Characterization of antibody specificity using immunoprecipitation and mass spectrometry

Verify capture of antibody target(s), identify isoforms, off-targets, and interacting proteins, and estimate relative antibody performance for a given target

By Bhavin Patel, M.D.; Greg Potts, Ph.D.; Leigh Foster, B.S.; Alex Behling, M.S.; John Bucci, M.S.; John Rogers, Ph.D.
08/19/2016

Antibodies are used in a broad range of research and diagnostic applications for the enrichment, detection, and quantitation of proteins and their modifications. Hundreds of thousands of antibodies are commercially available against thousands of proteins, which are used in a variety of applications, including western blotting (WB), immunofluorescence (IF), immunoprecipitation (IP), flow cytometry (FC), chromatin IP (ChIP), and enzyme-linked immunoassays (ELISA). These antibodies may be monoclonal, polyclonal, or recombinant from different organisms, and they may be used to interrogate biological systems and signaling pathways, diagnose disease, and assess responses to treatment [1].

Unfortunately, many antibodies are poorly characterized, both initially and between manufacturing lots. This is due to three challenges: 1) the sheer number of available protein targets and antibodies can be overwhelming to consider validating, so efficient and thorough characterization strategies are required; 2) the complex nature of proteins, binding interactions, physiochemical conditions (such as native versus denatured epitope conformations), and biology for each antibody application and model system; and 3) the lack of consistent standard approaches and criteria to assess antibody selectivity [2, 3]. There is a great need to verify that antibodies recognize their intended targets and to ensure that the reagents are fit-for-purpose in a given application. Multiple recommendations for antibody validation have been proposed, and databases of consolidated antibody annotation and performance "scoring" information are freely accessible (e.g., Antibodypedia, http://www.proteinatlas.org/learn/method/antibodypedia, [4]). A variety of antibody validation criteria have been proposed, including: 1) the verification of antibody specificity with genetic knockdowns or blocking peptides; 2) validation of antibody detection results with different biological systems (e.g., target localization, expression in model systems, etc.); 3) correlation of antibody results between methods; and 4) demonstration of reproducibility between samples, labs, and manufacturing lots [3–5].

Recently, mass spectrometric (MS) approaches have been proposed for antibody target verification [6, 7]. Despite the cost and technical requirements of MS, of all existing validation methods, mass spectrometry has the unique ability to identify the actual antibody target(s), isoforms, posttranslational modifications, and target-associated proteins present within a sample. Only MS can identify and characterize antibodies with this level of depth and specificity. Unlike western blotting, ELISA, and other standard immunological methods that use blocking reagents (such as milk or bovine serum albumin, BSA) to minimize background protein binding, MS detects all proteins, specific and nonspecific, in a prepared sample. For example, immunoprecipitation with an immobilized antibody on a bead or resin is a common approach to enrich a target protein and associated proteins from a lysate or biofluid. Due to the low abundance of a specific target relative to common background proteins, nonspecifically bound proteins may overwhelm and interfere with or prevent the detection of a low-abundance target. Even with stringent wash conditions and optimized reagents, dozens to hundreds of background proteins are typically observed in an immunoprecipitated sample that is analyzed by mass spectrometry. Thus, while MS holds great promise for contextualizing targets amidst potential interacting proteins, the MS results from an immunoprecipitated sample can be difficult to filter and interpret since nonspecific background proteins will also be present. In particular, an assurance of antibody selectivity for a native protein from a real biological system is particularly challenging to demonstrate.

To address these issues and to assess the binding specificity of Invitrogen™ antibodies produced by Thermo Fisher Scientific, here we describe a new approach to antibody target verification. Through the use of optimized sample preparation reagents and methods, high-resolution MS instrumentation, and a novel data analysis pipeline, we have created a comprehensive workflow to assess antibody specificity for its intended target using immunoprecipitation combined with mass spectrometry (IP-MS, Figure 1). The benefits of this IP-MS approach include: identification of the antibody target(s), isoforms, and modifications, quantitative assessment of antibody selectivity by calculating fold-enrichment of targets and off-targets for the assessment of antibody selectivity, and identification of interacting proteins.

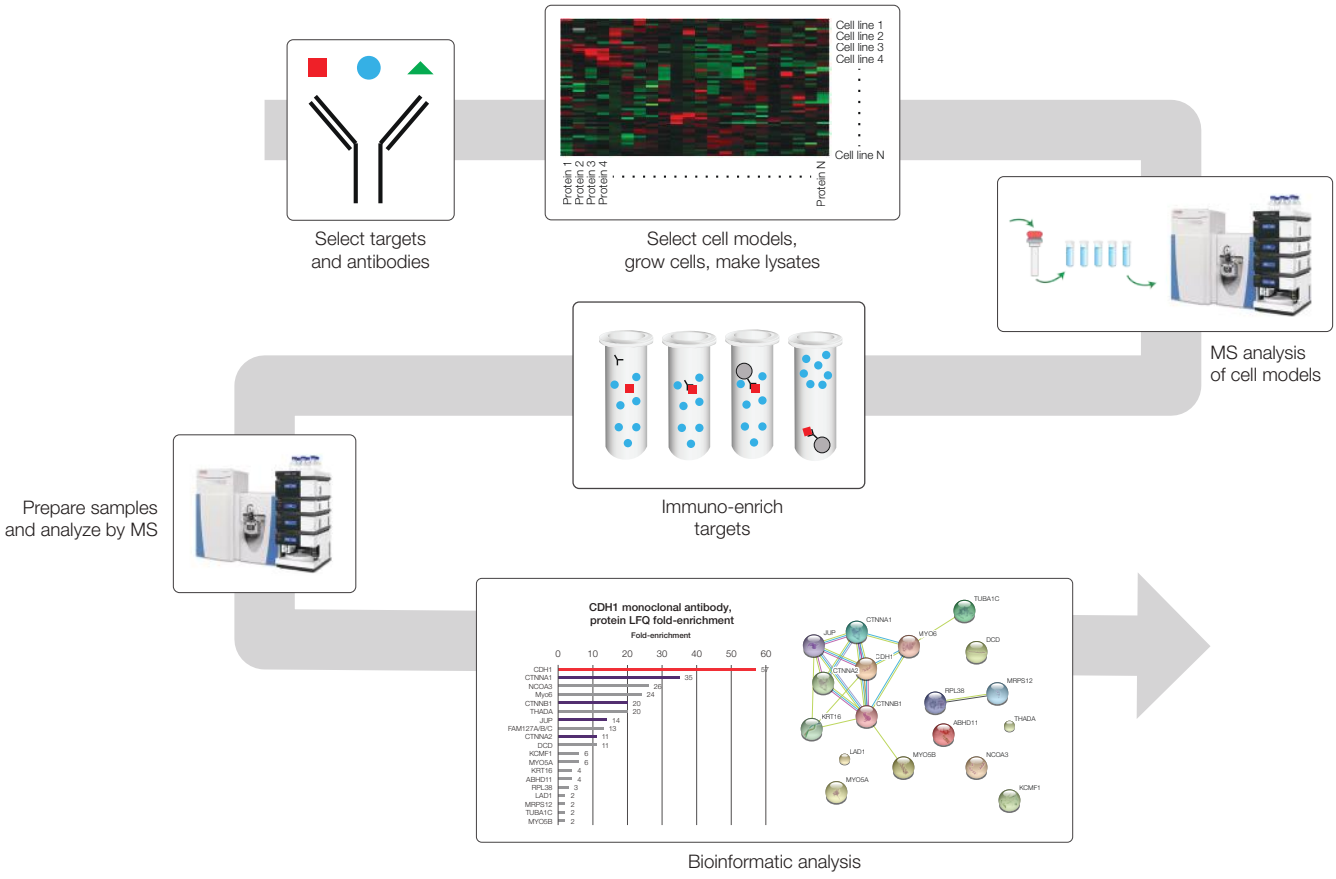


Figure 1. Experimental workflow for antibody verification by IP-MS analysis.

Results and discussion

Target and antibody selection

Thermo Fisher Scientific currently offers more than 48,000 antibodies to more than 20,000 proteins and protein modifications. The targets for these antibodies were prioritized based upon literature references, database mining, and consideration of signaling pathways and targeted genomic panels, such as Ion AmpliSeq™ panels for targeted gene amplification and next-generation DNA sequencing. For example, Figure 2 lists the top 1,000 most-referenced genes in the PubMed literature database. The *TP53* gene is the most highly referenced gene in PubMed, with more than 7,500 references. *TP53* encodes a key signaling protein with many known modifications and interacting proteins, and gain- and loss-of-function mutations in the *TP53* gene are involved in genetic instability and tumorigenesis. The next 999 most-referenced genes vary greatly in their reference frequency, so the distribution of antibody targets has a very long "tail". Antibody targets were chosen based on these considerations and criteria, and systematically verified through our IP-MS workflow.

Cell model selection

To identify candidate complementary cell lines for our selected protein targets, we utilized literature resources, public transcriptomic and proteomic databases, and biological samples likely to contain the most diverse and comprehensive set of protein targets to our prioritized pathways and panels. For instance, the CellMiner™ website is a public repository of exome sequences, transcriptomic, microRNA, and protein array expression analyses, and drug response results for the NCI60 cell lines (**http://discover.nci.nih.gov/cellminer/**).

