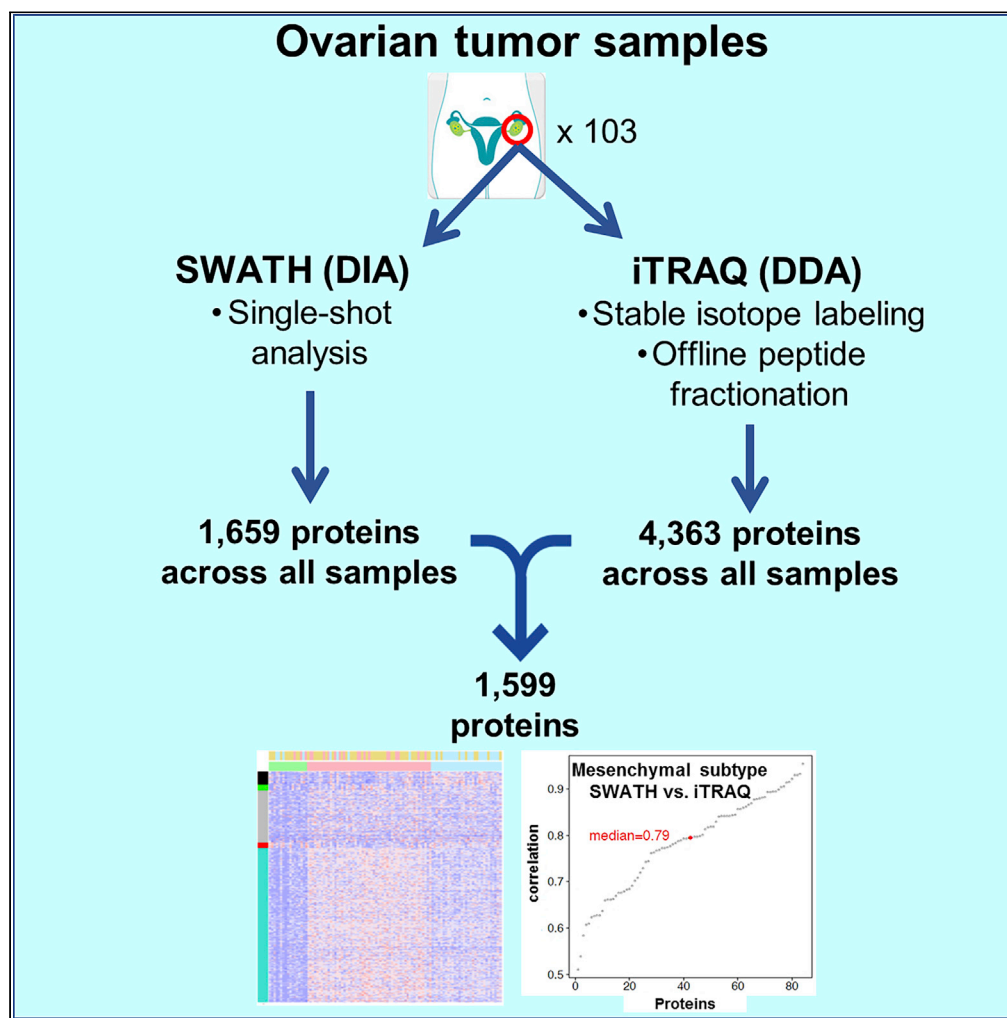


Article

Orthogonal Proteomic Platforms and Their Implications for the Stable Classification of High-Grade Serous Ovarian Cancer Subtypes



Stefani N. Thomas, Betty Friedrich, Michael Schnaubelt, Daniel W. Chan, Hui Zhang, Ruedi Aebersold

huizhang@jhu.edu (H.Z.)
aebersold@imsb.biol.ethz.ch (R.A.)

HIGHLIGHTS

SWATH-MS and iTRAQ-DDA are used to classify 103 high-grade serous ovarian cancer

SWATH-MS re-capitulates differentially expressed proteins in ovarian cancer subtypes

SWATH-MS is a robust proteomic approach for large-scale clinical proteomic studies

DATA AND CODE**AVAILABILITY**

PXD010437

Thomas et al., iScience 23, 101079
June 26, 2020 © 2020 The Author(s).
<https://doi.org/10.1016/j.isci.2020.101079>

Article

Orthogonal Proteomic Platforms and Their Implications for the Stable Classification of High-Grade Serous Ovarian Cancer Subtypes

Stefani N. Thomas,^{1,4} Betty Friedrich,^{2,4} Michael Schnaubelt,¹ Daniel W. Chan,¹ Hui Zhang,^{1,5,*} and Ruedi Aebersold^{2,3,*}

SUMMARY

The National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC) established a harmonized method for large-scale clinical proteomic studies. SWATH-MS, an instance of data-independent acquisition (DIA) proteomic methods, is an alternate proteomic approach. In this study, we used SWATH-MS to analyze remnant peptides from the original retrospective TCGA samples generated for the CPTAC ovarian cancer proteogenomic study. The SWATH-MS results recapitulated the confident identification of differentially expressed proteins in enriched pathways associated with the robust Mesenchymal high-grade serous ovarian cancer subtype and the homologous recombination deficient tumors. Hence, SWATH/DIA-MS presents a promising complementary or orthogonal alternative to the CPTAC proteomic workflow, with the advantages of simpler and faster workflows and lower sample consumption, albeit with shallower proteome coverage. In summary, both analytical methods are suitable to characterize clinical samples, providing proteomic workflow alternatives for cancer researchers depending on the context-specific goals of the studies.

INTRODUCTION

Advances in sample preparation workflows, mass spectrometry instrumentation, and data processing software have positioned proteomics to provide comprehensive insights into complex biological processes at a level close to the underlying biochemical mechanisms. Indeed, it is currently possible to routinely quantify >10,000 proteins in human cell proteomes (Beck et al., 2011; Nagaraj et al., 2011) and human tissue proteomes (Kim et al., 2014; Mertins et al., 2018; Wilhelm et al., 2014) using mass spectrometry-based platforms. The workflows for many of these large-scale proteomic studies entail extensive offline fractionation of the peptides generated from enzymatically digested proteins, followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Consequently, proteomic analysis of large cohorts of clinical specimens (>100) requires several months for data acquisition using the aforementioned workflows. Furthermore, because each fraction analyzed typically requires 1–5 µg of total peptides, the required quantity of the original tissue sample is in the milligram level. Although more rapid proteomic workflows have been developed (Anagnostopoulos et al., 2017; Hebert et al., 2014; Kulak et al., 2014; Richards et al., 2015), they have not yet been deployed for large-scale clinical proteomic studies.

Examples of large-scale clinical proteomic studies using the 2DLC-MS/MS workflow described above include the National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC) studies. CPTAC was formed to accelerate the understanding of the molecular basis of cancer through the application of large-scale proteogenomic analyses. Several hundred tumor tissue specimens from breast, ovarian, and colorectal cancer tissues previously analyzed by NCI's The Cancer Genome Atlas (TCGA) have also been characterized using proteomics, informed by genomics, resulting in the identification and quantification of proteins and phosphoproteins in cancer-associated cell signaling pathways and networks (Clark et al., 2020; Dou et al., 2020; Mertins et al., 2016; Zhang et al., 2014a, 2016a). These studies employed data-dependent acquisition (DDA) mass spectrometry, a mode of MS/MS data collection wherein a fixed number of precursor ions whose *m/z* values were recorded in a survey scan are selected for fragmentation using a pre-determined set of rules (Mann et al., 2001). DDA-based proteomic workflows have undergone considerable optimization to improve the reliability and reproducibility of the generated data in an effort to minimize the limitations due to the

¹Department of Pathology, Clinical Chemistry Division, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

²Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland

³Faculty of Science, University of Zürich, Zürich, Switzerland

⁴These authors contributed equally

⁵Lead Contact

*Correspondence: hui Zhang@jhu.edu (H.Z.), aebersold@imsb.biol.ethz.ch (R.A.)

<https://doi.org/10.1016/j.isci.2020.101079>



stochastic nature of precursor ion selection and low sampling efficiency resulting in missing values across datasets (Mertins et al., 2018; Revesz et al., 2018; Tabb et al., 2016; Zhou et al., 2017).

A relatively newer method termed data-independent acquisition (DIA) mass spectrometry has been gaining traction in large-scale proteomic studies (Vidova and Spacil, 2017). DIA mass spectrometry is an alternative to DDA that allows all ions within a selected mass range to be concurrently fragmented and analyzed by tandem mass spectrometry. Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS) is an example of a DIA acquisition method whose use in proteomic studies has increased considerably within the past 5 years. The SWATH-MS method acquires a complete and permanent digital fragment ion record for all detectable precursor ions of a sample (Collins et al., 2013; Gillet et al., 2012; Liu et al., 2013) and provides an iterative, targeted search strategy that determines the presence and quantity of tens of thousands of query peptides using reference fragment ion spectra for the query peptides as prior information (Rost et al., 2014). To support SWATH/DIA data analysis, several software tools have been developed and benchmarked (Navarro et al., 2016). As with any analytical methodology that has potential widespread use, several studies have been conducted to optimize and evaluate the performance of SWATH-MS (Li et al., 2017b; Rardin et al., 2015). In a multi-laboratory study including 11 sites worldwide, SWATH-MS was shown to have a linear dynamic range exceeding four orders of magnitude with an inter-laboratory coefficient of variation (CV) of $22.0 \pm 17.4\%$ (Collins et al., 2017), demonstrating that SWATH-MS is a reproducible method for large-scale protein quantification.

To assess the potential of SWATH-MS in addressing some of the common limitations of DDA proteomic workflows, comparative analyses of SWATH-MS and DDA using isobaric tags for relative and absolute quantitation (iTRAQ) have been conducted (Basak et al., 2015; Bourassa et al., 2015; Zhang et al., 2014b). Basak et al. concluded that SWATH-MS and iTRAQ DDA are complementary techniques with a 60% overlap of the high-confidence quantifiable proteins identified by both methods using *Saccharomyces cerevisiae* as a model system when incorporating offline peptide fractionation into the LC-MS workflow (Basak et al., 2015). In their study, Basak et al. incorporated first dimension separation using strong cation exchange chromatography wherein the peptides were fractionated into six fractions followed by second dimension separation using reversed-phase chromatography prior to LC-MS analysis.

In the current study, we utilized SWATH-MS to analyze peptides from 103 HGSOC tumors that were previously analyzed by iTRAQ DDA as part of the NCI CPTAC study (Zhang et al., 2016a). The iTRAQ DDA proteomic workflow resulted in the identification of 8,597 proteins from these tumors using a 24-fraction peptide separation method, whereas 2,914 proteins were quantified by SWATH-MS without peptide fractionation. We compared the two proteomic workflows on the basis of cost, robustness, complexity, ability to detect differential protein expression, and the elucidated biological information. Our analysis demonstrated that despite the greater than 2-fold difference in the analytical depth of iTRAQ DDA compared with SWATH-MS common differentially expressed proteins in enriched pathways associated with the HGSOC Mesenchymal subtype were identified by both workflows with 96% of the proteins quantified by SWATH-MS also quantified by iTRAQ DDA. We also showed that tumor subtype classification stability is sensitive to the number of samples that are analyzed. Lastly, our results indicated a conservation of the homologous recombination deficiency (HRD)-associated enriched DNA repair and chromosome organization pathways in the iTRAQ DDA and SWATH-MS datasets, thus indicating that some biological information for HGSOC could be consistently extracted from either dataset.

Taken together, compared with the DDA analysis of HGSOC using iTRAQ labeling and 2D fractionation, SWATH-MS is a robust proteomic method that can be used to re-capitulate common differentially expressed proteins in enriched pathways associated with the HGSOC Mesenchymal subtype (Zhang et al., 2016a). The SWATH-MS proteomic workflow is simpler, cheaper, and consumes less sample, but it results in shallower proteome coverage. The significantly lower number of proteins detected by SWATH-MS compared with the iTRAQ DDA workflow is mitigated by the streamlined and less complex workflow, the increased sample throughput (requiring ~80% less time), ~10-fold reduced sample requirements, and lower technical variability and attenuated signal compression. The SWATH-MS workflow therefore presents novel opportunities to enhance the efficiency of clinical proteomic studies with continuous method improvements.

Ludwig et al. provide an in-depth overview of the advantages and limitations of SWATH-MS in comparison with DDA (Ludwig et al., 2018). There are analytical challenges common to both workflows that should be taken into account when deciding on an optimal approach for a given biological or clinical problem. These challenges include sample complexity, dynamic range of the measured proteins, accuracy of the protein measurements, use of incomplete databases, and the selection of peptides for protein quantification. The strengths and weaknesses of SWATH-MS and iTRAQ DDA must be considered in the context of the afore-mentioned analytical challenges.

RESULTS

Study Design and Evaluation of Technical Performance of Analytical Platforms

Proteomic measurements of clinically annotated HGSOEC previously characterized by TCGA (Cancer Genome Atlas Research, 2011) were conducted using an iTRAQ DDA workflow entailing stable isotope labeling, offline fractionation, and LC-MS/MS analysis of each fraction (Zhang et al., 2016a). A total of 103 of these tumors were used for the current SWATH-MS analysis. An overview of the experimental design of our study is shown in Figure 1.

Protein was extracted from each tumor specimen followed by enzymatic digestion with trypsin. For the iTRAQ DDA workflow, the resulting peptides were labeled with 4-plex iTRAQ reagents followed by combination into analysis sets comprising the peptides from three tumors, each labeled with a distinct iTRAQ tag, and an iTRAQ-labeled reference pool comprising the peptides from most of the tumors. Each analysis set was subjected to offline fractionation into 24 concatenated fractions, and the fractions from each analytical set were sequentially analyzed using a DDA method on an LTQ-Orbitrap Velos mass spectrometer. An unfractionated aliquot of each analytical set was also analyzed by DDA-MS without fractionation. In comparison, the SWATH-MS workflow did not require stable isotope labeling. However, for the purpose of generating a spectral library to facilitate the targeted protein identification, each sample from 103 tumors was pooled, fractionated to 48 fractions and subjected to DDA analysis, which is an optional step for this workflow. We note that, although we chose in this study to create a sample-specific reference library, several publicly available reference libraries exist (Ludwig et al., 2018). The implications of using a generic human library on the false discovery rate (FDR) due to multiple hypothesis testing correction have been addressed (Rosenberger et al., 2017a). Another option is to avoid acquiring DDA runs to generate the spectral library by utilizing spectrum-centric scoring of DIA data based on the use of algorithms such as DIA-Umpire (Li et al., 2015; Tsou et al., 2015).

