Briefings in Bioinformatics, 2026, 27, bbag195
https://doi.org/10.1093/bib/bbag195
Published: 28 April 2026
Problem Solving Protocol

OXFORD

Building a credibility-based framework for target discovery: Perspectives from tRNA synthetase-linked metabolic diseases

Jaeyoung Choi[1,2], Ina Yoon[2,3], YounSung Jung[4], SeanKyo Han[4], EunHee Kang[4], TaeJin Ahn[5], Sunghoon Kim[1,3,6,7,*]

[1]Institute for Artificial Intelligence and Biomedical Research, Medicinal Bioconvergence Research Center, College of Pharmacy, Yonsei University, Incheon 21983, Republic of Korea
[2]Department of Pharmacy and Yonsei Institute of Pharmaceutical Sciences, College of Pharmacy, Yonsei University, Incheon 21983, Republic of Korea
[3]Department of Integrative Biotechnology, Yonsei University, Incheon 21983, Republic of Korea
[4]Department of Advanced Convergence, Graduate School, Handong Global University, Pohang 37554, Republic of Korea
[5]Department of Life Science, Handong Global University, Pohang 37554, Republic of Korea
[6]Institute for Convergence Research and Education in Advanced Technology, Yonsei University, Incheon 21983, Republic of Korea
[7]College of Medicine, Gangnam Severance Hospital, Yonsei University, Seoul 06273, Republic of Korea

*Corresponding author. 85 Songdogwahak-ro, Yeonsu-gu, Incheon 21983, Republic of Korea. E-mail: sunghoonkim@yonsei.ac.kr

Abstract

While target identification is essential for successful drug discovery, no systematic workflow exists to prioritize potential targets for a given indication. Therefore, this study aims to develop an information-based approach combining text mining, network analysis, and centrality-based prioritization. As a case study, we applied this workflow to identify metabolic disease targets potentially linked to aminoacyl-tRNA synthetases (ARSs). From 1,407,654 PubMed articles, potential ARS interactors and their disease associations were mined. Using these data, the ARS interactor–disease networks were constructed based on edge frequency and citation count. To assess the reliability of these linkages, we used five centrality indices with novel visualization tools and identified 94 high-credibility disease-associated ARS interactors. Among them, two targets (ESR1 and APP) were selected for experimental validation. Although demonstrated in ARS-mediated metabolic diseases, this approach can be similarly used to identify disease-associated factors with credibility scores within any target space of interest.

Keywords text mining, network analysis, therapeutic targets, aminoacyl-tRNA synthetases (ARSs), nutritional and metabolic diseases

Introduction

The success of drug discovery critically depends on the accurate identification of therapeutic targets, as these molecular entities form the foundation for the development of effective and safe treatments. Appropriate target selection markedly increases the likelihood of clinical success, whereas inadequate target identification frequently results in costly failures [1–5]. Despite its importance, drug target discovery remains one of the most challenging steps in drug development due to the inherent complexity of biological systems and disease mechanisms. Current approaches, including genomic, proteomic, and phenotypic screening, are often constrained by incomplete data integration, limited validation models, and the absence of scalable frameworks capable of reliably predicting target relevance [6–17]. Although bioinformatic strategies provide time- and cost-efficient alternatives, they are similarly limited in their ability to identify previously unexplored targets [18–22]. Overcoming these challenges requires the incorporation of novel biological perspectives during data collection and analysis. In this study, we leveraged the emerging biology of aminoacyl-tRNA synthetases (ARSs) to uncover new therapeutic targets relevant to metabolic diseases.

ARSs are essential enzymes responsible for attaching specific amino acids to their cognate tRNAs, thereby ensuring translational fidelity and proper protein synthesis. This canonical catalytic function has long been regarded as their primary role and is highly conserved across evolution [23–27]. However, accumulating evidence has revealed that ARSs possess diverse non-catalytic functions, including roles in cellular stress responses, immune regulation, and metabolic control. These findings challenge the traditional view of ARSs as static components of the translation machinery and instead position them as dynamic regulators of critical cellular processes [28–36].

Consistent with this expanded functional repertoire, ARS dysregulation has increasingly been linked to human diseases, particularly metabolic disorders characterized by impaired energy homeostasis and cellular metabolism. Emerging studies suggest that ARSs influence mitochondrial function, lipid metabolism, and glucose regulation, highlighting their potential roles as modulators of metabolic

Received: January 12, 2026. **Revised:** March 13, 2026. **Accepted:** March 24, 2026

pathways [37–43]. Despite these observations, the mechanistic connections between ARSs and metabolic diseases remain incompletely understood, underscoring the need for systematic investigation.

To examine how the focus of ARS research has evolved, we performed a Dirichlet multinomial regression (DMR) analysis of ARS-related literature (Supplementary Table 1). DMR extends latent Dirichlet allocation by incorporating external metadata, such as publication year, allowing topic distributions to be conditioned on temporal trends [44–47]. Our analysis revealed a steady increase in the prominence of topics related to "Disease" and "Pathway Regulator" over recent decades (Supplementary Fig. 1A). Although these topics remain less prevalent than traditional categories such as "Traditional Enzyme" and "Structure," their growth has been substantial. Between 2000 and 2010, the "Pathway Regulator" topic increased by over 120%, while the "Disease" topic rose by more than 250%, reflecting a clear shift in research priorities toward the non-canonical roles of ARSs in disease pathways and metabolic regulation.

To further dissect disease-related trends, we analyzed ARS-associated literature across MeSH disease categories, focusing on seven representative groups (complete results are provided in Supplementary Table 2). Among these, "Nutritional and Metabolic Diseases (C18)" consistently exhibited a notable presence (Supplementary Fig. 1B), consistent with the established involvement of ARSs in core metabolic processes such as amino acid metabolism and energy regulation. However, compared with categories such as Nervous System Diseases (C10), Musculoskeletal Diseases (C05), and Neoplasms (C04), the relative underrepresentation of metabolic diseases highlights a critical gap in ARS-focused research and suggests untapped potential for target discovery in this domain.

Building on these observations, we applied text mining approaches to systematically explore connections between ARSs and metabolic diseases within the biomedical literature. Importantly, we incorporated ARS interactors as intermediary nodes within molecular networks to capture indirect relationships that may not be apparent from ARS-centered analyses alone. Comparison of ARS-centered and interactor-centered disease networks revealed distinct patterns in disease category prominence. While Nervous System Diseases (C10) dominated the ARS-centered network and Neoplasms (C04) ranked third (Supplementary Fig. 2A), Neoplasms emerged as the most prominent category in the interactor-centered network, with Nervous System Diseases falling to second place (Supplementary Fig. 2B).

Notably, "Nutritional and Metabolic Diseases (C18)" maintained the same rank in both networks, yet its interpretation differed substantially. In the ARS-centered network, this category reflects the direct involvement of ARSs in metabolic regulation and energy homeostasis. In contrast, in the interactor-centered network, metabolic disease associations arise through indirect connections mediated by ARS interactors, integrating ARSs into broader systemic pathways. This distinction highlights how ARS interactors expand the functional landscape of ARSs beyond their canonical roles and underscores the value of network-based approaches for uncovering previously unrecognized therapeutic opportunities.

Results

Figure 1 provides an overview of the analytical workflow employed in this study to investigate the relationship between ARSs and metabolic diseases. The analysis began with large-scale literature collection from public databases, including PubMed and Web of Science [48, 49], focusing on the MeSH category "Nutritional and Metabolic Diseases" (C18) [50], which served as the foundation for subsequent analyses.

Named entity recognition (NER) and relation extraction (RE) were then applied to identify ARS interactors, disease entities, and their relationships within the collected literature [51–54]. The extracted associations were used to construct a graph-based network representing interactions between ARS interactors and metabolic diseases, with ARS specificity of the disease terms verified to ensure analytical reliability.

To prioritize key ARS interactors, multiple network centrality measures—including Degree (Cd), PageRank (Cp), Betweenness (Cb), Closeness (Cc), and Eigenvector Centrality (Ce)—were calculated [55–63]. These metrics were integrated using bump-chart thresholding to identify consistently high-ranking nodes [64], and the robustness of the results was assessed using overlap heatmap analysis.

Finally, the prioritized ARS interactors were incorporated into a refined network linking ARSs to metabolic diseases, enabling the identification of candidate drug targets. The biological relevance of the prioritization strategy was further supported by experimental validation. To facilitate reproducibility and scalability, the entire workflow was implemented as an automated system, allowing its application to additional disease contexts and biological questions.

