Sivanathan and Hu *BioData Mining* (2025) 18:32
https://doi.org/10.1186/s13040-025-00444-x

BioData Mining

RESEARCH Open Access

Learning the therapeutic targets of acute myeloid leukemia through multiscale human interactome network and community analysis

Suruthy Sivanathan[1*] and Ting Hu[2]

*Correspondence:
22ss59@queensu.ca

[1] School of Computing and the Department of Biomedical and Molecular Sciences, Queen's University, Goodwin Hall, Kingston K7L 2N8, Ontario, Canada
[2] School of Computing, Queen's University, Goodwin Hall, Kingston K7L 2N8, Ontario, Canada

Abstract

Acute myeloid leukemia (AML) is caused by proliferation of mutated myeloid progenitor cells. The standard chemotherapy regimen does not efficiently cause remission as there is a high relapse rate. Resistance acquired by leukemic stem cells is suggested to be one of the root causes of relapse. Therefore, there is an urgency to develop new drugs for therapy. Repurposing approved drugs for AML can provide a cost-friendly, time-efficient, and affordable alternative. The multiscale interactome network is a computational tool that can identify potential therapeutic candidates by comparing mechanisms of the drug and disease. Communities that could be potentially experimentally validated are detected in the multiscale interactome network using the algorithm CRank. The results are evaluated through literature search and Gene Ontology (GO) enrichment analysis. In this research, we identify therapeutic candidates for AML and their mechanisms from the interactome, and isolate prioritized communities that are dominant in the therapeutic mechanism that could potentially be used as a prompt for pre-clinical/translational research (e.g. bioinformatics, laboratory research) to focus on biological functions and mechanisms that are associated with the disease and drug. This method may allow for an efficient and accelerated discovery of potential candidates for AML, a rapidly progressing disease.

Keywords: Acute myeloid leukemia (AML), Drugs, Therapeutic targets, Human interactome, Networks, Community detection

Introduction

Acute myeloid leukemia (AML) originates from myeloid progenitor cells that have acquired mutations to proliferate and halt differentiation to allow the progressive accumulation of immature cells in the bone marrow [1, 2]. Standard therapy for AML is the combination of either daunorubicin or idarubicin with cytosine arabinoside [3]. The complete remission rate from this therapy is 60–85% for patients under 60 years and 40–60% for patients over 60 years [4]. However, there are many cases of relapse (adaptive resistance) and refractory disease (resistance to therapy occurring during primary treatment). Unfortunately, due to adaptive resistance, the 5-year overall survival rate is 40–50% for patients under 60 years and 15–20% for patients

over 60 years [5]. This detrimental effect caused by AML is suggested to stem from a subpopulation of resistant leukemia cells known as minimal residual disease (MRD). Leukemia stem cells located in the MRD are suggested to induce regrowth of AML in relapse [5]. Therefore, this indicates that standard therapy is not sufficient to combat AML and that there is a need for more therapeutic options.

Discovering new drugs to treat this genetically diverse disease through the drug development pipeline would lead to consumption of a lot of money (2.5 billion dollars USD) and time (13–15 years) [1, 2]. A solution to reduce cost and time is repurposing drugs that have already been approved. More specifically, prior knowledge of the safety, toxicity, and efficacy of approved drugs will reduce the cost and time to provide affordable therapeutic options to AML patients [1]. To discover these potential therapeutic candidates, the multiscale interactome network can be used.

The multiscale interactome network is a network of approved drugs-protein interactions, disease-protein interactions, protein-protein interactions, protein-biological function interactions, and biological function-biological function interactions. This network can be used to predict potential drug candidates for a particular disease based on the principle that the drug candidate could treat a disease if the mechanism of the drug affects the same functions that are affected by the disease's mechanism of action [6]. Hence, the candidate is predicted by comparing the visitation frequency of all the nodes (diffusion profile) in the context of the drug and disease. This is computed by a biased random walk that starts from either the drug or disease node. In other words, the multiscale interactome generates computed mechanisms of the drug and disease which are then compared for similarities to be determined as a possible candidate for repurposing.

Protein-protein interactions that occur in a cell are modular. The structural organization of a cell facilitates the environment and functions of proteins. Furthermore, proteins can work as a "community" to elicit a general function (e.g., transcription of DNA). These communities could also include a pathway of proteins that are involved in regulatory mechanisms (e.g., regulating a hormone). In addition, these modular communities are interconnected forming a complex interaction. A community detection method, Communities through Directed Affiliations (CoDA), and a community prioritization metric, CRank can be used to detect important communities in the multiscale interactome network [8, 8]. The novelty of CRank is that it ranks communities based on specific metrics, with only an input of the network and list of communities detected by a community detection method, to indicate the most promising communities that can be experimentally validated. By combining the results from the multiscale interactome network with CRank, this paper attempts to evaluate the resulting candidates and predict the functional communities that their mechanisms belong to. AML is a complex hematologic disease which can benefit from a computational analysis for comprehension of its complexity to determine potential therapeutic options. This method could potentially act as a prompt to identify molecular/community interactions to research in pre-clinical/translational studies (e.g. bioinformatics, laboratory research) for repurposing in AML.

Methods

Discovering potential drug candidates with the multiscale interactome

The multiscale interactome was constructed using 17,660 human proteins, 9798 biological functions, 1661 drugs, and 840 diseases by Ruiz et al. [6]. Constructing this large network required the arrangement of data on the individual interactions between pairs of these entities. More specifically, curating data for drug-protein, disease-protein, protein-protein, protein-biological function, and biological function-biological function interactions. A detailed instruction of how the interactions were obtained is provided in the paper by Ruiz et al. [6]. However, in general, various databases specific to the interactions were used to determine them. For example, Drug Repurposing Hub for drug-protein interactions, DisGeNet for disease-protein interactions, BioGRID for protein-protein interactions, Gene Ontology for protein-biological function interactions, and Gene Ontology-Biological Processes for hierarchical biological function-biological function interactions. A set of criteria was used to select pairs to develop a network of good quality. These criteria included ensuring that the pairs were experimentally validated, the pairs were related to Homo sapiens, the protein-protein interactions were direct physical interactions, and the biological functions were relevant to a drug's mode of action (e.g., affecting the regulatory mechanisms of a cell). Furthermore, drug-disease pairs were also formulated with various databases (e.g., Drug Indication Database). Similarly, a set of criteria was used to refine this list of pairs. For example, this involved the selection of approved drugs, the selection of diseases not caused by infections, and that they were present in the databases that drug-protein and disease-protein interactions are derived from. It is important to note that the drug-disease pairs are not incorporated into the multiscale interactome. Instead, the multiscale interactome learns to predict this association through the medium of diffusion profiles.

Diffusion profiles of a drug or disease are determined using biased random walks in the multiscale interactome. Therefore, a random walker walks on the nodes in the network by starting from a drug or a disease node which are connected to protein nodes. From the protein node, the random walker will continue to walk to either another protein or a biological function node. Finally, the random walker will continue to walk to either another biological function or protein node from the biological function node (e.g. drug/disease $\Leftrightarrow$ protein $\Leftrightarrow$ protein $\Leftrightarrow$ biological function $\Leftrightarrow$ lower-level biological function $\Leftrightarrow$ higher-level biological function). The random walker chooses the next step based on edge weights that denote the relative importance of each type of node ($W = w_{drug}$, $w_{disease}$, $w_{protein}$, $w_{biologicalfunction}$, $w_{higher-levelbiologicalfunction}$, $w_{lower-levelbiologicalfunction}$). More specifically, its next move is determined by the relative weights of its neighboring connected nodes. For example, when the random walker is situated at the protein node, the probability that the random walker will choose another protein node is $w_{protein}$ /$w_{biologicalfunction}$. As a result, the random walker will start at either the disease or drug node and choose to visit the next node or completely restart the walk based on the relative edge weights.

When the walk is completed after multiple rounds, the diffusion profile will encompass the frequencies at which each protein and biological function node was visited for a particular drug or a disease. The comparison of the diffusion profiles of a drug ($r^{(c)}$) and disease ($r^{(d)}$) is used to predict if the drug can treat a specific disease based on the

similarity of the diffusion profiles. A higher similarity will indicate that the mechanisms of the disease and drug are potentially the same. This increases the likelihood that the drug can be used for treatment by affecting the mechanisms which are involved in the disease. Ruiz et al. tuned the hyperparameters to optimize the predictability of the model by selecting the optimal hyperparameters determined from a sweep. The resulting optimal edge weights are drug = 3.2071696595616364, disease = 3.541889556309463, protein = 4.396695660380823, biological function = 6.583155399238509, lower-level biological function = 4.4863053901688685, and higher-level biological function = 2.09685000906964, and $\alpha = 0.8595436247434408$, where α is the probability of continuing the walk. These optimal hyperparameters provided by Ruiz et al. are used in our research to predict potential drug candidates for acute myeloid leukemia.

Combining baseline metrics with diffusion profiles for promising results

The diffusion profiles of all the drugs contained in the multiscale interactome and of AML are obtained. Observing the diffusion profiles with no additional measurements does not provide significant results. As mentioned by Ruiz et al., baseline metrics are taken into consideration in this study to measure the similarity of the diffusion profiles between the drugs and AML. Five baseline metrics used: L2 norm, L1 norm, Canberra distance, Cosine similarity, and Correlation distance. When the results from these metrics are ranked from having the highest similarity to the lowest similarity, the drugs further down the list become less significant as their diffusion profiles become dissimilar to the diffusion profile of AML. Therefore, to focus on the top results, the top 10, top 20, and top 50 candidates from each baseline metric are selected. Then, within each range, the drugs that appear in more than one baseline metric are selected for further downstream calculations and analysis. Therefore, a total of 71 drugs were selected. The purpose of this procedure is to take into account that the results of each baseline metric will not be identical, thus the presence of a drug in multiple metrics for a particular range (e.g., top 10) amplifies its significance in AML. It is important to note that top 30 and top 40 drugs were not considered because more drugs are included further down the list with top 50 drugs. This adds more variability and could change the result of which drugs hold the top 3 position.

Calculating the product of $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ to determine significant drug candidates

After selecting the drug candidates based on similar diffusion profiles, the potential effectiveness of the candidates in AML needs to be considered to compare and select drugs for further analysis. This is implemented by determining the visitation frequency of the candidates in the diffusion profile of AML ($\mathbf{r}_i^{(d)}$) and the visitation frequency of AML in the diffusion profile of the candidate drugs ($\mathbf{r}_j^{(c)}$). In other words, the frequency at which the particular drug node is visited in the diffusion profile of AML and the frequency at which the AML node is visited in the diffusion profile of a particular candidate drug is retrieved. These values are multiplied relative to the drug to generate a value that represents the contribution of these two factors in influencing the potential significance of the drug in the context of AML. Therefore, a drug with a high product value will suggest that there were many visits involved. As a result, the drug will hold much relevance in the treatment of AML. In this study, the products for all three ranges of selected drugs

were individually determined (top 10, top 20, and top 50). Based on the values of the product, the drugs are ranked in each range. Then, the top 3 drugs from each range are selected and observed for further downstream analysis. This selection is performed to check the consistency of the position of the drugs. If there is a drug which holds the top position in all three ranges, it can be inferred that it possesses much therapeutic power in the treatment of AML.

