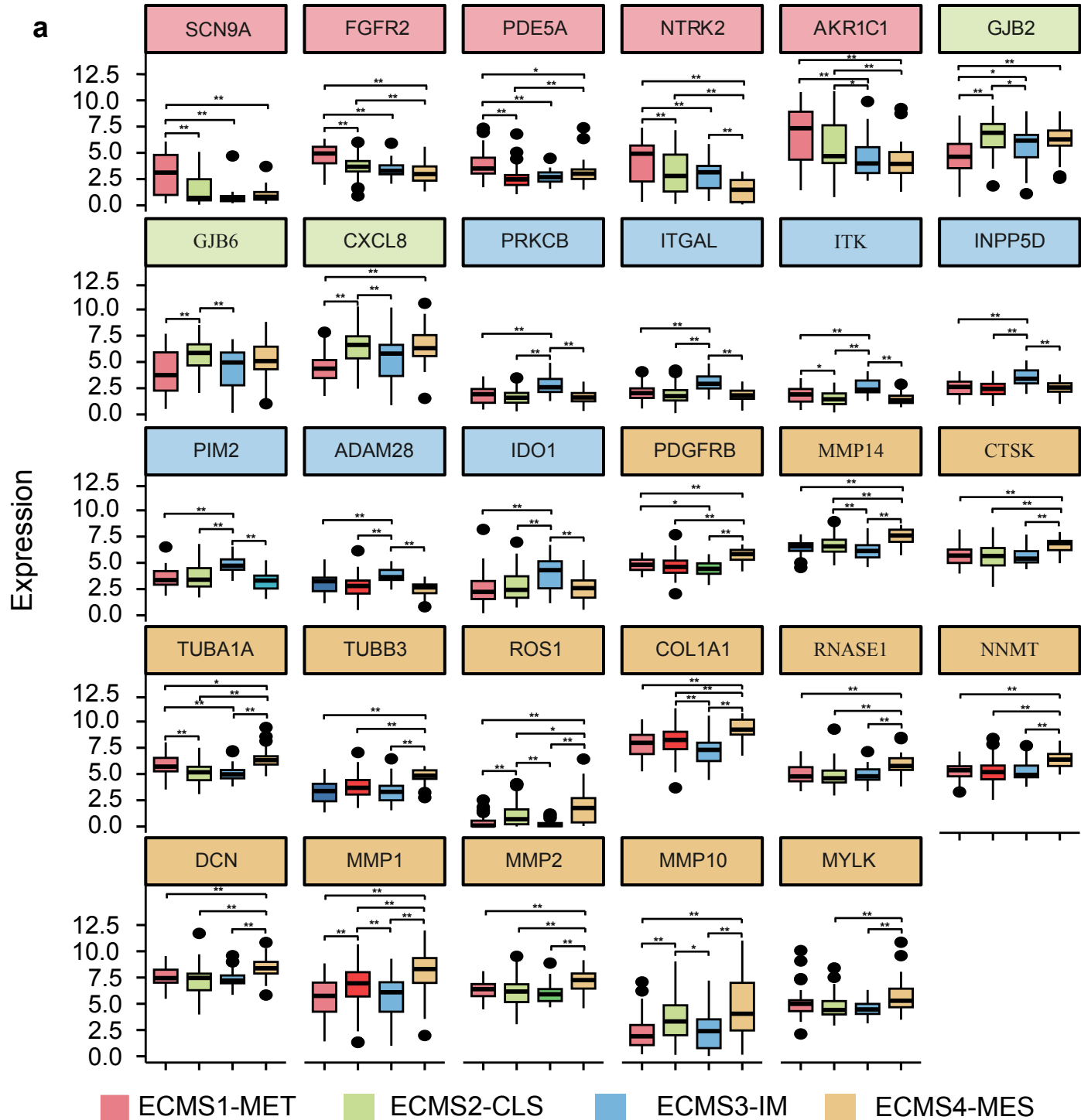
Supplementary Fig.10. ECMS prediction in the GSE53625 cohort. (a) Barplot showing the ECMS subtype distribution within each subtype defined by Wang et al¹⁹. (b) Heatmap showing the association between ECMS and the taxonomy proposed by Wang et al¹⁹. (c) KM plot illustrating the association between ECMSs and overall survival in the GSE53625 cohort.



