

A Proteomic Atlas of Human Milk: Uncovering Milk Functional Components

Marco Agostino Deriu

`marco.deriu@polito.it`

Politecnico di Torino

Harry Zaverdas

InSyBio PC

Syed Taimoor Hussain Shah

PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, 10129 Turin, Italy

Filip Stojceski

Dalle Molle Institute for Artificial Intelligence USI-SUPSI <https://orcid.org/0000-0001-8778-8395>

Konstantinos Panagiotopoulos

PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, 10129 Turin, Italy

Konstantina Liontou

InSyBio PC, Patras, 265 04, Greece

Seferina Mavroudi

University of Patras <https://orcid.org/0000-0003-0612-1009>

María Carmen Durán-Ruiz

Biomedical Research and Innovation Institute of Cadiz (INiBICA), Cadiz, Spain

Foteini Kiskira

InSyBio PC, Patras, 265 04, Greece

Athanasios Kalogeras

Industrial Systems Institute <https://orcid.org/0000-0001-5914-7523>

Christos Alexakos

ATHENA RC <https://orcid.org/0000-0002-8932-6781>

Gianvito Grasso

Dalle Molle Institute for Artificial Intelligence USI-SUPSI, Polo universitario Lugano-Campus Est, Lugano, Switzerland. <https://orcid.org/0000-0002-7761-222X>

Konstantinos Theofilatos

InSyBio PC <https://orcid.org/0000-0001-6799-0553>

Article

Keywords: Human Milk, Proteomics, Protein-Protein Interactions, Protein Regulatory Networks
Cardiovascular Diseases

Posted Date: August 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10089470/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Human milk is a complex and highly organized biofluid whose molecular composition underpins neonatal nutrition, immune protection, and developmental regulation. Despite extensive proteomic profiling, its global organizational structure remains fragmented and poorly integrated. Here, we performed a large-scale meta-analysis of mass spectrometry–based proteomics datasets to construct a unified, queryable human milk proteome atlas for system-level interpretation. By harmonizing datasets through standardized preprocessing and normalization, we defined a core human milk proteome comprising the most recurrent and consistently detected proteins and peptides. An established machine-learning-based method was applied to predict binary protein–protein interactions from this core set, enabling reconstruction of a dense interaction network and moving beyond isolated protein catalogues toward a systems-level representation of milk organization. Network analysis revealed functional modules related to immune defence, vesicle-mediated secretion, and proteostatic control, while protein regulatory networks were also reconstructed. These results highlight the systems-level relevance of milk proteins and their potential role in maternal–infant regulation in health and disease. The integrative framework and web resource (<http://atlas-galatea.insybio.com/>) provide a foundation for comparative proteomics and network-based studies of the regulatory architecture of human milk.

Main

Human milk represents one of nature's most sophisticated and precisely calibrated nutritional systems, providing optimal nourishment for infant development during the critical early months of life. Beyond its role as a nutrient source, human milk functions as a structured biological system delivering coordinated immune, metabolic, and developmental signals to the neonate [1].

The composition of human milk demonstrates extraordinary complexity, containing a wide range of interacting molecular constituents [2], that work synergistically to support infant growth and development. Rather than forming a static mixture, human milk shows dynamic variability influenced by maternal physiology, environmental factors, and lactation stage, indicating a feeding system with context-dependent and adaptive organisation and reflecting millions of years of evolutionary refinement [3]. The intricate balance of macronutrients, micronutrients, bioactive compounds, and immunological factors within human milk, influences not only immediate nutritional outcomes but also establishes foundational patterns for immune system development [4], cognitive function [5], and metabolic reprogramming [6]. Together, these observations imply that human milk functions as an integrated molecular system where components are arranged in coordinated structures rather than operating independently. Such coordinated organisation suggests that milk transmits structured regulatory information at the maternal–infant interface.

Human milk oligosaccharides (HMOs), a key component of human milk, are the third most abundant solid component of breast milk (after lactose and fat), exemplifying the networked nature of milk's biology [7]. Over 200 distinct HMO structures are known, synthesized in the mammary gland via complex

enzymatic pathways. Recent systems biology approaches have elucidated the biosynthetic network of HMOs in unprecedented detail. For example, Kellman et al. [8] integrated glycomics and mammary gland transcriptomics to construct an HMO biosynthesis network and predict the specific glycosyltransferase enzymes involved. In a complementary *in silico* study, McDonald et al. [7] demonstrated that a minimal set of 11 enzymatic activities could generate the vast majority of identified HMO structures, thereby mapping a reaction network covering most of the HMO glycome. Beyond biosynthesis, HMOs engage in molecular interaction networks with microbes and the immune system. HMOs are indigestible by the infant but serve as selective substrates (“prebiotics”) for beneficial gut bacteria (notably *Bifidobacterium* and *Lactobacillus* species), thereby shaping the infant gut microbiota. Through these networked interactions, HMOs contribute to protecting infants from intestinal disorders (like necrotizing enterocolitis) and guide the development of the immune system [8]. Collectively, these studies show how a single class of milk components can be seen as part of a coordinated biosynthetic and interaction network, rather than as isolated bioactive molecules.

Proteins are the central functional layer within human milk contributing to infant nutrition, weight gain control, immune protection, microbiome development, and infant growth [1]. Proteomic analysis techniques have been developed to handle the complexity and dynamic range of milk proteins, implementing sample selection and fractionation techniques to detect low-abundance yet functionally critical proteins [9–11]. Research focus is centered on exploring temporal changes in the milk proteome across lactation stages [12], comparing milk from different species to pinpoint evolutionary differences and species-specific functions [13], investigating milk proteins’ glycosylation patterns [14], their role in microbial colonization and immune stimulation [15], and -importantly- understanding how milk proteomes respond to maternal conditions and disease [16].

However, in contrast to the increasingly resolved network architecture of HMOs, the protein layer of human milk remains comparatively fragmented across studies, lacking a unified systems-level reconstruction of its interaction and regulatory organization. Advances in liquid chromatography–mass spectrometry (LC–MS/MS) and quantitative proteomics have enabled increasingly comprehensive characterization of the human milk proteome, capturing both abundant structural proteins and low-abundance bioactive factors [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28]. Proteomics analyses on human milk have been employed to characterise whey proteins, glycoproteins, extracellular vesicle components, and species-specific peptides, as well as to explore temporal variation across lactation and responses to maternal physiological or pathological conditions [29] [30] [31] [32] [33] [34] [35]. Moreover, differences between human milk and bovine-based formula milk have been studied through proteomics, leading to the identification of hundreds of human-specific milk peptides [36]. Two notable studies have used an abundant set of milk samples derived from the CHILD Cohort Study [37], which were used for unravelling network associations of proteins, peptides, and between them, and for detecting relations between the allergy status of mother and child through the proteome differences in allergen abundances [38]. The above mentioned studies have substantially expanded the catalog of identified milk proteins and peptides and highlighted their functional diversity. However, most investigations have remained

dataset-specific or functionally focused, providing detailed compositional insights without reconstructing an integrated systems-level architecture of protein interactions [39] [19].

Recent large-scale proteomics and peptidomics studies have revealed that milk proteins form distinct co-expression and interaction networks reflecting their origins and functions. Dekker et al. [40] applied network inference approaches to uncover tightly connected modules of proteins and peptides. These modules corresponded to shared secretory pathways, proteolytic processing events, and coordinated immune-related functions. Such findings suggest that certain milk proteins are organized into functional clusters rather than existing as independent units, reflecting shared biological origins (e.g., systemic transfer versus mammary synthesis) or hierarchical processing relationships, such as precursor–fragment cascades.

Using weighted correlation network analysis (WGCNA) [41], studies have delineated modules of co-varying milk proteins and peptides and linked them to maternal and infant attributes [42]. These modules reflect coordinated regulation across lactation stages and physiological contexts. Such network analyses of the milk proteome and peptidome not only map the interactions within milk (e.g., a lactoferrin-lactoferricin peptide network indicating enzyme cleavage of lactoferrin) but also provide biological context; for instance, the presence of a tightly connected cluster of immunoglobulins, lactoferrin, and lysozyme underscores a concerted immune defence role. Collectively, these studies indicate that the human milk proteome exhibits modular organization and coordinated regulation [40]; however, these reconstructions remain largely dataset-specific and do not yet provide a unified, systems-level interaction and regulatory architecture.

Although human milk proteomics is expanding quickly and modular co-expression analyses are emerging, the field still faces challenges such as fragmented datasets, varied preprocessing methods, and a lack of a unified interaction-resolved framework. As a consequence, the global architecture of protein–protein interactions and regulatory hierarchies within human milk remains incompletely defined. To tackle this challenge, we adopted a large-scale integrative approach that standardizes publicly available MS-based proteomics datasets through preprocessing, normalization, and filtering. This allows for comparison across studies and helps define a reliable core human milk proteome. From this core set, we reconstructed a protein–protein interaction landscape using a machine-learning–based predictive framework, generating a dense interaction network that substantially expands beyond existing curated repositories. This reconstruction revealed a highly interconnected modular structure in which immune-related proteins, including immunoglobulin components, emerge as central regulatory nodes. We further reconstructed a human milk interaction-derived network architecture and performed comparative analyses with established cardiometabolic regulatory networks, identifying shared central hubs such as IGHA1, IGKC, and IGHG1, and implying potentially therapeutic components of human milk.

Finally, selected high-centrality interaction pairs were modelled by means of protein–protein docking and subjected to molecular dynamics (MD) simulations to assess interface stability and conformational

behaviour, thereby integrating atomistic structural dynamics into the reconstructed systems-level interaction framework.

Results

The overall computational workflow for the meta-analysis of human milk MS data is presented in Fig. 1, including data mining from ProteomeXchange, network reconstruction and analysis, network comparison with networks related to cardiovascular diseases and additional Molecular Dynamics analysis/validation of selected PPIs from the reconstructed network.

The results are organized to progressively describe the molecular organization of human milk, moving from the characterization of its integrated proteome and the identification of a core proteomic set, to the analysis of protein associations and functional modules, and finally placing these molecular relationships in a broader biological context through cross-network comparisons and structural exploration of selected interactions.

The human milk proteome and its core components

Large-scale reanalysis of publicly available MS datasets enabled the reconstruction of a unified human milk proteome and the identification of its most consistently represented components across studies.

After retrieving selected ProteomeXchange datasets described in Supplementary Table 1, and re-analyzing them with MaxQuant and the latest human protein library, we ended up with a filtered list of 2,911 proteins. By keeping those quantified in at least 20% of the samples across all datasets, a filtered list of 645 IDs was created, which constitutes the reported “core” human milk proteome. The GO Term enrichment analysis of the full list of 2,911 proteins composing the human milk proteome (Supplementary Fig. 1) showed statistically significant enrichment in translation processes and their involvement in the formation of cytosolic ribosomes (top 3 GO Terms: ‘cytosolic ribosome’/ QValue 1.39e^{-44} / FE 5.68, ‘cytoplasmic translation’/ QValue 6.32e^{-44} / FE 5.67, and ‘translation’/ QValue 8.00e^{-24} / FE 2.96). In contrast, the filtered core proteome was predominantly enriched in immunological processes, while the filtered list was mostly enriched in immunological processes such as antigen binding, immune response and in the formation of immunoglobulin complexes (top 3 GO Terms: ‘antigen binding’/ QValue 1.11e^{-22} / FE 8.17, ‘immunoglobulin complex’/ QValue 2.53e^{-19} / FE 6.40, and ‘immune response’/ QValue 5.54e^{-15} / FE 3.32), suggesting that the recurrent components of human milk are strongly associated with immune-related functions.

The human milk PPI network and characteristics of protein functional clusters

From the filtered list of 645 IDs, 207,690 potential binary combinations were created. Among these possible pairs, 12,550 PPIs were identified as positive by applying the previously established EOA

method [43], forming the basis of the human milk protein interaction network analysed in this study. From these positive interactions, FASN-YWHAZ, LTF-CEL, and IGHG1-FASN were the top 3 milk-specific interactions in terms of Probability Score and Predicted Affinity. To explore the internal organization of the interaction network, clustering analysis was performed, identifying groups of proteins with shared topological and functional properties. The results of the network reconstruction, along with the top 3 clusters' GO Term enrichment plots, are presented in Fig. 2. Proteins were grouped into functional and topological categories using an expanded version of the matrisomeDB to consider milk-specific proteins and cellular protein groups.

