Supplementary Figures

Figure S1. The difference of ctDNA properties in patients with different clinical characteristics and distict histological subtypes.

Figure S2. Number of SNVs/Indels in blood and ctDNA level demonstrate significantly linear dependence in different histological subtypes.

Figure S3. Mutational and pathway landscape for the stage IV, untreated population.

Figure S4. Distribution of mutational types for frequently altered genes in different histological subtypes.

Figure S5. Percentage of pathway alterations in each sample.

Figure S6. Pathway members and interactions in the 10 selected pathways.

FigureS7. Samples with different *EGFR* driver mutations show little discrepancy about ctDNA properties and patient prognosis.

Figure S8. *EGFR* clonality and concurrent *TP53* mutation in ctDNA cohort from

stage lV, untreated population.

FigureS9. Concurrent mutant genes for different *EGFR* driver mutations in the stage IV, untreated cohort.

Figure S10. The ctDNA properties and pathway alterations in the stage lV, untreated, NSCLC subset with *RB1* mutations.

Figure S11. The interaction between somatic mutational events in the stage lV, untreated, NSCLC subset with *RB1* mutations.

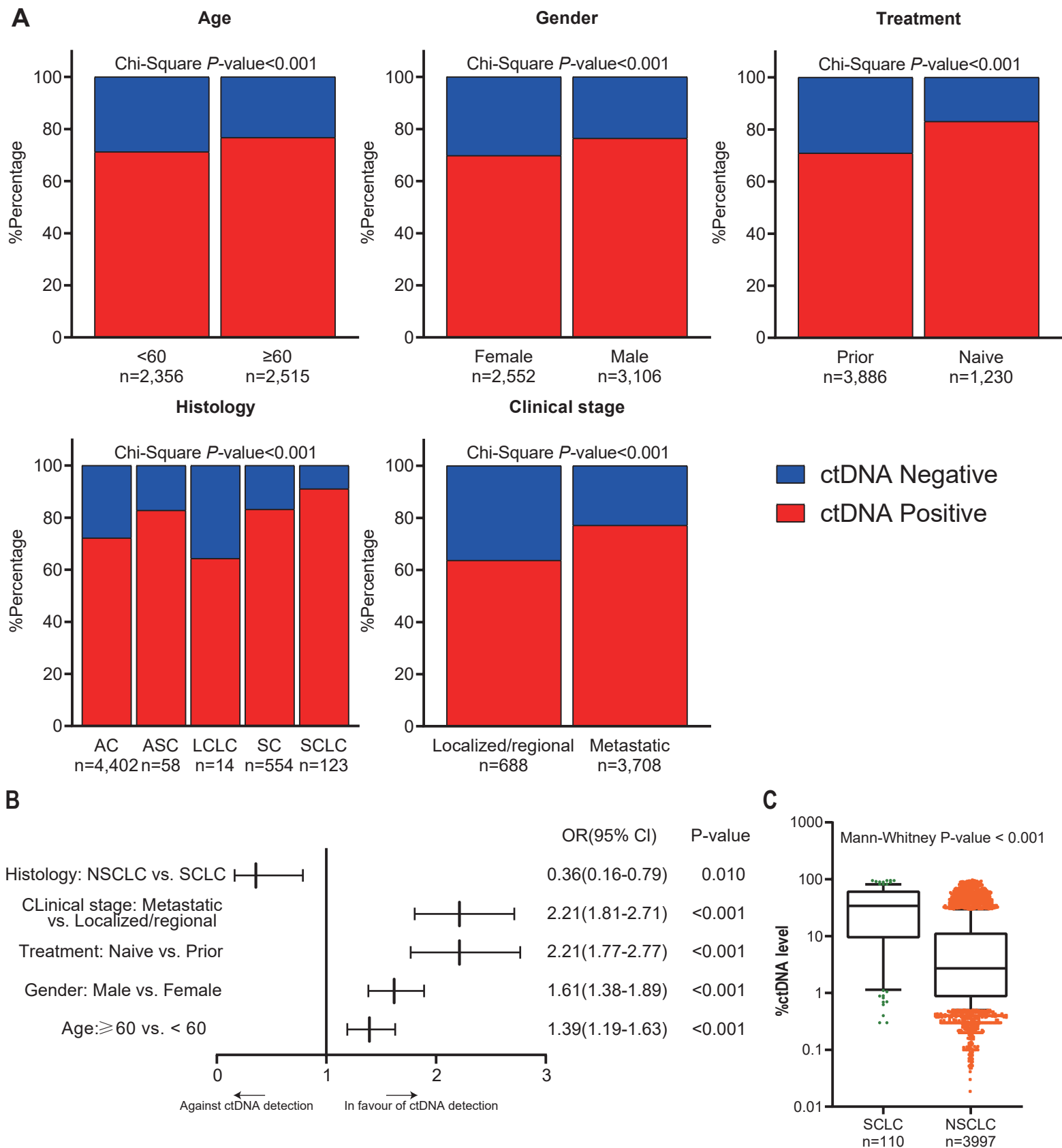Figure S12. Genomic concordance between paired tissue and blood samples.

