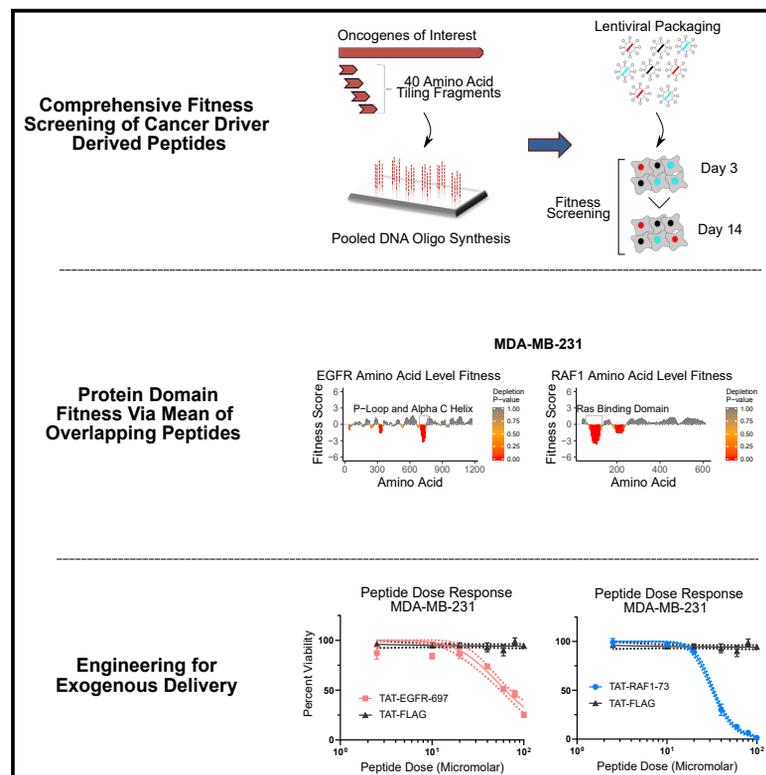


Cell Systems

Peptide-tiling screens of cancer drivers reveal oncogenic protein domains and associated peptide inhibitors

Graphical abstract



Authors

Kyle M. Ford, Rebecca Panwala, Dai-Hua Chen, Andrew Portell, Nathan Palmer, Prashant Mali

Correspondence

pmali@ucsd.edu

In brief

Understanding how protein domains impact cancer fitness is an ongoing challenge in oncology. Here, Ford et al. present a pooled screening methodology utilizing lentiviral overexpression of peptides comprehensively tiling cancer driver genes. This strategy identified bioactive domains and associated interfering peptides. Coupling cell-penetrating motifs to the peptides enabled exogenous drug-like delivery of these with dose-dependent anti-cancer activity.

Highlights

- Fitness screening of peptides extensively tiling cancer drivers in cancer cell lines
- *De novo* mapping of bioactive protein domains and associated interfering peptides
- Coupling cell-penetrating motifs to peptides enabled dose-dependent anticancer activity

Article

Peptide-tiling screens of cancer drivers reveal oncogenic protein domains and associated peptide inhibitors

Kyle M. Ford,¹ Rebecca Panwala,¹ Dai-Hua Chen,¹ Andrew Portell,¹ Nathan Palmer,² and Prashant Mali^{1,3,*}

¹Department of Bioengineering, University of California, San Diego, San Diego, CA 92093, USA

²Division of Biological Sciences, University of California, San Diego, San Diego, CA 92093, USA

³Lead contact

*Correspondence: pmali@ucsd.edu

<https://doi.org/10.1016/j.cels.2021.05.002>

SUMMARY

Gene fragments derived from structural domains mediating physical interactions can modulate biological functions. Utilizing this, we developed lentiviral overexpression libraries of peptides comprehensively tiling high-confidence cancer driver genes. Toward inhibiting cancer growth, we assayed ~66,000 peptides, tiling 65 cancer drivers and 579 mutant alleles. Pooled fitness screens in two breast cancer cell lines revealed peptides, which selectively reduced cellular proliferation, implicating oncogenic protein domains important for cell fitness. Coupling of cell-penetrating motifs to these peptides enabled drug-like function, with peptides derived from EGFR and RAF1 inhibiting cell growth at IC50s of 27–63 μ M. We anticipate that this peptide-tiling (PepTile) approach will enable rapid *de novo* mapping of bioactive protein domains and associated interfering peptides.

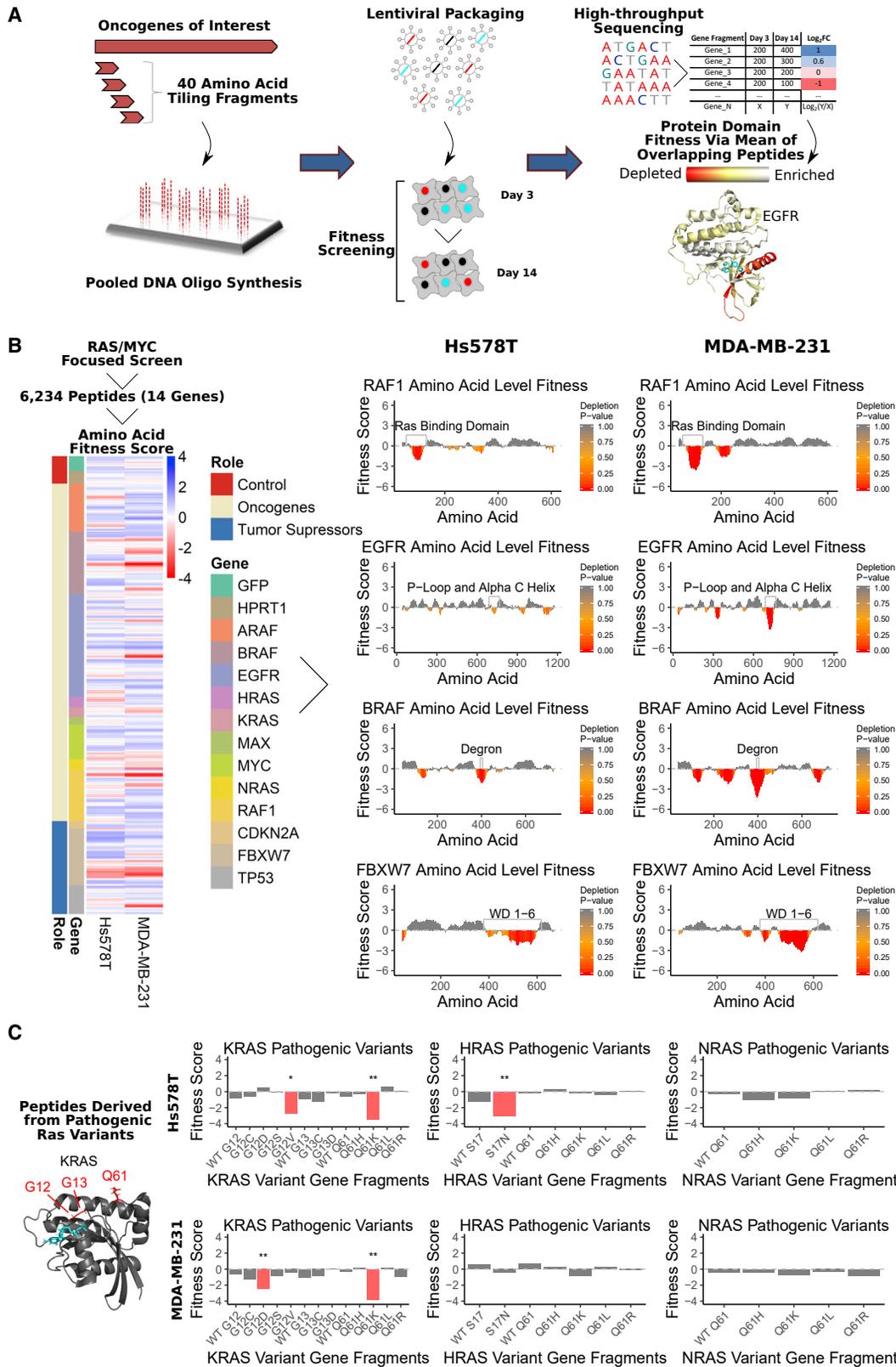
INTRODUCTION

Over the last decade, large-scale sequencing and functional genomic screening efforts have identified high-confidence lists of genes essential for cancer fitness. However, direct antagonism of many of these genes (Ras GTPases, transcription factors, cyclins, etc.) has proven challenging due to their reliance on large protein-protein interaction interfaces lacking a small molecule binding pocket to mediate signaling. Still, previous studies have demonstrated the feasibility of inhibiting hard to drug intracellular protein-protein interactions via direct transduction of protein/peptide therapeutics (Liu et al., 2010; Chang et al., 2013; Nim et al., 2016; Beaulieu et al., 2019). However, identifying and engineering protein/peptide therapeutics has classically relied on structure guided testing of individually produced protein variants. This process is time consuming and limited by the costs associated with direct peptide synthesis and recombinant production. Furthermore, target discovery itself is hindered in this context by the challenge of identifying therapeutically actionable protein-protein interaction interfaces. Subsequently, there is a compelling need for new technologies to identify and inhibit oncogenic signaling interfaces. With this in mind, here, we describe a modular oligonucleotide synthesis and sequencing-based screening protocol to identify bioactive peptides, which cause a slow growing phenotype, and corresponding protein-protein interaction domains implicated in driving cancer proliferation.

High-throughput screening strategies to identify novel proteins/peptides with a growth inhibition phenotype have been

previously explored, primarily in *Saccharomyces cerevisiae*. These studies include novel approaches to assay computationally defined C-terminal protein fragments (Nim et al., 2016), randomly digested genomic fragments (Ramer et al., 1992; Akada et al., 1997; Boyer et al., 2004), and, in a recent elegant approach, transposon-mediated fragmentation and overexpression of gene fragments to identify inhibitors of essential proteins in yeast (Dorrity et al., 2019). However, these libraries typically do not comprehensively cover protein-protein interaction interface regions for target proteins and often randomly generate gene fragments of various lengths and frame, hindering control of library composition. Consequently, these studies have been limited in their sensitivity, modularity, or ability to interrogate translatable phenotypes (Ramer et al., 1992; Akada et al., 1997; Boyer et al., 2004; Nim et al., 2016). As an alternative, purely computational methods to identify peptide self-inhibitors have been developed, but experimental screening is critical to progressively improving underlying structure-function predictions (London et al., 2010; Donsky and Wolfson, 2011; Zaidman and Wolfson, 2016; Han and Král, 2020).

To address these issues, we integrated lentiviral screening (Nim et al., 2016) and protein fragmentation (Dorrity et al., 2019) with array-based custom oligonucleotide pools (Kosuri and Church, 2014) to generate user-defined libraries of overexpressed peptide-coding gene fragments. We built our libraries using the target proteins as a scaffold from which to derive inhibitory sequences, synthesizing a comprehensive library of every possible overlapping 40-mer peptide for each target protein.



(legend on next page)

This strategy allows for modular library design, complete coverage of protein-protein interaction interfaces, and is supported by extensive previous research showing that fragmented or truncated proteins can function as inhibitors of the full-length protein (Herskowitz, 1987; Ramer et al., 1992; Akada et al., 1997; Barnard et al., 1998; Soucek et al., 2002; Boyer et al., 2004; Zhu et al., 2016; Nim et al., 2016; Bai et al., 2017; Yu et al., 2017; Dorrity et al., 2019). Furthermore, non-canonical translation of small ORFs overlapping protein coding genes has been shown to affect cell fitness, further supporting our strategy (Chen et al., 2020). We assayed these overexpression libraries via lentivirus-mediated pooled screening in two disease-relevant cell lines, interrogating over 65,000 peptides, tiling 65 cancer drivers and 579 mutant alleles. In contrast to contemporary approaches that employ libraries of genetically encoded functional perturbations that are agnostic to mechanism (CRISPR-Cas9 sgRNA, siRNA, etc.; Shalem et al., 2015; Doench, 2018; Ford et al., 2019), our approach enables rapid unbiased mapping of bioactive protein domains and associated interfering peptides.

RESULTS

Peptide-tiling-based map of protein domains implicated in proliferation via MAPK signaling

We first synthesized a pilot peptide library of oncogenes and associated effectors from the MAPK signaling pathways along with a panel of tumor suppressors and negative controls (Figures 1, S1, and S2, Table S1). RAS and MYC are two of the most frequently mutated/amplified oncogenes across a wide variety of malignancies, highlighting the medical need to identify functional inhibitors (Downward, 2003; Dang, 2012; Simanshu et al., 2017). Compounding this, RAS and MYC have proven challenging to drug via small molecules, due to their lack of a binding pocket and reliance on protein-protein interactions for signal transduction (Cox et al., 2014). Owing to their larger size and ability to form complex folded structures, we surmised that peptide biologics are likely suited to disrupting the protein-protein interactions through which RAS and MYC mediate cellular proliferation (Craik et al., 2013).

For every target protein in our library, we synthesized gene fragments via oligonucleotide pools coding for every possible overlapping 40-mer peptide within the proteins primary structure. Testing every overlapping 40-mer improves statistical power and allows for sensitive discrimination of similar peptide mo-

tifs, minimizing the required downstream optimization of inhibitors. To maximize the chance of identifying a peptide inhibitor of RAS or MYC signaling, we included gene fragments derived from the downstream RAS effectors ARAF, BRAF, and RAF1, as well as the negative regulator of MYC stability FBXW7. FBXW7 was of special interest due to its role in regulating the degradation of several other key oncogenes (Sato et al., 2015; Yeh et al., 2018). In addition to gene fragments derived from the wildtype (WT) RAS and MYC proteins, we also included fragments derived from pathogenic Ras variants that have been shown to have unique protein-protein interaction networks (Adhikari and Counter, 2018). Furthermore, we included gene fragments derived from EGFR (due to its role in proliferation and oncogenic signal transduction to Ras proteins), from the HRAS S17N dominant-negative and the MYC dominant-negative Omomyc (Soucek et al., 2002; Nassar et al., 2010). As negative controls, we included fragments derived from the green fluorescent protein (GFP) and hypoxanthine(guanine) phosphoribosyltransferase (HPRT1) (Gasperini et al., 2017). Finally, we included in the library two canonical tumor suppressor genes TP53 and CDKN2A. After removing duplicates, the final library consisted of 6,234 unique gene fragments, spanning 14 full-length genes. The pooled library of gene fragments was then synthesized as single-stranded oligonucleotides and cloned into a lentiviral vector, with an EF1 α -promoter-driving gene fragment transcription (Figures 1A, S1A, and S2; Table S1; STAR methods). An internal ribosomal entry site (IRES) was placed after the gene fragment stop codon to allow for co-translation of a puromycin acetyltransferase gene. This allowed for selection of transduced cells via the addition of puromycin to the cell culture media.

The library was then packaged into lentiviral particles that were used to transduce the MYC- and RAS-dependent Hs578T and MDA-MB-231 cell lines in duplicate (Figure 1A; STAR methods) (Eckert et al., 2004; Kang et al., 2014; Broad Institute, 2019). Genomic DNA was isolated 3 days after transduction, as well as 14 days after transduction to calculate peptide-specific log₂ fold changes. These peptide-specific log₂ fold change values were then used to calculate an amino-acid-level fitness score (Dorrity et al., 2019) via the mean of all fragments that overlap a particular codon (Figures 1A, 1B, and S1B-S1E; Table S2; STAR methods). The amino-acid-level fitness score was first calculated by taking the mean log₂ fold change of all overlapping peptides. For every residue in the

Figure 1. Peptide overexpression screening strategy and MAPK focused library

(A) Design of overlapping peptide expression library. Gene fragments coding for all possible overlapping 40-mer peptides were computed from target gene cDNA sequences. Peptide-coding sequences were then generated via chip-based oligonucleotide synthesis and cloned into a lentiviral plasmid vector. This plasmid library was in turn used to generate lentiviral particles via transient transfection. The lentiviral particles were then used to infect target mammalian cell lines at a low multiplicity of infection (MOI) to ensure only one peptide was expressed per cell. The cells were then grown for 2 weeks, with genomic DNA extracted at days 3 and 14. Next, peptide-coding gene fragments were PCR amplified from genomic DNA and sequenced to track peptide abundances and calculate log₂ enrichment and depletion. Peptides were mapped back to target gene coding sequences, and each codon/amino acid was given a fitness score defined as the Z-normalized mean log₂ fold change of all overlapping peptides.

(B) Resulting amino-acid-level fitness scores. Screening data from Hs578T and MDA-MB-231 cells shows conserved regions of peptide depletion, as well as cell line specific peptide depletion. The heatmap shows the fitness score for each amino acid position (sorted in ascending order from top to bottom) across all proteins assayed in the screen. On the right, plots showing the statistical likelihood of depletion are shown for RAF1, EGFR, BRAF, and FBXW7. Peptides overlapping amino acid positions with known functional roles are significantly depleted over the course of cell growth.

(C) The fitness effects of peptides derived from known pathogenic and dominant-negative Ras mutants. Peptides derived from KRASQ61K were significantly depleted in both cell lines, while peptides derived from HRAS S17N is depleted only in HRAS mutant Hs578T cells (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001) (see also Figures S1 and S2).

protein scaffolds, the mean \log_2 fold change values were then Z-score-normalized to yield a relative fitness score. This fitness score served as a way to map the results of individual peptides back to the original protein structure. Based on this, 2.6% (Hs578T) and 9.5% (MDA-MB-231) of residues tested had significantly depleted overlapping peptides, indicating that peptides derived from these positions were collectively more deleterious to cell fitness than a random sampling of peptides from the library (Figures S1D and S1E). There was good correlation between biological replicates, with the Hs578T and MDA-MB-231 amino acid scores having a Pearson correlation of 0.54 and 0.75, respectively.

In order to visualize protein motifs with a significant impact on cell fitness, the amino acid scores were superimposed along the primary amino acid sequence for each associated protein (Figure 1B). EGFR, BRAF, FBXW7, and RAF1 all had regions of significant depletion in one or both of the cell lines, corresponding to previously annotated protein function. Peptides derived from the P-loop and alpha C-helix of EGFR were depleted across both cell lines. The P-loop of EGFR is involved in ATP binding, while the conformationally sensitive autoinhibitory C-helix plays a regulatory role in controlling EGFR enzymatic activity (Yun et al., 2007; Ruan and Kannan, 2018). The EGFR alpha C-helix regulates EGFR activation by dynamic orientation toward the ATP-binding pocket (active state), or away from the ATP-binding pocket (inactive state). Supporting a functional role for this depleted EGFR domain in regulating cell fitness, this region of the EGFR gene (exon 19) is frequently deleted in cancer, comprising approximately 44% of activating EGFR mutations seen clinically (Kumar et al., 2008). Maintaining an active EGFR structural state critically depends on the positioning of the alpha C-helix structure, suggesting that overexpressed alpha-C-helix-derived peptides may be active participants in allosteric EGFR regulation. However, because alpha-C-helix motifs are ubiquitous in regulating kinase activity (Palmieri and Rastelli, 2013), homologous protein motifs on other structures may also be implicated in mediating EGFR-derived peptide bioactivity.

