SUPPORTING INFORMATION

Comprehensive characterization of protein-protein interactions perturbed by disease mutations

Cheng et al., *Nature Genetics* 2021

To whom correspondence should be addressed:

Joseph Loscalzo, M.D., Ph.D.

Brigham and Women's Hospital

75 Francis St.

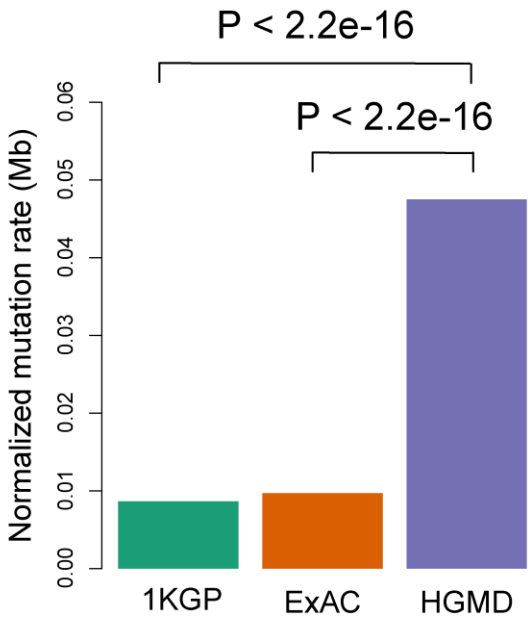
Boston, MA 02115

Phone: 617-732-6340; fax: 617-732-6439

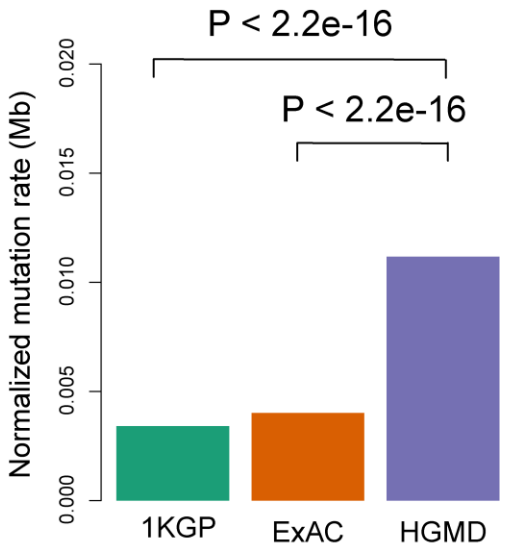
Email: jloscalzo@rics.bwh.harvard.edu

Supplemental information includes 22 Supplementary Figures (pdf file), one Supplementary Table (PDF), and two Supplementary Data files (excel files). All other supporting data are available: https://mutanome.lerner.ccf.org/

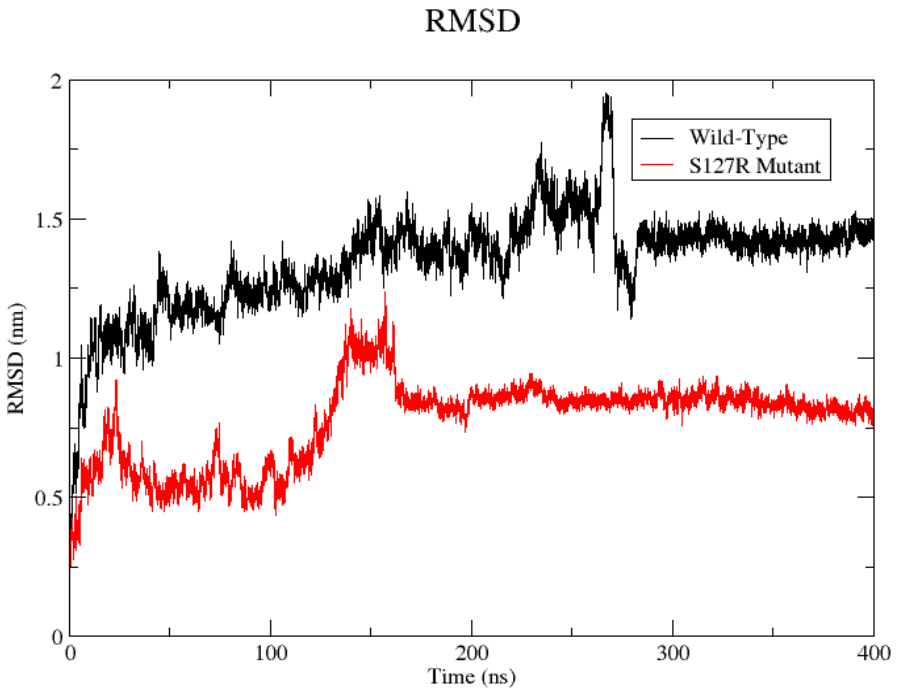
Supplementary Figures



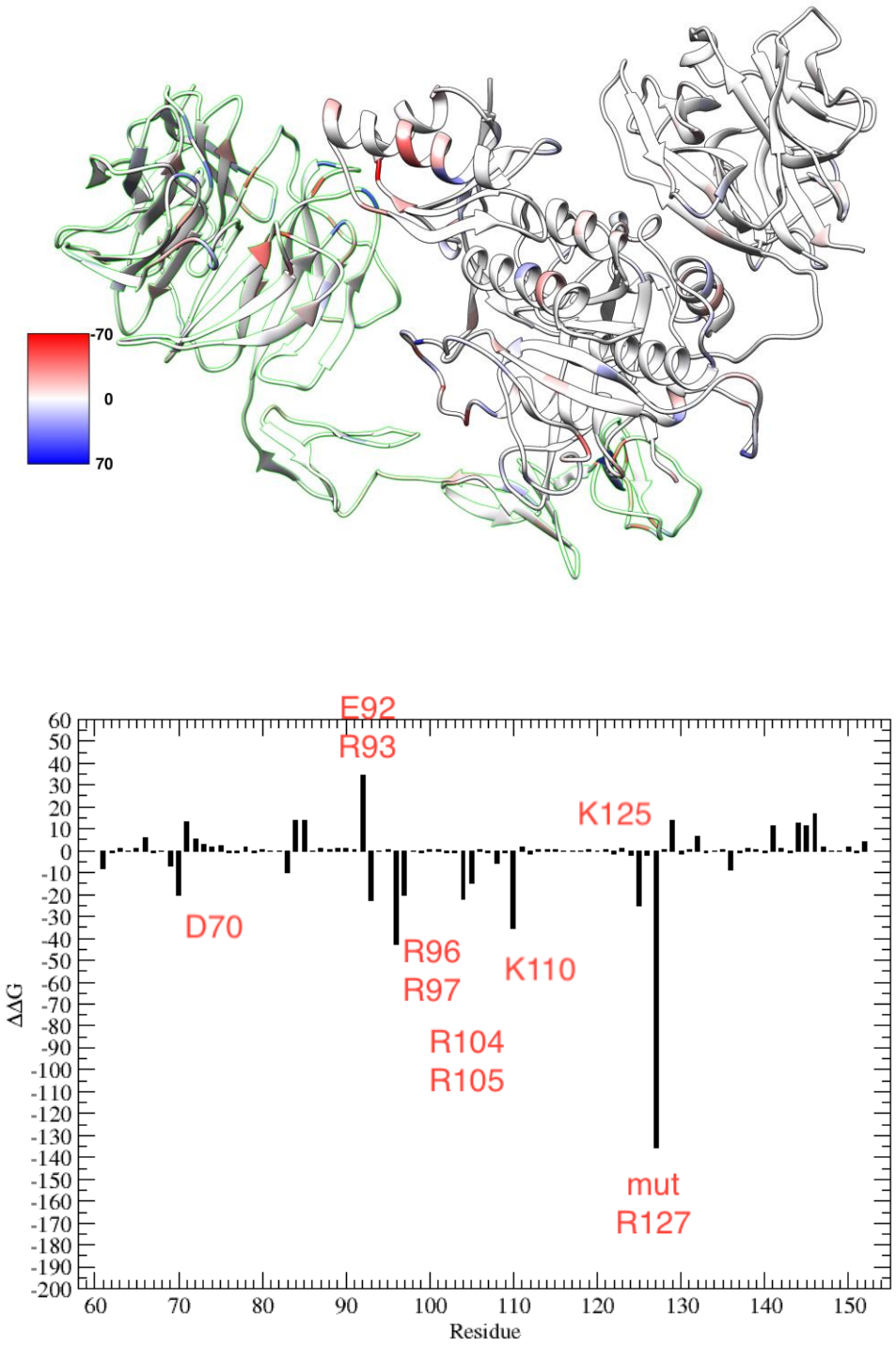
Supplementary **Figure 1**. Distribution of mutation burden at protein-protein interfaces for disease-associated germline mutations from HGMD in comparison to mutations from the 1,000 Genome Project (1KGP) and ExAC Project using 4,150 physical protein-protein interactions with known crystal structures from PDB database. P-value was calculated by two-tailed Fisher's test.



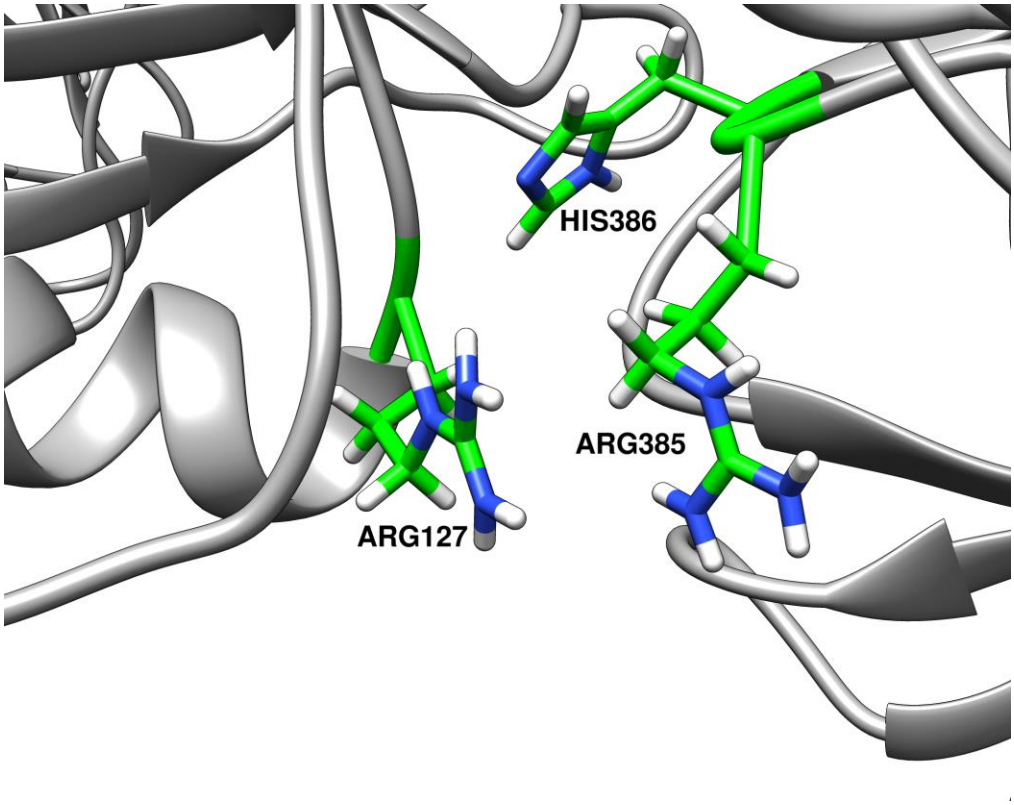
Supplementary **Figure 2**. Distribution of mutation burden at protein-protein interfaces for disease-associated germline mutations from HGMD in comparison to mutations from the 1,000 Genome Project (1KGP) and ExAC Project using interfaces from 8,230 physical protein-protein interactions identified by systematic, Y2H assay. P-value was calculated by two-tailed Fisher's test.



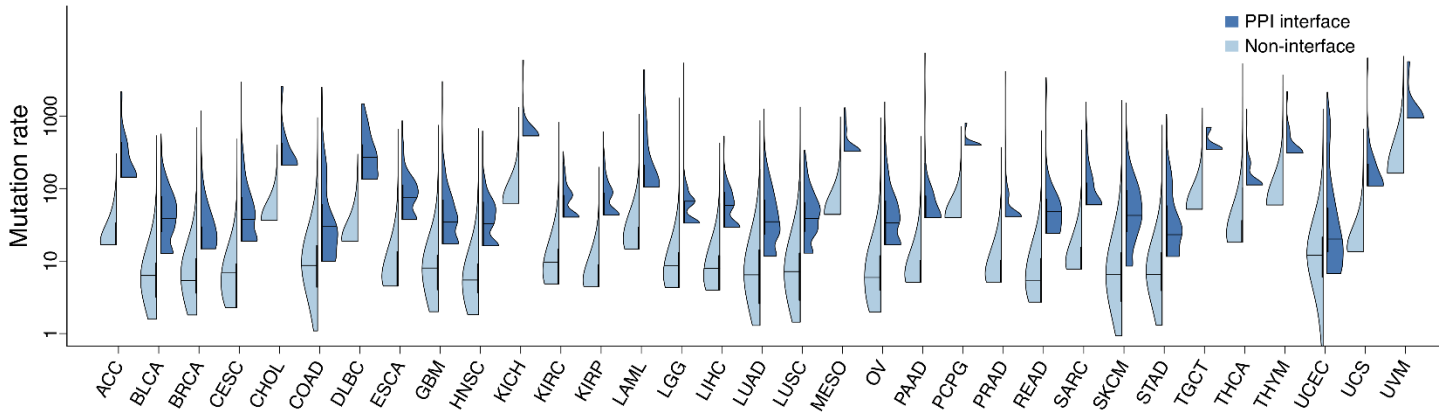
Supplementary **Figure 3**. Distribution of root-mean-squared deviation (RMSD) during 400 ns molecular dynamics simulation for two systems: PCSK9-LDLR wild-type vs. PCSK9-LDLR complex with p.Ser127Arg (S127R) on PCSK9. RMSD reveals stable systems after 150 ns molecular dynamics simulation for both wild-type and mutated systems.