Figure S1. The difference of ctDNA properties in patients with different clinical characteristics and distict histological subtypes. (A) ctDNA detectability varies in patients with different clinical characteristics. AC, adenocarcinoma; ASC, adenosquamous carcinoma; LCLC, large cell lung cancer; SC, squamous carcinoma; SCLC, small cell lung cancer. (B) Forest plot shows clinical factors and histological subtypes significantly influencing ctDNA detectability. Statistics is performed using Binary Logistic Regression and corresponding parameters are listed in the right part of the panel. NSCLC, non-small cell lung cancer. (C) ctDNA level fluctuate obviously between NSCLC and SCLC. P-value <0.05 is identified as statistical significance.
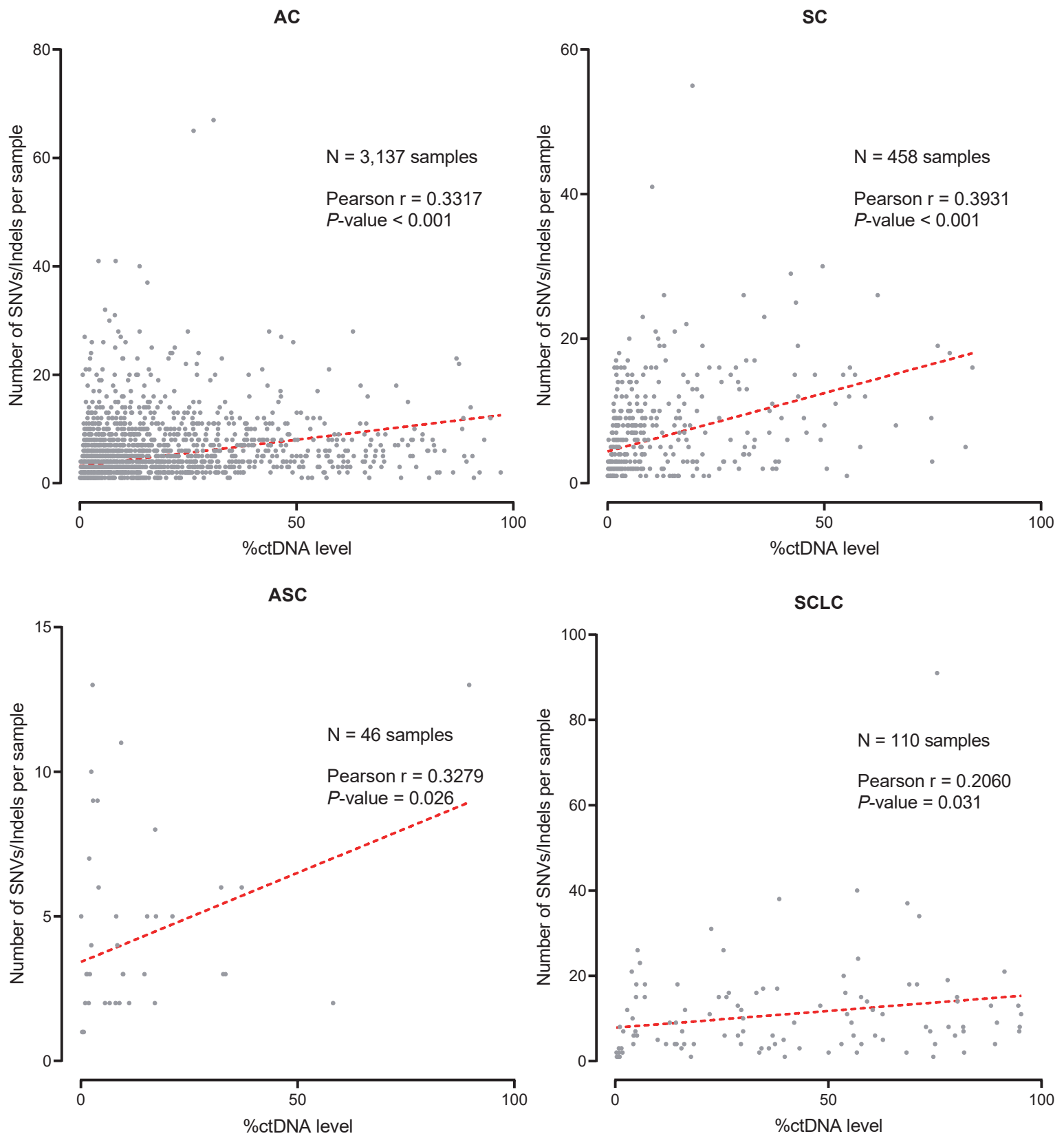
Figure S2. Number of SNVs/Indels in blood and ctDNA level demonstrate significantly linear dependence in different histological subtypes. ctDNA level indicates the maximal variant AF in each blood sample. Pearson's coefficient is used to evaluate the correlation and *P*-value <0.05 is identified as statistical significance.
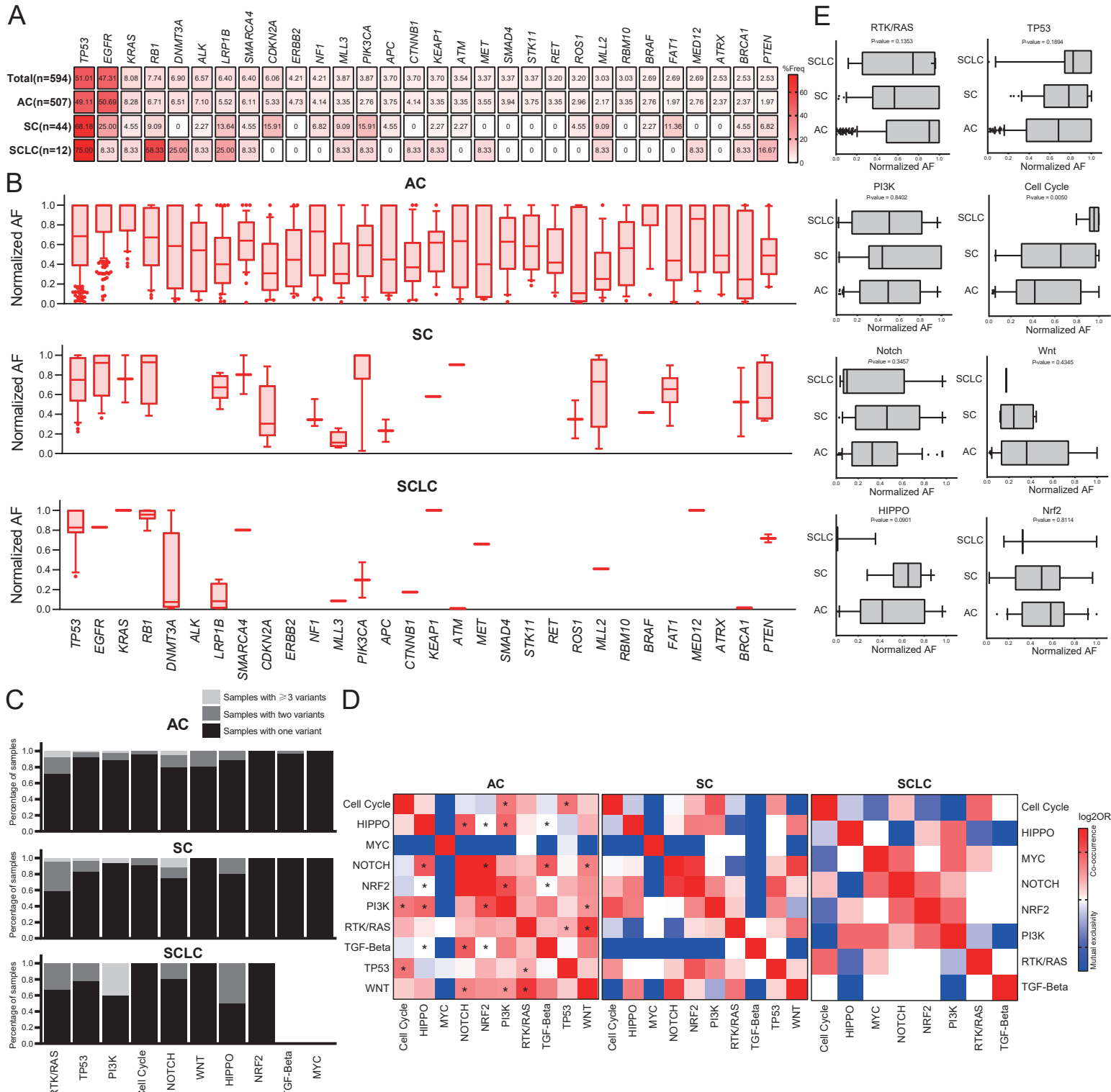
Figure S3. Mutational and pathway landscape for the stage IV, untreated population. (A) Frequently altered gene in this cohort and different subtypes. Color gradation indicates the prevalence of each mutant gene. AC, adenocarcinoma; SC, squamous carcinoma; SCLC, small cell lung cancer. (B) Bos plots illustrate the normalized