

# A Proteomic Connectivity Map for Characterizing the Tumor Adaptive Response to Small Molecule Chemical Perturbagens

Zhenzhen Zi<sup>1</sup>, Yajie Zhang<sup>1</sup>, Peng Zhang<sup>3,4</sup>, Qing Ding<sup>1</sup>, Michael Chu<sup>1</sup>, Yiwen Chen<sup>3</sup>, John D. Minna<sup>2</sup>, and Yonghao Yu<sup>1,5,\*</sup>

1. Department of Biochemistry, UT Southwestern Medical Center, Dallas, TX 75390, USA

2. Hamon Center for Therapeutic Oncology Research, UT Southwestern Medical Center, Dallas, TX 75390, USA

3. Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

4. Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

5. Lead Contact

\* Correspondence: [Yonghao.Yu@UTsouthwestern.edu](mailto:Yonghao.Yu@UTsouthwestern.edu)

## Contents

1. SI Methods
2. SI Table S1–S2
3. SI Figures S1–S5
4. SI Reference

## **Experimental Procedures**

### **Cell culture and IC<sub>50</sub> determinations**

HBEC30KT cells <sup>1</sup> were cultured in Keratinocyte-SFM (Gibco) medium, supplemented with Bovine Pituitary Extract and Epidermal Growth Factor 1–53 (Gibco). HCC4017 cells were cultured in RPMI-1640 (Sigma) supplemented with 10% FBS (Sigma). HEK293T and HEK293TD cells were cultured in DMEM (Sigma) supplemented with 10% FBS (Sigma). All the cells were cultured at 37 °C in 5% CO<sub>2</sub>. The IC<sub>50</sub> values were measured using the CellTiter-Glo Luminescent Cell Viability Assay Kit per the manufacturer's instructions.

We performed time course studies to establish 48 hours as a meaningful time point to fully capture the proteomic response to the drug treatment. Although this represents one single time point, it was chosen to achieve the broadest utility for the initial dataset. More focused studies could be performed to further explore the remodeling of the proteome at specific, alternative time scales. Except for Rapamycin (used at 20 nM), all other compounds were used at 1 μM. These concentrations have been shown in the literature to achieve satisfactory level of target inhibition in cellular studies.

### **Sample preparation and mass spectrometric analyses**

HCC4017 and HBEC30KT cells were treated with the compounds for 48 hours. The Rapamycin was used at 20 nM, and all other compounds were used at 1 μM. All compounds were purchased from Selleck. Cells were lysed with 1% SDS lysis buffer (1% SDS, 10 mM HEPES, pH 7.0, 2 mM MgCl<sub>2</sub>, 20 U ml<sup>-1</sup> universal nuclease). Protein concentrations were determined using the BCA assay (Thermo Fisher). Samples were reduced with 1 mM DTT for 10 min and alkylated with 50 mM iodoacetamide for 30 min at room temperature. The detergent was removed by methanol/chloroform precipitation. Proteins were digested using Lys-C for 2 h in 8 M urea (50 mM Tris, 10 mM EDTA, pH 7.5), then digested by trypsin overnight in 2 M urea (diluted with 100 mM ammonium carbonate). The peptides were desalted using Oasis HLB

solid-phase extraction cartridges, and then labeled with TMT-6plex reagents (Thermo Fisher) for 1 h at room temperature. The reactions were quenched with 5% hydroxylamine.

The labeled samples were lyophilized and reconstituted in buffer A (10 mM Ammonium formate, pH10.0), then centrifuged at 10,000 rpm for 3 min in spin-X centrifuge tube filters (Thermo-Fisher) prior to loading on a 300 Extend-C18 HPLC column (Agilent). Peptides were fractionated by HPLC at 0.2 mL min<sup>-1</sup> flow rate using a gradient (0min: 100% buffer A; 5min: 90% buffer A; 95min: 50% buffer A; 107min: 30% buffer A; 110min: 100% buffer B). Buffer B was made as 1% Ammonium formate, pH 10.0 and 90% Acetonitrile. The collected fractions were lyophilized, desalted, and analyzed by LC-MS/MS on an LTQ-Velos Pro Orbitrap Mass Spectrometer (Thermo Fisher).

MS/MS spectra were searched using the Sequest algorithm. The peptides were filtered to reach protein identification FDR of less than 1%. Protein TMT ratios were determined using the procedure described previously<sup>2</sup>. The signal-to-noise (SN) value of each protein was calculated by summing the SN across all identified peptides. Data were exported to Excel for further analysis.

Each set of TMT has a labeled DMSO control group. When we analyze the data, the result is normalized to Log<sub>2</sub>(Compound/DMSO) from the same set of TMT. The normalized data from each TMT set are combined, then the common proteins are extracted for further analysis. We also did duplicate experiments for confirmation of the data integrity.

