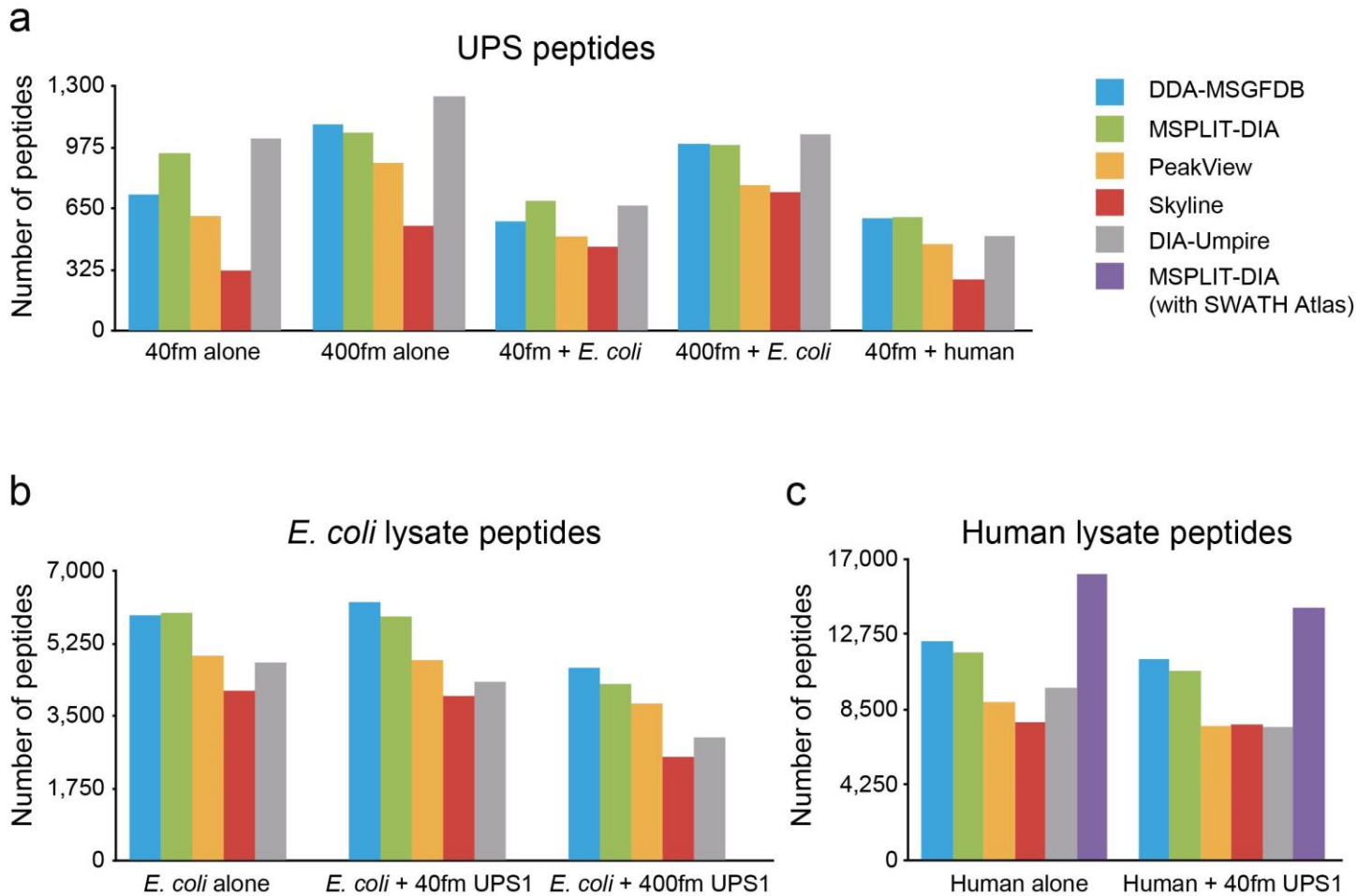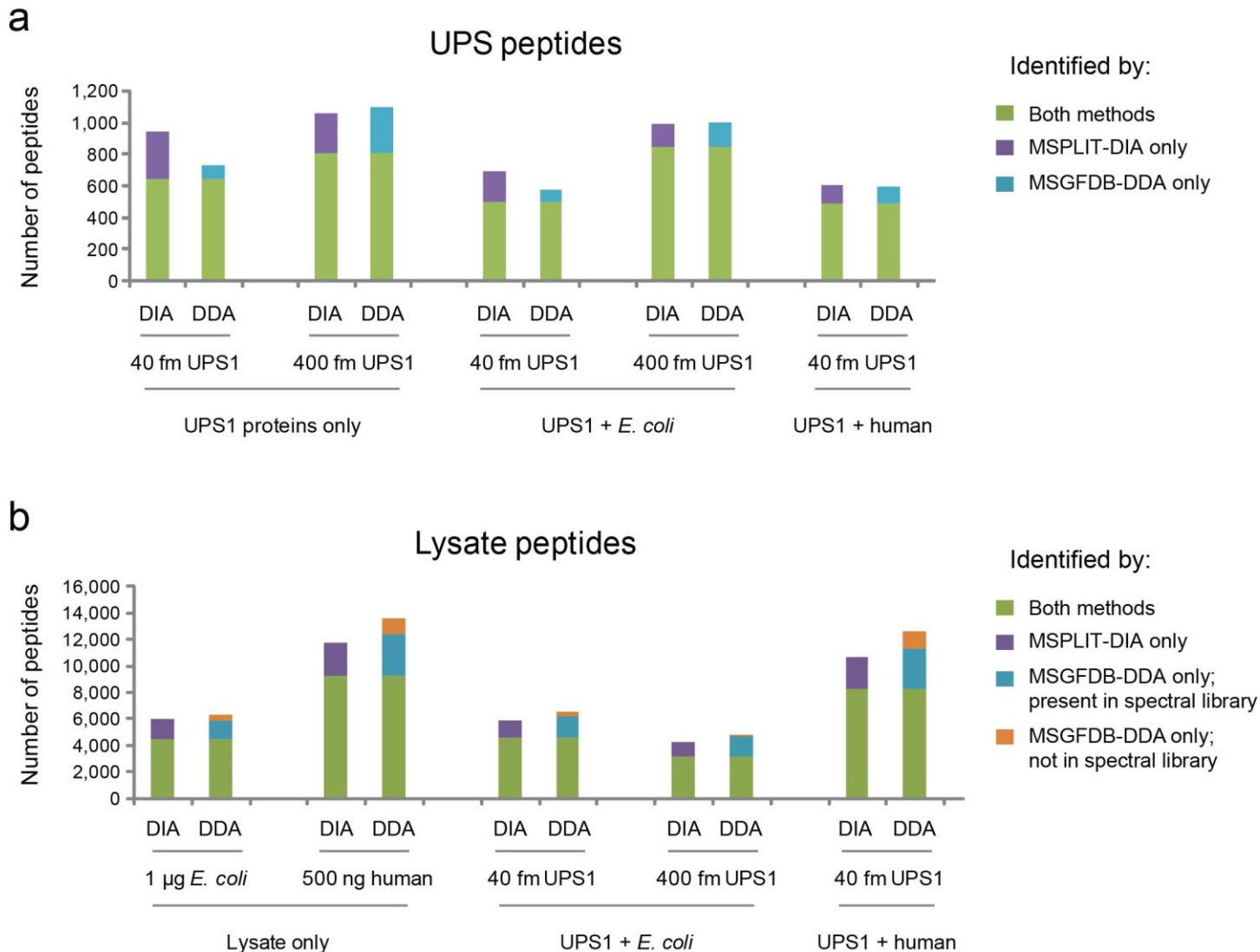**Supplementary Figure 1**