The CV of SWATH-MS data was assessed using two QC approaches (Figure S1). For the first method, peptides from HEK293 cells were analyzed in triplicate on three days for a total of nine runs between the DDA data acquisition runs for spectral library generation and the SWATH acquisition of the ovarian tumor data. The median CV for 3,855 quantified proteins was 8%, and the mean total CV (reflecting the intra- and inter-day CV) was 15% for the nine technical repeat analyses (Collins et al., 2013, 2017) (Figure S1A). For the second QC method, peptides from a control ovarian tumor were analyzed using a DDA method on the same 5600+ TripleTOF mass spectrometer that was used to acquire the SWATH data. These QC samples were run in duplicate immediately prior to the SWATH acquisition of the ovarian tumor samples, and again in duplicate 10 days later when half of the ovarian tumor sample data acquisition runs were completed. These measurements resulted in a total CV of 7% for 781 quantified proteins (Figure S1B). For both QC strategies, the duration of the LC gradient was identical to the gradient used for the SWATH analysis of the HGSOEC samples. Hence, the results from these QC strategies evaluated the analytical measurement/technical variability of the SWATH-MS platform that was used to acquire the data from the ovarian tumor proteins.

We used the 1,599 proteins quantified by iTRAQ DDA and SWATH-MS to compare the analytical differences between the two workflows with respect to relative protein abundances (Figure 2A), correlation of the normalized relative abundance of the quantified proteins (Figure 2B), and variability of the constituent peptides from the quantified proteins (Figure 2C; y axis indicates the peptide variability calculated by the standard deviation of the quantified peptides for each protein divided by the mean peptide intensity per protein). Relative protein abundances were determined based on the normalized \log_2 intensity ratios compared with the reference iTRAQ channel of each iTRAQ set for the iTRAQ data, and relative protein abundances in the SWATH-MS dataset were determined based on the \log_2 intensity of each protein as represented by the mean intensity of the constituent peptide ratios (peptide

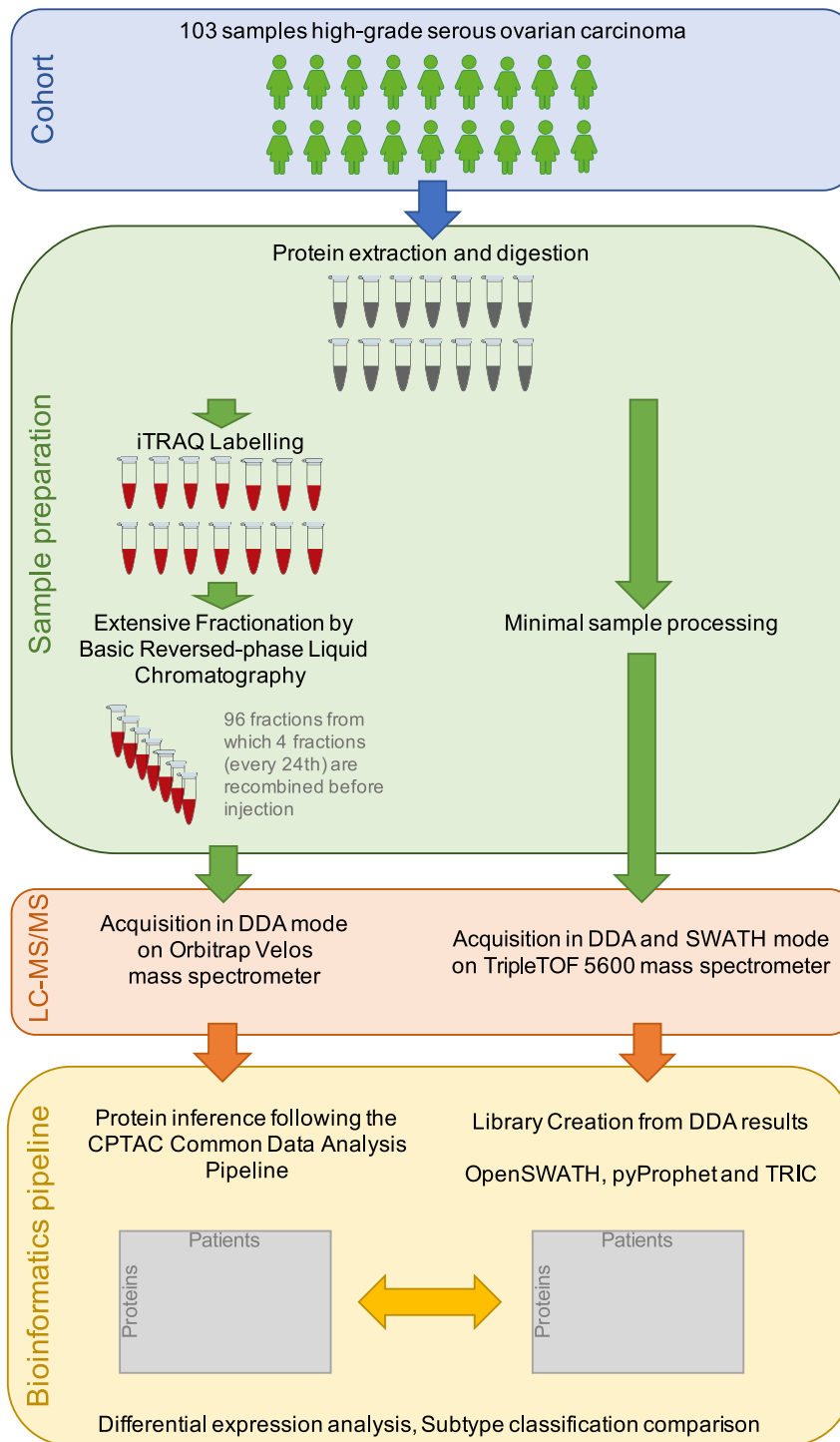


Figure 1. Experimental Design of Proteomic Analysis of High-Grade Serous Ovarian Tumors Using DIA Mass Spectrometry Instrument Platforms

A total of 103 clinically annotated ovarian high-grade serous carcinomas previously characterized by The Cancer Genome Atlas (TCGA) were processed for proteomic analysis using an iTRAQ DDA method wherein the samples were subjected to fractionation prior to data acquisition and also using a SWATH-MS method without fractionation prior to analysis. The data processing and bioinformatics pipeline enabled differential expression analysis and tumor subtype classification comparison.

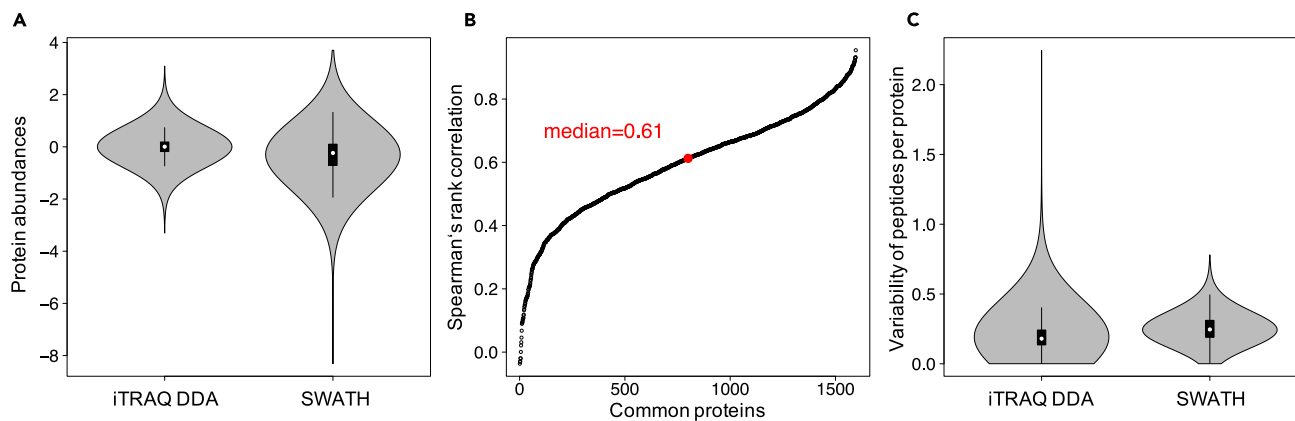


Figure 2. Comparison of iTRAQ DDA and SWATH-MS Proteomic Data Based on Completeness and Variability

The 1,599 proteins quantified by iTRAQ DDA and SWATH-MS were used for the following comparisons: (A) Distribution of relative protein abundance in the iTRAQ DDA and SWATH-MS datasets. (B) Spearman's rank correlation of proteins quantified by iTRAQ DDA and SWATH-MS. (C) Peptide variability calculated by the standard deviation of peptides per protein divided by mean peptide intensity per protein.

intensity divided by the mean peptide intensity) (Figure 2A). The median relative \log_2 relative protein abundance of the iTRAQ data was 0.01 compared with -0.23 in the SWATH-MS data. The compressed distribution of the quantified relative protein abundances of the iTRAQ DDA data reflects the well-documented phenomenon of ratio compression in iTRAQ-based relative quantification (Ow et al., 2011; Savitski et al., 2013) (Figure 2A).

Spearman's rank correlation was used to assess the strength of the association of the proteins quantified by iTRAQ DDA and SWATH-MS. The median ρ of 0.61 indicates a moderately positive correlation (Figure 2B). Among the factors that likely preclude the median ρ from being higher are the fundamental differences in protein quantification (\log_2 ratio of reporter ion intensities in the iTRAQ DDA data compared with the summed fragment ion intensities and normalized to the respective mean peptide intensities for the SWATH-MS data) and the differences in protein quantification wherein multiple peptides from the same protein were used to quantify each protein in the dataset. Additionally, for the majority of the quantified proteins, different constituent peptides and different numbers of unique peptides were identified in the iTRAQ DDA workflow compared with SWATH-MS. A direct comparison of the CV of each method was not possible because replicate analyses of the samples were not conducted; therefore, a CV-like score was calculated, which represents the biological variability of the sibling peptides from the quantified proteins. Because aliquots of identical samples were used for this analysis, the displayed variation reflects technical differences in the measurements and data handling of both methods. The median variability of the SWATH-MS data was higher than the iTRAQ DDA data (0.25 versus 0.18, respectively). As shown in the violin plots, the variability of the iTRAQ DDA data has a wider distribution than that of the SWATH-MS data (Figure 2C).

Both assessed proteomic methods yielded different numbers of quantified proteins: a cumulative total of 8,597 quantified proteins resulting from the iTRAQ DDA analysis and a cumulative total of 2,914 quantified proteins resulting from the SWATH-MS analysis. A total of $1 \mu\text{g}$ of each of the 24 fractionated peptide samples was used for the iTRAQ DDA analysis, and $1 \mu\text{g}$ of peptide samples from each tumor was injected for the SWATH-MS analysis. In iTRAQ DDA, we used the original quantified proteomic data from Zhang et al. (2016a), and we only used values obtained from unshared peptides with other proteins (proteotypic peptides). The SWATH-MS data were filtered on a peptide-query level FDR of 1% and a protein-level global FDR cutoff of 1% based on the method described by Rosenberger et al. for the statistical control of peptide and protein error rates in large-scale DIA analyses (Rosenberger et al., 2017a). In addition, only proteotypic peptides were used for quantification for this dataset.

For any proteomic dataset consisting of multiple analyses of biologically distinct samples, not every protein will be detected in every sample, and the resulting data matrix will have missing values. This can either be due to technical reasons, i.e., not every protein present in a sample will be identified owing to the stochastic nature of the sampling of ions by the mass spectrometer or reasons related to biological variability. We

therefore assessed the occurrence and distribution of missing values in the two datasets. [Figure S2](#) shows the distribution of the proteins with missing values in the iTRAQ and SWATH data. Here, we show the number of proteins with various percentages of missing values in the iTRAQ DDA ([Figure S2A](#)) and SWATH-MS datasets ([Figure S2B](#)), starting with proteins without any missing values across all 103 samples up to proteins with more than 91% missing values. In the iTRAQ dataset, approximately 50% of the proteins did not have any missing values, whereas in the SWATH-MS dataset, 11% of the proteins did not have any missing values and 50% of the proteins had ~30% missing values. The iTRAQ DDA dataset was directly filtered for the proteins without missing values (4,363). The same approach was not applicable in SWATH-MS owing to the sparsity of the matrix. Only working with complete measurements in SWATH-MS would also result in neglecting potential biological effects, where proteins might not be detected owing to down-regulation or absence in a subset of the samples. Owing to the instrument duty cycle and the stochastic nature of data acquisition, more missing values are expected to occur in low-abundance proteins than in high-abundance ones ([Wang et al., 2007](#)). We assessed this as well in the current dataset and could indeed observe this inverse correlation ([Figure S2C](#)). Hence, to distinguish technical and biological reasons for “missingness” we applied a filtering strategy following this assumption by allowing more missing values in high-abundance proteins and less missing values among the low-abundance proteins, yielding 1,659 proteins, instead of taking the 4,363 proteins without missing value from our previous iTRAQ based DDA approach. To enable the use of a complete matrix, an imputation approach was adopted based on that used in the Perseus software program ([Tyanova et al., 2016](#)). Imputation has been demonstrated to be an effective approach to address the challenge of missing values in SWATH-MS and related DIA approaches ([Collins et al., 2017](#); [Karpievitch et al., 2012](#); [Rost et al., 2016](#)).

The robust filtering approach used for the SWATH-MS data resulted in 1,659 proteins that were quantified with high confidence across all 103 tumors. Among these proteins, 1,599 were also quantified using iTRAQ-DDA, and these are the proteins that were used for all subsequent analyses, including the elucidation of the molecular subtype classification of the tumors based on their proteomic signatures.

High-Grade Serous Ovarian Carcinoma Subtype Classification Based on Proteomic Signatures

The rationale for the molecular subtyping of HGSOC is related to efforts to develop more specific and effective therapeutic strategies given the heterogeneity of this type of ovarian cancer. To assess the ability to classify the iTRAQ-DDA and SWATH-MS data into subtypes based on proteomic signatures, we used the same approach as reported by Zhang et al. wherein an unbiased molecular taxonomy of HGSOC was established using relative protein abundance data to identify subtypes that exhibit biological differences ([Zhang et al., 2016b](#)). The 1,599 proteins that were quantified in the iTRAQ-DDA and SWATH-MS datasets were used to classify the two datasets separately using mclust ([Fraley and Raftery, 2002](#)) based on the z-score-transformed relative protein abundances, and the emergent protein modules were characterized using weighted gene-correlation network analysis (WGCNA) ([Langfelder and Horvath, 2008](#)) and Reactome pathway enrichment. The heatmap resulting from the clustering analysis is shown in [Figure S3A](#) with the proteins listed in horizontal rows and the tumors listed in vertical columns. The colored vertical bars represent the transcriptome-based HGSOC subtypes ([Verhaak et al., 2013](#)) (“Original TCGA”), the proposed proteomic iTRAQ DDA subtypes (“Original CPTAC” ([Zhang et al., 2016a](#)); “iTRAQ DDA”), and the proposed proteomic SWATH-MS subtypes (“SWATH-MS”).

