







Genetic fusions favor tumorigenesis through degron loss in oncogenes

Jing Liu ^{1,10}, Collin Tokheim^{2,3,10}, Jonathan D. Lee ^{4,5,10}, Wenjian Gan ⁶, Brian J. North⁷, X. Shirley Liu ^{2,3}✉, Pier Paolo Pandolfi ^{4,8,9}✉ & Wenyi Wei ¹✉

Chromosomal rearrangements can generate genetic fusions composed of two distinct gene sequences, many of which have been implicated in tumorigenesis and progression. Our study proposes a model whereby oncogenic gene fusions frequently alter the protein stability of the resulting fusion products, via exchanging protein degradation signal (degron) between gene sequences. Computational analyses of The Cancer Genome Atlas (TCGA) identify 2,406 cases of degron exchange events and reveal an enrichment of oncogene stabilization due to loss of degrons from fusion. Furthermore, we identify and experimentally validate that some recurrent fusions, such as BCR-ABL, CCDC6-RET and PML-RARA fusions, perturb protein stability by exchanging internal degrons. Likewise, we also validate that EGFR or RAF1 fusions can be stabilized by losing a computationally-predicted C-terminal degron. Thus, complementary to enhanced oncogene transcription via promoter swapping, our model of degron loss illustrates another general mechanism for recurrent fusion proteins in driving tumorigenesis.

¹Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. ²Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02215, USA. ³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. ⁴Cancer Research Institute, Beth Israel Deaconess Cancer Center, Department of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. ⁵Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. ⁶Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA. ⁷Department of Biomedical Sciences, Creighton University, Omaha, NE 68178, USA. ⁸Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin 10124, Italy. ⁹Renown Institute for Cancer, Nevada System of Higher Education, Reno, NV 89502, USA. ¹⁰These authors contributed equally: Jing Liu, Collin Tokheim, Jonathan D. Lee. ✉email: xslu@ds.dfci.harvard.edu; PierPaolo.PandolfiDeRinaldis@renown.org; wwei2@bidmc.harvard.edu

Genetic alterations accumulate during the multistep processes of tumorigenesis, which lead to the transformation of normal cells into cancer cells^{1,2}. Large-scale tumor sequencing has enabled the systematic identification of gene fusions derived from chromosomal rearrangements. The most famous chromosomal rearrangement, t(9;22), was identified in 1960 as a hallmark of chronic myeloid leukemia (LCML) and subsequently named the Philadelphia chromosome³. The Philadelphia chromosome prompts the discovery of the BCR (breakpoint cluster region)-ABL fusion^{4,5} and the clinical application of imatinib as a targeted therapy for treating LCML patients⁶. To date, chromosomal rearrangements have been reported as frequent genetic drivers of several types of human cancer, such as ETS-related gene (ERG) fusions in prostate cancer⁷, RET or anaplastic lymphoma kinase (ALK) fusions in lung cancer^{8,9}, and fibroblast growth factor receptor 3 (FGFR3) fusions in bladder cancer¹⁰. According to a previous comprehensive analysis of The Cancer Genome Atlas (TCGA), there are more than 25,000 genetic fusion events, which might drive the development of approximately 16.5% of total cancer cases¹¹. However, the molecular mechanisms underlying how these gene fusions are oncogenic remain largely unclear for most of these cases.

Several mechanisms have been proposed to explain the oncogenicity of fusion proteins¹². One mechanism relies on transcriptional up-regulation due to promoter exchange between two genes, such as the fusion of ERG with the 5'-UTR of *TMPRSS2* (*transmembrane serine protease 2*) to trigger the transcription of fusion products in prostate cancer¹³. Another mechanism for the oncogenic nature of fusion proteins is the constitutive activation of kinases, often achieved by dimerization or oligomerization, such as for ABL, ALK, and RET fusions^{5,8,9,14}. A third mechanism is the loss of an auto-inhibitory segment, such as for BRAF fusions¹⁵. We hypothesized that altered protein stability could be an additional widespread mechanism for gene fusion events. While there have been a few instances characterized, such as *TMPRSS2-ETV1*¹⁶ and *TMPRSS2-ERG* fusions^{17,18}, altered protein stability has not been previously discussed as a major mechanism for the functional impact of gene fusions on tumorigenesis¹².

Intracellular protein homeostasis is strictly controlled by the balance between protein synthesis by the ribosome and protein degradation by the ubiquitin proteasome system (UPS)¹⁹. Proteins are targeted for 26S proteasome-mediated proteolysis by conjugation of a poly-ubiquitin chain onto lysine residues^{20,21}. The exquisite selectivity of the ubiquitination process on a cellular protein relies on its recognition by specific E3 ubiquitin ligase(s)^{22–24}. There are more than 600 E3 ligases encoded in the human genome, but only a few have been extensively characterized^{22–24}. The binding specificity of an E3 ligase is thought to be governed by short sequence motifs on the substrate, known as degrons^{25,26}, which are typically several amino acids long. Some E3 ligases display strong locational preference for degrons at the C-terminus or N-terminus of a protein^{27–29}, while other degrons can be found within the internal protein sequence²⁶.

While our recent analysis suggest that oncogenic point mutations frequently perturb the function of the UPS (~19% of cancer driver genes)³⁰, it remains unknown whether gene fusions also frequently alter protein stability. In this study, we identify 2406 fusion candidates with possible degron loss preferentially occurred in oncogenes (OG) from bioinformatics analysis across 33 cancer types, and further experimentally validate the increased protein stability resulting from loss of degrons in 5 fusions, indicating that altering protein stability due to degron loss is a general mechanism for cancer-related genetic fusions to promote tumorigenesis.

Results

A systematic computational analysis of degron loss. Previous reports of degron loss in gene fusions have largely focused on prostate cancer due to the high frequency of oncogenic fusions^{7,11,16}. The two most common fusion events in prostate cancers involve either ERG (>50% of primary tumor samples) or ETV family transcription factors (<10%). Notably, we and others have reported that through the fusions with *TMPRSS2* or other 5' partners, ERG loses an SPOP (speckle-type POZ protein) degron, thus leading to stabilization of the fusion protein (Supplementary Fig. 1a, b)^{17,18}. Similarly, ETVs also lose two COP1 degrons during fusion, which leads to escape from COP1-mediated degradation (Supplementary Fig. 1c–f)¹⁶. These studies prompted us to hypothesize that degron loss could be a general mechanism for genetic fusion events in driving tumorigenesis beyond prostate cancer (Fig. 1).

We therefore systematically analyzed 24,336 fusion genes reported across 9624 tumor samples in TCGA to discern the importance of a degron loss mechanism (Fig. 2a and Supplementary Data 1). Consistent with a likely substantial contribution of gain-of-function fusions towards tumorigenesis in TCGA, we found that fusions containing previously implicated oncogenes were enriched for in-frame fusions ($p < 5 \times 10^{-10}$, Fig. 2b) and preferentially retained functional protein domains ($p < 8 \times 10^{-6}$, Fig. 2c). To understand the specific contribution of degron loss, we systematically predicted internal degrons for E3 ligases with known motifs using a Random Forest machine learning model (Supplementary Fig. 2a–d, “Methods”). In addition, we also unbiasedly predicted C-terminal degrons using the deepdegron method that we previously developed³⁰ to identify degron motifs from the global protein stability assay³¹. Notably, among the highly recurrent fusions (>10 tumor samples), degron loss is significantly more enriched in oncogenes (Fig. 2d, e, 30.4%) than tumor suppressor genes (Fig. 2d, e, 14%, $p = 0.01$, Fisher’s exact test) or likely passenger genes (Fig. 2d, e, 13.2%, $p = 5 \times 10^{-6}$). These results were robust to the choice of threshold for recurrent fusions (Supplementary Fig. 2e). Moreover, fusions involving oncogenes displayed a clear bias for degron loss over degron gain (Fig. 2e). In contrast to oncogenes, fusions involving tumor suppressor genes had a trend towards degron gain, although the overall number of events was relatively low. Taken together, these results indicate that degron loss could be a major contributor to the oncogenicity of gene fusions.

We next systematically discovered the specific genes involved in fusions that preferentially underwent degron loss. By first analyzing internal degrons, we identified 47 genes where gene fusions led to more predicted loss of internal degrons than expected ($q < 0.1$, permutation test, “Methods”³²), which contained several known oncogenes, such as ABL1, RET, and IGF1R (insulin like growth factor 1 receptor) fusions (Fig. 2e, Supplementary Fig. 2f, g and Supplementary Data 2). Likewise, genes that are fusion partners to well-known oncogenes were also common (Fig. 2g and Supplementary Fig. 2h), such as the statistically significant degron loss for *CCDC6*, particularly when fused with RET ($p < 0.0001$, permutation test). This suggested a potential selection pressure to avoid protein degradation in both members of a fusion gene product. By further restricting our analysis to only previously implicated oncogenes ($q < 0.1$), we found additional internal degron loss events for rare oncogenic fusions containing *PDGFRA/FIP1L1* (platelet-derived growth factor receptor alpha/factor interacting with PAPOLA and CPSF1, Supplementary Fig. 2i) and a positive control *ETV1* fusion (Supplementary Fig. 1c and Supplementary Data 2), with additional fusions containing *ETV4* ($q = 0.16$, Supplementary Fig. 1e) and *ETV5* ($q = 0.13$, Supplementary Fig. 1f) at the borderline of statistical significance. Thus, it is plausible that with

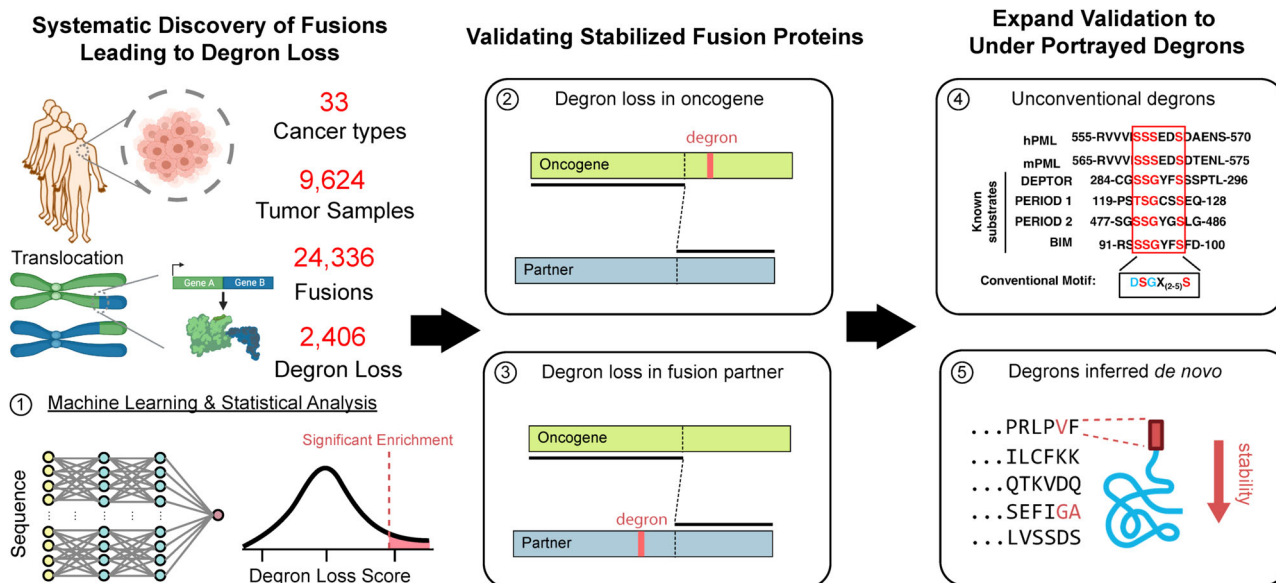


Fig. 1 Degron loss-prone genetic fusion events favor oncoprotein stabilization and tumorigenesis. A schematic diagram to illustrate the mechanism of how gene fusion leads to degron gain/loss event and thus perturbs proteins stability. Created with BioRender.com.

greater sample size, additional fusions leading to degron loss in genes not previously known to be oncogenes will be found.

Genetic fusions with degron loss are likely cancer type-specific.

