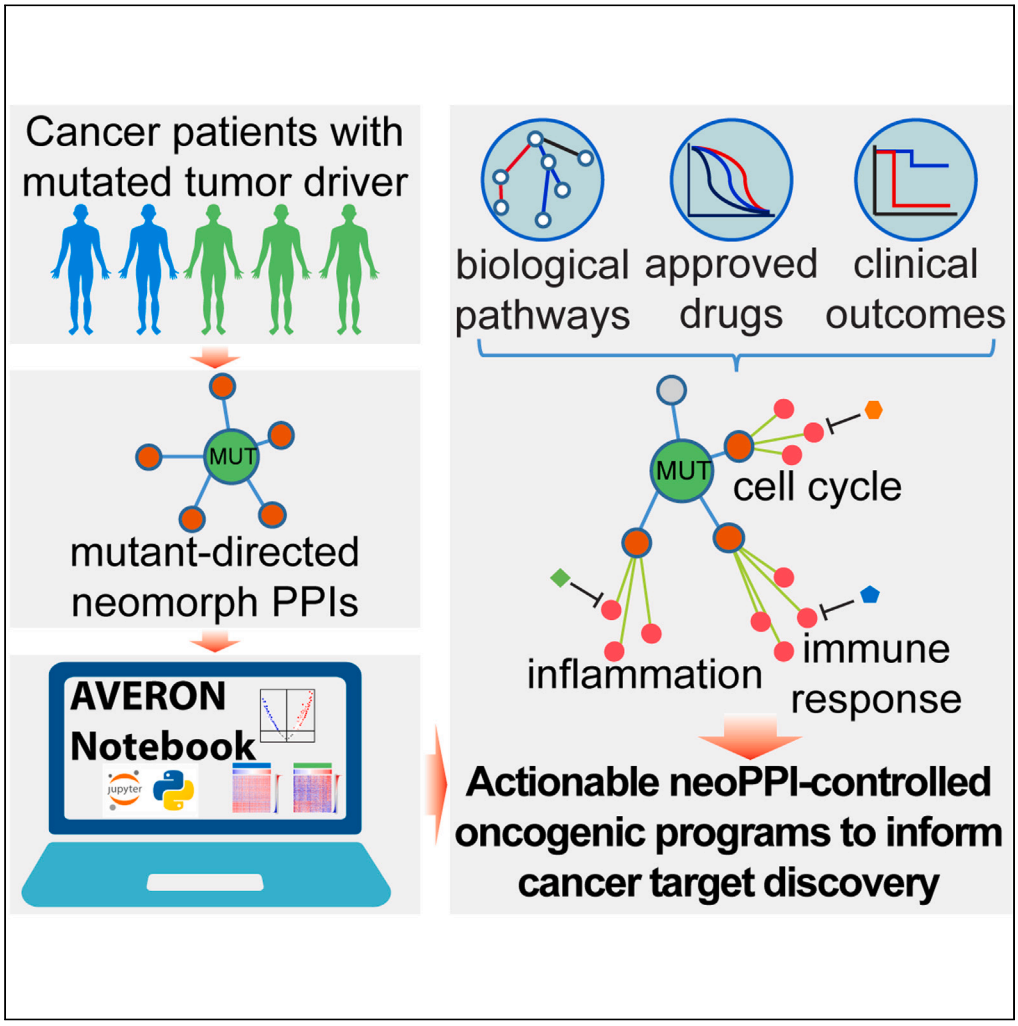
iScience

CellPress
OPEN ACCESS

Article

AVERON notebook to discover actionable cancer vulnerabilities enabled by neomorph protein-protein interactions



Hongyue Chen, Brian Revennaugh, Haian Fu, Andrey A. Ivanov

andrey.ivanov@emory.edu

Highlights

AVERON evaluates the biological functions and clinical significance of neomorph PPIs

Users can estimate and compare neoPPI levels across different cancer patient cohorts

AVERON uncovers neoPPI-regulated genes and maps them on defined pathways

AVERON identifies druggable clinically significant neoPPI-regulated genes

Chen et al., iScience 27, 110035
June 21, 2024
https://doi.org/10.1016/j.isci.2024.110035

iScience

CellPress
OPEN ACCESS

Article

AVERON notebook to discover actionable cancer vulnerabilities enabled by neomorph protein-protein interactions

Hongyue Chen,[1] Brian Revennaugh,[1] Haian Fu,[1,2,3,4] and Andrey A. Ivanov[1,2,3,5,*]

SUMMARY

Genomic alterations, such as missense mutations, often lead to the activation of oncogenic pathways and cell transformation by rewiring protein-protein interaction (PPI) networks. Understanding how mutant-directed neomorph PPIs (neoPPIs) drive cancer is vital to developing new personalized clinical strategies. However, the experimental interrogation of neoPPI functions in patients with cancer is highly challenging. To address this challenge, we developed a computational platform, termed AVERON for discovering actionable vulnerabilities enabled by rewired oncogenic networks. AVERON enables rapid systematic profiling of the clinical significance of neomorph PPIs across different cancer types, informing molecular mechanisms of neoPPI-driven tumorigenesis, and revealing therapeutically actionable neoPPI-regulated genes. We demonstrated the application of the AVERON platform by evaluating the biological functions and clinical significance of 130 neomorph interactions, experimentally determined for oncogenic BRAF^{V600E}. The AVERON application to broad sets of mutant-directed PPIs may inform new testable biological models and clinical strategies in cancer.

INTRODUCTION

Genomics alterations, such as missense mutations, play a crucial role in cancer initiation and progression.[1,2] Missense mutations can alter protein structure, activity, cellular localization, and other properties, affecting how proteins interact with each other. While some mutations can weaken existing interactions, others can induce neomorph protein-protein interactions (neoPPIs) that are not natural for the wild-type counterparts. Such mutation-induced changes in PPI networks produce system-wide effects on key oncogenic pathways.[3,4] Identifying variant-enabled neoPPIs and their oncogenic functions is vital to reveal much-needed tumor-specific molecular targets for therapeutic intervention.[5] Moreover, understanding how cancer driver mutations change oncogenic signaling networks through altered PPIs may reveal new strategies to target currently undruggable genes.[3,6,7]

The high-throughput PPI screening studies have revealed an emerging landscape of both loss-of-function and gain-of-function neoPPIs for both oncogenic and tumor suppressor mutations.[5,8] However, comprehensive experimental profiling of functional effects induced by mutant-directed neomorph PPIs and translation of the neoPPI networks into therapeutically actionable biological models is highly challenging. Many informatics tools, databases, and algorithms have been developed to inform the discovery of new targets and biomarkers in cancer.[9–11] For example, the OncoKB,[12] TumorPortal,[2] Mutation Assessor,[13] CIViC,[14] OncoMX,[15] BioMuta,[16] and other tools provide comprehensive annotations and prioritization of tumor driver mutations based on their oncogenic potential and clinical significance. Multiple approaches were proposed to either predict PPIs or estimate changes in protein binding affinities due to mutations.[17–19] To explore cancer-focused PPI networks integrated with cancer genomics, and clinical, pharmacological, and structural data, we have established the OncoPPi Portal.[20,21] Together, this informatics toolbox is highly instrumental in evaluating global biological and clinical responses to tumor driver mutations. However, no informatics tools were specially designed to identify druggable cancer dependencies acquired due to distinct neomorph protein-protein interactions at single amino acid residue resolution.

Here, we report the development of a novel computational platform, termed AVERON, for discovering **a**ctionable **v**ulnerabilities **e**nabled by **r**ewired **o**ncogenic **n**etworks. The AVERON provides the first computational environment to rapidly identify therapeutically actionable targets and pathways enabled by mutant-directed protein-protein interactions to inform target discovery in cancer.

[1]Department of Pharmacology and Chemical Biology, Emory University School of Medicine, Emory University, Atlanta, GA, USA
[2]Emory Chemical Biology Discovery Center, Emory University School of Medicine, Emory University, Atlanta, GA, USA
[3]Winship Cancer Institute, Emory University, Atlanta, GA, USA
[4]Department of Hematology, Medical Oncology Emory University, Atlanta, GA, USA
[5]Lead contact.
*Correspondence: andrey.ivanov@emory.edu
https://doi.org/10.1016/j.isci.2024.110035

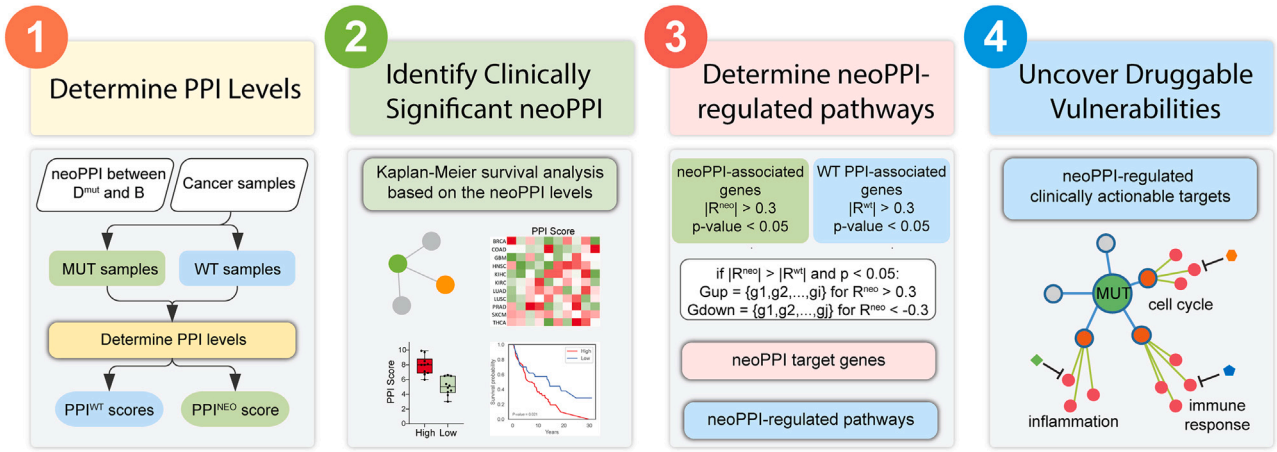


Figure 1. The AVERON Notebook design and utility

The AVERON Notebook provides a computational platform to uncover new actionable tumor dependencies on mutant-directed protein-protein interactions.