The Ras-binding domain (RBD) of RAF1 was also significantly depleted across both cell lines, presumably due to the peptides binding endogenous Ras proteins within the cell. This result is supported by previous research showing that chemically synthesized and recombinant RAF1 RBD mini proteins can bind Ras proteins with nanomolar affinity (Clark et al., 1996; Barnard et al., 1998; Becker et al., 2003). Ras-targeting peptides derived from RAF1 have also been shown to block oncogenic signaling, lending further credence to this hypothesis. While the RAF1 cysteine-rich domain (AA 139 to 184) has also been previously identified as a KRAS binder, this region does not correspond to significant peptide depletion in either breast cancer cell line. This result is potentially due to the orders of magnitude lower binding affinity of the cysteine-rich domain compared with the RBD (micromolar versus nanomolar affinity) (Williams et al., 2000).

FBXW7 had a broad region of depletion corresponding to WD repeats 1–6. Knockout screening via CRISPR-Cas9 has shown that FBXW7 is not essential in Hs578T or MDA-MB-231 cells, meaning it is unlikely that this depletion is due to direct inhibition of FBXW7 (Broad Institute, 2019). The WD repeats in FBXW7 mediate substrate binding and subsequent recruitment to the

E3 ubiquitin-protein ligase complex, suggesting that the highly depleted peptides are potentially interacting with one of the endogenous partners of FBXW7 (Hao et al., 2007). BRAF also had several significantly depleted regions dispersed across the primary sequence including one corresponding to a previously identified phospho-degron motif centered on amino acids 394–405 (Eisenhardt et al., 2016).

Toward the broader goal of identifying peptide inhibitors of KRAS function, we tested if peptides derived from pathogenic variants could function as more effective anti-proliferative proteins than their WT counterparts (Figure 1C). The 40-mer peptides derived from KRAS Q61K were significantly depleted across both cell lines, while WT peptides overlapping amino acid showed no effect on cell fitness. The full-length Q61K mutant is highly transforming because of a modified Ras/Raf interaction, which may play a role in the anti-proliferative activity of the Q61K derived fragments (Der et al., 1986; Buhman et al., 2007). Furthermore, peptides derived from the known HRAS S17N dominant-negative mutant showed selective depletion only in the mutant HRAS-driven Hs578T cell line, emphasizing the ability of this technology to discriminate fitness dependencies with a degree of specificity.

Large-scale peptide-tiling screens identify diverse peptides and domains that perturb cell fitness

In order to mine anti-proliferative peptide motifs in a more systematic fashion we next synthesized a library of 43,441 peptides (Figures 2A and S3A; Tables S1 and S3) derived from 65 key oncogenic driver genes with a high prevalence in TCGA-sequencing data (Bailey et al., 2018). This library covers ~20% of all high-confidence cancer drivers identified in a recent computational approach, allowing for a more comprehensive characterization of potential oncogene-derived peptide inhibitors of proliferation (Bailey et al., 2018). This expanded screen was performed in MDA-MB-231 cells and identified nearly an order of magnitude greater number of peptides with fitness defects (as measured by \log fold change), compared with those identified in the smaller pilot screen (Figure 2A). Building on this screen of cancer drivers, we also built a library of peptides derived from high-confidence cancer driver mutations identified via the Cancer Genome Atlas sequencing data (Bailey et al., 2018). This screen interrogated 579 mutant residues across 53 cancer driver genes, via 22,724 peptide-coding gene fragments (Figures 2B and S3B; Tables S1 and S4). Peptide names indicate the gene from which the peptide was derived, and the first amino acid they align to on the full-length structure.

We observed in most cases that mutant peptides had a similar effect on cell fitness compared with their WT counterparts (Figure 2C). This can be rationalized by the high degree of sequence homology (>97%) between WT peptides and single mutants. We then quantified how peptide depletion in the screen relates to bulk biophysical properties such as charge and hydrophobicity (Figure 2D). We found that peptide effects on cell fitness were not dependent on charge or hydrophobicity, indicating that highly charged or highly hydrophobic peptides do not result in false-positive cellular toxicities.

As in the pilot screen, we then sought to map peptides from the library back to the primary structure of the WT protein to visualize domains with a significant impact on cell fitness (Figures

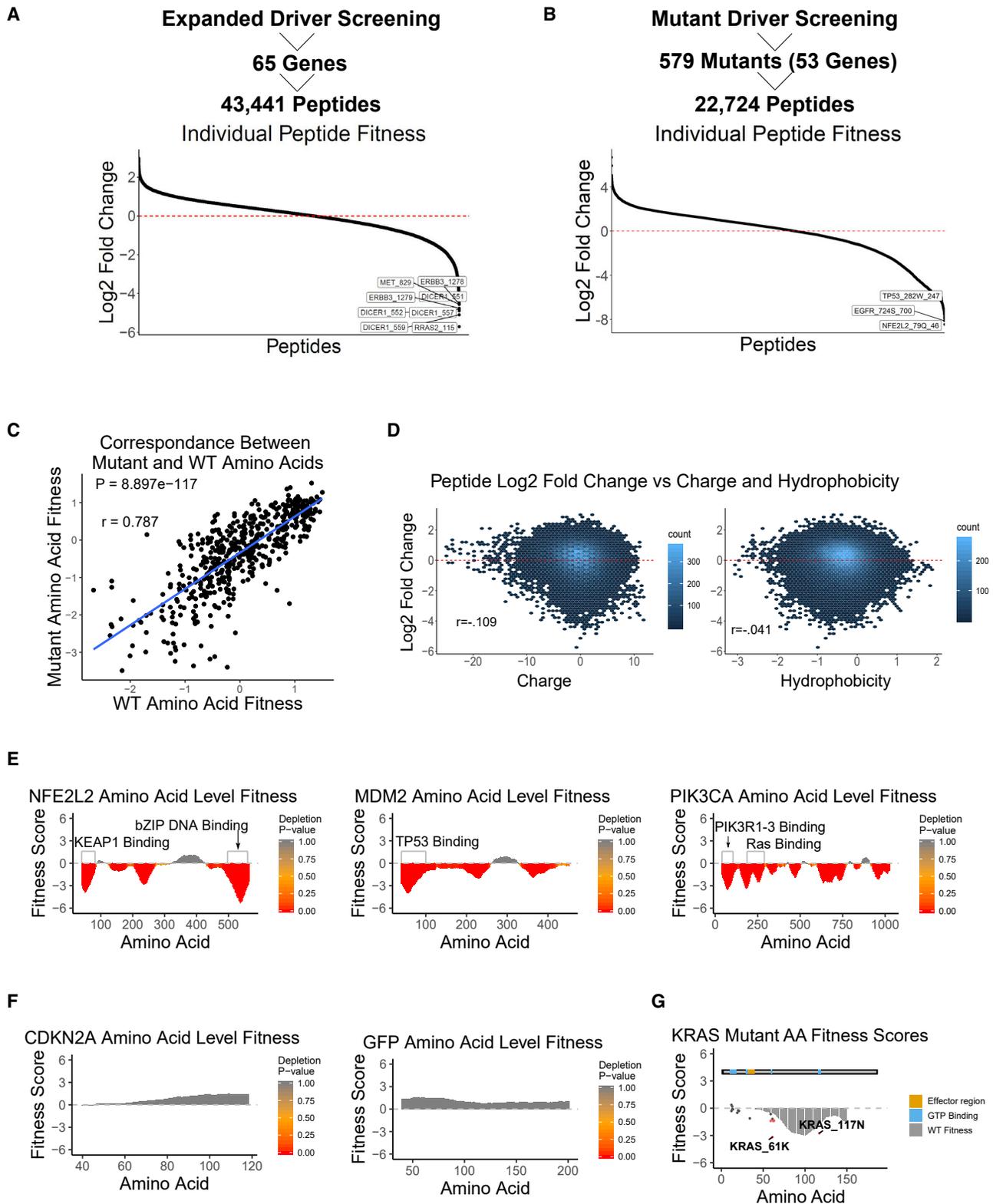


Figure 2. Expanded library screening enables more comprehensive evaluation of cancer driver derived peptides

(A) Plot of individual peptide enrichment/depletion for expanded screen. Peptides are centered around zero depletion, with a subpopulation being significantly deleterious to cells when overexpressed genetically. Peptides with \log_2 fold change values less than -4.5 are labeled. Cancer driver genes were hand curated from (Bailey et al., 2018) and (Santarius et al., 2010), with additional controls added from the pilot screen.

(legend continued on next page)

2E–2G, S4A, and S4B). We first examined the pattern of depletion for the transcription factor NFE2L2, the protein containing the most deleterious domain as scored by this screen (Figures S4C and 2E). Peptides derived from the DNA-binding domain, as well as the KEAP1-binding domain of NFE2L2 were highly depleted in the screen, consistent with the critical role these regions play in mediating NFE2L2 function (Cuadrado et al., 2019). NFE2L2 has been previously shown to support cellular proliferation and metastasis in MDA-MB-231 cells, supporting the conclusion that peptide mediated disruption of NFE2L2 function could be used to inhibit cell growth (Zhang et al., 2016). Neither the negative control GFP protein or the tumor suppressor CDKN2A showed significant depletion of any domain, highlighting the ability of this technology to discriminate bioactive peptide motifs (Figure 2F). While the majority of mutant peptides had similar fitness scores compared with WT peptides overlapping the same residues, some mutants such as PIK3CA956F, KRAS61K, and BRAF594N showed markedly more deleterious effects on cell fitness (Figures 2G and S4E).

We next investigated the fitness of peptides derived from MDM2. MDM2 is a negative regulator of TP53 function in the cell, and inhibition of the MDM2-TP53 PPI has been shown to effectively oppose cancer growth across a variety of malignancies (Liu et al., 2010; Chang et al., 2013; Zhao et al., 2015). In our screening data, peptides derived from the TP53-binding domain of MDM2 were significantly depleted consistent with previous reports that truncated MDM2 proteins containing only the N terminus function as dominant negatives (Kubbutat et al., 1999). However, interpreting the bioactivity of MDM2 derived peptides is made challenging by the highly contextual MDM2 and TP53 biological functions. For example, MDA-MB-231 cells contain a TP53 hotspot mutation (R280K) (Chavez et al., 2010) obfuscating if putative TP53-binding peptides are activating WT TP53 functions, or inhibiting oncogenic mutant TP53 functions (Hui et al., 2006). Given the TP53-binding domain of MDM2 occupies the transactivation domain of TP53, both hypotheses have a structural justification, highlighting the complex role TP53 plays in cancer etiology (Iwakuma and Lozano, 2003).

We then sought to investigate the fitness effect of peptides derived from PIK3CA. The PI3K-AKT-mTOR pathway is one of the most frequently dysregulated pathways in cancer, and PIK3CA plays a pivotal role in signal transduction along this pathway (Janku et al., 2018). The most critical region impacting cell fitness in PIK3CA corresponds to the adaptor-binding domain of the protein. PIK3CA activity is modulated by the binding of various adaptor proteins encoded by genes such as PIK3R1, PIK3R2, and PIK3R3. Supporting the hypothesis that

these peptides potentially inhibit proliferation via disruption of the PIK3CA/PIK3R1-3 complex, the corresponding PIK3CA-binding domain in PIK3R1 is also depleted. Additionally, the RBD of PIK3CA was also significantly depleted in this screen, implying Ras-PIK3CA cross-talk may impact cell fitness in MDA-MB-231 cells.

Next, we plotted the depleted domains for the miRNA-processing protein DICER1. Regions corresponding to binding sites for known DICER1 cofactors TARBP and PRKRA were heavily depleted, comprising some of the most deleterious peptides in the screen (Figure S4D). However, DICER1 activity is predicated not just on binding other proteins but also on binding RNA via helicase, RNase, and dsRNA-binding domains present throughout the protein structure (Gurtan et al., 2012). The deleterious nature of DICER1-derived peptides could therefore be attributed to protein-protein, as well as protein-RNA interactions. These data support the growing understanding of the oncogenic role miRNAs and other epigenetic regulators play in tumorigenesis (Rupaimoole and Slack, 2017).

ERBB4 had a pattern of depletion similar to EGFR (Figure S4D), with overexpression of peptides derived from the ERBB4 regulatory P-loop and alpha C-helix resulting in a significant fitness defect, highlighting the importance of this region in ERBB allosteric regulation and proliferative signaling (Bose and Zhang, 2009). This example also supports previous work suggesting that alpha C-helix displacement is a broadly shared (and therapeutically targetable) mechanism of regulating kinase activity (Palmieri and Rastelli, 2013). Further supporting this conclusion, alpha C-helix displacement has even seen clinical success in breast cancer via the small molecule EGFR/HER2/ERBB4 inhibitor Lapatanib (Palmieri and Rastelli, 2013).

Next, we sought to validate the anti-proliferative effects of select peptides identified as depleted in the screen via a complementary technology other than sequencing. Specifically, after transduction with putative anti-proliferative peptides derived from WT proteins, Hs578T cells and MDA-MB-231 cells were seeded in 96-well plates with proliferation measured via the colorimetric WST-8 assay (Figures 3A, S5A, and S5B; Table S6). All 12 peptides tested had significant growth defects when assayed in Hs578T and/or MDA-MB-231 cells compared with infection with the GFP control plasmid. EGFR-697 specifically was extremely harmful to cell growth in both cell lines. We similarly tested three peptides derived from the KRAS-Q61K-mutant protein (KRAS61K-24, KRAS61K-28, and KRAS61K-34), all of which significantly reduced cell growth in both cell lines (Figures S5A and S5B). To test the specificity of these perturbations, we transduced MCF-7 cells with RAF1-73 and EGFR-697. MCF-7

(B) Plot of individual peptide enrichment/depletion for mutant screen. 579 mutant cancer drivers covering 53 driver genes were assayed for growth inhibition as in (A). Peptides are centered around zero depletion, with a subpopulation being significantly deleterious to cells when overexpressed genetically. Peptides with \log_2 fold change values less than -8 are labeled.

(C) Correlation between WT and mutant amino acid fitness scores. There is a high correlation (Pearson $r = 0.787$) between WT and mutant amino acids.

(D) Plots showing the correlation between peptide depletion versus charge and hydrophobicity. There is little correlation between charge/hydrophobicity and peptide log fold change, indicating that gross physiochemical factors do not mediate peptide effects on fitness.

(E) Per position fitness scores for NFE2L2, MDM2, and PIK3CA. Select known PPIs are annotated on the plots, corresponding to regions of significant depletion.

(F) Per position fitness scores for the tumor suppressor CDKN2A and the negative control GFP. No regions of depletion are identified over the length of either protein.

(G) Fitness scores for mutant residues derived from KRAS. Functional regions sourced from UniProt are overlaid above WT fitness. Dots indicate mutant amino acid fitness scores. Red dots indicate mutant amino acid fitness scores that were significantly depleted during the pooled screen (BH-adjusted $p < 0.05$) (see also Figures S3 and S4).

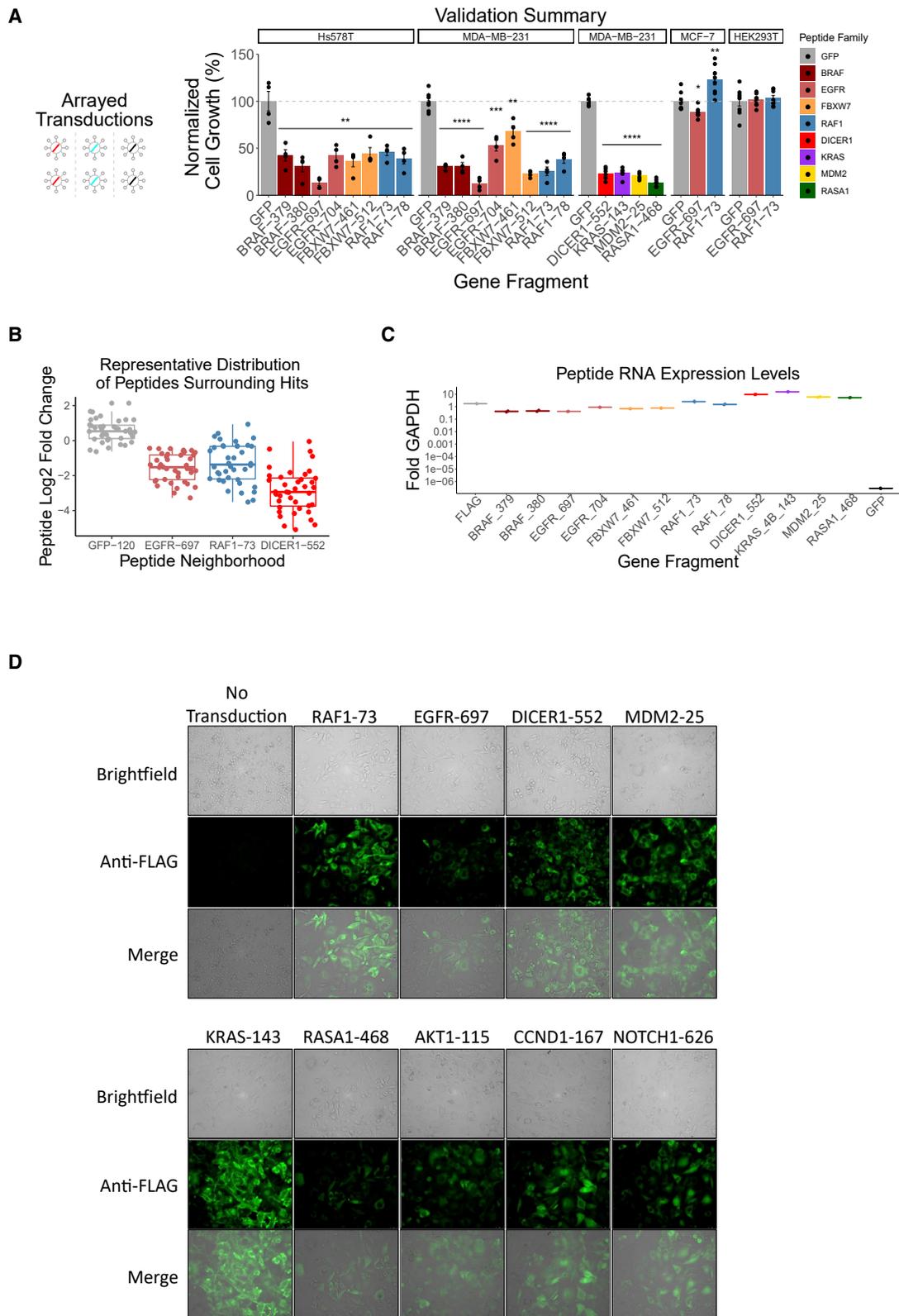


Figure 3. Validation of anti-proliferative peptide activity and expression

(A) *In-vitro*-arrayed validation of lentivirus delivered gene fragments derived from WT proteins. Peptides predicted to be deleterious to cell growth (by depletion in pooled screen) significantly inhibited proliferation relative to GFP control. Cell proliferation was measured via the WST-8 assay after one week of growth following

(legend continued on next page)

cells are Ras WT and not sensitive to RAF1 knockout; correspondingly, they show no fitness defect upon overexpression of the RAF1-73 peptide (Eckert et al., 2004; Broad Institute, 2019). Additionally, MCF-7 cells show a reduced fitness defect upon overexpression of EGFR-697, consistent with their status as an EGFR-negative cell line (Uhlen et al., 2017). As well, the EGFR-negative and Ras WT HEK293T cell line transduced with EGFR-697 and RAF1-73 showed no growth defects, further indicating that this screening methodology identifies context dependent inhibitors of cellular proliferation rather than generally toxic peptide motifs.