Example of a DIA multiplexed spectrum with three peptides identified by MSPLIT-DIA.

Supplementary Figure 2

Comparison of data analysis approaches on a standard protein mixture in isolation or spiked into complex backgrounds.

A standard 48-protein mixture (UPS1) was analyzed by itself or spiked into either an *E. coli* or human lysate background. Each sample was analyzed by both DDA and DIA. DDA data were analyzed using MSGFDB (Kim et al., *Mol Cell Proteomics*, 2010; blue bars) and DIA data were analyzed using MSPLIT-DIA (green bars), PeakView (orange bars), Skyline (red bars) and DIA-Umpire in untargeted identification mode (grey bars); unless stated otherwise, DIA analyses utilized in-house spectral libraries derived from all corresponding DDA runs (see **Supplementary Note**). PeakView (orange) and Skyline (red) employ the targeted extraction strategy most utilized for SWATH data analysis, while DIA-Umpire (grey) performs direct identification (using database searching) after the generation of pseudo-MS/MS spectra. The performance of MSPLIT-DIA for peptide identification in human lysate samples was greatly improved by employing the much more comprehensive publicly available human spectral library (SWATH-Atlas, Rosenberger et al., *Scientific Data*, 2014; purple bars). The number of identified non-redundant peptides derived from UPS1 (a) and *E. coli* (b) or human (c) lysate proteins are shown. All results are reported at 1% peptide-level FDR. In comparison with peptides detected by targeted extraction approaches, MSPLIT-DIA identified up to 63% more UPS1 peptides than could be detected by PeakView in a simple DIA run (40 fmol UPS1) and 41% more human peptides than PeakView in a much more complex DIA run when employing the same in-house library (UPS1 spiked into a human lysate background), with some variability observed due to signal suppression in complex samples (see **Supplementary Figure 4**). The increased sensitivity of MSPLIT-DIA was also detected in this analysis when comparing to Skyline; we note that analysis with OpenSWATH analysis was also attempted but due to software issues we could not resolve, the analysis never successfully completed and was thus not included.

a

UPS peptides

Identified by:
- Both methods (green)
- MSPLIT-DIA only (purple)
- MSGFDB-DDA only (blue)

b

Lysate peptides

Identified by:
- Both methods (green)
- MSPLIT-DIA only (purple)
- MSGFDB-DDA only; present in spectral library (blue)
- MSGFDB-DDA only; not in spectral library (orange)