(1) Based on the mRNA expression or proteomics data determined for individual binding partners in cancer samples, the AVERON calculates the levels of mutant and wild-type protein-protein complexes in terms of PPI^{neo} and PPI^{wt} scores. The PPI scores enable a differential neoPPI level analysis across cancer types and patient cohorts.

(2) The calculated neoPPI scores also provide a new metric to determine the correlations between neoPPI levels, rather than individual gene levels, and clinical outcomes.

(3) A comparative analysis of correlations between PPI^{neo} and PPI^{wt} scores and gene expression enables the determination of neoPPI positively and negatively regulated gene sets and defined oncogenic pathways.

(4) The survival analysis of neoPPI-regulated genes helps uncover neoPPI-enabled tumor vulnerability and clinically significant neoPPI-dependent genes with available approved drugs and clinical compounds. Together, the AVERON provides a computational approach to determine therapeutically actionable oncogenic programs regulated by mutant-enabled protein-protein interactions to inform new clinical strategies and biological models in cancer.

RESULTS

The AVERON notebook design and utility

Available genomics and proteomics data define the landscape of individual genes and proteins across different cancer types. However, the large-scale experimental determination of the levels and functions of protein-protein complexes in patients with cancer is highly challenging, and currently, such data are not available. To overcome this challenge and inform the discovery of **a**ctionable **v**ulnerabilities **e**nabled by **r**e-wired **o**ncogenic **n**etworks, we developed a set of informatics methods termed AVERON. Implemented as a widely used Python Jupyter Notebook, the AVERON provides tools specially designed to i) rapidly estimate the levels of experimentally determined PPIs in cancer samples, ii) determine clinical significance of individual neoPPIs and neoPPI subsets, iii) uncover neoPPI-dependent oncogenic pathways and cellular programs, and iv) identify clinically significant neoPPI-regulated genes with available clinical compounds and approved drugs to inform new therapeutic strategies (Figure 1). The AVERON is freely available on GitHub repository: https://github.com/aivanovlab/averon_notebook and through https://chemicalbiology.emory.edu/averon.html.

The AVERON enables the comparative profiling of neoPPI amounts across different cancer samples. Multiple studies have demonstrated the relationship between PPIs and mRNA expression of the interactors.[22,23] Therefore, we can expect that the amount of a protein-protein complex in a cancer sample can be approximated based on the amounts of individual binding partners. Accordingly, the AVERON estimates neoPPI levels by calculating PPI scores defined as the arithmetic mean of the log(x+1) transformed mRNA expression or protein concentration of the binding partners. A detailed procedure for PPI Score calculations is described in the STAR methods section. The PPI scores provide a new metric that reflects the neoPPI level, enabling subsequent differential analysis of neoPPI levels across cancer types and patient cohorts and neoPPI-based survival analysis to evaluate the clinical significance of protein-protein complexes rather than individual proteins.

To gain insights into molecular mechanisms underlying the neoPPI-mediated tumorigenesis, the AVERON determines the neoPPI-regulated genes. By establishing correlations between the neo*PPI* scores and gene expression in samples harboring the corresponding neoPPI-enabling mutation, the AVERON identifies gene sets either positively or negatively associated with the neoPPI level. The implemented pathway enrichment analysis maps the neoPPI-associated genes with defined signaling and metabolic pathways and tumor hallmarks, which can be visualized within the AVERON environment or using external programs, such as Cytoscape.[24]

Then, through embedded survival analysis, the AVERON enables unveiling new cancer vulnerabilities enabled by mutant-directed PPIs and determining specific neoPPI-dependent genes whose expression contributes to worsened clinical outcomes of patients with cancer. It further connects the identified clinically relevant genes with available approved drugs and compounds under clinical testing to inform new therapeutic approaches and drug repurposing (Figure 1).

Together, AVERON provides a computational platform to uncover biological functions and molecular mechanisms of neoPPI-driven tumorigenesis and reveal therapeutically actionable cancer vulnerabilities created by mutant-directed protein-protein interactions in cancer.

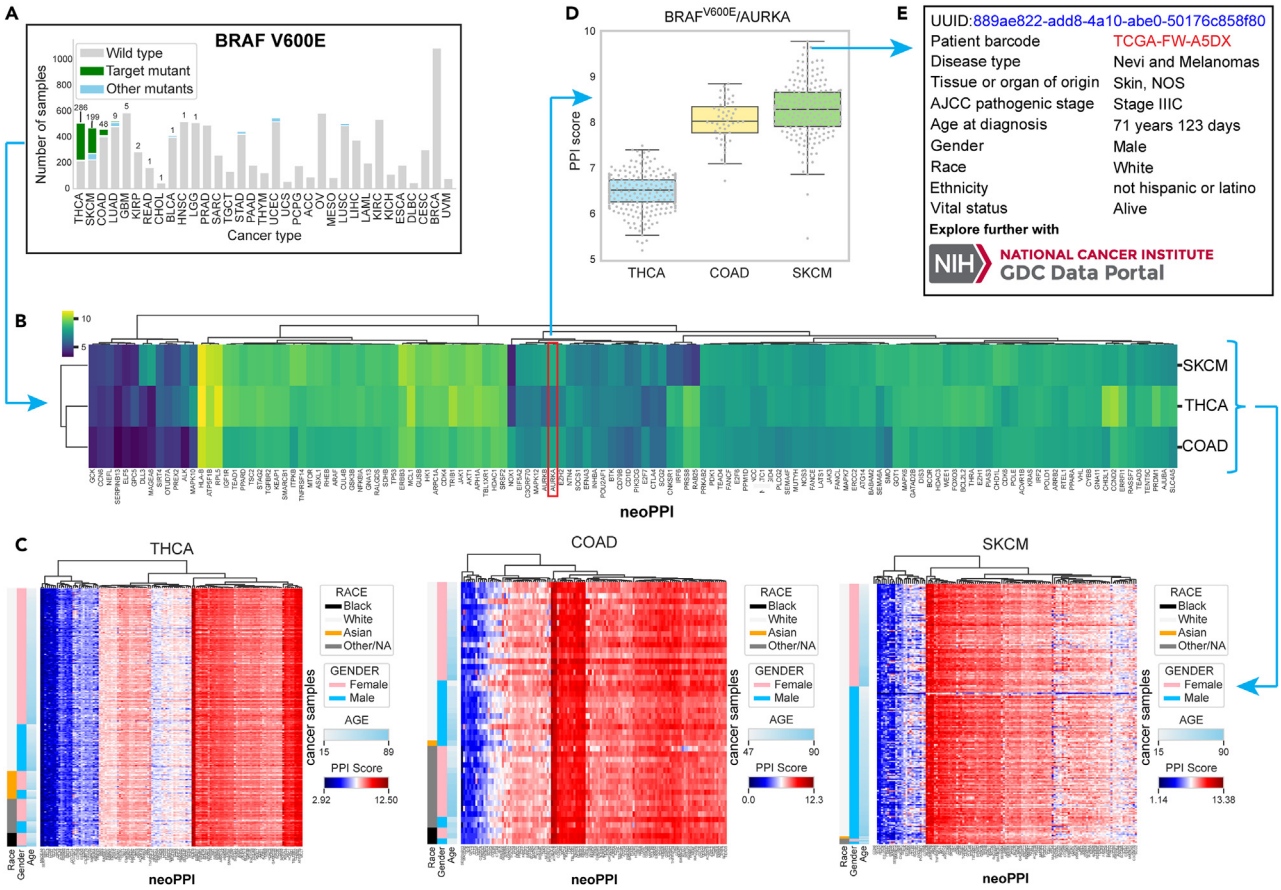


Figure 2. Evaluation and comparison of neoPPI levels across cancer samples

(A) The frequency of BRAF V600E mutation across different cancer types. See also Table S2.

(B) a heatmap shows the distribution of the averaged PPI scores calculated for multiple BRAF^{V600E} neoPPIs in skin melanoma (SKCM), colon (COAD), and thyroid (THCA) cancers.

(C) Detailed clusterograms help identify the clusters of cancer samples and neoPPIs with similar PPI levels. See also Figure S1, Tables S1, and S3.

