SEC-TMT facilitates quantitative differential analysis of protein interaction networks

Ella Doron-Mandel[1†,#], Benjamin J. Bokor[1†], Yanzhe Ma[1†], Lena A. Street[1], Lauren C. Tang[1], Ahmed A. Abdou[1], Neel H. Shah[2], George A. Rosenberger[3] & Marko Jovanovic[1#]

† Equally contributing authors
# Corresponding author

[1] Department of Biological Sciences, Columbia University, New-York, NY, USA
[2] Department of Chemistry, Columbia University, New-York, NY, USA
[3] Herbert Irving Cancer Research Center, Columbia University Medical Center, New-York, NY, USA

Correspondence to: ed2853@columbia.edu (E.D-M.), mj2794@columbia.edu (M.J.)

Abstract

The majority of cellular proteins interact with at least one partner or assemble into molecular-complexes to exert their function. This network of protein-protein interactions (PPIs) and the composition of macromolecular machines differ between cell types and physiological conditions. Therefore, characterizing PPI networks and their dynamic changes is vital for discovering novel biological functions and underlying mechanisms of cellular processes. However, producing an in-depth, global snapshot of PPIs from a given specimen requires measuring tens to thousands of LC-MS/MS runs. Consequently, while recent works made seminal contributions by mapping PPIs at great depth, almost all focused on just 1-2 conditions, generating comprehensive but mostly static PPI networks.

In this study we report the development of SEC-TMT, a method that enables identifying and measuring PPIs in a quantitative manner from only 4-8 LC-MS/MS runs per biological sample. This was accomplished by incorporating tandem mass tag (TMT) multiplexing with a size exclusion chromatography mass spectrometry (SEC-MS) work-flow. SEC-TMT reduces measurement time by an order of magnitude while maintaining resolution and coverage of thousands of cellular interactions, equivalent to the gold standard in the field. We show that SEC-TMT provides benefits for conducting differential analyses to measure changes in the PPI network between conditions. This development makes it feasible to study dynamic systems at scale and holds the potential to drive more rapid discoveries of PPI impact on cellular processes.

Introduction

The majority of cellular proteins interact with at least one partner or assemble into molecular-complexes to exert their function. Thus, a growing interest within the molecular and systems biology community is to globally map protein-protein interactions (PPIs) and their rewiring upon perturbation in an effort to better understand protein functions, molecular processes and cellular architecture. In line with this, multiple methods have been applied to the identification and quantification of PPI networks in living cells and tissues. Recently, seminal studies have used affinity purifications (AP) or proximity biotinylation (PB) on up to thousands of bait proteins to map the human cellular PPI networks in great depth – identifying nearly 150,000 interactions between ~15,000 proteins (spanning ~75% of the human genome)[1–5]. However, these approaches rely on targeting proteins-of-interest with antibodies or by genetically fusing affinity-tags or enzymes – limiting the resulting networks to the chosen bait proteins. Additionally, generating extensive PPI networks using bait-reliant methods requires measurement of thousands of samples from a given biological condition – a feat only feasible for specialized labs with abundant resources, such as dedicated mass spectrometer (MS) instrument(s).

Alternatively, discovery of PPIs is possible through protein correlation profiling (PCP) in co-fractionation (CF) experiments, where samples are fractionated based on protein biochemical properties (e.g., pH, isoelectric points, or size exclusion chromatography (SEC)) under physiological conditions that favor PPIs. Subsequent analysis of the correlation between the elution profiles of different proteins predicts their physical interaction[6–12]. Such CF methods are untargeted and are not dependent on antibody availability or the ability to transfect/ transduce the sample, enabling global analysis of PPIs from virtually any sample. However, CF methods are sensitive to missing datapoints, and traditional data-dependent acquisition (DDA) liquid chromatography tandem mass spectrometry (LC-MS/MS) methods are prone to data incompleteness due to their bias toward abundant proteins. To compensate, recent CF studies used data independent acquisition (DIA) LC-MS/MS methods and significantly improved coverage and completeness[10]. With the increased usage of SEC-DIA (also termed "SEC-SWATH") and other CF methods, multiple analysis tools have been developed - such as SECAT[13], CCprofiler[10], and PCprophet[14] - which focus on quantitative, error-rate controlled analyses of known protein interactions and molecular complexes, as well as other methods like EPIC[15] and PrInCE[16], which are more geared toward discovery of novel interactions.

Co-fractionation methods result in tens to hundreds of fractions per sample and thus, even the most comprehensive studies to date have been limited to profiling PPIs from just 1-2 conditions, thereby producing mostly static PPI networks. On the contrary, recent studies show that PPI networks are dynamic and highly dependent on the cell state[3,7,17–21]. In addition, PPIs are context dependent, varying between tissues and cell types[12]. Thus, to reveal how PPI dynamics shape biological function, more efficient methods that will enable larger studies like extensive time-courses or differentiating numerous biological conditions are still required.

To develop an efficient method to measure PPIs at scale with only a handful of LC-MS/MS runs we combined size exclusion chromatography (SEC) – a CF method that requires no more than 40 fractions to reliably quantify thousands of PPIs[22] – with multiplexing using Tandem Mass Tags (TMT). TMT is compatible with biochemical CF[7], and it allowed us to increase throughput and significantly reduce measurement time without compromising quantitation. The resulting "SEC-TMT" method enables global, differential quantitative analysis of PPIs from only 4-8 LC-MS/MS runs per biological sample, reducing

measurement time by an order of magnitude compared to the field's gold standard, while maintaining the same levels of coverage and resolution.

Results

Development of SEC-TMT – labeling and multiplexing design

We set out to test the feasibility of SEC-TMT by conducting SEC on HEK293 cells in two biological replicates and multiplexing the resulting 72 fractions per replicate with TMT18 reagents. To offset the low coverage of data dependent acquisition (DDA), SEC-adjacent fractions were multiplexed within the same TMT mix. This mixing scheme minimized the occurrence of missing data points per SEC elution peak since once a peptide was triggered for MS2 acquisition TMT reporter intensity values were, in most cases, assigned to all channels. In addition, we designed a "full-overlap" scheme in which we split every fraction in two and measured it twice, in different mixes (Figure 1, S1A). Such a design addresses the need for a shared reference channel between any two mixes for mix batch correction while maintaining low mix complexity. Indeed, we found that this approach improves signal and protein coverage compared to using a spike-in made of subsampling all fractions (data not shown) or having no-overlap between mixes (Figure 1B, S2A-B). We then used the false-discovery-rate (FDR) controlled analysis tool SECAT[13] to identify high-confidence PPIs and showed that the full-overlap scheme significantly increased the number of PPIs by nearly 3 fold compared to the non-overlap schemes (Figure S2C). While the 10% increase in protein IDs was expected from doubling the number of measurements, we were positively surprised by the large increase in PPIs recovered by the full-overlap scheme. Therefore, we concluded it justified doubling the number of LC-MS/MS runs and all subsequent experiments were conducted with the full-overlap scheme.

