Alternative splicing changes as drivers of cancer

Héctor Climente-González[1], Eduard Porta-Pardo[2], Adam Godzik[2], Eduardo Eyras[1,3]

[1]Pompeu Fabra University (UPF), E08003, Barcelona, Spain

[2]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, 92037, USA

[3]Catalan Institution of Research and Advanced Studies (ICREA), E08010, Barcelona, Spain

Contact information: eduardo.eyras@upf.edu

Summary

Alternative splicing changes are frequently observed in cancer and are starting to be recognized as important signatures for tumor progression and therapy. However, their functional impact and relevance to tumorigenesis remains mostly unknown. We carried out a systematic analysis to characterize the potential functional consequences of alternative splicing changes in thousands of tumor samples. This analysis reveals that a subset of alternative splicing changes affect protein domain families that are frequently mutated in tumors, potentially disrupt protein–protein interactions in cancer-related pathways, and are mutually exclusive with mutations in multiple cancer drivers. Moreover, there is a negative correlation between the number of these alternative splicing changes in a sample and the number of somatic mutations in drivers. We propose that a subset of the alternative splicing changes observed in tumors represents independent oncogenic processes and could potentially be considered alternative splicing drivers (AS-drivers).

Introduction

Alternative splicing provides the potential to generate diversity at RNA and protein levels from an apparently limited protein coding part of the genome (Yang et al., 2016). Besides being a critical mechanism during development, cell differentiation, and regulation of cell-type-specific functions (Norris and Calarco, 2012), it is also involved in multiple pathologies, including cancer

(Chabot and Shkreta, 2016). Many alternative splicing changes can essentially recapitulate cancer-associated phenotypes, for instance, by promoting angiogenesis (Vorlova *et al*, 2011), inducing cell proliferation (Yanagisawa et al., 2008), or avoiding apoptosis (Karni et al., 2007). Alternative splicing in tumors can appear as a consequence of somatic mutations that disrupt splicing regulatory motifs in exons and introns (Jung et al., 2015; Ward and Cooper, 2010), as well as through mutations or expression changes in core and auxiliary splicing factors, which impact the splicing of cancer-related genes (Alsafadi et al., 2016; Bechara et al., 2013; Madan et al., 2015; Sebestyén et al., 2016; Zong et al., 2014).

Alterations in alternative splicing are also emerging as relevant targets of therapy. This is the case with an exon-skipping event in *MET* observed in a number of lung cancer patients, resulting in a deletion of the protein region that inhibits its kinase catalytic activity (Kong-Beltran et al., 2006; Ma et al., 2003). Tumors that show an exon skipping in the proto-oncogene *MET* respond to *MET*-targeted therapies despite not having any other activating alteration in this gene (Frampton et al., 2015; Paik et al., 2015). Furthermore, alternative splicing is important in drug resistance. For instance, although an effective targeted treatment exists for patients with *BRAF* mutations in the kinase domain (Davies et al., 2002), a considerable number of non-responders express a *BRAF* isoform lacking exons 4–8, which encompass the RAS binding domain (Poulikakos et al., 2011). Small-molecule modulators of pre-mRNA splicing are capable of restoring the original *BRAF* splicing and reduce growth of therapy-resistant cells (Salton et al, 2015). Similarly, alternative splicing also impacts immunotherapy in cancer in relation to the aberrant activity of the splicing factor *SRSF3* (Sotillo et al., 2015). Thus, specific alterations in pre-mRNA splicing may provide a selective advantage in tumors and could potentially be direct targets of therapy. This also raises the question of whether splicing changes may act as cancer driver events.

Multiple studies have shown frequent splicing changes in tumors compared with normal tissues or during tumor progression and metastasis (Danan-Gotthold et al., 2015; Lu et al., 2015; Sebestyén et al., 2015; Trincado et al., 2016). However, the functional impact of these splicing changes and their possible role as drivers of cancer is not known yet. Alternative splicing changes can have diverse effects on the structure of the resulting protein and hence confer radical functional changes (Wang et al., 2005), remodel the network of protein–protein

interactions in a tissue-specific manner (Buljan et al., 2012; Ellis et al., 2012), and expand the protein interaction capabilities of genes (Yang et al., 2016). We hypothesized that a subset of splicing changes in tumors may trigger oncogenic mechanisms through the disruption of specific protein domains and protein–protein interactions.

Here we describe a systematic evaluation of the potential functional impact of recurrent alternative splicing changes observed in cancer samples. We describe splicing changes in terms of isoform switches in each tumor sample and determine the protein features and protein–protein interactions affected by them, and their relation to cancer drivers. Our analysis reveals a set of splicing isoform switches that affect protein domains from families frequently mutated in tumors, remodel the protein interaction network of cancer drivers, and occur in patients with no mutations in known cancer drivers. We propose that these isoform switches with driver-like properties, AS-drivers, play an important role in the neoplastic process independently of or in conjunction with existing mutations in cancer drivers.

Results

Isoform switches in cancer tend to reduce the protein coding potential

With the aim of defining potential alternative splicing drivers (AS-drivers) of cancer, we analyzed the expression of human transcript isoforms in 4,542 samples from 11 cancer types from TCGA to identify splicing alterations in each tumor sample. We used transcript isoforms, as they represent the endpoint of transcription and splicing and they ultimately determine the functional capacity of cells. For each gene and each patient sample we determined whether there was a switch between the most abundant transcript isoform in the normal samples and a different isoform that is the most abundant in that tumor sample, such that the change of relative abundance is higher than expected by the variability in normal samples and the gene shows no differential expression between tumor and normal samples. Additionally, we did not consider switches with a significant association with stromal or immune cell content, as we cannot be sure whether they are actually present in tumor cells (see Experimental Procedures for details).

We found a total of 8,979 different isoform switches in 6,978 genes that describe consistent changes in the transcriptome of the tumor samples and that would not be observable by simply measuring gene expression (Figure 1A and Table S1). Interestingly, tumor protein isoforms tend to be shorter than protein isoforms in normal tissues (Figure S1A). Moreover, while in most switches — 7,656 (85.27%) — both transcript isoforms code for protein, in the rest there is a significantly higher proportion of switches where only the normal transcript isoform is protein-coding (9.05% *vs.* 2.92%, binomial test p-value < 2.2e-16), suggesting that isoform switches in tumors are associated with the loss of protein coding capacity.

Isoform switches in cancer are frequently associated with protein feature losses

To determine the potential functional impact of the calculated isoform switches, we determined the protein features they affect (Experimental Procedures). Out of the 7,656 switches where both transcript isoforms code for proteins, 5,612 (73.30%) involve a change in at least one of the following functional features: Pfam domains, Prosite patterns, general disordered regions, disordered regions with potential to mediate protein–protein interactions, and protein loops (Figure 1B). We compared the switches that affect protein features with 100 sets of simulated switches, controlling for normal and tumor isoform expression. Remarkably, isoform switches in tumors have more protein feature losses than expected by chance (Fisher's exact test p-value < 2.84e-05, odds-ratio > 1.19), despite the fact that simulated switches also tend to have longer normal protein isoforms (Figure S1B). This indicates that isoform switches in cancer are strongly associated with the loss of protein function capabilities.

We focus on isoform switches that show a gain or loss in at least one protein feature, which we call *functional switches*, as they are likely to impact gene activity. There are 6,682 functional switches (Table S1), including 1070 for which only one of the isoforms codes for protein. Interestingly, functional switches are enriched in cancer drivers in some tumor types (Figure S1C). Among the top cancer driver genes with switches we identified a recurrent switch in *RAC1* (Figure 1C), which was linked before to metastasis (Zhou et al., 2012) and which we predict to gain an extra Ras family domain. We also found a recurrent switch in *TP53* that changes to a non-coding isoform and a switch in *ERBB2* that removes one of the receptor domains and does not coincide with those previously described (Jackson et al., 2013).

To characterize how functional switches may affect protein function, we calculated the enrichment of gain or loss for each domain families (Experimental Procedures). To ensure that this is attributed to a switch and not to the co-occurrence with another domain, we imposed a minimum of two distinct switches affecting each domain family. We detected 225 and 47 domain families exclusively lost or gained, respectively, and 16 that were both gained and lost, more frequently than expected by chance and in at least two different switches (Table S2). Functional categories of domain families with significant losses in switches include the regulation of protein activity and apoptosis (Figure 1D), suggesting effects on protein-protein interactions and cancer related pathways. To further characterize these functional switches, we calculated the proportions of cancer drivers annotated as oncogenes or tumor suppressors that contain domain families enriched in gains or losses using the reference proteome. Interestingly, domain families significantly gained in switches occur more frequently in oncogenes than in tumor suppressors (Wilcoxon test p-value = 9.55e-05), suggesting a similarity between isoform switches and oncogenic mechanisms in cancer.

