A Pan-Cancer assessment of alterations of ULK1 kinase, an upstream regulator of autophagy

Mukesh Kumar[1], Elena Papaleo[1,2]

1 Computational Biology Laboratory, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen

2 Translational Disease System Biology, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

Corresponding author: elenap@cancer.dk

Keywords: molecular dynamics, protein structure networks, kinase, TCGA, missense mutations, protein stability

ABSTRACT

Autophagy is a key clearance process for the cell to recycle damaged cellular components. Autophagy is also well known for its dual role in cancer, acting as both tumor suppressor or promoter in a context-dependent way. One important upstream regulator of autophagy is the kinase ULK1. Several structures of the kinase domain of ULK1 have been solved, but a comprehensive study, including molecular dynamics, is currently missing. Also, an exhaustive description of the landscape of ULK1 cancer-related alterations is presently lacking. We here exploited the large amount of data on more than 30 cancer types through The Cancer Genome Atlas and the Recount2 initiatives to identify and analyze the atlas of ULK1 alterations in terms of gene expression and mutations. Moreover, we predicted the effects of mutations on ULK1 function and structural stability using a computational workflow accounting for protein dynamics and different layers of changes that a mutation can induce in a protein.

We discovered that ULK1, along with ULK2 are down-regulated in gynecological tumors, suggesting an impairment of the upstream regulation of autophagy. In other cancer types, ULK2 can compensate for ULK1 downregulation and, in the majority of the cases, no marked changes in expression level have been found for both genes. We identified 36 missense mutations of ULK1 especially in uterine, colon, stomach and lung cancer that were co-occurring with mutations in a large number of ULK1 interactors, suggesting a pronounced effect of the upstream steps of autophagy in these cancer types. Moreover, our study allowed to pinpoint that more than 50% of the mutations of ULK1 found in the cancer samples affect protein stability, which is likely to affect the cellular level and turnover of the protein. Among the damaging

mutations for stability, A125T, F273V, L215P, F14L, and G12D are also predicted not to alter the functional state of the protein. Three mutations (S184F, D102N, and A28V) are predicted with the only impact on kinase activity, either modifying the functional dynamics of the protein or the capability to exert effects from distal site to the functional and catalytic regions. The framework here applied could be extended more broadly to other targets to help in the classification of mutational effects in disease, to prioritize variants for experimental validation or select the appropriate biological readouts for experiments.

INTRODUCTION

Autophagy is a highly conserved catabolic mechanism across eukaryotes to degrade different cellular components and molecules, including organelles, proteins and bacteria [1,2]. Autophagy initiates with the formation of the autophagosome, which later fuses to the lysosome, resulting in the degradation of the cargo and the release of cellular building blocks [3,4] .

At a basal level, autophagy contributes to maintain cellular homeostasis and the recycling of cellular components. Autophagy can also be induced as a response to different cellular stresses with a cytoprotective role. Defects in the autophagy machinery are often linked to diseases, including cancer, neurodegeneration and bacterial infections [5].

Autophagy is initiated through the ULK1 (Unc-51 like autophagy activating kinase 1) complex [6–8], which consists of the ULK1 protein kinase, the FIP200 (FAK family kinase interacting protein of 200 kDA) scaffold protein, and the HORMA (Hop/Rev7/Mad2) containing proteins ATG13 and ATG101 [9]. The ULK1 complex can integrate different signals to promote both bulk and selective autophagy [9]. A highly coordinated and conserved cascade of post-translational events, including phosphorylation and ubiquitination, of the ULK1 kinase complex acts as major switches for autophagy initiation [10–13]. In the majority of the cases, the autophagy initiation is regulated by the interplay between mTOR (mammalian Target of Rapamycin) and AMPK (AMP-activated protein kinase) kinase complexes, which perform a series of inhibitory or activatory phosphorylations of the ULK1 kinase complex in different physiological conditions [10,14–18]. Upon activation ULK1 directly phosphorylate the PI3K kinase complex including BECLIN-1, VPS34, and ATG14, which facilitate nucleation of phagophore membrane at the phagophore assembly sites [19–21]. ULK1 also phosphorylated AMBRA1, which is a positive regulator of the PI3K complex [21].

Mammals have five ULK1 homologs with ULK1 and ULK2 featuring the highest similarities and also functional redundancy [7], implying that both need to be inactivated for a substantial inhibition of autophagy [22].

In mitophagy, ULK1 can be recruited to the mitochondria by different autophagy adaptors, such as optineurin, NDP52 and FUNDC1 [23,24]. FUNDC1 is also phosphorylated by ULK1 [24]. By analogy to the yeast counterpart, ULK1 has been

suggested to play essential scaffolding roles for autophagosome formation and maturation [9].

ULK1 has been reported overexpressed or downregulated in different cancer types or subtypes [25–28]. Autophagy in general has a strong association with cancer and can act with a dual role in a context-dependent way, being both tumor suppressor or promoter [29,30]. AMPK-ULK1 mediated autophagy also induces resistance against bromodomain and extraterminal domain inhibitors, which are novel epigenetic therapeutics for acute myeloid leukemia [31].So it can be a pivotal target to curb the chemotherapeutic resistance in cancers.

ULK1 is a multi-domain serine-threonine kinase, enriched in disordered regions [32], which is a common trait of many scaffolding proteins. ULK1 has preference for serine as the phospho-acceptor residue in the substrate and for hydrophobic residues surrounding the phosphorylation site [21]. The kinase domain is located at the N-terminal region of the protein and conserved among yeast and mammals [9].

Several inhibitors of ULK1 have been used to study its function in autophagy [21,33–35]. These molecules also have potential to be used in cancer therapy since cancer 'addiction' to autophagy has been reported [36]. ULK1 is the only serine/threonine kinase of the autophagy pathway and it is involved in the first biochemical steps of autophagy, thus representing an amenable druggable target in the pathway. X-ray structures of ULK1 in complex with the inhibitors are available (PDB entries: 4WNO [34], 4WNP [34], 5CI7 [35], 6QAS [37], and 6MNH [38]).

The N-terminal kinase domain of ULK1 includes the activation loops (165-174 and 178-191) and the catalytic loop (136-145), which are required for its function. ULK1 also includes a proline/serine-rich region (279-828) and a cytoplasmic domain (829-1-51), which are involved in interactions with ATG13, FIP200 and other upstream regulators such as mTOR and AMPK [15,39]. On the contrary, the N-terminal kinase domain of ULK1 could interact with the LRKK2 protein [40]. An important activatory autophosphorylation site is located in the activation loop at T180 [9] which, upon phosphorylation can engage in salt bridges with the neighbouring arginine residues R137 and R170 [34].

Despite the large number of X-ray structures of the ULK1 kinase domain, no extensive molecular dynamics (MD) simulation studies of the protein have been undertaken. The availability of a conformational ensemble of a protein is of paramount importance to better understand its function and the effects of its alterations due to mutations [41–44]. We thus used all-atom and coarse-grain models to obtain an ensemble of conformations and account for ULK1 flexibility and dynamics. We then applied methods inspired by network theory to achieve a Protein Structure Network representation [45,46] of conformational ensemble of the protein, which can help in the identification of residues important for structural stability and function, along with to pinpoint possible and elusive effects triggered by distal sites with respect to the functional residues of the protein [47–49].

We combined these structural studies with a curation of different layers of alteration of ULK1 in more than 30 cancer studies available in The Cancer Genome Atlas (TCGA) [50], so that its changes in expression level or due to mutational events of

the protein itself or its interactome can be taken into account to clarify the atlas of alteration of the protein in different cancer types. The structural methods used in this study also allowed to provide a detailed assessment of the different effects that the mutations in ULK1 found in the cancer samples can cause to perturb its structural stability or activity. Our results can be exploited for a better selection of ULK1 as target in a certain cancer type and guide both experimental research in cancer cellular biology and more pharmacological or clinical-oriented efforts.

MATERIALS AND METHODS

All the inputs, scripts and main outputs of this study are available in the Github repository https://github.com/ELELAB/ULK1_mutations.

Expression levels of ULK1 in TCGA datasets

We downloaded and pre-processed level 3 harmonized RNA-Seq data (HTSeq count) for all the available datasets from TCGA. We downloaded the data in June 2019 from the Genomic Data Common (GDC) Portal using the GDCdownload function of *TCGAbiolinks* [51]. An overview of the analysed datasets is reported in **Table S1**. We employed the *TCGAbiolinks* function *GDCprepare* to obtain a Summarized Experiment object [52]. We removed outlier samples with the *TCGAanalyze_Preprocessing* function of *TCGAbiolinks* using a Spearman correlation cutoff of 0.6. We normalized the datasets for GC-content [53] and library size using the *TCGAanalyze_Normalization* function of *TCGAbiolinks*. Lastly, we filtered the normalized RNA-Seq data for low counts across samples using the function *TCGAanalyze_Filtering* with a 0.20 cutoff for quantile filtering. For the TCGA datasets where normal samples were missing, we used the unified dataset that integrates the Genotype-Tissue Expression (GTEx) datasets [54] of healthy samples and the TCGA data, as provided by the *Recount2* protocol [55]. We also employed this dataset as an additional source of information for the TCGA datasets with less than five normal samples (see Table S1). We used the *TCGAquery_Recount2* function of TCGAbiolinks [56] to query the GTEx and TCGA unified datasets. We carried out GC-content normalization and quantile filtering on the unified datasets, as described above.
Differential expression analyses have been carried out using *limma-voom* [57] as implemented within the *TCGAanalyze_DEA function* of *TCGAbiolinks* [56], along with *edgeR* to confirm the results, as we recently applied to another case study [58]. We included in the design matrix conditions (tumor vs normal) and the TSS (Tissue Source Site; the center where the samples are collected) or the Plates (where available) as source of batch-effects to assess the robustness of the estimate of changes in expression with respect to different correction factors. In all our DEA analyses, we defined as a cutoff to retain significant DE genes a log fold change

(logFC) >= 0.5 or <=-0.5, whereas a cutoff of 0.05 was used for the False Discovery Rate (FDR). We then retrieved the estimate logFC for ULK1 (ENSEMBL ID ENSG00000177169) and ULK2 (ENSG00000083290) in the different comparisons (**see Table S1**).

Curation and analyses of missense mutations of ULK1 from TCGA

We retrieved mutations for ULK1 from each TCGA cancer study using the *MuTect2* pipeline [59] as implemented in the *TCGAbiolinks* function *GDCquery_Maf*. We retained missense mutations in the kinase domain of ULK1 for the structural analysis. For each mutation, we also collected the following information: i) potential of cancer driver mutation with *Fathmm* [60]; ii) pathogenic potential with *PolyPhen* [61] and *SIFT* [62]; iii) *REVEL* score [63]; iv) interplay with post-translational modifications and functional short linear motifs using as a source of information *PhosphoSite* [64] and *ELM* [65], respectively; v) identification of the same mutation in *COSMIC* [66]. We verified that the mutations under investigation were not found in *ExAC* as natural polymorphisms with high frequency in the healthy population [67]. We also used *iSNO-AAPair* [68], *SNOSite* [69] and *NetPhos* [70] to predict *S*-nitrosylation or phosphorylation sites upon mutation to cysteine or phosphorylatable (serine, threonine and tyrosine) residues, respectively. We used the *NetPhos* predictor only for those mutations that were in solvent exposed sites upon analyzes of solvent accessibility of their sidechain with *NACCESS* (http://wolf.bms.umist.ac.uk/naccess/). Moreover, we verified that each mutation under investigation was the only one targeting the ULK1 gene in the sample where it was identified.

Interactome of ULK1 and co-occurrence of mutations

We retrieved the experimentally known ULK1 interactors through the *Integrated Interaction Database* (IID) version 2018-05 [71]. We then estimated the co-occurrence of mutations between ULK1 and each of these interactors, along with other ULKs kinases (i.e. ULK2, ULK3 and ULK4) with the *somaticInteractions* function of *maftools* R/Bioconductor package [72], which performs a pairwise Fisher's Exact test to detect significant pairs of genes.

Prediction of driver genes

We used the *oncodrive* function of *maftools* [72] to evaluate if ULK1 was predicted as driver gene in any of the cancer type under investigation. The function is based on the algorithm *oncodriveCLUST* [73].

Free energy calculations

We employed the *FoldX* energy function [74,75] to perform in silico saturation mutagenesis. Calculations with this empirical energy function resulted in an average ΔΔG (differences in ΔG between mutant and wild-type variant) for each mutation over five independent runs performed using: i) the X-ray structure of ULK1 (PDB entry 5CI7, [35]), ii) an ensemble of 20 representative conformations from the MD simulations with CHARMM22star or iii) with CHARMM27. The protocol is detailed in our previous publication [48]. We also performed a literature-based curation of mutations for which the effects have been studied experimentally and use them as a control of the quality of our predictions.

