Title

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

Journal

Cell Systems, 12(7)

ISSN

2405-4712

Authors

Ford, Kyle M
Panwala, Rebecca
Chen, Dai-Hua
et al.

Publication Date

2021-07-01

DOI

10.1016/j.cels.2021.05.002

Peer reviewed

Published in final edited form as:

Cell Syst. 2021 July 21; 12(7): 716–732.e7. doi:10.1016/j.cels.2021.05.002.

Peptide Tiling Screens of Cancer Drivers Reveal Oncogenic Protein Domains and Associated Peptide Inhibitors

Kyle M. Ford[a], **Rebecca Panwala**[a], **Dai-Hua Chen**[a], **Andrew Portell**[a], **Nathan Palmer**[b], **Prashant Mali**[a,†]

[a]Department of Bioengineering, University of California San Diego, San Diego, California, 92093 USA.

[b]Division of Biological Sciences, University of California San Diego, San Diego, California, 92093 USA.

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. Towards 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 27–63μM. We anticipate this peptide tiling (PepTile) approach will enable rapid de novo mapping of bioactive protein domains and associated interfering peptides.

Graphical Abstract

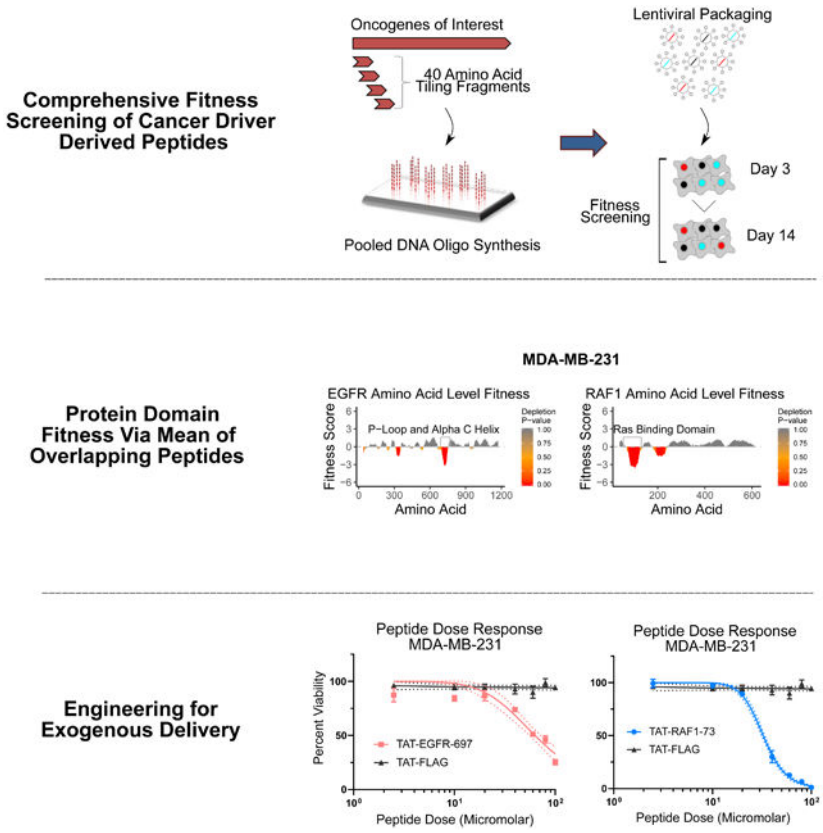
[†]Correspondence: pmali@ucsd.edu.

AUTHOR CONTRIBUTIONS

K.M.F. and P.M. conceived and planned the study, K.M.F., R.P., A.P. and D.C. performed experiments, K.M.F, N.P. and D.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.



ETOC Blurb:

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 that upon engineering for exogenous delivery enabled drug-like function with anti-cancer activity.

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, Elledge and Davis, 1992; Akada, Yamamoto and Yamashita, 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, Queitsch and Fields, 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, Elledge and Davis, 1992; Akada, Yamamoto and Yamashita, 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, Queitsch and Fields, 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 40mer peptide for each target protein. This strategy allows for modular library design, complete coverage of protein-protein interaction interfaces, and is supported by extensive previous research showing fragmented or truncated proteins can function as inhibitors of the full length protein(Herskowitz, 1987; Ramer, Elledge and Davis, 1992; Akada, Yamamoto and Yamashita, 1997; Barnard *et al.*, 1998; Soucek *et al.*, 2002; Boyer *et al.*, 2004; Zhu, Lu and Zhu, 2016; Nim *et al.*, 2016; Bai *et al.*, 2017; Yu *et al.*, 2017; Dorrity, Queitsch and Fields, 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, 2017; Ford, McDonald and Mali, 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 (Figure 1, Supplemental Figure 1–2, Supplemental Table 1). 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, Nissley and McCormick, 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 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 40mer peptide within the proteins primary structure. Testing every overlapping 40mer improves statistical power and allows for sensitive discrimination of similar peptide motifs, 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, Bellon and Nicot, 2018). In addition to gene fragments derived from the wildtype RAS and MYC proteins, we also included fragments derived from pathogenic Ras variants which 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, Singh and Garcia-Diaz, 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 6234 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 EF1a promoter driving gene fragment transcription (Figure 1A, Supplemental Figure 1A, Supplemental Figure 2, Supplemental Table 1, and 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 which were used to transduce the MYC and RAS dependent Hs578T and MDA-MB-231 cell lines in duplicate (Figure 1A and STAR Methods)(Eckert *et al.*, 2004; Kang *et al.*, 2014; Depmap Broad, 2019). Genomic

DNA was isolated 3 days after transduction, as well as 14 days after transduction to calculate peptide specific $\log_2$ fold changes. These peptide specific $\log_2$ fold change values were then used to calculate an amino acid level fitness score (Dorrity, Queitsch and Fields, 2019) via the mean of all fragments which overlap a particular codon (Figure 1A,B Supplemental Figure 1B–E, Supplemental Table 2 and STAR Methods). The amino acid level fitness score was first calculated by taking the mean $\log_2$ fold change of all overlapping peptides. For every residue in the 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 peptides derived from these positions were collectively more deleterious to cell fitness than a random sampling of peptides from the library (Supplemental Figure 1D–E). There was good correlation between biological replicates, with the Hs578T and MDA-MB-231 amino acid scores having a Pearson correlation of .54 and .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 towards 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 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 to the RBD (micromolar vs nanomolar affinity)(Williams *et al.*, 2000).

FBXW7 had a broad region of depletion corresponding to WD repeats 1–6. Knock-out screening via CRISPR-Cas9 has shown 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(Depmap Broad, 2019). The WD repeats in FBXW7 mediate substrate binding and subsequent recruitment to the E3 ubiquitin-protein ligase complex, suggesting 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).

