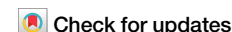**Article**

A multi-omics approach for biomarker discovery in neuroblastoma: a network-based framework

Check for updates

Rahma Hussein[1,3], Ahmed M. Abou-Shanab [1,3] & Eman Badr [1,2] ✉

Neuroblastoma (NB) is one of the leading causes of cancer-associated death in children. MYCN amplification is a prominent genetic marker for NB, and its targeting to halt NB progression is difficult to achieve. Therefore, an in-depth understanding of the molecular interactome of NB is needed to improve treatment outcomes. Analysis of NB multi-omics unravels valuable insight into the interplay between MYCN transcriptional and miRNA post-transcriptional modulation. Moreover, it aids in the identification of various miRNAs that participate in NB development and progression. This study proposes an integrated computational framework with three levels of high-throughput NB data (mRNA-seq, miRNA-seq, and methylation array). Similarity Network Fusion (SNF) and ranked SNF methods were utilized to identify essential genes and miRNAs. The specified genes included both miRNA-target genes and transcription factors (TFs). The interactions between TFs and miRNAs and between miRNAs and their target genes were retrieved where a regulatory network was developed. Finally, an interaction network-based analysis was performed to identify candidate biomarkers. The candidate biomarkers were further analyzed for their potential use in prognosis and diagnosis. The candidate biomarkers included three TFs and seven miRNAs. Four biomarkers have been previously studied and tested in NB, while the remaining identified biomarkers have known roles in other types of cancer. Although the specific molecular role is yet to be addressed, most identified biomarkers possess evidence of involvement in NB tumorigenesis. Analyzing cellular interactome to identify potential biomarkers is a promising approach that can contribute to optimizing efficient therapeutic regimens to target NB vulnerabilities.

Neuroblastoma (NB) is one of the most common solid malignant cancers in new births and has a sympathoadrenal origin[1,2]. NB patients are mostly diagnosed before age 10, and the age at which diagnosis occurs is inversely related to treatment outcomes[3,4]. Distinguishable from other pediatric tumors, infants with low-risk tumors have an 88% survival rate, which has been improved through advances in the cancer biology field. Moreover, NB encounters spontaneous regression or differentiates into benign ganglioneuromas[5,6]. Contrarily, the survival rate for older children, specifically those between 18 months and 12 years, is ~49%. However, in young adults (above 12 years), the survival rate plunges below 10% due to the predominance of extensive hematogenous metastasis as NB progresses[7]. Hence, this increases death cases despite the continuous multimodal

therapeutics[8,9]. The clinical treatment and presentation of high-risk NB bring on relapse and a refractory state after significant responses to the initial chemotherapy.

MYCN gene amplification is the best-known prognostic marker of NB[10]. It has been established to antagonize several oncosuppressive mRNAs, like p53[11], and miRNAs, such as miR-184[12], indicating that MYCN exerts both transcriptional and post-transcriptional interaction on its targets. Although MYCN is the most crucial target for NB therapy, its direct targeting is challenging because of its pleiotropic effect and the difficulty of blocking transcription factors[13]. As an alternative approach, drugs could be designed to inactivate MYCN partners or transcriptional targets[14]. Consequently, the construction of a comprehensive MYCN regulatory network of