Isoform switches and somatic mutations affect similar domain families

Isoform switches and somatic mutations are part of an intertwined continuum of alterations in cells that may be connected by a multitude of relationships. We conducted various comparisons using switches and *cis*-occurring mutations from whole exome and genome sequencing data, but could not find an association between them (Figures S2A-S2C) (see Supplemental Experimental Procedures). However, we observed that tumor samples with few genes with protein-affecting mutations (PAMs) tend to have many genes with functional switches, and vice versa, tumor samples with a low number of functional switches tend to have many genes with PAMs (Figure 2A). This suggests a complementarity between a domain change caused by a switch and protein affecting mutations. To investigate this, we calculated domain families enriched in protein affecting mutations (PAMs) (Experimental Procedures). We found that 76 domain families across 11 tumor types are enriched in mutations (Table S2). These domains enriched in mutations occur more frequently in cancer drivers compared to non-drivers (Wilcoxon test p-value < 2.2e-16), in agreement with recent analyses (Miller et al., 2015; Yang et al., 2015). When we compared the domain families enriched in somatic mutations with those enriched in gains or losses through switches, we found an overlap of 15 domain families, which is higher than expected by chance given the 5,307 domain families observed in a reference

proteome (Fisher's test p-value = 1.03e-05, odds ratio = 4.45). From the domain families enriched in mutations, 7 show only enrichment in losses, 6 show only enrichment in gains, and 2 show enrichment in both (Figure 2B) (Tables S2). Among the gains we find Cadherin domains related to switches in *CHD8, CDH26, FAT1, FAT2 and FAT3*. Among the losses, we find the Calcium-binding EGF domain, which is affected by various switches, including one in *NOTCH4*. A notable case involves the loss of the *TP53* DNA-binding domain and the *TP53* tetramerization motif. Although it only occurs in a single switch in *TP53*, its recurrence highlights the relevance of the alternative splicing of TP53 (Bourdon, 2007).

To further explore the similarity between the changes in functionality introduced by mutations and switches, we performed a GO-term enrichment analysis on the mutated and the switched domains separately using DcGO (Fang and Gough, 2013), and calculated the overlap between both set and compared it to the overlap obtained by randomly sampling the reference proteome. Notably, the observed overlap is higher than expected across the different ontologies and GO slim levels (Figure 2C). In particular, among the shared molecular functional categories, several are related to receptor activity and protein binding. This result supports the notion that switches and mutations affecting protein domains may impact similar functions in tumors. This also suggests switches that affect domains that are frequently mutated in tumors could be considered to have a relevant impact in cell function. A total of 855 functional switches in 701 genes (57 of them in 43 cancer drivers) affect domains that are enriched in mutations (Table S1) and represent potential AS-drivers.

Functional switches show mutual exclusion with driver mutations in cancer pathways

Another unmistakable sign that an alteration provides a selective advantage to the tumor is the mutual exclusion with other recurrent alterations in genes within the same pathway (Babur et al., 2015). We identified 254 functional switches that are mutually exclusive with somatic PAMs in three or more cancer drivers (Table S3). In fact, some switches tend to occur in patients that do not harbor mutations in known cancer drivers, i.e., pan-negative tumors (Saito et al., 2015). For instance, a switch in *PRX* shows mutual exclusion with PAMs in *BRAF*, *NRAS*, and *HRAS* in thyroid cancer and characterizes 14 (11,1%) of the pan-negative cases (Figure S2D). Moreover, 22 of these switches share a functional pathway with at least one cancer driver (Figure 2D).

These 22 switches include one in *PRDM1*, which shows mutual exclusion with mutations in *PTEN* in lung squamous cell carcinoma (LUSC) and *TP53* in lung adenocarcinoma (LUAD). These 22 functional switches that show mutual exclusion with drivers and share a functional pathway with one or more drivers may be indicative of alternative oncogenic processes and thus represent potential AS-drivers.

Isoform switches affect protein interactions with cancer drivers

Many of the frequently lost and gained domain families in functional switches are involved in protein binding activities, indicating a potential impact on protein–protein interactions (PPIs) in cancer. To analyze how our switches may affect the PPIs, we built a consensus PPI network with 8,142 nodes, each node representing a gene, using data from five different sources (Experimental Procedures) (Figure S3). To determine the effect of switches on the PPI network, we mapped PPIs to domain–domain interactions (DDIs) using a reference protein interaction network (Figure S4). From the 8,142 genes in the PPI network, 3,243 have at least one isoform switch, and for 1,896 isoform switches (in 1,488 genes) we were able to map at least one of their PPIs to a specific DDI. A total of 186 of these switches are located in 137 cancer drivers, with the remaining 1,710 in non-driver genes.

For each isoform switch, using the DDI information, we evaluated whether it would affect a PPI from the consensus network by matching the domains affected by the switch to the domains mediating the interaction, controlling for the expression of the isoforms predicted to be interaction partners. We found that 553 switches (29%) in 488 different genes affect domains that mediate protein interactions and likely affect such interactions. Most of these interaction-altering switches (n = 473, 85.5%) cause the loss of the domain that mediates the interaction, while a minority (n = 79, 14.3%) leads to a gain of the interacting domain. There is only one switch that leads to gain and loss of interactions with different partners. This is in *TAF9*, which loses a TIFIID domain and gains an AAA domain (Table S4).

Notably, switches in driver genes tend to alter PPIs more frequently than those in other genes, and they more frequently lose interactions (Figure 3A). From the 186 switches in drivers, 51 (27%) of them alter at least one interaction, either causing loss (42 switches) or gain (9 switches); with similar proportions in non-driver genes: 502 (29%) of them alter at least one interaction either by loss (431), gain (70) or both (1) (Figure S5A). Interestingly, switches that

affect domains from families enriched in mutations (Chi-square test p-value < 2.2e-16) or that show frequent mutual exclusion with cancer drivers (Chi-square test p-value = 1.42e-07, affect PPIs more frequently than other functional switches (Figure S5B). Notably, functional switches in genes annotated as direct interactors of drivers affect PPIs more frequently than the rest of functional switches (Fisher's exact test p-value<6.7e-16 OR>4.68 for all tests) (Figure 3B). Additionally, all functional pathways found enriched in PPI-affecting switches (Fisher's exact test corrected p-value < 0.05 and odds-ratio > 2) are related to cancer: *TP53* signaling, *WNT*, *RB1* and *MYC* pathways and the extrinsic pathway for apoptosis (Table S5), reinforcing the potential impact of isoform switches in cancer. We thus considered these 553 PPI-affecting switches as candidate AS-drivers.

Isoform switches remodel protein interaction networks in cancer

To further characterize the role of switches in remodeling the protein interaction network in cancer, we calculated modules in the PPI network (Blondel et al., 2008) using only interaction edges affected by switches (Experimental Procedures). This produced 197 modules involving 1584 genes (Table S6). Interestingly, one module is enriched in splicing factors (SFs) and RNA binding protein (RBP) genes (Module 9 in Table S6) and includes the cancer drivers *SF3B1*, *FUS*, *SYNCRIP*, *NUP98*, *EEF1A1* and *YBX1* (Figure 3C). The module contains a switch in *RBMX* that involves the skipping of two exons and the elimination of an RNA recognition motif (RRM) that would impact interactions with *SF3B1*, *EEF1A1* and multiple RBP genes (Figure 3C); and a switch in *TRA2B* that yields a non-coding transcript and would eliminate an interaction with *SF3B1* and multiple SFs. There is also a switch in *HNRNPC* that affects interactions with *SRSF12*; a switch in *TRA2A* linked to the gain of interactions with *CLK3* and *SRSF5*; and switches in *NXF1* and *RBMS2* that lose interactions with various SR protein coding genes and drivers *NUP98* and *SYNCRIP*. Consistent with a potential impact of switches in the regulation of RNA processing, the PPI-affecting switches show mutual exclusion with the cancer drivers (Figure 3D). Interestingly, this module also contains switches in the Importin genes *IPO11* and *IPO13*, which would affect interactions with the ubiquitin conjugating enzymes *UBE2E1*, *UBE2E3* and *UBE2I* (Figure 3C), and that show mutual exclusion across different tumor types (Figure 3D). These results indicate that the activity of RNA-processing factors may be altered through the disruption of their PPIs by alternative splicing.

We also found a module including multiple regulators of translation (module 43 in Table S6), with switches in *EIF4B*, *EIF3B* and *EIF4E* that affect interactions with the drivers *EIF4G1*, *EIF4A2* and *PABPC1* (Figure 3E). The switch in *EIF4B* causes the skipping of one exon, which we predict to eliminate an RRM domain and lose interactions with *EIF4G1* and *PABPC1*. The switch in *EIF3B* yields a non-coding transcript that loses multiple interactions. Although we did not predict any PPI change for the *EIF4E,* this switch loses eight predicted ANCHOR regions, suggesting a possible effect on other interactions. Besides frequent PAMs, *PABPC1* also present a functional switch that affects 2 disordered regions but does not affect any of the RRMs. In this case we did not predict any change in PPI and the possible functional impact remains to be discovered. These results, and the observed mutual exclusion between PAMs in *EIF4G1* and *PABPC1* and the identified PPI-affecting switches (Figure 3F), suggest that alternative splicing switches may impact translational regulation in tumors through the alteration of protein–protein interactions.

Isoform switches as drivers of cancer

Our results provide evidence that a subset of the alternative splicing switches, which we define as alternative splicing drivers, or AS-drivers, may be relevant for tumorigenesis. We define an AS-driver as a functional isoform switch that either (I) induces a gain or a loss of a protein domain from a family frequently mutated in cancer, (II) affects one or more PPIs, (III) displays mutual exclusion with drivers, or (IV) displays recurrence in cancer genomes beyond what is expected by chance. This definition yields 1875 potential AS-drivers (Figure 4A) (Table S1), with a large fraction of those affecting mutated domain families and/or PPIs (cases I and II from the list above, see Figure 4B).