(D) The boxplot shows the distribution of BRAF^{V600E}/AURKA neoPPI scores calculated for individual samples in different cancer types. Error bars indicate 95% confidence intervals. See also Figure S2.

(E) The AVERON provides on-the-fly annotation of patient samples with demographic data and diagnosis characteristics and hyperlinks the samples with detailed information available through the NCI GDC Data Portal for further exploration.

To exemplify the AVERON functionality and demonstrate its utility, we analyzed a set of 130 neoPPIs experimentally determined for BRAF V600E mutation[8] (Table S1), as described later in discussion.

The inference of neomorph protein-protein interaction levels from gene expression data

BRAF^{V600E} is a frequent oncogenic mutation in the thyroid (THCA), skin melanoma (SKCM), and colon (COAD) cancers (Figure 2A; Table S2). Recently, we found that this mutation leads to numerous neoPPIs, which either do not exist or are significantly weaker for BRAF wild type.[8] Here, we applied AVERON to determine which of those neoPPIs are most clinically important, identify mechanisms of BRAF^{V600E} neoPPI-dependent tumorigenesis, and inform new therapeutical approaches to control BRAF-driven cancers.

First, to evaluate the levels of BRAF^{V600E} neoPPIs in cancer samples, we calculated the PPI^{neo} scores. In this study, the PPI scores were calculated using the batch-normalized RNASeqV2 mRNA data along with mutation, and clinical data available through the NCI GDC Portal for TCGA PanCancer Atlas cohorts[25] as described in STAR methods.

The comparison of the calculated scores with the mRNA expression of the neomorph binding partners indicated that the PPI^{neo} score is a distinct parameter, which is not redundant with the binding partner expression. We found that for more than 30% of tested genes, the correlation between gene mRNA expression and PPI^{neo} scores was below 0.7, and for 20% the correlation was below 0.6 (Figure S1A). Furthermore, the overall correlation between PPI^{neo} scores calculated for different BRAF^{V600E} binding partners was also limited (on average, R = 0.45, Figure S1B), supporting the sensitivity of the PPI scores to the levels of different neoPPIs. Meanwhile, the correlation analysis revealed distinct communities of neoPPIs, co-regulated in patients with cancer. For example, we identified at least four distinct clusters of BRAF^{V600E} neoPPIs

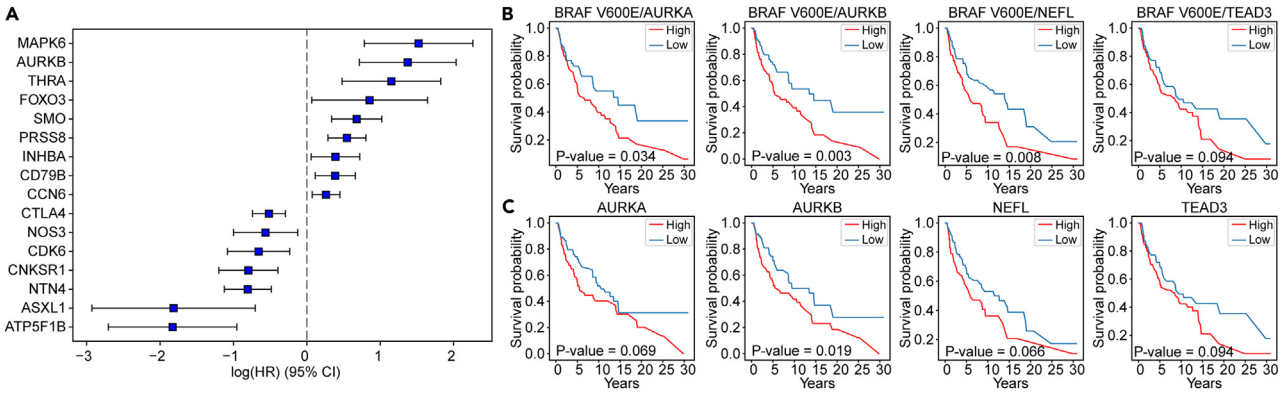


Figure 3. Identification of clinically significant neoPPIs in cancer

(A) The multivariate COX regression revealed a statistically significant (likelihood ratio test *p*-value <0.005) association between multiple BRAF^{V600E} neoPPIs and clinical outcomes of patients with SKCM. The Y axis indicates statistically significant BRAF^{V600E} neoPPIs served as the covariates in the model. The X axis shows the hazard ratio (HR) coefficients along with the 95% confidence intervals (CI). See also Table S4.

(B) Examples of BRAF^{V600E} neoPPIs whose high level correlates (*p*-value <0.1) with decreased SKCM patient survival in the univariate Kaplan-Meier analysis. The log rank *p*-values are shown. See also Figure S3 and Table S5.

(C) For multiple genes, the mRNA expression demonstrated a decreased statistical significance in the Kaplan-Meier analysis as compared to the neoPPI scores.

co-upregulated in patients with colon cancer (Figure S1C). *Cluster 1* was enriched (FDR <1%) in BRAF^{V600E} binding partners involved in JAK/STAT and interleukin signaling pathways. *Cluster 2* was enriched in genes regulated by E2F and involved in the G2-M checkpoint as well as the p53 pathway feedback loops 2 as defined in the Panther database. *Cluster 3* showed significant enrichment (FDR <1%) in the Ras pathway, interleukin signaling pathway, p53 pathway by glucose deprivation, and insulin/IGF pathway-protein kinase B signaling cascade. *Cluster 4* demonstrated the enrichment in IL-2/STAT5 signaling and B-cell activation. These results suggest that a single BRAF^{V600E} mutation can induce distinct groups of neomorph interactions that can work together to regulate multiple oncogenic pathways and transcriptional programs.

The calculated PPI^{neo} scores allowed us to evaluate the relative levels of individual neoPPIs across thyroid, skin, and colon cancers. The neoPPI levels appeared significantly different for different cancer types (Figures 2B and 2C). Among all tested neoPPIs, only the PPI^{neo} scores calculated for BRAF^{V600E} neoPPIs with GOT1 (Figure S2), POU2AF1, KRAS, CYBB, ACVR1B, and NEFL demonstrated a limited variation across the three cancer types with the Kruskal-Wallis H-test *p*-value >0.01 and the false discovery rate (FDR) > 1%. In contrast, most neoPPIs demonstrated significant variations in the PPI^{neo} scores (FDR $\ll$ 0.01%) in different cancer types (Table S3). For example, BRAF^{V600E} PPI^{neo} scores obtained for SNKSR1 and AJUBA were the highest in thyroid cancer, and the lowest in skin melanoma (Figure S2). BRAF^{V600E}/NOX1 PPI showed the highest level in colon cancer. The levels of BRAF^{V600E} neoPPIs with AURKA (Figure 2D), DDL3, and CHD1L (Figure S2), and multiple other genes showed the highest levels in patients with skin melanoma.

To enable detailed exploration of sample-associated metadata, we implemented on-the-fly annotations of individual tumor samples with demographic and diagnosis characteristics and hyperlinked the samples with sample annotations available through the NCI GDC Data Portal (Figure 2E).

Impact of neomorph protein-protein interactions on cancer patient clinical outcomes

The estimated levels of neoPPIs in cancer samples enable the evaluation of neoPPI clinical significance as potential biomarkers or targets for therapeutic discovery. To achieve this goal, we used the neoPPI scores as covariates for multivariate COX regression as well as univariate Kaplan-Meier analysis. For example, the application of the COX regression analysis to neoPPI scores calculated for BRAF^{V600E} neoPPI in patients with skin cutaneous melanoma (SKCM) resulted in a model containing 16 neoPPIs as statistically significant (*p*-value <0.05) covariates (Figure 3A; Table S4) and characterized by the likelihood ratio test *p*-value of 7.09×10^{-13} and the concordance of 0.79. Among the BRAF^{V600E} neoPPIs included in the model, ASXL1, ATP5F1B, CDK6, CNKSR1, CTLA4, NOS3, and NTN4 neoPPIs showed the hazard ratios (HR) < 1, suggesting that high levels of those neoPPIs may have tumor suppressive functions. In contrast, the elevation of AURKB, CCN6, CD79B, FOXO3, INHBA, MAPK6, PRSS8, SMO, and THRA neoPPIs was associated with increased risk (HR > 1).

In addition to the multivariate COX regression, the AVERON can determine the association of individual neoPPIs with clinical outcomes through the univariate Kaplan-Meier survival analysis of individual neoPPIs as described in STAR methods. The univariate analysis of BRAF^{V600E} neoPPIs confirmed a strong association of decreased patient survival with high levels of AURKAB and MAPK6 which showed the highest HR values of 3.96 and 4.5, respectively, in the COX regression model (Figure 3B). In addition, the univariate analysis revealed that high levels of BRAF^{V600E} neoPPIs with AURKA, CDK4, EFNA3, ERBB3, MAPK7, NEFL, RAB25, SERPINB13, and TEAD3 can contribute to worsened clinical outcomes of patients with skin melanoma with the log rank *p*-values <0.1 (Figure 3A). Among them, AURKA, AURKB, ERBB3, MAPK6, and NEFL demonstrated *p*-values $\leq$0.05 and FDR 25% (Table S5; Figure S3A). In colon cancer, BRAF^{V600E} neoPPIs with ASXL1, BCL2L2, BCOR, CCN6, DLL3, GNA11, JAK3, NFATC1, NOS3, NTN4, PLCG2, SOCS1, TNFRSF14, and TSC2 correlated with decreased

patient survival (p-value <0.1). Among them, GNA11 (p-value = 0.003, FDR 17%) and NTN4 (p-value 0.001, FDR 13%) demonstrated the highest statistical significance (Figure S3B; Table S5).