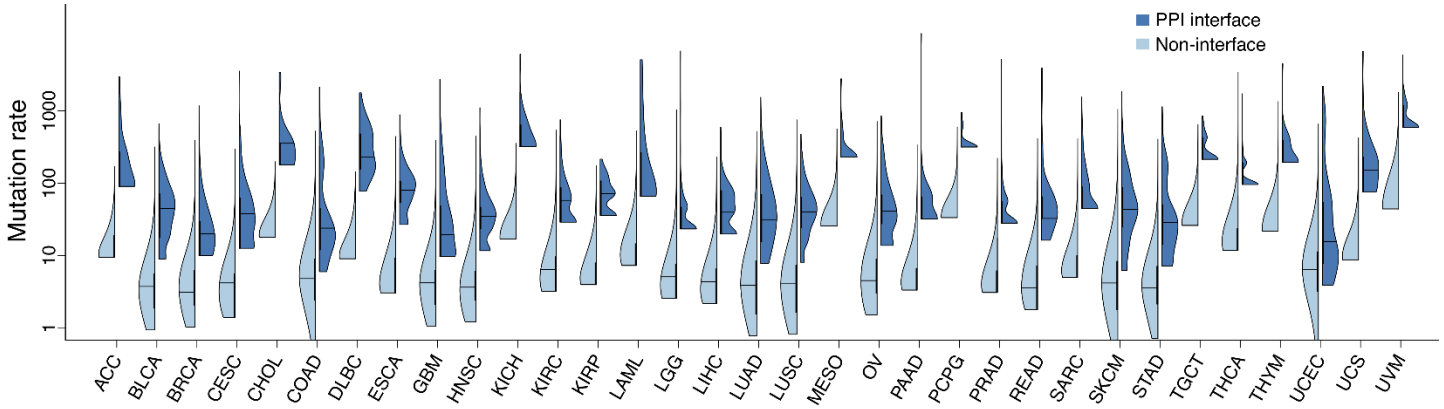
Supplementary **Figure 4.** Distribution of binding affinity (ΔΔG) for PCSK9-LDLR complex with p.Ser127Arg (S127R) on PCSK9, in the last 50 ns (350-400 ns, Supplementary **Figure 3**).



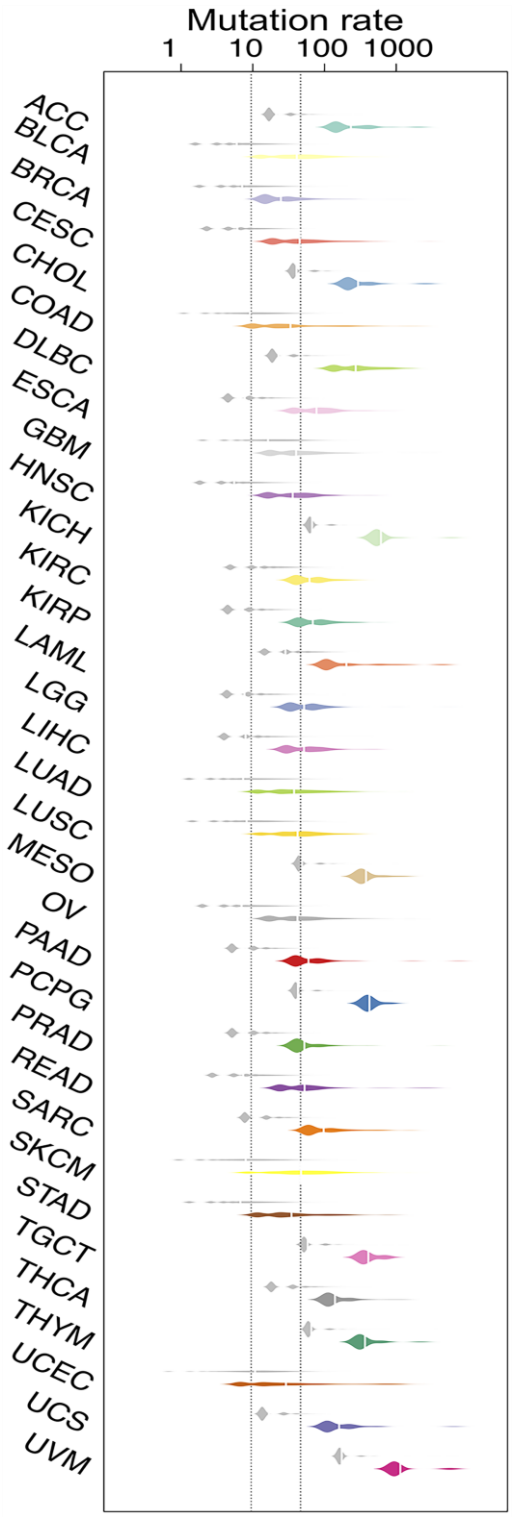
Supplementary **Figure 5.** The detailed binding model of PCSK9-LDLR complex with p.Ser127Arg (S127R) on PCSK9, in the last 50 ns (350-400 ns, Supplementary **Figure 3**).



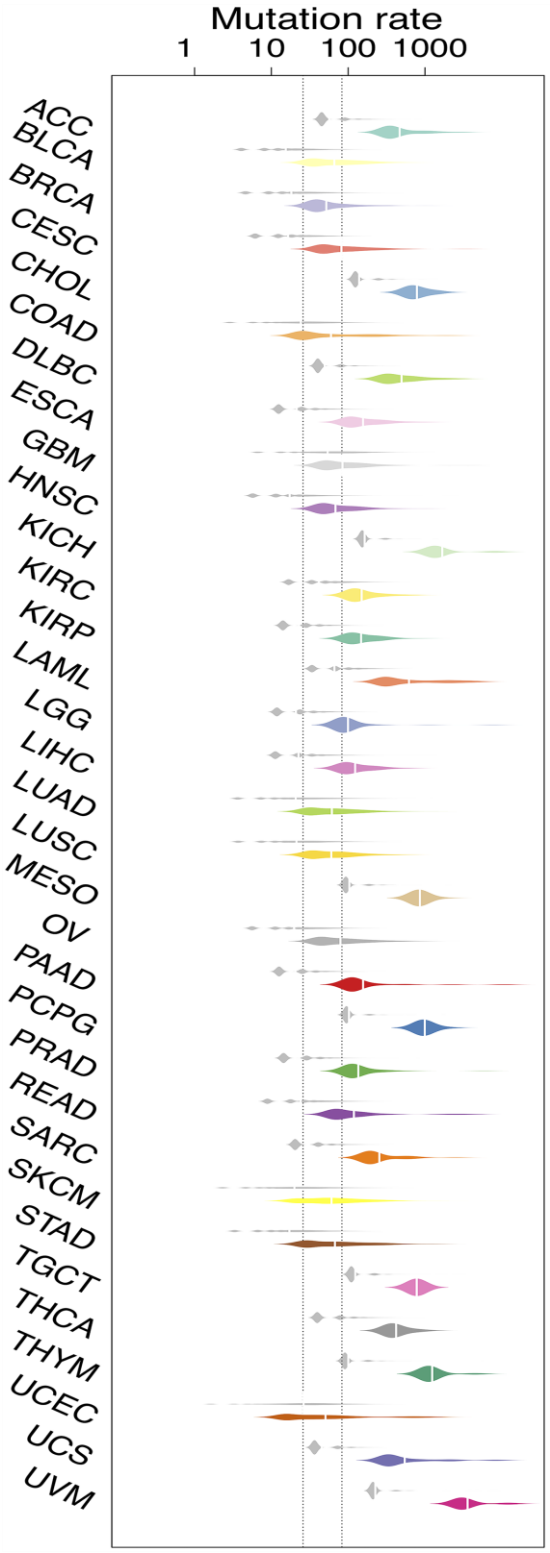
Supplementary **Figure 6.** Distribution of somatic missense mutations in protein-protein interfaces versus non-interfaces across 33 cancer types from TCGA using all physical protein-protein interactions with known crystal structures from the PDB database (www.rcsb.org). P-value < 2.2×10^{-16} (two-sided Wilcoxon test) for all 33 cancer types. 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.



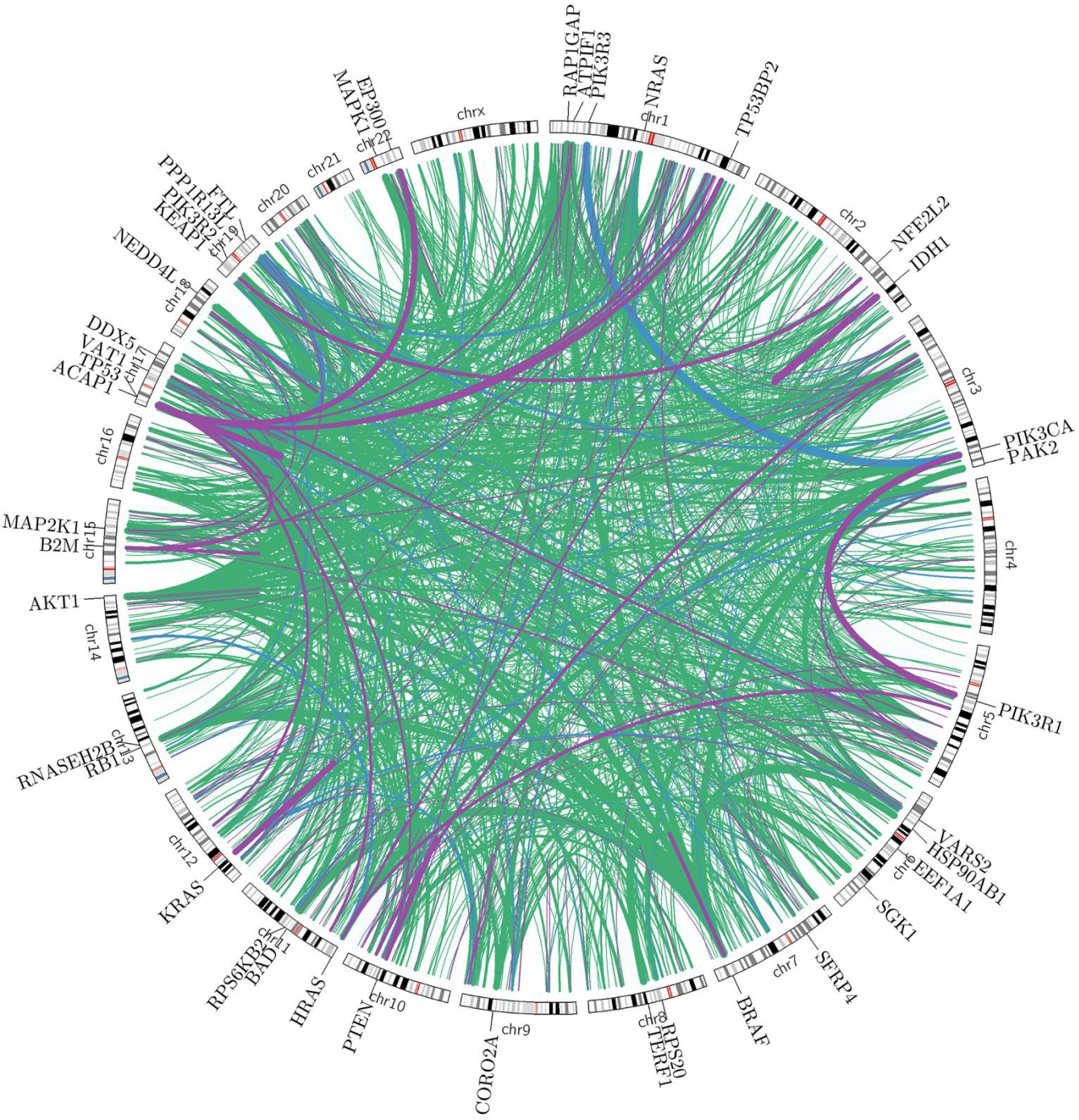
Supplementary **Figure 7.** Distribution of somatic missense mutations in protein-protein interfaces versus non-interfaces across 33 cancer types from TCGA using physical protein-protein interactions with computationally predicted interfaces alone from the ECLAIR database (http://interactomeinsider.yulab.org). P-value < 2.2×10^{-16} (two-sided Wilcoxon test) for all 33 cancer types. 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.



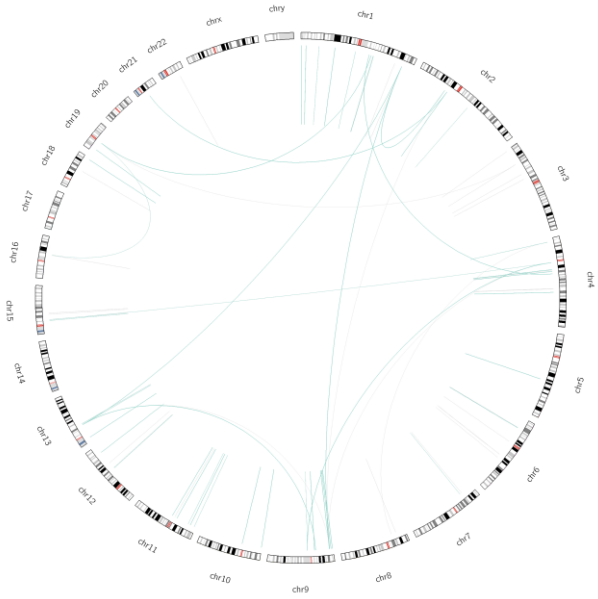
Supplementary **Figure 8.** Distribution of somatic missense mutations in protein-protein interfaces (color bins) versus non-interfaces (gray bins) across 33 cancer types from TCGA using 4,150 physical, binary protein-protein interactions with known crystal structures from PDB database. P-value < 2.2×10^{-16} (two-sided Wilcoxon test) for all 33 cancer types. The data are represented as a boxplot with a underlaid violin plot where the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent the interquartile range (IQR) ×1.5 and beyond the whiskers are outlier points.



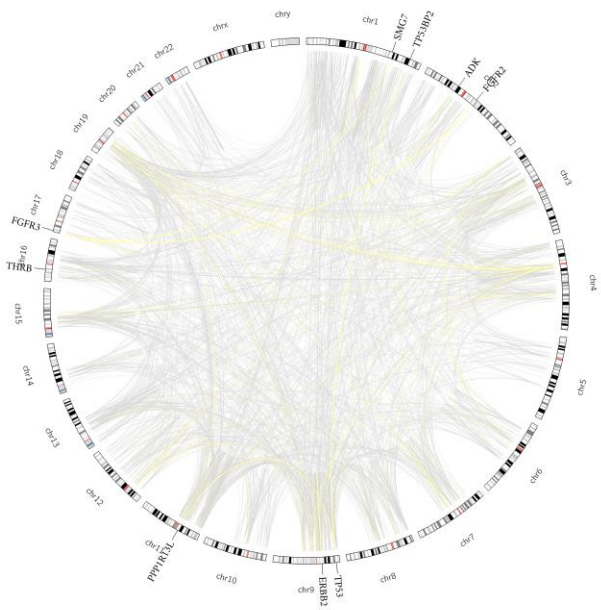
Supplementary **Figure 9**. Distribution of somatic missense mutations in protein-protein interfaces (color bins) versus non-interfaces (gray bins) across 33 cancer types from TCGA using interfaces from 8,230 physical protein-protein interactions identified by systematic, Y2H assay. P-value < 2.2×10^{-16} (two-sided Wilcoxon test) for all 33 cancer types. The data are represented as a boxplot with a underlaid violin plot where the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent the interquartile range (IQR) ×1.5 and beyond the whiskers are outlier points.



