Supplemental Material

The functional impact of alternative splicing in cancer

Héctor Climente-González[1,2,3,4], Eduard Porta-Pardo[5,6], Adam Godzik[5], Eduardo Eyras[1,7]

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

[2] MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 77300 Fontainebleau, France

[3] Institut Curie, 75248 Paris Cedex, France

[4] INSERM U900, 75248 Paris Cedex, France

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

[6] Barcelona Supercomputing Centre (BSC), E08034 Barcelona

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

Lead contact: eduardo.eyras@upf.edu

Supplemental Experimental Procedures

Analyzed 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 publications 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).

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 switches 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 available in Table S1.

Relation between transcript isoform switches and local alternative splicing events

Using SUPPA (Alamancos et al., 2015; Trincado et al., 2016), we calculated 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). 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

Given the total number of unique switches, S, the number of patients with one or more switches, P, and the total number of switches occurring in patients as N, the expected frequency of a switch was estimated as $f = N/(S \cdot P)$. We tested the significance of recurrence across patients in each tumor type using a binomial test with the observed patient count n and the expected frequency f:

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

Switches were considered significantly recurrent for an adjusted binomial test p-value < 0.05. On the other hand, we filtered out switches that were significantly lowly recurrent, i.e. they occurred in fewer patients than expected by chance. To measure this, we used the same test as above for recurrence. The switch was significantly lowly recurrent if $1\text{-}P(n)$ was significant and the expected frequency, $f = N/(S \cdot P)$ as defined above, was higher than the observed frequency, n/P, where n is the number of patients with that switch, and P is the number of patients with one or more switches. The cutoff for significance was 0.05 after adjusting for multiple testing, using all tumor types.

Functional Switches

From the 8,122 different switches found, for 6,937 (85,41%) of them both isoforms had an annotated protein, for 9.01% only the normal isoform had an annotated protein, and for 2.8% only the tumor isoform had an annotated protein (Table S1). A switch was defined as functional if both isoforms overlapped in genomic extent, i.e. shared transcribed locus, there was a change in the encoded protein (including cases where only one of the isoforms was coding) 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). For IUPred and ANCHOR we only considered changes involving at least 5 amino acids. Switches for which we could not map any protein feature were not considered functional despite the possible difference in coding sequences. Significance on the enrichment of protein features losses versus gains was calculated by comparing the number of gains and losses in switches with the numbers in simulated switches (SS): IUPRED (gains:7702, losses: 14425, SS-gains: 3401162, SS-losses: 6858382), ANCHOR (gains: 4707, losses: 14425, SS-gains: 2215263, SS-losses: 4512854),

Pfam (gains: 699, losses: 3052, SS-gains: 485611, SS-losses: 1421408), ProSite (gains: 605, losses: 2296, SS-gains: 331247, SS-losses: 975136).

Enrichment of Domain families in switches and mutations

To determine protein domain families significantly affected by switches we first calculated a reference proteome for each tumor type. Using genes with multiple transcripts, we selected those that had at least one isoform with TPM>1, and only kept the isoform with the highest median expression across the normal samples in the same tissue type. The proteins encoded by these isoforms were considered the reference proteome in each tumor type. We aggregated the reference proteomes from all tumor types to form a pan-cancer reference proteome. The expected frequency *f(a)* of a protein feature *a*, e.g. a Pfam domain family, that appears *m(a)* times was then measured as the proportion of this feature in the reference proteome:

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

where *b* runs over all protein features in the reference proteome, e.g. all Pfam domain families. 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 times feature *a* was gained or lost in switches and *n* is the total number of feature 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 *a* covering the proteome a number of amino acids *n(a)* was considered to be the proportion of this coverage:

$$g(a) = \frac{n(a)}{\sum_b n(b)}$$

where *b* runs over all protein features in the reference proteome. We aggregated all observed mutations falling within each domain family and calculated the expected probability of the observed mutations using a binomial test as:

$$P(a) = \frac{N!}{n!(N-n)!} g(a)^n (1 - g(a))^{N-n}$$

where now *n* is the number of mutations falling in domain family *a*, and *N* is the total number of mutations falling in all domain families considered. 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).

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 the 11 tumor types analyzed had mutation and RNA-seq data.

We also tested mutual exclusion of our isoform switches and 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 tested mutual exclusion between functional switches with individual mutational drivers in the same functional pathway using the same test. These results are provided in Table S3 before

multiple test correction. After multiple test correction none of these cases showed significant mutual exclusion.

Potential impact of isoform switches in protein interactions with cancer drivers

Functional switches were divided according to whether they occurred in tumor-specific drivers or not. For each tumor type we then calculated the proportion of protein-protein interactions (PPIs) that were gained, lost, or remained unaffected, and performed a Chi-Square test comparing the proportions for the tumor-specific drivers and the rest of genes. Individual Chi-square tests for each tumor type: BRCA p-val = 0.001, COAD p-val = 1.3e-17, KICH p-val = 6.4e-33, LUSC p-val = 1.1e-7, PRAD p-val = 1.5e-12. The tumor types KIRC, LUAD and THCA showed no significance. Samples from KIRP and LIHC had no PPI-affecting switches in drivers.

We further divided functional switches mapped to PPIs according to whether they affected a PPI or not. For each tumor type we calculated the proportion of functional switches that occurred in cancer drivers, in interactors of drivers, or in other genes, and calculated a Fisher's exact test comparing the PPIs affected by switches in driver-interactors and in other genes mapped to PPIs (non-drivers and non-driver-interactors). All cases were significant except for KIRC, LUAD and LUSC: BRCA p-val=1.05e-09, OR= 2.2; COAD p-val=4.5-21, OR=1.1; HNSC p-val=9.9e-02, OR=1.1; KICH pval=1.6e-35, OR=7.5, KIRC p-val=5.4e-01.3e-41, OR=1.07; KIRP p-val=2.e-21, OR=2.7; LIHC p-val=1.08e-28, OR=8.0; LUAD p-val=6.9e-01, OR=1.07; LUSC p-val=1, OR=1; PRAD, p-val=8.0e-06, OR=1.6; THCA, p-val=1.3e-27, OR=4.2.

Module and gene-set 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).

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, 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).

Calculation of driver-like properties

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). We thus calculated these properties for switches predicted as candidate AS-drivers. For each switch from Table S1 in each tumor type we calculated the centrality in the consensus PPI network using the degree_centrality() function from NetworkX (https://networkx.github.io) (Hagberg et al. 2008). We then compared the distribution of centrality values for switches described as AS-drivers and for the rest of functional switches, together (Mann-Whitney test p-value < 2.2e-16) or separating candidate AS-drivers according to their properties (see Figure S6A). We also considered for each switch from Table S1 (AS-driver or not) in each tumor type, what is the distance to the closest tumor-specific gene cancer driver, which we call closest driver distance (CDD). Every switch with CDD<=3 we labelled it as “Close to a driver”, or “Far from a driver” otherwise. With this, we calculated the proportion of AS-drivers and switches non-drivers according to the CDD.

