Figure S1

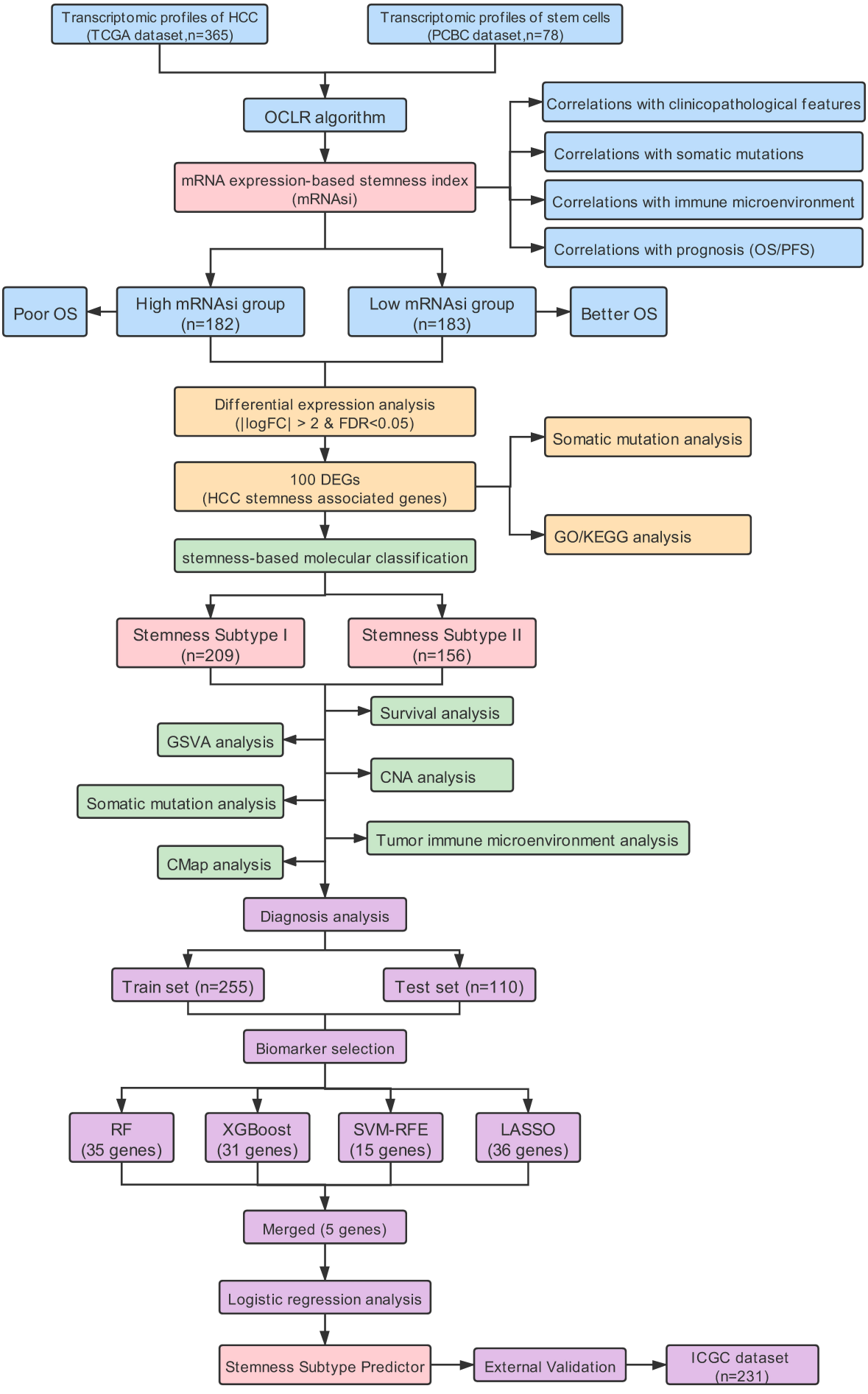
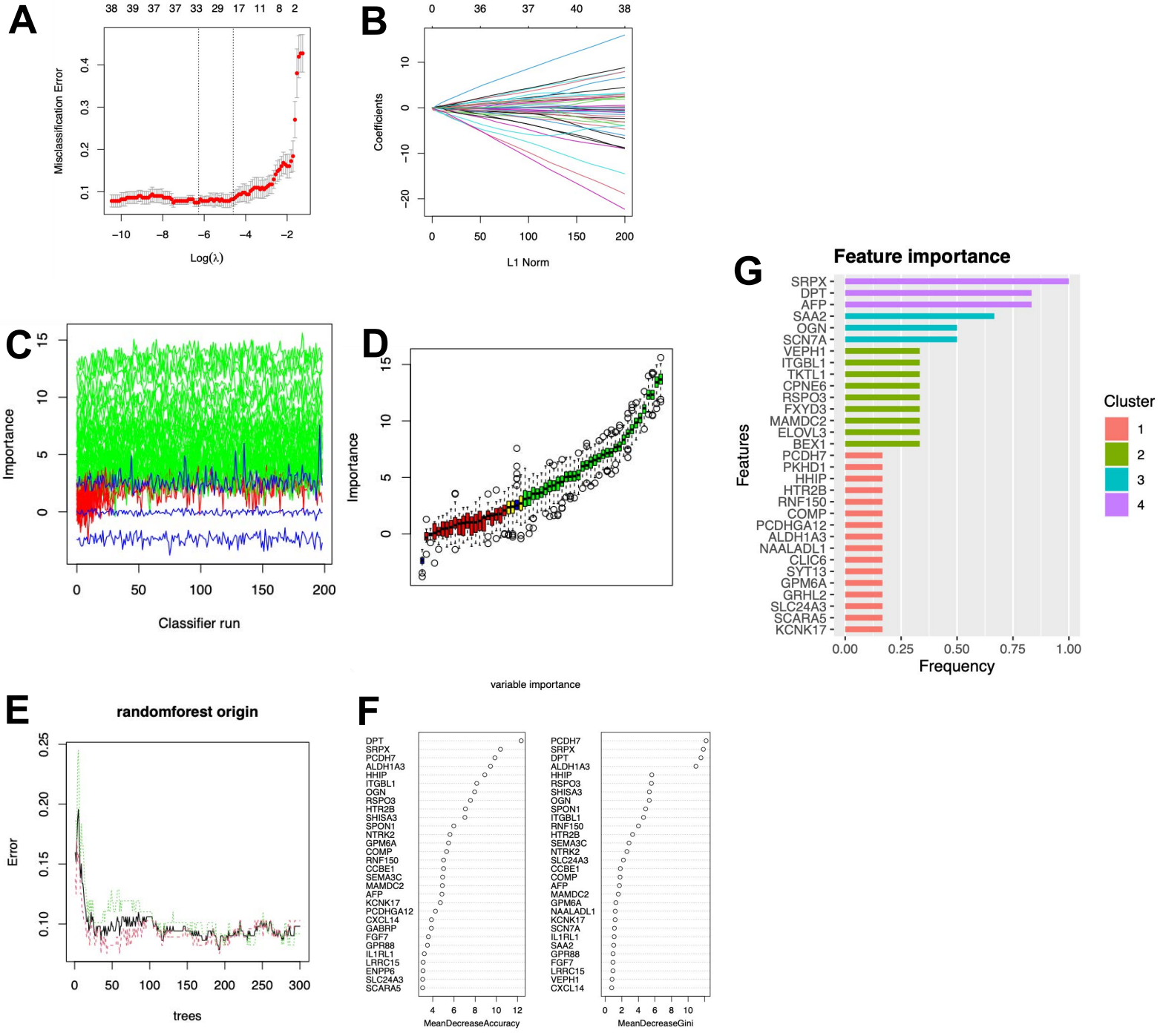