Towards 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 wildtype counterparts (Figure 1C). The 40mer peptides derived from KRAS Q61K were significantly depleted across both cell lines, while wild type peptides overlapping amino acid showed no effect on cell fitness. The full length Q61K mutant is highly transforming due to a modified Ras/Raf interaction which may play a role in the anti-proliferative activity of the Q61K derived fragments(Der, Finkel and Cooper, 1986; Buhrman, Wink and Mattos, 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 which 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 (Figure 2A, Supplemental Figure 3A, Supplemental Table 1,3) 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 to 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 (Figure 2B, Supplemental Figure 3B, Supplemental Table 1,4). 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 mutant peptides had a similar effect on cell fitness compared to their wildtype 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 (Figure 2E–G, Supplemental Figure 4A–B). We first examined the pattern of depletion for the transcription factor NFE2L2, the protein containing the most deleterious domain as scored by this screen (Supplemental Figure 4C, Figure 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 to WT peptides overlapping the same residues, some mutants such as PIK3CA956F, KRAS61K, and BRAF594N showed markedly more deleterious effects on cell fitness (Figure 2G, Supplemental Figure 4E).

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, Garimella and Lipkowitz, 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, Yap and Meric-Bernstam, 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 Ras binding domain 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 (Supplemental Figure 4D). 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. This data supports 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 (Supplemental Figure 4D), 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 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 (Figure 3A, Supplemental Figure 5A–B, and Supplemental Table 6). All 12 peptides tested had significant growth defects when assayed in Hs578T and/or MDA-MB-231 cells compared to infection with the GFP control plasmid. EGFR-697 specifically was extremely deleterious 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 (Supplemental Figure 5A–B). To test the specificity of these perturbations, we transduced MCF-7 cells with RAF1-73 and EGFR-697. MCF-7 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; Depmap Broad, 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 this screening methodology identifies context dependent inhibitors of cellular proliferation rather than generally toxic peptide motifs.

For individual peptides which 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 level 72 hours after transduction, indicating the EF1α promoter can drive robust expression of small peptides. Assuming the peptides are translated from their mRNA at a similar rate as GAPDH is (GAPDH has a cellular concentration of approximately .4μM), it can be estimated from the qPCR data that peptide molar concentrations in MDA-MB-231 cells range from .15–6.5 μM depending on the construct(Lazarev, Guzhova and Margulis, 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 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 (Supplemental Figure 5C). While the average length of a protein domain is predicted to be 100 amino acids(Lin and Zewail, 2012), we hypothesized based on the modular conformation of long proteins and prior work focused on dominant negatives that 40mer 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 to the GFP control, the parent peptides consistently caused slower cell growth than the shorter versions (Supplemental Figure 5D).

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 which 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, Hura and Rubin, 2012). Consistent with this hypothesis, it has been previously shown the addition of N-terminal domains of RB1 is sufficient to halt DNA replication in xenopus egg extracts(Borysov, Nepon-Sixt and Alexandrow, 2016). 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, Supplemental Table 7). This result highlights how higher order protein level features can inform observed peptide fitness, and gives 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; Depmap Broad, 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 (P=.79). This result stems from the fact that 1: not every gene which 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 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 >.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=.279; P=.045) between the predicted mutant TP53 functionality from the DMS data to the mutant TP53 peptide fitness. This comparison to DMS data indicates 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, Supplemental Table 8). The IC50s of

this peptide was 33.3μM for Hs578T and 63μM for MDA-MB-231. Moreover, RAF1-73 was also highly deleterious to cell growth, with IC50 values of 27.0μM and 32.6μM for Hs578T and MDA-MB-231 respectively. These IC50 values are comparable to the mean IC50 of all drugs tested on these cell lines in the Sanger Genomics of Drug Sensitivity Database (48.6μM for Hs578T and 54.0μM 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 23μM and 33μM respectively in MDA-MB-231 cells (Figure 5B). This result demonstrates how the high throughput nature of the PepTile screening strategy can identify diverse bioactive peptides which 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, this data further confirms the peptides identified in this screen are acting at the protein level and suggests that further engineering of these compounds could yield translationally relevant biopharmaceuticals.

As peptide constructs will likely require additional engineering to maximize efficacy towards 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 (Supplemental Figure 6, STAR Methods). Because this peptide production method (as well as the PepTile fitness screening strategy – See Supplemental Table 9) 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 Ras binding domain) 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.

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 (Supplemental Table 10). We identified 225 differentially expressed genes (BH adjusted P-value <.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 < .025) upregulation/downregulation in cells expressing EGFR-697 compared to control cells transduced with GFP (Figure 6B). Several metabolic pathways relating to oxidative phosphorylation and carbon metabolism were downregulated, consistent with the role of oncogenic EGFR signaling as a driver of metabolic alterations(Borlak, Singh and Gazzana, 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 of 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=.008 and FDR=.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 (Figure 6C–D). Both the genetically overexpressed and the exogenously delivered versions of EGFR-697 showed greater activity in cell lines with detectable EGFR expression. This data was benchmarked against a high dose of Erlotinib, showing a similar EGFR expression dependent change in sensitivity. Collectively, this data supports 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 (Supplemental Figure 7). 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 (Supplemental Figure 7A). 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 .63 and .67 respectively). A TM-Score greater than .5 corresponds to a P-value less than 5.5×10^{-7}, and is a widely used criterion for if 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 (Supplemental Figure 7B). 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 (Supplemental Figure 7C–D). However, a small subset of derived peptides modeled had structures diverging from that of the full length protein (minimum TM-score observed = .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 (Supplemental Figure 7E). Peptides derived from regions with alpha helixes, beta sheets, or both were largely predicted to fold into structures resembling the full length protein (mean TM-scores of .74, .82, .76 respectively). When examining the predicted structures least similar to the full length protein (TM-scores<.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 (Supplemental Figure 7F). 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 which 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 could

generate synthetic peptides better suited for inhibiting this type of interaction. Additionally, current DNA synthesis technology limits array synthesized DNA libraries to less than ~350bp (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.

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 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(‘Mechanism matters’, 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, Brambilla and Rossi, 2017); or alternatively, 2) coupled to chemical moieties such as polyethylene 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 which 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

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 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.

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 Supplemental Figure 2–3 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 Supplemental Table 5 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 H_2O 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, 1µl 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 hours 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 36µl of lipofectamine 2000, 3 µg of pMD2.G plasmid (Addgene #12259), 12 µg of pCMV deltaR8.2 plasmid (Addgene #12263), and 9 µg 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 8µg/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. 2µg/ml puromycin was used to select the Hs578T cells, and 3.5µg/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 2.5µl of primers PEP_3 and PEP_4 (10µM) per 50µl reaction. For each sample (i.e. time point and replicate) from the pilot library, 10 separate 50µl PCR reactions with 4µg of gDNA each (40µg 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 300μg of total gDNA per timepoint and replicate. The larger libraries were then sequenced with 100bp 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 (**F**ast **L**ength **A**djustment of **SH**ort 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, Rondón-Villarreal and Torres, 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 $\log_2$ 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 $\log_2$ fold change value (x) and Z-normalizing to the library wide amino acid $\log_2$ fold change standard deviation (σ) and mean (μ).