Connecting the interactome with community detection and community prioritization

The network of protein interactions and biological function interactions in the multiscale interactome might intrinsically contain communities that derive from the biological interactions. To identify these communities and provide a deeper analysis, a community detection method was used, Communities through Directed Affiliations (CoDA) [8]. This method detects cohesive and 2-mode communities in both directed and undirected networks by fitting a Directed Affiliation Network Model to an unlabeled directed network. A cohesive community refers to a community where the nodes within the community are densely linked to each other, while a 2-mode community refers to a community where the nodes within the community link to other nodes in a bipartite style instead of linking to each other. More specifically, the identification of cohesive communities in the multiscale interactome can encompass proteins with similar biological functions (e.g. transcription of DNA) and the identification of 2-mode communities can encompass proteins with similar regulatory functions (e.g. regulation of transcription) [8].

The algorithm for CoDA is written in C++, and the Ubuntu version of the executable was used to perform the community detection task. A text file containing the edgelist of the multiscale interactome was inputted with all the other parameters kept as default. The resulting output is two files where one contains the cohesive communities while the other contains the 2-mode communities. Although numerous communities are detected, the communities are predictions and not experimentally validated to be true. However, certain communities can be prioritized as statistically possible communities based on their structural integrity without the requirement of additional knowledge (e.g. gene expression data). This prioritization method is performed by CRank which uses four prioritization metrics: density, likelihood, boundary, and allegiance [8]. Density evaluates the strength of the connections within a community by observing the presence of densely nested communities. On the other hand, the likelihood metric evaluates the edges as a probability of being involved in a community. The boundary metric evaluates the quality of the boundary and how stable it is by observing how well the edges at the boundary distinguish the community and link to other parts of the network. Finally, the allegiance metric evaluates how prone the communities are to small alterations to the edges by determining the fraction of nodes in a community that have a higher probability of its edges with connections directing within the community than the probability of its edges with connections directing outside the community.

These four metrics are calculated for and rank each community by combining the scores determined in the original network and the slightly perturbed version of the original network in the prioritization metric formula. Next, the four metrics are combined to form the aggregate prioritization score by aggregating all four metrics based on the importance weight that was calculated to represent the contribution of each metric to

the aggregation score. An importance weight is necessary to calculate because the contribution of each metric varies for different networks as their importance in various networks varies. Therefore, the aggregation prioritization score is a weighted average of the ranked lists for density, likelihood, boundary, and allegiance. In general, the higher the aggregation score calculated by Crank, the more stable the detected communities are to be prioritized. The algorithm for CRank is written in C++ and the code ran with the input of the detected communities from CoDA (either cohesive or 2-mode) and the edge list of the multiscale interactome. The resulting output is a text file containing a list of the communities and their corresponding prioritization scores.

Results

Significant candidates were identified by calculating the product of $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ for drugs selected based on the five baseline metrics. To gain a better understanding of their mechanisms, the top 20 highly visited proteins and biological functions were identified for each candidate and AML. In addition, the genes that held importance in the context of the candidates and AML were determined by performing a calculation using the visitation frequencies of the proteins. The results from this calculation were used as a colour gradient in the constructed visual networks of each candidate and AML with the top 20 proteins and biological function nodes. Furthermore, to understand the broader functions of the mechanisms involved, the prioritized cohesive and 2-mode communities that were most frequently present in the top 20 nodes were identified (Fig. 1).

Results from calculating the product of $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ to determine significant drug candidates

Prior to the multiplication of $\mathbf{r}_i^{(d)}$ and $\mathbf{r}_j^{(c)}$, drugs are selected based on their frequency of presence in the top 10, top 20, and top 50 ranges of the five baseline metrics. This was performed by encoding a counter to record the total number of times a drug is present in the results of the five baseline metrics for each range. To ensure that there was no error, the counting was also performed manually on an excel sheet. Although this verified that the code performed correctly, it also showed that two drugs were represented by two nodes. More specifically, arsenic trioxide was represented as "arsenic-trioxide" and "DB01169", and megestrol acetate was represented as "megestrol-acetate" and "DB00351". To overcome this duplication, each candidate was represented by their DrugBank ID to maintain consistency with the other nodes.

After this refinement, the visitation frequencies of the selected drugs in the top10, top 20, and top 50 ranges in the AML diffusion profile were determined ($\mathbf{r}_i^{(d)}$). In the top 10 range, fostamatinib (1.8635e- 3), zinc (8.0001e- 4), and lasofoxifene (1.9427e-4) take the top 3 positions for $\mathbf{r}_i^{(d)}$. Similarly, these candidates take the same positions for the top 20 range for $\mathbf{r}_i^{(d)}$. In the top 50 range, fostamatinib, zinc, and L-glutamic acid (6.0801e- 4) take the top 3 positions. Subsequently, the visitation frequency of AML in the diffusion profile of the selected drugs in the top 10, top 20, and top 50 ranges were determined ($\mathbf{r}_j^{(c)}$). Compared to the calculation of $\mathbf{r}_i^{(d)}$, there was more variation in the candidates that were ranked at the top 3 positions for $\mathbf{r}_j^{(c)}$. The top 10 range consisted of lasofoxifene (2.3208e- 3), sunitinib (1.6211e- 3), and zinc (3.6870e-4) in the top 3 positions. In the top 20 range, lasofoxifene, vinblastine (2.1063e- 3),

and sunitinib take the top 3 positions. For the range of top 50, podophyllotoxin (2.9032e- 3), lasofoxifene, and vinblastine take the top positions.

Although there is a variation in the candidates that comprise the top 3 positions in $\mathbf{r}_i^{(d)}$ and $\mathbf{r}_j^{(c)}$ for all three ranges, it is the multiplication of these factors that will encompass their contribution and represent the relevance of the drug candidates in AML. The product values for each selected drug in the top 10 range reveal that fostamatinib (5.836730435016371e- 07), lasofoxifene (4.508462045960589e- 07), and zinc (2.949656116389859e- 07) acquire the top 3 positions. These drugs maintain their positions in the top 20 range. In the top 50 range, L-glutamic acid (7.622297565141564e- 07), fostamatinib, and lasofoxifene take the top 3 positions. When the top 3 candidates, derived from multiplication, for each range are compiled together, fostamatinib, lasofoxifene, zinc, and L-glutamic acid are observed to be the top candidates maintaining these positions. However, when identifying the product of $\mathbf{r}_i^{(d)}$ and $\mathbf{r}_j^{(c)}$ for all the drugs in the multiscale interactome without a baseline metrics-based selection process, the results are different. In this case, the top 3 drugs are gemtuzumab ozogamicin (1.5778308390603061e- 06), deslanoside (9.378050182994979e- 07), and L-glutamic acid (7.622297565141564e- 07).

Proteins and biological functions associated with the significant candidates

The benefit of the multiscale interactome is that its incorporation of the diffusion profile can be utilized to discover proteins and biological functions that hold importance in the mechanism of a drug or a disease. Therefore, the top 20 highly visited proteins and biological functions of fostamatinib, lasofoxifene, zinc, L-glutamic acid, and acute myeloid leukemia in their diffusion profiles were determined. A table summary of the discovered proteins and biological functions for the four candidates and AML can be found in Table 1.

Identifying important genes by simultaneously observing its frequency in the drug and disease

Although the top 20 highly visited proteins and biological functions were determined for the candidates, it does not consider how relevant the genes of the proteins are in the context of treating AML. To determine the proteins that may hold vital importance in the mechanism of treating AML, the treatment importance is calculated. This involves the multiplication of the visitation frequency of the protein node in the drug diffusion profile with its visitation frequency in the AML diffusion profile. Therefore, the treatment importance was calculated for all the proteins in the multiscale interactome for each drug.

In Table 1, proteins which have a high treatment importance are bolded (higher rank than 50). These proteins provide a hint towards the mode of action of a drug in the context of AML. Furthermore, Table 2 shows the treatment importance values of all the top AML proteins in each candidate and Table 3 shows the top 20 proteins with the highest treatment importance value in each candidate.

Table 1 The top 20 proteins and biological functions of the 4 candidates and AML

Fostamatinib	Lasofoxifene	
protein phosphorylation	**ESR1 (2)**	
protein autophosphorylation	**ESR2 (3)**	
intracellular signal transduction	**CNR2 (1)**	
positive regulation of transcription by RNA polymerase II	**GNA15 (4)**	
peptidyl-tyrosine phosphorylation	positive regulation of cytosolic calcium ion concentration	
positive regulation of transcription, DNA-templated	negative regulation of transcription by RNA polymerase II	
phosphorylation	positive regulation of transcription, DNA-templated	
HSP90 AA1 (15)	positive regulation of DNA binding transcription factor activity	
negative regulation of apoptotic process	positive regulation of transcription by RNA polymerase II	
peptidyl-serine phosphorylation	transcription by RNA polymerase II	
cellular protein modification process	regulation of cytosolic calcium ion concentration	
negative regulation of transcription by RNA polymerase II	cellular response to estradiol stimulus	
EGFR (12)	positive regulation of calcium ion transport into cytosol	
regulation of RNA splicing	negative regulation of gene expression	
NEK1 (337)	calcium ion transport into cytosol	
negative regulation of transcription, DNA-templated	negative regulation of DNA binding transcription factor activity	
positive regulation of gene expression	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G-protein coupled signaling pathway	
LRRK2 (18)	positive regulation of RNA polymerase II transcriptional preinitiation complex assembly	
NTRK1 (10)	negative regulation of transcription, DNA-templated	
APP (7)	negative regulation of production of miRNAs involved in gene silencing by miRNA	
Zinc	L-Glutamic Acid	Acute Myeloid Leukemia
CPN1 (979)	**GRIN1 (48)**	positive regulation of transcription by RNA polymerase II
FN1 (6)	GRIK2 (63)	positive regulation of transcription, DNA-templated
APOA1 (16)	GOT2 (70)	negative regulation of transcription, DNA-templated
C3 (40)	SLC1 A1 (55)	negative regulation of transcription by RNA polymerase II
IGHM (163)	GRIA2 (190)	MYC
APP (2)	SLC25 A22 (171)	RUNX3
CLU (12)	**GOT1 (29)**	STAT3
C5 (209)	GRIN2 A (151)	IDH1
TP53 (4)	EARS2 (156)	ANXA2
C8 A (533)	SLC1 A3 (51)	GATA1
complement activation	GLUD2 (145)	NPM1
A2M (47)	GLUD1 (100)	SPI1
APLP1 (116)	SLC1 A2 (131)	PSIP1
APOE (9)	calcium ion transmembrane import into cytosol	GATA2
KNG1 (95)	GLS (180)	KRAS
peptide cross-linking	response to ethanol	NSD1
C8G (179)	GRID2 (182)	GFI1
C8B (907)	**TAT (20)**	HSPB1
positive regulation of transcription by RNA polymerase II	FPGS (107)	EHMT2
HRNR (166)	GRM7 (166)	RUNX1

Table 2 Treatment importance of the top AML proteins in the context of the candidates