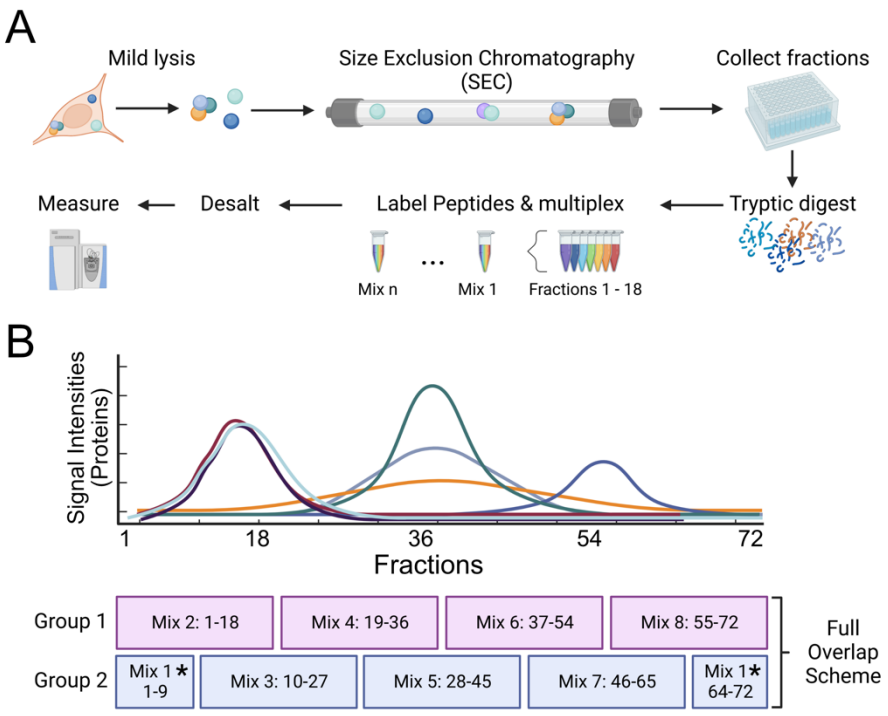
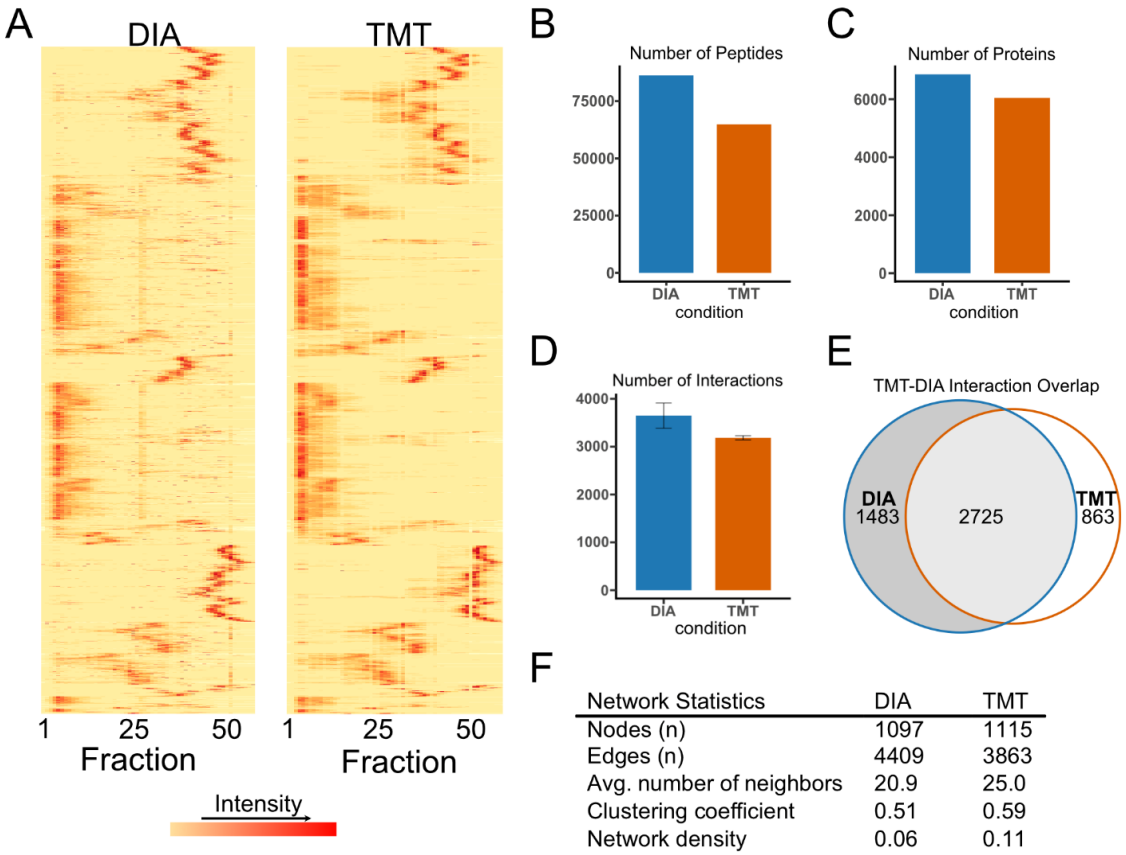


Figure 1: Overview of SEC-TMT experimental workflow
(A) Cells are lysed under physiological conditions, followed by fractionation on a size exclusion chromatography (SEC) column into ~90 fractions. After BCA quantification of the resulting fractions, protein-containing fractions are selected (54-72 fractions total) for further processing. Samples are then denatured, reduced, alkylated and digested using trypsin. The resulting peptides are directly labeled by TMT and multiplexed. TMTplexes are then desalted prior to being injected on the LC-MS/MS. (B) A "full overlap" mixing scheme was developed for SEC-TMT, in which each fraction is divided in two and each half is measured in two different mixes, keeping adjacent fractions together in a mix.

An additional hurdle in SEC-MS experiments is the high number of samples to process for LC-MS/MS. To overcome this, we modified the TMT labeling method (Figure 1A, and detailed in the "Materials and Methods" section) in a manner that obviates the need for peptide desalting in individual fractions prior to TMT labeling. Following digestion, peptides were immediately labeled in the adjusted digestion/labeling buffer and desalted only after multiplexing. This protocol resulted in a mean labeling efficiency of 98.3%, comparable to the classic protocol (mean 98.7%). Therefore, multiplexing did not only reduce measurement time, but also decreased sample processing.

SEC-TMT shows comparable performance to SEC-DIA in coverage and resolution, using considerably less measurement time

To evaluate the performance of SEC-TMT compared to the field's current gold standard, SEC-DIA, we performed two additional replicates of SEC from HEK cells, and measured each single fraction by label-free DIA. This comparison showed that SEC-TMT had strikingly similar elution patterns as SEC-DIA, evident by the positions of the elution peaks per protein (Figure 2A). As expected, the DIA dataset



Network Statistics	DIA	TMT
Nodes (n)	1097	1115
Edges (n)	4409	3863
Avg. number of neighbors	20.9	25.0
Clustering coefficient	0.51	0.59
Network density	0.06	0.11

Figure 2: ***SEC-TMT shows comparable performance to SEC-DIA in coverage and resolution***
(A) Heatmap representation comparing signals in SEC-DIA and SEC-TMT, for proteins measured in both. Columns represent fractions, rows represent different proteins, which are scaled from 0 to 1 so that the max elution peak per protein is represented in red. Rows in both heatmaps are arranged in the same order. (B-C) The number of (B) peptides and (C) proteins identified over two biological replicates of SEC-DIA and SEC-TMT. (D) The number of interactions identified by SECAT analysis (q-value < 0.05). Mean +/- std deviation, n = 5 SECAT runs. (E) The overlap of interactions between SEC-DIA and SEC-TMT, q-value < 0.05 in at least one condition and < 0.1 in the other, (see main text for details), in at least 3 SECAT runs. (F) Network statistics of the PPI networks from SEC-DIA and SEC-TMT (calculated with Cytoscape).

had overall more identifications (~20% at the peptide level and ~10% at the protein level, Figure 2B-C). Nevertheless, we found that the 10% coverage difference was worth the 90% decrease in LC-MS/MS runs.

We then ran SECAT on each dataset independently to identify interactions and compare the resulting networks. In both datasets, SECAT was able to build a classifier that successfully distinguished between decoy and targets based on their discriminant score (Figure S2D). In order to identify high-confidence interactions we ran SECAT five times on each dataset and set a strict cutoff of 5% FDR (interaction q-value < 0.05) required in at least three out of five runs. We initially observed that 34% of the interactions were mutual to both datasets, but found a large number of interactions unique at a q-value < 0.05 cutoff were very close to the cutoff in the other experimental setup (Figure S1F). Therefore, we adjusted the cutoff to include any interaction with a q-value between 0.05 and 0.1 (in at least 3 out of 5 SECAT runs), if its q-value was lower than 0.05 in the other dataset. With this adjusted cutoff we observed that 54% of interactions were identified in both datasets (2,725/5,071), while 29% (1,483/5,071) were unique to DIA and 17% (863/5,071) unique to TMT (Figure 2E). Analysis of the resulting PPI networks and their quantitative parameters showed that the DIA and TMT based networks are similar in their architecture (Figure 2F, S2E). Further inspection of the networks and the signals of specific complexes and their subunits revealed a few potential explanations for differences in the networks. In some cases, not all interactors were measured. In other cases, despite full coverage of the interactors, missing interactions may have stemmed from small interfering signals of individual proteins (e.g., CCT7, Figure 3A-C) or from not passing the SECAT cutoff (e.g., VBP1, Figure 3D-F).

Additionally, since a major interest of PPI studies lies in identifying novel interactions, we used a different analysis tool, EPIC[15], to identify reference-free interactions in SEC-TMT and SEC-DIA. This analysis resulted in 12,620 and 14,406 PPIs in SEC-TMT and SEC-DIA, respectively.

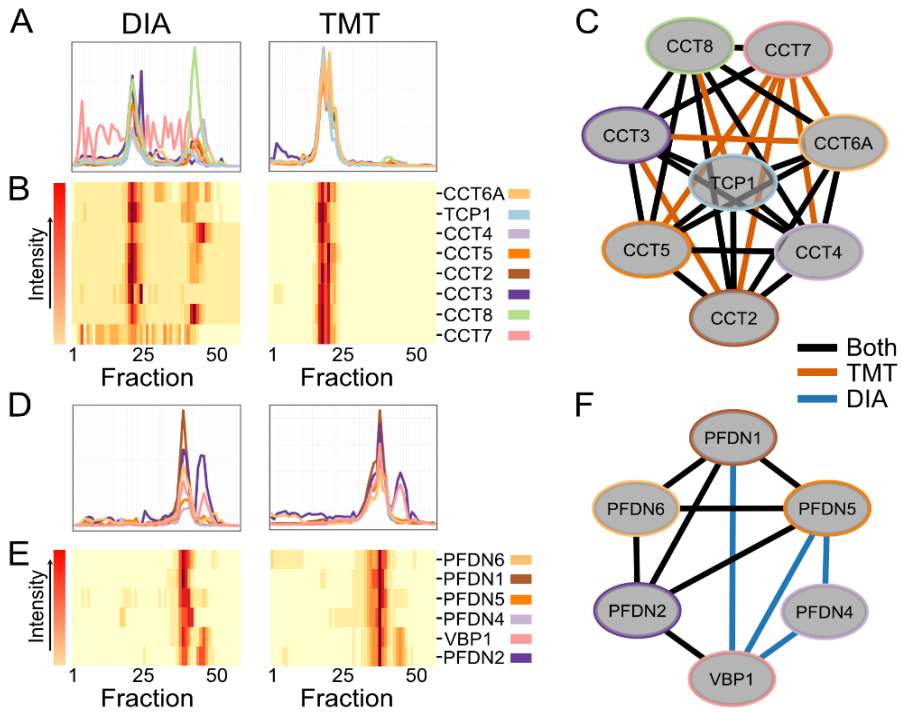


Figure 3: *Signal reproducibility between SEC-DIA and SEC-TMT for exemplary complexes* (A,D) SEC chromatograms for subunits of selected complexes (fractions on the x-axis, normalized signal intensity on the y-axis), color-coded as denoted in B and E. (B,E) Heatmaps representing the scaled elution profiles of all complex subunits presented in A and D. (C,F) network representation of complex-subunit interactions as identified by SECAT. Edge color represents whether it was identified in SEC-TMT only (orange), SEC-DIA only (blue), or both (black). A-C show the TRiC-CCT complex, D-F show the prefoldin complex.

