



## OPEN

## The mutational landscape of phosphorylation signaling in cancer

## SUBJECT AREAS:

CANCER

GENETICS

SYSTEMS BIOLOGY

BIOINFORMATICS

Jüri Reimand, Omar Wagih &amp; Gary D. Bader

The Donnelly Centre, University of Toronto, Canada.

Received  
17 July 2013Accepted  
23 August 2013Published  
2 October 2013

Correspondence and requests for materials should be addressed to J.R. (Juri.Reimand@utoronto.ca) or G.D.B. (Gary.Bader@utoronto.ca)

Somatic mutations in cancer genomes include drivers that provide selective advantages to tumor cells and passengers present due to genome instability. Discovery of pan-cancer drivers will help characterize biological systems important in multiple cancers and lead to development of better therapies. Driver genes are most often identified by their recurrent mutations across tumor samples. However, some mutations are more important for protein function than others. Thus considering the location of mutations with respect to functional protein sites can predict their mechanisms of action and improve the sensitivity of driver gene detection. Protein phosphorylation is a post-translational modification central to cancer biology and treatment, and frequently altered by driver mutations. Here we used our ActiveDriver method to analyze known phosphorylation sites mutated by single nucleotide variants (SNVs) in The Cancer Genome Atlas Research Network (TCGA) pan-cancer dataset of 3,185 genomes and 12 cancer types. Phosphorylation-related SNVs (pSNVs) occur in ~90% of tumors, show increased conservation and functional mutation impact compared to other protein-coding mutations, and are enriched in cancer genes and pathways. Gene-centric analysis found 150 known and candidate cancer genes with significant pSNV recurrence. Using a novel computational method, we predict that 29% of these mutations directly abolish phosphorylation or modify kinase target sites to rewire signaling pathways. This analysis shows that incorporation of information about protein signaling sites will improve computational pipelines for variant function prediction.

Cancer is a set of diseases characterized by somatically acquired cellular alterations that lead to selective advantages such as unrestricted growth, suppression of apoptosis and enhanced metabolism<sup>1</sup>. The complexity of cancer is observed at multiple levels of cellular organization, as somatic alterations in chromosomal copy numbers, epigenetic regulation and gene expression give rise to tumor types and subtypes with different biological and clinical properties<sup>2–4</sup>. In particular, high-throughput sequencing has revealed a complex landscape of somatic DNA mutations in cancer genomes<sup>5</sup>. Most cancer mutations are likely passengers that appear due to genetic, epigenetic and transcriptional instability, while few mutations, termed drivers, unlock oncogenic cell properties that lead to selective advantages and tumor development<sup>1,6</sup>. Cancer drivers are often discovered due to high mutation frequency across many tumors of a particular type, however combinations of rare mutations in related systems or pathways may be also responsible for tumorigenesis<sup>7,60</sup>. The accumulation of sequencing data from cancer genome projects<sup>8–16</sup> now enables the discovery of driver mutations relevant across multiple tumor types. The characterization of these pan-cancer drivers is important for establishing efficient multi-cancer therapies such as the mutant BRAF inhibition strategy applicable in melanoma and leukemia<sup>17,18</sup>.

Cellular signaling networks are complex systems of interacting proteins that are ultimately encoded in the genome. Analysis of disease mutations using network context will thus lead to better understanding of their mechanisms of action<sup>61</sup>. Protein phosphorylation, a reversible post-translational modification (PTM) at serine (S), threonine (T) and tyrosine (Y) residues, involves a system of sequence-specific kinases (writers), phosphatases (erasers) and reader proteins. Phosphorylation signaling can modulate protein activity, alter protein folding, and help mediate or inhibit interactions with other proteins. Phosphorylation is important in cancer and is involved in the control of proliferation, oncogenic kinase signaling<sup>19</sup>, transcriptional regulation<sup>20</sup>, and TP53 activity<sup>21</sup>, among other processes. Phosphorylation is also a pharmacologically targetable mechanism with multiple approved therapies available for cancer treatment<sup>17,18</sup>. We recently proposed that cancer may be driven by statistically significant and spatially specific mutations in protein sites involved in cellular phosphorylation signaling, and developed the ActiveDriver method to detect such mutations comprehensively<sup>7</sup>. ActiveDriver is a gene-centric method that identifies signaling sites where the mutation rate is significantly higher than expected from the entire gene sequence, thus suggesting the site's importance in tumor biology.

The recently available pan-cancer dataset of 3,185 tumor genomes and 12 cancer types from The Cancer Genome Atlas (TCGA) comprises the largest collection of somatic cancer mutations to date<sup>65</sup>. It involves four



times more samples and 24 times more SNVs than previous collections<sup>7</sup>, providing the opportunity to discover novel cancer driver genes across multiple cancer types. Here, we analyze the TCGA pan-cancer dataset of protein-coding missense single nucleotide variants (SNVs), as SNVs are easiest to interpret as specific alterations of signaling sites and are more reliably detected and abundant than other types of genetic mutations. We predict known and novel signaling-specific cancer driver genes, create a high-confidence collection of cancer mutations likely involved in altered cellular signaling, and propose numerous specific hypotheses to explain their functional effects.

## Results

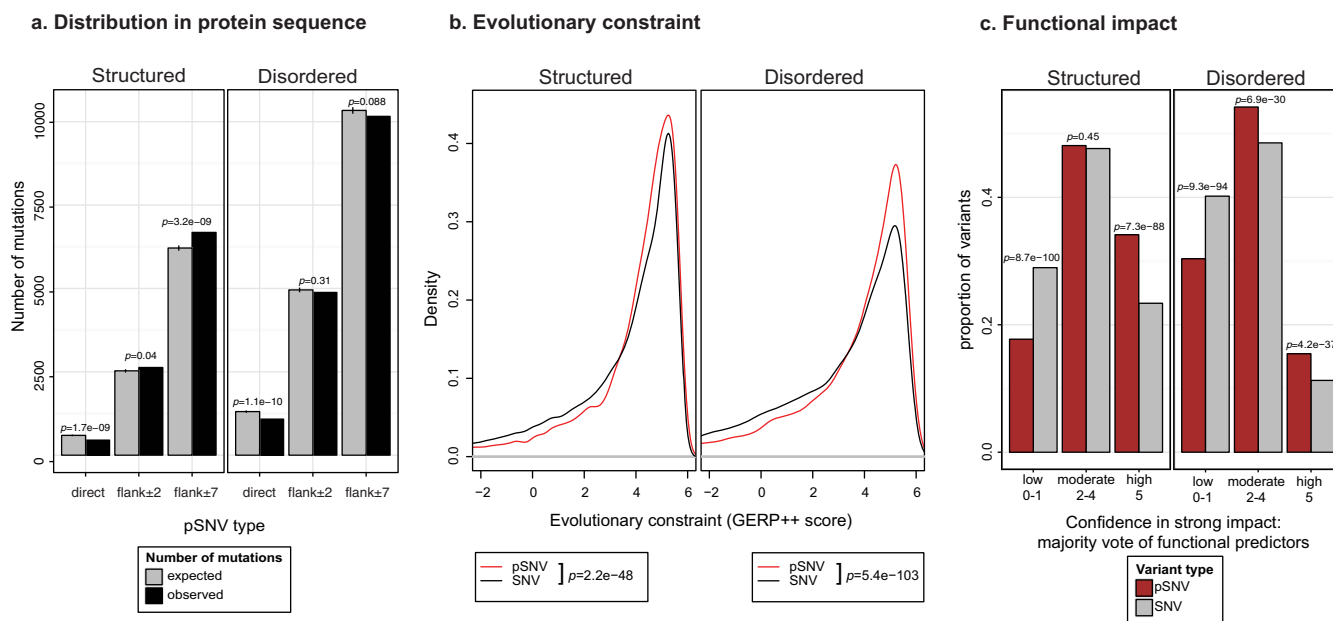
**Signaling sites are important in mutation function prediction.** To investigate cancer mutations in phosphorylation signaling, we collected 87,060 experimentally determined phosphosites in 10,185 human proteins and integrated these with 241,701 missense single nucleotide variants from the TCGA pan-cancer project (Supplementary Table 1). Including  $\pm 7$  residues of phosphosite flanking regions and covering 7% of protein sequence, we found 16,840 phosphorylation-related SNVs (pSNVs) in 5,859 genes and 89% of all samples – over 17 times more pSNVs than previously discovered<sup>7</sup> (Supplementary Tables 2–3). The observed large number of pSNVs helps us identify novel pan-cancer trends in phosphosite mutation distribution. We previously found that phosphosites are enriched in somatic mutations<sup>7</sup>. While this trend no longer holds with the larger pan-cancer dataset (Supplementary Fig. 1), we now identify a more specific signal – structured protein regions are enriched in pSNVs ( $p < 0.001$ , Poisson exact test), while disordered regions show no significant bias (Fig. 1a). pSNVs also show significantly stronger evolutionary constraint in 34 mammalian genomes<sup>22</sup> compared to other SNVs (Fig. 1b), suggesting that these signaling sites undergo negative selection. Further, we compared the predicted functional impact of mutations in a majority vote<sup>23</sup> of five state-of-the-art

