nature biotechnology

Article

https://doi.org/10.1038/s41587-022-01539-0

A proteome-wide atlas of drug mechanism of action

In the format provided by the authors and unedited

Supplemental Note #1: Initial Consideration for Library Profiling

Our overarching goals were to understand how modulators of different classes of proteins affect proteome remodeling, to draw conclusions about the activities of distinct chemical scaffolds, and to appreciate the effects of understudied compounds. Achieving these goals necessitated the use of a large screening library that covered a broad chemical space and included redundancies in the annotated targets.

The screening library ultimately selected for this work is comprised of 875 small molecules that, when all known targets are included, target 914 proteins that constitute approximately half of the targetable proteome[1] (**Fig. 1 and S1B**). The library includes compounds across all stages of clinical development, ranging from highly cited FDA approved drugs, such as indomethacin with more than 43,000 citations over 70 years, to largely undescribed tool compounds like UNC2881 with only one citation in PubMed. (**Fig. 1B and S1A**). Nearly 700 compounds share primary or secondary targets with other library members, allowing for interpretation of both target-induced proteome fingerprints and off-target polypharmacology (**Fig. S1C**).

This large library is comprised of compounds that are used at a broad range of effective concentrations, though many do not have annotated EC50 values for their target protein(s). Additionally, the "true" target of many drugs and tool compounds is often unknown[2], which complicates efforts in picking treatment concentrations. Moreover, the concentration required to engage enough protein in cells to elicit a functional response is not revealed by the *in vitro* IC50[3, 4]. For the fraction of library members that are cytotoxic, it is possible to treat cells at empirically measured concentrations that mirror proliferation IC50 values. However, this approach results in proteome fingerprints that are dominated by cell death pathways[5], limiting the interpretation of nuanced differences between proteome-measured MOAs.

Thus, we chose a single treatment concentration for the majority of the library, settling on the common screening concentration of 10 µM [6], that is still considered useful for MOA determination[7]. This concentration provided a middle ground for high potency compounds and low affinity ligands, while making a large library screen more tractable, in part, due to available automation. Moreover, by treating cells with concentrations that are sometimes higher than might be typical, we had the opportunity to discover new, biologically relevant targets that can be used for i) "off-label" tool compound usage, ii) drug repurposing efforts, and iii) leveraged polypharmacology for treating disease.

We balanced library size against screening across multiple cell lines, ultimately deciding to use HCT116 cells as a standalone model. This microsatellite instable colorectal cancer (CRC) cell line[8] expresses mutant KRAS (G13D), the third most common driver mutation across cancers[9, 10]. Moreover, CRC is the second leading cause of cancer related death,[11] highlighting the need to understand drug action in this cell model. Additionally, the protein-protein interaction landscape of HCT116 cells has been thoroughly defined through the BioPlex Network[12], providing a useful companion resource for interpreting protein-complex coregulation, while making the HCT116 proteome one of the most well studied. Although using a single genetic background limits the scope of protein targets, and thus signaling pathways, that can be perturbed by compounds in the library, our decision gives the ability to screen a library 10 times larger than ever reported, allowing us to incorporate many poorly characterized tool compounds. Moreover, we can now make more comparisons between compound-induced proteome fingerprints, which is essential for future MOA deconvolution efforts.

Supplemental Note #2: Characterization of Batch Effects and Normalization Methods

This supplementary note outlines how we analyzed and reduced batch effects across nearly 200 plexes. Cells were grown and treated in 22, 96-well plates (**Figure SN2-1**). Replicates were performed at the plate level. Cells grown in the first column in each plate were treated with DMSO. Cells in the rest of the columns were treated with compounds. Cells treated with toxic compounds were replaced with those treated with the control compound when needed. Samples in each row were analyzed in the same TMT plex. Samples in each column were labeled with the same TMT-11plex tag (**Figure SN1**). We first adjusted the TMT reporter ion intensities by making the total summed value of all the reporter ion intensities for all proteins in each sample the same across each plex. This is to account for protein loading variance, pipetting errors, *etc.* during the sample preparation. The rationale for this adjustment is that we intent to load equal amounts of proteins in each TMT channel. After this adjustment, $\log_2$ fold changes (compound vs DMSO) were obtained. We then analyzed the batch effects at plex, plate, channel, and replicate level using hierarchical clustering analysis (HCA) (**Figure SN2-2**), principal component analysis (PCA) (**Figure SN2-3**), and linear regression (**Figure SN2-4**).

Log_2 fold changes of proteins that were quantified in all TMT channels were used in the HCA and PCA (4,087 proteins, 1,628 TMT channels). As shown in the HCA results before median normalization (**Figure SN2-2**), samples were clustered by replicate batch, plex, and plate to some extent (highlighted by the red arrows), suggesting batch effects at these levels. The PCA results before median normalization also showed the batch effect at plex level (**Figure SN2-3**) as indicted by the red arrow. Previous studies in our lab have shown that per-protein batch effects exist in large-scale multiplexed datasets, where one protein is consistently up-or down-regulated in all samples in a TMT [13, 14]. In a large-scale multiplexed experiment where samples are randomly distributed across TMT channels, a protein being up- or down-regulated in all samples in a plex is unlikely and a per-protein mean or median centering normalization can be performed to remove such effect. In this work, the source of the per-protein batch effect can be random error in the measurement of the DMSO-treated samples, edge effect of the 96-well plate, *etc.* As compounds were randomly assigned in this work to evenly distribute the targets across all TMT plexes/channels, we expect that the median $\log_2$ fold change of a protein within each plex should be zero. We thus further normalize the data and investigated how this per-protein median normalization affected the batch effects. As shown in the HCA results after median normalization (Figure SN2), TMT channels were no longer grouped by replicate, plex, or plate and the per-protein median normalization reduced the small batch effects. The PCA results also showed that the plex level batch effect was also effectively removed by the per-protein median normalization (**Figure SN2-3**). In the linear regression analysis, the per-protein median normalization removed the plex and channel level batch effects, while it retained the primary target effect (**Figure SN2-4**). So before moving forward to perform further quantitative analysis, we did the per-protein median normalization and reported the normalized ratios in the viewer and the supplementary tables.

