Research Square

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

EV-PROTEOMICS: A Database-Driven Computational Framework for Refining Extracellular Vesicle-Associated Protein Profiles in Genotoxicity and Carcinogenicity Research

Nathalia Stephanie Oliveira Nascimento
Federal University of Minas Gerais (UFMG)
Adriana Oliveira Costa
Federal University of Minas Gerais (UFMG)
Wildiner Bruno Pereira Gonçalves
Federal University of Minas Gerais (UFMG)
Barbara Marra Rebuiti
Federal University of Minas Gerais (UFMG)
Ana Beatriz Souza Faria
Federal University of Minas Gerais (UFMG)
Carlos Alberto Tagliati
carlostagliati@ufmg.br

Federal University of Minas Gerais (UFMG)

Article

Keywords: New Approach Methodologies (NAMs), Subtractive proteomics, Mode of action (MoA), Computational toxicology, Biomarker discovery, Jaccard distance

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11551750/v1

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

Additional Declarations: No competing interests reported.

EV-PROTEOMICS: A Database-Driven Computational Framework for Refining Extracellular Vesicle-Associated Protein Profiles in Genotoxicity and Carcinogenicity Research

Nathalia Stephanie Oliveira Nascimento[1], Adriana Oliveira Costa[1], Wildiner Bruno Pereira Gonçalves[2],Barbara Marra Rebuiti[1], Ana Beatriz de Souza Faria[1], Carlos Alberto Tagliati*[1,2,3]

[1]Graduate Program in Clinical and Toxicological Analysis, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil.

[2]Graduate Program in Technological Innovation and Intellectual Property, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil.

[3]Interunit Graduate Program in Bioinformatics, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil.

*carlostagliati@ufmg.br

Abstract: Carcinogenicity assessment using rodent bioassays is constrained by interspecies differences, cost, duration, and ethical concerns, highlighting the need for human-relevant New Approach Methodologies (NAMs). Extracellular vesicles (EVs) are potential sources of molecular biomarkers, but heterogeneous EV-associated proteomic data can complicate the identification of mechanistically informative patterns. We present EV-PROTEOMICS, a database-driven computational framework for refining curated EV-associated protein profiles across predefined toxicological reference categories. Protein profiles were represented as binary data and processed using a subtractive refinement strategy. Twelve reference substances representing genotoxic and non-genotoxic categories were evaluated. Refinement reduced the genotoxicity-associated protein set from 245 to 119 proteins. Jaccard dissimilarity and Principal Coordinates Analysis showed that the proportion of variation represented by the first two axes increased from 31.2% in the initial dataset to 66.3% after refinement. Further mechanistic exploration prioritized candidates, including NDRG2 and NFE2L2, in profiles associated with aromatic amine/nitrosamine reference substances. These findings provide an in silico proof of concept for refining EV-associated protein datasets and generating candidate signatures for mechanistic toxicology research. Independent datasets and experimental studies are required to assess reproducibility and determine discriminatory or predictive performance.

Keywords: New Approach Methodologies (NAMs); Subtractive proteomics; Mode of action (MoA); Computational toxicology; Biomarker discovery; Jaccard distance

1. Introduction

The establishment of New Approach Methodologies (NAMs) has driven the transition from animal-based toxicology toward mechanistically oriented strategies grounded in human biology. When assessing carcinogenic potential, the limitations of translating *in vivo* findings to humans underscore the need for predictive approaches capable of reducing reliance on animal testing [1, 2]. Recent regulatory initiatives in both the United States and the European Union have promoted the development, validation, and incorporation of NAMs as alternatives or complements to conventional experimental models [3- 5].

Computational methods are critical components of NAMs, enabling the integration and interpretation of data to support predictive toxicological assessments.

In genetic toxicology, a central challenge lies in discriminating among different modes of action (MoAs), including mutagenicity, clastogenicity, and aneugenicity. Traditional phenotypic assays, such as the micronucleus test and the comet assay, are important for identifying genetic damage but have limitations in characterizing the underlying molecular mechanisms [6]. This distinction is critical for toxicological and regulatory interpretation; characterizing MoAs links early molecular alterations to key events and ultimate adverse outcomes, a principle central to the Adverse Outcome Pathway (AOP) framework [7, 8]. The identification of molecular patterns capable of discriminating between related mechanisms is therefore an essential step in developing Integrated Approaches to Testing and Assessment (IATA)[9].

A recent systematic scoping review by our group demonstrated that while mechanistic integration is increasing, it remains predominantly complementary to apical endpoints. Among 120 studies published between 2020 and 2025, 65% incorporated mechanistic components, whereas 91.7% remained anchored to established genotoxicity endpoints. Systems-level approaches based on omics and computational integration were less frequently represented, highlighting a gap between the growing demand for mechanistic resolution and the adoption of more complex molecular strategies [6].

This mapping review identified no studies investigating extracellular vesicles (EVs) as mechanistic biomarkers in genotoxicity assessments [6]. EVs are dynamic molecular compartments containing proteins, nucleic acids, lipids, and other molecules involved in cellular state regulation and intercellular communication[10]. Their molecular cargo reflects processes related to DNA damage responses, cell-cycle regulation, apoptosis, and cellular stress [11-14], underscoring their potential as sources of mechanistic information in genetic toxicology. Consequently, the EV proteome is a promising source of biomarkers for investigating molecular alterations following chemical exposure.

However, using EV proteomic data for mechanistic discrimination presents challenges arising from EV heterogeneity, variability across experimental protocols and analytical platforms, and the presence of nonspecific cellular responses [10, 15]. Proteins associated with generalized cellular stress can emerge under various toxicological conditions, obscuring mechanism-specific patterns. Strategies that prioritize recurrent, discriminative protein patterns—rather than relying solely on quantitative differences across heterogeneous studies, can enhance the extraction of mechanistically relevant signals.

One strategy to address this challenge is to analyze the recurrence of molecular features across reference sets, isolating shared components from those specifically associated with distinct mechanistic classes. Binary data representation allows the comparison of molecular profiles based on protein presence or absence, thereby reducing reliance on quantitative measurements from varying experimental setups.

Similarity metrics, such as the Jaccard index, can then be employed to assess the overlap between profiles and their clustering in relation to different toxicological classes [16].

To bridge this gap, this study investigates the computational component of EV-PROTEOMICS, an integrated platform designed to combine computational analysis, experimental evidence, human biological models, and mechanistic interpretation [17]. Specifically, this study explores whether computational refinement of curated EV-associated protein datasets can generate distinct candidate signatures across predefined toxicological reference categories. The analysis further examines the organization of these signatures using binary representation, Jaccard dissimilarity, ordination, protein–protein interaction networks, and functional enrichment. This work constitutes an exploratory in silico proof of concept intended to support the generation of mechanistic hypotheses for subsequent experimental evaluation.

2. Results

2.1. Initial Organization of Extracellular Vesicle Proteomic Profiles

Exploratory analysis of extracellular vesicle (EV)-derived proteomic profiles prior to refinement revealed substantial heterogeneity and overlap among the toxicological classes. Principal Coordinates Analysis (PCoA) using Jaccard dissimilarity demonstrated that the first two principal coordinates captured 31.2% of the total variance, with Axis 1 and Axis 2 accounting for 16.7% and 14.5%, respectively **(Figure 1a).**

Despite this overlap, profiles exhibited partial spatial organization in the multidimensional space, displaying a discernible trend toward separation between the genotoxic/mutagenic and control groups. Hierarchical clustering based on Jaccard distances corroborated this structure, revealing distinct clusters and greater relative similarity among profiles driven by shared mechanistic modes **(Figure 1b).** Collectively, these findings indicate that while raw proteomic data inherently contain heterogeneity and inter-group overlap, they still harbor potentially discriminative mechanistic information.

