

Widespread alteration of protein autoinhibition in human cancers

Ashwani Jha^{1,2}, Jennifer M. Bui^{1,2}, Dokyun Na³ and Jörg Gsponer^{1,2,*}

¹Michael Smith Laboratories, ²Department of Biochemistry and Molecular Biology, University of British Columbia. ³School of Integrative Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 156-756, Republic of Korea

*Jörg Gsponer, M.D.-Ph.D.
Department of Biochemistry and Molecular Biology
University of British Columbia
Vancouver, BC V6T 1Z4
Email: gsponer@mssl.ubc.ca

KEYWORDS:

Allostery, autoinhibition, protein regulation, cancer fusions, cancer missense mutations

ABSTRACT

Autoinhibition is a prevalent allosteric regulatory mechanism in signaling proteins as it prevents spurious pathway activation and primes for signal propagation only under appropriate inputs. Altered functioning of inhibitory allosteric switches underlies the tumorigenic potential of numerous cancer drivers. However, whether protein autoinhibition is altered generically in cancer cells remains elusive. Here, we reveal that cancer-associated missense mutations and fusion breakpoints are found with significant enrichment within inhibitory allosteric switches across all cancer types, which in the case of the fusion breakpoints is specific to cancer and not present in other diseases. Recurrently disrupted or mutated allosteric switches identify established and new cancer drivers. Cancer-specific mutations in allosteric switches are associated with distinct changes in signaling, and suggest molecular mechanisms for altered protein regulation, which in the case of ASK1, DAPK2 and EIF4G1 were supported by biophysical simulations. Our results demonstrate that autoinhibition-modulating genetic alterations are positively selected for by cancer cells, and that their study provides valuable insights into molecular mechanisms of cancer misregulation.

MAIN

Significant efforts in the cancer research community are underway in order to map all somatic molecular alterations in tumor cells such as single nucleotide variants (SNVs), copy number variations or gene fusions that initiate and drive cancer development¹⁻⁶. The targeted map covers a large number of cancer types and tissues including brain, lung, liver, pancreas, ovary, colon and blood. The currently available map has already revealed a picture of a very complex landscape of genetic alterations^{7,8}. Most alterations are likely to be functionally neutral and the result of an increased mutation rate in cancer cells, while a minority of driver alterations provide a selective advantage to cancer cells by helping establish and maintain one or multiple of the six hallmarks of cancer^{9,10}. Driver gene fusions, for instance, are responsible for the development of over 50% of all leukemias¹¹⁻¹³. For the specific case of chronic myeloid leukemia (CML), nearly all of its cases are caused by the chromosomal translocations that give rise to the BCR-ABL1 fusion protein¹⁴. Establishing functional association between driver alterations and the development of specific cancer types, as it is the case for the BCR-ABL1 gene fusion and CML, has improved biomarker development, diagnostics and clinical care significantly¹⁵⁻¹⁷.

Equally important as the discovery of drivers is the elucidation of the molecular mechanisms by which the drivers contribute to cancerous cell transformation and progression⁸. For the BCR-ABL1 fusion protein, the transforming potential is, at least in part, due to the disruption of ABL1's autoinhibition in the gene fusion product^{14,18,19}. In autoinhibition, in the most general sense, transient intramolecular interactions between distant parts of the same protein chain reduce or shut down protein function²⁰⁻²². Autoinhibition is a wide-spread form of allosteric protein function regulation, but it is most often seen in signaling proteins. In these proteins, the allosteric regulatory mechanism prevents spurious pathway activation and primes pathways to respond only to appropriate signals. Normally, biomolecular partners and post-translational modifications (PTMs) initiate and reinforce inhibitory intramolecular interactions, or, alternatively, trigger their disruption during signal transmission. As multiple allosteric effectors acting on the same autoinhibited protein can have synergistic or antagonistic effects, autoinhibition has emerged as ideal nexus for complex signal integration and activity-based signal propagation. In cancer cells, this type of allosteric protein activity regulation can be altered due to mutations that remove part of the autoinhibitory switch (e.g. BCR-ABL1 fusion protein) or modulate (strengthen or weaken) the inhibitory intramolecular interactions (e.g. SNVs in the autoinhibited kinase Akt1)^{14,20,23,24}. There has been significant progress and success in the identification of individual proteins that drive cancerous cell transformation due to altered autoinhibition, not least because re-establishing and stabilizing autoinhibition via small compounds has proven to be a successful strategy in cancer drug development^{18,25-27}. However, a systematic and robust analysis is needed to reveal whether mutations of allosteric inhibitory switches in cancer is a rare event, or whether altered allostery is highly prevalent and associated with signature molecular changes.

To close this knowledge gap, we integrated cancer genomics data with protein regulation information and systematically analyzed the association between cancer fusions /somatic cancer missense mutations and protein autoinhibition across multiple tumor types. Our data integration approach reveals a significant enrichment for cancer-associated missense mutations and fusion breakpoints within allosteric inhibitory switches, which in the case of the breakpoints is specific to cancer fusions and not seen in fusions associated with other diseases. Selection for allosteric switches that are recurrently altered by gene fusions or missense

mutations across and within cancer types identifies numerous well-established cancer drivers as well as putative new ones. Cancer missense mutations in allosteric switches are associated with specific transcriptional signatures of cancer, which provides insights into the molecular pathways that are activated as a result of altered allosteric protein regulation in cancers.

RESULTS

Breakpoints of cancer fusions are enriched within inhibitory allosteric regulatory switches

We sought to exploit recently established maps of genetic alterations, including gene fusions and SNVs, covering various human cancers coupled with data on allosteric protein regulation to examine the association between autoinhibition and cancer (**Fig. 1**). From 12'957 fusion genes collected from the literature (see methods), we identified 5'503 in-frame fusion proteins that were found in cancer cells and had the fusion breakpoint within protein coding regions. We matched these cancer fusion proteins with a set of 234 human proteins for which we found evidence in the literature that their activity is regulation via inhibitory allosteric switches (ASs). The analysis revealed that proteins found in cancer fusions have an enrichment for regulation via autoinhibition when compared to the entire human proteome (Fisher's exact test, $P < 4.42 \times 10^{-10}$)(**Fig. 2A**), proteins that, so far, have not been found in any gene fusion (Fisher's exact test, $P < 1.11 \times 10^{-16}$), and proteins that are part of gene fusions associated with neurodevelopmental disorders or other disease states (Fisher's exact test, $P < 0.02$) but not cancer (**Supplementary Fig. 1**). As some of these 234 proteins with inhibitory allosteric switches may have been characterized biochemically because they had previously been found in a cancer fusion (e. g. ABL1) and, as a result, their autoinhibition discovered, there may exist a study bias that, at least partially, explains the observed association. Therefore, we analyzed an additional group of 3'381 human proteins that have predicted allosteric regulatory switches (pASs). The predictions were made with an algorithm that we have previously developed and validated for the identification of intrinsically disordered ASs²⁸ (see methods for details). Consistent with the analysis of known autoinhibited proteins, we find that proteins found in cancer fusions are enriched in pAS when compared to the entire human proteome (Fisher's exact test, $P < 3.33 \times 10^{-16}$)(**Fig. 2B**) and proteins that have not been found in any gene fusion (Fisher's exact test, $P < 2.22 \times 10^{-16}$). As proteins involved in gene fusion have been shown to be enriched in intrinsic disorder, we repeated the analysis only for proteins that contain large disordered regions (IDRs). This control reveals that cancer fusions are also enriched in pAS when only considering proteins that have IDRs (**Supplementary Fig. 2**).

To probe for the relevance of these associations, we analyzed whether autoinhibitory switches get shortened or even removed as result of the gene fusion, i.e., whether breakpoints are located in the AS. Indeed, breakpoints in cancer gene fusions that involve autoinhibited proteins (**Fig. 2C**) are enriched in locations that potentially affect autoinhibition (Hypergeometric test, $P < 1.00 \times 10^{-99}$) compared to any other locations in the proteins. This enrichment holds also for the subset of fusions that involve known cancer driver genes (Hypergeometric test, $P < 1.00 \times 10^{-99}$). The enrichment appears reversed when considering only gene fusions that are not associated with cancer, even when these non-cancer fusions involve known cancer driver genes (**Fig. 2C**). Dividing driver genes in oncogenes and tumor suppressors reveals that only oncogenes have breakpoints enriched in locations that potentially affect autoinhibition (Hypergeometric test, $P < 1.00 \times 10^{-99}$), which is expected as driver alterations in oncogenes are activating while inactivating in tumor suppressors (**Supplementary Fig. 3**). We repeated the breakpoint analysis for proteins with predicted allosteric switches. This analysis reveals that breakpoints are also enriched in pASs when

compared to other disordered regions of the same cancer fusion proteins (**Fig. 2D**) (Hypergeometric test, $P < 1.00 \times 10^{-99}$) and cancer fusions containing known drivers (Hypergeometric test, $P < 1.00 \times 10^{-99}$), respectively. Interestingly, breakpoints are significantly enriched outside of pASs in fusions that are not associated with cancer (Hypergeometric test, $P < 3.33 \times 10^{-64}$ and $P < 1.10 \times 10^{-3}$, respectively). Consistent with the analysis on known AS, only oncogenes have breakpoints enriched in predicted pASs (Hypergeometric test, $P < 2.94 \times 10^{-34}$, **Supplementary Fig. 3**). We confirmed these enrichment results for pAS in control calculations that defined allosteric-altering fusions as those where breakpoints are not only in the pAS but also between the pAS and the protein domain which function is regulated (see methods and **Supplementary Fig. 4**).

Cancer missense mutations are enriched within inhibitory allosteric regulatory switches

Next, we mapped 7'799 missense mutations from the COSMIC database²⁹ on human proteins with AS to determine whether residues in these experimentally-mapped allosteric regulatory switches are frequently mutated in cancers. A significantly higher percentage of sequence positions in ASs have at least one cancer-associated mutation when compared to other locations in the same protein (**Fig. 3A**) (Hypergeometric test, $P < 4.10 \times 10^{-4}$). As some sequence positions are mutated multiple times, we also computed the average mutation rate, and found the rate significantly higher inside the ASs compared to outside (**Fig. 3B**) (T-test, $P < 0.04$). Recent reports have highlighted the potential importance of cancer-related mutations close to or at phosphorylation sites^{30,31}. To control for this possible co-variant, we calculated the average mutation rate at (**Fig. 3C**) or close to known phosphorylation sites (**Fig. 3D**) inside and outside of ASs. Phosphorylation sites and residues close-by in ASs have a significantly higher mutation rate than the equivalent sites outside ASs (T-test, $P < 0.02$ and $P < 4.09 \times 10^{-4}$, respectively). When the analysis was restricted to ASs in known cancer drivers (**Fig. 3A-D**), similar trends were observed but, probably due to the lower number of cases, only the mutation rate at phosphorylation sites inside ASs of drivers was found statistically higher than outside (**Fig. 3C**) (T-test, $P < 0.03$).

We repeated this analysis for the 3'381 human proteins with pASs. Consistent with the analysis of proteins with ASs, we find that a significantly higher percentage of residues in pAS of all human proteins or known drivers have cancer-associated mutation (Hypergeometric test, $P < 2.02 \times 10^{-26}$ and $P < 9.01 \times 10^{-11}$, respectively) (**Fig. 3A, bottom**) when compared to other disordered parts of the same proteins. In addition, the average mutation rate is higher within pASs, at phosphorylation sites or close to phosphorylation sites within pASs of all human proteins (**Fig. 3B-D, bottom**) (T-test, $P < 0.02$, $P < 0.04$, and $P < 1.74 \times 10^{-14}$) or known drivers (T-test, $P < 0.02$, $P < 0.05$, and $P < 1.46 \times 10^{-4}$), respectively, when compared to disordered regions outside pASs. We found similar trends when taking as a reference the mutation rates of all protein regions outside pASs and not only the disordered parts (**Supplementary Fig. 5**).

Significantly disrupted inhibitory allosteric switches identify known cancer genes and predict new ones.