Supplementary Figure 3

**Peptide identification overlaps between MSGFDB-DDA and MSPLIT-DIA.**

Peptides identified in i) UPS1 only, ii) UPS1 + *E. coli* or iii) UPS1 + human samples by MSPLIT-DIA and MSGFDB-DDA were compared. While the number of peptides detected by MSPLIT-DIA using sample-matched libraries was comparable to MSGFDB-DDA for medium-to-high abundance peptides, MSPLIT-DIA was up to 31% more sensitive for the identification of low abundance peptides, with the largest gains over MSGFDB-DDA occurring when UPS1 peptides are present at just 40 fmol UPS1, either by themselves or when spiked in an *E. coli* lysate background. The number of identified peptides from UPS1 proteins (a) and peptides from lysate proteins (b) are listed. Color coding in each bar indicates the peptides that were identified by both methods (green), identified only by MSGFDB-DDA (blue), or only by MSPLIT-DIA (purple). The characteristics of peptides identified by MSGFDB-DDA but not MSPLIT-DIA were further investigated: the quality of library spectra and dynamic range suppression of low abundance signals were found to affect peptide detectability in DIA data (see **Supplementary Figure 4** and **Supplementary Note Figure SN4**). Note that a small fraction of peptides (orange) were identified only by MSGFDB-DDA because they were not included in the spectral library used for MSPLIT-DIA since we enforced a global 1% peptide-level FDR for the identified peptides across all DDA runs in order to ensure the quality of the spectral library (see **Supplementary Note**). To better contrast the sensitivity of the peptide identification approaches in a comparable search space, we excluded peptides with no spectra in the in-house library when comparing results between MSGFDB-DDA and MSPLIT-DIA in **Supplementary Figure 2**.
