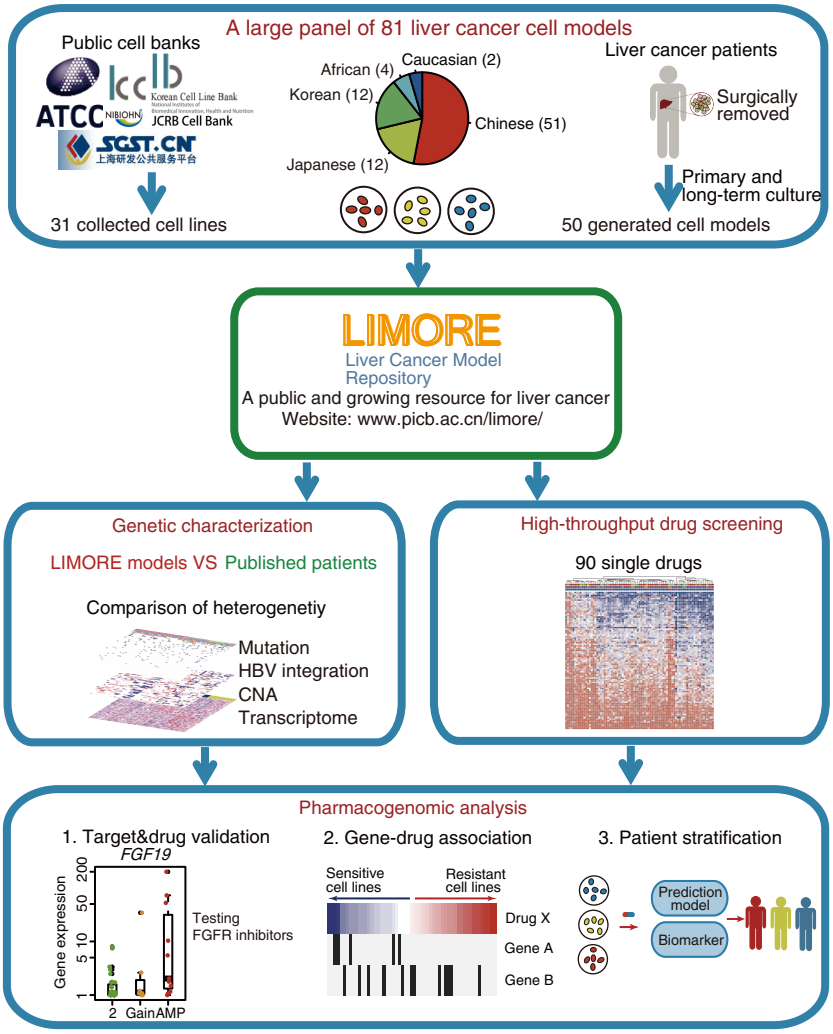
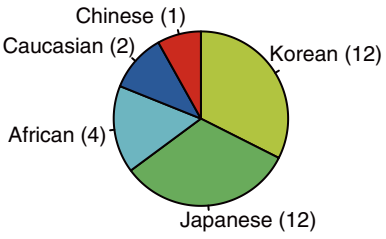
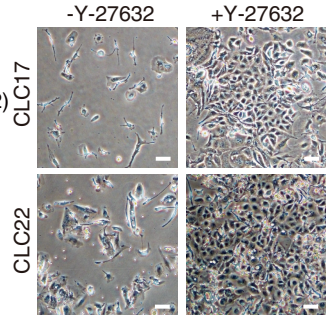
A


A large panel of 81 liver cancer cell models
Public cell banks
ATCC
Korean Cell Line Bank
JCRB Cell Bank
SCST.CN
31 collected cell lines
African (4)
Caucasian (2)
Korean (12)
Chinese (51)
Japanese (12)
Liver cancer patients
Surgically removed
Primary and long-term culture
50 generated cell models
LIMORE
Liver Cancer Model Repository
A public and growing resource for liver cancer
Website: www.picb.ac.cn/limore/
Genetic characterization
LIMORE models VS Published patients
Comparison of heterogenetiy
Mutation
HBV integration
CNA
Transcriptome
High-throughput drug screening
90 single drugs
Pharmacogenomic analysis
1. Target&drug validation
FGF19
Gene expression
Testing FGFR inhibitors
2 GainAMP
2. Gene-drug association
Sensitive cell lines
Resistant cell lines
Drug X
Gene A
Gene B
3. Patient stratification
Prediction model
Biomarker

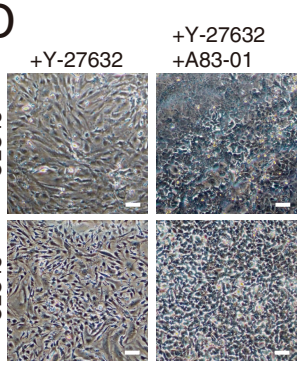
B


Chinese (1)
Caucasian (2)
Korean (12)
African (4)
Japanese (12)

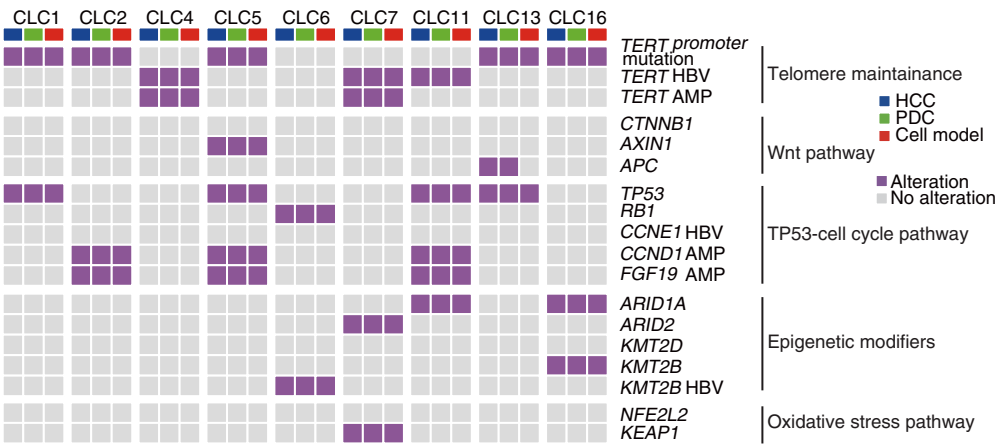
C


-Y-27632
+Y-27632
CLC17
CLC22

D


+Y-27632
+Y-27632
+A83-01
CLC43
CLC48

E


CLC1
CLC2
CLC4
CLC5
CLC6
CLC7
CLC11
CLC13
CLC16
TERT promoter mutation
TERT HBV
TERT AMP
CTNNB1
AXIN1
APC
TP53
RB1
CCNE1 HBV
CCND1 AMP
FGF19 AMP
ARID1A
ARID2
KMT2D
KMT2B
KMT2B HBV
NFE2L2
KEAP1
Telomere maintainance
Wnt pathway
TP53-cell cycle pathway
Epigenetic modifiers
Oxidative stress pathway
HCC
PDC
Cell model
Alteration
No alteration