Motivated by the observed associations, we decided to identify those human genes with recurrent fusion breakpoints in regions that would alter the regulatory effect of known or predicted allosteric switches, (p)ASs, (see Methods) and assess their cancer driver potential. Various methods exist that exploit high cancer mutation rates within genes or mutation clustering to identify drivers^{3,4,30,32-37}, but these methods often provide no or little insight into how mutations affect protein function and induce cancer formation. Focusing on allosteric

switches provides testable hypotheses on the molecular mechanism by which genetic alterations could induce cancerous cell changes. Thus, we calculated P-values of the binomial probabilities of finding k or more fusion breakpoints within regions that encode the extended (p)AS (see methods for details). Only a pan-cancer enrichment analysis (33 cancer types) was possible because of the low number of reported fusion events. We identified 37 genes with significantly disrupted allosteric switches (**Fig. 4A**). Most of these genes are expected to be oncogenes because truncation or removal of autoinhibitory switches, as a result of fusion, should lead to activation. Indeed, 13 of the 37 genes (ABL1, RET, BRAF, PDGFRA/B, ALK, FGFR1/2/3, MET, ERG, ETV6 and MYB⁸; **Fig. 4B** and **Supplementary Fig. 6**) are well-known oncogenes, which is a significant enrichment compared to the number of known oncogenes found in all cancer gene fusions (Fisher's exact test, $P < 2.14 \times 10^{-10}$) and cancer gene fusions involving proteins with (p)ASs (Fisher's exact test, $P < 5.44 \times 10^{-7}$). Importantly, disruption of autoinhibition is known to contribute to the oncogenic activity of several of these genes including ABL1¹⁴, BRAF³⁸, PDGFRA³⁹ and MYB⁴⁰. Interestingly, one of the 37 genes is the tumor suppressor NF-1. Neurofibromin, the protein product of NF-1, shuttles, like many tumor suppressors⁴¹, between cytoplasm and nucleus⁴². Nuclear import depends on a NLS⁴² in the C-terminus that also contains a pAS. As appropriate cellular localization of neurofibromin is key to its function⁴³, it is conceivable that a removal of the predicted pAS (and the NLS) as a result of fusion will affect its tumor suppressor activity.

We found literature evidence for strong associations with cancer for 19 other genes in the set of 37, highlighting their driver potential. These 19 genes include the known putative drivers SGK1, PRKCA/B, and JAK2, but also new ones such as DYRK1A/B and eIF4G1/3 (**Supplementary Table 1**). Dual specificity tyrosine-phosphorylation-regulated kinases 1A and B (DYRK1A/B) have been mainly studied in the context of Down syndrome but only recently for their oncogenic potential⁴⁴. Overexpression of DYRK1A as a result of the trisomy 21 has been linked to increased risk of leukemia in children⁴⁵. Moreover, in vitro experiments revealed that the C-terminal tail of DYRK1A contains an autoinhibitory region (**Supplementary Fig. 7A**) and cleavage of this autoinhibitory region by calpain I increase kinase activity in vitro⁴⁶. Thus, removal of the AS, as a result of gene fusion, may constitutively activate DYRK1A and contribute to cancerous cell transformation in certain tissues. Of interest are also the eukaryotic translation initiation factors eIF4G1/3. Overexpression of their interaction partner eIF4E has been found in many solid tumors, where eIF4E promotes the translation of various malignancy-associated mRNAs⁴⁷. The promotion of translation relies on eIF4E's cap-binding activity and the stimulation of helicase activity of the eIF4A/eIF4G complex. eIF4G1/3 has an AS that inhibits basal helicase activity, and eIF4E is recruited to relieve the inhibitory break and induce full helicase activity (**Supplementary Fig. 7B**). Moreover, truncation constructs of eIF4G1 that lack the AS have been shown to be able to support efficient translation of capped mRNA⁴⁸. Thus, it is conceivable that removal of the AS in eIF4G1/3 as a result of fusion may promote the translation of malignancy-associated mRNAs⁴⁹⁻⁵².

Overall, the 37 genes with significantly disrupted (p)AS across the 33 cancers are significantly enriched in tyrosine kinases (Hypergeometric Test, $P < 3.92 \times 10^{-12}$). Nine (RET, PDGFRA, PDGFRB, FGFR1-3, ALK, MET, NTRK3) of them are receptor tyrosine kinases (RTKs), which represents a significant enrichment compared the number of RTKs in all cancer gene fusions (Fisher's exact test, $P < 1 \times 10^{-99}$) and RTKs in cancer fusions involving genes that encode (p)ASs (Fisher's exact test, $P < 3.17 \times 10^{-6}$). This enrichment for RTKs is not a surprise because altered autoinhibition in RTKs, as a result of gene fusion, has been found to be a driver in several human cancers. More importantly, it has been demonstrated that shortening of the autoinhibitory juxtamembrane region or the C-terminal tail in RTKs can be sufficient to constitutively activate them and induce cell transformation^{39,53}. This said, it is

clear that RTKs, like other kinases and transcription factors, often get activated in cancer fusions because of the addition of dimerization domains that promotes activating homodimerization, a mechanism that can even overcome autoinhibition^{39,53}. We analyzed how often fusions add a dimerization domain to each of the 37 identified genes. On average, 47% of all fusion events that involve the 37 genes and generate a product with a disrupted/removed (p)AS occur with a fusion partner that encodes a dimerization domain. However, 53% of all fusion events involving these genes result in products that have a disrupted (p)AS and no readily identifiable, additional dimerization domain, which suggests that altered autoinhibition may provide a rational for known or unknown oncogenic activities of these fusion products.

Significantly mutated inhibitory allosteric switches are associated with molecular signatures of cancer

We also identified human genes with recurrent somatic cancer missense mutations in (p)AS. To this end, we mapped 18,845 somatic missense mutations from NIH-GDC⁵⁴ on genes with regions that encode (p)AS across 33 cancer types and calculated enrichments using a binomial test (see Methods for details). We identified 302 genes with significant mutated (p)AS, some of them in multiple cancer types (**Fig. 4C**). Although genes with recurrently mutated (p)AS are found in 29 different cancer types, several cancers including melanoma and lung adenocarcinoma have a higher burden of genes with recurrently mutated (p)AS (**Fig. 5A**). For those cancer types with more than 10 such genes, an enrichment analysis revealed that these genes have a significant association with cancer pathways (**Supplementary Table 2**). 29 of the 302 genes with recurrently mutated (p)AS are well-established drivers (**Supplementary Table 3**), which is a significant enrichment with respect to the number of known drivers in the human proteome (Fisher's Exact Test, $P < 1.45E^{-10}$) as well as compared to known drivers among proteins with (p)AS (Fisher's Exact Test, $P < 9.58E^{-6}$). A literature search further provided evidence that 136 of the 302 genes had been associated with cancer in previous studies, which is a significant 21-fold enrichment (Fisher Exact test, $P < 2.2 \times 10^{-16}$) as compared to all genes associated with somatic cancer from the COSMIC database (**Supplementary Table 3**).

Somatic missense mutations can disrupt inhibitory allosteric interaction and, as a result, lead to increased activity of the mutated protein. For 12 genes with significant mutated (p)AS (PTPN11⁵⁵, PDGFRA⁵⁶, KIT⁵⁷, AKT3²⁴, Vav1⁵⁸, BLK⁵⁹, CARD11^{60,61}, FLT3⁶², NLRP1⁶³, NOTCH2⁶⁴, EGFR⁶⁵ and SPATA13[ASEF2]⁶⁶ PLCG1⁶⁷), we found literature support for activity increasing effects of somatic cancer missense mutations in (p)AS. For other genes, there exists indirect support for altered allosteric inhibition, as is the case of PANX1 (**Fig. 4D**). Pannexin-1, the protein encoded by PANX1, forms a hexameric plasma membrane channel that in the active form releases ATP in the extracellular space. A truncation mutation, which is found enriched in metastatic breast cancer, enhances Pannexin-1 channel activity and drives metastatic efficiency, likely due to the removal of the autoinhibitory switch in the truncation mutant^{68,69}. Cancer missense mutations in this autoinhibitory switch of PANX1 may promote metastatic activity in a similar way. Although missense mutations in (p)AS can weaken intramolecular regulatory interactions, they may also strengthen them or affect interactions with allosteric effectors and as a result change protein activity, localization or stability. An apposite example is the well-known oncogenes β -catenin (CTNNB1), which we find to have recurrent cancer missense mutations in its pAS. Consistent with the prediction of the AS, structural and binding studies have revealed that the N-terminal tail (and C-terminal tail) of β -catenin transiently interacts with other parts of the protein and, thereby, modulates interactions with binding partners⁷⁰⁻⁷². However, the N-terminal tail also harbors multiple phosphorylation sites that are key to the regulation of β -catenin's ubiquitin-

mediated degradation and subcellular localization, and the oncogenic potential of missense mutations located in this N-terminal tail has been explained mainly by changes in protein stability and localization^{73,74}.

To investigate the potential impact of missense mutations in (p)AS more broadly and search for molecular signatures associated with these mutations, we decided to exploit RNA sequencing data (RNAseq) from NIH-GDC⁵⁴. RNAseq data from 19 cancers showed that in 140 instances mutations found in the (p)AS of 126 of the 302 genes are associated with more than 10 differentially expressed genes (FDR<5%; **Fig. 5B**). 49 of these differentially expressed genes sets are significantly enriched in cancer or cancer pathway annotations (**Supplementary Table 4**). Moreover, detailed analyses of the RNAseq data revealed so-far unknown connections between somatic cancer missense mutations and molecular signatures and provide new insights into unknown or less-studied mechanistic relations. For instance, we selected all transcription factors (NFATC4, GLI1, PPARG and LARP7) among the genes that have differentially expressed genes associated with mutations in their (p)AS and investigated whether the differential expression is, at least in part, the result of increased transcriptional activity of these transcription factors. To this end, we assessed co-expression (Pearson correlation) of these transcription factors and their literature-curated target genes. For all four transcription factors, co-expression is significantly higher (T-test, **Supplementary Table 5**) in patients with mutations inside (p)AS when compared to those with mutations outside the p(AS) of these transcription factors. In the specific case of the PPARG gene, which encodes the peroxisome proliferator-activated receptor gamma and plays an oncogenic role in different cancers⁷⁵⁻⁷⁷, missense mutations in its pAS in endometrial carcinoma are associated with significantly increased RNA expressions of DKK1 (T-Test, $P < 7.0E^{-3}$, **Fig. 5C**). Dickkopf-1 (DKK1) is a secreted Wnt/ β -catenin pathway antagonist, and high expression levels of DKK1 are associated with shorter survival in various cancers, likely due to its effect on cancers' metastatic behavior^{78,79}. Intriguingly, elevated levels of PPARG have recently also been associated with more aggressive metastatic prostate cancer behavior and poor prognosis⁷⁷. The mechanistic relation between mutations in the pAS of PPARG and the increased expression of DKK1 uncovered here may link the effect of both genes on the metastasizing behavior of certain cancers. In the case of the transcription factor GLI1, missense mutations in its pAS in stomach adenocarcinoma are associated with significantly elevated RNA levels of PREX1 (T-Test, $P < 2.1E^{-2}$, **Fig. 5C**). GLI1 and PREX1 have been associated with mechanisms that sustain cancer stem cells^{80,81}. Autocrine VEGF/neuropilin-2 signaling is critical for sustaining cancer stem cells in different cancers, and neuropilin-2 signaling has recently been shown to lead to the PREX1-mediated expression of ERK, which underlies the formation and function of cancer stem cells. Importantly, VEGF/neuropilin-2 signaling also contributes to the induction of GLI1 expression, which enhances neuropilin-2 expression and thus creates the autocrine loop. The link between GLI1 and PREX1 expression, due to mutations in GLI1's pAS, provides new insights into how the autocrine loop in the VEGF/ neuropilin-2 signaling sustains stem cell properties in cancers.