Thus, our final approach to normalization included 1) column normalization to equalize protein loading, 2) ratioing against the DMSO channel to all cross-plex analyses, and 3) median normalization of the log2-transformed ratios across each plex to correct for small differences in DMSO measurements.

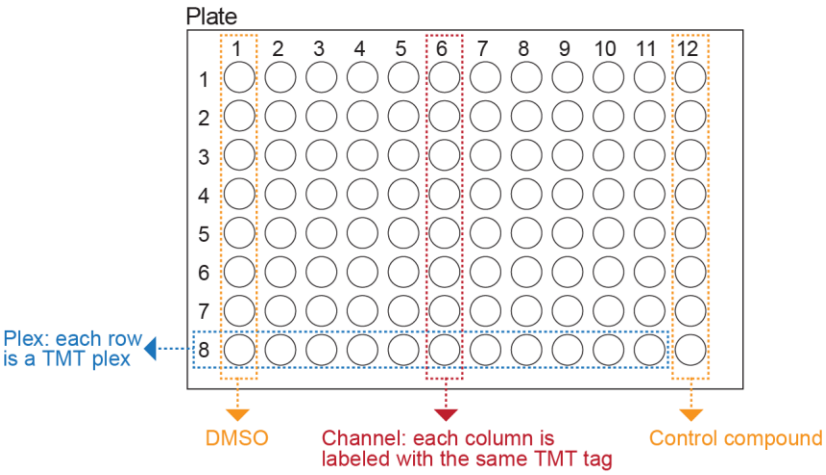


Figure SN2A. The plate layout. Cells in the first column in each 96-well plate were treated with DMSO. Cells in the last column in each plate were treated with the control compound. Cells in the rest of the columns were treated with compounds. Cells treated with toxic compounds were replaced with those treated with the control compound when

needed. Samples in each row were analyzed in the same TMT plex. Samples in each column were labeled with the same TMT-11plex tag. This figure explains the “Plate”, “Plex”, and “Channel” in the following Figure SN2, Figure SN3 and Figure SN4.

