Volume 16 | Issue 1 | July 2026 ISSN 2182-0287

Journal of Integrated OMICS

a methodological journal

JOURNAL OF INTEGRATED OMICS

A methodological Journal

Review Article

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

REVIEW ARTICLE | DOI: 10.5584/jiomics.v16i1.251

Bloodstain Proteomics as a Molecular Clock for Estimating Time Since Deposition, Biological Age, and Sex in Forensic Investigations

Raquel Fonseca [1,2], Inês F. Domingos [1,2], Hugo Miguel Santos [1,2], José Luis Capelo [1,2,*]

[1] OMICS and Analytical Development Group, BIOSCOPE Research Group, LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. [2] PROTEOMASS Scientific Society, 2825-466 Caparica, Portugal.

Available Online: July 2026

ABSTRACT

Blood is one of the most informative biological traces recovered from crime scenes. Beyond conventional DNA profiling, its proteome may provide contextual information concerning when a stain was deposited and broad biological characteristics of the donor. This critical review examines three interconnected applications of forensic blood proteomics: time since deposition estimation, biological-age estimation and biological-sex estimation. We evaluate evidence from haemoglobin oxidation, plasma ageing signatures and sex-associated proteins; examine environmental, substrate and donor-related confounders; and compare the statistical and machine-learning approaches used for prediction. Six proteins, ITIH3, HSPA9, SNAP91, FTL, XPO4 and NCF2, currently provide the strongest replicated feature-selection evidence for age association in dried bloodstains. We further propose a nanoLC-timsTOF diaPASEF workflow for future validation, combining trapped-ion mobility, rigorous quality control, and multi-output models to simultaneously estimate several forensic attributes. Although a unified assay has not yet been validated in a single representative bloodstain cohort, the available evidence supports further development of proteomics as a complement to DNA profiling for crime-scene reconstruction.

1. Introduction

Blood is the most frequently encountered biological trace at violent crime scenes and represents one of the most information-rich substrates available to forensic investigators [1,2]. Traditionally, bloodstain analysis has served two primary purposes: confirming that the substance is human blood and identifying its donor through DNA short tandem repeat (STR) profiling. Although conventional DNA profiling primarily addresses individualisation and can provide sex-associated information when sufficient DNA is recovered, it generally provides little or no information about time since deposition (TSD) or donor age, and it may fail when traces are highly degraded or environmentally compromised [3,4]. These gaps have motivated growing interest in complementary molecular approaches. RNA-based methods have been explored for body-fluid identification and for TSD estimation by exploiting measurable changes in transcript abundance and integrity in dried stains [5,6]. Spectroscopic methods, particularly Raman, FTIR, and UV-Vis spectroscopy, monitor the characteristic colour change from oxyhaemoglobin to methaemoglobin and haemichrome as a proxy for elapsed time [7-9]. Metabolomics approaches using high-resolution liquid chromatography-mass spectrometry (LC-HR-MS/MS) have identified circadian and time-varying metabolites in bloodstains [10,11]. However, none of these approaches has achieved widespread implementation in routine forensic practice because of persistent challenges involving reproducibility across diverse conditions, limited standardisation and validation, insufficient reference datasets, and the absence of integrated models addressing multiple forensic objectives simultaneously [12]. Proteomics, the large-scale, quantitative study of proteins, offers several advantages that make it particularly well suited to address these limitations. Proteins are generally more chemically stable than RNA and more informationally diverse than haemoglobin-centred spectroscopic targets [13,14]. The plasma proteome contains thousands of proteins whose abundances are associated with age [15,16], influenced by biological sex and hormonal status [17], and subject to systematic, time-dependent changes after desiccation [18]. Modern high-resolution mass spectrometry platforms operating in data-independent acquisition (DIA) mode can quantify hundreds to more than 1000 proteins from small dried -bloodstain samples, depending on sample quality and workflow depth [19-21]. This review synthesises the current evidence for proteomics-based forensic analysis of blood across three questions: (1) When was this blood deposited? (2) How old was the donor?

***Corresponding author:** José Luis Capelo, jlcm@fct.unl.pt

and (3) What is the biological sex of the donor? We describe the analytical workflows required, catalogue the protein markers identified so far, evaluate the machine learning models used to generate predictions, discuss confounders, and propose a roadmap toward integrated, multi-parameter forensic proteomics.

1.1. Scope and Literature Approach

This critical review synthesises peer-reviewed literature on proteomic and complementary molecular approaches to bloodstain analysis, with emphasis on time since deposition, donor age and biological sex estimation. Relevant publications available up to June 2026 were identified through targeted searches of major bibliographic databases, citation chaining and examination of key forensic-proteomics reviews. Priority was given to studies using human blood or bloodstains, quantitative mass spectrometry, independently assessed prediction models, and experimentally characterised environmental or substrate effects. Studies using other matrices or instrumental configurations were included when they informed analytical feasibility, marker biology or workflow design. Because the evidence base remains heterogeneous and limited in cohort size, the review is critical and narrative rather than a formal systematic review or meta-analysis.

2. Analytical Workflow for Dried Bloodstain Proteomics

2.1. Sample Collection and Extraction

The typical starting material for forensic bloodstain proteomics is a disc punched from a dried bloodstain deposited on a substrate such as fabric, tile, wood, paper or glass. Protein yields depend strongly on stain age, substrate porosity and environmental history. Extraction commonly uses 8M urea or 6M guanidinium hydrochloride in Tris-HCl buffer (pH 8.0), followed by sonication to maximise protein solubilisation from the stain matrix [18,21]. A critical consideration is substrate interference: fabrics in particular contribute abundant keratin peptides that can dominate the spectral space and suppress detection of low-abundance proteins. Strategies to mitigate this include protein precipitation (methanol/chloroform or acetone-based), filter-aided sample preparation (FASP), and short-gel electrophoretic pre-fractionation. The input protein amount recoverable from a standard 6 mm punch of a bloodstain typically falls in the range of 1-20 µg depending on stain load, which is compatible with current nano-flow LC-MS/MS systems but constrains the depth of proteome coverage achievable [19,22].

2.2. Proteolytic Digestion

Following protein extraction, samples may be processed by conventional in-solution digestion, filter-aided sample preparation (FASP), or ultrasound-assisted FASP (US-FASP). For in-solution digestion, proteins are reduced with dithiothreitol (10 mM, 56 degrees ^{o}C, 30 min), alkylated with iodoacetamide (20 mM, room temperature, 20 min in the dark), and digested overnight at 37 degrees ^{o}C with sequencing-grade trypsin at an enzyme-to-substrate ratio of approximately 1:50 (*w*/*w*). Alternatively, FASP combines protein digestion with detergent removal and sample clean-up [22]. The protein extract is loaded onto a centrifugal ultrafiltration device, typically with a 30 kDa molecular-weight cut-off, and washed with 8 M urea in Tris buffer. Proteins retained on the membrane are reduced and alkylated, followed by buffer exchange into ammonium bicarbonate or triethylammonium bicarbonate and digestion with sequencing-grade trypsin. US-FASP is a faster alternative in which ultrasonic energy accelerates reduction, alkylation and enzymatic digestion, substantially shortening sample-preparation time while retaining the clean-up advantages of conventional FASP [23]. The resulting peptides are recovered by centrifugation and additional membrane-washing steps. Trypsin predominantly cleaves peptide bonds on the C-terminal side of arginine and lysine residues, except when followed by proline, generating peptides suitable for reversed-phase chromatography and electrospray ionisation. Peptide mixtures obtained with any of these workflows are subsequently desalted using C18 StageTips, micro-solid-phase extraction devices or conventional solid-phase extraction before LC-MS/MS analysis.

2.3. LC-MS/MS Acquisition Strategies

Two broad acquisition strategies are employed in forensic bloodstain proteomics. Data-dependent acquisition (DDA) selects abundant precursor ions for fragmentation in real time. Although it generates high-quality tandem mass spectra and is useful for spectral-library construction, intensity-dependent precursor selection may undersample low-abundance peptides and reduce consistency across analytical runs [20]. On trapped ion mobility spectrometry-time-of-flight (timsTOF) platforms, DDA can be implemented as ddaPASEF, which combines trapped ion mobility separation with parallel accumulation-serial fragmentation to increase sequencing speed and sensitivity, while remaining dependent on precursor selection [24]. Data-independent acquisition (DIA), in which precursor ions within predefined m/z windows are systematically co-isolated and fragmented, provides more comprehensive and reproducible quantification across samples and is therefore particularly attractive for forensic bloodstain studies [19,20]. On timsTOF instruments, DIA can be combined with PASEF in the diaPASEF acquisition mode. This strategy synchronises precursor isolation with ion-mobility separation, adding inverse reduced ion mobility ($1/K_0$) to retention time and m/z, reducing spectral interference and improving sensitivity, quantitative accuracy and reproducibility [25]. These characteristics make diaPASEF a promising option for small or degraded forensic bloodstains, although its application to bloodstain ageing and multi-attribute prediction remains underexplored. The Thermo Scientific Orbitrap Astral, introduced in 2023, has further expanded DIA capabilities through increased acquisition speed, sensitivity and proteome depth. In a recent forensic pilot study, Gao et al. used Orbitrap Astral DIA to analyse bloodstains from 40 male donors stored for four years and identified 1655 proteins, of which 71 were associated with donor age and subsequently evaluated for age-prediction modelling [18].

2.4. Forensic Potential of Ion-Mobility-Enabled Mass Spectrometry

Although LC-diaPASEF has not yet been systematically validated for multi-attribute bloodstain profiling, the forensic applicability of the timsTOF platform has been demonstrated through its MALDI capabilities. Heaton *et al.* used a timsTOF fleX for MALDI profiling, tandem mass spectrometry and imaging of haemoglobin variants in human blood fingermarks, correctly assigning eight of nine samples in a blind set and mapping blood fingermarks at 50 micrometre spatial resolution [26]. Kennedy *et al.* used MALDI MSP and MSI to detect and image human blood in fingermarks on simulated painted walls after Ninhydrin or Acid Black-1 enhancement, confirming haem and haemoglobin peptide markers across depletion series and even detecting blood beneath paint despite polymer interference. Subsequent STR analysis showed DNA profiling remained feasible after MALDI MSI, demonstrating the incorporation capabilities into standard forensic laboratorial analyses [27]. Kennedy *et al.* also applied bottom-up MALDI proteomics to distinguish animal-blood species and provenance, using timsTOF fleX MALDI-MS/MS to confirm markers differentiating blood from wounded animals and packaged meat [28]. Igoh et al. also used a microTOF II for peptide mass fingerprinting and mascot search in order to screen animal species through the analysis of haemoglobin subunits peptides, in both fresh and dried blood samples [29]. In 2025, Kuckova *et al.* applied LC-MS/MS analysis to identify blood present in historical artworks taken from Japanese and Chinese lacquer artefacts dating from the 18th and 19th centuries using an ESI-Q/TOF Maxis Impact, advancing conservation science and addressing protein database limitations [30]. These studies support the broader forensic potential of ion-mobility-enabled instrumentation, while not constituting direct validation of LC-diaPASEF for TSD, age or sex prediction.

2.5. Confounding Variables

A recurring challenge across all three forensic applications reviewed here is the influence of environmental conditions on the protein composition and stability of dried bloodstains. Temperature is the most potent confounder: protein degradation rates increase substantially at elevated temperatures, and TSD models calibrated at 20 °C perform poorly when applied to stains aged at 37 °C [12,31]. Relative humidity affects the oxidation kinetics of haemoglobin, specifically the rate of conversion of methaemoglobin to haemichrome is humidity-dependent, while the initial oxyhaemoglobin-to-methaemoglobin conversion is relatively humidity-independent [9]. UV light exposure and substrate chemical composition introduce additional variability. A combined model trained across all environmental conditions typically achieves markedly lower R^2 than condition-specific models, underscoring the need for environmental metadata in forensic proteomics databases [31]. The main confounding variables are found in **Table 1**.

Table 1 | Principal confounding variables affecting forensic bloodstain proteomics and their implications for predictive-model design.

Variable	Effect on proteome	Relevance for forensic model
Temperature	Accelerates protein denaturation and degradation; shifts TSD kinetics	Must be recorded or estimated from scene; strong confounder for TSD
Humidity	Controls metHb → haemichrome rate; affects peptide solubility on re-extraction	Affects haemoglobin-based TSD but less critical for broad proteome models
Substrate	Introduces exogenous proteins (keratin from fabric); affects extraction yield	Substrate-aware clean-up or interference-control procedures may be required; substrate type should be annotated
UV/light	Photo-oxidation of tryptophan and methionine residues	Indoor vs. outdoor stains require separate calibration models
Donor variability	Individual plasma proteome differs by health status, medication, diet	Cohort diversity essential in training sets; N=40 (Gao *et al.*) still small

3. Time Since Deposition: Blood as a Molecular Clock

3.1. The Haemoglobin Oxidation Cascade

The most studied molecular marker for TSD in blood is haemoglobin, which constitutes approximately 90% of the dry weight of erythrocytes and therefore dominates the dried bloodstain proteome by mass. Upon exit from the vascular system, oxyhaemoglobin (HbO_2, Fe^{2+}, bright red) undergoes biphasic oxidation: a rapid first phase converts HbO_2 to methaemoglobin (metHb, Fe^{3+}, dark red-brown) within hours, followed by a slower second phase converting metHb to haemichrome (HC, dark brown), in which the haem iron coordinates to an additional amino acid ligand [7,9].

The kinetics of the HbO_2 → metHb → HC cascade are strongly temperature-dependent but the metHb → HC transition is additionally sensitive to humidity [9]. These optical changes underlie UV-Vis and Raman spectroscopy-based TSD approaches, which have been reviewed extensively [32,33]. Electrochemical methods, measuring the redox potential of haem species by differential pulse voltammetry, have also been applied over a two-week time series in five temperature conditions, identifying seven of twenty-seven voltammetric variables significantly correlated with TSD ($p < 0.033$) [34]. Although powerful and instrumentation-light, spectroscopic and electrochemical methods are necessarily anchored to a single analyte (haemoglobin) and cannot simultaneously address age or sex.

3.2. Proteome-Wide TSD Markers

The first systematic proteomics-based TSD study was reported by Schneider *et al.* in 2022 using bottom-up DIA LC-MS/MS on dried bloodstains sampled at multiple time points over a two-month period [19]. Principal component analysis (PCA) of the quantified protein matrix showed clear clustering of samples according to TSD. Partial least-squares discriminant analysis (PLS-DA) further confirmed that the proteome composition - not only haemoglobin - encodes temporal information. Key proteins showing systematic abundance changes included components of the complement system, coagulation cascade proteins, and acute-phase reactants, reflecting the progressive degradation and structural reorganisation of the dried blood matrix over time. This work established a critical proof of concept: TSD estimation from blood is not limited to haemoglobin spectroscopy but can be approached with the full proteomic toolkit, opening the door to simultaneous multi-attribute modelling. The two-month observation window, however, leaves longer TSD ranges, relevant to cold cases, unaddressed, and the single-donor, single-temperature study design constrains generalisability [19]. Metabolomics-based TSD work (using LC-HR -MS/MS) has complemented these findings by identifying 28 metabolites with significant circadian oscillations in bloodstains, enabling estimation not only of elapsed time but potentially of time -of-day of deposition [10]. Transcriptomics-based approaches show that mRNA degradation rates support TSD estimation within 2-6 weeks for indoor stains, with accuracy degrading significantly beyond this window [5,6]. Together, these data suggest that a multi -omic approach such as proteomics for longer-range TSD, metabolomics for within-day resolution, and RNA metrics for short -range accuracy, may ultimately be most powerful.

3.3. Protein Degradation Signatures and Chemical Modifications

Changes in protein abundance represent only one component of bloodstain ageing. Following deposition, proteins also undergo a series of spontaneous and environmentally driven chemical modifications that reflect progressive molecular degradation. These include oxidation of susceptible amino acid residues, deamidation, backbone fragmentation and the generation of semi-tryptic peptides, all of which have been recognised as characteristic features of protein ageing in proteomics studies [35,19]. Oxidation is among the earliest detectable modifications in dried blood proteins. Methionine and tryptophan residues are particularly susceptible to reactive oxygen species generated during prolonged exposure to atmospheric oxygen, light and environmental stressors, leading to predictable increases in oxidised peptide species over time [36]. Deamidation of asparagine and glutamine residues proceeds more gradually and has been investigated as a molecular indicator of protein ageing in archaeological, palaeoproteomic and forensic applications, although its dependence on temperature, humidity and protein structure currently limits its use as an independent chronological marker [37]. Progressive protein degradation may also generate endogenous cleavage products that appear as semi-tryptic peptides following bottom-up analysis. Unlike fully tryptic peptides produced during enzymatic digestion, these species reflect cleavage events that occurred before sample preparation and may therefore provide additional information regarding bloodstain ageing [19]. Liu *et al.* have found 4 molecular modifications (succinylation and phosphorylation modification of cystein and the modification of lysine with threonine) with significantly higher values in the old group than the young group, while the carbamylation of lysine was lower in the young group, suggesting an increase in modified proteins at older ages, causing changes in the function of those proteins [38]. DIA workflows are well suited to monitor these molecular changes because modified and unmodified peptide species can be quantified reproducibly across large sample cohorts. Recent advances in library-free database searching and prediction-based spectral matching have further improved the identification of modified peptides without requiring experimentally generated spectral libraries, facilitating more comprehensive characterisation of protein degradation products [39,40]. Consequently, future forensic workflows may benefit from integrating peptide-level degradation signatures with conventional protein abundance measurements to improve the robustness of time-since-deposition models across diverse environmental conditions. Although relatively few forensic studies have systematically evaluated post-translational modifications as direct predictors of TSD, their biochemical basis and compatibility with modern LC–MS/MS workflows make them an important direction for future research. Validation across larger donor cohorts, longer ageing periods and environmentally diverse conditions will be required to determine whether these modifications provide predictive information beyond that obtained from protein abundance alone.

3.4. Machine Learning Models for TSD

TSD estimation has been approached with several statistical frameworks. Linear mixed models have been used in electrochemical studies, treating the biological replicate (individual donor) as a random effect to improve model fit [34]. For proteomics data, dimensionality reduction followed by multivariate regression is more common, with PLS-DA scores along the first latent variable showing linear correlation with elapsed time in the Schneider study [19]. Colorimetric analyses of bloodstains have employed Random Forest regression, achieving R^2 values of up to 0.92 for individual environmental conditions but only 0.71 when all conditions are pooled, a warning that environmental stratification will be essential in any operational model [31].

4. Biological Age Estimation from Bloodstain Proteomics

4.1. The Plasma Proteome as an Aging Clock

The idea that proteins carry an aging signature predates proteomics. It has long been known that certain plasma proteins - such as albumin, immunoglobulins, and complement components - change in concentration with advancing age. The emergence of large-scale mass spectrometry-based plasma proteomics has transformed this insight into a quantitative framework. Studies involving thousands of participants have shown that more than one

-third of detectable plasma proteins change significantly with age [15,41]. Among the most consistently identified aging-associated proteins are GDF15 (a TGF-β superfamily member and hallmark of the senescence-associated secretory phenotype), CXCL9 (an inflammation marker), fetuin-A (hepatokine), and numerous components of the coagulation and complement systems [15,41,42]. Proteomic aging clocks, analogous to the Horvath epigenetic methylation clock, have now been validated in multiple independent cohorts. The ProteAge clock, built on UK Biobank plasma proteomics data, predicts biological age and mortality risk with high accuracy. Critically, proteomic clocks outperform small RNA-based clocks for age prediction when applied to comparable sample sizes [43], an observation directly relevant to the forensic setting where RNA yields from dried stains are often limiting.