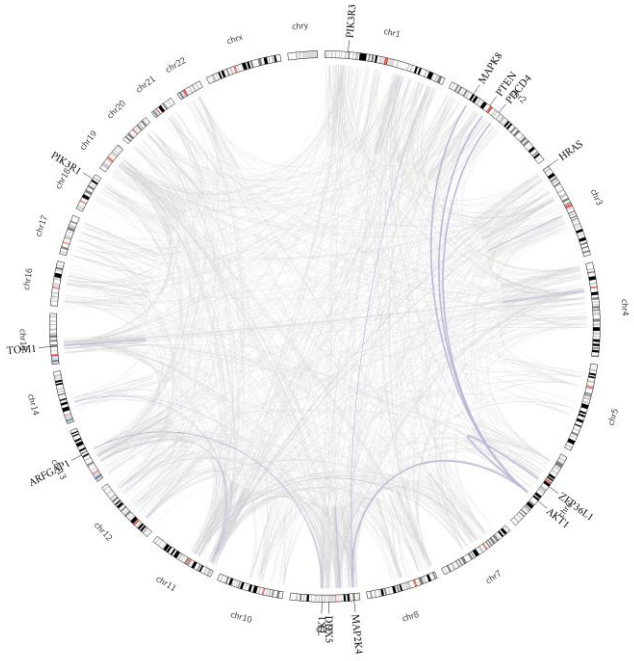
Supplementary **Figure 10.** Landscape of significant mutation-perturbed PPIs (termed putative oncoPPIs) which harbor a statistically significant excess number of missense mutations at PPI interfaces in pan-cancer analysis. The links in the circus plot represent putative oncoPPIs and weights (i.e., line thickness) of links indicate significance (adjusted p-value < 0.001, bold links reveal the most significant adjusted p-values [please see Supplementary **Data 1**]). The color of edges denotes the different types of PPI interfaces: PDB (purple), I3D (blue), and ECLAIR (green).



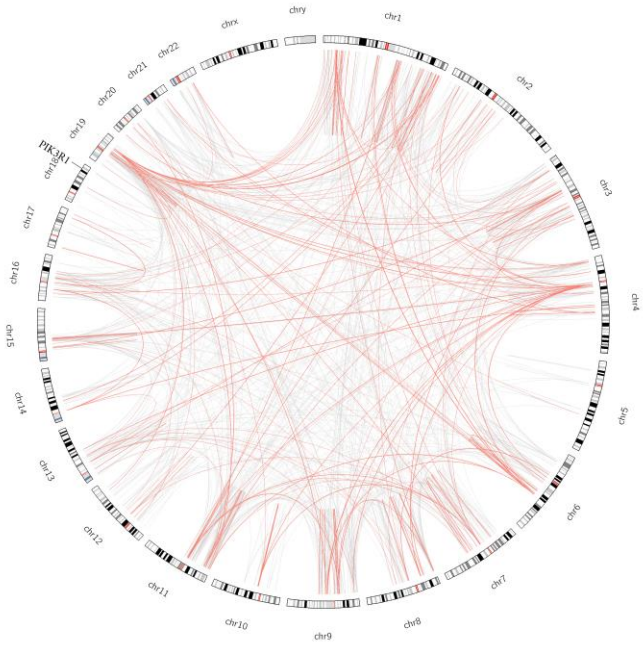
(ACC)



(BLCA)

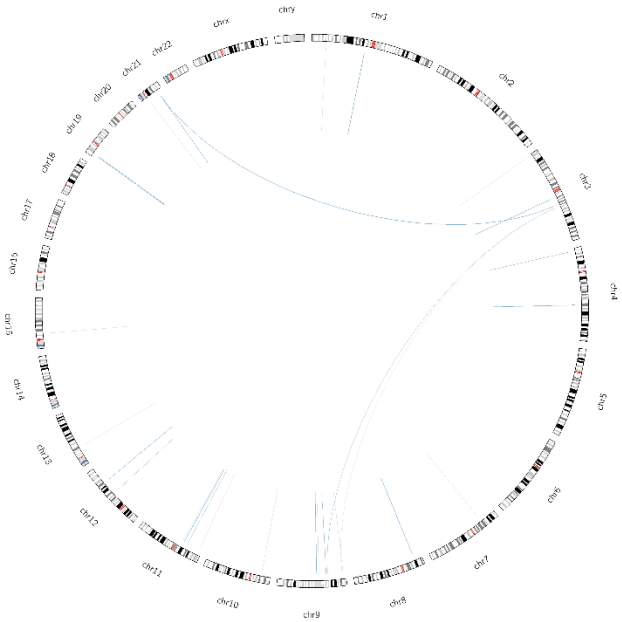


(BRCA)

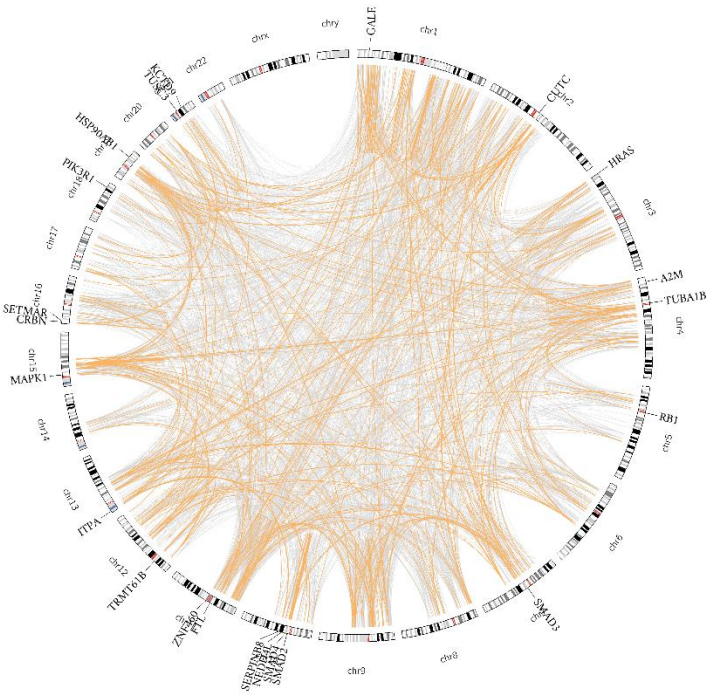


(CESC)

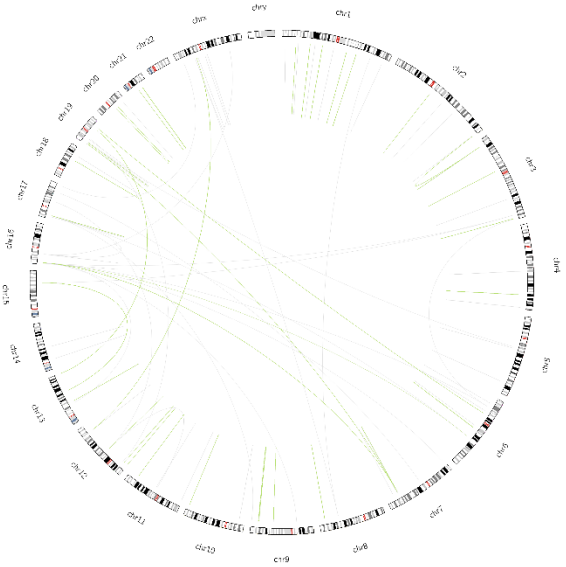
Supplementary **Figure 11**. Landscape of significant mutation perturbed PPIs (termed putative oncoPPIs) which harbor a statistically significant excess number of missense mutations at PPI interfaces in individual cancer types. The links in the circus plot represent putative oncoPPIs and weights (i.e., line thickness) of links indicate significance (p-value, bold links reveal the most significant p-value [see Online Methods]).



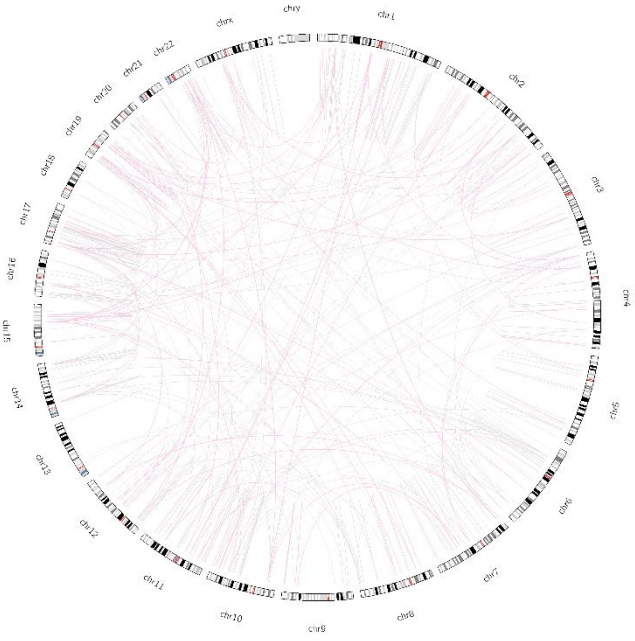
(CHOL)



(COAD)

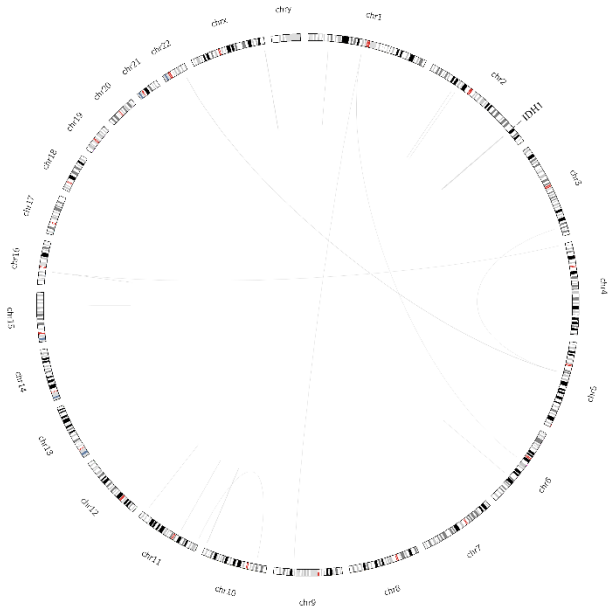


(DLBC)

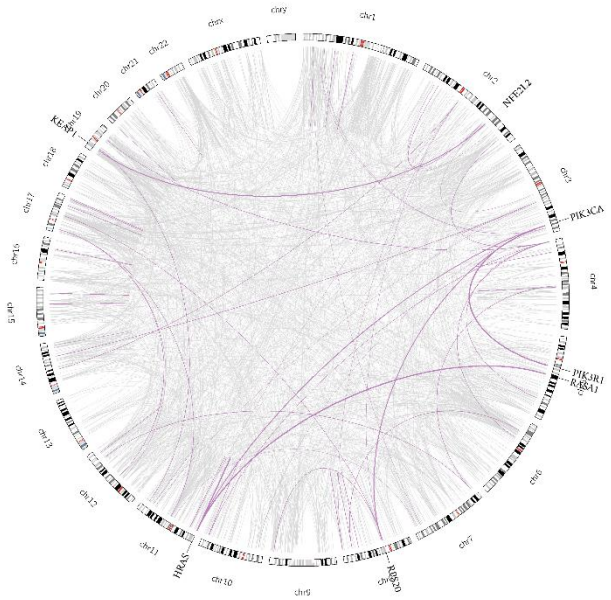


(ESCA)

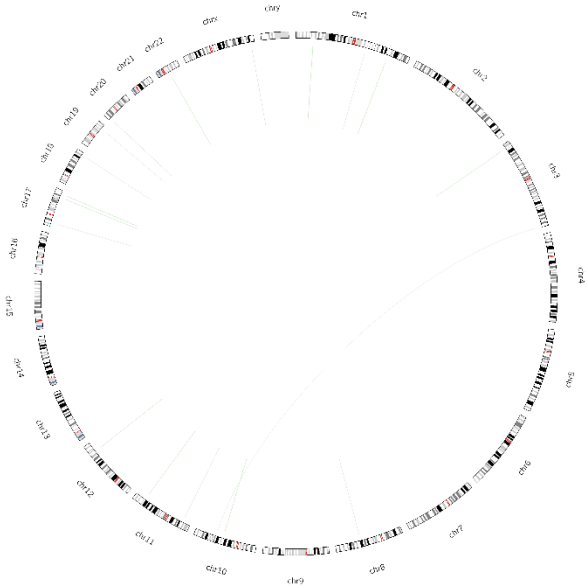
Supplementary **Figure 11** continued.