In order to further evaluate SEC-TMT as an interactome building method, we compared the coverage and the numbers of PPIs per LC-MS/MS run in SEC-TMT interactomes generated by SECAT and EPIC to their matching SEC-DIA interactomes and to published interactomes produced by a variety of methods[1–7,11–13,23] (Figure 4). SEC-TMT generated the largest number of PPIs per MS run, validating it as a useful method for generating large-scale interactomes in shorter time. In conclusion, SEC-TMT performed comparably to SEC-DIA and other published (non-SEC) methods in building the human PPI network while requiring an order of magnitude less measurement time.

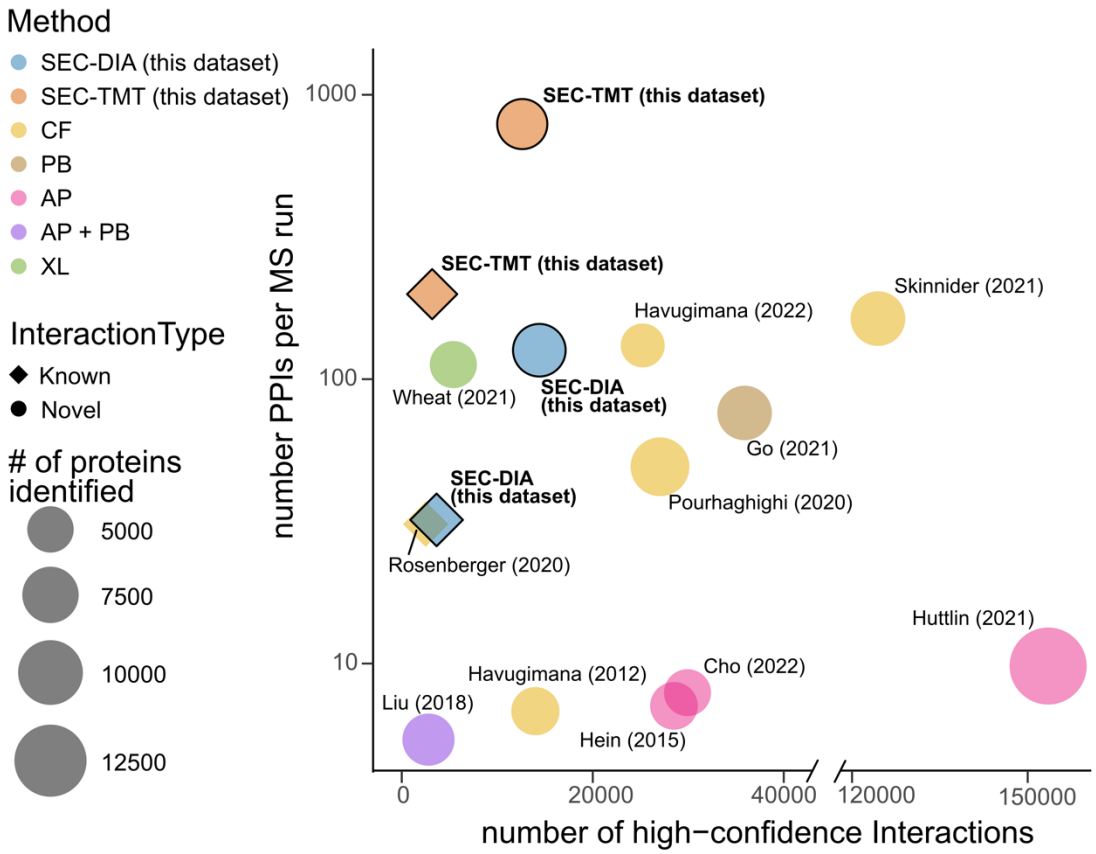


Figure 4: ***SEC-TMT performs on par with published methods, at an order-of-magnitude reduction in measurement time***

Evaluating the performance of MS-based methods for PPI identification through overall number of high-confidence interactions identified (x-axis, log2 scaled), number of interactions identified per MS run (y-axis, log10 scaled) and total number of protein identifications (symbol area). Symbol shape represents mode of analysis; allowing identification of novel interactions (circles), or a more conservative analysis based on query of known PPI databases (diamonds). Symbols are color coded based on the method used (CF = co-fractionation, including SEC, AP = affinity purifications, PB = proximity biotinylation, XL = crosslinking).

SEC-TMT enables quantitative differential comparison between multiple samples

The order-of-magnitude reduction in measurement time of SEC-TMT facilitates experimental designs with multiple conditions, and renders SEC-TMT a viable approach for studying context-dependent interactions and dynamic changes in PPI networks. Therefore, we evaluated the use of SEC-TMT for differential quantitative analysis using SECAT. To this end, we decided to compare the PPI network of HEK293 versus HCT116 cells, two commonly used cell lines.

In order to maximize our ability to perform quantitative differential analysis between two samples that differ in their global protein expression profiles, we minimized variability in protein coverage between the two cell-lines by multiplexing them together in the same mixes (Figure 5A, S1B). We hypothesized this multiplexing scheme would reduce the likelihood of identifying an interaction in one cell but not the other due to one of the interactors having missing values (if it was not triggered for MS2 in that sample).

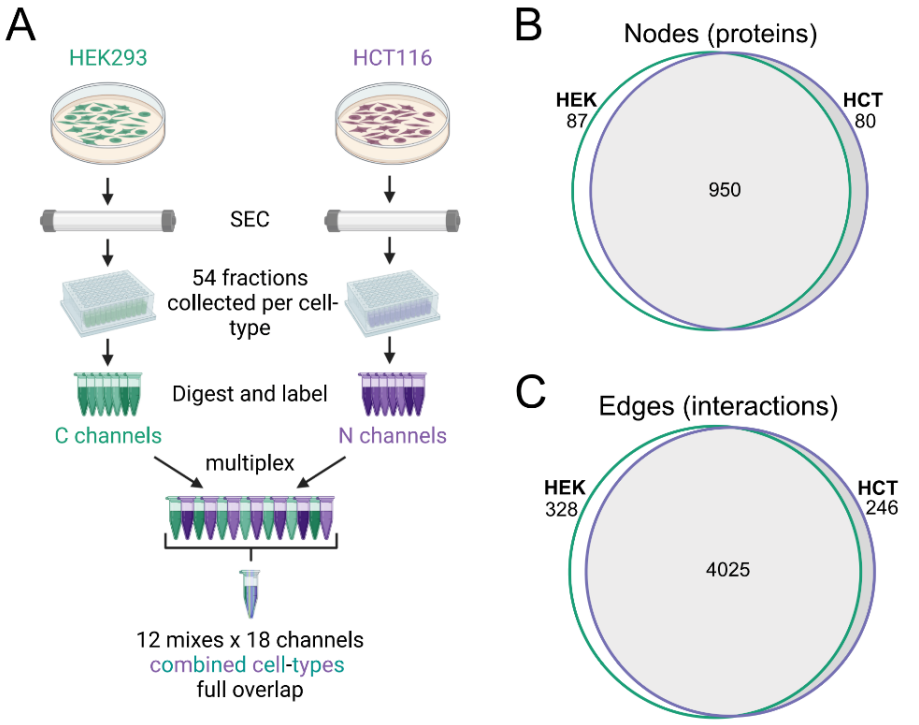


Figure 5: ***designing a TMT multiplexing scheme to maximize coverage of shared proteins and interactions between two different conditions***
(A) HEK and HCT cells were lysed, fractionated and labeled separately, in two biological replicates each. HEK fractions were labeled with the C-channels of the TMT reagent and HCT fractions with the N-channels of the TMT-channels to reduce signal bleed through between samples. HEK and HCT labeled fractions were then multiplexed together, maintaining a "full-overlap" scheme between adjacent mixes. (B-C) The overlap of nodes (proteins, B) or edges (interactions, C) in the PPI networks resulting from SECAT analysis.