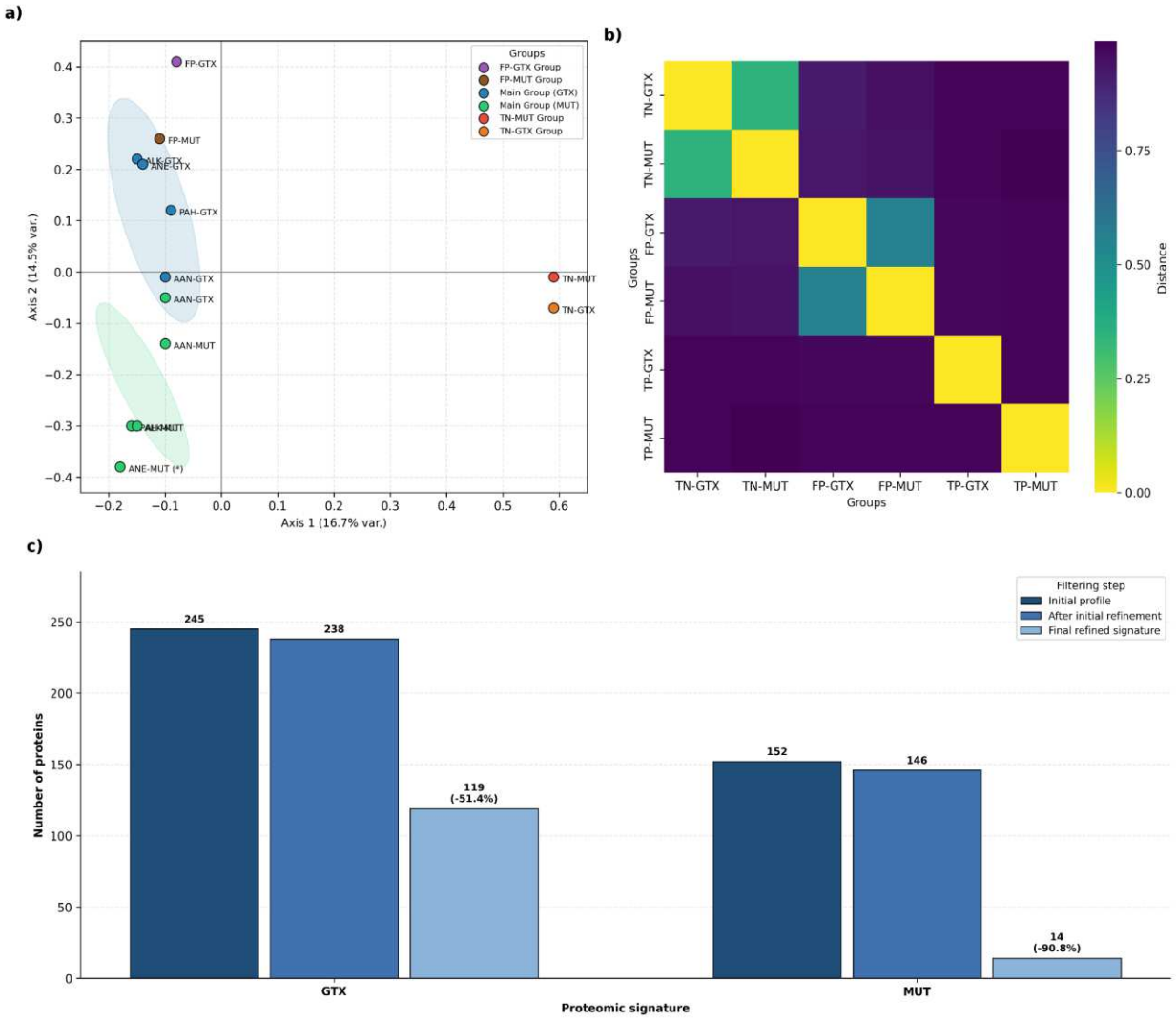


Figure 1. Computational refinement and structural organization of proteomic signatures. (A) Principal Coordinates Analysis (PCoA) using Jaccard dissimilarity, demonstrating enhanced separation and spatial organization of toxicological groups following computational refinement. (B) Hierarchical clustering heatmap based on Jaccard dissimilarity, illustrating pairwise similarities among experimental groups post-refinement. (C) Progressive reduction in protein counts across computational filtering stages for the genotoxic carcinogenicity (GTX) and mutagenicity (MUT) signatures. The final refined signatures comprised 119 proteins for GTX and 13 for MUT, reflecting reductions of 51.4% and 90.8%, respectively. Abbreviations: ALK, alkylating agents; AAN, aromatic amines and nitrosamines; PAH, polycyclic aromatic hydrocarbons; ANE, aneugenic compounds; FP, false positives; TN, true negatives; TP, true positives; MUT, mutagenicity; GTX, genotoxic carcinogenicity.

2.2. Exploratory Organization of Refined Proteomic Profiles Across Toxicological Reference Categories

Applying the EV-PROTEOMICS computational framework progressively reduced the number of proteins retained in the refined signatures. For the genotoxic carcinogenicity (GTX) set, the initial profile comprised 245 proteins, which decreased to 238 following the first refinement step, and ultimately to 119. Similarly, the initial mutagenicity (MUT) profile contained 153 proteins, which dropped to 146 and finally to 13 **(Figure 1c)**. The reduction was more pronounced in the MUT set, resulting in a smaller candidate signature. This reduction reflects the outcome of the refinement procedure but does not, by itself, establish improved biological specificity or predictive performance. The complete quantitative workflow detailing protein counts across processing stages is provided in Table S1 (Supplementary Material).

2.3. Profile Refinement Improves Discrimination Among Toxicological Classes

Post-refinement profile organization was evaluated via PCoA and hierarchical clustering using Jaccard dissimilarity. For the True Positive (TP), False Positive (FP), and True Negative (TN) groups, the first two principal coordinates captured 66.3% of the total variance, with 36.8% and 29.5% attributed to Axis 1 and Axis 2, respectively **(Figure 2a)**. Following refinement, the resulting profiles exhibited a more distinct spatial arrangement in the PCoA representation. This pattern describes the organization of the refined datasets under the analytical conditions used and does not independently establish classification accuracy or generalizability to previously unexamined substances. Hierarchical clustering corroborated these findings, revealing higher intra-group similarity and greater dissimilarity across distinct toxicological classes **(Figure 2c).**

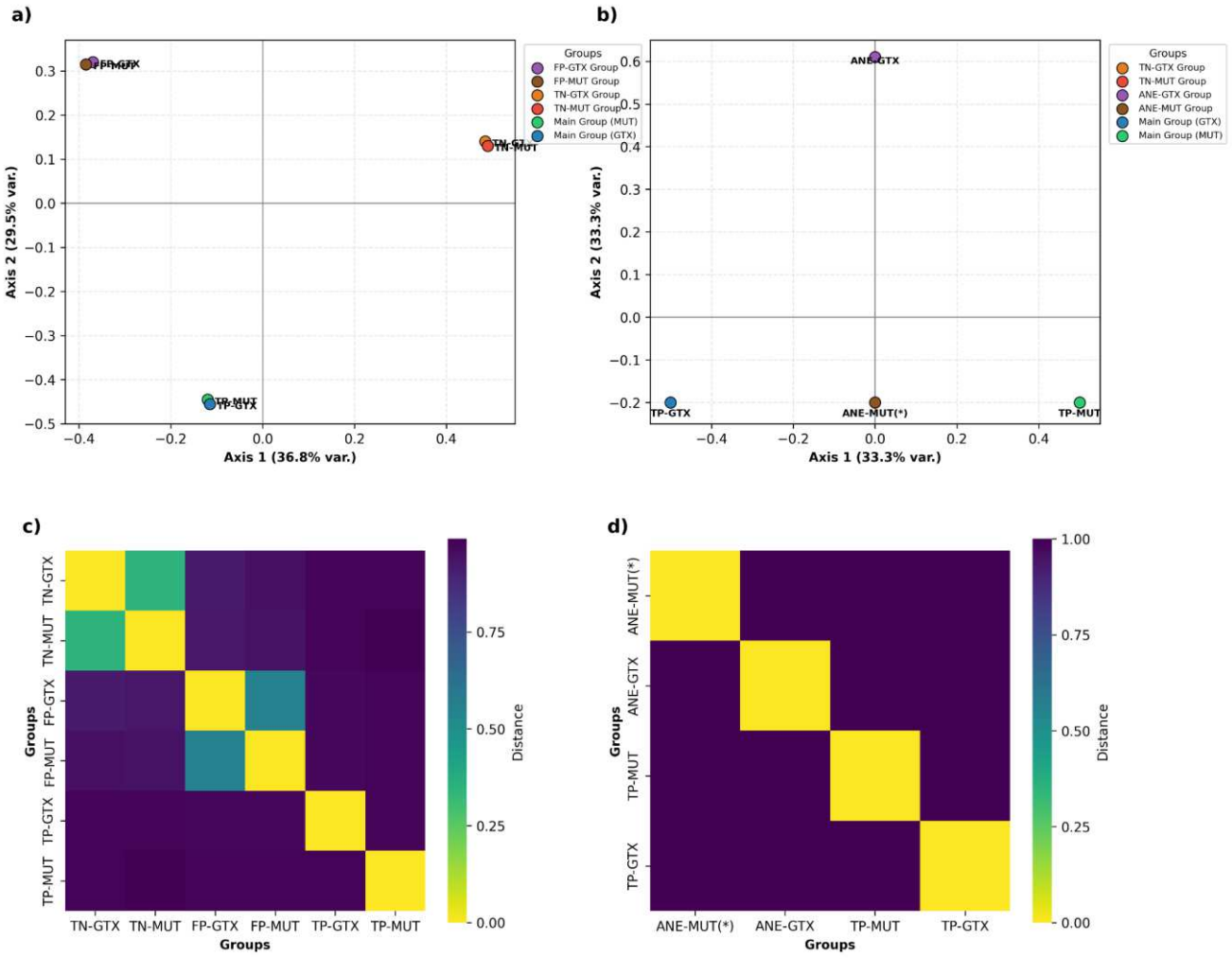


Figure 2. Discrimination of toxicological classes and genotoxic mechanisms based on refined proteomic profiles. (A) Principal Coordinates Analysis (PCoA) using Jaccard dissimilarity, illustrating the distribution of refined proteomic signatures across True Positive (TP), False Positive (FP), and True Negative (TN) groups. The first two principal coordinates capture 66.3% of the total variance (Axis 1: 36.8%; Axis 2: 29.5%). (B) Hierarchical clustering heatmap based on Jaccard dissimilarity, depicting pairwise similarities and dissimilarities among TP, FP, and TN profiles. (C) PCoA using Jaccard dissimilarity, demonstrating the separation of refined signatures associated with distinct genotoxic mechanisms, including broad genotoxicity (TP-GTX), mutagenicity (TP-MUT), and aneugenicity (ANE-GTX). The first two principal coordinates capture 66.6% of the total variance (Axis 1: 33.3%; Axis 2: 33.3%). The ANE-MUT* profile was retained for methodological comparison. (D) Hierarchical clustering heatmap based on Jaccard dissimilarity, displaying the structural relationships among the refined genotoxic mechanism-associated signatures. Yellow indicates higher similarity (lower dissimilarity), whereas dark purple indicates higher dissimilarity. Abbreviations: PCoA, Principal Coordinates Analysis; GTX, genotoxic carcinogenicity; MUT, mutagenicity; TP, True Positive; FP, False Positive; TN, True Negative; ANE, aneugenic compounds.