The possible relevance of these AS-drivers varies across samples and tumor types. Considering tumor specific mutational drivers (Mut-drivers) and our set of AS-drivers, we labeled each patient as AS-driver–enriched or Mut-driver–enriched according to whether the proportion of switched AS-drivers or mutated Mut-drivers was higher, respectively. This partition of the samples indicates that, although Mut-drivers are predominant in patients for most tumors types, AS-drivers seem relevant for a considerable number of patients across most tumor types, and particularly for kidney and prostate tumors (Figure 4C). Additionally, regardless of the tumor type, patients with many mutations in Mut-drivers tend to show a low number of switched AS-

drivers, and vice versa (Figure 4D), bearing resemblance with the proposed cancer genome hyperbola between mutations and copy number variants (Ciriello et al., 2013). This supports the notion that AS-drivers represent alternative, yet-unexplored oncogenic mechanisms that could provide a complementary route to induce similar effects as genetic mutations.

Discussion

We have identified consistent and recurrent alternative splicing switches in tumors, which we call AS-drivers, that impact the function of affected proteins by adding or removing protein domains that are frequently mutated in cancer or by disrupting protein–protein interactions with cancer drivers or in cancer related pathways. Moreover, we observe that patients with AS-drivers tend not to harbor mutations in cancer drivers. We propose a model by which pathways often altered in cancer through somatic mutations may be affected in a similar way by AS-drivers in some patients, and in particular, in pan-negative patients. Recently, an alternative splicing change in *NFE2L2* has been described to lead to the loss of a protein domain and the interaction with its negative regulator *KEAP1*, thereby providing an alternative mechanism for the activation of an oncogenic pathway (Goldstein et al., 2016). This example provides further support for a role of AS-drivers similar to mutations, expression or epigenetic changes in cancer drivers. Importantly, AS-drivers occur without gene expression changes in the host gene and thus provide an independent set of functional alterations not considered previously. Furthermore, our estimates of the number of potential AS-drivers have been very conservative; hence it is possible that many more remain to be described.

Functional domains and interactions might not always be entirely lost through a switch, as normal isoforms generally retain some expression in tumors. This could be partly due to the uncertainty in the estimate of transcript abundance from RNA sequencing or to the heterogeneity in the transcriptomes of tumor cells. Still, a relatively small change in transcript abundance could be enough to trigger an oncogenic effect (Bechara et al., 2013; Sebestyén et al., 2016). Additionally, a number of the AS-drivers define a switch from a protein-coding transcript to a non-coding one, possibly undergoing non-sense mediated decay. These can be considered a form of alternative splicing mediated gene expression regulation (Hansen et al., 2009), and will alter function in a similar way. The predicted impact on domains and interactions

could therefore be indicative of alterations on the regulatory networks with variable functional effects.

Our description in terms of isoform switches allows a better analysis of the protein features potentially gained or lost through splicing changes. However, this may have some limitation, as accurate determination of differential transcript usage in genes with many isoforms requires high coverage and sufficient samples per condition (Sebestyén et al., 2015). Another limitation of our approach is that we only recovered a small fraction of the entire set of protein-protein interactions taking place in the cell. For instance, we did not characterize those interactions possibly mediated through low complexity regions (Buljan et al., 2012; Ellis et al., 2012), hence we expect that many more interactions will be affected in tumors.

The origin of the observed splicing changes remains to be discovered. We did not find a general association with somatic mutations in *cis*. However, genetic alterations affecting splicing may involve small indels that are still hard to detect with the sequencing coverage available for these samples. Alternatively, it is possible that switches mostly occur through trans-acting alterations, such as the expression change in splicing factors (Sebestyén et al., 2016), which may be controlled by pathways often altered in tumors (Fu and Ares, 2014). On the other hand, multiple different alterations may trigger the same or similar splicing changes. For instance, mutations in *RBM10* or downregulation of *QKI* lead to the same splicing change in *NUMB* that promotes cell proliferation (Bechara et al., 2013; Zong et al., 2014). A faction of the somatic mutations in tumors are subclonal (Sottoriva et al., 2015), providing enough intra-tumor heterogeneity to allow AS-drivers to be present in a fraction of the cell population, which may allow their persistance in a fraction of the dividing tumor cells. Additionally, tumor cells display non-genetic variability, defining multiple stable states (Brock et al., 2009), which have been proposed to determine the fitness of cells and the progression of tumors independently of somatic mutations. Since natural selection acts on the phenotype rather than on the genotype, an interesting possibility is that AS-drivers define specific tumor phenotypes that might be closely related to those determined by the somatic mutations in drivers, thereby defining an advantageous phenotype such that the selective pressure to develop equivalent adaptations is relaxed. Accordingly, AS-drivers may play an important role in the neoplastic process independently of or in conjunction with the already characterized genetic alterations.

Experimental Procedures

Software and laboratory notebook

The pipeline developed to perform the analyses of isoform switches and the calculation of AS-drivers is freely available at https://bitbucket.org/regulatorygenomicsupf/smartas. The datasets used and the software to reproduce the analyses described in this work are available at https://github.com/hclimente/smartas.

Data

Estimated read counts for isoforms were obtained from the TCGA data portal (https://gdc.nci.nih.gov/). Only transcripts with TPM > 0.1 were considered expressed. Details on the mutation data collected is given in the Supplemental Experimental Procedures. We collected cancer drivers based from Intogen (Gundem et al., 2010) and from the TCGA papers for kidney renal papillary carcinoma (KIRP) and kidney chromophobe (KICH) (Davis et al., 2014; The Cancer Genome Atlas Research Network, 2016). This list included a total of 460 unique cancer driver genes, each one defined as a tumor-specific driver for one or more tumor types. The annotation of these genes as oncogenes or tumor suppressors was performed as in (Sebestyén et al., 2016) (see Supplemental Experimental Procedures for details).

Calculation of significant isoform switches per patient

We modeled splicing alterations in a gene as a switch between two transcript isoforms, one normal and one tumoral. For every transcript in every patient, we calculated $\Delta PSI = PSI_{tumor} - PSI_{ref}$, where PSI_{tumor} is the PSI value in a tumor sample and PSI_{ref} corresponds to the paired normal sample when available or to the median of the PSI distribution in the normal samples, otherwise. We considered significant those changes with $|\Delta PSI| > 0.05$ and empirical $p < 0.01$ in the comparison between normal and tumor samples. We only kept those cases for which the tumor isoform abundance was higher in the tumor than compared to the normal sample and the median abundance of the normal isoform was higher in normal samples compared to the tumor sample. Moreover, we discarded genes that either had an outlier expression in the tumor sample compared to normal tissues (empirical p-value < 0.025 or empirical p-value > 0.975) or

showed differential expression between the tumor samples with the switch and the normal samples (Wilcoxon test p-value < 0.01).

Candidate switches were defined per patient and for each gene. In some samples, different switches could appear for the same gene; hence, we discarded those switches that contradicted a more frequent switch in the same gene in the same tumor type. Moreover, we discarded any switch that affected a number of patients below the top 99% of the distribution of patient frequency of these contradictory switches. Thus, a switch in a patient sample is defined as a pair of transcripts in a gene with no expression change and with significant changes in opposite directions that show consistency across patients. We aggregated the calculated switches from the different tumor types to get the final list (Table S1). For the pan-cancer analyses, if a switch did not pass the frequency threshold in one tumor type but was significant in a different tumor type, that switch was also considered. Switches with significant association with stromal or immune cell content were discarded (see Supplemental Experimental Procedures).

Recurrence

We considered the number of different switches S and the number of patients with switches P, having in total N switches. We estimated the expected frequency of a switch as $f = N/(S \cdot P)$. For a given switch, we tested the significance of its recurrence across patients using a binomial test with the observed patient count and the expected frequency f. Switches were considered significantly recurrent for adjusted binomial test p-value < 0.05.

Simulated switches

We simulated switches between normal and tumor tissues by using genes with more than one expressed isoform. For each gene, we selected the isoform with the highest median expression across patients as the normal isoform and an arbitrary different transcript expressed in the tumor samples as the tumor isoform. For each gene, we generated a maximum of five such simulated switches.

Functional switches

A switch was defined as functional if both isoforms overlap in genomic extent and there is a change in the encoded protein, including cases where only one of the isoforms has a CDS, and

moreover there is a gain or loss of a protein feature: Pfam domains mapped with InterProScan (Jones et al., 2014), ProSite patterns (Gattiker et al., 2002); disordered regions from IUPred (Dosztanyi et al., 2005); disordered regions potentially involved in protein–protein interactions from ANCHOR (Dosztanyi et al., 2009); and protein loops (Bonet et al., 2014).

Domain families enriched in switches or mutations

To determine which protein domain families were significantly affected by switches, we first estimated their expected frequency of occurrence in switches. For this we calculated a reference proteome using the isoform with the highest median expression in the normal samples and with at least 0.1 TPM, for each expressed gene in each tumor type. For the enrichment analysis we used only genes with evidence of alternative splicing. We aggregated the representative proteomes from all tumor types to form a pan-cancer reference proteome. The expected frequency $f(a)$ for a feature a that appears $m(a)$ times was then measured as the proportion of this feature in the pan-cancer representative proteome:

$$f(a) = \frac{m(a)}{\sum_b m(b)}$$

We then calculated the expected probability using the binomial test:

$$P(a) = \frac{n!}{k!(n-k)!} f(a)^k (1 - f(a))^{n-k}$$

where k is the number of observations of the domain a being gained or lost and n is the total number of gains or losses due to switches. We selected cases with Benjamini-Hochberg (BH) adjusted p-value < 0.05. However, to ensure the specificity of the enrichment for each domain class, we considered only domains affected by at least two switches. To calculate domain families enriched in mutations, we considered the reference proteome in each tumor type as before. The expected mutation rate in a domain family is considered to be the proportion of the proteome length it covers. We aggregated all observed mutations falling within each family and calculated the expected probability of the observed mutations using a binomial test as before. After correcting for multiple testing, we kept those cases with a BH adjusted p-value < 0.05.