Molecular dynamics simulations

We carried out 1-µs molecular dynamics (MD) simulations for the human ULK1 kinase domains in explicit solvent using GROMACS software version 4.6 [76]. We used as starting structure the PDB entry 5CI7 after in silico retro-mutation of the phospho-Thr 180 to Thr, to provide a model of the unphosphorylated variant of the domain. We used two protein different force fields CHARMM22* [77], CHARMM27 [78] in combination with the TIP3P water model [79] to evaluate the robustness of our results with respect to different physical models.

We used a dodecahedral box applying periodic boundary conditions and a concentration of NaCl of 150mM, neutralizing the charges of the system. The simulated system (protein+water) accounted for 83077 atoms. The system was prepared by different steps of minimization, solvent equilibration, thermalization and pressurization. We carried out productive MD simulations in the canonical ensemble at 300 K using velocity rescaling with a stochastic term [80]. We applied the LINCS algorithm [81] to constrain the heavy atom bonds to use a time-step of 2 fs. We calculated long-range electrostatic interactions using the Particle-mesh Ewald (PME) summation scheme [82], whereas we truncated Van der Waals and short-range Coulomb interactions at 10 Å.

We verified the absence of artificial contacts between the periodic images of the protein in the simulations, which were always at a distance higher than 30 Å. We evaluated the quality of the conformational ensemble on 100 representative conformations equally spaced in time for each of the simulations using the machine-learning based approach implemented in *ResProx* [83] to predict the atomic resolution from structural ensembles of proteins. We obtained a predicted resolution of 1.53 +/- 0.30 and 1.66 +/- 0.12 Å for CHARMM22star and CHARMM27 simulations of ULK1 kinase domain, respectively. These values are very close to the resolution of the corresponding X-ray structure (1.74 Å), suggesting an overall quality of the MD-based ensemble.

We used Principal Component Analysis (PCA) of the covariance matrix of Cα atomic fluctuations to extract the principal motions from the MD simulations [84]. We performed the PCA on a concatenated trajectory of the two MD simulations of ULK1 to compare them in the same essential subspace.

CABS_flex ensembles of selected ULK1 mutant variants

For a selection of mutant variants of ULK1 (F14L, N96D, A101T, A125T, R137C, R137H and D138N) that have hub-behavior, we also collected conformational ensembles using the coarse-grained approach implemented in *CABS_flex* 2.0 [85].
We used as starting structures for *CABS_flex* calculations the models generated by *FoldX* during the mutational scan for each of these mutations. In particular, we selected the most representative models in terms of rotameric state of the mutated residue for each mutation. We then collected an ensemble of ten different conformations for each mutant variant to be used for the contact-based PSN analyses of hub residues described below, upon reconstruction of the corresponding full-atom models.

Protein structure networks

We employed a contact-based Protein Structure Network (PSN) to the MD ensemble as implemented in *Pyinteraph* [86]. We defined as hubs those residues of the network with at least three edges [45]. We used the node inter-connectivity to calculate the connected components, which are clusters of connected residues in the graph. We selected 5 Å as the optimal cutoff for the contact-based PSN using the *PyInKnife* pipeline [87]. The distance was estimated between the center of mass of the residue side chains. We removed spurious interactions during the simulations applying a persistence cutoff of 20% (i.e., each contact was included as an edge of the PSN only if occurring in 20% of the MD frames), as indicated in the original implementation of the method [86]. We applied a variant of the depth-first search algorithm to identify the shortest path of communication. We defined the shortest path as the path in which the two residues were non-covalently connected by the smallest number of intermediate nodes.

We also calculated the persistence of salt bridges and hydrogen bonds with *PyInteraph* and the corresponding networks. For salt-bridges, all the distances between atom pairs belonging to charged moieties of two oppositely charged residues were calculated. The charged moieties were considered as interacting if at least one pair of atoms was found at a distance shorter than 4.5 Å. In the case of aspartate and glutamate residues, the atoms forming the carboxylic group were considered. The NH3- and the guanidinium groups were employed for lysine and arginine, respectively. We also verified the consistency of the results with a cutoff of 5 Å. We applied a persistence cutoff to filter interactions of 20% also for these networks.

RESULTS AND DISCUSSION

Alterations in gene expression in ULK1 and ULK2 in different cancer types

Our study mostly focused on annotating the effects of mutations found in cancer samples in the ULK1 kinase domain. Nevertheless, it is also important to verify if the expression levels of the corresponding genes are also altered and can then (partially) compensate for the effect induced by the mutation. Thus, at first, we analyzed the changes in gene expression of ULK1 in the different cancer types available by TCGA and Recount2 initiatives, for a total of 30 different comparisons (**Figure 1, Table S1**). Due to the high functional redundancy of ULK1 and ULK2 [88], we also monitored the changes of ULK2 in the same cancer types. We used differential expression analyses based on linear or generalized linear models to estimate the changes in expression of all the genes in the dataset and then retrieved the estimate for ULK1 and ULK2 (**Figure 1**). We observed different changes depending on the cancer types. Brain, gynecological and esophagus cancer types feature a down-regulation of both genes, suggesting an impairment of their function in autophagy. In another group of tumors, compensatory effects are in play with one of the two down-regulated but the other up-regulated, as exemplified by LUSC (**Figure 1**). In other cases, only one of the two kinases is either up- or down-regulated, suggesting that the unaltered levels of the remaining one can also have a partial compensatory effect. No marked changes in expression of either ULK1 or ULK2 have been identified in the remaining eleven cancer types used in our study.

Functional elements of ULK1 kinase domain

The fold of the ULK1 kinase domain (residues 8-280) is similar to that of protein kinase A [89] and consist of catalytic and regulatory regions. It can be divided in to a small N-lobe (residues 8-92), a hinge region (93-EYCNGG-98) and a large C-lobe (99-280), as shown in **Figure 2A**. At first, we analyzed the available X-ray structure of ULK1 comparatively with other kinases to identify the most important functional elements of this kinase domain, which have been described in little details in the literature. This structural knowledge is of paramount importance to be able to appreciate the effects that the mutations found in cancer samples could exert on this protein. As the main reference for the comparison, we used c-SRC kinase for which a recent review article provides a complete overview of important functional and structural elements in kinases [90].
ULK1 kinase domain includes a long positively charged activation loop (165-174 and 178-191) that may play a role for substrate recognition and activity regulations (**Figure 2A**). ULK kinases share the long disordered activation loop, a feature which is rare in the rest of the kinome [91] with only five other kinases with insertions of comparable size (i.e., Sgk, Wee1, Clk, Mlk and JAK). The kinase domain can be activated by phosphorylation on Thr180 on the activation loop [9,92] **Figure 2A**). The activation loop of ULK1 also includes the invariant kinase DFG motif (Asp165-Phe166-Gly167) and extends up to the APE motif (Ala189-Pro190-Glu191). In active kinases, the activation loop generally forms a cleft for substrate binding. The bound substrates form specific interactions with the conserved HRD motif of kinases (His136-Arg137-Asp138 in ULK1) which occurs in the catalytic loop of (**Figure 2B**).

The active state generally exhibits a salt bridge interaction between a conserved Lys (K46 in ULK1) residue in the β3 strand and a glutamate residue (E63) in the C- helix (48-69, **Figure 2B**). This salt bridge is also conserved in the structures from the MD ensembles collected in this study (see below). A basic patch has been observed in ULK1 around K162, which is also an acetylation site important for protein activation [34]. This basic patch can be involved in interactions of ULK1 with membrane structures or ATG13, along with for the binding to its own C-terminal domain [34].
ULK1 is known to prefer serine in the target substrates [93], a trait that we predict related to the presence of the F168 as DFG+1 residue (**Figure 2B**), in agreement with the findings by Chen et al. [94] that large hydrophobic DFG+1 residues promote Ser phosphorylation.
Kinases often present a gatekeeper residue in the active site [95], which corresponds to Met92 in ULK1 (**Figure 2B**). Mutation of gatekeeper residues in other kinases has been associated with the development of chemotherapeutic resistance [96]. Such resistance occurs strategically by reducing the Km of ATP apparently increase the concentration required for therapeutic effect, which consequently develop resistance later [95]. Mutations in the gatekeeper position of kinases have been also exploited to implement a strategy for improving inhibitors potency, favouring large-to-small mutation at this site [97]. In this context, methionine (as observed in ULK1) is among the larger and bulky gatekeeper residues and in other kinases mutation of a threonine gatekeeper to methionine was associated to the development of drug resistance [98].
A Gly-rich loop in the proximity of the DFG motif is also important for kinase function [99] and we observed two glycine residues in this loop in ULK1 (G25 and G23).
The regulatory (RS0: D203, RS1: H136, RS2: F166, RS3: L67 and RS4: L78, **Figure 2C**) and the catalytic spines (L21, V30, A44, L145, I144, L146, I210 and C214, **Figure 2D**) observed for c-SRC are conserved in ULK1. The only exceptions are the substitution of L455 of c-SRC catalytic spine with C214 in ULK1, along with the substitution of the RS3 methionine of c-SRC with a leucine in ULK1.
A comparison of active and inactive regulatory spines (R-spines) of c-SRC showed that the RS3 residue in the C-helix of the dormant enzyme is displaced and the spine not properly aligned when compared to the active enzyme [90]. The R-spine consists of residues from both the N- and C-lobes. The histidine of the HRD motif and the phenylalanine of the DFG motif also contribute to the R-spine formation. The V30 and A44 of the catalytic spine of ULK1 should be the residues for the binding with the adenine group of ATP, whereas one of the leucine residues is important for the interaction with the adenine base. The catalytic and the regulatory spines control catalysis by dictating the positioning of the ATP and the substrate, respectively. Thus, their proper alignment is necessary for the assembly of an active kinase.

ULK1 microsecond dynamics

Kinases are characterized by very complex conformational changes and several dynamic elements [100,101], we thus collected all-atom MD simulations in explicit

solvent to provide the first description of the ensemble of conformations of the ULK1 kinase domain in solution. An ensemble of conformations as the one provided by MD is also useful in the structural analyses that we used for the annotation of the mutations found in ULK1 in cancer genomics study (see below), as we recently applied to other target proteins [48,102]. To rule out patterns that are depending from the physical model used in the simulations, we collected MD simulations of the ULK1 kinase domain with two different force fields (CHARMM22star and CHARMM27). Using a dimensionality reduction approach based on Principal Component Analysis (PCA), we compared the sampling [103] achieved with the two force fields (**Figure 3A**) and we identified a good overlap, suggesting that the two simulations give a consistent view on the ULK1 dynamics. We also evaluated the principal motions described by the first PC (**Figure 3B-C**) and we noticed that there is concerted motion between the loop 148-158 and a part of the activation loop (172-183), along with the loop adjacent to the catalytic lysine (in the region 35-41). The two disordered regions 148-158 and 172-183 feature closing motions toward the rest of the ULK1 structure. The conformational change in the first region seems to be triggered by the electrostatic interactions between R152, R153 in the first disordered region and a negatively charged residue (D102) on the helix in front of them. Simultaneously with this motion, a conformational change in the region 35-41 occurs where this loop moves apart from the L78 residue of the regulatory spine. The relevance of these conformational changes will be discussed with attention to the ULK1 missense mutations found in cancer samples in the sections below.

Missense mutations in cancer of ULK1 kinase domain

We retrieved the missense mutations in the coding region of ULK1 gene for each cancer study in TCGA (**Table S2, Figure 4A**). We identified the majority of the mutations in UCEC, COAD, STAD and LUSC, along with one or two mutations in other eleven cancer types. ULK1 is also predicted as a driver gene in UCEC, COAD and LUSC by *OncodriveCLUST* (**Table S2**), a method based on positional clustering and exploiting the notion that variants in cancer causing genes are enriched at few specific loci [73]. Most of them are cancer types in which there are no marked changes in gene expression of ULK1, with the exception of GBM and UCEC where both ULK1 and ULK2 genes are down-regulated. In total, we identified 36 different mutations distributed over the whole structure (**Figure 4B**) of which D138N of the HRD motif occurred in both PAAD and LUSC cancer types and the R137, R252 and R261 sites were mutated to different residues in different cancer types. Methods for fast annotation of damaging and neutral mutations poorly agree on the effect of each of them (**Figure S1**). Moreover, we noticed some of these methods identified several potential damaging mutations, which is likely uncommon in cancer where most of the mutations will have a passenger effect, whereas others are too strict and then to predict as damaging mutations the ones at functional sites (such as D138N) without considering the structural impact of the mutation itself.

We thus followed a workflow similar to the one that we recently applied to another protein [102] for a more comprehensive assessment and understanding of the effects induced by these mutations. We used different methods which can help to pinpoint effects more related to the structural stability of the protein or to its function. These aspects are dissected one by one in the following sections.

