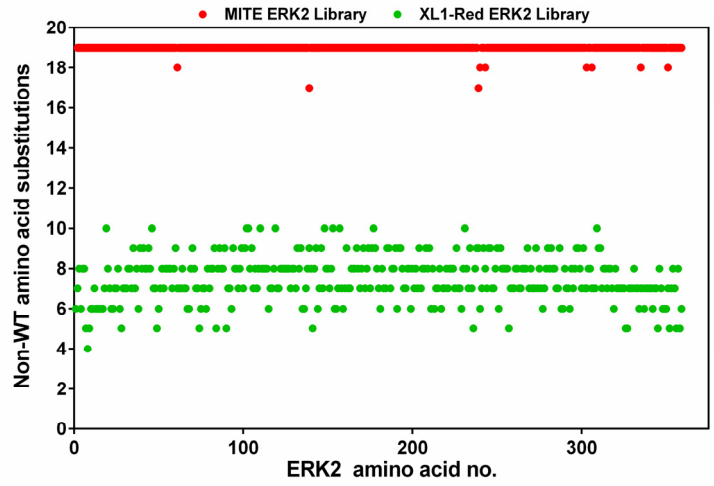
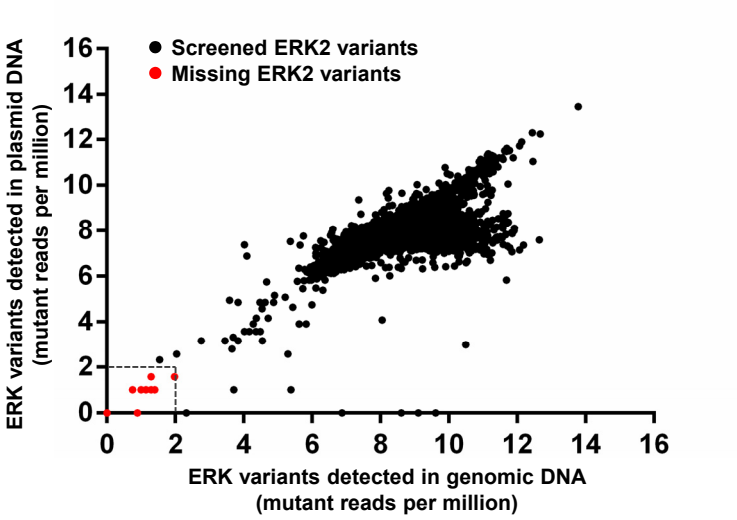
A



B



C

Missing ERK2 variants
H61F
Y139R
Y139W
H239T
H239W
I240V
I243L
E303H
Q306F
L335F
T351F