2.4. Exploratory Comparison of Proteomic Profiles Associated with Genotoxic Mechanisms

The ability to discriminate among distinct genotoxic mechanisms was evaluated by comparing the True Positive profiles with those of aneugenic compounds. Principal Coordinates Analysis (PCoA) using Jaccard dissimilarity demonstrated that the first two principal coordinates captured 66.6% of the total variance, with 33.3% attributed to each axis **(Figure 2c)**. Profiles associated with broad genotoxicity, mutagenicity, and aneuploidy exhibited distinct spatial separation.

Hierarchical clustering revealed greater dissimilarity between the aneuploidy-associated profile and the True Positive compound profiles **(Figure 2d)**. The ANE-MUT profile, derived from an in silico search using the term "mutagenicity" for aneugenic compounds, was retained for methodological evaluation and exhibited higher similarity to the analyzed chemical profiles. Collectively, these analyses identified differences in the organization of curated protein profiles associated with the predefined genotoxic mechanism categories. These findings support the generation of candidate mechanistic signatures but do not establish that the signatures are specific to the respective mechanisms or can independently classify new substances.

2.5. Minimal Overlap Among Chemical Classes Suggests Heterogeneity in Proteomic Signatures

Integrating profiles across distinct chemical classes within the True Positive group yielded 245 genotoxic carcinogenicity (GTX)-associated proteins and 153 mutagenicity (MUT)-associated proteins. Intersection analysis revealed minimal overlap among these chemical classes.

In the GTX set, approximately 89.4% of the identified proteins were exclusive to a single chemical class; transferrin (TF) was the only protein shared across all three evaluated classes **(Figure 3a)**. Conversely, no proteins were shared across all classes in the MUT set, resulting in a null central intersection **(Figure 3b)**. These findings suggest substantial heterogeneity among proteomic signatures derived from different chemical classes, indicating the lack of a universal mutagenicity signature within the computationally analyzed dataset.

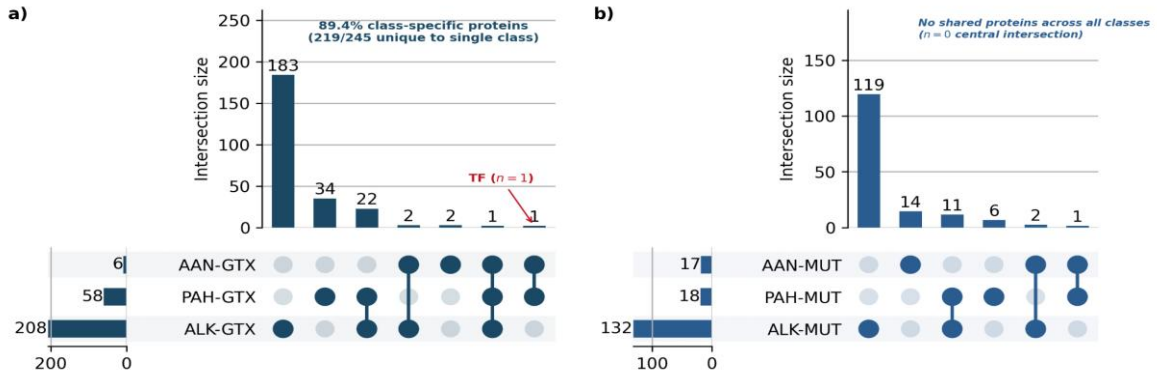


Figure 3. Intersection analysis of identified proteins across chemical classes. A) UpSet plot illustrating the distribution of 245 proteins from the genotoxic carcinogenicity (GTX) set across ALK-GTX, PAH-GTX, and AAN-GTX classes. Approximately 89.4% (219/245) of proteins are class-specific, with transferrin (TF, n = 1) identified as the sole protein in the triple intersection. B) UpSet plot depicting 153 proteins from the mutagenicity (MUT) set across ALK-MUT, PAH-MUT, and AAN-MUT classes, showing no shared proteins across all three classes (n = 0 central intersection).

2.6. Proteomic Signatures Exhibit Distinct Functional Organization Across Genotoxic Mechanisms

Protein–protein interaction (PPI) network and functional enrichment analyses were employed to investigate the biological organization of signatures associated with distinct genotoxic mechanisms.

The mutagenic/clastogenic signature network comprised 179 nodes and 415 interactions, exhibiting statistically significant enrichment ($p = 1.63 \times 10^{-13}$) **(Figure 4a)**. Key functional modules encompassed glutathione conjugation, ubiquitin activation and synthesis, and cytoplasmic stress granule formation **(Figure 4c)**. Conversely, the aneugenic network comprised 159 nodes and 816 interactions ($p < 1.0 \times 10^{-16}$). Its primary functional modules were associated with chromatin assembly, tubulin family proteins, and mitotic machinery components **(Figure 4b,c)**.

Functional enrichment analyses corroborated this distinction, highlighting a predominance of glutathione metabolism, detoxification, and stress response pathways in the mutagenic/clastogenic profile. In contrast, the aneugenic profile was significantly enriched for chromatin organization, nucleosome assembly, and cell division processes.

