Supplementary figures

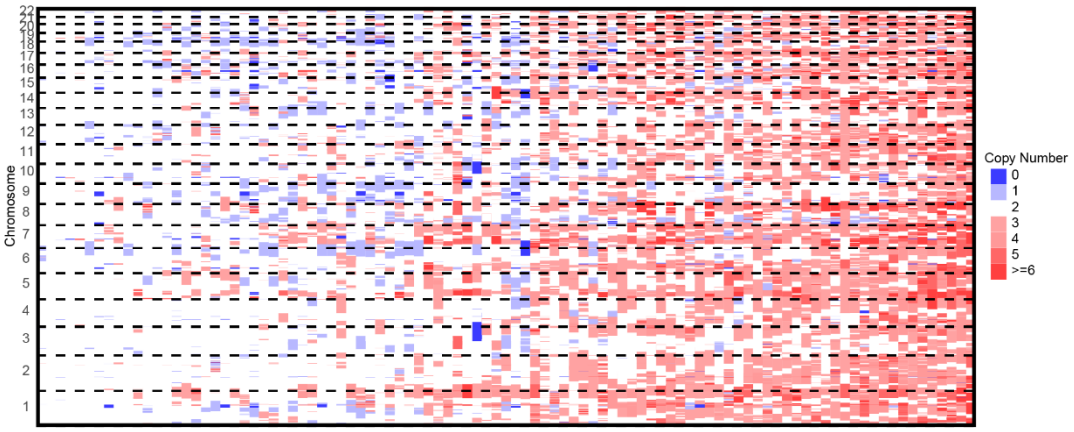

Figure S1. Focal level CNV in 97 LUAD patients in the HG cohort.
Heatmap showing the total segmented copy-number profile across all the patients in the HG cohort. The chromosomes are arranged vertically from bottom to top and patient samples are arranged from left to right according to FGA. Red and blue represent gain and loss, respectively.