Data collection

To investigate the relationship between ARSs and metabolic diseases, we collected 1,407,654 PubMed articles, focusing on the "Nutritional and Metabolic Diseases" category (Tree Number: C18, MeSH Unique ID: D009750) within MeSH. This category included nutritional disorder-related terms resulting from poor absorption or nutritional imbalances, and metabolic disorders caused by defects in the biosynthesis (anabolism) or breakdown (catabolism) of endogenous substances [50].

Simultaneously, we limited our dataset to publications up to 2022 to ensure analytical reliability and validity. Citation count was a key metric for assessing the academic impact of a study, and papers published from 2023 onward had few or no citations due to the limited time available for citation accumulation. Including these papers could compromise citation-based analyses and distort estimates of scientific influence. Major bibliographic databases, including PubMed, undergo annual updates, typically early in the year, providing fully curated datasets. To ensure analytical consistency, we used a complete, finalized dataset rather than newly added records that may not be fully integrated into these databases. Moreover, given the large-scale nature of this study—which processed over 1.4 million articles—adding recent publications would require reprocessing the entire dataset, demanding extensive computational resources and time for data preprocessing, network construction, and centrality analysis. Given these constraints, restricting our dataset to 2022 publications was the most feasible approach, ensuring analytical rigor and computational efficiency.

Furthermore, citation count for the retrieved articles was collected from Web of Science as an additional data source for higher-level analysis. This additional data provided an indicator of the academic influence and relevance of each article. However, platform discrepancies limited citation data to 982,358 of 1,407,654 articles (≈69.8% of the initial PubMed dataset). Articles without citation data were retained but excluded from citation-based analyses.

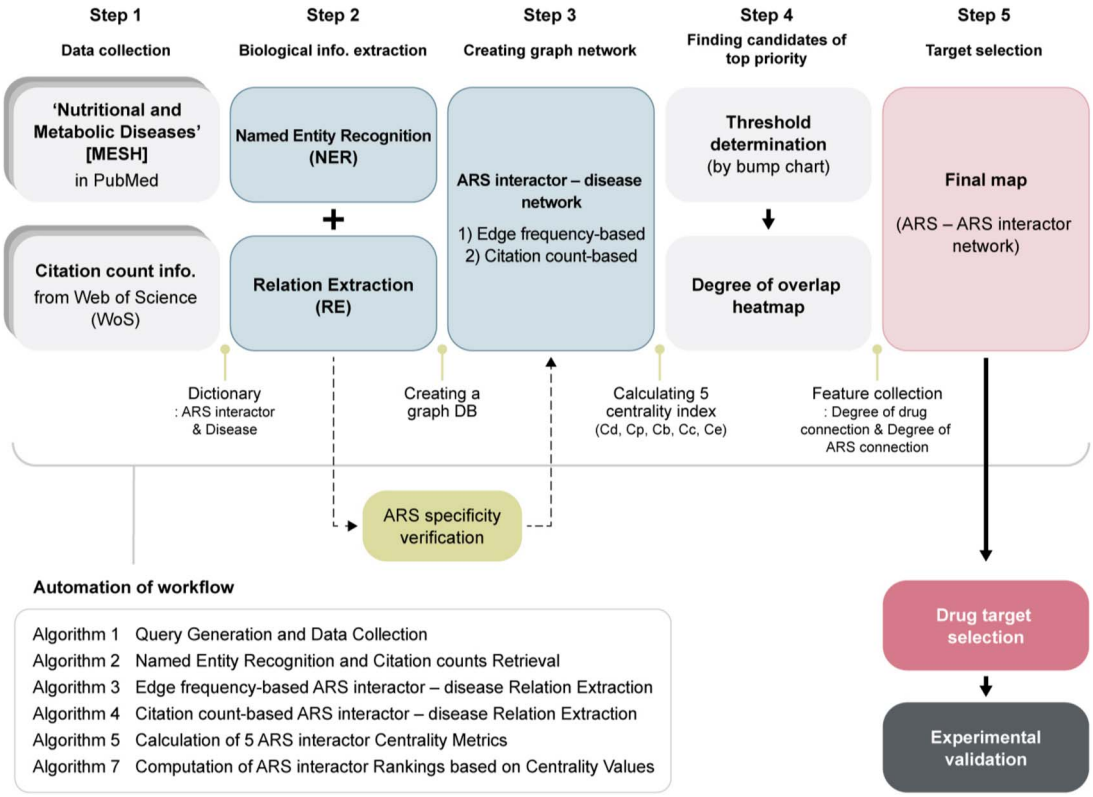

Figure 1 Systematic workflow to identify the potential drug targets associated with metabolic diseases focusing on the network of ARSs and their interactors. This figure outlines the stepwise process undertaken in this study to investigate the relationship between ARSs and metabolic diseases. Step 1 involves data collection from PubMed ("Nutritional and Metabolic Diseases" [MeSH] category) and Web of Science (citation information). In step 2, NER and RE are employed to identify ARS interactors and disease relationships. Step 3 focuses on constructing a graph network of ARS interactor–disease interactions based on edge frequency and citation data. Step 4 calculates five centrality indices and applies thresholding with bump charts and overlap heatmaps to prioritize key ARS interactors. Step 5 visualizes the final ARS–ARS interactor network, highlighting interactors of top priority for drug target selection. Selected candidates were subsequently evaluated through experimental validation. While this workflow was designed to establish a systematic framework for linking ARS biology with metabolic disease mechanisms, it can be broadly applied to identify novel druggable targets in various disease areas. To enhance usability, we automated each step using appropriate algorithms.

Building dictionaries

We constructed two dictionaries—an ARS interactor and a disease dictionary—for automated named entity recognition. These dictionaries were built from reliable databases to ensure accurate, consistent entity identification.

The ARS interaction dictionary draws from the "Interactions" section of each ARS in the NCBI Gene Database. This resource integrates data from five sources: BIND, BioGRID, EcoCyc, HIV-1 protein interactions, and HPRD. Each dictionary entry lists the gene identifier, gene name, supporting literature, and experimental verification method. These entries were organized to enable efficient ARS interactor recognition in text [65].

The disease dictionary was derived from the "Diseases" category ("C" category) of MeSH. Only the most granular, indivisible terms in the tree structure were selected as dictionary entries. Each entry included the disease term and tree number for precise identification in text [66].

As of September 2022, the ARS interactor and disease dictionaries contained 2,357 and 4,791 entries, respectively. These dictionaries supported NER and RE tasks, enabling comprehensive analysis of ARS–disease interactions, and can be found in the supplementary information.

Evaluation of metabolic disease-associated targets from the perspective of ARS biology

This study aims to identify diseases co-occurring with ARS interactors through text mining techniques and to determine the most central ARS interactors in ARS–disease pairs (Fig. 1). We recognized a potential bias where some diseases were paired with ARS interactors simply because they frequently appear in a nonspecific manner. In such cases, many ARS interactors may be linked to these diseases without a meaningful pathological association. Other entities beyond ARS interactors may also link to these diseases due to their high frequency in the literature.

To address this issue, we verified whether the observed relationships were ARS interactome-specific rather than driven by overall disease frequency. Verification was performed from two perspectives: edge frequency- and citation count-based ranking comparisons. Dictionary-based NER and RE were performed on 1,407,654 PubMed articles and citation data, followed by network analysis.

Edge frequency-based ranking comparison

The analysis aimed to compare disease frequency rankings in the C18 literature with their importance rankings in ARS interactor–

disease pairs. In the former, disease rankings were based on their frequent appearance in the literature. For the latter, a network of all ARS–disease pairs was constructed, and disease importance was measured using Cd, which represents total node connections. Diseases co-mentioned with ARS interactors in multiple documents had higher Cd values, reflecting their prominence in the ARS interactome. Using this approach, rankings of 2,245 diseases from 1,407,654 documents were calculated and compared.

Citation count-based ranking comparison

Disease frequency calculation based on citation count differed from the edge frequency approach. In edge frequency analysis, a disease occurrence was counted once per paper in which it appeared. In this analysis, each paper's citation count (n) from the Web of Science was also considered. The citation count contributed proportionally to the disease frequency.

Nevertheless, using raw citation counts posed a risk of bias, as older papers naturally accumulate more citations over time regardless of quality. To address this, we applied a correction formula:

$$\textit{Adjusted Citation Count} = \textit{Citation Count} / \left(2023 - \textit{Publication Year}\right).$$

This adjusted value was used as the citation-based disease frequency, ensuring a balanced evaluation of disease importance.

