

Computational analysis of protein–protein interactions of cancer drivers in renal cell carcinoma

Jimin Pei^{1,2,3} , Jing Zhang^{1,2,3} and Qian Cong^{1,2,3}

1 Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA

2 Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA

3 Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA

Keywords

cancer drivers; chromatin remodeling;
protein–protein interaction; renal cell
carcinoma; ubiquitination

Correspondence

Q. Cong, Eugene McDermott Center for
Human Growth and Development,
University of Texas Southwestern Medical
Center, Dallas, TX, 75390 USA
E-mail: qian.cong@utsouthwestern.edu(Received 16 June 2023, revised 30
October 2023, accepted 6 November 2023)

doi:10.1002/2211-5463.13732

Edited by Cláudio Soares

Renal cell carcinoma (RCC) is the most common type of kidney cancer with rising cases in recent years. Extensive research has identified various cancer driver proteins associated with different subtypes of RCC. Most RCC drivers are encoded by tumor suppressor genes and exhibit enrichment in functional categories such as protein degradation, chromatin remodeling, and transcription. To further our understanding of RCC, we utilized powerful deep-learning methods based on AlphaFold to predict protein–protein interactions (PPIs) involving RCC drivers. We predicted high-confidence complexes formed by various RCC drivers, including TCEB1, KMT2C/D and KDM6A of the COMPASS-related complexes, TSC1 of the MTOR pathway, and TRRAP. These predictions provide valuable structural insights into the interaction interfaces, some of which are promising targets for cancer drug design, such as the NRF2-MAFK interface. Cancer somatic missense mutations from large datasets of genome sequencing of RCCs were mapped to the interfaces of predicted and experimental structures of PPIs involving RCC drivers, and their effects on the binding affinity were evaluated. We observed more than 100 cancer somatic mutations affecting the binding affinity of complexes formed by key RCC drivers such as VHL and TCEB1. These findings emphasize the importance of these mutations in RCC pathogenesis and potentially offer new avenues for targeted therapies.

Kidney cancer ranks among the 10 most prevalent cancers in the United States [1] and caused about 180 000 deaths globally in 2020 [2]. Cases of the most common kidney cancer, renal cell carcinoma (RCC), have been on the rise in recent years [3,4]. While early diagnosis and intervention by urological procedures could successfully treat RCCs, the overall prognosis

remains poor, particularly for late-stage patients. Detection of early-stage RCCs is hindered by the lack of noticeable clinical symptoms in most cases. Furthermore, traditional drugs for RCCs, such as tyrosine kinase inhibitors (TKIs), have demonstrated limited efficacy. RCCs comprise a heterogeneous group of kidney tumors, with clear cell RCC (ccRCC) accounting

Abbreviations

ccRCC, clear cell renal cell carcinoma; chRCC, chromophobe RCC; COMPASS, COMplex of Proteins Associated with SET1; CUL3, Cullin 3; HIF- α , α -subunit of hypoxia-inducible factor; HIF- β , β -subunit of hypoxia-inducible factor; KEAP1, Kelch-like ECH-associated protein 1; KMT2, type 2 lysine methyltransferase; MTOR, mechanistic target of rapamycin; NCG, Network of Cancer Genes; NFE2L2, nuclear factor (erythroid-derived 2)-like 2; PAE, Predicted Aligned Error; PDB, Protein Data Bank; PIKK, phosphatidylinositol 3-kinase-related kinase; pLDDT, predicted Local Distance Difference Test; PMID, PubMed Identifier; PPI, protein–protein interaction; pRCC, papillary RCC; RCC, renal cell carcinoma; SAGA, Spt-Ada-Gcn5 acetyltransferase complex; SLiM, short linear motif; TKI, tyrosine kinase inhibitor; TRRAP, Transactivation/Transformation-domain Associated Protein; TSG, tumor suppressor gene.

for 70–80% of cases, papillary RCC accounting for 10–15% of cases, and chromophobe RCC accounting for 3–5% of cases [5]. Precision medicine has made significant progress in targeted therapy of RCCs by differentiating between these subtypes in genomic and pathway contexts [3,6]. It is crucial to advance our understanding of the molecular mechanisms underlying RCC, as the knowledge could pave the way for innovations in diagnosis and treatment strategies.

The development and progress of cancer often result from the deregulation of multiple cellular processes that leads to unrestricted proliferation and replication potentials, insensitivity and resistance to antigrowth signals and apoptosis, promotion of angiogenesis, and the capability of invasion and metastasis [7]. Protein complexes formed by protein–protein interactions (PPIs) play critical roles in diverse aspects of cellular processes [8], including those related to cancer. Targeting cancer-enabling PPIs has emerged as a promising strategy in drug development, aiming to disrupt these interactions and hinder tumor growth [9–11]. While large-scale high-throughput experimental studies have generated large datasets of human PPIs at the whole proteome level [12–14] or on cancer-related proteins [15,16], these datasets often suffer from high rates of false positives and false negatives with considerable inconsistencies among them [17,18].

Cancer is often driven by genetic alterations that activate oncogenes or deactivate tumor suppressor genes [19]. These alterations can manifest as structural changes in chromosomes, gene fusion events, changes in gene copy numbers, and mutations at specific nucleotide positions. Next-generation sequencing has been employed to identify genetic differences between cancer and normal cells [20]. Cancer genomic studies have expanded the landscape of cancer driver genes and their genetic alterations and advanced our understanding of the molecular mechanisms of cancer growth, metastasis, and drug resistance [21–25]. Detection of disease-causing genetic variations or somatic mutations in cancer cells by genome sequencing provides valuable information for making clinical decisions for cancer patients [26]. A significant fraction of germline and somatic mutations occurs at the PPI interfaces [27,28]. These mutations can have profound functional and clinical implications by disrupting protein complexes and interrupting essential cellular processes such as signal transduction and protein degradation [29,30]. Accurate prediction of PPI interfaces holds tremendous potential in interpreting the clinical significance of cancer mutations and identifying targets for cancer drug development. By understanding how these mutations affect protein interactions, we can gain deeper

insights into the underlying mechanisms of cancer and identify novel therapeutic avenues.

Deep-learning methods have revolutionized our ability to predict high-resolution protein structures [31] and model protein–protein interactions [32]. Methods such as AlphaFold [31] and RoseTTAFold [33] have been used to identify PPIs and model the structures of their interfaces with an accuracy better than large-scale experimental studies [34]. These deep-learning methods can be applied to screen confident PPIs from noisy high-throughput candidate PPIs identified in large-scale studies and obtain high-quality structure models for the protein complexes [32,34,35].

In this study, we investigate PPIs involving RCC driver proteins and the effects of cancer somatic mutations at the PPI interfaces. We collected a comprehensive set of RCC drivers from publicly available databases [36,37]. AF2Complex [38], FoldDock [39], and an in-house program [40–42], all based on AlphaFold, were utilized to predict protein complexes formed by these cancer drivers. These methods predicted hundreds of PPIs formed by various RCC drivers and offered novel structural insights into protein complexes with important roles in RCC progression, including complexes involving the protein interaction hub protein TCEB1, the KMT2C/D COMPASS complex, the TRRAP complexes, and the TSC complex of the MTOR pathway. We also identified RCC somatic mutations occurring at the interfaces of predicted and experimental structures of the protein complexes formed by RCC drivers. We evaluated their functional impact in terms of binding affinity changes. Cancer mutations affecting the binding affinity of protein complexes were mostly observed in several key RCC drivers, such as VHL and TCEB1, while few deleterious missense mutations were observed in other RCC drivers, such as those involved in chromatin remodeling.

Results and Discussion

Functional categories of RCC drivers

We collected a set of canonical and candidate RCC driver proteins from two resources: the Network of Cancer Genes (NCG) [36] and OncoVar [37] (see [Materials and methods](#)). Canonical cancer drivers mainly fall into several functional categories: protein degradation and ubiquitination, chromatin remodeling, transcription, MTOR pathway, signaling, cell cycle, and immunity (Table 1). The functional category of chromatin remodeling includes the largest number of NCG canonical RCC cancer drivers (bold names in Table 1). They

Table 1. Functional categories of RCC drivers. These drivers are classified in the NCG and OncoVar databases. NCG canonical cancer drivers are in bold. OncoVar cancer drivers are underlined. NCG candidate cancer drivers are in regular font. Oncogenes and tumor suppressors classified in NCG or OncoVar are labeled by (o) and (s), respectively. These drivers are classified according to broad functional categories.

