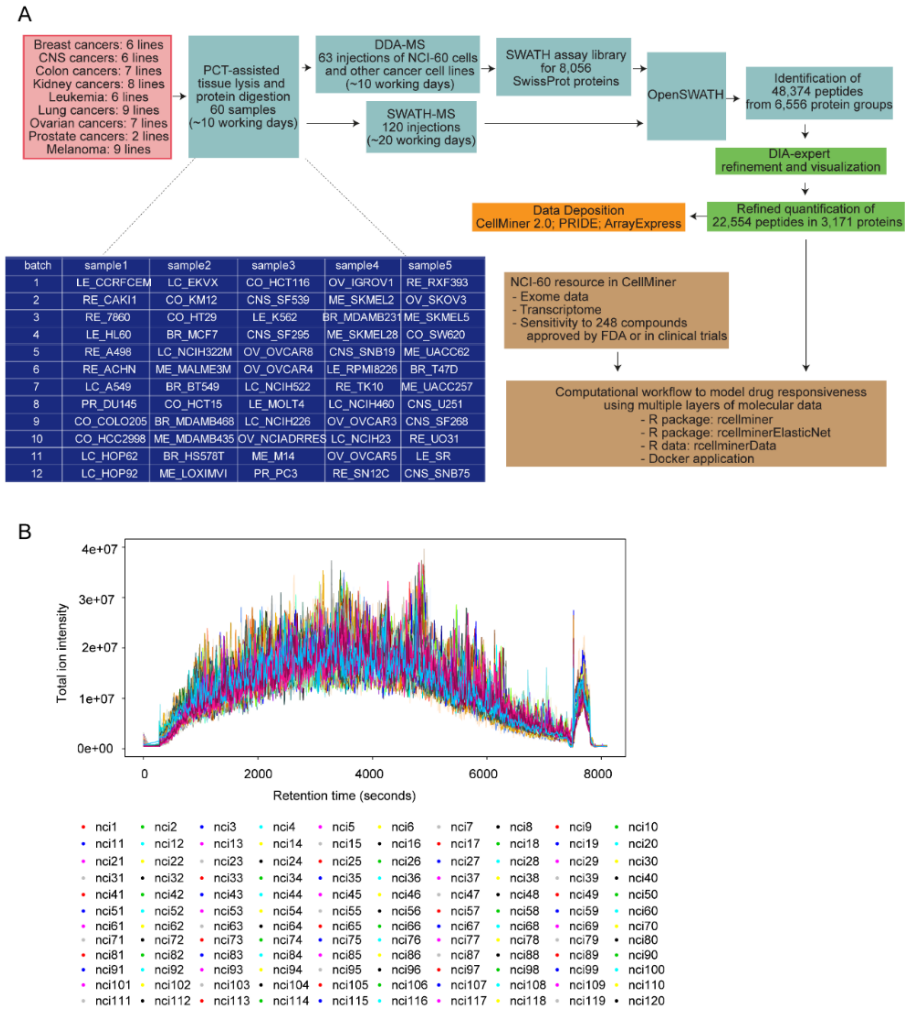
ISCI, Volume *21*

Supplemental Information

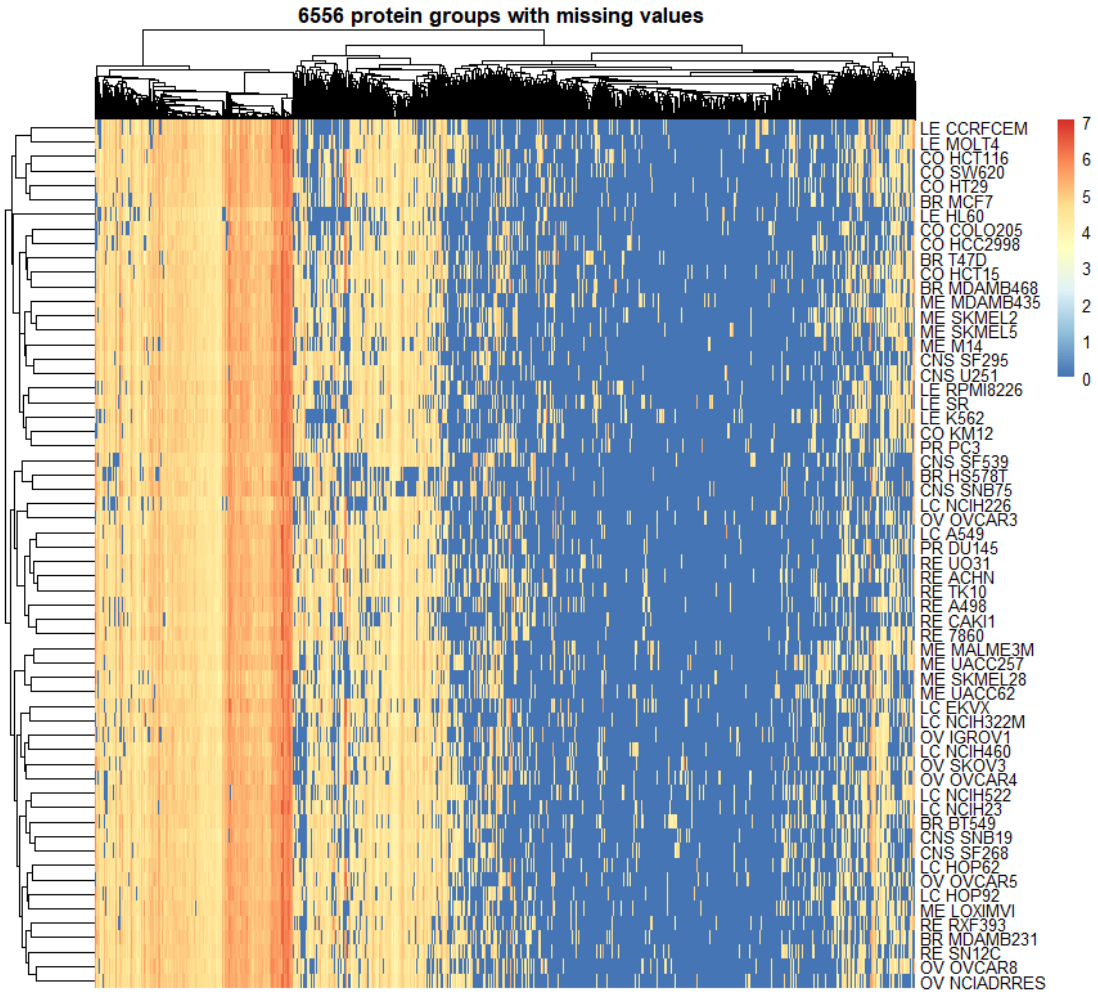
Quantitative Proteome Landscape of the NCI-60 Cancer Cell Lines

Tiannan Guo, Augustin Luna, Vinodh N. Rajapakse, Ching Chiek Koh, Zhicheng Wu, Wei Liu, Yaoting Sun, Huanhuan Gao, Michael P. Menden, Chao Xu, Laurence Calzone, Loredana Martignetti, Chiara Auwerx, Marija Buljan, Amir Banaei-Esfahani, Alessandro Ori, Murat Iskar, Ludovic Gillet, Ran Bi, Jiangnan Zhang, Huanhuan Zhang, Chenhuan Yu, Qing Zhong, Sudhir Varma, Uwe Schmitt, Peng Qiu, Qiushi Zhang, Yi Zhu, Peter J. Wild, Mathew J. Garnett, Peer Bork, Martin Beck, Kexin Liu, Julio Saez-Rodriguez, Fathi Elloumi, William C. Reinhold, Chris Sander, Yves Pommier, and Ruedi Aebersold

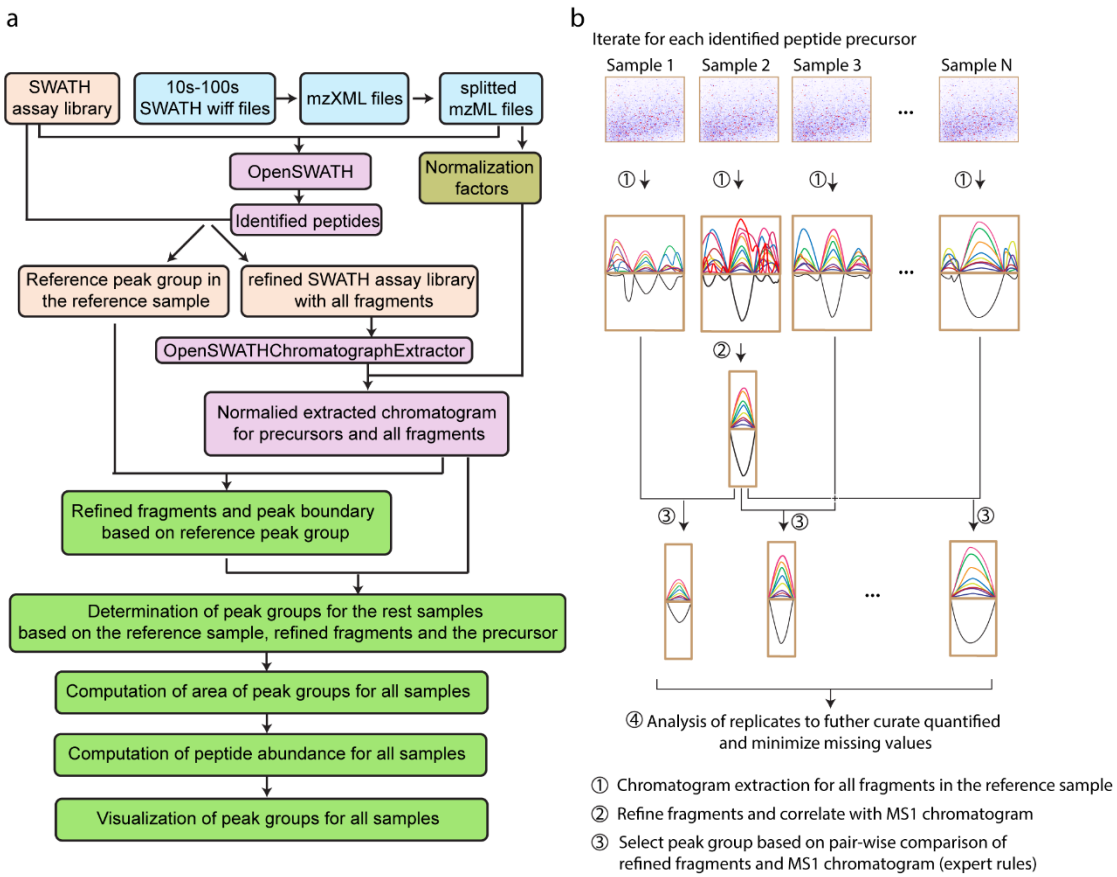
Supplementary Figures



Supplementary Figure 1, Related to Figure 1. Workflow for generating the NCI-60 proteome maps and predicting phenotypes. (A) Flowchart of experimental design. The NCI-60 cell pellets were divided into 12 batches, lysed and digested using the PCT method. The peptides were first analyzed in DDA mode to build a SWATH assay library. We also included DDA files from U2OS and HeLa cell digests. In total we performed 63 DDA injections either from whole cell lysate or fractionated samples. Each sample was analyzed in SWATH mode twice. The SWATH data were processed using software tools including OpenSWATH and DIA-expert in sequence. Our data were deposited in several public databases including CellMiner 2.0. Subsequently, we developed a computational workflow to model drug responsiveness using multiple layers of molecular data. The generation of a spectral library specifically for the NCI-60 cells consumed ca. 10 working days. For studies of this type this step is optional because similar results can be obtained from the use of publicly accessible, extensive human spectral libraries such as the pan-human library (Rosenberger et al., 2014). (**B**) raw mass spectrometric signal for the 120 SWATH runs. Total ion chromatography graphs are shown. The index of the 120 NCI SWATH files is explained in **Supplementary Table 1**.



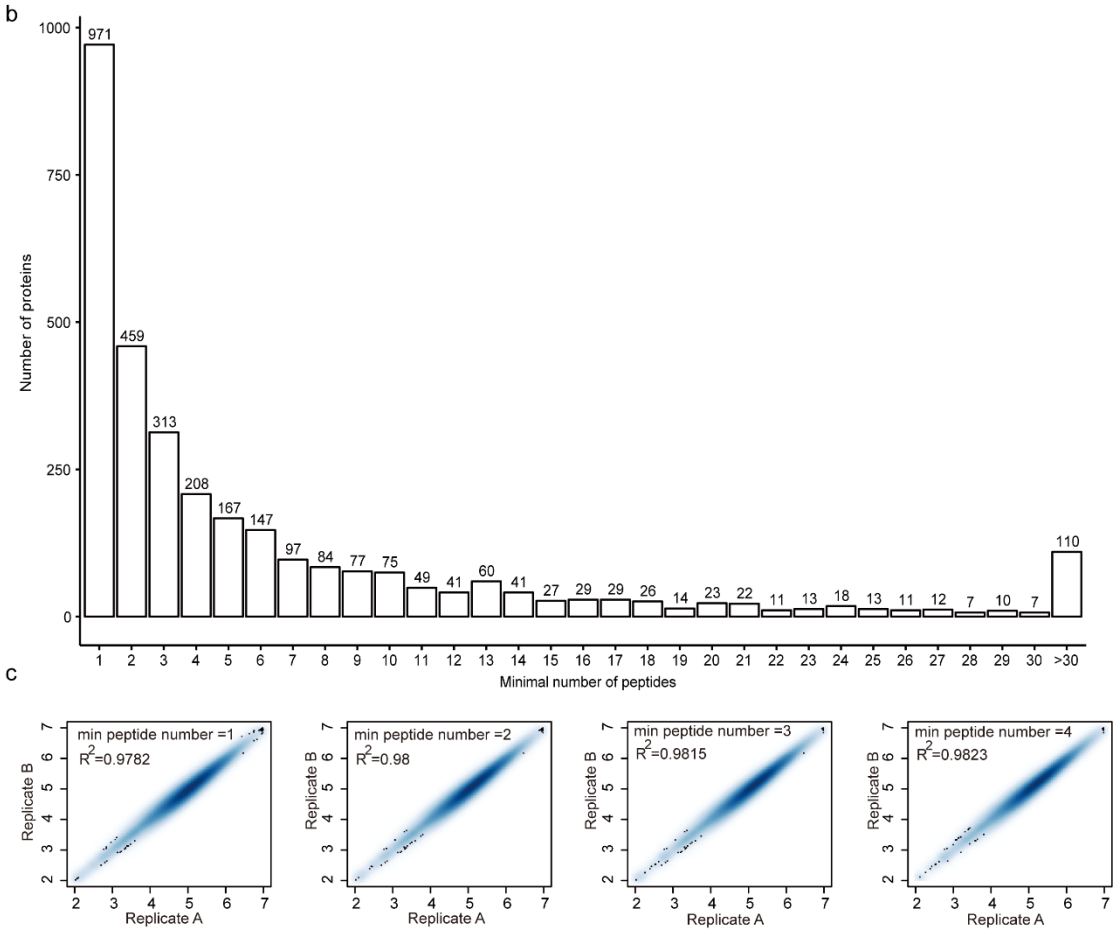
Supplementary Figure 2, Related to Figure 1. Unsupervised clustering of 6556 protein groups identified and quantified in the NCI-60 cells. Using the SWATH library containing 8056 protein groups, we displayed the identified and quantified protein groups after unsupervised clustering of both cells and proteins based on their log10 transformed intensity values.



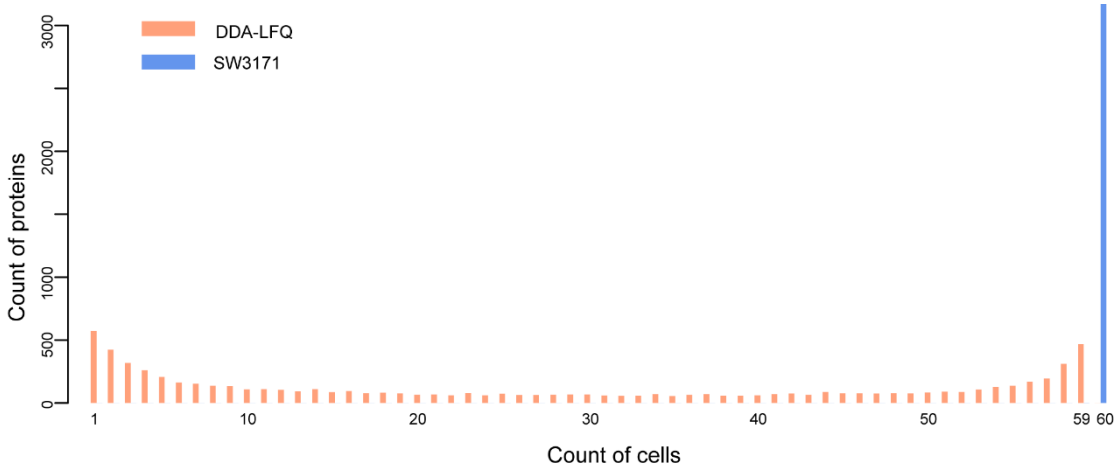
Supplementary Figure 3, Related to Figure 1. Design and implementation of DIA-expert, Related to Figure 1. (**a**) DIA-experts reads output data from OpenSWATH analysis of SWATH/DIA maps and then curates and visualizes quantitative ion chromatogram signals. (**b**) DIA-expert analyses each identified peptide precursor in a sample set. In Step ① it extracts ion chromatography signals for any number of fragments and the precursor ion chromatogram for all samples. In Step ②, it selects reference sample(s) from the sample set and refines non-contaminated chromatographic signals by learning the signal characteristics of the reference sample(s). In step 3 the system performs pair-wise comparisons of the reference sample(s) and a sample to be quantified based on the refined fragments ion set. Last, replicates of each sample and proteotypic peptides from the same protein were considered to exclude unreliably quantified peptides and minimize missing values for protein quantification across the entire data set.

a

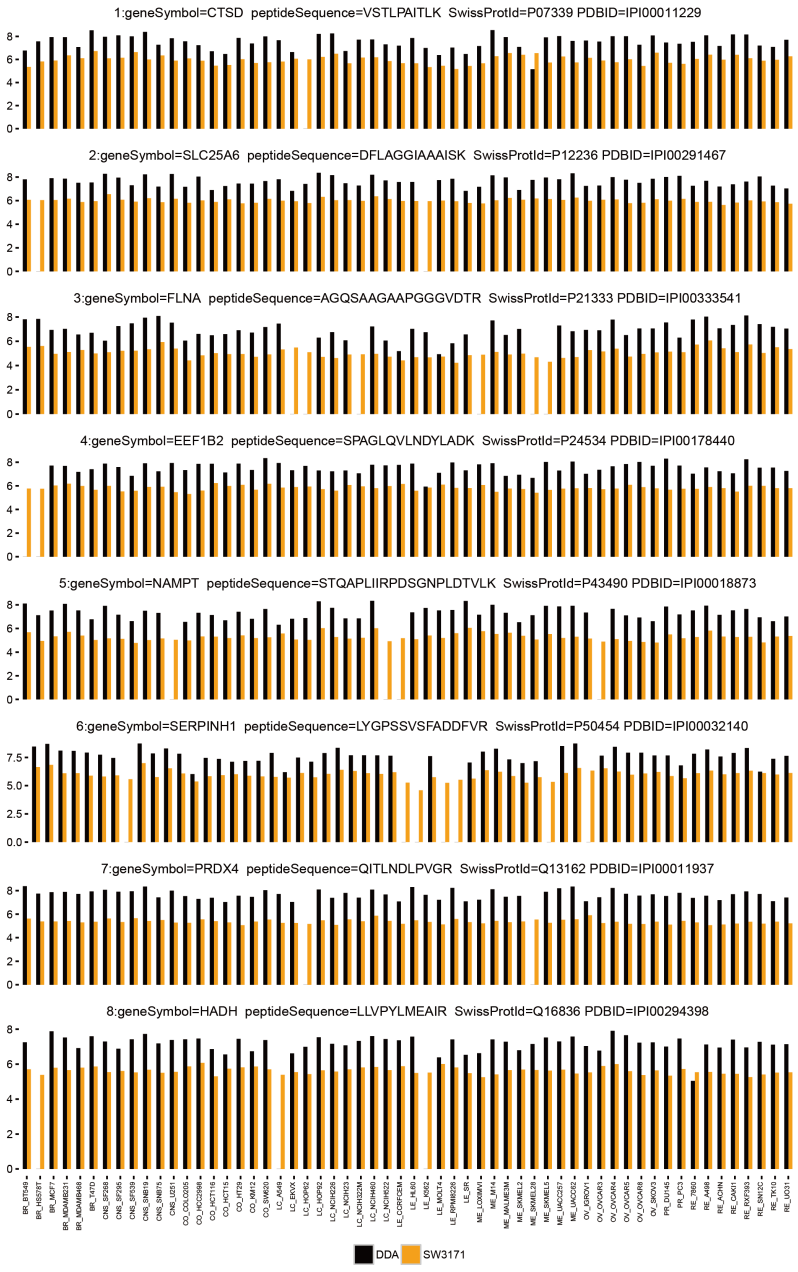
Min number of peptides	Number of proteins	R2
1	3171	0.9872
2	2200	0.9800
3	1741	0.9815
4	1428	0.9823



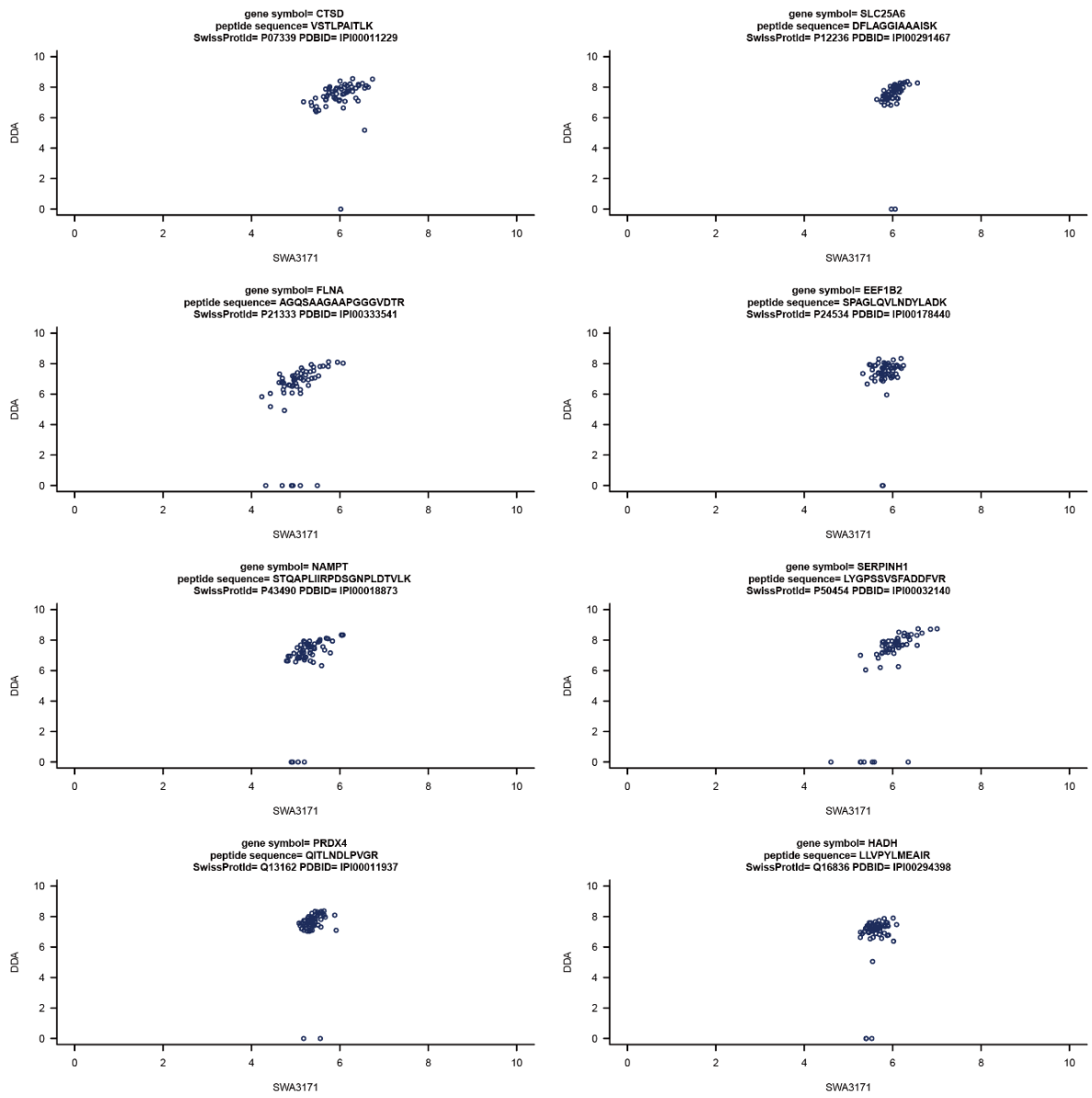
Supplementary Figure 4, Related to Figure 1. Quantitative accuracy of the NCI-60 proteome as a function of the number of peptides quantified per protein. (**a**) Number of proteins quantified when minimally 1, 2, 3 and 4 peptides were quantified per protein. The R2 values of technical replicates are computed. (**b**) Distribution of protein numbers based on increasing number of peptides. (**c**) The heatmap scatter plot of proteins quantified in two technical replicates when the minimal peptide number is limited to 1, 2, 3 and 4.