Functional enrichment analysis on the most central clusters (Clusters 3, 6, and 10, Fig. 2B) resulted in distinct functional enrichment patterns among the clustered proteins. These clusters highlight different functional contexts represented within the human milk proteome. Cluster 3 is highly enriched for proteasome-related processes (top 3 GO Terms: 'proteasome complex'/ P-value $1.83e^{-12}$ / FE 21.93, 'peptidase complex'/ P-value $1.83e^{-12}$ / FE 20.47, and 'endopeptidase complex'/ P-value $1.83e^{-12}$ / FE 20.47), including the proteasome core complex and endopeptidase activity, indicating potential involvement in protein degradation pathways. Cluster 6 shows significant enrichment in immune-related functions, particularly the immunoglobulin binding, suggesting a role in antigen processing and presentation (top 3 GO Terms: 'blood microparticle'/ P-value 0.00013/ FE 3.27, 'immunoglobulin complex'/ P-value 0.00034/ FE 5.42, and 'actin-based cell projection'/ P-value 0.014/ FE 6.93). In contrast, Cluster 10 is predominantly associated with vesicle-mediated transport and membrane organization, with enrichment in exocytic vesicles, plasma membrane transport, and synaptic vesicle components (top 3 GO Terms: 'synaptic vesicle'/ P-value 0.00012/ FE 30.7, 'exocytic vesicle'/ P-value 0.00018/ FE 24.56, and 'presynapse'/ P-value 0.00079/ FE 16.37). Supplementary Fig. 2 presents the functional enrichment analysis of all the clusters that had at least one significantly enriched (adjusted p-value < 0.05) GO Term cluster. Based on the protein groups enrichment results, these clusters do not comprise a specific protein group, but are rather generally associated as cellular proteins, which is more apparent in Clusters 3 and 10 (significant enrichment in the 'Other Cellular Proteins' group, with P-value 0.0006/ FE 2.07 and P-value 0.02/ FE 2.07, respectively). However, a subgroup of genes comprising Cluster 6 is found to be significantly (P-value 0.00015/ FE 4.23) involved in the Immunoglobulins protein group. The results of the Protein Groups enrichment results for the top-ranked clusters are presented in Supplementary Fig. 3. An overview of all the unclustered nodes can be found in Supplementary Fig. 4.

Functional associations within the human milk proteome

To explore how proteins within the human milk proteome are associated across samples, we re-analysed correlation patterns among the proteins identified in the dataset from Dekker et al. [38], using the regulatory network reconstruction pipeline described in the Methods section. In this work, Dekker et al. described coordinated protein associations in human milk, with a particular focus on extracellular matrix (ECM) components. In agreement with these observations, our analysis revealed consistent association patterns among several proteins detected in our dataset, including HSPG2 (Perlecan), COL14A1,

SERPINA1, and immunoglobulins (e.g., IGHA1, IGHG2, IGKC), which are highly abundant in human milk. The numerical values represent correlation coefficients between protein expression levels and their accompanying p-values, which indicate statistical significance. For instance, a strong correlation like HSPG2-COL14A1 = 0.7129 with a p-value $\approx 8.49\text{E-}11$, indicates a strong and statistically significant co-variation in their abundance profiles across samples. These associations highlight potential functional relationships between ECM components and immune-related proteins in human milk, suggesting their possible involvement in neonatal development, immune regulation, and gut epithelial support. The consistent involvement of immune-related proteins (e.g., IGKC, IGHG1, IGHA1) in strong correlations with ECM components suggests the presence of coordinated molecular patterns linking structural and immune-associated proteins within the milk proteome.

Shared hubs across distinct biological protein networks

Given that human milk is a secreted biofluid derived from maternal physiology, it is relevant to examine the extent to which its protein composition reflects broader molecular processes associated with the maternal organism. To address this, we compared the topology of the human milk network with previously published atherosclerotic [44] and ischemic heart failure-related [45] protein regulatory networks, with all 3 of them reconstructed using the exact same computational framework described in the Methods section, in order to identify recurrent central proteins and shared topological features .

The atherosclerosis-related network included proteins such as RAB6A, PCBP1, APOD, and COPE, together with proteins associated with mitochondrial function, oxidative stress, and extracellular matrix remodelling. The heart-failure network included extracellular matrix-related proteins and associated factors, spanning structural ECM components (e.g., COL6A1, LAMC1, LAMA5), immune-related proteins (e.g., IGHG1, IGHG3, CFH), and modulators of inflammation and fibrosis (e.g., CLU, POSTN, PRELP). As reported in that study, ECM proteins such as collagens (COL6A1/2/3), fibronectin (FN1), and decorin (DCN) were among the network components, together with complement- and immunoglobulin-related proteins including CFH, IGHG1, IGHA1, and IGKC, as well as proteins such as APOE, CLU, VCAN, ABI3BP, and ASPN.

To identify common hubs across systems, we performed a comparative network centrality analysis between the human milk and the heart disease PPI networks using the PageRank algorithm [46]. The PageRank score (PR) quantifies relative influence of a protein within the network topology, where higher values indicate nodes that sit on many well-connected paths [47]. Among the shared central proteins, IGLC2 (Immunoglobulin Lambda Constant 2) showed a PageRank score of 1.0 in the human milk network and was also highly ranked in the atherosclerosis network. Its centrality across both contexts is consistent with its established role in the constant region of immunoglobulin light chains and, more broadly, in adaptive immune function, including antibody production and mucosal immunity in milk [48], as well as inflammatory responses reported in atherosclerotic lesions [49–51]. Similarly, SERPINA1 (Alpha-1-antitrypsin) displayed a high PageRank in both networks, consistent with its known function as a regulatory anti-protease involved in tissue protection and modulation of proteolytic damage [52], and

suggesting relevance to both milk-associated immune-protection and protease-driven matrix remodelling in atherosclerosis. Another shared hub was IGKC (Immunoglobulin Kappa Constant), with a PageRank of 0.87 in the milk network and a strong presence in the atherosclerosis network, further supporting the recurrence of immunoglobulin-related proteins among central nodes across the analysed networks. Additional proteins such as APOA1, HP (Haptoglobin), and APCS also showed significant centrality in both networks, supported by low p-values and false discovery rates. APOA1, a key lipid transport and anti-inflammatory protein [53], has recognised roles in milk-associated metabolic support [54, 55] and in cholesterol metabolism and plaque-related processes in atherosclerosis [56]. HP is involved in oxidative stress management and has been linked to endothelial protection in atherosclerosis [57, 58], while its presence in the milk network may also be consistent with broader host-defence functions. Proteins such as COL18A1, FN1, and CLU were also highly ranked in both networks and are associated with extracellular matrix remodelling [59], lipid metabolism [60], and tissue homeostasis [61], indicating recurrence of these functional themes across the compared network topologies.

In the ischemic heart failure network reconstructed from proteomic data of failing heart tissue [90], CFH (Complement Factor H) was among the most influential nodes and also showed a dominant role in the human milk network, with a PageRank score of 1.0 in the heart failure network. This overlap was supported by highly significant p-values (≈ 0.0005) and false discovery rates ($\text{FDR} < 0.001$) in the comparison between the two networks. CFH is known to participate in immune regulation, particularly through complement pathway inhibition and inflammation resolution, both in cardiovascular settings [62] and in human milk, including reported functional interplay with lactoferrin [63]. Several additional proteins also showed high PageRank scores in both networks, indicating recurring centrality across distinct biological contexts. FGB (Fibrinogen beta chain) and PRELP (Prolargin), for example, scored 0.79 and 0.76 in the heart failure network, respectively, and were also highly ranked in human milk. Their recurrence is consistent with shared themes related to coagulation, extracellular matrix regulation, and tissue remodelling. FGB has been linked to vascular injury responses [64] and immune activation [65], whereas PRELP contributes to connective tissue integrity [66] and extracellular matrix structural organisation [67]. ABI3BP and OGN also ranked among the top proteins. ABI3BP has been associated with cellular senescence, extracellular remodelling, and cardiovascular disease [68], while OGN (Osteoglycin) contributes to collagen fibril formation [69] and matrix mineralisation [70]. Overall, the recurrence of these proteins among central nodes suggests that extracellular matrix- and immune-related components represent shared topological features across the analysed networks.

To further explore common structural features across networks, we performed a comparative clustering analysis across the three MCL-clustered biological networks (atherosclerosis, human milk, and heart failure) to identify conserved proteins consistently grouped within biologically significant clusters, defined here as clusters containing at least five proteins.

In addition, as shown in Fig. 4, HM proteome and cardiovascular disease signatures frequently occupy hub or bridge positions (Fig. 4A and 4B). More specifically, of particular interest are Clusters 1 and 2 in the HM-Atherosclerosis subfigure. Cluster 1 exhibits a subset of immunoglobulin genes (IGHG1, IGHG2,

IGHG3, IGKC) with high centrality scores (Betweenness Centrality 0.36–0.46) that are also common nodes between the 3 studied regulatory networks (Fig. 3). Cluster 2 presents a gene group, all of which are common signatures between HM and atherosclerosis. In this, PRELP and LUM genes are found to regulate FMOD. In the HM-Heart Failure subfigure, a similarly high centrality (Betweenness Centrality 0.10–0.43) of the same immunoglobulin genes is found on Cluster 1. In Cluster 2, the triplet of FMOD-PRELP-LUM occupies central positions (max. Betweenness Centrality 0.49), with the same directionality, while in Cluster 3 the same can be said about another triplet of laminin genes (LAMA5, LAMC1, LAMB2, Betweenness Centrality 0.20–0.53) that is indirectly co-regulated with a subset of collagen genes in the same cluster (COL6A1, COL6A2, COL6A3, COL14A1).

Expression summaries for the two overlap sets of signatures (HM-Atherosclerosis and HM-Heart Failure) demonstrate broad distribution across several tissues in the GTEx analysis (Fig. 4C and 4D). A subset of HM-Atherosclerosis common genes (IGHA1, IGKC, APOE, PRDX2, HTRA1, TIMP3, ANXA2, PRELP, LAMB2, AEBP1, LUM, FMOD, LAMC1, NID1) exhibits high TPM scores in many tissues, spanning the digestive, cardiovascular, endocrine, reproductive, and integumentary systems, while IGHA1, IGKC, APOE, PRDX2, and HTRA1 are also in brain tissues. The majority of these genes (except IGHA1 and HTRA1) that are expressed in all GTEx-available tissues are also common between the HM-Heart Failure and HM-Atherosclerosis networks.

Complementing the network and expression views, transcription-factor (TF) enrichment of the shared signatures pinpoints upstream regulators confirmed across multiple libraries through the ChEA3 platform (Fig. 4E and 4F). Specifically, the TFs found enriched in both common networks were FOXS1, HEYL, NR1H4, HNF4A, FOXC2, PRRX2, CREB3L3, TWIST2, PRRX1, and BNC2. Factors like FOXS1, FOXC2, TWIST2, and BNC2 suggest that the TFs of genes common to HM and cardiovascular diseases are likely linked to Epithelial-to-Mesenchymal Transition (EMT) or Tissue Fibrosis, while the inclusion of HNF4A, NR1H4, and CREB3L3 highlights a strong connection of said genes to liver metabolism (as also evident by the GTEx tissue expression analysis).

The meta-analysis of cardiovascular disease–associated proteins retained in the human milk regulatory network reveals a strong convergence into conserved regulatory modules. This convergence is consistently supported across all three analyses (regulatory network topology, GTEx-based tissue expression, and transcription factor enrichment) and highlights the predominance of immune-related and extracellular matrix–associated programs that are shared between human milk biology and cardiovascular disease pathology.

Molecular Dynamics

To elucidate the structural stability of the top-ranked PPIs identified by our ML-based pipeline, we performed all-atom MD simulations on three selected binary complexes, LDBH-YWHAZ, LDBH-NME2, and LDBH-MDH1. For each PPI pair, five distinct docking poses generated by HDOCKlite were subjected to 200 ns of simulation in three replicas. To quantify the interaction behaviour of these interactions, we

analysed the trajectories of the last 20 ns, calculating the average number of hydrogen bonds (H-bonds) and the buried surface area (BSA) for each pose (Fig. 5A and 5B). This analysis allowed us to identify the most conformations with the highest H-bonds and BSA for each complex.

For the LDHB-YWHAZ complex, model 5 demonstrated the most favorable interaction behavior. This model maintained an average of 16.88 ± 2.88 H-bonds and a BSA of 19.41 ± 2.25 nm². The high number of persistent H-bonds coupled with a substantial buried surface suggests a stable and potentially specific interaction between the glycolytic enzyme and the regulatory 14-3-3 protein. In the case of the LDHB-NME2 interaction, model 4 was identified as the one with the highest interaction metrics among the others. This complex exhibited an average of 9.83 ± 2.35 H-bonds and a BSA of 15.13 ± 0.77 nm². Finally, the LDHB-MDH1 complex showed the strongest binding interface in model 1. This model showed an average of 16.18 ± 2.43 H-bonds and the largest BSA among the three complexes at 20.55 ± 1.81 nm² (Fig. 5A). The extensive buried surface suggests that the interaction between LDHB and MDH1 involves a significant portion of their solvent-accessible surfaces.

Qualitative representations of the PPI pairs involved in the highest interacting models are provided in Fig. 5C, highlighting the spatial orientation of the interacting partners. The successful stabilization of these complexes *in silico* provides structural evidence supporting the biological plausibility of these PPIs within the human milk proteome. In addition, the root mean square deviation (rmsd) analyses of the top interacting PPIs are provided in Supplementary Fig. 5.

Web platform of Proteomic Atlas of Human Milk

To facilitate researchers in the exploration of the human milk proteome, a dedicated web interface was developed that integrates quantitative proteomics data, predicted PPI and regulatory networks, and curated external annotations into a single queryable framework. The interface is structured to provide both protein-centric and network-centric information, capturing all the levels of analysis of the present study. The main functionalities (Supplementary Fig. 6) include a home page that serves as a central access point, allowing users to search for individual proteins and their interaction networks. This design enables switching between molecular-level information and system-level information in the HM proteome. The protein search module provides structured access to all proteins identified in the meta-analysis, as well as to their standardized identifiers, sequence information, dataset coverage, and functional classification, as well as direct links to UniProt, HUGO, and RefSeq databases. The resources section supports data availability efforts, providing direct downloads of all major outputs generated in the study, including comprehensive protein lists, filtered protein sets used for network inference, and MaxQuant-derived datasets harmonized across studies, enabling researchers to independently validate and reanalyze the generated data. Network-level exploration is supported both in the predicted PPI network and in the regulatory network. The global PPI network view (Supplementary Fig. 6D) further displays the proteins grouped according to functional categories and clustered based on network topology. This allows users to identify densely connected modules and assess the distribution of biological functions across the network. In parallel, the interface supports targeted subnetwork

extraction (Supplementary Fig. 6E), enabling visualization of interaction neighborhoods for user-selected proteins, enabling hypothesis-driven analyses, such as examining the interaction context of specific groups of proteins of interest.