Co-occurrence of mutations in ULK1 interactome

We curated the ULK1 interactome mining the *IID* protein-protein interaction database [71] and estimate the co-occurrence of mutations among each of the 30 identified interactors and ULK1 mutations for each cancer type (**Table S1**). Indeed, to better appreciate the effect of the mutations identified for a protein in the cancer samples, the knowledge of the spectrum of alteration of the biological partners of this protein provides a more complete scenario. We found co-occurrence of mutations between ULK1 and its interactors in eleven out of the 14 cancer types in which the gene has been found mutated (see *Github* repository for more details on each of them). Among these, CESC, COAD, SKCM, STAD and especially GBM, PAAD and UCEC (**Figure 5**) are characterized by a marked presence of co-occurring mutations in a large number of members of the ULK1 interactome, especially the ones important for the upstream regulation of autophagy (such as AMBRA1, components of the mTOR and ULK1 complex, RB1CC1/FIP200, TBC1, AMPK subunits, members of the ATG8 family, ATG16L1, BECN-1, PDPK,I RGM, RAB1A, P62,MINK1, SDCBP, ATG13 ….). This suggests a pronounced effect of alterations related to ULK1 function and activity, and, more in general, upstream steps of autophagy, in these cancer types.

Interplay of the ULK1 mutations with post-translational modifications and functional motifs

We then evaluated each mutation site in the context of interplay with PTMs and overlap with functional short linear motifs (SLiMs), along with the potential of harbouring new PTMs sites (**Table S2**)
Moreover,we also evaluated if any of the mutations was located in the LC3 interaction region (LIR) of ULK1 [104–106], which is placed in a distal region with respect to the kinase domain. This analysis was pursued upon the observation we found members of the ATG8 family among the interactors with co-occurrence of mutations with ULK1 in some of the TCGA cancer studies (**Figure 5**).
LIRs are SLiMs for interaction between the LC3/GABARAP (ATG8) family members and other autophagy proteins and key mediators of autophagosome formation [107]. We recently found mutations of LC3B co-occurring with mutations in its LIR-containing interactors in cancer genomic data [102] and we thus aimed here to evaluate if the same happens for ULK1. We did not find any mutations in the surrounding of the ULK1 LIR in the TCGA datasets under investigation, thus suggesting that the recognition between ULK1 and the ATG8 family is not a major driver of its alterations in cancer.

The only mutation in the proximity of a SLiM is the C-terminal D279N for an IAP (Inhibitor of Apoptosis Protein) binding motif (IBM), which has not been characterized in ULK1 at the best of our knowledge. Interestingly, one of the ULK1 interactors, BIRC2/ (**Table S2**) is a cellular inhibitor of apoptosis and promote autophagy, and it associate to ULK1 during mitophagy [108]. We speculate that this interaction could be mediated by the IAP motif of ULK1 and that the mutation of D279N could impair it. The mutations have been found in uterine cancer, which is also one of the datasets with a co-occurrence of mutations between ULK1and BIRC2.
We also did not identify any mutations directly altering an experimentally validated PTM site. Nevertheless, we identified one solvent-exposed mutation site (S195) which is predicted as a phosphorylation site for DNAPK or ATM kinases. Mutation to proline could have the effect to abolish this modification.
We then analyzed the mutations to serine, threonine or tyrosine for their capability to introduce new phosphorylatable residues, along with mutations to cysteine for their possibility to introduce a redox-sensitive PTM, such as *S*-nitrosylation ([109] **Table S2**). We only predicted the mutation P250S to result in a phosphorylatable serine by PKC kinase and the R137C mutation of the HRD motif as a possible *S*-nitrosylation site.

Assessment of the impact on ULK1 protein stability of the missense mutations

One of the main effects that a mutation can have on a protein is to alter its structural stability, causing local misfolding and a higher propensity for the mutated variant to be targeted by pathways for protein clearance, such as proteasomal degradation [44,110,111]. In this scenario, the function of the protein will be also affected but as a result of compromised protein levels of the protein itself not necessarily an alteration of its capability to interact with the partners or be active, depending on the degree of the damage induced by the mutation.
We thus used a high-throughput saturation mutagenesis approach based on an effective empirical energy function implemented in *FoldX* to predict the effect on protein stability induced by all the possible substitution of each position of the ULK1 kinase structure [48]. This approach has the advantage of providing both the estimate of the damaging effect of the disease-related mutation of interest and a pre-computed list of predicted changes in stability for any other mutation of the protein. The latter is useful to identify important hotspots for the structure of ULK11, along with accessibility to pre-annotated effects of mutations which will be found in future genomics studies.
To overcome the inherent issues in local sampling of the *FoldX* free energy function, we used the MD-derived ensembles for the estimate of the changes in the free energy of unfolding upon mutation (**Table S3-S6**), as recently applied to other cases study [48,102,112]. We found a good agreement between the predicted ΔΔGs using the two MD-derived ensembles of conformations (**Table S3**). Moreover, we noticed that for some mutations the predicted damaging effect was a result of the usage of the X-ray structure only (**Table S3**), whereas the possibility to account for the

flexibility of the protein structure in the prediction resulted in neutral effect. One example of this behaviour is A28V (**Table S3**). We thus used the results of ΔΔGs from the calculation on the MD ensemble to classify the ULK1 mutations found in cancer patients (**Figure 4B**). Two mutations (S184F and V211I) resulted in a stabilization of the protein architecture, suggesting a better packing of the protein. S184 is located in the proximity of aromatic residues, including the one of the DFG motif, which is important for catalysis so it cannot be excluded that a mutation to S184F could has the negative effect of altering the functional dynamic of the kinase at this site.

We also noticed that some of the mutation sites are general hotspots for ULK1 structural stability (**Figure 6A**) since it is not only the mutation found in the cancer samples that is predicted with a high ΔΔG. These sites are sensitive to substitution with most of the other residues (i.e., F14, S56, L78, F81, A125, R137, A169, G183, and F273). I135 is also a stability hotspot with the I135V mutation found in the cancer sample as one of the few tolerated substitutions at this site, suggesting a neutral effect for it.

The availability of MD ensembles for ULK1 prompted us also to apply a Protein Structure Network (PSN) approach based on the persistence of side-chain contacts in the conformational ensemble [86,87] to estimate hub residues, which are often corresponding to important residues for protein stability (**Figure 6B-C**). We thus verified which of the mutation sites were found in correspondence or proximity of a hub, as an additional parameter to evaluate the impact that the mutation has on protein stability. Moreover, to account for the fact that the mutation in a hub can still retain its hub capability, we collected the same analyses on conformational ensembles derived for each of the mutant variants where the mutation site is a PSN hub (**Table S3**). We classified as damaging mutations according to this parameter only the ones for which the mutation was abolished the hub behaviour. In particular, only in one case (F14L) we found that the hub behaviour was maintained. Most of the mutation sites corresponding to hubs were overlapping with mutations associated to high ΔΔGs with the exception of D138N.

Another strategy to impair structural stability could be related to the loss of electrostatic interactions in the form of salt bridges or hydrogen bonds. We thus calculated the persistence of the salt bridges and their network in the MD ensembles and annotate which of the mutation site where likely to abolish these interactions (**Table 1**). Most of the mutations of residues involved in salt bridges are likely to have marginal effects since they conserve either the negatively charged nature of the wild-type residue or they are replaced by asparagine which could still give rise to interactions of electrostatic nature with the guanidinium group of arginine. The only mutations which could impair salt-bridge formation are R152L and D268H. R152L was only found to be involved in salt bridges in the CHARMM27 MD ensemble, whereas even an increase of the distance cutoff for the analyses with CHARMM22star simulations showed a loosely tendency to form electrostatic interactions. MD force fields are known to have limitations in overestimating salt bridge contributions in some cases [113,114] and according to a recent

benchmarking [114], we selected the CHARMM22star results for the annotation of the mutations.
We noticed that most of the predicted destabilizing mutations are located at buried or partially accessible sites of the protein (according to an estimate of solvent accessibility). SASA and thermodynamic stability are known to be inversely associated [115]. Moreover, in agreement with patterns reported by Tokuriki et al. [116], we observed that for some of the ULK1 mutation sites (such as A169) the destabilizing effect arises from the replacement of a hydrophobic (A) with a polar residue (T).
Overall 58%of the ULK1 mutations found in the different cancer types were predicted damaging for protein stability using at least one of the two criteria above.

Assessment of the impact of mutations on ULK1 function

Apart from effects on the stability, the availability of structure and dynamics of ULK1 allows also to infer possible effects induced by the mutations on its function.
We thus evaluated the occurrence of the mutations in proximity of the disordered regions interested by the low frequency motions underpinned by PCA (**Figure 3B-C)**. R152L and D102N are likely to impair and weaken, respectively the closing motions observed for the region 148-158, which is triggered by their electrostatic interaction. We cannot rule out that the presence of R153 in the R152L mutant variant can partially compensate for the mutation. On the other side, the closing lateral motion of the region 172-183 o the activation loop is surrounded by different mutation sites, such as M177 in the loop itself, S195 in the region where the loop bents and interact, and Y171, G183/S184 which act as hinges for the motion of these regions. It could thus be expected that these mutations could impair its functional dynamics. In addition, the regulatory spine residue L78 and F81 in the proximity of the other disordered loop (35-41) also involved in the concerted conformational changes are also mutation sites.
In particular, PSN approaches can also be used to infer functionally-damaging sites if the paths of communication between the mutation sites and other important functional sites for the kinase activity are taken into account. This analysis can indeed shed light on effects that can been transmitted long-range, often at the base of allostery [47,49,117]. We thus estimated all the shortest paths of communication between each mutation site and five important classes of residues for kinase function. In particular, we selected three groups of target residues: i) residues important for activity (K46, E63, M62, T180 and K162); ii) residues of the DFG, HRD, and APE motifs; iii) central residues of the C-helix (55-65); iv) the residues of the catalytic and v) regulatory spines. We selected only those paths conserved both in the CHARMM22star and CHARMM27 simulations (**Table S7**). We did not find any communication to the regulatory spine or to the APE motif. On the contrary, a subset of mutation site (A28, A101, D102, D138, N96, and R137) was communicating with at least two of the target areas of interest, often using multiple paths. This result suggests that substitutions at these sites could be detrimental for functional long-

range communication or it could increase it (if oncogenic) especially in cases in which the steric hindrance or the physico-chemical properties of the wild-type residue are not maintained upon mutation, as it is the case for A28V, A101T, R137C and R137H.

General assessment and classification of ULK1 missense mutations in TCGA

We integrated all the results collected by the analyses described above to provide an overall view on the several properties and layers of alterations that a mutation in a protein can cause and are ultimately connected to the alterations in its function at the cellular level. In particular, with our framework we can assess: i) effects on protein structural stability, which will impact on the protein levels and turnover in the cell; ii) interplay with post-translational modifications and emergence of new layers of regulation; iii) alterations of binding regions for biological partners; iv) co-occurrence of mutations with other biological partners; and v) long-range functional effects.
We then classified the mutations according to each of these properties as 'damaging' or 'neutral' (**Figure 8A**) and then rank them. The ranking allowed to identify mutations that are likely to be 'driver' or 'passenger', along with to identify if the effect is triggered more by a destabilization of the protein product or a stable protein variant with impaired functionality **(Figure 8B**).
We observed, as stated above, that more than 50% of the mutations of ULK1 found in the cancer samples have an effect on protein stability. This is often accompanied by a possible impact also on the native functional properties of the same variant. A minority of mutations are only damaging for stability (A125T, F273V, L215P, F14L and G12D) and do not alter the functional state of the protein. The detrimental effect on the protein stability observed by these mutations could affect the level and turnover of the protein at the cellular level. This effect will be ultimately dominating with respect to the effects that the same mutations can exert on the protein activity.
ULK1 stability can also be modulated by the interaction with ATG13 and FIP200, which bind to the cytoplasmic domain of the ULK1 [39], or by chaperonins like proteins, such as p32 [118]. Thus, these interactions can compensate for the loss of stability induced by some of the mutant variants. Interestingly, in several cases the cancer types where destabilizing mutations of ULK1 occurred, were also featuring co-occurrence of mutations in ATG13 or FIP200 (**Table S3**), suggesting an overall alteration of ULK1 stability. Three mutations (S184F, D102N, and A28V) are predicted with a possible impact only on kinase activity, either altering the functional dynamics of the protein or the capability to exert long range effects from distal site to the functional and catalytic regions.
We searched in the literature if the mutation sites under investigation have been subjects of experimental studies to validate our prediction. We identified a study reporting a mutation of S184 to alanine [119]. S184 is mutated to phenylalanine in the HNSC dataset and we observed a marginally stabilizing effect upon mutation. Moreover, we classified S184 with a functional impact and likely to impair the functional dynamics highlighted by PCA since it is one of the possible hinges for the

motions of the activation loop region 172-183. S184A and S184D have been reported to extremely inactivate the ULK1 kinase [119], thus supporting the predicted functional role more than an effect on structural stability. We also collected other mutations at other sites for which the functional impact has been studied experimentally and they are summarized in **Table S3.** Of interest, S174A mutation results in a hyperactive enzyme [119] and it is also located in the aforementioned region for conformational changes of the activation loop. Moreover, we noticed that all the other experimental mutations with a reported effect on ULK1 activity are predicted to have marginal effects on protein stability, with the exception of K46N, M92A and Y89A. These mutations result either in an inactivation of the kinase [120] or an impairment of phosphorylation of the ATG13 substrate [27]. Our calculations suggest that an additional detrimental effect can come from altering the native stability of the protein, an aspect which could deserve further investigation to verify the cellular levels and half-life of these variants to assess which one is the predominant effect.

