EVOSEP

An Evosep White Paper

The Era of Industrial Proteomics

Proteomics is reaching the maturity genomics had a decade ago - the next ten years of drug development, diagnostics, and clinical care will be built on proteomics.

Summary

Plasma proteomics has scaled by three orders of magnitude in five years, from studies of a few hundred participants to cohorts of half a million. Per-sample coverage has come along with it: thousands of proteins quantified at clinical throughput, on hardware that exists now. Artificial Intelligence (AI) in mass spectrometry has moved at a similar pace, though by a different mechanism – large pre-trained models that interpret spectra, predict peptide properties, and learn disease signatures from datasets that did not exist when the current generation of cohorts was being planned. The two lines of progress are starting to intersect, with consequences for how proteomics is performed, interpreted, and used.

Genomes indicate what could happen; proteomes describe what is happening. AI will turn that dynamic, functional readout into decisions about targets, trials, patients, therapies, and the biologics that emerge from manufacturing pipelines. Over the next decade, standardized LC–MS based proteomics, paired with AI, is likely to become the functional measurement layer of precision medicine: the system through which human biology is read at scale, hypotheses are generated, drugs are developed and quality-controlled, and clinical decisions are made.

For that to happen, several preconditions need to be in place. AI requires data at a scale and reproducibility through standardization in order to build strong models. Proteomics needs analytical methods capable of learning from billions of spectra and translating them into useful signals. And the cost per sample has to come down to the point where LC-MS proteomics is no longer the premium option but the default one. Artisanal workflows can't satisfy these requirements, and affinity panels – useful as they are within their range – can only fill part of the gap.

What's needed are LC–MS workflows that produce comparable data across sites, instruments, and years, AI models trained on that kind of data, and per-sample costs that become attractive compared to the affinity-based alternatives. None of these pieces are exotic on their own; the difficult part has been getting them into place at the same time. Standardization is the load-bearing

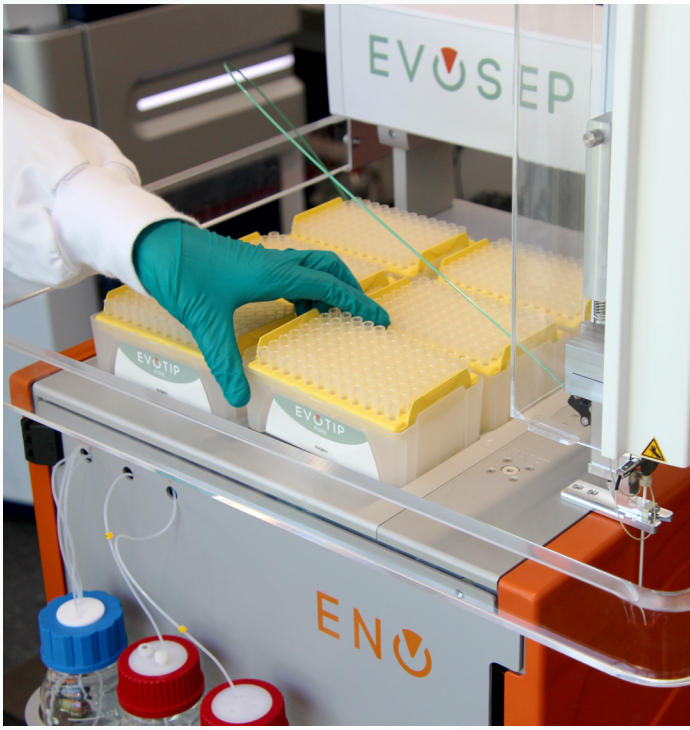
piece, and everything else depends on it. Without a high level of comparability, MS proteomics stays a cottage-industry research tool – powerful, but hard to scale into clinical or manufacturing settings. Comparable measurement is what converts the same workflow into something that clinical and manufacturing labs can build upon. Evosep was founded to drive exactly this transition.

1. The genomics precedent

Proteomics is undergoing a shift that has a recent precedent in genomics. Through the 2010s, sequencing costs collapsed, datasets became more standardized and shareable, and the AI models that were trained on that data – first early GWAS pipelines and later AlphaFold – finally had a consistent substrate to work with. Research that might once have taken a decade arrived in a few years, and human genetic evidence now roughly doubles the probability of clinical success in drug development[1].