Given that genetic fusions have been previously noted to exhibit tissue specificity, such as ETV family fusions in prostate cancer⁷ and ALK or MET fusions in lung cancer^{8,9,14}, we hypothesized that inclusion of cancer type-specificity would likely improve our statistical power. To this end, using low entropy as a metric for specificity (Fig. 3a), we found that genes involved in highly cancer type-specific fusions were significantly enriched for previously known oncogenic fusions ($p = 3.1 \times 10^{-16}$, Fischer's Exact test, Supplementary Fig. 3a and Supplementary Data 3). Interestingly, we observed statistical significance for the loss of internal degrons from several fusion genes, such as 5' EML4 fusions, 3' NSD1 fusions and the previously validated 3' ETV4 fusions, only when considered in conjunction with cancer type (Fig. 3b and Supplementary Data 3). Overall, degron loss contributes to many of the most highly recurrent gene fusions specific to particular cancer types (Fig. 3c), including PML-RARA in acute myeloid leukemia (LAML), EGFR-SEPT14 in gliomas, and Tmprs2-ERG in prostate cancers. Thus, by using an unbiased statistical approach, we found both known (e.g. ETV fusions) and previously unknown cases of gene fusions leading to degron loss.

As numerous gene fusions with degron loss exhibit cancer type-specificity, we sought to identify the corresponding E3 ligases likely involved in this specificity. These associations include APC/CDC20 for EML4-ALK fusions in lung cancers (Fig. 3d), SPOP for NUP98-NSD1 (Fig. 3e) and BCR-ABL fusions in LAML, and FBW7 (or FBXW7, F-box and WD repeat domain containing 7) for CCDC6-RET fusions in thyroid carcinomas (THCAs) (Supplementary Data 3). Indeed, the E3 ligase most frequently involved in degron loss is the known tumor suppressor gene SPOP (Supplementary Fig. 3b), which suggests a selective pressure to avoid protein degradation in a variety of cancer types.

Genetic fusions with degron loss are associated with downstream functional consequences.

Based on the above analyses that degron loss may lead to increases in the stability of fusion proteins, we hypothesized that tumors containing these fusions would be associated with an altered proteomic and subsequent

transcriptomic state of cancer cells. To validate this hypothesis, we first analyzed the abundance of 198 proteins measured by reverse phase protein arrays (RPPA) across the TCGA (Supplementary Data 4). Consistent with our finding of degron loss, 5' ERBB2 and 5' EGFR fusions had significantly higher expression levels and active phosphorylation of their respective proteins (Supplementary Fig. 3d). Furthermore, degron loss in CCDC6 fusions led to elevated levels of downstream effectors, including active phosphorylated forms of YAP, PKC, p38, and 4EBP1 (Supplementary Fig. 3d). Given the limited number of proteins assayed by RPPA, we next analyzed for potential downstream consequences on the transcriptome through modulating the activity of transcription factors (TFs). Since many oncogenic fusions are involved in protein signaling, we reasoned that TF activity could be best approximated by the expression of TF target genes. Here, TF target genes are defined by thousands of ChIP-seq profiles from the Cistrome database³³. Using the RABIT algorithm³⁴ to find coordinated differential expression of TF target genes, we found 113 significant associations between TF activity and fusion events (Supplementary Fig. 3e, f and Supplementary Data 5). In support of the reliability of our analysis, previous studies support several of the most significant associations identified, such as AR for ERG fusions, TTF1 for EML4-ALK fusions, and TAL1 for BCR-ABL fusions³⁵⁻³⁷. Interestingly, 5' EGFR fusions were significantly associated with increased STAT1 activity, suggesting that it is either a downstream consequence of EGFR kinase activity or an immunogenic consequence of a predicted fusion neoantigen^{11,38}. Cumulatively, our analyses indicate that fusion events undergoing degron loss have significant downstream functional consequences on both the proteome as well as the transcriptome.

BCR-ABL fusion leads to loss of the SPOP degron in ABL and stabilization of fusion protein.

Our systematic computational approach allowed us to potentially find, even for the most well-studied oncogenes, previously unknown degrons that were lost during fusion. For example, our analyses predicted that BCR-ABL fusions led to the loss of a SPOP degron originally found in the oncoprotein ABL1 (Fig. 4a). BCR-ABL is the gene fusion product of the Philadelphia chromosome found in LCML (Supplementary Fig. 4a, b)³⁻⁵, and it has been a therapeutic target for LCML

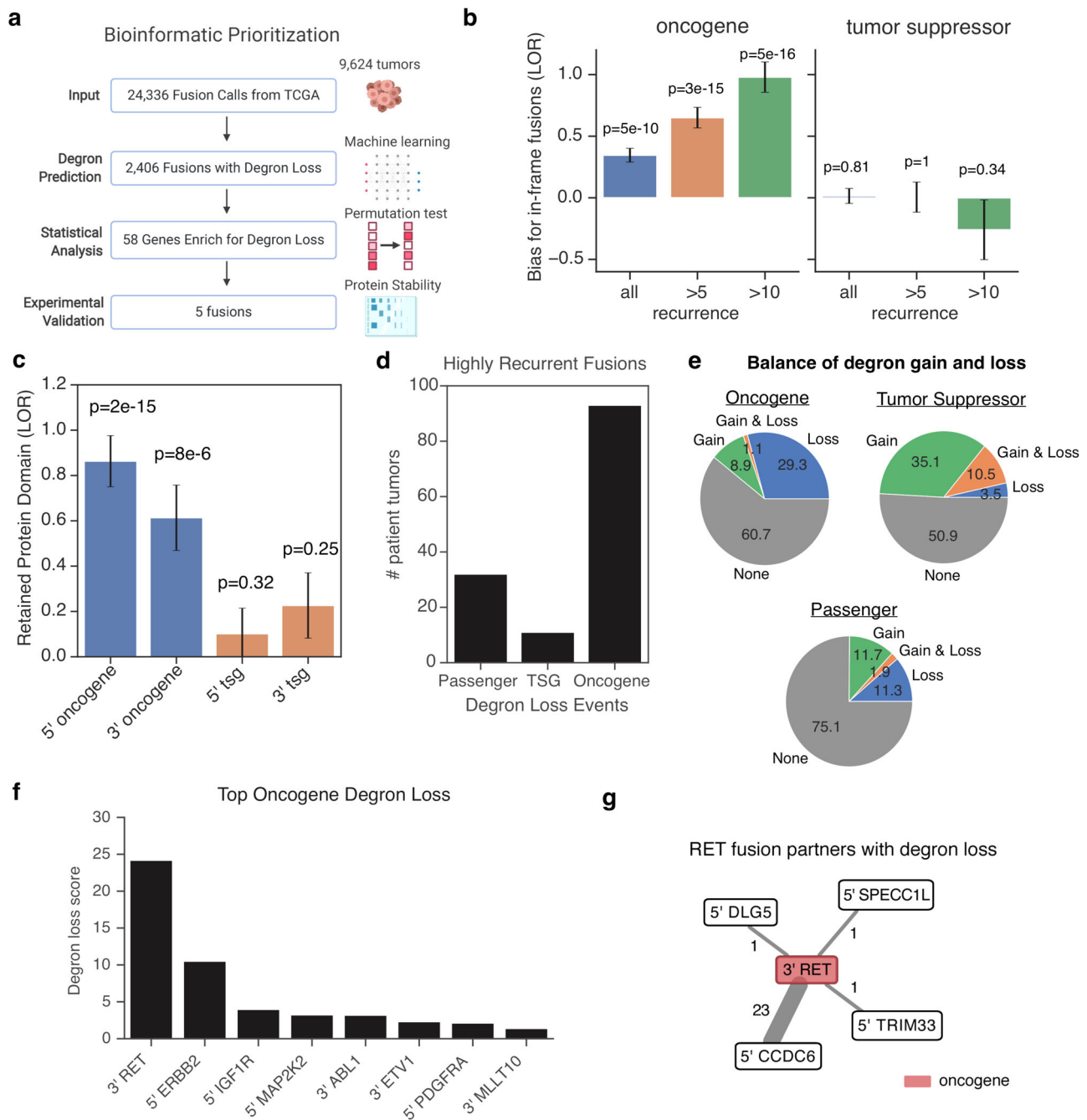


Fig. 2 Degron landscape of genetic fusions in cancer. **a** Combined analysis pipeline for validating cases of degnon loss in fusion genes. Created with BioRender.com. **b** Enrichment for in-frame gene fusions for oncogenes and tumor suppressor genes relative to likely passenger genes (STAR methods). Recurrence is the number of tumors that a fusion gene was present in the TCGA. Barplot displays the log odds ratio (LOR), with error bars indicating ± 1 SE. In order of bars from left to right, sample sizes (i.e. number of fusions) are $n = 17,848$, $n = 4254$, $n = 1316$, $n = 17,581$, $n = 3,799$, and $n = 925$. A two-sided Fisher's exact test was performed. **c** Enrichment analysis for whether gene fusions result in the retention of at least one protein domain from the original protein sequence. Barplot displays the log odds ratio (LOR), with error bars indicating ± 1 SE. In order of bars from left to right, sample sizes (i.e. number of fusions) are $n = 8911$, $n = 9139$, $n = 8935$, and $n = 8896$. A two-sided Fisher's exact test was performed. **d** The total number of tumors in the TCGA that contain highly recurrent fusion genes (a fusion found in >10 tumors). **e** Balance of degnon gain and loss in oncogene, tumor suppressor genes or passenger genes. **f** The top 10 most significant oncogenes that are biased towards loss of internal degnons in gene fusions. **g** Network diagram showing degnon loss in the fusion partners to oncogene RET. Line width represents the frequency of the event. Red indicates previously annotated oncogenes. Numbers indicate the recurrence of the fusion across TCGA. The relevant raw data are provided in Source Data.

treatment for decades^{6,39,40}. Our computational analysis predicted that the fusion between ABL and its 5' partner BCR leads to loss of a degnon recognized by SPOP (Fig. 4a, b), which is a substrate adaptor of the Cullin 3 family of E3 ligases. The putative SPOP degnon (17-LSSSS-21) is evolutionarily conserved in

human and mouse ABL1 protein sequence, and similar to several known SPOP substrates, including ERG, AR, and DEK (Fig. 4c)^{17,18,41}. This indicated that SPOP degnon loss was plausible for BCR-ABL fusions, and thus might complement a previously proposed mechanism of constitutive kinase activity⁴².