RESOURCE TABLE

Resource	Source	Identifier
Data		
Datasets analyzed (observed switches, functional implications, links to mutational data, etc.)	This paper	https://zenodo.org/record/824637
Supplementary tables with tab separated values (.tsv format)	This paper	https://github.com/hclimente/smartas/tree/master/results/supplementary_files
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,	(Gundem et al., 2010)	https://www.intogen.org/

LUSC, PRAD, THCA		
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/
Functional Pathways and Gene sets - Molecular Signatures Database (MSigDB)	(Liberzon et al., 2015)	http://software.broadinstitute.org/gsea/msigdb/index.jsp
Protein complexes (CORUM)	(Ruepp et al., 2009)	http://mips.helmholtz-muenchen.de/genre/proj/corum/index.html
Complexes related to RNA metabolism	(Akerman et al., 2015)	https://www.ncbi.nlm.nih.gov/pubmed/26047612
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	(Bonet et al.,	http://sbi.imim.es/archdb/

loops	2014).	
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		
Software to calculate isoform switches described in this work	This paper	https://bitbucket.org/regulatorygenomicsupf/smartas/
The software to reproduce the analyses carried out on isoform switches	This paper	https://github.com/hclimente/smartas
SUPPA: software for the	(Alamancos et al.,	https://github.com/comprna/SUPPA

calculation of alternative splicing events	2015)	
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/estimate project /
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/
iGraph: R package to find community structures in networks	(Blondel et al., 2008)	http://igraph.org/r/doc/cluster_louvain.html
NetworkX: software package for the study of networks.	(Hagberg et al., 2008)	https://networkx.github.io
UpSetR: software for plotting intersecting sets.	(Conway et al., 2017).	https://cran.r-project.org/package=UpSetR

Supplemental Figures

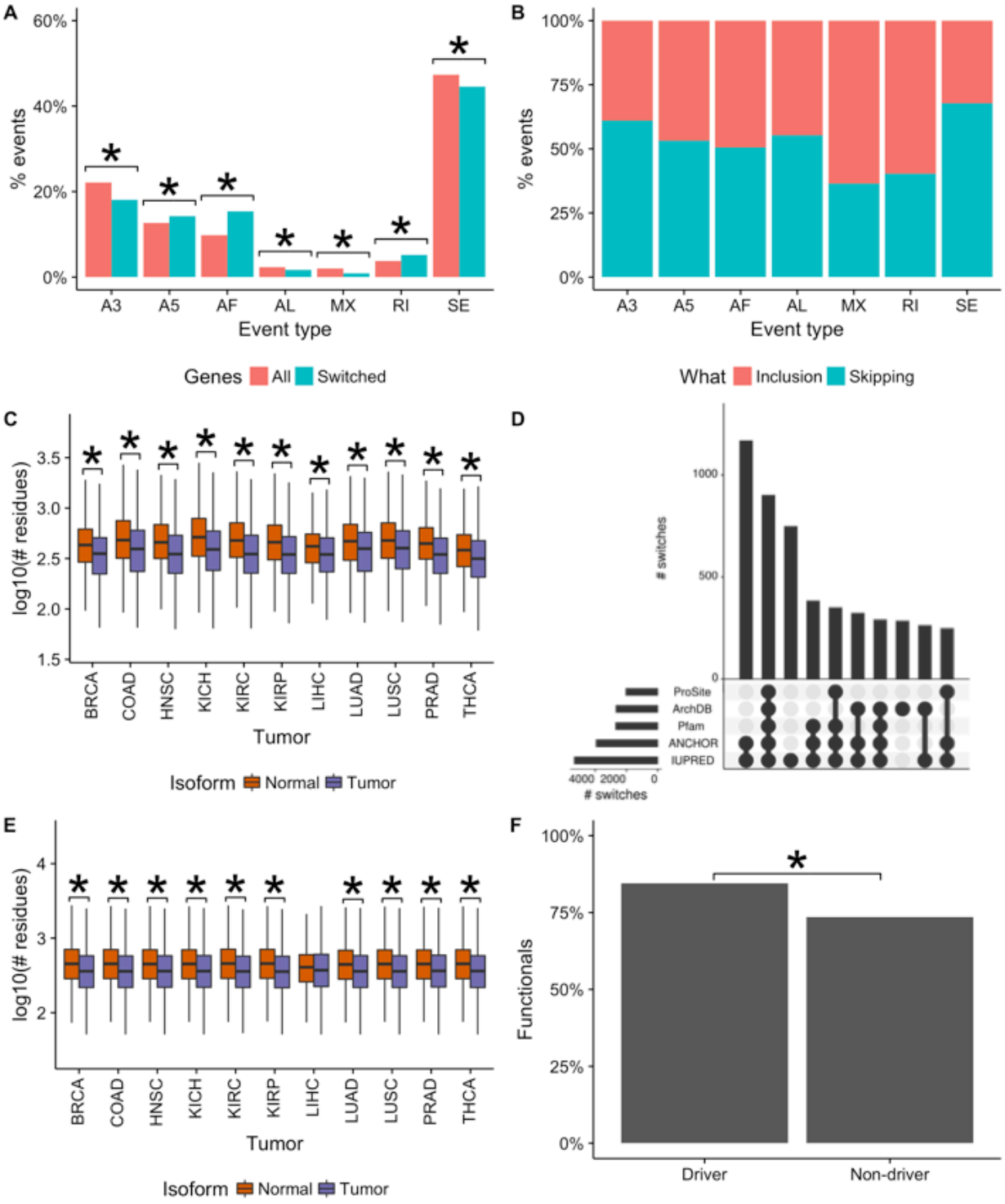

Figure S1. Properties of isoform switches Related to Figure 1. **(A)** Proportion of local alternative splicing event types (y-axis) described by the switches (blue) and by all genes in the annotation (red). These proportions are shown for events of type alternative 3' (A3) and 5' (A5) splice-site, alternative first (AF) and last (AL) exons, mutually exclusive exons (MX), intron retention events (RI) and exon cassette events (SE). Significance of the difference was determined with a Fisher's exact test for each event type using a contingency table with the counts of each event type and the rest of events in the two sets: switches and annotation **(B)** For each set of local alternative splicing events from the same type mapped to isoform switches, we indicate the proportion of cases that correspond to either inclusion (red) or exclusion (blue). For instance, inclusion for the A3 and A5 events correspond to the longer form, for AF events to the most upstream exon, to the most downstream exon for AL events, to the inclusion of the exon with the lowest coordinates for MX events, to the retention of the intron for RI events, and to the inclusion of the cassette exon for SE events. Blue corresponds to the opposite configuration.Further details of the description of the events can be found in https://github.com/comprna/SUPPA (Alamancos et al., 2015). **(C)** Distributions of the lengths of the tumor (purple) and normal (red) protein isoforms in the calculated isoform switches. The *y*-axis indicates the number of residues in log10 scale. **(D)** Overlap graph (Conway et al., 2017) of protein features affected in functional switches: Prosite patterns (Prosite), protein loops (ArchDB), Pfam domains (Pfam), disordered regions with potential to mediate protein–protein interactions (ANCHOR), and general disordered regions (IUPRED). The horizontal bars indicate the number of switches affecting each feature. The vertical bars indicate the number of switches in each intersection indicated by connected bullet points. **(E)** Distributions of the lengths of the tumor (purple) and normal (red) protein isoforms in the simulated transcript isoform switches. **(F)** Enrichment of functional switches in cancer drivers. We separated all switches (from Table S1) according to whether they are cancer drivers or non-drivers (in any tumor type), and whether they have functional switches or not. From the 6004 functional switches, ~4% are drivers, whereas from the 2118 non-functional switches, ~2% are drivers. Similarly, from all considered 278 drivers, ~84% are functional, whereas ~73% of the 7844 non-driver switches are functional. A Fisher's exact test produced a p-value = 2.034e-05 and odds-ratio = 1.965563 for the enrichment of functional switches in drivers (95 percent confidence interval: 1.409, 2.799).