Similarly, in edge frequency ranking, disease importance was measured using Cd, modified to incorporate the adjusted citation count. Previously, the link weight was fixed at one, whereas here the adjusted citation count served as the link weight. This adjustment ensured that disease centrality reflected connection frequency with ARS interactors and the academic impact of linking papers.

Ranking comparison results

Substantial ranking changes were observed in most diseases across the edge frequency- and citation count-based analyses (Supplementary Fig. 3), except a few—including diabetes mellitus, obesity, hyperglycemia, inflammation, and metabolic syndrome—which consistently ranked highest. This divergence in ranking underscores the unique influence of ARS interactors on disease relevance, indicating their links are not merely general literature trends. The findings highlight the specificity of ARS interactors in disease networks, underscoring the need to assess their importance in a context-specific biological framework. Consequently, this distinction supports the validity of the subsequent analyses, which further clarify the unique ARS interactor–disease relationships from an ARS-biology perspective.

Generation of edge frequency-based ARS interactor–disease network

ARS interactor–disease pair extraction

Using 1,407,654 PubMed articles in the "Nutritional and Metabolic Diseases" MeSH category and ARS interactor–disease dictionaries, we used NER and RE to identify ARS interactor–disease relationships. Terminology Unification and Abbreviation Disambiguation were employed to standardize terms and ensure consistency (Box 1). A relationship pair was identified only when the ARS interactor and disease co-occurred within the same sentence. This sentence-level constraint was adopted to prioritize extraction precision while minimizing spurious edges and potential false positives in the ARS interactor–disease network. A conservative lower-bound aggregation at the document level indicated that relaxing this constraint resulted in only a modest increase in edges (∼1.5%–1.6%) and unique pairs (∼3.5%–3.8%) (Supplementary Table 3), suggesting that the sentence-level rule maintains high extraction precision while minimally affecting overall network coverage. Each pair was counted once per article, even if repeated, to avoid overrepresenting frequently mentioned relationships. This systematic approach produced 175,903 unique pairs involving 1,329 ARS interactors and 2,245 diseases, suggesting diverse pathological implications of ARSs in metabolic diseases.

Box 1 Terminology unification and abbreviation disambiguation.

- **Terminology unification** consolidates various expressions referring to the same substance or disease into a single representation, preventing scattered information and improving analytical accuracy. Therefore, all expressions in the original text that match a specific term are replaced with a unique identifier (ID). These IDs are converted back to their standardized terms after extraction tasks are completed, effectively resolving inconsistencies.
- **Abbreviation disambiguation** addresses the challenge of distinguishing between identical abbreviations with different full names, a common source of false positives. This is achieved by identifying the full names of each abbreviation in the abstract and replacing all abbreviations with their respective full names. Then, only those full names found within a predefined dictionary are extracted as valid entities. This process ensures accurate linkage between abbreviations and their intended references, reducing errors in data extraction.

Edge frequency-based weight assignment

Based on the relation pairs created above, we summed the frequencies of each pair across the literature. For example, the "INS–Diabetes Mellitus, Type 2" pair appeared in 14,697 papers, resulting in a frequency of 14,697, regardless of the number of times it appeared in a single article. Consequently, 22,220 unique pairs were derived. These unique ARS interactor–disease pairs were visualized as a network map. This network included 1,329 ARS interactors and 2,245 diseases, with node size representing weighted degree centrality, indicating how often an ARS interactor or disease connected to others.

Disease grouping

The 2,245 diseases were grouped based on their corresponding MeSH categories. While we focused on C18 category diseases during literature extraction, others outside this category also appeared. To exclude them, diseases were grouped into higher MeSH categories and connections to all categories except C18 were removed. Through these processes, only the relations between true C18-group diseases and their ARS interactors were retained.

We observed that one individual disease was often matched to multiple higher-level categories. For example, in the "MDH2–Diabetes Mellitus, Experimental" pair, the disease belongs to Nutritional and Metabolic Diseases (C18) and Endocrine System Diseases (C19) categories. After disease grouping, the "MDH2–Diabetes Mellitus, Experimental" pair was divided into "MDH2–Nutritional and Metabolic

Diseases (C18)" and "MDH2–Endocrine System Diseases (C19)." Consequently, relation pairs increased from 175,903 to 373,191 after disease grouping.

Supplementary Fig. 4 (upper) illustrates the grouping results as a network, showing all 23 MeSH disease categories in varying proportions. Since only papers on Nutritional and Metabolic Diseases (C18) were analyzed, this field appeared proportionally larger than that of other disease groups. Among them, Endocrine System Diseases (C19), Pathological Conditions, Signs and Symptoms (C23), and Nervous System Diseases (C10) were prominent, suggesting a closer link to Nutritional and Metabolic Diseases (C18).

Focusing on the C18 category

Among the 373,191 pairs generated after disease grouping, 118,749 relations were connected to 297 diseases exclusively classified under C18, compared with 2,245 before grouping. Relation frequencies across the literature were summed to assign weights, and the final network was constructed only for C18. As stated above, node size reflected these weights, allowing visual identification of ARS interactors or diseases with greater importance in our analysis (Fig. 2A).

Generation of citation count-based ARS interactor–disease network

ARS interactor–disease pair extraction

From the PubMed data categorized under the Nutritional and Metabolic Diseases (C18) category in MeSH, 979,894 documents were utilized, all with Web of Science information and at least one citation, covering publications through 2022. Using the same biological information extraction approach applied in the edge frequency ARS interactor–disease network, the final dataset contained 146,352 unique relation pairs. These pairs involved 1,275 ARS interactors and 2,062 unique diseases, forming the basis for subsequent network construction and analysis.

Citation count-based weight assignment

The number of citations for each source paper, obtained from Web of Science, was used as the score for every extracted ARS interactor–disease relation pair. Overall, 2,464 documents with zero citations or missing citation data were removed from further analysis. Because assigning weight solely by citation count can bias results toward older publications, we corrected for this by applying the formula "total citations divided by the time elapsed from the publication year to 2023", as previously described. All calibrated citation values appearing across different documents were then summed, and a final weighted score was assigned to each ARS interactor–disease pair. Using these weighted scores, the network map was generated and contained 19,383 unique relation pairs.

In the network, the number of ARS interactors was 1,275, and that of diseases was 2,062, slightly lower than in the edge frequency network owing to the exclusion of papers with zero or missing citation data. Weighted degree centrality, reflecting calibrated citation values, was used to determine node size, enabling the relative importance of each ARS interactor or disease to be visually represented.

Disease grouping

The next step involved grouping the 2,062 individual diseases according to the MeSH tree structure. Supplementary Fig. 4 (lower) shows that nearly all disease categories appear, consistent with the pattern observed in the edge frequency analysis. However, Nutritional and Metabolic Diseases (C18) constituted the largest proportion. Citation count-based analysis additionally highlighted Nervous System Diseases (C10), Pathological Conditions, Signs and Symptoms (C23), and Endocrine System Diseases (C19) as top-ranking categories following C18. These findings were consistent with the edge frequency results, although the precise ranking among these categories differed.

Focusing on the C18 category

We selected 98,940 relations associated with the C18 category from the 312,344 relations obtained after disease grouping. The number of included diseases decreased to 286, compared to 297 in the edge frequency analysis, because papers with zero or missing citation data were excluded. The relation network between these 286 diseases and their matched ARS interactors was visualized using the calibrated citation counts as weights (Fig. 2B). Additionally, the cumulative calibrated citation scores were represented as the node size, indicating the relative importance of each disease or interactor.

Determination of target priority via the comparative analysis of five centrality network analyses

We compared the rankings of ARS interactors derived from the edge frequency- and citation count-based networks. In network analysis, centrality measures are commonly used to identify important nodes, with higher centrality values indicating greater structural importance within the network. To ensure reliable identification of disease-associated ARS interactors, we applied five widely used centrality measures: Cd, Cp, Cb, Cc, and Ce (Box 2) [55–63]. These metrics capture complementary aspects of network topology, including local connectivity, recursive importance, mediation of shortest paths, global reachability, and influence through highly connected neighbors, respectively. To assess potential redundancy among these metrics, we computed pairwise Spearman rank correlations among the five centrality-based rankings (Supplementary Fig. 5), which showed moderate to strong correlations but non-identical rankings, supporting partial complementarity among the metrics. These measures were then applied to rank ARS interactors identified in both the edge frequency- and citation count-based networks, and an identical set of 1,120 ARS interactor nodes was used to assess their relative importance within a consistent framework for direct comparison.