Protein interaction analysis

We created a consensus protein–protein interaction (PPI) network using data from PSICQUIC (del Toro et al. 2013), BIOGRID (Chatr-Aryamontri et al., 2015), HumNet (Lee et al., 2011), STRING (Szklarczyk et al., 2011), and a human interactome derived from the literature and experimental data (Rolland et al., 2014). The consensus network consisted of 8,142 nodes with 29,991 interactions, each found in at least four of these five sources. To find PPIs likely altered due to splicing change, we first mapped each PPI in a gene affected by isoform switches to a specific domain–domain interaction (DDI) that included one domain from the isoform expressed in the normal sample. We used information on domain–domain interactions from iPfam (Finn et al., 2014), DOMINE (Raghavachari et al., 2008), and 3did (Mosca et al., 2014). We only considered those PPIs that could be mapped to at least one DDI in either the normal or the tumor isoforms. We define a PPI as lost if it is mapped to one or more DDIs in the isoform expressed in the normal tissue but not in the isoform expressed in the tumor sample. When multiple domains mediate the same interaction, it is considered lost if at least one of these domains is lost in the switch. On the other hand, we define a PPI as gained if it can only be mapped to a DDI in the tumor-expressed isoform but not in the normal isoform.

Analysis of the interaction network affected by switches

We considered gene sets consisting of functional and cancer-related pathways (Liberzon et al., 2015), protein complexes (Ruepp et al., 2009) and complexes related to RNA metabolism (Akerman et al., 2015). The enrichment of PPI-affecting switches was performed with a Fisher's exact test based on the separation of switches into being in the gene set or not, and affecting PPIs or not. To calculate network modules, we considered the network defined only by PPIs that are affected by switches, whose nodes are PPI-affecting switches and their interaction partners. We calculated modules using the multi-level modularity optimization algorithm for finding community structures (Blondel et al., 2008) implemented in the iGraph R package (http://igraph.org/r/doc/cluster_louvain.html). For each of the gene sets used before, we calculated whether it was significantly included in a module using a binomial test to estimate the probability of finding by chance the observed number of genes with affected PPIs in an arbitrary list of genes of the same size as the gene set.

Mutation analysis

To identify switches significantly associated with pan-negative tumors, we considered the top 10 drivers according to their frequency of protein-affecting mutations in each tumor type. We tested the mutual exclusion between the patients affected by the switch and the patients with a PAM in at least the top three drivers using a one-tailed Fisher's test (Babur et al., 2015). From this set, we further considered functional switches that shared functional pathway with a driver. Mutual exclusion between PPI-affecting switches and target drivers was further measured with DENDRIX (Vandin et al., 2012). To test the validity of mutual-exclusion patterns observed, for each mutational driver we sampled the same number of PPI-affecting switches randomly from the same patients 100 times. None of the random combinations of switches showed significance in their mutual exclusion patterns with the mutational driver.

Author contributions

EE proposed the study. HCG performed the calculation of AS-drivers and the analysis of their properties. EPP carried out the analyses of protein-protein interactions as well as the stroma and immune cell content analysis. EE and AG supervised the analyses. EE and HCG wrote the manuscript with essential inputs from EPP and AG.

Acknowledgements

HCG and EE were supported by the MINECO and FEDER (BIO2014-52566-R), Consolider RNAREG (CSD2009-00080), AGAUR (SGR2014-1121), and the Sandra Ibarra Foundation for Cancer (FSI2013). EPP and AG were supported by the SBP CC grant (P30 CA030199). All authors thank The Cancer Genome Atlas project for making their results publicly available for further analysis.

References

Akerman, M., Fregoso, O.I., Das, S., Ruse, C., Jensen, M. a, Pappin, D.J., Zhang, M.Q., Krainer, A.R., 2015. Differential connectivity of splicing activators and repressors to the

human spliceosome. Genome Biol. 16, 119. doi:10.1186/s13059-015-0682-5

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., Dutertre, M., Stern, M.-H., 2016. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. 7, 10615. doi:10.1038/ncomms10615

Babur, Ö., Gönen, M., Aksoy, B.A., Schultz, N., Ciriello, G., Sander, C., Demir, E., 2015. Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. Genome Biol. 16, 45. doi:10.1186/s13059-015-0612-6

Bechara, E.G., Sebestyén, E., Bernardis, I., Eyras, E., Valcárcel, J., 2013. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. Mol. Cell 52. doi:10.1016/j.molcel.2013.11.010

Birzele, F., Csaba, G., Zimmer, R., 2008. Alternative splicing and protein structure evolution. Nucleic Acids Res. 36, 550–558. doi:10.1093/nar/gkm1054

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 10008, 6. doi:10.1088/1742-5468/2008/10/P10008

Bonet, J., Planas-Iglesias, J., Garcia-Garcia, J., Marín-López, M.A., Fernandez-Fuentes, N., Oliva, B., 2014. ArchDB 2014: Structural classification of loops in proteins. Nucleic Acids Res. 42. doi:10.1093/nar/gkt1189

Bourdon, J.-C., 2007. P53 and Its Isoforms in Cancer. Br. J. Cancer 97, 277–82. doi:10.1038/sj.bjc.6603886

Brock, A., Chang, H., Huang, S., 2009. Non-genetic heterogeneity--a mutation-independent driving force for the somatic evolution of tumours. Nat. Rev. Genet. 10, 336–42. doi:10.1038/nrg2556

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., Babu, M.M., 2012. Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. Mol. Cell 46, 871–883. doi:10.1016/j.molcel.2012.05.039

Chabot, B., Shkreta, L., 2016. Defective control of pre-messenger RNA splicing in human disease. J. Cell Biol. 212, 13–27. doi:10.1083/jcb.201510032

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K., Tyers, M., 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 43, D470–D478. doi:10.1093/nar/gku1204

Ciriello, G., Miller, M.L., Aksoy, B.A., Senbabaoglu, Y., Schultz, N., Sander, C., 2013. Emerging landscape of oncogenic signatures across human cancers. Nat. Genet. 45, 1127–1133. doi:10.1038/ng.2762

Danan-Gotthold, M., Golan-Gerstl, R., Eisenberg, E., Meir, K., Karni, R., Levanon, E.Y., 2015. Identification of recurrent regulated alternative splicing events across human solid tumors. Nucleic Acids Res. 43, 1–15. doi:10.1093/nar/gkv210

David, C.J., Manley, J.L., 2010. Alternative pre-mRNA splicing regulation in cancer: Pathways and programs unhinged. Genes Dev. doi:10.1101/gad.1973010

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B. a, Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G.J., Bigner, D.D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J.W.C., Leung, S.Y., Yuen, S.T., Weber, B.L., Seigler, H.F., Darrow, T.L., Paterson, H., Marais, R., Marshall,

C.J., Wooster, R., Stratton, M.R., Futreal, P.A., 2002. Mutations of the BRAF gene in human cancer. Nature 417, 949–954. doi:10.1038/nature00766

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S., Fahey, C.C., Hacker, K.E., Bhanot, G., Gordenin, D.A., Chu, A., Gunaratne, P.H., Biehl, M., Seth, S., Kaipparettu, B.A., Bristow, C.A., Donehower, L.A., Wallen, E.M., Smith, A.B., Tickoo, S.K., Tamboli, P., Reuter, V., Schmidt, L.S., Hsieh, J.J., Choueiri, T.K., Hakimi, A.A., Chin, L., Meyerson, M., Kucherlapati, R., Park, W.Y., Robertson, A.G., Laird, P.W., Henske, E.P., Kwiatkowski, D.J., Park, P.J., Morgan, M., Shuch, B., Muzny, D., Wheeler, D.A., Linehan, W.M., Gibbs, R.A., Rathmell, W.K., Creighton, C.J., 2014. The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. Cancer Cell 26, 319–330. doi:10.1016/j.ccr.2014.07.014

Dosztányi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–4. doi:10.1093/bioinformatics/bti541

Dosztányi, Z., Mészáros, B., Simon, I., 2009. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25, 2745–6. doi:10.1093/bioinformatics/btp518

Ellis, J.D., Barrios-Rodiles, M., ??olak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., Wrana, J.L., Blencowe, B.J., 2012. Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. Mol. Cell 46, 884–892. doi:10.1016/j.molcel.2012.05.037

Fang, H., Gough, J., 2013. DcGO: Database of domain-centric ontologies on functions, phenotypes, diseases and more. Nucleic Acids Res. 41. doi:10.1093/nar/gks1080

Finn, R.D., Miller, B.L., Clements, J., Bateman, A., 2014. IPfam: A database of protein family and domain interactions found in the Protein Data Bank. Nucleic Acids Res. 42. doi:10.1093/nar/gkt1210

Frampton, G.M., Ali, S.M., Rosenzweig, M., Chmielecki, J., Lu, X., Bauer, T.M., Akimov, M., Bufill, J.A., Lee, C., Jentz, D., Hoover, R., Ignatius Ou, S.H., Salgia, R., Brennan, T., Chalmers, Z.R., Jaeger, S., Huang, A., Elvin, J.A., Erlich, R., Fichtenholtz, A., Gowen, K.A., Greenbowe, J., Johnson, A., Khaira, D., McMahon, C., Sanford, E.M., Roels, S., White, J., Greshock, J., Schlegel, R., Lipson, D., Yelensky, R., Morosini, D., Ross, J.S., Collisson, E., Peters, M., Stephens, P.J., Miller, V.A., 2015. Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors. Cancer Discov. 5, 850–860. doi:10.1158/2159-8290.CD-15-0285

Fu, X.-D., Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. 15, 689–701. doi:10.1038/nrg3778

Gattiker, A., Gasteiger, E., Bairoch, A., 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. Appl. Bioinformatics 1, 107–8.