D

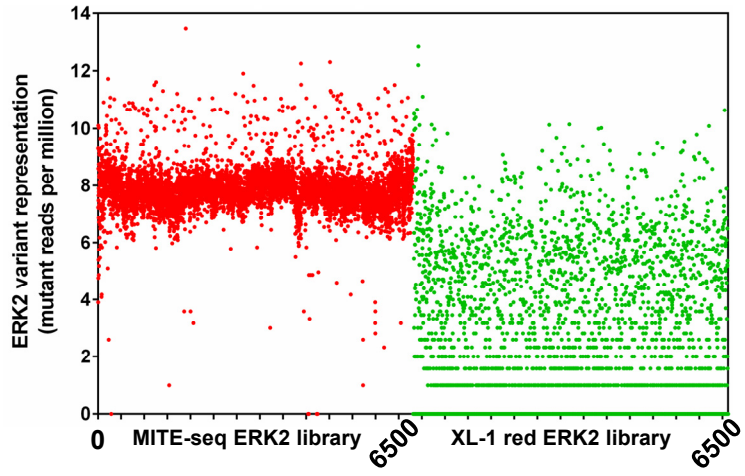


Figure S1

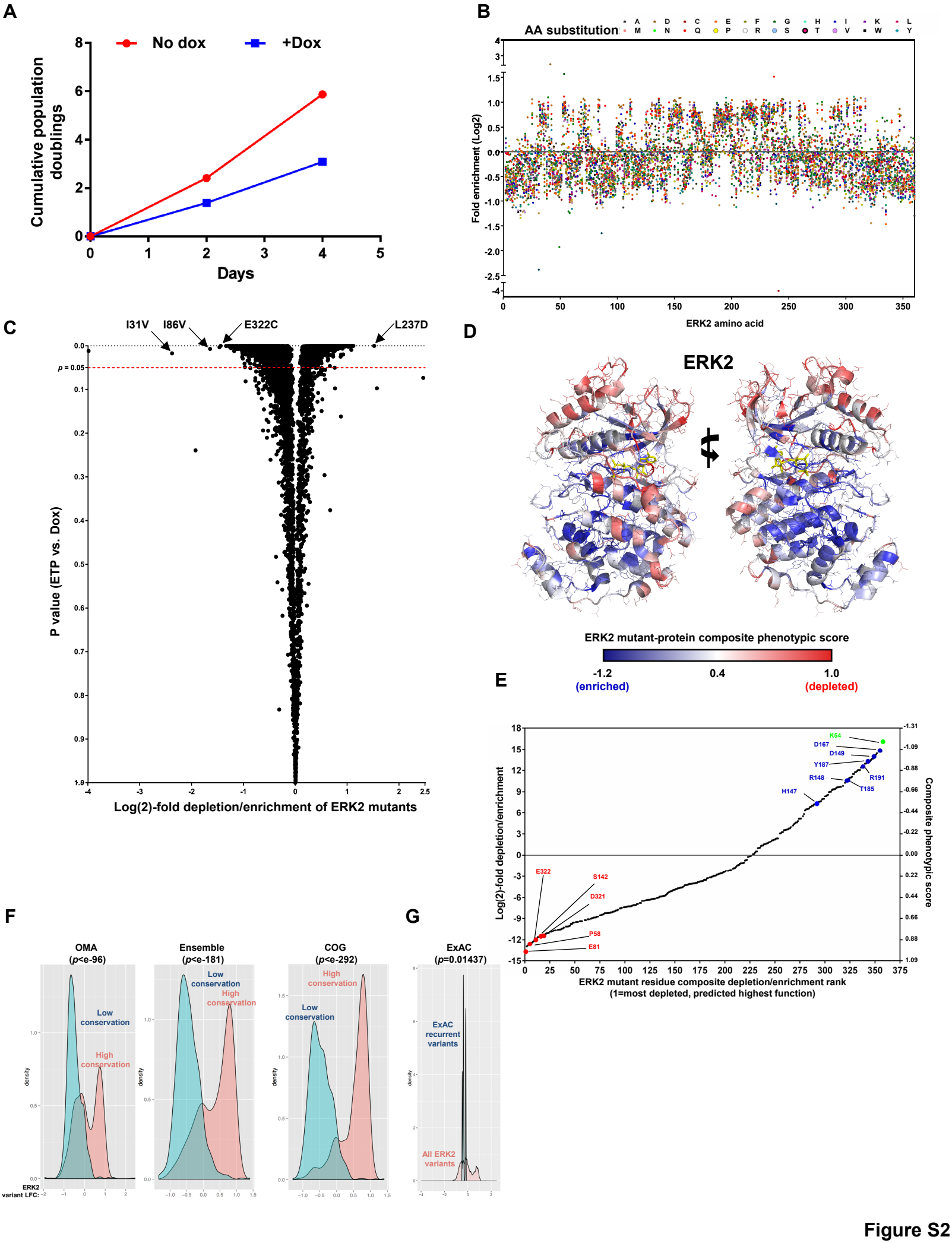

A
No dox
+Dox
Cumulative population doublings
Days
B
AA substitution
A M D N C Q E P F R G S H T I V K W L Y
Fold enrichment (Log2)
ERK2 amino acid
C
I31V
I86V
E322C
L237D
p = 0.05
P value (ETP vs. Dox)
Log(2)-fold depletion/enrichment of ERK2 mutants
D
ERK2
ERK2 mutant-protein composite phenotypic score
-1.2
(enriched)
0.4
1.0
(depleted)
E
Log(2)-fold depletion/enrichment
Composite phenotypic score
K54
D167
D149
Y187
R148
R191
T185
H147
E322
S142
D321
P58
E81
ERK2 mutant residue composite depletion/enrichment rank
(1=most depleted, predicted highest function)
F
OMA
(p<e-96)
Ensemble
(p<e-181)
COG
(p<e-292)
Low conservation
High conservation
density
ERK2 variant LFC:
G
ExAC
(p=0.01437)
ExAC recurrent variants
All ERK2 variants