For individual peptides that were significantly depleted, we saw consistent depletion of nearby peptides, supporting our strategy of using an amino-acid-level score to rank domains (Figure 3B). To understand the level of peptide expression achieved via our lentiviral constructs, we then performed qPCR on all peptides validated via the WST-8 assay (Figure 3C). We additionally generated 3xFLAG-tagged versions of several significantly depleted peptides to verify peptide constructs had robust protein translation when overexpressed via lentivirus (Figure 3D). The peptides tested showed strong expression at the RNA and protein levels 72 h after transduction, indicating that the EF1 α promoter can drive robust expression of small peptides. Assuming the peptides are translated from their mRNA at a similar rate as GAPDH (GAPDH has a cellular concentration of approximately 0.4 μ M), it can be estimated from the qPCR data that peptide molar concentrations in MDA-MB-231 cells range from 0.15–6.5 μ M depending on the construct (Lazarev et al., 2020).

We further tested three putatively enriched peptides derived from AKT1 (AKT1-115), NOTCH1 (NOTCH1-626), and CCND1 (CCND1-167) in MDA-MB-231 cells to verify that they conferred a growth advantage. All three peptides grew more rapidly than a control group transduced with GFP-coding lentivirus, confirming that if desired this methodology can be used to identify peptides with a pro-proliferative phenotype (Figure S5C). While the average length of a protein domain is predicted to be 100 aa (Lin and Zewail, 2012), we hypothesized based on the modular conformation of long proteins and prior work focused on dominant negatives that 40-mer peptides would be sufficient to fold into ordered structures. To experimentally examine the effect of peptide length on antiproliferative phenotype we transduced MDA-MB-231 cells 4 different-sized peptides centered on our identified hits RAF1-73 and EGFR-697. Although most of the peptides tested still had a growth disadvantage compared with the GFP control, the parent peptides consistently caused slower cell growth than the shorter versions did (Figure S5D).

After validating the bioactivity and expression of peptides identified in the screens, we then sought to extract higher order functional information from the dataset. First, we examined how peptide depletion corresponded to the 3D structure of RB1. The tumor suppressor RB1 contained domains that were highly deleterious to cell fitness. The N-terminal RbN domains were both highly depleted, potentially due to previously described allosteric interactions with the cell-cycle regulatory transcription factor E2F (Burke et al., 2012). Consistent with this hypothesis, it has been previously shown that the addition of N-terminal domains of RB1 is sufficient to halt DNA replication in xenopus egg extracts (Borysov et al., 2015). By overlaying the amino-acid-level fitness scores on the crystal structure for RB1, we found that the periodicity of the depletion profile correlates with the transition between the various alpha helices of the protein (Figure 4A; Table S7). This result highlights how higher-order protein-level features can inform observed peptide fitness and give new insights into the modular nature of the RB1 structure.

We next visualized how peptides from this screen impact cancer-driver-specific signaling networks (Figure 4B) using publicly available protein-protein interaction data from Interactome INSIDER (Meyer et al., 2018). Interactome INSIDER predicts protein-protein interaction interfaces via a random forest classifier built on experimental cocrystal structures, homology models, and co-evolution data. While the PepTile screening methodology is agnostic to mechanism of action (overexpressed peptides can interact with proteins, nucleic acids, lipids, small molecules, etc. within the cell), we chose to focus initially on protein-protein interactions owing to the availability of extensive databases of predicted and experimentally validated interactions. In the network presented in Figure 4B, edges indicate whether a protein interaction interface overlaps a region of peptides deleterious to cell fitness, and nodes are colored by gene fitness data sourced from DepMap CRISPR knockout screening (Meyer et al., 2018; Broad Institute, 2019). There was not a significant association between the DepMap CERES fitness score (an estimate of knockout fitness adjusted for copy-number variations) for a gene and the minimum peptide-derived domain fitness for that gene (Pearson $p = 0.79$). This result stems from the fact that (1) not every gene that is essential has modular domains from which a strongly bioactive peptide can be derived and (2) many genes with no fitness impact in CRISPR screens (such as the tumor suppressor RB1, FBXW7, or TP53) have interfaces from which deleterious peptides can be mined. Together, these analyses highlight the ability of peptides derived from protein-protein interaction interfaces to perturb cellular proliferation. 53.7% of Interactome INSIDER predicted physical interactions between

lentiviral transduction. Bar plots indicate mean, with error bars representing standard error (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Each panel represents a separately conducted experiment (hence the two MDA-MB-231 panels).

(B) Representative distributions of peptide level \log_2 fold change for all peptides overlapping several hits identified from the screen. In addition, we have included an arbitrarily selected region of the GFP protein to highlight a domain with no growth disadvantage. There is consistent depletion of the peptides surrounding hits, providing further justification for our strategy of averaging nearby peptides into an amino-acid-level score.

(C) qPCR validation of lentivirally delivered peptide expression levels relative to GAPDH internal control. MDA-MB-231 cells were transduced at an MOI of 4 in duplicate, with RNA extracted after 72 h. Expression levels of all peptide hits shown in the main text have been quantified at the RNA level, along with a non-targeting 3xFlag tag control peptide for reference. Also included is a negative control GFP transduction, lacking appropriate primer-binding sites for amplification.

(D) Validation of peptide expression via immunofluorescence. MDA-MB-231 cells were transduced (MOI of 4) with lentivirus coding for 3x FLAG-tagged peptides 72 h before immunostaining and imaging (see STAR methods). Expression levels of six antiproliferative peptides shown in the main text have been quantified at the protein level, along with untransduced MDA-MB-231 cells as a control. Additionally, the protein expression level of the three validated enriched peptides was tested. All peptides show robust expression, validating the protein-level expression of these small peptide constructs (see also Figure S5).

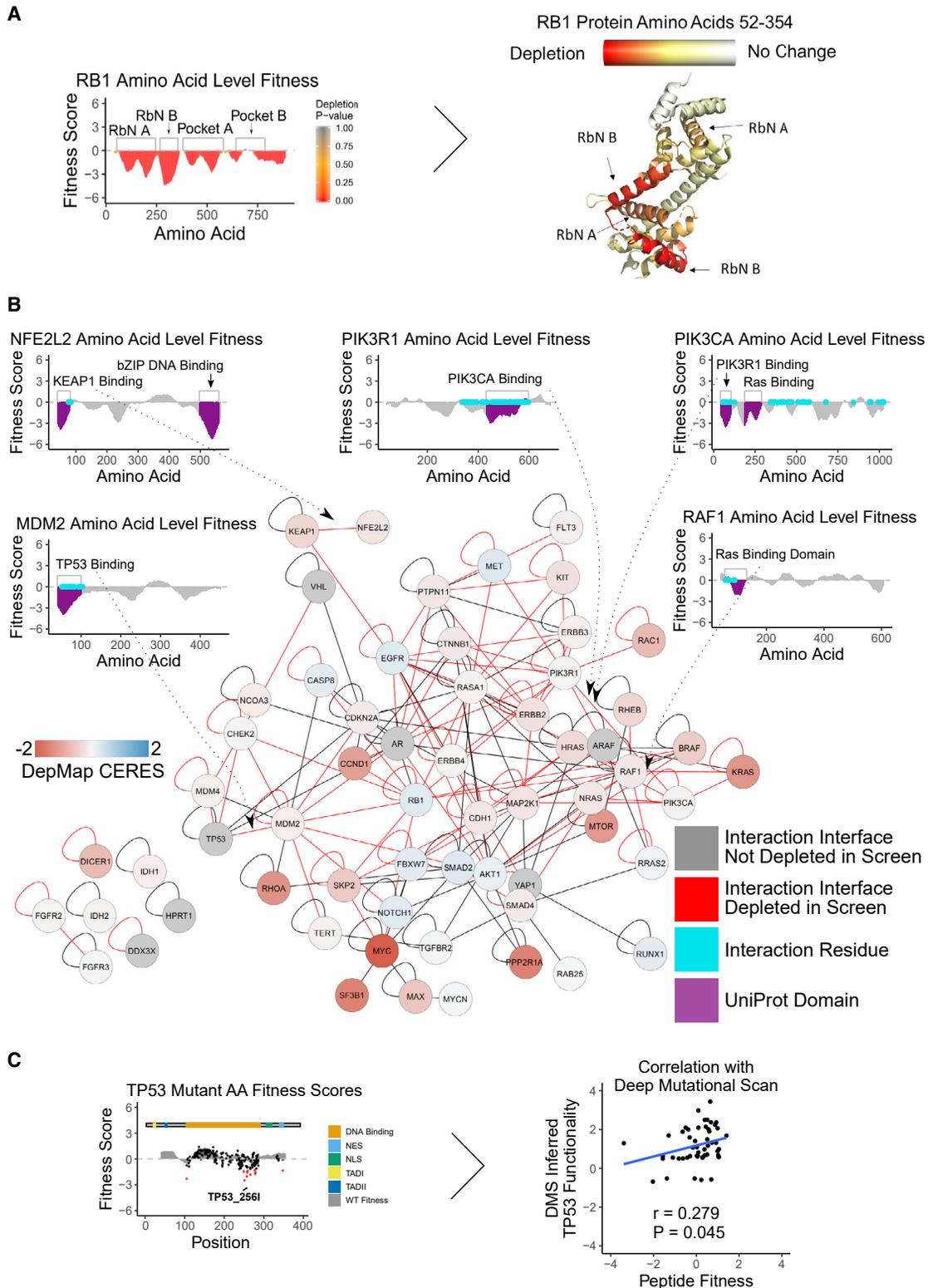


Figure 4. Anti-proliferative peptides derived from oncogenic interaction interfaces

(A) RB1 per position fitness scores mapped onto the RB1 N-terminal crystal structure. Regions of relatively high and low depletion appear to correspond to transitions between specific alpha helix's in the RB1 structure, illustrating how structural elements in the parent protein control peptide phenotype. (B) Network of potential interactions among cancer drivers in this gene set. Interaction data are sourced from Interactome INSIDER, with fitness data from DepMap CRISPR screening overlaid. Nodes colored in red are essential for cell fitness, while nodes colored in blue are non-essential or have increased growth rates upon

(legend continued on next page)

cancer driver genes assayed overlap regions with bioactive peptides, supporting the broad importance of modular interacting motifs in controlling cell fitness.

To further validate that our peptide overexpression platform can identify biophysical features relevant to the protein from which they were derived, we compared the mutant TP53 peptide data with existing TP53 deep mutational scan (DMS) data (Figure 4C) (Kotler et al., 2018). In this DMS dataset, TP53-null cells were transduced with a library of lentiviral particles coding for full-length mutant TP53 variants and subjected to competitive growth. After first filtering the DMS data for only TP53 mutants with a high magnitude of effect on cell fitness (absolute fitness value >0.5) we compared the fitness of the corresponding mutant peptides from our own screen. We surmised that given the highly dissimilar nature of the screening technologies, limiting the comparison to only high effect size mutants would allow for a clearer interpretation. Inferred TP53 functionality was defined as the inverse of the TP53 variant “relative fitness score,” insofar as synonymous, fully functional, TP53 mutants have highly negative fitness scores due to their activity as tumor suppressors. Even with the highly dissimilar screening modalities, we observed significant correlation (Pearson $r = 0.279$; $p = 0.045$) between the predicted mutant TP53 functionality from the DMS data to the mutant TP53 peptide fitness. This comparison to DMS data indicates that TP53 mutants expected to be functional (i.e., have structures consistent with appropriate ligand binding and cellular bioactivity) generate mutant peptides with greater bioactivity in the cell. Together, these results highlight a major utility of this approach i.e., the ability to interrogate user-defined peptide sequences as opposed to those present only in WT protein structures. Future assays could combine this peptide screening protocol with structural modeling to design and test rationally mutagenized peptide libraries with novel biophysical properties or improved target binding.

Engineering peptides for exogenous delivery

After validating the activity of these peptide constructs when overexpressed genetically, we investigated if peptides from our screen could function when repurposed as exogenously delivered drug-like molecules (Figure 5A). To test this, we chemically synthesized EGFR-697 as well as RAF1-73 and measured their ability to inhibit cell growth when conjugated to the TAT cell-penetrating protein transduction domain (Schwarze et al., 1999). EGFR-697 maintained its anti-proliferative effects when delivered exogenously, showing a dose-dependent impact on cell viability (Figure 5B; Table S8). The IC50s of this peptide was 33.3 μM for Hs578T and 63 μM for MDA-MB-231. Moreover, RAF1-73 was also highly deleterious to cell growth, with IC50 values of 27.0 and 32.6 μM for Hs578T and MDA-MB-231, respectively. These IC50 values are comparable with the mean IC50 of all drugs tested on these cell lines in the Sanger Geno-

mics of Drug Sensitivity Database (48.6 μM for Hs578T and 54.0 μM for MDA-MB-231 cells), contextualizing the relative activity of these peptides and the potential for this methodology (Yang et al., 2013). We also identified two additional peptides (RASA1-468 and MDM2-25) from the larger screen in MDA-MB-231 cells, which show cytotoxic activity when delivered exogenously. RASA1-468 is derived from the Pleckstrin homology domain of RASA1 (mediating various PPIs and interactions with phospholipids; Scheffzek and Welti, 2012), while MDM2-25 is derived from the p53-binding domain of MDM2 (Liu et al., 2010). These peptides had IC50s of 23 and 33 μM , respectively, in MDA-MB-231 cells (Figure 5B). This result demonstrates that how the high-throughput nature of the PepTile screening strategy can identify diverse bioactive peptides that maintain activity when conjugated to a cell-penetrating motif. We similarly anticipate there are many more unexplored hit peptides from the screen which could show anti-cancer activity when delivered exogenously. Collectively, these data further confirm that the peptides identified in this screen are acting at the protein level and suggest that further engineering of these compounds could yield translationally relevant biopharmaceuticals.

As peptide constructs will likely require additional engineering to maximize efficacy toward intracellular targets *in vivo*, we have also demonstrated a streamlined recombinant production protocol as a complement to the PepTile approach and general resource to accelerate the engineering of peptide therapeutics. This method was validated by the production of milligram-scale quantities of TAT conjugated 3xFLAG peptide, outperforming the costs associated with commercial peptide synthesis (Figure S6; STAR methods). Because this peptide production method (as well as the PepTile fitness screening strategy—see Table S9) requires inexpensive equipment and few specialized reagents, it is easily adaptable to labs of any scale, as well as automated medium throughput screening approaches.

Characterization of peptide function

We then sought to validate our hypothesis that the functionality of these putative inhibitory peptides was dependent on the role and structure of the WT protein domain they were derived from. Specifically, we explored whether the RAF1-73 peptide (derived from the RAF1-RBD) retained the ability of the full-length domain to bind activated Ras proteins. To evaluate this potential interaction, we co-transfected the constitutively active KRAS G12V mutant and 3xFLAG-RAF1-73 in HEK293T cells, then performed a co-immunoprecipitation using anti-FLAG agarose beads (Figure 6A). We chose to transfect with a constitutively active KRAS variant because the Ras-Raf interaction occurs only on activated Ras proteins. Western blot analysis of the immunoprecipitated protein complexes subsequently verified the protein-protein interaction between RAF1-73 and Ras.

knockout. Dark gray nodes indicate genes for which high-confidence CRISPR-based fitness data were not available. Edges indicate a predicted interaction interface between the cancer drivers. Red edges indicate interactions which overlap regions of significant peptide depletion (fitness score < -1.5 for interface amino acids). Arrows highlight example depleted peptide regions corresponding to specific oncogenic PPIs.

(C) Comparison of mutant fitness scores derived from peptide screening data, with fitness scores derived from DMS data in a TP53-null cell line (Kotler et al., 2018). After filtering out TP53 mutants with little effect on cell fitness in the DMS (absolute value of fitness scores < 0.5), inferred that TP53 functionality is significantly correlated with mutant-peptide-derived fitness (Pearson, $p = 0.045$), supporting the hypothesis that peptide screening can be used to identify functionally important residues in the context of cancer cell fitness (see also Figures S3 and S4).

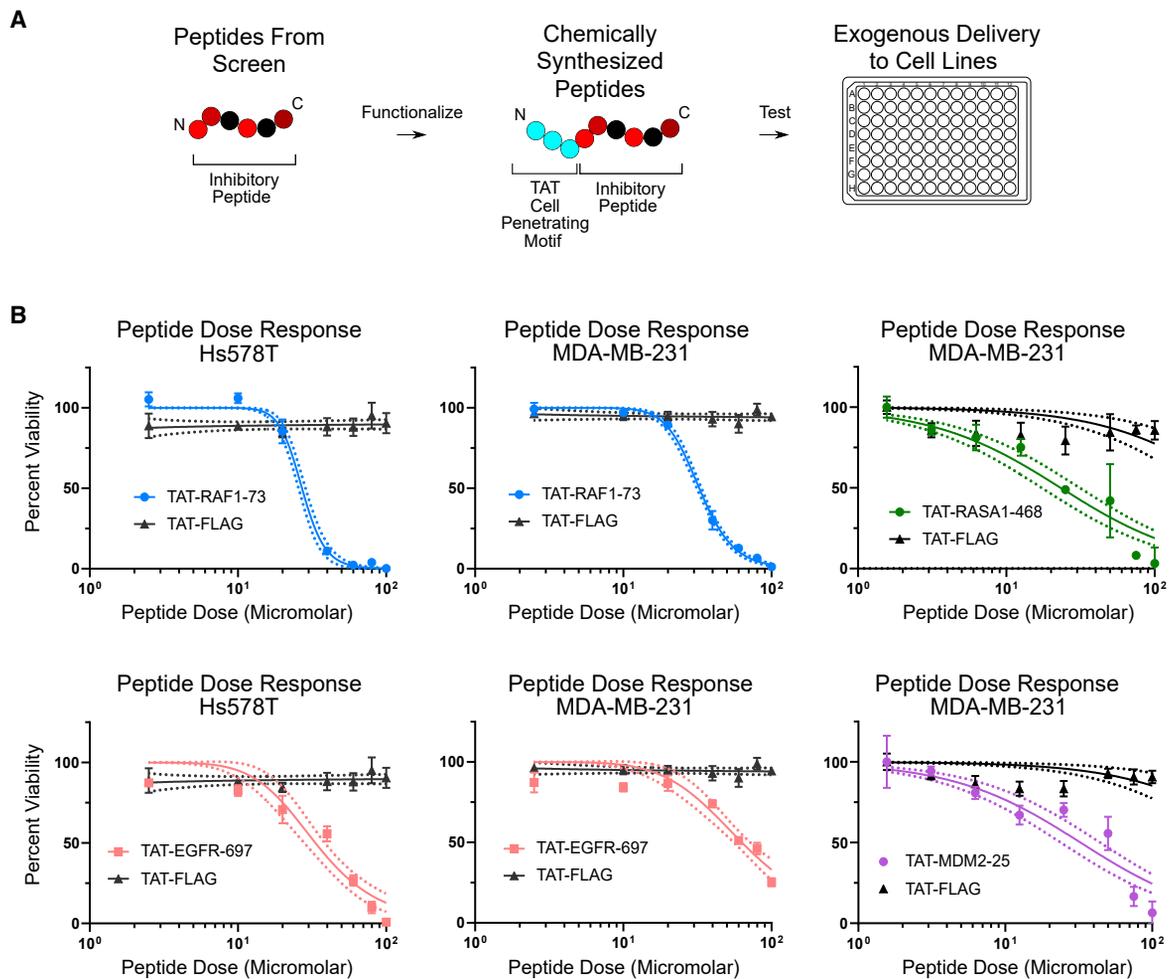


Figure 5. Cancer-driver-derived peptides have protein-level activity and potential drug-like function

(A) Overview of peptide functionalization for intracellular delivery. Hit peptides from the screen were conjugated to a TAT cell-penetrating motif and produced via solid phase peptide synthesis.