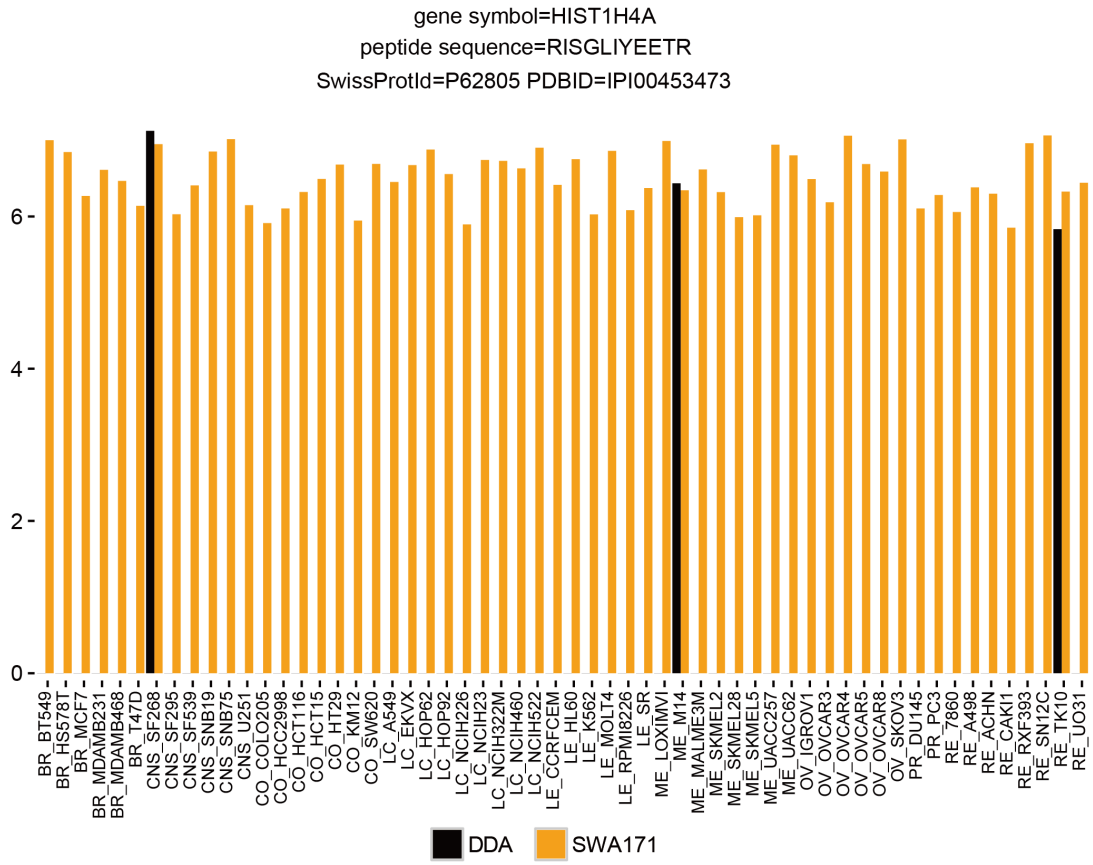
Supplementary Figure 5, Related to Figure 2. Count of proteins quantified in increasing number of cells. This plot shows the number of proteins quantified in the NCI-60 cells. DDA-LFQ denotes the LFQ-processed DDA data of the NCI-60 cells. SW3171 means the SWATH data set presented in this study. Most of the SW3171 proteins were quantified in all 60 cells. In DDA-LFQ data set (Gholami et al., 2013b), highest numbers of IPI protein groups were quantified in 1 and 59 cells.



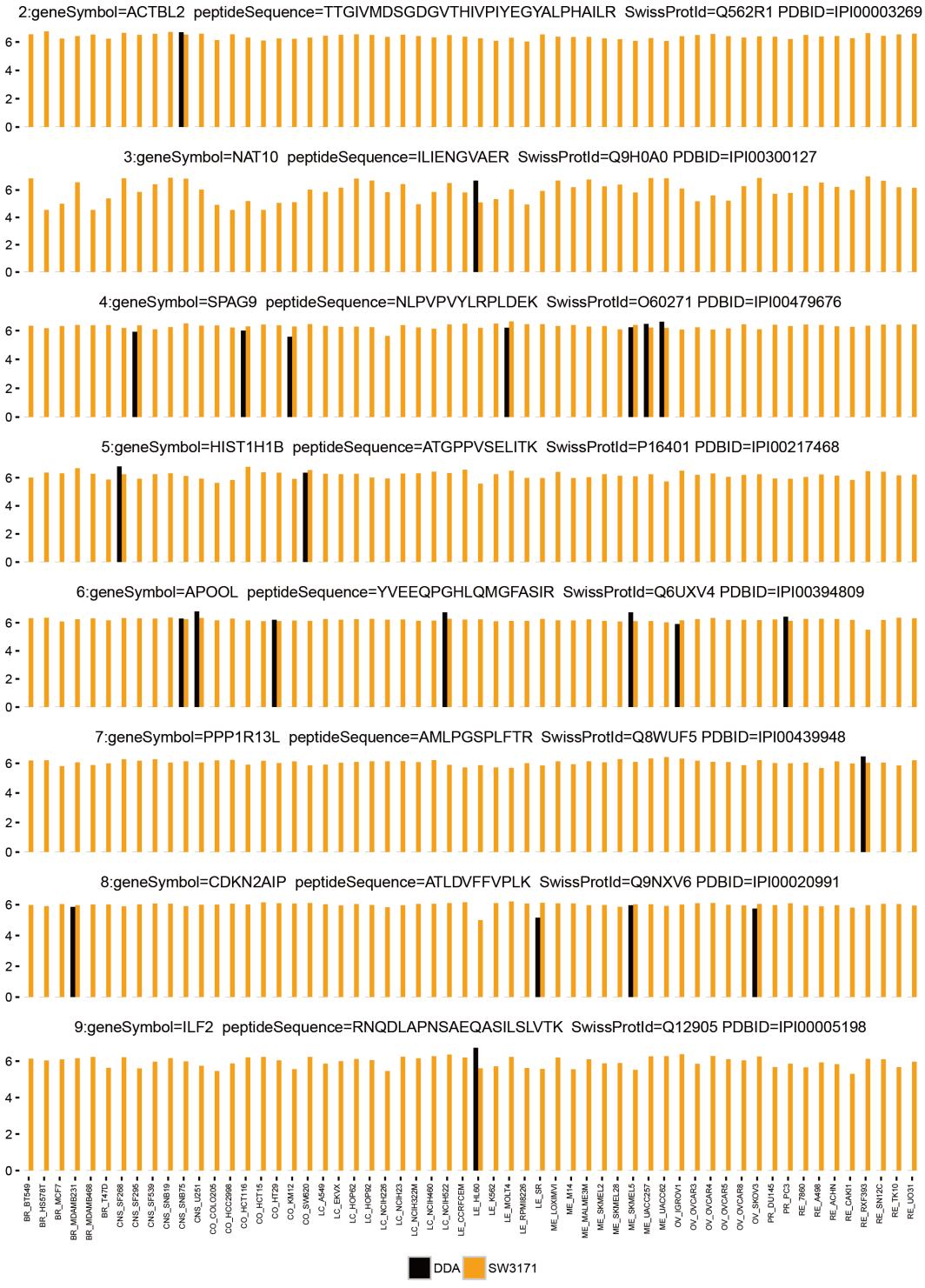
Supplementary Figure 6, Related to Figure 2. Comparison of 8 representative proteins which have been consistently quantified across nearly all NCI-60 cell lines by DDA. The data are shown in bar plots. Protein intensity values are log10 scaled



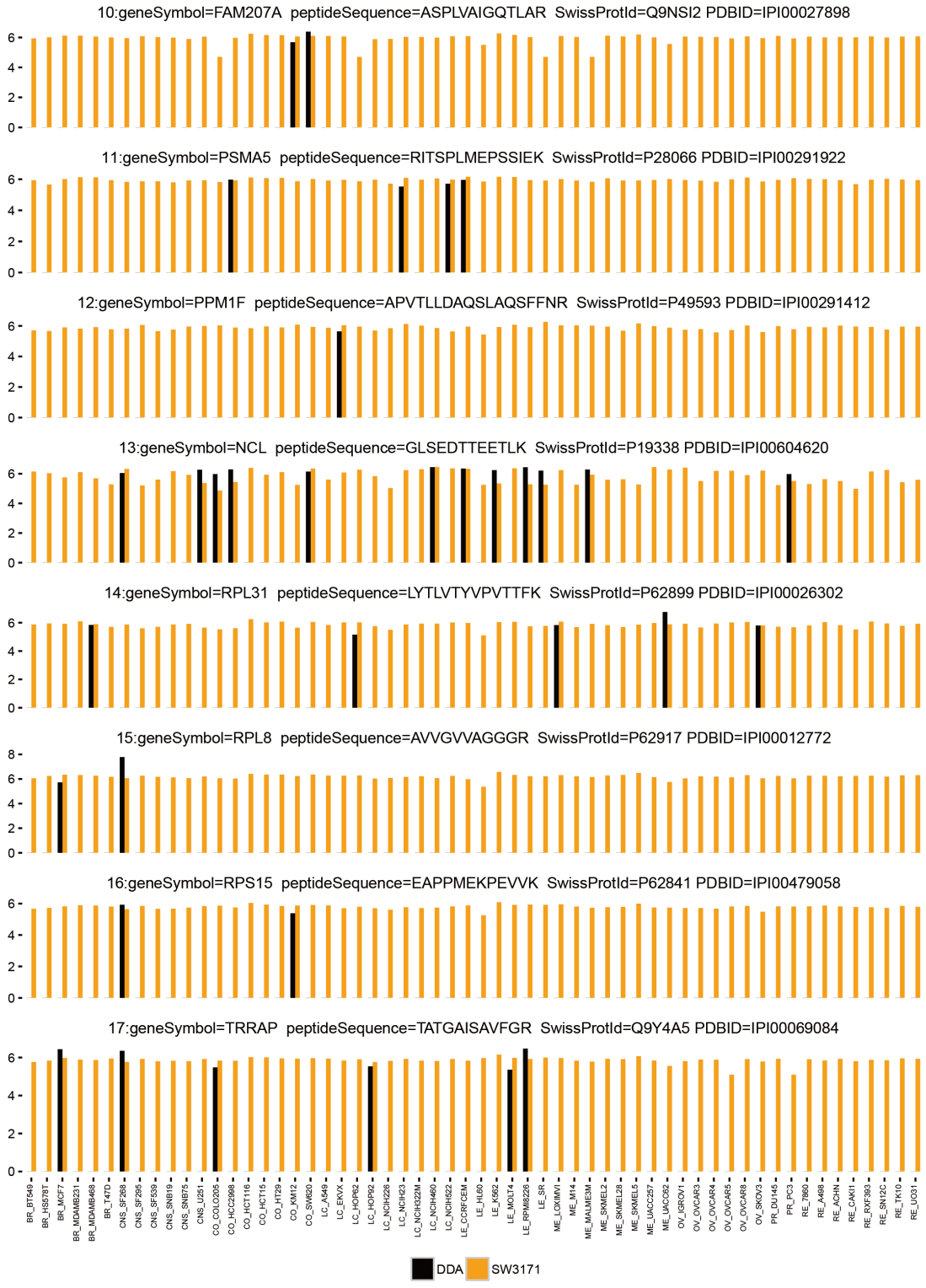
Supplementary Figure 7, Related to Figure 2. Comparison of 8 representative proteins which have been consistently quantified across nearly all NCI-60 cell lines by SWATH. The data are shown in scatter plots. Protein intensity values are log10 scaled.



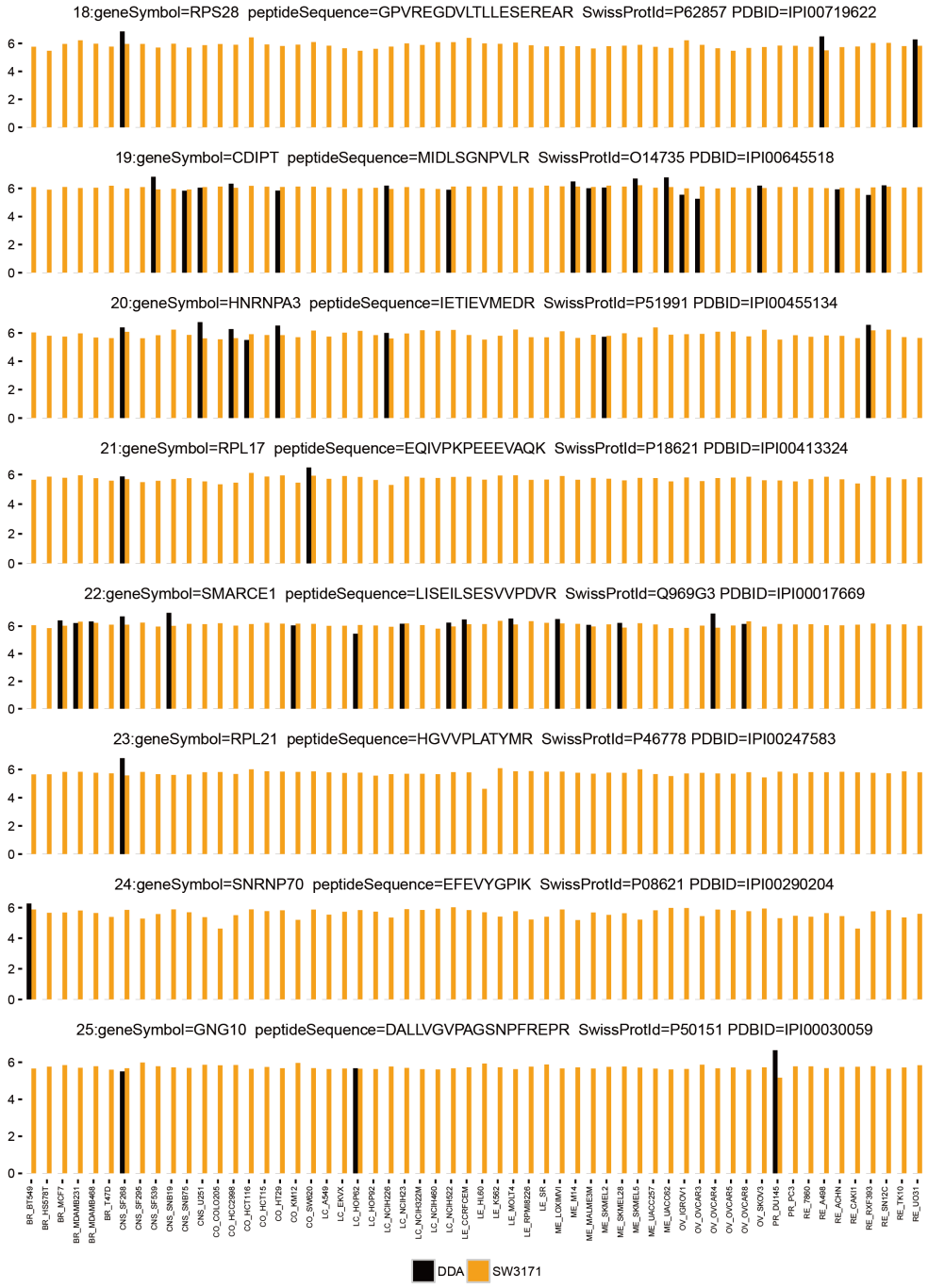
Supplementary Figure 8, Related to Figure 2. Bar plots for P62805, which is quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



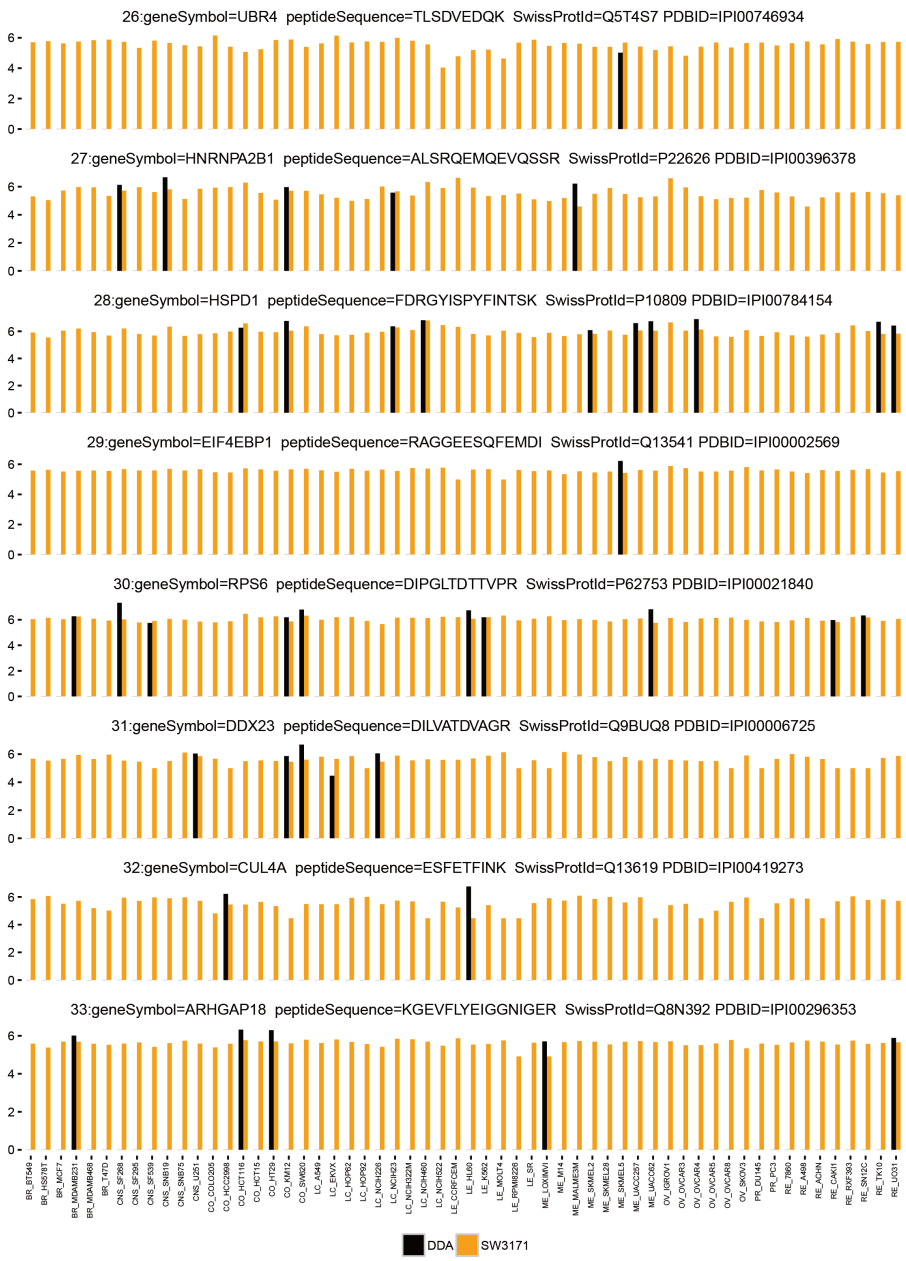
Supplementary Figure 9, Related to Figure 2. Bar plots for Q562R1, Q9H0A0, O60271, P16401, Q6UXV4, Q8WUF5, Q9NXV6 and Q12905, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



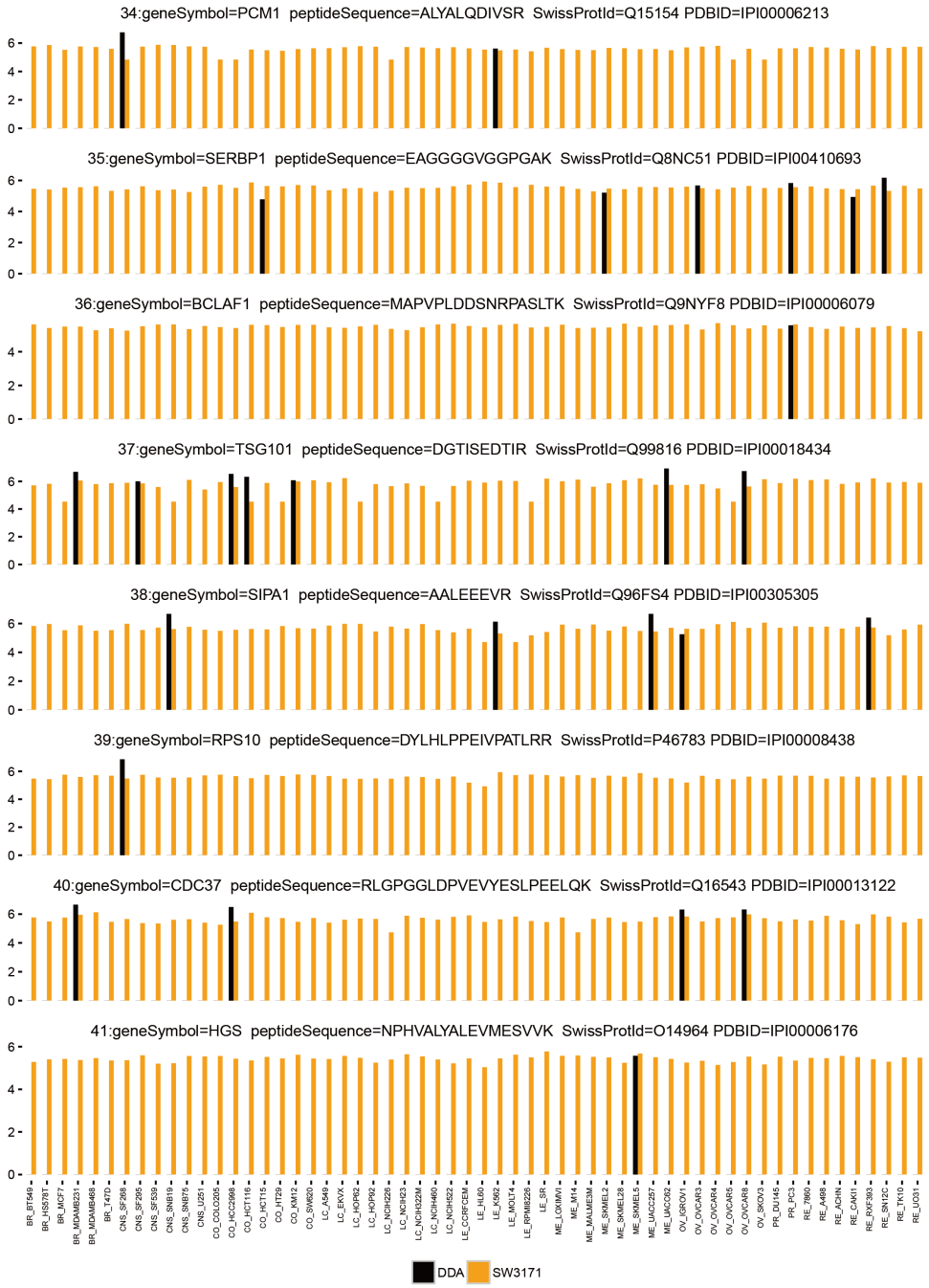
Supplementary Figure 10, Related to Figure 2. Bar plots for Q9NSI2, P28066, P49593, P19338, P62899, P62917, P62841, and Q9Y4A5, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