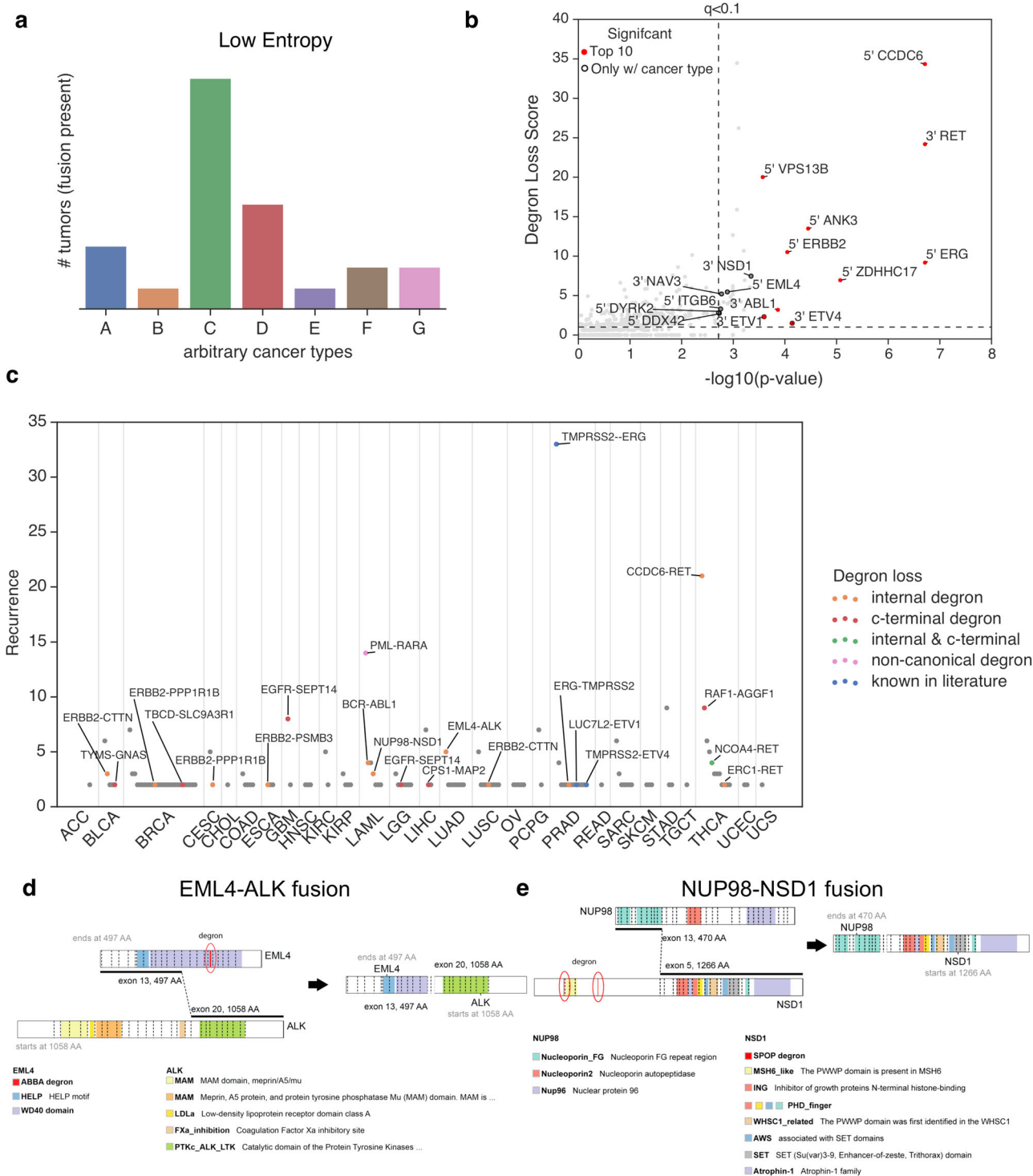
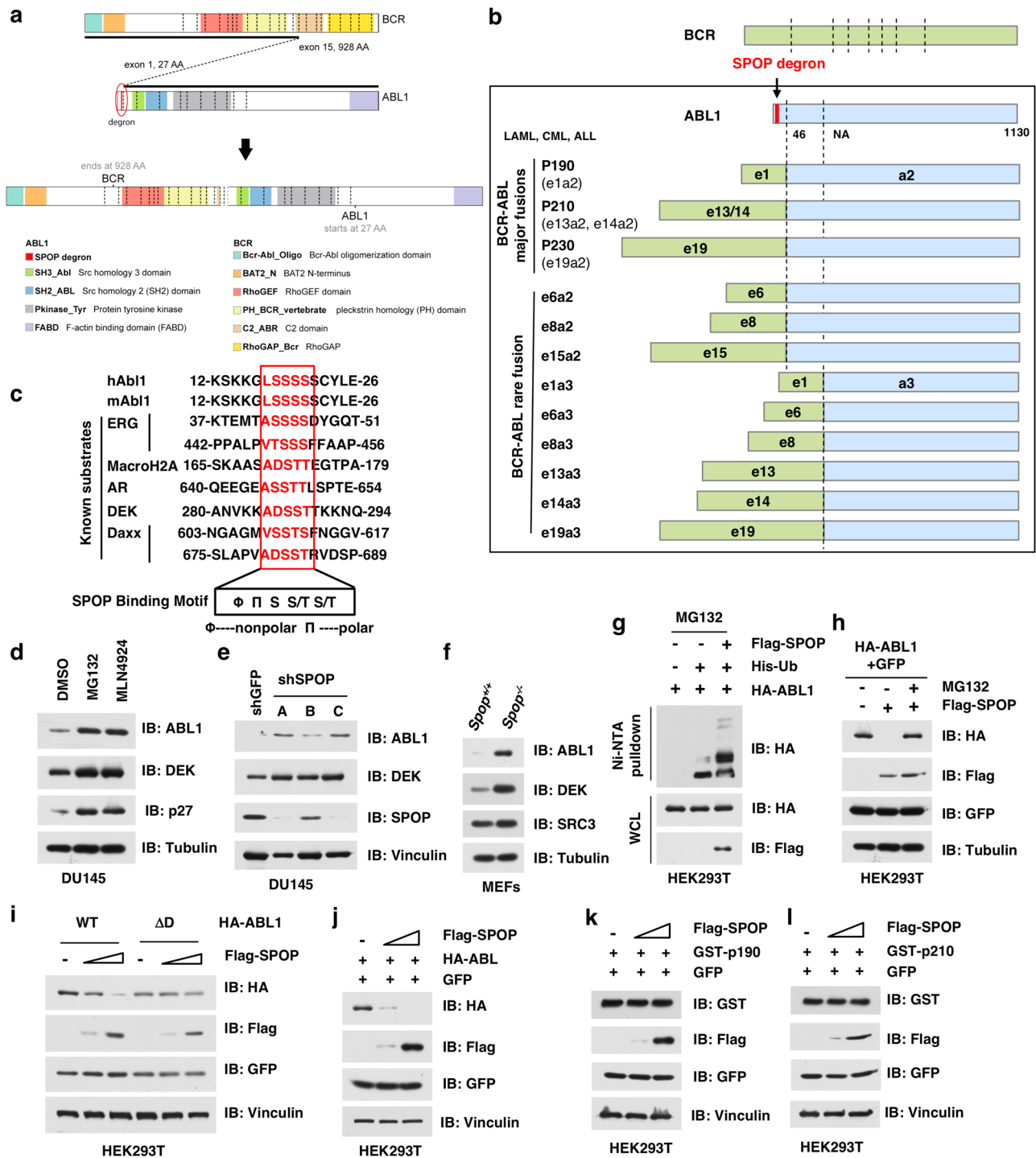


Fig. 3 Cancer type-specificity of genetic fusions reveals oncogenicity. **a** Conceptual diagram showing that a low entropy value that summarizes the distribution of a fusion across cancer types indicates the fusion gene is cancer type-specific. **b** Scatter plot showing the statistical significance of gene fusions for loss of internal degrons when also considering information on cancer type-specificity. X-axis represents the combined p value for both degron loss and cancer type-specificity (Fisher’s method), while the Y-axis represents the effect size for degron loss. P values were calculated using a one-sided permutation test. See “Methods” for additional details on calculation of p values. **c** Scatter plot indicating the most recurrent fusion genes for particular cancer types. Fusion genes with degron loss are labeled in color, including internal degrons, C-terminal degrons, and those degrons previously known in the literature. **d** A schematic diagram showing the fusion of *EML4* and *ALK* genes in lung adenocarcinoma. **e** A schematic diagram showing the fusion of *NUP98* and *NSD1* genes in lung adenocarcinoma. The relevant raw data are provided in Source Data.



First, we aimed to experimentally validate ABL1 as a bona fide substrate of the Cullin 3^{SPOP} E3 ligase. Indeed, similar to the known SPOP substrate ERG, the protein abundance of ABL1 increased in DU145 prostate cancer cells upon treatment with either the proteasome inhibitor MG132 or the neddylation inhibitor MLN4924 (Fig. 4d). Depletion of endogenous *Cullin 3* (Supplementary Fig. 4c) or *SPOP* (Fig. 4e) led to an increase of ABL1 protein abundance. Furthermore, *Spop*^{-/-} mouse embryonic fibroblasts (MEFs) had relatively higher protein abundance of ABL1 than wild-type (WT) MEFs (Fig. 4f), consistent with the positive control SPOP substrates DEK⁴¹ and SRC3 (ref. 43). As expected for abrogating protein degradation, the protein half-life

of ABL1 was dramatically longer in *Spop*^{-/-} MEFs than in WT MEFs (Supplementary Fig. 4d, e). Moreover, ectopic expression of SPOP promoted the ubiquitination and degradation of ABL1 protein, which could be largely inhibited by the proteasome inhibitor MG132, thus indicating a proteasome-dependent mechanism (Fig. 4g, h). To ensure the enhanced degradation of the ABL1 protein was due to an on-target mechanism, we evaluated whether cancer-derived mutations, including Y87C, F102C, W131G, and F133V (Supplementary Fig. 4f)⁴⁴, that abrogate SPOP binding to substrates would fail to promote ABL1 protein degradation. Notably, ectopic expression of WT SPOP, but not the SPOP mutants, could degrade ABL1 protein

Fig. 4 BCR-ABL escapes from SPOP-mediated degradation due to degron loss. **a** A schematic diagram showing the fusion of *BCR* and *ABL* genes in chronic myeloid leukemia (LCML). **b** BCR-ABL fusion leads to the loss of a SPOP degron (-LSSSS-) on the N-terminus of ABL protein, resulting in stabilization of BCR-ABL fusion proteins, including p190, p210, p230, and other rare fusions. **c** SPOP degron in ABL1 is conserved among species. Φ : nonpolar; Π : polar. **d** The degradation of ABL1 is proteasome- and Cullin E3 ligase-dependent. DU145 cells were treated with either 10 μ M MG132 or 1 μ M MLN4924 for 12 h. Cells were harvested, lysed, and immunoblotted for ABL1 with ERG and p27 as positive controls. **e** Depletion of SPOP leads to accumulation of ABL1 in DU145 cells. **f** *Spop*^{-/-} MEFs has relatively higher protein abundance of ABL1 than WT MEFs. **g** SPOP promotes the ubiquitination of ABL1 in cells. HEK293T cells were transfected with indicated constructs and treated with 30 μ M MG132 for 6 h. Cells were lysed under denaturing conditions and His-Ub-conjugated proteins were pulled down with Ni-NTA-resin, washed and IB for indicated proteins. **h** SPOP promotes the degradation of ABL1 in a proteasome-dependent manner. HEK293 cells were transfected with indicated constructs and treated with or without 10 μ M MG132 for 12 h, and then harvested, lysed, and immunoblotted for indicated proteins. **i** Deletion of the SPOP-binding motif in ABL1 leads to stabilization of ABL1 protein. Cells were transfected with indicated constructs for 48 h, followed by harvested, lysed, and immunoblotted for indicated proteins. **j-l** Compared with ABL-WT, BCR-ABL fusion proteins, both p190 and p210, which lack the SPOP degron, were relatively resistant to SPOP-mediated degradation. Cells were transfected increasing concentrations of Flag-SPOP together with either ABL-WT (**j**), BCR-ABL fusion proteins, both p190 (**k**) and p210 (**l**). Cells were then harvested, lysed, and IB of indicated proteins. Two independent experiments were performed for **d-l**. The relevant raw data and uncropped blots are provided in Source Data.

(Supplementary Fig. 4g). Taken together, these results support our computational prediction and indicate that ABL1 is likely a bona fide substrate of the SPOP E3 ligase.

We next sought to investigate whether BCR-ABL fusion proteins, named p190 and p210, could escape SPOP-mediated degradation in cells. This requires firstly excluding the possibility of another SPOP degron in ABL1 which is not lost in a fusion. We found that after deleting the predicted SPOP degron in the ABL1 protein, the resultant ABL1- Δ D mutant was relatively resistant to SPOP-mediated degradation in cells (Fig. 4i). Secondly, ectopic expression of SPOP degrades only WT ABL1 (Fig. 4j), but not BCR-ABL1 fusion proteins (Fig. 4k, l), indicating that BCR-ABL fusions escape from SPOP-mediated degradation via loss of the sole SPOP degron in the N-terminus of ABL1. Apart from the most frequent fusions, p190 and p210, there are several other low frequent fusions (e19a2) and rare fusions (e6a2, e8a2, e15a2, e1a3, e6a3, e8a3, e13a3, e14a3, and e19a3) in LCML, LAML, and acute lymphocytic leukemia (ALL, Fig. 4b). Notably, the SPOP degron in exon 1 of *ABL* is lost in all of these genetic fusions (Fig. 4b), suggesting a similar mechanism for these fusions in promoting tumorigenesis.

CCDC6-RET fusion escapes from FBW7-mediated degradation. Although the BCR-ABL fusion led to loss of a degron in the known oncoprotein ABL1, degron loss in the fusion partner to known oncogenes might also contribute towards increasing protein stability of fusion proteins. Our computational analysis showed that CCDC6-RET fusions were highly enriched for loss of predicted degrons in both fusion components, including FBW7 degrons in CCDC6 and a D-box degron in the oncogene RET (Fig. 5a). There are several variants of CCDC6-RET fusion, which contain N-terminal fragments of CCDC6 and C-terminus of RET, in thyroid carcinoma⁴⁵, non-small cell lung cancer⁹, and other cancer types⁴⁶ (Fig. 5b). Apart from *RET*, *CCDC6* also fuses with other genes, including *ROS1* (ref. ⁴⁷) and *PDGFRB*⁴⁸. Notably, the FBW7 degrons in CCDC6 were lost in all of these fusion proteins (Fig. 5b), suggesting an analogous mechanism of increasing protein stability.