For the comparison of protein abundance between HBEC30KT and HCC4017 cells. In the TMT sets, we included HBEC30KT cells that were treated (in parallel with HCC4017 cells) with either DMSO or compounds. For example, each TMT set contains one DMSO and Two compounds treatment (HBEC30KT DMSO; HBEC30KT Compound 1; HBEC30KT Compound 2; HCC4017 DMSO; HCC4017 Compound 1; HCC4017 Compound 2). The result is normalized to Log<sub>2</sub>. The proteome perturbation difference between HBEC30KT and HCC4017 cells of is reflected by Log<sub>2</sub>(HCC4017 DMSO/HBEC30KT DMSO) or Log<sub>2</sub>(Compound/DMSO) of the same set of TMT.

## Western blotting analysis and antibodies

Cells were lysed with the 1% SDS lysis buffer. Protein concentrations were measured using a BCA assay kit (Thermo Fisher). The same amount of protein per lane was loaded on an SDS-PAGE gel; after electrophoretic separation proteins were transferred to a 0.2- $\mu$ m NC membrane (GE Healthcare). The membrane was blotted with the primary antibodies overnight at 4 °C followed by incubation with the secondary antibody for 1 h at room temperature. Blots were developed using enhanced chemiluminescence and were exposed on autoradiograph film.

The following antibodies were used: anti-phospho-p70 S6 kinase (Thr389) (CST, catalog number 9234s), anti-phospho-S6 ribosomal protein (Ser240/244) (CST, 5364s), anti-phospho-p44/42 MAPK (Thr202/Tyr204) (CST, 4370s), anti-UBE2C (CST, 14234), anti-Hells (CST, 7998s), anti-MCM2 (CST, 3619s), anti-MCM7 (CST, 3735s), anti-SLC7A11 (CST, 12691s), anti-HSP90 (CST, 4877), anti-HA (CST, 3724), anti-GAPDH (Thermo-Fisher, AM4300), anti-C5orf22 (Sigma, R39433), anti-Flag (Sigma, F7425 and F1804), goat anti-rabbit IgG, HRP-linked (GE Healthcare, NA-9340), and goat anti-mouse IgG, HRP conjugate (Millipore, AP181P).

## Data analysis

Statistical analysis of MS data was performed using built-in functions of Microsoft Office Excel 365 Proplus and GraphPad Prism (version 8.01). Hierarchical clustering was performed using Cluster3.0 and visualized using Java Tree View. Gene ontology analysis was performed using DAVID (version 6.7). PCA analysis was performed using Multibase and visualized using 5dchart. Pearson correlation analyses were performed using Python version 2.7.11. Protein interaction network were performed in the String database and drawn with Cytoscape<sup>3</sup>.

## Co-immunoprecipitation

Cell pellets were collected and suspended in the BLB lysis buffer (10 mM KPO<sub>4</sub>, pH 7.6, 6 mM EDTA, pH 8.0, 10 mM MgCl<sub>2</sub>, 0.5 % NP-40, 0.1 % Brij-35, 0.1 % DOC, adjust pH to 7.4), containing EDTA-free protease inhibitor and phosphatase inhibitor (Roch). After incubation for 20 min at 4 °C, samples were centrifuged at 14,000g for 10 min at 4 °C. Supernatants were transferred to new tubes. Agarose beads were added, and samples were incubated for 4 h at

4 °C. The beads were washed three times with the BLB buffer, and proteins were eluted with 1 % SDS elution buffer. TAP-MS experiments were performed using the protocol as previously described <sup>4</sup>.

### **Plasmid cloning**

Clones in the pDONOR 221 vector were obtained from the Center for Human Growth and Development of UTSW. Open reading frames were subcloned into the pCDNA5-ZZvTEV-Flag vector (addgene), and the ZZ-TEV-ORF-Flag was transferred to the pLenti-6.3-V5-Dest vector (Thermo Fisher) to construct stable cell lines. pCDNA3 or pRK5 vectors (addgene) were used for transient transfections. shRNAs targeting *C5orf22* were purchased from Sigma (TRCN0000143468 and TRCN0000121615).

### **Construction of stable cell lines**

HEK293TD cells were cultured to 50–70% confluency in 10cm dishes, and then transfected with Lipo-2000 (Sigma). The Plenti or pLKO.1 construct, VSVG, and delta8.9 were used at 8 µg, 6 µg, and 4 µg, respectively. The medium was changed 6 h after transfection. Viruses were collected twice at 24 h and 48 h after transfection. Subsequently, 1 mL of virus was added to each well of HCC4017 cells in 6-well plates with Polybrene at a final concentration of 8 µg ml<sup>-1</sup>. After splitting the cells once, cells were infected with virus again using the same procedure. The medium was replaced after 48 h with a fresh growth medium containing 2 µg ml<sup>-1</sup> blasticidin or puromycin.