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

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:

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

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; Zeileis *et al.*, 2012; 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. 2μg/ml puromycin was used to select Hs578T and MCF-7 cells, while 3.5μg/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 Figure 6D and Supplemental Figure 5C–D, relative cell viability was determined via Crystal Violet staining. At the experimental endpoint cells were washed once with PBS, and subsequently incubated in 50μl 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 $(GS)_3$ linker sequence (Supplemental Table 8) 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 1μg of each indicated plasmid per well with 5μl 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 400µl 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 300µl of the lysate to 20µl of packed anti FLAG agarose beads (Millipore Sigma A2220) prewashed with TBS. The remaining 100µl 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 20µl 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 .2µm 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 µl cDNA, 400 nM of each primer (See Supplemental Table 5), 2X iTaq Universal SYBR Green Supermix (BioRad), with ultra pure water up to 20 µl. 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 housekeeping 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, Huber and Anders, 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, Chivian and Baker, 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 Supplemental Table 7). 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 (lDDT) as determined by DeepAccNet(Hiranuma *et al.*, 2021). Validated lDDT 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, Cherry and Waugh, 2009). Recombinant MBP fusions and TEV protease were cloned into the pET Champion vector (Thermo K630203) and expressed in T7 express E. coli (NEB C2566I). 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 .8, and then induced with .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, .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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

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, U54CA209891), and an NSF Graduate Research Fellowship (DGE-1650112) to K.M.F.

REFERENCES

Adhikari H and Counter CM (2018) 'Interrogating the protein interactomes of RAS isoforms identifies PIP5K1A as a KRAS-specific vulnerability', Nature Communications. doi: 10.1038/s41467-018-05692-6.

Akada R, Yamamoto J and Yamashita I (1997) 'Screening and identification of yeast sequences that cause growth inhibition when overexpressed', Molecular and General Genetics. doi: 10.1007/s004380050415.

Azeloglu EU and Iyengar R (2015) 'Signaling networks: Information flow, computation, and decision making', Cold Spring Harbor Perspectives in Biology. doi: 10.1101/cshperspect.a005934.

Bai Z et al. (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', Journal of Chemical Information and Modeling. doi: 10.1021/acs.jcim.6b00673.

Bailey MH et al. (2018) 'Comprehensive Characterization of Cancer Driver Genes and Mutations', Cell. doi: 10.1016/j.cell.2018.02.060.

Barnard D et al. (1998) 'In vitro inhibition of Ras-Raf association by short peptides', Biochemical and Biophysical Research Communications. doi: 10.1006/bbrc.1998.8746.

Beaulieu ME et al. (2019) 'Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy', Science Translational Medicine. doi: 10.1126/scitranslmed.aar5012.

Becker CFW et al. (2003) 'Total chemical synthesis of a functional interacting protein pair: The protooncogene H-Ras and the Ras-binding domain of its effector c-Raf1', Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.0831227100.

Benjamini Y and Hochberg Y (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', Journal of the Royal Statistical Society: Series B (Methodological). doi: 10.1111/j.2517-6161.1995.tb02031.x.

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. doi: 10.1186/s12864-015-1312-z.

Borysov SI, Nepon-Sixt BS and Alexandrow MG (2016) 'The N Terminus of the Retinoblastoma Protein Inhibits DNA Replication via a Bipartite Mechanism Disrupted in Partially Penetrant Retinoblastomas', Molecular and Cellular Biology. doi: 10.1128/mcb.00636-15.

Bose R and Zhang X (2009) 'The ErbB kinase domain: Structural perspectives into kinase activation and inhibition', Experimental Cell Research. doi: 10.1016/j.yexcr.2008.07.031.

Boyer J et al. (2004) 'Large-scale exploration of growth inhibition caused by overexpression of genomic fragments in Saccharomyces cerevisiae.', Genome biology.

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. doi: 10.1016/j.str.2007.10.011.

Burke JR, Hura GL and Rubin SM (2012) 'Structures of inactive retinoblastoma protein reveal multiple mechanisms for cell cycle control', Genes and Development. doi: 10.1101/gad.189837.112.

Chang YS et al. (2013) 'Stapled α-helical peptide drug development: A potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.1303002110.

Chavez KJ, Garimella SV and Lipkowitz S (2010) 'Triple negative breast cancer cell lines: One tool in the search for better treatment of triple negative breast cancer', Breast Disease. doi: 10.3233/BD-2010-0307.

Chen J et al. (2020) 'Pervasive functional translation of noncanonical human open reading frames', Science (New York, N.Y.) doi: 10.1126/science.aay0262.

Clark GJ et al. (1996) 'Peptides containing a consensus Ras binding sequence from Raf-1 and the GTPase activating protein NF1 inhibit Ras function', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.93.4.1577.

Cox AD et al. (2014) 'Drugging the undruggable RAS: Mission Possible?', Nature Reviews Drug Discovery. doi: 10.1038/nrd4389.

Craik DJ et al. (2013) 'The Future of Peptide-based Drugs', Chemical Biology and Drug Design. doi: 10.1111/cbdd.12055.

Cuadrado A et al. (2019) 'Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases', Nature Reviews Drug Discovery. doi: 10.1038/s41573-018-0008-x.

Dang CV (2012) 'MYC on the path to cancer', Cell. doi: 10.1016/j.cell.2012.03.003.

Depmap Broad (2019) DepMap Achilles 19Q1 Public., figshare. Fileset.

Der CJ, Finkel T and Cooper GM (1986) 'Biological and biochemical properties of human rasH genes mutated at codon 61', Cell. doi: 10.1016/0092-8674(86)90495-2.

Dobin A et al. (2013) 'STAR: Ultrafast universal RNA-seq aligner', Bioinformatics. doi: 10.1093/bioinformatics/bts635.

Doench JG (2017) 'Am I ready for CRISPR? A user's guide to genetic screens', Nature Reviews Genetics. Nature Publishing Group, 19(2), pp. 67–80. doi: 10.1038/nrg.2017.97.

Donsky E and Wolfson HJ (2011) 'PepCrawler: A fast RRT-based algorithm for high-resolution refinement and binding affinity estimation of peptide inhibitors', Bioinformatics. doi: 10.1093/bioinformatics/btr498.

Dorrity MW, Queitsch C and Fields S (2019) 'High-throughput identification of dominant negative polypeptides in yeast', Nature Methods. doi: 10.1038/s41592-019-0368-0.

Downward J (2003) 'Targeting RAS signalling pathways in cancer therapy', Nature Reviews Cancer. doi: 10.1038/nrc969.

Eckert LB et al. (2004) 'Involvement of ras activation in human breast cancer cell signaling, invasion, and anoikis', Cancer Research. doi: 10.1158/0008-5472.CAN-04-0396.

Eisenhardt AE et al. (2016) 'Phospho-proteomic analyses of B-Raf protein complexes reveal new regulatory principles', Oncotarget. doi: 10.18632/oncotarget.8427.

Ford K, McDonald D and Mali P (2019) 'Functional Genomics via CRISPR–Cas', Journal of Molecular Biology. doi: 10.1016/j.jmb.2018.06.034.

Frishman D and Argos P (1995) 'STRIDE: Protein secondary structure assignment from atomic coordinates', Proteins Structure Function and Genetics.

