Comprehensive characterization of protein–protein interactions perturbed by disease mutations

Feixiong Cheng [1,2,3,18], Junfei Zhao[4,5,18], Yang Wang[6,7,18], Weiqiang Lu [8,18], Zehui Liu[9],
Yadi Zhou[1], William R. Martin [1], Ruisheng Wang[10], Jin Huang[9], Tong Hao [6,7], Hong Yue [6,7],
Jing Ma[6,11], Yuan Hou[1], Jessica A. Castrillon[1], Jiansong Fang[1,12], Justin D. Lathia[2,3], Ruth A. Keri[3,13],
Felice C. Lightstone [14], Elliott Marshall Antman [15], Raul Rabadan [4,5], David E. Hill [6,7],
Charis Eng [1,2,3,16,17], Marc Vidal[6,7] and Joseph Loscalzo [10 ✉]

Technological and computational advances in genomics and interactomics have made it possible to identify how disease mutations perturb protein–protein interaction (PPI) networks within human cells. Here, we show that disease-associated germline variants are significantly enriched in sequences encoding PPI interfaces compared to variants identified in healthy participants from the projects 1000 Genomes and ExAC. Somatic missense mutations are also significantly enriched in PPI interfaces compared to noninterfaces in 10,861 tumor exomes. We computationally identified 470 putative oncoPPIs in a pan-cancer analysis and demonstrate that oncoPPIs are highly correlated with patient survival and drug resistance/sensitivity. We experimentally validate the network effects of 13 oncoPPIs using a systematic binary interaction assay, and also demonstrate the functional consequences of two of these on tumor cell growth. In summary, this human interactome network framework provides a powerful tool for prioritization of alleles with PPI-perturbing mutations to inform pathobiological mechanism- and genotype-based therapeutic discovery.

Interpretation of the clinical pathogenic effects of variants is crucial for the advancement of precision medicine. However, our ability to understand the functional and biological consequences of genetic variants identified by human genome sequencing projects is limited. Many computational approaches can identify only a small proportion of pathogenic variants with the high confidence required in clinical settings. Human genome sequencing studies have reported potential mutation–disease associations with the functional regions altered by somatic mutations, such as molecular drivers in cancers[1,2]. However, many important issues in the field remain unclear, including the phenotypic consequences of different mutations within the same gene and the same mutation across different cell types.

Recent efforts using systematic analyses of 1,000–3,000 missense mutations in Mendelian disorders[3,4] and ~2,000 de novo missense mutations in developmental disorders[5] demonstrate that disease-associated alleles commonly alter distinct PPIs rather than grossly affecting the folding and stability of proteins[3,4]. Network-based approaches provide new insights into disease–disease[6] and drug–disease[7–9] relationships within the human interactome. Nevertheless, the functional consequences of disease mutations on the comprehensive human interactome and their implications for therapeutic development remain understudied. Several studies have suggested that protein structure-based mutation enrichment analysis offers a potential tool for identification of possible cancer driver genes[10], such as hotspot mutation regions in three-dimensional (3D) protein structures[11–14]. Development of new computational and experimental approaches for the study of functional consequences of mutations at single–amino acid residue resolution is crucial for our understanding of the pleiotropic effects of disease risk genes, and offers potential strategies for accelerating precision medicine[15,16].

In this study, we investigated the network effects of disease-associated mutations at amino acid resolution within the 3D

[1]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. [2]Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, USA. [3]Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [4]Department of Systems Biology, Herbert Irving Comprehensive Center, Columbia University, New York, NY, USA. [5]Department of Biomedical Informatics, Columbia University, New York, NY, USA. [6]Center for Cancer Systems Biology (CCSB), Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [7]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [8]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. [9]Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai, China. [10]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [11]Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, Zhejiang, China. [12]Science and Technology Innovation Center, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China. [13]Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [14]Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Lab, Livermore, CA, USA. [15]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [16]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [17]Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA. [18]These authors contributed equally: Feixiong Cheng, Junfei Zhao, Yang Wang, Weiqiang Lu. ✉e-mail: jloscalzo@rics.bwh.harvard.edu

macromolecular interactome of structurally resolved and computationally predicted PPI interfaces. We provide evidence for widespread perturbations of PPIs in human diseases caused by both germline and somatic mutations identified in large-scale sequencing studies.

Results

Widespread network perturbations by germline mutations. To investigate the effects of disease-associated mutations at amino acid resolution on a PPI network, we constructed a structurally resolved human protein–protein interactome network by assembling three types of experimentally validated binary PPIs with experimental or predicted interface information: (1) PPIs with crystal structures from the RCSB protein data bank[17], (2) PPIs with homology modeling structures from Interactome3D[18] and (3) experimentally determined PPIs with computationally predicted interface residues from Interactome INSIDER[19] (Methods). In total, we collected 121,575 PPIs (edges or links) connecting 15,046 unique proteins (nodes). We found that disease-associated mutations from the Human Gene Mutation Database (HGMD)[20] were significantly enriched in PPI interfaces of the respective proteins compared to variants identified in individuals from the projects 1000 Genomes[21] ($P < 2.2 \times 10^{-16}$, two-tailed Fisher's test; Fig. 1a) and ExAC[22] ($P < 2.2 \times 10^{-16}$, two-tailed Fisher's test; Fig. 1a). We found the same level of enrichment for mutant interface residues with both crystal structures (Supplementary Fig. 1) and within the high-throughput systematic interactome identified by (unbiased) yeast two-hybrid (Y2H) screening assays[23] (Supplementary Fig. 2). Figure 1b reveals the global view of network perturbations in disease-associated germline mutations from the HGMD[20]. For example, multiple disease-associated gene products, including p53, LMNA, CFTR, HBA and GJB2, have PPIs altered by multiple interface, disease-associated mutations.

Proprotein convertase subtilisin/kexin type 9 (PCSK9), first discovered by human genetic screening studies in 2003, has generated great interest in genomics-informed drug discovery for cardiovascular disease[24]. We therefore investigated whether the PCSK9 allele carrying a p.Ser127Arg substitution perturbs the interaction between PCSK9 and LDLR (low-density lipoprotein receptor protein) (Fig. 1c). We performed molecular dynamics simulations (Supplementary Fig. 3) to predict whether the binding affinity between p.Ser127Arg PCSK9 and LDLR would be increased ($545 \, \text{kJ} \, \text{mol}^{-1}$) compared with wild type (WT) ($691 \, \text{kJ} \, \text{mol}^{-1}$; Supplementary Fig. 4). We focused on the interaction between the beta-propeller region of LDLR and the noncovalently-bound propeptide (residues 61–152) of PCSK9. The binding affinity ($\Delta\Delta G$) of p.Ser127Arg relative to that of WT is predicted to change by $-14 \, \text{kJ} \, \text{mol}^{-1}$, suggesting that the strength of interaction with LDLR is perturbed by the p.Ser127Arg substitution on PCSK9.

We next focused on the propeptide of PCSK9, where the total change in binding affinity by p.Ser127Arg is predicted to be altered by $-211 \, \text{kJ} \, \text{mol}^{-1}$ (Supplementary Fig. 4). The region centered on the p.Ser127Arg substitution is key to the increased binding affinity in the mutant PCSK9 (ref. [25]). While interactions between the propeptide of PCSK9 and the beta-propeller of LDLR do exist in the WT system, they do not involve the region surrounding residue 127 (Supplementary Fig. 4). Much of the change in binding affinity, on a per-residue basis, is due to a steep increase in electrostatic interaction energy with the substituted residue (Arg127), which accounts for the greatest contribution to the significant change in overall binding affinity (Supplementary Fig. 4). For example, a number of arginine residues in the alpha-helix (Leu88-Arg105) distal to the interface between the beta-propeller of LDLR and the propeptide are predicted to exhibit an increase in their binding affinity due to an increase in electrostatic interactions. This increase in electrostatic interactions stems from a reduction of roughly 15 Å in the

distance between the center of the helix and the interaction region, measured from the alpha-carbon of Arg86 in PCSK9 and Arg385 of LDLR (Supplementary Fig. 5). For the PCSK9 p.Ser127Arg–LDLR complex, the combination of the extra length of the side chain, in addition to the charged guanidinium functionality, would allow interactions with the side chains of Arg385 and His386 on LDLR. In summary, combining human interactome analyses and computational biophysical modeling supports an interaction perturbation model for p.Ser127Arg, in agreement with our notion of PPI-perturbing alleles.

PPI-perturbing alleles in somatic mutations. We next turned to an investigation of the somatic mutation load between PPI interface and noninterface regions. We inspected 1,750,987 missense somatic mutations from 10,861 tumor exomes across 33 cancer types from The Cancer Genome Atlas (TCGA) in the interface regions of 121,575 PPIs. We found a significantly higher somatic mutation burden at PPI interfaces compared to noninterfaces across all 33 cancer types ($P < 2.2 \times 10^{-16}$, two-tailed Wilcoxon test; Fig. 2a). For breast cancer, the average missense mutation burden leading to amino acid substitutions is 20 per 1 million residues in interface regions, significantly higher than that in noninterface regions (4 per 1 million, five-fold enrichment; $P < 2.2 \times 10^{-16}$, two-tailed Wilcoxon test). We also found the same trend where somatic mutations are highly enriched in both crystal-structure-derived (Supplementary Fig. 6) and computationally inferred (Supplementary Fig. 7) PPI interfaces compared to noninterface regions across all 33 cancer types. We further performed the same mutation burden analysis in structurally resolved, unbiased PPIs. We also found a higher mutation load at the interface residues of the physical human interactome using co-crystal structures only (Supplementary Fig. 8) and unbiased, binary PPIs identified by Y2H with available co-crystal-structure-derived interfaces and computationally predicted interfaces (Supplementary Fig. 9), supporting the robustness of the analysis. We investigated the cumulative distribution of deleterious amino acid substitutions between PPI interface and noninterface regions. Deleterious substitutions quantified by both SIFT (Fig. 2b) and PolyPhen-2 (Fig. 2c) scores are significantly enriched at PPI interfaces compared to noninterfaces. Thus, widespread interaction perturbations caused by somatic mutations may contribute to tumorigenesis. Following this analysis, we next pursued the identification of putative oncoPPIs (also termed significantly mutated edges) by systematically exploring the mutation burden between PPI interfaces versus noninterfaces across 10,861 tumor exomes.

Systematic identification of oncoPPIs. Based on the observation that somatic missense mutations are enriched at PPI interfaces (Fig. 2a) and that mutations at PPI interfaces are more likely to be deleterious than those at the noninterfaces (Fig. 2b,c), we prioritized putative oncoPPIs using a binomial statistical model (Methods). In total, we investigated the somatic mutations in 10,861 tumor–normal pairs across 33 cancer types/subtypes in the TCGA database. All abbreviations for 33 cancer types/subtypes are provided in Methods. We selected putative oncoPPIs using subject matter expertise based on a combination of factors: (1) strength of the prediction (a higher adjusted P value is shown in Supplementary Data 1); (2) availability of crystal-structure-derived PPI interfaces; (3) novelty of the predicted oncoPPIs; and (4) availability of sufficient patient survival and pharmacogenomics data for meaningful evaluation. Applying these criteria resulted in 470 putative oncoPPIs among a total of 15,357 PPIs (Supplementary Fig. 10 and Supplementary Data 1). Among 470 pan-cancer putative oncoPPIs, we highlighted 13 with documented crystal-structure-based interface mutations, including KEAP1–KFE2L2, SPOP–H2AFY and FGF1–FGFR3 (Extended Data Fig. 1).