4.2. Protein Equalisation

One of the leading challenges in bottom-up proteomics is the presence of highly abundant proteins, which dominate complex biological samples and mask the detection of low-abundance proteins with potential biological or forensic significance [44]. The wide dynamic range of protein concentrations in blood exceeds the analytical capacity of current LC–MS/MS platforms, causing peptides from abundant proteins to preferentially ionize and be selected for fragmentation, thereby reducing proteome coverage and limiting the identification of low-abundance biomarkers [45]. In whole blood, haemoglobin and albumin are two of the most abundant proteins, making them the primary targets for depletion to improve the detection of low-abundance proteins. Several studies have combined abundant protein depletion or equalisation strategies with MALDI-TOF-MS to improve proteomic profiling. Notably, Warden et al, enhanced peptidome profiles with minimal albumin interference and successfully detected the 10.5 kDa protein rKringle5 in plasma after albumin depletion [46]; Fernández *et al.* precipitated disulfide-rich proteins and produced a depleted serum enriched in immunoglobulins [47]; and Araújo *et al.* identified 6 potential protein biomarkers following major protein equalization, consistent with previous reports for bladder cancer [48]. More recently, Domingos *et al.* extended this concept by applying Ultra High-Resolution Quadrupole Time-of-Flight LC-MS/MS analysis to independently investigate both the supernatant and pellet fractions in multiple myeloma samples. This approach highlighted that analysing the high- and low-abundance protein fractions separately provides complementary quantitative proteomic information, with dysregulated proteins from each fraction converging on biological pathways involved in chaperone-mediated autophagy and protein folding [49].

4.3. Proteomic Age Estimation from Dried Bloodstains

The landmark study translating plasma aging proteomics to forensic bloodstains is that of Gao *et al.* [18]. The authors analysed dried bloodstains from 40 healthy male donors spanning an age range of 10-79 years, stored at room temperature for four years, using a Thermo Scientific Orbitrap Astral instrument in DIA mode. A total of 1655 proteins were quantified in bloodstain extracts. Of these, 71 showed significant correlation with donor age (Pearson $|r| > 0.3$, $p < 0.05$): 26 positively correlated and 45 negatively correlated with age. Feature selection was performed using two complementary approaches: LASSO (Least Absolute Shrinkage and Selection Operator) regression, which identified 18 age-associated proteins, and the Boruta algorithm (a Random Forest wrapper), which identified 10. Six proteins appeared in both feature sets, representing the most robustly supported age markers: ITIH3 (inter -alpha-trypsin inhibitor heavy chain 3), HSPA9 (mitochondrial heat shock protein 70), SNAP91 (synaptosomal-associated protein 91), FTL (ferritin light chain), XPO4 (exportin-4), and NCF2 (neutrophil cytosol factor 2). These results are further highlighted in **Table 2**. The best-performing model, a Random Forest trained on the Boruta-selected features, achieved $R^2 = 0.70$ and mean absolute error (MAE) = 9.14 years on an independent test set [18]. These results are promising but must be contextualised. An MAE of approximately nine years is comparable to the current performance of DNA methylation-based forensic age clocks operating on degraded samples, but substantially worse than those achieved with high-quality fresh blood (MAE 2-4 years). The all-male cohort (N = 40) limits inference about sex-related variation in the age proteome, and the four-year storage condition, while realistic for cold-case scenarios, confounds TSD and donor-age effects in a way that future studies must disentangle.

Table 2 | Six proteins identified by both LASSO and Boruta feature selection as age-associated markers in dried bloodstains in Gao *et al.* [18].

Protein	Full name	Direction with age	Biological relevance
ITIH3	Inter-alpha-trypsin inhibitor heavy chain 3	Negative	Acute-phase reactant; hyaluronan matrix stabilisation; decreases with chronic inflammation of aging
HSPA9	Mitochondrial Hsp70 / mortalin	Positive	Mitochondrial chaperone; accumulates with oxidative stress in aging cells
SNAP91	Synaptosomal-associated protein 91	Negative	Involved in clathrin-mediated endocytosis; altered neuronal function with aging
FTL	Ferritin light chain	Positive	Iron storage; elevated with aging; oxidative stress marker
XPO4	Exportin-4	Negative	Nuclear export of eIF5A and Smad3; links to senescence pathways
NCF2	Neutrophil cytosol factor 2	Positive	NADPH oxidase component; increased reactive oxygen species production in aging neutrophils

4.3. Comparisson with other Biological Age Estimation Approaches

DNA methylation clocks (Horvath, GrimAge, PCPhenoAge) represent the current reference standard for forensic age estimation, with MAE values of 3-5 years achieved on fresh blood. Their performance on aged or damaged DNA is, however, substantially worse, and the requirement for bisulphite conversion introduces additional degradation of already compromised material [50]. Using a multiplex methylation SNaPshot assay, Dias et al. developed an age prediction model based on four CpG sites (ELOVL2, FHL2, C1orf132 and TRIM59) that explained 92.5% of the variation in chronological age (adjusted R^2 = 0.925) with a MAD of 4.97 years across combined blood samples from living and deceased individuals [51]. Telomere length analysis, while feasible from bloodstains, provides population-level age trends rather than individual-level predictions and has high inter-individual variability [50]. RNA-based age clocks are impractical for long TSD stains due to RNA instability. Proteomics-based age estimation occupies a complementary niche: proteins are more stable than both RNA and methylated DNA under adverse storage conditions, the approach is compatible with the same LC-MS/MS workflow used for TSD and sex analysis (enabling multi-attribute analysis from a single extraction), and the protein panel can be continuously refined as larger forensic cohorts are studied. The main current limitations are the relatively small training cohorts published to date and the need to account for TSD-related protein changes that may confound age signals.

5. Biological Sex Estimation from Blood Proteomics

5.1. Protein Markers for Biological Sex in Blood

Biological sex leaves a strong imprint on the plasma proteome through several mechanisms: (i) differential expression of Y-chromosome-encoded proteins absent in females; (ii) hormonal regulation of liver-produced secreted proteins; and (iii) immune and inflammatory protein expression differences that are consistent across populations. Testosterone, oestrogen, and their receptors modulate the transcription of hundreds of genes encoding plasma proteins, many of which are quantifiable by LC-MS/MS [17,52]. The most studied sex-specific protein in forensic contexts is amelogenin, which produces two isoforms encoded by the X (AMELX) and Y (AMELY) chromosomes. AMELY detection in enamel peptides has enabled proteomic sex determination from teeth and bone in archaeological remains, and the approach has been validated on heavily degraded material [53]. However, amelogenin is not expressed in blood, making it inapplicable to bloodstain analysis. For blood specifically, the 2025 study by Alex *et al.* provides the most detailed characterisation, translated to **Table 3** [17]. The authors analysed LC-MS/MS proteomes from 100 whole blood samples of known biological sex (50 male, 50 female) and applied gradient-boosting classifiers (with L1 and L2 regularisation, analogous to LASSO and ridge regression). Cross-validated classification achieved high overall accuracy, with pregnancy zone protein (PZP) and ceruloplasmin (CP) peptides among the most discriminating features. PZP is normally expressed in high amounts only during pregnancy but shows residual sex-differential expression in non-pregnant females at steady state [54]; CP is higher in females due to oestrogen-mediated transcriptional upregulation [55].

5.2. Challenges and Limitations

Despite these promising markers, the Alex *et al.* study explicitly cautions that proteomics-based sex estimation from blood faces technical and biological obstacles that may limit its forensic utility [17]. Principal among these is the very large inter-individual variability in plasma protein concentrations - driven by health status, medication use, hormonal contraception, age, and diet - that partially obscures sex-related differences. The dynamic range of the blood proteome spans more than ten orders of magnitude, with highly abundant proteins (haemoglobin, albumin, immunoglobulins) routinely suppressing detection of low-abundance sex-specific markers [56]. An additional concern is the effect of TSD on sex-discriminating proteins. If stain age differentially affects the stability of female- versus male-enriched proteins, a classifier trained on fresh blood will give biased predictions when applied to aged stains. To date, no study has systematically addressed this interaction. The authors of the sex estimation study therefore recommend that forensic proteomics research prioritise source attribution (identifying the body fluid and tissue of origin) and TSD estimation - where proteomic signals

Table 3 | Key proteins proposed for biological-sex estimation in forensic proteomics, with mechanisms and forensic applicability.

Protein	Sex enrichment	Mechanism	Forensic utility
Pregnancy zone protein (PZP)	Female	Oestrogen-regulated; elevated in pregnancy but detectable at lower levels in non-pregnant females	High sensitivity, moderate specificity; confounded by hormonal status
Ceruloplasmin (CP)	Female	Oestrogen upregulates hepatic CP expression via transcription factor binding	Widely quantified in DIA workflows; moderate sex-ratio
Sex hormone-binding globulin (SHBG)	Female > Male	Regulates free testosterone; higher in females and in males with low testosterone	Useful but affected by age and BMI
Prostate-specific antigen (PSA, KLK3)	Male	Tissue-specific to prostate; detectable in blood at low ng/mL range	Useful in aged males; absent in females
Amelogenin-Y (AMELY)	Male (tissue)	Y-chromosome-encoded enamel protein	Validated in bone/tooth; not expressed in blood

are stronger and more reproducible - over phenotypic profiling of donors, where the gain above existing immunological and DNA-based methods is currently modest [17].

6. Integration: Toward a Multi-Attribute Forensic Proteomic Workflow

6.1. Simultaneous Multi-Parameter Prediction

The three forensic attributes reviewed here, SD, donor age and biological se, are not analytically independent. Signals potentially informative of all three attributes may be recoverable from the same proteomic dataset generated from a single LC-MS/MS analysis of a bloodstain extract, although simultaneous prediction has not yet been prospectively demonstrated in one validated bloodstain cohort. This creates both an opportunity and a statistical challenge. A unified workflow could preserve material by producing several investigative outputs from one punched disc. However, potentially informative features may overlap: age-associated proteins such as FTL and HSPA9 may also change with TSD; sex-associated proteins such as PZP may vary with age and hormonal status; and TSD-associated acute-phase proteins may be influenced by both age and sex [13,17,18]. Multi-output machine learning frameworks, such as multi-task Random Forest, multi-response LASSO, or deep learning architectures trained on shared latent representations, offer a principled way to jointly model all three attributes while accounting for their dependencies. Such frameworks have been applied in clinical proteomics but have not yet been deployed in forensic bloodstain research, representing a clear priority for the field.

6.2. Machine Learning Approaches

Several machine learning approaches have been applied across the studies reviewed for forensic blood proteomics. Random Forest regression and classification have been used for age estimation and TSD modelling [18,31], offering robustness to outliers and non-linear relationships while handling high-dimensional proteomics data through feature importance scoring. LASSO and elastic-net regression have been employed for age marker selection and sex classification [17,18], performing simultaneous feature selection and regularisation with interpretable coefficients that map directly to protein contributions. PLS-DA has been applied to visualise proteome-wide separation among predefined TSD groups [19], whereas continuous elapsed-time prediction requires regression approaches such as PLS regression. Gradient boosting methods, including XGBoost and LightGBM with L1/L2 regularisation, have been applied to sex classification [17] and generally outperform Random Forest on tabular data when sample sizes are sufficient. Finally, linear mixed models have been used in electrochemical and RNA-based TSD studies [11,34], and are appropriate when repeated measures from the same donor are available, with donor treated as a random effect. Nevertheless, quantitative proteomics data require extensive preprocessing before statistical modelling to minimise technical variation while preserving biologically meaningful differences. Protein intensities are typically normalised using median, total-ion-current or variance-stabilising approaches to reduce systematic differences arising from sample preparation and LC-MS/MS performance [57]. Missing values, which commonly result from stochastic peptide detection or low-abundance proteins, must be handled carefully because inappropriate imputation can introduce bias into downstream prediction models [58,59]. Frequently applied strategies include filtering proteins with excessive missingness, imputing values based on local similarity or low-abundance distributions or employing machine-learning algorithms capable of accommodating incomplete datasets directly [59]. Prior to model training, protein identifications are generally filtered using a peptide- and protein-level false discovery rate (FDR) of 1%, with quantification based on proteotypic peptides consistently detected across samples [60]. Where analyses are performed over extended acquisition periods or across multiple batches, batch-effect correction and quality-control monitoring using pooled reference samples are recommended to ensure analytical reproducibility and minimise instrument-related variation [59,61]. These preprocessing steps are well established in quantitative proteomics and are essential to produce robust forensic prediction models with reproducible performance across laboratories and analytical platforms. For DIA workflows, software platforms such as DIA-NN, and FragPipe implement many of these preprocessing procedures, including interference correction, protein inference, normalisation and false-discovery-rate control, thereby generating quantitative protein matrices suitable for downstream statistical and machine-learning analyses [40,62]. A consistent limitation across all published models is the small size of training cohorts, typically fewer than 100 individuals. Given the high within-population variability of plasma protein concentrations, reliable model training for forensic applications will likely require datasets of several hundred to several thousand individuals. The establishment of open, curated forensic proteomics reference databases - analogous to existing DNA STR frequency databases - is therefore an urgent priority for the field.

6.3. Proposed Workflow

Based on the evidence reviewed, and as depicted in **Figure 1**, we propose the following integrated workflow for multi-attribute forensic bloodstain proteomics using trapped ion mobility spectrometry coupled to high-resolution mass spectrometry:

1. **Scene documentation:** Photograph, measure, and record the stain substrate, estimated stain dimensions, sampling location, and relevant environmental conditions, including temperature, humidity, light exposure, and potential UV exposure history, when determinable.
2. **Stain extraction:** Excise one or more 6 mm discs from the bloodstain. Extract proteins in 8M urea buffer using brief probe or bath sonication. Determine protein concentration using a BCA assay and normalise the input to a constant protein mass whenever sufficient material is available.
3. **Tryptic digestion:** Reduce disulfide bonds, alkylate cysteine residues, and digest proteins with sequencing-grade trypsin.

Desalt the resulting peptides using C18 micro-solid-phase extraction, dry under vacuum, and reconstitute in 0.1% formic acid.

4. **nanoLC-timsTOF analysis:** Separate peptides by nano-flow reversed-phase liquid chromatography at approximately 300 nL/min using a 60-90 min gradient. Analyse peptides using a timsTOF instrument operated in diaPASEF mode, combining high-resolution mass spectrometry with trapped ion mobility separation. Optimise the ion-mobility range, collision-energy ramp, precursor isolation scheme, accumulation time, and duty cycle for blood-derived peptide mixtures.
5. **Ion-mobility-aware data processing and protein quantification:** Process diaPASEF data using software compatible with Bruker trapped ion mobility data, such as DIA -NN, Spectronaut, or FragPipe. Perform either library-free analysis using predicted spectra and ion-mobility values or library-based analysis using a project-specific spectral library generated from pooled reference blood samples analysed by ddaPASEF or diaPASEF. Protein quantification should be based on consistently detected proteotypic peptides, with a minimum of two unique or protein-group-specific peptides per reported protein. Apply appropriate between-run normalisation and control peptide- and protein-level false-discovery rates at 1%.
6. **Four-dimensional feature extraction:** Integrate peptide intensity, retention time, mass-to-charge ratio and inverse reduced ion mobility ($1/K_0$). Where appropriately derived or calibrated, collisional cross-section values may also be used. Ion mobility should serve as an additional identification and quality-control dimension to improve discrimination between co-eluting peptides, reduce interference and assess analytical consistency across stains, substrates and ageing conditions.
7. **Multi-attribute modelling:** Apply pre-validated Random Forest, LASSO, or other suitable multivariate models for time since deposition, donor age, and biological sex in parallel. Models may incorporate protein abundance, selected peptide abundance, and ion-mobility-derived variables. Generate point predictions, class probabilities, and confidence or prediction intervals, and cross-reference outputs to assess internal consistency among the predicted attributes.
8. **Quality control:** Include pooled blood reference samples, extraction blanks, digestion blanks, retention-time or ion-mobility reference standards, and replicate injections. Monitor peptide identifications, protein-group numbers, precursor intensity, mass accuracy, retention-time stability, ion-mobility stability, missing-value rates, and carryover. Samples failing predefined analytical criteria should be flagged or excluded from predictive modelling.
9. **Reporting:** Report point estimates, confidence or prediction intervals, model probabilities, and analytical reliability flags. Relevant limitations should include low protein yield, high levels of missing data, poor digestion efficiency, substrate-related interference, environmental conditions outside the model training range, and ion-mobility or retention-time deviations from quality-control limits.

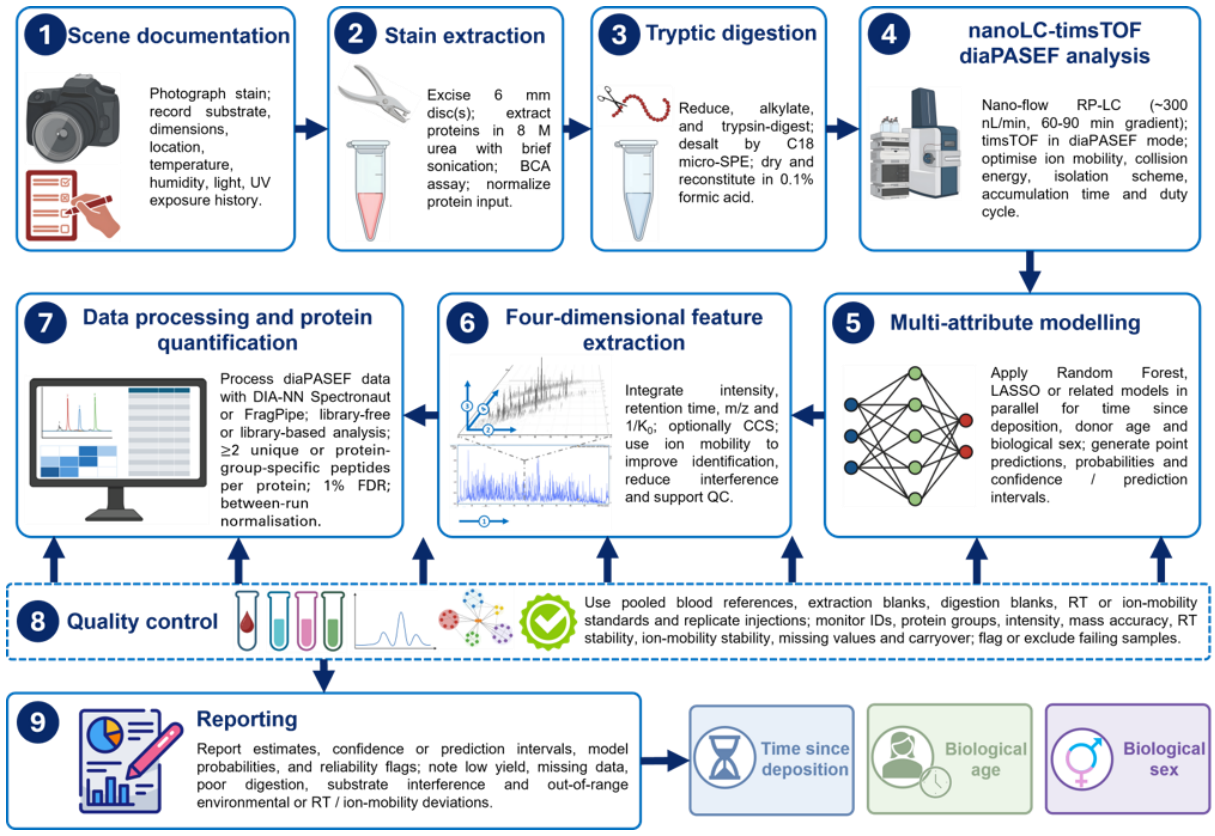

Figure 1 | Proposed workflow for multi-attribute forensic bloodstain proteomics.