methods (SIFT<sup>24</sup>, PolyPhen2<sup>25</sup>, LRT<sup>26</sup>, PhyloP<sup>27</sup>, MutationTaster<sup>28</sup>) and found that pSNVs are more likely disruptive to protein function than other protein-coding mutations (Fig. 1c). According to another measure of pSNV importance, 1,427 direct pSNVs replace the central phosphorylated residue and thus disrupt phosphorylation; such mutations are under-represented on the whole, although frequently seen in known cancer genes such as *TP53* and *CTNNB1* (79 cancer genes,  $p = 4.2e - 18$ , Fisher's exact test). In total, we predict a specific signaling mechanism for 29% of pSNVs (4,800) through either direct pSNVs or kinase network rewiring (see below). Collectively, these global observations suggest that pSNVs represent a subset of protein-coding missense mutations with greater functional impact and relevance in cancer compared to other SNVs.

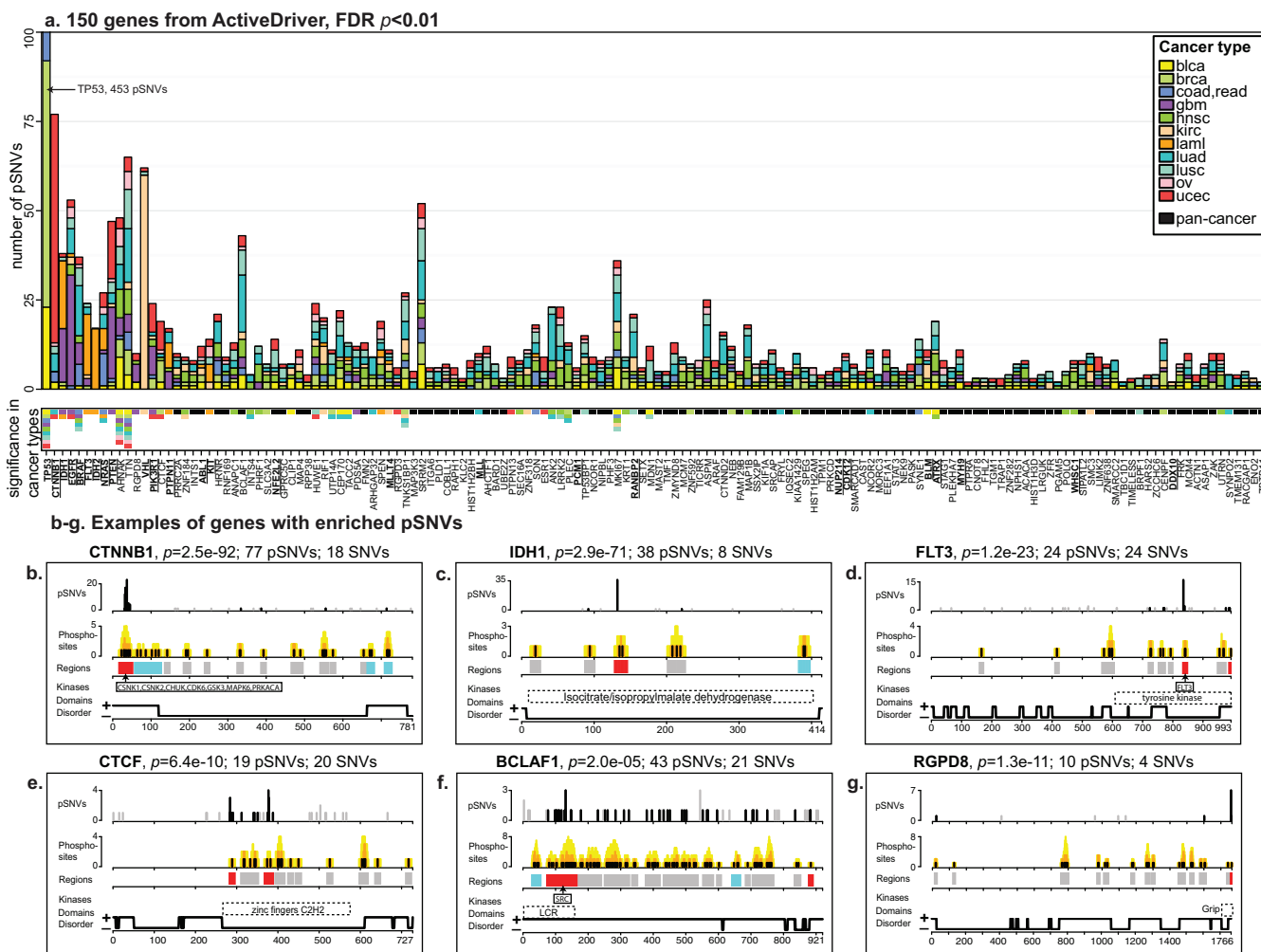
Interestingly, 40% of our mechanistic predictions (1,937 pSNVs) are considered neutral by most functional predictors, showing that incorporation of signaling information will greatly improve coverage of existing methods. We also observe that disordered regions comprising 37% of protein sequence carry the majority of phosphosites (71%) and thus represent an important but typically untreated factor in mutation analysis – one that ActiveDriver considers (Supplementary Fig. 2). While variants in disordered regions are generally less conserved than in structured regions, disordered pSNVs form a subset of higher evolutionary constraint and potential functional impact (Fig. 1b–c). In particular, 49% (977) of pSNVs in hallmark cancer genes occur in disordered regions. Thus, extending conservation-weighted functional mutation prediction methods to consider protein disorder will improve their coverage.

## ActiveDriver reveals pan-cancer driver genes with enriched pSNVs.

Next we used our ActiveDriver mutation significance model<sup>7</sup> to predict a list of cancer driver genes specifically enriched in pSNVs. Pan-cancer analysis revealed 150 genes (FDR  $p < 0.01$ ; Fig. 2a, Supplementary Table 4), many of which are known cancer genes ( $n = 26$ ,  $p = 6.1e - 15$ , Fisher's exact test). Most



**Figure 1 | Properties of phosphorylation-related single nucleotide variants (pSNVs) in structured and disordered protein sequence.** (a). Phosphosites in structured protein regions are significantly enriched in pSNVs (left panel), while disordered regions show no significant bias. Direct pSNVs altering phosphorylated residues are less common than expected. P-values are shown on top of bars (Poisson exact test). (b). Evolutionary constraint of pSNVs (red) is significantly greater than for other SNVs (grey) for structured as well as disordered protein sequence. Significance p-values are shown in the below plots (Wilcoxon test). (c). Prediction of variant functional impact with five methods shows that pSNVs (red) are more frequently considered harmful by multiple methods. Predictor majority vote established confidence in functional impact, as low confidence (0–1 methods predict strong impact), moderate confidence (2–4 methods) and high confidence (5 methods). Bar height reflects proportion of variants. Significance p-values are shown above bars (binomial test).