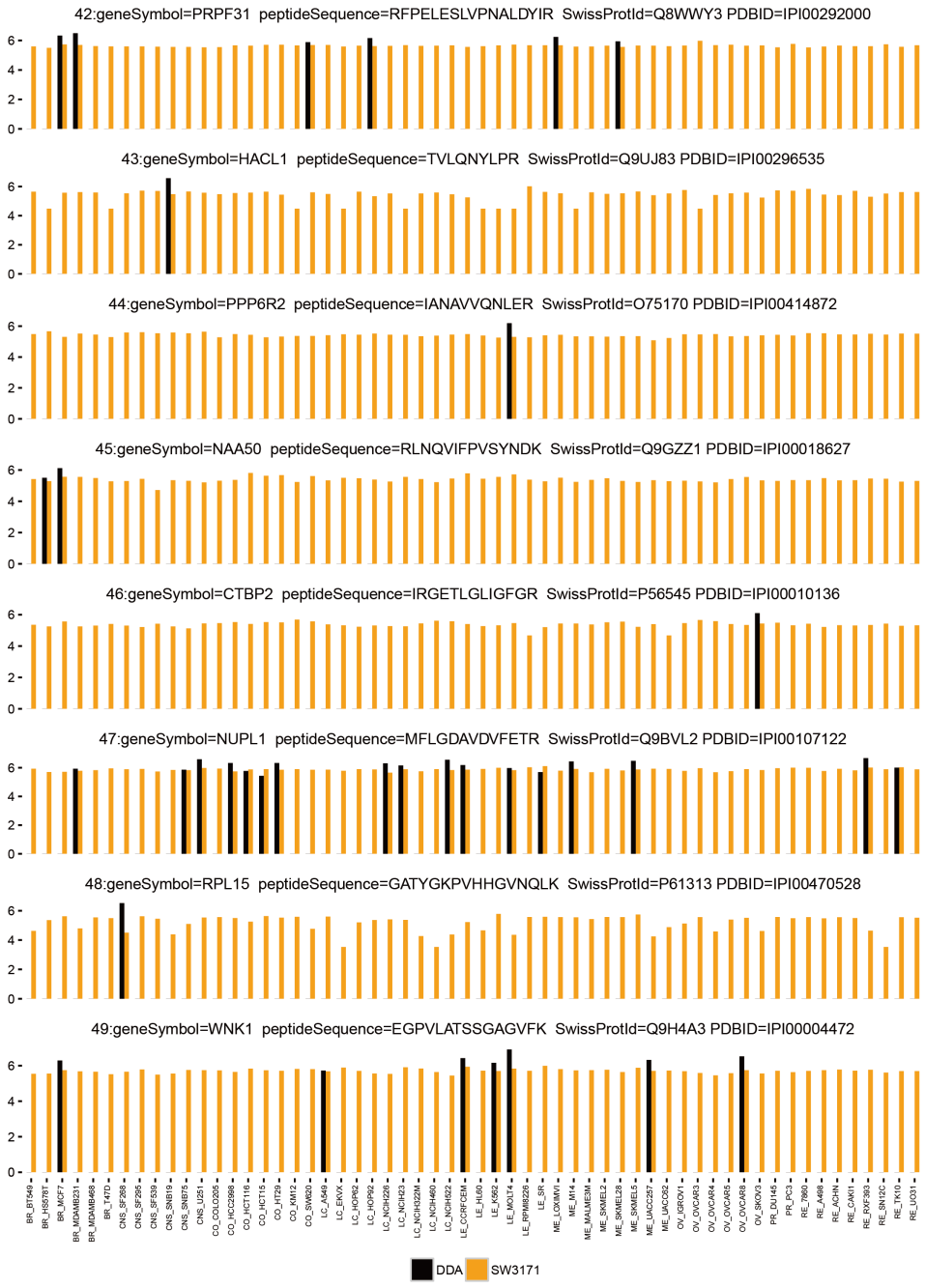
Supplementary Figure 11, Related to Figure 2. Bar plots for P62875, O14735, P51991, P18621, Q969G3, P46778, P08621, and P50151, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



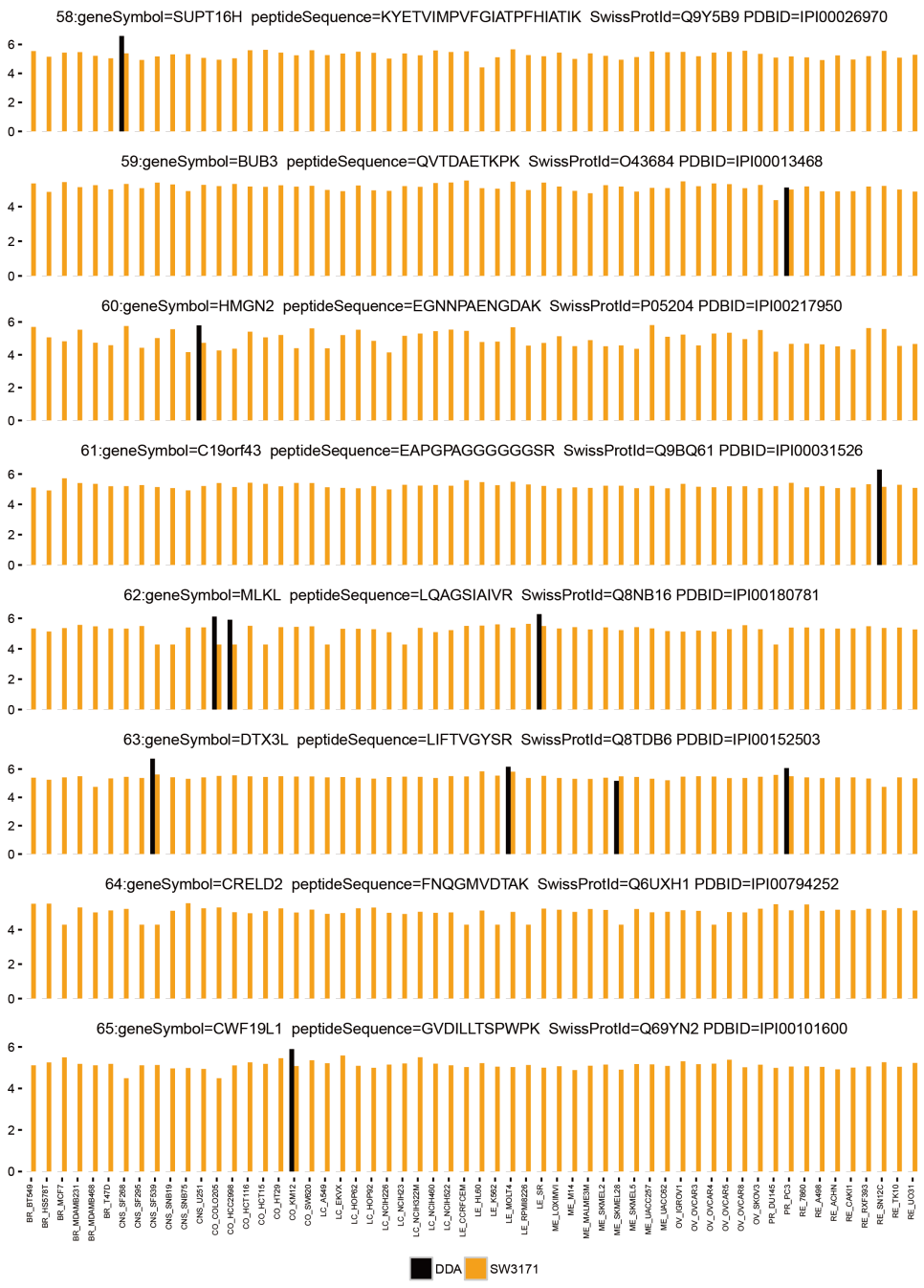
Supplementary Figure 12, Related to Figure 2. Bar plots for Q5T4S7, P22626, P10809, Q13541, P62753, Q9BUQ8, Q13619 and Q8N392, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



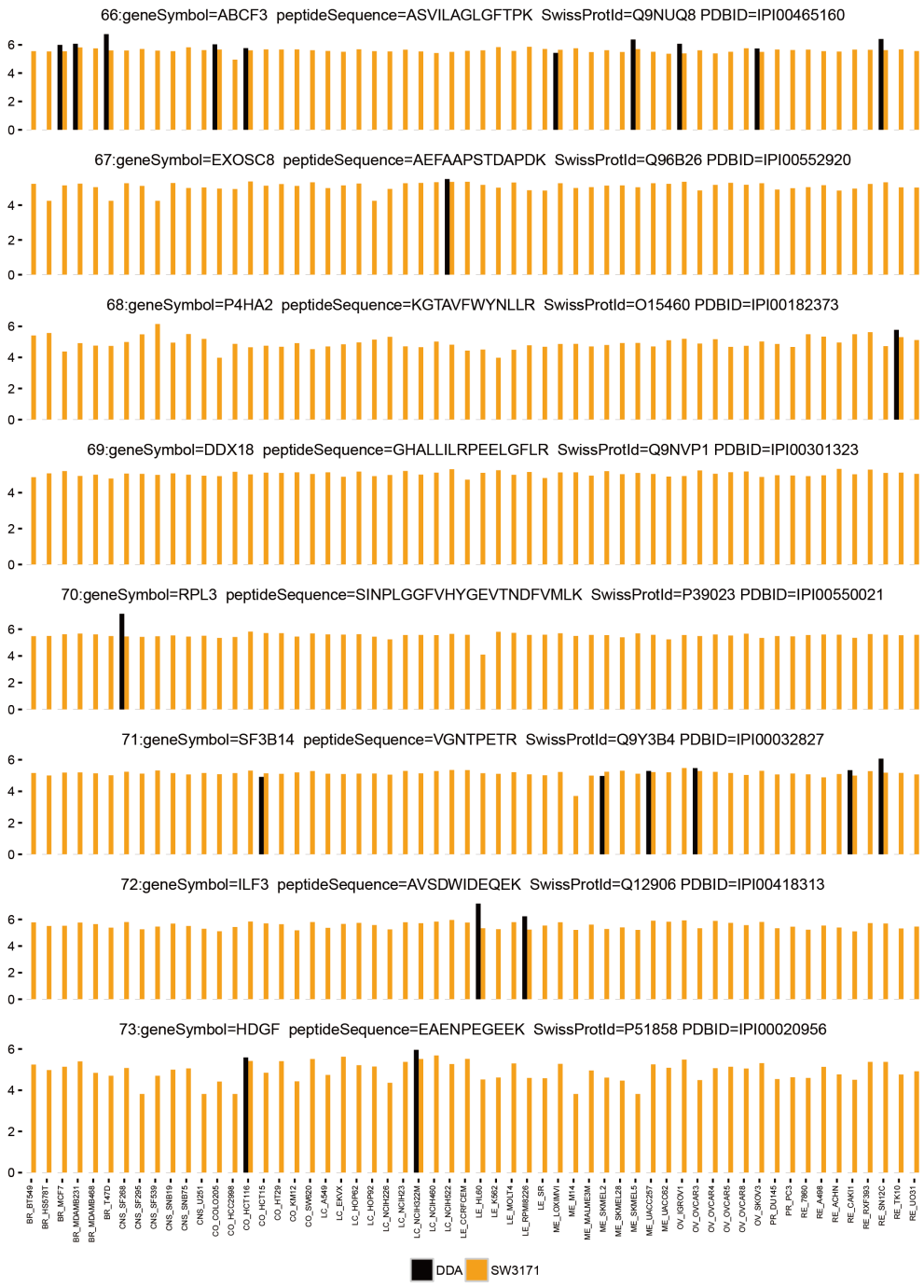
Supplementary Figure 13, Related to Figure 2. Bar plots for Q15154, Q8NC51, Q9NYF8, Q99816, Q96FS4, P46783, Q16543, O14964, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



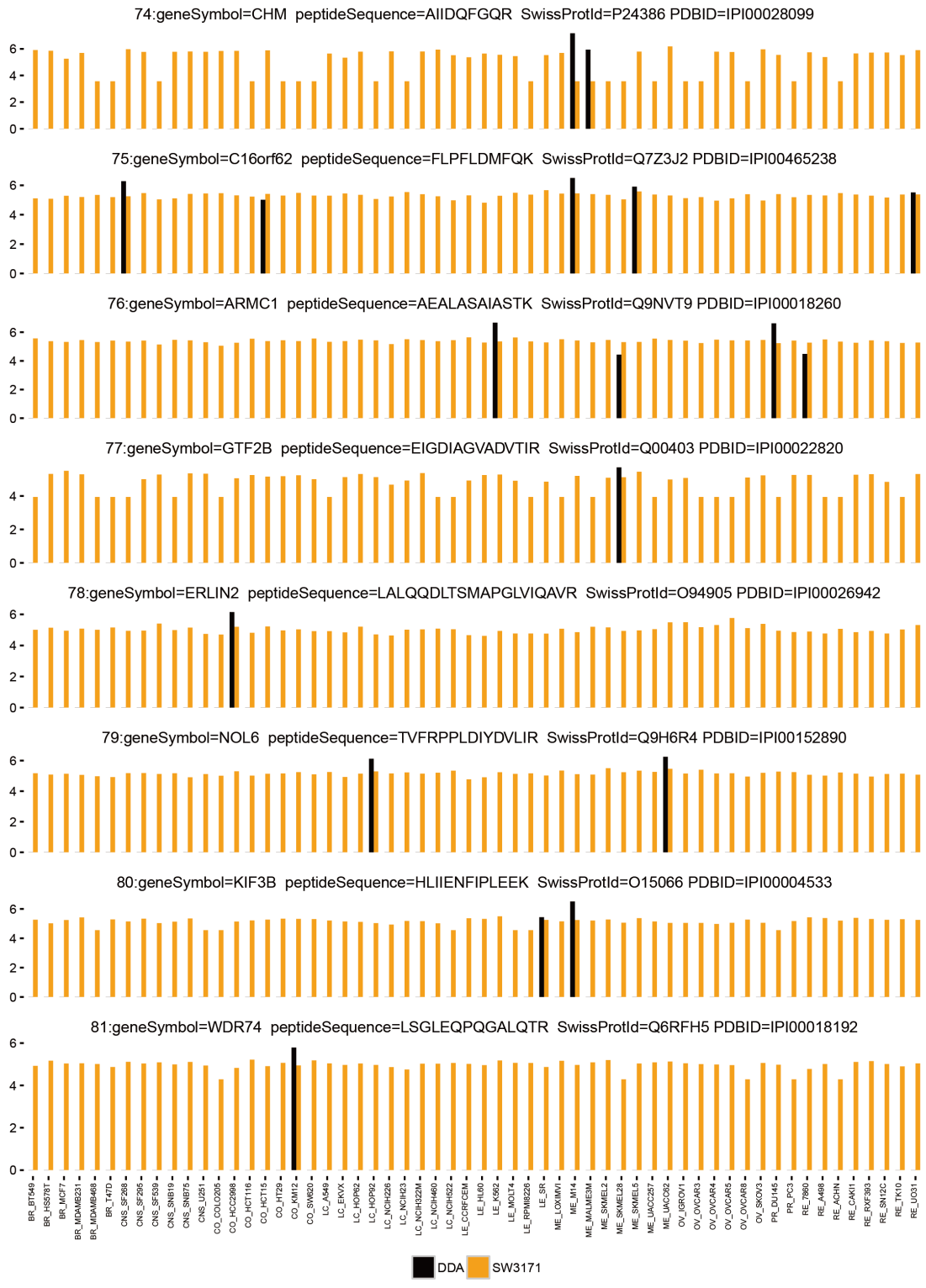
Supplementary Figure 14, Related to Figure 2. Bar plots for Q8WWY3, Q9UJB3, O75170, Q9GZZ1, P56545, Q9BVL2, P61313 and Q9H4A3, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



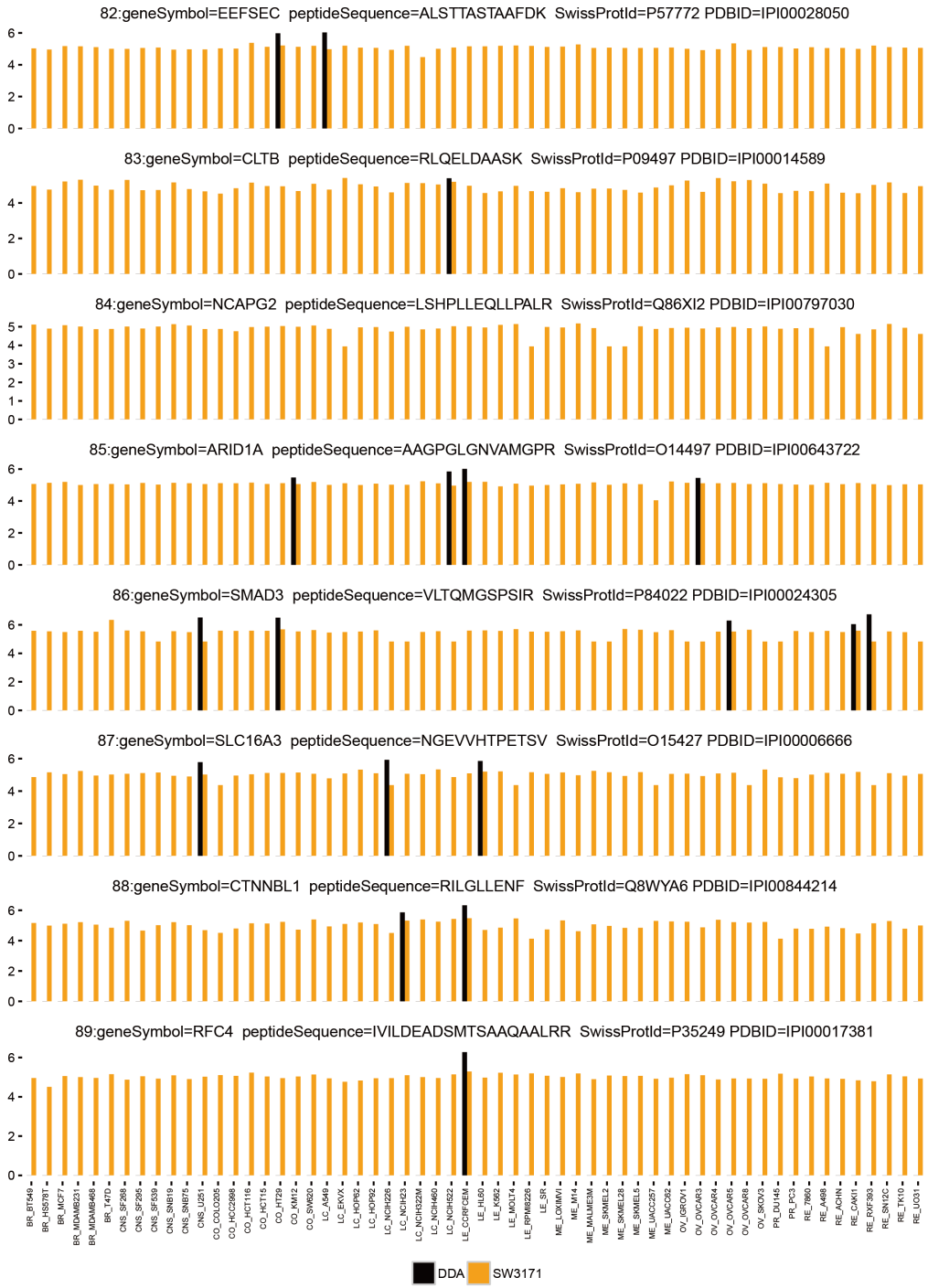
Supplementary Figure 15, Related to Figure 2. Bar plots for Q9Y5B9, Q436B4, P05204, Q9BQ61, Q8NB16, Q8TDB6, Q6UXH1, and Q69YN2, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



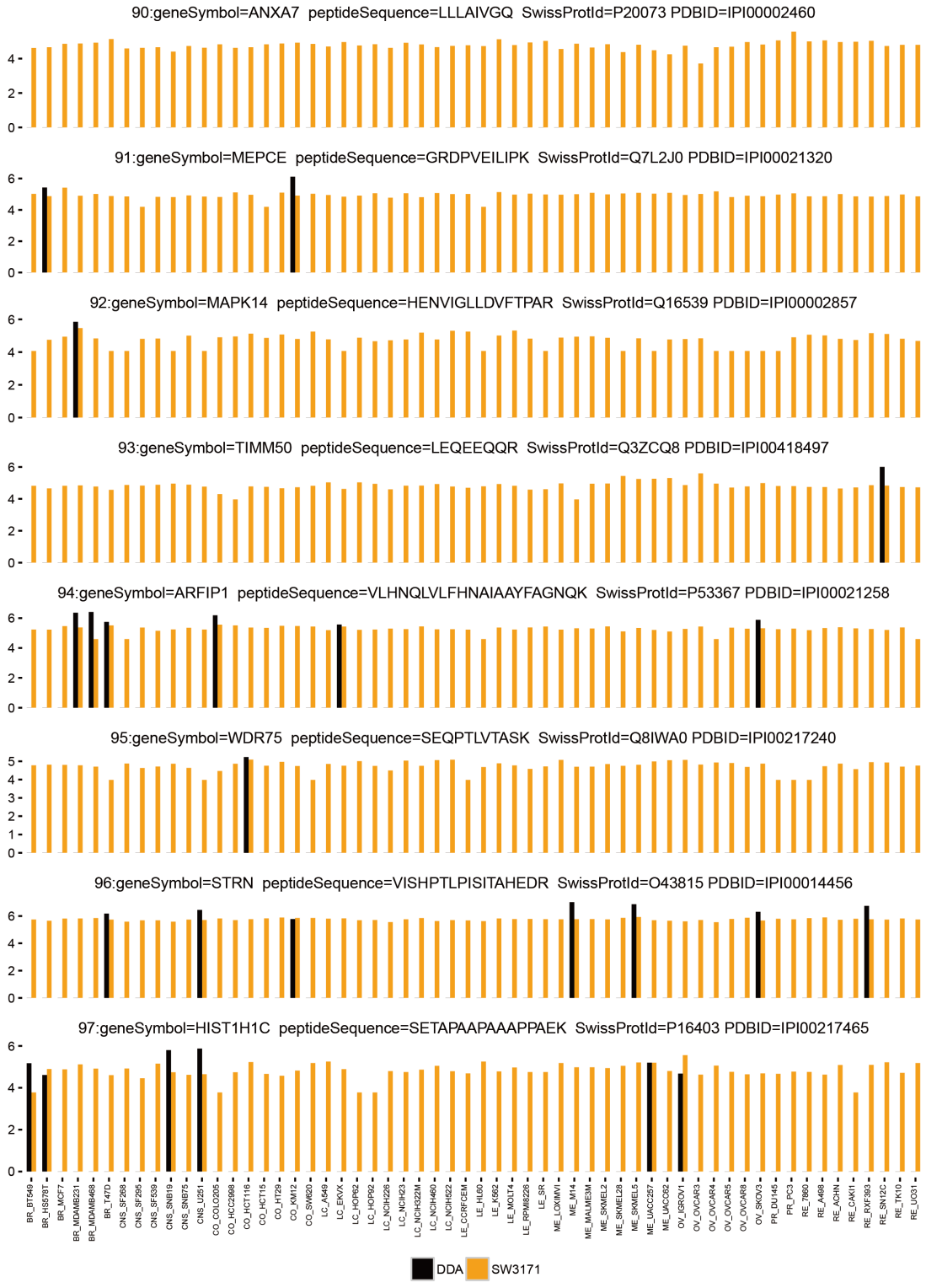
Supplementary Figure 16, Related to Figure 2. Bar plots for Q9NUQ8, Q96B26, O15460, Q9NVP1, P39D23, Q9Y3B4, Q12906, and P51858, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



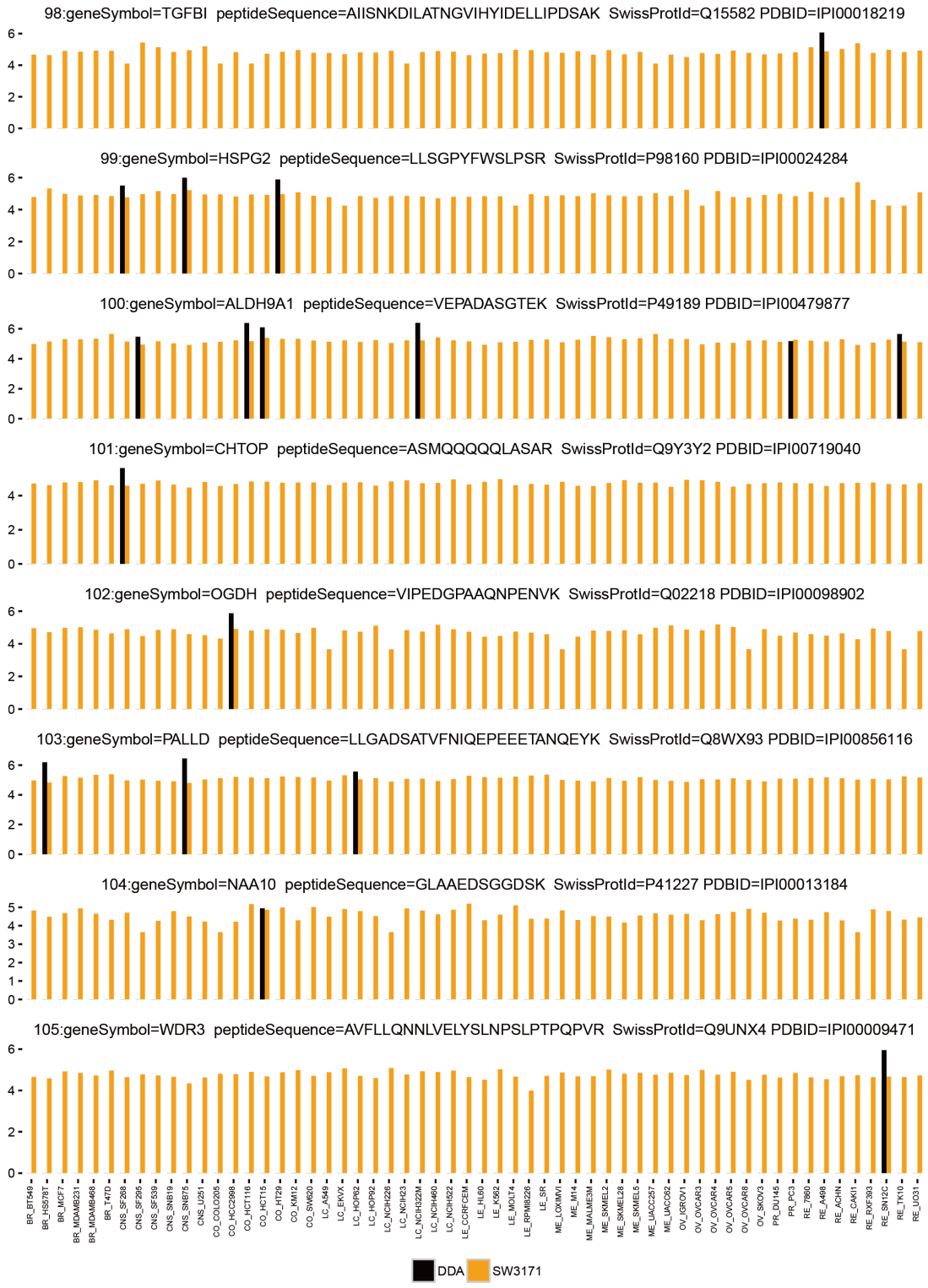
Supplementary Figure 17, Related to Figure 2. Bar plots for Bar plots for P24386, Q7Z3J2, Q9NVT9, Q00403, Q94905, Q9H6R4, O15066, and Q6RFH5, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



