# 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

Lead contact: 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 revealed 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 number of genes in 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), alternative splicing 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; Supek et al., 2014), 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; Darman et al., 2015; 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 mutations in the kinase domain of *BRAF* (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 to tumor cells and could represent 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., 2016, 2015; Trincado et al., 2016). However, the functional impact of most 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 present a systematic evaluation of the potential functional impact of recurrent alternative splicing changes observed in cancer samples. We described splicing changes in terms of transcript isoform switches in each tumor sample and determined the protein features and protein–protein interactions affected by them, and their relation to cancer drivers. Our analysis revealed a set of isoform switches that affect protein domains from families frequently mutated in tumors, remodel the protein interaction network of cancer drivers, and tend to 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 (STAR Methods). We used transcript isoforms to describe splicing changes, 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 an isoform switch, defined as a differential transcript isoform usage between a tumor sample and the normal samples. Each isoform switch was thus defined per patient as a pair of transcripts, which we named the tumor and the normal isoforms, such that the change in relative abundance of the tumor isoform between the normal samples and the tumor sample was higher than expected by the variability in normal samples, and such that the gene showed 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 could not be sure whether they were actually present in tumor cells (STAR Methods).

In all patients we found a total of 8,122 different isoform switches in 6,442 genes that described 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). Using SUPPA (Alamancos et al., 2015) we calculated the relation of the calculated switches with local alternative splicing (AS) events (STAR Methods). From the 8122 switches, 5667 (69.7%) were mapped to one or more local alternative splicing events. Comparing to the expected proportion of event types, we observed an enrichment of events of type alternative 5'ss (A5), alternative first exon (AF) and retained intron (RI), and a depletion of events of type alternative 3'ss (A3), alternative last exon (AL), mutually exclusive exons (MX) and exon-cassette (SE) (Figure S1A). We further calculated which of the two forms of the event corresponded to the tumor isoform (STAR Methods). RI events were enriched towards the retention of the intron, in agreement with previous observations (Dvinge and Bradley, 2015), whereas SE events were enriched for the skipping of the cassette exon (Figure S1B). Interestingly, 30,3% of the switches were not mapped to any local alternative splicing event, indicating that our description in terms of transcripts provides a wider spectrum of RNA variations than that described by local alternative splicing events.

## Isoform switches in cancer are frequently associated with protein feature losses

We next studied the proteins encoded by the transcripts involved in the switches. Interestingly, tumor protein isoforms tended to be shorter than protein isoforms in normal tissues (Figure S1C). Moreover, while for most switches — 6,937 (85,41%) — both transcript isoforms coded for protein, the rest of them had a significantly higher proportion of cases for which only the normal transcript isoform was protein-coding (9.01% *vs.* 2.8%, binomial test p-value < 2.2e-16) (Table S1), suggesting that isoform switches in tumors are associated with the loss of protein coding capacity. To determine the potential functional impact of the calculated isoform switches, we determined the protein features they affected (STAR Methods). Out of the 6,937 switches where both transcript isoforms coded for proteins, 5,047 (72.75%) involved 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 S1D). Interestingly, protein features are more frequently lost than gained (Figure 1B) and a large number of switches only affected disordered regions, with more than a half of them being related to protein-protein interactions (Figure S1D). We compared the switches that affect protein features with 100 sets of simulated switches, controlling for normal and tumor isoform expression (STAR Methods). Remarkably, isoform switches in tumors had more protein feature losses than expected by chance (Fisher's exact test p-value < 1.45e-08, odds-ratio > 1.29), despite the fact that for the simulated switches the normal protein isoform also tended to be longer than the tumor protein isoform (Figure S1E). This indicates that isoform switches in cancer are strongly associated with the loss of protein function capabilities.

We focused on isoform switches that showed a gain or loss in at least one protein feature, which we called *functional switches*, as they were likely to impact gene activity. There were 6004 (73,9%) functional switches (Table S1). These functional switches included 729 (8,9%) and 228 (2,8%) cases for which only the normal or the tumor isoform, respectively, coded for a protein with one or more predicted protein features. Interestingly, functional switches were enriched in cancer drivers (Fisher's exact test p-value = 2.034e-05, OR = 1.965) (Figure S1F). Among the top cancer driver genes with switches we identified a recurrent switch in *RAC1* (Figure 1C), which was linked before to tumor initiation and progression (Zhou et al., 2012) and which we predicted to gain an extra Ras family domain. We also found a recurrent switch in *TP53* that changed to a non-coding isoform and a switch in *ERBB2* that removed one of the receptor domains and does not coincided with splicing changes previously described for this gene in tumors (Jackson et al., 2013).

To characterize how functional switches affect protein function, we calculated the enrichment of gains or losses of specific domain families (STAR Methods). To ensure that this was attributed to a switch and not to the co-occurrence of two domains, for a domain to be considered for enrichment we imposed a minimum of two switches in different genes affecting the domain. We detected 220 and 41 domain families exclusively lost or gained, respectively, and 13 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 included the regulation of protein activity (Figure 1D), suggesting effects on protein-protein interactions. To further characterize these functional switches, we calculated the proportion of cancer drivers annotated as oncogenes or tumor suppressors that contained domain families

enriched in gains or losses using the reference proteome. From the 69 cancer drivers that contained domains enriched in gains, 58 (84%) correspond to oncogenes (Fisher's exact test p-value = 0.0066, odds-ratio 0.40). Although we did not find an enrichment of domain losses in tumor suppressors, domain families significantly gained in switches occur more frequently in oncogenes than in tumor suppressors (Wilcoxon test p-value = 0.0009) (STAR Methods). These results suggest a similarity between our functional 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 our switches and *cis*-occurring mutations from whole exome (WES) and whole genome (WGS) sequencing data (STAR Methods). The frequencies of genes or samples with functional switches were generally similar to those for protein-affecting mutations (PAMs) and smaller than the frequencies for all mutations from WGS data (Figures S2A and S2B), indicating a similar prevalence of switches and PAMs but not for switches and WGS mutations Since in our approach we calculated switches per patient, we were able to study how these distribute across patients in relation to mutations. We thus tested the co-occurrence of somatic mutations and switches across patients (STAR Methods) (Figures S2C-F). The largest associations with WGS mutations included a switch in the cancer gene driver *CUX1*, although involving a low number of patients (Figure S2D). When we considered the top cases according to the number of patients with both mutations and switches, the cancer gene drivers *TP53* and *NTRK2* and the cadherin gene *FAT3* appeared among the top ones (Figures S2E and S2F), as well as *FAM19A5*, which we already described as a switch before between lung adenocarcinoma and paired normal samples (Sebestyén et al., 2015).

Overall we observed a low association of mutations and switches (Figure S2G). On the other hand, tumor samples that had with few genes with protein-affecting mutations (PAMs) tended to have many genes with functional switches, and vice versa, tumor samples with a low number of functional switches tended to have many genes with PAMs (Figure 2A). This suggested a complementarity between protein affecting mutations and switches affecting protein domains. To investigate this, we calculated domain families enriched in PAMs (STAR Methods). We

found that 76 domain families across 11 tumor types were 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 = 5.637e-06, odds ratio = 4.71) (STAR Methods). From the domain families enriched in mutations, 7 showed only enrichment in losses, 6 showed only enrichment in gains, and 2 showed enrichment in both (Figure 2B) (Tables S2). Among the gains, we found Cadherin domains related to switches in *CHD8, CDH26, FAT1, FAT2 and FAT3*; whereas among the losses, we found the Calcium-binding EGF domain, which is affected by various switches, including one in *NOTCH4*. A notable case involved the loss of the *TP53* DNA-binding domain and the *TP53* tetramerization motif. Although it only occurs in a single switch in *TP53*, its high recurrence in patients 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 an enrichment analysis of Gene Ontologies (GO) on the domains enriched in mutations and switches separately (STAR Methods). We then calculated the overlap between both set and compared it to the overlap obtained by randomly sampling the reference proteome and the complete list of functional switches. Notably, the observed overlap is higher than expected at 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 that switches that affect domains frequently mutated in tumors could be considered to have a relevant impact in cell function. A total of 754 functional switches in 634 genes (47 of them in 37 cancer drivers) affected domains that are enriched in mutations (Table S1) and thus they represented 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 292 functional switches that were mutually exclusive with somatic PAMs in three or more cancer drivers (Fisher's test p-value < 0.05) (Table S3) (STAR Methods). Moreover, 16 of them showed mutual exclusion with at least one driver with which it shared a pathway. These 16 switches included one in *COL9A3,* which showed mutual exclusion with *MET* mutations in kidney renal papillary cell carcinoma (kirp), and one in *PRDM1*, which showed mutual exclusion with mutations in *TP53* in lung adenocarcinoma (LUAD) (Figure 2D). In lung squamous cell carcinoma (lusc), the same switch in *PRDM1* showed mutual exclusion with *PTEN,* which is in the same pathway, although this was the only mutual exclusion in this tumor type (Table S3). As these switches tended to occur in patients that do not harbor mutations in known cancer drivers (Figure S2D), they could be indicative of alternative oncogenic processes and potentially explain pan-negative tumors (Saito et al., 2015); hence we considered them potential AS-drivers.

## Isoform switches affect protein interactions with cancer drivers

Many of the frequently lost and gained domain families in functional switches were 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 used data from five different sources to build a consensus PPI network with 8,142 nodes, each node representing a gene (Figure S3) (STAR Methods). To determine the effect of switches on the PPI network, we mapped PPIs from this consensus PPI network to domain–domain interactions (DDIs), which we then mapped to the transcript isoforms involved in the switches (Figure S4). From the 8,142 genes in the PPI network, 3,243 had at least one isoform switch, and for 1,688 isoform switches (in 1,355 genes) we were able to map at least one of the PPIs to a specific DDI. A total of 162 of these switches were located in 123 cancer drivers, with the remaining 1,526 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 477 switches (28.3%) in 423 different genes affected domains that mediated protein interactions and thus likely impacted such interactions. Most of these interaction-altering switches (*n* = 414, 86.8%) caused the loss of the domain that mediates the interaction, while a minority (*n* = 64, 13.2%) led to a gain of the interacting domain. There is only one switch that led to gains and losses of interactions with different partners. This is in *TAF9*, which lost a TIFIID domain and gained an AAA domain (Table S4).