### **RNA Extraction and real-time RT-PCR**

HCC4017 cells were lysed with TRIzol (Thermo Fisher Scientific) after paclitaxel treatment for 48 h or after treatment with shRNAs. Total RNA was extracted according to the manufacturer's instructions. The RNA was reverse-transcribed with SuperScript™ III One-Step RT-PCR System (Thermo Fisher). Real-time RT-PCR experiments were performed using a Power SYBR™ Green PCR Master Mix (Thermo Fisher). Raw data were normalized to the internal control (*Actin*) and presented as relative expression levels. All primers are listed in Table S2.

Table S1. Compounds and their IC<sub>50</sub> values against HCC4017 cells

No.	Inhibitor	Target	IC <sub>50</sub> ( $\mu$ M)	Log <sub>10</sub> (IC <sub>50</sub> )
1*	NVP-BEZ235	PI3K, mTOR	0.03	-1.49
2*	GDC-0941	PI3K	0.19	-0.72
3*	MK2206	AKT	29.60	1.47
4*	GSK2334470	PDK1	3.20	0.50
5*	Rapamycin	mTORC1	0.002	-2.70
6*	KU-0063794	mTORC1/2	0.25	-0.61
7*	CHIR-99021	GSK	8.47	0.93
8*	Sorafenib	BRAF	57201	4.76
9*	PLX4720	BRAFV600E	0.16	-0.81
10*	Trametinib	MEK	0.02	-1.62
11	Erlotinib	EGFR	37.30	1.57
12	Gefitinib	EGFR	61344	4.79
13	Lapatinib	EGFR/ErbB2	23017	4.36
14	NVP-TAE684	ALK	0.08	-1.09
15	R428	AXL	74243	4.87
16	BMS536924	IGF1R	16.40	1.21
17	BMS754807	IGF1R	0.72	-0.14
18	NVP-AEW541	IGF1R	1.09	0.04
19	GSK-1838705A	IGF1R	4.05	0.61
20	OSI-906	IGF1R	3.16	0.50
21	Vatalanib	VEGFR	0.53	-0.27
22	Tandutinib	PDGFR	14465	4.16
23	Su6668	PDGFR	1.48	0.17

24	Sunitinib	PDGFR	34464	4.54
25	Imatinib	PDGFR	260.90	2.42
26	Paclitaxel	Microtubules	0.005	-2.34
27	Pemetrexed	Purine and pyrimidine synthesis	0.18	-0.76
28	Carbo plantin	DNA repair	0.35	-0.46
29	Gemcitabine	Nucleoside analog	1.63	0.21
30	Etoposide	Topoisomerase	38	1.58
31	Ruxolitinib	JAK1/2	2996	3.48
32	PF-573228	FAK	6.50	0.81
33	Sch727965	CDK	0.003	-2.56
34	VX680	Aurora	1.53	0.18
35	Dasatinib	Abl/Src	0.03	-1.49

All the compounds were used to treat HCC4017 cells. Compounds labeled with an asterisk were also used to treat HBEC30KT cells.

Table S2 Primers for Q-RT-PCR experiments

	Forward Primer	Reverse Primer
<i>C5orf22</i>	TGGAGGATCATCAGGAGGTTCT	GGAGGTCTGGATGTGAGTCG
<i>IL-1A</i>	GACGCACTTGTAGCCACGTA	GGCCATCTTGACTTCTTTGCT
<i>IL-1B</i>	TCGCCAGTGAAATGATGGCT	TGGAAGGAGCACTTCATCTGTT
<i>BRCA1</i>	TGTCTCCACAAAGTGTGACCA	TGTTTGCATACTCCAAACCTGTG
<i>Actin</i>	AAATCGTGCGTGACATTAAGGA	GCCATCTCTTGCTCGAAGTC



Supplemental Figures

Figure S1.