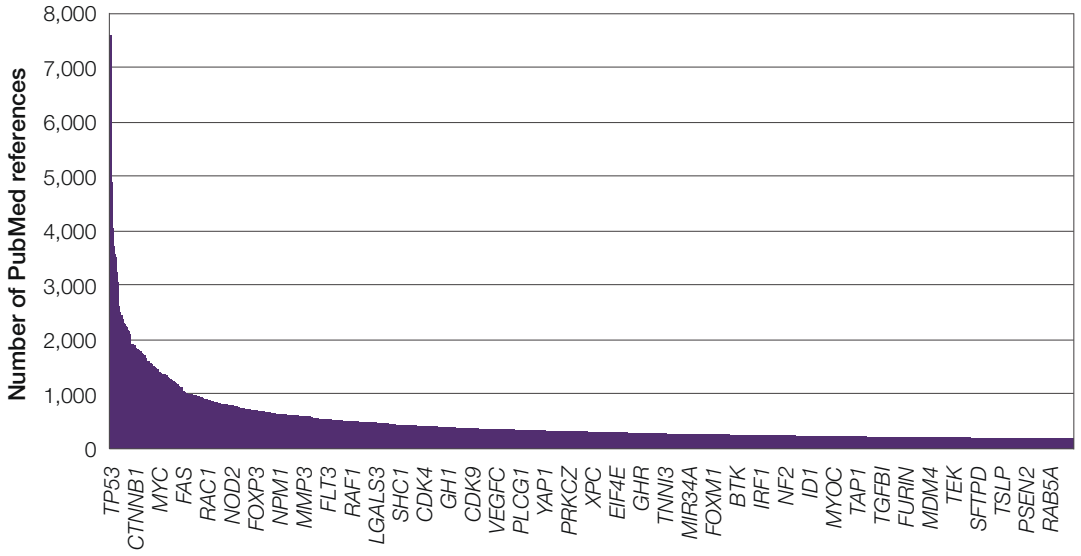


Figure 2. Representative list of the top 1,000 most-referenced genes in PubMed as of January 2016.

Hierarchical clustering of RNA transcript expression Z-scores for 22 targets in the Ion AmpliSeq™ Colon and Lung Cancer Panel across the NCI60 cell lines allowed us to identify 4 cell lines likely to express these 22 proteins (Figure 3A). Additionally, the ProteomicsDB and PRIDE websites are public repositories of MS-related proteomics data, including protein and peptide identifications, posttranslational modifications, and supporting spectral evidence (**https://www.proteomicsdb.org/** and **https://www.ebi.ac.uk/pride/archive/**). A combination of these resources was used to select 12 complementary cell lines likely to express more than 90% of the top 1,000 most-referenced genes for deep proteome analysis and IP-MS antibody verification.

Unfractionated and fractionated proteome analysis

These 12 selected human cell lines were grown in recommended conditions in order to generate protein lysates from each cell line for MS-based proteome analysis. Using optimized sample preparation methods and instrumentation, we generated proteome data of unfractionated and fractionated samples to confirm that the cell lines expressed our selected target proteins. Briefly, lysate proteins were solubilized, proteolytically digested with trypsin, and prepared for LC-MS analysis. Digested protein samples were directly analyzed by LC-MS (unfractionated) as well as fractionated using high-pH reverse-phase columns to improve the depth of proteome coverage in subsequent LC-MS analyses. Each unfractionated protein digest identified 3,800–4,500 unique protein groups, and each fractionated protein digest identified 7,500–9,000 unique protein groups. To assess our overall protein sequence coverage and overlap between cell lines, the identified proteins from each cell line were then compared to determine pairwise correlation scores. These correlations ranged from 75–95%. When the overall protein identifications from the 5 least-correlated cell lines were compared, 3,611 proteins were consistently identified, with an additional 900–1,500 proteins uniquely observed in each of the cell lines (Figure 3B).

To rapidly and effectively screen IP samples for the presence or absence of target proteins, we used the MS peptide peak areas of the top 3 peptides observed per protein, the number of unique peptides identified, and the spectral count detected with Thermo Scientific™ Proteome Discoverer™ Software (Cat. No. IQLAAEGABSFAKJMAUH). In addition to providing a comprehensive list of protein identifications from multiple cell lines, we also quantified the relative abundance of these proteins using label-free quantitation (LFQ) values from MaxQuant software (**http://www.biochem.mpg.de/5111795/maxquant**). Using these metrics, target protein expression was ranked and compared across the 12 cell lines. IP-MS antibody verification used this protein expression data to select cell lines that expressed target proteins at medium to low abundance. For example, E-cadherin (CDH1) and N-cadherin (CDH2) had complementary expression patterns across the 12 selected cells lines (Figure 4A). CDH1 was detected only in unfractionated HCT116, LNCaP, and MCF7 cells, while CDH2 was only seen in unfractionated A549, BT549, HEK293, and Hs578T cells.

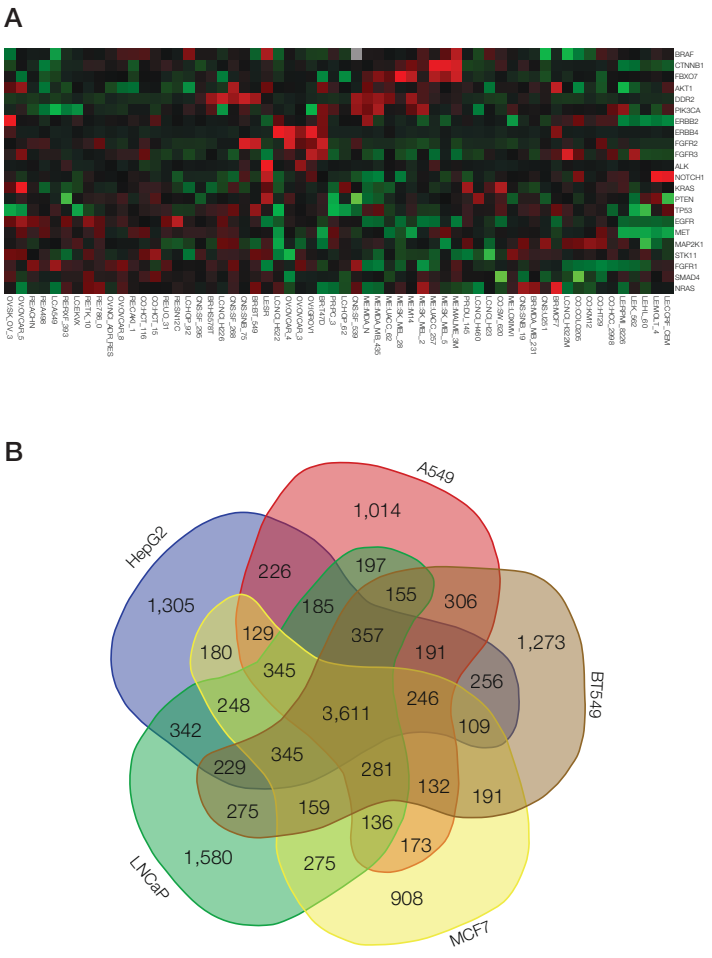


Figure 3. Selection of cell lines for antibody verification by IP-MS. (A) RNA expression Z-scores from the NCI60 cell line panel were hierarchically clustered for 22 genes in the Ion Ampliseq Colon and Lung Cancer Panel. **(B)** Venn diagram of the number of proteins identified from five NCI60 cell lines with mass spectrometric analysis of fractionated peptides from each lysate.

Both isoforms were detectable in several cell lines after fractionation and deep proteome MS analysis of these fractions (Figure 4A). Furthermore, cell lines with high expression of protein targets were deemed inappropriate models for antibody testing. Target proteins were ranked by their protein LFQ values in order to visualize protein target abundance within the context of the whole proteome and select appropriate cells lines for antibody screening. For example, CDH1 was ranked 1,184 of 4,638 proteins by protein LFQ in unfractionated MCF7 lysate and 956 of 7,176 proteins in the fractionated lysate (Figure 4B-a and Figure 4B-b). CDH2 was ranked 1,411 of 4,541 proteins in unfractionated A549 lysate and 1,521 of 7,252 proteins in the fractionated lysate, while CDH1 was only detectable in A549 lysate after fractionation (rank 3,965 of 7,252 proteins, Figure 4B-c and Figure 4B-d). This protein expression information was invaluable for the selection of cell models and assessment of isoform-specific and pan-specific antibody selectivity. As a result, one or more cell lines for each target protein were chosen for this study based upon MS identification and target protein abundance.