(B) *In vitro* testing with chemically synthesized peptides ($n = 3-4$). Chemically synthesized hit peptides conjugated to a cell-penetrating TAT protein transduction motif were added to cells at 0–100 μM . A 3 \times FLAG peptide conjugated to TAT served as the negative control. Cell viability was measured 24 h later by the WST-8 assay, indicating that TAT functionalized hit peptides can effectively inhibit the growth of Hs578T and MDA-MB-231 cells in a dose-dependent manner. Dotted lines indicate 95% confidence intervals for nonlinear fit. TAT-RAF1-73 and TAT-EGFR-697 were tested on the same plate, hence identical negative control measurements (see also Figure S6).

Next, we performed a similar experiment investigating the potential interaction between full-length EGFR and EGFR-697, confirming detectable co-immunoprecipitation of the EGFR-697 peptide with the full-length EGFR protein (Figure 6A). In order to better understand how the EGFR-697 peptide was perturbing the cells, we conducted RNA sequencing on Hs578T cells modified via lentivirus to overexpress EGFR-697 (Table S10). We identified 225 differentially expressed genes (BH-adjusted p value < 0.05) and performed gene set enrichment analysis (GSEA) to identify upregulation and downregulation of genetic pathways (Sergushichev, 2016). We tested 239 KEGG pathways corresponding to cell signaling and metabolism, with 22 pathways showing highly significant (false discovery rate < 0.025) upregulation/downregulation in cells expressing EGFR-697 compared with control cells transduced with GFP (Figure 6B). Several metabolic pathways relating to oxidative phos-

phorylation and carbon metabolism were downregulated, consistent with the role of oncogenic EGFR signaling as a driver of metabolic alterations (Borlak et al., 2015; Li et al., 2015; Lanning et al., 2017). Furthermore, genes relating to DNA replication were also downregulated, consistent with the observed slow growing phenotype. In addition to performing GSEA on KEGG pathways, we also tested a set of curated genes from the Molecular Signatures Database comprised genes significantly downregulated/upregulated in H1975 cells upon treatment with an irreversible EGFR inhibitor (Kobayashi et al., 2006). We chose to test against these gene sets derived from EGFR inhibition experiments because they describe the putative transcriptomic effects of perturbing EGFR at the protein level. EGFR-697 transduction in Hs578T cells resulted in downregulation of genes identified as downregulated in response to chemical EGFR inhibition and upregulation of genes identified as upregulated (FDR = 0.008 and

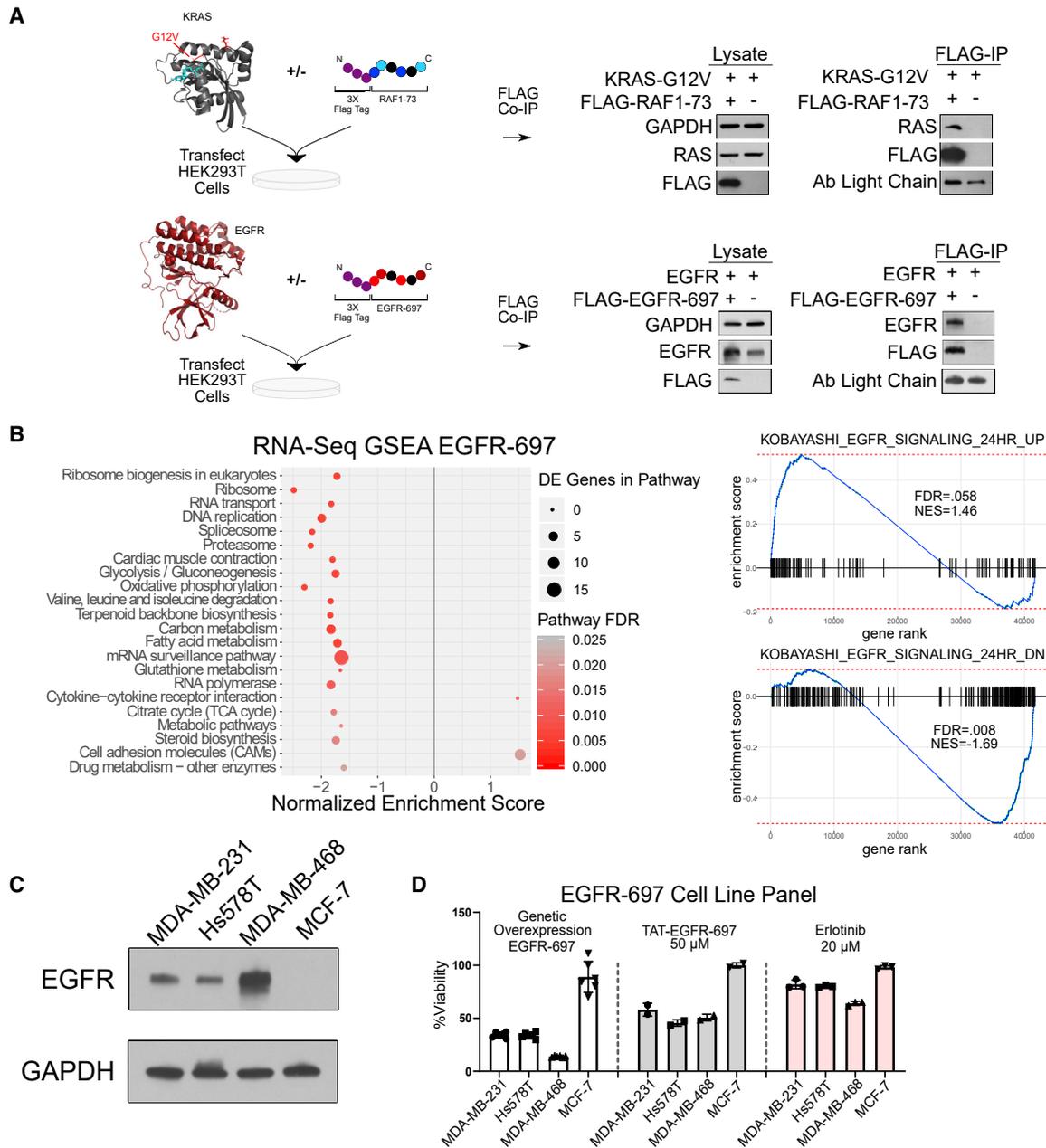


Figure 6. Cancer-driver-derived peptides show context-dependent activity

(A) Peptide mechanism explored via co-immunoprecipitation. 3X-Flag-tagged RAF1-73 derived from the RBD of RAF1 pulls down activated Ras when immunoprecipitated, indicating retention of WT domain biological functionality. Analogously, the 3X-FLAG-tagged EGFR-697 peptide pulls down the co-transfected full-length EGFR protein confirming a protein-level interaction between the two proteins.

(B) Results of RNA sequencing on EGFR-697 expressing Hs578T cells. EGFR-697 overexpression results in significant growth arrest, and differential expression of 225 genes, as well as significant downregulation of pathways relevant to cellular proliferation. Additional GSEA analysis revealed a transcriptional phenotype consistent with perturbed signaling along the EGFR pathway. Gene set “KOBAYASHI_EGFR_SIGNALING_24HRS_DN” is a gene set composed of genes downregulated upon treatment with an irreversible EGFR inhibitor in H1975 cells (Kobayashi et al., 2006). Treatment with EGFR-697 peptide results in significant downregulation of this gene set in Hs578T cells. The “KOBAYASHI_EGFR_SIGNALING_24HRS_UP” is a gene set from the same experiment highlighting genes that are upregulated upon EGFR inhibition. This gene set is significantly upregulated upon EGFR-697 overexpression. The vertical lines on the plot each represent a gene in the gene set, with their location representing their position in the ranked list of genes from the RNA sequencing data (ranked by DESeq2’s shrunken log fold change; Love et al., 2014). NES is the normalized enrichment score, quantifying the extent genes within the given gene set are up or downregulated in the RNA sequencing data. FDR is the false discovery rate for that enrichment score.

(C) EGFR expression levels of breast cancer cell lines quantified via western blot. MCF-7 cells show no detectable expression of EGFR.

(D) Breast cancer cell line panel treated with genetically overexpressed EGFR-697, synthesized TAT-EGFR-697 and erlotinib. Cell viabilities were determined via crystal violet staining of live cells after 7 days for the genetically overexpressed constructs, or 24 h for the exogenously delivered molecules. For the genetically

(legend continued on next page)

0.058, respectively). To provide further confidence that EGFR-697 is acting in an EGFR-dependent manner, we tested the effects of genetically overexpressed EGFR-697 and TAT-EGFR-697 in a panel of breast cancer cell lines with varying levels of EGFR expression (Figures 6C and 6D). Both the genetically overexpressed and the exogenously delivered versions of EGFR-697 showed greater activity in cell lines with detectable EGFR expression. These data were benchmarked against a high dose of erlotinib, showing a similar EGFR-expression-dependent change in sensitivity. Collectively, these data support the hypothesis the EGFR-697 peptide perturbs breast cancer cells in an EGFR-dependent manner. However, the exact mechanism of this interaction and the extent of off-target interactions will need further study.

As a final analysis of peptide function, we have explored computationally the predicted structure of peptides derived from RAF1 and EGFR (Figure S7). We first examined whether the individual hit peptides RAF1-73 and EGFR-697 had modeled structures resembling that of the WT domain they were derived from (Figure S7A). Peptide structures were generated using TrRosetta, a highly accurate protein structural prediction software (Yang et al., 2020). Both RAF1-73 and EGFR-697 were predicted to fold into structures highly similar to that of the WT protein (TM scores of 0.63 and 0.67, respectively). A TM score greater than 0.5 corresponds to a p value less than 5.5×10^{-7} and is a widely used criterion when two protein structures have the same fold (Xu and Zhang, 2010). Subsequently, we comprehensively modeled 957 peptides derived from RAF1 and EGFR, which had available overlapping crystal structures on PDB. We found that the vast majority (>75%) of the peptide models derived from RAF1 and EGFR had predicted structures highly similar to that of the full-length protein (Figure S7B). All models predicted from trRosetta had confidence scores (predicted local distance difference Test outputted by DeepAccNet) greater than 0.58, indicating high stereochemical plausibility of the predicted models (Figures S7C and S7D). However, a small subset of derived peptides modeled had structures diverging from that of the full-length protein (minimum TM score observed = 0.305). To evaluate the variation in TM scores among the fragments, we analyzed the TM scores of each fragment as a function of its secondary structure. We found that secondary structure in the full-length protein is not a strong driver of predicted peptide conformational similarity to WT folding (Figure S7E). Peptides derived from regions with alpha helices, beta sheets, or both were largely predicted to fold into structures resembling the full-length protein (mean TM scores of 0.74, 0.82, and 0.76, respectively). When examining the predicted structures least similar to the full-length protein (TM scores < 0.5), we found that secondary structure of the peptides was consistent with the full-length structure in 79% of low similarity RAF1 peptides and 71% of EGFR peptides. This suggests that the low TM scores were attributed to differences in the angle of certain

amino acids rather than the misfolding of secondary structures (Figure S7F). Given the diversity of peptides tested, some peptides which deplete in this screen may fold into structures dissimilar to that of the full-length protein from which they are derived (just as some sgRNA or siRNA have unexpected off-targets), underscoring the need for robust downstream validations of screen results.

DISCUSSION

Overall, we have demonstrated a comprehensive screening platform that enables the identification of peptide inhibitors of cancer cell growth. This methodology is scalable due to the ease of oligonucleotide synthesis, simple to perform, and highly precise, allowing users to interrogate protein sequences with single-amino-acid resolution. Because the library of peptide-coding gene fragments is user defined and custom synthesized, this strategy is easily adaptable to diverse studies where a selection strategy can be devised to enrich or deplete cells with the phenotype of interest.

Studies on signal transduction in the mammalian cell often consider proteins as a series of nodes within a network for simplicity (Azeloglu and Iyengar, 2015). The results presented here also highlight that signal transduction is highly dependent on tight control of numerous modular functional units within proteins to mediate information flow and maintain cell fitness. Supporting this conclusion, peptide mediated perturbations to the endogenous interaction network of proteins and their diverse ligands (proteins, small molecules, DNA/RNA, etc.) can strongly impact cellular growth rates. Ongoing efforts to comprehensively map protein functional domains are thus critical to understanding disease-relevant cell signaling programs. Furthermore, we find that functional domains within proteins can serve as a promising source of bioactive peptides with which to perturb signaling and protein-protein interactions.

However, PepTile as implemented has several limitations which future technology development can iteratively work to improve. First, tiling libraries are likely unsuited for inhibiting protein interactions mediated by residues close in physical space, but far apart in the full-length ORF. In the future, using structural modeling to inform library design can generate synthetic peptides better suited for inhibiting this type of interaction. Additionally, current DNA synthesis technology limits array synthesized DNA libraries to less than ~350 bp (with increasing error rates as the size of the DNA increases). Moving forward, improvements in DNA synthesis will open new avenues for screening more complex peptide and protein therapeutics efficiently. As well, PepTile is currently agnostic to any post-translational modifications which may be essential for peptide function. Advances in high-throughput protein-level analysis will additionally allow for a more rapid and accurate characterization of peptide mechanism.

overexpressed EGFR-697, after 7 days of growth there was a significant association between EGFR expression levels and cell lines viability relative to a GFP transduced control (Pearson $p < 0.0001$, $r = -0.803$). EGFR expression levels were quantified based on the pixel intensity of the western blot data shown in (C), relative to the GAPDH internal control. At 50 μM , the cell lines with detectable EGFR expression show a reduction in viability after 24 h of exposure to TAT-EGFR-697. In contrast, EGFR-negative MCF7 cells show no reduction in viability. Cell viabilities are normalized to a PBS-vehicle-treated control on the same plate. Cells expressing EGFR at detectable levels have greater sensitivity to erlotinib (24-h treatment) than non-EGFR-expressing MCF7 cells. Cell viabilities for erlotinib-treated cells are normalized to DMSO-treated cells on the same plate. Data indicate mean \pm standard deviation (see also Table S10).

Peptides expressed outside the context of the native protein may in some cases have bioactivity not consistent with the function of the parent protein. Peptides derived from highly hydrophobic or transmembrane domains, domains with high homology to other proteins, those bearing reactive moieties such as cysteines, or peptides with a high net charge could result in non-specific binding/aggregation within the cell. This possibility highlights the importance of downstream validation of peptide hits, and the broader challenge of identifying the mechanism underlying biological phenotypes (Editorial, 2010). Furthermore, peptides mined via the screens will likely have only moderate binding affinities and bioavailability, and to improve activity systematic mutagenesis may be required. To this end, WT peptide screening could be followed up with a smaller secondary screen mutagenizing hit compounds to identify semi-synthetic binders with higher affinity to the target protein, better bioavailability, or other improved functional characteristics.

Inhibitory peptides have immense potential as both research tools and therapeutics. Direct inhibition of protein activity without genetic alteration opens unique screening avenues with which to probe protein function. For example, protein-protein interaction networks could be more precisely perturbed via inhibitory peptides contacting a specific protein surface than by complete genetic knockdown. The ability to identify protein regions associated with cell fitness can also serve to complement traditional drug development efforts, such as determining critical residues for inhibition via small molecules or antibodies. Additionally, this screening resource identifies inhibitory peptides that are immediately translatable, bypassing the need for additional high-throughput screens to identify candidate molecules. Functionally, peptides can be (1) readily made cell permeable via coupling of cell-penetrating motifs to enable drug-like function (Guidotti et al., 2017) or, alternatively, (2) coupled to chemical moieties such as poly-ethylene glycol (PEG) or protein domains with naturally long serum half-life such as Fc, transferrin, or albumin to improve persistence in circulation (Strohl, 2015). In this study, with minimal engineering we developed two drug-like peptides that opposed triple-negative breast cancer cell growth *in vitro* as effectively as some FDA-approved small molecules targeting the same proteins (Yang et al., 2013). Advances in biologics delivery will further improve the translational relevance of this strategy. We anticipate a future role for this method of peptide inhibitor screening in both basic research and drug development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Design of peptide coding gene fragment libraries
 - Cancer driver gene fragment cloning
 - Lentivirus production

- Fitness screening in mammalian cell lines
- HTS library preparation and sequencing
- Processing of sequencing files
- Calculation of amino acid level fitness scores
- Validating highly depleted gene fragments
- Crystal violet viability measurements
- Engineering peptides for exogenous delivery
- Co-immunoprecipitation
- Western blotting
- qPCR
- Immunofluorescence
- RNA-seq of highly depleted fragments
- Network visualization
- Computational modeling of peptide structure
- Recombinant peptide production

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cels.2021.05.002>.

ACKNOWLEDGMENTS

We thank members of the Mali lab for advice with experiments and analyses. We also thank Kristen Jepsen and Benjamin Henson for advice and help with next generation sequencing. This work was generously supported by UCSD Institutional Funds, NIH grants (R01HG009285, RO1CA222826, RO1GM123313, and U54CA209891), an NHLBI training grant to A.P. (T32 HL 105373) and an NSF Graduate Research Fellowship (DGE-1650112) to K.M.F.

AUTHOR CONTRIBUTIONS

K.M.F. and P.M. conceived and planned the study; K.M.F., R.P., A.P., and D.-H.C. performed experiments; K.M.F., N.P., and D.-H.C. analyzed data and generated figures; K.M.F. and P.M. wrote the manuscript, and all authors assisted in editing the manuscript.

DECLARATION OF INTERESTS

The authors have filed a patent based on this work. P.M. is a scientific co-founder of Shape Therapeutics, Boundless Biosciences, Seven Therapeutics, Navega Therapeutics, and Engine Biosciences. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies.

Received: May 29, 2020
Revised: February 9, 2021
Accepted: March 30, 2021
Published: May 28, 2021

REFERENCES

- Adhikari, H., and Counter, C.M. (2018). Interrogating the protein interactomes of RAS isoforms identifies PIP5K1A as a KRAS-specific vulnerability. *Nat. Commun.* 9, 3646.
- Akada, R., Yamamoto, J., and Yamashita, I. (1997). Screening and identification of yeast sequences that cause growth inhibition when overexpressed. *Mol. Gen. Genet.* 254, 267–274.
- Azeloglu, E.U., and Iyengar, R. (2015). Signaling networks: information flow, computation, and decision making. *Cold Spring Harb. Perspect. Biol.* 7, a005934.
- Bai, Z., Hou, S., Zhang, S., Li, Z., and Zhou, P. (2017). Targeting self-binding peptides as a novel strategy to regulate protein activity and function: a case

study on the proto-oncogene tyrosine protein kinase C-Src. *J. Chem. Inf. Model.* **57**, 835–845.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e18.

Barnard, D., Sun, H., Baker, L., and Marshall, M.S. (1998). In vitro inhibition of Ras-Raf association by short peptides. *Biochem. Biophys. Res. Commun.* **247**, 176–180.

Beaulieu, M.E., Jauset, T., Massó-Vallés, D., Martínez-Martín, S., Rahl, P., Maltais, L., Zacarias-Fluck, M.F., Casacuberta-Serra, S., Serrano Del Pozo, E., Fiore, C., et al. (2019). Intrinsic cell-penetrating activity propels oncomyc from proof of concept to viable anti-myc therapy. *Sci. Transl. Med.* **11**.

