Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Comparative analysis of plasma sample preparation methods for high-throughput mass spectrometry-based proteomics and application to colorectal cancer

**Jianing Liu**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Catarina Almeida-Marques**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Berend Gagestein**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Liva Stalidzane**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Lijie Xu**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Thang V. Pham**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Jaco C. Knol**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Sander R. Piersma**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Irene V Bijnsdorp**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Ka Wan Li**
Amsterdam UMC, location Vrije Universiteit Amsterdam
**Connie R. Jimenez**
c.jimenez@amsterdamumc.nl

Amsterdam UMC, location Vrije Universiteit Amsterdam

---

**Research Article**

**Keywords:**

**Posted Date:** October 20th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11240936/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Title page

# Comparative analysis of plasma sample preparation methods for high-throughput mass spectrometry-based proteomics and application to colorectal cancer

Jianing Liu[1,2*], Catarina Almeida-Marques[1,2*], Berend Gagestein[1,2], Liva Stalidzane[1,2], Lijie Xu[1,2], Thang V. Pham[1,2], Jaco C. Knol[1,2], Sander R. Piersma[1,2], Irene V Bijnsdorp[1,2,3], Ka Wan Li[1], Connie R. Jimenez[1,2#]

[1] Amsterdam UMC, location Vrije Universiteit Amsterdam, Department of Medical Oncology, OncoProteomics Laboratory, Amsterdam, The Netherlands

[2]Cancer Center Amsterdam, Biomarkers and Imaging, Amsterdam, The Netherlands

[3]Amsterdam UMC, location Vrije Universiteit Amsterdam, Urology, Amsterdam, The Netherlands

* *Shared first authors*

\# *Correspondence: c.jimenez@amsterdamumc.nl*

## Abstract

**Background**

In recent years, multiple methods for high-throughput plasma proteome sample preparation have emerged. This study aimed to compare four scalable and affordable plasma sample preparation approaches coupled to mass spectrometry-based profiling and evaluate their applicability to clinical samples.

**Methods**

The following plasma sample preparation methods were included: in-solution digestion (ISD), suspension trapping (S-Trap), perchloric acid precipitation (perCA), and extracellular vesicle (EV) enrichment using hyper-porous strong-anion exchange magnetic microparticles (Mag-Net). Aliquots from a pooled human plasma sample were analyzed in workflow triplicate using data-independent acquisition mass spectrometry. Furthermore, a plasma cohort of patients with colorectal cancer was processed using ISD and Mag-Net to evaluate their clinical applicability.

**Results**

Mag-Net achieved the greatest proteome depth (~1000 proteins per sample) and the highest protein identification reproducibility (data completeness = 87.4%), whereas global plasma proteomics by ISD showed the best quantitative reproducibility (CV = 8.5%). Across all workflows, 1130 proteins were identified in all three technical replicates of at least one workflow, of which 399 were shared by all four workflows. Functional analyses revealed differential enrichment of largely overlapping functions. In line with previous work, Mag-Net enriched for extracellular vesicles.

Two workflows, Mag-Net (best in-depth workflow) and ISD (best global workflow), were subsequently applied to a cohort of 30 colorectal cancer (CRC) patients and 29 healthy controls. Both workflows successfully distinguished CRC from healthy samples and recovered known CRC-associated proteins. Pathway and functional analyses revealed CRC-associated biological hallmarks: MYC targets, MTORC1

signalling and mitotic spindle. Random forest analysis identified workflow-specific candidate biomarker panels, with combined top 20 protein models achieving Area Under Curve (AUC) values of 0.985 and 0.964 for Mag-Net and ISD, respectively. Importantly, multiple top ranked biomarker candidates were supported by external CRC tissue and plasma proteomics datasets.

**Conclusions**

Together, our findings demonstrate that both global (ISD) and in-depth plasma proteomics (Mag-Net) are suitable, reproducible workflows for clinical applications that warrant further exploration for biomarker discovery in disease.

## Keywords

Plasma proteome; Sample preparation workflows; Extracellular vesicle enrichment; Data-independent acquisition mass spectrometry; Colorectal cancer

## Background

Plasma is one of the most clinically relevant biofluids and serves as a cornerstone for biomarker discovery and disease monitoring. Plasma collection is minimally invasive, provides a systemic snapshot of physiological and pathological states, and in view of its strong translational potential, plasma is a highly attractive biological matrix for clinical proteomics [1]. In recent years, with the advent of high-throughput, high-resolution mass spectrometers [2, 3] and novel sample preparation workflows [4], both global plasma proteomics and in-depth plasma proteomics using depletion- and

enrichment workflows have emerged as valuable approaches for plasma proteomics and biomarker discovery [5, 6].

In this study, we evaluated four scalable and affordable methods for plasma proteomics: In-solution digestion (ISD), as a fast and reproducible global proteomics approach [7]; Suspension trapping filter (S-Trap) that employs a filter-based mechanism to efficiently capture, digest and clean-up SDS-solubilized proteins [8] as an alternative global proteomics approach; Perchloric acid precipitation (perCA) as a cost-effective depletion strategy that selectively precipitates abundant plasma proteins, facilitating the identification of low-abundance, tissue-derived proteins that are otherwise difficult to detect [9]; Strong anion-exchange beads (Mag-Net) which base on magnetic affinity capture to enrich extracellular vesicles and other membrane-bound particles, while depleting abundant plasma proteins [10, 11].

Considering that each workflow offers distinct advantages, we aimed for a direct comparison to determine how these strategies compare in terms of proteome depth, reproducibility, and applicability in plasma proteomics of disease. To this end, we performed a workflow comparison of ISD, S-Trap, perCA, and Mag-Net using pooled human plasma analyzed in triplicate by single-shot DIA-MS (Data-Independent Acquisition mass spectrometry). Moreover, the best performing global (ISD) and enrichment/depletion (Mag-Net) plasma proteomics workflows were applied to a colorectal cancer (CRC) case–control cohort, underscoring the potential of both workflows for biomarker discovery in clinical proteomics.

# MATERIALS AND METHODS

## Reference plasma samples

Reference plasma samples were obtained from healthy individuals with written informed consent to use their biospecimens for research purposes. Equal volumes of each plasma sample from each donor were pooled, aliquoted, and stored at −80 °C until use. Plasma was thawed on ice immediately before processing.

## Clinical plasma samples

Plasma samples were collected from patients with colorectal cancer (CRC, n = 30) and healthy controls (n = 29), which were provided by the Liquid Biopsy Center of the Amsterdam UMC, and the application was reviewed by the scientific and ethical committee (043-LBC-2022: 2022.0432 and 2022.0246). Written informed consent was obtained from all participants.

## In-solution protein digestion

Five µL of plasma was mixed with 55 µL of lysis buffer containing 8 M urea (catalog no. U5378, Sigma-Aldrich), 10 mM dithiothreitol (DTT; catalog no. D9163, Sigma-Aldrich) and 50 mM ammonium bicarbonate (ABC; catalog no. A6141, Sigma-Aldrich). Proteins were denatured and reduced at 30 °C for 1h, followed by alkylation with iodoacetamide (IAA; catalog no. I1149, Sigma-Aldrich) at a final concentration of 20 mM for 15 min at room temperature in the dark. Samples were diluted to a final volume of 500 µL using 50 mM ABC to reduce the urea concentration to <1M, and 50 µL of each diluted sample was transferred to a new low-binding microtube (catalog no. 30108116, Eppendorf). Proteins were digested overnight at 37 °C with 30 µL trypsin

(catalog no. V5111, Promega; enzyme: protein ratio of 1:10). Digestion was quenched with 20 µL 10% formic acid (FA; prepared from Biosolve formic acid, catalog no. 069141).