We next investigated putative oncoPPIs identified across 33 individual cancer types. In total, we found 2,708 unique, putative
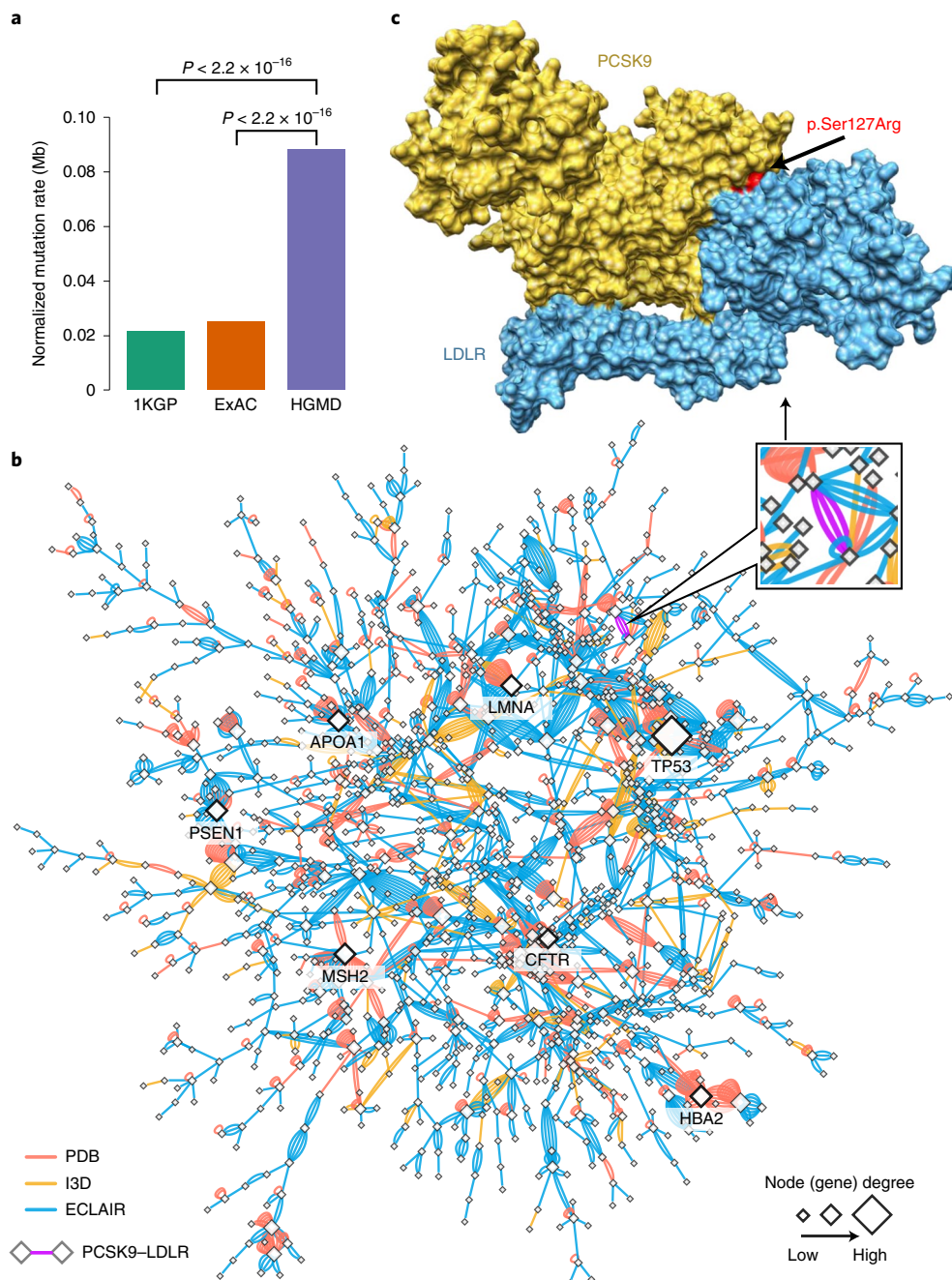
Fig. 1 | Proof of concept of protein–protein interaction–perturbing alleles in human diseases. a, Distribution of mutation burden at protein–protein interfaces for disease-associated germline mutations from HGMD in comparison to mutations from the projects 1000 Genome (1KGP) and ExAC. *P* values were calculated by two-tailed Fisher's test. **b**, A subnetwork highlighting the disease network module for all human-disease-associated mutations at protein–protein interfaces. An edge denotes at least one disease-associated mutation from HGMD at the interfaces of experimentally identified binary PPIs. Three types of protein–protein interface are illustrated: (1) PPIs with crystal structures (PDB), (2) PPIs with homology models (I3D) and (3) experimentally determined PPIs with computationally predicted interface residues (ECLAIR) (Methods). A PPI with multiple types of evidence of protein–protein interface-associated mutations is denoted by multiple edges (links). Node size is measured by degree (connectivity). **c**, Example of a PPI-perturbing alteration (p.Ser127Arg in PCSK9) affecting complexes PCSK9 and LDLR (PDB ID: 3M0C). The protein 3D images were prepared by PyMOL (https://pymol.org/2/).

oncoPPIs across 29 cancer types, including 2,338 heterodimers and 370 homodimers, that reached a level of significance (Fig. 2d; FDR *q* < 0.05; Methods and Supplementary Data 1), with at least one putative oncoPPI (Supplementary Fig. 11). Among the 10,861 TCGA tumor samples analyzed, 4,405 (40%) harbor at least one putative oncoPPI. Figure 3 illustrates the landscape of putative oncoPPIs across the 33 cancer types. The top five oncoPPIs (heterodimers) with known structural information are

BRAF–MAP2K1, PIK3R1–PIK3CA, TP53–EP300, TP53–TP53BP2 and KEAP1–NFE2L2. Among these complexes, 3D structural analysis revealed that p.Val600Glu on BRAF may perturb the BRAF–MAP2K1 interaction (Supplementary Fig. 12), consistent with a previous study[26]. The specific PPI-perturbing somatic mutations among these leading oncoPPIs are p.Arg132His in IDH1, p.Val600Glu in BRAF, p.His1047Arg in PIK3CA, p.Gln209Leu in GNA11 and p.Phe133Leu in SPOP (Fig. 3). In summary, many
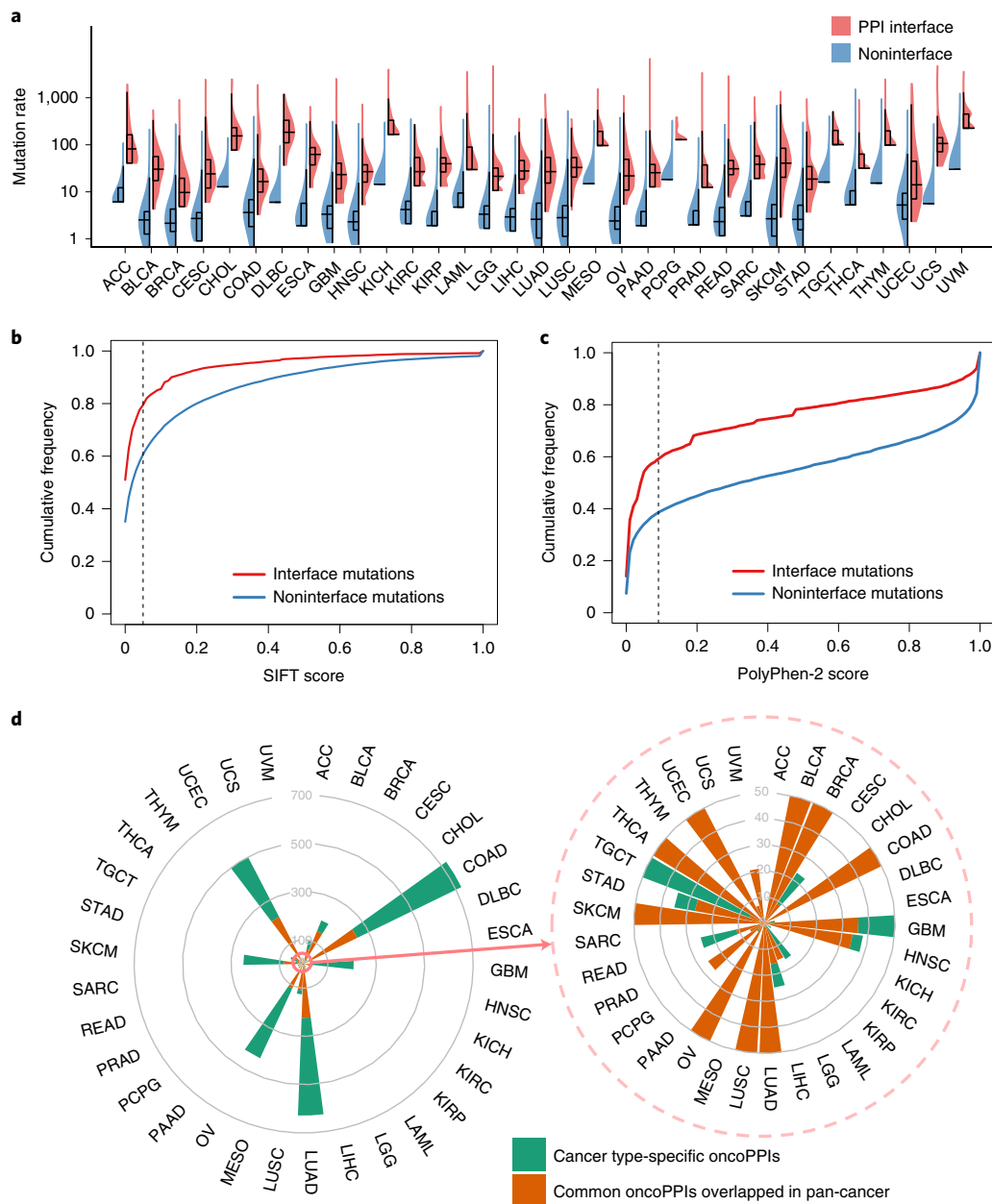
Fig. 2 | Network perturbation by missense somatic mutations in human cancers. a, Distribution of missense mutations in protein–protein interfaces versus noninterfaces across 33 cancer types/subtypes from TCGA. The data are represented as violin plots with underlaid boxplots, where the middle line is the median, the lower and upper edges of the rectangle are the first and third quartiles and the lower and upper whiskers of the violin plot represent the interquartile range (IQR) ×1.5. **b,c**, Cumulative frequencies of SIFT (**b**) and PolyPhen-2 scores (**c**) for protein–protein interface mutations (red) versus noninterface (blue) mutations. **d**, Circos plot illustrating the landscape of significant oncoPPIs harboring a statistically significant excess number of missense mutations at PPI interfaces across 33 cancer types/subtypes (adjusted $P < 0.05$; Methods). Bars denote the number of putative oncoPPIs across each cancer type/subtype. Overlapped oncoPPIs between individual cancer type and pan-cancer analysis (adjusted $P < 0.001$) are highlighted in orange. Detailed data are provided in Supplementary Data 1. All abbreviations for each cancer type/subtype are provided in the main text.

known driver mutations are commonly located in regions that are part of the interaction interface of one or the other binding partner proteins, indicating the potential for widespread interaction perturbation in human cancers (Fig. 3).

Pharmacogenomic landscape of PPI-perturbing mutations. We next examined whether putative oncoPPIs can predict drug responses (Fig. 4a). We used analysis of variance (ANOVA) to determine whether there is a significant difference in the response of specific cell lines by comparing the PPI interface-mutated and WT

interface groups. Utilizing analysis of the drug pharmacogenomic profiles of >1,000 cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC) database (Methods), we found that interface-predicted mutations of oncoPPIs are highly correlated with sensitivity or resistance to multiple therapeutic agents (Supplementary Data 2). Figure 4b shows that oncoPPIs correlate with the sensitivity or resistance of 66 clinically investigational or approved anticancer agents. For example, we found that PPI-perturbing mutations in *SNAI1* and *ACTN2* were responsible for resistance to foretinib (a c-Met and VEGFR2 kinase inhibitor[27]; Supplementary Fig. 13),
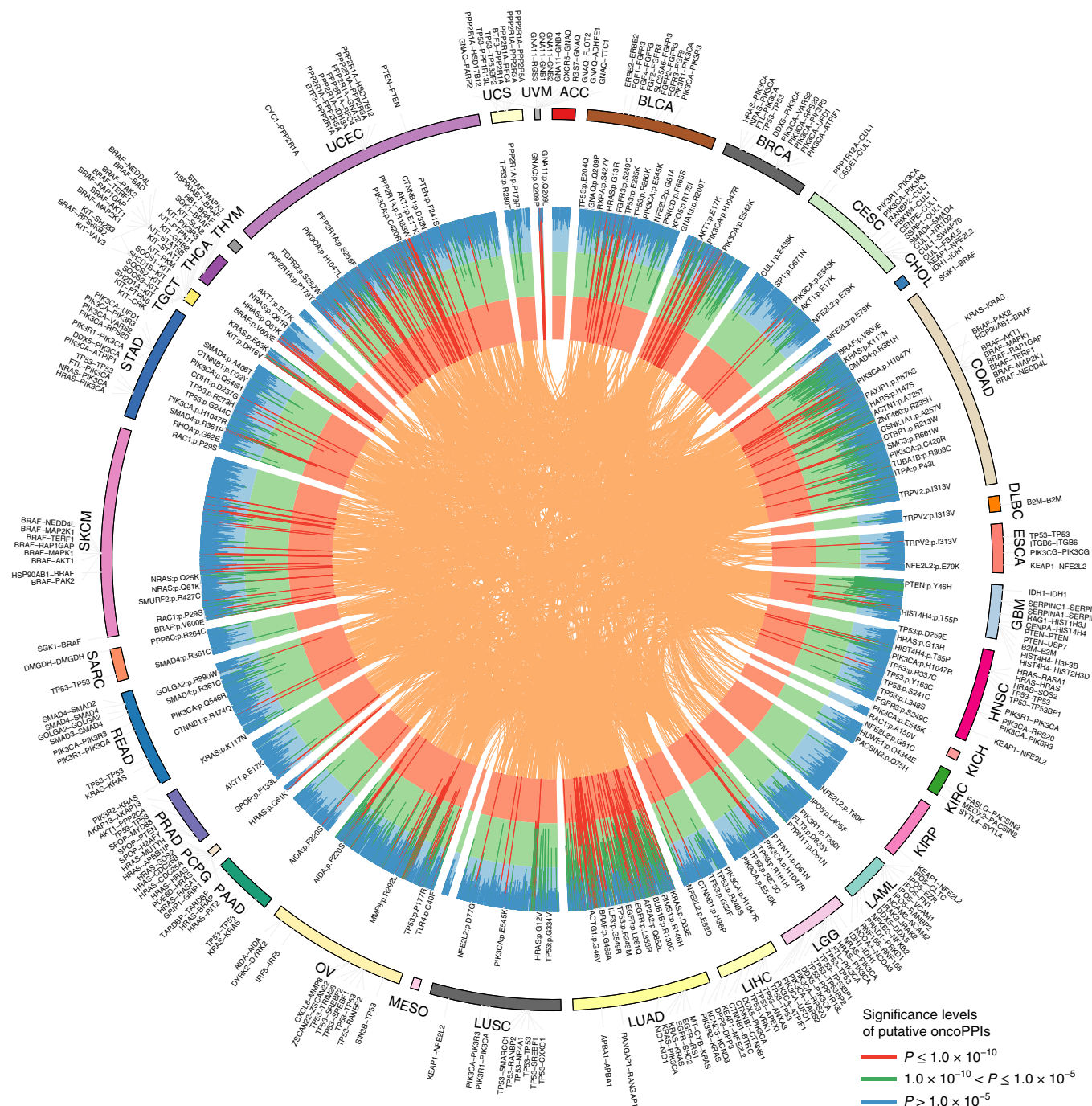
Fig. 3 | Landscape of protein–protein interaction–perturbing mutations across 33 cancer types. Circos plot displaying significant putative oncoPPIs (Methods) harboring a statistically significant excess number of missense mutations at PPI interfaces across 33 cancer types. Putative oncoPPIs with various significance levels (Methods) are plotted in the three inner layers. The links (edges, orange) connecting two oncoPPIs indicate two cancer types sharing the same oncoPPIs. Selected significant oncoPPIs and their related mutations are plotted on the outer surface. The length of each line is proportional to $-\log_{10}(P)$. All oncoPPIs and PPI-perturbing mutations are freely available at https://mutanome.lerner.ccf.org/.