**Figure 2 | Genes with significant pSNVs in pan-cancer genomes.** (a). ActiveDriver predictions of pan-cancer driver genes ( $n = 150$ , FDR  $p < 0.01$ ). Top panel shows pSNV counts in pan-cancer genomes. Genes are ordered from left to right according to statistical significance, and colors indicate cancer types. Bottom panel shows summary of per-cancer pSNV analysis. Colored boxes show individual cancer types with pSNV enrichment, and black boxes represent genes found in pan-cancer data that were not significant in individual cancer types. Well-recognized cancer genes are in bold and underlined. (b–g). Examples of genes with enriched pSNVs. Six tracks of protein sequence information are shown, from top to bottom: (i) counts of pSNVs (black) and other SNVs (gray); (ii) location of phosphosites (black) and flanking regions (orange - within two amino acid positions, yellow - within 3–7 positions); (iii) ActiveDriver prediction of phosphosite regions as pSNV-enriched (red), pSNV-depleted (blue), and not significant (gray); (iv–v) kinases and protein domains associated to pSNV-enriched regions (LCR - low complexity region); (vi) regions corresponding to disordered (+) and ordered (–) protein sequence.

genes (86), including seven known cancer genes, are only identified in the combined dataset but not in any individual cancer type, emphasizing the utility of integrated mutation analysis. Analysis of pSNVs recovers established driver mutations, validating our results, and provides novel signaling-related insight into functional consequence of many SNVs.

ActiveDriver reveals many known and novel cancer genes with specific pSNVs. For instance, beta-catenin (*CTNNB1*) is a Wnt-activated oncogene in lung and liver cancer that is degraded in non-tumor cells via phosphorylation of its N-terminus<sup>20</sup>. ActiveDriver identifies N-terminal mutations in five cancer types as highly significant ( $n = 73$ , FDR  $p = 2.5e - 92$  from ActiveDriver; Fig. 2b), confirming constitutive activation of *CTNNB1* via disrupted phosphorylation as a prevalent driving mechanism of cancer. As another example, isocitrate dehydrogenase 1 (*IDH1*) is a metabolic enzyme with frequent R132H mutations in leukemia and glioblastoma that associate to altered DNA methylation<sup>29</sup>. ActiveDriver suggests a novel hypothesis that 36 pSNVs modify the phosphosite Y135 recently observed in multiple proteomics datasets (Ref. 30, unpublished data at www.phosphosite.org; FDR

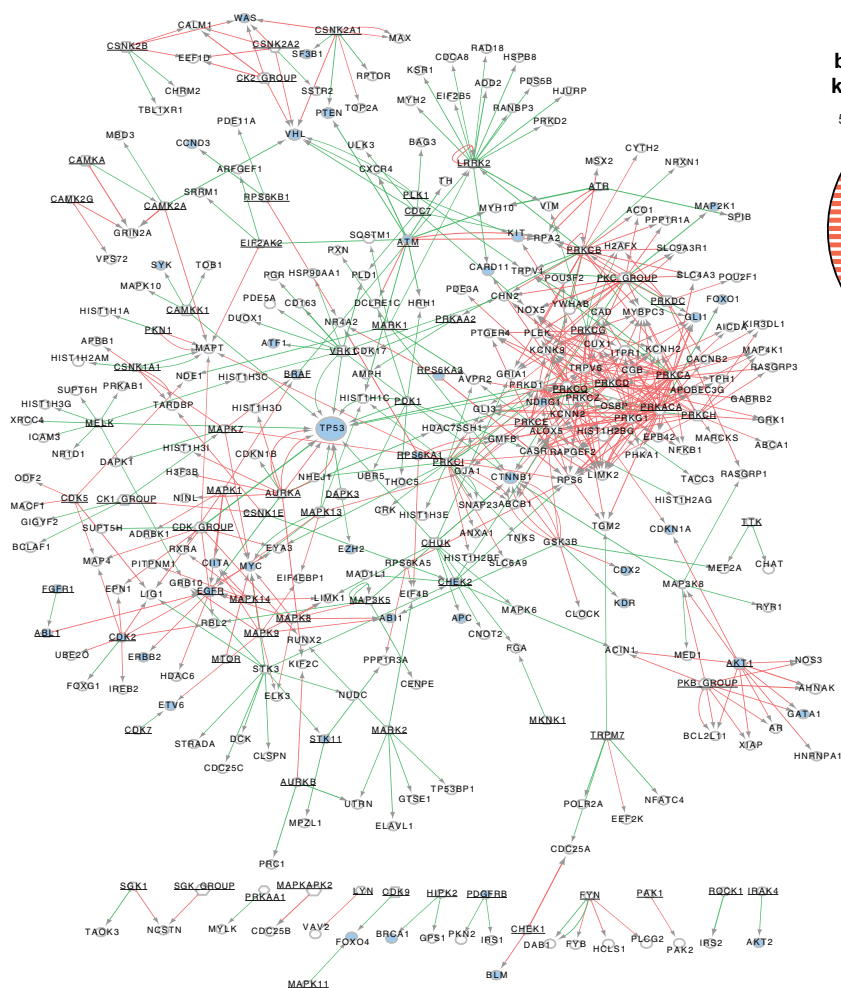
$p = 2.9e - 71$ ; Fig. 2c). Kinase proteins are enriched in the predicted list of cancer drivers ( $n = 14$ ,  $p = 1.0e - 07$ , Fisher's exact test), indicating that pSNVs also modify regulators in signaling networks. The tyrosine kinases *FLT3*<sup>19</sup> (Fig. 2d) and *KIT*<sup>31</sup> are the most prominent examples of significant activation loop pSNVs in leukemia that lead to increased kinase activity and acquired drug resistance, while we also observe similar mutations in other kinases, such as *HCK* and *CHEK2*. In addition to oncogenic kinases, pSNVs also modify transcription factors (TFs). For instance, the multifunctional TF and candidate tumor suppressor *CTCF* is differentially phosphorylated in its DNA-binding domains during cell cycle<sup>32,33</sup>. ActiveDriver highlights 19 pSNVs at these sites in six cancer types (FDR  $p = 6.4e - 10$ ; Fig. 2e), providing phospho-mechanistic insight into the earlier observation that cancer mutations in *CTCF* DNA-binding domains alter transcription of proliferative genes<sup>34</sup>. The candidate cancer gene *BCLAF1* (Bcl-2-associated transcription factor 1) is another phosphorylated TF with frequent pSNVs in sites of *SRC* onco-kinase (Fig. 2f). Finally, *RGPD8* is an example of a poorly characterized gene with recurrent and specific pSNVs of four cancer types in the Grip domain related to Golgi

signaling and protein secretion (Fig. 2g). In summary, analysis of phospho-mutations in cancer helps predict novel cancer driver genes and signaling-related mechanisms for known cancer genes.