Notably, switches in driver genes tended to alter PPIs more frequently than those in non-drivers, and they more frequently lost interactions (Chi-squared test p-value = 1.87e-15) (Figure 3A). From the 162 switches in drivers, 41 (25.3%) of them altered at least one interaction, either causing loss (33 switches) or gain (8 switches). Interestingly, switches that affected domains from families enriched in mutations or that showed frequent mutual exclusion with mutational drivers also affected PPIs significantly more frequently than other functional switches (Chi-square test p-value < 2.2e-16 and p-value = 6.8e-08, respectively) (Figure S5). Notably, functional switches in genes annotated as direct interactors of drivers affected PPIs more frequently than the rest of functional switches (Fisher's exact test p-value<1.81e-03 OR>3.57 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 (Table S5), reinforcing the potential impact of isoform switches in cancer. We thus considered these 477 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 (STAR Methods). This produced 179 modules involving 1405 genes (Table S6). These modules were tested for the enrichment of genes belonging to specific protein complexes (Ruepp et al., 2009), complexes related to RNA-processing and splicing (Akerman et al., 2015) and cancer-related pathways (Liberzon et al., 2015) (Table S6). Further, we inspected the modules to identify those that included cancer drivers, had at least one switch

for which both isoforms are protein coding, and involved both PPI gains and losses. Among these cases, we found one module enriched in splicing factor (SF) and RNA binding protein (RBP) genes (Module 11 in Table S6) that included the cancer drivers *SF3B1*, *FUS*, *SYNCRIP*, *EEF1A1* and *YBX1* (Figure 3C). The module contained a switch in *RBMX* involving 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 yielded a non-coding transcript and would eliminate an interaction with *SF3B1* and multiple SFs. We also found a switch in *HNRNPC, TRA2A, NXF1 and RBMS2* that lost interactions with various SR protein coding genes. Consistent with a potential functional impact, the PPI-affecting switches showed mutual exclusion with the mutational cancer drivers (Figure 3D). Interestingly, this module also contained switches in the Importin genes *IPO11* and *IPO13*, which would affect interactions with the ubiquitin conjugating enzymes *UBE2E1*, *UBE2E3* and *UBE2I*, (Figure 3C), and showed mutual exclusion across different tumor types (Figure 3D). These results indicate that the activity of RNA-processing factors may be altered in cancer through the disruption of their PPIs by alternative splicing.

We also found a module including multiple regulators of translation (module 28 in Table S6), with switches in *EIF4B*, *EIF3B* and *EIF4E* that affected interactions with the drivers *EIF4G1*, *EIF4A2* and *PABPC1* (Figure 3E). The switch in *EIF4B* caused the skipping of one exon, which we predicted to eliminate an RRM domain and lose interactions with drivers *EIF4G1* and *PABPC1*. The switch in *EIF3B* yielded a non-coding transcript that would lose multiple interactions. Although we did not predict any PPI change for *EIF4E,* this switch lost eight predicted ANCHOR regions, suggesting a possible effect on yet to be described interactions. Besides frequent PAMs, *PABPC1* also presented 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 in tumors through the alteration of protein–protein interactions of the translational regulators.

## Isoform switches as drivers of cancer

Our results provide evidence that a subset of the alternative splicing switches show similar properties to mutational drivers. We hypothesized that these switches may be relevant for tumorigenesis, and defined them as alternative splicing drivers, or AS-drivers, if they either (I) induced a gain or a loss of a protein domain from a family frequently mutated in cancer, (II) affected one or more PPIs, (III) displayed mutual exclusion with drivers, or (IV) displayed recurrence in cancer genomes beyond what is expected by chance. This definition yielded 1662 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). Genes relevant to a given tumor type usually participate in the same pathways and therefore lie close to each other in the PPI network, and tend to show high centrality in the network (Jonsson and Bates, 2006; Taylor et al., 2009; Wachi et al., 2005). To validate our AS-drivers, we thus calculated their centrality and their distance to mutational drivers (STAR Methods). In our consensus PPI network, AS-drivers showed greater centrality (Mann-Whitney test p-value < 2.2e-16, W = 90999000) (Figure S6A) and closer distances to tumor-specific drivers (Fisher's exact test p-value < 2.2e-16, odds-ratio = 1.55) (Figure S6B) compared to the rest of switches.

The possible relevance of these AS-drivers varied 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 indicated that, although Mut-drivers are predominant in patients for most tumor types, AS-drivers seemed 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 tended to show a low number of switched AS-drivers, and vice versa (Figure 4D). Interestingly, copy number alteration (CNA) drivers show a similar anti-correlation (Figure S6C), bearing resemblance with the proposed cancer genome hyperbola between mutations and CNAs (Figure S6D) (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 transcript isoform switches, 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 or gaining protein–protein interactions – possibly also altering the formation of protein complexes – with cancer drivers or in cancer related pathways. Moreover, we observed that patients with AS-drivers tended not to harbor mutations in cancer drivers and the other way around. We proposed 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 definition of AS-drivers as complementary alterations with similar impacts to mutations, expression or epigenetic changes in cancer drivers. Importantly, AS-drivers could occur without gene expression changes in the host gene and thus provide an independent set of functional alterations not considered previously in cancer expression studies. As our estimates of the number of potential AS-drivers and their impact have been very conservative, 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 regulatory networks with variable functional effects.

Our description in terms of transcript isoform switches, rather than using local alternative splicing events, allowed a better analysis of the possible transcriptome variations and the protein features potentially gained or lost through splicing changes. However, this approach may have some limitations. Accurate determination of differential transcript usage in genes with many isoforms requires high coverage and sufficient samples per condition (Sebestyén et al., 2015). In our approach, we used information of the variability across normal samples to determine the significance of an isoform change per patient. On the other hand, since we used annotated transcript isoforms we may have missed unannotated transcripts that are specific to the tumor or transcripts that are not yet correctly annotated. 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 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 elucidated. We did not find a general association with somatic mutations in *cis*. It is possible that the genetic alterations affecting splicing involve small copy-number alterations or indels. For instance, a recurrent small deletion in TP73 was found recently in association to exon skipping in small cell lung cancer patients using WGS data (George et al., 2015). However, these alterations are still hard to detect with WES and WGS data and more targeted searches or deeper sequencing may be necessary to fully uncover such cases at genome-scale in other tumors. An alternative possibility is that the majority of the switches described occur through *trans* acting alterations, such as the expression change in splicing factors (Sebestyén et al., 2016). For instance, the oncogenic switch in *RAC1* (Zhou et al., 2012) is regulated by expression changes of various splicing factors (Gonçalves et al., 2009; Pelisch et al., 2012), which are controlled by pathways often altered in tumors (Fu and Ares, 2014).

Alterations in different splicing factors 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). The intra-tumor heterogeneity may thus allow recapitulating similar transcriptome phenotypes. It has been proposed that a large faction of the somatic mutations in tumors are subclonal (Sottoriva et al., 2015). This would provide enough heterogeneity to allow the persistence of AS-drivers in a

fraction of the dividing cell population. Additionally, tumor cells also display non-genetic variability (Brock et al., 2009), which could define multiple stable states and 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 adal1ptations 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.

# Author contributions

EE proposed the study. HCG performed the calculation of transcript isoform switches, their analyses, the calculation of AS-drivers and the study of their properties. EPP built the consensus protein-protein interaction network, did the mapping to domain-domain interactions and performed the stromal and immune cell content analysis per sample. 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

Alamancos, G.P., Pagés, A., Trincado, J.L., Eyras, E., Pages, A., Trincado, J.L., Bellora, N., Eyras, E., 2015. Leveraging transcript quantification for fast computation of alternative splicing profiles. RNA 21, 1521–1531. doi:10.1261/rna.051557.115

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, 720–33. 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

Butts, C.T., 2008. network : A Package for Managing Relational Data in R. J. Stat. Softw. 24, 1–36. doi:10.18637/jss.v024.i02

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

Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: An R Package for the Visualization of Intersecting Sets and their Properties 2–5. doi:10.1101/120600

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

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E.B., Chan, B., Colla, S., Corson, L., Feala, J., Fekkes, P., Ichikawa, K., Keaney, G.F., Lee, L., Kumar, P., Kunii, K., MacKenzie, C., Matijevic, M., Mizui, Y., Myint, K., Park, E.S., Puyang, X., Selvaraj, A., Thomas, M.P., Tsai, J., Wang, J.Y., Warmuth, M., Yang, H., Zhu, P., Garcia-Manero, G., Furman, R.R., Yu, L., Smith, P.G., Buonamici, S., 2015. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3' Splice Site Selection through Use of a Different Branch Point. Cell Rep. 13, 1033–45. doi:10.1016/j.celrep.2015.09.053

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

del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L., Hermjakob, H., 2013. A new reference implementation of the PSICQUIC web service. Nucleic Acids Res. 41, W601-6. doi:10.1093/nar/gkt392

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

Dvinge, H., Bradley, R.K., 2015. Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 7, 45. doi:10.1186/s13073-015-0168-9

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., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279-85. doi:10.1093/nar/gkv1344

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

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C.Y., Jia, M., De, T., Teague, J.W., Stratton, M.R., McDermott, U., Campbell, P.J., 2015. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 43, D805–D811. doi:10.1093/nar/gku1075

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

Fredriksson, N.J., Ny, L., Nilsson, J.A., Larsson, E., 2014. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat. Genet. 46, 1–7. doi:10.1038/ng.3141

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.

George, J., Lim, J.S., Jang, S.J., Cun, Y., Ozretic, L., Kong, G., Leenders, F., Lu, X., Fernandez-Cuesta, L., Bosco, G., Muller, C., Dahmen, I., Jahchan, N.S., Park, K.-S., Yang, D., Karnezis, A.N., Vaka, D., Torres, A., Wang, M.S., Korbel, J.O., Menon, R., Chun, S.-M., Kim, D., Wilkerson, M., Hayes, N., Engelmann, D., Putzer, B., Bos, M., Michels, S., Vlasic, I., Seidel, D., Pinther, B., Schaub, P., Becker, C., Altmuller, J., Yokota, J., Kohno, T.,

Iwakawa, R., Tsuta, K., Noguchi, M., Muley, T., Hoffmann, H., Schnabel, P.A., Petersen, I., Chen, Y., Soltermann, A., Tischler, V., Choi, C., Kim, Y.-H., Massion, P.P., Zou, Y., Jovanovic, D., Kontic, M., Wright, G.M., Russell, P.A., Solomon, B., Koch, I., Lindner, M., Muscarella, L.A., la Torre, A., Field, J.K., Jakopovic, M., Knezevic, J., Castanos-Velez, E., Roz, L., Pastorino, U., Brustugun, O.-T., Lund-Iversen, M., Thunnissen, E., Kohler, J., Schuler, M., Botling, J., Sandelin, M., Sanchez-Cespedes, M., Salvesen, H.B., Achter, V., Lang, U., Bogus, M., Schneider, P.M., Zander, T., Ansen, S., Hallek, M., Wolf, J., Vingron, M., Yatabe, Y., Travis, W.D., Nurnberg, P., Reinhardt, C., Perner, S., Heukamp, L., Buttner, R., Haas, S.A., Brambilla, E., Peifer, M., Sage, J., Thomas, R.K., 2015. Comprehensive genomic profiles of small cell lung cancer. Nature 524, 47–53. doi:10.1038/nature14664

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

Gonçalves, V., Matos, P., Jordan, P., 2009. Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways. Hum. Mol. Genet. 18, 3696–3707. doi:10.1093/hmg/ddp317

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

Hagberg, A.A., Schult, D.A., Swart, P.J., 2008. Exploring network structure, dynamics, and function using NetworkX. Proc. 7th Python Sci. Conf. (SciPy 2008) 11–15.