---

[1]Biomedical Sciences Program, University of Science and Technology, Zewail City of Science and Technology, Giza 12578, Egypt. [2]Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt. [3]These authors contributed equally: Rahma Hussein, Ahmed M. Abou-Shanab.
✉e-mail: emostafa@zewailcity.edu.eg
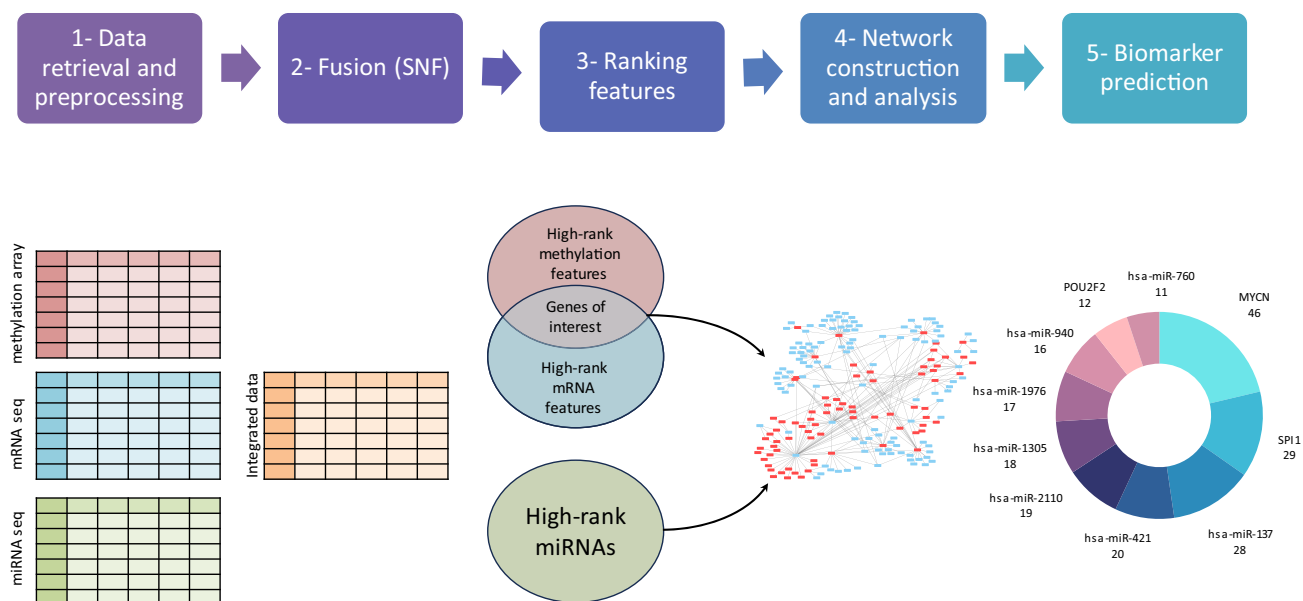
Fig. 1 | **The proposed workflow.** A block diagram for predicting NB candidate biomarkers.

various regulatory interaction types is needed. Overall, further investigations on the whole NB cell interactome orchestration are required to unravel candidate biomarkers that could be targeted to prevent NB tumorigenesis.

High-throughput technologies have enabled the global analysis of biological molecules, yielding vast amounts of omics data yearly[15]. Having multi-omics datasets for the same cohort of patients provides insight into biological interactions at different levels. It can lead to a better understanding of the molecular mechanism network that primes disease development, such as cancers. A characteristic of high-throughput multi-omics data is high dimensionality. For instance, a typical RNA-seq experiment may produce expression data for tens of thousands of transcripts. Machine learning methods help identify essential features in high-dimensional datasets and predict potential disease biomarkers. Biomarker identification contributes to understanding the underlying biology of the disease, such as identifying associated driver genes or finding potential drug targets[16].