Becker, C.F.W., Hunter, C.L., Seidel, R., Kent, S.B., Goody, R.S., and Engelhard, M. (2003). Total chemical synthesis of a functional interacting protein pair: the protooncogene H-Ras and the Ras-binding domain of its effector c-Raf1. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5075–5080.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* **57**, 289–300.

Borlak, J., Singh, P., and Gazzana, G. (2015). Proteome mapping of epidermal growth factor induced hepatocellular carcinomas identifies novel cell metabolism targets and mitogen activated protein kinase signalling events. *BMC Genomics* **16**, 124.

Borysov, S.I., Nepon-Sixt, B.S., and Alexandrow, M.G. (2015). The N terminus of the retinoblastoma protein inhibits DNA replication via a bipartite mechanism disrupted in partially penetrant retinoblastomas. *Mol. Cell. Biol.* **36**, 832–845.

Bose, R., and Zhang, X. (2009). The ErbB kinase domain: structural perspectives into kinase activation and inhibition. *Exp. Cell Res.* **315**, 649–658.

Boyer, J., Badis, G., Fairhead, C., Talla, E., Hantraye, F., Fabre, E., Fischer, G., Hennequin, C., Koszul, R., Lafontaine, I., et al. (2004). Large-scale exploration of growth inhibition caused by overexpression of genomic fragments in *Saccharomyces cerevisiae*. *Genome Biol* **5**, R72.

Broad Institute (2019). DepMap Achilles 19Q1 public. <https://depmap.org/portal/download/>.

Buhrman, G., Wink, G., and Mattos, C. (2007). Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. *Structure* **15**, 1618–1629.

Burke, J.R., Hura, G.L., and Rubin, S.M. (2012). Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control. *Genes Dev* **26**, 1156–1166.

Chang, Y.S., Graves, B., Guerlavais, V., Tovar, C., Packman, K., To, K.H., Olson, K.A., Kesavan, K., Gangurde, P., Mukherjee, A., et al. (2013). Stapled α -helical peptide drug development: a potent dual inhibitor of MDM2 and MDX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. USA* **110**, E3445–E3454.

Chavez, K.J., Garimella, S.V., and Lipkowitz, S. (2010). Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. *Breast Dis* **32**, 35–48.

Chen, J., Brunner, A.D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., and Weissman, J.S. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science* **367**, 1140–1146.

Clark, G.J., Drugan, J.K., Terrell, R.S., Bradham, C., Der, C.J., Bell, R.M., and Campbell, S. (1996). Peptides containing a consensus Ras binding sequence from Raf-1 and the GTPase activating protein NF1 inhibit Ras function. *Proc. Natl. Acad. Sci. USA* **93**, 1577–1581.

Cox, A.D., Fesik, S.W., Kimmelman, A.C., Luo, J., and Der, C.J. (2014). Drugging the undruggable RAS: mission Possible? *Nat. Rev. Drug Discov.* **13**, 828–851.

Craik, D.J., Fairlie, D.P., Liras, S., and Price, D. (2013). The future of peptide-based drugs. *Chem. Biol. Drug Des.* **81**, 136–147.

Cuadrado, A., Rojo, A.I., Wells, G., Hayes, J.D., Cousin, S.P., Rumsey, W.L., Attucks, O.C., Franklin, S., Levonen, A.-L., Kensler, T.W., and Dinkova-Kostova, A.T. (2019). Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases. *Nat. Rev. Drug Discov.* **18**, 295–317.

Dang, C.V. (2012). MYC on the path to cancer. *Cell* **149**, 22–35.

Der, C.J., Finkel, T., and Cooper, G.M. (1986). Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell* **44**, 167–176.

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.

Doench, J.G. (2018). Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* **19**, 67–80.

Donsky, E., and Wolfson, H.J. (2011). PepCrawler: A fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics* **27**, 2836–2842.

Dorrity, M.W., Queitsch, C., and Fields, S. (2019). High-throughput identification of dominant negative polypeptides in yeast. *Nat. Methods* **16**, 413–416.

Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3**, 11–22.

Eckert, L.B., Repasky, G.A., Ulkū, A.S., McFall, A., Zhou, H., Sartor, C.I., and Der, C.J. (2004). Involvement of ras activation in human breast cancer cell signaling, invasion, and anoikis. *Cancer Res* **64**, 4585–4592.

Eisenhardt, A.E., Sprenger, A., Röring, M., Herr, R., Weinberg, F., Köhler, M., Braun, S., Orth, J., Diedrich, B., Lanner, U., et al. (2016). Phospho-proteomic analyses of B-Raf protein complexes reveal new regulatory principles. *Oncotarget* **7**, 26628–26652.

Ford, K., McDonald, D., and Mali, P. (2019). Functional genomics via CRISPR-Cas. *J. Mol. Biol.* **431**, 48–65.

Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579.

Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192–205.

Guidotti, G., Brambilla, L., and Rossi, D. (2017). Cell-penetrating peptides: From basic research to clinics. *Trends Pharmacol. Sci.* **38**, 406–424.

Gurtan, A.M., Lu, V., Bhutkar, A., and Sharp, P.A. (2012). In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. *RNA* **18**, 1116–1122.

Han, Y., and Král, P. (2020). Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. *ACS Nano* **14**, 5143–5147.

Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P. (2007). Structure of a Fbw7-Skp1-Cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* **26**, 131–143.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219–222.

Hiranuma, N., Park, H., Baek, M., Anishchenko, I., Dauparas, J., and Baker, D. (2021). Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat. Commun.* **12**, 1340.

Hui, L., Zheng, Y., Yan, Y., Bargonetti, J., and Foster, D.A. (2006). Mutant p53 in MDA-MB-231 breast cancer cells is stabilized by elevated phospholipase D activity and contributes to survival signals generated by phospholipase D. *Oncogene* **25**, 7305–7310.

Iwakuma, T., and Lozano, G. (2003). MDM2, an introduction. *Mol. Cancer Res.* **1**, 993–1000.

Janku, F., Yap, T.A., and Meric-Bernstam, F. (2018). Targeting the PI3K pathway in cancer: are we making headway? *Nat. Rev. Clin. Oncol.* **15**, 273–291.

Kang, J., Sergio, C.M., Sutherland, R.L., and Musgrove, E.A. (2014). Targeting cyclin-dependent kinase 1 (CDK1) but not CDK4/6 or CDK2 is selectively lethal to MYC-dependent human breast cancer cells. *BMC Cancer* **14**, 32.

- Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32, W526–W531.
- Kobayashi, S., Shimamura, T., Monti, S., Steidl, U., Hetherington, C.J., Lowell, A.M., Golub, T., Meyerson, M., Tenen, D.G., Shapiro, G.I., et al. (2006). Transcriptional profiling identifies cyclin D1 as a critical downstream effector of mutant epidermal growth factor receptor signaling. *Cancer Res* 66, 11389–11398.
- Kosuri, S., and Church, G.M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* 11, 499–507.
- Kotler, E., Shani, O., Goldfeld, G., Lotan-Pompan, M., Tarcic, O., Gershoni, A., Hopf, T.A., Marks, D.S., Oren, M., and Segal, E. (2018). A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. *Mol. Cell* 71, 178–190.e8.
- Kubbutat, M.H.G., Ludwig, R.L., Levine, A.J., and Vousden, K.H. (1999). Analysis of the degradation function of Mdm2. *Cell Growth Differ* 10, 87–92.
- Kumar, A., Petri, E.T., Halmos, B., and Boggon, T.J. (2008). Structure and clinical relevance of the epidermal growth factor receptor in human cancer. *J. Clin. Oncol.* 26, 1742–1751.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lanning, N.J., Castle, J.P., Singh, S.J., Leon, A.N., Tovar, E.A., Sanghera, A., MacKeigan, J.P., Filipp, F.V., and Graveel, C.R. (2017). Metabolic profiling of triple-negative breast cancer cells reveals metabolic vulnerabilities. *Cancer Metab* 5, 6.
- Lazarev, V.F., Guzhova, I.V., and Margulis, B.A. (2020). Glyceraldehyde-3-phosphate dehydrogenase is a multifaceted therapeutic target. *Pharmaceutics* 12, 416.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* 15, 554.
- Li, X., Lu, Y., Lu, H., Luo, J., Hong, Y., and Fan, Z. (2015). AMPK-mediated energy homeostasis and associated metabolic effects on cancer cell response and resistance to cetuximab. *Oncotarget* 6, 11507–11518.
- Lin, M.M., and Zewail, A.H. (2012). Hydrophobic forces and the length limit of foldable protein domains. *Proc. Natl. Acad. Sci. USA* 109, 9851–9856.
- Liu, M., Li, C., Pazgier, M., Li, C., Mao, Y., Lv, Y., Gu, B., Wei, G., Yuan, W., Zhan, C., et al. (2010). D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms. *Proc. Natl. Acad. Sci. USA* 107, 14321–14326.
- London, N., Raveh, B., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? *Proteins* 78, 3140–3149.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963.
- Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 29, 2722–2728.
- Martz, C.A., Ottina, K.A., Singleton, K.R., Jasper, J.S., Wardell, S.E., Peraz-Penton, A., Anderson, G.R., Winter, P.S., Wang, T., Alley, H.M., et al. (2014). Systematic identification of signaling pathways with potential to confer anti-cancer drug resistance. *Sci. Signal* 7, ra121. <https://doi.org/10.1126/scisignal.aaa1877>.
- Meyer, M.J., Beltrán, J.F., Liang, S., Fragosa, R., Rumack, A., Liang, J., Wei, X., and Yu, H. (2018). Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* 15, 107–114.
- Nassar, N., Singh, K., and Garcia-Diaz, M. (2010). Structure of the dominant negative S17N mutant of Ras. *Biochemistry* 49, 1970–1974.
- Editorial. (2010). Mechanism matters. *Nat. Med.* 16, 347.
- Nim, S., Jeon, J., Corbi-Verge, C., Seo, M.H., Ivarsson, Y., Moffat, J., Tarasova, N., and Kim, P.M. (2016). Pooled screening for antiproliferative inhibitors of protein-protein interactions. *Nat. Chem. Biol.* 12, 275–281.
- Osorio, D., Rondón-Villarreal, P., and Torres, R. (2015). Peptides: a package for data mining of antimicrobial peptides. *R J* 7.
- Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2017). Biostrings: efficient manipulation of biological strings, R package version 2.46.0. <https://bioconductor.org/packages/Biostrings>.
- Palmieri, L., and Rastelli, G. (2013). α C helix displacement as a general approach for allosteric modulation of protein kinases. *Drug Discov. Today* 18, 407–414.
- Ramer, S.W., Elledge, S.J., and Davis, R.W. (1992). Dominant genetics using a yeast genomic library under the control of a strong inducible promoter. *Proc. Natl. Acad. Sci. USA* 89, 11589–11593.
- Ruan, Z., and Kannan, N. (2018). Altered conformational landscape and dimerization dependency underpins the activation of EGFR by α C- β 4 loop insertion mutations. *Proc. Natl. Acad. Sci. USA* 115, E8162–E8171.
- Rupaimoole, R., and Slack, F.J. (2017). MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.* 16, 203–222.
- Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* 10, 59–64.
- Sato, M., Rodriguez-Barrueco, R., Yu, J., Do, C., Silva, J.M., and Gautier, J. (2015). MYC is a critical target of FBXW7. *Oncotarget* 6, 3292–3305.
- Scheffzek, K., and Welti, S. (2012). Pleckstrin homology (PH) like domains - Versatile modules in protein-protein interaction platforms. *FEBS Lett* 586, 2662–2673.
- Schwarze, S.R., Ho, A., Vocero-Akbani, A., and Dowdy, S.F. (1999). In vivo protein transduction: delivery of a biologically active protein into the mouse. *Science* 285, 1569–1572.
- Sergushichev, A.A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv*.
- Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 16, 299–311.
- Simanshu, D.K., Nissley, D.V., and McCormick, F. (2017). RAS proteins and their regulators in human disease. *Cell* 170, 17–33.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432. <https://doi.org/10.1093/bioinformatics/btq675>.
- Soucek, L., Jucker, R., Panacchia, L., Ricordy, R., Tatò, F., and Nasi, S. (2002). Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res* 62, 3507–3510.
- Strohl, W.R. (2015). Fusion proteins for half-life extension of biologics as a strategy to make biobetters. *BioDrugs* 29, 215–239.
- Tropea, J.E., Cherry, S., and Waugh, D.S. (2009). Expression and purification of soluble His⁶-tagged TEV protease. *Methods Mol. Biol.* 498, 297–307.
- Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhorji, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., et al. (2017). A pathology atlas of the human cancer transcriptome. *Science* 357.
- Wickham, H. (2011). ggplot2. *WIREs. Comp. Stat.* 3, 180–185.
- Zou, J., Maeder, M.L., Mali, P., Pruetz-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., et al. (2009). Gene Targeting of a Disease-Related Gene in Human Induced Pluripotent Stem and Embryonic Stem Cells. *Cell Stem Cell* 5, 97–110. <https://doi.org/10.1016/j.stem.2009.05.023>.
- Wickham, H., et al. (2019). Package “dplyr”. A grammar of data manipulation. R package version 0.8.0.1.
- Williams, J.G., Drugan, J.K., Yi, G.S., Clark, G.J., Der, C.J., and Campbell, S.L. (2000). Elucidation of binding determinants and functional consequences of Ras/Raf-cysteine-rich domain interactions. *J. Biol. Chem.* 275, 22172–22179.

- Xu, J., and Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* 26, 889–895.
- Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA* 117, 1496–1503. <https://doi.org/10.1073/pnas.1914677117>.
- Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 41, D955–D961.
- Yeh, C.H., Bellon, M., and Nicot, C. (2018). FBXW7: a critical tumor suppressor of human cancers. *Mol. Cancer* 17, 115.
- Yu, J., Wang, S., Yu, J., Liu, C., Xu, F., Wang, S., Yi, Y., and Yin, Y. (2017). Structure-based rational design of self-inhibitory peptides to disrupt the intermolecular interaction between the troponin subunits C and I in neuropathic pain. *Bioorg. Chem.* 73, 10–15.
- Yun, C.H., Boggan, T.J., Li, Y., Woo, M.S., Greulich, H., Meyerson, M., and Eck, M.J. (2007). Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* 11, 217–227.
- Zaidman, D., and Wolfson, H.J. (2016). PinaColada: peptide-inhibitor ant colony ad-hoc design algorithm. *Bioinformatics* 32, 2289–2296.
- Zhang, C., Wang, H.-J., Bao, Q.-C., Wang, L., Guo, T.-K., Chen, W.-L., Xu, L.-L., Zhou, H.-S., Bian, J.-L., Yang, Y.-R., et al. (2016). NRF2 promotes breast cancer cell proliferation and metastasis by increasing RhoA/ROCK pathway signal transduction. *Oncotarget* 7, 73593–73606.
- Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins* 57, 702–710.
- Zhao, Y., Aguilar, A., Bernard, D., and Wang, S. (2015). Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 inhibitors) in clinical trials for cancer treatment. *J. Med. Chem.* 58, 1038–1052.
- Zhu, J., Lu, M., and Zhu, L. (2016). Rational derivation of CETP self-binding helical peptides by π - π stacking and halogen bonding: therapeutic implication for atherosclerosis. *Bioorg. Chem.* 68, 259–264.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-GAPDH rabbit mAb	Cell Signaling	cat#2118S; RRID:AB_561053
anti-EGFR rabbit mAb	Cell Signaling	cat#4267S; RRID:AB_2246311
anti-RAS rabbit mAb	Abcam	cat#ab108602; RRID:AB_10891004
anti-FLAG mouse mAb	Millipore	cat#F1804; RRID:AB_262044
Anti-MouseIgG (HRP Conjugated)	Cell Signaling	cat#7076S; RRID:AB_330924
Anti-RabbitIgG (HRP Conjugated)	Cell Signaling	cat#7074S; RRID:AB_2099233
Anti-Mouse IgG (Dylight 488)	Abcam	cat#ab96871; RRID:AB_10680543
Anti-FLAG agarose beads	Millipore	cat#A2220-1ML; RRID:AB_10063035
Bacterial and virus strains		
StbI4	Thermo-Fisher	cat#11635018
T7 Express	New England BioLabs	cat#C2566H
Chemicals, peptides, and recombinant proteins		
Custom Synthesized Peptides	Genscript	cat#SC1208
Puromycin	Thermo Fisher	cat#A1113802
IPTG	Thermo Fisher	cat#15529019
Erlotinib	Sigma	cat#SML2156-50MG
Lipofectamine 2000	Thermo-Fisher	cat#11668030
KAPA Hifi Hotstart Ready Mix	Roche	cat#7958935001
Ampure XP beads	Beckman	cat# A63881
Genomic DNA Isolation Kit	Qiagen	cat# 69504
SeaBlock	Thermo	cat# 37527
RNA Purification Kit	Qiagen	cat# 74104
ProtoScript	NEB	cat# E6560S
RNA Sequencing Kit	NEB	cat# E7530L
PCR Purification Kit	Qiagen	cat# 28104
EcoRI	NEB	cat# R0101S
Gibson MasterMix	NEB	cat# E2611S
Plasmid DNA Maxiprep Kit	Qiagen	cat# 12963
iTaq Universal SYBR Green Supermix	BioRad	cat# 1725120
Polybrene	Millipore-Sigma	cat#TR-1003-G
Triton x-100	Millipore-Sigma	cat#X100-5ML
NP-40	Thermo-Fisher	cat#FNN0021
.05% Trypsin-EDTA	Thermo-Fisher	cat#25300062
DMEM	Thermo-Fisher	cat#10566016
.25% Trypsin-EDTA	Thermo-Fisher	cat#25200056
FBS	Thermo-Fisher	cat#10082147
SDS	BioRad	cat#1610418
Tris-Glycine	BioRad	cat#1610734
Ni-NTA Resin	Thermo-Fisher	cat#88223
Terrific Broth	Thermo-Fisher	cat#22711022
Carbenicillin	Teknova	cat#C2199
Critical commercial assays		
WST-8 Cell Viability Dye	Dojindo	cat#CK04-05
Deposited data		

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
TP53 Deep Mutational Scan Data	(Kotler et al., 2018)	https://doi.org/10.1016/j.molcel.2018.06.012
Interactome INSIDER	(Meyer et al., 2018)	http://interactomeinsider.yulab.org/
Sequencing Data: Peptide Overexpression Screens	This Paper	NCBI Sequencing Read Archive: PRJNA720162

Experimental models: Cell lines

MDA-MB-231	ATCC	cat#HTB-26
MDA-MB-468	ATCC	cat#HTB-132
Hs578T	ATCC	cat#HTB-126
MCF-7	ATCC	cat#HTB-22
HEK293T	ATCC	cat#CRL-3216

Oligonucleotides

PEP_01:GGCTAGGTAAGCTTGATA TCGGCCACCATG	IDT	25 nmole DNA Oligo
PEP_02:GGCGGCACTGTTTAAACA AGCCCGTCAGTAG	IDT	25 nmole DNA Oligo
PEP_03:ACACTCTTCCCTACAC GACGCTCTCCGATCTGCTTGA TATCGGCCACCATG	IDT	25 nmole DNA Oligo
PEP_04:GACTGGAGTTCAGACGT GTGCTCTCCGATCTCACTGTTT AACAAGCCCGTCAGTAG	IDT	25 nmole DNA Oligo
GAPDH_F: ACAGTCAGCCGCATCTTCTT	IDT	25 nmole DNA Oligo
GAPDH_R: ACGACCAAATCCGTTGACTC	IDT	25 nmole DNA Oligo
EF1a_seq:TTCTCAAGCCTCAGACAGTGG	IDT	25 nmole DNA Oligo

Recombinant DNA

pET Champion	Thermo-Fisher	cat#K630203
pCMV delta R8.2	Unpublished (Trono Lab)	Addgene cat#12263
pMD2.G	Unpublished (Trono Lab)	Addgene cat#12259
pEPIP	This Paper	N/A
Kras (G12V)-pcw107	Martz et al., 2014	Addgene cat#64602
pEGIP	Zou et al., 2009	Addgene cat#26777

Software and algorithms

ggplot2	https://cran.r-project.org/web/packages/ggplot2/index.html	Version 3.3
DESeq2	https://bioconductor.org/packages/release/bioc/html/DESeq2.html	Version 3.12
Dplyr	https://cran.r-project.org/web/packages/dplyr/index.html	Version 1.05
RcppRoll	https://cran.r-project.org/web/packages/RcppRoll/	Version 0.3.0
Readr	https://cran.r-project.org/web/packages/readr/index.html	Version 1.4.0
Peptides	https://cran.r-project.org/web/packages/Peptides/index.html	Version 2.4.3
Biostrings	https://bioconductor.org/packages/release/bioc/html/Biostrings.html	Version 3.12
ggrepel	https://cran.r-project.org/web/packages/ggrepel/index.html	Version 0.9.1
hexbin	https://cran.r-project.org/web/packages/hexbin/index.html	Version 1.28.2

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pheatmap	https://cran.r-project.org/web/packages/pheatmap/index.html	Version 1.0.12
fgsea	https://bioconductor.org/packages/release/bioc/html/fgsea.html%20-%20version%203.12	Version 3.12
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	Version 2.4.2
MaGeCK	https://sourceforge.net/p/mageck/wiki/Home/	Version 0.5.9
Robetta	https://rosetta.bakerlab.org/	N/A
STRIDE	http://webclu.bio.wzw.tum.de/stride/	N/A
PyMol	https://pymol.org/2/	Version 2.3.3
FLASH	https://ccb.jhu.edu/software/FLASH/	Version 1.2.11
STAR	https://github.com/alexdobin/STAR	Version 2.7.8.a

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prashant Mali (pmali@eng.ucsd.edu).