AF of different mutant genes in AC, SC, and SCLC. Centre line, median; box limits, upper and lower quartiles; whiskers, 10%-90% data range. (C) Percentage of samples harboring specific numbers of variant within the same pathways. (D) Mutual exclusivity (blue) and co-occurrence (red) among pathway alterations. The co-occurrence and mutual exclusivity of one pathway (assumed as A) with another pathway (assumed as B) was estimated via odds ratio (OR) and q-value derived from Benjamini-Hochberg FDR correction procedure. OR = (Neither * Both) / (A Not B * B Not A). Those pathway pairs with OR >4 or <0.25 and q-value <0.05 are identified as significantly enriched co-pathways or mutually exclusive pathways which are labelled with asterisks. (E) Normalized AFs of pathway alterations vary among different AC, SC, and SCLC. Centre line, median; box limits, upper and lower quartiles; whiskers, 10%-90% data range. Statistics is performed using One-way ANOVA and P-value <0.05 is identified as statistical significance.
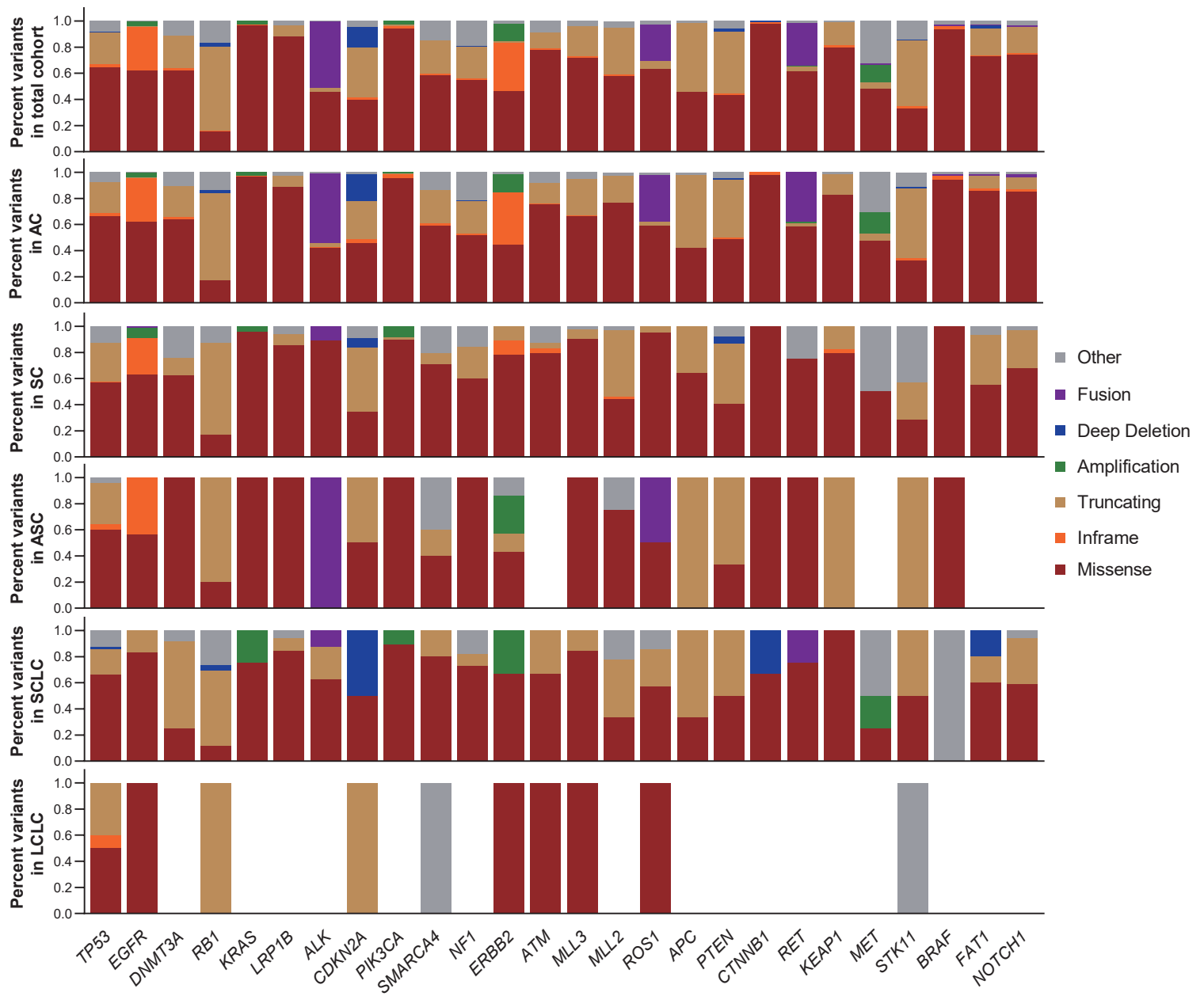
Figure S4. Distribution of mutational types for frequently altered genes in different histological subtypes. Top 26 genes in the total cohort are illustrated.
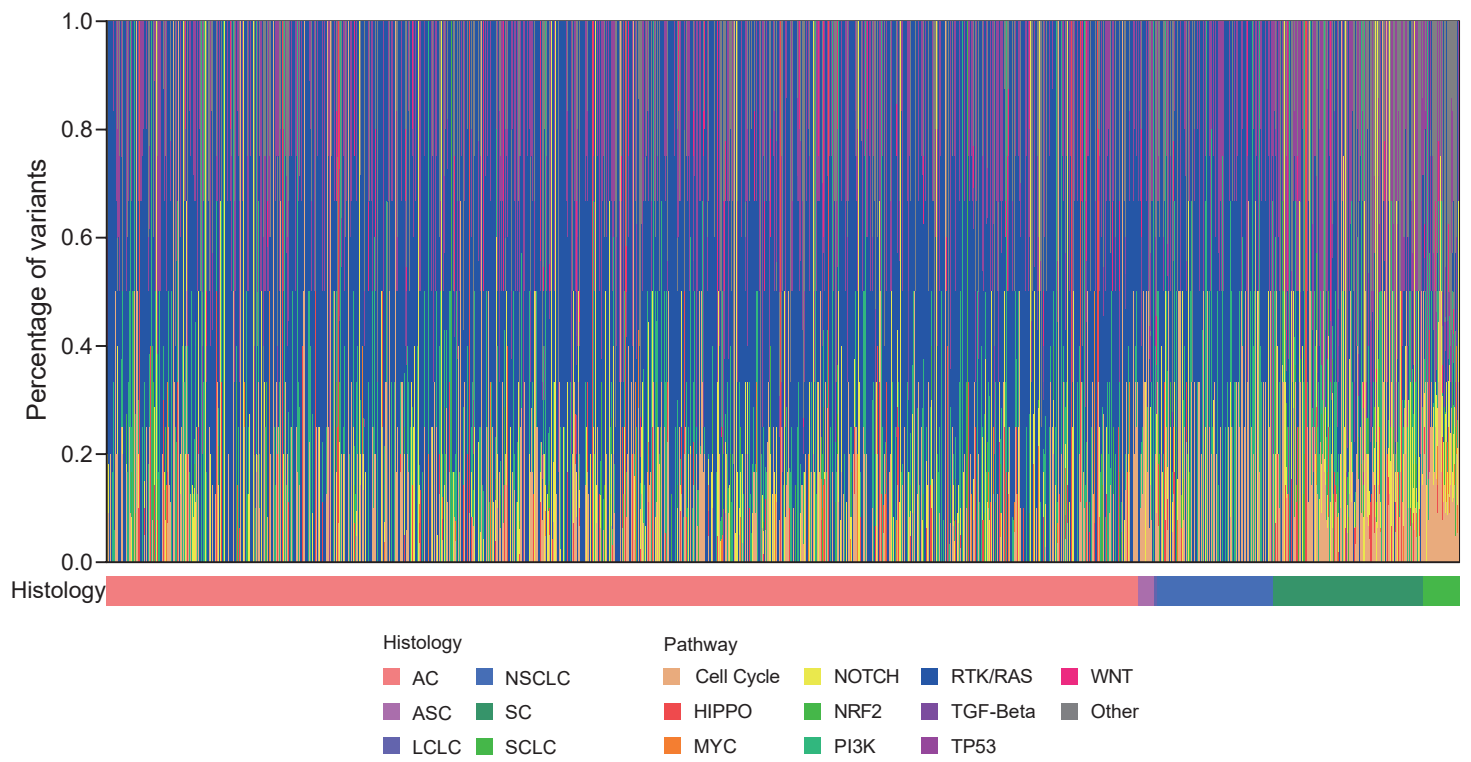
Figure S5. Percentage of pathway alterations in each sample. The histological subtype is displayed below in the chart.
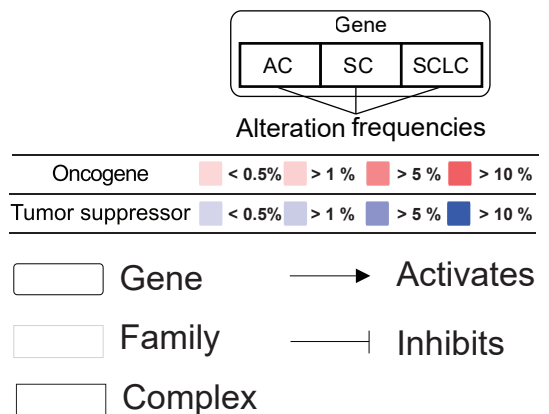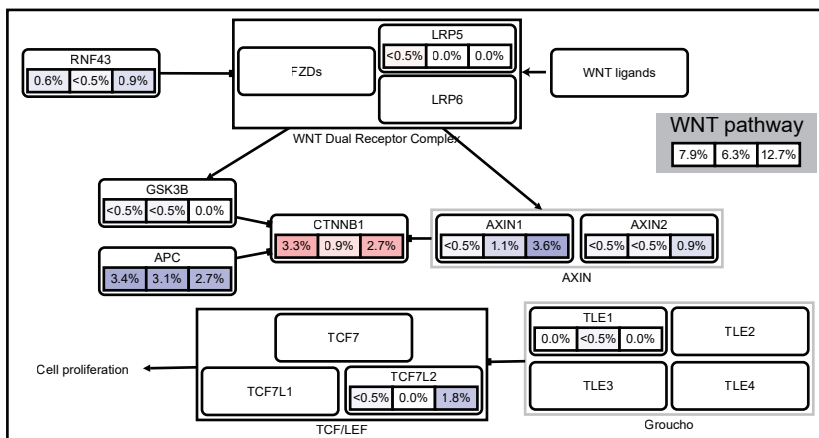
RTK/RAS pathway