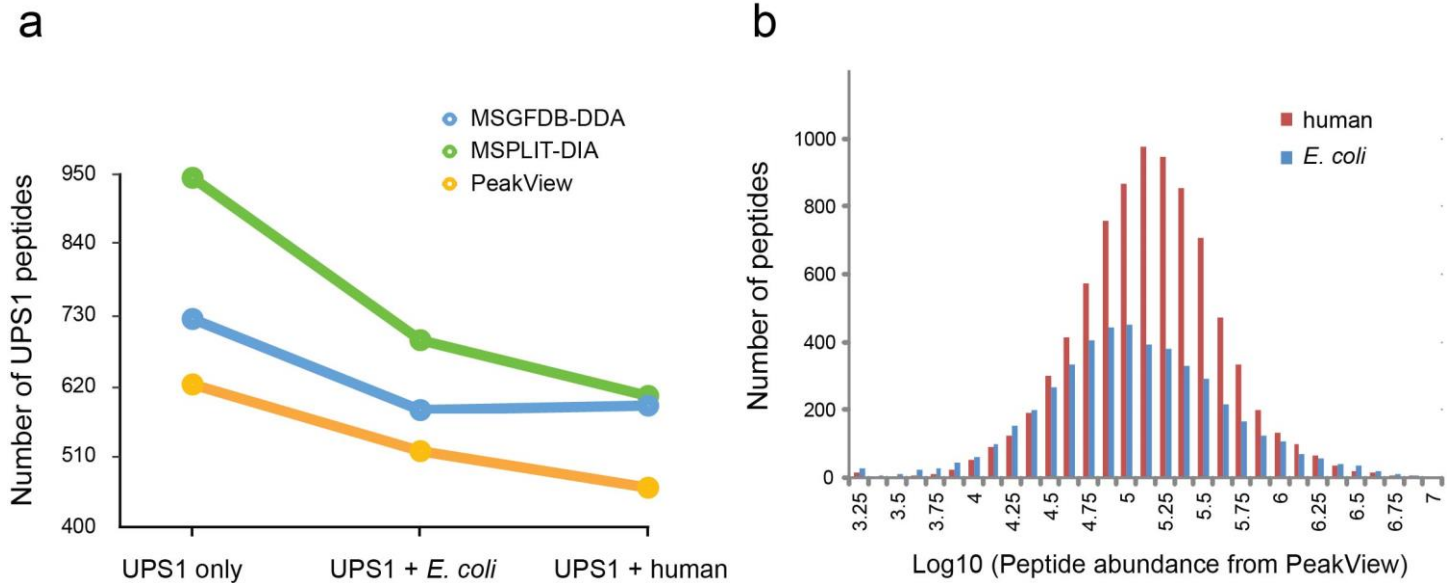
Supplementary Figure 4

Signal suppression in complex samples.

(a) Number of UPS1 peptides detected by MSPLIT-DIA, MSGFDB-DDA and PeakView at 1% FDR across samples of different complexity. In general MSPLIT-DIA identifies about the same or more UPS1 peptides than either MSGFDB-DDA or PeakView. However MSPLIT-DIA's higher sensitivity also appears to be more affected by dynamic range signal suppression for lower abundance signals than MSGFDB-DDA (see **Supplementary Figure 2**). These dynamic range effects seem to scale with sample complexity when comparing MSPLIT-DIA to MSGFDB-DDA (30%→19%→1% peptide gains in UPS1-only → UPS1+*E. coli* → UPS1+human) but somewhat stabilize when comparing with PeakView (52%→33%→30% peptide gains in UPS1-only → UPS1+*E. coli*→ UPS1+human). A possible explanation for this effect is that the number of high abundance peptides (precursor intensity >10e6) in both the *E. coli* and human lysates is surprisingly similar: 346 peptides in *E. coli* and 386 peptides in human (b). As such, it is likely that a substantial fraction of the observed dynamic range signal suppression is due to the co-occurrence of a relatively small number of these high abundance peptides in the same SWATH isolation windows. In contrast, MSGFDB-DDA appears to be less affected by these factors and thus reaches a level of performance comparable to MSPLIT-DIA in human lysate samples. Nevertheless, even though signal suppression seems to have a comparable effect on MSPLIT-DIA and PeakView, we note that MSPLIT-DIA continues to be significantly more sensitive than PeakView in higher complexity samples.
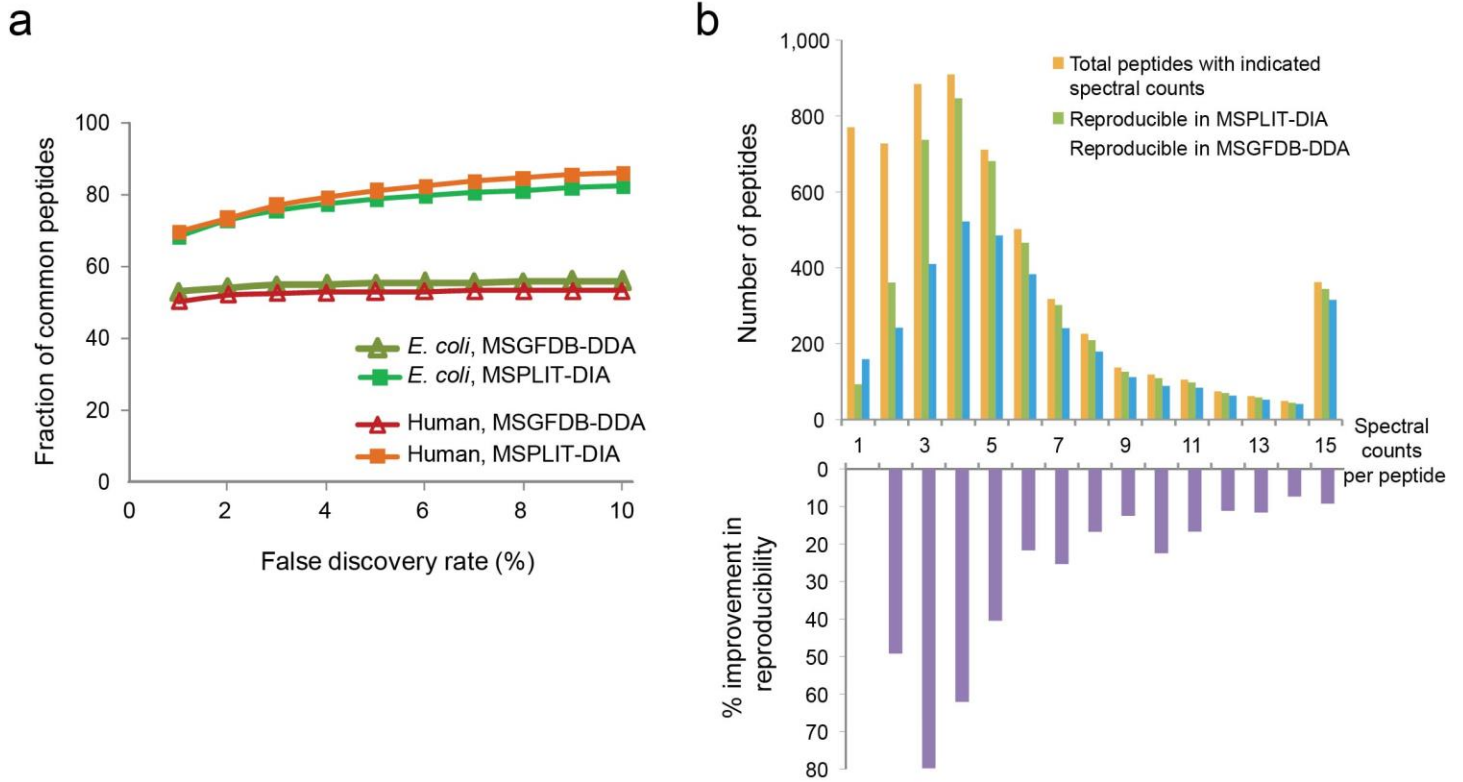
Supplementary Figure 5