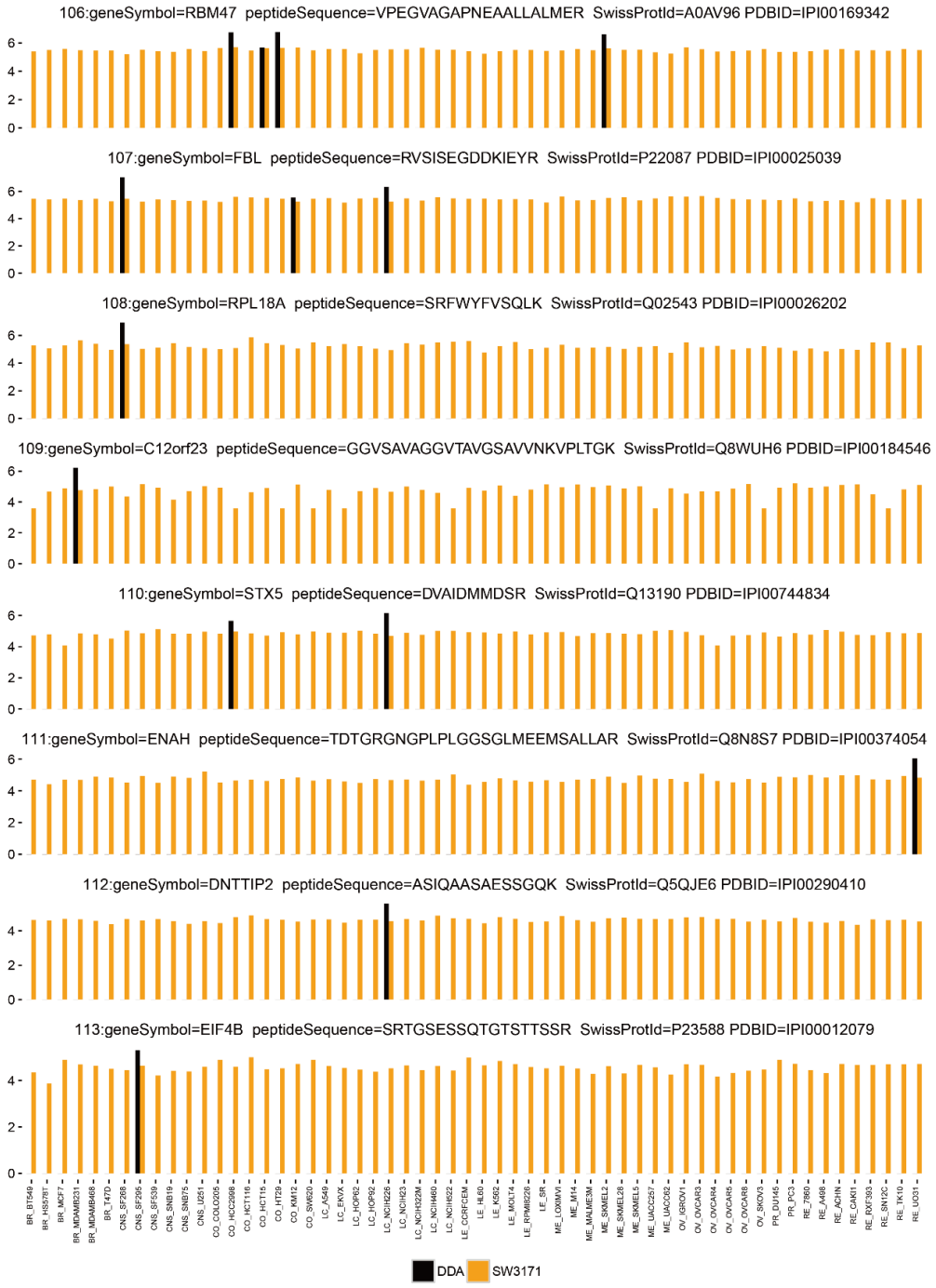
Supplementary Figure 18, Related to Figure 2. Bar plots for P57772, P09497, Q86X12, O14497, P84022, O15427, Q8WYA6, and P359249, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



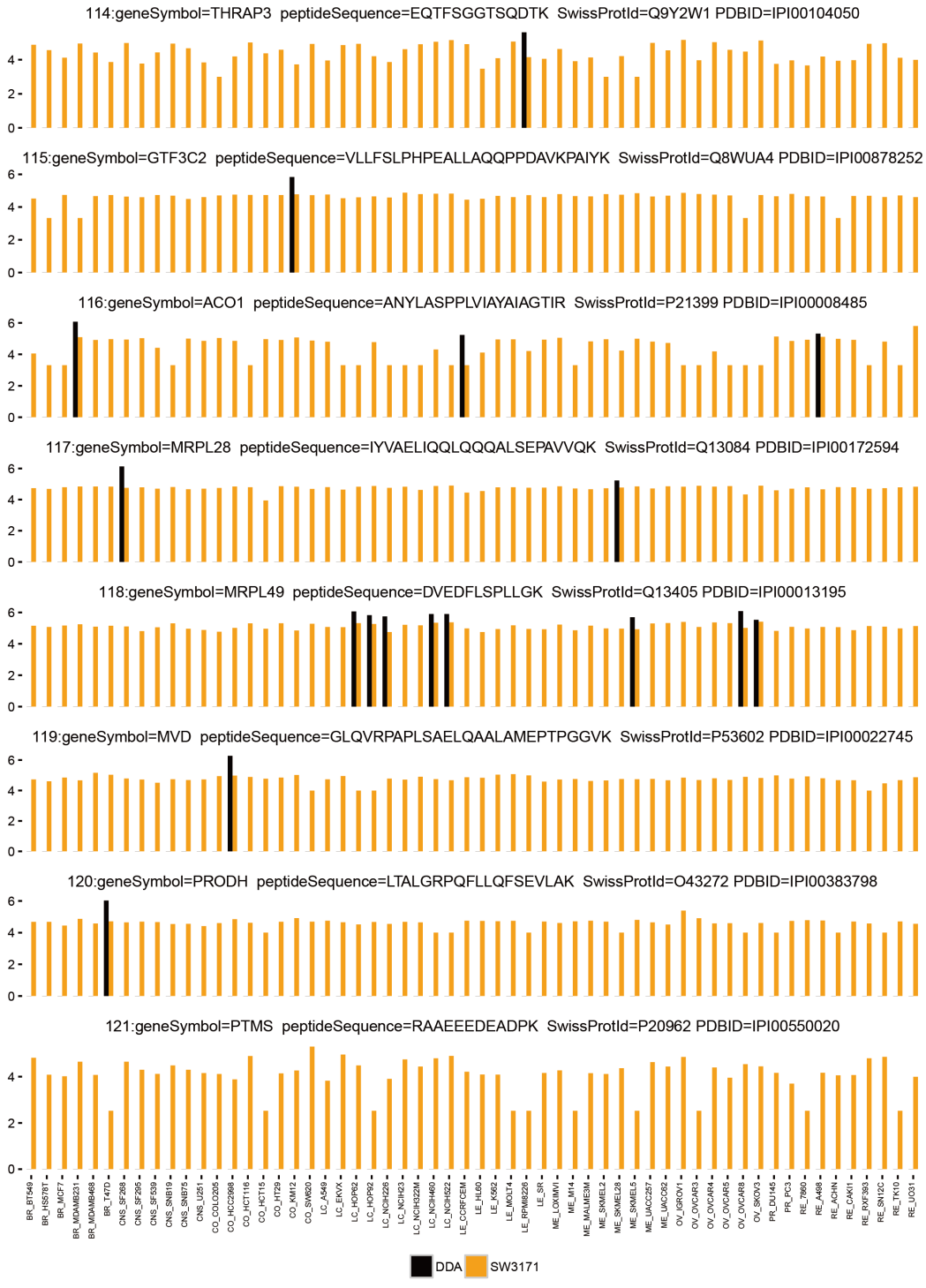
Supplementary Figure 19, Related to Figure 2. Bar plots for P20073, Q7L2J0, Q16539, Q3ZCQ8, P53367, Q8IWA0, 043815 and P16403, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



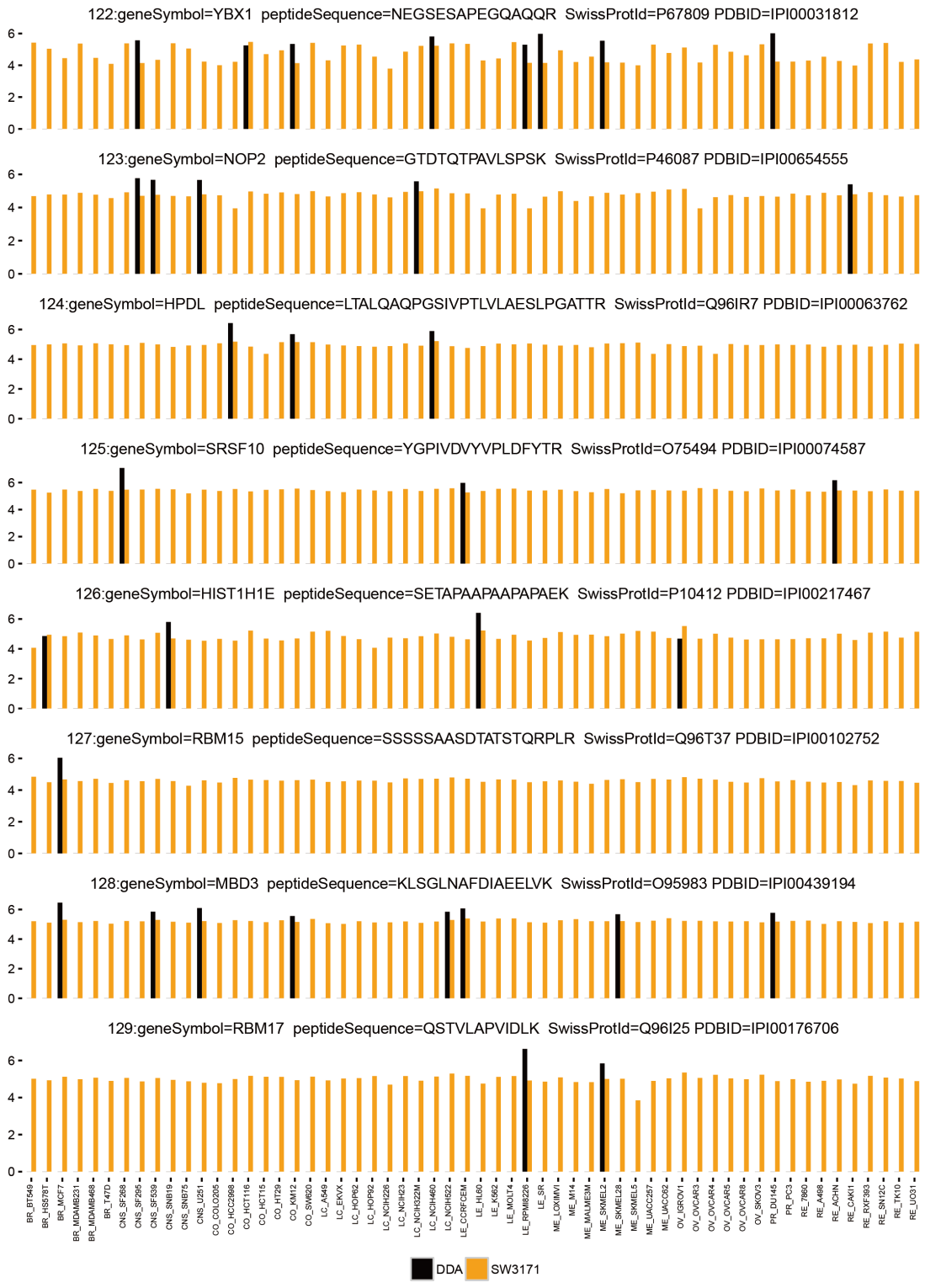
Supplementary Figure 20, Related to Figure 2. Bar plots for Q15582, P98160, P49189, Q9Y3Y2, Q02218, Q8WX93, P41227, and Q9UNX4, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



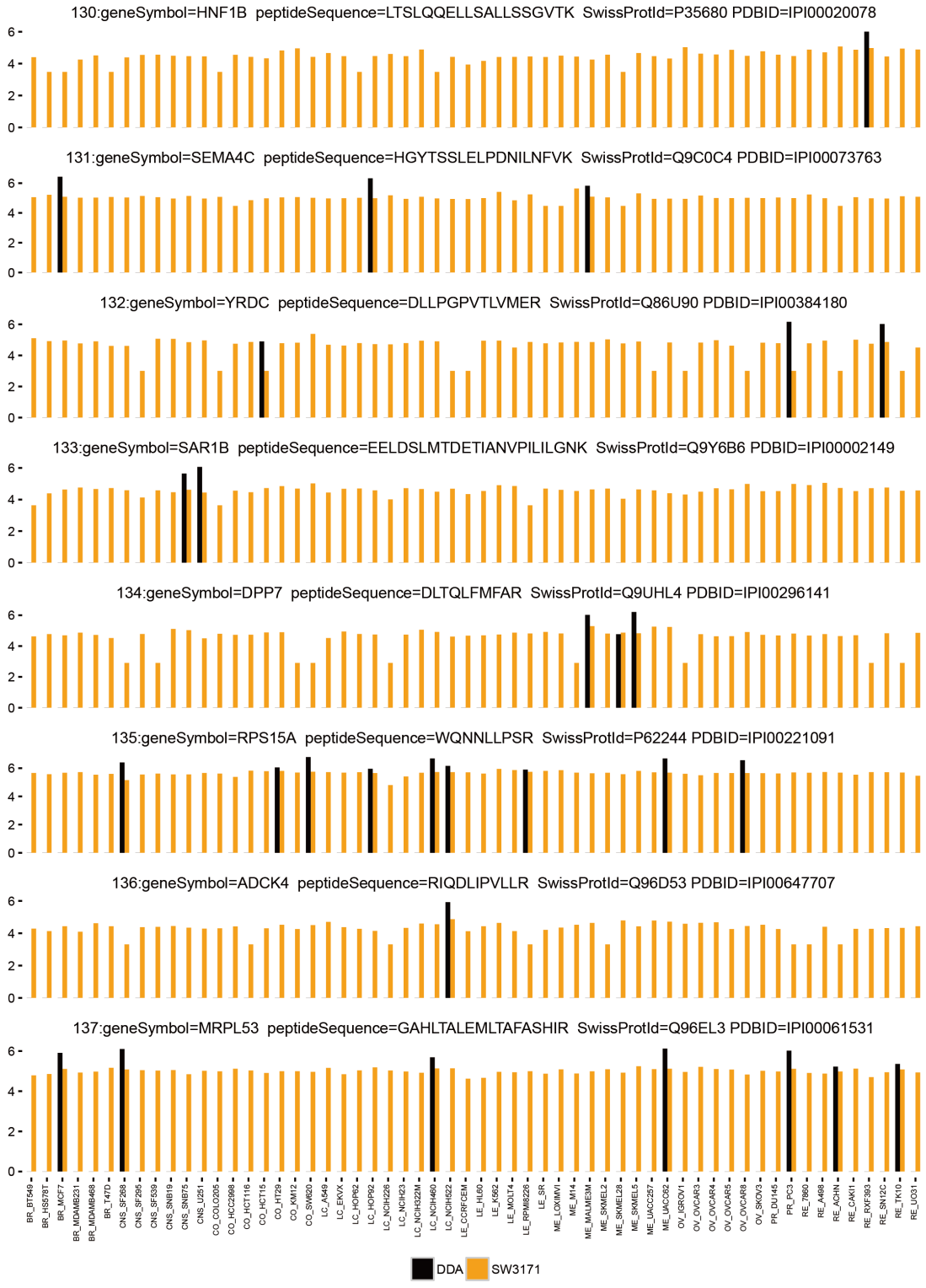
Supplementary Figure 21, Related to Figure 2. Bar plots for A0AV96, P22087, Q02543, Q8WUH6, Q13190, Q8N8S7, Q5QJE6 and P23588, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



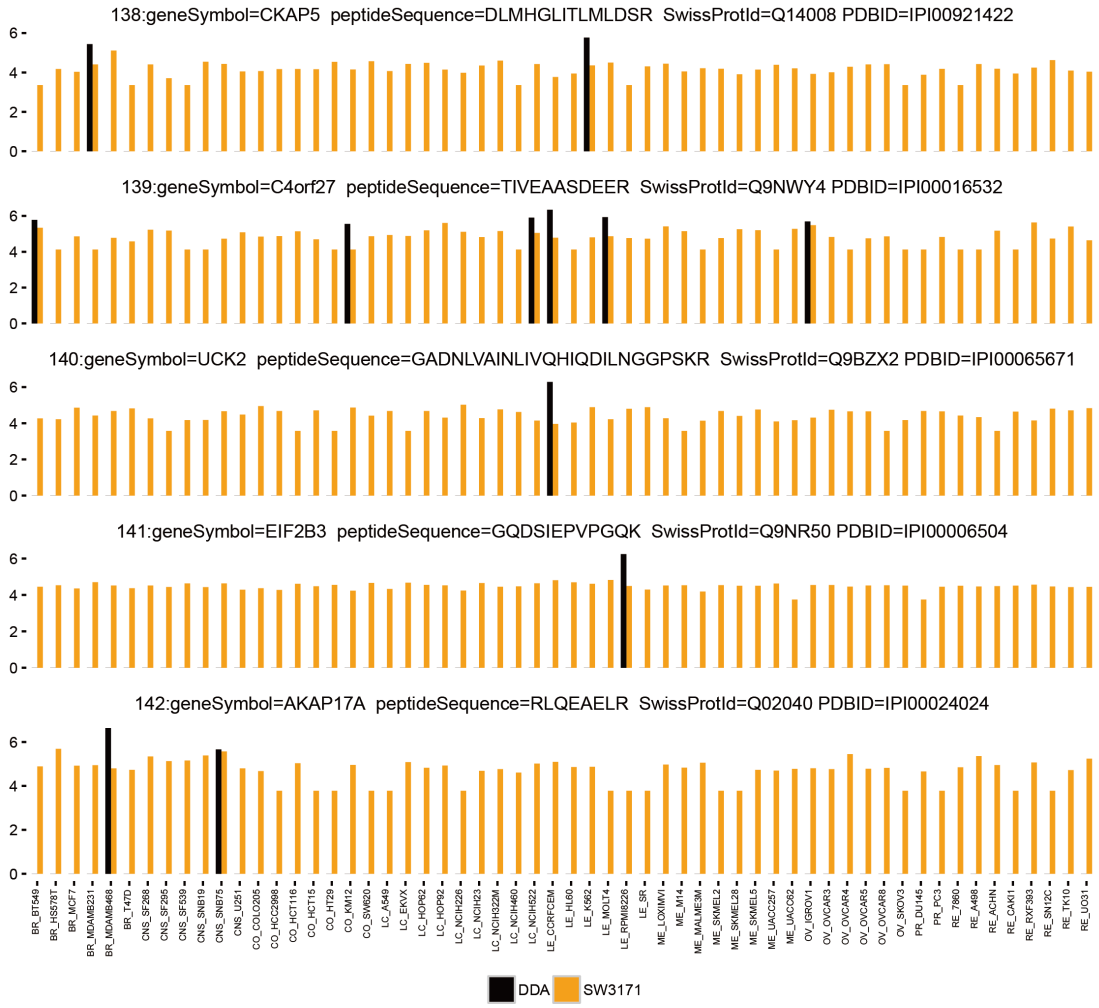
Supplementary Figure 22, Related to Figure 2. Bar plots for Q9Y2W1, Q8WUA4, P21399, Q13084, P53602, O43272 and P20962, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



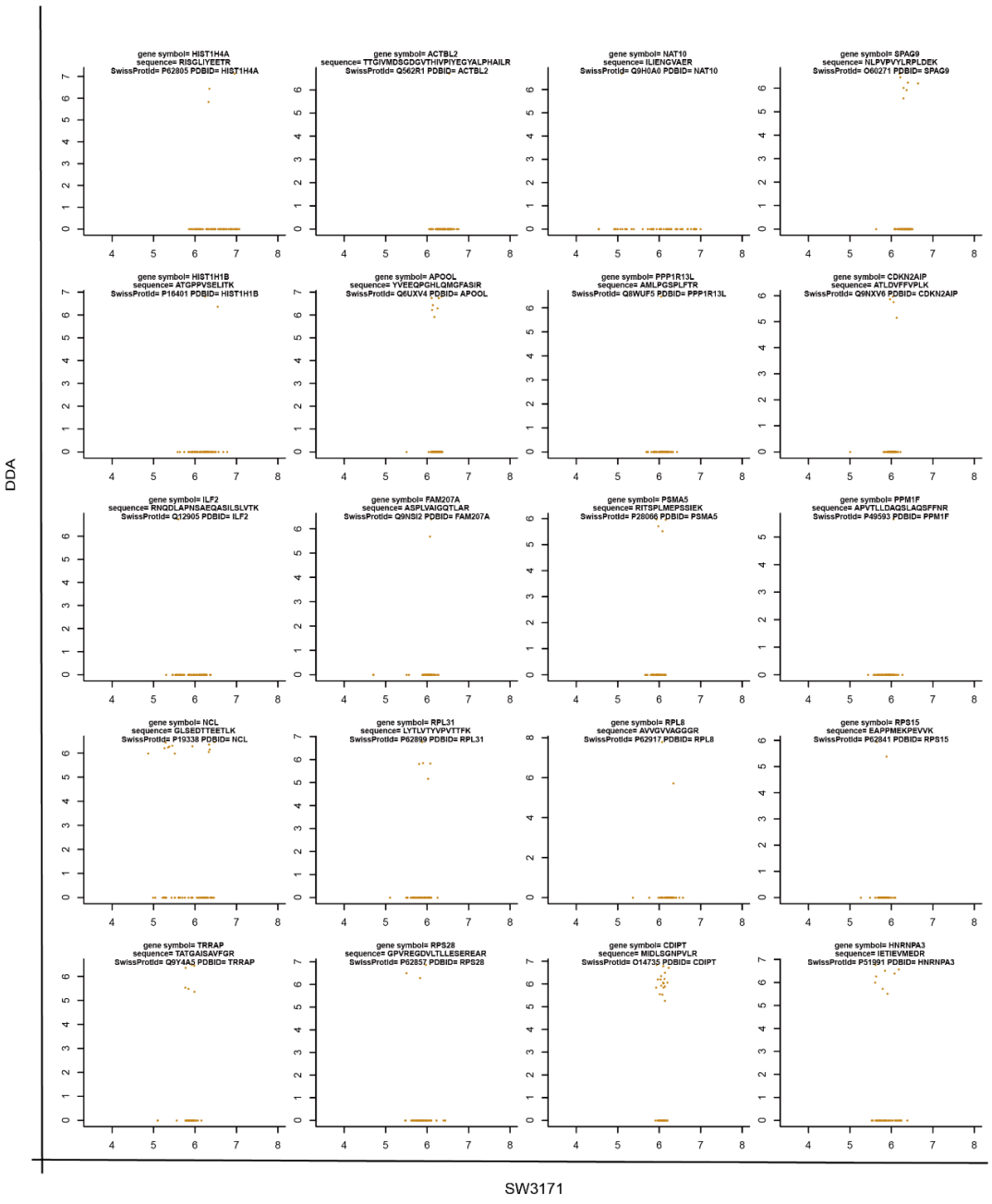
Supplementary Figure 23, Related to Figure 2. Bar plots for P67809, P46087, Q96IR7, O75494, P10412, Q96T37, O95983, and Q96I25, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