consistent with a previous study[28]. In addition, PPI-perturbing mutations in GNAI2 (encoding G protein subunit alpha-I2 (ref. [29])) that directly disrupt interactions with RGS20 and TRIP6 were associated with resistance to several chemotherapeutic agents, including gemcitabine and tamoxifen (Supplementary Fig. 13).

We further investigated their correlation with anticancer drug response by analyzing the data from in vivo compound screens between ~1,000 patient-derived tumor xenograft (PDXs) models and 62 medications[30]. In total, we found 2,808 significant correlations ($P < 0.05$, ANOVA; Methods) between 49 medications

and 1,411 putative oncoPPIs (Fig. 4c). Amino acid substitutions in vinculin (VCL), located at the interface between VCL and fragile X mental retardation syndrome–related protein 1 (FXR1) are significantly correlated with resistance to encorafenib, an FDA-approved BRAF inhibitor for the treatment of melanoma[31], compared to patients without VCL–FXR1-perturbing mutations. Importantly, FXR1–BRAF fusion has been found in glioma[32,33], which may explain the correlation of the response to encorafenib with interface substitutions that disrupt VCL–FXR1 (Fig. 4d). Interface substitutions that disrupt BRAF–MAP2K1 were significantly associated

with response to combination therapy with ribociclib (a CDK4/CDK6 inhibitor[34]) and encorafenib in PDX models, suggesting potential pharmacogenomic biomarkers for rational development of combination therapy in cancer. In summary, PPI-perturbing mutations offer potential as pharmacogenomic biomarkers in both cancer cell lines and PDX models, an observation that warrants further detailed investigation using preclinical and clinical data.

PPI-perturbing alleles in histone H4 complex. We next investigated the correlation between patient survival and oncoPPIs. We found that interface substitutions of p53 or arginine-rich splicing factor 1 (SRSF1) (ref. [35]) in p53–SRSF1 were significantly associated with poor survival rate in bladder urothelial carcinoma (BLCA) ($P = 6.1 \times 10^{-3}$, log-rank test), breast invasive carcinoma (BRCA) ($P = 6.4 \times 10^{-4}$) and colon adenocarcinoma (COAD) ($P = 7.2 \times 10^{-3}$) among 33 cancer types (Extended Data Fig. 2 and Supplementary Fig. 14). Nevertheless, mutations in p53 alone are modestly associated with poor survival rate in BRCA ($P = 0.03$, log-rank test) but are not associated with BLCA survival rate ($P = 0.79$) and COAD ($P = 0.11$; Extended Data Fig. 2). Amino acid substitutions of histone acetyltransferase p300 (EP300) (ref. [36]) or nuclear transcription factor Y subunit beta (NFYB) at the interfaces of EP300–NFYB significantly correlated with poor survival rate in melanoma patients ($P = 0.019$, log-rank test; Supplementary Fig. 15). For COAD, PPI-perturbing mutations in plasminogen (PLG) or mothers against decapentaplegic homolog 4 (SMAD4) are highly correlated with poor survival ($P = 0.0025$, log-rank test; Supplementary Fig. 15).

Histone H4, encoded by *HIST1H4A*, is one of the five main histone proteins involved in gene regulation and chromatin structure[37]. Figure 5a shows multiple potential PPI-perturbing mutations in histone H4 in complex with death-associated protein 6 (DAXX), H3 histone family member 3A (H3F3A) and centromere protein A (CENPA). We found a high mutation burden of the histone H4 complex in multiple cancer types (Fig. 5b), especially for BLCA, head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC) and uterine corpus endometrial carcinoma (UCEC). Figure 5c illustrates several selected H4 interface substitutions of the histone H4 complex. Interface substitutions of HIST1H4A or H3F3A (encoding histone H3.3 (ref. [38])) in H3.3–H4 interfaces are significantly associated with poor survival in COAD (Fig. 5e) and response to multiple anticancer drugs, including paclitaxel and BMS-754807 (Fig. 5f). We found multiple interface substitutions between histone H4 and DAXX that are potentially involved in tumorigenesis and drug responses (Fig. 5c). For example, PPI-perturbing mutations in histone H4 that disrupt the DAXX interaction are significantly associated with poor survival in COAD and LUSC, and are further associated with drug responses in those malignancies (Fig. 5e,f). Thus, PPI-perturbing alleles in the histone H4 complex provide another example of clinically relevant mechanisms of PPI-perturbing alleles.

Experimental validation of PPI-perturbing alleles. To test PPI-perturbing alleles experimentally, we selected and cloned 13 high-confidence oncoPPIs by using our previously established binary interaction mapping vectors[39]. We selected 23 somatic missense mutations (Supplementary Table 1) across these 13 oncoPPIs for experimental validation, using subject matter expertise based on a combination of factors: (1) interface mutations with crystal-structure evidence; (2) PPI-perturbing mutations significantly correlated with drug response and patient survival; and (3) mutations affecting the interaction that can be detected by Y2H assay[39].

We first tested the impact of these mutations on the corresponding 13 oncoPPIs using our well-established Y2H assay[3,39]. As shown in Fig. 6, among 23 tested mutations, 17 (74%) led to loss of PPIs or reduced the detected effects of PPIs while six (26%) maintained those interactions predicted to be affected by the mutation (Supplementary Table 1). Our experimental results are consistent with the PPI test results of disease mutations in our previous study[3], in which approximately two-thirds of disease mutations were PPI perturbing. Importantly, this previous study did not identify the location of the mutation in the protein tertiary structure.

Among the tested alterations, p.Met146Lys (Fig. 6b) in arachidonate 5-lipoxygenase (ALOX5) disrupts its interaction with MAD1L1, a mitotic spindle assembly checkpoint protein. Both ALOX5 and MAD1L1 have been reportedly involved in tumorigenesis and/or tumor progression of several cancer types[40,41]. Another example is the p.Arg382Trp alteration in homeobox and leucine zipper encoding (HOMEZ) that alters the interaction between HOMEZ and early B-cell factor 1 (EBF1). We performed Zdock protein docking analysis[42] of the effect of p.Arg382Trp on the HOMEZ–EBF1 interaction (Supplementary Fig. 16). We computationally constructed the homology structure of the HOMEZ and EBF1 complexes from the monomer structures of the HOMEZ homeobox domain (Protein Data Bank (PDB): 2ECC) and EBF1 IPT/TIG domain (PDB: 3MQI), respectively. According to the docking structure model with the best predicted score (Fig. 6c and Supplementary Fig. 16), Arg382 is located at the binding interface of HOMEZ–EBF1, forming one salt bridge and one hydrogen bond with Asp285 and Asn286 in EBF1, respectively. Interestingly, p.Arg382Trp disrupts the salt bridge and hydrogen bond and further alters surface topography due to the difference in both size and shape between Arg and Trp, which contributes to the binding free energy loss of the protein complex. By superimposing a homeobox DNA complex structure onto the HOMEZ–EBF1 complex model (Supplementary Fig. 16), we observed that HOMEZ contains two distinct binding interfaces of its homeobox domain that interact simultaneously with DNA and EBF1. Although p.Arg382Trp disrupts the interaction of HOMEZ and EBF1, it may also alter protein–DNA interaction.

We next focused on the RHOA–ARHGDIA interaction because it has an available co-crystal structure (Supplementary Fig. 17). In the RHOA–ARHGDIA system, the p.Pro75Ser substitution causes a shift in the secondary structure of the region. Using molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) to calculate interaction enthalpy, we observed a difference of >100 kJ mol⁻¹ incident in the mutant protein, indicating a significant loss of interaction energy inherent in the mutation and consistent with our experimental data (Fig. 6b,d). RHOA is a well-known oncogene product in which it was reported that multiple mutations were

Fig. 4 | Pharmacogenomics landscape of protein–protein interaction–perturbing alleles. a, Experimental design of pharmacogenomics predicted by PPI-perturbing alleles. **b,** Drug responses evaluated by putative oncoPPIs harboring a statistically significant excess number of missense mutations at PPI interfaces, following a binomial distribution across 66 selected anticancer therapeutic agents in cancer cell lines. Each node denotes a specific oncoPPI. Node size denotes *P* values computed by two-tailed ANOVA (Methods). Effect size was quantified by Cohen's statistic using the difference between two means divided by a pooled s.d. for the data. Node color denotes three different types of PPI (Fig. 1b). **c,** Drug responses evaluated by oncoPPIs in the PDX models. **d,e,** Highlighted examples of drug response (the area under the dose–response curve (AUC)) to encorafenib and its combinations (LEE011 and encorafenib) predicted by interface mutations on VCL–FXR1 (**d**; *n* = 3 mutated cell lines; *n* = 73 WT cell lines) and BRAF-MAP2K1 (**e**; *n* = 14 mutated cell lines; *n* = 19 WT cell lines). *P* values were calculated by using two-tailed ANOVA. Data are represented as a boxplot with an underlaid violin plot in which the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent IQR × 1.5 and beyond the whiskers are outlier points.

probably pathogenic in various types of cancer, including lymphoma and adenocarcinoma[43]. Its interaction with ARHGDIA is important for the inactivation and stabilization of RHOA. Loss of