AML proteins	Fostamatinib	Lasofoxifene	Zinc	L-Glutamic acid
MYC	**8**	**6**	**5**	**2**
RUNX3	102	97	237	276
STAT3	**11**	**9**	**34**	**32**
IDH1	336	280	256	**6**
ANXA2	**23**	**15**	**21**	**30**
GATA1	**41**	**16**	63	56
NPM1	**16**	**10**	**15**	**12**
SPI1	134	**38**	132	139
PSIP1	183	**41**	156	164
GATA2	54	**33**	80	68
KRAS	**13**	**17**	**20**	**5**
NSD1	105	**32**	200	218
GFI1	182	80	190	153
HSPB1	**17**	**19**	**32**	**11**
EHMT2	136	**28**	124	123
RUNX1	62	52	**42**	67

The treatment importance values of the top acute myeloid leukemia proteins are shown in the table according to the respective top 4 candidates. The bolded numbers have a treatment importance rank that is one to 50

Visualizing the interactome of the drug candidates and AML

Interpretation of the mechanism of a drug from a list of highly visited biological functions and proteins is difficult to process. However, a visual representation of a network of proteins and biological functions from the candidate and AML provides a comprehensive format for interpretation. Cytoscape is a tool that facilitates the visualization of network interactions with the provided graph file (.graphml) [8]. Therefore, a subgraph was induced between the top 20 highly visited proteins and biological functions of each candidate and AML. The protein nodes were given a triangular shape, the biological function nodes were given a rectangular shape, and the drug and disease nodes were given a hexagonal shape. In addition, the visitation frequency of all the nodes in the AML diffusion profile and the candidate diffusion profile were added respectively to the nodes in the Cytoscape software. A colour gradient can be formed using the style tool in Cytoscape for either the AML diffusion profile or the candidate diffusion profile. This enables a visual representation of the visitation frequencies of the nodes in either diffusion profiles. By switching between the two diffusion profiles, the network can be formatted to represent if the nodes are interacting more with either the drug or the disease. The visual representation that results from this formatting for all four candidates can be seen in Figs. 2, 3, 4 and 5. Furthermore, the treatment importance value of each protein node is added to the software to determine the key proteins that are potentially involved in the therapeutic mechanism in AML (Figs. 2c, 3, 4 and 5c).

Detecting the prioritized communities involved in the top 20 proteins and biological functions

To observe the possible cellular processes and functions involved with the candidates, the frequency of the top 20 nodes in the highly prioritized communities were observed. This was performed by selecting the cohesive and 2-mode communities

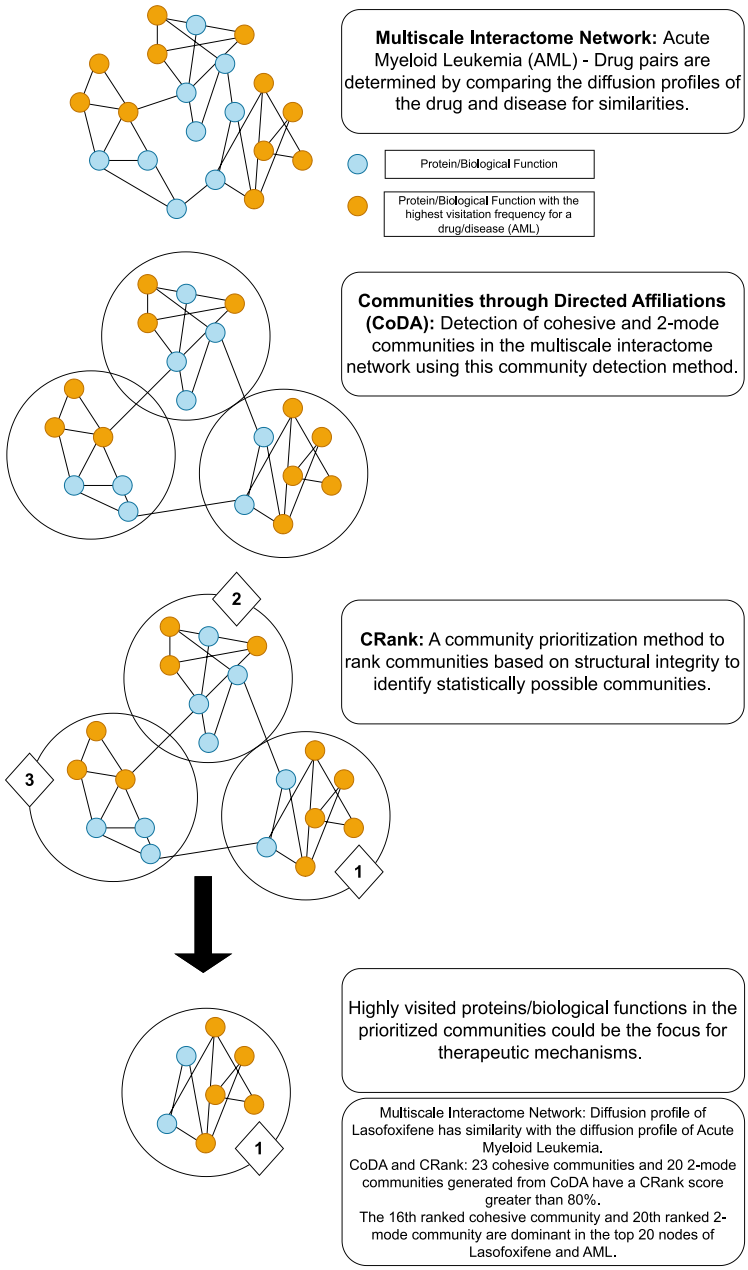

Fig. 1 Flowchart demonstrating a general overview of the method being assessed to determine potential therapeutic targets for acute myeloid leukemia

that had a CRank score greater than 80%. The results of this selection process led to 23 communities in cohesive communities and 20 communities in 2-mode communities. To efficiently evaluate the presence of the nodes within these communities, each community was saved as a text file. Furthermore, tables were created for cohesive and 2-mode communities which contained the community name, the CRank information, the text file name, and the nodes within the community (Supplementary Table 1 - Supplementary Table 2). However, the table for cohesive communities indicated that there were two communities with the name '7157'. This is the result from the names

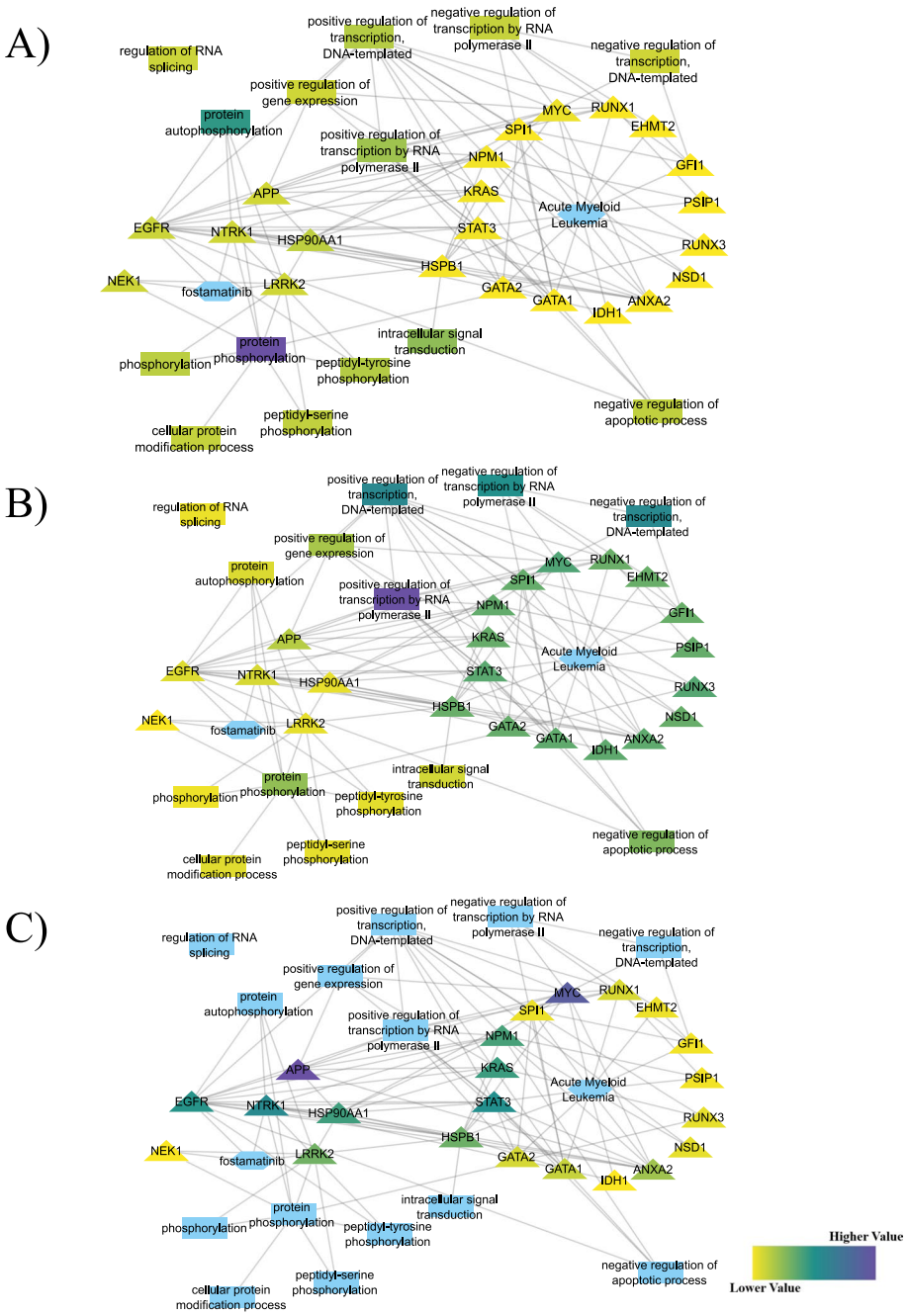

Fig. 2 Subgraphs for fostamatinib with colour gradient representation of diffusion profiles and treatment importance. **a** The colour gradient represents the diffusion profile in fostamatinib. **b** The colour gradient represents the diffusion profile in AML. **c** The colour gradient represents the treatment importance

being determined by the first node in the communities. To overcome this, after determining the location of the 20 nodes of each candidate and AML in the cohesive and 2-mode communities by looping through the text files, the cohesive community table was used to distinguish the two cohesive communities. This was executed by manually replacing the name to '715712' for the community ranked in twelfth place in each