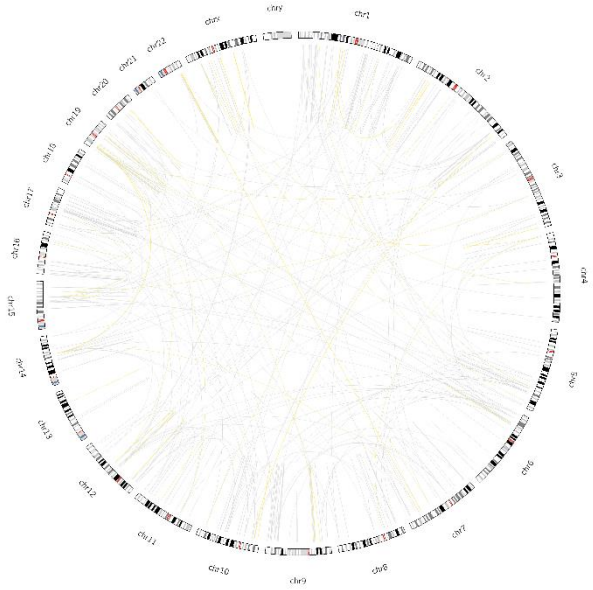
(GBM)



(HNSC)

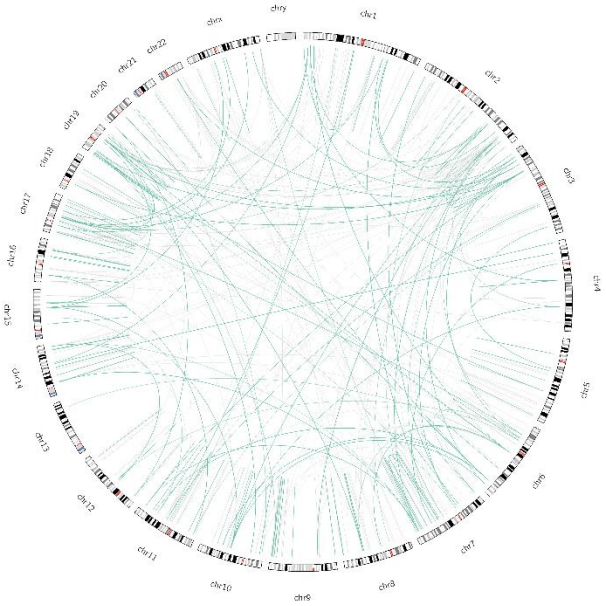


(KICH)

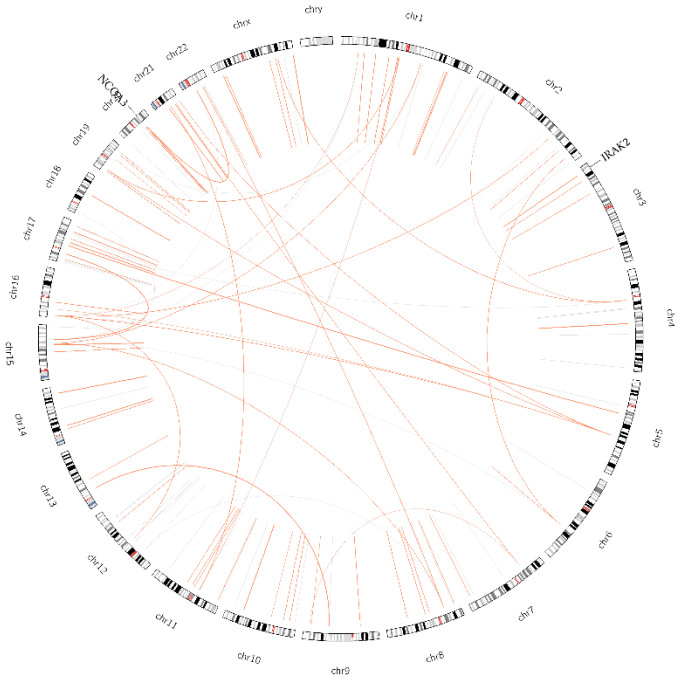


(KIRC)

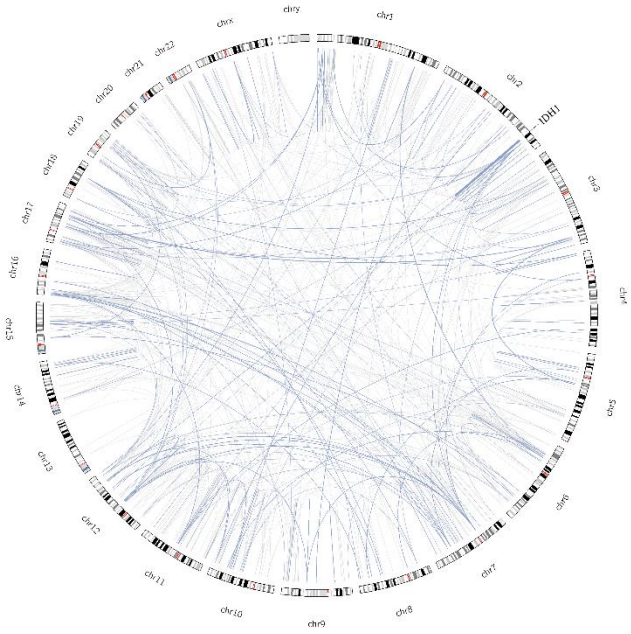
Supplementary **Figure 11** continued.



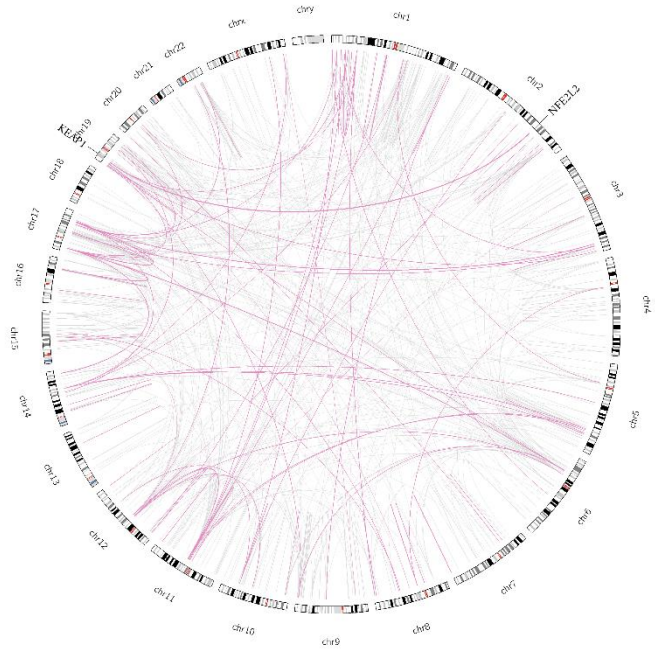
(KIRP)



(LAML)

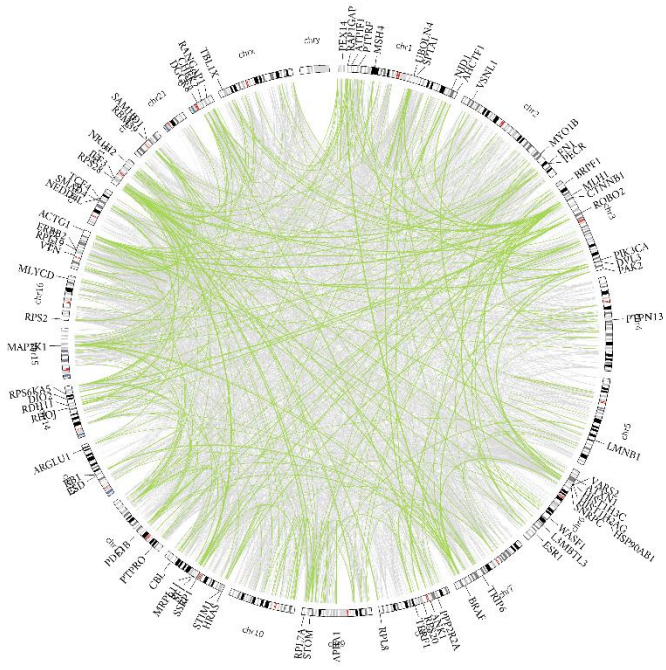


(LGG)

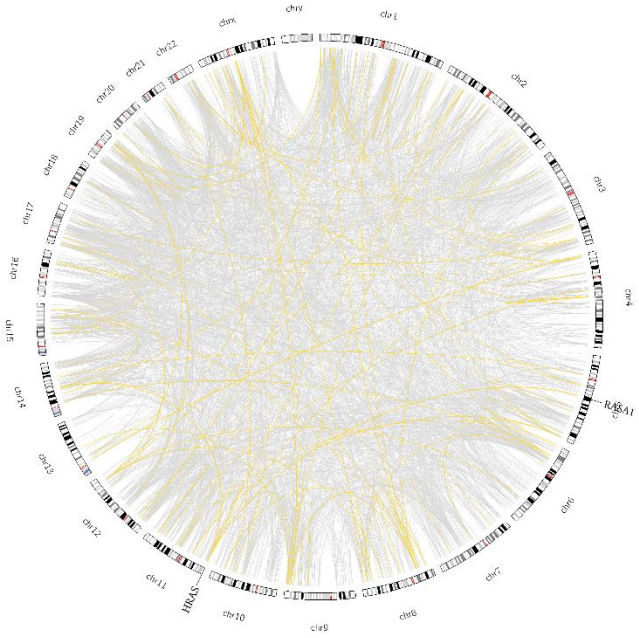


(LIHC)

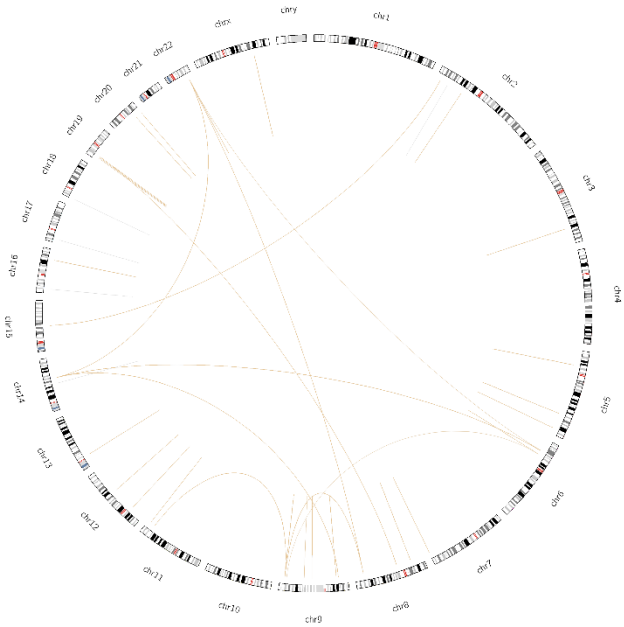
Supplementary **Figure 11** continued.



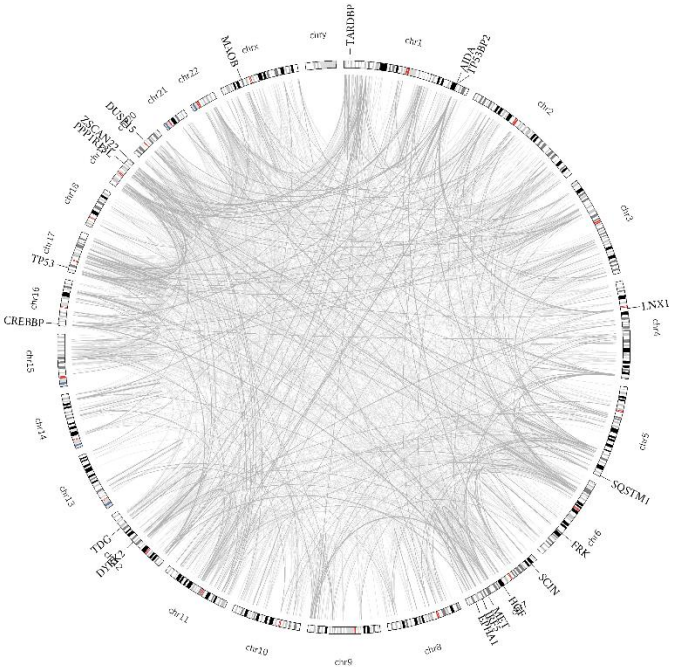
(LUAD)



(LUSC)

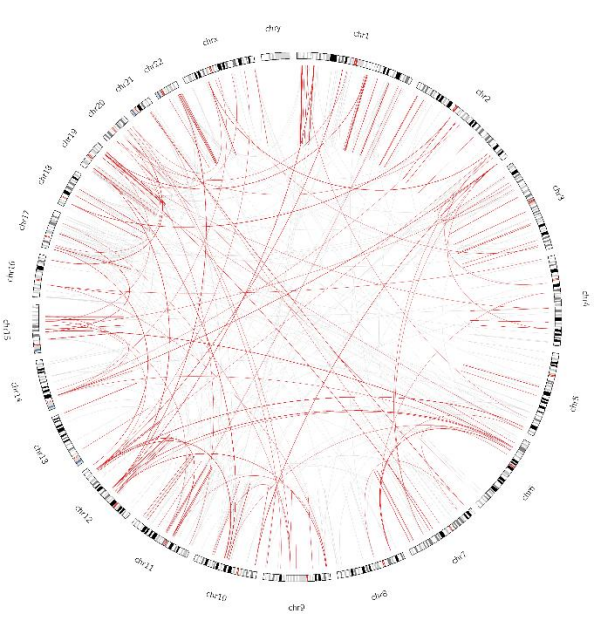


(MESO)

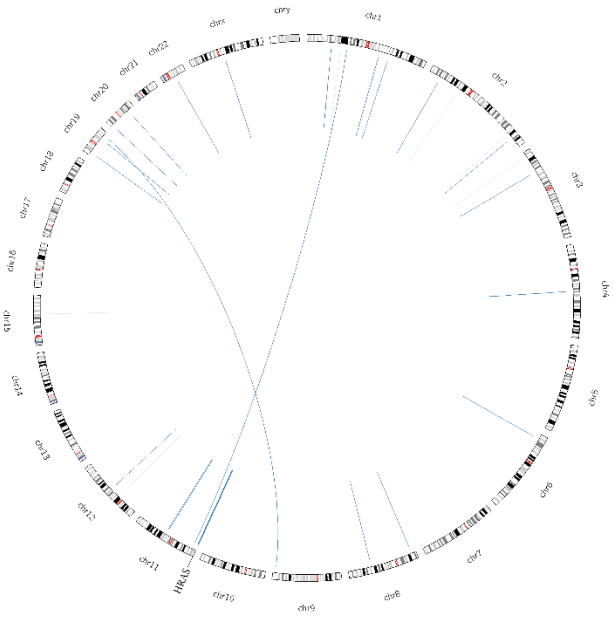


(OV)

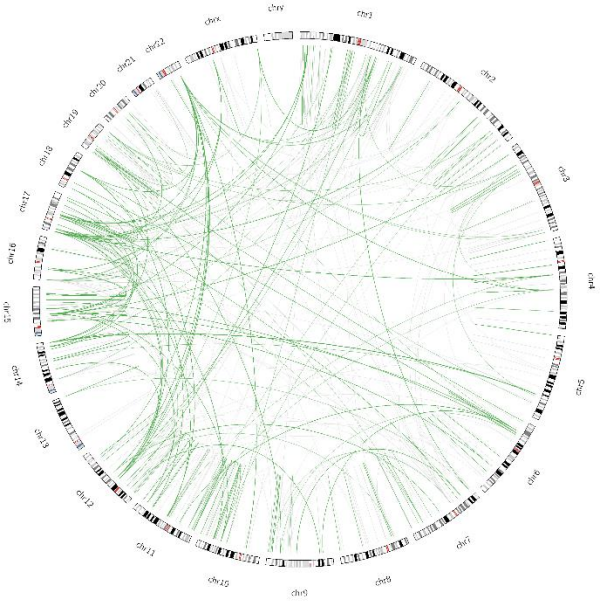
Supplementary **Figure 11** continued.