the RHOA–ARHGDIA interaction could, therefore, lead to tumor cell proliferation and metastasis[44,45]. These observations suggest that p.Pro75Ser is a potential functional PPI-perturbing mutation
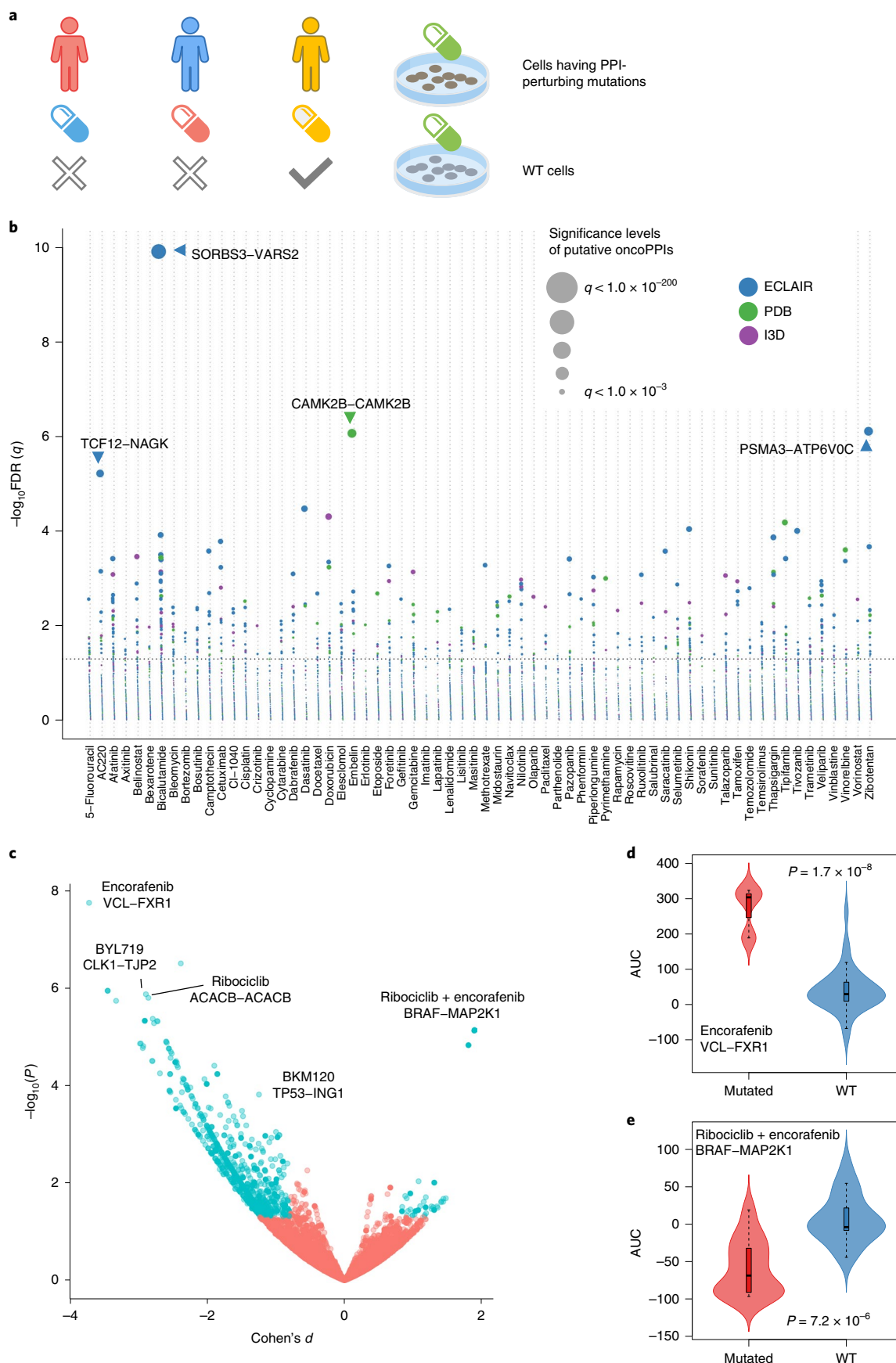
that alters the RHOA–ARHGDIA interaction in cancer cells. In summary, our experimental assays and computational biophysical analyses identify network perturbations by PPI-perturbing mutations that can potentially lead to the discovery of new pathogenic molecular mechanisms.

Functional validation of PPI-perturbing alleles. We next turned to functional validation using two selected systems: (1) the RXRA p.Ser427Phe alteration at the RXRA–PPARG interface and (2) the ALOX5 p.Met146Lys alteration at the ALOX5–MAD1L1 interface (Fig. 6b). RXRA is a member of the nuclear receptor superfamily and plays critical roles in pathologic processes of multiple diseases, including oncogenesis[46]. Our oncoPPI analysis revealed that p.Ser427Phe in RXRA plays a potential role in tumorigenesis, including pancreatic carcinogenesis (Fig. 7a). To demonstrate an oncogenic role of p.Ser427Phe in pancreatic cancer, we transfected the WT and p.Ser427Phe mutant RXRA into pancreatic cancer cells (Supplementary Fig. 18). We observed that p.Ser427Phe promoted tumor cell growth and clone formation in two pancreatic cancer cell lines, Capan-2 and SW1990 (Fig. 7b–d). It has been reported that p.Ser427Phe in RXRA simulated peroxisome proliferator-activated receptors (PPARs) to drive urothelial proliferation, and that a PPAR-specific antagonist can block mutant RXRA-driven cell proliferation[47]. To test this hypothesis, Capan-2 and SW1990 cells transfected with WT or RXRA p.Ser427Phe were treated with GSK0660, a potent PPARβ/δ antagonist. As shown in Fig. 7e, p.Ser427Phe-expressing Capan-2 cells are modestly susceptible to GSK0660 (half-maximal inhibitory concentration $(IC_{50}) = 1.11\,\mu M$) as compared to empty vector (EV, $IC_{50} = 8.41\,\mu M$) or WT ($IC_{50} = 2.51\,\mu M$)-transfected cells. A similar result was also obtained using the SW1990 cell line (EV, $IC_{50} = 6.99\,\mu M$; WT: $IC_{50} = 2.84\,\mu M$; p.Ser427Phe, $IC_{50} = 1.80\,\mu M$; Fig. 7e,f). Taken together, these data show that the RXRA–PPARG-perturbing alteration p.Ser427Phe promotes pancreatic cancer cell growth and sensitivity to PPAR antagonists.

ALOX5, a key enzyme in the biosynthesis of leukotrienes[48], plays roles in tumorigenesis and tumor progression[49]. Our Y2H assay showed that p.Met146Lys in ALOX5 (Fig. 7g) perturbed the physical interaction between ALOX5 and MAD1L1. To examine the functional role of p.Met146Lys in ALOX5 in cancer cell proliferation, we generated ALOX5 p.Met146Lys using standard site-directed mutagenesis (Supplementary Fig. 19). We next expressed WT and p.Met146Lys mutant ALOX5 in two lung cancer cell lines, H1299 and H460 (Supplementary Fig. 19). As shown in Fig. 7h–j, we found that p.Met146Lys significantly promotes cell proliferation and clone formation of cell lines H1299 and H460. Taken together, these experiments provide proof-of-concept evidence for the functional consequences of PPI-perturbing alleles in cancer.

Discussion

Previous studies have demonstrated that the human protein–protein interactome provides a powerful network-based tool to quantify disease–disease[6] and drug–disease[7–9] relationships; however, the functional network consequences of disease-associated mutations remain largely unknown. In this study, we developed a human structurally resolved macromolecular interactome framework for comprehensive identification of PPI-perturbing alleles in human disease. We showed widespread PPI network perturbations altered by both disease-associated germline and somatic mutations. Using a binomial statistical model, we identified 470 putative oncoPPIs harboring a statistically significant excess number of missense mutations at PPI interfaces (oncoPPIs) in pan-cancer analysis using a combination of factors, and validated the selected predictions experimentally. We demonstrated that network-predicted oncoPPIs were highly correlated with patient survival and drug resistance/sensitivity in human cancer cell lines, and also in patient-derived xenografts, offering actionable prognostic markers and pharmacogenomic biomarkers for potential clinical guidance. Taken together, these findings provide network-medicine-based, fundamental pathogenic molecular mechanisms and offer potential disease-specific targets for genotype-informed therapeutic discovery.

Our systematic network strategy provides a practical approach to identification of the potential, functional consequences of candidate disease alleles by perturbing PPI networks. PPI-perturbing mutations are significantly associated with poor survival rate in cancer patients, while mutations in the gene alone did not typically correlate with patient survival (Extended Data Fig. 2). PPI-perturbing mutations were significantly correlated with drug sensitivity or resistance, but mutations in a gene alone typically failed to predict drug responses (Fig. 4 and Supplementary Fig. 20). We found that proteins involved in oncoPPIs do not directly overlap with known drug targets (Supplementary Fig. 21). One possible explanation is that oncoPPIs influence downstream or upstream-network-associated protein targets of these drugs. In support of this view, we found that known drug targets did overlap with the neighbors of oncoPPIs (Supplementary Fig. 21) rather than with oncoPPIs directly, supporting the network proximity analysis of drug–disease relationships in the human interactome, as demonstrated in previous studies[8,9].

We found that gene expression of oncoPPIs is unlikely to be cancer type specific (Supplementary Fig. 22). This conclusion is consistent with our recent human interactome analysis showing no significant enrichment for PPIs between causal disease proteins and tissue-specific expressed proteins[39]. One possible explanation for this finding is that PPIs are more likely to be altered by somatic coding mutations that alter physical binding affinity. For example, we found that p.Met146Lys specifically perturbed the interaction between ALOX5 and MAD1L1 (Fig. 6b and Supplementary Fig. 19). Previous studies have shown low or no correlation between protein expression or activity and gene expression[50]. There are, of course, many factors that influence the correlation between protein expression or activities and messenger RNA abundance, including post-translational modification of proteins, RNA editing and others[50,51].

We acknowledge several potential limitations in the current study. Different tissue collection protocols and sequencing approaches, and variant calling and filtering approaches from TCGA, may generate the potential risk of a significant false-positive rate. Although we found the same level of enrichment for mutant interface residues using both crystal structures and analysis of the high-throughput systematic interactome identified by unbiased Y2H assays[23], some potential noise in the computationally inferred