Proteomics is following the same curve, on a measurement that is in many respects more biologically informative. E.g. the UK Biobank Pharma Proteomics Project (UKB-PPP) profiled approximately 54,000 participants on the Olink Explore 3072 affinity-based platform, measuring ~2,923 unique proteins, and showed that population-scale proteomics could generate productive biology at speed[2]. In January 2025, UK Biobank announced an expansion of that program to 500,000 participants, with an additional 100,000 repeat samples and a target of up to 5,400 proteins measured in each of 600,000 samples on Olink Explore HT[3]. This will be the largest prospective molecular dataset ever assembled on humans. However, the affinity based platform also sets the analytical ceiling for the data: affinity panels report only what their binders were designed to detect, while the next layer of proteomic biology – isoforms, post-translational modifications (PTMs), novel proteins, proteoforms that change with disease state or treatment response – sits largely outside that ceiling.

The same dynamic that makes the proteome informative in precision medicine is what makes it analytically demanding. Unlike a genome – measured once, useful for a lifetime – a proteome has to be measured repeatedly, comparably across sites, instruments, and years, and at a resolution capable of distinguishing proteoforms that may differ by a single post-translational modification. Where those conditions are met, modern AI methods open analytical possibilities that were not practical even a few years ago. Where they are not, AI cannot compensate, because the data set the ceiling on what can be learned.



2. AI has arrived in MS proteomics

Until recently, machine learning in mass spectrometry meant relatively small, purpose-built models – retention-time predictors, fragmentation models, rescoring classifiers. That generation of tools is now being augmented, and in some workflows replaced, by foundation models in the modern sense: large architectures pre-trained on tens to hundreds of millions of spectra and fine-tuned for downstream tasks. Some recent examples:

- DIA-BERT, a transformer pre-trained on 276 million peptide precursors drawn from existing DIA-MS files, reports a 51% increase in protein identifications and a 22% increase in peptide precursors compared with the established DIA-NN baseline across five human cancer datasets[4]. Spectral library coverage used to be the main practical limit on protein identification. Foundation models are starting to change that.
- The Sanders et al. tandem MS model[5], which was trained on de novo sequencing data, produces learned spectrum representations that transfer to downstream tasks where labeled data have historically been scarce – including quality prediction, chimericity detection, and PTM calling.
- Peptide-property prediction models (Prosit, AlphaPeptDeep, MS2PIP, and successors)[6] have made in silico spectral libraries a feasible alternative to empirical ones, to the point where many Data Independent Analysis (DIA) workflows are now built around them. The analytical method, in those cases, is itself partly a model.

AI is also moving upstream into instrument operation. iDIA-QC[7] achieves AUCs of 0.91 for LC faults and 0.97 for MS faults in real-time quality monitoring. And Deep Visual Proteomics (DVP), originally developed in the Mann laboratory[8], couples a vision-transformer foundation model to laser microdissection and mass spectrometry. The DVP spin-out Resolute Biosciences (formerly OmicVision Biosciences), in collaboration with MD Anderson Cancer Center, recently used the platform to detect KRAS G12A peptides in histologically normal pancreatic ducts before any visible pathology[9]. The DVP pipeline now operates on an Evosep Eno coupled to an Orbitrap Astral, running 40–120 samples-per-day Whisper-zoom standardized gradients. That this prominent frontier AI-proteomics workflow runs on standardized LC is exactly what is required when the data has to be comparable enough for foundation models to learn from.

