Cell, Volume *177*

Supplemental Information

Identifying Epistasis in Cancer Genomes:

A Delicate Affair

Joris van de Haar, Sander Canisius, Michael K. Yu, Emile E. Voest, Lodewyk F.A. Wessels, and Trey Ideker

Supplemental Figures

Identifying epistasis in cancer genomes: a delicate affair

Joris van de Haar, Sander Canisius, Michael K. Yu, Emile E. Voest, Lodewyk F.A. Wessels, and Trey Ideker

Contents

Supplemental Figure Legends

Figure S1: Mutation Load Association (MLA) spans a broad range in 13 tumor cohorts, is dependent on tumor type, and enriches for cancer genes at the low end, Related to Figure 2.

A. Boxplots showing the MLA distribution for all 13 cohorts analyzed. Boxes, whiskers, and dots indicate quartiles, 1.5 interquartile ranges, and outliers outside these ranges, respectively. Across all cohorts, MLAs range from negative to strongly positive.

B. Scatter plots of MLAs calculated on the TCGA cohort (x-axis), versus the MSKCC cohort (y-axis), for all tumor types with cohorts included from both databases. Linear fits and 95% confidence intervals (based on bootstrapping) are plotted.

C. Histogram of signed $\log_{10}$ P-values of Pearson (panel B) and Spearman (panel C) correlations between gene association with tumor mutation load (MLA) for every pair of included cohorts. Signed $\log_{10}$ P-values were calculated as ($-\log_{10}$ P-values) x (sign of correlation coefficient). Orange denotes cohort pairs of the same tumor type, gray denotes cohort pairs of different tumor types.

D. Cancer gene enrichment for the 25 genes with the lowest MLA for all TCGA cohorts. P-value range $1.0 \times 10^{-2} - 7.9 \times 10^{-12}$ (two-sided Fisher's exact test).

BLCA: Bladder Urothelial Carcinoma, BRCA: Breast Invasive Carcinoma, LUAD: Lung Adenocarcinoma, SCKM: Skin Cutaneous Melanoma, COADREAD: Colorectal Adenocarcinoma, STAD: Stomach Adenocarcinoma, LUSC: Lung Squamous Cell Carcinoma, UCEC: Uterine Corpus Endometrial Carcinoma, TCGA: The Cancer Genome Atlas, MSKCC: Memorial Sloan Kettering Cancer Center.

Figure S2: MLA anti-correlates with the number of significant findings on the DISCOVER test for mutual exclusivity across 12 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (DISCOVER P-value<0.05), plotted against their association with tumor mutation load (MLA). Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

Figure S3: MLA anti-correlates with the number of significant findings on the MEMo test for mutual exclusivity across 13 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (MEMo P-value<0.05), plotted against their association with tumor mutation load (MLA). Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

Figure S4: MLA anti-correlates with the number of significant findings on the WExT test for mutual exclusivity across 13 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (WExT P-value<0.05), plotted against their association with tumor mutation load (MLA). Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

Figure S5: MLA anti-correlates with the number of significant findings on the Fisher's exact test for mutual exclusivity across 13 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (Fisher's exact test P-value<0.05), plotted against their association with tumor mutation load (MLA). Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

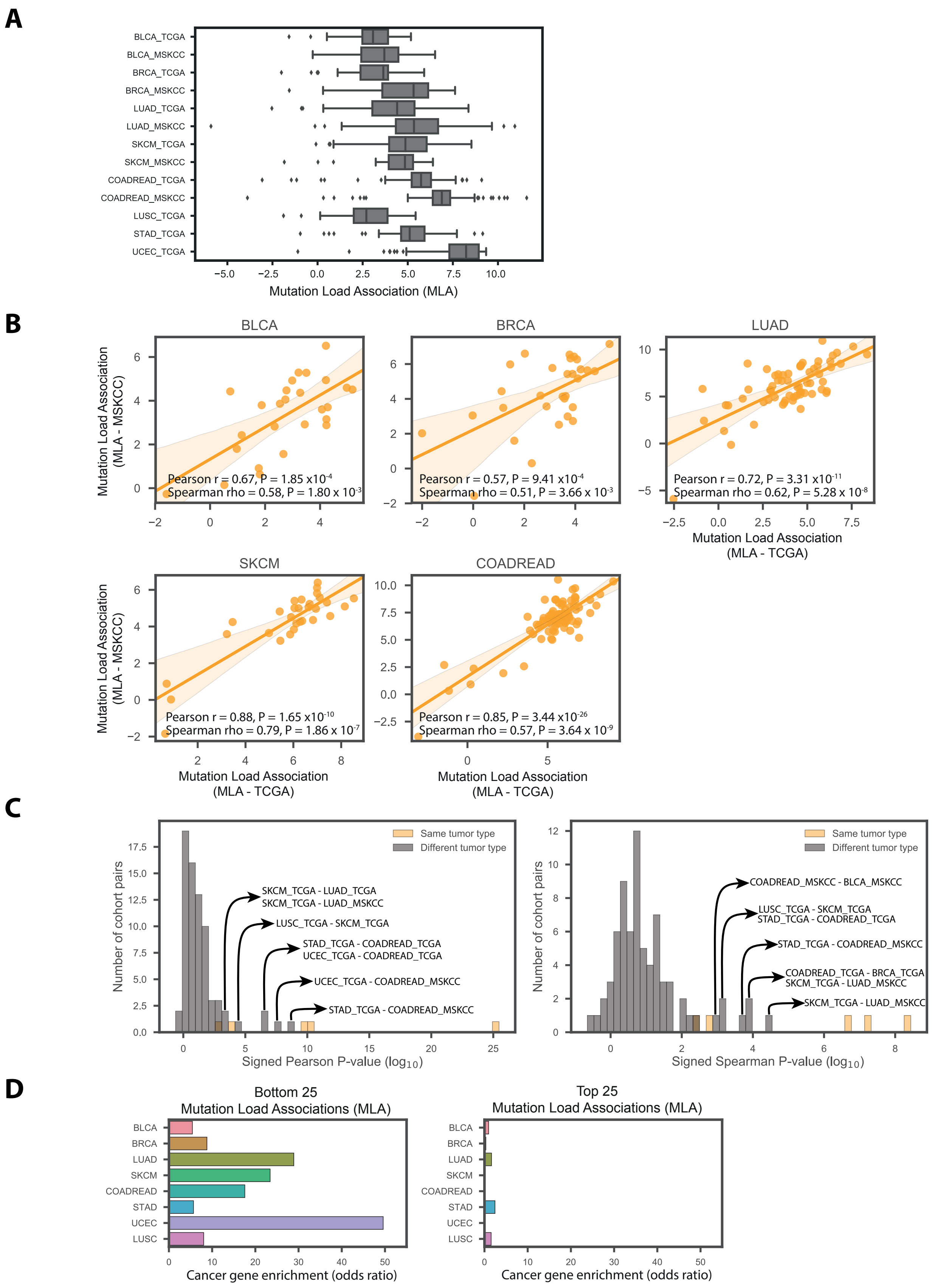
Figure S6: MLA anti-correlates with the number of significant findings on the MEGSA test for mutual exclusivity across 13 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (MEGSA P-value<0.05), plotted against their association with tumor mutation load (MLA). Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