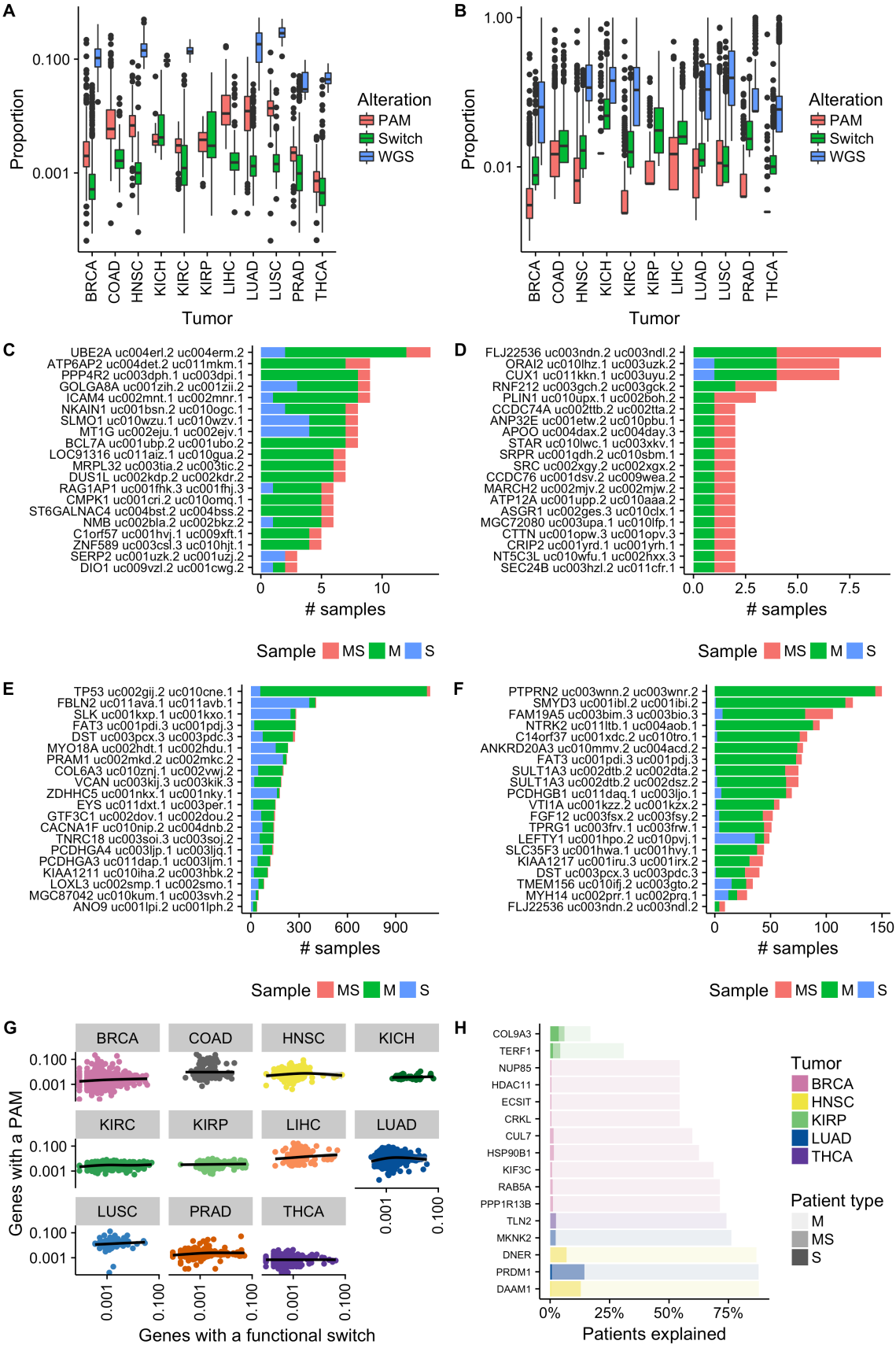
A
Proportion
0.100
0.001
BRCA COAD HNSC KICH KIRC KIRP LIHC LUAD LUSC PRAD THCA
Tumor
Alteration
PAM
Switch
WGS
B
1.00
0.01
C
UBE2A uc004erl.2 uc004erm.2
ATP6AP2 uc004det.2 uc011mkm.1
PPP4R2 uc003dph.1 uc003dpi.1
GOLGA8A uc001zih.2 uc001zii.2
ICAM4 uc002mnt.1 uc002mnr.1
NKAIN1 uc001bsn.2 uc010ogc.1
SLMO1 uc010wzu.1 uc010wzv.1
MT1G uc002eju.1 uc002ejv.1
BCL7A uc001ubp.2 uc001ubo.2
LOC91316 uc011aiz.1 uc010gua.2
MRPL32 uc003tia.2 uc003tic.2
DUS1L uc002kdp.2 uc002kdr.2
RAG1AP1 uc001fhk.3 uc001fhj.3
CMPK1 uc001cri.2 uc010omq.1
ST6GALNAC4 uc004bst.2 uc004bss.2
NMB uc002bla.2 uc002bkz.2
C1orf57 uc001hvj.1 uc009xft.1
ZNF589 uc003csl.3 uc010hjt.1
SERP2 uc001uzk.2 uc001uzj.2
DIO1 uc009vzl.2 uc001cwg.2
0 5 10
samples
Sample MS M S
D
FLJ22536 uc003ndn.2 uc003ndl.2
ORAI2 uc010lhz.1 uc003uzk.2
CUX1 uc011kkn.1 uc003uyu.2
RNF212 uc003gch.2 uc003gck.2
PLIN1 uc010upx.1 uc002boh.2
CCDC74A uc002ttb.2 uc002tta.2
ANP32E uc001etw.2 uc010pbu.1
APOO uc004dax.2 uc004day.3
STAR uc010lwc.1 uc003xkv.1
SRPR uc001qdh.2 uc010sbm.1
SRC uc002xgy.2 uc002xgx.2
CCDC76 uc001dsv.2 uc009wea.2
MARCH2 uc002mjv.2 uc002mjw.2
ATP12A uc001upp.2 uc010aaa.2
ASGR1 uc002ges.3 uc010clx.1
MGC72080 uc003upa.1 uc010lfp.1
CTTN uc001opw.3 uc001opv.3
CRIP2 uc001yrd.1 uc001yrh.1
NT5C3L uc010wfu.1 uc002hxx.3
SEC24B uc003hzl.2 uc011cfr.1
0.0 2.5 5.0 7.5
samples
Sample MS M S
E
TP53 uc002gij.2 uc010cne.1
FBLN2 uc011ava.1 uc011avb.1
SLK uc001kxp.1 uc001kxo.1
FAT3 uc001pdi.3 uc001pdj.3
DST uc003pcx.3 uc003pdc.3
MYO18A uc002hdt.1 uc002hdu.1
PRAM1 uc002mkd.2 uc002mkc.2
COL6A3 uc010znj.1 uc002vwj.2
VCAN uc003kij.3 uc003kik.3
ZDHHC5 uc001nkx.1 uc001nky.1
EYS uc011dxt.1 uc003per.1
GTF3C1 uc002dov.1 uc002dou.2
CACNA1F uc010nip.2 uc004dnb.2
TNRC18 uc003soi.3 uc003soj.2
PCDHGA4 uc003ljp.1 uc003ljq.1
PCDHGA3 uc011dap.1 uc003ljm.1
KIAA1211 uc010iha.2 uc003hbk.2
LOXL3 uc002smp.1 uc002smo.1
MGC87042 uc010kum.1 uc003svh.2
ANO9 uc001lpi.2 uc001lph.2
0 300 600 900
samples
Sample MS M S
F
PTPRN2 uc003wnn.2 uc003wnr.2
SMYD3 uc001ibl.2 uc001ibi.2
FAM19A5 uc003bim.3 uc003bio.3
NTRK2 uc011ltb.1 uc004aob.1
C14orf37 uc001xdc.2 uc010tro.1
ANKRD20A3 uc010mmv.2 uc004acd.2
FAT3 uc001pdi.3 uc001pdj.3
SULT1A3 uc002dtb.2 uc002dta.2
SULT1A3 uc002dtb.2 uc002dsz.2
PCDHGB1 uc011daq.1 uc003ljo.1
VTI1A uc001kzz.2 uc001kzx.2
FGF12 uc003fsx.2 uc003fsy.2
TPRG1 uc003frv.1 uc003frw.1
LEFTY1 uc001hpo.2 uc010pvj.1
SLC35F3 uc001hwa.1 uc001hvy.1
KIAA1217 uc001iru.3 uc001irx.2
DST uc003pcx.3 uc003pdc.3
TMEM156 uc010ifj.2 uc003gto.2
MYH14 uc002prr.1 uc002prq.1
FLJ22536 uc003ndn.2 uc003ndl.2
0 50 100 150
samples
Sample MS M S
G
BRCA COAD HNSC KICH
KIRC KIRP LIHC LUAD
LUSC PRAD THCA
0.100
0.001
Genes with a PAM
Genes with a functional switch
H
COL9A3
TERF1
NUP85
HDAC11
ECSIT
CRKL
CUL7
HSP90B1
KIF3C
RAB5A
PPP1R13B
TLN2
MKNK2
DNER
PRDM1
DAAM1
0% 25% 50% 75%
Patients explained
Tumor
BRCA
HNSC
KIRP
LUAD
THCA
Patient type
M
MS
S