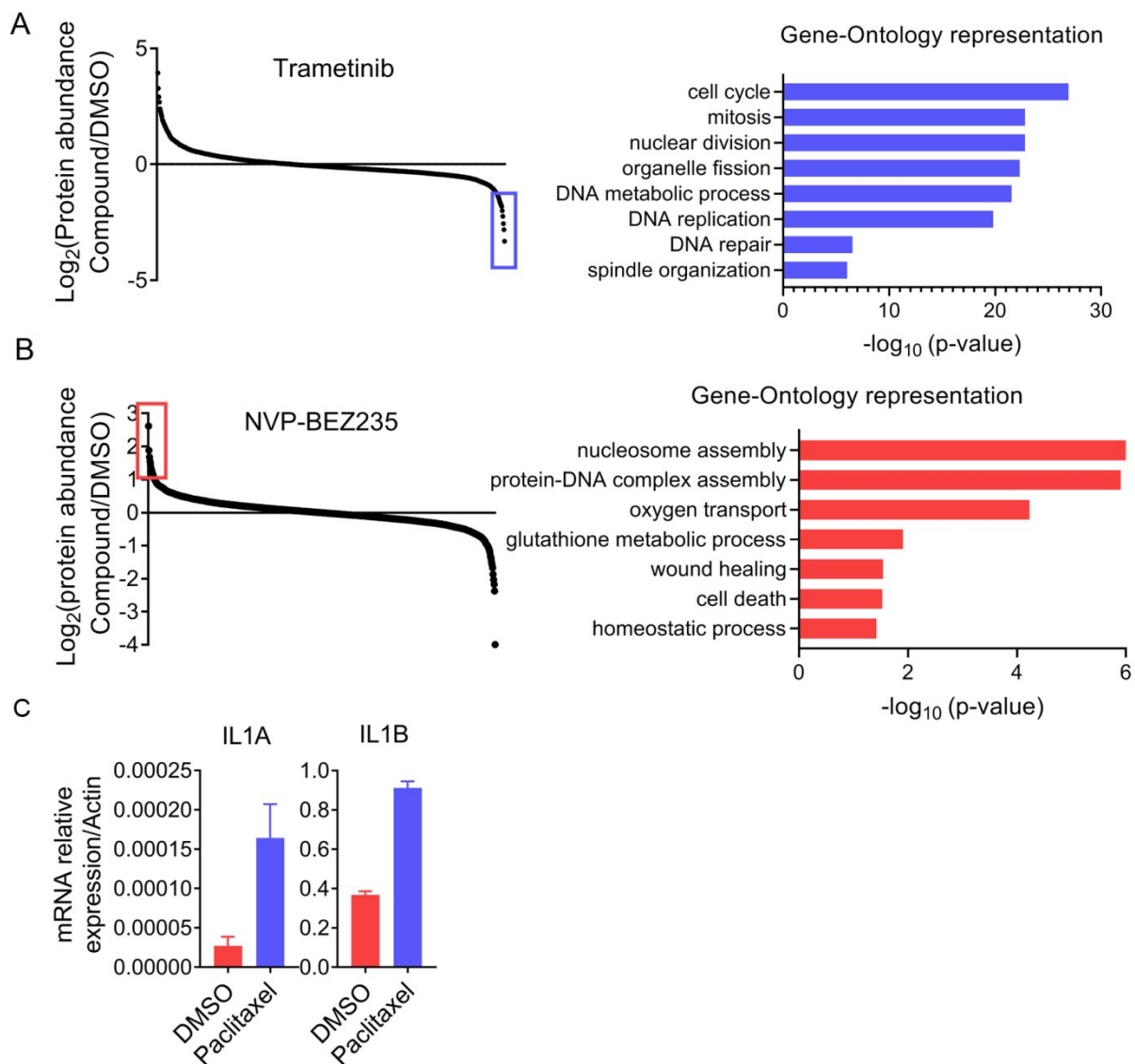


Figure S1. Effects of NVP-BEZ235 and Trametinib on the proteome of HCC4017 cells.

(A) Log<sub>2</sub> plot analysis of protein abundances in HCC4017 cells treated with Trametinib. DMSO treated cells were used as the control. The blue box highlights the proteins that were down-regulated by more than 2-fold after Trametinib treatment. These proteins were then subject to GO analysis.

(B) Log<sub>2</sub> plot analysis of protein abundances in HCC4017 cells treated with NVP-BEZ235. DMSO-treated cells were used as the control. The red box highlights the proteins that were up-regulated by more than 2-fold after NVP-BEZ235 treatment. These proteins were then subject to GO analysis.

(C) *IL1A* and *IL1B* expression as measured by real-time RT-PCR analyses Data are shown as mean ± SEM (n=3).