Figure S2

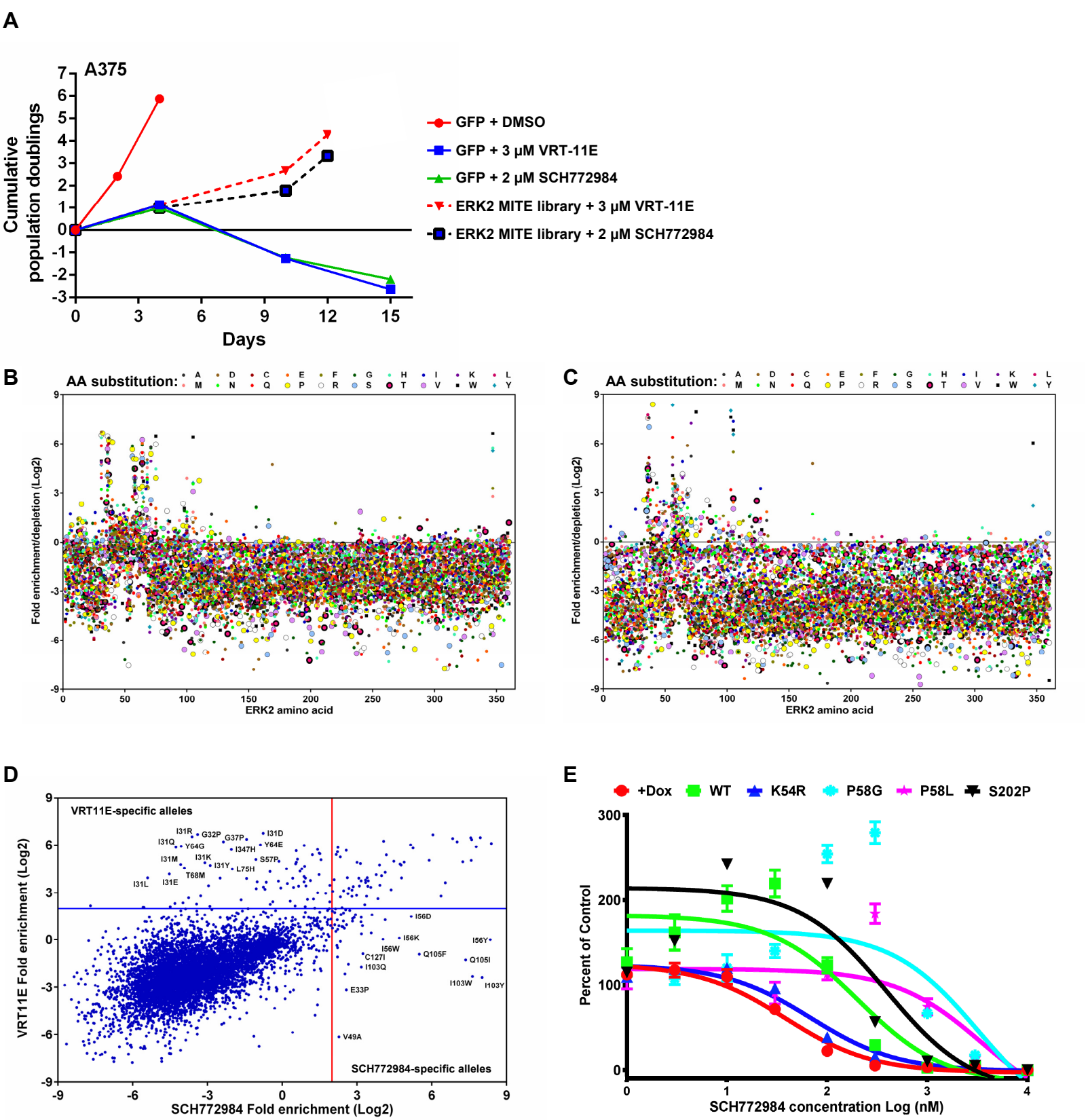
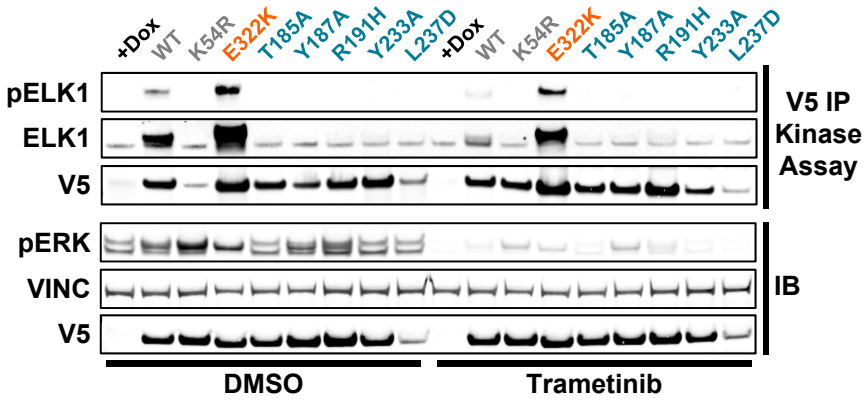

A
A375
Cumulative population doublings
Days
GFP + DMSO
GFP + 3 μM VRT-11E
GFP + 2 μM SCH772984
ERK2 MITE library + 3 μM VRT-11E
ERK2 MITE library + 2 μM SCH772984
B
AA substitution:
Fold enrichment/depletion (Log2)
ERK2 amino acid
C
AA substitution:
Fold enrichment/depletion (Log2)
ERK2 amino acid
D
VRT11E-specific alleles
SCH772984-specific alleles
VRT11E Fold enrichment (Log2)
SCH772984 Fold enrichment (Log2)
E
+Dox
WT
K54R
P58G
P58L
S202P
Percent of Control
SCH772984 concentration Log (nM)