Reproducibility of peptide identification.

(a) Reproducibility in identification was assessed by the fraction of peptides that could be consistently identified across four runs at different FDR thresholds. (b) Peptide reproducibility was further evaluated using replicate runs of the *E. coli* lysate samples. We defined the overall set of identified peptides to include peptides that were identified in at least one DDA run and at least one DIA run. A peptide is reproducibly identified if it is identified at 1% FDR in all four DDA runs (DDA reproducibility) or all four DIA runs (DIA reproducibility). Peptide abundance was estimated by the average spectral count for a particular peptide across all DIA runs. The total number of peptides identified at each abundance level (yellow bars) and the number of peptides reproducibly identified in DIA (green bars) and DDA (blue bars) are shown on the top panels. The bottom panel shows the gains in reproducibility as the percentage increase in peptides identifications of MSPLIT-DIA in comparison to MSGFDB-DDA. See **Fig. 1d** for an analysis of human lysate samples.

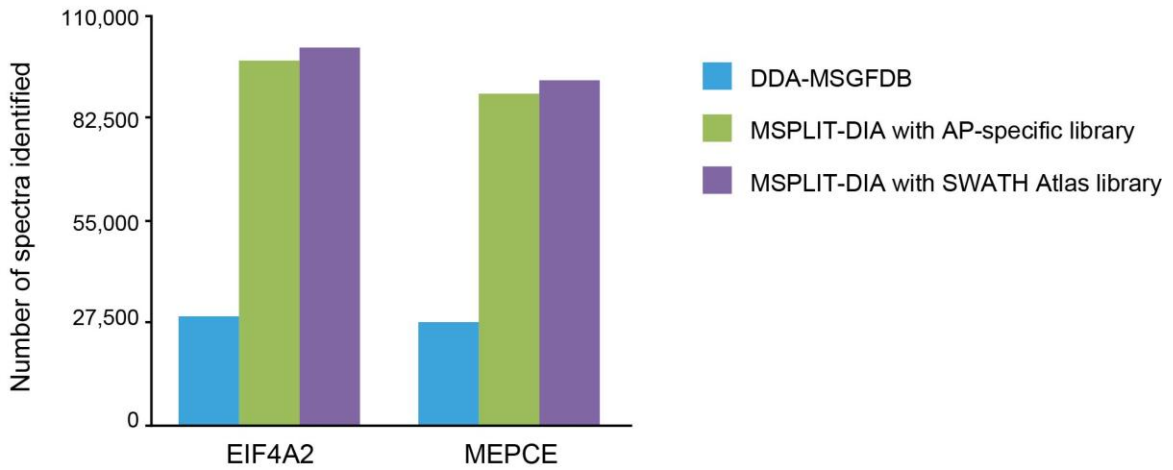**a** Peptide identifications

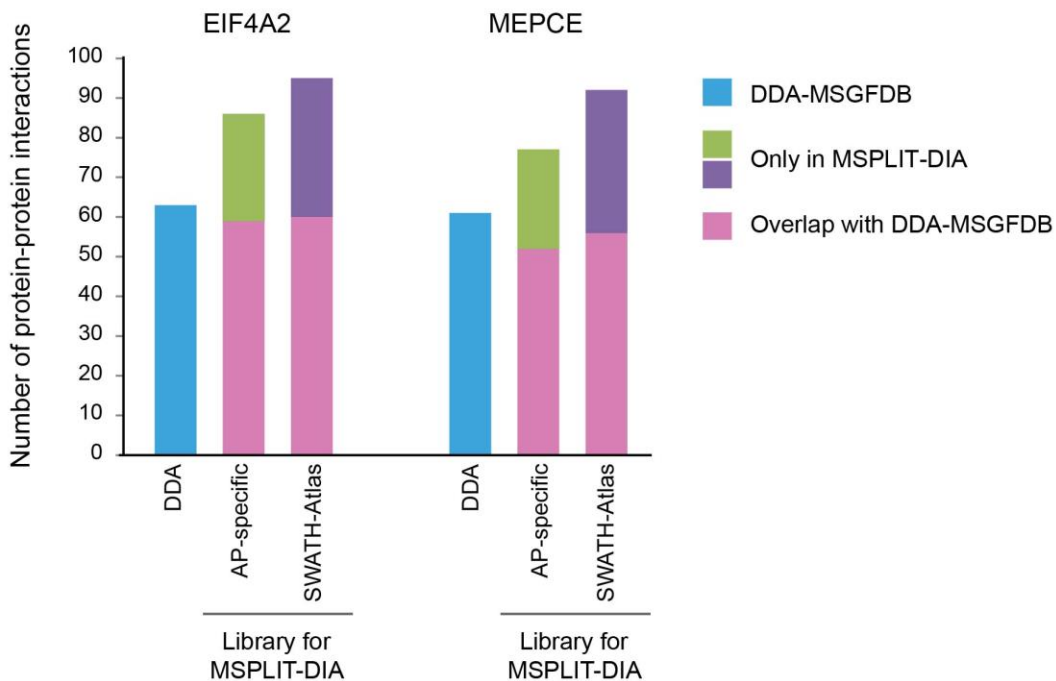**b** Protein identifications
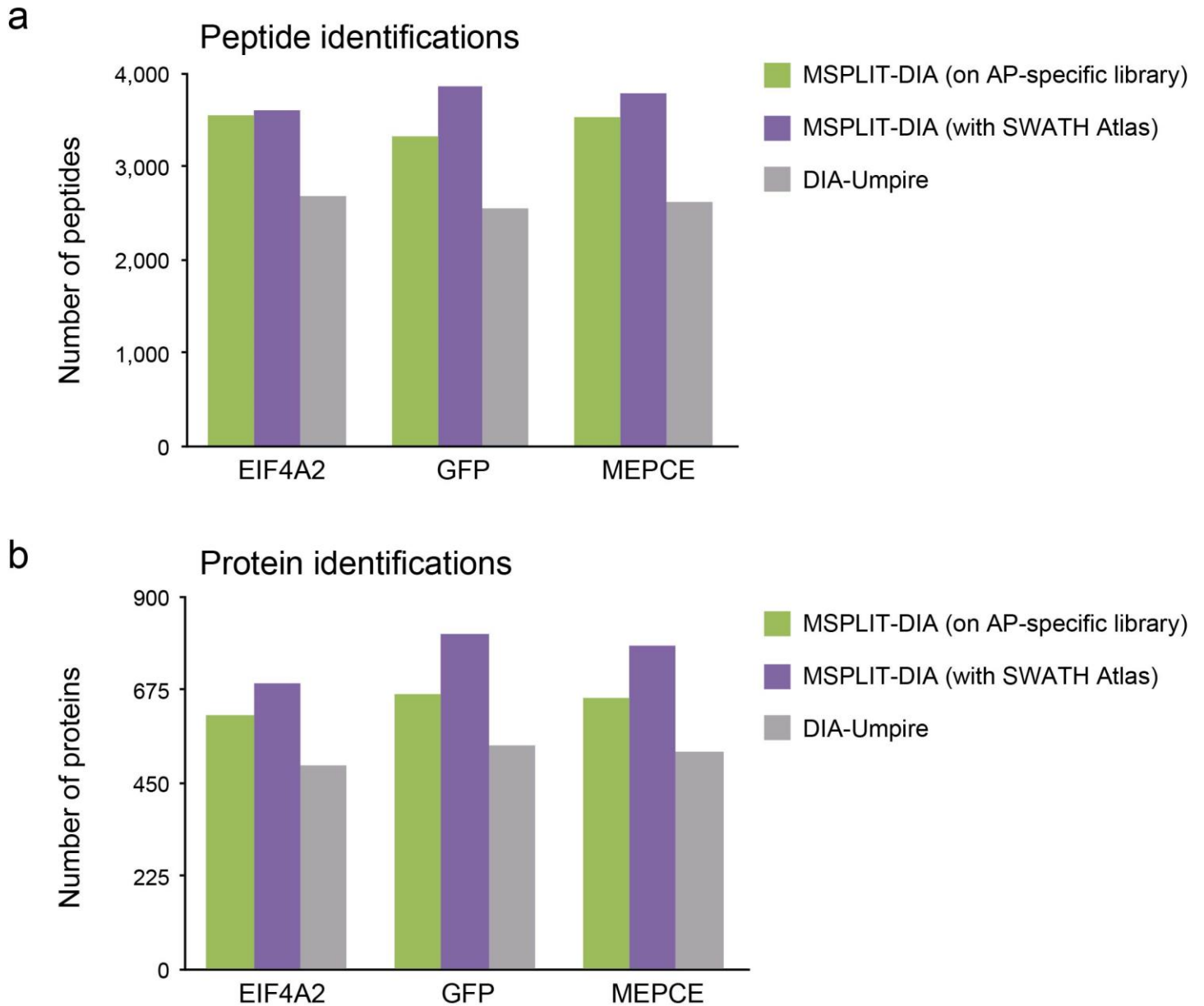
Supplementary Figure 6

Peptide and protein identification in affinity purified samples.