Figure S2. Properties of functional isoform switches in tumors. Related to Figure 2. **(A)** Proportion of genes in $\log_{10}$ scale (*y*-axis) with either of these three alterations: isoform switches (red), protein-affecting mutations (PAMs) from whole Exome sequencing (WES) data (green), and any mutation type from whole genome sequencing (WGS) data (blue). **(B)** Proportion of samples in $\log_{10}$ scale (*y*-axis) with either of these three alterations: isoform switches (red), PAMs from WES data (green), and any mutation type from WGS data (blue). **(C-F)** Potential associations between mutations and switches. We show the top 20 cases according to the Jaccard score for the association of mutations (M) and switches (S) using WES (C) and WGS (D) data. We also show the top 20 cases according to the number of MS samples for WES (E) and WGS (F) data. For each gene and isoform (y axis), we show the number of patients for which we observed a mutation only (M), a switch only (S), or the co-occurrence of both (MS). **(G)** Lack of correlation between mutations and switches. For each tumor type, each dot represents a sample according to the number of genes with a functional switch (*x*-axis) and the number of genes with protein-affecting mutations (PAMs) (*y*-axis). **(H)** Functional switches that potentially characterize pan-negative tumor samples. For each switch along the *y*-axis, we represent the proportion of patients from a given tumor type (*x*-axis) that harbor mutations in a tumor-specific mutational driver (M), have the switch (S), or have both (MS). The switches are ranked from the bottom of the *y*-axis according to the total number of patients explained. Only the top 30 cases are shown. Each case is color-coded according to tumor type.

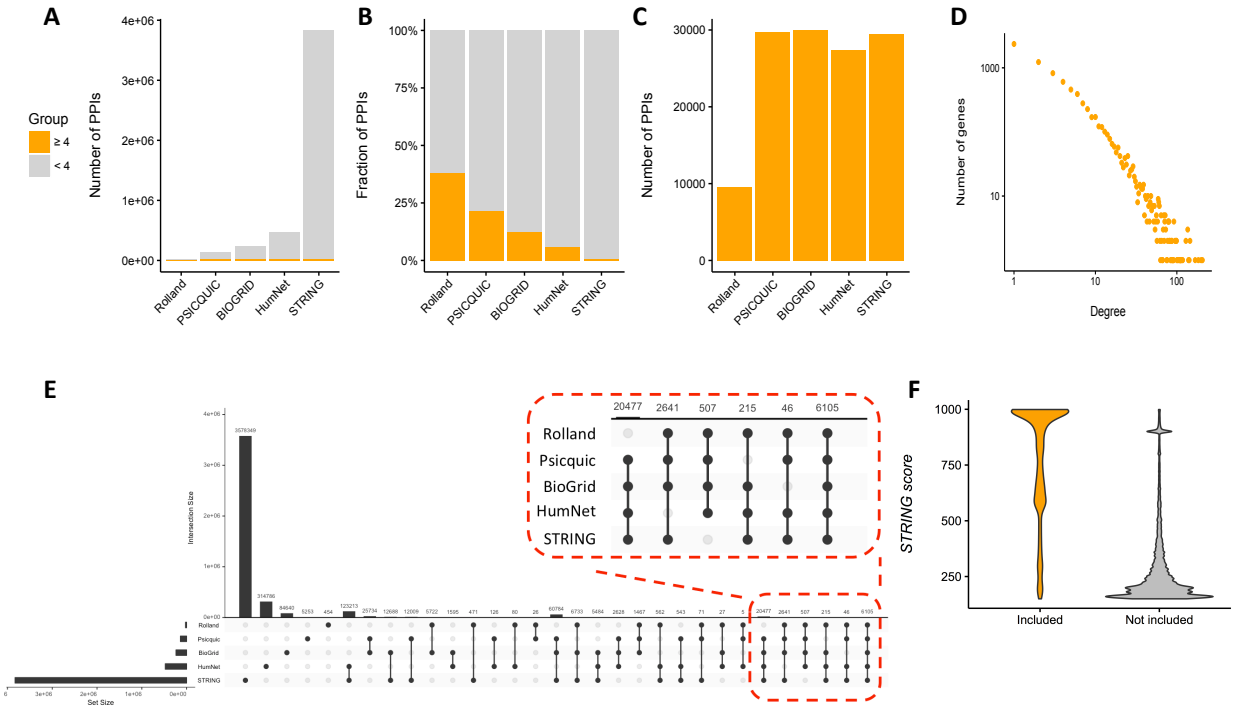
A
Group
≥ 4
< 4
Number of PPIs
4e+06
3e+06
2e+06
1e+06
0e+00
Rolland
PSICQUIC
BIOGRID
HumNet
STRING
B
Fraction of PPIs
100%
75%
50%
25%
0%
C
Number of PPIs
30000
20000
10000
0
D
Number of genes
1000
10
1
10
100
Degree
E
Interaction Size
Set Size
20477
2641
507
215
46
6105
Rolland
Psicquic
BioGrid
HumNet
STRING
F
STRING score
1000
750
500
250
Included
Not included