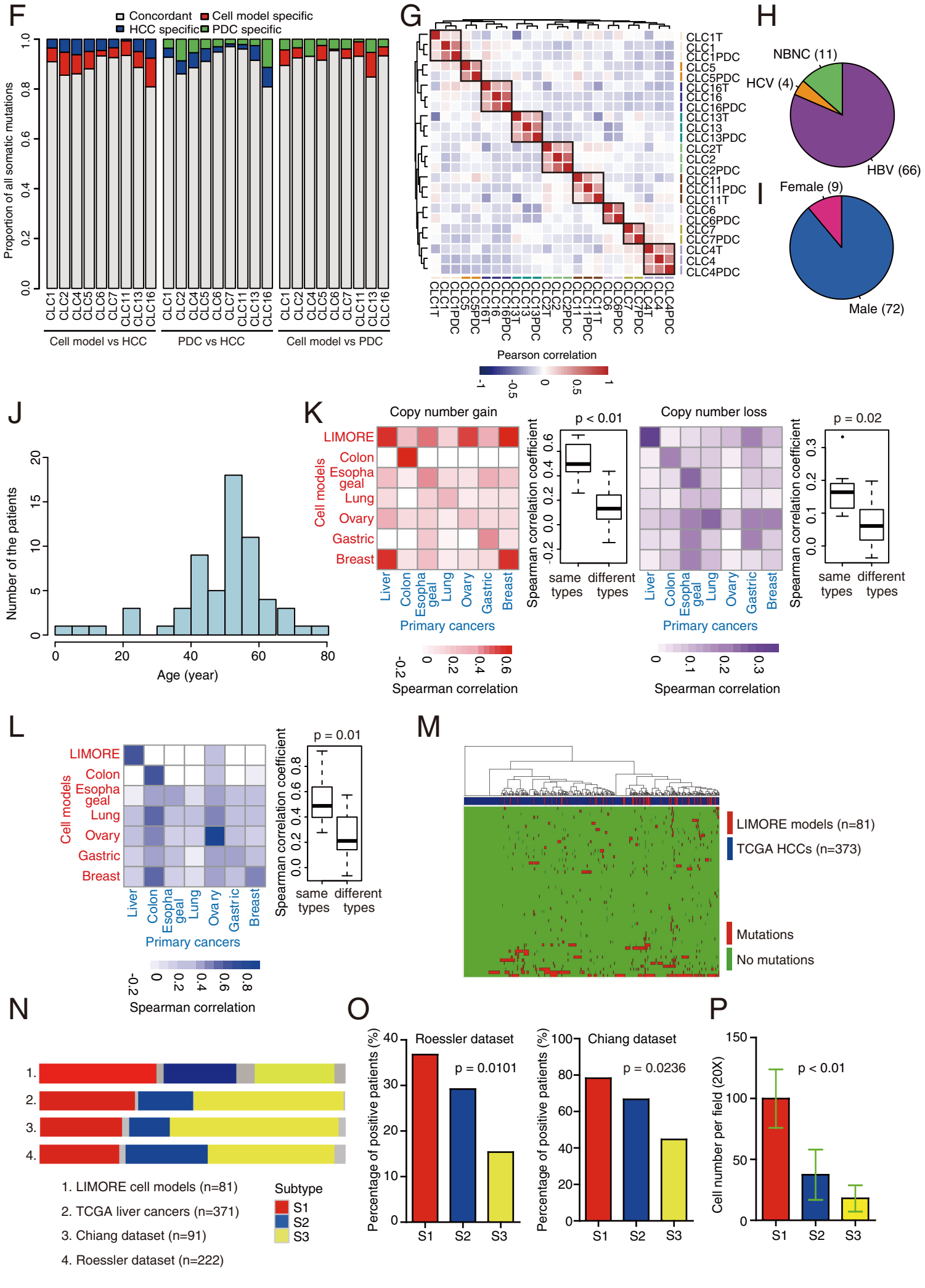

F
Concordant
Cell model specific
HCC specific
PDC specific
Proportion of all somatic mutations
Cell model vs HCC
PDC vs HCC
Cell model vs PDC
G
Pearson correlation
H
NBNC (11)
HCV (4)
HBV (66)
I
Female (9)
Male (72)
J
Number of the patients
Age (year)
K
Copy number gain
Copy number loss
Cell models
LIMORE
Colon
Esophageal
Lung
Ovary
Gastric
Breast
Primary cancers
Spearman correlation
Spearman correlation coefficient
p < 0.01
p = 0.02
same types
different types
L
p = 0.01
M
LIMORE models (n=81)
TCGA HCCs (n=373)
Mutations
No mutations
N
1. LIMORE cell models (n=81)
2. TCGA liver cancers (n=371)
3. Chiang dataset (n=91)
4. Roessler dataset (n=222)
Subtype
S1
S2
S3
O
Roessler dataset
p = 0.0101
Chiang dataset
p = 0.0236
Percentage of positive patients (%)
P
p < 0.01
Cell number per field (20X)

Figure S1. Overview of LIMORE platform and epidemiological and intertumor heterogeneities between cell models and primary cancers, Related to Figure 1.

(A) Schematic outline for Liver Cancer Model Repository (LIMORE). Current version of LIMORE consists of 81 liver cancer cell models, including 31 collected publically available cell models, and 50 generated models from surgically resected Chinese HCCs. A web platform is provided to help the community to use LIMORE (www.picb.ac.cn/limore/ or http://limore.sibcb.ac.cn/). All cell models have been extensively characterized using next-generation sequencing and high-throughput drug screening. Using these datasets, pharmocogenomic landscape was constructed and interrogated to explore gene-drug associations, including biomarkers to improve sorafenib response. CNA, copy number alteration. AMP, amplification.

(B) Pieplot shows population distribution of 31 collected publically available liver cancer cell models.

(C) Representative pictures of early passages of patient-derived cells grown in media with or without Y-27632. Epithelial cells proliferated in media with Y-27632. Scale bars, 100 μm.

(D) Representative pictures of early passages of patient-derived cells grown in media with Y-27632 or with both Y-27632 and A83-01. Epithelial cells were enriched in media with both Y-27632 and A83-01, while cancer associated fibroblast overgrew in media with Y-27632 alone. Scale bars, 100 μm.