7. Current Challenges and Future Directions

Several challenges must be addressed before forensic bloodstain proteomics reaches operational maturity:

7.1. Cohort size and diversity

All published forensic bloodstain proteomics studies have cohort sizes below 100 individuals, and several are limited to a single sex or narrow age range. Larger, population-diverse cohorts are needed, and reference datasets should span a range of health conditions, medications, and ethnicities to avoid systemic bias in predictions.

7.2. Disentangling TSD from donor age

Both TSD and donor biological age affect the bloodstain proteome, but in potentially overlapping ways. A study design that varies both dimensions independently, collecting stains from donors across the age spectrum and ageing each stain across multiple time points, is needed to build models that can deconvolve these two effects.

7.3. Environmental calibration

Protein degradation kinetics are non-linearly dependent on temperature and humidity. Stains recovered from outdoor scenes, vehicles, or tropical climates will have experienced very different environmental histories than laboratory-aged standards. Environmental sensor data co-deposited at scenes (temperature loggers, humidity records) or computational reconstruction from meteorological data could partially address this, but calibration across a wider range of conditions is required.

7.4. Substrate diversity

Most published studies use cotton fabric, glass, or filter paper as substrates. Real crime scenes involve wool, synthetic textiles, carpet, soil-contaminated surfaces, and mixed-substrate stains. Each substrate introduces different extraction efficiencies, different levels of contaminant proteins, and different physical protection of stain proteins from environmental degradation.

7.5. Legal and ethical framework

Inferring age, sex, or phenotypic characteristics from biological traces raises significant legal and ethical questions, particularly regarding the permissible scope of forensic intelligence. Thus, distinguishing information used to generate investigative leads from evidence admissible in court. Regulatory frameworks for forensic DNA phenotyping are still evolving in most jurisdictions, and proteomics-based phenotyping will require equivalent debate and legislation.

7.6. Integration with existing forensic workflows

Forensic laboratories operate under strict accreditation requirements (ISO/IEC 17025:2017), and any new analytical method must be validated according to SWGMAT or equivalent standards before operational use. This includes establishing limits of detection, specificity, precision, accuracy, and robustness across the intended operational range of conditions.

8. Conclusions

Mass spectrometry-based proteomics has emerged as a credible and powerful platform for the multi-attribute forensic analysis of bloodstains. The three principal attributes reviewed here - time since deposition, biological age, and biological sex - are each supported by proof-of-concept evidence and can, in principle, be investigated within a single integrated analytical workflow. In the approach proposed here, the analytical core is based on nanoLC coupled to timsTOF mass spectrometry operating in diaPASEF mode, which combines data-independent acquisition with trapped ion mobility separation. This four-dimensional framework, incorporating mass-to-charge ratio, retention time, ion mobility, and signal intensity, offers increased precursor selectivity, reduced spectral interference, and improved analytical reproducibility for complex and potentially degraded forensic samples. The required informatics infrastructure is also increasingly mature. Software platforms such as DIA-NN, Spectronaut and FragPipe can process diaPASEF data using project-specific spectral libraries or library-free prediction strategies, while established statistical and machine-learning environments support multivariate modelling of forensic attributes. Although the analytical feasibility of deep and reproducible bloodstain proteomics is increasingly well established, a unified diaPASEF-based multi-attribute assay still requires direct experimental validation. Larger and more diverse donor cohorts, systematic assessment of environmental and substrate effects, harmonised quality-control criteria and genuinely multi-attribute study designs are now required. The six proteins consistently associated with donor age in dried bloodstains - ITIH3, HSPA9, SNAP91, FTL, XPO4, and NCF2 - provide a rational starting panel for the development of proteomic age-estimation models. The feasibility of estimating time since deposition using proteome-wide DIA measurements has also been demonstrated, although future studies must extend the validated time range and account for temperature, humidity, UV exposure, substrate type, and other environmental variables. Biological sex estimation using blood proteomics remains less accurate than established genetic approaches, but it may still provide useful complementary information as one component of a broader multi-parameter forensic profile. The addition of trapped ion mobility is particularly relevant for forensic applications because it provides an orthogonal dimension for peptide identification and quality control. Ion-mobility or collisional-cross-section information may improve the discrimination of co-eluting species, increase confidence in peptide assignments, and support the detection of analytical drift or matrix-related interference. However, ion-mobility-derived variables must themselves be rigorously validated across instruments, laboratories, substrates, and ageing conditions before they can be incorporated into forensic prediction models. Forensic proteomics is not currently positioned to replace DNA profiling for individualisation. Its value lies instead in complementing genetic analysis by

providing contextual information about when a bloodstain was deposited and about broad biological characteristics of the donor. A validated timsTOF-based diaPASEF workflow could therefore contribute to crime-scene reconstruction beyond the question of identity. With appropriate investment in reference-cohort development, environmental modelling, inter-laboratory validation, standardisation, and machine-learning integration, multi-attribute bloodstain proteomics may become a practical component of forensic investigation over the coming decade.

Acknowledgements

The authors acknowledge the support of the BIOSCOPE Group, LAQV-REQUIMTE, NOVA School of Science and Technology, NOVA University Lisbon.

Funding

PThis work was funded by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the projects UID/50006/2025, UID/PRR/50006/2025, UID/PRR2/50006/2025, and LA/P/0008/2020 of the Associated Laboratory for Green Chemistry - LAQV-REQUIMTE (https://doi.org/10.54499/UID/50006/2025, https://doi.org/10.54499/UID/PRR/50006/2025, https://doi.org/10.54499/UID/PRR2/50006/2025, and https://doi.org/10.54499/LA/P/0008/2020). H. M. Santos acknowledges that his research contract is funded by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of project LA/P/0008/2020 of the Associated Laboratory for Green Chemistry - LAQV-REQUIMTE (DOI: 10.54499/LA/P/0008/2020). R. F. is funded by the FCT, I.P., PhD grant 2025.05016.BD, and I. F. D. is funded by the FCT/MCTES PhD grant 2024.00745.BD and DOI identifier http://doi.org/10.54499/2024.00745.BD. The PROTEOMASS Scientific Society is acknowledged for the funding provided to the Laboratory for Biological Mass Spectrometry - Isabel Moura through grants PM001/2024, PM001/2019, and PM003/2016.

References

[1] K. Virkler, I.K. Lednev, Forensic Science International 188 (2009) 1–17. DOI: 10.1016/j.forsciint.2009.02.013

[2] M.A. Jobling, P. Gill, Nat Rev Genet 5 (2004) 739–751. DOI: 10.1038/nrg1455

[3] B. Budowle, A. Van Daal, BioTechniques 44 (2008) 603–610. DOI: 10.2144/000112806

[4] M. Kayser, P. De Knijff, Nat Rev Genet 12 (2011) 179–192. DOI: 10.1038/nrg2952

[5] M. Bauer, D. Patzelt, J Forensic Sci 47 (2002) 1278–1282

[6] S. Harbison, R. Fleming, RRFMS (2016) 11. DOI: 10.2147/RRFMS.S57994

[7] G. Edelman, V. Manti, S.M. Van Ruth, T. Van Leeuwen, M. Aalders, Forensic Science International 220 (2012) 239–244. DOI: 10.1016/j.forsciint.2012.03.009

[8] N.U. Kamaruzaman, A.K. Mohd Jamil, M.F. Khyasudeen, K.H. Low, Forensic Science International 378 (2026) 112676. DOI: 10.1016/j.forsciint.2025.112676

[9] R.H. Bremmer, D.M. De Bruin, M. De Joode, W.J. Buma, T.G. Van Leeuwen, M.C.G. Aalders, PLoS ONE 6 (2011) e21845. DOI: 10.1371/journal.pone.0021845

[10] T.D. Schneider, T. Kraemer, A.E. Steuer, Drug Testing and Analysis 15 (2023) 840–852. DOI: 10.1002/dta.3480

[11] F. Cheng, W. Li, Z. Ji, J. Li, W. Hu, M. Zhao, D. Yu, H. Simayijiang, J. Yan, Forensic Science International: Genetics 66 (2023) 102910. DOI: 10.1016/j.fsigen.2023.102910

[12] M.A. Sacco, S. Gualtieri, A.P. Tarallo, L. Calanna, R. La Russa, I. Aquila, IJMS 25 (2024) 7469. DOI: 10.3390/ijms25137469

[13] A. Raj T, A. G.B., A. M, A. E. M., Egypt J Forensic Sci 15 (2025) 75. DOI: 10.1186/s41935-025-00484-8

[14] R.E. Keane, R.J. Tidy, G.J. Parker, J.P.A. Gummer, C. Priddis, WIREs Forensic Science 6 (2024) e1516. DOI: 10.1002/wfs2.1516

[15] B. Lehallier, D. Gate, N. Schaum, T. Nanasi, S.E. Lee, H. Yousef, P. Moran Losada, D. Berdnik, A. Keller, J. Verghese, S. Sathyan, C. Franceschi, S. Milman, N. Barzilai, T. Wyss-Coray, Nat Med 25 (2019) 1843–1850. DOI: 10.1038/s41591-019-0673-2

[16] T. Tanaka, N. Basisty, G. Fantoni, J. Candia, A.Z. Moore, A. Biancotto, B. Schilling, S. Bandinelli, L. Ferrucci, eLife 9 (2020) e61073. DOI: 10.7554/eLife.61073

[17] S. Alex, R. Almey, R.S. Dennis, O. Tytgat, R. Bouwmeester, D. Deforce, M.D. Puit, M. Dhaenens, L.D. Clerck, J. Proteome Res. 24 (2025) 4888–4899. DOI: 10.1021/acs.jproteome.5c00098

[18] N. Gao, D. Yu, J. Xu, J. Hao, J. Liu, J. Yan, BMC Genomics 26 (2025) 970. DOI: 10.1186/s12864-025-12164-x

[19] T.D. Schneider, B. Roschitzki, J. Grossmann, T. Kraemer, A.E. Steuer, Anal. Chem. 94 (2022) 10695–10704. DOI: 10.1021/acs.analchem.2c01009

[20] K.L. Nielsen, J.K. Lassen, I.M. M. Løber, F. Skovbo, K. Frisch, T.P. Czaja, B. Khakimov, S.B. Engelsen, M. Johannsen, P. Villesen, J. Proteome Res. 25 (2026) 3671–3680. DOI: 10.1021/acs.jproteome.6c00199

[21] N.J. Martin, J. Bunch, H.J. Cooper, J. Am. Soc. Mass Spectrom. 24 (2013) 1242–1249. DOI: 10.1007/s13361-013-0658-1

[22] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, Nat Methods 6 (2009) 359–362. DOI: 10.1038/nmeth.1322

[23] L.B. Carvalho, J.-L. Capelo-Martínez, C. Lodeiro, J.R. Wiśniewski, H.M. Santos, Anal. Chem. 92 (2020) 9164–9171. DOI: 10.1021/acs.analchem.0c01470

[24] F. Meier, A.-D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Molecular & Cellular Proteomics 17 (2018) 2534–2545. DOI: 10.1074/mcp.TIR118.000900

[25] F. Meier, A.-D. Brunner, M. Frank, A. Ha, I. Bludau, E. Voytik, S. Kaspar-Schoenefeld, M. Lubeck, O. Raether, N. Bache, R. Aebersold, B.C. Collins, H.L. Röst, M. Mann, Nat Methods 17 (2020) 1229–1236. DOI: 10.1038/s41592-020-00998-0

[26] C. Heaton, M. Witt, L. Cole, J. Eyre, S. Tazzyman, R. McColm, S. Francese, Analyst 146 (2021) 4290–4302. DOI: 10.1039/D1AN00578B

[27] K. Kennedy, R. Bengiat, C. Heaton, Y. Herman, C. Oz, M.L. Elad, L. Cole, S. Francese, Forensic Science International 323 (2021) 110774. DOI: 10.1016/j.forsciint.2021.110774

[28] K. Kennedy, L. Cole, M. Witt, M. Sealey, S. Francese, Molecules 27 (2022) 2039. DOI: 10.3390/molecules27072039

[29] A. Igoh, M. Miura, S. Miyaishi, Anal. Chem. (2024) acs.analchem.3c04314. DOI: 10.1021/acs.analchem.3c04314

[30] S. Kuckova, T. Matecha, J. Santrucek, A. Meledina, D. Straka, K. Harcarikova, T. Dostal, L. Dvorak, V. Pitthard, S. Miklin-Kniefacz, Anal. Methods 17 (2025) 7755–7763. DOI: 10.1039/D5AY01156F

[31] Y.-R. Lee, S. Lee, S. Kwon, J. Lee, H.-G. Kang, Environmental Research 216 (2023) 114743. DOI: 10.1016/j.envres.2022.114743

[32] K. Virkler, I.K. Lednev, Forensic Science International 181 (2008) e1–e5. DOI: 10.1016/j.forsciint.2008.08.004

[33] T. Bergmann, D. Labudde, A. Thomas, S. Grabherr, Forensic Science International 374 (2025) 112530. DOI: 10.1016/j.forsciint.2025.112530

[34] M. Tiessen, H.M. Fruehwald, E.B. Easton, T. Stotesbury, Front. Anal. Sci. 2 (2022) 900483. DOI: 10.3389/frans.2022.900483

[35] R. Aebersold, M. Mann, Nature 537 (2016) 347–355. DOI: 10.1038/nature19949

[36] E.R. Stadtman, R.L. Levine, Annals of the New York Academy of Sciences 899 (2000) 191–208. DOI: 10.1111/j.1749-6632.2000.tb06187.x

[37] S.S. Adav, Proteomes 13 (2025) 24. DOI: 10.3390/proteomes13020024

[38] Y. Liu, X. Pan, M. Zhao, Y. Gao, Sci Rep 10 (2020) 14998. DOI: 10.1038/s41598-020-72196-z

[39] S. Gessulat, T. Schmidt, D.P. Zolg, P. Samaras, K. Schnatbaum, J. Zerweck, T. Knaute, J. Rechenberger, B. Delanghe, A. Huhmer, U. Reimer, H.-C. Ehrlich, S. Aiche, B. Kuster, M. Wilhelm, Nat Methods 16 (2019) 509–518. DOI: 10.1038/s41592-019-0426-7

[40] V. Demichev, C.B. Messner, S.I. Vernardis, K.S. Lilley, M. Ralser, Nat Methods 17 (2020) 41–44. DOI: 10.1038/s41592-019-0638-x

[41] L.-Z. Ma, W.-S. Liu, Y. He, Y. Zhang, J. You, J.-F. Feng, L. Tan, W. Cheng, J.-T. Yu, Journal of Advanced Research 80 (2026) 691–704. DOI: 10.1016/j.jare.2025.05.004

[42] M.A. Argentieri, S. Xiao, D. Bennett, L. Winchester, A.J. Nevado-Holgado, U. Ghose, A. Albukhari, P. Yao, M. Mazidi, J. Lv, I. Millwood, H. Fry, R.S. Rodosthenous, J. Partanen, Z. Zheng, M. Kurki, M.J. Daly, A. Palotie, C.J. Adams, L. Li, R. Clarke, N. Amin, Z. Chen, C.M. Van Duijn, Nat Med 30 (2024) 2450–2460. DOI: 10.1038/s41591-024-03164-7

[43] J. Salignon, O.R. Faridani, T. Miliotis, G.E. Janssens, P. Chen, B. Zarrouki, R. Sandberg, P. Davidsson, C.G. Riedel, Aging (2023). DOI: 10.18632/aging.204787

[44] J.P. Da Costa, P.S.M. Santos, R. Vitorino, T. Rocha-Santos, A.C. Duarte, TrAC Trends in Analytical Chemistry 93 (2017) 171–182. DOI: 10.1016/j.trac.2017.05.014

[45] U. Distler, H.B. Yoo, O. Kardell, D. Hein, M. Sielaff, M. Scherer, A.M. Jozefowicz, C. Leps, D. Gomez-Zepeda, C. Von Toerne, J. Merl-Pham, T.K. Barth, J. Tüshaus, P. Giesbertz, T. Müller, G. Kliewer, K. Aljakouch, B. Helm, H. Unger, D.L. Frey, D. Helm, L. Schwarzmüller, O. Popp, D. Qin, S.I. Wudy, L.R. Sinn, J. Mergner, C. Ludwig, A. Imhof, B. Kuster, S.F. Lichtenthaler, J. Krijgsveld, U. Klingmüller, P. Mertins, F. Coscia, M. Ralser, M. Mülleder, S.M. Hauck, S. Tenzer, Nat Commun 16 (2025) 8774. DOI: 10.1038/s41467-025-64501-z

[46] S.E. Warder, L.A. Tucker, T.J. Strelitzer, E.M. McKeegan, J.L. Meuth, P.M. Jung, A. Saraf, B. Singh, J. Lai-Zhang, G. Gagne, J.C. Rogers, Analytical Biochemistry 387 (2009) 184–193. DOI: 10.1016/j.ab.2009.01.013

[47] C. Fernández, H.M. Santos, C. Ruíz-Romero, F.J. Blanco, J. Capelo-Martínez, Electrophoresis 32 (2011) 2966–2974. DOI: 10.1002/elps.201100183

[48] J.E. Araújo, H. López-Fernández, M.S. Diniz, P.M. Baltazar, L.C. Pinheiro, F.C. Da Silva, M. Carrascal, P. Videira, H.M. Santos, J.L. Capelo, Talanta 180 (2018) 36–46. DOI: 10.1016/

j.talanta.2017.11.063

[49] I.F. Domingos, L.B. Carvalho, C. Lodeiro, R. Gerivaz, G. Prag, E. Micaglio, E. Muchtar, H.M. Santos, J.L. Capelo, Talanta 279 (2024) 126589. DOI: 10.1016/j.talanta.2024.126589

[50] J. Foox, D. Bezdan, P. Vijay, K. Getz, K. Ratanachai, J.W. Davis, K. Booher, X. Yang, C. Meydan, C.E. Mason, Forensic Genomics 1 (2021) 83–86. DOI: 10.1089/forensic.2021.0005

[51] H.C. Dias, C. Cordeiro, J. Pereira, C. Pinto, F.C. Real, E. Cunha, L. Manco, Forensic Science International 311 (2020) 110267. DOI: 10.1016/j.forsciint.2020.110267

[52] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M.A. Kamat, B.P. Prins, S.K. Wilcox, E.S. Zimmerman, A. Chi, N. Bansal, S.L. Spain, A.M. Wood, N.W. Morrell, J.R. Bradley, N. Janjic, D.J. Roberts, W.H. Ouwehand, J.A. Todd, N. Soranzo, K. Suhre, D.S. Paul, C.S. Fox, R.M. Plenge, J. Danesh, H. Runz, A.S. Butterworth, Nature 558 (2018) 73–79. DOI: 10.1038/s41586-018-0175-2

[53] N.A. Stewart, R.F. Gerlach, R.L. Gowland, K.J. Gron, J. Montgomery, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 13649–13654. DOI: 10.1073/pnas.1714926115

[54] J.H. Cater, J.R. Kumita, R. Zeineddine Abdallah, G. Zhao, A. Bernardo-Gancedo, A. Henry, W. Winata, M. Chi, B.S.F. Grenyer, M.L. Townsend, M. Ranson, C.S. Buhimschi, D.S. Charnock-Jones, C.M. Dobson, M.R. Wilson, I.A. Buhimschi, A.R. Wyatt, Proc. Natl. Acad. Sci. U.S.A. 116 (2019) 6101–6110. DOI: 10.1073/pnas.1817298116

[55] R.B. Middleton, M.C. Linder, Archives of Biochemistry and Biophysics 302 (1993) 362–368. DOI: 10.1006/abbi.1993.1224

[56] N.L. Anderson, N.G. Anderson, Molecular & Cellular Proteomics 1 (2002) 845–867. DOI: 10.1074/mcp.R200007-MCP200

[57] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W. Qian, B.-J.M. Webb-Robertson, R.D. Smith, M.S. Lipton, J. Proteome Res. 5 (2006) 277–286. DOI: 10.1021/pr050300l

[58] C. Lazar, L. Gatto, M. Ferro, C. Bruley, T. Burger, J. Proteome Res. 15 (2016) 1116–1125. DOI: 10.1021/acs.jproteome.5b00981

[59] T. Välikangas, T. Suomi, L.L. Elo, Briefings in Bioinformatics (2017). DOI: 10.1093/bib/bbx054

[60] R.J. Chalkley, K.R. Clauser, Molecular & Cellular Proteomics 11 (2012) 3–14. DOI: 10.1074/mcp.R111.015305

[61] L. Gatto, K.S. Lilley, Bioinformatics 28 (2012) 288–289. DOI: 10.1093/bioinformatics/btr645

[62] F. Da Veiga Leprevost, S.E. Haynes, D.M. Avtonomov, H.-Y. Chang, A.K. Shanmugam, D. Mellacheruvu, A.T. Kong, A.I. Nesvizhskii, Nat Methods 17 (2020) 869–870. DOI: 10.1038/s41592-020-0912-y

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

REVIEW ARTICLE | DOI: 10.5584/jiomics.v16i1.252

Proteomics in Pancreatic Ductal Adenocarcinoma: Current Advances and Translational Perspectives

Raquel A. P. F. Correia[1,2], Hugo M. Santos[1,2], José L. Capelo[1,2,*]

[1]OMICS and Analytical Development Group, BIOSCOPE Research Group, LAQV REQUIMTE, Chemistry Department, NOVA School of Science and Technology, Universidade NOVA de Lisboa, NOVA FCT, 2825-466, Portugal. [2]PROTEOMASS Scientific Society, Caparica, 2825-466, Portugal.