Figure S7: Alteration Load Association anti-correlates with the number of significant findings on the DISCOVER test for mutual exclusivity across 8 cancer cohorts, Related to Figure 3.

Scatterplots of the percentage of a gene's pairwise interactions (DISCOVER P-value<0.05), plotted against their association with alteration load (ALA). Alterations are mutations and high-level copy number variations. Acronyms indicate cancer types and clinical cohorts, as defined in Figure S1.

Figure S1



A
BLCA_TCGA
BLCA_MSKCC
BRCA_TCGA
BRCA_MSKCC
LUAD_TCGA
LUAD_MSKCC
SKCM_TCGA
SKCM_MSKCC
COADREAD_TCGA
COADREAD_MSKCC
LUSC_TCGA
STAD_TCGA
UCEC_TCGA
−5.0 −2.5 0.0 2.5 5.0 7.5 10.0
Mutation Load Association (MLA)
B
BLCA
Pearson r = 0.67, P = 1.85 x10^{-4}
Spearman rho = 0.58, P = 1.80 x 10^{-3}
BRCA
Pearson r = 0.57, P = 9.41 x10^{-4}
Spearman rho = 0.51, P = 3.66 x 10^{-3}
LUAD
Pearson r = 0.72, P = 3.31 x10^{-11}
Spearman rho = 0.62, P = 5.28 x 10^{-8}
SKCM
Pearson r = 0.88, P = 1.65 x10^{-10}
Spearman rho = 0.79, P = 1.86 x 10^{-7}
COADREAD
Pearson r = 0.85, P = 3.44 x10^{-26}
Spearman rho = 0.57, P = 3.64 x 10^{-9}
Mutation Load Association (MLA - MSKCC)
Mutation Load Association (MLA - TCGA)
C
Same tumor type
Different tumor type
Number of cohort pairs
SKCM_TCGA - LUAD_TCGA
SKCM_TCGA - LUAD_MSKCC
LUSC_TCGA - SKCM_TCGA
STAD_TCGA - COADREAD_TCGA
UCEC_TCGA - COADREAD_TCGA
UCEC_TCGA - COADREAD_MSKCC
STAD_TCGA - COADREAD_MSKCC
Signed Pearson P-value ($\log_{10}$)
COADREAD_MSKCC - BLCA_MSKCC
LUSC_TCGA - SKCM_TCGA
STAD_TCGA - COADREAD_TCGA
STAD_TCGA - COADREAD_MSKCC
COADREAD_TCGA - BRCA_TCGA
SKCM_TCGA - LUAD_MSKCC
SKCM_TCGA - LUAD_MSKCC
Signed Spearman P-value ($\log_{10}$)
D
Bottom 25
Mutation Load Associations (MLA)
Top 25
Mutation Load Associations (MLA)
BLCA
BRCA
LUAD
SKCM
COADREAD
STAD
UCEC
LUSC
0 10 20 30 40 50
Cancer gene enrichment (odds ratio)