The correlation of the WGCNA-derived protein modules with the proteomic subtypes derived from the iTRAQ DDA and SWATH-MS datasets is shown in [Figure S4](#). The enriched Reactome pathways in the WGCNA-derived modules in the SWATH-MS data include ECM organization, immune response, metabolism, complement cascade and fibrin clot formation, and gene expression and translation. In the SWATH-MS dataset, the tumors with proteins exhibiting a significantly positive correlation (Pearson correlation coefficient $\rho = 0.52$; p value = 2×10^{-8}) with the ECM organization Reactome pathway were assigned to the Mesenchymal subtype. A significantly positive correlation (Pearson correlation coefficient $\rho = 0.54$; $p = 5 \times 10^{-9}$) was also observed among the proteins assigned to the gene expression and translation ontology.

Based on the observation highlighted by the dashed box in [Figure 3](#) indicating a large degree of similarity among the genomic and proteomic profiles of the Mesenchymal subtype tumors, we conducted a down-sampling analysis in an effort to evaluate the stability of the Mesenchymal subtype. In this analysis, the number of tumors used for the cluster analysis was systematically decreased to determine the sample number cutoff below which the enrichment of the Reactome pathways was no longer significant. The Mesenchymal

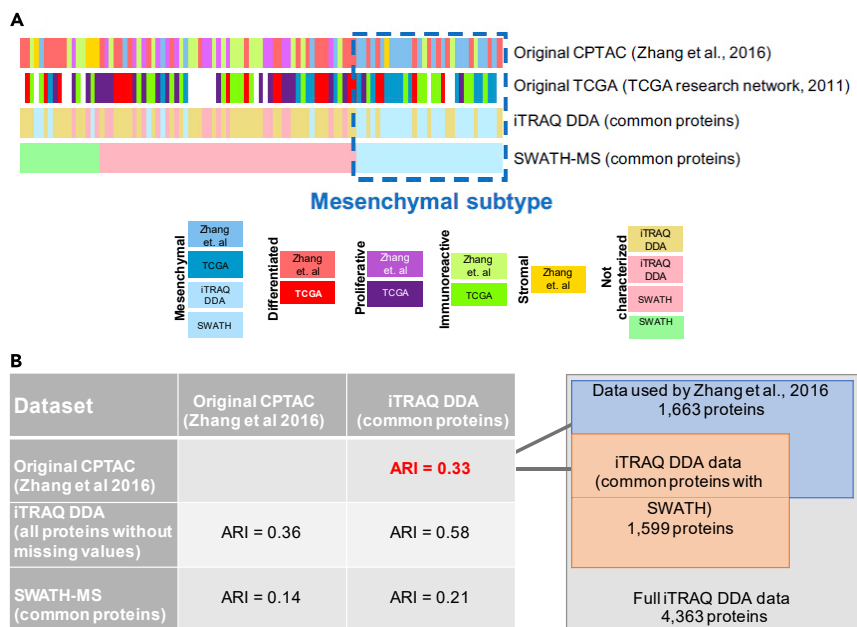


Figure 3. Protein (iTRAQ DDA, SWATH-MS) and mRNA-Based High-Grade Serous Ovarian Cancer Subtype Classification

(A) Tumor subtype classification comparison based on gene expression (TCGA study) and relative protein abundance (original CPTAC study [Zhang et al., 2016a], and the iTRAQ DDA and SWATH-MS data from the current analysis). Subtype classification: blue, Mesenchymal; red, Differentiated; purple, Proliferative; green, Immunoreactive; yellow, Stromal.
(B) Subtype classification agreement among the genomic and proteomic data assessed by the Adjusted Rand Index (ARI).

cluster was the only cluster for which the enrichment of specific Reactome pathways remained significant ($p < 0.01$, Fisher's exact test) when the sample number was reduced to 60% (Figure S3B). These results suggest that the SWATH-MS and iTRAQ DDA proteomic signatures of the Mesenchymal subtype tumors are consistent and robust.

Because the protein and mRNA components of the same ovarian tumors were analyzed, we were able to compare the robustness of the Mesenchymal subtype across different analytical methods. Figure 3A compares the subtyping of the tumors based on the five protein-based subtypes (Differentiated "D," Immunoreactive "I," Proliferative "P," Mesenchymal "M," and Stromal "S") that emerged from the original iTRAQ DDA-based analysis of these tumors (Zhang et al., 2016a), the four mRNA-based subtypes (Differentiated, Proliferative, Mesenchymal, and Immunoreactive) that emerged from the genomic analysis (Cancer Genome Atlas Research, 2011), and the three protein-based subtypes that resulted from the clustering analysis of the iTRAQ DDA and SWATH-MS data using the 1,599 proteins that were quantified using these two analytical approaches. Based on visual analysis, the Mesenchymal subtype tumors (represented by shades of blue) exhibit the highest degree of classification agreement among the four analytical approaches. Although the accurate characterization of the molecular subtypes of HGSOc is challenging, it is widely accepted that the Mesenchymal subtype is defined by the increased expression of extracellular matrix proteins and desmoplasia (Chen et al., 2018). Compared with the other data types, the SWATH-MS data facilitates the discrete partitioning of one of the five proteomic subtypes, the Mesenchymal subtype tumors, from our previous five proteomic subtypes (Figure 3A).

Influence of Sample Size and Analytical Depth on Tumor Molecular Subtype Classification Stability

We used the Adjusted Rand Index (ARI) to quantitatively assess the agreement among the classification of the tumors based on the proteomic analyses (iTRAQ DDA and SWATH-MS) (Figure 3B). ARI values range from 0 to 1, with 1 indicating clustering results that are identical and 0 indicating clusters that are devoid of similarity. Although we did not anticipate an ARI of 1 when comparing the iTRAQ DDA and SWATH-MS datasets using the commonly quantified 1,599 proteins owing to the inherent noise in biological data, we did not

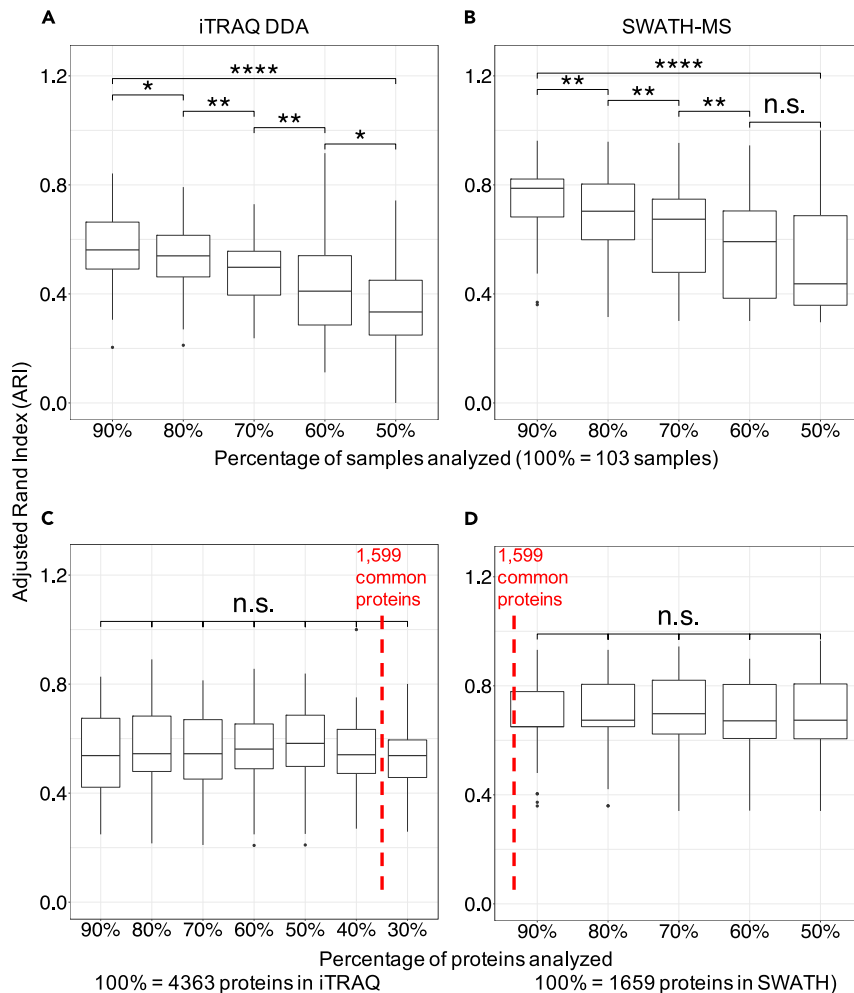


Figure 4. Proteomic Cluster Stability as a Function of Sample Number and Protein Number in the iTRAQ DDA and SWATH-MS Datasets

Cluster stability measured by the Adjusted Rand Index (ARI) as a function of the percentage of tumors analyzed in the iTRAQ DDA (A) and SWATH-MS (B) datasets. (C and D) Cluster stability assessed by ARI as a function of the percentage of quantified proteins in the iTRAQ DDA (C) and SWATH-MS (D) datasets. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$. n.s. non-significant

expect the relatively low ARI of 0.21. To further explore the similarity among the various proteomic datasets mentioned in Figure 3A, we calculated the ARI values that resulted from changing the numbers of proteins included in the analysis. The highest ARI was 0.58, which was obtained by comparing the complete iTRAQ DDA dataset without any missing values (4,363) with the 1,599 proteins in the iTRAQ DDA dataset that were also quantified by SWATH-MS. This ARI was unexpectedly low given the expected level of similarity when comparing a subset to the full set of exactly the same data. Conversely, the lowest ARI, 0.14, was obtained when comparing the full iTRAQ DDA dataset of 8,597 proteins with the 1,599 proteins that were quantified in common using SWATH-MS. This low ARI value of 0.14 is likely reflective of the fundamental differences in protein quantification between iTRAQ DDA and SWATH-MS.

We further examined the unexpected lack of similarity when comparing the subset of the iTRAQ data to the full dataset by adopting a systematic bootstrapping approach. Bootstrapping was used to randomly select a fraction of either the 103 samples analyzed by iTRAQ DDA and SWATH-MS (Figures 4A and 4B) or the proteins that were quantified by iTRAQ DDA (4,363) and SWATH-MS (1,659) (Figures 4C and 4D). ARI values based on mclust classification were used to determine the robustness of the clusters. This bootstrapping was repeated 100 times for each subset of the proteins or samples. The resulting classifications were compared with the result

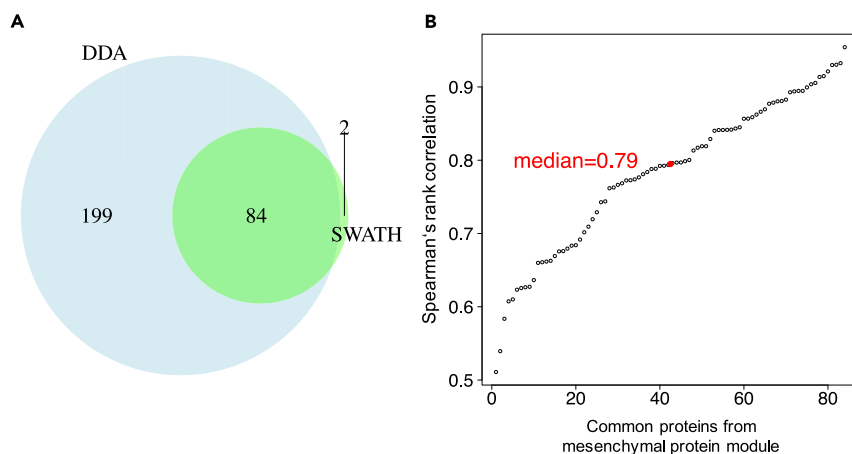


Figure 5. Comparison of Proteins Comprising the Mesenchymal Subtype Tumors in the iTRAQ DDA and SWATH-MS Data

(A) High overlap of proteins characterizing the Mesenchymal subtype in the SWATH MS data and the iTRAQ DDA data. (B) Strong positive correlation of 84 proteins commonly comprising the Mesenchymal subtype in the iTRAQ DDA and SWATH-MS data (median $\rho = 0.79$).

from the complete initial dataset, and the distribution of ARI values is shown in Figure 4 for the iTRAQ DDA (left column, A, C) and SWATH-MS datasets (right column, B, D). A lack of statistical significance among the ARI values (n.s.) indicates cluster stability, whereas cluster instability is indicated by statistically significant changes in ARI. When we bootstrapped along the dimension of sample number, the average ARI decreased dramatically with the reduced number of samples, even when comparing 90% versus 80% of the samples in the iTRAQ DDA and SWATH-MS datasets (Figures 4A and 4B, respectively). However, the corresponding median ARI values of the incrementally reduced sample sizes in the SWATH-MS dataset systematically exceeded the median ARI values of the iTRAQ DDA dataset (Figure 4B). These results indicate that the fidelity of molecular clusters is rather sensitive to the number of samples included in the analysis and that the clusters identified from the SWATH-MS data generally have slightly higher confidence. In comparison, when we bootstrapped along the dimension of the number of quantified proteins (Figures 4C and 4D) the ARI was refractory to the number of proteins included in the clustering. The difference among the ARI values did not reach statistical significance at a significance level of 0.05 (i.e., the clusters were stable) when the number of quantified proteins was down-sampled to 30% and 50% of the initial group of proteins in the iTRAQ DDA and SWATH-MS datasets, respectively. Hence, the resulting clusters are largely insensitive to the number of included proteins.

The sensitivity of molecular subtype clustering to the number of samples included in the analysis has been shown previously (Levine and Domany, 2001; Monti et al., 2003); however, this phenomenon has not been demonstrated in a systematic manner for proteomics data. Increasing the number of samples included in subtype classification analyses renders the classifications more robust, whereas increasing the numbers of proteins by employing proteomics methods such as 2DLC-MS/MS iTRAQ DDA does not provide an added benefit in terms of cluster stability. Our bootstrapping approach clearly demonstrated how easily classification results can be perturbed by experimental variables such as sample size, with a negative impact on the ability to extract robust biological content.