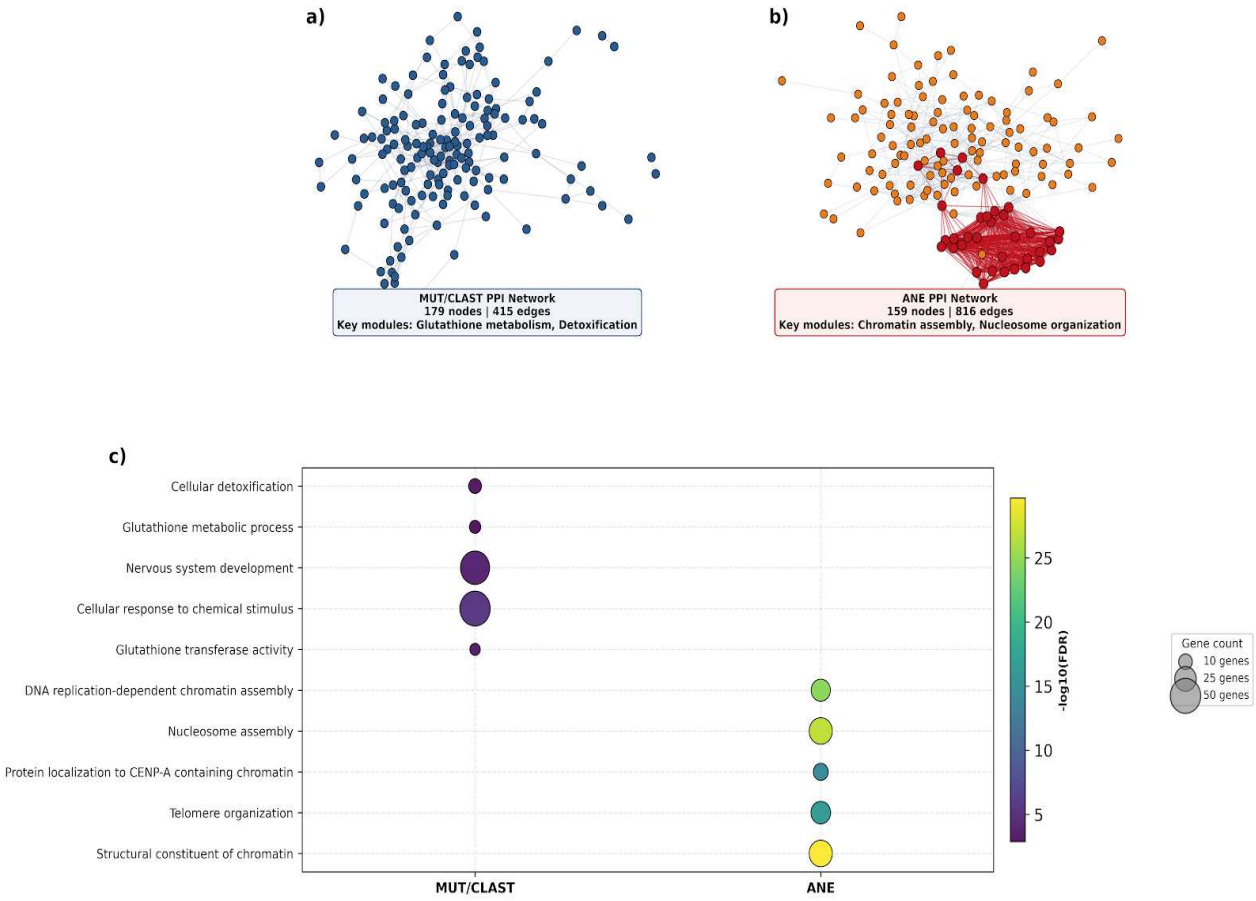


Figure 4. Functional network organization and comparative biological enrichment across distinct genotoxic mechanisms. A) Protein-protein interaction (PPI) network constructed using STRING and visualized in Python for mutagenicity/clastogenicity signatures (MUT/CLAST: 179 nodes, 415 edges), highlighting functional modules associated with glutathione conjugation, ubiquitin activation, and stress granule assembly. B) PPI network constructed using STRING and visualized in Python for aneugenic signatures (ANE: 159 nodes, 816 edges), displaying a dense topological cluster associated with chromatin assembly, tubulin components, and mitotic machinery. C) Comparative dot plot showing Gene Ontology (GO) biological processes enriched in MUT/CLAST *versus* ANE mechanisms. Circle size represents the number of identified proteins (gene count), and color intensity represents statistical significance, expressed as $-\log_{10}$(FDR). Abbreviations: PPI, protein-protein interaction; MUT/CLAST, mutagenicity/clastogenicity; ANE, aneugenicity; GO, Gene Ontology; FDR, false discovery rate.

2.7. Genotoxic Carcinogenicity Signature Exhibits Functional Coherence

Analysis of the genotoxic carcinogenicity-associated signature revealed a PPI network comprising 92 proteins and 198 interactions. This network exhibited significantly greater connectivity than expected by chance ($p = 1.99 \times 10^{-11}$), indicating robust functional coherence among the identified constituents. MCL clustering identified modules involved in glutathione conjugation, gap junction

formation, ribosomal components, PI3K/AKT signaling, mitochondrial processes, and the regulation of cell survival and cellular stress responses **(Figure 5a,d)**.

Functional enrichment analyses highlighted a predominance of pathways linked to glutathione metabolism, glutathione transferase activity, and mitochondrial organization. Significant enrichment was also observed for cellular components mapping to extracellular vesicles ($FDR = 7.95 \times 10^{-19}$) and exosomes ($FDR = 8.33 \times 10^{-19}$), alongside the outer mitochondrial membrane ($FDR = 1.0 \times 10^{-4}$) **(Figure 5d).** Furthermore, pathway analysis indicated a convergence toward xenobiotic biotransformation and adaptive chemical stress responses, encompassing cytochrome P450-mediated metabolism, glutathione conjugation, and chemical carcinogenesis pathways **(Figure 5c).**

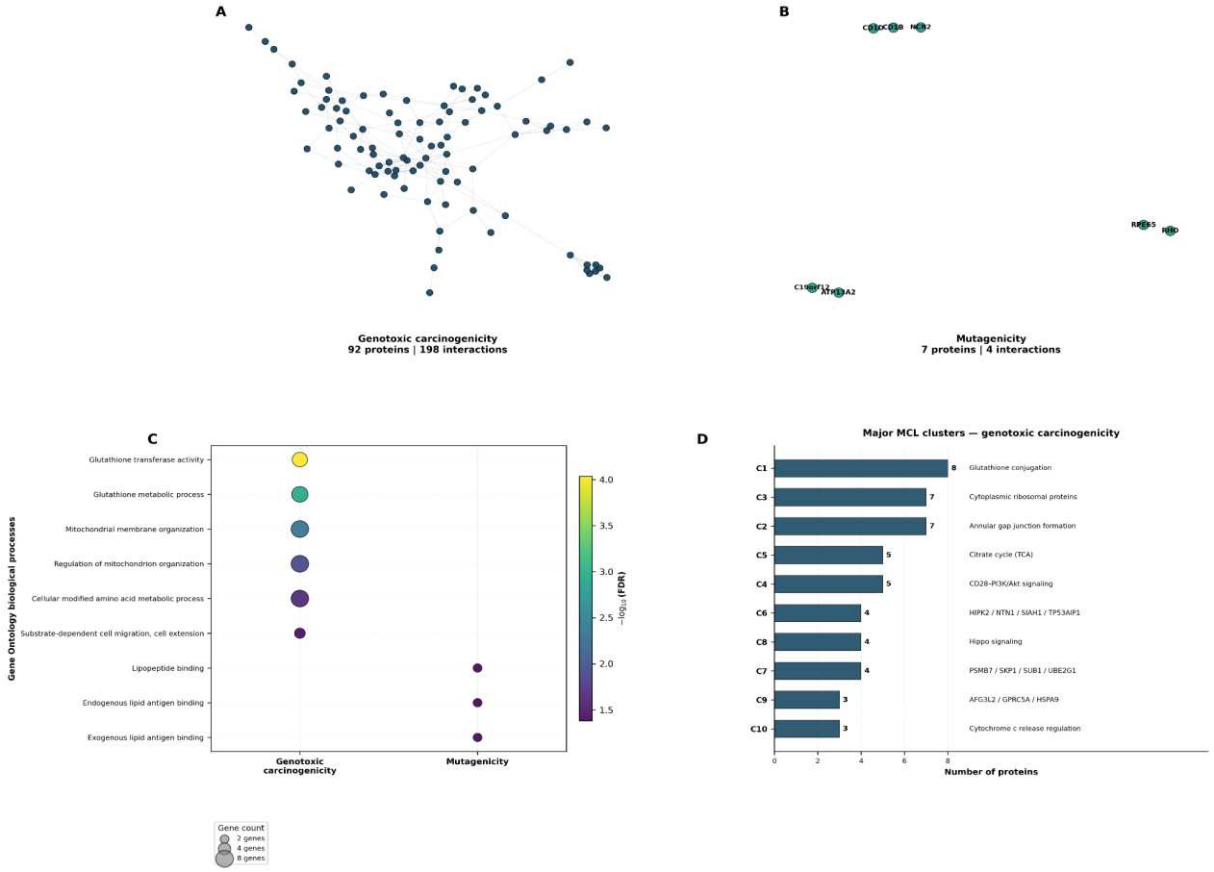


Figure 5. Functional organization of proteomic signatures associated with genotoxic carcinogenicity and mutagenicity. (A) Protein–protein interaction (PPI) network of the genotoxic carcinogenicity-associated signature (92 nodes, 198 edges). The network exhibits significantly greater connectivity than expected by chance ($p = 1.99 \times 10^{-11}$) and highlights enriched pathways linked to glutathione metabolism, mitochondrial organization, and xenobiotic metabolism. (B) PPI network of the mutagenicity-associated signature (7 nodes, 4 edges), displaying functional enrichment for lipid antigen binding and immunoregulatory interactions. (C) Functional enrichment analysis comparing both signatures, highlighting selected Gene Ontology (GO) biological processes and molecular functions. (D)

MCL clustering of the genotoxic carcinogenicity-associated PPI network, revealing functional modules involved in glutathione conjugation, gap junction formation, ribosomal components, PI3K/AKT signaling, mitochondrial processes, and cellular stress responses. PPI network construction and functional enrichment analyses were performed using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) version 12.0.

3. DISCUSSION

This study established an *in silico* proof of concept (PoC) for the EV-PROTEOMICS platform, designed to investigate whether extracellular vesicles (EVs) exhibit proteomic signatures capable of discriminating among different MoAs related to genotoxicity and carcinogenicity. The results provide exploratory evidence that the refined protein sets exhibit different patterns of organization across the predefined genotoxic, non-genotoxic carcinogenic, and aneugenic reference categories. These observations support the feasibility of using computational refinement to generate candidate mechanism-associated signatures, although their specificity and discriminatory performance remain to be independently established.

Exploratory analyses using PCoA and hierarchical clustering initially revealed that the data exhibited non-random structural organization. The observed spatial separation delineated distinct patterns associated with biological conditions and MoAs. Along Axis 1, genotoxic carcinogen and mutagen profiles clustered with their respective class controls, whereas the TN and FP groups exhibited distinct topological patterns. For AAN, ALK, ANE, and PAH, the GTX *versus* MUT condition appeared to drive profile organization more strongly than the chemical identity of the class itself. Consequently, for these classes, the generalized genotoxic or mutagenic response likely constitutes a more robust shared proteomic signal than the specific chemical identity of the compounds.

This observation formed the rationale for the pipeline's refinement strategy, which relied on comparisons among biologically defined groups and the application of filters designed to attenuate non-specific signals. Subsequent analyses demonstrated that this refinement concentrated the proteomic signal into signatures with enhanced discriminatory capacity across the evaluated MoAs.