Materials availability

Plasmids generated in this study will be distributed via Addgene and upon request.

Data and code availability

- Source data related to this manuscript can be found in the supplemental files, with raw sequencing data publicly accessible on the NCBI Sequence Read Archive via accession number SRA: [PRJNA720162](https://www.ncbi.nlm.nih.gov/sra/PRJNA720162).
- This paper does not report original code.
- Additional scripts used to generate the figures reported in this paper are available in the packages listed in the key resources table, and their specific use is described in the [STAR methods](#).
- Any additional information required to reproduce this work is available from the lead contact.

METHOD DETAILS

Design of peptide coding gene fragment libraries

Peptide coding gene fragments from target genes were composed of the DNA coding sequence for all 40mer amino acids from the genes/mutants listed in [Figures S2](#) and [S3](#) and the main text. For fitness screens the 5' and 3' ends of each gene fragment were modified to contain a start and stop codon, as well as ~20bp of DNA homologous to the expression plasmid for downstream Gibson cloning.

Cancer driver gene fragment cloning

Peptide coding gene fragment libraries were synthesized as pooled single stranded oligonucleotides by Custom Array. These oligonucleotides were then PCR amplified using KAPA-HiFi (Kapa Biosystems) to generate double stranded gene fragments compatible with Gibson cloning. 50µl PCR reactions were set up with 25ng of pooled oligonucleotide template and 2.5 µl of primers PEP_1 and PEP_2 (10µM). The thermal cycler was programmed to run at 95C for 3 minutes, followed by 12 cycles of 98C for 20 seconds, 65C for 15 seconds, and 72C for 45 seconds. This was followed by a final 5-minute extension at 72C. PCR products were then purified using the QIAquick PCR purification kit. See [Table S5](#) for primer sequences.

The peptide overexpression vector pEPIP was generated from a modified pEGIP (Addgene #26777). The vector was modified to remove the GFP insert, insert an EcoRI cloning site, and add primer binding regions with which to amplify the libraries for HTS. To clone the gene fragment libraries into the expression vector, pEPIP was first digested with EcoRI (NEB) for 3 hours at 37C. The linearized vector was then column purified using the QIAquick PCR purification kit. Subsequently, Gibson assembly was used to clone the gene fragment libraries into the pEPIP vector. For each reaction, 10µl of Gibson Reaction MasterMix (NEB) was combined with 100ng of the vector and 50ng of the double stranded gene fragment library, with H₂O up to 20µl. The Gibson reactions were then incubated at 50C for 1hr and transformed via electroporation into 200µl of ElectroMAX Stbl4 competent cells per 10,000 library

elements (Invitrogen) according to the manufacturer's protocol. The Stbl4 cells were then resuspended in 4mL of SOC media and placed at 37C with shaking for 1hr to recover. After recovering, 1 μ l of the SOC/cell suspension was spread on LB-carbenicillin plates to calculate library coverage, with the remaining SOC/cells used to inoculate a 100ml culture of LB-carbenicillin. Greater than 200 fold library coverage was obtained to ensure all gene fragments were well represented. After 16 hr of incubation at 37C with shaking, plasmid DNA was isolated via a Qiagen Plasmid Plus Maxi Kit.

Lentivirus production

Replication deficient lentiviral particles were produced in HEK293T cells (ATCC) via transient transfection. HEK293T cells were grown in DMEM media (Gibco) supplemented with 10%FBS (Gibco). The day before transfection, HEK293T cells were seeded in a 15cm dish at ~40% confluency. The day of transfection, the culture media was changed to fresh DMEM plus 10% FBS. At the same time, 3ml of Optimem reduced serum media (Life Technologies) was mixed with 36 μ l of lipofectamine 2000, 3 μ g of pMD2.G plasmid (Addgene #12259), 12 μ g of pCMV deltaR8.2 plasmid (Addgene #12263), and 9 μ g of the gene fragment plasmid library. After 30 minutes of incubation, the plasmid/lipofectamine mixture was added dropwise to the HEK293FT cells. Supernatant containing viral particles was harvested 48 and 72 hours after transfection and concentrated to 1ml using Amicon Ultra-15 centrifugal filters with a cutoff 100,000 NMWL (Millipore). The viral particles were then aliquoted and frozen at -80C until further use.

Fitness screening in mammalian cell lines

Hs578T cells and MDA-MB-231 cells were cultured in DMEM media supplemented with 10% FBS. Cells were transduced with the peptide coding gene fragment library at an MOI <3 to ensure each cell received a single construct. Viral transduction was performed in media containing 8 μ g/ml polybrene to improve transduction efficiency. For each cell line, screening was conducted with two biological replicates. 24 hours after transduction the cell culture media was changed back to DMEM without polybrene supplementation. 48 hours after transduction, the cell culture media was changed to DMEM containing puromycin to select for transduced cells. 2 μ g/ml puromycin was used to select the Hs578T cells, and 3.5 μ g/ml puromycin was used to select the MDA-MB-231 cells. In the pilot screens, more than 6,000,000 cells (from each cell line) were transduced to ensure greater than 1000-fold coverage of the library. The cells were cultured for 14 days after transduction, with genomic DNA isolated via a Qiagen DNeasy Blood and Tissue Kit at days 3 and 14. For the larger screens, the number of cells transduced was scaled up accordingly.

HTS library preparation and sequencing

Peptide coding gene fragments for each time point and replicate were then amplified from the genomic DNA using Kapa HiFi. The fragments serve as their own barcodes for downstream abundance calculations. Illumina compatible libraries were prepared using 2.5 μ l of primers PEP_3 and PEP_4 (10 μ M) per 50 μ l reaction. For each sample (i.e. time point and replicate) from the pilot library, 10 separate 50 μ l PCR reactions with 4 μ g of gDNA each (40 μ g total) were performed to ensure adequate library coverage. Thermal cycling parameters were identical to those used to amplify the gene fragment oligos, with the exception that the gDNA required 26 cycles to amplify. Ampure XP beads were used to purify all samples for sequencing. NEBNext Multiplexed Oligos for Illumina (NEB) were then used to index the samples, and 150bp single end reads were then generated via an Illumina HiSeq2500. Greater than 500-fold sequencing depth was used to ensure accurate abundance quantitation. For the larger libraries, the number of PCR reactions was scaled to process 300 μ g of total gDNA per timepoint and replicate. The larger libraries were then sequenced with 100-bp paired end reads generated via an Illumina HiSeq4000.

Processing of sequencing files

To quantify peptide coding gene fragment relative abundance, the library definition text file (containing gene fragment names and sequences) was first converted into Fasta format. This Fasta file was then used to build a Bowtie2 index file. For the pilot library, raw FASTQ reads were directly mapped to the library index file via Bowtie2 (Langmead and Salzberg, 2012). For the expanded libraries paired end reads were first merged into a single FASTQ file via FLASH (Fast Length Adjustment of SHort reads)(Magoč and Salzberg, 2011). For both libraries, reads with insertion or deletion mutations were removed to eliminate spurious data resulting from out of frame gene fragments, retaining 35-40% of total reads. Reads aligning to mutant peptides were filtered to retain only perfect matches (to prevent miscalling of mutant alleles). The resulting SAM files were then compressed to BAM files via SAMtools (Li et al., 2009). Following this, the count and test modules in MAGeCK were used to determine the median normalized peptide coding gene fragment abundances from the alignment files and individual peptide log fold change and depletion p-values.(Li et al., 2009, 2014). Following this, the R packages "Peptides" and "Biostrings" were used to determine peptide biophysical parameters such as charge and hydrophobicity(Osorio et al., 2015; Pagès et al., 2017).

Calculation of amino acid level fitness scores

After generating the peptide count files, all downstream analysis was performed in R. For each amino acid residue in the overall protein structure, an amino acid level log fold change was calculated by taking the mean log₂ fold change of all overlapping peptides with greater than 30 raw counts in both replicates of the day 3 timepoint. Then, for every residue in the protein scaffolds, a normalized fitness score was calculated by taking this mean log₂ fold change value (x) and Z-normalizing to the library wide amino acid log₂ fold change standard deviation (σ) and mean (μ).

$$\text{Fitness Score} = Z = \frac{X - \mu}{\sigma}$$

To identify amino acid positions which were significantly depleted, a one tailed permutation test was performed. The approximate permutation distribution of amino acid fitness scores was generated by randomly shuffling the labels of all gene fragments in the screen. This shuffled data was subsequently used to recalculate the amino acid fitness scores. This resampling procedure was then repeated $N=10,000$ times, with the P values for each amino acid position calculated by the following:

$$P = \frac{1 + \sum_{i=1}^N [\text{Fitness}_{\text{Permuted}} < \text{Fitness}_{\text{Observed}}]}{N \text{ permutations}}$$

These P values were then adjusted for multiple comparison testing by the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). The R packages “ggplot2”, “hexbin”, “ggrepel”, “dplyr”, and “RcppRoll” were used to generate publication quality figures (Wickham, 2011; ; Wickham et al., 2019).

Validating highly depleted gene fragments

All cell lines used were cultured in DMEM media supplemented with 10% FBS. The fitness impact of highly depleted peptides was tested in an arrayed format via a WST-8 (Dojindo) cell growth assay. Highly depleted peptide coding gene fragments were synthesized by Twist Biosciences, cloned directly into the pEPIP vector, and subsequently packaged into lentiviral particles. Cells were transduced at an MOI of 4, and switched to puromycin containing media after 48 hours. Following 24 hours of puromycin selection, 1,500 cells were seeded per well as biological replicates in a 96 well plate. All experimental groups for Hs578T cells had $n=4$. For the first set of validations in MDA-MB-231 cells, all experimental groups had $n=4$, with the exception of the GFP control which had $n=8$. For the second panel of experiments (DICER1-552, etc.) all experimental groups had $n=6$. For HEK293T and MCF-7 cells all experimental groups had $n=8$. $2\mu\text{g/ml}$ puromycin was used to select Hs578T and MCF-7 cells, while $3.5\mu\text{g/ml}$ puromycin was used to select MDA-MB-231 and HEK293T cells. Cell growth was then quantified via absorbance at 450nm following 1.5hrs of incubation with WST-8 reagent. A two-tailed P value was then calculated via an unpaired t-test with Welch’s correction.

Crystal violet viability measurements

In Figures 6D, S5C, and S5D, relative cell viability was determined via Crystal Violet staining. At the experimental endpoint cells were washed once with PBS, and subsequently incubated in $50\mu\text{l}$ of crystal violet stain solution (.5% w/v Crystal Violet, 20% v/v methanol in DI water) for 15 minutes. Following this, excess crystal violet was removed from the plates via five immersions in 2 liters of DI water. The plates were allowed to dry overnight, and the next morning the crystal violet stain was solubilized with 1% v/v SDS in DI water, and relative cell numbers were quantified via absorbance at 595nm.

Engineering peptides for exogenous delivery

Peptides shown in Figure 5B were fused to an N-terminal cell penetrating motif via a $(\text{GS})_3$ linker sequence (Table S8) and chemically synthesized by GenScript’s Custom Peptide Synthesis service at crude purity. For dose response experiments, cells were plated in 96 well plates ($n=4$) at 50% confluency and peptides were added at the indicated concentrations with cell viability quantified after 24hrs via the WST-8 assay. Cell viability was normalized to that of an untreated control on the same plate.

Co-immunoprecipitation

HEK293T cells were seeded in 6 well plates to be 75% confluent on the day of transfection. Transfections were performed with $1\mu\text{g}$ of each indicated plasmid per well with $5\mu\text{l}$ of Lipofectamine 2000 according to the manufacturers protocol. For the RAF1-73 experiments, 48 hours after transfection, cells were washed twice with ice cold PBS and lysed for 30 minutes in ice cold $400\mu\text{l}$ TBS buffer containing .5% Triton x-100, 1mM EDTA, and Halt Protease Inhibitor Cocktail (Thermo Fisher 78429). The supernatant was then clarified by centrifugation at 14,000G for 15 minutes. Following this, immunoprecipitation of FLAG tagged constructs was performed by adding $300\mu\text{l}$ of the lysate to $20\mu\text{l}$ of packed anti FLAG agarose beads (Millipore Sigma A2220) prewashed with TBS. The remaining $100\mu\text{l}$ of lysate was stored at -80C for later analysis. The bead-lysate mixture was then mixed end over end at 4C for 2 hours. After binding to the beads, the bead-protein complexes were washed three times with 1ml lysis buffer and eluted with $20\mu\text{l}$ of 2x SDS-PAGE Laemmli loading buffer (BioRad 1610737). The EGFR-697 Co-IP experiments were performed identically, with the exception that .75% NP-40 was used instead of Triton x-100 for cell lysis.

Western blotting

For the RAF1-73 Co-IP experiments proteins were first separated on 4-20% polyacrylamide gels (BioRad 4561094) under denaturing conditions in Tris-Glycine-SDS (BioRad 1610732) for 1 hour at 100V. Following this, proteins were transferred to $.2\mu\text{m}$ nitrocellulose membranes (BioRad 1620112) for 30 minutes at 100V in Tris-Glycine buffer (BioRad 1610734) containing 30% methanol. Membranes were then blocked for 1 hour in TBS-T (Cell Signaling 9997) containing 5% non fat dry milk (BioRad 1706404XTU). The EGFR-697 experiments and EGFR expression level testing were performed identically, with the exception that the transfer voltage was reduced to 30V and performed overnight at 4C . Primary antibodies were then added (diluted 1:1000 in TBS-T+ 5% milk) and incubated overnight at 4C with gentle agitation. The following day the membranes were washed three times in TBS-T and incubated for 1 hour with

HRP conjugated secondary antibodies (diluted 1:10,000 in TBS-T + 5% milk) at room temp. The membranes were then washed again three times with TBS-T and developed using SuperSignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher 34577).

qPCR

Cells were plated the day before transduction at approximately 20% confluency. On the day of transduction, cells were transduced with the appropriate lentiviral constructs at an MOI of 4 and allowed to grow for 72 hours. RNA was subsequently isolated with an RNEasy Kit (Qiagen) with on column DNase I treatment. Following this, cDNA was generated using the ProtoScript II First Strand cDNA Synthesis Kit (NEB) and diluted up to 1:4 with nuclease-free water. The qPCR reactions were setup as: 2 μ l cDNA, 400 nM of each primer (See [Table S5](#)), 2X iTaq Universal SYBR Green Supermix (BioRad), with ultra pure water up to 20 μ l. The qPCR was performed using a CFX Connect Real Time PCR Detection System (Bio-Rad) with the following parameters: 95°C for 3 min; 95°C for 3 s; 60°C for 20s, for 40 cycles. All experiments were performed in duplicate and results were normalized against a house-keeping gene, GAPDH. Relative mRNA expression levels (normalized to GAPDH) were determined by the comparative cycle threshold (Ct) method.

Immunofluorescence

Cells were plated the day before transduction at approximately 20% confluency. On the day of transduction, cells were transduced with the appropriate lentiviral constructs at an MOI of 4 and allowed to grow for 72 hours. Following this, the cells were washed twice with PBS and fixed for 30 minutes at room temperature with 4% paraformaldehyde. Cells were then washed three times with PBS and blocked for 1 hour at room temp with PBS plus 5% Sea Block (Thermo Fisher PI37527X3) and 2% Triton x-100. The blocking buffer was then aspirated and replaced with blocking buffer plus anti-FLAG primary antibody at a 1:500 dilution. The primary antibody was then allowed to bind overnight at 4C. The following day, the cells were washed three times with PBS, and incubated for 1 hour with a secondary anti-mouse IgG antibody conjugated to DyLight 488 (diluted 1:200). The cells were then washed three times with PBS and subsequently imaged via fluorescence microscopy.

RNA-seq of highly depleted fragments

RNA sequencing was performed on Hs578T cells 6 days after transduction with lentivirus expressing gene fragments of interest. Two biological replicates were sequenced for each experimental condition. Total RNA was isolated from cells via an RNEasy Kit (Qiagen) with on column DNase I treatment. An NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490S) was then used to deplete rRNA. Subsequently, an NEBNext Ultra RNA Library Prep Kit (E7530S) was used to generate Illumina compatible RNA sequencing libraries. Sequencing was performed on an Illumina HiSeq4000, with paired end 100bp reads. Reads were aligned to the human reference transcriptome via the STAR aligner, and differential gene expression was performed using DESeq2. Differential expression was tested in reference to a control group transduced with lentivirus coding for GFP. Following this, the R package “fgsea” was used to conduct GSEA pre-ranked analysis ([Dobin et al., 2013](#); [Love et al., 2014](#); [Sergushichev, 2016](#)). Genes were ranked via the shrunken log fold change values outputted by DESeq2.

Network visualization

Network of protein-protein interactions was generated using publicly available data from Interactome INSIDER ([Meyer et al., 2018](#)). Edges were drawn for all high confidence interaction interfaces calculated from PDB structures, homology models, and the “Very High” and “High” interface potential categories from ECLAIR. Node color was based on fitness scores for each gene available via DepMap CRISPR knockout screening. The CERES normalized gene effects were used to quantify the fitness impact of a given knockout. Visualization was then performed in CytoScape ([Smoot et al., 2011](#)).