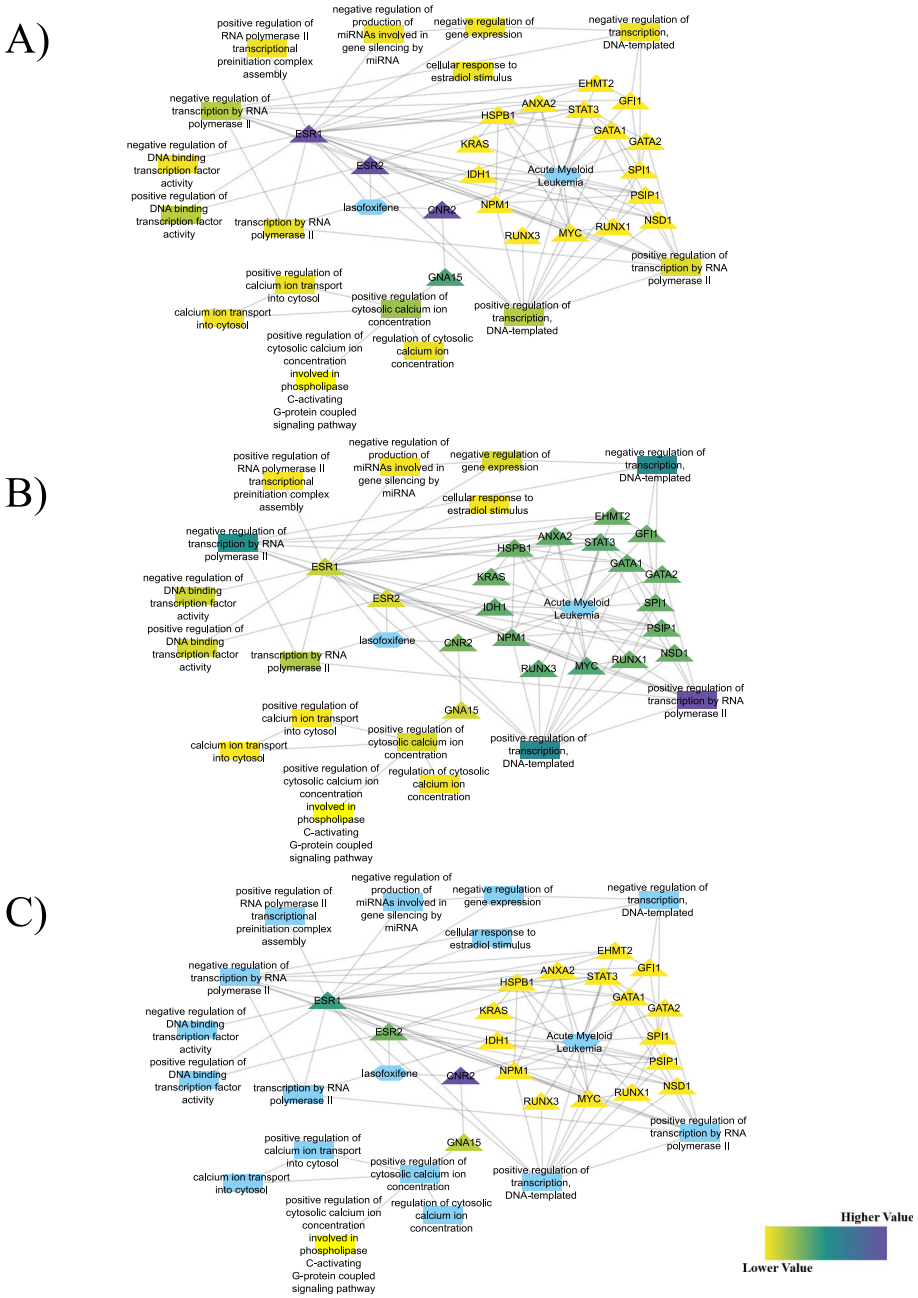

Fig. 3 Subgraphs for lasofoxifene with colour gradient representation of diffusion profiles and treatment importance. **a** The colour gradient represents the diffusion profile in lasofoxifene. **b** The colour gradient represents the diffusion profile in AML. **c** The colour gradient represents the treatment importance

candidate's location csv file if the nodes were present in the 'nodes' column for that community in the cohesive community table (Supplementary Table 1).

These cohesive community location files and the edited cohesive community table along with the files for 2-mode communities were used to determine the frequency of the top 20 nodes within the respective communities and were visualized as graphs (Figs. 6-7). It is important to note that each community was given a number that reflected their prioritization rank from CRank. Finally, the most frequent

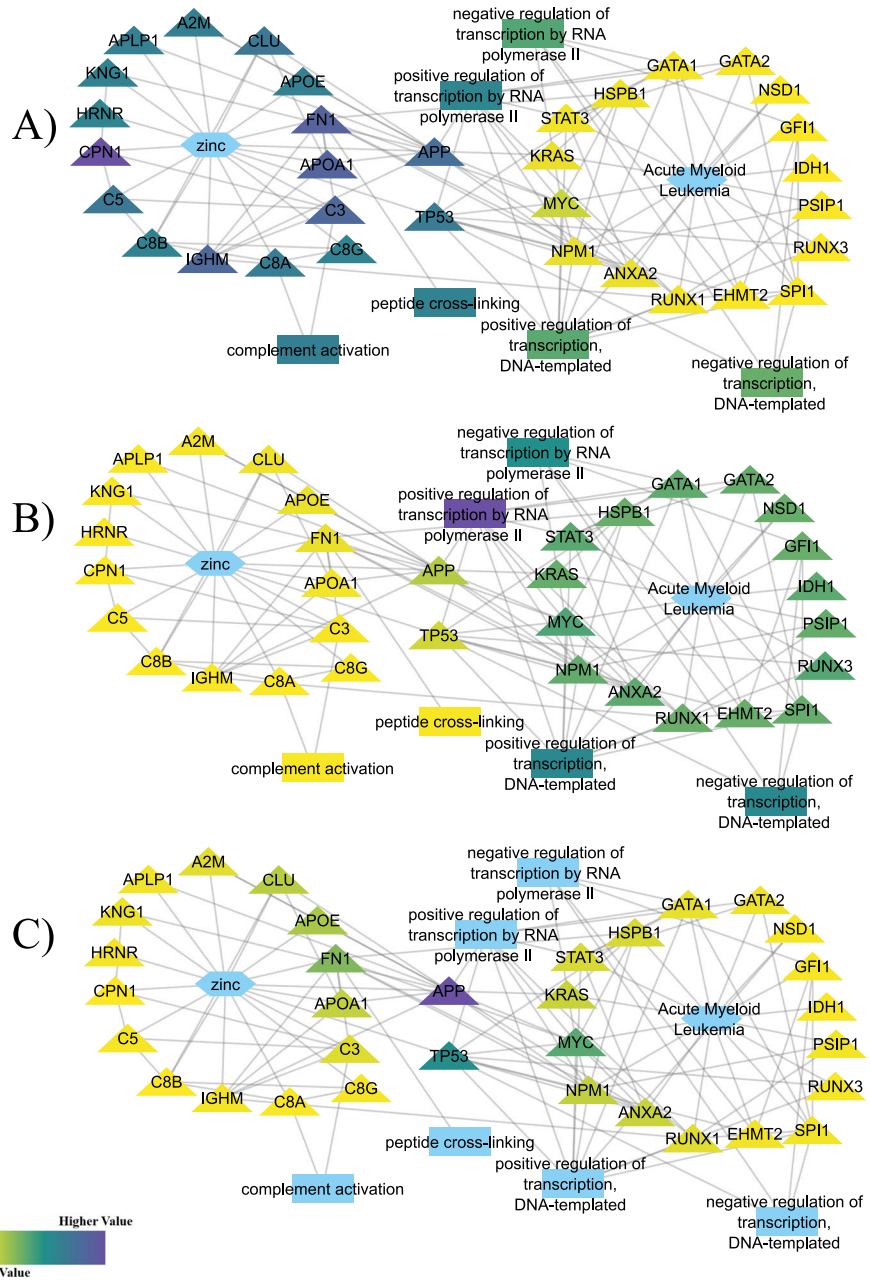

Fig. 4 Subgraphs for zinc with colour gradient representation of diffusion profiles and treatment importance. **a** The colour gradient represents the diffusion profile in zinc. **b** The colour gradient represents the diffusion profile in AML. **c** The colour gradient represents the treatment importance

communities from each graph of the candidates and AML for 2-mode and cohesive communities were searched in the "GO Enrichment Analysis" to determine the major biological processes involved in the community.

Discussion

The application of diffusion profiles from the multiscale interactome to baseline metrics and multiplication of specific visitation frequencies resulted in four significant candidates for acute myeloid leukemia (AML) and they are fostamatinib,

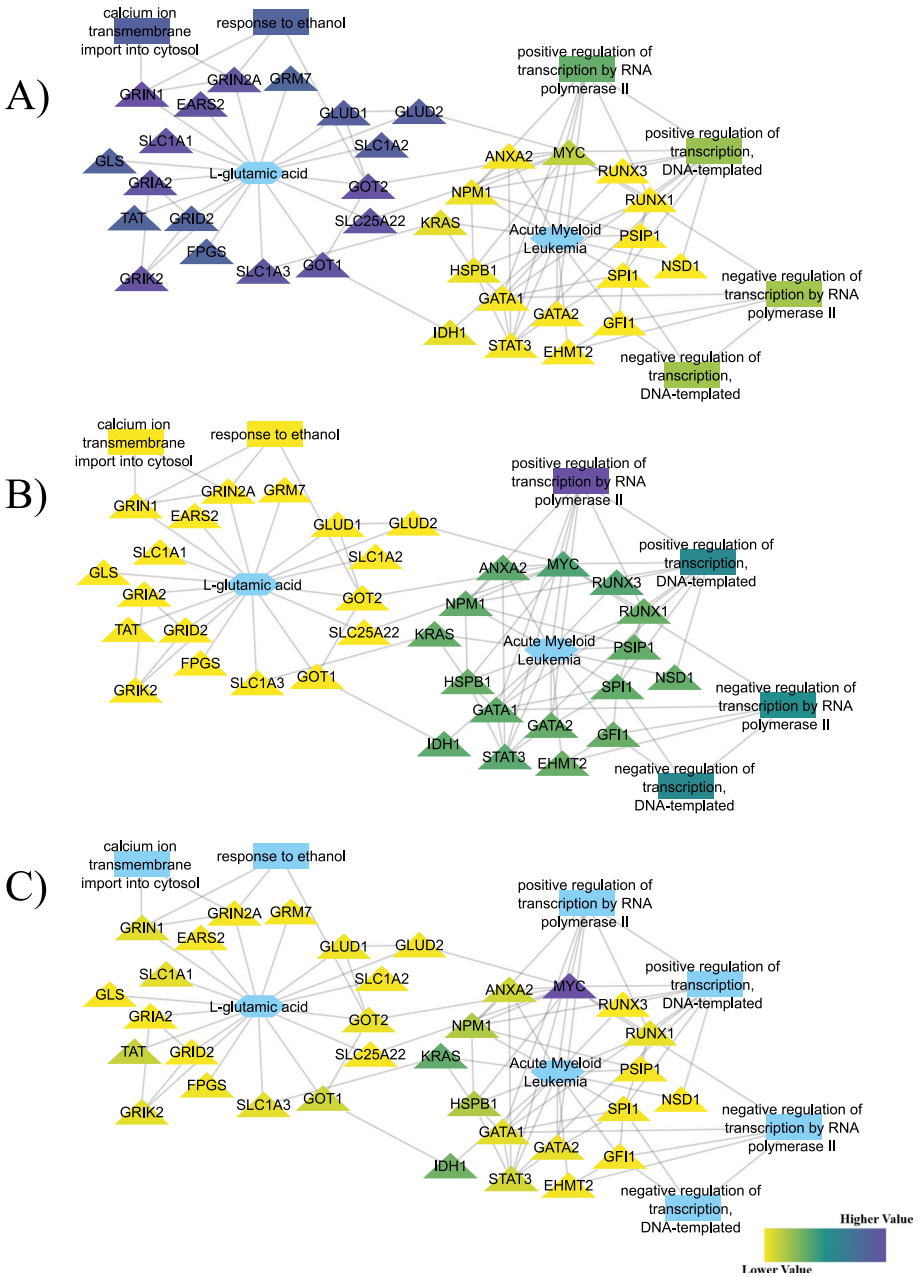

Fig. 5 Subgraphs for L-glutamic acid with colour gradient representation of diffusion profiles and treatment importance. **a** The colour gradient represents the diffusion profile in L-glutamic acid. **b** The colour gradient represents the diffusion profile in AML. **c** The colour gradient represents the treatment importance

lasofoxifene, zinc, and L-glutamic acid. However, when the baseline metrics were discarded in the selection process, the resulting top 3 candidates are gemtuzumab ozogamicin, deslanoside, and L-glutamic acid. Although gemtuzumab ozogamicin and deslanoside might have therapeutic potential in AML, there is higher confidence for the candidates that are selected with the baseline metrics because it focuses on