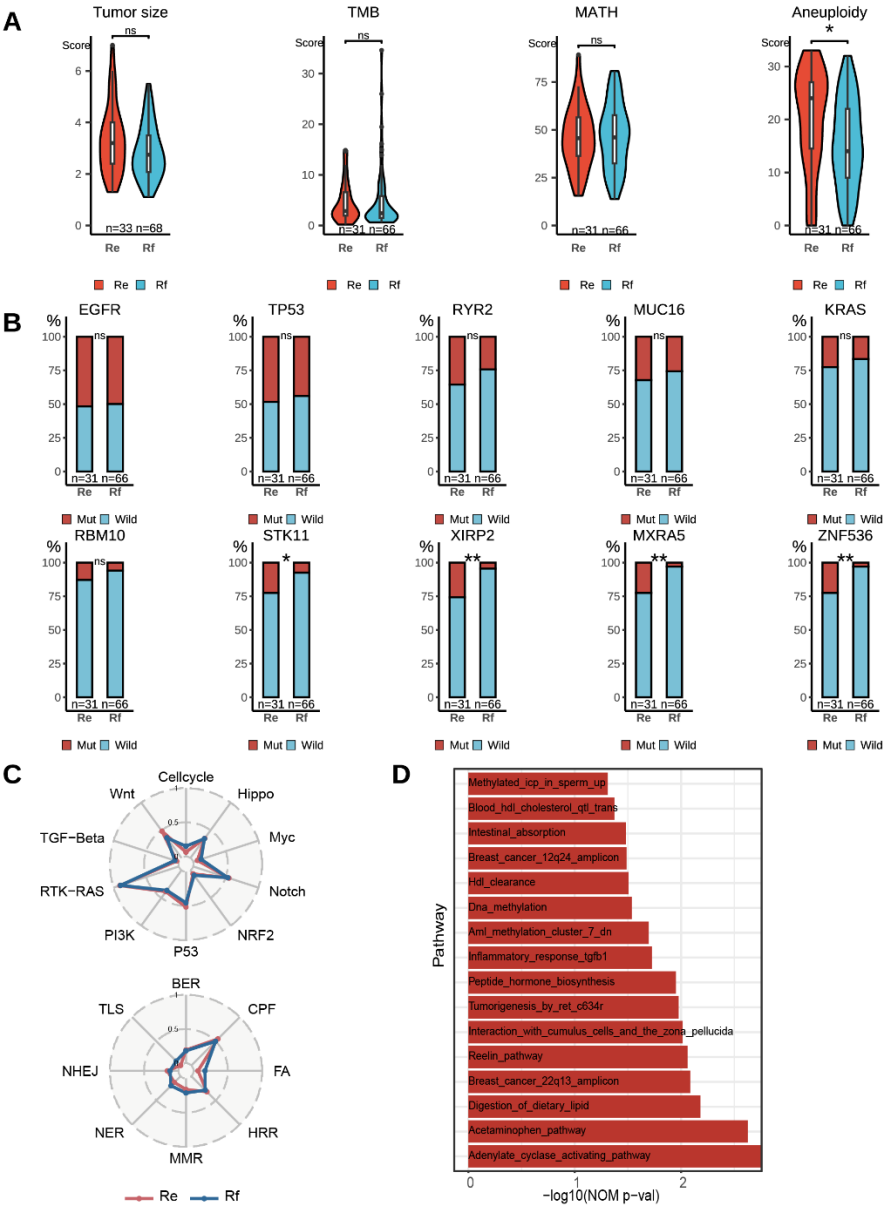

Figure S2. Comparison of clinical and molecular features between relapse (Re) and recurrence-free (Rf) patients.

(A) Comparison of clinical (tumor size) and molecular features (TMB, MATH and aneuploidy) between recurrence (Re) and recurrence-free (Rf) patients.

(B) Comparison of mutational frequencies of top genes in the HG cohort between Re and Rf patients.

(C) Comparison of the Re (red) and Rf patients (blue) in the WES based-ten canonical oncogenic signaling pathways (top) and DNA damage repair (DDR) pathways (bottom). The scale represents the percentage of the cohort with alterations in the respective pathways. A tumor is considered 'altered' in a specific pathway when ⩾1 gene is altered in the corresponding pathway.

(D) Significantly enriched pathways in the Re (red) patients determined by the gene set enrichment analysis (GSEA). Molecular Signatures Database (MSigDB) of curated gene sets (C2) was used for enrichment analysis. A nominal p-value < 1% was used as cutoff.

Note: To calculate p-values, fisher's test was utilized for categorical variables, and Wilcoxon test for continuous variables. ns: p >0.05 (not marked on the figure C), *: 0.01 < p <= 0.05, **: 0.001 < p <= 0.01, ***: p <= 0.001.