**Suspension trapping (S-Trap) and protein digestion**

Five µL of each plasma sample was diluted with 195 µL lysis buffer (5% SDS,100 mM Tris, pH 8.0), and 50 µL was transferred to a new low-binding tube. Samples were reduced by adding 5 µL 55 mM DTT and incubating at 65 °C for 15 min, and alkylated by adding 5 µL 120 mM IAA and incubating at room temperature for 15 min in the dark. After acidification with 5.5 µL of 12% phosphoric acid, samples were mixed with 400 µL suspension buffer (90% methanol, 100 mM Tris, pH 8.0) and transferred to plasmid DNA micro-spin columns (HiPure, Magen Biotechnology, China), which act as suspension trapping filters. Columns were washed three times by addition of 400 µL washing buffer and centrifugation at 1400 x *g* for 1-2 min to remove SDS contaminants. On-filter digestion was performed overnight at 37 °C in a humidified incubator using 50 µL trypsin (enzyme: protein ratio of 1:10). Peptides were eluted by addition of (i) 80 µL 0.1% formic acid (FA) in water and (ii) 80 µL 0.1% FA in acetonitrile (ACN; catalog no. 012041, Biosolve), each time collecting eluates by centrifugation at 1400 x *g* for 1 min. Combined eluates (i) and (ii) were dried in a SpeedVac (Eppendorf, Germany).

**Strong anion exchange (SAX) bead-based protein capture (Mag-Net)**

Per sample to be processed, 25 µL MagReSyn SAX magnetic beads (Resyn Biosciences, Edenvale, South Africa; 20 mg/ml) were equilibrated three times with 500 µL wash buffer (50 mM Bis-Tris propane, pH 6.3, 150 mM NaCl), each time collecting the beads with a magnetic separator and removing supernatant. Each sample consisted of 50 µL of plasma mixed with 50 µL of freshly prepared binding buffer (100 mM Bis-Tris propane, pH 6.3, 150 mM NaCl, supplemented with protease inhibitor cocktail; catalog no. 11836170001, Roche), and was incubated with the equilibrated beads for 30 min at 1100 rpm shaking. After incubation, the beads were magnetically separated, the supernatant was discarded, and the beads were washed three times with 500 µL wash buffer. Bound proteins were denatured in 150 µL of lysis buffer (5% SDS, 100 mM Tris, pH 8.0), and 50 µL of the resulting suspension was transferred for reduction, alkylation, and tryptic digestion according to the suspension trapping protocol.

**Perchloric acid precipitation**

Fifty µL of plasma was diluted with 450 µL of Milli-Q water, and mixed with 25 µL of 70% perchloric acid to precipitate high-abundance plasma proteins through denaturation. The mixture was incubated at -20 °C for 15 min and then centrifuged at 3200 × *g* for 60 min at 4 °C. Subsequently, 400 µL of the supernatant was collected, supplemented with 40 µL of 1% trifluoroacetic acid (TFA; catalog no. 85183, Thermo Fisher) to adjust the solvent conditions for reversed-phase binding and loaded onto Oasis HLB cartridges (1 cc, 10 mg sorbent, Waters, Etten-Leur, Netherlands), which had been preconditioned with 1 mL of methanol and equilibrated twice with 1 mL of 0.1% TFA. The cartridges were washed twice with 1 mL of 0.1% TFA, and retained

proteins were eluted with 100 µL of 90% acetonitrile containing 0.1% TFA. Eluates were dried in a SpeedVac and resuspended in 40 µL 50 mM ABC. Proteins were digested overnight at 37 °C with 5 µL trypsin (enzyme: protein ratio of 1:10). The digestion was quenched with 5 µL 10% FA.

**LC-MS/MS acquisition**

The peptide solution (400 ng) was transferred to an Evotip and separated on an Evosep One liquid chromatography system (Evosep, Denmark) with the 30 samples-per-day gradient. Peptides were resolved on a 15 cm × 150 µm C18 column (1.5 µm beads, PepSep, Bruker Daltonics) and introduced via a nano-electroSPRAY source into a timsTOF HT mass spectrometer (Bruker Daltonics). Data were acquired in DIA-PASEF mode optimized for plasma proteomics, covering an m/z range of 300–1200 and an ion mobility range of 1.5–0.7 Vs $cm^{-2}$, with 10 DIA-PASEF scans per cycle (cycle time ~1.1 s). Collision energy was linearly ramped from 59 eV at 1.6 Vs $cm^{-2}$ to 20 eV at 0.6 Vs $cm^{-2}$.

**Data processing and statistical analysis**

Raw DIA files were processed with DIA-NN v1.8.1 [12] with a project-specific spectral library generated using FragPipe [13] against the UniProt reviewed human canonical and isoform reference proteome (release 2023_03). Neural network interference correction and cross-run normalization were enabled, and protein-level quantities were reported by the MaxLFQ algorithm using fragment intensities [14].

For the workflow comparison (ISD, S-Trap, perCA, Mag-Net) and the clinical cohort (CRC vs healthy), differential protein abundance was assessed with the limma package in R [15], with significance defined as raw $p < 0.05$. Unsupervised hierarchical clustering was performed using 1 - Spearman correlation for samples and Euclidean distance for proteins with complete linkage.

Bioinformatic analyses included Gene Ontology enrichment and hallmark pathway activity scoring (MSigDB gene sets [16] evaluated using ssGSEA implemented in GSVA package [17]), performed in R to characterize biological differences across workflows and between clinical groups.

For biomarker prioritization, significantly regulated proteins (raw $p < 0.05$) were ranked by random forest analysis [18], and receiver-operating characteristic (ROC) curves were generated to evaluate classification accuracy of individual and combined protein panels.

# Results

### Study design and outline of plasma workflows

We designed a two-stage study to evaluate plasma proteomics strategies (Figure 1). In the first stage, four workflows representing both whole plasma (ISD and S-Trap) and enrichment/depletion approaches (Mag-Net and perCA) were systematically benchmarked in technical triplicates (Figure 1A). In the second stage, the two best-performing workflows were applied to a CRC versus healthy cohort to assess translational relevance (Figure 1B). This design enabled us to both compare

methodological performance and demonstrate applicability in a clinical context.