Fifty-four SEC fractions were collected from each cell line, in two biological replicates. Every 9 consecutive fractions from both cells were multiplexed together into one mix while still maintaining a "full overlap" between every two mixes, yielding 12 mixes of 18 channels per biological replicate (Figure S1B). Overall, we measured 53,816 peptides covering 5,357 proteins in both cell types. We then analyzed the data using SECAT to identify high confidence PPIs (at 5% FDR in at least one condition and 10% in the other, as described above, Figure S3C) and quantified their abundance in order to investigate the differences in PPIs between the two cell-types. The resulting network had 1,117 nodes and 4,599 edges of which 85% and 87.5% were shared, respectively (Figure 5B-C). While the high overlap in node identification is expected from the multiplexing scheme (the same proteins in HEK and HCT were measured and quantified), we were surprised to see such agreement in PPIs between two different cell types. Therefore, we wanted to find which overlapping interactions showed quantitative changes between the two samples.

To evaluate the reliability of our SEC-TMT differential quantification, we measured the global protein expression of each unfractionated original sample using DIA and plotted the resulting ratio changes in protein expression (HCT/HEK) against the ratio of the "total abundance" (HCT/HEK) calculated by SECAT directly from our fractionated SEC-TMT sample (Figure S3A) and observed a high-degree positive correlation between the values ($R^2 = 0.58$), indicating that SEC-TMT enabled accurate quantification.

Next, we turned to quantitative comparison between the HEK and HCT PPI networks. To estimate the extent of proteins participating in differential interactions between the two cell lines, we used the differential ratio of SECAT's interactor abundances (HCT/HEK), defined as the abundance ratio (HCT/HEK) within the interaction region for any protein participating in an interaction (such that a given protein with multiple interaction partners will have several interactor abundance values assigned to it - one for each of its interactions). We defined an interactor as "significantly differential" using a cutoff of 50% up- or down-regulation (absolute log2 ratio > 0.58) and identified 249 unique proteins in interactions upregulated in HCT cells and 208 in HEK cells (Figure 6A). As expected, we observed a strong correlation between the fold changes in interactor-abundances (HCT/HEK) and the fold changes in global expression as measured by DIA (R^2 = 0.63, Figure 6B), suggesting that the majority of interaction differences are driven by differential expression of the interacting proteins. Subsequent GO term enrichment analysis[24] of the resulting enriched protein lists showed that these proteins represent distinct molecular functions in each cell type (Figure 6C-D), such as the ribosome, cell adhesion, and actin binding in HCT cells and DNA catalytic activity, helicases, ligases and deacetylases in HEK cells.

Lastly, we quantified the network differences on the edge-level using the "complex abundance" parameter, which quantifies the ratio (HCT/HEK) of the summed abundances of both interactors in the interaction region. Based on the distribution of these ratios we set a cutoff of 50% change (absolute log ratio > 0.58, Figure S3B) to call an edge quantitatively "differential" between the two conditions. We then mapped these differences to scale the edge color in our network (see figure S4 for the full network representation). A handful of complexes were found to have many differential edges between HEK cells and HCT cells (Figure 6E-F), representing potentially altered molecular functions in line with the above node-level GO analysis. For example, ribosomal proteins were driving the enrichment in GO terms "Structural constituent of ribosome" and "rRNA binding" in HCT cells and the GINS complex subunits are part of the GO term "catalytic activity acting on DNA" enriched in HEK cells. Others, however, were mostly quantified as un-changed, as seen in the example of the COP9 signalosome[3] (Figure 6G). Taken together, our data show SEC-TMT is compatible with downstream quantitative analyses geared towards identifying potential protein- and interaction-state differences in the PPI network.

Discussion

In this study we developed SEC-TMT, a multiplexed SEC-MS method for PPI identification and differential quantification, and showed that it performs as well as state-of-the-art label free SEC-MS methods, in a fraction of the time. Additionally, we displayed how careful multiplexing design can overcome coverage issues associated with TMT-based acquisition and position SEC-TMT as especially well-suited for quantifying differences in PPIs between samples. The strong reduction in measurement time combined with the quantitative advantages shown here are expected to bolster comparative studies of PPIs.

One of the most straight forward advantages of SEC-TMT is the vastly reduced instrument time, an improvement that can makes global protein interaction experiments accessible in non-specialized labs. In addition, SEC-TMT opens the possibility for larger study designs, comparing multiple conditions and exploring how PPI networks changes over a time-course or under multiple experimental and physiological conditions, to better understand the ramifications of PPI dynamics.

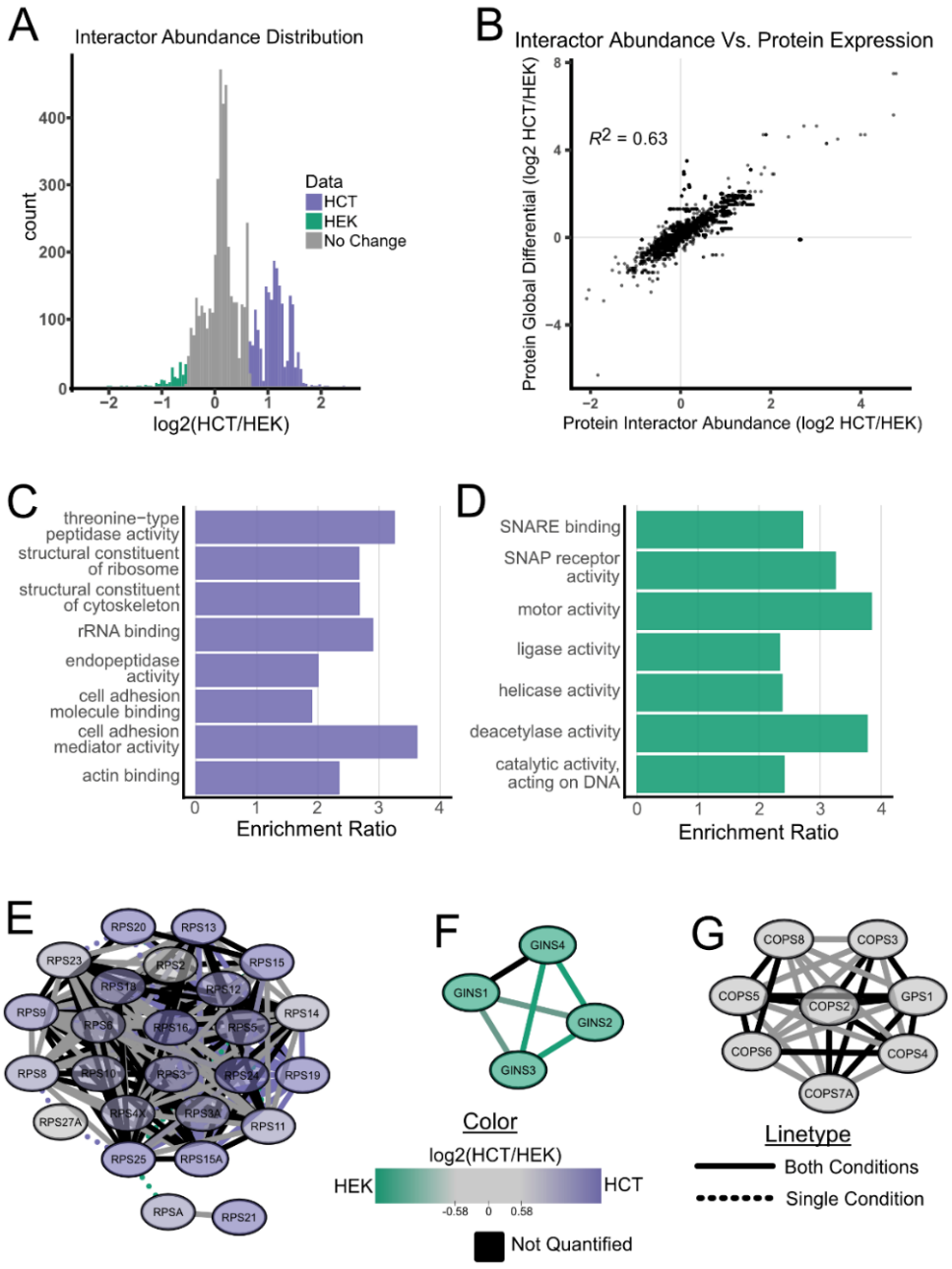


Figure 6. ***SEC-TMT enables quantitative differential comparison between multiple samples***
(A) Distribution of Log2 ratios of interactor abundance values between HCT and HEK cells for all nodes in the mutual network. Colors represent changes greater than 50% toward either HCT (purple) or HEK (green), or unchanged (gray). (B) Scatterplot of interactor abundance ratios against ratios of total expression as measured by DIA (log2, HCT/HEK) (C-D) GO terms (molecular function, non-redundant) found to be enriched in the list of nodes with differential interactor abundances (color codes as shown in B) at FDR < 0.1 (as analyzed by WebGestalt). (E-G) Examples of protein interaction networks for the small ribosomal subunit (E), the GINS complex (F), and the COP9 signalosome complex (G). Edge line type represents interactions that were only identified in a single condition (dotted, purple for HCT, green for HEK) or in both (solid). For the solid edges, edge color is scaled by ratio (HCT/HEK) of complex abundance values as shown in the scale bar (purple - stronger in HCT, green - stronger in HEK, gray - unchanged). Solid black edges represent interactions that were identified as significant in both cell-lines but were not assigned a complex abundance value by SECAT. Nodes are colored on the same scale, based on HCT/HEK ratios of global expression as measured by DIA quantification of the non-fractionated original sample.