Discussion

The present study aimed to meta-analyze existing mass spectrometry-based proteomics datasets of human milk by applying consistent preprocessing, normalization, filtering, and imputation workflows, ultimately generating a queryable repository. By further applying an ML-based pipeline to reconstruct protein-protein interaction (PPI) and clustering algorithms the goal was to detect key hubs of HM-related proteins and their interactions. The network relationships of these milk-specific proteins provided a framework for guiding future experimental validation of their functional roles and their involvement in already established molecular pathways. In addition, using the largest publicly available human milk proteomics dataset (Dekker et al.), a protein regulatory network was reconstructed and compared with atherosclerotic and heart tissue-related networks to identify disease-specific proteins and clusters. Based on the combination of protein centrality, statistical significance (p-value and False Discovery Rate), and PageRank (PR) score, the most prominent shared proteins between the human milk and atherosclerosis-related networks were IGLC2, SERPINA1, and IGKC, while CFH, FGB, and PRELP were shared between the human milk and heart tissue networks, and IGHA1, IGKC, and IGHG1 were common to all three networks. These results provided important insights into the molecular pathways mediated by human milk and established a foundation for exploring its potential therapeutic properties.

The MaxQuant-based preprocessing of the ProteomeXChange datasets led to the identification of 2,911 unique human milk proteins with a filtering criterion ($\geq 20\%$ identification across datasets), which restricted this to 645 consistently detected proteins, representing the “core” human milk proteome. It is worth noting the complexity of human milk across populations, which, coupled with the different sample conditions and software preprocessing configurations, restricts a consensus on the consistency of the HM proteome. An earlier study[71] trying to merge 38 publicly available proteomic datasets of human milk identified a similar number of 2,698 unique proteins, though this presented a significant shift from proteins identified in milk-derived EVs (1,963). Subsequently, the exponential development of proteomics through LC-MS/MS analysis changed the landscape of the HM proteome. The most recent work from Dekker et al[72, 73] has analyzed a thorough dataset of 300 mother-child dyads from the CHILD cohort (<https://www.childstudy.ca>), identifying 1,690 proteins that, upon filtering for consistency in the mother-child allergy groups, reached 687, which was surprisingly similar to the filtered network proteins in the present study (645). Even though different filters and different datasets/groups were used as input in MaxQuant, more than half of the filtered proteins (377) were common, indicating a core proteome that can be assessed and meta-analyzed for its biological significance.

The enrichment of translation- and ribosome-related GO terms in the full set of 2,911 proteins likely indicates involvement in secretion-related processes, which can be attributed to milk’s cellular origin (epithelial and immune cells). Specifically, the lactating mammary gland consists of branching ductal

networks made up of epithelial cells that form extensive lobulo-alveolar clusters, serving as the site of milk production and secretion[74, 75], and subsequently, HM is expected to be enriched in proteins involved in this process. The filtered core proteome of 645 proteins shows a clear functional shift toward immune-related processes, in accordance with the known biological specificity of human milk in neonatal immune protection and passive immunity transfer[75, 76]. This enrichment in antigen binding, immunoglobulin complex formation, and immune response terms is well characterized in HM enrichment studies[77], but also underscores that consistency filtering helps narrow down functional proteins actively secreted into milk from background cellular components, helping identify HM's core functions.

From these 645 proteins, 12,550 positive HM PPIs were identified using the EOA methodology, and were used as input for the network reconstruction process and for the clustering algorithms. From the positive interactions, FASN-YWHAZ, LTF-CEL, and IGHG1-FASN being the top 3 milk-specific interactions in terms of Probability Score and Predicted Affinity. The network clustering was performed using the MCODE algorithm for clustering based on a Silhouette Coefficient (SC) criterion (see relevant Supplementary Table), leading to a 645-node highly interconnected system. The clusters were individually assessed based on their per-cluster SC, and GO term enrichment analysis was performed to reveal the cluster-specific functions in comparison to all other network proteins. The most biologically significant clusters (in terms of GO term enrichments of adj. p-val < 0.05) appeared to be Clusters 3,6, and 10.

Cluster 3, enriched in proteasome and peptidase proteins, potentially reflects the proteostatic control mechanisms essential to secretory mammary epithelial cells[75]. During lactation, these cells synthesize and secrete large quantities of milk proteins via the exocytotic pathway, requiring strict protein folding and degradation systems to maintain secretion efficiency[78, 79]. Proteasome and peptidase activity help remove misfolded or excess proteins before secretion, thereby sustaining milk proteostasis. Consistent with this, the mammary gland possesses a regulated proteolytic system, including enzymes such as plasmin and cathepsin D, which manage protein turnover and peptide generation within milk[80, 81]. Cluster 3, being highly homogenous in milk-specific proteins (see Fig. 2), is likely a subset of the quality-control and degradation machinery that ensures protein homeostasis and prevents the accumulation of misfolded proteins in the secretory pathway. It is worth noting that two of the proteins in the cluster (PA28 α and PA28 β subunits, products of PSME1 and PSME2 genes) form the PA28 (or 11S) complex that characterizes the eukaryotic immunoproteasomes[82]. So far, to our knowledge, no studies have described the functionality of immunoproteasome in human milk and its interactions with other proteasomal proteins, which can lead to potentially novel findings about an HM-specific interplay between proteostatic control and HM's innate immunity mechanisms.

Cluster 6 was found enriched in immunoglobulin binding and immune processes, which is expected, given the similar enrichment of the 645 proteins that were used as input to the network and the immunological abilities of HM that were described thoroughly. This cluster suggests a networked immune module possibly responsible for antigen presentation, antibody transport, or immune complex stabilization within milk. But further study is needed to unravel specific interactions of the antigen-

binding/ immunoglobulin complex-specific proteins that form this cluster (IGKV2-30, IGKV3-11, IGHA1,IGHM, IGHG1, PIGR, IGHV3-7, IGKV3-15, IGKV2D-28) with others within it that have a high Fold Enrichment score and take part in actin-based cell projection and filopodium formation (MSN, EZR, B4GALT1, DYNC1H1, ACTC1). Filopodia are believed to promote viral infection and enhance cell-to-cell viral spread by increasing transmission speed, enabling immune evasion, helping viruses overcome physiological barriers, and enhancing the infectivity of target cells. These indirect associations between filopodia proteins and immunoglobulins could be the groundwork for further study of the immunological block of filopodia-based viral spread, which could also take place in HM.

Cluster 10 is another highly homogenous cluster, as 4/5 proteins are milk-specific and involved in vesicle-mediated transport or form exocytic vesicles, representing the well-characterized vesicle-driven secretion mechanisms of HM, which include exosomes or microvesicles, which are known carriers of bioactive molecules from mother to infant[75]. The existence of Ubiquitin carboxyl-terminal hydrolase isozyme L3 (UCHL3), a deubiquitination enzyme, in this interacting cluster could be further studied. Ubiquitination (and deubiquitination) is well known to regulate many steps in the secretory pathway, including cargo sorting, vesicle formation, and endocytosis/exocytosis[83]. Some studies have suggested a positive role of deubiquitination enzymes (DUBs) in cell signaling, as DUBs use their influence in cell architecture to promote cell signaling[84]. Further exploring DUBs such as UCHL3 in relation to signaling mechanisms, such as the EV-mediated one that is crucial in HM.

The network reconstruction of the HM-specific PPIs showcased a set of functionally coherent and homogeneous clusters that cover major themes of HM function, such as immune protection and EV formation. Each of these clusters represents a distinct but interconnected facet of human milk biology, from proteostatic regulation (Cluster 3) and immune defense mechanisms (Cluster 6) to EV-mediated secretion and trafficking (Cluster 10). Clusters 3, 6, and 10 underscore the robustness of the proposed methodology to identify biologically significant modules in the reconstructed network, and this consistency paves the way for further experimental study and meta-analysis of the clustered interactions, aimed at elucidating the molecular architecture of human milk and its role in maternal–infant immunological and metabolic communication.

The human milk protein regulatory network reveals a finely tuned system of molecular interactions that underpin the biological richness and functional complexity of this vital secretion. Given the scale and density of the human milk protein regulatory network (1459 nodes and 132962 edges), we prioritized Markov Clustering (MCL) over other algorithms as its flow-based expansion inflation procedure is well-suited to large, noisy biological interaction graphs, where communities correspond to regions of high random-walk retention. Unlike many heuristics, MCL has a single, interpretable knob, the inflation parameter, that directly controls cluster granularity and enables transparent, reproducible tuning. MCL has a long track record of strong performance on network clustering problems [85–87], and high-performance implementations (e.g., HipMCL[88]) demonstrate scalability to large numbers of nodes and edges, properties that match our network’s size and our need for well-separated modules. We benchmarked candidate inflations (≈ 1.5 -3.0) and selected 2.7 as the settings that best balanced

resolution and stability: it maximized graph modularity, yielded consistent partitions across bootstraps (high adjusted Rand index), and was robust to modest edge-weight perturbations. Practically, MCL is fast, memory-efficient, and produces non-overlapping, compact modules that are easy to interpret, making it an appropriate choice for downstream functional analyses of our ARACNe-AP/SIREN-derived network.

Using the ARACNe-AP pipeline followed by SIREN-based directionality assignment and MCL clustering (inflation rate 2.7), the network decomposes into tightly knit modules that reflect specific biological tasks. A key observation is the centrality of immunoglobulin-related proteins such as IGHA1, IGKC, and IGHG2, which consistently appear in dominant clusters and exhibit strong mutual interactions. These proteins form the backbone of passive immunity conferred to the neonate, highlighting the network's primary focus on immune protection.

IGHA1 [89] was consistently identified in the biggest cluster across atherosclerosis, human milk, and heart failure networks. Its uniform clustering highlights its fundamental immunological role, likely involving mucosal immunity and antibody-mediated response, consistent with its known distribution in secretions and involvement in systemic inflammation.

IGHG1[90] and IGKC[91], were also highly central across human milk and cardiovascular tissues. IGHA1 is known to be correlated with multiple cardiovascular conditions, such as diabetic angiopathy, congestive heart failure, and hypertension, influencing vascular elasticity and heart contractility[92, 93]. Studies have also shown that the aortic valves of clonal hematopoiesis patients (shown to be linked with cardiovascular disorders[94]) contain higher transcription levels of IGHG1 than those without[95], and more interestingly, was shown to be upregulated in advanced-stage carotid atherosclerosis[96]. scRNA-seq and mRNA-seq differential expression analyses have shown upregulation of IGKC in endothelial cell populations in atherosclerotic plaques compared to normal tissue[97], while in microarray expression data from atheroma plaques showed it to be significantly downregulated[98], potentially indicating that the regulation of IGKC is correlated with atherosclerosis pathology. This triangulated cluster overlap reveals how immunoglobulin components retain strong modular integration regardless of the surrounding biological system, whether circulating in the vasculature during atherosclerotic stress, contributing to immune defense in milk, or participating in inflammatory cascades in heart failure. Their consistent inclusion in robust clusters strengthens their candidacy as core biomarkers or cross-condition immune mediators, highlighting the value of multi-network systems biology in uncovering conserved molecular scaffolds that underpin physiological resilience or disease propagation.

Equally prominent are clusters dominated by extracellular matrix (ECM) components, including COL14A1, FN1, and HSPG2, which play crucial roles in structural support and cellular adhesion. The co-localization of immune and ECM proteins within clusters suggests a coordinated mechanism of mucosal defence and tissue development, emphasizing the dual immuno-structural function of human milk. Clusters also encompass proteins involved in lipid transport, antimicrobial defence, and growth signalling, each representing a microdomain of co-regulated activity. The fine-grained clustering enabled

by the high inflation parameter ensures that these functional units are distinctly resolved, providing an interpretable framework for understanding how diverse protein functions are compartmentalized within the milk proteome. These insights reinforce the idea that human milk is a highly organized biological fluid, where immune surveillance, nutrient transport, and developmental cues are orchestrated through a modular protein network.

Interestingly, the human milk network also contains several proteins that are classically associated with cardiovascular disease (CVD), suggesting a broader systemic relevance of milk-borne molecules. Comparative analyses with atherosclerosis and heart failure regulatory networks, reconstructed and analyzed with the same workflow, revealed overlapping high-centrality proteins, such as CFH, FN1, CLU, and APOA1, which retain influential positions across all three conditions. Their presence within major clusters in the milk network implies that these proteins are not only essential for infant health but may also be key regulators in systemic immune and inflammatory processes. For example, CFH (Complement Factor H), identified as the most central node in the heart failure network, also emerges prominently in the milk interactome, where it likely contributes to immune modulation and pathogen clearance at mucosal surfaces.