Figure S3. Protein-protein interaction network. Related to Figure 3**. (A)** Consensus protein–protein interaction (PPI) network. We used data from five different sources: PSICQUIC, BIOGRID, HumNet, STRING, and (Rolland et al., 2014). These networks vary in their size, connectivity, and origin, with PSICQUIC, BIOGRID, and Rolland being experimental networks and HumNet and STRING being functional networks. To build our consensus network, we used only those interactions that were defined in at least four different networks (shown in orange). **(B)** Fraction of each network included in the consensus network, with the data from (Rolland et al., 2014) having over 30% of its interactions and STRING less than 5%. **(C)** Number of interactions from each network included in the consensus network. **(D)** Degree distribution of the consensus network. For each number of PPI connections (*x*-axis), we give the number of genes with this degree (*y*-axis). **(E)** Highlighted in red are the PPIs considered for our analysis. Despite the fact that the dataset published in Rolland et al. was obtained through a search for new protein-protein interactions, many interactions in Rolland et al. are also present in the other PPI databases, with only 454 unique to Rolland et al. The plot also shows that even though many interactions are only present in STRING, most of them are not taken into account in our analysis. Plot performed with UpSetR (Conway et al., 2017). 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**. (F)** STRING PPIs included in our analysis (present in at least three other databases) are enriched for high-scoring interactions.

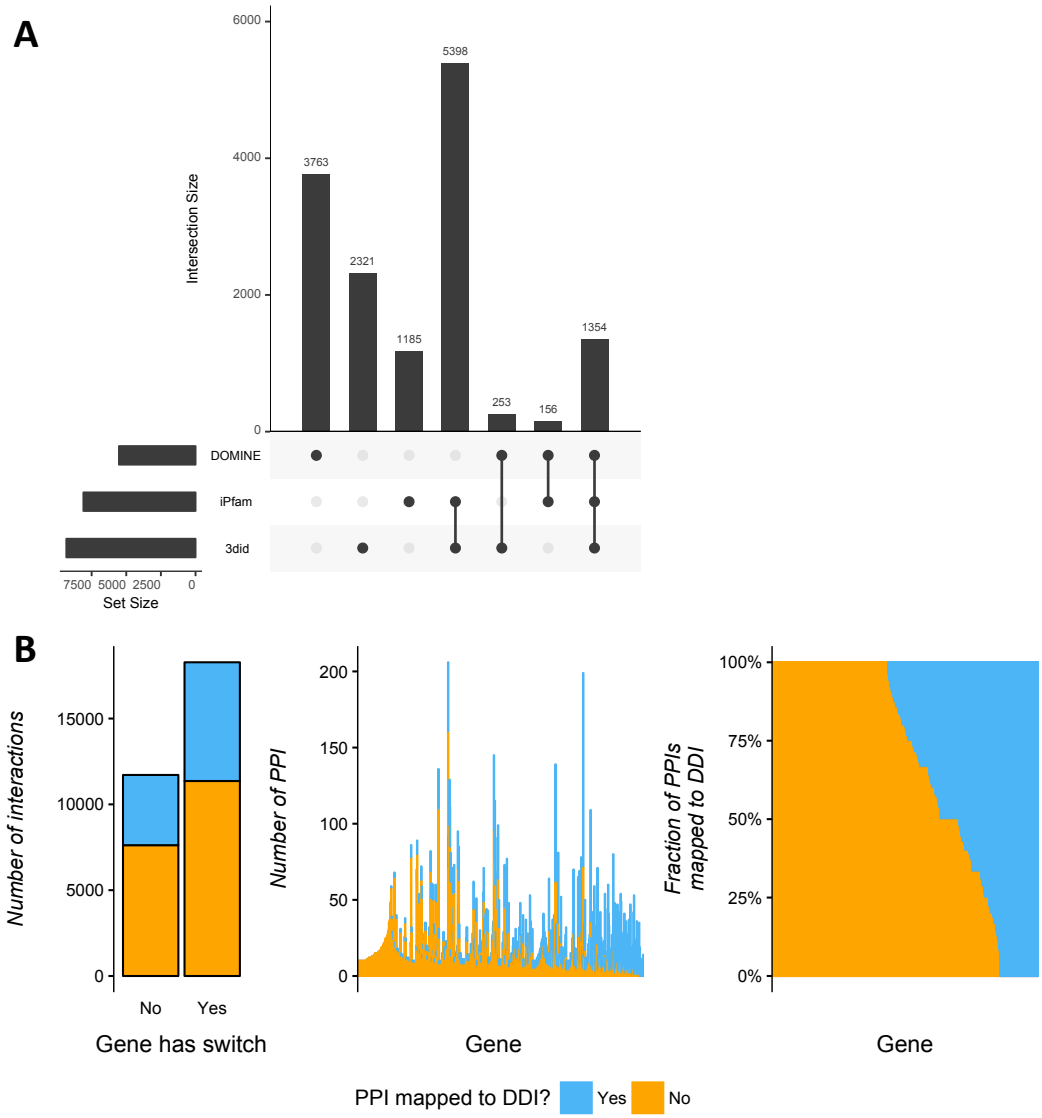
A
6000
4000
2000
0
Intersection Size
3763
2321
1185
5398
253
156
1354
DOMINE
iPfam
3did
7500 5000 2500 0
Set Size
B
Number of interactions
15000
10000
5000
0
No
Yes
Gene has switch
Number of PPI
200
150
100
50
0
Gene
Fraction of PPIs mapped to DDI
100%
75%
50%
25%
0%
Gene
PPI mapped to DDI?
Yes
No

Figure S4. Protein-protein interactions assignment to functional isoform switches. Related to Figure 3. (A) Number of domain–domain interactions (DDIs) analyzed, separated by source: 3did, iPfam, DOMINE. The plot shows the number of cases in each source (horizontal bars) and the intersections between the sources (vertical bars), which are indicated by connected bullet points (B) Mapping of switches to protein-protein interactions (PPIs). Left panel: From a total of 29991 PPIs, 11008 of them were mapped to DDIs, 6917 of them in genes with switches whereas 4091 are in genes without switches. The rest of the 18983 PPIs did not map to DDIs: 11361 corresponded to genes with switches, and 7622 to genes without switches. Middle panel: Absolute number of PPI interactions mapped (blue) or not mapped (orange) to a DDI in each gene (only genes with at least 10 PPIs are depicted). Genes are sorted according to the fraction of interactions that could be mapped to DDIs. The picture shows no correlation between the degree of a gene and the fraction of interactions mapped. Right panel: Fraction of PPIs mapped to DDIs per gene. Genes are sorted according to the fraction of PPIs successfully mapped to DDIs.