(E) Heatmap shows landscapes of genetic alterations for HCC driver genes in the matched patient-derived cells (PDCs) at early-stage passages (< 10 passages, PDC), established cell models (around 20 passages) and primary HCCs. AMP, amplification. HBV, HBV integration.

(F) Paired comparison of the matched cell models, primary HCCs and PDCs using protein-altering somatic mutations.

(G) Heatmap shows Pearson correlation coefficient matrix of gene expression in matched cell

models, primary HCCs and PDCs. Please be noted, that genetic data (E and F) and transcriptomic data (G) for 4 cell models (CLC5, CLC11, CLC13 and CLC16) were adapted from previously published study (Qiu et al., 2016).

(H-J) Distributions of virus infection status (H), gender (I) and age (J) of liver cancer patients which all the 81 liver cancer cell models were derived from. NBNC, non-HBV and non-HCV.

(K and L) Spearman correlations of CNAs (K) and somatic mutations (L) between cell models and primary cancers from 7 types of cancers. Boxplot shows the coefficients between cell models and primary cancers from the same types or different types. For box-and-whisker plot, the box indicates IQR, the line in the box indicates the median, the whiskers indicate points within Q3+1.5×IQR and Q1−1.5×IQR and the points beyond whiskers indicate outliers. Statistical significance was determined by unpaired Student's t-test.

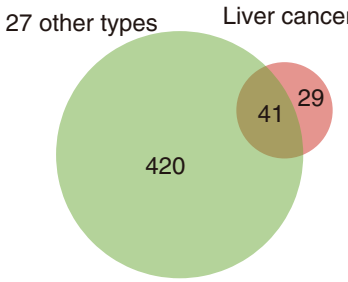
(M) Hierarchical clustering analysis of whole exome somatic mutations at the gene level between LIMORE (n=81) and TCGA liver cancers (n=373) using Ward algorithm.

(N) Molecular transcriptome classification of LIMORE models and 3 published cohorts of primary liver cancers using a HCC classification system. The red, blue and yellow show that the samples are classified into S1, S2 and S3 subtype with high confidence (FDR < 0.25), respectively. The grey indicates the classification of the samples with FDR > 0.25. FDR, false discovery rate.

(O) Venous metastasis (left) and vascular invasion (right) of primary liver cancers from different transcriptome subtypes. Statistical significance was determined by Chi-square test.

(P) Migration abilities of LIMORE models from different transcriptome subtypes, as determined by *in vitro* Transwell assay. Data are presented as mean±SEM. Statistical significance was determined by Kruskal–Wallis test.

A

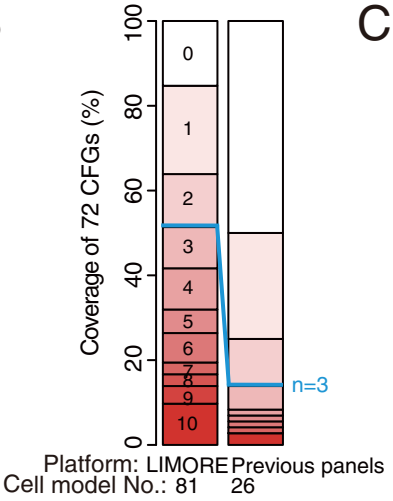


Other types:
FBXW7
KRAS
BRCA1
CHEK2
MAP4K3
EGFR
ERBB2
RUNX1
IDH2
SMO
MYC
ALK
...

Shared:
TP53
CTNNB1
CDKN2A
CDKN2B
ARID1A
ARID2
PTEN
RB1
APC
AXIN1
NFE2L2
...

Liver cancer:
ALB
G6PC
RPS6KA3
HNF4A
...

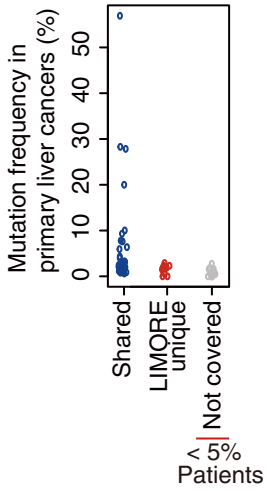
B



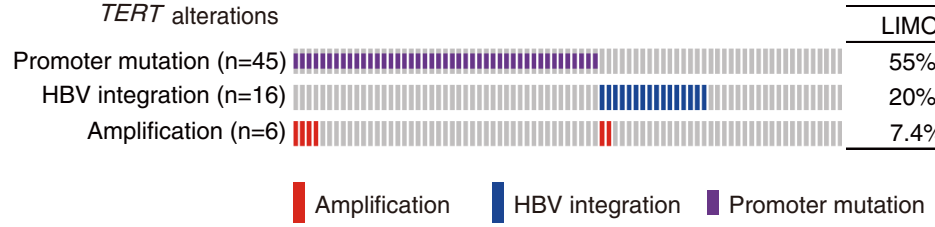
C

Mutations	LIMORE	Previous panels
CTNNB1 activating	8	3
TERT HBV	16	3
RB1	5	1
PIK3CA	2	0
RPS6KA3	2	0

D

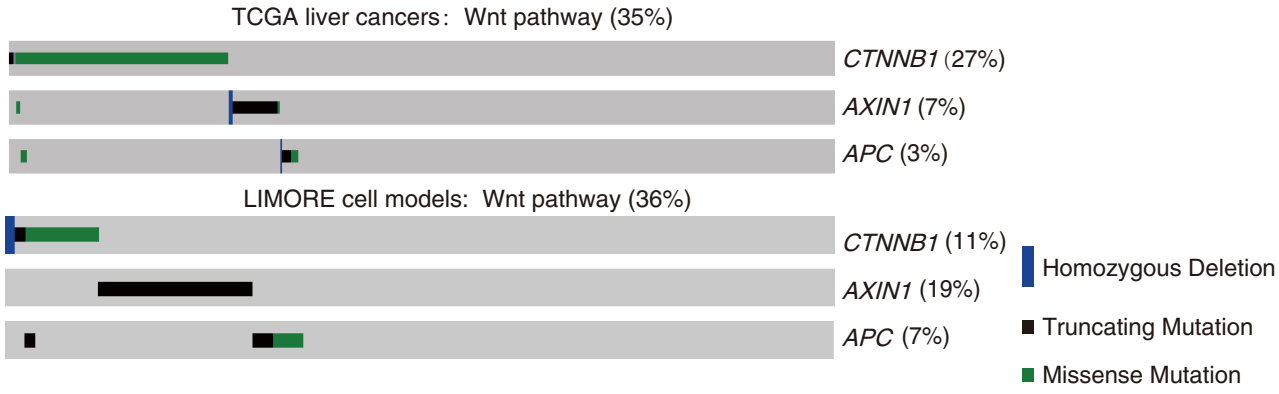


E



	LIMORE	HCC patients
Promoter mutation (n=45)	55%	43%
HBV integration (n=16)	20%	22%
Amplification (n=6)	7.4%	7.6%