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

Jonsson, P.F., Bates, P.A., 2006. Global topological features of cancer proteins in the human interactome. Bioinformatics 22, 2291–7. doi:10.1093/bioinformatics/btl390

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

Pelisch, F., Khauv, D., Risso, G., Stallings-Mann, M., Blaustein, M., Quadrana, L., Radisky, D.C., Srebrow, A., 2012. Involvement of hnRNP A1 in the matrix metalloprotease-3-

dependent regulation of Rac1 pre-mRNA splicing. J. Cell. Biochem. 113, 2319–2329. doi:10.1002/jcb.24103

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, W.K., Voss, T., Shapiro, B.A., Poulikakos, P.I., Misteli, T., 2015. Inhibition of vemurafenib-resistant melanoma by interference with pre-mRNA splicing. Nat. Commun. 6, 7103. doi:10.1038/ncomms8103

Schroeder, M.P., Rubio-Perez, C., Tamborero, D., Gonzalez-Perez, A., Lopez-Bigas, N., 2014. OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action, in: Bioinformatics. doi:10.1093/bioinformatics/btu467

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, 1345–1356. doi:10.1093/nar/gku1392

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

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., Lehner, B., 2014. Synonymous mutations

frequently act as driver mutations in human cancers. Cell 156, 1324–35. doi:10.1016/j.cell.2014.01.051

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

Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., Wrana, J.L., 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat. Biotechnol. 27, 199–204. doi:10.1038/nbt.1522

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., Sebestyen, E., Pages, A., Eyras, E., 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

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz Jr., L.A., Kinzler, K.W., 2013. Cancer Genome Landscapes. Science (80-. ). 339, 1546–1558. doi:10.1126/science.1235122

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

Wachi, S., Yoneda, K., Wu, R., 2005. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 21, 4205–8. doi:10.1093/bioinformatics/bti688

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

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

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D. a, Carter, S.L., Getz, G., Stemke-Hale, K., Mills, G.B., Verhaak, R.G.W., 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. 4, 2612. doi:10.1038/ncomms3612

Zhao, M., Kim, P., Mitra, R., Zhao, J., Zhao, Z., 2015. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 1–9. doi:10.1093/nar/gkv1268

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

# Figures

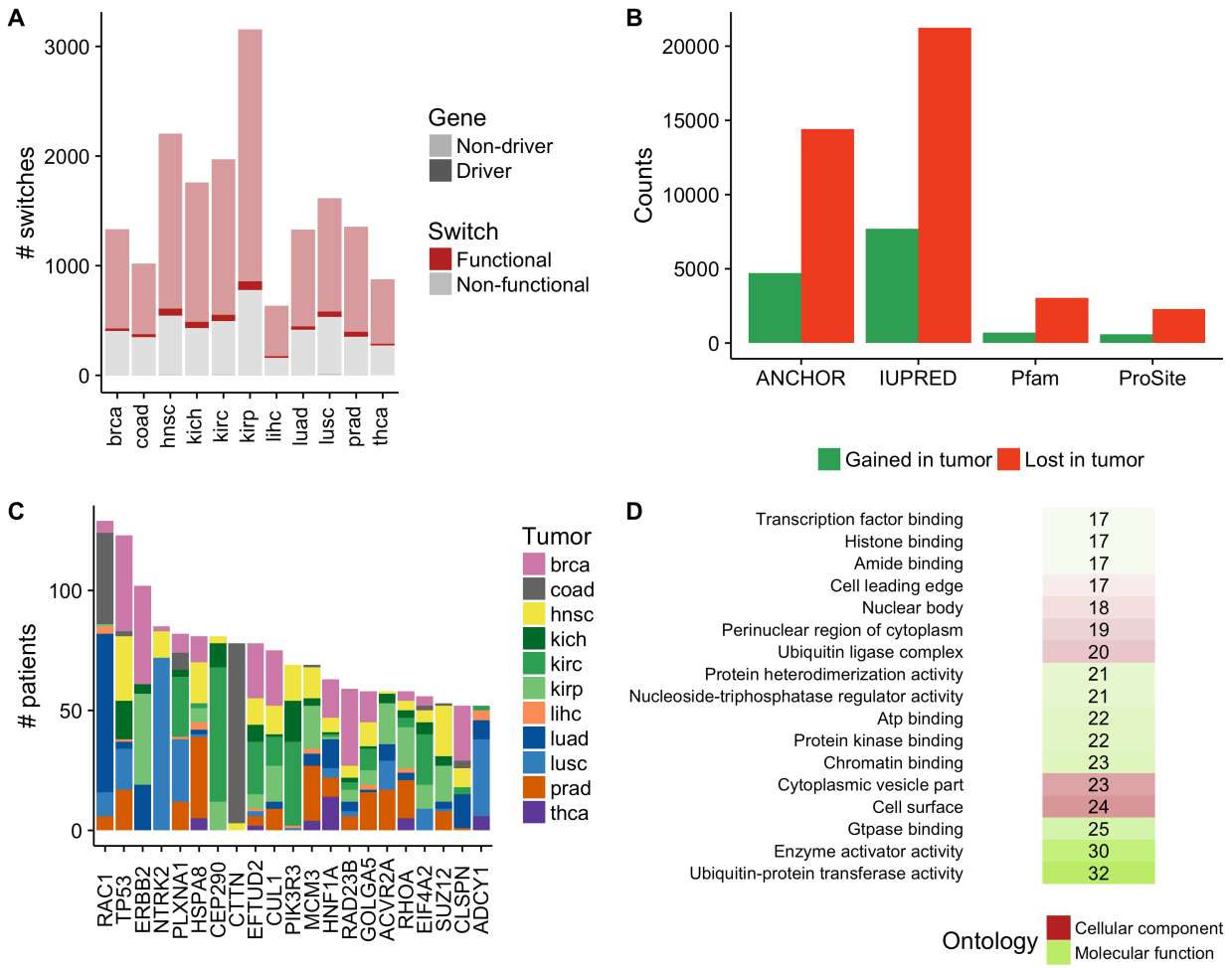



**Figure 1. (A)** Number of isoform switches (*y* axis) calculated in each tumor type, separated according to whether the switches affect an annotated protein feature (Functional) or not (Non-functional) and whether they occur in cancer gene drivers (Driver) or not (Non-driver). **(B)** Number of different protein feature gains and losses in functional switches for each of the protein annotations considered. **(C)** Top 20 functional switches in cancer drivers according to the total number of patients in which they occur. Tumor types are 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)** Cellular component (red) and Molecular function (green) ontologies associated with protein domain families that are significantly lost in functional isoform switches (Binomial test -

Benjamini-Hochberg adjusted p-value < 0.05). For each functional category, we give the number of isoform switches in which a protein domain family related to this category is lost. The color shade relates to this number.

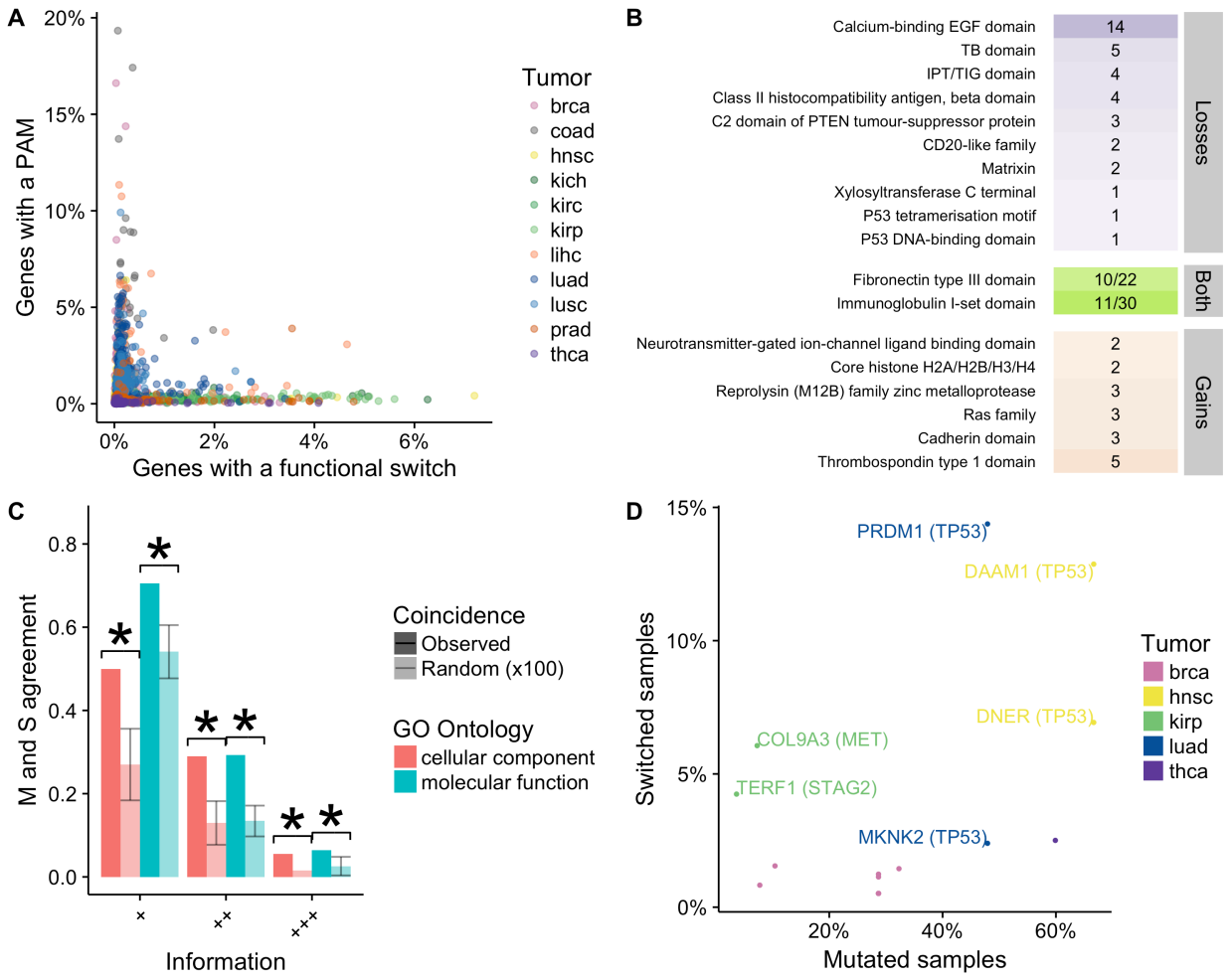



**Figure 2. (A)** For each patient sample, color-coded according to the tumor type, we indicate the proportion of all genes with protein-affecting mutations (PAMs) (*y* axis) and the proportion of genes with multiple transcript isoforms that present a functional isoform switch in the same sample (*x* axis). **(B)** Domain families that are significantly lost or gained in functional isoform switches 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 (*y* axis) measured in terms of the functional categories of the protein domains they affect (*x* axis), using two gene ontologies (GOs) at three 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 of domains 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 *y-axis* indicates the percentage of samples where the switch occurs and *x*-axis indicates the percentage of samples where the driver is mutated in the same tumor type.