**Figure 1**

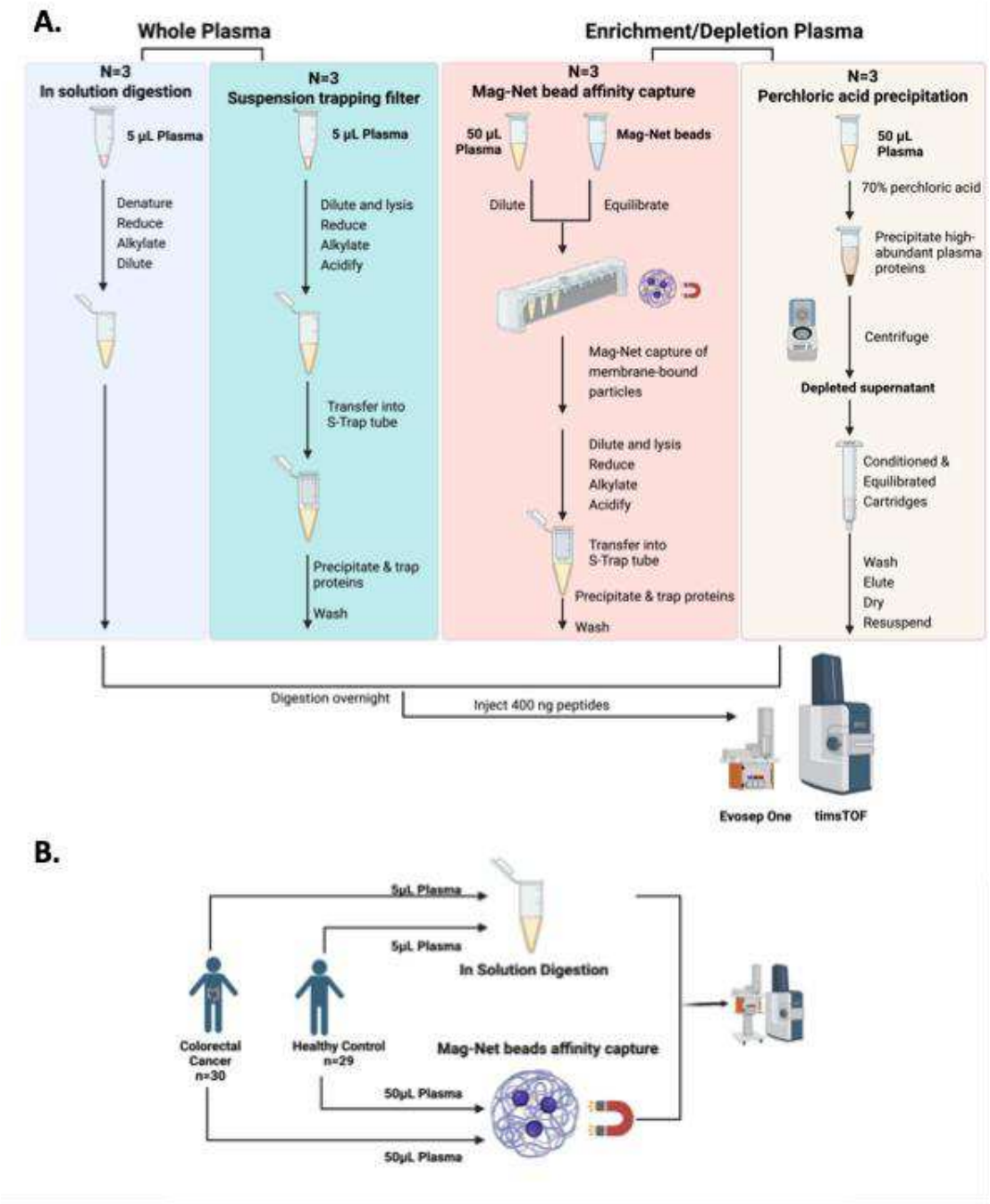



**Figure 1. Study design and plasma proteomics workflows**

(A) Experimental design for benchmarking four plasma proteomics workflows. ISD and S-Trap were performed on 5 µL plasma, Mag-Net on 50 µL plasma for extracellular

vesicle (EV) enrichment, and perCA on 50 µL plasma for depletion of high-abundance proteins. Following overnight digestion, peptides were analyzed by DIA using an Evosep One coupled to a timsTOF HT mass spectrometer. Created with BioRender.com.

(B) Clinical study design. Plasma samples from colorectal cancer (CRC) patients (n = 30) and healthy controls (n = 29) were analyzed using ISD and Mag-Net workflows followed by differential abundance, pathway, and biomarker analyses. Created with BioRender.com.

**Comparative performance of four plasma proteomics workflows**

To assess the impact of different plasma preparation strategies on identified proteome, we evaluated different aspects, including the number of identified proteins, reproducibility, data correlation and workflow specific proteins. Mag-Net yielded the highest (n = 976) average number of identified proteins per replicate and the largest set consistently detected in all three replicates (n = 907), demonstrating superior proteome coverage. ISD (n = 798 with 721 reproducibly identified proteins) identified a similar number proteins as perCA (771/713 proteins), and more than S-Trap (n = 725/622 proteins) (Figure 2A). Reproducibility was further assessed based on coefficients of variation (%CV) and data completeness (Figure 2A). Mag-Net achieved the highest reproducibility of identifications (87.4% of proteins consistently detected across all replicates), while ISD provided the most robust quantification reproducibility with the lowest average CV (8.5%). S-Trap exhibited intermediate performance in both indexes, and perCA showed the lowest reproducibility overall.

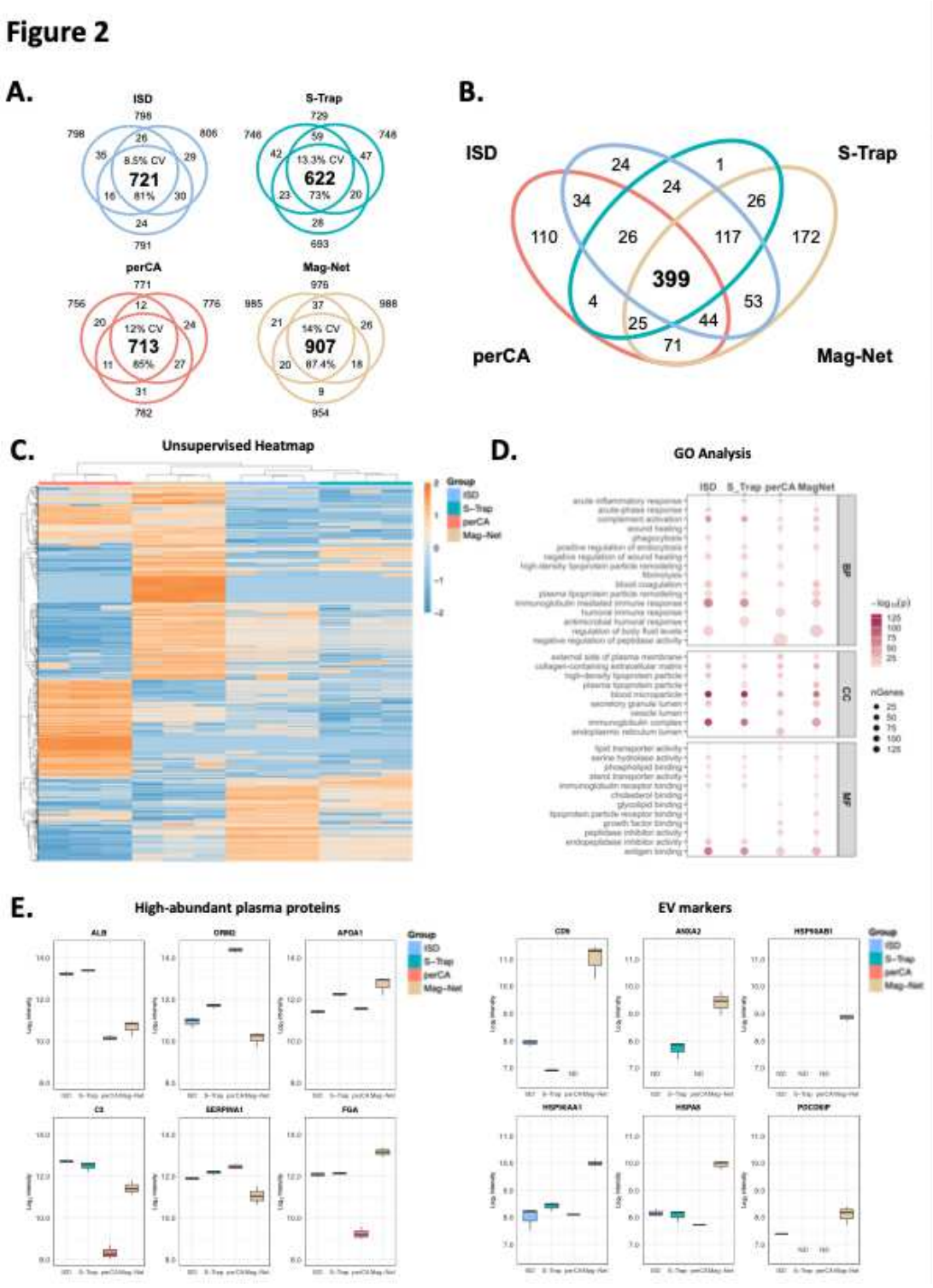
To investigate shared and workflow-specific proteome components, proteins reproducibly identified across all three replicates were compared using Venn analysis (Figure 2B). A substantial core plasma proteome (n = 399) was detected by all four workflows. Mag-Net contributed the largest unique subset (n = 172), followed by perCA (n = 110). In contrast, ISD (n = 24) and S-Trap (n = 1) yielded relatively few unique identifications, underscoring that enrichment/depletion-based methods recover overlapping and distinct protein populations beyond conventional whole-plasma workflows.

Unsupervised hierarchical clustering of normalized protein intensities of all identified grouped samples according to workflow, further emphasizes that each workflow uncovers overlapping and distinct plasma proteome features (Figure 2C). As expected, the two global plasma proteomics workflows (ISD and S-Trap) yield more similar proteomes, whereas Mag-Net and perCA formed distinct branches characterized by workflow-specific protein abundance patterns.

Gene Ontology (GO) enrichment analyses, performed separately using proteins identified in all three technical replicates for each workflow, further highlighted the biological information obtained using the different plasma proteomics workflows (Figure 2D). Overall, multiple GO terms related to inflammation, coagulation and immune response were shared across workflows (Figure 2D). Only a limited number of GO terms were restricted to one or two workflows. For example, perCA was uniquely enriched for endoplasmic reticulum lumen and growth factor binding among others. Surprisingly, this unbiased functional data mining did not yield strong enrichment of

(extracellular) vesicle-related signatures for the Mag-Net workflow. Together, these results demonstrate differential enrichment of largely overlapping biology across four workflows.