**Figure S2.**

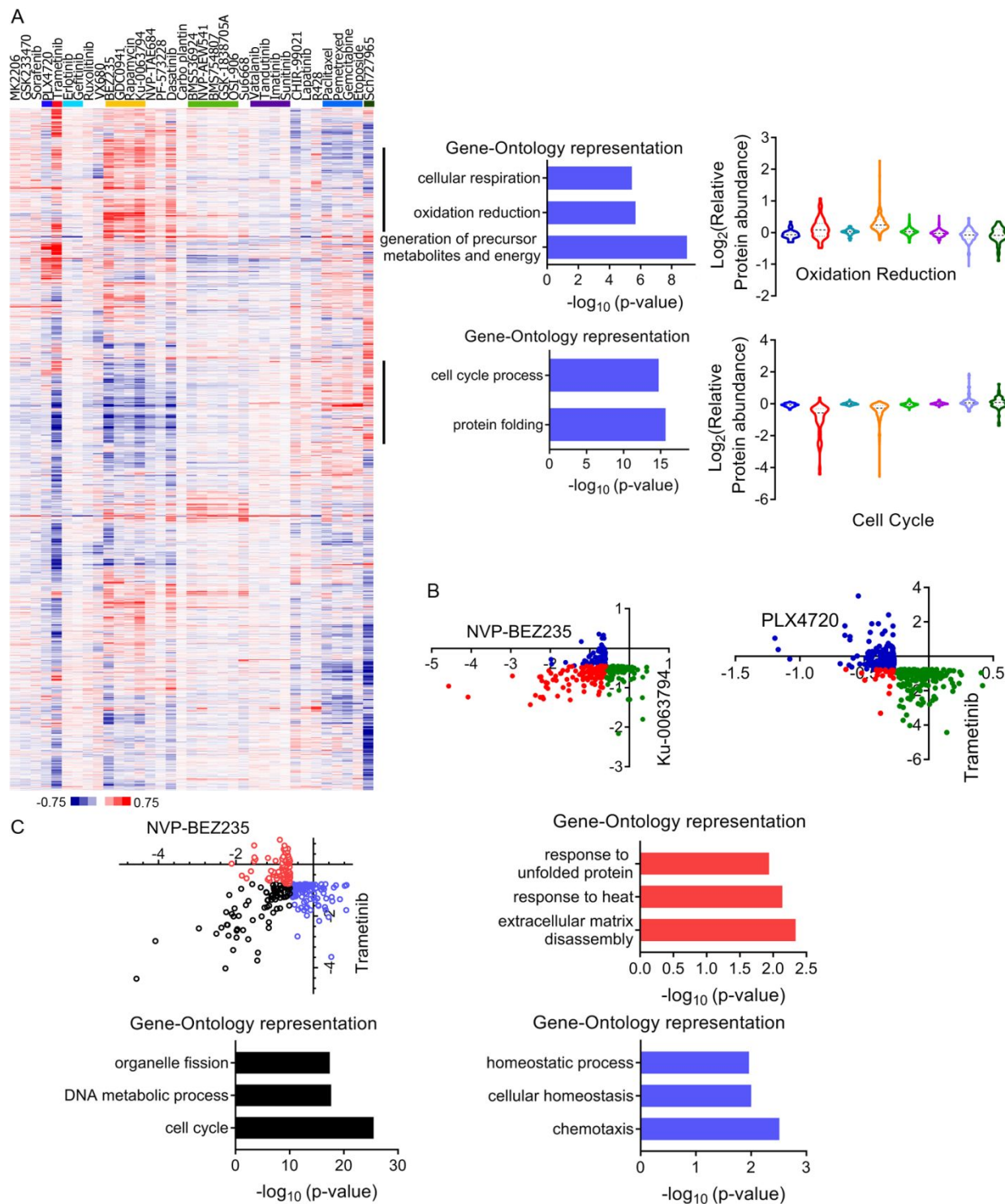


Figure S2. **Hierarchical clustering and PCA analysis of drug-correlation profiling**

(A) Hierarchical clustering of protein expression profiles of HCC4017 cells upon the treatment of the 35 compounds (as shown in Fig. 3A). Two representative protein groups were subject to GO analysis, with

the top enriched biological processes indicated. Distribution of protein abundances in each drug cluster for proteins involved in oxidation-reduction and cell cycle regulation.

(B) Cross reference analysis of the top 200 downregulated proteins between the indicated pairs of inhibitors. Proteins that were down-regulated in HCC4017 cells upon the treatment with both inhibitors are represented by red dots. Proteins that were down-regulated only by the inhibitor on the X and Y axis are represented by blue and green dots, respectively. Results are shown in a  $\text{Log}_2$  scale.

(C) Cross-reference analysis of the top 200 downregulated proteins in HCC4017 cells treated with NVP-BEZ235 and Trametinib. Proteins that were downregulated by both compounds (black circles) are subject to GO analysis. GO analyses are also performed on proteins that were down-regulated by NVP-BEZ235 (red dots) and Trametinib (blue dots).