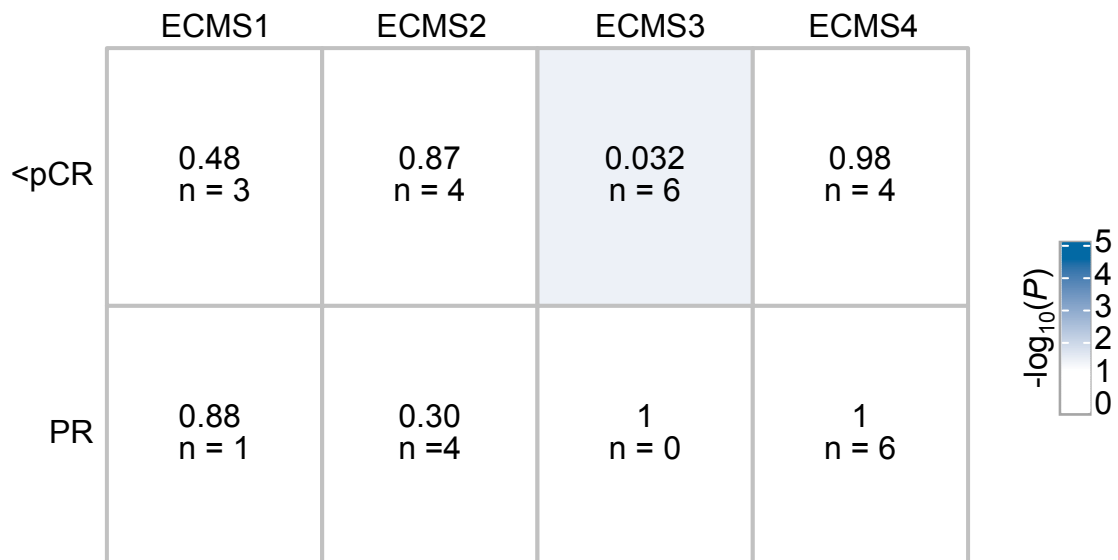
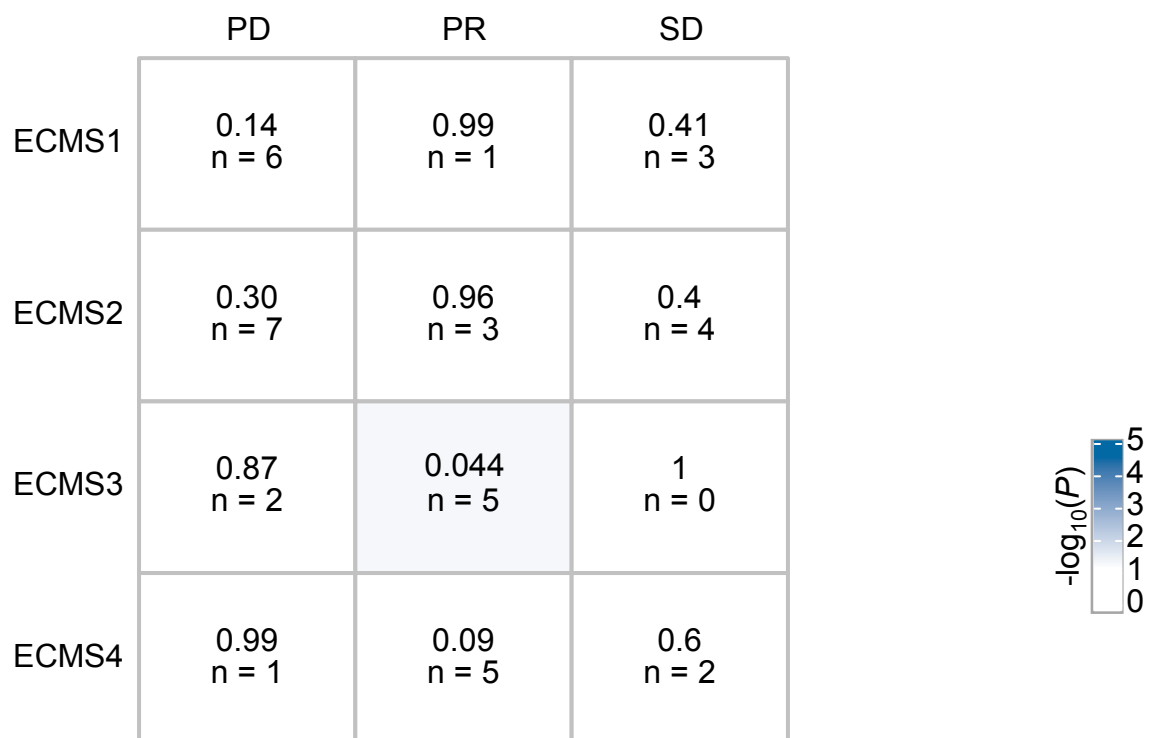
Supplementary Fig. 11. ECMS survival analysis of TCGA cohort. Association of ECMS with disease-free survival in the TCGA cohort.