Figure S2

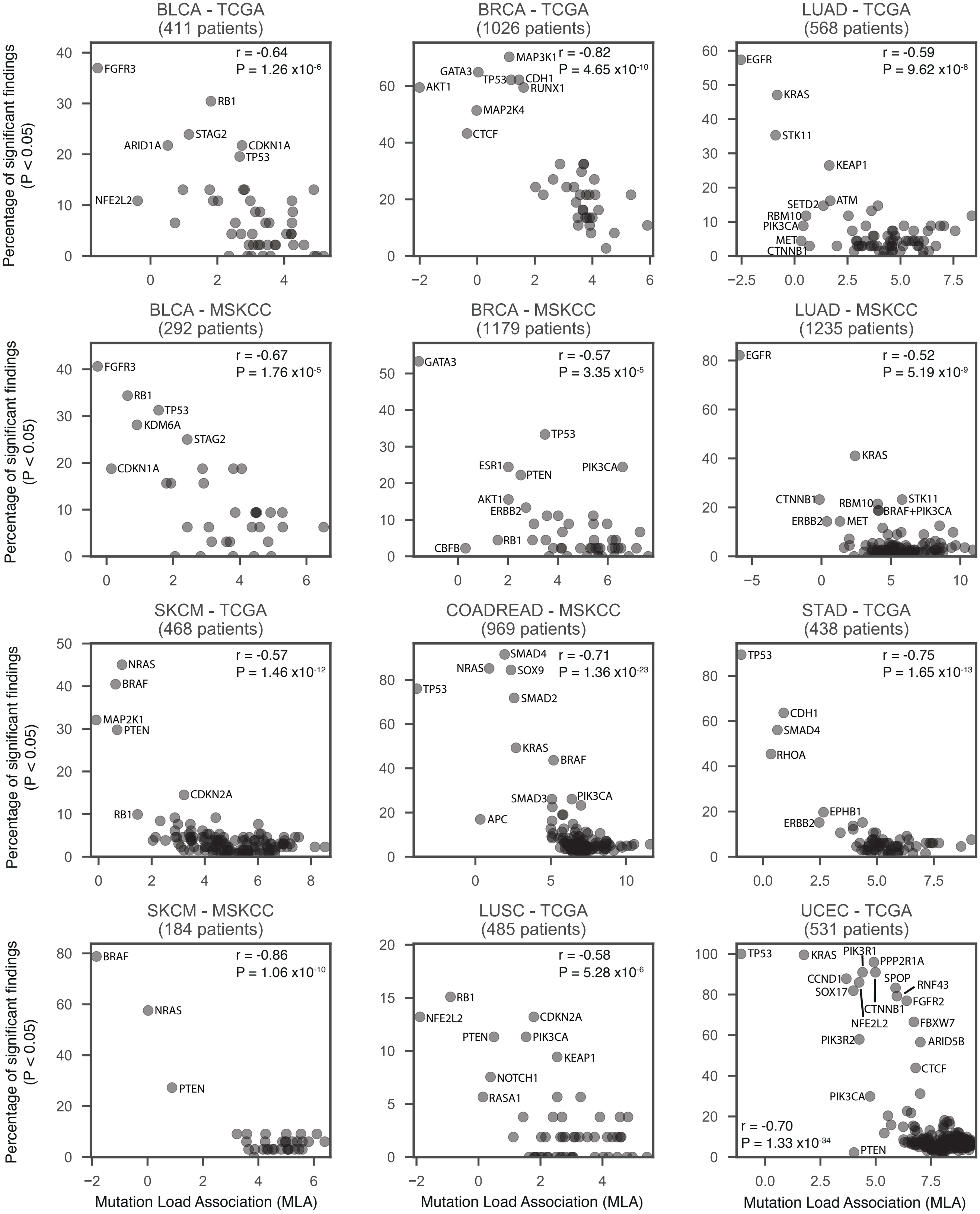
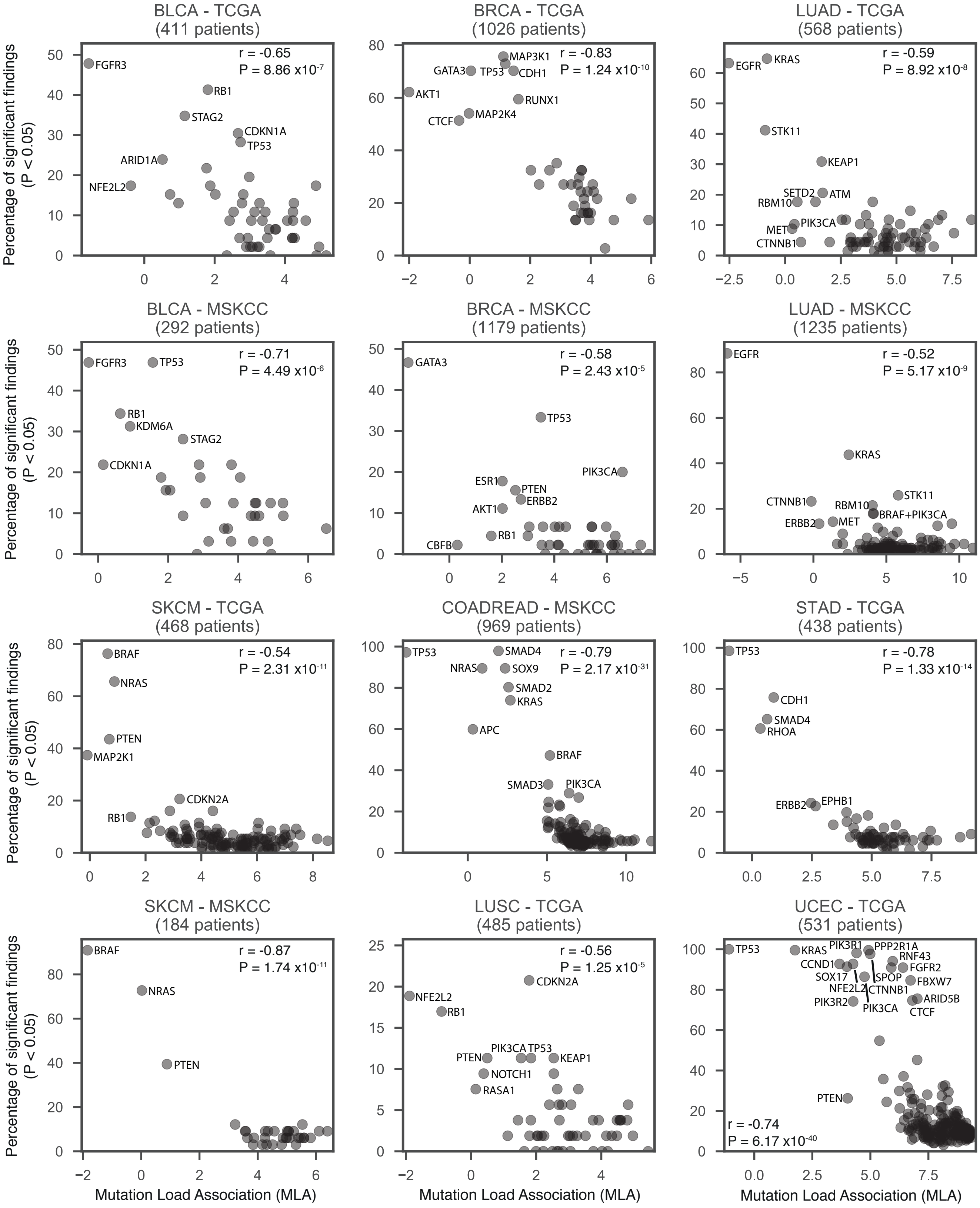