Computational modeling of peptide structure

To computationally predict 40-mer peptide structures, amino acid sequences for RAF1 and EGFR peptides were submitted to the Robetta service, a protein structure prediction service hosted by the Baker Lab at University of Washington ([Kim et al., 2004](#)). TrRosetta, a deep learning-based structure prediction method, was used for all submissions to the server ([Yang et al., 2020](#)). Regions of the protein of interest with available crystal structures from the RCSB Protein Data Bank were fragmented and used to evaluate the folded structure of the computationally modeled fragments (see [Table S7](#)). PyMOL was then used to visualize the predicted structures as well as the available crystal structures from the database. To evaluate the similarity between the modeled peptides and those from the crystal structure, the TM score (template modeling score) was used ([Zhang and Skolnick, 2004](#)). To evaluate the TM-scores of the fragments as a function of the secondary structure of the native protein, we extracted the structural annotations of the RAF1 and EGFR proteins from the PDB structure files available on RCSB. We then defined a fragment as containing a secondary structure if it had a minimum overlap of 3 amino acids with the corresponding annotated regions. A minimum overlap of 3 was chosen as the shortest annotated secondary structure in the native proteins is an alpha helix containing 3 amino acids. The confidence scores of the predicted peptide structures were given as the predicted Local Distance Difference Test (IDDT) as determined by DeepAccNet ([Hiranuma et al., 2021](#)). Validated IDDT baseline scores for proteins with the wrong fold are 0.20 with a mean absolute deviation of 0.04 ([Mariani et al., 2013](#)). The secondary structures of both the native structure and predicted structured were assigned through STRIDE ([Frishman and Argos, 1995](#)).

Recombinant peptide production

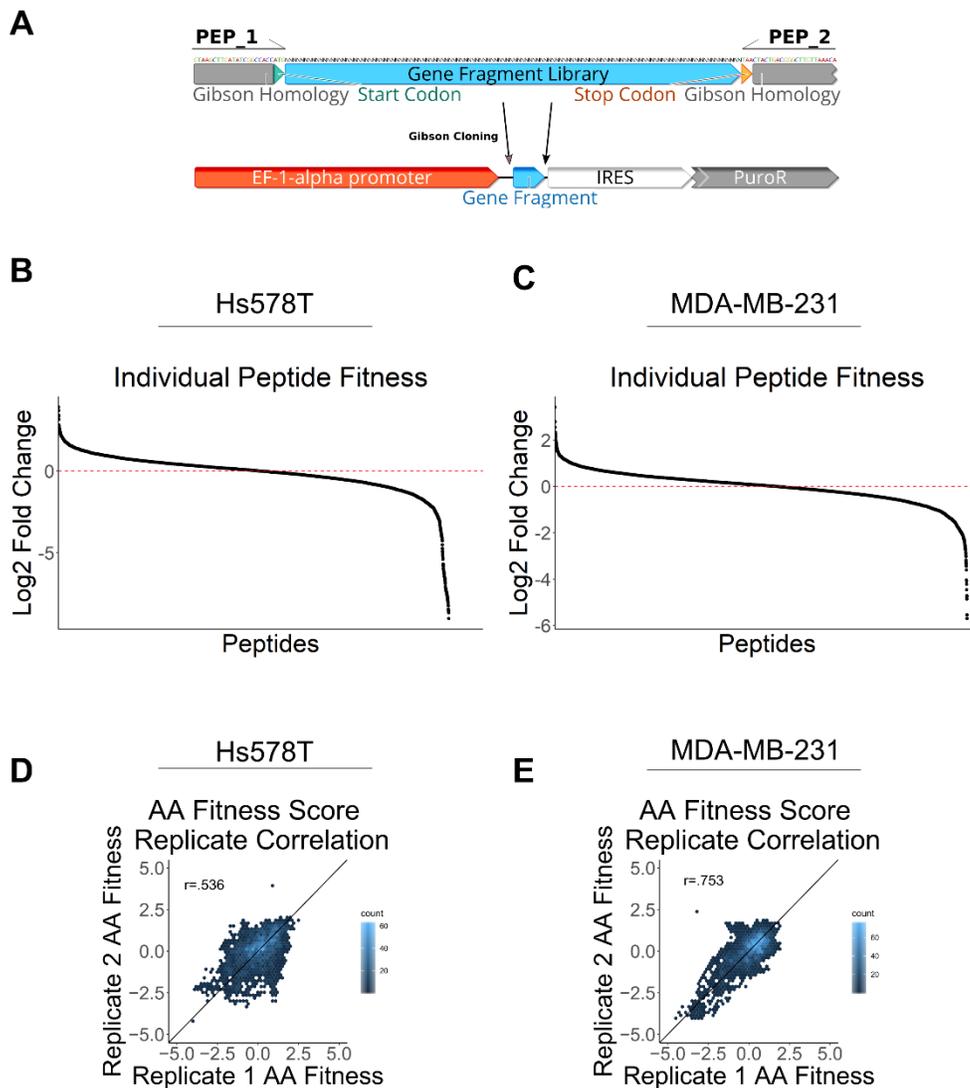
Recombinant production protocol was adapted from (Tropea et al., 2009)(Tropea et al., 2009). Recombinant MBP fusions and TEV protease were cloned into the pET Champion vector (Thermo K630203) and expressed in T7 express E. coli (NEB C25661). Constructs were ordered as gBlocks from IDT and cloned directly into the vector via Gibson Assembly. To produce high yield MBP-peptide fusions and TEV protease, a 10mL starter culture of E. coli was grown for 14 hours at 37C in TB media. This starter culture was then used to induce a 1L culture of TB media. This culture was grown at 37C until an OD of 0.8, and then induced with 0.5mM IPTG. The cells were subsequently grown overnight at 25C, following which the cells were pelleted and stored at -20C. To isolate recombinant proteins, cells were first lysed via mechanical disruption with mortar and pestle in liquid nitrogen and resuspended in binding buffer (50mL 50mM sodium phosphate, 200mM NaCl, 10% glycerol, and 25mM imidazole at pH 8.0). Cell lysate was then clarified via centrifugation for 30 minutes at 20,000g. Following this, the soluble fraction of the lysate was applied via gravity flow to 5mL of a pre-equilibrated Ni-NTA resin (Thermo 88221). The resin was subsequently washed with 15 column volumes of binding buffer, and eluted with 50mM sodium phosphate, 200mM NaCl, 10% glycerol and 250mM imidazole at pH 8.0. Purified TEV protease and the MBP-peptide fusions were subsequently dialyzed into cleavage buffer (50mM sodium phosphate, 200mM NaCl, pH 7.4) using Amicon 3kD MWCO centrifugal spin filters (Millipore UFC800324). Cleavage reactions were set up in cleavage buffer containing 2mg/mL MBP-peptide fusion, 0.2mg/mL TEV protease, and 1mM DTT (added fresh). This reaction was allowed to proceed overnight at 25C. The following day, the cleavage reaction was diluted 1:8 with binding buffer and applied over a pre-equilibrated Ni-NTA resin to remove the TEV protease and MBP proteins (1mL resin per 5mg fusion protein). The flow through (containing purified peptide) was subsequently dialyzed into PBS and concentrated to 5mg/mL.

Cell Systems, Volume 12

Supplemental information

**Peptide-tiling screens of cancer drivers
reveal oncogenic protein domains
and associated peptide inhibitors**

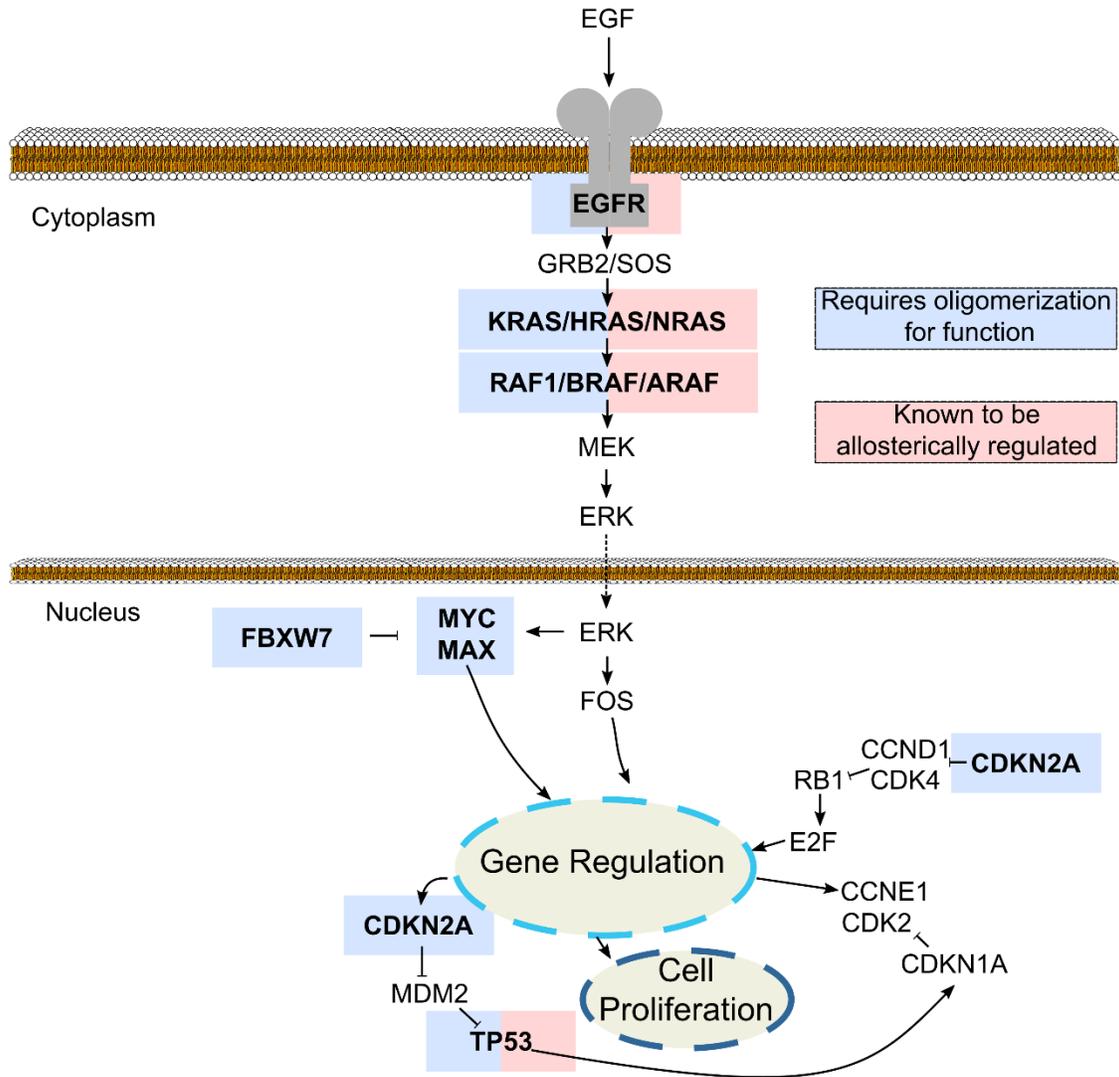
Kyle M. Ford, Rebecca Panwala, Dai-Hua Chen, Andrew Portell, Nathan Palmer, and Prashant Mali



SUPPLEMENTAL FIGURE 1. Cloning strategy and MAPK focused screen overall analyses:

(A) Detailed overview of library construction. Library was ordered as single stranded DNA oligos from Custom Array, and subsequently amplified via PCR to generate gene fragment libraries compatible with Gibson assembly cloning. This library was then cloned into pEPIP, with library coverage determined via high throughput sequencing. **(B-C)** Initial analysis for pooled pilot screen in Hs578T and MDA-MB-231 cells. The majority of peptides tested did not drop out during the fitness screen, although the distribution of peptide log fold change values is skewed towards depletion rather than enrichment. **(D-E)** The computed fitness scores for the amino acid positions showed good correlation between replicates in both Hs578T and MDA-MB-231. ($r=0.536$ and

$r=.753$ respectively). The majority of amino acid positions scored have no significant depletion, with a small subset having a detectable impact on fitness. **(Related to Figure 1)**



SUPPLEMENTAL FIGURE 2. Overview of MAPK focused peptide overexpression library:

Peptide coding sequences derived from the WT coding sequences of key genes within the MAPK signaling pathway were synthesized as an oligonucleotide pool and subsequently cloned into a lentiviral overexpression vector. Proteins within the MAPK signaling pathway drive cellular proliferation through a cascade of physical interactions with proteins, nucleic acids, and other effector molecules within cells. **(Related to Figure 1)**

A

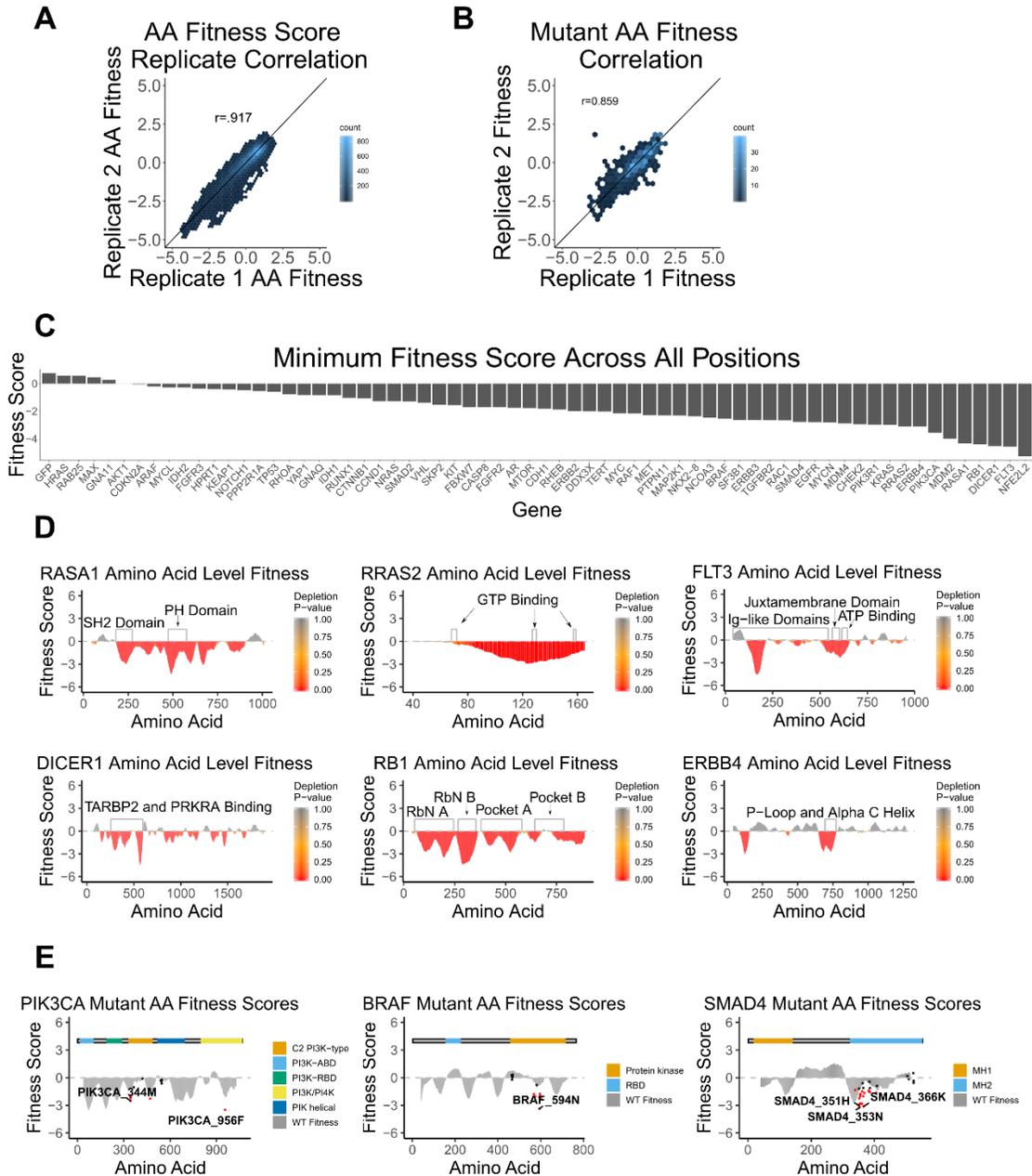
Controls	Apoptosis	Cell Cycle	Chromatin Histone Modifiers	Genome Integrity	MAPK Signaling	Metabolism	NOTCH Signaling	Other Signaling
HPRT1	CASP8	CCND1	NCOA3	CHEK2	ARAF	IDH1	NOTCH1	AR
GFP		CDKN2A RB1		MDM2 MDM4 TERT TP53	BRAF HRAS KRAS MAP2K1 NRAS RAB25 RAF1 RRAS2	IDH2		CDH1 GNA11 GNAQ KEAP1 PTPN11 RAC1 RHOA
PI3K Signaling	Protein Homeostasis	RNA Abundance	RTK Signaling	Splicing	TGFB Signaling	TOR Signaling	Transcription Factor	Wnt/B-Catenin Signaling
AKT1	FBXW7	DDX3X	EGFR	SF3B1	SMAD2	MTOR	MAX	CTNNB1
PIK3CA	SKP2	DICER1	ERBB2		SMAD4	RHEB	MYC	
PIK3R1	VHL		ERBB3		TGFBR2		MYCL	
PPP2R1A			ERBB4				MYCN	
			FGFR2				NFE2L2	
			FGFR3				NKX2-8	
			FLT3				RUNX1	
			KIT				YAP1	
			MET					
			RASA1					

B

Genes	# of Mutants	Genes	# of Mutants	Genes	# of Mutants
Controls		NOTCH Signaling		RTK Signaling	
HPRT1	NA	NOTCH1	3	EGFR	19
GFP	NA	Other Signaling		ERBB2	3
Apoptosis		CDH1	4	ERBB3	3
CASP8	3	GNA11	4	ERBB4	3
Cell Cycle		GNAQ	4	FGFR2	3
CDKN2A	3	KEAP1	9	FGFR3	2
RB1	1	PTPN11	15	FLT3	2
Chromatin Histone Modifiers		RAC1	8	KIT	9
EP300	8	RHOA	16	MET	4
KMT2C	2	PI3K Signaling		RASA1	1
SETD2	5	AKT1	3	Splicing	
Genome Integrity		PIK3CA	12	SF3B1	16
CHEK2	6	PIK3R1	11	TGFB Signaling	
TP53	195	PPP2R1A	13	SMAD2	2
MAPK Signaling		Protein Homeostasis		SMAD4	31
BRAF	16	FBXW7	13	TGFBR2	3
HRAS	11	VHL	24	TOR Signaling	
KRAS	16	SPOP	2	MTOR	9
MAP2K1	4	RNA Abundance		RHEB	1
NRAS	9	DDX3X	3	Transcription Factor	
RRAS2	2	DICER1	9	NFE2L2	9
Metabolism		ZFP36L2	1	RUNX1	2
IDH1	2			Wnt/B-Catenin Signaling	
IDH2	5			CTNNB1	15

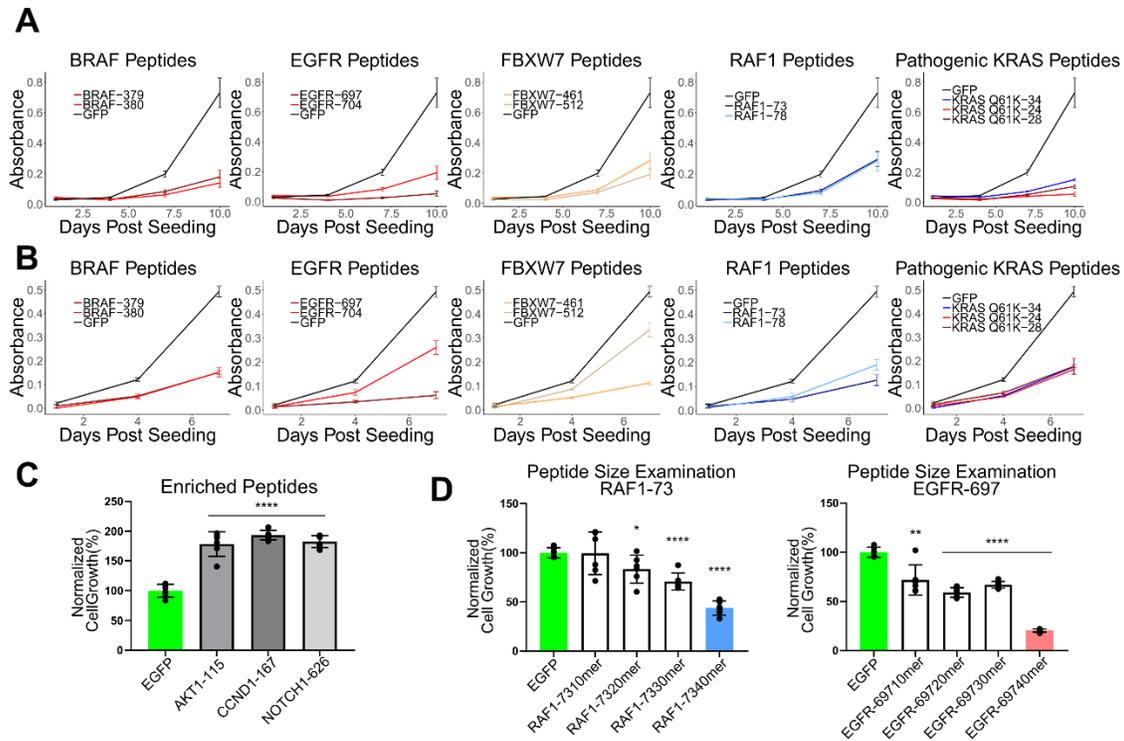
SUPPLEMENTAL FIGURE 3. Library composition for secondary expanded cancer driver screens: (A) Table detailing all the peptides assayed in the expanded wildtype driver screen.