One major challenge in this study was that multiple nodes frequently shared identical centrality values, resulting in duplicated rankings. As summarized in Supplementary Table 4, the proportion of nodes belonging to tied groups ranged from 24.6% to 93.3% across the 10 metric–network combinations, indicating that ties were common and could reduce the effective resolution of the ranking if left unresolved. To address this issue, we developed a deterministic tie-breaking strategy termed Cross Reference, which resolves ties in a primary ranking using the complementary ranking derived from the alternate network weighting. Specifically, ties in edge frequency rankings were resolved using citation-based rankings, whereas ties in citation-based rankings were resolved using edge frequency rankings. For example, in the degree centrality (Cd) edge frequency ranking, RBM8A and FUS exhibited identical edge frequency values and

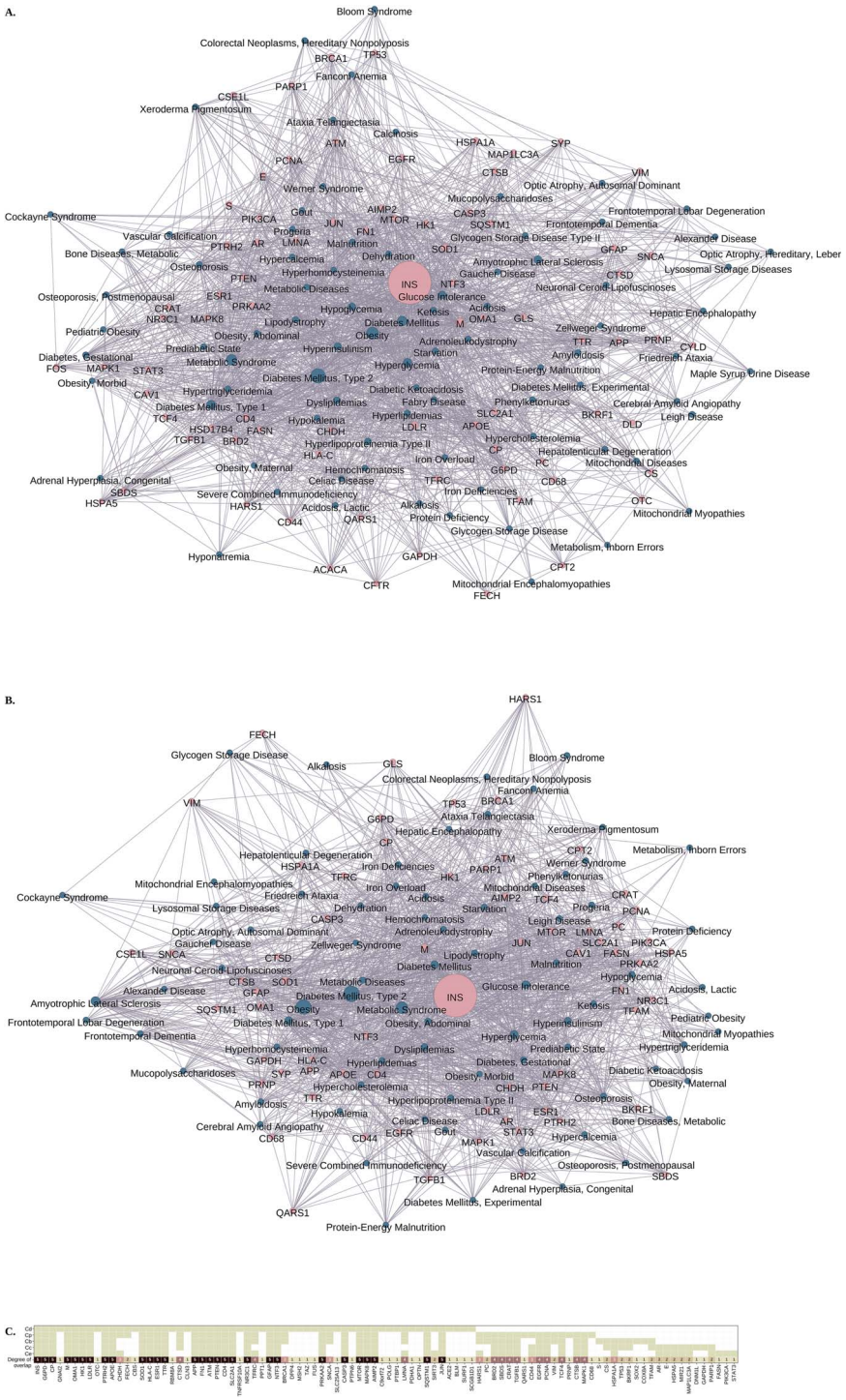

Figure 2 The networks of ARS interactors associated with nutritional and metabolic diseases. (A) Edge frequency-based ARS interactor–disease network. ARS interactor–disease pairs extracted from 1,407 654 PubMed articles in the field of "Nutritional and Metabolic Diseases (C18)" were used (175,903 pairs) for analysis. Weights were assigned by adding up the frequencies of specific pairs that appeared in different literature, and 2,245 unique diseases included in the entire set of ARS interactor–disease pairs were matched to the upper disease categories of MeSH to which they each belong. Among these, only the relations with "Nutritional and Metabolic Diseases (C18)" were retained, and they were used in this network. Here, the size of the node is a weighted degree centrality value that indicates how many times a specific ARS interactor or disease has been linked with other elements, and the thickness of the link represents how many papers a specific ARS interactor–disease pair has appeared in the entire literature population. Nodes with a weighted degree centrality of 20 or more are shown in this figure for visual simplification. **(B) Citation count-based ARS interactor–disease network.** Among the 1,407,654 documents extracted above, 979,894 that had citation count at least one in Web of Science were selected. The final weight was given to the ARS interactor–disease pairs by adding the corrected citation count. Through grouping 2,062 individual diseases within the entire pair according to the MeSH tree structure, only relations connected to the "Nutritional and Metabolic Diseases (C18)" were left. The size of the node is the cumulative sum of the corrected citation scores, and the thickness of the link is the sum of all corrected citations of papers in which a specific ARS interactor–disease pair appears. Also, only nodes with a weighted degree centrality of at least 20 were used in the graph for visibility. **(C) Degree of centrality overlap heatmap.** (Upper) A heatmap was generated for 94 ARS interactors selected based on threshold determination using 5 different centrality analysis methods (Cd, Cp, Cb, Cc, and Ce; see Box 2 for details). (Lower) The color intensity represents the degree of overlap for each ARS interactor across the five centrality analysis methods.

Box 2 Overview of five centrality measures.

Measure	Definition	Formula	Key references
Degree centrality	The total number of connections (edges) a node has in the network.	$DC(v) = \sum_{u \in N(v)} 1$	[61]
PageRank	A measure of node importance based on the probability of visiting the node in a random walk across the network.	$PR(v) = \frac{1-d}{N} + d \cdot \sum_{u \in N(v)} \frac{PR(u)}{d_u}$	[58, 59, 63]
Betweenness centrality	The fraction of shortest paths between any two nodes that pass through a given node.	$BC(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$	[57, 60]
Closeness centrality	The reciprocal of the average shortest path length from a node to all other nodes in the network.	$CC(v) = \frac{1}{\sum_{u \neq v} d(v,u)}$	[55, 62]
Eigenvector centrality	A measure of node influence based on connections to other influential nodes, considering the quality of connections rather than just quantity.	$EC(v) = \frac{1}{\lambda} \sum_{u \in N(v)} A_{vu} \cdot EC(u)$	[56]

Formula explanation

1. Degree centrality ($DC(v)$):
   - Sum of all direct connections ($N(v)$) to a node (v).
2. PageRank ($PR(v)$):
   - d: Damping factor (commonly set to 0.85).
   - d_u: Out-degree of node u.
   - N: Total number of nodes.
3. Betweenness centrality ($BC(v)$):
   - $\sigma_{st}(v)$: Number of shortest paths between nodes s and t passing through v.
   - σ_{st}: Total number of shortest paths between s and t.
4. Closeness centrality ($CC(v)$):
   - Reciprocal of the sum of shortest path distances ($d(v, u)$) from node v to all other nodes.
5. Eigenvector centrality ($EC(v)$):
   - A_{vu}: Adjacency matrix entry indicating the connection between nodes v and u.
   - λ: Eigenvalue of the adjacency matrix.

therefore shared rank 17 prior to tie-breaking. Referring to the complementary citation-based ranking (RBM8A: 8; FUS: 7) deterministically resolved this tie, resulting in final positions of rank 17 for FUS and rank 18 for RBM8A. Conversely, when ties occur in citation-based rankings, the same procedure is applied in the opposite direction using the complementary edge frequency ranking. By systematically resolving duplicated values in this manner, Cross Reference increased the effective resolution of the ranking while preserving the primary centrality structure and affecting only nodes with identical scores.