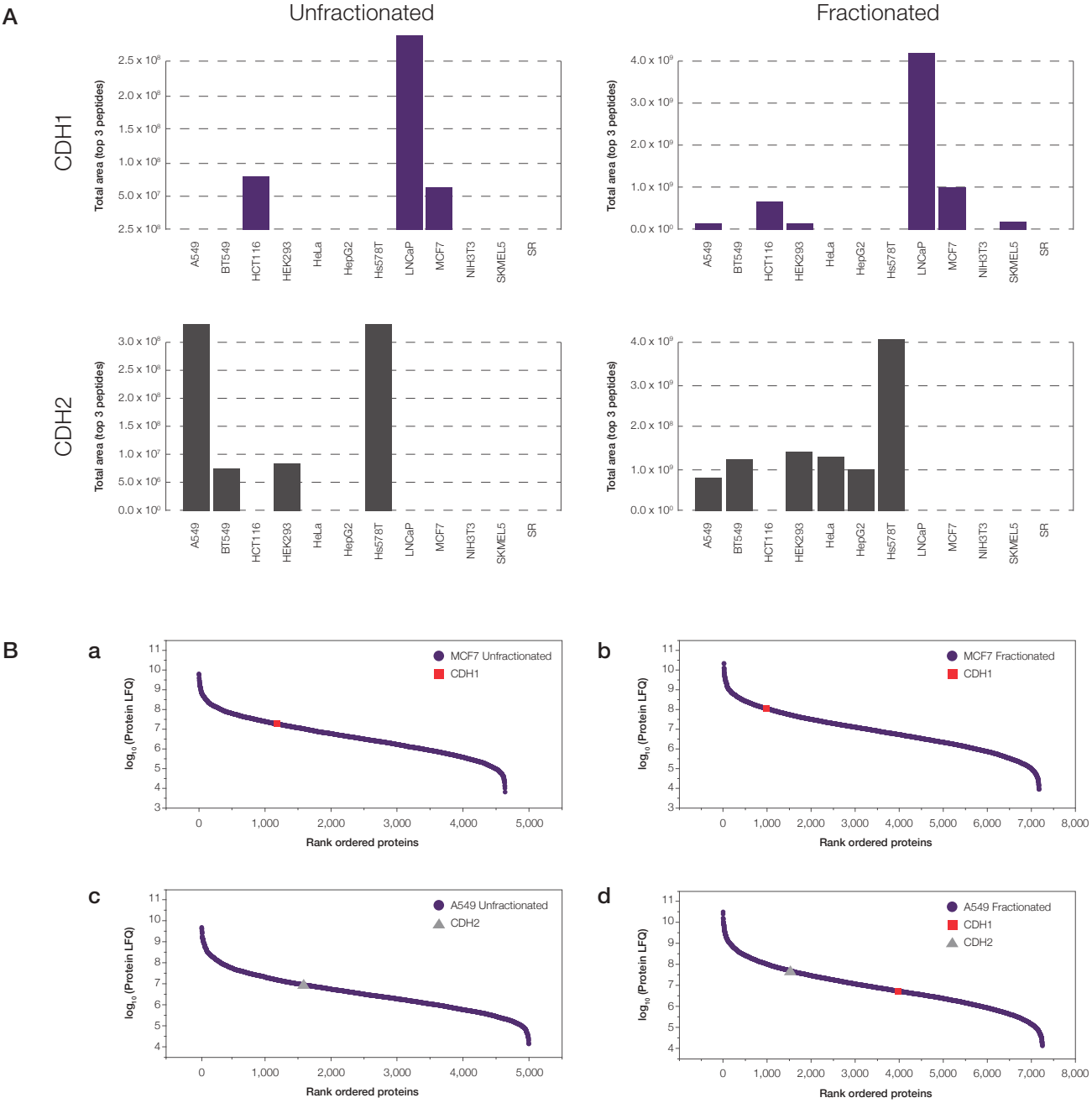


Figure 4. Cell line selection for CDH1 and CDH2. (A) Comparison of E-cadherin (CDH1) and N-cadherin (CDH2) protein expression across twelve cell lines without or with fractionation for deeper proteome analysis. **(B)** Distribution of expressed proteins detected in unfractionated and fractionated MCF7 and A549 cells, highlighting the expression levels of CDH1 and CHD2.

Target identification by IP-MS

We utilized protein expression profiles to assist in the assessment of antibodies by IP-MS. The key benefit of verifying an antibody's target by IP-MS is identification of not only the native target protein, but also its isoforms, posttranslational modifications, and interacting proteins. Historically, the results of this target identification can be assessed in several ways, including the number of unique peptides, protein sequence coverage, number of spectra observed for peptides from the target protein (spectral count), or integrated MS signal intensities from a subset or all of the detected peptides, as described above. The relative performance of various antibodies for the same target can be easily compared regardless of the measurement approach. For example, immunoprecipitations with 10 antibodies validated by a combination of IP and western blot to TP53 protein were assessed in replicates using the protein LFQ values (Figure 5). Results of replicate IP samples were highly reproducible, and all 10 antibodies showed reproducible MS signals (CV <25% across replicates). This IP-MS approach assesses antibody fit-for-purpose, provides definitive evidence of target protein capture, and readily permits antibody comparisons that may indicate relative antibody affinity.

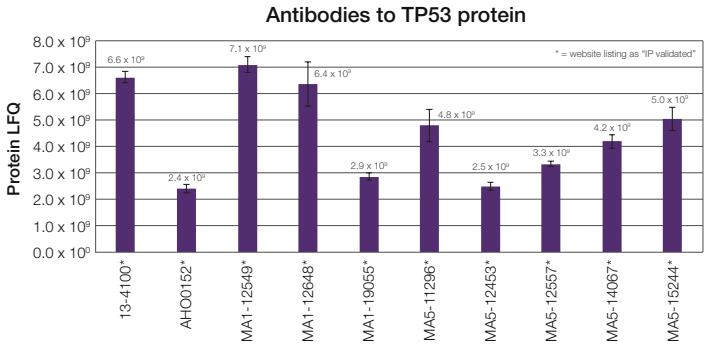
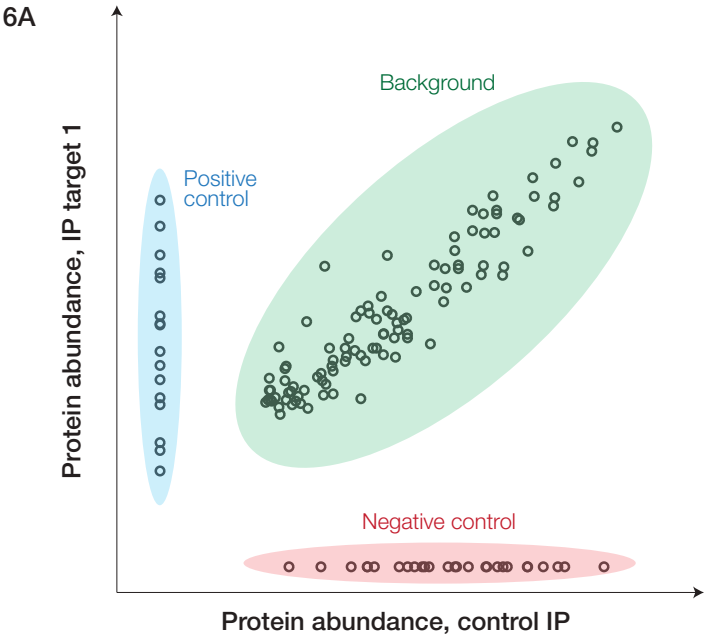


Figure 5. Reproducibility of TP53 antibody verification by IP-MS. TP53 protein was immunoprecipitated from BT-549 cell lysate with the indicated antibodies in replicate and quantified by MS using the protein label-free quantitation (LFQ) values of TP53 peptides. Asterisks highlight the antibodies annotated as IP validated by IP–western blot.

Background subtraction with scatter plots

As mentioned previously, protein IP with immobilized antibodies is a common method for targeted protein enrichment, but tens to hundreds of background proteins are commonly identified by mass spectrometry even after stringent washing conditions. Quantitative tools to analyze protein affinity capture results, such as COMPASS, SAINT, and Perseus, offer sophisticated scoring methods and data visualization for filtering protein identifications, but the implementation of these methods can be challenging and the results difficult to interpret [7, 9–11]. To better understand these background proteins and more easily identify specifically captured versus nonspecific proteins, we attempted to simplify the data representation by using protein LFQ values to quantitatively compare the proteins immunoprecipitated with a specific antibody versus a negative control antibody (Figure 6A). The resulting scatter plot of MS intensities had three clusters: 1) specifically captured proteins that were only observed with the test antibody after IP (Figure 6A, y-axis); 2) nonspecifically captured proteins only observed with the negative control antibody IP (Figure 6A, x-axis), and; 3) proteins distributed along the scatterplot diagonal, which represented common and highly abundant background proteins observed in



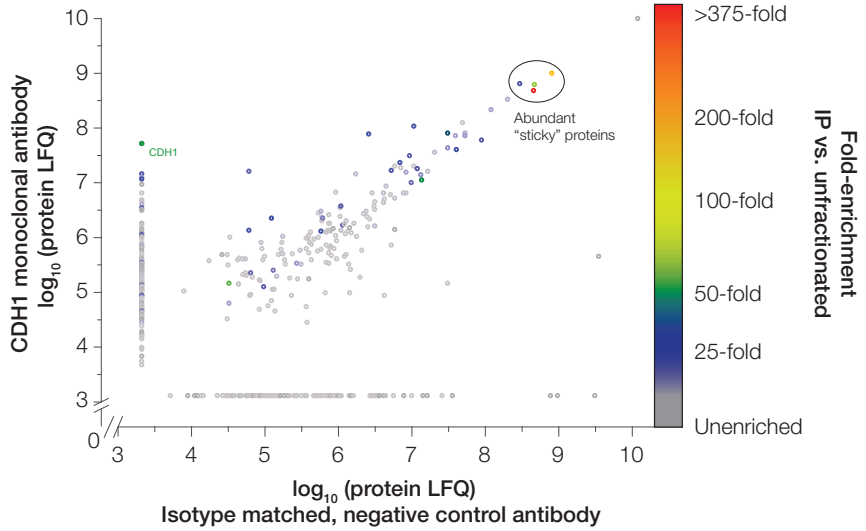
both IPs (Figure 6A, diagonal). This approach could be easily adapted to compare an antibody of interest to multiple negative control antibodies to remove common, nonspecifically bound proteins. To provide additional insight, these scatter plot results were colored based upon the fold-enrichment calculations described below (Figure 6B). Interestingly, some of the common background proteins along the diagonal were significantly enriched with both antibodies. These could be due to specific binding to the magnetic bead resin or antibody isotype, and may depend on the cell type used in the sample preparation. This scatter plot approach typically eliminated more than 90% of the identified proteins as nonspecific binders. Databases of common background proteins observed in affinity purification (AP) experiments are available (e.g., **http://crapome.org**, [12]), but these background proteins may vary by cell type and AP technique. The scatter plot with fold-enrichment shows the intended target(s) and interacting partners (Figure 6B, y-axis) in the context to background proteins and thus provides an estimate of antibody selectivity for IP.