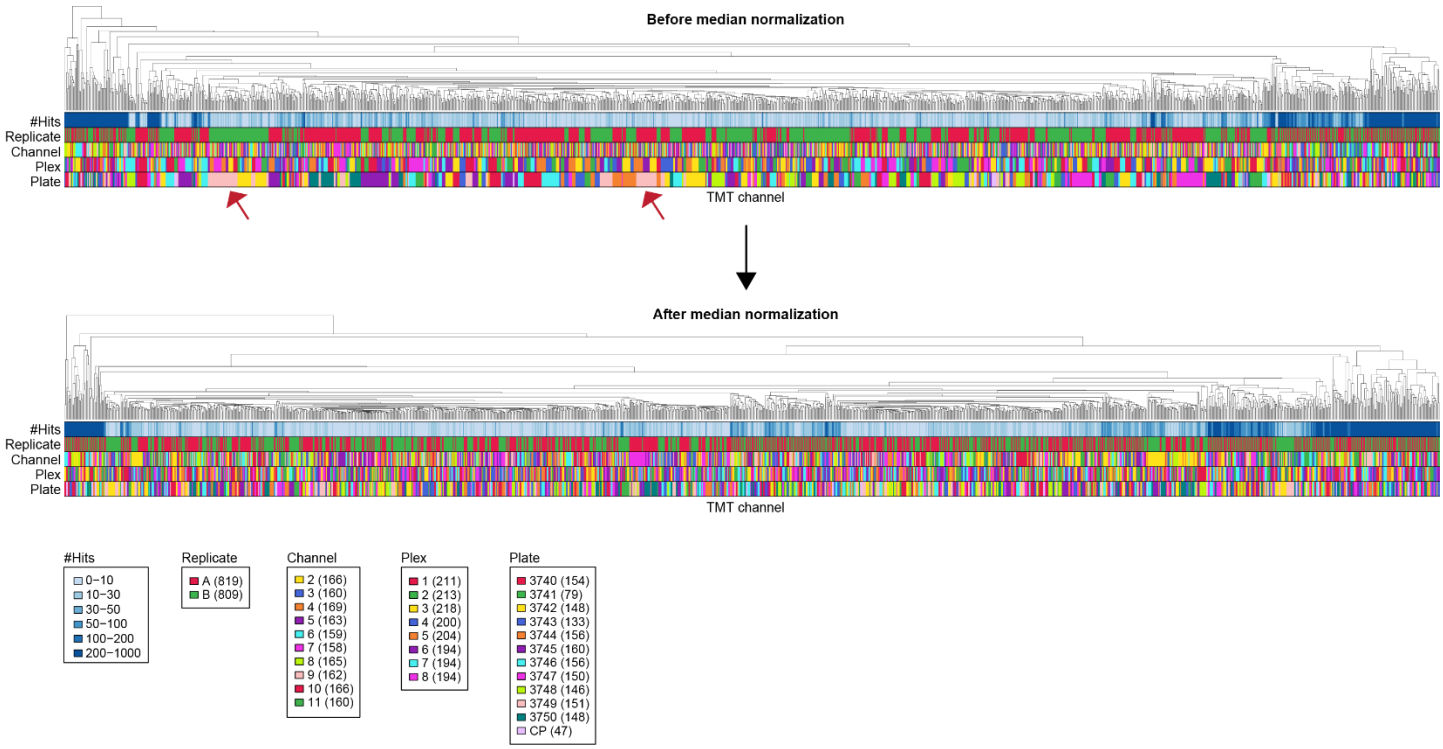


Figure SN2B. Hierarchical clustering analysis of the data before and after per-protein median normalization. As highlighted by the red arrows in the heatmap before median normalization, there are batch effects at replicate, plex and plate level. As shown in the heatmap map after median normalization, per-protein normalization reduced the batch effects. “#Hits” indicates the number of significantly regulation proteins in each TMT channel. “Replicate” indicates the replicate batch. “Channel”, “Plex”, and “Plate” are explained in Figure SN1. Numbers in the parentheses in the figure legend indicate the number of TMT channels per color. Log_2 fold changes (compound vs DMSO) were used to perform the hierarchical clustering analysis.

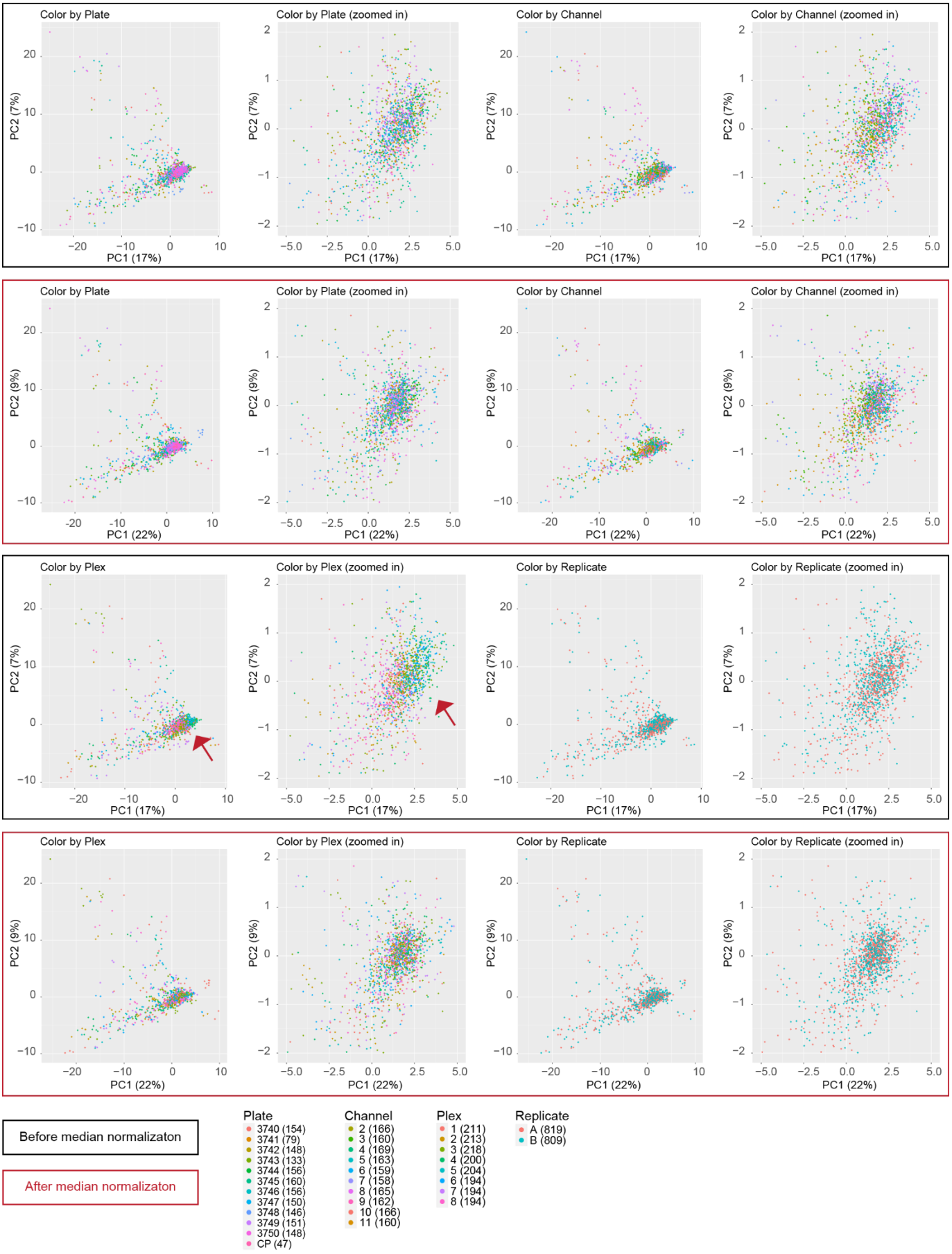


Figure SN2C. Principal component analysis of the data before and after per-protein median normalization. As highlighted by the red arrows, there is batch effect at plex level. The per-protein normalization reduced the batch effect. “Replicate”

indicates the replicate batch. “Channel”, “Plex”, and “Plate” are explained in Figure SN1. Numbers in the parentheses in the figure legend indicate the number of TMT channels per color. Log_2 fold changes (compound vs DMSO) were used to perform the principal component analysis.