**Figure S3.**

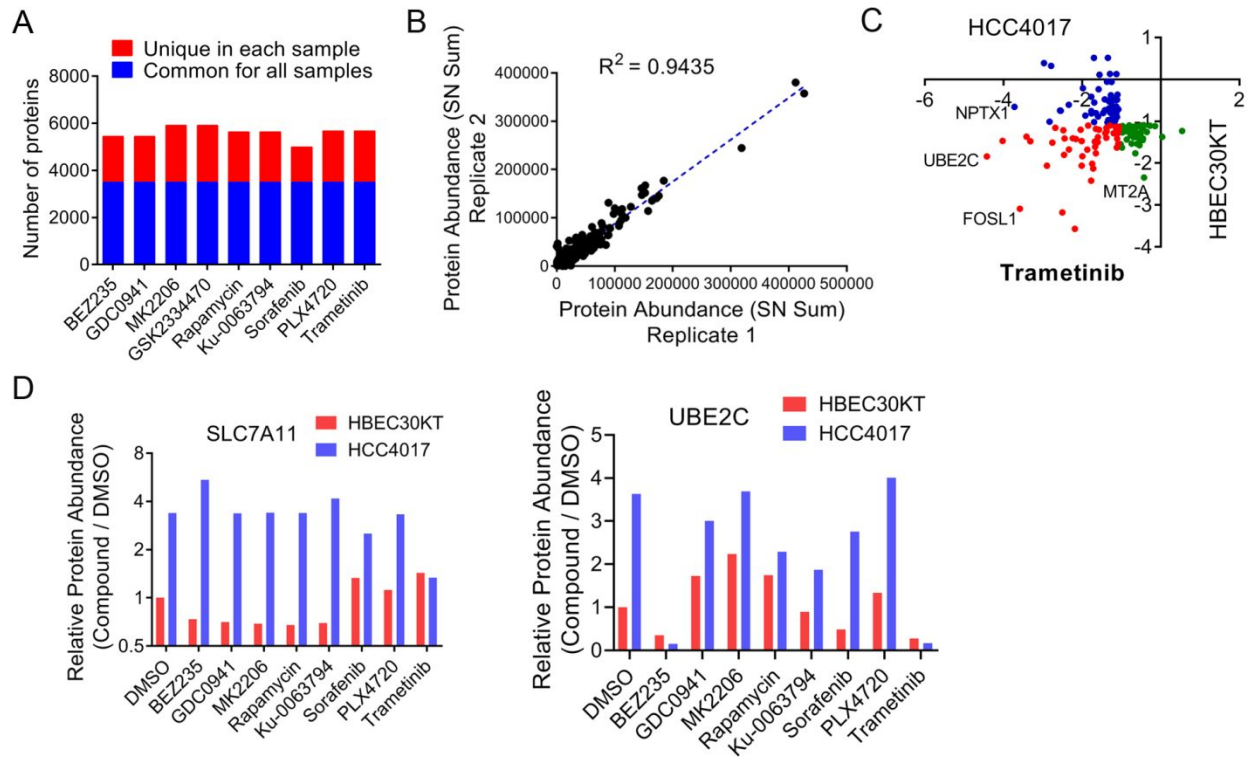


Figure S3. **Context-selectivity of the proteomic response to drug treatment**

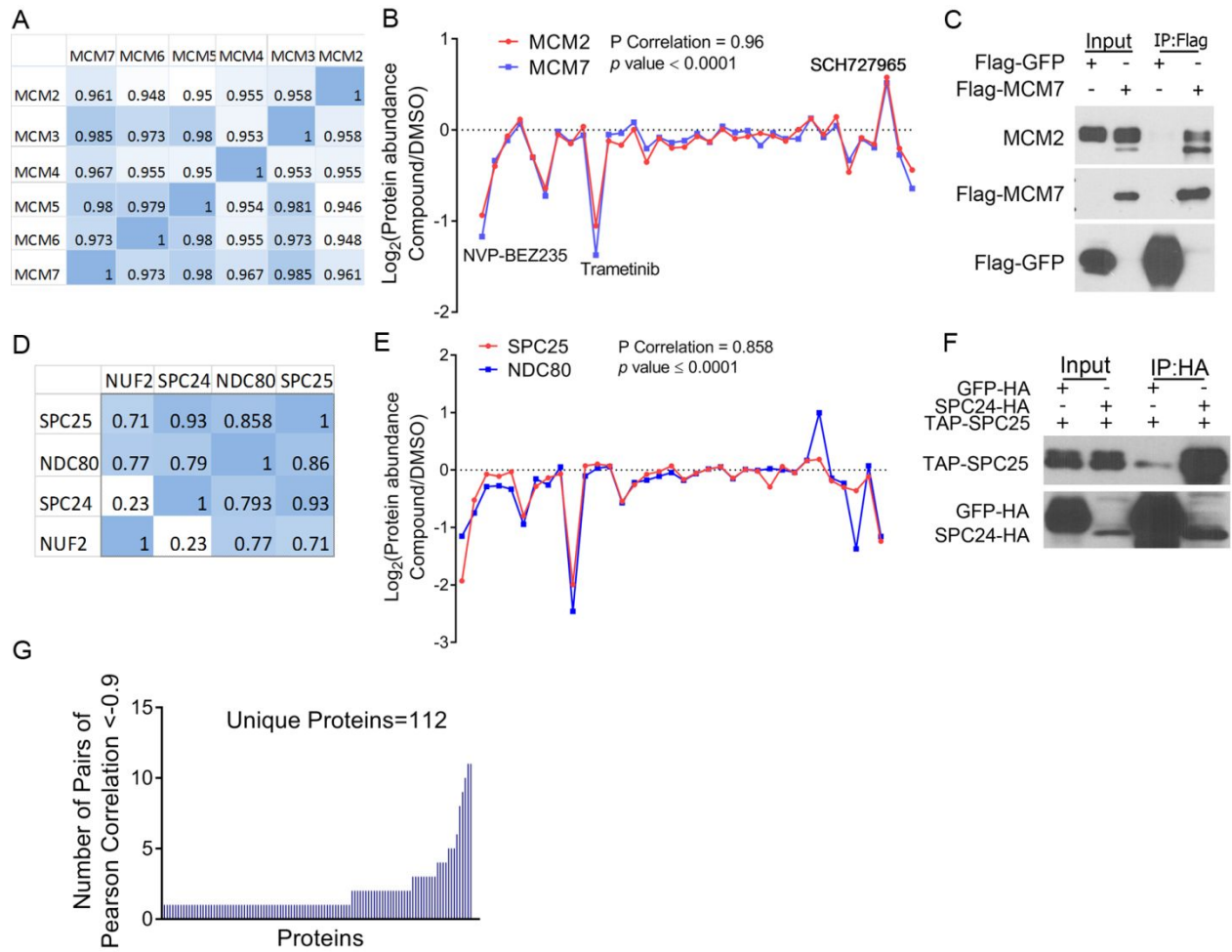
(A) Number of proteins identified in all TMT sets (blue bars) and number of unique proteins identified in each TMT set (red bars) for the HBEC30KT compound treatment experiments.