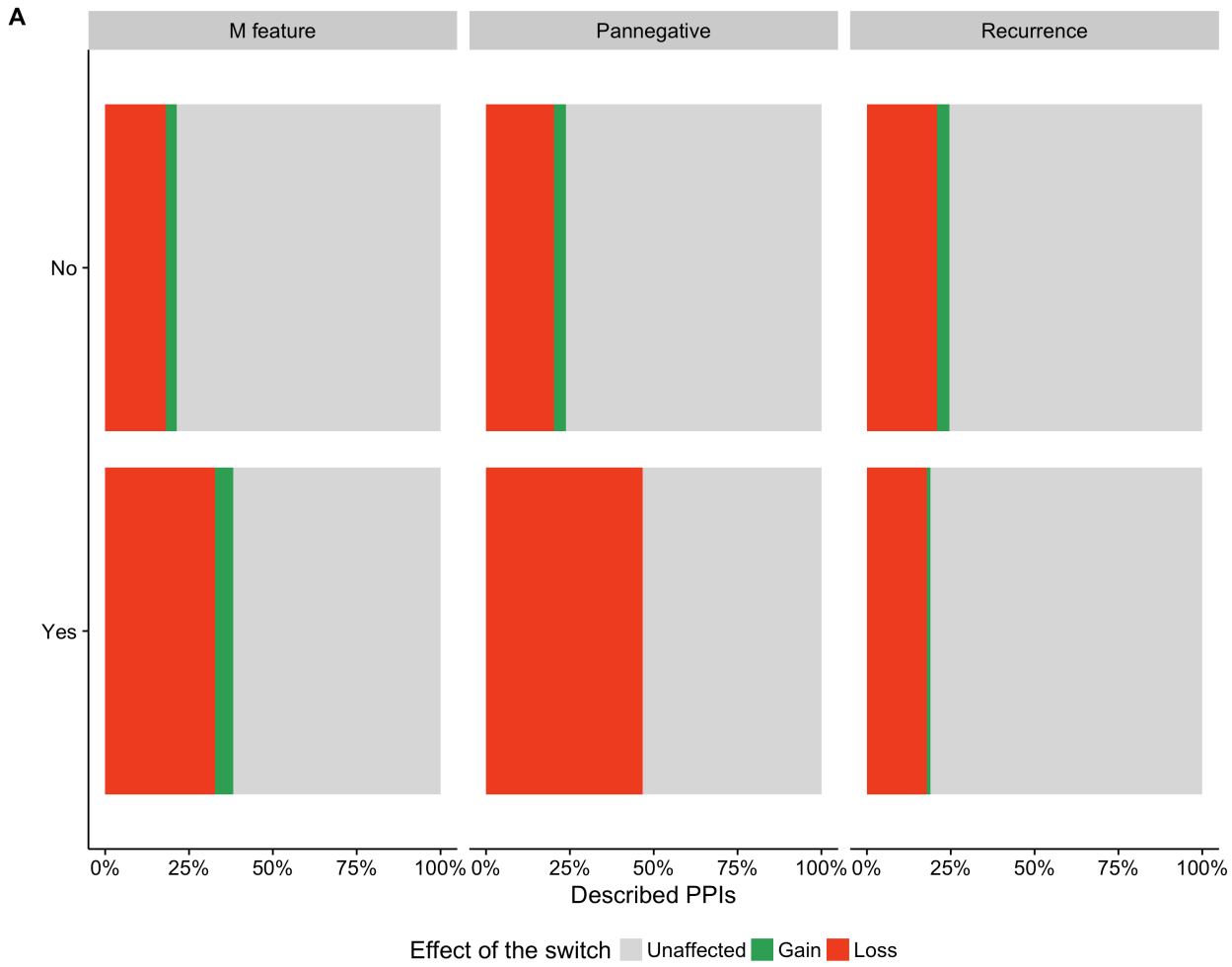

Figure S5. Properties of switches that affect protein-protein interactions. Related to Figure 3. Comparison of proportions of functional switches that affect protein-protein interactions (PPIs). In the left panel, functional switches are divided according to whether they affect domains frequently mutated in cancer (M feature) (Yes) or not (No). In the middle panel, functional switches are divided according to whether the switch has significant mutual exclusion with tumor-specific drivers (Pannegative). In the right panel, functional switches are divided according to whether they are recurrent (Yes) or not (No). In each subset we plot the proportion of PPIs that are kept unaffected (gray), lost (red), or gained (green). Using these three categories and the two values for each feature, M feature and Pannegative associate frequently with PPI-affecting switches (Chi-square test p-value < 2.2e-16 and p-value = 6.8e-08, respectively).