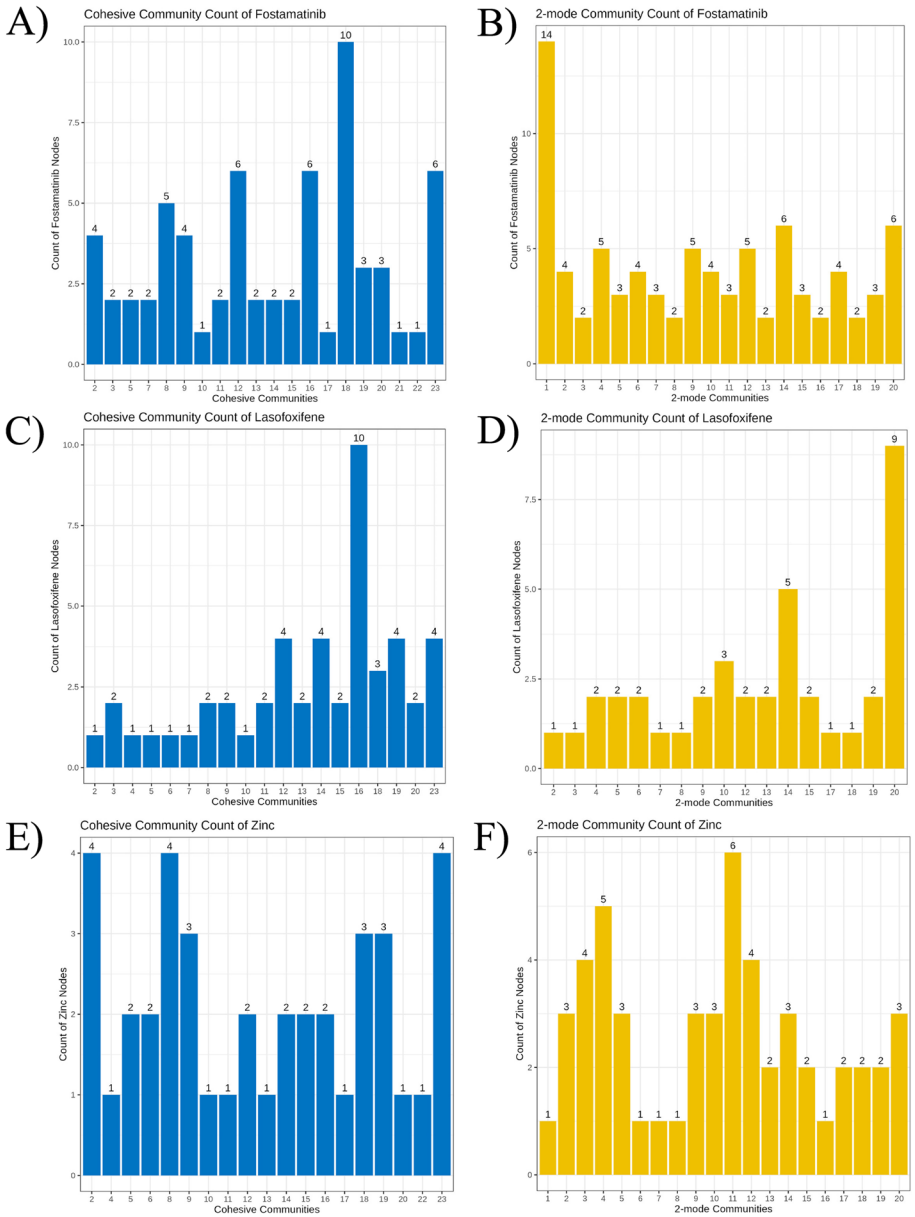

Fig. 6 Location of top 20 nodes in the cohesive and 2-mode communities. **a** Location of top nodes of fostamatinib in top cohesive communities. The 18^{th} ranked cohesive community has the highest frequency of nodes. **b** Location of top nodes of fostamatinib in top 2-mode communities. The 1^{st} ranked 2-mode community has the highest frequency of nodes. **c** Location of top nodes of lasofoxifene in top cohesive communities. The 16^{th} ranked cohesive community has the highest frequency of nodes. **d** Location of top nodes of lasofoxifene in top 2-mode communities. The 20^{th} ranked 2-mode community has the highest frequency of nodes. **e** Location of top nodes of zinc in top cohesive communities. The 2^{nd}, 8^{th}, and 23^{rd} ranked cohesive communities have the highest frequency of nodes. **f** Location of top nodes of zinc in 2-mode communities. The 11^{th} ranked 2-mode community has the highest frequency of nodes

the similarity between the drugs and AML diffusion profiles. A higher similarity indicates that similar proteins and biological functions were visited in AML and the candidate. This implies that mechanisms of the candidate and AML are similar, thus

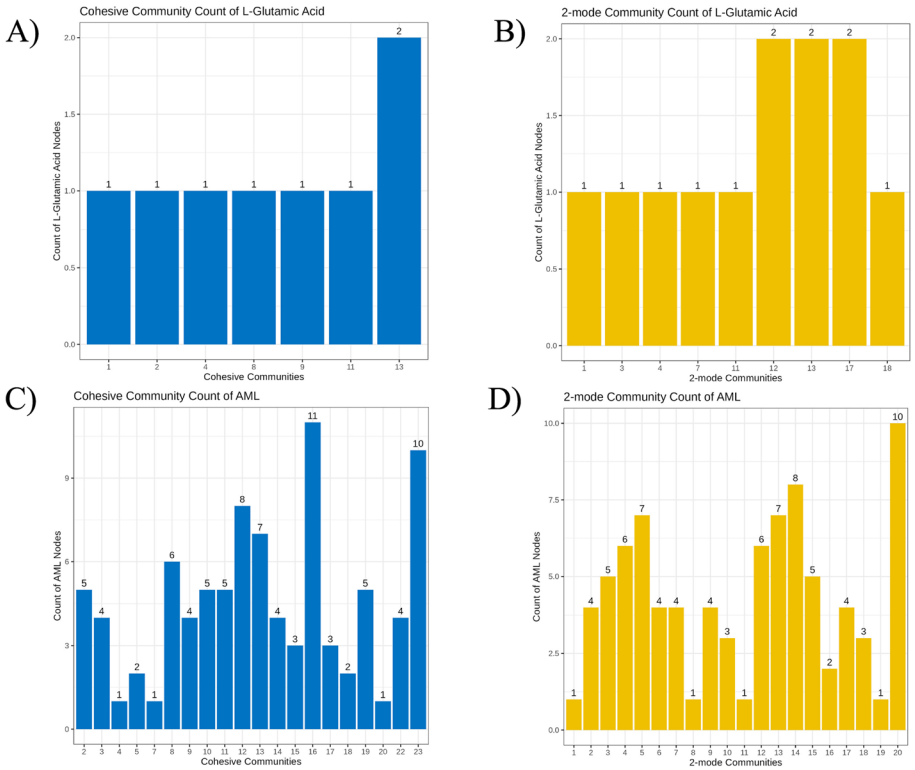

Fig. 7 Location of top 20 nodes in the cohesive and 2-mode communities (continued). **a** Location of top nodes of L-glutamic acid in top cohesive communities. The 13^{th} ranked cohesive community has the highest frequency of nodes. **b** Location of top nodes of L-glutamic acid in top 2-mode communities. The 12^{th}, 13^{th}, and 17^{th} ranked 2-mode communities have the highest frequency of nodes. **c** Location of top nodes of AML in top cohesive communities. The 16^{th} ranked cohesive community has the highest frequency of nodes. **d** Location of top nodes of AML in 2-mode communities. The 20^{th} ranked 2-mode community has the highest frequency of nodes

there is higher confidence that the candidates could impact AML by affecting the mechanisms of AML.

Fostamatinib

Fostamatinib achieves the first rank in the top 10 and top 20 selected range and second rank in the top 50 selected range for $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ multiplication. Therefore, the high rank demonstrates that fostamatinib is a strong therapeutic candidate for AML. The induced subgraph for fostamatinib and AML indicates that the drug interacts with tyrosine kinase signalling pathways ("protein phosphorylation", "protein autophosphorylation", "cellular protein modification process", "peptidyl-tyrosine phosphorylation") that are important for the immune system and AML for functions such as cell proliferation and cytokine signalling ("EGFR", "NTRK1", "STAT3", "KRAS", "MYC") (Fig. 2). A general examination of the subgraph illustrates that the top proteins and biological functions of fostamatinib encompass receptors and their subsequent protein signalling cascade ("EGFR", "NTRK1", "protein phosphorylation"), whereas for AML it encompasses the regulation of genes by transcription factors ("MYC", "STAT3", "RUNX1/3", "positive regulation of gene expression", "negative regulation of transcription by RNA polymerase II").

Adjusting the colourscale of the subgraph to reflect the treatment importance values helps focus on the nodes that are important in the context of fostamatinib and AML (Fig. 2c). This resulted in the identification of EGFR, NTRK1, HSP90AA1, APP, KRAS, STAT3, NPM1, and MYC as the important proteins. Fostamatinib forms edges with EGFR and NTRK1 which implies that it interacts with these receptors. According to the DrugBank database, fostamatinib acts as an inhibitor when interacting with EGFR and NTRK1 [8]. A study has identified crosstalk between EGFR and NTRK1 to cause transactivation to occur between these receptors in monocytes. In other words, the activation of the EGFR receptor resulted in the phosphorylation and activation of NTRK1 (TrkA) [8]. Furthermore, significant expression of NTRK1 was identified in myeloid leukemia cell lines and high EGFR expression was identified in its phosphorylated state in AML blasts in a subset of AML patients [8, 8]. According to the study, the inhibition of EGFR limited the activation of NTRK1 and the inhibition of NTRK1 limited the activation of EGFR [8].