CONCLUSIONS

The assessment of the different effects that a mutation can exert on a protein explored in this study and the subsequent classification of the mutations can provide a useful complement to cancer genomics studies. For example, it allows to identify mutations that are likely to be ‘driver’ or ‘passenger’, along with to predict if the effect is triggered more by a destabilization of the protein product or a protein variant with impaired functionality. Moreover, our combined approach for mutation assessment could also benefit for the prioritization and selection of mutant variants for cellular experimental validation. Indeed, it can suggest how to select the proper readout for experimental validation. As an example, in a case where the mutation is predicted to be damaging for stability, experiments to estimate its cellular levels and half-life could be used, along with readouts to evaluate if the changes are due to proteasomal degradation or other clearance mechanisms. On the other side, if a mutation is predicted to result in a variant which is as stable as the wild-type, but the effect is more related to its function, experiments to evaluate its interactions in the cell with the biological partners, its regulation by PTMs or cellular assays to evaluate the effects on the pathways where interactors of the target protein are involved would be the most suitable choice. Moreover, we showed how the structural analyses used here benefit of the integration of bioinformatic tools to assess the changes in expression level of the target gene along with changes in other genes that can have compensatory effects, as we exemplified for ULK2. In addition, the extension of the analyses to the protein target interactome in terms of understanding co-occurrence of alterations and synergic effects that can arise from them allow a comprehensive view and to pinpoint interesting alterations at the molecular level. We here showed how our workflow can help in the study of an underappreciated key kinase of the autophagy pathway, ULK1. We discovered that in the majority of the cases the gene expression levels are not altered or can be compensated by an up-regulation of the homologous kinase ULK2, whereas more than 30 different

missense mutations altering the coding region of the gene have been identified. These mutations co-occur with mutations in ULK1 interactors fundamental for the upstream regulation of autophagy, suggesting an impairment of this process in cancer types such as uterine, stomach, skin, glioblastoma and colon cancers. Moreover, our study allowed to pinpoint that more than 50% of the mutations of ULK1 found in the cancer samples have an effect on protein stability, which is likely to have a more pronounced effect that the residual effect on protein activity, especially if it cannot be compensated by interactions with regulators of cellular ULK1 stability, which are also altered in the samples under investigation. We identified three mutations (S184F, D102N, and A28V) that predicted with only impact on kinase activity, either altering the functional dynamics of the protein or the capability to exert long range effects from distal site to the functional and catalytic regions. Future studies will be required to understand if these mutations have an inhibitory or activatory role on the kinase. The framework here applied could be more broadly extended to other targets of interest, as we recently started to apply, to help in the classification of mutational effects, along with prioritizing the variants for experimental validation and a specific biological readout.

ACKNOWLEDGMENTS

The authors would like to thank Matteo Lambrughi and Matteo Tiberti for technical assistance and fruitful discussion. The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The project was supported by Danmarks Grundforskningsfond (DNRF125) and a Carlsberg Foundation Distinguished Fellowship (CF18-0314). Moreover, the project has been supported by a Netaji Subhash ICAR international fellowship, Govt. of India to MK to work in EP group. The calculations described in this paper were performed using the DeiC National Life Science Supercomputer Computerome at DTU (Denmark), and DECI-PRACE 14th HPC Grant for calculations on Archer (UK).

Author contributions

Conceptualization: EP; Data Curation: MK, EP; Formal Analysis: MK, EP; Funding Acquisition: MK, EP; Investigation: MK, EP; Methodology: EP; Project Administration: EP; Resources: EP; Supervision: EP; Validation: MK, EP; Visualization: EP; Writing-Original Draft: EP; Writing-Review and Editing: MK, EP.

References

1. Reggiori F, Klionsky DJ. Autophagy in the Eukaryotic Cell. Eukaryot Cell. 2002;1: 11–21. doi:10.1128/ec.01.1.11-21.2002
2. Galluzzi L, Baehrecke EH, Ballabio A, Boya P, Bravo San Pedro JM, Cecconi F, et al. Molecular definitions of autophagy and related processes. EMBO J.

2017;36: 1811–1836. doi:10.15252/embj.201796697
3. Yin Z, Pascual C, Klionsky D. Autophagy: machinery and regulation. Microb Cell. 2016;3: 457–465. doi:10.15698/mic2016.12.546
4. Yu L, Chen Y, Tooze SA. Autophagy pathway: Cellular and molecular mechanisms. Autophagy. 2018;14: 207–215. doi:10.1080/15548627.2017.1378838
5. Mizushima N, Levine B, Cuervo AM, Klionsky DJ. Autophagy fights disease through cellular self-digestion. Nature. 2008;451: 1069–1075. doi:10.1038/nature06639
6. Noda NN, Fujioka Y. Atg1 family kinases in autophagy initiation. Cell Mol Life Sci. Springer Basel; 2015;72: 3083–3096. doi:10.1007/s00018-015-1917-z
7. Zachari M, Ganley IG. The mammalian ULK1 complex and autophagy initiation. Essays Biochem. 2017;61: 585–596. doi:10.1042/ebc20170021
8. Mizushima N. The role of the Atg1/ULK1 complex in autophagy regulation. Curr Opin Cell Biol. Elsevier Ltd; 2010;22: 132–139. doi:10.1016/j.ceb.2009.12.004
9. Lin MG, Hurley JH. Structure and function of the ULK1 complex in autophagy. Curr Opin Cell Biol. Elsevier Ltd; 2016;39: 61–68. doi:10.1016/j.ceb.2016.02.010
10. Antonioli M, Di Rienzo M, Piacentini M, Fimia GM. Emerging Mechanisms in Initiating and Terminating Autophagy. Trends Biochem Sci. Elsevier Ltd; 2016;xx: 1–14. doi:10.1016/j.tibs.2016.09.008
11. Dorsey FC, Rose KL, Coenen S, Prater SM, Cavett V, Cleveland JL, et al. Mapping the phosphorylation sites of Ulk1. J Proteome Res. 2009;8: 5253–5263. doi:10.1021/pr900583m
12. Alers S, Löffler AS, Wesselborg S, Stork B. The incredible ULKs. Cell Commun Signal. 2012;10: 7. doi:10.1186/1478-811X-10-7
13. Nazio F, Carinci M, Valacca C, Bielli P, Strappazzon F, Antonioli M, et al. Fine-tuning of ULK1 mRNA and protein levels is required for autophagy oscillation. J Cell Biol. 2016;215: 841–856. doi:10.1083/jcb.201605089
14. Egan DF, Shackelford DB, Mihaylova MM, Gelino S, Kohnz RA, Mair W, et al. Phosphorylation of ULK1 (hATG1) by AMP-activated protein kinase connects energy sensing to mitophagy. Science (80-). 2011;331: 456–461. doi:10.1126/science.1196371
15. Kim J, Kundu M, Viollet B, Guan K-L. AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. Nat Cell Biol. Nature Publishing Group; 2011;13: 132–41. doi:10.1038/ncb2152
16. Nazio F, Strappazzon F, Antonioli M, Bielli P, Cianfanelli V, Bordi M, et al. mTOR inhibits autophagy by controlling ULK1 ubiquitylation, self-association and function through AMBRA1 and TRAF6. Nat Cell Biol. 2013;15: 406–16. doi:10.1038/ncb2708
17. Wong P-M, Puente C, Ganley IG, Jiang X. The ULK1 complex: sensing nutrient signals for autophagy activation. Autophagy. 2013;9: 124–37. doi:10.4161/auto.23323
18. Puente C, Hendrickson RC, Jiang X. Nutrient-regulated phosphorylation of ATG13 inhibits starvation-induced autophagy. J Biol Chem. 2016;291: 6026–6035. doi:10.1074/jbc.M115.689646
19. Russell RC, Tian Y, Yuan H, Park HW, Chang YY, Kim J, et al. ULK1 induces autophagy by phosphorylating Beclin-1 and activating VPS34 lipid kinase. Nat Cell Biol. Nature Publishing Group; 2013;15: 741–750. doi:10.1038/ncb2757

20. Park JM, Jung CH, Seo M, Otto NM, Grunwald D, Kim KH, et al. The ULK1 complex mediates MTORC1 signaling to the autophagy initiation machinery via binding and phosphorylating ATG14. Autophagy. 2016;12: 547–564. doi:10.1080/15548627.2016.1140293
21. Egan DF, Chun MGH, Vamos M, Zou H, Rong J, Miller CJ, et al. Small Molecule Inhibition of the Autophagy Kinase ULK1 and Identification of ULK1 Substrates. Mol Cell. Elsevier Inc.; 2015;59: 285–297. doi:10.1016/j.molcel.2015.05.031
22. Mcalpine F, Williamson LE, Tooze SA, Chan EYW. Regulation of nutrient-sensitive autophagy by uncoordinated 51-like kinases 1 and 2. Autophagy. 2013;9: 361–373. doi:10.4161/auto.23066
23. Lazarou M, Sliter DA, Kane LA, Sarraf SA, Wang C, Burman JL, et al. The ubiquitin kinase PINK1 recruits autophagy receptors to induce mitophagy. Nature. 2015;524: 309–314. doi:10.1038/nature14893
24. Chen G, Hu Z, Li L, Zhang X, Zhou C, Zhang L, et al. ULK1 translocates to mitochondria and phosphorylates FUNDC1 to regulate mitophagy. EMBO Rep. 2014;15: 566–575. doi:10.1002/embr.201438501
25. Lu J, Zhu L, Zheng L ping, Cui Q, Zhu H huan, Zhao H, et al. Overexpression of ULK1 Represents a Potential Diagnostic Marker for Clear Cell Renal Carcinoma and the Antitumor Effects of SBI-0206965. EBioMedicine. The Authors; 2018;34: 85–93. doi:10.1016/j.ebiom.2018.07.034
26. Wu D hao, Wang TT, Ruan DY, Li X, Chen ZH, Wen JY, et al. Combination of ULK1 and LC3B improve prognosis assessment of hepatocellular carcinoma. Biomed Pharmacother. Elsevier; 2018;97: 195–202. doi:10.1016/j.biopha.2017.10.025
27. Zhang L, Fu L, Zhang S, Zhang J, Zhao Y, Zheng Y, et al. Discovery of a small molecule targeting ULK1-modulated cell death of triple negative breast cancer in vitro and in vivo. Chem Sci. 2017;8: 2687–2701. doi:10.1039/C6SC05368H
28. Yun M, Bai HY, Zhang JX, Rong J, Weng HW, Zheng ZS, et al. ULK1: A promising biomarker in predicting poor prognosis and therapeutic response in human nasopharygeal carcinoma. PLoS One. 2015;10: 1–15. doi:10.1371/journal.pone.0117375
29. Martinet W, Agostinis P, Vanhoecke B, Dewaele M, Meyer GRYDE. Autophagy in disease□: a double-edged sword with therapeutic potential. Clin Sci. 2009;712: 697–712. doi:10.1042/CS20080508
30. Singh SS, Vats S, Chia AYQ, Tan TZ, Deng S, Ong MS, et al. Dual role of autophagy in hallmarks of cancer. Oncogene. Springer US; 2018;37: 1142–1158. doi:10.1038/s41388-017-0046-6
31. Jang JE, Eom JI, Jeung HK, Cheong JW, Lee JY, Kim JS, et al. AMPK-ULK1-mediated autophagy confers resistance to BET inhibitor JQ1 in acute myeloid leukemia stem cells. Clin Cancer Res. 2017;23: 2781–2794. doi:10.1158/1078-0432.CCR-16-1903
32. Mercer TJ, Gubas A, Tooze SA. A molecular perspective of mammalian autophagosome biogenesis. J Biol Chem. 2018;293: 5386–5395. doi:10.1074/jbc.R117.810366
33. Petherick KJ, Conway OJL, Mpamhanga C, Osborne SA, Kamal A, Saxty B, et al. Pharmacological inhibition of ULK1 kinase blocks mammalian target of rapamycin (mTOR)-dependent autophagy. J Biol Chem. 2015;290: 11376–11383. doi:10.1074/jbc.C114.627778
34. Lazarus MB, Novotny CJ, Shokat KM. Structure of the human autophagy