Finally, boxplot analyses of selected proteins provided additional insight into workflow-dependent detection patterns for high-abundance plasma proteins and canonical EV markers (Figure 2E). As expected, high-abundance plasma proteins generally showed the highest intensities using ISD and S-Trap, whereas lower intensities were observed using Mag-Net and perCA. In line with previous work that positioned the Mag-Net workflow as a plasma EV capture workflow [11], EV markers were most abundantly detected in Mag-Net, while ISD, S-Trap, and perCA identified only a subset of EV-associated proteins at much lower intensities.




247

248 **Figure 2. Comparative performance of four plasma proteomics workflows**

249 (A) Proteome depth, data completeness, and quantitative reproducibility across

250 workflows.

(B) Four-way Venn diagram showing shared and workflow-specific proteins reproducibly identified across technical triplicates.

(C) Unsupervised hierarchical clustering of all identified proteins across workflows.

(D) Gene Ontology enrichment analysis of reproducibly identified proteins. BP, Biological Process; CC, Cellular Component; MF, Molecular Function.

(E) Representative high-abundance plasma proteins and EV markers across workflows. Data are shown as $log_2$-transformed protein intensities. ND = not detected.

**Application of whole plasma and EV-enriched plasma proteomics to clinical CRC samples**

Having established the comparative performance of the four workflows, we next selected Mag-Net and ISD for clinical evaluation. Mag-Net provided the greatest proteome depth and the largest number of reproducibly identified proteins, whereas ISD was the best-performing global plasma proteomics method that achieved best quantitative reproducibility overall and similar proteome depth as compared to the depletion workflow (perCA) and better depth as compared to S-Trap. To this end, we profiled clinical CRC (n = 30) and healthy (n = 29) plasma samples (Figure 1B) and evaluated their clinical utility for biomarker discovery.

A total of 1547 proteins and 747 proteins were identified in the Mag-Net and ISD datasets, respectively, across all plasma samples. Unsupervised hierarchical clustering using all identified proteins with a data presence over 10% in each dataset (1544 proteins for Mag-Net and 736 proteins for ISD) showed partial separation between CRC

and healthy control samples in both datasets (Figure 3A). These results demonstrate that both workflows capture CRC-associated signatures.

We evaluated whether known CRC-associated protein features could be recovered. As representative examples, several proteins previously implicated in CRC biology, including ALCAM, IGFBP3, NOTCH1, NRAS, VWF, and YWHAE, were examined (Figure 3B) [19–24]. All proteins were detected using Mag-Net, while ISD of full plasma detected 4 of the 6 proteins. Importantly, when detected in both workflows, the direction of regulation was the same. For example, VWF and YWHAE were upregulated and IGFBP3 was down-regulated in CRC in both datasets. NOTCH1 (down) and NRAS (up) were uniquely identified in the Mag-Net dataset. Of note, the clinically used biomarker CEA was not detected in either dataset.

We further investigated whether EV-associated signals could be detected in plasma by checking well established EV markers, including CD9, CD63, FLOT1, HSP90AA1, HSPA8, and PDCD6IP (Figure 3C) [10,25]. Interestingly, multiple EV-associated proteins, including CD9, CD63, FLOT1, HSP90AA1, HSPA8, and PDCD6IP, showed increased abundance in CRC samples. HSP90AA1 and HSPA8 were significantly elevated in both workflows, whereas CD9, CD63, FLOT1, and PDCD6IP reached significance primarily in the Mag-Net dataset. Together, these observations indicate that EV-associated proteins are enriched using the Mag-Net workflow and exhibit biomarker potential.

**Figure 3**

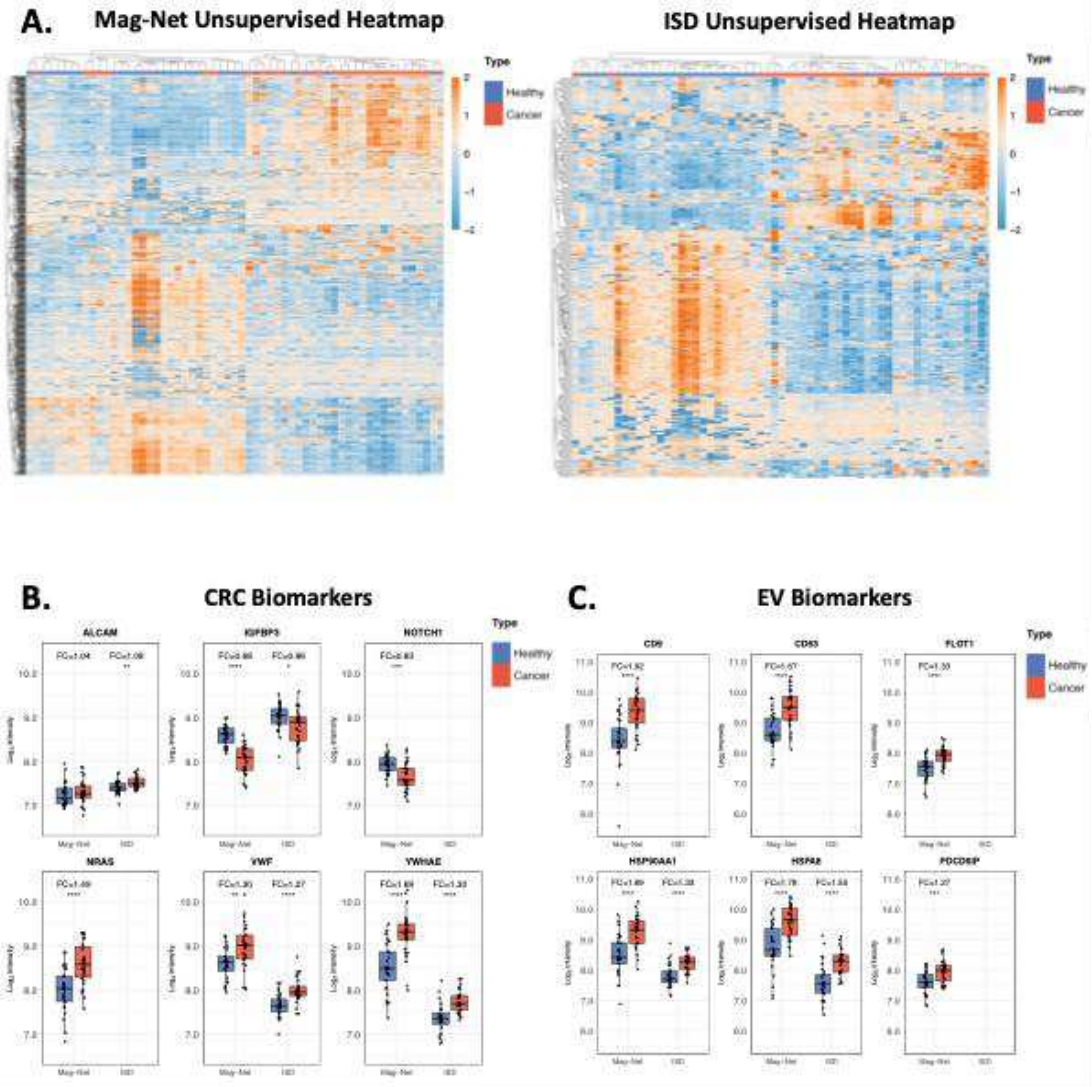



**Figure 3. Whole-plasma and EV-enriched proteomics reveal CRC-associated protein signatures**

(A) Unsupervised hierarchical clustering of CRC and healthy plasma samples analyzed using Mag-Net and ISD workflows.

(B) Representative CRC-associated proteins detected in Mag-Net and ISD datasets. Data are shown as $\log_2$-transformed protein intensities. Statistical significance is

299 indicated as *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

300 (C) Representative extracellular vesicle-associated proteins detected in CRC and
301 healthy plasma samples. Data are shown as $log_2$-transformed protein intensities.
302 Statistical significance is indicated as in panel B.