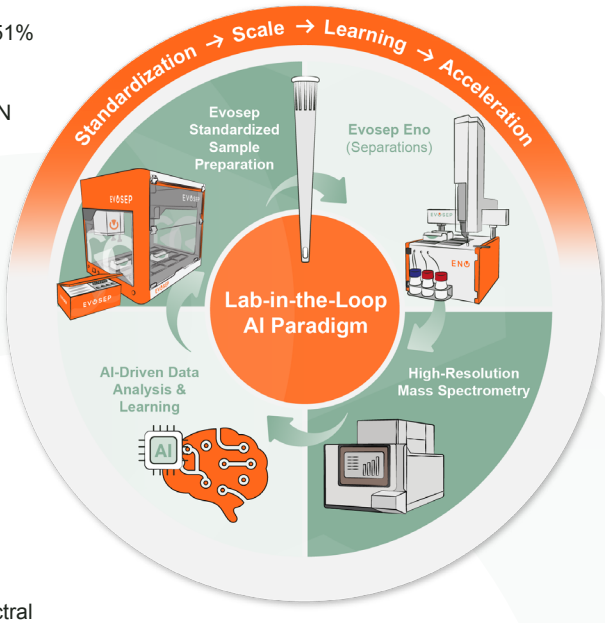


Figure 1. This workflow follows a sample through a standardized preparation and LC separation into a mass spectrometer for unbiased peptide measurement. AI models interpret the results and feed back into the design of subsequent experiments. Each stage informs the others, and the iteration allows the workflow to learn.

What proteomics analysis is undergoing is structurally similar to earlier shifts in language modelling, computer vision, and protein-structure prediction: performance is now governed primarily by the consistency and scale of the underlying data, not necessarily by the inventiveness of the model. The implication is that investments in data infrastructure may matter more, in the next phase of this field, than investments in algorithm development.

3. Standardization is the gating constraint

The foundation models described above are only as accurate as the spectra they learn from. In LC–MS based proteomics, most of the unwanted variance comes in upstream from MS, at sample preparation and liquid chromatography. How a sample is prepared will change which peptides the MS instrument sees. Protocol variation and batch effects can be pronounced without consistent and standardized sample preparation. Retention-time drift, gradient inconsistency, variation in sample load: every form of LC variability shows up downstream. No statistical method recovers what the mass spectrometer didn't measure consistently in the first place.

Evosep defines standardization as a turnkey ecosystem where hardware, consumables, and methods are pre-optimized to work together out of the box. This approach transforms sample preparation and separation from a variable to a constant, allowing different labs to achieve identical results without the need for manual fine-tuning.

Small single-site studies have always handled this variance with careful operation and tight batch design. Population-scale studies can't. They run for years, across many labs, with the kind of variability no amount of batch design absorbs. Train a model on one site's data and retention-time changes alone can stop it from generalizing to another. Running the same model for two years will require stable materials and supplies. Multi-site meta-analyses, the ones that drive therapeutic target discovery, fail when labs don't measure the same way.

Standardized LC–MS workflows now hit inter-assay CVs below 15% across more than a thousand samples per iteration, while throughputs per LC-MS system range from many tens to several hundreds of samples per day depending on the application. At that level of reproducibility and speed, MS proteomics should now scale to clinical cohorts, biopharma manufacturing QC, and all cross-site consortia requirements.

3.1 The cost case for standardization

Cost gets less attention than AI in this story, but it moves adoption just as much. Per-sample LC–MS running cost has been falling faster than affinity platforms can match. A research lab running e.g. 18 samples per day on LC–MS today pays roughly $120 per sample fully loaded. In a year, or so, the same workflow at 40 samples per day on the same MS hardware lands in the $30 range. That's a four-fold drop. It doesn't come

from new mass spectrometers; it comes from standardizing the upstream workflow on hardware that already exists.

The comparison with affinity proteomics is where users tend to be surprised. Olink Explore HT, the higher-coverage version of the Olink platform family now being deployed for the UKB-PPP expansion to 5,400 proteins, runs at $387–$823 per sample at academic and commercial rates (e.g. Vanderbilt rate sheet)[10]. The original UKB-PPP data was generated on Olink Explore 3072 at ~2,923 proteins. SomaScan and Alamar sit in similar per-sample ranges. Standardized MS-based plasma proteomics already delivers per-sample costs that are several-fold below those numbers at today's throughputs, and as workflow throughput and kit standardization mature, MS pulls into a roughly order-of-magnitude cost advantage.

MS measures what's there directly, including isoforms and PTMs. It costs less per sample. And it produces the kind of reproducibility AI models need. Those three advantages together will likely pull large cohort programs and clinical translation off affinity panels and onto MS. That shift will happen as standardized workflows take hold.