Fold-enrichment to assess selectivity and identify interaction partners

While the IP-MS approach provides verification of intended protein targets, it also identifies all other proteins present in an IP sample. This includes the IP antibody and carrier proteins, like BSA, in addition to interacting proteins and abundant nonspecific proteins. These additional proteins present a challenge to verifying antibody selectivity and performance. To better quantify the performance and selectivity of an antibody and to normalize the results of antibodies against different targets across experiments and cell models, we utilized the concept of protein fold-enrichment. Calculations of fold-enrichment are commonly used to assess and optimize protein purification methods, and this approach can be used to assess an antibody's ability to enrich its native target and interaction partners relative to background proteins from a biological matrix. The formula we used for this analysis is:

$$\text{Fold enrichment} = \frac{\left(\frac{\text{Target protein abundance in IP}}{\text{Total protein abundance in IP}}\right)}{\left(\frac{\text{Target protein abundance in whole lysate}}{\text{Total protein abundance in whole lysate}}\right)}$$

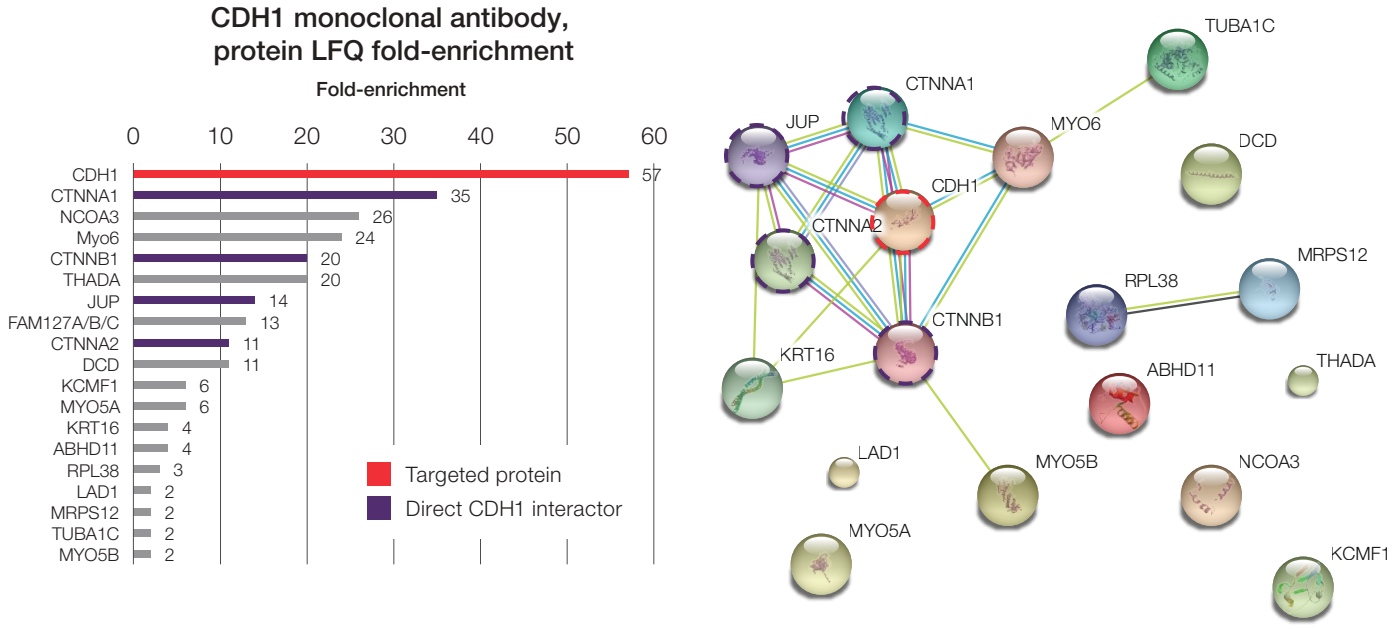
6B



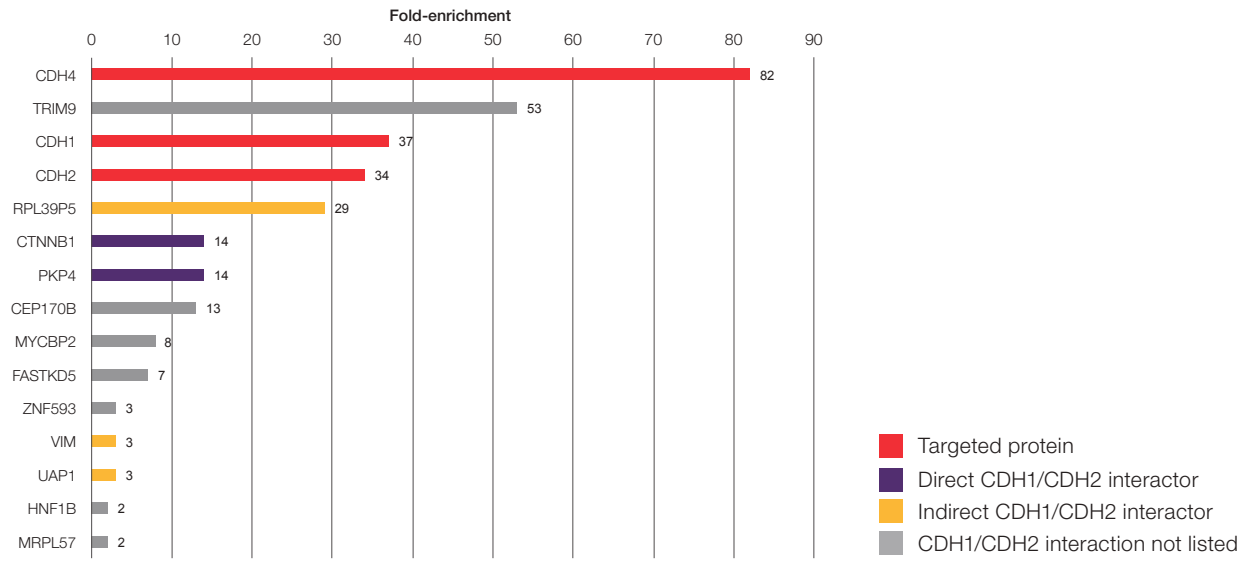
Cell lines were chosen for antibody assessment based upon a deep, MS-based proteome analysis, and more than 10,000 protein groups were identified and quantified in whole cell lysates. Using this data, antibody performance could be assessed quantitatively by calculating the fold-enrichment of all proteins identified in an IP sample. In this manner, the antibody target could be verified and the performance of different antibodies to the same target could be compared. For example, proteins immunoprecipitated from MCF7 cells with an antibody validated by a combination of IP and western blot to CDH1 were compared to proteins immunoprecipitated with an isotype-matched negative control antibody (Figure 6B and 6C). CDH1 was only identified with the IP-validated antibody, and fold-enrichment calculations identified a small subset of proteins that were also specifically enriched with this anti-CDH1 antibody, including alpha1-, alpha2-, and beta1-catenin (CTNNA1, CTNNA2, CTNNB1) and plakoglobin (JUP, also known as gamma-catenin, Figure 6C). These enriched proteins are known CDH1 interaction partners documented in BioGRID (http://thebiogrid.org/) and STRING (http://string-db.org/) protein interaction databases (Figure 6C).

In another example, proteins immunoprecipitated from A549 cells with a pan-specific anti-cadherin antibody were compared to proteins immunoprecipitated with another isotype-matched negative control antibody that did not immunoprecipitate CDH1 (Figure 6D). The pan-specific anti-cadherin antibody enriched R-cadherin (CDH4), E-cadherin (CDH1), and N-cadherin (CDH2) by 30- to 80-fold. Interestingly, the TRIM9 protein was also enriched 50-fold, suggesting potential cross-reactivity with TRIM9.

6C



6D



A previous bioinformatic analysis of TRIM9 and its related proteins highlighted regions of structural similarity to the cadherin superfamily of proteins, potentially explaining the capture of TRIM9 protein with this pan-specific antibody [13]. Further bioinformatic analysis of the specifically immunoprecipitated and enriched proteins identified several known protein interaction partners related to the catenin complex and cell adhesion (Figure 6E and 6F).

