**A streamlined mass spectrometry-based proteomics workflow for**

**large scale FFPE tissue analysis**

Coscia, Doll, *et al. J Pathol* DOI: 10.1002/path.5420

Reference numbers refer to the main text list

**Figure S1. Mass spectrometry-based proteomic workflow for FFPE tissues**

(A)FFPE tissue collection from glass slides by macro-dissection. H&E stained tissue sections are collected by scraping with a scalpel and transferred into PCR tubes in a small droplet (3 µl) of water. (B)Comparison of SDS and TFE-based protocols. Left panel: Example pictures of OvCa tissue section replicates of the same tumor (10 µm thick) used to compare SDS and TFE- based protocols. Red boxes indicate the collected areas by scraping (approximately 5 mm x 5 mm). Right panel: Total peptide yield (µg) after clean-up by StageTips measured on a Nanodrop instrument. (C)Quantitativecomparison of SDS and TFE-based protocols. Left panel: UpSetplot showing intersections of protein quantification across the two protocols. Right panel: pairwise proteome comparison of tissue replicates processed with the SDS or TFE-based protocol. Protein levels are median centered and log10 transformed with a Pearson correlation (R) of 0.96. **(**D)Glioma FFPE tissue used for protocol comparison (TFE, SDS and RapiGest). For each protocol, three areas (labeled 1–3) from three consecutive 10 µm sections of the same tissue were collected, processed and analyzed by LC-MS/MS. (E)Protein identifications from major cellular compartments (‘cytosol’, ‘nucleus’, ‘plasma membrane’ and ‘extracellular region’, Gene Ontology Cellular Component, GOCC) using different lysis buffers (TFE, SDS and RapiGest) and tissues (ovarian cancer and glioma). Percentages are the number of quantified proteins per compartment over all quantified proteins in the corresponding sample. The comparison to FrFr OvCa tissue [50] is included. (F)Comparison ofrelative protein abundance between OvCa and glioma samples processed with the TFE, SDS or RapiGest-based protocol. Box plots show relative protein abundance (sample median normalized) of ‘Chromatin binding’ and ‘DNA binding’ (Gene Ontology Molecular Function, GOMF) proteins. OvCa proteomes obtained from FrFr tissue [50] is plotted for relative comparison to non-formalin fixed tissue. (G)Open modification search (pFIND) results of FFPE tissues. The table summarizes the most abundant peptide modifications and tryptic miscleavage rates of OvCa tissues (left) and glioma (right) processed with the TFE, SDS or RapiGest-based protocol. Previous OvCa studies are included for comparison. FrFr OvCa [50], FFPE OvCa [12]. (H)Comparison of lysine methylation rates between SDS- and TFE-based protocols. The ratios of lysine methylated peptide intensities to unmodified base peptide intensities are shown as a boxplot. (I)Comparison of ACN and TFE-based protocols. Pictures of three OvCa tissue section replicates of the same tumor (10 µm thick) used to compare ACN and TFE-based protocols. Red boxes indicate areas before and after collection by scraping (approximately 5 mm x 5 mm). (J)Proteomic results for two OvCa tissue samples processed in replicates with the ACN or TFE-based protocol. Sample 1 corresponds to sections shown in panel F. (K)Proteome correlation map (Pearson r) for two OvCa tissue samples processed in replicates with the ACN or TFE-based protocol.

**Figure S2. FFPE tissue workflow is broadly applicable across tissue types**

(A)Streamlined FFPE tissue workflow in a 96-well format.The upper panel shows how 96 samples are processed with our tissue workflow. The lower panel depicts PCR tubes with H&E stained tissue in lysis buffer before and after sonication. (B)Total ion current of the mass spectrometric analysis of tissue samples collected by LCM or macro-dissection. (C) Principal component analysis (PCA) of ten FFPE tissue based on their proteomic expression profiles. Upper panel: The first and second component segregate the different FFPE tissues and account for 50% and 15% of the variability, respectively. Samples collected by LCM and macro-dissection are depicted as rectangles and dots, respectively. Ellipses indicate samples of same tissue origin. Proteins driving the segregation between the different FFPE tissues are depicted in the lower panel. Several known protein markers for the corresponding tissues are highlighted. (D)Proteome correlation matrix of five laser microdissected OvCa tissues obtained from two patients with high-grade serous OvCa. Depicted values are Pearson correlations. FT: invasive Fallopian tube cancer, OM: omental OvCa metastasis, Ov: invasive OvCa, GBM: glioma, Urachus: urachal carcinoma. (E)Total number of quantified protein groups for macro-dissected tissues in 100 min single-shot DDA analysis. Grey bar charts represent the protein groups only identified by MS/MS, orange bar plots the protein groups identified by matching (‘match-between-runs’), and red bar plots the protein groups identified by ‘BoxCar’ acquisition.

**Figure S3. Immunohistochemical control stains**

Positive control tissue: normal colon (CDX2), skin (CD44). Negative control tissue: liver (CDX2 and CD44). Note, biliary epithelial cells and hepatocytes are CD44 negative and the staining observed is due to lymphocytes, endothelial cells, and hepatic stellate cells (HSCs). Staining control: diluent instead of primary antibody. Scale bars,100 µm for CDX2, 50 µm for CD44 images.

**Figure S4. Reproducible and streamlined proteomic analysis of adenoma tissues of varying archival time**

(A)Proteome correlation matrix of three injection replicates (#1–3) of the same adenoma sample, with corresponding Pearson correlations. (B) Total number of quantified protein groups in the adenoma tissue cohort (N=118) in relation to data completeness. (C) Consensus Cumulative Distribution Function (CDF) plot. The cumulative distribution functions of the consensus matrix are plotted for each number of clusters, k, (indicated by colors), estimated by a histogram of 100 bins. The plot allows the determination of an optimal number of clusters, k, where the CDF reaches an approximate maximum with consensus and cluster confidence at a maximum. (D) Delta area plot showing the relative change in area under the CDF curve comparing k and k − 1. For k = 2, no k -1 is present, explaining why the total area under the curve rather than the relative increase is plotted. This plot enables the determination of the relative increase in consensus and k at which there is no or little increase. (E)Intensity of commonly and uniquely quantified peptides (upper panel) and protein groups (lower panel) in the three archival time groups. (F)Effect of archival time on the number of quantified peptides. Upper panel: Box plots showing total ion currents (TIC) of the samples used for pairwise comparison. Samples of different archival groups (6–9 years and 14–20 years) were matched by their median TIC (calculated for the main peptide elution window 10–80 min in the 100 min gradient). Lower panel: Bar charts showing the number of quantified peptides for each sample. Sample pairs matched by TIC are depicted next to each other. (G) Scatter plot showing the number of quantified peptides *versus* archival years. A linear model-based trend line is shown including 95% confidence intervals (grey area). (H) Total number of quantified proteins in FFPE tissues archived with different storage times. All proteins that were quantified in each group with at least 50% valid values are shown in the left panel, whereas the 3,000 most abundant proteins in the dataset (calculated by median protein level across all 98 samples) are plotted in the right panel.