Two different spectral libraries were searched to test the robustness of MSPLIT-DIA to spectral libraries built from different sources for the analysis of interactions by AP-MS. The "AP-specific" spectral library was generated by paired acquisition of DDA data on the same samples analyzed by DIA; the "SWATH-Atlas" is a large generic human spectral library for the AB SCIEX 5600+ Triple TOF instrument that was recently made available by another group (Rosenberger et al., *Scientific Data*, 2014). The same DIA files were also searched using the untargeted identification module of DIA-Umpire. (a) Number of unique peptides identified (averaged; 3 biological replicates); (b) Number of proteins identified (averaged; 3 biological replicates). Please note that these samples were run with a 50 ms MS1 scan time, which is suboptimal for DIA-Umpire (250 ms MS1 acquisition was performed in Tsou et al., *Nature Methods*, 2015, and for the lysate samples used in this manuscript).
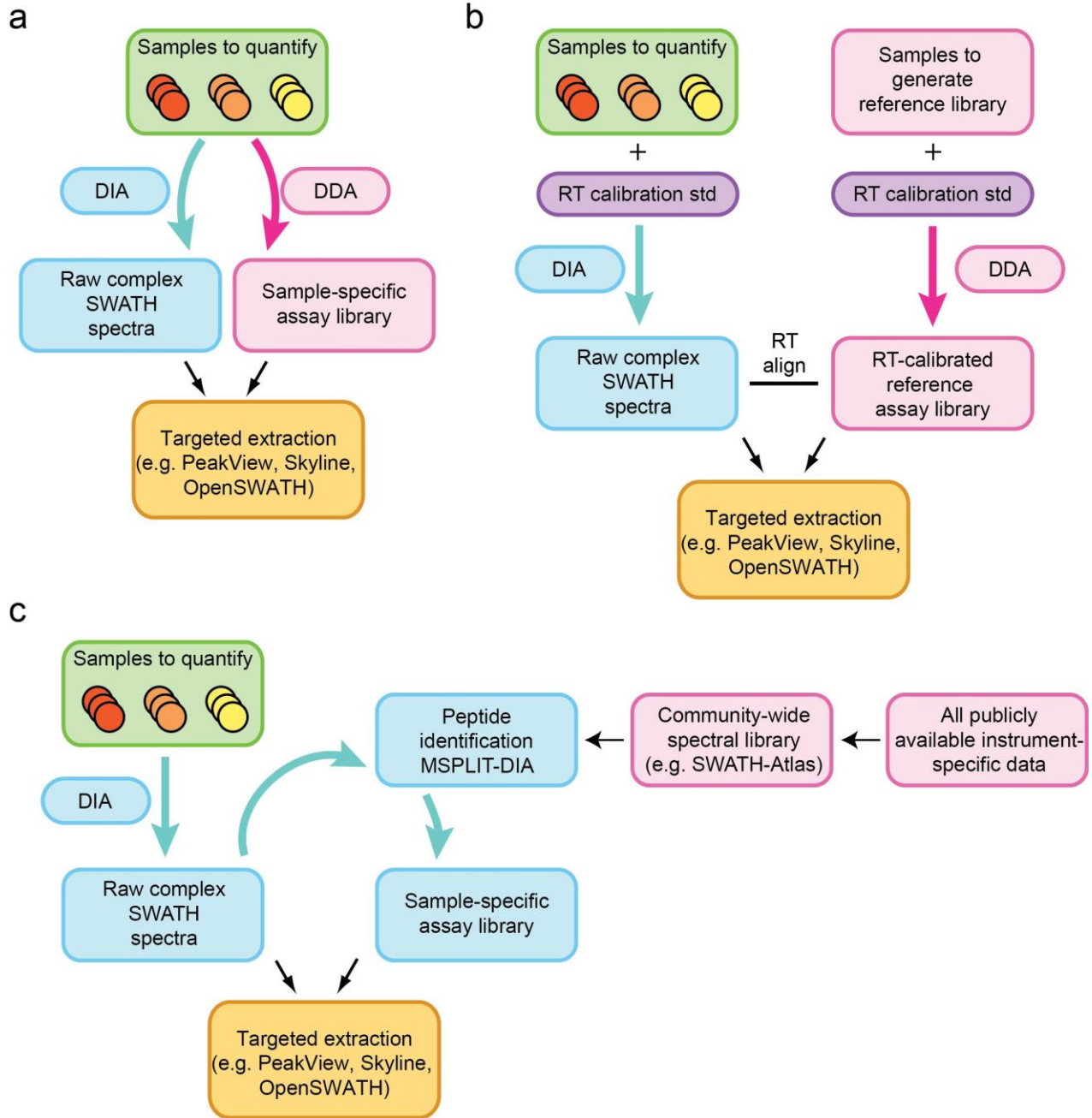
Supplementary Figure 7