Amyloid precursor protein (APP) is a membrane protein that can be released into active soluble fragments upon cleavage. The overexpression of APP in a subset of AML patients was determined to be an indicator for the formation of tumours of leukemic cells in regions other than the blood or bone marrow (Extramedullary leukemia). A study by Rocha et al. demonstrated that APP interacts with EGFR ligands, epidermal growth factor (EGF) and heparin-binding EGF-like growth factor (HB-EGF) [8, 8, 8]. In addition, they discovered that the overexpression of APP in HeLa and neuroblastoma cells, with the addition of EGF, significantly increased the amount of phosphorylated ERK. Along with the RAS/MEK/ERK1/2 pathway, PI3 K/AKT and JAK/STAT are the primary signalling cascades downstream of EGFR [8]. Jiang et al. observed the regulatory effects of microRNA - 144 on APP in an AML cell line [8]. The study showed that microRNA - 144 negatively regulated APP by reducing the expression of the APP protein. Furthermore, interfering with the expression of APP resulted in a decrease in cell migration and expression of phosphorylated ERK and c-Myc. Therefore, the researchers hypothesized that the APP/p-ERK/c-Myc/MMP- 2 pathway plays a role in cell migration. Myc is the top protein node for AML and represents the highest treatment importance in the subgraph in the context of AML and fostamatinib (Table 1, Fig. 2c). The activation of Myc is known to contribute to the development of leukemia from hematopoietic stem cells (leukemogenesis). More specifically, it is involved in the activation of genes that play a role in self-renewal of leukemia stem cells and cell growth [8].

In general, it can be inferred from the subgraph that the effect of fostamatinib toward AML could potentially function through pathways involved in cell proliferation, growth, and migration. Moreover, in table 3, which shows the 20 proteins with the highest treatment importance values for all the candidates, Kit and Jak2 are at the top 2 positions for fostamatinib and AML. Mutations in Kit and Jak2 have been found to induce proliferation by increasing tyrosine kinase activity [8]. Currently, fostamatinib is used to treat rheumatoid arthritis and immune thrombocytopenia by inhibiting the spleen tyrosine kinase (SYK). SYK is known to be present in hematopoietic cells and recent publications in 2020 and 2022 evaluated the therapeutic possibility of fostamatinib in AML cells which demonstrated the ability of SYK inhibitors to decrease proliferation of AML cells [8, 8, 8].

Table 3 Top 20 treatment importance ranked proteins in the candidates

Fostamatinib	Lasofoxifene	Zinc	L-glutamic Acid
JAK2	CNR2	S100 A8	GMPS
KIT	ESR1	APP	MYC
DAPK1	ESR2	ESR1	HNRNPL
MET	GNA15	TP53	APP
CSF1R	APP	MYC	KRAS
FLT3	MYC	FN1	IDH1
APP	RGS2	HDAC1	EGFR
MYC	LPAR1	MDM2	TRIM25
HNRNPL	STAT3	APOE	ADCY7
NTRK1	NPM1	GSN	NTRK1
STAT3	TP53	SH3GL1	HSPB1
EGFR	TRIM25	CLU	NPM1
KRAS	HNRNPL	EGFR	ELAVL1
TRIM25	EP300	PARP1	GRM1
HSP90 AA1	ANXA2	NPM1	SGK1
NPM1	GATA1	APOA1	PHB
HSPB1	KRAS	HNRNPL	CD44
LRRK2	CEBPA	JUP	ESR2
SRC	HSPB1	HDAC4	CEBPA
TP53	TGFB1	KRAS	TAT

The top 20 proteins with the highest treatment importance value in each candidate is depicted in this table

Lasofoxifene

Lasofoxifene has been intended for the treatment of postmenopausal osteoporosis and vaginal atrophy. Its mechanism of action involves the selective interaction with ERα (ESR1) and ERβ (ESR2) receptors [8]. Depending on the type of tissue, lasofoxifene can have an activating (agonist) or inhibitory effect (antagonist) on the estrogen receptor. For example, lasofoxifene acts as an agonist in the bone and an antagonist in the mammary gland and uterus [8]. Aside from the intended use of lasofoxifene, it also has prospects in treating ER-positive breast cancer as study has shown its ability to limit tumour growth and metastases [8]. Although there has been no preclinical studies related to the application of lasofoxifene in acute myeloid leukemia, a recent paper in 2018 investigated the effect of lasofoxifene and other modulators on B-cell and T-cell development (adaptive immunity) [8].

Lasofoxifene achieved second rank in the top 10 and top 20 selected range and third rank in the top 50 selected range for $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ multiplication. The subgraph depicts that the biological functions associated with lasofoxifene are related to regulating calcium homeostasis ("positive regulation of cytosolic calcium ion concentration", "positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G-protein coupled signaling pathway") and gene expression (positive regulation of RNA polymerase II transcriptional preinitiation complex assembly", "negative regulation of production of miRNAs involved in gene silencing by miRNA) (Fig. 3). In the context of treatment importance, the subgraph indicates ESR1, ESR2, CNR2, and GNA15 to be valuable in the context of the drug and AML (Fig. 3c).

ESR1 and ESR2 mediate the expression of specific genes by interacting with a DNA binding site known as the estrogen response element [8, 8, 8]. There is differential expression of these estrogen receptors in different tissue types and can have different effects. For example, ESR1 is predominantly expressed in liver, kidney, breast, and ovary, whereas ESR2 is predominantly expressed in the bone, prostate, ovary and central nervous system [8]. Furthermore, ESR1 is mainly involved in promoting growth such as in the uterine tissue and ESR2 is primarily involved in limiting proliferation such as in the bone marrow stem cells. There is emerging interest in investigating the effect of estrogen receptors in AML as disruption of its expression has been observed in this disease. In addition, men are more likely to be diagnosed with lymphomas or lymphocytic leukemia and this implies a potential influence of the sex-hormone, estrogen, in preventing the formation of these cancers [8].

In the subgraph, according to Table 3, the protein that ESR1 interacts with which has the highest treatment importance is Myc (Fig. 3). Study has shown that ESR1 is involved with mediating and increasing the transcription of genes regulated by c-Myc through interaction near the estrogen responsive promoter region with estradiol, a type of estrogen, and c-Myc [8, 8]. Rehman et al. have discovered that increased estrogen levels in breast cancer patients corresponded to elevated expression of c-Myc in leukocytes present in the peripheral breast cancer blood. Moreover, when comparing ESR1 expression levels in leukocytes of healthy subjects and patients, there was a significantly greater relative expression of ESR1 in the patients. Although the leukocytes are found in the peripheral blood of breast cancer, these findings can be informative in the context of AML [8]. However, there has also been evidence that a greater ratio value of ESR2/ESR1 increased the sensitivity of AML cells of patients to a drug, diosmetin, that targets ESR2 [8]. A study observed that diosmetin increased the level of intracellular reactive oxygen species in osteosarcoma cells which provides insight into how the drug promotes cell death [8]. According to Table 3, Stat3 has the highest treatment importance compared to the other proteins that interact with ESR2. Ning et al. have shown that diosmetin is able to cause apoptosis in osteosarcoma cells through the inhibition of the stat3/c-Myc signalling pathway which is involved in cell proliferation [8]. Therefore, the potential involvement of this pathway in the context of lasofoxifene and AML could be considered in future research.

Figure 3c depicts CNR2 (cannabinoid receptor 2), a G-protein coupled receptor, as having the highest treatment importance. In Table 3, it takes the first rank for treatment importance. Connected to CNR2 is GNA15 (guanine nucleotide binding protein, alpha 15) which is the alpa subunit ($G\alpha 15$) of the heterotrimeric G15 protein. This complex relays the signal of a G-protein coupled receptor to the activation of phospholipase C that results in an increase of the calcium ion in the cytosol [8, 8]. The biological functions connected to GNA15 in the subgraph are reflective of this function and are included in the top 20 proteins and biological functions of lasofoxifene (Fig. 3, Table 1). CNR2 is reported to be mainly located in immune cells and GNA15 has been identified to be expressed in hematopoietic stem cells [8, 8]. Lasofoxifene acts as an inverse agonist on the CNR2 to decrease the activity of the receptor [8]. Although current research does not focus on the interaction of CNR2

with lasofoxifene in the context of AML, Meritxell et al. demonstrated that CNR2 is enriched in multiple myeloid leukemia cell lines and AML blasts. Furthermore, they discovered that SR144528, an inverse agonist, was able to offset the migration of a myeloid cell line that was promoted by the activation of CNR2 by 2-arachidonoylglycerol [8]. Ultimately, lasofoxifene has multiple targets through which it can cause an effect on AML, thus, further preclinical and clinical research is required.

Zinc & L-glutamic acid

Zinc, a micronutrient, achieved the third rank in the top 10 and top 20 selected range of $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ multiplication, whereas, L-glutamic acid, an amino acid, achieved the first rank in the top 50 selected range and the third rank without the influence of baseline metrics for $\mathbf{r}_i^{(d)} * \mathbf{r}_j^{(c)}$ multiplication. The high similarity that is observed between the diffusion profile of these candidates and AML suggests the importance and the profound involvement of nutrients in the context of disease.

Zinc is involved in cell growth, differentiation, proliferation, and regulating apoptosis through enacting various functions in the cell by being incorporated in the catalytic region of proteins and stabilizing protein structures by interacting with zinc binding regions (zinc finger). Therefore, zinc partakes in a wide variety of roles by being associated with proteins such as transcription, regulation, repairing DNA, and chromatin structure [8, 8].

The impact of zinc on AML and other cancers has not been largely researched. Zinc deficiency has been identified in AML patients and thus signifies its importance. Costa et al. observed the influence of zinc on the DNA damage response pathway in AML cells [8]. They discovered that exhausting zinc increased cell death in normal lymphocytes and the supplementation of zinc caused cytotoxicity in AML cells. Furthermore, supplementing zinc with genotoxic agents caused a protective effect against the lymphocytes while this caused enhancement of cytotoxicity in AML cells. These results show the possibility of using zinc in a therapeutic strategy in AML where normal cells are protected. Nathani et al. exemplified this combination of a drug and zinc by researching the effect of treating colon cancer cells with zinc and berberine. Their research demonstrated that the growth of cells was inhibited with zinc and lower concentrations of berberine which suggested a synergistic effect of this combination therapy [42]. Clusterin (CLU), which is indicated in the top 20 list of proteins and biological functions of zinc and ranked in twelfth place for the top 20 treatment importance (Table 1, Fig. 4, Table 3), has been implicated to play a role in the effects of this combination therapy. More specifically, lower levels of CLU are present upon the treatment with this combination therapy. Since CLU has been implicated to be involved in cell survival, when lower levels were detected, it resulted in the apoptosis of the cells. As a result, it can be hypothesized from these studies that CLU might cause similar effects when treating AML cells with a combination of a drug and zinc.

TP53, a well-known tumour suppressor, attains the fourth rank in treatment importance and is present in the top 20 proteins and biological functions of zinc (Table 1, Fig. 4, Table 3). This transcription factor is associated with zinc because it contains a zinc finger. Furthermore, mutations in TP53 have been detected in AML, and mutations in this protein have been linked to various cancers since its functions are related to cell

cycle, apoptosis, and DNA repair [8]. The involvement of zinc with TP53 demonstrates the significance of this micronutrient in cancer.

In the subgraph for zinc (Fig. 4), there is a connection formed between zinc, APOE, and APP. In treatment importance, APP is ranked in the second place and APOE is ranked in the ninth place (Table 3). The APP protein is known to interact with zinc and the cleavage of APP results in the amyloid-β protein. A complex of amyloid-β protein, APOE, and zinc has been studied in the context of Alzheimer's disease [8, 42]. Oh et al. observed that in the presence of sufficient amounts of zinc, APOE and amyloid-β formed a complex, whereas when zinc was depleted, the aggregation of the complex was reduced [42]. Although this complex is specifically studied in the brain, the significant effects of similar interactions can be researched in the future for AML.