In addition, SEC-TMT opens potential avenues for previously unexplored types of protein interactions studies. By multiplexing 18 fractions together, the effective initial input amount required per fraction is theoretically reduced by a factor of 18 (or 9 in the full-overlap scheme), potentially making it possible to investigate rare samples. Alternatively, the reduced input may allow enrichment of chemical modifications (e.g, phosphorylation, acetylation, and others) downstream from SEC and enable post translational modification (PTM) mapping on top of the SEC data layer – opening the possibility of distinguishing between the interactions of different protein isoforms to analyze how PTMs affect PPIs.

However, there are some caveats and disadvantages to consider when choosing between label-free or TMT based SEC analyses. First, the cost of the TMT labeling reagent is high and it increases the price of these experiments significantly compared to label free SEC-MS (however, this is to a large part off-set by the reduced instrument time needed). Second, even with the full overlap multiplexing scheme developed here, SEC-TMT still does not reach the same coverage as SEC-DIA and is biased towards analysis of the more abundant proteins in the sample. However, the provided "full-overlap" mixing schemes largely mitigate issues with coverage. Third, SEC-TMT also requires more sample processing steps, including labeling, generation and measurement of labeling controls, and additional signal processing and data handling steps before using PPI analysis algorithms like SECAT, EPIC, PRINCE, or CCprofiler. All of these additional steps require time and are potentially prone to the introduction of errors and should be performed with consideration. However, using the direct TMT labeling protocol described here, sample processing is considerably reduced and our scripts for the additional data processing steps are available.

In conclusion, this study laid the ground for multiplexing CF samples to significantly reduce the required measurement time and offer advantages for differential PPI quantification. We hope that this advancement will pave the way for more labs to explore PPI networks in a dynamic manner in order to enhance our understanding of the molecular architecture of the cell and provide new insights into molecular functions and regulatory mechanisms involved in protein regulation.

Materials and Methods

Cell culture

HEK293XT cells (Takara Bio Lenti-X 293T, #632180) were a provided by the Yeo lab at UC San Diego (SEC-DIA versus SEC-TMT experiments), or purchased from ATCC (ATCC, CRL-3216, in HEK vsersus HCT experiments), HCT116 cells were provided by the Prives lab at Columbia University. HEK cells were cultured in DMEM (containing L-glutamine and Sodium pyruvate) and HCT in McCoy's 5A media. Both media were supplemented with 10% Fetal Bovine Serum and Penicillin (100 U/mL) Streptomycin (100 µg/mL). Cells were grown to 80-90% confluency and were harvested at passages 6-20.

Sample preparation for SEC

SEC sample preparation was as previously described in Bludau et al. 2020[25]. Cells (25-40 million per sample) were harvested by scraping in ice cold PBS, washed and pelleted. Pellets were flash frozen in liquid nitrogen and stored in -80 °C. Upon thawing, cell pellets were lysed in cold lysis buffer (for TMT versus DIA comparisons: 150mM NaCl, 50mM Tris pH 7.5, 1% IGPAL-CA-630, 5% Glycerol; for HEK-HCT experiments: 50 mM HEPES pH 7.5, 150 mM NaCl, 0.5% NP40) supplemented with 50mM NaF, 2mM Na_3VO_4, 1mM PMSF, and 1X protease inhibitor cocktail (Sigma), followed by 10-30 minutes incubation on ice with intermittent vortexing. Cell lysates was then pre-cleared by 10 minutes centrifugation at 10,000g (4 °C) followed by 20 minutes of ultracentrifugation at 100,000 g, 4 °C. To dilute detergents in the buffer, samples underwent buffer exchange on Amicon® ultra-0.5 centrifugal filter with 30 kDa molecular weight cutoff (Sigma) into 50 mM HEPES pH 7.5, 150 mM NaCl and 50 mM NaF in iterative steps of no larger than 1:3 dilutions. The final dilution ratio of the original lysis buffer to detergent free buffer was 1:50. The cell lysate was further cleared by 5 minutes of centrifugation at 17,000 g, 4 °C. The concentration of the supernatant was measured by Nanodrop spectrophotometer (Thermo Scientific) and adjusted to 20-50mg/ml. Two mg of lysate were loaded on the SEC column per run.

SEC

Size exclusion was conducted on an Agilent 1260 Infinity II system operated with Agilent OpenLAB ChemStation software (version C.01.09). Two mg of cell lysate at 20-50mg/ml were loaded onto a Yarra SEC-4000 column (Phenomenex 00H-4514-K0, 3µm silica particles, 500A pores, column dimensions: 300 x 7.8mm) and fractionated in SEC running buffer (50 mM HEPES pH 7.5, 150 mM NaCl) at a flow rate of 1ml/min (first TMT experiment) or 0.5ml/min (all other experiments) and 100µL fractions were collected between minutes 6.5 to 16 or 11 to 30, respectively into 96 Well DeepWell Polypropylene Microplates (Thermo Scientific).

Protein digestion and desalting

Following SEC fractionation protein concentration was measured using the Pierce™ BCA protein assay kit (Thermo Scientific) based on the manufacturer's instructions. Equal volumes (~80µL) from each of the fractions containing proteins (54 - 72) were subsequently processed. Proteins were denatured by incubation with an equal volume of urea buffer containing 8 M urea, 75 mM NaCl, 50 mM HEPES (pH 8.5) and 1 mM EDTA at 25 °C, 600 rpm for 20 mins in 96 Well DeepWell Polypropylene Microplates (Thermo Scientific). Proteins were then reduced with 5 mM DTT at 25 °C, 600 rpm for 45 minutes and then alkylated with 10 mM iodoacetamide (IAA) at 25 °C, 600 rpm for 45 mins in the dark. Proteins were then diluted in a ratio of 1:3 with 50 mM HEPES (pH 8.5) to lower the urea concentration less than 2M, and digested with

trypsin enzyme (Promega) at 25 °C, 600 rpm overnight using 1:50 (enzyme: substrate) ratio. Digested peptides were acidified using formic acid and desalted on in-house packed C18 StageTips (two plugs) on top of 96 Well DeepWell Polypropylene Microplates as elaborated in Rappsilber et al., 2007[26]. For DIA measurements, dried peptides were reconstituted to a final concentration of 0.5µg/µL with 3% acetonitrile/ 0.2% formic acid. For TMT labeling purposes dried peptides were reconstituted in 50 mM HEPES (pH 8.5). An aliquot of 0.2mg of each non-fractionated sample (for HEK and HCT global protein expression analysis using DIA) was processed in a similar manner.

For the HEK-HCT dataset we used a direct labeling method. Following fraction selection based on BCA measurements, proteins were denatured by incubation at 95 °C, 600 rpm for 10 mins, followed by two cycles of 1 minute bath sonication. After samples cooled down to room temperature, proteins were reduced with 5 mM DTT at 25 °C, 600 rpm for 45 minutes and then alkylated with 10 mM iodoacetamide (IAA) at 25 °C, 600 rpm for 45 mins in the dark. Proteins were then diluted in a ratio of 1:3 with 50 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS), pH 9.0. pH was adjusted to ~ 8.2, and samples were subsequently digested with trypsin enzyme (Promega) at 25 °C, rpm 600 overnight using 1:50 (enzyme: substrate) ratio.

TMT labeling

For SEC-TMT experiments used to compare to SEC-DIA samples were digested and desalted as elaborated above. The resulting peptides were reconstituted with 50 mM HEPES (pH 8.5). For direct labeling peptides were labeled in the adjusted digestion buffer (50mM EPPS, pH adjusted to ~8.2). For all samples, peptides were labeled by addition of TMTpro™ 18 plex reagents (Thermo Scientific) into the sample at a ratio of 1:3 (peptide: TMT) by mass in a final volume of 29% acetonitrile. The labeling reaction was incubated at 25 °C, 600 rpm for 1 hour before being quenched with a final concentration of 0.3% hydroxylamine. Samples were then pooled as described in the pooling scheme and dried at least half of the volume to lower the acetonitrile concentration to less than 5%. The labeled peptides were then acidified using formic acid (pH <3) and desalted on C18 StageTips (two plugs)[26]. The desalted peptides were dried and resuspended in 3% acetonitrile/ 0.2% formic acid for subsequent liquid chromatography-tandem mass spectrometry (LC-MS/MS) processing.