Goldstein, L.D., Lee, J., Gnad, F., Klijn, C., Schaub, A., Reeder, J., Daemen, A., Bakalarski, C.E., Holcomb, T., Shames, D.S., Hartmaier, R.J., Chmielecki, J., Seshagiri, S., Gentleman, R., Stokoe, D., 2016. Recurrent Loss of NFE2L2 Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers. Cell Rep. doi:10.1016/j.celrep.2016.08.010

Gundem, G., Perez-Llamas, C., Jene-Sanz, A., Kedzierska, A., Islam, A., Deu-Pons, J., Furney, S.J., Lopez-Bigas, N., 2010. IntOGen: integration and data mining of multidimensional oncogenomic data. Nat. Methods 7, 92–93. doi:10.1038/nmeth0210-92

Hansen, K.D., Lareau, L.F., Blanchette, M., Green, R.E., Meng, Q., Rehwinkel, J., Gallusser, F.L., Izaurralde, E., Rio, D.C., Dudoit, S., Brenner, S.E., 2009. Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in Drosophila. PLoS Genet. 5. doi:10.1371/journal.pgen.1000525

Jackson, C., Browell, D., Gautrey, H., Tyson-Capper, A., 2013. Clinical significance of HER-2 splice variants in breast cancer progression and drug resistance. Int. J. Cell Biol. doi:10.1155/2013/973584

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240. doi:10.1093/bioinformatics/btu031

Jung, H., Lee, D., Lee, J., Park, D., Kim, Y.J., Park, W.-Y., Hong, D., Park, P.J., Lee, E., 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat. Genet. 47, 1242–1248. doi:10.1038/ng.3414

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., Krainer, A.R., 2007. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat. Struct. Mol. Biol. 14, 185–93. doi:10.1038/nsmb1209

Kong-Beltran, M., Seshagiri, S., Zha, J., Zhu, W., Bhawe, K., Mendoza, N., Holcomb, T., Pujara, K., Stinson, J., Fu, L., Severin, C., Rangell, L., Schwall, R., Amier, L., Wickramasinghe, D., Yauch, R., 2006. Somatic mutations lead to an oncogenic deletion of Met in lung cancer. Cancer Res. 66, 283–289. doi:10.1158/0008-5472.CAN-05-2749

Lee, I., Blom, U.M., Wang, P.I., Shim, J.E., Marcotte, E.M., 2011. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res. 21, 1109–1121. doi:10.1101/gr.118992.110

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. 1, 417–425. doi:10.1016/j.cels.2015.12.004

Lu, Z., Huang, Q., Park, J.W., Shen, S., Lin, L., Tokheim, C.J., Henry, M.D., Xing, Y., 2015. Transcriptome-wide landscape of pre-mRNA alternative splicing associated with metastatic colonization. Mol. Cancer Res. 13, 305–18. doi:10.1158/1541-7786.MCR-14-0366

Ma, P.C., Kijima, T., Maulik, G., Fox, E.A., Sattler, M., Griffin, J.D., Johnson, B.E., Salgia, R., 2003. c-MET mutational analysis in small cell lung cancer: Novel juxtamembrane domain

mutations regulating cytoskeletal functions. Cancer Res. 63, 6272–6281.

Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., Miyano, S., Thol, F., Ganser, A., Yang, H., Haferlach, T., Ogawa, S., Koeffler, H.P., 2015. Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat. Commun. 6, 6042. doi:10.1038/ncomms7042

Miller, M.L., Reznik, E., Gauthier, N.P., Aksoy, B.A., Korkut, A., Gao, J., Ciriello, G., Schultz, N., Sander, C., 2015. Pan-Cancer Analysis of Mutation Hotspots in Protein Domains. Cell Syst. 1, 197–209. doi:10.1016/j.cels.2015.08.014

Mosca, R., Céol, A., Stein, A., Olivella, R., Aloy, P., 2014. 3did: A catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 42. doi:10.1093/nar/gkt887

Norris, A.D., Calarco, J.A., 2012. Emerging roles of alternative pre-mRNA splicing regulation in neuronal development and function. Front. Neurosci. doi:10.3389/fnins.2012.00122

Paik, P.K., Drilon, A., Fan, P.-D., Yu, H., Rekhtman, N., Ginsberg, M.S., Borsu, L., Schultz, N., Berger, M.F., Rudin, C.M., Ladanyi, M., 2015. Response to MET Inhibitors in Patients with Stage IV Lung Adenocarcinomas Harboring MET Mutations Causing Exon 14 Skipping. Cancer Discov. 5, 842–849. doi:10.1158/2159-8290.CD-14-1467

Poulikakos, P.I., Persaud, Y., Janakiraman, M., Kong, X., Ng, C., Moriceau, G., Shi, H., Atefi, M., Titz, B., Gabay, M.T., Salton, M., Dahlman, K.B., Tadi, M., Wargo, J.A., Flaherty, K.T., Kelley, M.C., Misteli, T., Chapman, P.B., Sosman, J.A., Graeber, T.G., Ribas, A., Lo, R.S., Rosen, N., Solit, D.B., 2011. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature 480, 387–90. doi:10.1038/nature10662

Raghavachari, B., Tasneem, A., Przytycka, T.M., Jothi, R., 2008. DOMINE: A database of protein domain interactions. Nucleic Acids Res. 36. doi:10.1093/nar/gkm761

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., Kamburov, A., Ghiassian, S.D., Yang, X., Ghamsari, L., Balcha, D., Begg, B.E., Braun, P., Brehme, M., Broly, M.P., Carvunis, A.R., Convery-Zupan, D., Corominas, R., Coulombe-Huntington, J., Dann, E., Dreze, M., Dricot, A., Fan, C., Franzosa, E., Gebreab, F., Gutierrez, B.J., Hardy, M.F., Jin, M., Kang, S., Kiros, R., Lin, G.N., Luck, K., Macwilliams, A., Menche, J., Murray, R.R., Palagi, A., Poulin, M.M., Rambout, X., Rasla, J., Reichert, P., Romero, V., Ruyssinck, E., Sahalie, J.M., Scholz, A., Shah, A.A., Sharma, A., Shen, Y., Spirohn, K., Tam, S., Tejeda, A.O., Trigg, S.A., Twizere, J.C., Vega, K., Walsh, J., Cusick, M.E., Xia, Y., Barabási, A.L., Iakoucheva, L.M., Aloy, P., De Las Rivas, J., Tavernier, J., Calderwood, M.A., Hill, D.E., Hao, T., Roth, F.P., Vidal, M., 2014. A proteome-scale map of the human interactome network. Cell 159, 1212–1226. doi:10.1016/j.cell.2014.10.050

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., Mewes, H.W., 2009. CORUM: The comprehensive resource of mammalian protein complexes-2009. Nucleic Acids Res. 38. doi:10.1093/nar/gkp914

Saito, M., Shimada, Y., Shiraishi, K., Sakamoto, H., Tsuta, K., Totsuka, H., Chiku, S., Ichikawa, H., Kato, M., Watanabe, S.I., Yoshida, T., Yokota, J., Kohno, T., 2015. Development of lung adenocarcinomas with exclusive dependence on oncogene fusions. Cancer Res. 75, 2264–2271. doi:10.1158/0008-5472.CAN-14-3282

Salton M, Kasprzak WK, Voss T, Shapiro BA, Poulikakos PI, M.T., 2015. Inhibition of vemurafenib-resistant melanoma by interference with pre-mRNA splicing. Nat. Commun. Dis. 6:7103. doi:10.1002/jcp.24872.The

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M.A., Valcárcel, J., Eyras, E., 2016. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. 26. doi:10.1101/gr.199935.115

Sebestyén, E., Zawisza, M., Eyras, E., 2015. Detection of recurrent alternative splicing switches

in tumor samples reveals novel signatures of cancer. Nucleic Acids Res. 43. doi:10.1093/nar/gku1392

Shen, S., Wang, Y., Wang, C., Wu, Y.N., Xing, Y., 2016. SURVIV for survival analysis of mRNA isoform variation. Nat. Commun. 7, 11548. doi:10.1038/ncomms11548

Sotillo, E., Barrett, D.M., Black, K.L., Bagashev, A., Oldridge, D., Wu, G., Sussman, R., Lanauze, C., Ruella, M., Gazzara, M.R., Martinez, N.M., Harrington, C.T., Chung, E.Y., Perazzelli, J., Hofmann, T.J., Maude, S.L., Raman, P., Barrera, A., Gill, S., Lacey, S.F., Melenhorst, J.J., Allman, D., Jacoby, E., Fry, T., Mackall, C., Barash, Y., Lynch, K.W., Maris, J.M., Grupp, S.A., Thomas-Tikhonenko, A., 2015. Convergence of acquired mutations and alternative splicing of CD19 enables resistance to CART-19 immunotherapy. Cancer Discov. 5, 1282–1295. doi:10.1158/2159-8290.CD-15-1020

Sottoriva, A., Kang, H., Ma, Z., Graham, T.A., Salomon, M.P., Zhao, J., Marjoram, P., Siegmund, K., Press, M.F., Shibata, D., Curtis, C., 2015. A Big Bang model of human colorectal tumor growth. Nat. Genet. 47, 209–216. doi:10.1038/ng.3214

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., Von Mering, C., 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39. doi:10.1093/nar/gkq973

The Cancer Genome Atlas Research Network, 2016. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N. Engl. J. Med. 374, 135–45. doi:10.1056/NEJMoa1505917