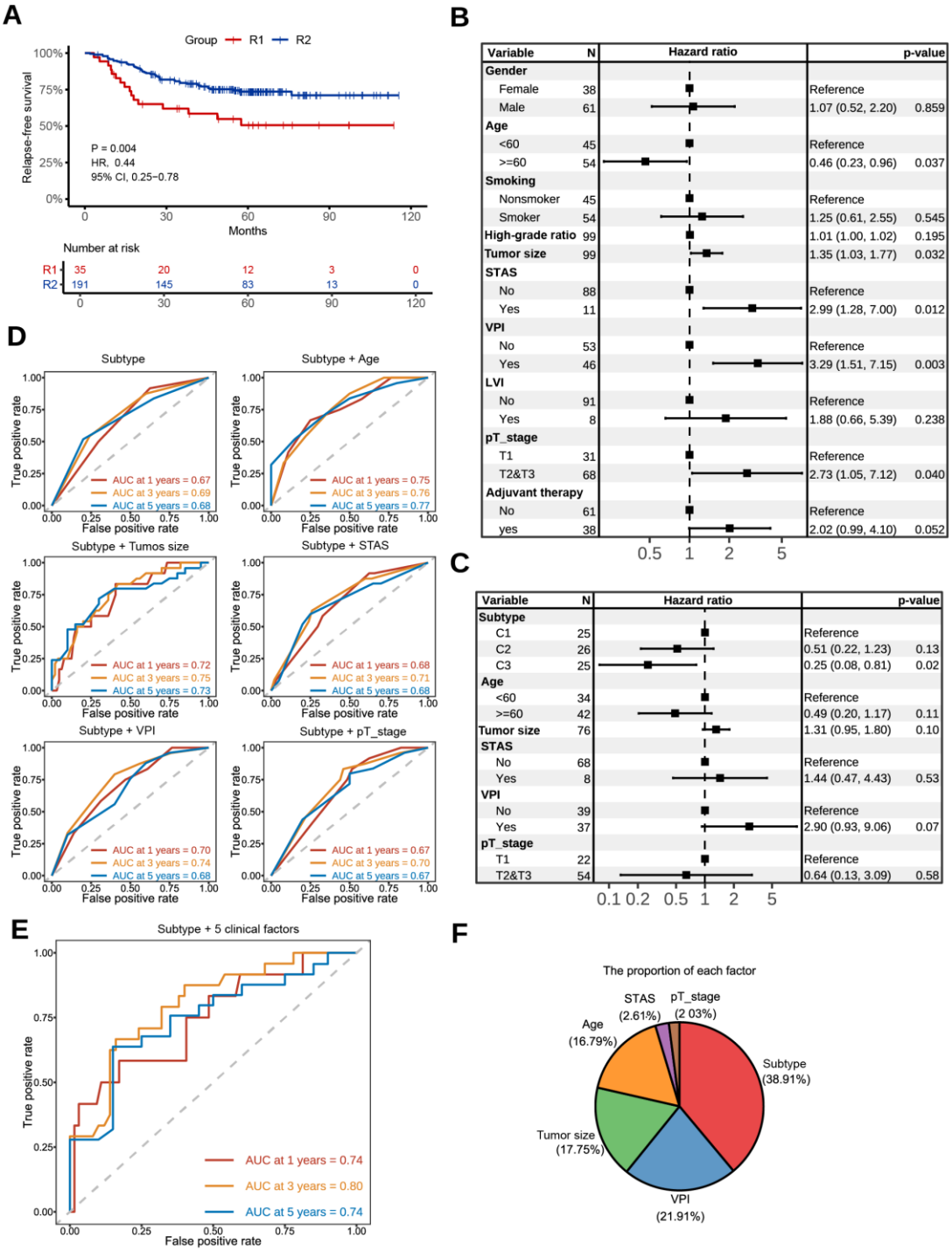

Figure S3. Predictive performance of the molecular subtypes.
(A) Validation of the single-sample transcriptomic classifier in an independent cohort (GSE31210), with Kaplan-Meier survival curves comparing RFS for RNA-based subtypes (R1 and R2).
(B) Univariate Cox regression analysis of clinical parameters, including age, gender, smoking, High-grade pattern ratio, tumor size, STAS, VPI, LVI, pT stage, and receipt of adjuvant treatment. Hazard ratios, 95% confidence interval (CI), and p-values are indicated.
(C) Multivariate Cox regression analysis of molecular subtypes combined with all five clinical parameters. Hazard ratios, 95% confidence interval (CI) and p-values are indicated.

(D) Area under curve (AUC) analysis for molecular subtypes alone or in combination with each of the five significant clinical parameters (age, tumor size, STAS, VPI, and pT stage) to predict 1-, 3-, and 5-year disease recurrence risk.
(E) AUC analysis to show the combined model including molecular subtypes and all five clinical factors in predicting 1-, 3-, and 5-year disease recurrence risks.
(F) Pie chart showing the contribution of each variant in the combined model.