Given these computational predictions, we expected the putative FBW7 degrons in CCDC6 would be similar to those found in previously known substrates. Sequence alignment showed that the predicted FBW7 degrons ((pT/pS)PXX(pS/pT), p indicating phosphorylation) were conserved in both human and mouse CCDC6, consistent with several known FBW7 substrates, such as c-Myc^{49,50}, c-Jun⁵¹ and cyclin E⁵² (Fig. 5c). The recognition by FBW7 is known to dependent on the phosphorylation of serine or threonine residues within its degron motif⁵³. As expected, large-scale phospho-proteomics data (<https://www.phosphosite.org>)⁵⁴ have detected phosphorylation on

residues within the putative FBW7 degrons (Thr-357, Ser-361, Thr-380, Ser-384, and Thr-427), further supporting CCDC6 as a potential substrate of FBW7.

To experimentally assess whether CCDC6-RET escapes FBW7-mediated degradation, we aimed to first validate CCDC6 as a bona fide FBW7 substrate. We found that the CCDC6 protein levels were relatively higher in *FBXW7* (also known as *FBW7*) null DLD1 and HCT116 cells, compared with respective WT parental control cells (Fig. 5d). *FBW7* is frequently mutated and inactivated in colorectal cancer (CRC), and *FBW7* mutant CRC cells have relatively lower FBW7 expression and higher abundance of FBW7 substrates such as MCL1 (ref. ⁵⁵). Thus, we further assessed CCDC6 protein levels in a panel of CRC cells with either WT or mutant *FBW7*, and found that *FBW7*-mutant cells trend to have relatively higher abundance of CCDC6 protein than *FBW7*-WT cells (Fig. 5e and Supplementary Fig. 4h). These data together indicate that CCDC6 is a ubiquitin substrate of FBW7. More importantly, compared with WT-CCDC6, CCDC6-RET fusion protein escaped recognition by FBW7 (Fig. 5f), leading to stabilization of the resultant fusion product in the in vivo ubiquitination assay (Fig. 5g). In keeping with this notion, depletion of *FBW7* extended the half-life of CCDC6 protein in a cycloheximide (CHX) chasing assay (Supplementary Fig. 4i, j).

Unlike a prior report of CCDC6 as a substrate of FBW7⁵⁶, our findings support the relevance of FBW7 degron loss in CCDC6 fusions. Interestingly, given that CCDC6-RET fusions are predicted to generate neoantigens (Supplementary Fig. 4k)¹¹, an increase of CCDC6-RET protein stability might also reduce the generation of antigenic peptides derived from proteasomal degradation⁵⁷, thus evading an otherwise strong immune response ($p = 0.02$, likelihood ratio test; Supplementary Fig. 4l). To assess how loss of FBW7 degrons in the CCDC6 protein impact tumorigenesis, we further generated a DLD1 cell line that stably expresses either WT CCDC6 or CCDC6-RET fusion protein (Fig. 5h). We found that the CCDC6-RET-expressing cell line were more clonogenic than the WT-CCDC6-expressing cells in a colony formation assay (Fig. 5i, j) and resulted in larger tumors in a mouse xenograft model (Fig. 5k, l). Together, these data indicate that loss of FBW7 degrons in the CCDC6-RET fusion elevates its oncogenic phenotype.

PML-RARA escapes from β -TRCP-mediated degradation. Our systematic bioinformatic analyses of internal degrons relied on previously reported motifs for E3 ligases. However, we and others have validated degrons that may sometimes have unconventional motifs, such as the β -TRCP (F-box/WD repeat-containing protein 1A, FBXW1) degron in Twist (sSspvS)⁵⁸, PER1(tSgcsS)⁵⁹, and CHK1 (tSggcS)⁶⁰. Given drugs that induce protein degradation of PML-RARA lead to high response rates in acute

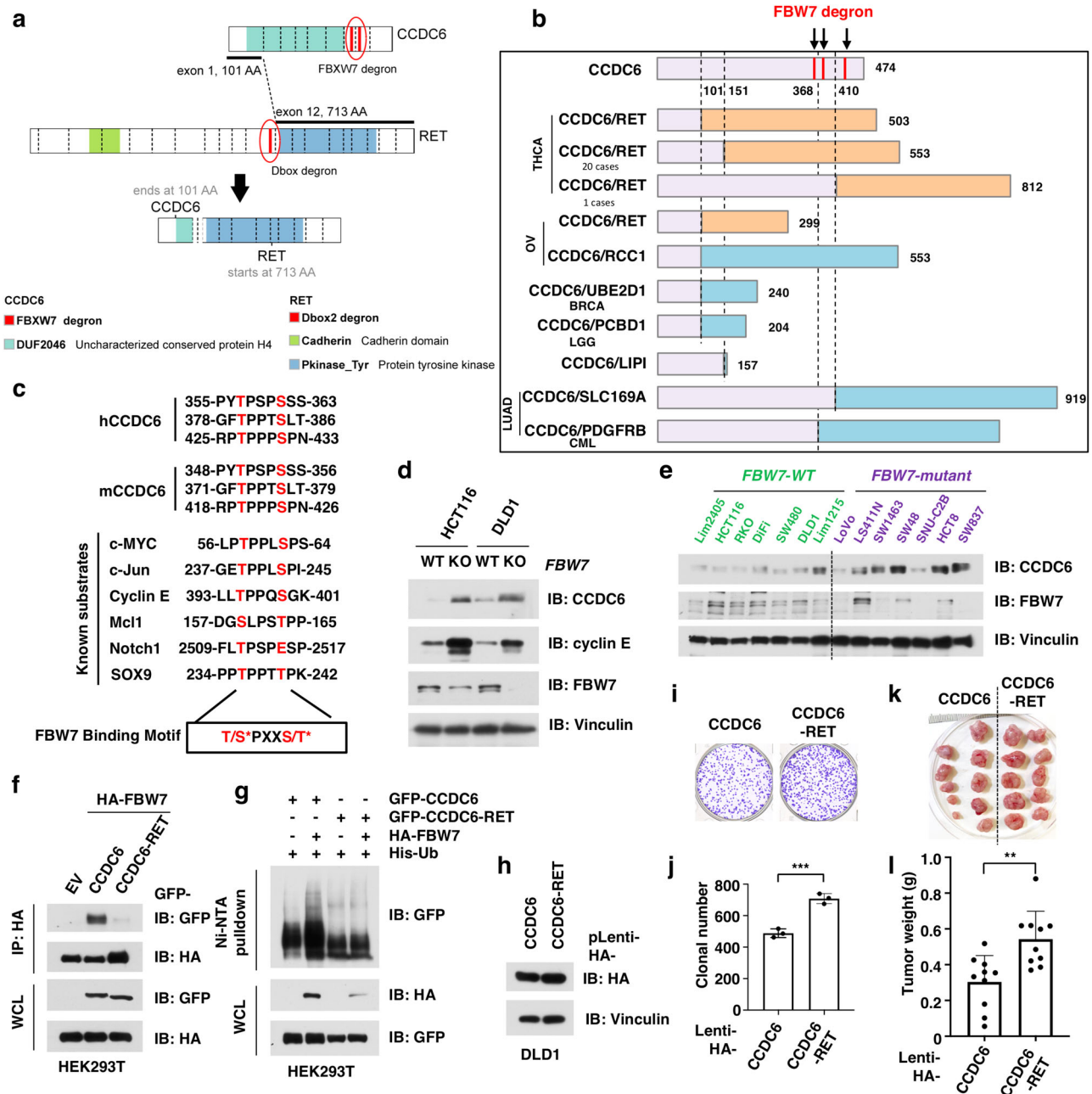


Fig. 5 CCDC6-RET escapes from FBW7-mediated degradation due to degron loss during genetic fusion. a A schematic diagram shows the fusion of *CCDC6* and *RET* genes in thyroid carcinoma (THCA). **b** CCDC6-RET fusion events remove the FBW7 degron from the CCDC6 protein, leading to the stabilization of CCDC6-RET fusion proteins. **c** FBW7 degron of CCDC6 is evolutionarily conserved among species. *Phosphorylation. **d** Knockout of *FBW7* in DLD1 and HCT116 cells leads to accumulation of CCDC6 protein. DLD1 or HCT116 WT and *FBW7* deleted cells were harvested, lysed, and immunoblotted for indicated proteins. **e** Colorectal cancer (CRC) cells with *FBW7*-mutant exhibited relatively higher protein abundance of CCDC6, compared with those cells with WT *FBW7*. **f** CCDC6-RET fusion protein escaped from the recognition by the FBW7 E3 ligase. Cells were transfected with HA-FBW7 together with either GFP-tagged CCDC6 and CCDC6-RET fusion constructs. Cells were treated with 10 μM MG132 for 12 h, followed by harvested, lysed, and immunoprecipitated (IP) with anti-HA. Inputs and immunoprecipitated material were immunoblotted for GFP and HA. **g** CCDC6-RET fusion protein escaped from FBW7-mediated ubiquitination. Cells were transfected with His-Ub, HA-FBW7 together with either GFP-tagged CCDC6 and CCDC6-RET fusion constructs. Cells were treated with 30 μM MG132 for 6 h, followed by lysed under denaturing conditions. The His-Ub-conjugated proteins were pulled down with Ni-NTA-resin, washed, and immunoblotted for indicated proteins. **h** Immunoblot analysis for DLD1 stable cell lines that ectopically express either wild-type (WT) CCDC6 or CCDC6-RET fusion protein. **i, j** CCDC6-RET-expressing cells had greater clonogenicity than WT-CCDC6-expressing cells in vitro. The DLD1 stable cell lines as in **h** were subjected for colony formation assay (**i**) and statistical analysis (**j**). $n = 3/\text{group}$. **k, l** CCDC6-RET-expressing cells developed larger tumors than WT-CCDC6-expressing cells in vivo. The DLD1 stable cell lines as in **h** were subjected for mouse xenograft assay (**k**) and statistical analysis (**l**). $n = 10 \text{ mice/group}$. Data are presented as mean \pm SD. ** $p < 0.01$; *** $p < 0.001$; by two-tailed Student's *t*-test. Two independent experiments were performed for **d** and **h**. The relevant raw data and uncropped blots are provided in Source Data.

promyelocytic leukemia (APL)^{61–65}, we hypothesized that the PML-RARA fusion may escape protein degradation through degron loss, but was missed in our systematic analysis. Interestingly, when using an unconventional β -TRCP degron motif (SSSxxS) reported from a previous study⁵⁸, we found PML-RARA may lead to loss of a degron that is originally found in the PML protein (560-SSSEDS-565) (Supplementary Fig. 5a–c). Among all the genetic fusions observed in TCGA, PML-RARA is the second most frequent fusion event and is preferentially found in LAML¹¹. Although not included in TCGA, nearly all APLs contain a PML-RARA fusion (95% of cases), which is caused by the reciprocal translocation t(15;17)(q24;q21)⁶⁶ (Supplementary Fig. 5a). Depending on the exact location of the translocation, PML-RARA fusion yields two major fusion proteins, namely PML-RARA-s and PML-RARA-l (Supplementary Fig. 5b).

Because of the high prevalence and therapeutic relevance of PML-RARA fusions, we next sought to experimentally validate PML as a bona fide substrate of β -TRCP and thereby implicate the predicted degron loss mechanism. Indeed, depleting endogenous *BTRC* (also known as β -TRCP), but not other F-box E3 ligase we tested, induced the accumulation of the endogenous PML protein (Supplementary Fig. 5d). In addition, depletion of β -TRCP extended the half-life of PML protein (Supplementary Fig. 5e). Consistent with the required phosphorylation of a β -TRCP degron, all four Serine residues were observed to be phosphorylated in a previous unbiased screen^{67,68}. Furthermore, depletion of *CSNK2A1* (also known as *CKII*)⁶⁹ also led to the accumulation of PML protein (Supplementary Fig. 5f), indicating that CKII is a potential kinase for PML. Using an in vitro phosphorylation assay, we found that mutation of serine residues within the putative β -TRCP degron (PML-4A) abolished the phosphorylation mediated by CKII kinase (Supplementary Fig. 5g). Moreover, the non-phosphorylated PML mutant (PML-4A) lost the interaction with β -TRCP, thus becoming resistant to β -TRCP-mediated degradation (Supplementary Fig. 5h). Taken together, these results indicate that PML is likely a bona fide substrate of β -TRCP, and loss of a β -TRCP degron likely renders greater stability to PML-RARA fusions.