L-glutamic acid (glutamate) is found in the glutaminolysis pathway and can be converted from glutamine [8]. In AML, this pathway along with other metabolic pathways have been observed to be disrupted. Furthermore, AML is termed as a "glutamine addicted" cancer where it relies on its metabolic contribution of glutamine for proliferation and survival [42, 8]. In the subgraph, it is depicted that Myc has the highest treatment importance (Fig. 5c). Wise et al. discovered that this glutamine addiction can be caused by Myc by triggering mitochondrial glutaminolysis. They observed that Myc was able to shift the glucose dependent glioma cells to glutamine dependent cells by triggering the expression of glutamine transporters which was discovered to be mediated by the binding of Myc to the promoter elements of the transporter genes. This resulted in increased expression of the transporter mRNA [42]. In the subgraph, it can be observed that Myc is connected to proteins that partake in mitochondrial glutaminolysis (Fig. 5). Therefore, this suggests a possibility for a similar mechanism to occur in AML. In addition, KRAS, which displays a high treatment importance in the subgraph, has been found to increase the expression of certain amino acid transporters that have resulted in a higher uptake of glutamine in KRAS mutated colorectal cancer cells [8] (Fig. 5c). As a result, the possibility of similar mechanisms occurring in AML with this mutated enzyme can be further researched.

Glutamate can be converted into α-ketoglutarate by the enzyme glutamate dehydrogenase [8]. Furthermore, α-ketoglutarate can be generated from isocitrate with the enzyme isocitrate dehydrogenase (IDH) which is associated with citrate metabolism [42]. In the subgraph, IDH1 is depicted to have a high treatment importance (Fig. 5c). It has been observed that 8% percent of AML cases contain mutations in IDH1 and 12% percent of the cases contain mutations in IDH2 [8]. IDH1 is located in the cytoplasm and IDH2 is located in the mitochondria. These mutations cause the generation of the oncometabolite 2-hydroxyglutarate instead of generating α-ketoglutarate. The consequence of this alternate metabolite is that it can interfere with α-ketoglutarate dependent methylation of histones and DNA and inhibit the differentiation of immature hematopoietic progenitor cells to mature cells. Currently, there are IDH1 and IDH2 inhibitors being clinically improved and developed to unblock the myeloid differentiation. Promising results have been demonstrated in patients containing mutations in IDH. Furthermore, there is also prospect in using these inhibitors alongside allogeneic hematopoietic cell transplantation for AML [42, 42].

Biological processes in cohesive and 2-mode communities

Understanding the major biological processes involved in the most frequently occurring communities of the multiscale interactome network in the top 20 proteins and biological functions of AML and the candidates could potentially provide a starting point in preclinical research for discovery of mechanisms that are significant towards therapy. Therefore, the communities that consisted of the highest number of nodes from the top 20 biological functions and proteins for AML and the candidates were subjected to GO Enrichment Analysis. The resulting top 3 biological processes are summarized based on the resulting biological processes and subprocesses.

Cohesive community

Fostamatinib has 10 out of 20 of its top 20 nodes in community 18 (Fig. 6a). According to the GO Enrichment Analysis, this 18 th ranked community is associated with regulation of calcium ion transport ("negative regulation of calcium ion export across plasma membrane"), complex assembly in nucleus ("telomerase holoenzyme complex assembly"), and regulating cell development ("regulation of type B pancreatic cell development"). Lasofoxifene and AML have their top 20 nodes frequently located in community 16. Lasofoxifene has 10 out of its top 20 nodes located in community 16, whereas AML has 11 of its top 20 nodes located in this community (Figs. 6c, 7c). The enrichment analysis showed that this community is associated with regulation of chromosome/organelle organization ("negative regulation of chromosome condensation), cellular response to organonitrogen compound and chemical stimulus ("cellular response to anisomycin"), and anatomical structure morphogenesis and development ("fungiform papilla formation"). Zinc has three communities that are equally frequent in its top 20 nodes. Community 2, 8, and 23 each have 4 nodes present (Fig. 6e). Based on the enrichment analysis, community 2 is involved in regulation of calcium ion transport ("negative regulation of calcium ion export across plasma membrane"), miRNA/RNA transport and localization ("miRNA transport"), and regulating the cell cycle process ("regulation of G1 to G0 transition"). Community 8 is associated with microtubule-based process and organization ("regulation of centriole-centriole cohesion), chaperone-mediated autophagy and cellular catabolic/metabolic processes ("protein targeting to lysosome involved in chaperone-mediated autophagy"), and regulation of oxidoreductase/catalytic activity ("positive regulation of superoxide dismutase activity"). Finally, community 23 for zinc is involved in translesion/DNA repair synthesis ("error-free translesion synthesis"), nucleotide-excision repair/DNA incision ("nucleotide-excision repair, DNA incision, 5'-to lesion"), and DNA strand elongation/DNA replication ("lagging strand elongation"). L-glutamic acid has fewer nodes from the top 20 nodes located in the selected cohesive communities. Therefore, community 13, which is the most frequent community, has 2 out of the top 20 nodes of L-glutamic acid (Fig. 7a). The biological processes associated with this community are regulation of calcium ion transport ("negative regulation of calcium ion export across plasma membrane"), regulation of chromosome/organelle organization ("negative regulation of chromosome condensation"), and histone deacetylation/modification ("histone H4 deacetylation").

2-mode community

In comparison to cohesive communities, 2-mode communities contain nodes that make connections with nodes from other communities rather than with nodes in their communities. Therefore, these biological processes could reflect similar regulatory biological functions for the nodes within the same community. Fostamatinib has 14 out of its top 20 nodes in community 1 (Fig. 6b). According to the GO Enrichment Analysis, this 1 st ranked community is involved in complex assembly in nucleus ("telomerase holoenzyme complex assembly"), response to oxygen-containing compound/chemical ("response to hydrostatic pressure"), and anatomical structure development/morphogenesis ("trachea formation"). Lasofoxifene and AML have their nodes frequently located in community 20. Lasofoxifene has 9 out of its top 20 nodes located in community 20, whereas AML contains 10 out of its top 20 nodes located in this community (Figs. 6d, 7d). The results of the enrichment analysis show that community 20 is associated with anatomical structure morphogenesis and development ("fungiform papilla formation"), regulation of muscle adaptation and response to stimulus ("regulation of muscle atrophy"), and multicellular organismal process and tissue/epithelium development ("hair follicle placode formation"). Zinc has 6 out of its top 20 nodes located in community 11(Fig. 6f). This 11 th ranked community is involved in response to stimulus/chemical ("response to actinomycin D"), organic substance metabolic process ("tetrahydrobiopterin biosynthetic process"), and regulation of striated/cardiac muscle cell apoptotic process ("positive regulation of cardiac muscle cell apoptotic process"). As mentioned previously, L-glutamic acid has fewer of its top 20 nodes located in the selected 2-mode communities. Furthermore, L-glutamic acid has three equally frequent communities; it has 2 out of its top 20 nodes in community, 12, 13, and 17 (Fig. 7b). According to the enrichment analysis, community 12 is associated with positive regulation of DNA replication/DNA metabolic process ("positive regulation of mitochondrial DNA replication"), microtubule-based process and organization ("regulation of centriole-centriole cohesion), and chaperone-mediated autophagy and cellular catabolic/metabolic processes ("protein targeting to lysosome involved in chaperone-mediated autophagy"). Community 13 is involved in regulation of chromosome/organelle organization ("negative regulation of chromosome condensation"), "negative regulation of calcium ion export across plasma membrane", and response to chemical/stimulus ("response to fungicide"). Finally, community 17 for L-glutamic acid is associated with regulation of cellular component biogenesis/cellular process ("regulation of extracellular exosome assembly"), anatomical structure/tube development and multicellular organism development ("determination of digestive tract left/right asymmetry"), and endocytosis/vesicle-mediated transport ("G protein-coupled receptor internalization").

Conclusions

The multiscale interactome network identified fostamatinib and lasofoxifene as strong candidates for the treatment of acute myeloid leukemia. In addition, zinc and L-glutamic acid were identified as an important micronutrient and an amino acid, respectively, that has a significant contribution to this disease. Literature search of

the networks resulting from the diffusion profiles of the top 4 candidates showed promising results that the candidates' mechanism of actions could potentially play a role in the mechanism of AML. However, some proteins seemed to have more significant functions in other tissue types. As a result, fostamatinib and lasofoxifene can be further researched as a structural lead compound, and the therapeutic candidates that are ranked further down the list can be analyzed for therapeutic potential in AML. Furthermore, the accuracy of the predictions can be increased by adding tissue specificity (e.g., "Multiscale Interactome + Uberon + Cell Ontology") [6]. When the multiscale interactome network was subjected to community detection (CoDA) and community prioritization (CRank), top communities that could potentially be validated were identified for 2-mode and cohesive communities. Furthermore, the frequency of the top 20 proteins and biological functions of the 4 candidates in the top communities were detected. Communities with the highest count of nodes for each candidate were selected for GO Enrichment Analysis. Biological processes that were given to the communities seemed appropriate with respect to the candidates that represented the communities. This is because L-glutamic acid, a well-known candidate, had a general biological process of histone deacetylation/modification for the 13 th ranked cohesive community and positive regulation of DNA replication/DNA metabolic process in the mitochondria for the 12 th ranked 2-mode community. As mentioned previously, the mechanisms of L-glutamic acid are involved in metabolism of the mitochondria and modifications in the histones for DNA expression. However, in this study, only the top 20 nodes were considered, which is a significantly small number of nodes to identify significant communities. Hence, future studies can focus on using more than the top 100 proteins and biological functions to identify significant communities. In addition, along with a candidate's treatment importance data, the nodes and network within these communities could be observed to identify community specific pathways important in the context of the drug and disease. Alternatively, searching for "druggable communities" enriched in drug targets from the detected communities could be considered as another avenue for further research [8].

Repurposing various drugs for acute myeloid leukemia could occur more efficiently with the computational selection process of the multiscale interactome network and network analysis through analyzing prioritized community networks. However, the resistance observed in acute myeloid leukemia is caused by the development of diversified leukemic stem cells during therapy. As a result, as mentioned by Van et al., the resistant cells within the MRD should be researched at the single-cell level to enhance the knowledge of resistant mechanisms and provide better therapy to AML patients [5]. Zitnik et al. constructed a cell-cell similarity network from single-cell RNA sequencing data of neuronal cells and have detected prioritized communities using Crank [8]. Therefore, a future application would be to detect prioritized communities in cell-cell networks of AML which can then be correlated to markers and genes that are linked to diffusion profile/mechanism of drugs identified by the multiscale interactome. This can enable the discovery of targeted therapy for specific cell types.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13040-025-00444-x.

Supplementary Material 1.

Authors' information
Not applicable.

Acknowledgements
Not applicable.

Authors' contributions
S.S. and T.H. conceived and designed the research. S.S. performed the implementation and wrote the main manuscript. All authors reviewed the manuscript.

Funding
This research is supported by Natural Sciences and Engineering Research Council of Canada, Discovery Grant RGPIN-2023 - 03302 to T.H.

Data availability
Data that is generated for analysis in this manuscript can be accessed from https://github.com/SuruthyS/learning_the_targets_of_AML.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare no competing interests.