Functional category	RCC drivers
Protein ubiquitination and degradation	VHL (s), BAP1 (s), CUL3 (s), KEAP1 (s), TCEB1/ELOC, CUL7, ASB15, WDFY3
Chromatin remodeling	ARID1A (s), ARID2 (s), KDM5C (s), KDM6A (s), KMT2C (s), KMT2D (s), PBRM1 (s), SETD2 (s), SMARCB1 (s), DNMT3A(o), CHD3
Transcription	CRTC1 (o), ERCC2 (s), NF2 (s), NFE2L2 (o), TP53 (s), TRRAP (o), HIF1A (o), RB1 (s), MAML2 (o), MAX (s), NONO , SFPQ(s), SPEN(s), TFE3(o), TFEB(o), CNOT1, MED13, TRIM37, ZFXH4, ZNF469, ZNF765, ZNF804A
MTOR pathway	FLCN (s), MTOR (o), TSC1 (s), TSC2 (s), TMEM127(s), RHEB
Other signaling pathways	FAT1 (s), FAT3 , LRP1B (s), MET (o), PIK3CA (o), PTEN (s), PTK6(o), TRIO
Cell cycle	ATM (s), STAG2 (s), CDKN1A (s), PRCC , CCNB2
Immunity	CARD11 (o), BTNL3, NAV3
Others	CLTC(s), FH(s), HNRNPM, NEFH, AHNK, AKAP13, AKAP9, ANPEP, C2CD4C, CCDC168, COL11A1, CPQ, CSMD3, CUBN, FAAH2, FAM111B, FMN2, GABRB2, LRRK2, MICALCL, MSR1, NPNT, PDHB, PDXDC1, PIEZO2, PLEC, PNKD, RADIL, RYR1, SHANK1, SLC1A3, SLC9A3R1, SLITRK6, SPTBN4, SYNE2, TXNIP

include large, modular proteins with enzymatic domains that perform post-translational modifications of histones such as SETD2, KDM5C, KDM6A, KMT2C, and KMT2D, as well as nonenzymatic subunits of large chromatin remodeling complexes such as PBRM1, ARID1A, ARID2, and SMARCB1 [43]. Several canonical RCC drivers (BAP1, CUL3, KEAP1, and VHL) are involved in ubiquitin-related pathways such as protein degradation. In particular, *VHL* is the most frequently mutated gene in ccRCC (the dominating subtype of RCC) [44]. Four NCG canonical cancer drivers (FLCN, MTOR, TSC1, and TSC2) and two OncoVar cancer drivers (TMEM127 and RHEB) are involved in the MTOR pathway that regulates metabolism and is frequently perturbed in RCCs [45]. The functional categories of transcription and signaling are two other major contributors of NCG canonical cancer drivers of RCCs, while the functional categories of cell cycle and immunity are less represented in RCCs (Table 1). One characteristic of RCCs compared with other cancer types is that most canonical drivers of RCCs are encoded by tumor suppressor genes (TSGs, marked by (s) in Table 1). Only seven NCG canonical drivers are encoded by oncogenes (marked by (o) in Table 1), including three protein/lipid kinases—MTOR, MET, and PIK3CA.

Comparison of AlphaFold-based methods of PPI predictions for RCC drivers

We applied three protein complex prediction methods (AF2Complex [38], FoldDock [39], and an in-house program based on the contact probability of AlphaFold,

namely AF-contact [40–42]) to predict the protein–protein interactions involving RCC drivers. All these methods were developed based on AlphaFold, and they mainly differ in the evaluation of complex structures. Our method, AF-contact, relies on the inter-residue distograms (probability distribution over a series of distance bins) produced by AlphaFold, and the sum of probability in bins corresponding to $< 12 \text{ \AA}$ is used as the inter-residue contact probability. AF-contact utilizes the intermolecular contact probability for ranking the protein complex models (see Materials and methods). AF2Complex relies on an interface score that combines PAE (Predicted Aligned Error) estimates of interface residues and the size of the interface. FoldDock has a scoring metric that combines the interface pLDDT (predicted Local Distance Difference Test) score and the number of interface contacts.

We applied these methods to the 3595 candidate PPIs involving RCC drivers (see below), and we compared their performance by examining the fraction of pairs with experimentally determined structures in PDB among the top-ranking predictions. While the set of candidate PPIs may still contain a considerable fraction of false PPIs, those supported by experimental structures are highly confident and can be used as positive controls. We ranked the predicted PPIs by each method's PPI confidence metric (contact probability for AF-contact, interface score for AF2Complex, and pDockQ for FoldDock), and the numbers of predicted PPIs by each method that are supported by PDB structures at each rank, up to top 1000, are shown in Fig. 1A. These curves show that the three programs perform similarly for top-scoring pairs. Among the top

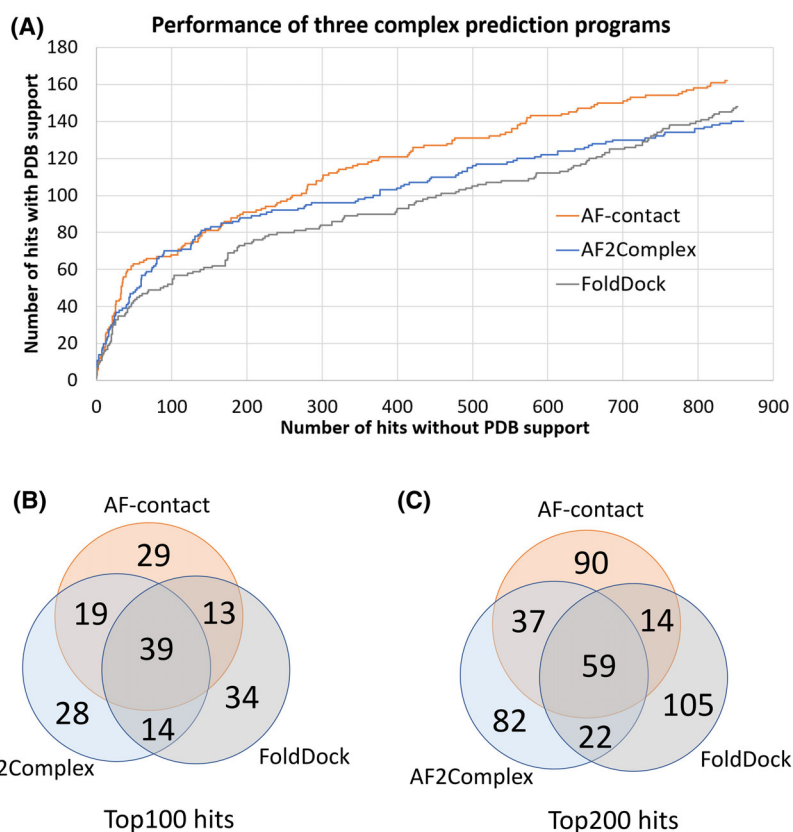


Fig. 1. Comparison of protein complex prediction by three programs (AF-contact, AF2Complex, and FoldDock). (A) The numbers of predicted PPIs supported by experimental structures in PDB among the top 1000 predictions of the three programs. (B) Venn diagram of the top 100 predicted PPIs by the three programs. (C) Venn diagram of the top 200 predicted PPIs by the three programs.

50 predicted PPIs, AF-contact, AF2Complex, and FoldDock have 30, 30, and 29 pairs supported by experimental structures, respectively. Among the top 100 predictions, AF-contact has 60 pairs supported by PDB structures, while AF2Complex and FoldDock have 49 and 48 such pairs, respectively. The number of pairs with experimental structures grows much slower after the top 100 hits (Fig. 1A). AF-contact performs slightly better as the number of PDB-supported predictions is mostly higher than the other two methods at different ranks.

We also compared the top-scoring PPIs predicted by the three programs, revealing a high consistency between them. For example, 39 PPIs (~40%) were predicted by all three programs among their top 100 predictions (Fig. 1B). Approximately one-third of the top 100 predicted PPIs generated by each program was unique and did not overlap with the top 100 predictions of the other two programs. The consistency between the programs decreased when more predictions were included in the comparison (Fig. 1C). Among the top 200 predictions, only 59 (~30%) were consistently found by all three pairs. AF-contact and AF2Complex appear to show more consistent results with each other and share more top predictions.

Prediction of PPIs for RCC drivers and assessment of mutation effects

Candidate human PPIs involving RCC drivers were obtained from the BioGRID database [46]. Since most of these PPIs were identified from large-scale experiments prone to false positives [17,18], we selected 3595 candidate PPI pairs involving RCC drivers based on publication-weighted scores (see Materials and methods). Our method AF-contact predicted high-confidence models for 211 pairs of proteins (299 domain pairs with contact probability > 0.9). Among those, 85 protein pairs are supported by experimental structures. The majority of these predictions are supported with high confidence by the other two programs: AF2Complex and FoldDock (Table S1). Most of the examples discussed below are supported by two or more programs.

We mapped the RCC somatic mutations to predicted and experimentally determined PPI interfaces involving RCC drivers. These mutations are from the COSMIC database [23] and a large collection of somatic mutations from 148 RCC patients [47]. The effects of mutations found at PPI interfaces were accessed by BindProfX with the FoldX as the force field [48,49]. Predicted free energy change associated

with each mutation at the PPI interface is provided in Table S2, and mutations predicted to destabilize protein complexes remarkably ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$) are highlighted in yellow. These predictions were used to interpret the functional impact of cancer somatic mutations discussed below.