Comprehensive analysis of C-terminal degron loss upon oncogenic gene fusion. The loss of a non-canonical degron in PML-RARA highlights that, even for well-studied E3 ligases like β -TRCP, our current knowledge of degron motifs is largely incomplete. This dearth of knowledge may lead to conclusions that overlook the role of degron loss in fusion events. Thus, we hypothesized that unbiased learning of degron motifs from data would reveal additional cases of degron loss in gene fusions. Although systematic profiling of degrons across the entire proteome has not yet been performed, a previous global protein stability (GPS) assay has systematically measured all C-terminal protein sequences for protein stability, which led to the discovery of several novel degron motifs³¹. We therefore leveraged a machine learning model trained on the GPS assay (deepDegron)³⁰ to score whether gene fusions preferentially lead to C-terminal degron loss (Fig. 6a and Supplementary Fig. 6a). We found gene fusions overall were substantially enriched for C-terminal degron loss, with statistical significance further improved by including cancer type information (Supplementary Fig. 6b and Supplementary Data 6). 5' EGFR and 5' RAF1 fusions yielded the highest scores for C-terminal degron loss among the 16 statistically significant genes (Fig. 6b). EGFR and RAF1 fusions additionally displayed substantial cancer type-specificity, with 65% of 5' EGFR fusions occurring in gliomas (Supplementary Fig. 6c) and 69% of 5' RAF1 fusions occurring in thyroid carcinomas (THCAs) (Supplementary Fig. 6d). Interestingly,

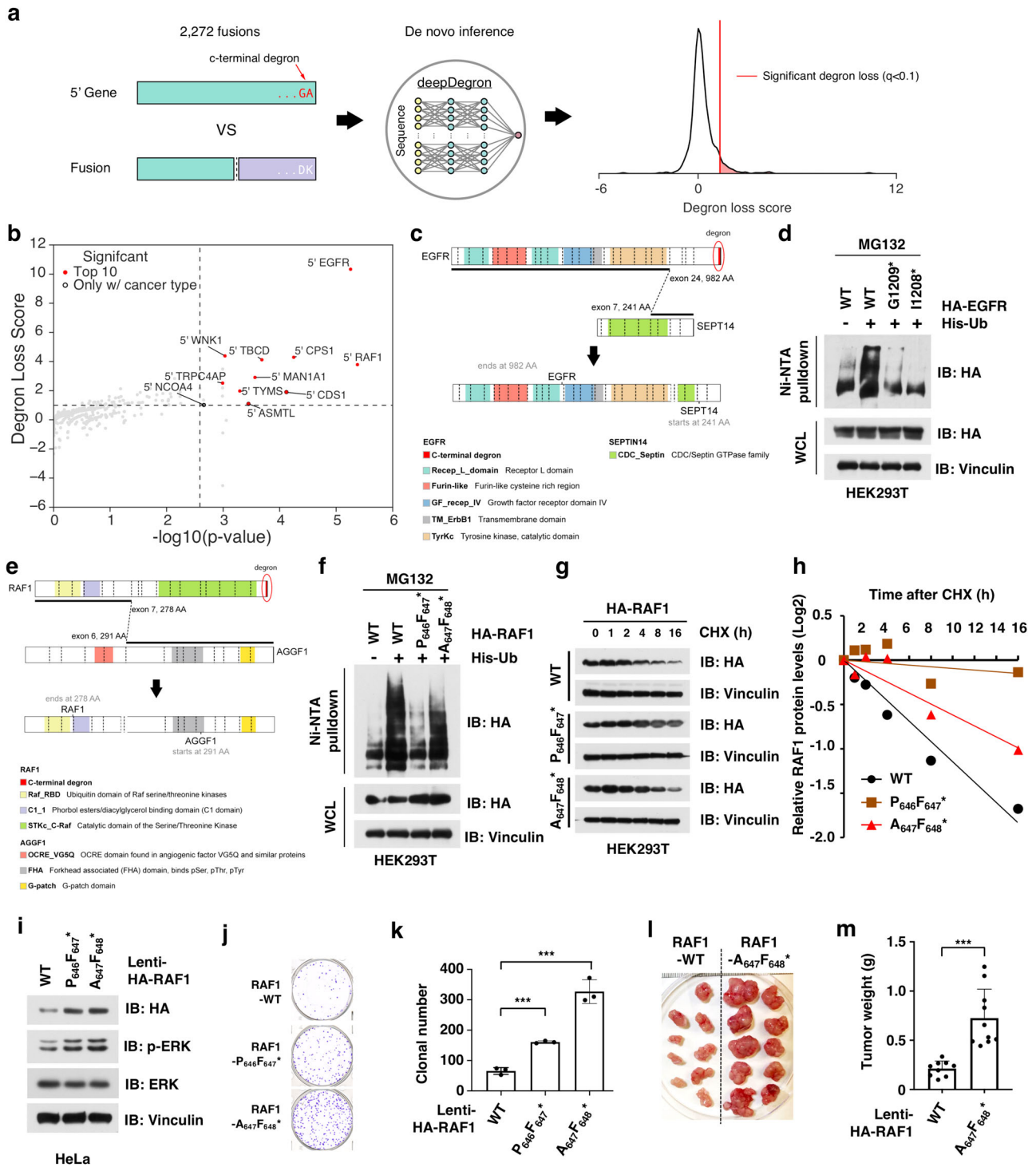
C-terminal and internal degrons can be simultaneously lost in a gene fusion, as observed for 5' NCOA4 fused with 3' RET (Supplementary Data 1 and Supplementary Fig. 6e).

EGFR-SEPT14 is the most frequent EGFR fusion and occurs mostly in glioblastoma (GBM) and low-grade gliomas (LGG). EGFR-SEPT14 fusions result in loss of a putative C-terminal degron (-GA*, Fig. 6c and Supplementary Fig. 6f), which is evolutionarily conserved among species (Supplementary Fig. 6g). To experimentally validate the key role of the -GA* motif in controlling the protein stability of EGFR protein, we generated two EGFR mutants with either deletion of the last alanine residue (G₁₂₀₉*) or glycine-alanine dipeptide (I₁₂₀₈*, S6F, Supplementary Fig. 6h). Notably, WT EGFR underwent significant ubiquitination, but both EGFR mutants resulted in a dramatic reduction in ubiquitination (Fig. 6d). This supports our computational finding that a C-terminal degron (-GA*) is lost in EGFR genetic fusions, which likely lead to increase stability of the resultant fusion proteins.

Among RAF1 fusions, RAF1-AGGF1 is the most frequent fusion, with 3' partners TRAK1 and PHC3 being observed less frequently. Our computational analysis predicts a putative C-terminal degron in RAF1 that is evolutionarily conserved among species (-Vx*, x means any amino acid, Fig. 6e and Supplementary Fig. 6i, j). Notably, all RAF1 fusions result in the loss of this putative C-terminal degron. To experimentally validate this finding, we mutated the putative RAF1 degron, by either deletion of the valine residue (P₆₄₆F₆₄₇*) or substitution of the valine to alanine (A₆₄₇F₆₄₈*, Supplementary Fig. 6k). Compared to WT RAF1, both mutants exhibited relatively less ubiquitination (Fig. 6f) and an extended protein half-life (Fig. 6g, h). To further assess whether loss of the C-terminal degron in RAF1 affects tumorigenesis, we generated a HeLa cell line that stably express either WT RAF1 or the degron loss mutants of RAF1 (P₆₄₆F₆₄₇* and A₆₄₇F₆₄₈*, Fig. 6i). Cells expressing the degron loss mutant forms of RAF1 were more clonogenic than those expressing WT-RAF1 in vitro in a colony formation assay (Fig. 6j, k). Furthermore, the RAF1 mutant-expressing cells (A₆₄₇F₆₄₈*) generated larger tumors in a mouse xenograft model than those expressing WT-RAF1 (Fig. 6l, m). These experimental results support our computational prediction that RAF1 loses a C-terminal degron (-Vx*) during fusion events, a process likely rendering greater stability to the fusion protein to facilitate tumorigenesis.

Discussion

While oncogenic gene fusions in human cancers have been extensively cataloged^{11,70}, the molecular mechanisms underlying their oncogenicity is incompletely understood. By analyzing more than 9000 tumors across 33 cancer types, we provide a systematic analysis of genetic fusions that demonstrate the prevalence of degron loss as a mechanism to increase the resultant protein stability. Among the 2406 fusion events that are predicted by machine learning to undergo degron loss, we experimentally validated five highly recurrent oncogenic gene fusions for altered protein stability and oncogenicity, thus more than doubling the number of previously validated cases^{16–18}. Prior systematic studies have largely focused on transcriptional over-expression of gene fusions caused by the exchange of promoters or enhancers^{11,13}. Our results suggest that degron loss is a complementary and generally applicable mechanism by which genetic fusions increase protein expression levels and thus promote tumorigenesis. We note that degron loss is not necessarily mutually exclusive with other previously proposed mechanisms such as promoter swapping, and therefore might act in concert with them to explain the oncogenicity of a gene fusion. For



example, genetic fusions that lead to loss of the C-terminal degrons (such as those for RAF1 or EGFR) might simultaneously promote the kinase activity through a similar mechanism of dimerization or oligomerization^{5,8,9,14}.

Despite our study providing the most comprehensive examination of degnon loss for genetic fusions to date, many instances of degnon loss may still be missed for a couple of reasons. First, our analysis still has limited statistical power in identifying enrichment for degnon loss in rare fusion events. For example, a previously validated KEAP1 degnon in IKBKB⁷¹ was lost in HOOK3-IKBKB fusions in breast cancer (Supplementary Fig. 3b), but this fusion event did not surpass our stringent false discovery

rate cutoff. Secondly, given the incomplete knowledge of degnon motifs, we further prioritized likely true degrons by employing machine learning and ensuring requisite post-translational modifications. However, these stringent criteria may also lead to false negatives in degnon motifs, such as the lack of a previously reported phosphorylation event in CCDC6 preventing the accurate prediction of a third FBW7 degnon. Further basic science efforts to decipher additional degnon motifs coupled with an increased throughput of tumor sequencing will be necessary to provide a complete landscape of degnon loss for oncogenic fusions.

Fig. 6 Systematic analysis of C-terminal degron loss. **a** Conceptual diagram showing that fusions resulting in C-terminal degron loss were identified by the deepDegron method. **b** Top fusions enriched for C-terminal degron loss. X-axis represents the combined *p* value for both degron loss and cancer type-specificity (Fisher's method), while the Y-axis represents the effect size for degron loss. *P* values were calculated using a one-sided permutation test. See "Methods" for additional details on calculation of *p* values. **c** Diagram of an EGFR-SEPT14 fusion leading to loss of the C-terminal -GA* degron. **d** Removal of C-terminal degron (-GA*) lead to less ubiquitination of EGFR. WT and mutant HA-tagged EGFR proteins were expressed with His-Ub in HEK293T cells and treated with 10 μM MG132 for 12 h. Cells were lysed under denaturing conditions and His-Ub-conjugated proteins were pulled down with Ni-NTA-resin, washed, and immunoblotted for indicated proteins. **e** Diagram of an RAF1-AGGF1 fusion leading to loss of the C-terminal -Vx* degron. **f** Removal of C-terminal degron (-Vx*) lead to less ubiquitination of RAF1. WT and mutant HA-RAF1 were expressed with His-Ub in HEK293T cells and subsequently treated and pulled down as described in **d**. **g, h** Removal of C-terminal led to extended protein half-life of RAF1 in cycloheximide (CHX) assay. WT and mutant HA-RAF1 proteins were expressed in HEK293T cells and treated with CHX for indicated time. Cells were harvested, lysed, and immunoblotted for indicated proteins (**g**) and quantification (**h**). **i** Immunoblot analysis for HeLa stable cell lines that express either WT RAF1 or RAF1 mutants without -Vx* degron. **j, k** The stable cells expressing the RAF1 mutants without -Vx* degron had greater clonogenicity than WT RAF1-expressing cells in vitro. The stable cell lines as in Fig. 6i were subjected for colony formation assay (**j**) and statistical analysis (**k**). *n* = 3/group. Two independent experiments were performed. **l, m** RAF1 mutant-expressing cells developed larger tumors than WT RAF1-expressing cells in vivo. The stable cell lines as in **i** were subjected for mouse xenograft assay (**l**) and statistical analysis (**m**). *n* = 9 mice for the RAF1-WT group and *n* = 10 mice for the RAF1-mutant group. Data are presented as mean ± SD. ****p* < 0.01; by two-tailed Student's *t*-test. Two independent experiments were performed for **d, f**, and **g**. The relevant raw data and uncropped blots are provided in Source Data.