Trincado, J.L., Sebestyén, E., Pagés, A., Eyras, E., 2016. The prognostic potential of alternative transcript isoforms across human tumors. Genome Med. 8, 85. doi:10.1186/s13073-016-0339-3

Vandin, F., Upfal, E., Raphael, B.J., 2012. De novo discovery of mutated driver pathways in

cancer. Genome Res. 22, 375–385. doi:10.1101/gr.120477.111

Vorlová, S., Rocco, G., Lefave, C. V, Jodelka, F.M., Hess, K., Hastings, M.L., Henke, E., Cartegni, L., 2011. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. Mol. Cell 43, 927–39. doi:10.1016/j.molcel.2011.08.009

Wang, P., Yan, B., Guo, J.-T., Hicks, C., Xu, Y., 2005. Structural genomics analysis of alternative splicing and application to isoform structure modeling. Proc. Natl. Acad. Sci. U. S. A. 102, 18920–5. doi:10.1073/pnas.0506770102

Ward, A.J., Cooper, T.A., 2010. The pathobiology of splicing. J. Pathol. doi:10.1002/path.2649

Yanagisawa, M., Huveldt, D., Kreinest, P., Lohse, C.M., Cheville, J.C., Parker, A.S., Copland, J.A., Anastasiadis, P.Z., 2008. A p120 catenin isoform switch affects rho activity, induces tumor cell invasion, and predicts metastatic disease. J. Biol. Chem. 283, 18344–18354. doi:10.1074/jbc.M801192200

Yang, F., Petsalaki, E., Rolland, T., Hill, D.E., Vidal, M., Roth, F.P., 2015. Protein Domain-Level Landscape of Cancer-Type-Specific Somatic Mutations. PLoS Comput. Biol. 11. doi:10.1371/journal.pcbi.1004147

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y.A., Murray, R.R., Spirohn, K., Begg, B.E., Duran-Frigola, M., MacWilliams, A., Pevzner, S.J., Zhong, Q., Trigg, S.A., Tam, S., Ghamsari, L., Sahni, N., Yi, S., Rodriguez, M.D., Balcha, D., Tan, G., Costanzo, M., Andrews, B., Boone, C., Zhou, X.J., Salehi-Ashtiani, K., Charloteaux, B., Chen, A.A., Calderwood, M.A., Aloy, P., Roth, F.P., Hill, D.E., Iakoucheva, L.M., Xia, Y., Vidal, M., 2016. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. Cell 164, 805–817. doi:10.1016/j.cell.2016.01.029

Zhou, C., Licciulli, S., Avila, J.L., Cho, M., Troutman, S., Jiang, P., Kossenkov, a V, Showe, L.C., Liu, Q., Vachani, a, Albelda, S.M., Kissil, J.L., 2012. The Rac1 splice form Rac1b

promotes K-ras-induced lung tumorigenesis. Oncogene 32, 903–909. doi:10.1038/onc.2012.99

Zong, F.Y., Fu, X., Wei, W.J., Luo, Y.G., Heiner, M., Cao, L.J., Fang, Z., Fang, R., Lu, D., Ji, H., Hui, J., 2014. The RNA-Binding Protein QKI Suppresses Cancer-Associated Aberrant Splicing. PLoS Genet. 10. doi:10.1371/journal.pgen.1004289

Figure Legends

Figure 1. (A) Number of isoform switches (*y* axis) calculated in each tumor type, separated according to whether the switches are functional and affect cancer drivers. **(B)** Overlap graph of protein features affected in functional switches: Prosite patterns (Prosite), protein loops (ArchDB), Pfam domains (Pfam), disordered regions with potential to mediate protein–protein interactions (ANCHOR), and general disordered regions (IUPRED). The horizontal bars indicate the number of switches affecting each feature. The vertical bars indicate the number of switches in each intersection indicated by connected bullet points. **(C)** Top 20 functional switches in cancer drivers according to the total number of patients in which they occur. Tumor type is indicated by color: breast carcinoma (brca), colon adenocarcinoma (coad), head and neck squamous cell carcinoma (hnsc), kidney chromophobe (kich), kidney renal clear-cell carcinoma (kirc), kidney papillary cell carcinoma (kirp), liver hepatocellular carcinoma (lihc), lung adenocarcinoma (luad), lung squamous cell carcinoma (lusc), prostate adenocarcinoma (prad), and thyroid carcinoma (thca). **(D)** Biological processes (Slim level 2) associated with protein domain families that are significantly lost in functional isoform switches. For each functional category, we give the number of isoform switches in which a protein domain family related to this category is lost.

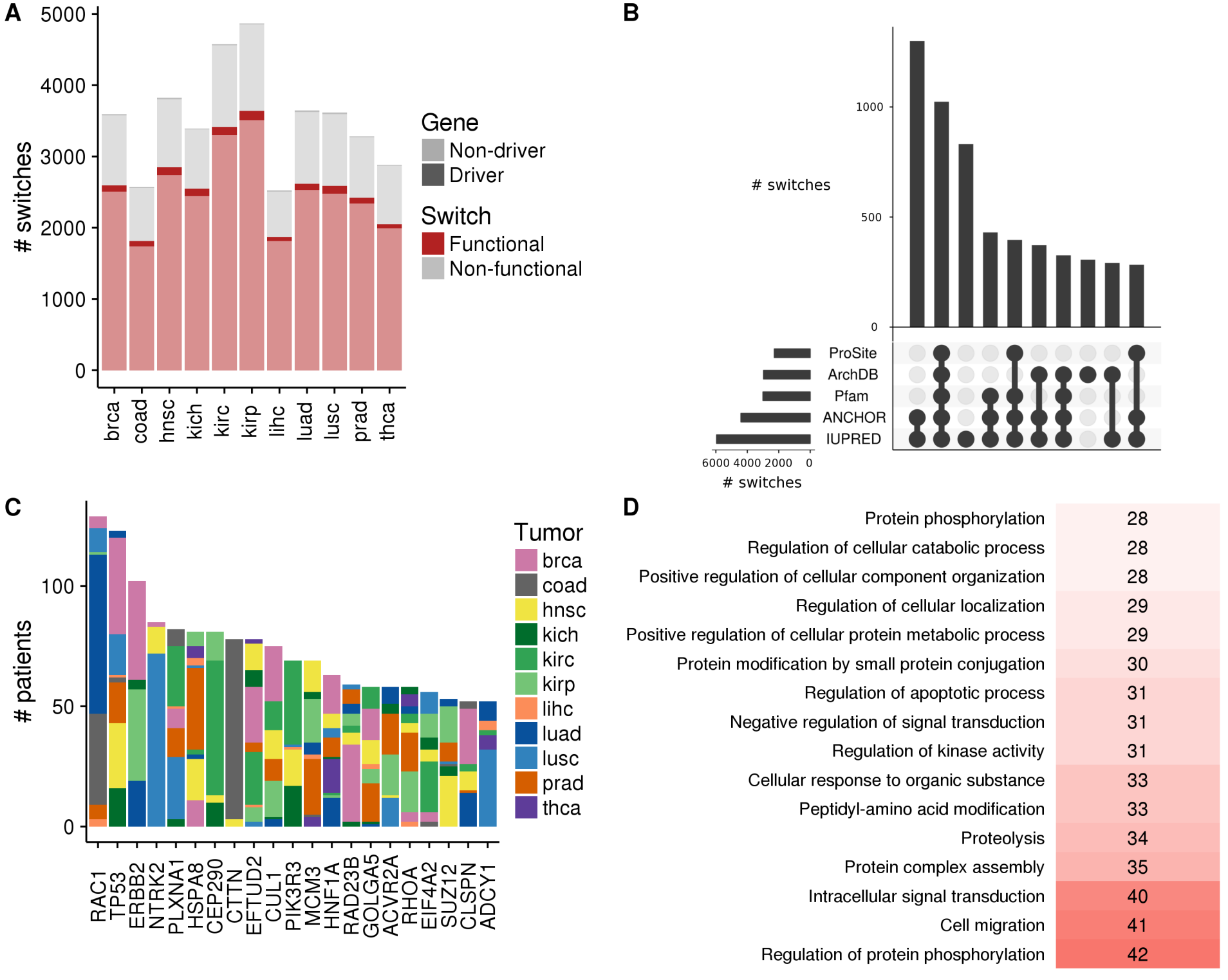
Figure 2. (A) For each patient sample, color-coded according to the tumor type, we indicate the proportion of all genes with protein-affecting mutations (PAMs) (*y* axis) and the proportion of genes with multiple transcript isoforms that present a functional isoform switch in the same sample (*x* axis). **(B)** Domain families that are significantly lost or gained in functional isoform