303 To further characterize the biological processes in a direct comparison of both
304 workflows, GO Biological Process enrichment analysis was performed using
305 differentially abundant proteins in CRC as identified in each workflow (Figure 4A and
306 B). This analysis showed differential processes overlapping in both workflows and that
307 Mag-Net workflow yields more distinct differential biology. Common processes
308 upregulated in CRC for example were response to stress, immune system process, and
309 anatomical structure morphogenesis, while uniquely identified cancer-associated
310 processes included various transport- and organelle-related terms (Mag-Net) and
311 animal organ development, apoptotic cell clearance and regulation of humoral response
312 (ISD).

313 To explore whether CRC-associated biology is traceable in plasma, Hallmark
314 enrichment analysis was performed using ssGSEA across CRC and healthy plasma
315 samples. All significantly differential Hallmarks are shown in Figure 4C and D. Among
316 the most significant Hallmarks with higher ssGSEA scores in CRC were MYC targets,
317 fatty acid metabolism, bile acid metabolism, MTORC1 and oxidative phosphorylation
318 in the Mag-Net dataset, whereas mitotic spindle, PI3K–AKT–mTOR signaling, apical
319 junction, MYC target and MTORC1 signaling were increased in the ISD dataset.

Overall, more significantly altered Hallmarks were identified in the Mag-Net dataset than in the ISD.

Collectively, Hallmark pathway analysis of colorectal cancer and healthy control plasma samples processed using ISD and Mag-Net revealed shared and workflow-specific CRC-associated biological processes, while both workflows identified proteins with potential biomarker relevance.

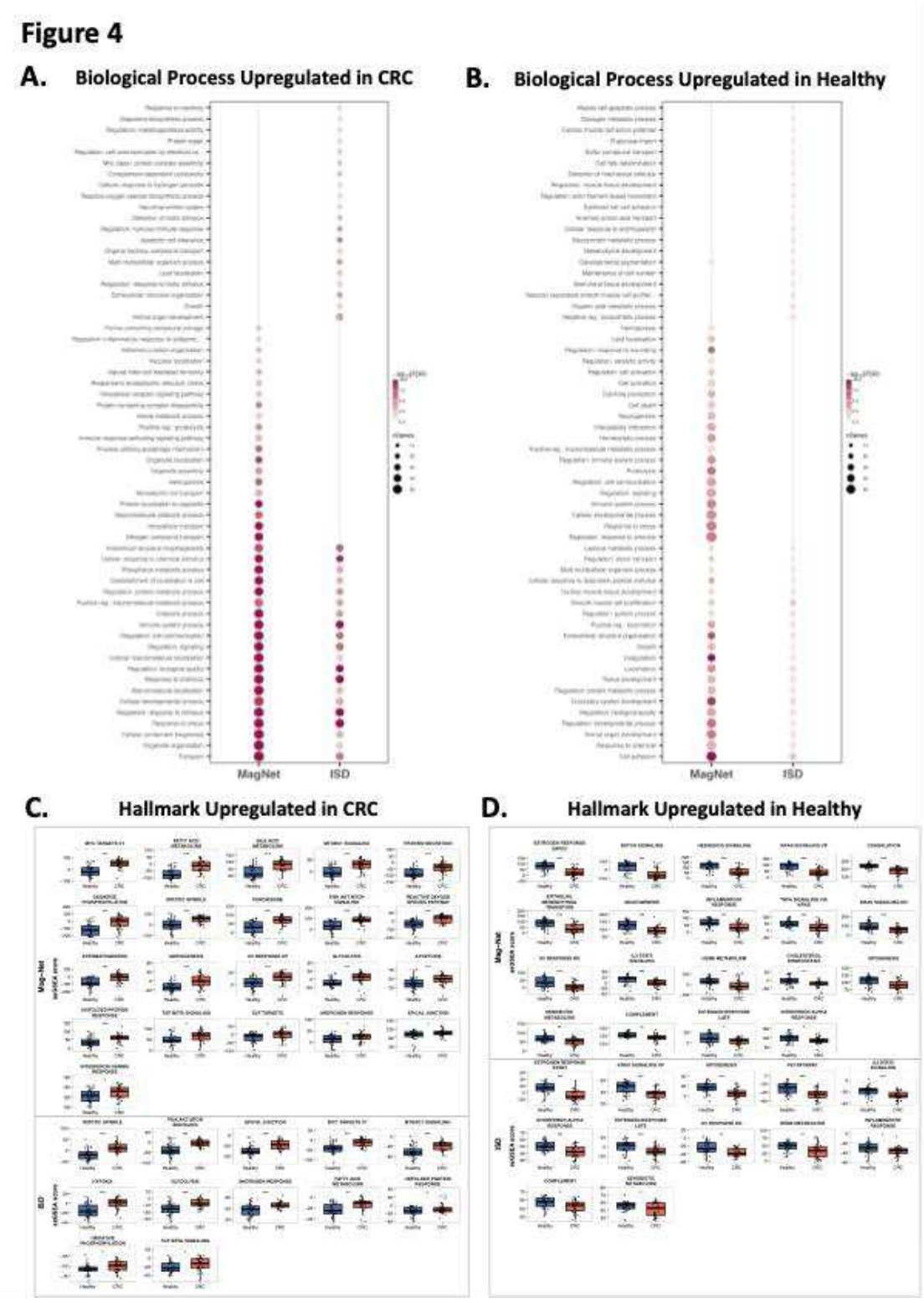



**Figure 4. Functional characterization of CRC-associated plasma proteomes.**

(A–B) Gene Ontology Biological Process enrichment analysis of proteins with significantly higher abundance in CRC (A) or healthy controls (B). Analyses were

performed separately for the Mag-Net and ISD datasets. Dot size represents the number of proteins associated with each term, and color represents enrichment significance (−log10 false discovery rate). (C–D) Hallmark pathway analysis using single-sample Gene Set Enrichment Analysis. Box plots show significantly altered Hallmark pathways with higher enrichment scores in CRC (C) or healthy controls (D) in the Mag-Net and ISD datasets. Points represent individual samples. Statistical significance is indicated as in Figure 3.

**Identification and evaluation of candidate CRC biomarkers**

To pinpoint promising CRC-associated proteins for classification, random forest (RF) analysis was performed independently on the Mag-Net and ISD datasets. Distinct top ranked protein panels of differential proteins were identified for each workflow (Figure 5A). Most (15/20, Mag-Net) and more than half (12/20, ISD) top-ranked proteins were supported as CRC biomarkers by independent evidence from external tissue (CPTAC [26]) and plasma (HPA [27]) proteomic resources. Detailed biological annotations and cancer relevance for all top-ranked proteins are provided in Supplementary Table 1. In addition, Supplementary Figures 1 and 2 show the substantial direction-concordant overlap across the complete differential plasma proteomes with external proteomics data from CPTAC and HPA.

The diagnostic performance of these 20 ranked proteins by RF was subsequently evaluated by receiver operating characteristic (ROC) analysis (Figure 5B). The combined twenty-protein models showed strong discriminatory performance in both

Mag-Net and ISD datasets, with AUCs of 0.985 and 0.964, respectively. Because feature selection and model evaluation were performed within the same cohort, these performance estimates should be considered exploratory and require validation in independent patient cohorts.

To further investigate the performance of the top 20 plasma EV proteins in an external cohort, we re-analyzed three publicly available plasma EV proteomics datasets (Supplementary Figure 3). Details of the studies are provided in Supplementary Table 2 [28, 29, 37]. Proteins detected in more than 10% of samples were retained, and differential abundance between CRC and healthy controls was determined using limma (raw $p < 0.05$), analysis details are provided in Supplementary Table 3. The Wang *et al* study was selected for further evaluation of Mag-Net derived protein panel based on its larger cohort size, broader MS-based protein coverage, and greater overlap with the Mag-Net dataset. Nine of the Mag-Net top-20 RF features were quantified in the Wang dataset. As a combined protein-panel, these nine proteins distinguished CRC from healthy controls in the independent Wang cohort with an AUC of 0.932 (Figure 5C). The same panel achieved an AUC of 0.992 in the Mag-Net cohort (Figure 5C).

Together, these findings indicate that Mag-Net captures a CRC associated EV protein signature with discriminatory potential across independent cohorts and EV isolation workflows.