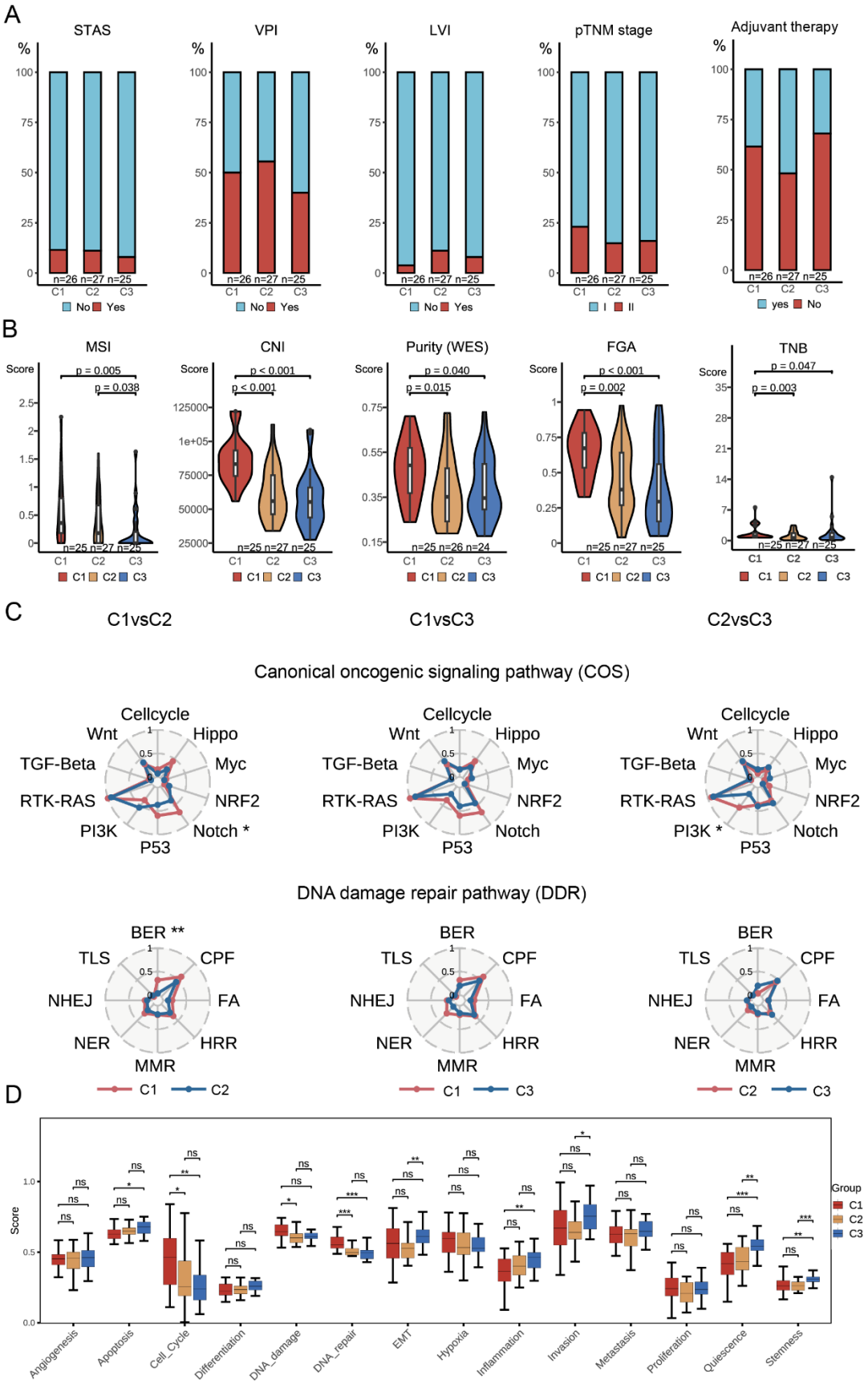
A
STAS
VPI
LVI
pTNM stage
Adjuvant therapy
%
n=26 n=27 n=25
C1 C2 C3
No Yes
I II
yes No
B
MSI
CNI
Purity (WES)
FGA
TNB
Score
p = 0.005
p = 0.038
p < 0.001
p < 0.001
p = 0.040
p = 0.015
p < 0.001
p = 0.002
p = 0.047
p = 0.003
n=25 n=27 n=25
n=25 n=26 n=24
C1 C2 C3
C
C1vsC2
C1vsC3
C2vsC3
Canonical oncogenic signaling pathway (COS)
Cellcycle
Hippo
Myc
NRF2
Notch *
P53
PI3K
PI3K *
RTK-RAS
TGF-Beta
Wnt
DNA damage repair pathway (DDR)
BER **
BER
TLS
CPF
FA
HRR
MMR
NER
NHEJ
C1 C2
C1 C3
C2 C3
D
Score
Group
C1
C2
C3
Angiogenesis
Apoptosis
Cell_Cycle
Differentiation
DNA_damage
DNA_repair
EMT
Hypoxia
Inflammation
Invasion
Metastasis
Proliferation
Quiescence
Stemness
ns

Figure S4. Further comparison of the three molecular subtypes.
(A) Comparison of clinical parameters among three subtypes (C1, C2, and C3). Clinical parameters, including STAS, VPI, LVI, pTNM stage and receipt of adjuvant therapy, were analyzed.
(B) Comparison of Molecular features among three subtypes (C1, C2 and C3). Molecular features such as MSI, CNI, WES-based purity, FGA, and neoantigen burden were also assessed.
(C) Comparison of ten canonical oncogenic signaling (COS) pathways (top) and DNA damage repair (DDR) pathways (bottom) among the three subtypes (C1 vs C2, C2 vs C3 and C1 vs C3).
(D) The box plot showing statistical differences in 14 functional states from 25 cancer types (including stemness, invasion, metastasis, proliferation, EMT, angiogenesis, apoptosis, cell cycle, differentiation, DNA damage, DNA repair, hypoxia, inflammation and quiescence) among three subtypes C1, C2 and C3. The x-axis represents 14 functional states, and the y-axis represents the enrichment score of each functional state.
Note: To calculate p-values, fisher's test was utilized for categorical variables, and Wilcoxon test for continuous variables. ns: $p > 0.05$ (not marked on the figures A, B, and C), *: $0.01 < p <= 0.05$, **: $0.001 < p <= 0.01$, ***: $p <= 0.001$.