An analysis of the similarity between differentially expressed genes sets associated with mutations in (p)AS of gene pairs revealed a pronounced concordance in expression changes within certain tumor types (**Fig. 5D**). An enrichment analysis revealed that the sets of overlapping differentially expressed genes in uterine corpus endometrial carcinoma, melanoma, stomach adenocarcinoma, bladder urothelial carcinoma, and lung squamous cell carcinoma are significantly associated with cancer pathways (**Supplementary Table 6**). The concordance in expression changes as a result of mutations in (p)AS of different genes suggests the existence of potential functional relationships. Indeed, for some of the gene pairs that display concordance in expression changes, mechanistic relationships are well known such as the regulation of PPARG expression by NFATC4⁸², the regulatory impact of CDC20 on

CTNNB1⁸³ or of CTNNB1 on CCND2⁸⁴. Moreover, some differentially expressed genes do not only overlap between pairs of genes with mutated (p)AS but multiple such genes. Nearly half (45%) of the genes with bladder urothelial carcinoma-associated mutations in their (p)AS change the expression of the genes SAA1, NT5E and KRT14. KRT14 has recently been shown to mark a subpopulation of bladder basal cells that play a pivotal role in tumorigenesis⁸⁵. Thus, our analysis reveals that mutations in (p)AS are associated with specific molecular signatures of cancers.

Molecular simulations validate altering effects of missense mutations in AS

Missense mutations in AS can enhance tumorigenesis via multiple effects including the direct weakening (in oncogenes) or strengthening (in tumor suppressors) of autoinhibition and the modulation of the interaction with allosteric effectors. We investigated the molecular effects of missense mutations in three genes with significantly mutated AS for which structures of switches are available: ASK1, DAPK2 and EIF4G1. ASK1⁸⁶ and DAPK2⁸⁷ have been shown to have tumor suppressor effects and missense mutations are expected to reduce activity, while EIF4G1 promotes the translation of malignancy-associated mRNAs⁴⁷ and, thus, missense mutations in its AS are expected to enhance activity. Microsecond molecular dynamics (MD) simulations of wild-type and mutant sequences revealed molecular changes that are consistent with these hypotheses.

ASK1 (MAP3K5) belongs to the apoptosis signal-regulating kinases that are part of the p38 and JNK MAP kinase pathways. Inactivating mutations in ASK1 contribute to the development of melanoma⁸⁶. A recent study determined the structure of the AS in Ask1, which consists of seven tetratricopeptide repeats and a pleckstrin homology domain (**Fig. 6A**), and revealed that the mutation R395E opens the compact structure of this AS, thereby decreasing kinase activity⁸⁸. Large-scale MD simulations of AS mutations R300W and V591M, which were found in esophageal carcinoma, demonstrate opening of the compact structure of the autoinhibitory switch, as monitored by distances between residues in its tetratricopeptide repeat core (**Fig. 6A, right**), which suggests that they may result in a loss of kinase activity similar to the R395E substitution studied previously.

DAPK2 is a pro-apoptotic Ca²⁺/calmodulin (CaM) regulated serine/threonine kinase that, like its closest homolog DAPK1, is a tumor suppressor⁸⁷. Phosphorylation of S299 in its autoinhibitory switch places negative charges next to E123 and, thus, leads, most likely as a result of disrupted autoinhibition, to increased kinase activity⁸⁹. MD simulations reveal that the uterine corpus endometrial carcinoma-associated R296H mutation changes the salt-bridge network and enables new interactions between the H296 and E123 (**Fig. 6B**), which may reduce the activating effect of the phosphorylation of nearby S299. Moreover, energy calculations suggest that the mutation may strengthen interactions between the inhibitory AS and the kinase domain (**Fig. 6B, right**). Together, these analysis results of the wild type and R296H mutation simulations are consistent with the hypothesis that the R296H mutation reduces activity/activation of DAPK2.

eIF4G1 is, as mentioned before, a eukaryotic translation initiation factor with helicase activity that is inhibited by an AS^{47,49}. Binding of the oncogenic protein eIF4E to the AS of eIF4G1 relieves autoinhibition and promotes mRNA translation. The glioblastoma-associated D613V mutation is located at the interface between eIF4G1 and its allosteric effector eIF4E. Microsecond simulations of the eIF4G1-eIF4E complex suggest that this mutation leads to the loss of intramolecular interactions between eIF4G1 residues R611 and V613 as well as the gain of new intermolecular interactions between R611 in eIF4G1 and D143 in eIF4E (**Fig. 6C**). Moreover, the D613V mutation is likely to stabilize the interaction between eIF4G1 and eIF4E as suggested by calculations of the enthalpy of binding (**Fig. 6C, right**). This stabilization of

the interaction is consistent with the hypothesis that the D613V mutations enhances the relief of autoinhibition and, thus, increases helicase activity.

DISCUSSION

Autoinhibition is a prevalent allosteric regulatory mechanism in signaling proteins because it prevents spurious pathway activation and enables signal transmission only upon integration of appropriate inputs²⁰⁻²². Therefore, it does not come as a surprise that pronounced effects on cell signaling and regulation, including effects that contribute to cancerous cell transformation and progression, have been observed when autoinhibition is modulated via mutations in individual signaling proteins^{19,57,58,90-92}. Our systems-level analysis now reveals that cancer fusion breakpoints and cancer-associated missense mutations are found with significant enrichment within inhibitory allosteric switches across cancer genomes. Importantly, the enrichment is for fusion breakpoints tied to oncogenic activity, consistent with the expected effect of removal or truncation of an inhibitory allosteric switch. This “specificity” in the enrichment of cancer fusion breakpoints within switches of oncogenes, thus, not found in tumor suppressor genes or fusion proteins associated with non-cancer diseases, highlights the potential predictive benefits of integrative studies that combine protein regulation information with cancer genomics data. Our finding that missense mutations rates are higher within allosteric switches compared to outside may be perceived as non-surprising in the light of the previous observation that cancer mutations are enriched at or close to phosphosites^{30,31} and the fact that inhibitory allosteric switches often contain phosphosite clusters that are involved in the regulation of switch function^{22,28}. However, cancer mutation rates at or close to phosphosites are significantly higher within allosteric switches compared to outside, which demonstrates that not all phosphosites are equal and that particularly those within allosteric switches are altered in tumor cells. Allosteric switches often act as signal integrating platforms where multiple allosteric effectors contribute to signal transmission, also via (de)phosphorylation or mediated via phosphorylation⁹³. Thus, the multi-functionality of these integrative platforms may make them more susceptible to interference that provides selective advantages to cancer cells.

We exploited the enrichment of cancer missense mutations and fusion breakpoints in allosteric switches for the identification of potential cancer driver genes, which follows the rationale of previously developed methods that assess the compositional bias of somatic cancer mutations in human genes^{30,32-37}. However, our approach complements existing ones as it focusses specifically on protein regulatory sites, which are often short and located in intrinsically disordered regions and as such missed by other approaches. As a result, we are also able to identify potential drivers that have only few somatic gene alterations in regions that encode short allosteric switches (**Supplementary Fig. 8**). When using cancer fusion data, we identify predominantly known cancer drivers and genes that have been associated with cancer development or progression previously, lending confidence to the overall approach and the driver potential of the selected genes that have not been studied in the cancer context before. Moreover, identified known drivers are nearly all, as expected, oncogenes. It is important to note that different mechanisms can make a fusion gene a driver, including changes in expression, localization or activity, the latter most often due to oligomerization and the removal of inhibitory allosteric switches^{6,11,53}. It has been demonstrated that removal of inhibitory switches can be sufficient to confer cell transforming potential to certain drivers but that such removal is not absolutely necessary³⁹. Autoinhibition can even be overcome by enforced oligomerization³⁹. Thus, our approach will miss drivers that have inhibitory switches but get activated because they oligomerize as a result of gene fusion and not because the switch is removed or shortened. For those cases where the fusion product has dimerization domains as well as an altered allosteric switch, oligomerization and reduced autoinhibition could synergistically enhance oncogene activity as demonstrated for individual cases previously⁵³.

Genes with significant enrichment for somatic cancer missense mutations in allosteric switches are also strongly associated with cancer development and contain numerous established drivers. Predicting the molecular phenotype that results from these mutations is, however, more complex than predicting the one that results from switch disruption in gene fusions. Mutations in allosteric switches can weaken or strengthen intramolecular regulatory interactions, thereby increasing or reducing protein activity, or modulate interactions with allosteric regulators^{61,91}. Indeed, biophysical simulations suggest that cancer-associated missense mutations in the AS of the tumor suppressors ASK1 and DAPK2 reduce activity of these genes, while they increase the activating interactions between that the oncogenes eIF4E and eIF4G1. It is important to note that missense mutations, as well as truncations as a result of gene fusion, in signal-integrating ASs can also affect other aspects of protein regulation than allostery such as cell localization or protein half-life^{74,94}, which in turn can promote cancer development. Besides providing leads for the potential molecular effects in oncogenes or tumor suppressors, our analysis demonstrates that missense mutations in allosteric switches are associated with specific transcriptional signatures of cancer. These include cancer-specific and concordant gene expression changes such the concomitant expression modulation of KRT14 as a result of bladder urothelial carcinoma-associated mutations in allosteric switches of multiple genes and specific regulatory relationships between mutated transcription factors and cancer genes.

The presented approach for the identification of potential drivers has some limitations. First, as it relies on annotations for allosteric protein regulation and as these are sparse, we also exploited predictions of allosteric switches. Thus, predictions of drivers can only be as accurate as the ones of allosteric switches. However, new proteomics methods for the mapping of structural changes that occur in signaling cascades have been developed recently that may provide new high-throughput means to identify allosteric switches and help improve predictions of these⁹⁵. Second, we only used cancer fusions and missense mutations, excluding all other types of mutations, which precludes a detailed comparison with other driver-identification approaches that include all types of mutations. Other mutation types can easily be included in the future. Finally, cancer tissue specific gene expression and splicing will determine the functional relevance of a putative driver and the presence of the allosteric switch in it. Thus, predictions need to be combined with cancer specific transcriptomics data to assess biological relevance further.

In summary, our study reveals that cancer fusions and missense mutations have a statistical preference for proteins with inhibitory allosteric switches, respectively the residues therein, which demonstrates that altered autoinhibition is positively selected for by cancer cells. We show that selection of genes with significantly mutated allosteric switches provides a fruitful way to identify cancer drivers. This new way of identifying drivers is particularly attractive in the light of recent successes in the design and clinical use of precision medicines that stabilize autoinhibited states of oncogenes^{25,26,96,97}.

METHODS

Data collections and processing

Proteins with experimentally-characterized inhibitory allosteric switches (AS) were identified in a literature search as described previously^{22,28}. Boundaries of AS were taken from the original papers in which either assays on deletion constructs or detailed structural characterizations were used to map autoinhibitory protein regions. 234 human proteins with

AS were found in the literature search. Proteins with predicted AS (pAS) were identified by using the predictor Cis-regPred²⁸. Cis-regPred identifies intrinsically disordered allosteric switches based on protein sequence information as well as publically available proteomics data. Features that Cis-regPred uses to identify pAS include, among others, the amino acid compositional bias, intrinsic disorder prediction scores, enrichment for splicing sites and post-translational modifications of autoinhibitory regions. In addition, electrostatic complementarity between the allosteric switch and other parts of the same proteins is used. Cis-regPred has been trained and validated on independent sets of proteins with autoinhibitory regions that have been mapped by experimental means. Although trained exclusively on inhibitory allosteric switches (autoinhibitory protein regions), Cis-regPred has been shown to identify also activating allosteric switches. This said, most of the pAS identified by Cis-regPred are likely to be inhibitory in nature. Intrinsic disorder prediction scores from Cis-regPred were also used to identify intrinsically disordered protein regions.

Fusion genes were collected from the databases FusionCancer⁹⁸ and ChimerDB 3.0⁹⁹ as well as from Yoshihara *et al*¹⁰⁰. Only fusion genes were considered in which the fusion breakpoint is within the protein coding region. For this analysis, genomic co-ordinates of fusion proteins breakpoints were mapped to protein co-ordinates from annotation file downloaded from Ensembl¹⁰¹. In addition, to identify functional fusion proteins, only fusion proteins were considered with functional domain(s). Domains were identified based on annotations provided by PFAM¹⁰² and InterPro¹⁰³. Overall, 12'957 fusion genes were collected, and 5'503 in-frame fusion proteins identified that were found in cancer cells and had the fusion breakpoint within protein coding regions. 756 proteins were identified that occurred only in non-cancer fusions.