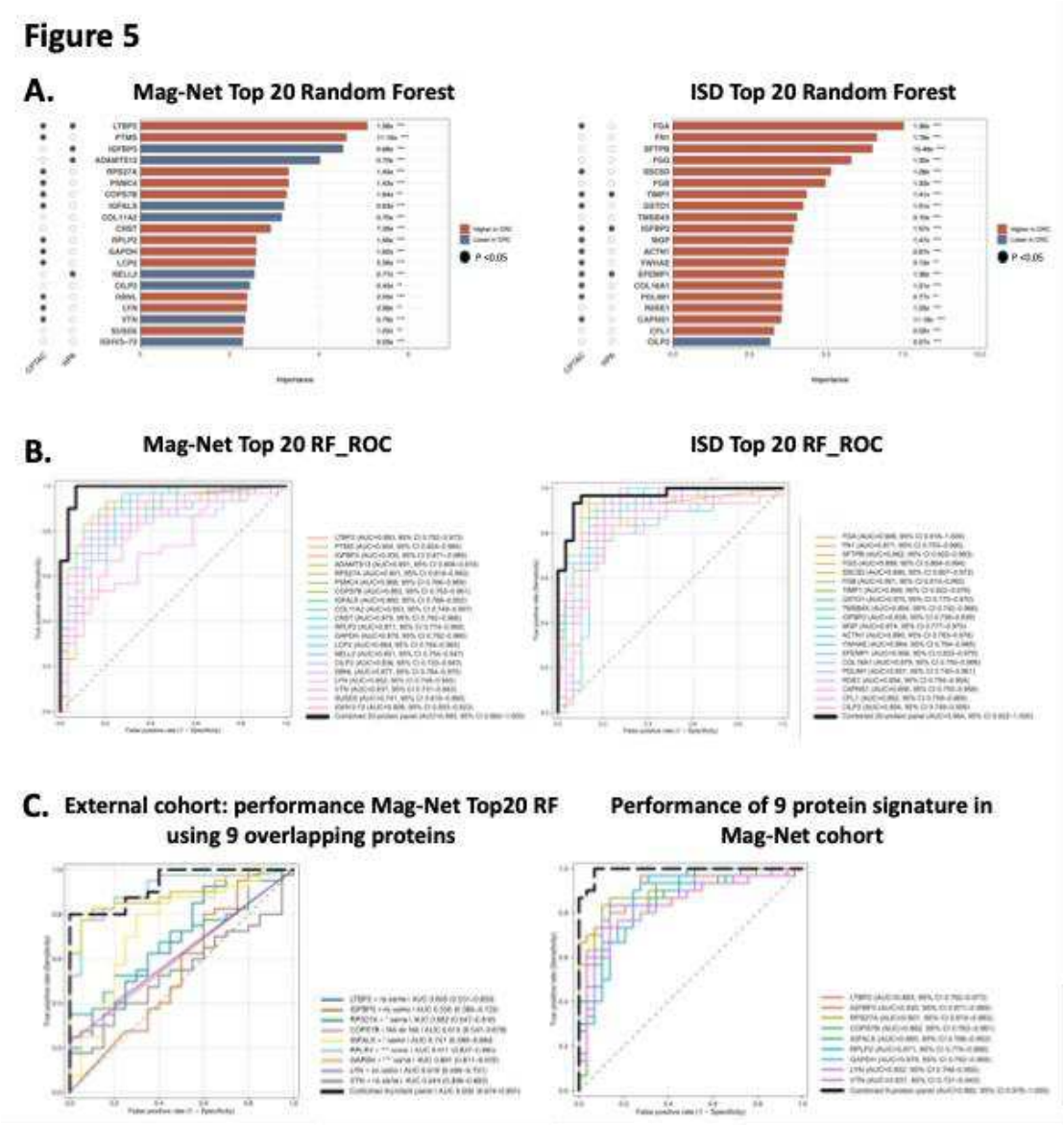



**Figure 5. Identification and evaluation of candidate CRC biomarkers**

(A) Random Forest ranking of the top 20 discriminative proteins identified in the Mag-Net and ISD datasets. Proteins were ranked according to RF variable importance. Bar colors indicate higher (red) or lower (blue) abundance in CRC relative to healthy controls. Fold-change values and significance levels from the differential analysis are shown beside each protein. Filled circles indicate proteins significantly altered ($p < 0.05$) in the CPTAC or HPA dataset; open circles indicate proteins that were not significantly

altered or were unavailable in the corresponding dataset.

(B) ROC curves for the 20 highest-ranked RF proteins from the Mag-Net and ISD workflows. Curves show the performance of each protein individually and of the combined twenty-protein RF panel. AUC values and corresponding 95% confidence intervals are provided in the legends. The dashed diagonal line indicates performance expected by chance, and the black curve represents the combined 20-protein panel.

(C) ROC analysis of the Mag-Net RF-ranked proteins (top 20) available in the external cohort. The shared protein panel was fixed before ROC analysis and was not reselected according to its performance in the external data. Individual-protein and combined-panel ROC curves are shown for the in-house Mag-Net discovery cohort (right) and for the same proteins in the external cohort (left). AUC values and corresponding 95% confidence intervals are provided in the legends. The dashed diagonal line indicates performance expected by chance, and the black dashed curve represents the combined protein panel.

## Discussion

The present study demonstrates that plasma preparation shapes the plasma proteome and thereby biological and disease information. Whole-plasma (ISD and S-Trap), enrichment-based (Mag-Net), and depletion-based (perCA) workflows yielded overlapping (399 proteins) yet different circulating proteomes. Mag-Net outperformed the other workflows in terms of depth, providing the highest number of reproducibly identified proteins (907 versus 622-721). Furthermore, the ISD workflow provided

399 highest reproducibility of protein quantitation (%CV 8.5 versus 12-14). When applied
400 to the CRC cohort, both ISD and Mag-Net yielded discriminatory CRC related proteins
401 in a 30 samples per day analysis that balances throughput and depth.

402 In line with a previous plasma proteomics methods comparison, Mag-Net outperformed
403 the other workflows in terms of depth. In addition, Mag-Net workflow yielded the best
404 enrichment of EV markers, underscoring this method as a plasma EV capture approach
405 [30]. Importantly, EV-associated markers exhibited significantly elevated levels
406 observed in patients with CRC. These findings align with previous literature
407 demonstrating that tumorigenesis leads to an increased rate of EV secretion into
408 circulation, resulting in higher circulating EV concentrations in blood plasma [31, 32].
409 Because EV-associated and cancer-specific markers are also elevated when applying
410 ISD directly to full plasma, complete EV isolation may not be strictly necessary for
411 downstream clinical diagnostic testing. Nevertheless, given that isolating EVs yields a
412 deeper analytical depth of coverage and superior sample separation, EV isolation
413 remains highly recommended for initial biomarker discovery workflows, whereas
414 direct ISD can serve as an efficient, streamlined approach for subsequent clinical
415 validation.

416 Both ISD and Mag-Net successfully distinguished CRC from healthy plasma samples
417 using top 20 ranked proteins identified in each workflow, indicating that both the global
418 plasma proteome and the plasma EV proteome harbor cancer-associated information.
419 Moreover, both workflows identified well-established proteins associated with CRC

biology including IGFBP3, VWF, and YWHAE - that are implicated in CRC progression, coagulation, and signaling [19, 24, 33]. Their detection in plasma demonstrates the potential of unbiased proteomics to broaden the biomarker spectrum beyond conventional assays. Beyond the identification of individual proteins, pathway-level analyses showed that the observed proteomic alterations reflected biological processes relevant to CRC such as inflammation, immune regulation, extracellular matrix remodeling, cellular communication, metabolic adaptation, and cellular stress responses. These findings are in agreement with previous proteomic studies of CRC, including tissue, plasma, and plasma EV proteomics, further supporting the biological relevance of the observed proteomic alterations [28, 34, 35]. The prominent representation and differential abundance of CRC-associated EV proteins further emphasizes the importance of vesicle biology in CRC [10, 19, 28]. Interestingly, the identification of non-secreted and nuclear pathways, such as MYC targets and mitotic spindle components, in plasma via Mag-Net aligns with growing evidence that extracellular vesicles selectively encapsulate and transport intracellular oncogenic signaling molecules into circulation [34-37].