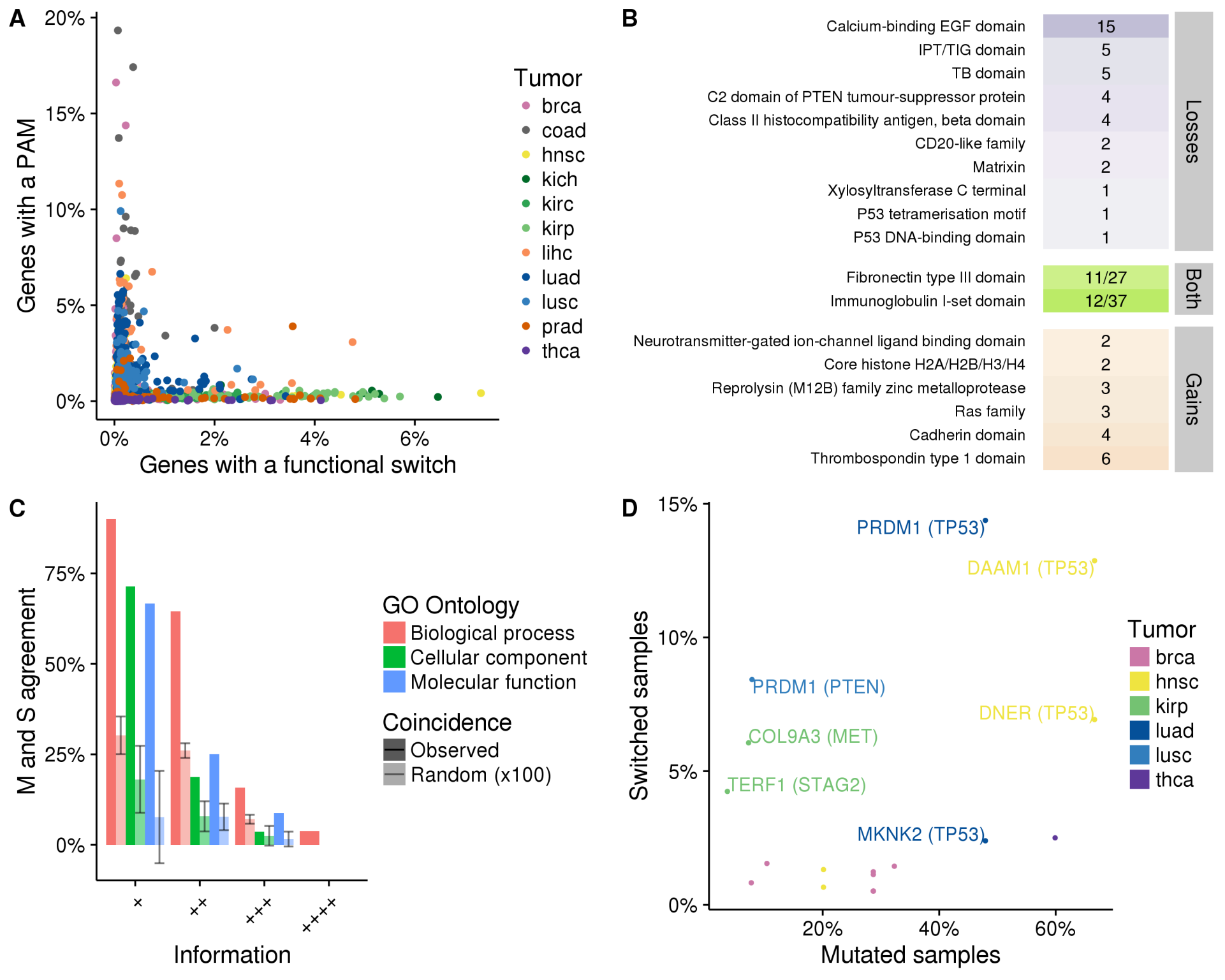
switches and that are also significantly enriched in protein-affecting mutations in tumors. For each domain class, we indicate the number of different switches in which they occur. We include here the loss of the P53 DNA-binding and P53 tetramerization domains, which only occur in the switch in *TP53*. **(C)** Agreement between protein-affecting mutations and functional switches measured (*y* axis) in terms of the functional categories of the protein domains they affect (*x* axis), using the three gene ontologies (GOs) and at four different GO Slim levels, from most specific (++++) to least specific (+). Random occurrences (plotted in light color) were calculated by sampling 100 times the same number of domain families affected by functional switches and the same number affected by protein-affecting mutations. Agreement is calculated as the percentage of the union of functional categories that are common to both sets. **(D)** Pairs formed by a cancer driver (in parentheses) and a functional switch that belong to the same pathway and show mutual exclusion between PAMs and switches across patients in at least one tumor type (color-coded by tumor type). The graph indicates the percentage of samples where the switch occurs (*y* axis) and the percentage of samples where the driver is mutated in the same tumor type (*x* axis).

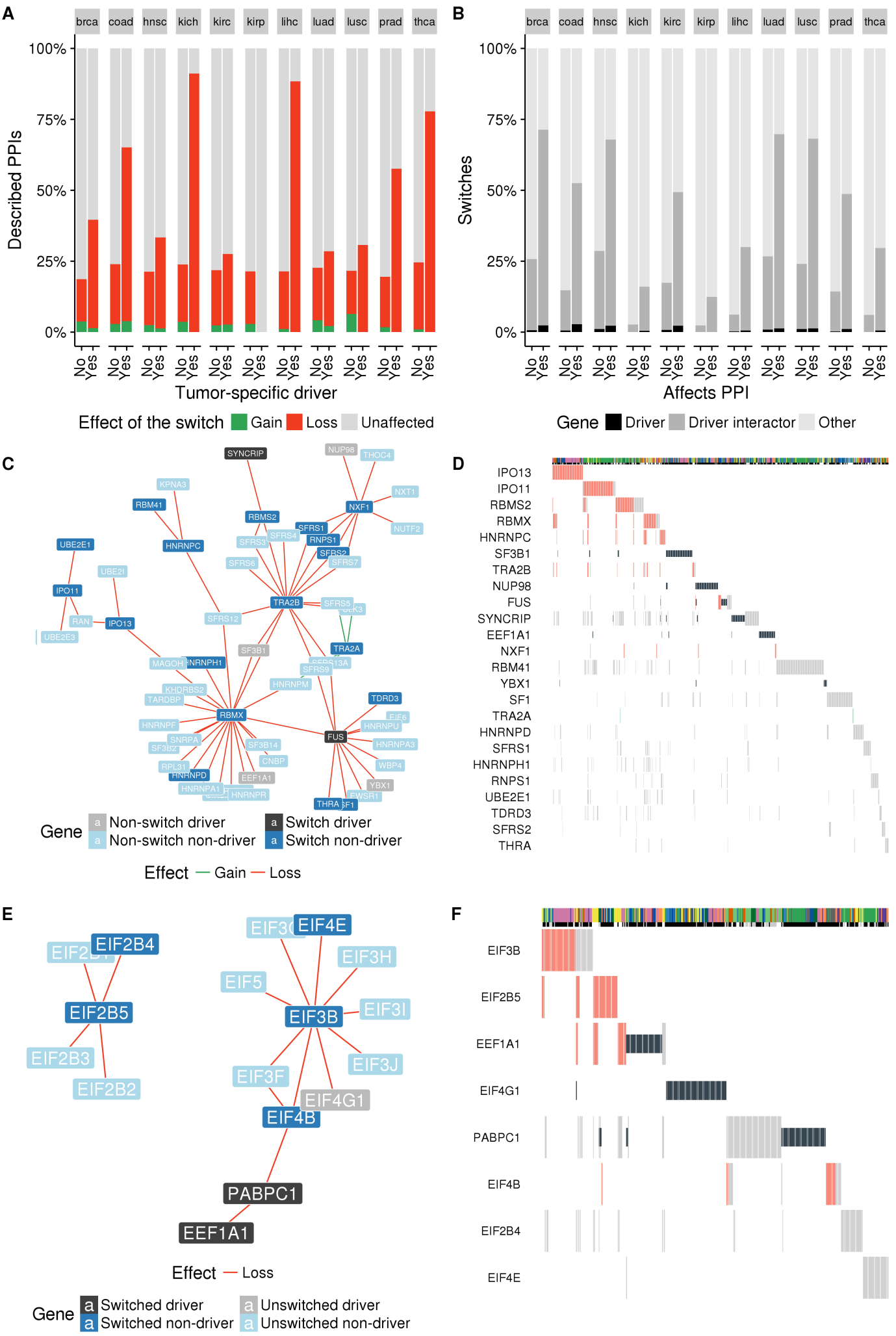
Figure 3. (A) Functional switches are divided according to whether they occur in tumor-specific drivers (yes) or not (no). For each tumor type we plot the proportion of protein-protein interactions (PPIs) described (*y* axis) that are kept intact (gray), lost (red), or gained (green). KIRP does not show PPI-affecting switches in drivers. **(B)** Functional switches are divided according to whether they affect a PPI (yes) or not (no). For each tumor type we plot the proportion of functional switches (*y* axis) that occur in cancer drivers (black), in driver interactors (dark gray), or in other genes (light gray). **(C)** A network module with PPIs predicted to be lost (red) or gained (green) by isoform switches. Cancer drivers are indicated in gray or black if they have a functional switch. Other genes are indicated in light blue, or dark blue if they have a functional switch. We do not indicate unaffected interactions. **(D)** Oncoprint for the samples that present the switches and protein-affecting mutations (PAMs) in drivers from (C). Mutations are indicated in black. PPI-affecting switches are indicated in red (loss) and green (gain). Other switches with no predicted effect on the PPI are depicted in gray. The top panel indicates the tumor type of each sample by color (same color code as in previous figures). The second top panel indicates whether the sample harbors a PAM in a tumor-specific driver (black) or not (gray). In white we indicate that no mutation data is available for that sample. **(E)** Same as in (C)