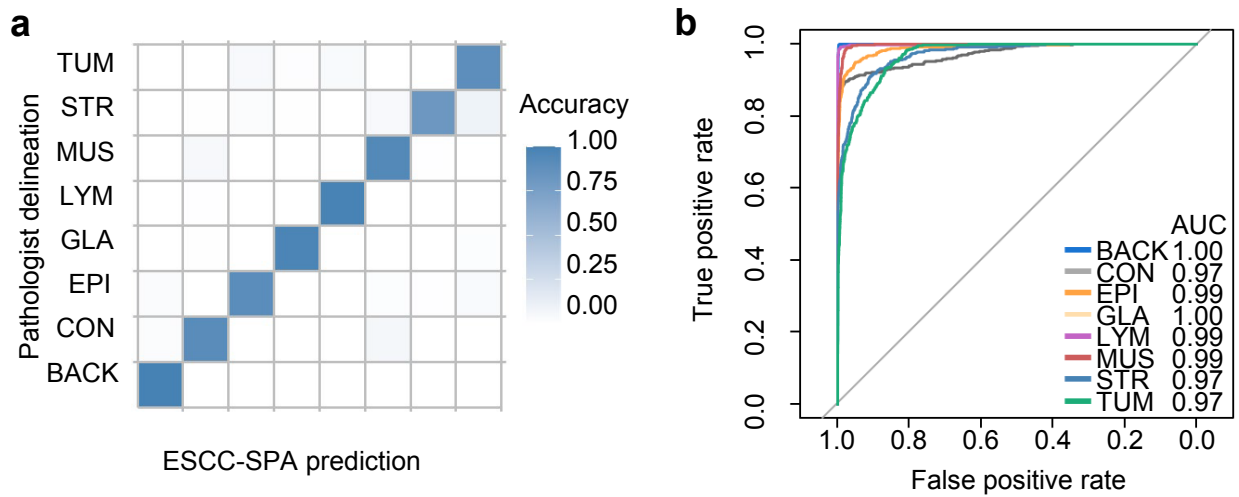
Supplementary Fig. 12. Chemotherapy sensitivity analysis. Violinplots showing the difference of chemotherapy sensitivity for ECMS. (a) Cisplatin, (b) Vinorelbine, (c) Paclitaxel and (d) Docetaxel. (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon signed-rank tests)



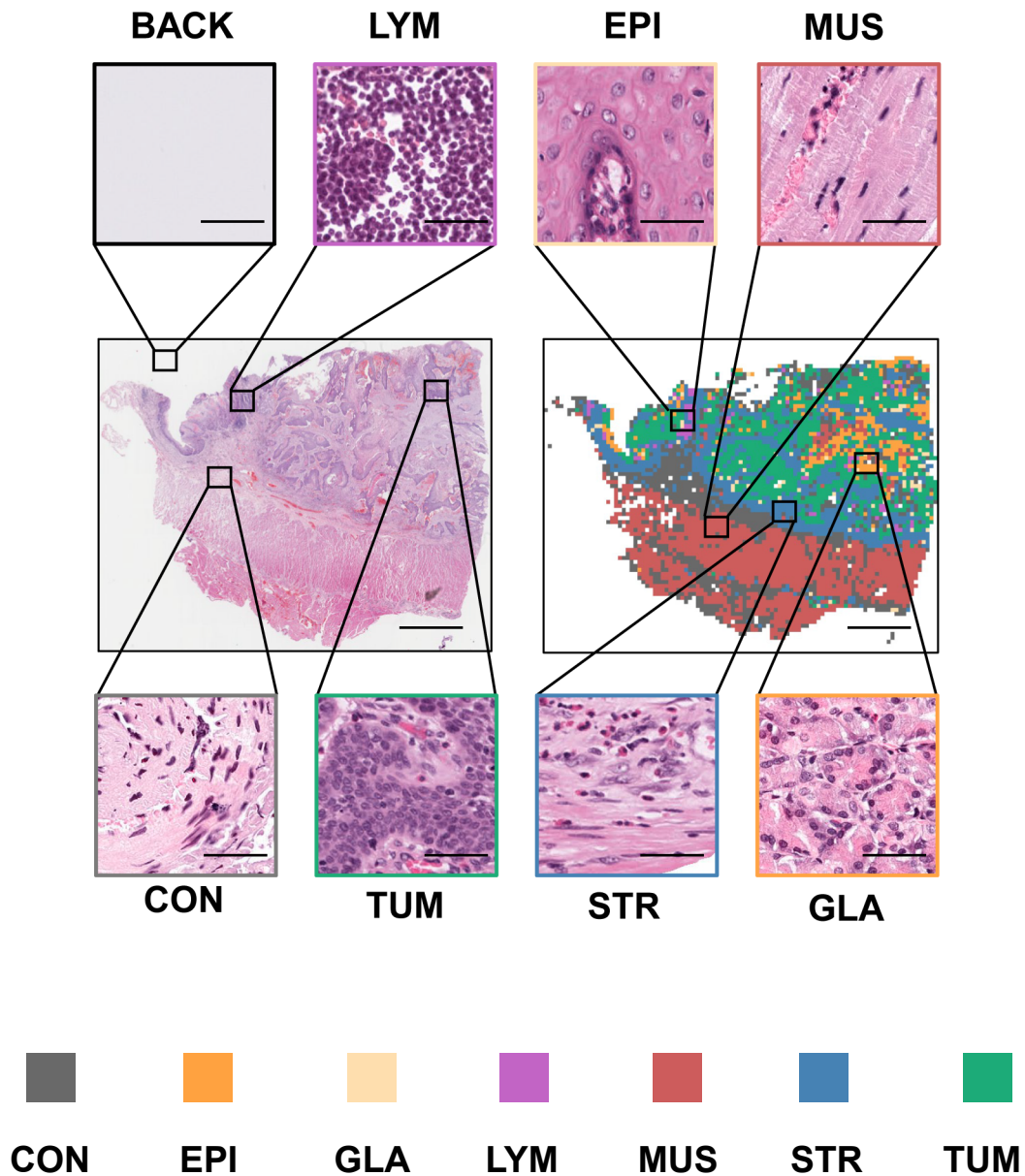
Supplementary Fig. 13. Potential drug target analysis. (a) Boxplots showing the expression of drug targets distribution across the four ECMSs. (b) Scatterplot showing the correlation between drug sensitivity AUC and FGFR2 expression. (c) Barplot showing the proportion of ERBB2 amplification between ECMS2 and non-ECMS2. (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon signed-rank tests)