Similarly, FN1 (Fibronectin 1) and CLU (Clusterin) participate in clusters associated with tissue remodelling and lipid metabolism functions equally relevant to neonatal tissue development and vascular repair mechanisms in CVD. These shared proteins act as molecular bridges between physiological (milk-based) and pathological (CVD-related) environments, highlighting the functional plasticity of immune-ECM mediators. Their clustering within human milk reflects a sophisticated system where evolutionarily conserved proteins serve dual roles in both protection and structural regulation. The identification of such CVD-related hubs in milk reinforces the hypothesis that maternal milk not only nourishes but may also impart systemic resilience through the transfer of multifunctional proteins. This finding opens new avenues for exploring milk-derived components as potential biomarkers or therapeutic leads for cardiovascular conditions, as well as for understanding maternal-neonatal molecular crosstalk in health and disease.

The regulatory network meta-analysis (Fig. 4) further argues that proteins shared between the HM and cardiovascular disease tissue networks are organized into coherent modules. The localization of shared proteins within high-centrality regions of the network is an indication of conserved regulatory programs that are apparent in both maternal milk secretion and cardiovascular disease states. This theory is based on the assumption that hub and bridge nodes often exert significant influence over system behaviour[99]. At first, immunoglobulin-centered clusters are present in all regulatory networks, especially IGHA1, IGKC, and IGHG1, further linking HM immunity with chronic inflammatory processes characteristic of atherosclerosis and heart failure. In parallel, extracellular matrix-related modules, particularly the PRELP-LUM-FMOD axis and laminin-collagen subclusters, emerge as conserved features across the shared networks. These ECM proteins are regulators of collagen fibrillogenesis, tissue stiffness, and fibrotic remodelling, while being linked in cardiovascular pathology[100–104], and their apparent regulation in human milk suggests that similar molecular systems may guide intestinal

maturation and tissue organization in the neonate. Importantly, this network-level convergence is independently validated by GTEx expression analyses, which show that shared regulatory genes are consistently expressed across cardiovascular, digestive, endocrine, and neural tissues, indicating systemic physiological relevance. TF enrichment further also shows that the shared genes are regulated by TF associated with epithelial-to-mesenchymal transition and fibrosis (FOXC2, TWIST2, PRRX1/2) [105–107], as well as metabolic and lipid regulatory pathways (HNF4A, NR1H4, CREB3L3)[108, 109], all of which are drivers of inflammatory remodelling and metabolic dysregulation in cardiovascular disease[110–112]. Summarizing, the regulatory network analyses support a model in which human milk may preserve immune and extracellular matrix regulatory systems that are also activated during cardiovascular pathology, providing a framework for understanding how milk-derived proteins may function as regulators that support neonatal development while also participating in inflammatory and remodelling processes later in life.

To complement the network-level findings, we performed MD simulations on top-scoring milk-specific PPIs to validate their structural stability and uncover mechanistic details that cannot be inferred from network analysis alone. Among the identified pairs, LDHB-centered interactions emerged as dominant among top scoring milk-specific PPIs. In particular, the LDHB–YWHAZ complex displayed strong stability, supported by extensive hydrogen bonding and a large buried surface area, suggesting a regulatory interaction between lactate metabolism and 14-3-3–mediated signalling[113]. Similarly, the LDHB–NME2 interaction is supported by existing experimental evidence, reinforcing the robustness of our computational predictions[114]. These findings indicate that key metabolic enzymes in human milk may participate in structured regulatory complexes, linking energy metabolism with broader signalling pathways. Additional analyses on these PPIs are provided in the Supplementary Discussion 1.

Several limitations of the proposed methodologies should be acknowledged. First, the meta-analysis of the ProteomeXChange datasets relied on integrating samples acquired under heterogeneous experimental conditions, and these methodological discrepancies likely introduced batch effects and platform-specific biases that could not be fully mitigated despite standardized preprocessing and normalization. Moreover, only proteins identified and quantified across $\geq 20\%$ of the samples were included in downstream analyses, which may have excluded low abundance but biologically relevant proteins. Second, ML–based reconstruction of protein–protein interaction and regulatory networks provided valuable insights, but these computational networks are inherently limited by the incompleteness and potential false-positivity of available interaction repositories and the predictive nature of ML models; network edges do not reflect confirmed biochemical interactions, and experimental validation remains essential. Furthermore, the regulatory network reconstruction using ARACNe-AP and SIREN was based on proteomic abundance profiles, which capture correlation but not causality. Third, the current analysis did not include temporal or longitudinal dimensions of milk composition, which are known to vary across lactation stages and maternal physiological states. Future work involves expanding this framework through multi-omics integration, as recent systems biology initiatives, such as the Phoenix Study [115], and multi-omic milk network analyses by Chandran et al. [116] and others [117,

118], have demonstrated the value of integrating milk's proteome with its metabolome, microbiome, and immunome to identify novel biomarkers predictive of infant immune development and disease risk.

Conclusively, this study represents the first large-scale meta-analysis of human milk proteomics datasets integrated through a standardized preprocessing and normalization framework and complemented by ML-based network reconstruction. By harmonizing raw human milk samples from 8 ProteomeXChange studies, we defined a robust “core” human milk proteome of 645 consistently quantified proteins. The applied EOA pipeline was able to predict 12,550 binary PPIs from these 645 proteins, with LDBH-YWHAZ, LDBH-NME2, and LDBH-MDH1 being the top-scoring milk-specific interactions in terms of probability score and predicted affinity. Their structural stability was characterized thoroughly with MD analysis, showcasing significant stabilization of the complexes and further confirming their biological significance. The critical role of LDHB in the HM function and its dysregulation in disease states serve as an indication of the importance of these interactions, whose further exploration may provide insights into novel therapeutic markers contained in the HM proteome. The reconstructed PPI network from these 12,550 PPIs further verified the key functional properties of human milk by showcasing distinct clusters enriched in immune defence, vesicle-mediated secretion, and proteostatic control terms. Furthermore, the protein regulatory networks of human milk, atherosclerosis, and heart failure highlighted proteins such as IGHA1, IGKC, and IGHG1 as central mediators that bridge milk-specific and cardiovascular-related molecular systems. These findings emphasize that human milk proteins not only play a role in neonatal immunity but also exhibit systemic biological relevance, potentially linking maternal and infant cardiovascular and immune health.

Finally, by incorporating these findings in a unified publicly available web interface, the aim was to enhance reproducibility and data reuse by providing direct access to all resources used in constructing the atlas. In this atlas, users can query for proteins and peptides identified in the proteomic datasets and browse interactively in the PPI and regulatory networks, ultimately offering a structured entry point for researchers in proteomics, nutrition, and nutraceutical science to investigate the molecular composition of human milk and its potential systemic implications.

Methods

Proteomics Data Description

Proteomics datasets of Human milk were mined from ProteomeXchange, ending up with 8 datasets as presented in Supplementary Table 1, which included a total of 1,175 raw human milk samples in the CHILD repository. The largest dataset retrieved was PXD034806, which includes a Liquid Chromatography Mass Spectrometry (LC-MSMS) dataset of 367 human milk samples from the ChildDB. This dataset was initially used to study the proteome relation to allergies of the mother and the allergy development of the infant. Raw data were retrieved for each dataset, and after filtering out additional samples from fractionation protocols (such as in the case of the PXD021974 dataset) and common

samples between the PXD036477 and PXD034806 datasets, a total of 753 out of the initial 1,175 raw samples were meta-analysed with the workflow described below.

Proteomics Data Preprocessing

The MaxQuant/Andromeda [119] v.2.6.6.0 software was selected for preprocessing, with the following parameters: protein false discovery rate (FDR) cutoff of 0.01, minimum peptide length for unspecific search 8 and maximum of 25, variable modifications of acetylation of the peptide N-term, deamidation of the side chains of asparagine and glutamine, and oxidation of methionine, Trypsin/P digestion with a maximum of 2 missed cleavages, and match between runs option activated. The raw data were searched against human proteins retrieved from UniProt (20,421 reviewed UniprotKB/Swiss-Prot proteins as of December 2024). The steps of the preprocessing the quantitative data included the control for strict FDR ratios (5% at peptide and protein levels), filtering proteins with more than 30% missing values across all samples, normalizing against the total ion intensity, and imputing missing values using k-NN impute. Regarding the proteins selected for protein-protein interaction analysis, we applied a more stringent filtering layer before imputation, keeping proteins identified in at least 20% of the samples across all proteomics datasets of human milk.

Protein Regulatory Network Reconstruction and Analysis

The reconstruction of the protein regulatory networks was conducted with ARACNe-AP algorithm [120, 121] for network inference, complemented by statistical correction for multiple testing, integration of biological directionality through the SIREN method [122]. The pipeline begins with the application of the ARACNe-AP algorithm, which infers regulatory interactions among genes or proteins based on mutual information (MI) [123] computed from protein abundance profiles. To enhance robustness, a bootstrapping approach was employed by performing multiple ARACNe runs with different random seeds. The resulting networks were then aggregated, and a three-step filtering strategy was applied to retain only strong and biologically meaningful interactions. First, a threshold was applied to the aggregated network obtained from the bootstrap process. Second, a dynamic threshold was assigned individually to each protein based on the average strength of its interactions. Third, a global upper-bound threshold defined by ARACNe was applied to further restrict retained interactions. To address the problem of multiple hypothesis testing inherent in large-scale network inference, the Benjamini-Hochberg (BH) procedure [124] was applied and only interactions with corrected p-values below 0.05 were retained for further analysis. Once the corrected undirected network was obtained, biological directionality was inferred using the SIREN method. Directionality scores were calculated and appended to the network, resulting in a directed interaction map in which edges have both magnitude, denoted as MI in the algorithm, and orientation. Finally, to avoid redundancy, duplicate interactions, including bidirectional edges with the same inferred directionality, were evaluated, and among mirrored edges only the one with the stronger directionality score was retained. Self-loops, defined as edges connecting a node to itself, were also removed.

Analyses on the constructed networks were performed using the InSyBio BioNets tool [125], part of the InSyBio Suite (<https://suite.insybio.com/>), for network metric calculation, clustering using the Markov Clustering (MCL), Restricted Neighborhood Search Clustering (RNSC), and ClusterOne algorithms, as well as network comparison. Custom R and Python scripts were used for downstream analysis and visualization of pathway and functional enrichment results. Pathway enrichment analysis was performed using the DAVID web tool [126], the GTEx portal [127] (<https://www.gtexportal.org/>) was used to investigate the cell-type specificity of the expression profiles of proteins of interest, and the ChEA3 portal [128] (<https://maayanlab.cloud/chea3/>) was used for transcription factor enrichment analysis.

Protein-Protein Interaction Network Reconstruction

The Protein-Protein Interaction (PPI) network reconstruction, based on the human milk proteins identified in the mined human milk MS data, was performed using the predicted interactions from the Evolutionary Optimization Algorithm-based (EOA) methodology suggested by Zaverdas, Stojceski, et al [43], in the context of the taste receptor interactome network reconstruction. This methodology is based on a hybrid combination of an evolutionary heuristic optimization method with machine learning classification and regression methods to develop classification models for the prediction of physical binary positive PPIs based on 61 functional, orthology, sequence, co-expression, and structural features. This method also provides a regression model to estimate PPI binding affinity, and, along with the classifier's prediction probability, this is used to rank the predicted PPIs.

The network reconstruction was carried out using a list of proteins identified in all the proteomics datasets. This list was reduced by filtering out proteins that were quantified in less than 20% of the samples in all the datasets where they have been identified. Upon filtering, the final reduced list is used for creating binary combinations (PPIs) between them. All the possible combinations (potential binary PPIs) are retrieved from this list, and then the EOA methodology is used for identifying the positive PPIs from the multitude of potential ones, which in turn were used in the network reconstruction phase.

Network visualization was performed with the Cytoscape [129] v.3.10.1 software and network clustering carried out using the Markov Clustering (MCL) [130], Molecular Complex Detection (MCODE) [131], and Leiden [132] algorithms. The parameters for the algorithms were: Granularity 4, 2.5, and 2 for the MCL; for MCODE, haircut and fluff were applied, node score cutoff was set to 0.3, k-core to 2, max depth to 200, and degree coefficient to 3 and 2; and for Leiden, resolution was set to 0.1 and 0.5. Clustering assessment metrics were applied to evaluate the results and choose the optimal clustering algorithm and its parameters in order to evaluate the biological relevance of the findings. Global and intra-cluster metrics were computed by analyzing the clustered subgraph of the network, calculating topology-based measures (e.g., density, degree, clustering coefficient) using NetworkX v.3.4.2 and Cytoscape-exported attributes, and assessing overall cluster separation using Node2Vec embeddings followed by silhouette scoring based on Cytoscape-derived cluster labels. For assessing each cluster in the network, the per-cluster Density, Inertia, and Silhouette Coefficient were calculated by constructing subgraphs for each cluster from the original network. Then, the per cluster sub-graph density was calculated using

NetworkX, and node-level topological features (e.g., betweenness, closeness, clustering coefficient) were standardized via z-score normalization.

From the clustering metrics, a final network is chosen based on the Silhouette Coefficient and Density Metrics (for the rationale, see Supplementary Notes 1). For the selected network, a per-cluster Gene Ontology (GO) term and Protein Groups enrichment analysis was performed against all the network's proteins, in order to find in which biological processes, molecular functions, cellular components, and protein groups each cluster is significantly enriched in regard to the other clusters and genes in the network. GO Term enrichment was also performed in both the full and filtered lists of proteins, which initialized the network creation (Supplementary Fig. 1). For better focus on the biological relevance of the findings, all the graphs created contain clusters and genes with an adjusted p-value of enrichment being < 0.05 .

