The hidden geometry of the human interactome in hyperbolic space

Dissertation zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Aimilia- Christina Vagiona

geb. am 14.05.1989 in Serres, Greece

Mainz, 2025

Dekan: [redacted]

1. Berichterstatter: [redacted]

2. Berichterstatter: [redacted]

Tag der mündlichen Prüfung:

Aimilia Christina Vagiona

The hidden geometry of the human interactome in hyperbolic space

Thesis, October 2025

Reviewers: [redacted]

Supervision: [redacted]

Johannes Gutenberg University Mainz

[redacted] Computational Biology and Data Mining

Institute of Organismic and Molecular Evolution (iOME)

Faculty of Biology

Hanns-Dieter-Hüsch-Weg 15

55128 Mainz

Acknowledgements

[redacted]

Table of Contents

1. Zusammenfassung

Proteine bilden dynamische Netzwerke von Interaktionen, die nahezu jeden Aspekt der Zellfunktion unterstützen und regulieren von Signalübertragung und Regulation bis hin zur Bildung makromolekularer Komplexe. Die Darstellung dieser Protein-Protein-Interaktionen (PPIs) in Netzwerken bietet eine leistungsstarke Möglichkeit, die zelluläre Organisation zu untersuchen. Allerdings bleiben herkömmliche Darstellungen von Protein-Protein-Interaktionsnetzwerken (PPINs) begrenzt: Sie sind oft unvollständig, statisch und auf binäre Assoziationen ausgerichtet, wodurch sie nur einen Teil der Komplexität molekularer Systeme widerspiegeln.

Diese Arbeit untersucht das Konzept der hyperbolischen Geometrie als Ansatz, um die verborgene Struktur des menschlichen Interaktoms aufzudecken. Durch die Einbettung von Proteinen in den hyperbolischen Raum wird es möglich, sowohl ihre hierarchische Organisation als auch ihre funktionale Ähnlichkeit auf eine Weise darzustellen, die der biologischen Realität näherkommt. Radiale Koordinaten erfassen evolutionäre Konservierung und Konnektivität, während die Winkelproximität funktionale Ähnlichkeit widerspiegelt. Zusammengenommen liefern diese Koordinaten aussagekräftige Merkmale, die über klassische Netzwerkmaße hinausgehen.

Mit diesem Ansatz wurden verschiedene Aspekte von Proteinfunktionen untersucht. Krankheitsmodule konnten innerhalb der Geometrie des Interaktoms identifiziert und interpretiert werden, was Einblicke in die Rolle von Interaktionsstörungen bei pathologischen Prozessen ermöglicht. Direktionale regulatorische Beziehungen, wie etwa Phosphorylierungen, wurden durch die Integration hyperbolischer Merkmale mit maschinellen Lernverfahren vorhergesagt und zeigten, dass geometriebasierte Ansätze biologische Zusammenhänge erfassen können. Über paarweise Interaktionen hinaus offenbarte die Untersuchung von Protein-Triplets, wie die Netzwerksgeometrie mit kooperativem und kompetitivem Binden zusammenhängt ein Befund, der durch Strukturmodellierungen mit Hilfe von AlphaFold3 gestützt wurde. Schließlich wurde der Ansatz auf die Multi-Omics-Integration ausgeweitet, indem Proteine, Metabolite und Lipide in eine gemeinsame hyperbolische Karte eingebettet wurden. Dies ermöglichte die Identifizierung von Cross-Layer-Modulen mit Relevanz für kardiovaskuläre und metabolische Erkrankungen.

Insgesamt zeigen die Ergebnisse, dass hyperbolische Einbettungen eine biologisch aussagekräftige Repräsentation molekularer Netzwerke bieten. Durch die Verbindung von Topologie mit Strukturbiologie und Multi-Omics-Daten etabliert diese Arbeit eine geometrische Sichtweise der Systembiologie, die unser Verständnis der zellulären Organisation erweitert und neue Perspektiven für die Untersuchung von Krankheitsmechanismen und therapeutischen Ansätzen eröffnet.

2. Summary

Proteins form dynamic networks of interactions that support and regulate almost every aspect of cellular function, from signaling and regulation to the formation of macromolecular assemblies. Mapping these protein-protein interactions (PPIs) into networks provides a powerful way to study cellular organization. However, traditional representations of protein-protein interaction networks (PPINs) remain limited: they are often incomplete, static, and biased toward binary associations, offering only a partial view of the complexity of molecular systems.

This thesis explores hyperbolic geometry as an approach to uncover the hidden structure of the human interactome. By embedding proteins into hyperbolic space, it becomes possible to represent both their hierarchical organization and functional similarity in a way that is more relevant to biological reality. Radial coordinates capture evolutionary conservation and connectivity, while angular proximity reflects functional similarity, together providing meaningful features that go beyond traditional network measures.

This method was applied to investigate diverse aspects of protein biology. Disease modules could be identified and interpreted within the geometry of the interactome, offering insights into how perturbations in interactions contribute to pathology. Directional regulatory relationships, such as phosphorylation, were predicted by integrating hyperbolic features with machine learning approaches, revealing that geometry-based approaches can capture biological relationships. Moving beyond pairwise interactions, the study of protein triplets revealed how network geometry relates to cooperative and competitive binding, a finding supported by structural modeling (AlphaFold3). Finally, by extending the approach to multi-omics integration, proteins, metabolites, and lipids were embedded into a hyperbolic map, enabling the discovery of cross-layer modules with relevance for cardiovascular and metabolic disease.

Together, these findings demonstrate that hyperbolic embeddings provide a biologically meaningful representation of molecular networks. By bridging topology with structural biology and multi-omics data, this thesis establishes a geometric view of systems biology that advances our understanding of cellular organization and offers new perspectives for studying disease mechanisms and therapeutic opportunities.

3. Introduction

3.1 Protein-Protein Interaction Networks

Proteins are fundamental molecules in the cell, performing essential cellular functions that range from catalyzing metabolic and enzymatic reactions and coordinating intracellular signaling pathways, to organizing cytoskeletal and membrane-associated complexes and directing transcriptional and epigenetic gene regulation [1–3]. These functions rarely occur in isolation. Instead, proteins commonly operate through physical and functional interactions with other proteins, forming Protein-Protein Interactions (PPIs) that are important for coordinating complex biological processes. The comprehensive map of these interactions within a cell or organism is referred to as the interactome. This large-scale interaction landscape can be systematically represented as a Protein-Protein Interaction Network (PPINs), which models proteins as nodes and their associations as edges [4–6].

Protein-protein interaction networks are typically constructed through a combination of high-throughput experimental techniques that aim to systematically identify physical associations between proteins under various biological conditions. Among the most widely used methods is the yeast two-hybrid (Y2H) assay, a genetic screening technique that detects direct, binary interactions between protein pairs. In this method, protein coding sequences are fused to DNA-binding and activation domains, and physical interaction between the proteins reassembles a functional transcription factor, leading to reporter gene expression in yeast cells [7]. Y2H is particularly suited for discovering novel and transient interactions, especially those that occur in the nucleus, although it can yield false positives due to its artificial context and lack of post-translational modifications. Another foundational technique is affinity purification coupled with mass spectrometry (AP-MS), which isolates protein complexes from native cellular environments. In AP-MS, a tagged "bait" protein is expressed in cells, and its interacting partners are co-purified through biochemical isolation and then identified by mass spectrometry [8]. This approach is highly effective for identifying stable and physiologically relevant multi-protein assemblies, especially those that participate in macromolecular machines such as the ribosome, proteasome, or chromatin remodeling complexes. Variants like tandem affinity purification (TAP) or proximity labeling methods (e.g., BioID, APEX) further enhance the specificity and depth of interaction profiling [9–11].

To organize and interpret the large amount of data generated by these experimental approaches, protein-protein interactions are systematically stored in specialized PPI databases, which vary in their scope and methodology. These resources are generally classified into three categories. Primary databases such as BioGRID [12], IntAct [13], and DIP [14] curate experimentally validated interactions directly from the literature, often including detailed annotations about the experimental methods, organism, and context in which each interaction was detected. In contrast,

meta-databases or secondary, including HIPPIE [15,16] and Mentha [17], act as integrative platforms that merge data from multiple primary sources, standardize identifiers, and frequently assign confidence scores based on the type and reproducibility of experimental evidence. These resources do not generate new predictions but offer a consensus score of known interactions. Finally, predictive databases, such as STRING [18], combine experimental interaction data with computational predictions based on genomic context, co-expression, evolutionary conservation, and literature mining. These platforms aim to provide a more comprehensive interactome, often including confidence metrics or interaction probabilities to guide biological interpretation (Figure 1).

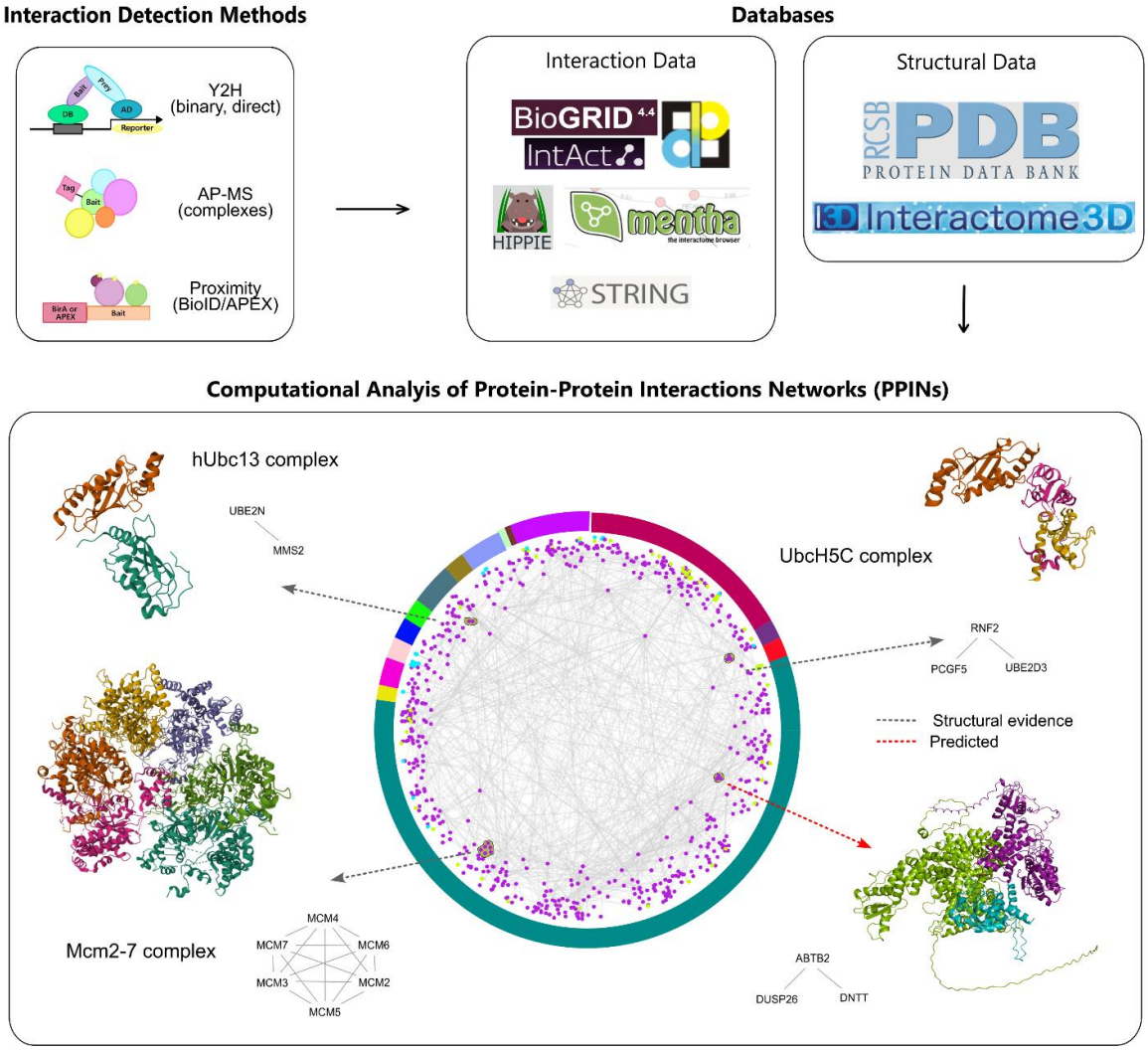


Figure 1. Overview of protein-protein interaction network construction and analysis*. Experimental methods such as Y2H, AP-MS, and BioID/APEX provide large-scale datasets of protein-protein interactions (PPIs). These data are systematically curated in primary databases (e.g., BioGRID, IntAct), meta-databases (e.g., HIPPIE, Mentha), and predictive resources (e.g., STRING). Structural information from repositories such as the Protein Data Bank (PDB) and Interactome3D enables validation and interpretation of PPIs at the molecular level. Computational network analysis can then reveal binary complexes or higher-order assemblies, and structurally validated or predicted triplets, providing insights into cellular organization and function.*

However, current representations of protein-protein interaction networks (PPINs) come with important limitations. Many interaction datasets are incomplete and biased toward proteins that are easier to detect or study experimentally, leading to an uneven representation of the interactome, especially for the proteins that are hard to detect experimentally [19]. Moreover, most PPINs are static, failing to capture the dynamic properties of protein interactions that occur in vivo [20,21] . Another major challenge is that most interaction data describe binary associations, which do not fully represent the complexity of the cellular environment, where proteins typically act as part of multi-subunit assemblies or participate in higher-order cooperative mechanisms. The dense molecular environment inside the cell complicates the analysis of protein interactions, as proteins frequently interact nonspecifically under physiological conditions. While often this is considered as noise, such interactions can influence functional associations by altering concentration or protein conformations [22]. This highlights the necessity of discriminating functionally relevant protein interactions from incidental ones when analyzing PPINs. Addressing these challenges requires complementary approaches, such as network geometry, which can capture latent topological functional organization that are not evident in traditional network representations.

Despite these challenges, PPINs remain fundamental tools for systems biology. They support the discovery of new protein functions, enable the identification of disease-associated modules, provide a foundation for the identification and evaluation of potential therapeutic targets, and support functional annotation [23–25]. In addition, integrating PPINs with gene expression, phenotypic, or post-translational modification data allows the construction of condition-specific interaction maps that more faithfully capture the biological context [26]. Understanding how PPINs are constructed, what biological aspects they represent, and where their limitations lie is necessary for understanding protein function in health and disease.

3.2 Network Properties and Centralities

The behavior of most complex systems, such as biological systems emerges from the coordinated activity of many components that interact with each other through pairwise interactions [27]. This complexity is well represented through network models, in which biological entities are captured as nodes and their interactions as connecting edges [28]. Among the various types of biological networks, protein-protein interaction networks (PPINs) are particularly informative for studying cellular function [29]. Over the past two decades, network-based analysis of PPIs has become a foundational tool in systems biology, enabling researchers to explore the structural and functional organization of the proteome [30].

The key to network-based analysis lies in the application of graph theory, a branch of mathematics that provides computational tools to model and analyze complex systems as networks. Graph theory enables not only the visualization of network structure but also the quantification of node properties and global topological features [31]. This approach allows researchers to also

characterize the role of each protein within the network, identify critical components, and detect functional modules or communities [32]. One of the central goals in biological network analysis is to quantify the topological significance of individual proteins within the interactome. This is commonly achieved using centrality measures, a class of graph-theoretical metrics designed to capture different aspects of a node's influence or position in the network [33]. These metrics are often used as indicators of biological importance, as proteins with topologically central roles often correspond to essential genes, regulatory hubs, or integrators of multiple pathways (Figure 2)[34].

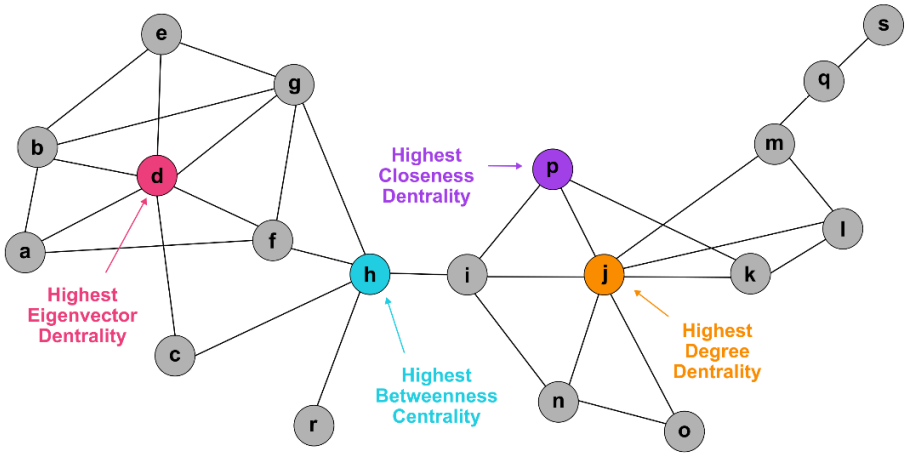


Figure 2: Diverse centrality measures. *Example network highlighting nodes with the highest values for different centrality measures. Node j (orange) has the highest degree centrality, node h (cyan) the highest betweenness centrality, node p (purple) the highest closeness centrality, and node d (pink) the highest eigenvector centrality. These metrics capture supplementary aspects of node importance within the interactome.*

Degree centrality provides a straightforward way to assess a node's connectivity within a network, and it quantifies the number of direct interactions (edges) a node has. In the context of protein-protein interaction networks, it corresponds to the number of direct physical or functional partners of a protein. In graph theory, the degree centrality is defined as:

$$DC_{(i)} = \frac{d_{(i)}}{x-1}$$

where $d_{(i)}$ is the degree (number of neighbors) of a node i and x is the total number of nodes [35].

Biologically, proteins with high degree centrality often act as network hubs, playing essential roles in cellular processes and contributing significantly to network robustness [36]. These hub proteins are frequently evolutionarily conserved and are implicated in multiple functional pathways, making them key targets for functional annotation, disease association studies, and drug discovery efforts [37,38].

While degree centrality focuses on immediate connectivity, betweenness centrality highlights a protein's role in global communication. Mathematically, the betweenness centrality is defined as:

$$BC_{(i)} = \sum_{\substack{s,t \in V \\ s \neq t \neq i}} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths between nodes s and t and $\sigma_{st}(i)$ is the number of those paths that pass-through node i. This metric captures the influence node has over the flow of information or interaction potential within the network [39].

In the context of protein-protein interaction networks (PPINs), proteins with high betweenness centrality are often considered bridging nodes. These proteins may not have the highest number of direct interactions (degree centrality), but they play crucial roles in maintaining network connectivity by linking otherwise distant regions of the network [40]. As such, their removal or dysfunction can disrupt communication or signal transduction across the network [41]. Proteins with high betweenness centrality are frequently enriched for regulatory roles, such as kinases and transcription factors, and have been associated with essential cellular processes including cell cycle regulation and stress response [34,42]. From a systems biology perspective, betweenness centrality offers an insightful viewpoint for identifying proteins that could serve as potential drug targets, especially when therapeutic strategies aim to disrupt specific signaling pathways without broadly impairing network integrity [37,43]. In disease networks, proteins with perturbed betweenness profiles may reflect rewiring of cellular processes due to mutations or environmental changes [44]. Despite being relatively easy to interpret, betweenness centrality is computationally more demanding than simpler measures like degree, especially in large networks [45] . It is also sensitive to noise and incompleteness in the network, which are common issues in experimental interactome data [46]. As such, its use is often complemented by other centrality measures to robustly identify functionally significant proteins [47].

Another key measure of node importance in biological networks is closeness centrality, which reflects how close a protein is, on average, to all other proteins in the network. Unlike degree or betweenness centrality, closeness emphasizes the global efficiency of information spread from a given node throughout the entire network [48].

$$CC_{(i)} = \frac{x-1}{\sum_{j \in V, j \neq i} d(i,j)}$$

where $d(i,j)$ is the shortest path distance between nodes i and j [49].

When applied to PPIs, high closeness centrality implies that a protein can quickly interact, either directly or indirectly, with many other proteins. This is biologically meaningful, as such proteins often play central roles in coordinating cellular functions, including signal transduction, transcriptional regulation, and stress response [50]. Proteins with high closeness values may act as efficient communicators or integrators of information across functional modules. Because they are topologically proximal to many nodes, they can facilitate rapid response and influence multiple biological pathways simultaneously [51]. In disease settings, shifts in closeness centrality can indicate loss of global network connectivity or impaired information flow due to mutations or dysregulation [52].

Not all nodes in a network are important simply because they have many connections; some are crucial because they connect to other important nodes. Eigenvector centrality captures this principle by assigning higher scores to nodes that are linked to other highly connected and central nodes, unlike degree centrality, which treats all neighbors equally. The principle behind eigenvector centrality is simple: a node's score depends on the scores of its neighbors. Formally, it is defined as:

$$EC_{(i)} = \frac{1}{\lambda}\sum_{j\in V} A_{ij}\, EC_{(j)}$$

where A is the adjacency matrix of the graph, and λ is the largest eigenvalue [53]. This formulation makes eigenvector centrality well suited for identifying core nodes embedded in cohesive subnetworks, rather than just local hubs.

In protein-protein interaction networks, centrality measures such as eigenvector centrality highlight proteins that are not only well connected but strategically positioned within influential neighborhoods. These proteins often participate in regulatory assemblies, multi-subunit complexes, or pathways with broad cellular influence. For example, transcription factors, chaperones, or kinases commonly show high eigenvector centrality because of their regulatory roles, despite not having the most direct connections. From a systems biology perspective, eigenvector centrality reveals whether a protein's importance is only local (highly connected but limited broad influence) or global (strategically placed within the core of the interactome). This makes it particularly useful for identifying candidate disease genes, potential drug targets, or evolutionarily conserved functional modules [54–56]. Collectively, these centrality measures provide a quantitative basis for identifying proteins that influence network structure, mediate signal integration, or act as critical intermediates in cellular processes.

3.3 Network embedding in Hyperbolic space with PS model

Recent advances in network science have proposed that complex systems, including biological networks, may be shaped not only by topological relationships but also by an underlying geometric structure. Traditional topological measures such as degree centrality, clustering coefficient, or path length provide useful attributes, but they often fail to capture the latent geometry that forms how these systems are organized and evolve. Many real-world networks, including protein-protein interaction networks (PPINs), are characterized by scale-free degree distributions, meaning that most proteins interact with only a few partners while a minority functions as hubs with notably many connections. This distribution follows an approximate power-law and is common across cellular systems [57,58]. Hubs in PPINs are often essential genes, multifunctional regulators and are typically more evolutionarily conserved than non-hub proteins, highlighting the biological significance of network heterogeneity [59]. In addition to scale-free organization, PPINs show high clustering and modularity: proteins participating in the same

complex or pathway tend to interact with one another, resulting in densely connected local clusters that are themselves structured hierarchically [60].

Together, heterogeneous connectivity, strong clustering, and modular hierarchy imply that PPINs are organized according to hidden geometric principles rather than by random structure. Hyperbolic geometry has emerged as a natural framework to represent such networks. Unlike Euclidean space, hyperbolic space has constant negative curvature, which allows the amount of space to expand exponentially with distance from the center. This property makes it particularly well suited for embedding hierarchical, tree-like structures while preserving structural relationships as well as global hierarchy and local clustering, providing a geometrically interpretable latent space [61].

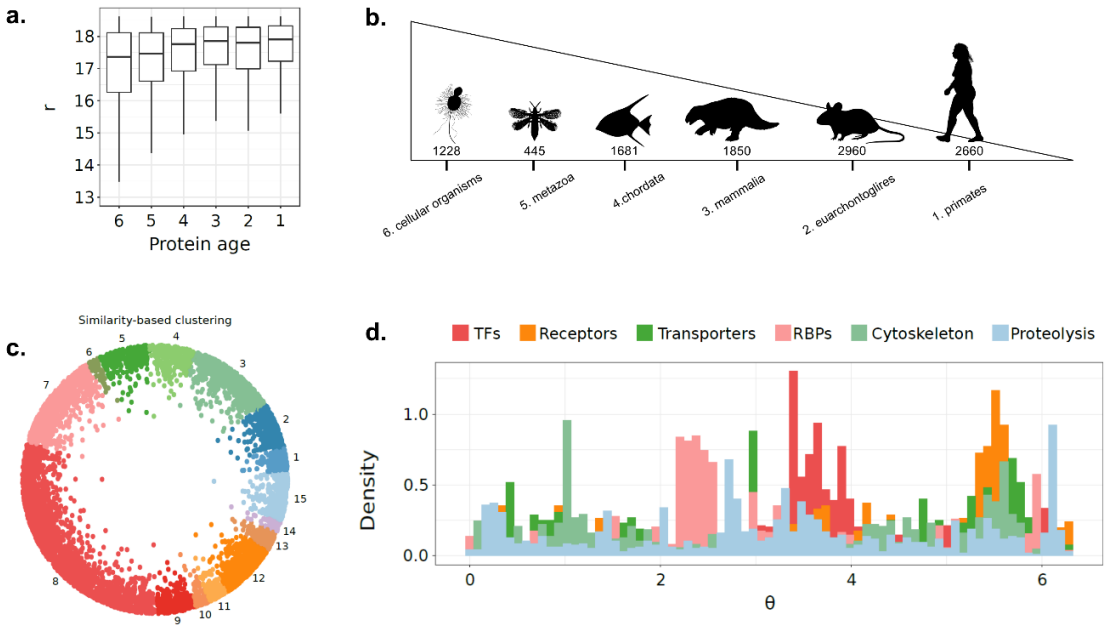


Figure 3: The human protein interaction network embedded in hyperbolic space and interpretation of hyperbolic coordinates. ***(a)*** *Distribution of inferred radial coordinates (r) for proteins across six evolutionary age groups. Proteins from older groups (e.g., cellular organisms) tend to occupy smaller radial values, reflecting higher popularity (hub status), while younger proteins (e.g., primates) are positioned at larger radial values.* ***(b)*** *Representation of protein age groups and the number of proteins in each group.* ***(c)*** *Angular layout of proteins in hyperbolic space, where proximity in angle (θ) reflects functional similarity.* ***(d)*** *Density of angular coordinates for different molecular classes (TFs: transcription factors; RBPs: RNA-binding proteins; transporters; receptors; cytoskeleton; proteolysis), showing that proteins with similar functions cluster within specific angular sectors. Adapted from [63].*

The Popularity-Similarity (PS) model provides a generative explanation for how complex networks can be embedded and evolve in hyperbolic space [62]. In this model, each node is assigned to polar coordinates (r_i, θ_i). The radial coordinate (r_i) encodes popularity or seniority: nodes that entered the system earlier or that accumulate many links are assigned to smaller radial values, reflecting their hub status. Conversely, nodes with fewer connections positioned at larger radial values. The angular coordinate (θ_i) represents similarity: nodes positioned at nearby angles share functional or structural characteristics and are more likely to connect (Figure 3).

The probability that two nodes interact depends on their hyperbolic distance, which integrates both radial and angular contributions [61,64]. For two nodes i and j with coordinates (r_i, θ_i) and (r_j, θ_j), the distance can be approximated as:

$$d_{ij} \approx r_i + r_j + 2\ln\left(\frac{\theta_{ij}}{2}\right),$$

where θ_{ij} is the angular separation between the nodes. This formula captures the trade-off between popularity and similarity: nodes are more likely to connect if they are either highly connected (small radial distance) or functionally related (small angular separation).

As a result, the PS model naturally reproduces properties of real networks, including scale-free degree distributions, high clustering, and modular hierarchical organization [61]. When applied to PPINs, proteins with small radial coordinates tend to be evolutionarily conserved or essential, while angular proximity corresponds to functional similarity, such as components of the same complex or pathway [63]. Thus, hyperbolic embeddings derived from the PS model not only replicate the observed structure of protein interaction networks but also offer a principled way to predict missing links and uncover functional associations.

Embedding a real biological network under the PS model involves inferring the set of radial and angular coordinates that maximize the likelihood of observing the network structure. Several computational strategies have been proposed for inferring these coordinates, including maximum-likelihood estimation and manifold learning approaches such as the LaBNE+HM algorithm [65]. The resulting embeddings provide compact, low-dimensional representations of large networks, allowing analysis and visualization. This model has been extensively validated across synthetic and real networks. Krioukov et al. demonstrated that hyperbolic embeddings capture essential statistical properties of scale-free networks, including degree distributions and clustering, with high accuracy [61]. Papadopoulos et al. [62,64] showed that PS and PSO (Popularity-Similarity-Optimization) models can generate networks that match the hierarchical modularity of real-world data. More recent work has extended hyperbolic embedding methods in several directions. The non-uniform PSO (nPSO) model integrates heterogeneous angular distributions to better reproduce community structure together with scale-free and hierarchical organization [66]. Hyperbolic graph neural network approaches applies embedding principles in deep learning, improving the representation of hierarchy and modularity in large-scale biological networks [67]. Lizotte et al. introduced a Bayesian approach to hyperbolic embedding that measures uncertainty in node positions, provides a probabilistic view of network geometry, and shows that multiple embeddings may equally explain the same network [68].

The advantages of hyperbolic embedding extend beyond structural description, as the hyperbolic features encode hierarchical depth and functional similarity and can therefore be directly applied in predictive studies. Within the PS model, hyperbolic distance provides a numerical metric of proximity in the latent space, supporting applications such as link prediction, module detection,

and functional annotation. In PPINs, proteins that are hyperbolically close but not experimentally confirmed to interact are strong candidates for missing interactions and may share functional relationships, such as participation in the same pathway or complex, thereby guiding the design of experiments to validate predicted interactions. Several studies have demonstrated the applicability of hyperbolic distance to biological networks. Alanis-Lobato et al. [63] showed that latent geometry in the human interactome can be used to recover missing protein-protein interactions and reconstruct signaling pathways; Tang et al. [69] introduced HI-PPI, a model which leverages hierarchical embedding in hyperbolic space to improve PPI prediction; Vagiona et al. [70] implemented hyperbolic distances to classify protein triplets into cooperative or competitive interactions; and Pogány et al. [71] demonstrated that embedding biological hierarchies in hyperbolic space improves the representation of functional organization in interaction networks.

Despite their advantages, hyperbolic embeddings show several limitations. The estimation of polar coordinates is computationally demanding for large interactomes, and outcomes can vary depending on the optimization algorithm or the modeling assumptions regarding network evolution. In addition, noise in protein-protein interaction data, including false positives and sampling biases, may reduce the stability of embeddings. Nevertheless, the Popularity-Similarity model provides informative coordinates and captures fundamental properties of complex networks, establishing hyperbolic embedding as a reliable and biologically meaningful tool for the analysis of PPINs.

3.4 Functional and Structural Context of Protein Interactions

While network-based models and hyperbolic embeddings provide powerful approaches to capture latent organizational principles of protein-protein interaction networks (PPINs), a comprehensive understanding of protein interactions requires integrating these abstract representations with functional and structural annotations. Proteins interact not only because they are connected in a graph, but because they perform molecular functions and assemble into structurally stable complexes. Classical PPINs, however, typically encode interactions as undirected, binary edges, and therefore fail to capture critical aspects such as directionality, cooperativity, competition, and structural compatibility. Overcoming these limitations is essential to gain a more holistic overview and a deeper understanding of how proteins interact within cellular environments.

At the functional level, protein interactions support a wide range of cellular processes. Some interactions are structural, helping to form stable assemblies such as the ribosome, proteasome, or chromatin remodeling complexes [72]. Others are regulatory or catalytic, for example when enzymes bind to substrates or when inhibitors modulate the activity of their targets [73]. PPIs also play central roles in signal transduction, coordinating proteins within signaling cascades and in gene expression, where transcription factors interact with co-regulators or components of the transcriptional machinery [27]. These interactions can be either stable, forming long-lived

complexes, or transient, depending on the cellular or signaling state. Some are conditional, occurring only in specific compartments, time points, or environmental conditions [6].

Many protein interactions are influenced or enabled by post-translational modifications (PTMs). PTMs are covalent or enzymatic alterations of proteins that occur after translation. They are commonly grouped into categories such as (i) addition of functional or chemical groups (acetylation, methylation, phosphorylation), (ii) conjugation of polypeptide chains (ubiquitination, SUMOylation), (iii) attachment of complex molecules (palmitoylation, glycosylation), and (iv) amino acid modifications (proteolytic cleavage) [74]. These chemical modifications can regulate protein activity, localization, stability, and binding capacity [75]. Among them, phosphorylation and dephosphorylation are the most studied and widespread, carried out by kinases and phosphatases, respectively. These processes introduce directionality into protein interactions by distinguishing regulators from targets, since a kinase modifies a substrate without being modified in return [76].

Large-scale datasets of PTMs are systematically curated within specialized databases. For phosphorylation, databases such as PhosphoSitePlus [77] and Phospho.ELM [78] provide experimentally validated modification sites with functional annotations. Expanded resources integrate multiple types of modifications, offering both curated data and associations with enzymes and pathways [79,80]. Beyond experimental annotation, numerous predictive algorithms have been developed to infer PTM sites based on sequence motifs, structural context, or evolutionary conservation. Computational predictors extend PTM coverage by suggesting putative modification sites that have not yet been experimentally validated [81,82].

Although PTMs can regulate protein interactions, they can only occur if the proteins are structurally compatible. Whether two proteins can bind depends on the three-dimensional arrangement of their binding domains, interface accessibility, and conformational flexibility. Some interactions occur through well-defined interfaces, while others involve intrinsically disordered regions or short linear motifs that adopt structure only upon binding [83,84]. High-resolution experimental methods such as X-ray crystallography and cryo-EM have provided detailed conformational information for thousands of protein complexes, which are systematically stored in the Protein Data Bank (PDB) [85]. To extend such information beyond experimentally resolved cases, databases like Interactome3D annotate PPIs with structural details by mapping interactions to PDB entries and, when necessary, generating homology-based models to increase coverage [86]. More recently, advances in deep learning, particularly AlphaFold-Multimer, have allowed large-scale prediction of protein complexes structures, expanding structural coverage and supporting experimental approaches [87,88].

Structural context not only determines whether proteins can interact, but also how interactions are organized within multi-protein systems. A single protein may bind with multiple partners at distinct sites, enabling simultaneous cooperative binding, or at overlapping interfaces, leading to competition. This distinction has important functional consequences but is generally absent from

standard network representations, which list interactions as independent binary pairs without specifying structural compatibility [89]. Traditional interactome maps rarely include residue-level details such as which amino acids mediate binding or how conformational changes influence stability. This thesis addresses these gaps by predicting directional (de-)phosphorylation-related interactions from functional annotations and distinguishing cooperative from competitive triplets using structural features.

3.5 Applications to Disease Biology

Biological networks are often perturbed in disease, leading to altered interactions and disrupted signaling pathways. This is observed across diverse conditions, including neurodegenerative diseases, cardiovascular disorders, and cancer, where key proteins loose or gain interactions that contribute to pathological states [23,44]. One way to study these changes is through the concept of disease modules, groups of proteins that are interconnected in the interactome and collectively contribute to disease phenotypes [37,90]. These modules may consist of both known disease-associated genes and their interacting partners, forming functionally sub-networks. Identification of these modules requires network-level analysis that accounts for the connectivity and organization of proteins, rather than focusing on individual genes. [91].

Disease modules have been identified across a range of conditions (Figure 4). In cancer, abnormal PPIs play a central role in tumor initiation and progression by disrupting normal cellular control mechanisms. These changes often involve perturbed protein expression, post-translational modifications, or mutations that affect binding specificity and interaction dynamics. As a result, key signaling modules become dysregulated, activating mitogenic pathways, suppressing immune recognition, and promoting resistance to apoptosis and targeted therapies [92]. In cardiovascular disease, network-level disruptions in protein interactions disrupt endothelial function, mitochondrial regulation, and stress response pathways, contributing to vascular dysfunction and myocardial remodeling [93,94]. These molecular alterations are linked to dysregulation of lipid and metabolite homeostasis, processes that increase vascular inflammation, promote oxidative stress, and disrupt myocardial energy metabolism. Together, these perturbations highlight that cardiovascular pathology emerges from multi-omics level dysregulation, including proteins, lipids, and metabolites, rather than from single gene or pathway defects [95]. Neurodegenerative diseases, such as Alzheimer's, Parkinson's, Huntington's disease (HD), and Spinocerebellar Ataxias (SCAs), are marked by progressive loss of neuronal function and also involve dysfunction in core protein interaction modules [96]. These include pathways responsible for synaptic transmission, protein homeostasis, transcriptional regulation, and RNA metabolism [97,98]. Mutations in disease genes often lead to misfolded or aggregated proteins that interfere with both direct partners and the structure of broader functional assemblies.

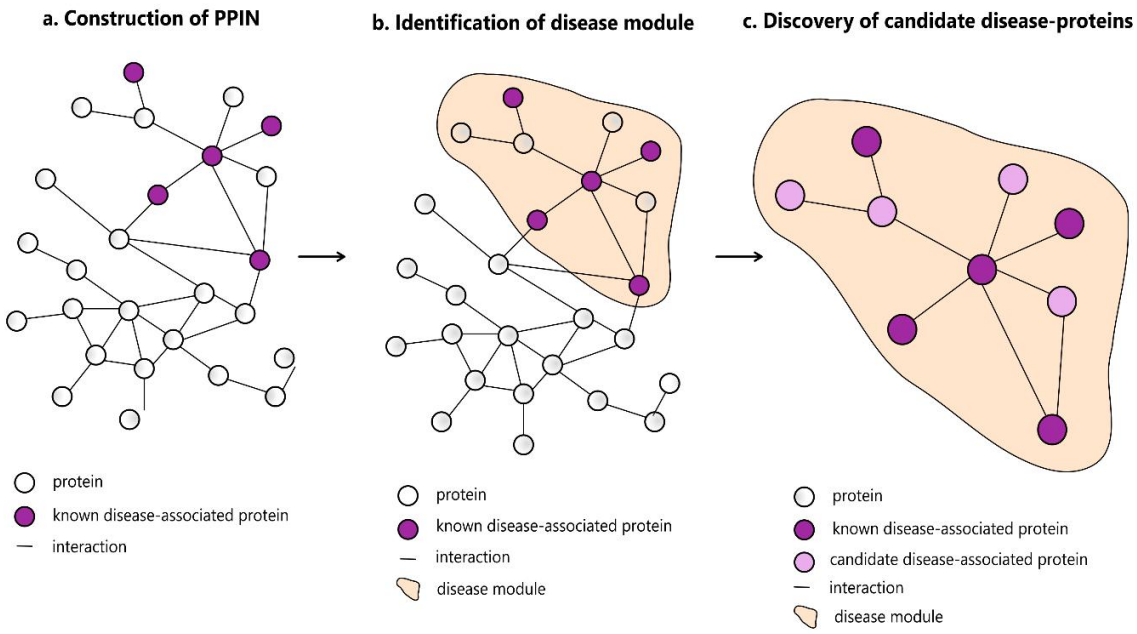


Figure 4. Schematic representation of disease module identification in protein-protein interaction networks. (a) *Construction of the protein-protein interaction network (PPIN), where proteins are represented as nodes and edges indicate interactions. Known disease-associated proteins are highlighted in purple.* ***(b)*** *Disease-associated proteins cluster together, forming a disease module (shaded area) within the network.* ***(c)*** *Expansion of the module allows the identification of candidate disease-associated proteins (light purple). Panels (a-c) illustrate different stages of analysis within the same underlying network.*

Among neurodegenerative disorders, Huntington's disease (HD) is an autosomal dominant disease caused by a CAG trinucleotide repeat expansion in the HTT gene, which leads to an abnormally long polyglutamine tract in the huntingtin protein (mHTT). The mutant protein is prone to misfolding and aggregation and disrupts multiple cellular processes including transcriptional regulation, axonal transport, and DNA repair. At the network level, mHTT perturbs protein interaction modules by binding abnormally to transcriptional regulators such as CREB-binding protein (CBP), TAFII130, and REST, leading to global transcriptional dysregulation in neurons [99,100]. These effects extend beyond direct binding partners, affecting entire nuclear complexes and gene regulatory networks. Furthermore, mHTT impairs the assembly of functional complexes involved in mitochondrial homeostasis, proteostasis, and synaptic function, contributing to neuronal dysfunction and degeneration, particularly in the striatum and cortex [101]. Huntington's disease is not only characterized by the toxic properties of mutant huntingtin but also by the complex influence of its interactors. Paralogous proteins are of particular interest, as they can share structural similarity while displaying different functional effects. Network-based analyses of huntingtin interactors have revealed cases where one paralog increases polyglutamine toxicity, while the other member of the pair plays a protective role. For example, paralog pairs such as MID1/PML, IKBKB/IKKA, and IKBKG/OPTN were shown to have opposing influences on neuronal survival, suggesting that the balance of paralog activity can critically shape disease progression [102].

A similar principle applies to Spinocerebellar Ataxia Type 1 (SCA1), another polyglutamine expansion disorder in which mutant ataxin-1 perturbs nuclear protein interaction networks. SCA1 is caused by a CAG repeat expansion in the ATXN1 gene, resulting in an elongated polyglutamine tract in the ataxin-1 protein. The mutant protein misfolds and forms nuclear aggregates, particularly in cerebellar Purkinje cells. At the molecular level, mutant ataxin-1 disrupts protein-protein interaction networks by abnormally binding to RNA processing and transcription complexes, including splicing factors such as RBM17 and U2AF65, and transcriptional regulators like CIC [103,104]. Petrakis and colleagues identified 21 human proteins that modulate ataxin-1 aggregation or toxicity, many of which contain coiled-coil domains that promote aggregation and enhance proteotoxicity [105]. These altered interactions destabilize larger nuclear assemblies involved in RNA metabolism, ribosome biogenesis, and stress responses. Moreover, phosphorylation at serine 776 stabilizes mutant ATXN1, promotes its pathogenic interactions, and drives disease progression, highlighting this post-translational modification as a critical determinant of SCA1 pathogenesis [106]. Together, these findings underscore how mutant ataxin-1 perturbs functional modules in the nuclear proteome, contributing to progressive cerebellar dysfunction in SCA1.

Recent work has demonstrated that embedding protein interaction networks into hyperbolic space provides a powerful framework for studying disease modules. The geometric representation captures both hierarchy and functional similarity, enabling the identification of spatially disease-associated regions in the interactome [107]. This approach has been successfully applied to prioritize therapeutic targets in infectious disease [108]and to investigate the organization of genetic modifiers in neurodegenerative disorders such as Huntington's disease [102]. Collectively, these studies establish hyperbolic embeddings as a valuable methodology for analyzing the organization of disease modules within protein interaction networks.

3.6 Multi-Omics Integration

A system-level understanding of cellular biology requires integration across multiple molecular layers. While PPIs capture the core of cellular organization, they represent only one dimension of molecular complexity. Genomics, transcriptomics, proteomics, metabolomics, and lipidomics offer complementary perspectives that together yield a more comprehensive view of functional regulation. Genomics characterizes the cell's genetic material, describing DNA sequence variation, structural rearrangements, and epigenetic modifications that influence transcriptional regulation [109]. Genomic studies can detect inherited variants or somatic mutations of DNA that contribute to the development of genetic diseases and affect disease risk [110]. Transcriptomics reflects the dynamic expression of genes, measured as RNA abundance across conditions, cell types, or tissues. It provides insight into gene regulation, alternative splicing, and differential gene expression across biological contexts that cannot be inferred from genomic data alone [111]. Proteomics describes protein abundance, post-translational modifications, and interactions, metabolomics reflects the

dynamic biochemical state of the cell through small-molecule intermediates and metabolic pathway dynamics, and lipidomics characterizes lipid molecules that are critical for membrane architecture, energy storage, and signaling [2,112,113]. Integration of these layers allows the detection of coordinated molecular changes that remain hidden when individual layers are analyzed in isolation [114,115].

Recent experimental advances have made multi-omics integration increasingly accessible. For example, mass spectrometry-based protocols now allow proteins, metabolites, and lipids to be extracted and quantified together from the same biological sample, reducing technical variation and enabling multi-layered profiling from limited material [116]. In addition, mass spectrometry imaging (MSI) provides spatially resolved maps of proteins, metabolites, and lipids directly within tissue samples. Unlike bulk profiling, MSI retains spatial information, enabling the identification of molecular distributions within defined regions of a tissue sample [117] . This spatial resolution is particularly important because protein abundance, metabolic activity, and lipid composition vary strongly between organs and cell types. Adding tissue-level information improves multi-omics analysis, allowing the detection of modules that are specific to certain biological contexts or tissue areas [118].

The development of computational tools for multi-omics integration has further expanded the scope of these analyses. Statistical and machine learning approaches are widely used to collectively analyze proteomic, metabolomic, and lipidomic datasets, uncovering latent correlations and functional modules across molecular layers [119,120]. Beyond these approaches, graph theory provides a natural way to integrate heterogeneous molecular layers. Different entities, such as genes, transcripts, proteins, metabolites, and lipids, can be represented as nodes, while edges capture their associations, whether derived from experimental evidence, correlation, or established biological knowledge. In this way, graph-based models allow the integration of diverse omics layers into a single network representation, enabling the detection of cross-layer interactions and coordinated regulation. Network analysis can then be used to identify hubs, tissue-specific modules, or disease-associated subnetworks that emerge from the combined molecular landscape [121,122]. Embedding such multi-layered networks into hyperbolic space preserves both functional similarity and hierarchical organization, facilitating the discovery of context-specific modules and improving predictions of functional associations (Figure 5) [63].