VHL mutations affect binding affinities in multiple protein complexes

Biallelic inactivation of the tumor suppressor gene *VHL* is a hallmark in most ccRCCs [50]. *VHL* inactivation is often due to the loss of chromosome 3p and point mutations. VHL is the substrate-binding subunit of an E3 ubiquitin ligase that recognizes and targets the α -subunit of hypoxia-inducible factor (HIF- α) for protein degradation in physiological conditions with normal oxygen level [51]. In hypoxic conditions, HIF- α is not hydroxylated or targeted by VHL for degradation. The stable HIF- α forms a dimer with the β -subunit (HIF- β , also called ARNT) that functions as a transcription factor to promote the expression of target genes involving many cellular processes in adaptation to hypoxic conditions such as glycolysis, glucose transport, and angiogenesis. These functions support the critical role of VHL in regulating tumor progression in ccRCC.

VHL forms protein complexes with TCEB1 (ELOC), ELOB, and CUL2. While the VHL-TCEB1 interaction was detected by AlphaFold, the VHL-CUL2 interaction was not detected. We mapped several deleterious mutations of VHL to the interface of the experimentally determined structure of the VHL-TCEB1-CUL2 complex [52] (PDB id: 4wqo, Fig. 2A). These mutations are located in the small helical SOCS-box domain of VHL.

Experimental structures have been determined for the complexes of VHL-HIF1 α and VHL-EPAS1

(HIF2 α) [53,54]. Both HIF1 α and EPAS1 utilize a short and extended segment with a hydroxylated proline to interact with the immunoglobulin domain of VHL at similar locations (Fig. 2B,C). AlphaFold did not detect the interactions of VHL-HIF1 α and VHL-EPAS1 above the probability score cutoff (0.9), possibly due to the small interfaces and the inability of AlphaFold to incorporate modified residues (hydroxylated prolines) in the modeling process. Several deleterious VHL mutations were mapped in the VHL-HIF1 α interface (PDB id: 1lm8) and the VHL-EPAS1 interface (PDB id: 6i7q; Fig. 2B,C) [53,54].

TCEB1 mutations in multiple PPI interfaces

TCEB1 (ELOC, elongin C) is a critical component of the VHL complex that targets HIF- α for degradation. Recently, a distinct group of kidney cancers characterized by the loss of heterozygosity at the *TCEB1* location in chromosome 8 and hotspot mutations in *TCEB1* were identified by integrated sequence analysis [55]. These tumors exhibited different morphological and immunohistochemical characteristics compared with ccRCC tumors caused by *VHL* alterations. TCEB1 is a SOCS-box binding protein [56]. Indeed, AlphaFold predicted the interactions between TCEB1 and numerous SOCS-box-containing proteins, such as VHL, SOCS1, ELOA, ASB1, PRAME1, and WSB1 (8 proteins shown in Fig. 3). The SOCS box is a short helical domain found in various proteins with different domain contents. Interaction through the SOCS box facilitates targeting these proteins by the adapter TCEB1 to E3 ubiquitin ligases for degradation [56]. While the VHL-TCEB1 interaction has been well studied, the roles of TCEB1 interactions with other SOCS-box-containing proteins have not been fully explored. Deleterious RCC mutations in TCEB1 such as

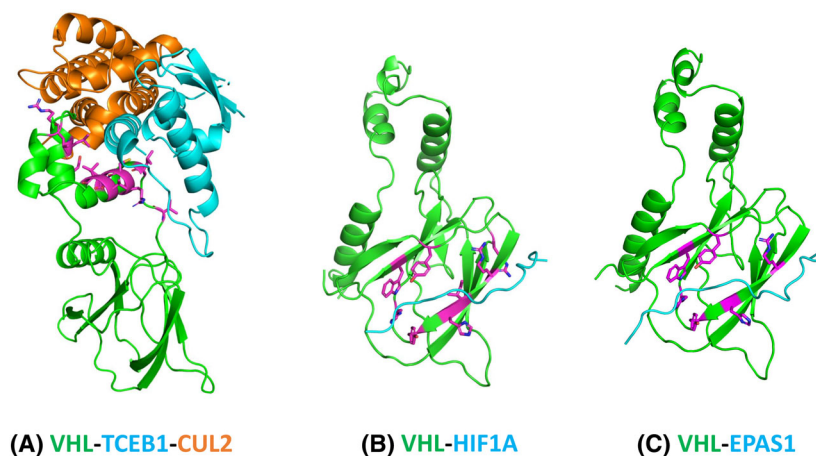


Fig. 2. Protein–protein interactions of VHL. (A) The VHL-TCEB1-CUL2 complex (based on PDB id: 4wqo). (B) The VHL-HIF1A complex (based on PDB id: 1lm8). (C) The VHL-EPAS1 complex (based on PDB id: 6i7q). Predicted destabilizing interface mutations ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$) on VHL were shown in magenta with the sidechains displayed in sticks.

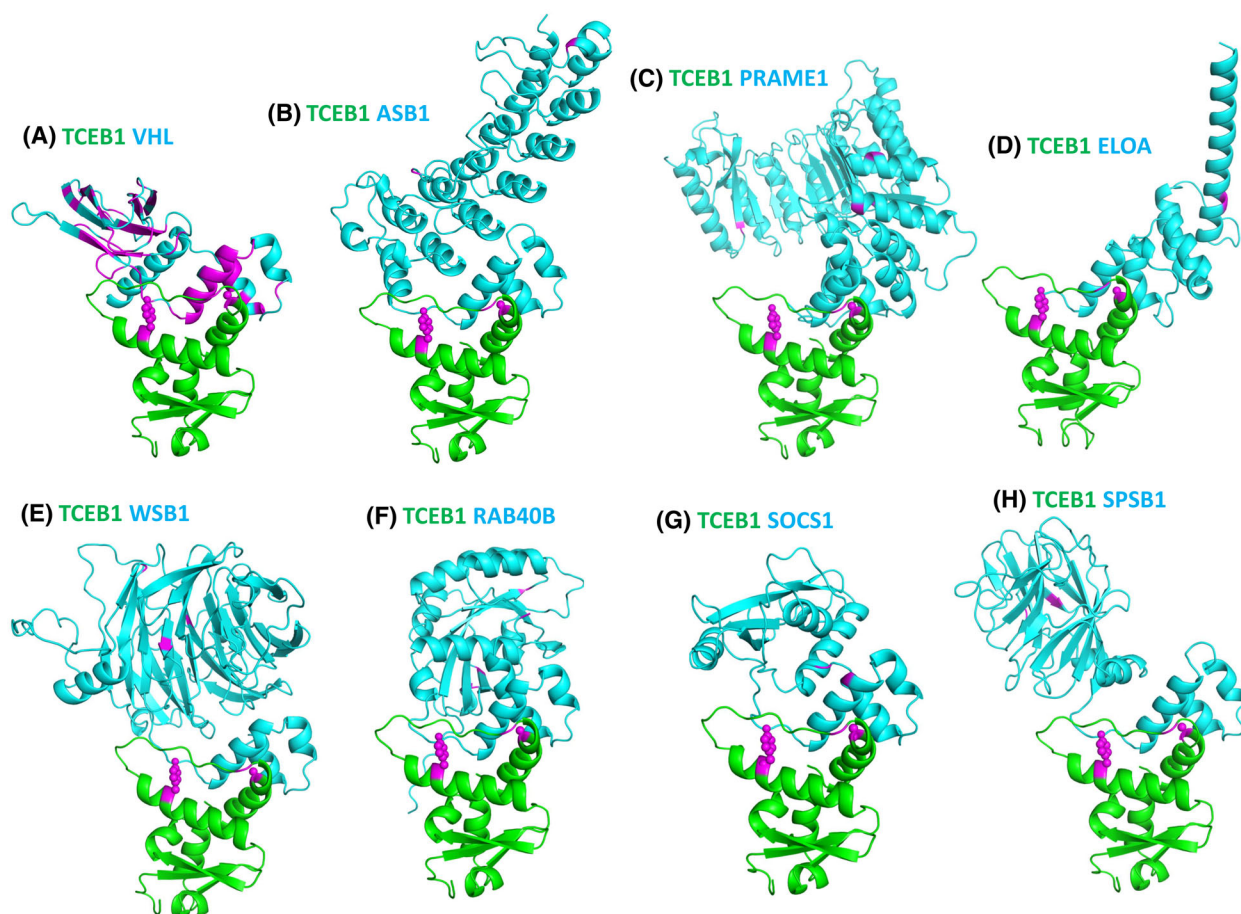


Fig. 3. Protein–protein interactions of TCEB1 predicted by AlphaFold. (A) The TCEB1-VHL complex. (B) The TCEB1-ASB1 complex. (C) The TCEB1-PRAME1 complex. (D) The TCEB1-ELOA complex. (E) The TCEB1-WSB1 complex. (F) The TCEB1-RAB40B complex. (G) The TCEB1-SOCS1 complex. (H) The TCEB1-SPSB1 complex. RCC mutations in these proteins were shown in magenta. Sidechains of predicted destabilizing interface mutations ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$) on TCEB1 were shown in spheres.