Importantly, we observed notable differences in the association of clinical outcomes with neoPPI levels and mRNA expression of the neo-binding partners. For example, the $BRAF^{V600E}$/AURKA neoPPI score showed a log rank p-value of 0.034, while the association of AURKA expression with decreased patient survival was characterized a by p-value of 0.069. The p-value of 0.066 obtained for NEFL expression contrasted with the NEFL neoPPI p-value of 0.008. In addition, the p-value of 0.094 obtained for TEAD3 neoPPI increased to p-value = 0.243, shown by TEAD3 expression, and the AURKB neoPPI p-value = 0.003 also increased to 0.019 for AURKB expression, which can be rejected with a stringent statistical cutoff of 0.01. Thus, the estimated PPI levels may uncover new clinically significant cancer dependencies hidden for individual gene expression analysis.

Neomorph protein-protein interaction-dependent oncogenic programs

To get mechanistic insights into how neoPPIs promote tumorigenesis, the AVERON identifies genes that can be regulated through individual neoPPIs. We hypothesize that the expression of neoPPI-regulated genes correlates with the neoPPI level. Accordingly, the AVERON conducts a systematic Pearson correlation analysis to reveal genes that show a statistically stronger positive or negative correlation with the neoPPI scores than with the wild-type PPI scores (see STAR methods for details).

The analysis of $BRAF^{V600E}$ neoPPI-regulated genes (Table S6) demonstrated that neoPPI-regulated gene sets can vary in both size and composition. For example, most (85) of $BRAF^{V600E}$ neoPPIs were strongly associated with the upregulation of 10–100 genes, 41 neoPPIs were predicted to upregulate more than 100 genes, and for 4 neoPPIs less than 10 genes showed a statistically significant positive correlation (Figure S4A). The overlap between the signature genes identified for different neoPPIs was also limited as indicated by the averaged Jaccard index of 3.1% (Figure S4B). These data suggest that different neoPPIs induced by the same mutation, such as $BRAF^{V600E}$, can activate distinct sets of genes leading to the differential regulation of diverse oncogenic programs.

To uncover such neoPPI-dependent oncogenic pathways, we implemented pathway enrichment analysis (PEA) using the MSigDB Cancer Hallmarks gene sets[26,27] along with signaling and metabolic pathways defined in the KEGG[28] and the Reactome[29] databases (Figure 4A; Table S7; Figure S5).

The PEA application to $BRAF^{V600E}$ neoPPIs revealed their role in the regulation of multiple major oncogenic pathways. For example, genes upregulated through $BRAF^{V600E}$ neoPPIs with PLCG2, CD79B, BTK, CNKSR1, ARRB2, IRF2, CD1D, and RALGD1 were associated with the interferon-gamma response, inflammatory response, and allograft rejection (Figure S3C; Table S7). The SRSF2, MUTYH, HDAC3, FANCF, GATAD2B, and ASXL1 neoPPI-dependent genes showed enrichment in the Apoptosis hallmark gene set defined in the MSigDB (Figure 4B). The $BRAF^{V600E}$/KEAP1 neoPPI was associated with the upregulation of genes involved in WNT/β-Catenin signaling (Figure 4B), supporting the previously reported cross-talk between KEAP1/NRF2 and WNT/β-Catenin pathways.[30,31] Genes controlled by $BRAF^{V600E}$ neoPPIs with AKT1, POLE, E2F6, AURKA, E2F7, SMARCB1, ATP5F1B, ERCC2, TEAD4, BCOR, SEMA6A, SDHB, CDK4, EZH2, and SMO showed the enrichment in G2M Checkpoint (Figure 4B) and E2F Targets hallmark gene sets. In addition, multiple neoPPIs, including AKT1, POLE, SMARCB1, ATP5F1B, ERCC2, TEAD4, and AURKA demonstrated the enrichment in the target genes of the major tumor driver transcription factor MYC (Figure S4C; Table S7).

Uncovering clinically significant neomorph protein-protein interaction-enabled cancer vulnerabilities

To prioritize the most clinically significant neoPPI-regulated genes, AVERON identifies genes whose high expression correlates with worsened clinical outcomes in the Kaplan-Meier survival analysis. Such genes may represent new vulnerabilities enabled by mutant-directed PPIs and serve as new therapeutic targets in mutant-driven cancers. Using this approach, we found that a total of 314 genes upregulated by $BRAF^{V600E}$ neoPPIs strongly correlate with decreased clinical outcomes (p-value <0.05) in patients with SKCM with $BRAF^{V600E}$ mutation. We further prioritized a set of 110 "high confidence" genes whose high expression correlates with decreased patient survival with the FDR <1% (Table S5). Thus, the pharmacological inhibition of those genes may provide new opportunities for therapeutic intervention in BRAF-driven skin melanomas. While some genes may represent novel promising targets for therapeutic discoveries, we found that 25 $BRAF^{V600E}$ neoPPI-regulated genes whose high expression correlates with decreased clinical outcomes in patients with SKCM have already been validated as the druggable targets with available small molecule inhibitors, including approved drugs (Figure 5A). For example, ingenol mebutant was approved for protein kinase C (PKC) the Epsilon isoform of which, appeared to be regulated by $BRAF^{V600E}$/PRSS8 neoPPI. The $BRAF^{V600E}$/PRSS8 neoPPI-upregulated ribonucleotide reductase (RPM2) can be controlled by FDA-approved gemcitabine, fludarabine, clofarabine, and hydroxyurea. We found that $BRAF^{V600E}$ neoPPI with TP53 correlates with an increased expression of ERBB2, one of the major oncogenes and a well-validated therapeutic target, with multiple FDA-approved drugs[32] (Figure 5B). Previous studies have demonstrated that TP53 mutations inhibit TP53 function as a suppressor of ERBB2 expression.[33,34] However, in patients with SKCM, TP53 mutations and $BRAF^{V600E}$ mutations are mutually exclusive, and the majority of patients with SKCM with $BRAF^{V600E}$ have TP53 wild type. Our data suggest, that oncogenic BRAF V600E mutation may induce its interaction with TP53, and inhibit TP53 tumor suppressor activity resulting in ERBB2 upregulation. Thus, we can hypothesize that patients with SKCM with elevated $BRAF^{V600E}$/TP53 neoPPI may benefit from targeting ERBB2 with FDA-approved drugs, for example, trastuzumab.[35] Multiple drugs have also been approved for cyclin-dependent kinase 4 (CDK4). While CDK4 can directly interact with $BRAF^{V600E}$, the AVERON predicted that CDK4 expression can be controlled through other neoPPIs, including $BRAF^{V600E}$ neoPPIs with POLE, TEAD4, and AURKA (Figure 5B). Thus, the tumor dependency on those neoPPIs can be targeted by CDK4 inhibitors. The $BRAF^{V600E}$/AURKA neoPPI was among the neoPPs, that showed a significant negative impact on SKCM patient clinical outcomes (Figure 6A) and which we