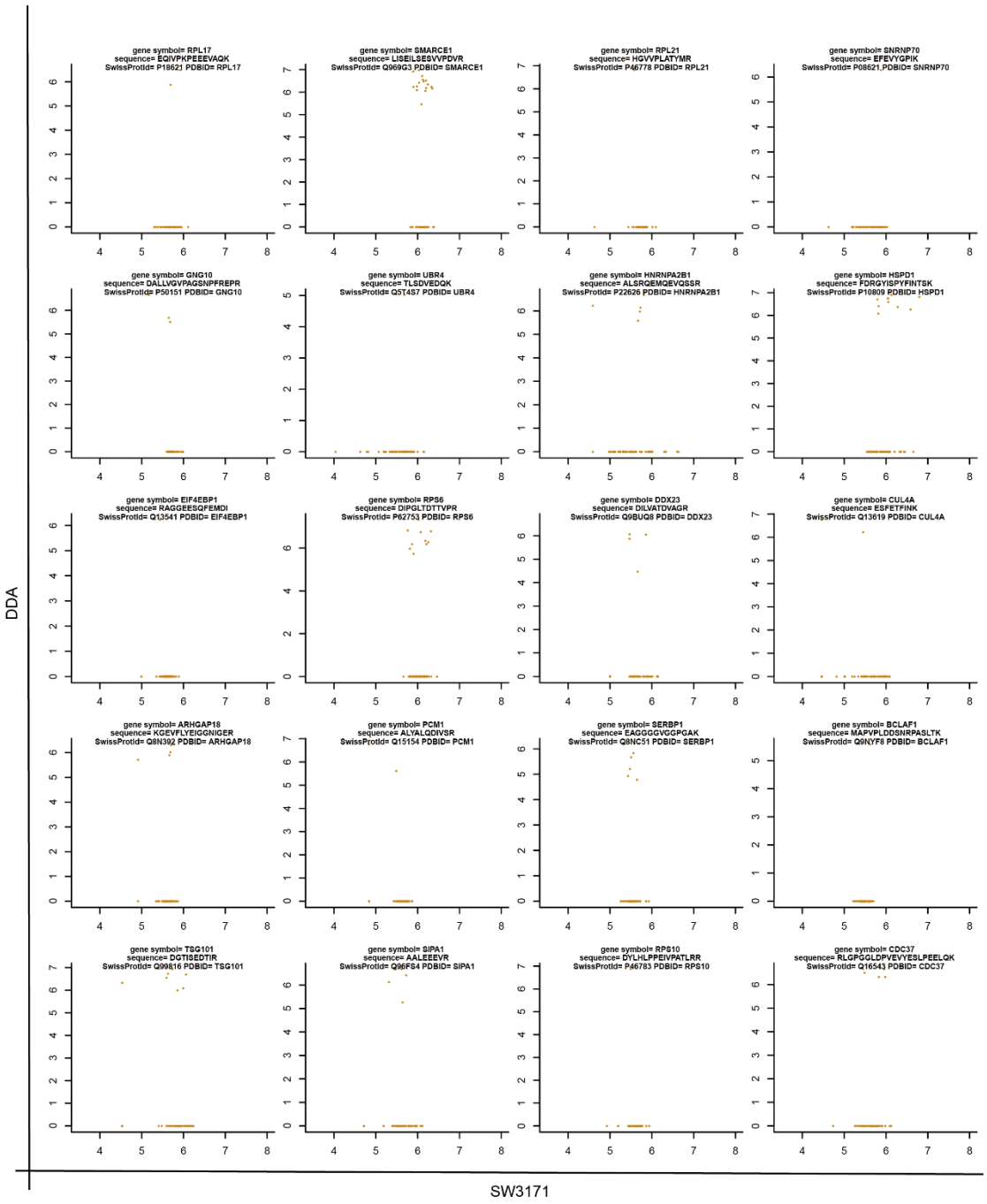
Supplementary Figure 24, Related to Figure 2. Bar plots for P35680, Q9C0C4, Q86U90, Q9Y6B6, Q9UHL4, P62244, Q96D53 and Q96EL3, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.



Supplementary Figure 25, Related to Figure 2. Bar plots for Q14008, Q9NWY4, Q9BZX2, Q9NR50, and Q02040, which are quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA.

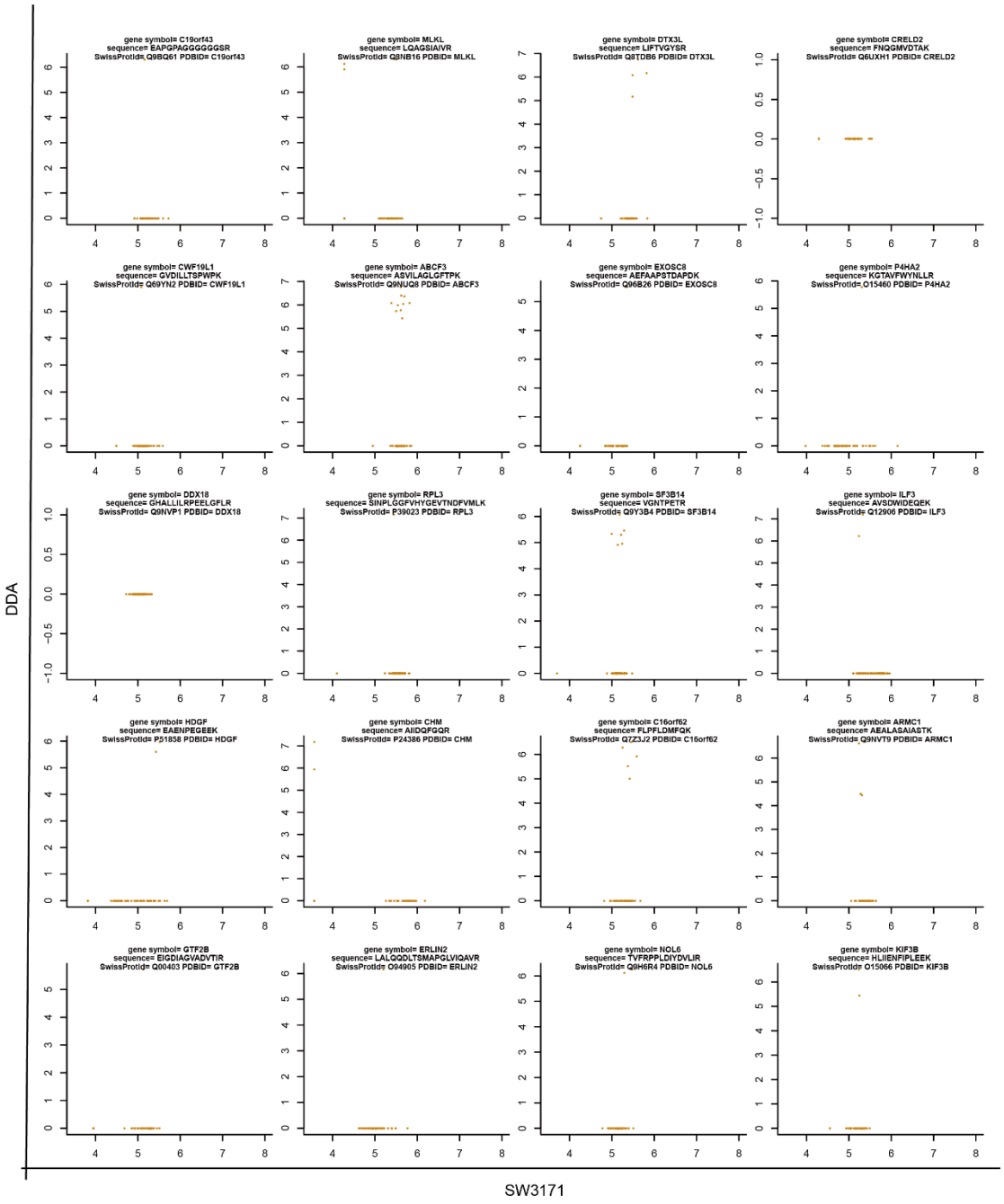


Supplementary Figure 26, Related to Figure 2. Scatter plots for P62805, Q562R1, Q9H0A0, O60271, P16401, Q6UXV4, Q8WUF5, Q9NXV6 , Q12905,Q9NSI2, P28066, P49593, P19338, P62899, P62917, P62841, , Q9Y4A5,P62875, O14735, and P51991 , which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.

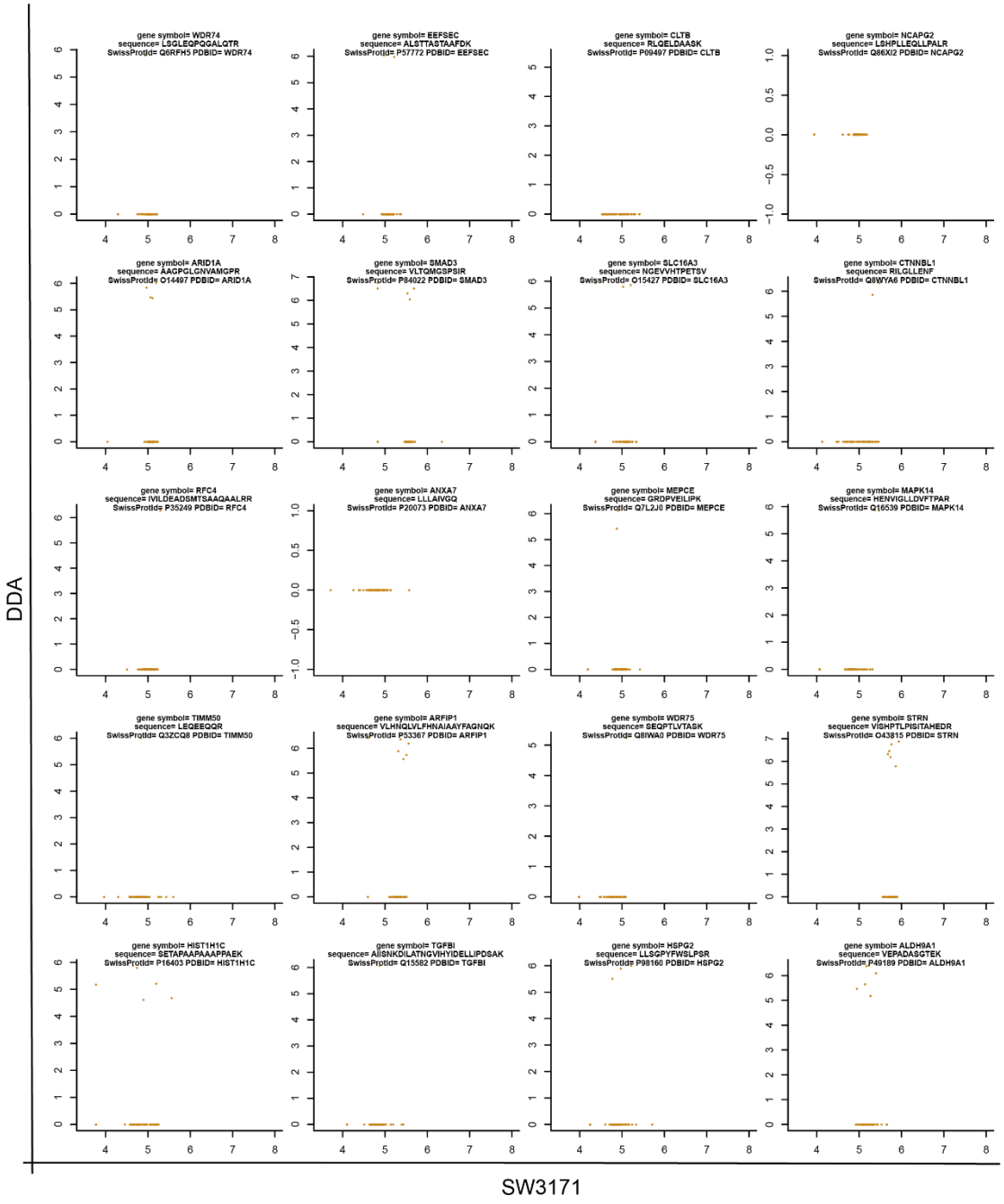


Supplementary Figure 27, Related to Figure 2. Scatter plots for P18621, Q969G3, P46778, P08621, P50151, Q5T4S7, P22626, P10809, Q13541, P62753, Q9BUQ8, Q13619 , Q8N392, Q15154, Q8NC51, Q9NYF8, Q99816, Q96FS4, P46783, and Q16543, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.

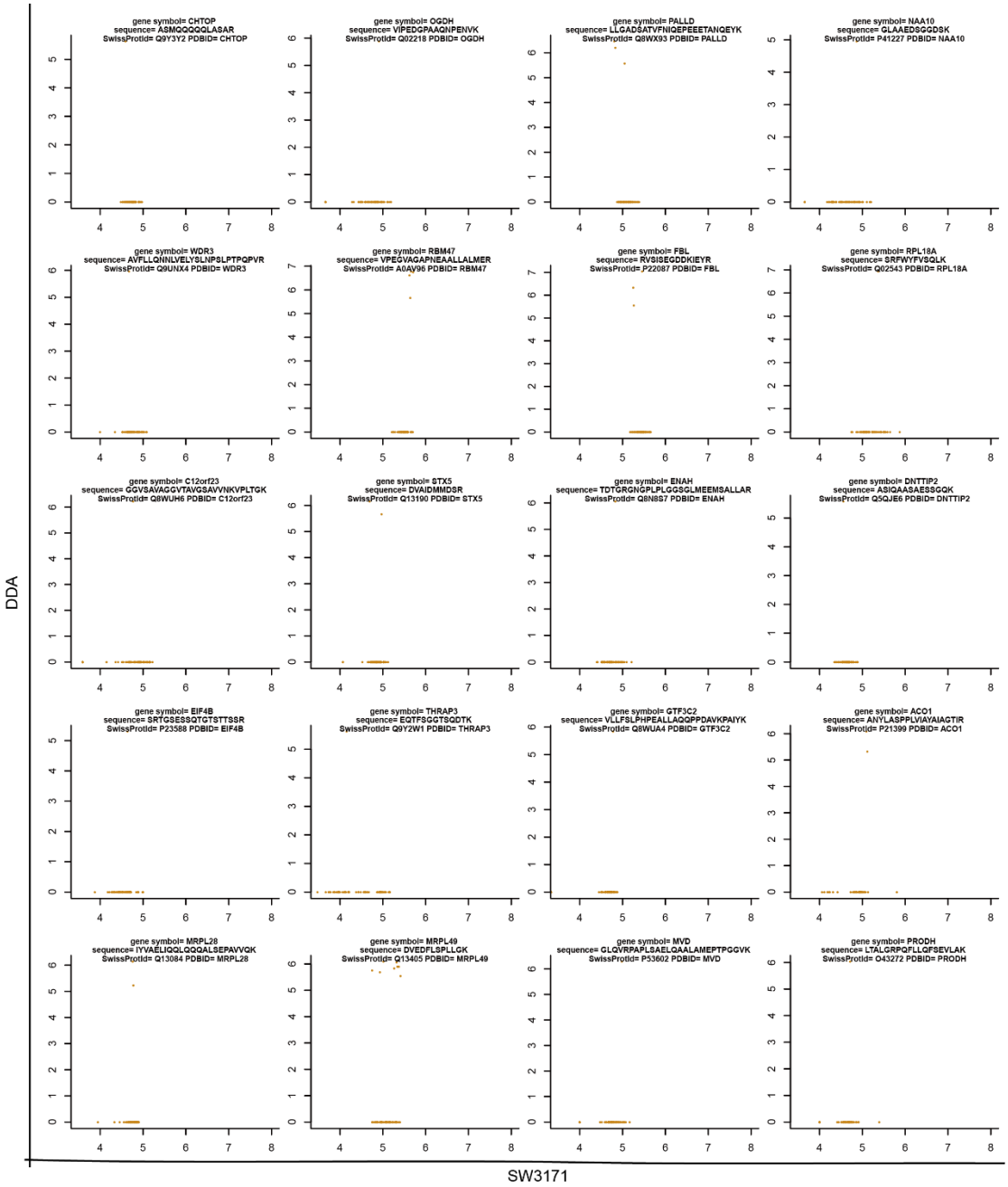
Supplementary Figure 28, Related to Figure 2. Scatter plots for O14964,Q8WWY3, Q9UJB3, O75170, Q9GZZ1, P56545, Q9BVL2, P61313, Q9H4A3,Q9NPE3, Q06023, Q92576, Q969S3, Q9H089, Q13435, Q8N5N7, Q9Y5B9, Q436B4, and P05204, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.



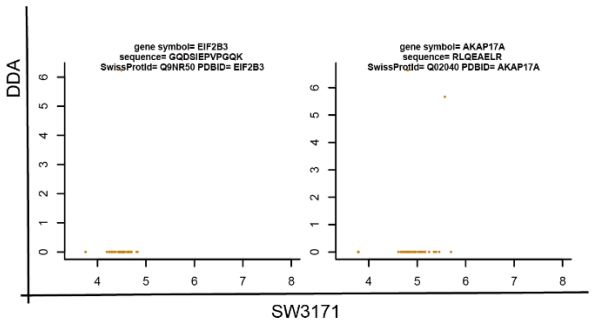
Supplementary Figure 29, Related to Figure 2. Scatter plots for Q9BQ61, Q8NB16, Q8TDB6, Q6UXH1, Q69YN2, Q9NUQ8, Q96B26, O15460, Q9NVP1, P39D23, Q9Y3B4, Q12906, , P51858, P24386, Q7Z3J2, Q9NVT9, Q00403, Q94905, Q9H6R4, and O15066, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.



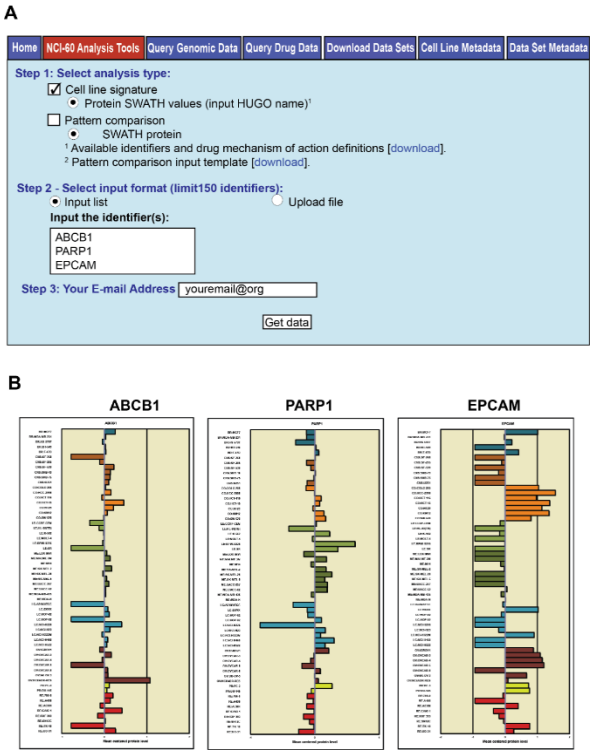
Supplementary Figure 30, Related to Figure 2. Scatter plots for Q6RFH5, P57772, P09497, Q86X12, O14497, P84022, O15427, Q8WYA6, P359249, P20073, Q7L2J0, Q16539, Q3ZCQ8, P53367, Q8IWA0, 043815 , P16403,Q15582, P98160, and P49189, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.



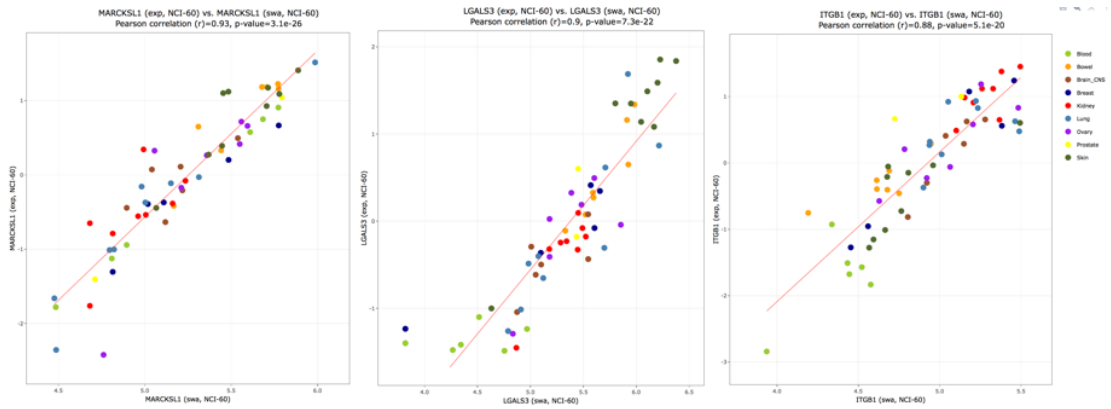
Supplementary Figure 31, Related to Figure 2. Scatter plots for Q9Y3Y2, Q02218, Q8WX93, P41227, Q9UNX4, A0AV96, P22087, Q02543, Q8WUH6, Q13190, Q8N8S7, Q5QJE6 , P23588,Q9Y2W1, Q8WUA4, P21399,Q13084, P53602, and O43272, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.



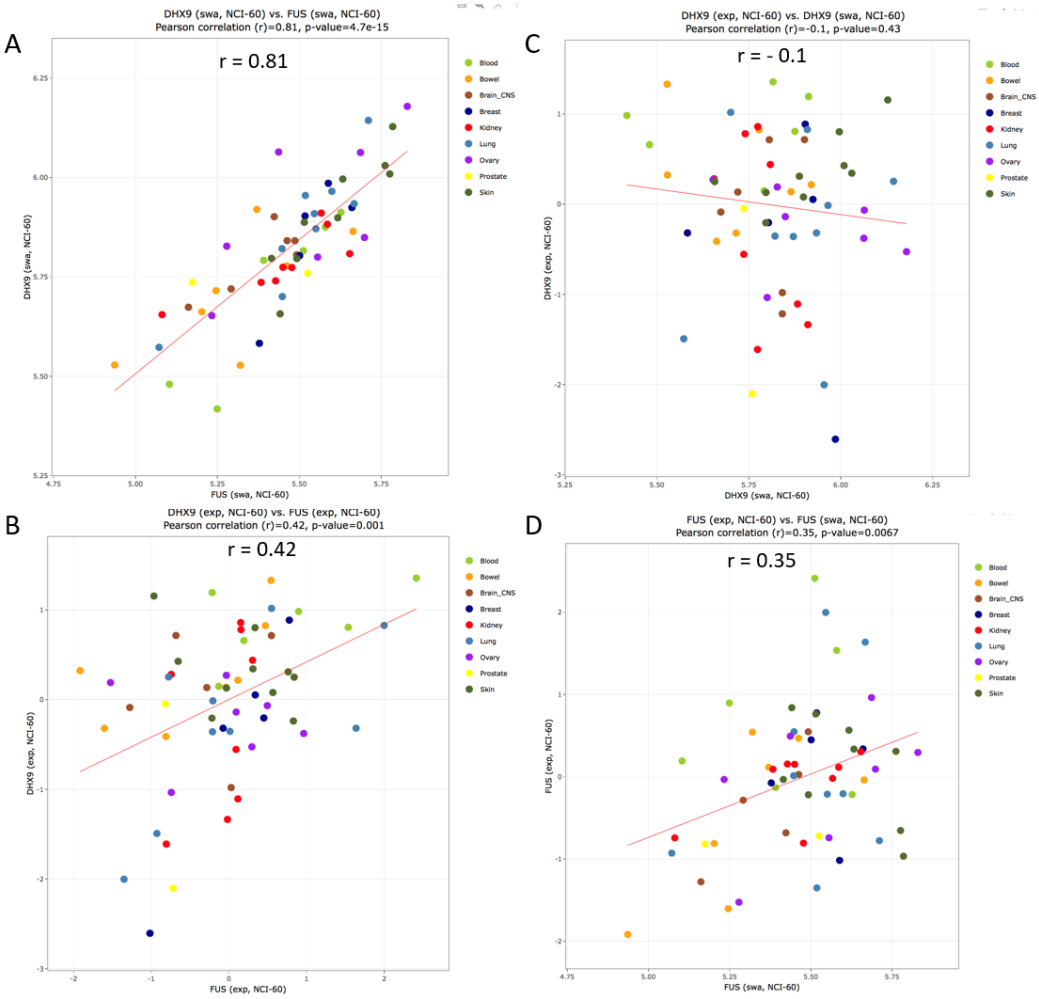
Supplementary Figure 32, Related to Figure 2. Scatter plots for P20962,P67809, P46087, Q96IR7, O75494, P10412, Q96T37, O95983, Q96I25,P35680, Q9C0C4, Q86U90, Q9Y6B6, Q9UHL4, P62244, Q96D53, Q96EL3, Q14008, Q9NWY4, Q9BZX2, Q9NR50, and Q02040, which are all quantified cross all NCI-60 cell lines by SWATH but not quantified by DDA. This figure shows the data completeness difference of the two data sets.