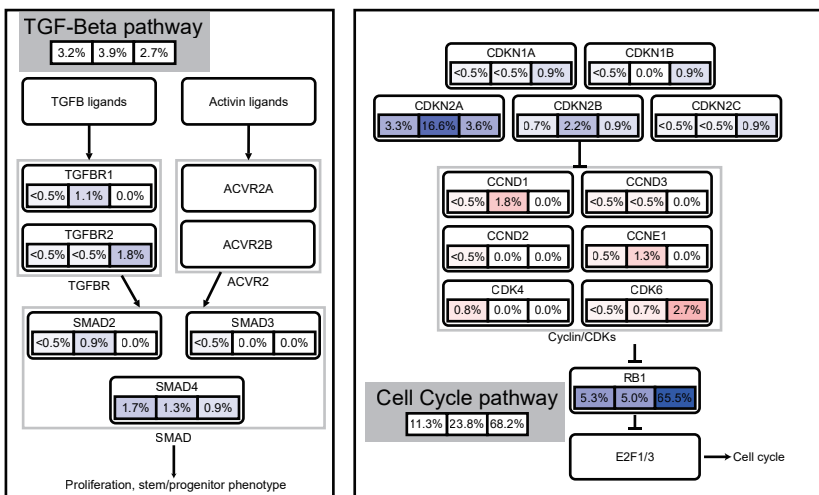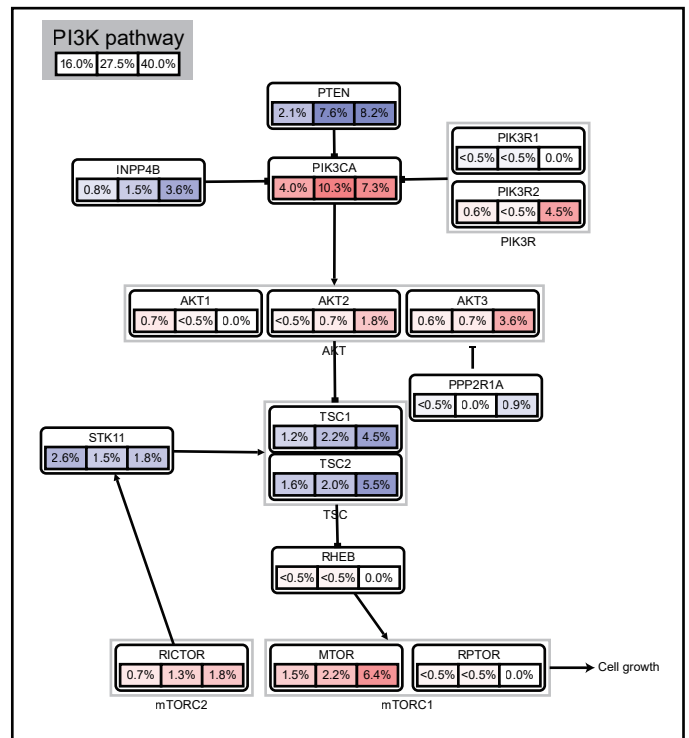**RTK/RAS pathway:** 76.6% 46.9% 54.5%