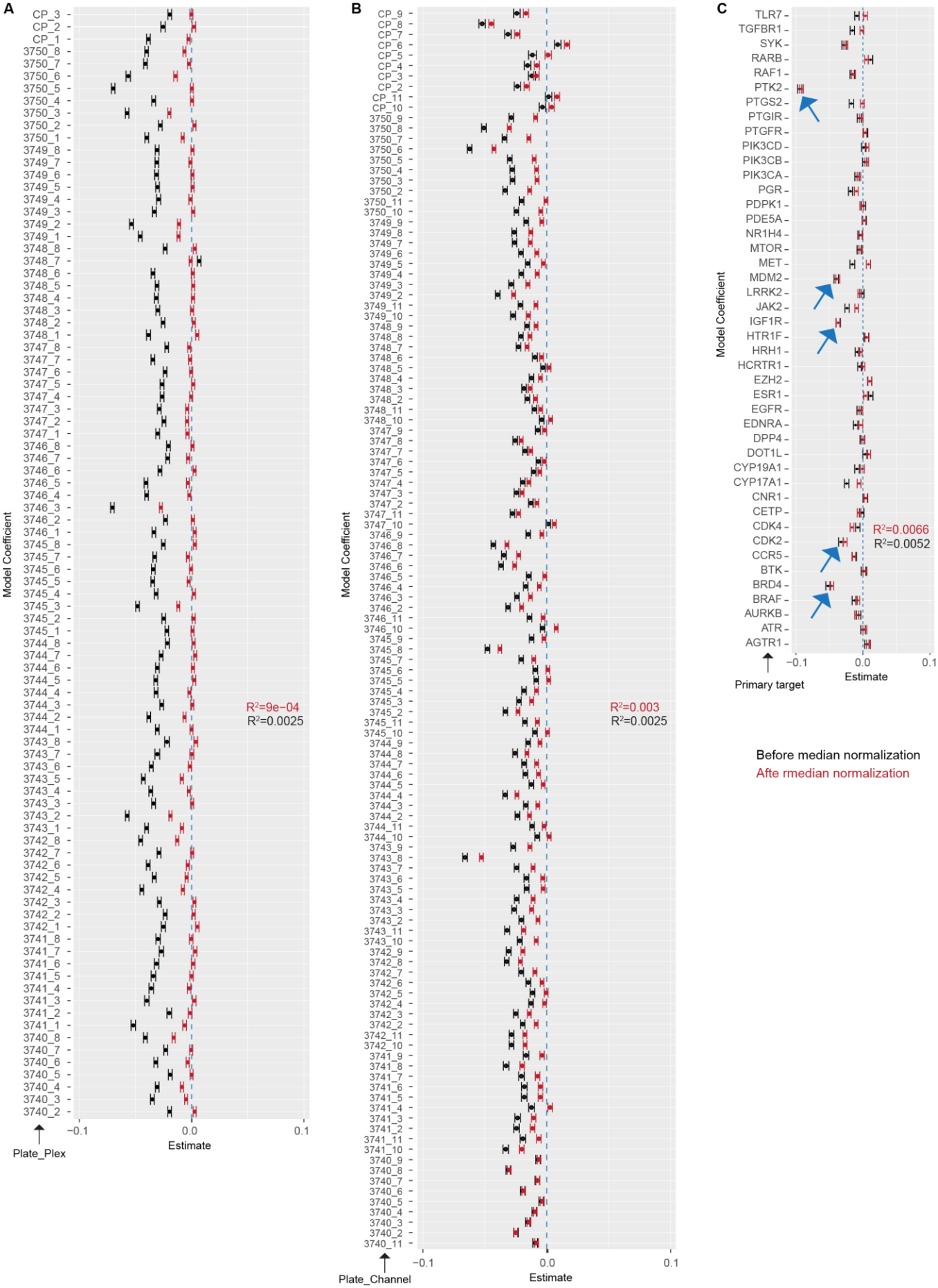

A
Model Coefficient
Plate_Plex
Estimate
−0.1
0.0
0.1
R²=9e−04
R²=0.0025
B
Model Coefficient
Plate_Channel
Estimate
−0.1
0.0
0.1
R²=0.003
R²=0.0025
C
Model Coefficient
Primary target
Estimate
−0.1
0.0
0.1
R²=0.0066
R²=0.0052
Before median normalization
Afte rmedian normalization

Figure SN2D. Linear model coefficients plots along with associated R^2 values for the models. Models were fit on the entire protein quantitative data ($\log_2$ fold changes, compound vs DMSO, n = ~10 million) using dummy variables for plex, channel, and primary target. The per-protein median normalization reduced batch effects at both plex (A) and channel level (B), especially the plex level. The primary target-based effect remained after the per-protein median normalization as highlighted by the blue arrows (C). “Channel”, “Plex”, and “Plate” are explained in Figure SN1. In all plots, data points represent the 95% confidence interval.