Bump charts and threshold determination

Following the methodology described above, bump charts were generated for Cd, Cp, Cb, Cc, and Ce (Supplementary Fig. 6). A bump chart lists the elements ranked by two criteria in the left and right columns and connects the identical elements to show whether their relative importance changes across the two viewpoints [64]. In this analysis, 1,120 ARS interactors present in both networks were placed in the left column according to edge frequency-based centrality and in the right column according to citation count-based centrality. Each interactor was then connected across the two columns. The slope and direction of these connecting lines illustrate the degree of ranking inconsistency between the two criteria, allowing visual assessment of how differently each interactor is prioritized under edge frequency versus citation count information.

We aligned the five bump charts generated from the different centrality methods to identify ARS interactors that consistently maintained high rankings from the edge frequency and citation count perspectives. The relative positions of the highest-ranked interactors showed minimal variation between the two columns. Based on the overall line patterns, we selected the top 50 interactors in each chart (Supplementary Fig. 6; dark grey line) because the upper-ranking region showed relatively limited positional fluctuation between the two ranking criteria. Consistent with this observation, quantitative stability analysis revealed increasing rank divergence beyond the top 50 region (median $\Delta = 4.0$ for ranks 31–50, 7.0 for ranks 51–70, and 11.0 for ranks 71–100), whereas the top 50 interactors maintained strong concordance between the two ranking schemes (mean prefix Spearman $\rho = 0.924$; Supplementary Table 5). Because the top 50 sets differed slightly between edge frequency- and citation count-based rankings, we included 59, 52, 54, 53, and 54 ARS interactors in the Cd, Cp, Cb, Cc, and Ce charts, respectively. Combining all the selected interactors across the 5 charts resulted in 94 ARS interactors.

Degree of overlap heatmap

We constructed a heatmap for the 94 selected ARS interactors to examine the degree of overlap among interactors identified by the 5 centrality methods (Supplementary Table 6 and Fig. 2C). In this heatmap, the x-axis lists the ARS interactors and the y-axis lists the centrality methods, with each selected interactor marked by color in the corresponding cell. For example, all cells for INS were colored because it appeared among the top 50 interactors in all 5 charts

(overlap = 5). In contrast, only one cell for GNAI2 was colored since it was selected in the Cd chart only (overlap = 1). The overall degree of overlap for each interactor is represented by color intensity and summarized numerically in the bottom row.

Feature collection of 94 ARS interactors

Using the ARS interactors selected as high-ranking interactors for C18-group diseases, we generated a connection map with ARSs. This map helps to identify potential drug targets for C18 diseases by focusing on ARS biology. To achieve this goal, we designed a network structure that visually distinguishes drug targets based on the following criteria.

(1) Degree of disease connection: focusing on the Nutritional and Metabolic Diseases (C18) section of MeSH, the ARS interactors inherently reflect this information through the selection processes applied in the current study.

(2) Degree of drug connection: to evaluate the druggability of the selected ARS interactors, we considered whether each interactor had previously been used as a drug target. We retrieved relevant publications from PubMed by searching for the ARS interactor together with the terms "drug or drugs." For each ARS interactor, the NCBI Gene database provided the Official Symbol, Official Full Name, and synonyms, which were combined using "OR." This gene-related phrase was then connected with "drug or drugs" using AND. For example, the query for albumin was ("ALB" OR "albumin" OR "HSA" OR "FDAHT" OR "PRO0883" OR "PRO0903" OR "PRO1341") AND ("drug" OR "drugs"), which was submitted to PubMed to identify publications connecting albumin to drugs.

(3) Degree of ARS connection: ARS-interactor pairs were created using the interactor information collected when constructing the ARS interactor dictionary at the beginning of the study. The 94 previously selected interactors were then reconstructed from an ARS interactor-centric perspective. Interactors connected to multiple ARSs were assigned a higher ARS connection degree, and those with more connections were prioritized in the final target selection. Supplementary Table 7 presents the data collected in early 2023 on the extent of drug and ARS linkages.

Network map between ARSs and their interactors selected for metabolic diseases

We constructed a network connecting ARSs with the 94 ARS interactors selected through thresholding across 5 centrality criteria. ARS node sizes were uniform, while ARS interactor node sizes varied according to drug connection degree (Supplementary Table 7), with the largest set to 100%. The network comprises 359 connections, and connected nodes are positioned proximally (Fig. 3A). Among the 94 ARS interactors, 29 showed the highest overlap (overlap = 5), reflecting high confidence in their pathological association with C18-group diseases. Based on this, an additional network map was generated using these 29 top-ranked ARS interactors following the same principle as the initial network (Fig. 3B).

The two network maps—one with 94 interactors and another with 29—highlighted different prominent targets. In the larger network, EGFR and TP53 were most prominent, followed by MTOR, TGFB1, STAT3, and ESR1, whereas in the smaller network, ESR1, MTOR, and APP were more noticeable.

Experimental validation of the network map

Among the 29 ARS interactors identified as potential druggable targets, estrogen receptor 1 (ESR1) and amyloid beta precursor protein (APP) were selected for experimental validation of their interactions with ARSs (Fig. 3B). Both proteins were consistently included in the highest overlap tier (degree of overlap = 5), showed high drug-related connectivity, and exhibited broad ARS interaction profiles (Supplementary Table 7), while representing distinct biological contexts relevant to metabolic disease. To explore their disease relevance, ESR1- and APP-associated diseases connected from Fig. 2A and B were extracted and mapped to higher-level metabolic disease categories. In the resulting ARS interactor–disease network map, ESR1 was associated with metabolic bone diseases, glucose metabolism disorders, and metabolic syndrome, while APP was related to proteostasis deficiencies, metabolic inborn errors, and glucose metabolism disorders (Fig. 4A and B). Since these interactions were identified through proteomics-based or reconstituted complex approaches, we sought to validate them in a cellular context using a co-immunoprecipitation assay. For APP, two isoforms—APP695 and APP751, named for their amino acid lengths—were tested. APP695 is predominantly neuronal, whereas APP751 is mainly expressed in nonneuronal cells, indicating possible distinct pathological roles. While testing all ARS interactions was not feasible, ESR1 showed potential interactions with MARS1 (methionyl-tRNA synthetase 1), while APP interacted with IARS1 (isoleucyl-tRNA synthetase 1) and AIMP1 (ARS-interacting multifunctional protein 1) (Fig. 4C). These interactions warrant further investigation to elucidate their potential roles in ESR1- and APP-related metabolic diseases.

Automation of workflow

Considering the general applicability of this workflow for identifying new druggable targets, we developed an automated system to systematically analyze relationships between ARS interactors and metabolic diseases, thereby enhancing efficiency and scalability. By automating the entire process—from literature search and biological information extraction to network construction and centrality analysis—this framework was employed to streamline data processing and minimize manual intervention. This automated pipeline improves research efficiency and scalability, thereby allowing consistent analysis across multiple disease categories or biological groups.

Discussion

Target identification remains a critical challenge in drug discovery, particularly for complex diseases involving multiple molecular interactions. Traditional genomic, proteomic, and phenotypic screening approaches are often limited in their ability to integrate large-scale biomedical literature and to prioritize novel, functionally relevant targets. In this study, we developed an information-driven pipeline that combines text mining and network analysis to systematically identify key interactors within the ARS-mediated metabolic disease landscape. By analyzing over 1.4 million PubMed articles using the terminology normalization and abbreviation disambiguation procedures described in Box 1 and applying centrality-based prioritization, we identified and experimentally validated novel ARS interactors, including ESR1 and APP, with potential roles in metabolic disorders. These results demonstrate how computational–experimental integration can uncover previously unrecognized disease mechanisms and

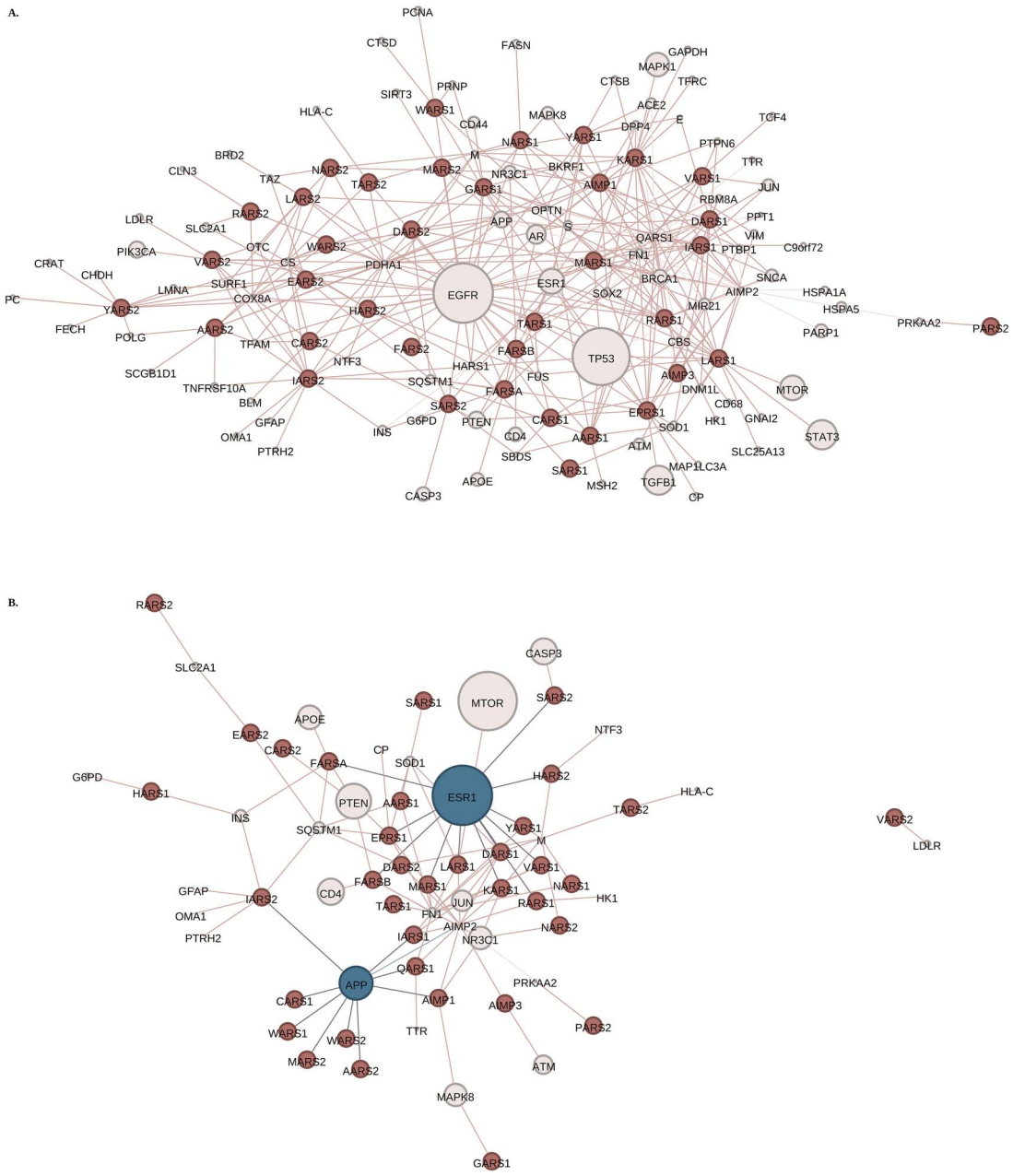

Figure 3 Network integrating ARS connectivity and druggability. (A) Network map between ARSs and 94 ARS interactors selected from 5 centrality analyses. Using the selected 94 candidates, the final network was organized to reflect three features. Degree of drug connection was embedded in the node size of ARS interactors. The size of the interactor node with the largest drug connection was set to 100, and the rest were scaled based on the corresponding percentage. On the other hand, the size of the ARS nodes and the thickness of the links were displayed as constant. The connected nodes were laid out close to each other on the network. **(B) Network map between ARSs and 29 ARS interactors selected from all the 5 centrality** analyses. The ARS network was constructed using 29 ARS interactors identified across all 5 centrality analysis methods to provide a more focused picture. Based on this map, we selected ESR1 and APP for further analysis of their association with metabolic diseases and for experimental verification.

provide a scalable strategy for drug target prioritization in complex biological systems.

Beyond the specific interactions identified here, this automated pipeline offers a flexible framework for expanding ARS research through large-scale meta-analyses and cross-disease investigations. Rather than focusing exclusively on Nutritional and Metabolic Diseases (C18), the approach can be extended to examine cross-disease relationships involving Endocrine System Diseases (C19) and Neoplasms (C04). In the longer term, integration with machine learning or AI-based predictive models may further enhance automated target discovery and improve mechanistic insight. In addition, the development of a self-updating pipeline that continuously incorporates newly published literature would enable the system to adapt to the rapidly evolving biomedical research landscape.

Figure 4 Integrative association of ESR1 and APP with metabolic diseases and experimental validation of their interaction with ARSs. (A) ESR1–metabolic disease network. (Edge frequency-based) This network visualization represents the association between ESR1 and metabolic diseases based on edge frequency analysis. Each edge connects ESR1 to a specific metabolic disease, with the edge thickness reflecting the frequency of co-occurrence in the analyzed PubMed dataset. A higher frequency suggests a stronger textual association between ESR1 and the disease. The node size of "Bone Diseases, Metabolic" is larger, indicating its prominence in the dataset, implying a significant literature-based link between ESR1 and metabolic bone diseases. This approach helps identify frequently mentioned connections between ARS interactors and metabolic disease categories, providing a literature-driven prioritization of potential targets. **(Citation count-based)** This network visualization presents the relationship between ESR1 and metabolic diseases, constructed using citation count-based analysis. In this approach, the strength of associations is determined by the cumulative citation count of the publications supporting each connection, ensuring that highly cited research contributes more to the network structure. As shown, "Bone Diseases, Metabolic" again appears as a major association, reflecting its substantial academic impact and reinforcing its potential biological relevance. This analysis provides a complementary perspective by weighting associations based on their scholarly influence, helping to filter out less impactful co-occurrences from the dataset. **(B) APP–metabolic disease network. (Edge frequency-based)** Unlike in ESR1, "Proteostasis Deficiencies" and "Metabolism, Inborn Errors" are most prominent, suggesting that they are frequently co-mentioned with APP in studies of metabolic diseases. **(Citation count-based)** Compared with the edge frequency-based network, "Proteostasis Deficiencies" remains the most prominent association, reinforcing the academic importance of APP's role in protein quality control and degradation pathways. In addition, we can also see the relative salience of "Glucose Metabolism

Importantly, this framework also supports experimental validation by enabling the prioritization of druggable targets for downstream biochemical and functional assays. Through multistep analyses and experimental confirmation, we identified disease-associated interactions between ARSs and key metabolic disease-related proteins. Specifically, ESR1 interacted with MARS1, while APP interacted with IARS1 and AIMP1, revealing previously unrecognized molecular connections that may contribute to disease pathogenesis. Given the established roles of ESR1 in metabolic bone disease, glucose metabolism disorders, and metabolic syndrome, its interaction with MARS1 suggests a potential role for ARSs in hormone-responsive metabolic regulation. Similarly, the association of APP with multiple ARSs suggests that ARSs may modulate APP-related pathways involved in proteostasis imbalance, inborn metabolic errors, and glucose metabolism disorders. While further studies are required to elucidate the precise molecular mechanisms underlying these interactions, the present findings establish a robust and broadly applicable framework for efficient drug target identification across diverse disease contexts.

Despite these strengths, several methodological limitations should be considered. The current framework constructs literature-derived networks primarily through sentence-level co-occurrence, which enables scalable analysis across large corpora but does not explicitly capture interaction directionality, polarity, or long-range linguistic dependencies. Although dictionary normalization procedures were implemented to mitigate alias-related ambiguity—including canonical symbol mapping and abbreviation disambiguation—residual ambiguity may still occur in rare cases. In addition, literature-based analyses can be influenced by publication or citation bias toward extensively studied genes, and MeSH indexing practices may evolve over time as the biomedical literature expands. To evaluate the robustness of the framework in the context of such temporal changes, we conducted an additional analysis using publications from 2023 to 2024 within the same MeSH category. Notably, 88 of the 94 prioritized ARS interactors (93.6%) were still observed in the recent corpus, and the majority of Tier 1 candidates remained highly ranked, indicating that the prioritized interaction landscape is largely stable over time (Supplementary Table 8). This observation further supports our decision to conservatively restrict the publication years included in the literature corpus—even at the cost of some reduction in recall—to enhance the reliability and interpretability of the resulting ARS-centered interaction landscape.