(B) Quantitative reproducibility of the TMT experiments. The TMT reporter ion intensities (signal to noise, SN) from all the peptides of a protein were summed (from HBEC30KT treated with DMSO) and plotted for the two biological replicate experiments.

(C) Cross-reference analysis of the top 100 down-regulated proteins in HCC4017 and HBEC30KT upon Trametinib treatment. Proteins that were down-regulated in both cell lines are represented by red dots. Proteins that are selectively down-regulated in HCC4017 cells and HBEC30KT cells are represented by blue and green dots, respectively. Results are shown as  $\text{Log}_2$ -transformed values ( $\text{Log}_2(\text{Compound}/\text{DMSO})$ ).

(D) SLC7A11 and UBE2C protein abundances as measured by the TMT experiments. Protein abundances were normalized to that in HBEC30KT cells treated with DMSO.

Figure S4.



**Figure S4. Identification of protein covariance networks.**

(A) Pearson correlation values for each pair of the MCM complex members.

(B)  $\text{Log}_2(\text{Compound}/\text{DMSO})$  values for MCM2 (red) and MCM7 (blue) in HCC4017 cells treated with the compound panel.

(C) Validation of the interaction between MCM2 and MCM7. Western blotting analysis of the immunoprecipitates of HCC4017 cells that stably express TAP-MCM7. TAP-GFP was used as a negative control.

(D) Pearson correlation values for each pair of the NDC80 complex members.