---

Fig. 5 | Protein–protein interaction–perturbing alleles in histone H4 complex. a, A highlighted PPI-perturbing mutation network for the histone H4 complex in human cancer. **b**, Somatic mutation landscape of histone H4 complex across 18 selected cancer types with the highest somatic mutation rate. **c**, Selected PPI-perturbing mutations (highlighted in red) in the histone H4 complex. **d**, Interface mutations (highlighted in red) between histone H4 and DAXX. **e**, Interface mutations (MUT) of histone H4 complex are significantly correlated with survival in COAD and LUSC. *P* values were calculated by two-tailed log-rank test. **f**, Interface mutations of histone H4 complex are significantly correlated with anticancer drug responses, including paclitaxel (*n* = 16 mutated cell lines, *n* = 411 WT cell lines), BMC-754807 (an IGF-1R inhibitor) (*n* = 32 mutated cell lines, *n* = 895 WT cell lines) and EHT-1864 (a Rho inhibitor) (*n* = 36 mutated cell lines, *n* = 928 WT cell lines). *P* values were calculated by using two-tailed ANOVA. The data are represented as a boxplot with an underlaid violin plot in which the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent IQR × 1.5 and beyond the whiskers are outlier points. **c,d**, Protein 3D images were prepared by PyMOL.
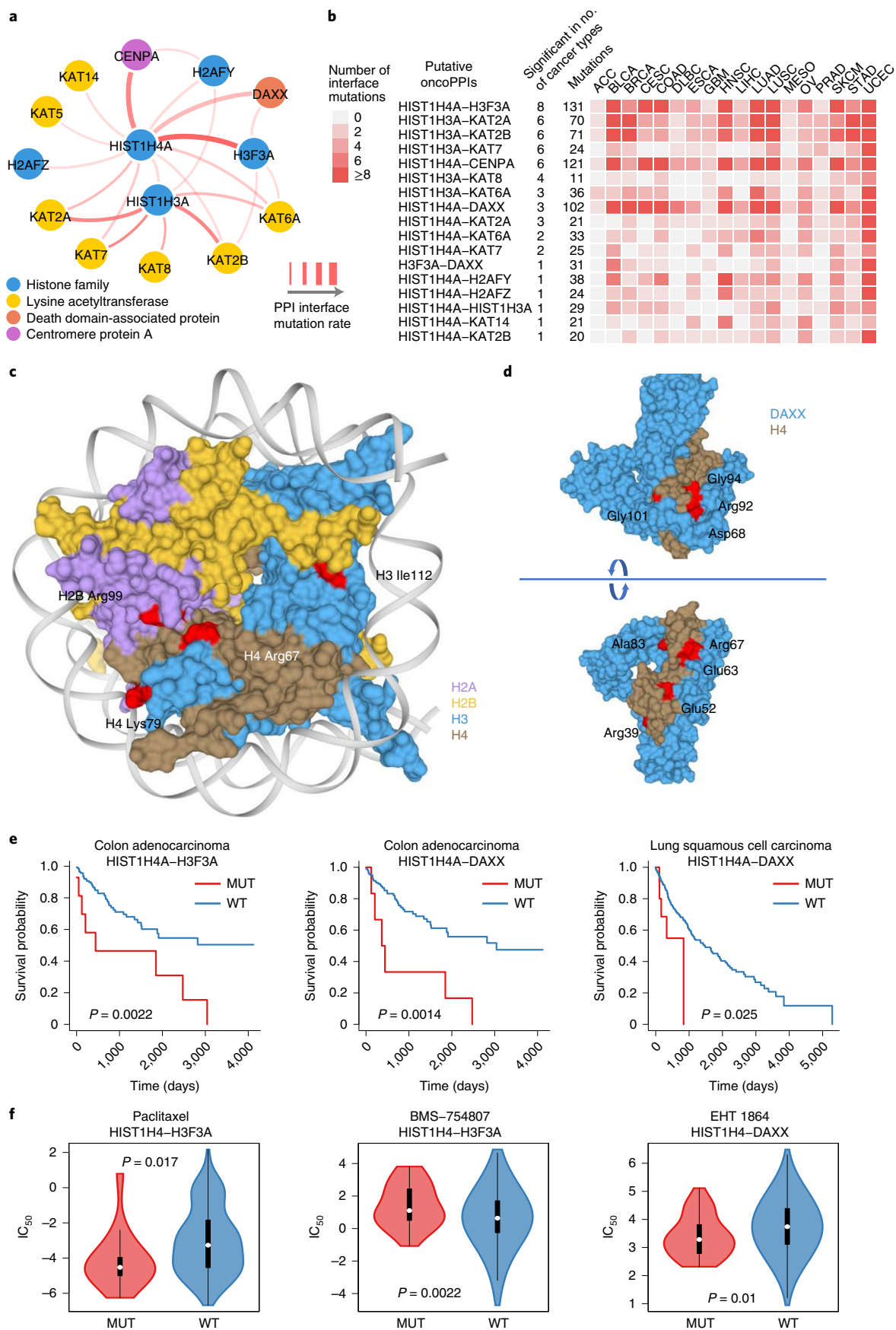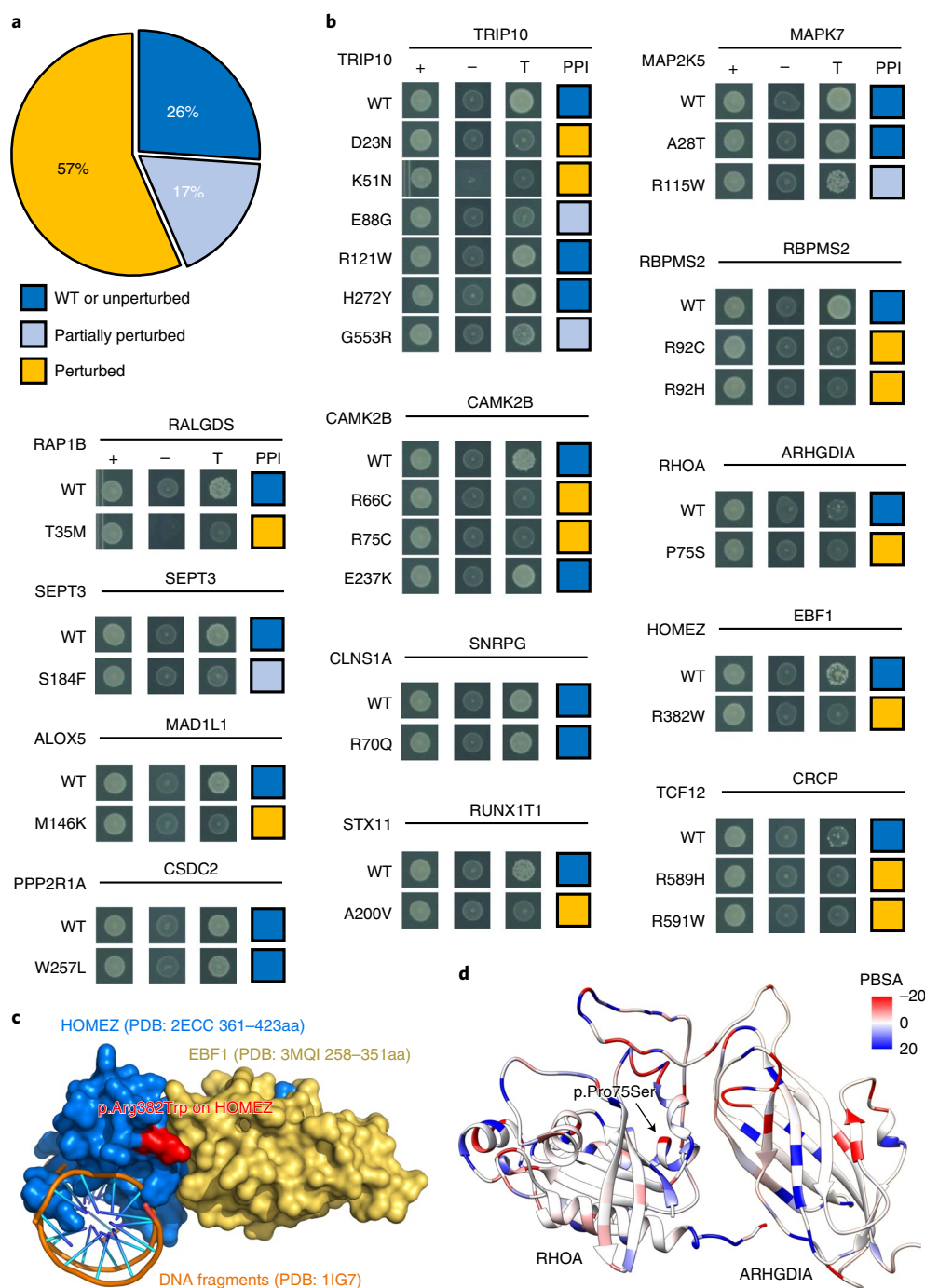
Fig. 6 | Experimental investigation of alleles with perturbed physical protein–protein interactions. a, Distribution of three types of mutational consequences on PPIs: unperturbed, partially perturbed and perturbed. **b**, Y2H readouts of oncoPPIs with and without mutations. +, selection for existence of AD and DB plasmids carrying ORFs for PPI testing; –, selection for autoactivators; T, selection for PPIs. Growth indicates interaction while no growth suggests no interaction (Methods and Supplementary Table 1). **c**, HOMEZ–EBF1 complex model location of the interface alteration p.Arg382Trp on HOMEZ. The complex model was built by Zdock protein docking simulation (Supplementary Fig. 16). **d**, Distribution of calculated binding affinity (PBSA) of the RHOA–ARHGDIA complex (PDB ID: 1CC0) directed by the p.Pro75Ser alteration on RHOA. Color bar indicates binding affinity (Methods), from high (blue) to low (red). **c,d**, Protein 3D images were prepared by PyMOL.

PPI interfaces may exist. Although we validated WT PPIs in Y2H assays using the PacBio sequencing approach (Methods), several WT PPI pairs (such as RHOA–ARHGDIA) showed a weak Y2H signal. Further experimental validation is, therefore, needed in the future.

We compiled a comprehensive, structurally resolved interactome network based on our sizeable efforts. Importantly, the incompleteness of the human interactome may limit coverage for unknown disease proteins or mutations. In the future, our approach may help direct facilitatation of biological interpretation of mutations and inform disease-driven PPI allele identification in multiple ongoing and future human genome sequencing efforts, including TopMed[52], PVDOMICS[53], International Cancer Genome Consortium[54], All of Us[55] and many others. Taken together, this work suggests that we
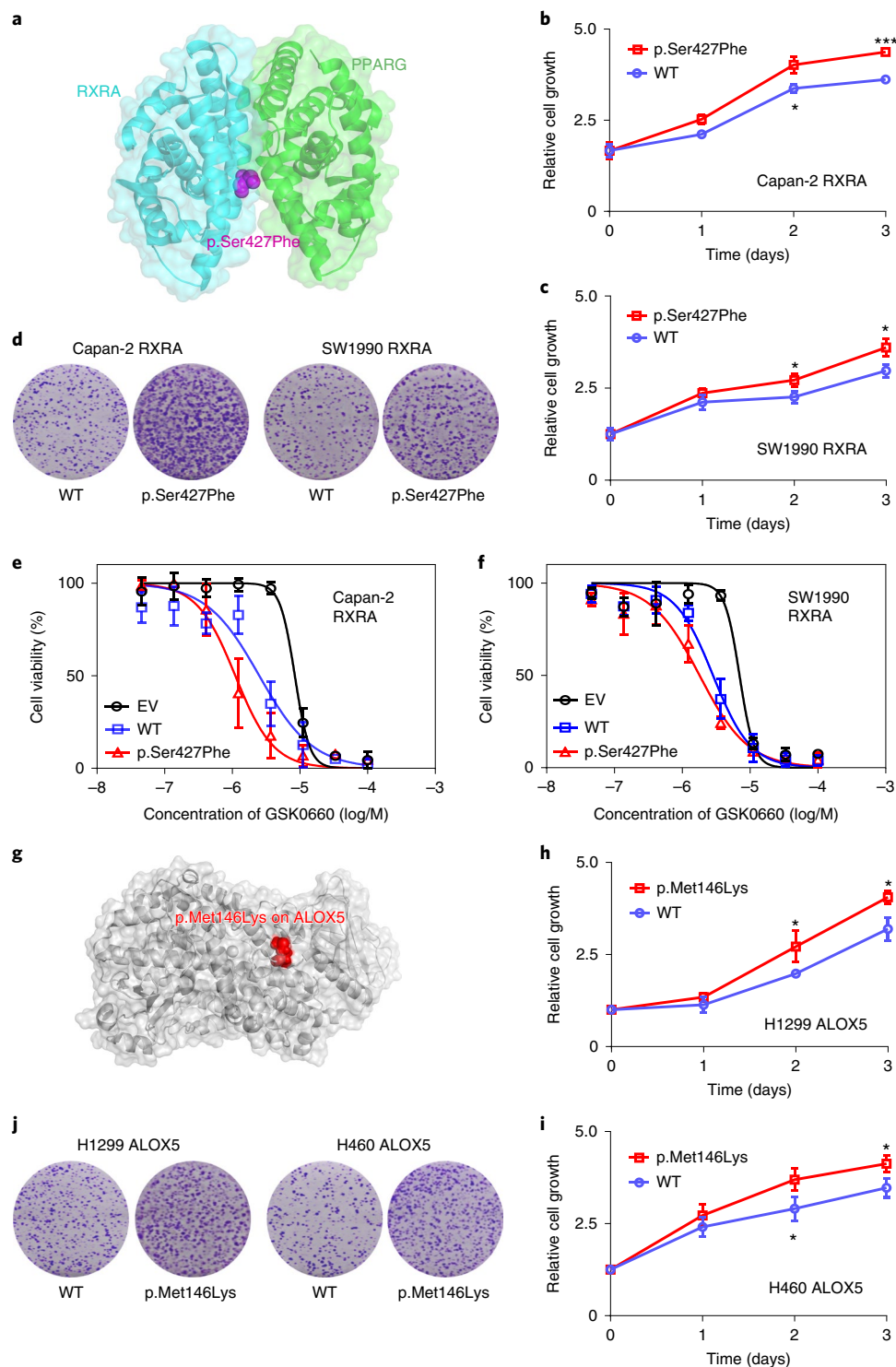
Fig. 7 | Mutants of RXRA and ALOX5 promote cancer cell growth. a, Structure of the RXRA–PPARG complex. **b,c**, Relative cell growth of Capan-2 (**b**) and SW1990 cells (**c**) transfected with pCDNA3-RXRA WT or pCDNA3-RXRA p.Ser427Phe. Cell proliferation was measured by MTS assay at 24-h intervals for 72 h. Graphs show the mean ± s.d. derived from three independent experiments. Two-tailed Student's t-test was used to test for statistical significance, *$P < 0.01$, ***$P < 0.001$. **d**, For the colony formation assay, cells were maintained in normal medium containing 10% FBS for 14 days then fixed and stained with crystal violet. **e,f**, Suppression of WT and mutant RXRA-driven cell proliferation by GSK0660, a potent PPARβ/δ antagonist with (log/M) defined as the $\log_{10}$ of the molar concentration of GSK0660. Capan-2 (**e**) and SW1990 cells (**f**) were transfected with pCDNA3 EV, pCDNA3-RXRA WT or pCDNA3-RXRA p.Ser427Phe, then treated with various concentrations of GSK0660 for 72 h. The graphs show mean ± s.d. derived from three independent experiments. **g**, An example of a perturbed allele, encoding p.Met146Lys, on an ALOX5 crystal structure (PDB ID: 3V98). **h,i**, Relative cell growth of H1299 (**h**) and H460 cells (**i**) transfected with pCDNA3-ALOX5 WT or pCDNA3-ALOX5 p.Met146Lys. Cell proliferation was measured by the MTS method at 24-h intervals for 72 h (Methods). The graphs show mean ± s.d. derived from three independent experiments. Student's t-test was used to test for statistical significance, *$P < 0.01$. **j**, For the colony formation assay, cells were maintained in normal medium containing 10% FBS for 14 days, then fixed and stained with crystal violet. Error bars denote s.d. of $n = 3$ independent experiments. **a,g**, Protein 3D images were prepared by PyMOL.

can minimize the translational gap between genomics and clinical medicine, and provide a clear path from network medicine to precision medicine in the process. If broadly applied, this human 3D interactome network analysis framework could prioritize actionable prognostic and pharmacogenomic biomarkers for personalized treatments and offer disease-allele-specific targets for genotype-informed protein–protein inhibitor discovery.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-00774-y.