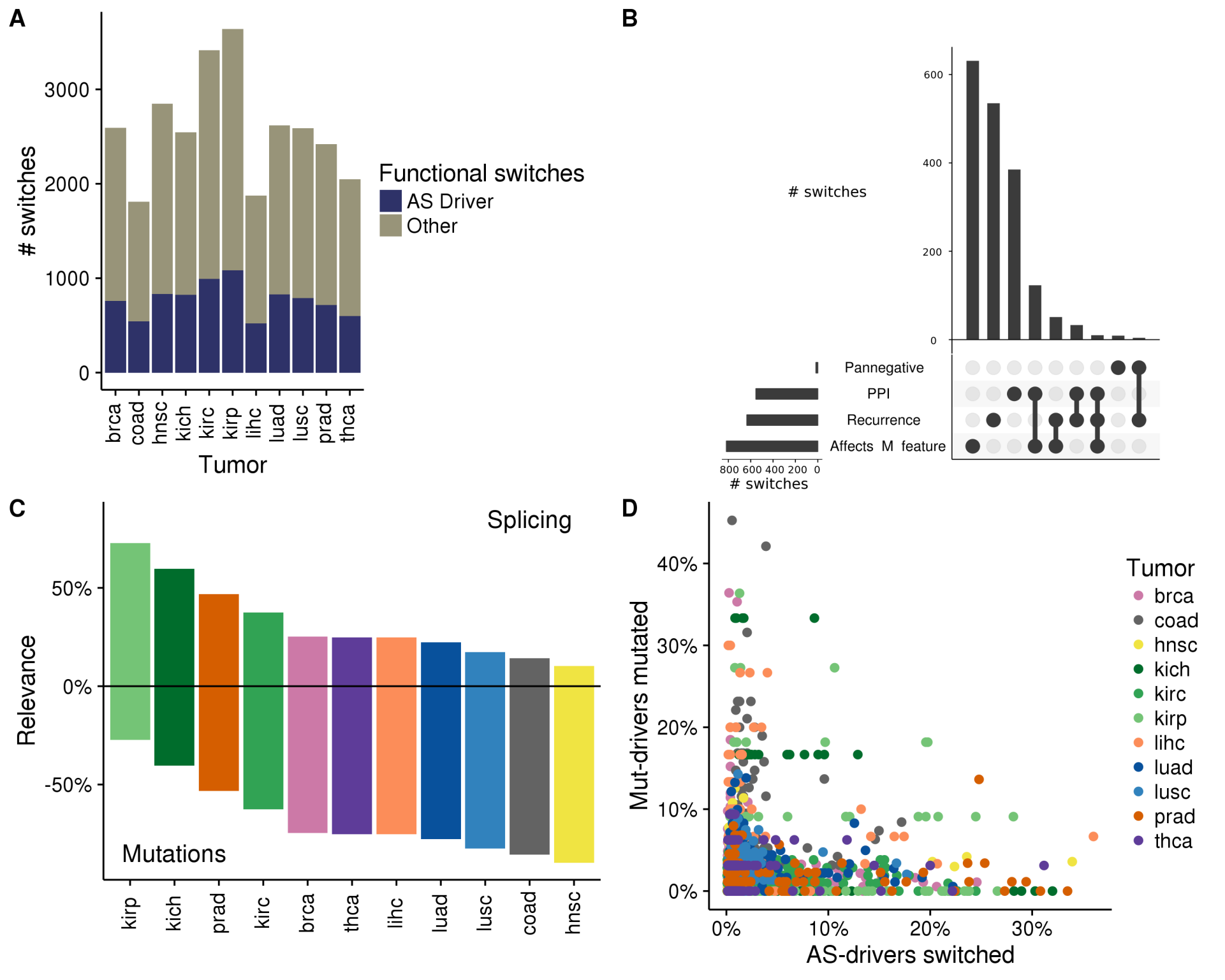
for the network module containing genes from the translation initiation complex. **(F)** Oncoprint for the switches and drivers from (E). Colors are as in (D).

Figure 4. (A) Number of functional switches and AS-drivers detected in each tumor type (see text for definitions). **(B)** Candidate AS-drivers grouped according to their properties: disruption of protein–protein interactions (PPIs), significant recurrence across patients (Recurrence), gain or loss of a protein feature that is frequently mutated in tumors (Affects M_feature), mutual exclusion and sharing pathway with cancer drivers (Pan-negative). The horizontal bars indicate the number of switches for each property. The vertical bars indicate the number of switches in each of the intersections indicated by connected bullet points. **(C)** Classification of samples according to the relevance of AS-drivers or Mut-drivers in each tumor type. For each tumor type (*x* axis), the positive *y* axis shows the percentage of samples that have a proportion of switched AS-drivers higher than the proportion of mutated Mut-drivers. The negative *y* axis shows the percentage of samples in which the proportion of mutated Mut-drivers is higher than the proportion of switched AS-drivers. Only patients with mutation and transcriptome data are shown. **(D)** Each of the patients from (C) is represented according to the percentage of mutated Mut-drivers (*y* axis) and the percentage of switched AS-drivers (*x* axis).


A
switches
5000
4000
3000
2000
1000
0
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
Gene
Non-driver
Driver
Switch
Functional
Non-functional
B
switches
1000
500
0
ProSite
ArchDB
Pfam
ANCHOR
IUPRED
6000 4000 2000 0
switches
C
patients
100
50
0
RAC1
TP53
ERBB2
NTRK2
PLXNA1
HSPA8
CEP290
CTTN
EFTUD2
CUL1
PIK3R3
MCM3
HNF1A
RAD23B
GOLGA5
ACVR2A
RHOA
EIF4A2
SUZ12
CLSPN
ADCY1
Tumor
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
D
Protein phosphorylation 28
Regulation of cellular catabolic process 28
Positive regulation of cellular component organization 28
Regulation of cellular localization 29
Positive regulation of cellular protein metabolic process 29
Protein modification by small protein conjugation 30
Regulation of apoptotic process 31
Negative regulation of signal transduction 31
Regulation of kinase activity 31
Cellular response to organic substance 33
Peptidyl-amino acid modification 33
Proteolysis 34
Protein complex assembly 35
Intracellular signal transduction 40
Cell migration 41
Regulation of protein phosphorylation 42


A
Genes with a PAM
20%
15%
10%
5%
0%
0%
2%
4%
6%
Genes with a functional switch
Tumor
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
B
Calcium-binding EGF domain 15
IPT/TIG domain 5
TB domain 5
C2 domain of PTEN tumour-suppressor protein 4
Class II histocompatibility antigen, beta domain 4
CD20-like family 2
Matrixin 2
Xylosyltransferase C terminal 1
P53 tetramerisation motif 1
P53 DNA-binding domain 1
Losses
Fibronectin type III domain 11/27
Immunoglobulin I-set domain 12/37
Both
Neurotransmitter-gated ion-channel ligand binding domain 2
Core histone H2A/H2B/H3/H4 2
Reprolysin (M12B) family zinc metalloprotease 3
Ras family 3
Cadherin domain 4
Thrombospondin type 1 domain 6
Gains
C
M and S agreement
75%
50%
25%
0%
×
××
×××
××××
Information
GO Ontology
Biological process
Cellular component
Molecular function
Coincidence
Observed
Random (x100)
D
Switched samples
15%
10%
5%
0%
20%
40%
60%
Mutated samples
PRDM1 (TP53)
DAAM1 (TP53)
PRDM1 (PTEN)
DNER (TP53)
COL9A3 (MET)
TERF1 (STAG2)
MKNK2 (TP53)
Tumor
brca
hnsc
kirp
luad
lusc
thca


A
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
Described PPIs
100%
75%
50%
25%
0%
No
Yes
Tumor-specific driver
Effect of the switch
Gain
Loss
Unaffected
B
Switches
Affects PPI
Gene
Driver
Driver interactor
Other
C
SYNCRIP
NUP98
THOC4
KPNA3
NXT1
RBM41
NXF1
RBMS2
SFRS1
NUTF2
UBE2E1
HNRNPC
SFRS3
SFRS4
RNPS1
SFRS2
SFRS6
SFRS7
UBE2I
IPO11
TRA2B
SFRS5
RAN
IPO13
SFRS12
UBE2E3
SF3B1
TRA2A
MAGOH
HNRNPH1
SFRS13A
SFRS9
HNRNPM
KHDRBS2
TARDBP
TDRD3
RBMX
EIF6
HNRNPF
HNRNPU
SNRPA
FUS
HNRNPA3
SF3B14
CNBP
RPL31
WBP4
HNRNPD
EEF1A1
YBX1
HNRNPA1
HNRNPR
THRA
Gene
Non-switch driver
Non-switch non-driver
Switch driver
Switch non-driver
Effect
Gain
Loss
D
IPO13
IPO11
RBMS2
RBMX
HNRNPC
SF3B1
TRA2B
NUP98
FUS
SYNCRIP
EEF1A1
NXF1
RBM41
YBX1
SF1
TRA2A
HNRNPD
SFRS1
HNRNPH1
RNPS1
UBE2E1
TDRD3
SFRS2
THRA
E
EIF2B4
EIF2B5
EIF2B3
EIF2B2
EIF4E
EIF3H
EIF5
EIF3B
EIF3I
EIF3J
EIF3F
EIF4G1
EIF4B
PABPC1
EEF1A1
Effect
Loss
Gene
Switched driver
Switched non-driver
Unswitched driver
Unswitched non-driver
F
EIF3B
EIF2B5
EEF1A1
EIF4G1
PABPC1
EIF4B
EIF2B4
EIF4E


A
switches
3000
2000
1000
0
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
Tumor
Functional switches
AS Driver
Other
B
switches
600
400
200
0
Pannegative
PPI
Recurrence
Affects M feature
800 600 400 200 0
switches
C
Splicing
Relevance
50%
0%
-50%
Mutations
kirp
kich
prad
kirc
brca
thca
lihc
luad
lusc
coad
hnsc
D
Mut-drivers mutated
40%
30%
20%
10%
0%
0%
10%
20%
30%
AS-drivers switched
Tumor
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca