Genomics, Proteomics & Bioinformatics, 2025, **23(2)**, qzaf037
https://doi.org/10.1093/gpbjnl/qzaf037
Advance access publication: 29 April 2025
Method

DyNDG: Identifying Leukemia-related Genes Based on Time-series Dynamic Network by Integrating Differential Genes

Jin A (阿瑾)[1], Ju Xiang (项炬)[2,*], Xiangmao Meng (孟祥茂)[3], Yue Sheng (盛岳)[4,5], Hongling Peng (彭宏凌)[4,5], Min Li (李敏)[1,*]

[1]School of Computer Science and Engineering, Central South University, Changsha 410083, China
[2]School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China
[3]School of Computer Science & School of Cyberspace Science, Xiangtan University, Xiangtan 411105, China
[4]Department of Hematology, The Second Xiangya Hospital, Central South University, Changsha 410011, China
[5]Hunan Engineering Research Center of Cell Immunotherapy for Hematopoietic Malignancies, Changsha 410011, China

*Corresponding authors: limin@mail.csu.edu.cn (Li M), xiangju@csust.edu.cn (Xiang J).

Handling Editor: Edwin Wang

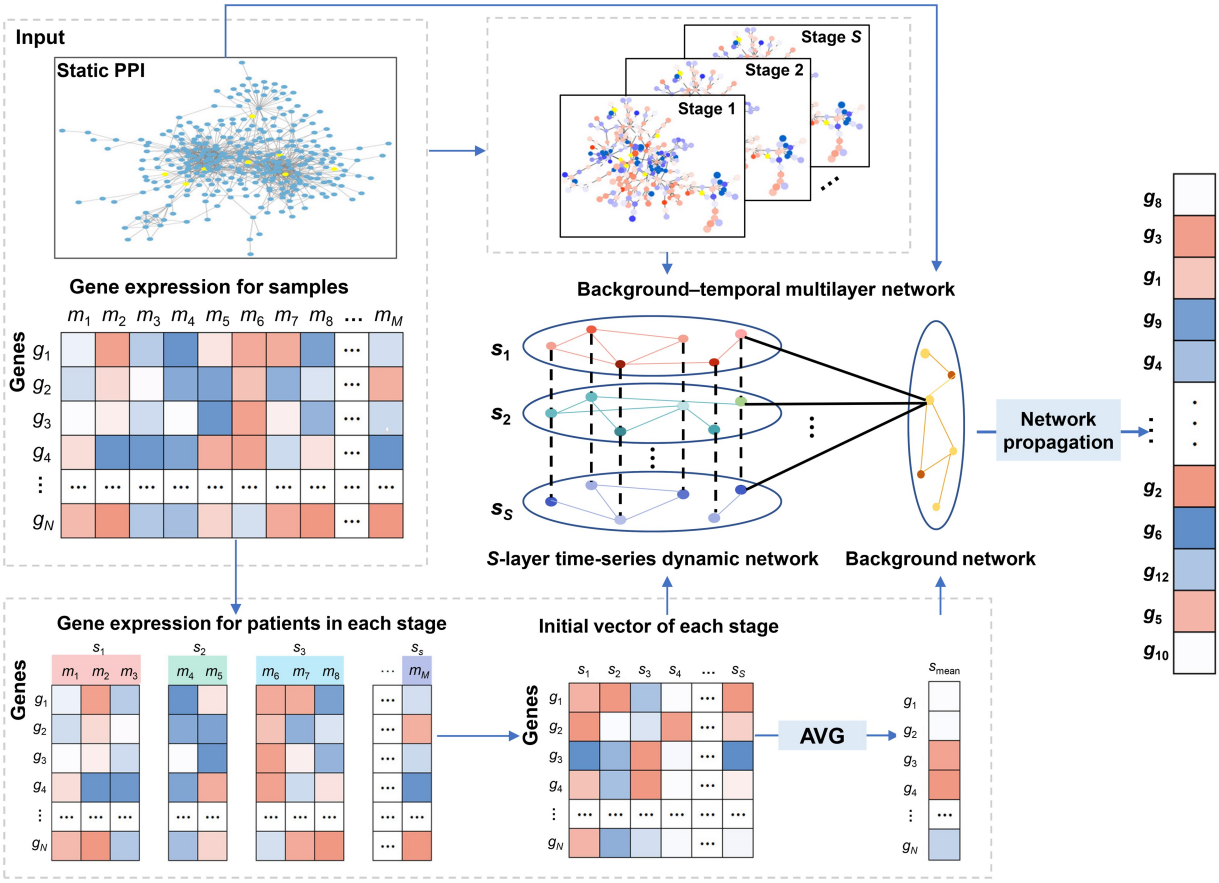
Abstract

Leukemia is a malignant disease characterized by progressive accumulation with high morbidity and mortality rates, and investigating its disease genes is crucial for understanding its etiology and pathogenesis. Network propagation methods have emerged and been widely employed in disease gene prediction, but most of them focus on static biological networks, which hinders their applicability and effectiveness in the study of progressive diseases. Moreover, there is currently a lack of special algorithms for the identification of leukemia disease genes. Here, we proposed a novel Dynamic Network-based model integrating Differentially expressed Genes (DyNDG) to identify leukemia-related genes. Initially, we constructed a time-series dynamic network to model the development trajectory of leukemia. Then, we built a background–temporal multilayer network by integrating both the dynamic network and the static background network, which was initialized with differentially expressed genes at each stage. To quantify the associations between genes and leukemia, we extended a random walk process to the background–temporal multilayer network. The results demonstrate that DyNDG achieves superior accuracy compared to several state-of-the-art methods. Moreover, after excluding housekeeping genes, DyNDG yields a set of promising candidate genes associated with leukemia progression or potential biomarkers, indicating the value of dynamic network information in identifying leukemia-related genes. The implementation of DyNDG is available at both https://ngdc.cncb.ac.cn/biocode/tool/BT7617 and https://github.com/CSUBioGroup/DyNDG.

Key words: Leukemia; Dynamic network; Random walk; Differentially expressed gene; Disease gene prediction.

Received: 6 January 2024; **Revised:** 22 February 2025; **Accepted:** 26 April 2025.

Graphical abstract



Introduction

Blood disorders often give rise to pathological conditions that extend beyond the blood and may lead to dysfunction in other vital organs, posing a threat to human health. People may suffer from various types of blood conditions and blood cancers. Common blood disorders include anemia (such as hemolytic anemia), bleeding disorders (such as hemophilia and blood clots), and blood cancers (such as leukemia, lymphoma, and myeloma). Leukemia, specifically, is a malignant blood disease characterized by the overproduction of immature or abnormal white blood cells, which eventually suppresses the production of normal blood cells and causes symptoms associated with cytopenia. According to the updated global cancer statistics 2024 released by the World Health Organization's International Agency for Research on Cancer (IARC), there were 9.7 million cancer deaths worldwide, including over 305,033 deaths from leukemia, while the number of leukemia deaths in China reached 50,100 [1]. The occurrence and development of leukemia are closely linked to the mutation and abnormal expression of genes that regulate cell growth in the bone marrow [2], although genetic, behavioral, and environmental factors collectively contribute to leukemia [3]. Therefore, investigating the causative genes of leukemia holds significant importance to uncover its pathogenesis. Identifying the causative genes of leukemia is an extremely challenging task due to the pleiotropy of genes, the genetic heterogeneity of diseases, and the limited number of study subjects [4–6].

With the continuous explosion of biological data, computational methods for predicting potential pathogenic genes have emerged, playing a significant role in disease research, prevention, and detection [7]. The functions of biomolecular components in cells are often interdependent rather than acting independently. Diseases typically arise from disruptions in a complex biomolecular network caused by genetic variants, pathogens, and epigenetic changes, rather than solely from abnormal expression of individual genes [8]. Therefore, numerous classical disease–gene prediction methods based on biomolecular networks have been proposed [9–16], among which network propagation has gained popularity due to its remarkable performance [17–21].

However, most traditional network-based methods are based on static biomolecular network structures. In contrast, cellular systems are highly dynamic and respond dynamically to external changes in different time points, spaces, and conditions [22]. Thus, the relationships between biomolecules are constantly changing in response to time and conditions, and the molecular mechanisms underlying disease onset and progression are closely related to this dynamic nature. Traditional static networks lose the dynamic information due to their highly averaged and idealized structures, hindering the further advancement of network-based methods [23,24]. Some researchers have already shifted their attention from static biological networks to dynamic biological networks [25]. The key challenge in constructing dynamic biological

networks is how to determine the dynamic features of biomolecule expression at different time points. Initially, de Lichtenberg et al. [26] proposed that persistently expressed proteins are not dynamic, while cyclically expressed proteins are only expressed at their highest levels in the expression cycle. Hegde et al. [27] suggested using the average expression value of each region in the network as the threshold for judging which proteins are actually expressed in that region. Tang et al. [28] discovered that periodically expressed genes often exhibit peak expression levels greater than a fixed constant by studying the dynamic protein interaction network in yeast cells. In contrast to the former fixed threshold method, Zhang et al. [29] introduced a *k*-sigma method based on the 3-sigma rule by designing an activity threshold for each gene based on its dynamic expression. The development of dynamic protein network construction methods, which simulate the operational rules of real biological systems, effectively overcomes the limitations of static protein network-based analysis methods and plays a significant role in protein complex identification [30,31], protein function prediction [32,33], and biomarker identification [34,35]. Numerous studies have demonstrated that dynamic protein networks can yield superior results in various related issues since they can better reflect the dynamic properties of biological processes over time and in response to external environments [36,37]. It is necessary to simulate the dynamic changes of biomolecular networks during disease progression to reveal the associations between diseases and genes.

Therefore, we proposed the Dynamic Network-based model integrating Differentially Expressed Genes (DyNDG) for identifying leukemia-related genes. We took three common leukemias: chronic myeloid leukemia (CML), chronic lymphocytic leukemia (CLL), and acute myeloid leukemia (AML) as research subjects. For each leukemia, DyNDG first generates a time-series dynamic network using expression data over stages of disease development, and then constructs a background–temporal multilayer network by integrating both the static network structure information and the dynamic information of leukemia development. Moreover, initialized by differentially expressed genes (DEGs), DyNDG extends a random walk process into the multilayer network to extract scores of leukemia-related genes. Results on the three types of leukemia and three control sets demonstrate that considering the time-series dynamics of leukemia development through DyNDG significantly enhances the accuracy of prioritizing leukemia-related genes compared to popular methods. Directly using predicted leukemia-related genes as drug targets may lead to interference and toxic side effects on normal cells. In order to minimize and avoid this possible harm, we selected higher-ranked genes from the predicted candidate gene list, excluded housekeeping genes from them, and yielded a set of promising candidate genes associated with leukemia. Through the integration of multiple analysis methods, we aimed to further understand the functions and regulatory mechanisms of these genes in the development of leukemia, providing valuable references and assistance for researchers and doctors.

Method

Datasets

Time-series expression data for leukemia

Gene Expression Omnibus (GEO) database is a public functional genomics data repository, which includes different gene expression datasets under different designs and conditions. Time-series expression data for CML, CLL, and AML were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (GEO: GSE47927, GSE2403, and GSE122917, respectively). We preprocessed the data in the following three steps: (1) mapping of gene IDs and gene symbols; (2) filtering genes encoding proteins; and (3) cleaning the data. Subsequently, the expression data were subjected to pathological analysis to classify the stages of leukemia development.

The development of leukemia is a multi-factor and multi-step cancerous process. Leukemia is generally categorized into two types: acute leukemia and chronic leukemia, based on the differentiation and maturation status of leukemic cells and the natural progression of the disease [38]. AML, as a typical representative of acute leukemia, progresses rapidly and its disease stages are temporarily described as untreated, in remission, refractory, or recurrent [39], because there is currently no standard disease staging system for it. Chronic leukemia develops slowly and has a natural course of several years. In particular, CLL follows the Rai system, which classifies CLL into "low-risk group (stage 0), intermediate-risk group (stages I and II), and high-risk group (stages III and IV)" [40]. CML is categorized into three disease phases: chronic phase (CP), accelerated phase (AP), and blast phase (BP) [41]. **Table 1** provides detailed information on the time-series staging sample statistics for GSE47927, GSE2403, and GSE122917, based on their corresponding staging systems.

Biological networks

STRING is a database that covers the largest number of species and contains the most extensive information on protein–protein interactions (PPIs). It collects and integrates known and predicted protein–protein association data from a variety of sources, including automated text mining of scientific literature, computational interaction predictions based on coexpression analyses, interaction experiments, and known complexes/pathways from curated sources [42]. HumanNet is an integrated human gene network database for disease research. It was constructed by incorporating new data types, expanding data sources, and utilizing improved network inference algorithms [43]. Due to the rich PPIs collected in STRING (https://cn.string-db.org/) and the focus of HumanNet (https://staging2.inetbio.org/humannetv3/) on disease research, we chose to use data from the two databases to construct the static PPI networks, respectively. Further details regarding the static PPI networks can be found in Table S1.

Table 1 Details of time-series staging sample statistics

Disease	Data	Period	Symbol	No. of samples
CML	GSE47927	Normal	Normal	15
		Chronic phase	CP	24
		Accelerated phase	AP	18
		Blast phase	BP	10
CLL	GSE2403	Low-risk	Rai 0	8
		Intermediate-risk	Rai I and Rai II	7
		High-risk	Rai III and Rai IV	6
AML	GSE122917	Normal	0	3
		Untreated	1	3
		Recurrent	2	3

Note: CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; AML, acute myeloid leukemia.

Disease genes for leukemia

Known leukemia-related genes were obtained from the MalaCards database (https://www.malacards.org/) (MCID: LKM071, LKM063, and LKM061) [44], which is an integrated database of human diseases and their annotations. There are 527 AML-related genes, 204 CML-related genes, and 339 CLL-related genes collected from the MalaCards database. We filtered these genes using protein-coding genes curated from the Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) database (https://housekeeping.unicamp.br/?download) [45].

Framework of DyNDG

Here, we proposed a dynamic network-based model, DyNDG, for predicting leukemia-related genes. DyNDG involves constructing the time-series dynamic biological networks and the background–temporal multilayer biological network, followed by network propagation on the multilayer biological network (**Figure 1**). The original inputs consist of a gene expression matrix and a static PPI network (Figure 1A). After constructing the time-series dynamic network (Figure 1B), the S-layer time-series dynamic network and the background network are combined to form the background–temporal multilayer biological network (Figure 1C). The initial probability vector of genes in the multilayer network is derived through differential gene expression analysis (see Figure 1D and the section "Network propagation on background–temporal multilayer biological network" for details). Subsequently, a network propagation process on the background–temporal multilayer biological network, which is initialized by DEGs, is performed to generate the predictive scores for leukemia-related genes (Figure 1E).

Construction of time-series dynamic networks

The disease progression of leukemia involves multiple stages, and the biological networks associated with leukemia undergo dynamic changes throughout these stages. The gene expression data capturing the progression of leukemia were obtained by investigating patient populations at different stages of the disease, which can provide valuable insights into the dynamic gene changes that occur as leukemia progresses. The biological network constructed for each stage

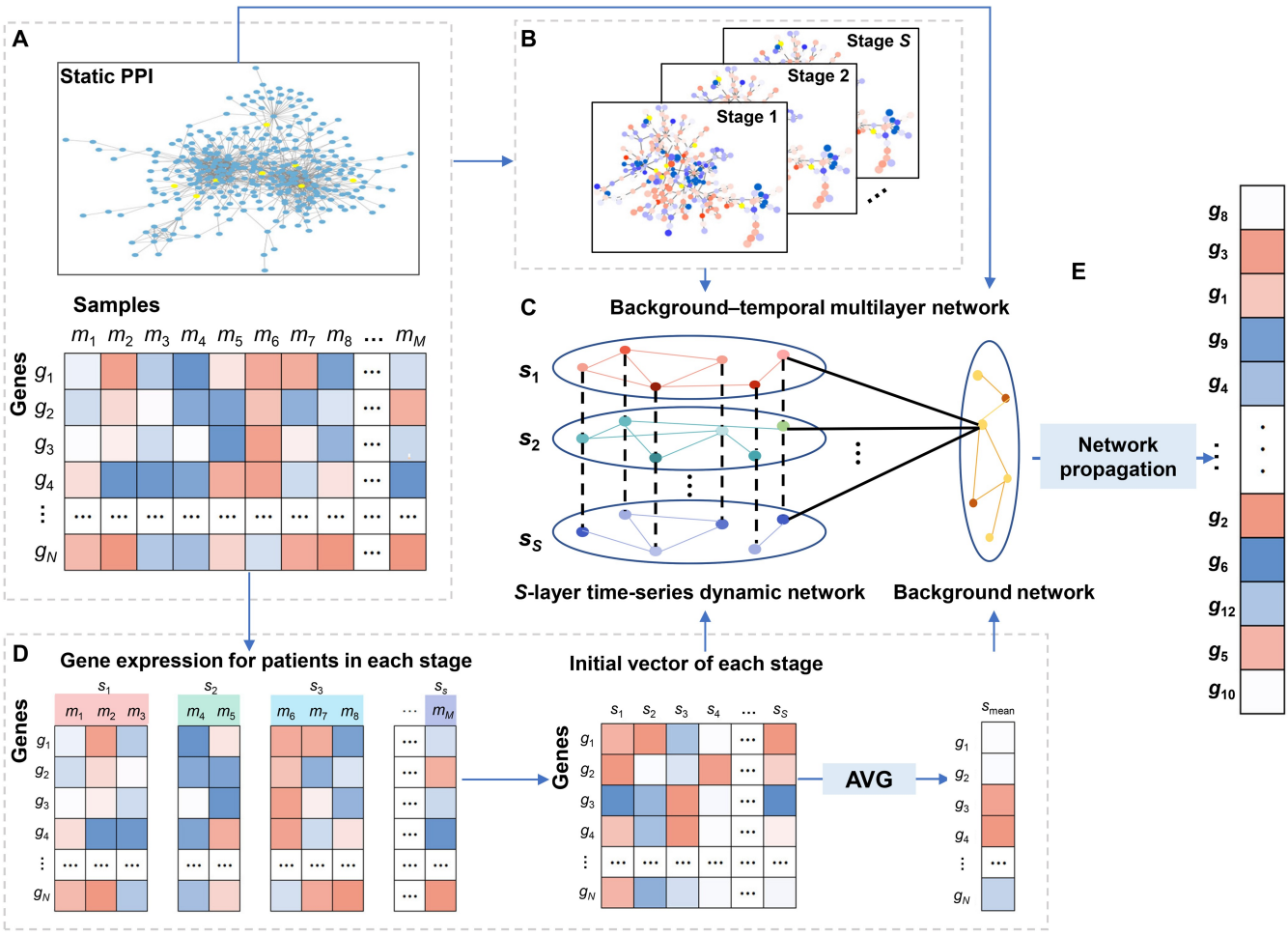


Figure 1 The schematic diagram of DyNDG

A. The original inputs to DyNDG. **B**. The *S*-layer time-series dynamic network. **C**. The background–temporal multilayer biological network. The initial probability vector for each stage is applied to corresponding gene nodes in the corresponding-stage network layer. Smean is applied to corresponding gene nodes in the background network layer. **D**. Obtaining the initial probability vector of gene nodes for each stage through differential expressed gene analysis from time-series gene expression data of patients. Smean means taking the arithmetic mean of the initial probability vectors across all stages. Smean is used as the initial probability vector of gene nodes in the background network layer. **E**. The vector of scores for leukemia-related genes is obtained as the output of network propagation on the background–temporal multilayer biological network. DyNDG, Dynamic Network-based model integrating Differentially Expressed Genes; PPI, protein–protein interaction; AVG, average.

encompasses the gene interaction information from all patients within that stage. Given the inherent noise in gene expression data and the diverse expression patterns across genes, we employed the k-sigma method to calculate the active probability of each gene across various patient samples [29]. This approach effectively distinguishes the active level of genes in each patient sample when constructing the biological network for each disease stage. The gene expression matrix $G^{N\times M}$ contains expression profiles of M patient samples from S disease stages, where N is the number of genes. For each patient sample m, $G_m(g_n)$ denotes the gene expression value of gene g_n in this patient. For all M patient samples, define $\overline{G(g_n)}$ and $\sigma(g_n)$ as the algorithmic mean and standard deviation of gene expression values for g_n, respectively. We calculated k-sigma threshold (where k represents the number of sigma) of each gene g_n by:

$$Active\_Th_k(g_n) = \overline{G(g_n)} + k\cdot\sigma(g_n)\cdot\left(1-\frac{1}{1+\sigma^2(g_n)}\right),\quad k=1,\ 2,\ \text{or}\ 3 \tag{1}$$

$$\overline{G(g_n)} = \frac{\sum_{m=1}^{M} G_m(g_n)}{M} \tag{2}$$

$$\sigma^2(g_n) = \frac{\sum_{m=1}^{M}\left(G_m(g_n)-\overline{G(g_n)}\right)^2}{M-1} \tag{3}$$

$Active\_Th_k(g_n)$ represents k active threshold of g_n. The active probability of a protein corresponding to gene g_n in the patient sample m is calculated as follows [29]:

$$Active\_Pr_m(g_n) = \begin{cases} 0.99 & if\ \ G_m(g_n)\geq Active\_Th_3(g_n) \\ 0.95 & if\ \ Active\_Th_3(g_n)>G_m(g_n)\geq Active\_Th_2(g_n) \\ 0.68 & if\ \ Active\_Th_2(g_n)>G_m(g_n)\geq Active\_Th_1(g_n) \\ 0 & if\ \ G_m(g_n)<Active\_Th_1(g_n) \end{cases} \tag{4}$$

In general, the active probability value of a protein can be used as a measure of its active level in a patient sample. In order to construct the time-series dynamic network, we initially constructed the PPI network $Net_m\left(V_m^{1\times N},\ E_m,\ AP_m^{1\times N},\ A_m\right)$ for each patient sample m; $V_m^{1\times N}=\{v_1,\ldots,v_i,\ldots,v_j,\ldots,v_N\}$ is the set of N gene/protein nodes; E_m is the set of interactions; $AP_m^{1\times N}$ is the active probability vector of N gene/protein nodes; A_m represents the adjacency matrix which represents the confidence of the interactions between genes. Reweight A_m to A'_m using $AP_m^{1\times N}$ as follows:

$$A'_m[v_i,v_j] = AP_m[v_i]\times A_m[v_i,v_j]\times AP_m[v_j] \tag{5}$$

For the sake of concise and compact expression, a calculation method called "Broadcasting Multiplication" can be employed to describe the reweighting operation. The calculation process of Equation 5 can be found in section 1 of File S1.

Divide M samples into S leukemia stages, and the number of samples in each stage s is M_s. The adjacency matrix $A^{[s]}(s=1,\ldots,S)$ of leukemia stage s can be calculated as follows:

$$A^{[s]} = \frac{\sum_{m=1}^{M_s} A'_m}{M_s} \tag{6}$$

Finally, we constructed the time-series dynamic networks based on the adjacency matrix $A^{[s]}$ of each leukemia stage. Figure 2 shows the process of building the time-series dynamic networks. Firstly, the k-sigma threshold for each gene g_n is calculated from the gene expression matrix. The active probability vector of proteins corresponding to genes is obtained by using this threshold. By integrating the gene expression matrix and the static PPI network, we obtained the initial network for each patient sample. Further, new networks of patient samples are obtained by reweighting the edges with the active probabilities of the two corresponding genes using "Broadcasting Multiplication". The time-series dynamic networks corresponding to multiple leukemia stages are obtained by taking the "Union" of the patient sample networks. Note that the networks for all stages have the same number of genes, but the genes in different networks may be interconnected with different weights.

Construction of background–temporal multilayer biological network

The comprehensiveness and richness of biological networks significantly impact the ability to predict leukemia genes. To capture more realistic network changes throughout the course of leukemia, building upon the time-series dynamic networks simulating the dynamic alterations in biological networks, we proposed a background–temporal multilayer biological network framework. This framework integrates the dynamic networks and the static network into a multilayer network, effectively amalgamating information from both static and dynamic networks. The multilayer network is constructed by connecting shared gene nodes across different network layers, for jumping to a different layer through the same node when conducting propagation on it. The specific steps are as follows. Firstly, we constructed a S-layer temporal dynamic network by connecting the S time-series dynamic network layers of neighboring disease stages through shared nodes. Secondly, we connected the background network layer constructed by the static PPI network with each layer of the S-layer dynamic network by linking the corresponding gene counterparts. In the end, we obtained a connected multilayer network $Net_{multi}=(V_{multi},\ E_{multi},\ A_{multi})$ with $L(=S+1)$ layers, where S is the number of layers in the aforementioned dynamic network; $A_{multi}=\{A^{[1]},\ldots,A^{[S]},A^{[L]}\}$ is the set of adjacency matrixes for the L layers, and $A^{[L]}$ is the adjacency matrix of static PPI network; $V_{multi}=\left\{v_i^{[l]}\mid i=1\sim N,\ l=1\sim L\right\}$ is the set of nodes (note that each gene has a counterpart node in each network layer); $E_{multi}=\left\{edge\left(v_i^{[l]},v_j^{[l]}\right)\mid i=1\sim N-1, j=i+1\sim N,\ l=1\sim L,\ A^{[l]}\left[v_i^{[l]},v_j^{[l]}\right]\neq 0\right\}\cup\left\{edge\left(v_i^{[\alpha]},v_i^{[\beta]}\right)\mid i=1\sim N,\ \alpha=1\sim S-1,\ \beta=\alpha+1\right\}\cup\left\{edge\left(v_i^{[\alpha]},v_i^{[\beta]}\right)\mid i=1\sim N,\ \alpha=1\sim S,\ \beta=L\right\}$ is the set of interlayer connections and intralayer PPIs. Note that all the edges here are undirected.

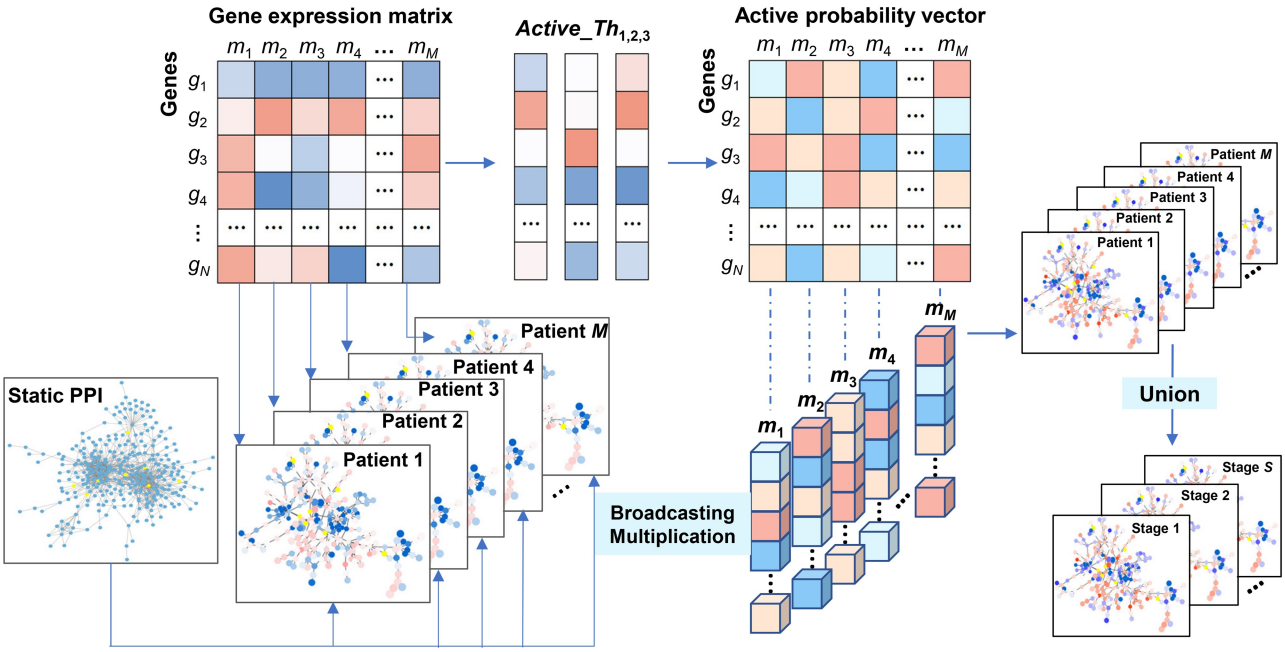


Figure 2 The schematic diagram of time-series dynamic network construction
The process of constructing time-series dynamic network. Inputs consist of a gene expression matrix and a static PPI network. Union indicates network union of patient samples at the same stage.

Complex information about leukemia progression flows through edges in the multi-layer network. The background–temporal multilayer biological network consists of two parts: the static network and the time-series dynamic network. The relevant details of the background–temporal multilayer biological network are presented in Table S2.

Network propagation on background–temporal multilayer biological network

To identify leukemia-related genes, we applied a network propagation process on the background–temporal multilayer biological network Net_{multi}. During the walking process, the particles in the Net_{multi} either restart with a certain probability, or walk along a randomly selected (intra-layer or inter-layer) edge to a neighboring node (corresponding one of two possible actions: walk within the same layer, or jump to the counterparts of the same node in a different layer). Finally, as the random walk process concludes after a certain period, the scoring value of each node in the network will be obtained. During the random walk process, the scoring vector P_{t+1} of the nodes at step $t+1$ can be obtained as follows:

$$P_{t+1} = (1-\gamma)T_L^c P_t + \gamma P_{RS} \tag{7}$$

where $\gamma \in (0,1)$ is the probability of restart, T_L^c is the column-normalized transition matrix of the multilayer network, $P_{RS} \in R^{NL\times 1}$ represents the initial scoring vector of NL nodes in the L-layer network, and $P_t \in R^{NL\times 1}$ represents the scoring vector of NL nodes at step t.

Different stages of disease development are often accompanied by changes in gene expression levels and dynamic alterations in the network structure of gene interactions. However, most existing disease gene prediction methods do not take these biological phenomena into account. To address this, we proposed incorporating differential information during leukemia development as prior knowledge in the network propagation model. Specifically, for the gene expression matrix divided into S disease stages, we conducted differential gene expression analysis between each stage and the others by using the limma [46] R package to obtain the T-statistic values of the genes in the corresponding layer of each stage. The absolute value of these T-statistic values is used to determine the initial scoring vector of genes, as they contain valuable information that effectively distinguishes each stage from the others. For the temporal multilayer network, we use the absolute T-statistics of the genes calculated in each of the S stages as the initial probability for each stage, thus incorporating stage-specific differential gene expression information. For the background network, we calculated the average of the aforementioned gene scores obtained from the differential analysis across all stages as the initial scores for each gene. By incorporating the T-statistic values into the background–temporal multilayer biological network, we can enhance the network with more comprehensive stage-specific information. As a result, we got the initial scoring vector $P_{RS} \in R^{NL\times 1}$ in the L-layer network, which incorporates differential information of gene expression during disease progression as prior knowledge into the network.

Then, we constructed the $NL \times NL$ transition matrix T_L of the L-layer network by

$$T_L = \begin{bmatrix} (1-\delta)B & \delta J^T \\ \delta J & T_S \end{bmatrix} \tag{8}$$

where $\delta \in [0,\ 1]$, $e = (1,1,\ldots\ldots,1)^T \in R^{S\times 1}$, $J = \frac{1}{S} e \otimes I$, and I is a $N \times N$ identity matrix. Here, B is the adjacency matrix of the background network, δ controls the probability of jumping between two types of network layers, and J is the inter-network transition matrix that controls the jumping

from the background network layer to each layer of the time-series dynamic network. Let T_S be the $NS \times NS$ transition matrix of the S-layer temporal dynamic network by

$$T_S = \begin{Bmatrix} (1-\mu)A^{[1]} & \mu I & 0 & \cdots & 0 \\ \mu I & (1-\mu)A^{[2]} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & (1-\mu)A^{[S-1]} & \mu I \\ 0 & \cdots & 0 & \mu I & (1-\mu)A^{[S]} \end{Bmatrix} \quad (9)$$

where $\mu \in [0,\ 1]$ controls the transition probability between the temporal dynamic network layers and $A^{[s]}$ $(s = 1 \sim S)$ represents the adjacency matrix of each layer of the S-layer dynamic network. The matrix T_S controls the inter-layer and intra-layer jumps of the S-layer temporal dynamic network. We obtained T_L^c by column normalization of matrix T_L. The example and parameter description of Equation 9 can be found in section 2 of File S1.

We plunged the column-normalized matrix T_L^c and initial scoring vector P_{RS} into the aforementioned iterative Equation 7, which is iteratively updated until the difference between P_t and P_{t+1} falls below 10^{-6}. Upon convergence, gene nodes in the background–temporal multilayer network are assigned a steady-state scoring vector P_∞. In the background network, the steady-state scoring vector P_∞^B gives a measure of how strongly a gene is associated with leukemia. If $P_\infty^B(v_i) > P_\infty^B(v_j)$, the gene corresponding to node v_i is more closely related to leukemia. Then genes are sorted in descending order based on P_∞^B, generating a ranking list of leukemia-related genes. Therefore, the higher-ranked genes can be considered as promising genes related to leukemia.

Evaluation methods

Evaluation strategy and metrics

We conducted a systematic evaluation of the proposed DyNDG model, which involves constructing a background–temporal multilayer network and predicting leukemia-related genes on this network. The set of candidate genes consists of both a test gene set and a control gene set. The test set is composed of known disease genes. We considered three different schemes for constructing the control set. (1) Artificial Linked Interval Control Set (ALI): In this scheme, resembling linkage analysis or association studies, we selected 99 control genes for each test gene from the genes located closest to the test gene on the same chromosome. (2) Randomized Control Set (RC): This scheme mimics the scenario in exome sequencing studies. We randomly selected 99 control genes for each test gene from the entire genome. (3) Whole Genomic Control Set (WG): In this scheme, all genes in the network, excluding the known disease genes, were considered as control genes. For each leukemia, let Set_t denote the test gene set, let Set_c denote the control gene set, while $Set = \{Set_t, Set_c\}$ represents the set of candidate genes. Combining the ranking list of genes obtained by DyNDG with Set can obtain the top k ranking list R_k of candidate genes. We can quantitatively calculate four common metrics, *i.e.*, Topk_Precision, Topk_Recall, area under the receiver operating characteristic curve (AUROC), and area under the precision–recall curve (AUPRC), to evaluate the proposed method.

Comparison algorithms

Network-based methods have gained significant popularity in predicting disease genes and have demonstrated remarkable prediction results. To compare the performance, we considered two categories of network-based disease gene prediction algorithms (Table S3). For each category, we selected representative algorithms: (1) algorithms based on the single network: Random Walk with Restart (RWR) [19]; (2) algorithms based on multiple networks: methods utilizing multiple independent networks such as Random Walk with Restart integrating the Discounted Rating System (RWRDRS) [47], and methods based on multiple interconnected networks like Random Walk with Restart on the Merged Graph (RWRMG) [18] and Random Walk with Restart on Multiplex Graphs (RWRMP) [20]. Furthermore, it is a common strategy to leverage the differences in gene expression levels between normal and disease stages in predicting disease genes. Hence, we also included two representative algorithms: (1) the T-test method, which generates disease association measures similar to network-based methods but without utilizing network information; and (2) the Disease-Specific Network Enhancement Prioritization (DiSNEP) method, which enhances a general gene network into a disease-specific network and then prioritizes gene associations on the enhanced network [48].

Functional enrichment analysis

Functional enrichment analysis is a technique widely used in bioinformatics to identify biological functions or processes that are overrepresented in a given set of genes or proteins. This analysis helps in understanding the biological significance of a gene list or a set of DEGs by linking them to known biological functions, pathways, or processes. We used the clusterProfiler R package [49] to perform Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on the top 1% ranked genes in the predicted leukemia-related candidate gene list to investigate their associations with leukemia-related functions and metabolic pathways.

Results and discussion

DyNDG greatly improves predictive ability

To evaluate the performance of DyNDG quantitatively, we generated the set of leukemia-related candidate genes by the decreasing order of the predictive scores, and the evaluation of DyNDG was conducted by four common metrics (*i.e.*, Topk_Precision, Topk_Recall, AUROC, and AUPRC), along with three kinds of the control sets: ALI, RC, and WG.

In identifying leukemia-related genes, our DyNDG method significantly outperforms the single-network-based method RWR [19] based on the static network upon Topk_Recall and Topk_Precision metrics, because of incorporating temporal dynamic network information of disease progression, while RWR demonstrates better performance than the classical T-test method (**Figure 3**). Multi-network-based methods have been designed to operate on multiple independent or interconnected networks, enabling effective utilization of more information from these networks. Here, we also compared our DyNDG method with three multi-network-based methods: RWRDRS [47], RWRMG [18], and RWRMP [20]. The comparison results in **Figure 4**A–C, respectively, show that DyNDG outperforms these multi-network-based

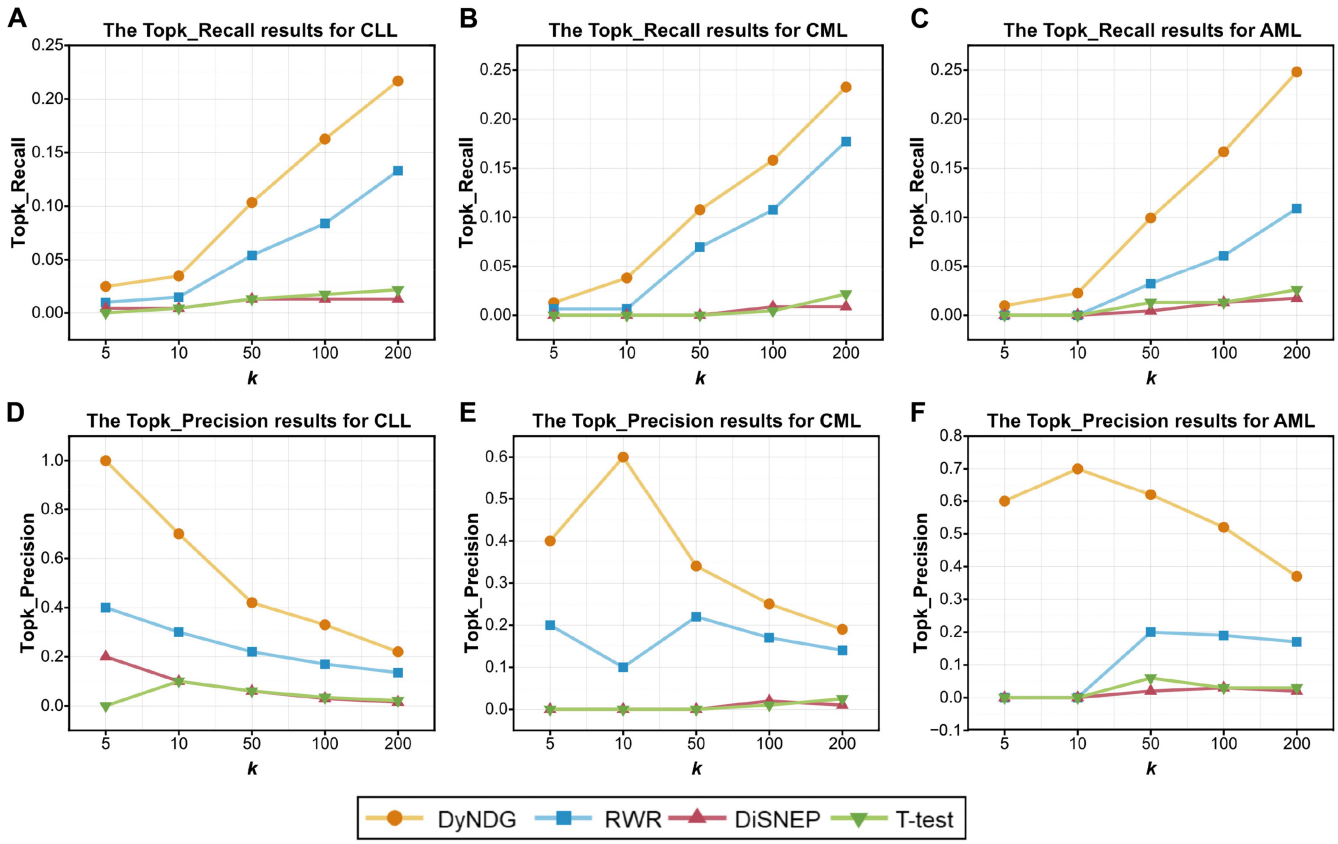


Figure 3 Performance comparison of DyNDG, RWR, T-test, and DiSNEP on WG using Topk_Recall and Topk_Precision
A.–**C**. The Topk_Recall results for CLL (A), CML (B), and AML (C). **D**.–**F**. The Topk_Precision results for CLL (D), CML (E), and AML (F). When $k = 5$, 10, 50, 100, and 200, DyNDG consistently exhibits higher Topk_Recall and Topk_Precision values than other methods. RWR, Random Walk with Restart; DiSNEP, Disease-Specific Network Enhancement Prioritization; WG, Whole Genomic Control Set; CLL, chronic lymphocytic leukemia; CML, chronic myeloid leukemia; ALL, acute myeloid leukemia.

methods in identifying CLL-related genes, CML-related genes, and AML-related genes, in terms of Topk_Recall performance. It can be observed that increasing the k value appropriately leads to a noticeable improvement in Topk_Recall. It is indeed a difficult task to search for disease-related genes in almost full gene set. The recall scores of all other algorithms are also relatively low. Additionally, if the candidate range for search/prediction is limited, the recall of the algorithm will usually improve, including utilizing more known information. The results in Figure 4D–F also demonstrate that DyNDG can more precisely predict CLL-related genes, CML-related genes, and AML-related genes, respectively, in terms of Topk_Precision performance. Figure S1 shows the overlaps of the top 100 genes predicted by different methods for CLL, CML, and AML. Upon analyzing the overlaps of candidate genes predicted by DyNDG and other methods, we observed more than 50% overlap between DyNDG and other methods, with the majority of the overlapping predictions coming from multi-network methods. Additionally, DyNDG recommended some genes that were not identified by other algorithms, which further illustrates the potential of DyNDG to complement and enhance existing predictive approaches. Figures S2–S5 also show that DyNDG achieves the best on three leukemia datasets and different kinds of the control sets upon Topk_Recall and Topk_Precision, re-affirming that DyNDG can obtain more reliable and stable results, independent to the types of control sets.

Moreover, the results of the performance comparison between DyNDG and other methods, using the AUROC and AUPRC metrics, are shown in Table 2. DyNDG significantly outperforms all the comparison algorithms in terms of AUROC and AUPRC, demonstrating a notable advancement in identification of leukemia disease genes.

Effect of parameters and data

We systematically examined the impact of parameters (*i.e.*, δ, μ, and γ) on the performance of the DyNDG model in three types of leukemia. Details of the analysis can be found in section 3 of File S1, and the results are presented in Figures S6–S11. δ and μ regulate the transitions of the background–temporal multilayer network. The results indicate that the model performance remains relatively stable as the parameters δ and μ vary, which suggests that the multilayer network framework can effectively integrate background network knowledge and temporal dynamic network information, ensuring the robustness of the model. A lower restart probability γ corresponds to better model performance, indicating that the background–temporal multilayer network contains rich information related to leukemia disease genes and is crucial for predictive performance. Therefore, we chose $\delta = 0.5$, $\mu = 0.5$, and $\gamma = 0.1$ as the parameter combination that yields the relatively best model performance.

Additionally, we investigated the influence of different static network data on the performance of DyNDG. Further information about this study can be found in section 4 of File S1.

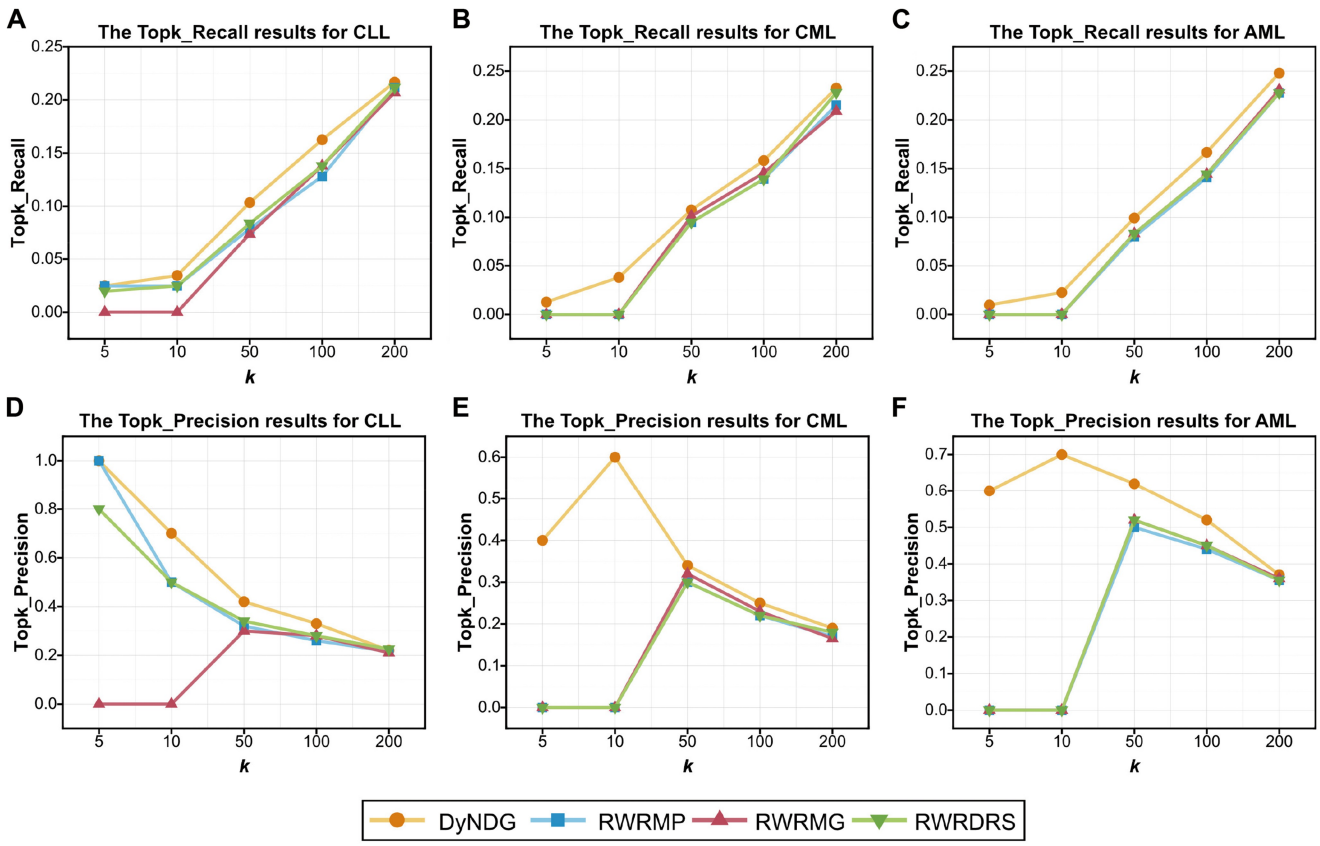


Figure 4 Performance comparison of DyNDG, RWRDRS, RWRMG, and RWRMP on WG using Topk_Recall and Topk_Precision
A.–**C**. The Topk_Recall results for CLL (A), CML (B), and AML (C). **D**.–**F**. The Topk_Precision results for CLL (D), CML (E), and AML (F). When $k = 5$, 10, 50, 100, and 200, DyNDG consistently outperforms other methods in terms of Topk_Recall and Topk_Precision. RWRDRS, Random Walk with Restart integrating the Discounted Rating System; RWRMG, Random Walk with Restart on the Merged Graph; RWRMP, Random Walk with Restart on Multiplex Graphs.

The results in Figure S12 show that DyNDG performs better when using the static PPI network of STRING [42], which provides a broader and more diverse range of PPI information. As shown in Figures S13 and S14, we also studied the differences in the initial probability distribution between different layers of the background–temporal multilayer network. The initial probabilities of the background network comprehensively consider the characteristics of the temporal multilayer networks at various stages. The differences in the initial probability distribution across different stages of disease progression indicate that the T-statistic distributions obtained from differential gene analysis at each stage can effectively represent the characteristics of each stage.

To assess the contribution of different components of the DyNDG model to the predictive power, we performed two different experiments: (1) DyNDG_static, where the random walk process was performed only in the background network; and (2) DyNDG_dynet, where the random walk process was conducted exclusively in the dynamic network (detailed experimental setup can be found in section 5 of File S1). Results (Figure S15) show that the background network plays a crucial role in the model which builds upon universal interaction patterns in biomolecules, providing a comprehensive background and a solid foundation for our model. Simultaneously, the introduction of the dynamic network enhances the predictive capability of the model, particularly in simulating dynamic changes in biological networks at different leukemia stages and capturing potential gene associations. This suggests that in the prediction of leukemia-related genes, the collaboration of static and dynamic networks contributes to the overall improvement in predictive performance of the model.

Comprehensive analysis of candidate disease genes for AML, CLL, and CML

In this study, we further conducted case studies and comprehensive analysis for specific leukemia (*i.e.*, AML, CLL, and CML). Specifically, we applied DyNDG to each of the three leukemias and obtained a ranked list of genes for each of them. Then, we individually filtered each of the three gene lists, by removing the genes reported in MalaCards [44] that are associated with the respective types of leukemia. Additionally, we excluded the human housekeeping genes collected from the Housekeeping and Reference Transcript Atlas (HRT Atlas) database [50]. As a result, we obtained three lists of candidate genes associated with their respective types of leukemia.

Integrative analysis of functional annotations and clinical survival outcomes in AML

AML is the most common and lethal adult acute leukemia, characterized by its aggressive nature, poor prognosis, and high susceptibility to relapse after treatment [51]. Through the application of DyNDG, we identified AML-related genes that have the potential to drive AML progression and contribute to poor prognosis in patients.

Table 2 Performance comparison of DyNDG and other methods using AUROC and AUPRC metrics

Leukemia	Type	Method	AUROC	AUPRC
CLL	Single-network	RWR	0.8080	0.0646
	Multi-network	RWRMP	0.7967	0.1396
		RWRMG	0.8394	0.1078
		RWRDRS	0.7941	0.1433
	Differential gene expression	T-test	0.7116	0.0224
		DiSNEP	0.7177	0.0231
		DyNDG	**0.8875**	**0.1753**
CML	Single-network	RWR	0.7881	0.0676
	Multi-network	RWRMP	0.8487	0.0910
		RWRMG	0.8308	0.0920
		RWRDRS	0.8556	0.0940
	Differential gene expression	T-test	0.5798	0.0162
		DiSNEP	0.5712	0.0154
		DyNDG	**0.8983**	**0.1260**
AML	Single-network	RWR	0.7493	0.0745
	Multi-network	RWRMP	0.8422	0.1887
		RWRMG	0.8445	0.1918
		RWRDRS	0.8433	0.1927
	Differential gene expression	T-test	0.5478	0.0155
		DiSNEP	0.5402	0.0145
		DyNDG	**0.8797**	**0.2334**

Note: DyNDG, Dynamic Network-based model integrating Differentially Expressed Genes; AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; RWR, Random Walk with Restart; DiSNEP, Disease-Specific Network Enhancement Prioritization; RWRDRS, Random Walk with Restart integrating the Discounted Rating System; RWRMG, Random Walk with Restart on the Merged Graph; RWRMP, Random Walk with Restart on Multiplex Graphs.

Functional enrichment analysis conducted on the top 1% ranked genes in the AML-related candidate gene list revealed the KEGG pathways and GO terms most associated with AML, as shown in **Figure 5A and B**, respectively. The most relevant KEGG pathways consist of cancer-related pathways (*e.g.*, cell cycle and MAPK signaling pathways), signaling pathways regulating pluripotency of stem cells, and leukemia-related pathways (*e.g.*, leukocyte transendothelial migration and platelet activation). The GO enrichment analysis shows many AML-related biological processes (*e.g.*, leukocyte cell–cell adhesion, leukocyte proliferation, and cellular response to inorganic substance), molecular functions (*e.g.*, catalytic activity, acting on DNA and cyclic nucleotide-dependent protein kinase activity), and cellular components (*e.g.*, chromosomal region). To further investigate the candidate genes enriched in these AML-related pathways, we obtained a set of multi-pathway enrichment genes as shown in Figure 5C (corresponding relationships between GO term IDs and descriptions, as well as between KEGG pathway IDs and descriptions, can be found in Table S4).

By conducting survival analysis, we observed significant differences in overall survival among patient groups with different expression levels of many multi-pathway-enriched genes (*e.g.*, *TLR4*, *MAPK8*, *SHH*, and *PIK3R1*) (Figure 5D). *TLR4* is expressed by AML cells, several bone marrow stromal cells, and non-leukemic cells involved in inflammation. *TLR4* in leukemia cells is of significant importance for the growth and development of leukemia cells in human AML, and targeting *TLR4* may have direct and indirect effects on leukemogenesis. A study has indicated that high expression of *TLR4* is associated with a decrease in survival rates after intensified anti-leukemia treatment [52]. This suggests that high expression of *TLR4* may be related to lower survival rates in AML patients, which is consistent with our survival analysis results of *TLR4* shown in Figure 5D. *MAPK8*, also known as *JNK1*, is a member of the MAPK family. The aberrant activation of the MAPK signaling pathway is closely associated with the proliferation, survival, drug resistance, and metastatic capacity of AML cells [53]. Therefore, as a member involved in regulating the MAPK signaling pathway, *MAPK8* may play an important role in key biological processes regulating AML. Studies have shown that in AML, certain anticancer drugs may exert their anticancer effects by activating *MAPK8* to inhibit cell proliferation in AML cell lines [54]. In the clinical survival analysis presented in Figure 5D, it was observed that low expression of *MAPK8* may be associated with lower survival rates. This could be attributed to the aberrant activation or impaired function of the MAPK signaling pathway resulting from the lower expression of *MAPK8*, ultimately leading to reduced survival rates in AML patients. These findings provide important clues for further investigating the role of *MAPK8* in the development and treatment of AML. *SHH* has been implicated in the maintenance of cancer stem cells (CSCs), playing a critical role in the development of drug resistance and disease relapse in AML. Research findings suggest that *SHH* could serve as a prognostic marker for CSCs in AML, providing valuable insights into disease outcomes [55]. Furthermore, in the presence of lipopolysaccharide/tumor necrosis factor-α/interferons (LPS/TNF-α/IFNs), *SHH* antagonists have been utilized for the treatment of AML patients [56]. Previous evidence suggests that *SHH* may be associated with the occurrence, progression, and treatment of AML. As shown in Figure 5D, a significant difference in survival probability was observed between patient groups with different levels of *SHH* expression, providing clinical evidence for the pivotal role of *SHH* in the development of AML. *PIK3R1* is the gene encoding the p85α subunit of the PI3K regulatory subunit, playing a critical regulatory role in the PI3K/Akt signaling pathway. Previous studies have indicated that *PIK3R1* is an actionable gene in AML [57]. The PI3K/Akt signaling pathway regulated by *PIK3R1* is involved in various cellular processes such as cell survival, proliferation, and metabolism in normal cells. Its aberrant activation is associated with the development and progression of multiple tumor types, including AML. In 50%–80% of AML patients, constitutive activation of the PI3K/Akt pathway has been detected, which is associated with reduced overall survival [58]. Therefore, the abnormal expression of *PIK3R1*, which regulates the PI3K/Akt pathway, may be implicated in the occurrence and progression of AML. Figure 5D illustrates the relationship between *PIK3R1* expression and the survival rate of AML. The identification of these genes provides valuable insights into the potential molecular mechanisms, pathogenesis, and key factors influencing the prognosis of AML. It offers opportunities to identify new therapeutic targets and has the potential to improve prognosis assessment and develop personalized treatment strategies.

Integrative analysis of functional annotations and networks in CLL

CLL is one of the most prevalent types of leukemia, and there is currently no definitive cure available. Despite advancements in treatment with novel therapies and targeted drugs, there is a continuing need for new targeted treatments and

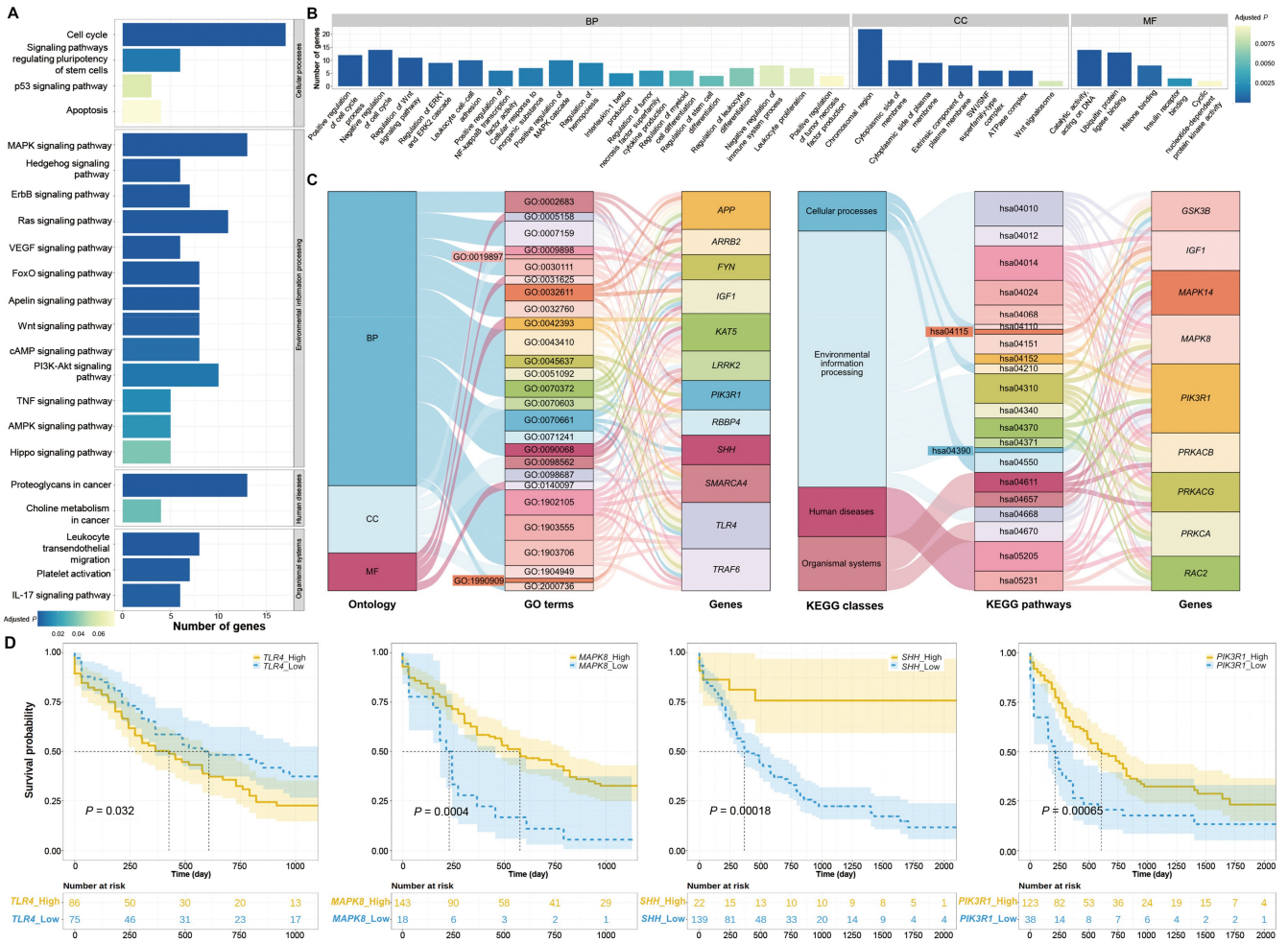


Figure 5 DyNDG predicts AML-related genes closely associated with clinical survival

A. The bar chart showcases the results of KEGG pathway enrichment analysis, highlighting the most relevant KEGG pathways associated with AML. **B**. The bar chart illustrates the outcomes of GO enrichment analysis, highlighting the most relevant GO terms associated with AML. **C**. The Sankey diagrams show the associations between GO terms and multi-pathway enriched genes, as well as the associations between KEGG pathways and multi-pathway enriched genes. **D**. Kaplan–Meier overall survival curves of AML patients grouped by the averaged expression of *TLR4*, *MAPK8*, *SHH*, and *PIK3R1*, respectively (with the median value as the threshold). *P* values were calculated by the log-rank test. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; BP, biological process; CC, cellular component; MF, molecular function.

innovative strategies to improve efficacy and survival rates [59]. By applying DyNDG to CLL, we predicted CLL-related genes that can serve as references for discovering new therapeutic targets and developing personalized treatment strategies.

Functional enrichment analysis conducted on the top 1% candidate genes revealed the KEGG pathways and GO terms most relevant to CLL, as shown in **Figure 6A and B**, respectively. The KEGG pathways most relevant to CLL include cancer-related pathways (such as cell cycle, MAPK signaling pathway, and transcriptional misregulation in cancer), B cell receptor signaling pathway, and pathways associated with chronic leukemias (such as chronic myeloid leukemia and leukocyte transendothelial migration). GO enrichment analysis revealed numerous CLL-related biological processes (such as leukocyte cell–cell adhesion, leukocyte proliferation, lymphocyte proliferation, and regulation of tumor necrosis factor production), molecular functions (such as kinase regulatory activity, SMAD binding, and G protein-coupled receptor binding), and cellular components (such as chromosomal regions). Then, we obtained a set of multi-pathway enriched genes, which are enriched in multiple GO terms and KEGG pathways, through Sankey diagrams as shown in Figure 6C (enriched pathways are identified by their respective GO term IDs and KEGG pathway IDs, which can be found in Table S4, along with their descriptions).

To explore the criticality of candidate genes enriched in CLL-related pathways in the PPI network, we conducted network analysis using Cytoscape [60]. Based on human static PPI network sourced from the STRING database [42], we analyzed the node degree centrality of candidate genes enriched in CLL-related pathways and CLL-related genes reported in MalaCards [44]. Subsequently, we generated a network diagram as shown in Figure 6D. It is well-known that genes with high-degree centrality play a crucial role in regulating and influencing information transmission, network stability, functional integration, and disease relevance within the PPI network. In the static PPI network, candidate genes enriched in CLL-related pathways exhibit degrees similar to known CLL-related genes. This suggests that the candidate genes may have similar network influence as the known CLL-related genes and could play a crucial role in the progression of CLL. Many of these candidate genes even surpass certain known

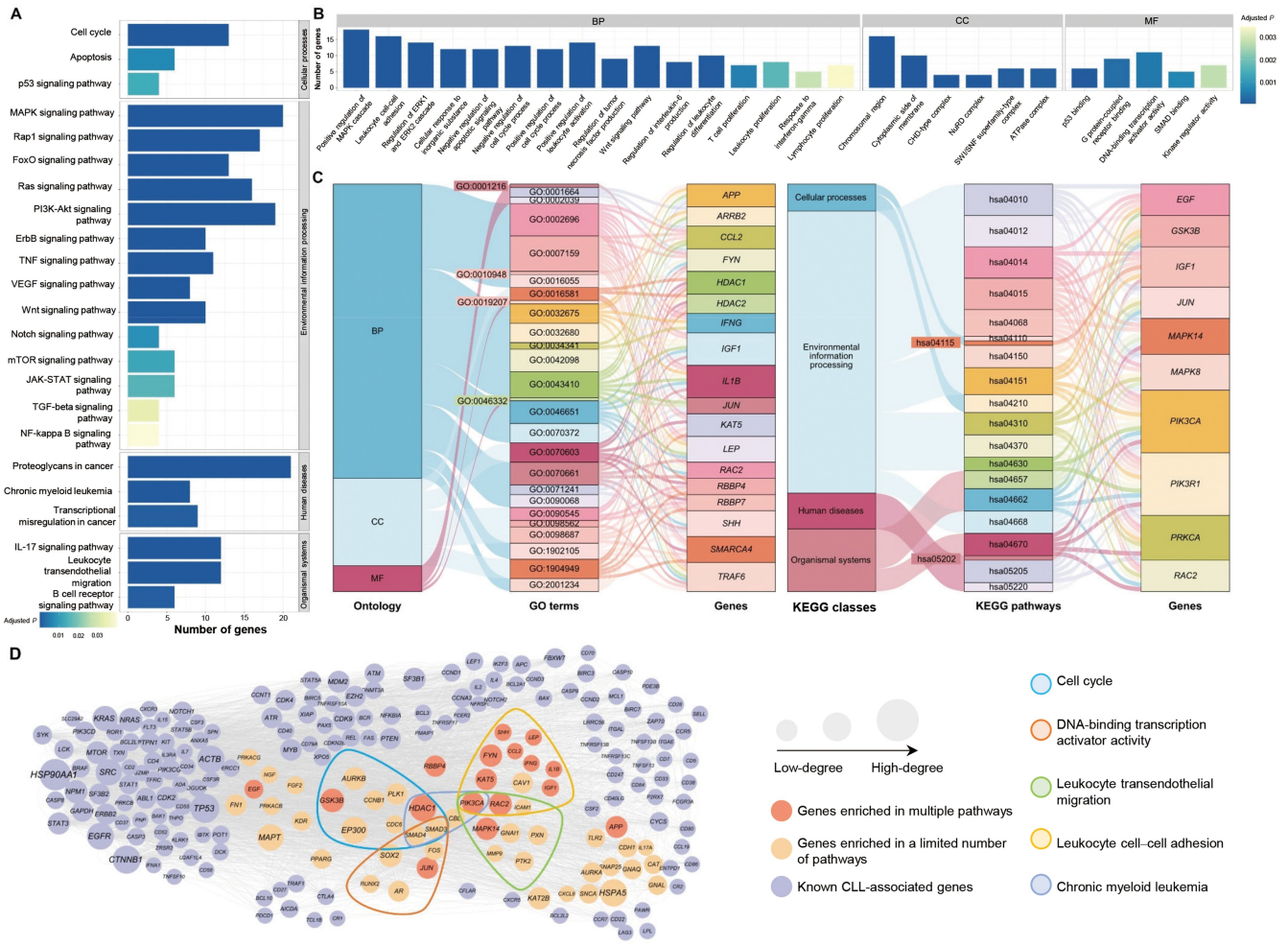


Figure 6 DyNDG predicts CLL-related genes possessing network hub characteristics

A. The bar chart showcases the results of KEGG pathway enrichment analysis, highlighting the most relevant KEGG pathways associated with CLL. **B**. The bar chart illustrates the outcomes of GO enrichment analysis, highlighting the most relevant GO terms associated with CLL. **C**. The Sankey diagrams show the associations between GO terms and multi-pathway enriched genes, as well as the associations between KEGG pathways and multi-pathway enriched genes. **D**. The network graph reflects the degree centrality of candidate genes enriched in CLL-related pathways and known CLL-related genes.

CLL-related genes in terms of degree centrality, indicating that they may have a greater importance in regulating the network than some of the known key genes. *HDAC1*, *PIK3CA*, and *RAC2* are genes that are enriched in multiple pathways and exhibit high-degree centrality in the PPI network. Lastly, we conducted literature research on *HDAC1*, *PIK3CA*, and *RAC2*. *HDAC1* is one of the components of the histone deacetylase complex, and elevated *HDAC* enzyme activity has been found to be associated with the development of leukemia and other cancers. Significantly elevated levels of *HDAC1* expression have been observed in CLL [61], and studies have indicated that *HDAC1* can act as a transcriptional activator in CLL, promoting CLL cell survival and progression [62]. *PIK3CA* is a component of the PI3K pathway and has long been described as an oncogene. Inhibition of PI3K subtype associated with *PIK3CA* has been shown to be potentially valuable in CLL [63]. *RAC2* is a member of the RAC subfamily of Rho GTPases and is closely associated with oncogenic signaling. Existing studies have found that *RAC2* can activate two known PLCγ2 mutations in CLL, suggesting that *RAC2* may play a significant role in the survival and proliferation of CLL cells carrying these mutations [64]. Research has also indicated that upregulation of *RAC2* expression may be associated with resistance to ibrutinib in CLL [65]. In conclusion, *HDAC1*, *PIK3CA*, and *RAC2* may have significant biological functions and clinical implications in CLL. They could potentially serve as potential therapeutic targets or biomarkers for predicting CLL progression and treatment response.

Integrative analysis of functional annotations and differential expression in CML

CML is a chronic leukemia caused by the fusion of the *BCR* and *ABL* genes. Inhibiting the activity of the BCR–ABL fusion protein with tyrosine kinase inhibitors (TKIs) can effectively control the proliferation and survival of CML cells, leading to sustained clinical and molecular remission in the majority of patients. However, long-term use of TKIs may result in side effects, and some patients may develop resistance or intolerance to TKI treatment [66]. By applying DyNDG to CML, predicting genes associated with its disease progression may provide valuable insights for studying new targets in CML, addressing the issue of drug resistance, and exploring novel therapeutic approaches for CML. Functional enrichment analysis conducted on the top 1% candidate genes revealed the KEGG pathways and GO terms most related to CML, as shown in Figure 7A and B, respectively.

The KEGG pathways most relevant to CML include cancer-related pathways (such as cell cycle, PI3K-Akt signaling pathway, and proteoglycans in cancer), the pathways closely associated with the immunity (such as Th1 and Th2 cell differentiation), and pathways associated with chronic leukemias (such as chronic myeloid leukemia and leukocyte transendothelial migration). GO enrichment analysis revealed numerous CML-related biological processes (such as leukocyte cell–cell adhesion, leukocyte proliferation, myeloid leukocyte differentiation, and leukocyte mediated immunity), molecular functions (such as DNA-binding transcription factor binding and ubiquitin-like protein ligase binding), and cellular components (such as chromosomal regions). In order to identify key genes for further research, we generated Sankey diagrams, as shown in Figure 7C, to obtain genes enriched in multiple GO terms and KEGG pathways (IDs and corresponding descriptions of GO terms and KEGG pathways can be found in Table S4).

Then, by studying the expression level changes of these key genes in different pathological stages of CML, we found significant differences in the expression of many genes across different stages of CML development. Specifically, *IL1B*, *IL6*, *TRAF6*, and *PRKCA* show significantly higher expression levels in the CML_BP stage than in other stages (Figure 7D), indicating their association with CML disease progression and their potential as specific markers for the late-stage of CML. The extensive previous research can provide support for our analysis. IL-1β is an important member of the IL-1 family and serves as a crucial cytokine involved in various inflammatory processes and certain anti-tumor physiological mechanisms. The increase in *IL1B* levels may promote CML resistance to TKIs by enhancing cell viability and facilitating cell migration [67]. Recent studies have shown that the development of CML is accompanied by elevated levels of *IL1B* [68]. IL-6, a pleiotropic cytokine, plays an important role in some processes associated with leukemia such as immune response and hematopoiesis. *IL6* has been regarded as a valuable prognostic marker in CML and may be correlated with the transformation of different phases in CML because *IL6* expression in CML-BP is significantly upregulated compared with that in CML-CP and CML-AP [69]. TRAF6 is a member of the superfamily of TRAF proteins. It is not only overexpressed in many cancer tissues but also closely associated with tumor cell proliferation, migration, and

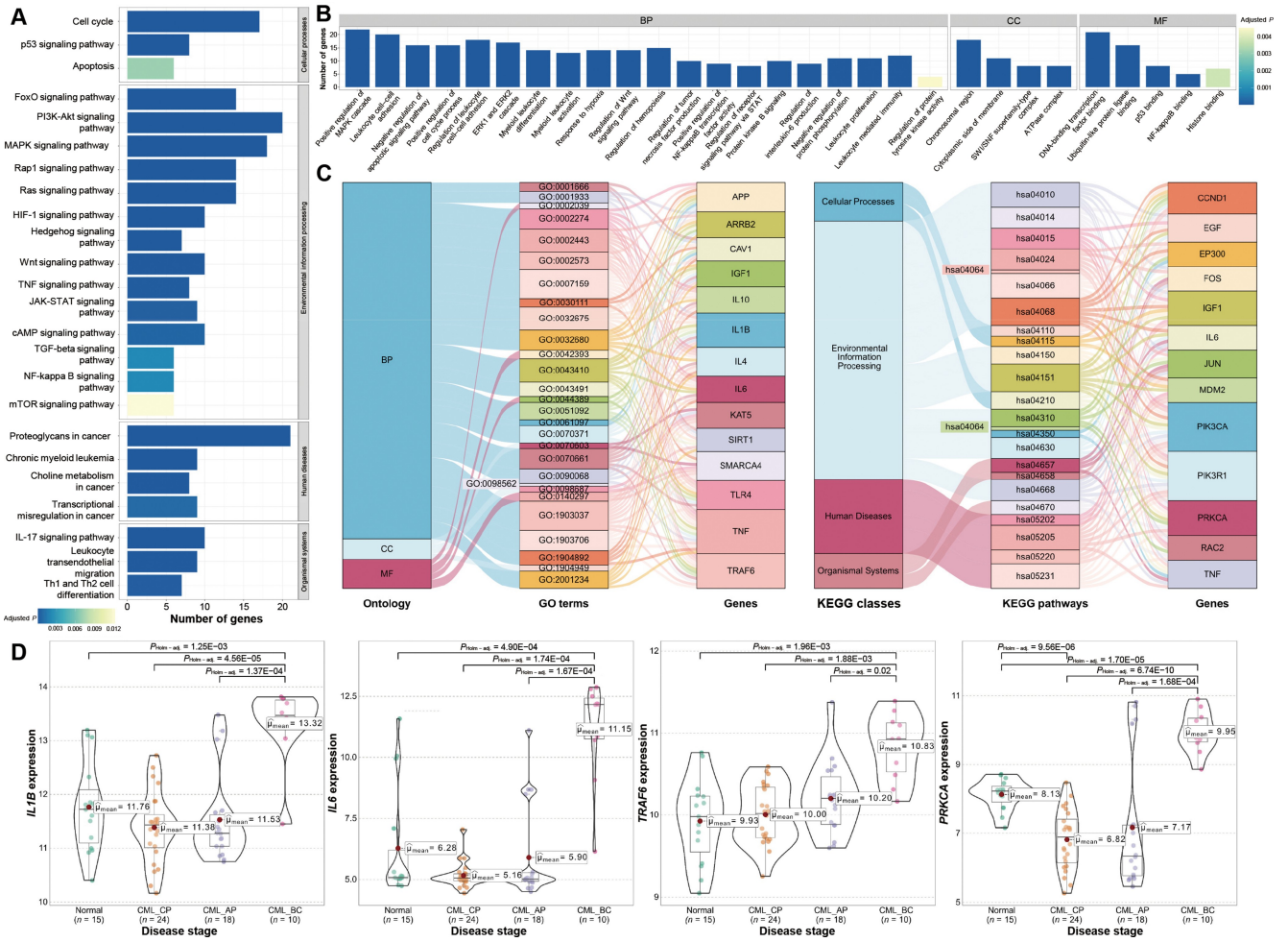


Figure 7 DyNDG predicts CML-related genes closely associated with disease progression
A. The bar chart showcases the results of KEGG pathway enrichment analysis, highlighting the most relevant KEGG pathways associated with CML. **B**. The bar chart illustrates the outcomes of GO enrichment analysis, highlighting the most relevant GO terms associated with CML. **C**. The Sankey diagrams show the associations between GO terms and multi-pathway enriched genes, as well as the associations between KEGG pathways and multi-pathway enriched genes. **D**. The violin plots display the significant differences between CML disease progression stages for *IL1B*, *IL6*, *TRAF6*, and *PRKCA*, respectively. CP, chronic phase; AP, accelerated phase; BP, blast phase.

apoptosis. Studies have shown that interfering with or inhibiting the role of *TRAF6* in tumor-related signaling pathways may provide new therapeutic approaches for cancer treatment [70]. In patients with CML undergoing imatinib treatment, downregulation of *TRAF6* leads to higher levels of cellular apoptosis, indicating the significant regulatory role of *TRAF6* in CML cell survival and treatment response [71]. *PRKCA*, which encodes PKCα, is an important subtype of the protein kinase C (PKC) family. Aberrant regulation of different PKC subtypes is associated with the development of many human diseases. In CML cells, PKCα, as a classical PKC isoform, exhibits significantly reduced kinase activity. This indicates that PKCα may be functionally impaired or inactivated in CML, potentially contributing to abnormal proliferation and pathological progression of CML cells [72].

Genetic dependence analysis in AML, CLL, and CML

To verify the effectiveness of our method DyNDG, we utilized the independent DepMap database [73] and investigated the roles of predicted candidate genes in different types of leukemia cell lines. Specifically, we analyzed three types of leukemia cell lines: AML, CLL, and CML, focusing on the top 10 genes in our predicted candidate gene lists. Notably, the "Gene effect" values for most of these genes are below 0, with some even lower than −1 (**Figure 8A–C**), which suggests that the majority of the top 10 predicted genes are likely essential in the corresponding leukemia cell lines, as knocking out or down these genes can affect leukemia cell proliferation. Almost all of the top 10 predicted genes are associated with at least one type of leukemia cell lines, where their knockout or knockdown shows significant effects. Some of these genes have been implicated in the diagnosis and prognosis of leukemia. For example, it has been shown that high expression of *HSPA8* is often associated with poorer survival rates in AML patients [74]; *HDAC1* can activate driver genes in CLL, thereby promoting the survival and progression of CLL [62]; *EP300* is a promising tumor therapeutic target. The mutations of *EP300* frequently occur in various types of hematologic malignancies, including pediatric ALL (both primary and relapsed), aggressive natural killer-cell leukemia (ANKL), and CML-CP, through diverse mechanisms [75]. These findings not only confirm the efficacy of our predictive method but also highlight the potential of these predicted candidate genes as therapeutic targets. Moreover, we examined the "Gene effect" of known disease genes for AML, CLL, and CML in the DepMap database. As shown in Figure S16A–C, most known disease genes display a significant "Gene effect" on leukemia cell proliferation, while a few show no measurable impact in the available cell line datasets. These findings are consistent with the results of our analysis of the predicted candidate genes. This further confirms the

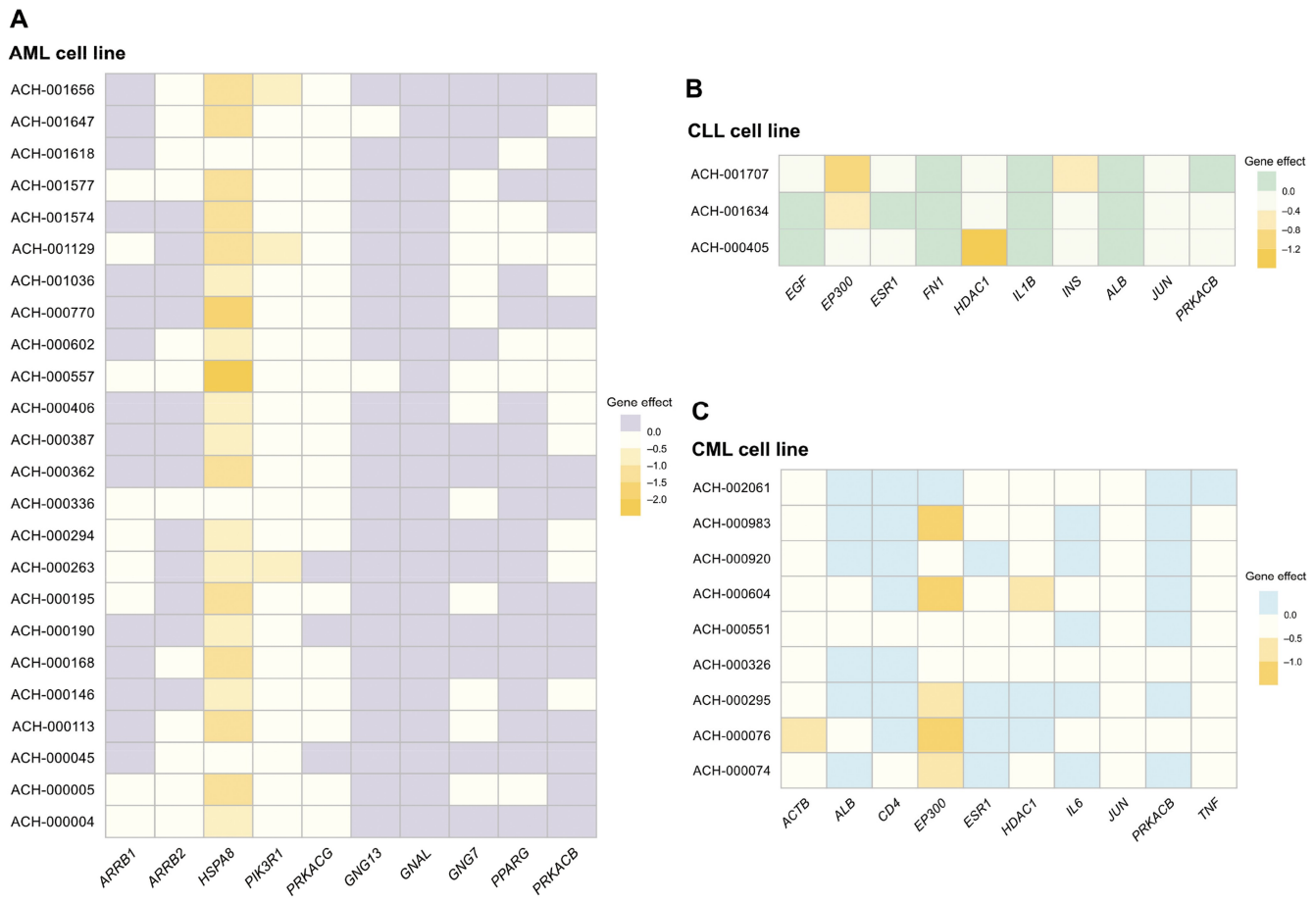


Figure 8 Gene effect profiles of the top 10 predicted candidate genes for three types of leukemia
The heatmaps show the "Gene effect" values for the top 10 DyNDG-predicted candidate genes in AML cell lines (**A**), CLL cell lines (**B**), and CML cell lines (**C**).

potential relevance of our predicted candidate genes to leukemia and validates the effectiveness of our prediction method.

Conclusion

The occurrence and development of blood diseases, such as leukemia, are closely related to gene mutations and abnormal expression. While existing methods for predicting disease genes have made significant contributions, most of them rely on static networks, limiting the improvement of their predictive ability. Addressing how to integrate dynamic information during disease development is a crucial issue that needs to be solved. Here, we proposed the dynamic network-based model called DyNDG for predicting leukemia-related genes, which consists of three main steps: (1) constructing a time-series dynamic network; (2) building a background–temporal multilayer biological network; and (3) applying a novel network propagation approach on the background–temporal multilayer network to generate a ranked list of leukemia-related genes.

The DyNDG model was evaluated on CLL-related, CML-related, and AML-related datasets. Performance was assessed using four metrics: Topk_Precision, Topk_Recall, AUROC, and AUPRC. Comparing to classical single-network and multi-network methods as well as differential gene expression-based methods, DyNDG shows superior performance in predicting leukemia-related genes.

Finally, we processed the DyNDG-predicted candidate genes for AML, CLL, and CML, respectively, by excluding known leukemia-related genes and housekeeping genes, to generate three candidate gene lists for the three types of leukemia. We systematically analyzed the candidate genes for the three types of leukemia from different perspectives by integrating various analysis methods. The results demonstrate that the predicted candidate genes possess significant research value across various biological contexts. These findings further emphasize DyNDG's comprehensive predictive capability, multifunctionality, and robustness in accurately identifying disease genes across multiple types of leukemia.

Code availability

The implementation of DyNDG is available at https://github.com/CSUBioGroup/DyNDG. The code has also been submitted to BioCode at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB) (BioCode: BT007617), which is publicly accessible at https://ngdc.cncb.ac.cn/biocode/tool/BT7617.

Data availability

The web server of DyNDG can be freely accessed at https://csuligroup.com/DyNDG.

CRediT author statement

Jin A: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. **Ju Xiang:** Formal analysis, Methodology, Writing – original draft. **Xiangmao Meng:** Data curation, Formal analysis. **Yue Sheng:** Supervision. **Hongling Peng:** Supervision. **Min Li:** Formal analysis, Writing – review & editing, Methodology, Supervision. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (https://doi.org/10.1093/gpbjnl/qzaf037).

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 62225209) to Min Li, the National Natural Science Foundation of China (Grant No. 62472051) to Ju Xiang, and the High Performance Computing Center of Central South University, China. We thank Dr. Linconghua Wang for her invaluable assistance and inspiring discussions regarding the analysis of experimental results.

ORCID

0009-0002-6258-1093 (Jin A)
0000-0002-3045-5706 (Ju Xiang)
0000-0002-7966-551X (Xiangmao Meng)
0000-0002-9424-123X (Yue Sheng)
0000-0003-2770-5150 (Hongling Peng)
0000-0002-0188-1394 (Min Li)

References

[1] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74:229–63.
[2] Puckett Y, Chan O. Acute lymphocytic leukemia. Treasure Island: StatPearls Publishing; 2022.
[3] Chang A, Schulz PJ, Wenghin Cheong A. Online newspaper framing of non-communicable diseases: comparison of Mainland China, Taiwan, Hong Kong and Macao. Int J Environ Res Public Health 2020;17:5593.
[4] Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. Am J Hum Genet 2001;69:1357–69.
[5] Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–7.
[6] Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. Genetics 2003;164:829–33.
[7] Luo P, Tian LP, Ruan J, Wu FX. Disease gene prediction by integrating PPI networks, clinical RNA-seq data and OMIM data. IEEE/ACM Trans Comput Biol Bioinform 2019;16:222–32.
[8] Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011;12:56–68.
[9] Yue X, Wang Z, Huang J, Parthasarathy S, Moosavinasab S, Huang Y, et al. Graph embedding on biomedical networks: methods, applications and evaluations. Bioinformatics 2020;36:1241–51.

[10] Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol Syst Biol 2008;4:189.
[11] Sun PG, Gao L, Han S. Prediction of human disease-related gene clusters by clustering analysis. Int J Biol Sci 2011;7:61–73.
[12] Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein–protein interactions. J Med Genet 2006;43:691–8.
[13] Lage K, Karlberg EO, Størling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 2007;25:309–16.
[14] Ata SK, Wu M, Fang Y, Ou-Yang L, Kwoh CK, Li XL. Recent advances in network-based methods for disease gene prediction. Brief Bioinform 2021;22:bbaa303.
[15] Hu K, Hu JB, Tang L, Xiang J, Ma JL, Gao YY, et al. Predicting disease-related genes by path structure and community structure in protein–protein networks. J Stat Mech 2018;2018:100001.
[16] Xiang J, Meng X, Zhao Y, Wu FX, Li M. HyMM: hybrid method for disease-gene prediction by integrating multiscale module structure. Brief Bioinform 2022;23:bbac072.
[17] Li Y, Patra JC. Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. Bioinformatics 2010;26:1219–24.
[18] Li Y, Li J. Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data. BMC Genomics 2012;13:S27.
[19] Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet 2008;82:949–58.
[20] Valdeolivas A, Tichit L, Navarro C, Perrin S, Odelin G, Levy N, et al. Random walk with restart on multiplex and heterogeneous biological networks. Bioinformatics 2019;35:497–505.
[21] Zhang Y, Xiang J, Tang L, Li J, Lu Q, Tian G, et al. Identifying breast cancer-related genes based on a novel computational framework involving KEGG pathways and PPI network modularity. Front Genet 2021;12:596794.
[22] Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: it's about time. Brief Bioinform 2010;11:15–29.
[23] Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, et al. Dynamic protein–protein interaction wiring of the human spliceosome. Mol Cell 2012;45:567–80.
[24] Carneiro DG, Clarke T, Davies CC, Bailey D. Identifying novel protein interactions: proteomic methods, optimisation approaches and data analysis pipelines. Methods 2016;95:46–54.
[25] Wang J, Peng X, Peng W, Wu FX. Dynamic protein interaction network construction and applications. Proteomics 2014;14:338–52.
[26] de Lichtenberg U, Jensen LJ, Brunak S, Bork P. Dynamic complex formation during the yeast cell cycle. Science 2005;307:724–7.
[27] Hegde SR, Manimaran P, Mande SC. Dynamic changes in protein functional linkage networks revealed by integration with gene expression data. PLoS Comput Biol 2008;4:e1000237.
[28] Tang X, Wang J, Liu B, Li M, Chen G, Pan Y. A comparison of the functional modules identified from time course and static PPI network data. BMC Bioinformatics 2011;12:339.
[29] Zhang Y, Lin H, Yang Z, Wang J. Construction of dynamic probabilistic protein interaction networks for protein complex identification. BMC Bioinformatics 2016;17:186.
[30] Ou-Yang L, Dai DQ, Li XL, Wu M, Zhang XF, Yang P. Detecting temporal protein complexes from dynamic protein–protein interaction networks. BMC Bioinformatics 2014;15:335.
[31] Lei X, Wang F, Wu FX, Zhang A, Pedrycz W. Protein complex identification through Markov clustering with firefly algorithm on dynamic protein–protein interaction networks. Information Sciences 2016;329:303–16.
[32] Xiao Q, Wang J, Peng X, Wu FX, Pan Y. Identifying essential proteins from active PPI networks constructed with dynamic gene expression. BMC Genomics 2015;16:S1.
[33] Li M, Chen X, Ni P, Wang J, Pan Y. Identifying essential proteins by purifying protein interaction networks. 12th International Symposium on Bioinformatics Research and Applications 2016;9683:106–16.
[34] Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep 2012;2:342.
[35] Liu R, Wang X, Aihara K, Chen L. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev 2014;34:455–78.
[36] Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009;27:199–204.
[37] Faisal FE, Milenković T. Dynamic networks reveal key players in aging. Bioinformatics 2014;30:1721–9.
[38] Chennamadhavuni A, Lyengar V, Shimanovsky A. Leukemia. Treasure Island: StatPearls Publishing; 2022.
[39] PDQ Adult Treatment Editorial Board. Acute Myeloid Leukemia Treatment (PDQ®)–Health Professional Version. [Internet]. Bethesda, MD: National Cancer Institute (US); 2022, https://www.cancer.gov/types/leukemia/hp/adult-aml-treatment-pdq.
[40] PDQ Adult Treatment Editorial Board. Chronic Lymphocytic Leukemia Treatment (PDQ®)–Patient Version. [Internet]. Bethesda, MD: National Cancer Institute (US); 2022, https://www.cancer.gov/types/leukemia/patient/cll-treatment-pdq.
[41] PDQ Adult Treatment Editorial Board. Chronic Myeloid Leukemia Treatment (PDQ®)–Patient Version. [Internet]. Bethesda, MD: National Cancer Institute (US); 2022, https://www.cancer.gov/types/leukemia/patient/cml-treatment-pdq.
[42] Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res 2022;51:D638–46.
[43] Kim CY, Baek S, Cha J, Yang S, Kim E, Marcotte EM, et al. HumanNet v3: an improved database of human gene networks for disease research. Nucleic Acids Res 2022;50:D632–9.
[44] Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res 2017;45:D877–87.
[45] Seal RL, Braschi B, Gray K, Jones TE, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res 2023;51:D1003–9.
[46] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[47] Li Y, Patra JC. Integration of multiple data sources to prioritize candidate genes using discounted rating system. BMC Bioinformatics 2010;11:S20.
[48] Ruan P, Wang S. DiSNEP: a disease-specific gene network enhancement to improve prioritizing candidate disease genes. Brief Bioinform 2021;22:bbaa241.
[49] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation (Camb) 2021;2:100141.
[50] Hounkpe BW, Chenou F, de Lima F, De Paula EV. HRT Atlas v1. 0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res 2021;49:D947–55.
[51] Li K, Du Y, Cai Y, Liu W, Lv Y, Huang B, et al. Single-cell analysis reveals the chemotherapy-induced cellular reprogramming and novel therapeutic targets in relapsed/refractory acute myeloid leukemia. Leukemia 2023;37:308–25.
[52] Bruserud Ø, Reikvam H, Brenner AK. Toll-like receptor 4, osteoblasts and leukemogenesis; the lesson from acute myeloid leukemia. Molecules 2022;27:735.
[53] He M, Cai X, Zeng Y. MAPK signaling pathway was dysregulated in acute myeloid leukemia with *RUNX1* mutations. Res Sq 2021; https://www.researchsquare.com/article/rs-502740/v1.

[54] Hsiao PC, Hsieh YH, Chow JM, Yang SF, Hsiao M, Hua KT, et al. Hispolon induces apoptosis through JNK1/2-mediated activation of a caspase-8,-9, and-3-dependent pathway in acute myeloid leukemia (AML) cells and inhibits AML xenograft tumor growth *in vivo*. J Agric Food Chem 2013;61:10063–73.
[55] Su YC, Li SC, Wu YC, Wang LM, Chao KS, Liao HF. Resveratrol downregulates interleukin-6-stimulated sonic hedgehog signaling in human acute myeloid leukemia. Evid Based Complement Alternat Med 2013;2013:547430.
[56] Lu FL, Yu CC, Chiu HH, Liu HE, Chen SY, Lin S, et al. Sonic hedgehog antagonists induce cell death in acute myeloid leukemia cells with the presence of lipopolysaccharides, tumor necrosis factor-α, or interferons. Invest New Drugs 2013;31:823–32.
[57] Ibanez M, Carbonell-Caballero J, Such E, Garcia-Alonso L, Liquori A, Lopez-Pavia M, et al. The modular network structure of the mutational landscape of acute myeloid leukemia. PLoS One 2018;13:e0202926.
[58] Darici S, Alkhaldi H, Horne G, Jørgensen HG, Marmiroli S, Huang X. Targeting PI3K/Akt/mTOR in AML: rationale and clinical evidence. J Clin Med 2020;9:2934.
[59] Patel K, Pagel JM. Current and future treatment strategies in chronic lymphocytic leukemia. J Hematol Oncol 2021;14:69.
[60] 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;13:2498–504.
[61] Wang JC, Kafeel MI, Avezbakiyev B, Chen C, Sun Y, Rathnasabapathy C, et al. Histone deacetylase in chronic lymphocytic leukemia. Oncology 2012;81:325–9.
[62] Lai TH, Ozer HG, Gasparini P, Nigita G, Distefano R, Yu L, et al. HDAC1 regulates the chromatin landscape to control transcriptional dependencies in chronic lymphocytic leukemia. Blood Adv 2023;7:2897–911.
[63] Byrd JC, Woyach JA, Johnson AJ. Translating PI3K-delta inhibitors to the clinic in chronic lymphocytic leukemia: the story of CAL-101 (GS1101). Am Soc Clin Oncol Educ Book 2012;32:691–4.
[64] Durand-Onaylı V, Haslauer T, Härzschel A, Hartmann TN. Rac GTPases in hematological malignancies. Int J Mol Sci 2018;19:4041.
[65] Shaffer AL 3rd, Phelan JD, Wang JQ, Huang D, Wright GW, Kasbekar M, et al. Overcoming acquired epigenetic resistance to BTK inhibitors. Blood Cancer Discov 2021;2:630–47.
[66] Minciacchi VR, Kumar R, Krause DS. Chronic myeloid leukemia: a model disease of the past, present and future. Cells 2021;10:117.
[67] Lee CR, Kang JA, Kim HE, Choi Y, Yang T, Park SG. Secretion of IL-1β from imatinib-resistant chronic myeloid leukemia cells contributes to *BCR–ABL* mutation-independent imatinib resistance. FEBS Lett 2016;590:358–68.
[68] Bütow M, Testaquadra FJ, Baumeister J, Maié T, Chatain N, Jaquet T, et al. Targeting cytokine-induced leukemic stem cell persistence in chronic myeloid leukemia by IKK2-inhibition. Haematologica 2023;108:1179–85.
[69] Sharma K, Singh U, Rai M, Shukla J, Gupta V, Narayan G, et al. Interleukin 6 and disease transformation in chronic myeloid leukemia: a Northeast Indian population study. J Cancer Res Ther 2020;16:30–3.
[70] LI M, ZHOU LJ. Research progress on the relationship between TRAF6 and tumor. Biotechnology Bulletin 2017;33:24–31.
[71] Han SH, Korm S, Han YG, Choi SY, Kim SH, Chung HJ, et al. GCA links TRAF6-ULK1-dependent autophagy activation in resistant chronic myeloid leukemia. Autophagy 2019;15:2076–90.
[72] Mencalha AL, Corrêa S, Abdelhay E. Role of calcium-dependent protein kinases in chronic myeloid leukemia: combined effects of PKC and BCR–ABL signaling on cellular alterations during leukemia development. Onco Targets Ther 2014;7:1247–54.
[73] Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell 2017;170:564–76.e16.
[74] Li J, Ge Z. High HSPA8 expression predicts adverse outcomes of acute myeloid leukemia. BMC Cancer 2021;21:475.
[75] Zhu Y, Wang Z, Li Y, Peng H, Liu J, Zhang J, et al. The role of CREBBP/EP300 and its therapeutic implications in hematological malignancies. Cancers 2023;15:1219.

Genomics, Proteomics & Bioinformatics, 2025, 23, 1–17
https://doi.org/10.1093/gpbjnl/qzaf037
Method