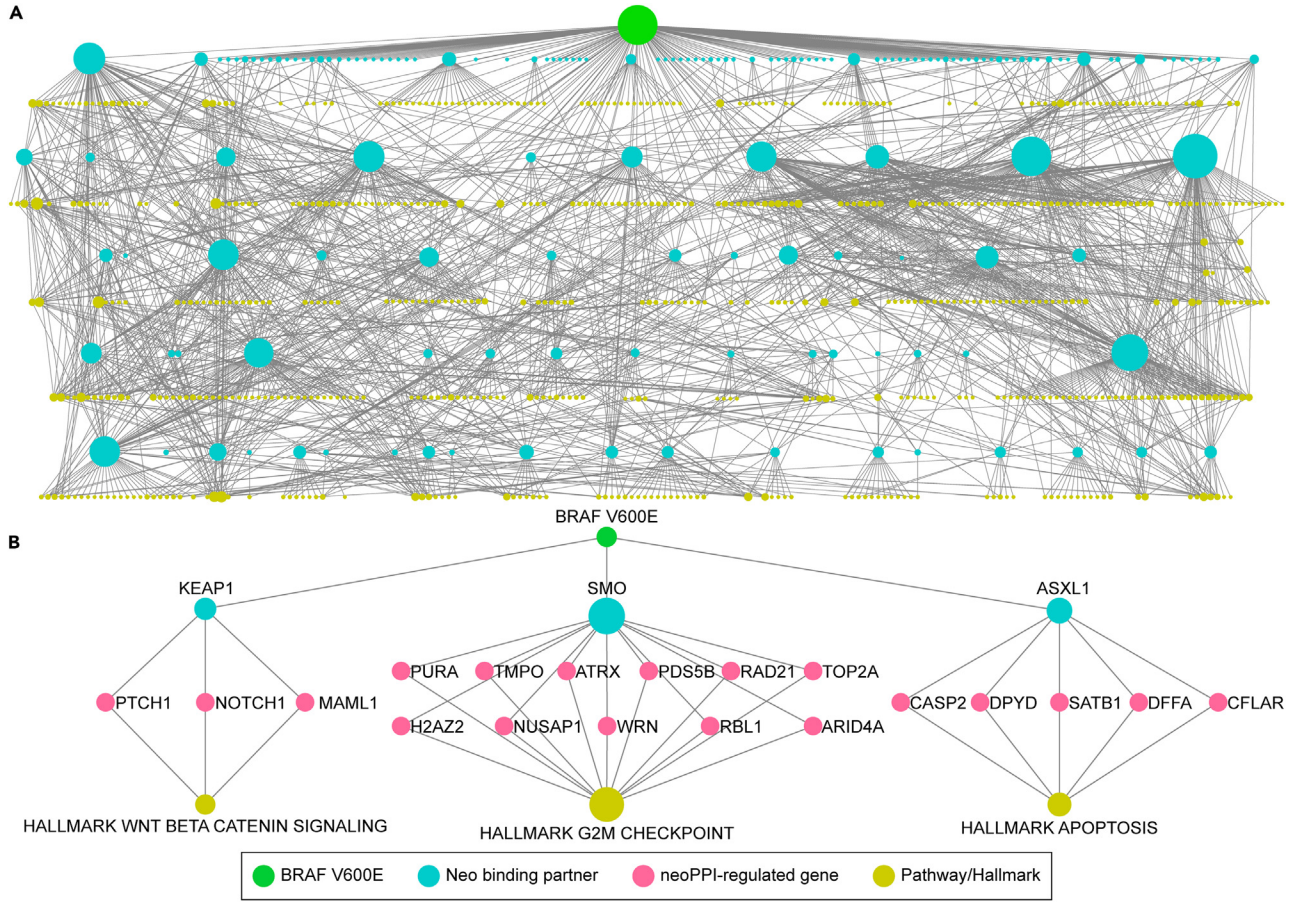


Figure 4. Discovery of neoPPI-dependent oncogenic programs

(A) The network of BRAF^{V600E} neoPPI-dependent pathways illustrates the diversity of signaling and metabolic pathways that can be controlled through different BRAF^{V600E} neoPPIs. BRAF^{V600E} is shown in green, the neo-binding partners are colored in cyan, and the neoPPI-regulated pathways defined in MSigDB, KEGG, and Reactome databases are shown in yellow. Pathways that show enrichment for BRAF^{V600E} neoPPI-regulated genes with FDR <5% are shown.

(B) Representative examples of BRAF^{V600E} neoPPI-regulated pathways. The neoPPI-regulated genes are shown in pink. See also Figures S4, S5, Tables S6, and S7.

have experimentally confirmed in multiple assays.[8] The AVERON showed that besides CDK4 regulation, the BRAF^{V600E}/AURKA neoPPI level correlates with the expression of more than 60 genes (Figure 6A) that are enriched in MYC- and E2F-target genes and cell cycle-regulating genes (Figures 6B and 6C). For instance, PLK1 expression correlated much stronger with BRAF^{V600E}/AURKA PPI^{neo} scores calculated for samples with BRAF^{V600E} mutation (R = 0.669, Figure 5D) than with PPI^{wt} scores obtained for BRAF wild-type samples (R = 0.395, Figure 6E).

The analysis of the patient survival data revealed that high PLK1 expression contributes to decreased survival of patients with SKCM with BRAF^{V600E} mutation with the log-rang *p*-value = 0.016 (Figure 6F). PLK1 is a well-studied kinase and an established target for therapeutic intervention.[36] For example, the AVERON search for PLK1 inhibitors provided multiple small molecule compounds, including onvansertib, the FDA Fast Track designated drug.[37] These findings led us to a new model by which BRAF^{V600E} mutation induces the neomorph PPI with AURKA leading to the upregulation of PLK1 and PLK1-dependent MYC activation,[38] which in turn contribute to cancer cell growth and poor clinical outcomes in skin cancer (Figure 6G). The model also suggests that patients with SKCM with BRAF^{V600E} mutation and high levels of AURKA can benefit from PLK1 suppression.

DISCUSSION

Dysregulation of the PPI networks by oncogenic genomic alterations, such as mutations, can lead to the perturbation of major oncogenic pathways and the acquisition of tumor hallmarks. The uncovering of the most biologically and clinically significant mutant-directed PPIs may significantly expand the understanding of molecular mechanisms of tumorigenesis and facilitate the discovery of new clinical strategies and predicting markers in cancer. However, the experimental determination of the protein-protein complexes and their functions in patients with cancer is highly challenging, and today, such information is unavailable for an arbitrary PPI. To overcome these critical challenges, we designed a computational approach, termed the AVERON Notebook, to systematically evaluate biological functions and clinical significance of neomorph protein-protein interactions enabled by single tumor-driving mutations.

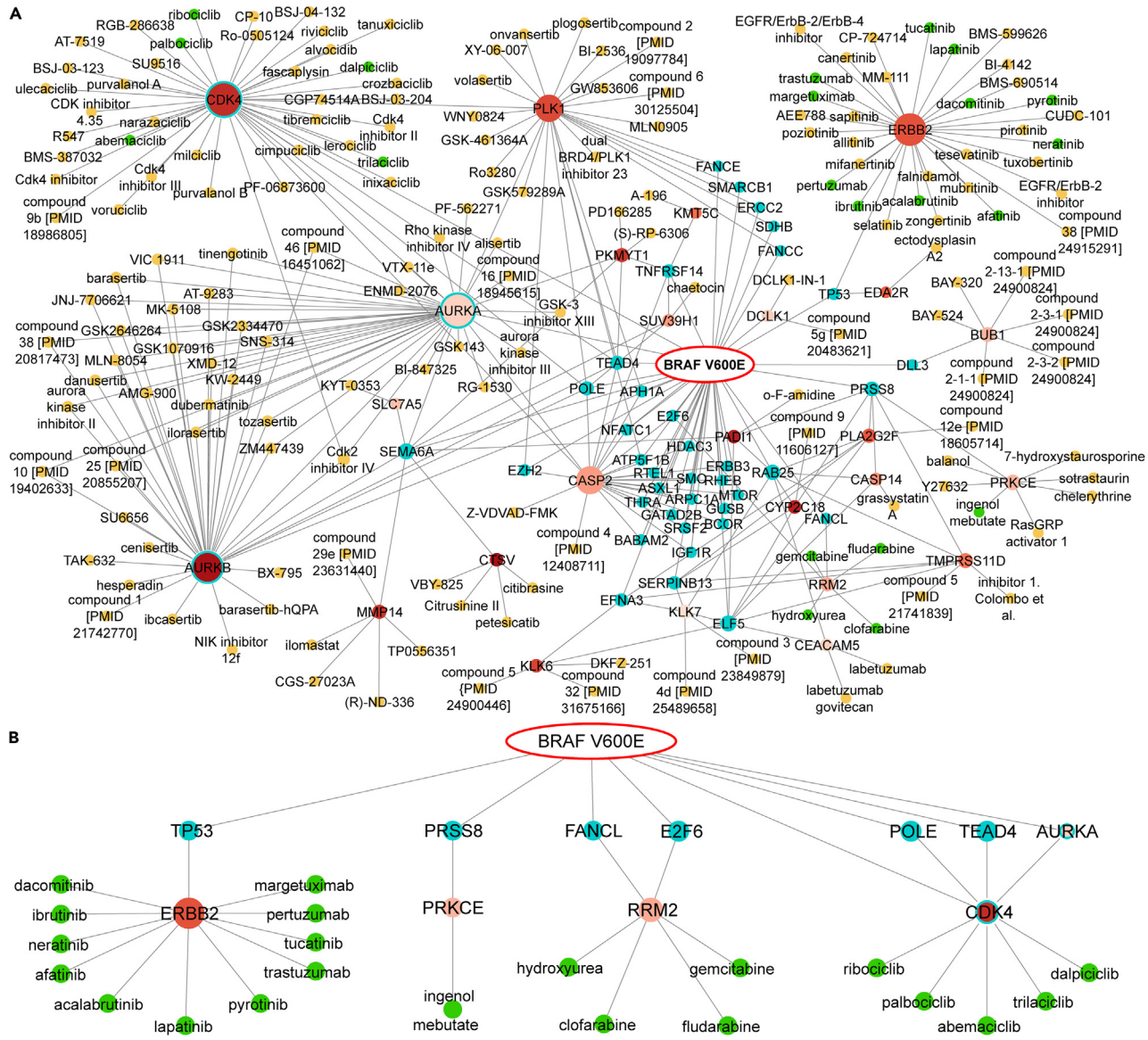


Figure 5. Discovery of druggable and clinically significant neoPPI-regulated genes

(A) The network shows clinically significant genes (red) upregulated by BRAF^{V600E} neoPPIs (cyan) in patients with skin melanoma. The gradient of red color indicates the log rank *p*-values, ranging from 0.05 (light red) to 0.0005 (strong red). CDK4, AURKA, and AURKB represent both neo-binding partners and neoPPI-regulated genes, as highlighted with a cyan border. The yellow circles indicate available small molecule inhibitors and the green circles show the approved drugs.

(B) A subnetwork of BRAF^{V600E} neoPPI-regulated genes with FDA-approved drugs.