Molecular Dynamics

To elucidate and characterize the HM-related PPIs identified using the ML-based EOA pipeline, we selected the top interactions from the multitude of positively predicted ones for further Molecular Dynamics (MD) analysis. The selection of the PPI pairs for the MD simulation was based on the following criteria: 1) the involvement of at least one milk-specific protein; 2) Both proteins to have an intrinsic disorder score (calculated using the DisPredict v3.0 [133] software) below 0.5; 3) To be among the top 20 interactions in terms of the average score of the classifier's probability and the computationally predicted affinity; 4) Both proteins should have PDB structures available and not derived from homology modelling; 5) To limit the computational power required for the simulation, we decided to consider PPIs composed by proteins with fewer than 350 amino acids. Given that L-lactate dehydrogenase B chain (LDBH, UniProt ID: P07195) appeared in 13 of the top 20 PPIs, it was identified as the most recurrent interacting protein. Consequently, we selected LDBH as the fixed anchor and prioritized the three top-ranked PPIs involving this protein. The selected interactors with LDBH were 14-3-3 protein zeta/delta (YWHAZ - Uniprot ID: P63104), nucleoside diphosphate kinase B (NME2 - Uniprot ID: P22392), and cytoplasmic malate dehydrogenase (MDH1 - Uniprot ID: P40925). The atomistic structure of LDBH, YWHAZ, NME2, and MDH1 proteins was collected from the PDB IDs 7DBJ [134], 2002 [135], 3BBB [136], and 7RM9 [137], respectively.

These three PPI pairs were docked using HDOCKlite v1.1[138, 139] and the first 5 docked poses were selected as starting points for the simulations. Each docked model was placed at the center of a dodecahedral simulation box with a minimum distance of 1.5 nm between periodic images and solvated using TIP3P [140] water molecules. The final solvated systems, following the addition of Na⁺ and Cl⁻ ions to a physiological concentration of 0.15 M, contained between 90,000 and 120,000 particles in total. The Amber19SB [141] force field was used to define protein topology through an all-atom approach, and each system was minimized using the steepest descent method for 5000 steps. We then performed the equilibration procedure through one MD simulation of 500 ps under the NVT ensemble and a subsequent MD simulation of 500 ps under the NPT ensemble. For the equilibration protocol, the v-rescaling [142] temperature coupling algorithm with a time constant of 1.0 ps was applied to keep the temperature at

310 K. C-rescale [143] semi-isotropic pressure coupling algorithm with a reference pressure of 1 bar and a time constant of 5.0 ps was employed. Then, all systems were simulated for the production run in the NPT ensemble with 4 fs time steps [144], using v-rescaling thermostat and c-rescale barostat, with a time constant of 1.0 ps and 5.0 ps, respectively. Each starting pose was simulated for 200 ns in 3 replicas, for a total of 9 μ s of sampled time. Electrostatic interactions were calculated by applying the particle-mesh Ewald [145] (PME) method and van der Waals interactions [146] were defined within a cut-off of 1.0 nm. A time step of 4 fs was achieved through the application of a hydrogen mass repartitioning scheme[144], in which the minimum atomic mass was scaled by a factor of 3, in combination with LINCS constraints[147]. Trajectories were collected every 20 ps and the Visual Molecular Dynamics [148] (VMD) package was employed to visually inspect the simulated systems. GROMACS 2025 [149, 150] package was used for simulations and data analysis. The last 20 ns of the 200 ns production run of each simulation were considered for the analyses.

Web platform of Proteomic Atlas of Human Milk

The application utilizes PHP v7.4.33 as the backend programming language, built on the Yii v1.1.23 framework, a straightforward framework compatible with legacy systems. Networks were reconstructed using the Cytoscape software as discussed in section 2.4 and integrated using its dedicated JSON format (cyjs) designed for web-based visualization. The front-end was developed using JavaScript for interactivity, with HTML5 and CSS managing the structure and styling. PostgreSQL v16.9 serves as the database management system.

Declarations

Acknowledgements

This work was developed with the funding support from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement Action (No. 101183057), through the GALATEA (diGiTAL twin frAmework for TailorEd infAnt nutrition) project, and more details about it are available at: <https://galateahe.eu/>.

Acknowledgment of the GTEx Portal use:

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 09/01/2025.

Conflict of Interest

The authors declare no competing interests.

Code Availability

All the proteomics and network data generated in this study are publicly available in the web application created for this purpose as described in the Data Availability section (Human Milk Proteomic Atlas web interface).

Data Availability

All data generated and analyzed in this study, as well as the networks in an interactive form, are available through the Human Milk Proteomic Atlas web interface and can be accessed at: <http://atlas-galatea.insybio.com/>.

References

1. Perrella S, Gridneva Z, Lai CT, et al. Human milk composition promotes optimal infant growth, development and health. *Seminars in Perinatology* 2021; 45: 151380.
2. Talbert JA, Townsend SD. Human milk as a complex natural product. *Nat Prod Rep* 2025; 42: 406–420.
3. Ma J, Palmer DJ, Geddes D, et al. Human Milk Microbiome and Microbiome-Related Products: Potential Modulators of Infant Growth. *Nutrients* 2022; 14: 5148.
4. Christian P, Smith ER, Lee SE, et al. The need to study human milk as a biological system. *The American Journal of Clinical Nutrition* 2021; 113: 1063–1072.
5. Chiurazzi M, Cozzolino M, Reinelt T, et al. Human Milk and Brain Development in Infants. *Reprod Med* 2021; 2: 107–117.
6. Suwaydi MA, Gridneva Z, Perrella SL, et al. Human Milk Metabolic Hormones: Analytical Methods and Current Understanding. *IJMS* 2021; 22: 8708.
7. McDonald AG, Mariethoz J, Davey GP, et al. In silico analysis of the human milk oligosaccharide glycome reveals key enzymes of their biosynthesis. *Sci Rep* 2022; 12: 10846.
8. Kellman BP, Richelle A, Yang J-Y, et al. Elucidating Human Milk Oligosaccharide biosynthetic genes through network-based multi-omics integration. *Nat Commun* 2022; 13: 2455.
9. Boschetti E, Righetti PG. Low-Abundance Protein Enrichment for Medical Applications: The Involvement of Combinatorial Peptide Library Technique. *IJMS* 2023; 24: 10329.
10. Coscia A, Orrù S, Di Nicola P, et al. Detection of cow's milk proteins and minor components in human milk using proteomics techniques. *The Journal of Maternal-Fetal & Neonatal Medicine* 2012; 25: 49–51.
11. Vahkal B, Kraft J, Ferretti E, et al. Review of Methodological Approaches to Human Milk Small Extracellular Vesicle Proteomics. *Biomolecules* 2021; 11: 833.
12. Zhang L, De Waard M, Verheijen H, et al. Changes over lactation in breast milk serum proteins involved in the maturation of immune and digestive system of the infant. *Data in Brief* 2016; 7: 362–365.

13. Ning J, Yang M, Liu W, et al. Proteomics and Peptidomics As a Tool to Compare the Proteins and Endogenous Peptides in Human, Cow, and Donkey Milk. *J Agric Food Chem* 2023; 71: 16435–16451.
14. Guan B, Liu D, Cao X, et al. New insights into site-specific N -glycosylation profiling differences of serum proteomes between colostrum and mature breast milk. *Food Frontiers* 2023; 4: 2024–2044.
15. Kalbermatter C, Fernandez Trigo N, Christensen S, et al. Maternal Microbiota, Early Life Colonization and Breast Milk Drive Immune Development in the Newborn. *Front Immunol* 2021; 12: 683022.
16. Roncada P, Stipetic LH, Bonizzi L, et al. Proteomics as a tool to explore human milk in health and disease. *Journal of Proteomics* 2013; 88: 47–57.
17. Damm M, Nussold C, Cantillo D, et al. Can electromagnetic fields influence the structure and enzymatic digest of proteins? A critical evaluation of microwave-assisted proteomics protocols. *Journal of Proteomics* 2012; 75: 5533–5543.
18. Zhang D, Xie Y, Mrozek MF, et al. Raman Detection of Proteomic Analytes. *Anal Chem* 2003; 75: 5703–5709.
19. Haas J, Mizaikoff B. Advances in Mid-Infrared Spectroscopy for Chemical Analysis. *Annual Rev Anal Chem* 2016; 9: 45–68.
20. Blanco M, Villarroya I. NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry* 2002; 21: 240–250.
21. Aernouts B, Van Beers R, Watté R, et al. Visible and near-infrared bulk optical properties of raw milk. *Journal of Dairy Science* 2015; 98: 6727–6738.
22. Bao Y, Newburg DS. Capillary electrophoresis of acidic oligosaccharides from human milk. *Electrophoresis* 2008; 29: 2508–2515.
23. Imperiale S, Morozova K, Ferrentino G, et al. Analysis of milk with liquid chromatography–mass spectrometry: a review. *Eur Food Res Technol* 2023; 249: 861–902.
24. Label-Based and Label-Free Strategies for Protein Quantitation. In: *Methods in Molecular Biology*. New York, NY: Springer New York, pp. 31–43.
25. Isotope-Coded Affinity Tags for Protein Quantification. In: *New and Emerging Proteomic Techniques*. New Jersey: Humana Press, pp. 151–158.
26. Kang U-B, Yeom J, Kim H, et al. Quantitative Analysis of mTRAQ-Labeled Proteome Using Full MS Scans. *J Proteome Res* 2010; 9: 3750–3758.
27. Koehler CJ, Strozynski M, Kozielski F, et al. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *J Proteome Res* 2009; 8: 4333–4341.
28. Liu J, Zhou Y, Shan Y, et al. A Multiplex Fragment-Ion-Based Method for Accurate Proteome Quantification. *Anal Chem* 2019; 91: 3921–3928.
29. Liao Y, Alvarado R, Phinney B, et al. Proteomic Characterization of Human Milk Whey Proteins during a Twelve-Month Lactation Period. *J Proteome Res* 2011; 10: 1746–1754.
30. Grapov D, Lemay DG, Weber D, et al. The Human Colostrum Whey Proteome Is Altered in Gestational Diabetes Mellitus. *J Proteome Res* 2015; 14: 512–520.

31. Van Herwijnen MJC, Zonneveld MI, Goerdayal S, et al. Comprehensive Proteomic Analysis of Human Milk-derived Extracellular Vesicles Unveils a Novel Functional Proteome Distinct from Other Milk Components. *Molecular & Cellular Proteomics* 2016; 15: 3412–3423.
32. Zhu J, Lin Y-H, Dingess KA, et al. Quantitative Longitudinal Inventory of the N-Glycoproteome of Human Milk from a Single Donor Reveals the Highly Variable Repertoire and Dynamic Site-Specific Changes. *J Proteome Res* 2020; 19: 1941–1952.
33. Wang Z, Zhang N, Wang W, et al. Glycoproteomics Analysis Reveals Differential Expression of Site-Specific Glycosylation in Human Milk Whey during Lactation. *J Agric Food Chem* 2021; 69: 6690–6700.
34. Zhang L, De Waard M, Verheijen H, et al. Changes over lactation in breast milk serum proteins involved in the maturation of immune and digestive system of the infant. *Data in Brief* 2016; 7: 362–365.
35. Zhu J, Dingess KA, Mank M, et al. Personalized Profiling Reveals Donor- and Lactation-Specific Trends in the Human Milk Proteome and Peptidome. *The Journal of Nutrition* 2021; 151: 826–839.
36. Lopez C, Adelfio A, Wall AM, et al. Human milk and infant formulae: Peptide differences and the opportunity to address the functional gap. *Current Research in Food Science* 2020; 3: 217–226.
37. Subbarao P, Anand SS, Becker AB, et al. The Canadian Healthy Infant Longitudinal Development (CHILD) Study: examining developmental origins of allergy and asthma: Table 1. *Thorax* 2015; 70: 998–1000.
38. Dekker PM, Azad MB, Boeren S, et al. The human milk proteome and allergy of mother and child: Exploring associations with protein abundances and protein network connectivity. *Front Immunol* 2022; 13: 977470.
39. Zhu J, Dingess KA. The Functional Power of the Human Milk Proteome. *Nutrients* 2019; 11: 1834.
40. Dekker PM, Boeren S, Saccenti E, et al. Network analysis of the proteome and peptidome sheds light on human milk as a biological system. *Sci Rep* 2024; 14: 7569.
41. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
42. Dekker PM, Boeren S, van Goudoever JB, et al. Exploring Human Milk Dynamics: Interindividual Variation in Milk Proteome, Peptidome, and Metabolome. *J Proteome Res* 2022; 21: 1002–1016.
43. Zaverdas H, Stojceski F, Romero-Zaliz R, et al. Unravelling the human taste receptor interactome: machine learning and molecular modelling insights into protein-protein interactions. *npj Sci Food*; 9. Epub ahead of print 1 July 2025. DOI: 10.1038/s41538-025-00478-9.
44. Theofilatos K, Stojkovic S, Hasman M, et al. Proteomic atlas of atherosclerosis: the contribution of proteoglycans to sex differences, plaque phenotypes, and outcomes. *Circulation research* 2023; 133: 542–558.
45. Theofilatos K, Stojkovic S, Hasman M, et al. Proteomic Atlas of Atherosclerosis: The Contribution of Proteoglycans to Sex Differences, Plaque Phenotypes, and Outcomes | Circulation Research.