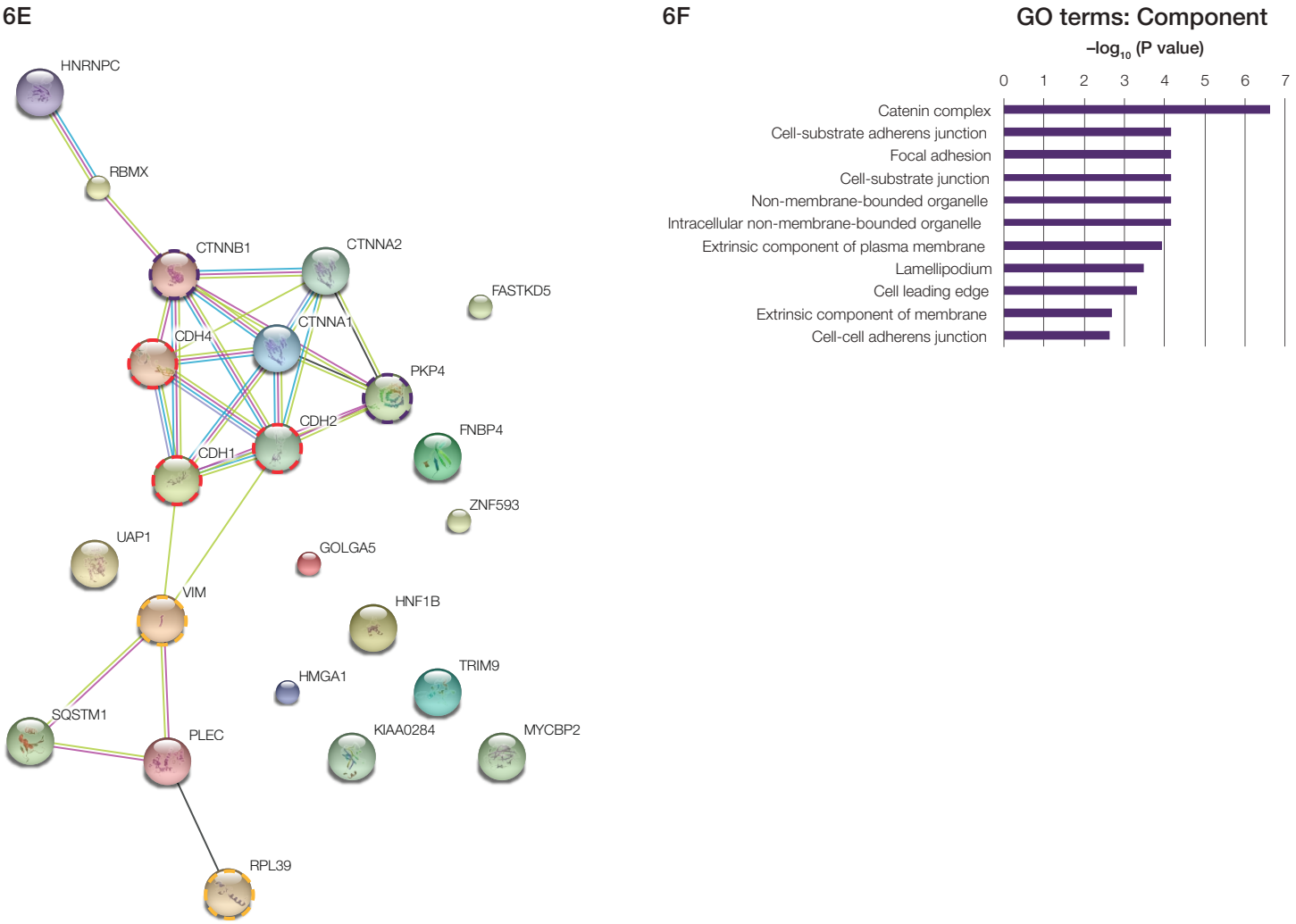
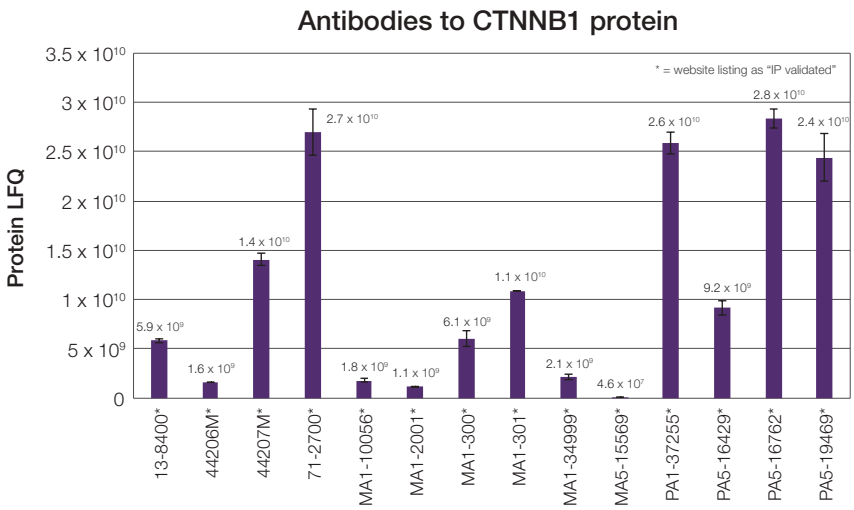


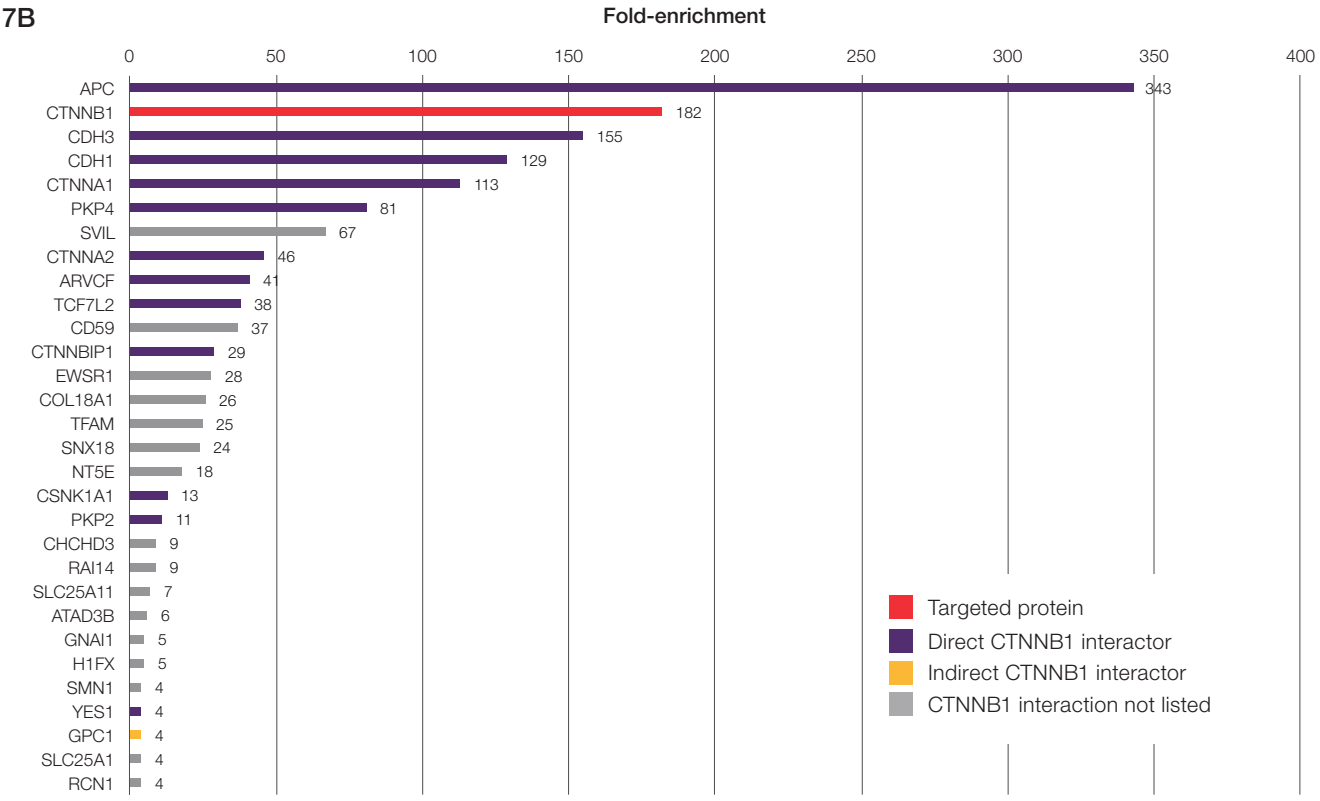
Figure 6. Filtering and visualization of specific proteins captured and quantified by IP-MS. (A) Pictorial diagram of the clusters of proteins quantified after immunoprecipitation with positive and negative control antibodies. **(B)** Scatterplot results showing the proteins captured by CDH1 monoclonal antibody and isotype-matched, negative control antibody and quantified with MS. Fold-enrichment results relative to MCF7 lysates using LFQ quantitation are indicated with colored circles. **(C)** CDH1 was enriched >50-fold using anti-CDH1 monoclonal antibody. Identified proteins were submitted to STRING for interactome analysis. CDH1 is highlighted in dark red, and STRING-assigned interactors are highlighted in blue. **(D)** Enrichment of cadherin targets and additional proteins from A549 cell lysate with pan-cadherin antibody and annotation of known interactors. **(E)** Interaction diagram from STRING database highlighting enriched proteins identified using the pan-cadherin antibody. **(F)** Analysis of Gene Ontology (GO) term enrichment based upon the list of specifically enriched proteins identified using pan-cadherin antibody.