Our finding that fusion proteins preferentially escape protein degradation by degron loss suggests that tumors may be particularly sensitive to degradation of oncogenic fusions. Indeed, the standard of care for APL harboring the PML-RARA fusion is either all-*trans*-retinoic acid (ATRA) or arsenic trioxide, both of which lead to the degradation of the PML-RARA fusion protein^{61–65}. Given recent advance in the development of compounds that induce targeted protein degradation such as PROTACs (PROteolysis TARgeting Chimeras)⁷², other fusions besides PML-RARA that undergo degron loss could become efficacious therapeutic targets. Notably, compounds that specifically degrade the BCR-ABL and ALK fusions protein have been developed^{73–75}. An additional theoretical benefit of degrading fusion proteins is the possibility to overcome acquired resistance mutations to previously used inhibitors, such as imatinib for BCR-ABL^{76,77} and crizotinib for EML4-ALK fusions⁷⁸. Because not all gene fusions undergo degron loss, our analysis may help prioritize the most promising targets for further PROTAC drug development. However, there are numerous questions that deserve further attention. For example, how can we understand the combinatorial impact of degron loss with other simultaneous mechanisms involved in gene fusions? Are there differences in the functional consequences of pharmacological inhibition versus degradation of stable fusion proteins? Could induced degradation of otherwise stable fusion proteins increase the presentation of neoantigens that yield an immune response against cancer? Future studies of gene fusions that combine mechanistic and bioinformatic insights may reveal the answers to these and more questions.

Methods

Human cell lines and culture conditions. Human embryonic kidney 293 (HEK293), HEK293T, HeLa, DU145, and LNCaP cells were purchased from American Type Culture Collection (ATCC). *Spop*^{+/+} and *Spop*^{-/-} MEFs were kind gifts from Dr. Nicholas Mitsiades (Baylor College of Medicine). The panel of colon cancer cell lines (Lim2405, RKO, DiFi, SW480, Lim1215, LoVo, LS411N, SW1463, SW48, SNU-C2B, HCT8, and SW837) were obtained from Dr. Lin Zhang (University of Pittsburgh), and HCT116-FBW7-KO, HCT116 WT, and DLD1-FBW7-KO, DLD1-WT cell lines were kind gifts from Dr. Bert Vogelstein (John Hopkins University). HEK293, HEK293T, HeLa cells, *Spop*^{+/+}, and *Spop*^{-/-} MEFs were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 100 units of penicillin and 100 μg/ml streptomycin. DU145, LNCaP, HCT116, DLD1, Lim2405, RKO, DiFi, SW480, Lim1215, LoVo, LS411N, SW1463, SW48, SNU-C2B, HCT8, SW837, HCT116-FBW7-KO, and DLD1-FBW7-KO cells were cultured in RPMI1640 containing 10% FBS, 100 Units of penicillin and 100 μg/ml streptomycin.

General cloning. Expression vectors HA-ABL1, HA-FBW7, and HA-RAF1 were constructed by cloning the corresponding cDNAs into pcDNA3-HA vector. Flag-SPOP, Flag-SPOP-Y87C, Flag-SPOP-F102C, and Flag-SPOP-W131G were constructed

as previous described¹⁷. Myc-β-TRCP1 was constructed as previous describe⁷⁹. GFP-CCDC6 (571577), GFP-CCDC6/RET (572024), and HA-EGFR (703594) were purchased from MRC PPU (University of Dundee). HA-ABL1-ΔD, Flag-PML-S518A, Flag-PML-4A, Flag-PML-5A, HA-EGFR-G1029*, HA-EGFR-I1028*, HA-RAF1-P646A647*, and HA-RAF1-A647A648* were constructed using the Site-Directed Mutagenesis Kit (Agilent) following the manufacturer's instructions. GST-PML-WT, GST-PML-4A, and GST-PML-S518A were constructed by cloning the corresponding cDNA into pGEX-GST-4T1 vector. pLenti-HA-CCDC6, pLenti-HA-CCDC6-RET, pLenti-HA-RAF1, pLenti-HA-RAF1-P646A647*, and pLenti-HA-RAF1-A647A648* were constructed by cloning the corresponding cDNAs into pLenti-puro vector. The primers for site mutation are as below: PML-S518A-f: 5'-GCACCTCCAAGGCCAGT CGCACACCCACCTGG-3'; PML-S518A-r: 5'-CCAGGTGGGGTGGTGGCGACT GCCTTGGAGGTGC-3'; PML-4A-f: 5'-CGCGTGTGGTATCGCCGCCGCGGAA GAGCCAGATGCGCGAAAACCTCG-3'; PML-4A-r: 5'-CGAGTTTTCGGCATCTCGG TCTTCCGCGCGCGGATCCACACAACCGC-3'; ABL1-ΔD-f: 5'-GCAATCCAA GAAGGGGAGCTGTTATCTGGAAG-3'; ABL1-ΔD-r: 5'-CTTCCAGATAACAGC TCCCTTCTTGATTTCG-3'; EGFR-G1029-f: 5'-CAGTGAATTTATTGGATGAG CGGCGCTTACC-3'; EGFR-G1029-r: 5'-GGTAAGCGCGGCTCATCCAATAAA TTCACTG-3'; EGFR-I1028-f: 5'-CAAAGCAGTGAATTTATTGGAGCGCGCCGT- TACCC-3'; EGFR-I1028-r: 5'-GGTAAGCGCGGCTCAAATAAATTCAGTCTT- TG-3'; RAF1-A647A648-f: 5'-CCCCGAGGTCCTATGTTCTAGTTGACTTTG- CACC-3'; RAF1-A647A648-r: 5'-GGTGAAGTCAACTAGAACATAGGCAGCC TCGGGG-3'; RAF1-P646A647-f: 5'-CCCCGAGGTCCTTTCTAGTTGACTTTG- CACTG-3'; RAF1-P646A647-r: 5'-CAGGTGCAAAGTCAACTAGAAAGGCAGCC TCGGGG-3'. The shRNA vectors for SPOP were purchased from Sigma (TRCN000122224, TRCN0001139181, TRCN000145024).

Antibodies. The anti-ABL1 (2862, 1:1000), anti-p27 (3686, 1:1000), anti-DEK (13962, 1:1000), anti-SRC3 (2126, 1:1000), anti-CUL3 (2759, 1:1000), anti-GST (2625, 1:2000), anti-β-TRCP (4394, 1:1000), anti-p-ERK(9101, 1:1000), and anti-ERK (4695, 1:1000) antibodies were obtained from Cell Signaling Technology. Anti-ERG (EPR3864, 1:1000) antibody was obtained from Abcam. Anti-SPOP (16750-1-AP, 1:1000) antibody was obtained from Proteintech. Anti-GFP (A-11122, 1:5000) antibody was obtained from Thermo Fisher. Anti-FBW7 (A301-720A, 1:1000) and anti-PML (A301-167A, 1:1000) were obtained from Bethyl Laboratories. Anti-CCDC6 (sc-100309, 1:1000) and anti-α Tubulin (sc-8035, 1:2000) antibodies were obtained from Santa Cruz Biotechnology. Mouse monoclonal anti-HA.11 epitope tag (clone 16B12, 901513, 1:1000) was obtained from BioLegend. Anti-Vinculin (V9131, 1:50000), rabbit polyclonal anti-HA (H6908, 1:3000), mouse monoclonal ANTI-FLAG® M2 (F3165, 1:5000), rabbit polyclonal ANTI-FLAG® (F7425, 1:3000), anti-mouse IgG (whole molecule)-peroxidase (A4416, 1:5000), and anti-rabbit IgG (whole molecule)-peroxidase (A4914, 1:5000) were obtained from Sigma-Aldrich. Mouse monoclonal ANTI-FLAG® M2 affinity agarose gel (A2220) and mouse monoclonal anti-HA-agarose (A2095) were obtained from Sigma-Aldrich.

Annotation of fusion consequence. We annotated the protein sequence consequence of fusions using the software tool AGFusion⁸⁰. To provide consistent annotation, we chose the Matched Annotation from NCBI and EMBL-EBI (MANE select transcripts v0.9) from GENCODE when possible⁸¹, or otherwise the longest transcript that is consistent with the fusion junction. Transcript annotations were based on Ensembl release 95 using pyensembl (<https://github.com/openvax/pyensembl>). Of 25,664 fusions reported in TCGA, 24,239 fusions could be annotated. The PFAM database was used to annotate the impact of fusions on protein domains⁸². Code used to analyze fusion genes can be found on github (https://github.com/ctokheim/fusion_pipeline).

Annotation of cancer driver genes. A consensus among multiple sources was used to annotate previously implicated cancer driver genes, which included OncoKB (<https://www.oncokb.org/>), downloaded 4/2020⁸³, The Cancer Genome Atlas (TCGA)⁸⁴, and the Cancer Gene Census (CGC), downloaded 4/9/2020⁸⁵. For CGC, we excluded genes with only support for germline mutations. For OncoKB, we only used genes that were annotated by OncoKB, rather than including additional genes from other sources. To further distinguish oncogenes versus tumor suppressor genes, we annotated based on the evidence from at least one source and no conflicting interpretations. Given that TCGA has cancer type-specific assessments of oncogene and tumor suppressor genes, we chose based on the most frequent annotation across cancer types.

Enrichment for in-frame fusions. To analyze whether fusions containing driver genes are biased towards in-frame fusions, we analyzed the odds ratio of in-frame vs out-of-frame fusions. The in-frame status of fusions was determined by the annotation software AGFusion (<https://github.com/murphyjcj/AGFusion>)⁸⁰. The log odds ratio was calculated separately for oncogene and tumor suppressor gene fusions, relative to putative passenger fusions that do not contain a gene previously implicated in cancer. In cases where a fusion is composed of both an oncogene and a tumor suppressor, the fusion gene was regarded as an oncogene. The standard error for the log odds ratio was calculated using a normal approximation⁸⁶.

Protein domain analysis. To analyze whether putatively oncogenic fusions preferentially retain protein domains, we compared the odds ratio that a fusion retained at least one protein domain for implicated driver genes (oncogenes or tumor suppressors) to passenger genes. We used domains from PFAM to annotate whether fusion retained or lost protein domains. Protein domains needed to be at least 25 amino acids long. For cases where the fusion junction interrupted a protein domain, we considered a protein domain as retained in the fusion gene if greater than 50% of the protein sequence was included.

Motif search for internal degrons. We first curated known degn motifs from eukaryotic linear motifs (ELM) database and other literatures²⁵ (Supplementary Data 7). Each motif is represented as a regular expression which describes the allowable amino acid residues at each position. Motifs were then searched against the protein translation of GENCODE transcripts using the python “re” package. When multiple transcripts were available for a gene, the MANE select transcript (v0.9) was used. Some degn motifs require not only a particular protein sequence, but also that certain residues have appropriate post-translational modifications (PTM). Towards this end, we collected all available PTMs in the PhosphoSitePlus database⁷⁴ and filtered motif sequence matches for any requisite PTMs (phosphorylation or acetylation). For the non-standard BTRC degn, we used the regular expression “SSxxS”. The motif search revealed 32,804 hits across 8623 genes involved in TCGA fusions.

Machine learning prioritization of internal degn motifs. Because motif instances may happen by chance in the proteome, we wanted further prioritize motifs that are a biologically plausible degn. Previously, we developed a model to predict the potential of a motif to be a degn using a Random Forest algorithm. The model was trained on 83 features from the SNVBox database^{87,88} to distinguish previously reported degrons ($n = 186$)²⁶ from random other sequences within the same set of proteins ($n = 186$). Features spanned characterization of evolutionary conservation to biophysical features of amino acid residues within a protein. To summarize features across the multiple amino acid residues in a motif, we took the average of each feature. Evaluated using 20-fold cross-validation, performance as measured by the area under the receiver operating characteristic curve (auROC) was 0.8 out of 1.0 ($p = 2 \times 10^{-25}$, Mann–Whitney U test).

Internal degn motif filtering. Because degn motifs are generally short, motif matches can happen by chance across the proteome. We therefore filtered motifs that had low potential to actually being a degn according to a Random Forest algorithm (score ≤ 0.6 out of 1.0), see section entitled “Machine learning prioritization of internal degn motifs”. This resulted in keeping 2485 high-likelihood degn motifs for downstream analysis.

C-terminal degn motif. In contrast to internal degn motifs from the ELM database, C-terminal degn motifs were defined based on de novo inference from the Global Protein Stability (GPS) assay³¹ as previously described³⁰. Briefly, the c-terminal sequence of every protein in the proteome is ranked by a degn potential score by the deepDegron method. A binomial model is then used to test for motifs that are statistically enriched in high scoring sequences ($q < 0.05$). This revealed 236 C-terminal degn motifs. Note that C-terminal degn motifs may partially overlap so multiple motif matches in a protein sequence are regarded as the same as a single motif match. All C-terminal degn motifs can also be found in Supplementary Data 8. Documentation for deepDegron is available on readthedocs (<https://deepdegn.readthedocs.io/>) and source code is available on github (<https://github.com/ctokheim/deepDegron>).