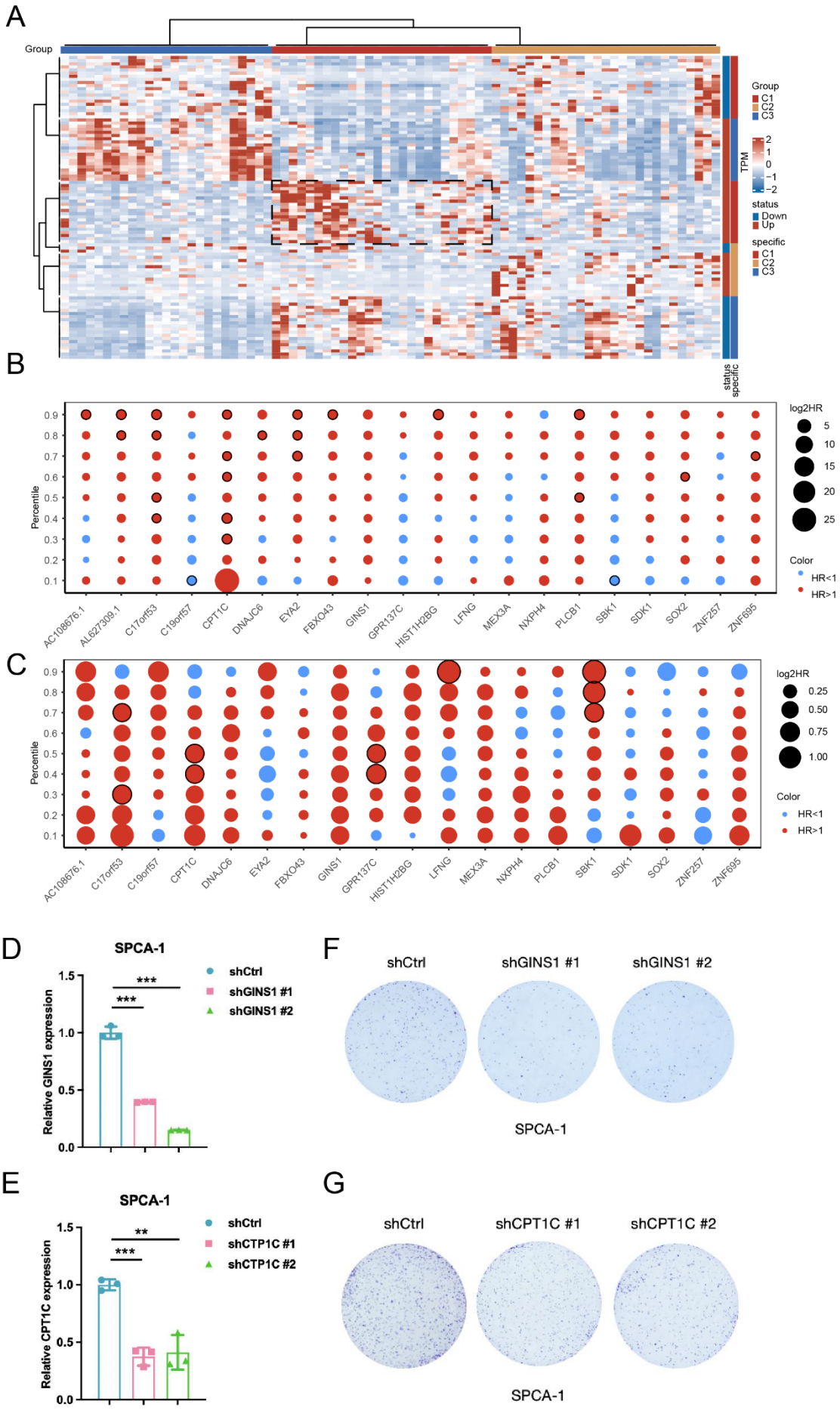
A
Group
Group
C1
C2
C3
TPM
2
1
0
−1
−2
status
Down
Up
specific
C1
C2
C3
status
specific
B
Percentile
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
log2HR
5
10
15
20
25
Color
HR<1
HR>1
AC108676.1
AL627309.1
C17orf53
C19orf57
CPT1C
DNAJC6
EYA2
FBXO43
GINS1
GPR137C
HIST1H2BG
LFNG
MEX3A
NXPH4
PLCB1
SBK1
SDK1
SOX2
ZNF257
ZNF695
C
Percentile
log2HR
0.25
0.50
0.75
1.00
Color
HR<1
HR>1
D
SPCA-1
Relative GINS1 expression
1.5
1.0
0.5
0.0
***
***
shCtrl
shGINS1 #1
shGINS1 #2
F
shCtrl
shGINS1 #1
shGINS1 #2
SPCA-1
E
SPCA-1
Relative CPT1C expression
1.5
1.0
0.5
0.0
**
***
shCtrl
shCTP1C #1
shCTP1C #2
G
shCtrl
shCPT1C #1
shCPT1C #2
SPCA-1