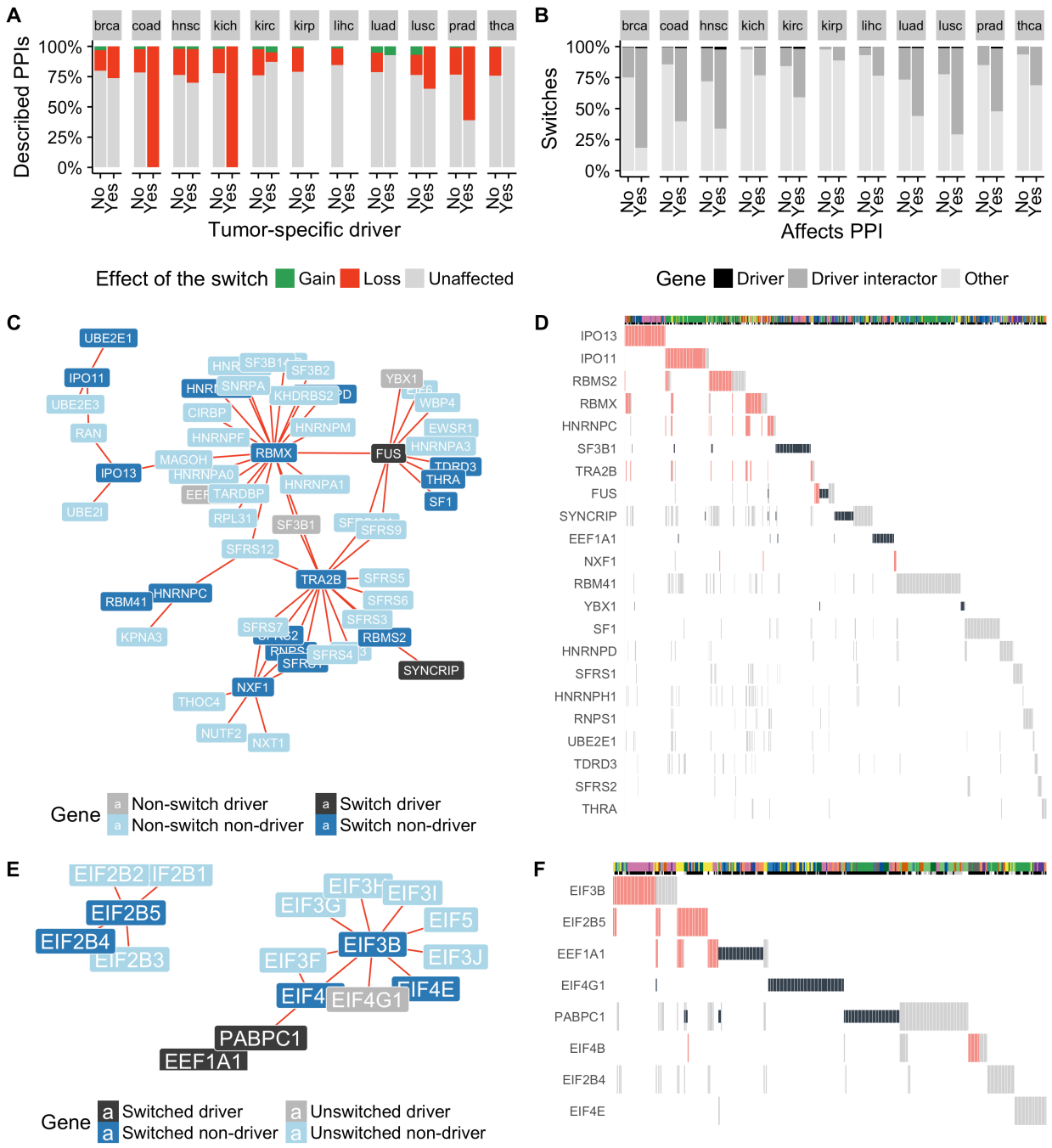



**Figure 3. (A)** Functional switches are divided according to whether they occur in tumor-specific drivers (yes) or not (no). For each tumor type we plot the proportion of protein-protein interactions (PPIs) (*y* axis) that are gained (green), lost (red), or remain unaffected (gray). Chi-squared test of all the PPIs affected by switches in drivers and non-drivers separating gains and losses: p-value = 1.87e-15. Individual Chi-square tests for each tumor type: brca p-val = 1.012e-

3, coad p-val = 1.34e-17, kich p-val = 6.414e-33, lusc p-val = 1.176e-7, prad p-val = 1.515e-12. The types kirc, luad and thca show no significance. Samples from kidney renal papillary cell carcinoma (kirp) and liver hepatocellular carcinoma (lihc) do 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 interactors of drivers (dark gray), or in other genes (light gray). Fisher's exact test of the PPIs affected by switches in driver-interactors and in other genes (non-drivers and non-driver-interactors): brca p-val=4.70e-58, OR= 6.39; coad p-val=1.84e-28, OR=5.59; hnsc p-val=4.40e-53, OR=5.12; kich pval=6.73e-16, OR=7.11, kirc p-val=4.35e-41, OR=4.68; kirp p-val= 5.69e-16, OR=6.15; lihc p-val=6.93e-17, OR=5.77; luad p-val=9.18e-61, OR=6.35; lusc p-val=4.76e-59, OR=6.47; prad, p-val= 1.00e-31, OR=5.18; thca, p-val=2.69e-19, OR=6.02. **(C)** A network module with PPIs predicted to be lost (red). Cancer drivers are indicated in black or gray if they have a functional switch or not, respectively. Other genes are indicated in dark blue or light blue if they have a functional switch or not, respectively. We do not show unaffected interactions. **(D)** OncoPrint for the samples that present protein-affecting mutations (PAMs) in drivers or switches from (C). Mutations are indicated in black and PPI-affecting switches are indicated in red (loss in this case). 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), or whether no mutation data is available for that sample (white). **(E)** A network module containing genes from the translation initiation complex with PPIs predicted to be lost (red) by isoform switches. Cancer drivers are indicated in black or gray if they have a functional switch or not, respectively. Other genes are indicated in dark blue or light blue if they have a functional switch or not, respectively. We do not show unaffected interactions. **(F)** OncoPrint for the switches and drivers from (E). Colors are as in (D).

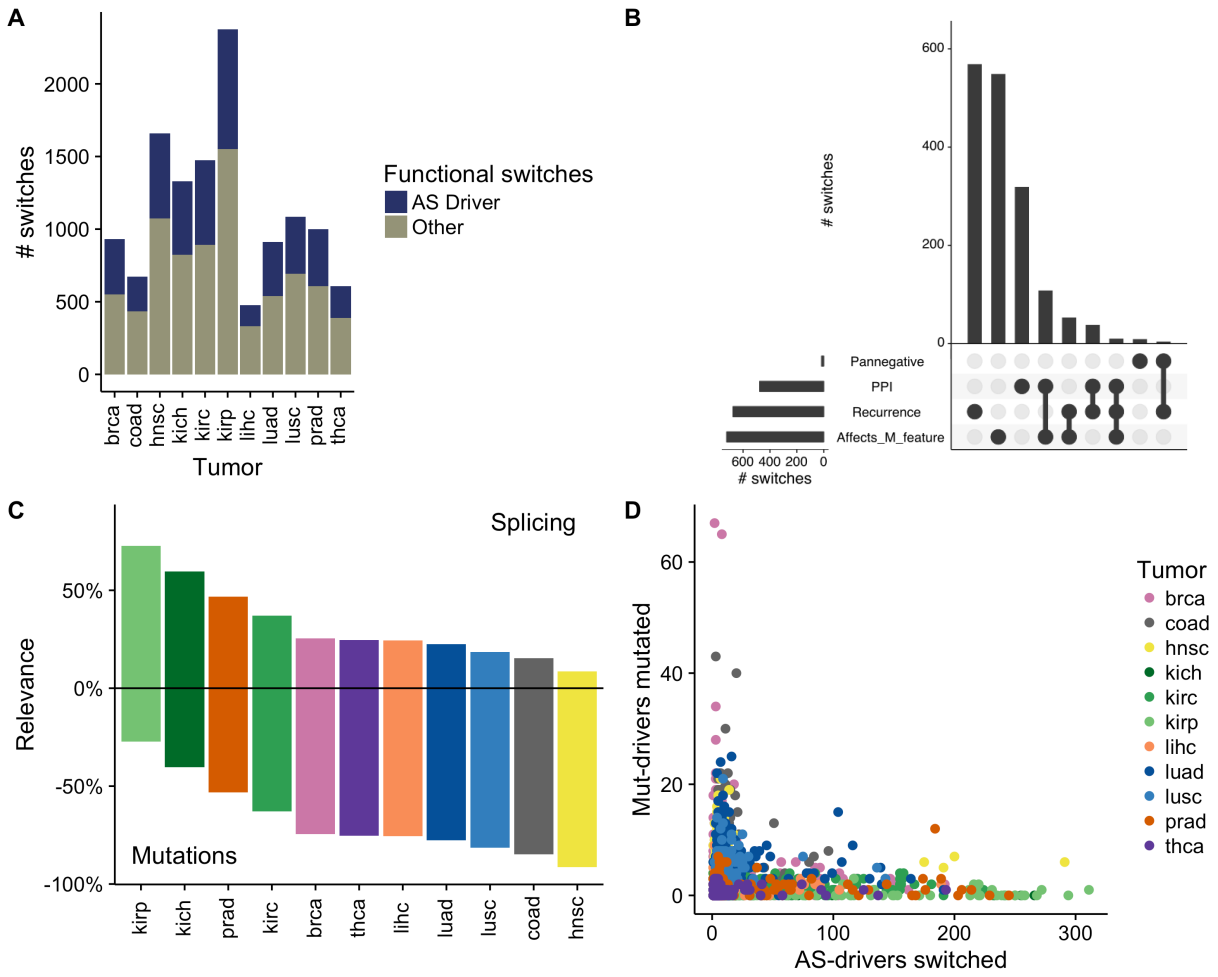



**Figure 4. (A)** Number of functional switches and AS-drivers detected in each tumor type (see text for definitions). **(B)** Candidate AS-drivers grouped according to their properties: disruption of protein–protein interactions (PPIs), significant recurrence across patients (Recurrence), gain or loss of a protein feature that is frequently mutated in tumors (Affects M_feature), mutual exclusion and sharing pathway with cancer drivers (Pannegative). 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 (Conway et al., 2017). **(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).