Gasperini M et al. (2017) 'CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions', American Journal of Human Genetics. doi: 10.1016/j.ajhg.2017.06.010.

Guidotti G, Brambilla L and Rossi D (2017) 'Cell-Penetrating Peptides: From Basic Research to Clinics', Trends in Pharmacological Sciences. doi: 10.1016/j.tips.2017.01.003.

Gurtan AM et al. (2012) 'In vivo structure-function analysis of human Dicer reveals directional processing of precursor miRNAs', RNA. doi: 10.1261/rna.032680.112.

Han Y and Král P (2020) 'Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2', ACS Nano. doi: 10.1021/acsnano.0c02857.

Hao B et al. (2007) 'Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases', Molecular Cell. doi: 10.1016/j.molcel.2007.02.022.

Herskowitz I (1987) 'Functional inactivation of genes by dominant negative mutations', Nature. doi: 10.1038/329219a0.

Hiranuma N et al. (2021) 'Improved protein structure refinement guided by deep learning based accuracy estimation', Nature Communications. doi: 10.1038/s41467-021-21511-x.

Hui L et al. (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. doi: 10.1038/sj.onc.1209735.

Iwakuma T and Lozano G (2003) 'MDM2, An Introduction', Molecular Cancer Research.

Janku F, Yap TA and Meric-Bernstam F (2018) 'Targeting the PI3K pathway in cancer: Are we making headway?', Nature Reviews Clinical Oncology. doi: 10.1038/nrclinonc.2018.28.

Kang J et al. (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. doi: 10.1186/1471-2407-14-32.

Kim DE, Chivian D and Baker D (2004) 'Protein structure prediction and analysis using the Robetta server', Nucleic Acids Research. doi: 10.1093/nar/gkh468.

Kobayashi S et al. (2006) 'Transcriptional profiling identifies cyclin D1 as a critical downstream effector of mutant epidermal growth factor receptor signaling', Cancer Research. doi: 10.1158/0008-5472.CAN-06-2318.

Kosuri S and Church GM (2014) 'Large-scale de novo DNA synthesis: Technologies and applications', Nature Methods, 11(5), pp. 499–507. doi: 10.1038/nmeth.2918. [PubMed: 24781323]

Kotler E et al. (2018) 'A Systematic p53 Mutation Library Links Differential Functional Impact to Cancer Mutation Pattern and Evolutionary Conservation', Molecular Cell. doi: 10.1016/j.molcel.2018.06.012.

Kubbutat MHG et al. (1999) 'Analysis of the degradation function of Mdm2', Cell Growth and Differentiation.

Kumar A et al. (2008) 'Structure and clinical relevance of the epidermal growth factor receptor in human cancer', Journal of Clinical Oncology. doi: 10.1200/JCO.2007.12.1178.

Langmead B and Salzberg SL (2012) 'Fast gapped-read alignment with Bowtie 2', Nature Methods. doi: 10.1038/nmeth.1923.

Lanning NJ et al. (2017) 'Metabolic profiling of triple-negative breast cancer cells reveals metabolic vulnerabilities', Cancer & Metabolism. doi: 10.1186/s40170-017-0168-x.

Lazarev VF, Guzhova IV and Margulis BA (2020) 'Glyceraldehyde-3-phosphate dehydrogenase is a multifaceted therapeutic target', Pharmaceutics. doi: 10.3390/pharmaceutics12050416.

Li H et al. (2009) 'The Sequence Alignment/Map format and SAMtools', Bioinformatics. doi: 10.1093/bioinformatics/btp352.

Li W et al. (2014) 'MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens', Genome Biology, 15(12), p. 554. doi: 10.1186/s13059-014-0554-4. [PubMed: 25476604]

Li X et al. (2015) 'AMPK-mediated energy homeostasis and associated metabolic effects on cancer cell response and resistance to cetuximab', Oncotarget. doi: 10.18632/oncotarget.3432.

Lin MM and Zewail AH (2012) 'Hydrophobic forces and the length limit of foldable protein domains', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.1207382109.

Liu M et al. (2010) 'D-peptide inhibitors of the p53-MDM2 interaction for targeted molecular therapy of malignant neoplasms', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.1008930107.

London N et al. (2010) 'Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions?', Proteins: Structure, Function and Bioinformatics. doi: 10.1002/prot.22785.

Love MI, Huber W and Anders S (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', Genome Biology. doi: 10.1186/s13059-014-0550-8.

Magoč T and Salzberg SL (2011) 'FLASH: Fast length adjustment of short reads to improve genome assemblies', Bioinformatics. doi: 10.1093/bioinformatics/btr507.

Mariani V et al. (2013) 'IDDT: A local superposition-free score for comparing protein structures and models using distance difference tests', Bioinformatics. doi: 10.1093/bioinformatics/btt473.

'Mechanism matters' (2010) Nature Medicine. doi: 10.1038/nm0410-347.

Meyer MJ et al. (2018) 'Interactome INSIDER: a structural interactome browser for genomic studies', Nature Methods. doi: 10.1038/nmeth.4540.

Nassar N, Singh K and Garcia-Diaz M (2010) 'Structure of the dominant negative S17N mutant of Ras', Biochemistry. doi: 10.1021/bi9020742.

Nim S et al. (2016) 'Pooled screening for antiproliferative inhibitors of protein-protein interactions', Nature Chemical Biology. doi: 10.1038/nchembio.2026.

Osorio D, Rondón-Villarreal P and Torres R (2015) 'Peptides: A package for data mining of antimicrobial peptides', R Journal. doi: 10.32614/rj-2015-001.

Pagès H et al. (2017) Biostrings: Efficient manipulation of biological strings, R package version 2.46.0

Palmieri L and Rastelli G (2013) 'αc helix displacement as a general approach for allosteric modulation of protein kinases', Drug Discovery Today. doi: 10.1016/j.drudis.2012.11.009.

Ramer SW, Elledge SJ and Davis RW (1992) 'Dominant genetics using a yeast genomic library under the control of a strong inducible promoter', Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.89.23.11589.

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 ', Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1803152115.

Rupaimoole R and Slack FJ (2017) 'MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases', Nature Reviews Drug Discovery. doi: 10.1038/nrd.2016.246.

Santarius T et al. (2010) 'A census of amplified and overexpressed human cancer genes', Nature Reviews Cancer. doi: 10.1038/nrc2771.

Sato M et al. (2015) 'MYC is a critical target of FBXW7', Oncotarget. doi: 10.18632/oncotarget.3203.

Scheffzek K and Welti S (2012) 'Pleckstrin homology (PH) like domains - Versatile modules in protein-protein interaction platforms', FEBS Letters. doi: 10.1016/j.febslet.2012.06.006.

Schwarze SR et al. (1999) 'In vivo protein transduction: Delivery of a biologically active protein into the mouse', Science. doi: 10.1126/science.285.5433.1569.

Sergushichev AA (2016) 'An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation', bioRxiv. doi: 10.1101/060012.