4. Where MS and affinity platforms differ

The analytical difference between MS-based proteomics and affinity platforms (Olink, SomaScan, Alamar) is obvious but consequential. Affinity platforms quantify a pre-specified panel, typically 1,000 to 11,000 proteins, using antibody or aptamer reagents that target a single epitope per protein. Within that panel, they are technically reproducible and commercially mature, but their coverage stops at the edge of the assay design.

AI-driven discovery faces several limitations, particularly in clinical settings where success depends entirely on biological resolution.

Affinity platforms are blind to anything outside what they were built to detect. Novel proteins, isoforms, proteoforms, and post-translational modifications are absent from the data the AI model sees. Phosphorylation, glycosylation, ubiquitination, and proteolytic cleavage – modifications that distinguish active proteins from inactive ones, that determine drug response, and that mark the difference between healthy and diseased tissue – do not appear in an affinity readout.

Genetic variants that alter a protein's binding site (epitope) can be mistaken for changes in protein levels rather than changes in measurement, a problem reported in every published comparison between affinity-based protein platforms and other measurement methods to date. A 2025 Nature Genetics study using nanoparticle enrichment mass spectrometry identified genetic signals affecting protein abundance that affinity-based platforms failed to detect, and estimated that up to ~30% of strong genetic associations reported by affinity-based platforms may reflect binding artefacts rather than true differences in protein levels—an upper-bound estimate, with alternative explanations possible in individual cases[11]. For drug discovery, this suggests that a meaningful share of target hypotheses derived from affinity-based data may not translate biologically, and mass spectrometry is currently the most direct way to distinguish true abundance changes from measurement artefacts.

Affinity coverage scales with reagent development, not with biological discovery. Each new protein needs a new validated binder. Isoforms usually stay inaccessible even when a binder exists.

Mass spectrometry has the opposite profile. It measures whatever peptides are in the sample, including the ones that distinguish isoforms and the ones carrying modifications. Coverage grows with LC separation and MS sensitivity, not reagent development. Quantification is sequence-specific and accurate despite genetic variation. With sufficient sensitivity, the same workflow can quantify any protein in the sample, even those not yet annotated. Proteoform resolution and PTM awareness are where the field is going next. So are measurements that track how a treatment changes protein dynamics. Affinity panels weren't built for either.

What matters here is less the contest between platforms and more the question of capability. Affinity platforms remain useful for high-throughput monitoring of known biology where the panel is well chosen and the question is well bounded. For everything beyond that – and increasingly for the use cases affinity platforms currently cover – LC-MS provides answers at lower cost and with higher information content.

5. Lab-in-the-loop proteomics in practice

In practice, proteomics-plus-AI takes the shape of a lab-in-the-loop in which AI models recommend hypotheses or next experiments, laboratory systems execute them, and the resulting data retrain the models. A 2025 review in Royal Society Open Science describes the most capable self-driving laboratories as systems that "automate nearly the entire scientific method through repeated closed-loop experimentation"[12].

In proteomics, several closed loops are already operational or nearly operational:

- Discovery loop. The discovery loop is visible in Deep Visual Proteomics, where a vision-transformer model selects the cells of interest, a laser microdissector isolates them, MS quantifies their proteome, and the results refine the model's understanding of tissue heterogeneity. The pancreatic-cancer-precursor work described in Section 2 is one example; similar loops are emerging in other areas of oncology, neurology, and immunology.

- Clinical stratification loop. Closer to the clinic, AI models developed using thousands of patient plasma proteomes propose biomarker panels that are then validated in new cohorts, with the validation results feeding back into the training data. A 2025 Behçet's disease study published in Advanced Science is an early clinical example – an XGBoost diagnostic model trained on a hybrid DIA-MS plus customizable antibody microarray platform reached AUC 0.984 in training and 0.967 in validation[13].

- Drug discovery loop. The drug discovery loop runs on human genetic and proteomic evidence used to prioritize targets and stratify trial populations. The Pharma Proteomics Project consortium, involving more than a dozen major pharma companies, exists because proteomics-plus-genetics evidence materially improves the probability of clinical success – enough to justify pre-competitive collaboration among companies that would normally guard their target lists.

These loops all depend on a measurement layer that doesn't drift. When the LC–MS is reproducible across sites and instruments and stable across years, each new batch of data trains the model further. Without that reproducibility, each batch adds variability the model has to filter out instead. Clinical-grade proteomics can't run on that.