# METHODS

| Datasets and databases | Source | Identifier |
|---|---|---|
| Datasets analysed (observed switches, functional implications, links to mutational data, etc.) | This paper | DOI: 10.5281/zenodo.439846 (https://zenodo.org/record/439846) |
| Human reference genome hg19 assembly | Genome Reference Consortium | http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/ |
| TCGA level 3 data for RNA-seq (read counts for isoforms), mutation and copy number variation (CNV) data for brca, coad, hnsc, kich, kirc, kirp, lihc, luad, lusc, prad, thca | TCGA data portal | https://gdc-portal.nci.nih.gov/ |
| Mutations from whole genome sequencing for 306 samples from brca, coad, hnsc, kich, kirc, luad, lusc, prad, thca | (Fredriksson et al., 2014) | https://www.synapse.org/#!Synapse:syn2882200 |
| List of cancer drivers per tumor type for brca, coad, hnsc, kirc, lihc, luad, lusc, prad, thca | (Gundem et al., 2010) | https://www.intogen.org/ |
| Cancer drivers for papillary renal-cell carcinoma (kirp) | (The Cancer Genome Atlas Research Network, 2016) | DOI:10.1056/NEJMoa1505917 |
| Cancer drivers for chromophobe renal-cell carcinoma (kich) | (Davis et al., 2014) | DOI:10.1016/j.ccr.2014.07.014 |
| COSMIC: list of oncogenes and tumor suppressors per tumor type | (Forbes et al., 2015), | http://cancer.sanger.ac.uk/cosmic |
| TSGene: database of tumor suppressors | (Zhao et al., 2015) | https://bioinfo.uth.edu/TSGene/ |
| Pfam: protein domain families | (Finn et al., 2016) | https://www.ebi.ac.uk/services/teams/pfam |
| ProSite: protein domain patterns | (Gattiker et al., 2002) | http://prosite.expasy.org/ |
| ArchDB: database of protein loops | (Bonet et al., 2014). | http://sbi.imim.es/archdb/ |

| | | |
|---|---|---|
| PSICQUIC: database of protein-protein interactions | (del-Toro et al., 2013) | https://www.ucl.ac.uk/functional-gene-annotation/psicquic |
| BIOGRID: database of protein-protein interactions | (Chatr-Aryamontri et al., 2015) | https://thebiogrid.org/ |
| HumNet: database of protein-protein interactions | (Lee et al., 2011) | http://www.functionalnet.org/humannet/about.html |
| STRING: database of protein-protein interactions | (Szklarczyk et al., 2011) | http://string-db.org/ |
| Dataset of protein-protein interactions | (Rolland et al., 2014) | DOI: http://dx.doi.org/10.1016/j.cell.2014.10.050 |
| iPfam: database of domain-domain interactions | (Finn et al., 2014) | http://ipfam.org/ |
| DOMINE: database of domain-domain interactions | (Raghavachari et al., 2008) | http://domine.utdallas.edu/cgi-bin/Domine |
| 3did: database of domain-domain interactions | (Mosca et al., 2014) | http://3did.irbbarcelona.org/ |
| Software and algorithms | Source | Identifier |
| Software to calculate isoform switches and AS-drivers | This paper | https://bitbucket.org/regulatorygenomicsupf/smartas/ |
| The software to reproduce the analyses of switches and AS-drivers. | This paper | https://github.com/hclimente/smartas |
| SUPPA: software for the calculation of alternative splicing events | (Alamancos et al., 2015) | https://github.com/comprna/SUPPA |
| Domain-centric analysis of Gene Ontologies | (Fang and Gough, 2013) | http://supfam.org/SUPERFAMILY/dcGO/ |
| OncodriveROLE: method to predict whether a cancer gene driver is an oncogene or a tumor suppressor | (Schroeder et al., 2014) | http://bg.upf.edu/oncodrive-role |
| ESTIMATE: method to measure the stromal and immune cell content in a sample | (Yoshihara et al., 2013) | https://sourceforge.net/projects/estimateproject / |
| IUPred: prediction of protein disordered regions | (Dosztanyi et al., 2005) | http://iupred.enzim.hu/ |

| | | |
|---|---|---|
| ANCHOR: prediction of disordered regions with potential for protein-protein interactions | (Dosztanyi et al., 2009) | http://anchor.enzim.hu/ |
| R 3.3 | The R project | https://www.r-project.org/ |
| intergraph 2.0-2 - R library for network analysis | Integraph | https://cran.r-project.org/web/packages/intergraph/index.html |
| igraph 1.0.1<br>-R library for network analysis | (Blondel et al., 2008) | http://igraph.org/r/ |
| network 1.13.0 - R library for network analysis | (Butts, 2008) | https://cran.r-project.org/web/packages/network/index.html |
| ggnetwork 0.5.1<br>- R library for network plotting | Ggnetwork | https://github.com/briatte/ggnetwork |
| cowplot 0.7.0 - R library for plotting | Cowplot | https://github.com/wilkelab/cowplot |
| ggplot2 2.2.1- R library for plotting | ggplot2 | http://ggplot2.org/ |
| UpSetR - R library for plotting | (Conway et al., 2017). | https://cran.r-project.org/package=UpSetR |
| tidyverse 1.0.0 - R library for data manipulation | Tidyverse | http://tidyverse.org/ |
| magrittr 1.5 - R library for code development | Magrittr | https://cran.r-project.org/package=magrittr/vignettes/magrittr.html |
| R scripts from wisdom repository | This paper | https://github.com/hclimente/wisdom |
| find.me - R library for plotting | This paper | https://github.com/hclimente/find.me |
| ggstars – R library for plotting | This paper | https://github.com/hclimente/ggstars |
| Anaconda 4.3.1 - Python libraries for large-scale data processing | Continuum Analytics | https://www.continuum.io/why-anaconda |
| NetworkX: Python software package for the study of networks. | (Hagberg et al., 2008) | https://networkx.github.io |

**Data**

Estimated RNA sequencing (RNA-seq) read counts per transcript isoform were obtained from the TCGA data portal (https://gdc.nci.nih.gov/) for a total of 4442 samples for 11 tumor types: 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). Only transcripts with expression TPM > 0.1 were considered. Tumor specific mutational and copy-number alteration drivers were collected from Intogen (Gundem et al., 2010) and from the TCGA papers for kidney chromophobe (kich) (Davis et al., 2014) and kidney renal papillary carcinoma (kirp) (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. These genes were annotated as oncogenes or tumor suppressors using the annotations provided by COSMIC (Forbes et al., 2015), Vogelstein et al. (Vogelstein et al., 2013), and by the TSGene database (Zhao et al., 2015). Unlabeled cases were predicted with OncodriveROLE (Schroeder et al., 2014) using cutoffs 0.3 (loss-of-function class) and 0.7 (activating class).

**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 each transcript, the relative abundance per sample was calculated by normalizing its abundance in TPM units by the sum of abundances of all transcripts in the same gene, which we called PSI. Then, for each transcript and sample we calculated the change in relative abundance as $\Delta PSI = PSI_{tumor} - PSI_{ref}$, where $PSI_{tumor}$ is the PSI value of the transcript in a tumor sample and $PSI_{ref}$ is the normal reference value, which corresponds to the value in the paired normal sample when available or to the median of the PSI distribution in the normal samples of the same tissue type, otherwise. We considered significant those changes with $|\Delta PSI| > 0.05$ in the comparison between normal and tumor samples and with empirical $p < 0.01$ in the comparison of the observed $|\Delta PSI|$ value with the distribution of $|\Delta PSI|$ values obtained by comparing the normal samples pairwise without repetition. We only kept those

cases for which the tumor isoform PSI was higher than the normal isoform in the tumor sample, and the normal isoform PSI in the normal sample was higher than the value for the tumor isoform. Moreover, we discarded genes that either had an outlier expression in the tumor sample compared to normal tissues – had expression below the bottom 2.5% or above the 97.5% of the values of normal expression – or showed differential expression between the tumor samples with the switch and the normal samples (Wilcoxon test p-value < 0.01 using the gene TPM values).

Candidate switches were defined per patient and per gene. In some samples, different switches could appear in the same gene. 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. Additionally, we filtered out switches that were significantly lowly recurrent, i.e. they occurred in fewer patients than expected by chance (adjusted p-value < 0.05 binomial test using all tumor types). This imposed extra restrictions in the switches and none of the reported switches occured in less than 5 samples. 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 a minimum number of 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.

**Comparison with stromal and immune signatures**

To determine whether the observed switches merely reflected the cellular content of the samples, we measured the significant association with stromal and immune cell content using ESTIMATE (Yoshihara et al., 2013). For each switch we performed a Wilcoxon test to compare the ESTIMATE scores between patients with and without the switch. After correcting for multiple testing (Benjamini-Hochberg method), we found 1108 and 473 exclusively associated (FDR < 0.05) with stromal and immune cell content, respectively; and 306 associated with both. These were eliminated from the final set of isoform switches (Table S1).

**Relation between transcript isoform switches and local alternative splicing events**

We calculated from the annotation the possible local alternative splicing events of type alternative 3' (A3) and 5' (A5) splice-site, intron retention (RI), exon skipping (SE), mutually exclusive exons (MX), alternative first exon (AF) and alternative last exon (AL) using SUPPA (Alamancos et al., 2015). SUPPA provides for each alternative splicing event the set of transcript isoforms that contribute to either form of the event. We thus were able to determine whether each pair of isoforms describing a switch corresponded to one or more local alternative splicing events, and which of the two forms of the event corresponded to the tumor and the normal isoform. For instance, we calculated whether an isoform switch describing an exon cassette (SE) event corresponded to an increase or decrease of exon inclusion in the tumor sample. Accordingly, if the tumor isoform contained the alternative exon and the normal isoform did not contain it, the event would correspond to inclusion in tumor. Similarly, if the tumor isoform did not have the exon but the normal isoform did, the event would indicate skipping in the tumor sample.