A primary finding of this PoC was the computational capacity to distinguish genotoxic (TP) from non-genotoxic (FP) carcinogens utilizing EV proteomic profiles. The PPI network constructed from the 119 TP-specific proteins exhibited highly significant functional integration ($p = 1.99 \times 10^{-11}$), demonstrating that these constituents do not represent a random assembly, but rather a biologically cohesive network.

Functional analysis of the TP signature identified modules enriched for xenobiotic metabolism, glutathione pathways, and CYP450 activity. These processes align with cellular responses to compounds that undergo bioactivation and generate reactive species. Glutathione is a central component of cellular redox defense, mediating xenobiotic detoxification and reactive oxygen species (ROS) neutralization, largely via glutathione S-transferases [18- 20]. The prominent representation of these pathways in the genotoxic carcinogen signature is thus biologically consistent with a robust defense response against chemical and oxidative stress.

The observed response also encompassed processes linked to TP53-mediated DNA damage responses, PI3K/AKT signaling, the Hippo pathway, and fatty acid metabolism. This multifaceted profile suggests that the TP signature does not exclusively represent a detoxification response, but rather integrates diverse components of the cellular reaction to genotoxic stress. The TP53 axis orchestrates cell-cycle arrest, DNA repair, and apoptosis, whereas PI3K/AKT signaling promotes cell survival. The crosstalk between these pathways constitutes a critical component of the cellular response to genetic damage [21-23].

The enrichment of xenobiotic metabolism processes is particularly consistent with exposure to genotoxic carcinogens requiring metabolic bioactivation. CYP450 enzymes generate electrophilic intermediates and ROS, which are subsequently neutralized by the phase II response, including the glutathione system (18, 24, 25]. Therefore, the identified signature likely reflects the integration of bioactivation, oxidative stress, and coordinated cellular defense mechanisms.

In contrast, the non-genotoxic carcinogen (FP) signature exhibited a distinctly different functional profile. While both groups displayed oxidative stress-associated responses, the FP group showed significantly stronger enrichment for protein folding and proteolysis—including ubiquitin-mediated proteolysis—as well as pathways involving proteoglycans in cancer and cell–cell junction assembly. This distinction is critical, as it indicates that oxidative stress alone is insufficient to explain the computational discrimination observed between the groups.

Consequently, the results suggest that EV-PROTEOMICS does not act merely as a non-specific oxidative stress sensor. In the TP group, oxidative stress co-occurred with xenobiotic metabolism and genomic damage responses, whereas in the FP group, it was linked to protein homeostasis and cellular remodeling. Proteotoxic stress and disrupted protein homeostasis can drive non-genotoxic carcinogenesis via mechanisms distinct from direct DNA sequence damage [26].

This functional divergence aligns with the mechanistic heterogeneity of non-genotoxic carcinogens. These agents promote carcinogenesis through alterations in gene expression and cell growth regulation, encompassing chronic proliferation, hormonal imbalances, oxidative stress generation, disrupted

intercellular communication, and epigenetic modifications [27]. Approximately 12% (45/371) of carcinogens classified by the IARC in Groups 1, 2A, and 2B are estimated to be non-genotoxic [28].

Analysis of the mutagenicity signature further reinforced this distinction. Whereas the TP-MUT group exhibited a reduced signature (13 proteins) associated primarily with lipid antigen presentation and natural killer (NK) cell activation pathways, the FP-MUT group mounted a much broader response (1,340 proteins) involving cytokine, TNF, interleukin, and complement pathways. Although these findings indicate distinct functional profiles, the interpretation of these biological processes must remain hypothesis-generating, given the *in silico* nature of the analysis.

Therefore, the comparison between TP and FP indicates that the curated and refined protein profiles exhibit different functional associations within the analyzed reference dataset. These differences generate hypotheses regarding the molecular processes represented by the respective profiles. However, the present analysis does not establish that EV-PROTEOMICS can reliably classify genotoxic and non-genotoxic carcinogens outside this dataset.

Beyond detoxification and DNA damage responses, the TP-GTX signature was significantly enriched for processes governing membrane organization and the outer mitochondrial membrane. This suggests that the EV-captured response encompasses secondary effects of genotoxic exposure, extending beyond direct genetic damage mechanisms.

The interplay among oxidative stress, mitochondrial dysfunction, genotoxicity, and carcinogenesis is well established [29-31]. Elevated mitochondrial oxidative stress contributes to DNA damage and genomic instability [32]. For instance, exposure to polycyclic aromatic hydrocarbons (PAHs) can disrupt redox homeostasis and induce mitochondrial oxidative damage, altering mitochondrial DNA copy numbers [32].

Furthermore, mitochondrial alterations are linked to the metabolic reprogramming characteristic of tumor cells. The Warburg effect—defined by a preference for aerobic glycolysis—drives this metabolic adaptation [31, 33]. While these mechanisms provide biological context for the enriched processes, our findings do not establish a causal link between the EV signature and these metabolic shifts.

The enrichment of pathways mediating platinum-based drug resistance and chemical carcinogenesis further contextualizes the TP signature within oncology. Activating defense systems, such as glutathione metabolism, can confer resistance to genotoxic chemotherapeutics [34]. Consequently, the identified signature may reflect an initial adaptive response sharing components with mechanisms later exploited by tumor cells.

The true negative (TN) group served as a critical control to evaluate the platform's ability to distinguish non-specific cellular stress from genuine genotoxic responses. This group included TBHQ and menthol—compounds classified as non-carcinogenic and non-genotoxic *in vivo*, despite occasionally yielding positive results under specific *in vitro* conditions [35, 36].

The TN signature featured pathways involved in AGE-RAGE signaling, glutathione metabolism, NFE2L2 degradation, and SOD1 activity. This profile indicates a redox response lacking the TP53 axis activation characteristic of genotoxic carcinogens. TBHQ activates the NFE2L2/NRF2 pathway and phase II enzymes (e.g., GST and NQO1), thereby inducing antioxidant defense mechanisms [37, 38]. Similarly, menthol elicits antioxidant responses, increasing SOD and glutathione levels alongside Nrf2 activation [37, 39].

Accordingly, detecting oxidative stress responses in the TN group without a concurrent genomic damage signature underscores the platform's specificity. It demonstrates that redox stress alone is insufficient to classify an exposure as genotoxic.

However, the identification of processes such as "shear stress" may reflect inherent limitations in the underlying databases and semantic protein annotations. This highlights the necessity of cautiously interpreting *in silico* results and experimentally validating prioritized biomarkers.

A second critical dimension of the PoC was discriminating between mutagenic/clastogenic and aneugenic agents. The ANE-GTX signature was significantly enriched for structural proteins governing cell division, notably tubulins and histones, which organized into discrete modules mediating chromatin assembly and tubulin dynamics.

This profile directly aligns with the aneugenic mode of action, wherein disrupted microtubule and mitotic spindle dynamics impair chromosome segregation, leading to micronucleus formation. Classic aneugens, such as colchicine and paclitaxel, operate via this exact mechanism [40- 42].

In contrast, the TP-GTX signature featured pathways driving glutathione conjugation, ubiquitin synthesis, and cytoplasmic stress granule localization. Therefore, while the aneugenic signature was structurally oriented toward cell division and chromatin architecture, the TP signature was predominantly defined by metabolic and cellular stress responses.

This distinction is particularly valuable because the *in vitro* micronucleus assay (OECD TG 487; 2023) is a phenotypic test that cannot intrinsically specify the micronucleus's origin. Distinguishing between clastogenic and aneugenic mechanisms traditionally requires complementary tools, such as FISH or CREST [43- 44].

Toxicologically, this differentiation is crucial because mutagenicity, clastogenicity, and aneugenicity represent fundamentally distinct genetic damage mechanisms [44- 46]. Clarifying the MoA directly informs risk assessment: aneugenic effects can sometimes support the establishment of health-based reference values, whereas clastogenic and mutagenic effects pose far greater challenges for defining safety thresholds [44, 47].

Principal Coordinates Analysis (PCoA) corroborated this differentiation. The genotoxicity MoA-specific PCoA captured 66.6% of the total variance. The ANE-GTX group occupied a distinct spatial position, while the TP-GTX, TP-MUT, and ANE-MUT(*) profiles clustered closely together. The Jaccard dissimilarity heatmap revealed a concordant pattern, highlighting profound dissimilarity between the ANE-GTX group and the other profiles.

Collectively, these results suggest that the curated protein profiles associated with aneugenic reference substances exhibit functional associations that differ from those observed in the mutagenic/clastogenic reference profiles. These patterns provide a basis for further mechanistic investigation but require independent validation before they can be interpreted as mechanism-specific signatures.

Although the TP-MUT signature exhibited reduced dimensionality (13 proteins and four interactions), it showed statistically significant enrichment in the PPI analysis ($p = 0.0436$). The identified modules included processes related to antigen presentation, the retinoid cycle, and the proteins ATP13A2 and C19orf12.