a**GSE45670 (n=28)****b****ICB cohort (n=39)**

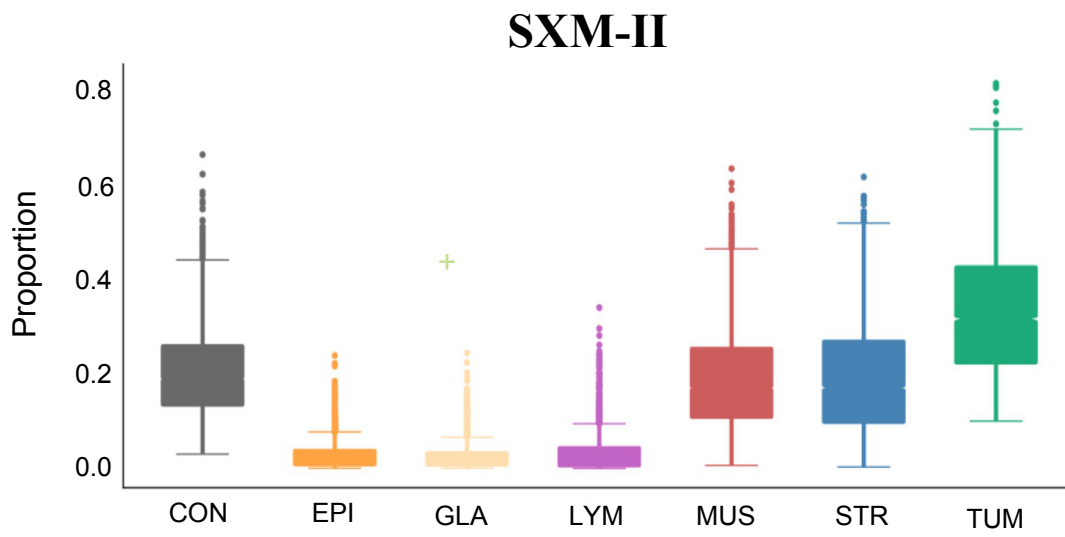
Supplementary Fig. 14. Associations between ECMS and (a) neo-chemoradiotherapy and (b) immunotherapy. pCR: pathological complete response. <pCR, less than pCR; PR, partial response; SD, stable disease; PD, progressive disease.



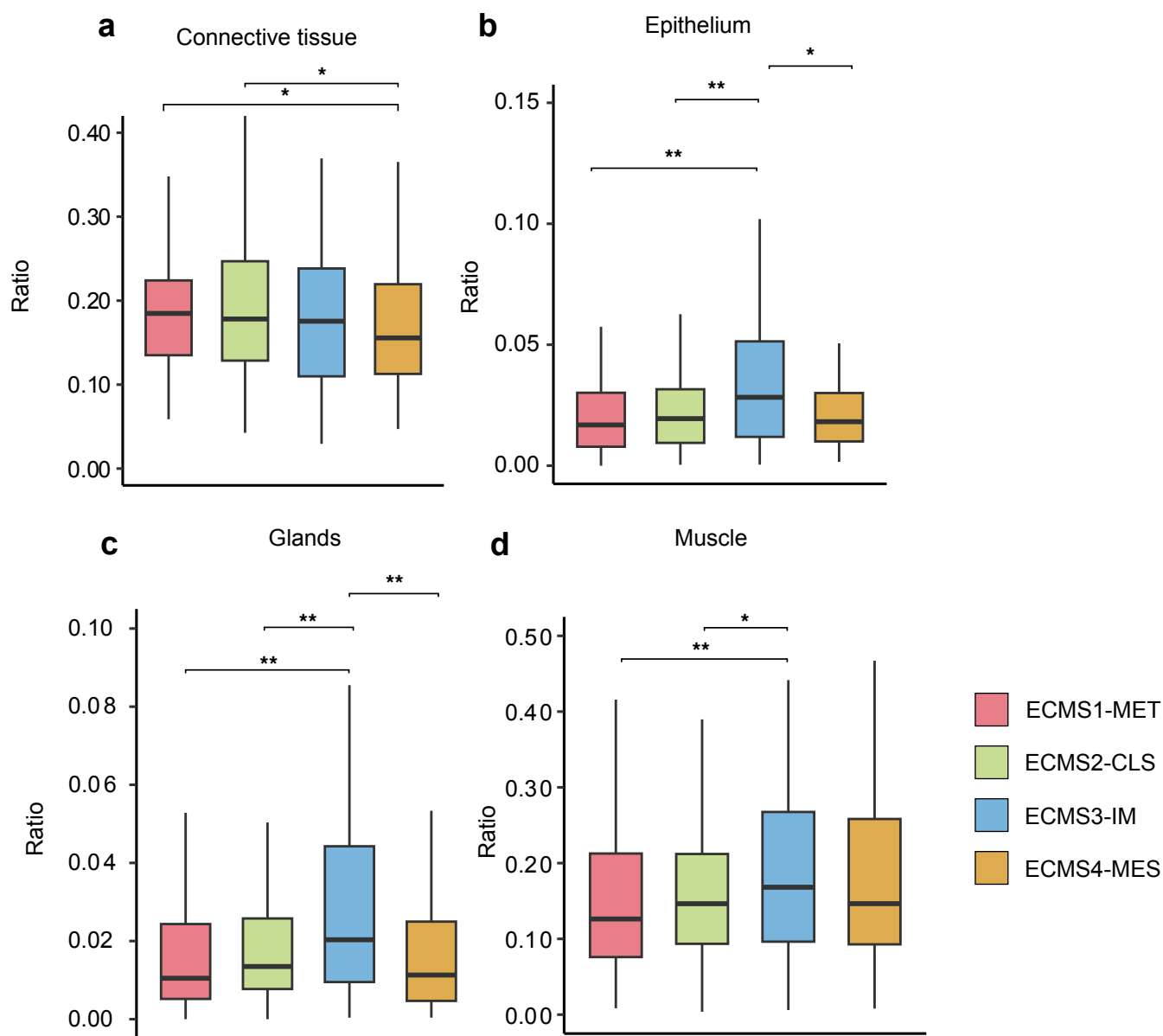
Supplementary Fig. 15. Evaluation of the performance of tissue classification based on a trained deep CNN. (a) Heatmap showing the confusion matrix between ESCC-SPA prediction and the delineation by pathologists. **(b)** ROC curves showing the prediction performance for different tissue types.