F



G

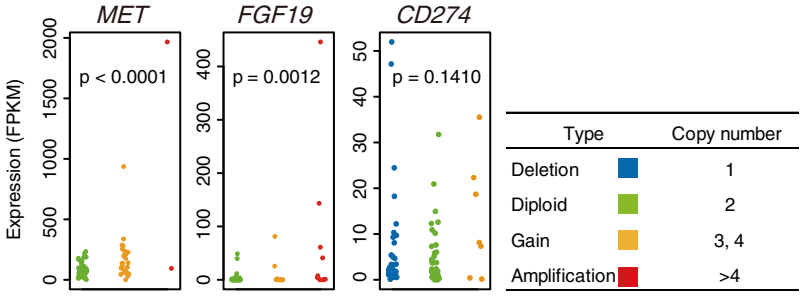


Figure S2. Comparison of cancer functional genes, Related to Figure 2.

(A) Venn diagram shows comparison of mutational cancer functional genes (CFGs) between LIMORE and reported IntOGen pan-cancer database (www.intogen.org). Highlighted are representative recurrently mutational genes specifically in liver cancers or 27 other types of cancers.

(B) Coverage of CFG alterations by liver cancer models from LIMORE or previous panels. The number in the bar indicates how many cell models harbor alterations of a CFG.

(C) Representation of key CFGs by liver cancer models in LIMORE and previous cell panels.

(D) Mutational frequencies of CFGs in primary liver cancer cohorts with mutational data available.

(E) Patterns of multiple types of alterations in *TERT* gene in LIMORE (left) and the comparison of mutational frequencies to primary liver cancers (right).

(F) Oncoprint plot of Wnt-related alterations (*CTNNB1*, *AXIN1* and *APC*) in TCGA liver cancers (n=373) and LIMORE (n=81). Truncating mutations included nonsense, frameshift and splicing mutations. The alteration frequencies are indicated. The visualization was performed using Oncoprint tool in cBioPortal (www.cbioportal.org).

(G) Evaluation of potential targets in liver cancers and their mRNA expression levels, including *FGF19* or *MET* amplifications. Gene expression levels in RNA-Seq were presented as FPKM (fragments per kilobase of exon per million fragments mapped). Statistical significance was determined by Kruskal–Wallis test.