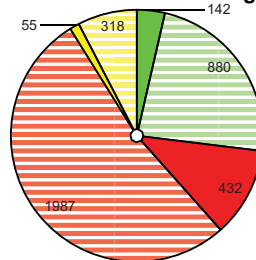
**Gain-of-signaling and loss-of-signaling pSNVs cause oncogenic rewiring of the kinase network.** As most pSNVs occur in phosphosite flanking sequence, we sought to characterize their impact on kinase binding specificity, which may affect site phosphorylation. We first developed a high-confidence set of sequence patterns recognized by 96 kinases modeled as position weight matrices (PWMs), based on known kinase binding sites (Supplementary Fig. 3). We next predicted which kinases recognize known phosphosites and how this changes after pSNV mutation. We identified 3,814 significant network-rewiring mutations – losses or gains in kinase-substrate signaling – that alter 11,802 potential kinase-substrate interactions and cover 23% of all pSNVs ( $p < 0.05$ ; Supplementary Table 5). A high-confidence network includes 392

pSNVs in 534 interactions, and comprises only top-scoring kinase binding sites for signaling gain and experimentally determined kinase-substrate signaling loss (Fig. 3a). Most inferred network-rewiring pSNVs confer loss-of-signaling, but many added interactions and change-of-signaling events are also predicted (Fig. 3b). Several examples of well-studied mutations are apparent in the rewiring analysis, which help validate our results. N-terminal phosphorylation of CTNNB1 at S37 involves the most frequent rewiring with 23 direct pSNVs and 23 flanking pSNVs (Fig. 3c) leading to predictions of gained signaling (with MAP3K8, PRKCE kinases) and lost signaling (TBK1). Notably, the gain-of-signaling mutation G34R may involve Wnt-MAPK signaling of CTNNB1 activation<sup>35</sup>, or the protein kinase C pathway that degrades CTNNB1 independently of Wnt<sup>36</sup>. As another example, phosphorylation of TP53 by AURKA kinase at S215 is known to inhibit TP53 tumor suppression function<sup>21</sup>. We predict that AURKA-TP53 signaling at that site is disrupted by six direct and seven flanking pSNVs (Fig. 3d),

### a. High-confidence network rewiring events



### b. Number of pSNVs in kinase network rewiring



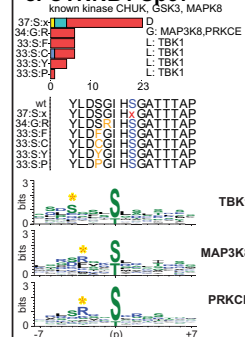
Rewiring events (total 3814)  
 Gained Lost Lost+gained  
 Kinase site Known Predicted

Interaction  
 Lost Gained  
 Protein  
 ● Substrate  
 ● Kinase  
 Annotation  
 ● Cancer gene  
 ○ Other

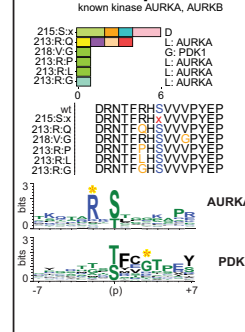
Cancer type  
 ● brca  
 ● brca  
 ● coad\_read  
 ● gbm  
 ● hnscc  
 ● kirc  
 ● laml  
 ● liuad  
 ● luad  
 ● ov  
 ● ucec

wildtype residue  
 central p-residue  
 direct mutation  
 flanking mutation  
 ★ rewiring mutation

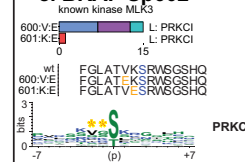
### c. CTNNB1 Sp37



### d. TP53 Sp215



### e. BRAF Sp602



**Figure 3 | pSNVs in kinase network rewiring.** (a). The high-confidence network of gain-of-signaling (green edges) and loss-of-signaling events (red edges) induced by pSNVs was predicted for 96 kinase binding models ( $p < 0.05$ ). Node size corresponds to number of network-rewiring pSNVs, and edge weight shows predicted impact of pSNV to kinase binding. Known cancer genes are shown as blue nodes and kinases are underlined. (b). Total number of pSNVs predicted to induce network rewiring (red – loss-of-signaling; green – gain-of-signaling; yellow – switch-of-signaling). pSNVs affecting known kinases (solid colors) and predicted kinases (striped colors) are shown separately. (c–e). Examples of phosphosites with network-rewiring pSNVs. Each panel contains three plots from top to bottom. (i) Bar chart shows pSNV counts in different cancer types and rewiring events for kinases (L – loss; G – gain; D – direct pSNV). Phosphosite and its experimentally determined kinases are shown on top. (ii) Block of 15-mer peptides shows pSNVs in phosphosite sequences (wildtype (wt) amino acid – black; central residue – blue; flanking pSNV – orange; direct pSNV – red). (iii) Predicted kinase binding models as position weight matrices (PWMs), with phosphosite sequence on X-axis ( $\pm 7$  residues, central residue indicated with ‘p’), and information content on Y-axis. Letter height corresponds to prevalence of amino acid at given position. Red asterisks in kinase binding models denote residues affected by network-rewiring pSNVs.