The predominance of CD1 proteins, particularly CD1D and CD1B, alongside NCR2, suggests an immune response associated with mutagenic stress. CD1D and CD1B are involved in lipid antigen presentation, whereas NCR2 is associated with NK cell activation. Current literature indicates that NK cells and unconventional T cells may play distinct roles in tumor elimination or progression, depending on the signaling context [48].

Oxidative stress and lipid peroxidation linked to genotoxic exposure likely contribute to immunoregulatory signals detectable in EVs. In this context, EVs could act as carriers of proteins, lipids, and other components mediating cellular communication.

However, this interpretation represents a pipeline-generated hypothesis rather than causal evidence. The absence of classical DNA repair pathways may partially result from the relevance filter itself, as nuclear proteins are generally less represented in EV databases like Vesiclepedia and ExoCarta [49]. Furthermore, immune pathway enrichment may reflect semantic correlations derived from the data mining literature. Therefore, while CD1D, CD1B, and NCR2 remain promising candidates for validation, the current data cannot conclusively establish this axis as specific to mutagenicity. Further experimental validation is required.

Comparing the exploratory and refined analyses provided further evidence supporting the subtractive workflow's utility. The initial PCoA explained 31.2% of the total variance, whereas the refined analysis, following filter application and group organization by MoA, explained 66.3%.

The increase in the proportion of variance represented by the first two PCoA axes indicates a change in the structure of the dissimilarity matrix following refinement. However, because the initial and refined analyses involve different feature sets and potentially different profile structures, these percentages should not be interpreted as a direct measure of improved discriminatory performance. The observed spatial organization is therefore considered exploratory.

A similar outcome emerged in the genotoxicity-specific analysis. The PCoA explained 66.6% of the variance and distinctly separated ANE-GTX from the other groups. The ANE-MUT(*) profile did not cluster with its expected biological group (ANE-GTX), but rather with TP-GTX and TP-MUT. We interpreted this behavior as a potential artifact reflecting general genotoxicity rather than an aneugenicity-specific signature.

Removing the ANE-MUT(*) signal via the corresponding filter acts as a signature-purification step, though it requires subsequent experimental validation. Similarly, the convergence toward individual markers in certain groups, such as NDRG2 for aromatic amines and ITGB7 for PAH-MUT, highlights candidate prioritization by the pipeline rather than the discovery of universal biomarkers.

Chemical class analysis demonstrated that the filters used to differentiate between genotoxic/non-genotoxic carcinogenicity and mutagenicity/clastogenicity versus aneugenicity provided complementary insights.

Alkylating agents displayed complex networks in the genotoxic carcinogenicity analysis, including pathways linked to the p53/HIPK2 axis and PI3K/Akt. When filtered specifically for the mutagenic/clastogenic MoA, the network remained extensive and showed enrichment for the hematopoietic cell lineage. This aligns with the known relevance of myelosuppression in alkylating agent toxicity, though this pipeline-identified association serves as *in silico* mechanistic evidence.

PAHs exhibited a signature dominated by glutathione conjugation, reflecting xenobiotic metabolism involvement (Bansal & Simon, 2018). Modules involving RNF2, SUB1, and SUMO2 suggest an interface between the genotoxic response and epigenetic regulation, corroborating PAH carcinogenicity models that involve both DNA adduct formation and epigenetic alterations [50]. The presence of ribosomal proteins also indicated ribotoxic stress and potential p53–RPL5/RPL11–MDM2 axis engagement [51].

However, upon refinement for the mutagenic/clastogenic MoA, the PAH network lost statistical significance ($p = 0.21$), leaving only a weak inflammatory signal centered on IL1B and P2RX7. This implies that the primary proteomic response to PAHs links to broad genotoxic carcinogenicity and metabolic stress, while the specific mutagenic/clastogenic component is subtler or masked.

Aromatic amines/nitrosamines (AANs) displayed a distinct pattern. While their profile in the genotoxic carcinogenicity analysis was restricted (including NDRG2 and NFE2L2), refinement for the mutagenic/clastogenic MoA revealed a highly interconnected, statistically significant network ($p = 2.56 \times 10^{-6}$). This network enriched for tumor suppressors and proto-oncogenes, including APC, FHIT, and MYC. Ultimately, multilevel analysis successfully isolated distinct proteomic response components for each chemical class.

These results support the potential of EVs as a source of mechanistic biomarkers within a New Approach Methodologies (NAMs) toxicological framework. The primary contribution of this PoC is not defining universal signatures for each class, but demonstrating that an integrated workflow, combining data mining, filtering, and functional analysis of EV-associated proteins, can generate discriminative patterns for various MoAs.

We propose a conceptual framework for future investigation in which EV-associated proteomic data could be evaluated for broad genotoxicity-associated patterns, followed by the exploration of candidate signatures associated with distinct mechanisms. This sequence represents a prospective development strategy rather than a validated classification workflow. Its feasibility will depend on independent datasets, experimental confirmation, and evaluation of performance across chemically diverse substances.

A primary limitation of this proof of concept is the limited number of analyzed compounds. The workflow evaluated 12 substances selected from established reference lists, restricting generalizability and precluding the interpretation of the resulting signatures as universal chemical-class profiles. The findings should therefore be interpreted as exploratory separation among predefined toxicological reference categories. Broader chemical coverage and evaluation using independent substances are necessary to assess the robustness and discriminatory performance of the resulting profiles.

Additionally, functional associations derived from *in silico* protein–protein interaction networks, enrichment analyses, and semantic mining cannot establish causality. This limitation is particularly relevant when interpreting the TP-MUT signature and the ANE-MUT(*) result, for which database representation and annotation biases may influence the observed patterns.

Some groups also yielded signatures containing very few proteins, in some cases converging on individual markers. Although such reduction may support candidate prioritization, isolated markers

cannot independently characterize a mode of action (MoA). Experimental studies are required to establish their detectability in EV samples, assess their behavior following chemical exposure, and determine whether they contribute to reproducible discrimination among reference categories.

An additional limitation is that the proprietary refinement algorithm is not fully disclosed in this publication. Although the general analytical rationale and downstream analyses are described, the omission of implementation-specific decision rules limits independent reconstruction of the complete refinement procedure. Accordingly, the findings should be interpreted as an evaluation of the resulting curated profiles and their exploratory organization, rather than as an independent validation of the complete computational platform.

This study provides an in silico proof of concept for computational refinement of curated EV-associated protein profiles and prioritization of candidate biomarkers for mechanistic toxicology research. The analyses identified patterns associated with the predefined toxicological reference categories and prioritized candidates, including NDRG2 and NFE2L2, for further investigation. These findings remain exploratory: neither the candidate biomarkers nor the discriminatory or predictive performance of the complete EV-PROTEOMICS platform has been independently validated.

Further studies should assess the robustness of the identified patterns using independent substances and datasets and experimentally investigate candidate proteins in EV samples following chemical exposure. Such work will be necessary to determine whether the approach can contribute to human biology-based New Approach Methodologies (NAMs) for toxicological assessment.

4. Materials and Methods

This section describes the development and application of the computational component of the integrated EV-PROTEOMICS platform, based on a strategy for refining proteomic profiles to discriminate among different genotoxic modes of action (MoAs) and investigate molecular signatures associated with carcinogenic potential. Overall, the computational component was structured into sequential stages of proteomic profile processing, refinement, and interpretation (Figure 6).

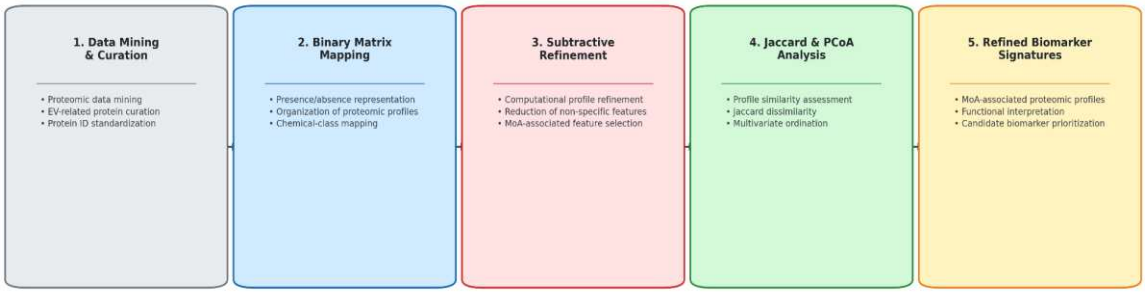


Figure 6. General workflow of the computational component of the EV-PROTEOMICS platform. The computational component was structured into sequential steps involving proteomic data mining and curation, organization of profiles into binary matrices, computational data refinement, similarity analysis and multivariate ordination, and definition of proteomic signatures associated with the investigated modes of action. The final step includes functional assessment of the signatures and prioritization of potential biomarkers.