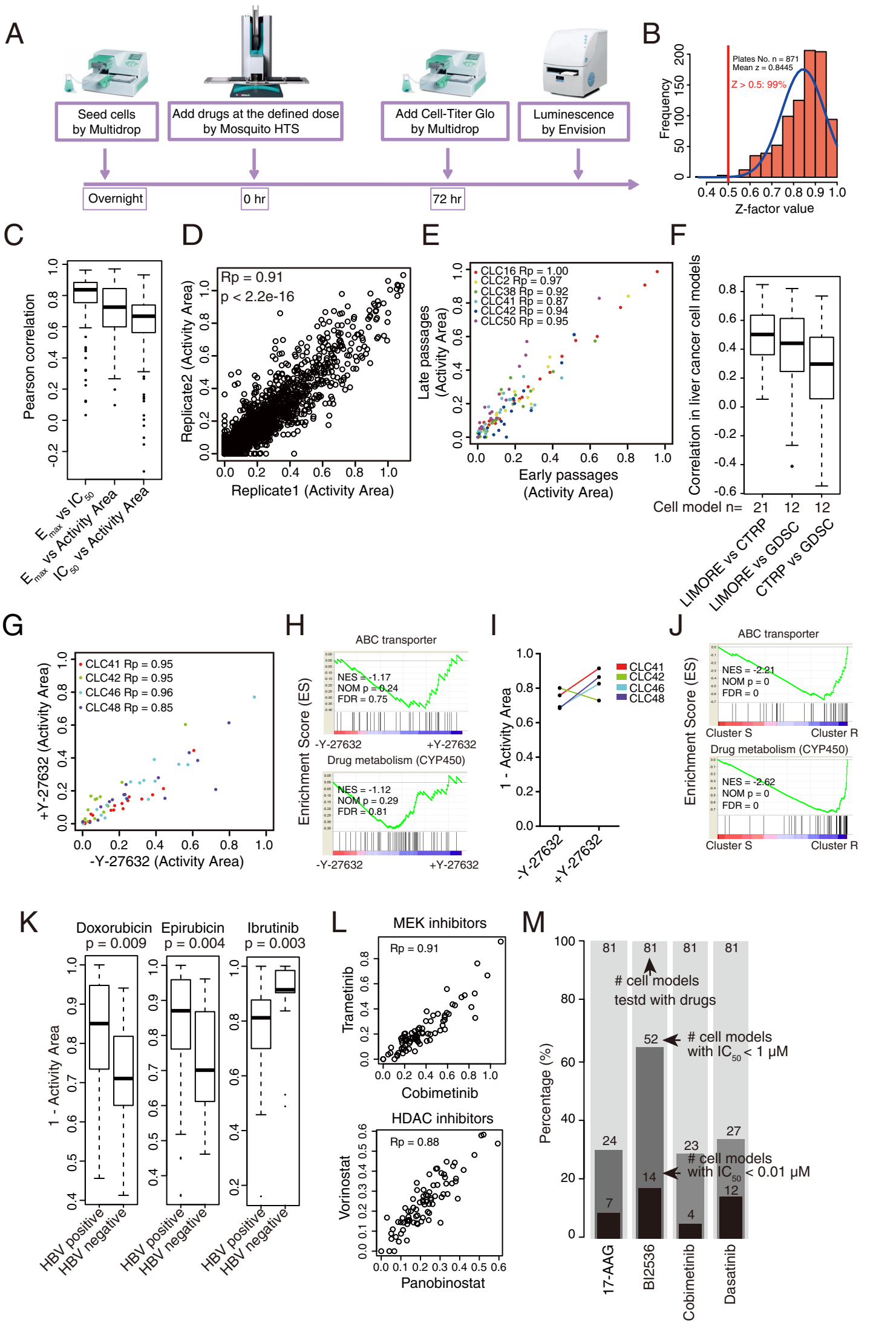

A
Seed cells by Multidrop
Add drugs at the defined dose by Mosquito HTS
Add Cell-Titer Glo by Multidrop
Luminescence by Envision
Overnight
0 hr
72 hr
B
Plates No. n = 871
Mean z = 0.8445
Z > 0.5: 99%
Frequency
Z-factor value
C
Pearson correlation
Emax vs IC50
Emax vs Activity Area
IC50 vs Activity Area
D
Rp = 0.91
p < 2.2e-16
Replicate2 (Activity Area)
Replicate1 (Activity Area)
E
CLC16 Rp = 1.00
CLC2 Rp = 0.97
CLC38 Rp = 0.92
CLC41 Rp = 0.87
CLC42 Rp = 0.94
CLC50 Rp = 0.95
Late passages (Activity Area)
Early passages (Activity Area)
F
Correlation in liver cancer cell models
Cell model n= 21 12 12
LIMORE vs CTRP
LIMORE vs GDSC
CTRP vs GDSC
G
CLC41 Rp = 0.95
CLC42 Rp = 0.95
CLC46 Rp = 0.96
CLC48 Rp = 0.85
+Y-27632 (Activity Area)
-Y-27632 (Activity Area)
H
ABC transporter
NES = -1.17
NOM p = 0.24
FDR = 0.75
-Y-27632
+Y-27632
Drug metabolism (CYP450)
NES = -1.12
NOM p = 0.29
FDR = 0.81
Enrichment Score (ES)
I
CLC41
CLC42
CLC46
CLC48
1 - Activity Area
-Y-27632
+Y-27632
J
ABC transporter
NES = -2.21
NOM p = 0
FDR = 0
Cluster S
Cluster R
Drug metabolism (CYP450)
NES = -2.62
NOM p = 0
FDR = 0
Enrichment Score (ES)
K
Doxorubicin p = 0.009
Epirubicin p = 0.004
Ibrutinib p = 0.003
1 - Activity Area
HBV positive
HBV negative
L
MEK inhibitors
Rp = 0.91
Trametinib
Cobimetinib
HDAC inhibitors
Rp = 0.88
Vorinostat
Panobinostat
M
81
81
81
81
cell models testd with drugs
52
cell models with IC50 < 1 μM
24
7
14
23
4
27
12
cell models with IC50 < 0.01 μM
Percentage (%)
17-AAG
BI2536
Cobimetinib
Dasatinib

Figure S3. The high-throughput drug screening, Related to Figure 3.

(A) Schematic outline of the high-throughput drug screening protocol.

(B) The histogram shows the distribution of Z-prime values from 871 screening 384-well plates.

(C) Boxplot shows Pearson correlations of Activity Area (AA), IC_{50} and E_{max} for 90 drugs tested in LIMORE.

(D) Scatterplot shows the Pearson correlation of AA values for biological replicates of drug screening data.

(E) Scatterplot shows the Pearson correlation of AA values for 6 LIMORE models (> 20 passages) and the matched early passage cells (< 10 passages).

(F) Boxplot shows Spearman correlations of drug responses for 38 drugs in liver cancer cell models among different datasets. Liver cancer cell models in CCLE and CTRP were combined as one dataset (CTRP) and drug results in CTRP were used. Drug response in CTRP was presented as area-under-curve (AUC) value. Drug response in GDSC is presented as IC_{50} value.

(G) Scatterplot shows the Pearson correlation of AA values for 4 cell models analyzed in media with or without Y-27632.

(H) Gene sets involved in drug transporter or metabolism are not enriched in cell models cultured with Y-27632, as demonstrated by Gene Set Enrichment Analysis (GSEA). NES, normalized enrichment score in GSEA. FDR, false discovery rate.

(I) Scatterplot shows 1-Activity Area values of paclitaxel in 4 cell models analyzed in media with or without Y-27632.

(J) Gene sets enriched in Cluster R cell models, as demonstrated by Gene Set Enrichment Analysis (GSEA). Gene sets involved in drug transporter or metabolism are shown.

(K) Boxplots show 1-Activity Area values of doxorubicin, epirubicin and ibrutinib in HBV positive

(n=66) and negative (n=15) LIMORE models. Statistical significance was determined by Wilcoxon rank-sum test.

(L) Scatterplots show the Pearson correlation of AA values for drugs targeting MEK and HDAC, respectively.

(M) The number of LIMORE models for each cutoff of IC_{50} values of the four indicated drugs.

For box-and-whisker plot, the box indicates IQR, the line in the box indicates the median, the whiskers indicate points within Q3+1.5×IQR and Q1−1.5×IQR and the points beyond whiskers indicate outliers.