LC-MS/MS

LC-MS/MS analysis was performed on a Q-Exactive HF. 5µL of total peptides were analyzed on a Waters M-Class UPLC using a C18 25cm Thermo EASY-Spray column (2um, 100A, 75um x 25cm) or IonOpticks Aurora ultimate column (1.7um, 75um x 25cm) coupled to a benchtop ThermoFisher Scientific Orbitrap Q Exactive HF mass spectrometer. Peptides were separated at a flow rate of 400 nL/min with the following gradients: 70 minutes (SEC-DIA), 160 minutes (SEC-TMT and DIA runs for non-fractionated samples), all including sample loading and column equilibration times. For DIA runs MS1 Spectra were measured with a resolution of 120,000, an AGC target of $5e^6$ and a mass range from 350 to 1650 m/z. 15 isolation windows of 87 m/z were measured at a resolution of 30,000, an AGC target of $3e^6$, normalized collision energies of 22.5, 25, 27.5, and a fixed first mass of 200 m/z. For DDA runs MS1 Spectra were measured with a resolution of 120,000, an AGC target of $3e^6$ and a mass range from 300 to 1800 m/z. Top12 MS2 spectra were acquired at a resolution of 60,000, an AGC target of $1e^5$, an isolation window of 0.8m/z, normalized collision energies of 28, and a fixed first mass of 110 m/z.

Data analysis

Searches

Proteomics raw data were analyzed using the directDIA method on SpectroNaut v16.0 for DIA runs or SpectroMine (3.2.220222.52329) for DDA runs (Biognosys) using a human UniProt database (Homo sapiens, UP000005640), under BSG factory settings modified without automatic cross-run normalization or imputation for SEC runs. Cross run median normalization and global imputation were used for global expression analysis (HEK-HCT non fractionated samples). Peptide spectral matches (PSMs), peptides and protein group data were exported for subsequent analysis.

signal processing

The peptide intensities were spread out along 57 SEC fractions. Peptides were filtered by being proteotypic and non-decoy. Empty or NA measurments were converted to zeros[22], and a single uniprot ID was assigned to each peptide. In TMT experiments, peptide reporter intensity values were normalized to their respective MS1 peak intensity. In experiments conducted with the full overlap TMT mixing scheme, TMT batch effects were corrected based on the signal in the common fractions between any two adjacent mixes. A normalization factor was calculated by dividing the peptide fraction intensities of mix [n+1] by mix [n], then taking the median of all the peptides and the mean of all the overlapping fractions in common between the mixes. Mix [n+1] was then normalized to mix [n] by multiplying all intensities by the normalization factor. Lastly, the peptide intensities of overlapping fractions were averaged.

SECAT

SECAT was used to identify previously reported protein interactions[13]. Replicates were analyzed in the same run to leverage the predictive power of the classifier. SECAT analysis was conducted on the processed peptide level signal (as mentioned above) using the default (SECAT provided) positive and negative interaction networks for the training step, and a target database of STRING's human interactions (9606.protein.links.v11.5) for the query step. The default SECAT parameters were set except for a 'pi0_lambda' of 0.4 0 0 0, an 'ss_initial_fdr' of 0.5 and 'ss_iteration_fdr' of 0.2 during the 'learn' step. Additionally, the 'export_tables' option of the SECAT 'learn' step was used to export tables for extracting the STRING target and learning interactions along with their scores. The HEK-HCT data was also quantified by setting HEK as the 'control_condition', and using a 'maximum_interaciton_qvalue' of 0.1 for the quantify and export steps.

The networks were obtained by setting a q-value cutoff of 0.05 on the exported network tables. For interactions that made the $q<0.05$ cutoff, a second cutoff was set to allow the interaction from the other condition if it was at least $q<0.1$. The HEK-HCT differential networks were obtained by merging the networks with the differential edge and protein tables to obtain the interactor abundance and complex abundance values for the obtained interactions.

EPIC

The EPIC tool was used to identify high confidence interactions allowing the discovery of novel interactions[15]. Replicates were analyzed in the same run to leverage the predictive power of the classifier. Peptides were collapsed to the protein level by adding the top three peptide intensities for each protein.

Proteins that eluted in only one fraction were filtered out. Pairwise protein-protein similarities were then computed using the Pearson Correlation-Coefficent (with and without noise), Jaccard, Apex, Mutual Information, and Euclidean metrics respectively. A cutoff of 0.5 for the features was chosen prior to analysis by a Random Forest Classifier which was trained on reference complexes generated using CORUM, INTact, and GO human proteins. The classifier was trained using an 80/20 cross validation split to minimize variance across runs and maximize predictive capabilities. Finally, de-novo protein-protein interactions were found by querying the classifier and reporting every interaction above 50% confidence as an interaction. To further benchmark the classifier a precision-recall graph was generated by varying the confidence of the classifier and reporting the metrics, the intersection of the precision and recall occurs at 60% confidence. However, as we are trying to minimize false positives, we picked a higher confidence of 80% (as previously reported by Pourhaghighi et al., 2020[11]) which has less but more precise interactions.

Network analysis (cytoscape)

The cytoscape networks were generated by importing the interaction networks (full network and specific CORUM complex networks) and then importing the node tables to get gene names and log2 fold changes for each protein. The log2 fold changes of complex abundances were set to color the edges and a separate global protein expression differential was measured and calculated to get a log2 fold change for each protein. The linetype was set by translating the protein identification on either HEK, HCT, or both conditions into a column in the network import table. Network statistics were extracted from cytoscape using the default parameters.

GO term enrichment analysis

Gene ontology enrichment analysis was performed using the WebGetalt (http://www.webgestalt.org) platform using the over-representation analysis (ORA) on molecular function non redundant terms[24]. Enriched set was compared to a background list containing all the proteins identified in the experiment.

Acknowledgements

B.J.B. was supported by NSF-GRFP (Award DGE2036197). L.C.T. was supported by NSF-GRFP (Award DGE2036197). M.J. is funded by the NIH (R35GM128802; R01AG071869 and R01HG012216), NSF (Award 2224211) and Columbia startup funding. We thank the Yeo lab (UC San Diego) for kindly gifting us HEK293T cells, and the Prives lab (Columbia University) for their kind gift of HCT116 cells. We thank Cassandra A. Chartier for her assistance with HPLC.

Declaration of interests

The authors declare no competing interests.

References

1. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).
2. Go, C. D. *et al.* A proximity-dependent biotinylation map of a human cell. *Nature* **595**, 120–124 (2021).
3. Huttlin, E. L. *et al.* Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell* **184**, 3022-3040.e28 (2021).
4. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
5. Liu, X. *et al.* An AP-MS- and BioID-compatible MAC-tag enables comprehensive mapping of protein interactions and subcellular localizations. *Nat. Commun.* **9**, 1188 (2018).
6. Havugimana, P. C. *et al.* A Census of Human Soluble Protein Complexes. *Cell* **150**, 1068–1081 (2012).
7. Havugimana, P. C. *et al.* Scalable multiplex co-fractionation/mass spectrometry platform for accelerated protein interactome discovery. *Nat. Commun.* **13**, 4043 (2022).
8. Kirkwood, K. J., Ahmad, Y., Larance, M. & Lamond, A. I. Characterization of Native Protein Complexes and Protein Isoform Variation Using Size-fractionation-based Quantitative Proteomics. *Mol. Cell. Proteomics* **12**, 3851–3873 (2013).
9. Larance, M., Ahmad, Y., Kirkwood, K. J., Ly, T. & Lamond, A. I. Global Subcellular Characterization of Protein Degradation Using Quantitative Proteomics. *Mol. Cell. Proteomics MCP* **12**, 638–650 (2013).
10. Heusel, M. *et al.* Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.* **15**, e8438 (2019).
11. Pourhaghighi, R. *et al.* BraInMap Elucidates the Macromolecular Connectivity Landscape of Mammalian Brain. *Cell Syst.* **10**, 333-350.e14 (2020).
12. Skinnider, M. A. *et al.* An atlas of protein-protein interactions across mouse tissues. *Cell* **184**, 4073-4089.e17 (2021).
13. Rosenberger, G. *et al.* SECAT: Quantifying Protein Complex Dynamics across Cell States by Network-Centric Analysis of SEC-SWATH-MS Profiles. *Cell Syst.* **11**, 589-607.e8 (2020).
14. Fossati, A. *et al.* PCprophet: a framework for protein complex prediction and differential analysis using proteomic data. *Nat. Methods* **18**, 520–527 (2021).
15. Hu, L. Z. *et al.* EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* **16**, 737–742 (2019).
16. Skinnider, M. A., Cai, C., Stacey, R. G. & Foster, L. J. PrInCE: an R/Bioconductor package for protein–protein interaction network inference from co-fractionation mass spectrometry data. *Bioinformatics* **37**, 2775–2777 (2021).
17. Kelmer Sacramento, E. *et al.* Reduced proteasome activity in the aging brain results in ribosome stoichiometry loss and aggregation. *Mol. Syst. Biol.* **16**, e9596 (2020).
18. Heusel, M. *et al.* A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133-155.e6 (2020).
19. Kim, M. *et al.* A protein interaction landscape of breast cancer. *Science* **374**, eabf3066 (2021).
20. Swaney, D. L. *et al.* A protein network map of head and neck cancer reveals PIK3CA mutant drug sensitivity. *Science* **374**, eabf2911 (2021).
21. Bludau, I. *et al.* Systematic detection of functional proteoform groups from bottom-up proteomic datasets. *Nat. Commun.* **12**, 3810 (2021).
22. Skinnider, M. A. & Foster, L. J. Meta-analysis defines principles for the design and analysis of co-fractionation mass spectrometry experiments. *Nat. Methods* **18**, 806–815 (2021).