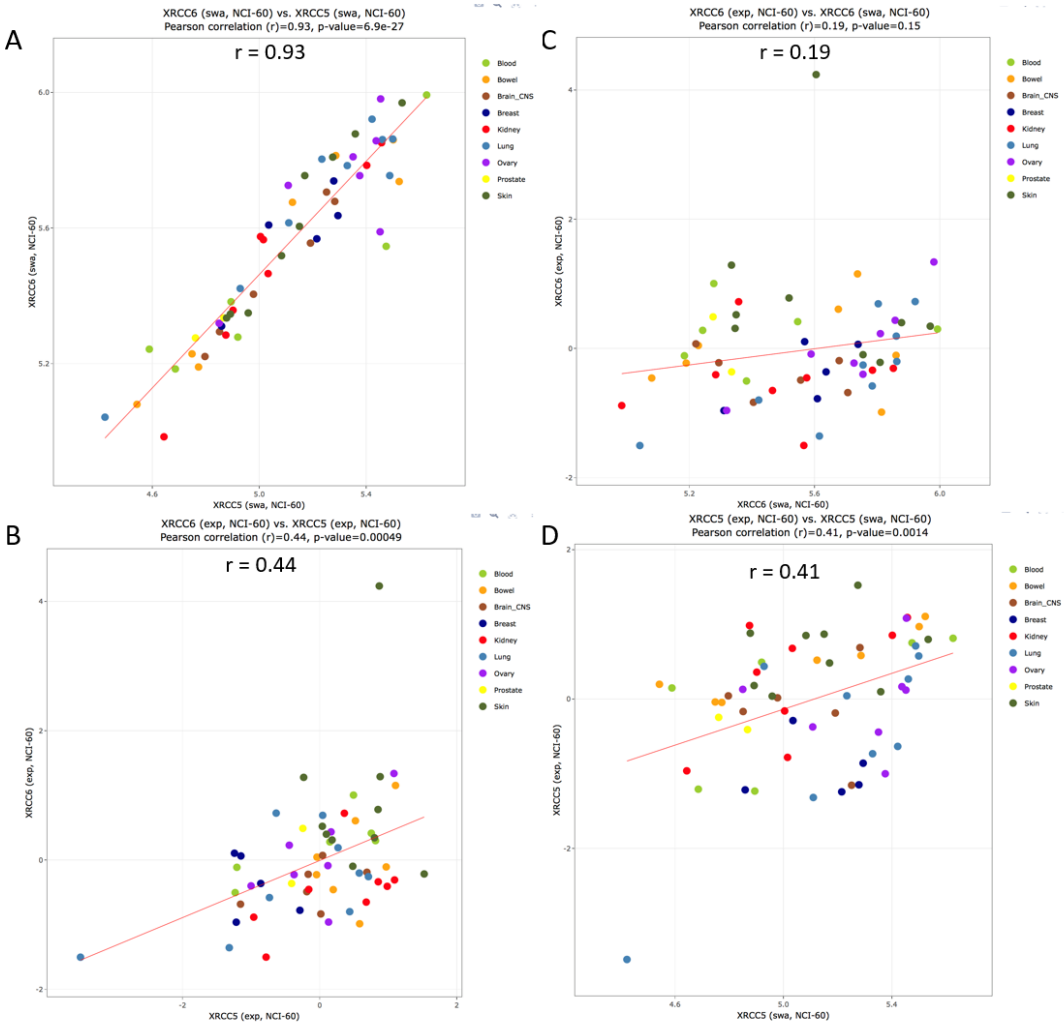
Supplementary Figure 33, Related to Figure 2. Access to NCI-60 proteotype in Cellminer. To facilitate data access, visualization, and comparison with other forms of genomic and pharmacological data for the NCI-60 cancer cell lines, we have incorporated the SWATH data within CellMiner 4 5. The CellMiner web site allows the data to be retrieved or used in several ways 6. (**A**) The "Download Data Sets" tab allows either the total 3,171 proteins, or the 22,554 peptides data sets to be downloaded. This data will primarily be of use in computational biology pipelines. The "Query Genomic Data" tab allows up to 150 proteins or peptides to be accessed (using the "Gene" or "Peptide" pull downs), queryable by gene name or peptide peak identifier, chromosomal or genomic location. Data is sent in both Excel (.xls) and text (.txt) format. The "NCI-60 Analysis Tools" tab (**A**) provides "Cell line signatures". To obtain "Cell line signatures" for genes, select "Cell line signature" in Step 1, and then "Protein SWATH values". In Step 2, up to 150 genes of interest may be input by either typing in the gene names in the "Input the identifier" box, or uploading them as a text or Excel file using the "Upload file" radio button. In Step 3, enter your e-mail address, and click "Get data". Results will be sent by e-mail for each gene, with a link to download the results. This file contains three worksheets: i) tabular mean centered protein levels ratios as a both a bar plot and tabular data, and the peptide peak information for that gene ii) "Bin protein levels" with a histogram of the protein levels and iii) and "Footnotes". (**B**) provides examples of three genes of interest. These "Cell line signatures" can also be used as input for the Pattern Comparison tool (also within the NCI-60 Analysis tools section) which provides correlated molecular and compound activity data. All available gene and peptide identifiers are available as a list within the "Available identifiers and drug mechanism of action definitions" as a download within the "NCI-60 Analysis Tools" tab.



Supplementary Figure 34, Related to Figure 2. CellminerCDB snapshot views of three genes with highest correlation between expression in SWATH and transcriptome data: Myristyolated Alanine-Rich C Kinase (MARCKSL1), Galectin 3 (LGALS3) and Integrin β1 (ITGB1) (see Supplementary Table 3). Data are snapshots from https://discover.nci.nih.gov/CellMinerCDB.

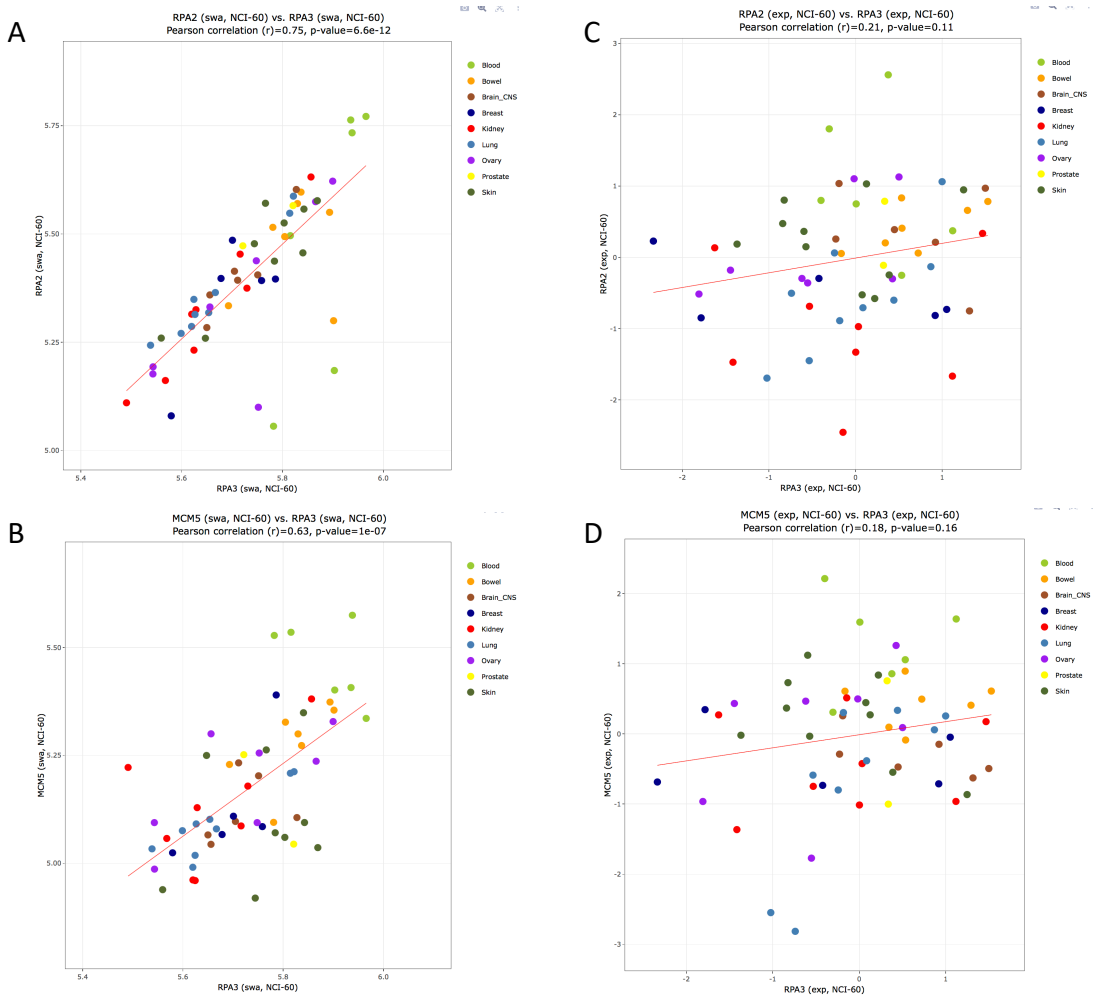


Supplementary Figure 35, Related to Figure 2 and 4. CellminerCDB view of the stoichiometric relationship of the expression of two RNA binding proteins DHX9 (RNAse A) and FUS (Fused in Sarcoma) in SWATH and transcriptome data. https://discover.nci.nih.gov/CellMinerCDB snapshots showing: **A.** The high stoichiometric correlation for both DHX9 and FUS across the NCI-60. **B.** The lower stoichiometric relationship between DHX9 and FUS transcripts. **C.** The lack of correlation between DHX9 protein and transcripts across the NCI-60. **D.** The low stoichiometric relationship between FUS protein and transcripts across the NCI-60.



Supplementary Figure 36, Related to Figure 2. CellminerCDB view of XRCC5 and XRCC6 expression in SWATH and transcriptome data.

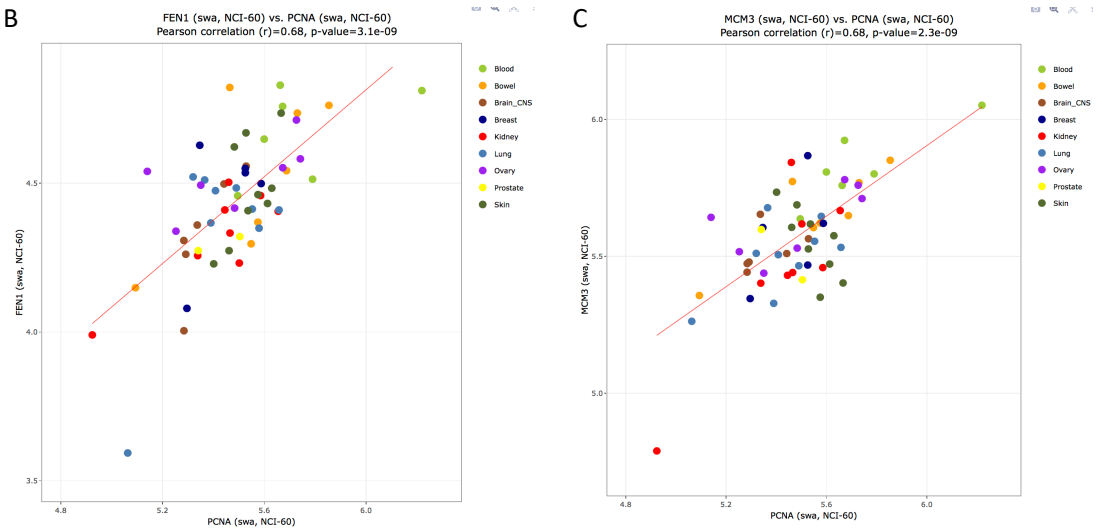
https://discover.nci.nih.gov/CellMinerCDB snapshots showing: **A.** The high stoichiometric correlation for both Ku subunits XRCC6 (KU70) and XRCC5 (KU80) across the NCI-60. **B.** The lower stoichiometric relationship between XRCC6 and XRCC5 transcripts. **C.** The lack of correlation between XRCC6 protein and transcripts across the NCI-60. **D.** The lower stoichiometric relationship between XRCC5 protein and transcripts across the NCI-60.



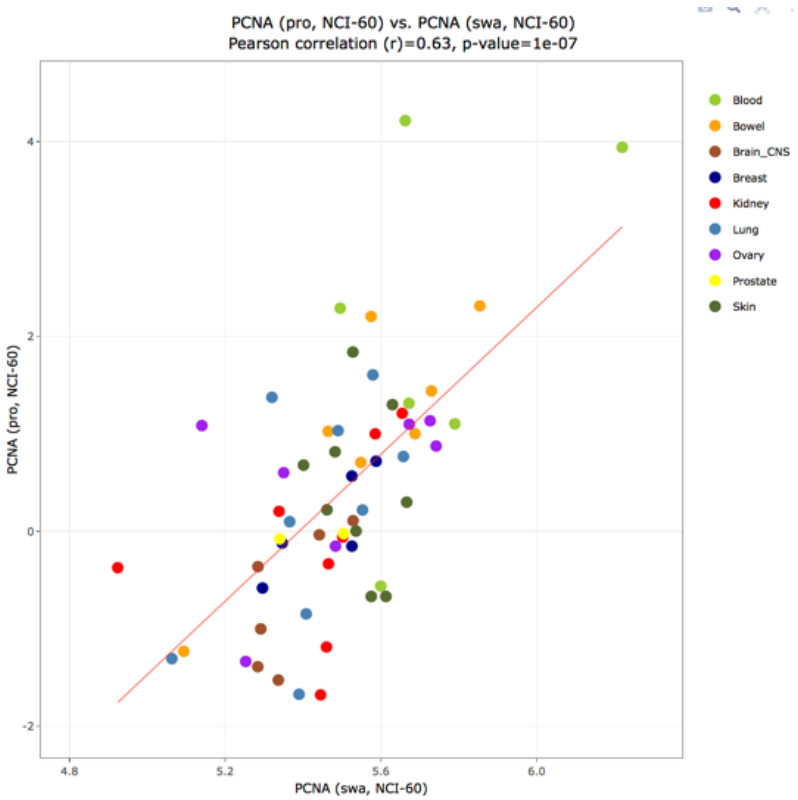
Supplementary Figure 37, Related to Figure 2. CellminerCDB snapshot views showing co-expression of replication proteins: RPA3 with RPA2 (A) and MCM5 (C) whereas transcripts do not show significant correlations (C-D). Data are snapshots from https://discover.nci.nih.gov/CellMinerCDB.

A

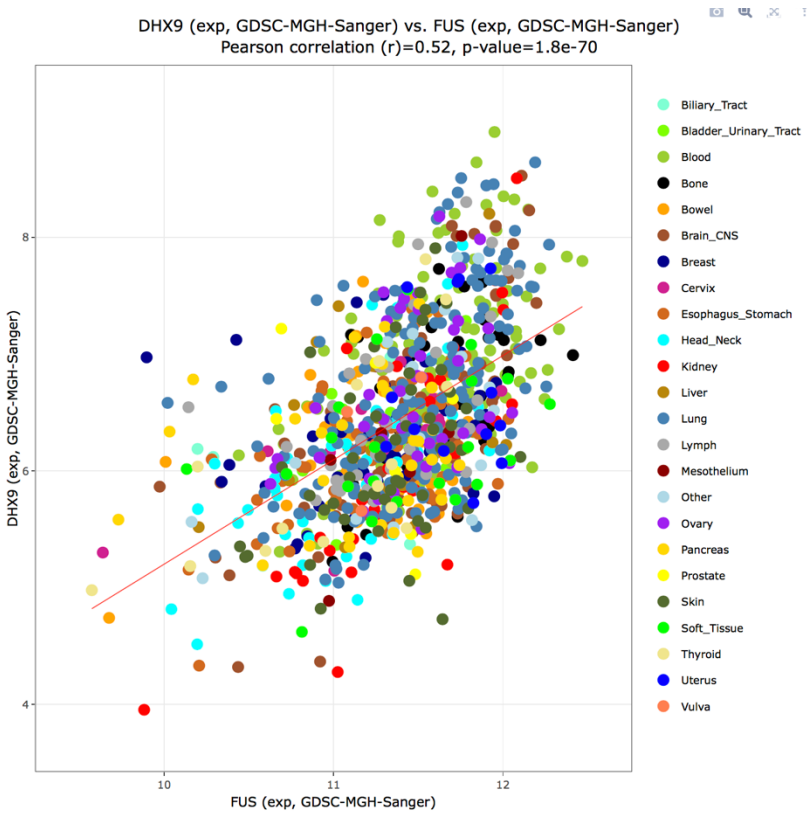
SWA	All	All	All	All	All	All
swa	PCNA	20pter-p12	1	0	0	DNA Damage Response (DDR); DDR (BER); DDR (DNA replication)
swa	MCM3	6p12	0.684	2.3e-9	0.0000904	DDR (DNA replication)
swa	FEN1	11q12	0.68	3.15e-9	0.0000904	DNA Damage Response (DDR); DDR (BER); DDR (DNA replication)
swa	MTHFD1	14q24	0.661	1.19e-8	0.00025	water-soluble vitamin metabolic process;cellular amino acid biosynthetic process
swa	ATIC	2q35	0.65	2.5e-8	0.000448	purine ribonucleoside monophosphate biosynthetic process;organ regeneration



Supplementary Figure 38, Related to Figure 2. CellminerCDB snapshots showing co-expression of replication proteins determined by SWATH and detailing the coexpression of FEN1 with PCNA and of PCNA with the replication helicase protein MCM3. A. The "Compare pattern" tool was used with PCNA as the "x-axis entry". Snapshots from https://discover.nci.nih.gov/CellMinerCDB showing only the top correlates with RPA3 including MCM3 and FEN1. **B.** Stoichiometric relationship between PCNA and FEN1 proteins across the NCI-60. **C.** Stoichiometric relationship between PCNA and MCM3 proteins across the NCI-60.



Supplementary Figure 39, Related to Figure 2. CellminerCDB snapshot showing reproducible expression of PCNA determined by SWATH and RPPA (Reverse Phase Protein Array). The snapshot from https://discover.nci.nih.gov/CellMinerCDB shows PCNA protein levels across the NCI-60.

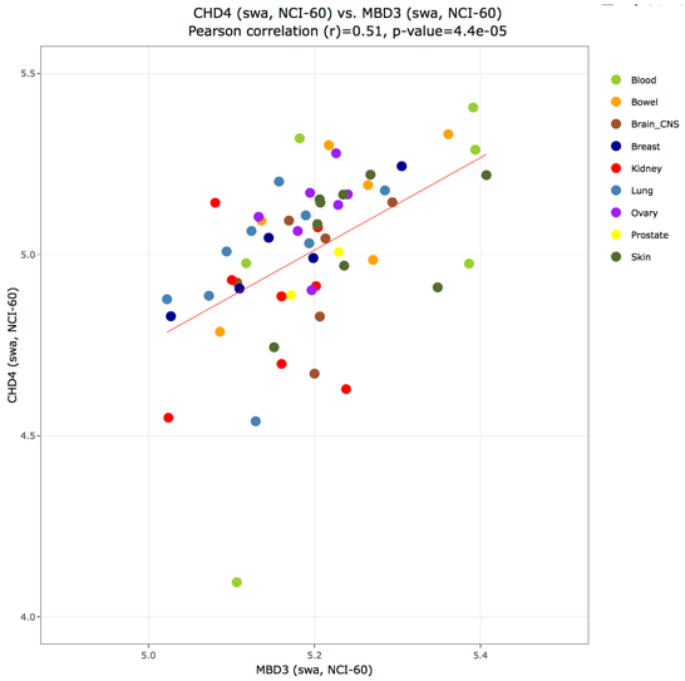


Supplementary Figure 40, Related to Figure 2 and 4. CellminerCDB snapshot of FUS and DHX9 transcript expression (https://discover.nci.nih.gov/CellMinerCDB) **across the MGH-Sanger cell lines**.

SWATH											
Pearsons correlation											
Identifier	RBBP7	RBBP4	MTA3	MTA1	HDAC2	HDAC1	GATAD2B	GATAD2A	MBD3	CHD4	ZMYND8
RBBP7	1	0.226	-0.031	0.064	-0.01	0.197	0.127	-0.053	-0.019	0.12	-0.136
RBBP4	0.178	1	-0.091	0.245	0.271	0.539	0.073	0.18	0.168	0.343	-0.063
MTA3	-0.245	-0.099	1	0.325	0.039	0.076	-0.036	-0.254	-0.159	-0.083	0.016
MTA1	0.228	0.047	-0.173	1	0.457	0.214	0.147	-0.017	0.125	0.106	-0.364
HDAC2	0.133	0.026	-0.043	**0.518**	1	0.419	-0.067	0.307	0.286	0.351	-0.118
HDAC1	**0.391**	**0.497**	-0.127	0.189	-0.004	1	0.05	0.36	0.049	0.497	0.027
GATAD2B	0.24	0.29	-0.088	0.226	0.219	0.271	1	0.218	-0.069	0.366	0.047
GATAD2A	0.274	0.179	0.061	**0.306**	0.277	0.294	0.095	1	0.248	0.451	0.297
MBD3	**0.316**	**0.447**	-0.086	**0.443**	**0.313**	**0.434**	**0.438**	**0.421**	1	0.171	-0.084
CHD4	0.269	0.238	-0.201	**0.311**	0.138	**0.382**	**0.402**	**0.419**	**0.506**	1	0.316
ZMYND8	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	1

Transcripts											
Pearsons correlation											
Identifier	RBBP7	RBBP4	MTA3	MTA1	HDAC2	HDAC1	GATAD2B	GATAD2A	MBD3	CHD4	ZMYND8
RBBP7	1	0.226	-0.031	0.064	-0.01	0.197	0.127	-0.053	-0.019	0.12	-0.136
RBBP4	0.226	1	-0.091	0.245	0.271	0.539	0.073	0.18	0.168	0.343	-0.063
MTA3	-0.031	-0.091	1	0.325	0.039	0.076	-0.036	-0.254	-0.159	-0.083	0.016
MTA1	0.064	0.245	**0.325**	1	0.457	0.214	0.147	-0.017	0.125	0.106	-0.364
HDAC2	-0.01	0.271	0.039	**0.457**	1	0.419	-0.067	0.307	0.286	0.351	-0.118
HDAC1	0.197	**0.539**	0.076	0.214	**0.419**	1	0.05	0.36	0.049	0.497	0.027
GATAD2B	0.127	0.073	-0.036	0.147	-0.067	0.05	1	0.218	-0.069	0.366	0.047
GATAD2A	-0.053	0.18	-0.254	-0.017	0.307	**0.36**	0.218	1	0.248	0.451	0.297
MBD3	-0.019	0.168	-0.159	0.125	0.286	0.049	-0.069	0.248	1	0.171	-0.084
CHD4	0.12	**0.343**	-0.083	0.106	**0.351**	**0.497**	**0.366**	**0.451**	0.171	1	0.316
ZMYND8	-0.136	-0.063	0.016	-0.364	-0.118	0.027	0.047	0.297	-0.084	**0.316**	1

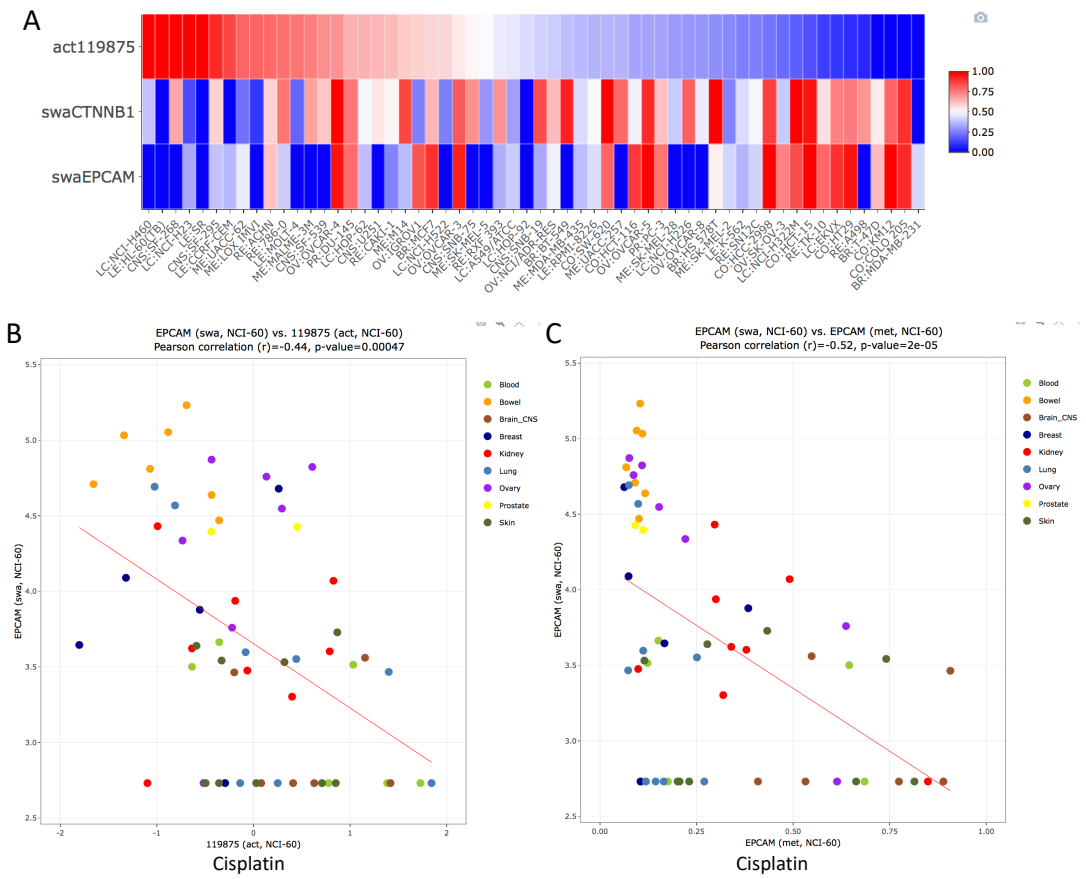
Supplementary Figure 41, Related to Figure 2. Snapshot views showing co-expression of the NuRF (Nucleosome Remodeling Factors) proteins determined by SWATH (top) and transcripts (bottom). The "Cross-correlations" tool of CellMiner was used with the listed proteins or genes (left column). Snapshots from the Excel files obtained from http://discover.nci.nih.gov/cellminer.