Circulation Research; 133. Epub ahead of print 30 August 2023. DOI: 10.1161/CIRCRESAHA.123.322590.

46. Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 1998; 30: 107–117.
47. Zhang P, Wang T, Yan J. PageRank centrality and algorithms for weighted, directed networks. *Physica A: Statistical Mechanics and its Applications* 2022; 586: 126438.
48. Brandtzaeg P. Mucosal immunity: integration between mother and the breast-fed infant. *Vaccine* 2003; 21: 3382–3388.
49. Sartore G, Piarulli F, Ragazzi E, et al. Circulating Factors as Potential Biomarkers of Cardiovascular Damage Progression Associated with Type 2 Diabetes. *Proteomes* 2024; 12: 29.
50. Zhang Z, Chen Z. Exploring the causal links between inflammation-related genes and atherosclerosis through Mendelian randomization analysis. *Medicine* 2025; 104: e42584.
51. Shao X, Hou X, Zhang X, et al. Integrated single-cell RNA-seq analysis reveals the vital cell types and dynamic development signature of atherosclerosis. *Front Physiol* 2023; 14: 1118239.
52. Mazzuca C, Vitiello L, Travaglini S, et al. Immunological and homeostatic pathways of alpha -1 antitrypsin: a new therapeutic potential. *Front Immunol* 2024; 15: 1443297.
53. Cochran BJ, Ong K-L, Manandhar B, et al. APOA1: a Protein with Multiple Therapeutic Functions. *Curr Atheroscler Rep* 2021; 23: 11.
54. Zhang X, Jiang B, Ji C, et al. Quantitative Label-Free Proteomic Analysis of Milk Fat Globule Membrane in Donkey and Human Milk. *Front Nutr* 2021; 8: 670099.
55. Zhang X, Jiang G, Ji C, et al. Comparative Whey Proteome Profiling of Donkey Milk With Human and Cow Milk. *Front Nutr* 2022; 9: 911454.
56. Abe RJ, Abe J, Nguyen MTH, et al. Free Cholesterol Bioavailability and Atherosclerosis. *Curr Atheroscler Rep* 2022; 24: 323–336.
57. Purushothaman K-R, Purushothaman M, Levy AP, et al. Increased Expression of Oxidation-Specific Epitopes and Apoptosis Are Associated With Haptoglobin Genotype. *Journal of the American College of Cardiology* 2012; 60: 112–119.
58. Schaer CA, Deuel JW, Bittermann AG, et al. Mechanisms of haptoglobin protection against hemoglobin peroxidation triggered endothelial damage. *Cell Death Differ* 2013; 20: 1569–1579.
59. Naba A. Mechanisms of assembly and remodelling of the extracellular matrix. *Nat Rev Mol Cell Biol* 2024; 25: 865–885.
60. Hossain N, Jeong Y. Protective Genes Against Alzheimer's Disease: Case Review and Therapeutic Implications. *Dement Neurocogn Disord* 2025; 24: 75.
61. Zhang Y, Lv X, Chen L, et al. The role and function of CLU in cancer biology and therapy. *Clin Exp Med* 2022; 23: 1375–1391.
62. Hok KD, Rich HE, Shadid A, et al. Functional Roles of the Complement Immune System in Cardiac Inflammation and Hypertrophy. *IJMS* 2025; 26: 9931.

63. Gori A, Brindisi G, Daglia M, et al. Exploring the Role of Lactoferrin in Managing Allergic Airway Diseases among Children: Unrevealing a Potential Breakthrough. *Nutrients* 2024; 16: 1906.
64. Luyendyk JP, Schoenecker JG, Flick MJ. The multifaceted role of fibrinogen in tissue injury and inflammation. *Blood* 2019; 133: 511–520.
65. Hsieh JY, Smith TD, Meli VS, et al. Differential regulation of macrophage inflammatory activation by fibrin and fibrinogen. *Acta Biomaterialia* 2017; 47: 14–24.
66. Lewis M. PRELP, collagen, and a theory of Hutchinson–Gilford progeria. *Ageing Research Reviews* 2003; 2: 95–105.
67. Jacobson KR, Lipp S, Acuna A, et al. Comparative Analysis of the Extracellular Matrix Proteome across the Myotendinous Junction. *J Proteome Res* 2020; 19: 3955–3967.
68. Delfín DA, DeAguero JL, McKown EN. The Extracellular Matrix Protein ABI3BP in Cardiovascular Health and Disease. *Front Cardiovasc Med* 2019; 6: 23.
69. Nulali J, Zhan M, Zhang K, et al. Osteoglycin: An ECM Factor Regulating Fibrosis and Tumorigenesis. *Biomolecules* 2022; 12: 1674.
70. Ge G, Seo N-S, Liang X, et al. Bone Morphogenetic Protein-1/Tolloid-related Metalloproteinases Process Osteoglycin and Enhance Its Ability to Regulate Collagen Fibrillogenesis. *Journal of Biological Chemistry* 2004; 279: 41626–41633.
71. Van Herwijnen MJC, Zonneveld MI, Goerdal S, et al. Comprehensive Proteomic Analysis of Human Milk-derived Extracellular Vesicles Unveils a Novel Functional Proteome Distinct from Other Milk Components. *Molecular & Cellular Proteomics* 2016; 15: 3412–3423.
72. Dekker PM, Azad MB, Boeren S, et al. The human milk proteome and allergy of mother and child: Exploring associations with protein abundances and protein network connectivity. *Front Immunol*; 13. Epub ahead of print 13 October 2022. DOI: 10.3389/fimmu.2022.977470.
73. Dekker PM, Boeren S, Saccenti E, et al. Network analysis of the proteome and peptidome sheds light on human milk as a biological system. *Sci Rep*; 14. Epub ahead of print 30 March 2024. DOI: 10.1038/s41598-024-58127-2.
74. Linzell JL, Peaker M. Mechanism of milk secretion. *Physiological Reviews* 1971; 51: 564–597.
75. Zhu J, Dingess KA. The Functional Power of the Human Milk Proteome. *Nutrients* 2019; 11: 1834.
76. Akira S, Uematsu S, Takeuchi O. Pathogen Recognition and Innate Immunity. *Cell* 2006; 124: 783–801.
77. Newburg DS. Innate Immunity and Human Milk. *The Journal of Nutrition* 2005; 135: 1308–1312.
78. Truchet S, Honvo-Houéto E. Physiology of milk secretion. *Best Practice & Research Clinical Endocrinology & Metabolism* 2017; 31: 367–384.
79. Dallas DC, Murray NM, Gan J. Proteolytic Systems in Milk: Perspectives on the Evolutionary Function within the Mammary Gland and the Infant. *J Mammary Gland Biol Neoplasia* 2015; 20: 133–147.
80. Dallas DC, Guerrero A, Khaldi N, et al. Extensive *in vivo* Human Milk Peptidomics Reveals Specific Proteolysis Yielding Protective Antimicrobial Peptides. *J Proteome Res* 2013; 12: 2295–2304.

81. Dingess KA, Van Den Toorn HWP, Mank M, et al. Toward an efficient workflow for the analysis of the human milk peptidome. *Anal Bioanal Chem* 2019; 411: 1351–1363.
82. Gomes AV. Genetics of Proteasome Diseases. *Scientifica* 2013; 2013: 1–30.
83. Foot N, Henshall T, Kumar S. Ubiquitination and the Regulation of Membrane Proteins. *Physiological Reviews* 2017; 97: 253–281.
84. Clague MJ, Urbé S. Integration of cellular ubiquitin and membrane traffic systems: focus on deubiquitylases. *The FEBS Journal* 2017; 284: 1753–1766.
85. Selvitopi O, Hussain MT, Azad A, et al. Optimizing High Performance Markov Clustering for Pre-Exascale Architectures. Epub ahead of print 2020. DOI: 10.48550/ARXIV.2002.10083.
86. Van Dongen S, Abreu-Goodger C. Using MCL to Extract Clusters from Networks. In: Van Helden J, Toussaint A, Thieffry D (eds) *Bacterial Molecular Networks*. New York, NY: Springer New York, pp. 281–295.
87. Kalecky K, Cho Y-R. PrimAlign: PageRank-inspired Markovian alignment for large biological networks. *Bioinformatics* 2018; 34: i537–i546.
88. Azad A, Pavlopoulos GA, Ouzounis CA, et al. HipMCL: a high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. *Nucleic Acids Research* 2018; 46: e33–e33.
89. Brandtzaeg P. Secretory IgA: Designed for Anti-Microbial Defense. *Front Immunol*; 4. Epub ahead of print 2013. DOI: 10.3389/fimmu.2013.00222.
90. Vidarsson G, Dekkers G, Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front Immunol*; 5. Epub ahead of print 20 October 2014. DOI: 10.3389/fimmu.2014.00520.
91. Tiller T, Meffre E, Yurasov S, et al. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *Journal of Immunological Methods* 2008; 329: 112–124.
92. Cavassan NRV, Camargo CC, De Pontes LG, et al. Correlation between chronic venous ulcer exudate proteins and clinical profile: A cross-sectional study. *Journal of Proteomics* 2019; 192: 280–290.
93. Rusanov VB, Pastushkova LH, Goncharova AG, et al. Reflection of Heart Rate Physiological Regulation Parameters in the Urinary Proteome in Healthy Young Males. *Hum Physiol* 2020; 46: 182–190.
94. Fuster JJ. Clonal hematopoiesis and cardiovascular disease in cancer patients and survivors. *Thrombosis Research* 2022; 213: S107–S112.
95. Viecei Dalla Sega F, Palumbo D, Fortini F, et al. Transcriptomic profiling of calcified aortic valves in clonal hematopoiesis of indeterminate potential carriers. *Sci Rep* 2022; 12: 20400.
96. Lin M, Zhao L, Zhao W, et al. Dissecting the mechanism of carotid atherosclerosis from the perspective of regulation. *International Journal of Molecular Medicine* 2014; 34: 1458–1466.

97. Liu X, Li L, Yin Y, et al. Single-cell transcriptomic, transcriptomic, and metabolomic characterization of human atherosclerosis. *Ann Transl Med* 2022; 10: 1215–1215.
98. Dong J, Song C, Zhang L, et al. Identified key genes related to carotid atheroma plaque from gene expression chip. *Artificial Cells, Nanomedicine, and Biotechnology* 2017; 45: 1132–1137.
99. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004; 5: 101–113.
100. Shu CC, Flannery CR, Little CB, et al. Catabolism of Fibromodulin in Developmental Rudiment and Pathologic Articular Cartilage Demonstrates Novel Roles for MMP-13 and ADAMTS-4 in C-terminal Processing of SLRPs. *IJMS* 2019; 20: 579.
101. Dupuis LE, Doucette L, Rice AK, et al. Development of myotendinous-like junctions that anchor cardiac valves requires fibromodulin and lumican. *Developmental Dynamics* 2016; 245: 1029–1042.
102. Lama P, Tamang BK. Regulatory Role and Significance of Class II SLRPs: Fibromodulin, Lumican, Osteoadherin, PRELP, and Keratocan. In: Lama P (ed) *The Biological Role of Small Leucine-Rich Proteoglycans (SLRPs) in Matrix Homeostasis and Diseases*. Singapore: Springer Nature Singapore, pp. 101–120.
103. Faarvang A-SA, Overgaard M, Jensen PS, et al. Arterial collagen IV and laminin levels associate with type 2 diabetes, glycemic control, major adverse cardiac events, and mortality in coronary bypass patients. *Cardiovasc Diabetol* 2025; 24: 458.
104. Zhuang J, Xie L, Chen H, et al. Correlation of Serum Laminin Levels with Cardiac Function and In-Hospital Prognosis in Patients with Atrial Fibrillation. *Int Heart J* 2023; 64: 172–177.
105. Hollier BG, Tinnirello AA, Werden SJ, et al. FOXC2 Expression Links Epithelial–Mesenchymal Transition and Stem Cell Properties in Breast Cancer. *Cancer Research* 2013; 73: 1981–1992.
106. Sipos F. Epithelial-to-mesenchymal and mesenchymal-to-epithelial transitions in the colon. *WJG* 2012; 18: 601.
107. Du W, Liu X, Yang M, et al. The Regulatory Role of PRRX1 in Cancer Epithelial-Mesenchymal Transition. *OTT* 2021; Volume 14: 4223–4229.
108. Velazquez JJ, LeGraw R, Moghadam F, et al. Gene Regulatory Network Analysis and Engineering Directs Development and Vascularization of Multilineage Human Liver Organoids. *Cell Systems* 2021; 12: 41-55.e11.
109. Cai D, Liu H, Zhao R. Nuclear Receptors in Hepatic Glucose and Lipid Metabolism During Neonatal and Adult Life. *CPPS* 2017; 18: 548–561.
110. Cheng W, Li X, Liu D, et al. Endothelial-to-Mesenchymal Transition: Role in Cardiac Fibrosis. *J Cardiovasc Pharmacol Ther* 2021; 26: 3–11.
111. Sweeney DW, Shen M-C, Farber SA. Creb3l3 deficiency promotes intestinal lipid accumulation and alters ApoB-containing lipoprotein kinetics. *Journal of Lipid Research* 2025; 66: 100833.
112. Lefebvre P, Cariou B, Lien F, et al. Role of Bile Acids and Bile Acid Receptors in Metabolic Regulation. *Physiological Reviews* 2009; 89: 147–191.