Detection of protein-protein interactions using DDA vs. DIA

Peptides were identified using MSPLIT-DIA or MSGFDB-DDA at 1% FDR. The identified peptides and their respective spectral counts were analyzed by SAINTexpress (Teo et al., *J Proteomics*, 2014) to detect protein-protein interactions by identifying proteins with significant enrichment over a control GFP sample. (a) Number of MS/MS spectra identified by MSGFDB-DDA (blue) or MSPLIT-DIA (AP-specific library, green; SWATH-Atlas library, purple). The absence of dynamic exclusion in DIA alongside periodic sampling leads to 3 to 4–fold increase in spectral counts over DDA for these affinity purified samples. (b) High confidence (≤1% FDR) protein-protein interactions as scored by SAINTexpress. MSPLIT-DIA was able to identify over 90% of the interacting proteins detected by MSGFDB-DDA (blue bars) and an additional 40% of interactors not identified by DDA (also see **Fig. 1e**). MSPLIT-DIA performs robustly with spectral libraries from different sources (in-house AP-specific or generic SWATH-Atlas) in scoring interactions with SAINTexpress. MSPLIT-DIA thus essentially eliminates the need to construct AP-specific libraries as previously proposed (Lambert et al., *Nature Methods*, 2013) when analyzing DIA data from AP-MS studies (see main text for comments).
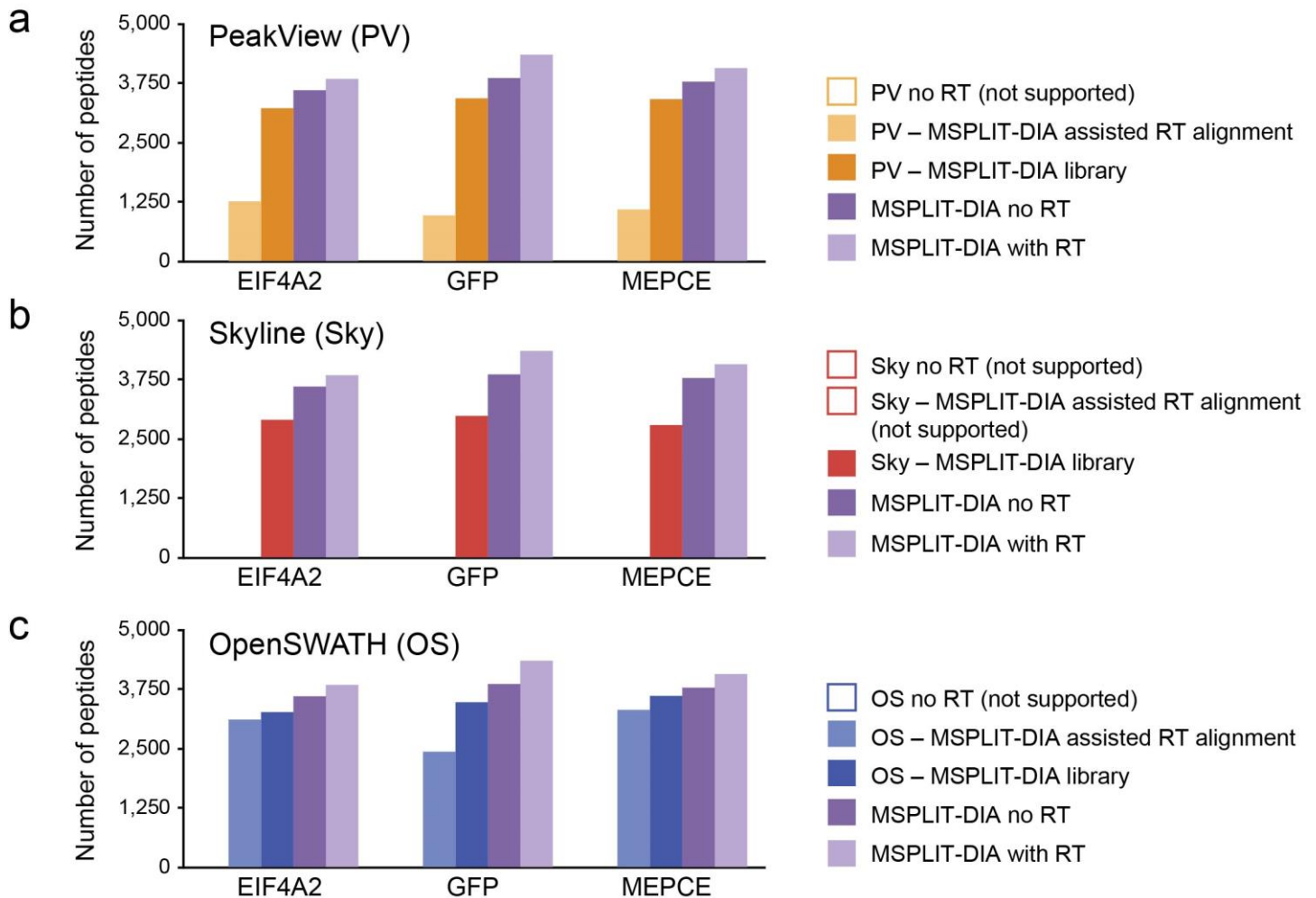
Supplementary Figure 8

MSPLIT-DIA improves targeted extraction from DIA data.