Available Online: August 2026

ABSTRACT

Background: Pancreatic ductal adenocarcinoma (PDAC) remains one of the most lethal malignancies, with a median overall survival below one year in most clinical settings and limited options for early detection. Although comprehensive genomic profiling has established the dominant mutational landscape, including near-universal KRAS alterations and frequent alterations in TP53, CDKN2A and SMAD4, tumour phenotypes are ultimately executed at the protein level. Mass spectrometry-based proteomics has therefore become a central platform for PDAC research, enabling quantitative assessment of signalling pathway activity, characterisation of the tumour microenvironment, and identification of circulating biomarkers that may reflect tumour biology more directly than individual genomic alterations.

Methods: We conducted a narrative review of literature on mass spectrometry-based proteomics in PDAC, covering discovery and translational studies published up to 31 December 2025. PubMed and Scopus were searched using the terms 'PDAC proteomics', 'pancreatic cancer mass spectrometry', 'pancreatic cancer biomarkers', 'glycoproteomics', 'phosphoproteomics', 'extracellular vesicle proteomics' and related combinations. Studies were selected according to methodological relevance, sample type and translational context, encompassing tissue proteomics, plasma- and serum-based proteomics, secretome analysis, extracellular-vesicle proteomics, glycoproteomics, phosphoproteomics and spatially resolved proteomics. The THBS2/CA19-9 biomarker panel was used as a running case study to illustrate the translational arc from secretome-guided discovery to prediagnostic evaluation in independent cohorts.

Results: Tissue proteogenomics, led primarily by Clinical Proteomic Tumor Analysis Consortium initiatives, has defined protein-anchored molecular subtypes and signalling dependencies that cannot be predicted from genomic data alone, although translation into circulating biomarkers is constrained by secretion dynamics, plasma dilution and systemic background. Blood-based approaches remain technically demanding because of the extreme dynamic range of the plasma proteome. The THBS2/CA19-9 panel illustrates both the potential and the limitations of circulating strategies, showing strong diagnostic performance at clinical presentation but substantially attenuated sensitivity in prediagnostic samples. Secretome-guided discovery and EV proteomics provide biologically informed frameworks with growing translational potential, and recent multi-protein EV signatures incorporating ALPPL2 and THBS2 show promise for early detection. Glycoproteomic strategies targeting tumour-associated sialylation, fucosylation and truncated O-glycans offer biologically coherent routes to improving specificity beyond bulk CA19-9 quantification. Phosphoproteomics provides a direct readout of kinase-network activity, while mass spectrometry imaging, laser microdissection-coupled proteomics and complementary spatial platforms can help resolve the tumour–stroma admixture that confounds bulk analyses.

Conclusions: Proteomics provides functional and mechanistic insights into PDAC biology that are inaccessible through genomic or transcriptomic approaches alone. Translational progress requires precise alignment between biological compartment and clinical intent, rigorous validation in multiple independent cohorts under real-world confounding conditions, and integration of protein, glycan and nucleic-acid signals. Spatial proteomics and integrative multi-omics panels are particularly promising for resolving tumour–stroma heterogeneity and improving early detection in high-risk populations.

Keywords: Pancreatic ductal adenocarcinoma, PDAC; Proteomics, Mass spectrometry, Biomarkers, Glycoproteomics, Phosphoproteomics, Extracellular vesicles, Spatial proteomics, CA19-9, Liquid biopsy.

***Corresponding author:** José Luis Capelo, jlcm@fct.unl.pt

1. Background

Pancreatic ductal adenocarcinoma (PDAC) remains one of the most lethal malignancies, with a median overall survival below one year in most clinical settings. This dismal prognosis reflects both its complex molecular architecture and the limited efficacy of current therapeutic strategies. Comprehensive genomic studies have shown that PDAC is shaped by recurrent alterations across key molecular pathways, including KRAS, TGF-β, WNT, NOTCH and ROBO/SLIT signalling, G1/S transition, SWI/SNF, chromatin modification, DNA repair and RNA processing, establishing a recurrent but heterogeneous molecular framework for the disease [1, 2]. These advances have refined molecular-subtype classification and clarified dominant oncogenic and differentiation-related programmes, underscoring the biological diversity of PDAC and its relevance to therapeutic development. Despite these advances, clinical translation has remained modest because tumour phenotypes are ultimately executed at the protein level. Translation, degradation, subcellular localisation and post-translational modifications (PTMs) determine the functional state of signalling networks. This is particularly relevant in PDAC, where desmoplastic remodelling, inflammatory signalling and therapeutic resistance are strongly mediated by extracellular-matrix (ECM) components, secreted factors and phosphorylation-dependent pathways. Proteomics can therefore complement genomic analyses by quantifying pathway activity, characterising stromal and immune contributions within the tumour microenvironment, and identifying circulating biomarkers that may more closely reflect functional tumour biology [3-6]. Contemporary proteomic research in PDAC incorporates several complementary analytical layers: global proteome profiling, phosphoproteomics, glycoproteomics and glycomics, secretome analysis, EV proteomics, and spatially resolved proteomics. This review examines each strategy in relation to its methodological basis, principal findings and realistic translational contribution.

2. Mass Spectrometry-Based Methods in PDAC Proteomics

As depicted in **Figure 1** and further detailed in Table 1 available in SM1, large-scale proteogenomic initiatives, most notably those conducted by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), have demonstrated the value of integrating proteomic, phosphoproteomic and glycoproteomic data with genomic and transcriptomic profiles. These studies have refined molecular-subtype classification and revealed protein-level signalling features that are not fully captured by transcriptome-based analyses [7, 8]. A central analytical challenge in PDAC is its low neoplastic cellularity, together with substantial stromal and immune infiltration in bulk tumour specimens [9, 10]. This cellular heterogeneity profoundly influences protein quantification and downstream biological interpretation [9, 10]. Mass spectrometry (MS)-based proteomics supports biomarker discovery, molecular characterisation and analysis of tumour tissue, plasma, serum and EVs [11, 12]. Most discovery studies use high-resolution liquid chromatography–tandem mass spectrometry (LC–MS/MS) with label-free quantification (LFQ), either through data-dependent acquisition (DDA) or data-independent acquisition (DIA), including sequential window acquisition of all theoretical fragment ion spectra (SWATH-MS) workflows. Isobaric labelling strategies, such as tandem mass tags (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ), provide an alternative route to multiplexed analysis [11, 12]. These workflows require careful experimental design, normalisation and control of batch effects. DIA is increasingly favoured in clinical proteomics because it improves quantitative completeness across large cohorts and is scalable to biobank-level studies [11, 12]. Comprehensive methodological reviews describe the implementation of these strategies across PDAC sample types [3, 11-13]. Although discovery proteomics frequently yields candidate biomarker signatures, successful clinical translation depends on stringent validation. Targeted proteomic approaches provide an established route from discovery to clinically deployable assays. These include selected reaction monitoring, also known as multiple reaction monitoring (SRM/MRM), on triple-quadrupole instruments and parallel reaction monitoring (PRM) on high-resolution platforms, usually supported by stable-isotope-labelled peptides and rigorous quality-control frameworks [14, 15]. Targeted MS is particularly relevant in PDAC because proposed biomarkers require validation in larger, independent and clinically heterogeneous cohorts, especially in the presence of confounding conditions such as pancreatitis and diabetes [16, 17]. Candidate discovery is rarely the principal bottleneck; robust validation under real-world clinical conditions ultimately determines translational success [16-18].

3. Tissue Proteomics and Proteogenomics

Tissue proteomics provides the most informative context for mechanistic discovery and therapeutic-target identification in PDAC. Proteogenomic studies have shown that proteomic layers refine tumour classification and identify signalling events that cannot be predicted from RNA expression alone. Large-scale CPTAC analyses revealed coordinated alterations in metabolic enzymes, ECM and stromal programmes, and phosphorylation-driven signalling networks, supporting the existence of protein-defined tumour subgroups with potential therapeutic implications [7]. Broader pan-cancer proteogenomic resources further confirmed that protein abundance and PTMs frequently diverge from transcriptomic patterns, underscoring the added value of direct proteomic measurement for pathway inference and target nomination [8]. More recent studies integrating tissue and serum proteomics continue to map PDAC phenotypes and nominate candidate biomarkers that overlap across compartments [13, 19]. The biological meaning of tissue-proteomic measurements is strongly dependent on sample composition. Bulk analyses of tumour tissue reflect a mixture of neoplastic, stromal, immune and normal pancreatic-cell proteins. Without accounting for neoplastic cellularity and cellular composition, proteins identified as putative tumour biomarkers may instead reflect stromal abundance rather than tumour-cell-intrinsic biology. Integrating proteomic measurements with genomic, transcriptomic and histopathological information can improve interpretation in the context of cellular heterogeneity [7-10]. Importantly, biological relevance in tumour tissue does not automatically translate into a clinically actionable

circulating biomarker. Secretion dynamics, vascular access, plasma dilution, proteolytic degradation and systemic background determine detectability in blood. Thus, mechanistic relevance in tissue does not necessarily confer analytical detectability or clinical utility in plasma or serum.

4. Blood-Based Proteomics: Plasma and Serum

Plasma- and serum-based proteomics offer clear advantages for early detection, risk stratification and longitudinal monitoring. However, the plasma proteome is one of the most analytically challenging biological matrices, exhibiting a dynamic range exceeding ten orders of magnitude and substantial inter-individual and pre-analytical variability [20, 21]. Contemporary workflows therefore require strict pre-analytical standardisation, including control of time to centrifugation, storage conditions, anticoagulant choice and freeze–thaw exposure [20, 22]. Abundant-protein depletion, enrichment and fractionation can improve proteome depth, but may introduce technical variability through sample loss and unintended co-depletion of low-abundance proteins bound to high-abundance carriers [20-23]. Large-scale LC–MS/MS workflows, including DIA-based pipelines, additionally require randomisation, quality control, normalisation, assessment and correction of batch effects where necessary, and orthogonal validation by targeted MS or immunoassays [14, 15, 20, 22]. Even under optimised conditions, performance observed at or shortly before clinical diagnosis frequently declines in prediagnostic samples, highlighting the importance of longitudinal cohort design and sampling of high-risk populations [17, 18]. The THBS2/CA19-9 biomarker panel illustrates this distinction. Initial studies reported strong diagnostic performance in symptomatic or newly diagnosed PDAC cohorts [16, 24].

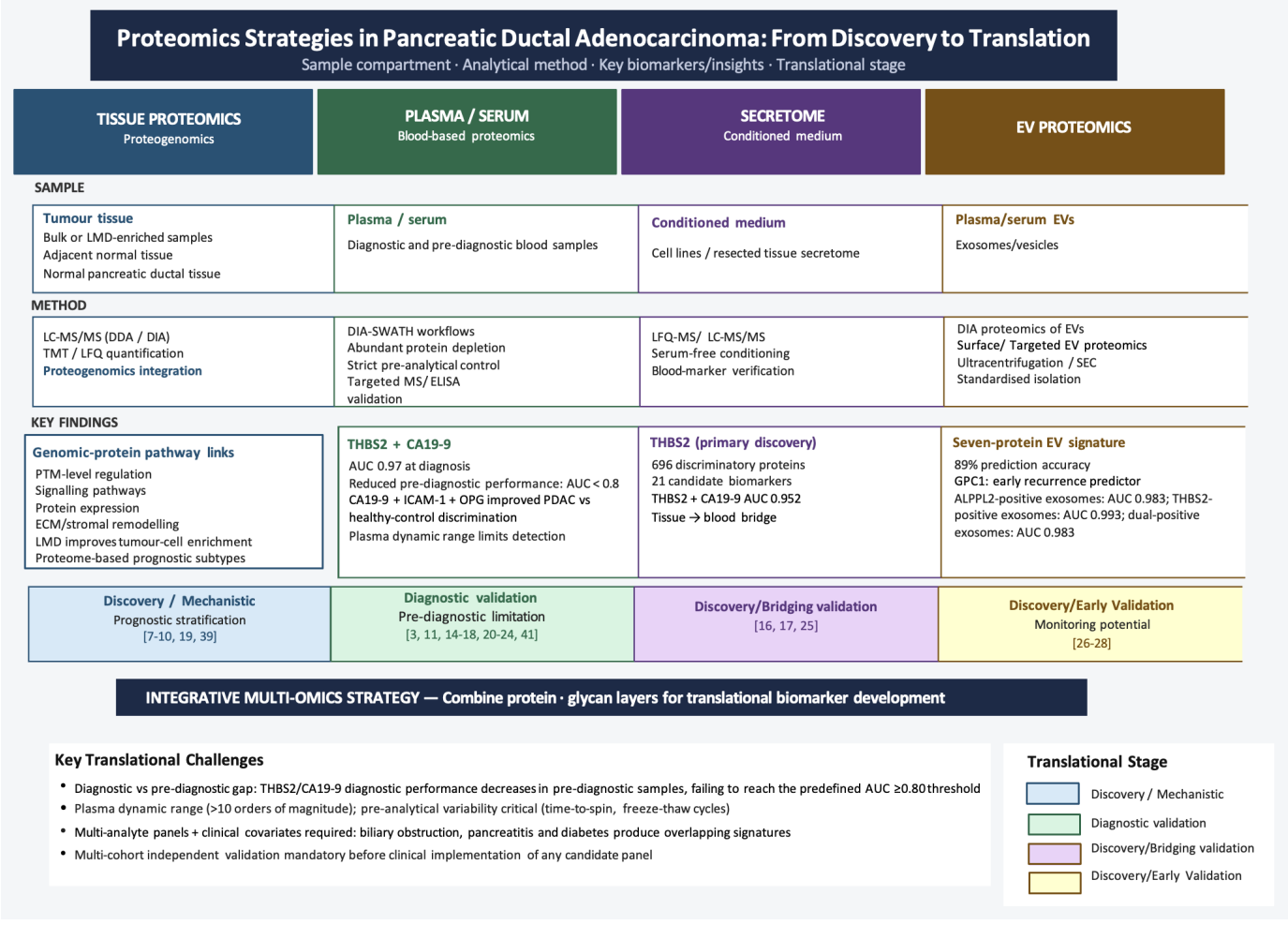

Figure 1a | Overview of proteomic and complementary analytical strategies in pancreatic ductal adenocarcinoma (PDAC), organised according to sample compartment, analytical method, key biomarkers or biological insights, and stage of translational development. Digital spatial profiling (DSP) and enzyme-linked immunosorbent assay (ELISA) are included as complementary non-mass spectrometry methods. Abbreviations: ALPPL2, alkaline phosphatase, placental-like 2; AUC, area under the curve; CA19-9, carbohydrate antigen 19-9; DDA, data-dependent acquisition; DIA, data-independent acquisition; ELISA, enzyme-linked immunosorbent assay; EV, extracellular vesicle; GPC1, glypican-1; ICAM-1, intercellular adhesion molecule 1; LC–MS/MS, liquid chromatography–tandem mass spectrometry; LFQ, label-free quantification; LMD, laser microdissection; MS, mass spectrometry; OPG, osteoprotegerin; PTM, post-translational modification; SWATH, sequential window acquisition of all theoretical fragment ion spectra; THBS2, thrombospondin-2; TMT, tandem mass tags.

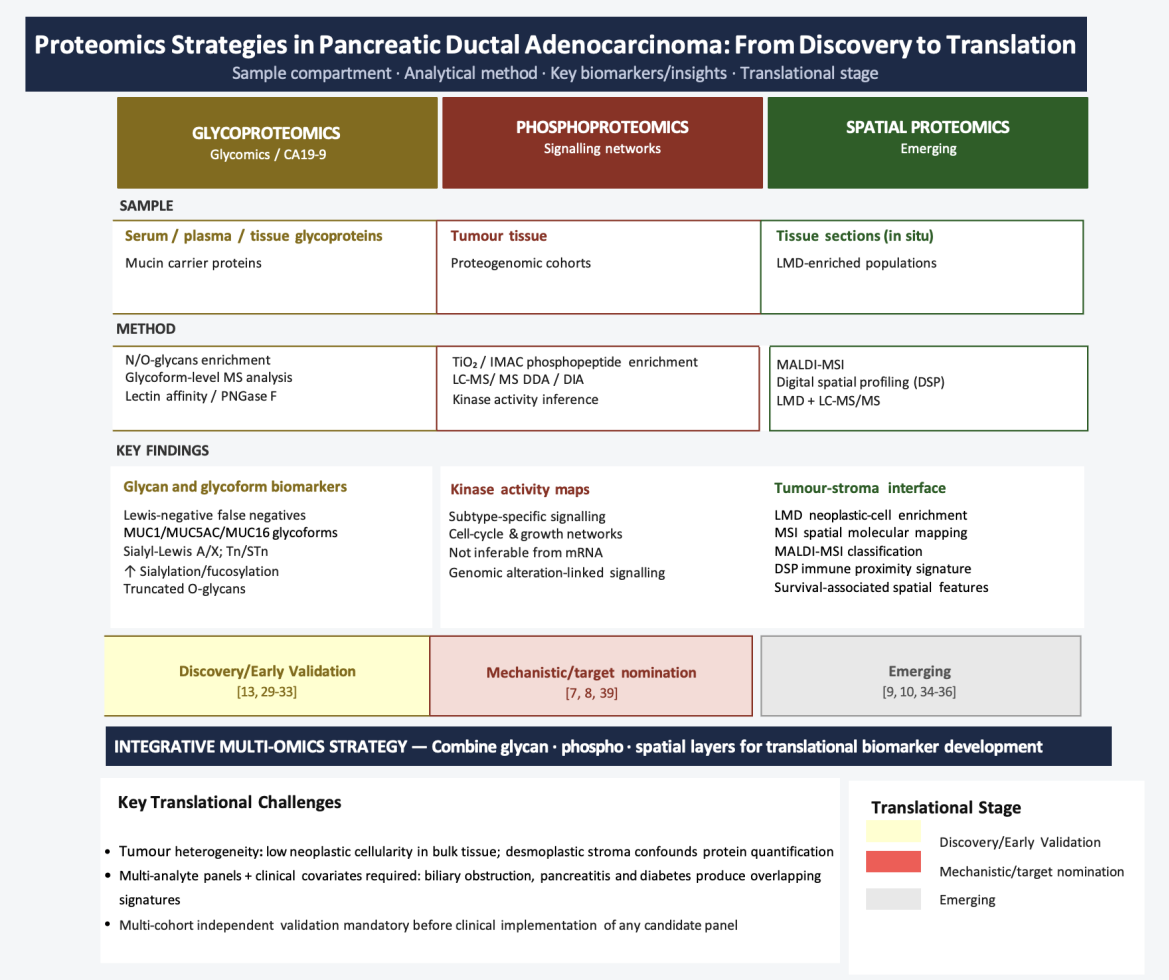

Figure 1b | Overview of proteomic and complementary analytical strategies in pancreatic ductal adenocarcinoma (PDAC), organised according to sample compartment, analytical method, key biomarkers or biological insights, and stage of translational development. Digital spatial profiling (DSP) and enzyme-linked immunosorbent assay (ELISA) are included as complementary non-mass spectrometry methods. Abbreviations: DDA, data-dependent acquisition; DIA, data-independent acquisition; DSP, digital spatial profiling; IMAC, immobilised metal affinity chromatography; LC–MS/MS, liquid chromatography–tandem mass spectrometry; LMD, laser microdissection; MALDI-MSI, matrix-assisted laser desorption/ionisation mass spectrometry imaging; MS, mass spectrometry; PNGase F, peptide-N-glycosidase F; SEC, size-exclusion chromatography.