113. Chang C-C, Zhang C, Zhang Q, et al. Upregulation of lactate dehydrogenase a by 14-3-3 ζ leads to increased glycolysis critical for breast cancer initiation and progression. *Oncotarget* 2016; 7: 35270–35283.
114. Jovanović S, Jovanović A, Crawford RM. M-LDH Serves as a Regulatory Subunit of the Cytosolic Substrate-channelling Complex in Vivo. *Journal of Molecular Biology* 2007; 371: 349–361.
115. Wu J, Gan J, Zeng G, et al. Investigation of Human Milk as a Biological System in a Multicenter Mother–Infant Cohort: Protocol Design and Cohort Profile of the Phoenix Study. *Nutrients* 2024; 16: 2892.
116. Chandran D, Warren K, McKeone D, et al. The Association between Infant Colic and the Multi-Omic Composition of Human Milk. *Biomolecules* 2023; 13: 559.
117. Jung SY. Multi-Omics Data Analysis Uncovers Molecular Networks and Gene Regulators for Metabolic Biomarkers. *Biomolecules* 2021; 11: 406.
118. Ecologies, synergies, and biological systems shaping human milk composition—a report from “Breastmilk Ecology: Genesis of Infant Nutrition (BEGIN)” Working Group 2. *The American Journal of Clinical Nutrition* 2023; 117: S28–S42.
119. Cox J, Neuhauser N, Michalski A, et al. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J Proteome Res* 2011; 10: 1794–1805.
120. Lachmann A, Giorgi FM, Lopez G, et al. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* 2016; 32: 2233–2235.
121. Margolin AA, Nemenman I, Basso K, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 2006; 7: S7.
122. Khosravi P, Gazestani VH, Pirhaji L, et al. Inferring interaction type in gene regulatory networks using co-expression data. *Algorithms Mol Biol* 2015; 10: 23.
123. Meyer PE, Bontempi G. Information-Theoretic Gene Selection In Expression Data. In: Elloumi M, Zomaya AY (eds) *Biological Knowledge Discovery Handbook*. Wiley, pp. 399–420.
124. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995; 57: 289–300.
125. Theofilatos K, Dimitrakopoulos C, Alexakos C, et al. InSyBio BioNets: an efficient tool for network-based biomarker discovery. *EMBnet j* 2016; 22: 871.
126. Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research* 2022; 50: W216–W221.
127. The GTEx Consortium, Aguet F, Anand S, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020; 369: 1318–1330.
128. Keenan AB, Torre D, Lachmann A, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research* 2019; 47: W212–W224.

129. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 2003; 13: 2498–2504.
130. Van Dongen S. Graph Clustering Via a Discrete Uncoupling Process. *SIAM J Matrix Anal & Appl* 2008; 30: 121–141.
131. Vardakas G, Papakostas I, Likas A. Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters. Epub ahead of print 2024. DOI: 10.48550/ARXIV.2402.00608.
132. Traag VA, Waltman L, Van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 2019; 9: 5233.
133. Iqbal S, Hoque MT. DisPredict: A Predictor of Disordered Protein Using Optimized RBF Kernel. *PLoS ONE* 2015; 10: e0141551.
134. Shibata S, Sogabe S, Miwa M, et al. Identification of the first highly selective inhibitor of human lactate dehydrogenase B. *Sci Rep* 2021; 11: 21353.
135. Ottmann C, Yasmin L, Weyand M, et al. Phosphorylation-independent interaction between 14-3-3 and exoenzyme S: from structure to pathogenesis. *EMBO J* 2007; 26: 902–913.
136. Dexheimer TS, Carey SS, Zuohe S, et al. NM23-H2 may play an indirect role in transcriptional activation of *c-myc* gene expression but does not cleave the nuclease hypersensitive element III1. *Molecular Cancer Therapeutics* 2009; 8: 1363–1377.
137. McCue WM, Finzel BC. Structural Characterization of the Human Cytosolic Malate Dehydrogenase I. *ACS Omega* 2022; 7: 207–214.
138. Yan Y, Zhang D, Zhou P, et al. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Research* 2017; 45: W365–W373.
139. Yan Y, Tao H, He J, et al. The HDock server for integrated protein–protein docking. *Nat Protoc* 2020; 15: 1829–1852.
140. Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 1983; 79: 926–935.
141. Tian C, Kasavajhala K, Belfon KAA, et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* 2020; 16: 528–552.
142. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* 2007; 126: 014101.
143. Bernetti M, Bussi G. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* 2020; 153: 114107.
144. Hopkins CW, Le Grand S, Walker RC, et al. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* 2015; 11: 1864–1874.
145. Essmann U, Perera L, Berkowitz ML, et al. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 1995; 103: 8577–8593.

146. Johnson JK, Zollweg JA, Gubbins KE. The Lennard-Jones equation of state revisited. *Molecular Physics* 1993; 78: 591–618.
147. Hess B, Bekker H, Berendsen HJC, et al. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 1997; 18: 1463–1472.
148. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996; 14: 33–38.
149. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015; 1–2: 19–25.
150. Kutzner C, Páll S, Fechner M, et al. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J Comput Chem* 2019; 40: 2418–2431.

Figures

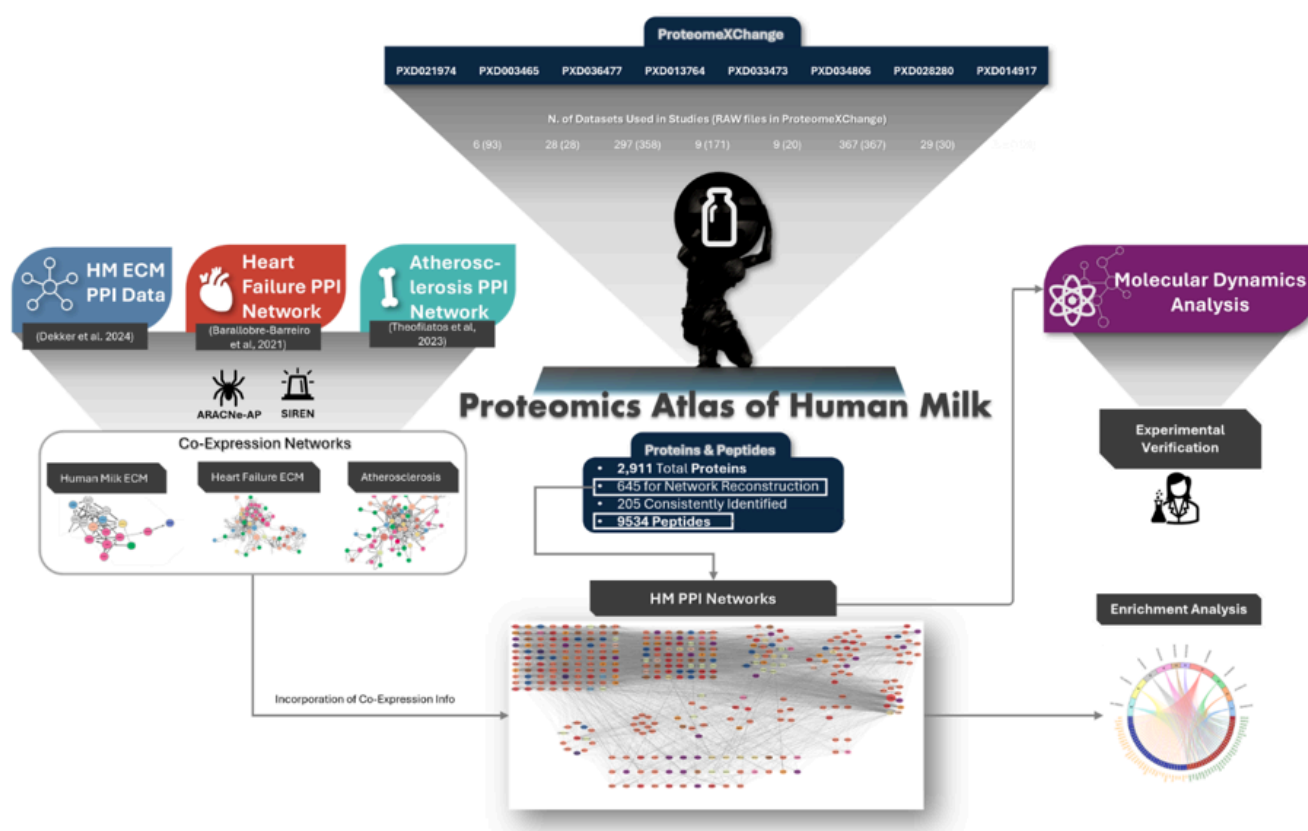


Figure 1

Overview of meta-analysis of publicly available proteomics data of human milk. Overall workflow for proteomics data analysis, regulatory and Protein-Protein Interaction network reconstruction, and Molecular Dynamics Analysis of selected interactions, which led to the establishment of the Proteomics Atlas of Human Milk.

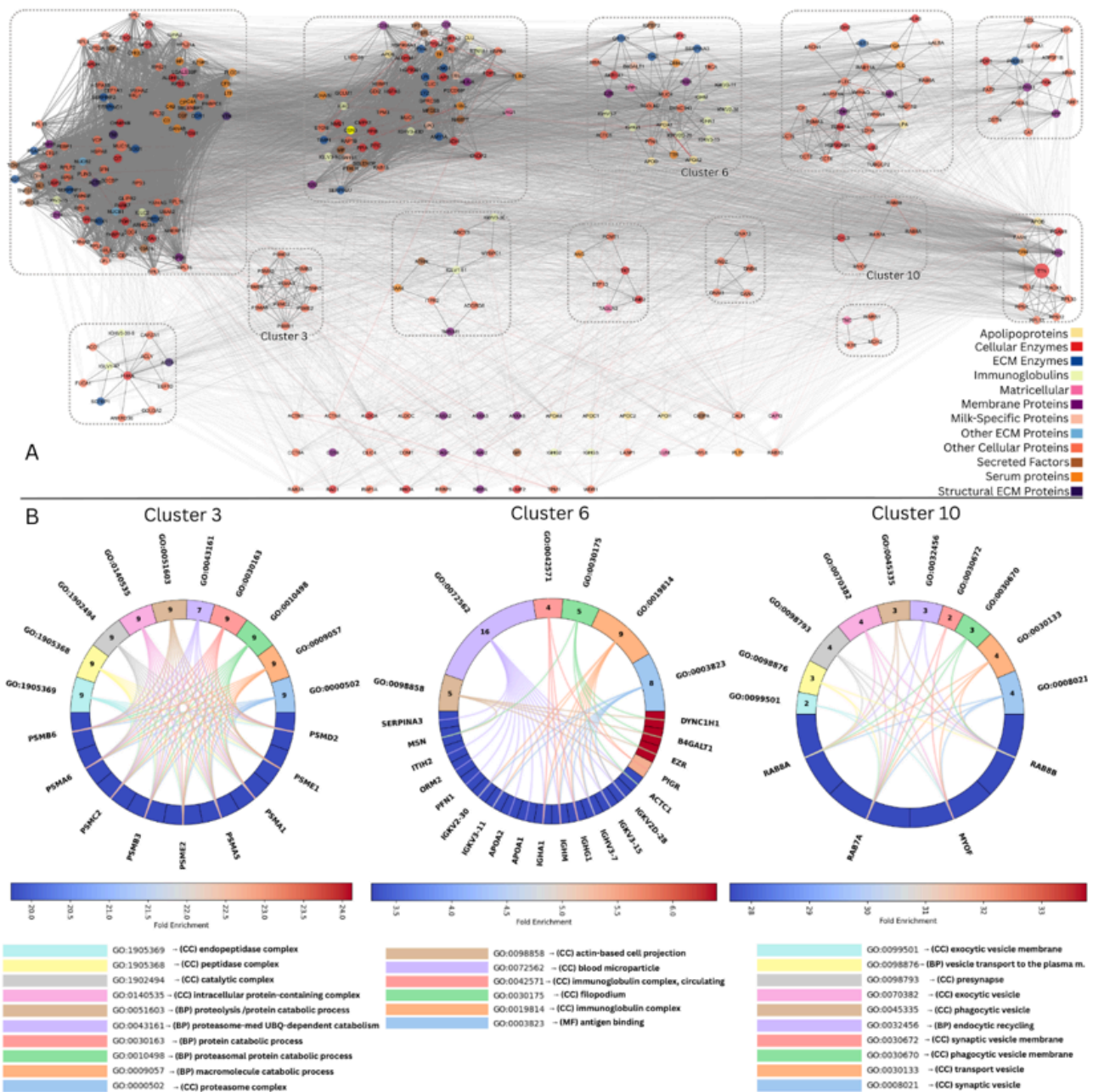


Figure 2

Reconstructed Protein-Protein Interaction Network of Human Milk. A) PPI network of Human Milk Proteins visualized with Cytoscape. Edge width is analogous to the corresponding edge weight in the same network. Node size is analogous to the Betweenness Centrality of the node, while node color corresponds to the protein groups according to the legend on the bottom right. Clusters are enclosed in grey dotted lines. The unclustered nodes are the ones involved in the HM Co-Expression Network's edges, while the rest were filtered out. B) Circular graphs of the GO Term enrichment analysis results of the 3 top-ranked clusters based on the Silhouette-Coefficient. On the bottom half of each circle are the significantly enriched Genes of each cluster, ranked from left to right with ascending Fold Enrichment values (also colored from blue to red, respectively). Only Genes with adjusted p-value<0.05 are depicted

in the graphs. GO Term clusters are present on the top half of the circle with distinct colors that are mapped below the graph to the GO Term description. Inside each GO term cluster cell, the number of Genes involved in the cluster is noted.