Importantly, many top ranked proteins in our data were previously reported as colon cancer enriched proteins in tissue (CPTAC [26]) or plasma (HPA [27]). Additional support for the biomarker relevance of the identified candidates was provided through integration with external plasma EV proteomic datasets [28, 29, 37]. Importantly, nine of the Mag-Net top-20 RF features were quantified in the independent Wang cohort and retained strong collective discriminatory performance with an AUC of 0.932. Together,

these observations indicate that clinically relevant disease-associated information can be recovered using both global and enrichment-based plasma proteomics strategies and highlight the value of integrating multiple proteomics resources to contextualize candidate biomarkers and prioritize targets for future validation studies.

We note four main limitations of the current study. First, this is not a comprehensive comparison of all state-of-art plasma proteomics sample preparation methods [38, 39], but rather focused on affordable ones that may be more promising for large scale studies. Second, the clinical cohort remains relatively modest in size, and larger independent cohorts will be required to establish the robustness and generalizability of the identified biomarkers. Third, despite its widespread clinical use, carcinoembryonic antigen (CEA) was not detected in the ISD or Mag-Net workflows [40]. This observation likely reflects a combination of factors, including low circulating abundance and extensive glycosylation [41]. Fourth, while concordance with external proteomic resources provides supporting evidence, prospective validation using orthogonal methodologies will be required before clinical implementation.

## Conclusions

In summary, this study provides both a methodological and biological framework for plasma proteomics-based biomarker discovery. By systematically comparing whole-plasma and EV-enriched workflows and subsequently applying both approaches to a clinical CRC cohort, we demonstrate that these strategies should be viewed as complementary approaches. Future studies may benefit from integrating whole-plasma

and EV-enriched measurements within the same analytical workflow, allowing complementary biological layers to be interrogated simultaneously.

## Abbreviations

ABC: Ammonium bicarbonate

AUC: Area under the curve

BP: Biological process

CC: Cellular component

CPTAC: Clinical Proteomic Tumor Analysis Consortium

CRC: Colorectal cancer

CV: Coefficient of variation

DIA: Data-independent acquisition

DTT: Dithiothreitol

ECM: Extracellular matrix

EV: Extracellular vesicle

FA: Formic acid

FDR: False discovery rate

GO: Gene Ontology

HLB: Hydrophilic–lipophilic balance

HPA: Human Protein Atlas

IAA: Iodoacetamide

ISD: In-solution digestion

LC–MS/MS: Liquid chromatography–tandem mass spectrometry

MF: Molecular function

mTORC1: Mechanistic target of rapamycin complex 1

PCA: Principal component analysis

perCA: Perchloric acid precipitation

PIC: Protease inhibitor cocktail

RF: Random forest

ROC: Receiver operating characteristic

SAX: Strong anion exchange

S-Trap: Suspension trapping

ssGSEA: Single-sample Gene Set Enrichment Analysis

TFA: Trifluoroacetic acid

## Declarations

### Ethics approval and consent to participate

Plasma samples of 30 CRC patients and 29 healthy controls were used to explore the potential of the best performing EV and global plasma proteomics workflows to identify candidate biomarkers in CRC. EDTA plasma was provided by the Liquid Biopsy center of the Amsterdam UMC, and the application was reviewed by the scientific and ethical committee (043-LBC-2022: 2022.0432 and 2022.0246).

## Consent for publication

Not applicable

## Availability of data and materials

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [42] partner repository with the dataset identifier PXD084723.

## Competing interests

The authors declare no competing interests.

## Authors' contributions

CRJ designed the study. JL and CAM performed the experiments. JL conducted the downstream data analysis, interpreted the results, and drafted the manuscript. BG performed the mass spectrometry measurements. TVP processed the raw mass spectrometry data using DIA-NN. All authors reviewed and approved the final manuscript.

## Funding

Jianing Liu is supported by a PhD scholarship from the China Scholarship Council (CSC).

## Acknowledgements

Jianing Liu is supported by the China Scholarship Council (CSC). Amsterdam UMC is acknowledged for providing the proteomics infrastructure.

## References

1. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst. 2016;2:185–95. https://doi.org/10.1016/j.cels.2016.02.015

2. Vitko D, Chou W-F, Nouri Golmaei S, Lee J-Y, Belthangady C, Blume J, et al. timsTOF HT Improves Protein Identification and Quantitative Reproducibility for Deep Unbiased Plasma Protein Biomarker Discovery. J Proteome Res. 2024;23:929–38. https://doi.org/10.1021/acs.jproteome.3c00646

3. Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen JV. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. J Proteome Res. 2018;17:727–38. https://doi.org/10.1021/acs.jproteome.7b00602

4. Kardell O, von Toerne C, Merl-Pham J, König A-C, Blindert M, Barth TK, et al. Multicenter Collaborative Study to Optimize Mass Spectrometry Workflows of Clinical Specimens. J Proteome Res. 2024;23:117–29. https://doi.org/10.1021/acs.jproteome.3c00473

538 5. Niu L, Thiele M, Geyer PE, Rasmussen DN, Webel HE, Santos A, et al. Noninvasive
539 proteomic biomarkers for alcohol-related liver disease. Nat Med. 2022;28:1277–87.
540 https://doi.org/10.1038/s41591-022-01850-y

541 6. Keshishian H, Burgess MW, Specht H, Wallace L, Clauser KR, Gillette MA, et al.
542 Quantitative, multiplexed workflow for deep analysis of human blood plasma and
543 biomarker discovery by mass spectrometry. Nat Protoc. 2017;12:1683–701.
544 https://doi.org/10.1038/nprot.2017.054

545 7. Woo J, Zhang Q. A Streamlined High-Throughput Plasma Proteomics Platform for
546 Clinical Proteomics with Improved Proteome Coverage, Reproducibility, and
547 Robustness. J Am Soc Mass Spectrom. 2023;34:754–62.
548 https://doi.org/10.1021/jasms.3c00022

549 8. Thanou E, Koopmans F, Pita-Illobre D, Klaassen RV, Özer B, Charalampopoulos I,
550 et al. Suspension TRAPping Filter (sTRAP) Sample Preparation for Quantitative
551 Proteomics in the Low µg Input Range Using a Plasmid DNA Micro-Spin Column:
552 Analysis of the Hippocampus from the 5xFAD Alzheimer's Disease Mouse Model.
553 Cells. 2023;12:1242. https://doi.org/10.3390/cells12091242

554 9. Viode A, van Zalm P, Smolen KK, Fatou B, Stevenson D, Jha M, et al. A simple,
555 time- and cost-effective, high-throughput depletion strategy for deep plasma
556 proteomics. Sci Adv. 2023;9:eadf9717. https://doi.org/10.1126/sciadv.adf9717

557 10. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of

extracellular vesicles. Nat Rev Mol Cell Biol. 2018;19:213–28. https://doi.org/10.1038/nrm.2017.125

11. Wu CC, Tsantilas KA, Park J, Plubell D, Sanders JA, Naicker P, et al. Enrichment of extracellular vesicles using Mag-Net for the analysis of the plasma proteome. Nat Commun. 2025;16:5447. https://doi.org/10.1038/s41467-025-60595-7

12. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020;17:41–4. https://doi.org/10.1038/s41592-019-0638-x

13. Yu F, Teo GC, Kong AT, Fröhlich K, Li GX, Demichev V, et al. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. Nat Commun. 2023;14:4154. https://doi.org/10.1038/s41467-023-39869-5

14. Pham TV, Henneman AA, Jimenez CR. iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. Bioinformatics. 2020;36:2611–3. https://doi.org/10.1093/bioinformatics/btz961

15. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3. https://doi.org/10.2202/1544-6115.1027

16. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25. https://doi.org/10.1016/j.cels.2015.12.004

578 17. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for
579 microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
580 https://doi.org/10.1186/1471-2105-14-7

581 18. Breiman L. Random Forests. Machine Learning. 2001;45:5–32.
582 https://doi.org/10.1023/A:1010933404324