References

1. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 (2018).
2. Cheng, F., Liang, H., Butte, A. J., Eng, C. & Nussinov, R. Personal mutanomes meet modern oncology drug discovery and precision health. *Pharmacol. Rev.* **71**, 1–19 (2019).
3. Sahni, N. et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).
4. Ng, P. K. et al. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* **33**, 450–462 (2018).
5. Chen, S. et al. An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nat. Genet.* **50**, 1032–1040 (2018).
6. Menche, J. et al. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
7. Cheng, F. et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* **9**, 2691 (2018).
8. Cheng, F., Kovacs, I. A. & Barabasi, A. L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 (2019).
9. Cheng, F. et al. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nat. Commun.* **10**, 3476 (2019).
10. Porta-Pardo, E., Garcia-Alonso, L., Hrabe, T., Dopazo, J. & Godzik, A. A pan-cancer catalogue of cancer driver protein interaction interfaces. *PLoS Comput. Biol.* **11**, e1004518 (2015).
11. Gao, J. et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* **9**, 4 (2017).
12. Niu, B. et al. Protein-structure-guided discovery of functional mutations across 19 cancer types. *Nat. Genet.* **48**, 827–837 (2016).
13. Tokheim, C. et al. Exome-scale discovery of hotspot mutation regions in human cancer using 3D protein structure. *Cancer Res.* **76**, 3719–3731 (2016).
14. Kamburov, A. et al. Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc. Natl Acad. Sci. USA* **112**, E5486–E5495 (2015).
15. Mosca, R. et al. dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods* **12**, 167–168 (2015).
16. Wang, X. et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* **30**, 159–164 (2012).
17. Rose, P. W. et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **45**, D271–D281 (2017).
18. Mosca, R., Ceol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **10**, 47–53 (2013).
19. Meyer, M. J. et al. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* **15**, 107–114 (2018).
20. Stenson, P. D. et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
21. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
22. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
23. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
24. Mullard, A. Nine paths to PCSK9 inhibition. *Nat. Rev. Drug Discov.* **16**, 299–301 (2017).
25. Pandit, S. et al. Functional analysis of sites within PCSK9 responsible for hypercholesterolemia. *J. Lipid Res.* **49**, 1333–1343 (2008).
26. Diedrich, B. et al. Discrete cytosolic macromolecular BRAF complexes exhibit distinct activities and composition. *EMBO J.* **36**, 646–663 (2017).
27. Zillhardt, M. et al. Foretinib (GSK1363089), an orally available multikinase inhibitor of c-Met and VEGFR-2, blocks proliferation, induces anoikis, and impairs ovarian cancer metastasis. *Clin. Cancer Res.* **17**, 4042–4051 (2011).
28. Wang, Y., Shi, J., Chai, K., Ying, X. & Zhou, B. P. The role of Snail in EMT and tumorigenesis. *Curr. Cancer Drug Targets* **13**, 963–972 (2013).
29. Raymond, J. R. Jr., Appleton, K. M., Pierce, J. Y. & Peterson, Y. K. Suppression of GNAI2 message in ovarian cancer. *J. Ovarian Res.* **7**, 6 (2014).
30. Gao, H. et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* **21**, 1318–1325 (2015).
31. Koelblinger, P., Thuerigen, O. & Dummer, R. Development of encorafenib for BRAF-mutated advanced melanoma. *Curr. Opin. Oncol.* **30**, 125–133 (2018).
32. Blessing, M. M. et al. Novel BRAF alteration in desmoplastic infantile ganglioglioma with response to targeted therapy. *Acta Neuropathol. Commun.* **6**, 118 (2018).
33. Zhang, J. et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* **45**, 602–612 (2013).
34. Tripathy, D., Bardia, A. & Sellers, W. R. Ribociclib (LEE011): mechanism of action and clinical impact of this selective cyclin-dependent kinase 4/6 inhibitor in various solid tumors. *Clin. Cancer Res.* **23**, 3251–3262 (2017).
35. Anczukow, O. et al. SRSF1-regulated alternative splicing in breast cancer. *Mol. Cell* **60**, 105–117 (2015).
36. Yan, G. et al. Selective inhibition of p300 HAT blocks cell cycle progression, induces cellular senescence, and inhibits the DNA damage response in melanoma cells. *J. Invest. Dermatol.* **133**, 2444–2452 (2013).
37. Urdinguio, R. G. et al. Chromatin regulation by histone H4 acetylation at lysine 16 during cell death and differentiation in the myeloid compartment. *Nucleic Acids Res.* **47**, 5016–5037 (2019).
38. Yuen, B. T. & Knoepfler, P. S. Histone H3.3 mutations: a variant path to cancer. *Cancer Cell* **24**, 567–574 (2013).
39. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
40. Wang, Y. et al. ALOX5 exhibits anti-tumor and drug-sensitizing effects in MLL-rearranged leukemia. *Sci. Rep.* **7**, 1853 (2017).
41. Tsukasaki, K. et al. Mutations in the mitotic check point gene, MAD1L1, in human cancers. *Oncogene* **20**, 3301–3305 (2001).
42. Pierce, B. G. et al. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).
43. Nagata, Y. et al. Variegated RHOA mutations in adult T-cell leukemia/lymphoma. *Blood* **127**, 596–604 (2016).
44. Liang, L. et al. Loss of ARHGDIA expression is associated with poor prognosis in HCC and promotes invasion and metastasis of HCC cells. *Int. J. Oncol.* **45**, 659–666 (2014).
45. Lu, W. et al. Downregulation of ARHGDIA contributes to human glioma progression through activation of Rho GTPase signaling pathway. *Tumour Biol.* **37**, 15783–15793 (2016).
46. Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors, RXR, and the big bang. *Cell* **157**, 255–266 (2014).
47. Halstead, A. M. et al. Bladder-cancer-associated mutations in RXRA activate peroxisome proliferator-activated receptors to drive urothelial proliferation. *eLife* **6**, e30862 (2017).
48. Radmark, O. & Samuelsson, B. 5-Lipoxygenase: mechanisms of regulation. *J. Lipid Res.* **50**, S40–S45 (2009).
49. Pidgeon, G. P. et al. Lipoxygenase metabolism: roles in tumor progression and survival. *Cancer Metastasis Rev.* **26**, 503–524 (2007).
50. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
51. Han, L. et al. The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell* **28**, 515–528 (2015).
52. Sarnowski, C. et al. Impact of rare and common genetic variants on diabetes diagnosis by hemoglobin A1c in multi-ancestry cohorts: the trans-omics for precision medicine program. *Am. J. Hum. Genet.* **105**, 706–718 (2019).
53. Hemnes, A. R. et al. PVDOMICS: a multi-center study to improve understanding of pulmonary vascular disease through phenomics. *Circ. Res.* **121**, 1136–1139 (2017).
54. Consortium, I. T. P.-C. Ao. W. G. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
55. Reardon, S. Giant study poses DNA data-sharing dilemma. *Nature* **525**, 16–17 (2015).

Methods

Building the human protein–protein interactome. To build a comprehensive human binary protein–protein interactome, we assembled three types of experimental evidence: (1) PPIs with crystal structures from the RCSB protein data bank (https://www.rcsb.org)[17]; (2) PPIs with homology modeling structures from Interactome3D (v.2017.12, https://interactome3d.irbbarcelona.org)[18]; and (3) experimentally determined binary PPIs with computationally predicted interface residues from Interactome INSIDER (v.2018.3, http://interactomeinsider.yulab.org)[19]. For crystal structures and homology models of PPIs, any residue that is at the surface of a protein (≥15% exposed surface) and whose solvent-accessible surface area decreases by ≥1.0 Å[2] in complex is considered to be at the interface. In addition, we also assembled computationally predicted interfaces using the ECLAIR classifier for experimentally identified PPIs from Interactome INSIDER[19]. Genes were mapped to their Entrez ID based on the NCBI database (https://www.ncbi.nlm.nih.gov)[56], as well as on their official gene symbols based on GeneCards (http://www.genecards.org/). The resulting human binary interactome constructed in this way includes 121,575 PPIs (edges or links) connecting 15,046 unique proteins (nodes). All PPIs are experimentally validated PPIs derived from different types of experimental evidence, as described in the original study[19].

Collection and preparation of genome sequencing data. We downloaded the tumor–normal pairwise somatic mutation data for patients from TCGA GDC Data Portal (https://portal.gdc.cancer.gov)[57] using the R package TCGA-assembler (v.2, http://www.compgenome.org/TCGA-Assembler/)[58] for 33 cancer types/subtypes. These 33 major cancer types include acute myeloid leukemia (LAML), adrenocortical carcinoma (ACC), BLCA, BRCA, cervical carcinoma (CESC), cholangiocarcinoma (CHOL), colon and rectal adenocarcinoma (COAD/READ), diffuse large B-cell lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma (GBM), HNSC, kidney chromophobe carcinoma, kidney renal clear cell carcinoma, kidney papillary cell carcinoma, low grade glioma, liver hepatocellular carcinoma (LIHC), LUAD, LUSC, mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic ductal adenocarcinoma (PAAD), paraganglioma and pheochromocytoma (PCPG), prostate adenocarcinoma (PRAD), sarcoma (SARC), rectal adenocarcinoma (READ), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA), testicular germ cell cancer (TGCT), thymoma (THYM), UCEC, uterine carcinosarcoma (UCS) and uveal melanoma (UVM).

Disease-associated missense mutations were downloaded from HGMD (http://www.hgmd.cf.ac.uk/ac/index.php)[20]. Population-based missense mutations were obtained from the databases 1000 Genomes (phase 3, 2,504 individuals, https://www.internationalgenome.org)[21] and ExAC (v.0.3.1, 60,706 individuals, https://gnomad.broadinstitute.org)[22]. We downloaded putative somatic mutations for 1,001 cancer cell lines from GDSC (http://www.cancerrxgene.org/). The list of genomic variants found in these cell lines by whole-exome sequencing was also obtained from GDSC. The sequencing variants were identified by comparison to a reference genome. The resulting variants were then filtered using data from the NHLBI GO Exome Sequencing Project and the 1000 Genomes Project to remove sequencing artifacts and germline variants[59]. In addition, we used ANNOVAR (https://doc-openbio.readthedocs.io/projects/annovar/en/latest/)[60] to map these somatic mutations to identify the corresponding amino acid changes via RefSeq ID. The functional impact of nonsynonymous single-nucleotide variants was measured by both SIFT[61] and PolyPhen-2 scores[62]. For this analysis, we obtained SIFT and PolyPhen-2 scores from the ANNOVAR annotation database. We then converted RefSeq ID to UniProt ID using a UniProt ID mapping tool (http://www.uniprot.org/uploadlists/).

Significance test of PPI interface mutations. A PPI in which there is significant enrichment in interface mutations in one or the other of the two protein-binding partners across individuals will be defined as an oncoPPI. For each gene g_i and its PPI interfaces, we assume that the observed number of mutations for a given interface follows a binomial distribution, binomial (T, p_{g_i}), in which T is the total number of mutations observed in one gene and p_{g_i} is the estimated mutation rate for the region of interest under the null hypothesis that the region was not recurrently mutated. Using length (g_i) to represent the length of the protein product of gene g_i, for each interface, we computed the P value—the probability of observing $>k$ mutations around this interface out of T total mutations observed in this gene—using the following equation:

$$P(X \geq k) = 1 - P(X < k) = 1 - \sum_{x=0}^{k-1} \binom{T}{x} p_{g_i}^x \left(1 - p_{g_i}\right)^{T-x} \quad (1)$$

in which $p_{g_i} = \frac{\text{length of interface}}{\text{length}(g_i)}$. Finally, we set the minimal P value across all the interfaces in a specific protein as the representative P value of its coding gene g_i, denoted $P(g_i)$. The significance of each PPI is defined as the product of P values of the two proteins (gene products). All P values were adjusted for multiple testing using the Bonferroni correction.

Cancer cell line annotation. We downloaded the annotation file of the cancer cell lines: molecular and drug-response data availability, microsatellite instability (MSI) status, growth properties and media, and TCGA and COSMIC tissue classification, from GDSC (http://www.cancerrxgene.org/). The details of this annotation have been described previously[59].

Drug sensitivity data. Natural log half-maximal inhibitory concentration (IC$_{50}$) and the area under the dose–response curve (AUC) values for all screened cell line–drug combinations were downloaded from GDSC. After applying the data preparation procedure described in a previous study[59], a total of 251 drugs tested in 1,074 cancer cell lines with 224,510 data points were used. In addition, we collected anticancer drug response data from in vivo compound screens between ~1,000 PDXs and 62 treatments across six indications[30].

ANOVA model. For each drug, we constructed a drug-response vector consisting of n IC$_{50}$ values from treatment of n cell lines. Next, a drug-response vector was modeled as a linear combination of the tissue of origin of the cell lines, screening medium, growth properties and the status of a genomic feature:

$$\text{IC}_{50} = \text{Mut} + \text{Tissue} + \text{Medium} + \text{MSI} \quad (2)$$

where Mut is mutations and MSI is microsatellite instability (including small indels).

In this study, considering the data sparsity, we performed only pan-cancer analysis. A genomic feature–drug pair was tested only if the final drug-response vector contained at least three positive and at least three negative cell lines. The effect size was quantified through Cohen's d statistic using the difference between two means divided by a pooled s.d. for the data. The resulting P values were corrected by the Benjamini–Hochberg method[63]. All statistical analyses were performed using the R package v.3.2.3 (http://www.r-project.org/) and two-tailed tests.

Pathway enrichment analysis. We used ClueGO[64] for enrichment analysis of genes in the canonical Kyoto Encyclopedia of Genes and Genomes pathways. A hypergeometric test was performed to estimate statistical significance, and all P values were adjusted for multiple testing using the Bonferroni correction (adjusted P values).