Shalem O et al. (2015) 'High-throughput functional genomics using CRISPR-Cas9', 16(5), pp. 299–311. doi: 10.1038/nrg3899.High-throughput.

Simanshu DK, Nissley DV and McCormick F (2017) 'RAS Proteins and Their Regulators in Human Disease', Cell. doi: 10.1016/j.cell.2017.06.009.

Smoot ME et al. (2011) 'Cytoscape 2.8: New features for data integration and network visualization', Bioinformatics. doi: 10.1093/bioinformatics/btq675.

Soucek L et al. (2002) 'Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis', Cancer Research.

Strohl WR (2015) 'Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters', BioDrugs. doi: 10.1007/s40259-015-0133-6.

Tropea JE, Cherry S and Waugh DS (2009) 'Expression and purification of soluble His6-tagged TEV protease', Methods in Molecular Biology. doi: 10.1007/978-1-59745-196-3_19.

Uhlen M et al. (2017) 'A pathology atlas of the human cancer transcriptome', Science. doi: 10.1126/science.aan2507.

Wickham H (2011) 'ggplot2', Wiley Interdisciplinary Reviews: Computational Statistics. doi: 10.1002/wics.147.

Wickham H et al. (2019) 'Package "dplyr". A Grammar of Data Manipulation.', R package version 0.8.0.1

Williams JG et al. (2000) 'Elucidation of binding determinants and functional consequences of Ras/Raf-cysteine-rich domain interactions', Journal of Biological Chemistry. doi: 10.1074/jbc.M000397200.

Xu J and Zhang Y (2010) 'How significant is a protein structure similarity with TM-score = 0.5?', Bioinformatics. doi: 10.1093/bioinformatics/btq066.

Yang J et al. (2020) 'trRosetta', Proceedings of the National Academy of Sciences of the United States of America.

Yang W et al. (2013) 'Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells', Nucleic Acids Research. doi: 10.1093/nar/gks1111.

Yeh CH, Bellon M and Nicot C (2018) 'FBXW7: A critical tumor suppressor of human cancers', Molecular Cancer. doi: 10.1186/s12943-018-0857-2.

Yu Junmin et al. (2017) 'Structure-based rational design of self-inhibitory peptides to disrupt the intermolecular interaction between the troponin subunits C and I in neuropathic pain', Bioorganic Chemistry. doi: 10.1016/j.bioorg.2017.05.004.

Yun CH et al. (2007) 'Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity', Cancer Cell. doi: 10.1016/j.ccr.2006.12.017.

Zaidman D and Wolfson HJ (2016) 'PinaColada: Peptide-inhibitor ant colony ad-hoc design algorithm', Bioinformatics. doi: 10.1093/bioinformatics/btw133.

Zeileis A et al. (2012) 'Package " zoo "', Cran.

Zhang C et al. (2016) 'NRF2 promotes breast cancer cell proliferation and metastasis by increasing RhoA/ROCK pathway signal transduction', Oncotarget. doi: 10.18632/oncotarget.12435.

Zhang Y and Skolnick J (2004) 'Scoring function for automated assessment of protein structure template quality', Proteins: Structure, Function and Genetics. doi: 10.1002/prot.20264.

Zhao Y et al. (2015) 'Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 inhibitors) in clinical trials for cancer treatment', Journal of Medicinal Chemistry. doi: 10.1021/jm501092z.

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', Bioorganic Chemistry. doi: 10.1016/j.bioorg.2016.08.012.

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

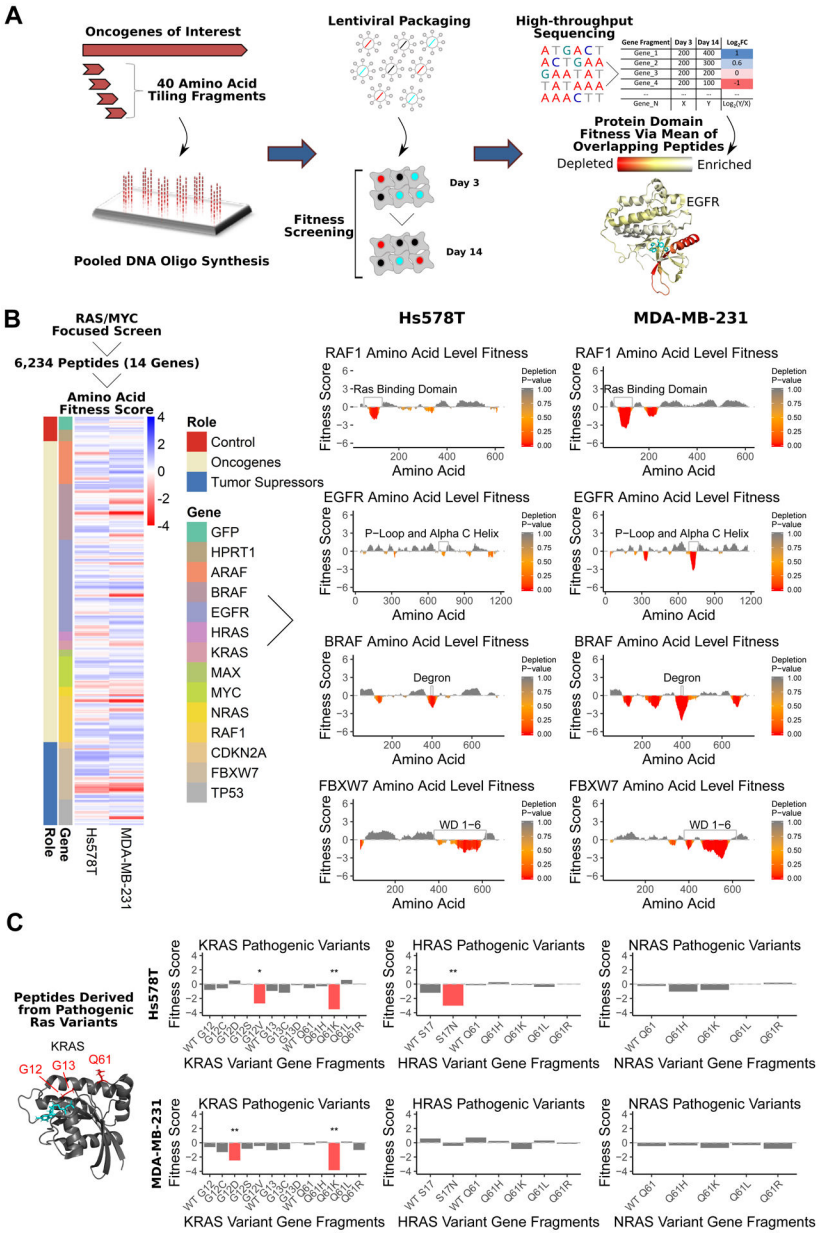


FIGURE 1. Peptide overexpression screening strategy and MAPK focused library:

(A) Design of overlapping peptide expression library. Gene fragments coding for all possible overlapping 40mer 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 MOI to ensure only one peptide was expressed per cell. The cells were then grown for two weeks, with genomic DNA extracted at day 3 and day 14. Next, peptide coding gene fragments were PCR amplified from genomic DNA and sequenced to track peptide abundances and calculate $\log_2$ 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 $\log_2$ 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<.05,**P<.01,***P<.001,****P<.0001). (See also Figure S1–2)