Statistical test for degn loss in fusion genes. A permutation-based approach was used to determine whether fusions preferentially lead to degn loss. Since a gene may have different fusion partners that all lead to degn loss (e.g. ETV fusions, Fig. 1), we chose to measure enrichment separately for 5' and 3' genes. For internal degrons, each gene involved in a fusion received a degn loss score, representing the sum of scores for degrons lost in the fusion. The degn loss score represents both the confidence that the degn exists and the frequency by which it is lost in fusion events. Basically, each predicted degn in a protein sequence receives a score from a Random Forest machine learning model that reflects the confidence in the prediction. The score of that degn is then summed each time a fusion event leads to its loss. Likewise, for c-terminal degn analysis, each 5' gene involved in a fusion received a delta degn potential score, representing the difference in degn potential scores between the 5' gene and 3' gene of a fusion. Only fusions resulting in in-frame fusions were analyzed, as the loss of degn in an out-of-frame fusion would not lead to increased activity of the fusion product. Additionally, as the previously reported validation rate of fusion calls is 63%¹¹, we only analyzed genes involved in at least two fusions to mitigate the impact of spurious calls. The sum of degn loss scores for a gene across multiple fusions was then calculated as the test statistic. The observed scores were then compared to 10,000 permutations, where degn loss scores per fusion were randomly shuffled and the gene-based test statistic was recalculated. The p value for a gene's observed test statistic is calculated as the fraction of permutations that have an equal or greater test statistic. Genes were regarded as statistically significant based on the false discovery rate ($q < 0.1$) using the Benjamini–Hochberg method³². Given that oncogene fusions display a significant bias towards in-frame mutations, we only considered genes as degn loss candidates if they had at least 50% of fusions as in-frame. To further prioritize oncogene fusions that have degn loss in the oncogene itself rather than the partner gene, we also included analyses of only genes with a retained protein domain and a restricted hypothesis test analyzing only previously implicated oncogenes.

Statistical test for cancer type-specificity of fusion genes. Similar to the statistical test for degn loss, we also used a permutation test to evaluate whether genes were involved in fusions preferentially found in particular cancer types. To quantify cancer type-specificity, we used entropy,

$$h_g = -\sum_{c \in C_g} p_c \log_2 p_c \quad (1)$$

where h_g is the entropy for gene g , c reflects a particular cancer type, C_g reflects all cancer types with fusions containing gene g , and p_c reflects the fraction of fusions for gene g found in cancer type c . Lower entropy values represent higher cancer type-specificity. We randomly shuffled the labels for cancer types of the fusions 10,000 times, and recomputed h_g^i . The corresponding p value was calculated as the fraction of permutations i that had an entropy equal to or lower than the observed entropy. Genes were regarded as statistically significant based on the false discovery rate ($q < 0.1$) using the Benjamini–Hochberg method³².

Lollipop diagrams. Lollipop diagrams displaying fusion genes were generated using ProteinPaint (<https://pecan.stjude.cloud/proteinpaint>)⁸⁹. Fusion junctions were submitted according to their genomic coordinates. Protein domains are shown as colored boxes along the protein sequence.

Immunoblots and immunoprecipitation (IP). Cells were lysed in EBC buffer (50 mM Tris pH 7.5, 120 mM NaCl, 0.5% NP-40) supplemented with protease inhibitors (Thermo Fisher) and phosphatase inhibitors (phosphatase inhibitor cocktail set I and II, Calbiochem). The lysates were then resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotted with indicated antibodies. For IP, 0.5–1 mg lysates were incubated with the appropriate beads for 4 h at 4 °C. Immuno-complexes were washed four times with NETN buffer (20 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40) before being resolved by SDS-PAGE and immunoblotted for indicated proteins. These primary antibodies were diluted in 5% BSA in TBST and secondary antibodies were diluted in 5% non-fat milk for immunoblotting analysis. The Quantity One software was used for the quantification of protein band intensity, and graphic and statistical analyses were generated using GraphPad 8.

In vitro kinase assays. PML in vitro kinase assays were performed as previous reported⁹⁰. Briefly, GST-PML-WT, GST-PML-4A, and GST-PML-S518A were expressed in BL21 *E. coli* and purified using Glutathione Sepharose 4B according to the manufacturer's instructions (Thermo). One microgram of GST-PML-WT, or GST-PML-4A, or GST-PML-S518A protein were incubated with ³²P-ATP in the absence or presence of CKII kinase in kinase assay buffer (10 mM HEPES, pH 8.0, 10 mM MgCl₂, 1 mM dithiothreitol, 0.1 mM ATP). The reaction was initiated by the addition of 10× kinase assay buffer in a volume of 30 μL for 45 min at 30 °C followed by the addition of SDS-PAGE sample buffer to stop the reaction before resolved by SDS-PAGE.

In vivo ubiquitination assays. Denatured in vivo ubiquitination assays were performed as previously described¹⁷. Briefly, HEK293T cells were transfected with indicated constructs. Forty-eight hours after transfection, 30 μ M MG132 was added to block proteasome degradation for 6 h and then cells were harvested in denatured buffer (6 M guanidine-HCl, pH 8.0, 0.1 M Na₂HPO₄/NaH₂PO₄, 10 mM imidazole). After sonication, the ubiquitinated proteins were purified by incubation with Ni-NTA matrices for 3 h at room temperature. The pull-down products were washed sequentially twice in buffer A, twice in buffer A/TI mixture (buffer A: buffer TI = 1:3, v/v) and once in buffer TI (25 mM Tris-HCl, pH 6.8, 20 mM imidazole). The poly-ubiquitinated proteins were separated by SDS-PAGE for immunoblot analyses.

Protein half-life cycloheximide (CHX) chasing assays. To measure the half-life of ABL1 protein, a CHX-based assay was performed following our previously described experimental procedures⁹⁰. Briefly, cells were treated with 200 μ g/ml CHX for indicated time before harvest for immunoblot analysis of indicated proteins.

Colony formation assays. Stable cell lines were seeded into six-well plates in medium (1,000 cells/well) and cultured for 2–3 weeks until colonies are visible. Then, the colonies were washed once with PBS, fixed with fixation buffer (10% acetic acid, 10% methanol) for 20 min, and then stained with staining solution (0.4% crystal violet, 20% ethanol) for 10 min. After staining, the plates were washed with distilled water and air-dried, and then colonies were counted for statistical analysis.

Mouse xenograft assays. Five- to six-week-old male nude mice were purchased from Taconic (#NCRNU) for xenograft studies. A total of 1 \times 10⁶ cells were resuspended in 100 μ l PBS solution and injected subcutaneously into the mice (n = 9 or 10 mice for each group) as described previously⁷⁹. At the end of experiment, mice were sacrificed and tumors were dissected for imaging and weighing. All mouse experiments were approved by the Institutional Animal Care and Use Committee (IACUC, RN150D) at Beth Israel Deaconess Medical Center (BIDMC). The Institute is committed to the highest ethical standards of care for animals used for the purpose of continued progress in the field of human cancer research. All mice were housed in a pathogen-free environment at BIDMC animal facility and were handled in strict accordance with the “Guide for the Care and Use of Laboratory Animals” and the applicable institutional regulations.

Association of CCDC6-RET with leukocyte fraction. To analyze whether CCDC6-RET fusions were associated with leukocyte infiltration, we utilized a previous estimate of immune infiltration for TCGA tumors⁹¹. A likelihood ratio test was performed after adjusting for tumor purity from ABSOLUTE (downloaded from <https://gdc.cancer.gov/about-data/publications/pancanatlas>)⁹², tumor mutation burden, and cancer type.

Association of fusions events with protein abundance (RPPA). To analyze whether fusion events were associated with an altered proteome, we correlated the mutation status of fusion genes with protein abundance from reverse phase protein array (RPPA) in TCGA⁹³. A Wald test was performed after adjustment for cancer type. Only fusions present in at least three tumors were considered.

Association of fusions events with transcription factor activity. We hypothesized that fusion events may be associated with altered activity of transcription factors. To quantify activity, we leveraged thousands of transcription factor ChIP-seq profiles in Cistrome DB to identify target genes³³. Computational analysis was then carried out as previously performed³⁰. Briefly, we first analyzed fusion events for differentially expressed genes, after adjusting for tumor purity and tumor subtype. RABIT³⁴ was then used to infer the transcription factor regulators that explain the differentially expressed genes by using the transcription factor target genes defined by Cistrome DB. Associations were regarded as significant at a family wise error rate of 0.01. Analysis only considered fusions with at least three events in a cancer type and transcription factors not deemed to be an outlier (see below).

Defining outlier transcription factors. ChIP-seq data defining the target genes of transcription factors can be of inconsistent quality. We reasoned that ChIP-seq datasets that consistently arise as explaining differentially expressed genes for nearly all fusion events may reflect data artifacts. We therefore performed outlier analysis using robust covariance estimation (scikit learn python package)⁹⁴, assuming a gaussian distribution and a significant contamination rate of 0.01 (Supplementary Fig. 3f).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data are available in the article, Supplementary Information, or Supplementary Data 1–8. The full list of recurrent genetic fusions, full list of genes and oncogenes with internal and

C-terminal degron loss, full list of protein abundance of fused genes, full list of downstream transcription factors due to genetic fusions are included in the Supplementary Data. The original gene fusion calls were obtained from Supplementary Data 1 of Gao et al.¹¹. The subsequently annotated and processed gene fusion data for downstream statistical analysis is available on GitHub (https://github.com/ctokheim/fusion_pipeline). The output from the analysis can be found in the Supplementary Data. All data used in the analyses described in this study are freely available within the public database, including TCGA (<https://www.cancer.gov/tcga>), OncoKB (<https://www.oncoKB.org/>), CGC (<https://cancer.sanger.ac.uk/census>), Uniprot (<https://www.uniprot.org/>), PFAM (<http://pfam.xfam.org/>), ELM (<http://elm.eu.org/>), and PhosphoSitePlus (<https://www.phosphosite.org/>). Source data are provided with this paper.

Code availability

Custom code for this manuscript is available on GitHub (https://github.com/ctokheim/fusion_pipeline) and is archived on Zenodo⁹⁵. The README file in the GitHub repository describes how to reproduce the analysis. The code uses python 3 and exact version numbers of dependencies are listed in the environment configuration file. The deepDegron code to analyze c-terminal degrons is also freely available on GitHub (<https://github.com/ctokheim/deepDegron>).

Received: 30 January 2021; Accepted: 27 October 2021;

Published online: 18 November 2021

References

- Vogelstein, B. & Kinzler, K. W. The multistep nature of cancer. *Trends Genet.* **9**, 138–141 (1993).
- Knudson, A. G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* **68**, 820–823 (1971).
- Nowell, P. & Hungerford, D. A minute chromosome in chronic granulocytic leukemia. *Science* **132**, 1488–1501 (1960).
- Lugo, T. G., Pendergast, A.-M., Muller, A. J. & Witte, O. N. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science* **247**, 1079–1082 (1990).
- Ren, R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat. Rev. Cancer* **5**, 172–183 (2005).
- Druker, B. J. et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344**, 1031–1037 (2001).
- Kumar-Sinha, C., Tomlins, S. A. & Chinnaiyan, A. M. Recurrent gene fusions in prostate cancer. *Nat. Rev. Cancer* **8**, 497–511 (2008).
- Lipson, D. et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat. Med.* **18**, 382–384 (2012).
- Takeuchi, K. et al. RET, ROS1 and ALK fusions in lung cancer. *Nat. Med.* **18**, 378–381 (2012).
- Wu, Y. M. et al. Identification of targetable FGFR gene fusions in diverse cancers. *Cancer Discov.* **3**, 636–647 (2013).
- Gao, Q. et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* **23**, 227–238 e223 (2018).
- Tuna, M., Amos, C. I. & Mills, G. B. Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. *Oncotarget* **10**, 2095–2111 (2019).
- Bastus, N. C. et al. Androgen-induced TMPRSS2:ERG fusion in nonmalignant prostate epithelial cells. *Cancer Res.* **70**, 9544–9548 (2010).
- Soda, M. et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
- Palanisamy, N. et al. Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat. Med.* **16**, 793–798 (2010).
- Vitari, A. C. et al. COP1 is a tumour suppressor that causes degradation of ETS transcription factors. *Nature* **474**, 403–406 (2011).
- Gan, W. et al. SPOP promotes ubiquitination and degradation of the ERG oncoprotein to suppress prostate cancer progression. *Mol. Cell* **59**, 917–930 (2015).
- An, J. et al. Truncated ERG oncoproteins from TMPRSS2-ERG fusions are resistant to SPOP-mediated proteasome degradation. *Mol. Cell* **59**, 904–916 (2015).
- Ciechanover, A. The ubiquitin-proteasome proteolytic pathway. *Cell* **79**, 13–21 (1994).
- Pohl, C. & Dikic, I. Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science* **366**, 818–822 (2019).
- Komander, D. & Rape, M. The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203–229 (2012).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Zhou, W., Wei, W. & Sun, Y. Genetically engineered mouse models for functional studies of SKP1-CUL1-F-box-protein (SCF) E3 ubiquitin ligases. *Cell Res.* **23**, 599–619 (2013).