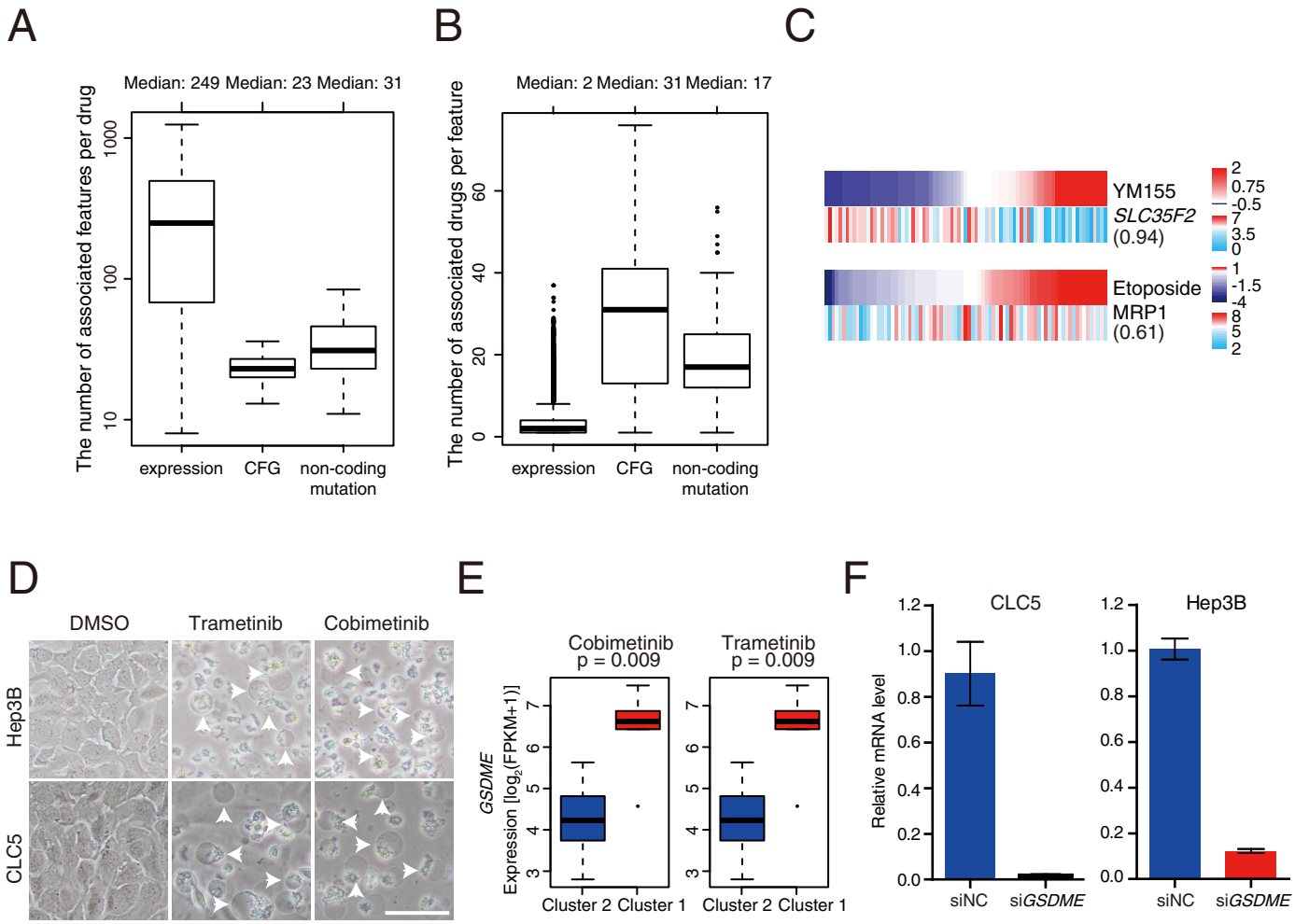

A
Median: 249 Median: 23 Median: 31
The number of associated features per drug
1000
100
10
expression
CFG
non-coding mutation
B
Median: 2 Median: 31 Median: 17
The number of associated drugs per feature
60
40
20
0
expression
CFG
non-coding mutation
C
YM155
SLC35F2
(0.94)
Etoposide
MRP1
(0.61)
2
0.75
-0.5
7
3.5
0
1
-1.5
-4
8
5
2
D
DMSO
Trametinib
Cobimetinib
Hep3B
CLC5
E
Cobimetinib
p = 0.009
Trametinib
p = 0.009
GSDME
Expression [log2(FPKM+1)]
Cluster 2 Cluster 1
Cluster 2 Cluster 1
F
CLC5
Hep3B
Relative mRNA level
1.2
1.0
0.8
0.6
0.4
0.2
0.0
siNC
siGSDME
siNC
siGSDME

Figure S4. Pharmacogenomic analysis in LIMORE, Related to Figure 4.

(A and B) Boxplots show the number of associated features per drug (A) and the number of associated drugs per feature (B). Features with EN score > 0.60 were included.

(C) Heatmaps show the gene-drug associations for YM155 and etoposide. Rank-ordered sensitivity values are indicated as upper heatmap with corresponding features plotted below. The number in the parentheses indicates the EN score.

(D) Morphology characteristics of LIMORE models treated with Cobimetinib (10 μM) or Trametinib (0.5 μM) for 48 hr. Scale bars, 100 μm.

(E) Boxplots show the RNA-seq expression values for *GSDME* in Cluster 1 or 2 subgroup of LIMORE models. Statistical significance was determined by unpaired Student's t-test.

(F) Knockdown efficiency of *GSDME* siRNA in CLC5 and Hep3B as determined by Real-time qRCR. Data are presented as mean±SD. siNC, nontargeting siRNA control.

For box-and-whisker plot, the box indicates IQR, the line in the box indicates the median, the whiskers indicate points within Q3+1.5×IQR and Q1−1.5×IQR and the points beyond whiskers indicate outliers.