6. Significance for pharma R&D and clinical translation

The pharma industry is under sustained pressure. Clinical development is highly attritional – the BIO "Clinical Development Success Rates 2011–2020" report puts the overall likelihood of approval following Phase I at 7.9%, with Phase II as the largest hurdle at a 28.9% transition rate, and an average of 10.5 years from Phase I to approval[14]. Independent peer-reviewed analyses put overall clinical failure rates at roughly 90%[15]. Much of this attrition reflects the difficulty of validating targets and predicting human responses from preclinical models – gaps that human-relevant molecular evidence is increasingly being asked to fill.

Meanwhile, regulators are moving toward human-relevant methods. The FDA Modernization Act 2.0 (December 2022) made certain non-animal alternatives acceptable for IND submissions[16]. In April 2025, the FDA announced a plan to

phase out animal testing requirements for monoclonal antibodies and other drugs. The agency is encouraging New Approach Methodologies (NAMs), including AI-based computational models and human organoid systems[17]. In March 2026, the FDA issued draft guidance with validation principles for NAMs in drug development. It's the first concrete framework for replacing animal studies with human-focused evidence[18]. Parallel moves are underway in the UK and EU.

Proteomics is well equipped to address both pressures. Plasma proteomics at scale has already produced concrete findings:

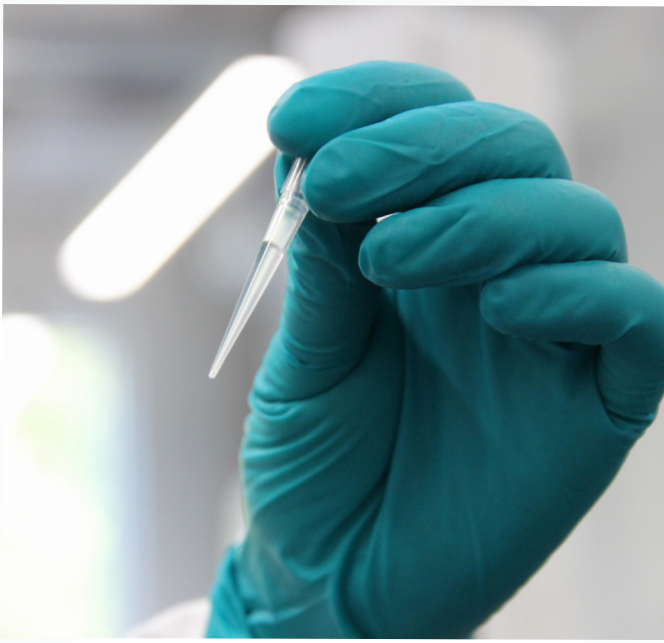
- Plasma proteomics modestly improves prediction of major cardiovascular events beyond traditional risk factors in primary prevention (UK Biobank, European Journal of Preventive Cardiology 2024)[19], with statistically significant but numerically small AUC gains (up to 0.035) over SCORE2.
- Plasma protein patterns predict dementia risk up to 15 years before diagnosis (Nature Aging 2024)[20].
- An atlas of 53,026 individuals identified 168,100 protein–disease associations across 1,066 disease endpoints[21].
- Early-warning protein panels exist for 19 cancer types, with better-than-clinical prediction for 67 health conditions, from the UKB-PPP pilot alone[3].

These outcomes emerged primarily from affinity datasets analyzed with AI and establish what can be done in principle. The next generation of work – proteoform-resolved, PTM-aware, dynamic across treatment, and capable of distinguishing closely related variants of the same protein – will require MS at comparable scale and reproducibility.

The organizations that build that infrastructure now will be the ones positioned for the next wave of target discovery, patient stratification, and companion diagnostics.

7. The same infrastructure serves biopharma manufacturing

Discovery and clinical proteomics are not the only beneficiaries of standardization. The same workflow – sample preparation, LC, MS, and AI-augmented quantification – is essentially what biopharma quality control performs on biologic products at every stage of development and manufacturing, with minor sample-specific adaptations. Identity, purity, sequence integrity, glycosylation profile, oxidation, deamidation, fragmentation, host-cell-protein content, aggregation – every one of these is a peptide- or protein-level measurement that LC–MS handles natively, and that has historically been done in artisanal, method-bespoke workflows in biopharma analytical labs.