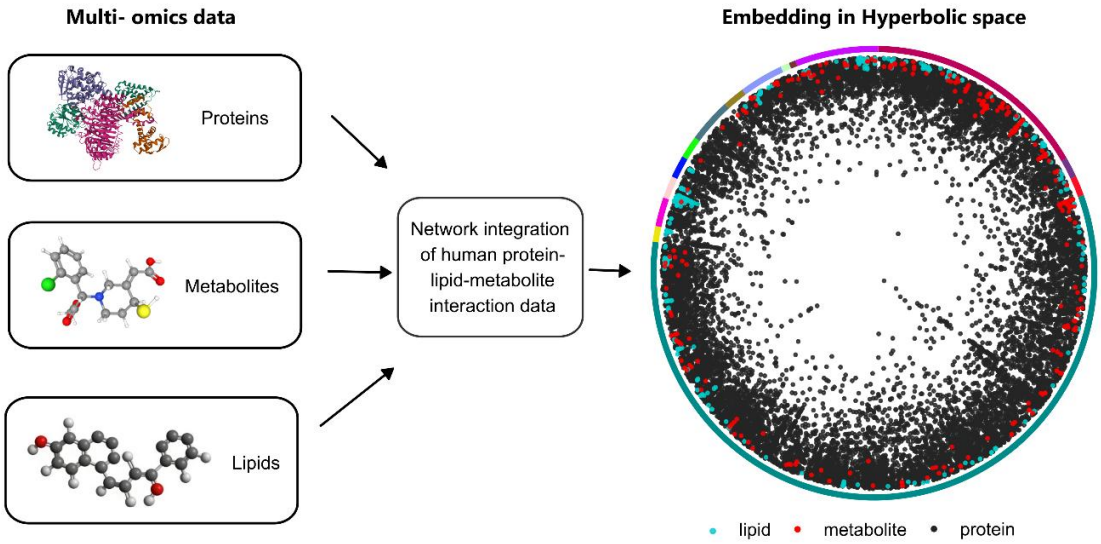


Figure 5. Integration of multi-omics data into a hyperbolic network representation. *Proteins, metabolites, and lipids were integrated into a single network of human interactions. The resulting network was embedded in hyperbolic space, where different molecular entities (proteins, metabolites, lipids) are positioned according to latent geometric relationships that capture hierarchical and functional organization.*

Multi-omics integration has been successfully applied in diverse disease contexts, including cancer, cardiovascular disease, and neurodegeneration, where perturbations span multiple regulatory layers [123,124]. For example, integrated proteomic and metabolomic profiling has uncovered metabolic reprogramming in tumors, while lipidomics combined with proteomics has revealed mechanisms of inflammatory signaling in cardiovascular disorders [125,126]. These studies highlight how multi-omics integration improves mechanistic understanding by linking molecular changes to higher-order network organization.

4. Scientific questions

Despite extensive progress in mapping and analyzing protein-protein interactions, key questions remain about their functional specificity, structural compatibility, and biological context within the human proteome. The central aim of this thesis is to address these challenges using network-based approaches and structural analysis. In addition, collaborative work on multi-omics integration highlights the potential of combining network models with omics data to study disease mechanisms. A common principle across all chapters is the use of hyperbolic network embedding. Hyperbolic geometry offers a latent space uniquely suited to capture the hierarchical and structured organization of biological systems, more effective than Euclidean or purely topological approaches in representing network structure. Within this space, radial coordinates encode popularity or seniority while angular coordinates capture functional similarity, establishing a framework to explore diverse aspects of protein biology.

In the first chapter of this thesis, we apply hyperbolic embedding to the human protein interaction network (hPIN) and investigate its ability to reveal latent structure in the context of Huntington's disease (HD). We focus on interactors of huntingtin (HTT) and identify pairs of paralogs positioned in distinct regions of the hPIN. While these paralogs are expected to interact similarly with HTT, their differing network locations point to divergent roles in pathways and complexes. By evaluating protein pairs with opposing effects on HD, we uncover common partners of positive and negative effectors, which we propose as candidate modulators with potential impact in HD models.

In the following section (Chapter 2), we extend the approach to directed protein-protein interactions, focusing on phosphorylation and dephosphorylation. By integrating hyperbolic coordinates with classical centrality measures, we trained a random forest model to predict directed (de-)phosphorylation relationships. To evaluate the model, we applied it to proteomic data from Spinocerebellar Ataxia type 1 (SCA1), a neurodegenerative disease in which abnormal phosphorylation of ataxin-1 has been implicated in promoting toxic protein aggregation. This analysis identified phosphorylation interactions associated with disease and demonstrated the value of hyperbolic embedding for studying post-translationally regulated networks.

In Chapter 3, we move beyond binary protein interactions to higher-order motifs, focusing on triplets of proteins that share a common partner. Starting from a hyperbolically embedded human protein interaction network (hPIN), we identified open triangles and annotated them with structural data from experimentally validated complexes. Using this dataset, we trained a random forest classifier to distinguish cooperative from competitive triplets based on topological, geometric, and biological features. We further show that hyperbolic angular separation reflects binding compatibility, and structural modeling confirms whether partners occupy distinct or overlapping binding regions. This analysis highlights how higher-order network motifs provide insight into the organization and function of molecular complexes.

In Chapter 4, we expand the scope to multi-omics integration by embedding a combined protein-lipid-metabolite network into hyperbolic space. This network serves as the foundation for a software tool that enables users to input molecules from one omics layer and retrieve ranked associations in the other two. Incorporating biochemical knowledge and tissue-specific data, the framework captures interactions between molecular layers and identifies disease-relevant modules. We highlight its relevance through applications to cardiovascular disease, where lipid and metabolic dysregulation play a central role, and to functional enrichment analysis of lipidomic profiles.

Together, these chapters highlight how hyperbolic embedding serves as the common key that links topological analysis, functional prediction, structural modeling, and multi-omics integration. By applying this geometric framework, the thesis develops computational approaches that move beyond static protein-protein interactions, toward biologically interpretable and disease-relevant models of the human interactome.

Chapter 1

Analysis of Huntington's Disease Modifiers Using the Hyperbolic Mapping of the Protein Interaction Network

Aimilia-Christina Vagiona, Pablo Mier, Spyros Petrakis and Miguel A. Andrade-Navarro

Article published in *International Journal of Molecular Sciences, vol. 23, 10, 5853. (19 May 2022), doi: https://doi.org/10.3390/ijms23105853*

International Journal of
Molecular Sciences

MDPI

Article

Analysis of Huntington's Disease Modifiers Using the Hyperbolic Mapping of the Protein Interaction Network

Aimilia-Christina Vagiona [1], Pablo Mier [1], Spyros Petrakis [2] and Miguel A. Andrade-Navarro [1,*]

[1] Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University, Hans-Dieter-Hüsch-Weg 15, 55128 Mainz, Germany; avagiona@uni-mainz.de (A.-C.V.); munoz@uni-mainz.de (P.M.)
[2] Institute of Applied Biosciences/Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece; spetrak@certh.gr
* Correspondence: andrade@uni-mainz.de

Abstract: Huntington's disease (HD) is caused by the production of a mutant huntingtin (HTT) with an abnormally long poly-glutamine (polyQ) tract, forming aggregates and inclusions in neurons. Previous work by us and others has shown that an increase or decrease in polyQ-triggered aggregates can be passive simply due to the interaction of proteins with the aggregates. To search for proteins with active (functional) effects, which might be more effective in finding therapies and mechanisms of HD, we selected among the proteins that interact with HTT a total of 49 pairs of proteins that, while being paralogous to each other (and thus expected to have similar passive interaction with HTT), are located in different regions of the protein interaction network (suggesting participation in different pathways or complexes). Three of these 49 pairs contained members with opposite effects on HD, according to the literature. The negative members of the three pairs, MID1, IKBKG, and IKBKB, interact with PPP2CA and TUBB, which are known negative factors in HD, as well as with HSP90AA1 and RPS3. The positive members of the three pairs interact with HSPA9. Our results provide potential HD modifiers of functional relevance and reveal the dynamic aspect of paralog evolution within the interaction network.

Keywords: Huntington's disease; paralogy; protein–protein interaction

Citation: Vagiona, A.-C.; Mier, P.; Petrakis, S.; Andrade-Navarro, M.A. Analysis of Huntington's Disease Modifiers Using the Hyperbolic Mapping of the Protein Interaction Network. *Int. J. Mol. Sci.* **2022**, *23*, 5853. https://doi.org/10.3390/ijms23105853

Academic Editors: Luis M. Valor and Antonio Campos-Caro

Received: 28 April 2022
Accepted: 19 May 2022
Published: 23 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Huntington's disease (HD) is one of nine autosomal dominant neurodegenerative disorders caused by the expansion of a CAG trinucleotide repeat. For HD, this expansion is located in the first exon of the huntingtin gene (htt) and results in an abnormally long poly-glutamine (polyQ) tract within the N-terminus of the huntingtin protein [1]. Expansion of CAG repeats results in the production of mutant proteins, which aggregate and form inclusions within neurons [2]. PolyQs with lengths above 40 amino acids cause mutant HTT proteins to misfold, form aggregates, become toxic, and cause disease [3].

The mechanism of polyQ-mediated toxicity is still under study; however, there is evidence supporting aberrant protein–protein interactions in the pathogenesis of HD [4–6]. Several lines of evidence support that expanded HTT is processed into N-terminal fragments that form inclusions in the cytoplasm and nucleus [7,8]. Many proteins, such as ubiquitin, heat shock proteins, and transcription factors, localize to polyQ inclusions [9–11].

Reports are accumulating on a variety of positive or negative effects that the expression or inhibition of multiple proteins have on HD's progression or its effects [12–18]. While these positive or negative effects may be actively due to specific functions (e.g., the phosphorylation of particular residues in HTT [19,20]), previous work suggests that the interactions of proteins with polyQ-caused aggregates can passively trigger both the increase and decrease of aggregates [21,22].

We hypothesized that, given the large size of HTT and its large number of interactors [4], it should be possible to explore the relatively complex network of the interactions surrounding HTT. We previously exploited this possibility to demonstrate the existence of multiple partners of HTT that use similar modes of interaction [23]. Here, we reasoned that paralogous expansions in the set of proteins interacting with HTT with divergent effects could be used to pinpoint active functions with relevance in HD. It is well known that gene duplication and speciation events, followed by mutation, can lead to functional changes, meaning that proteins with high sequence similarity may not have the same function [24,25]. In particular, several lines of work report the opposite effect of pairs of paralogs, revealing a functional diversity [26]. Favaro et al. found that two very similar proteins, PSD-93, and its paralog PSD-95, although they share similar functional domains and have evolved through the duplication of a single ancestral gene, have opposite roles in glutamatergic synapse maturation [27].

We assumed that the identification of pairs of paralogs interacting with HTT with opposite effects on HD might reveal active functions relevant to HD, under the assumption that these paralogs might interact identically with HTT, but their different effect on HD would arise from different interactions with other functional components of the protein interaction network. To maximize the divergence in protein interactions, we would need to account for the entire human protein interaction network (hPIN) in an unbiased approach. This is facilitated by techniques that project networks in a geometric space where closeness means a higher connection probability.

Since proteins are very complex machines, and their interactions with other proteins form a very complicated molecular system, their study as a protein–protein interaction network has gained traction in recent years [28]. Several algorithms and models support the existence of a hidden geometry underlying the structure and topology of complex systems, such as the human protein–protein interaction network [29]. The Popularity-Similarity (PS) model assumes that clustering and the hierarchy of complex networks arise from trade-offs between node popularity and similarity [30]. Additionally, Alanis-Lobato et al. found that the embedding of the hPIN to hyperbolic space has biological interpretations in terms of the PS model. They realized that the radial positioning of the nodes encapsulates information about protein conservation and evolution, while their angular positioning captures the functional and spatial organization of proteins in the cell [29]. This mapping may also lead to a better understanding of complex human disorders [31].

Motivated by these results, we followed a step-by-step computational filtering strategy, starting from a large protein–protein interaction (PPI) dataset embedded in the hyperbolic disc to obtain a network that consists of HTT interactors. This mapping enabled us to select pairs of paralogs of HTT interactors located in different regions of the hPIN. Proteins in each of these pairs are expected to interact similarly with HTT, but their different positions in the hPIN suggest their different involvement in pathways or complexes. The evaluation of protein pairs with opposed effects on HD was interpreted to find common partners for positive or negative effectors, which we propose as potential candidates for powerful effects in HD models.

2. Results

2.1. Human Protein Interaction Network Embedding to Hyperbolic Disc (hPIN)

In the first step of our analysis, we created a protein–protein interaction network from the HIPPIE database with high-quality interactions formed with a confidence score of $\geq$0.71 [32,33]. The largest connected component (LCC) of the hPIN is comprised of 93,140 interactions between 13,076 proteins. The resulting network was embedded into the two-dimensional hyperbolic plane H^2 using LaBNE+HM [34–36], and the hyperbolic coordinates were inferred for each protein of the network (Supplementary Table S1). We then proceeded to analyze the topological and geometrical properties of the hPIN.

2.2. Identification of Protein Clusters in the Angular Dimension

The similarity component of the PSM (angular coordinates of the nodes in H^2) abstracts the characteristics that make a node similar to others [29]; neighboring proteins play a role in similar biological processes [31]. To explore the biological meaning of the angular dimension, we identified big gaps between consecutive inferred angles and determined 12 protein clusters in the hPIN (Figure 1; Supplementary Figure S1; see Section 4 for details). From the biological point of view, the angles capture the functional organization of the cell, supported by the GO term annotations of the proteins in each cluster. As an example, the overrepresented biological process of cluster 1 is *protein lipidation*. Proteins agglomerate in similarity-based clusters since each of them is enriched in different aspects of the GO BP terms.

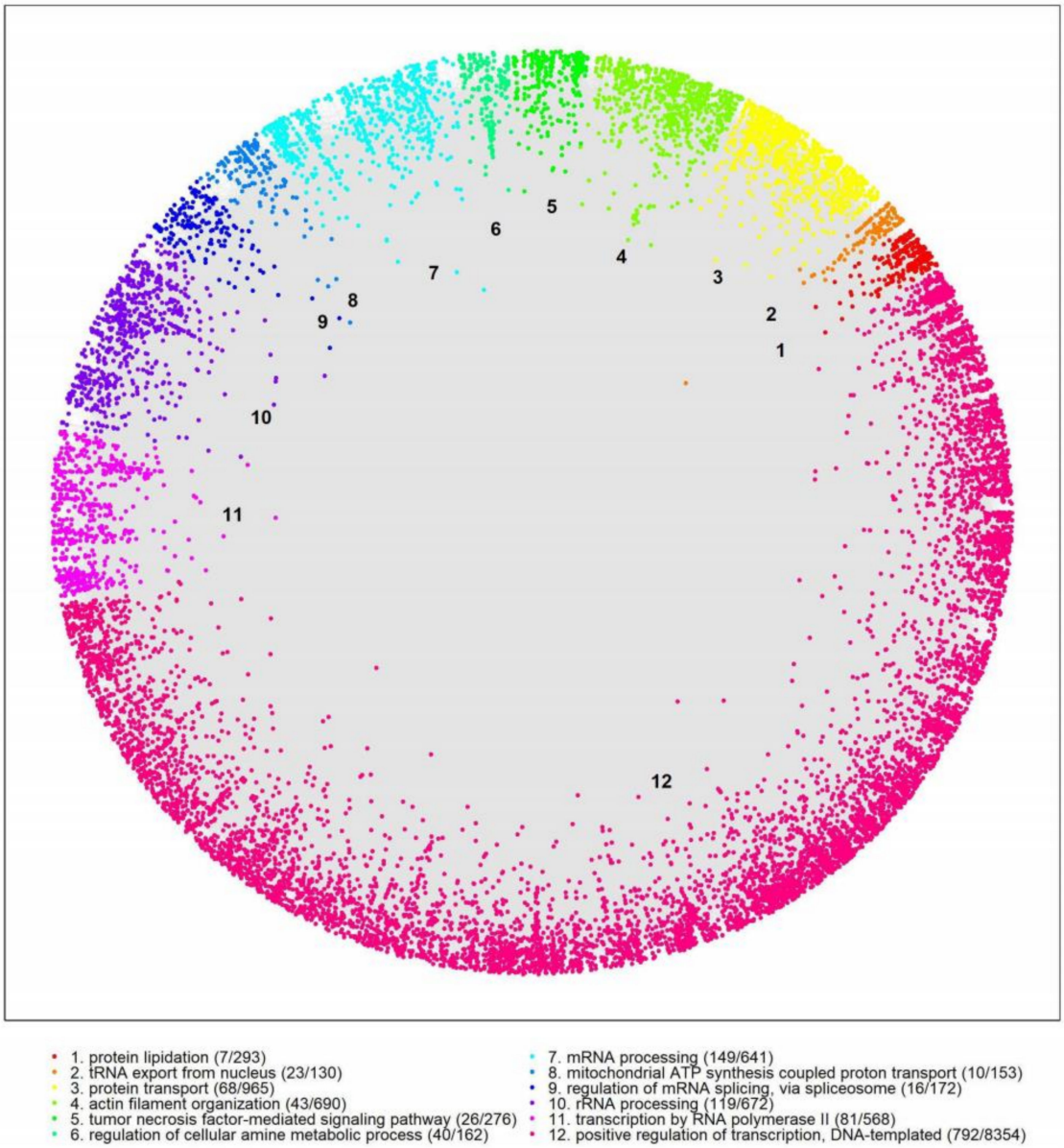


Figure 1. Human protein–protein interaction network embedded in the hyperbolic disc. Protein clusters in different colors were identified by big gaps separating groups of proteins in the angular dimension of the hyperbolic space. The overrepresented biological function in each cluster was determined via GO enrichment analysis (BP: Biological Process). Each cluster was assigned a numeric identifier (1–12). For each protein cluster, the number of proteins that are associated with the GO BP terms and the number of proteins in each cluster are shown.

2.3. HTT-Interactors in the Hyperbolic Disc

Starting from a large human protein interaction network with 13,076 proteins, we performed an interaction network filtering procedure in order to limit the dataset to a relatively small network, focusing on the HTT protein and its interactors (HttPIN). We observed, in this network, 382 proteins that are directly linked to HTT (Figure 2; Supplementary Table S2).

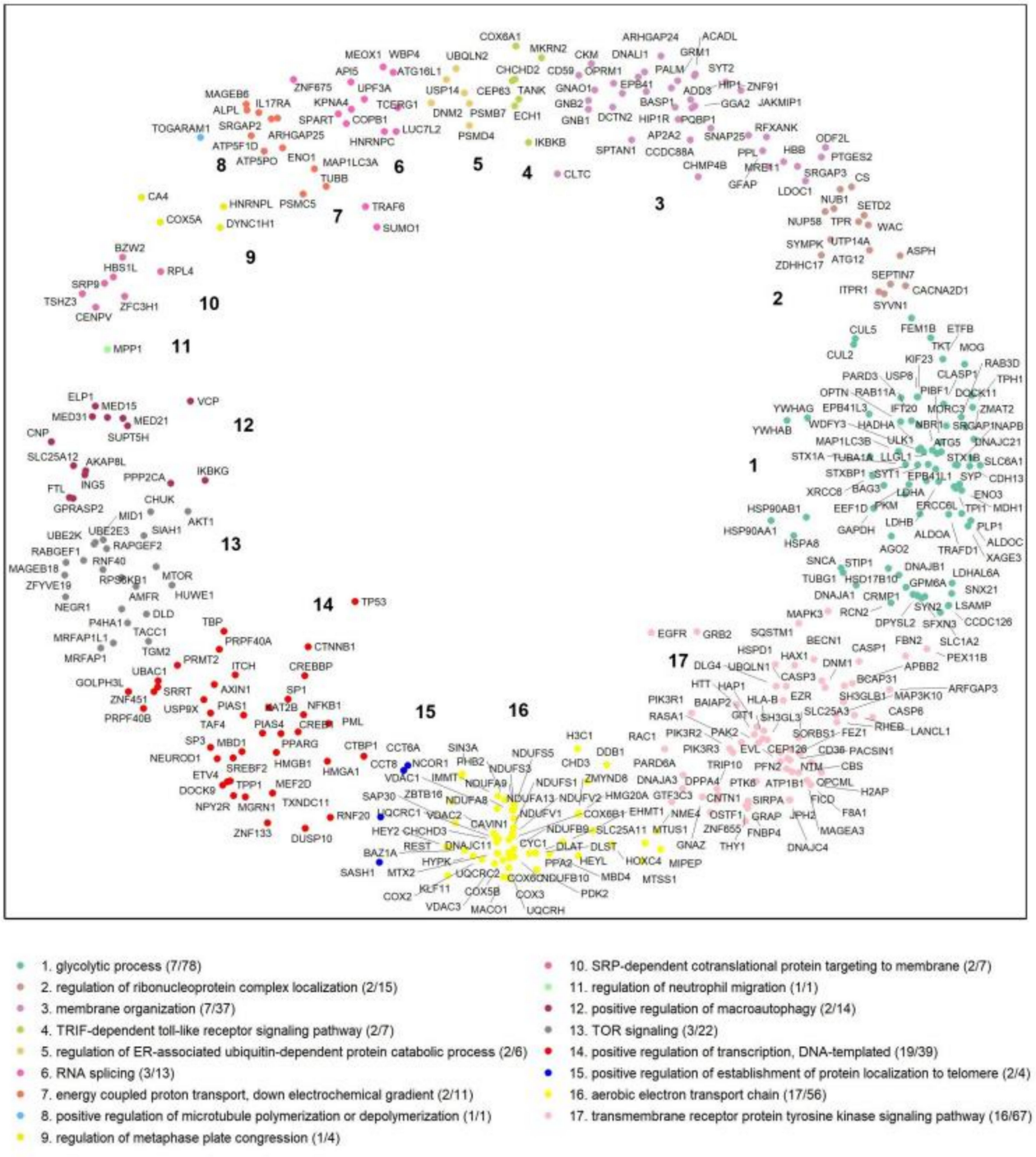


Figure 2. HTT interactors in the H^2. The different clusters were identified by big gaps separating groups of proteins in the angular dimension of the hyperbolic space. Each cluster was assigned a numeric identifier (1–17). The number of proteins/genes that are associated with the GO BP terms and the number of proteins/genes in each cluster are shown.

To reveal the functional modularity of the HTT interactors, following the procedure used above to cluster the hPIN, we partitioned the angular dimension of the nodes of the HttPIN into several sectors according to large gaps between the consecutive inferred angles of the HTT interactors (Supplementary Figure S2; see Section 4 for details). As it is shown in Figure 2, proteins agglomerate into 17 sectors. The overrepresented biological function in each cluster was determined through GO enrichment analysis and points to the heterogeneity of the clusters, since no common GO BP terms were observed between the sectors.

2.4. Paralog Pairs of HTT Interactors

To further investigate our hypothesis, we looked for HTT interactors with paralogs having opposite effects on HD. Previous computational analysis in yeast highlighted the value of studying protein–protein interaction networks to examine the functional divergence among duplicated gene products [37,38]. From all 382 HTT interactors, we obtained 87 paralogous pairs (Supplementary Table S3). Considering that the geometrical properties of the hyperbolic disc capture biologically relevant features, such as function, we speculated that paralogous proteins in different clusters could have divergent effects on the disease. Therefore, we selected paralog pairs of proteins detected in different clusters (Supplementary Table S4). Figure 3 shows this network, which consists of 74 nodes and 49 paralog pairs. Overall, 87 paralog pairs interact with HTT (Supplementary Table S3); 49 of them are located in different clusters, while 38 are in the same cluster.

2.5. Effects of Paralog Pairs on HD

We then reviewed the literature to find paralog pairs with opposite effects on HD (Supplementary Table S4). More specifically, DNAJC21 and its paralogs: DNAJC11, DNAJC4, DNAJA1, and DNAJA3 are all members of the DnaJ heat shock protein family (Hsp40). Previous studies in animal models have shown that Hsp40 chaperones are protective of neurodegeneration [39]. The overexpression of Hsp40 proteins can suppress polyQ aggregation, and, hence, they are critical for cell survival [40,41]. In addition, CCT8 and CCT6A, which are paralog proteins of HSPD1, have a protective role on HD. In fact, the upregulation of CCT8 has been linked to a mechanism that protects from polyQ aggregation, while the knockdown of CCT6A led to stimulating the aggregation of expanded polyglutamine and mutant huntingtin in cellular models [42,43]. Moreover, an increase in the UBQLN1 expression protects against HTT-polyQ-induced cell death and toxicity. Likewise, UBQLN2 significantly decreases in both wild-type and polyglutamine-expanded full-length HTT levels in cellular and animal models [12]. Finally, the TUBB protein interacts more strongly with mutant HTT than with the wild-type. This event blocks intracellular transport, suggesting a pathogenetic mechanism in HD [16].

We next focused on pairs of paralogs located in different regions of the hPIN and with opposite effects on HD according to the literature (Table 1).

Table 1. List of paralog pairs with opposite effects on HD.

Genes	Proteins	Effects on HD	References
MID1/PML	E3 ubiquitin-protein ligase Midline-1/Protein PML	Mutant HTT recruits MID1 protein complex resulting in overproduction of polyQ HTT; PML plays a protective role against neuronal toxicity associated with polyQ proteins.	[13,44–46]
IKBKB/IKKA(CHUK)	Inhibitor of nuclear factor kappa B kinase subunit beta/Inhibitor of nuclear factor kappa-B kinase subunit alpha	Inhibition of IKBKB may promote neuronal survival in HD; IKKA has a protective role in preventing HTT proteolysis.	[47]
IKBKG/OPTN	NF-kappa-B essential modulator/Optineurin	Inhibition of IKBKG activity reduces HTT-polyQ toxicity; OPTN has a protective effect on polyQ neurotoxicity associated with mutant HTT.	[48–50]

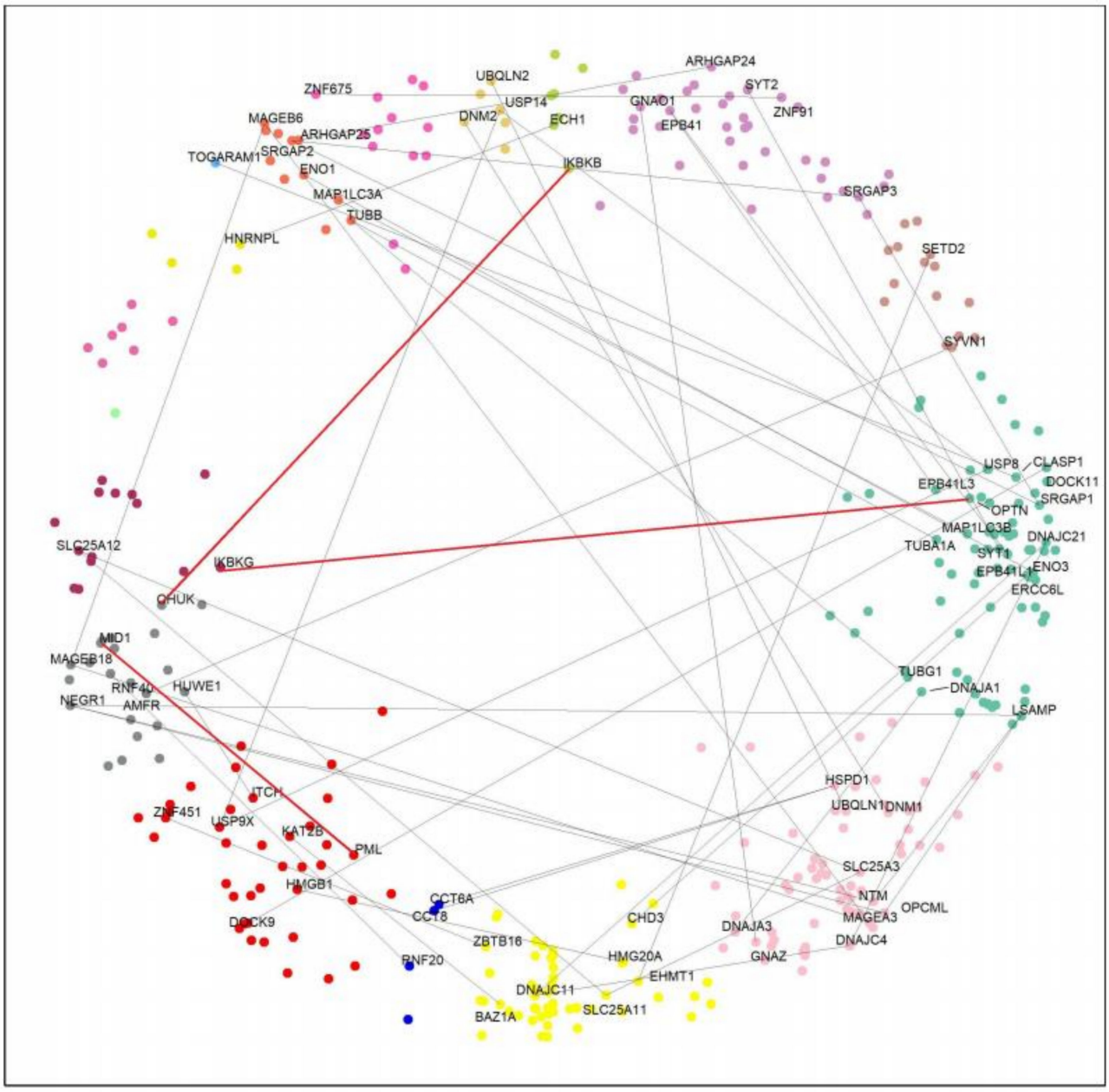


Figure 3. Paralog pairs of HTT interactors in different clusters. Nodes in different colors display protein agglomeration in angular similarity-based sectors. The nodes with the Gene Symbol represent the 49 paralog pairs located in different clusters. Paralog pairs are connected by an edge. Red edges indicate the three pairs of paralogs with opposite effects on HD (see text for details).

We detected three paralog pairs with opposite roles on HD. Notably, MID1/PML, IKBKB/IKKA(CHUK), and IKBKG/OPTN are paralogous pairs with experimental evidence suggesting their different effect on HD. MID1 leads to an aberrant overproduction of the mutant polyglutamine protein, inhibition of IKBKB has a protective effect on neurodegeneration, and IKBKG binds mutant HTT contributing to HD neurotoxicity [46,49,51]. The potential participation of IKBKB and IKBKG in the pathogenesis of HD was discussed in a computational analysis [52]. Differently, several lines of evidence support that PML, IKKA, and OPTN play a protective role against neuronal toxicity associated with HD [45,48–50].

2.6. Common Interactors between Positive and Negative Paralogs

Finally, we hypothesized that the existence of common interaction partners of the three negative paralogs could reveal functions that negatively influence HD and whose inhibition could have therapeutic effects. The three negative paralogs are IKBKB, IKBKG, and MID1. Using the HIPPIE database, we obtained their common interactors, besides HTT. Filtering interactions with a confidence score of $\geq$0.71 resulted in only PPP2CA. At a confidence score of $\geq$0.63, TUBB, HSP90AA1, and RPS3 were also found (see Figure 4).

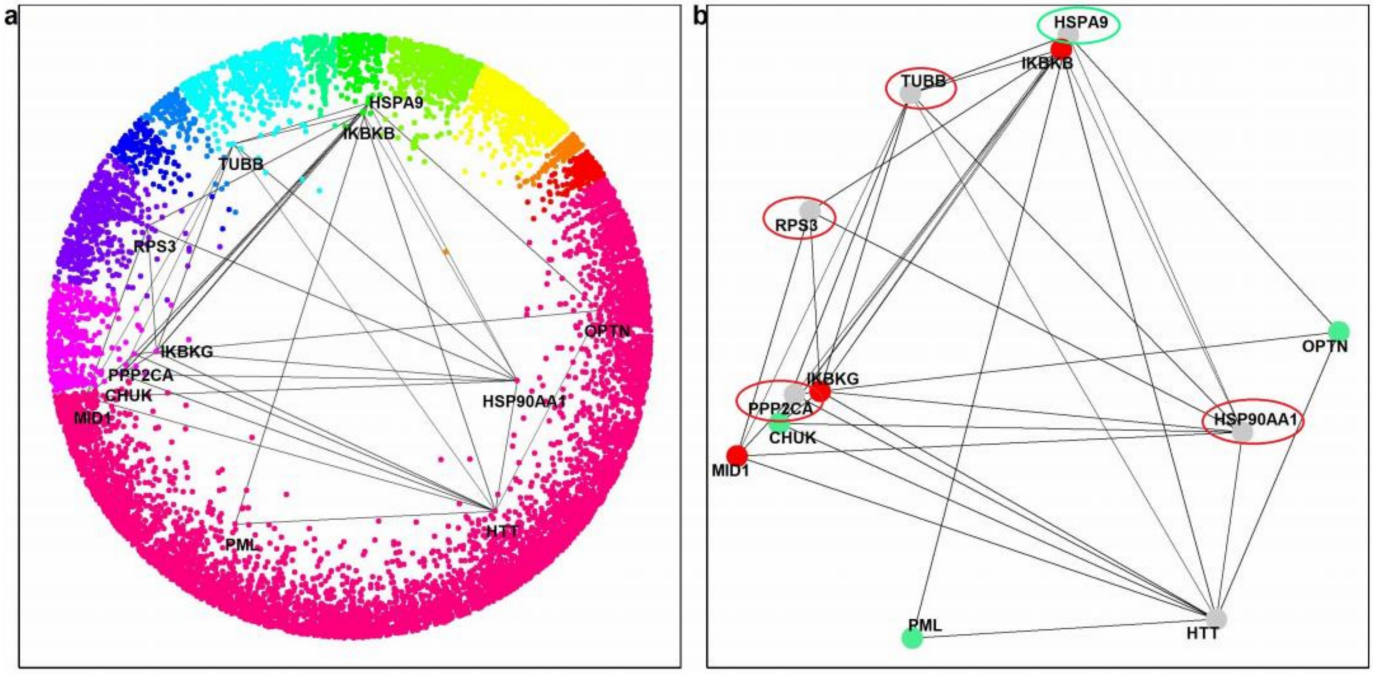


Figure 4. Positive and negative effectors on HD and closely connected components. (**a**) Position of the positive and negative paralogs and their common interacting partners in the hyperbolic disc. (**b**) Red nodes represent the negative paralogs, and the nodes surrounded by the red ellipses are their common interactors. Green nodes represent the positive paralogs, and the node surrounded by the green ellipse represents their common interactor.

We used the same approach with all three positive partners to find proteins that might be effective as a therapy for HD. The three positive paralogs are OPTN, PML, and CHUK. Using the HIPPIE database and a confidence score of $\geq$0.49, besides HTT, only one partner was found, chaperon HSPA9 (also known as mortalin), which is not a direct interactor of HTT (Figure 4).

Finally, we checked the connectivity of the common interacting partners of the paralogs with HTT. All five (including RPS3 and HSPA9, which do not bind directly to HTT) rank very high on the list of human proteins ordered by the number of interactions with HTT-interactors (among the top 4% of 10,914 ranked proteins; Table 2).

Table 2. Common interacting partners between negative and positive paralogs.

Paralogs	Effect on HD	Common Interactors between Paralogs	Confidence Score	Interaction of Common Interactors with HTT	Ranking by # of Interactions with HTT-Interactors *
MID1, IKBKB, IKBKG	Negative	PPP2CA	$\geq$0.72	Yes	402 (20)
		TUBB	$\geq$0.63	Yes	81 (41)
		HSP90AA1	$\geq$0.63	Yes	51 (47)
		RPS3	$\geq$0.63	No	295 (23)
PML, OPTN, CHUK	Positive	HSPA9	$\geq$0.49	No	47 (48)

* Out of 10,914 proteins. Number of interactions with HTT interactors indicated in parentheses.

3. Discussion

The two-dimensional hyperbolic embedding of the human protein interaction network has been shown to be both relevant and useful. We previously showed (i) that the radial coordinates of nodes (proteins) correlate with protein age, with older proteins occupying more central positions, and (ii) that proteins with related biological functions and cellular localizations cluster together along the angular coordinates [29]. The reason for these distributions, in terms of the protein interaction network, is that older proteins have more interactions, and their corresponding nodes are shifted towards the center of the map, while proteins with similar functions tend to be part of the same complexes and pathways and, therefore, because they interact or have common interactors, they tend to be pulled together towards the same region of the map.

In this paper, the latent geometry of the hPIN proves useful, namely in the network-based analysis of huntingtin's interactors. Considering pairs of paralogs that (a) both interact with huntingtin, (b) are located in different regions of the hPIN, and (c) have opposite effects on HD, we found three pairs that correspond to these criteria. Particularly, MID1 is an aberrant interaction partner of HTT. Its binding leads to the induction of an aberrant translation of the mutant HTT mRNA. MID1 assembles a protein complex with its interaction partners, PP2A and 40S ribosomal S6 kinase (S6K), and recruits this complex to the mutant HTT mRNA. This recruitment induces translation in a CAG repeat length-dependent manner, resulting in a toxic gain of function [46]. The translational induction by MID1 has also been found in models of other CAG repeat diseases [53]. Heinz and colleagues found that blocking the interaction between MID1 and the mutant HTT mRNA is a promising therapeutic approach [13]. On the other hand, the MID1 paralog protein, PML, can associate directly with polyQ proteins and preferentially with the pathogenic form, recognizing structures or regions that are commonly found in misfolded proteins [45]. Misfolded nuclear proteins that are selectively recognized by PML are marked with poly-SUMO2/SUMO3 chains. RNF4, which is a SUMO-dependent E3 ubiquitin ligase, binds to the poly-SUMO2/SUMO3 chains via tandem SUMO interacting motifs (SIM) and ubiquitylates, the protein, which leads to its proteasomal degradation [44]. This relay system likely provides a critical link between misfolded proteins and may play an important role in protecting against neurodegeneration.

Another paralog pair with opposite effects on HD is IKBKB/CHUK(IKKA). Most of the IKKA and IKBKB molecules in the cell are part of IKK complexes. The IKK complex also contains a regulatory subunit called IKKγ or NEMO [54]. Concurrently, DNA damage is an important factor in the development of neurotoxicity and a potential regulator of HD pathology [55]. It was shown that the induction of DNA damage has opposite effects on this paralog pair, increasing the activity of IKBKB while decreasing the activity of IKKA in the neurons [47]. The increased activity of IKBKB is also involved in several neurodegenerative disorders, including HD, Alzheimer's disease (AD), and Parkinson's disease (PD) [49,56,57]. The IKBKB activation by the DNA damage promotes HTT cleavage, and by increasing IKKA or reducing IKBKB, blocks this event. In the context of neuronal DNA damage, IKBKB activation is deleterious, and its inhibition may be protective in HD and potentially in other neurodegenerative disorders where DNA damage plays a role [47]. Moreover, the inactivation of IKBKB prevents the development of metabolic abnormalities induced by mutant HTT in the hypothalamus [58].

IKBKG, which is the regulatory module of the IKK complex, binds to mutant HTT through polyQ and polyP regions. This binding activates the IKK complex and promotes the activation and nuclear localization of the nuclear factor kappa (NF-κB) [49]. Activated NF-κB is involved in neuronal injury and in pathological conditions, such as HD [59,60]. The inhibition of NF-κB may have a protective effect on excitotoxicity, apoptosis, and neurodegeneration and, therefore, NF-κB inhibitors may deserve investigation for their potential role in HD [51]. Optineurin (OPTN) is one of a number of HTT-interacting proteins [4] that promotes neuronal survival by counteracting the glutamate-induced neurotoxicity in diseases, demonstrating its neuroprotective role [48,61]. Shen and colleagues also showed

that OPTN decreased the misfolded protein aggregates, mainly through polyUbK63-linked autophagy, while OPTN mutations lead to diseases by altering the protein quality control and degradation machinery [50].

We next focus on common partners of the three negative proteins: TUBB, PPP2CA, HSP90AA1, and RPS3. It is interesting to note that TUBB is part of our paralog interacting set with a negative effect [16]. Furthermore, PPP2CA, HSP90AA1, and RPS3 are involved in the pathogenesis of another polyQ disease named SCA1. These proteins are members of a protein–protein interaction network, which is affected by the gradual aggregation of the relevant polyQ-expanded protein, ataxin-1, and the degeneration of Purkinje neurons in animal models [62]. HSP90AA1 interacts with the N-terminal of HTT and recruits the deubiquinating enzyme USP19 [13]. Additionally, this protein participates in a chaperome network, safeguarding proteostasis, and is repressed in the brain of patients with neurodegenerative diseases, including HD [63].

PPP2CA is an interesting candidate because it dephosphorylates S421 in HTT and, blocking its activity, was found to protect striatal neurons from NMDA-induced cell death [64]. This protein not only regulates translation of HTT mRNAs through the MID1-PP2A complex [46] but may also induce apoptotic cell death through the activation of the mTOR/PI3K/Akt pathway [65]. The selection of the ribosomal protein RPS3 in this network suggests a new avenue for exploration. Ribosomal proteins preferentially interact with the mutant HTT [66], suggesting the participation of the protein translation machinery in the pathogenesis of polyQ diseases [67]. The shuttling of RPS3 from the cytoplasm to the nucleus can be induced by toxic DNA damage [68] and to mitochondria by increased ROS levels [69]. Using the same approach with the three positive proteins, we found only one common partner: chaperone HSPA9 (also known as mortalin), which is not a direct interactor of HTT. However, its positive effect may be mediated by other members of the heat shock protein 70 family, including HSPA8, which was previously shown to preferentially interact with HTT [66]. The interactions of HSPA9 with OPTN, PML, and CHUK are reported by non-specific works [70–72] but could also suggest a positive effect. One high-throughput non-specific interaction study links it also to the negative HD modifiers, IKBKB, IKBKG, and TUBB [71]. We take recent work linking HSPA9 to roles in the control of peroxisomal function [73] and neuronal stress detection [74] and its downregulation in animal models of Alzheimer's disease and patient's brains [75] as a suggestion that this could be a relevant protein for the control of HD.

4. Materials and Methods

4.1. Construction of the hPIN

The hPIN is a subset of the Human Integrated Protein–Protein Interaction rEference (HIPPIE) [32,33]. HIPPIE retrieves interactions between human proteins from major expert-curated databases and calculates a score for each one, reflecting its combined experimental evidence. This score is a combination of the number of studies that detect an interaction, the quality of experimental techniques used to measure an interaction, and the number of non-human organisms in which an interaction was reproduced. The raw version of this network is available in the Download section of the HIPPIE database [32,33]. In this study, only interactions with a confidence score of $\geq$0.71 that belong to the largest connected component (LCC) in release 2.2 were considered (N = 13,076 nodes and L = 93,140 edges). The 0.71-network was preferred because it has a high percentage of edges supported by more than one experiment (70%).

4.2. Mapping the hPIN to Hyperbolic Space

In order to embed the hPIN into the two-dimensional hyperbolic plane, we used the R package "NetHypGeom," which implements the LaBNE + HM algorithm [35]. This algorithm combines manifold learning [34,35] and maximum likelihood estimation [36] to uncover the hidden geometry of complex networks. The PS model has a geometrical interpretation in hyperbolic space (H^2), where nodes that join the system connect with the

existing ones that are hyperbolically closest to them [30,36,76]. The N nodes of the network lie within a hyperbolic disc with a radius of R~N, where the radial coordinate of a node, r_i, represents the popularity dimension with nodes that joined the system first being close to the disc's center. The angular coordinate, θ_i, represents the similarity dimension. The network was embedded in the two-dimensional hyperbolic plane using the LaBNE + HM algorithm to infer the hyperbolic coordinates of each protein, with parameters γ = 2.74, T = 0.8, and $w = 2\pi$.

4.3. Clustering in the Similarity Dimension

To cluster proteins in the similarity dimension, we sorted the nodes increasingly by their angular coordinates and computed the difference between θ_i and θ_{i+1} to identify large gaps between groups of proteins. The gap size, g, that was chosen to separate protein clusters produces sectors with a minimum of ten components (g = 0.011344, see Supplementary Figure S1). Then, we applied an ad hoc rule, where clusters with less than 100 proteins were merged clockwise with the consecutive one to avoid redundancy. We then carried out Gene Ontology (GO) enrichment analysis [77] for the proteins in each sector of the hPIN, using the nodes of the hPIN as our background set. Only GO Biological Process (BP) terms enriched at the 0.05 significance level (p-value) were kept.

4.4. HTT Interactors in the Hyperbolic Space

We then obtained the list of human HTT interactors from the HIPPIEv2.2 database [32,33] and identified their position in the hyperbolic disc. We created groups of proteins in the HttPIN based on the angular similarity dimension of the HTT interactors. To determine the start and the end of each group, proteins were sorted increasingly by their inferred angular coordinate, θ, and the difference between θ_i and θ_{i+1} was computed. The gap size = 0.059198 was chosen (Supplementary Figure S2). The enriched GO BP terms for each group were determined and the ones enriched at the 0.05 significance level (p-values) were extracted.

4.5. Paralog Pairs and Common Interacting Partners

From the HTT interacting proteins dataset, we detected pairs of paralogs. This information was derived from the Ensembl BioMart database [78], using the human genome assembly, GRCh38.p13. We focused on pairs of paralogous proteins located in different clusters in the H^2 to explore functional interpretations based on the angular similarity dimension. We then conducted a literature review to identify pairs with opposite effects on HD. For the latter analysis, we used the HIPPIEv2.2 database [32,33] to obtain common interacting partners between pairs of paralogs with negative and positive effects on HD, applying different confidence scores.