Nevertheless, in our SWATH-MS analysis, we were able to extract one robust HGSOC subtype, which was characterized by the increased relative abundance of proteins with extracellular matrix functions (Figure S3A). The robustness of this presumptive Mesenchymal subtype as a function of sample number was demonstrated using different subsets of samples in the SWATH-MS dataset using a Fisher's exact test (Figure S3B). A fundamental premise of this stability analysis is that only the most robust information can be reproducibly extracted from the data. The emergence of the Mesenchymal phenotype as a robust subtype in the iTRAQ DDA and SWATH-MS data indicates that biologically relevant content can be robustly extracted from these orthogonal/complementary analytical methods. The proteins characterizing the Mesenchymal cluster in the SWATH data (86 proteins) and the iTRAQ DDA data (283 proteins) were enriched in extracellular matrix organization function, show an overlap of 84 proteins (Figure 5A), and have a significantly higher correlation of the relative abundance of constituent proteins compared with the

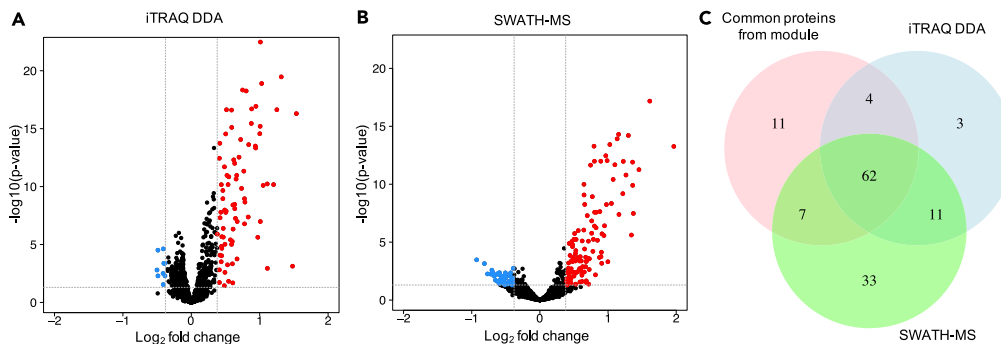


Figure 6. Group Comparison of Proteins in Mesenchymal Subtype versus the Non-mesenchymal Subtypes

Up-regulated proteins among the Mesenchymal subtype tumors are indicated in red and the down-regulated proteins are indicated in blue. The fold-change cutoff was 1.3, and the significance cutoff was $p < 0.05$.

(A) iTRAQ DDA data.

(B) SWATH-MS data.

(C) Comparison of all proteins assigned to the ECM module based on the original proteomic classification (Zhang et al., 2016a) versus proteins assigned to the ECM module in the iTRAQ DDA data and the SWATH-MS data.

proteins in the entire dataset of 1,599 proteins (Figure 5B, median $\rho = 0.79$ compared with Figure 2B, median $\rho = 0.61$). The relatively high correlation of the relative abundance of proteins comprising the Mesenchymal cluster is also evidenced by the overlap among the Mesenchymal subtype tumors (Figure 3A). These results suggest that the robustness of the clusters increases with increasing quantification accuracy rather than by increasing the number of quantified proteins.

To determine whether the same proteins from the iTRAQ DDA and SWATH-MS datasets would be identified in a group comparison of Mesenchymal subtype samples versus the other samples, we used the sample subtype annotations from the Zhang et al. iTRAQ DDA study (Zhang et al., 2016a). The \log_2 -fold changes of these proteins in the iTRAQ DDA and SWATH-MS datasets were determined as well as the associated p value (Figure 6). Proteins with significantly increased or decreased relative abundance (fold-change cut-off: 1.3; $p < 0.05$) are indicated in red or blue, respectively. The significantly upregulated proteins (Figures 6A and 6B, red dots) in these comparisons showed a high overlap with the 84 common proteins from the ECM protein module (Figure 6C) as defined by the Zhang et al. iTRAQ DDA study (Zhang et al., 2016a). Well-established cancer markers such as Fibronectin 1 and Thrombospondins that are known for contributing to the metastatic progression of tumors (Hu et al., 2017; Incardona et al., 1993; Kenny et al., 2014; Mitra et al., 2011; Ricciardelli et al., 2016) were among those proteins, confirming the characterization of these samples as Mesenchymal subtype tumors. Although there is a high overlap of these proteins between the iTRAQ DDA and SWATH datasets, 33 of the up-regulated proteins were not identified as such in the iTRAQ DDA dataset. Nevertheless, most of those 33 proteins are part of the extracellular region and exhibit enrichment in extracellular matrix organization as well. Also, of note, the quantitative dimension of iTRAQ DDA is narrower than that of the SWATH-MS data, which influences the extent of similarity among the proteins quantified by each method.

Since the proteomic subtypes other than the Mesenchymal subtype identified in the CPTAC and TCGA studies could not be clearly distinguished by the classification analyses of the datasets used in this study, we performed a similar differential expression analysis on those groups using the 1,599 proteins that were quantified in the iTRAQ DDA and SWATH-MS datasets (Figure S5). As expected, the p values of those comparisons were considerably less significant than those resulting from the analyses of the Mesenchymal subtype. The Differentiated and Stromal subtypes did not lead to any conclusive results in these analyses based on the functional enrichment of the proteins with increased or decreased relative abundances, whereas the Proliferative and Immunoreactive subtypes were characterized by several differentially expressed proteins with rather high p values (low significance). The WGCNA analyses of the respectively abundant proteins identified protein modules related to immune response and gene expression and translation (Figure S4). However, those modules did not show high associations with any of the patient groups.

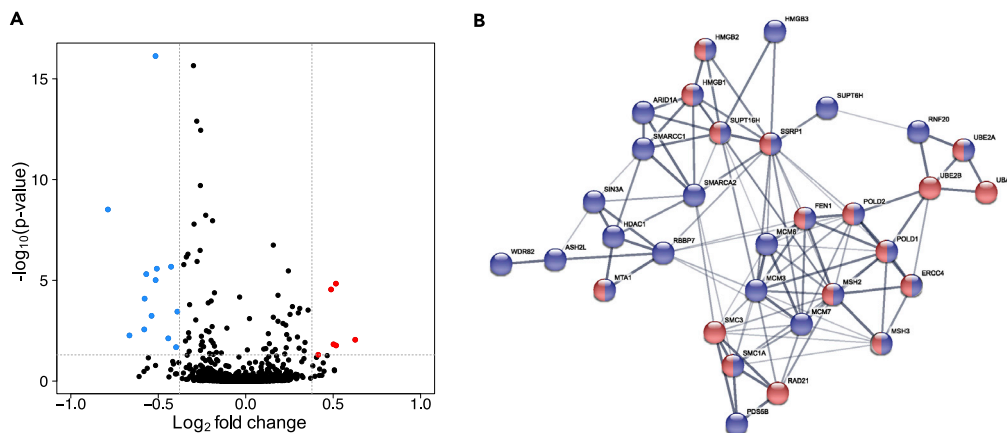


Figure 7. Group Comparison of Proteins in HRD versus Non-HRD Tumors Analyzed by SWATH-MS

(A) Relative abundance of up- (red) and down- (blue) regulated proteins in the HRD compared with the non-HRD tumors. The \log_2 -fold change cutoff was 1.3 and the significance cutoff was $p < 0.05$.

(B) Functional enrichment analysis using STRING. Blue nodes: Chromosome organization proteins. Red nodes: DNA repair proteins. Blue and red nodes: Proteins with roles in chromosome organization and DNA repair.

Homologous Recombination Deficiency-Related Proteomic Signature of HGSOC Identified in SWATH-MS Data

Previous studies have used gene expression and mutation profiles to characterize molecular subtypes of high-grade serous ovarian cancer to identify patients who respond well to poly-ADP ribose polymerase (PARP) inhibitor treatment (Lheureux et al., 2017; Tothill et al., 2008). Homologous recombination deficiency (HRD) is associated with a higher sensitivity toward PARP inhibitor treatment and therefore with a better prognosis for the respective patients. The initial iTRAQ DDA analysis of the HGSOC tumors resulted in the identification of a well-defined network of proteins with roles in histone acetylation that differentiated HRD from non-HRD tumors (Zhang et al., 2016a). Thus, we conducted an analysis to determine whether similar HRD-related features could be identified from the SWATH-MS data in the current study.

A group comparison of HRD versus non-HRD patients using mapDIA (Teo et al., 2015) revealed several differentially expressed proteins (Figure 7A). Because none of these proteins could be directly linked to DNA repair mechanism-related functions, we used a network propagation approach (Hofree et al., 2013) to extend the comparatively limited analytical depth of the SWATH measurements. As an input we used a network obtained from STRING (Szklarczyk et al., 2015) filtered for highly confident experimental evidence (physical interactions with a score higher than 800) with 4,424 nodes. \log_2 -fold changes obtained from the group comparison approach were mapped on this network, and these signals were propagated over the network to identify signals accumulating within a subnetwork. The top 5% (221 proteins) of the resulting positive and negative scores were used for further investigation and functional enrichment.

Further verifying that proteins with roles in histone acetylation differentiate HRD from non-HRD tumors (Zhang et al., 2016a), DNA repair and chromosome organization were among the GO terms that were identified using the previously described network propagation approach based on functional enrichment analysis. The respective sub-network contained proteins previously identified as belonging to the HRD-associated protein network, including histone deacetylase 1 (HDAC1; Figure 7B) and a histone-binding protein RBBP7 (Zhang et al., 2016a) (Figure 7B). HDAC1 and RBBP7 are among the blue nodes representing chromosome organization proteins. Another protein identified in this sub-network was a DNA mismatch repair protein that is a well-known marker for ovarian cancer, MSH2 (Maresca et al., 2015; Stewart et al., 2017; Xiao et al., 2014; Zhao et al., 2018). Additionally, subunits of the tumor suppressor complex SWI/SNF, including ARID1A, SMARCC1, and SMARCA2, were among the proteins in the enriched chromosome organization network. The interaction of SWI/SNF with PARP and BRCA1 has been demonstrated previously by *in vitro* studies using yeast two-hybrid screening, affinity purification followed by western blotting, and co-immunoprecipitation purification followed by LC-MS/MS (Bochar et al., 2000; Harte et al., 2010; Hill et al., 2004). The identification of SWI/SNF complex components in the

enriched chromosome organization network in our study and the results from *in vitro* studies indicating the interaction of these proteins with PARP and BRCA1 support the well-known role of these proteins in the control of homologous recombination during DNA repair.

In summary, the analysis of the SWATH data was able to provide not only confirmation of some previously identified signatures from the iTRAQ DDA data in a larger untargeted network approach, but also important mechanistic insights into HRD-related pathways as discovered by iTRAQ labeling and LC-MS/MS using DDA (Zhang et al., 2016a).

DISCUSSION

HGSOC is among the cancer types that have been proteogenomically characterized by CPTAC. The large-scale proteome analytical workflow employed by CPTAC uses isobaric tagging, offline peptide fractionation, and LC-MS/MS. This workflow is considered a reference method for the comparison of tissue protein relative abundance across large sample cohorts. However, recently, DIA proteomic methods exemplified by SWATH-MS have been developed, which are simpler, cheaper, and consume less sample than the reference method but provide shallower proteome coverage.

Although HGSOC is the most common histological subtype of ovarian cancer, there is a considerable amount of tumor heterogeneity (Arend et al., 2018), thus underscoring the need for comprehensive characterization of the molecular subtypes of this lethal disease. Previous studies have shown that somatic mutations (Wiegand et al., 2010), genetic (Gates et al., 2010) and environmental risk factors (Goode et al., 2010), and the clinical response rates to platinum- or taxane-based therapy (Sugiyama et al., 2000) vary considerably among the HGSOC molecular subtypes. As such, elucidating the molecular characteristics of HGSOC could facilitate the development of more targeted and effective therapies (Leong et al., 2015; Vaughan et al., 2011).

DDA mass spectrometry workflows have been employed to comprehensively characterize HGSOC (Coscia et al., 2016; Li et al., 2017a; Xie et al., 2017; Zhang et al., 2016a) with varying numbers of molecular subtypes emerging from these analyses. Functional genomic studies have identified various numbers of distinct HGSOC molecular subtypes with clinical relevance and pathways that are responsible for growth control in epithelial ovarian cancer (Cancer Genome Atlas Research, 2011; Tan et al., 2013). The discordance among these subtypes results from the varied sample sizes and analytical criteria used to conduct these studies (Helland et al., 2011; Tothill et al., 2008; Verhaak et al., 2013). As we show in this study, the discordance also results from the low robustness of observed patterns as a function of sample size and the numbers of protein used for the analysis. Our current study provides orthogonal evidence of the proteomic signatures of the HGSOC Mesenchymal subtype and the fidelity of the HGSOC Mesenchymal subtype, which is more sensitive to sample number compared with the number of quantified proteins. This has implications for the design of future large-scale clinical proteomic studies of cancer types where molecular subtyping is a predominant goal.

In this study, we compared the results obtained by the reference large-scale DDA proteomic method versus SWATH-MS using aliquots of peptide samples generated for the CPTAC HGSOC study (Zhang et al., 2016a). The results indicate that iTRAQ DDA and SWATH-MS confidently identified differentially expressed proteins in enriched pathways associated with the Mesenchymal subtype of HGSOC tumors as evidenced by the (1) high degree of overlap of up-regulated extracellular matrix-related proteins (Figure 5A), (2) strongly positive median correlation ($\rho = 0.79$) among the proteins comprising this subtype from both proteomic workflows (Figure 5B), and (3) statistically significant stability of this subtype in the context of the number of tumors included in the clustering analysis (Figure S3B). The robustness of the Mesenchymal subtype with respect to molecular subtype cluster stability could be a signature of the decreased survival of patients whose tumors express this molecular signature compared with an Immunoreactive signature. A study conducted using a cohort of 174 HGSOC patients from the Mayo Clinic with long-term clinical follow-up observed statistically significantly worse survival of patients whose tumor samples expressed a Mesenchymal-like signature upon the analysis of a set of 1,850 genes (Konecny et al., 2014). A similar trend of worse survival for patients with Mesenchymal subtype tumors compared with those with Immunoreactive subtype tumors was observed upon the analysis of a separate cohort of 185 HGSOC patients (Bonome et al., 2008).

In addition to its association with worse survival and low rates of resection, the Mesenchymal subtype of HGSOC is associated with a significant contribution from the cancer-associated stroma (Tothill et al., 2008; Zhang et al., 2019). Cancer-associated stroma comprises extracellular matrix proteins, tissue remodeling proteases, and other proteins that comprise a scaffold for tissue organization and integrity (Davidson et al., 2014). These proteins are highly abundant, which could explain why they are readily detectable across distinct analytical workflows resulting in the consistent classification of the Mesenchymal subtype.

The notion of discrete HGSOC subtypes that are mutually exclusive is not universally accepted. Verhaak et al. suggested that an individual tumor could be represented by multiple signatures based on different levels of pathway activation (Verhaak et al., 2013). This concept has been supported by Konecny et al. who proposed a multi-dimensional approach to subtyping where molecular subtypes lie on a spectrum with partly overlapping causes (Konecny et al., 2014). Additional large-scale clinical proteomic studies of HGSOC tumors that are designed to address issues related to tumor heterogeneity would be beneficial in enhancing the resolution of molecular subtyping.

Our SWATH-MS analysis confirmed the proteomic signature of HRD established by iTRAQ DDA analysis wherein a sub-network of BRCA1- or BRCA2-related proteins displayed co-expression patterns differentiating HRD from non-HRD patients (Zhang et al., 2016a). Many of the proteins in these identified modules have roles in histone acetylation or deacetylation. Inhibitors of PARP and histone deacetylase inhibitors have emerged as novel classes of anti-cancer drugs to treat HR-related ovarian cancer associated with BRCA1/2 mutations (Bryant et al., 2005; Farmer et al., 2005; Yano et al., 2018; Yuan et al., 2017). Thus, the proteomic characterization of HGSOC tumor tissue biopsies from patients who receive PARP inhibitor treatment could be beneficial in further elucidating the molecular mechanisms that are implicated in HRD.