BLCA - TCGA
(411 patients)
r = -0.64
P = 1.26 x10-6
FGFR3
RB1
STAG2
ARID1A
CDKN1A
TP53
NFE2L2
BRCA - TCGA
(1026 patients)
r = -0.82
P = 4.65 x10-10
MAP3K1
GATA3
TP53
CDH1
RUNX1
AKT1
MAP2K4
CTCF
LUAD - TCGA
(568 patients)
r = -0.59
P = 9.62 x10-8
EGFR
KRAS
STK11
KEAP1
SETD2
ATM
RBM10
PIK3CA
MET
CTNNB1
BLCA - MSKCC
(292 patients)
r = -0.67
P = 1.76 x10-5
FGFR3
RB1
TP53
KDM6A
STAG2
CDKN1A
BRCA - MSKCC
(1179 patients)
r = -0.57
P = 3.35 x10-5
GATA3
TP53
ESR1
PTEN
PIK3CA
AKT1
ERBB2
RB1
CBFB
LUAD - MSKCC
(1235 patients)
r = -0.52
P = 5.19 x10-9
EGFR
KRAS
CTNNB1
RBM10
STK11
BRAF+PIK3CA
ERBB2
MET
SKCM - TCGA
(468 patients)
r = -0.57
P = 1.46 x10-12
NRAS
BRAF
MAP2K1
PTEN
CDKN2A
RB1
COADREAD - MSKCC
(969 patients)
r = -0.71
P = 1.36 x10-23
SMAD4
NRAS
SOX9
TP53
SMAD2
KRAS
BRAF
SMAD3
PIK3CA
APC
STAD - TCGA
(438 patients)
r = -0.75
P = 1.65 x10-13
TP53
CDH1
SMAD4
RHOA
EPHB1
ERBB2
SKCM - MSKCC
(184 patients)
r = -0.86
P = 1.06 x10-10
BRAF
NRAS
PTEN
LUSC - TCGA
(485 patients)
r = -0.58
P = 5.28 x10-6
RB1
NFE2L2
CDKN2A
PTEN
PIK3CA
KEAP1
NOTCH1
RASA1
UCEC - TCGA
(531 patients)
TP53
KRAS
PIK3R1
PPP2R1A
CCND1
SPOP
SOX17
RNF43
FGFR2
CTNNB1
NFE2L2
FBXW7
PIK3R2
ARID5B
CTCF
PIK3CA
r = -0.70
P = 1.33 x10-34
PTEN
Percentage of significant findings (P < 0.05)
Mutation Load Association (MLA)

Figure S3



BLCA - TCGA
(411 patients)
r = -0.65
P = 8.86 x10^{-7}
FGFR3
RB1
STAG2
CDKN1A
TP53
ARID1A
NFE2L2
BRCA - TCGA
(1026 patients)
r = -0.83
P = 1.24 x10^{-10}
MAP3K1
GATA3
TP53
CDH1
AKT1
RUNX1
CTCF
MAP2K4
LUAD - TCGA
(568 patients)
r = -0.59
P = 8.92 x10^{-8}
EGFR
KRAS
STK11
KEAP1
SETD2
ATM
RBM10
PIK3CA
MET
CTNNB1
BLCA - MSKCC
(292 patients)
r = -0.71
P = 4.49 x10^{-6}
FGFR3
TP53
RB1
KDM6A
STAG2
CDKN1A
BRCA - MSKCC
(1179 patients)
r = -0.58
P = 2.43 x10^{-5}
GATA3
TP53
PIK3CA
ESR1
PTEN
ERBB2
AKT1
RB1
CBFB
LUAD - MSKCC
(1235 patients)
r = -0.52
P = 5.17 x10^{-9}
EGFR
KRAS
STK11
CTNNB1
RBM10
BRAF+PIK3CA
ERBB2
MET
SKCM - TCGA
(468 patients)
r = -0.54
P = 2.31 x10^{-11}
BRAF
NRAS
PTEN
MAP2K1
CDKN2A
RB1
COADREAD - MSKCC
(969 patients)
r = -0.79
P = 2.17 x10^{-31}
TP53
SMAD4
NRAS
SOX9
SMAD2
KRAS
APC
BRAF
SMAD3
PIK3CA
STAD - TCGA
(438 patients)
r = -0.78
P = 1.33 x10^{-14}
TP53
CDH1
SMAD4
RHOA
ERBB2
EPHB1
SKCM - MSKCC
(184 patients)
r = -0.87
P = 1.74 x10^{-11}
BRAF
NRAS
PTEN
LUSC - TCGA
(485 patients)
r = -0.56
P = 1.25 x10^{-5}
CDKN2A
NFE2L2
RB1
PIK3CA
TP53
PTEN
KEAP1
NOTCH1
RASA1
UCEC - TCGA
(531 patients)
TP53
KRAS
PIK3R1
PPP2R1A
RNF43
CCND1
FGFR2
SOX17
SPOP
FBXW7
NFE2L2
CTNNB1
ARID5B
PIK3R2
PIK3CA
CTCF
PTEN
r = -0.74
P = 6.17 x10^{-40}
Percentage of significant findings
(P < 0.05)
Mutation Load Association (MLA)