However, performance was substantially reduced when the panel was evaluated in samples collected before diagnosis [18]. This discrepancy emphasises that early detection encompasses at least two distinct clinical problems: distinguishing PDAC from benign pancreatic disease at the time of diagnosis and detecting preclinical disease months before clinical manifestation. Study design must therefore align precisely with the intended clinical window [17, 18].

5. Secretome Proteomics

Secretome proteomics provides a rational bridge between tissue discovery and blood-based validation [25]. By profiling proteins actively secreted or shed by tumour cells and cells in the tumour microenvironment, this strategy enriches for candidates with a greater likelihood of being detectable in the circulation [25]. Thrombospondin-2 (THBS2) emerged through a secretome-oriented strategy and subsequently showed additive diagnostic value when combined with CA19-9 in plasma [16, 24]. The sequential workflow of secretome discovery, plasma verification and multi-cohort validation represents a biologically informed framework for improving translational plausibility in PDAC biomarker development [16, 24, 25].

6. Extracellular-Vesicle Proteomics

Extracellular vesicles, including small EV populations commonly described as exosomes, represent a promising liquid-biopsy compartment. EVs can concentrate tumour-derived proteins and protect molecular cargo from enzymatic degradation, potentially improving analytical sensitivity. Large-scale proteomic resources of circulating EVs are being developed to distinguish PDAC from benign pancreatic conditions and healthy controls [26]. Early enthusiasm surrounding individual EV-associated markers, particularly glypican-1-positive vesicles, stimulated rapid growth in the field. Subsequent studies, however, suggested that serum and exosomal glypican-1 may be more context-dependent and may relate to recurrence risk rather than providing a robust stand-alone diagnostic marker [27]. More recent studies have moved towards

multi-protein EV signatures incorporating markers such as ALPPL2 and THBS2, with greater emphasis on standardised isolation and analytical protocols [28]. Collectively, EV proteomics is converging on the broader principles of robust clinical proteomics: multi-analyte panels, rigorous controls and independent validation. The realistic contribution of proteomics to PDAC is therefore shaped by both sample type and clinical intent. Tissue proteomics provides direct insight into pathway activity, tumour–stroma interactions and therapeutic-target nomination; plasma and serum proteomics test whether candidate biomarkers remain detectable under stringent systemic constraints; secretome analysis increases the probability of identifying tumour- or stroma-derived proteins with properties compatible with detection in blood; and EV proteomics offers a tumour-enriched compartment with emerging translational potential. Much of the translational gap in PDAC proteomics stems not solely from analytical limitations, but also from inadequate alignment between biological compartment, analytical strategy and intended clinical use.

7. Glycoproteomics and Glycomics

PDAC is characterised by profound remodelling of the tumour glycome, including increased sialylation, enhanced fucosylation and expression of truncated O-glycans, particularly on mucins and other secreted glycoproteins. These changes are mechanistically linked to tumour progression, immune modulation and metastatic potential, reinforcing the biological relevance of glycomic and glycoproteomic approaches [29, 30]. CA19-9 (carbohydrate antigen 19-9) remains the most widely used serum biomarker in PDAC; however, it is a glycan epitope, sialyl-Lewis A, rather than a protein marker. This molecular identity underlies both its diagnostic value and its limitations. CA19-9 lacks specificity because concentrations may increase in benign conditions such as biliary obstruction and inflammatory pancreatic disease [3, 30]. Furthermore, Lewis-antigen-negative individuals cannot synthesise CA19-9, resulting in false-negative findings independent of tumour presence [31]. Total CA19-9 concentrations therefore reflect a composite signal influenced by tumour-associated glycosylation, inflammation and Lewis-antigen status. MS-based glycoproteomics can move beyond total CA19-9 quantification by enabling protein- and site-specific characterisation of glycoforms. These approaches may characterise glycan structures on defined carrier proteins, including mucins, and may potentially help distinguish tumour-associated glycosylation patterns from changes associated with benign inflammatory conditions [30, 32, 33]. This structural resolution allows interrogation of glycosylation changes that are biologically aligned with PDAC pathogenesis and may improve diagnostic specificity relative to measurements of total protein abundance alone. Reviews of the available evidence indicate that incorporating quantitative glycan characteristics or tumour-associated glycoforms may improve discrimination between PDAC and control or benign -disease groups in selected cohorts [30, 32, 33]. Nevertheless, large, multicentre validation studies, particularly in early-stage and prediagnostic populations, remain essential before clinical utility can be established. Overall, glycoproteomics represents one of the most biologically coherent strategies for the early detection of PDAC. By targeting tumour-associated glycosylation patterns, rather than relying exclusively on changes in total protein abundance that may also arise from inflammation, glycomic and glycoproteomic approaches interrogate a central biological axis in PDAC. [29, 30, 33].

8. Phosphoproteomics and Signalling-Network Analysis

Phosphoproteomics enables large-scale analysis of protein phosphorylation and provides a functional readout of intracellular signalling. In PDAC, this approach has been applied within proteogenomic frameworks to characterise oncogenic programmes and connect genomic alterations with downstream protein activity. Studies integrating genomic, transcriptomic, proteomic and phosphoproteomic layers have shown that phosphorylation data reveal kinase-network activity and pathway-activation states associated with molecular features of PDAC [7, 8]. These analyses connect mutations and molecular subtypes with downstream regulatory mechanisms, yielding functional insights that cannot be inferred from genomic data or protein abundance alone. Phosphoproteomic profiling has been particularly informative for nominating kinase-regulated vulnerabilities with potential therapeutic relevance, illustrating how PTM analysis adds functional resolution to PDAC biology [7, 8].

9. Spatially Resolved Proteomics

Spatially resolved proteomics is an emerging strategy designed to address the pronounced cellular heterogeneity and tumour–stroma admixture that characterise PDAC. Unlike bulk analyses, which yield averaged signals across mixed cell populations, spatial approaches localise protein expression to defined anatomical or cellular compartments. This capability is particularly relevant in PDAC, where neoplastic epithelium may constitute a minority of cells in a tumour specimen and the remaining signal is dominated by desmoplastic stroma, immune infiltrates and pancreatic parenchyma [9, 10]. Several complementary spatial technologies, including MS-based and antibody-based approaches, have been applied to PDAC. Mass spectrometry imaging, particularly matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI), enables untargeted mapping of peptides and proteins directly from tissue sections at spatial resolutions ranging from tens to hundreds of micrometres [34, 35]. Multiplexed platforms such as DSP enable spatial comparison of tumour, stromal and immune protein expression in PDAC [36]. Laser microdissection followed by high-sensitivity LC–MS/MS provides a complementary MS-based route to deep proteomic profiling of physically isolated cell populations, generating genuinely compartment-specific measurements rather than computationally deconvolved estimates [9]. In PDAC, spatially resolved approaches are particularly relevant for characterising the tumour–stroma interface, quantifying protein gradients between desmoplastic stroma and neoplastic epithelium, and distinguishing tumour-cell-intrinsic alterations from signals driven by the abundant stromal compartment. These capabilities address one of the most persistent sources of interpretive error in bulk tissue proteomics. Although

spatial proteomics in PDAC remains less mature than spatial transcriptomics, its integration into multi-compartment workflows should refine mechanistic interpretation, biomarker specificity and identification of stromal therapeutic targets.

10. Proteomics in Precision Oncology

Proteomics is increasingly considered a complementary diagnostic layer within precision-oncology frameworks, particularly when genomic profiling does not reveal actionable alterations or fails to explain tumour behaviour fully. Because proteomic measurements capture protein abundance, pathway activity and post-translational regulation, they can provide functional information on signalling networks that is not apparent from DNA sequencing alone. Reviews addressing integration of proteomics into clinical decision-making highlight its potential role in molecular tumour boards, where multi-omics data are interpreted jointly to identify pathway-level vulnerabilities in individual patients [37]. Curated transcriptomic, proteogenomic and other open molecular resources also enable cross-study comparison and independent validation, both of which are critical to biomarker maturation. Open, well-annotated platforms allow investigators to compare findings across cohorts, harmonise analytical pipelines and evaluate candidates in independent datasets [38, 39]. Plasma proteome-wide and proteogenomic studies additionally seek to connect circulating protein concentrations with genetic-risk variants and molecular pathways involved in PDAC development, thereby identifying proteins that may serve as mechanistic biomarkers or therapeutic targets [40].

11. Limitations and Methodological Challenges

Several methodological and biological factors limit the robustness and reproducibility of proteomics-based biomarker studies in PDAC. Important variability arises during the pre-analytical phase. In plasma proteomics, the interval between blood collection and processing, storage temperature, repeated freeze–thaw cycles and haemolysis can alter protein composition and abundance, potentially obscuring biological differences between disease and control samples. Strict standardisation of sample handling is therefore essential to minimise technical artefacts [20-22]. Interpretation of tissue-proteomic data is further complicated by the complex cellular composition of PDAC tumours. Without accounting for neoplastic cellularity and cellular composition, proteins identified as putative tumour biomarkers may reflect stromal signals rather than neoplastic epithelial biology. Proteogenomic integration and spatially informed approaches can help address this problem, but they require matched multi-omic or compartment-resolved data that are not always available [7-10]. A persistent limitation of biomarker research is reliance on a single discovery cohort. Candidates that perform well in initial discovery cohorts frequently show reduced performance in independent validation cohorts, particularly when clinically relevant pancreatic and hepatobiliary control conditions, including chronic pancreatitis, biliary obstruction and diabetes, are included. These conditions produce overlapping molecular signatures that complicate discrimination between malignant and non-malignant states. Validation across multiple independent cohorts is therefore essential before a candidate can be considered clinically relevant [17, 18, 22]. An important distinction must also be maintained between diagnostic and prediagnostic performance. Many studies evaluate patients with clinically diagnosed PDAC and may therefore produce overly optimistic estimates of diagnostic accuracy. Biomarkers that perform well at diagnosis do not necessarily detect disease during asymptomatic stages relevant to screening. Longitudinal studies in high-risk or prediagnostic populations are critical for determining whether candidates can identify PDAC before clinical presentation [18]. Finally, individual biomarkers are unlikely to provide adequate sensitivity and specificity for PDAC detection given the biological heterogeneity of the disease. Multi-analyte panels that combine complementary molecular features have greater diagnostic potential. The THBS2/CA19-9 panel illustrates this principle, showing improved performance relative to either marker alone in diagnostic settings [16, 17]. Its reduced performance in prediagnostic samples nevertheless emphasises the need for longitudinal validation before biomarker panels can be translated into early-detection strategies [18].

12. Future Directions

Hybrid panels combining circulating proteins with cfDNA, including tumour-derived ctDNA, may improve early-stage PDAC discrimination [41]. Integration with glycosylation characteristics and clinical covariates such as age, bilirubin concentration and new -onset diabetes represents an additional direction that requires prospective validation [30]. Advances in EV-isolation methods and MS workflows, particularly DIA, are expected to improve the reproducibility and scalability of EV-based proteomic analyses [11, 26, 28]. Proteome-wide association studies that link circulating protein concentrations with genetic-risk variants represent another emerging direction for identifying mechanistic biomarkers and therapeutic targets [40]. A sustained focus on early-stage PDAC and precursor lesions remains a critical unmet need. Much of the current proteomic literature is based on clinically diagnosed or advanced tumours, potentially biasing biomarker discovery towards late-stage disease biology. Targeted studies of early lesions are required to identify molecular signatures that emerge during the earliest phases of tumour development [42]. Translation of proteomic biomarkers into clinical practice will probably require standardised targeted-MS assays validated through multicentre and interlaboratory studies. SRM/MRM and PRM allow specific and reproducible quantification of selected proteins, and the establishment of such assays represents a critical step towards clinical implementation of proteomics-based diagnostics in PDAC [14, 15].

13. Conclusions

MS-based proteomics has become a central platform for PDAC research, providing functional and mechanistic information that genomic and transcriptomic approaches cannot fully capture.

Across tissue, plasma, serum, secretome, EV, glycoproteomic, phosphoproteomic and spatial domains, proteomic strategies have expanded understanding of PDAC biology and generated a growing set of candidate biomarkers, although few have reached clinical implementation. The central lessons are consistent: translational success requires alignment between biological compartment and clinical intent; individual biomarkers are unlikely to capture the heterogeneous biology of PDAC; validation in multiple cohorts under clinically realistic confounding conditions is essential; and the distinction between diagnostic and prediagnostic performance must be addressed explicitly in study design. The THBS2/CA19-9 trajectory illustrates both the promise and the limitations of the current paradigm, with strong performance at diagnosis but attenuated performance in prediagnostic samples. Integrative multi-omics strategies, spatially resolved proteomics, and hybrid molecular panels combining protein, glycan and nucleic-acid signals are among the most promising directions. These approaches may help achieve reliable early detection in high-risk populations, the clinical setting in which proteomics could have its greatest impact in PDAC.

Competing interests

The authors declare that they have no competing interests.

Funding

PROTEOMASS Scientific Society, #PM001/2019, #PM003/2016. Fundação para a Ciência e a Tecnologia (FCT/MCTES), UIDB/50006/ 2020, UIDP/50006/2020, LA/P/0008/2020, SFRH/BD/144222/2019, SFRH/BD/120537/2016. University of Pittsburgh Hillman Cancer Center shared resource facilities (Cancer Genomics Facility and The Health Science Tissue Bank) supported in part by award P30CA047904.

Authors contributions

R.A.P.F.C., J.L.C. and H.M.S. designed the review. R.A.P.F drafted the first version. J.L.C Supervised and corrected all subsequent versions done by R.A.P.F.C. and drafted the final one. H.M.S supervised and gave valuable suggestions to the final version. The authors declare that artificial intelligence-assisted tools were used solely for language editing and grammar checking during manuscript preparation. All scientific content, interpretation, and conclusions are the sole responsibility of the authors.

Acknowledgements

This work was funded by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., under the projects UID/50006/2025, UID/PRR/50006/2025 and LA/P/0008/2020 of the Associated Laboratory for Green Chemistry - LAQV REQUIMTE (https://doi.org/10.54499/UID/50006/2025, https://doi.org/10.54499/UID/PRR/50006/2025 and https://doi.org/10.54499/LA/P/0008/2020), and by the Project Peptide Boolean Biosensors: Integrating Proteomics, Computing, and Biosensing for Next-Generation Diagnostics, https://doi.org/10.54499/2024.13101.PEX, funded by FCT, I.P. H.M. Santos acknowledges his research contract that is funded by national funds from FCT – Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project LA/P/0008/2020 of the Associated Laboratory for Green Chemistry - LAQV REQUIMTE (DOI: 10.54499/LA/P/0008/2020). PROTEOMASS Scientific Society is acknowledged by the funding provided to the Laboratory for Biological Mass Spectrometry – Isabel Moura (#PM001/2025, #PM001/2024, and #PM001/2019).

References

[1] Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531(7592):47–52. doi:10.1038/nature16965.

[2] Taherian M, Wang H, Wang H. Pancreatic ductal adenocarcinoma: molecular pathology and predictive biomarkers. Cells. 2022;11(19):3068. doi:10.3390/cells11193068.

[3] Brand RE, Nolen BM, Zeh HJ, Allen PJ, Eloubeidi MA, Goldberg M, et al. Serum biomarker panels for the detection of pancreatic cancer. Clin Cancer Res. 2011;17(4):805–816. doi:10.1158/1078-0432.CCR-10-0248.

[4] Cai F, Gu Y, Ling Y, Yi G, Zang S, Su T, et al. Proteomics in pancreatic cancer. Biomark Res. 2025;13:93. doi:10.1186/s40364-025-00805-y.

[5] Vellan CJ, Jayapalan JJ, Yoong BK, Abdul-Aziz A, Mat-Junit S, Subramanian P. Application of proteomics in pancreatic ductal adenocarcinoma biomarker investigations: a review. Int J Mol Sci. 2022;23(4):2093. doi:10.3390/ijms23042093.

[6] Huang P, Gao W, Fu C, Tian R. Functional and clinical proteomic exploration of pancreatic cancer. Mol Cell Proteomics. 2023;22(7):100575. doi:10.1016/j.mcpro.2023.100575.

[7] Cao L, Huang C, Cui Zhou D, Hu Y, Lih TM, Savage SR, et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell. 2021;184(19):5031–5052.e26. doi:10.1016/j.cell.2021.08.023.

[8] Li Y, Dou Y, Da Veiga Leprevost F, Geffen Y, Calinawan AP, Aguet F, et al. Proteogenomic data and resources for pan-cancer analysis. Cancer Cell. 2023;41(8):1397–1406. doi:10.1016/j.ccell.2023.06.009.

[9] Li QK, Hu Y, Chen L, Schnaubelt M, Zhou DC, Li Y, et al. Neoplastic cell enrichment of tumor tissues using coring and laser microdissection for proteomic and genomic analyses of pancreatic ductal adenocarcinoma. Clin Proteomics. 2022;19:36. doi:10.1186/s12014-022-09373-x.

[10] Iovanna J, Fraunhoffer N, Urrutia R, Dusetti N. Understanding the heterogeneity of pancreatic ductal adenocarcinoma. Transl Oncol. 2025;60:102479. doi:10.1016/j.tranon.2025.102479.

[11] Sun X, Wang Q, Li J, Zhang H. Mass spectrometry-based proteomics technology in pancreatic cancer research. J Pancreatol. 2024;7(3):145–158. doi:10.1097/JP9.0000000000000152.

[12] Ramalhete L, Vigia E, Araújo R, Marques HP. Proteomics-driven biomarkers in pancreatic cancer. Proteomes. 2023;11(3):24. doi:10.3390/proteomes11030024.

[13] Lih TM, Cao L, Minoo P, Omenn GS, Hruban RH, Chan DW, et al. Detection of pancreatic ductal adenocarcinoma-associated proteins in serum. Mol Cell Proteomics. 2024;23(1):100687. doi:10.1016/j.mcpro.2023.100687.