5. Conclusions

We approached the heterogeneity of the measurements of the effects of various proteins in HD models by providing a common framework for evaluation. Our hypothesis is that strong HD modifiers should produce collective effects through multiple pathways and complexes. These strong functional effectors may be obscured by the supposedly abundant but weaker effects of proteins that influence HD aggregates by their passive interaction with HTT. To avoid this problem, we focused on pairs of HTT interactor paralogs occupying divergent positions in the protein interaction network mapped to H^2 and found pairs with opposite effects on HD. We then explored the components closely connected to the positive or negative effectors. Our findings confirm proteins with relevant effects in HD and suggest RPS3 and HSPA9 as non-direct interactors of HTT that could have a negative and positive effect in HD, respectively. With our approach, we have shown how the interaction network connects the effects of HD modifiers to the literature, and the finer details of each experiment can ultimately be examined to make sense of these results and select or discard ideas for experimental work.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23105853/s1.

Author Contributions: Conceptualization, A.-C.V., P.M. and M.A.A.-N.; methodology, A.-C.V. and M.A.A.-N.; software, A.-C.V.; validation, A.-C.V. and M.A.A.-N.; formal analysis, A.-C.V.; investigation, A.-C.V.; writing—original draft preparation, A.-C.V. and M.A.A.-N.; writing—review and editing, A.-C.V., P.M., S.P. and M.A.A.-N.; supervision, M.A.A.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Gregorio Alanis-Lobato for fruitful discussions and comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Macdonald, M. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **1993**, *72*, 971–983. [CrossRef]
2. Ordway, J.M.; Tallaksen-Greene, S.; Gutekunst, C.-A.; Bernstein, E.M.; Cearley, J.A.; Wiener, H.W.; Dure, L.S.; Lindsey, R.; Hersch, S.M.; Jope, R.S.; et al. Ectopically Expressed CAG Repeats Cause Intranuclear Inclusions and a Progressive Late Onset Neurological Phenotype in the Mouse. *Cell* **1997**, *91*, 753–763. [CrossRef]
3. Kaminosono, S.; Saito, T.; Oyama, F.; Ohshima, T.; Asada, A.; Nagai, Y.; Nukina, N.; Hisanaga, S.I. Suppression of Mutant Huntingtin Aggregate Formation by Cdk5/p35 through the Effect on Microtubule Stability. *J. Neurosci.* **2008**, *28*, 8747–8755. [CrossRef] [PubMed]
4. Harjes, P.; Wanker, E.E. The hunt for huntingtin function: Interaction partners tell many different stories. *Trends Biochem. Sci.* **2003**, *28*, 425–433. [CrossRef]
5. Kaltenbach, L.S.; Romero, E.; Becklin, R.R.; Chettier, R.; Bell, R.; Phansalkar, A.; Strand, A.; Torcassi, C.; Savage, J.; Hurlburt, A.; et al. Huntingtin Interacting Proteins Are Genetic Modifiers of Neurodegeneration. *PLoS Genet.* **2007**, *3*, e82. [CrossRef]
6. Li, S.-H.; Li, X.-J. Huntingtin–protein interactions and the pathogenesis of Huntington's disease. *Trends Genet.* **2004**, *20*, 146–154. [CrossRef]
7. DiFiglia, M.; Sapp, E.; Chase, K.O.; Davies, S.W.; Bates, G.P.; Vonsattel, J.P.; Aronin, N. Aggregation of Huntingtin in Neuronal Intranuclear Inclusions and Dystrophic Neurites in Brain. *Science* **1997**, *277*, 1990–1993. [CrossRef]
8. Hoffner, G.; Island, M.L.; Djian, P. Purification of neuronal inclusions of patients with Huntington's disease reveals a broad range of N-terminal fragments of expanded huntingtin and insoluble polymers: Huntingtin fragments and polymers in inclusions. *J. Neurochem.* **2005**, *95*, 125–136. [CrossRef]
9. Boutell, J.M.; Thomas, P.; Neal, J.W.; Weston, V.J.; Duce, J.; Harper, P.S.; Jones, A.L. Aberrant Interactions of Transcriptional Repressor Proteins with the Huntington's Disease Gene Product, Huntingtin. *Hum. Mol. Genet.* **1999**, *8*, 1647–1655. [CrossRef]
10. Nucifora, F.C.; Sasaki, M.; Peters, M.F.; Huang, H.; Cooper, J.K.; Yamada, M.; Takahashi, H.; Tsuji, S.; Troncoso, J.; Dawson, V.L.; et al. Interference by Huntingtin and Atrophin-1 with CBP-Mediated Transcription Leading to Cellular Toxicity. *Science* **2001**, *291*, 2423–2428. [CrossRef]
11. Suhr, S.T.; Senut, M.-C.; Whitelegge, J.P.; Faull, K.F.; Cuizon, D.B.; Gage, F.H. Identities of Sequestered Proteins in Aggregates from Cells with Induced Polyglutamine Expression. *J. Cell Biol.* **2001**, *153*, 283–294. [CrossRef] [PubMed]
12. Gerson, J.E.; Safren, N.; Fischer, S.; Patel, R.; Crowley, E.V.; Welday, J.P.; Windle, A.K.; Barmada, S.; Paulson, H.L.; Sharkey, L.M. Ubiquilin-2 differentially regulates polyglutamine disease proteins. *Hum. Mol. Genetics* **2020**, *29*, 2596–2610. [CrossRef] [PubMed]
13. Heinz, A.; Schilling, J.; van Roon-Mom, W.; Krauß, S. The MID1 Protein: A Promising Therapeutic Target in Huntington's Disease. *Front. Genet.* **2021**, *12*, 761714. [CrossRef] [PubMed]
14. Hyrskyluoto, A.; Bruelle, C.; Lundh, S.H.; Do, H.T.; Kivinen, J.; Rappou, E.; Reijonen, S.; Waltimo, T.; Petersén, Å.; Lindholm, D.; et al. Ubiquitin-specific protease-14 reduces cellular aggregates and protects against mutant huntingtin-induced cell degeneration: Involvement of the proteasome and ER stress-activated kinase IRE1. *Hum. Mol. Genet.* **2014**, *23*, 5928–5939. [CrossRef]
15. Min, H.J.; Ko, E.A.; Wu, J.; Kim, E.S.; Kwon, M.K.; Kwak, M.S.; Choi, J.E.; Lee, J.E.; Shin, J.-S. Chaperone-like Activity of High-Mobility Group Box 1 Protein and Its Role in Reducing the Formation of Polyglutamine Aggregates. *J. Immunol.* **2013**, *190*, 1797–1806. [CrossRef]

16. Smith, R.; Bacos, K.; Fedele, V.; Soulet, D.; Walz, H.A.; Obermüller, S.; Lindqvist, A.; Björkqvist, M.; Klein, P.; Önnerfjord, P.; et al. Mutant huntingtin interacts with -tubulin and disrupts vesicular transport and insulin secretion. *Hum. Mol. Genet.* **2009**, *18*, 3942–3954. [CrossRef]
17. Wang, H.Q.; Xu, Y.X.; Zhao, X.Y.; Zhao, H.; Yan, J.; Sun, X.B.; Guo, J.-C.; Zhu, C.-Q. Overexpression of F0F1-ATP synthase α suppresses mutant huntingtin aggregation and toxicity in vitro. *Biochem. Biophys. Res. Commun.* **2009**, *390*, 1294–1298. [CrossRef]
18. Yang, H.; Liu, C.; Zhong, Y.; Luo, S.; Monteiro, M.J.; Fang, S. Huntingtin Interacts with the Cue Domain of gp78 and Inhibits gp78 Binding to Ubiquitin and p97/VCP. *PLoS ONE* **2010**, *5*, e8905. [CrossRef]
19. Cariulo, C.; Azzollini, L.; Verani, M.; Martufi, P.; Boggio, R.; Chiki, A.; Deguire, S.M.; Cherubini, M.; Gines, S.; Marsh, J.L.; et al. Phosphorylation of huntingtin at residue T3 is decreased in Huntington's disease and modulates mutant huntingtin protein conformation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E10809–E10818. [CrossRef]
20. Mishra, R.; Hoop, C.L.; Kodali, R.; Sahoo, B.; van der Wel, P.C.A.; Wetzel, R. Serine Phosphorylation Suppresses Huntingtin Amyloid Accumulation by Altering Protein Aggregation Properties. *J. Mol. Biol.* **2012**, *424*, 1–14. [CrossRef]
21. Petrakis, S.; Rasko, T.; Russ, J.; Friedrich, R.P.; Stroedicke, M.; Riechers, S.-P.; Muehlenberg, K.; Möller, A.; Reinhardt, A.; Vinayagam, A.; et al. Identification of Human Proteins That Modify Misfolding and Proteotoxicity of Pathogenic Ataxin-1. *PLoS Genet.* **2012**, *8*, e1002897. [CrossRef] [PubMed]
22. Petrakis, S.; Schaefer, M.H.; Wanker, E.E.; Andrade-Navarro, M.A. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *BioEssays* **2013**, *35*, 503–507. [CrossRef] [PubMed]
23. Kastano, K.; Mier, P.; Andrade-Navarro, M.A. The Role of Low Complexity Regions in Protein Interaction Modes: An Illustration in Huntingtin. *Int. J. Mol. Sci.* **2021**, *22*, 1727. [CrossRef] [PubMed]
24. Jensen, L.J.; Ussery, D.W.; Brunak, S. Functionality of System Components: Conservation of Protein Function in Protein Feature Space. *Genome Res.* **2003**, *13*, 2444–2449. [CrossRef]
25. Zallot, R.; Harrison, K.J.; Kolaczkowski, B.; De Crécy-Lagard, V. Functional Annotations of Paralogs: A Blessing and a Curse. *Life* **2016**, *6*, 39. [CrossRef] [PubMed]
26. Serlidaki, D.; van Waarde, M.A.W.H.; Rohland, L.; Wentink, A.S.; Dekker, S.L.; Kamphuis, M.J.; Boertien, J.M.; Brunsting, J.F.; Nillegoda, N.B.; Bukau, B.; et al. Functional diversity between HSP70 paralogs caused by variable interactions with specific co-chaperones. *J. Biol. Chem.* **2020**, *295*, 7301–7316. [CrossRef]
27. Favaro, P.D.; Huang, X.; Hosang, L.; Stodieck, S.; Cui, L.; Liu, Y.-z.; Engelhardt, K.-A.; Schmitz, F.; Dong, Y.; Löwel, S.; et al. An opposing function of paralogs in balancing developmental synapse maturation. *PLoS Biol.* **2018**, *16*, e2006838. [CrossRef]
28. Gosak, M.; Markovič, R.; Dolenšek, J.; Slak Rupnik, M.; Marhl, M.; Stožer, A.; Perc, M. Network science of biological systems at different scales: A review. *Phys. Life Rev.* **2018**, *24*, 118–135. [CrossRef]
29. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M. The latent geometry of the human protein interaction network. *Bioinformatics* **2018**, *34*, 2826–2834. [CrossRef]
30. Papadopoulos, F.; Kitsak, M.; Serrano, M.Á.; Boguñá, M.; Krioukov, D. Popularity versus similarity in growing networks. *Nature* **2012**, *489*, 537–540. [CrossRef]
31. Härtner, F.; Andrade-Navarro, M.A.; Alanis-Lobato, G. Geometric characterisation of disease modules. *Appl. Netw. Sci.* **2018**, *3*, 10. [CrossRef] [PubMed]
32. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **2017**, *45*, D408–D414. [CrossRef] [PubMed]
33. Schaefer, M.H.; Fontaine, J.F.; Vinayagam, A.; Porras, P.; Wanker, E.E.; Andrade-Navarro, M.A. HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. *PLoS ONE* **2012**, *7*, e31826. [CrossRef] [PubMed]
34. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M.A. Efficient embedding of complex networks to hyperbolic space via their Laplacian. *Sci. Rep.* **2016**, *6*, 30108. [CrossRef] [PubMed]
35. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M.A. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Appl. Netw. Sci.* **2016**, *1*, 10. [CrossRef]
36. Papadopoulos, F.; Aldecoa, R.; Krioukov, D. Network geometry inference using common neighbors. *Phys. Rev. E* **2015**, *92*, 022807. [CrossRef]
37. Musso, G.; Zhang, Z.; Emili, A. Retention of protein complex membership by ancient duplicated gene products in budding yeast. *Trends Genet.* **2007**, *23*, 266–269. [CrossRef]
38. Wagner, A. The Yeast Protein Interaction Network Evolves Rapidly and Contains Few Redundant Duplicate Genes. *Mol. Biol. Evol.* **2001**, *18*, 1283–1292. [CrossRef]
39. Wacker, J.L.; Zareie, M.H.; Fong, H.; Sarikaya, M.; Muchowski, P.J. Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1215–1222. [CrossRef]
40. Chai, Y.; Koppenhafer, S.L.; Bonini, N.M.; Paulson, H.L. Analysis of the Role of Heat Shock Protein (Hsp) Molecular Chaperones in Polyglutamine Disease. *J. Neurosci.* **1999**, *19*, 10338–10347. [CrossRef]
41. Rodríguez-González, C.; Lin, S.; Arkan, S.; Hansen, C. Co-chaperones DNAJA1 and DNAJB6 are critical for regulation of polyglutamine aggregation. *Sci. Rep.* **2020**, *10*, 8130. [CrossRef] [PubMed]
42. Kitamura, A.; Kubota, H.; Pack, C.G.; Matsumoto, G.; Hirayama, S.; Takahashi, Y.; Kimura, H.; Kinjo, M.; Morimoto, R.; Nagata, K. Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nat. Cell Biol.* **2006**, *8*, 1163–1169. [CrossRef] [PubMed]

43. Langfelder, P.; Cantle, J.P.; Chatzopoulou, D.; Wang, N.; Gao, F.; Al-Ramahi, I.; Lu, X.-H.; Ramos, E.M.; El-Zein, K.; Zhao, Y.; et al. Integrated genomics and proteomics define huntingtin CAG length–dependent networks in mice. *Nat. Neurosci.* **2016**, *19*, 623–633. [CrossRef] [PubMed]
44. Gärtner, A.; Muller, S. PML, SUMO, and RNF4: Guardians of Nuclear Protein Quality. *Mol. Cell* **2014**, *55*, 1–3. [CrossRef]
45. Guo, L.; Giasson, B.I.; Glavis-Bloom, A.; Brewer, M.D.; Shorter, J.; Gitler, A.D.; Yang, X. A Cellular System that Degrades Misfolded Proteins and Protects against Neurodegeneration. *Mol. Cell* **2014**, *55*, 15–30. [CrossRef]
46. Krauß, S.; Griesche, N.; Jastrzebska, E.; Chen, C.; Rutschow, D.; Achmüller, C.; Dorn, S.; Boesch, S.M.; Lalowski, M.; Wanker, E.; et al. Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1–PP2A protein complex. *Nat. Commun.* **2013**, *4*, 1511. [CrossRef]
47. Khoshnan, A.; Ko, J.; Tescu, S.; Brundin, P.; Patterson, P.H. IKKα and IKKβ Regulation of DNA Damage-Induced Cleavage of Huntingtin. *PLoS ONE* **2009**, *4*, e5768. [CrossRef]
48. Anborgh, P.H.; Godin, C.; Pampillo, M.; Dhami, G.K.; Dale, L.B.; Cregan, S.P.; Truant, R.; Ferguson, S.S.G. Inhibition of Metabotropic Glutamate Receptor Signaling by the Huntingtin-binding Protein Optineurin. *J. Biol. Chem.* **2005**, *280*, 34840–34848. [CrossRef]
49. Khoshnan, A.; Ko, J.; Watkin, E.E.; Paige, L.A.; Reinhart, P.H.; Patterson, P.H. Activation of the I B Kinase Complex and Nuclear Factor- B Contributes to Mutant Huntingtin Neurotoxicity. *J. Neurosci.* **2004**, *24*, 7999–8008. [CrossRef]
50. Shen, W.-C.; Li, H.-Y.; Chen, G.-C.; Chern, Y.; Tu, P.-h. Mutations in the ubiquitin-binding domain of OPTN/optineurin interfere with autophagy-mediated degradation of misfolded proteins by a dominant-negative mechanism. *Autophagy* **2015**, *11*, 685–700. [CrossRef]
51. Gupta, S.; Sharma, B. Pharmacological benefit of I1-imidazoline receptors activation and nuclear factor kappa-B (NF-κB) modulation in experimental Huntington's disease. *Brain Res. Bull.* **2014**, *102*, 57–68. [CrossRef] [PubMed]
52. Kalathur, R.K.R.; Giner-Lamia, J.; Machado, S.; Barata, T.; Ayasolla, K.R.S.; Futschik, M.E. The unfolded protein response and its potential role in Huntington's disease elucidated by a systems biology approach. *F1000Research* **2015**, *4*, 103. [CrossRef] [PubMed]
53. Griesche, N.; Schilling, J.; Weber, S.; Rohm, M.; Pesch, V.; Matthes, F.; Auburger, G.; Krauss, S. Regulation of mRNA Translation by MID1: A Common Mechanism of Expanded CAG Repeat RNAs. *Front. Cell. Neurosci.* **2016**, *10*, 226. [CrossRef] [PubMed]
54. Häcker, H.; Karin, M. Regulation and Function of IKK and IKK-Related Kinases. *Sci. STKE* **2006**, *2006*, re13. [CrossRef] [PubMed]
55. De Luca, G.; Russo, M.T.; Degan, P.; Tiveron, C.; Zijno, A.; Meccia, E.; Ventura, I.; Mattei, E.; Nakabeppu, Y.; Crescenzi, M.; et al. A Role for Oxidized DNA Precursors in Huntington's Disease–Like Striatal Neurodegeneration. *PLoS Genet.* **2008**, *4*, e1000266. [CrossRef] [PubMed]
56. Ghosh, A.; Roy, A.; Liu, X.; Kordower, J.H.; Mufson, E.J.; Hartley, D.M.; Ghosh, S.; Mosley, R.L.; Gendelman, H.E.; Pahan, K. Selective inhibition of NF-κB activation prevents dopaminergic neuronal loss in a mouse model of Parkinson's disease. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18754–18759. [CrossRef]
57. Mattson, M.P.; Culmsee, C.; Yu, Z.; Camandola, S. Roles of Nuclear Factor κB in Neuronal Survival and Plasticity. *J. Neurochem.* **2001**, *74*, 443–456. [CrossRef]
58. Soylu-Kucharz, R.; Khoshnan, A.; Petersén, Å. IKKβ signaling mediates metabolic changes in the hypothalamus of a Huntington disease mouse model. *iScience* **2022**, *25*, 103771. [CrossRef]
59. Napolitano, M.; Zei, D.; Centonze, D.; Palermo, R.; Bernardi, G.; Vacca, A.; Calabresi, P.; Gulino, A. NF-kB/NOS cross-talk induced by mitochondrial complex II inhibition: Implications for Huntington's disease. *Neurosci. Lett.* **2008**, *434*, 241–246. [CrossRef]
60. Yakovleva, T.; Bazov, I.; Watanabe, H.; Hauser, K.F.; Bakalkin, G. Transcriptional control of maladaptive and protective responses in alcoholics: A role of the NF-κB system. *Brain Behav. Immun.* **2011**, *25*, S29–S38. [CrossRef]
61. De Marco, N.; Buono, M.; Troise, F.; Diez-Roux, G. Optineurin Increases Cell Survival and Translocates to the Nucleus in a Rab8-dependent Manner upon an Apoptotic Stimulus. *J. Biol. Chem.* **2006**, *281*, 16147–16156. [CrossRef] [PubMed]
62. Vagiona, A.-C.; Andrade-Navarro, M.A.; Psomopoulos, F.; Petrakis, S. Dynamics of a Protein Interaction Network Associated to the Aggregation of polyQ-Expanded Ataxin-1. *Genes* **2020**, *11*, 1129. [CrossRef] [PubMed]
63. Brheme, M.; Voisine, C.; Rolland, T.; Wachi, S.; Soper, J.H.; Zhu, Y.; Orton, K.; Villella, A.; Garza, D.; Vidal, M.; et al. A chaperone subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* **2014**, *9*, 1135–1150. [CrossRef] [PubMed]
64. Metzler, M.; Gan, L.; Mazarei, G.; Graham, R.K.; Liu, L.; Bissada, N.; Lu, G.; Leavitt, B.R.; Hayden, M.R. Phosphorylation of Huntingtin at Ser421 in YAC128 Neurons Is Associated with Protection of YAC128 Neurons from NMDA-Mediated Excitotoxicity and Is Modulated by PP1 and PP2A. *J. Neurosci.* **2010**, *30*, 14318–14329. [CrossRef]
65. Zhou, H.; Luo, W.; Zeng, C.; Zhang, Y.; Wang, L.; Yao, W.; Nie, C. PP2A mediates apoptosis or autophagic cell death in multiple myeloma cell lines. *Oncotarget* **2017**, *8*, 80770–80789. [CrossRef]
66. Haenig, C.; Atias, N.; Taylor, A.K.; Mazza, A.; Schaefer, M.H.; Russ, J.; Riechers, S.-P.; Jain, S.; Coughlin, M.; Fontaine, J.-F.; et al. Interactome Mapping Provides a Network of Neurodegenerative Disease Proteins and Uncovers Widespread Protein Aggregation in Affected Brains. *Cell Rep.* **2020**, *32*, 108050. [CrossRef]
67. Laidou, S.; Alanis-Lobato, G.; Pribyl, J.; Raskó, T.; Tichy, B.; Mikulasek, K.; Tsagiopoulou, M.; Oppelt, J.; Kastrinaki, G.; Lefaki, M.; et al. Nuclear inclusions of pathogenic ataxin-1 induce oxidative stress and perturb the protein synthesis machinery. *Redox Biol.* **2020**, *32*, 101458. [CrossRef]

68. Yadavilli, S.; Hegde, V.; Deutsch, W.A. Translocation of human ribosomal protein S3 to sites of DNA damage is dependant on ERK-mediated phosphorylation following genotoxic stress. *DNA Repair* **2007**, *6*, 1453–1462. [CrossRef]
69. Kim, Y.; Kim, H.D.; Kim, J. Cytoplasmic ribosomal protein S3 (rpS3) plays a pivotal role in mitochondrial DNA damage surveillance. *Biochim. Biophys. Acta BBA Mol. Cell Res.* **2013**, *1833*, 2943–2952. [CrossRef]
70. Bell, R.; Hubbard, A.; Chettier, R.; Chen, D.; Miller, J.P.; Kapahi, P.; Tarnopolsky, M.; Sahasrabuhde, S.; Melov, S.; Hughes, R.E. A Human Protein Interaction Network Shows Conservation of Aging Processes between Human and Invertebrate Species. *PLoS Genet.* **2009**, *5*, e1000414. [CrossRef]
71. Bouwmeester, T.; Bauch, A.; Ruffner, H.; Angrand, P.O.; Bergamini, G.; Croughton, K.; Cruciat, C.; Eberhard, D.; Gagneur, J.; Ghidelli, S.; et al. A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. *Nat. Cell Biol.* **2004**, *6*, 97–105. [CrossRef] [PubMed]
72. Miyamoto-Sato, E.; Fujimori, S.; Ishizaka, M.; Hirai, N.; Masuoka, K.; Saito, R.; Ozawa, Y.; Hino, K.; Washio, T.; Tomita, M.; et al. A Comprehensive Resource of Interacting Protein Regions for Refining Human Transcription Factor Networks. *PLoS ONE* **2010**, *5*, e9289. [CrossRef] [PubMed]
73. Jo, D.S.; Park, S.J.; Kim, A.K.; Park, N.Y.; Kim, J.B.; Bae, J.E.; Park, J.H.; Shin, J.H.; Chang, J.W.; Kim, P.K.; et al. Loss of HSPA9 induces peroxisomal degradation by increasing pexophagy. *Autophagy* **2020**, *16*, 1989–2003. [CrossRef] [PubMed]
74. Burbulla, L.F.; Schelling, C.; Kato, H.; Rapaport, D.; Woitalla, D.; Schiesling, C.; Schulte, C.; Sharma, M.; Illig, T.; Bauer, P.; et al. Dissecting the role of the mitochondrial chaperone mortalin in Parkinson's disease: Functional impact of disease-related variants on mitochondrial homeostasis. *Hum. Mol. Genet.* **2010**, *19*, 4437–4452. [CrossRef] [PubMed]
75. Park, S.J.; Shin, J.H.; Jeong, J.I.; Song, J.H.; Jo, Y.K.; Kim, E.S.; Lee, E.H.; Hwang, J.J.; Lee, E.K.; Chung, J.S.; et al. Down-regulation of Mortalin Exacerbates Aβ-mediated Mitochondrial Fragmentation and Dysfunction. *J. Biol. Chem.* **2014**, *289*, 2195–2204. [CrossRef] [PubMed]
76. Krioukov, D.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; Boguna, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **2010**, *82*, 036106. [CrossRef]
77. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
78. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [CrossRef]

Chapter 2

Prediction of Protein Interactions with Function in Protein (De-) Phosphorylation

Aimilia-Christina Vagiona, Sofia Notopoulou, Zbyněk Zdráhal, Mariane Gonçalves-Kulik, Spyros Petrakis and Miguel A. Andrade-Navarro

Article published in *PLOS ONE, vol. 20, 3, e0319084. (3 March 2025),*

doi: https://doi.org/10.1371/journal.pone.0319084

PLOS ONE

RESEARCH ARTICLE

Prediction of protein interactions with function in protein (de-)phosphorylation

Aimilia-Christina Vagiona[1], Sofia Notopoulou[2], Zbyněk Zdráhal[3], Mariane Gonçalves-Kulik[1], Spyros Petrakis[2], Miguel A. Andrade-Navarro[1]*

1 Faculty of Biology, Insitute of Organismic and Molecular Evolution, Johannes Gutenberg University, Biozentrum I, Mainz, Germany, **2** Institute of Applied Biosciences/Centre for Research and Technology Hellas, Thessaloniki, Greece, **3** Central European Institute of Technology, Masaryk University, Brno, Czech Republic

* andrade@uni-mainz.de

OPEN ACCESS

Citation: Vagiona A-C, Notopoulou S, Zdráhal Z, Gonçalves-Kulik M, Petrakis S, Andrade-Navarro MA (2025) Prediction of protein interactions with function in protein (de-) phosphorylation. PLoS ONE 20(3): e0319084. https://doi.org/10.1371/journal.pone.0319084

Editor: Hilary A. Coller, UC Los Angeles: University of California Los Angeles, UNITED STATES OF AMERICA

Received: September 10, 2024

Accepted: January 28, 2025

Published: March 3, 2025

Data availability statement: Results were deposited in the PRIDE Archive (https://www.ebi.ac.uk/pride/archive, accession number: PXD038393).

Funding: The author(s) received no specific funding for this work.

Abstract

Protein–protein interactions (PPIs) form a complex network called "interactome" that regulates many functions in the cell. In recent years, there is an increasing accumulation of evidence supporting the existence of a hyperbolic geometry underlying the network representation of complex systems such as the interactome. In particular, it has been shown that the embedding of the human Protein-Interaction Network (hPIN) in hyperbolic space (H^2) captures biologically relevant information. Here we explore whether this mapping contains information that would allow us to predict the function of PPIs, more specifically interactions related to post-translational modification (PTM). We used a random forest algorithm to predict PTM-related directed PPIs, concretely, protein phosphorylation and dephosphorylation, based on hyperbolic properties and centrality measures of the hPIN mapped in H^2. To evaluate the efficacy of our algorithm, we predicted PTM-related PPIs of ataxin-1, a protein which is responsible for Spinocerebellar Ataxia type 1 (SCA1). Proteomics analysis in a cellular model revealed that several of the predicted PTM-PPIs were indeed dysregulated in a SCA1-related disease network. A compact cluster composed of ataxin-1, its dysregulated PTM-PPIs and their common upstream regulators may represent critical interactions for disease pathology. Thus, our algorithm may infer phosphorylation activity on proteins through directed PPIs.

Introduction

Protein–protein interactions (PPIs) play crucial roles in fundamental processes in living cells [1,2]. PPIs in cells form a complicated network which has been named "interactome" [3]. By coordinating the activity of many proteins and protein complexes, the interactome performs many functions, including signal transduction, cell growth and differentiation, catalytic metabolic reactions, activation or suppression of a protein, transportation of molecules, etc [4,5]. Studying PPIs can help to reveal the underlying molecular machinery in cells [6]. Aberrant PPIs are associated with a wide range of human diseases, including cancer, infectious diseases and neurodegenerative diseases [7,8]. Recent studies indicate that targeting and restoring

Competing interests: The authors have declared that no competing interests exist.

dysregulated PPIs is a promising strategy for drug development for therapeutic intervention [9,10].

Several studies support that complex networks, such as the interactome, are well suited to be modeled using hyperbolic geometry, a space whose mathematical properties naturally lead to the emergence of networks with scale invariance and strong clustering [11–13]. The Popularity-Similarity (PS) model provides a geometric interpretation in hyperbolic space (H^2) and assumes that the clustering and hierarchy of complex networks arise from tradeoffs between popularity and similarity of nodes [14]. Basically, in the PS model, the network nodes are situated within a circle at polar coordinates. The network nodes have a radial coordinate that represents their popularity or seniority, the angular coordinate reflects the similarity between nodes, and the hyperbolic distance between nodes abstracts an optimization process in which new nodes connect to nodes that are popular and similar.

Alanis-Lobato et al. found that the embedding of the human Protein-Interaction Network (hPIN) in hyperbolic space has biological interpretations in terms of the PS model. The radial positioning of the nodes encapsulates information about the conservation and the evolution of proteins, corresponding to popularity and seniority, where nodes closed to the center of the circle represent proteins that evolved earlier and had more time to receive connections from newer proteins situated in the periphery of the circle. The angular positioning reflects the functional similarity between proteins and is driven by interactions in pathways and protein complexes, thus capturing the functional and spatial organization in the cell [15]. This mapping can also lead to a better understanding of complex human disorders [16,17].

Information on protein interactions can be obtained by a variety of experimental methods and this data is systematically stored in specialized databases [18–21]. However, little is known about the function of many of these interactions, especially those obtained by high-throughput methods, like yeast-two-hybrid. Following our findings about the biological properties encapsulated in the mapping of the hPIN in hyperbolic space, here we explore if this mapping also contains information that would allow us to predict the function of PPIs.

PPIs may result in the post-translational modification (PTM) of one of the interacting proteins. PTMs are considered as covalent or enzymatic modifications of a protein occurring after protein synthesis. They are classified into different groups such as the addition of functional groups/chemical groups (acetylation, methylation, phosphorylation), the addition of a polypeptide chain (ubiquitination, SUMOylation), the addition of other complex molecules (palmitoylation, glycosylation), and amino acid modifications (proteolytic cleavage) [22].

In this work, we applied a machine learning method (random forest, RF) to predict whether PPIs result in PTMs using properties extracted from the mapping of the hPIN in hyperbolic space (Fig 1). To validate the potency of our algorithm, we predicted PTM-related protein interactions (PTM-PPIs) of ataxin-1, a protein implicated in Spinocerebellar ataxia type 1 (SCA1). SCA1 is a severe neurodegenerative disease caused by CAG-trinucleotide repeat expansions (> 39) in the ATXN1 gene. These mutations induce misfolding of polyQ-expanded ataxin-1, leading to its accumulation into toxic intranuclear inclusions in human neurons [23]. The exact mechanism of protein aggregation remains unknown. However, recent evidence indicates that abnormal PTMs in ataxin-1, especially phosphorylation, significantly accelerate the aggregation process [24]. Proteomics analysis in a cellular model of polyQ-expanded ataxin-1 aggregation enabled the construction of a perturbed hPIN [25]. The SCA1 hPIN network contained 12 out of 32 predicted PTM-PPIs directly related to common upstream regulators. A compact cluster composed of ataxin-1, its dysregulated PTM-PPIs and their upstream regulators highly correlated to SCA1, suggesting that it might represent a crucial part of disease pathology.

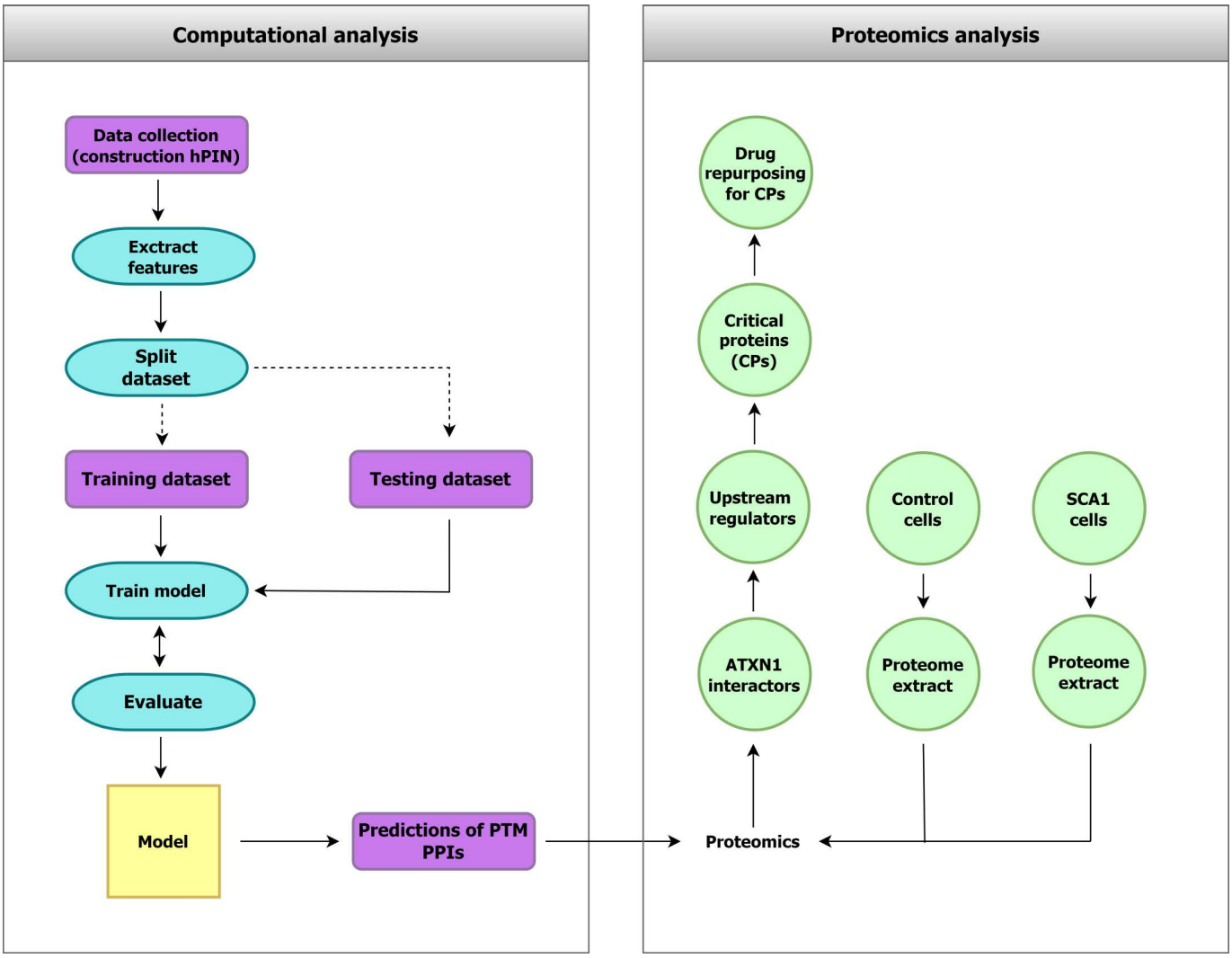


Fig 1. Structure of the work presented in this manuscript. Left: computational prediction of PTM-related directed protein interactions (PTM-PPIs). A random forest model was constructed to predict PTM-PPIs based on hyperbolic topological features and network properties extracted from the hPIN. **Right:** quantitative proteomics data revealed dysregulated proteins in a cellular model of SCA1. The results were interpreted using predicted PTM-PPIs, providing directions for future experimental research on SCA1 pathogenesis.

https://doi.org/10.1371/journal.pone.0319084.g001

Materials and methods

Human protein interaction network construction

The hPIN is a subset of release 2.3 of the Human Integrated Protein–Protein Interaction rEference (HIPPIE; (26,27)). HIPPIE retrieves interactions between human proteins from major expert-curated databases and calculates a score for each one, reflecting its combined experimental evidence. The raw version of this network is available in the Download section of the HIPPIE database. In this study, the hPIN was constructed using interactions with confidence score ≥ 0.71 (selects for a high percentage of interactions supported by at least two publications [17]). After discarding self-interactions and extracting the network's largest connected component (LCC) we obtained an hPIN consisting of 15,587 proteins (nodes; S1 Table) with 186,196 interactions (edges; S2 Table).

Mapping the human protein interactome in hyperbolic space

We embedded the hPIN in the two-dimensional hyperbolic plane using the R package "NetHypGeom", which implements the LaBNE + HM algorithm [28]. This algorithm combines manifold learning and maximum likelihood estimation to model the geometry of complex networks [29,30]. The PS model has a geometrical interpretation in hyperbolic space (H^2) where nodes that join the system connect with the existing ones that are hyperbolically closest to them [12,14,30]. The network was embedded in H^2 to infer the hyperbolic coordinates of each protein, with parameters $\gamma = 2.97$, $T = 0.83$, and $w = 2\pi$. The 15,587 nodes of the hPIN lie within a hyperbolic disc where the radial coordinate of a node, r_i, represents the popularity dimension with nodes that joined the system first being close to the disc's center. The angular coordinate, θ_i, represents the similarity dimension [25–27].

Clustering in the similarity dimension

To cluster proteins in the similarity dimension, we computed the difference between consecutive angular coordinates to identify big gaps. The nodes were sorted increasingly by their inferred angles θ, and the difference between θ_i and θ_{i+1} was computed to identify the largest gaps between protein clusters in the similarity dimension. Gap size ($g = 0.0077$) that produces sectors with a minimum of three components, was chosen. The same process was followed to subcluster the proteins of the first sector, with a minimum of five components in each subcluster and gap size, $g = 0.0042$. (S1 Fig). To determine the start and the end of each cluster, we chose gap sizes g that produced clusters with a minimum number of members (3 and 5 respectively) because this allowed us to perform meaningful functional enrichment analysis of each group of proteins. We carried out Gene Ontology (GO) enrichment analysis for the proteins in each sector of the hPIN, using the nodes of the network as background set. Only GO Biological Process (BP) terms enriched at a significance level (p-value) of 0.05 or less were kept. Neighboring clusters with similar biological functions were merged to avoid redundancy.

Selection of experimentally known phosphorylation and dephosphorylation PPIs

The functional associations for the interactions within the hPIN were extracted from multiple providers using the PSIQUIC webservice [31]. PSIQUIC enables access to molecular interaction databases supporting the PSI-MI format, which provides a hierarchical structure describing protein interactions. Specifically, we considered the interactions annotated with the children terms of the PSI-MI category 0414 *"enzymatic reaction"* and particularly we focused on two of them: PSI-MI category 0217 *"phosphorylation reaction"* and 0203 *"dephosphorylation reaction"*. The frequency of use of other terms (e.g., ubiquitination, methylation, acetylation) were too low for our purposes.

From the interactions annotated as phosphorylation or dephosphorylation, we selected those for which we were able to determine the direction of PTM activity from an effector protein (protein kinase or protein phosphatase) to a target according to the annotations of the interacting proteins. To identify effector proteins we used KinaseMD [32] and the human DEPhOsphorylation Database (DEPOD) [33]. We discarded cases in which neither protein was identified as a putative effector, or both proteins were identified as one effector type (protein kinase or protein phosphatase), because it is not possible to identify the direction of the PTM-related interaction in these. We obtained a total of 295 PPIs as PTM-related directed interactions (from effector protein to target protein; training dataset; S3 Table). Two cases involved a protein kinase and a protein phosphatase mutually acting on each other. These 295 PPIs were used as positives and the rest were used as negatives to train our model.

Feature extraction

We used a total of 14 features to train a classifier to detect PTM-related directed PPIs. Given a directed PPI to test, one node is taken as effector and the other as target according to the direction being tested. Six properties are taken for effector and target: two are their hyperbolic coordinates (r and theta), and the other four are measures of centrality. In network analysis, centrality measures evaluate the importance of a node based on certain parameters [34]. As measures of centrality, we used degree centrality (DC), betweenness centrality (BC), closeness centrality (CC) and eigenvector centrality (EC). DC is the number of immediate neighbours of a given node [35]. BC computes the significance of a node by calculating the fraction of all shortest paths that pass through it [36]. CC defines the proximity of a node to all the others [37] and EC reflects the influence of a node in a network [38]. The remaining two properties are defined for the edge: hyperbolic distance between the interacting proteins and r difference (absolute value). The values used are available in S1 and S2 Tables.

Model development and evaluation

Model developing was done using the "caret" package in R [39]. For the primary model building we used k-fold repeated cross validation in a training partition (70%) and validated the model in a leave out external validation sample (30%). We used the random forest (RF) algorithm [40] from the "caret" package to train our model. Five-fold repeated cross-validation was used (*repeats* = 10) to identify optimal hyperparameters. The parameter values were varied, and optimal values were chosen based on the accuracy (*mtry* = 14, *ntrees* = 500). We report accuracy scores in the 5-fold repeated cross validation (*repeats* = 10) samples. To address class imbalance, we used the under-sampling technique while sampling for cross validation. The importance of each feature was then calculated. This study implemented a ROC curve to determine the efficacy of the RF model. The receiver operator curve (ROC) represents the relationship between false positive rate and the true positive rate in a plane for each cut-off value used to define positive and negative classification results [41]. We then calculated the area under the curve (AUC) value, which describes the classifier's ability to discriminate between positive and negative results. It is a standard measurement of prediction quality and is commonly used to compare performance of models [42].

Comparison of predictions by alternative methods

We used kinase-substrate predictions by two alternative methods to add support to our predictions: PhosD [43] and Phosformer-ST [44]. We obtained a set of predictions from PhosD using a score threshold of 0.5 (1852 predictions). Of those, only 1062 overlapped with PPIs in our HIPPIE dataset. Phosformer-ST assigns scores to serine/threonine phosphorylation sites. To be able to compare this approach with ours, we re-assigned the predictions at a protein level, indicating as phosphorylated proteins that contain at least one peptide with a score above 0.5. Isoforms were removed from the comparison set, since the tools' predictions are based on sequence fragments, and they can diverge among isoforms. This resulted in a set of 451,724 predictions. Of those, only 961 overlapped with PPIs in our HIPPIE dataset.

Mass spectrometry (MS) analysis

The generation of Tet-On YFP-ATXN1(Q82) mesenchymal stem cells (MSCs) has been previously described [45]. Cells were cultured for 10 days in the presence or absence of doxycycline. For protein extraction and solubilization, technical triplicates of cells were vigorously shaken at 95 °C with hot SDT buffer and centrifuged at high speed. Protein solutions were

loaded into a polyacrylamide gel and stained by Coomassie Brilliant Blue G-250 (CBB-G250) for sample quality control. A cut-off filter of 10 kDa was used for FASP sample processing, which includes protein reduction by dithiothreitol and alkylation by iodoacetamide to prevent disulfide bond formation following incubation of the samples in presence of trypsin at 37 °C for 18 hours. Extraction with ethyl acetate solvent was used for the removal of any potential SDS traces from the resulting peptide mixture. Liquid Chromatography with tandem mass spectrometry (LC-MS/MS) analysis of peptide mixture was performed using the Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) followed by the Orbitrap Q-Exactive HF X system (Thermo Fisher Scientific). The analytical column outlet was linked to the Digital PicoView 550 (New Objective) ion source, coupled with the Active Background Ion Reduction Device (ABIRD, ESI Source Solutions). MS data were acquired in a data-dependent strategy selecting up to the top 20 precursors.

MS data processing

Raw data obtained from MS were processed on MaxQuant (version 1.6.3.3) with Andromeda search engine utilization. Peptide sequences were annotated on the UniProtKB database (version 20180912, Human) and MaxQuant contamination databases (downloaded with the given version). Mass tolerances for peptides and MS/MS fragments were 4.5–10 ppm and 0.05 Da, respectively. Oxidation of methionine, deamidation (N, Q) and N-terminal acetylation were set as optional protein modifications, while carbamidomethylation (C) was set as fixed protein modification. Two enzyme miss-cleavages were permitted for the final annotation. Peptides and proteins with false discovery rate (FDR; q-value) < 1% were considered. The MaxQuant label-free quantification algorithm (MaxLFQ) was applied for global data normalization (minimal ratio count 1) and the MaxQuant protein group list was further analyzed via KNIME Analytics Platform (v.3.7.1). Results were deposited in the PRIDE Archive (https://www.ebi.ac.uk/pride/archive, accession number: PXD038393).

Construction of SCA1 PPI networks and enrichment analysis

Proteins were annotated in the HIPPIE database for the retrieval of high-confidence protein-protein interaction (PPI) scores (≥ 0.71). A PPI network was constructed in Cytoscape [46] and proteins were further clustered in functional communities using the GLay algorithm [47]. The layout of the network was designed in the Gephi platform [48]. For the prediction of upstream regulatory kinases, proteins were annotated in the X2K Appyters platform using the KEA3 database [49]. Predicted kinases were filtered with an overall score lower than 75. Over-representation analyses for KEGG biological pathways, human diseases (Jensen diseases), proteomics signatures (ProteomicsDB) and cell types and tissues (Descartes) were performed using the EnrichR package [50].