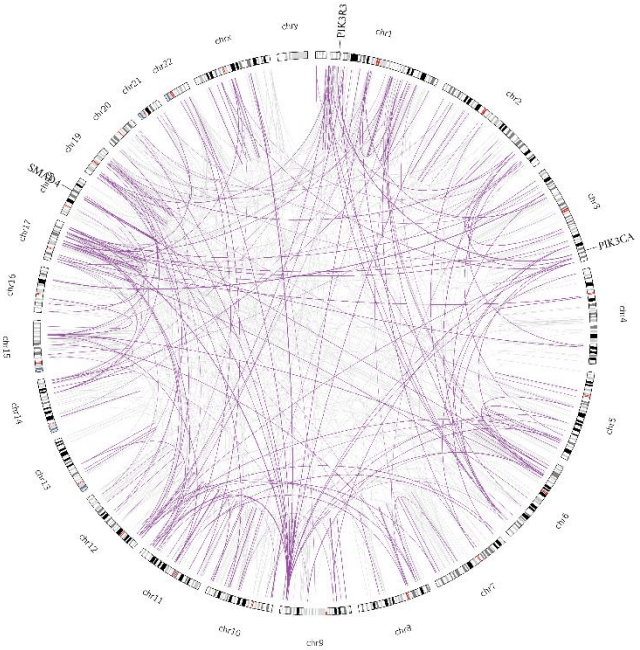
(PAAD)



(PCPG)

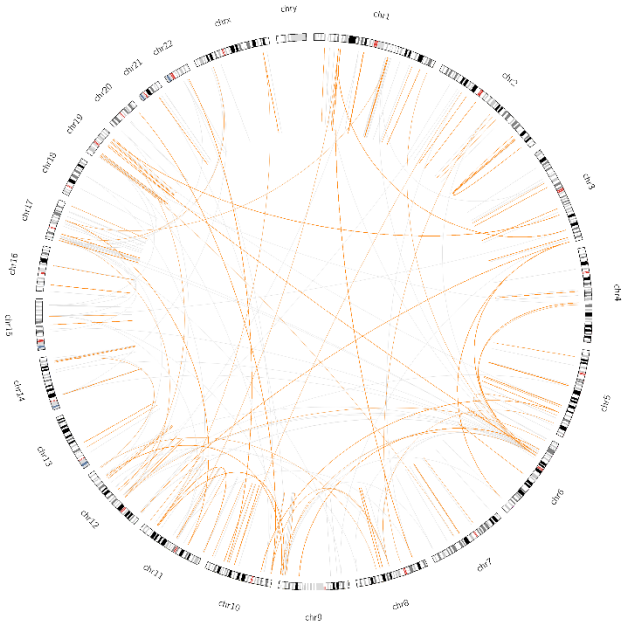


(PRAD)

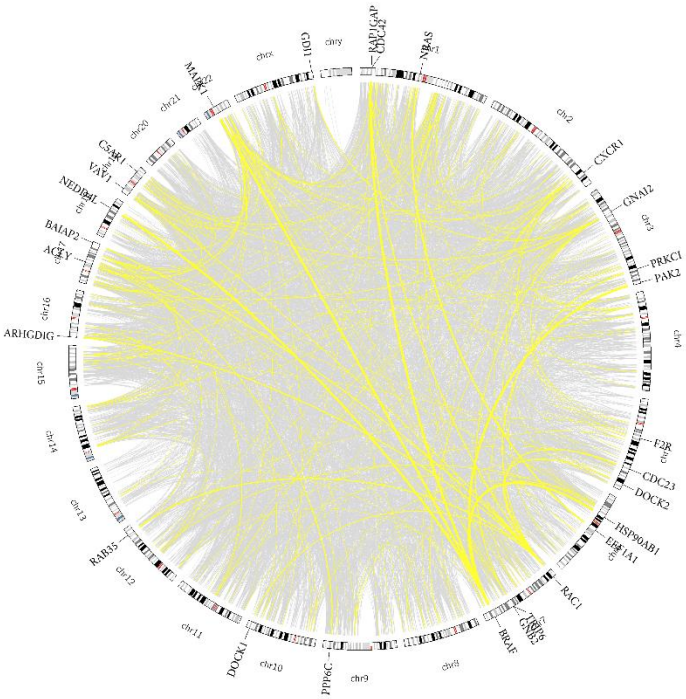


(READ)

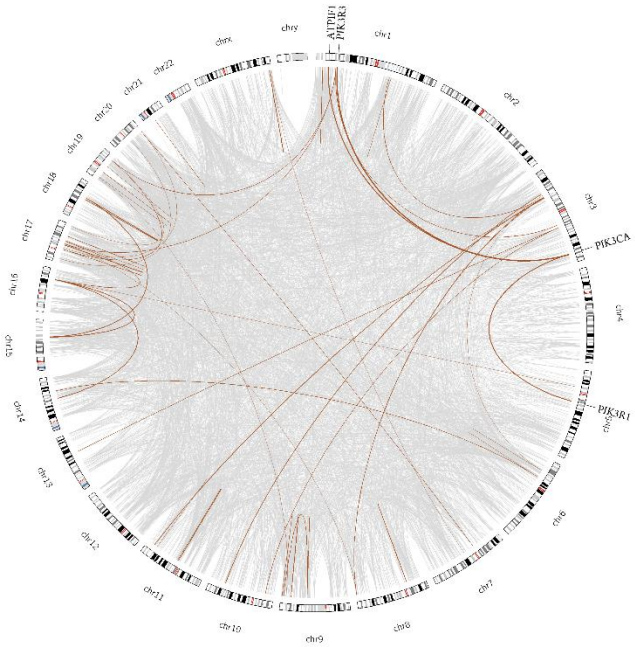
Supplementary **Figure 11** continued.



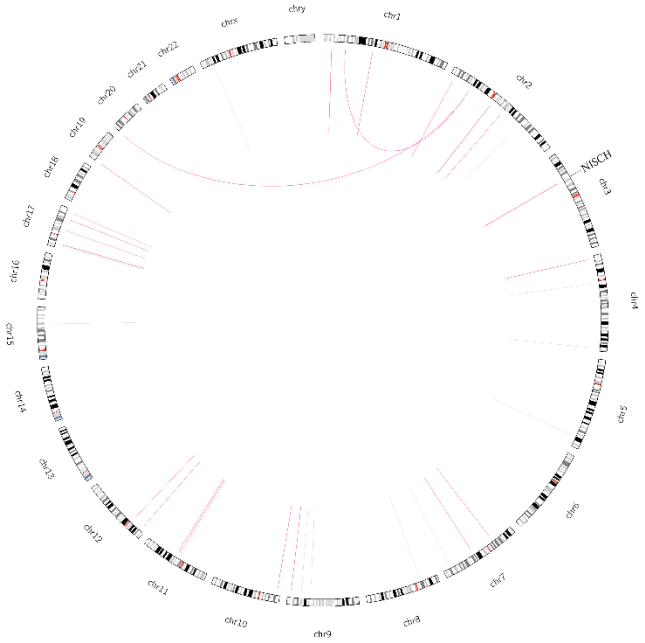
(SARC)



(SKCM)

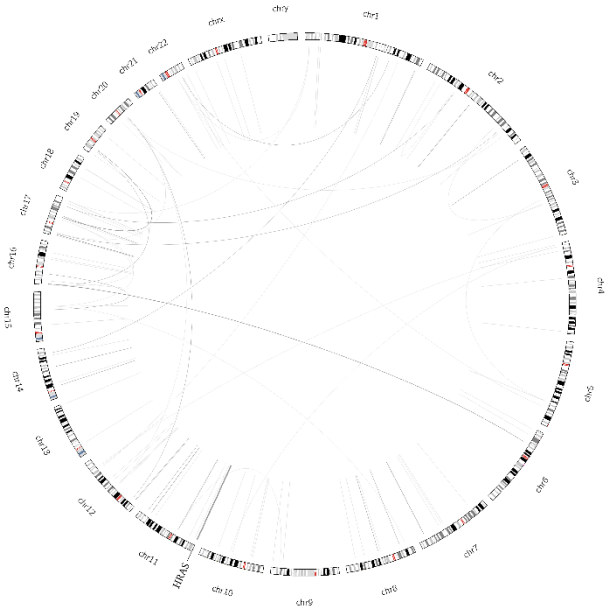


(STAD)

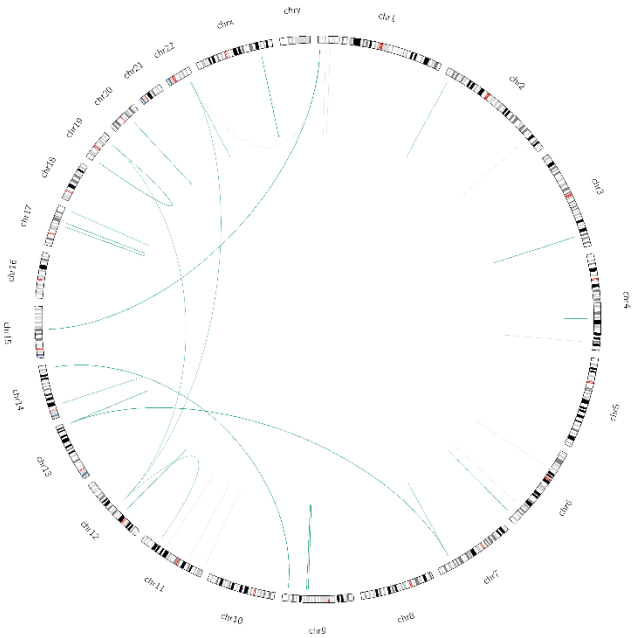


(TGCT)

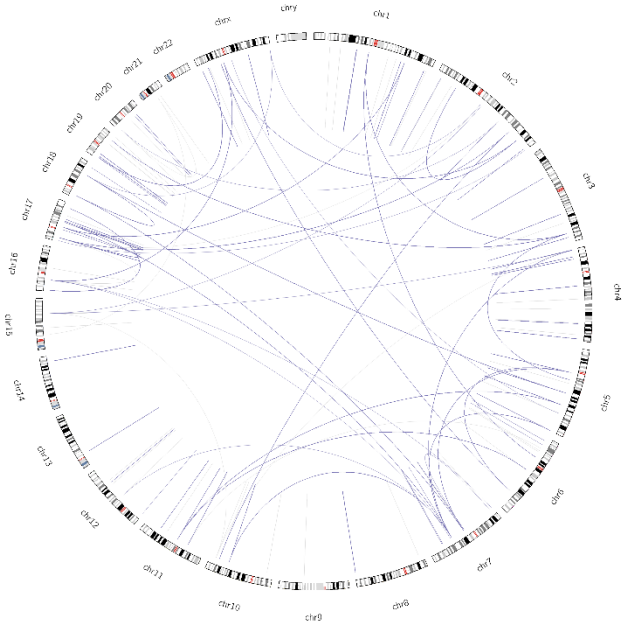
Supplementary **Figure 11** continued.



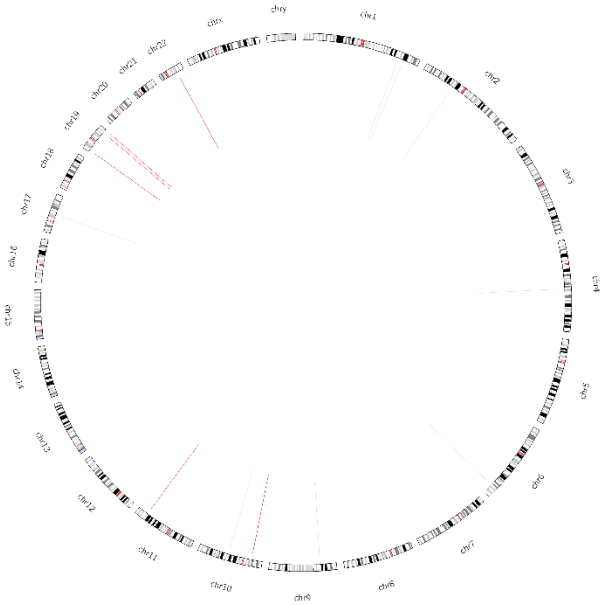
(THCA)



(THYM)

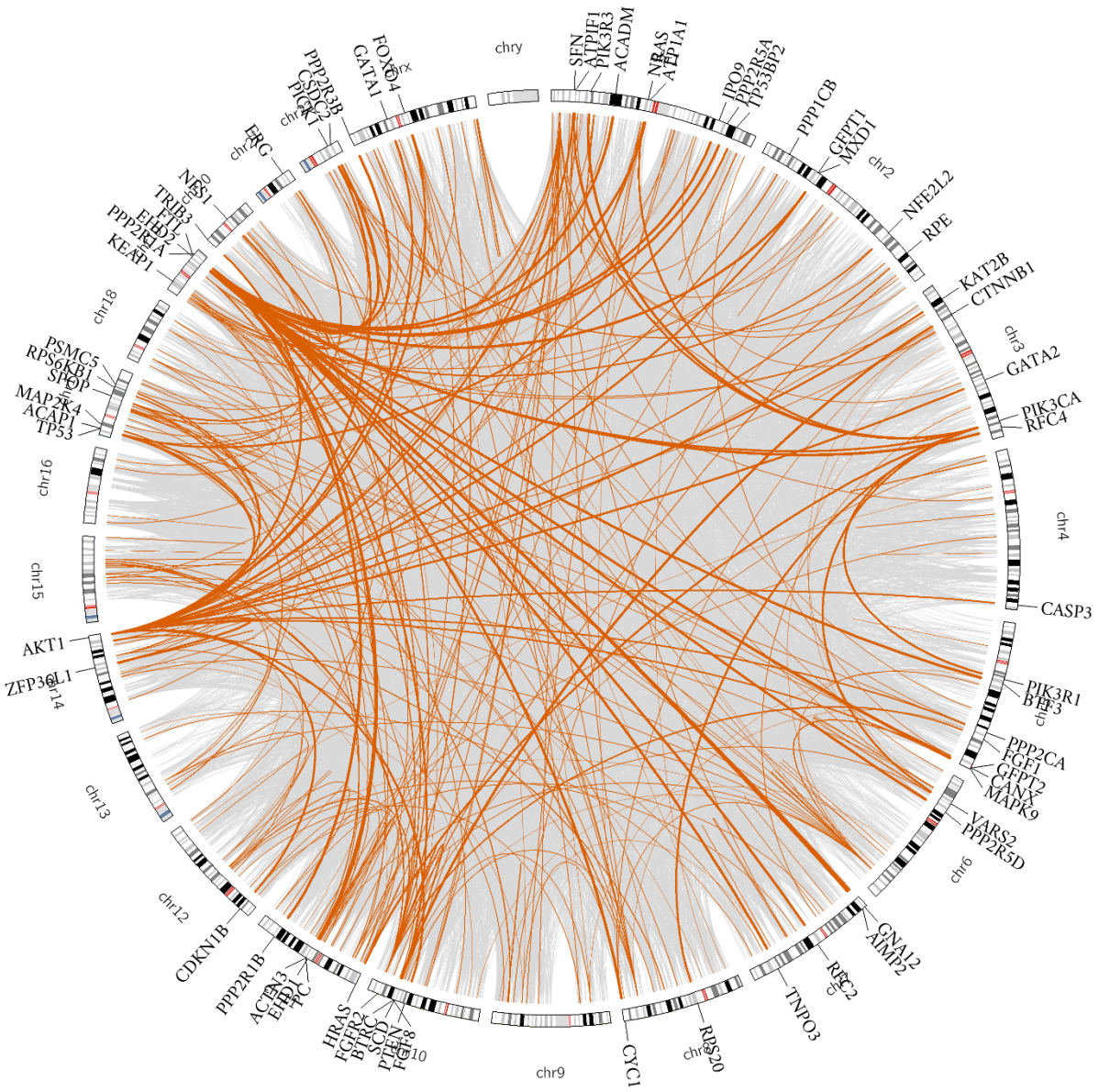


(UCS)