(a) Generation of targeted extraction libraries as in Lambert et al. (*Nature Methods*, 2013): the same samples are each analyzed twice, once by DDA for the generation of a reference spectral library and once by DIA for quantification. While analyzing the same sample by DIA and DDA on the same instrument largely helps with issues of retention time alignment and spectral similarity, this comes at the cost of doubling the acquisition time and the amount of sample required. (b) Use of external retention time calibration standards spiked in both the sample to quantify by DIA and the DDA sample used to build the reference spectral library (as in Escher et al., *Proteomics*, 2012) can help with normalizing the retention times, but the local alignment may not be perfect if different chromatography systems are utilized. (c) MSPLIT-DIA circumvents issues in (a) and (b), by first identifying the peptides present in the sample at a fixed FDR, then generating a sample-specific assay library from the identified peptides. Importantly, since the peptides were identified in DIA, their true retention time in the DIA sample is known, circumventing the need for external calibration standards.
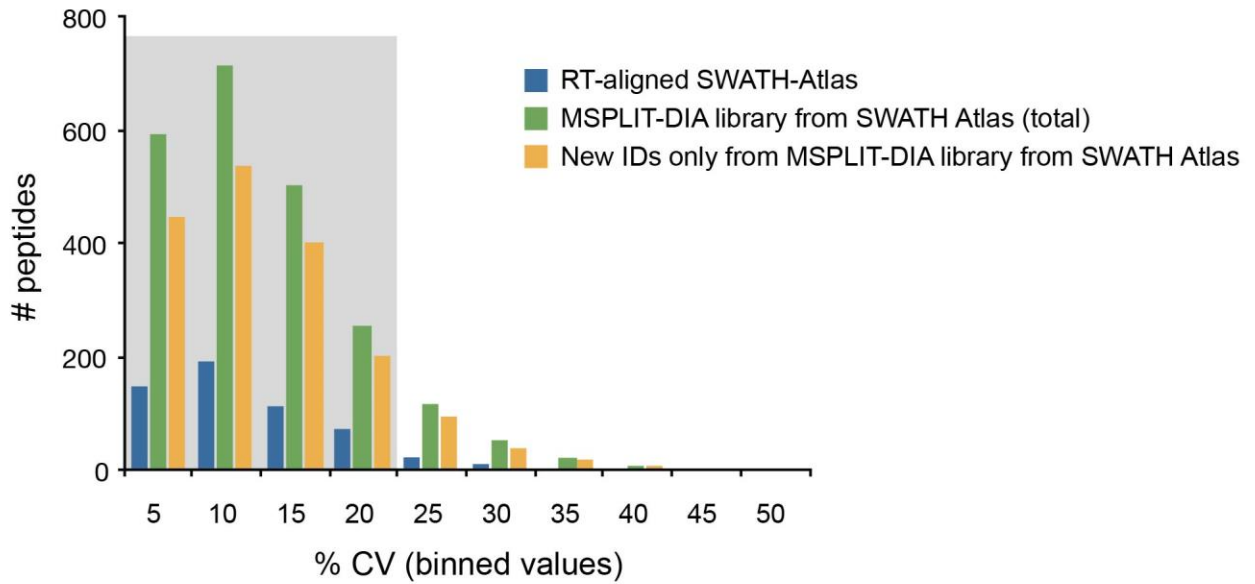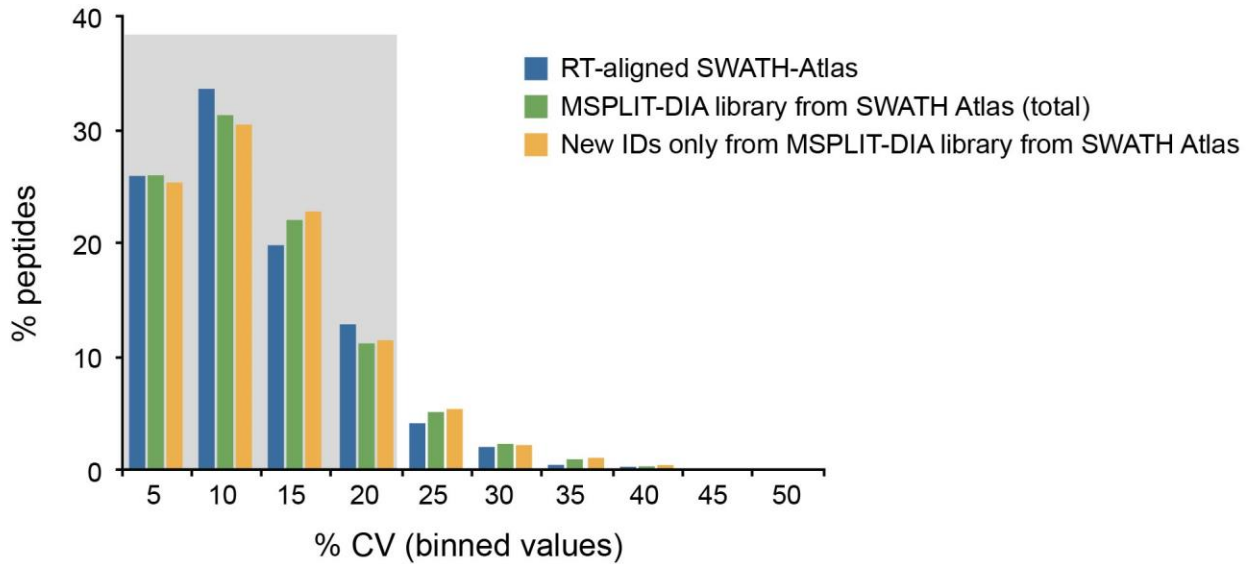
Supplementary Figure 9

MSPLIT-DIA improves peptide detection by targeted methods in AP-MS samples using proteome-scale spectral libraries.

We benchmarked the performance of PeakView (a; PV), Skyline (b; Sky) and OpenSWATH (c; OS) in analyzing DIA data from affinity purification (AP) samples using the SWATH-Atlas library, a publicly available proteome-scale human spectral library (Rosenberger et. al. *Scientific Data,* 2014). Targeted approaches were not suitable* to analyze DIA data in a very large search space (i.e., without retention-time (RT) information and using this large organism-scale spectral-library) while MSPLIT-DIA was able to identify thousands of peptides (dark purple bars). As described in **Supplementary Note** and illustrated in **Supplementary Figure 8**, MSPLIT-DIA can be used in conjunction with targeted approaches to improve the sensitivity of peptide extraction in DIA data. In brief, peptides identified by MSPLIT-DIA without RT information can be used to align retention times between a DIA run and the spectral-library. This allows targeted approaches to search only within a small RT range, enabling PeakView (light orange) and OpenSWATH (light blue) to successfully extract information for a subset of MSPLIT-identified peptides ("MSPLIT-DIA-assisted RT alignment"). In this mode, MSPLIT-DIA essentially provides a computational alternative to the "iRT-peptides" approach (Escher et al. *Proteomics,* 2012) that was previously shown to be effective, while eliminating the need to spike-in iRT reference peptides in every single MS injection. In addition, MSPLIT-DIA can also be used to generate sample-specific assay libraries for targeted approaches. This reduces the number of peptides in the assay library by ~50 fold (from ~200,000 peptides to ~4,000 peptides) while also providing an accurate RT where the peptides were detected in DIA data, thus leading to further improvement in the number of peptides that can be extracted by targeted approaches ("MSPLIT-DIA library" shown by the darker orange, red and blue, respectively, in panels a-c). Although MSPLIT-DIA was designed to operate without requiring RT information, when normalized retention times are available in the library, MSPLIT-DIA can also use these to filter out false positive matches and thus further improve the number of identified peptides (light purple bars).

*For very large searches, targeted tools took a very long time to process single data file (e.g. > 4 days). In order to make the analysis practical, if a tool did not finish analyzing a particular data file in 48 hours, we reported it as 'not supported' (see legend).

Supplementary Figure 10

Reproducibility of quantification by targeted extraction after MSPLIT-DIA analysis.

Using the EIF4A2 triplicate samples shown in **Supplementary Figure 9**, we evaluated the reproducibility in quantification by PeakView after performing the MSPLIT-assisted retention time alignment of the SWATH-Atlas library (blue bars), or after generating a MSPLIT-DIA sub-library from only those identified peptides (green bar). The peptides which could only be identified by the smaller MSPLIT-DIA library are shown in orange. The percentage CV (binned values) for peptides which could be detected across the three replicates is shown in total number of peptides (a), or as a percentage of the peptides within a CV bin (b).