Drug repurposing and repositioning

Protein lists were uploaded in the L1000FWD platform [51] according to their pattern of dysregulation and hits were sorted based on their combined score. Mechanism of action and drug target identification were studied on the PHAROS, DrugBank and Reactome platforms. Evidence for drug safety and usage was obtained from the International Clinical Trials Registry Platform.

Results

Prediction of phosphorylation and dephosphorylation directed PPIs

To predict PTM-related directed PPIs (PTM-PPIs) we considered the entire dataset of human PPIs mapped in hyperbolic space (hPIN; see Methods for details). In this space, the angular

coordinates (theta) represent the similarity between the nodes in terms of interacting partners, and shorter distance to the center (r) corresponds to nodes with higher connectivity. The angular coordinate of the nodes in the hyperbolic plane reflects characteristics that make a node similar to the others. From a biological point of view, proteins agglomerate in the angular dimension of the H^2 capturing functional organization [15,17,52]. To investigate the biological meaning of the theta coordinates, we find proteins grouped in clusters by identifying gaps between consecutive inferred angles (S1 Fig, see Methods for details). This resulted in 24 clusters in the hPIN. The proteins are grouped in a similarity-based manner as each cluster is found to be enriched with various aspects of the GO biological process (Fig 2).

To predict PPIs as directed PTM-PPIs we collected a dataset of 295 experimentally supported interactions involving protein phosphorylation or dephosphorylation. We selected PPIs for which one of the interacting proteins is a putative effector (kinase or phosphatase) while the other is not (see Methods for details). We assume that this gives us a good estimate of the directionality of the interaction. Interestingly, the distribution of nodes corresponding to effectors and targets is different from that of the background proteome in the hPIN (Fig 2A). For example, effectors and targets appear to be depleted (less frequent than background) in clusters 1.13 and 1.14 associated with mitochondrial functions, while targets are enriched in cluster 1.1 associated with mRNA processing and effectors in cluster 1.7 associated with protein ubiquitination (Fig 2B). These results suggest that the hPIN provides information that could be used to discriminate effectors and targets of protein phosphorylation or dephosphorylation.

We chose 14 features to train a random forest (RF) model, six of them assigned to each of the two interacting nodes (r and theta coordinates in the hyperbolic map and four measurements of centrality), and two regarding the edge (hyperbolic distance and radial difference between the nodes) (See Methods for details).

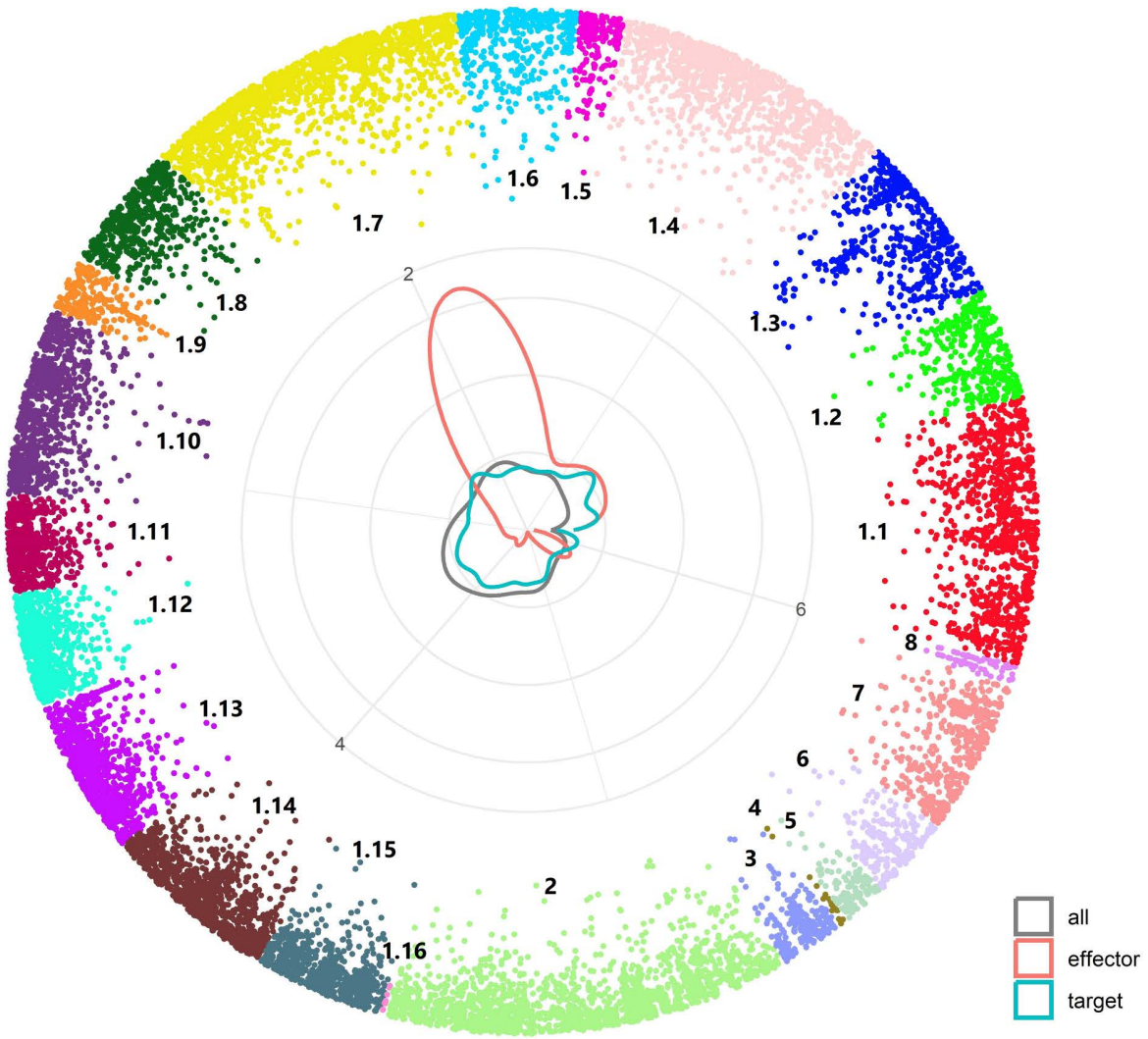
Regarding these features, we could appreciate significant differences in their distributions for effectors, targets and background (Fig 3A). Regarding hyperbolic coordinates, the effectors of the 295 positive directed-PPIs had in general shorter radius than the background proteins. This was also the case for targets, although with a less pronounced difference. This indicates that targets and effectors are more interconnected than other proteins, which agrees with their contribution to signaling pathways, stronger for the effectors. Regarding the angular dimension, both effectors and targets have maxima at a position different to the background, with effectors having a more pronounced grouping around theta = 1.8, which corresponds to cluster 1.7 (see also circular plot in Fig 2).

Regarding centrality measures, targets and effectors have higher values than background. For the distributions with maxima at zero values of EC, BC and DC, targets seem to have a larger number of low values than effectors. For the Gaussian distribution of CC, again effectors have slightly higher values than targets. Together with the observations of shorter radius this is in accordance to the higher connectivity of effectors and targets, with effectors slightly more connected and central than targets.

Regarding the edges, the radius differences between the connected nodes are higher in PTM-PPIs than background. This suggests that these interactions have a greater capacity to connect highly and lowly connected regions of the hPIN. The hyperbolic distances between nodes are slightly lower than for background PPIs. This could reflect that these PTM-PPIs participate in closely connected pathways and signaling networks.

To train the RF, we used 70% of the 295 interactions with 5-fold cross-validation, while the remaining 30% were used as the validation set. We performed the cross validation independently 10 times. The model with the highest accuracy was chosen as the final prediction model and it was validated on the test set, showing an accuracy of 74% (S2A Fig; see Methods

A



B

Cluster	GO BP Terms	Cluster	GO BP Terms
1.1	mRNA processing	1.13	mitochondrial translational elongation
1.2	regulation of gene silencing by miRNA	1.14	mitochondrial respiratory chain complex assembly
1.3	regulation of nucleobase-containing compound metabolic process	1.15	sphingolipid metabolic process
1.4	alpha-amino acid biosynthetic process	1.16	regulation of angiotensin levels in blood
1.5	cell cycle G1/S phase transition	2	regulation of cell cycle G2/M phase transition
1.6	post-translational protein modification	3	intraciliary retrograde transport
1.7	protein ubiquitination	4	N-terminal protein amino acid modification
1.8	positive regulation of establishment of protein localization to telomere	5	positive regulation of tyrosine phosphorylation of STAT protein
1.9	regulation of cellular amine metabolic process	6	positive regulation of lipopolysaccharide-mediated signaling pathway
1.10	metanephric renal vesicle morphogenesis	7	chemokine-mediated signaling pathway
1.11	protein transport	8	regulation of DNA-templated transcription, initiation
1.12	protein insertion into ER membrane by stop-transfer membrane-anchor sequence		

Fig 2. Properties of the hPIN. (A) Positions of nodes. Colors indicate node clusters. The circular plot at the center indicates the density of nodes in the angular dimension for all nodes (gray), and for the nodes of 295 directed PTM-PPIs with experimental evidence used for training a predictor: effectors (red) and targets (blue) (see text for details). (B) Top enriched GO BP term for each cluster.

https://doi.org/10.1371/journal.pone.0319084.g002

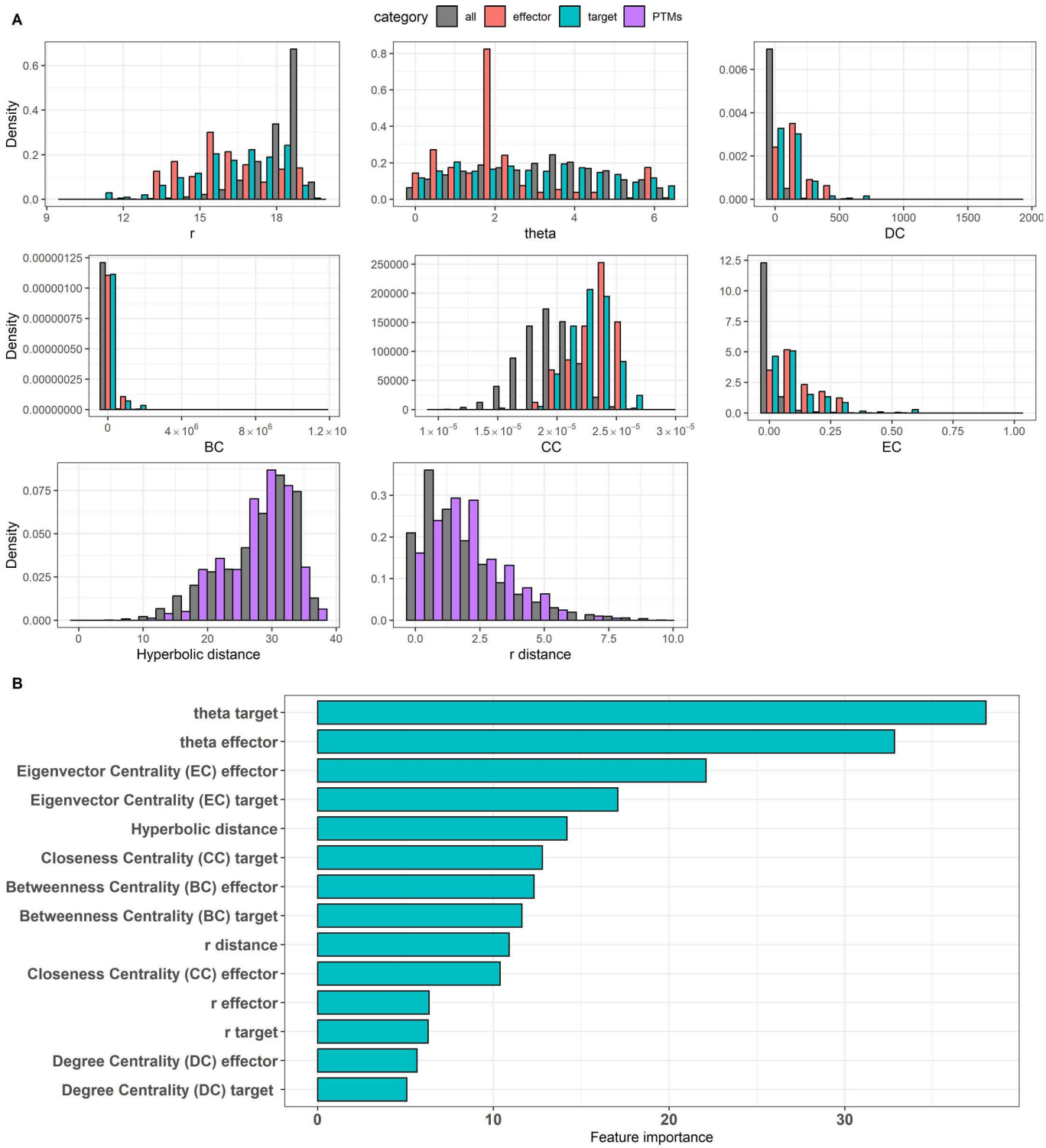


Fig 3. Properties of the 14 features that were used for building the model. (A) Distributions of r, theta, degree centrality (DC), betweenness centrality (BC), closeness centrality (CC) and eigenvector centrality (EC), hyperbolic distance and r distance. (B) Feature importance values indicate the impact of each predictor on the prediction model. For this study, the angular coordinates of the of effector and target proteins in the hyperbolic space are the most highly significant predictors of the model, confirming the importance of the embedding of the hPIN in H2.

https://doi.org/10.1371/journal.pone.0319084.g003

for details). A random forest (RF) classifier with 500 trees was able to produce satisfactory results. Finally, 74% sensitivity and 80% specificity were calculated from the confusion matrix. The receiver operating characteristic curve (ROC) had an AUC value of 0.87, indicating that the classifier could effectively find directed PTM-related PPIs based on topological and network properties of the interacting partners in the hPIN (S2B Fig). Additionally, the Precision-Recall curve (S2C Fig) was computed to assess the model's performance in the context of the imbalanced dataset, providing further insights into the classifier's ability to correctly identify the minority class (directed PTM-related PPIs).

Regarding the contribution of the features to the predictions, we observed that the angular coordinate of the target is the most important feature, closely followed by the angular coordinate of the effector (Fig 3B). The high relevance of the angular coordinates of effectors and targets is related to the fact that the angular positioning of the hyperbolic mapping captures the functional organization of the proteins in the cell.

EC of effector and target (which represents the importance of a node based on the links to important nodes) were the next features in order of importance. These were followed by the hyperbolic distance between the pairs of interacting proteins.

The next feature in order of importance was the CC of the target (which seemed to be much more important than that of the effector). This feature measures the central position of the node with respect to the entire network. The next features were the BC of the effector and of the target (representing how often a node is on paths between other nodes). We then have the r difference, the CC of the effector and, with much less importance, the r of the effector and of the target, suggesting that the distance to the center (which reflects the evolutionary age of the protein) is not very informative. Finally, the least important features are the DC values of the target and of the effector, which represent how well a node is directly connected to most nodes in the network.

It is interesting that the values of theta, and more marginally, the hyperbolic distance and the r distance between the nodes of PPIs were relevant features for the prediction. These results indicate that the embedding of the hPIN in hyperbolic space, which assigns these r and theta values to each node, can be useful to identify PTM-directed protein interactions. In particular, the contrast of the distributions of theta values of effectors and targets in the training dataset with the functions enriched in the corresponding clusters is revealing (see Fig 2A). While the maximum accumulation of effectors happens in a region of cluster 1.7 associated with the GO term "*protein ubiquitination*", a wider maximum happens for effectors and targets around clusters 1.1–1.3 enriched with terms "*mRNA processing*", "*regulation of gene silencing by miRNA*" and "*regulation of nucleobase-containing compound metabolic process*". The latest includes the synthesis of DNA and RNA. The distribution of targets is more similar to the background than that of the effectors, suggesting that PTM regulation targets all cell processes. Differently, effectors have a tendency to occupy tighter angular regions of the map, suggesting an association with regulatory mechanisms of control. The association with protein degradation (ubiquitination) seems to be a salient feature, which agrees with mechanisms known to stop active kinases [53]. These distributions of theta values have biological significance, which explains why they had the best predictive value.

Regarding the centrality measures, which are independent of the hyperbolic mapping, it can be seen that while EC, particularly of the effectors, plays an important role, DC does not seem to contribute so much. In any case, all features receive non-null values, suggesting that they all have predictive value.

To get a better understanding of how important are the features for the predictions of (de-) phosphorylation directed PPIs, we trained five RF models masking different features each time and we evaluated the overall contribution of the attributes in terms of information gain.

More specifically, we built the model on the 14 features and we determined the significance of each variable in the predictions using the *varImp ()* function from the random forest classifier. This method tracks the changes in model statistics for each predictor and accumulates the reduction in the statistic when each predictor's feature is added to the model. This total reduction is used as the variable importance measure. We conducted the prediction process while masking different features and we observed the changes in the performance metrics of the model. We created five datasets, starting with 14 variables and then we removed them one by one in order of their importance (Fig 3B). ROC curves analysis of the different datasets showed that using 14 features, AUC has the higher value; while removing them, the AUC is reducing (S3 Fig). This finding indicates the importance of the hyperbolic properties and centrality measures in classifying the directed (de-) phosphorylation related PPIs. The prediction model was applied to the entire set of edges. As we predict directional PTM-PPIs (with an effector and a target) all edges were tested in both directions for a total of n = 2 x 186,198 = 372,396 evaluations. The model produces a probability of the interaction being a directed PTM-PPI or not. A total of 117,655 directed interactions received a score>= 0.5 and 6,790 a score>= 0.9 (S4 Table).

The table of predictions allows easily to find the best predictions as target or effector for every protein in the network. For example, MAPK3 (MAP kinase-activated protein kinase 3; UniProt ID MAPK3_HUMAN) has a total of 9 edges; none of them were part of the experimentally verified set used in the training. Regardless, the predictions make sense: 8/9 have a score>= 0.5 for MAPK3 as effector (the best one is for HSPB1 as a target; score = 0.942), and there is only one with a score>= 0.5 for MAPK3 as target (with PRKY; PRKY_HUMAN), which is precisely the one with score < 0.5 for MAPK3 as effector. PRKY is a putative serine/threonine protein kinase with very little experimental information and its prediction as effector over MAPK3 is modest (score = 0.59), but the fact that it is annotated as protein kinase makes the prediction plausible.

We see that the classifier has trouble assigning the correct direction of the interaction. For example, the edge CHK2 (CHK2_HUMAN) RB (retinoblastoma; RB_HUMAN), which was used as positive for training, is highly evaluated with RB as target (score = 0.968), but also with RB as effector (score = 0.78). This indicates that the predictions need to be taken with care but also suggests that the scores can be compared.

To evaluate whether our predictions collectively are meaningful from a biological point of view, we performed a Gene Ontology (GO) enrichment analysis of proteins predicted even once as effectors (n = 12,115, Fig 4). Even considering that this is a large number of proteins, regarding GO Biological Process terms, these proteins are enriched in terms such as *"ubiquitin-dependent protein catalytic process"*, *"protein ubiquitination"*, *"protein phosphorylation"*, and "*cellular protein modifications*". Additionally, GO Molecular Function terms like *"protein serine/threonine kinase activity"*, *"protein kinase binding"* and *"kinase binding"* are also enriched. For comparison, we computed the enrichment for proteins predicted even once as not being effectors (n = 13,784; the two lists overlap in 10,314 proteins) and most of these terms received less significant p-values. This functional analysis supports the good performance of our prediction model.

Support of the predictions by other methods

To add support to the predictions of our model, we verified which of our predictions were detected by two alternative approaches: PhosD [43] and Phosformer-ST [44] (see Methods for details). PhosD is a kinase-substate prediction tool based on protein domains. Phosformer-ST is a machine learning tool that uses serine/threonine phosphorylation sites comprised of 15-mer peptides, assigning scores for these regions. A total of 788 and 535 predictions were

Fig 4. Enrichment analysis of predicted effectors. Left, negative predictions. Right, positive predictions. Colors from pink to grey indicate p-values from high to low; a lower p value suggests the proteins are more enriched in GO Biological Process and Molecular Function terms related to post-translational modifications. GO BP and MF terms are more enriched in the PTMs class, which supports the good performance of the model.

https://doi.org/10.1371/journal.pone.0319084.g004

supported by PhosD and Phosformer-ST, respectively, with an overlap of the three methods for 24 predictions. The detailed prediction results can be found in S4 Table.

Proteomics analysis highlights dysregulated biological pathways in a SCA1 cellular model

The results presented above suggest that the hyperbolic mapping of the hPIN provides a predictive value for direct PTM-PPIs. However, the interpretation of individual prediction scores remains complex. Therefore, we hypothesized that these predictive insights can collectively help us understand perturbations of the hPIN, which could be particularly valuable in identifying therapeutic mechanisms for human diseases. This is especially relevant for complex neurodegenerative diseases, in which the normal interactome is reportedly disrupted and abnormal PTMs can promote a plethora of pathological events. To test this hypothesis, we analyzed proteomics data from a SCA1 cellular model, in which the hPIN is perturbed due to the accumulation of inclusions of polyQ-expanded ataxin-1.

In particular, proteome alterations driven by the accumulation of mutant ataxin-1 were studied in Tet-On YFP-ATXN1(Q82) MSCs, a previously characterized cellular model of protein aggregation [45]. These cells contain insoluble intranuclear inclusions of polyQ-expanded ataxin-1 with a β-sheet conformation, an event that characterizes late-stage SCA1. Global proteome profiling was performed in inclusions-containing cells (SCA1, n = 3) and control cells (CTL, n = 3; see Methods for details). As a result, 3,926 proteins were identified and 3,179 of them were quantified in all six samples. The two conditions were efficiently discriminated

by principal component analysis, using as a criterion the variance in protein representation in each group (S4A Fig).

To create a representative protein network for SCA1 cellular pathology, we filtered 805 dysregulated proteins by a $|\log2\ FC| \geq 0.5$ and adj. p-value ≤ 0.05 (S4B-C Fig) and retrieved their high confidence interaction scores using the HIPPIE platform [26,27]. As a result, a complex PPI network of 636 significantly dysregulated proteins was generated, representing proteome alterations due to the accumulation of polyQ-expanded ataxin-1 inclusions (S5 Table).

The PPI network was further divided into smaller communities of densely interacting proteins, which outline functional modules (see Methods for details). Ataxin-1 was detected in the largest community (C1), which was highly associated with neurodegeneration and neuronal-related terms (S5A Fig). Enrichment analysis for biological pathways on the next 4 largest communities revealed a significant implication of spliceosome, lysosome and ribosome, as well as metabolic pathways (C2-C5, respectively) (S5A Fig). Interestingly, clustering and analysis of a control PPI network generated from a randomly selected protein dataset (n = 805) did not result in similar enrichment terms, indicating that the SCA1 PPI network and sub-communities are not generic but strongly associate with polyQ aggregation.

PTM-PPIs of ataxin-1 are components of the SCA1 network

To date, the effect of PTM-PPIs on the aggregation of mutant ataxin-1 remains unknown. Therefore, we sought to identify potential PTM-PPIs of ataxin-1 which may be involved in polyQ protein aggregation and eventually SCA1 pathogenesis. In the SCA1 PPI network, ataxin-1 directly interacted with 21 proteins. Implementation of our algorithm suggested that 13 of them may have a post-translational modification activity. Specifically, four of these proteins (gene names: *ANP32A*, *EIF3F*, *GSPT1* and *USP7*) were downregulated, while nine of them (gene names: *DNAJB6*, *HSPB1*, *PHPT1*, *SNCA*, *SQSTM1*, *SUMO1*, *TBL1XR1*, *TRIP6* and *TPM3*) were upregulated in SCA1 cells (S6 Table). The enzymatic activity of phosphohistidine phosphatase 1 (*PHPT1*), small ubiquitin-related modifier 1 (*SUMO1*), Ubiquitin-specific-processing protease 7 (*USP7*) and the chaperone proteins (*SNCA*, *HSPB1* and DNAJB6) have been previously described, while no such information exists for the rest (7/13) of the predicted proteins [54–57]. PTM-PPIs of ataxin-1 mainly clustered in community C1 (associated with neurodegeneration, eight proteins) and to a lesser extent in C2 (associated with spliceosome; four proteins) and C4 (associated with lysosome; one protein) (S5A Fig).

In an attempt to find a link among the three major communities containing the PTM-PPIs of ataxin-1 (C1, C2 and C4), we searched for potential common upstream regulators. To do so, the proteins of each cluster were considered substrates and were annotated using the KEA3 database for the prediction of regulatory kinases (see Methods for details). According to the results, 21 kinases were identified as potential common regulators for all three communities. Interestingly, three of them (*MAPK1*, *MAPK3* and *CDK4*) were indeed significantly dysregulated in SCA1 cells (S6 Table), suggesting their potential impact on regulating the C1, C2 and C4 communities. These kinases interact with various components of clusters C1 and C2 but not with proteins of C4, while none of them directly interacts with ataxin-1 (S5B Fig). Remarkably, no significantly dysregulated kinases were identified when repeating the analysis for a randomly sampled test PPI network, underscoring the specificity of the identified kinases to the SCA1-related network.

Identification and restoration of critical components of the SCA1 PPI network

SCA1-related cellular pathology might be driven by a few specific components scattered within the disease PPI network. To address this hypothesis, we generated a sub-network

consisting of ataxin-1, its predicted PTM-PPIs (n = 13) and their three common upstream kinases (MAPK1, MAPK3 and CDK4) (S6 Table; see Methods for details). Interestingly, all three identified kinases were connected to ataxin-1 through α-synuclein (*SNCA*), a protein associated with several neurodegenerative diseases and particularly Parkinson's disease (Fig 5A). Enrichment analysis for rare diseases (see Methods for details) indicated that this sub-network is associated with cerebellar degeneration terms, including Spinocerebellar Ataxia and the formation of nuclear inclusion bodies. This result suggests that these proteins are critical for the disease and their dysregulation may underlie SCA1-related pathological events (Fig 5B).

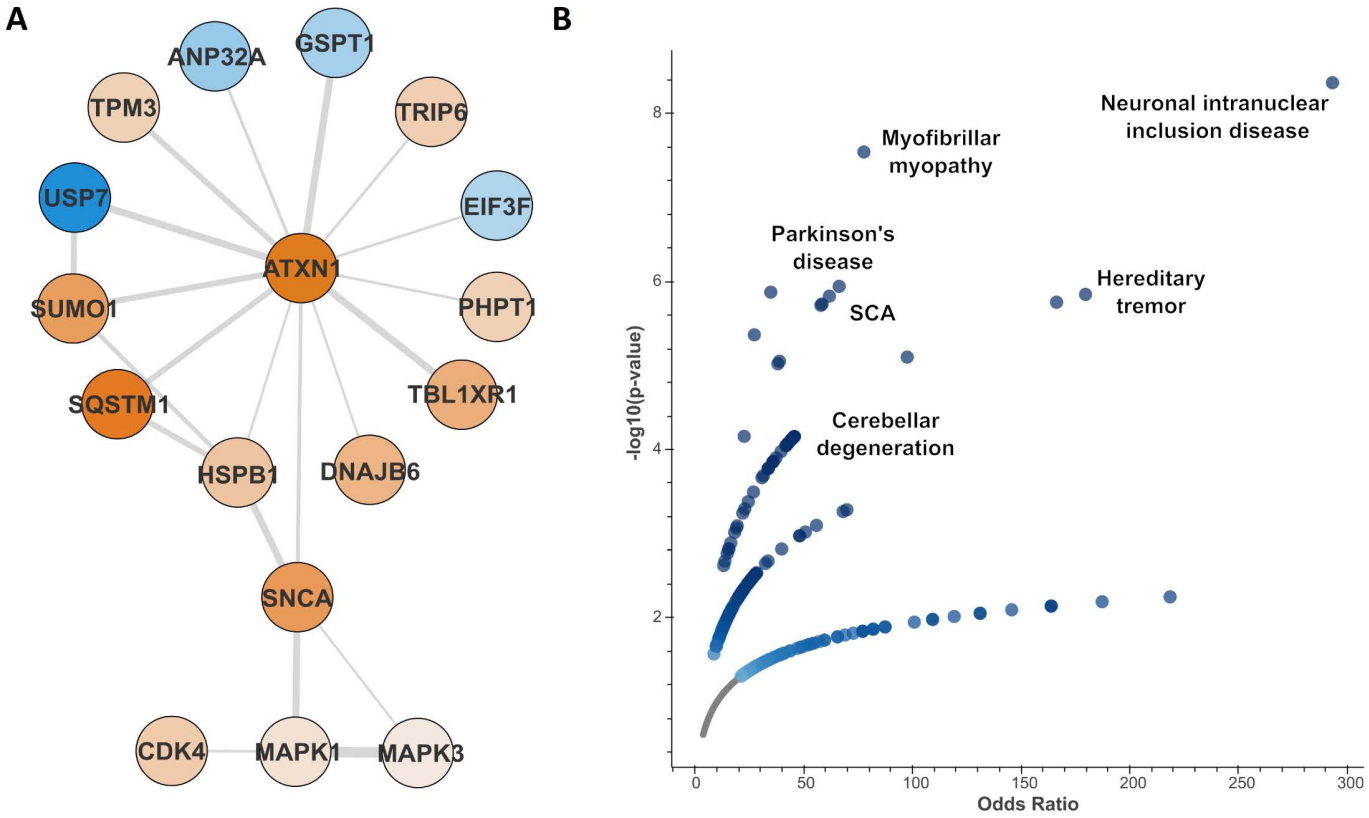
Therefore, restoration of their dysregulation pattern might mitigate disease progression. To this end, we searched for candidate drugs potent to increase the levels of downregulated proteins and decrease those of upregulated ones. Their reverse score indicated the overlap between the input proteins and the altered signature after drug administration. Hits with at least a 25% reverse score were considered significant candidates (see Methods for details). Then, they were sorted by descending combined score, considering the reverse score, p-value and Z-score. From this analysis, we identified four known drugs (artesunate, linifanib, budesonide and betamethasone) and three novel compounds (BRD-K54687541, BRD-K71265179 and BRD-A08662020) as potential treatment approaches (Fig 5C). These agents might mitigate polyQ-expanded ataxin-1-associated neuropathology in SCA1 cells potentially leading to the development of novel therapeutic strategies against the disease.

Discussion

Machine learning has shown good performance in extracting rules from massive biological data. Here we present a computational method that implements machine learning based on the random forest algorithm and trains a model to predict directed PTM-PPIs, concretely, phosphorylation and dephosphorylation interactions between a target and an effector. Several lines of work approach PPI prediction through various computational methods [58–64] but currently not much research has been performed on predicting the function of PPIs. The representation of the human protein interaction network in the two-dimensional hyperbolic plane has been shown to be both meaningful and useful: inferred node coordinates uncover information about protein evolution and function, whereas hyperbolic distances can be used to identify potential protein interactions [15]. In this study, we report another scenario, in which hyperbolic properties together with metrics from network analysis are used to predict directed PTM-PPIs.

The result that the angular (theta) coordinates of targets and effectors were the more predictive features is of particular relevance, considering that they are superior to network measures that are not depending on the hyperbolic mapping. The fact that theta of target is more predictive that than that of the effector is consistent with targets being responsible of narrower functions (signaling, cell cycle control, cell differentiation), while effectors, with a more upstream position on a regulatory network, could be expected to have more general functions, and therefore, less restrictions to take various angular positions in the hyperbolic map as core components of signaling cascades [65].

Our results evaluate various centrality measures suggesting that while all of them have predictive value, degree centrality may be the less informative compared to eigenvector, closeness and betweenness, at least, when evaluating phosphorylation and dephosphorylation. The complete list of predicted effectors turned out to be explainable from a biological point of view according to functional enrichment analysis.



C

Drug candidate	Reverse score	Administration
artesunate	25%	Malaria
BRD-K54687541	25%	Unknown
betamethasone	25%	cerebral edema, congenital adrenal hyperplasia
budesonide	25%	Crohn's disease, asthma, allergies
BRD-K71265179	25%	Unknown
Linifanib	25%	hematologic malignancies, solid tumors
BRD-A08662020	25%	Unknown

Fig 5. Critical proteins associated with SCA1 pathology. (A) A PPI network consisting of ataxin-1, its 13 PTM-PPIs and the 3 dysregulated kinases (n = 17). Node color corresponds to fold change (blue for down- and orange for upregulated proteins, respectively) and edge width to HIPPIE interaction score [24]. **(B)** Enrichment analysis for rare diseases highlighted an association of the critical proteins with SCA-related terms. **(C)** Known drugs which may restore the levels of at least 25% of the dysregulated proteins, ranked according to their combined score (p-value, adj. p-value, z-score).

https://doi.org/10.1371/journal.pone.0319084.g005

To illustrate how to use the collective predictions for studying disease progression, we predicted PTM-PPIs in a SCA1 disease model with intense hPIN perturbation. The cell model employed here recapitulates key pathological features of SCA1, one of several polyQ diseases that are caused by expanded CAG repeats encoding a long polyQ tract in the respective proteins and lead to neurodegeneration [23,45,66].

Two possible factors contributing to selective neuronal impairment are the abnormal subcellular localization of polyQ proteins and the change of folding and function [67]. PTMs are shown to regulate properties including their intracellular localization and functions [68]. Consequently, understanding the effect of PTMs in polyQ diseases may yield important insight into mechanisms behind neuronal damage and more specifically in SCA1. We identified proteins that could be at the center of dysregulated phosphorylation networks and generated a disease-specific PPI network enriched with our PTM-PPI predictions. Among them, we identified α-synuclein (SNCA), a protein that translocates between the cytoplasm and the nucleus. SNCA may also function as a chaperone as it shares physical and functional homology with the 14-3-3 protein family which are responsible for ataxin-1 translocation from the cytoplasm to the nucleus [69–72]. Interestingly, SNCA is also involved in the pathogenesis of Parkinson's disease and may impart its effect in line with the key upstream kinases of the SCA1 disease network [73].

To evaluate the translational impact of the predicted outcome, we searched for existing drugs that could revert the production of these proteins as potential therapeutics against the disease. By implementing a network-based drug repurposing analysis, we identified 7 potential drug candidates. Artesunate was the top hit, acting as a protein synthesis inhibitor and a glucocorticoid receptor agonist. We have previously shown that the accumulation of mutant ataxin-1 disrupts ribosome assembly and causes proteome instability [45,74]. Therefore, regulation of translation may have a therapeutic effect in SCA1 cells. Interestingly, artesunate is currently evaluated as a therapeutic agent for Friedreich's ataxia (FA), suggesting that it might be also relevant for the treatment of other similar disorders, including SCA1 [75]. Furthermore, the predicted drug candidates betamethasone and budesonide also act as glucocorticoid receptor agonists. Although there is no direct evidence for SCA1, activation of glucocorticoid receptors seems to attenuate the aggregation of polyQ-expanded ataxin-3 and huntingtin proteins in SCA3 and Huntington's disease (HD), respectively [76,77].

Our work supports the value of the hPIN and of their hyperbolic mapping for the prediction of the function of directed PTM-PPIs. The method was limited to detect phosphorylation and dephosphorylation, the most common PTMs, as directed interactions between a regulatory protein and its target; more complex interactions can be expected since regulatory proteins are often multi-domain proteins with a multiplicity of sites for their own regulation: our approach cannot be expected to capture those without an appropriate training dataset. In addition, we focused on phosphorylation and dephosphorylation, without making a distinction between them and, most importantly, without considering other less frequent but relevant PTMs such as ubiquitination, methylation or acetylation. Even within these limitations, we were able to apply our predictions to provide a proteome-wide set of scored interactions that we used to suggest therapeutic actions against a neurodegenerative disease. Our predictions should find applicability in combination with many other experimental and computational datasets.

Supporting information

S1 Fig. Identification of big gaps between inferred protein angles. Proteins were sorted increasingly by their inferred angular coordinates θ and the difference between θi and θi + 1 was computed. The peaks correspond to gap sizes in the angular dimension and hint at the presence of similarity-based clusters. To determine the beginning and end of each cluster in the hPIN, we chose the gap size (g = 0.0077, line in red color) that produced clusters with a minimum of three components. The same process was followed to subcluster the first sector into 15 smaller clusters using a smaller gap size (g = 0.042, line in blue), This allowed us to perform meaningful enrichment analysis of each group of proteins.
(JPG)

S2 Fig. Evaluation of the model. (**A**) Accuracy for all the models after 5- fold cross validation repeated 10 times. (**B**) The ROC of the model confirms a satisfactory classification performance. (**C**) Precision-Recall curve, providing additional performance evaluation.
(JPG)

S3 Fig. ROC curves and AUC scores to compare the classifier performance of the different data sets containing various number of features. In total 14 features related to hyperbolic properties and centrality measures were used to predict phosphorylation and dephosphorylation directed PPIs.
(JPG)

S4 Fig. Profile of dysregulated proteins in SCA1 cells containing polyQ inclusions. (A) SCA1 cells were efficiently discriminated from control cells (CTL) using PCA. **(B)** Volcano plot depicting 449 significantly downregulated proteins (blue color) and 356 significantly upregulated proteins (red color) in SCA1 cells [selection criteria (log2FC ≤ |0.5|, adj. p-value ≤ 0.05)]. The top 10 dysregulated proteins are highlighted in the plot. **(C)** Heatmap plot according to Euclidean distance indicates two distinct groups of up- and down-regulated proteins (red and blue color, respectively) in SCA1 and control cells.
(JPG)

S5 Fig. Connectivity of dysregulated proteins in SCA1 cells. (A) PPI network of significantly dysregulated proteins in SCA1 cells. Proteins were clustered into dense communities representing functional modules. Enrichment analysis on each cluster indicated a strong association with neurodegeneration (C1), spliceosomal (C2) and lysosomal (C4) activity, ribosome assembly (C3) and metabolic pathways (C5). Ataxin-1 directly interacts with 21 proteins, 13 of which are predicted as PTM-PPIs and participate in C1, C2 and C4. Both PTM and non-PTM PPIs of ataxin-1 are highlighted with red and yellow color, respectively. **(B)** Identification of regulatory kinases for C1, C2 and C4 clusters, which contain the PTM-PPIs of ataxin-1. MAPK1, MAPK3 and CDK4 are significantly dysregulated in SCA1 cells.
(JPG)

S1 Table. Nodes of the hPIN. Columns indicate protein identifiers (UniProtKB), hyperbolic coordinates (r, theta), and centralities (Degree DC, Betweenness BC, Closeness CC and Eigenvector EC).
(XLSX)

S2 Table. Edges of the hPIN. Columns indicate protein identifiers (UniProtKB; p1, p2), hyperbolic distance, r difference.
(XLSX)

S3 Table. Training dataset of experimentally known phosphorylation and dephosphorylation PPIs. Columns indicate effector protein identifier (UniProtKB; p1), effector type, and target protein identifier (UniProtKB; p2).
(XLSX)

S4 Table. Prediction scores of directed PTM-PPIs. Columns indicate predicted effector and target protein identifiers (UniProtKB; p1, p2), score of the prediction of our method and classification of our method, PhosD and Phosformer-ST.
(XLSX)

S5 Table. Significantly dysregulated proteins (|log2FC| ≤ 0.5, adj. p-value ≤ 0.05) in SCA1 (805 proteins). Columns indicate protein identifier (UniProtKB ID and gene name), log2FC value, p-value, adjusted p-value and cluster number of the proteins SCA1 PPI network: values

are (i) C1-C5 or unclustered for the 636 strongly connected proteins or (ii) blank for the remaining 169 less connected proteins.
(XLSX)

S6 Table. Components of a critical PPI network involved in SCA1 pathogenesis. Columns indicate protein identifiers (UniProtKB ID and gene name), log2FC value and adjusted p-value.
(XLSX)

Author contributions

Conceptualization: Miguel A. Andrade-Navarro.

Data curation: Aimilia-Christina Vagiona.

Formal analysis: Mariane Gonçalves-Kulik.

Methodology: Aimilia-Christina Vagiona.

Supervision: Miguel A. Andrade-Navarro.

Validation: Spyros Petrakis.

Visualization: Aimilia-Christina Vagiona, Sofia Notopoulou.

Writing – original draft: Aimilia-Christina Vagiona, Sofia Notopoulou, Zbyněk Zdráhal, Mariane Gonçalves-Kulik, Spyros Petrakis, Miguel A. Andrade-Navarro.

Writing – review & editing: Zbyněk Zdráhal, Mariane Gonçalves-Kulik, Spyros Petrakis, Miguel A. Andrade-Navarro.

References

1. Rual J-F, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;437(7062):1173–8. https://doi.org/10.1038/nature04209 PMID: 16189514
2. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell. 2005;122(6):957–68. https://doi.org/10.1016/j.cell.2005.08.029 PMID: 16169070
3. Koh G, Porras P, Aranda B, Hermjakob H, Orchard S. Analyzing protein–protein interaction networks. J Proteome Res. 2012;11(4):2014–31.
4. Arkin MR, Whitty A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. Curr Opin Chem Biol. 2009;13(3):284–90. https://doi.org/10.1016/j.cbpa.2009.05.125 PMID: 19553156
5. Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell. 2015;163(3):712–23. https://doi.org/10.1016/j.cell.2015.09.053 PMID: 26496610
6. De Las Rivas J, Fontanillo C. Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell. Brief Funct Genomics. 2012;11(6):489–96. https://doi.org/10.1093/bfgp/els036 PMID: 22908212
7. Kanhaiya K, Czeizler E, Gratie C, Petre I. Controlling directed protein interaction networks in cancer. Sci Rep. 2017;7(1):10327. https://doi.org/10.1038/s41598-017-10491-y PMID: 28871116
8. Rosell M, Fernández-Recio J. Hot-spot analysis for drug discovery targeting protein-protein interactions. Expert Opin Drug Discov. 2018;13(4):327–38. https://doi.org/10.1080/17460441.2018.1430763 PMID: 29376444
9. Macalino S, Basith S, Clavio N, Chang H, Kang S, Choi S. Evolution of in silico strategies for protein-protein interaction drug discovery. Molecules. 2018;23(8):1963.
10. Scott DE, Bayly AR, Abell C, Skidmore J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat Rev Drug Discov. 2016;15(8):533–50. https://doi.org/10.1038/nrd.2016.29 PMID: 27050677
11. Cannistraci CV, Alanis-Lobato G, Ravasi T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. Bioinformatics. 2013;29(13):i199–209. https://doi.org/10.1093/bioinformatics/btt208 PMID: 23812985

12. Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguñá M. Hyperbolic geometry of complex networks. Physical Review E. 2010;82(3):036106. https://doi.org/10.1103/PhysRevE.82.036106
13. Serrano MA, Boguna M, Sagues F. Uncovering the hidden geometry behind metabolic networks [Internet]. arXiv; 2011 [cited 2022 Oct 31]. Available from: http://arxiv.org/abs/1109.1934
14. Papadopoulos F, Kitsak M, Serrano MÁ, Boguñá M, Krioukov D. Popularity versus similarity in growing networks. Nature. 2012;489(7417):537–40. https://doi.org/10.1038/nature11459 PMID: 22972194
15. Alanis-Lobato G, Mier P, Andrade-Navarro M. The latent geometry of the human protein interaction network. Bioinformatics. 2018;34(16):2826–34.
16. Härtner F, Andrade-Navarro MA, Alanis-Lobato G. Geometric characterisation of disease modules. Appl Netw Sci. 2018;3(1):10. https://doi.org/10.1007/s41109-018-0066-3 PMID: 30839777
17. Vagiona A-C, Mier P, Petrakis S, Andrade-Navarro MA. Analysis of Huntington's Disease modifiers using the hyperbolic mapping of the protein interaction network. Int J Mol Sci. 2022;23(10):5853. https://doi.org/10.3390/ijms23105853 PMID: 35628660
18. Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. Nat Methods. 2013;10(8):690–1. https://doi.org/10.1038/nmeth.2561 PMID: 23900247
19. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, et al. MINT: the Molecular INTeraction database. Nucleic Acids Res. 2007;35(Database issue):D572–4. https://doi.org/10.1093/nar/gkl950 PMID: 17135203
20. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research. 2011;39(Database):D691–7.
21. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, Willems A, et al. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. 2021;30(1):187–200. https://doi.org/10.1002/pro.3978 PMID: 33070389
22. Xu H, Wang Y, Lin S, Deng W, Peng D, Cui Q, et al. PTMD: a database of human disease-associated post-translational modifications. Genomics Proteomics Bioinformatics. 2018;16(4):244–51. https://doi.org/10.1016/j.gpb.2018.06.004 PMID: 30244175
23. Buijsen R, Hu M, Sáez-González M, Notopoulou S, Mina E, Koning W, et al. Spinocerebellar ataxia type 1 characteristics in patient-derived fibroblast and iPSC-derived neuronal cultures. Movement Disorders. 2023. https://doi.org/10.1002/mds.29446
24. Ju H, Kokubu H, Lim J. Beyond the glutamine expansion: influence of posttranslational modifications of ataxin-1 in the pathogenesis of spinocerebellar ataxia type 1. Mol Neurobiol. 2014;50(3):866–74. https://doi.org/10.1007/s12035-014-8703-z PMID: 24752589
25. Nolfi-Donegan D, Braganza A, Shiva S. Mitochondrial electron transport chain: oxidative phosphorylation, oxidant production, and methods of measurement. Redox Biol. 2020;37:101674. https://doi.org/10.1016/j.redox.2020.101674 PMID: 32811789
26. Alanis-Lobato G, Andrade-Navarro M, Schaefer M. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. Nucleic Acids Research. 2017;45(D1):D408–14.
27. Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: Integrating protein interaction networks with experiment based quality scores. PLoS One. 2012;7(2):e31826. https://doi.org/10.1371/journal.pone.0031826 PMID: 22348130
28. Alanis-Lobato G, Mier P, Andrade-Navarro MA. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. Appl Netw Sci. 2016;1(1):10. https://doi.org/10.1007/s41109-016-0013-0 PMID: 30533502
29. Alanis-Lobato G, Mier P, Andrade-Navarro MA. Efficient embedding of complex networks to hyperbolic space via their Laplacian. Sci Rep. 2016;6:30108. https://doi.org/10.1038/srep30108 PMID: 27445157
30. Papadopoulos F, Aldecoa R, Krioukov D. Network geometry inference using common neighbors. Phys Rev E Stat Nonlin Soft Matter Phys. 2015;92(2):022807. https://doi.org/10.1103/PhysRevE.92.022807 PMID: 26382454
31. del-Toro N, Dumousseau M, Orchard S, Jimenez RC, Galeota E, Launay G, et al. A new reference implementation of the PSICQUIC web service. Nucleic Acids Res. 2013;41(Web Server issue):W601–6. https://doi.org/10.1093/nar/gkt392 PMID: 23671334
32. Hu R, Xu H, Jia P, Zhao Z. KinaseMD: kinase mutations and drug response database. Nucleic Acids Research. 2021;49(D1):D552–61.
33. Duan G, Li X, Köhn M. The human DEPhOsphorylation database DEPOD: a 2015 update. Nucleic Acids Research. 2015;43(D1):D531–5.
34. Borgatti SP. Centrality and network flow. Social Networks. 2005;27(1):55–71. https://doi.org/10.1016/j.socnet.2004.11.008

35. Freeman LC. Centrality in social networks conceptual clarification. Social Networks. 1978;1(3):215–39. https://doi.org/10.1016/0378-8733(78)90021-7

36. Freeman LC. A Set of measures of centrality based on betweenness. Sociometry. 1977;40(1):35. https://doi.org/10.2307/3033543

37. Sabidussi G. The centrality of a graph. Psychometrika. 1966;31(4):581–603. https://doi.org/10.1007/BF02289527 PMID: 5232444

38. Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks. 2001;23(3):191–201. https://doi.org/10.1016/s0378-8733(01)00038-7

39. Kuhn M. Building predictive models in R using the caret package. J Stat Soft. 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05

40. Svetnik V, Liaw A, Tong C, Culberson J, Sheridan R, Feuston B. Random forest: A classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58.

41. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition. 1997;30(7):1145–59. https://doi.org/10.1016/s0031-3203(96)00142-2

42. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Global Ecol Biogeography. 2007;17(2):145–51. https://doi.org/10.1111/j.1466-8238.2007.00358.x

43. Qin G, Li R, Zhao X. PhosD: inferring kinase–substrate interactions based on protein domains. Bioinformatics. 2017;33(8):1197–204.

44. Zhou Z, Yeung W, Soleymani S, Gravel N, Salcedo M, Li S, et al. Using explainable machine learning to uncover the kinase-substrate interaction landscape. Bioinformatics. 2024 Jan 19;btae033.

45. Laidou S, Alanis-Lobato G, Pribyl J, Raskó T, Tichy B, Mikulasek K, et al. Nuclear inclusions of pathogenic ataxin-1 induce oxidative stress and perturb the protein synthesis machinery. Redox Biol. 2020;32101458. https://doi.org/10.1016/j.redox.2020.101458 PMID: 32145456

46. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

47. Su G, Kuchinsky A, Morris J, States D, Meng F. GLay: community structure analysis of biological networks. Bioinformatics. 2010;26(24):3135–7.

48. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and manipulating networks. Proceedings of the International Conference on Weblogs and Social Media (ICWSM). 2009;3(1):361–2.

49. Kuleshov M, Xie Z, London A, Yang J, Evangelista J, Lachmann A. KEA3: improved kinase enrichment analysis via data integration. Nucleic Acids Research. 2021;49(W1):W304–16.

50. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128. https://doi.org/10.1186/1471-2105-14-128 PMID: 23586463

51. Wang Z, Lachmann A, Keenan A, Ma'ayan A. L1000FWD: fireworks visualization of drug-induced transcriptomic signatures. Bioinformatics. 2018;34(12):2150–2.

52. Zahra NUA, Vagiona AC, Uddin R, Andrade-Navarro MA. Selection of multi-drug targets against drug-resistant Mycobacterium tuberculosis XDR1219 using the hyperbolic mapping of the protein interaction network. IJMS. 2023;24(18):14050.

53. Lu Z, Hunter T. Degradation of activated protein kinases by ubiquitination. Annu Rev Biochem. 2009;78435–75. https://doi.org/10.1146/annurev.biochem.013008.092711 PMID: 19489726

54. Hoffman L, Jensen C, Beckerle M. Phosphorylation of the small heat shock protein HspB1 regulates cytoskeletal recruitment and cell motility. Molecular Biology of the Cell. 2022;33(11):ar100.

55. Månsson C, Kakkar V, Monsellier E, Sourigues Y, Härmark J, Kampinga HH, et al. DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. Cell Stress Chaperones. 2014;19(2):227–39. https://doi.org/10.1007/s12192-013-0448-5 PMID: 23904097

56. Martin DR, Dutta P, Mahajan S, Varma S, Stevens SM. Structural and activity characterization of human PHPT1 after oxidative modification. Sci Rep. 2016 Apr 1;6(1):23658.

57. Matsuzaki S, Lee L, Knock E, Srikumar T, Sakurai M, Hazrati L. SUMO1 affects synaptic function, spine density and memory. Sci Rep. 2015;5(1):10730. https://doi.org/10.1038/srep10730

58. Luo X, Wang L, Hu P, Hu L. Predicting protein-protein interactions using sequence and network information via variational graph autoencoder. IEEE/ACM Trans Comput Biol Bioinform. 2023;20(5):3182–94. https://doi.org/10.1109/TCBB.2023.3273567 PMID: 37155405

59. Li X, Han P, Chen W, Gao C, Wang S, Song T, et al. MARPPI: boosting prediction of protein-protein interactions with multi-scale architecture residual network. Brief Bioinform. 2023;24(1):bbac524. https://doi.org/10.1093/bib/bbac524 PMID: 36502435

60. Wang X, Yang W, Yang Y, He Y, Zhang J, Wang L, et al. PPISB: a novel network-based algorithm of predicting protein-protein interactions with mixed membership stochastic blockmodel. IEEE/ACM Trans Comput Biol Bioinform. 2023;20(2):1606–12. https://doi.org/10.1109/TCBB.2022.3196336 PMID: 35939453

61. Su X-R, Hu L, You Z-H, Hu P-W, Zhao B-W. Multi-view heterogeneous molecular network representation learning for protein-protein interaction prediction. BMC Bioinformatics. 2022;23(1):234. https://doi.org/10.1186/s12859-022-04766-z PMID: 35710342

62. Chen C, Zhang Q, Yu B, Yu Z, Lawrence PJ, Ma Q, et al. Improving protein-protein interactions prediction accuracy using XGBoost feature selection and stacked ensemble classifier. Comput Biol Med. 2020;123:103899. https://doi.org/10.1016/j.compbiomed.2020.103899 PMID: 32768046

63. Hashemifar S, Neyshabur B, Khan A, Xu J. Predicting protein–protein interactions through sequence-based deep learning. Bioinformatics. 2018;34(17):i802–10.

64. Zhang L, Yu G, Xia D, Wang J. Protein–protein interactions prediction based on ensemble deep neural networks. Neurocomputing. 2019;324:10–9. https://doi.org/10.1016/j.neucom.2018.02.097

65. Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, et al. A directed protein interaction network for investigating intracellular signal transduction. Sci Signal. 2011;4(189):rs8. https://doi.org/10.1126/scisignal.2001699 PMID: 21900206

66. Tejwani L, Lim J. Pathogenic mechanisms underlying spinocerebellar ataxia type 1. Cell Mol Life Sci. 2020;77(20):4015–29. https://doi.org/10.1007/s00018-020-03520-z PMID: 32306062

67. Sambataro F, Pennuto M. Cell-autonomous and non-cell-autonomous toxicity in polyglutamine diseases. Prog Neurobiol. 2012;97(2):152–72. https://doi.org/10.1016/j.pneurobio.2011.10.003 PMID: 22061202

68. Sambataro F, Pennuto M. Post-translational modifications and protein quality control in motor neuron and Polyglutamine diseases. Front Mol Neurosci [Internet]. [cited 2022 Nov 11] 2017 Mar 31;10. Available from: http://journal.frontiersin.org/article/10.3389/fnmol.2017.00082/full

69. Rekas A, Ahn KJ, Kim J, Carver JA. The chaperone activity of α-synuclein: Utilizing deletion mutants to map its interaction with target proteins. Proteins. 2012;80(5):1316–25. https://doi.org/10.1002/prot.24028 PMID: 22274962

70. Ostrerova N, Petrucelli L, Farrer M, Mehta N, Choi P, Hardy J, et al. alpha-Synuclein shares physical and functional homology with 14-3-3 proteins. J Neurosci. 1999;19(14):5782–91. https://doi.org/10.1523/JNEUROSCI.19-14-05782.1999 PMID: 10407019

71. Shirasaka M, Kuwata K, Honda R. α-Synuclein chaperone suppresses nucleation and amyloidogenesis of prion protein. Biochem Biophys Res Commun. 2020;521(1):259–64. https://doi.org/10.1016/j.bbrc.2019.10.120 PMID: 31635805

72. Lai S, O'Callaghan B, Zoghbi H, Orr H. 14-3-3 binding to ataxin-1(ATXN1) regulates its dephosphorylation at Ser-776 and transport to the nucleus. J Biol Chem. 2011;286(40):34606–16.

73. Siddiqui I, Pervaiz N, Abbasi A. The Parkinson disease gene SNCA: Evolutionary and structural insights with pathological implication. Scientific Reports. 2016;6:24475.

74. Gkekas I, Vagiona A-C, Pechlivanis N, Kastrinaki G, Pliatsika K, Iben S, et al. Intranuclear inclusions of polyQ-expanded ATXN1 sequester RNA molecules. Front Mol Neurosci. 2023;16:1280546. https://doi.org/10.3389/fnmol.2023.1280546 PMID: 38125008

75. Rufini A, Malisan F, Condò I, Testi R. Drug repositioning in friedreich ataxia. Front Neurosci. 2022;16:814445. https://doi.org/10.3389/fnins.2022.814445 PMID: 35221903

76. Diamond M, Robinson M, Yamamoto K. Regulation of expanded polyglutamine protein aggregation and nuclear localization by the glucocorticoid receptor. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(2):657–61. https://doi.org/10.1073/pnas.97.2.657

77. Duarte-Silva S, Da Silva JD, Monteiro-Fernandes D, Costa MD, Neves-Carvalho A, Raposo M, et al. Glucocorticoid receptor-dependent therapeutic efficacy of tauroursodeoxycholic acid in preclinical models of spinocerebellar ataxia type 3. J Clin Invest. 2024;134(5):e162246. https://doi.org/10.1172/JCI162246 PMID: 38227368

Chapter 3

Unraveling Cooperative and Competitive Interactions in Protein Triplet Complexes in the Human Proteome

Aimilia-Christina Vagiona, Pablo Mier and Miguel A. Andrade-Navarro

Article published in *Scientific Reports, vol. 15, 1, 32548. (15 September 2025),*

doi: https://doi.org/10.1038/s41598-025-19264-4

scientific reports

Check for updates

OPEN

Unraveling cooperative and competitive interactions within protein triplets in the human interactome

Aimilia-Christina Vagiona[1], Pablo Mier[2] & Miguel A. Andrade-Navarro[1]✉

Knowledge of protein–protein interactions (PPIs) is essential for understanding cellular function, yet most network analyses focus on binary interactions. Higher-order motifs such as protein triplets can reveal cooperative or competitive relationships but are difficult to distinguish systematically. We present a computational framework to classify protein triplets in the human protein interaction network (hPIN) as cooperative or competitive. The hPIN was embedded in hyperbolic space using the LaBNE + HM algorithm, and a Random Forest classifier was trained on structurally validated triplets from Interactome3D, achieving high accuracy (AUC = 0.88). Angular and hyperbolic distances were key predictive features. Predicted cooperative triplets were enriched in paralogous partners, indicating that paralogs often bind together to a shared protein using non-overlapping surfaces. The model proved to be effective when tested on a new dataset. AlphaFold 3 modeling supported these predictions, showing that cooperative partners bind at distinct sites, while competitive ones overlap. Our results demonstrate the value of hyperbolic geometry for capturing functional organization in protein complexes.

Proteins are the essential units of life, regulating crucial biological processes ranging from molecular transport to signal transduction[1]. These functions depend on a highly coordinated network of protein–protein interactions (PPIs), collectively referred to as the interactome, which controls cellular organization and functionality[2,3]. Understanding these interactions is key to unraveling biological mechanisms and designing therapeutic strategies for various diseases[4,5].

Protein–protein interaction networks (PPINs) are primary constructed from binary interactions, which represent direct physical contact between two proteins[6]. Most large-scale interactome maps rely on pairwise interaction data, as these are the most accessible and well-characterized through traditional experimental approaches[7]. High-throughput techniques such as yeast two-hybrid (Y2H) assays and affinity purification coupled with mass spectrometry (AP-MS) have been essential in mapping the human interactome[8,9]. However, many biological processes extend beyond pairwise PPIs, as proteins often function in multi-protein assemblies rather than simple one-to-one interactions, in order to carry out essential tasks[10,11]. Among these multi-protein complexes, protein triplets represent a crucial class of higher-order interactions. Triplet interactions provide a framework for understanding cooperative and competitive dynamics, which influence the structural and functional stability of protein complexes[12,13].

Cooperative interactions occur when multiple proteins work together synergistically to enhance stability or function, such as in multiprotein enzyme complexes or transcription factor binding events[14,15]. In contrast, competitive interactions arise when two proteins compete for the same binding partner, modulating signaling pathways or enzymatic activity based on cellular conditions[16]. Accordingly, proteins interacting with many partners can have either multiple interfaces or just one interface. Some of these interfaces are shared by different partners, resulting in mutually exclusive bindings; other interfaces are used by only one partner such that interactions with different partners can occur simultaneously. Thus, to judge whether two partners can interact with the common protein simultaneously, the key is to know whether they share an interaction interface[17].

Despite their importance, higher-order interactions remain difficult to study using traditional experimental methods, which are primarily designed to detect pairwise interactions. This limitation highlights the need

[1]Insitute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University, Biozentrum I, Hans-Dieter-Hüsch-Weg 15, 55128 Mainz, Germany. [2]Andalusian Center for Developmental Biology (CABD, UPO-CSIC-JA), Faculty of Experimental Sciences (Genetics Area), Universidad Pablo de Olavide, 41013 Seville, Spain. ✉email: andrade@uni-mainz.de

for computational approaches that can infer triplet interactions and predict whether they are cooperative or competitive. By integrating network properties and advanced mathematical models, such as hyperbolic embeddings, researchers can gain a more comprehensive understanding of how protein complexes form and function in the cellular environment[18].

Several studies have demonstrated that complex networks, including the human protein–protein interaction network (hPIN) can be effectively modeled using hyperbolic geometry[19–21]. The Popularity-Similarity (PS) model provides a geometric framework in which network nodes are positioned within a two-dimensional hyperbolic space (H^2), represented as a disk. In this model, the radial coordinate of a node captures its popularity and evolutionary age, with older, highly connected proteins positioned closer to the center, while newly emerging proteins occupy the periphery. The angular coordinate encodes functional similarity, grouping proteins involved in shared biological processes or pathways. Alanis-Lobato et al. demonstrated that embedding the hPIN in hyperbolic space provides biologically meaningful insights, with radial positioning reflecting protein conservation and seniority, and angular positioning capturing functional and spatial organization within the cell[22]. Such mapping approaches can contribute to a deeper understanding of complex human disorders[23–25].

Motivated by the importance of higher-order protein interactions, we developed a computational framework to distinguish cooperative from competitive triplets in the human protein interaction network (hPIN). Starting from a high-confidence PPI network embedded in hyperbolic space, we identified open triangles, triplets of proteins where two interact with a shared partner but not with each other, and annotated them using structural data. Using this annotated dataset, we trained and evaluated a Random Forest algorithm to classify protein triplets as either cooperative or competitive based on topological, geometric and biological features. This classification is related to the previously defined "party" and "date" hubs, which distinguish proteins based on co-expression and temporal interaction patterns[26]. However, we adopt the terms cooperative and competitive to emphasize structural criteria, specifically, whether the two partners can bind the shared protein simultaneously at distinct interfaces (cooperative) or must do so mutually exclusively due to overlapping interfaces (competitive). To assess the predictive power of our approach, we performed structural validation using AlphaFold 3, confirming distinct spatial configurations between predicted cooperative and competitive triplets. By exploring multi-protein interactions, we provide a deeper insight into how molecular complexes are organized and operate within biological systems.

Results and discussion

Construction of the hPIN into the hyperbolic space and structural annotations of cooperative and competitive relationships

To investigate cooperative and competitive interactions between proteins in triplets, where one protein interacts with another two proteins, we first constructed a high-confidence human protein–protein interaction network (hPIN) using experimentally supported data (Fig. 1a). For this, we retrieved all human PPIs from the HIPPIE database and filtered interactions with a confidence score ≥ 0.71, as this threshold ensures that the majority of the interactions are validated through multiple independent sources. The resulting network comprised 15,319 proteins and 187,791 interactions. To uncover the latent geometry underlying the hPIN, we embedded the hPIN into the two-dimensional hyperbolic plane (H^2) using the LaBNE + HM algorithm, which integrates manifold learning with maximum likelihood estimation (see Methods for details). Each protein was assigned with a set of hyperbolic coordinates: a radial coordinate (r) representing its topological centrality where a shorter distance to the center corresponds to nodes with higher connectivity, and an angular coordinate (theta) indicating its similarity to other nodes based on interacting partners[18,22,25]. This hyperbolic embedding allowed us to extract geometric and topological features essential for the classification of cooperative versus competitive interactions within protein triplets.

To investigate the diversity of protein structures involved in cooperative triplets, we analyzed residue-level annotations from the Interactome3D database[27]. We identified triplets within experimentally resolved complexes, focusing on open triangle configurations in which a central "common" protein binds two partners (V1 and V2) that do not interact directly. Such motifs are indicative of potential cooperative relationships, especially when the binding regions of the common protein are distinct or complementary. Across all PDB complexes containing at least three proteins, we observed a broad distribution in the number of proteins per complex, with most structures comprising fewer than 15 unique proteins (Fig. 1b)[28].

These complexes were then computationally evaluated to identify open triangles. We found that while many PDB entries contained only a few open triangles, some complexes presented high structural complexity, producing up to 100 triplets (Fig. 1c). Large well-studied assemblies such as ribosomes, proteasomes and chaperones, where many subunits are resolved together within the same structural experiment, result in large amounts of triplets[29,30]. Subunits of such large complexes participate in multiple triplets, which could be a source of bias in analyses of their properties.

After characterizing the structural diversity of complexes, we proceeded to map the identified triplets into the hPIN. This allowed us to extract structurally supported triplets with topological relevance, which served as cooperative examples for modeling. Specifically, we overlapped the common–V1 and common–V2 interaction pairs from each structurally validated triplet with the interactions present in the hPIN (Fig. 1d). To avoid the redundancy that could be introduced by counting proteins from large complexes multiple times, as mentioned above, we retained only one cooperative triplet per common interactor (see Methods for details). This filtering resulted in a final, non-redundant set of 211 cooperative triplets, distributed across 352 PDB complexes, each annotated with residue-level interface information derived from Interactome3D. These well-supported triplets formed the positive class used to train our classification model (see Data Availability section to access these data).

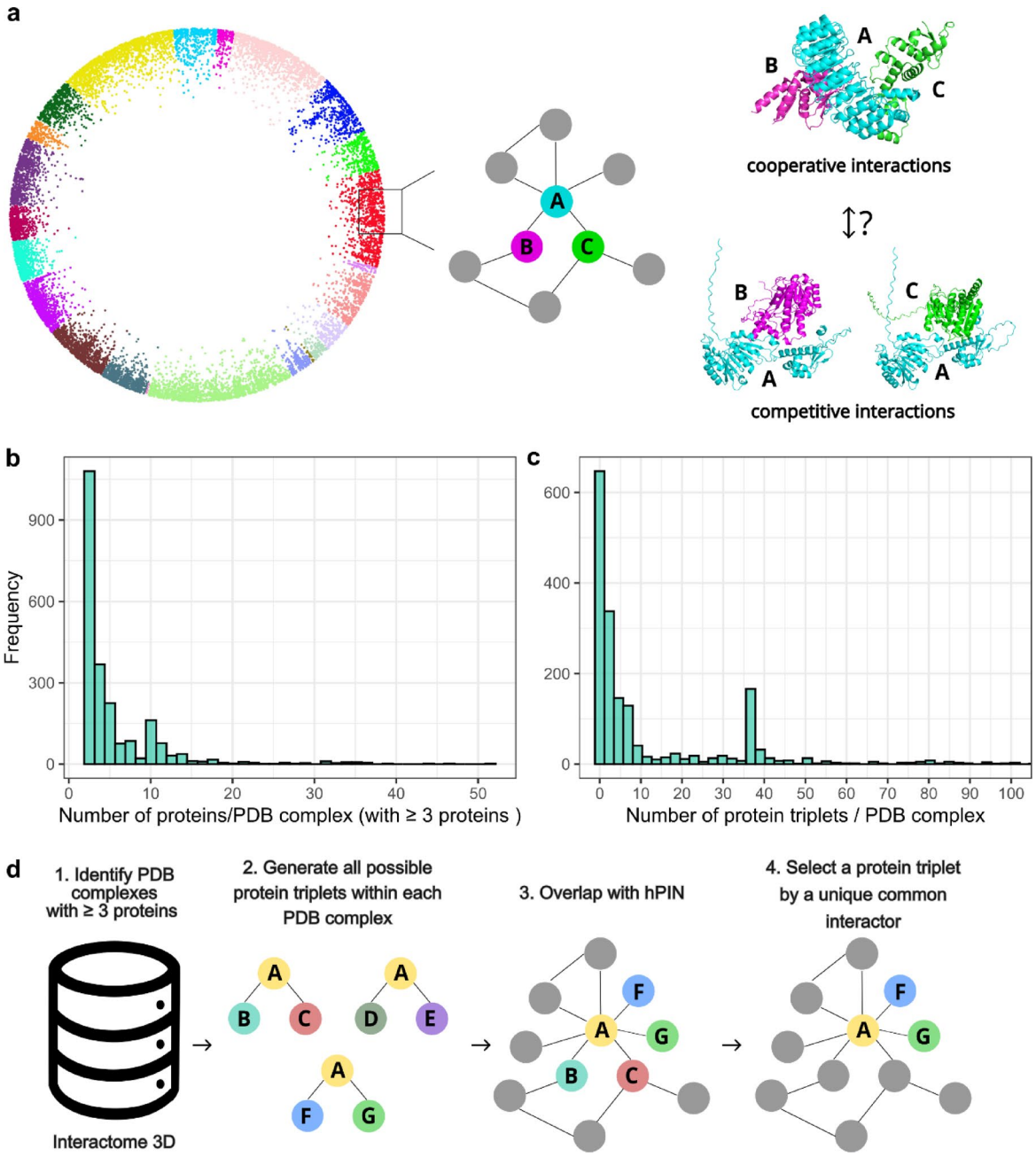


Fig. 1. Construction of the human protein interaction network and identification of structurally supported cooperative triplets. (**a**) Overview of the high-confidence human protein–protein interaction network (hPIN) constructed from the HIPPIE database (confidence score ≥ 0.71) and embedded into two-dimensional hyperbolic space using the LaBNE + HM algorithm. Node colors represent functional clusters identified using the angular gap-based clustering method described in our previous study[18], where proteins grouped by angular proximity reflect functional similarity. (**b**) Distribution of the number of proteins per PDB complex with at least three unique proteins, based on residue-level annotations from Interactome3D. (**c**) Distribution of the number of structurally supported open triangle motifs (i.e., cooperative triplets) per PDB complex. A subset of complexes contributes disproportionately high numbers of triplets, reflecting structurally rich assemblies. (**d**) Mapping of structurally supported triplets into the hPIN. Only one cooperative triplet was retained per common interactor to ensure non-redundancy, resulting in a curated set of 211 cooperative triplets used for model training.

Predicting cooperative triplets in the human protein interaction network

To systematically predict cooperative interactions among protein triplets, we constructed a dataset composed of positive and negative cases. The 211 structurally supported triplets identified from Interactome3D[27] served as the positive class, representing cooperative configurations in which a central protein (common interactor) binds

two partners without a direct interaction between them (V1 and V2). As a "noisy" negative class, we selected open triangles from the hPIN that lacked structural support. These contain both positives and negatives. To eliminate any potential positional bias (for example, due to labels), we randomized the assignment of V1 and V2 within each triplet.

To build our model, we extracted for each triplet a set of topological and geometric features. Topological and hyperbolic features have been previously shown to be predictive of protein interaction with function in protein (de-)phosphorylation[16]. These included hyperbolic coordinates and centrality measures (degree, closeness, betweenness, eigenvector) for each of the three proteins in a triplet, and distances, angular and radial differences for each pairwise relationship (common-V1, common-V2, V1–V2).

We also considered biological features: presence of a disordered region, and subcellular location, for each of the three proteins in a triplet. From a biological perspective, we assumed that the formation of stable cooperative complexes requires that interacting proteins are co-localized in the same subcellular compartment[31]. In addition, proteins with well-defined structural domains are often capable of forming stable binding interfaces, making structural order a key indicator for such interactions[32].

In contrast, intrinsically disordered proteins (IDPs), or proteins with intrinsically disordered regions (IDRs), lack a well-defined tertiary structure under physiological conditions. While this structural flexibility can facilitate transient interactions such as in cellular signaling, it is generally less favorable for stable, multimeric complex formation[33]. The final feature matrix comprised 42 features per triplet (see Methods for details).

We trained multiple machine learning models—Random Forest (RF), Support Vector Machine (SVM), Logistic Regression, Decision Trees, and k-Nearest Neighbors (kNN)—and evaluated their performance using a 70/30 train-test split. Specifically, 70% of the dataset was used for model training and internal validation, while the remaining 30% was reserved for testing. To address the class imbalance between cooperative (positive) and competitive (negative) triplets, we applied random undersampling to the majority class in the training set prior to model training, resulting in a balanced dataset of 296 samples. For the training procedure, we performed fivefold cross-validation repeated 10 times.

Among all tested classifiers, the Random Forest (RF) model (trained with 500 trees) achieved the highest performance, with a mean accuracy of 0.80, F1-score of 0.89, sensitivity of 0.80, and specificity of 0.80 on the held-out test set (Fig. 2a). ROC curve analysis with an AUC of 0.88 (Fig. 2b) indicated a superior ability to distinguish cooperative and competitive interactions between triplets of protein complexes. As our model predicts the interaction type within protein triplets, each unique triplet was evaluated once, resulting in a total of $n = 17{,}165{,}561$ triplet evaluations across the human protein interaction network. The model produces a probability score indicating whether a given triplet interaction is predicted to be cooperative or competitive. A total of 3,405,136 triplets received a score ≥ 0.5, and 311,557 triplets were classified with high confidence, receiving a score ≥ 0.9 (Supplementary Table S4).

To gain insight into which features were most predictive of cooperative interactions, we examined feature importance scores derived from the Random Forest model (Fig. 2c). The most important feature was the angular difference between V1 and V2, closely followed by the hyperbolic distance between V1 and V2 proteins. The high importance of the angular and hyperbolic geometrical features aligns with previous observations that angular positioning in hyperbolic space captures the functional organization of the proteome[22].

Other top-ranking features included the betweenness (BC), degree (DC) and closeness (CC) centrality of the common protein as well as its radial coordinate. BC represents proteins that act like bridges in the network, DC values indicate how well a node is directly connected to the other nodes and CC measures the central position of the proteins in the network. Regarding the radial coordinate, in the hyperbolic embedding, proteins closer to the center (with lower radial values) are usually more connected in the network and often older in evolutionary terms.; they can be informative for predicting cooperation or competition between the proteins.

Together, these results highlight the value of hyperbolic embedding, which captures latent geometric and topological properties of the protein interaction network that cannot be easily inferred from biological features alone. Although biological attributes like subcellular localization and structural order add meaningful information, it was the geometric features derived from the network, especially angular and hyperbolic distances, that played the most critical role in distinguishing cooperative from competitive triplets. This demonstrates that embedding the hPIN in hyperbolic space enhances our ability to capture functional relationships and improve predictions of protein interaction types.

Biological interpretation of predicted triplets

To investigate the biological relevance of our model's predictions, we focused on the subset of predicted cooperative and competitive triplets where all three proteins had low degree centrality (degree ≤ 10).

This filtering step reduced the influence of highly connected hub proteins and allowed us to focus on triplets where cooperative assembly might be more functionally focused. From this subset, we retained 1595 triplets and divided them into four quantiles based on their predicted probability scores, with Q1 representing low-scoring (competitive-like) and Q4 high-scoring (cooperative-like) predictions (Supplementary Table S5).

In addition to this subset, we evaluated the model on an external validation set derived from the most recent Interactome3D release, containing structurally supported triplets absent from the training data to ensure an independent test of predictive performance. Comparing the distribution of predicted scores across the all-predictions set, the training set, the external validation set, and the low DC subset showed distinct distributions (Fig. 3a). The all-predictions set was skewed toward lower scores (competitive), whereas the other three datasets showed a shift toward higher scores (cooperative), with the external validation subset showing the strongest enrichment. This finding confirms the model's ability to detect cooperative assemblies in previously unseen structural data.

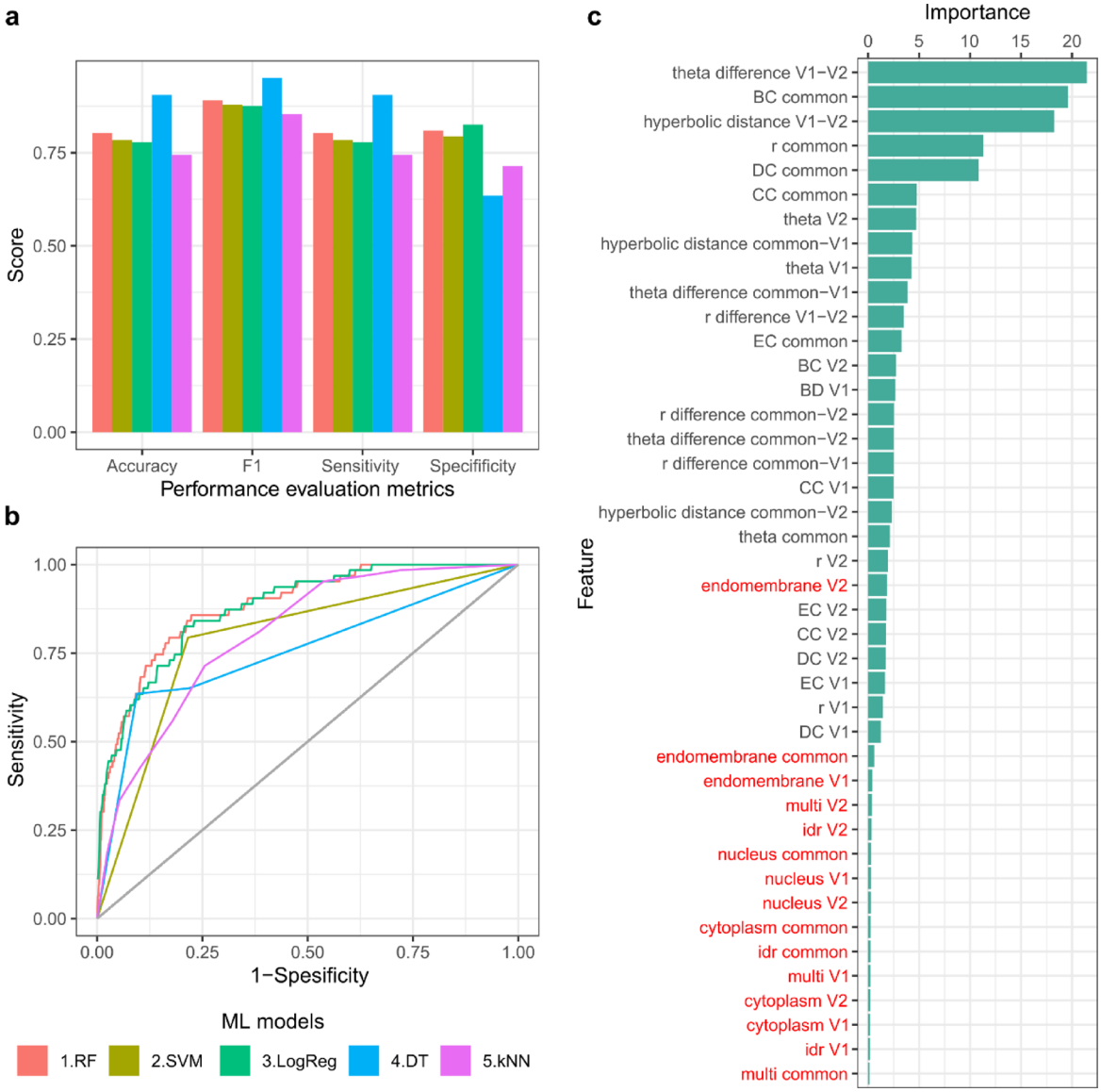


Fig. 2. Predicting cooperative protein triplets in the human protein interaction network using machine learning. (**a**) Performance metrics (accuracy, F1-score, sensitivity, and specificity) of five machine learning models trained to classify cooperative versus competitive triplets. Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LogReg), Decision Trees (DT), and k-Nearest Neighbors (kNN). The Random Forest (RF) classifier showed the highest overall performance. (**b**) Receiver Operating Characteristic (ROC) curves for all models. The RF model achieved the best performance with an area under the curve (AUC) of 0.89, indicating strong discriminatory power. (**c**) Feature importance scores from the RF classifier, ranked by mean decrease in accuracy. The most informative features included the angular difference between V1 and V2, the betweenness centrality of the common protein, and other geometric and topological features derived from the hyperbolic embedding and network structure. Biological features (labels marked in red) were less important.

Next, we assessed the paralogous relationship between V1 and V2 in each triplet. Based on the idea that paralogous proteins share structural and functional similarity, we initially expected them to appear more often in competitive interactions, where they could compete for the same binding site on the common interactor. Instead, our results revealed the opposite pattern. The proportion of paralogous V1–V2 pairs was highest in Q4 and decreased gradually toward Q1 (Fig. 3b), indicating that paralog pairs are more likely to occur in cooperative triplets. This finding is consistent with previous studies showing that paralogs are frequently retained within the same protein complexes after duplication, contributing to cooperative function, structural stability, and functional redundancy rather than direct competition[34,35].

Finally, we examined the angular (theta) difference between V1 and V2, the top-ranked feature from our model. Triplets in Q4 exhibited the smallest angular separation, with theta differences progressively increasing toward Q1 (Fig. 3c). This pattern suggests that cooperative triplets often involve proteins that are functionally

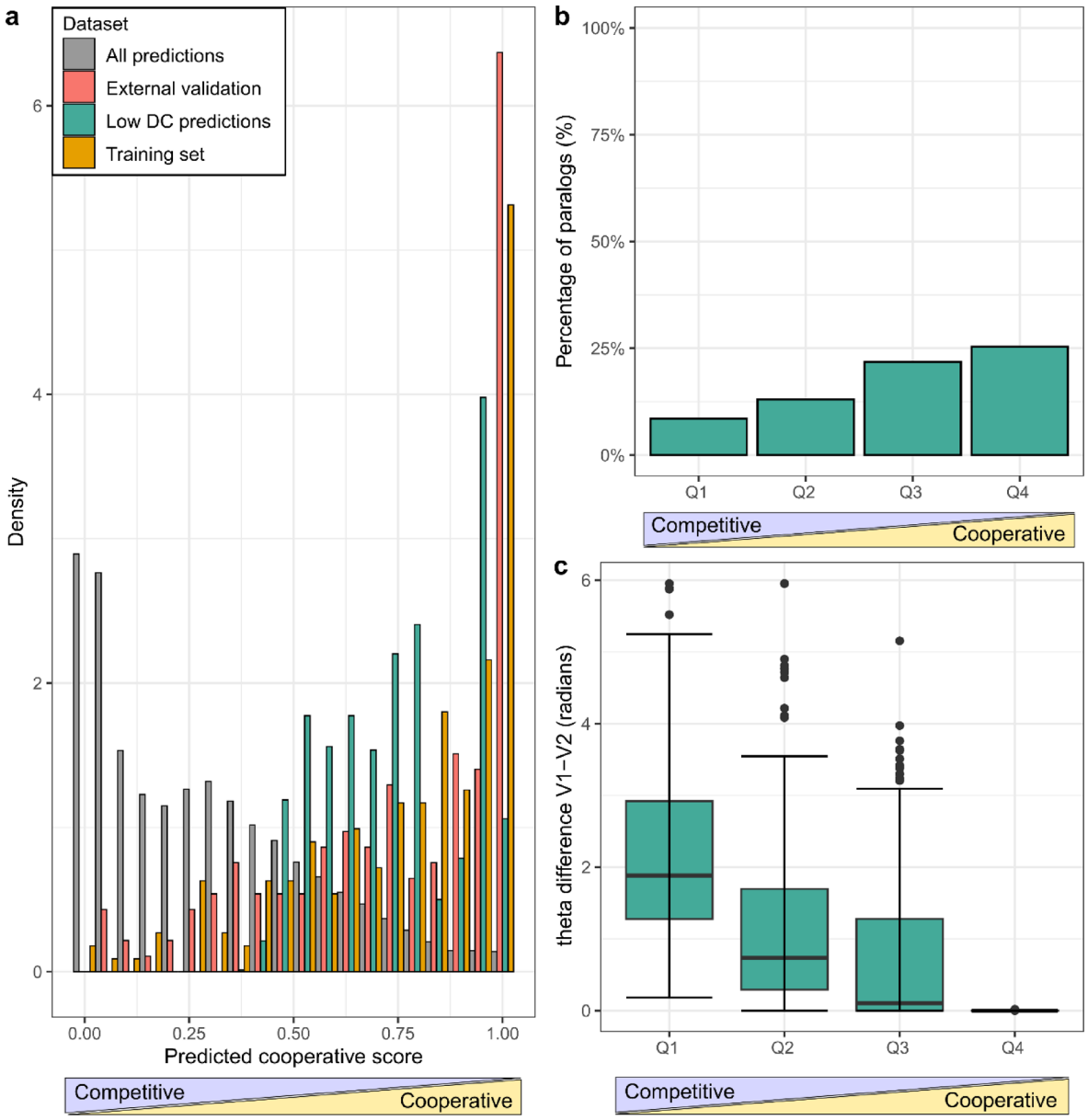


Fig. 3. Biological interpretation of predicted triplets based on model score distributions, paralogy, and angular separation. (**a**) Distribution of predicted cooperative scores across datasets (legend). The all-predictions set is enriched toward low scores (competitive-like), whereas the training, external validation, and low-degree common (DC) subsets concentrate near high scores (cooperative-like). (**b**) Proportion of triplets in which V1 and V2 are paralogs across prediction-score quartiles (Q1 = low/competitive, Q4 = high/cooperative). Paralogous pairs are more prevalent in Q4, indicating that similar proteins are more often predicted to act cooperatively. (**c**) Distribution of angular (theta) differences between V1 and V2 across quartiles. Angular separation decreases toward Q4; cooperative triplets tend to show smaller theta differences (functional proximity), whereas competitive-like triplets display larger separations.

similar in the hyperbolic space, consistent with paralogs or related proteins acting together within the same complex. In contrast, larger angular separations, indicative of greater functional diversity, were more frequently observed in predicted competitive triplets. These findings support the idea that cooperation is favored between proteins sharing similar functional and structural contexts, as is typical for subunits within protein complexes. This observation can be compared with the L3 principle described by Kovács et al., where proteins connected by many paths of length three tend to share interaction partners and often bind the same interface[36].

Structural evaluation of predictions using AlphaFold

To evaluate whether our predictions are meaningful from a structural point of view, we selected four representative triplets for structural modeling: two from the top-ranked cooperative predictions and two from the bottom-ranked competitive predictions, based on their respective classification scores. For each triplet, we

generated AlphaFold 3 models of the two binary interactions (common–V1 and common–V2) yielding eight predicted complexes in total (Fig. 4).

We report both pTM and ipTM values as measures of model confidence. While pTM values reflect the confidence in the overall structural fold of the predicted complex, ipTM values focus specifically on the accuracy of the predicted interaction interface, making them particularly informative for our analysis. To further assess the nature of the interfaces, we manually reviewed the predicted structures in PyMOL, examining the positioning of residues at the binding interface. Triplets with non-overlapping binding sites on the common protein were interpreted as cooperative, whereas overlapping or adjacent binding surfaces indicated potential competition between V1 and V2.

The triplet composed of Q96D31 (*ORAI1*) binding to O75185 (*ATP2C2*) and Q13507 (*TRPC3*) was predicted by our model to be cooperative, based on a high classification score, together with favorable pTM and ipTM values for both binary complexes (Fig. 4a). More specifically, ATP2C2 interacts primarily with transmembrane helices of ORAI1 (ipTM = 0.14; pTM = 0.55), while TRPC3 binds to an extracellular region (ipTM = 0.14; pTM = 0.60), with no interface overlap between the two partners. More specifically, ATP2C2 interacts primarily with the cytoplasmic domains of ORAI1, while TRPC3 binds to an extracellular region, with no interface overlap between the two partners. Such an observed spatial configuration aligns with cooperative interactions. ORAI1 is the pore-forming subunit of CRAC channels, central to store-operated calcium entry (SOCE) in immune cells. Its C-terminal domain plays a critical role in STIM1-mediated gating of the channel, and mutations in ORAI1 can cause severe combined immunodeficiency[37,38]. ATP2C2 is an ATPase localized to the Golgi apparatus, where it maintains calcium and manganese homeostasis in the secretory pathway[39]. Although a direct interaction between ATP2C2 and ORAI1 has not previously been described, their co-localization in calcium-regulating compartments suggests a possible functional relationship. TRPC3 is a diacylglycerol-sensitive, non-selective cation channel activated downstream of G protein–coupled receptors (GPCRs) and receptor tyrosine kinases (RTKs)[40]. TRPC3 participates in receptor-operated calcium entry and has been reported to form complexes with ORAI1 and STIM1, integrating different calcium influx pathways[41]. Taken together, the non-overlapping interaction interfaces of ATP2C2 and TRPC3 on ORAI1, combined with their distinct but complementary roles in calcium signaling, support a cooperative mode of interaction. ORAI1 may link multiple calcium signaling pathways through interactions with ATP2C2 and TRPC3 at separate binding sites.

A similar pattern was observed in the second cooperative example of Q8N961 (*ABTB2*) binding to Q9BV47 (*DUSP26*) and P04053 (*DNTT*), where the two partners were positioned at distinct, non-overlapping sites on the common interactor, as revealed by AlphaFold modeling. DUSP26 engaged one end of ABTB2 (ipTM = 0.37, pTM = 0.68), whereas DNTT bound at a different, non-overlapping site (ipTM = 0.26, pTM = 0.56), with clear spatial separation between the binding sites, supporting a cooperative mode of interaction. ABTB2 (Ankyrin repeat and BTB/POZ domain-containing protein 2) contains ankyrin repeat and BTB/POZ domains, which are structural motifs that mediate protein–protein interactions. While direct studies of ABTB2's interaction partners remain limited, genetic evidence links it to reduced apoptosis in cancer cells[42]. This finding points to a possible role for ABTB2 in promoting cell survival under stress conditions. DUSP26 inactivates MAPK1 (*ERK2*) and MAPK3 (*ERK1*), thereby modulating downstream MAP kinase signaling events. It interacts with the heat shock transcription factor Hsf4b, and through this association, places Hsf4b within a regulatory circuit of the MAP kinase pathway, influencing transcriptional responses to cellular stress[43]. DNTT is a template-independent DNA polymerase that adds non-templated nucleotides during V(D)J recombination, a crucial process for generating diversity in adaptive immune receptors[44]. This analysis indicates that ABTB2 can bind both partners simultaneously, potentially allowing parallel regulation of MAP kinase mediated stress responses and V(D)J recombination (Fig. 4b).