Y79C/S/F/N and A100P are mapped to interfaces of these complexes (Fig. 3), suggesting that these mutations could interfere with the degradation of various TCEB1 targets besides VHL. The specific roles of various TCEB1-interacting SOCS-box binding proteins in RCC progression remain to be studied.

Structural insight into the COMPASS-related complexes

RCC drivers KMT2C and KMT2D belong to the family of type 2 lysine methyltransferases (KMT2) involved in histone modification and chromatin remodeling [57]. They perform post-translational methylation of histone 3 lysine 4 (H3K4), which is frequently associated with active transcription. KMT2C (4911 amino acid residues) and KMT2D (5537 amino acid residues) are large, modular proteins with methyltransferase activity driven by a

conserved SET domain and contain a large fraction of disordered regions. They are part of the COMPASS (COMplex of Proteins Associated with SET1)-related complexes conserved from fungi to metazoans [58,59]. The other subunits of the COMPASS complexes include NCOA6 (also named ASC2), KDM6A (also named UTX), PAXIP1 (also named PTIP), PAGR1, NCOA1, WDR5, RBBP5, ASHL2, and DPY-30. KDM6A, a demethylase, is another canonical cancer driver of ccRCC [60].

AlphaFold detected multiple interactions involving short linear motifs (SLiMs) in disordered regions of KMT2C/KMT2D and several other COMPASS-related complex subunits such as KDM6A, NCOA6, and PAXIP1 (Fig. 4). Three SLiMs in KMT2C (residues 1167–1196, 1349–1388, and 4275–4298) interact with the demethylase domain of KDM6A (Fig. 4), suggesting that KMT2C could regulate the enzymatic

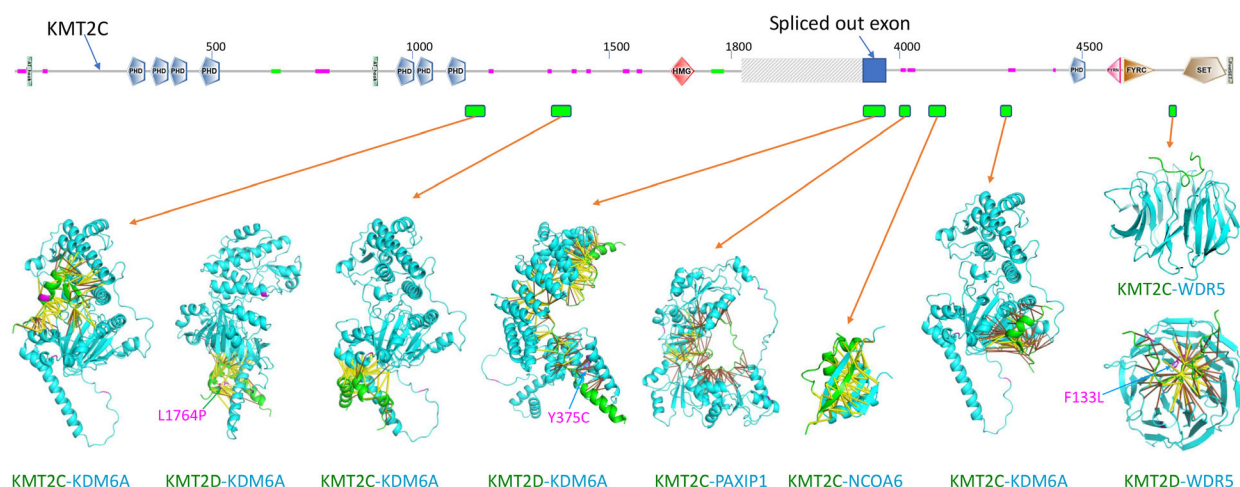


Fig. 4. Protein–protein interactions of KMT2C/D in the COMPASS-related complex. The domain diagram of KMT2C is shown with globular domains labeled according to the SMART database. The long and disordered region is omitted and replaced by a crossbar. The extra exon spliced out in the canonical form of KMT2C is shown as a blue rectangular box. The interaction regions are labeled by green boxes. KMT2C/D regions in the structural models were shown in green. The interaction partners of KMT2C/D are shown in cyan. Positions containing RCC missense mutations are shown in magenta. Contacts with scores above 0.9 are shown in yellow sticks, and contacts with scores between 0.5 and 0.9 are shown in orange sticks.

activity of KDM6A. A SLiM in KMT2C (residues 4073–4120) forms a tightly packed small globular domain with residues 88–139 of NCOA6. Another SLiM in KMT2C (residues 3995–4040) interacts with four C-terminal BRCT domains of the COMPASS complex subunit PAXIP1 (residues 589–1069). In addition, KMT2C uses a C-terminal SLiM (residues 4703–4716) to interact with the WD40 domain of WDR5, where the complex structure has been solved [61].

A region (residues 4543–4618) in KMT2D, a paralog of KMT2C, is predicted to interact with the N-terminal helical repeat domain of KDM6A (residue 38–405). Interestingly, the corresponding homologous region in KMT2C is encoded by an exon (shown in the KMT2C domain diagram in Fig. 4) that is spliced out in the canonical isoform of KMT2C (UniProt accession: [Q8NEZ4](#)). The protein segment encoded by this exon is present in two other KMT2C isoforms (UniProt: [Q8NEZ4-2](#) and [Q8NEZ4-3](#), 58 residues inserted between residues 3890 and 3891 of the canonical isoform). This suggests that the activity of the COMPASS complex could be modulated by alternative splicing.

Few RCC mutations were mapped to the interaction interfaces involving KMT2C (4 mutations) and KMT2D (6 mutations) despite their large sizes (Table S2). Three such mutations in the interaction interfaces of KMT2D (KMT2D L1764P, KDM6A Y375C, and WDR5 F133L) were predicted to be destabilizing ($\Delta\Delta G > 1.4$ kcal·mol^{−1}) according to free energy change calculations (labeled in Fig. 4). Due to the extensive interfaces involving the

large subunits in the COMPASS-related complexes, single amino acid mutations may not have a significant impact in disrupting their interactions. Inactivating mutations (nonsense, frameshift, or splice site disruption) in the subunits of COMPASS-related complexes such as KMT2C, KMT2D, and KDM6A [62,63] could in general have more dominant effects in the progress of RCCs than missense mutations.

Protein complexes involving NRF2 (NFE2L2) and KEAP1

NRF2 (also named NFE2L2 for nuclear factor (erythroid-derived 2)-like 2) is a transcription factor that regulates the expression of multiple genes involved in cellular stress response [64]. NRF2 is targeted by protein degradation through its interaction with the Kelch-like ECH-associated protein 1 (KEAP1) [65], which is an adaptor protein that bridges NRF2 and Cullin 3 (CUL3)-based E3-ubiquitin ligase complex. Disruption of the NRF2-KEAP1 interaction prevents NRF2 from degradation, and upregulation of NRF2 is a major driving event in certain RCCs and other types of cancers [66,67]. NRF2, KEAP1, and CUL3 are classified as canonical RCC drivers in the NCG database. NRF2 utilizes two small SLiMs in disordered regions (residues 17–40 and residues 69–84) to interact with the beta-propeller domain of KEAP1 [68,69]. Several mutations in these regions [70], such as those involving Asp29, Leu30, and Gly31, have been found in RCCs. These mutations were predicted to

destabilize the NRF2-KEAP1 complex according to the calculation of free energy change ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$; Fig. 5). We also identified two predicted destabilizing RCC mutations in CUL3 (R59S and Y62H) in the experimentally determined interface of KEAP1 and CUL3 (Fig. 5) [71].

The transcriptional activity of NRF2 depends on the formation of the NRF2-MAFK heterodimer [72]. AlphaFold predicted the interaction between NRF2 and MAFK with high contact probability (> 0.99). The complex model of NRF2 and MAFK demonstrates the interaction between the coiled coils from the two proteins, which forms a leucine zipper heterodimer. Moreover, the C terminus of NRF2 forms a β -hairpin that also interacts with the coiled-coil of MAFK (Fig. 5). No cancer somatic mutations with deleterious effects were identified in the NRF2-MAFK interface in our datasets, reinforcing the idea that the NRF2-MAFK interaction is critical for activating gene expressions for cancer progression. The newly discovered interface involving the NRF2 C-terminal β -hairpin presents a potential target for developing anticancer drugs. By disrupting the NRF2-MAFK interaction, the transcriptional activity of NRF2 could be attenuated, offering a promising avenue for RCC drug design.

Protein complexes in the mTOR pathway

The mTOR pathway is often perturbed in chromophobe RCC [73]. MTOR (mechanistic target of

rapamycin), an evolutionarily conserved serine/threonine kinase, forms complexes such as mTORC1 and mTORC2 with other proteins to regulate cell growth and survival in response to nutritional status and various growth factors and stress signals [74]. MTOR phosphorylates downstream targets such as S6K1 to promote cell growth. Therefore, MTOR and its activator RHEB have been classified as oncogenes, while several negative regulators of the mTORC1 complex, such as TSC1, TSC2, TMEM127, and FCLN, are considered tumor suppressors [75–77].