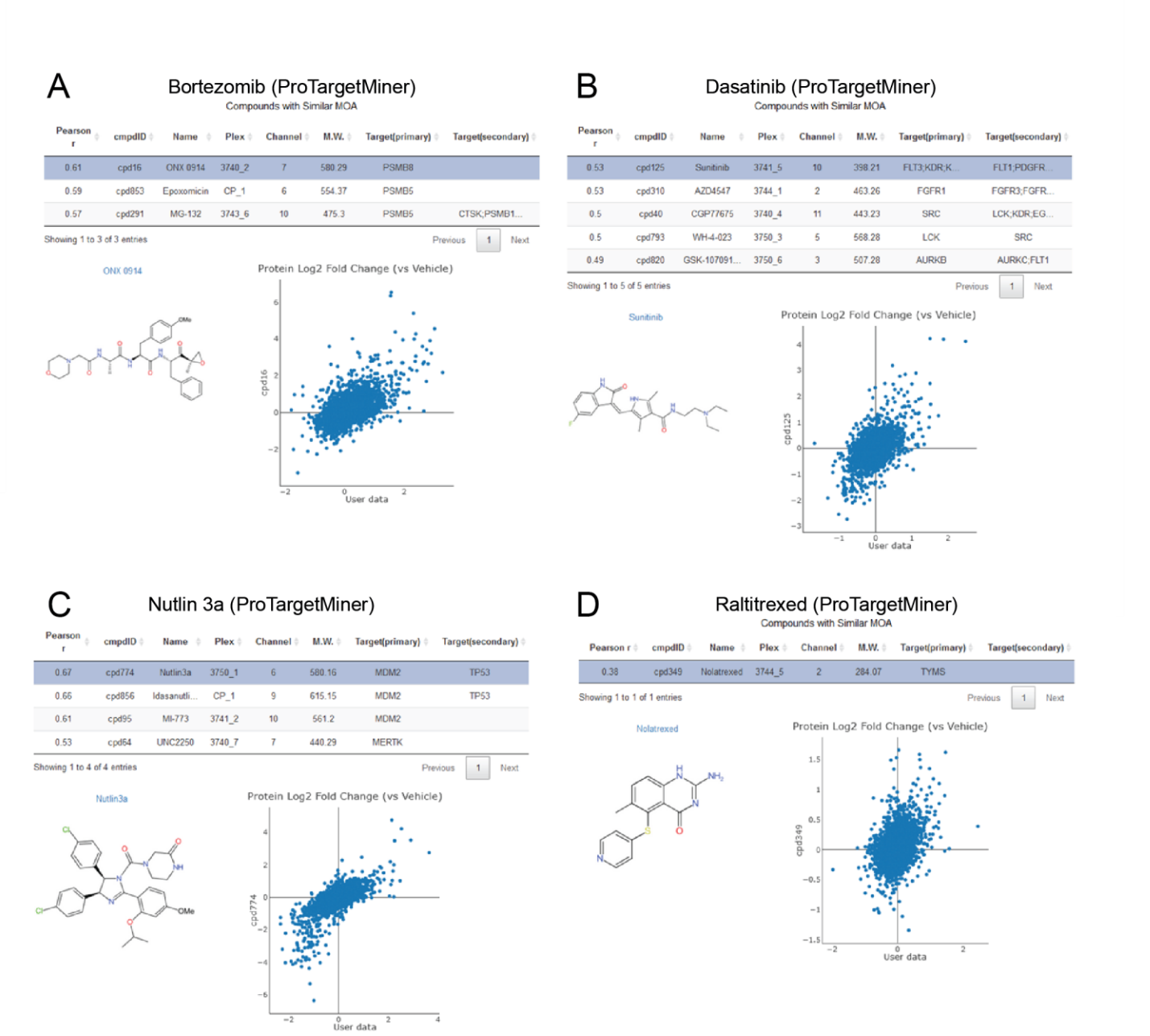
Supplemental Note #3: Proteome fingerprints are reproducible across concentrations

We next investigated the changes in proteome remodeling for eight highly cited compounds at different concentrations (**Fig. S9**). We analyzed global protein expression changes at a repeat concentration of 10 µM as well as a lower concentration of 500 nM or 1 µM. Despite 10-20-fold lower treatment concentrations, proteome changes were well conserved for this subset of compounds (**Fig. S9A-D**). We then calculated compound similarity for the rescreened compounds against the entire 875 compound MOA dataset using a Pearson correlation matrix (**Fig. S9B**). We found that the most highly related proteome fingerprint for rescreened compounds at each concentration was the same compound or another which targets the same protein, as was the case for Pictilisib and JQ-1 (**Fig. S9C**). Of the eight compounds investigated, only simvastatin proteome fingerprints were poorly correlated across the 20-fold treatment range. However, the target of this compound, HMGCR, was strongly upregulated at both 500 nM and 10 µM; a feature which is helpful in defining the MOA of simvastatin and possibly other HMGCR inhibitors (**Fig. S9E**).

We were encouraged by the results from the 1 µM treatment of the MDM2 inhibitor Nutlin-3a, which is used in cells at concentrations of 5-20 µM [15-17]. Despite 1 µM being much lower than the effective concentration for eliciting a change in the cell growth phenotype, this concentration was still highly correlated to Nutlin-3a from the MOA set. The proteome fingerprint was also highly similar to the more potent idasanutlin [18], which required a lower treatment concentration of 1 µM in the main screen due to toxicity at 10 µM. These findings suggest that treating cells with compounds at concentrations outside their “preferred” usage range may not strongly affect the interpretation of the correlated proteome fingerprints. The basis for this robustness seems to be that while the magnitudes of the proteome changes induced scale according to drug concentration, the proteins altered in response to drug treatment are largely the same, regardless of the dose. While the slope of the resulting regression may change (**Fig. S9D**), the associated correlation remains high.