In contrast, two additional triplets classified by the model as competitive provide further structural and biological support for mutually exclusive interactions. The first competitive triplet consists of Q5T8I9 (*HENMT1*) interacting with Q9H903 (*MTHFD2L*) and Q9BY49 (*PECR*) (Fig. 3c). HENMT1 is an RNA methyltransferase that catalyzes the 2′-O-methylation of the 3′ terminal nucleotide of PIWI-interacting RNAs (piRNAs), a modification that enhances their stability and protects them from exonucleolytic degradation[45]. MTHFD2L is a mitochondrial enzyme with dual redox cofactor specificity that functions as a methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase, participating in the mitochondrial one-carbon metabolism pathway. It is expressed in both adult and embryonic tissues, contributing to nucleotide biosynthesis and cellular metabolic homeostasis[46]. PECR is a peroxisomal enzyme that catalyzes the final step in the NADPH-dependent reduction of trans-2-enoyl-CoAs during fatty acid chain elongation, contributing to lipid metabolism and peroxisomal function[47]. AlphaFold 3 structural modeling revealed that MTHFD2L and PECR bind to overlapping or closely adjacent regions on the surface of HENMT1. Interface analysis showed an ipTM of 0.21 (pTM = 0.61) for the HENMT1–MTHFD2L complex and an ipTM of 0.19 (pTM = 0.58) for HENMT1–PECR, with significant spatial overlap between the binding footprints. This suggests that concurrent binding is prevented by overlap of the interaction interfaces. Given the distinct biological functions of MTHFD2L (mitochondrial one carbon metabolism) and PECR (peroxisomal fatty acid metabolism), such competition may limit HENMT1's capacity to simultaneously regulate both metabolic pathways.

In the second competitive triplet, Q9Y6W8 (*ICOS*) binds to P08582 (*MELTF*) and O75144 (*ICOSLG*) (Fig. 4d). ICOS is a T-cell co-stimulatory receptor that plays a non-redundant role in adaptive immunity, particularly in T-cell activation and differentiation[48]. Its established ligand, ICOSLG, modulates T-cell activation and trafficking in the tumor microenvironment, contributing to cancer progression and immune regulation[49]. MELTF is a GPI-anchored transferrin-family protein involved in iron binding and transport at the cell surface. Recent studies suggest that MELTF expression can attenuate malignant melanoma progression in both murine models and human patients[50]. AlphaFold 3 modeling indicated that ICOS–ICOSLG forms a higher-confidence interface (ipTM = 0.64; pTM = 0.64) on the ICOS Ig-like head, consistent with its known physiological interaction,

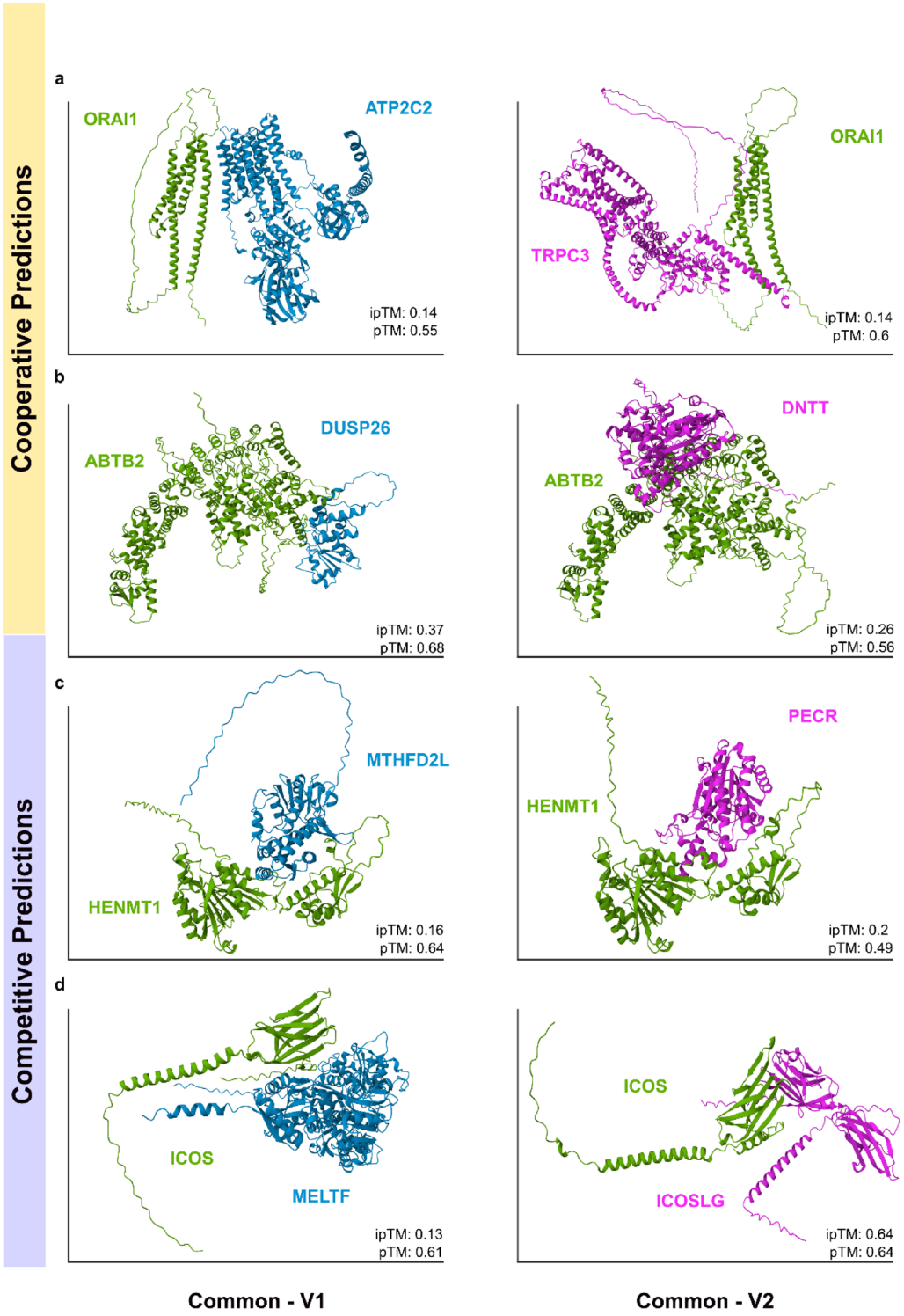


Fig. 4. Structural validation of model predictions using AlphaFold. Representative examples of predicted cooperative (top two rows) and competitive (bottom two rows) triplets. Each panel shows a binary complex between the common protein and one of its partners, modeled with AlphaFold 3. In cooperative cases, the two partners bind to distinct, non-overlapping regions on the common protein: (**a**) O75185 (ATP2C2) and Q13507 (TRPC3) binding to Q96D31 (ORAI1); (**b**) Q9BV47 (DUSP26) and P04053 (DNTT) binding to Q8N961 (ABTB2). In competitive cases, both partners share overlapping or adjacent binding interfaces with the common protein: (**c**) Q9H903 (MTHFD2L) and Q9BY49 (PECR) binding to Q5T8I9 (HENMT1); (**d**) P08582 (MELTF) and O75144 (ICOSLG) binding to Q9Y6W8 (ICOS).

which has been partially structurally characterized[51]. By contrast, the ICOS–MELTF interface had much lower prediction confidence (ipTM = 0.13; pTM = 0.61). Structural mapping places MELTF's binding footprint on the same binding site of the ICOS Ig-domain used by ICOSLG, producing overlapping interfaces, indicating that the two cannot bind simultaneously. Given the established role of ICOSLG as the physiological ligand of ICOS and the distinct functions of MELTF, the observed structural overlap and the higher-confidence ICOS–ICOSLG interface indicate that the two ligands compete for the same binding site, with ICOS.

While our model generally performs well in distinguishing cooperative from competitive triplets, there are cases where caution is warranted. For example, the triplet with P78508 (*KCNJ10*) binding to Q15049 (*MLC1*) and Q96SZ5 (*ADO*) was predicted as competitive based on the classification score. However, AlphaFold 3 structural modeling suggests that MLC1 and ADO bind to different regions on the surface of KCNJ10, suggesting potential cooperation rather than competition engagement (model not shown). This observation shows that even when the model assigns a cooperative or competitive label based on a prediction score, structural modeling can reveal overlapping interfaces that are not captured by network features alone and highlights the importance of structural validation.

Conclusions

In this study, we systematically explored cooperative and competitive interactions among protein triplets by combining structural data, network topology, hyperbolic embeddings, and machine learning-based prediction. Our results provide novel insights into the structural and functional determinants that differentiate cooperative from competitive assemblies in the human protein interaction network (hPIN). By embedding the hPIN into hyperbolic space, we were able to capture latent geometric properties of the network that traditional biological or topological features alone could not fully reveal. The angular separation between proteins emerged as the most informative feature for predicting interaction type, highlighting the value of hyperbolic geometry in modeling functional organization within complex biological networks. This finding is consistent with previous work showing that angular proximity in hyperbolic space reflects functional similarity, while radial positioning captures evolutionary and topological hierarchy.

Our machine learning models, particularly the Random Forest classifier, demonstrated strong performance in distinguishing cooperative from competitive triplets, achieving high accuracy, F1-score, and AUC values. We observed that biological context such as subcellular localization and the presence of intrinsically disordered regions did not contribute to predictive success as much as geometrical and topological features. Cooperative triplets were enriched in paralogous partners, reflecting their ability to associate with the common protein through non-overlapping interfaces. Importantly, evaluation on an external dataset of newly released structurally supported triplets confirmed the performance of the model and its ability to generalize to previously unseen data.

Validation through AlphaFold 3 demonstrated that cooperative interactions involve distinct interfaces, while competitive interactions are characterized by significant interface overlap. The identification of cooperative modules involved in processes such as ion transport, immune response and cellular stress, strengthens the biological significance of our results. Conversely, competitive triplets showcase the complexity of competition-driven regulatory mechanisms in metabolic pathways and T-cell activation.

A limitation of our approach is the assumption that open triplets not supported by structural complexes represent competitive interactions. While this enabled large-scale modeling, it may introduce false negatives if some of these triplets are cooperative but lack structural evidence. A more accurate strategy would involve identifying cases where two proteins bind overlapping interfaces of a common partner across different PDB entries. However, assembling such a dataset would require extensive structural alignment and is not currently manageable across the whole human interactome. As an alternative, filtering triplets based on the number of L3 pathways connecting V1 and V2 may help improving the training dataset. These refinements represent promising directions for future work.

In conclusion, our study demonstrates that integrating network-based features, hyperbolic embeddings, and machine learning provides a powerful framework for investigating higher-order protein interactions. Our findings demonstrate the potential of geometric approaches as powerful tools for exploring the complexity of biological networks.

Materials and methods

Human protein interaction network construction

The human protein–protein interaction network (hPIN) used in this study was built using version 2.3 of the Human Integrated Protein–Protein Interaction rEference (HIPPIE) database[52,53]. HIPPIE integrates experimentally supported physical interactions from major expert-curated resources and assigns a confidence score to each interaction based on evidence type, reproducibility, and publication quality. In this study, the hPIN was constructed using interactions with confidence score ≥ 0.71 because the majority of the edges are supported by more than one experiment[18,24]. Self-interactions were removed, and only proteins within the largest connected component (LCC) were kept. The final network comprises 15,319 proteins and 185,791 high-confidence interactions (Supplementary Tables S1 and S2).

Mapping the hPIN into the hyperbolic space

We embedded the hPIN in the two-dimensional hyperbolic plane using the R package "NetHypGeom", which implements the LaBNE + HM algorithm[54]. The algorithm combines manifold learning and maximum likelihood estimation to uncover the hidden geometry of complex networks[55,56]. LaBNE + HM expects a connected network as input, typically the LCC. The other components cannot be mapped due to the lack of adjacency

information relative to the LCC. The Popularity–Similarity (PS) model describes the growth of complex networks in hyperbolic space, where radial coordinates reflect node popularity or evolutionary age, and angular coordinates capture functional similarity. In this model, new nodes preferentially connect to existing nodes that are hyperbolically closest to them[20,57]. The network was embedded in H2 to infer the hyperbolic coordinates of each protein, with parameters $\gamma = 2.97$, $T = 0.83$, and $w = 2\pi$. The 15,319 nodes of the hPIN lie within a hyperbolic disc where the radial coordinate of a node, ri, represents the popularity dimension with nodes that joined the system first being close to the disc's center. The angular coordinate, θi, represents the similarity dimension[22].

Identification of cooperative interactions and structural annotations

We then identified all the possible open triangles within the hPIN, resulting in approximately 17 million unique triplets. An open triangle is defined as a triplet of proteins in which a central node (common interactor) interacts with two others (V1 and V2), which do not interact. We focused exclusively on open triangles because in closed triangles (where all three proteins interact), no central or common node can be uniquely defined as all nodes are equivalent. This simplified structure allowed a more effective way of modeling cooperative and competitive interaction types.

To annotate structural support for these triplets, we used PPI data from the Interactome3D database, which includes residue-level annotations of experimentally resolved structures deposited in the Protein Data Bank (PDB)[27]. Model-based predictions and self-interactions were excluded from the analysis. Protein complexes were grouped by PDB ID, and only those involving at least three proteins were considered. Within each complex, all possible protein triplets were generated, and the number of observed pairwise interactions per triplet was counted. Triplets in which exactly two of the three possible pairwise interactions were present were classified as cooperative interactions in protein triplet complexes. We then integrated the Interactome3D-derived cooperative triplets with the full set of hPIN open triangles to determine the final set of positive cases used for model training and evaluation. To avoid overrepresentation of proteins, present in many complexes or in complexes with many subunits, only one validated triplet per common interactor was retained, by keeping the first occurrence based on dataset order (Supplementary Table S3). All remaining open triangles in the hPIN were considered non-cooperative (or competitive).

Feature extraction and randomization

Each protein triplet in the dataset was annotated with a total of 42 features, capturing both network topology and biological context at the node and edge levels. For every protein within a triplet (common interactor, V1, and V2) we extracted 11 features, resulting in 33 features per triplet. These included the two hyperbolic coordinates (r and theta) and four network centrality measures: Degree Centrality (DC)[58], Closeness Centrality (CC)[59], Betweenness Centrality (BC)[60], and Eigenvector Centrality (EC)[61]. These centralities quantify different aspects of a node's importance within the network, such as how connected it is (DC), how close it is to all other nodes (CC), how often it lies on shortest paths (BC), and how influential its neighbors are (EC). To incorporate biological relevance, we further assigned each protein five additional features: the presence of Intrinsically Disordered Regions (IDRs), obtained from the DisProt database v9.7[62], and the subcellular localization (nucleus, cytoplasm, endomembrane, and multi-localized) retrieved from the Human Protein Atlas database v24.0[63].

In addition, three more features were computed for each of the edges between the possible protein pairs in the triplet (common–V1, common–V2, V1–V2): hyperbolic distance (hd), radial difference (rd), and angular difference (thetad), resulting in 9 edge-level features per triplet. To avoid positional bias during model training, the order of V1 and V2 was swapped in a random selection of 50% of the triplets. This ensured that the model learned from meaningful topological and biological properties rather than any artificial ordering patterns. The complete list of features and their descriptions is provided in Supplementary Tables S1 and S2. To access the complete dataset that was used for the modeling see Data Availability section.

Model development and evaluation

Model development was performed using the caret package in R[64]. The primary classification algorithm applied was Random Forest (RF)[65], selected for its robustness and effectiveness in handling high-dimensional biological data. To benchmark performance, we additionally trained and evaluated four other machine learning algorithms: Support Vector Machine (SVM)[66], Logistic Regression (LogReg)[67], Decision Tree (DT)[68], and k-Nearest Neighbors (kNN)[69]. The dataset was split into a training set (70%) and a test set (30%), and class imbalance was addressed by applying random undersampling of the majority class during training. Model training was conducted using fivefold cross-validation repeated 10 times to ensure robust hyperparameter optimization. For the Random Forest model, parameters such as the number of variables randomly sampled at each split ($n = 22$) and the number of decision trees ($n = 500$) were tuned based on cross-validation performance. Model evaluation was carried out on the independent test set using multiple performance metrics, such as Receiver Operating Characteristic (ROC) curves[70] and the Area Under the Curve (AUC)[71] to assess the model's ability to distinguish between cooperative and competitive interactions. In addition, we calculated accuracy, precision, recall (sensitivity), and F1-score to comprehensively evaluate classification performance, particularly in the context of class imbalance[72]. Finally, we computed feature importance scores using the Random Forest model to identify the most informative topological and biological features contributing to the classification task.

Homology annotations and external evaluation for biological validation of predicted cooperative and competitive protein triplet complexes

Following the model application to the complete dataset of open triangles in the hPIN, we annotated the predicted interactions with additional biological features to support downstream analysis. Specifically, we integrated information about the paralogy between V1 and V2 proteins and protein sequence lengths. Paralogy

between V1 and V2 was assessed using curated human data (GRCh38.p14) from the Ensembl database release 113[73]. In addition, to test the predictive score of the model, we evaluated predictions on an external dataset of structurally supported triplets obtained from the most recent Interactome3D release[27]. Triplets were extracted from newly available complexes following the same criteria as for the training data, and any overlapping cases with the training set were excluded. This dataset provided an independent benchmark of previously unseen triplets for model validation.

Structural validation of predicted cooperative and competitive protein triplet complexes using AlphaFold 3

Following the prediction of protein triplet complexes as either cooperative or competitive, we proceeded with structural modeling to evaluate whether the predicted interaction types were supported by structural evidence. For each protein triplet, the individual protein sequences were retrieved from the UniProtKB database[74] and used as input for AlphaFold 3[75]. Instead of modeling the entire triplet as a single complex, we performed pairwise structure prediction by separately analyzing the common-V1 and common-V2 protein pairs. Specifically, we examined whether V1 and V2 interacted with the same or overlapping regions on the common interactor protein. Overlap in binding sites was interpreted as evidence of potential competitive binding, while distinct binding interfaces were indicative of potential cooperative interactions. We used PyMOL to visualize the predicted structures, which allowed us to manually examine where the proteins were binding and how they were positioned in 3D space[76].

Data availability

All supplementary data supporting the findings of this study, including node- and edge-level features, structural annotations, and prediction scores, are available at: https://github.com/avagiona/Cooperative-Competitive-interactions-within-protein-triplets.

Received: 8 July 2025; Accepted: 8 September 2025
Published online: 15 September 2025

References

1. Toyama, B. H. et al. Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971–982. https://doi.org/10.1016/j.cell.2013.07.037 (2013).
2. Barabási, A.-L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113. https://doi.org/10.1038/nrg1272 (2004).
3. Koh, G. C. K. W., Porras, P., Aranda, B., Hermjakob, H. & Orchard, S. E. Analyzing protein-protein interaction networks. *J. Proteome Res.* **11**, 2014–2031. https://doi.org/10.1021/pr201211w (2012).
4. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986–998. https://doi.org/10.1016/j.cell.2011.02.016 (2011).
5. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68. https://doi.org/10.1038/nrg2918 (2011).
6. Rual, J.-F. et al. Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178. https://doi.org/10.1038/nature04209 (2005).
7. Stumpf, M. P. H. et al. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci.* **105**, 6959–6964. https://doi.org/10.1073/pnas.0708078105 (2008).
8. Rigaut, G. et al. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030–1032. https://doi.org/10.1038/13732 (1999).
9. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340**, 245–246. https://doi.org/10.1038/340245a0 (1989).
10. Luo, F. et al. Modular organization of protein interaction networks. *Bioinformatics* **23**, 207–214. https://doi.org/10.1093/bioinformatics/btl562 (2007).
11. Nooren, I. M. A. & Thornton, J. M. Diversity of protein–protein interactions. *EMBO J.* **22**, 3486–3492. https://doi.org/10.1093/emboj/cdg359 (2003).
12. Benson, A. R., Gleich, D. F. & Leskovec, J. Higher-order organization of complex networks. *Science* **353**, 163–166. https://doi.org/10.1126/science.aad9029 (2016).
13. Battiston, F. et al. Networks beyond pairwise interactions: Structure and dynamics. *Phys. Rep.* **874**, 1–92. https://doi.org/10.1016/j.physrep.2020.05.004 (2020).
14. Patel, M. S., Nemeria, N. S., Furey, W. & Jordan, F. The pyruvate dehydrogenase complexes: Structure-based function and regulation*. *J. Biol. Chem.* **289**, 16615–16623. https://doi.org/10.1074/jbc.R114.563148 (2014).
15. Oeckinghaus, A., Hayden, M. S. & Ghosh, S. Crosstalk in NF-κB signaling pathways. *Nat. Immunol.* **12**, 695–708. https://doi.org/10.1038/ni.2065 (2011).
16. Mintseris, J. & Weng, Z. Structure, function, and evolution of transient and obligate protein–protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10930–10935. https://doi.org/10.1073/pnas.0502667102 (2005).
17. Li, H., Zhou, Y. & Zhang, Z. Competition-cooperation relationship networks characterize the competition and cooperation between proteins. *Sci. Rep.* **5**, 11619. https://doi.org/10.1038/srep11619 (2015).
18. Vagiona, A.-C. et al. Prediction of protein interactions with function in protein (de-)phosphorylation. *PLoS ONE* **20**, 0319084. https://doi.org/10.1371/journal.pone.0319084 (2025).
19. Cannistraci, C. V., Alanis-Lobato, G. & Ravasi, T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* **29**, i199-209. https://doi.org/10.1093/bioinformatics/btt208 (2013).
20. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguñá, M. Hyperbolic geometry of complex networks. *Phys. Rev. E* **82**, 036106. https://doi.org/10.1103/PhysRevE.82.036106 (2010).
21. Serrano MA, Boguna M, Sagues F. Uncovering the hidden geometry behind metabolic networks 2011.
22. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. The latent geometry of the human protein interaction network. *Bioinformatics* **34**, 2826–2834. https://doi.org/10.1093/bioinformatics/bty206 (2018).
23. Härtner, F., Andrade-Navarro, M. A. & Alanis-Lobato, G. Geometric characterisation of disease modules. *Appl. Netw. Sci.* **3**, 10. https://doi.org/10.1007/s41109-018-0066-3 (2018).
24. Vagiona, A.-C., Mier, P., Petrakis, S. & Andrade-Navarro, M. A. Analysis of huntington's disease modifiers using the hyperbolic mapping of the protein interaction network. *IJMS* **23**, 5853. https://doi.org/10.3390/ijms23105853 (2022).

25. Zahra, N. U. A., Vagiona, A.-C., Uddin, R. & Andrade-Navarro, M. A. Selection of multi-drug targets against drug-resistant mycobacterium tuberculosis XDR1219 using the hyperbolic mapping of the protein interaction network. *IJMS* **24**, 14050. https://doi.org/10.3390/ijms241814050 (2023).
26. Han, J.-D.J. et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93. https://doi.org/10.1038/nature02555 (2004).
27. Mosca, R., Céol, A. & Aloy, P. Interactome3D: Adding structural details to protein networks. *Nat. Methods* **10**, 47–53. https://doi.org/10.1038/nmeth.2289 (2013).
28. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D complex: A structural classification of protein complexes. *PLoS Comput. Biol.* **2**, 155. https://doi.org/10.1371/journal.pcbi.0020155 (2006).
29. Schweitzer, A. et al. Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proc. Natl. Acad. Sci.* **113**, 7816–7821. https://doi.org/10.1073/pnas.1608050113 (2016).
30. Anger, A. M. et al. Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80–85. https://doi.org/10.1038/nature12104 (2013).
31. Park, S. et al. Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. *Mol. Syst. Biol.* **7**, 494. https://doi.org/10.1038/msb.2011.29 (2011).
32. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84**, 551–575. https://doi.org/10.1146/annurev-biochem-060614-034142 (2015).
33. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29. https://doi.org/10.1038/nrm3920 (2015).
34. Dandage, R. & Landry, C. R. Paralog dependency indirectly affects the robustness of human cells. *Mol. Syst. Biol.* **15**, 8871. https://doi.org/10.15252/msb.20198871 (2019).
35. van Dam, T. J. P. & Snel, B. Protein complex evolution does not involve extensive network rewiring. *PLoS Comput. Biol.* **4**, 1000132. https://doi.org/10.1371/journal.pcbi.1000132 (2008).
36. Kovács, I. A. et al. Network-based prediction of protein interactions. *Nat. Commun.* **10**, 1240. https://doi.org/10.1038/s41467-019-09177-y (2019).
37. Palty, R., Stanley, C. & Isacoff, E. Y. Critical role for Orai1 C-terminal domain and TM4 in CRAC channel gating. *Cell Res.* **25**, 963–980. https://doi.org/10.1038/cr.2015.80 (2015).
38. Feske, S. et al. A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185. https://doi.org/10.1038/nature04702 (2006).
39. Vanoevelen, J. et al. The secretory pathway Ca2+/Mn2+-ATPase 2 is a golgi-localized pump with high affinity for Ca2+ Ions*. *J. Biol. Chem.* **280**, 22800–22808. https://doi.org/10.1074/jbc.M501026200 (2005).
40. Venkatachalam, K. & Montell, C. TRP channels. *Ann. Rev. Biochem.* **76**, 387–417. https://doi.org/10.1146/annurev.biochem.75.103004.142819 (2007).
41. Ambudkar, I. S., de Souza, L. B. & Ong, H. L. TRPC1, Orai1, and STIM1 in SOCE: Friends in tight spaces. *Cell Calcium* **63**, 33–39. https://doi.org/10.1016/j.ceca.2016.12.009 (2017).
42. Gong, Y. et al. ABTB2 Regulatory variant as predictor of epirubicin-based neoadjuvant chemotherapy in luminal a breast cancer. *Front. Oncol.* **10**, 571517. https://doi.org/10.3389/fonc.2020.571517 (2020).
43. Hu, Y. & Mivechi, N. F. Association and regulation of heat shock transcription factor 4b with both extracellular signal-regulated kinase mitogen-activated protein kinase and dual-specificity tyrosine phosphatase DUSP26. *Mol. Cell. Biol.* **26**, 3282–3294. https://doi.org/10.1128/MCB.26.8.3282-3294.2006 (2006).
44. Kuznetsova, A. A. et al. Insight into the mechanism of DNA synthesis by human terminal deoxynucleotidyltransferase. *Life Sci. Alliance.* **5**, 202201428. https://doi.org/10.26508/lsa.202201428 (2022).
45. Gainetdinov, I. et al. Terminal modification, sequence, length, and piwi-protein identity determine piRNA stability. *Mol. Cell* **S1097–2765**(21), 00746–00752. https://doi.org/10.1016/j.molcel.2021.09.012 (2021).
46. Shin, M., Bryant, J. D., Momb, J. & Appling, D. R. Mitochondrial MTHFD2L is a dual redox cofactor-specific methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase expressed in both adult and embryonic tissues. *J. Biol. Chem.* **289**, 15507. https://doi.org/10.1074/jbc.M114.555573 (2014).
47. Das, A. K., Uhler, M. D. & Hajra, A. K. Molecular cloning and expression of mammalian Peroxisomal*trans*-2-Enoyl-coenzyme a reductase cDNAs*. *J. Biol. Chem.* **275**, 24333–24340. https://doi.org/10.1074/jbc.M001168200 (2000).
48. Dillon, S. R. et al. Non-redundant roles of t cell costimulation pathways in inflammatory arthritis revealed by dual blockade of icos and cd28 with acazicolcept (ALPN-101). *Arthritis Rheumatol.* **75**, 1344–1356. https://doi.org/10.1002/art.42484 (2023).
49. Dong, Y. et al. ICOSLG-associated immunological landscape and diagnostic value in oral squamous cell carcinoma: A prospective cohort study. *Front. Cell Dev. Biol.* https://doi.org/10.3389/fcell.2023.1257314 (2023).
50. Jandova, J. & Wondrak, G. T. Melanotransferrin (MELTF, MFI2, CD228) expression attenuates malignant melanoma progression in the A375-Luc2 murine metastasis model and human patients. *J. Invest. Dermatol.* **144**, 2820-2823.e6. https://doi.org/10.1016/j.jid.2024.05.028 (2024).
51. Rujas, E., Cui, H., Sicard, T., Semesi, A. & Julien, J.-P. Structural characterization of the ICOS/ICOS-L immune complex reveals high molecular mimicry by therapeutic antibodies. *Nat. Commun.* **11**, 5066. https://doi.org/10.1038/s41467-020-18828-4 (2020).
52. Schaefer, M. H. et al. HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS ONE* **7**, 31826. https://doi.org/10.1371/journal.pone.0031826 (2012).
53. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414. https://doi.org/10.1093/nar/gkw985 (2017).
54. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. A. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Appl. Netw. Sci.* **1**, 10. https://doi.org/10.1007/s41109-016-0013-0 (2016).
55. Alanis-Lobato, G., Mier, P. & Andrade-Navarro, M. A. Efficient embedding of complex networks to hyperbolic space via their laplacian. *Sci. Rep.* **6**, 30108. https://doi.org/10.1038/srep30108 (2016).
56. Papadopoulos, F., Aldecoa, R. & Krioukov, D. Network geometry inference using common neighbors. *Phys. Rev. E* **92**, 022807. https://doi.org/10.1103/PhysRevE.92.022807 (2015).
57. Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540. https://doi.org/10.1038/nature11459 (2012).
58. Freeman, L. C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1**, 215–239. https://doi.org/10.1016/0378-8733(78)90021-7 (1978).
59. Bavelas, A. Communication patterns in task-oriented groups. *J. Acoustical Soc. Am.* **22**, 725–730. https://doi.org/10.1121/1.1906679 (1950).
60. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35. https://doi.org/10.2307/3033543 (1977).
61. Bonacich, P. Factoring and weighting approaches to status scores and clique identification. *J. Mathe. Sociol.* **2**, 113–120. https://doi.org/10.1080/0022250X.1972.9989806 (1972).
62. Aspromonte, M. C. et al. DisProt in improving function annotation of intrinsically disordered proteins. *Nucleic Acids Res.* **52**, D434–D441 (2024).
63. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321. https://doi.org/10.1126/science.aal3321 (2017).
64. Kuhn, M. Building predictive models in *R* using the **caret** package. *J. Stat. Soft.* https://doi.org/10.18637/jss.v028.i05 (2008).

65. Svetnik, V. et al. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958. https://doi.org/10.1021/ci034160g (2003).
66. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297. https://doi.org/10.1007/BF00994018 (1995).
67. Cox, D. R. The Regression analysis of binary sequences. *J. Roy. Stat. Soc. Ser. B* **20**, 215–242 (1958).
68. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106. https://doi.org/10.1007/BF00116251 (1986).
69. Cover, T. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **13**, 21–27. https://doi.org/10.1109/TIT.1967.1053964 (1967).
70. Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **30**, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2 (1997).
71. Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: A misleading measure of the performance of predictive distribution models. *Glob. Ecol Biogeogr.* **17**, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x (2008).
72. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.* **45**, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002 (2009).
73. Dyer, S. C. et al. Ensembl 2025. *Nucleic. Acids Res.* **53**, D948–D957. https://doi.org/10.1093/nar/gkae1071 (2025).
74. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617. https://doi.org/10.1093/nar/gkae1010 (2025).
75. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. https://doi.org/10.1038/s41586-024-07487-w (2024).
76. Seeliger, D. & de Groot, B. L. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.* **24**, 417–422. https://doi.org/10.1007/s10822-010-9352-6 (2010).

Author contributions

Conceptualization, A.V. and M.A.A. Investigation, A.V. Methodology, all authors. Writing original draft and figures, A.V. All authors reviewed and edited the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. P.M. is supported by Beatriz Galindo Senior grant BG23/00060 (financed by the Spanish Ministry of Science, Innovation and Universities).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-19264-4.

Correspondence and requests for materials should be addressed to M.A.A.-N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 4

A Map of the Lipid-Metabolite-Protein Network to Aid Multi-Omics Integration

Uchenna Alex Anyaegbunam, **Aimilia-Christina Vagiona**, Vincent ten Cate, Katrin Bauer, Thierry Schmidlin, Ute Distler, Stefan Tenzer, Elisa Araldi, Laura Bindila, Philipp Wild and Miguel A. Andrade-Navarro

Article published in *Biomolecules, vol. 15, 4, 484. (26 March 2025),*

doi: https://doi.org/10.3390/biom15040484

biomolecules

MDPI

Article

A Map of the Lipid–Metabolite–Protein Network to Aid Multi-Omics Integration

Uchenna Alex Anyaegbunam [1], Aimilia-Christina Vagiona [1], Vincent ten Cate [2,3,4], Katrin Bauer [2,4,5], Thierry Schmidlin [6,7], Ute Distler [6,7], Stefan Tenzer [6,7], Elisa Araldi [2,4,5,8], Laura Bindila [9], Philipp Wild [2,3,4] and Miguel A. Andrade-Navarro [1,*]

1 Computational Biology and Data Mining Group (CBDM), Institute of Organismic and Molecular Evolution (iOME), Johannes Gutenberg University, 55122 Mainz, Germany
2 Preventive Cardiology and Preventive Medicine, Department of Cardiology, University Medical Center, Johannes-Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany
3 Clinical Epidemiology and Systems Medicine, Center for Thrombosis and Hemostasis (CTH), University Medical Center, 55131 Mainz, Germany
4 German Center for Cardiovascular Research (DZHK), Partner Site Rhine Main, University Medical Center, Johannes-Gutenberg University Mainz, 55131 Mainz, Germany
5 Computational Systems Medicine, Center for Thrombosis and Hemostasis (CTH), 55131 Mainz, Germany
6 Institute of Immunology, University Medical Center, Johannes-Gutenberg University Mainz, 55131 Mainz, Germany
7 Research Centre for Immunotherapy (FZI), University Medical Center, Johannes-Gutenberg University Mainz, 55131 Mainz, Germany
8 Systems Medicine Laboratory, Department of Medicine and Surgery, University of Parma, 43121 Parma, Italy
9 Institute of Physiological Chemistry, University Medical Center, 55131 Mainz, Germany
* Correspondence: andrade@uni-mainz.de

Abstract: The integration of multi-omics data offers transformative potential for elucidating complex molecular mechanisms underlying biological processes and diseases. In this study, we developed a lipid–metabolite–protein network that combines a protein–protein interaction network and enzymatic and genetic interactions of proteins with metabolites and lipids to provide a unified framework for multi-omics integration. Using hyperbolic embedding, the network visualizes connections across omics layers, accessible through a user-friendly Shiny R (version 1.10.0) software package. This framework ranks molecules across omics layers based on functional proximity, enabling intuitive exploration. Application in a cardiovascular disease (CVD) case study identified lipids and metabolites associated with CVD-related proteins. The analysis confirmed known associations, like cholesterol esters and sphingomyelin, and highlighted potential novel biomarkers, such as 4-imidazoleacetate and indoleacetaldehyde. Furthermore, we used the network to analyze empagliflozin's temporal effects on lipid metabolism. Functional enrichment analysis of proteins associated with lipid signatures revealed dynamic shifts in biological processes, with early effects impacting phospholipid metabolism and long-term effects affecting sphingolipid biosynthesis. Our framework offers a versatile tool for hypothesis generation, functional analysis, and biomarker discovery. By bridging molecular layers, this approach advances our understanding of disease mechanisms and therapeutic effects, with broad applications in computational biology and precision medicine.

Keywords: lipids; metabolites; proteins; cardiovascular diseases; multi-omics integration; lipid–metabolite–protein network; omics

Academic Editor: Jianlin Cheng

Received: 5 February 2025
Revised: 13 March 2025
Accepted: 20 March 2025
Published: 26 March 2025

Citation: Anyaegbunam, U.A.; Vagiona, A.-C.; ten Cate, V.; Bauer, K.; Schmidlin, T.; Distler, U.; Tenzer, S.; Araldi, E.; Bindila, L.; Wild, P.; et al. A Map of the Lipid–Metabolite–Protein Network to Aid Multi-Omics Integration. *Biomolecules* **2025**, *15*, 484. https://doi.org/10.3390/biom15040484

1. Introduction

The integration of multi-omics data offers immense potential for discovering intricate molecular mechanisms in biological systems [1,2]. For example, combining lipidomics and transcriptomics has identified critical interactions in prostate cancer, such as the specificity of sphingosine in distinguishing cancerous from benign conditions and its downstream signaling implications [3,4]. Similarly, the integration of proteomics with genomic and transcriptomic data has been instrumental in pinpointing potential cancer driver genes and their pathways, as shown in colon cancer studies [1,5]. However, practical challenges remain, including the heterogeneity of data types, variations in data quality, and the high dimensionality of omics datasets, which complicate data alignment and biological interpretation [6]. These challenges hinder the merging of diverse omics layers—such as proteomics, lipidomics, and metabolomics—into a unified framework [6–9]. Specific examples include difficulties in aligning multi-omics data due to different scales and formats, and limitations in the interpretability of integrated models when applied to toxicological studies [2,6]. Furthermore, the lack of standardization in experimental protocols and data preprocessing often results in sparsely connected datasets with reduced biological coherence [7].

These issues have been addressed with various strategies, such as early integration and hierarchical integration approaches, deep learning methods, and network-based frameworks [10–31]. For instance, early integration models combine omics data into a single representation, enabling the simultaneous analysis and identification of cross-modal patterns. Frameworks such as those proposed by [10,11] directly combine data from different omics layers, thereby enhancing predictive accuracy for tasks like biomarker discovery—a strategy comprehensively reviewed by [12]. Hierarchical integration models leverage the strengths of individual layers before integrating them in a stepwise manner to enhance interpretability and predictive power [12]. This approach, exemplified by dual-path graph attention auto-encoders [13] and unsupervised neural networks, like UMINT [14], preserves spatial or single-cell resolution while mitigating issues related to data sparsity and high dimensionality. Complementary frameworks, such as MUON [15], further refine this multi-tiered analysis. Deep learning approaches leverage adaptive training to each omics layer independently and transformer architectures to optimize cross-modality interaction learning to maximize data utility. Methods like CustOmics [16] and scmFormer [17] have been successfully applied to large-scale proteomic–transcriptomic integration. These approaches have not only improved predictive performance in complex phenotypes—as highlighted in studies on neurodegenerative diseases [18] and multimodal deep learning reviews [19]—but have also benefited from transfer learning strategies exemplified by scJoint [20]. Additionally, scCross [21] addresses omics integration problems by overcoming modality discrepancies, data scarcity in less robust modalities, and complex cross-modal alignment challenges, thereby enabling effective cross-modal data generation, simulation, and in silico cellular perturbation analysis.

Network-based frameworks exploit graph-theoretical models to capture the intricate interdependencies among molecular entities. Graph-linked embedding techniques, like GLUE [22], facilitate regulatory inference from single-cell data, while graph convolutional networks employed in MoGCN [23] offer refined analysis for cancer subtype classification. Hypergraph integration networks, such as MORE [24], enhance biomedical classification and biomarker identification, and executable network models [25] provide dynamic platforms for simulating multi-omics interactions. Additional frameworks, like TEMINET [26] and scBridge [27], further address cellular heterogeneity. TEMINET builds disease-specific networks from intra-omics features and then leverages graph attention networks within a multi-level framework to capture comprehensive collective information beyond simple

pairwise interactions, while scBridge iteratively selects scATAC-seq cells with minimal omics differences and integrates them with scRNA-seq.

In this study, we present a network embedded in hyperbolic space that integrates human lipids, metabolites, and proteins to aid multi-omics integrative research requiring the interpretation of the associations among these three particular molecular types. Accordingly, the nodes of the network are proteins, lipids, and metabolites. Edges connect proteins with interacting proteins, metabolites, and lipids. There are several algorithms and models supporting that the topology of complex systems, such as interactomes, is shaped by an underlying hidden hyperbolic geometry [32,33]. Modeling complex systems in hyperbolic geometry offers a novel way to analyze their hierarchical and clustering properties, which is more appropriate for densely connected networks and has implications for understanding how nodes interact and how disease effects propagate in networks [34,35].

To facilitate multi-omics integration, this network is used as a backend database for software implementation that considers three molecular types: proteins, lipids, and metabolites. Using this tool, a user can input a set of molecules of one of the three molecular types (for example, dysregulated proteins or proteins with a common biological function) and retrieve a ranked list of the other two molecular types (for example, lipids and metabolites). First, we demonstrate the utility of the tool for literature discovery from an omics layer in the context of studying a disease. Here, we focus on cardiovascular diseases, a major area of interest due to the complexity and interrelationships of lipid and metabolic dysregulation involved [36–40]. Secondly, we show how our method offers opportunities for functional enrichment analysis of lipidomic profiles.

2. Methods

2.1. Construction of a Protein, Lipid, and Metabolite Network

We started from the protein–protein interaction network of the HIPPIE database (version 2.3) [41]. This database assigns confidence scores (from 0, low, to 1, high) to interactions between human proteins based on the type and amount of experimental information available in biomedical publications [42]. In this study, only protein–protein interactions with a confidence score of $\geq$0.71 were kept (following [34]) because this selects a high percentage of interactions supported by at least two publications. The network's largest connected component (subset of connected nodes) was obtained. This consisted of 15,412 nodes (proteins) and 185,793 edges. Proteins not connected to this subset are not mappable by definition and were ignored.

Lipids (detectable in human plasma, as defined in [43]) related to the proteins in this initial network were then added using information from enzymatic reactions sourced from the SwissLipid database (967 interactions, [44]), and links from lipids to gene loci from a GWAS (652 interactions, [45]). This resulted in the addition of 940 lipids to the network.

In addition, metabolites (detectable in human plasma, according to [46]) related to proteins were then added using enzymatic reaction information from PubChem (1471 interactions, [47]). This resulted in the addition of 273 metabolites to the network. In total, the network consists of 16,625 nodes (Supplementary Table S1) and 188,883 edges (Supplementary Table S2).

2.2. Hyperbolic Embedding of the Network

To facilitate visual interpretation and to reflect the overall connectivity of the network, we applied hyperbolic mapping techniques. We embedded the multi-omics network in the two-dimensional hyperbolic plane using the R package "NetHypGeom" (version 1.0), which implements the LaBNE + HM algorithm [48]. This approach combines maximum likelihood estimation and manifold learning to decipher the underlying hyperbolic ge-

ometry of complex networks [49,50]. The popularity–similarity (PS) model provides a geometric interpretation in hyperbolic space (H^2) and assumes that the clustering and hierarchy of complex networks arise from trade-offs between the popularity and similarity of nodes [51]. In this embedding, the hyperbolic distance reflects their similarity, such that closer pairs of nodes in the hyperbolic space are more likely to have connections in the network [50–52]. The network was embedded in H^2 to infer the hyperbolic coordinates of each node, with parameters γ = 2.98, T = 0.84, and w = 2π. In the embedding, the 16,625 nodes of the multi-omics network lie within a hyperbolic disk where the radial coordinate of a node, r_i, represents the popularity dimension with nodes that joined the system first being close to the disk's center. The angular coordinate of a node, θ_i, represents the similarity dimension. The hyperbolic coordinates (r and θ) were used to compute the hyperbolic distance between nodes in the hyperbolic map. This representation can be interpreted as molecular relationships within the context of specific biological processes.

2.3. Clustering in the Angular Similarity Dimension

To cluster nodes in the similarity dimension, we computed the difference between consecutive angular coordinates to identify gaps. The nodes were sorted according to their inferred angles θ in increasing order, and the difference between θ_i and θ_{i+1} was computed to identify the largest gaps between clusters in the similarity dimension. To determine the start and the end of each cluster, we chose gap sizes that produced clusters with a minimum number of 10 protein members because this allowed us to perform meaningful functional enrichment analysis of each sector.

2.4. Evaluation of Molecular Relationships Between Omics Layers

The hyperbolic network offers the possibility of measuring distances (in hyperbolic space) among all nodes (proteins, lipids, and metabolites), which can be interpreted as functional relationships. We implemented a user-friendly Shiny R (version 1.10.0) software package ((https://github.com/uchealex/Omint) (accessed on 24 March 2025)) that ranks molecules between omics layers based on their hyperbolic distances. The method uses as input a dataset of molecules of one of the three molecular types (for example, a list of differentially expressed proteins), and provides a list of all of the molecules in each of the other two molecular types (for example, lipids or metabolites), ranked according to their distances to the input dataset.

The ranking score is generated by taking the reciprocal of the sum of the n smallest distances (n = 3, by default) between the nodes corresponding to the user-defined subset (for example, proteins) and the nodes of the other two molecular types (for example, lipids or metabolites). Therefore, higher scores indicate stronger associations between the nodes of the subset and the nodes of the other molecular types. The algorithmic implementation of our method is based on the pseudo-code shown in Box 1.

2.5. Functional Enrichment Analyses

We carried out Gene Ontology (GO) enrichment analysis for the proteins in each sector of the multi-omics network, using the proteins of the complete network as a background set. For functional enrichment analysis, we used the Enrichr R Package (version 3.4). Enrichr is a comprehensive tool that provides access to multiple gene set libraries and databases, enabling the identification of significantly enriched biological pathways and functional terms [53]. Only GO biological process (BP) terms enriched at a significance level (p-value, multiple testing corrected) of 0.05 or less were retained. A similar methodology was used to compute enrichment analysis for sets of proteins associated with empagliflozin treatment.

Box 1. Algorithm for cross-omics association ranking using hyperbolic geometry. This schematic outlines the computational workflow for identifying molecular associations between omics layers. The algorithm (1) processes user-input molecular identifiers of one omics layer, (2) calculates hyperbolic distances between all valid input molecules and all items in the other (non-input) layers, (3) computes association scores as the reciprocal sum of the n smallest distances to the items in each of the non-input layers (default n = 3), and (4) generates ranked lists of associated molecules in the non-input layers with evidence trails.