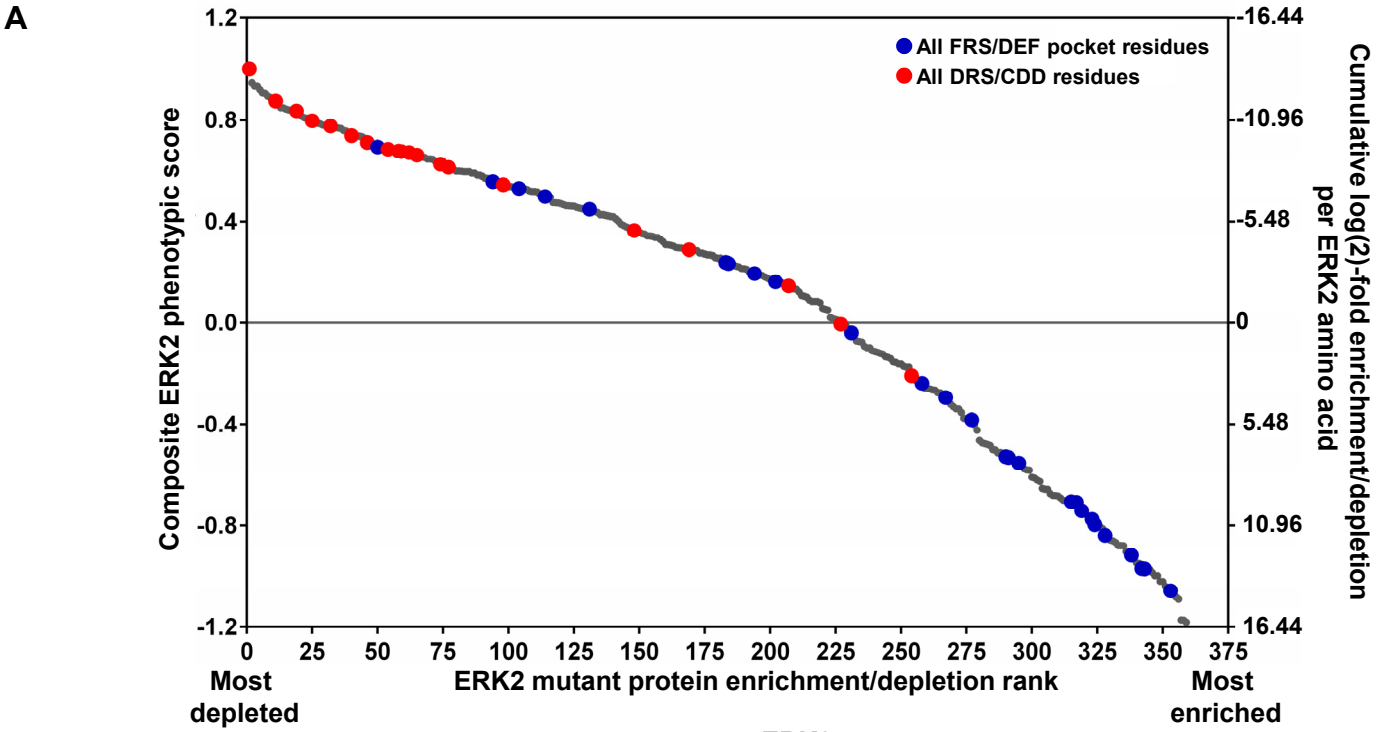
Figure S3

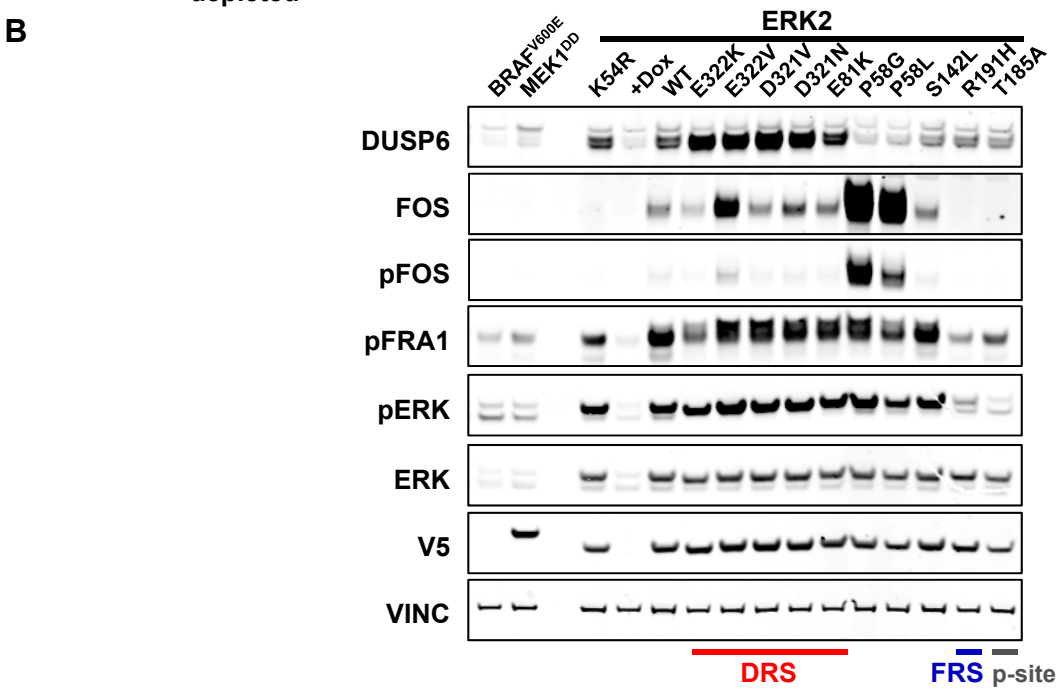
A



Negative/wild-type controls | Predicted GOF ERK2 mutants | Predicted LOF ERK2 mutants

Figure S4


A
All FRS/DEF pocket residues
All DRS/CDD residues
Composite ERK2 phenotypic score
Cumulative log(2)-fold enrichment/depletion per ERK2 amino acid
ERK2 mutant protein enrichment/depletion rank
Most depleted
Most enriched


B
ERK2
BRAFV600E
MEK1DD
K54R
+Dox
WT
E322K
E322V
D321V
D321N
E81K
P58G
P58L
S142L
R191H
T185A
DUSP6
FOS
pFOS
pFRA1
pERK
ERK
V5
VINC
DRS
FRS
p-site