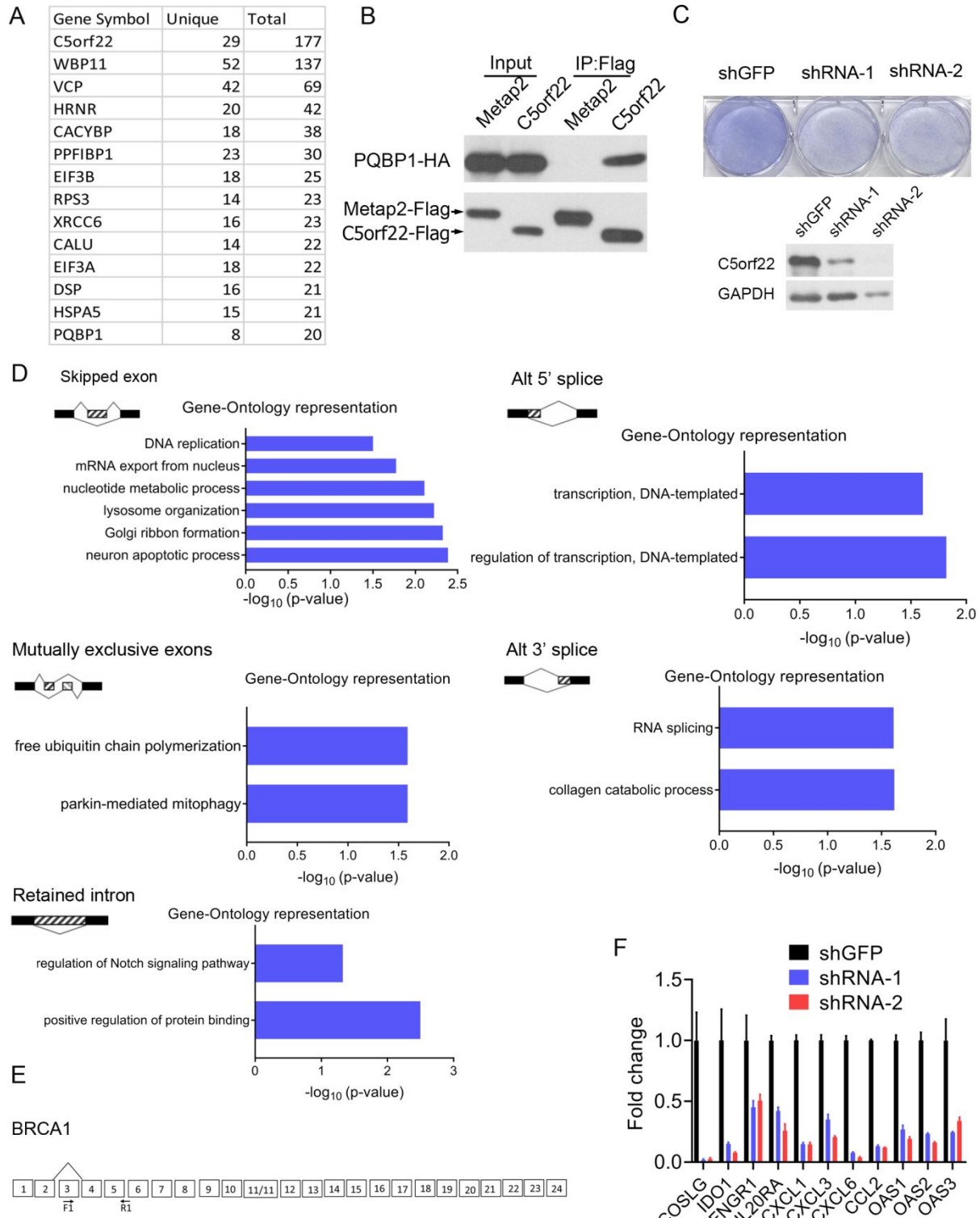
(E) Covariation patterns between Spc25 (red) and Ndc80 (blue) in drug-treated HCC4017 cells.

(F) Validation of the interaction between SPC25 and SPC24. Western blotting analysis of the immunoprecipitates of HEK293T cells that were co-transfected with TAP-SPC25 and SPC24-HA or TAP-SPC25 and GFP-HA. GFP-HA was used as a negative control.

(G) Pearson correlation analysis of the protein covariation patterns. Protein pairs with a Pearson correlation coefficient of less than  $-0.9$  are shown.



**Figure S5.**



**Figure S5. C5orf22 is a novel component of WBP11/PQBP1 spliceosome complex**

(A) Validation of the interaction among C5orf22, WBP11 and PQBP1. C5orf22 (with a ZZ-Flag Tag) was purified from cells, and the immunoprecipitates were trypsin-digested and were analyzed by LC-MS/MS experiments. The identified proteins and their spectral counts (total and unique peptides) are shown.

(B) Validation of the binding between C5orf22 and PCBP1 by co-immunoprecipitation experiments. Flag-Metap2 was used as the negative control.

(C) Crystal violet staining of Control and C5orf22 depletion HCC4017 cells. Efficiency of *C5orf22* depletion in HCC4017 cells, as shown by Western blotting experiment. shGFP was used as the negative control.

(D) GO analyses of the genes whose splicing is altered as a result of C5orf22 depletion.

(E) Alternative splicing event for *BRCA1* pre-mRNAs with primers used for real-time RT-PCR analyses shown.

(F) Real-time RT-PCR analyses of the immunomodulatory genes in *C5orf22*-depleted HCC4017 cells. shGFP HCC4017 cells were used as the control.

## References

- (1) Kim, H. S., Mendiratta, S., Kim, J., Pecot, C. V., Larsen, J. E., Zubovych, I., Seo, B. Y., Kim, J., Eskiocak, B., Chung, H., McMillan, E., Wu, S., De Brabander, J., Komurov, K., Toombs, J. E., Wei, S., Peyton, M., Williams, N., Gazdar, A. F., Posner, B. A., Brekken, R. A., Sood, A. K., Deberardinis, R. J., Roth, M. G., Minna, J. D., and White, M. A. (2013) Systematic identification of molecular subtype-selective vulnerabilities in non-small-cell lung cancer, *Cell* *155*, 552–566.
- (2) Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., Villen, J., Haas, W., Sowa, M. E., and Gygi, S. P. (2010) A tissue-specific atlas of mouse protein phosphorylation and expression, *Cell* *143*, 1174–1189.
- (3) 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.
- (4) Tsai, A., and Carstens, R. P. (2006) An optimized protocol for protein purification in cultured mammalian cells using a tandem affinity purification approach, *Nat Protoc* *1*, 2820–2827.