Supplemental Note #4: Querying user data on the MOA Website

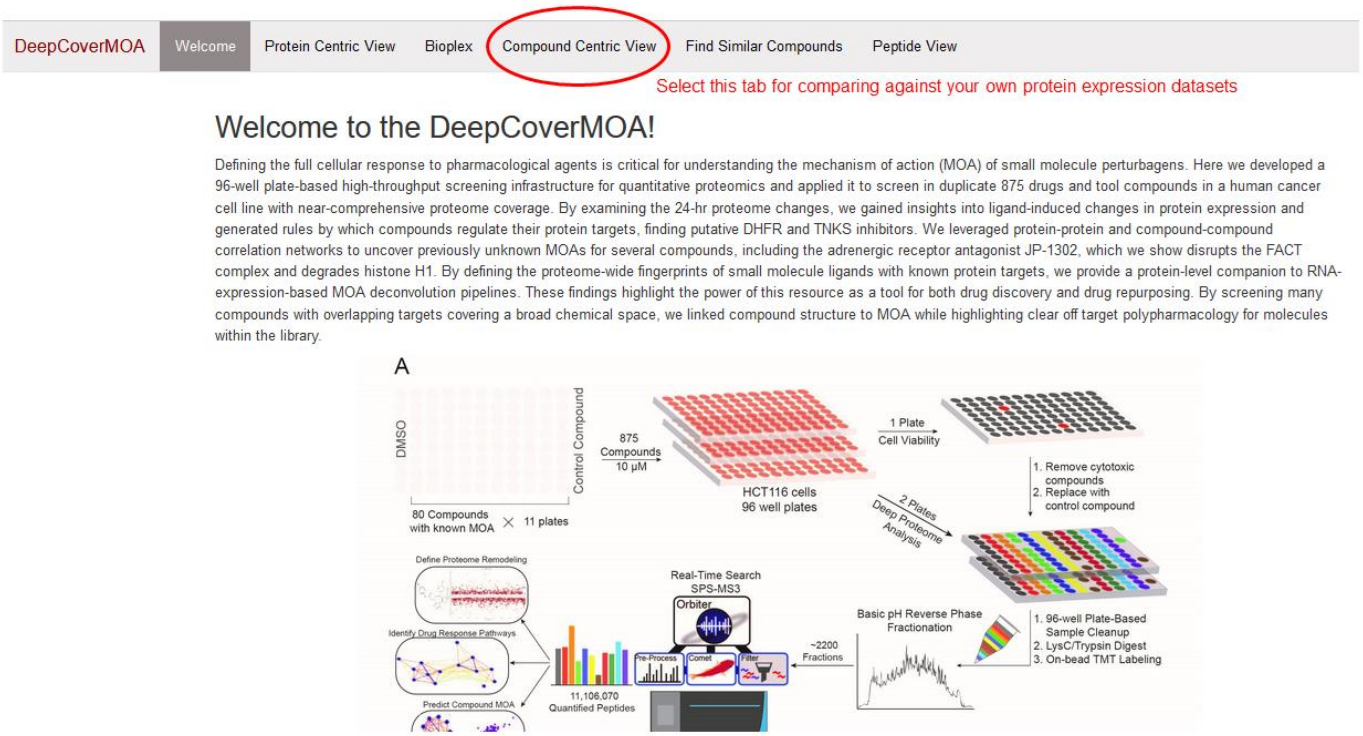
To further characterize the robustness of correlation-based similarity assignments, we used an external dataset of compound induced proteome changes [19]. Though these compounds were screened at different concentrations, in a cell line representing a different cancer type and a different genetic background, and acquired using a different MS method, the proteome fingerprints were consistent with our dataset (**SN4A**). Thus, our correlation-based analysis of proteome fingerprints remains able to correctly identify the most highly related compounds, despite substantial changes in compound concentration and screening conditions, preserving the utility of this resource for MOA deconvolution. We have included a feature on the website for users to query their own proteome expression data against our MOA dataset. (https://wren.hms.harvard.edu/DeepCoverMOA/) A tutorial for using this feature with user data is available below. (**SN4B**)