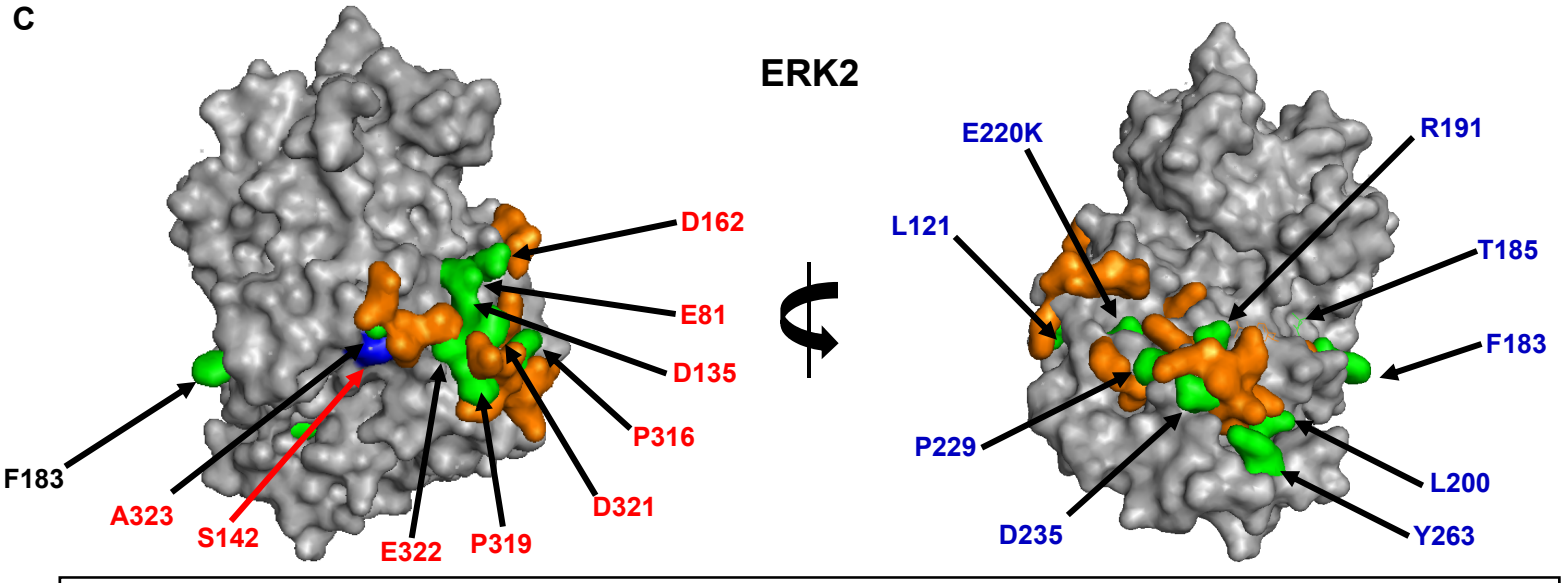

C
ERK2
D162
E81
D135
P316
D321
P319
E322
S142
A323
F183
E220K
R191
L121
T185
F183
P229
L200
D235
Y263
Non-recruitment site-residues
Recruitment site residues
Somatically-mutated recruitment site residues
Depleted/Enriched residues

Figure S5

Supplemental Figure and Table Title and Legends

Figure S1. Related to Figure 1. A saturation mutagenesis library of ERK2 mutants.

(A) Comparison of the number of detectable non-wild-type substitutions in ERK2 cloning libraries generated in XL-1 red repair-deficient bacteria (green circles) or using MITE (red circles). Data points represent the mean of 4 biological replicates for the ERK2 MITE library and a single replicate of the ERK2 XL-1 red library.
(B) Comparison of ERK2 variants detected in expression-plasmid DNA versus genomic DNA of transduced A375 cells. Variants highlighted in red (n=11) were undetectable in both conditions. Data points represent the mean of 4 biological replicates for genomic DNA detection.
(C) List of 11 ERK2 mutants highlighted in red in (B) that were undetectable in both expression-plasmid DNA and in genomic DNA of transduced A375 cells.
(D) Comparison of ERK2 variant representation in libraries generated using XL-1 red repair-deficient bacteria or MITE. Data points represent the mean of 4 biological replicates for the ERK2 MITE library and a single replicate of the ERK2 XL-1 red library.

Figure S2. Related to Figure 1. Damaging ERK2 mutations are enriched in internal and evolutionarily conserved residues.

(A) Cumulative population doubling of A375 cells transduced with the ERK2 mutant library in the presence (+Dox) or absence (No dox) of 1 µg/ml doxycycline.
(B) Log_2 fold-enrichment of individual ERK2 mutants across residues 2-360 of ERK2 in the presence of 1 µg/ml doxycycline. Each point represents the mean of 4 biological replicates and each substitution is colored by amino acid.
(C) Volcano plot of ERK2 log_2 fold-depletion/enrichment versus p-value (students T-test, 4 replicates). Mutants below the red dashed line have a p-value > 0.05 and are not statistically significant. See also **Figure S5**, **Table S1**.
(D) ERK2 3-dimensional protein structure colored according to the impact of residue mutation on ERK2 activity. Red indicates inferred GOF activity, whereas blue indicates inferred LOF activity.
(E) Scatter-plot showing that ERK2 mutant enrichment corresponds to LOF mutants, whereas ERK2 mutant enrichment corresponds to GOF mutants.
(F) Density plot of ERK2 mutant activity reveals different distributions for residues that are highly conserved versus residues that are lowly conserved in evolution across three datasets of ERK2 homologues.
(G) ERK2 mutations found in the germline of healthy humans have no impact on cellular proliferation.

Figure S3. Related to Figure 2 and Figure 3. Comprehensive identification of ERK2 mutants that are resistant to ERK-directed kinase inhibitors.