Genes were sourced from Bailey et al. 2018 and Santarius et al. 2010 (Santarius *et al.*, 2010), comprising diverse cancer associated signaling pathways and processes. **(B)** Table detailing all the peptides assayed in the mutant screen. Mutant genes cover a wide range of signaling pathways and molecular functions. **(Related to Figure 2-4)**



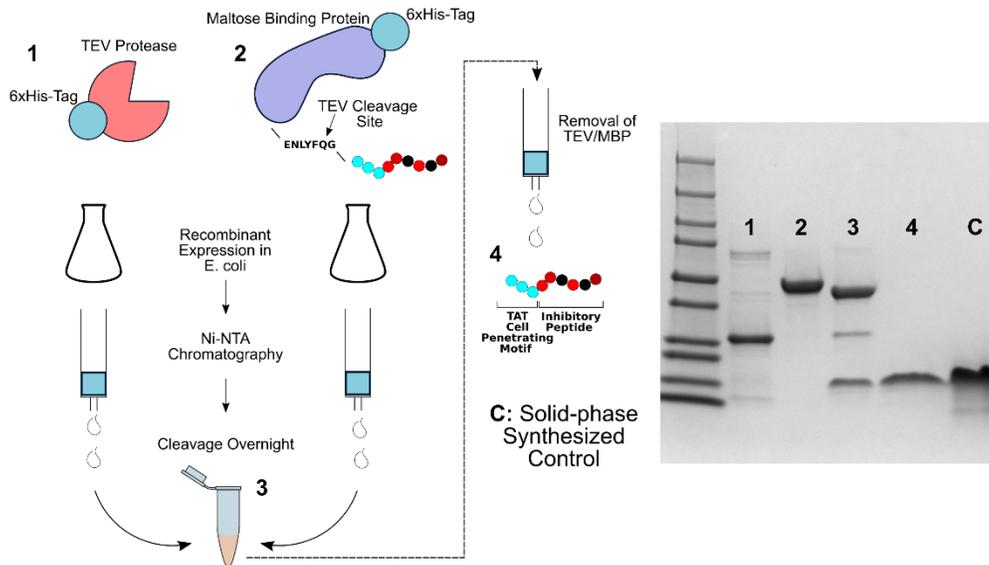
SUPPLEMENTAL FIGURE 4. Quality control metrics and amino acid level fitness plots for expanded cancer driver screens: (A) Computed per position amino acid scores had good correlation between replicates (Pearson correlation = .917), with reproducibility exceeding that of the pilot screen. (B) Replicate correlation for the mutant peptide screen. Screen shows high degree of reproducibility (Pearson correlation = .859). (C) The fitness score for the most deleterious residue in each full-length protein is plotted for each gene. GFP and HPRT1 controls show little effect on cell fitness across protein structure. (D) Per position fitness scores for RASA1,

RRAS2, FLT3, DICER1, RB1, and ERBB4. Select PPIs are annotated on the plots, corresponding to regions of significant depletion (**E**) Plot of wild type (gray bars) and mutant amino acid fitness scores (points) for PIK3CA, BRAF, and SMAD4. Dots indicate mutant amino acid fitness scores at the specified positions. Dots labeled in red were significantly (BH adjusted P value < .05) depleted in the pooled screen. (**Related to Figures 2-4**)

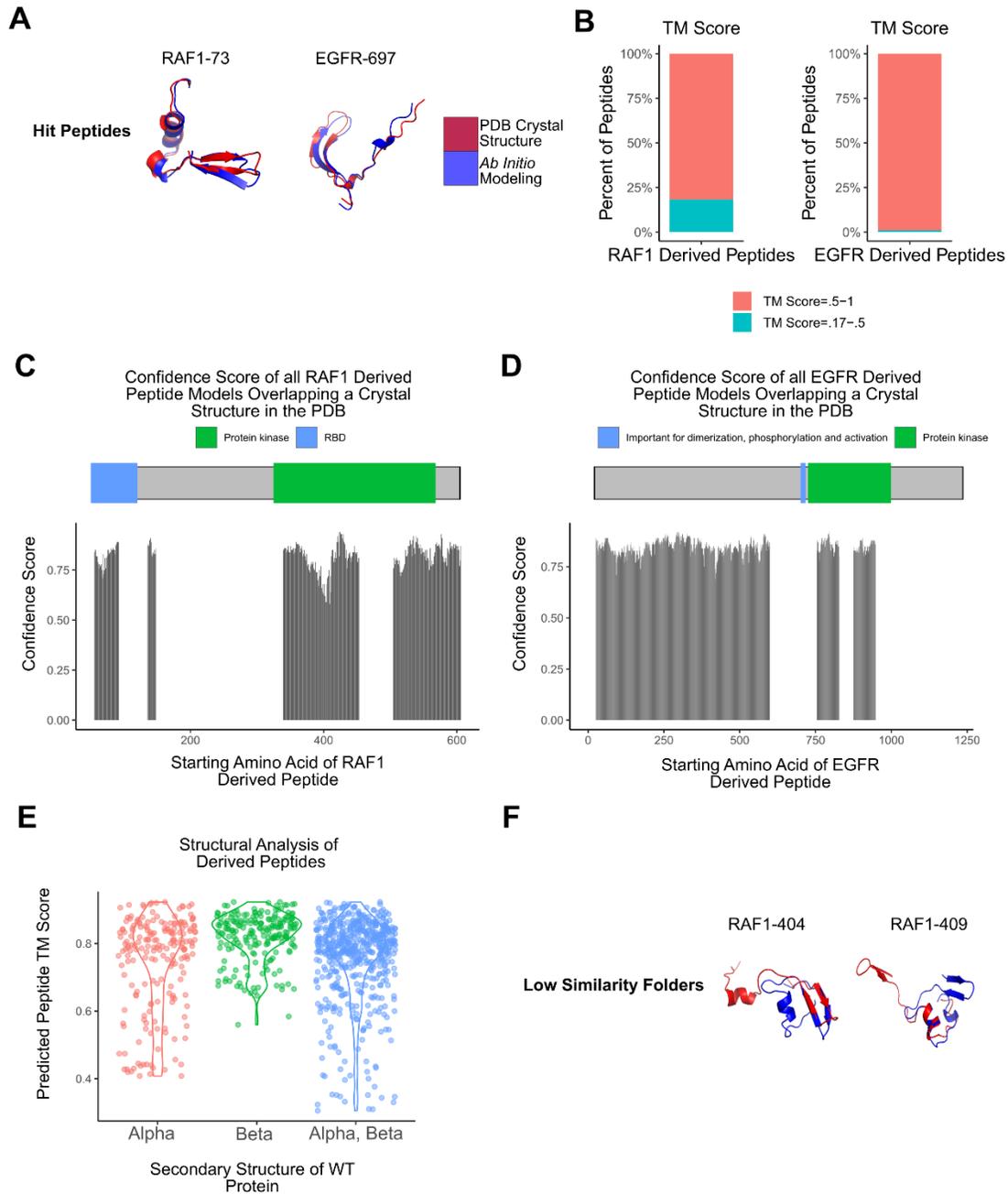


SUPPLEMENTAL FIGURE 5. Validation of hit peptide activity: (A) Growth kinetics in Hs578T for individual peptide variants shown in **Figure 3A**. Cell growth was quantified via the WST-8 proliferation assay. Results are from the same experiment split into multiple plots for ease of visualization, hence identical GFP controls for each peptide group. Arrayed validation of lentivirally delivered gene fragments derived from KRAS mutants is also shown. KRAS61K mutant peptides predicted to be deleterious to cell growth significantly inhibited growth ($P < .05$, as measured at the 7 day time point). **(B)** Growth kinetics in MDA-MB-231 for individual peptide variants shown in **Figure 3A**. Cell growth was quantified via the WST-8 proliferation assay. Results are from the same experiment split into multiple plots for ease of visualization, hence identical GFP controls for each peptide group. Arrayed validation of lentivirally delivered gene fragments derived from KRAS mutants is also shown. KRAS61K mutant peptides predicted to be deleterious to cell growth significantly inhibited growth ($P < .05$, as measured at the 7 day time point). **(C)** Significantly enriched peptides identified from the larger screen in MDA-MB-231 cells were tested in an arrayed format to validate the growth advantage phenotype. Cells were transduced with lentivirus to overexpress each construct, selected with puromycin and subsequently seeded into a 96 well plate to quantitate relative growth rates. After seven days the relative cell numbers for each construct were then measured via crystal violet staining. Bar plots show mean with error bars showing standard deviation, statistical tests comparing cell growth

relative to GFP control (* $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$). **(D)** Effect of varying peptide length on cell fitness. Peptides centered on the previously identified hits RAF1-73 and EGFR-697 were overexpressed via lentiviral transduction in MDA-MB-231 cells. After 7 days of competitive growth, relative cell numbers were quantified via crystal violet staining. Bar plots show mean with error bars showing standard deviation, statistical tests comparing cell growth relative to GFP control (* $P < .05$, ** $P < .01$, *** $P < .001$, **** $P < .0001$). **(Related to Figure 3)**



SUPPLEMENTAL FIGURE 6. Recombinant production of peptides for exogenous delivery: Peptide production protocol to facilitate translation of peptide hits. Tagless peptides conjugated to cell penetrating protein TAT were produced at high purity via fusion to Maltose Binding Protein (MBP), and subsequent cleavage by TEV protease. The protocol makes use of no specialized instruments, and is easily adaptable to alternative cell penetrating motifs or peptide constructs. Ladder has bands marking 10,15,20,25,37,50,75,100,150,and 250kD. **(Related to Figure 5)**



SUPPLEMENTAL FIGURE 7. Peptide structural analysis: (A) Structural alignments of predicted peptide structures to experimentally resolved crystal structures of the full length protein (modeled using Tr Rosetta – **STAR Methods**). Shown are two hit peptides (RAF1-73 and EGFR-697). (B) Template modeling scores (TM-Score) for all peptides derived from RAF1 and EGFR. We comprehensively modeled 957 total peptides derived from RAF1/EGFR which had available overlapping crystal structures on RCSB (see **Supplemental Table 7**). We found that all modeled peptides had structural similarities with the WT structure greater than random chance (TM-score

>.17), and over 75% of the modeled peptides in both proteins had approximately the same fold as the WT structure (TM-score >.5). **(C)** Confidence scores for the predicted RAF1 peptide models outputted by TrRosetta. Confidence scores shown are the predicted Local Distance Difference Test (IDDT) values for the protein as determined by DeepAccNet(Hiranuma *et al.*, 2021). **(D)** Confidence scores for the predicted EGFR peptide models outputted by TrRosetta. Confidence scores shown are the predicted Local Distance Difference Test (IDDT) values for the protein as determined by DeepAccNet(Hiranuma *et al.*, 2021). **(E)** Predicted peptide TM-Score as a function of the secondary structure of the full length protein. Peptides were binned into groups based on their overlap (>3 amino acids minimum) with structural elements on the full length protein. **(F)** Shown are 2 representative low similarity folders (TM-Scores .305, and .347 respectively) derived from RAF1. Secondary structure of the full length protein is largely retained, however the orientation of secondary structural elements is different from the full length WT. **(Related to STAR Methods)**

SUPPLEMENTAL TABLE 5. Primers (Related to Figure 1-2, STAR methods)

Name	Description	Sequence
PEP_01	Used to amplify initial oligo pool and individually synthesized cancer driver gene fragments for cloning. Additionally used for qPCR of overexpressed peptides.	GGCTAGGTAAGCTTGATA TCGGCCACCATG
PEP_02	Used to amplify initial oligo pool and individually synthesized cancer driver gene fragments for cloning. Additionally used for qPCR of overexpressed peptides.	GGCGGCACTGTTTAACAA GCCCCGTCAGTAG
PEP_03	Used to amplify cancer driver gene fragments for high throughput sequencing.	ACACTCTTTCCCTACACGA CGCTCTTCCGATCTGCTT GATATCGGCCACCATG
PEP_04	Used to amplify cancer driver gene fragments for high throughput sequencing.	GACTGGAGTTCAGACGTG TGCTCTTCCGATCTCACT GTTTAACAAGCCCCGTCAG TAG
GAPDH _F	Used for qPCR of overexpressed peptides.	ACAGTCAGCCGCATCTTC TT
GAPDH _R	Used for qPCR of overexpressed peptides.	ACGACCAAATCCGTTGAC TC
EF1a_s eq	Used for Sanger sequencing of constructs cloned into peptide expression vectors.	TTCTCAAGCCTCAGACAG TGG

SUPPLEMENTAL TABLE 6. Peptides Validated via Lentiviral Overexpression (Related to Figure 3)

Gene Name	Amino Acid Sequence
BRAF-379	MIDDLIRDQGFRGDGGSTTGLSATPPASLPGSLTNVKALQK
BRAF-380	MDDLIRDQGFRGDGGSTTGLSATPPASLPGSLTNVKALQKS
EGFR-697	MEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGE
EGFR-704	MLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVA
FBXW7-461	MTSTVRCMHLHEKRVSRSRDLRVWDIETGQCLHVLGMGH
FBXW7-512	MRRVVSAGYDFMVKVWDPETETCLHTLQGHTNRVYSLQFDG
RAF1-73	MRNGMSLHDCLMKALKVRGLQPECCAVFRLLEHKGGKKARL
RAF1-78	MLHDCLMKALKVRGLQPECCAVFRLLEHKGGKKARLDWNTD
KRAS61K-24	MIQNHVDEYDPTIEDSYRKQVVIDGETCLLDILDITAGKEE
KRAS61K-28	MFVDEYDPTIEDSYRKQVVIDGETCLLDILDITAGKEEYSAM
KRAS61K-34	MPTIEDSYRKQVVIDGETCLLDILDITAGKEEYSAMRDQYMR
DICER1-552	MRARAPISNYIMLADTDKIKSFEEDLKTYKAIEKILRNKCS
KRAS-143	METSAKTRQGVDDAFYTLVREIRKHKEKMSKDGGKKKKKSK
MDM2-25	METLVRPKPLLLKLLKSVGAQKDTYTMKEVLFYLGQYIMTK
RASA1-468	MKDAFYKNIVKKGYYLLKKGKGRWKNLYFILEGSDAQLIYF
AKT1-115	MEEEMDFRSGSPSDNSGAEEMEVS LAKPKHRVTMNEFEYLK
CCND1-167	MKMPEAEENKIIRKHAQTFVALCATDVKFISNPPSMVAAG
NOTCH1-626	MLCFCLKGTTGPNCEINLDDCASSPCDSGTCLDKIDGYECA

SUPPLEMENTAL TABLE 7. Crystal Structures (Related to STAR Methods)

Protein	PDB Crystal Structure ID	DOI
EGFR	5JEB, 1M14, 1XKK, 3QWQ	10.1038/nchembio.2171, 10.1074/jbc.M207135200, 10.1158/0008-5472.CAN-04- 1168, 10.1016/j.str.2011.11.016
RB1	2QDJ	10.1016/j.molcel.2007.08.023
RAF1	1GUA, 7JHP, 70MV	10.1038/nsb0896-723, 10.2210/pdb7JHP/pdb, 10.1038/nature08833

SUPPLEMENTAL TABLE 8. Chemically Synthesized Peptides (Related to Figure 5)

Name	Amino Acid Sequence
TAT-EGFR-697	GRKKRRQRRRPPQGSGSGSMEAPNQALLRILKETEFKKIKVLGSGAFG TVYKGLWIPEGE
TAT-RAF1-73	GRKKRRQRRRPPQGSGSGSMRNGMSLHDCLMKALKVRGLQPECCA VFRLLEHKGKKARL
TAT-FLAG	GRKKRRQRRRPPQGSGSGSDYKDHDGDYKDHDIDYKDDDDK
TAT-RASA1-468	GRKKRRQRRRPPQGSGSGSMKDAFYKNIVKKGILLKKGKGRWKNL YFILEGSDAQLIYF
TAT-MDM2-25	GRKKRRQRRRPPQGSGSGSMETLVRPKPLLLKLLKSVGAQKDTYTMK EVLFYLGQYIMTK

SUPPLEMENTAL TABLE 9. Overall Cost Analysis (Related to STAR Methods)

Item	Vendor	Price (\$)
Oligonucleotide Synthesis (~12,000 constructs)	Custom Array	2400
Cell Culture Media (DMEM + FBS +Trypsin)	Thermo Fisher (10566016,16140071,25200056)	215
Genomic DNA Isolation Kit (8 columns)	Qiagen (69504)	28
Polymerase for Sequencing Library Construction (1mL)	Kapa HiFi HotStart Ready Mix (Roche KK2602)	112
AMPure XP Beads (1mL)	Beckman (A63881)	20
PE100 Sequencing (2 time points, 2 replicates per time point)	Core Facility (NovaSeq S4, 50,000,000 reads)	140
Total	NA	2915