One of the strengths of the iTRAQ DDA workflow is the ability to achieve deep proteome coverage that exceeds that of the SWATH workflow by almost 3-fold. A total of 8,597 proteins were quantified by iTRAQ DDA compared with 2,914 proteins by SWATH-MS. These numbers represent the aggregate numbers of quantified proteins. However, after employing filtering strategies to restrict the data to only the proteins that were quantified across all 103 tumors for each workflow, these numbers decreased to 4,363 and 1,659, respectively. The group of 1,599 proteins that were quantified by both proteomic workflows was used to compare the performance of iTRAQ DDA and SWATH-MS.

In addition to their analytical performance, there are also considerable differences in the resource characteristics of iTRAQ DDA and SWATH-MS, including sample requirement, sample throughput, and cost. Hence, it is clear that iTRAQ DDA for comprehensive proteome profiling is a substantially more resource-intensive workflow compared with SWATH-MS.

Based on the concordance between the iTRAQ DDA and SWATH-MS results that we have shown in this study, SWATH-MS, which is considerably less resource intense than iTRAQ DDA, can be reliably deployed in the proteomic analysis of clinical specimens. The clinical utility of future large-scale translational proteomic studies, regardless of the employed analytical methodology, can be strengthened by the use of large sample cohorts that have undergone comprehensive pathology review.

Although our study was focused on the performance of SWATH-MS, other DIA methods are currently available. Our SWATH-MS analysis was conducted using a 5600+ TripleTOF mass spectrometer; however, a newer generation of this mass spectrometer platform exists with an increased linear dynamic range and enhanced detection system, which could result in improved instrument performance with respect to the number of quantifiable proteins per analytical run. A recent DIA study conducted by Bruderer et al. using a quadrupole ultra-high field Orbitrap mass spectrometer resulted in the identification of more than 6,000 proteins in human cell lines and more than 7,000 proteins in mouse tissues (Bruderer et al., 2017). Of note, the use of isobaric chemical labeling strategies, including 10- and 11-plex tandem mass tags, has greatly facilitated the multiplexing capabilities of large-scale clinical proteomic studies resulting in the reproducible quantification of >10,000 proteins (Krieger et al., 2019; Mertins et al., 2018).

In contrast to the original CPTAC HGSOC study, we did not assess protein phosphorylation in the current study. However, it should be noted that SWATH/DIA-MS has been shown to be compatible with the analysis of protein phosphorylation patterns and specific software tools supporting such analyses have been developed (Rosenberger et al., 2017b).

Limitations of the Study

As expected, the development of new analytical instruments and methods often enables an expanded breadth and/or depth of analytical measurement. After the performance of these instruments and methods has been optimized and validated, the short-comings of previously existing methods become evident. It is as yet unknown whether SWATH-MS and other DIA proteomic methods will have an increased prevalence in clinical proteomic analyses. However, our current study provides compelling orthogonal evidence that SWATH-MS elucidates some of the common biological signatures of the Mesenchymal subtype of HGSOc.

Novel biological insights beyond those that were initially gleaned from the iTRAQ DDA data (Zhang et al., 2016a) and confirmed by the SWATH-MS data could be obtained via additional bioinformatics and statistical analyses. However, the goal of this study, as reflected in the experimental design, was not to reveal novel biological insights, but rather to demonstrate that the SWATH-MS workflow presents novel opportunities to enhance the efficiency of clinical proteomic studies and to disseminate large-scale cancer proteomic studies beyond the realm of large research consortia, thus potentially accelerating the use of this powerful proteomic approach to cancer biology.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

DATA AND CODE AVAILABILITY

All the raw data from SWATH-MS measurements, along with the input spectral library and OpenSWATH results, can be freely downloaded from the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) with the dataset identifier: PXD010437 via the PRIDE partner (Vizcaino et al., 2016).

The iTRAQ DDA data are available on the NCI CPTAC Data Portal (<https://cptac-data-portal.georgetown.edu/cptac/s/S026>); mass spectrometry site: Johns Hopkins University. Other data are available from the corresponding authors upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101079>.

ACKNOWLEDGMENTS

We acknowledge contributions from Vlad Petyuk who provided scripts and methods from the previously published CPTAC study. We also thank Ludovic Gillet, Isabell Bludau, Ben Collins, and the computational proteomics team at ETH Zürich for their important contributions in discussions. We are grateful to Sciex (Christie Hunter, Mark Cafazzo, and George Manning) for the 5600+ TripleTOF mass spectrometer and access to open-source SWATH data processing software. Funding for the study was provided by the United States National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (NCI CPTAC, U24CA160036 and U24CA210985 to H.Z. and D.W.C.) and the Early Detection Research Network (NCI EDNRN, U01CA152813 to H.Z. and R. A.), the European Research Council, European Union (AdG-233226 Proteomics v.3.0 and AdG-670821 Proteomics 4D to R.A.), and the Swiss National Science Foundation, European Union (grant 31003A_166435 to R.A.).

AUTHOR CONTRIBUTIONS

S.N.T. performed the experiments with sample preparation and SWATH-MS analysis; B.F. conducted computational analyses for SWATH data and bioinformatics analysis; S.N.T. and M.S. performed and assisted in computational analyses with iTRAQ DDA data. R.A., H.Z., and D.W.C. conceived the study, designed the experiments, and supervised the participants. All authors participated in the preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 24, 2019

Revised: June 19, 2019

Accepted: April 14, 2020

Published: June 26, 2020

REFERENCES

- Anagnostopoulos, A.K., Stravopodis, D.J., and Tsangaris, G.T. (2017). Yield of 6,000 proteins by 1D nLC-MS/MS without pre-fractionation. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 1047, 92–96.
- Arend, R.C., Londono, A.I., Montgomery, A.M., Smith, H.J., Dobbin, Z.C., Katre, A.A., Martinez, A., Yang, E.S., Alvarez, R.D., Huh, W.K., et al. (2018). Molecular response to neoadjuvant chemotherapy in high-grade serous ovarian carcinoma. *Mol. Cancer Res.* 16, 813–824.
- Basak, T., Bhat, A., Malakar, D., Pillai, M., and Sengupta, S. (2015). In-depth comparative proteomic analysis of yeast proteome using iTRAQ and SWATH based MS. *Mol. Biosyst.* 11, 2135–2143.
- Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011). The quantitative proteome of a human cell line. *Mol. Syst. Biol.* 7, 549.
- Bochar, D.A., Wang, L., Beniya, H., Kinev, A., Xue, Y., Lane, W.S., Wang, W., Kashanchi, F., and Shiekhhattar, R. (2000). BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* 102, 257–265.
- Bonome, T., Levine, D.A., Shih, J., Randonovich, M., Pise-Masison, C.A., Bogomolnij, F., Ozburn, L., Brady, J., Barrett, J.C., Boyd, J., et al. (2008). A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. *Cancer Res.* 68, 5478–5486.
- Bourassa, S., Fournier, F., Nehme, B., Kelly, I., Tremblay, A., Lemelin, V., Lamarche, B., Couture, P., and Droit, A. (2015). Evaluation of iTRAQ and SWATH-MS for the quantification of proteins associated with insulin resistance in human duodenal biopsy samples. *PLoS One* 10, e0125934.
- Bruderer, R., Bernhardt, O.M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., and Reiter, L. (2017). Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol. Cell. Proteomics* 16, 2296–2309.
- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434, 913–917.
- Cancer Genome Atlas Research, N. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615.
- Chen, G.M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D.M.A., Parmigiani, G., Birrer, M., Haibe-Kains, B., and Waldron, L. (2018). Consensus on molecular subtypes of high-grade serous ovarian carcinoma. *Clin. Cancer Res.* 24, 5037–5047.
- Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.M., Chang, H.Y., et al. (2020). Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* 180, 207.
- Collins, B.C., Gillet, L.C., Rosenberger, G., Rost, H.L., Vichalkovski, A., Gstaiger, M., and Aebersold, R. (2013). Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* 10, 1246–1253.
- Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., Chan, D.W., Gibson, B.W., Gingras, A.C., Held, J.M., et al. (2017). Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* 8, 291.
- Coscia, F., Watters, K.M., Curtis, M., Eckert, M.A., Chiang, C.Y., Tyanova, S., Montag, A., Lastra, R.R., Lengyel, E., and Mann, M. (2016). Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 7, 12645.
- Davidson, B., Trope, C.G., and Reich, R. (2014). The role of the tumor stroma in ovarian cancer. *Front. Oncol.* 4, 104.
- Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al. (2020). Proteogenomic characterization of endometrial carcinoma. *Cell* 180, 729–748.e26.
- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434, 917–921.
- Fraley, C., and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* 97, 611–631.
- Gates, M.A., Rosner, B.A., Hecht, J.L., and Tworoger, S.S. (2010). Risk factors for epithelial ovarian cancer by histologic subtype. *Am. J. Epidemiol.* 171, 45–53.
- Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11, O111.016717–O111.016717-17.
- Goode, E.L., Chenevix-Trench, G., Song, H., Ramus, S.J., Notaridou, M., Lawrenson, K., Widschwendter, M., Vierkant, R.A., Larson, M.C., Kjaer, S.K., et al. (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat. Genet.* 42, 874–879.
- Harte, M.T., O'Brien, G.J., Ryan, N.M., Gorski, J.J., Savage, K.I., Crawford, N.T., Mullan, P.B., and Harkin, D.P. (2010). BRD7, a subunit of SWI/SNF complexes, binds directly to BRCA1 and regulates BRCA1-dependent transcription. *Cancer Res.* 70, 2538–2547.
- Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014). The one hour yeast proteome. *Mol. Cell. Proteomics* 13, 339–347.
- Helland, A., Anglesio, M.S., George, J., Cowin, P.A., Johnstone, C.N., House, C.M., Sheppard, K.E., Etemadmoghadam, D., Melnyk, N., Rustgi, A.K., et al. (2011). Deregulation of MYCN, LIN28B and LET7 in a molecular subtype of aggressive high-grade serous ovarian cancers. *PLoS One* 6, e18064.
- Hill, D.A., de la Serna, I.L., Veal, T.M., and Imbalzano, A.N. (2004). BRCA1 interacts with dominant negative SWI/SNF enzymes without affecting homologous recombination or radiation-induced gene activation of p21 or Mdm2. *J. Cell. Biochem.* 91, 987–998.
- Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108–1115.
- Hu, C., Wen, J., Gong, L., Chen, X., Wang, J., Hu, F., Zhou, Q., Liang, J., Wei, L., Shen, Y., et al. (2017). Thrombospondin-1 promotes cell migration, invasion and lung metastasis of osteosarcoma through FAK dependent pathway. *Oncotarget* 8, 75881–75892.
- Incardona, F., Calvo, F., Fauvel-Lafeve, F., Legrand, Y., and Legrand, C. (1993). Involvement of thrombospondin in the adherence of human breast-adenocarcinoma cells: a possible role in the metastatic process. *Int. J. Cancer* 55, 471–477.
- Karpievitch, Y.V., Dabney, A.R., and Smith, R.D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* 13 (Suppl 16), S5.
- Kenny, H.A., Chiang, C.Y., White, E.A., Schryver, E.M., Habis, M., Romero, I.L., Ladanyi, A., Penicka, C.V., George, J., Matlin, K., et al. (2014). Mesothelial cells promote early ovarian cancer metastasis through fibronectin secretion. *J. Clin. Invest.* 124, 4614–4628.
- Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581.