AlphaFold predicted several complexes involving proteins in the MTOR pathway, such as MTOR-RPTOR, MTOR-MLST8, and MTOR-DEPTOR, most of which are supported by experimental structures [78]. The TSC complex, comprising TSC1, TSC2, and TBC1D7, is a major suppressor of MTOR. TSC1 possesses an N-terminal helical repeat domain, disordered regions in the middle, and a C-terminal coiled-coil domain (Fig. 6). A recent structure of TSC1 in complex with TSC2 and TBC1D7 showed that the coiled-coil region of TSC1 forms a homodimer that interacts with the elongated homodimer of TSC2 (Fig. 6). Part of the coiled-coil region (residues 936–960) of TSC1 also interacts with TBC1D7 (Fig. 6). Interestingly, AlphaFold predicted an additional interaction site between TSC1 and TSC2 with high contact probability (> 0.99): A short sequence motif from TSC1 (residues 342–360, Fig. 6) embedded in the disordered region interacts with TSC2's HEAT repeat

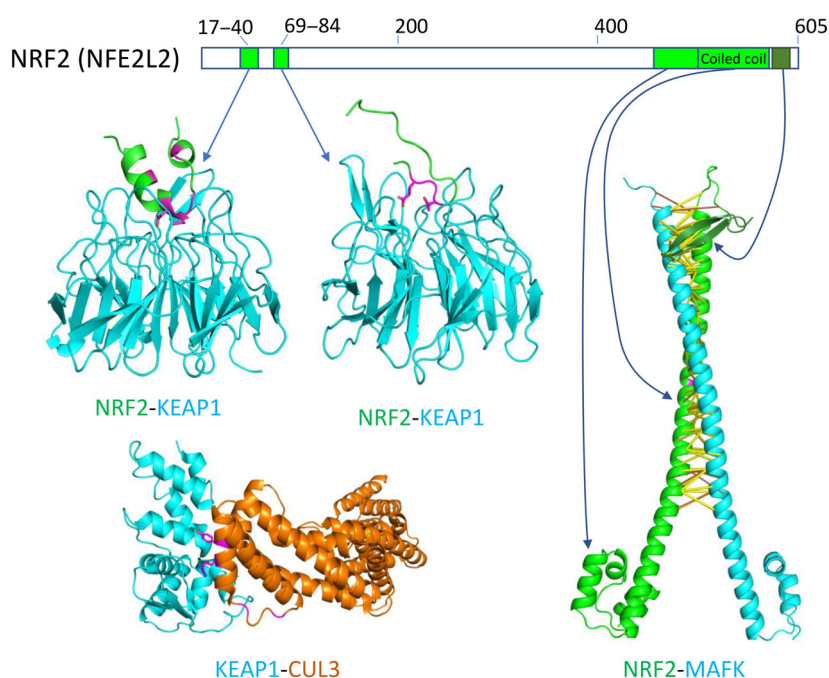


Fig. 5. Protein–protein interactions of NRF2, KEAP1, and CUL3. NRF2 regions involved in PPIs are shown in green or dark green (C-terminal β -hairpin). The NRF2-KEAP1 interactions are shown based on two experimental structures [PDB ids: 3wn7 (upper left) and 2flu (middle)]. The KEAP1-CUL3 interactions are shown based on an experimental structure [PDB id: 5nlb (lower left)]. RCC mutations in these proteins are shown in magenta. Sidechains of predicted destabilizing interface mutations ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$) are shown in sticks. Contacts with scores above 0.9 are shown in yellow sticks, and contacts with scores between 0.5 and 0.9 are shown in orange sticks.

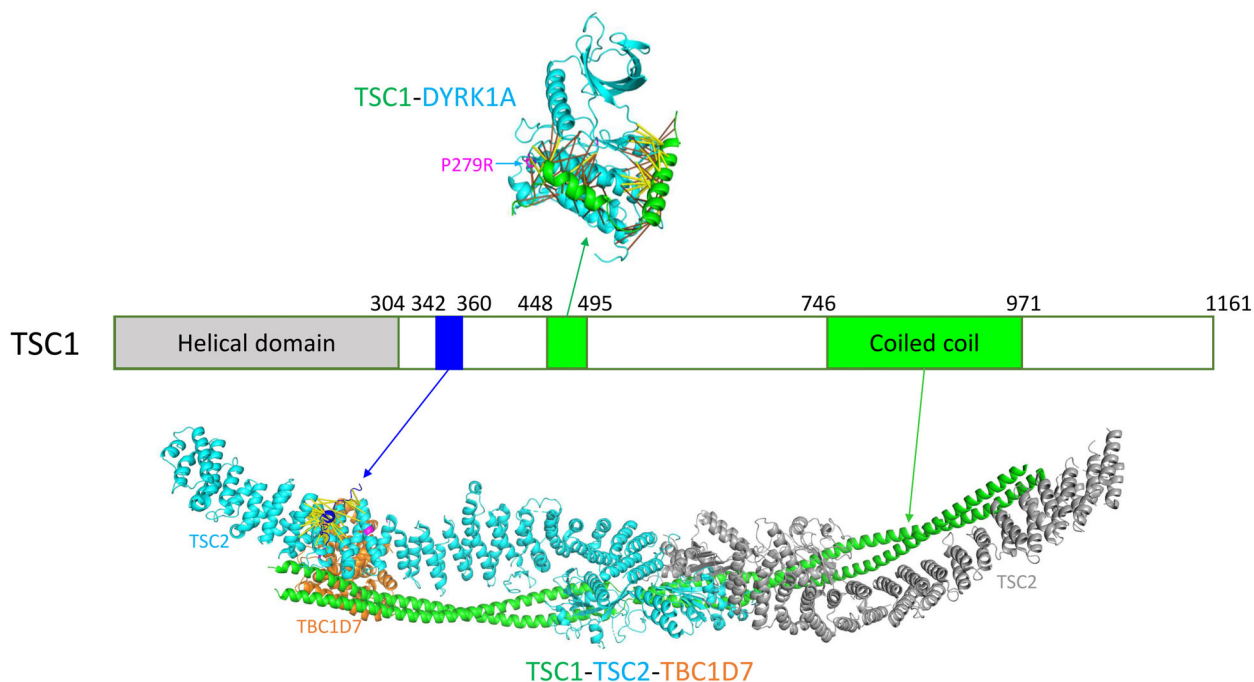


Fig. 6. Protein–protein interactions of TSC1 in the mTOR pathway. The domain structure of TSC1 is shown with the N-terminal helical domain and the coiled-coil region labeled. The bottom structure is based on an experimental structural complex between TSC1, TSC2, and TBC1D7 (PDB id: 7dl2). Two copies of the coiled-coil region of TSC1 form a dimer and are shown in green. The two subunits of the TSC2 dimer are shown in cyan and gray. A conserved region from 342 to 360 in TSC1 (shown in blue) was predicted to interact with TSC2, and the contacts between them with scores above 0.9 are shown in yellow sticks. The top structure shows the AlphaFold model of the complex between TSC1 (green) and DYRK1A (cyan). Contacts with scores above 0.9 are shown in yellow sticks, and contacts with scores between 0.5 and 0.9 are shown in orange sticks.

domain. This short sequence motif contains several highly conserved positions, such as W347, P349, and C353, that are mapped to the interaction interface, suggesting that the interaction through this motif could be functionally important for the TSC1–TSC2 complex.

AlphaFold predicted the interaction between TSC1 and the protein kinase DYRK1A. This interaction has been reported in three high-throughput experimental studies in BioGRID [46]. Moreover, it has been shown that DYRK1A regulates the mTOR pathway through binding to the subunits of the TSC complex [79]. However, the structural details of this interaction have not been revealed experimentally. AlphaFold model showed that the DYRK1A-interacting segment in TSC1 lies in the disordered region. It forms two alpha helices surrounding the helical subdomain (C-lobe) of the DYRK1A kinase domain (Fig. 6). The functional impact of this interaction has yet to be determined. One RCC somatic mutation in DYRK1A (P279R) is mapped to the interaction interface and was predicted to destabilize ($\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$) this complex based on predicted free energy changes upon binding (Fig. 6).

Structural insight into TRRAP complexes

Transactivation/tTransformation-domain Associated Protein (TRRAP) is a PIKK (phosphatidylinositol 3-kinase-related kinase) family pseudokinase involved in signal transduction and transcriptional regulation [80]. TRRAP is a canonical cancer driver suggested by NCG and has been classified as an oncogene gene product in the COSMIC database [81]. Two other members in the PIKK family are also RCC drivers: ATM (a tumor suppressor) and MTOR (an oncogene product). TRRAP is a main component of two large macromolecular histone-acetylation complexes: the Spt-Ada-Gcn5 acetyltransferase complex (SAGA) [82] and the nucleosome acetyltransferase of H4 complex (NuA4 or TIP60) [83].