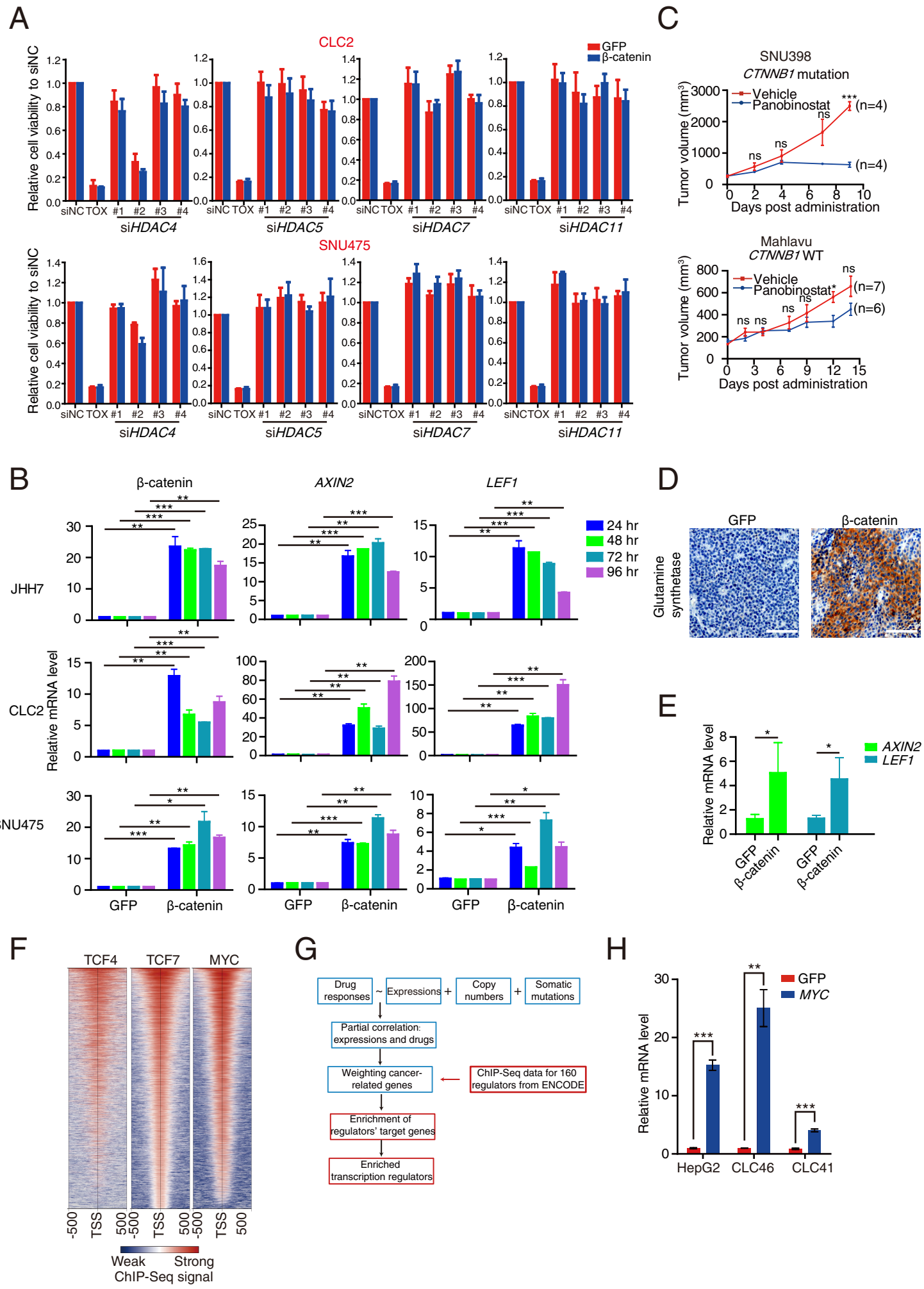

A
CLC2
GFP
β-catenin
Relative cell viability to siNC
siNC TOX #1 #2 #3 #4
siHDAC4
siHDAC5
siHDAC7
siHDAC11
SNU475
B
β-catenin
AXIN2
LEF1
24 hr
48 hr
72 hr
96 hr
JHH7
CLC2
SNU475
Relative mRNA level
GFP
β-catenin
C
SNU398
CTNNB1 mutation
Vehicle
Panobinostat
(n=4)
(n=4)
Tumor volume (mm3)
Days post administration
Mahlavu
CTNNB1 WT
(n=7)
(n=6)
D
GFP
β-catenin
Glutamine synthetase
E
AXIN2
LEF1
Relative mRNA level
F
TCF4
TCF7
MYC
-500
TSS
500
Weak
Strong
ChIP-Seq signal
G
Drug responses
Expressions
Copy numbers
Somatic mutations
Partial correlation: expressions and drugs
Weighting cancer-related genes
ChIP-Seq data for 160 regulators from ENCODE
Enrichment of regulators' target genes
Enriched transcription regulators
H
GFP
MYC
Relative mRNA level
HepG2
CLC46
CLC41

Figure S5. Synthetic lethal interactions with Wnt activation, Related to Figure 5.

(A) Relative cell viabilities of CLC2 and SNU475 cells after β-catenin overexpression with siRNA mediated knock-down of *HDAC4*, *HDAC5*, *HDAC7*, *HDAC11* for 72 hr. GFP is the control group. siNC, nontargeting siRNA control. TOX, transfection control used to assess transfection efficiency. Data are presented as mean±SD. Experiments were biologically repeated in triplicate and one representative result is shown.

(B) mRNA levels of β-catenin and target genes *AXIN2*, *LEF1* measured by qPCR in JHH7, CLC2 and SNU475 cells at different time points. Data are presented as mean±SD.

(C) Tumor growth curves of SNU398 and Mahlavu treated with panobinostat or vehicle. Data are presented as mean±SEM.

(D) Representative pictures of glutamine synthetase (GS) staining on JHH7-GFP and JHH7-β-catenin derived cancers. n=4 mice for JHH7-GFP and n=3 mice for JHH7-β-catenin. Scale bars, 100 μM.

(E) mRNA levels of β-catenin target genes *AXIN2* and *LEF1* measured by qPCR in JHH7-GFP and JHH7-β-catenin derived cancers. n=4 mice for JHH7-GFP and n=3 mice for JHH7-β-catenin. Data are presented as mean±SD.

(F) Heatmap shows the co-binding of TCF4, TCF7 and MYC on TCF4 bound promoter regions (4733 peaks) as determined by ChIP-Seq profiles in HepG2. TSS, transcription start site.

(G) Outline for drug response-associated transcription regulator analysis by integrating RNA-Seq data from LIMORE and ChIP-Seq data from ENCODE.

(H) mRNA levels of *MYC* measured by qPCR in HepG2, CLC46 and CLC41 cells. Data are presented as mean±SD.

*p < 0.05, **p < 0.01, ***p < 0.001 by unpaired Student's t-test.