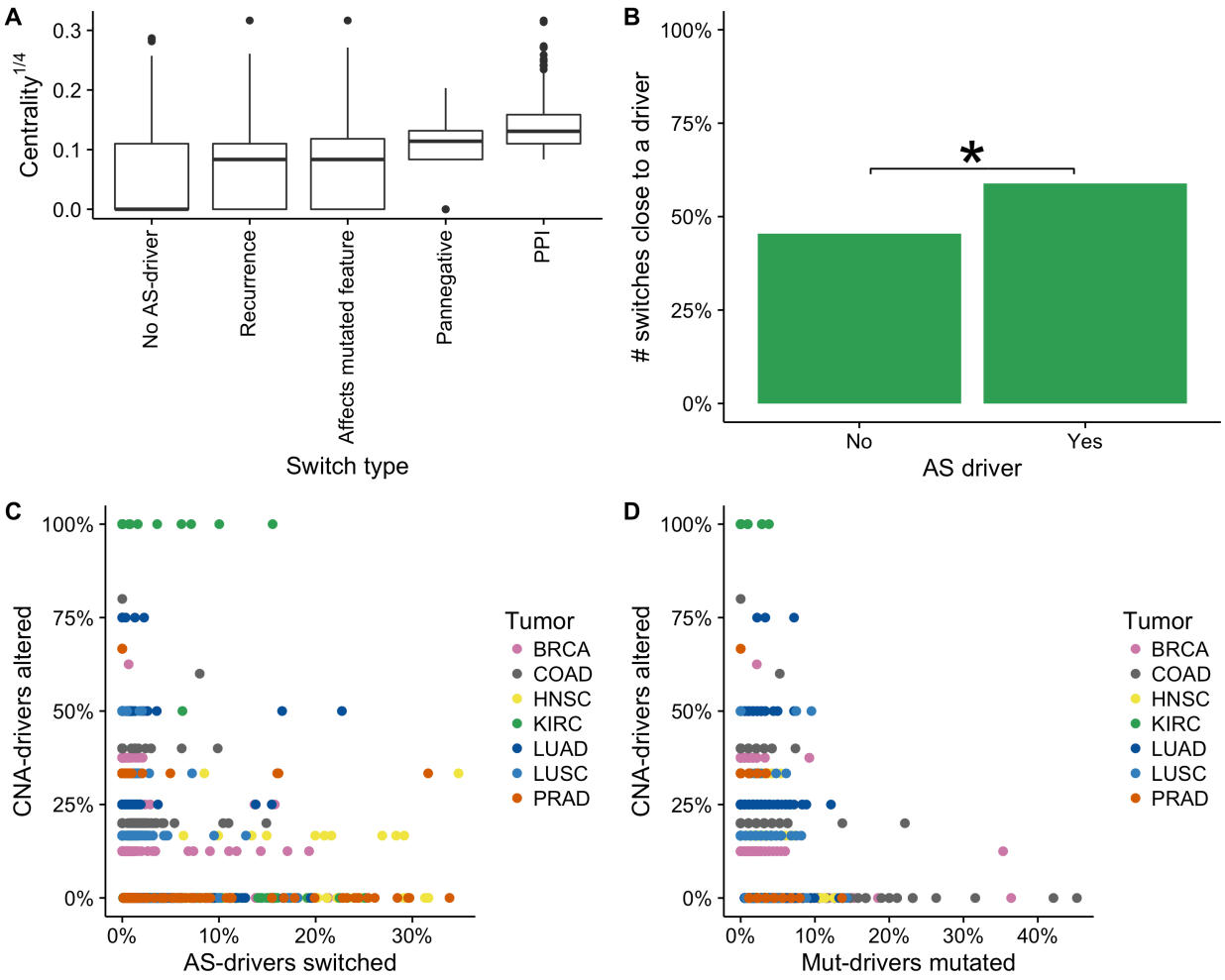

Figure S6. Potential AS-drivers. Related to Figure 4. **(A)** We show the distribution of centrality values (*y* axis) for functional isoform switches (*x* axis), labeled as candidate AS-drivers (*n* = 1662), separated according to their properties, and for the rest of functional switches (No AS-driver) (*n* = 4342). A Mann-Whitney test comparing all candidate AS-drivers with the rest of functional switches (*No AS-driver*) yielded a p-value < 2.2e-16. Comparing each subset of AS-drivers with the *No AS-driver* set yielded p=0.019 for the significantly recurrent AS-drivers (Recurrence), p=1.05e-4 for the AS-drivers that affect domains frequently mutated in cancer (Affects mutated feature), p=0.0023 for pannegative cases, and 2.70e-151 for cases that affect PPIs. **(B)** We show the proportion of potential AS-drivers and of other switches according to the closest driver distance (CDD). CDD is calculated as the distance to the closest tumor-specific cancer gene driver in the consensus PPI network. Every switch with CDD<=3 was labeled as

“Close to a driver”. Otherwise, it was labeled “Far from a driver”. A Fisher’s exact test on the proportion of potential AS-drivers and other isoform changes (no AS-drivers) that are close or far from a driver, gave an enrichment of potential AS-drivers close to drivers (p-value < 2.2e-16, odds-ratio = 1.5). **(C)** Each patient is colored by tumor type and represented according to the percentage of tumor-specific copy number alteration (CNA) driver genes amplified in that sample (*y* axis) and the percentage of potential AS-drivers occurring in the same sample (*x* axis). **(D)** Each patient is colored by tumor type and represented according to the percentage of tumor-specific CNA driver genes amplified in that sample (*y* axis) and the percentage of mutational drivers mutated in the same sample (*x* axis)

Supplementary Tables

Table S1. Isoform switches. Related to Figure 1. Provided as a text file with tab-separated values (.tsv). This table contains the list of identified isoform switches used for this analysis, including functional and nonfunctional ones, and indicating which ones might be potential AS-drivers. The table provides the following information:

Column number	Column label	Description
1	GeneId	Entrez gene id
2	Symbol	HGNC gene symbol
3	Normal_transcript	UCSC transcript id
4	Tumor_transcript	UCSC transcript id
5	Normal_protein	Uniprot_ID (None if not known)
6	Tumor_protein	Uniprot_ID (None if not known)
7	DriverAnnotation	“Driver” if it’s a driver, “d1” if it’s an interactor of a driver, and “Nothing” otherwise
8	IsFunctional	1 if it is functional as defined in the article, 0 otherwise
9	Driver	1 if it is a driver, 0 otherwise
10	Druggable	1 if it is a target of a known drug according to DGIdb (http://dgidb.genome.wustl.edu/)
11	CDS_Normal	1 if the normal transcript has an annotated CDS, 0 otherwise
12	CDS_Tumor	1 if the tumor transcript has an annotated CDS, 0 otherwise
13	CDS_change	1 if the CDS changes between the tumor and normal transcripts
14	UTR_change	1 if the 5’3 or 3’ UTRs change between the tumor and normal transcripts
15	Tumors	Tumor types in which the switch appears (BRCA, COAD, etc…)
16	Number_samples	Number of samples in which the switch appears
17	Percentage_samples	Percentage of samples from the total studied across all tumor types in which the switch appears
18	Samples	IDs of samples in which the switch appears
19	Recurrence	1 if it is recurrent, 0 otherwise
20	PPI	1 if the switch affects a PPI in every tumor type where it appears; 0 otherwise. All PPIs affected by switches per tumor type are in Supp. File 3.
21	Affects_mutated_feature	1 if the switch leads to a gain or loss of a domain that is enriched in mutations in tumors, 0 otherwise
22	Pannegative	Number of cancer drivers from the same pathway with which the switch shows mutual exclusion
23	Potential_AS_driver	1 if 19,20,21 or 22 is equal to 1, 0 otherwise
24	MS.pam	Samples with co-occurrence of switch and PAM in the same gene
25	M.pam	Samples with PAMs only
26	S.pam	Samples with Switches
27	N.pam	Rest of samples
28	p.pam.me	p-value of the mutual exclusion test

29	MS.mut	Samples with co-occurrence of switch and WGS mutations
30	M.mut	Samples with WGS mutations only
31	S.mut	Samples with Switches
32	N.mut	Rest of samples
33	p.mut.o	p-value of the co-occurrence of mutations and switches

Table S2. Mutation and domain gain/loss enrichments in protein domain families. Related to Figure 2. Provided as a text file with tab-separated values (.tsv). This table contains the information about the Protein domain families that are significantly enriched in mutations as well as gains or losses in isoform switches. The information provided for each domain family is the following:

Column number	Column label	Description
1	Pfam_id	PFAM ID for the domain family
2	Name	Name of the domain family
3	p_switch_gain	P-value for the gain-test
4	adjp_switch_gain	Adjusted P-value for the gain-test
5	p_switch_loss	P-value for the loss-test
6	adjp_switch_loss	Adjusted P-value for the loss-test
7	p_mutation	P-value for the mutation-test
8	adjp_mutation	Adjusted P-value for the mutation-test
9	Switches_where_gained	Number of switches where domain family is gained
10	Switches_where_lost	Number of switches where domain family is lost

Table S3. Mutual exclusion analysis between switches and cancer drivers. Related to Figure 2. Provided as a text file with tab-separated values (.tsv). This table contains the analysis of mutual exclusion between functional switches, *global mutual exclusion*, and mutational drivers in the same pathway, *local mutual exclusion*. Switches present global mutual exclusion if they exhibit an extreme mutually exclusive pattern (p_mut_ex < 0.05) with at least 3 of the most frequent tumor drivers for a certain cancer type (Number_ME_drivers >= 3). Switches present local mutual exclusion if they exhibit an extreme mutually exclusive pattern (p_me_pathway_driver < 0.05) with a driver from the same pathway (indicated in Same_pathway_driver). .