In a final example, 8 antibodies to beta-catenin (CTNNB1) were compared with six immunoprecipitation and western blot–validated (IP-WB–validated) CTNNB1 antibodies using IP-MS (Figure 7A). All six previously validated antibodies successfully captured the target, and an additional 8 antibodies not previously validated for IP also captured the target. As an example, monoclonal antibody 44207M enriched CTNNB1 from HCT116 cells over 150-fold, along with many known protein interaction partners (Figure 7B).

7A



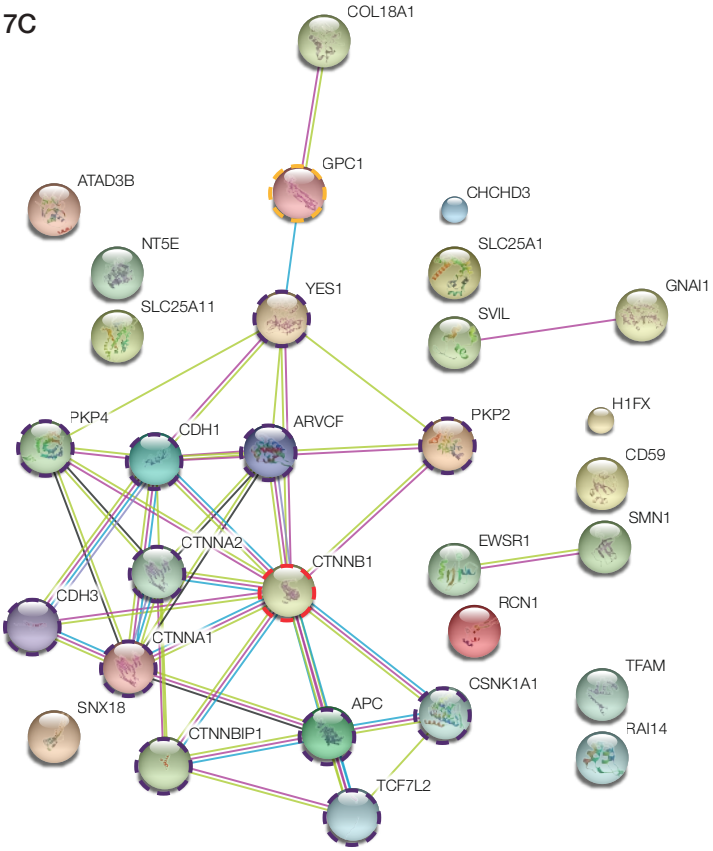
7B



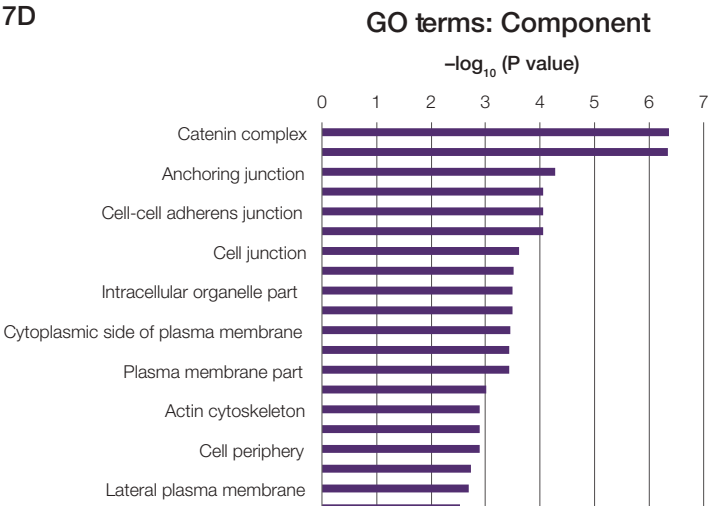
Bioinformatic analysis of the specifically captured and enriched proteins revealed many components of the catenin complex and cell-cell junctions (Figure 7C and 7D). The enrichment of various cadherins cross-validated the previous results that showed capture of beta-catenin with anti-cadherin antibodies (Figure 6C and 6D). Most of the CTNNB1-interacting proteins enriched with 44207M antibody were also seen with several other antibodies, including adenomatous polyposis coli (APC) protein (Figure 7E). APC promotes rapid degradation of CTNNB1, and both proteins play a key role in colorectal cancer [14].

The beta-catenin antibodies were screened with HCT116 cell lysate based upon the MS-based proteome expression profile of beta-catenin, and HCT116 is derived from a colorectal cancer tumor. The higher fold-enrichment of APC than beta-catenin is interesting and illustrates a potential limitation of this approach, as the enrichment of very low-abundance proteins may result in disproportionately high fold-enrichment values. For example, in several cases we observed known interaction partners that were not seen in the MS-based proteome profiling studies, making it impossible to calculate fold-enrichment for these proteins.

7C



7D



7E

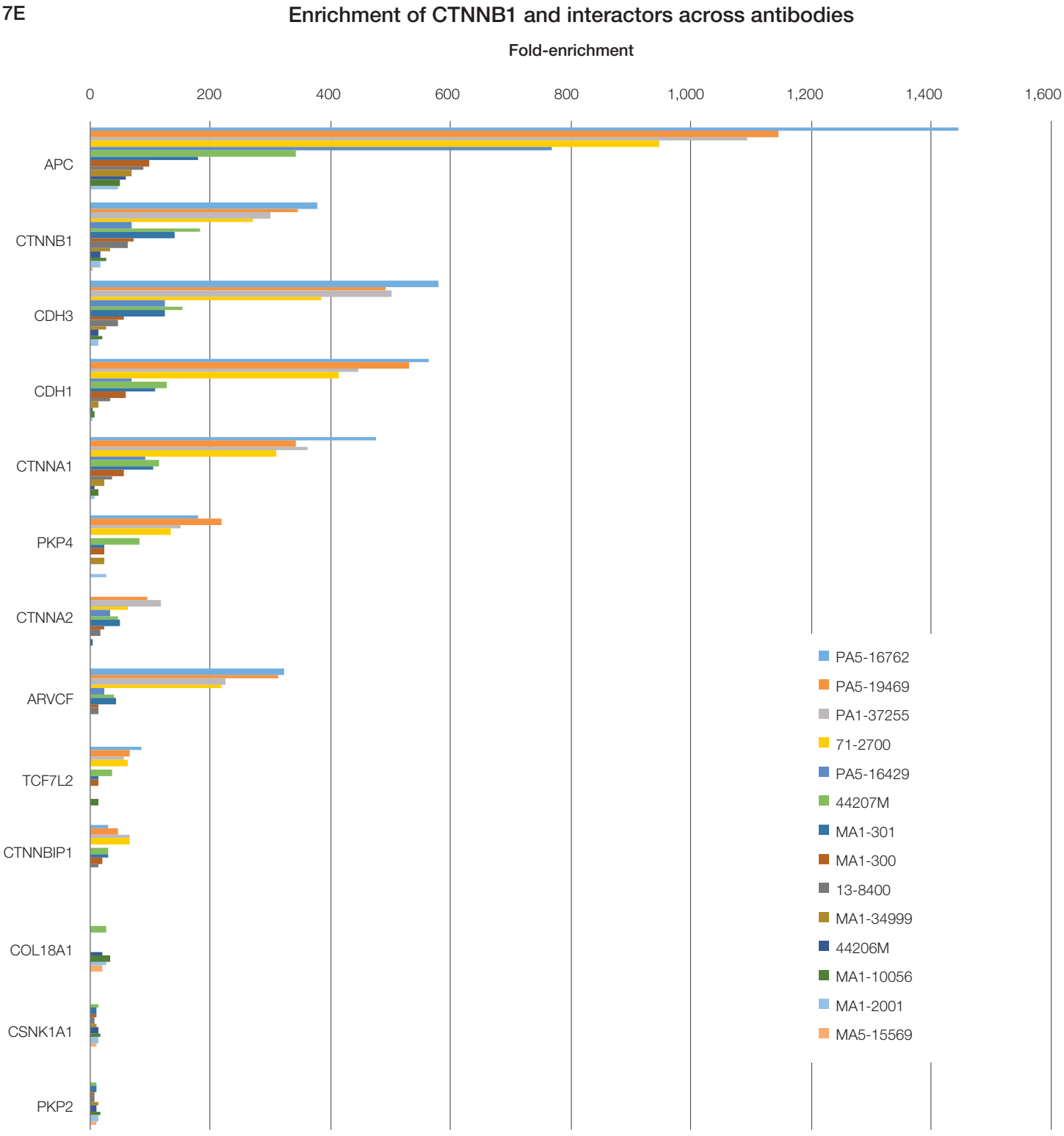


Figure 7. Comparison of antibodies to immunoprecipitate CTNNB1. (A) CTNNB1 protein was immunoprecipitated from HCT116 cell lysate with the indicated antibodies in replicates and quantified by MS using the protein label-free quantitation (LFQ) values. **(B)** Enrichment of CTNNB1 target and additional proteins with antibody 44207M and annotation of known interactors. **(C)** Interaction diagram from STRING database highlighting enriched proteins. Proteins are highlighted according to their status as assigned by the STRING database. **(D)** Analysis of Gene Ontology (GO) term enrichment based upon the list of specifically enriched proteins. **(E)** Enrichment of CTNNB1 target and interacting proteins with multiple positive antibodies.