For most PPIs, their concentration in cancer samples is unknown, and precise calculation of the complex concentration would require information about the dynamics and kinetics of the interaction as well as the protein stability, post-translation modification state, and the states of other binding partners. Today, such information is mainly unavailable, especially in patients with cancer. However, for experimentally determined complexes, we can assume that the higher the concentrations of each binding partner, the higher the concentration of their complex. Thus, the relative amounts of protein-protein complexes can be estimated as the averaged levels of two interactors, as implemented in AVERON with the PPI scores. Here, we showed that the PPI scores can provide a new metric distinct from the amounts of individual binding partners. Accordingly, we used the PPI scores in the AVERON informatics pipeline to enable the rapid evaluation of relative neoPPI levels across multiple cancer types and the impact of neoPPIs on cancer patient clinical outcomes. As one example, a systematic profiling of 130 experimentally determined BRAF^{V600E} neoPPIs revealed that distinct neoPPIs induced by V600E mutation may differently impact clinical outcomes in different cancer types. For instance, the effects of high levels of BRAF^{V600E}/AURKA neoPPI on skin melanoma patient survival can be more prominent than in patients with thyroid or colon cancer. In contrast, other BRAF^{V600E} neoPPIs, such as NTN4 or GNA11, can contribute to

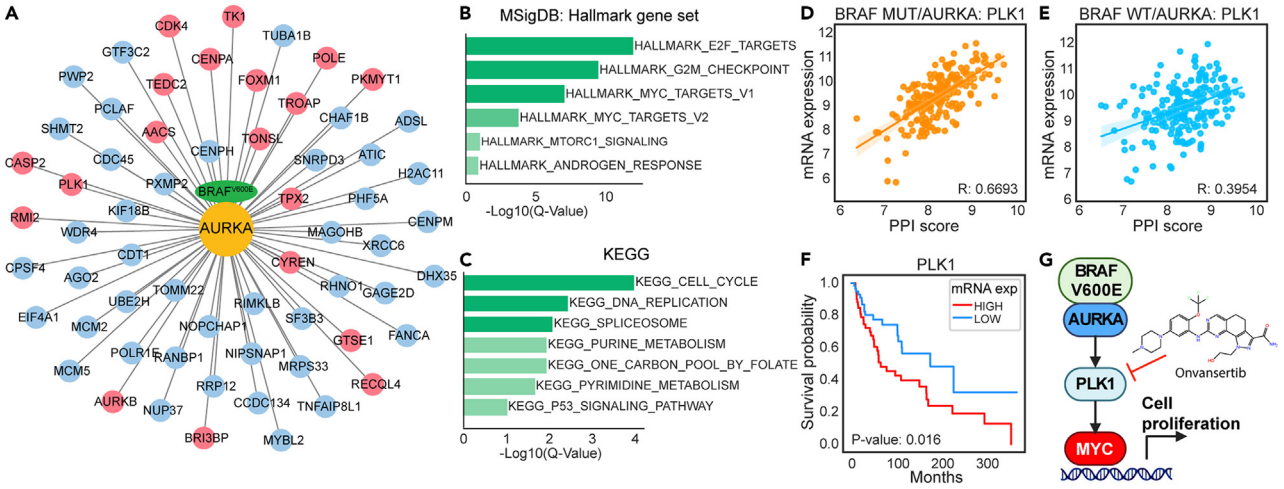


Figure 6. AVERON informs new biological models to control tumor growth and survival

(A) The network shows $BRAF^{V600E}$/AURKA neoPPI-regulated genes. Genes whose expression correlates with worsened clinical outcomes in patients with SKCM harboring BRAF V600E mutation are highlighted in red.

(B) $BRAF^{V600E}$/AURKA neoPPI-regulated genes showed the enrichment in MSigDB cancer hallmark gene sets and (C) pathways defined in the KEGG database.

(D) PLK1 expression correlates stronger with the $BRAF^{V600E}$/AURKA neoPPI levels in $BRAF^{V600E}$ mutant samples than with (E) $BRAF^{WT}$/AURKA PPI levels in BRAF wild-type samples.

(F) High PLK1 expression correlates with decreased survival of patients with SKCM with $BRAF^{V600E}$ mutation.

(G) A working model suggests that $BRAF^{V600E}$/AURKA neoPPI promotes MYC-dependent skin cancer growth by enhancing PLK1 expression and increasing cancer dependency on PLK1, which can be targeted by PLK1 inhibitors.

decreased survival of patients with colon cancer. Such associations between neoPPI levels and clinical outcomes can provide valuable guidance to select appropriate models for detailed experimental studies, which otherwise is unavailable. Moreover, the multivariate COX regression analysis allowed us to identify a group of 16 neoPPIs, which may serve as predictors for SKCM clinical outcomes.

To help understand how neoPPIs contribute to tumorigenesis, AVERON identifies neoPPI-dependent gene sets and oncogenic programs. For instance, the AVERON analysis suggests that BRAFV600E mutation may produce a broad network effect dysregulating multiple oncogenic programs through distinct neomorph PPIs or neoPPI communities, including the regulation of cell cycle, apoptosis, immune response, and others. These data provide experimentally testable hypotheses to guide the development of new biological models.

Despite the advances in PPI targeting by small molecules for therapeutic development, a systematic discovery of clinically suitable PPI inhibitors remains challenging. A discovery of druggable and clinically significant neoPPI-regulated may provide a powerful alternative to target neoPPI-enabled cancer vulnerabilities. Toward this goal, AVERON enables the systematic profiling of neoPPI-regulated genes whose high expression correlates with decreased cancer patient survival. The genes are further connected with available clinical compounds and approved drugs. As one example, our data suggests new connectivity between $BRAF^{V600E}$ mutation and PLK1/MYC-dependent cell proliferation mediated through $BRAF^{V600E}$/AURKA neoPPI in patients with SKCM. PLK1 is a well-validated therapeutic target and a druggable kinase, and inhibiting PLK1 with available clinical compounds, such as onvansertib, can be beneficial for patients with skin melanoma with $BRAF^{V600E}$ mutation and high AURKA expression.

In summary, the AVERON Notebook provides the first tools to uncover cancer dependencies created by mutant-directed protein-protein interactions. Its application to $BRAF^{V600E}$ neoPPIs has revealed new mechanisms by which V600E mutation can drive tumorigenesis and suggested specific genes as promising targets to control BRAF-driven cancers. We believe that AVERON will provide a valuable resource to identify new molecular mechanisms of oncogenic signaling, determine new cancer dependencies on specific mutant-directed PPIs, and inform new personalized clinical strategies in mutant-driven cancers.

Limitations of the study

In some cases, a limited number of samples with a driver mutation can decrease the statistical power of the AVERON analysis. The integration of multiple datasets from different sources could overcome this common limitation. Since the experimental data on the concentrations of protein complexes in cancer samples is unavailable, the AVERON approximates the neoPPI levels based on the mRNA expression or protein concentrations available for the single proteins. Although this is one of the most appropriate strategies currently available, such an approximation cannot fully recapitulate the actual amounts and dynamics of PPIs in patients with cancer. A limited agreement between clinical data and the data derived from available cancer models can also challenge a direct experimental validation of AVERON predictions. Nevertheless, we believe that AVERON will provide a highly informative platform to identify i) molecular mechanisms of neoPPI-dependent tumorigenesis, ii) patient cohorts with the highest dependency on individual neoPPIs and their signature genes, and iii) drug repurposing and combinations to control mutant-driven cancers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - AVERON notebook implementation
  - Evaluation of neoPPI levels in cancer patients
  - Identification of neoPPI-regulated genes
  - Pathway enrichment analysis
  - Survival analysis
  - Integration of pharmacological data
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110035.

ACKNOWLEDGMENTS

This work has been supported in part by NCI's Informatics Technology for Cancer Research (ITCR) Program (R21CA274620, A.A.I.), Winship Cancer Institute #IRG-17-181-06 from the American Cancer Society (A.A.I.). NCI Emory Lung Cancer SPORE (P50CA217691, H.F.), NCI P01CA257906 (H.F.), Career Enhancement Program (A.A.I., P50CA217691), Winship Cancer Institute (NIH 5P30CA138292), Mary Kay Ash Foundation Cancer Research Award (A.A.I).

AUTHOR CONTRIBUTIONS

AAI: Conceptualization; AAI, HC: methodology, investigation, formal analysis, and visualization; AAI, HF, and HC: writing – review and editing; AAI and HC: writing – original draft; AAI, HC, and BR: software and data curation; AAI and HF: funding acquisition.

DECLARATION OF INTERESTS

There are no conflicts of interest to disclose.