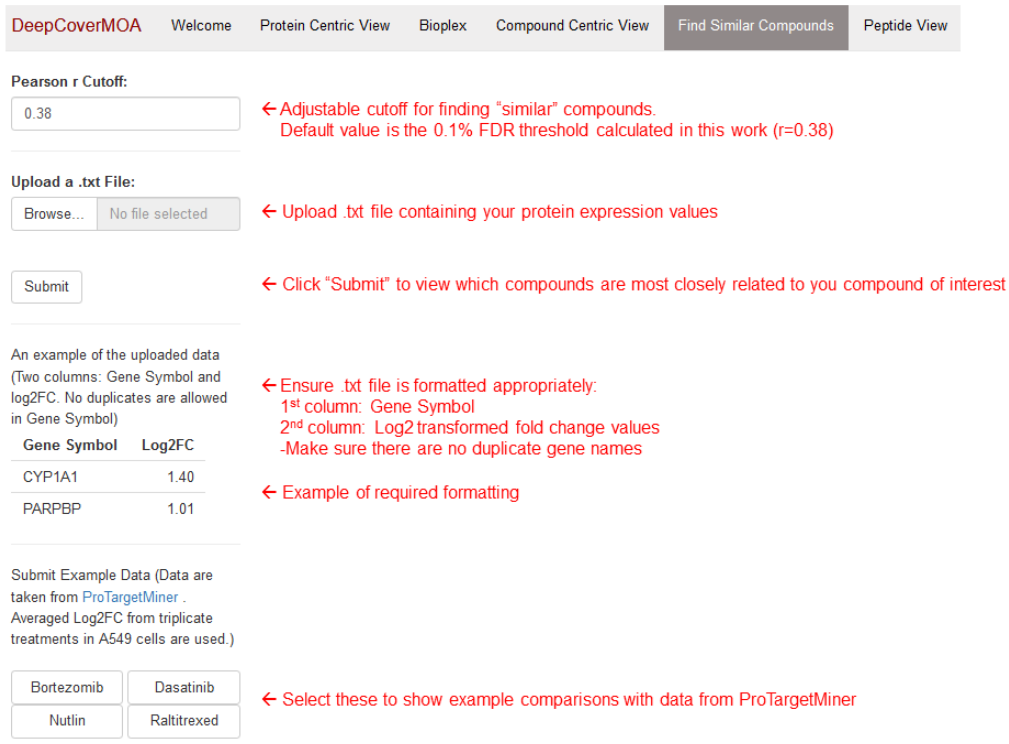
SN4A. Protein fingerprints from this study are compatible with those from external datasets. Using the MOA viewer "Find Similar Proteins" tool (Supplementary Note 1), protein fingerprints corresponding to four compounds from the ProTargetMiner [19] dataset were compared against the 875 compound MOA library. **(A)** Bortezomib – Targets the proteasome**, (B)** Dasatinib – inhibits many tyrosine kinases**, (C)** Nutlin-3a – inhibits MDM2**, (D)** Raltitrexed – Inhibits TYMS.

SN4B: Walkthrough of uploading a user-collected dataset for MOA comparison with the MOA library.

Step 1:



Step 2:



Step 3:

Table of all compounds with Pearson (r) over user set point

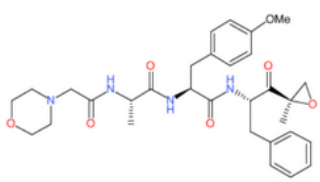
Compounds with Similar MOA

Pearson r	cmpdID	Name	Plex	Channel	M.W.	Target(primary)	Target(secondary)
0.61	cpd16	ONX 0914	3740_2	7	580.29	PSMB8	
0.59	cpd853	Epoxomicin	CP_1	6	554.37	PSMB5	
0.57	cpd291	MG-132	3743_6	10	475.3	PSMB5	CTSK;PSMB1...

Showing 1 to 3 of 3 entries

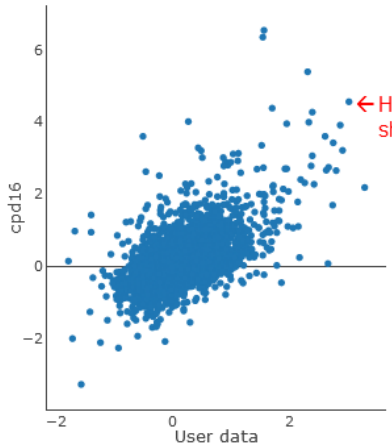
Previous 1 Next

ONX 0914



Chemical Structure of selected compound from above table

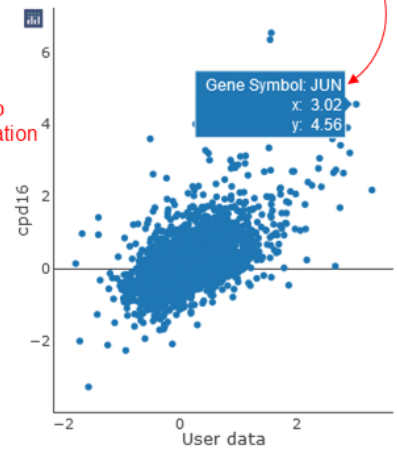
Protein Log2 Fold Change (vs Vehicle)