Future methodological developments may further enhance the interpretability and biological resolution of this framework. In particular, integrating transformer-based biomedical language models for semantic relation extraction and document-level contextual integration could enable more precise identification of interaction types and biological context. In addition, claim classification and structured knowledge extraction approaches may facilitate the transition from association-based networks toward mechanism-oriented ARS signaling networks, ultimately enabling deeper integration with multiomics data. Together, these developments may further strengthen literature-driven systems approaches for uncovering biologically meaningful targets in complex disease networks.

Key Points

- Prioritizing disease associations using text mining and graph centrality, mature concepts in systems biology but underutilized in target discovery.
- Highlighting the novel biology of human aminoacyl-tRNA synthetases as a case study, which is often overlooked outside of rare diseases or basic translation research.
- Demonstrating scalability to any protein family or therapeutic area with a clear validation path.

Acknowledgements

The authors thank all contributors and collaborators involved in this study.

Author contributions

J.C. conceived and designed the study, carried out all experiments and analyses, and wrote the manuscript. I.Y. provided biological and methodological advice throughout the project, performed validation experiments, and contributed to the sections Experimental validation of the network map and part of the DISCUSSION. Y.J., S.H. and E.K. developed the automation part of the target scoring system and T.A. supervised and guided the automation process. Overall, S.K. revised the manuscript and supervised the entire project. All authors approved the final version of the manuscript for submission.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Funding

This work was supported by the National Research Foundation of Korea, funded by the Korea government (Ministry of Science and ICT) (grant numbers NRF-2021R1A3B1076605, NRF-2021R1C1C10133,32, and RS-2026-25498878); the Korea Institute for Advancement of Technology, funded by the Ministry of Trade, Industry and Energy (grant number RS-2024-00418203); the Global Learning and Academic Research Institution for Master's, PhD students, and Postdoctoral

Disorders" in this picture. **(C) Experimental validation of the interaction with ARSs. (Left) ESR1–ARS interaction.** This figure presents western blot analysis of co-immunoprecipitation (Co-IP) experiments assessing interactions between GFP-tagged estrogen receptor 1 (ESR1) and various ARSs. ESR1 was immunoprecipitated using an anti-GFP antibody, and co-precipitated ARS proteins were detected with their respective antibodies. Among the four tested ARSs, MARS1 (methionyl-tRNA synthetase 1) was specifically co-precipitated with GFP-ESR1 but not with GFP alone, suggesting a potential direct or indirect interaction with ESR1. Input controls confirmed the expression of the indicated proteins. **(Right) APP–ARS interaction.** Two different forms of amyloid precursor protein (APP695 and APP751) were expressed as flag-tagged fusion proteins and subjected to Co-IP experiments with different ARSs and AIMPs (ARS-interacting multifunctional proteins). Among the five tested proteins, IARS1 (isoleucyl-tRNA synthetase 1) and AIMP1 (ARS-interacting multifunctional protein 1) were co-precipitated with both forms of flag-tagged APP, but not with the flag tag alone. Input controls confirmed the expression of the indicated proteins.

Researchers (LAMP) Program of the National Research Foundation of Korea, funded by the Ministry of Education (grant number RS-2024-00442483); and the Scale-up TIPS Program (grant number RS-2025-25467058) funded by the Ministry of SMEs and Startups.

Data availability

The datasets underlying this article will be shared on reasonable request to the corresponding author.

References

1. Lindsay MA. Target discovery. *Nat Rev Drug Discov* 2003;**2**:831–8. https://doi.org/10.1038/nrd1202
2. Butcher SP. Target discovery and validation in the post-genomic era. *Neurochem Res* 2003;**28**:367–71. https://doi.org/10.1023/A:1022349805831
3. Emmerich CH, Gamboa LM, Hofmann MCJ *et al.* Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat Rev Drug Discov* 2021;**20**:64–81. https://doi.org/10.1038/s41573-020-0087-3
4. Tabana Y, Babu D, Fahlman R *et al.* Target identification of small molecules: an overview of the current applications in drug discovery. *BMC Biotechnol* 2023;**23**:44. https://doi.org/10.1186/s12896-023-00815-4
5. Jia ZC, Yang X, Wu YK *et al.* The art of finding the right drug target: emerging methods and strategies. *Pharmacol Rev* 2024;**76**:896–914. https://doi.org/10.1124/pharmrev.123.001028
6. Schirle M, Bantscheff M, Kuster B. Mass spectrometry-based proteomics in preclinical drug discovery. *Chem Biol* 2012;**19**:72–84. https://doi.org/10.1016/j.chembiol.2012.01.002
7. Jones LH, Bunnage ME. Applications of chemogenomic library screening in drug discovery. *Nat Rev Drug Discov* 2017;**16**:285–96. https://doi.org/10.1038/nrd.2016.244
8. Moffat JG, Vincent F, Lee JA *et al.* Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat Rev Drug Discov* 2017;**16**:531–43. https://doi.org/10.1038/nrd.2017.111
9. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. *Nat Rev Mol Cell Biol* 2019;**20**:285–302. https://doi.org/10.1038/s41580-018-0094-y
10. Labib M, Kelley SO. Single-cell analysis targeting the proteome. *Nat Rev Chem* 2020;**4**:143–58. https://doi.org/10.1038/s41570-020-0162-7
11. Haley B, Roudnicky F. Functional genomics for cancer drug target discovery. *Cancer Cell* 2020;**38**:31–43. https://doi.org/10.1016/j.ccell.2020.04.006
12. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for large population-based studies. *Nat Rev Genet* 2021;**22**:19–37. https://doi.org/10.1038/s41576-020-0268-2
13. Hughes RE, Elliott RJR, Dawson JC *et al.* High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. *Cell Chem Biol* 2021;**28**:338–55. https://doi.org/10.1016/j.chembiol.2021.02.015
14. Meissner F, Geddes-McAlister J, Mann M *et al.* The emerging role of mass spectrometry-based proteomics in drug discovery. *Nat Rev Drug Discov* 2022;**21**:637–54. https://doi.org/10.1038/s41573-022-00409-3
15. Trajanoska K, Bherer C, Taliun D *et al.* From target discovery to clinical drug development with human genetics. *Nature* 2023;**620**:737–45. https://doi.org/10.1038/s41586-023-06388-8
16. Khodosevich K, Dragicevic K, Howes O. Drug targeting in psychiatric disorders - how to overcome the loss in translation? *Nat Rev Drug Discov* 2024;**23**:218–31. https://doi.org/10.1038/s41573-023-00847-7
17. Guo T, Steen JA, Mann M. Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature* 2025;**638**:901–11. https://doi.org/10.1038/s41586-025-08584-0
18. Peck D, Crawford ED, Ross KN *et al.* A method for high-throughput gene expression signature analysis. *Genome Biol* 2006;**7**:R61. https://doi.org/10.1186/gb-2006-7-7-r61
19. Yang Y, Adelstein SJ, Kassis AI. Target discovery from data mining approaches. *Drug Discov Today* 2012;**17**:S16–23. https://doi.org/10.1016/j.drudis.2011.12.006
20. Csermely P, Korcsmaros T, Kiss HJ *et al.* Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013;**138**:333–408. https://doi.org/10.1016/j.pharmthera.2013.01.016
21. Shamsi A, Khan MS, Yadav DK *et al.* Structure-based drug-development study against fibroblast growth factor receptor 2: molecular docking and molecular dynamics simulation approaches. *Sci Rep* 2024;**14**:19439. https://doi.org/10.1038/s41598-024-69850-1
22. Yan R, Zheng C, Qian S *et al.* The ZNF263/CPT1B axis regulates fatty acid beta-oxidation to affect cisplatin resistance in lung adenocarcinoma. *Pharmacogenomics J* 2024;**24**:33. https://doi.org/10.1038/s41397-024-00355-w
23. McNeil MR, Schimmel PR. Effect of transfer ribonucleic acid on the rate law and mechanism of the adenosine triphosphate–pyrophosphate isotope exchange reaction of an aminoacyl transfer ribonucleic acid synthetase. *Arch Biochem Biophys* 1972;**152**:175–9. https://doi.org/10.1016/0003-9861(72)90205-6
24. Ibba M, Soll D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* 2000;**69**:617–50. https://doi.org/10.1146/annurev.biochem.69.1.617
25. Alexander RW, Schimmel P. Domain-domain communication in aminoacyl-tRNA synthetases. *Prog Nucleic Acid Res Mol Biol* 2001;**69**:317–49. https://doi.org/10.1016/S0079-6603(01)69050-0
26. Yang XL, Skene RJ, McRee DE *et al.* Crystal structure of a human aminoacyl-tRNA synthetase cytokine. *Proc Natl Acad Sci U S A* 2002;**99**:15369–74. https://doi.org/10.1073/pnas.242611799
27. Park SG, Schimmel P, Kim S. Aminoacyl tRNA synthetases and their connections to disease. *Proc Natl Acad Sci U S A* 2008;**105**:11043–9. https://doi.org/10.1073/pnas.0802862105
28. Guo M, Schimmel P, Yang XL. Functional expansion of human tRNA synthetases achieved by structural inventions. *FEBS Lett* 2010;**584**:434–42. https://doi.org/10.1016/j.febslet.2009.11.064
29. Kim S, You S, Hwang D. Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nat Rev Cancer* 2011;**11**:708–18. https://doi.org/10.1038/nrc3124
30. Yao P, Fox PL. Aminoacyl-tRNA synthetases in medicine and disease. *EMBO Mol Med* 2013;**5**:332–43. https://doi.org/10.1002/emmm.201100626
31. Yao P, Poruri K, Martinis SA *et al.* Non-catalytic regulation of gene expression by aminoacyl-tRNA synthetases. *Top Curr Chem* 2014;**344**:167–87. https://doi.org/10.1007/128_2013_422
32. Lee EY, Lee HC, Kim HK *et al.* Infection-specific phosphorylation of glutamyl-prolyl tRNA synthetase induces antiviral immunity. *Nat Immunol* 2016;**17**:1252–62. https://doi.org/10.1038/ni.3542
33. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. *Nat Rev Mol Cell Biol* 2018;**19**:45–58. https://doi.org/10.1038/nrm.2017.77
34. Kanaji T, Vo MN, Kanaji S *et al.* Tyrosyl-tRNA synthetase stimulates thrombopoietin-independent hematopoiesis accelerating recovery