7'799 cancer mutations and 17'549 phosphorylated residues for the initial enrichment analysis were downloaded from COSMIC database²⁹ and phosphositePlus¹⁰⁴. Gene expression data (level 3), 18'845 somatic mutations for the identification of recurrently mutated (p)AS, and mapped read count data were downloaded from NCI-Genome Data Commons⁵⁴.

Fusion breakpoint and cancer mutation enrichment analysis

The hypergeometric test was used to assess the enrichment of breakpoints/junctions of fusion genes within (p)AS. Thus, the number of breakpoints disrupting (p)AS regions was compared to the number of breakpoints outside the region, using the number of residues in and outside of (p)AS for normalization. In a second enrichment analysis, the boundary of the (p)AS was extended by including the linker between the (p)AS and the regulated functional domain of each protein. This was done as breakpoints may not fall within the (p)AS region but between the (p)AS and the regulated domain, thus resulting in the removal of the (p)AS but keeping the functional domain intact. For known AS, boundaries of AS and the regulated domain are clearly defined, which enables identification of the "extended AS". For predicted AS, it is not a priori clear which domain or sequence region is regulated by the (p)AS, particularly for multidomain proteins. Therefore, the following rules were applied. 1. For single domain proteins with pAS, the unique domain present was considered the domain that is regulated by the pAS. 2. For multidomain proteins, only those proteins were considered that have a kinase, DNA-binding or guanine nucleotide exchange factor domain. These domains were assumed to be regulated by the pAS.

To assess the enrichment of for COSMIC cancer mutations within (p)AS, the hypergeometric test was applied as explained for the fusion breakpoints. Averaged mutations rates per residue inside and outside of (p)AS were compare using the T-test.

Identification of genes with significantly altered (p)AS

For the identification of significantly mutated (p)AS, 18'845 somatic mutations from 10,177 patients were downloaded from NCI GDC⁵⁴. Cancer-specific missense mutations were mapped to (p)AS regions using BEDtools¹⁰⁵. Then P-values of the binomial probabilities of finding k or more missense mutations within regions that encode the (p)AS were calculated under the assumption that all residues have the same probability of being mutated. Thus, given the total number of mutations in a protein, its length and the length of the (p)AS, one can calculate the binomial probabilities of finding k or more missense mutations within regions that encode the (p)AS under the null hypothesis that the probability of mutations is equal across the gene, an approach also used by e-Driver³⁶. Consistent with the null hypothesis, no significant difference in the rate of coding-silent mutations inside and outside of (p)AS was detected (**Supplementary Fig. 9**). Moreover, several algorithms have recently been used and compared in the identification of cancer drivers, including e-Driver¹⁰⁶. Different quality metrics were applied to compare methods and eliminate outlier predictions. One of these metrics is the divergence of observed from theoretically expected P values, which, when large, may indicate that a methods statistical assumption is not well satisfied. E-Driver produced less often outliers, when analyzing cancer-specific data, than many of the other established driver-identification methods, thus providing confidence for the statistical approach used. Once P values for all (p)AS in each gene of a specific cancer were calculated, the Benjamini–Hochberg false discovery rate algorithm was applied to correct for multiple testing.

For fusion breakpoints, the same procedure was used but missense mutations replaced by breakpoint locations and (p)AS replaced by extended (p)AS (for extended (p)AS definition see above). A cancer-specific analysis of fusion data did not provide any hits for significantly altered (p)AS, thus a pan-cancer analysis was carried out instead. Genes with significantly altered (p)AS as a result of gene fusion were identified and ranked after correction for multiple testing using the Benjamini–Hochberg false discovery rate algorithm.

Dimerization domain analysis

A list of dimerization domains was compiled based on PFAM¹⁰² and InterPro¹⁰³ annotations. These dimerization domains were then mapped to the fusion proteins and assessed how often fusion involving proteins with (p)AS lead to the addition of a new dimerization domain.

Differential expression and downstream analysis

Mapped raw reads count data downloaded from NCI-Genome Data Commons⁵⁴ was used for the identification of genes differentially expressed between patients with mutations inside and outside the (p)AS. Patients having mutations within as well as outside the (p)AS region were removed from the analyses. Further, only those mutations were selected for which at least 2 replicates exist. This was done to reduce the variance in the data and improve the identification of differentially expressed genes. This selection criteria resulted in 79 BLCA, 33 BRCA, 31 CESC, 83 COAD, 12 ESCA, 33 HNSC, 4 KIRC, 4 KIRP, 5 LAML, 90 LIHC, 165 LUAD, 137 LUSC, 10 OV, 5 PAAD, 16 PRAD, 8 READ, 297 SKCM, 78 STAD and, 306 UCEC tissues/samples which have matched count and mutation data. After initial filtering, DESeq2 was used to identify differentially expressed genes by taking adjusted P-value of $\ll 0.05$. Further, we only considered genes with recurrently mutated (p)AS that have at least 10 differentially expressed genes. These filtering result in 140 gene-cancer combinations consisting of 126 unique genes in 19 cancer tissues. Enrichment of differentially expressed genes was done by EnrichR¹⁰⁷ and PathwAX¹⁰⁸. Then the overlap of these sets of differentially expressed genes within cancers was determined. This was done by iterating the differentially

expressed (DE) genes starting from top 10% of genes in each cancer followed by 10% increment in each iteration till all the genes were covered (100%). Each iteration in a cancer set was compared with iteration in other cancer set thus, resulting in 10 X 10 steps. For each steps log odd ratio and p-value was calculated using Fisher's Exact test. For clustering log odd ratio of most significant (lowest) p-value was extracted for each pair-wise overlapping DE genes set. their log odd ratio calculated and Fisher's exact test applied to find the significance of overlap.

Expression correlation between transcription factors with significantly mutated (p)AS and their target genes

(p)AS containing transcription factors (TFs) that are enriched in somatic cancer missense mutation were identified as described above. Mutations mapping to these TFs were mapped to corresponding gene expression data sets (FPKM values) that were downloaded from NCI GDC. For the mapping, unique IDs of mutations and RNA-seq expression data were matched. The effect of mutation on TF activity was assessed by comparing expression data from patients with mutations within and outside the (p)AS. Targets of transcription factors involved in this study was downloaded from GTRD database¹⁰⁹. Once all the targets of TFs were downloaded gene expression profile based co-expression between TF and target genes was calculated as Pearson Correlation Coefficient (PCC). P-values of co-expression (PCC) were calculated using "T-test". To find the significant change in PCC between mutations within AI/CRE region to mutation outside AI/CRE region, T-test was applied.

Statistical analyses

Statistical analyses were performed using the JAVA apache maths library "R". Gene Ontology and pathway enrichment analyses were done using EnrichR¹⁰⁷ and PathwAX¹⁰⁸.

Molecular dynamics simulations

We performed 2.5 μ s molecular dynamics (MD) simulations for each studied system of both wild-type and mutants. Prior MD production, missing electron loop densities of crystal structures were reconstructed using Rosetta loop remodelling. Next, protonation states for all titratable residues in the resulting structures were determined. The effective pKa for each titratable group was computed using the WHATIF pKa calculation package as described previously¹¹⁰. The linearized Poisson-Boltzmann equation was computed using a dielectric constant of four for the protein using Adaptive Poisson-Boltzmann electrostatic solver¹¹¹. Molecular dynamics simulations were carried out using the modified AMBER ff14SB force field¹¹². The simulations began with 5000 steps of steepest descent energy minimization were followed by the reassignments of velocities from the Maxwell distribution at 300 K every 1 ps for 1000 ps, and by a final equilibration of the system for 2.0 ns. After equilibration, the molecular dynamics trajectories were integrated using the GPU-accelerated *pmemd* engine in Amber14. The simulations were conducted using the isobaric-isothermal ensemble¹¹³ at 300 K and 1 atmosphere and using long-range non-bonded interactions with a 12 Å residue-based cutoff. Long-range electrostatic forces were calculated using the particle-mesh Ewald sum¹¹⁴. Bonds to hydrogen atoms were maintained with the SHAKE algorithm¹¹⁵ and an integration step size of two femtoseconds was used.

ACKNOWLEDGEMENTS

We thank the Canadian Institute of Health Research (CIHR), Genome Canada, Genome BC and the Michael Smith Foundation for Health Research (MSFHR) for their valuable support. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014R1A1A1003444).

FIGURE CAPTIONS

Figure 1. Systematic characterization of altered allosteric protein regulation in cancer. Shown is the flowchart for the integration and analysis of cancer genomics and protein regulation data. In-frame cancer fusion proteins and their breakpoints as well as missense mutations from 33 cancer types are matched with human proteins that have known or predicted allosteric switches (AS). The analysis contains three steps: (1) The enrichment of cancer-associated genetic alterations within AS is assessed. (2) Genes with significantly altered AS are identified. (3) Molecular consequences of altered allosteric regulation are investigated and associated molecular signatures identified.

Figure 2. Involvement of allosterically regulated proteins in cancer fusion

(A and B) Enrichment of proteins with known (A) or predicted (B) allosteric switches among proteins that are part of cancer fusions (Fisher's exact test).

(C and D) Enrichment of fusion breakpoints within known (C) and predicted (D) allosteric switches. Left, enrichment for all cancer fusions as well as the subset of cancer fusions that involves known driver genes. Right, enrichment for non-cancer fusions as well as the subset of non-cancer fusions that involves cancer driver genes. (Hypergeometric test)

Figure 3. Missense mutations of allosteric switches

(A) Percentage of residues within and outside of (p)AS (top AS and bottom pAS) that are hit by missense mutation (hypergeometric test).

(B) Averaged mutation rates for residues within and outside of (p)AS. The distributions of averaged mutation rates were compared using Student's T-test.

(C) Averaged mutation rates for phosphorylation sites within and outside of (p)AS.

(D) Averaged mutation rates for phosphorylation sites and adjacent residues (± 5 residues) within and outside of (p)AS.

Figure 4. Genes with recurrently altered allosteric switches

(A) 39 genes with recurrent fusion breakpoints in their (p)AS across 33 cancer types ranked according to the FDR-corrected P values. Cancers in which breakpoints were found in the (p)AS are indicated by the color. Known cancer driver genes are highlighted in bold.

(B) Structural details of the gene MYB and their changes in the MYB-NFIB fusion that is found in breast invasive carcinoma. MYB encodes the transcriptional activator Myb that contains a DNA-binding domain (DBD) and a large C-terminal segment that acts as allosteric switch (AS). In the MYB-NFIB fusion, the fusion breakpoint is located in the AS (highlighted by the red arrow), which, as a result, gets shortened.

(C) Top ranked genes with recurrent missense mutations from a specific cancer in their (p)AS. Genes are ranked according to the FDR-corrected P values. Cancer types in which these

missense mutations were found are indicated by the color. Known cancer driver genes are highlighted in bold.

(D) Structural details of the gene PANX1 with the location of missense mutations in its AS.

PANX1 encodes the plasma membrane channel pannexin-1 that contains an innexin domain with four transmembrane segments (TMS) at the N-terminus and an inhibitory allosteric switch at the C-terminus.

Figure 5. Recurrently mutated (p)AS are associated with distinct molecular signatures

(A) Frequency distribution of recurrently mutated (p)AS among different cancer types. The 15 cancer types with the highest number of recurrently mutated (p)AS are shown.

(B) Matched RNAseq data from 19 cancer types revealed that mutations in (p)AS of 140 different proteins are associated with more than 10 differentially expressed genes (FDR<0.05). The number of differentially expressed genes is color coded in log scale.

(C) RNAseq signals for DKK1 in endometrial carcinoma and PREX1 in stomach adenocarcinoma are shown. DKK1 and PREX1 were identified as top differentially expressed genes with highest log fold changes when comparing RNAseq data for mutations within and outside the pAS of PPARG and GLI1, respectively. The red lines represent signals for samples with mutations in the pAS of PPARG and GLI1.

(D) Overlap between differentially expressed genes as a result of mutations in the (p)AS of 140 proteins. Overlapping differentially expressed genes were identified using Fisher's exact test and clustered based on log odd ratios of the overlap. The heat map reveals that cancer-specific mutations located in (p)AS lead to overlapping gene expression changes.

Figure 6. Molecular dynamics simulations of wild type and mutated allosteric switches.