The NuA4 complex plays critical roles in transcriptional activation, primarily through acetylating histones H4 and H2A. Another key component of the NuA4 complex is EP400. AlphaFold predicted the interaction between TRRAP and EP400 with a high probability (score > 0.99). The interaction regions in TRRAP are from the HEAT domain of helical repeats and the FAT

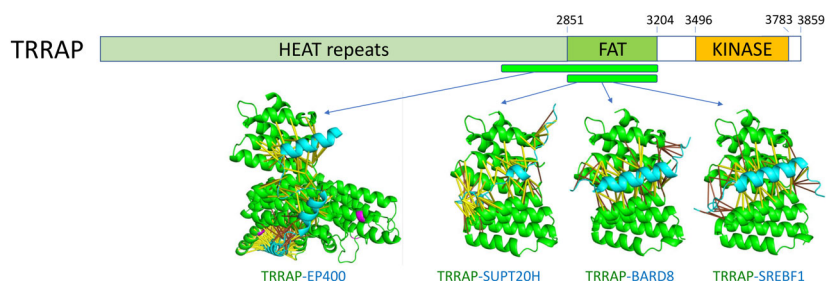


Fig. 7. Protein–protein interactions of TRRAP. The domain structure of TRRAP is shown with globular domains labeled. TRRAP is shown in green in the protein complex structures. RCC mutations are shown in magenta. Contacts with scores above 0.9 are shown in yellow sticks, and contacts with scores between 0.5 and 0.9 are shown in orange sticks.

domain (Fig. 7). The region of interactions from EP400 does not form a globular domain. Instead, it contains several alpha helices that wrap around TRRAP helical repeats (Fig. 7).

AlphaFold also predicted the interaction between TRRAP and the SAGA complex core subunit SUPT20H—an α -helix in SUPT20H docks onto the FAT domain of TRRAP (Fig. 7). The exact location on the TRRAP FAT domain is used in the interactions with two other proteins (BARD8 and SREBF1) and is also part of the interface between TRRAP and EP400 (Fig. 7). No somatic cancer mutations were identified in these TRRAP interfaces in our datasets. Since TRRAP has been classified as an oncogene product and could play a vital role in the transcriptional activation of genes promoting cell growth and cancer development, disrupting these interfaces could serve as a promising direction for cancer therapeutics targeting the TRRAP complexes.

Conclusion

RCCs encompass a diverse group of commonly occurring kidney tumors. Numerous cancer driver proteins associated with different RCC subtypes have been identified, and the majority of them are encoded by tumor suppressor genes. RCC drivers can be categorized into major functional groups, including chromatin remodeling, protein degradation, and transcription regulation. To gain deeper insights into the function of these drivers, we employed AlphaFold to predict and model their interactions with other proteins. AlphaFold predictions revealed high-confidence interactions involving various RCC drivers, including TCEB1, NRF2, subunits of the COMPASS-related complexes, the TSC complexes in the MTOR pathway, and complexes involving TRRAP. Notably, many of these predicted complexes have not been structurally characterized. Therefore, these findings offer novel and valuable structural insights into the mechanisms underlying the

functions of these complexes and their role in RCC development. A recurring scheme among these interactions is the utilization of short, linear motifs in disordered regions to interact with globular domains in other proteins. Cancer somatic missense mutations from large datasets of RCC genome sequencing were mapped to the interfaces of predicted and experimentally determined structures of the complexes. Cancer somatic mutations significantly impacted the stability of complexes formed by key RCC drivers, including VHL, TCEB1, and NRF2, highlighting the critical role of these mutations in RCC development. In contrast, relatively fewer deleterious missense mutations were found in other RCC drivers, especially those involved in chromatin remodeling. Furthermore, our analysis revealed a scarcity of cancer somatic mutations in the interfaces of complexes involving two oncogene products, NRF2 (the NRF2-MAF1 interface) and TRRAP (interfaces involving the FAT domain of TRRAP), suggesting critical functions of these protein complexes in cancer cells. This observation holds promise for developing cancer therapeutics targeting the disruption of these interfaces.

Materials and methods

Collection of RCC drivers and identification of candidate RCC PPIs from databases

A set of RCC driver proteins were obtained from NCG [36] and OncoVar [37]. NCG has a manually curated set of cancer driver genes for various cancer types. Canonical drivers and candidate drivers classified by NCG were retrieved from these cancer subtypes: chromophobe RCC (chRCC), ccRCC, non-clear cell RCC, papillary RCC (pRCC), and renal cancer (all histology). Cancer drivers were also retrieved from the OncoVar database for these RCC subtypes: ccRCC (KIRC), pRCC (KIRP), and chRCC (KICH). The combined set has 99 cancer drivers mapped to human-curated entries in the UniProt database [84]. We obtained human PPIs involving these cancer

drivers from the BioGRID database [46], which links each PPI to PubMed Identifiers (PMIDs), representing experimental studies supporting this interaction. We assigned a weight for each publication based on its associated number of interactions. The weight is defined as the reciprocal of the square root of the associated interaction number. The score for each interaction is defined as the sum of the weights of its associated PMIDs. We selected PPIs involving RCC drivers with scores no < 0.3 as candidate PPIs.

Generation of protein sequence alignments

We used a sequence database of 49 102 568 proteins from 2568 representative or reference eukaryotic proteomes (one per species) downloaded from NCBI. For each human protein (referred to as the target protein below), the corresponding orthologous group at the eukaryotes level in OrthoDB [85] was identified. Human proteins that were not classified in OrthoDB or belonged to the Zinc finger C2H2-type group (OrthoDB group: 1318335at2759), the Reverse transcriptase, RNaseH-like domain group (OrthoDB group: 583605at2759), and the pentatricopeptide repeat group (OrthoDB group: 1344243at2759) were ignored as these groups contain a large number (over 58,000) of paralogous sequences from gene duplications. The set of OrthoDB orthologs for any target protein was clustered by CD-HIT [86] at a 40% identity level. For each CD-HIT cluster, we selected one representative sequence that showed the lowest BLAST [87] e-value to the target protein. The representative sequences were then used as queries to search against the eukaryotic proteomes to identify homologous proteins to the target protein (e-value cutoff: 0.001).

For each target protein, we identified a single best hit from each organism, if available, that satisfies two criteria: (a) its sequence identity to the target protein is above 35%; (b) it shows the highest sequence similarity to the target protein among the combined BLAST hits found by multiple representative sequences. These best hits and the target protein were aligned by MAFFT [88] (with the – auto option). We labeled each protein by the organism it belongs to in the resulting MSAs and removed positions that were gaps in the human proteins. We concatenated the aligned sequences of the same organism from two MSAs to construct the joint MSA for any two human proteins. In cases where a sequence is missing from an organism, we used gaps to replace that sequence and added these cases at the end of the joint MSA.

Modeling protein complex structures with AlphaFold

We parsed large proteins into domains based on DeepMind's models of monomeric human proteins from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>) as described previously [42].

AlphaFold was used to model protein complexes by inputting joint MSAs of two protein (or domain) alignments with a gap of 200 residues in between the alignments. Protein pairs with combined length ≤ 1500 were directly modeled. Larger proteins were split into domains as described above, and all pairs of domains were modeled if none had a combined length > 1500 . AlphaFold network outputs the 'logits' of 'distogram', which are the probability distributions for the distances between the C β atoms of two residues to be in a serial of distance bins (64 bins of equal size, from 2 to 22 Å). Residue–residue contact probability was calculated as the sum of the probabilities for the first 32 distance bins, corresponding to distances between 2 and 12 Å. For a pair of proteins, the matrix (m) of contact probability is of the shape $(L_1 + L_2)$ by $(L_1 + L_2)$, where L_1 and L_2 are the length of the first and second proteins, respectively. To investigate the interprotein contacts, we extracted the sub-matrix $m' = m[0 : L_1][L_1 : L_1 + L_2]$. The highest residue–residue contact probability in this matrix m' is used as the contact probability for a pair of proteins.

We also applied two complex prediction methods developed based on AlphaFold: AF2Complex [38] and FoldDock [39]. These two methods mainly differ in the scoring of the complex. AF2Complex's metric is an interface score that depends on the predicted aligned error of interface residue pairs and the interface size. FoldDock relies on the pDockQ score based on the predicted LDDT (pLDDT) scores of interface residues and the number of contacts in the predicted interface. In comparison, our in-house program (AF-contact) ranks the complex models by the maximum intermolecular contact probabilities. We evaluated the performance of the three programs by comparing their predicted PPIs with the highest confidence to available experimental structural complexes (described below).