In regulated manufacturing, standardization isn't optional: two sister sites running the same method have to produce comparable results, because if they don't, the regulator treats the gap as a finding. Drift across years inside a single lab creates the same problem on a longer timescale. Throughput and per-sample cost are tight too. Every released batch needs an analytical package, including stability time points across the product's shelf life. Slow decisions show up in cost of goods. The years of comparability data eventually become part of the product's regulatory file.

AI is reaching biopharma QC for the same reasons. PTM mapping, glycoform quantification, sequence verification, and host-cell-protein identification are all tasks where modern peptide-property and spectrum-interpretation models materially outperform traditional rule-based pipelines. Because the same standardized LC–MS infrastructure serves both discovery proteomics and biopharma QC, AI built for one carries over to the other. A QC lab running standardized workflows on the same hardware as its discovery group inherits the AI tooling.

Biopharma manufacturing is therefore in need of the same industrial proteomics infrastructure, applied to a different sample matrix and decision context. Regulatory expectations are stricter, and unit volumes per customer are higher, but the workflow underneath is the same. Customers who realize this earliest will run the same Evosep-anchored workflow in their analytical development, QC release, and stability-monitoring labs that they run in their discovery and translational labs – with the same standardized methods, comparability data, and AI tools backing both sides.

8. Evosep as infrastructure

Evosep was designed on a specific premise: clinical and translational proteomics would only become real once sample preparation and LC separations were an order of magnitude more robust and faster than research-grade workflows. That premise predates the AI wave by years, but it has now become a precondition for it.

The Evosep Eno system delivers standardized gradients from 30 to 500 samples per day. Disposable Evotip sample loading has eliminated the column-clogging and carry-over issues that used to haunt LC separation. Inter-assay CVs are tight enough that data from different labs, instruments, and years can be compared directly. At the throughput and robustness modern AI proteomics demands, the separation is not a source of variance.

The bigger claim is this: the labs, consortia, and companies that will shape the next decade of precision medicine and biopharma manufacturing are the ones already treating their measurement layer as infrastructure. That has been Evosep's working bet for a decade, and the field is now converging on the same view.

9. The decade ahead

A few trajectories over the next decade are clear enough to commit to.

Foundation models for mass spectra will likely become to proteomics what AlphaFold has been to structural biology: an interpretive layer that shifts what's possible, not a replacement for the measurement itself. Identification limits will move. PTM identification, isoform resolution, and cross-study integration will all improve by orders of magnitude. The improvements will compound on themselves.

As standardized MS proteomics gets cheaper per sample, it will pull large-cohort, clinical, and biopharma QC programs onto MS. That will happen even where affinity workflows are entrenched. Better biological resolution and lower cost per sample together make a technology displacement that wins on every dimension at once. That is rare.

The future of AI-driven proteomics relies on transforming LC-MS from a localized art into a universally comparable utility. Rigid platform standardization is what unlocks the power to reliably merge and compare data across different sites, distinct studies, and years of longitudinal timepoints.

The regulatory environment will pull proteomics-plus-AI into drug development

by making the alternatives less attractive. As FDA NAM guidance matures and animal-testing requirements recede, the pressure to produce human-relevant molecular evidence – the kind MS proteomics is uniquely positioned to generate – will only grow. Companies with standardized proteomics infrastructure already in place will hold a structural advantage in IND-enabling packages and in CMC submissions.

The story isn't really about a new tool or a new model. It's about a measurement technology that took three decades of work to bring to maturity. That system has now reached the reproducibility threshold and the cost level where modern AI can operate on it. And its biological advantages are landing in routine use. Genomics crossed a similar threshold around 2010. Proteomics is crossing it now. The organizations across academia, industry, the clinic, and manufacturing that understand this convergence and build for it will shape precision medicine in 2035.

Standardization may not be the exciting part of this story, but it's what makes everything else possible.