Cloning of disease mutations. We generated the predicted disease mutants by implementing a site-directed mutagenesis pipeline as described in the pairwise test for identification of perturbed interactions section (described below). For each mutation, two 'primary PCRs' were performed to generate DNA fragments containing the mutation, and a 'stitch PCR' was performed to fuse the two fragments to obtain the mutated ORF. For primary PCRs, two universal primers (E2E forward and E2E reverse) and two ORF-specific internal forward and reverse primers were used. The two ORF-specific primers contained the desired nucleotide change. The fragments generated by the primary PCRs were fused by the stitch PCR using the universal primers to generate the mutated ORF. The final product was a full-length ORF containing the mutation of interest. All mutated ORFs were cloned into a Gateway donor vector, pDONR223, by BP reaction followed by bacterial transformation and selection using spectinomycin. Two single colonies were picked for each transformant. All picked colonies were transferred into pDEST-AD and pDEST-DB by LR reaction followed by bacterial transformation and selection using ampicillin. The plasmids were then extracted, purified and transformed into Y8930 yeast strain for pairwise testing.

Pairwise test for identification of perturbed interactions. The pairwise test was performed in 96-well format. The ORFs were inoculated in SC-Leu and SC-Trp media overnight and mated in YEPD media the following day. All WT and mutant alleles in pDEST-DB were mated with their interacting partner in pDEST-AD (DB-ORFxAD-ORF), as well as in pDEST-AD but without insertion of the ORF (DB-ORFxAD-empty). After incubation at 30 °C overnight, mated yeasts were transferred to SC-Leu-Trp medium to select for diploids. The following day, diploid yeasts were spotted on SC-Leu-Trp-His + 1 mM 3AT and SC-Leu-Trp medium, to control for mating success.

After 3 days of growth at 30 °C, each spot on plates was scored with a growth score ranging from 0 to 4, 0 being no growth, 1 being one or two colonies, 2 being some colonies, 3 being many colonies and 4 being a large consolidated spot in which no individual colonies could be distinguished. Pairs for which the SC-Leu-Trp spot was scored as 3 or 4, and the 3AT spot was valid (yeasts were spotted and no contamination or other experimental failure), were considered as being successfully tested. A successfully tested pair can be further classified as positive, negative or autoactivator depending on the growth scores of DB-ORFxAD-ORF and DB-ORF-AD-empty on SC-Leu-Trp-His + 1 mM 3AT plates. If the growth score of DB-ORFxAD-ORF was 0, the pair was classified as negative; if the growth score of DB-ORFxAD-ORF – DB-ORFxAD-empty was ≥2, the pair was classified as positive (Supplementary Table 1); otherwise the pair was classified as autoactivator. Pairs were scored blindly with respect to their identity using in-house software.

In parallel, we made lysates of all SC-Leu-Trp plates to perform duplex PCR using barcoded AD/DB and Term primers, followed by pooling and sequencing with the PacBio Sequel system. We used the SMRT tools (v.5.1.0) and ISO-SEQ

(v.3.1) software packages to analyze raw sequencing results. The pipeline includes five main steps to obtaining high-quality sequences: (1) generation of circular consensus (CCS) reads; (2) demultiplexing and primer removal; (3) classification of full-length CCS reads; (4) clustering of full-length non-chimeric reads; and (5) polishing of cluster sequences. Polished sequences were then aligned to ORF sequences using BLAST. Colonies with the exact full-length sequence as expected (with, and only with, the expected mutations, fully covered by polished reads) were considered as sequence confirmed.

Only pairs (1) successfully tested, (2) classified as positive or negative, (3) for which the WT allele was classified as positive with growth score ≥2 and (4) that were sequence confirmed were considered for all further analysis. An interaction was considered perturbed by an allele if its growth score was ≤1 and that was less than the growth score of the corresponding WT pair by at least 2. Otherwise, an interaction was considered partially perturbed by an allele if the growth score of the WT pair was greater than the growth score for that interaction with the respective allele by 1.

System construction for molecular simulation. Crystal structures (PDBs: 1CC0 and 3M0C) were accessed from the RCSB PDB protein data bank. Co-crystalized ions were retained from the structure. Nonterminal missing loops were reconstructed, where required, using Modeller9.18 within UCSF Chimera. Protonation states for charged residues were determined using PROPKA 2.0. Mutations and preparation of the system for molecular dynamics simulation were accomplished using the quick molecular dynamics simulator module of CHARMM-GUI. Following a processing step, including the addition of hydrogens and patching the terminal regions, a water box using TIP3 water molecules with edges at least 12 Å from the protein was added. The system was neutralized to a NaCl concentration of 150 mM.

Simulation parameters. Molecular dynamics simulations were carried out using GROMACS (v.2018.2)[7] on the Pitzer computing cluster at the Ohio Supercomputer Center. Initial minimizations of the systems were carried out using steepest descent until the energy of the system reached machine precision. Following minimization, an NVT (i.e., constant number of particles, volume and temperature) equilibration step with positional restraints of 400 kJ mol^{-1} nm^{-2} on backbone atoms and 40 kJ mol^{-1} nm^{-2} on side chain atoms was run using a timestep of 2 fs for 500,000 steps, yielding 1 ns of equilibration. Finally, NPT (i.e., constant number of particles, pressure and temperature) dynamics were run with no positional restraints for 400 ns using the same 2-fs timestep from equilibration, after which the system was determined by its root-mean-squared deviation (RMSD) to be reasonably well equilibrated.

Hydrogen atoms were constrained using the a linear constraint solver algorithm implemented in GROMACS (v.2018.2)[7]. Temperature coupling to 310.15 °K was done separately for protein and water/ions using a Nose–Hoover thermostat and a 1-ps coupling constant. For NPT dynamics simulation, isotropic pressure coupling to 1 bar was done using a Parrinello–Rahman barostat with a coupling constant of 5.0 ps and compressibility of 4.5×10^{-5} bar^{-1}. The pair list cutoff was constructed using the Verlet scheme, updated every 20 evaluations with a cutoff distance of 12 Å. Particle mesh Ewald electrostatics were chosen to describe coulombic interactions using the same cutoff as in the pair list. Van der Waals forces were smoothly switched to zero between 10 and 12 Å using a force-switch modifier to the cutoff scheme.

Postprocessing and RMSD plots were generated using standard GROMACS tools. MM/PBSA energies were calculated on 1,001 frames over the final 100 ns of each simulation using g_mmpbsa, which uses an adaptive Poisson–Boltzmann solver to determine polar and nonpolar contributions to binding energy. Briefly, binding free energy can be expressed as

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{protein1}} + G_{\text{protein2}}) \quad (3)$$

where complex refers to the protein–protein complex and proteins 1 and 2 the respective proteins in the complex. The individual free energies for each component above are determined by

$$G_x = \langle E_{\text{MM}} \rangle + \langle G_{\text{solvation}} \rangle - \text{TS} \quad (4)$$

where E_{MM} is vacuum molecular mechanics energy, $G_{\text{solvation}}$ the solvation energy and TS the entropic contribution. Entropic contributions were not included owing to computational cost and evidence that the inclusion of the entropy term does not always improve the accuracy of calculations. Molecular mechanics energy and solvation energy can be further broken down into their component energies:

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{nonbonded}} = E_{\text{bonded}} + E_{\text{vdW}} + E_{\text{elec}} \quad (5)$$

$$G_{\text{solvation}} = G_{\text{polar}} + G_{\text{nonpolar}} \quad (6)$$

Here E_{bonded} is 0, since we used the single trajectory approach. E_{vdW} and E_{elec} are the van der Waals and electrostatic contributions to vacuum binding, respectively, while G_{polar} and G_{nonpolar} are electrostatic and nonelectrostatic contributions to solvation energy, respectively.

Expression vector construction. pCDNA3-RXRA was generated using standard molecular cloning methods. pcDNA3-ALOX5 was kindly provided by C. D. Funk (Department of Biochemistry, Queen's University, Canada). Site-directed mutagenesis were performed using the KOD-Plus-Mutagenesis Kit (TOKOYO, catalog no. SMK-101) according to the manufacturer's instructions. RXRA p.Ser427Phe and ALOX5 p.Met146Lys were generated from vectors pCDNA3-RXRA and pcDNA3-ALOX5, respectively. All generated plasmids were confirmed by Sanger sequencing.

Cell culture and transfection. Human cancer cell lines (Capan-2, SW1990, H1299 and H460) were obtained from the American Type Culture Collection. All cells were cultured in DMEM (Gibco, catalog no. 11995040) supplemented with 10% fetal bovine serum (FBS; Gibco, catalog no. 10099-141) and maintained under an atmosphere containing 5% CO_2 at 37 °C. All cell lines were negative for mycoplasma. Pancreatic cancer cell lines (Capan-2 and SW1990) were transfected with EV, pcDNA3-RXRA WT or pcDNA3-RXRA p.Ser427Phe, and lung cancer cell lines (H1299 and H460) were transfected with EV, pcDNA3-ALOX5 WT or pcDNA3-ALOX5 p.Met146Lys using Lipofectamine 2000 (Invitrogen, catalog no. 11668019).

Cell proliferation assay. Cell viability was determined using CellTiter 96 AQueous Non-Radioactive Cell Proliferation Assay (MTS; Promega, catalog no. G5421) according to the manufacturer's recommendation. In brief, treated cancer cells were seed into 96-well plates at a density of 3,000–5,000 cells per well and incubated for the indicated time. Next, 20 µl per well of combined phenazine methosulfate solution (PMS) was added and absorbance was recorded at 490 nm using the microplate reader Synergy 2 (BioTek).

Immunoblotting. Cells were lysed with RIPA lysis buffer (20 mM Tris-HCl, 37 mM NaCl$_2$, 2 mM EDTA, 1% Triton-X, 10% glycerol, 0.1% SDS and 0.5% sodium deoxycholate) with protease and phosphatase inhibitors (Roche). Protein samples were quantified (Pierce BCA Protein Assay Kit, Thermo Fisher Scientific), subjected to SDS–polyacrylamide gel electrophoresis and transferred to polyvinylidene difluoride membranes. Membranes were incubated with primary antibodies, including RXRA (1:1,000; Proteintech, catalog no. 21218-1-AP) and ALOX5 (1:1,000; Abclonal, catalog no. A2877) and subsequent secondary antibody: horseradish peroxidase–conjugated Affinipure Goat Anti-Rabbit IgG(H+L) (1:3,000; Proteintech, catalog no. SA00001-2).

Colony formation assay. Transfected cells were seeded into six-well plates at a density of 3,000 cells per well in 2 ml of DMEM medium supplemented with 10% FBS. The medium was replaced every 3 days. After 14 days, viable colonies were fixed in 4% paraformaldehyde and stained with 0.1% crystal violet at room temperature. Formed colonies were photographed with an inverted fluorescence microscope (Olympus).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All mapping interface mutations, network-predicted oncoPPIs across pan-cancer and 33 individual cancer types, the human protein–protein interactome and predicted drug responses and patient survival analysis are freely available at the websites https://mutanome.lerner.ccf.org/ and https://github.com/ChengF-Lab/oncoPPIs. Publicly available databases used in the present study include the RCSB protein data bank (https://www.rcsb.org), Interactome3D (v.2017.12, https://interactome3d.irbbarcelona.org), Interactome INSIDER (v.2018.3, http://interactomeinsider.yulab.org), GeneCards (http://www.genecards.org/), NCBI (https://www.ncbi.nlm.nih.gov), TCGA GDC Data Portal (https://portal.gdc.cancer.gov), HGMD (http://www.hgmd.cf.ac.uk/ac/index.php), 1000 Genomes (phase 3, 2,504 individuals, https://www.internationalgenome.org), ExAC database (v.0.3.1, 60,706 individuals, https://gnomad.broadinstitute.org) and GDSC (http://www.cancerrxgene.org/).

Code availability

All codes written for and used in this study are available from https://github.com/ChengF-Lab/oncoPPIs.

References

56. Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
57. Grossman, R. L. et al. Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).
58. Zhu, Y., Qiu, P. & Ji, Y. TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* **11**, 599–600 (2014).
59. Iorio, F. et al. A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754 (2016).

60. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
61. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
62. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
63. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).
64. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).

Author contributions

J.L. and F.C. conceived the study. F.C., J.Z., Y.W., W.L. and Z.L. performed experiments and data analysis. Y.Z., W.R.M., H.Y., J.H., J.M., R.W., T.H., D.E.H., J.A.C., J.F., Y.H., J.D.L., R.A.K., F.C.L., E.M.A., R.R., C.E. and M.V. interpreted the data analysis. F.C., J.Z., Y.W., W.L. and J.L. drafted the manuscript and critically revised the manuscript. All authors critically revised and gave final approval of the manuscript.

Competing interests

J.L. is the scientific cofounder of Scipher Medicine, Inc., a startup company that uses network medicine to identify biomarkers for disease and specific pathway targets for drug development. M.V. is a shareholder and scientific advisor of seqWell, Inc. and founder and scientific advisor of Gloucester Marine Genomics Institute, Inc. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-00774-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-020-00774-y.

Correspondence and requests for materials should be addressed to J.L.