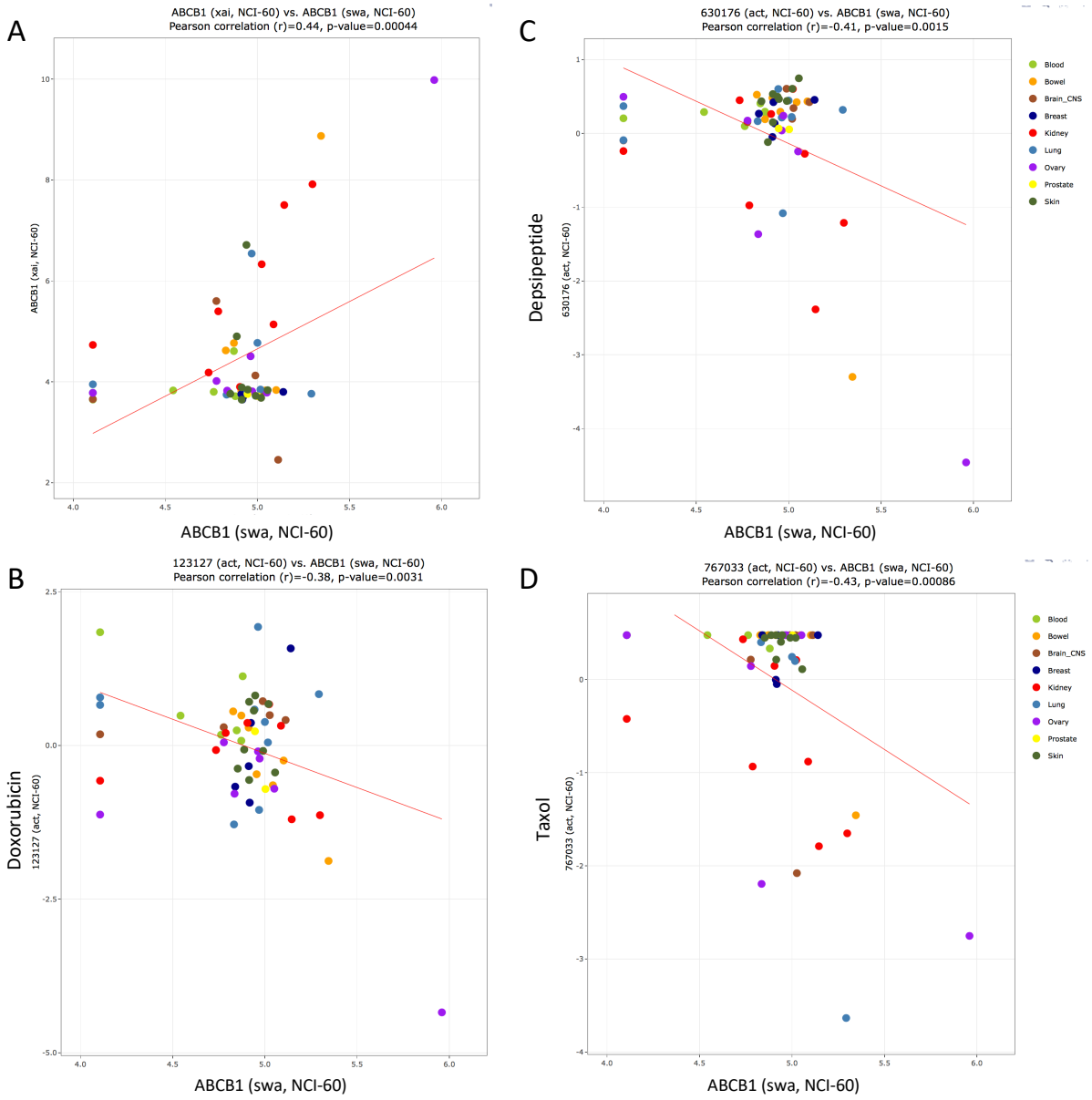
Supplementary Figure 42, Related to Figure 2. CellminerCDB snapshot showing stoichiometry expression of MBD3 and CHD4 determined by SWATH. The https://discover.nci.nih.gov/CellMinerCDB snapshot show MBD3 and CHD4 protein levels across the NCI-60.

Data Type	ID	Location	Correlation	P-Value	FDR	Annotation
swa	All	All	All	All	All	All
swa	CTNNB1	3p21	1	0	0	Apoptosis; Cell Signaling; Oncogenes
swa	CTNND1	11q11	0.814	4.42e-15	2.78e-10	morphogenesis of a polarized epithelium;adherens junction organization
swa	CTNNA1	5q31.2	0.784	2.18e-13	9.15e-9	Apoptosis; Tumor Suppressors
swa	RAB6A	11q13.3	0.531	0.0000154	0.0303	peptidyl-cysteine methylation;protein localization in Golgi apparatus
swa	CTNNA2	2p12-p11.1	0.523	0.0000216	0.0321	radial glia guided migration of Purkinje cell;muscle cell differentiation
swa	TROVE2	1q31	-0.505	0.0000457	0.0384	transcription from RNA polymerase III promoter
swa	SLC1A5	19q13.3	0.502	0.0000502	0.0395	Solute Carriers
swa	S100A16	1q21	0.49	0.0000803	0.0455	response to calcium ion
swa	COL5A2	2q14-q32	0.466	0.000199	0.0606	collagen fibril organization;skin development
swa	CD9	12p13.3	0.459	0.000258	0.0661	response to water deprivation;platelet degranulation
swa	PGK1	Xq13.3	-0.457	0.000276	0.0684	glucose metabolic process;gluconeogenesis
swa	RAC2	22q13.1	-0.456	0.000282	0.069	GTP catabolic process;actin cytoskeleton organization
swa	SIPA1	11q13	0.456	0.000286	0.069	Apoptosis
swa	EPCAM	2p21	0.454	0.000301	0.0697	positive regulation of cell proliferation;ureteric bud development
swa	DHRS7	14q23.1	0.453	0.000313	0.0702	
swa	INF2	14q32.33	0.447	0.000388	0.0756	cellular component organization;actin cytoskeleton organization
swa	IBA57	1q42.13	0.443	0.000442	0.079	glycine catabolic process;heme biosynthetic process
swa	VAPB	20q13.33	0.436	0.000555	0.0845	endoplasmic reticulum unfolded protein response;small molecule metabolic process
swa	NDUFAF3	3p21.31	0.433	0.000619	0.0871	mitochondrial respiratory chain complex I assembly
swa	DSP	6p24	0.431	0.000656	0.0873	cell-cell adhesion;peptide cross-linking
swa	ANXA2	15q22.2	0.43	0.000683	0.0889	positive regulation of vesicle fusion;fibrinolysis

Supplementary Figure 43, Related to Figure 2 and 4. CellminerCDB snapshot showing co-expression of cell adhesion proteins determined by SWATH. The "Compare pattern" tool was used with β-catenin (CTNNB1) as the "x-axis entry" (https://discover.nci.nih.gov/CellMinerCDB). Only the top correlates are shown among over 3,000 proteins in the database.



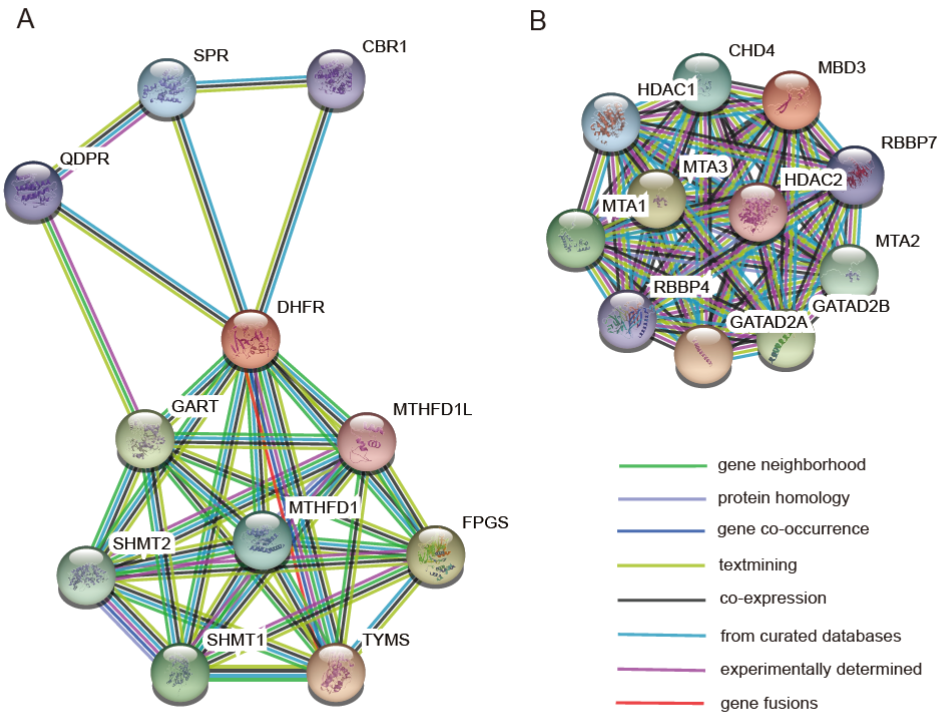
Supplementary Figure 44, Related to Figure 4. Predictive protein biomarkers for cisplatin (NSC119875) activity. The snapshots from https://discover.nci.nih.gov/CellMinerCDB show: **A.** Results obtained with the "Regression Model" tool of CellMinerCDB using cisplatin as "Response Identifier" for the query. **B.** Significant negative correlation between EPCAM protein expression determined by SWATH and activity of cisplatin. **C.** Highly significant negative correlation between EPCAM protein expression and EPCAM promoter methylation across the NCI-60.



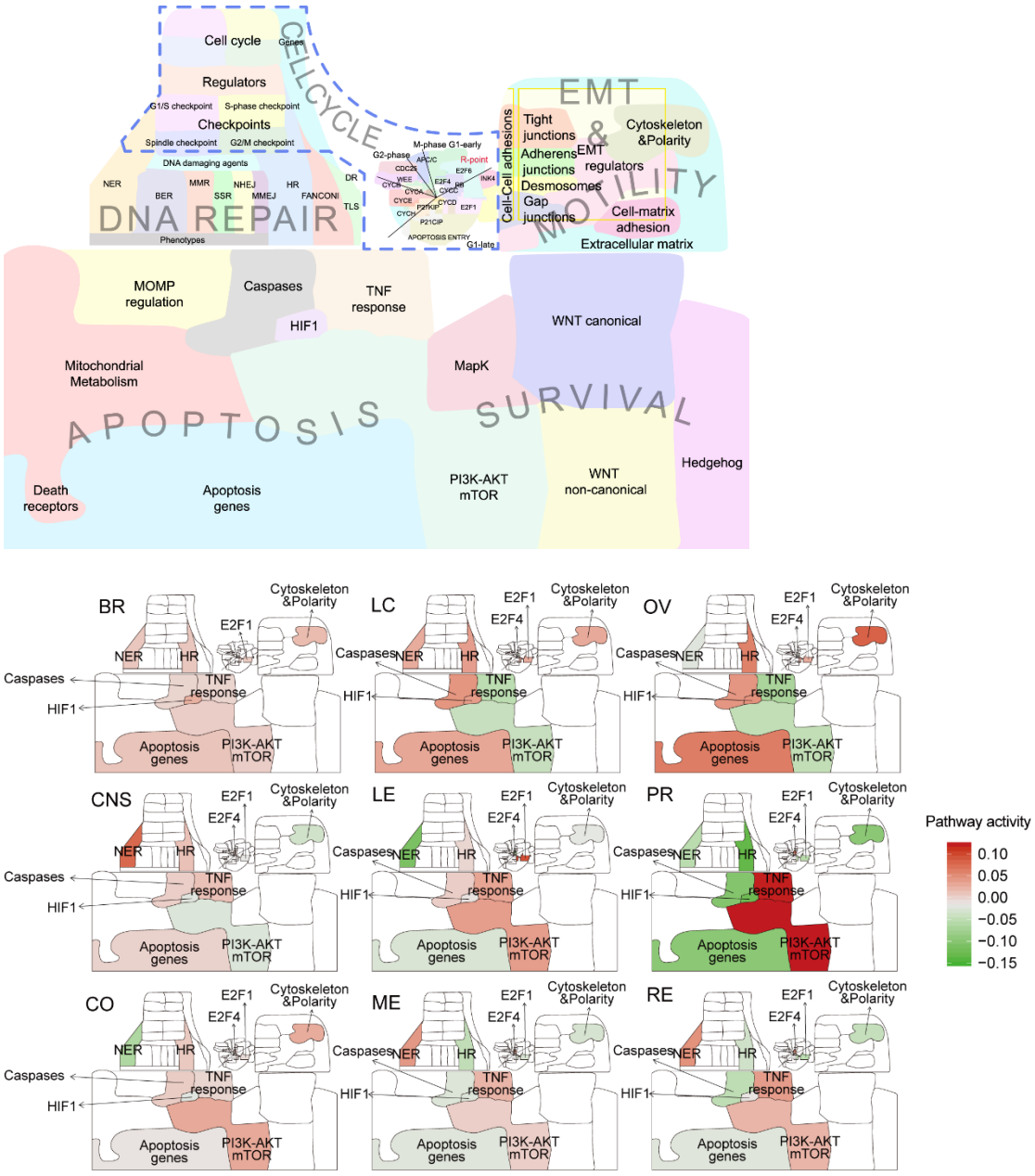
Supplementary Figure 45, Related to Figure 4. ABCB1 (PGP; P-glycoprotein) protein levels across the NCI-60 and prediction of drug response. https://discover.nci.nih.gov/CellMinerCDB snapshots showing: **A.** The correlation between ABCB1 protein and gene expression across the NCI-60. **B-D.** The significant negative correlations between ABCB1 protein levels and response to doxorubicin, depsipeptide and taxol across the NCI-60.

Data Type	ID	Location	Correlation	P-Value	FDR	Annotation
swa	All	All	All	All	All	All
swa	RPA3	7p22	1	0	0	DNA Damage Response (DDR); DDR (MMR); DDR (DNA replication)
swa	RPA2	1p35	0.752	6.58e-12	2.07e-7	DNA Damage Response (DDR); DDR (MMR); DDR (DNA replication)
swa	MCM5	22q13.1	0.628	1.02e-7	0.000985	DDR (DNA replication)
swa	MCM7	7q21.3-q22.1	0.625	1.22e-7	0.00102	DDR (DNA replication)
swa	PFDN2	1q23.3	0.615	2.25e-7	0.00142	protein folding;cellular protein metabolic process
swa	ITGB1	10p11.2	-0.612	2.56e-7	0.00153	Apoptosis
swa	DUT	15q21.1	0.6	5.02e-7	0.00208	DNA Damage Response (DDR)
swa	NUDC	1p35-p34	0.598	5.63e-7	0.00208	mitotic cell cycle;cytokinesis
swa	LRRC47	1p36.32	0.598	5.85e-7	0.00208	
swa	MCM3	6p12	0.595	6.71e-7	0.00208	DDR (DNA replication)
swa	TBCA	5q14.1	0.592	7.73e-7	0.00214	'de novo' posttranslational protein folding;protein folding
swa	PAICS	4q12	0.591	8.21e-7	0.00217	purine base metabolic process;purine nucleotide biosynthetic process
swa	UBE2K	4p14	0.588	9.67e-7	0.0023	ubiquitin-dependent protein catabolic process;protein ubiquitination
swa	RFC4	3q27	0.588	9.71e-7	0.0023	DNA Damage Response (DDR); DDR (MMR); DDR (DNA replication)
swa	PLEC	8q24	-0.585	0.00000116	0.00257	cell junction assembly;apoptotic process
swa	EIF3G	19p13.2	0.583	0.00000126	0.0026	cellular protein metabolic process;translation
swa	PFDN6	6p21.3	0.581	0.00000143	0.0027	
swa	HMGB2	4q31	0.581	0.00000144	0.0027	DNA Damage Response (DDR); DDR (BER)
swa	RFC2	7q11.23	0.578	0.00000163	0.00296	DNA Damage Response (DDR); DDR (MMR); DDR (DNA replication)
swa	ETF1	5q31.1	0.576	0.0000018	0.00305	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay;cellular protein metabolic process
swa	PSMF1	20p13	0.575	0.00000186	0.00305	S phase of mitotic cell cycle;apoptotic process

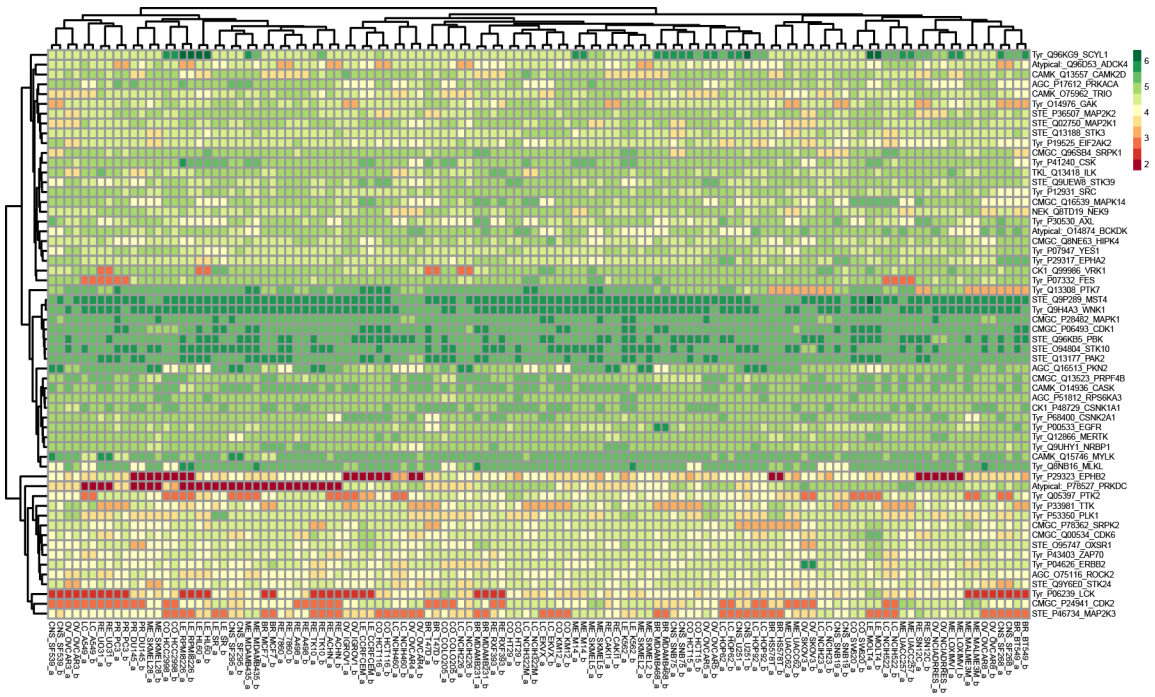
Supplementary Figure 46, Related to Figure 2. CellminerCDB snapshot views showing co-expression of replication proteins determined by SWATH. The "Compare pattern" tool was used with RPA3 as the "x-axis entry". Snapshots from https://discover.nci.nih.gov/CellMinerCDB showing the top correlates with RPA3.



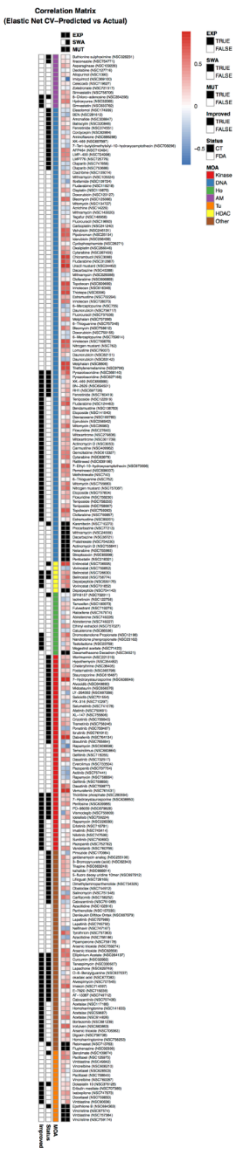
Supplementary Figure 47, Related to Figure 2. Proteins interacting with DHFR and MBD3. Proteins interacting with DHFR (A) and MBD3 (B) from STRING.



Supplementary Figure 48, Related to Figure 3. Global cancer signaling pathway maps based on Atlas of Cancer Signaling Network (ACSN) pathways (Kuperstein et al., 2015) (www. acsn.curie.fr). The annotations of the pathway map are shown in the upper panel. ROMA representation of the pathway activities are shown in the lower panel.



Supplementary Figure 49, Related to Figure 3. Fifty-eight protein kinases quantified in the NCI-60 cell panel. Expression of 58 protein kinases, represented by log10 transformed protein intensity values, in the NCI-60. Values are clustered without supervision across both proteins and cell lines.



Supplementary Figure 50, Related to Figure 3. Predictive power of different omics data combinations for the activity of 224 compounds, based on elastic net (multivariate linear regression) modeling of the drug response. Each column indicates the input data gene expression and mutation alone and in combination with proteomic abundances; each row represents a compound. The color indicates the predictive power, measured by Pearson correlation of cross-validation predicted and observed drug response values. Rows specifying compound-specific response prediction accuracies are sorted by mechanism of action and additional annotations are provided 1) whether the inclusion of the SWATH data improved the overall model and 2) the clinical status of the compound whether FDA approved or in clinical trial.

Supplementary Table Legends

All the supplementary tables are provided as separate Excel spreadsheet.

Supplementary Table 1, Related to Figure 1. Quantitative proteome maps of the NCI-60 cell lines. (A) Information for PCT-SWATH analysis of the NCI-60 cells (B) List of peptide precursors appeared in the library. (C) Quantitative values of 22,554 peptide precursors in the NCI-60 cells in duplicates. (D) Quantitative values of 3,171 proteins in the NCI-60 cells in duplicates. (E) Averaged protein intensity in the NCI-60 cells.

Supplementary Table 2, Related to Figure 2. Count of proteins in each cell line in the DDA data and the SWATH data. This table shows the count of IPI protein group number from the DDA data set quantified using iBAQ and LFQ algorithms 1, and the count of SwissProt proteotypic proteins from the SWATH data as reported in this data(Gholami et al., 2013a).

Supplementary Table 3, Related to Figure 2. Correlation between NCI-60 transcript expression and SWATH-MS protein expression for indicated gene.

Supplementary Table 4, Related to Figure 2. Stoichiometry of 101 protein complexes in the NCI-60 proteotype. Average Abundance (log10) means the averaged log 10 scaled protein abundance signal for proteins in a complex. Standard Deviation means the standard deviation of log 10 scaled protein abundance signal for proteins in a complex. Average Pearson Correlation means the averaged Pearson corelation value for each pair of proteins in a complex.

Supplementary Table 5, Related to Figure 2. The activity of apoptosiswas found significantly higher in ovarian cell lines. (A) The modules that show a significant dispersion are reported here. (B) A t-test is performed for cell lines from one cancer type vs. all other cancer cell lines.

Supplementary Table 6, Related to Figure 3. Cellminer data for the NCI-60 cells used in this study. (A) Exome data of the NCI-60 cells. (B) Log2 scaled mRNA expression data of the NCI-60 cells. (C) Common features at three different levels, i.e. DNA, mRNA and protein.