from thrombocytopenia. *Proc Natl Acad Sci U S A* 2018;**115**:E8228–35. https://doi.org/10.1073/pnas.1807000115
35. Kwon NH, Fox PL, Kim S. Aminoacyl-tRNA synthetases as therapeutic targets. *Nat Rev Drug Discov* 2019;**18**:629–50. https://doi.org/10.1038/s41573-019-0026-3
36. Yoon I, Nam M, Kim HK *et al.* Glucose-dependent control of leucine metabolism by leucyl-tRNA synthetase 1. *Science* 2020;**367**:205–10. https://doi.org/10.1126/science.aau2753
37. Suomalainen A, Battersby BJ. Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat Rev Mol Cell Biol* 2018;**19**:77–92. https://doi.org/10.1038/nrm.2017.66
38. Minton DR, Nam M, McLaughlin DJ *et al.* Serine catabolism by SHMT2 is required for proper mitochondrial translation initiation and maintenance of formylmethionyl-tRNAs. *Mol Cell* 2018;**69**:610–621.e5. https://doi.org/10.1016/j.molcel.2018.01.024
39. Boutoual R, Meseguer S, Villarroya M *et al.* Defects in the mitochondrial-tRNA modification enzymes MTO1 and GTPBP3 promote different metabolic reprogramming through a HIF-PPARgamma-UCP2-AMPK axis. *Sci Rep* 2018;**8**:1163. https://doi.org/10.1038/s41598-018-19587-5
40. Thandapani P, Kloetgen A, Witkowski MT *et al.* Valine tRNA levels and availability regulate complex I assembly in leukaemia. *Nature* 2022;**601**:428–33. https://doi.org/10.1038/s41586-021-04244-1
41. Sung Y, Yu YC, Han JM. Nutrient sensors and their crosstalk. *Exp Mol Med* 2023;**55**:1076–89. https://doi.org/10.1038/s12276-023-01006-z
42. Delaunay S, Helm M, Frye M. RNA modifications in physiology and disease: towards clinical applications. *Nat Rev Genet* 2024;**25**:104–22. https://doi.org/10.1038/s41576-023-00645-2
43. Yoon I, Kim U, Choi J *et al.* Disease association and therapeutic routes of aminoacyl-tRNA synthetases. *Trends Mol Med* 2024;**30**:89–105. https://doi.org/10.1016/j.molmed.2023.10.006
44. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003;**3**:993–1022.
45. Rosen-Zvi M, Griffiths T, Steyvers M *et al.* The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI 2004)*. pp. 487–494, 2004.
46. Mimno D, McCallum A. Topic models conditioned on arbitrary features with Dirichletmultinomial regression. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI 2008)*, 24, 411–418, 2008.
47. Chen J, Li H. Variable selection for sparse Dirichlet-multinomial regression with an application to microbiome data analysis. *Ann Appl Stat* 2013;**7**:7. https://doi.org/10.1214/12-AOAS592
48. National Library of Medicine (US). PubMed [Internet]. Bethesda (MD): National Library of Medicine (US); available from: https://pubmed.ncbi.nlm.nih.gov/ (accessed May 9, 2025).
49. Clarivate. Web of Science [Internet]. London (UK): Clarivate; available from: https://www.webofscience.com/ (accessed May 9, 2025).
50. National Library of Medicine (US). Nutritional and Metabolic Diseases (C18). In: *MeSH Browser [Internet].* Bethesda (MD): National Library of Medicine (US); available from: https://meshb.nlm.nih.gov/record/ui?ui=D009750
51. Tjong Kim Sang, EF, De Meulder F. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003* (pp. 142–147). 2003.
52. Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction. *J Mach Learn Res* 2003;**3**:1083–106.
53. Ratinov L, Roth D. Design challenges and misconceptions in named entity recognition. In: Stevenson S, Carreras X, (eds.), *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*; Boulder, Colorado: Association for Computational Linguistics. p. 147–155, 2009.
54. Mintz M, Bills S, Snow R *et al.* Distant supervision for relation extraction without labeled data. In: Su K-Y, Su J, Wiebe J, Li H, (eds.), *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore: Association for Computational Linguistics; 2009, pp. 1003–11.
55. Bavelas A. Communication patterns in task-oriented groups. *J Acoust Soc Am* 1950;**22**:725–30. https://doi.org/10.1121/1.1906679
56. Bonacich P. Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 1972;**2**:113–20. https://doi.org/10.1080/0022250X.1972.9989806
57. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry* 1977;**40**:35–41. https://doi.org/10.2307/3033543
58. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* 1998;**30**:107–17. https://doi.org/10.1016/S0169-7552(98)00110-X
59. Page L, Brin S, Motwani R *et al. The PageRank Citation Ranking: Bringing Order to the Web*, Stanford, CA: Stanford InfoLab; 1999, Technical Report No.: 1999–66.
60. Brandes U. A faster algorithm for betweenness centrality. *J Math Sociol* 2001;**25**:163–77. https://doi.org/10.1080/0022250X.2001.9990249
61. Freeman LC. *Centrality in Social Networks: Conceptual Clarification, Social Network: Critical Concepts in Sociology*. Scott J, (ed.), Londres: Routledge, 2002;**1**:215–39. https://doi.org/10.1016/0378-8733(78)90021-7.
62. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc Netw* 2010;**32**:245–51. https://doi.org/10.1016/j.socnet.2010.03.006
63. Gleich DF. PageRank beyond the web. *SIAM Rev* 2015;**57**:321–63. https://doi.org/10.1137/140976649
64. Friendly M, Sigal M, Harnanansingh D. The milestones project: a database for the history of data visualization. In: Kimball MA, Kostelnick C, (eds.), *Visible Numbers: Essays on the History of Statistical Graphics*, London (UK): Routledge; 2017, pp. 219–34.
65. National Center for Biotechnology Information (NCBI). *Gene database [Internet]*. Bethesda (MD): National Library of Medicine (US); available from: https://www.ncbi.nlm.nih.gov/gene/ (accessed May 9, 2025).
66. National Library of Medicine (US). Diseases [C]. In: *MeSH Browser [Internet]*. Bethesda (MD): National Library of Medicine (US); available from: https://meshb.nlm.nih.gov/treeView (accessed May 9, 2025).

Briefings in Bioinformatics, 2026, **27**, bbag195
https://doi.org/10.1093/bib/bbag195
Problem Solving Protocol