(A) Cumulative population doubling of A375 cells expressing a control ORF (GFP) or the ERK2 mutant library treated with either DMSO, VRT-11E (3 µM), SCH772984 (2 µM). Data points represent the mean of 4 biological replicates for VRT-11E and 6 biological replicates SCH772984.
(B) Log_2 fold-enrichment of individual ERK2 mutants across residues 2-360 of ERK2 in the presence of VRT-11E. Each point represents the mean of 4 biological replicates and each substitution is colored by amino acid.
(C) Log_2 fold-enrichment of individual ERK2 mutants across residues 2-360 of ERK2 in the presence of SCH772984. Each point represents the mean of 6 biological replicates and each substitution is colored by amino acid.
(D) Comparative log_2 fold-enrichment of individual ERK2 mutants in the presence of VRT-11E versus SCH772984. Data points represent the mean of 4 biological replicates for VRT-11E and 6 biological replicates for SCH772984.
(E) Proliferation assay of A375 expressing the indicated ERK2 constructs following treatment with the ERK inhibitor SCH772984.

Figure S4. Related to Figure 4. Identification and validation of unappreciated, tumor-associated GOF and LOF ERK2 mutants.

(A) *In vitro* kinase assay using V5-tagged ERK2 mutants immunoprecipitated from A375 cells treated with DMSO or 4 nM trametinib for 4 hours. Kinase activity is determined by the capacity to phosphorylate

purified ELK1. (pERK, phosphorylated ERK; VINC, vinculin; V5 IP KA, immunoprecipitated V5 epitope followed by kinase assay).

Figure S5. Related to Figure 6. Alteration of residues within the ERK2 DRS are associated with increased activity, whereas mutations in residues within the FRS are associated with decreased activity.

(A) Distribution of ERK2 D- and FRS mutants ranked by ERK2 activity. FRS mutants are noted in blue, DRS mutants in red. Data points represent the mean of 4 biological replicates.
(B) DRS ERK2 mutants are phosphorylated and induce DUSP6 expression, whereas FRS mutants behave like kinase impaired ERK2 (pERK, phosphorylated ERK; VINC, vinculin).
(C) Ser-142 of ERK2 is proximal to the DRS and its disruption reduces the capacity of DUSP6 to dephosphorylate mutant ERK2

Table S1. Related to Figure 1, Figure 2, Figure 3, Figure 4. Comprehensive phenotypic characterization of ERK2 mutants

Annotation of each ERK2 mutant with all raw and processed phenotypic data described in this study

Table S2. Related to Figure 1. Previously-characterized ERK2 mutants

List of previously identified ERK2 GOF and LOF mutants, references and their behavior in the functional screen presented here.

Table S3. Related to Figure 1, Figure 2, Figure 3. Comprehensive phenotypic characterization of ERK2 residues

Annotation of each ERK2 residue with protein domain, somatic mutations and residue-level phenotypic data described in this study

Table S4. Related to Figure 2, Figure 3. ERK2 drug-resistant mutants

A list of ERK2 mutants that confer resistance (LFC>2) to either VRT-11E, SCH772984 or both.

Table S5. Related to Figure 2, Figure 3. ERK2 drug-resistant residues

A list of ERK2 residues that, when mutated, confer resistance to either VRT-11E, SCH772984 or both inhibitors.

Table S6. Related to Figure 2, Figure 3. ERK2 residues that form contacts with VRT-11E or SCH772984.

A list of ERK2 residues that form contacts with VRT-11E or SCH772984.

Table S7. Related to Figure 4. Tumor-associated ERK2 mutants

Accounting of all aggregated ERK2 cancer-associated mutants and their sources.

EXTENDED EXPERIMENTAL PROCEDURES

ERK2 ORF synthesis

The *MAPK1* open reading frame and an in-frame Sh *ble* gene was synthesized and cloned into pUC57-KAN (GeneScript, Piscataway, NJ, USA) to create the wild-type cloning template (pUC57-MAPK1-ble) for subsequent mutagenesis by integrated tiles (MITE). Plasmid was isolated using the QIAGEN Miniprep kit (QIAGEN, Hilden, Germany) and sequence confirmed by Sanger sequencing.

Tile Pool Preparation

Non-overlapping oligonucleotide pools were synthesized using a CustomArray B3 synthesizer. Tile pools were generated by emulsifying the oligos with the tile-specific primer set and 2x Phusion Master Mix (NEB, Ipswich, MA, USA). Emulsions were split into 50 µl reaction volumes, thermal cycled (95°C for 1 minute; 36 cycles of 95°C 10 seconds, 59°C for 20 seconds, 72°C for 30 seconds and a final extension of 72°C for 7 minutes), re-pooled and broken. Emulsion making and breaking was performed using the Micella DNA Emulsion and Purification kit (EURx, Gdansk, Poland) according to manufacturer's specifications. The purified reactions were then electrophoresed on a 2% agarose gel, and the 150 base tiles were extracted by cutting, and purified using the QIAGEN Gel Extraction kit and protocol (QIAGEN, Hilden, Germany).

Entry vector backbone preparation

Primer sets flanking each tile region of *MAPK1* were combined with pUC57-MAPK1-ble in separate High Fidelity Phusion (NEB, Ipswich, MA, USA) PCR reactions. Each mix was thermal cycled (98°C for 1 minute; 28 cycles of 98°C 10 seconds, 60°C for 20 seconds, 72°C for 140 seconds and a final extension of 72°C for 7 minutes), electrophoresed on a 1% agarose gel and the appropriate nucleic acid band was cut from the gel and extracted as described above. Purified fragments were then DpnI treated and purified using a DNA Clean & Concentrator™-5 kit according to the manufacturer's guidelines (Zymo Research, Irvine, CA, USA).