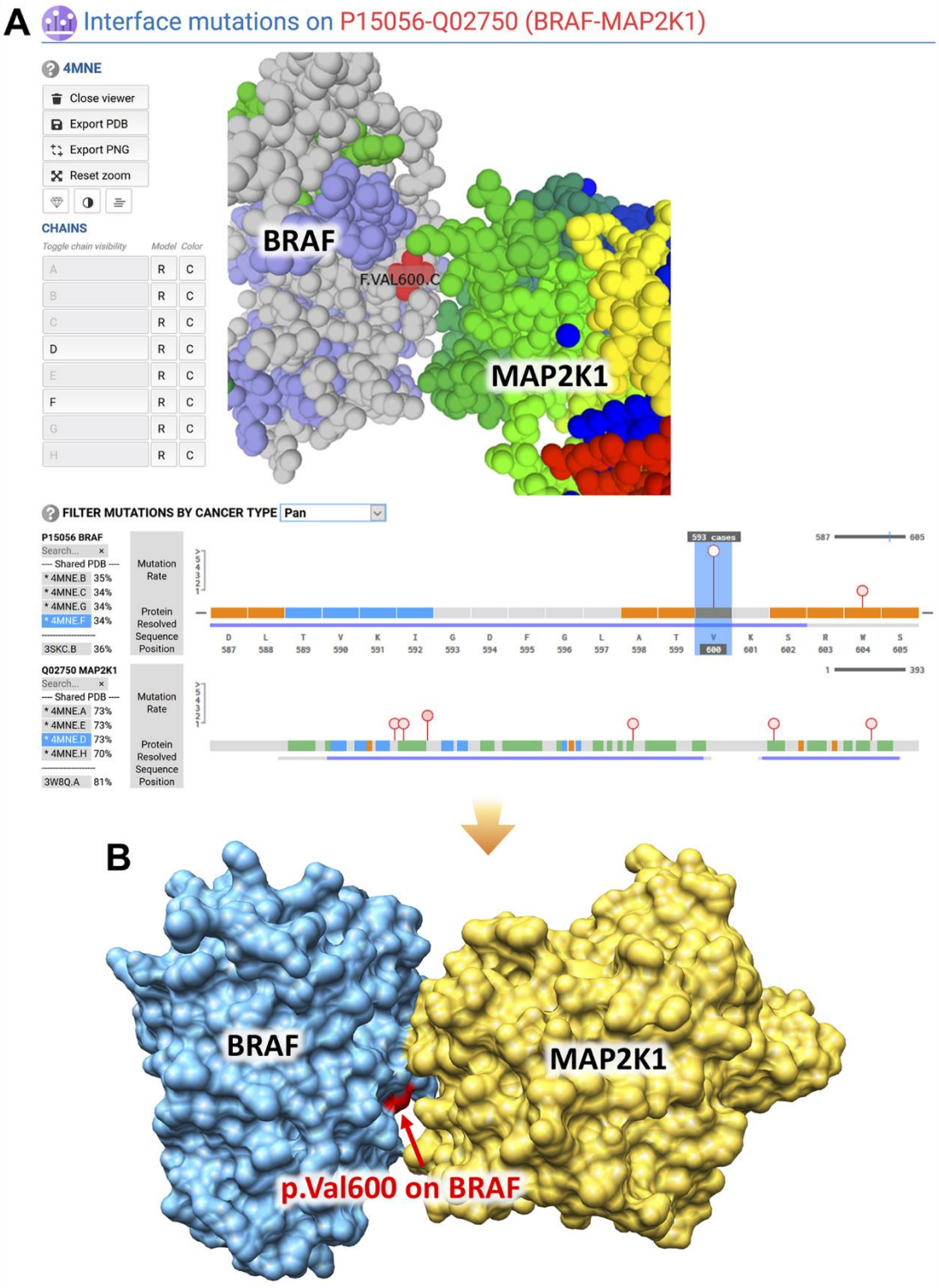
(UCS)

Supplementary **Figure 11** continued.

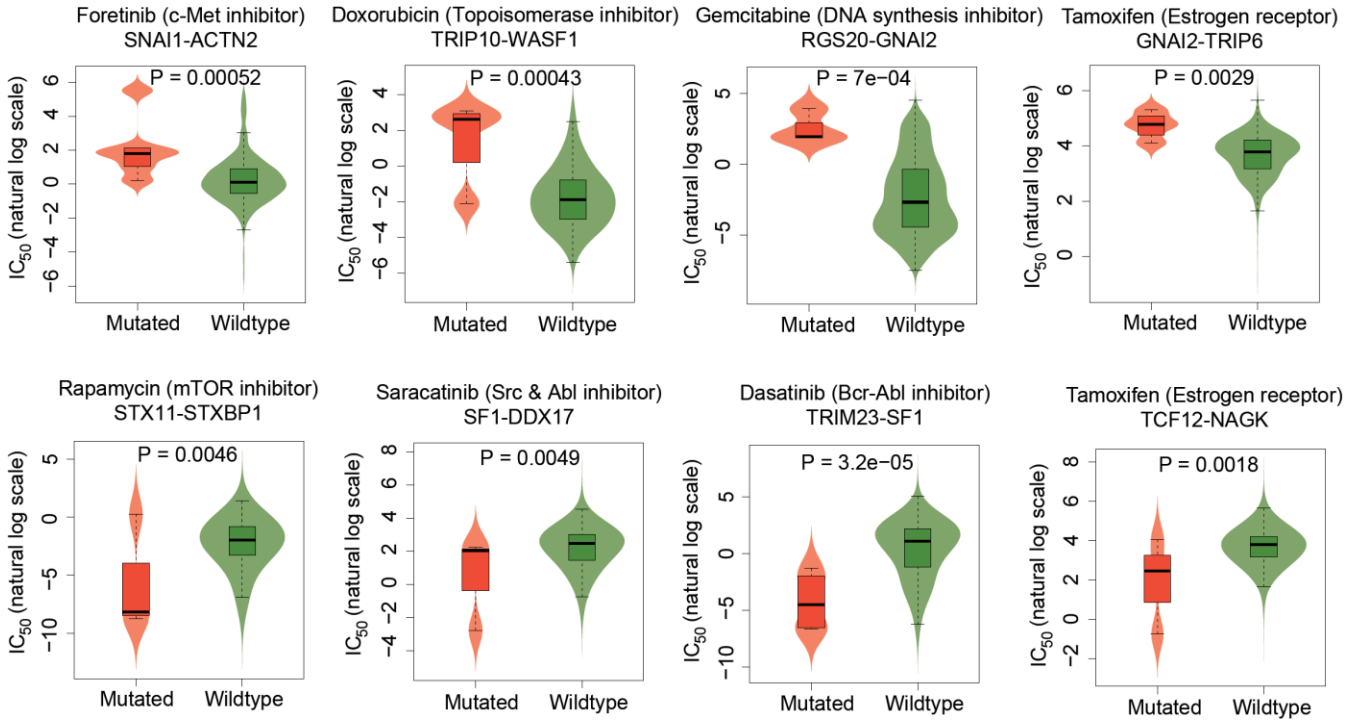


(UCEC)

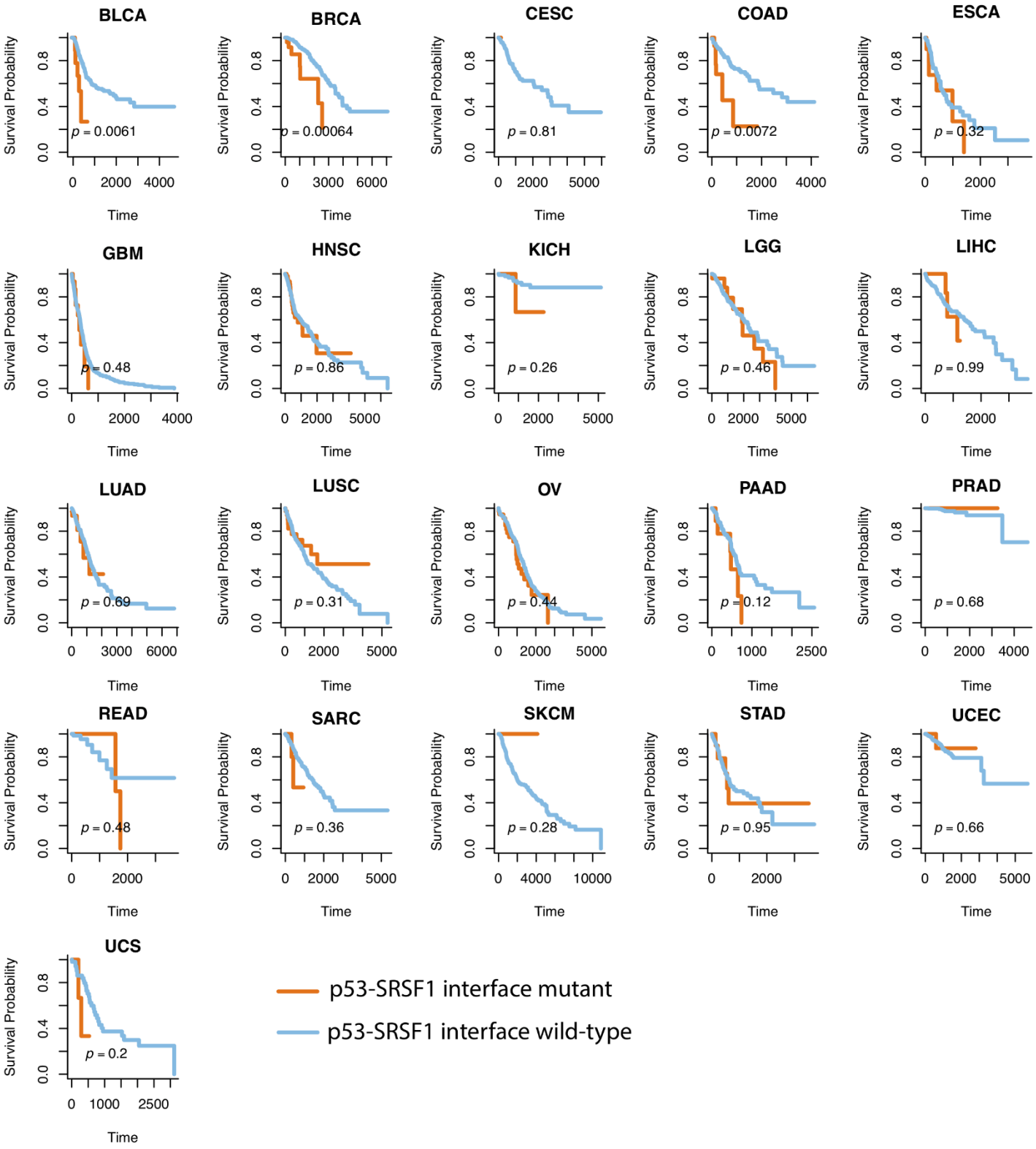
Supplementary **Figure 11** continued.



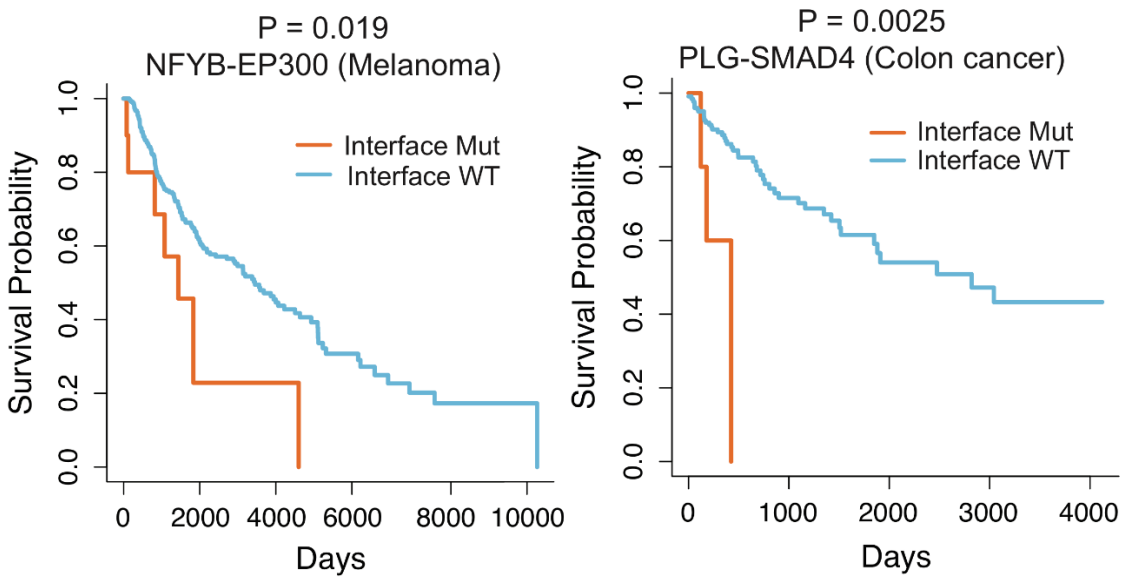
Extended Figure 12. p.Val600Glu on BRAF is a BRAF-MAP2K1 interface mutation. BRAF-MAP2K1 interface mutation visualization by (**A**) the mutanome tool (https://mutanome.lerner.ccf.org) and PyMOL (https://pymol.org/2/) (**B**).