24. Bernassola, F., Karin, M., Ciechanover, A. & Melino, G. The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell* **14**, 10–21 (2008).
25. Kumar, M. et al. ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* **48**, D296–D306 (2020).
26. Mészáros, B., Kumar, M., Gibson, T. J., Uyar, B. & Dosztányi, Z. Degrons in cancer. *Sci. Signal.* **10**, eaak9982 (2017).
27. Varshavsky, A. N-degron and C-degron pathways of protein degradation. *Proc. Natl Acad. Sci. USA* **116**, 358–366 (2019).
28. Park, S. E. et al. Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. *Science* **347**, 1249–1252 (2015).
29. Chen, S. J., Wu, X., Wadas, B., Oh, J. H. & Varshavsky, A. An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, eaal3655 (2017).
30. Tokheim, C. et al. Systematic characterization of mutations altering protein degradation in human cancers. *Mol. Cell* **81**, 1292–1308 (2021).
31. Koren, I. et al. The eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* **173**, 1622–1635 e1614 (2018).
32. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
33. Zheng, R. et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735 (2019).
34. Jiang, P., Freedman, M. L., Liu, J. S. & Liu, X. S. Inference of transcriptional regulation in cancers. *Proc. Natl Acad. Sci. USA* **112**, 7731–7736 (2015).
35. Yu, J. et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**, 443–454 (2010).
36. Inamura, K. et al. EML4-ALK lung cancers are characterized by rare other mutations, a TTF-1 cell lineage, an acinar histology, and young onset. *Mod. Pathol.* **22**, 508–515 (2009).
37. Godavarthy, P. S. et al. The vascular bone marrow niche influences outcome in chronic myeloid leukemia via the E-selectin–SCL/TAL1–CD44 axis. *Haematologica* **105**, 136–147 (2020).
38. Castro, F., Cardoso, A. P., Goncalves, R. M., Serre, K. & Oliveira, M. J. Interferon-gamma at the crossroads of tumor immune surveillance or evasion. *Front. Immunol.* **9**, 847 (2018).
39. Peggs, K. & Mackinnon, S. Imatinib mesylate—the new gold standard for treatment of chronic myeloid leukemia. *N. Engl. J. Med.* **348**, 1048–1050 (2003).
40. Schiffer, C. A. BCR-ABL tyrosine kinase inhibitors for chronic myelogenous leukemia. *N. Engl. J. Med.* **357**, 258–265 (2007).
41. Theurillat, J. P. et al. Prostate cancer. Ubiquitylome analysis identifies dysregulation of effector substrates in SPOP-mutant prostate cancer. *Science* **346**, 85–89 (2014).
42. Hantschel, O. et al. A myristoyl/phosphotyrosine switch regulates c-Abl. *Cell* **112**, 845–857 (2003).
43. Li, C. et al. Tumor-suppressor role for the SPOP ubiquitin ligase in signal-dependent proteolysis of the oncogenic co-activator SRC-3/AIB1. *Oncogene* **30**, 4350–4364 (2011).
44. Barbieri, C. E. et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
45. Grieco, M. et al. PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas. *Cell* **60**, 557–563 (1990).
46. Cerrato, A., Visconti, R. & Celetti, A. The rationale for druggability of CCDC6-tyrosine kinase fusions in lung cancer. *Mol. Cancer* **17**, 46 (2018).
47. Seo, J. S. et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119 (2012).
48. Drechsler, M., Hildebrandt, B., Kundgen, A., Germing, U. & Royer-Pokora, B. Fusion of H4/D10S170 to PDGFRbeta in a patient with chronic myelomonocytic leukemia and long-term responsiveness to imatinib. *Ann. Hematol.* **86**, 353–354 (2007).
49. Thompson, B. J. et al. The SCFFBW7 ubiquitin ligase complex as a tumor suppressor in T cell leukemia. *J. Exp. Med.* **204**, 1825–1835 (2007).
50. Yada, M. et al. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *EMBO J.* **23**, 2116–2125 (2004).
51. Wei, W., Jin, J., Schlisio, S., Harper, J. W. & Kaelin, W. G. Jr. The v-Jun point mutation allows c-Jun to escape GSK3-dependent recognition and destruction by the Fbw7 ubiquitin ligase. *Cancer Cell* **8**, 25–33 (2005).
52. Koepf, D. M. et al. Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science* **294**, 173–177 (2001).
53. Davis, R. J., Welcker, B. & Clurman, B. E. Tumor suppression by the Fbw7 ubiquitin ligase: mechanisms and opportunities. *Cancer Cell* **26**, 455–464 (2014).
54. Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–520 (2015).
55. Tong, J., Tan, S., Zou, F., Yu, J. & Zhang, L. FBW7 mutations mediate resistance of colorectal cancer to targeted therapies by blocking Mcl-1 degradation. *Oncogene* **36**, 787–796 (2017).
56. Zhao, J., Tang, J., Men, W. & Ren, K. FBXW7-mediated degradation of CCDC6 is impaired by ATM during DNA damage response in lung cancer cells. *FEBS Lett.* **586**, 4257–4263 (2012).
57. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
58. Zhong, J., Ogura, K., Wang, Z. & Inuzuka, H. Degradation of the transcription factor Twist, an oncoprotein that promotes cancer metastasis. *Discov. Med* **15**, 7–15 (2013).
59. Shirogane, T., Jin, J., Ang, X. L. & Harper, J. W. SCFBeta-TRCP controls clock-dependent transcription via casein kinase 1-dependent degradation of the mammalian period-1 (Per1) protein. *J. Biol. Chem.* **280**, 26863–26872 (2005).
60. Ma, Y. et al. SCFBeta-TrCP ubiquitinates CHK1 in an AMPK-dependent manner in response to glucose deprivation. *Mol. Oncol.* **13**, 307–321 (2019).
61. Shen, Z. X. et al. All-trans retinoic acid/As2O3 combination yields a high quality remission and survival in newly diagnosed acute promyelocytic leukemia. *Proc. Natl Acad. Sci. USA* **101**, 5328–5335 (2004).
62. Zhu, J. et al. Retinoic acid induces proteasome-dependent degradation of retinoic acid receptor alpha (RARalpha) and oncogenic RARalpha fusion proteins. *Proc. Natl Acad. Sci. USA* **96**, 14807–14812 (1999).
63. Zhu, J. et al. Arsenic-induced PML targeting onto nuclear bodies: implications for the treatment of acute promyelocytic leukemia. *Proc. Natl Acad. Sci. USA* **94**, 3978–3983 (1997).
64. Yoshida, H. et al. Accelerated degradation of PML-retinoic acid receptor alpha (PML-RARA) oncoprotein by all-trans-retinoic acid in acute promyelocytic leukemia: possible role of the proteasome pathway. *Cancer Res.* **56**, 2945–2948 (1996).
65. Raelson, J. V. et al. The PML/RAR alpha oncoprotein is a direct molecular target of retinoic acid in acute promyelocytic leukemia cells. *Blood* **88**, 2826–2832 (1996).
66. Kakizuka, A. et al. Chromosomal translocation (t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. *Cell* **66**, 663–674 (1991).
67. Stehmeier, P. & Muller, S. Phospho-regulated SUMO interaction modules connect the SUMO system to CK2 signaling. *Mol. Cell* **33**, 400–409 (2009).
68. Percherancier, Y. et al. Role of SUMO in RNF4-mediated promyelocytic leukemia protein (PML) degradation: sumoylation of PML and phospho-switch control of its SUMO binding domain dissected in living cells. *J. Biol. Chem.* **284**, 16595–16608 (2009).
69. Scaglioni, P. P. et al. A CK2-dependent mechanism for degradation of the PML tumor suppressor. *Cell* **126**, 269–283 (2006).
70. Yoshihara, K. et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845–4854 (2015).
71. Lee, D. F. et al. KEAP1 E3 ligase-mediated downregulation of NF-kappaB signaling by targeting IKKbeta. *Mol. Cell* **36**, 131–140 (2009).
72. Schapira, M., Calabrese, M. F., Bullock, A. N. & Crews, C. M. Targeted protein degradation: expanding the toolbox. *Nat. Rev. Drug Discov.* **18**, 949–963 (2019).
73. Zhang, C. et al. Proteolysis targeting chimeras (PROTACs) of anaplastic lymphoma kinase (ALK). *Eur. J. Med. Chem.* **151**, 304–314 (2018).
74. Kong, X. et al. Drug discovery targeting anaplastic lymphoma kinase (ALK). *J. Med. Chem.* **62**, 10927–10954 (2019).
75. Tong, B. et al. A nimbolide-based kinase degrader preferentially degrades oncogenic BCR-ABL. *ACS Chem. Biol.* **15**, 1788–1794 (2020).
76. Yang, Y. et al. Global PROTAC Toolbox for degrading BCR-ABL overcomes drug-resistant mutants and adverse effects. *J. Med. Chem.* **63**, 8567–8583 (2020).
77. Burslem, G. M. et al. Targeting BCR-ABL1 in chronic myeloid leukemia by PROTAC-mediated targeted protein degradation. *Cancer Res.* **79**, 4744–4753 (2019).
78. Choi, Y. L. et al. EML4-ALK mutations in lung cancer that confer resistance to ALK inhibitors. *N. Engl. J. Med.* **363**, 1734–1739 (2010).
79. Inuzuka, H. et al. Phosphorylation by casein kinase I promotes the turnover of the Mdm2 oncoprotein via the SCF(beta-TRCP) ubiquitin ligase. *Cancer Cell* **18**, 147–159 (2010).
80. Murphy, C. & Elemento, O. AGFusion: annotate and visualize gene fusions. Preprint at <https://www.biorxiv.org/content/10.1101/080903v1> (2016).
81. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
82. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
83. Chakravarty, D. et al. OncoKB: a Precision Oncology Knowledge Base. *JCO Precis Oncol.* **2017**, <https://doi.org/10.1200/PO.17.00011> (2017).
84. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 e318 (2018).
85. Sondka, Z. et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
86. Bland, J. M. & Altman, D. G. Statistics notes: the odds ratio. *BMJ* **320**, 1468 (2000).

87. Wong, W. C. et al. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics* **27**, 2147–2148 (2011).
88. UniProt-Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
89. Zhou, X. et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet.* **48**, 4–6 (2016).
90. Inuzuka, H. et al. SCF FBW7 regulates cellular apoptosis by targeting MCL1 for ubiquitylation and destruction. *Nature* **471**, 104 (2011).
91. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830 e814 (2018).
92. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
93. Li, J. et al. TCPA: a resource for cancer functional proteomics data. *Nat. Methods* **10**, 1046–1047 (2013).
94. Rousseeuw, P. J. & Driessen, K. V. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223 (1999).
95. Tokheim, C. Genetic fusions favor tumorigenesis through degron loss in oncogenes, fusion_pipeline, <https://doi.org/10.5281/zenodo.5565550> (2021).

Acknowledgements

We thank the Liu and Wei lab members for suggestions and comments on this work. This work was supported by R35CA253027 (to W.W.) and Breast Cancer Research Foundation BCRF-20-100 (to X.S.L.). C.T. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRQ-04-20).

Author contributions

The idea was conceived by W.W., X.S.L. and P.P.P.; C.T. designed and performed the bioinformatics analysis; J.L., W.G. and B.J.N. designed and performed most of the experiments with assistance from Y.L.; J.L. and C.T. wrote the manuscript. W.W., X.S.L. and P.P.P. supervised the study and edited the manuscript. All authors commented on the manuscript.

Competing interests

W.W. and P.P.P. are co-founders and stockholders of the Rekindle Therapeutics. X.S.L. is a cofounder, board member, SAB member, and consultant of GV20 Oncotherapy and its subsidiaries; stockholder of BMY, TMO, WBA, ABT, ABBV, and JNJ; and received research funding from Takeda, Sanofi, and Novartis. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26871-y>.

Correspondence and requests for materials should be addressed to X. Shirley Liu, Pier Paolo Pandolfi or Wenyi Wei.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021