Figure S4

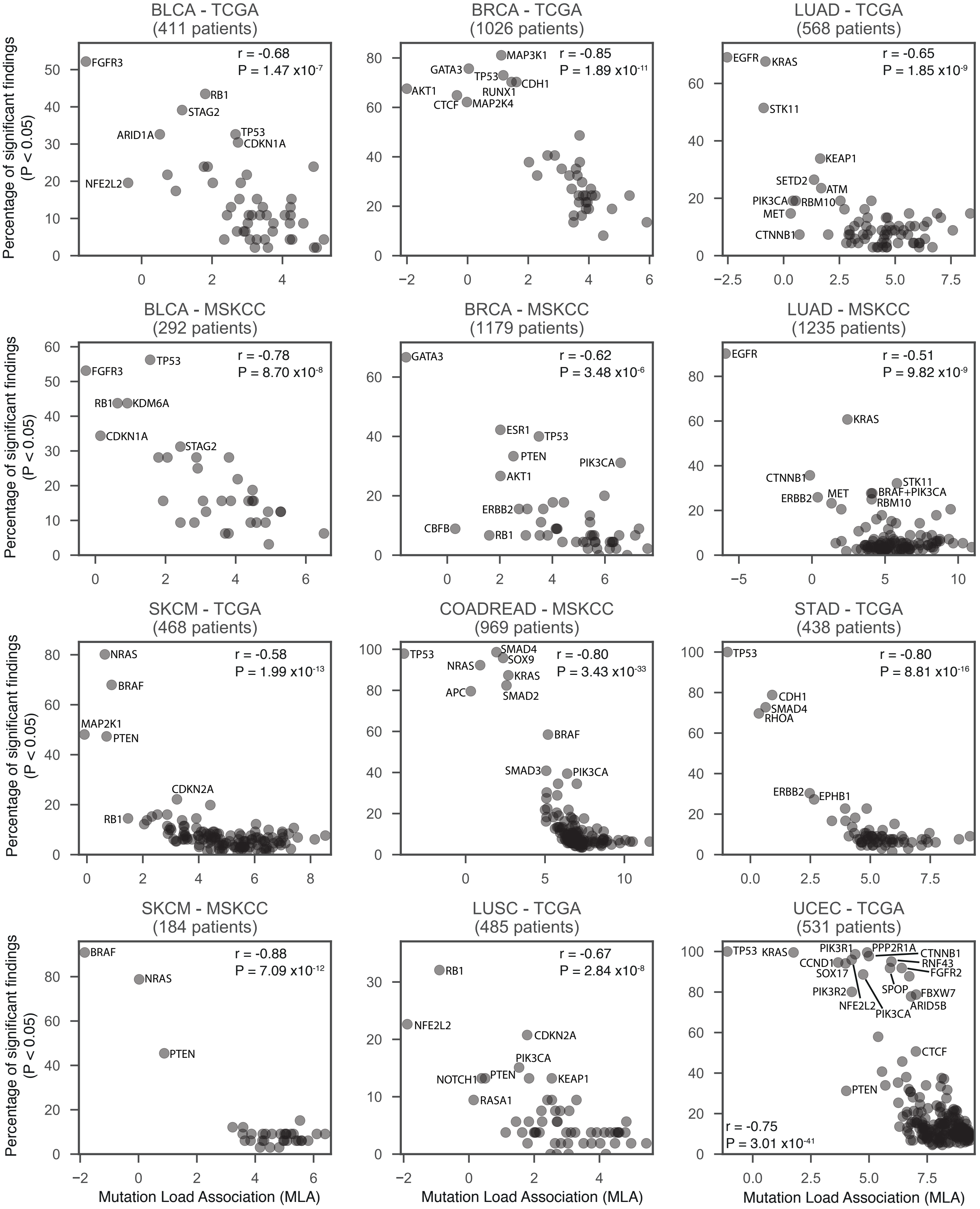

BLCA - TCGA
(411 patients)
$r = -0.68$
$P = 1.47 \times 10^{-7}$
FGFR3
RB1
STAG2
ARID1A
TP53
CDKN1A
NFE2L2
BRCA - TCGA
(1026 patients)
$r = -0.85$
$P = 1.89 \times 10^{-11}$
MAP3K1
GATA3
TP53
CDH1
RUNX1
AKT1
CTCF
MAP2K4
LUAD - TCGA
(568 patients)
$r = -0.65$
$P = 1.85 \times 10^{-9}$
EGFR
KRAS
STK11
KEAP1
SETD2
ATM
PIK3CA
RBM10
MET
CTNNB1
BLCA - MSKCC
(292 patients)
$r = -0.78$
$P = 8.70 \times 10^{-8}$
TP53
FGFR3
RB1
KDM6A
CDKN1A
STAG2
BRCA - MSKCC
(1179 patients)
$r = -0.62$
$P = 3.48 \times 10^{-6}$
GATA3
ESR1
TP53
PTEN
PIK3CA
AKT1
ERBB2
CBFB
RB1
LUAD - MSKCC
(1235 patients)
$r = -0.51$
$P = 9.82 \times 10^{-9}$
EGFR
KRAS
CTNNB1
STK11
MET
BRAF+PIK3CA
ERBB2
RBM10
SKCM - TCGA
(468 patients)
$r = -0.58$
$P = 1.99 \times 10^{-13}$
NRAS
BRAF
MAP2K1
PTEN
CDKN2A
RB1
COADREAD - MSKCC
(969 patients)
$r = -0.80$
$P = 3.43 \times 10^{-33}$
TP53
SMAD4
SOX9
NRAS
KRAS
APC
SMAD2
BRAF
SMAD3
PIK3CA
STAD - TCGA
(438 patients)
$r = -0.80$
$P = 8.81 \times 10^{-16}$
TP53
CDH1
SMAD4
RHOA
ERBB2
EPHB1
SKCM - MSKCC
(184 patients)
$r = -0.88$
$P = 7.09 \times 10^{-12}$
BRAF
NRAS
PTEN
LUSC - TCGA
(485 patients)
$r = -0.67$
$P = 2.84 \times 10^{-8}$
RB1
NFE2L2
CDKN2A
PIK3CA
NOTCH1
PTEN
KEAP1
RASA1
UCEC - TCGA
(531 patients)
TP53
KRAS
PIK3R1
PPP2R1A
CTNNB1
CCND1
RNF43
SOX17
FGFR2
SPOP
PIK3R2
FBXW7
NFE2L2
ARID5B
PIK3CA
CTCF
PTEN
$r = -0.75$
$P = 3.01 \times 10^{-41}$
Percentage of significant findings
($P < 0.05$)
Mutation Load Association (MLA)