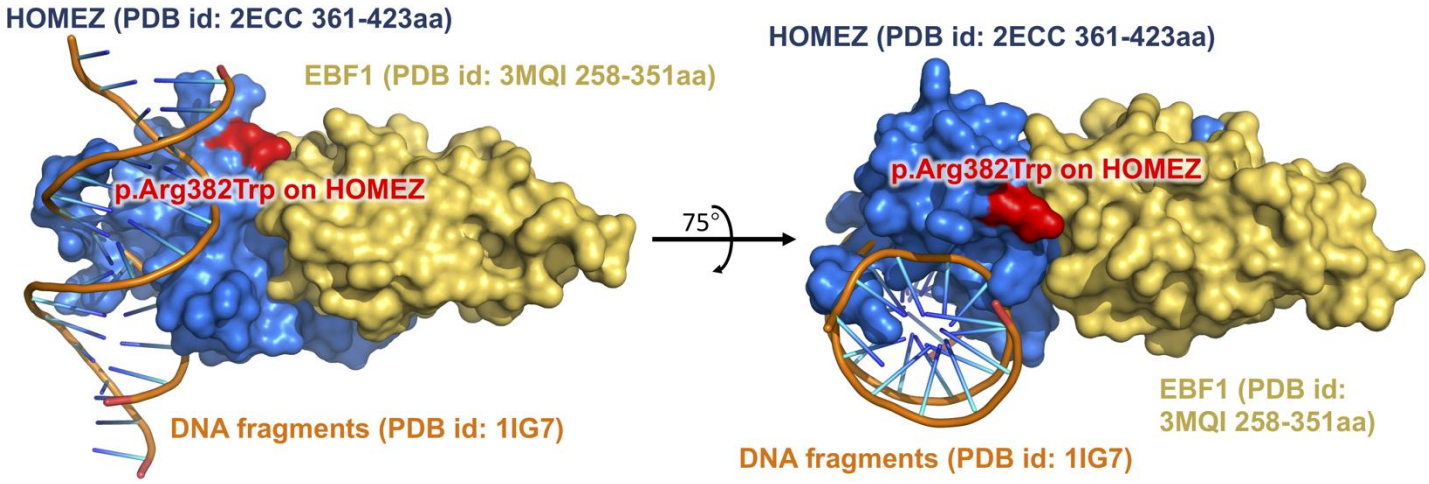
Supplementary **Figure 13**. Eight selected examples of drug responses predicted by mutation perturbed PPIs (oncoPPIs) that harbor a statistically significant excess number of missense mutations at PPI interfaces as predicted by a binomial distribution. All oncoPPI-predicted drug responses are freely available at the website: https://mutanome.lerner.ccf.org/. The p-value (P) was calculated by two-side ANOVA test. The data are represented as a boxplot with a underlaid violin plot where the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent the interquartile range (IQR) ×1.5 and beyond the whiskers are outlier points. Foretinib (n=6 mutated cell lines; n=969 wild type cell lines); Doxorubicin (n=4 mutated cell lines; n=927 wild type cell lines); Gemcitabine (n=3 mutated cell lines; n=923 wild type cell lines); Tamoxifen [GNAI2-TRIP6] (n=4 mutated cell lines; n=962 wild type cell lines); Rapamycin (n=3 mutated cell lines; n=386 wild type cell lines); Saracatinib (n=3 mutated cell lines; n=433 wild type cell lines); Dasatinib (n=4 mutated cell lines; n=419 wild type cell lines); Tamoxifen [TCF12-NAGK] (n=3 mutated cell lines; n=963 wild type cell lines).



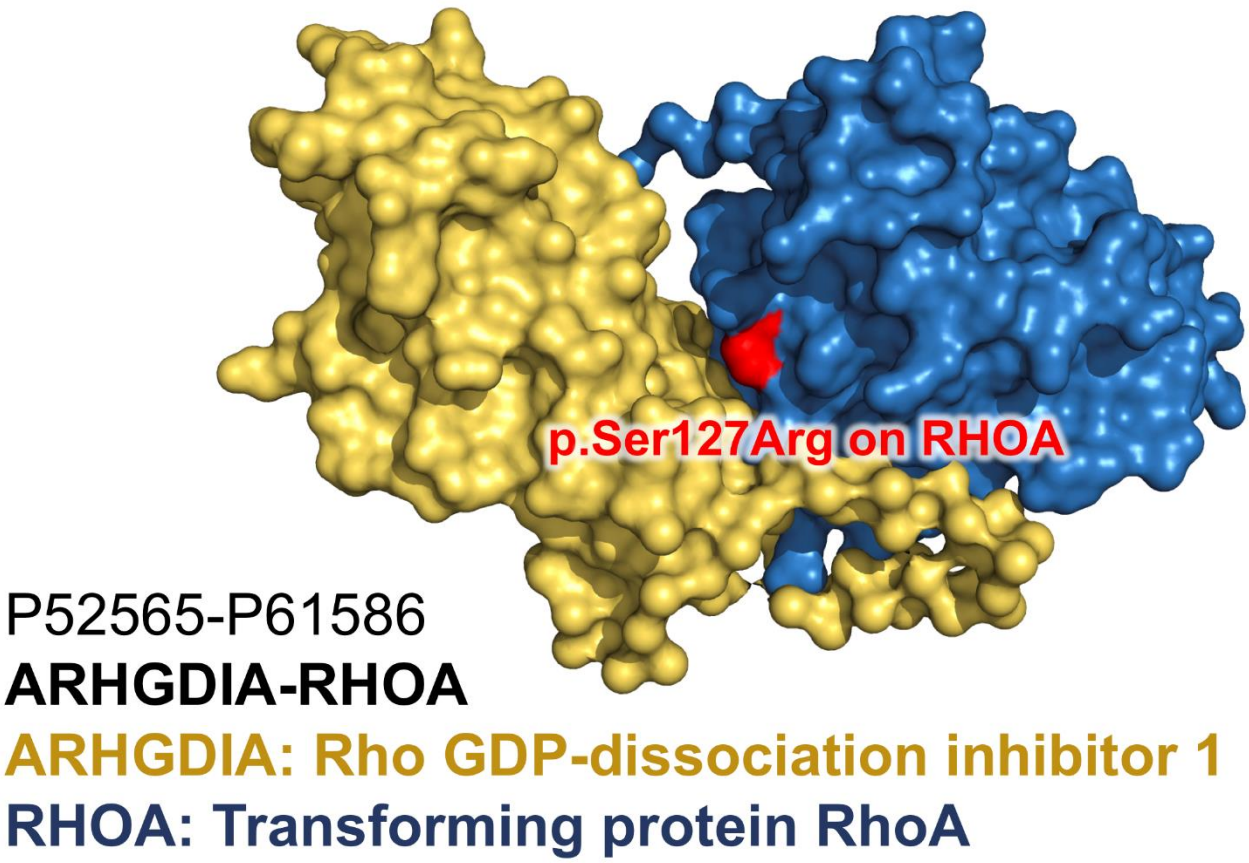
Supplementary **Figure 14.** Survival analyses of p53-SRSF1 perturbing-mutations across 33 cancer types from the TCGA. Time: Days. The p-value (P) was computed by Log-rank test.



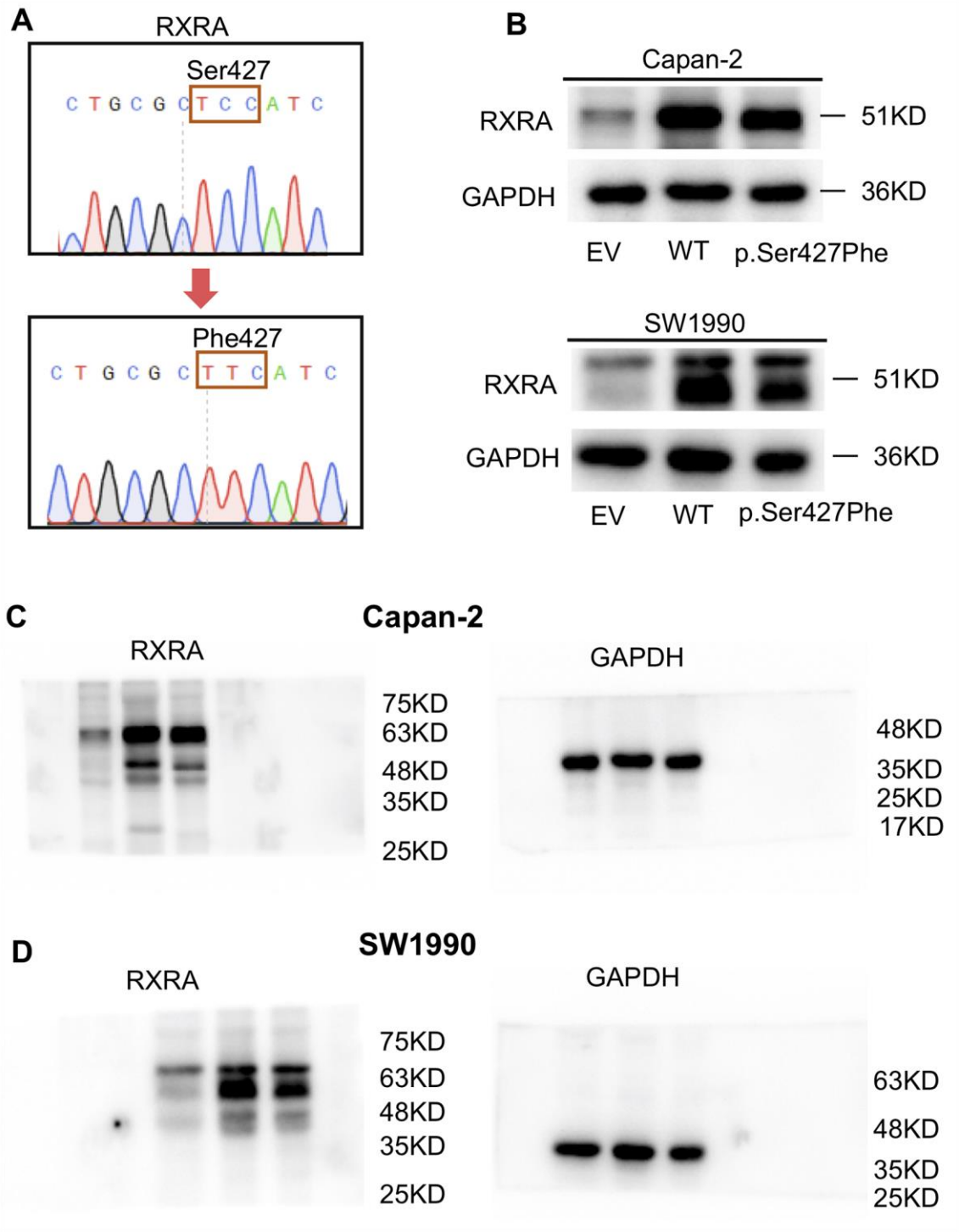
Supplementary **Figure 15.** Two selected examples of patient survival analyses predicted by mutation-perturbed PPIs (oncoPPIs) that harbor a statistically significant excess number of missense mutations at PPI interfaces as predicted by a binomial distribution. WT: wild-type and Mut: Mutation. The p-value (P) was computed by Log-rank test. All oncoPPI-predicted survival analyses for 33 cancer types are freely available at the website: https://mutanome.lerner.ccf.org/.



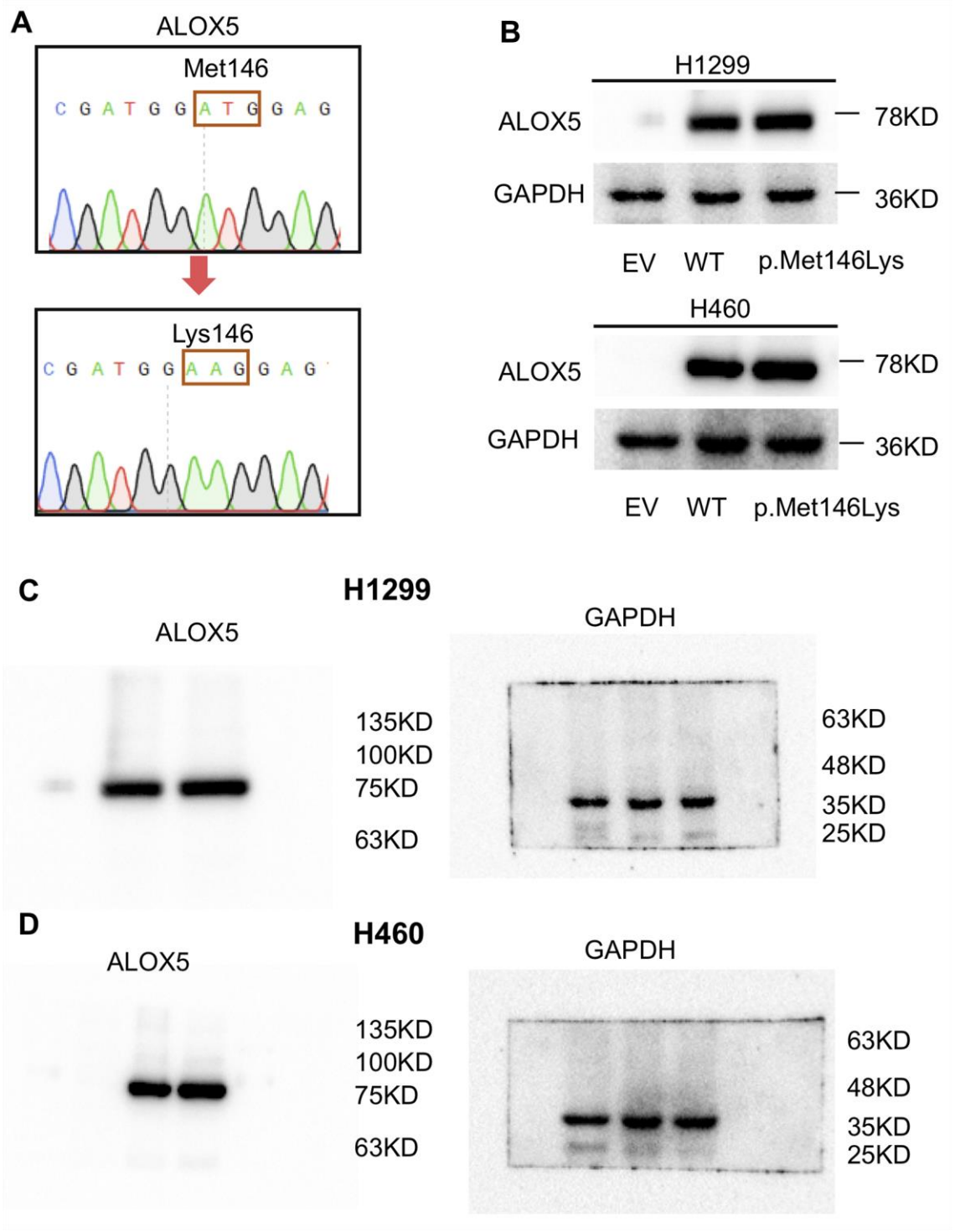
Supplementary **Figure 16.** HOMEZ-EBF1 complex model and the location of the interface mutation, p.Arg382Trp on HOMEZ. The complex model was built by Zdock protein docking analysis (Pierce *et al., Bioinformatics* 2014, **30**, 1771-1773). The images were built by PyMOL (https://pymol.org/2/)