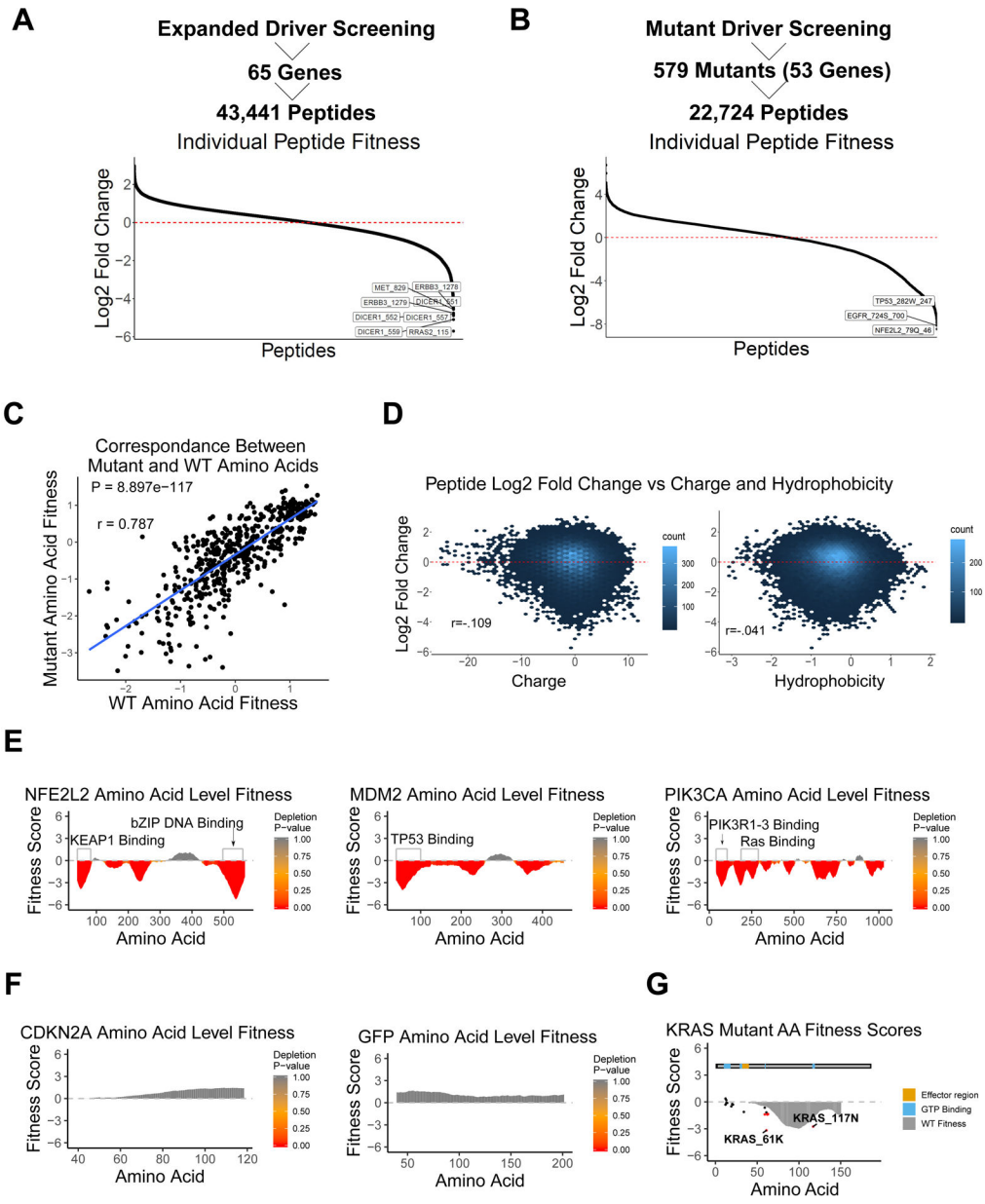


FIGURE 2. Expanded library screening enables more comprehensive evaluation of cancer driver derived peptides:

(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. **(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=.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 which were significantly depleted during the pooled screen (BH adjusted P<.05). (See also Figure S3–4)

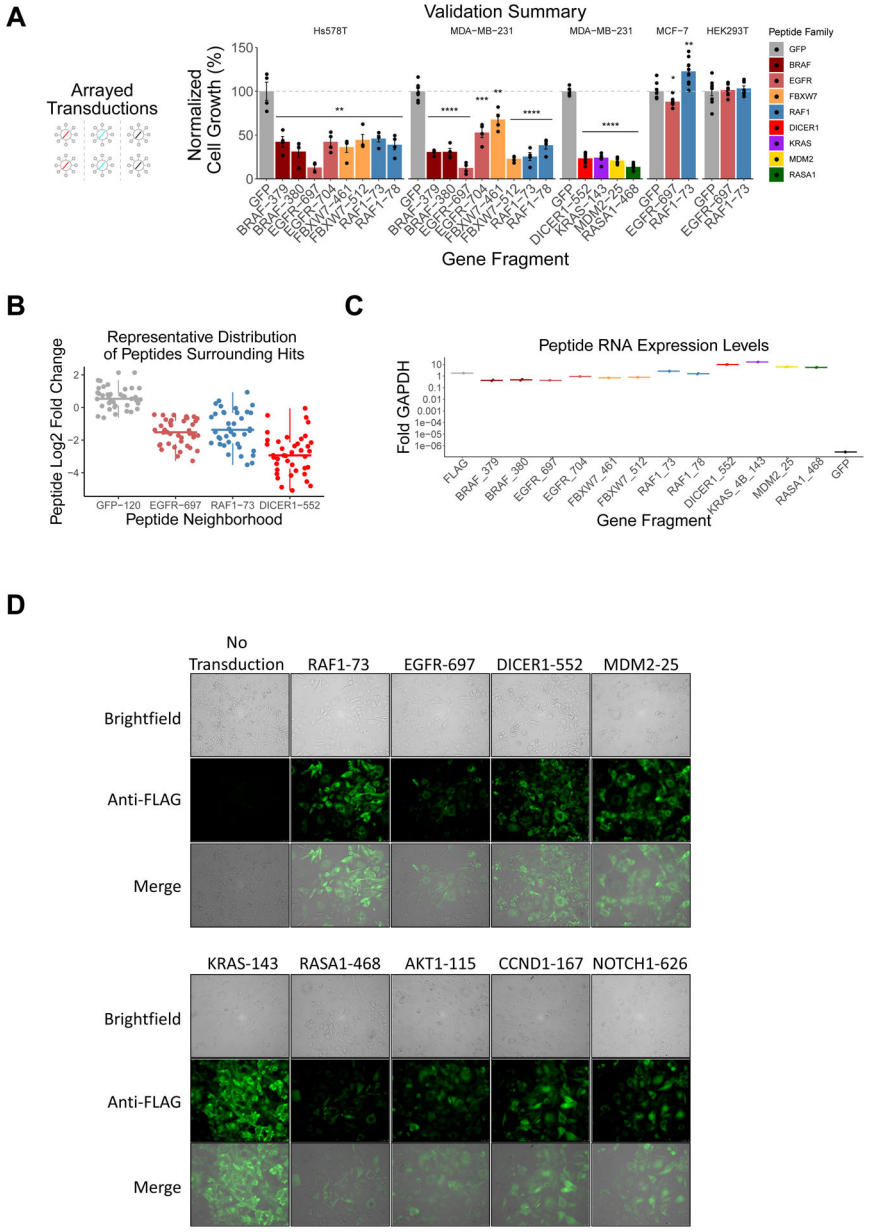


FIGURE 3. Validation of anti-proliferative peptide activity and expression

(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 lentiviral transduction. Bar plots indicate mean, with error bars representing standard error (*P<.05,**P<.01,***P<.001,****P<.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 hours. 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 hours 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)