4.1 Selection of the Reference Set

The framework's performance was evaluated using a set of 12 chemical substances (Table 1), selected based on reference criteria established by Kirkland et al. (2008, 2016). These substances were stratified into four toxicological and mechanistic reference groups:

- **True Positives (TP):** Reference substances with established evidence of genotoxicity, mutagenicity, and carcinogenicity;
- **False Positives (FP):** Substances that induce carcinogenesis via non-genotoxic mechanisms, serving as a mechanistic contrast group;
- **Aneugens (ANE):** Genotoxic agents used as references for mechanisms involving altered chromosome segregation or mitotic spindle dynamics;
- **True Negatives (TN):** Substances used as specificity controls, classified as negative for genotoxicity and *in vivo* carcinogenicity according to the adopted criteria.

The TP, FP, ANE, and TN designations were used exclusively as reference categories for data organization and computational comparisons in this study; they do not reflect diagnostic assay results or classifications generated by the proposed framework.

Table 1. Groups of substances used for *in silico* evaluation of the computational component of the EV-PROTEOMICS platform.

Group	Substances	Mechanistic category	Use in the analysis
I. True Positives (TP)	**Alkylating agents:** Cyclophosphamide, EMS. **Aromatic hydrocarbons:** 2-aminoanthracene, Benzo(a)pyrene. **Aromatic amines/Nitrosamines:** Dimethylnitrosamine (NDMA), 2-amino-3-methylimidazo[4,5-f]quinoline (IQ).	Genotoxicity/mutagenicity associated with carcinogenicity	Positive reference
II. False Positives (FP)	Progesterone, Methyl carbamate	Carcinogenicity via non-genotoxic mechanisms	Mechanistic contrast
III. Aneugens (ANE)	Colchicine, Taxol (Paclitaxel)	Aneugenicity	Reference for MoA discrimination
IV. True Negatives (TN)	L-Menthol, TBHQ (tert-butylhydroquinone)	No evidence of genotoxicity/carcinogenicity under the employed reference conditions	Negative control/specificity

Adapted [52].

4.2 Mining and Curation of Proteomic Data

Proteomic data were extracted from the GeneCards and UniProt databases using structured queries that combined the names of the evaluated substances with terms related to genotoxicity, DNA damage, and mutagenicity. The retrieved data underwent sequential curation to standardize protein identifiers and select proteins with experimental evidence of presence in extracellular vesicles, cross-referencing the Vesiclepedia and ExoCarta databases. Curated data were organized into individual proteomic profiles and subsequently consolidated according to the predefined toxicological groups. The integration, prioritization, and filtering procedures applied post-curation are conceptually described in Section 2.3, without disclosing the specific operational criteria that comprise the proprietary EV-PROTEOMICS architecture.

4.3. Computational Strategy for Proteomic Refinement

The computational framework operates as a proteomic refinement strategy designed to reduce shared features across toxicological categories while prioritizing those specific to the investigated mechanisms. Profiles were organized by predefined toxicological categories and subjected to sequential comparative filtering against reference sets. This sequential process generated progressively refined signatures tailored to discriminate among the evaluated toxicological profiles.

This approach was applied across two primary mechanistic axes: (i) distinguishing True Positive (TP) from False Positive (FP) profiles to contrast genotoxic versus non-genotoxic carcinogenicity; and (ii) differentiating mutagenic/clastogenic from aneugenic profiles using the TP and ANE groups as mechanistic references. Complementary analyses evaluated signature organization and specificity

across distinct chemical classes within the TP group. The specific decision criteria, operational sequence, and algorithmic parameters constitute proprietary elements of the EV-PROTEOMICS architecture and are omitted from this report.

4.4. Quantitative Workflow of the Refinement Pipeline

To ensure workflow traceability, protein counts were tracked and recorded at every analytical stage, from initial profile consolidation by chemical class to final signature refinement. Data for each processing stage are provided in Table S1 (Supplementary Material). Downstream results, including retained protein counts for the GTX and MUT signatures and their mechanistic comparisons, are detailed in the Results section.

4.5. Computational Representation of Proteomic Profiles

4.5.1 Definition of Proteomic Sets

Let $S=\{s_1,\ldots,s_n\}$ denote the set of substances. Each substance s_i is associated with a set of proteins P_i. Proteomic profiles for the toxicological groups were generated by consolidating the proteins identified for all substances within each group. For subsequent analyses, these profiles were treated as discrete sets of molecular features, enabling comparative analyses based on protein presence or absence. For each toxicological group G_k, the consolidated proteomic profile was defined as the union of the protein sets associated with its constituent substances:

$$P_{G_k} = \bigcup_{s_i \in G_k} P_i$$

Consequently, each toxicological profile was represented as a discrete set of unique proteins identified post-curation. This representation facilitated the comparison of proteomic compositions across groups based strictly on protein presence or absence.

4.5.2 Binary Matrix Representation

The proteomic profiles were encoded into a binary presence/absence matrix, where rows represent individual proteins and columns represent toxicological profiles or reference groups. For each protein p_j and profile G_k, the matrix elements X_{jk} were defined as follows:

$$x_{jk} = \begin{cases} 1, & \text{if protein } p_j \in G_k, \\ 0, & \text{otherwise.} \end{cases}$$

This formatting enabled comparative analyses independent of relative protein abundance, relying exclusively on protein occurrence across the evaluated sets.

4.5.3 Jaccard Index

Proteomic profile similarity was assessed using the Jaccard index, which quantifies the overlap between distinct sets. For any two sets A and B, the Jaccard coefficient is defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

$$d_J(A,B) = 1 - J(A,B)$$

These values were used to characterize proteomic profile similarities and subsequently construct a dissimilarity matrix for ordination analysis.

4.5.4 Principal Coordinates Analysis (PCoA)

The resulting dissimilarity matrix was subjected to Principal Coordinates Analysis (PCoA) to project the relationships among profiles into a lower-dimensional space. This exploratory analysis facilitated the visualization of similarity and dissimilarity patterns across the investigated toxicological categories.

4.6 Computational Implementation

All analyses were performed in Python 3.10 using the NumPy, SciPy, pandas, scikit-learn, and Matplotlib libraries. The computational pipeline was executed within a Google Colab/Jupyter Notebook environment. The pipeline was modularly structured, encompassing data mining and curation, proteomic profile construction, set refinement, binary representation, similarity analysis, and ordination. The algorithms, parameters, decision criteria, and integration rules governing the proprietary stages of the pipeline are omitted from this report. Only the methodological principles and analyses essential for data interpretation are presented herein.

4.7 Network Analysis and Functional Enrichment

Protein interaction networks were analyzed using STRING (version 12.0), with *Homo sapiens* specified as the reference organism (NCBI taxonomy identifier: 9606). The minimum required interaction confidence score was set to 0.400 (medium confidence). Default STRING evidence settings were retained. For multiple-protein queries, the network analysis was based on associations among the submitted proteins, without the intentional addition of first- or second-shell interactors. The numbers of

input proteins and proteins represented in the resulting networks were considered separately when interpreting network topology and protein–protein interaction (PPI) enrichment.

Functional enrichment analyses were performed using the STRING functional enrichment module, covering Gene Ontology Biological Process (GO-BP), KEGG pathways, and Reactome pathways. Over-representation was assessed using a hypergeometric test, and p-values were adjusted for multiple testing using the Benjamini–Hochberg procedure. Terms with an adjusted false discovery rate (FDR) below 0.05 were considered statistically significant. The default species-specific proteome was used as the reference background. Enrichment analyses were based on the submitted protein sets, without intentional expansion of the input lists with additional network interactors. Functional enrichment results were interpreted separately from PPI enrichment statistics, which assess whether the observed number of network associations exceeds the number expected by chance.

Protein profiles were curated using GeneCards, UniProt, Vesiclepedia, and ExoCarta, with queries and data retrieval conducted during 2025. Exact database retrieval dates and release versions are reported only where recoverable from the original records. Protein identifiers were harmonized where possible to support cross-database comparisons. Evidence that a protein had previously been reported in association with extracellular vesicles was interpreted as evidence of prior EV association, rather than as direct evidence of exposure-induced changes in EV protein abundance.

The proprietary signature-refinement procedure is described in terms of its inputs, general analytical stages, and outputs. Internal decision rules and implementation-specific parameters are not disclosed; conventional database resources, statistical procedures, and network-analysis settings are reported to support critical assessment of the results.

5. Competing interests

The authors declare no conflict of interest (COI) that could inappropriately influence the work reported in this paper.

6. Data Availability

The executable demonstration version of the EV-PROTEOMICS framework is provided for evaluation by the reviewers. The underlying protein datasets are not publicly available but may be made available in a blinded format by the corresponding author upon reasonable request.

7. Author Contributions

N.S.O.N.: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **A.O.C.:** Data curation, Formal analysis, Funding acquisition, Writing – review & editing. **W.B.P.G.:** Data curation, Formal analysis, Writing – review & editing. **B.M.R.:** Data

curation. **A.B.S.F.:** Data curation. **C.A.T.:** Conceptualization, Supervision, Writing – review & editing. All authors read and approved the final version of the manuscript.

8. Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Gemini (Google) and ChatGPT in order to enhance the clarity, grammar, and formal coherence of the manuscript, particularly in refining non-native English expressions for scientific precision, and to structure the Declaration and Acknowledgement sections according to editorial standards. After using this tool/service, the author(s) reviewed and edited the content as needed, performing rigorous scientific accuracy checks, and take(s) full responsibility for the content of the published article.