Algorithm: Cross-omics Relationship Ranking via Hyperbolic Distance
Input:

- User-defined molecule subset (S): List of identifiers of one omics layer (proteins, lipids, or metabolites)
- Hyperbolic coordinates dataset: Precomputed (r, θ) for all molecules across layers
- n: Number of smallest distances to consider (default = 3)

Output:

- Ranked lists of molecules of non-input omics layers, sorted by association strength

Procedure:

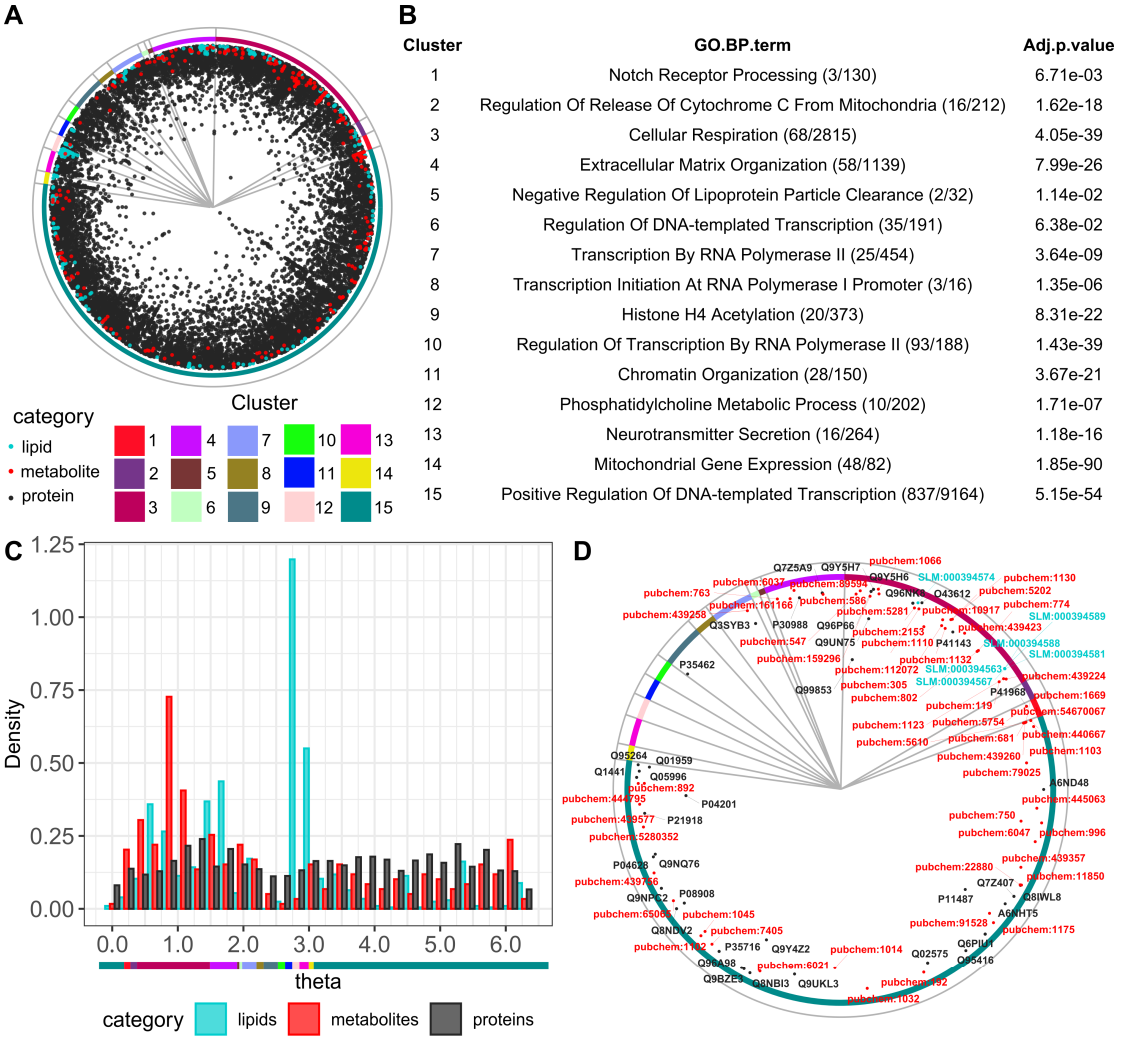
1. Input Processing:
   a. Receive user input (S) and parameter n
   b. Validate identifiers in S against reference database
   c. Filter valid subset (S_valid) = S ∩ database_entries
2. Distance Matrix Construction:
   For each target omics layer T ∉ input layer:
   a. Initialize distance matrix D with dimensions $|S\_valid| \times |T|$
   b. Compute hyperbolic distances between all pairs (s ∈ S_valid, t ∈ T):
   $distance(s,t) = \operatorname{acosh}[\cosh(r_s)\cosh(r_t) - \sinh(r_s)\sinh(r_t)\cos(\Delta\theta)]$
   where $\Delta\theta = \pi - |\pi - |\theta_s - \theta_t||$
3. Association Score Calculation:
   For each molecule t in target layer T:
   a. Collect all distances from S_valid to t: {$distance(s_1,t), \ldots, distance(s_k,t)$}
   b. Identify n smallest distances: $d_1 \leq d_2 \leq \ldots \leq d_n$
   c. Compute association score: $score(t) = 1/(\Sigma^{n}_{\{i=1\}} d_i)$
4. Ranking and Output:
   a. Sort all molecules in T by descending score(t)
   b. Generate evidence strings for top associations:
   evidence(t) = [si:di (sorted)] for i = 1...n
   c. Return ranked list: (t, score(t), evidence(t)) ∀ t ∈ T

3. Results

To provide a holistic view of molecular interactions across multiple omics layers, including lipids, metabolites, and proteins, we constructed a network by combining public data describing experimental interactions among them. The network was embedded in hyperbolic space (see Section 2 for details). By using hyperbolic mapping, we enable researchers to explore molecular relationships in a visually intuitive and biologically meaningful way.

In hyperbolic space, nodes have polar coordinates (r, θ). Nodes with shorter distances to the center (low r values) correspond to nodes with higher connectivity. The similarity in the θ coordinates between nodes reflects their similar interactions with other nodes [32,34,35]. To exploit the biological meaning of the θ coordinates in the multi-omics network, we clustered the elements of the network into groups based on the identified gaps between consecutive angles and determined 15 clusters (see Section 2 for details; Figure 1A). To evaluate the functions represented in each sector, we determined the enrichment in GO biological process terms in the proteins of each cluster (Figure 1B). The distribution of proteins, lipids, and metabolites in the angular dimension highlights how

lipids and metabolites agglomerate in similarity-based clusters. For example, the angular distribution of lipids has a maximum of around $\theta = 2.8$, which corresponds to cluster 12. Accordingly, cluster 12 is enriched in proteins annotated with the term Phosphatidylcholine Metabolic Process, indicating the biological significance of the similarity component of the map (Figure 1C). The angular distribution of metabolites has a maximum of around $\theta = 1.0$ in cluster 3, which is enriched in proteins annotated with the term Cellular Respiration; this reflects the central role of cellular respiration in cellular metabolism (Figure 1C).



Cluster	GO.BP.term	Adj.p.value
1	Notch Receptor Processing (3/130)	6.71e-03
2	Regulation Of Release Of Cytochrome C From Mitochondria (16/212)	1.62e-18
3	Cellular Respiration (68/2815)	4.05e-39
4	Extracellular Matrix Organization (58/1139)	7.99e-26
5	Negative Regulation Of Lipoprotein Particle Clearance (2/32)	1.14e-02
6	Regulation Of DNA-templated Transcription (35/191)	6.38e-02
7	Transcription By RNA Polymerase II (25/454)	3.64e-09
8	Transcription Initiation At RNA Polymerase I Promoter (3/16)	1.35e-06
9	Histone H4 Acetylation (20/373)	8.31e-22
10	Regulation Of Transcription By RNA Polymerase II (93/188)	1.43e-39
11	Chromatin Organization (28/150)	3.67e-21
12	Phosphatidylcholine Metabolic Process (10/202)	1.71e-07
13	Neurotransmitter Secretion (16/264)	1.18e-16
14	Mitochondrial Gene Expression (48/82)	1.85e-90
15	Positive Regulation Of DNA-templated Transcription (837/9164)	5.15e-54

Figure 1. Visualization and functional characterization of the human protein–lipid–metabolite interaction network. (**A**) Network integration of multi-omics data embedded in the hyperbolic space (see Section 2 for details). Node colors represent different molecular classes (proteins in black, lipids in cyan, metabolites in red). The circular sectors indicate clusters (numbered 1 to 15) identified by the gaps between consecutive nodes in the angular dimension of the hyperbolic space (see Section 2 for details). Node coordinates are available in Supplementary Table S1. Edges (not represented) are available in Supplementary Table S2). (**B**) The over-represented biological function in each cluster was determined via GO enrichment analysis (BP: biological process) to reveal the biological relevance based on the term with the lowest adjusted *p*-value of enrichment (see Section 2 for details). The numbers in parentheses indicate how many proteins had the indicated GO term out of all the proteins in the respective cluster. (**C**) The distribution of θ coordinates for each omics layer: proteins, lipids, and metabolites. The different maxima positions of lipids and metabolites in the angular dimension indicate that they are grouped in different clusters. The angular coordinates of the 15 clusters are indicated with a colored bar for reference under the plot. (**D**) Brain-related proteins, lipids, and metabolites in the hyperbolic map (identifiers from UniProt, SLM, and PubChem, respectively).

To test the coherence across omics layers, we annotated lipids (using SwissLipids [44]) and metabolites (using HMDB [54]) present in the brain and proteins specifically expressed in the brain (according to all brain tissues represented in GTEx [55]). This resulted in a selection of 6 lipids, 55 metabolites, and 99 proteins. Their distributions overlapped across clusters and were absent in a few clusters on the top left part of the map (Figure 1D). The lipids are grouped in the cellular respiration cluster, while proteins and metabolites are more widely distributed.

Interestingly, while cluster 13 is enriched in neurotransmitter secretion proteins, this cluster is devoid of brain-specific molecules. This reflects the fact that the machinery for vesicle fusion and regulated secretion uses proteins that are expressed in neural and non-neural tissues. For example, one of these proteins is Unc-13 homolog A (UNC13A, a.k.a. SYT14L). While this protein has neural functions and expression, and has been shown to extend dendrite length, it also regulates melanocyte differentiation [56]. Another example is synaptotagmin-1 (SYT1), which is well-known for neural functions but has also been associated with endocytosis in pancreatic beta-cells [57]. These observations indicate that our map can be used to obtain useful insight into biological function by leveraging known associations among proteins, lipids, and metabolites.

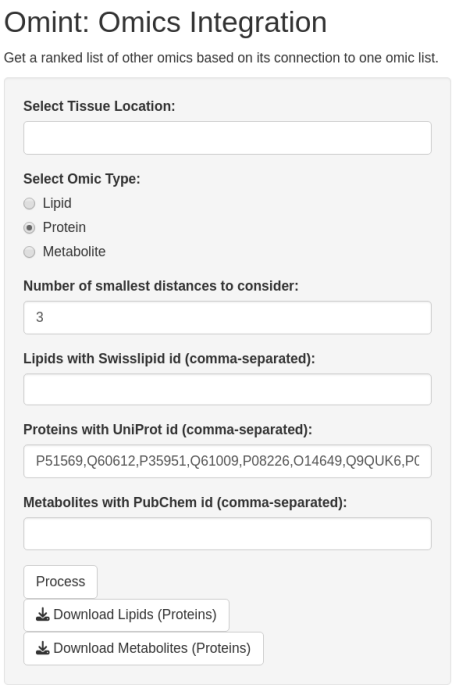
3.1. Software Implementation

To take advantage of the possibilities that our map offers to exploit biological information connected across proteins, lipids, and metabolites, we developed a user-friendly software implementation that utilizes the Shiny R framework to allow users to input a list of molecules of one molecular type (subset) and retrieve ranked lists of the other two molecular types according to their distances to the input subset in the hyperbolic map (Figure 2). For instance, a user can input a list of proteins and retrieve ranked lipids or metabolites based on their proximity to the input proteins in the network (see Section 2 for details).

To use the software, the user first indicates what type of molecule is provided as the input. Secondly, the user chooses the number of closest molecules from the user-defined subset to be used to rank the other sets (the default number used here was three). Finally, the user inputs a list of molecules (subset) and clicks the "process" button to produce and display ranked lists of each of the other two molecular types. These lists can be downloaded by clicking the corresponding "download" buttons.

Optionally, the user can filter the analysis for proteins expressed in a selection of tissues. For this, we systematically annotated proteins for their messenger RNA (mRNA) tissue-specific expression using data from the Human Protein Atlas Database [58], considering expression profiles across 50 tissues. Specifically, each protein of the network was mapped to its corresponding tissue-specific RNA expression profile, setting a cut-off of 10 nTMP (normalized transcripts per million). This value is generally considered a moderate to high expression level, reducing false positive cases and improving the biological significance of tissue annotations.

The software we developed enhances the utility of our protein–lipid–metabolite network by allowing researchers to explore multi-omics connections, enabling hypothesis generation and the identification of potential biomarkers. In the next sections, we illustrate the use of the method with two use cases.



Lipid	score	evidence
SLM:000000470	1.672e-02	P55157:9.26936, P05091:22.06172, Q16635:28.46229
SLM:000390712	1.648e-02	P02649:8.03236, P02652:24.27141, P02647:28.3725
SLM:000087698	1.533e-02	P02649:12.80965, P02652:24.18225, P02647:28.25481
SLM:000489933	1.520e-02	P02649:13.19065, P02652:24.24332, P02647:28.3725
SLM:000488109	1.495e-02	P02649:14.30232, P02652:24.22607, P02647:28.37799
SLM:000394588	1.490e-02	P02649:14.42331, P02652:24.3299, P02647:28.37368
SLM:000390694	1.476e-02	P02649:15.17792, P02652:24.1931, P02647:28.37493
SLM:000500359	1.474e-02	P02649:15.14028, P02652:24.34733, P02647:28.36761
SLM:000088786	1.463e-02	P02649:15.75941, P02652:24.19339, P02647:28.40246
SLM:000000239	1.460e-02	P55157:19.49757, P05091:21.05302, Q16635:27.94602
SLM:000391259	1.457e-02	P02649:16.10565, P02652:24.14574, P02647:28.37803
SLM:000394589	1.451e-02	P02649:16.15284, P02652:24.39682, P02647:28.36692
SLM:000035988	1.447e-02	P05091:17.58681, P55157:22.96174, Q16635:28.57194
SLM:000067899	1.446e-02	P05091:17.61498, P55157:22.96166, Q16635:28.57473

Metabolite	score	evidence
pubchem:800	1.496e-02	P05091:16.63596, P55157:22.31284, Q16635:27.89449
pubchem:751	1.495e-02	P15121:10.95172, P19320:26.26761, P10275:29.66154
pubchem:96215	1.490e-02	P05091:16.09958, P55157:22.86455, Q16635:28.14591
pubchem:802	1.437e-02	P05091:19.39229, P55157:22.03031, Q16635:28.15948
pubchem:11850	1.429e-02	P15121:13.49447, P19320:26.53788, P10275:29.93165
pubchem:5780	1.393e-02	P15121:15.85999, P19320:26.27804, P10275:29.67148
pubchem:1150	1.384e-02	P55157:21.16552, P05091:22.67245, Q16635:28.40491
pubchem:774	1.382e-02	P55157:21.41934, P05091:22.66518, Q16635:28.28343
pubchem:2519	1.358e-02	P10632:14.41711, P04114:29.57628, P51589:29.66344
pubchem:1826	1.341e-02	P55157:23.03036, P05091:23.31383, Q16635:28.22797

Figure 2. Software implementation for omics integration. Our method was implemented using the Shiny R framework to allow users to input a list of molecules of one type (proteins, lipids, or metabolites) and retrieve molecules of the other two molecular types ranked by their distances to the input subset in the hyperbolic map.

3.2. Literature Discovery: Cardiovascular Disease Case Study

It is well known that cardiovascular diseases (CVDs) are accompanied by the dysregulation of certain proteins, lipids, and metabolites. To find lipids and metabolites with relevance to CVD, we used a list of proteins selected for their importance in the literature on lipid research in CVD by [43] (Supplementary Table S4) as input to our software to derive a ranked list of lipids and metabolites closely associated with these proteins (Supplementary Tables S5 and S6) (Figure 3A). In short, the proteins were selected for their enzymatic connections to lipids associated with CVD, as addressed in PubMed records with a focus on physiological research on humans and mice (mentioning plasma, heart, or myocardium).

Table 1. Lipids associated with CVD proteins. This table summarizes the results of testing our method using proteins linked to CVD, identifying lipids associated with these conditions. The top four lipids identified through our analysis have been cited more than 600 times in relation to CVD in the scientific literature. The publications were obtained from PubMed using a Boolean search of each lipid in conjunction with the term "cardiovascular disease" (example: "cholesterol esters" AND "cardiovascular disease").

Lipid	Score	Name	Number of Publications
SLM:000000470	0.01672	Cholesterol esters	184
SLM:000390712	0.01648	Sphingomyelin	121
SLM:000087698	0.01533	Phosphatidylcholine	295
SLM:000489933	0.01520	Ganglioside	15

Table 2. Metabolite associations with CVD proteins. This table summarizes the results of testing our method using proteins linked to CVDs, identifying metabolites associated with these conditions. The top ten metabolites identified through our analysis have been cited over 850 times in relation to CVD in the scientific literature. The publications were obtained from PubChem by filtering the literature associated with a given metabolite (in the Literature section of the metabolite record in PubChem) using the search term "cardiovascular disease".

Metabolite	Score	Name	Number of Publications
Pubchem:800	0.01496	Indoleacetaldehyde	2
Pubchem:751	0.01495	Glyceraldehyde	35
Pubchem:96215	0.01490	4-imidazoleacetate	1
Pubchem:802	0.01437	Indole-3-acetate	24
Pubchem:11850	0.01429	Galactitol	4
Pubchem:5780	0.01393	Sorbitol	46
Pubchem:1150	0.01384	Tryptamine	7
Pubchem:774	0.01382	Histamine	166
Pubchem:2519	0.01358	Caffeine	562
Pubchem:1826	0.01341	5-hydroxyindoleacetate	10

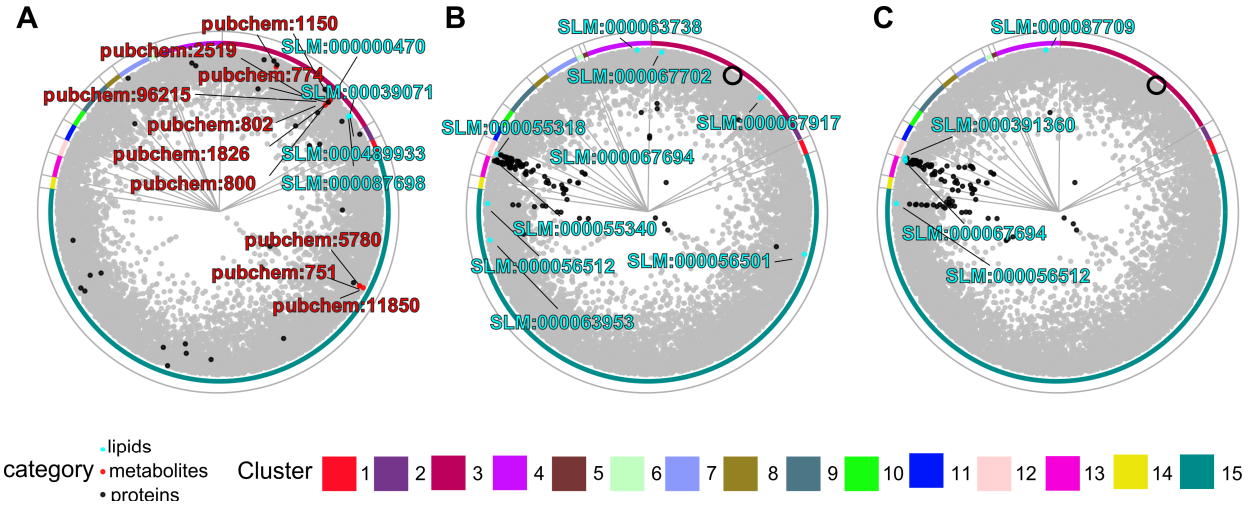


Figure 3. Application of the hyperbolic map to infer associations among molecular types. Different views of the human protein–lipid–metabolite interaction network (nodes in grey) highlighting (**A**) a subset of proteins related to the literature investigating lipids in CVD, used as input to score associated lipids (top four shown; cyan; see also Table 1; full data in Supplementary Table S4) and metabolites (top ten shown; red; see also Table 2; full data in Supplementary Table S5), and (**B**,**C**) subsets of lipids (cyan) that were dysregulated at two time points of empagliflozin treatment (one and twelve weeks, respectively), used as input to obtain related proteins (top 100 shown, black; full data in Supplementary Table S6; the black empty circle indicates the position of SGLT2, the protein inhibited by empagliflozin).

The four top-ranked lipids (cholesterol esters, sphingomyelin, phosphatidylcholine, and ganglioside) had over 600 citations in publications relating to CVD (Table 1). The ten top-ranked metabolites had over 850 citations in publications in relation to CVD (Table 2). The metabolites with the most citations were caffeine and histamine. Altogether, our method is able to identify lipids and metabolites that have been previously implicated in pathways relevant to cardiovascular health. Furthermore, highly ranked metabolites

with low literature presence, such as 4-imidazoleacetate and indoleacetaldehyde, could be unknown potential biomarkers of CVD.

3.3. Functional Enrichment Analysis of Lipid Signatures

Empagliflozin is an antidiabetic drug that works as an inhibitor of the sodium-glucose co-transporter-2 (SGLT2) [59]. Lipid signatures of treatment with empagliflozin at two time points (one week and twelve weeks after treatment) are available from the EmDia study, a placebo-controlled randomized phase 3 clinical study on the effect of empagliflozin on left ventricular diastolic function [60,61]. To add information on the drug's temporal effects, we used as input to our tool the list of one- and twelve-week empagliflozin signature lipids present in our map (Supplementary Tables S7 and S8) to score human proteins for their association with them (Supplementary Tables S9 and S10). The map positions of the signature lipids and the top 100 scored proteins (cyan and black, respectively; Figure 3B,C) suggest a shift where, at the one-week time point, some dysregulated lipids (and some of the top scoring proteins) are located in the same or neighboring clusters as SGLT2 in the map (black circle, Figure 3B,C), whereas, at the twelve-week time point, most of the mapped dysregulated lipids are close to the clusters of the top-scoring proteins on the left of the map.

To interpret this temporal shift in protein function, the top 100 proteins at each time point (Figure 3B,C) were used for GO enrichment analysis (Figure 4; Supplementary Tables S11 and S12). For the one-week time point, the primary enriched terms were phosphatidylcholine metabolic process, phosphatidylethanolamine metabolic process, and sphingosine metabolic process. These pathways highlight the initial biochemical shifts triggered by empagliflozin treatment, with early alterations emphasizing general phospholipid metabolism.

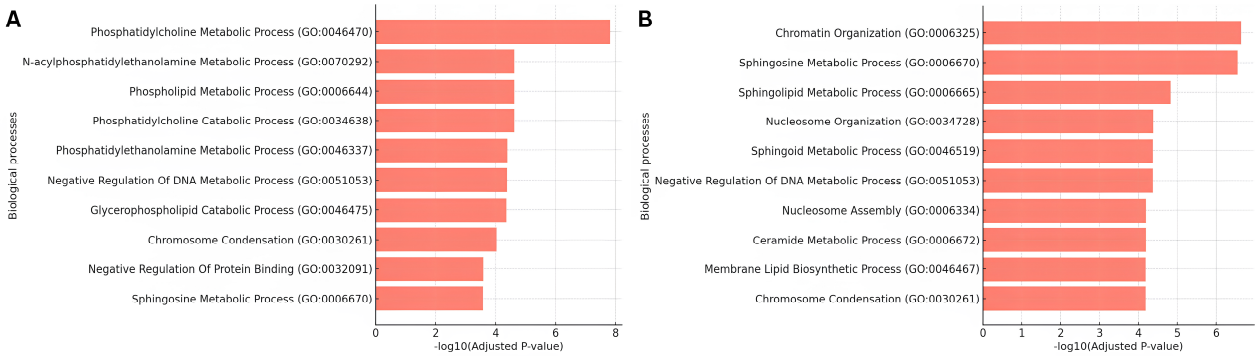


Figure 4. Temporal Gene Ontology (GO) enrichment analysis of lipid signatures following empagliflozin treatment. (**A**) Enriched GO biological processes for proteins associated with lipid signatures at one week and (**B**) twelve weeks after empagliflozin treatment, showing shifts from phospholipid metabolism to sphingolipid and ceramide metabolism. Functional enrichment was carried out using the top 100 proteins derived from lipid signatures. Full lists are given in Supplementary Tables S7 and S8, respectively.

By twelve weeks, the enrichment profile shifted to emphasize the sphingosine metabolic process and the ceramide metabolic process. The sustained enrichment of sphingosine metabolism at both time points underscores its central role in the long-term effects of empagliflozin. This aligns with experimental findings in the EmDia study [60,61] showing an increase in sphingomyelins and ceramides, which are integral to sphingosine metabolism. The emergence of ceramide metabolism at twelve weeks does not only suggest an increase in ceramides, but also a progressive impact on sphingolipid biosynthesis, reflecting adaptive or downstream metabolic responses to prolonged treatment. These

findings highlight the potential of integrating temporal lipidomic data with the proteomic layer to achieve a comprehensive understanding of drug effects and metabolic adaptations.

In addition, other pathways not strictly related to lipid synthesis and metabolism were observed. In particular, we observed enrichment in genes related to the negative regulation of DNA metabolic process and chromosome organization and condensation at both times (e.g., TERF1, ENPP7, H3C1, and histones H1-5, H1-4, H1-1, and H1-0), while genes enriched in terms related to nucleosome and assembly organization and chromatin organization appeared at the second time point (e.g., H2AX, H2BC9, RBBP4, and H2BC21). These results show potential additional mechanisms that lead lipid-mediated empagliflozin effects. Our method provides directions of research for these results. For example, we can see the signature lipids that led to the selection of gene RBBP4 (UniProt AC Q09028), which ranked better in the second time point than in the first one (compare Supplementary Tables S9 and S10). Of the genes selected among the top 100 at the second time point but not at the first, many are in cluster 11 (chromatin organization; see Supplementary Table S1 and Figure 3C), and RBBP4 is one of them. These results could point to an epigenetic effect of empagliflozin aided by lipids.

Our method points to proteins associated with the lipid signatures using the map geometry, which is itself based on network connectivity. Therefore, it is possible to further substantiate the associations by examining the connections in the network (provided in Supplementary Table S2) with external tools (for example, Cytoscape [62]) to study the paths connecting associated lipids and proteins. For example, in this case study, the protein with the best association score at the second time point (serine palmitoyltransferase 3, UniProt AC Q9NUV7) had already been ranked in the top hundred at the first time point, but then the presence of ceramide (d43:1) (SLM:000391360) in the lipid signature at the second time point pushed it to the top because the protein is directly connected to the lipid by a GWAS interaction [45].

4. Discussion and Conclusions

Multi-omics data integration has emerged as a transformative approach to understanding complex molecular mechanisms and their dynamic changes in biological systems. In this study, we present a lipid–metabolite–protein network that leverages hyperbolic mapping to provide an intuitive and biologically meaningful representation of molecular interactions across omics layers. Coupled with a user-friendly software implementation, our framework enhances the ability to explore multi-omics connections and identify biologically relevant associations, with demonstrated utility in literature discovery and functional analysis. The method was implemented as a Shiny application freely available online in a GitHub repository [63]. Users can upload omics datasets (e.g., SwissLipid, UniProt, or PubChem IDs) and obtain ranked results based on hyperbolic distance metrics. The implementation supports interactive processing, detailed result visualization, and downloadable CSV files for further analysis. The software is compatible with R ($\geq$4.0.0) and requires only Shiny and dplyr libraries. Input customization is supported through configurable file paths. The results include ranked lists of related entities and valid subsets for each omics type, presented in an intuitive interface.

4.1. Multi-Omics Integration for Disease Insights

Our case study on cardiovascular diseases (CVDs) shows the value of combining different types of biological data with our method. We found lipids and metabolites linked to heart disease-related proteins, many of which are already well-known (like cholesterol and caffeine). We also found less-studied molecules (like 4-imidazoleacetate) that could

be new targets for future research. Our method both confirms what we already know and suggests new avenues to explore.

With a second case study, a temporal analysis of empagliflozin's lipid signatures, we illustrated how to use our method to study how the drug affects the body. Early on, it mainly changes how certain fats (phosphatidylcholine and phosphatidylethanolamine) are processed. Later, it increasingly affects sphingolipid production, specifically increasing sphingomyelins and ceramides. This suggests sphingosine metabolism is key to the drug's long-term effects. Future research could combine lipid and protein data to see how these changes influence other processes like inflammation, giving us a better understanding of how empagliflozin works.

The two topics used above as examples of application of our approach were specifically chosen because they are related to clinical research that benefits from the determination of functional associations across proteins, lipids, and metabolites. The method can be applied with complete generality to any situation where a researcher has a set of molecules (proteins, lipids, or metabolites) for which similar investigation requirements hold. Naturally, if the input data are enriched in functions and pathways affected in the condition studied, the spatial organization of the map is expected to provide nodes related to similar functions in the other molecular layers. Ultimately, a researcher can examine the network connectivity in detail, for example, to find the connections between a particular protein highlighted by a lipid signature used as input and the lipids in the signature.

4.2. Limitations of Our Approach

Our approach is based on the assumption that there is an underlying network of connections among proteins, lipids, and metabolites that is condition-independent, and that parts of this network are activated in different cell types and situations. We are aware that we do not yet fully know this underlying network because there are connections that have not been observed experimentally and because the network does not include all regulatory mechanisms. Regardless, we have shown that it is useful to provide a network linking proteins, lipids, and metabolites, even if limited by incomplete data, as it provides the opportunity to relate expression patterns among these molecular types.

4.3. Advantages and Future Directions

We have provided the first method that facilitates analyses seeking to integrate datasets of proteins, lipids, or metabolites, in a transparent manner, using a unique map in which functionally related molecules occupy nearby positions. We have implemented software to use this map with protein, lipid, or metabolite datasets provided by researchers. We have shown that our tool enables biological relevant discoveries when it is necessary to explore connections across these molecular layers, particularly in the context of evaluating experimental data.

Our approach addresses several challenges inherent to multi-omics integration. The use of hyperbolic mapping provides an intuitive way to visualize and interpret complex networks, while the software implementation facilitates the exploration of molecular associations across omics layers. By combining these capabilities, our framework offers a robust platform for studying disease mechanisms and drug effects, and for biomarker discovery.

Unlike frameworks like GLUE [22] or scCross [21], which focus on broad multi-omics alignment (e.g., transcriptomics and chromatin accessibility), our strategy specifically uses enzymatic reactions and genetic associations from domain-specific databases. This bypasses the limitations of methods reliant on paired single-cell data or generic pathway maps, enabling the direct interrogation of biochemical interactions often overlooked in broader analyses, and circumventing alignment challenges posed by non-overlapping molecular lay-

ers in conventional multi-omics datasets. On the aspect of enrichment analysis, traditional enrichment tools, like MOPA or ActivePathways, rely on pathway-level statistics or Euclidean distances, which may fail to capture hierarchical dependencies [64,65]. Directional methods, like DPM, penalize inconsistent omics interactions (e.g., inverse mRNA–protein correlations) but lack spatial context. Our approach instead quantifies connectivity through hyperbolic embeddings, enabling nuanced insights into how molecular types interact (e.g., proximity in metabolic networks) rather than whether they align directionally [66].

In our applications illustrating how our method can be used to obtain biological insights, we made some pragmatic choices, such as combining information from three neighbors to score the relationships between different omics layers, or focusing on the top 100 scoring proteins when evaluating the lists of proteins ranked by lipid signatures: our method is flexible in how these thresholds are used, and we envision that different applications might allow exploratory analyses, while others might require statistical support for choosing thresholds. Our method offers the possibility to easily change these thresholds and evaluate the effects on the results. We understand that facilitating the application of our method to different biological and clinical scenarios requires this flexibility and transparency, and with the examples given, we have provided guidelines to achieve this successfully.

Future work could expand our framework by integrating additional regulatory layers, such as miRNAs targeting transcripts, or transcription factor interactions with genes, to capture a more comprehensive view of molecular interactions. Additionally, applying our approach to other disease contexts or therapeutic interventions could validate its generalizability and uncover new biological insights. Finally, incorporating deep learning methods could further improve the performance of our method, as exemplified by other integrative omics approaches, such as CustOmics [16] and scmFormer [17].

In conclusion, our lipid–metabolite–protein network and associated software implementation provide a powerful means for multi-omics integration, offering novel insights into disease biology and drug effects. By bridging the gap between molecular layers, this approach facilitates the discovery of functionally relevant associations, enables hypothesis generation, and supports the identification of potential biomarkers. As multi-omics continues to evolve, frameworks like ours will play a critical role in advancing precision medicine and systems biology.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/biom15040484/s1, Table S1. Nodes of the multi-omics network. Columns indicate the protein, lipid, or metabolite identifiers, the polar coordinates of the node in the map, the cluster number, and whether it was associated with the brain (see Section 2 for details). Table S2. Edges of the multi-omics network. Columns indicate the protein, lipid, or metabolite identifiers, the category of each node, and the database that each edge is obtained from. Table S3. Protein tissue annotations. Columns indicate the protein identifier, the tissue name, and the nTPM. Obtained from [58]. Table S4. Proteins relevant for research on lipids in CVD. UniProt identifiers used as input for the analysis presented in Section 3.2. Literature discovery: Cardiovascular disease case study. Obtained from [43]. Table S5. Lipids scored for associations with CVD proteins. Columns indicate lipid identifiers, scores, and evidence (three closest proteins and distances to them). Table S6. Metabolites scored for associations with CVD proteins. Columns indicate metabolite identifiers, scores, and evidence (three closest proteins and distances to them). Table S7. Lipids dysregulated after one week of treatment with empagliflozin. Used as input in Section 3.3. Functional enrichment analysis of lipid signatures. Columns indicate lipid names and SwissLipids IDs. Obtained from the EmDia study [60,61]. Note that only 9 were in our map (Figure 3B). Table S8. Lipids dysregulated after twelve weeks of treatment with empagliflozin. Used as input in Section 3.3. Functional enrichment analysis of lipid signatures. Columns indicate lipid names and SwissLipids IDs. Obtained from the

EmDia study [60,61]. Note that only 3 were in our map (Figure 3C). Table S9. Proteins scored by their associations with lipids dysregulated after one week of treatment with empagliflozin. Columns indicate protein identifiers, scores, and evidence (three closest lipids and distances to them). Table S10. Proteins scored by their associations with lipids dysregulated after twelve weeks of treatment with empagliflozin. Columns indicate: protein identifiers, scores, and evidence (three closest lipids and distances to them). Table S11. Enriched biological processes (GO BP) for lipid signatures at one week after empagliflozin treatment. Columns indicate GO BP terms, overlap (genes in set versus all possible), *p*-values, adjusted *p*-values, odds ratios, combined scores, and genes. Table S12. Enriched biological processes (GO BP) for lipid signatures at twelve weeks after empagliflozin treatment. Columns indicate GO BP terms, overlap (genes in set versus all possible), *p*-values, adjusted *p*-values, odds ratios, combined scores, and genes.

Author Contributions: Conceptualization, M.A.A.-N.; data curation, U.A.A., A.-C.V. and M.A.A.-N.; formal analysis, U.A.A., A.-C.V. and M.A.A.-N.; funding acquisition, P.W. and M.A.A.-N.; investigation, U.A.A., A.-C.V. and M.A.A.-N.; methodology, U.A.A., A.-C.V. and M.A.A.-N.; resources, U.A.A., A.-C.V., K.B., T.S., E.A., L.B. and M.A.A.-N.; software, U.A.A., A.-C.V. and M.A.A.-N.; supervision, M.A.A.-N.; validation, U.A.A., V.t.C., K.B., T.S., U.D., S.T., E.A. and M.A.A.-N.; visualization, U.A.A., A.-C.V. and M.A.A.-N.; writing—original draft, U.A.A., A.-C.V. and M.A.A.-N.; writing—review and editing, U.A.A., A.-C.V., V.t.C., K.B., T.S., U.D., S.T., E.A., L.B., P.W. and M.A.A.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Federal Ministry of Education and Research: 03ZU1202AB; Federal Ministry of Education and Research: 03ZU1202EC; Federal Ministry of Education and Research: Forschungskernen für Massenspektrometrie in der Systemmedizin (MSCoreSys).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data necessary to reproduce our results is provided with this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vasaikar, S.; Huang, C.; Wang, X.; Petyuk, V.A.; Savage, S.R.; Wen, B.; Dou, Y.; Zhang, Y.; Shi, Z.; Arshad, O.A.; et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **2019**, *177*, 1035–1049.e19. [CrossRef] [PubMed]
2. Sun, Y.V.; Hu, Y.J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv. Genet.* **2016**, *93*, 147–190. [CrossRef] [PubMed]
3. Linda, A. Heffernan-Stroud, Lina, M. Obeid, Chapter Seven—Sphingosine Kinase 1 in Cancer. In *Advances in Cancer Research*; Norris, J.S., Ed.; Academic Press: Cambridge, MA, USA, 2013; Volume 117, pp. 201–235, ISBN 9780123942746. [CrossRef]
4. Xu, X.; Liu, B.; Zou, P.; Zhang, Y.; You, J.; Pei, F. Silencing of LASS2/TMSG1 enhances invasion and metastasis capacity of prostate cancer cell. *J. Cell. Biochem.* **2014**, *115*, 731–743. [CrossRef]
5. Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M.C.; Zimmerman, L.J.; Shaddox, K.F.; Kim, S.; et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, *513*, 382–387. [CrossRef] [PubMed]
6. Canzler, S.; Schor, J.; Busch, W.; Schubert, K.; Rolle-Kampczyk, U.E.; Seitz, H.; Kamp, H.; von Bergen, M.; Buesen, R.; Hackermüller, J. Prospects and challenges of multi-omics data integration in toxicology. *Arch. Toxicol.* **2020**, *94*, 371–388. [CrossRef]
7. Wörheide, M.A.; Krumsiek, J.; Kastenmüller, G.; Arnold, M. Multi-omics integration in biomedical research—A metabolomics-centric review. *Anal. Chim. Acta* **2021**, *1141*, 144–162. [CrossRef]
8. López de Maturana, E.; Alonso, L.; Alarcón, P.; Martín-Antoniano, I.A.; Pineda, S.; Piorno, L.; Calle, M.L.; Malats, N. Challenges in the Integration of Omics and Non-Omics Data. *Genes.* **2019**, *10*, 238. [CrossRef]
9. Fondi, M.; Liò, P. Multi-omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiol. Res.* **2015**, *171*, 52–64. [CrossRef]
10. Wu, Y.; Liu, Q.; Xie, L. Hierarchical multi-omics data integration and modeling predict cell-specific chemical proteomics and drug responses. *Cell Rep. Methods* **2023**, *3*, 100452. [CrossRef]

11. Zhang, W.; Mou, M.; Hu, W.; Lu, M.; Zhang, H.; Zhang, H.; Luo, Y.; Xu, H.; Tao, L.; Dai, H.; et al. MOINER: A Novel Multiomics Early Integration Framework for Biomedical Classification and Biomarker Discovery. *J. Chem. Inf. Model.* **2024**, *64*, 2720–2732. [CrossRef]
12. Picard, M.; Scott-Boyer, M.P.; Bodein, A.; Périn, O.; Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3735–3746. [CrossRef] [PubMed]
13. Lv, T.; Zhang, Y.; Liu, J.; Kang, Q.; Liu, L. Multi-omics integration for both single-cell and spatially resolved data based on dual-path graph attention auto-encoder. *Brief. Bioinform.* **2024**, *25*, bbae450. [CrossRef]
14. Maitra, C.; Seal, D.B.; Das, V.; De, R.K. Unsupervised neural network for single cell Multi-omics INTegration (UMINT): An application to health and disease. *Front. Mol. Biosci.* **2023**, *10*, 1184748. [CrossRef]
15. Bredikhin, D.; Kats, I.; Stegle, O. MUON: Multimodal omics analysis framework. *Genome Biol.* **2022**, *23*, 42. [CrossRef]
16. Benkirane, H.; Pradat, Y.; Michiels, S.; Cournède, P.H. CustOmics: A versatile deep-learning based strategy for multi-omics integration. *PLoS Comput. Biol.* **2023**, *19*, e1010921. [CrossRef]
17. Xu, J.; Huang, D.S.; Zhang, X. scmFormer Integrates Large-Scale Single-Cell Proteomics and Transcriptomics Data by Multi-Task Transformer. *Adv. Sci.* **2024**, *11*, e2307835. [CrossRef]
18. Kodam, P.; Sai Swaroop, R.; Pradhan, S.S.; Sivaramakrishnan, V.; Vadrevu, R. Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. *Sci. Rep.* **2023**, *13*, 3695. [CrossRef]
19. Stahlschmidt, S.R.; Ulfenborg, B.; Synnergren, J. Multimodal deep learning for biomedical data fusion: A review. *Brief. Bioinform.* **2022**, *23*, bbab569. [CrossRef]
20. Lin, Y.; Wu, T.-Y.; Wan, S.; Yang, J.Y.H.; Wong, W.H.; Wang, Y.X.R. scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. *Nat. Biotechnol.* **2022**, *40*, 703–710. [CrossRef]
21. Yang, X.; Mann, K.K.; Wu, H.; Ding, J. scCross: A deep generative model for unifying single-cell multi-omics with seamless integration, cross-modal generation, and in silico exploration. *Genome Biol.* **2024**, *25*, 198. [CrossRef]
22. Cao, Z.J.; Gao, G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nat. Biotechnol.* **2022**, *40*, 1458–1466. [CrossRef]
23. Li, X.; Ma, J.; Leng, L.; Han, M.; Li, M.; He, F.; Zhu, Y. MoGCN: A Multi-Omics Integration Method Based on Graph Convolutional Network for Cancer Subtype Analysis. *Front. Genet.* **2022**, *13*, 806842. [CrossRef] [PubMed]
24. Wang, Y.; Wang, Z.; Yu, X.; Wang, X.; Song, J.; Yu, D.J.; Ge, F. MORE: A multi-omics data-driven hypergraph integration network for biomedical data classification and biomarker identification. *Brief. Bioinform.* **2024**, *26*, bbae658. [CrossRef]
25. Palshikar, M.G.; Min, X.; Crystal, A.; Meng, J.; Hilchey, S.P.; Zand, M.S.; Thakar, J. Executable Network Models of Integrated Multiomics Data. *J. Proteome Res.* **2023**, *22*, 1546–1556. [CrossRef] [PubMed]
26. Luo, H.; Liang, H.; Liu, H.; Fan, Z.; Wei, Y.; Yao, X.; Cong, S. TEMINET: A Co-Informative and Trustworthy Multi-Omics Integration Network for Diagnostic Prediction. *Int. J. Mol. Sci.* **2024**, *25*, 1655. [CrossRef]
27. Li, Y.; Zhang, D.; Yang, M.; Peng, D.; Yu, J.; Liu, Y.; Lv, J.; Chen, L.; Peng, X. scBridge embraces cell heterogeneity in single-cell RNA-seq and ATAC-seq data integration. *Nat. Commun.* **2023**, *14*, 6045. [CrossRef] [PubMed]
28. Liu, T.; Salguero, P.; Petek, M.; Martinez-Mira, C.; Balzano-Nogueira, L.; Ramšak, Ž.; McIntyre, L.; Gruden, K.; Tarazona, S.; Conesa, A. PaintOmics 4: New tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases. *Nucleic Acids Res.* **2022**, *50*, W551–W559. [CrossRef]
29. Li, J.; Ni, Q.; He, G.; Huang, J.; Chao, H.; Li, S.; Chen, M.; Hu, G.; Whelan, J.; Shou, H. SoyOD: An Integrated Soybean Multi-omics Database for Mining Genes and Biological Research. *Genom. Proteom. Bioinform.* **2025**, *22*, qzae080. [CrossRef]
30. Ning, L.; Zhou, Y.L.; Sun, H.; Zhang, Y.; Shen, C.; Wang, Z.; Xuan, B.; Zhao, Y.; Ma, Y.; Yan, Y.; et al. Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. *Nat. Commun.* **2023**, *14*, 7135. [CrossRef]
31. Maghsoudi, Z.; Nguyen, H.; Tavakkoli, A.; Nguyen, T. A comprehensive survey of the approaches for pathway analysis using multi-omics data integration. *Brief. Bioinform.* **2022**, *23*, bbac435. [CrossRef]
32. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M. The latent geometry of the human protein interaction network. *Bioinformatics* **2018**, *34*, 2826–2834. [CrossRef] [PubMed]
33. Cannistraci, C.V.; Alanis-Lobato, G.; Ravasi, T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* **2013**, *29*, i199–i209. [CrossRef] [PubMed]
34. Vagiona, A.C.; Mier, P.; Petrakis, S.; Andrade-Navarro, M.A. Analysis of Huntington's Disease Modifiers Using the Hyperbolic Mapping of the Protein Interaction Network. *Int. J. Mol. Sci.* **2022**, *23*, 5853. [CrossRef] [PubMed]
35. Zahra, N.U.A.; Vagiona, A.C.; Uddin, R.; Andrade-Navarro, M.A. Selection of Multi-Drug Targets against Drug-Resistant Mycobacterium tuberculosis XDR1219 Using the Hyperbolic Mapping of the Protein Interaction Network. *Int. J. Mol. Sci.* **2023**, *24*, 14050. [CrossRef]