**Recurrence**

We defined the total number of unique different switches as $S$ and the number of patients with switches as $P$, and the total number of switches occurring in patients as $N$. Thus, the expected frequency of a switch is $f = N/(S \cdot P)$. For a given switch, we then tested the significance of its recurrence across patients in each tumor type 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 the normal samples 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 was a change in the encoded protein, including cases where only one of the isoforms has a coding DNA sequence (CDS), and moreover there was a gain or loss of a protein feature: Pfam domains (Finn et al., 2016) 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). Switches between a protein-coding isoform and non-coding isoform or between two protein-coding isoform for which we could not map any protein feature on the proteins were not be labelled as functional.

**Domain families enriched in switches or mutations**

To determine which protein domain families were significantly affected by switches, we first calculated a reference proteome for each tumor type. We selected genes that had an isoform with at least 0.1 TPM, and from each of these genes we only kept the isoform with the highest median expression across the normal samples of the same tissue type. The proteins encoded by these isoforms were considered the reference proteome in each tumor type. We then aggregated the reference proteomes from all tumor types to form a pan-cancer reference proteome using only genes annotated with multiple transcripts. The expected frequency $f(a)$ of a protein feature $a$ that appears $m(a)$ times was then measured as the proportion of this feature in the pan-cancer representative proteome:

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

where $b$ runs over all protein features in the reference proteome. We then calculated the expected probability of a protein feature to be affected by a switch 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. Additionally, to ensure the specificity of the enrichment for each domain class, we considered only domain families affected in 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 of a domain family was considered to be the proportion of the proteome length it covered. 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. GO enrichment analysis was performed using DcGO (Fang and Gough, 2013). We considered significant those cases with FDR < 0.01 (hypergeometric test).

**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 an isoform switch, we first mapped each PPI in a gene to a specific domain–domain interaction (DDI). We used information on domain–domain interactions from iPfam (Finn et al., 2014), DOMINE (Raghavachari et al., 2008), and 3did (Mosca et al., 2014). Domains involved in DDIs were then mapped to specific protein isoforms. For the genes with switches, we then considered those PPIs that could be mapped to DDIs involving domains mapped in either the normal or the tumor isoforms. 3,242 genes with 4,219 switches mapped to one or more interactions in the consensus network, and 1,688 isoform switches (in 1,355 genes) were mapped to at least one specific DDI. We defined 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. If multiple domains mediated the same interaction, it was considered lost if at least one

of these domains was lost in the switch. On the other hand, we defined a PPI as gained if it was mapped to a DDI only in the tumor isoform but not in the normal isoform. For each switch from Table S1 in each tumor type we calculated its centrality in the consensus PPI network and the distance to the closest tumor-specific cancer gene driver using NetworkX (Hagberg et al., 2008).

**Analysis of the interaction network affected by switches**

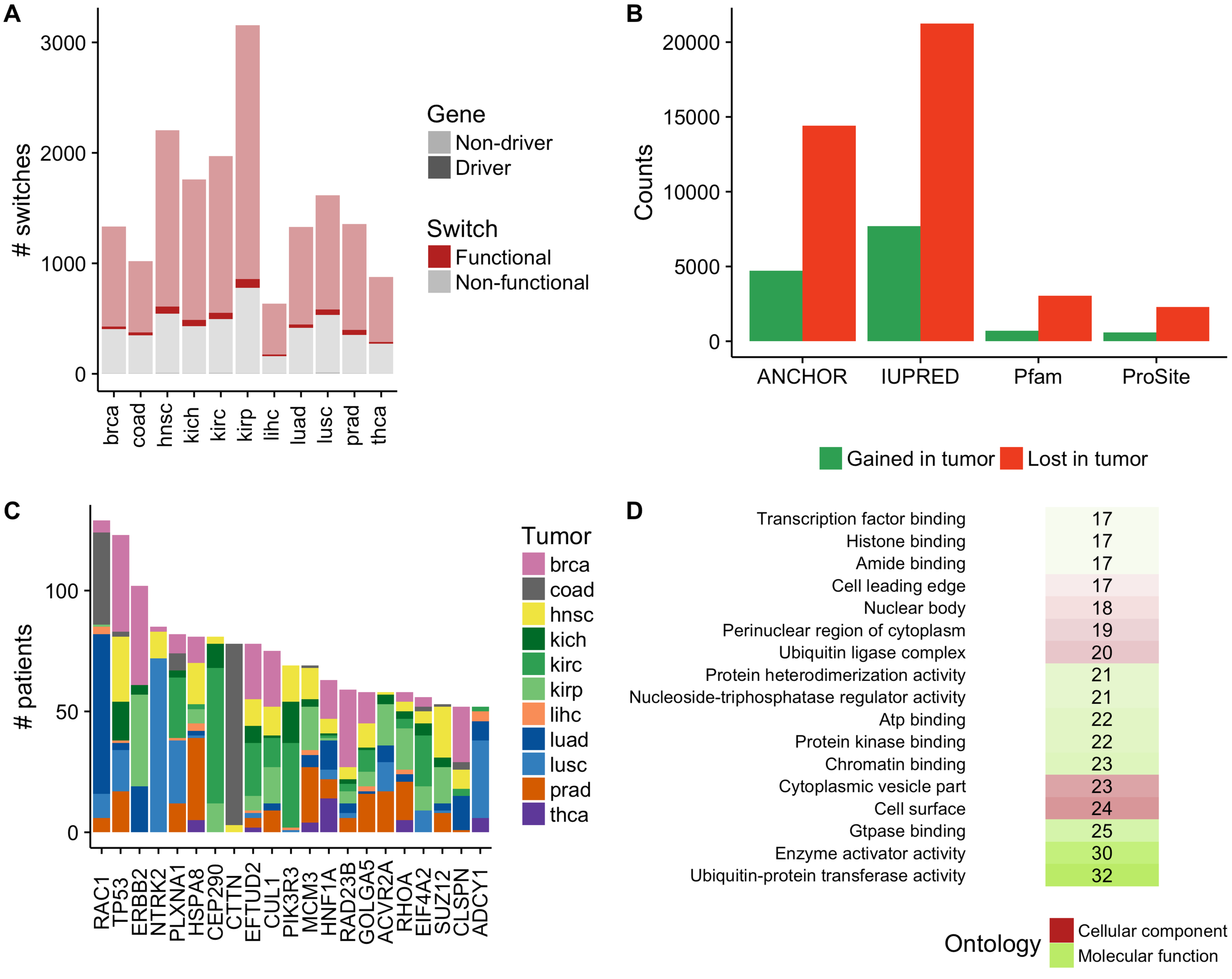
We considered gene sets consisting of functional and cancer-related pathways (Liberzon et al., 2015), protein complexes (Ruepp et al., 2009) and complexes related to RNA metabolism (Akerman et al., 2015). We calculated the enrichment of PPI-affecting switches in each gene set using a Fisher's exact test based on the separation of switches into being in the gene set or not, and affecting PPIs or not (Table S5). Additionally, considering the network formed by the PPIs between genes that are either gained or lost through an isoform switch, i.e. we only used the connections that are lost or gained, 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 any of the modules 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 (Table S6).

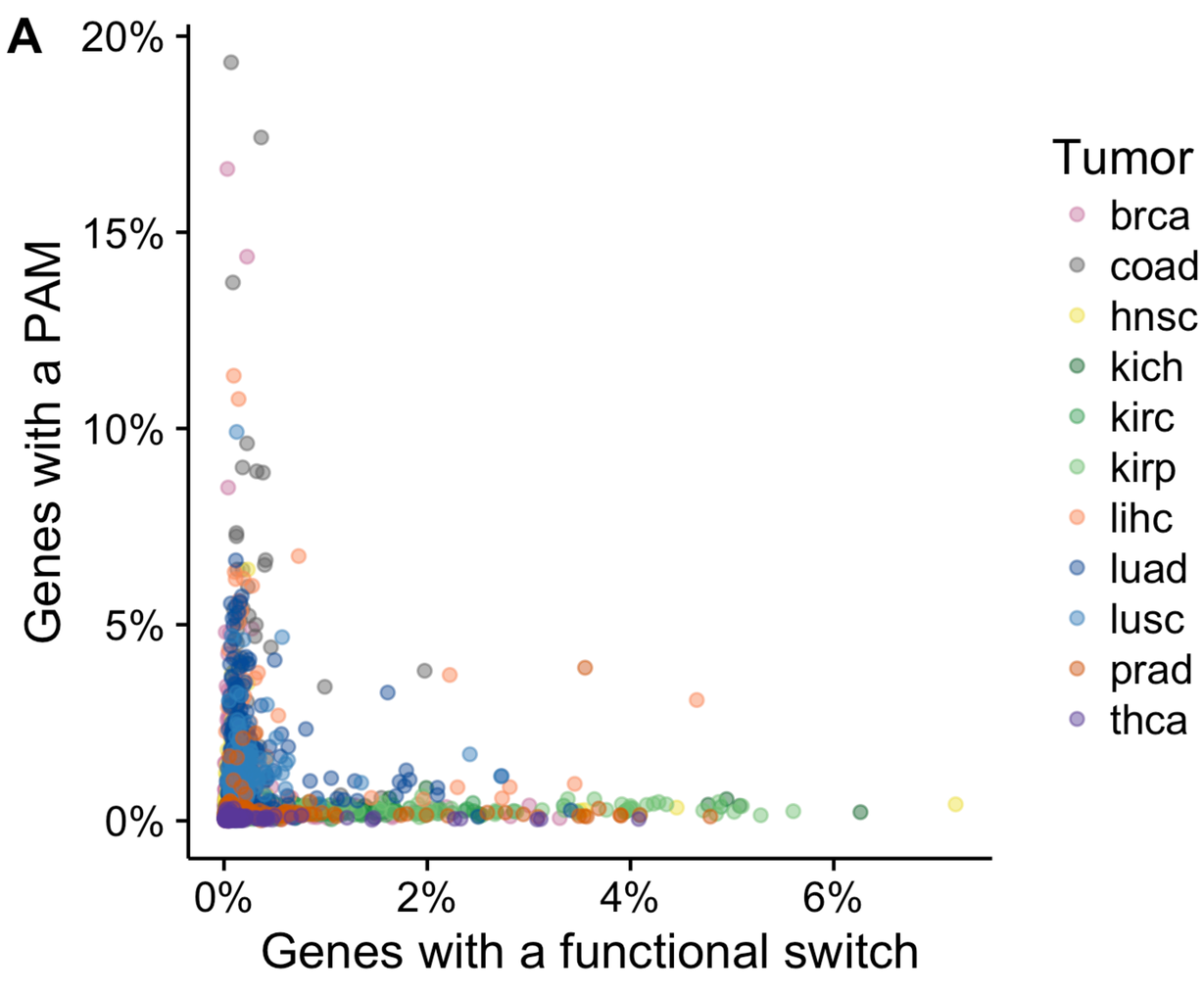
**Mutation and copy number analysis**

Mutation information was downloaded from the TCGA data portal for all tumor samples in the form of MAF files containing Level 2 somatic mutation calls from whole exome data. Additionally, we used somatic mutations from whole genome sequence (WGS) data (Fredriksson et al., 2014) for 306 of the samples studied. For copy number alterations (CNAs), as done before (Sebestyén et al., 2016), we used CNA regions overlapping at least the full gene locus. We considered a CNA loss if the score was smaller than $\log_2(1/2)$, which means at least 1 copy is lost; and a CNA gain, if the score was larger than $\log_2(3/2)$, which means at least 1 copy is gained.