Gene	AC	SC	SCLC
EGFR	51.5%	10.9%	5.5%
ERBB2	4.3%	2.0%	2.7%
ERBB3	0.6%	0.7%	1.8%
ERBB4	1.1%	3.7%	4.5%
MET	2.8%	2.2%	3.6%
PDGFRA	1.1%	2.0%	3.6%
FGFR1	1.1%	2.4%	2.7%
FGFR2	0.5%	0.9%	0.9%
FGFR3	0.7%	1.8%	0.9%
FGFR4	0.8%	0.7%	3.6%
KIT	1.2%	2.6%	6.4%
IGF1R	0.6%	0.9%	1.8%
RET	3.2%	1.8%	3.6%
ROS1	2.6%	4.2%	6.4%
ALK	5.8%	3.9%	6.4%
FLT3	0.7%	1.1%	2.7%
NTRK1	1.5%	1.1%	1.8%
NTRK2	<0.5%	<0.5%	0.0%

RTKs

Gene	AC	SC	SCLC
CBL	1.1%	0.7%	3.6%
ERRFI1	<0.5%	<0.5%	0.9%
SOS1	<0.5%	0.0%	0.9%
NF1	4.6%	5.2%	9.1%

RASA1

Gene	AC	SC	SCLC
PTPN11	<0.5%	0.9%	0.9%
KRAS	6.8%	4.8%	6.4%
HRAS	<0.5%	0.9%	0.0%
NRAS	0.6%	1.1%	0.0%

RIT1 — RAS

Gene	AC	SC	SCLC
RAC1	<0.5%	0.0%	0.0%
ARAF	0.5%	<0.5%	0.9%
BRAF	2.3%	3.1%	0.9%
RAF1	<0.5%	0.7%	0.9%

RAF

Gene	AC	SC	SCLC
MAPK1	<0.5%	<0.5%	0.9%
MAP2K1	0.8%	0.0%	1.8%
MAP2K2	<0.5%	<0.5%	0.9%

MEK

Proliferation — Cell survival — Translation

TP53 pathway

TP53 pathway: 48.5% 65.1% 80.9%

DNA replication stress

Gene	AC	SC	SCLC
MDM2	0.8%	0.7%	0.9%
MDM4	<0.5%	0.7%	0.9%
ATM	3.7%	5.0%	4.5%
CHEK2	1.1%	2.6%	0.9%
TP53	45.4%	61.4%	80.0%

RPS6KA3

Oncogenic stress — Cell survival, proliferation — Senescence, apoptosis

NRF2 pathway

NRF2 pathway: 3.1% 15.9% 10.0%

Gene	AC	SC	SCLC
KEAP1	2.0%	7.4%	6.4%
CUL3	<0.5%	2.4%	2.7%
NFE2L2	0.7%	6.3%	2.7%

Oxidative Stress Response

Notch pathway

Notch pathway: 9.1% 21.2% 44.6%

Gene	AC	SC	SCLC
FBXW7	0.7%	4.8%	8.2%

CUL1 — CLEAVED NOTCH — MALM3 — KAT2B

Gene	AC	SC	SCLC
CREBBP	1.4%	1.8%	12.7%
EP300	1.1%	2.8%	7.3%

HES-X — HEY-X — NRARP

Gene	AC	SC	SCLC
NCOR1	0.7%	1.3%	1.8%
SPEN	<0.5%	<0.5%	0.0%
KDM5A	0.7%	1.8%	3.6%
NCOR2	<0.5%	0.0%	0.0%
NOTCH1	1.7%	5.9%	14.6%
NOTCH2	1.0%	2.0%	6.4%
NOTCH3	1.3%	4.4%	10.0%
NOTCH4	1.2%	1.5%	1.8%

Cell growth, apoptosis

HIPPO pathway

HIPPO pathway: 1.6% 2.6% 10.9%

Gene	AC	SC	SCLC
FAT1	1.9%	5.7%	10.9%
FAT2	2.0%	3.3%	5.5%
FAT3	<0.5%	0.0%	1.8%
FAT4	<0.5%	<0.5%	0.0%
NF2	0.8%	1.5%	0.0%

LATS1/2 — TAZ — YAP1 — TEAD2

Cell proliferation and differentiation

MYC pathway

MYC pathway: 1.6% 2.6% 10.9%

Gene	AC	SC	SCLC
MYC	0.9%	<0.5%	6.4%
MYCN	<0.5%	0.7%	0.9%

MYCL — MYC/MAX Complex

Gene	AC	SC	SCLC
MAX	<0.5%	<0.5%	0.0%
MGA	<0.5%	<0.5%	0.0%

MAX/MGA Complex

Cell growth, proliferation, apoptosis

TGF-Beta pathway

TGF-Beta pathway: 3.2% 3.9% 2.7%

TGFB ligands — Activin ligands

Gene	AC	SC	SCLC
TGFBR1	<0.5%	1.1%	0.0%
TGFBR2	<0.5%	<0.5%	1.8%

ACVR2A — ACVR2B — TGFBR — ACVR2

Gene	AC	SC	SCLC
SMAD2	<0.5%	0.9%	0.0%
SMAD3	<0.5%	0.0%	0.0%
SMAD4	1.7%	1.3%	0.9%

SMAD

Proliferation, stem/progenitor phenotype

Cell Cycle pathway

Cell Cycle pathway: 11.3% 23.8% 68.2%

Gene	AC	SC	SCLC
CDKN1A	<0.5%	<0.5%	0.9%
CDKN1B	<0.5%	0.0%	0.9%
CDKN2A	3.3%	16.6%	3.6%
CDKN2B	0.7%	2.2%	0.9%
CDKN2C	<0.5%	<0.5%	0.9%
CCND1	<0.5%	1.8%	0.0%
CCND3	<0.5%	<0.5%	0.0%
CCND2	<0.5%	0.0%	0.0%
CCNE1	0.5%	1.3%	0.0%
CDK4	0.8%	0.0%	0.0%
CDK6	<0.5%	0.7%	2.7%