Received: 25 July 2024 Accepted: 3 April 2025
Published: 2 May 2025

References

1. Andresen V, Gjertsen BT. Drug repurposing for the treatment of acute myeloid leukemia. Front Med. 2017;4:211.
2. Wojcicki AV, Kadapakkam M, Frymoyer A, Lacayo N, Chae HD, Sakamoto KM. Repurposing drugs for acute myeloid leukemia: A worthy cause or a futile pursuit? Cancers. 2020;12(2):441.
3. Zhang J, Gu Y, Chen B. Mechanisms of drug resistance in acute myeloid leukemia. OncoTargets Ther. 2019;12:1937.
4. Niu J, Peng D, Liu L. Drug Resistance Mechanisms of Acute Myeloid Leukemia Stem Cells. Front Oncol. 2022;12:896426.
5. Van Gils N, Denkers F, Smit L. Escape from treatment; the different faces of leukemic stem cells and therapy resistance in acute myeloid leukemia. Front Oncol. 2021;11:659253.
6. Ruiz C, Zitnik M, Leskovec J. Identification of disease treatment mechanisms through the multiscale interactome. Nat Commun. 2021;12(1):1796.
7. Yang J, McAuley J, Leskovec J. Detecting cohesive and 2-mode communities indirected andundirected networks. In Proceedings of the 7th ACM international conference on Web search and data mining. 2014 (pp. 323-332).
8. Zitnik M, Sosič R, Leskovec J. Prioritizing network communities Nat Commun. 2018;9(1):2544.
9. 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.
10. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(suppl_1):D668–72.
11. El Zein N, D'Hondt S, Sariban E. Crosstalks between the receptors tyrosine kinase EGFR and TrkA and the GPCR, FPR, in human monocytes are essential for receptors-mediated cell activation. Cell Signal. 2010;22(10):1437–47.
12. Li Z, Beutel G, Rhein M, Meyer J, Koenecke C, Neumann T, et al. High-affinity neurotrophin receptors and ligands promote leukemogenesis. Blood J Am Soc Hematol. 2009;113(9):2028–37.
13. Mahmud H, Kornblau SM, Ter Elst A, Scherpen FJ, Qiu YH, Coombes KR, et al. Epidermal growth factor receptor is expressed and active in a subset of acute myeloid leukemia. J Hematol Oncol. 2016;9(1):1–3.
14. da Rocha JF, Bastos L, Domingues SC, Bento AR, Konietzko U, da Cruz E Silva OA, et al. APP binds to the EGFR ligands HB-EGF and EGF, acting synergistically with EGF to promote ERK signaling and neuritogenesis. Mol Neurobiol. 2021;58:668–88.

15. Jiang L, Yu G, Meng W, Wang Z, Meng F, Ma W. Overexpression of amyloid precursor protein in acute myeloid leukemia enhances extramedullary infiltration by MMP-2. Tumor Biol. 2013;34:629–36.
16. Jiang L, Meng W, Yu G, Yin C, Wang Z, Liao L, et al. MicroRNA-144 targets APP to regulate AML1/ETO+ leukemia cell migration via the p-ERK/c-Myc/MMP-2 pathway. Oncol Lett. 2019;18(2):2034–42.
17. Ahmadi SE, Rahimi S, Zarandi B, Chegeni R, Safa M. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. J Hematol Oncol. 2021;14:1–49.
18. Kahraman CY, Sincan G, Tatar A. Next-Generation Sequencing Panel Test in Myeloid Neoplasms and Evaluation with the Clinical Results. Eurasian J Med. 2022;54(2):181.
19. Cooper N, Ghanima W, Hill QA, Nicolson PL, Markovtsov V, Kessler C. Recent advances in understanding spleen tyrosine kinase (SYK) in human biology and disease, with a focus on fostamatinib. Platelets. 2023;34(1):2131751.
20. Brattås MK, Hemsing AL, Rye KP, Hatfield KJ, Reikvam H. Heterogeneity of Patient-Derived Acute Myeloid Leukemia Cells Subjected to SYK In Vitro Inhibition. Int J Mol Sci. 2022;23(23):14706.
21. Polak A, Bialopiotrowicz E, Krzymieniewska B, Wozniak J, Stojak M, Cybulska M, et al. SYK inhibition targets acute myeloid leukemia stem cells by blocking their oxidative metabolism. Cell Death Dis. 2020;11(11):956.
22. Lainé M, Fanning SW, Chang YF, Green B, Greene ME, Komm B, et al. Lasofoxifene as a potential treatment for therapy-resistant ER-positive metastatic breast cancer. Breast Cancer Res. 2021;23(1):54.
23. Swan VJ, Hamilton CJ, Jamal SA. Lasofoxifene in osteoporosis and its place in therapy. Adv Ther. 2010;27:917–32.
24. Andersson A, Törnqvist AE, Moverare-Skrtic S, Bernardi AI, Farman HH, Chambon P, et al. Roles of activating functions 1 and 2 of estrogen receptor α in lymphopoiesis. J Endocrinol. 2018;236(2):99–109.
25. Mal R, Magner A, David J, Datta J, Vallabhaneni M, Kassem M, et al. Estrogen receptor beta (ERβ): a ligand activated tumor suppressor. Front Oncol. 2020;10:587386.
26. Liu Z, Wong HM, Chen X, Lin J, Zhang S, Yan S, et al. MotifHub: Detection of trans-acting DNA motif group with probabilistic modeling algorithm. Comput Biol Med. 2024;168:107753.
27. Chen N, Yu J, Liu Z, Meng L, Li X, Wong KC. Discovering DNA shape motifs with multiple DNA shape features: generalization, methods, and validation. Nucleic Acids Res. 2024;52(8):4137–50.
28. Ma Z, Xia Y, Hu C, Yu M, Yi H. Quercetin promotes the survival of granulocytic myeloid-derived suppressor cells via the ESR2/STAT3 signaling pathway. Biomed Pharmacother. 2020;125:109922.
29. Roma A, Spagnuolo PA. Estrogen receptors alpha and beta in acute myeloid leukemia. Cancers. 2020;12(4):907.
30. Rehman N, Khan S, Manzoor S, Abubakar M, Sami R, Alharthy SA, et al. Estrogen Induces c-myc Transcription by Binding to Upstream ERE Element in Promoter. Appl Sci. 2022;12(14):6853.
31. Rota SG, Roma A, Dude I, Ma C, Stevens R, MacEachern J, et al. Estrogen receptor β is a novel target in acute myeloid leukemia. Mol Cancer Ther. 2017;16(11):2618–26.
32. Ning R, Chen G, Fang R, Zhang Y, Zhao W, Qian F. Diosmetin inhibits cell proliferation and promotes apoptosis through STAT3/c-Myc signaling pathway in human osteosarcoma cells. Biol Res. 2021;54.
33. Giannone F, Malpeli G, Lisi V, Grasso S, Shukla P, Ramarli D, et al. The puzzling uniqueness of the heterotrimeric G15 protein and its potential beyond hematopoiesis. J Mol Endocrinol. 2010;44(5):259–69.
34. Salter DM. 8 - Connective tissue responses to mechanical stress. In: Hochberg MC, Silman AJ, Smolen JS, Weinblatt ME, Weisman MH, editors. Rheumatology (Sixth Edition). 6th ed. Philadelphia: Mosby; 2015. pp. 61–64. https://doi.org/10.1016/B978-0-323-09138-1.00008-5.
35. Scott CE, Tang Y, Alt A, Burford NT, Gerritz SW, Ogawa LM, et al. Identification and biochemical analyses of selective CB2 agonists. Eur J Pharmacol. 2019;854:1–8.
36. Kumar P, Song ZH. CB2 cannabinoid receptor is a novel target for third-generation selective estrogen receptor modulators bazedoxifene and lasofoxifene. Biochem Biophys Res Commun. 2014;443(1):144–9.
37. Alberich Jorda M, Rayman N, Tas M, Verbakel SE, Battista N, Van Lom K, et al. The peripheral cannabinoid receptor Cb2, frequently expressed on AML blasts, either induces a neutrophilic differentiation block or confers abnormal migration properties in a ligand-dependent manner. Blood. 2004;104(2):526–34.
38. Leoni G, Rosato A, Perozzi G, Murgia C. Zinc proteome interaction network as a model to identify nutrient-affected pathways in human pathologies. Genes Nutr. 2014;9(6):1–9.
39. Costa MI, Lapa BS, Jorge J, Alves R, Carreira IM, Sarmento-Ribeiro AB, et al. Zinc prevents DNA damage in normal cells but shows genotoxic and cytotoxic effects in acute myeloid leukemia cells. Int J Mol Sci. 2022;23(5):2567.
40. Nathani S, Mishra R, Katiyar P, Sircar D, Roy P. Zinc acts synergistically with berberine forenhancing its efficacy as an anti-cancer agent by inducing clusterin-dependent apoptosis in HT-29 colorectal cancer cells. Biol Trace Elem Res. 2022;201:3755–3773.
41. Wild K, August A, Pietrzik CU, Kins S. Structure and synaptic function of metal binding to the amyloid precursor protein and its proteolytic fragments. Front Mol Neurosci. 2017;10:21.
42. Oh SB, Kim JA, Park S, Lee JY. Associative interactions among zinc, apolipoprotein E, and amyloid-β in the amyloid pathology. Int J Mol Sci. 2020;21(3):802.
43. Ubuka T. Subchapter 132A - Glutamic acid. In: Ando H, Ukena K, Nagata S, editors. Handbook of Hormones (Second Edition). second edition ed. San Diego: Academic Press; 2021. pp. 1063–1065. https://doi.org/10.1016/B978-0-12-820649-2.00296-5.
44. Bhingarkar A, Vangapandu HV, Rathod S, Hoshitsuki K, Fernandez CA. Amino acid metabolic vulnerabilities in acute and chronic myeloid leukemias. Front Oncol. 2021;11:694526.
45. Kandasamy P, Zlobec I, Nydegger DT, Pujol-Giménez J, Bhardwaj R, Shirasawa S, et al. Oncogenic KRAS mutations enhance amino acid uptake by colorectal cancer cells via the hippo signaling effector YAP1. Mol Oncol. 2021;15(10):2782–800.
46. Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, Pfeiffer HK, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc Natl Acad Sci. 2008;105(48):18782–7.
47. Wu N, Yang M, Gaur U, Xu H, Yao Y, Li D. Alpha-ketoglutarate: physiological functions and applications. Biomol Ther. 2016;24(1):1.

48. Kunadt D, Stasik S, Metzeler KH, Röllig C, Schliemann C, Greif PA, et al. Impact of IDH1 and IDH2 mutational subgroups in AML patients after allogeneic stem cell transplantation. J Hematol Oncol. 2022;15(1):1–15.
49. Issa GC, DiNardo CD. Acute myeloid leukemia with IDH1 and IDH2 mutations: 2021 treatment algorithm. Blood Cancer J. 2021;11(6):107.
50. Lu C, Ward PS, Kapoor GS, Rohle D, Turcan S, Abdel-Wahab O, et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. Nature. 2012;483(7390):474–8.
51. Hase T, Kikuchi K, Ghosh S, Kitano H, Tanaka H. Identification of drug-target modules in the human protein-protein interaction network. Artif Life Robot. 2014;19:406–13.

Publisher's Note