References

1. Nelson MR, Tipney H, Painter JL, et al (2015) The support of human genetic evidence for approved drug indications. Nat Genet 47:856–860
2. Sun BB, Chiou J, Traylor M, et al (2023) Plasma proteomic associations with genetics and health in the UK Biobank. Nature 622:329–338
3. UK Biobank (2025) Disease prediction and new drugs: why UK Biobank's huge new protein project matters. https://www.ukbiobank.ac.uk/research-stories/disease-prediction-and-new-drugs-why-uk-biobanks-huge-new-protein-project-matters/. Accessed 13 May 2026
4. Liu Z, Liu P, Sun Y, Nie Z, Zhang X, Zhang Y, Chen Y, Guo T (2025) DIA-BERT: pre-trained end-to-end transformer models for enhanced DIA proteomics data analysis. Nat Commun 16:3530
5. Sanders J, Yilmaz M, Russell J, Bittremieux W, Fondrie W, Riley N, Oh S, Noble W (2025) Foundation model for mass spectrometry proteomics. arXiv:2505.10848
6. Angelis J, Schröder EA, Xiao Z, Gabriel W, Wilhelm M (2025) Peptide property prediction for mass spectrometry using AI: An introduction to state of the art models. Proteomics. https://doi.org/10.1002/pmic.202400398
7. Gao H, Zhu Y, Wang D, et al (2025) iDIA-QC: AI-empowered data-independent acquisition mass spectrometry-based quality control. Nat Commun 16:892
8. Mund A, Coscia F, Kriston A, et al (2022) Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol 40:1231–1240. https://doi.org/10.1038/s41587-022-01302-5
9. Min J, Schweizer L, Zonderland G, et al (2025) AI-powered Deep Visual Proteomics reveals critical molecular transitions in pancreatic cancer precursors. bioRxiv. https://doi.org/10.1101/2025.07.07.663528
10. Vanderbilt University Medical Center High-throughput biomarker core rates. https://medsites.vumc.org/highthroughputbiomarkercore/pricing. Accessed 13 May 2026
11. Suhre K, Chen Q, Halama A, et al (2025) A genome-wide association study of mass spectrometry proteomics using a nanoparticle enrichment platform. Nat Genet 57:2987–2996
12. Tobias A, Wahab A (2025) Autonomous 'self-driving' laboratories: a review of technology and policy implications. R Soc Open Sci. https://doi.org/10.1098/rsos.250646
13. Cheng L, Li M, Bai Z, Yu X, Zheng W, Li Y, Liu Y (2025) Artificial intelligence-driven proteomics identifies plasma protein signatures for diagnosis and stratification of Behçet's disease. Adv Sci. https://doi.org/10.1002/advs.202510061
14. Biotechnology Innovation Organization (BIO), Informa Pharma Intelligence, QLS Advisors (2021) Clinical development success rates and contributing factors 2011–2020.
15. Sun D, Gao W, Hu H, Zhou S (2022) Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B 12:3049–3062
16. Adashi EY, O'Mahony DP, Cohen IG (2023) The FDA modernization act 2.0: Drug testing in animals is rendered optional. Am J Med 136:853–854

EVOSEP

17. U.S. Food and Drug Administration (2025) FDA announces plan to phase out animal testing requirement for monoclonal antibodies and other drugs. https://www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-other-drugs. Accessed 13 May 2026

18. U.S. Food and Drug Administration (2026) FDA releases draft guidance on alternatives to animal testing in drug development. https://www.fda.gov/news-events/press-announcements/fda-releases-draft-guidance-alternatives-animal-testing-drug-development. Accessed 13 May 2026

19. Royer P, Björnson E, Adiels M, Josefson R, Hagberg E, Gummesson A, Bergström G (2024) Large-scale plasma proteomics in the UK Biobank modestly improves prediction of major cardiovascular events in a population without previous cardiovascular disease. Eur J Prev Cardiol 31:1681–1689

20. Guo Y, You J, Zhang Y, et al (2024) Plasma proteomic profiles predict future dementia in healthy adults. Nature Aging. https://doi.org/10.1038/s43587-023-00565-0

21. Deng Y-T, You J, He Y, et al (2025) Atlas of the plasma proteome in health and disease in 53,026 adults. Cell 188:253-271.e7

EVOSEP

Next generation LC-MS based proteomics

WP-001A 26/05

evosep.com