[14] Shi T, Song E, Nie S, Rodland KD, Liu T, Qian WJ, Smith RD. Advances in targeted proteomics and applications to biomedical research. Proteomics. 2016;16(15–16):2160–2182. doi:10.1002/pmic.201500449.

[15] Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high-resolution and high-mass-accuracy quantitative targeted proteomics. Mol Cell Proteomics. 2012;11(11):1475–1488. doi:10.1074/mcp.O112.020131.

[16] Kim J, Bamlet WR, Oberg AL, Chaffee KG, Donahue G, Cao XJ, et al. Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. Sci Transl Med. 2017;9(398):eaah5583. doi:10.1126/scitranslmed.aah5583.

[17] Le Large TYS, Meijer LL, Paleckyte R, Boyd LNC, Kok B, Wurdinger T, et al. Combined expression of plasma thrombospondin-2 and CA19-9 for diagnosis of pancreatic cancer and distal cholangiocarcinoma: a proteome approach. Oncologist. 2020;25:e634–e643. doi:10.1634/theoncologist.2019-0680.

[18] Udgata S, Takenaka N, Bamlet WR, Oberg AL, Yee SS, Carpenter EL, et al. THBS2/CA19-9 detecting pancreatic ductal adenocarcinoma at diagnosis underperforms in prediagnostic detection: implications for biomarker advancement. Cancer Prev Res (Phila). 2021;14(2):223–232. doi:10.1158/1940-6207.CAPR-20-0403.

[19] Aref AT, Grealey J, Pathan M, Noor Z, Anees A, Azad AKM, et al. Mapping the proteomic landscape of pancreatic cancer: prognostic insights and subtype stratification. Cancer Res Commun. 2025;5(10):1879–1893. doi:10.1158/2767-9764.CRC-25-0229.

[20] Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, Baker MS, et al. Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. J Proteome Res. 2019;18(12):4085–4097. doi:10.1021/acs.jproteome.9b00503.

[21] Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002;1(11):845–867. doi:10.1074/mcp.R200007-MCP200.

[22] Nakayasu ES, Gritsenko M, Piehowski PD, Gao Y, Orton DJ, Schepmoes AA, et al. Tutorial: best practices and considerations for mass-spectrometry-based biomarker discovery. Nat Protoc. 2021;16(8):3737–3760. doi:10.1038/s41596-021-00566-6.

[23] Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, Liebler DC. Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res. 2010;9(10):4982–4991. doi:10.1021/pr100646w.

[24] Byrling J, Hilmersson KS, Ansari D, Andersson R, Andersson B. Thrombospondin-2 as a diagnostic biomarker for distal cholangiocarcinoma and pancreatic ductal adenocarcinoma. Clin Transl Oncol. 2022;24(2):297–304. doi:10.1007/s12094-021-02685-8.

[25] Xue H, Lü B, Lai M. The cancer secretome: a reservoir of biomarkers. J Transl Med. 2008;6:52. doi:10.1186/1479-5876-6-52.

[26] Bockorny B, Muthuswamy L, Huang L, Hadisurya M, Lim CM, Tsai LL, et al. A large-scale proteomics resource of circulating extracellular vesicles for biomarker discovery in pancreatic cancer. eLife. 2024;12:RP87369. doi:10.7554/eLife.87369.

[27] Zhao J, Guo M, Song Y, Liu S, Liao R, Zhang Y, et al. Serum exosomal and serum glypican-1 are associated with early recurrence of pancreatic ductal adenocarcinoma. Front Oncol. 2022;12:992929. doi:10.3389/fonc.2022.992929.

[28] Halder K, Borazanci EH, Jameson GS, Lin W, Vrana A, Cridebring D, et al. Exosomal ALPPL2 and THBS2 as biomarkers for early detection and disease monitoring of pancreatic ductal adenocarcinoma. Br J Cancer. 2025;133(9):1335–1343. doi:10.1038/s41416-025-03167-2.

[29] Guo Y, Jia W, Yang J, Zhan X. Cancer glycomics offers potential biomarkers and therapeutic targets in the framework of 3P medicine. Front Endocrinol. 2022;13:970489. doi:10.3389/fendo.2022.970489.

[30] Pienkowski T, Wawrzak-Pienkowska K, Tankiewicz-Kwedlo A, Ciborowski M, Kurek K, Pawlak D. Leveraging glycosylation for early detection and therapeutic target discovery in pancreatic cancer. Cell Death Dis. 2025;16(1):227. doi:10.1038/s41419-025-07517-z.

[31] Parra-Robert M, Santos VM, Canis SM, Pla XF, Fradera JMA, Porto RM. Relationship between CA 19.9 and the Lewis phenotype: options to improve diagnostic efficiency. Anticancer Res. 2018;38

(10):5883–5888. doi:10.21873/anticanres.12919.

[32] Silva MLS. Capitalizing glycomic changes for improved biomarker-based cancer diagnostics. Explor Target Antitumor Ther. 2023;4(3):366–395. doi:10.37349/etat.2023.00140.

[33] Bertok T, Pinkeova A, Lorencova L, Datkova A, Hires M, Jane E, Tkac J. Glycoproteomics of gastrointestinal cancers and its use in clinical diagnostics. J Proteome Res. 2025;24(6):2584–2599. doi:10.1021/acs.jproteome.5c00095.

[34] Zhang H, Lu KH, Ebbini M, Huang P, Lu H, Li L. Mass spectrometry imaging for spatially resolved multi-omics molecular mapping. npj imaging. 2024;2(1):20. doi:10.1038/s44303-024-00025-3.

[35] Kanter F, Lellmann J, Thiele H, Kalloger S, Schaeffer DF, Wellmann A, et al. Classification of pancreatic ductal adenocarcinoma using MALDI mass spectrometry imaging combined with neural networks. Cancers. 2023;15(3):686. doi:10.3390/cancers15030686.

[36] Mi H, Sivagnanam S, Betts CB, Liudahl SM, Jaffee EM, Coussens LM, et al. Quantitative spatial profiling of immune populations in pancreatic ductal adenocarcinoma reveals tumor microenvironment heterogeneity and prognostic biomarkers. Cancer Res. 2022;82(23):4359–4372. doi:10.1158/0008-5472.CAN-22-1190.

[37] Thiery J, Fahrner M. Integration of proteomics in the molecular tumor board. Proteomics. 2024;24(12–13):e2300002. doi:10.1002/pmic.202300002.

[38] Torre-Healy LA, Kawalerski RR, Oh K, Chrastecka L, Peng XL, Aguirre AJ, et al. Open-source curation of a pancreatic ductal adenocarcinoma gene expression analysis platform (pdacR) supports a two-subtype model. Commun Biol. 2023;6(1):163. doi:10.1038/s42003-023-04461-6.

[39] Tong Y, Sun M, Chen L, Wang Y, Li Y, Li L, et al. Proteogenomic insights into the biology and treatment of pancreatic ductal adenocarcinoma. J Hematol Oncol. 2022;15(1):168. doi:10.1186/s13045-022-01384-3.

[40] Feng H, Chen Z, Li J, Feng J, Yang F, Meng F, et al. Unveiling circulating targets in pancreatic cancer: Insights from proteogenomic evidence and clinical cohorts. iScience. 2025;28(3):111693. doi:10.1016/j.isci.2024.111693.

[41] Berger AW, Schwerdel D, Reinacher-Schick A, Uhl W, Algül H, Friess H, et al. A blood-based multi marker assay supports the differential diagnosis of early-stage pancreatic cancer. Theranostics. 2019;9(5):1280–1287. doi:10.7150/thno.29247.

[42] Huang Y, Sun C, Gao X, Zhai S, Wang G, Zhang F. Proteomic insights into early pancreatic ductal adenocarcinoma biology and screening. Discov Oncol. 2025;16(1):1531. doi:10.1007/s12672-025-03317-1.

PERSPECTIVE ARTICLE

JOURNAL OF INTEGRATED OMICS
A METHODOLOGICAL JOURNAL
HTTP://WWW.JIOMICS.COM

PERSPECTIVE ARTICLE | DOI: 10.5584/jiomics.v16i1.253

Integrating Omics into Endocannabinoid Research: a Paradigm Shift Toward a System-Level Understanding of a Dynamic Biological Network

Giulia Bottai [1,2], Laura Bertarini [1,3], Federica Pellati [1,*]

[1]Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 103, 41125 Modena, Italy. [2]Health Innovative Products and Technologies (HIP-TECH) PhD Program, Department of Life Sciences, University of Modena and Reggio Emilia, Via G. Campi 103, 41125 Modena, Italy. [3]Clinical and Experimental Medicine (CEM) PhD Program, Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Via G. Campi 287, 41125 Modena, Italy.

Available Online: Septemper 2026

ABSTRACT

Originally identified as the molecular machinery responsible for mediating the effects of cannabis-derived compounds, the endocannabinoid system (ECS) is now recognized as having a role in the regulation of many biological processes. Its contribution to physiological homeostasis relies on the coordinated activity of cannabinoid receptors, endogenous lipid mediators, metabolic enzymes and interconnected signaling pathways. Despite substantial advances in ECS research, several aspects of its regulation and functional relevance remain not completely understood. Many studies have traditionally focused on individual molecular components, involving either cannabinoid receptors or specific enzymes responsible for endocannabinoid metabolism. Although these approaches have provided essential mechanistic knowledge, they may not fully represent the dynamic complexity of the ECS, whose activity varies according to genetic background, metabolic state, environmental factors, cellular identity, tissue context, and disease progression. The integration of omics may provide a broader framework for investigating this complex system. Genomics, transcriptomics, proteomics, lipidomics, and metabolomics offer complementary molecular perspectives that allow the ECS to be examined beyond individual pathways and within a systems biology context. These approaches may contribute to the identification of regulatory mechanisms, molecular signatures associated with diseases, potential biomarkers, and factors influencing therapeutic response. The convergence of endocannabinoid biology and multi-omics technologies may support a more comprehensive representation of the ECS as a multidimensional and context-dependent biological network. This perspective paper discusses the current contribution of omics to the study of ECS physiology and dysfunction, together with their potential applications and the methodological and translational challenges that still need to be addressed.

Keywords: Endocannabinoid system; Endocannabinoidome; Cannabinoid signaling; Omics; System biology.

1. From Cannabinoid Receptors to a Complex Regulatory Network

The discovery of the endocannabinoid system (ECS) has broadened our understanding of cellular communication and physiological regulation. For decades, cannabinoids were investigated primarily for their psychoactive and pharmacological effects. The subsequent identification of cannabinoid receptors and endogenous ligands revealed the existence of an intrinsic signaling system that extends far beyond the pharmacological effects of cannabis-derived molecules. The ECS is now recognized as a fundamental regulator of biological homeostasis, contributing to maintaining the equilibrium between different physiological processes, including synaptic transmission, immune and inflammatory responses, appetite regulation, energy metabolism and stress adaptation [1,2]. The core components of the ECS include cannabinoid receptors, endogenous cannabinoid ligands, and the enzymes involved in their biosynthesis and degradation (**Figure 1**) [1,2].

Cannabinoid receptors CB1 and CB2 are the most extensively characterized elements of this system [3]. CB1 is among the most abundant G protein-coupled receptors in the central nervous system (CNS), where it regulates neurotransmitter release and contributes to synaptic plasticity, learning, memory, and emotional behavior [2,3].

***Corresponding author:** Federica Pellati, federica.pellati@unimore.it, +39 059 2058565

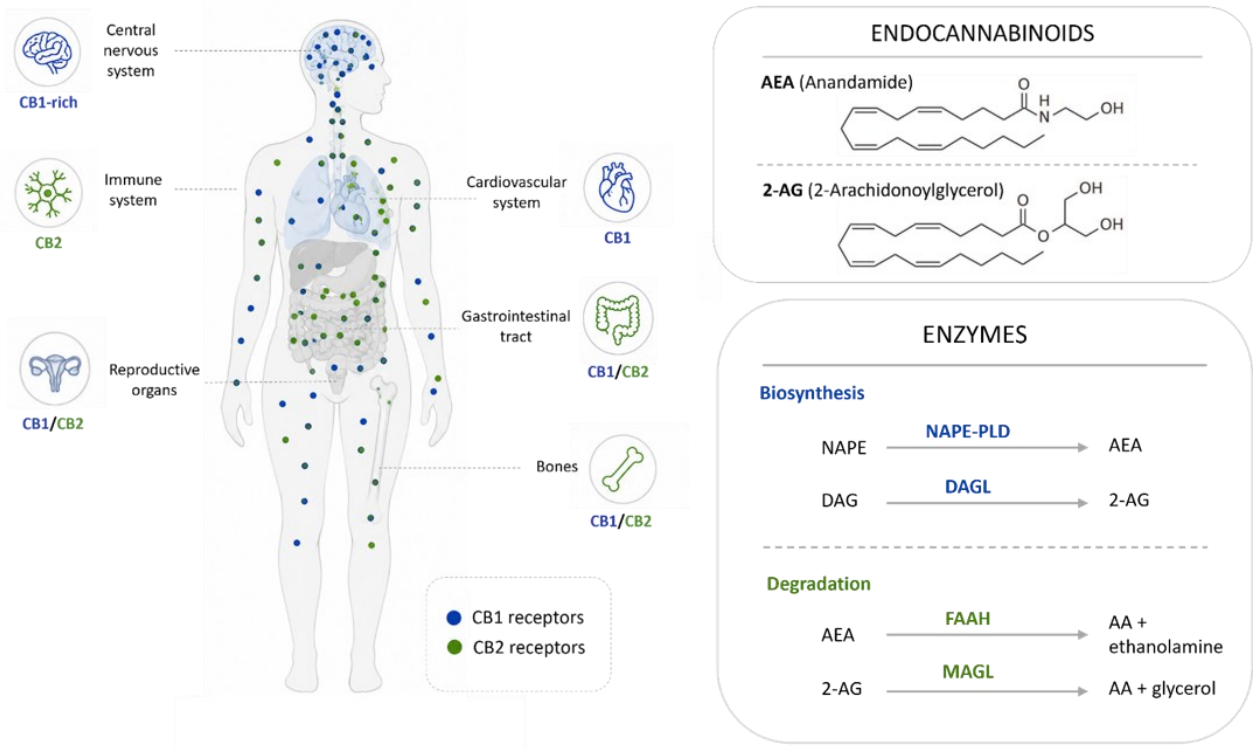

Figure 1 | Tissue distribution and principal mediators and enzymes of the endocannabinoid system. Distribution of cannabinoid receptors CB1 and CB2 across major human organs and tissues, with prominent enrichment of CB1 in the CNS and close association of CB2 with immune and peripheral compartments. Anandamide (AEA) and 2-arachidonoylglycerol (2-AG), the two best-characterized endocannabinoids, are also shown as key endogenous mediators of ECS signaling. The figure also summarizes the main enzymes involved in endocannabinoid biosynthesis and degradation: N-acyl phosphatidylethanolamine phospholipase D (NAPE-PLD) and diacylglycerol lipase (DAGL) contribute to the biosynthesis of AEA and 2-AG, respectively, whereas fatty acid amide hydrolase (*FAAH*) and monoacylglycerol lipase (MAGL) are major enzymes involved in their degradation.

CB2, initially described as an immune-associated receptor, is now also recognized for its involvement in immune cell function, inflammatory responses, and neuroimmune interactions [1,3,4]. The endogenous ligands of these receptors, referred to as endocannabinoids, include anandamide (AEA) and 2-arachidonoylglycerol (2-AG), the two most extensively investigated members of a broader family of bioactive lipid mediators [4]. Unlike classical neurotransmitters, which are synthesized in advance and stored in vesicles, endocannabinoids are mainly produced on demand in response to cellular stimulation (**Figure 2**). This mechanism supports rapid and localized modulation of cellular activity, while also making endocannabinoid activity strongly dependent on tissue context, cellular state and metabolic conditions [4].

Alterations in ECS signaling have been described in a broad range of pathological conditions, including obesity and diabetes, neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD), psychiatric conditions, inflammatory, neoplastic and cardiovascular disorders [1,4,5]. However, the functional significance of these alterations remains difficult to define. The ECS does not operate as an isolated signaling pathway but is embedded within broader processes involving lipid metabolism, mitochondrial function, immune signaling, hormonal regulation, and gene-expression networks. Consequently, a variation in a single ECS component may reflect a wider molecular response rather than an isolated defect in endocannabinoid signaling [4].

Changes in endocannabinoid levels may arise from altered biosynthesis or degradation, differences in precursor availability, or remodeling of related lipid pathways [4,6,7]. In cultured bovine adipocytes, for example, modulation of lipolytic pathways was accompanied by altered release of endocannabinoids and related lipid mediators, indicating a close relationship between cellular lipid metabolism and ECS signaling [8]. Receptor-expression data are similarly influenced by the identity and activation state of the cells being studied. In mouse brain, CB2 mRNA has been detected mainly in a small proportion of microglial cells and, to a lesser extent, in neurons, while its expression varied according to the type and timing of the inflammatory stimulus [9]. These findings also illustrate both the value and the limitations of approaches focused on individual molecular targets. Targeted methodologies remain essential for defining receptor function, enzyme activity, and signaling mechanisms, but they may be less suited to evaluating coordinated changes across different molecular levels.

A new conceptual framework is required, capable of capturing the ECS as a dynamic biological network rather than as a collection of independent components. In this context, omics sciences offer a powerful strategy to investigate a biological system across multiple layers of molecular organization. By simultaneously analyzing thousands of molecular features, omics technologies can identify complex molecular patterns associated with physiological regulation and pathological states [10].

In this view, the integration of omics into ECS research does not represent simply a technological improvement, but a paradigm shift. It enables the field to move beyond the isolated study of individual ECS components and toward an understanding of how the broader molecular ecosystem surrounding the ECS contributes to biological outcomes.

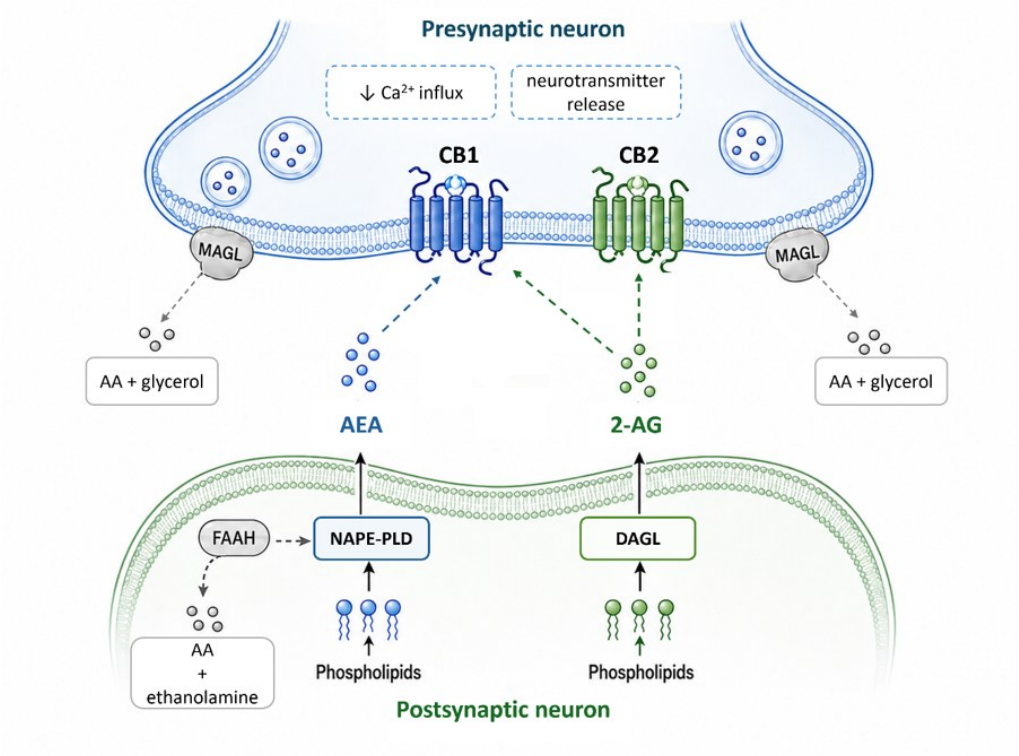

Figure 2 | Retrograde endocannabinoid signaling. Endocannabinoids are produced on demand in the postsynaptic neuron from membrane phospholipid precursors, with AEA mainly generated through NAPE-PLD and 2-AG through DAGL. After their retrograde diffusion across the synaptic cleft, AEA and 2-AG activate presynaptic cannabinoid receptors, primarily CB1, leading to reduced Ca^{2+} influx and neurotransmitter release, particularly at GABAergic and glutamatergic synapses. Signal termination is mediated by enzymatic degradation, mainly through *FAAH* for AEA and MAGL for 2-AG.