23. Wheat, A. *et al.* Protein interaction landscapes revealed by advanced in vivo cross-linking–mass spectrometry. *Proc. Natl. Acad. Sci.* **118**, e2023360118 (2021).
24. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
25. Bludau, I. *et al.* Complex-centric proteome profiling by SEC-SWATH-MS for the parallel detection of hundreds of protein complexes. *Nat. Protoc.* **15**, 2341–2386 (2020).
26. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

Supplementary Figures

A

HEK-TMT Mix Scheme

channel \ mix	mix1	mix2	mix3	mix4	mix5	mix6	mix7	mix8
126C	1	10	19	28	37	46	55	64
127N	64	1	10	19	28	37	46	55
127C	2	11	20	29	38	47	56	65
128N	65	2	11	20	29	38	47	56
128C	3	12	21	30	39	48	57	66
129N	66	3	12	21	30	39	48	57
129C	4	13	22	31	40	49	58	67
130N	67	4	13	22	31	40	49	58
130C	5	14	23	32	41	50	59	68
131N	68	5	14	23	32	41	50	59
131C	6	15	24	33	42	51	60	69
132N	69	6	15	24	33	42	51	60
132C	7	16	25	34	43	52	61	70
133N	70	7	16	25	34	43	52	61
133C	8	17	26	35	44	53	62	71
134N	71	8	17	26	35	44	53	62
134C	9	18	27	36	45	54	63	72
135	72	9	18	27	36	45	54	63

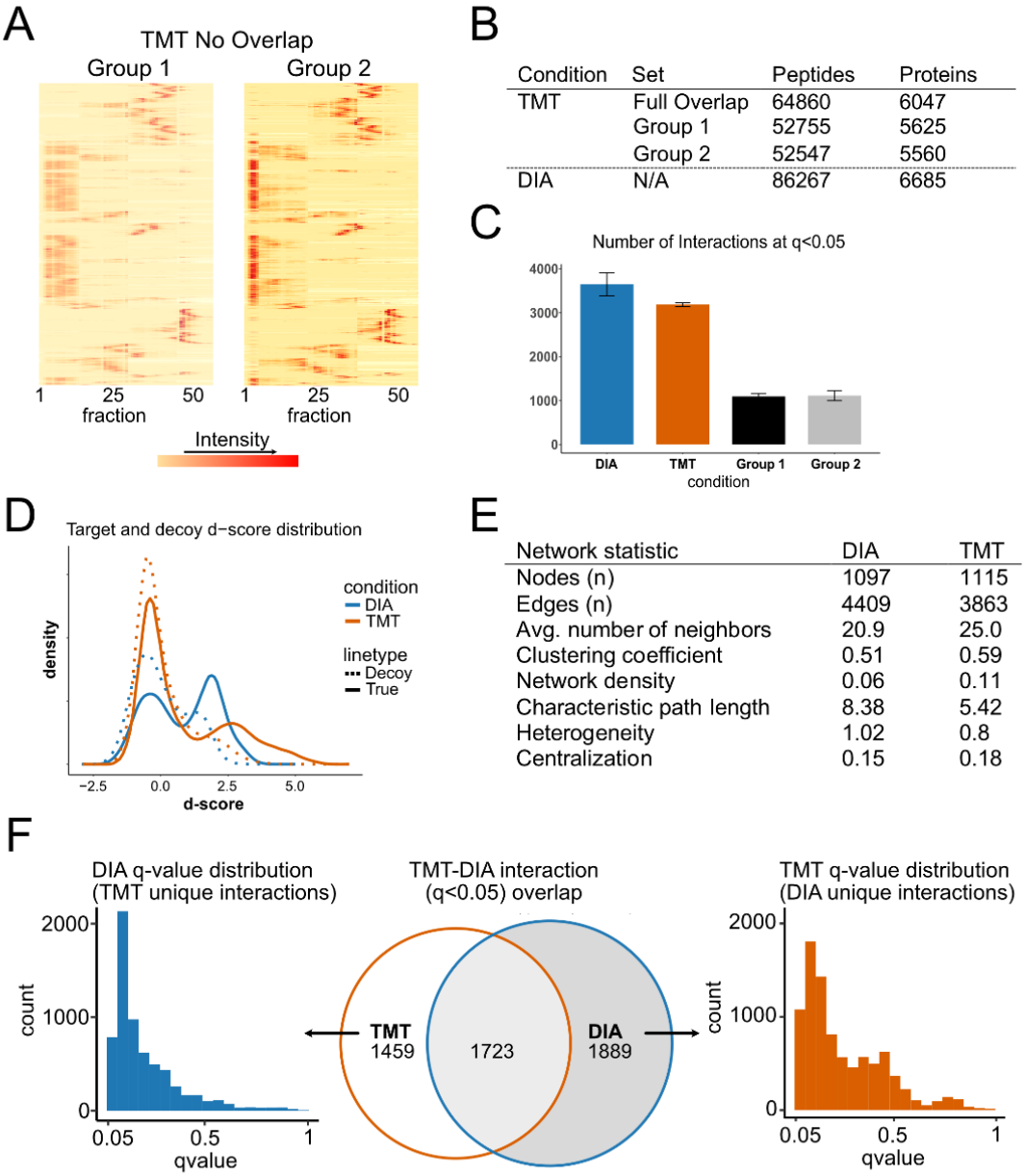
B

HEK-HCT Mix Scheme

channel \ mix	Cell-type	mix1	mix2	mix3	mix4	mix5	mix6	mix7	mix8	mix9	mix10	mix11	mix12
126	HEK	1	10	10	19	19	28	28	37	37	46	46	1
127N	HCT	1	10	10	19	19	28	28	37	37	46	46	1
127C	HEK	2	11	11	20	20	29	29	38	38	47	47	2
128N	HCT	2	11	11	20	20	29	29	38	38	47	47	2
128C	HEK	3	12	12	21	21	30	30	39	39	48	48	3
129N	HCT	3	12	12	21	21	30	30	39	39	48	48	3
129C	HEK	4	13	13	22	22	31	31	40	40	49	49	4
130N	HCT	4	13	13	22	22	31	31	40	40	49	49	4
130C	HEK	5	5	14	14	23	23	32	32	41	41	50	50
131N	HCT	5	5	14	14	23	23	32	32	41	41	50	50
131C	HEK	6	6	15	15	24	24	33	33	42	42	51	51
132N	HCT	6	6	15	15	24	24	33	33	42	42	51	51
132C	HEK	7	7	16	16	25	25	34	34	43	43	52	52
133N	HCT	7	7	16	16	25	25	34	34	43	43	52	52
133C	HEK	8	8	17	17	26	26	35	35	44	44	53	53
134N	HCT	8	8	17	17	26	26	35	35	44	44	53	53
134C	HEK	9	9	18	18	27	27	36	36	45	45	54	54
135	HCT	9	9	18	18	27	27	36	36	45	45	54	54

Supplemental Figure : Mixing schemes

(A) TMT-18 mixing scheme for the HEK cells only experiments with the rows representing channels and columns representing the mixing of the TMT channels for MS runs. (B) TMT-18 mixing scheme for the HEK cells versus HCT cells experiment with the rows representing channels and columns representing the mixing of the TMT channels for MS runs.