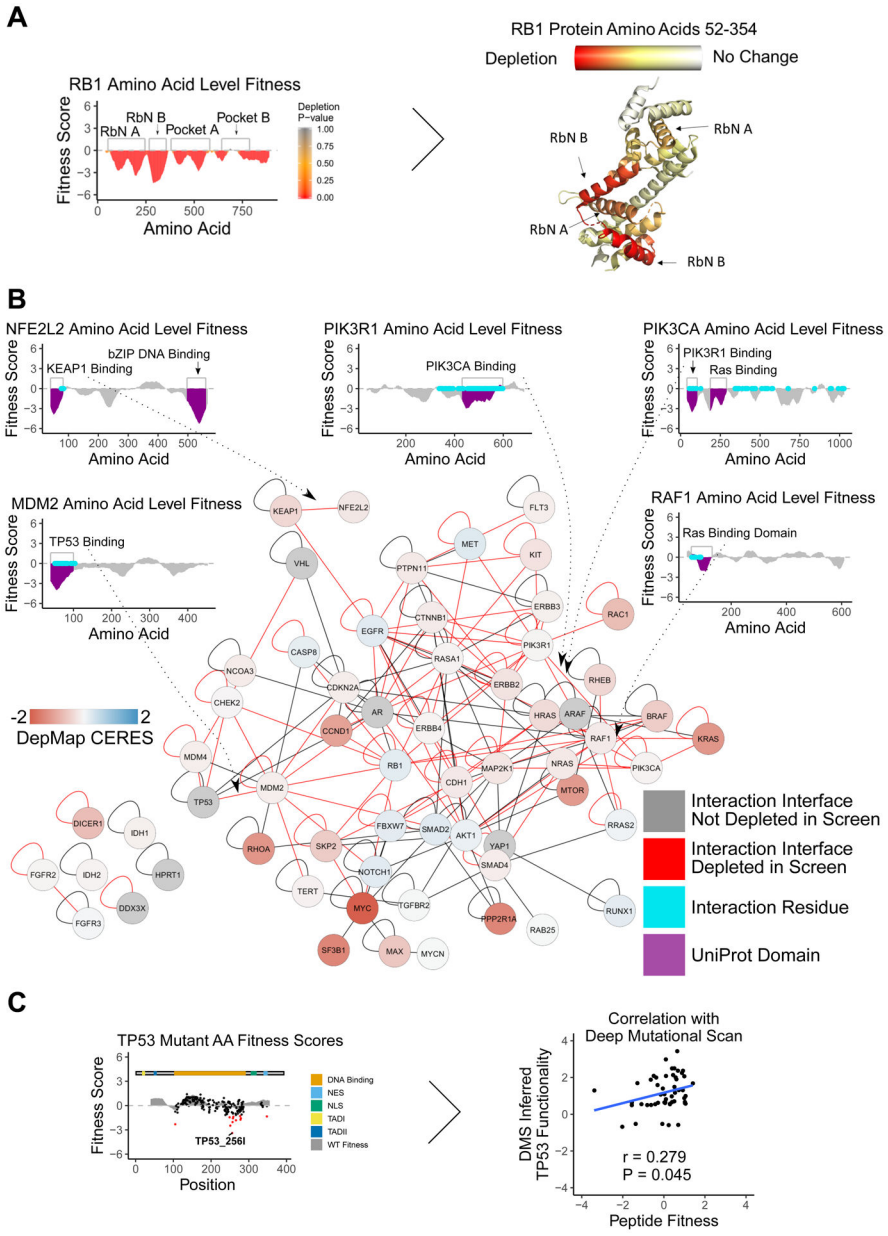


FIGURE 4. Anti-proliferative peptides derived from oncogenic interaction interfaces:
(A) RB1 per position fitness scores mapped onto the RB1 N terminal crystal structure. Regions of relatively high and low depletion appear to correspond to transitions between specific alpha helix's in the RB1 structure, 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 is 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 knockout. Dark gray nodes indicate genes for which high confidence CRISPR based fitness data was 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 deep mutational scan data in a TP53 null cell line (Kotler *et al.*, 2018). After filtering out TP53 mutants with little effect on cell fitness in the deep mutational scan (absolute value of fitness scores <.5), inferred TP53 functionality is significantly correlated with mutant peptide derived fitness (Pearson, P=.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 S3–4)