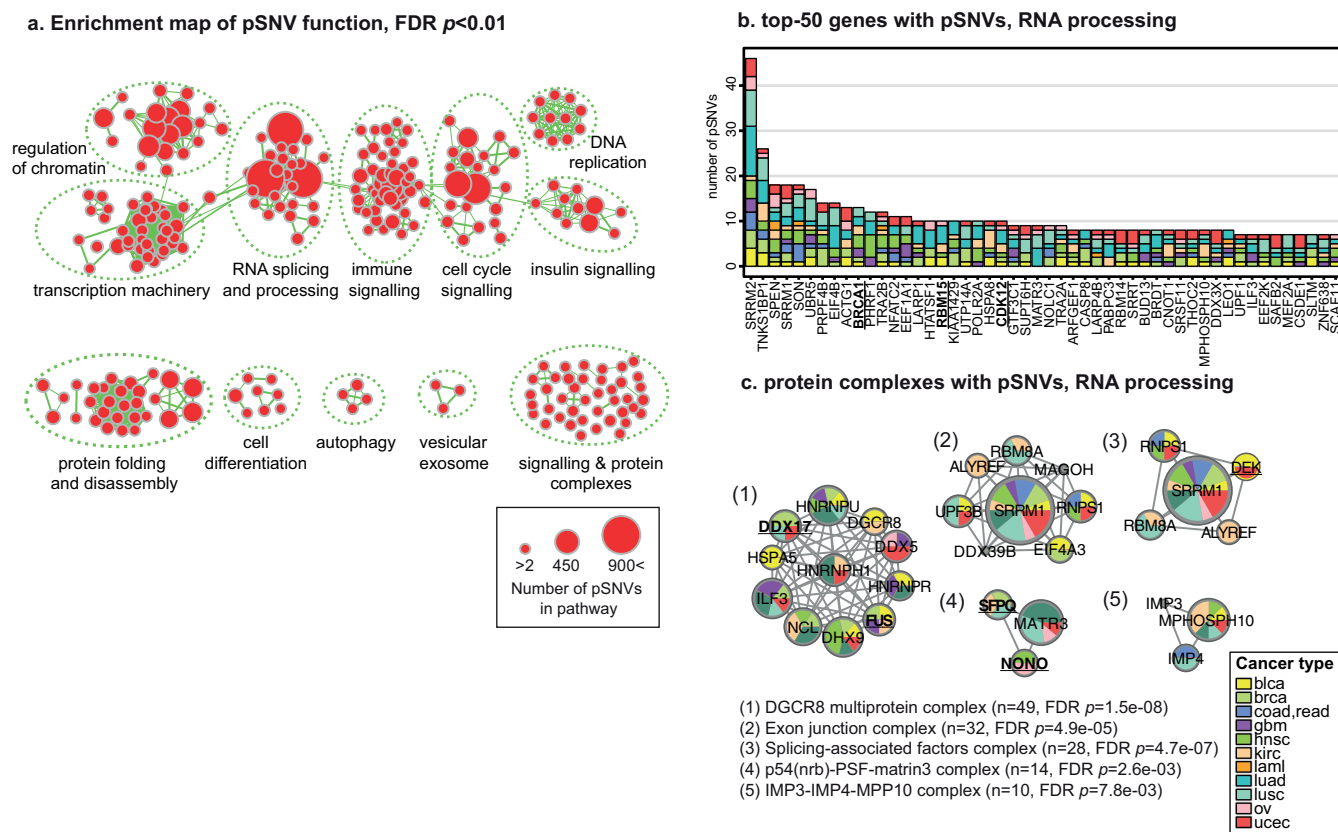


leading to increased activity of TP53. This potentially explains our earlier observation of improved survival of ovarian cancer patients with TP53 pSNVs<sup>7</sup> and encourages further investigation of the site as a prognostic biomarker. Next, our analysis predicts that the BRAF V600E kinase activation mutation<sup>37</sup> implicated in therapy of melanoma<sup>17</sup> and leukemia<sup>18</sup> interferes with phosphorylation by PRKCI at S602 (Fig. 3e). Phosphorylation of wildtype BRAF at S602 by MLK3 is known to activate cell proliferation pathways, however BRAF V600E tumor cells proliferate independently of MLK3<sup>38</sup>. We therefore speculate that BRAF signaling may involve PRKCI or another unrecognized kinase in a combinatorial or competitive mechanism with MLK3. Such kinase binding site analysis shows that flanking pSNVs may frequently lead to oncogenic rewiring of signaling networks, thus identifying specific mechanistic hypotheses about cancer-driving mutations.

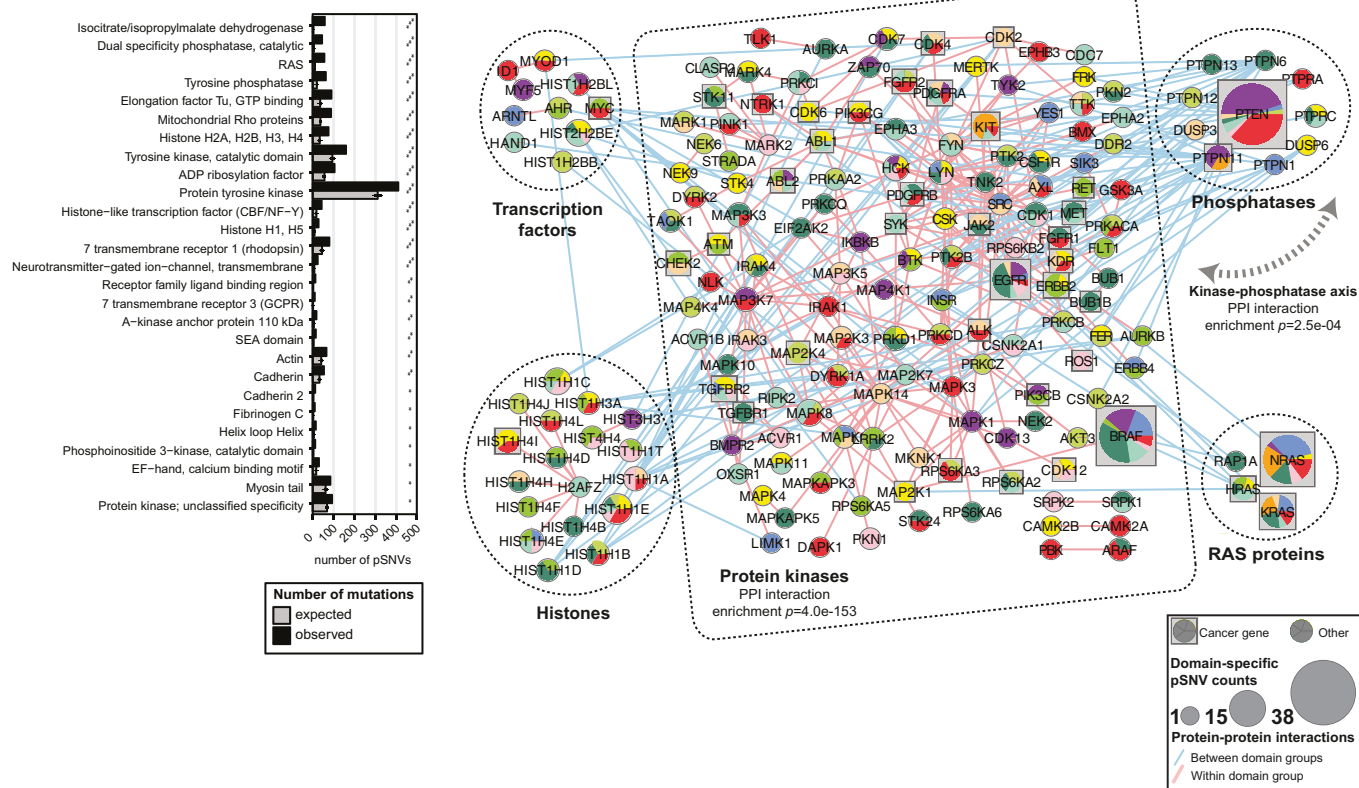
**Mutated signaling is central to hallmark cancer pathways and transcription regulatory networks.** Cancer is a disease of pathways driven by systematic alterations in hallmark processes<sup>1</sup>. Thus many infrequent but specific pSNVs may affect different components of underlying systems and lead to tumorigenic cell properties. To establish the significance of altered signaling at the pathway level, we conducted a functional enrichment analysis of somatic mutations. We focused on a stringently filtered subset of pathways specifically enriched in pSNVs and not in other SNVs, and revealed multiple phosphorylation-related functional themes with pan-cancer significance (FDR  $p < 0.01$ , Poisson exact test; Fig. 4a, Supplementary Fig. 4, Supplementary Table 6). The major functional themes are related to regulation of gene expression on epigenetic, transcriptional and post-transcriptional levels, cell cycle

and differentiation, and immune and insulin signaling. For example, the RNA splicing and processing theme involves 2,009 pSNVs in 652 genes (Fig. 4b), including 38 known cancer genes such as *BRCA1*, *RBM15*, *CDK12*, *CDC73* and *TOP1* (FDR  $p = 1.3e - 06$ , Fisher's exact test), as well as candidate cancer genes. For instance, the transcription factor SPEN, a negative regulator of Notch proliferative signaling pathway<sup>39</sup>, has 19 pSNVs in ten cancer types (FDR  $p = 2.2e - 03$  from ActiveDriver). Enrichment analysis also highlights pSNVs in protein complexes of RNA regulation (Fig. 4c), such as the DGCR8 multiprotein complex<sup>40</sup> of micro-RNA processing (49 pSNVs) and the exon junction complex<sup>41</sup> of mRNA splicing and post-transcriptional regulation (32 pSNVs).

As oncogenic signaling mutations appear in central gene regulatory processes, we hypothesized that pSNVs alter interactions and activity of regulatory protein domains. To investigate this in detail, we studied 5,662 protein domains for specific pSNV enrichment, and found 27 unique domains with frequent signaling mutations ( $\geq 25$  pSNVs, FDR  $p < 0.05$ , Poisson exact test, Fig. 5a, Supplementary Table 7). pSNVs are most abundant in kinase domains (446 pSNVs in 197 kinases), as kinases are regulated by phosphorylation in complex hierarchical networks<sup>7</sup>. In addition, pSNV enrichment is apparent in RAS, phosphatase, histone, and transcription regulatory domains. Interestingly, neurotransmitter genes such as GABA receptors were enriched in pSNVs in ligand-binding and transmembrane domains. Protein-protein interaction (PPI) network analysis of domain-specific pSNVs emphasizes the centrality of altered kinase signaling (Fig. 5b, Supplementary Table 8). In particular, the group of kinase proteins involves more within-group PPIs than expected ( $n = 253$ ,  $p = 4.0e - 153$ , Poisson exact test), and the phosphorylation reader-writer system of kinases and phosphatases is also highly



**Figure 4 | Pathways and processes with enriched pSNVs.** (a). Enrichment map of over-represented functional themes for pan-cancer pSNVs (FDR  $p < 0.01$ ). Functional enrichments also apparent in global SNV analysis have been filtered. Nodes represent pathways and processes, and edges show overlap of corresponding gene sets. Major functional themes are circled and annotated. (b). Top-50 RNA splicing and processing genes with frequent pSNVs. Colors denote cancer types. (c). Protein complexes with enriched pSNVs from the theme of RNA splicing and processing. Nodes correspond to complex subunits, and node size corresponds to pSNV count. Known cancer genes are in bold and underlined.