initiating kinase ULK1 in complex with potent inhibitors. ACS Chem Biol. 2015;10: 257–261. doi:10.1021/cb500835z
35. Lazarus MB, Shokat KM. Discovery and structure of a new inhibitor scaffold of the autophagy initiating kinase ULK1. Bioorganic Med Chem. Elsevier Ltd; 2015;23: 5483–5488. doi:10.1016/j.bmc.2015.07.034
36. Galluzzi L, Pietrocola F, Bravo-San Pedro JM, Amaravadi RK, Baehrecke EH, Cecconi F, et al. Autophagy in malignant transformation and cancer progression. EMBO J. 2015;34: 856–880. doi:10.15252/embj.201490784
37. Chaikuad A, Koschade SE, Stolz A, Zivkovic K, Pohl C, Shaid S, et al. Conservation of structure, function and inhibitor binding in UNC-51-like kinase 1 and 2 (ULK1/2). Biochem J. 2019; BCJ20190038. doi:10.1042/BCJ20190038
38. Nicolaou CA, Humblet C, Hu H, Martin EM, Dorsey FC, Castle TM, et al. Idea2Data: Toward a New Paradigm for Drug Discovery. ACS Med Chem Lett. 2019;10: 278–286. doi:10.1021/acsmedchemlett.8b00488
39. Ganley IG, Lam DH, Wang J, Ding X, Chen S, Jiang X. ULK1·ATG13·FIP200 complex mediates mTOR signaling and is essential for autophagy. J Biol Chem. 2009;284: 12297–12305. doi:10.1074/jbc.M900573200
40. Zhu Y, Wang C, Yu M, Cui J, Liu L, Xu Z. ULK1 and JNK are involved in mitophagy incurred by LRRK2 G2019S expression. Protein Cell. 2013;4: 711–721. doi:10.1007/s13238-013-3910-3
41. Nussinov R, Tsai C, Jang H. Protein ensembles link genotype to phenotype. PLoS Comput Biol. 2019;15: e1006648. doi:10.1371/journal.pcbi.1006648
42. Nussinov R. Precision medicine review□: rare driver mutations and their biophysical classification. Biophys Rev. Biophysical Reviews; 2019;11: 5–19.
43. Naganathan AN. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. Curr Opin Struct Biol. Elsevier Ltd; 2019;54: 1–9. doi:10.1016/j.sbi.2018.09.004
44. Stein A, Fowler DM, Hartmann-Petersen R, Lindorff-Larsen K. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. Trends Biochem Sci. Elsevier Ltd; 2019; 1–14. doi:10.1016/j.tibs.2019.01.003
45. Papaleo E. Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity. Front Mol Biosci. 2015;2: 28. doi:10.3389/fmolb.2015.00028
46. Di Paola L, Giuliani A. Protein contact network topology: a natural language for allostery. Curr Opin Struct Biol. 2015;31: 43–8. doi:10.1016/j.sbi.2015.03.001
47. Lambrughi M, De Gioia L, Gervasio FL, Lindorff-Larsen K, Nussinov R, Urani C, et al. DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. Nucleic Acids Res. 2016;44: 9096–9109. doi:10.1093/nar/gkw770
48. Nygaard M, Terkelsen T, Olsen AV, Sora V, Salamanca J, Rizza F, et al. The mutational landscape of the oncogenic MZF1 SCAN domain in cancer. Front Mol Biosci. 2016; doi:10.3389/fmolb.2016.00078
49. Papaleo E, Saladino G, Lambrughi M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. Chem Rev. 2016;116: 6391–6423. doi:10.1021/acs.chemrev.5b00623
50. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. Nature Publishing Group; 2013;45: 1113–1120. doi:10.1038/ng.2764
51. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al.

TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2015;44: gkv1507-. doi:10.1093/nar/gkv1507
52. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods. Nature Publishing Group; 2015;12: 115–121. doi:10.1038/Nmeth.3252
53. Risso D, Schwartz K, Sherlock G, Dudoit S. GC-Content Normalization for RNA-Seq Data. 2011;
54. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project. Biopreserv Biobank. 2015;13: 307–308. doi:10.1089/bio.2015.29031.hmm
55. Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, et al. Reproducible RNA-seq analysis using recount2. Nat Biotechnol. 2017;35: 319–321. doi:10.1038/nbt.3838
56. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLOS Comput Biol. 2019;15: e1006701. doi:10.1371/journal.pcbi.1006701
57. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15: R29. doi:10.1186/gb-2014-15-2-r29
58. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140. doi:10.1093/bioinformatics/btp616
59. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. Nature Publishing Group; 2013;31: 213–219. doi:10.1038/nbt.2514
60. Shihab HA, Gough J, Cooper DN, Day INM, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. Bioinformatics. 2013;29: 1504–1510. doi:10.1093/bioinformatics/btt182
61. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; 1–41. doi:10.1002/0471142905.hg0720s76
62. Ng PC, Henikoff S. SIFT□: predicting amino acid changes that affect protein function. 2003;31: 3812–3814. doi:10.1093/nar/gkg509
63. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99: 877–885. doi:10.1016/j.ajhg.2016.08.016
64. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015;43: D512–D520. doi:10.1093/nar/gku1267
65. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. Sci Signal. 2013;6: rs7. doi:10.1126/scisignal.2003345
66. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019;47: D941–D947. doi:10.1093/nar/gky1015
67. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: An empirical approach to

evaluating population data for clinical variant interpretation. Genome Med. Genome Medicine; 2017;9: 1–14. doi:10.1186/s13073-017-0403-7
68. Xu Y, Shao X, Wu L, Deng N, Chou K. iSNO-AAPair: incorporating amino acid pairwise coupling into PseAAC for predicting cysteine S-nitrosylation sites in proteins. PeerJ. 2013; 1–18. doi:10.7717/peerj.171
69. Lee TY, Chen YJ, Lu TC, Huang H Da, Chen YJ. Snosite: Exploiting maximal dependence decomposition to identify cysteine S-Nitrosylation with substrate site specificity. PLoS One. 2011;6. doi:10.1371/journal.pone.0021849
70. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 2004;4: 1633–1649. doi:10.1002/pmic.200300771
71. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: Context-specific physical protein-protein interactions in human, model organisms and domesticated species. Nucleic Acids Res. Oxford University Press; 2019;47: D581–D589. doi:10.1093/nar/gky1037
72. Mayakonda A, Koeffler HP. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. bioRxiv. 2016; 052662. doi:10.1101/052662
73. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST□: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics. 2013;29: 2238–2244. doi:10.1093/bioinformatics/btt395
74. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. J Mol Biol. 2002;320: 369–387. doi:10.1016/S0022-2836(02)00442-4
75. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005;33: W382-8. doi:10.1093/nar/gki387
76. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4:□ Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008;4: 435–447. doi:10.1021/ct700301q
77. Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? Biophys J. 2011;100: L47-9. doi:10.1016/j.bpj.2011.03.051
78. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem. 2004;25: 1400–15. doi:10.1002/jcc.20065
79. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. AIP Publishing; 1983;79: 926. doi:10.1063/1.445869
80. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. AIP Publishing; 2007;126: 014101. doi:10.1063/1.2408420
81. Hess B, Bekker H, Berendsen H, Fraaije J. LINCS: A linear constraint solver for molecular simulations. J Comput Chem. JOHN WILEY & SONS INC; 1993;12: 1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
82. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys. AIP; 1995;103: 8577. doi:10.1063/1.470117

83. Berjanskii M, Zhou J, Liang Y, Lin G, Wishart DS. Resolution-by-proxy: a simple measure for assessing and comparing the overall quality of NMR protein structures. J Biomol NMR. 2012;53: 167–80. doi:10.1007/s10858-012-9637-2
84. Daidone I, Amadei A. Essential dynamics: foundation and applications. Wiley Interdiscip Rev Comput Mol Sci. 2012;2: 762–770. doi:10.1002/wcms.1099
85. Kuriata A, Gierut AM, Oleniecki T, Ciemny MP, Kolinski A, Kurcinski M, et al. CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. Nucleic Acids Res. 2018;46: W338–W343. doi:10.1093/nar/gky356
86. Tiberti M, Invernizzi G, Lambrughi M, Inbar Y, Schreiber G, Papaleo E. PyInteraph□: a framework for the analysis of interaction networks in structural ensembles of proteins. J Chem Inf Model. 2014;54: 1537–1551.
87. Viloria JS, Allega MF, Lambrughi M, Papaleo E. An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass. Sci Rep. Springer US; 2017;7: 1–11. doi:10.1038/s41598-017-01498-6
88. Lee E, Tournier C. The requirement of uncoordinated 51-like kinase 1 (ULK1) and ULK2 in the regulation of autophagy. Autophagy. 2011;7: 689–695. doi:10.4161/auto.7.7.15450
89. Noda NN, Fujioka Y. Atg1 family kinases in autophagy initiation. Cell Mol Life Sci. Springer Basel; 2015;72: 3083–3096. doi:10.1007/s00018-015-1917-z
90. Roskoski R. Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors This paper is dedicated to the memory of Prof. Donald F. Steiner (1930-2014) - Advisor, mentor, and discoverer of proinsulin. Pharmacol Res. Elsevier Ltd; 2015;94: 9–25. doi:10.1016/j.phrs.2015.01.003
91. Lazarus MB, Novotny CJ, Shokat KM. Structure of the human autophagy initiating kinase ULK1 in complex with potent inhibitors. ACS Chem Biol. 2015;10: 257–261. doi:10.1021/cb500835z
92. Bach M, Larance M, James DE, Ramm G. The serine/threonine kinase ULK1 is a target of multiple phosphorylation events. Biochem J. 2011;440: 283–291. doi:10.1042/BJ20101894
93. Papinski D, Kraft C. Regulation of Autophagy By Signaling Through the Atg1 / ULK1 Complex. J Mol Biol. The Authors; 2016; 1725–1741. doi:10.1016/j.jmb.2016.03.030
94. Chen C, Hak Ha B, Thevenin A, Lou HJ, Zhang R, Yip KY, et al. Identification of a Major Determinant for Serine-Threonine Kinase Phosphoacceptor Specificity. Mol Cell. 2013;53: 140–147. doi:10.1016/j.molcel.2013.11.013
95. Liu Q, Sabnis Y, Zhao Z, Zhang T, Buhrlage SJ, Jones LH, et al. Developing Irreversible Inhibitors of the Protein Kinase Cysteinome. Chem Biol. Elsevier Ltd; 2013;20: 146–159. doi:10.1016/j.chembiol.2012.12.006
96. Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. Mol Cancer. 2018;17: 48. doi:10.1186/s12943-018-0804-2
97. Garske AL, Peters U, Cortesi AT, Perez JL, Shokat KM. Chemical genetic strategy for targeting protein kinases based on covalent complementarity. Proc Natl Acad Sci U S A. 2011;108. doi:10.1073/pnas.1111239108
98. Suda K, Onozato R, Yatabe Y, Mitsudomi T. EGFR T790M mutation: a double role in lung cancer cell survival? J Thorac Oncol. 2009;4: 1–4. doi:10.1097/JTO.0b013e3181913c9f
99. Modi V, Dunbrack RL. Defining a new nomenclature for the structures of active and inactive kinases. Proc Natl Acad Sci U S A. 2019;116: 6818–6827.

doi:10.1073/pnas.1814279116
100. Saladino G, Gervasio FL. Modeling the effect of pathogenic mutations on the conformational landscape of protein kinases. Curr Opin Struct Biol. Elsevier Ltd; 2016;37: 108–114. doi:10.1016/j.sbi.2016.01.005
101. Ahuja LG, Taylor SS, Kornev AP. Tuning the " Violin " of Protein Kinases: The Role of Dynamics-Based Allostery. IUBMB. 2019;71: 685–696. doi:10.1002/iub.2057
102. Fas BA, Kumar M, Sora V, Mashkoor M, Lambrughi M, Tiberti M, et al. The conformational and mutational landscape of the ubiquitin-like marker for the autophagosome formation in cancer. bioarXiv. 2019; doi:10.1101/635284
103. Martín-García F, Papaleo E, Gomez-Puertas P, Boomsma W, Lindorff-Larsen K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. PLoS One. 2015;10: e0121114. doi:10.1371/journal.pone.0121114
104. Grunwald DS, Otto NM, Park J-M, Song D, Grunwald DS, Otto NM, et al. GABARAPs and LC3s have opposite roles in regulating ULK1 for autophagy induction. Autophagy. Taylor & Francis; 2019;0: in press, doi: 10.1080/15548627.2019.1632620. doi:10.1080/15548627.2019.1632620
105. Alemu EA, Lamark T, Torgersen KM, Birgisdottir AB, Larsen KB, Jain A, et al. ATG8 family proteins act as scaffolds for assembly of the ULK complex: Sequence requirements for LC3-interacting region (LIR) motifs. J Biol Chem. 2012;287: 39275–39290. doi:10.1074/jbc.M112.378109
106. Kraft C, Kijanska M, Kalie E, Siergiejuk E, Lee SS, Semplicio G, et al. Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy. EMBO J. Nature Publishing Group; 2012;31: 3691–3703. doi:10.1038/emboj.2012.225
107. Birgisdottir AB, Lamark T, Johansen T. The LIR motif - crucial for selective autophagy. J Cell Sci. 2013;126: 3552–62. doi:10.1242/jcs.120477
108. Mukhopadhyay S, Naik PP, Panda PK, Sinha N, Bhutia SK. Serum starvation induces anti-apoptotic cIAP1 to promote mitophagy through ubiquitination. Biochem Biophys Res Commun. Elsevier Ltd; 2016;479: 940–946. doi:10.1016/j.bbrc.2016.09.143
109. Bignon E, Allega MF, Lucchetta M, Tiberti M, Papaleo E. Computational Structural Biology of S-nitrosylation of Cancer Targets. Front Oncol. 2018;8. doi:10.3389/fonc.2018.00272
110. Nielsen S V., Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, et al. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. Nichols KE, editor. PLOS Genet. 2017;13: e1006739. doi:10.1371/journal.pgen.1006739
111. Scheller R, Stein A, Nielsen S V., Marin FI, Gerdes A-M, Di Marco M, et al. Toward mechanistic models for genotype-phenotype correlations in phenylketonuria using protein stability calculations. Hum Mutat. 2019; doi:10.1002/humu.23707
112. Papaleo E, Parravicini F, Grandori R, De Gioia L, Brocca S. Structural investigation of the cold-adapted acylaminoacyl peptidase from Sporosarcina psychrophila by atomistic simulations and biophysical methods. Biochim Biophys Acta - Proteins Proteomics. Elsevier B.V.; 2014;1844: 2203–2213. doi:10.1016/j.bbapap.2014.09.018
113. Jónsdóttir LB, Ellertsson B, Invernizzi G, Magnúsdóttir M, Thorbjarnardóttir SH, Papaleo E, et al. The role of salt bridges on the temperature adaptation of aqualysin I, a thermostable subtilisin-like proteinase. Biochim Biophys Acta -