2. The Endocannabinoid System and its Multiple Molecular Dimensions

A central feature of the ECS is its context dependence. Rather than operating as a linear signaling cascade with uniform downstream effects, it acts as a modulatory network whose activity is dynamically regulated according to cellular state, tissue microenvironment, physiological and environmental conditions [2,11]. The ECS can be viewed as a regulatory system responsive to both internal physiological states and external influences [2]. Dietary fatty acid intake, stress exposure, gut microbiome composition and aging have been associated with changes in distinct aspects of ECS signaling [12-14]. Although these studies address different dimensions of the ECS regulation, they collectively indicate that endocannabinoid signaling is closely shaped by the biological context in which it operates [12-14]. The close relationship between the ECS and metabolism provides a clear example of this biological context dependence. Endocannabinoids are synthesized from membrane phospholipid-derived precursors, and they are therefore closely connected with cellular lipid metabolism [6]. Changes in nutritional status, fatty acid composition and energy balance can influence endocannabinoid production, while ECS signaling can modulate appetite, adipose-tissue function, insulin sensitivity, and mitochondrial metabolism, creating a bidirectional relationship between metabolic state and ECS signaling [6,11]. This interplay extends beyond systemic metabolic status to the local intestinal environment, where ECS activity is closely interconnected with the gut microbiota. Germ-free mice have displayed marked alterations in intestinal endocannabinoidome signaling across different intestinal regions and developmental stages, supporting a role for microbial composition in the regulation of this lipid network [15]. These findings also suggest that changes in the intestinal ECS should be interpreted in relation to the metabolic and microbial environment in which they occur.

Context-dependent regulation is also evident in immune response. In human LPS-stimulated peripheral blood mononuclear cells, pharmacological activation of CB2 modulated intracellular signaling and cytokine release, providing evidence that ECS signaling can influence inflammatory responses [16]. However, the biological effects of CB2 activation may vary according to cellular state and inflammatory context.

The interaction between the ECS and the nervous system provides a further example of context-specific regulation. Endocannabinoids regulate neurotransmitter release and synaptic plasticity through retrograde signaling mechanisms [17]. Beyond synaptic regulation,

alterations in ECS signaling have also been associated with neuroinflammatory, cognitive and neurodegenerative processes [1,4], as well as with neuropsychiatric disorders [5]. Taken together, these observations highlight the highly interconnected nature of the ECS, whose activity emerges from the interplay between multiple biological processes and is shaped by the molecular and cellular context in which it operates. Capturing this complexity requires approaches capable of examining the ECS beyond individual molecular components and pathways.

3. Omics Technologies as an Integrative Framework for Endocannabinoid Research

The development of high-throughput molecular technologies has changed the way complex biological systems are investigated. Rather than focusing exclusively on predefined molecular targets, omics allow for the simultaneous exploration of multiple molecular layers, including genes, transcripts, proteins, lipids, and metabolites [18]. This multilayered perspective is especially relevant to ECS research, where biological responses cannot be fully interpreted by considering individual components [18,19]. The major omics disciplines provide complementary perspectives:

- **Genomics** may identify either genetic variants affecting ECS components or influencing individual responses to cannabinoid-based interventions.
- **Transcriptomics** can reveal changes in the expression of receptors, enzymes, and signaling-related genes associated with ECS activity.
- **Proteomics** provides information on protein abundance, localization, interactions, and post-translational regulation, thereby linking gene expression to functional molecular mechanisms.
- **Lipidomics** allows for the characterization of endocannabinoids, endocannabinoid-like mediators, lipid precursors, and related bioactive lipid pathways.
- **Metabolomics** can describe the broader small-molecule changes associated with ECS modulation and help connect endocannabinoid signaling with cellular and systemic metabolism.

When integrated, these approaches can reveal coordinated molecular signatures linking ECS-related components to biological pathways and phenotypic outcomes (**Figure 3**). In this sense, multi-omics does not only increase the amount of measurable information but provides a framework for interpreting endocannabinoid signaling within the broader molecular context in which it operates [18,19].

4. Genomics: Understanding the Genetic Basis of Endocannabinoid Regulation

Genome represents a fundamental layer of biological complexity that may influence ECS function and contribute to interindividual variability in the endocannabinoid signaling. Variants in genes encoding cannabinoid receptors, such as *CNR1* and *CNR2*, as well as enzymes involved in endocannabinoid biosynthesis and degradation, including *DAGLA*, *MGLL*, and *FAAH*, have been investigated as potential modifiers of ECS activity and disease-related phenotypes [20,21].

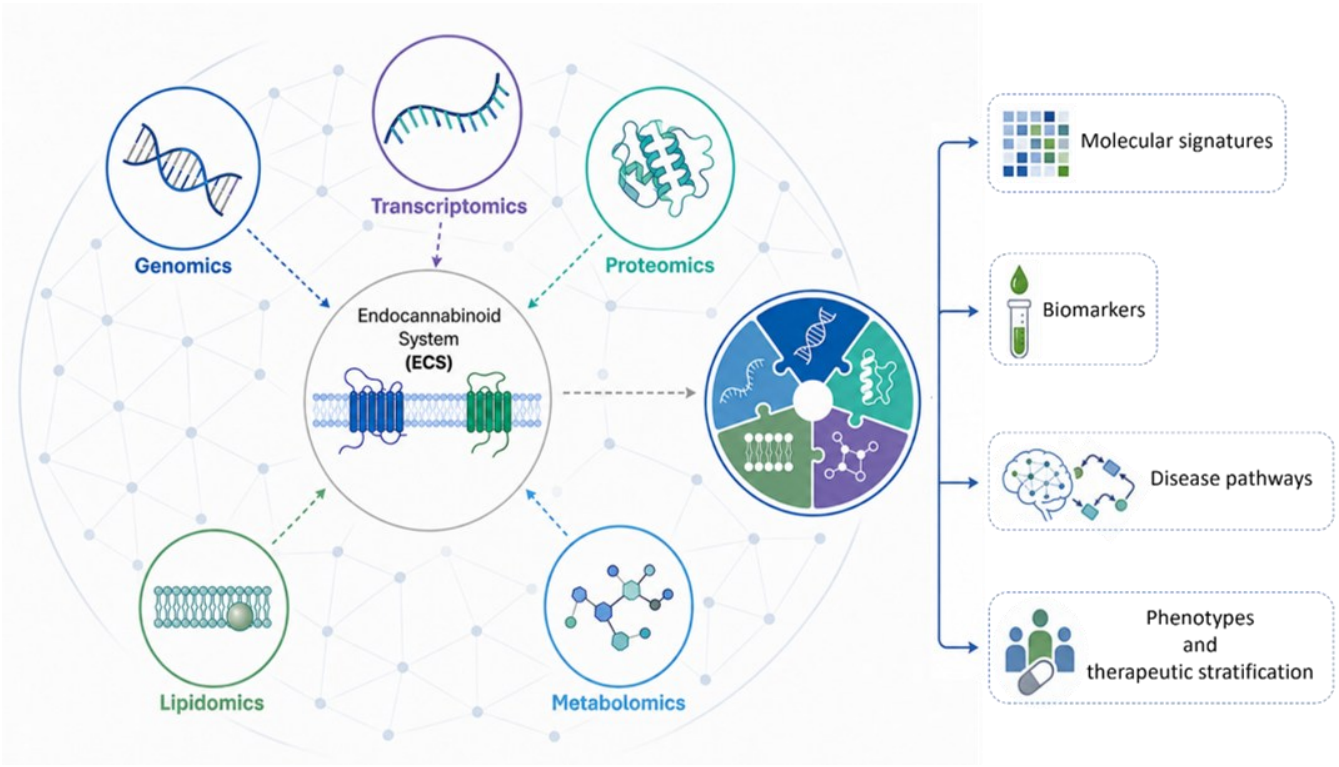

Figure 3 | Multi-omics integration in ECS research.

In a large cohort of individuals undergoing molecular testing for neurological disorders, rare variants in core ECS genes have been analyzed, and significant associations with specific neurological phenotypes were described for *CNR1* and *DAGLA* [20]. Rare *CNR1* variants have been associated with pain sensitivity, particularly migraine/headache, sleep and memory disorders, alone or in combination with anxiety, whereas rare *DAGLA* variants have been associated with seizures and neurodevelopmental disorders, including autism and abnormalities of brain morphology [20]. In parallel, a cross-disorder genome-wide association study (GWAS) meta-analysis has investigated ECS-related genes across psychiatric disorders, including major depressive disorder, bipolar disorder, attention-deficit/hyperactivity disorder, autism spectrum disorder, and schizophrenia, and it has identified *DAGLA* as the strongest gene-level association [21]. These studies suggest that genetic variability within ECS-associated pathways may contribute to neurological and psychiatric disease susceptibility [20,21]. Supporting this concept, expression quantitative trait locus (eQTL) analyses have shown that genetic variation influences the expression of a large proportion of ECS-related genes, with particularly widespread effects on endocannabinoid-catabolizing enzymes and marked tissue-specificity, highlighting a potential genetic contribution to interindividual and tissue-specific variability in ECS regulation [22]. However, these findings should be interpreted with caution, since genetic associations do not necessarily imply causality, and some signals may vary across disorders, cohorts or models applied. Independent replication in well-characterized populations and functional studies is required to clarify whether the identified variants produce measurable effects either on ECS signaling or disease biology [20,21]. Pharmacogenetics represents a promising extension of genomic research in the context of cannabinoid-based therapies. Interindividual differences in therapeutic efficacy, tolerability, and adverse effects may partly reflect genetic variation affecting cannabinoid receptors, endocannabinoid-metabolizing enzymes, and pharmacokinetic pathways involved in cannabinoid absorption, distribution, metabolism, and elimination (ADME) [23]. This is relevant because pharmacogenetic variability may influence both pharmacodynamic targets, such as *CNR1*, *CNR2*, and *FAAH*, and drug-metabolizing systems involved in the metabolism of exogenous cannabinoids [23].

In the future, genomic profiling could contribute to patient stratification and help identify molecular signatures associated with treatment response, tolerability and susceptibility to adverse reactions. However, the current evidence remains preliminary, and genotype-guided approaches for cannabinoid-based therapies have not yet reached a level of validation suitable for routine clinical implementation [23].

Nevertheless, genomics alone cannot capture the full complexity and dynamic nature of ECS regulation. The presence of a genetic variant does not necessarily translate into a measurable functional effect, since its biological consequences need to be interpreted in the broader context of gene expression, protein regulation, and metabolic state [18].

These highlights the need to integrate genomic information with transcriptomic, proteomic, lipidomic and metabolomic data to determine how genetic variation ultimately translates into functional changes in ECS activity.

5. Transcriptomics: Mapping the Cellular Regulation of the Endocannabinoid System

While genomics defines the genetic potential of a biological system, transcriptomics describes how this potential is dynamically expressed under specific physiological, pathological and environmental conditions. The ability to quantify RNA transcripts at a genome-wide scale has greatly expanded the characterization of cellular responses and complex biological networks [24].

In the context of ECS research, transcriptomic approaches enable the investigation of genes encoding cannabinoid receptors, enzymes involved in endocannabinoid biosynthesis and degradation, and molecular components of associated signaling pathways. A relevant contribution of transcriptomic analysis is the identification of cell- and context-specific patterns of ECS regulation [25]. A targeted microfluidic RT-qPCR analysis of astrocytes and microglia, purified from female mice during experimental autoimmune encephalomyelitis (EAE), has revealed both shared and cell-specific alterations in transcripts encoding cannabinoid receptors and enzymes involved in AEA and 2-AG signaling [25]. These changes varied across different stages of EAE, indicating that the transcriptional regulation of ECS components may depend on both cell identity and pathological stage [25]. This example is particularly informative because it shows that glial ECS-related transcriptional changes are not uniform but rather they are shaped by the inflammatory and cellular context in which they occur. Similarly, transcriptomic profiling of human endometrium revealed cycle-dependent changes in multiple genes involved in endocannabinoid synthesis, degradation and transport, further highlighting the dynamic regulation of ECS components according to physiological context [26].

Single-cell RNA sequencing provides an additional level of resolution by examining gene-expression profiles in individual cells rather than averaging signals from heterogeneous cell populations [24]. In the developing mouse paraventricular nucleus of the hypothalamus, the combination of single-cell RNA sequencing together with in situ hybridization and immunohistochemistry has been used to investigate the spatiotemporal distribution of genes and proteins involved in AEA and 2-AG signaling in distinct neuronal subpopulations and astrocytes [27]. This study illustrates the value of transcriptomic approaches for localizing ECS-related components during development, while also supporting the need for complementary molecular analyses to determine whether these components are functionally active [27].

Changes in transcript abundance do not necessarily correspond to changes in protein expression, enzymatic activity and endocannabinoid levels. Transcriptomic findings should be integrated with proteomics, lipidomics, and metabolomics to determine whether transcriptional alterations translate into functional changes in ECS activity [18].

6. Proteomics: Moving from Molecular Expression to Functional Regulation

Although transcriptomics provides information about gene regulation, proteomics offers a complementary and often closer view of the functional state of biological systems. Protein abundance, localization, stability, molecular interactions, and post-translational modifications can influence cellular activity independently of changes in transcript abundance [28].

In ECS research, proteomics can support the investigation of cannabinoid receptors, metabolic enzymes, and associated signaling pathways at the protein level [29]. Specifically, mass spectrometry (MS)-based proteomics enables large-scale identification and quantification of proteins, providing a broad overview of protein-level changes that may accompany ECS regulation or dysregulation [28,29]. Importantly, proteomic approaches can also provide functional information beyond protein abundance. Activity-based protein profiling coupled with chemoproteomics has been used to map the activity of endocannabinoid hydrolases across distinct mouse brain regions, revealing marked region-specific differences in enzymes, such as DAGL-α, FAAH and MAGL [30]. The disagreement observed between enzyme activity and protein abundance further illustrates that transcriptional and proteomic measurements alone may not fully capture ECS function, emphasizing the importance of integrating abundance-based omics with functional activity profiling [30].

Proteomics can also contribute to the characterization of protein–protein interactions [28,31]. This aspect is particularly relevant for cannabinoid receptors, which operate within multiprotein molecular environments containing proteins involved in intracellular signaling, membrane-associated processes, trafficking, and receptor regulation. In this context, affinity purification combined with MS has been used to characterize CB1 receptor-associated proteins in mouse hippocampal glutamatergic neurons and GABAergic interneurons, revealing both shared and cell-type-dependent receptor-associated protein networks [31]. This example illustrates how proteomics can move ECS research beyond the measurement of receptor expression, providing information on the molecular partners that may influence receptor localization, signaling, and functional output.

Specialized proteomics adds a protein-level and functionally informative layer to genomic and transcriptomic data, helping to clarify which ECS-related proteins are present and how they may participate in molecular complexes and cellular networks [28]. In the ECS, this level of information may help clarify how cannabinoid receptors and metabolic enzymes are organized within signaling complexes and how this organization may influence receptor regulation, enzymatic activity, and downstream cellular responses [28,31]. Integration of proteomics with lipidomic and metabolomic data may help determine whether protein-level alterations are accompanied by changes in endocannabinoid availability and downstream metabolic responses, although mechanistic experiments remain necessary to establish causality [18,28].

7. Lipidomics: the Central Role of Lipid Networks in Endocannabinoid Biology

Among the different omics disciplines, lipidomics occupies a particularly central position in ECS research, as endocannabinoids arise from lipid metabolic pathways and function within a broader network of structurally and functionally related mediators [7]. Although AEA and 2-AG remain the most extensively investigated endocannabinoids, their biological significance cannot be fully understood when considered alone. Lipidomics makes it possible to profile these molecules together with their precursors, metabolites, and related bioactive lipids, thereby providing a more integrated view of the lipid networks that shape ECS activity [7,32]. This broader perspective is particularly relevant because endocannabinoid signaling is influenced by substrate and precursor availability, as well as by the activity of interconnected biosynthetic and degradative pathways [7].

Dietary fatty acids can modify the availability of lipid precursors and, consequently, they can influence the production of endocannabinoids and related mediators [6,12]. Omega-3- and omega-6-derived polyunsaturated fatty acids specifically contribute to distinct families of bioactive lipids, including endocannabinoid-related compounds, with potential implications for metabolic and inflammatory regulation [32,33]. In humans, dietary intake of specific fatty acids has been associated with circulating levels of 2-AG and several related N-acylethanolamines and monoacylglycerols, supporting a link between dietary lipid composition and endocannabinoidome profiles [12].

Lipidomics is also particularly informative in inflammatory and inflammation-associated metabolic disorders, where endocannabinoid alterations are not isolated [6,33]. Rather, they may be part of a broader remodeling of lipid mediator pathways, including eicosanoids, specialized pro-resolving mediators, lysoglycerophospholipids, sphingolipids, and other bioactive lipids involved in different phases of inflammatory signaling and resolution [33]. In this context, changes in AEA, 2-AG and related N-acylethanolamines should be interpreted within the surrounding lipid profile, since their biological meaning may depend on the inflammatory and metabolic context in which they occur [33].

MS imaging adds a spatial dimension to lipid analysis by enabling the direct visualization of lipid distributions within tissue sections. This approach has been applied to map the cerebral distribution of 2-AG and to identify region-specific alterations associated with chronic restraint stress [34].

More recently, matrix-assisted laser desorption/ionization–mass spectrometry imaging (MALDI-MSI), particularly MALDI-2, was used to investigate the spatial modulation of major endocannabinoids in a mouse model combining mild traumatic brain injury (mTBI) and AD-related pathology, revealing region- and genotype-dependent changes in AEA, 2-AG and oleoylethanolamide (OEA) levels [35]. Spatially resolved lipid analysis may be particularly valuable in neurological and other complex disorders in which molecular changes are confined to specific anatomical regions or tissue microenvironments.

8. Metabolomics: Connecting ECS Activity with Metabolic Function

While lipidomics focuses on the comprehensive analysis of lipids, metabolomics provides a broader representation of the small-molecule biochemical state of cells, tissues and biological fluids. In the context of the ECS, metabolomics can help characterize the metabolic changes associated with altered endocannabinoid signaling and place variations in their levels within a wider biochemical context [18,36]. This perspective is particularly relevant in metabolic disorders, where ECS dysregulation is unlikely to occur in isolation, but it may be accompanied by coordinated changes across metabolically active tissues and pathways [37,38]. For instance, in a mouse model of diet-induced obesity, high-fat feeding has increased circulating levels of AEA and 2-AG, together with changes in the expression of enzymes involved in endocannabinoid synthesis in white and brown adipose tissue depots [37]. In addition, adipocyte-specific CB1 deletion in mice has shown protection against diet-induced obesity-associated metabolic alterations, reducing adiposity, improving insulin sensitivity, increasing energy expenditure, and promoting adipose-tissue remodeling [38]. These findings indicate that ECS-related metabolic changes should be interpreted considering both circulating lipid mediator profiles and tissue-specific metabolic regulation. In humans, circulating 2-AG levels have been positively associated with triglyceride levels and body fat percentage, supporting a relationship between endogenous cannabinoid signaling and metabolic phenotype [39]. However, these associations do not establish whether altered ECS activity contributes to metabolic dysfunction, or it develops because of it [39]. This distinction is particularly relevant in metabolomic studies, where measured metabolites may represent both drivers and readouts of biological adaptation.