Column number	Column label	Description
1	GeneId	Entrez gene ID
2	Symbol	HGNC gene symbol
3	Normal_transcript	UCSC transcript id
4	Tumor_transcript	UCSC transcript id
5	Tumor	Tumor type (BRCA, COAD, etc…)
6	p_mut_ex	P-value for the test for mutual exclusion (ME) with mutational drivers
7	Number_ME_drivers	Number of drivers with mutual exclusion (ME)
8	MS_mut_ex	Number of samples with mutation (M) and switch (S)
9	M_mut_ex	Number of samples with only M
10	S_mut_ex	Number of samples with only S
11	N_mut_ex	Number of samples without M or S
12	ME_drivers	HGNC gene symbols for the ME drivers
13	Same_pathway_driver	Pathways shared with ME drivers
14	p_me_pathway_driver	P-value for the test for mutual exclusion (ME) with drivers in the same pathway
15	MS_me_pathway_driver	Number of samples with mutation (M) and switch (S)
16	M_me_pathway_driver	Number of samples with only M
17	S_me_pathway_driver	Number of samples with only S
18	N_me_pathway_driver	Number of samples without M or S

Table S4. Protein features and protein-protein interactions affected by isoform switches. Related to Figure 3. Provided as a text file with tab-separated values (.tsv). This table contains the proteins features and protein–protein interactions affected in each functional switch. The column descriptions are:

Column number	Column label	Description
1	Tumor	Tumor type (BRCA, COAD, etc…)
2	GeneId	Entrez gene ID
3	Symbol	HGNC gene symbol
4	Normal_transcript	UCSC transcript id
5	Tumor_transcript	UCSC transcript id
6	Feature_type	Pfam, Prosite, IUPRED, ANCHOR
7	Feature_id	ID for the protein feature if available
8	Feature_name	Name of Feature if available, positions in protein for IUPRED and

		ANCHOR
9	Observation	Gained_in_tumor/Lost_in_tumor/No_change
10	Normal_isoform_order	Domain copy this corresponds to / total copies in normal isoform
11	Tumor_isoform_order	Domain copy this corresponds to / total copies in tumor isoform
12	GeneId_partner	Entrez ID of the protein-protein interaction partner
13	Symbol_partner	HGNC symbol of the protein-protein interaction partner
14	Transcript_partner	Transcripts identified as coding the interaction partner
15	Pfam_id_partner	PFAM ID for the domain mediating the interaction
16	Effect_on_interaction	Unaffected/Gain/Loss/NA(no interaction data)

Table S5. Pathways enriched in PPI-affecting switches. Related to Figure 3. Provided as a text file with tab-separated values (.tsv). This table contains the gene sets that are enriched in isoform switches that are predicted to affect protein-protein interactions. The enrichment tests is a Fisher's exact test based on the separations of switches being in the pathway or not, and affecting PPIs or not. We have tested Pathways, Complexes and gene sets-related to mRNA-metabolism. Only Pathways showed enrichment after multiple-test correction. The column descriptions are:

Column number	Column label	Description
1	Geneset_type	Pathway/Complex/mRNA_regulation
2	Geneset	Name of the gene set
3	Number_drivers	Number of drivers in the gene set.
4	p	Fisher's exact test p-value
5	adjp	p-value corrected for multiple testing
6	OR	Odds-ratio
7	eOR	Estimated odds-ration using with pseudocounts
8	Switched_genes	Genes in the gene set that have a PPI-affecting switch

Table S6. Gene modules with protein-protein interactions affected by isoform switches. Related to Figure 3. Provided as a text file with tab-separated values (.tsv). This table contains modules with high density of affected interactions: sets of genes that are connected in the network of protein-protein interactions and many of their interactions are affected by the isoform switches and separately from other genes in the PPI network. We provide a test for assigning a complex or pathway based on the intersection of the complex/pathway to the module (see Experimental Procedures for details). The column descriptions are:

Column number	Column label	Description
1	Module	Module number
2	Module_components	Genes in the module (calculated from the network of protein-protein interactions affected by isoform switches)
3	Geneset	Name of complex/pathway compared to the module (NA if none was assigned)
4	Geneset_size	Number of genes in the complex/pathway (NA if none was assigned)
5	p	p-value from binomial test for the intersection of the gene set (Complex/Pathway) to the module
6	Intersection	Number of genes from the gene set that are in the module
7	Number_drivers	Number of cancer drivers in the module
8	padj	p-value corrected for multiple testing

References

Akerman, M., Fregoso, O.I., Das, S., Ruse, C., Jensen, M. a, Pappin, D.J., Zhang, M.Q., and Krainer, A.R. (2015). Differential connectivity of splicing activators and repressors to the human spliceosome. Genome Biol. *16*, 119.

Alamancos, G.P., Pagés, A., Trincado, J.L., Bellora, N., and Eyras, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. RNA *21*, 1521–1531.

Babur, Ö., Gönen, M., Aksoy, B.A., Schultz, N., Ciriello, G., Sander, C., and Demir, E. (2015). Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. Genome Biol. *16*, 45.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *10008*, 6.

Bonet, J., Planas-Iglesias, J., Garcia-Garcia, J., Marín-López, M.A., Fernandez-Fuentes, N., and

Oliva, B. (2014). ArchDB 2014: Structural classification of loops in proteins. Nucleic Acids Res. *42*.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2015). The BioGRID interaction database: 2015 update. Nucleic Acids Res. *43*, D470–D478.

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

Davis, C.F., Ricketts, C.J., Wang, M., Yang, L., Cherniack, A.D., Shen, H., Buhay, C., Kang, H., Kim, S., Fahey, C.C., et al. (2014). The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. Cancer Cell *26*, 319–330.

del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L., et al. (2013). A new reference implementation of the PSICQUIC web service. Nucleic Acids Res. *41*, W601-6.

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433–3434.

Dosztányi, Z., Mészáros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics *25*, 2745–2746.

Fang, H., and Gough, J. (2013). DcGO: Database of domain-centric ontologies on functions, phenotypes, diseases and more. Nucleic Acids Res. *41*.

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

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., et al. (2016). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. *44*, D279-85.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., et al. (2015). COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. *43*, D805–D811.

Fredriksson, N.J., Ny, L., Nilsson, J.A., and Larsson, E. (2014). Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. Nat. Genet. *46*, 1–7.

Gattiker, A., Gasteiger, E., and Bairoch, A. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. Appl. Bioinformatics *1*, 107–108.

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

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

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

Jonsson, P.F., and Bates, P.A. (2006). Global topological features of cancer proteins in the human interactome. Bioinformatics *22*, 2291–2297.

Lee, I., Blom, U.M., Wang, P.I., Shim, J.E., and Marcotte, E.M. (2011). Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res. *21*, 1109–1121.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

Mosca, R., Céol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: A catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. *42*.

Raghavachari, B., Tasneem, A., Przytycka, T.M., and Jothi, R. (2008). DOMINE: A database of protein domain interactions. Nucleic Acids Res. *36*.

Rolland, T., Taşan, M., Charloteaux, B., Pevzner, S.J., Zhong, Q., Sahni, N., Yi, S., Lemmens, I., Fontanillo, C., Mosca, R., et al. (2014). A proteome-scale map of the human interactome network. Cell *159*, 1212–1226.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman,

G., Montrone, C., and Mewes, H.W. (2009). CORUM: The comprehensive resource of mammalian protein complexes-2009. Nucleic Acids Res. *38*.

Schroeder, M.P., Rubio-Perez, C., Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2014). OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. In Bioinformatics, p.

Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M.A., Valcárcel, J., and Eyras, E. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res. *26*, 732–744.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. (2011). The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. *39*.

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

The Cancer Genome Atlas Research Network (2016). Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. N. Engl. J. Med. *374*, 135–145.

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., and Eyras, E. (2016). SUPPA2 provides fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz Jr., L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science (80-.). *339*, 1546–1558.

Wachi, S., Yoneda, K., and Wu, R. (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics *21*, 4205–4208.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D. a, et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

Zhao, M., Kim, P., Mitra, R., Zhao, J., and Zhao, Z. (2015). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. 1–9.