Proteins Proteomics. Elsevier B.V.; 2014;1844: 2174–2181. doi:10.1016/j.bbapap.2014.08.011
114. Ahmed MC, Papaleo E, Lindorff-larsen K. How well do force fields capture the strength of salt bridges in proteins? PeerJ. 2018;6: e4967. doi:10.7717/peerj.4967
115. Tokuriki N, Stricher F, Serrano L, Tawfik DS. How Protein Stability and New Functions Trade Off. 2008;4: 35–37. doi:10.1371/journal.pcbi.1000002
116. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The Stability Effects of Protein Mutations Appear to be Universally Distributed. J Mol Biol. 2007;369: 1318–1332. doi:10.1016/j.jmb.2007.03.069
117. Invernizzi G, Tiberti M, Lambrughi M, Lindorff-Larsen K, Papaleo E. Communication Routes in ARID Domains between Distal Residues in Helix 5 and the DNA-Binding Loops. de Groot BL, editor. PLoS Comput Biol. 2014;10: e1003744. doi:10.1371/journal.pcbi.1003744
118. Jiao H, Su G, Dong W, Zhang L, Xie W, Yao L, et al. Chaperone-like protein p32 regulates ULK1 stability and autophagy. Cell Death Differ. 2015;22: 1812–1823. doi:10.1038/cdd.2015.34
119. Loska SL. Regulation of ULK1 in autophagy. PhD Thesis, Univ Manchester, Fac Life Sci. 2012;
120. Tomoda T, Bhatt RS, Kuroyanagi H, Shirasawa T, Hatten ME. A mouse serine/threonine kinase homologous to C. elegans UNC51 functions in parallel fiber formation of cerebellar granule neurons. Neuron. 1999;24: 833–46. Available: http://www.ncbi.nlm.nih.gov/pubmed/10624947

Figure legends.

Figure 1. Changes in gene expression for ULK1 and ULK2 in the 30 cancer datasets under investigation. In the plot, we reported the logFC comparing tumor versus normal samples, as estimated by *limma-voom* and including in the design matrix information on the TSS. If the logFC > 0.05 the gene is considered up-regulated in the tumor with respect to the normal samples, whereas if the logFC is <-0.05, the gene is down-regulated. When the logFC is set to 0 in the plot, it means that the gene is not differentially expressed according to the DEA.

Figure 2. Structural and functional element of the ULK1 kinase domain, described by similarity with the c-SRC kinase. We compared the PDB entries of ULK1 (4WNO) and c-Src (1Y57) by a structural alignment (**Figure S1**) and infer the corresponding important elements of the ULK1 kinase domain. **A)** The N- (residues 8-92) and C- (residues 99-280) lobes are shown, along with the hinge region (93-EYCNGG-98) of ULK1. The region of the activation loop (163-200) is highlighted in light blue, along with a stick representation for the side chains of the phosphorylation site T180, the DFG and APE motifs. **B)** The residues important for catalysis and substrate binding are shown. In particular, the catalytic lysine (K46) and its salt bridge with E63 in the N-lobe are shown as orange sticks, along with the gatekeeper methionine (M92). F168 adjacent to the DFG motif, which is likely to be important for substrate specificity, along with the HDR motif are also highlighted as sticks. The residues of the aligned regulatory spine (**C**) and of the catalytic spine (**D**) are shown as spheres. The catalytic lysine with dots and sticks as a reference of the active site.

Figure 3. Principal component analysis of MD simulations of ULK1 kinase domain. A) The overlap between the sampling described by the two simulations of ULK1 carried out with two different force fields is shown along the two first principal components of PCA as a visual reference. The two different simulations shown a certain degree of overlap and sampled similar regions of the PCA subspace described by PC1 and PC2, with some deviations as often encountered in comparisons of MD simulations of the same system with different force fields or even different replicates of the same system. **B-C)** Conformational changes described by the first principal component, which interest three disordered regions of the protein with concerted motions. As a reference, the mutation sites in this area (discussed in the main text) and the residues important for catalysis or the regulatory spine are shown in green and orange, respectively.

Figure 4. Missense mutations in the ULK1 kinase domain in TCGA datasets. A) The number of mutations found in the coding region corresponding to the ULK1 kinase domain are showed for each cancer type. **B)** The mutation sites are highlighted on the 3D structure, along with the sub-domain composition (N-, C-lobe and hinge in orange, marine and light grey, respectively). The residues of the DFG and HDR motifs are shown as sticks, whereas the catalytic lysine with sticks and dots.

Figure 5. Co-occurrence of mutations in ULK1 interactome. We used, as a description of the ULK1 interactome, the interactions annotated in *IID* and estimate the co-occurrence of mutations in the TCGA datasets under investigation. GBM (A), PAAD (B), UCEC (E) and SKCM (D) are shown here. The other data are available in the *Github* repository associated to this publication.

Figure 6. ULK1 mutations in TCGA cancer samples altering protein stability. We used two different estimates of the impact on protein stability, based on the calculations of changes in free energy of unfolding with an empirical scoring function (**A**) and the notion of hub residues in a Protein Structure Network (**B-C**). **A**) A heatmap with the estimated changes in free energy upon a deep mutational scanning of all the mutation sites is reported. As an example, we showed the results for the MD-ensemble with CHARMM22star force field. The results of the other mutational scans are reported as TableS4-S6. More than 50% of the ULK1 mutations are predicted to destabilize the protein structure and located in sites that are general hotspots for maintenance of the native fold. **B**) Hub prediction are in fair agreement using different methods to generate the ensemble of conformations, i.e. MD simulations with two different force fields (CHARMM22star and CHARMM27), along with a coarse-grained sampling method with a posteriori reconstruction of a full atom model (CABS-Flex), which overestimate some of the hubs in the C-terminal domain of the protein. **C**) Hub-behavior of the mutation site in the MD ensemble with CHARMM22star force field, as an example.

Figure 7. Shortest paths of communication between the mutation sites and key functional or regulatory regions of ULK1 kinase domain. The whole set of results is available in Table S7. We reported the data for CHARMM22star MD simulations for sake of clarity, which were very similar to the ones for the simulation carried out with CHARMM27. If a mutation site featured more than one path of communication for the same class of target residues, only the path with lower path length is taken into account in this figure. We used as

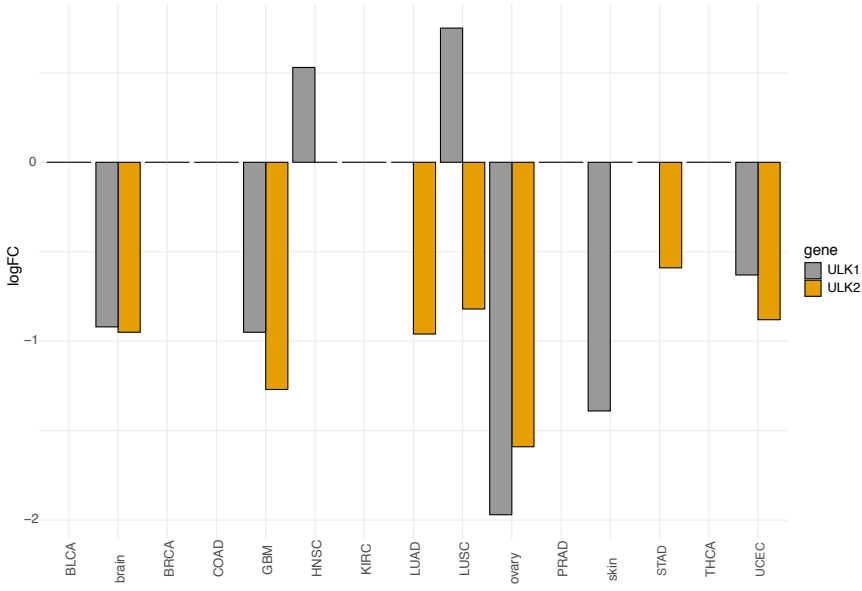
target residues: i) residues important for activity (K46, E63, M62, T180 and K162); ii) residues of the DFG, HRD, and APE motifs; iii) central residues of the C-helix (55-65); iv) the residues of the catalytic and v) regulatory spines. No paths were found from the mutation sites to the regulatory spine or the APE motif in both the MD simulations.

Figure 8. Classification of ULK1 missense mutations found in cancer genomics studies. A) The different analyses used in this study have been aggregated to associate the potential of damaging or neutral effect of each mutation. We used descriptors that account for either protein stability (purple) or function (blue). Mutations altering one of these properties are indicated as black dots. The diagram allows to link each mutation with a specific effect which could also guide the selection of the more appropriate set up for experimental validation. **B-C)** The heatmaps with the results of ranking on a collective score of damaging potential for the mutations in term of function (B) or stability (C). Darker the color more damaging the mutation is predicted to be.

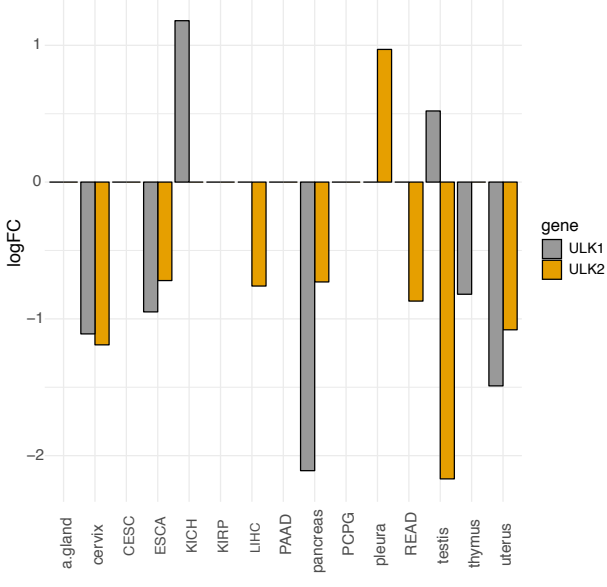
Table 1. Salt bridges involving ULK1 mutation sites characterized by charged residues, along with their persistence in the MD ensemble. R137, R252 and R261 are not reported since we did not find any stable salt bridges involving these residues in the simulations.