Figure S5

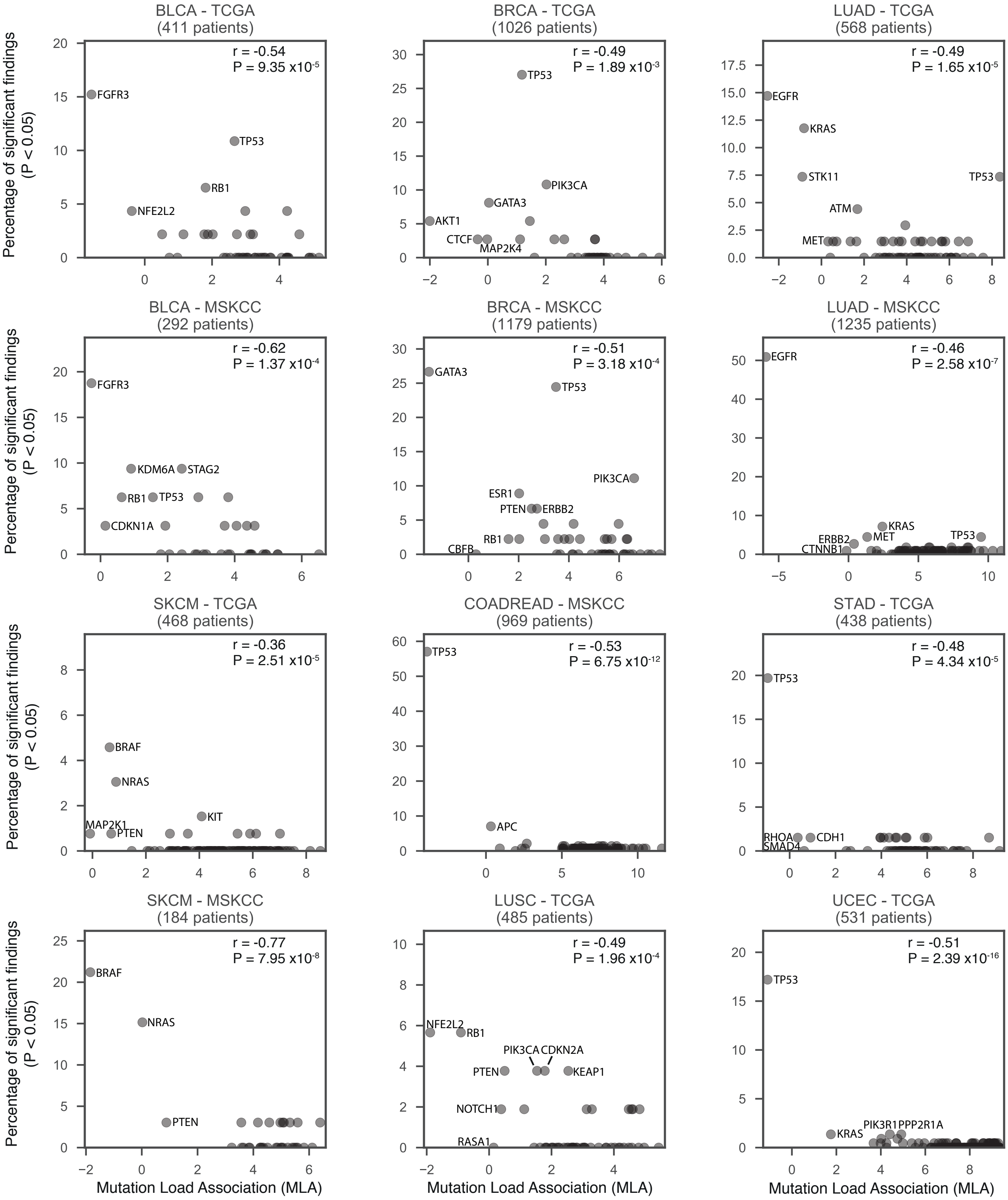

BLCA - TCGA
(411 patients)
r = -0.54
P = 9.35 x10^{-5}
FGFR3
TP53
RB1
NFE2L2
BRCA - TCGA
(1026 patients)
r = -0.49
P = 1.89 x10^{-3}
TP53
PIK3CA
GATA3
AKT1
CTCF
MAP2K4
LUAD - TCGA
(568 patients)
r = -0.49
P = 1.65 x10^{-5}
EGFR
KRAS
STK11
TP53
ATM
MET
BLCA - MSKCC
(292 patients)
r = -0.62
P = 1.37 x10^{-4}
FGFR3
KDM6A
STAG2
RB1
TP53
CDKN1A
BRCA - MSKCC
(1179 patients)
r = -0.51
P = 3.18 x10^{-4}
GATA3
TP53
PIK3CA
ESR1
PTEN
ERBB2
RB1
CBFB
LUAD - MSKCC
(1235 patients)
r = -0.46
P = 2.58 x10^{-7}
EGFR
KRAS
ERBB2
MET
CTNNB1
TP53
SKCM - TCGA
(468 patients)
r = -0.36
P = 2.51 x10^{-5}
BRAF
NRAS
KIT
MAP2K1
PTEN
COADREAD - MSKCC
(969 patients)
r = -0.53
P = 6.75 x10^{-12}
TP53
APC
STAD - TCGA
(438 patients)
r = -0.48
P = 4.34 x10^{-5}
TP53
RHOA
SMAD4
CDH1
SKCM - MSKCC
(184 patients)
r = -0.77
P = 7.95 x10^{-8}
BRAF
NRAS
PTEN
LUSC - TCGA
(485 patients)
r = -0.49
P = 1.96 x10^{-4}
NFE2L2
RB1
PIK3CA
CDKN2A
PTEN
KEAP1
NOTCH1
RASA1
UCEC - TCGA
(531 patients)
r = -0.51
P = 2.39 x10^{-16}
TP53
KRAS
PIK3R1
PPP2R1A
Percentage of significant findings (P < 0.05)
Mutation Load Association (MLA)