- Konecny, G.E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K.R., Dering, J., Ginther, C., Chen, H.W., Dowdy, S., Cliby, W., et al. (2014). Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J. Natl. Cancer Inst.* *106*, dju249.
- Krieger, J.R., Wybenga-Groot, L.E., Tong, J., Bache, N., Tsao, M.S., and Moran, M.F. (2019). Evosep One enables robust deep proteome coverage using tandem mass tags while significantly reducing instrument time. *J. Proteome Res.* *18*, 2346–2353.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* *11*, 319–324.
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.
- Leong, H.S., Galletta, L., Etemadmoghadam, D., George, J., Australian Ovarian Cancer Study, Kobel, M., Ramus, S.J., and Bowtell, D. (2015). Efficient molecular subtype classification of high-grade serous ovarian cancer. *J. Pathol.* *236*, 272–277.
- Levine, E., and Domany, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Comput.* *13*, 2573–2593.
- Lheureux, S., Lai, Z., Dougherty, B.A., Runswick, S., Hodgson, D.R., Timms, K.M., Lanchbury, J.S., Kaye, S., Gourley, C., Bowtell, D., et al. (2017). Long-term responders on olaparib maintenance in high-grade serous ovarian cancer: clinical and molecular characterization. *Clin. Cancer Res.* *23*, 4086–4094.
- Li, Y., Zhong, C.Q., Xu, X., Cai, S., Wu, X., Zhang, Y., Chen, J., Shi, J., Lin, S., and Han, J. (2015). Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files. *Nat. Methods* *12*, 1105–1106.
- Li, Q.K., Shah, P., Tian, Y., Hu, Y., Roden, R.B.S., Zhang, H., and Chan, D.W. (2017a). An integrated proteomic and glycoproteomic approach uncovers differences in glycosylation occupancy from benign and malignant epithelial ovarian tumors. *Clin. Proteomics* *14*, 16.
- Li, S., Cao, Q., Xiao, W., Guo, Y., Yang, Y., Duan, X., and Shui, W. (2017b). Optimization of acquisition and data-processing parameters for improved proteomic quantification by sequential window acquisition of all theoretical fragment ion mass spectrometry. *J. Proteome Res.* *16*, 738–747.
- Liu, Y., Huttenhain, R., Collins, B., and Aebersold, R. (2013). Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert Rev. Mol. Diagn.* *13*, 811–825.
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* *14*, e8126.
- Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* *70*, 437–473.
- Maresca, L., Spugnese, L., Lodovichi, S., Cozzani, C., Naccarato, A.G., Tancredi, M., Collavoli, A., Falaschi, E., Rossetti, E., Aretini, P., et al. (2015). MSH2 role in BRCA1-driven tumorigenesis: a preliminary study in yeast and in human tumors from BRCA1-VUS carriers. *Eur. J. Med. Genet.* *58*, 531–539.
- Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* *534*, 55–62.
- Mertins, P., Tang, L.C., Krug, K., Clark, D.J., Gritsenko, M.A., Chen, L., Clauser, K.R., Clauss, T.R., Shah, P., Gillette, M.A., et al. (2018). Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat. Protoc.* *13*, 1632–1661.
- Mitra, A.K., Sawada, K., Tiwari, P., Mui, K., Gwin, K., and Lengyel, E. (2011). Ligand-independent activation of c-Met by fibronectin and alpha(5) beta(1)-integrin regulates ovarian cancer invasion and metastasis. *Oncogene* *30*, 1566–1576.
- Monti, S., Tamayo, P., Mesirov, J., and Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* *52*, 91–118.
- Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* *7*, 548.
- Navarro, P., Kuharev, J., Gillet, L.C., Bernhardt, O.M., MacLean, B., Rost, H.L., Tate, S.A., Tsou, C.C., Reiter, L., Distler, U., et al. (2016). A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* *34*, 1130–1136.
- Ow, S.Y., Salim, M., Noirel, J., Evans, C., and Wright, P.C. (2011). Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *Proteomics* *11*, 2341–2346.
- Rardin, M.J., Schilling, B., Cheng, L.Y., MacLean, B.X., Sorensen, D.J., Sahu, A.K., MacCoss, M.J., Vitek, O., and Gibson, B.W. (2015). MS1 peptide ion intensity chromatograms in MS2 (SWATH) data independent acquisitions. Improving post acquisition analysis of proteomic experiments. *Mol. Cell. Proteomics* *14*, 2405–2419.
- Revesz, A., Rokob, T.A., Jeanne Dit Fouque, D., Turiak, L., Memboeuf, A., Karoly, V., and Drahos, L. (2018). Selection of collision energies in proteomics mass spectrometry experiments for best peptide identification: study of Mascot score energy dependence reveals double optimum. *J. Proteome Res.* *17*, 1898–1906.
- Ricciardelli, C., Lokman, N.A., Ween, M.P., and Oehler, M.K. (2016). WOMEN IN CANCER THEMATIC REVIEW: ovarian cancer-peritoneal cell interactions promote extracellular matrix processing. *Endocr. Relat. Cancer* *23*, T155–T168.
- Richards, A.L., Hebert, A.S., Ulbrich, A., Bailey, D.J., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2015). One-hour proteome analysis in yeast. *Nat. Protoc.* *10*, 701–714.
- Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C.L., Liu, Y., MacCoss, M.J., MacLean, B.X., Nesvizhskii, A.I., Pedrioli, P.G.A., et al. (2017a). Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat. Methods* *14*, 921–927.
- Rosenberger, G., Liu, Y., Rost, H.L., Ludwig, C., Buil, A., Bensimon, A., Soste, M., Spector, T.D., Dermitzakis, E.T., Collins, B.C., et al. (2017b). Inference and quantification of peptidofoms in large sample cohorts by SWATH-MS. *Nat. Biotechnol.* *35*, 781–788.
- Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., Woltski, W., Collins, B.C., Malmstrom, J., Malmstrom, L., et al. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* *32*, 219–223.
- Rost, H.L., Liu, Y., D’Agostino, G., Zanella, M., Navarro, P., Rosenberger, G., Collins, B.C., Gillet, L., Testa, G., Malmstrom, L., et al. (2016). TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat. Methods* *13*, 777–783.
- Savitski, M.M., Mathieson, T., Zinn, N., Sweetman, G., Doce, C., Becher, I., Pachel, F., Kuster, B., and Bantscheff, M. (2013). Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *J. Proteome Res.* *12*, 3586–3598.
- Stewart, C.J., Bowtell, D.D., Doherty, D.A., and Leung, Y.C. (2017). Long-term survival of patients with mismatch repair protein-deficient, high-stage ovarian clear cell carcinoma. *Histopathology* *70*, 309–313.
- Sugiyama, T., Kamura, T., Kigawa, J., Terakawa, N., Kikuchi, Y., Kita, T., Suzuki, M., Sato, I., and Taguchi, K. (2000). Clinical characteristics of clear cell carcinoma of the ovary: a distinct histologic type with poor prognosis and resistance to platinum-based chemotherapy. *Cancer* *88*, 2584–2589.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* *43*, D447–D452.
- Tabb, D.L., Wang, X., Carr, S.A., Clauser, K.R., Mertins, P., Chambers, M.C., Holman, J.D., Wang, J., Zhang, B., Zimmerman, L.J., et al. (2016). Reproducibility of differential proteomic technologies in CPTAC fractionated Xenografts. *J. Proteome Res.* *15*, 691–706.
- Tan, T.Z., Miow, Q.H., Huang, R.Y., Wong, M.K., Ye, J., Lau, J.A., Wu, M.C., Bin Abdul Hadi, L.H., Soong, R., Choolani, M., et al. (2013). Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol. Med.* *5*, 1051–1066.
- Teo, G., Kim, S., Tsou, C.C., Collins, B., Gingras, A.C., Nesvizhskii, A.I., and Choi, H. (2015). mapDIA: preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J. Proteomics* *129*, 108–120.

- Tohill, R.W., Tinker, A.V., George, J., Brown, R., Fox, S.B., Lade, S., Johnson, D.S., Trivett, M.K., Etemadmoghadam, D., Locandro, B., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin. Cancer Res.* 14, 5198–5208.
- Tsou, C.C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.C., and Nesvizhskii, A.I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* 12, 258–264.
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740.
- Vaughan, S., Coward, J.I., Bast, R.C., Jr., Berchuck, A., Berek, J.S., Brenton, J.D., Coukos, G., Crum, C.C., Drapkin, R., Etemadmoghadam, D., et al. (2011). Rethinking ovarian cancer: recommendations for improving outcomes. *Nat. Rev. Cancer* 11, 719–725.
- Verhaak, R.G., Tamayo, P., Yang, J.Y., Hubbard, D., Zhang, H., Creighton, C.J., Fereday, S., Lawrence, M., Carter, S.L., Mermel, C.H., et al. (2013). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Invest.* 123, 517–525.
- Vidova, V., and Spacil, Z. (2017). A review on mass spectrometry-based quantitative proteomics: targeted and data independent acquisition. *Anal. Chim. Acta* 964, 7–23.
- Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456.
- Wang, P., Tang, H., Fitzgibbon, M.P., McIntosh, M., Coram, M., Zhang, H., Yi, E., and Aebersold, R. (2007). A statistical alignment method for chromatographic alignment of LC-MS data. *Biostatistics* 8, 357–367.
- Wiegand, K.C., Shah, S.P., Al-Agha, O.M., Zhao, Y., Tse, K., Zeng, T., Senz, J., McConechy, M.K., Anglesio, M.S., Kaloger, S.E., et al. (2010). ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* 363, 1532–1543.
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587.
- Xiao, X., Melton, D.W., and Gourley, C. (2014). Mismatch repair deficiency in ovarian cancer – molecular characteristics and clinical implications. *Gynecol. Oncol.* 132, 506–512.
- Xie, H., Wang, W., Sun, F., Deng, K., Lu, X., Liu, H., Zhao, W., Zhang, Y., Zhou, X., Li, K., et al. (2017). Proteomics analysis to reveal biological pathways and predictive proteins in the survival of high-grade serous ovarian cancer. *Sci. Rep.* 7, 9896.
- Yano, M., Yasuda, M., Sakaki, M., Nagata, K., Fujino, T., Arai, E., Hasebe, T., Miyazawa, M., Miyazawa, M., Ogane, N., et al. (2018). Association of histone deacetylase expression with histology and prognosis of ovarian cancer. *Oncol. Lett.* 15, 3524–3531.
- Yuan, Z., Chen, S., Sun, Q., Wang, N., Li, D., Miao, S., Gao, C., Chen, Y., Tan, C., and Jiang, Y. (2017). Olaparib hydroxamic acid derivatives as dual PARP and HDAC inhibitors for cancer therapy. *Bioorg. Med. Chem.* 25, 4100–4109.
- Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al. (2014a). Proteogenomic characterization of human colon and rectal cancer. *Nature* 513, 382–387.
- Zhang, F., Lin, H., Gu, A., Li, J., Liu, L., Yu, T., Cui, Y., Deng, W., Yan, M., Li, J., et al. (2014b). SWATH- and iTRAQ-based quantitative proteomic analyses reveal an overexpression and biological relevance of CD109 in advanced NSCLC. *J. Proteomics* 102, 125–136.
- Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al. (2016a). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* 166, 755–765.
- Zhang, Z., Huang, K., Gu, C., Zhao, L., Wang, N., Wang, X., Zhao, D., Zhang, C., Lu, Y., and Meng, Y. (2016b). Molecular subtyping of serous ovarian cancer based on multi-omics data. *Sci. Rep.* 6, 26001.
- Zhang, Q., Wang, C., and Cliby, W.A. (2019). Cancer-associated stroma significantly contributes to the mesenchymal subtype signature of serous ovarian cancer. *Gynecol. Oncol.* 152, 368–374.
- Zhao, C., Li, S., Zhao, M., Zhu, H., and Zhu, X. (2018). Prognostic values of DNA mismatch repair genes in ovarian cancer patients treated with platinum-based chemotherapy. *Arch. Gynecol. Obstet.* 297, 153–159.
- Zhou, J.Y., Chen, L., Zhang, B., Tian, Y., Liu, T., Thomas, S.N., Chen, L., Schnaubelt, M., Boja, E., Hiltke, T., et al. (2017). Quality assessments of long-term quantitative proteomic analysis of breast cancer Xenograft tissues. *J. Proteome Res.* 16, 4523–4530.

iScience, Volume 23

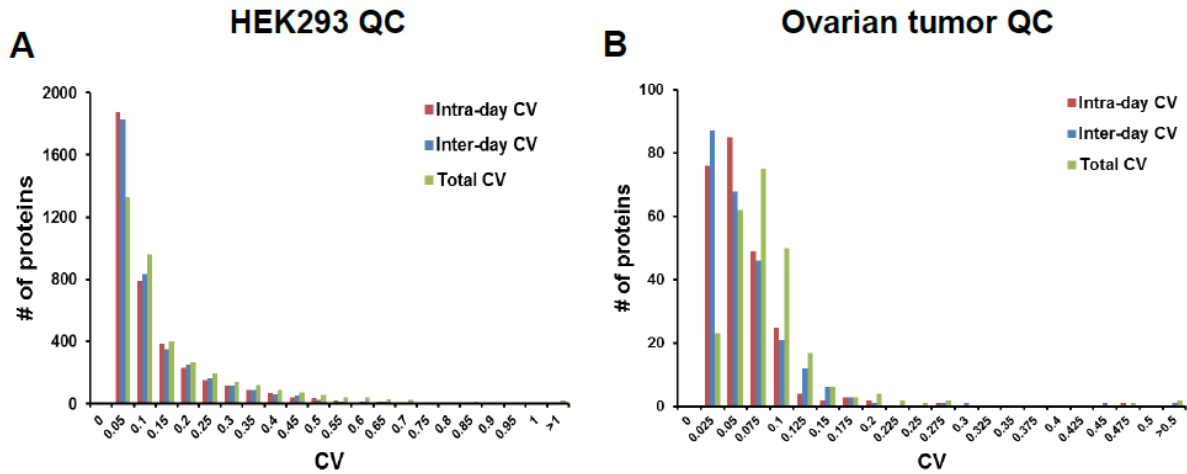
Supplemental Information

Orthogonal Proteomic Platforms and Their Implications for the Stable Classification of High-Grade Serous Ovarian Cancer Subtypes

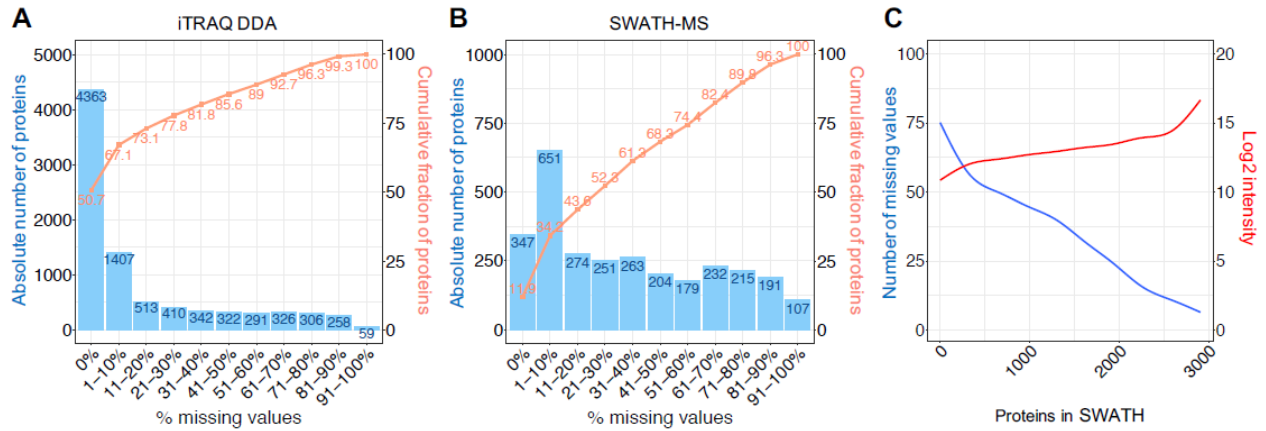
Stefani N. Thomas, Betty Friedrich, Michael Schnaubelt, Daniel W. Chan, Hui Zhang, and Ruedi Aebersold