Peer review information *Nature Genetics* thanks Leng Han, Ulrich Stelzl and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
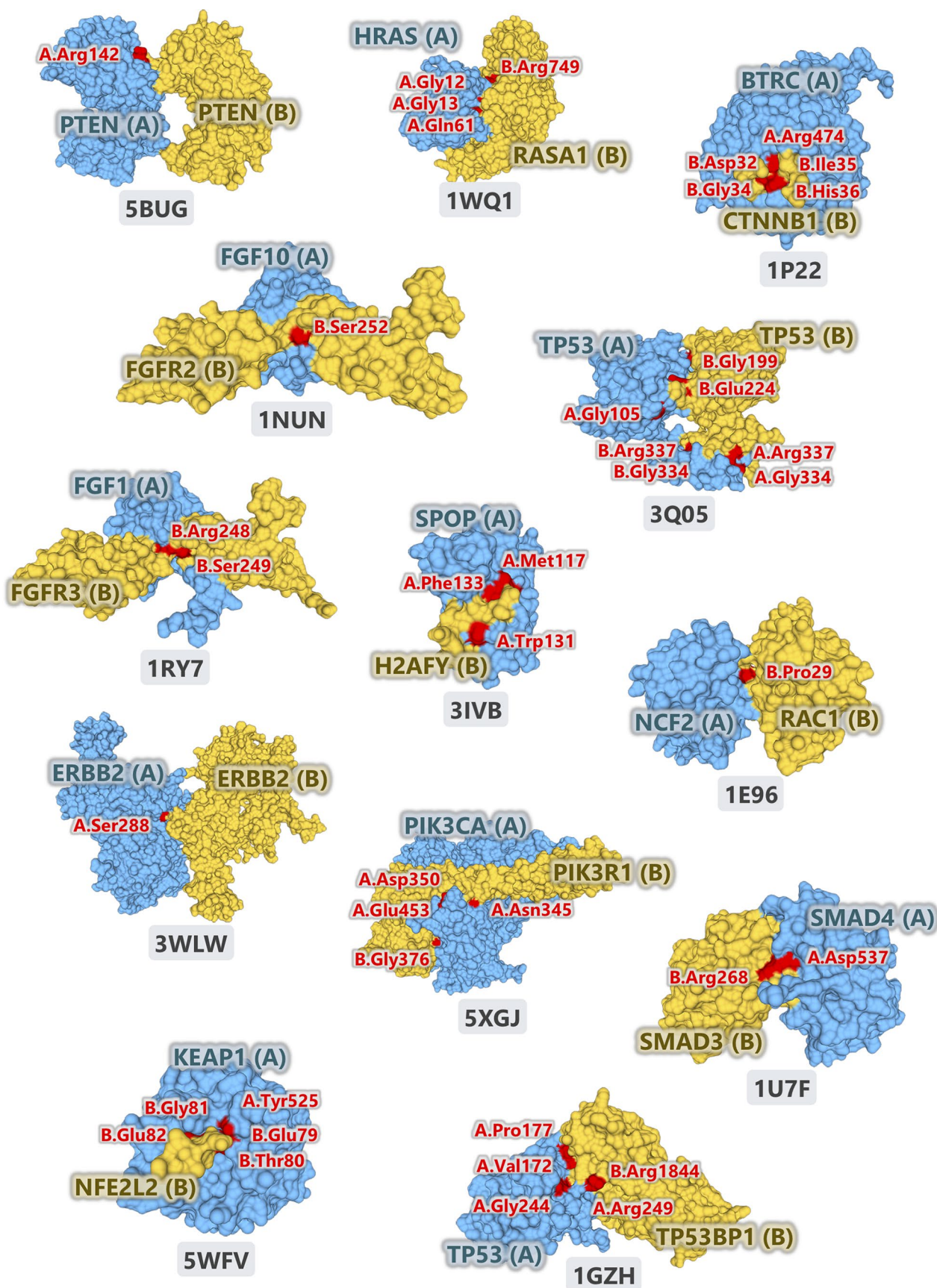
Extended Data Fig. 1 | The 13 selected pan-cancer oncoPPIs with crystal structure-based PPI interface mutations. The images were prepared by PyMOL (https://pymol.org/2/) using the Protein Data Bank (PDB) IDs (highlighted in figures) downloaded from PDB database (https://www.rcsb.org). Structural views of all oncoPPIs in pan-cancer and individual cancer types/subtypes are freely available: https://mutanome.lerner.ccf.org/.
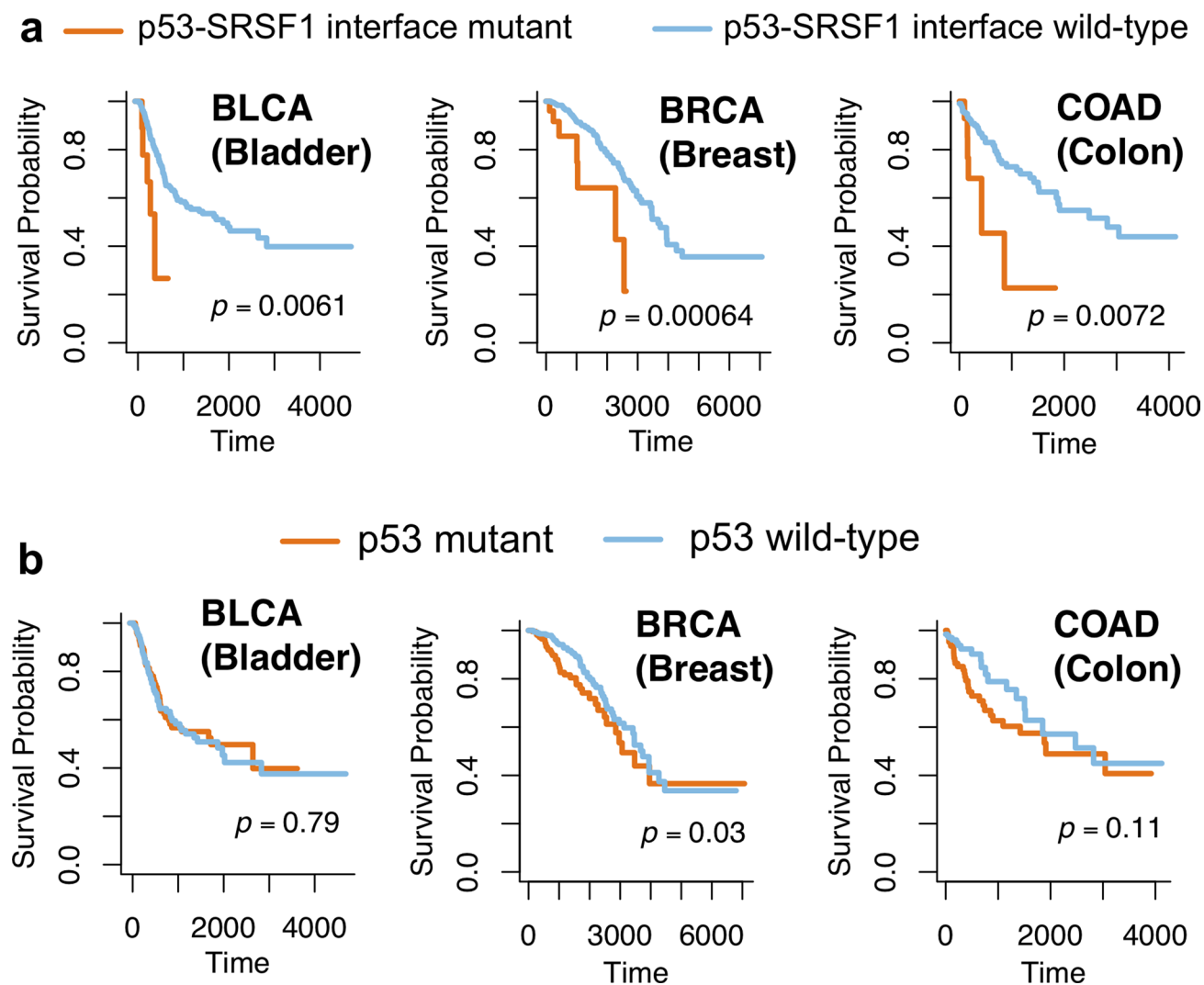
Extended Data Fig. 2 | Survival analyses of p53-SRSF1 PPI perturbing-mutations and p53 mutations alone. Three exemplary cancer types, including bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), and colon adenocarcinoma (COAD), are illustrated. Survival analyses of p53-SRSF1 PPI perturbing-mutations across other cancer types/subtypes are provided in Supplementary Fig. 14. The p-value (P) was computed by log-rank test. All oncoPPI-predicted survival analyses for 33 cancer types/subtypes are freely available at the following website: https://mutanome.lerner.ccf.org/.

nature research

Corresponding author(s):   Joseph Loscalzo

Last updated by author(s):   Dec 18, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

We downloaded the tumor-normal pairwise somatic mutation data for patients from TCGA GDC Data Portal using R package TCGA-assembler (v2, http://www.compgenome.org/TCGA-Assembler/). Population-based missense mutations were obtained from the 1000 Genomes Project (phase 3, 2,504 individuals) and from ExAC database (v0.3.1, 60,706 individuals). We downloaded putative somatic mutations for 1,001 cancer cell lines from the Genomics of Drug Sensitivity in Cancer (GDSC, http://www.cancerrxgene.org/). The resulting variants were then filtered using the data from NHLBI GO Exome Sequencing Project and the 1000 Genomes Project to remove sequencing artefacts and germline variants. We assembled three types of experimental PPI evidences: (1) PPIs with crystal structures from the RCSB protein data bank (https://www.rcsb.org), (2) PPIs with homology modeling structures from Interactome3D (https://interactome3d.irbbarcelona.org), and (3) experimentally determined binary PPIs with computationally predicted interface residues from Interactome INSIDER (v2018.3, http://interactomeinsider.yulab.org).

Data analysis

We used ANNOVAR (v4.0, https://doc-openbio.readthedocs.io/projects/annovar/en/latest/) to map somatic mutations in the protein sequences for identifying the corresponding amino acid changes via RefSeq ID. The functional impact of nonsynonymous SNVs (single nucleotide variants) was measured by both SIFT and PolyPhen-2 scores. We then converted RefSeq ID to UniProt ID using a UniProt ID mapping tool (http://www.uniprot.org/uploadlists/). The effect size was quantified through Cohen's d statistic using the difference between two means divided by a pooled standard deviation for the data. All statistical analyses were performed using the R package (v3.2.3, http://www.r-project.org/). We used ClueGO for enrichment analysis of genes in the canonical KEGG pathways. Molecular dynamics simulations were carried out using GROMACS v2018.2 on the Pitzer computing cluster at the Ohio Supercomputer Center. Hydrogen atoms were constrained using the a LINear Constraint Solver (LINCS) algorithm implemented in GROMACS (v2018.2) We used SMRT tools (v5.1.0) and ISO-SEQ (v3.1) software packages to analyze raw sequencing results. The protein 3D images were prepared by PyMOL (https://pymol.org/2/). Violin plots were performed by R/ggplot2.

Code availability. All codes written for and used in this study are available from https://github.com/ChengF-Lab/oncoPPIs and other codes are available from the corresponding author upon reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Data availability. All mapping interface mutations, network-predicted oncoPPIs across pan-cancer and 33 individual cancer types, the human protein-protein interactome, and predicted drug responses and patient survival analysis are freely available at the website: https://mutanome.lerner.ccf.org/ and https://github.com/ChengF-Lab/oncoPPIs. Publicly available databases used in the present study include the RCSB protein data bank (https://www.rcsb.org), Interactome3D (v2017.12, https://interactome3d.irbbarcelona.org), Interactome INSIDER (v2018.3, http://interactomeinsider.yulab.org), GeneCards (http://www.genecards.org/), NCBI (https://www.ncbi.nlm.nih.gov), TCGA GDC Data Portal (https://portal.gdc.cancer.gov), HGMD (http://www.hgmd.cf.ac.uk/ac/index.php), 1000 Genomes (phase 3, 2,504 individuals, https://www.internationalgenome.org), ExAC database (v0.3.1, 60,706 individuals, https://gnomad.broadinstitute.org), GDSC (http://www.cancerrxgene.org/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was based on the number of available, qualified tumor samples for this study.
Data exclusions	No data were excluded in our analysis.
Replication	All attempts at replications are successful for all computational work in our manuscript. Three independent experiments (replications) were performed in cell-based functional assays.
Randomization	Randomization is not relevant to our computational analysis.
Blinding	Blinding is not relevant to our computational analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies: RXRA (1:1000, Proteintech, Cat. 21218-1-AP) and ALOX5 (1:1000, Abclonal, Cat. A2877). Secondary antibody: HRP-conjugated Affinipure Goat Anti-Rabbit IgG(H+L) (1:3000, Proteintech, Cat. SA00001-2).
Validation	All antibodies were validated by the manufacturers prior to being used. RXRA: http://www.ptgcn.com/products/RXRA-Antibody-21218-1-AP.htm, ALOX-5: https://abclonal.com.cn/catalog/A2877, Secondary antibody:

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human cancer cell lines (Capan-2, SW1990, H1299 and H460) were obtained from American Type Culture Collection (ATCC).
Authentication	All cell lines were authenticated by STR profiling before used
Mycoplasma contamination	All cell lines were negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.