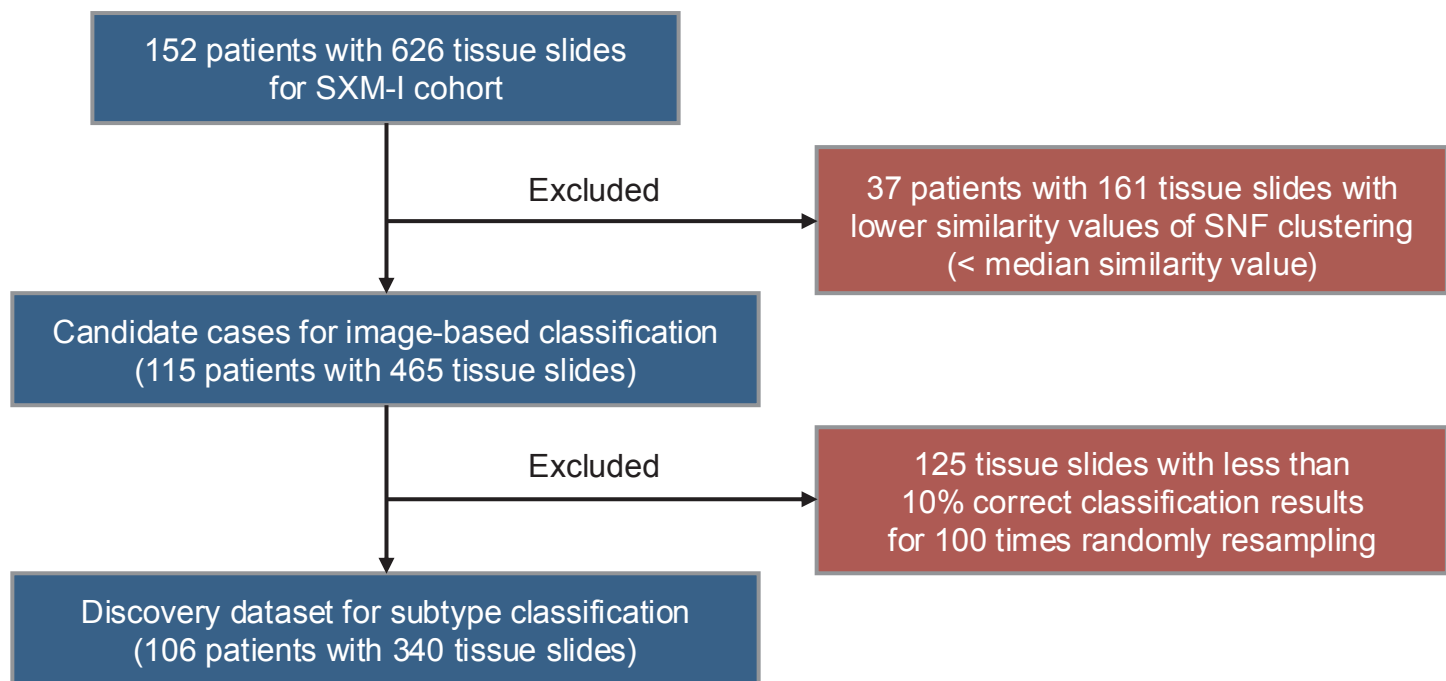
Supplementary Fig. 16. Representative examples illustrating automated delineation of H&E images based on tissue classification (Con, Connective; Epi, Epithelial; Mus, Muscle; Str, Stroma; Tum, Tumor; BACK, background; LYM, lymphocytes; STR, stroma; GLA, gland. The scale bar for the main overview image represents 2 mm, while the scale bar for the zoomed-in detailed inset represents 0.25 mm).