Figure S6

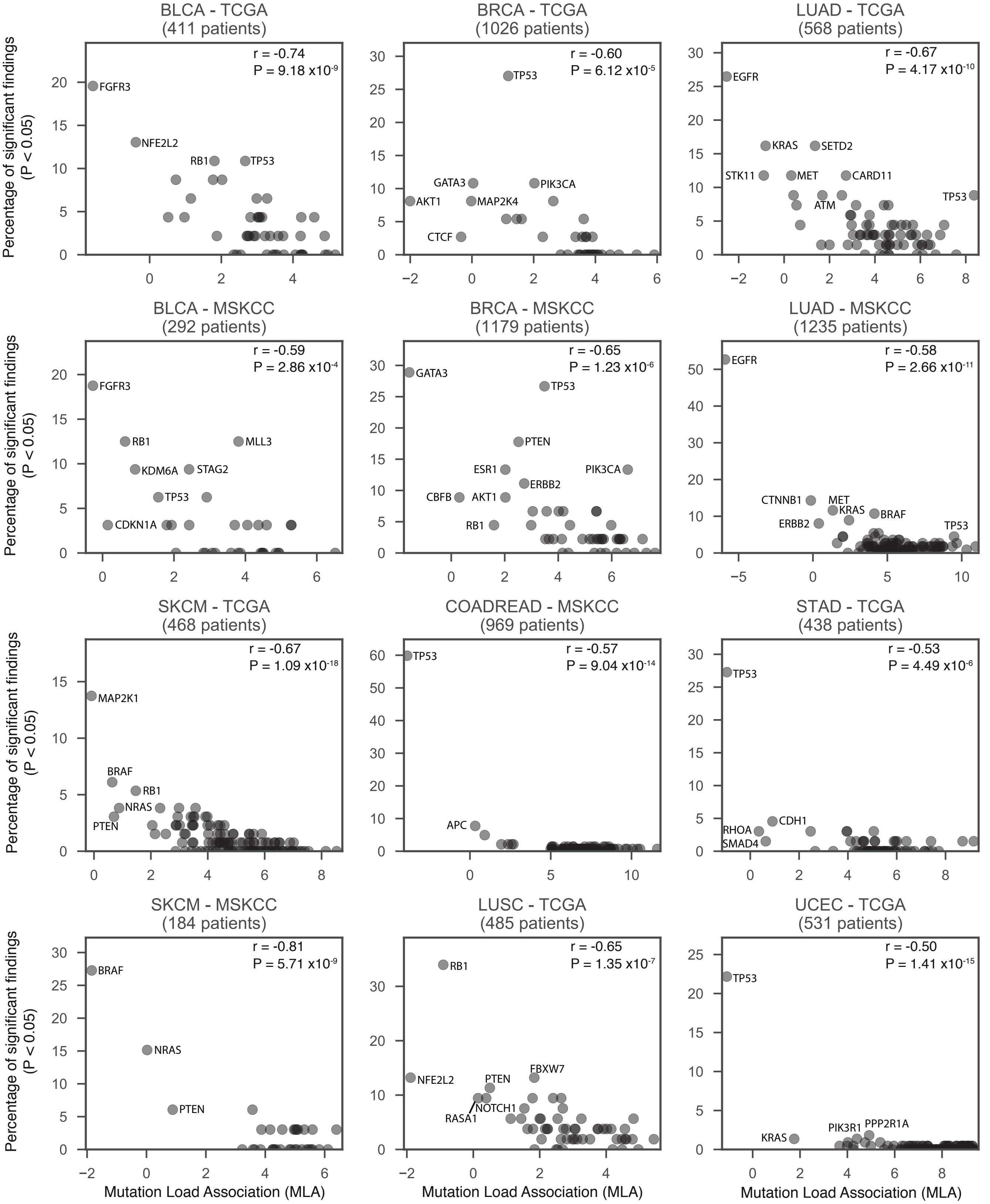
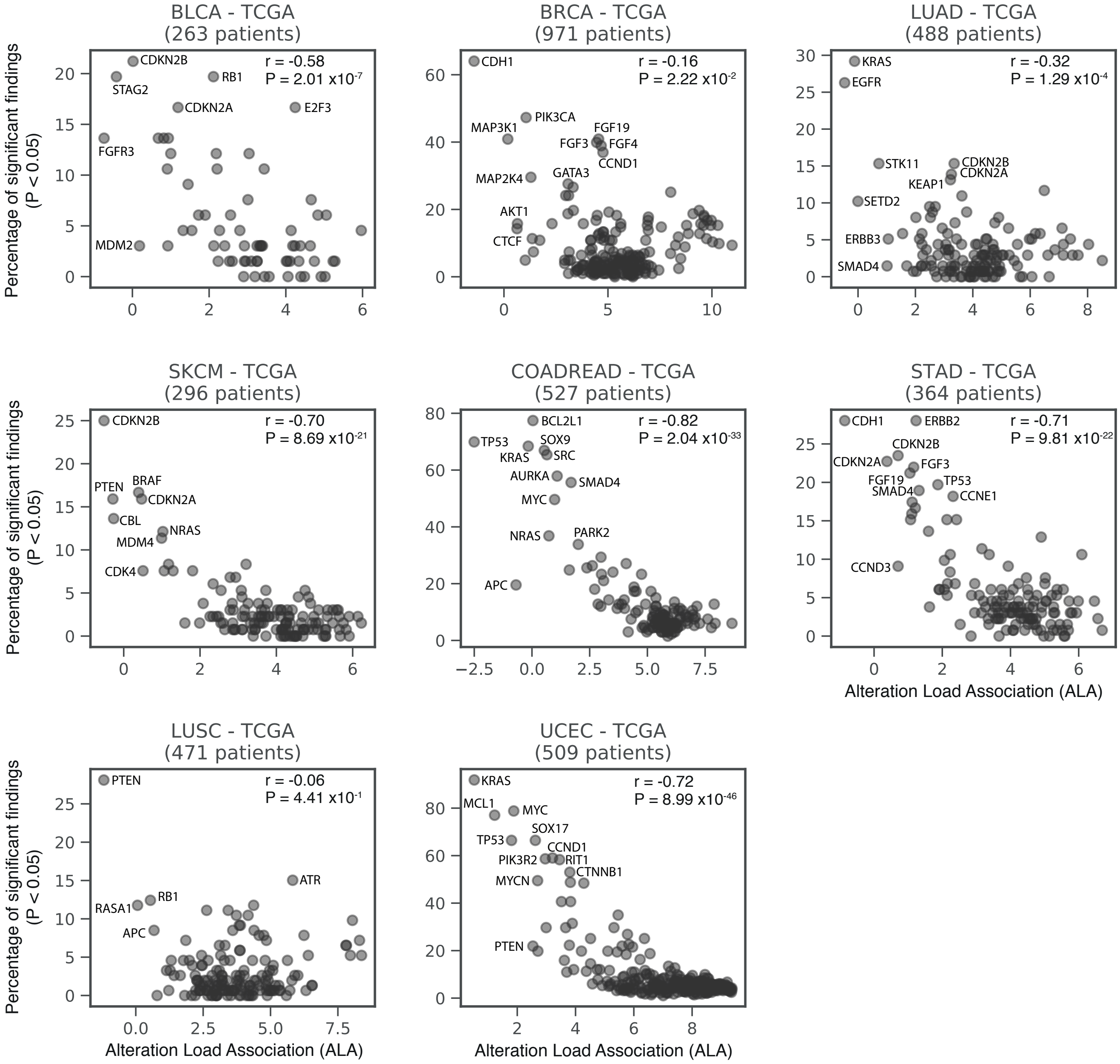