Data from user input .txt file

← Hover over point to show more information

Gene Sybol
X: L2FC from user input data
Y: L2FC from MOA library dataset



References – Supplementary Notes 1-4

1. Uhlen, M., P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Bjorling, and F. Ponten, (2010), Towards a knowledge-based Human Protein Atlas. Nat Biotechnol *28*. 1248-50.
2. Santos, R., O. Ursu, A. Gaulton, A.P. Bento, R.S. Donadi, C.G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T.I. Oprea, and J.P. Overington, (2017), A comprehensive map of molecular drug targets. Nat Rev Drug Discov *16*. 19-34.
3. Dubach, J.M., E. Kim, K. Yang, M. Cuccarese, R.J. Giedt, L.G. Meimetis, C. Vinegoni, and R. Weissleder, (2017), Quantitating drug-target engagement in single cells in vitro and in vivo. Nat Chem Biol *13*. 168-173.
4. Bunnage, M.E., E.L. Chekler, and L.H. Jones, (2013), Target validation using chemical probes. Nat Chem Biol *9*. 195-9.
5. Ruprecht, B., J. Di Bernardo, Z. Wang, X. Mo, O. Ursu, M. Christopher, R.B. Fernandez, L. Zheng, B.D. Dill, H. Wang, Y. Xu, A. Liaw, J.D. Mortison, N. Siriwardana, B. Andresen, M. Glick, J.R. Tata, V. Kutilek, I. Cornella-Taracido, and A. Chi, (2020), A mass spectrometry-based proteome map of drug action in lung cancer cell lines. Nat Chem Biol *16*. 1111-1119.
6. Hughes, J.P., S. Rees, S.B. Kalindjian, and K.L. Philpott, (2011), Principles of early drug discovery. Br J Pharmacol *162*. 1239-49.
7. Weiss, W.A., S.S. Taylor, and K.M. Shokat, (2007), Recognizing and exploiting differences between RNAi and small-molecule inhibitors. Nat Chem Biol *3*. 739-44.
8. Boland, C.R. and A. Goel, (2010), Microsatellite instability in colorectal cancer. Gastroenterology *138*. 2073-2087 e3.
9. Bailey, M.H., C. Tokheim, E. Porta-Pardo, S. Sengupta, D. Bertrand, A. Weerasinghe, A. Colaprico, M.C. Wendl, J. Kim, B. Reardon, P.K. Ng, K.J. Jeong, S. Cao, Z. Wang, J. Gao, Q. Gao, F. Wang, E.M. Liu, L. Mularoni, C. Rubio-Perez, N. Nagarajan, I. Cortes-Ciriano, D.C. Zhou, W.W. Liang, J.M. Hess, V.D. Yellapantula, D. Tamborero, A. Gonzalez-Perez, C. Suphavilai, J.Y. Ko, E. Khurana, P.J. Park, E.M. Van Allen, H. Liang, M.C.W. Group, N. Cancer Genome Atlas Research, M.S. Lawrence, A. Godzik, N. Lopez-Bigas, J. Stuart, D. Wheeler, G. Getz, K. Chen, A.J. Lazar, G.B. Mills, R. Karchin, and L. Ding, (2018), Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*. 371-385 e18.
10. Prior, I.A., F.E. Hood, and J.L. Hartley, (2020), The Frequency of Ras Mutations in Cancer. Cancer Res *80*. 2969-2974.
11. Siegel, R.L., K.D. Miller, A. Goding Sauer, S.A. Fedewa, L.F. Butterly, J.C. Anderson, A. Cercek, R.A. Smith, and A. Jemal, (2020), Colorectal cancer statistics, 2020. CA Cancer J Clin *70*. 145-164.
12. Huttlin, E.L., R.J. Bruckner, J. Navarrete-Perea, J.R. Cannon, K. Baltier, F. Gebreab, M.P. Gygi, A. Thornock, G. Zarraga, S. Tam, J. Szpyt, B.M. Gassaway, A. Panov, H. Parzen, S. Fu, A. Golbazi, E. Maenpaa, K. Stricker, S. Guha Thakurta, T. Zhang, R. Rad, J. Pan, D.P. Nusinow, J.A. Paulo, D.K. Schweppe, L.P. Vaites, J.W. Harper, and S.P. Gygi, (2021), Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell *184*. 3022-3040 e28.
13. Nusinow, D.P. and S.P. Gygi, (2020), A Guide to the Quantitative Proteomic Profiles of the Cancer Cell Line Encyclopedia. bioRxiv 2020.02.03.932384.
14. Nusinow, D.P., J. Szpyt, M. Ghandi, C.M. Rose, E.R. McDonald, 3rd, M. Kalocsay, J. Jane-Valbuena, E. Gelfand, D.K. Schweppe, M. Jedrychowski, J. Golji, D.A. Porter, T. Rejtar, Y.K. Wang, G.V. Kryukov, F. Stegmeier, B.K. Erickson, L.A. Garraway, W.R. Sellers, and S.P. Gygi, (2020), Quantitative Proteomics of the Cancer Cell Line Encyclopedia. Cell *180*. 387-402 e16.
15. Arya, A.K., A. El-Fert, T. Devling, R.M. Eccles, M.A. Aslam, C.P. Rubbi, N. Vlatkovic, J. Fenwick, B.H. Lloyd, D.R. Sibson, T.M. Jones, and M.T. Boyd, (2010), Nutlin-3, the small-molecule inhibitor of MDM2, promotes senescence and radiosensitises laryngeal carcinoma cells harbouring wild-type p53. Br J Cancer *103*. 186-95.
16. Valente, L.J., B.J. Aubrey, M.J. Herold, G.L. Kelly, L. Happo, C.L. Scott, A. Newbold, R.W. Johnstone, D.C. Huang, L.T. Vassilev, and A. Strasser, (2016), Therapeutic Response to Non-genotoxic Activation of p53 by Nutlin3a Is Driven by PUMA-Mediated Apoptosis in Lymphoma Cells. Cell Rep *14*. 1858-66.
17. Villalonga-Planells, R., L. Coll-Mulet, F. Martinez-Soler, E. Castano, J.J. Acebes, P. Gimenez-Bonafe, J. Gil, and A. Tortosa, (2011), Activation of p53 by nutlin-3a induces apoptosis and cellular senescence in human glioblastoma multiforme. PLoS One *6*. e18588.

18. Ding, Q., Z. Zhang, J.J. Liu, N. Jiang, J. Zhang, T.M. Ross, X.J. Chu, D. Bartkovitz, F. Podlaski, C. Janson, C. Tovar, Z.M. Filipovic, B. Higgins, K. Glenn, K. Packman, L.T. Vassilev, and B. Graves, (2013), Discovery of RG7388, a potent and selective p53-MDM2 inhibitor in clinical development. J Med Chem *56*. 5979-83.
19. Saei, A.A., C.M. Beusch, A. Chernobrovkin, P. Sabatier, B. Zhang, U.G. Tokat, E. Stergiou, M. Gaetani, A. Vegvari, and R.A. Zubarev, (2019), ProTargetMiner as a proteome signature library of anticancer molecules for functional discovery. Nat Commun *10*. 5715.