When combined with lipidomic and other omics data, metabolomics may provide a relevant functional correlation between ECS regulation and cellular metabolism [18,36]. A further example is provided by the investigation of serotonergic pathways alongside endocannabinoidome alterations in a mouse model of mTBI and AD-related pathology. MALDI-MSI revealed genotype- and injury-dependent changes in serotonin and related metabolites, highlighting the interplay between neurotransmitter metabolism and ECS-related lipid mediators [40]. This approach illustrates how broader metabolic profiling can expand the interpretation of ECS dysregulation beyond endocannabinoid pathways alone.

Future metabolomic studies should be designed to measure endocannabinoids and related lipid mediators together with the metabolic features that define the phenotype under investigation, rather than treating them as stand-alone circulating markers. Depending on the biological model or clinical cohort, this may include alterations in lipid and glucose metabolism, insulin sensitivity, adipose-tissue activity and inflammatory status. This approach would help clarify whether ECS-related changes reflect a general state of metabolic imbalance or are linked to more specific metabolic alterations and tissue-dependent mechanisms [37-39].

9. Multi-Omics Integration to Interpret ECS Networks

The integration of omics sciences represents one of the most promising directions for the endocannabinoid research, but its value should not be understood simply as the simultaneous collection of larger datasets. In our view, its main contribution is conceptual: by connecting molecular layers that are often investigated separately, multi-omics can help interpret the ECS as a context-dependent regulatory network rather than as a collection of isolated molecular markers. Endocannabinoid signaling emerges from the interaction between genetic background, cell-specific transcriptional programs, protein abundance and activity, lipid mediator availability, metabolic state, cellular identity, and tissue microenvironment. For this reason, the integration of omics approaches may provide a more informative framework than the search for isolated ECS biomarkers alone and may help identify coordinated ECS-related signatures associated with specific physiological states, disease contexts and therapeutic responses [18,41].

This is particularly evident in neurological and neuroinflammatory disorders. In EAE, cell-resolved gene-expression analysis has shown that ECS-related transcriptional changes vary according to glial cell type and disease phase [25]. In human multiple sclerosis lesions, metabolomic profiling, integrated with single-nucleus RNA-sequencing data, linked alterations in lipid metabolism, including endocannabinoid and monoacylglycerol pathways with transcriptomics-based cellular profiles [42]. Other neurological conditions further illustrate the importance of connecting ECS-related alterations with disease heterogeneity and functional consequences. In AD, targeted plasma profiling of oxylipins, endocannabinoids and related lipid mediators in the so-called Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort have revealed changes across disease stages and explored their relationships with sex, apolipoprotein E (APOE) genotype, cognitive performance, and progression from mild cognitive impairment to AD [43]. Although this approach does not represent a complete multi-omics workflow, it shows how ECS-related lipid mediators can be incorporated into stratified molecular analyses of disease heterogeneity [43].

A nice example of an integrated study on the endocannabinoidome has been performed on a mouse model, combining mild traumatic brain injury (mTBI) and AD-related pathology [40]. The integration of targeted lipidomics, gene expression analysis and MALDI-MSI has revealed region- and genotype-dependent alterations, including increased 2-AG levels in the cortex and hippocampus, and linked these changes to alterations in ECS-related enzymes and molecular targets [40]. These findings highlight how multi-omics integration can provide complementary spatial and mechanistic insights into endocannabinoidome dysregulation beyond those achievable with a single omics layer [40]. In rare early-onset PD, biallelic loss-of-function variants in *DAGLB* have been linked to impaired DAGLB-dependent 2-AG synthesis and nigral dopaminergic neuron dysfunction [44]. This genetics-to-function example further illustrates that ECS-related findings become more informative when molecular alterations are

connected to downstream cellular and physiological consequences; however, this evidence should be interpreted within the specific context of DAGLB-associated PD [44].

A similar need for contextual interpretation emerges in metabolic and microbiota-related disorders [6,12,45]. In humans, dietary fatty acid intake and microbiota composition have been associated with circulating profiles of endocannabinoids and related lipid mediators [12]. In genetically obese and diabetic mice, targeted lipidomics and qPCR-based transcriptomics have revealed intestinal segment-specific alterations in endocannabinoidome and oxylipin signaling, together with changes in inflammatory markers and gut microbiota composition [45]. These observations support the idea that ECS-related lipid networks should be interpreted within a broader host–microbiota-metabolism axis, rather than as isolated metabolic readouts [45].

Cancer represents another context in which the biological meaning of ECS signaling strongly depends on cellular composition and microenvironmental conditions. ECS components may be expressed by both tumor and immune cells, and their association with cancer progression can vary according to tumor type and immune context [46]; therefore, a universal pro- or antitumor role of ECS activation cannot be assumed [46]. In melanoma, the combined analysis of The Cancer Genome Atlas (TCGA) bulk tumor gene-expression data, a previously published single-cell RNA-sequencing dataset, and RNAscope on human tissue sections has shown that *CNR2* expression is predominantly associated with tumor-infiltrating B cells [47]. In a murine melanoma model, Cnr2 deficiency has been associated with enhanced tumor growth, impaired differentiation and activation of tumor-infiltrating B cells, increased regulatory T-cell frequencies, and reduced CD8+ T-cell infiltration [47].

Taken together, these examples indicate that multi-omics integration can help move ECS research from the description of individual molecular alterations toward a more context-aware interpretation of disease mechanisms and therapeutic variability [18,41].

10. Challenges and Future Perspectives: Toward a System-Level Understanding of the ECS

Despite the considerable potential of omics-based research, several challenges must be addressed before these approaches can provide a comprehensive and clinically useful understanding of the ECS. Multi-omics studies generate high-dimensional and heterogeneous datasets that differ in molecular coverage, analytical sensitivity, scale, and preprocessing requirements, while missing values, batch effects, and platform-specific biases further complicate their integration and interpretation [18,41,48]. Robust experimental design, standardized procedures for sample collection and analysis, and adequate statistical analysis are essential to improve reproducibility across studies [18,41]. Computationally identified associations should also be independently replicated and validated through targeted biochemical analyses, cellular models and functional perturbation experiments before they can be interpreted as causal mechanisms [18,41,48]. These requirements are especially relevant to the ECS, whose molecular organization is highly dynamic and varies across tissues, cell types, physiological states, metabolic and environmental contexts [1,2,11].

Artificial intelligence, statistical learning and network-based modelling may facilitate multi-omics integration, missing-data handling, and the identification of complex relationships across molecular layers. Their reliability will nevertheless depend on data quality, adequate sample size, appropriate model selection, interpretability, and independent validation [48]. Omics technologies should therefore complement, rather than replace, classical molecular and functional approaches.

The future of endocannabinoid research is unlikely to depend solely on the discovery of additional molecular components, but increasingly on the ability to understand how established and emerging elements of the ECS interact across biological scales. This transition may improve the identification of disease mechanisms, molecular signatures and therapeutic opportunities and support more precise patient stratification and ECS-targeting strategies.

References

[1] M. Şerban, C. Toader, R.A. Covache-Busuioc, Int. J. Mol. Sci. 26 (2025) 11132. DOI:10.3390/ ijms262211132

[2] H.C. Lu, K. Mackie, Biol. Psychiatry Cogn. Neurosci. Neuroimaging 6 (2021) 607–615. DOI: 10.1016/j.bpsc.2020.07.016

[3] A. Varghese, J. Liu, W. Guo, T.A.A. Patterson, H. Hong, Front. Chem. Biol. 5 (2026) 1830007. DOI: 10.3389/fchbi.2026.1830007

[4] P. Simankowicz, J. Stępniewska, J. Clin. Med. 14 (2025) 2851. DOI: 10.3390/jcm14082851

[5] T. Mušić, T.L. Turnšek, Biomedicines 14 (2026) 968. DOI: 10.3390/biomedicines14050968

[6] M.M. Almeida, C.P. Dias-Rocha, C. Calviño, I.H. Trevenzoli, Mol. Cell. Endocrinol. 542 (2022) 111522. DOI: 10.1016/ j.mce.2021.111522

[7] J. Keereetaweep, K.D. Chapman, Neural Plast. 2016 (2016) 2426398. DOI: 10.1155/2016/2426398

[8] M.N. Myers, M. Chirivi, J.C. Gandy, J. Tam, M. Zachut, G.A. Contreras, J. Anim. Sci. Biotechnol. 15 (2024) 103. DOI: 10.1186/ s40104-024-01062-z

[9] W. Grabon, A. Ruiz, N. Gasmi, C. Degletagne, B. Georges, A. Belmeguenai, J. Bodennec, S. Rheims, G. Marcy, L. Bezin, J. Neuroinflammation 21 (2024) 206. DOI: 10.1186/s12974-024-03202-8

[10] M. Zarid, J. Genome Biotechnol. Genet. 1 (2026) 8. DOI: 10.3390/jgbg1020008

[11] A. Busquets-García, J.P. Bolaños, G. Marsicano, Nat. Metab. 4 (2022) 848–855. DOI: 10.1038/s42255-022-00600-1

[12] S. Castonguay-Paradis, S. Lacroix, G. Rochefort, L. Parent, J. Perron, C. Martin, B. Lamarche, F. Raymond, N. Flamand, V. Di Marzo, A. Veilleux, Sci. Rep. 10 (2020) 15975. DOI: 10.1038/s41598-020-72861-3

[13] A. Demaili, A. Portugalov, M. Maroun, I. Akirav, K. Braun, J. Bock, Front. Cell. Neurosci. 18 (2024) 1474992. DOI: 10.3389/fncel.2024.1474992

[14] Y. Park, B.A. Watkins, Vitam. Horm. 115 (2021) 129–172. DOI: 10.1016/bs.vh.2020.12.007

[15] C. Manca, B. Boubertakh, N. Leblanc, T. Deschênes, S. Lacroix, C. Martin, A. Houde, A. Veilleux, N. Flamand, G.G. Muccioli, F. Raymond, P.D. Cani, V. Di Marzo, C. Silvestri, J. Lipid Res. 61 (2020) 70–85. DOI: 10.1194/jlr.ra119000424

[16] A. Capozzi, D. Caissutti, V. Mattei, F. Gado, S. Martellucci, A. Longo, S. Recalchi, V. Manganelli, G. Riitano, T. Garofalo, M. Sorice, C. Manera, R. Misasi, Molecules 27 (2022) 64. DOI: 10.3390/molecules27010064

[17] B.L. Winters, C.W. Vaughan, Neuropharmacology 197 (2021) 108736. DOI: 10.1016/j.neuropharm.2021.108736

[18] M.A. Wörheide, J. Krumsiek, G. Kastenmüller, M. Arnold, Anal. Chim. Acta 1141 (2021) 144–162. DOI: 10.1016/j.aca.2020.10.038

[19] P. Berman, L. Sulimani, A. Gelfand, K. Amsalem, G.M. Lewitus, D. Meiri, Talanta 219 (2020) 121336. DOI: 10.1016/j.talanta.2020.121336

[20] D.R. Smith, C.M. Stanley, T. Foss, R.G. Boles, K. McKernan, PLoS One 12 (2017) e0187926. DOI: 10.1371/journal.pone.0187926

[21] H.K. Kim, V.F. Gonçalves, M.I. Husain, D.J. Müller, B.H. Mulsant, G. Zai, S. Kloiber, Psychiatry Res. 330 (2023) 115563. DOI: 10.1016/j.psychres.2023.115563

[22] K. Tanaka, A.A. Amoako, S. Mortlock, P.A.W. Rogers, S.J. Holdsworth-Carson, J.F. Donoghue, W.T. Teh, G.W. Montgomery, B. McKinnon, J. Cannabis Res. 7 (2025) 29. DOI: 10.1186/s42238-025-00275-x

[23] S. Hryhorowicz, M. Walczak, O. Zakerska-Banaszak, R. Słomski, M. Skrzypczak-Zielińska, Eur. J. Drug Metab. Pharmacokinet. 43 (2018) 1–12. DOI: 10.1007/s13318-017-0416-z

[24] R. Stark, M. Grzelak, J. Hadfield, Nat. Rev. Genet. 20 (2019) 631–656. DOI: 10.1038/s41576-019-0150-2

[25] Á. Moreno-García, A. Bernal-Chico, T. Colomer, A. Rodríguez-Antigüedad, C. Matute, S. Mato, Biomolecules 10 (2020) 1228. DOI: 10.3390/biom10091228

[26] K. Tanaka, A.A. Amoako, S. Mortlock, P.A.W. Rogers, S.J. Holdsworth-Carson, J.F. Donoghue, W.T. Teh, G.W. Montgomery, B. McKinnon, Sci. Rep. 12 (2022) 9400. DOI: 10.1038/s41598-022-13488-4

[27] E.O. Tretiakov, Z. Hevesi, C. Böröczky, A. Alpár, T. Harkany, E. Keimpema, Cells 14 (2025) 788. DOI: 10.3390/cells14110788

[28] R. Aebersold, M. Mann, Nature 537 (2016) 347–355. DOI: 10.1038/nature19949

[29] G. Seabra, A.C.B. Falvella, P.C. Guest, D. Martins-de-Souza, V. de Almeida, Proteomics 18 (2018) e1700270. DOI: 10.1002/pmic.201700270

[30] M.P. Baggelaar, A.C.M. van Esbroeck, E.J. van Rooden, B.I. Florea, H.S. Overkleeft, G. Marsicano, F. Chaouloff, M. van der Stelt, ACS Chem. Biol. 12 (2017) 852–861. DOI: 10.1021/acschembio.6b01052

[31] T. Mattheus, K. Kukla, T. Zimmermann, S. Tenzer, B. Lutz, J. Proteome Res. 15 (2016) 3585–3601. DOI: 10.1021/acs.jproteome.6b00339

[32] W. Röhrig, S. Achenbach, B. Deutsch, M. Pischetsrieder, J. Lipid Res. 60 (2019) 1475–1488. DOI: 10.1194/jlr.d094680

[33] V. Chiurchiù, A. Leuti, M. Maccarrone, Front. Immunol. 9 (2018) 38. DOI: 10.3389/fimmu.2018.00038

[34] Q. Zhai, A. Islam, B. Chen, H. Zhang, D.H. Chi, M. Al Mamun, Y. Takahashi, N. Sato, H. Yamasue, Y. Nakajima, Y. Nagashima, F. Sano, T. Sato, T. Kahyo, M. Setou, Cells 12 (2023) 393. DOI: 10.3390/cells12030393

[35] E. Salviati, F. Guida, D. La Gioia, F. Merciai, S. Maione, V. Di Marzo, P. Campiglia, F. Piscitelli, E. Sommella, Talanta 290 (2025) 127811. DOI: 10.1016/j.talanta.2025.127811

[36] T. Jendoubi, J. Engel, Metabolites 11 (2021) 184. DOI: 10.3390/metabo11030184

[37] E.N. Kuipers, V. Kantae, B.C.E. Maarse, S.M. van den Berg, R. van Eenige, K.J. Nahon, A. Reifel-Miller, T. Coskun, M.P.J. de Winther, E. Lutgens, S. Kooijman, A.C. Harms, T. Hankemeier, M. van der Stelt, P.C.N. Rensen, M.R. Boon, Front. Physiol. 9 (2019) 1913. DOI: 10.3389/fphys.2018.01913

[38] I.R. De Azua, G. Mancini, R.K. Srivastava, A.A. Rey, P. Cardinal, L. Tedesco, C.M. Zingaretti, A. Sassmann, C. Quarta, C. Schwitter, A. Conrad, N. Wettschureck, V.K. Vemuri, A. Makriyannis, J. Hartwig, M. Mendez-Lago, L. Bindila, K. Monory, A. Giordano, S. Cinti, G. Marsicano, S. Offermanns, E. Nisoli, U. Pagotto, D. Cota, B. Lutz, J. Clin. Invest. 127 (2017) 4148–4162. DOI: 10.1172/jci83626

[39] V. Kantae, K.J. Nahon, M.E. Straat, L.E.H. Bakker, A.C. Harms, M. Van Der Stelt, T. Hankemeier, I.M. Jazet, M.R. Boon, P.C.N. Rensen, Sci. Rep. 7 (2017) 7558. DOI: 10.1038/s41598-017-07980-5

[40] F. Guida, M. Iannotta, A. Lauritano, R. Infantino, E. Salviati, R. Verde, L. Luongo, E.M. Sommella, F.A. Iannotti, P. Campiglia, S. Maione, V. Di Marzo, F. Piscitelli, Acta Neuropathol. Commun. 12 (2024) 113. DOI: 10.1186/s40478-024-01820-0

[41] S. Graw, K. Chappell, C.L. Washam, A. Gies, J. Bird, M.S. Robeson, S.D. Byrum, Mol. Omics 17 (2021) 170–185. DOI: 10.1039/d0mo00041h

[42] D.C. Ladakis, E. Pedrini, M.I. Reyes-Mantilla, M. Sanjayan, M.D. Smith, K.C. Fitzgerald, C.A. Pardo, D.S. Reich, M. Absinta, P. Bhargava, Neurol. Neuroimmunol. Neuroinflamm. 11 (2024) e200219. DOI: 10.1212/nxi.0000000000200219

[43] K. Borkowski, C. Yin, A. Kindt, N. Liang, E. de Lange, C. Blach, J.W. Newman, R. Kaddurah-Daouk, T. Hankemeier, Alzheimer's Res. Ther. 18 (2026) 21. DOI: 10.1186/s13195-025-01939-9

[44] Z. Liu, N. Yang, J. Dong, W. Tian, L. Chang, J. Ma, J. Guo, J. Tan, A. Dong, K. He, J. Zhou, R. Cinar, J. Wu, A.G. Salinas, L. Sun, M. Kumar, B.T. Sullivan, B.B. Oldham, V. Pitz, M.B. Makarious, J. Ding, J. Kung, C. Xie, S.L. Hawes, L. Wang, T. Wang, P. Chan, Z. Zhang, W. Le, S. Chen, D.M. Lovinger, C. Blauwendraat, A.B. Singleton, G. Cui, Y. Li, H. Cai, B. Tang, Nat. Commun. 13 (2022) 3495. DOI: 10.1038/s41467-022-31168-9

[45] F. Suriano, C. Manca, N. Flamand, M. Van Hul, N.M. Delzenne, C. Silvestri, P.D. Cani, V. Di Marzo, Cells 12 (2023) 411. DOI: 10.3390/cells12030411

[46] M. Kienzl, J. Kargl, R. Schicho, Int. J. Mol. Sci. 21 (2020) 8929. DOI: 10.3390/ijms21238929

[47] T. Gruber, S. Robatel, M. Kremenovic, L. Bäriswyl, J. Gertsch, M. Schenk, Cancers 13 (2021) 1934. DOI: 10.3390/cancers13081934

[48] J.E. Flores, D.M. Claborne, Z.D. Weller, B.J.M. Webb-Robertson, K.M. Waters, L.M. Bramer, Front. Artif. Intell. 6 (2023) 1098308. DOI: 10.3389/frai.2023.1098308