(A) Left, allosteric switch of wild type ASK1 in ribbon representation (red) with mutated residues R300 and V591 highlighted in green. Middle, representative structures of the wild type (red) and mutated (magenta and purple) allosteric switches after 2.5 μ s MD simulations. Structures were aligned on the PH domain (top left), to help visualizing the opening of the tetratricopeptide repeat core in the mutants. Right, distances between residues R322 in W509 and the corresponding distributions, which were used to monitor compactness of the allosteric switch.

(B) Left, ribbon representation of wild type DAPK2 with kinase domain in blue and allosteric switch in red. The mutated residue R296 is highlighted in green. Middle, structural detail of the interface between kinase domain and allosteric switch after 2.5 μ s of MD simulations of wild type and R296H mutant. Right, distribution of the distance between E123 and R296 (H296) in wild type and mutant simulation. Distribution of enthalpy of binding between kinase and switch for wild type and mutant simulation.

(C) Left, ribbon representation of the allosteric switch of wild type eIF4G1 (red) in complex with eIF4E (grey). The mutated residue D613 is highlighted in green. Middle, structural detail of the interface between eIF4G1 and eIF4E after 2.5 μ s of MD simulations of wild type and D613V mutant. Right, distribution of enthalpy of binding between eIF4G1 and eIF4E for wild type and mutant simulation.

REFERENCES

- 1 Futreal, P. A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183, doi:10.1038/nrc1299 (2004).
- 2 Schlabach, M. R. *et al.* Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620-624, doi:10.1126/science.1149200 (2008).
- 3 Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214-218, doi:10.1038/nature12213 (2013).
- 4 Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501, doi:10.1038/nature12912 (2014).
- 5 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
- 6 Stransky, N., Cerami, E., Schalm, S., Kim, J. L. & Lengauer, C. The landscape of kinase fusions in cancer. *Nat Commun* **5**, 4846, doi:10.1038/ncomms5846 (2014).
- 7 Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724, doi:10.1038/nature07943 (2009).
- 8 Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).
- 9 Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer. *Annual review of pathology* **10**, 25-50, doi:10.1146/annurev-pathol-012414-040312 (2015).
- 10 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
- 11 Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* **15**, 371-381, doi:10.1038/nrc3947 (2015).
- 12 Lobato, M. N. *et al.* Modeling chromosomal translocations using conditional alleles to recapitulate initiating events in human leukemias. *Journal of the National Cancer Institute. Monographs*, 58-63, doi:10.1093/jncimonographs/lgn022 (2008).
- 13 Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res* **44**, 4487-4503, doi:10.1093/nar/gkw282 (2016).
- 14 Faderl, S. *et al.* The biology of chronic myeloid leukemia. *The New England journal of medicine* **341**, 164-172, doi:10.1056/nejm199907153410306 (1999).
- 15 Woessner, D. W., Lim, C. S. & Deininger, M. W. Development of an effective therapy for chronic myelogenous leukemia. *Cancer journal (Sudbury, Mass.)* **17**, 477-486, doi:10.1097/PPO.0b013e318237e5b7 (2011).
- 16 Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing Genome-Driven Oncology. *Cell* **168**, 584-599, doi:10.1016/j.cell.2016.12.015 (2017).
- 17 Haslem, D. S. *et al.* Precision oncology in advanced cancer patients improves overall survival with lower weekly healthcare costs. *Oncotarget* **9**, 12316-12322, doi:10.18632/oncotarget.24384 (2018).
- 18 Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* **112**, 831-843 (2003).
- 19 Nagar, B. *et al.* Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859-871 (2003).
- 20 Pufall, M. A. & Graves, B. J. Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol* **18**, 421-462 (2002).

- 21 Schlessinger, J. Signal transduction. Autoinhibition control. *Science* **300**, 750-752, doi:10.1126/science.1082024 (2003).
- 22 Trudeau, T. *et al.* Structure and intrinsic disorder in protein autoinhibition. *Structure* **21**, 332-341, doi:10.1016/j.str.2012.12.013 (2013).
- 23 Reindl, C. & Spiekermann, K. From kinases to cancer: leakiness, loss of autoinhibition and leukemia. *Cell Cycle* **5**, 599-602, doi:10.4161/cc.5.6.2586 (2006).
- 24 Parikh, C. *et al.* Disruption of PH-kinase domain interactions leads to oncogenic activation of AKT in human cancers. *Proc Natl Acad Sci U S A* **109**, 19368-19373, doi:10.1073/pnas.1204384109 (2012).
- 25 Chen, Y. N. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148-152, doi:10.1038/nature18621 (2016).
- 26 Tap, W. D. *et al.* Structure-Guided Blockade of CSF1R Kinase in Tenosynovial Giant-Cell Tumor. *The New England journal of medicine* **373**, 428-437, doi:10.1056/NEJMoa1411366 (2015).
- 27 Peterson, J. R. & Golemis, E. A. Autoinhibited proteins as promising drug targets. *J Cell Biochem* **93**, 68-73 (2004).
- 28 Yeon, J. H., Heinkel, F., Sung, M., Na, D. & Gsponer, J. Systems-wide Identification of cis-Regulatory Elements in Proteins. *Cell Syst* **2**, 89-100, doi:10.1016/j.cels.2016.02.004 (2016).
- 29 Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805-811, doi:10.1093/nar/gku1075 (2015).
- 30 Reimand, J. & Bader, G. D. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Mol Syst Biol* **9**, 637, doi:10.1038/msb.2012.68 (2013).
- 31 Creixell, P. *et al.* Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* **163**, 202-217, doi:10.1016/j.cell.2015.08.056 (2015).
- 32 Tamborero, D., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* **29**, 2238-2244, doi:10.1093/bioinformatics/btt395 (2013).
- 33 Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nature methods* **10**, 1081-1082, doi:10.1038/nmeth.2642 (2013).
- 34 Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res* **40**, e169, doi:10.1093/nar/gks743 (2012).
- 35 Tokheim, C. J., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Karchin, R. Evaluating the evaluation of cancer driver genes. *Proc Natl Acad Sci U S A* **113**, 14330-14335, doi:10.1073/pnas.1616440113 (2016).
- 36 Porta-Pardo, E. & Godzik, A. e-Driver: a novel method to identify protein regions driving cancer. *Bioinformatics* **30**, 3109-3114, doi:10.1093/bioinformatics/btu499 (2014).
- 37 Araya, C. L. *et al.* Identification of significantly mutated regions across cancer types highlights a rich landscape of functional molecular alterations. *Nat Genet* **48**, 117-125, doi:10.1038/ng.3471 (2016).
- 38 Shin, C. H., Grossmann, A. H., Holmen, S. L. & Robinson, J. P. The BRAF kinase domain promotes the development of gliomas in vivo. *Genes & cancer* **6**, 9-18, doi:10.18632/genesandcancer.48 (2015).
- 39 Stover, E. H. *et al.* Activation of FIP1L1-PDGFRalpha requires disruption of the juxtamembrane domain of PDGFRalpha and is FIP1L1-independent. *Proc Natl Acad Sci U S A* **103**, 8078-8083, doi:10.1073/pnas.0601192103 (2006).

- 40 George, O. L. & Ness, S. A. Situational awareness: regulation of the myb transcription factor in differentiation, the cell cycle and oncogenesis. *Cancers* **6**, 2049-2071, doi:10.3390/cancers6042049 (2014).
- 41 Fabbro, M. & Henderson, B. R. Regulation of tumor suppressors by nuclear-cytoplasmic shuttling. *Exp Cell Res* **282**, 59-69 (2003).
- 42 Vandembroucke, I., Van Oostveldt, P., Coene, E., De Paepe, A. & Messiaen, L. Neurofibromin is actively transported to the nucleus. *FEBS Lett* **560**, 98-102, doi:10.1016/s0014-5793(04)00078-x (2004).
- 43 Koliou, X., Fedonidis, C., Kalpachidou, T. & Mangoura, D. Nuclear import mechanism of neurofibromin for localization on the spindle and function in chromosome congression. *J Neurochem* **136**, 78-91, doi:10.1111/jnc.13401 (2016).
- 44 Fernandez-Martinez, P., Zahonero, C. & Sanchez-Gomez, P. DYRK1A: the double-edged kinase as a protagonist in cell growth and tumorigenesis. *Molecular & cellular oncology* **2**, e970048, doi:10.4161/23723548.2014.970048 (2015).
- 45 Malinge, S. *et al.* Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *The Journal of clinical investigation* **122**, 948-962, doi:10.1172/jci60455 (2012).
- 46 Jin, N. *et al.* Truncation and Activation of Dual Specificity Tyrosine Phosphorylation-regulated Kinase 1A by Calpain I: A MOLECULAR MECHANISM LINKED TO TAU PATHOLOGY IN ALZHEIMER DISEASE. *J Biol Chem* **290**, 15219-15237, doi:10.1074/jbc.M115.645507 (2015).
- 47 Bhat, M. *et al.* Targeting the translation machinery in cancer. *Nat Rev Drug Discov* **14**, 261-278, doi:10.1038/nrd4505 (2015).
- 48 Ali, I. K., McKendrick, L., Morley, S. J. & Jackson, R. J. Truncated initiation factor eIF4G lacking an eIF4E binding site can support capped mRNA translation. *Embo j* **20**, 4233-4242, doi:10.1093/emboj/20.15.4233 (2001).
- 49 Feoktistova, K., Tuvshintogs, E., Do, A. & Fraser, C. S. Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. *Proc Natl Acad Sci U S A* **110**, 13339-13344, doi:10.1073/pnas.1303781110 (2013).
- 50 Badura, M., Braunstein, S., Zavadil, J. & Schneider, R. J. DNA damage and eIF4G1 in breast cancer cells reprogram translation for survival and DNA repair mRNAs. *Proc Natl Acad Sci U S A* **109**, 18767-18772, doi:10.1073/pnas.1203853109 (2012).
- 51 Braunstein, S. *et al.* A hypoxia-controlled cap-dependent to cap-independent translation switch in breast cancer. *Molecular cell* **28**, 501-512, doi:10.1016/j.molcel.2007.10.019 (2007).
- 52 Ho, J. J. D. *et al.* Systemic Reprogramming of Translation Efficiencies on Oxygen Stimulus. *Cell reports* **14**, 1293-1300, doi:10.1016/j.celrep.2016.01.036 (2016).
- 53 Medves, S. & Demoulin, J. B. Tyrosine kinase gene fusions in cancer: translating mechanisms into targeted therapies. *Journal of cellular and molecular medicine* **16**, 237-248, doi:10.1111/j.1582-4934.2011.01415.x (2012).
- 54 Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *The New England journal of medicine* **375**, 1109-1112, doi:10.1056/NEJMp1607591 (2016).
- 55 Tartaglia, M. *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet* **34**, 148-150, doi:10.1038/ng1156 (2003).
- 56 Hirota, S. *et al.* Gain-of-function mutations of platelet-derived growth factor receptor alpha gene in gastrointestinal stromal tumors. *Gastroenterology* **125**, 660-667 (2003).
- 57 Chan, P. M., Ilangumaran, S., La Rose, J., Chakrabarty, A. & Rottapel, R. Autoinhibition of the kit receptor tyrosine kinase by the cytosolic juxtamembrane region. *Mol Cell Biol* **23**, 3067-3078 (2003).