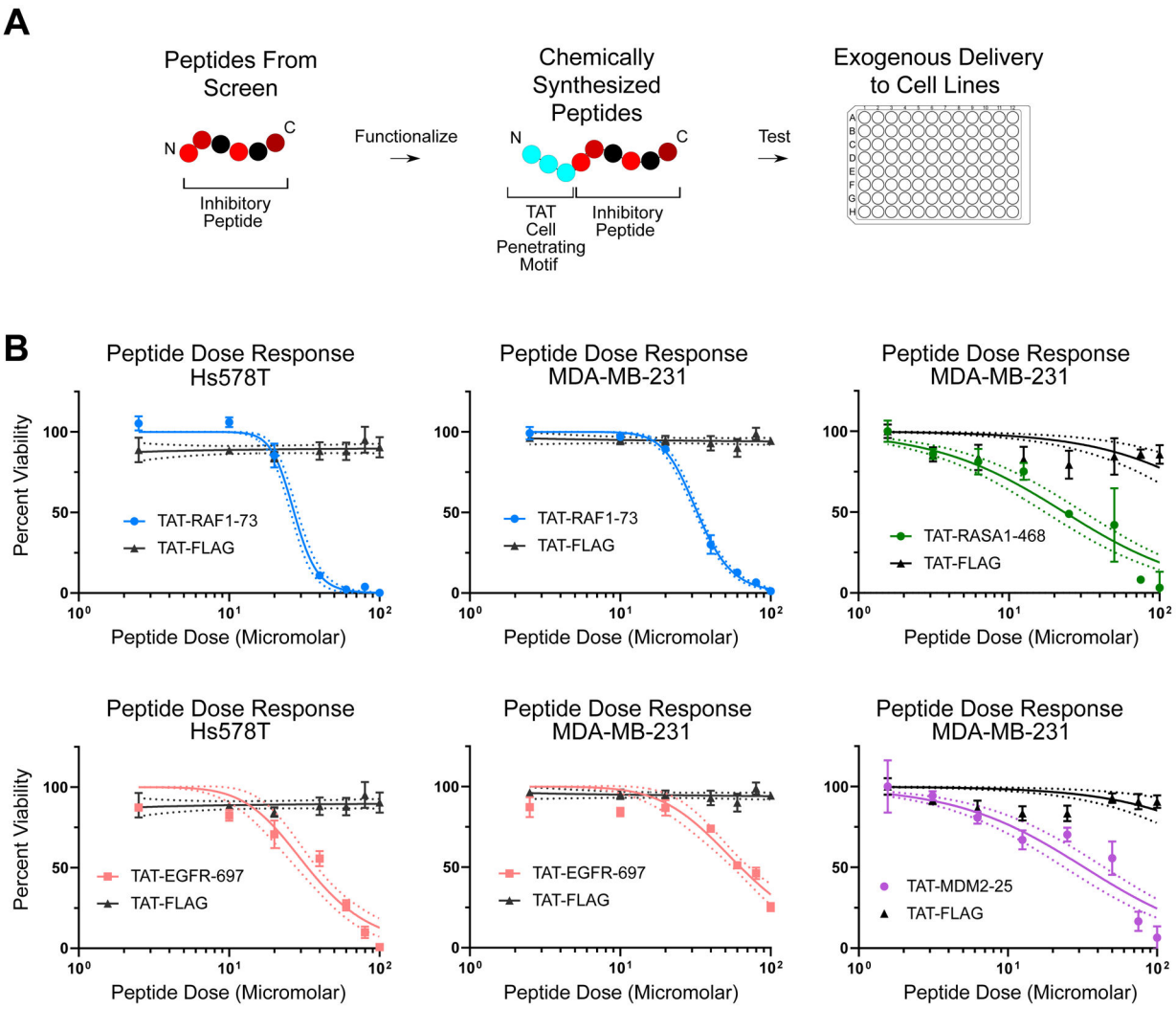


FIGURE 5. Cancer driver derived peptides have protein level activity and potential drug-like function:

(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 0–100μM. A 3x FLAG peptide conjugated to TAT served as the negative control. Cell viability was measured 24 hours later by the WST-8 assay, indicating 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)

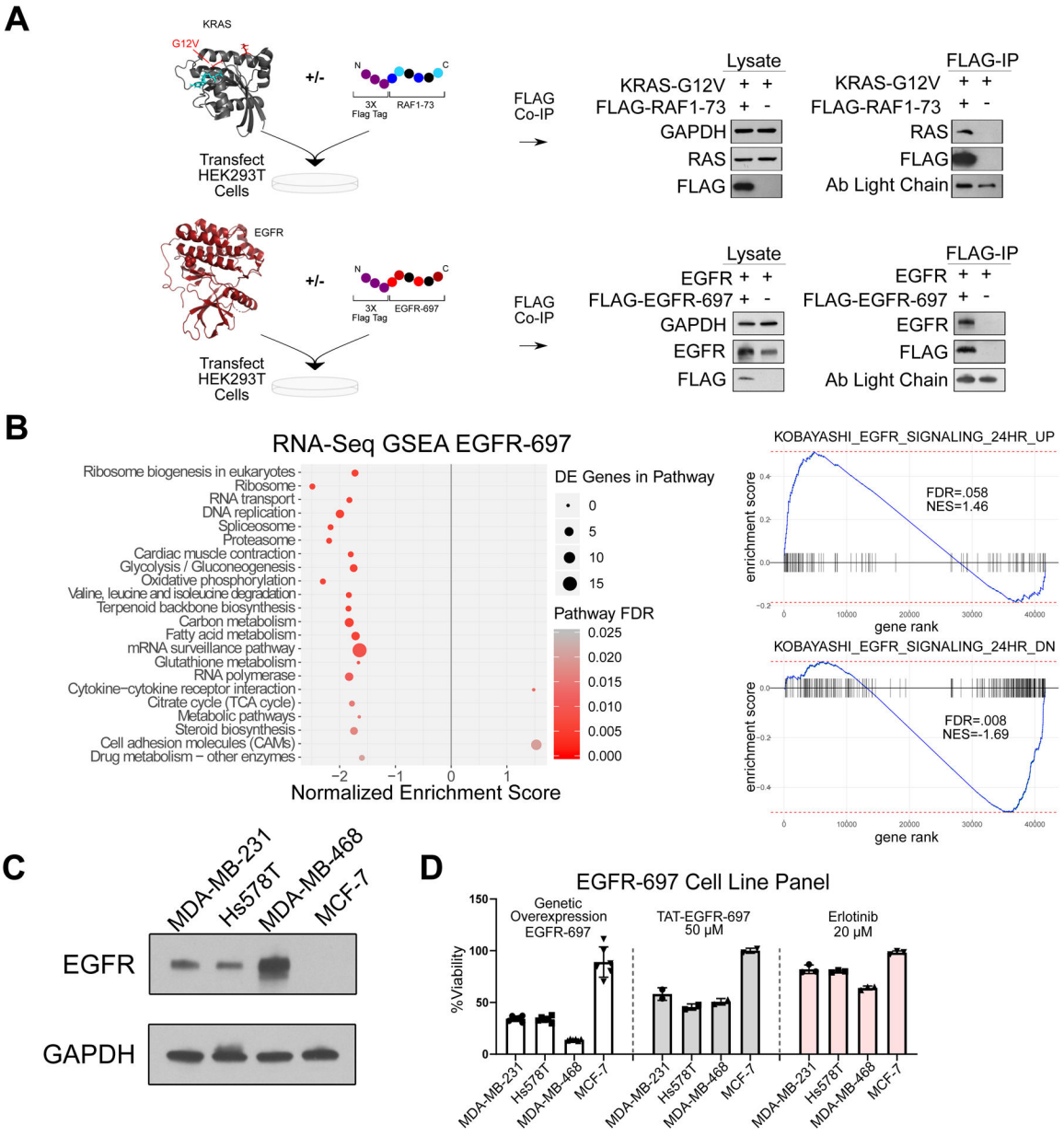


FIGURE 6. Cancer driver derived peptides show context dependent activity

(A) Peptide mechanism explored via co-immunoprecipitation. 3X-Flag tagged RAF1-73 derived from the Ras binding domain 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 which 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, Huber and Anders, 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 24hrs for the exogenously delivered molecules. For the genetically overexpressed EGFR-697, after 7 days of growth there was a significant association between EGFR expression levels and a cell lines viability relative to a GFP transduced control ($P<0.0001, r=-.803$). 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 50μM, the cell lines with detectable EGFR expression show a reduction in viability after 24hrs 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 (24hr treatment) than non-EGFR expressing MCF7 cells. Cell viabilities for Erlotinib treated cells are normalized to DMSO treated cells on the same plate. Data indicates mean ± standard deviation. (See also Supplemental Table 10).