Analysis of somatic missense mutations in RCC PPI interfaces

We collected genome-level somatic mutations of RCC from the COSMIC database [81] and a recent cohort study of RCCs [47]. We used BindProfX [48] with FoldX energy function [49] to predict the free energy change of mutations mapped to protein–protein interaction interfaces in experimentally determined protein complex structures and AlphaFold models with confident predictions (contact probability > 0.9). Experimental structural complexes were obtained from bioassemblies of entries in the RCSB PDB database. We considered all pairwise interactions with at least 10 inter-residue contacts at the interface (a contact is defined as two residues from different chains within a distance of < 8 Å). Difference between the wild-type and mutant in free energy changes ($\Delta\Delta G$) above $1.4 \text{ kcal}\cdot\text{mol}^{-1}$ is considered to have a destabilizing effect on the binding affinity of the complex. The $1.4 \text{ kcal}\cdot\text{mol}^{-1}$ cutoff is more stringent

than the absolute value of the threshold suggested as stabilizing mutations ($-1 \text{ kcal}\cdot\text{mol}^{-1}$) by the BindProfX program [48] (<https://zhanggroup.org/BindProfX/help.html>). The free energy difference of $1.36 \text{ Kcal}\cdot\text{mol}^{-1}$ corresponds to about a 10-fold change (1 order of magnitude) in binding affinity at room temperature. Other studies have used similar cutoff values ranging from 1 to $1.5 \text{ Kcal}\cdot\text{mol}^{-1}$ when applying the BindProfX program [89,90]. Of the 996 RCC somatic mutations mapped to experimental structures or AlphaFold models, about 27% (268 mutations) have $\Delta\Delta G > 1.4 \text{ kcal}\cdot\text{mol}^{-1}$ (Table S2). Additional 88 mutations with $\Delta\Delta G$ in between $1 \text{ kcal}\cdot\text{mol}^{-1}$ and $1.4 \text{ kcal}\cdot\text{mol}^{-1}$ (Table S2) may also have destabilizing effect on the binding affinity of the complex.

Acknowledgements

QC is a Southwestern Medical Foundation-endowed scholar. JZ is supported by a training grant RP210041 from the Cancer Prevention and Research Institute of Texas. This research is partly supported by grant I-2095-20220331 to QC from Welch Foundation and the SPORE CEP/DRP Award to QC from the Kidney Cancer Program of University of Texas Southwestern Medical Center. The authors thank Dr Payal Kapur and Hua Zhong for providing the UT Southwestern Kidney Cancer Program WES datasets and Dr Nick V. Grishin and Dr James Brugarolas for helpful discussions.

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

The structural models of AlphaFold-predicted complexes (in PDB format files and in PyMOL sessions) are available at: <http://conglab.swmed.edu/RCC>.

Author contributions

JP, JZ, and QC contributed to conceptualization and methodology; JP contributed to formal analysis and investigation and writing—original draft preparation; QC contributed to writing—review and editing and supervision.

References

- 1 Siegel RL, Miller KD and Jemal A (2018) Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7–30.
- 2 Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **71**, 209–249.
- 3 Barata PC and Rini BI (2017) Treatment of renal cell carcinoma: current status and future directions. *CA Cancer J Clin* **67**, 507–524.
- 4 Padala SA, Barsouk A, Thandra KC, Saginala K, Mohammed A, Vakiti A, Rawla P and Barsouk A (2020) Epidemiology of renal cell carcinoma. *World J Oncol* **11**, 79–87.
- 5 Rini BI, Campbell SC and Escudier B (2009) Renal cell carcinoma. *Lancet* **373**, 1119–1132.
- 6 Madan A and Nanus DM (2022) Genomics and precision medicine: kidney cancer. In *Genomic and Precision Medicine* (Ginsburg GS, Willard HF, Strickler JH and McKinney MS, eds), pp. 285–295. Elsevier, Boston.
- 7 Hanahan D and Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646–674.
- 8 Stelzl U and Wanker EE (2006) The value of high quality protein-protein interaction networks for systems biology. *Curr Opin Chem Biol* **10**, 551–558.
- 9 Ivanov AA, Khuri FR and Fu H (2013) Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol Sci* **34**, 393–400.
- 10 Wells JA and McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009.
- 11 Mullard A (2012) Protein-protein interaction inhibitors get into the groove. *Nat Rev Drug Discov* **11**, 173–175.
- 12 Rolland T, Tasan M, Charlotiaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R *et al.* (2014) A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226.
- 13 Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–1178.
- 14 Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R *et al.* (2019) The BioGRID interaction database: 2019 update. *Nucleic Acids Res* **47**, D529–D541.
- 15 Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, Webber P, McMillan E, Rusnak L, Pham C *et al.* (2017) The OncoPPI network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat Commun* **8**, 14356.
- 16 Kim M, Park J, Bouhaddou M, Kim K, Rojc A, Modak M, Southeray M, McGregor MJ, O'Leary P, Wolf D *et al.* (2021) A protein interaction landscape of breast cancer. *Science* **374**, eabf3066.

- 17 Kuchaiev O, Rasajski M, Higham DJ and Przulj N (2009) Geometric de-noising of protein-protein interaction networks. *PLoS Comput Biol* **5**, e1000454.
- 18 Mackay JP, Sunde M, Lowry JA, Crossley M and Matthews JM (2007) Protein interactions: is seeing believing? *Trends Biochem Sci* **32**, 530–531.
- 19 Porta-Pardo E, Valencia A and Godzik A (2020) Understanding oncogenicity of cancer driver genes and mutations in the cancer genomics era. *FEBS Lett* **594**, 4233–4246.
- 20 Chin L, Hahn WC, Getz G and Meyerson M (2011) Making sense of cancer genomic data. *Genes Dev* **25**, 534–555.
- 21 Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr and Kinzler KW (2013) Cancer genome landscapes. *Science* **339**, 1546–1558.
- 22 Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B *et al.* (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, e18.
- 23 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* **47**, D941–D947.
- 24 International Cancer Genome, Hudson C, Anderson TJ, Artez W, Barker A, Bell AD, Bernabe C, Bhan RR, Calvo MK, Eerola F *et al.* (2010) International network of cancer genome projects. *Nature* **464**, 993–998.
- 25 Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C and Stuart JM (2013) The cancer genome atlas pan-cancer analysis project. *Nat Genet* **45**, 1113–1120.
- 26 Nakagawa H, Wardell CP, Furuta M, Taniguchi H and Fujimoto A (2015) Cancer whole-genome sequencing: present and future. *Oncogene* **34**, 5943–5950.
- 27 Porta-Pardo E, Garcia-Alonso L, Hrabec T, Dopazo J and Godzik A (2015) A pan-cancer catalogue of cancer driver protein interaction interfaces. *PLoS Comput Biol* **11**, e1004518.
- 28 Nishi H, Tyagi M, Teng S, Shoemaker BA, Hashimoto K, Alexov E, Wuchty S and Panchenko AR (2013) Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS One* **8**, e66273.
- 29 Engin HB, Kreisberg JF and Carter H (2016) Structure-based analysis reveals cancer missense mutations target protein interaction interfaces. *PLoS One* **11**, e0152929.
- 30 Xiong D, Lee D, Li L, Zhao Q and Yu H (2022) Implications of disease-related mutations at protein–protein interfaces. *Curr Opin Struct Biol* **72**, 219–225.
- 31 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
- 32 Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, Zhu W, Dunham AS, Albanese P and Keller A (2023) Towards a structurally resolved human protein interaction network. *Nat Struct Mol Biol* **30**, 216–225.
- 33 Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876.
- 34 Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR *et al.* (2021) Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805.
- 35 Cong Q, Anishchenko I, Ovchinnikov S and Baker D (2019) Protein interaction networks revealed by proteome coevolution. *Science* **365**, 185–189.
- 36 Repana D, Nulsen J, Dressler L, Bortolomeazzi M, Venkata SK, Tourn A, Yakovleva A, Palmieri T and Ciccarelli FD (2019) The network of cancer genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol* **20**, 1.
- 37 Wang T, Ruan S, Zhao X, Shi X, Teng H, Zhong J, You M, Xia K, Sun Z and Mao F (2021) OncoVar: an integrated database and analysis platform for oncogenic driver variants in cancers. *Nucleic Acids Res* **49**, D1289–D1301.
- 38 Gao M, Nakajima An D, Parks JM and Skolnick J (2022) AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun* **13**, 1744.
- 39 Bryant P, Pozzati G and Elofsson A (2022) Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* **13**, 1265.
- 40 Zhang J, Pei J, Durham J, Bos T and Cong Q (2022) Computed cancer interactome explains the effects of somatic mutations in cancers. *Protein Sci* **31**, e4479.
- 41 Pei J, Zhang J and Cong Q (2021) Human mitochondrial protein complexes revealed by large-scale coevolution analysis and deep learning-based structure modeling. *bioRxiv* [10.1101/2021.09.14.460228](https://doi.org/10.1101/2021.09.14.460228) [PREPRINT]
- 42 Pei J, Zhang J, Wang XD, Kim C, Yu Y and Cong Q (2022) Impact of Asp/Glu-ADP-ribosylation on protein-protein interaction and protein function. *Proteomics* **23**, e2200083.
- 43 Hakimi AA, Chen Y-B, Wren J, Gonen M, Abdel-Wahab O, Heguy A, Liu H, Takeda S, Tickoo SK and