Supplemental Information

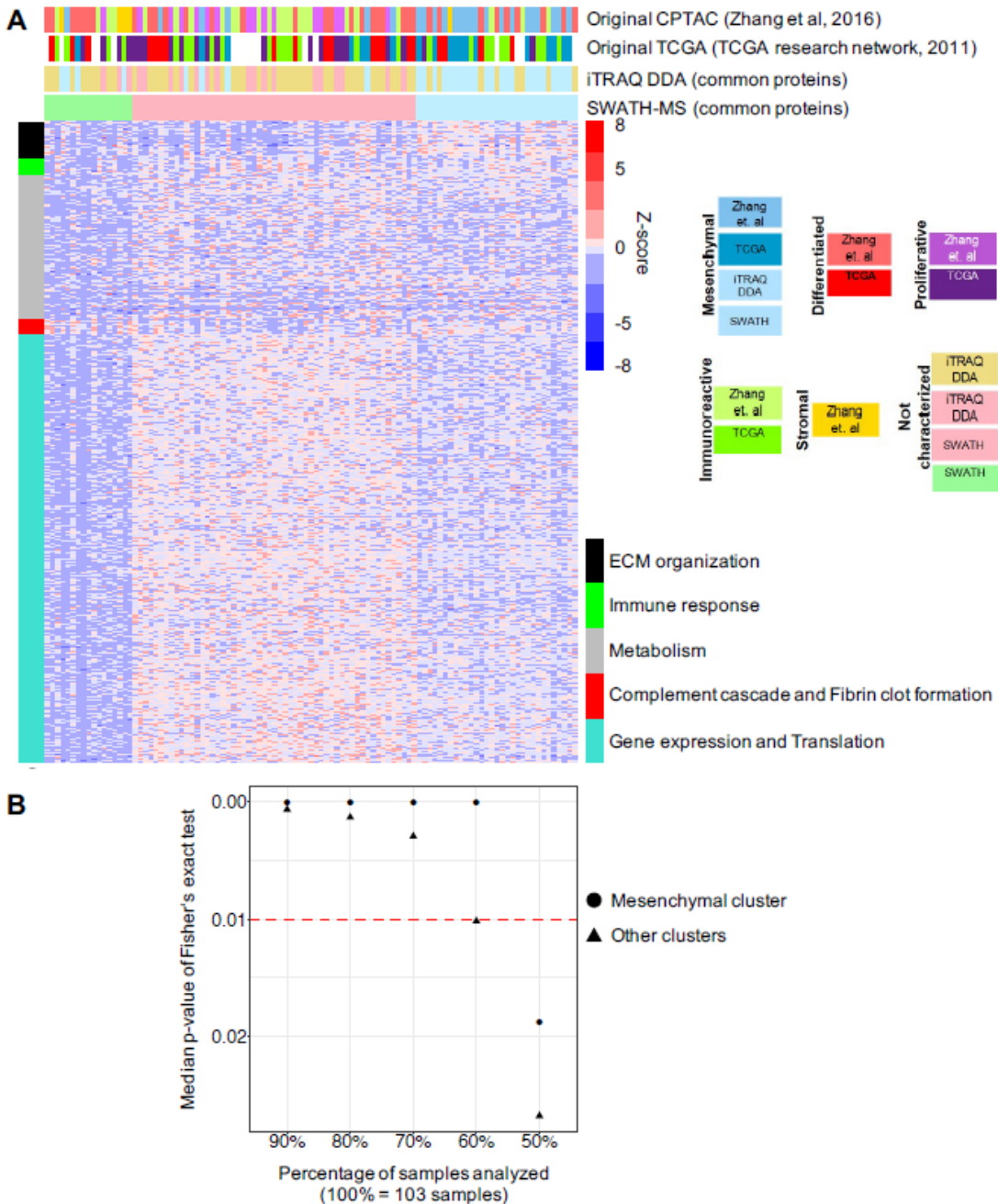
Supplementary Figures



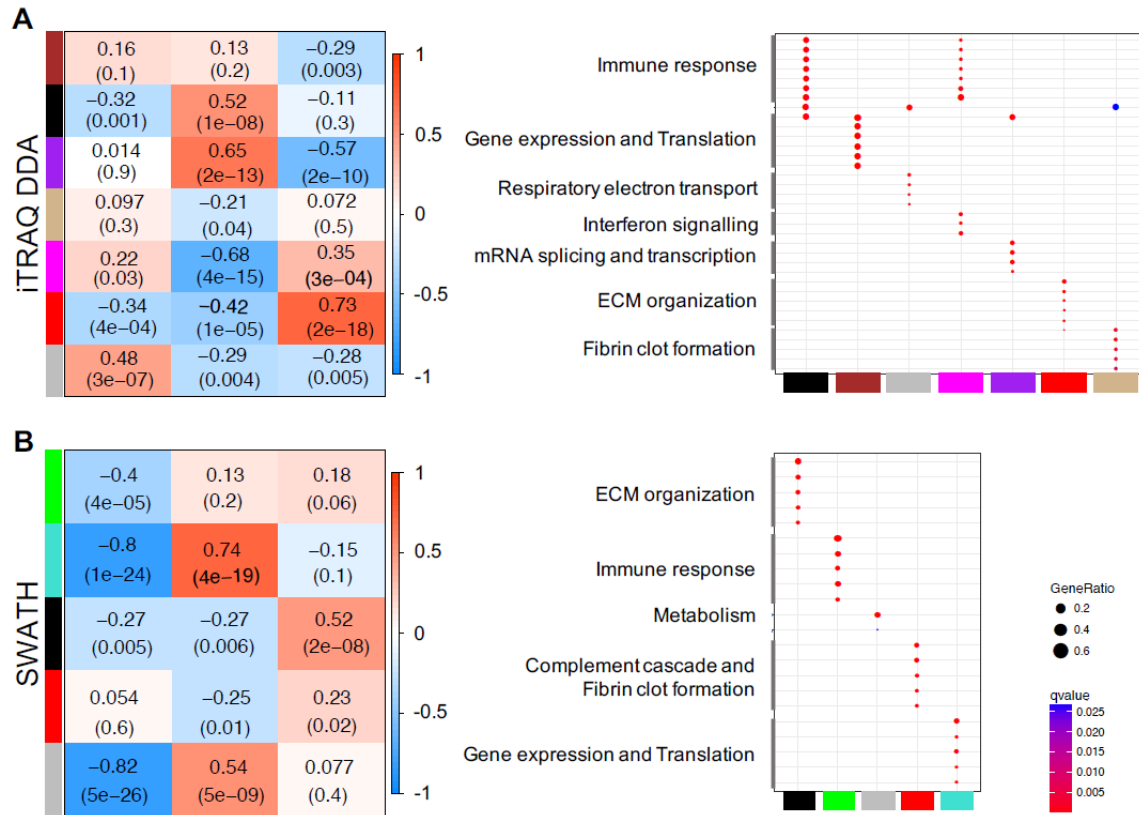
Supplementary Figure S1, Related to Figure 2. Distribution of quantified proteins from QC strategies used to assess technical variability. a) Distribution of quantified protein CVs from HEK293 lysate analyzed via SWATH (n=3 technical replicates per day for 3 days; total = 9 technical replicates) prior to the analysis of the ovarian tumors. Median total CV=15%; 3,855 quantified proteins. **b)** Distribution of quantified protein CVs from a control ovarian tumor analyzed in duplicate via DDA prior to and mid-way through the ovarian tumor SWATH analyses (n=4 technical replicates). Median total CV=7%; 781 quantified proteins.



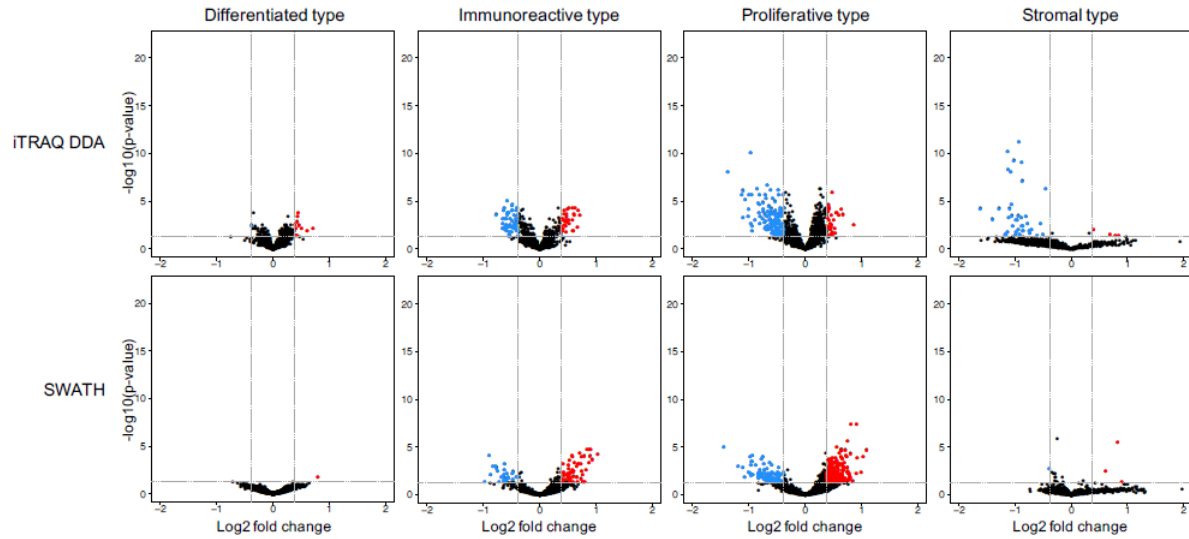
Supplementary Figure S2, Related to Figure 2. Completeness of iTRAQ DDA and SWATH-MS proteomic data. a) Analysis of missing values in the iTRAQ DDA dataset. **b)** Analysis of missing values in the SWATH-MS dataset. **c)** Relationship between signal intensity and number of missing values for the proteins quantified by SWATH-MS.



Supplementary Figure S3, Related to Figure 3. Protein (iTRAQ DDA, SWATH-MS) and mRNA-based high-grade serous ovarian cancer subtype classification. a) Tumor subtype classification based on the associated driving protein modules. Numbers indicate z-scores. Subtype classification: Blue – Mesenchymal; Red – Differentiated; Purple – Proliferative; Green – Immunoreactive; Yellow – Stromal. **b)** Stability of the Mesenchymal compared to the non-Mesenchymal subtypes assessed by bootstrapping and Fisher's exact test.



Supplementary Figure S4, Related to Figure 3. Analysis of tumor subtype analysis according to data type. Correlation of weighted gene co-expression network analysis (WGCNA)-derived protein modules with proteomic subtypes derived from the iTRAQ DDA data (**a**) and the SWATH-MS data (**b**). The numbers in each box indicate the Pearson correlation coefficient (-1 to 1), and the numbers in parentheses indicate the p-value corresponding to the significance of each association. The enrichment of Reactome pathways in the WGCNA-derived modules in the iTRAQ DDA and SWATH-MS data is also shown. The significance of association is denoted by colored dots indicating the q-value, and the number of proteins mapped to a given ontology as a proportion of the total number of proteins in the module (GeneRatio) is indicated by the size of the dot. The WGCNA-derived protein modules for the iTRAQ DDA data are as follows: magenta – gene expression; black – immune response; purple – transcription; tan – erythrocyte and platelet; magenta – cytokine signaling; red – ECM interaction; gray – metabolism. The WGCNA-derived protein modules for the SWATH data are as follows: green – immune response; turquoise – gene expression; black – ECM interaction; red – complement cascade; gray – metabolism.



Supplementary Figure S5, Related to Figure 3. Differential relative protein abundance analysis of proteins in the Differentiated, Immunoreactive, Proliferative and Stromal subtype tumors from the iTRAQ DDA and SWATH-MS datasets. Up-regulated proteins are indicated in red and the down-regulated proteins are indicated in blue. The log₂ fold-change cut-off was 1.3 and the significance cut-off was set at $p < 0.05$.

Transparent Methods

STAR Methods

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Human ovarian tumor tissue	National Cancer Institute Clinical Proteomic Tumor Analysis Consortium	https://proteomics.cancer.gov/programs/cptac
Chemicals, Peptides, and Recombinant Proteins		
Urea	Sigma-Aldrich	Cat# U0631
HPLC-grade water	J.T. Baker	Cat# 4218-03
Triethylammonium bicarbonate	Sigma-Aldrich	Cat# T7408
Ammonium bicarbonate	Sigma-Aldrich	Cat# A6141
Tris (2-carboxyethyl) phosphine (TCEP)	Thermo Fisher Scientific	Cat# 20490
Iodoacetamide	Sigma-Aldrich	Cat# A3221
Sequencing-grade modified trypsin	Promega	Cat# V511X
Formic acid	Sigma-Aldrich	Cat# 56302
Trifluoroacetic acid	Sigma-Aldrich	Cat# 91707
Acetonitrile	J.T. Baker	Cat# 9829-03
iTRAQ reagent kit – 4plex	Sciex	Cat# 4352135
Critical Commercial Assays		
BCA bicinchoninic protein assay kit	Pierce	Cat# 23325
Deposited Data		
iTRAQ DDA data	NCI CPTAC Data Portal	https://cptac-data-portal.georgetown.edu/cptac/s/S026 (Mass spectrometry site: Johns Hopkins University)
SWATH data, input spectral library, OpenSWATH results	ProteomeXchange Consortium	http://proteomecentral.proteomexchange.org Dataset identifier: PXD010437
Software and Algorithms		
Programming code for bioinformatics pipeline	Github repository	bfriedrichgrube/OC_CPTAC_iTRAQ_SWATH
msconvert	ProteoWizard v3.0.8851	http://proteowizard.sourceforge.net/download.html
X!Tandem 2013.06.25.1	Duncan et al., 2005	https://www.thegpm.org/tandem/

OMSSA 2.1.9	Geer et al., 2004	https://omictools.com/omssa-tool
Comet 2015.02 rev. 3	Eng et al., 2013	http://comet-ms.sourceforge.net/
Trans-Proteomic Pipeline v. 4.7	Keller et al., 2005; Lam et al., 2007	http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
OpenSWATH	Rost et al., 2014	http://www.openswath.org/en/latest/
SpectraST	Keller et al., 2005; Lam et al., 2007	http://tools.proteomecenter.org/wiki/index.php?title=SpectraST
TRIC	Rost et al., 2016	http://proteomics.ethz.ch/tric/
mapDIA v. 3.0.2	Teo et al., 2015	http://mapdia.sourceforge.net/Main.html
Other		
SepPak tC18 Vac cartridges	Waters	Cat# WAT054925
Polysulfoethyl A strong cation exchange columns	Glygen	Cat# TT3SSA

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hui Zhang (huizhang@jhu.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tumor samples The tumor specimens were obtained through The Cancer Genome Atlas (TCGA) Biospecimen Core Resource, and they were previously genomically characterized (Cancer Genome Atlas Research, 2011). As previously described (Zhang et al., 2016), the biospecimens were obtained from newly diagnosed patients with ovarian serous adenocarcinoma who were undergoing surgical resection and did not receive prior treatment, including chemotherapy or radiotherapy, for their disease. Per the TCGA study, all specimens were obtained from patients with appropriate consent from the relevant institutional review board. Frozen tissue specimens were extracted and used for subsequent proteomic analysis.

METHOD DETAILS

Protein extraction and in-solution digestion Approximately 50mg of each tumor tissue specimen was sonicated in 1.5mL of 8M urea, 0.8M NH_4HCO_3 , pH 8.0. Protein concentration was determined using a BCA assay (Thermo Fisher Scientific). Protein disulfide bonds were reduced with 10mM tris (2-carboxyethyl) phosphine (TCEP) for 1h at 37°C, followed by alkylation with 12mM iodoacetamide for 1h at RT in the dark. After dilution 1:4 with deionized water, proteins were digested with sequencing-grade modified trypsin (Promega, Madison, WI) (1:50 enzyme:protein, weight/weight) for 12h at 37°C. This was followed by the addition of an aliquot of the same amount of trypsin and incubation overnight at 37°C. The digested samples were acidified with 10% trifluoroacetic acid (TFA) to pH<3, de-salted using strong cation exchange and C18 solid-phase extraction (SPE) columns (Waters, Milford, MA) and dried using a Speed-Vac.