Conclusions

These examples illustrate some of the benefits of antibody target verification with IP-MS. The IP-MS approach uniquely verifies antibody capture performance by directly identifying peptide sequences from their putative targets. Beyond merely identifying the presence or absence of an antibody's target, the IP-MS workflow enables the characterization of antibody selectivity by identifying the other proteins that are present in a sample following IP. This adds to the power of the overall method by identifying both off-target proteins and observing the presence of potential interacting partners that bind to antibody targets. Compared to other methods that characterize protein-protein interactions, the IP-MS approach is unique for its ability to filter and calculate protein enrichment. Further, IP-MS is capable of identifying target proteins, off-targets, and interactors in their native states (requiring no N- or C-terminal tags) and expression levels in biologically relevant cell lines. In the future, these workflows and analyses may be further extended to calculate fold-enrichment at the peptide level to assess the specificity of antibodies to targeted posttranslational modification sites (e.g., phosphorylation, ubiquitination, and acetylation). Antibody selectivity measurements could also be used to help map antibody epitopes, protein conformations, or proteoforms that may have distinct protein interactions. These measurements could also help expedite the selection of complementary antibodies that may be combined in sandwich-type assays.

Methods

Unless otherwise stated, products identified with catalog numbers (Cat. No.) are from Thermo Fisher Scientific.

Cell culture

All cell lines were purchased from ATCC and grown in the condition noted in Table 1. All media and cell growth products were purchased from Thermo Fisher Scientific, including trypsin (Cat. No. 25200-056) and HBSS (Cat. No. 14175-079), and all media was supplemented with 10% FBS (Cat. No. 16000-036), 1X penicillin-streptomycin (Cat. No. 15140-163), and insulin Tocris Bioscience, (Cat. No. 12585014), if needed. Cells were grown to ~80% confluency at passage 12–18 before lysis with Pierce™ IP Lysis Buffer (Cat. No. 87788) and 1:100 Halt™ Protease and Phosphatase Inhibitor Cocktail (Cat. No. 78445). If cells underwent stimulation, cells were starved in 0.1% charcoal-stripped FBS (Cat. No. SH30068.01) for 24 hours before stimulation with 100 ng/mL of hIGF (Cell Signaling Technology, Cat. No. 8917SF) or 100 nM insulin (Cat. No. 12585014) for 15 min and then lysed immediately. Protein concentration was determined with the Pierce™ BCA Protein Assay Kit (Cat. No. 23225) using a Multiskan™ GO instrument for measurement; aliquots were stored at –80°C until use.

Table 1. Cell models used and growth conditions.

Cell model	Tissue type	Medium	Insulin	Stimulation	Cat. No. (media)
HCT116	Colon	McCoy's 5A	NA	± IGF	16600-082
A549	Lung	Hamm's F-12K	NA	± IGF	21127-022
MCF7	Breast	DMEM	10 µg/mL	NA	11995-040
HepG2	Liver	MEM	NA	± insulin	11095-072
LNCaP	Prostate	RPMI-1640	NA	± IGF	11875-085
NIH3T3	Fibroblast	DMEM	NA	NA	11995-040
BT-549	Breast	RPMI-1640	0.023I U/mL	NA	11875-085
SK MEL 5	Skin	DMEM	NA	NA	11995-040
Hs 578T	Breast	DMEM	10 µg/mL	NA	11995-040
SR	Lymphoblast	RPMI-1640	NA	NA	11875-085
HeLa	Cervical	DMEM	NA	NA	11995-040
HEK293	Kidney	DMEM	NA	NA	11995-040

Lysate sample prep for unfractionated and fractionated proteome analysis

200–800 µg of lysate was further processed for analysis by mass spectrometry using the Pierce™ Mass Spec Sample Prep Kit for Cultured Cells (Cat. No. 84840) as stated in the instruction booklet with proper scale-up of reagents. After the final drying step, samples were reconstituted in 0.1% trifluoroacetic acid (TFA) and cleaned of incompatible salts, detergents, and other reagents using the Pierce™ High pH Reverse-Phase Peptide Fractionation Kit (Cat. No. 84868) with a custom protocol involving column conditioning, 3 washes with 0.1% TFA, and 3 elution steps with 50% acetonitrile and 0.1% TFA. Samples were dried in a vacuum concentrator and reconstituted in 200 µL of 0.1% TFA. 5 µL of sample in 45 µL of water (1:5 dilution) was aliquoted and the Pierce™ Quantitative Fluorometric Peptide Assay (Cat. No. 23290) was performed to determine peptide concentration as described in the instruction booklet.

For fractionation, 100 µg of digested peptide sample was fractionated with the Pierce™ High pH Reverse-Phase Peptide Fractionation Kit (Cat. No. 84868), following the instruction booklet with the exception of a custom fractionation profile, as noted in Table 2.

Fractionated samples were dried in a vacuum concentrator and reconstituted in 20 µL of 2% acetonitrile and 0.1% formic acid. Peptide concentration was measured with the Pierce Quantitative Fluorometric Peptide Assay (Cat. No. 23290) using 8 µL of sample in 16 µL of water (1:3 dilution). Unfractionated and fractionated samples were transferred into an autosampler vial for LC-MS analysis.

LC-MS analysis of unfractionated and fractionated lysate samples

2 µg of unfractionated and fractionated samples were analyzed by nanoLC-MS/MS on a Thermo Scientific™ Dionex™ UltiMate™ 3000 RSLCnano System and Thermo Scientific™ Q Exactive™ HF Hybrid Quadrupole-Orbitrap Mass Spectrometer using a Thermo Scientific™ EASY-Spray™ column (50 cm x 75 µm ID, PepMap C18, 2 µm particles, 100 Å pore size, Cat. No. ES803). The column temperature was maintained at 60°C using an Easy Spray Ion Source (Cat. No. ES081) interfaced online with the mass spectrometer. Mobile phase A (0.1% Formic acid in water, LC-MS grade) and Mobile phase B (0.1% Formic acid in Acetonitrile (ACN), LC-MS grade) were used to buffer the pH in the two running buffers. The total gradient was 210 min followed by a 30 min washout and reequilibration. In detail, the flow rate started at 300 nL/min and 2% ACN with a linear increase to 20% ACN over 170 min followed by a 40 min linear increase to 32% ACN. The washout followed with a flow rate set to 400 nL/min at 95% ACN for 4 min followed by a 24 min reequilibration at 2% ACN.

The Q Exactive HF instrument (located in Bremen, Germany) was freshly cleaned and calibrated using Tune (version 2.5 build 2042) instrument control software. Spray voltage was set to 1.9 kV, S-lens RF level at 60, and heated capillary at 275°C. Full scan resolutions were set to 120,000 at m/z 200. Full scan target was 1×10^6 with a maximum IT fill time of 60 ms. Mass range was set to 400–1,600. Target value for fragment scans was set at 1×10^5, and intensity threshold was kept at 5×10^4. Isolation width was set at 2.0 Th. The normalized collision energy was set at 27. Peptide match was set to preferred, and isotope exclusion was utilized. All data was acquired in profile mode using positive polarity.

Table 2. Custom fractionation profile.

Fraction	Acetonitrile %	Acetonitrile (100%) µL	Triethylamine (0.1%) µL
1	5.0%	50	950
2	6.25%	62.5	937.5
3	7.5%	75	925
4	8.75%	87.5	912.5
5	10.0%	100	900
6	15.0%	150	850
7	20.0%	200	800
8	50.0%	500	500

IP-MS sample preparation

Antibodies against TP53, CDH1, CDH2, and CTNNB1 were purchased from Thermo Fisher Scientific (refer to Figures 5, 6B, 6C, 6D, 7A). The Pierce™ MS-Compatible Magnetic IP Kit (Protein A/G) (Cat. No. 90409) was used to screen and verify antibodies as described in the instruction manual. 500 µg lysate and 3 µg or recommended dilutions of antibody were used for all experiments. IP eluates were dried in a vacuum concentrator and samples were spiked with green fluorescent protein (GFP) as a digestion indicator and then processed by an in-solution digestion method as recommended in the instruction manual for Cat. No. 90409. Dried digested samples were resuspended in 13 µL of 4% acetonitrile and 0.2% formic acid and transferred into autosampler vials before LC-MS Analysis.

LC-MS analysis of IP-MS samples

The IP-MS samples were analyzed by nanoLC-MS/MS using a Thermo Scientific™ Dionex™ UltiMate™ 3000 RSLCnano System coupled to a Thermo™ Scientific™ Q Exactive™ HF Hybrid Quadrupole-Orbitrap Mass Spectrometer or Thermo™ Scientific™ Q Exactive™ Plus Orbitrap Mass Spectrometer. 7 µL of tryptic digest samples were desalted on-line using the Nano Trap Column (100 µm i.d. x 2 cm, packed with Acclaim PepMap100 C18, 5 µm, 100Å, Cat. No. 164564), and separated using an EASY-Spray PepMap C18 column (15 cm x 75 µm ID, 3 µm particles, 100 Å pore size, Cat. No. ES800), with a total gradient time of 62 min. In detail, the flow rate started at 300 nL/min and 3% ACN with a linear increase to 25% ACN over 55 min followed by a 7 min linear increase to 40% ACN. The column was washed with a flow rate set to 600 nL/min at 95% ACN for 3 min followed by a 5 min re-equilibration at 3% ACN.