Received: February 14, 2024
Revised: April 30, 2024
Accepted: May 16, 2024
Published: May 20, 2024

REFERENCES

1. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. Nature *578*, 82–93. https://doi.org/10.1038/s41586-020-1969-6.
2. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501. https://doi.org/10.1038/nature12912.
3. Ivanov, A.A., Khuri, F.R., and Fu, H. (2013). Targeting protein-protein interactions as an anticancer strategy. Trends Pharmacol. Sci. *34*, 393–400. https://doi.org/10.1016/j.tips.2013.04.007.
4. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558. https://doi.org/10.1126/science.1235122.
5. Cheng, F., Zhao, J., Wang, Y., Lu, W., Liu, Z., Zhou, Y., Martin, W.R., Wang, R., Huang, J., Hao, T., et al. (2021). Comprehensive characterization of protein-protein interactions perturbed by disease mutations. Nat. Genet. *53*, 342–353. https://doi.org/10.1038/s41588-020-00774-y.
6. Arkin, M.R., Tang, Y., and Wells, J.A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. Chem. Biol. *21*, 1102–1114. https://doi.org/10.1016/j.chembiol.2014.09.001.
7. Li, Z., Ivanov, A.A., 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. https://doi.org/10.1038/ncomms14356.
8. Mo, X., Niu, Q., Ivanov, A.A., Tsang, Y.H., Tang, C., Shu, C., Li, Q., Qian, K., Wahafu, A., Doyle, S.P., et al. (2022). Systematic discovery of mutation-directed neo-protein-protein interactions in cancer. Cell *185*, 1974–1985.e12. https://doi.org/10.1016/j.cell.2022.04.014.
9. Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. *50*, W345–W351. https://doi.org/10.1093/nar/gkac247.
10. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat. Genet. *38*, 500–501. https://doi.org/10.1038/ng0506-500.
11. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S.,

Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564–576.e16. https://doi.org/10.1016/j.cell.2017.06.010.

12. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis. Oncol. *1*, 1–16. https://doi.org/10.1200/PO.17.00011.
13. Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. *39*, e118. https://doi.org/10.1093/nar/gkr407.
14. Griffith, M., Spies, N.C., Krysiak, K., McMichael, J.F., Coffman, A.C., Danos, A.M., Ainscough, B.J., Ramirez, C.A., Rieke, D.T., Kujan, L., et al. (2017). CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat. Genet. *49*, 170–174. https://doi.org/10.1038/ng.3774.
15. Dingerdissen, H.M., Bastian, F., Vijay-Shanker, K., Robinson-Rechavi, M., Bell, A., Gogate, N., Gupta, S., Holmes, E., Kahsay, R., Keeney, J., et al. (2020). OncoMX: A Knowledgebase for Exploring Cancer Biomarkers in the Context of Related Cancer and Healthy Data. JCO Clin. Cancer Inform. *4*, 210–220. https://doi.org/10.1200/CCI.19.00117.
16. Dingerdissen, H.M., Torcivia-Rodriguez, J., Hu, Y., Chang, T.C., Mazumder, R., and Kahsay, R. (2018). BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. Nucleic Acids Res. *46*, D1128–D1136. https://doi.org/10.1093/nar/gkx907.
17. Rodrigues, C.H.M., Pires, D.E.V., and Ascher, D.B. (2021). mmCSM-PPI: predicting the effects of multiple point mutations on protein-protein interactions. Nucleic Acids Res. *49*, W417–W424. https://doi.org/10.1093/nar/gkab273.
18. Romero-Molina, S., Ruiz-Blanco, Y.B., Mieres-Perez, J., Harms, M., Münch, J., Ehrmann, M., and Sanchez-Garcia, E. (2022). PPI-Affinity: A Web Tool for the Prediction and Optimization of Protein-Peptide and Protein-Protein Binding Affinity. J. Proteome Res. *21*, 1829–1841. https://doi.org/10.1021/acs.jproteome.2c00020.
19. Wang, M., Cang, Z., and Wei, G.W. (2020). A topology-based network tree for the prediction of protein-protein binding affinity changes following mutation. Nat. Mach. Intell. *2*, 116–123. https://doi.org/10.1038/s42256-020-0149-6.
20. Ivanov, A.A., Revennaugh, B., Rusnak, L., Gonzalez-Pecchi, V., Mo, X., Johns, M.A., Du, Y., Cooper, L.A.D., Moreno, C.S., Khuri, F.R., and Fu, H. (2018). The OncoPPi Portal: an integrative resource to explore and prioritize protein-protein interactions for cancer target discovery. Bioinformatics *34*, 1183–1191. https://doi.org/10.1093/bioinformatics/btx743.
21. Ivanov, A.A. (2020). Explore Protein-Protein Interactions for Cancer Target Discovery Using the OncoPPi Portal. Methods Mol. Biol. *2074*, 145–164. https://doi.org/10.1007/978-1-4939-9873-9_12.
22. Su, L., Liu, G., Guo, Y., Zhang, X., Zhu, X., and Wang, J. (2022). Integration of Protein-Protein Interaction Networks and Gene Expression Profiles Helps Detect Pancreatic Adenocarcinoma Candidate Genes. Front. Genet. *13*, 854661. https://doi.org/10.3389/fgene.2022.854661.
23. Bhardwaj, N., and Lu, H. (2005). Correlation between gene expression profiles and protein-protein interactions within and across genomes. Bioinformatics *21*, 2730–2738. https://doi.org/10.1093/bioinformatics/bti398.
24. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.
25. Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., and Staudt, L.M. (2016). Toward a Shared Vision for Cancer Genomic Data. N. Engl. J. Med. *375*, 1109–1112. https://doi.org/10.1056/NEJMp1607591.
26. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.
27. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.
28. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361. https://doi.org/10.1093/nar/gkw1092.
29. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655. https://doi.org/10.1093/nar/gkx1132.
30. Nault, J.C., Rebouissou, S., and Zucman Rossi, J. (2015). NRF2/KEAP1 and Wnt/β-catenin in the multistep process of liver carcinogenesis in humans and rats. Hepatology *62*, 677–679. https://doi.org/10.1002/hep.27828.
31. Fragoulis, A., Schenkel, J., Schröder, N., Brandt, E.F., Weiand, M., Neu, T., Ramadori, P., Caspers, T., Kant, S., Pufe, T., et al. (2022). Nrf2 induces malignant transformation of hepatic progenitor cells by inducing beta-catenin expression. Redox Biol. *57*, 102453. https://doi.org/10.1016/j.redox.2022.102453.
32. Zhu, K., Yang, X., Tai, H., Zhong, X., Luo, T., and Zheng, H. (2024). HER2-targeted therapies in cancer: a systematic review. Biomark. Res. *12*, 16. https://doi.org/10.1186/s40364-024-00565-1.
33. Román-Rosales, A.A., Garcia-Villa, E., Herrera, L.A., Gariglio, P., and Diaz-Chavez, J. (2018). Mutant p53 gain of function induces HER2 over-expression in cancer cells. BMC Cancer *18*, 709. https://doi.org/10.1186/s12885-018-4613-1.
34. Narisawa-Saito, M., Handa, K., Yugawa, T., Ohno, S., Fujita, M., and Kiyono, T. (2007). HPV16 E6-mediated stabilization of ErbB2 in neoplastic transformation of human cervical keratinocytes. Oncogene *26*, 2988–2996. https://doi.org/10.1038/sj.onc.1210118.
35. Dawood, S., Broglio, K., Buzdar, A.U., Hortobagyi, G.N., and Giordano, S.H. (2010). Prognosis of women with metastatic breast cancer by HER2 status and trastuzumab treatment: an institutional-based review. J. Clin. Oncol. *28*, 92–98. https://doi.org/10.1200/JCO.2008.19.9844.
36. Liu, Z., Sun, Q., and Wang, X. (2017). PLK1, A Potential Target for Cancer Therapy. Transl. Oncol. *10*, 22–32. https://doi.org/10.1016/j.tranon.2016.10.003.
37. Wang, D., Veo, B., Pierce, A., Fosmire, S., Madhavan, K., Balakrishnan, I., Donson, A., Alimova, I., Sullivan, K.D., Joshi, M., et al. (2022). A novel PLK1 inhibitor onvansertib effectively sensitizes MYC-driven medulloblastoma to radiotherapy. Neuro Oncol. *24*, 414–426. https://doi.org/10.1093/neuonc/noab207.
38. Xiao, D., Yue, M., Su, H., Ren, P., Jiang, J., Li, F., Hu, Y., Du, H., Liu, H., and Qing, G. (2016). Polo-like Kinase-1 Regulates Myc Stabilization and Activates a Feedforward Circuit Promoting Tumor Cell Survival. Mol. Cell *64*, 493–506. https://doi.org/10.1016/j.molcel.2016.09.016.
39. Perkel, J.M. (2018). Why Jupyter is data scientists' computational notebook of choice. Nature *563*, 145–146. https://doi.org/10.1038/d41586-018-07196-1.
40. Mah, C.K., Wenzel, A.T., Juarez, E.F., Tabor, T., Reich, M.M., and Mesirov, J.P. (2018). An accessible, interactive GenePattern Notebook for analysis and exploration of single-cell transcriptomic data. F1000Res. *7*, 1306. https://doi.org/10.12688/f1000research.15830.2.
41. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Roy. Stat. Soc. B *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
42. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.
43. Harding, S.D., Sharman, J.L., Faccenda, E., Southan, C., Pawson, A.J., Ireland, S., Gray, A.J.G., Bruce, L., Alexander, S.P.H., Anderton, S., et al. (2018). The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. Nucleic Acids Res. *46*, D1091–D1106. https://doi.org/10.1093/nar/gkx1121.
44. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., and Frederic, J. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas, F. Loizides and B. Schmidt, eds., pp. 87–90.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Mutation-Directed Neo-Protein-Protein Interactions	Mo et al.[8]	https://doi.org/10.1016/j.cell.2022.04.014
MSigDB	Liberzon et al.[27]	https://www.gsea-msigdb.org/gsea/msigdb
KEGG	Kanehisa et al.[28]	https://www.genome.jp/kegg
REACTOME	Fabregat et al.[29]	https://reactome.org
IUPHAR database	Harding. et al.[43]	https://www.guidetopharmacology.org
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	ACC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	BLCA
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	BRCA
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	CESC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	CHOL
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	COAD
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	DLBC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	ESCA
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	GBM
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	HNSC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	KICH
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	KIRC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	KIRP
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	LAML
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	LGG
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	LIHC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	LUAD
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	LUSC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	MESO
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	OV
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	PAAD
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	PCPG
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	PRAD
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	READ
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	SARC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	SKCM
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	STAD
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	TGCT
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	THCA
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	THYM
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	UCEC
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	UCS
TCGA Pan-Cancer RNA-seq, Mutation, and clinical data	NCI GDC Data Portal	UVM
Ligand-target interactions	IUPHAR/BPS Guide to Pharmacology database	https://blog.guidetopharmacology.org/2024/03/27/database-release-2024-1/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
AVERON Notebook	This paper	https://github.com/aivanovlab/averon_notebook
Python 3	Python Software Foundation	https://www.python.org
Jupyter Notebook environment	Ragan-Kelley et al.[44]	www.jupyter.org
MSigDB	Liberzon et al.[27]	https://www.gsea-msigdb.org/gsea/msigdb
Cytoscape	Shannon et al.[24]	https://cytoscape.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Andrey A. Ivanov (andrey.ivanov@emory.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