- 58 Razanadrakoto, L. *et al.* Mutation of Vav1 adaptor region reveals a new oncogenic activation. *Oncotarget* **6**, 2524-2537, doi:10.18632/oncotarget.2629 (2015).
- 59 Petersen, D. L. *et al.* B-lymphoid tyrosine kinase (Blk) is an oncogene and a potential target for therapy with dasatinib in cutaneous T-cell lymphoma (CTCL). *Leukemia* **28**, 2109-2112, doi:10.1038/leu.2014.192 (2014).
- 60 Lenz, G. *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676-1679, doi:10.1126/science.1153629 (2008).
- 61 Lamason, R. L., McCully, R. R., Lew, S. M. & Pomerantz, J. L. Oncogenic CARD11 mutations induce hyperactive signaling by disrupting autoinhibition by the PKC-responsive inhibitory domain. *Biochemistry* **49**, 8240-8250, doi:10.1021/bi101052d (2010).
- 62 Frohling, S. *et al.* Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer cell* **12**, 501-513, doi:10.1016/j.ccr.2007.11.005 (2007).
- 63 Zhong, F. L. *et al.* Germline NLRP1 Mutations Cause Skin Inflammatory and Cancer Susceptibility Syndromes via Inflammasome Activation. *Cell* **167**, 187-202.e117, doi:10.1016/j.cell.2016.09.001 (2016).
- 64 Gordon, W. R. *et al.* Structural basis for autoinhibition of Notch. *Nat Struct Mol Biol* **14**, 295-300, doi:10.1038/nsmb1227 (2007).
- 65 Yun, C. H. *et al.* Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer cell* **11**, 217-227, doi:10.1016/j.ccr.2006.12.017 (2007).
- 66 Kawasaki, Y. *et al.* Identification and characterization of Asef2, a guanine-nucleotide exchange factor specific for Rac1 and Cdc42. *Oncogene* **26**, 7620-7267, doi:10.1038/sj.onc.1210574 (2007).
- 67 Behjati, S. *et al.* Recurrent PTPRB and PLCG1 mutations in angiosarcoma. *Nat Genet* **46**, 376-379, doi:10.1038/ng.2921 (2014).
- 68 Furlow, P. W. *et al.* Mechanosensitive pannexin-1 channels mediate microvascular metastatic cell survival. *Nature cell biology* **17**, 943-952, doi:10.1038/ncb3194 (2015).
- 69 Sandilos, J. K. *et al.* Pannexin 1, an ATP release channel, is activated by caspase cleavage of its pore-associated C-terminal autoinhibitory region. *J Biol Chem* **287**, 11303-11311, doi:10.1074/jbc.M111.323378 (2012).
- 70 Xing, Y. *et al.* Crystal structure of a full-length beta-catenin. *Structure* **16**, 478-487, doi:10.1016/j.str.2007.12.021 (2008).
- 71 Wahl, J. K. r. *et al.* The amino- and carboxyl-terminal tails of (beta)-catenin reduce its affinity for desmoglein 2. *J Cell Sci* **113 (Pt 10)**, 1737-1745 (2000).
- 72 Choi, H. J., Huber, A. H. & Weis, W. I. Thermodynamics of beta-catenin-ligand interactions: the roles of the N- and C-terminal tails in modulating binding affinity. *J Biol Chem* **281**, 1027-1038, doi:10.1074/jbc.M511338200 (2006).
- 73 Dar, M. S. *et al.* Terminal regions of beta-catenin are critical for regulating its adhesion and transcription functions. *Biochim Biophys Acta* **1863**, 2345-2357, doi:10.1016/j.bbamcr.2016.06.010 (2016).
- 74 Dar, M. S., Singh, P., Mir, R. A. & Dar, M. J. Betaeta-catenin N-terminal domain: An enigmatic region prone to cancer causing mutations. *Mutation research* **773**, 122-133, doi:10.1016/j.mrrev.2017.06.001 (2017).
- 75 Saez, E. *et al.* Activators of the nuclear receptor PPARgamma enhance colon polyp formation. *Nature medicine* **4**, 1058-1061, doi:10.1038/2042 (1998).
- 76 Goldstein, J. T. *et al.* Genomic Activation of PPARG Reveals a Candidate Therapeutic Axis in Bladder Cancer. *Cancer Res* **77**, 6987-6998, doi:10.1158/0008-5472.can-17-1701 (2017).

- 77 Ahmad, I. *et al.* Sleeping Beauty screen reveals Pparg activation in metastatic prostate cancer. *Proc Natl Acad Sci U S A* **113**, 8290-8295, doi:10.1073/pnas.1601571113 (2016).
- 78 Mazon, M., Masi, D. & Carreau, M. Modulating Dickkopf-1: A Strategy to Monitor or Treat Cancer? *Cancers* **8**, doi:10.3390/cancers8070062 (2016).
- 79 Zhuang, X. *et al.* Differential effects on lung and bone metastasis of breast cancer by Wnt signalling inhibitor DKK1. *Nature cell biology* **19**, 1274-1285, doi:10.1038/ncb3613 (2017).
- 80 Goel, H. L. *et al.* GLI1 regulates a novel neuropilin-2/alpha6beta1 integrin based autocrine pathway that contributes to breast cancer initiation. *EMBO molecular medicine* **5**, 488-508, doi:10.1002/emmm.201202078 (2013).
- 81 Goel, H. L. *et al.* P-Rex1 Promotes Resistance to VEGF/VEGFR-Targeted Therapy in Prostate Cancer. *Cell reports* **14**, 2193-2208, doi:10.1016/j.celrep.2016.02.016 (2016).
- 82 Yang, T. T., Xiong, Q., Enslen, H., Davis, R. J. & Chow, C. W. Phosphorylation of NFATc4 by p38 mitogen-activated protein kinases. *Mol Cell Biol* **22**, 3892-3904 (2002).
- 83 Hadjihannas, M. V., Bernkopf, D. B., Bruckner, M. & Behrens, J. Cell cycle control of Wnt/beta-catenin signalling by conductin/axin2 through CDC20. *EMBO Rep* **13**, 347-354, doi:10.1038/embor.2012.12 (2012).
- 84 Jiang, X. *et al.* Targeting beta-catenin signaling for therapeutic intervention in MEN1-deficient pancreatic neuroendocrine tumours. *Nat Commun* **5**, 5809, doi:10.1038/ncomms6809 (2014).
- 85 Papafotiou, G. *et al.* KRT14 marks a subpopulation of bladder basal cells with pivotal role in regeneration and tumorigenesis. *Nat Commun* **7**, 11914, doi:10.1038/ncomms11914 (2016).
- 86 Stark, M. S. *et al.* Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nat Genet* **44**, 165-169, doi:10.1038/ng.1041 (2011).
- 87 Schlegel, C. R. *et al.* DAPK2 regulates oxidative stress in cancer cells by preserving mitochondrial function. *Cell death & disease* **6**, e1671, doi:10.1038/cddis.2015.31 (2015).
- 88 Weijman, J. F. *et al.* Structural basis of autoregulatory scaffolding by apoptosis signal-regulating kinase 1. *Proc Natl Acad Sci U S A* **114**, E2096-e2105, doi:10.1073/pnas.1620813114 (2017).
- 89 Isshiki, K., Matsuda, S., Tsuji, A. & Yuasa, K. cGMP-dependent protein kinase I promotes cell apoptosis through hyperactivation of death-associated protein kinase 2. *Biochem Biophys Res Commun* **422**, 280-284, doi:10.1016/j.bbrc.2012.04.148 (2012).
- 90 Katzav, S., Cleveland, J. L., Heslop, H. E. & Pulido, D. Loss of the amino-terminal helix-loop-helix domain of the vav proto-oncogene activates its transforming potential. *Mol Cell Biol* **11**, 1912-1920 (1991).
- 91 Hata, A., Lo, R. S., Wotton, D., Lagna, G. & Massague, J. Mutations increasing autoinhibition inactivate tumour suppressors Smad2 and Smad4. *Nature* **388**, 82-87, doi:10.1038/40424 (1997).
- 92 Yang, H. *et al.* Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. *Nature* **552**, 368-373, doi:10.1038/nature25023 (2017).
- 93 Lim, W. A. The modular logic of signaling proteins: building allosteric switches from simple binding domains. *Curr Opin Struct Biol* **12**, 61-68 (2002).
- 94 Mazhab-Jafari, M. T. *et al.* Oncogenic and RASopathy-associated K-RAS mutations relieve membrane-dependent occlusion of the effector-binding site. *Proc Natl Acad Sci U S A* **112**, 6625-6630, doi:10.1073/pnas.1419895112 (2015).

- 95 Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* **32**, 1036-1044, doi:10.1038/nbt.2999 (2014).
- 96 Lapiere, J. M. *et al.* Discovery of 3-(3-(4-(1-Aminocyclobutyl)phenyl)-5-phenyl-3H-imidazo[4,5-b]pyridin-2-yl)pyridin -2-amine (ARQ 092): An Orally Bioavailable, Selective, and Potent Allosteric AKT Inhibitor. *J Med Chem* **59**, 6455-6469, doi:10.1021/acs.jmedchem.6b00619 (2016).
- 97 Skora, L., Mestan, J., Fabbro, D., Jahnke, W. & Grzesiek, S. NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. *Proc Natl Acad Sci U S A* **110**, E4437-4445, doi:10.1073/pnas.1314712110 (2013).
- 98 Wang, Y., Wu, N., Liu, J., Wu, Z. & Dong, D. FusionCancer: a database of cancer fusion genes derived from RNA-seq data. *Diagnostic pathology* **10**, 131, doi:10.1186/s13000-015-0310-4 (2015).
- 99 Lee, M. *et al.* ChimerDB 3.0: an enhanced database for fusion genes from cancer transcriptome and literature data mining. *Nucleic Acids Res* **45**, D784-d789, doi:10.1093/nar/gkw1083 (2017).
- 100 Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845-4854, doi:10.1038/onc.2014.406 (2015).
- 101 Yates, A. *et al.* Ensembl 2016. *Nucleic Acids Res* **44**, D710-716, doi:10.1093/nar/gkv1157 (2016).
- 102 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285, doi:10.1093/nar/gkv1344 (2016).
- 103 Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* **45**, D190-d199, doi:10.1093/nar/gkw1107 (2017).
- 104 Hornbeck, P. V. *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-270, doi:10.1093/nar/gkr1122 (2012).
- 105 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 106 Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e318, doi:10.1016/j.cell.2018.02.060 (2018).
- 107 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 108 Ogris, C., Helleday, T. & Sonnhammer, E. L. PathwAX: a web server for network crosstalk based pathway annotation. *Nucleic Acids Res* **44**, W105-109, doi:10.1093/nar/gkw356 (2016).
- 109 Yevshin, I., Sharipov, R., Valeev, T., Kel, A. & Kolpakov, F. GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. *Nucleic Acids Res* **45**, D61-d67, doi:10.1093/nar/gkw951 (2017).
- 110 Vriend, G. WHAT IF: a molecular modeling and drug design program. *Journal of molecular graphics* **8**, 52-56, 29 (1990).
- 111 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).
- 112 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of chemical theory and computation* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).
- 113 Andersen, H. C. Molecular Dynamics Simulations at Constant Pressure and-or Temperature *Journal of Chemical Physics* **72**, 2384-2393 (1980).

- 114 Darden, T., York, D. & Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **98**, 10089-10092 (1993).
- 115 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical-integration of cartesian equations of motion of a system with constraints - Molecular dynamics of n-alkanes. *J Comput Phys* **23**, 327-341 (1977).