Shotgun proteomics using iTRAQ data-dependent acquisition (DDA) Relative quantitative proteomic analysis was conducted using 4-plex isobaric tags for relative and absolute quantitation (iTRAQ) reagents (Sciex) as previously described (Zhang et al., 2016). Peptides (500µg) were dissolved in 150µL of 0.5M triethylammonium bicarbonate, pH 8.5 and combined with 5U of 4-plex iTRAQ reagent dissolved in ethanol followed by a 2h incubation at RT, quenching with 10% TFA, and de-salting using C18 SPE columns. iTRAQ channel 114 was used to label the reference sample which was created by pooling an aliquot from each individual tumor sample. Offline basic reversed phase liquid chromatography (bRPLC) was conducted using a Zorbax extend 4.6 x 100mm C-18 column (Agilent) and an Agilent 1220 Infinity HPLC system to reduce the sample complexity prior to mass spectrometry analysis. A total of 96 fractions were collected and concatenated into 24 fractions. The fractions were dried in a Speed-Vac and stored at -80°C until analysis by LC-MS/MS using an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific). Peptides were loaded onto a 2cm guard column (Thermo Scientific) and separated on a 75µm x 15cm Acclaim PepMap100 column (Thermo Scientific) using a Dionex Ultimate 3000 RSLC nano chromatography system (Thermo Scientific). The LC gradient profile was 2-22% B for 70min, 22-29% B for 8min, 29-95% B for 4min, and 95% B for 8min, where mobile phase B was 0.1% formic acid in acetonitrile, and mobile phase

A was 0.1% formic acid in water. Orbitrap full MS spectra were collected from 400-1800 m/z at a resolution of 30,000 followed by data-dependent MS/MS (7,500 resolution) of the ten most abundant ions. Charge-state screening was enabled to prevent the acquisition of MS/MS spectra for ions with unassigned or single charges. Dynamic exclusion (40s duration) was enabled to minimize the repeated acquisition of previously acquired MS/MS spectra.

Proteomic analysis using SWATH mass spectrometry SWATH-MS measurements were conducted at the Johns Hopkins University using a Sciex 5600+ TripleTOF mass spectrometer interfaced with an Eksigent ekspert nanoLC 425 cHiPLC system. Peptides (1 μ g) were loaded onto a 6mm x 200 μ m ChromXP C18-CL 3 μ m, 120Å trap column followed by separation on a 75 μ m x 15cm ChromXP C18-CL 3 μ m, 120Å Nano cHiPLC column using a 120min method (90min gradient from 3-35% B – 0.1% formic acid in acetonitrile) at a flow rate of 300 nL/min. To create the spectral library for the SWATH-MS data analysis, each sample was run individually (1 μ g peptides per injection) using a data-dependent data acquisition (DDA) method wherein MS spectra were acquired across a range of 400-1800 m/z followed by the acquisition of MS/MS spectra of the top 30 most intense precursor ions with a charge state of z=2-5. The spectral library was also comprised of mass spectrometry data acquired from a fractionated (48 fractions) pool of peptides from all 103 tumors. Each of the 48 fractions from the pooled sample was analyzed using the same DDA method described above. SWATH data of the individual tumors were acquired using a variable window strategy wherein the sizes of the precursor ion selection windows were inversely related to m/z density. The average window width for precursor ion selection was 12 m/z with a range of 6-25 m/z. The collision energy was optimized for each window according to the calculation for a charge 2+ ion centered in the window with a spread of 5 eV. The MS accumulation time was 250ms and the MS/MS accumulation time for fragment ions accumulated in high sensitivity mode was 50ms, resulting in a total duty cycle of approximately 3.5s. To assess the analytical precision of the proteomics measurements, peptides from trypsin-digested HEK293 cell proteins were analyzed via SWATH-MS in triplicate immediately prior to the SWATH-MS analysis of the 103 individual tumor samples. Additionally, an ovarian cancer tumor sample separate

from the TCGA collection was used as a QC sample. This specimen was analyzed using the DDA method described above in duplicate prior to the SWATH-MS analysis of the 103 individual tumor samples and 10 days later in duplicate mid-way through the analysis of the individual tumor samples. Instrument performance was assessed daily by monitoring the peak area of 5 peptides from a trypsin-digested *E. coli* β -Galactosidase LC-MS standard that was injected every day a sample was run on the mass spectrometer.

iTRAQ DDA data processing iTRAQ DDA data from the Johns Hopkins CPTAC center was analyzed by the Common Data Analysis Pipeline (CDAP) and downloaded from the NCI CPTAC Data Portal (<https://cptac-data-portal.georgetown.edu/cptac/s/S026>); mass spectrometry site: Johns Hopkins University (Edwards et al., 2015; Ellis et al., 2013). Values obtained from proteotypic peptides were chosen for further analysis. The protein matrix contained 8597 proteins. The protein matrix was filtered for proteins without missing values (4363 proteins). PSM files were additionally downloaded from the CPTAC Data Portal and processed following the CDAP to the peptide level data in order to investigate peptide variability.

SWATH-MS data processing Raw mass spectrometry measurements obtained from the TripleTOF 5600+ in DDA and SWATH mode were converted to mzXML file format using msconvert (ProteoWizard v3.0.8851) (Chambers et al., 2012). DDA measurements from all 103 samples and 48 fractions of a pooled sample were searched with X!Tandem (2013.06.25.1) (Duncan et al., 2005), OMSSA (2.1.9) (Geer et al., 2004) and Comet (2015.02 rev. 3) (Eng et al., 2013). Identified peptides were processed through the Trans-Proteomic Pipeline (TPP v.4.7 Polar Vortex rev 0) using PeptideProphet, iProphet and ProteinProphet scoring (FDR <0.01) and SpectraST (Keller et al., 2005; Lam et al., 2007). The assay library was built according to (Schubert et al., 2015) using a set of 113 common internal retention time standards (ciRTs) (Parker et al., 2015). The resulting library contained 8897 proteins supported by proteotypic peptides.

SWATH data was analyzed using OpenSWATH (Rost et al., 2014) from OpenMS (v.1.10.0) with the previously described sample-specific assay library. FDR was controlled

using PyProphet (v.0.0.19) allowing for a peptide-query FDR of 1% and protein-FDR of 1% as described by Rosenberger et al. (Rosenberger et al., 2017). Runs were aligned for improved quantification using TRIC (msproteomicstools master branch from GitHub commit c10a2b8) (Rost et al., 2016).

OpenSWATH results were annotated and formatted using the R package SWATH2stats (v.1.0.3) (Blattmann et al., 2016) prior to quantile normalization (preprocessCore v.1.32.0) (Bolstad, 2017) and batch-wise mean-centering batch correction of known sample preparation batches. Fragment-level data was then fed into mapDIA (v.3.0.2) (Teo et al., 2015) in order to filter for outliers and proteins with at least 3 fragments per peptide and 2 peptides per protein. The 3 most intense fragments per peptide were summed and the resulting peptide intensities were divided by the mean peptide intensity (as an internal reference) over all runs to obtain a similar data structure as in iTRAQ DDA. Peptide ratios were averaged per protein. A total of 2914 proteins were quantified in this manner across all 103 samples.

An additional filtering and imputation step was conducted to mitigate the missing values in the SWATH data. **Figure S2B** shows the number of missing values in the SWATH data and their occurrence dependent on the protein abundance represented by \log_2 intensity (**Figure S2C**). As expected, there is an inverse relationship between protein abundance and missing values. Missing values in low abundance proteins are more likely to occur due to technical reasons, while missing values amongst high abundance proteins potentially have biological significance which might be important for further analysis. We therefore filtered the protein matrix based on the occurrence of abundance-dependent missing values. We confined the number of missing values using a stricter threshold for low abundance proteins than high abundance proteins according to the following rules: no missing values were allowed for the lowest abundant 10% of proteins, 10% missing values were allowed for the next 10% abundant proteins and so on until 90% missing values were allowed for the highest abundant proteins. A total of 1659 proteins remained that passed these filtering steps. The missing values from those proteins were filled using an imputation method based on what is used by Perseus software (Tyanova et al., 2016)

wherein random numbers forming a normal distribution were generated fitting in the left end of the distribution of the measured values, representing values within the noise.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatics pipeline

Bioinformatic analyses were performed in R (v.3.2.2 “Fire Safety”).

Technical evaluation. A total of 1599 proteins common to the iTRAQ DDA and SWATH data sets were evaluated for their correlation and variability (**Figure 2**). The distribution of values in both data sets was plotted using the R package vioplot (v.0.2) (Adler, 2005). Spearman’s rank correlation was computed for each protein across all 103 samples in both data sets. The resulting Spearman’s rho values were ordered and the median was calculated. For peptide variability, a CV-like score was calculated for the peptides from each protein (standard deviation of peptides per protein divided by mean peptide intensity per protein). The distribution of the resulting CVs was plotted with vioplot.

Molecular classification comparison. Protein matrices were transformed into z-scores and a molecular classification was obtained by running a model-based clustering with mclust from the R package mclust under default settings (v.5.2) (Fraley and Raftery, 2002; Fraley et al., 2012). Protein modules characterizing sample groups were obtained with weighted gene-correlation network analysis (WGCNA v.1.51) (Langfelder and Horvath, 2008, 2012) and functional enrichment with clusterProfiler (v.2.4.3) (Yu et al., 2012) (**Figure S4**). The resulting molecular classifications were compared based on their similarity using the Adjusted Rand Index (ARI), a measure for classification similarity, implemented within the R package mclust previously used. The heat map displayed in **Figure S3A** was generated based on SWATH z-scores using pheatmap (v.1.0.8) (Kolde, 2015). The stability of the classification results was evaluated by bootstrapping. Different fractions of samples or proteins were drawn 100 times and classified using mclust with default settings, and the similarity to the molecular classification of the full data set was compared with the ARI (**Figure 4**).

In order to assess the stability of the mesenchymal class a Fisher's exact test was performed on the mesenchymal class and the other resulting classes with a bootstrapping approach by drawing a fraction of samples 100 times (**Figure S3B**). Group comparison of the original classes from CPTAC (Zhang et al., 2016) were obtained by a student's t-test (**Figure 6** and **S4**).

HRD group comparison. Group comparison to identify the differentially expressed proteins in the HRD vs. non-HRD patients was conducted using mapDIA (v.3.0.2) (Teo et al., 2015) on fragment ion-level data (**Figure 7A**). The resulting fold changes were used as input signal in a network propagation approach (Hofree et al., 2013) called Network Smoothing (R package BioNetSmooth v.1.0.0) (Chokkalingam et al., 2016).

As an input network the STRING database for human (Taxon 9606 v.10.5) was downloaded and filtered for interactions with experimental evidence and a score >800. The top 5% of negative and positive scores were used for further investigation of functional enrichment of Gene Ontology using the STRING database (STRING-db.org v.10.5 accessed on 2018/03/14).

DATA AND SOFTWARE AVAILABILITY

Mass spectrometry data

All the raw data from SWATH-MS measurements, along with the input spectral library and OpenSWATH results can be freely downloaded from the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) with the dataset identifier: PXD010437 via the PRIDE partner (Vizcaino et al., 2016).

Reviewer account details:

Username: reviewer95375@ebi.ac.uk

Password: 9VoPWx2g

The iTRAQ DDA data are available on the NCI CPTAC Data Portal (<https://cptac-data-portal.georgetown.edu/cptac/s/S026>); mass spectrometry site: Johns Hopkins University. Other data are available from the corresponding authors upon request.

Programming codes

All code used for the downstream bioinformatics pipeline can be obtained from the github repository ([bfriedrichgrube/OC_CPTAC_iTRAQ_SWATH](https://github.com/bfriedrichgrube/OC_CPTAC_iTRAQ_SWATH)).

SUPPLEMENTARY REFERENCES

Adler, D. (2005). *vioplot: Violin plot*. R package version 0.2.

Blattmann, P., Heusel, M., and Aebersold, R. (2016). *SWATH2stats: An R/Bioconductor Package to Process and Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools*. *PLoS One* *11*, e0153160.

Bolstad, B. (2017). *preprocessCore: A collection of pre-processing functions*. R package version 1.32.0.

Cancer Genome Atlas Research, N. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* *474*, 609-615.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., *et al.* (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* *30*, 918-920.

Chokkalingam, M., Debes, C., and Beyer, a. (2016). *BioNetSmooth: Spreads signal on interaction network to identify altered subnetworks*. R package v. 1.0.0.

Duncan, D.T., Craig, R., and Link, A.J. (2005). *Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem*. *J Proteome Res* *4*, 1842-1847.

Edwards, N.J., Oberti, M., Thangudu, R.R., Cai, S., McGarvey, P.B., Jacob, S., Madhavan, S., and Ketchum, K.A. (2015). *The CPTAC Data Portal: A Resource for Cancer Proteomics Research*. *J Proteome Res* *14*, 2707-2713.

Ellis, M.J., Gillette, M., Carr, S.A., Paulovich, A.G., Smith, R.D., Rodland, K.K., Townsend, R.R., Kinsinger, C., Mesri, M., Rodriguez, H., *et al.* (2013). Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium. *Cancer Discov* *3*, 1108-1112.

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

Fraley, C., and Raftery, A. (2002). *Model-based clustering, discriminant analysis, and density estimation*. *Journal of the American Statistical Association* *97*, 611-631.

Fraley, C., Raftery, A., Murphy, T., and Scrucca, L. (2012). *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation* University of Washington Department of Statistics Technical Report No 597.

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

Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat Methods* 10, 1108-1115.

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

Kolde, R. (2015). pheatmap: Pretty Heatmaps. R package version 1.0.8.

Lam, H., Deutsch, E.W., Eddes, J.S., Eng, J.K., King, N., Stein, S.E., and Aebersold, R. (2007). Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7, 655-667.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

Langfelder, P., and Horvath, S. (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* 46.

Parker, S.J., Rost, H., Rosenberger, G., Collins, B.C., Malmstrom, L., Amodei, D., Venkatraman, V., Raedschelders, K., Van Eyk, J.E., and Aebersold, R. (2015). Identification of a Set of Conserved Eukaryotic Internal Retention Time Standards for Data-independent Acquisition Mass Spectrometry. *Mol Cell Proteomics* 14, 2800-2813.

Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C.L., Liu, Y., MacCoss, M.J., MacLean, B.X., Nesvizhskii, A.I., Pedrioli, P.G.A., *et al.* (2017). Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. *Nat Methods* 14, 921-927.

Rost, H.L., Liu, Y., D'Agostino, G., Zanella, M., Navarro, P., Rosenberger, G., Collins, B.C., Gillet, L., Testa, G., Malmstrom, L., *et al.* (2016). TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. *Nat Methods* 13, 777-783.

Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmstrom, J., Malmstrom, L., *et al.* (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 32, 219-223.

Schubert, O.T., Gillet, L.C., Collins, B.C., Navarro, P., Rosenberger, G., Wolski, W.E., Lam, H., Amodei, D., Mallick, P., MacLean, B., *et al.* (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc* 10, 426-441.

Teo, G., Kim, S., Tsou, C.C., Collins, B., Gingras, A.C., Nesvizhskii, A.I., and Choi, H. (2015). mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J Proteomics* 129, 108-120.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13, 731-740.

Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., *et al.* (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* *44*, D447-456.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* *16*, 284-287.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., *et al.* (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell*.