Cyclin/CDKs

Gene	AC	SC	SCLC
RB1	5.3%	5.0%	85.5%

E2F1/3 — Cell cycle

PI3K pathway

PI3K pathway: 16.0% 27.5% 40.0%

Gene	AC	SC	SCLC
PTEN	2.1%	7.6%	8.2%
INPP4B	0.8%	1.5%	3.6%
PIK3CA	4.0%	10.3%	7.3%
PIK3R1	<0.5%	<0.5%	0.0%
PIK3R2	0.6%	<0.5%	4.5%

PIK3R

Gene	AC	SC	SCLC
AKT1	0.7%	<0.5%	0.0%
AKT2	<0.5%	0.7%	1.8%
AKT3	0.6%	0.7%	3.6%

AKT

Gene	AC	SC	SCLC
PPP2R1A	<0.5%	0.0%	0.9%
STK11	2.6%	1.5%	1.8%
TSC1	1.2%	2.2%	4.5%
TSC2	1.6%	2.0%	5.5%

TSC

Gene	AC	SC	SCLC
RHEB	<0.5%	<0.5%	0.0%
RICTOR	0.7%	1.3%	1.8%
MTOR	1.5%	2.2%	6.4%
RPTOR	<0.5%	<0.5%	0.0%

mTORC2 — mTORC1 — Cell growth

WNT pathway

WNT pathway: 7.9% 6.3% 12.7%

Gene	AC	SC	SCLC
RNF43	0.6%	<0.5%	0.9%
LRP5	<0.5%	0.0%	0.0%

FZDs — LRP6 — WNT ligands — WNT Dual Receptor Complex

Gene	AC	SC	SCLC
GSK3B	<0.5%	<0.5%	0.0%
CTNNB1	3.3%	0.9%	2.7%
AXIN1	<0.5%	1.1%	3.6%
AXIN2	<0.5%	<0.5%	0.9%
APC	3.4%	3.1%	2.7%

AXIN

Gene	AC	SC	SCLC
TLE1	0.0%	<0.5%	0.0%
TCF7L2	<0.5%	0.0%	1.8%

TCF7 — TCF7L1 — TCF/LEF — TLE2 — TLE3 — TLE4 — Groucho

Cell proliferation

Legend

Gene		
AC	SC	SCLC

Alteration frequencies

Oncogene: < 0.5% > 1% > 5% > 10%
Tumor suppressor: < 0.5% > 1% > 5% > 10%
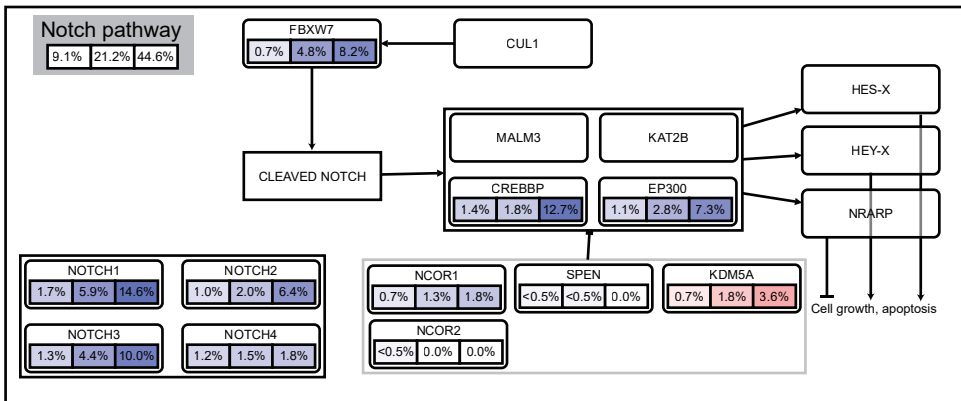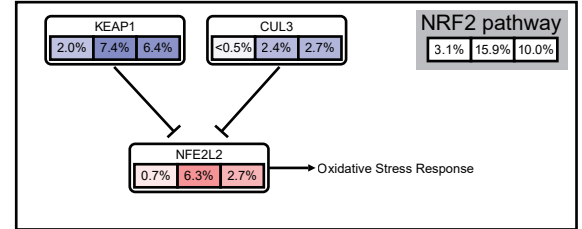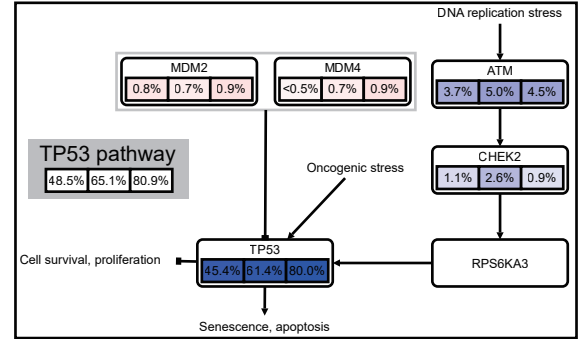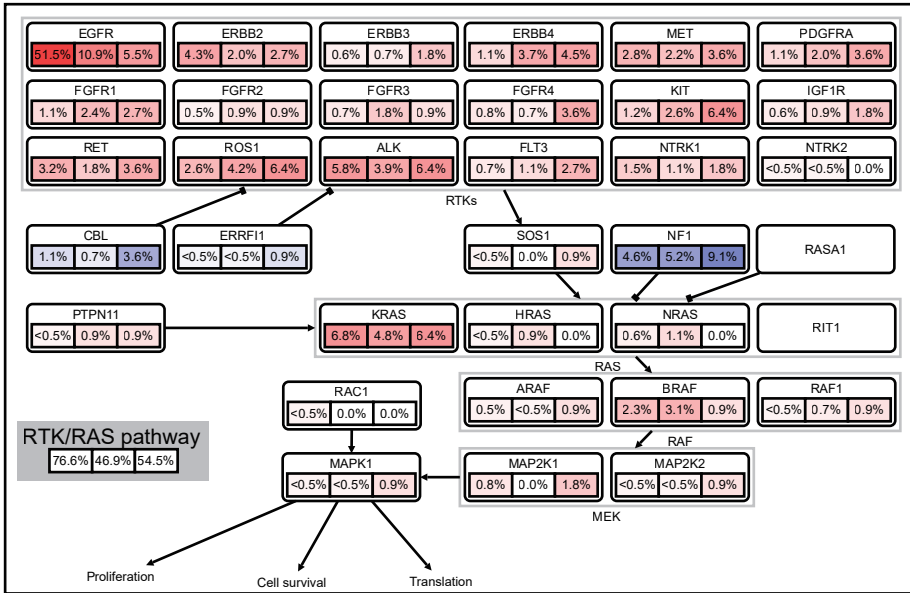
Gene — Family — Complex
Activates → — Inhibits ⊣