**a. Protein domains with enriched pSNVs**    **b. Protein-protein interactions and domain-specific pSNVs in the signaling-chromatin-transcription axis**


**Figure 5 | pSNVs in structural protein domains.** (a). SMART and Pfam protein domains with significant enrichment of phosphosite mutations (FDR  $p < 0.05$ ; domains with  $\geq 25$  pSNVs shown). Asterisks indicate level of statistical significance. (b). Protein-protein interaction (PPI) network of proteins with significant enrichment of domain-specific pSNVs. Nodes are grouped by domain type, as kinases (phospho-writers, center), phosphatases (phospho-erasers; top right), histones (chromatin regulation; bottom left), transcription factors (TFs, regulation of gene expression; top left), and RAS proteins (signaling, bottom right). Edge color denotes type of interaction (red – interaction within domain-specific group; blue – interaction between domain groups). Node size corresponds to number of domain-specific pSNVs, and node coloring reflects cancer types. Significant enrichment of PPIs is observed within the kinase group and in the kinase-phosphatase axis ( $p$ -values, Poisson exact test).

interacting ( $n = 52$ ,  $p = 2.5e - 04$ ). The domain-centric network includes 499 specific pSNVs in 183 proteins, including 42 known cancer genes ( $p = 1.5e - 28$ , Fisher's exact test). In summary, pathway and domain enrichment analysis shows that multiple regulatory systems and novel cancer hallmark processes are extensively altered by specific, potentially targetable mutations in phosphorylation signaling.

## Discussion

Our analysis demonstrates the extent of signaling-related mutations in a dozen important cancer types. A quarter of pSNVs are predicted to directly disrupt phosphorylation or involve kinase rewiring at the sequence level.

Our developed mechanistic hypotheses potentially depend on many factors of the cellular context. For instance, when observing a mutated signaling site, we assume that the site is actually phosphorylated in the tumor of study. While this information cannot be confirmed in general, we rely on the principle of evolutionary selection in cancer to gain confidence in our predictions. A major indicator of cancer driver mutations is their frequent, statistically significant recurrence in multiple tumor samples. The observation of repeated random events in precisely the same genomic locus is statistically so unlikely that we instead suggest that the site is under positive selection due to its role in tumor growth. The stronger the observed selection signal, the higher our confidence in calling the gene a cancer driver, regardless of our lack of knowledge about the context of the expressed protein in tumor cells. This is a standard and powerful approach in the cancer genomics field.

One limitation of the current analysis is that we restrict observed signaling alterations to protein-coding SNVs that comprise a minority of all cancer mutations known to affect protein function. We are thus underestimating the extent of mutated signaling in tumor cells caused by other mechanisms, for example genomic deletions such as the EGFRvIII isoform in glioblastoma, or translocations such as BCR-ABL in leukemia. However, the specific focus on protein-coding variants provides us with a high-confidence set of mutations that are simple to interpret in the context of cell signaling and lead to novel hypotheses that can be experimentally tested. In particular, tumor cell lines can be infected with mutated proteins using lentiviral technology, and phosphorylation-specific antibodies and mass spectrometry can be used to compare site-specific phosphorylation in mutated cell lines and controls. The functional consequence of such mutations can be characterised in phenotypic readouts such as growth assays or drug response screens. The impact of mutations on kinase binding specificity can be explored in vitro protein-binding microarrays, or in vivo, for example, by measuring kinase target phosphorylation after kinase mutation.

We previously performed ActiveDriver analysis on 800 cancer samples of eight cancer types<sup>7</sup>. Here we extend our analysis to 3,185 tumor genomes and 12 cancer types. As a result we predict 54 additional cancer-specific drivers and 82 genes only seen in pan-cancer analysis. In a related analysis, we combined these ActiveDriver results with predictions from MuSiC<sup>62</sup>, OncodriveFM<sup>63</sup>, and OncoDriveClust<sup>64</sup> to establish a high-confidence set of pan-cancer driver genes<sup>66</sup>.



The major goal of our analysis is to mechanistically explain the impact of phospho-mutations. Previously we could only specify that mutation of the central S, T, or Y amino acid to a non-STY residue would cause a loss of signaling. Here, we introduce a novel method to predict cancer-specific kinase rewiring events that considers how mutations in the phosphosite flanking region affect protein function. This approach increases the coverage of potential mechanistic explanations from 8% to 29% of all pSNVs. For our kinase rewiring analysis, we only cover 20% of human kinases and do not include information about kinase expression in a given tumor sample. Additional kinase specificity and tumor-specific protein expression data will improve the coverage and accuracy of these predictions.

In summary, this analysis leads to three major conclusions. First, incorporation of signaling information improves the sensitivity and coverage of variant function prediction and helps discovery of cancer driver mutations. Second, protein disorder is an important factor that functional mutation prediction methods should account for. Third, somatic alteration in phosphorylation signaling is a pan-cancer phenomenon, at least in the 12 cancer types studied here. As phosphorylation is a targetable mechanism with proven rational agents such as kinase inhibitors, consideration of pan-cancer pSNVs in therapy development may help find additional treatment strategies applicable to multiple cancer types.

## Methods

**Protein data.** Sequences for completed human refGene genes (hg19) were translated to proteins with the Annovar toolset for functional annotation of genetic variants<sup>42</sup> and filtered to retain the longest isoform for every gene. Pseudo-autosomal genes and genes with non-standard chromosomal annotations were discarded. Protein disorder was predicted with the DISOPRED2 toolset<sup>43</sup> using default parameters.

Experimentally validated and published phosphorylation sites in human proteins were retrieved from three databases (HPRD<sup>44</sup>, PhosphositePlus<sup>45</sup>, Phospho.ELM<sup>46</sup>). Phosphorylated sites with  $\pm 7$  residues were matched exactly to the longest isoforms of annotated proteins, allowing multiple matches per sequence. Sites with overlapping flanking sequences were merged into continuous regions. SMART<sup>46</sup> and Pfam<sup>47</sup> protein domains were predicted with the SMART website. Physical human protein-protein interactions (PPI) were retrieved from the BioGRID database<sup>48</sup> and auto-interactions were discarded. The list of 555 canonical cancer genes was compiled from earlier review papers<sup>6,49–51</sup> using the CancerGenes<sup>52</sup> and Cancer Gene Census<sup>6</sup> databases.

**Mutation data.** The cleaned and filtered pan-cancer mutation annotation file for 12 cancer types was retrieved from The Synapse database (www.synapse.org; ID syn1729383). Hypermutated samples (71) were discarded. DNA mutations were translated to protein sequence with Annovar<sup>42</sup>. Only missense single nucleotide variants (SNVs) were retained and other mutation types including stop codon mutations were discarded. Mutations were mapped to phosphosites with ActiveDriver<sup>7</sup>. pSNVs were classified as direct (mutation at central phosphorylated residue, S/T/Y), proximal flanking (mutation within 2 residues of the central residue), and distal flanking (mutation within 7 residues). The set of direct mutations was further filtered – 105 serine-threonine mutations (S > T, T > S) were not considered direct, as these are known to be equivalent in terms of kinase specificity. Phosphosite mutations and other supplementary data accompanying our analysis are stored in the Synapse database (www.synapse.org, ID syn2237931).