To measure the association between switches and mutations we measured a Jaccard score. For each gene with a switch, given the number of patients with only switches (*S*), only mutations (*M*) or both (*MS*), the Jaccard score was defined as $MS/(M+S+MS)$. The Jaccard score calculation was carried out using protein-affecting mutations (PAMs) for WES datasets, for all mutation types for WGS datasets. In each case we only used patients that had RNA-seq and mutation data and we compared the splicing pattern of the patient with its own mutation information. For WGS, 306 patients from 8 of the 11 tumor types considered had mutation and RNA-seq data, whereas for WES data, 3755 patient samples from all 11 tumor types analysed had mutation and RNA-seq data.

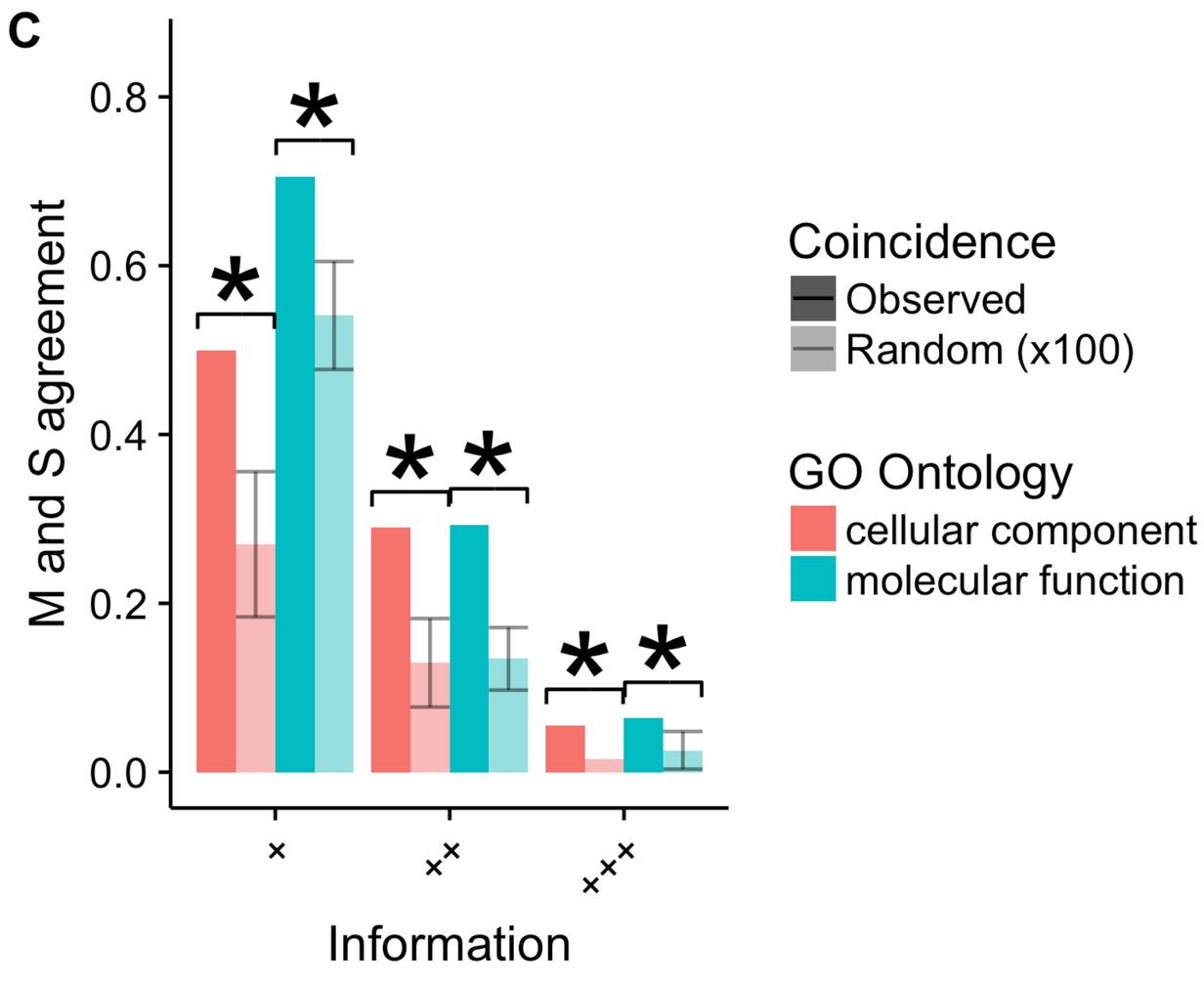
To identify switches significantly associated with pan-negative tumors, we considered the top 10 drivers according to their frequency of protein-affecting mutations (PAMs) in each tumor type. We tested the mutual exclusion between the patients affected by the switch and the patients with a PAM in at 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 with which they were also mutually exclusive according to the same test.

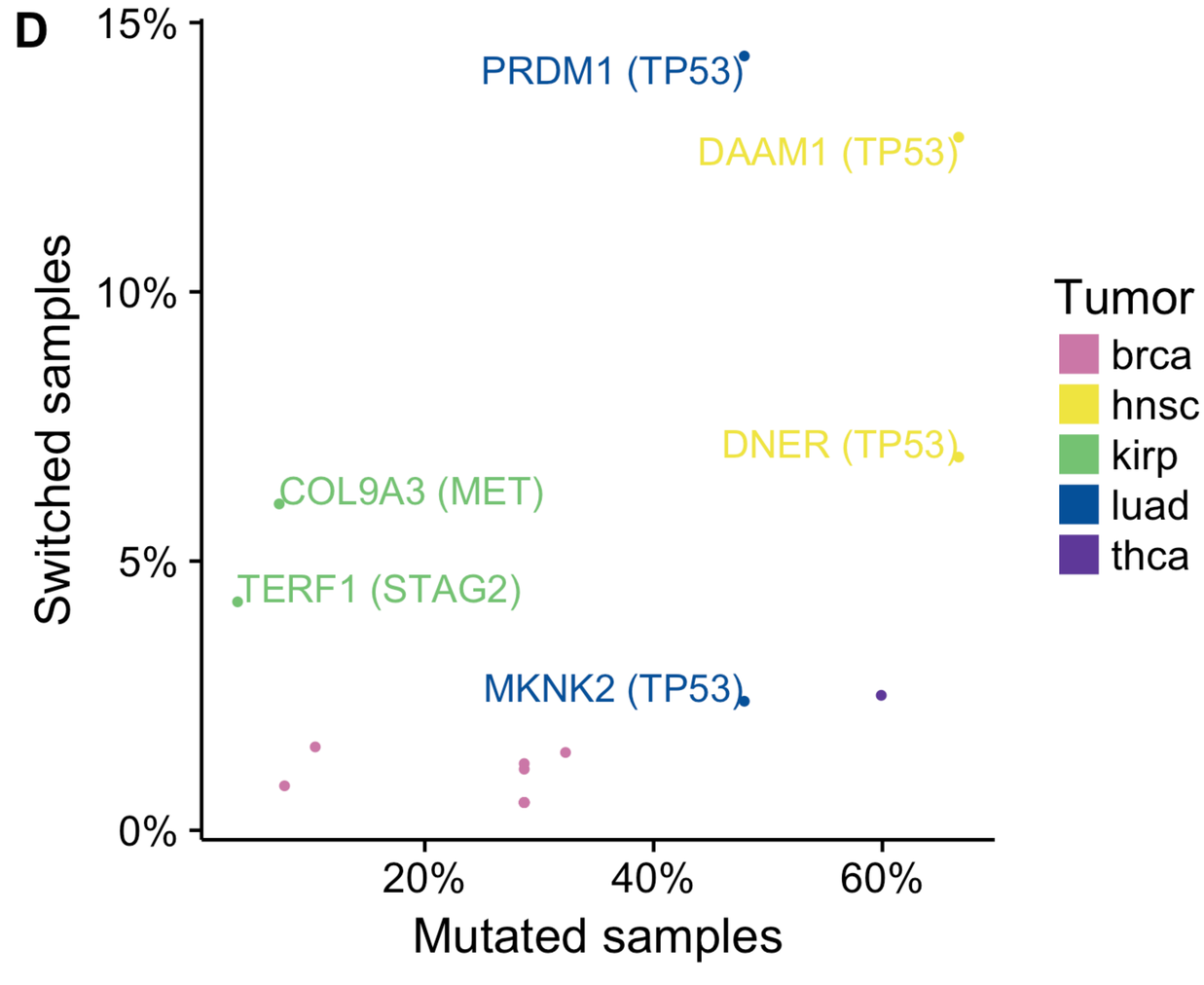

A
# switches
3000
2000
1000
0
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
Gene
Non-driver
Driver
Switch
Functional
Non-functional
B
Counts
20000
15000
10000
5000
0
ANCHOR
IUPRED
Pfam
ProSite
Gained in tumor
Lost in tumor
C
# patients
100
50
0
RAC1
TP53
ERBB2
NTRK2
PLXNA1
HSPA8
CEP290
CTTN
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
Transcription factor binding 17
Histone binding 17
Amide binding 17
Cell leading edge 17
Nuclear body 18
Perinuclear region of cytoplasm 19
Ubiquitin ligase complex 20
Protein heterodimerization activity 21
Nucleoside-triphosphatase regulator activity 21
Atp binding 22
Protein kinase binding 22
Chromatin binding 23
Cytoplasmic vesicle part 23
Cell surface 24
Gtpase binding 25
Enzyme activator activity 30
Ubiquitin-protein transferase activity 32
Ontology
Cellular component
Molecular function