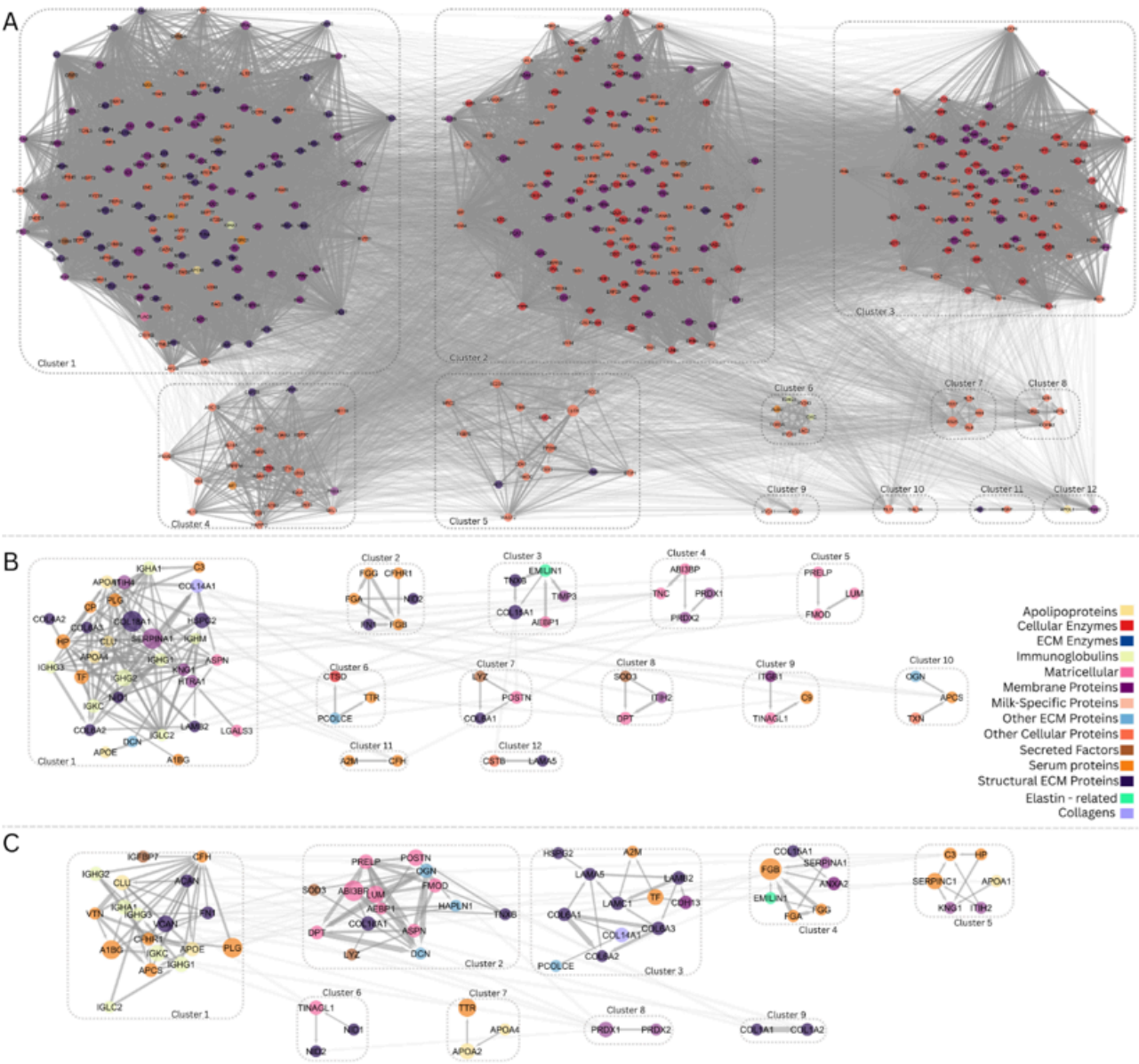


Figure 3

Human Milk and Heart Disease-related protein regulatory networks. A) Human Milk (top 12 clusters), B) Atherosclerosis, C) Heart Failure. Edge width is analogous to the corresponding Mutual Information (MI) index of the regulatory interaction, computed from ARACNe-AP reconstruction. Node size is analogous to the Betweenness Centrality of the node, while node color corresponds to the protein groups according to the legend on the middle-right. Clusters are enclosed in grey dotted lines. The unclustered nodes were filtered out.

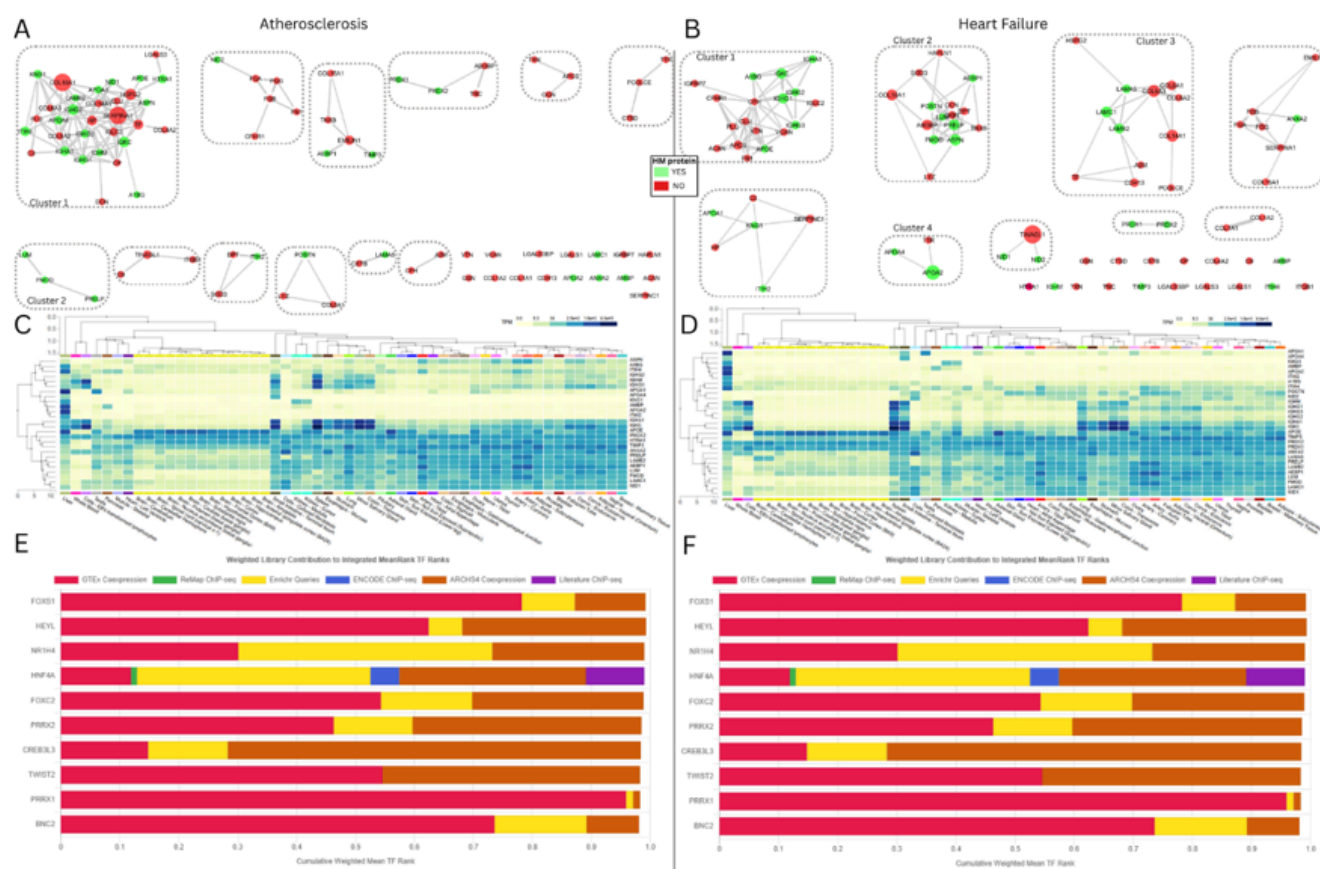


Figure 4

Regulatory network, tissue expression, and transcriptional enrichment of signatures shared between the human milk (HM) proteome and cardiovascular disease tissues. Regulatory networks are derived from the interactions of the common protein signatures of the HM regulatory network and (A) the atherosclerosis or (B) heart failure network. Green color represents the common proteins between HM and the respective network, while red represents the non-common ones that are interactors in these networks. GTEx-based tissue expression profiles of proteins (C) shared between human milk and atherosclerosis signatures, and (D) shared between human milk and heart failure signatures. Transcription factor enrichment analysis of proteins shared between (E) human milk and atherosclerosis signatures and (F) human milk and heart failure signatures, using the ChEA3 platform.

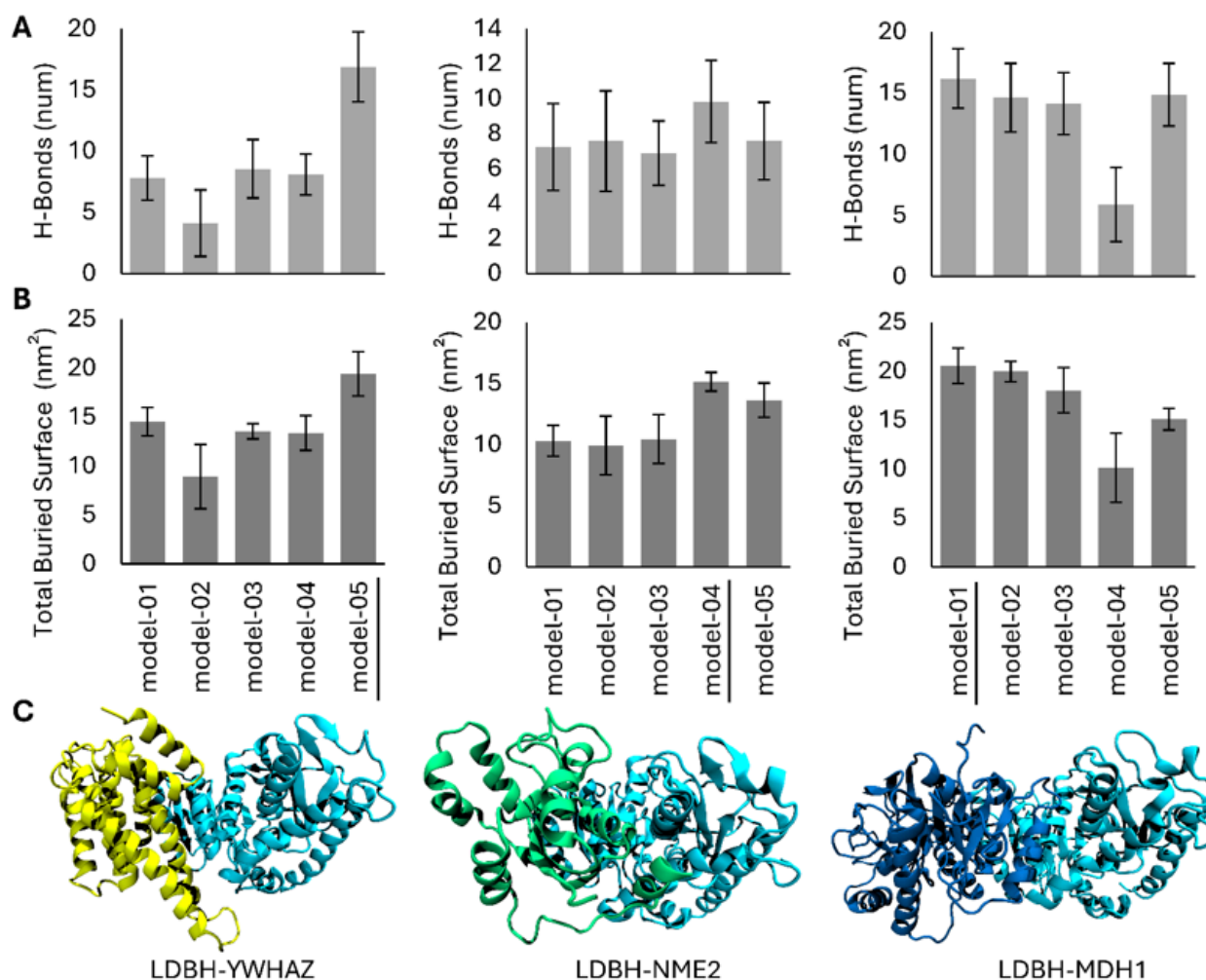


Figure 5

Hydrogen Bonds and Buried Surface Area calculations for each pose of the top-ranked Human Milk PPIs. A) Hydrogen bond analysis of the 5 docking models for each of the 3 selected PPI pairs: LDHB–YWHAZ, LDHB–NME2, and LDHB–MDH1. The results are reported as the average and standard deviation of the 3 replicas on each docking model. B) Total buried surface analysis of the five docking models for each of the three selected PPI pairs: LDHB–YWHAZ, LDHB–NME2, and LDHB–MDH1. The results are reported as average and standard deviation of the 3 replicas on each docking model. C) Qualitative representations of the 3 PPI pairs involved in the highest interacting models.

Supplementary Files

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

- [HumanMilkManuscriptSuppInfoNatFoodfinal.docx](#)