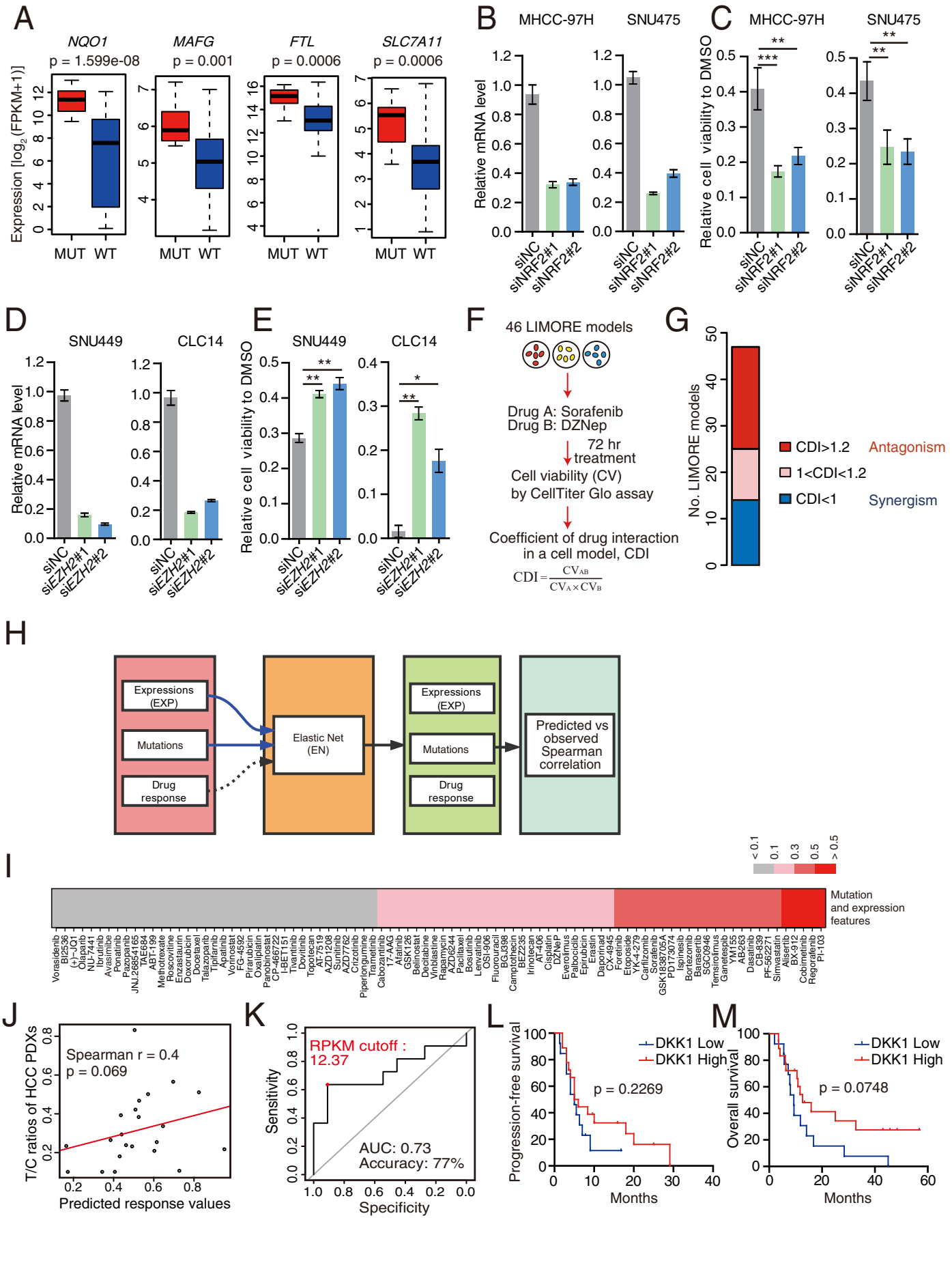

A
NQO1
p = 1.599e-08
MAFG
p = 0.001
FTL
p = 0.0006
SLC7A11
p = 0.0006
Expression [log2(FPKM+1)]
MUT WT
B
MHCC-97H
SNU475
Relative mRNA level
siNC
siNRF2#1
siNRF2#2
C
MHCC-97H
SNU475
Relative cell viability to DMSO
D
SNU449
CLC14
Relative mRNA level
siEZH2#1
siEZH2#2
E
SNU449
CLC14
Relative cell viability to DMSO
F
46 LIMORE models
Drug A: Sorafenib
Drug B: DZNep
72 hr treatment
Cell viability (CV) by CellTiter Glo assay
Coefficient of drug interaction in a cell model, CDI
CDI = CVAB / (CVA × CVB)
G
No. LIMORE models
CDI>1.2 Antagonism
1<CDI<1.2
CDI<1 Synergism
H
Expressions (EXP)
Mutations
Drug response
Elastic Net (EN)
Predicted vs observed Spearman correlation
I
< 0.1
0.1
0.3
0.5
> 0.5
Mutation and expression features
Vorasidenib
BI2536
(+)-JQ1
Olaparib
NU-7441
Ibrutinib
Avasimibe
Ponatinib
Pazopanib
JNJ-26854165
TAE684
ABT-199
Methotrexate
Roscovitine
Enzastaurin
Doxorubicin
Docetaxel
Talazoparib
Tipifarnib
Apatinib
Vorinostat
PHA-665752
Piracubicin
Oxaliplatin
Panobinostat
CP-466722
I-BET151
Tivantinib
Dovitinib
Topotecan
AT-7519
AZD1208
Sunitinib
AZD7762
Crizotinib
Piperlongumine
Trametinib
Cabozantinib
17-AAG
Afatinib
GSK126
Belinostat
Decitabine
Vinblastine
Rapamycin
AZD6244
Paclitaxel
Bosutinib
Lenvatinib
OSI-906
Fluorouracil
BGJ398
Camptothecin
BEZ235
Irinotecan
AT-406
Cisplatin
DZNeP
Everolimus
Pelitinib
Epirubicin
Erastin
Daporinad
CX-4945
Foretinib
Etoposide
YK-4-279
Carfilzomib
Sorafenib
GSK1838705A
PD173074
Ispinesib
Bortezomib
Barasertib
SGC0946
Temsirolimus
Ganetespib
YM155
AB263
Dasatinib
CB-839
PF-562271
Simvastatin
Alisertib
BX-912
Cobimetinib
Regorafenib
PI-103
J
T/C ratios of HCC PDXs
Spearman r = 0.4
p = 0.069
Predicted response values
K
RPKM cutoff : 12.37
AUC: 0.73
Accuracy: 77%
Sensitivity
Specificity
L
DKK1 Low
DKK1 High
p = 0.2269
Progression-free survival
Months
M
DKK1 Low
DKK1 High
p = 0.0748
Overall survival
Months

Figure S6. Prediction models and biomarkers for sorafenib sensitivity, Related to Figure 6.

(A) Boxplots show the mRNA levels of NRF2 downstream genes in LIMORE models with or without *KEAP1* mutation. MUT, mutation. Statistical significance was determined by unpaired Student's t-test.

(B-E) Knockdown efficiencies of NRF2 siRNA in MHCC-97H and SNU475 (B) and *EZH2* siRNA in SNU449 and CLC14 (D) were determined by qPCR. The relative cell viabilities of MHCC-97H and SNU475 with NRF2 knockdown (C) or SNU449 and CLC14 with *EZH2* knockdown (E) were detected after 6 μM sorafenib treatment for 72 hr. Data are presented as mean±SD. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by unpaired Student's t-test. siNC, nontargeting siRNA control. Experiments were biologically repeated in triplicate and one representative result is shown.

(F) Overview of sorafenib and DZNep combination screening in 46 LIMORE models. Single-dosed DZNep (10 μM) and sorafenib (5 μM) was simultaneously added to treat LIMORE models. The relative cell viability was calculated for the drug A, drug B and combination of A and B, respectively. The combination effect was evaluated by the coefficient of drug interaction (CDI).

(G) Barplot shows the distribution of combination effects of DZNep and sorafenib in 46 LIMORE models. For the CDI, 1 indicates drug additivity; CDI > 1 means the antagonistic effects of two drugs; CDI < 1 means the synergistic effects.

(H) Strategies to build prediction models for drug responses using elastic net (EN) regression in LIMORE. The prediction models were constructed in 54 training cell models and validated in 27 independent cell models.

(I) Heatmap shows Spearman correlation between predicted and detected drug responses for sorafenib in 27 LIMORE models.

(J) Scatterplot shows the Spearman correlation of predicted response values and treatment-to-control (T/C) ratios of patient-derived xenografts (PDXs). 22 HCC PDXs were treated with

sorafenib (40 mg/kg) and T/C ratios were monitored. n=10 mice for each PDX.

(K) ROC (receiver operating characteristics) curve of *DKK1* expression levels and sorafenib responses in 22 HCC PDXs.

(L and M) Kaplan-Meier plots for progression-free (L) and overall (M) survival. 31 HCC patients whose serum samples collected before sorafenib treatment were analyzed. Statistical significance was determined by Log-rank test.

For box-and-whisker plot, the box indicates IQR, the line in the box indicates the median, the whiskers indicate points within Q3+1.5×IQR and Q1−1.5×IQR and the points beyond whiskers indicate outliers.