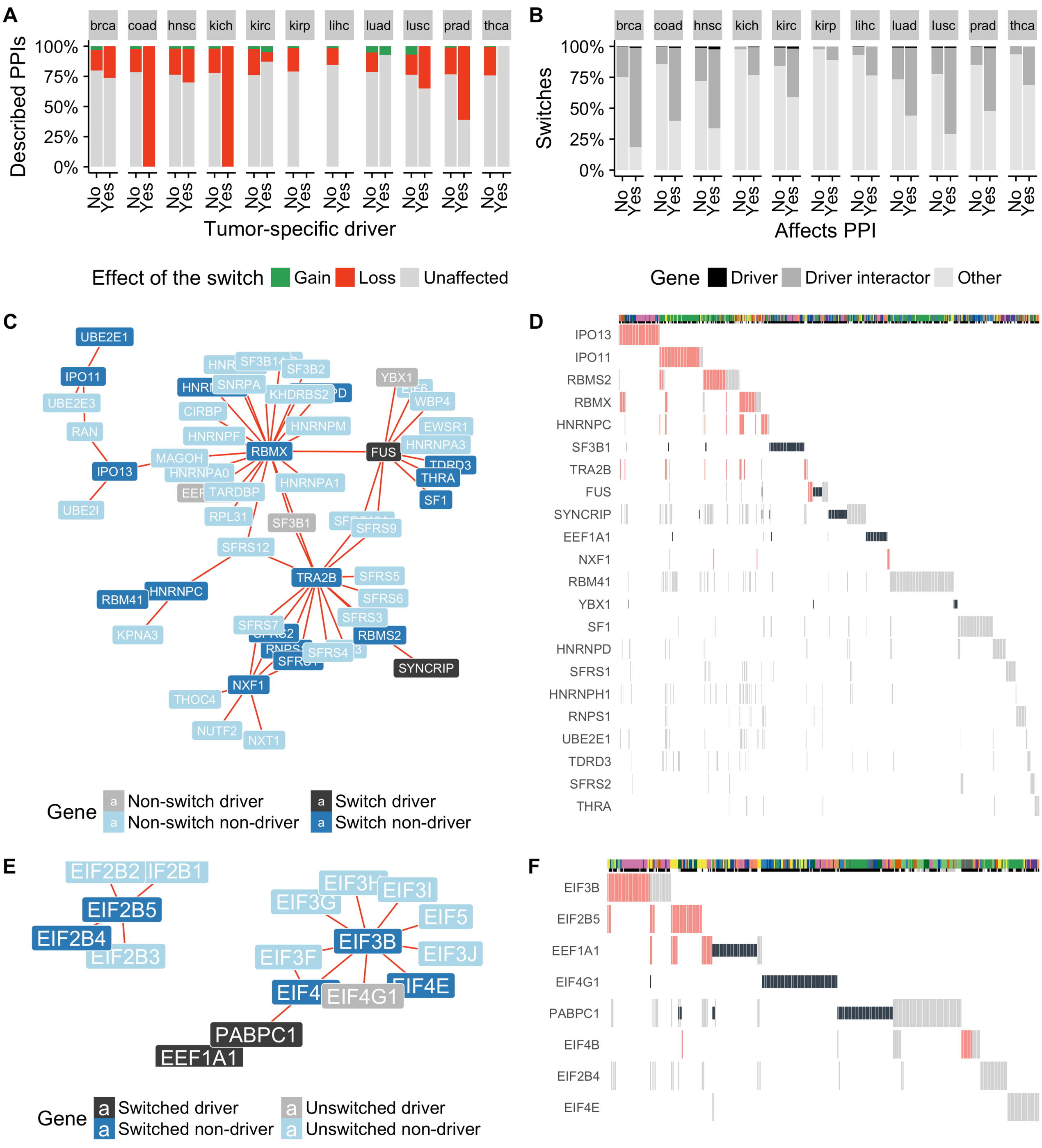


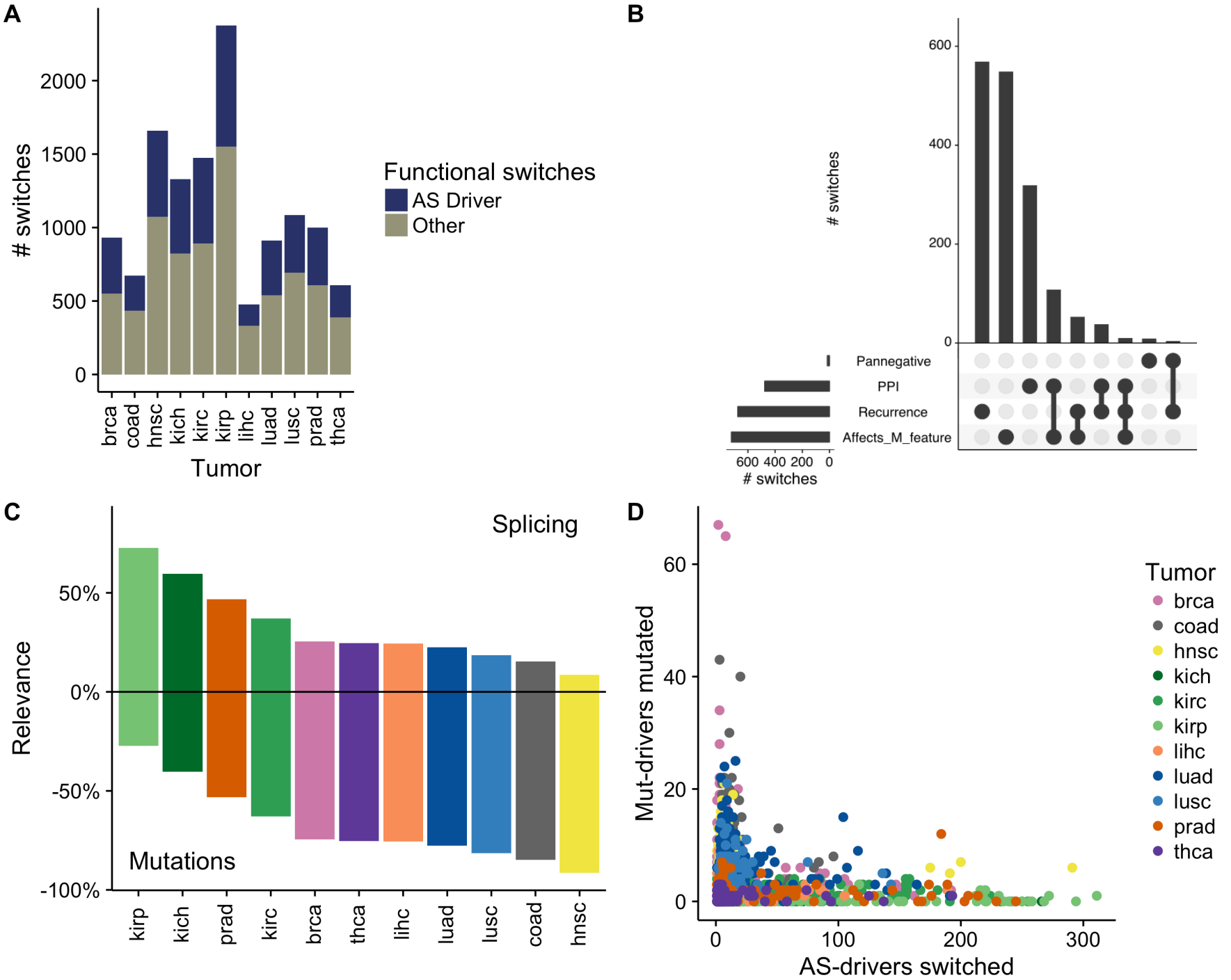

B

| | | |
|---|---|---|
| Calcium-binding EGF domain | 14 | Losses |
| TB domain | 5 | |
| IPT/TIG domain | 4 | |
| Class II histocompatibility antigen, beta domain | 4 | |
| C2 domain of PTEN tumour-suppressor protein | 3 | |
| CD20-like family | 2 | |
| Matrixin | 2 | |
| Xylosyltransferase C terminal | 1 | |
| P53 tetramerisation motif | 1 | |
| P53 DNA-binding domain | 1 | |
| Fibronectin type III domain | 10/22 | Both |
| Immunoglobulin I-set domain | 11/30 | |
| Neurotransmitter-gated ion-channel ligand binding domain | 2 | Gains |
| Core histone H2A/H2B/H3/H4 | 2 | |
| Reprolysin (M12B) family zinc metalloprotease | 3 | |
| Ras family | 3 | |
| Cadherin domain | 3 | |
| Thrombospondin type 1 domain | 5 | |




A
Described PPIs
100%
75%
50%
25%
0%
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca
No
Yes
Tumor-specific driver
Effect of the switch
Gain
Loss
Unaffected
B
Switches
Affects PPI
Gene
Driver
Driver interactor
Other
C
UBE2E1
IPO11
UBE2E3
RAN
IPO13
UBE2I
HNRNPM
HNRNPF
RBMX
MAGOH
HNRNPA0
HNRNPA1
TARDBP
RPL31
SF3B1
SF3B2
SNRPA
KHDRBS2
CIRBP
YBX1
WBP4
EWSR1
HNRNPA3
FUS
THRA
SF1
SFRS9
SFRS12
TRA2B
SFRS5
SFRS6
SFRS3
RBMS2
SYNCRIP
HNRNPC
RBM41
KPNA3
SFRS7
SFRS4
NXF1
THOC4
NUTF2
NXT1
Gene
Non-switch driver
Non-switch non-driver
Switch driver
Switch non-driver
D
IPO13
IPO11
RBMS2
RBMX
HNRNPC
SF3B1
TRA2B
FUS
SYNCRIP
EEF1A1
NXF1
RBM41
YBX1
SF1
HNRNPD
SFRS1
HNRNPH1
RNPS1
UBE2E1
TDRD3
SFRS2
THRA
E
EIF2B2
EIF2B1
EIF2B5
EIF2B4
EIF2B3
EIF3H
EIF3I
EIF3G
EIF5
EIF3B
EIF3F
EIF3J
EIF4G1
EIF4E
PABPC1
EEF1A1
Gene
Switched driver
Switched non-driver
Unswitched driver
Unswitched non-driver
F
EIF3B
EIF2B5
EEF1A1
EIF4G1
PABPC1
EIF4B
EIF2B4
EIF4E


A
# switches
2000
1500
1000
500
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
600 400 200 0
# switches
C
Splicing
Relevance
50%
0%
-50%
-100%
Mutations
kirp
kich
prad
kirc
brca
thca
lihc
luad
lusc
coad
hnsc
D
Mut-drivers mutated
60
40
20
0
0
100
200
300
AS-drivers switched
Tumor
brca
coad
hnsc
kich
kirc
kirp
lihc
luad
lusc
prad
thca