- Reuter VE (2013) Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. *Eur Urol* **63**, 848–854.
- 44 Cowey CL and Rathmell WK (2009) VHL gene mutations in renal cell carcinoma: role as a biomarker of disease outcome and drug efficacy. *Curr Oncol Rep* **11**, 94–101.
 - 45 Pantuck AJ, Seligson DB, Klatte T, Yu H, Leppert JT, Moore L, O'Toole T, Gibbons J, Beldegrun AS and Figlin RA (2007) Prognostic relevance of the mTOR pathway in renal cell carcinoma: implications for molecular patient selection for targeted therapy. *Cancer* **109**, 2257–2267.
 - 46 Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F *et al.* (2021) The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* **30**, 187–200.
 - 47 Elias R, Tcheuyap VT, Kaushik AK, Singla N, Gao M, Reig Torras O, Christie A, Mulgaonkar A, Woolford L, Stevens C *et al.* (2021) A renal cell carcinoma tumorgraft platform to advance precision medicine. *Cell Rep* **37**, 110055.
 - 48 Xiong P, Zhang C, Zheng W and Zhang Y (2017) BindProfX: assessing mutation-induced binding affinity change by protein Interface profiles with pseudo-counts. *J Mol Biol* **429**, 426–434.
 - 49 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F and Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* **33**, W382–W388.
 - 50 Hsieh JJ, Le VH, Oyama T, Ricketts CJ, Ho TH and Cheng EH (2018) Chromosome 3p loss-orchestrated VHL, HIF, and epigenetic deregulation in clear cell renal cell carcinoma. *J Clin Oncol* **36**, JCO2018792549.
 - 51 Baldewijns MM, van Vlodrop IJ, Vermeulen PB, Soetekouw PM, van Engeland M and de Bruine AP (2010) VHL and HIF signalling in renal cell carcinogenesis. *J Pathol* **221**, 125–138.
 - 52 Nguyen HC, Yang H, Fribourgh JL, Wolfe LS and Xiong Y (2015) Insights into Cullin-RING E3 ubiquitin ligase recruitment: structure of the VHL-EloBC-Cul2 complex. *Structure* **23**, 441–449.
 - 53 Min JH, Yang H, Ivan M, Gertler F, Kaelin WG Jr and Pavletich NP (2002) Structure of an HIF-1 α - pVHL complex: hydroxyproline recognition in signaling. *Science* **296**, 1886–1889.
 - 54 Tarade D, Robinson CM, Lee JE and Ohh M (2018) HIF-2 α -pVHL complex reveals broad genotype-phenotype correlations in HIF-2 α -driven disease. *Nat Commun* **9**, 3359.
 - 55 Hakimi AA, Tickoo SK, Jacobsen A, Sarungbam J, Sfakianos JP, Sato Y, Morikawa T, Kume H, Fukayama M, Homma Y *et al.* (2015) TCEB1-mutated renal cell carcinoma: a distinct genomic and morphological subtype. *Mod Pathol* **28**, 845–853.
 - 56 Kamura T, Sato S, Haque D, Liu L, Kaelin WG Jr, Conaway RC and Conaway JW (1998) The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. *Genes Dev* **12**, 3872–3881.
 - 57 Rao RC and Dou Y (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer* **15**, 334–346.
 - 58 Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF and Shilatifard A (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci USA* **98**, 12902–12907.
 - 59 Lavery WJ, Barski A, Wiley S, Schorry EK and Lindsley AW (2020) KMT2C/D COMPASS complex-associated diseases [K(CD)COM-ADs]: an emerging class of congenital regulopathies. *Clin Epigenetics* **12**, 10.
 - 60 Guo X and Zhang Q (2017) The emerging role of histone demethylases in renal cell carcinoma. *J Kidney Cancer VHL* **4**, 1–5.
 - 61 Zhang P, Lee H, Brunzelle JS and Couture JF (2012) The plasticity of WDR5 peptide-binding cleft enables the binding of the SET1 family of histone methyltransferases. *Nucleic Acids Res* **40**, 4237–4246.
 - 62 Yan L, Zhang Y, Ding B, Zhou H, Yao W and Xu H (2019) Genetic alteration of histone lysine methyltransferases and their significance in renal cell carcinoma. *PeerJ* **7**, e6396.
 - 63 Schulz WA, Lang A, Koch J and Greife A (2019) The histone demethylase UTX/KDM6A in cancer: progress and puzzles. *Int J Cancer* **145**, 614–620.
 - 64 He F, Ru X and Wen T (2020) NRF2, a transcription factor for stress response and beyond. *Int J Mol Sci* **21**, 4777.
 - 65 Taguchi K and Yamamoto M (2020) The KEAP1-NRF2 system as a molecular target of cancer treatment. *Cancers (Basel)* **13**, 46.
 - 66 Rojo de la Vega M, Chapman E and Zhang DD (2018) NRF2 and the hallmarks of cancer. *Cancer Cell* **34**, 21–43.
 - 67 Clerici S and Boletta A (2020) Role of the KEAP1-NRF2 Axis in renal cell carcinoma. *Cancers (Basel)* **12**, 3458.
 - 68 Lo SC, Li X, Henzl MT, Beamer LJ and Hannink M (2006) Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* **25**, 3605–3617.
 - 69 Fukutomi T, Takagi K, Mizushima T, Ohuchi N and Yamamoto M (2014) Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degron and Keap1. *Mol Cell Biol* **34**, 832–846.

- 70 Kerins MJ and Ooi A (2018) A catalogue of somatic NRF2 gain-of-function mutations in cancer. *Sci Rep* **8**, 12846.
- 71 Adamson RJ, Payne NC, Bartual SG, Mazitschek R and Bullock AN (2023) Structural and biochemical characterization establishes a detailed understanding of KEAP1-CUL3 complex assembly. *Free Radic Biol Med* **204**, 215–225.
- 72 Katsuoka F and Yamamoto M (2016) Small Maf proteins (MafF, MafG, MafK): history, structure and function. *Gene* **586**, 197–205.
- 73 Shuch B, Vourganti S, Friend JC, Zehngbot LM, Linehan WM and Srinivasan R (2012) Targeting the mTOR pathway in chromophobe kidney cancer. *J Cancer* **3**, 152–157.
- 74 Sarbassov DD, Ali SM and Sabatini DM (2005) Growing roles for the mTOR pathway. *Curr Opin Cell Biol* **17**, 596–603.
- 75 Nobukuni T and Thomas G (2004) The mTOR/S6K signalling pathway: the role of the TSC1/2 tumour suppressor complex and the proto-oncogene Rheb. *Novartis Found Symp* **262**, 148–154; discussion 154–9, 265–8.
- 76 Deng Y, Qin Y, Srikantan S, Luo A, Cheng ZM, Flores SK, Vogel KS, Wang E and Dahia PLM (2018) The TMEM127 human tumor suppressor is a component of the mTORC1 lysosomal nutrient-sensing complex. *Hum Mol Genet* **27**, 1794–1808.
- 77 Ramirez Reyes JM, Cuesta R and Pause A (2021) Folliculin: a regulator of transcription through AMPK and mTOR signaling pathways. *Front Cell Dev Biol* **9**, 667311.
- 78 Walchli M, Berneiser K, Mangia F, Imseng S, Craigie LM, Stüttfeld E, Hall MN and Maier T (2021) Regulation of human mTOR complexes by DEPTOR. *Elife* **10**, e70871.
- 79 Menon VR and Litovchick L (2015) DYRK1A kinase regulates mTOR signaling via modulating the TSC complex. *Cancer Res* **75**, 1022.
- 80 Murr R, Vaissiere T, Sawan C, Shukla V and Herceg Z (2007) Orchestration of chromatin-based processes: mind the TRRAP. *Oncogene* **26**, 5358–5372.
- 81 Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I and Forbes SA (2018) The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* **18**, 696–705.
- 82 Koutelou E, Hirsch CL and Dent SY (2010) Multiple faces of the SAGA complex. *Curr Opin Cell Biol* **22**, 374–382.
- 83 Doyon Y, Selleck W, Lane WS, Tan S and Cote J (2004) Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Mol Cell Biol* **24**, 1884–1896.
- 84 UniProt C (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489.
- 85 Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simao FA and Zdobnov EM (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* **47**, D807–D811.
- 86 Fu L, Niu B, Zhu Z, Wu S and Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152.
- 87 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- 88 Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780.
- 89 Horn S, Au M, Basel-Salmon L, Bayrak-Toydemir P, Chapin A, Cohen L, Elting MW, Graham JM, Gonzaga-Jauregui C, Konen O *et al.* (2019) De novo variants in PAK1 lead to intellectual disability with macrocephaly and seizures. *Brain* **142**, 3351–3359.
- 90 Begemann A, Sticht H, Begtrup A, Vitobello A, Faivre L, Banka S, Alhaddad B, Asadollahi R, Becker J, Bierhals T *et al.* (2021) New insights into the clinical and molecular spectrum of the novel CYFIP2-related neurodevelopmental disorder and impairment of the WRC-mediated Actin dynamics. *Genet Med* **23**, 543–554.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Top 1000 protein complexes predicted by AF-contact, AF2Complex, and FoldDock.

Table S2. Differences of free energy changes of mutations in the interfaces of AlphaFold-predicted protein complexes and experimentally determined protein complexes.