Figure S6. Pathway members and interactions in the 10 selected pathways. Oncogenes and tumor suppressor genes are illustrated with red and blue, respectively. Color intensity indicates the frequency of alteration within the entire dataset. Blank boxes represent genes not covered in our sequencing panel.
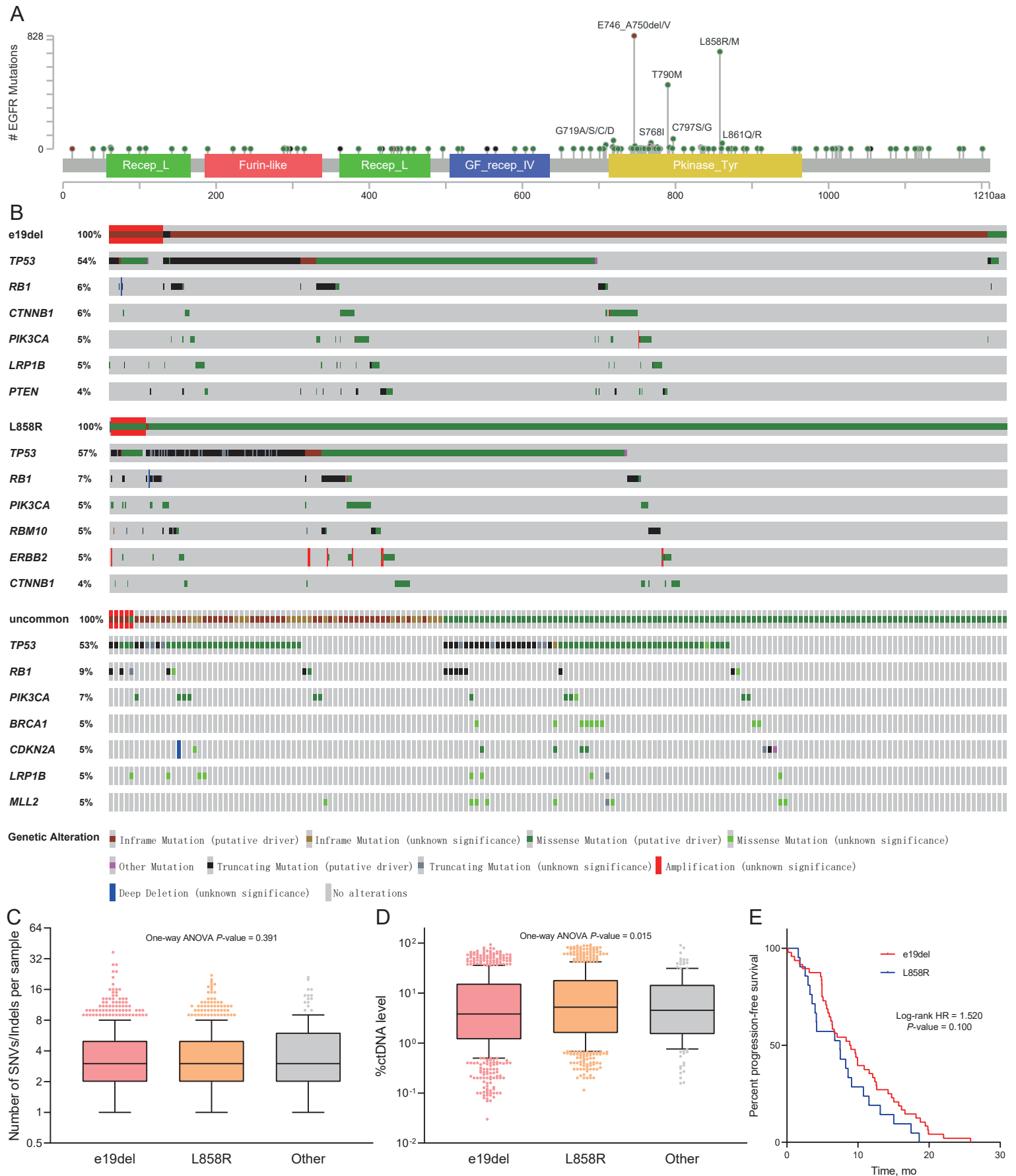
Figure S7. Samples with different *EGFR* driver mutations show little discrepancy about ctDNA properties and patient prognosis. (A) Lollipop chart illustrates the overview of all *EGFR* mutations. (B) Concurrent mutant genes for different *EGFR* driver mutations. (C) The number of SNVs/Indels in blood is similar among subsets with different *EGFR* driver mutations. (D) Samples with *EGFR* L858R show elevated ctDNA level compared with other *EGFR*-mutant samples. (E) Patients with e19del and L858R in blood ctDNA demonstrate similar PFS. Kaplan-Meier survival analysis is used to evaluated prognosis of different subgroups and *P*-value <0.05 is identified as statistical significance.
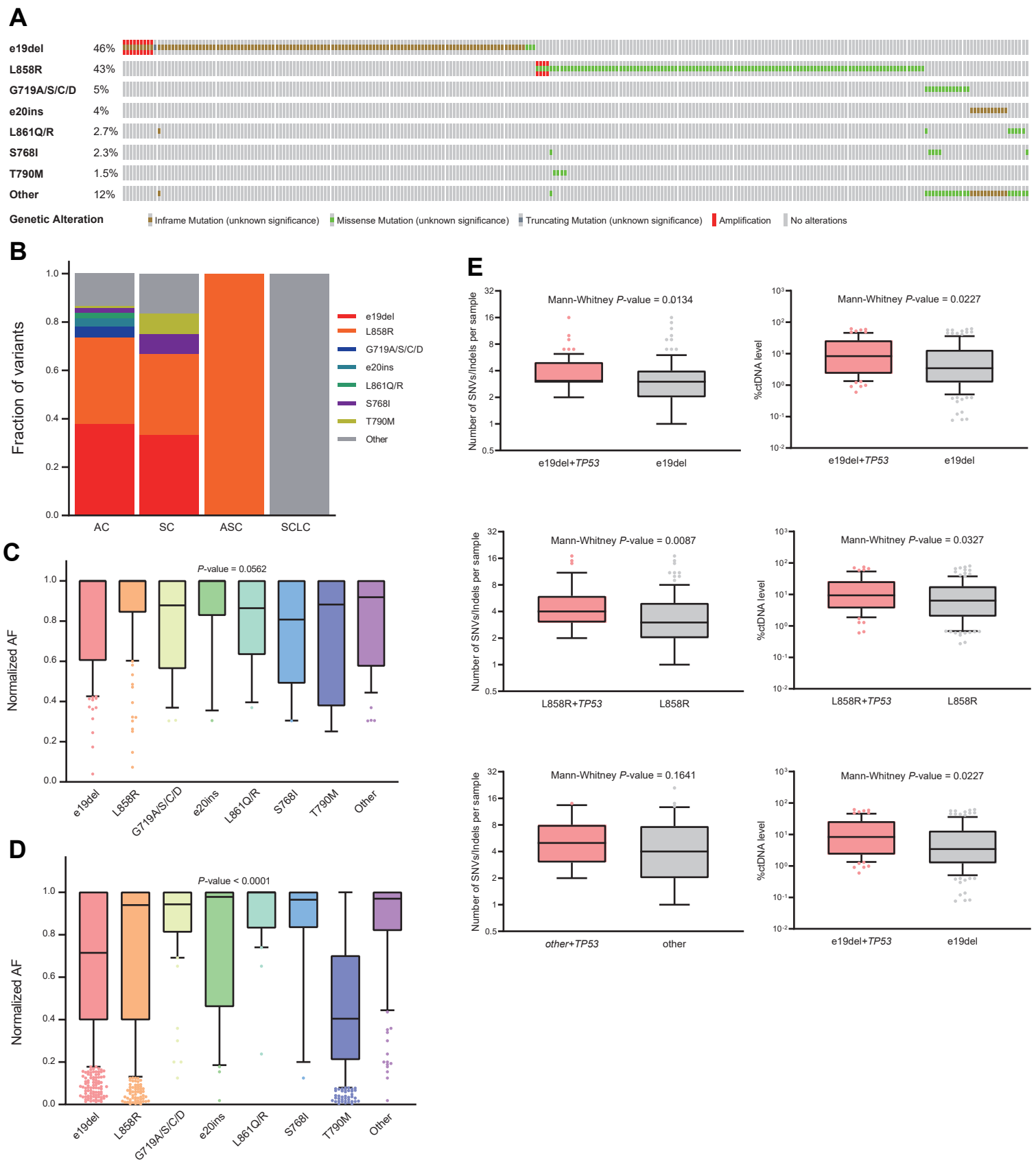
Figure S8. *EGFR* clonality and concurrent *TP53* mutation in ctDNA cohort from stage IV, untreated population. (A) Heatmap shows the presence of different mutation in *EGFR*-mutant samples. e19del, exon 19 deletion; AMP, amplifica-tion; e20ins, exon 20 insertion. (B) Distribution of different *EGFR* mutations among diverse *EGFR*-mutant subtypes. AC, adenocarcinoma; SC, squamous carcinoma; ASC, adenosquamous carcinoma; SCLC, small cell lung cancer. (C) and (D) Normalized AFs of different *EGFR* mutations in the stage IV, untreated cohort and the stage IV, treated cohort, respectively. Centre line, median; box limits, upper and lower quartiles; whiskers, 10%-90% data range. Statistics is performed using One-way ANOVA and P-value <0.05 is identified as statistical significance. (E) The number of SNVs/Indels in blood and ctDNA level are increased in samples with concurrent EGFR driver events and *TP53* mutations compared with *EGFR*-only samples. Centre line, median; box limits, upper and lower quartiles; whiskers, 10%-90% data range. Statistics is performed using Mann-Whitney U test and P-value <0.05 is identified as statistical significance.