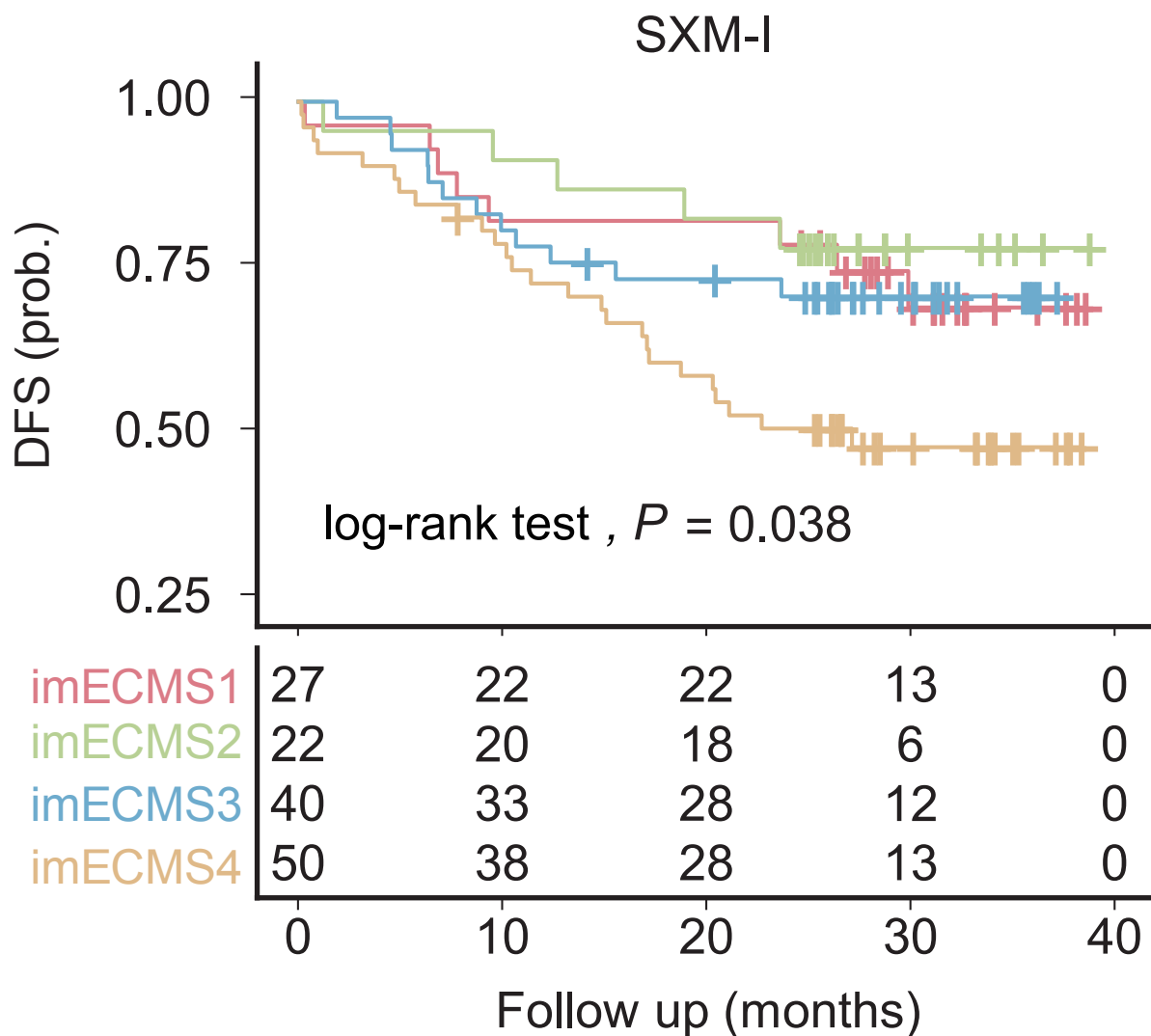
Supplementary Fig. 17. The distributions of relative abundance of different tissue types in the SXM-II cohort. (Con, Connective; Epi, Epithelial; Lym, Lymphocytes; Mus, Muscle; Str, Stroma; Tum, Tumor)