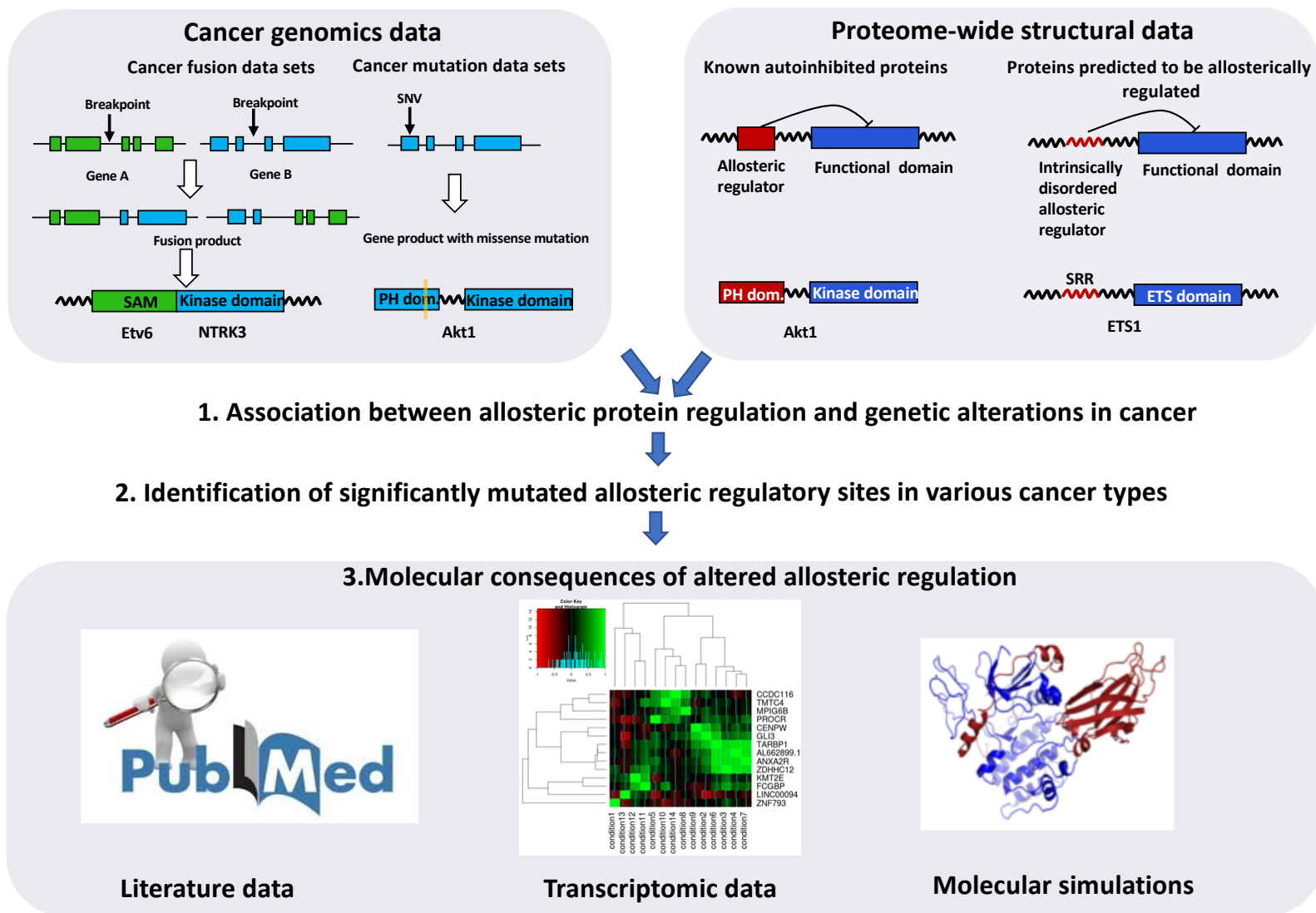
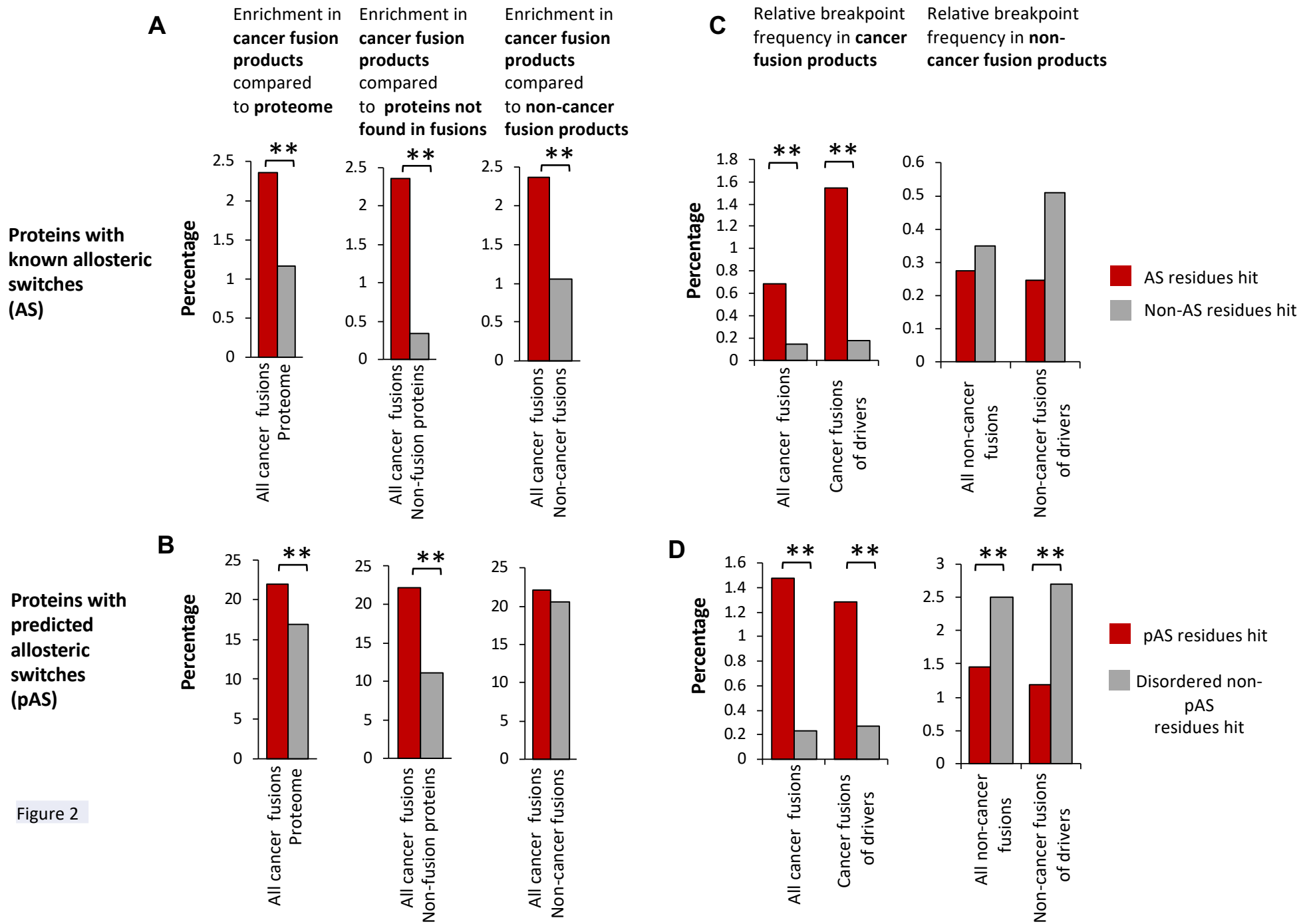
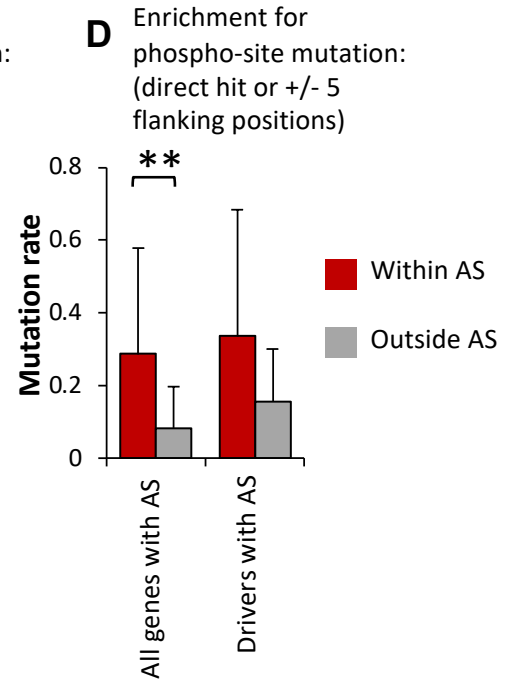
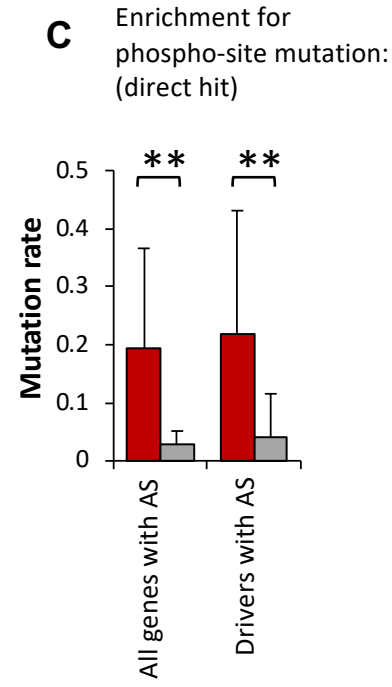
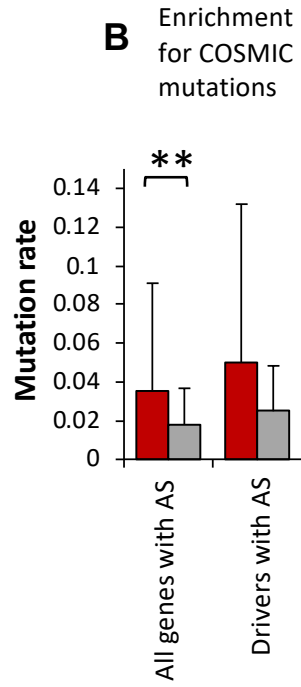
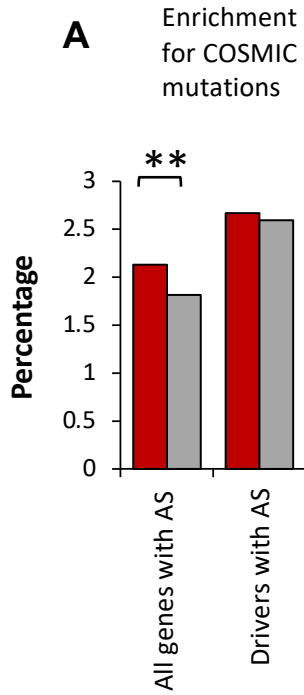


Figure 1



Proteins with known allosteric switches (AS)



Proteins with predicted allosteric switches (pAS)

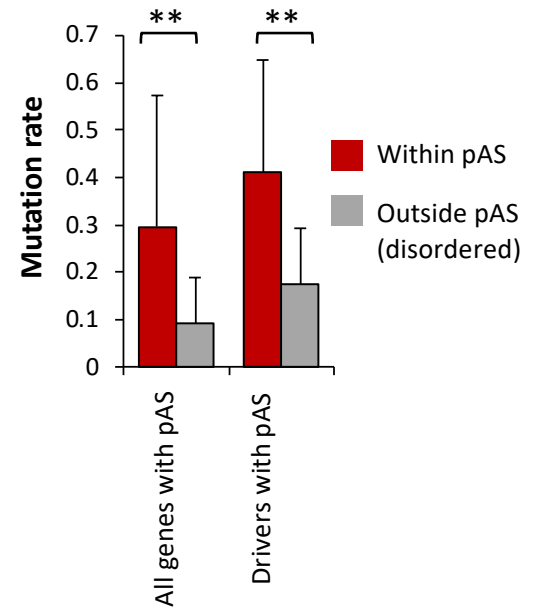
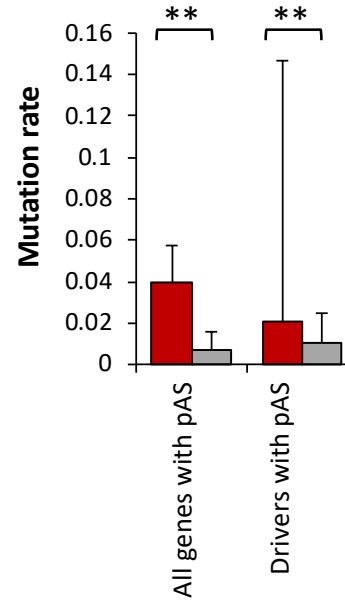
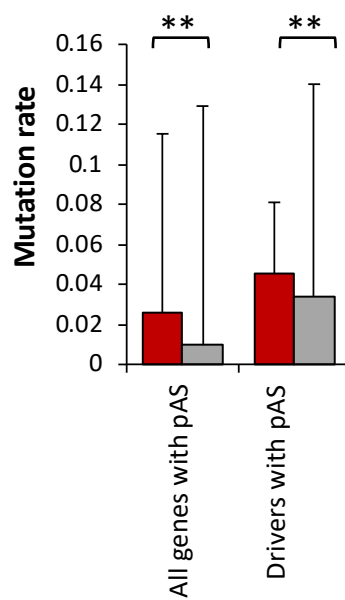
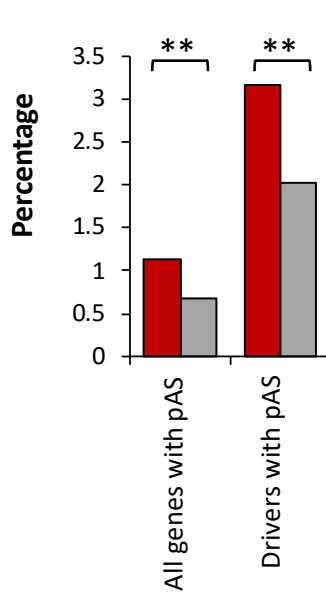