All original code has been deposited on GitHub and is publicly available as of the date of publication. The link is provided in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The AVERON Notebook uses cancer genomics and clinical data from TCGA Pan Cancer Atlas available for over 10,000 tumors and 33 cancer types through the NCI Genomic Data Commons Data Portal.[25]

METHOD DETAILS

AVERON notebook implementation

The AVERON Notebook is available on GitHub https://github.com/aivanovlab/averon_notebook and at https://chemicalbiology.emory.edu/averon.html. To streamline the availability of the AVERON to the scientific community, we implement it as an open-source freely available Python Jupyter Notebook, a format widely used in different areas of cancer data analysis.[39,40] The AVERON uses the following Python libraries: *glob, ipycytoscape, IPython, ipywidgets, json, lifelines, math, matplotlib, numpy, os, pandas, pygtop, requests, scikit_posthocs, scipy, seaborn, sklearn, statsmodels, sys, tabulate, tkinter, warnings.*

Evaluation of neoPPI levels in cancer patients

We define a neoPPI as a physical interaction between a mutated tumor driver protein (D^{mut}), and a binding partner (B), which demonstrates a significantly weaker or no interaction with the wild-type counterpart (D^{wt}). D^{mut} can correspond to a single point mutation or a set of mutations in the tumor driver D. The AVERON is based on neoPPIs, experimentally determined in cancer cells. In this study, we analyzed 130 BRAF V600E neoPPIs experimentally determined in HEK293T cells using the NanoLuc-based bioluminescence resonance energy transfer (BRET^n) platform. To estimate the levels of neoPPIs in patients the AVERON uses the mRNA expression data. The NCI Genomic Data Commons (GDC) Data Portal provides genomics data for 33 cancer types and more than 23,500 genes.[25] We use the batch-normalized RNASeqV2 mRNA data along with mutation, DNA copy number, and clinical data available through the GDC for TCGA PanCancer Atlas cohorts. Genes with zero expression in more than 30% of samples are excluded from the analysis. The AVERON also allows calculating the PPIScores based on the proteomics data available through the NCI Clinical Proteomic Tumor Analysis Consortium.

First, for a given cancer type, we identify samples with the D^{mut} mutation ($samples^{mut}$) and without the mutation ($samples^{wt}$). Samples with deletions of either D or B are eliminated from the analysis. Then, we calculate the neoPPI scores based on the expression of D and B in $samples^{mut}$ and $samples^{wt}$ samples. We have determined that the arithmetic mean of the log(x+1) transformed mRNA expression of D and B outperforms other ways of data scaling and transformation such as the geometric mean of the untransformed data or z-scaling. The neoPPI scores are calculated for each mutated sample i as $PPI_i^{neo} = \frac{D_i^{mut}+B_i}{2}, i \in samples^{mut}$. Similarly, the AVERON calculates the wild-type PPI scores for the wild-type samples $j \in samples^{wt}$ as $PPI_j^{wt} = \frac{D_j^{wt}+B_j}{2}$, where and D_j^{wt}, B_j are the log2(x+1) transformed mRNA expression of D^{wt} and B.

Identification of neoPPI-regulated genes

We hypothesize that neoPPIs exert oncogenic functions by regulating the expression of specific gene sets. To determine genes whose expression correlates with the neoPPI levels in the mutant samples we use the PPI^{neo} scores. We determine Pearson correlations between the PPI^{neo} scores and the expression of all human genes in $samples^{mut}$. To avoid artificial correlations, we eliminate the outliers using Cook's distance method. By default, if in $samples^{mut}$ gene g has an uncorrected p-value $p_g < 0.05$ and correlation $R_g^{neo} \geq |0.3|$, then we will say that the expression of gene g can be associated with the neoPPI level. The AVERON Notebook provides a user interface to adjust these parameters. The positive correlation indicates that gene g can be upregulated through the neoPPI, and the negative correlation indicates its downregulation.

If a gene g has an uncorrected p-value $p_g < 0.05$ and a correlation $R_g^{wt} \geq |0.3|$ in $samples^{wt}$, it can be also associated with the wild-type PPI^{wt}. The AVERON can identify genes with more significant correlations with PPI^{neo} scores than PPI^{wt} scores. For this purpose, the R_g^{neo} and R_g^{wt} are transformed into z-scores Z_g^{neo} and Z_g^{wt} using Fisher's Z-transformation, and z-test statistics are obtained as $Z_g = \frac{Z_g^{neo} - Z_g^{wt}}{\sqrt{\frac{1}{N_{mut} - 3} + \frac{1}{N_{wt} - 3}}}$, where N_{mut} and N_{wt} are the sizes of $samples^{mut}$ and $samples^{wt}$ after the elimination of outliers. The p-value (converted from Z_g) $p < 0.05$ and $\left|R_g^{neo}\right| > \left|R_g^{wt}\right|$ indicate that gene g is regulated by the neoPPI. $R_g^{neo} > 0.3$ indicates positively regulated genes $G_{up} = \{g_1, g_2, \ldots, g_i\}$. $R_g^{neo} < -0.3$ will indicate negatively regulated genes $G_{down} = \{g_1, g_2, \ldots, g_j\}$.

Pathway enrichment analysis

The AVERON can conduct enrichment analysis based on any gene sets provided in the Gene Matrix Transposed (GMT) format. For example, in this study, we used the pathway gene sets publicly available through the MSigDB databases.[27] The statistical significance of the pathway enrichment is evaluated through Fisher's exact test and the Benjamini-Hochberg procedure to control for the false discovery rate (FDR).[41] The p-value <0.05 and FDR <5% are used as thresholds for statistical significance.

Survival analysis

The survival analysis implemented in the current version of the AVERON is based on the clinical outcomes of cancer patients available through the NCI GDC Portal for the TCGA PanCancer Atlas cohorts.[42] The analysis is performed for the patients with the corresponding mutation(s) of the tumor driver gene. The AVERON determines the association between survival times of patients harboring the mutation(s) in the tumor driver gene and either calculated neoPPI levels or expression of neoPPI-regulated genes. Both the multivariate COX regression analysis and univariate Kaplan-Meier analysis implemented in the AVERON Notebook are based on the lifelines 0.27.8 Python library. For the Kaplan-Meier analysis, we compare the survival times of patients with high (above the median) and low (below the median) PPI^{neo} scores. To determine neoPPI-regulated genes that may contribute to poor clinical outcomes, the AVERON compares the survival times of patients with high and low expression of the target gene. With default parameters, the high expression of a gene g is defined as the expression above the 67th percentile, and the low expression is defined as the gene expression below the 33rd percentile in $samples^{mut}$. The AVERON Notebook allows the user to change those thresholds. The statistical significance of the correlation between neoPPI levels and patient survival is determined in terms of the log rank test p-values and represented with the Kaplan-Meier plot.

Integration of pharmacological data

The AVERON links neoPPI-target genes with clinical and FDA-approved compounds. The IUPHAR database release v. 2024.1 includes 3,067 human targets, 12,590 ligands, including 1,981 approved drugs, and 3,704 ligands with clinical use summaries. The compound-protein associations available through the IUPHAR database[43] are determined using the pyGtoP Python wrapper for the IUPHAR API. For each gene, the target-ligand pairs are identified based on the standard gene symbol. The search is conducted for the human genes. The ligand names associated with the target gene are extracted along with the IUPHAR ligand IDs, ligand type, and approval status. The data can be exported to a comma-separated (csv) file for further analysis. In addition, the total number of associated ligands as well as the number of approved and not approved ligands is counted for each gene and saved in a separate table.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using custom Python scripts implemented in AVERON Notebook. Details of all statistical analyses can be found above in the relevant subsections of the method details section.