583 19. Ye Y, Fang J, Wang Y, Xu B, Li Z, Chang L, et al. YWHAE affects proliferation,
584 migration and apoptosis of colorectal cancer by regulating extracellular vesicles
585 secretion and Wnt/β-catenin signaling pathway. Transl Cancer Res. 2025;14:2260–73.
586 https://doi.org/10.21037/tcr-24-1910

587 20. Kalantari E, Taheri T, Fata S, Abolhasani M, Mehrazma M, Madjd Z, et al.
588 Significant co-expression of putative cancer stem cell markers, EpCAM and CD166,
589 correlates with tumor stage and invasive behavior in colorectal cancer. World J Surg
590 Oncol. 2022;20:15. https://doi.org/10.1186/s12957-021-02469-y

591 21. Senousy MA, Shaker OG, Ayeldeen G, Radwan AF. Association of lncRNA MEG3
592 rs941576 polymorphism, expression profile, and its related targets with the risk of
593 obesity-related colorectal cancer: potential clinical insights. Sci Rep. 2024;14:10271.
594 https://doi.org/10.1038/s41598-024-60265-6

595 22. Gonulcu SC, Unal B, Bassorgun IC, Ozcan M, Coskun HS, Elpek GO. Expression
596 of Notch pathway components (Numb, Itch, and Siah-1) in colorectal tumors: A
597 clinicopathological study. World J Gastroenterol. 2020;26:3814–33.

598 https://doi.org/10.3748/wjg.v26.i26.3814

599 23. Maffeis V, Nicolè L, Cappellesso R. RAS, Cellular Plasticity, and Tumor Budding
600 in Colorectal Cancer. Front Oncol. 2019;9:1255.
601 https://doi.org/10.3389/fonc.2019.01255

602 24. Patmore S, Dhami SPS, O'Sullivan JM. Von Willebrand factor and cancer;
603 metastasis and coagulopathies. Journal of Thrombosis and Haemostasis.
604 2020;18:2444–56. https://doi.org/10.1111/jth.14976

605 25. Théry C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R,
606 et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a
607 position statement of the International Society for Extracellular Vesicles and update of
608 the MISEV2014 guidelines. J Extracell Vesicles. 2018;7:1535750.
609 https://doi.org/10.1080/20013078.2018.1535750

610 26. Clinical Proteomic Tumor Analysis Consortium (CPTAC) | NCI Genomic Data
611 Commons [Internet]. https://gdc.cancer.gov/about-gdc/contributed-genomic-data-
612 cancer-research/clinical-proteomic-tumor-analysis-consortium-cptac. Accessed 24 July
613 2024

614 27. The human blood atlas - The Human Protein Atlas.
615 https://www.proteinatlas.org/humanproteome/blood. Accessed 16 June 2025

616 28. Wang J, Gu C-Z, Wang P-X, Xian J-R, Wang H, Shang A-Q, et al. Integrative
617 proteomic profiling of tumor and plasma extracellular vesicles identifies a diagnostic

618 biomarker panel for colorectal cancer. Cell Rep Med. 2025;6:102090.
619 https://doi.org/10.1016/j.xcrm.2025.102090

620 29. Zhang J, Gao Z, Xiao W, Jin N, Zeng J, Wang F, et al. A simplified and efficient
621 extracellular vesicle-based proteomics strategy for early diagnosis of colorectal cancer.
622 Chem Sci. 2024;15:18419–30. https://doi.org/10.1039/d4sc05518g

623 30. Järvinen E, Liu X, Varjosalo M, Keskitalo S. Automated Mag-Net Enrichment
624 Unlocks Deep and Cost-Effective LC-MS Plasma Proteomics. J Proteome Res.
625 2026;25:618–32. https://doi.org/10.1021/acs.jproteome.5c00420

626 31. Peinado H, Alečković M, Lavotshkin S, Matei I, Costa-Silva B, Moreno-Bueno G,
627 et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-
628 metastatic phenotype through MET. Nat Med. 2012;18:883–91.
629 https://doi.org/10.1038/nm.2753

630 32. Wei P, Wu F, Kang B, Sun X, Heskia F, Pachot A, et al. Plasma extracellular vesicles
631 detected by Single Molecule array technology as a liquid biopsy for colorectal cancer.
632 J Extracell Vesicles. 2020;9:1809765. https://doi.org/10.1080/20013078.2020.1809765

633 33. Bao S, Li Y, Liang Y, Dai J, Huang H, Ma B. A novel metabolite-interacting protein
634 (MIP)-based molecular subtyping construction and validation of IGFBP3 as a MIP-
635 related oncogene in colorectal cancer. Genes Dis. 2024;12:101272.
636 https://doi.org/10.1016/j.gendis.2024.101272

637 34. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic

638 Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell.
639 2019;177:1035-1049.e19. https://doi.org/10.1016/j.cell.2019.03.030

640 35. Urbiola-Salvador V, Jabłońska A, Miroszewska D, Kamysz W, Duzowska K,
641 Drężek-Chyła K, et al. Mass Spectrometry Proteomics Characterization of Plasma
642 Biomarkers for Colorectal Cancer Associated With Inflammation. Biomark Insights.
643 2024;19:11772719241257739. https://doi.org/10.1177/11772719241257739

644 36. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of
645 exosomes. Science. 2020;367:eaau6977. https://doi.org/10.1126/science.aau6977

646 37. Mohamedali A, Heng B, Amirkhani A, Krishnamurthy S, Cantor D, Lee PJM, et al.
647 A Proteomic Examination of Plasma Extracellular Vesicles Across Colorectal Cancer
648 Stages Uncovers Biological Insights That Potentially Improve Prognosis. Cancers
649 (Basel). 2024;16:4259. https://doi.org/10.3390/cancers16244259

650 38. Beimers WF, Overmyer KA, Sinitcyn P, Lancaster NM, Quarmby ST, Coon JJ.
651 Technical Evaluation of Plasma Proteomics Technologies. J Proteome Res.
652 2025;24:3074–87. https://doi.org/10.1021/acs.jproteome.5c00221

653 39. Katsavelis D, van der Hart MGC, Wolters JC, Permentier HP, Horvatovich P,
654 Cremers TIFH. Impact of Sample Preparation Strategies on the Quantitative Accuracy
655 of Low-Abundance Serum Proteins in Shotgun Proteomics. J Proteome Res.
656 2025;24:4792–803. https://doi.org/10.1021/acs.jproteome.5c00572

657 40. Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: is it

clinically useful? Clin Chem. 2001;47:624–30. https://doi.org/10.1093/clinchem/47.4.624

41. Illiano A, Pinto G, Melchiorre C, Carpentieri A, Faraco V, Amoresano A. Protein Glycosylation Investigated by Mass Spectrometry: An Overview. Cells. 2020;9:1986. https://doi.org/10.3390/cells9091986

42. Perez-Riverol Y, Bandla C, Kundu DJ, Kamatchinathan S, Bai J, Hewapathirana S, et al. The PRIDE database at 20 years: 2025 update. Nucleic Acids Res. 2025;53:D543–53. https://doi.org/10.1093/nar/gkae1011

**Additional information**

Additional file 1 (PPTX). Supplementary Figures 1-3. Comparisons of CRC-associated protein alterations identified using Mag-Net and ISD with the CPTAC colorectal tissue dataset (Supplementary Figure 1) and the HPA plasma dataset (Supplementary Figure 2), and cross-study comparisons of CRC-associated plasma EV protein signatures (Supplementary Figure 3).

Additional file 2 (DOCX). Supplementary Table 1. Biological roles and reported cancer relevance of the top 20 random forest-ranked proteins identified in the Mag-Net and ISD datasets.

Additional file 3 (XLSX). Supplementary Table 2. Characteristics of the three published plasma EV proteomics studies included in the cross-study comparison, including cohort composition, sample sizes, EV isolation methods, analytical platforms, and data

679 processing software.

680 Additional file 4 (XLSX). Supplementary Table 3. Cross-study comparison of CRC-
681 associated proteins identified in the Mag-Net cohort. Separate worksheets summarize
682 proteins with increased or decreased abundance in CRC, together with their detection,
683 direction of change, and statistical significance in the three external studies.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1SupplementaryFigure13.pptx
- Additionalfile1SupplementaryTable3.xlsx
- Additionalfile1SupplementaryTable1.docx
- Additionalfile1SupplementaryTable2.xlsx