Figure 3

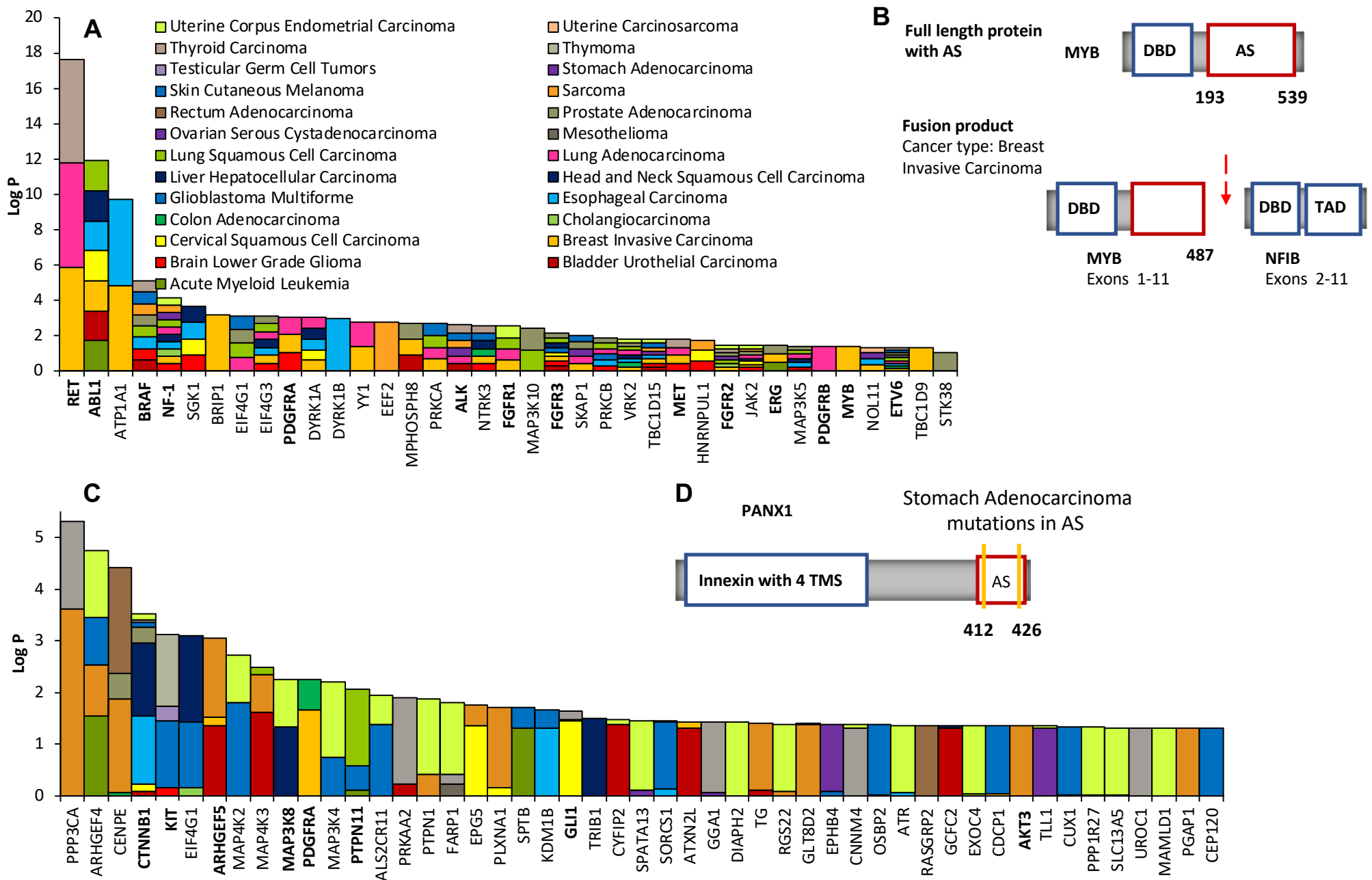


Figure 4

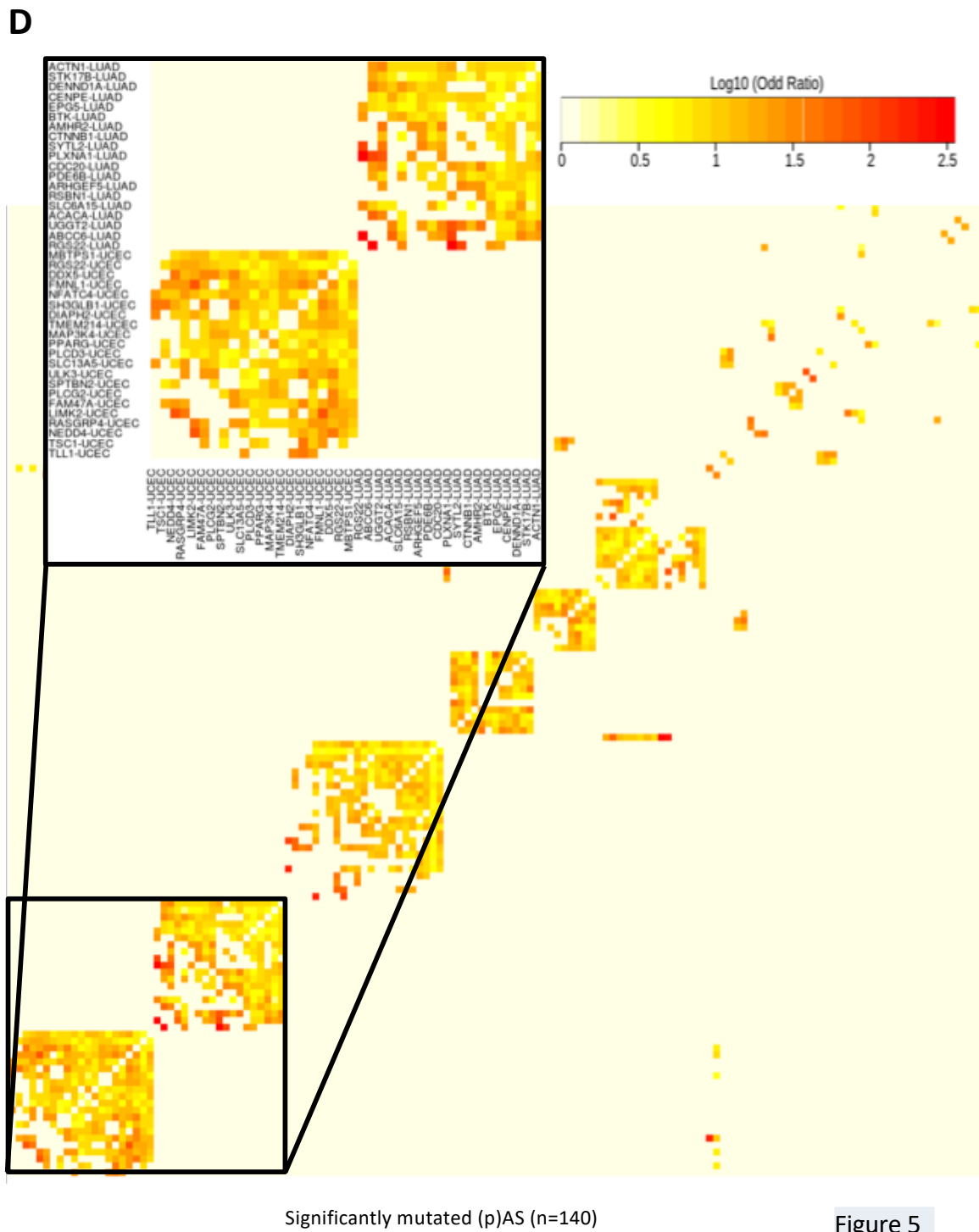
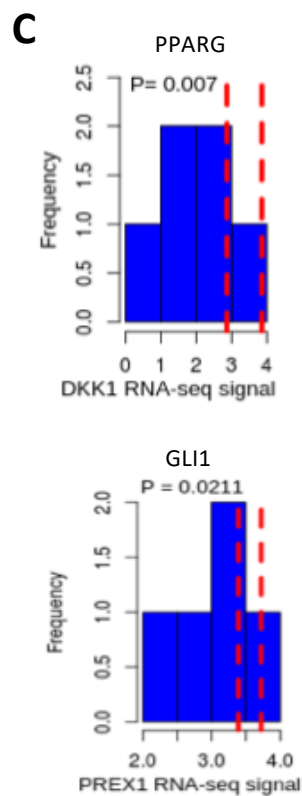
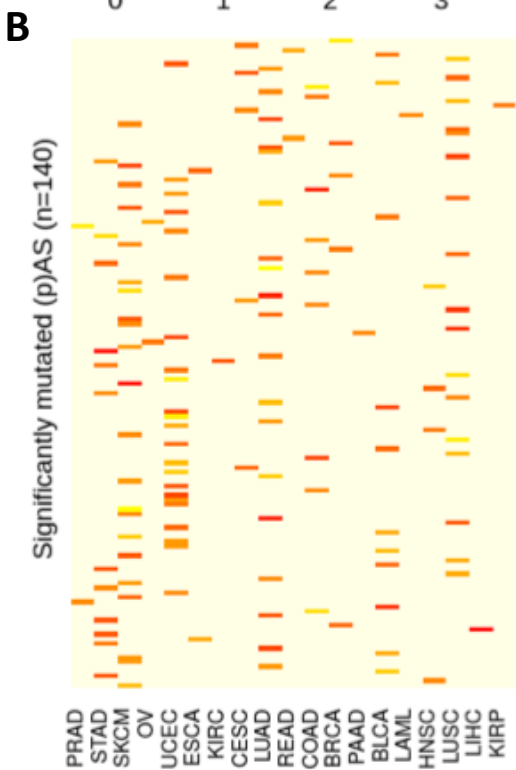
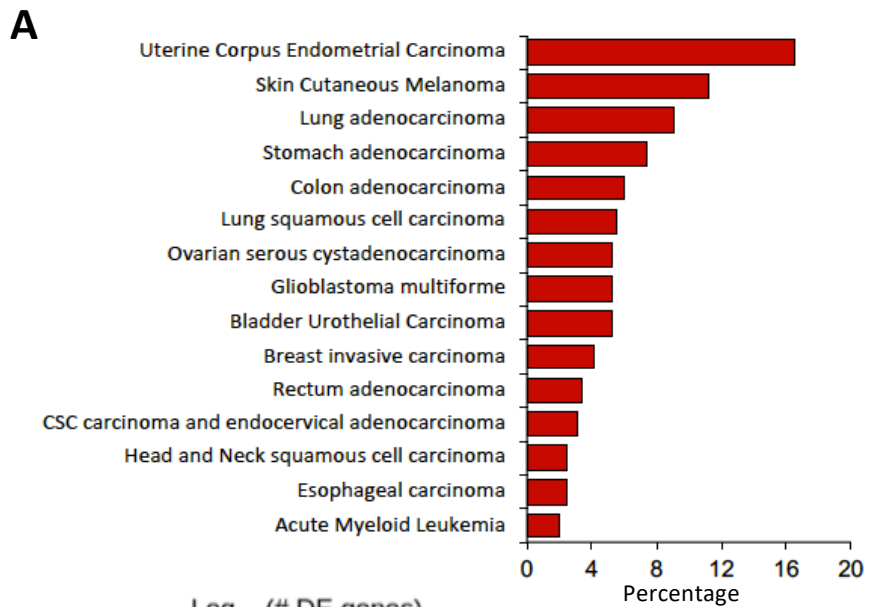


Figure 5