Entry clone generation

Entry vector backbone and oligo tile pools were combined at a ratio of 1:5 and mixed with NEBuilder HiFi DNA Master Mix (NEB, Ipswich, MA, USA) in 20 µl reactions. The mixes were incubated at 50°C for 15 minutes, and then purified using the DNA Clean & Concentrator™-5 as previously described. The purified plasmid DNA was transformed into TG1 Electrocompetent cells (Lucigen, Middleton, WI, USA), and recovered in 975 µl of Recovery Medium with shaking at 37°C for 1 hour. 250 µl of the 1 hour transformation mix was then added to 50 mL of Luria broth containing 25 µg/µl of Zeocin (ThermoFisher Scientific, Waltham, MA, USA), and grown at 37°C with shaking overnight. Cells were harvested by centrifugation, and plasmid DNA was isolated using the QIAGEN Midi kit (QIAGEN, Hilden, Germany). Cloning libraries were then restriction enzyme cloned into pLiX403 (AddGene), a doxycycline-inducible lentiviral vector.

Cell lines and reagents

A375 cells were grown in RPMI-1640 (Cellgro), 10% FBS and 1% penicillin/streptomycin. HEK293T were grown in DMEM (Cellgro), 10% FBS and 1% penicillin/streptomycin. All cell lines were acquired via the Cancer Cell Line Encyclopedia (http://www.broadinstitute.org/ccle/home) and validated via Fluidigm SNP-based fingerprinting. Dabrafenib (PubChem ID: 44462760), trametinib (PubChem ID: 11707110), VRT-11E (PubChem ID: 11591420), and SCH772984 (PubChem ID: 24866313) were purchased from Selleck Chemicals. All small molecules were dissolved in DMSO.

Quantification of variant enrichment and depletion in high throughput screens

Non-wild-type mutations identified through massively-parallel sequencing were normalized to the total read counts per amino acid position. The rate of false-mutation calls were assessed via sequencing of a wild-type ERK2 library. Background mutations were then subtracted from the overall mutation abundance. In this assay, sequencing cannot disambiguate wild-type reads that come from fully wild-type ERK2 molecules from those emanating from the non-mutated portion of an ERK2 mutant. As such, wild-type reads are uninformative and were discarded. For each of 4-6 biological replicates, a $\log_2$ fold-change was calculated between initial, un-induced variant abundance in transduced A375 relative to doxycycline-alone ($LOG2_{[\text{un-induced }]} - LOG2_{[\text{DOX induced}]}$), doxycycline with SCH772984 ($LOG2_{[\text{un-induced }]} - LOG2_{[\text{DOX induced+SCH772984}]}$)and doxycycline with VRT-11E ($LOG2_{[\text{un-induced }]} - LOG2_{[\text{DOX induced+VRT11E}]}$). Log-fold change for each variant was calculated by the average of 4-6 replicates. Composite, residue-level annotation of ERK2 activity was calculated by summing the $\log_2$ fold-change of each substitution at every residue of ERK2 followed by normalization to the most depleted residue (Glu-81).

Lentiviral ERK2 expression in A375 cells

For lentiviral production, HEK293T cells (8×10^5 cells/6-cm dish) were transfected with 1 µg of pLiX403 ERK2 variant expression plasmid, 900 ng psPAX2 and 100 ng VSV-G using 6 µl Fugene6 transfection reagent (Promega). After 72 hours, viral supernatant was harvested, aliquoted and placed at -80°C for a minimum of 24 hours prior to transduction.

To create stable cell lines, A375 cell were infected with lentivirus at a dilution of 1:5 (virus: growth media) in 6-well plates (4.0×10^5 cells/well) in the presence of 7.5 µg/ml polybrene. Plates were then centrifuged at 2250 RPM and 37°C for 60 minutes. After centrifugation, media was removed and replaced with complete growth media. After incubation for 24-48 hours, cells were moved from the 6-well plate to 10-cm dishes and selected with puromycin at a concentration of 1 µg/ml. To induce expression of ERK mutants in A375, cells were incubated with doxycycline (Sigma-Aldrich) at a final concentration of 1 µg/ml.

Assessment of the residue-specific evolutionary selection of ERK2

The assessment of per-residue evolutionary conservation level is based on a likelihood-based estimation of the substitution rate across multiple homologs. Three dataset of homologs were compiled based on (1) OMA ("Orthologous MAtrix") database (Altenhoff et al., 2015), OMA Group 179474, spanning 69 proteins; (2) Ensembl database (Yates et al., 2016), ENSGT00550000074298 group, spanning 452 proteins including ERK2 paralogs based on EnsemblCompara GeneTrees algorithm (Vilella et al., 2009); (3) COG ("Clusters of Orthologous Groups") database, KOG0660 group ("Mitogen-activated protein kinase"), spanning 876 proteins including ERK2 paralogs, eggNOG v4.0 (Powell et al., 2014).

For each of these datasets of homologs the Multiple Sequence Alignment (MSA) of the member proteins was used for the likelihood-based evolutionary analysis. The evolutionary selection for each residue was computed with an empirical Bayesian algorithm (Mayrose et al., 2004) which explicitly model residue-specific evolutionary rates as implemented in ConSurf web-server (Ashkenazy et al., 2010). Residues were defined as "High conservation" or "Low conservation" with evolutionary rates below 10 percentile or above 90 percentile among all of the protein residues, respectively.