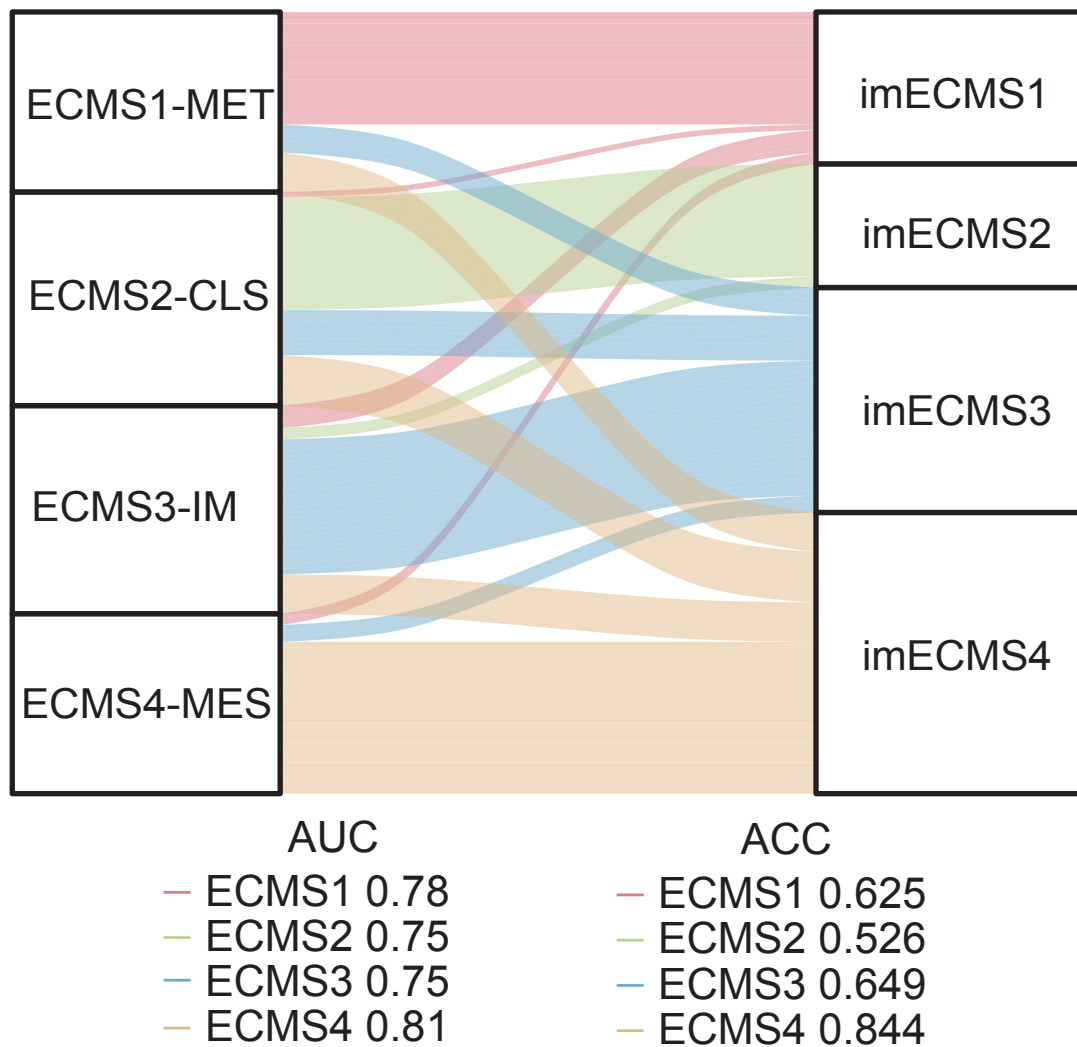
Supplementary Fig. 18. Boxplots comparing the distributions of whole-slide SOFs between ECMS subtypes. Boxplots comparing the relative abundance of (a) connective tissue, (b) normal epithelium, (c) glands and (d) muscle tissue between the four ECMSs. (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon signed-rank tests)