Transcriptomic profiles of HCC (TCGA dataset,n=365)
Transcriptomic profiles of stem cells (PCBC dataset,n=78)
OCLR algorithm
mRNA expression-based stemness index (mRNAsi)
Correlations with clinicopathological features
Correlations with somatic mutations
Correlations with immune microenvironment
Correlations with prognosis (OS/PFS)
Poor OS
High mRNAsi group (n=182)
Low mRNAsi group (n=183)
Better OS
Differential expression analysis (|logFC| > 2 & FDR<0.05)
100 DEGs (HCC stemness associated genes)
Somatic mutation analysis
GO/KEGG analysis
stemness-based molecular classification
Stemness Subtype I (n=209)
Stemness Subtype II (n=156)
Survival analysis
GSVA analysis
CNA analysis
Somatic mutation analysis
Tumor immune microenvironment analysis
CMap analysis
Diagnosis analysis
Train set (n=255)
Test set (n=110)
Biomarker selection
RF (35 genes)
XGBoost (31 genes)
SVM-RFE (15 genes)
LASSO (36 genes)
Merged (5 genes)
Logistic regression analysis
Stemness Subtype Predictor
External Validation
ICGC dataset (n=231)

Figure S2


A
38 39 37 37 33 29 17 11 8 2
Misclassification Error
Log(λ)
B
0 36 37 40 38
Coefficients
L1 Norm
C
Importance
Classifier run
D
Importance
variable importance
G
Feature importance
SRPX
DPT
AFP
SAA2
OGN
SCN7A
VEPH1
ITGBL1
TKTL1
CPNE6
RSPO3
FXYD3
MAMDC2
ELOVL3
BEX1
PCDH7
PKHD1
HHIP
HTR2B
RNF150
COMP
PCDHGA12
ALDH1A3
NAALADL1
CLIC6
SYT13
GPM6A
GRHL2
SLC24A3
SCARA5
KCNK17
Features
Frequency
Cluster
1
2
3
4
E
randomforest origin
Error
trees
F
DPT
SRPX
PCDH7
ALDH1A3
HHIP
ITGBL1
OGN
RSPO3
HTR2B
SHISA3
SPON1
NTRK2
GPM6A
COMP
RNF150
CCBE1
SEMA3C
MAMDC2
AFP
KCNK17
PCDHGA12
CXCL14
GABRP
FGF7
GPR88
IL1RL1
LRRC15
ENPP6
SLC24A3
SCARA5
MeanDecreaseAccuracy
PCDH7
SRPX
DPT
ALDH1A3
HHIP
RSPO3
SHISA3
OGN
SPON1
ITGBL1
RNF150
HTR2B
SEMA3C
NTRK2
SLC24A3
CCBE1
COMP
AFP
MAMDC2
GPM6A
NAALADL1
KCNK17
SCN7A
IL1RL1
SAA2
GPR88
FGF7
LRRC15
VEPH1
CXCL14
MeanDecreaseGini