The Q Exactive HF and Q Exactive Plus instruments (located in Bremen, Germany) were freshly cleaned and calibrated. Spray voltage was set to 1.9 kV, S-lens RF level at 60, and heated capillary at 275 °C. Full scan resolutions were set to 70,000 at m/z 200 (Q Exactive Plus) and 60,000 at m/z 200 (Q Exactive HF). The full scan automatic gain control (AGC) target was set to 3×10^6 with a maximum IT fill time of 50 ms for the Q Exactive Plus and 1×10^6 with a maximum IT fill time of 60 ms for the Q Exactive HF. The mass ranges for both instruments were set to 400–1,600 m/z. The target AGC value for fragment scans were set at 1×10^5, and the intensity threshold were kept at 1×10^4 (Q Exactive HF) and 3.3×10^3 (Q Exactive Plus). Instrument isolation widths were set at 1.2 Th for Q Exactive HF and 2.0 Th for Q Exactive Plus. The normalized collision energy was set at 27 for both instruments. Peptide match was set to preferred, and isotope exclusion was utilized. All data was acquired in profile mode using positive polarity.

RNA and protein profile data analyses

RNA expression Z-scores were retrieved from CellMiner, hierarchically clustered with Cluster 3.0, and displayed with Java TreeView 1.16r4. The Venn diagram of protein identification results was generated by an online tool at **http://bioinformatics.psb.ugent.be/webtools/Venn/**.

MS data analysis and visualization

MS data obtained from unfractionated lysate, fractionated lysates, and deep proteome analysis and initial IP sample screens were analyzed using Proteome Discoverer 1.4 (release 1.14). A custom human proteome database (UniProt, assembled Feb 2016) was utilized for deep proteome analysis, while a combined database of human proteome and mouse/rat/rabbit IgGs, (UniProt, assembled July 2016) was used for database search IP screens. The IP database also included the recombinant protein A/G and GFP protein sequences. Trypsin was selected as the enzyme used for digestion. During automated searching, concatenated target/decoy databases were generated to validate peptide-spectral matches (PSMs) and filter identifications to a 1% false discovery rate (FDR). MS spectra were searched using 20 ppm precursor mass tolerance and 0.03 Da fragment tolerance. The data was searched with a static modification of carbamidomethylation of cysteine residues, and dynamic modifications including the acetylation of protein N-termini, oxidation of methionine residues, and phosphorylation of serine, threonine, and tyrosine residues.

Protein groups of unfractionated, fractionated, and IP-MS samples data were exported and custom software was used to extract the unique peptide sequences, number of PSMs, and top 3 peptide peak areas for each identified protein. Top 3 peptide peak area was used to determine relative abundance of specific proteins across multiple cell lines.

Fractionated proteome data was curated to determine the total number of proteins identified from each cell line. Specific protein targets were compared between cell lines using custom software to extract the number of peptide spectral matches (PSMs), unique peptide sequences, and both summed and averaged peptide peak areas or peptide intensities. Cell lines were selected for IP using these metrics to determine cell lines which expressed protein targets at a moderate abundance within 2 standard deviations of the mean protein intensity.

After testing multiple antibodies for the same target with replicates, IP data was first searched and screened using PD 1.4 using a combined database of the human proteome and mouse/rat/rabbit IgGs in addition to recombinant protein A/G and GFP sequences. The number of unique peptides, top 3 highest peptide intensities, PSMs, and total background proteins observed in each IP were assessed in order to verify the performance of each antibody to isolate its putative target.

Following PD 1.4 analysis, samples with detectable target were searched using MaxQuant 1.5.3.51 to obtain relative quantification of peptides and proteins and compare these protein abundances from replicate IP samples to unfractionated and fractionated proteome lysate samples. Curated contaminant proteins were also added to the database search. A target-decoy database was generated during automated searching and resulting peptide and protein identifications were filtered to a 1% FDR. Data was searched using group-specific parameters with a multiplicity of one, trypsin as the enzyme used for digestion, a maximum of 2 missed cleavages, fixed modification of carbamidomethylation of cysteine residues, and variable modifications including the acetylation of protein N-termini and oxidation of methionine residues. Label-free quantification (LFQ) was performed using a minimum LFQ ratio count of 2 and fast LFQ. Spectra were searched using a 20 ppm first search peptide tolerance and a 4.5 ppm main search peptide tolerance. MS/MS spectra were analyzed with a 20 ppm fragment match tolerance. Protein quantification was defined using a minimum threshold of 2 ratios, using unique and razor peptides for quantification. Large LFQ values were stabilized and required MS/MS for LFQ comparisons.

Once MaxQuant output was obtained, the data was manually analyzed to compare the intensities and LFQ values obtained across the unfractionated, fractionated, and replicate IP samples. Protein LFQ values were used to generate scatterplots to characterize the specificity of antibodies used in IP. For these scatterplots, LFQ values were plotted to compare the relative abundances of proteins identified in a “test” IP (plotted on the y-axis) to those proteins identified in a negative control IP for an unrelated target (plotted on the x-axis). The negative control antibody was selected for comparison either because the antibody recognized a different target or did not identify the target that was pulled down by the test IP. Plotting the relative abundances of proteins from each test and negative IP led to three distinct regions of the scatterplot, where proteins identified in both IPs were considered nonspecific “background” proteins along the diagonal of the scatterplot. Those proteins uniquely identified in the test IP were observed as aligned along the y-axis, while those proteins identified only in the negative control IP were aligned along the x-axis of the plot. Proteins observed uniquely in the test IP were color-coded according to their fold-enrichment versus deep proteome samples. Fold-enrichment and scatterplot calculations were performed by a custom web application to streamline the data analysis and generation of graphs for IP verification. LFQ values for replicate IPs were utilized to further filter the data for those proteins which were observed reproducibly across replicates. A 25% CV cutoff was used to filter proteins which were not reproducibly identified or quantified across replicate IP samples.

To calculate fold enrichment, for each MS run searched, the LFQ abundance of each protein was extracted and divided by the summed abundance of all proteins identified to obtain a “fraction” of that protein’s relative abundance versus every other protein identified in the sample. The relative fraction of the protein’s abundance in an IP sample was then compared to the fraction of the protein in the deep proteome samples to observe whether this fraction increased, decreased, or stayed the same relative to the other proteins that were identified in each IP. In this way, a fold-enrichment was calculated for every protein in the IP samples, and this calculation was used to characterize the enrichment of putative antibody targets and known target-protein interactors. These fold-enrichment calculations were performed using both protein LFQ.

invitrogen

Proteins which were observed uniquely in test IPs and exhibited a >1-fold enrichment compared to deep proteome analysis were submitted to the STRING database (**http://string-db.org**) to probe known target-protein interactions. Protein interactions were selected against the *Homo sapiens* proteome. Proteins were plotted according to their known interactors using text mining, experimental verification, database annotation, co-expression, gene fusion, and co-occurrence data. Data was plotted with nodes representing proteins uniquely identified in the test IP and edges representing evidence of protein-protein interactions. Protein fold-enrichment bar charts were color coded according to whether the identified protein was the putative antibody target or listed as a direct interactor with the target via the STRING database. Proteins were also color coded to represent whether they were indirect interactors (i.e., listed as interacting with annotated target interactors) or were not listed as interacting via the STRING database. Network statistics from the STRING database were downloaded with enriched GO terms for cellular component, biological processes, molecular function, KEGG pathways, Pfam annotations, and InterPro classifications.

References

1. Lipman, N.S., et al., Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources. *ILAR Journal*, 2005. 46(3):258–268.
2. Madhusoodanan, J. Validating Antibodies: An Urgent Need. 2014; Available from: http://www.the-scientist.com/?articles.view/articleNo/41539/title/Validating-Antibodies--An-Urgent-Need/.
3. Bordeaux, J., et al., Antibody validation. *Biotechniques*, 2010. 48(3):197–209.
4. Bjorling, E. and M. Uhlen, Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol Cell Proteomics*, 2008. 7(10):2028–2037.
5. Pauly, D. and K. Hanack, How to avoid pitfalls in antibody use. *F1000Res*, 2015. 4.
6. Bostrom, T., et al., Investigating the applicability of antibodies generated within the human protein atlas as capture agents in immunoenrichment coupled to mass spectrometry. *J Proteome Res*, 2014. 13(10):4424–4435.
7. Marcon, E., et al., Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. (1548–7105 (Electronic)).
8. Cox, J. and M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotech*, 2008. 26(12):1367–1372.
9. Keilhauer, E.C., M.Y. Hein, and M. Mann, Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification Mass Spectrometry (AP-MS). *Mol Cell Proteomics*, 2015. 14(1):120–135.
10. Teo, G., et al., SAINTexpress: Improvements and additional features in Significance Analysis of INTeractome software. *Journal of Proteomics*, 2014. 100:37–43.
11. Sowa, M.E., et al., Defining the human deubiquitinating enzyme interaction landscape. *Cell*, 2009. 138(2):389–403.
12. Mellacheruvu, D., et al., The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Meth*, 2013. 10(8):730–736.
13. Short, K.M. and T.C. Cox, Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *J Biol Chem*, 2006. 281(13):8970–8980.
14. Ilyas M., I.P. Tomlinson, The interactions of APC, E-cadherin and beta-catenin in tumour development and progression. *J Pathol*, 1997. Jun;182(2):128–137.

Find out more at **thermofisher.com/antibodyvalidation**

ThermoFisher SCIENTIFIC