BLCA - TCGA
(411 patients)
r = -0.74
P = 9.18 x10^{-9}
FGFR3
NFE2L2
RB1
TP53
BRCA - TCGA
(1026 patients)
r = -0.60
P = 6.12 x10^{-5}
TP53
GATA3
PIK3CA
AKT1
MAP2K4
CTCF
LUAD - TCGA
(568 patients)
r = -0.67
P = 4.17 x10^{-10}
EGFR
KRAS
SETD2
STK11
MET
CARD11
ATM
TP53
BLCA - MSKCC
(292 patients)
r = -0.59
P = 2.86 x10^{-4}
FGFR3
RB1
MLL3
KDM6A
STAG2
TP53
CDKN1A
BRCA - MSKCC
(1179 patients)
r = -0.65
P = 1.23 x10^{-6}
GATA3
TP53
PTEN
ESR1
PIK3CA
ERBB2
CBFB
AKT1
RB1
LUAD - MSKCC
(1235 patients)
r = -0.58
P = 2.66 x10^{-11}
EGFR
CTNNB1
MET
KRAS
BRAF
ERBB2
TP53
SKCM - TCGA
(468 patients)
r = -0.67
P = 1.09 x10^{-18}
MAP2K1
BRAF
RB1
NRAS
PTEN
COADREAD - MSKCC
(969 patients)
r = -0.57
P = 9.04 x10^{-14}
TP53
APC
STAD - TCGA
(438 patients)
r = -0.53
P = 4.49 x10^{-6}
TP53
CDH1
RHOA
SMAD4
SKCM - MSKCC
(184 patients)
r = -0.81
P = 5.71 x10^{-9}
BRAF
NRAS
PTEN
LUSC - TCGA
(485 patients)
r = -0.65
P = 1.35 x10^{-7}
RB1
NFE2L2
FBXW7
PTEN
NOTCH1
RASA1
UCEC - TCGA
(531 patients)
r = -0.50
P = 1.41 x10^{-15}
TP53
PIK3R1
PPP2R1A
KRAS
Percentage of significant findings
(P < 0.05)
Mutation Load Association (MLA)

Figure S7


BLCA - TCGA
(263 patients)
r = -0.58
P = 2.01 x10-7
CDKN2B
STAG2
RB1
CDKN2A
E2F3
FGFR3
MDM2
BRCA - TCGA
(971 patients)
r = -0.16
P = 2.22 x10-2
CDH1
PIK3CA
MAP3K1
FGF19
FGF3
FGF4
CCND1
GATA3
MAP2K4
AKT1
CTCF
LUAD - TCGA
(488 patients)
r = -0.32
P = 1.29 x10-4
KRAS
EGFR
STK11
CDKN2B
CDKN2A
KEAP1
SETD2
ERBB3
SMAD4
SKCM - TCGA
(296 patients)
r = -0.70
P = 8.69 x10-21
CDKN2B
PTEN
BRAF
CDKN2A
CBL
NRAS
MDM4
CDK4
COADREAD - TCGA
(527 patients)
r = -0.82
P = 2.04 x10-33
BCL2L1
TP53
SOX9
KRAS
SRC
AURKA
SMAD4
MYC
NRAS
PARK2
APC
STAD - TCGA
(364 patients)
r = -0.71
P = 9.81 x10-22
CDH1
ERBB2
CDKN2B
CDKN2A
FGF3
FGF19
SMAD4
TP53
CCNE1
CCND3
LUSC - TCGA
(471 patients)
r = -0.06
P = 4.41 x10-1
PTEN
ATR
RB1
RASA1
APC
UCEC - TCGA
(509 patients)
r = -0.72
P = 8.99 x10-46
KRAS
MCL1
MYC
SOX17
TP53
CCND1
PIK3R2
RIT1
CTNNB1
MYCN
PTEN
Percentage of significant findings (P < 0.05)
Alteration Load Association (ALA)