Figure S9. Concurrent mutant genes for different *EGFR* driver mutations in the stage IV, untreated cohort.
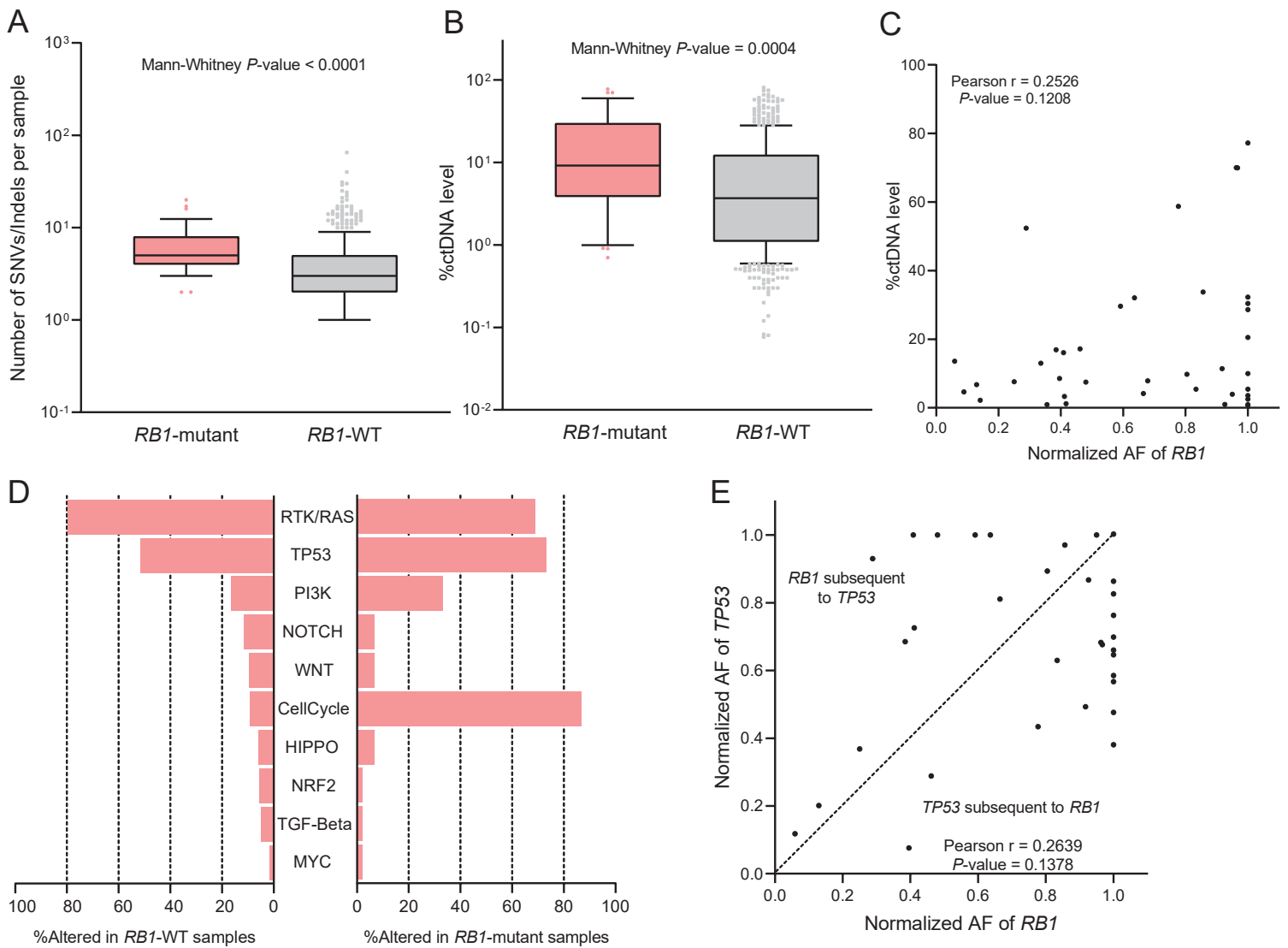
Figure S10. The ctDNA properties and pathway alterations in the stage IV, untreated, NSCLC subset with *RB1* mutations. (A) and (B) *RB1*-mutant NSCLC express superior number of SNVs/Indels and ctDNA level compared with *RB1*-wild type NSCLC. Centre line, median; box limits, upper and lower quartiles; whiskers, 10%-90% data range. Statistics is performed using Mann-Whitney U test and P-value <0.05 is identified as statistical significance. WT, wild type. (C) The normalized AFs of *RB1* mutations are not significantly correlated with the ctDNA level of corresponding samples. Pearson's coefficient is used to evaluate the correlation and P-value <0.05 is identified as statistical significance. (D) Frequencies of pathway alterations in RB1-mutant and *RB1*-wild type NSCLC. Statistics is performed using Chi-square test and P-value <0.05 is identified as statistical significance which is labelled with asterisks. (E) Correlation between normalized AFs of concurrent *RB1* and *TP53* mutations. Pearson's coefficient is used to evaluate the correlation and P-value <0.05 is identified as statistical significance.
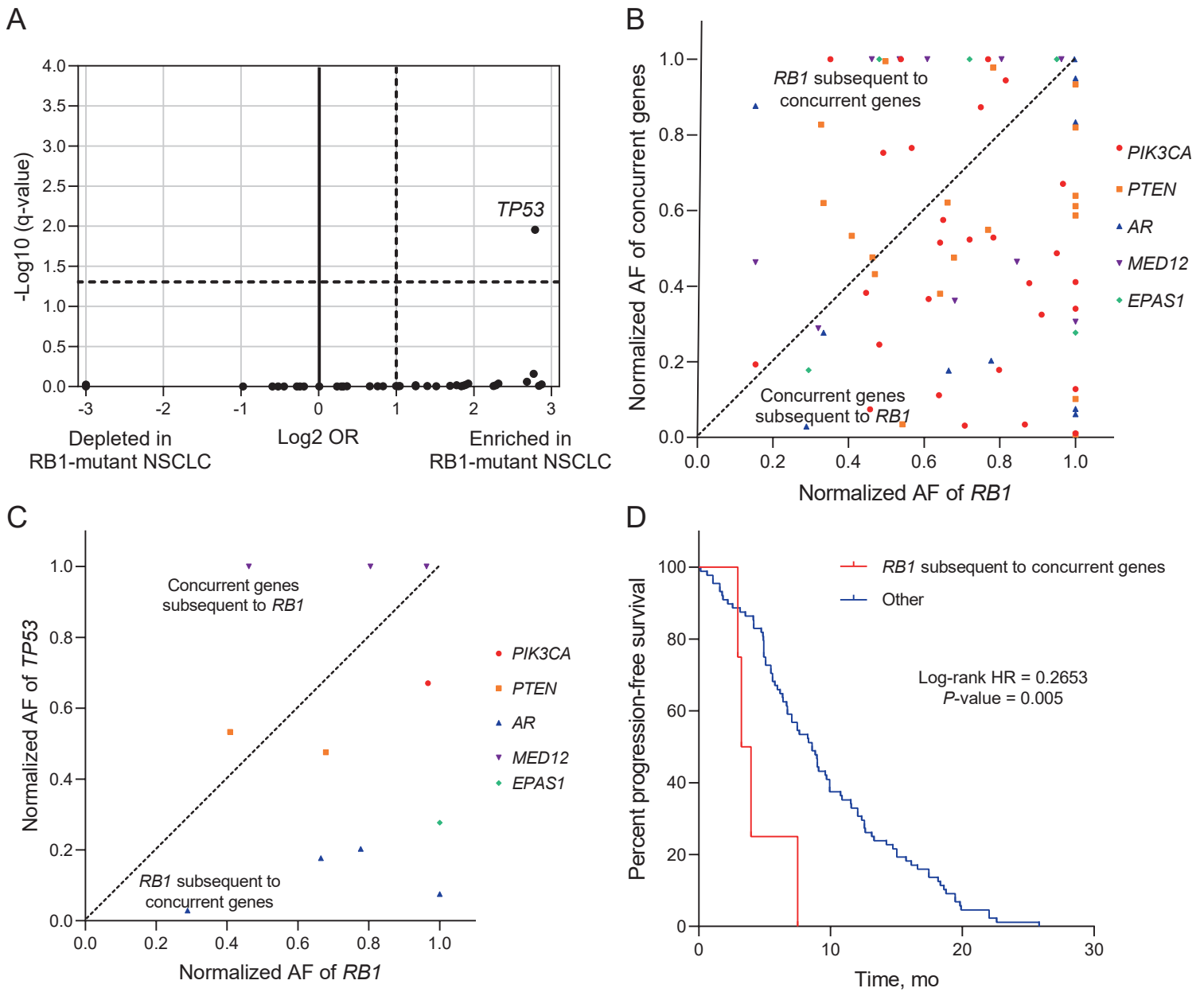
Figure S11. The interaction between somatic mutational events in the stage IV, untreated, NSCLC subset with RB1 mutations. (A) Mutual exclusivity and co-occurrence among mutant genes. The co-occurrence and mutual exclusivity of one gene (assumed as A) with another gene (assumed as B) was estimated via odds ratio(OR) and q-value derived from Benjamini-Hochberg FDR correction procedure. OR= (Neither * Both) / (A Not B*B Not A). Those gene pairs with OR >2 or <0.5 and q-value <0.05 are identified as significantly enriched co-genes or mutually exclusive genes. (B) and(C) Correlation between normalized AFs of concurrent RB1 and other mutations in NSCLC cohort and the stage IV, untreated NSCLC subset, respectively. (D) PFS is significantly deficient for patients with RB1 clonality < concurrent genes. Pearson's coefficient is used to evaluate the correlation and P-value <0.05 is identified as statistical significance. Kaplan-Meier survival analysis is used to evaluated prognosis of different subgroups and P-value <0.05 is identified as statistical significance.
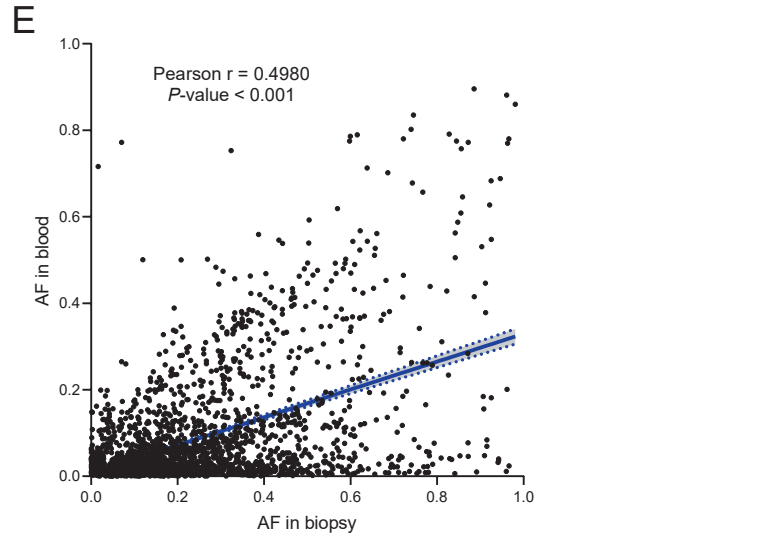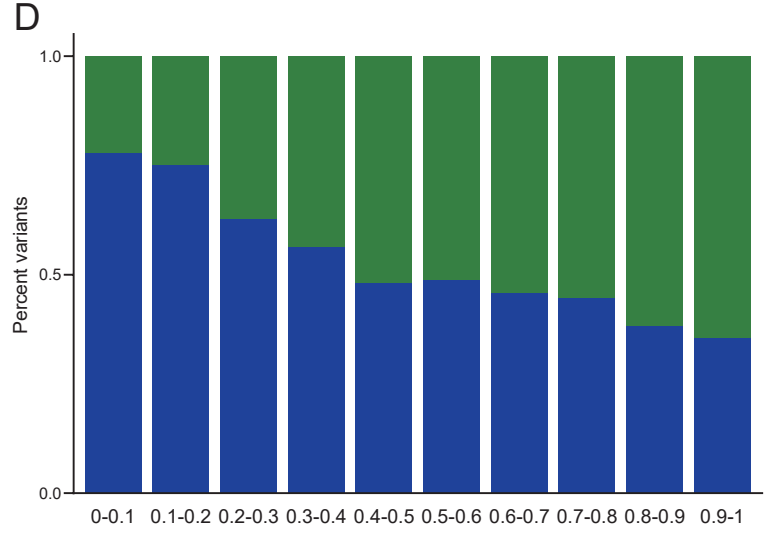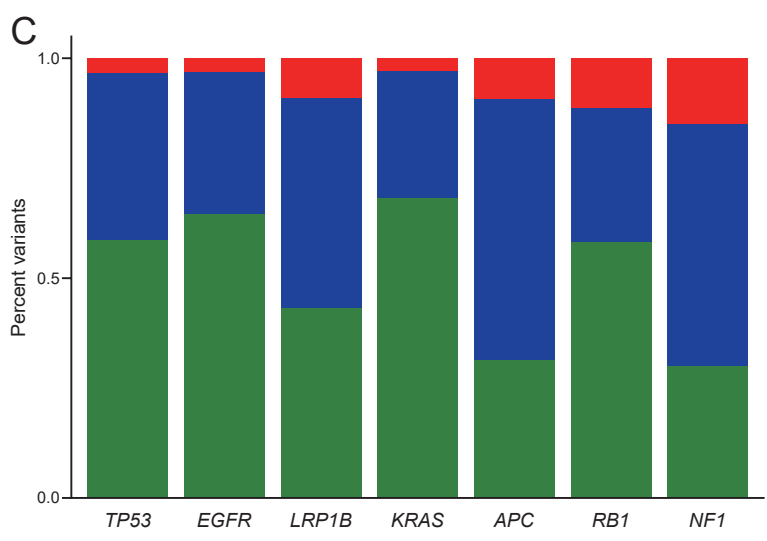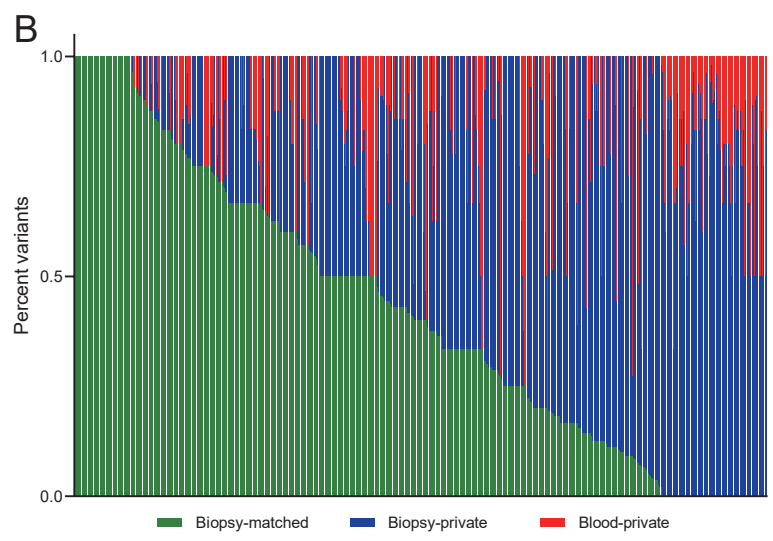
Figure S12. Genomic concordance between paired tissue and blood samples. (A)Genomic landscape of all tissue samples corresponding to blood samples in the primary cohort. (B-D) Distribution of different mutations defined by the presence in paired tissue and blood samples according to patients (B), mutant genes (C), and clonality range of tissue samples (D). (E) AFs of biopsy-matched mutations in paired samples show linear dependence.Pearson's coefficient is used to evaluate the correlation and *P*-value <0.05 is identified as statistical significance.