RNA isolation and sequencing

For RNA isolation, cells expressing ERK mutants were plated in 6-well plates (400,000 cells/well) in duplicate. The following day, ERK expression was induced with doxycycline as above. Twenty-four hours post-induction, half of the replicates were treated with 4nM trametinib for 4 hours. Subsequently, RNA isolation was performed using an RNeasy mini kit (Qiagen). Briefly, cells were lysed and lysate was cleared of cellular debris. RNA was precipitated from supernatant using 70% ethanol, then supernatant was applied to a spin column. The column was washed several times, then RNA was eluted from the column. Illumina RNASeq libraries were prepared from 100 ng of bulk RNA using the TruSeq RNA Sample Prep Kit (v2, Illumina). mRNA-seq libraries were sequenced on an Illumina NextSeq 500 sequencer both in biological and technical duplicates (total of 4 replicates per condition) 75 bp paired-end reads. Reads were mapped with STAR aligner (Dobin et al., 2013) with default parameters against hg19 of the human genome and build 12 of the GENCODE human genome annotation. Sample quality was assessed with the minimal number of mapped reads in pairs >4M (median 6.14 M), minimal number of different genes detected >16K (median of 17.67K). Differential expression analysis performed using *limma* package (Ritchie et al., 2015) with default parameters.

Principle component analysis of ERK2 mutant transcriptomes

Principle Components Analysis (PCA) was performed with the R implementations of the function prcomp() as part of the "stats" package using the default parameters. The PCA dimensionally reduction and was performed with the 5,000 top variable genes across all samples which were rank-normalized. The visualization of projection of samples to PC1 and PC2 was performed with ggfortify and ggplot2.

Antibodies used for immunoblots

Primary antibiodies recognizing vinculin (Sigma, V9131, 1:10,000), V5 epitope (Invitrogen, 37-7500, 1:1000) and HA epitope (Covance, MMS-101P, 1:1000). The following antibodies were purchased from Cell Signaling Technology and used at 1:1000 dilution: DUSP6 (#3058), phospho-ERK1/2 (Thr202/Tyr204, #4094), phospho-FRA1 (Ser265, #5841), phospho-cFOS (Ser32, #5348), ELK1 (#9182) and ERK1/2 (#9107). The following antibodies were purchased from Santa Cruz Biotechnology and used at 1:1000 dilution: FRA1 (# sc-28310), MEK1/2 (sc-81504), cFOS (sc-8047) and BRAF (sc-5284). After incubation with the appropriate secondary antibody (anti-rabbit or anti-mouse, IRDye-linked; 1:10,000 dilution; IRDye 800CW, IRDye 680, LiCOR), proteins were imaged using an Odyssey CLx scanner (LiCOR), converted to black and white and exported as JPEGS using LiCOR Image Studio. For quantitation of phosphorylated ERK, background phosphorylated ERK signal was quantified using LiCOR Image Studio and each mutant and condition normalized to that of wild-type ERK co-expressed with GFP.

Co-transfections

Transient transfections were performed in HEK293T in 6-well plates (4.0×10^5 cells/well). Twenty-four hours after plating, cells were transfected with 1 µg of ERK2 variant expression plasmid, 1 µg of $BRAF^{V600E}$ expression plasmid and 1 µg of either GFP or DUSP6 expression plasmid using 6 µl Fugene6 transfection reagent (Promega). Forty-eight hours after transfection, cells were washed once with ice-cold PBS and lysed with 1% NP-40 buffer [150 mM NaCl, 50 mM Tris pH 7.5, 2 mM EDTA pH 8, 25 mM NaF and 1% NP-40] containing 2x protease inhibitors (Roche) and 1x Phosphatase Inhibitor Cocktails I and II (CalBioChem). Immunoblotting was then performed as described above.

IP-kinase assays

For analysis of ERK activity, ERK-mutant expressing A375 were seeded at 50% confluence in 10-cm dishes in duplicate. Twenty-four hours after plating, all cells were incubated for a further 24 hours with doxycycline to induce ERK expression. After 24 hours of doxycycline exposure, DMSO or trametinib was added at a concentration of 4 nM to half of the replicates and incubated for 4 hours. After 4 hours of trametinib treatment cells were harvested in 1X lysis buffer (Cell Signaling #9803). ERK2 mutants were immunoprecipitated with 20 µl of V5-agarose beads (Sigma) from 200 µl of the cell lysate, followed by 18 hours of incubation on a rotator at 4 °C. Subsequently, beads were washed twice in ice-cold lysis buffer, then twice again in 1X kinase assay buffer (Cell Signaling, #9802). Next, beads were incubated at 30 °C for 30 min in 50 µl of kinase assay buffer with 250 µg ELK-1 fusion protein (Cell Signaling) and 100 µM ATP (Cell Signaling). Reactions were stopped by addition of 3X SDS loading buffer. Samples were reduced and denatured, then subjected to immunoblotting as described above.

SUPPLEMENTAL REFERENCES

Altenhoff, A.M., Skunca, N., Glover, N., Train, C.M., Sueki, A., Pilizota, I., Gori, K., Tomiczek, B., Muller, S., Redestig, H., *et al.* (2015). The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res *43*, D240-249.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Research *38*, W529-W533.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Mayrose, I., Graur, D., Ben-Tal, N., and Pupko, T. (2004). Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. Mol Biol Evol *21*, 1781-1791.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldon, T., Rattei, T., Creevey, C., Kuhn, M., *et al.* (2014). eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Research *42*, D231-D239.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y.F., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research *43*.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res *19*, 327-335.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., *et al.* (2016). Ensembl 2016. Nucleic Acids Research *44*, D710-D716.