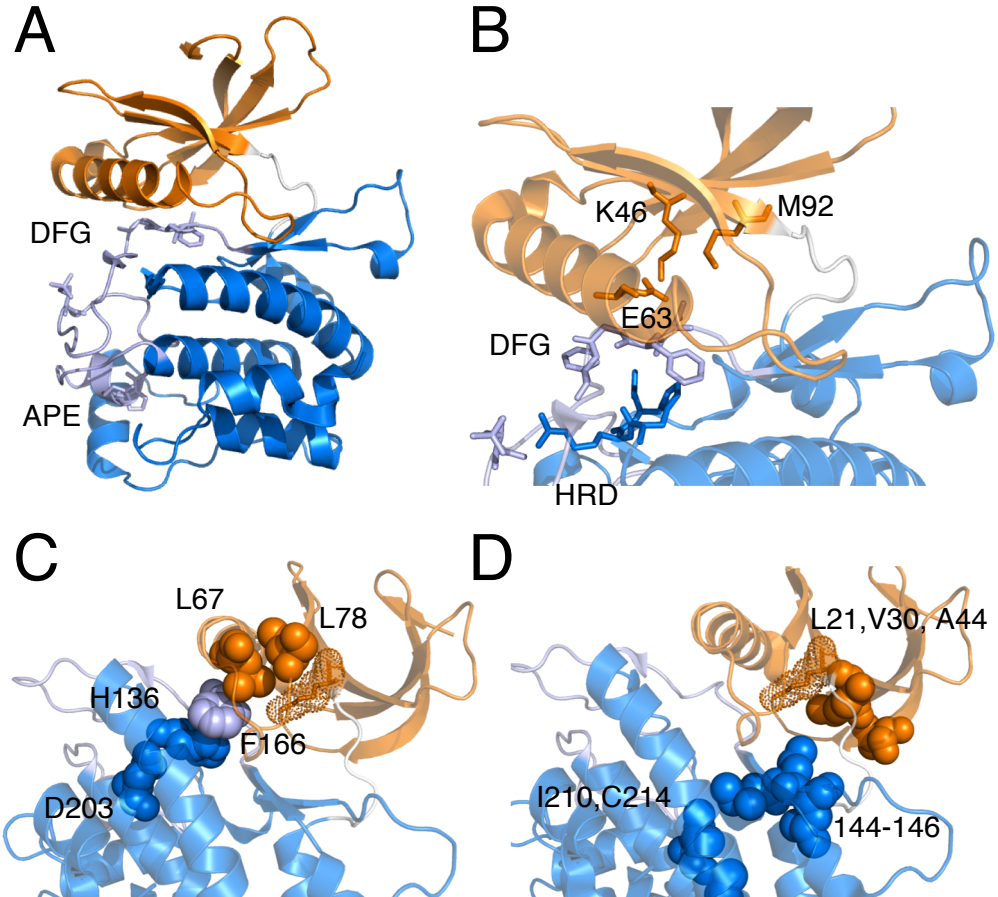
MUTATION	**CHARMM22star (%)**	**CHARMM27 (%)**
D102N		R152-D102-R153(67.5, 81.0)
D113E	D113-R116 (44.3)	D113-R116(55.8)
D138N	D138-K140 (99.8)	D138-K140(99.8)
R152L		R152-D102(67.5)
E246D		E246-R245(22.7)
D268H	D268-K201(98.8)	D268-K201(93.0)
D279N	D279-R116(63.2)	D279-R116(59.4)

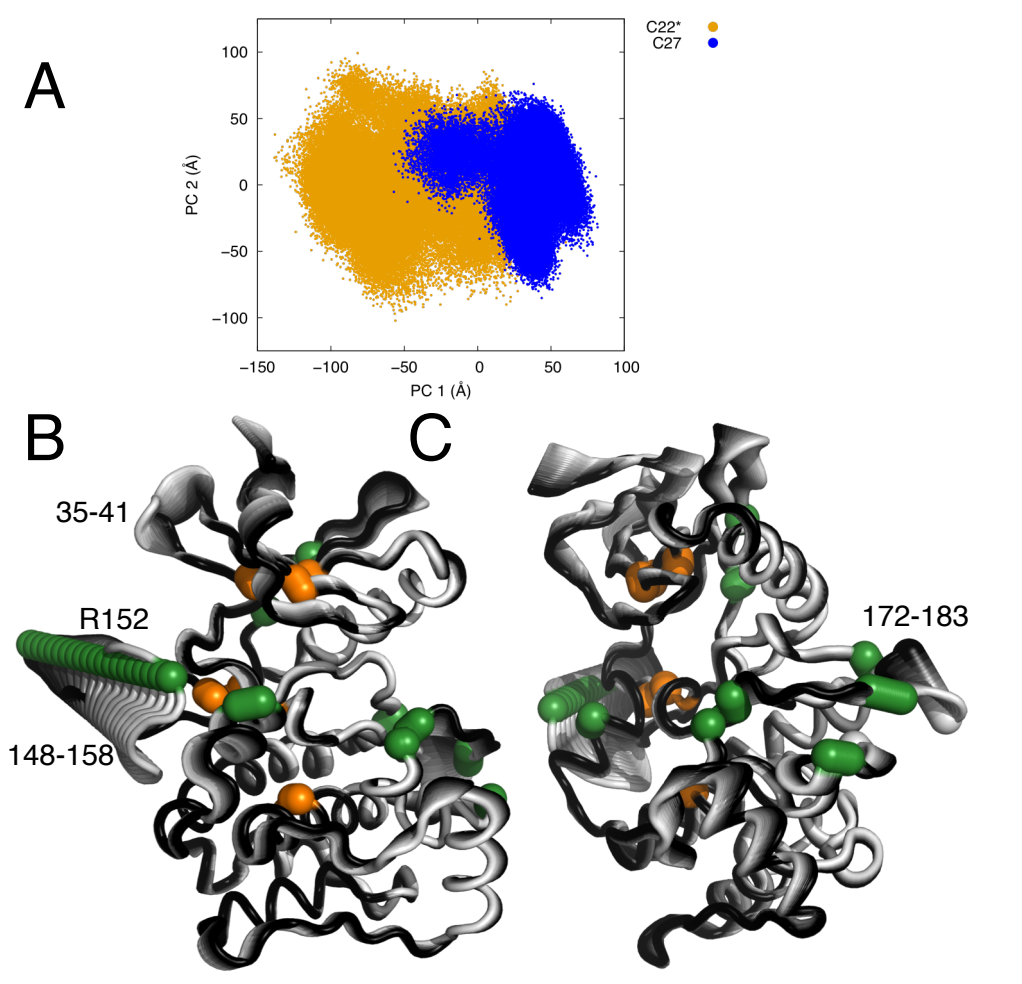
A


logFC
0
−1
−2
BLCA
brain
BRCA
COAD
GBM
HNSC
KIRC
LUAD
LUSC
ovary
PRAD
skin
STAD
THCA
UCEC
gene
ULK1
ULK2

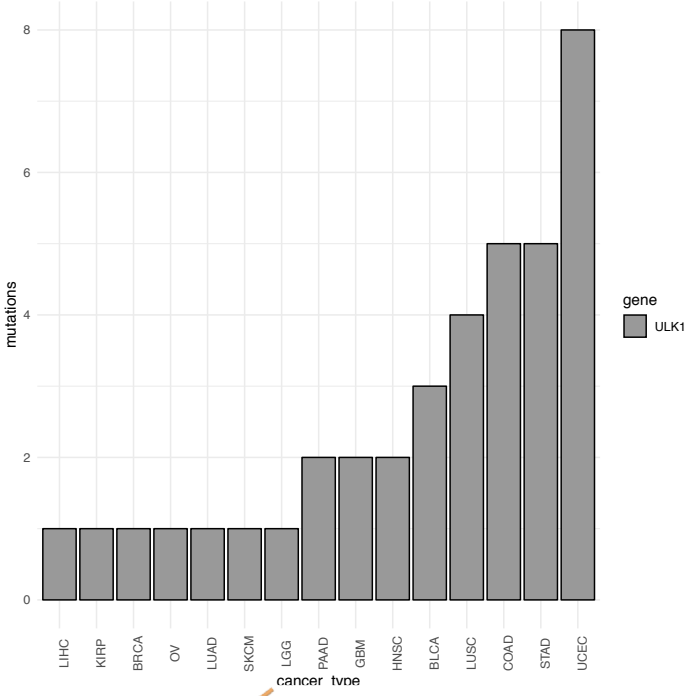
B


logFC
1
0
−1
−2
a.gland
cervix
CESC
ESCA
KICH
KIRP
LIHC
PAAD
pancreas
PCPG
pleura
READ
testis
thymus
uterus
gene
ULK1
ULK2