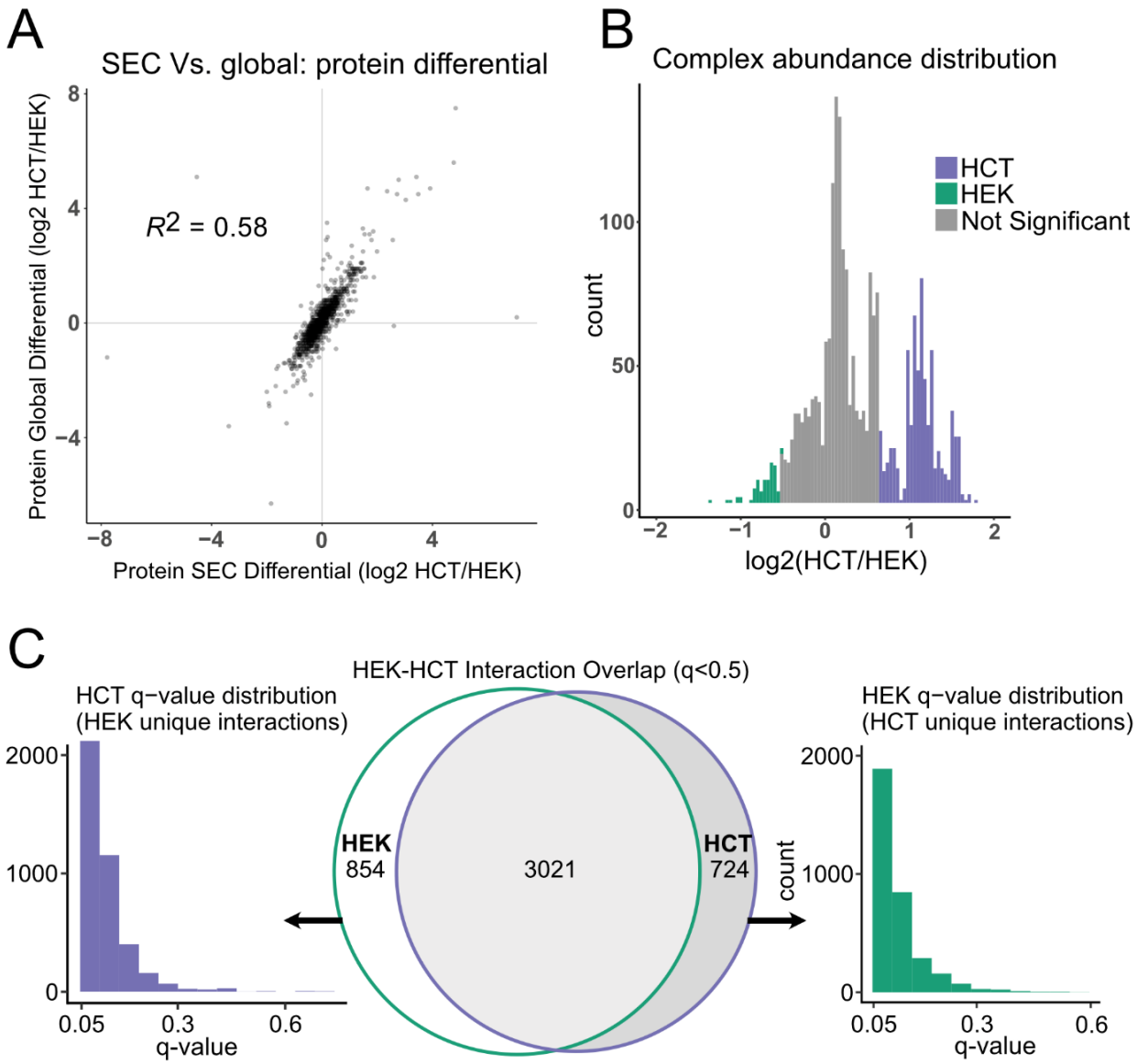


Condition	Set	Peptides	Proteins
TMT	Full Overlap	64860	6047
	Group 1	52755	5625
	Group 2	52547	5560
DIA	N/A	86267	6685

Network statistic	DIA	TMT
Nodes (n)	1097	1115
Edges (n)	4409	3863
Avg. number of neighbors	20.9	25.0
Clustering coefficient	0.51	0.59
Network density	0.06	0.11
Characteristic path length	8.38	5.42
Heterogeneity	1.02	0.8
Centralization	0.15	0.18

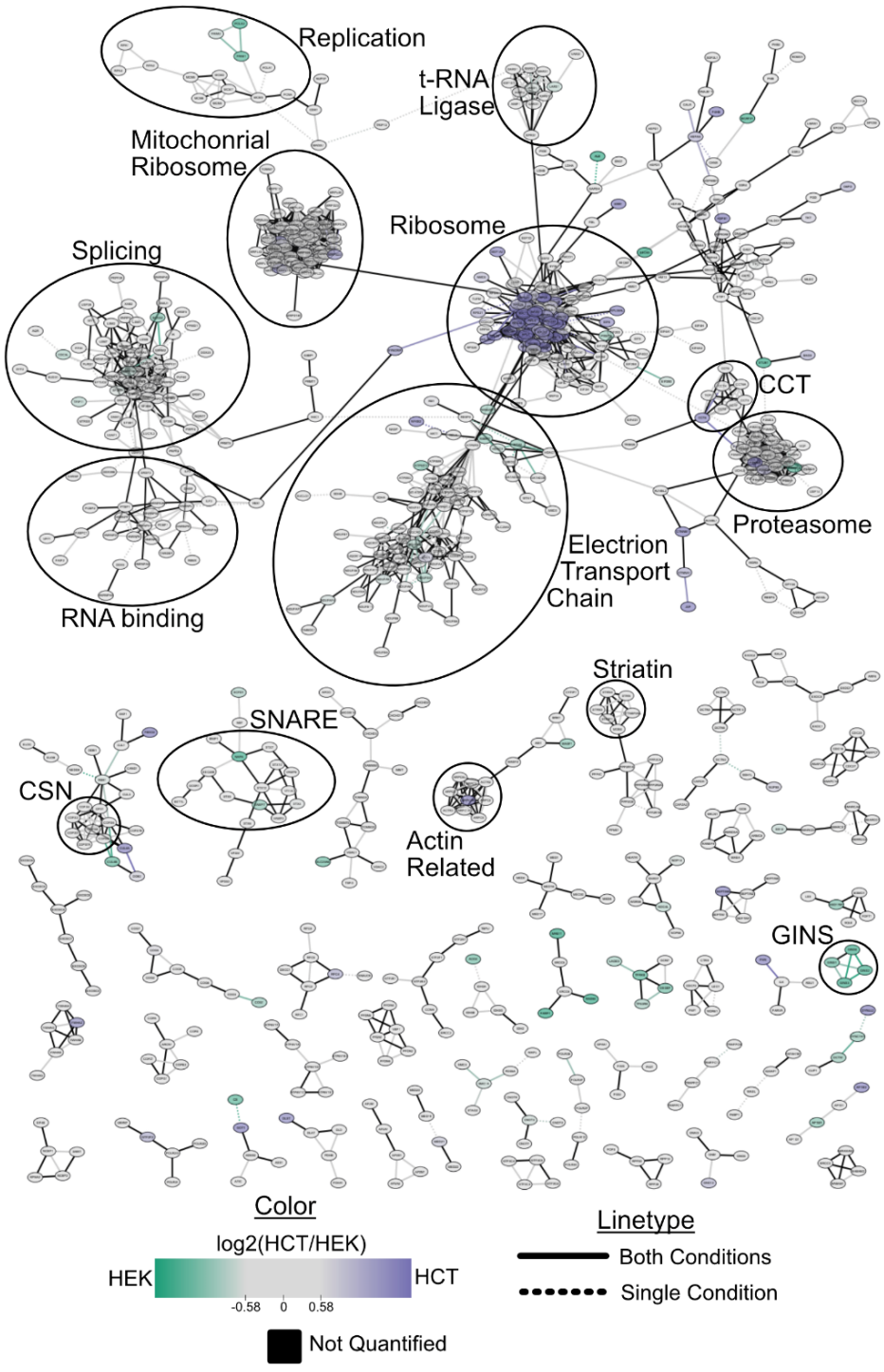
Supplementary Figure 2 - related to figure 2

(A) Heatmap representation of signals in SEC-TMT in the “no-overlap” multiplexing scheme as explained in figure 1B, for proteins measured in all 4 conditions (SEC-DIA, SEC-TMT full overlap and both groups of SEC-TMT no overlap). Columns represent fractions, rows represent different proteins and are scaled from 0 to 1 so that the max elution peak per protein is represented in red. Rows are in the same order as in figure 2A. (B) The overall peptide and protein identifications of each SEC-DIA and SEC-TMT biological replicate. (C) The number of interactions identified in SEC-DIA (blue), SEC-TMT full overlap (orange), SEC-TMT no overlap group 1 (black), SEC-TMT no overlap group 2 (gray). Mean +/- std deviation, n = 5 SECAT runs. (D) The density distribution of SECAT’s classifier determined d-scores for both the true positive (solid line) and decoy (dotted line) interactions for SEC-DIA (blue) and SEC-TMT full overlap (orange). (E) Extended list of network statistics from Cytoscape for both the SEC-DIA and SEC-TMT networks. (F) Interaction overlap in SEC-TMT and SEC-DIA using a strict q-value < 0.05 cutoff. For the interactions that were found to be significant in one dataset but not the other (“unique”) , the distribution of q-values in the non-significant dataset were plotted. Blue for the distribution of “TMT-unique” interactions in the DIA dataset, and orange for the “DIA-unique” interactions in the TMT dataset. Q-values were subsequently adjusted based on these distributions to include any interactions with q-value < 0.1, if its q-value was < 0.05 in the other condition (in 3 out of 5 SECAT runs).



Supplementary Figure 3 - related to figure 5

(A) Scatterplot comparing the differential expression of proteins as determined in the DIA measurements of the non-fractionated original sample (y-axis) compared to their "total abundance" in SEC-TMT as calculated by SECAT (x-axis). (B) Distribution of Log2 HCT/HEK ratios of complex abundance values for edges quantified in the mutual network. Colors represent change greater than 50% toward either HCT (purple) or HEK (green), unchanged in gray. (C) Interaction overlap in HEK and HCT using a strict q-value < 0.05 cutoff. For the interactions that were found to be significant in one condition but not the other ("unique"), the distribution of q-values in the non-significant dataset were plotted. Purple for the distribution of "HEK-unique" interactions in the HCT dataset, and green in the "HCT-unique" interactions in the HEK dataset.



Supplementary Figure 4 - related to figure 6

HEK-HCT annotated network. Interaction network containing all clusters with at least four nodes. The nodes and edges are colored as green (HEK), gray (not significant), or purple (HCT) along a spectrum with absolute cutoffs of the complex abundance log2(HCT/HEK) fold change of at least 0.58 to be considered significant. Black = interaction not quantified by SECAT. The linetypes signify whether the interaction was detected in both conditions (solid) or in only HCT or HEK (dotted) using the q-value < 0.1 in both conditions but < 0.05 in at least one cutoff, as explained in the text. Specific protein clusters were manually circled and labeled by function and complex.