Supplementary Table 7, Related to Figure 3. Elastic net results.

Transparent Methods

PCT-assisted sample preparation for MS analyses

The NCI-60 cells were obtained as frozen, non-viable cell pellets from the Developmental Therapeutics Program (DTP), National Cancer Institute (NCI-NIH) and processed using Barocycler® NEP2320 (PressureBioSciences Inc, South Easton, MA). The IDs of the NCI-60 cells in our study matching to the IDs in Cellminer and a previous proteomic study by the Kuster group are provided in **Supplementary Table 1**. Briefly, cell pellets were lysed in a buffer containing 8M urea, 0.1M ammonium bicarbonate, and Complete^{TM} protease inhibitor using barocycler program (20 seconds 45 kpsi, 10 seconds 0 kpsi, 120 cycles) at 35°C (Guo et al., 2015). Whole cell lysates were sonicated for 25 seconds with 1 min interval on ice for 3 times. Cellular debris was removed by centrifugation and sample protein concentration was determined by BCA assay prior to protein reduction with 10 mM TCEP for 20 min at 35°C, and alkylation with 40 mM iodoacetamide in the dark for 30 min at room temperature. Lys-C digestion (1/50, w/w) was performed in 6 M urea using PCT program (25 seconds 25 kpsi, 10 seconds 0 kpsi 75 cycles) at 35°C; whereas trypsin digestion (1/30, w/w) was performed in further diluted urea (1.6M) using PCT program (25 seconds 25 kpsi, 10 seconds 0 kpsi, 160 cycles) at 35°C. Digestion was stopped by acidification with trifluoroacetic acid to a final pH of around 2 before C18 column desalting using SEP-PAK C18 cartridges (Waters Corp., Milford, MA, USA).

Off-gel electrophoresis

To create a comprehensive spectral library for SWATH-MS analysis, we pooled 20-40% of desalted peptide solutions from each NCI-60 sample and performed off-gel fractionation. Briefly, pooled peptides were resolubilised in OGE buffer containing 5% (v/v) glycerol, 0.7% (v/v) acetonitrile (ACN) and 1% (v/v) carrier ampholytes mixture (IPG buffer pH 3.0-10.0, GE Healthcare). Fractionation was performed on a 3100 OFFGEL (OGE) Fractionator (Agilent Technologies) using a 24 cm pH3-10 IPG strip (Immobilised pH Gradient strip from GE Healthcare) according to manufacturer's instructions using a program of 1 h rehydration at a maximum of 500 V, 50 µA and 200 mW followed by separation at a maximum of 8000 V, 100 µA and 300 mW until 50 kVh were reached. Each of 24 fraction was recovered and cleaned up by C18 reversed-phase MicroSpin columns (The Nest Group Inc.). Based on the sample complexity (based on Nanodrop, A280 measurement), for each strip, the following fractions were pooled into 12 samples for MS injections: pool 1 (fraction 1-2), pool 2 (fraction 3), pool 3 (fraction 4), pool 4 (fraction 5), pool 5 (fraction 6-7), pool 6 (fraction 8-9), pool 7 (fraction 10-11), pool 8 (fraction 12-15), pool 9 (fraction 16-19), pool 10 (fraction 20-

21), pool 11 (fraction 22), pool 12 (fraction 23-24). Those were injected in quadruplicate, resulting in 48 DDA injections of fractionated samples.

DDA MS for spectral library generation

For spectral library generation, a SCIEX TripleTOF 5600 System mass spectrometer was operated essentially as described before (Schubert et al., 2015): all samples were analyzed on an Eksigent nanoLC (AS-2/1Dplus or AS-2/2Dplus) system coupled with a SWATH-MS-enabled AB SCIEX TripleTOF 5600 System. The HPLC solvent system consisted of buffer A (2% ACN and 0.1% formic acid, v/v) and buffer B (95% ACN with 0.1% formic acid, v/v). Samples were separated in a 75 µm diameter PicoTip emitter (New Objective) packed with 20 cm of Magic 3 µm, 200A C18 AQ material (Bischoff Chromatography). The loaded material was eluted from the column at a flow rate of 300 nL min^{-1} with the following gradient: linear 2 - 35% B over 120 min, linear 35 - 90% B for 1 min, isocratic 90% B for 4 min, linear 90 - 2% B for 1 min and isocratic 2% solvent B for 9 min. The mass spectrometer was operated in DDA mode using a top20 method, with 500 ms and 150 ms acquisition time for the MS1 and MS2 scans respectively, and 20 s dynamic exclusion for the fragmented precursors. Rolling collision energy using the following equation (0.0625 × m/z - 3.5) with a collision energy spread of 15 eV was used for fragmentation regardless of the charge state of the precursors, to mimic as close as possible the fragmentation conditions of the precursors in SWATH-MS mode. Altogether, we had 66 DDA-MS injections, including the 48 OGE samples and another 18 pooled peptide samples from the unfractionated cell lysate of the NCI-60 cells.

Spectral and assay library generation

All raw instrument data were centroided using Proteowizard msconvert (version 2.0). The assay library was generated using an established protocol (Schubert et al., 2015). In short, the shotgun data sets were searched individually using X!Tandem (Craig and Beavis, 2003) (2011.12.01.1) with k-score plugin (MacLean et al., 2006), Myrimatch (Tabb et al., 2007) (2.1.138), OMSSA (Geer et al., 2004) (2.1.8) and Comet (Eng et al., 2013) (2013.02r2) against the reviewed UniProtKB/Swiss-Prot (2014_02) protein sequence database containing 20,270 proteins appended with 11 iRT peptides and decoy sequences. Carbamidomethyl was used as a fixed modification and oxidation as the variable modification. Maximally two missed cleavages were allowed. Peptide mass tolerance was set to 50 ppm, fragment mass error to 0.1 Da. The search identifications were combined and statistically scored using PeptideProphet (Keller et al., 2002) and iProphet (Shteynberg et al., 2011) available within the Trans-Proteomics Pipeline (TPP) toolset (version 4.7.0) (Keller et al., 2005). MAYU

(Reiter et al., 2009) (v. 1.07) was used to determine the iProphet cutoff (0.999354) corresponding to a protein FDR of 1.03%. SpectraST was used in library generation mode with CID-QTOF settings and iRT normal-isation at import against the iRT Kit (Escher et al., 2012) peptide sequences (-c_IRTirt.txt -c_IRR) and a consensus library was consecutively generated. An in-house python script, spec-trast2tsv.py31 (msproteomicstools 0.2.2) was then used to generate the assay library with the following settings: -l 350,2000 -s b,y -x 1,2 -o 6 -n 6 -p 0.05 -d -e -w swath32.txt -k openswath (fragment ions between 350 and 2000 m/z, b and y ions authorized, fragment charges 1+ and 2+, 6 most intense transitions, precision of fragment ion retrieved 0.05 Da, exact fragment ion mass calculated, exclude fragments in the swath window). The OpenSWATH tool, ConvertTSVToTraML converted the TSV file to TraML format; Open-SwathDecoyGenerator generated the decoy assays in shuffle mode and appended them to the TraML assay library. In this study, we built a SWATH assay library containing 86,209 proteotypic peptide precursors in 8,056 proteotypic SwissProt proteins. This library is supplied in PRIDE project PXD003539.

SWATH-MS

The SWATH-MS data acquisition in a Sciex TripleTOF 5600 mass spectrometer was performed as described before (Gillet et al., 2012), using 32 windows of 25 Da effective isolation width (with an additional 1 Da overlap on the left side of the window) and with a dwell time of 100 ms to cover the mass range of 400 - 1200 *m/z* in 3.3 s. The collision energy for each window was set using the collision energy of a 2+ ion centered in the middle of the window (equation: 0.0625 x *m/z* - 3.5) with a spread of 15 eV. The sequential precursor isolation window setup was as follows: [400-425], [424-450], [449-475], …, [1174-1200].

Protein identification using OpenSWATH

We analyzed the SWATH data using OpenSWATH software (Rost et al., 2014) using parameters as described previously (Ori et al., 2016). We identified 48,374 peptides from 6,556 protein groups from the NCI-60 panel with < 1% false discovery rate at both peptide and protein level evaluated by OpenSWATH (Rost et al., 2014)and Mayu (Reiter et al., 2009) (supplied in PRIDE project PXD003539).

DIA-expert analyses

The DIA-expert software read OpenSWATH output result file which contains statistical scores (*i.e.* mProphet score or mScore) indicating the confidence of identification for each

peptide precursor in each sample, and from there selected the sample in which a peptide precursor was identified with highest confidence. It then obtained extracted ion chromatograms (XICs) for the target peptide precursor and all associated annotated *b* and *y* fragments in the reference sample, and refined fragments based on the peak shape of each fragment and its peak boundary. The refined fragments and precursor XIC traces from each of the rest samples were subsequently compared with the reference peak group using empirical expert rules, based on which the best matched peak group in each sample was picked and visualized. Duplicated measurements were used to evaluate the accuracy of peptide and protein quantification. The protein quantity was normalized based on total ion chromatography of the MS1 spectra from each raw SWATH file. All codes are provided in Github https://github.com/tiannanguo/dia-expert.

PRM analysis

PRM quantification strategy was used to quantify selected proteins. Biognosys-11 iRT peptides (Biognosys, Schlieren, CH) were spiked into peptide samples at the final concentration of 10% prior to MS injection for RT calibration. Peptides were separated at 300 nL/min along a 45min 8–35% linear LC gradient (buffer A: 2% ACN, 0.1% formic acid; buffer B: 20% ACN, 0.1% formic acid). The Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer was operated in the MS/MS mode with time-scheduled acquisition for 54 peptides in a +/- 5 min retention time window. The full MS mode was measured at resolution 60,000 at m/z 200 in the Orbitrap, with AGC target value of 3E6 and maximum IT of 55ms. Target ions were submitted to MS/MS in the HCD cell (1.2 amu isolation width, 30% normalized collision energy). MS/MS spectra were acquired at resolution 30,000 (at m/z 200) in the Orbitrap using AGC target value of 2E5, a max IT of 100ms.

Quantitative proteomics and transcriptomics analysis of protein complexes components

Technical replicates were averaged to generate the NCI-60 proteotypes. To assess the coverage of protein complexes by NCI-60 proteotypes, we first retrieved a large resource of mammalian protein complexes assembled from CORUM (Ruepp et al., 2010), COMPLEAT (Vinayagam et al., 2013) and literature-curated complexes (Ori et al., 2013; Ori et al., 2016). This resource contains 2,041 proteins as members of 279 distinct complexes and it is available at http://variablecomplexes.embl.de/. 101 complexes were represented in the NCI-60 proteotypes with at least 5 members quantified. These complexes, in total, contain 1,045 distinct proteins quantified in the NCI-60 proteotypes. Pearson's correlation coefficient was calculated for all the pairwise comparisons of 3,171 proteins across the NCI-60 cell lines. All

pairwise comparisons were classified into two categories: either two proteins were members of the same complex or not. Average abundance, standard deviation and average Pearson correlation of each complex were calculated based on the abundance of complex members in the NCI-60 proteotypes.

An extended list of protein-protein interactions (PPIs) was generated based on information acquired from 6 resources: 1) 17,556 PPIs were retrieved from the CORUM database of human protein complexes (Ruepp et al., 2010); 2) 16'345 PPIs were composed from the interaction pairs annotated as 'complex' members in the Reactome database (Fabregat et al., 2018); 3) 12,664 PPIs were retrieved from the STRING database (Szklarczyk et al., 2015) considering only high confidence interactions (score ≥ 700). 4) 1'378 interaction pairs were obtained from Interactome3D (Mosca et al., 2013). These interactors corresponded to experimentally observed interactions with a support in the form of structural data or structural models. 5) 309 PPIs were obtained by considering interactions identified in at least 3 independent APMS experiments. For this, we included studies deposited in the BioGrid database (Chatr-Aryamontri et al., 2017), interactions listed in the BioPlex portal (Huttlin et al., 2015), and interactions observed in the large-scale Polycomb (Hauri et al., 2016) and Kinome studies (manuscript in preparation). 6) 122 PPIs were retrieved from the EMBL-EBI complex portal (Park et al., 2017). The latter (smallest) set of interactions is manually curated and of high confidence.

Combining information from the different databases, a list of 35,693 unique interactions (encompassing 1,766 proteins) was generated. The Spearman coefficient of correlation of protein abundances (Spearman's r) and the associated p-value were calculated for all the 5,026,035 protein pairs that can be formed from the 3171 proteins measured in the proteomics dataset. For this, the cor.mtest function from the package corrplot was applied with the Benjamini-Hochberg correction for multiple testing. Distribution of pairwise correlation values for three different sets was visualized with the density plots. The sets represented pairs found to interact in the respective database, all background NCI60 pairs (common to all analysis) and protein pairs that were both measured by NCI60 and present in the respective database, but not reported as interacting. The mean correlation values between the datasets were compared with the Wilcoxon test in R.

Pairwise correlation analysis of the mRNA levels was based on the expression data retrieved from the CellMiner. Cell lines with missing values (CNS.SF_539, ME.MDA_N and LC.NCI_H23) and transcripts for which the matching proteins were not measured were excluded from the analysis. Therefore, the final analysis was performed on a complete matrix with 57 cell lines and 2,835 transcripts. The Spearman's r and associated p-values were calculated as above for the 4,017,195 mRNA pairs that can be formed from the 2835

measured transcripts. Distribution of correlation values was compared between the set of true interaction partners and the corresponding background sets as described above.

Pathway activity analysis

The activity of pathways, as they are described in ACSN, has been computed using ROMA (Martignetti et al., 2016). Among all the modules defined in ACSN, only 11 show a significant dispersion over the data set: AKT_MTOR, HR (Homologous Recombination), NER (nucleotide Excision Repair), TNF response, Death Receptors regulators, Apoptosis, caspases, E2F3 and E2F4 targets, HIF1 and cytoskeleton polarity. For these modules, the mean activity score for each type of cancer cell lines was computed and mapped onto the atlas (from bright green for low values to bright red for high values). To assess module differential activity between proteotypes, we computed a *t*-test on the activity scores in cell lines of a cancer type versus the activity of all other cancer cell lines. The definition of genes composing each module can be found in http://acsn.curie.fr

Drug sensitivity prediction using elastic net

The elastic net regularized regression algorithm was applied to predict drug response for 240 FDA-approved or investigational NSC-designated compounds. Some widely studied drugs are represented by more than one NSC identifier, with each identifier associated with a distinct compound sample and series of NCI-60 drug activity assays. For each compound, two sets of input data were evaluated. These included NCI-60 mRNA expression, gene-level mutation alone and in combination with SWATH-MS protein expression. mRNA expression data was available for 25,040 genes, and derived from CellMiner (discover.nci.nih.gov/cellminer and discover.nci.nih.gov/cellminecdb) (Rajapakse et al., 2018; Reinhold et al., 2012; Reinhold et al., 2015; Reinhold et al., 2017), with missing values imputed using the impute.knn function (with default parameters) of the Bioconductor impute package. Gene-level mutation profiles were available for 9,307 genes, and were obtained from CellMiner exome sequencing data, with values indicating the percent conversion to a variant form for the case of expected function-impacting alterations (frameshift, nonsense, splice-sense, missense mutations by SIFT/PolyPhen2 analysis). SWATH-MS based protein expression data from the current study was also included.

Elastic net analysis was done using the glmnet R package (Friedman et al., 2010). The elastic net analysis was conducted using a multi-step pipeline involving cross-validations performed in a nested manner. The “outer” cross-validation is a leave-one-out cross validation that is conducted over all computational steps present in the “inner” pipeline, and it is used to

validate model performance. The “inner” cross-validation are conducted to select elastic net hyperparameters (alpha and lambda) and for predictor set trimming, using data from a set of ~59 cell lines.

The elastic net parameters alpha and lambda were selected by minimizing the cross-validation error (average of 10 replicates of 10-fold cross-validation) within the “inner” pipeline. The selected alpha and lambda parameters were then applied to 200 runs of the elastic net algorithm, each using a random data subset derived from 90% of the available cell lines. The 200 resulting coefficient vectors were then averaged, and predictors were ranked by the magnitude of their average coefficient weight. To select a limited number of predictors with potential to generalize to new data, top k-element predictor sets (by average coefficient weight magnitude) were evaluated using standard linear regression and 10-fold cross-validation. The appropriate k was set to the smallest value yielding a cross-validation error within one standard deviation of the minimum cross-validation error.

To obtain a robust estimate of performance on unseen data, leave-one-out cross-validation was applied to the overall procedure as part of the “outer” pipeline. Specifically, drug response for each cell line was predicted using an elastic net model derived using the remaining held out data (and the steps outlined above). The vector of predicted response values was then correlated with the actual response values, with the Pearson’s correlation coefficient providing an estimate of the predictive value of the applied input data combination. More details of the elastic net algorithm are provided in Supplementary Note 6.

Elastic net analysis was done using the rcellminerElasticNet R package (https://bitbucket.org/cbio_mskcc/rcellminerelasticnet), which facilitates the application of the glmnet R package (which provides the elastic net algorithm code) to data from the rcellminer and rcellminerData packages (Luna et al., 2016). rcellminerElasticNet also provides utility functions for summarizing and visualizing elastic net results.

Results for the elastic net analysis are available from this URL: https://discover.nci.nih.gov/cellminerreviewdata/swath_analysis/swathOutput_062316_all.tar.gz. This compressed file contains results for the analysis run with all features and selected common features. Each drug compound has three files for each combination of molecular features used in a particular run of the elastic net algorithm: 1) a knitr report R Markdown (.Rmd) file containing the code that was run, 2) an RData (.Rdata) file containing the results of each elastic net run (see elasticNet() documentation in the rcellminerElasticNet package), 3) the rendered knitr report as a webpage (.html).

Beyond the knitr report containing code, the elastic net pipeline is made reproducible using a Docker image. Docker (www.docker.com) is an emerging platform for conducting reproducible research in the biomedical research community. All necessary software and

dependencies to run the described analysis have been embedded in the available Docker container to provide readers an environment that runs on all major operating systems (including Windows, OSX, and Linux), making Docker containers self-contained, portable, and capable of performing at levels similar to the host system.

The Docker container is available at the Docker Hub repository: cannin/swath (https://hub.docker.com/r/cannin/swath/). Key dependencies installed, include: RStudio Server (https://www.rstudio.com/), rcellminer/rcellminerData (Luna et al., 2016), and rcellminerElasticNet. With these installed dependencies, readers have the opportunity to 1) re-run analysis for specific drug compounds and modify the code in order to extend the analysis using RStudio Server, a web-based version of the RStudio R editor, and 2) use an R Shiny app web-based data explorer to further understand described results. Instructions on the usage of the Docker container are located at the rcellminerElasticNet project page (https://bitbucket.org/cbio_mskcc/rcellminerelasticnet).

Supplementary References

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A.*, et al.* (2017). The BioGRID interaction database: 2017 update. Nucleic acids research *45*, D369-D379.

Craig, R., and Beavis, R.C. (2003). A method for reducing the time required to match protein sequences with tandem mass spectra. Rapid Commun Mass Spectrom *17*, 2310-2316.

Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics *13*, 22-24.

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M.J., and Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics *12*, 1111-1121.

Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., Wu, G., Stein, L., D'Eustachio, P., and Hermjakob, H. (2018). Reactome graph database: Efficient access to complex pathway data. PLoS Comput Biol *14*, e1005968.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw *33*, 1-22.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004). Open mass spectrometry search algorithm. J Proteome Res *3*, 958-964.

Gholami, A.M., Hahne, H., Wu, Z., Auer, F.J., Meng, C., Wilhelm, M., and Kuster, B. (2013a). Global proteome analysis of the NCI-60 cell line panel. Cell Rep *4*, 609-620.

Gholami, A.M., Hahne, H., Wu, Z.X., Auer, F.J., Meng, C., Wilhelm, M., and Kuster, B. (2013b). Global Proteome Analysis of the NCI-60 Cell Line Panel. Cell Rep *4*, 609-620.

Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & cellular proteomics : MCP *11*, O111 016717.

Guo, T., Kouvonen, P., Koh, C.C., Gillet, L.C., Wolski, W.E., Rost, H.L., Rosenberger, G., Collins, B.C., Blum, L.C., Gillessen, S.*, et al.* (2015). Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. Nature medicine.

Hauri, S., Comoglio, F., Seimiya, M., Gerstung, M., Glatter, T., Hansen, K., Aebersold, R., Paro, R., Gstaiger, M., and Beisel, C. (2016). A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep *17*, 583-595.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K.*, et al.* (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell *162*, 425-440.

Keller, A., Eng, J., Zhang, N., Li, X.J., and Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol *1*, 2005 0017.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*, 5383-5392.

Kuperstein, I., Bonnet, E., Nguyen, H.A., Cohen, D., Viara, E., Grieco, L., Fourquet, S., Calzone, L., Russo, C., Kondratova, M*., et al.* (2015). Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis *4*, e160.
Luna, A., Rajapakse, V.N., Sousa, F.G., Gao, J., Schultz, N., Varma, S., Reinhold, W., Sander, C., and Pommier, Y. (2016). rcellminer: exploring molecular profiles and drug response of the NCI-60 cell lines in R. Bioinformatics *32*, 1272-1274.
MacLean, B., Eng, J.K., Beavis, R.C., and McIntosh, M. (2006). General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. Bioinformatics *22*, 2830-2832.
Martignetti, L., Calzone, L., Bonnet, E., Barillot, E., and Zinovyev, A. (2016). ROMA: Representation and Quantification of Module Activity from Target Expression Data. Front Genet *7*, 18.
Mosca, R., Ceol, A., and Aloy, P. (2013). Interactome3D: adding structural details to protein networks. Nat Methods *10*, 47-53.
Ori, A., Banterle, N., Iskar, M., Andres-Pons, A., Escher, C., Khanh Bui, H., Sparks, L., Solis-Mezarino, V., Rinner, O., Bork, P*., et al.* (2013). Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. Mol Syst Biol *9*, 648.
Ori, A., Iskar, M., Buczak, K., Kastritis, P., Parca, L., Andres-Pons, A., Singer, S., Bork, P., and Beck, M. (2016). Spatiotemporal variation of mammalian protein complex stoichiometries. Genome Biol *17*, 47.
Park, Y.M., Squizzato, S., Buso, N., Gur, T., and Lopez, R. (2017). The EBI search engine: EBI search as a service-making biological data accessible for all. Nucleic acids research *45*, W545-W549.
Rajapakse, V.N., Luna, A., Yamade, M., Loman, L., Varma, S., Sunshine, M., Iorio, F., Sousa, F.G., Elloumi, F., Aladjem, M.I*., et al.* (2018). CellMinerCDB for Integrative Cross-Database Genomics and Pharmacogenomics Analyses of Cancer Cell Lines. iScience *10*, 247-264.
Reinhold, W.C., Sunshine, M., Liu, H.F., Varma, S., Kohn, K.W., Morris, J., Doroshow, J., and Pommier, Y. (2012). CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. Cancer Res *72*, 3499-3511.
Reinhold, W.C., Sunshine, M., Varma, S., Doroshow, J.H., and Pommier, Y. (2015). Using CellMiner 1.6 for Systems Pharmacology and Genomic Analysis of the NCI-60. Clinical cancer research : an official journal of the American Association for Cancer Research *21*, 3841-3852.
Reinhold, W.C., Varma, S., Sunshine, M., Rajapakse, V., Luna, A., Kohn, K.W., Stevenson, H., Wang, Y., Heyn, H., Nogales, V*., et al.* (2017). The NCI-60 Methylome and Its Integration into CellMiner. Cancer Res *77*, 601-612.
Reiter, L., Claassen, M., Schrimpf, S.P., Jovanovic, M., Schmidt, A., Buhmann, J.M., Hengartner, M.O., and Aebersold, R. (2009). Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. Molecular & cellular proteomics : MCP *8*, 2405-2417.

Rosenberger, G., Koh, C.C., Guo, T., Röst, H.L., Kouvonen, P., Collins, B.C., Heusel, M., Liu, Y., Caron, E., and Vichalkovski, A. (2014). A repository of assays to quantify 10,000 human proteins by SWATH-MS. Scientific data *1*, 140031.
Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., Wolskit, W., Collins, B.C., Malmstrom, J., Malmstrom, L.*, et al.* (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature Biotechnology *32*, 219-223.
Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic acids research *38*, D497-501.
Schubert, O.T., Gillet, L.C., Collins, B.C., Navarro, P., Rosenberger, G., Wolski, W.E., Lam, H., Amodei, D., Mallick, P., MacLean, B.*, et al.* (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc *10*, 426-441.
Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Molecular & cellular proteomics : MCP *10*, M111 007690.
Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P.*, et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic acids research *43*, D447-452.
Tabb, D.L., Fernando, C.G., and Chambers, M.C. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res *6*, 654-661.
Vinayagam, A., Hu, Y., Kulkarni, M., Roesel, C., Sopko, R., Mohr, S.E., and Perrimon, N. (2013). Protein complex-based analysis framework for high-throughput data sets. Science signaling *6*, rs5.