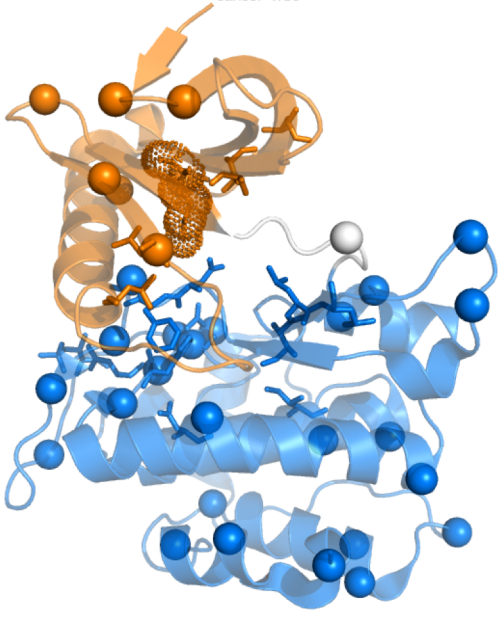

A
DFG
APE
B
K46
M92
E63
DFG
HRD
C
L67
L78
H136
F166
D203
D
L21,V30, A44
I210,C214
144-146

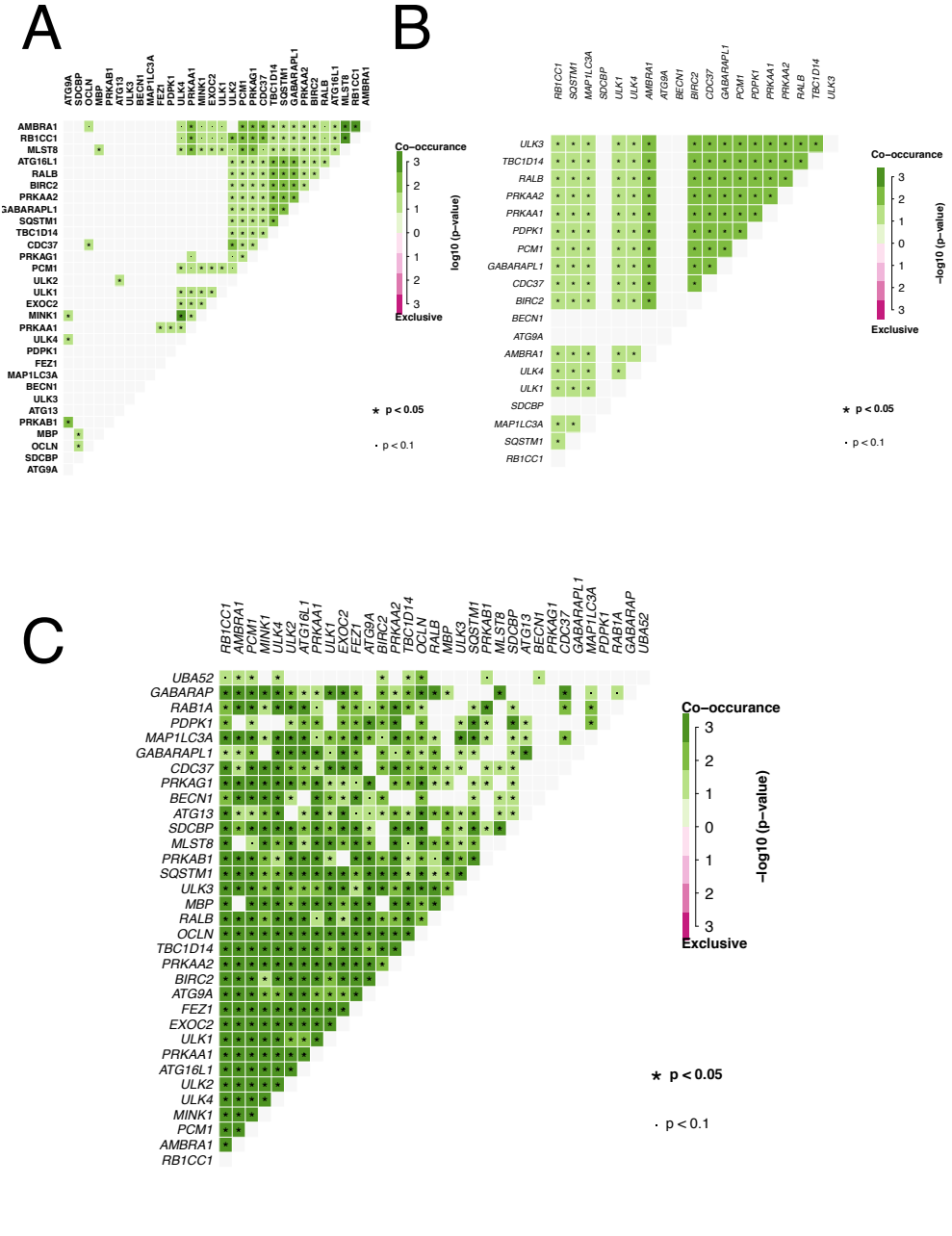

A
C22*
C27
PC 2 (Å)
PC 1 (Å)
100
50
0
−50
−100
−150
B
C
35-41
R152
148-158
172-183

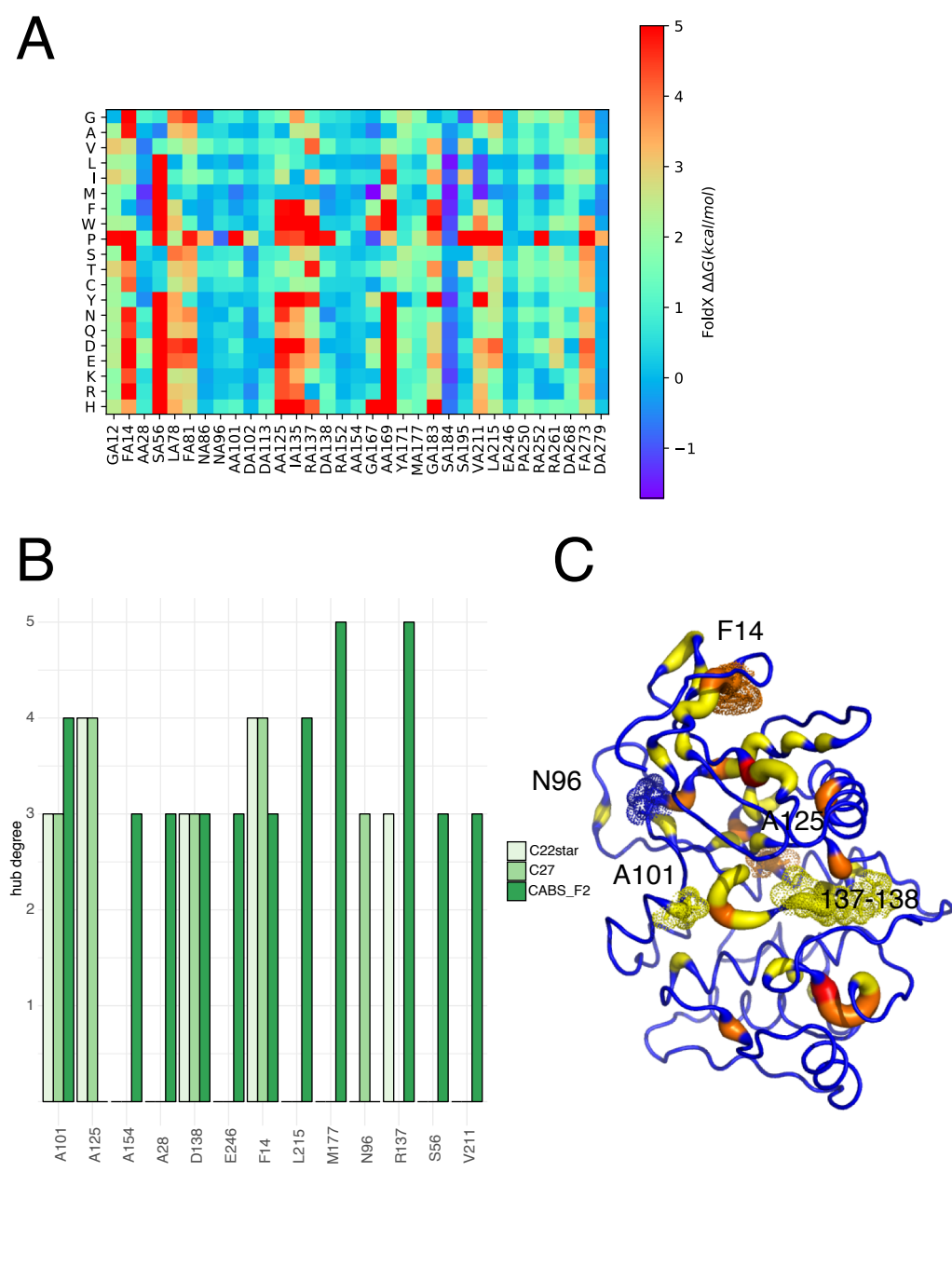
A


mutations
0
2
4
6
8
LIHC
KIRP
BRCA
OV
LUAD
SKCM
LGG
PAAD
GBM
HNSC
BLCA
LUSC
COAD
STAD
UCEC
cancer_type
gene
ULK1

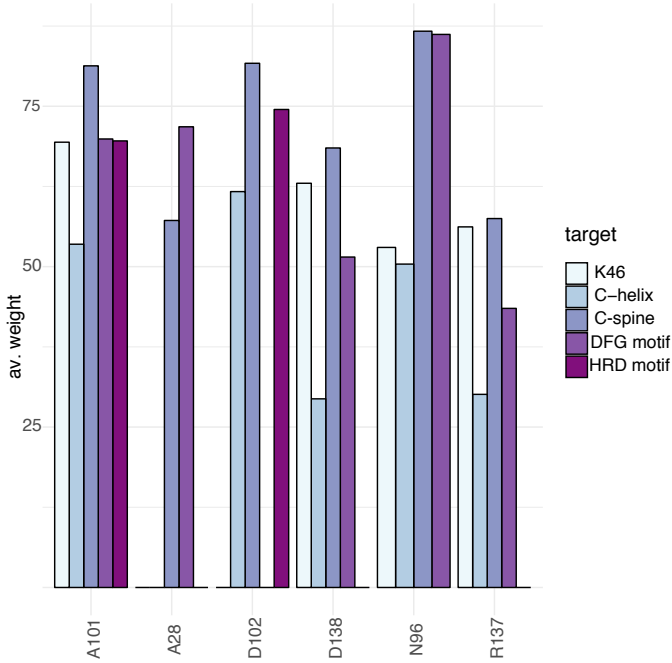
B



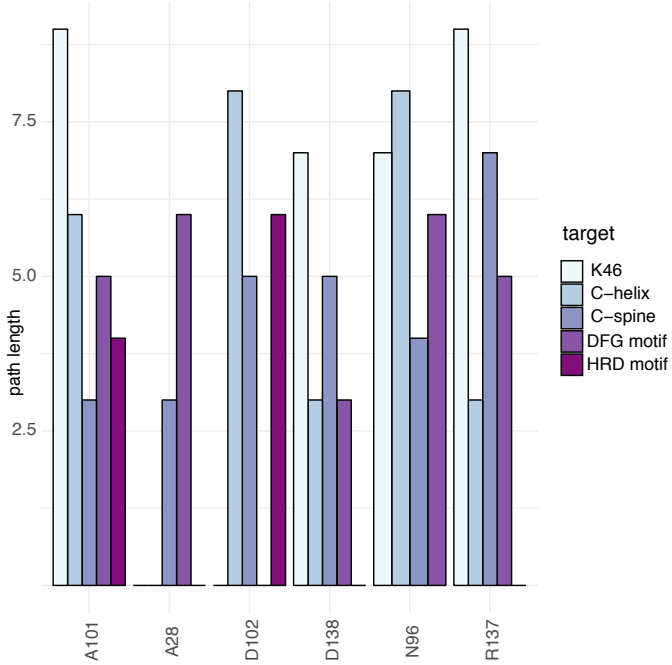

A
B
C
Co−occurance
Co−occurance
−log10 (p−value)
log10 (p−value)
Exclusive
Exclusive
★ p < 0.05
· p < 0.1
★ p < 0.05
· p < 0.1
★ p < 0.05
· p < 0.1

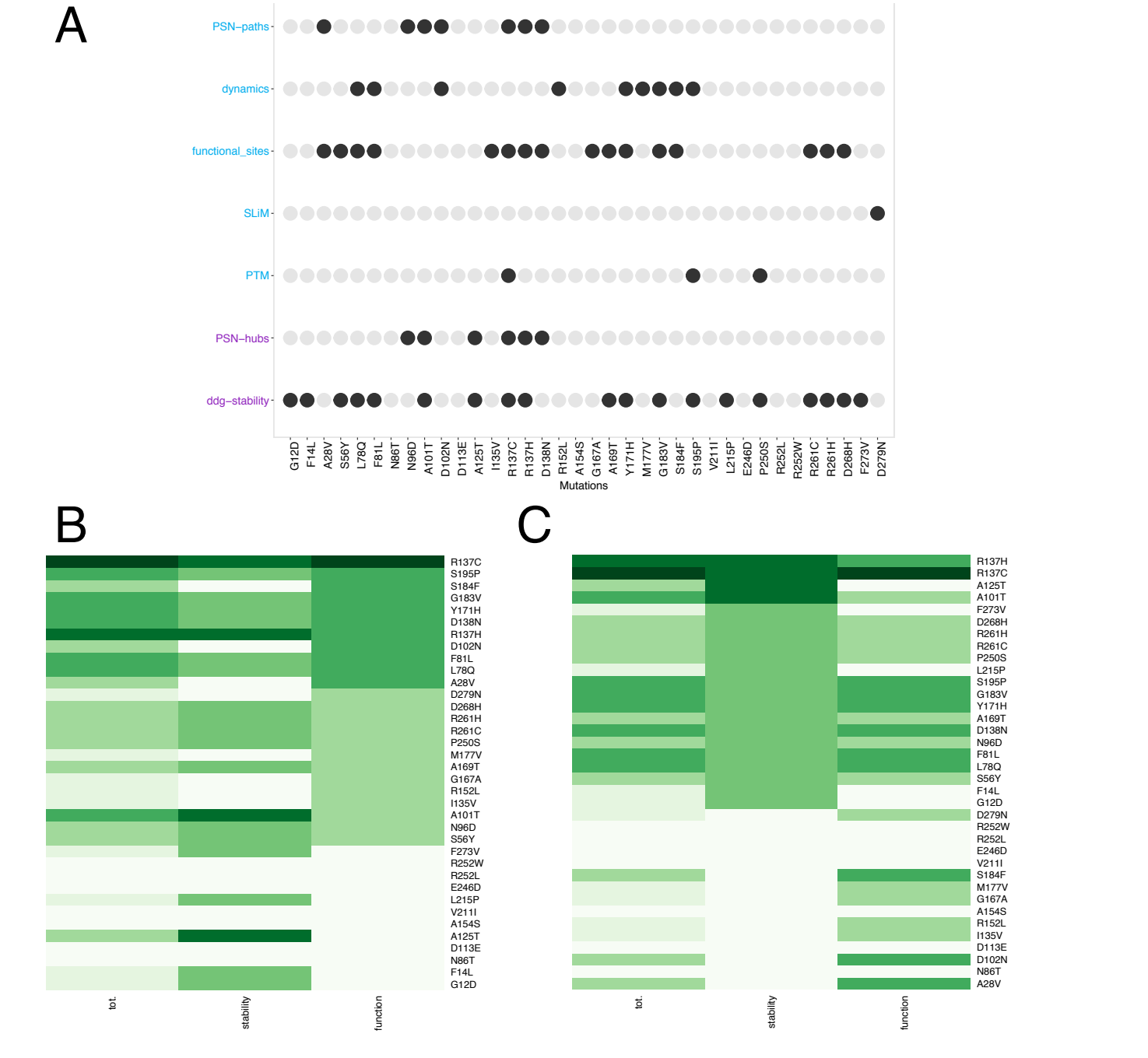

A
G
A
V
L
I
M
F
W
P
S
T
C
Y
N
Q
D
E
K
R
H
GA12
FA14
AA28
SA56
LA78
FA81
NA86
NA96
AA101
DA102
DA113
AA125
IA135
RA137
DA138
RA152
AA154
GA167
AA169
YA171
MA177
GA183
SA184
SA195
VA211
LA215
EA246
PA250
RA252
RA261
DA268
FA273
DA279
5
4
3
2
1
0
−1
FoldX ΔΔG(kcal/mol)
B
hub degree
5
4
3
2
1
A101
A125
A154
A28
D138
E246
F14
L215
M177
N96
R137
S56
V211
C22star
C27
CABS_F2
C
F14
N96
A125
A101
137-138

A


75
50
25
av. weight
A101
A28
D102
D138
N96
R137
target
K46
C−helix
C-spine
DFG motif
HRD motif

B


7.5
5.0
2.5
path length
A101
A28
D102
D138
N96
R137
target
K46
C−helix
C−spine
DFG motif
HRD motif


A
PSN−paths
dynamics
functional_sites
SLiM
PTM
PSN−hubs
ddg−stability
G12D
F14L
A28V
S56Y
L78Q
F81L
N86T
N96D
A101T
D102N
D113E
A125T
I135V
R137C
R137H
D138N
R152L
A154S
G167A
A169T
Y171H
M177V
G183V
S184F
S195P
V211I
L215P
E246D
P250S
R252L
R252W
R261C
R261H
D268H
F273V
D279N
Mutations
B
R137C
S195P
S184F
G183V
Y171H
D138N
R137H
D102N
F81L
L78Q
A28V
D279N
D268H
R261H
R261C
P250S
M177V
A169T
G167A
R152L
I135V
A101T
N96D
S56Y
F273V
R252W
R252L
E246D
L215P
V211I
A154S
A125T
D113E
N86T
F14L
G12D
tot.
stability
function
C
R137H
R137C
A125T
A101T
F273V
D268H
R261H
R261C
P250S
L215P
S195P
G183V
Y171H
A169T
D138N
N96D
F81L
L78Q
S56Y
F14L
G12D
D279N
R252W
R252L
E246D
V211I
S184F
M177V
G167A
A154S
R152L
I135V
D113E
D102N
N86T
A28V
tot.
stability
function