**Statistical analysis of pSNVs.** The distribution of SNVs with respect to phosphosites was assessed with the Poisson exact test by comparing the mutation frequencies (average number of SNVs per residue) in phosphosite-related and non-related sequence, using all proteins in the dataset. Robust expected range of pSNVs was sampled from the Poisson distribution (median  $\pm$  median absolute deviation). Distribution of phosphosites in disordered and structured regions was assessed similarly with background mutation rates estimated from disordered and structured sequence, respectively. GERP++ scores for evolutionary constraint of mutations in human and 33 other mammalian species<sup>22</sup>, as well as functional predictions from five methods (SIFT<sup>24</sup>, PolyPhen<sup>25</sup>, LRT<sup>25</sup>, PhyloP<sup>27</sup>, MutationTaster<sup>28</sup>) were retrieved from Annovar<sup>42</sup>. Evolutionary constraint of pSNVs and other SNVs was compared with the non-parametric Wilcoxon test as well as custom permutation tests ( $p < 1e-6$ ). Prediction of functional impact of mutations was carried out as a majority vote of the five methods, using the cutoff criteria as defined in the dbNSFP database of human non-synonymous SNPs<sup>23</sup>. Each mutation was scored by how many methods considered it harmful, and was binned as low (0–1), moderate (2–4) and high confidence (5), as previously proposed<sup>23</sup>. Statistical significance of pSNVs in bins was assessed with the binomial test using bin frequencies of other SNVs as background. Significantly phospho-mutated genes were computed with ActiveDriver<sup>7</sup> (false discovery rate (FDR)  $p < 0.01$ ). Two genes with significance only from depleted pSNVs were discarded (PIK3CA, KRAS).

**Kinase sequence binding models.** Kinase amino acid binding specificities were modeled as position weight matrices (PWMs) with 15 positions. PWM columns represent sequence positions, rows represent amino acids, and values represent probabilities of amino acids being present in a particular position of a kinase binding site. Initial PWMs were constructed from 15,659 sequence-specific kinase binding site annotations collected from HPRD, PhosphositePlus and Phospho.ELM. Kinase models with less than 20 binding sequences were discarded. PWMs were then refined using a strategy that discarded outlier sequences (cutoff  $p = 0.2$ ) from each kinase-specific sequence set and constructed the PWM iteratively until convergence (no further sequences below cutoff). PWMs were discarded if the refining procedure provided less than 20 sequences as result. Refined PWMs were assessed in tenfold cross-validation experiments in which 80% of known refined kinase binding sequences were used for PWM construction and the remaining 20% served as the positive test set. The negative test set for validation using target sequences from other kinases. Area Under Receiver Operator Curve (AUROC) statistics were used for PWM evaluation and low-quality PWMs were discarded (AUROC  $< 0.75$ ). The remaining high-confidence 96 PWMs (Supplementary Fig. 3) corresponding to 7,606 sequences were used to evaluate pSNVs impact on kinase binding to phosphosites.

**Analysis of kinase network rewiring.** To assess the similarity between a given sequence and a PWM, we adopted the Matrix Similarity Score (MSS) of the MATCH DNA binding model<sup>53</sup> to amino acids. We excluded the central S/T/Y residue (column 8 in PWM) from scoring to focus on flanking sequences. To establish the statistical significance of MSS values for a particular PWM, we computed an empirical background distribution of MSS scores. The background comprised all phosphosites with known kinases that were not associated to the kinase of the particular PWM. The significance  $p$ -value of a given MSS score was determined as the fraction of background sequences with equal or greater MSS. Impact of mutations on kinase binding was assessed using empirical  $p$ -values. A pSNV was considered to lead to loss-of-signaling of a kinase if a given reference phosphosite had a significant  $p$ -value of MSS ( $pRef < 0.05$ ) and the corresponding mutated phosphosite had a non-significant MSS  $p$ -value ( $pMut > 0.1$ ) for the PWM of that kinase. Gain-of-signaling pSNVs were predicted similarly ( $pRef > 0.1$  and  $pMut < 0.05$  for a PWM). A pSNV was considered to rewire a binding site (lead to switch-of-signaling) if both gain and loss events with different PWMs were predicted for the same pSNV.

**Protein domain analysis.** 5,662 protein domains from SMART and Pfam databases were matched to human protein sequences. Infrequent domains ( $< 3$  proteins) and infrequently mutated domains ( $< 3$  proteins with domain-specific pSNVs) were filtered. The remaining 525 domains were tested for pSNV enrichment with one-tailed Poisson exact tests, by comparing domain-specific pSNV counts in domain-associated sequence with expected mutation rate of all phosphorylation-associated protein sequence. Results were filtered for significance (FDR  $p < 0.05$ ) and further filtered manually for redundancy. The domain network comprises protein-protein interactions (PPI) between proteins with significantly phospho-mutated domains. Proteins were grouped by domain type (kinases - TyrKc, STYKc, PI3Kc, Pkinase\_Tyr; phosphatases - Pfam:DSPc, Pfam:Y\_phosphatase; histones - Pfam:Histone, Pfam:Linker\_histone; transcription factors - Pfam:HLH, Pfam:CBFD\_NFYB\_HMF) and assessed for enrichment of PPI within groups and between groups using the one-tailed Poisson exact test. The test considered number of interactions for a list of proteins (e.g. interactions within a list of phospho-mutated proteins with kinase domains), given the average number of interactions for all proteins in our set.

**Pathway enrichment analysis.** Functional gene lists representing pathways and processes were retrieved from the g:Profiler web server<sup>54</sup> and filtered to exclude large lists ( $> 1,000$  genes) and small lists ( $< 3$  genes). Gene Ontology<sup>55</sup> terms, Reactome<sup>56</sup> pathways and protein complexes from the CORUM database<sup>57</sup> were used for enrichment tests and other data sources were discarded. Gene lists with mutations in only single genes or samples were also discarded. Pathway enrichment analysis of mutations in 14,135 processes, complexes and pathways was carried out with one-tailed Poisson tests. Separate pathway analyses were carried out for SNVs and pSNVs, and both mutation groups were tested for cancer type-specific and pan-cancer pathway enrichments. Multiple testing corrections with FDR were applied separately for each data source (GO, Reactome, CORUM) mutation type (pSNV, SNV), and cancer type (12 cancer types plus pan-cancer dataset). The tests compared number of mutations within a functional gene list, given the average number of mutations across all genes (SNV analysis) or across genes with phosphosites (pSNV analysis). The final list of significantly mutated phospho-specific pan-cancer pathways included results from the pan-cancer analysis (FDR  $p < 0.01$ ) that were not significant in the SNV analysis of individual cancer types (FDR  $p > 0.1$ ) or the SNV analysis of the pan-cancer dataset (FDR  $p > 0.1$ ). Processes and pathways were visualized as Enrichment Maps<sup>58</sup> with the Cytoscape software<sup>59</sup>.

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674, doi:10.1016/j.cell.2011.02.013 (2011).
2. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488**, 49–56, doi:10.1038/nature11327 (2012).
3. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692, doi:10.1016/j.cell.2007.01.029 (2007).



4. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352, doi:10.1038/nature10983 (2012).
5. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558, doi:10.1126/science.1235122 (2013).
6. Futreal, P. A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177–183, doi:10.1038/nrc1299 (2004).
7. 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).
8. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068, doi:10.1038/nature07385 (2008).
9. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615, doi:10.1038/nature10166 (2011).
10. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70, doi:10.1038/nature11412 (2012).
11. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525, doi:10.1038/nature11404 (2012).
12. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337, doi:10.1038/nature11252 (2012).
13. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059–2074, doi:10.1056/NEJMoa1301689 (2013).
14. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73, doi:10.1038/nature12113 (2013).
15. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075, doi:10.1038/nature07423 (2008).
16. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160, doi:10.1126/science.1208130 (2011).
17. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med* **364**, 2507–2516, doi:10.1056/NEJMoa1103782 (2011).
18. Tiacci, E. *et al.* BRAF mutations in hairy-cell leukemia. *N Engl J Med* **364**, 2305–2315, doi:10.1056/NEJMoa1014209 (2011).
19. Smith, C. C. *et al.* Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature* **485**, 260–263, doi:10.1038/nature11016 (2012).
20. Morin, P. J. *et al.* Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787–1790 (1997).
21. Liu, Q. *et al.* Aurora-A abrogation of p53 DNA binding and transactivation activity by phosphorylation of serine 215. *J Biol Chem* **279**, 52175–52182, doi:10.1074/jbc.M406802200 (2004).
22. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025, doi:10.1371/journal.pcbi.1001025 (2010).
23. Liu, X., Jian, X. & Boerwinkle, E. dbSNFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* **32**, 894–899, doi:10.1002/humu.21517 (2011).
24. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073–1081, doi:10.1038/nprot.2009.86 (2009).
25. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–249, doi:10.1038/nmeth0410-248 (2010).
26. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res* **19**, 1553–1561, doi:10.1101/gr.092619.109 (2009).
27. Siepel, A., Pollard, K. S. & Haussler, D. In *Proceedings of the 10th annual international conference on Research in Computational Molecular Biology* 190–205 (Springer-Verlag, Venice, Italy, 2006).
28. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575–576, doi:10.1038/nmeth0810-575 (2010).
29. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478, doi:10.1038/nature10860 (2012).
30. Gu, T. L. *et al.* Survey of tyrosine kinase signaling reveals ROS kinase fusions in human cholangiocarcinoma. *PLoS One* **6**, e15640, doi:10.1371/journal.pone.0015640 (2011).
31. DiNitto, J. P. *et al.* Function of activation loop tyrosine phosphorylation in the mechanism of c-Kit auto-activation and its implication in sunitinib resistance. *J Biochem* **147**, 601–609, doi:10.1093/jb/mvq015 (2010).
32. Dephousse, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* **105**, 10762–10767, doi:10.1073/pnas.0805139105 (2008).
33. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3, doi:10.1126/scisignal.2000475 (2010).
34. Filippova, G. N. *et al.* Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res* **62**, 48–52 (2002).
35. Bikkavilli, R. K., Feigin, M. E. & Malbon, C. C. p38 mitogen-activated protein kinase regulates canonical Wnt-beta-catenin signaling by inactivation of GSK3beta. *J Cell Sci* **121**, 3598–3607, doi:10.1242/jcs.032854 (2008).
36. Gwak, J. *et al.* Protein-kinase-C-mediated beta-catenin phosphorylation negatively regulates the Wnt/beta-catenin pathway. *J Cell Sci* **119**, 4702–4709, doi:10.1242/jcs.03256 (2006).
37. Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
38. Chadee, D. N. & Kyriakis, J. M. MLK3 is required for mitogen activation of B-Raf, ERK and cell proliferation. *Nat Cell Biol* **6**, 770–776, doi:10.1038/ncb1152 (2004).
39. Oswald, F. *et al.* SHARP is a novel component of the Notch/RBP-1/Jkappa signalling pathway. *EMBO J* **21**, 5417–5426 (2002).
40. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Exp Cell Res* **313**, 4196–4207, doi:10.1016/j.yexcr.2007.07.020 (2007).
41. Lejeune, F., Ishigaki, Y., Li, X. & Maquat, L. E. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J* **21**, 3536–3545, doi:10.1093/emboj/cdf345 (2002).
42. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).
43. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139, doi:10.1093/bioinformatics/bth195 (2004).
44. Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Res* **37**, D767–772, doi:10.1093/nar/gkn892 (2009).
45. 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).
46. Dinkel, H. *et al.* Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res* **39**, D261–267, doi:10.1093/nar/gkq1104 (2011).
47. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290–301, doi:10.1093/nar/gkr1065 (2012).
48. Chattri-Aryamontri, A. *et al.* The BioGRID interaction database: 2013 update. *Nucleic Acids Res* **41**, D816–823, doi:10.1093/nar/gks1158 (2013).
49. Hahn, W. C. & Weinberg, R. A. Modelling the molecular circuitry of cancer. *Nat Rev Cancer* **2**, 331–341, doi:10.1038/nrc795 (2002).
50. Mitelman, F. Recurrent chromosome aberrations in cancer. *Mutat Res* **462**, 247–253 (2000).
51. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat Med* **10**, 789–799, doi:10.1038/nm1087 (2004).
52. Higgins, M. E., Claremont, M., Major, J. E., Sander, C. & Lash, A. E. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* **35**, D721–726, doi:10.1093/nar/gkl811 (2007).
53. Kel, A. E. *et al.* MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31**, 3576–3579 (2003).
54. Reimand, J., Arak, T. & Vilo, J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* **39**, W307–315 (2011).
55. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29, doi:10.1038/75556 (2000).
56. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–622 (2009).
57. Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes--2009. *Nucleic Acids Res* **38**, D497–501, doi:10.1093/nar/gkp914 (2010).
58. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984, doi:10.1371/journal.pone.0013984 (2010).
59. Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**, 2366–2382, doi:10.1038/nprot.2007.324 (2007).
60. Gonzalez-Perez, A. *et al.* Computational approaches to identify functional genetic variants in cancer genomes. *Nat. Methods.* **10**, 723–9. doi: 10.1038/nmeth.2562. (2013).
61. Reimand, J. *et al.* Domain-mediated protein interaction prediction: From genome to network. *FEBS Lett.* **586**, 2751–63. doi: 10.1016/j.febslet.2012.04.027 (2012).
62. Dees, N. D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–98, doi: 10.1101/gr.134635.111. (2012).
63. Gonzalez-Perez, A. *et al.* Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169, doi: 10.1093/nar/gks743. (2012).
64. Tamborero, D. *et al.* OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics.* **15**, 2238–44, doi: 10.1093/bioinformatics/btt395 (2013).
65. Stuart, M. J. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
66. Tamborero, D. *et al.* Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650; DOI: 10.1038/srep02650 (2013).

## Acknowledgements

We would like to thank Cyriac Kandoth, Michael D. McLellan and Li Ding for their work on TCGA pan-cancer data analysis. We are grateful to Nuria Lopez-Bigas, Abel Gonzalez-Perez, Mona Meyer and Liis Uusküla-Reimand for their beneficial comments and feedback on the manuscript. This work was supported by the Canadian Institutes of Health Research (CIHR) grant MOP-84324 and the National Resource for Network Biology (NRNB) under award numbers P41 RR031228 and GM103504. We gratefully acknowledge





the contributions from the TCGA Research Network and its TCGA Pan-Cancer Analysis Working Group (contributing consortium members are listed in **Supplementary Note 1**).

### Author contributions

J.R. designed the study, performed data analysis and interpretation, prepared figures and wrote the manuscript. O.W. led analysis of kinase network rewiring, and designed and implemented the related method. G.D.B. supervised the study and wrote the manuscript. All authors helped finalize the manuscript and approved the final version.

### Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Reimand, J., Wagih, O. & Bader, G.D. The mutational landscape of phosphorylation signaling in cancer. *Sci. Rep.* 3, 2651; DOI:10.1038/srep02651 (2013).



This work is licensed under a Creative Commons Attribution 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0>