REFERENCES

1. Ball, N. et al. A framework for chemical safety assessment incorporating new approach methodologies within REACH. *Arch. Toxicol.* **96**, 743-766 (2022).
2. Smirnova, L. et al. 3S—systematic, systemic, and systems biology and toxicology. *ALTEX* **35**, 139-162 (2018).
3. U.S. Food and Drug Administration. *Roadmap to reducing animal testing in preclinical safety studies*. (U.S. Department of Health and Human Services, 2025).
4. U.S. Food and Drug Administration. *General considerations for the use of new approach methodologies in drug development: draft guidance for industry*. (U.S. Department of Health and Human Services, 2026).
5. European Commission. *Roadmap towards phasing out animal testing for chemical safety assessments*. Communication from the Commission C(2026)3497 final. (European Commission, 2026).
6. Nascimento, N. S. O., Costa, A. O., Rebuiti, B. M. & Tagliati, C. A. Mechanistic biomarkers in genotoxicity testing: a systematic mapping review of their integration with apical endpoints. *Toxicol. Lett.* **421**, 111921 (2026).
7. Ankley, G. T. et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.* **29**, 730-741 (2010).
8. OECD. *Guidance document for the scientific review of adverse outcome pathways*. OECD Series on Testing and Assessment No. 344 (OECD Publishing, 2021).
9. OECD. *Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA)*. OECD Series on Testing and Assessment No. 260 (OECD Publishing, 2017).
10. Welsh, J. A. et al. Minimal information for studies of extracellular vesicles (MISEV2023): from basic to advanced approaches. *J. Extracell. Vesicles* **13**, e12404 (2024).

11. Wang, Z. & Lieberman, P. M. The crosstalk of telomere dysfunction and inflammation through cell-free TERRA containing exosomes. *RNA Biol.* **13**, 690-695 (2016).
12. Kawata, R. et al. Exploration of small RNA biomarkers for testicular injury in the serum exosomes of rats. *Toxicology* **440**, 152490 (2020).
13. Thakur, K. et al. Extracellular vesicle-derived DNA vs. cfDNA as a biomarker for the detection of colon cancer. *Genes (Basel)* **12**, 1171 (2021).
14. Parolini, I. et al. Intraluminal vesicle trafficking is involved in the secretion of base excision repair protein APE1. *FEBS J.* **291**, 2849-2875 (2024).
15. Tkach, M. & Théry, C. Communication by extracellular vesicles: where we are and where we need to go. *Cell* **164**, 1226-1232 (2016).
16. Koeneman, S. H. & Cavanaugh, J. E. An improved asymptotic test for the Jaccard similarity index for binary data. *Stat. Probab. Lett.* **184**, 109375 (2022).
17. Nascimento, N. S. O. et al. Integração de abordagens computacionais e experimentais para desenvolvimento de uma plataforma NAM mecanística em toxicologia genética. *BIOINFO* **5**, 5 (2025).
18. Bansal, A. & Simon, M. C. Glutathione metabolism in cancer progression and treatment resistance. *J. Cell Biol.* **217**, 2291-2298 (2018).
19. Nissar, S., Sameer, A. S., Rasool, R., Chowdri, N. A. & Rashid, F. Glutathione S-transferases: biochemistry, polymorphism and role in colorectal carcinogenesis. *J. Carcinog. Mutagen.* **8**, 1000287 (2017).
20. Valenti, G. E., Tasso, B., Traverso, N., Domenicotti, C. & Marengo, B. Glutathione in cancer progression and chemoresistance: an update. *Redox Exp. Med.* **2023** (2023).
21. Chen, M. et al. A p53-phosphoinositide signalosome regulates nuclear AKT activation. *Nat. Cell Biol.* **24**, 1099–1113 (2022).
22. Duan, L. & Maki, C. G. The IGF-1R/AKT pathway determines cell fate in response to p53. *Transl. Cancer Res.* **5**, 664-675 (2016).
23. Abraham, A. G. & O'Neill, E. PI3K/Akt-mediated regulation of p53 in cancer. *Biochem. Soc. Trans.* **42**, 798-803 (2014).
24. Guengerich, F. P. Cytochrome P450 and chemical toxicology. *Chem. Res. Toxicol.* **21**, 70-83 (2008).
25. Jennings, P., Limonciel, A., Felice, L. & Leonard, M. O. An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.* **87**, 49-72 (2013).
26. Deshaies, R. J. Proteotoxic crisis, the ubiquitin-proteasome system, and cancer therapy. *BMC Biol.* **12**, 94 (2014).
27. Benigni, R., Bossa, C. & Tcheremenskaia, O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. *Chem. Rev.* **113**, 2940-2957 (2013).

28. Hwang, S. H. et al. Predicting carcinogenic mechanisms of non-genotoxic carcinogens via combined analysis of global DNA methylation and in vitro cell transformation. *Int. J. Mol. Sci.* **21**, 5387 (2020).
29. Du, H., Xu, T., Yu, S., Wu, S. & Zhang, J. Mitochondrial metabolism and cancer therapeutic innovation. *Signal Transduct. Target Ther.* **10**, 245 (2025).
30. Hammad, M. et al. Roles of oxidative stress and Nrf2 signaling in pathogenic and non-pathogenic cells: a possible general mechanism of resistance to therapy. *Antioxidants* **12**, 1371 (2023).
31. Vyas, S., Zaganjor, E. & Haigis, M. C. Mitochondria and cancer. *Cell* **166**, 555-566 (2016).
32. Reddam, A., McLarnan, S. & Kupsco, A. Environmental chemical exposures and mitochondrial dysfunction: a review of recent literature. *Curr. Environ. Health Rep.* **9**, 631-649 (2022).
33. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *J. Gen. Physiol.* **8**, 519-530 (1927).
34. Galluzzi, L. et al. Molecular mechanisms of cisplatin resistance. *Oncogene* **31**, 1869-1883 (2012).
35. Fowler, P. et al. Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. II. Importance of accurate toxicity measurement. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **747**, 104-117 (2012).
36. Kirkland, D. et al. Updated recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **795**, 7-30 (2016).
37. Li, R. et al. Tert-butylhydroquinone mitigates Carbon Tetrachloride induced Hepatic Injury in mice. *Int. J. Med. Sci.* **17**, 2095-2103 (2020).
38. Khezerlou, A., Akhlaghi, A. P., Alizadeh, A. M., Dehghan, P. & Maleki, P. Alarming impact of the excessive use of tert-butylhydroquinone in food products: a narrative review. *Toxicol. Rep.* **9**, 1066-1075 (2022).
39. Liu, Y. et al. L-Menthol alleviates cigarette smoke extract induced lung injury in rats by inhibiting oxidative stress and inflammation via nuclear factor kappa B, p38 MAPK and Nrf2 signalling pathways. *RSC Adv.* **8**, 9353-9363 (2018).
40. Jordan, M. A. & Wilson, L. Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer* **4**, 253-265 (2004).
41. Fenech, M. et al. “Micronuclei and Disease” special issue: aims, scope, and synthesis of outcomes. *Mutat. Res. Rev. Mutat. Res.* **788**, 108384 (2021).
42. Gisselsson, D. et al. When the genome plays dice: circumvention of the spindle assembly checkpoint and near-random chromosome segregation in multipolar cancer cell mitoses. *PLoS One* **3**, e1871 (2008).

43. OECD. *Test No. 487: in vitro mammalian cell micronucleus test*. OECD Guidelines for the Testing of Chemicals, Section 4 (OECD Publishing, 2023).
44. More, S. J. et al. Guidance on aneugenicity assessment. *EFSA J.* **19**, e06770 (2021).
45. Corvi, R. & Madia, F. In vitro genotoxicity testing—can the performance be enhanced? *Food Chem. Toxicol.* **106**, 600-608 (2017).
46. Kirsch-Volders, M. et al. Risks of aneuploidy induction from chemical exposure: twenty years of collaborative research in Europe from basic science to regulatory implications. *Mutat. Res. Rev. Mutat. Res.* **779**, 126-147 (2019).
47. Nohmi, T. Thresholds of genotoxic and non-genotoxic carcinogens. *Toxicol. Res.* **34**, 281-290 (2018).
48. Laub, A. R. et al. Unconventional T cells in anti-cancer immunity. *Front. Immunol.* **16**, 1618393 (2025).
49. Zha, D. et al. Proteomic profiling of fallopian tube-derived extracellular vesicles using a microfluidic tissue-on-chip system. *Bioengineering (Basel)* **10**, 423 (2023).
50. Das, D. N. & Ravi, N. Influences of polycyclic aromatic hydrocarbon on the epigenome toxicity and its applicability in human health risk assessment. *Environ. Res.* **213**, 113677 (2022).
51. Zheng, J. et al. Structure of human MDM2 complexed with RPL11 reveals the molecular basis of p53 activation. *Genes Dev.* **29**, 1524-1534 (2015).
52. Kirkland, D., Kasper, P., Müller, L., Corvi, R. & Speit, G. Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. *Mutat. Res.* **653**, 99-108 (2008).