Figure S5. Representative differentially expressed genes (DEGs) in three Integrative subtypes and the biological functions of GINS1 and CPT1C.

(A) Heatmap of gene expression for representative DEGs in each of the three molecular subtypes (C1, C2, and C3) (detailed in the Methods). The box with black dash line highlighted those genes significantly up-regulated in the C1 subtypes comparing to the other two subtypes.

(B, C) Bubble plots showing hazard ratios (HR) and p-values of RFS/DFS across percentiles from 0.1 to 0.9 for genes in the HG cohort (B) and TCGA cohort (C). Dots indicate log(2) HR values, with blue representing HR < 1 and red representing HR > 1. Dots outlined in black indicate gene expression groups with significantly different RFS.

(D, E) The GINS1 (E) and CPT1C (F) knockdown efficiency was analyzed by qRT-PCR in SPCA-1 cells (n = 3).

(F, G) Respective images of Figure 5I and K.

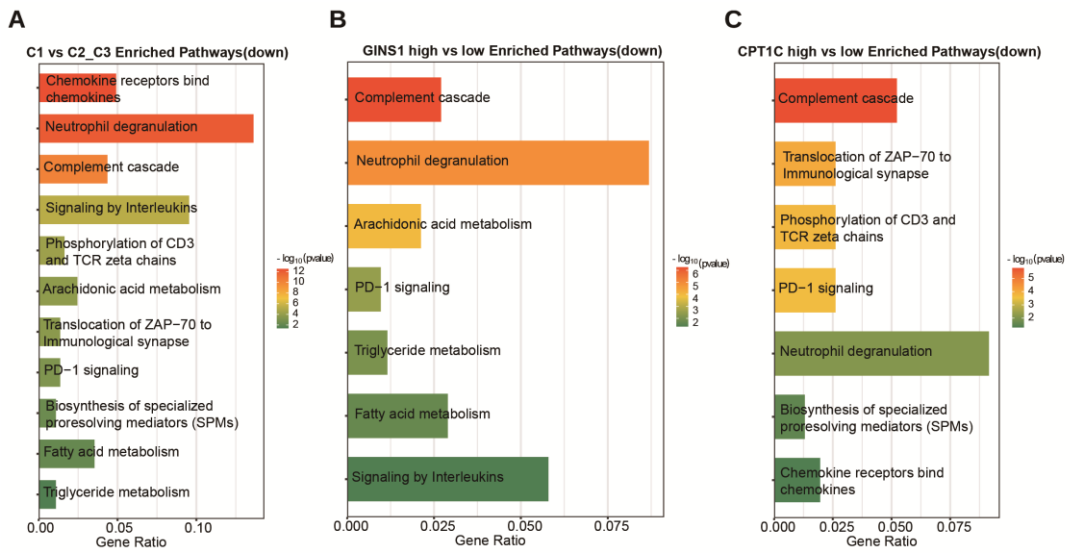

Figure S6. Pathway enrichment analysis for downregulated DEGs using the Reactome Pathway database.

(A) Pathways exhibiting significant differences between the C1 subtypes and the combined C2 and C3 subtypes.

(B) Pathways exhibiting significant differences between high and low GINS1 expression groups.

(C) Pathways exhibiting significant differences between high and low CPT1C expression groups.

Note: X-axis: GeneRatio (proportion of enriched genes), reflecting the degree of pathway enrichment. Y-axis: Pathway description, arranged in descending order of significance. Color: -log10(p-value), with a deeper red indicating higher significance.