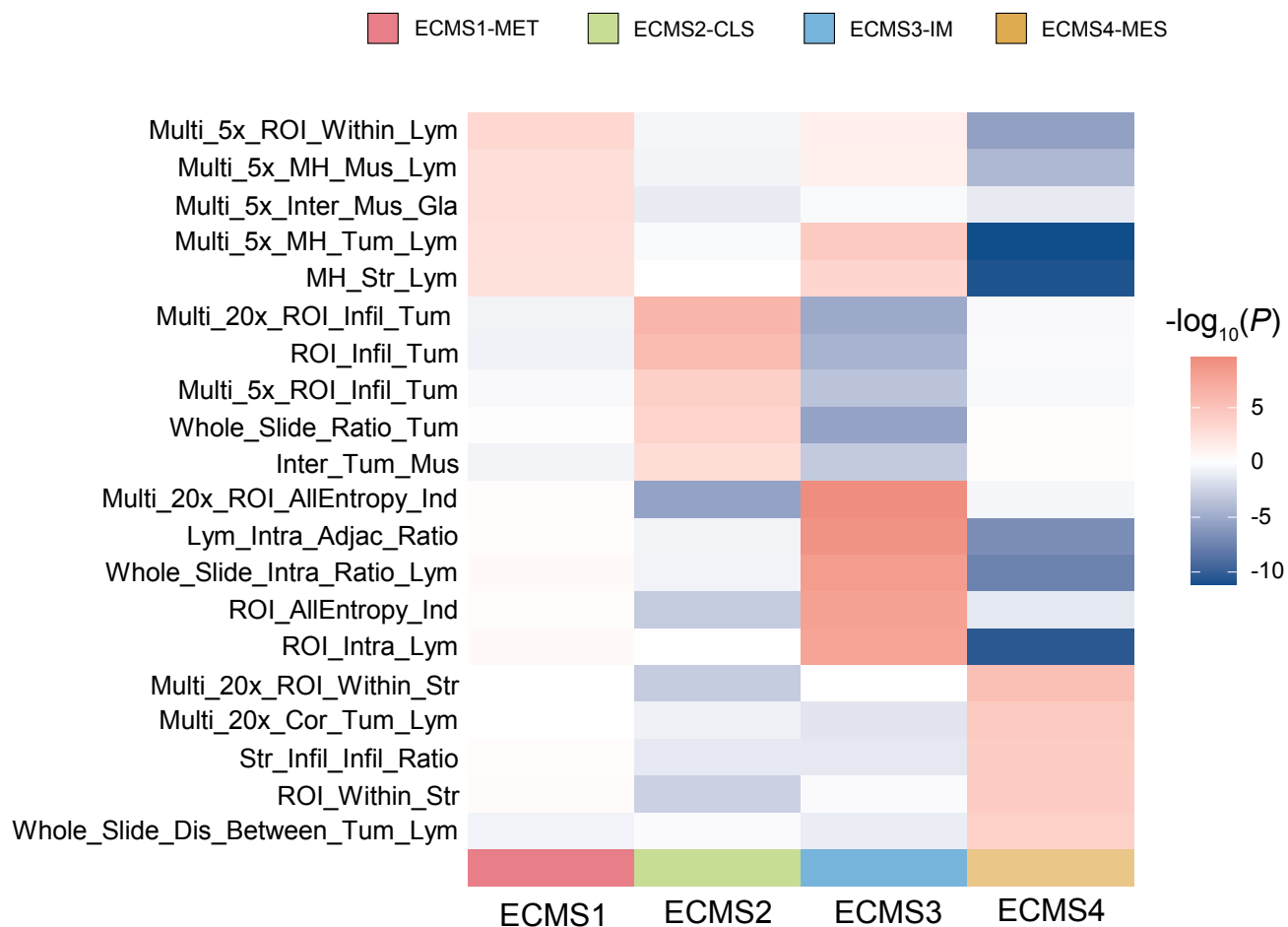
Supplementary Fig. 19. A flowchart of sample inclusion and exclusion criteria for sub-type classification in the discovery cohort



Supplementary Fig. 20. ECMS survival analysis of SXM-I. KM plot showing associations between predicted subtypes using the image-based classifier with disease-free survival in the SXM-I cohort.



Supplementary Fig. 21. Consistency between imECMS classification and ECMS classification
 Patient-level alluvial plot comparing molecular and image-based subtype predictions in the SXM-I cohort. AUC, area under the curve; ACC, accuracy.



Supplementary Fig. 22. ECMS SOFs. Heatmap of subtype-specific SOFs in the SXM-I cohort

Supplementary Data (Separate file):

Supplementary Data 1. The seven published ESCC subtyping systems

Supplementary Data 2. Clinical characteristics of patients with ESCC

Supplementary Data 3. ECMS classification results for the SXM-I, GSE53625, ICB, GSE45670 and TCGA cohorts

Supplementary Data 4. Somatic mutations identified for each ECMS. Chi-squared test was performed by two-by-two contingency table (one subtype vs others) with yellow denoting $P < 0.1$ and red $P < 0.05$.

Supplementary Data 5. Gene copy number variations were identified for each ECMS. Chi-squared test was performed by two-by-two contingency table (one subtype vs others) with sky blue denoting under-representation and red over-representation.

Supplementary Data 6. Differential mRNA expression levels across ECMS groups. Sky blue denotes significant low expression and red denotes significant high expression.

Supplementary Data 7. Differential gene promoter methylation levels across ECMS groups. Sky blue denotes significant hypomethylation and red denotes significant hypermethylation.

Supplementary Data 8. Differential miRNA expression levels across ECMS groups. Sky blue denotes significant low expression and red denotes significant high expression.

Supplementary Data 9. Potential drug for subtype specific gene. Genes were identified by expression and amplification analysis in Supplementary Fig. 13.

Supplementary Data 10. The number of image patches of the eight tissue types in the training and testing datasets

Supplementary Data 11. The number of unique spatial organization features for each subtype

Supplementary Data 12. The spatial organization features of all slides for all patients in

the SXM-I cohort

Supplementary Data 13. Wilcox signed-rank test for subtype-specific SOFs in the SXM-I cohort

Supplementary Data 14. The spatial organization features of all slides for all patients in the SXM-II cohort

Supplementary Data 15. The spatial organization features of all slides for all patients in the SXM-III cohort