36. Deprince, A.; Haas, J.T.; Staels, B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. *Mol. Metab.* **2020**, *42*, 101092. [CrossRef]
37. de Lima, E.P.; Moretti, R.C., Jr.; Torres Pomini, K.; Laurindo, L.F.; Sloan, K.P.; Sloan, L.A.; Castro, M.V.M.d.; Baldi, E., Jr.; Ferraz, B.F.R.; de Souza Bastos Mazuqueli Pereira, E.; et al. Glycolipid Metabolic Disorders, Metainflammation, Oxidative Stress, and Cardiovascular Diseases: Unraveling Pathways. *Biology* **2024**, *13*, 519. [CrossRef]
38. Taube, A.; Schlich, R.; Sell, H.; Eckardt, K.; Eckel, J. Inflammation and metabolic dysfunction: Links to cardiovascular diseases. *Am. J. Physiol. Heart Circ. Physiol.* **2012**, *302*, H2148–H2165. [CrossRef]
39. Doran, S.; Arif, M.; Lam, S.; Bayraktar, A.; Turkez, H.; Uhlen, M.; Boren, J.; Mardinoglu, A. Multi-omics approaches for revealing the complexity of cardiovascular disease. *Brief. Bioinform.* **2021**, *22*, bbab061. [CrossRef]
40. Zhong, S.; Li, L.; Shen, X.; Li, Q.; Xu, W.; Wang, X.; Tao, Y.; Yin, H. An update on lipid oxidation and inflammation in cardiovascular diseases. *Free Radic. Biol. Med.* **2019**, *144*, 266–278. [CrossRef]
41. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* **2017**, *45*, D408–D414. [CrossRef]
42. Schaefer, M.H.; Fontaine, J.F.; Vinayagam, A.; Porras, P.; Wanker, E.E.; Andrade-Navarro, M.A. HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS ONE* **2012**, *7*, e31826. [CrossRef] [PubMed]
43. Anyaegbunam, U.A.; More, P.; Fontaine, J.F.; Cate, V.T.; Bauer, K.; Distler, U.; Araldi, E.; Bindila, L.; Wild, P.; Andrade-Navarro, M.A. A Systematic Review of Lipid-Focused Cardiovascular Disease Research: Trends and Opportunities. *Curr. Issues Mol. Biol.* **2023**, *45*, 9904–9916. [CrossRef]
44. Aimo, L.; Liechti, R.; Hyka-Nouspikel, N.; Niknejad, A.; Gleizes, A.; Götz, L.; Kuznetsov, D.; David, F.P.; van der Goot, F.G.; Riezman, H.; et al. The SwissLipids knowledgebase for lipid biology. *Bioinformatics* **2015**, *31*, 2860–2866. [CrossRef]
45. Cadby, G.; Giles, C.; Melton, P.E.; Huynh, K.; Mellett, N.A.; Duong, T.; Nguyen, A.; Cinel, M.; Smith, A.; Olshansky, G.; et al. Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. *Nat. Commun.* **2022**, *13*, 3124. [CrossRef] [PubMed]
46. Wu, Z.; Bezwada, D.; Cai, F.; Harris, R.C.; Ko, B.; Sondhi, V.; Pan, C.; Vu, H.S.; Nguyen, P.T.; Faubert, B.; et al. Electron transport chain inhibition increases cellular dependence on purine transport and salvage. *Cell Metab.* **2024**, *36*, 1504–1520.e9. [CrossRef] [PubMed]
47. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380. [CrossRef]
48. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M.A. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. *Appl. Netw. Sci.* **2016**, *1*, 10. [CrossRef]
49. Alanis-Lobato, G.; Mier, P.; Andrade-Navarro, M.A. Efficient embedding of complex networks to hyperbolic space via their Laplacian. *Sci. Rep.* **2016**, *6*, 30108. [CrossRef]
50. Papadopoulos, F.; Aldecoa, R.; Krioukov, D. Network geometry inference using common neighbors. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2015**, *92*, 022807. [CrossRef]
51. Papadopoulos, F.; Kitsak, M.; Serrano, M.Á.; Boguñá, M.; Krioukov, D. Popularity versus similarity in growing networks. *Nature* **2012**, *489*, 537–540. [CrossRef]
52. Krioukov, D.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; Boguñá, M. Hyperbolic geometry of complex networks. Physical review. *E Stat. Nonlinear Soft Matter Phys.* **2010**, *82 Pt 2*, 036106. [CrossRef] [PubMed]
53. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [CrossRef]
54. Wishart, D.S.; Guo, A.; Oler, E.; Wang, F.; Anjum, A.; Peters, H.; Dizon, R.; Sayeeda, Z.; Tian, S.; Lee, B.L.; et al. HMDB 5.0: The Human Metabolome Database for 2022. *Nucleic Acids Res.* **2022**, *50*, D622–D631. [CrossRef]
55. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef] [PubMed]
56. Yoo, J.C.; Lim, T.Y.; Park, J.S.; Hah, Y.S.; Park, N.; Hong, S.G.; Park, J.Y.; Yoon, T.J. SYT14L, especially its C2 domain, is involved in regulating melanocyte differentiation. *J. Dermatol. Sci.* **2013**, *72*, 246–251. [CrossRef]
57. Tsuboi, T.; McMahon, H.T.; Rutter, G.A. Mechanisms of dense core vesicle recapture following "kiss and run" ("cavicapture") exocytosis in insulin-secreting cells. *J. Biol. Chem.* **2004**, *279*, 47115–47124. [CrossRef]
58. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef]
59. Zinman, B.; Lachin, J.M.; Inzucchi, S.E. Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *N. Engl. J. Med.* **2016**, *374*, 1094. [CrossRef] [PubMed]

60. Prochaska, J.H.; Jünger, C.; Schulz, A.; Arnold, N.; Müller, F.; Heidorn, M.W.; Baumkötter, R.; Zahn, D.; Koeck, T.; Tröbs, S.O.; et al. Effects of empagliflozin on left ventricular diastolic function in addition to usual care in individuals with type 2 diabetes mellitus-results from the randomized, double-blind, placebo-controlled EmDia trial. *Clin. Res. Cardiol. Off. J. Ger. Card. Soc.* **2023**, *112*, 911–922. [CrossRef]
61. Jünger, C.; Prochaska, J.H.; Gori, T.; Schulz, A.; Binder, H.; Daiber, A.; Koeck, T.; Rapp, S.; Lackner, K.J.; Münzel, T.; et al. Rationale and design of the effects of EMpagliflozin on left ventricular DIAstolic function in diabetes (EmDia) study. *J. Cardiovasc. Med.* **2022**, *23*, 191–197. [CrossRef]
62. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
63. Anyaegbunam, U.A. Omint. Available online: https://github.com/uchealex/Omint (accessed on 24 March 2025).
64. Jeon, J.; Han, E.Y.; Jung, I. MOPA: An integrative multi-omics pathway analysis method for measuring omics activity. *PLoS ONE* **2023**, *18*, e0278272. [CrossRef] [PubMed]
65. Paczkowska, M.; Barenboim, J.; Sintupisut, N.; Fox, N.S.; Zhu, H.; Abd-Rabbo, D.; Mee, M.W.; Boutros, P.C.; PCAWG Drivers and Functional Interpretation Working Group; Reimand, J.; et al. Integrative pathway enrichment analysis of multivariate omics data. *Nat. Commun.* **2020**, *11*, 735. [CrossRef] [PubMed]
66. Slobodyanyuk, M.; Bahcheli, A.T.; Klein, Z.P.; Bayati, M.; Strug, L.J.; Reimand, J. Directional integration and pathway enrichment analysis for multi-omics data. *Nat. Commun.* **2024**, *15*, 5690. [CrossRef] [PubMed]

5. General Discussion

Large-scale efforts have expanded the list of human protein-protein interactions, yet current maps still capture only a part of the interactome and are hard to interpret biologically [19,127]. This thesis used hyperbolic embeddings to map proteins into a coordinate system that captures both hierarchical depth and functional proximity. The embedding was applied to address several biological problems. First, they enabled the separation of disease modifiers. Second, they supported the prediction of directed (de-)phosphorylation interactions and the classification between cooperative and competitive binding motifs, while also integrating other omics layers.

5.1 Insights into disease biology

In Chapter 1, a hyperbolically embedded human interactome was used to investigate whether paralog pairs that interact with huntingtin (HTT) but exhibit opposite effects on polyQ toxicity also occupy different neighborhoods in latent space. The analysis showed that paralog pairs with opposite phenotypes tended to occupy different angular positions and radial depths, suggesting functional divergence despite sequence similarity. This pattern aligns with evidence that closely related proteins can modify aggregation and toxicity in opposite ways (e.g., coiled-coil domain paralogs or chaperone family members). More generally, these findings support the idea that modifier effects are shaped not only by sequence but also by structural motifs and their position in the broader protein interaction network [105]. A few specific examples highlight this pattern. The MID1/PML paralogs, both HTT interactors, fall into distinct angular sectors consistent with opposite mechanisms. More specifically, MID1 brings PP2A and S6K to the mutant HTT mRNA, which upregulates translation and aggravating toxicity. In contrast, PML scaffolds SUMOylation of misfolded nuclear proteins, which are then recognized by RNF4, a SUMO-targeted ubiquitin ligase that polyubiquitinates them for proteasomal degradation, reducing proteotoxic stress [128,129]. Similarly, under DNA damage the IKK paralogs diverge: IKKβ activation promotes HTT cleavage and neurotoxicity, whereas IKKα (CHUK) counteracts this response and reduces toxicity [130]. Analysis of shared interactors identified PPP2CA, HSP90AA1, RPS3, and TUBB as candidate modifiers: PPP2CA relates to HTT S421 dephosphorylation, HSP90AA1 to chaperone-mediated proteostasis, RPS3 to DNA-damage/stress signaling, and TUBB to microtubule transport. All these pathways implicated in HD and other polyQ disorders [131–134].

While Chapter 1 we focused on paralog divergence in HD, Chapter 2 extended the framework to a different polyQ disease context, spinocerebellar ataxia type 1 (SCA1), where directed PTM-PPIs are especially relevant. In this chapter, directed phosphorylation and dephosphorylation interactions were predicted using a model that combines hyperbolic coordinates with network features, and the approach was evaluated in a SCA1 context. Whereas prior computational work focused on undirected PPI prediction [135–137], our results show that direction and regulatory role can be inferred with useful accuracy when node positions in the hyperbolic map are included

[63]. Notably, the angular (θ) coordinates of targets and effectors appeared as the most informative features, outperforming purely topological measures. The stronger contribution of target θ relative to effector θ is consistent with the observation that targets often cluster within functionally well-defined network neighborhoods, such as signaling, cell-cycle, or differentiation modules [27], whereas upstream effectors, including kinases and transcription factors, act across broader sectors of the interactome [138]. Among network centralities, eigenvector (EC), closeness (CC), and betweenness (BC) added informative value, whereas degree was comparatively less informative for these directed interactions. This highlights that the functionality of a protein depends more on the nature and positional context of its interactions than on degree centrality alone [47].

We then applied the predictions to disease biology using a SCA1 cell model that reflects polyQ characteristics [139]. Our analysis of proteomics data demonstrated that polyQ-expanded ataxin-1 drives widespread alterations in the human protein interaction network (hPIN), with dysregulated proteins clustering in functional modules related to neurodegeneration, spliceosome, lysosome, and ribosome pathways [140]. Integration of our predictive post-translational modifications of protein interactions (PTM-PPI) identified 13 potential PTM-mediated interactors of ataxin-1, including known modifiers such as PHPT1, SUMO1, USP7, SNCA, HSPB1, and DNAJB6 [141–143]. Network analysis revealed that three dysregulated kinases (MAPK1, MAPK3, and CDK4) target α-synuclein (SNCA), a protein implicated in Parkinson's disease that links phosphorylation-dependent regulation to nuclear localization control [144]. This finding suggests that kinase-substrate pathways may share common mechanisms of protein aggregation in neurodegenerative diseases. Finally, network-based drug repurposing identified artesunate, budesonide, and betamethasone as candidate drugs that could reduce SCA1-related protein changes by influencing glucocorticoid receptor signaling and protein production, pointing to these processes as possible treatment targets in polyQ diseases [145,146].

Thus, hyperbolic embedding of the human interactome not only enables prediction of directed PTM-PPIs but also provides biological insights into disease mechanisms and potential therapeutic targets in polyQ neurodegeneration.

5.2 Higher-order motifs and structural context of PPIs

Chapter 3 extended the analysis of protein-protein interactions beyond pairs to triplets, addressing whether two partners of a common protein can bind simultaneously (cooperative) or compete for overlapping interfaces (competitive). Machine learning models trained in geometric, topological, and biological features showed that cooperative and competitive triplets could be discriminated with high accuracy. The Random Forest classifier achieved the best performance (AUC ≈ 0.88), confirming that higher-order interaction configurations can be predicted systematically. Feature importance analysis revealed that the angular separation between the two partner proteins (V1 and V2) was the most predictive variable. Smaller angular distances enriched for cooperative

assemblies, whereas larger separations were associated with competitive interactions. Biologically, this suggests that cooperative triplets tend to involve functionally related that can bind distinct interfaces, whereas competitive triplets involve functionally divergent proteins competing for overlapping surfaces on the common interactor [147].

Biological attributes such as subcellular localization and intrinsic disorder content contributed less to predictive accuracy, although they remain relevant to functional interpretation. Interestingly, paralogous pairs were enriched in the highest-scoring cooperative triplets, suggesting that paralogs are frequently retained within the same complex to support cooperative assembly and functional redundancy. This observation aligns with evolutionary studies showing that gene duplication can promote cooperative diversification rather than direct competition within complexes [147,148]. Structural validation with AlphaFold confirmed the predictive model of cooperative triplets corresponded to distinct, non-overlapping interfaces, whereas competitive triplets showed overlapping or adjacent binding surfaces. These findings offer structural validation of the predictive features identified by the model.

Taken together, these results demonstrate that hyperbolic embedding of the human protein interaction network captures latent geometric principles that distinguish cooperative from competitive assemblies, extending classical concepts such as “date” and “party” hubs [149] into a quantitative approach. A limitation of this study is the assumption that open triplets without structural support are competitive, which may underestimate the actual frequency of cooperative interactions in cases where structural data are missing. Despite this, our approach provides a scalable method for studying higher-order protein assemblies and shows that geometric and topological features of the network, rather than biological annotations alone, are key features of multiprotein complex organization.

5.3 Multi-omics integration and systems biology

In Chapter 4, the protein interaction network is expanded to include lipids and metabolites that interact with the proteins in the network. These additional layers are embedded into hyperbolic space to integrate proteomic, lipidomic, and metabolomic data and to investigate how molecular dysregulations across these three levels align within functional modules.

In the multi-omics network, proteins, lipids, and metabolites were positioned in the angular similarity dimension clustered into biological modules, such as respiration and phospholipid associated sectors. These clusters reflect known biochemical relationships while also suggesting novel associations. For example, lipids and metabolites with established cardiovascular roles were positioned near disease-related proteins, whereas poorly characterized molecules appeared in similar neighborhoods, pointing to potential biomarker roles. Next, we examined a drug-response scenario to evaluate how lipidomic alterations can be positioned within the network. Empagliflozin, a sodium-glucose cotransporter 2 (SGLT2) inhibitor with proven cardioprotective effects, was used

as a case study. Temporal lipidomics profiling revealed that early treatment effects were enriched in phospholipid metabolism, whereas later time points showed shifts toward sphingolipid and ceramide metabolism [150,151]. These lipidomic signatures correspond to clinical evidence indicating changes into the hyperbolic network, we could link them to candidate protein effectors, thereby suggesting mechanisms on how drug effects the molecules.

6. General conclusion and outlook

Across Chapters 1-4, a central result of this thesis is that the hyperbolic embedding of the human protein interaction network can be linked to biological meaning in different contexts. The same geometric map supported the identification of disease modifiers, the prediction of directed (de-) phosphorylation interactions, the classification of cooperative versus competitive binding motifs, and the integration of proteomics, lipidomics, and metabolomics.

At the same time, the methodology reflects the biases and gaps of the datasets it is derived from. First, the interactome maps are incomplete and skewed toward well-studied proteins and assays, and embeddings necessarily reflect these biases [19]. Second, defining negatives is challenging because many classified as non-interactions are simply untested or unobserved, which introduces noise into supervised learning (Chapter 3) and complicates the labeling of triplets (Chapter 4). Third, the current network is static, whereas many cooperative, competitive, or regulatory relationships are state-dependent and vary across conditions or cell types. Fourth, structural data is incomplete: although AlphaFold provides valuable predictions, it lacks conformational flexibility, especially in intrinsically disordered regions (IDRs). Finally, current embedding approaches can be computationally demanding, especially for large or multi-layer networks, and results may vary depending on inference method and parameter choices. Improving scalability and efficiency will be important for applying these methods to expanding omics datasets.

This work suggests several promising lines of investigation. One promising next step is to generate context-specific embeddings such as tissue and cell-type specific maps. Future studies should extend the multi-omics integration developed in Chapter 4 by incorporating additional data layers such as transcriptomics, epigenomics, or spatial omics. Embedding these into the same geometric space would provide a more complete map of molecular interactions. Another important direction is the construction of disease-specific networks. Comparing embeddings derived from healthy and diseased states, such as neurodegeneration or cancer, could reveal how network geometry reorganizes under pathological conditions and identifies modules or interactions that are selectively rewired in diseases.

7. References

[1] Tyers M, Mann M. From genomics to proteomics. Nature 2003;422:193–7. https://doi.org/10.1038/nature01510.

[2] Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature 2016;537:347–55. https://doi.org/10.1038/nature19949.

[3] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell 2018;172:650–65. https://doi.org/10.1016/j.cell.2018.01.029.

[4] Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein interaction network: a resource for annotating the proteome. Cell 2005;122:957–68. https://doi.org/10.1016/j.cell.2005.08.029.

[5] Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A Proteome-Scale Map of the Human Interactome Network. Cell 2014;159:1212–26. https://doi.org/10.1016/j.cell.2014.10.050.

[6] Nooren IMA, Thornton JM. Diversity of protein–protein interactions. The EMBO Journal 2003;22:3486–92. https://doi.org/10.1093/emboj/cdg359.

[7] Fields S, Song O. A novel genetic system to detect protein–protein interactions. Nature 1989;340:245–6. https://doi.org/10.1038/340245a0.

[8] Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002;415:141–7. https://doi.org/10.1038/415141a.

[9] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17:1030–2. https://doi.org/10.1038/13732.

[10] Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. Journal of Cell Biology 2012;196:801–10. https://doi.org/10.1083/jcb.201112098.

[11] Hung V, Lam SS, Udeshi ND, Svinkina T, Guzman G, Mootha VK, et al. Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. ELife 2017;6:e24463. https://doi.org/10.7554/eLife.24463.

[12] Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Research 2019;47:D529–41. https://doi.org/10.1093/nar/gky1079.

[13] Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Research 2014;42:D358–63. https://doi.org/10.1093/nar/gkt1115.

[14] Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. Nucleic Acids Research 2004;32:D449–51. https://doi.org/10.1093/nar/gkh086.

[15] Schaefer MH, Fontaine J-F, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. PLoS ONE 2012;7:e31826. https://doi.org/10.1371/journal.pone.0031826.

[16] Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. Nucleic Acids Res 2017;45:D408–14. https://doi.org/10.1093/nar/gkw985.

[17] Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. Nat Methods 2013;10:690–1. https://doi.org/10.1038/nmeth.2561.

[18] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide

experimental datasets. Nucleic Acids Research 2019;47:D607–13. https://doi.org/10.1093/nar/gky1131.
[19] Luck K, Kim D-K, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. Nature 2020;580:402–8. https://doi.org/10.1038/s41586-020-2188-x.
[20] Guo MG, Sosa DN, Altman RB. Challenges and opportunities in network-based solutions for biological questions. Brief Bioinform 2022;23:bbab437. https://doi.org/10.1093/bib/bbab437.
[21] Shin W-H, Kumazawa K, Imai K, Hirokawa T, Kihara D. Current Challenges and Opportunities in Designing Protein–Protein Interaction Targeted Drugs. Adv Appl Bioinform Chem 2020;13:11–25. https://doi.org/10.2147/AABC.S235542.
[22] Wodak SJ, Pu S, Vlasblom J, Séraphin B. Challenges and rewards of interaction proteomics. Mol Cell Proteomics 2009;8:3–18. https://doi.org/10.1074/mcp.R800014-MCP200.
[23] Vidal M, Cusick ME, Barabási A-L. Interactome Networks and Human Disease. Cell 2011;144:986–98. https://doi.org/10.1016/j.cell.2011.02.016.
[24] Menche J, Sharma A, Kitsak M, Ghiassian S, Vidal M, Loscalzo J, et al. Uncovering disease-disease relationships through the incomplete human interactome. Science 2015;347:1257601. https://doi.org/10.1126/science.1257601.
[25] Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. Nature 2017;545:505–9. https://doi.org/10.1038/nature22366.
[26] Agamah FE, Bayjanov JR, Niehues A, Njoku KF, Skelton M, Mazandu GK, et al. Computational approaches for network-based integrative multi-omics analysis. Front Mol Biosci 2022;9. https://doi.org/10.3389/fmolb.2022.967205.
[27] Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101–13. https://doi.org/10.1038/nrg1272.
[28] Newman MEJ. The Structure and Function of Complex Networks. SIAM Rev 2003;45:167–256. https://doi.org/10.1137/S003614450342480.
[29] De Las Rivas J, Fontanillo C. Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell. Briefings in Functional Genomics 2012;11:489–96. https://doi.org/10.1093/bfgp/els036.
[30] Ideker T, Galitski T, Hood L. A NEW APPROACH TO DECODING LIFE: Systems Biology. Annual Review of Genomics and Human Genetics 2001;2:343–72. https://doi.org/10.1146/annurev.genom.2.1.343.
[31] Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. BioData Mining 2011;4:10. https://doi.org/10.1186/1756-0381-4-10.
[32] Lesne A. Complex Networks: from Graph Theory to Biology. Lett Math Phys 2006;78:235–62. https://doi.org/10.1007/s11005-006-0123-1.
[33] Koschützki D, Schreiber F. Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. Gene Regul Syst Bio 2008;2:GRSB.S702. https://doi.org/10.4137/GRSB.S702.
[34] Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2. https://doi.org/10.1038/35075138.
[35] Freeman LC. Centrality in social networks conceptual clarification. Social Networks 1978;1:215–39. https://doi.org/10.1016/0378-8733(78)90021-7.
[36] Zotenko E, Mestre J, O'leary D, Przytycka T. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. PLoS Computational Biology 2008;4:e1000140. https://doi.org/10.1371/journal.pcbi.1000140.
[37] Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. Proceedings of the National Academy of Sciences 2007;104:8685–90. https://doi.org/10.1073/pnas.0701361104.

[38] Estrada E. Virtual identification of essential proteins within the protein interaction network of yeast. Proteomics 2006;6:35–40. https://doi.org/10.1002/pmic.200500209.
[39] Freeman LC. A Set of Measures of Centrality Based on Betweenness. Sociometry 1977;40:35–41. https://doi.org/10.2307/3033543.
[40] Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. PLOS Computational Biology 2007;3:e59. https://doi.org/10.1371/journal.pcbi.0030059.
[41] Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol 2005;2005:96–103. https://doi.org/10.1155/JBB.2005.96.
[42] Hahn MW, Kern AD. Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. Molecular Biology and Evolution 2005;22:803–6. https://doi.org/10.1093/molbev/msi072.
[43] Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. Nature 2000;406:378–82. https://doi.org/10.1038/35019019.
[44] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56–68. https://doi.org/10.1038/nrg2918.
[45] Brandes U. A faster algorithm for betweenness centrality*. The Journal of Mathematical Sociology 2001;25:163–77. https://doi.org/10.1080/0022250X.2001.9990249.
[46] Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, et al. Estimating the size of the human interactome. Proceedings of the National Academy of Sciences 2008;105:6959–64. https://doi.org/10.1073/pnas.0708078105.
[47] Koschützki D, Schreiber F. Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. Gene Regul Syst Bio 2008;2:193–201. https://doi.org/10.4137/grsb.s702.
[48] Sabidussi G. The centrality index of a graph. Psychometrika 1966;31:581–603. https://doi.org/10.1007/BF02289527.
[49] Bavelas A. Communication Patterns in Task-Oriented Groups. The Journal of the Acoustical Society of America 1950;22:725–30. https://doi.org/10.1121/1.1906679.
[50] del Sol A, Fujihashi H, O'Meara P. Topology of small-world networks of protein-protein complex structures. Bioinformatics 2005;21:1311–5. https://doi.org/10.1093/bioinformatics/bti167.
[51] Guimerà R, Nunes Amaral LA. Functional cartography of complex metabolic networks. Nature 2005;433:895–900. https://doi.org/10.1038/nature03288.
[52] Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009;27:199–204. https://doi.org/10.1038/nbt.1522.
[53] Bonacich P. Factoring and weighting approaches to status scores and clique identification. The Journal of Mathematical Sociology 1972;2:113–20. https://doi.org/10.1080/0022250X.1972.9989806.
[54] Estrada E, Rodríguez-Velázquez JA. Subgraph centrality in complex networks. Phys Rev E 2005;71:056103. https://doi.org/10.1103/PhysRevE.71.056103.
[55] del Rio G, Koschützki D, Coello G. How to identify essential genes from molecular networks? BMC Systems Biology 2009;3:102. https://doi.org/10.1186/1752-0509-3-102.
[56] Jalili M, Salehzadeh-Yazdi A, Asgari Y, Arab SS, Yaghmaie M, Ghavamzadeh A, et al. CentiServer: A Comprehensive Resource, Web-Based Application and R Package for Centrality Analysis. PLOS ONE 2015;10:e0143111. https://doi.org/10.1371/journal.pone.0143111.
[57] Barabasi AL, Albert R. Emergence of scaling in random networks. Science 1999;286:509–12. https://doi.org/10.1126/science.286.5439.509.
[58] Albert R. Scale-free networks in cell biology. J Cell Sci 2005;118:4947–57. https://doi.org/10.1242/jcs.02714.

[59] Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science 2002;296:750–2. https://doi.org/10.1126/science.1068696.
[60] Albert R. Statistical mechanics of complex networks. Rev Mod Phys 2002;74:47–97. https://doi.org/10.1103/RevModPhys.74.47.
[61] Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguñá M. Hyperbolic geometry of complex networks. Phys Rev E 2010;82:036106. https://doi.org/10.1103/PhysRevE.82.036106.
[62] Papadopoulos F, Kitsak M, Serrano MÁ, Boguñá M, Krioukov D. Popularity versus similarity in growing networks. Nature 2012;489:537–40. https://doi.org/10.1038/nature11459.
[63] Alanis-Lobato G, Mier P, Andrade-Navarro M. The latent geometry of the human protein interaction network. Bioinformatics 2018;34:2826–34. https://doi.org/10.1093/bioinformatics/bty206.
[64] Papadopoulos F, Psomas C, Krioukov D. Network Mapping by Replaying Hyperbolic Growth. IEEE/ACM Transactions on Networking 2015;23:198–211. https://doi.org/10.1109/TNET.2013.2294052.
[65] Alanis-Lobato G, Mier P, Andrade-Navarro MA. Manifold learning and maximum likelihood estimation for hyperbolic network embedding. Appl Netw Sci 2016;1:10. https://doi.org/10.1007/s41109-016-0013-0.
[66] Muscoloni A, Cannistraci CV. A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities. New J Phys 2018;20:052002. https://doi.org/10.1088/1367-2630/aac06f.
[67] Yang M, Zhou M, Li Z, Liu J, Pan L, Xiong H, et al. Hyperbolic Graph Neural Networks: A Review of Methods and Applications. 2022. https://doi.org/10.48550/arXiv.2202.13852.
[68] Lizotte S, Young J-G, Allard A. Symmetry-driven embedding of networks in hyperbolic space. Commun Phys 2025;8:199. https://doi.org/10.1038/s42005-025-02122-0.
[69] Tang T, Shen T, Jiang J, Li W, Wang P, Yuan S, et al. Prediction of protein–protein interaction based on interaction-specific learning and hierarchical information. BMC Biology 2025;23:236. https://doi.org/10.1186/s12915-025-02359-9.
[70] Vagiona A-C, Mier P, Andrade-Navarro MA. Unraveling cooperative and competitive interactions within protein triplets in the human interactome. Scientific Reports 2025;15:32548. https://doi.org/10.1038/s41598-025-19264-4.
[71] Pogány D, Antal P. Towards explainable interaction prediction: Embedding biological hierarchies into hyperbolic interaction space. PLOS ONE 2024;19:e0300906. https://doi.org/10.1371/journal.pone.0300906.
[72] Marsh JA, Teichmann SA. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. Annual Review of Biochemistry 2015;84:551–75. https://doi.org/10.1146/annurev-biochem-060614-034142.
[73] Seet BT, Dikic I, Zhou M-M, Pawson T. Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 2006;7:473–83. https://doi.org/10.1038/nrm1960.
[74] Xu H, Wang Y, Lin S, Deng W, Peng D, Cui Q, et al. PTMD: A Database of Human Disease-associated Post-translational Modifications. Genomics, Proteomics & Bioinformatics 2018;16:244–51. https://doi.org/10.1016/j.gpb.2018.06.004.
[75] Deribe YL, Pawson T, Dikic I. Post-translational modifications in signal integration. Nat Struct Mol Biol 2010;17:666–72. https://doi.org/10.1038/nsmb.1842.
[76] Pawson T, Scott JD. Protein phosphorylation in signaling – 50 years and counting. Trends in Biochemical Sciences 2005;30:286–90. https://doi.org/10.1016/j.tibs.2005.04.013.
[77] Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 2012;40:D261–70. https://doi.org/10.1093/nar/gkr1122.

[78] Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, et al. Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Res 2011;39:D261-267. https://doi.org/10.1093/nar/gkq1104.
[79] Li J, Jia J, Li H, Yu J, Sun H, He Y, et al. SysPTM 2.0: an updated systematic resource for post-translational modification. Database (Oxford) 2014;2014:bau025. https://doi.org/10.1093/database/bau025.
[80] Li Z, Li S, Luo M, Jhong J-H, Li W, Yao L, et al. dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. Nucleic Acids Res 2022;50:D471–9. https://doi.org/10.1093/nar/gkab1017.
[81] Wang D, Zeng S, Xu C, Qiu W, Liang Y, Joshi T, et al. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. Bioinformatics 2017;33:3909–16. https://doi.org/10.1093/bioinformatics/btx496.
[82] Chen M, Zhang W, Gou Y, Xu D, Wei Y, Liu D, et al. GPS 6.0: an updated server for prediction of kinase-specific phosphorylation sites in proteins. Nucleic Acids Res 2023;51:W243–50. https://doi.org/10.1093/nar/gkad383.
[83] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 2015;16:18–29. https://doi.org/10.1038/nrm3920.
[84] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. Journal of Molecular Biology 1999;293:321–31. https://doi.org/10.1006/jmbi.1999.3110.
[85] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42. https://doi.org/10.1093/nar/28.1.235.
[86] Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods 2013;10:47–53. https://doi.org/10.1038/nmeth.2289.
[87] Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer 2022:2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034.
[88] Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, et al. Towards a structurally resolved human protein interaction network. Nat Struct Mol Biol 2023;30:216–25. https://doi.org/10.1038/s41594-022-00910-8.
[89] Halakou F, Kilic ES, Cukuroglu E, Keskin O, Gursoy A. Enriching Traditional Protein-protein Interaction Networks with Alternative Conformations of Proteins. Sci Rep 2017;7:7180. https://doi.org/10.1038/s41598-017-07351-0.
[90] Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Uncovering disease-disease relationships through the incomplete interactome. Science 2015;347:1257601. https://doi.org/10.1126/science.1257601.
[91] Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 2015;47:569–76. https://doi.org/10.1038/ng.3259.
[92] Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell 2011;144:646–74. https://doi.org/10.1016/j.cell.2011.02.013.
[93] Loscalzo J, Barabasi A-L. Systems biology and the future of medicine. WIREs Systems Biology and Medicine 2011;3:619–27. https://doi.org/10.1002/wsbm.144.
[94] Zhou X, Menche J, Barabási A-L, Sharma A. Human symptoms–disease network. Nat Commun 2014;5:4212. https://doi.org/10.1038/ncomms5212.
[95] Ghazal R, Wang M, Liu D, Tschumperlin DJ, Pereira NL. Cardiac Fibrosis in the Multi-Omics Era: Implications for Heart Failure. Circ Res 2025;136:773–802. https://doi.org/10.1161/CIRCRESAHA.124.325402.
[96] Gitler AD, Dhillon P, Shorter J. Neurodegenerative disease: models, mechanisms, and a new hope. Dis Model Mech 2017;10:499–502. https://doi.org/10.1242/dmm.030205.

[97] Soto C, Pritzkow S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. Nat Neurosci 2018;21:1332–40. https://doi.org/10.1038/s41593-018-0235-9.
[98] Goedert M, Eisenberg DS, Crowther RA. Propagation of Tau Aggregates and Neurodegeneration. Annual Review of Neuroscience 2017;40:189–210. https://doi.org/10.1146/annurev-neuro-072116-031153.
[99] Zuccato C, Valenza M, Cattaneo E. Molecular Mechanisms and Potential Therapeutical Targets in Huntington's Disease. Physiological Reviews 2010;90:905–81. https://doi.org/10.1152/physrev.00041.2009.
[100] Shirasaki DI, Greiner ER, Al-Ramahi I, Gray M, Boontheung P, Geschwind DH, et al. Network Organization of the Huntingtin Proteomic Interactome in Mammalian Brain. Neuron 2012;75:41–57. https://doi.org/10.1016/j.neuron.2012.05.024.
[101] Saudou F, Humbert S. The Biology of Huntingtin. Neuron 2016;89:910–26. https://doi.org/10.1016/j.neuron.2016.02.003.
[102] Vagiona A-C, Mier P, Petrakis S, Andrade-Navarro MA. Analysis of Huntington's Disease Modifiers Using the Hyperbolic Mapping of the Protein Interaction Network. IJMS 2022;23:5853. https://doi.org/10.3390/ijms23105853.
[103] Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, Hill DE, et al. Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. Nature 2008;452:713–8. https://doi.org/10.1038/nature06731.
[104] Lasagna-Reeves CA, Rousseaux MW, Guerrero-Munoz MJ, Vilanova-Velez L, Park J, See L, et al. Ataxin-1 oligomers induce local spread of pathology and decreasing them by passive immunization slows Spinocerebellar ataxia type 1 phenotypes. ELife 2015;4:e10891. https://doi.org/10.7554/eLife.10891.
[105] Petrakis S, Raskó T, Russ J, Friedrich RP, Stroedicke M, Riechers S-P, et al. Identification of Human Proteins That Modify Misfolding and Proteotoxicity of Pathogenic Ataxin-1. PLOS Genetics 2012;8:e1002897. https://doi.org/10.1371/journal.pgen.1002897.
[106] Nitschke L, Coffin SL, Xhako E, El-Najjar DB, Orengo JP, Alcala E, et al. Modulation of ATXN1 S776 phosphorylation reveals the importance of allele-specific targeting in SCA1. JCI Insight 2021;6. https://doi.org/10.1172/jci.insight.144955.
[107] Härtner F, Andrade-Navarro MA, Alanis-Lobato G. Geometric characterisation of disease modules. Appl Netw Sci 2018;3:10. https://doi.org/10.1007/s41109-018-0066-3.
[108] Zahra NUA, Vagiona A-C, Uddin R, Andrade-Navarro MA. Selection of Multi-Drug Targets against Drug-Resistant Mycobacterium tuberculosis XDR1219 Using the Hyperbolic Mapping of the Protein Interaction Network. IJMS 2023;24:14050. https://doi.org/10.3390/ijms241814050.
[109] Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 2008;9:387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359.
[110] Long E, Wan P, Chen Q, Lu Z, Choi J. From function to translation: Decoding genetic susceptibility to human diseases via artificial intelligence. Cell Genomics 2023;3. https://doi.org/10.1016/j.xgen.2023.100320.
[111] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63. https://doi.org/10.1038/nrg2484.
[112] Wenk MR. The emerging field of lipidomics. Nat Rev Drug Discov 2005;4:594–610. https://doi.org/10.1038/nrd1776.
[113] Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 2016;17:451–9. https://doi.org/10.1038/nrm.2016.25.
[114] Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol 2017;18:83. https://doi.org/10.1186/s13059-017-1215-1.

[115] Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights 2020;14:1177932219899051. https://doi.org/10.1177/1177932219899051.
[116] Coman C, Solari FA, Hentschel A, Sickmann A, Zahedi RP, Ahrends R. Simultaneous Metabolite, Protein, Lipid Extraction (SIMPLEX): A Combinatorial Multimolecular Omics Approach for Systems Biology. Mol Cell Proteomics 2016;15:1453–66. https://doi.org/10.1074/mcp.M115.053702.
[117] Alexandrov T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. Annu Rev Biomed Data Sci 2020;3:61–87. https://doi.org/10.1146/annurev-biodatasci-011420-031537.
[118] Giesen C, Wang HAO, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 2014;11:417–22. https://doi.org/10.1038/nmeth.2869.
[119] Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: tools, advances and future approaches. J Mol Endocrinol 2019;62:R21–45. https://doi.org/10.1530/JME-18-0055.
[120] Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Computational Biology 2017;13:e1005752. https://doi.org/10.1371/journal.pcbi.1005752.
[121] Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol 2012;8:565. https://doi.org/10.1038/msb.2011.99.
[122] Bersanelli M, Mosca E, Remondini D, Castellani G, Milanesi L. Network diffusion-based analysis of high-throughput data for the detection of differentially enriched modules. Sci Rep 2016;6:34841. https://doi.org/10.1038/srep34841.
[123] Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 2018;14:e8124. https://doi.org/10.15252/msb.20178124.
[124] Cristescu R, Lee J, Nebozhyn M, Kim K-M, Ting JC, Wong SS, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449–56. https://doi.org/10.1038/nm.3850.
[125] Nurmohamed NS, Kraaijenhof JM, Mayr M, Nicholls SJ, Koenig W, Catapano AL, et al. Proteomics and lipidomics in atherosclerotic cardiovascular disease risk prediction. Eur Heart J 2023;44:1594–607. https://doi.org/10.1093/eurheartj/ehad161.
[126] Iannelli F, Lombardi R, Costantini S, Roca MS, Addi L, Bruzzese F, et al. Integrated proteomics and metabolomics analyses reveal new insights into the antitumor effects of valproic acid plus simvastatin combination in a prostate cancer xenograft model associated with downmodulation of YAP/TAZ signaling. Cancer Cell International 2024;24:381. https://doi.org/10.1186/s12935-024-03573-1.
[127] Wright SN, Colton S, Schaffer LV, Pillich RT, Churas C, Pratt D, et al. State of the interactomes: an evaluation of molecular networks for generating biological insights. Mol Syst Biol 2025;21:1–29. https://doi.org/10.1038/s44320-024-00077-y.
[128] Müller S, Matunis MJ, Dejean A. Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. EMBO J 1998;17:61–70. https://doi.org/10.1093/emboj/17.1.61.
[129] Krauß S, Griesche N, Jastrzebska E, Chen C, Rutschow D, Achmüller C, et al. Translation of HTT mRNA with expanded CAG repeats is regulated by the MID1–PP2A protein complex. Nat Commun 2013;4:1511. https://doi.org/10.1038/ncomms2514.
[130] Khoshnan A, Ko J, Tescu S, Brundin P, Patterson PH. IKKα and IKKβ Regulation of DNA Damage-Induced Cleavage of Huntingtin. PLoS ONE 2009;4:e5768. https://doi.org/10.1371/journal.pone.0005768.

[131] Metzler M, Gan L, Mazarei G, Graham RK, Liu L, Bissada N, et al. Phosphorylation of Huntingtin at Ser421 in YAC128 Neurons Is Associated with Protection of YAC128 Neurons from NMDA-Mediated Excitotoxicity and Is Modulated by PP1 and PP2A. Journal of Neuroscience 2010;30:14318–29. https://doi.org/10.1523/JNEUROSCI.1589-10.2010.
[132] Hay DG, Sathasivam K, Tobaben S, Stahl B, Marber M, Mestril R, et al. Progressive decrease in chaperone protein levels in a mouse model of Huntington's disease and induction of stress proteins as a therapeutic approach. Hum Mol Genet 2004;13:1389–405. https://doi.org/10.1093/hmg/ddh144.
[133] Yadavilli S, Hegde V, Deutsch WA. Translocation of human ribosomal protein S3 to sites of DNA damage is dependant on ERK-mediated phosphorylation following genotoxic stress. DNA Repair 2007;6:1453–62. https://doi.org/10.1016/j.dnarep.2007.04.009.
[134] Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, et al. A Protein Interaction Network Links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's Disease n.d.:13.
[135] Sledzieski S, Singh R, Cowen L, Berger B. D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. Cell Syst 2021;12:969-982.e6. https://doi.org/10.1016/j.cels.2021.08.010.
[136] Chen M, Ju CJ-T, Zhou G, Chen X, Zhang T, Chang K-W, et al. Multifaceted protein–protein interaction prediction based on Siamese residual RCNN. Bioinformatics 2019;35:i305–14. https://doi.org/10.1093/bioinformatics/btz328.
[137] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2023;51:D638–46. https://doi.org/10.1093/nar/gkac1000.
[138] Yeger-Lotem E, Sharan R. Human protein interaction networks across tissues and diseases. Front Genet 2015;6. https://doi.org/10.3389/fgene.2015.00257.
[139] Nethisinghe S, Pigazzini ML, Pemble S, Sweeney MG, Labrum R, Manso K, et al. PolyQ Tract Toxicity in SCA1 is Length Dependent in the Absence of CAG Repeat Interruption. Front Cell Neurosci 2018;12:200. https://doi.org/10.3389/fncel.2018.00200.
[140] Latouche M, Fragner P, Martin E, El Hachimi KH, Zander C, Sittler A, et al. Polyglutamine and polyalanine expansions in ataxin7 result in different types of aggregation and levels of toxicity. Mol Cell Neurosci 2006;31:438–45. https://doi.org/10.1016/j.mcn.2005.10.013.
[141] Ek P, Ek B, Zetterqvist Ö. Phosphohistidine phosphatase 1 (PHPT1) also dephosphorylates phospholysine of chemically phosphorylated histone H1 and polylysine. Ups J Med Sci 2015;120:20–7. https://doi.org/10.3109/03009734.2014.996720.
[142] Johnson ES. Protein modification by SUMO. Annu Rev Biochem 2004;73:355–82. https://doi.org/10.1146/annurev.biochem.73.011303.074118.
[143] Nicholson B, Suresh Kumar KG. The multifaceted roles of USP7: new therapeutic opportunities. Cell Biochem Biophys 2011;60:61–8. https://doi.org/10.1007/s12013-011-9185-5.
[144] Dong-Chen X, Yong C, Yang X, Chen-Yu S, Li-Hua P. Signaling pathways in Parkinson's disease: molecular mechanisms and therapeutic interventions. Signal Transduct Target Ther 2023;8:73. https://doi.org/10.1038/s41392-023-01353-3.
[145] Bassi S, Tripathi T, Monziani A, Di Leva F, Biagioli M. Epigenetics of Huntington's Disease. In: Delgado-Morales R, editor. Neuroepigenomics in Aging and Disease, vol. 978, Cham: Springer International Publishing; 2017, p. 277–99. https://doi.org/10.1007/978-3-319-53889-1_15.
[146] Duarte-Silva S, Da Silva JD, Monteiro-Fernandes D, Costa MD, Neves-Carvalho A, Raposo M, et al. Glucocorticoid receptor-dependent therapeutic efficacy of tauroursodeoxycholic acid in preclinical models of spinocerebellar ataxia type 3. J Clin Invest 2024;134:e162246. https://doi.org/10.1172/JCI162246.

[147] Pereira-Leal JB, Levy ED, Teichmann SA. The origins and evolution of functional modules: lessons from protein complexes. Philos Trans R Soc Lond B Biol Sci 2006;361:507–17. https://doi.org/10.1098/rstb.2005.1807.
[148] Musso G, Zhang Z, Emili A. Retention of protein complex membership by ancient duplicated gene products in budding yeast. Trends in Genetics 2007;23:266–9. https://doi.org/10.1016/j.tig.2007.03.012.
[149] Han J-DJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 2004;430:88–93. https://doi.org/10.1038/nature02555.
[150] Prochaska JH, Jünger C, Schulz A, Arnold N, Müller F, Heidorn MW, et al. Effects of empagliflozin on left ventricular diastolic function in addition to usual care in individuals with type 2 diabetes mellitus-results from the randomized, double-blind, placebo-controlled EmDia trial. Clin Res Cardiol 2023;112:911–22. https://doi.org/10.1007/s00392-023-02164-w.
[151] Borodzicz-Jażdżyk S, Jażdżyk P, Łysik W, Cudnoch-Jędrzejewska A, Czarzasta K. Sphingolipid metabolism and signaling in cardiovascular diseases. Front Cardiovasc Med 2022;9:915961. https://doi.org/10.3389/fcvm.2022.915961.