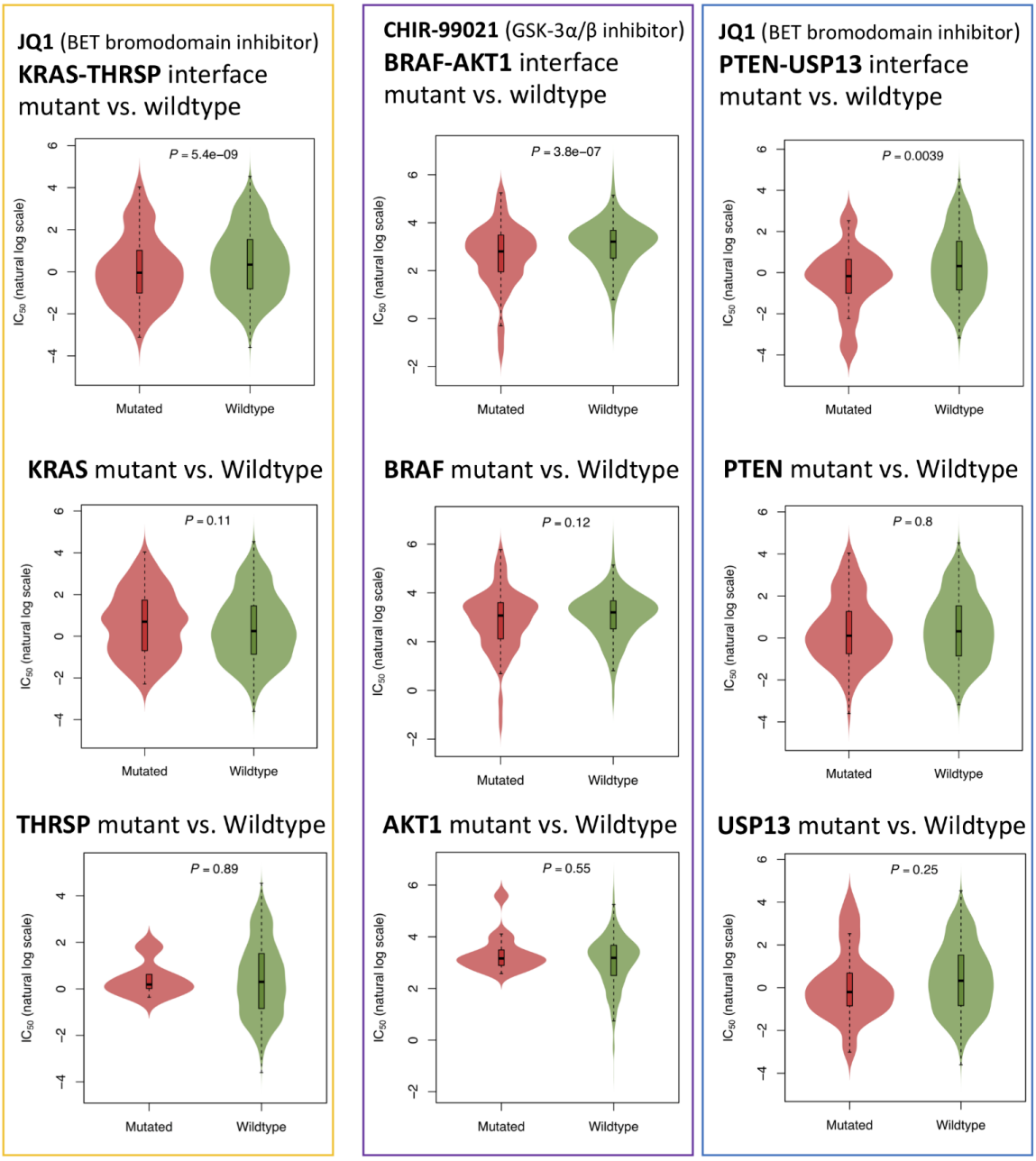
Supplementary **Figure 17.** RHOA-ARHGDIA complex (PDB id: 1CC0) and the location of the interface mutation, p.Pro75Ser on RHOA. The three-dimensional model was built by PyMOL (https://pymol.org/2/).



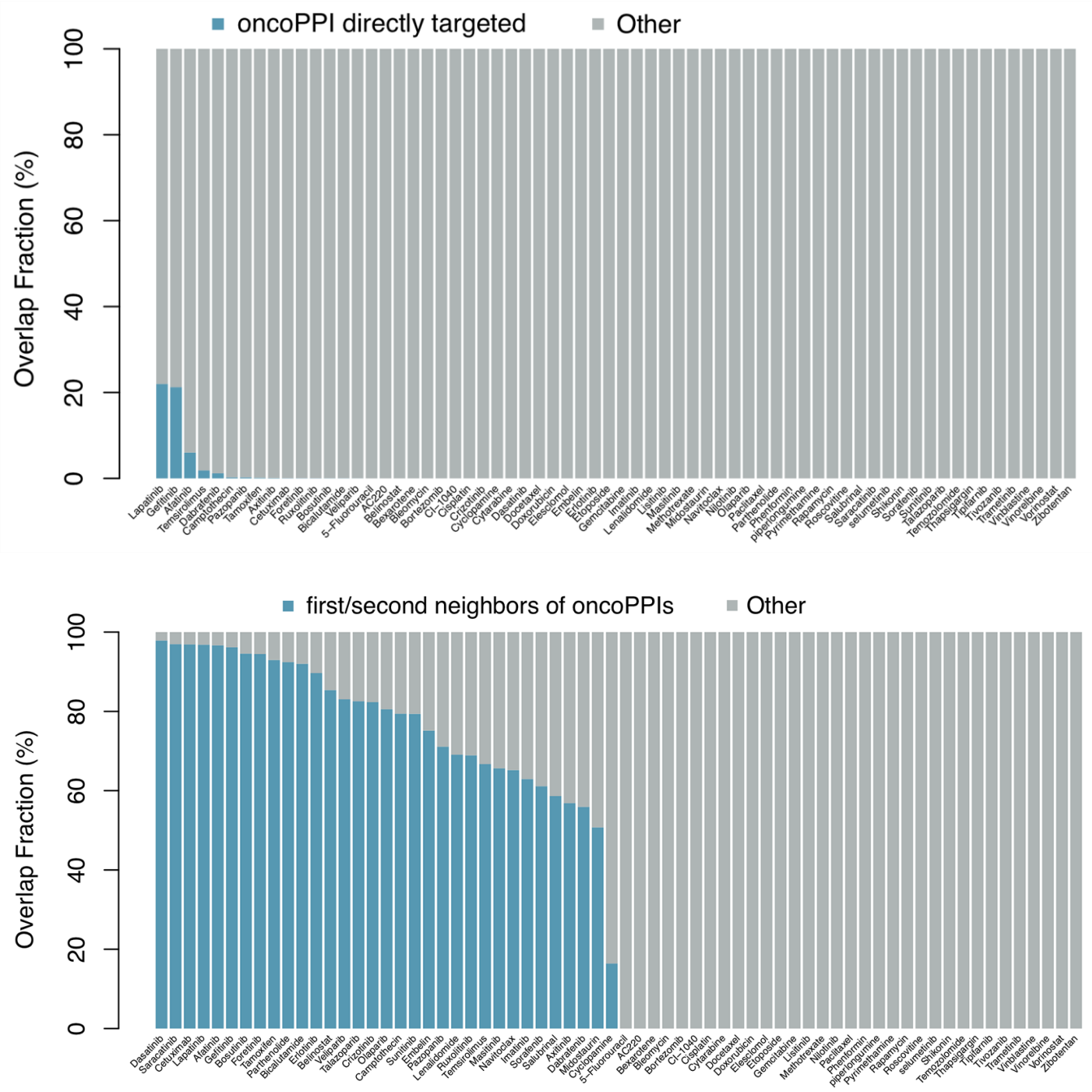
Supplementary **Figure 18.** Transfection of the wild-type (WT) and p.Ser427Phe mutant RXRA in two pancreatic cancer cells. (**A**) Schematic sequence of RXRA WT and mutant p.Ser427Phe. (**B**) Capan-2 and SW1990 cells were transfected with pCDNA3 empty vector (EV), pCDNA3-RXRA WT or pCDNA3-RXRA p.Ser427Phe for 48 hrs and RXRA expression was detected by Western blotting. (**C** and **D**) Uncropped images for Western blots in B.



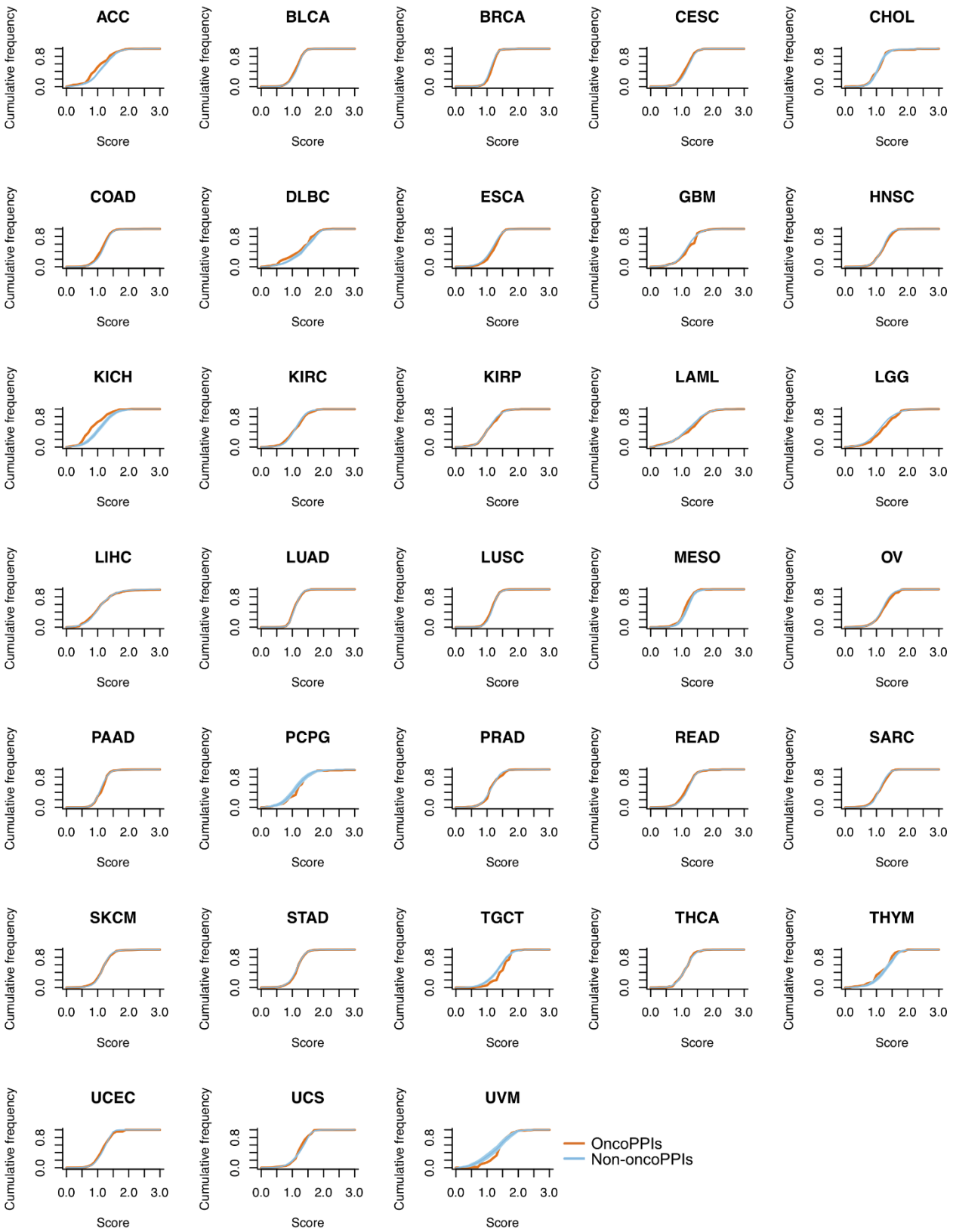
Supplementary **Figure 19.** Transfection of the wild-type (WT) and p.Met146Lys mutant ALOX5 in two non-small cell lung cancer cell line. (**A**) Schematic sequence of ALOX5 WT and mutant ALOX5 p.Met146Lys. (**B**) Western blot probing for ALOX5 protein in H1299 and H460 cells transfected with pCDNA3 empty vector (EV), pCDNA3-ALOX5 WT or pCDNA3-ALOX5 p.Met146Lys for 48 hrs. (**C** and **D**) Uncropped images for Western blots in **B**.



Supplementary **Figure 20.** Correlation of drug responses predicted by protein-protein interaction (PPI)-perturbing mutations and mutations in genes alone. PPI perturbing mutations (three selected oncoPPIs [KRAS-THRSP, BRAF-AKT1, and PTEN-USP13) are significantly associated with drug responses quantified by IC50, while mutations in genes alone failed to predict drug responses for all three selected oncoPPIs. The p-value (P) was calculated by two-side ANOVA test. The data are represented as a boxplot with a underlaid violin plot where the middle line is the median, the lower and upper edges of the box are the first and third quartiles, the whiskers represent the interquartile range (IQR) ×1.5 and beyond the whiskers are outlier points.



Supplementary **Figure 21.** Overlap analysis between drug targets and oncoPPIs. Targets for drugs listed in x-axis were collected from the GDSC database (www.cancerrxgene.org). We demonstrated that perturbing mutations on oncoPPIs are highly correlated with drug responses (Figure 4). However, we found very little overlap or no overlap between drug targets and proteins involving in oncoPPIs. Yet. we found significant overlap between drug targets and neighborhoods of proteins involving in oncoPPIs.



Supplementary **Figure 22.** Cancer type-specific expression of oncoPPI across 33 cancer types/subtypes. We applied the z-score measurement published in Sonawane et al. *Cell reports* 2017 on the expression data of 33 cancer types to calculate a tissue specific score for oncoPPI-cancer pairs.