Several studies have utilized machine learning techniques for studying NB. A deep neural network used NB gene expression data to classify patients according to their International Neuroblastoma Staging System (INSS) stage[17]. A Concatenated Diagnostic-Relapse Prognostic deep learning model was used for NB survival prediction. The model consisted of an autoencoder coupled with a multi-task classifier. It was trained on NB transcription data to predict the Overall Survival (OS) and Event-Free Survival (EFS)[18]. In[19], copy number alteration data was incorporated with expression data. A deep learning autoencoder combined with k-means clustering was utilized. The model identified two high-risk NB subtypes, each showing distinct survival outcomes. The authors in ref. 20 developed an integrative network fusion for end-point prediction of NB. It was used to analyze expression data and copy number data. Few reports tackled integrative omics approaches in identifying effectors of MYCN, critical regulators, and potential therapeutic targets in the NB[21,22]. Similarity network fusion (SNF) was used to integrate NB multi-omics data. A deep neural network and recursive feature elimination were then used on the combined data to predict patient survival. SNF achieved better integration for multi-omics data than feature-level fusion. It also showed proficiency in managing data heterogeneity and high dimensionality[23]. In another NB multi-omics study, epigenomic profiling using hmTOP-seq and uTOP-seq was performed. Transcriptomic profiling was conducted using mRNA-seq. Integrative epigenomic and transcriptomic data analysis was used to discover different cell signatures attributed to other NB cell subpopulations[24]. An NB metastasis signature consisting of 18 genes was identified through analysis of NB multi-omics. RNA-seq and copy number variation data for primary and recurrent tumors were compared. Survival analysis was utilized to identify prognostic biomarker genes[25].

Here, we aim to identify NB biomarkers utilizing a multi-omics data integrative approach. Data describing NB's molecular nature at the DNA methylation, mRNA, and miRNA levels were analyzed. Ranked SNF was utilized to select essential genes and miRNAs relevant to NB. Finally, network analysis was employed to identify their interaction and suggest potential biomarkers. The identified candidate biomarkers were further analyzed for their potential use as prognostic and diagnostic markers.

Results
Overview of the proposed framework
Multi-omics data were obtained for 99 patients. The data types included were mRNA-seq, miRNA-seq, and methylation array data. Each data type was utilized to construct a patient similarity matrix. The three similarity matrices were then integrated using the SNF technique, producing a single fused similarity matrix. The Ranked Similarity Network Fusion method (rSNF) was then used to rank the features of each data type. The top 10% of high-rank features from all data types were first filtered to determine candidate biomarkers. Then, the common genes from methylation and mRNA-seq data were considered essential genes. Next, the interactions between the identified set of essential genes and the previously determined high-rank miRNAs were explored by building an interaction network. To create the network, TF-miRNA and miRNA-target interactions between the identified genes and the filtered miRNAs were retrieved from public experimental databases and integrated to construct a regulatory network. Finally, the maximal clique centrality (MCC) was performed to identify hub nodes as potential biomarkers (Fig. 1).

Parameter tuning and data integration
Different values have been evaluated iteratively for SNF hyperparameters (T, k, α) to choose the best values for SNF convergence. Figure 2 illustrates the relative change between fused graphs obtained in consecutive iterations. For the T parameter, each iteration is present using one more fusion iteration step where $(T_{i+1} = T_i + 1)$, as shown in Fig. 2 (a,b), shows the effect of utilizing an increasing set of nearest neighbors $(k_{i+1} = k_i + 1)$. Different values of α have also been evaluated (Fig. 2c). As shown in Fig. 2, $T = 15$, $k = 20$, and $\alpha = 0.5$ were sufficient for convergence. Spectral clustering was performed on the fused graph using 2:7 clusters. The number of clusters was set to $c = 4$, as Fig. 3 shows, which resulted in the highest ratio between minimum intra-cluster similarity and maximum inter-cluster similarity.
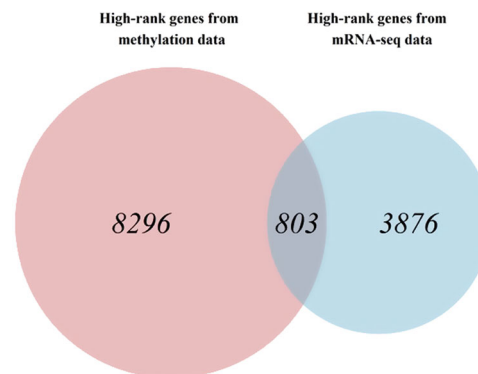
a

b

c
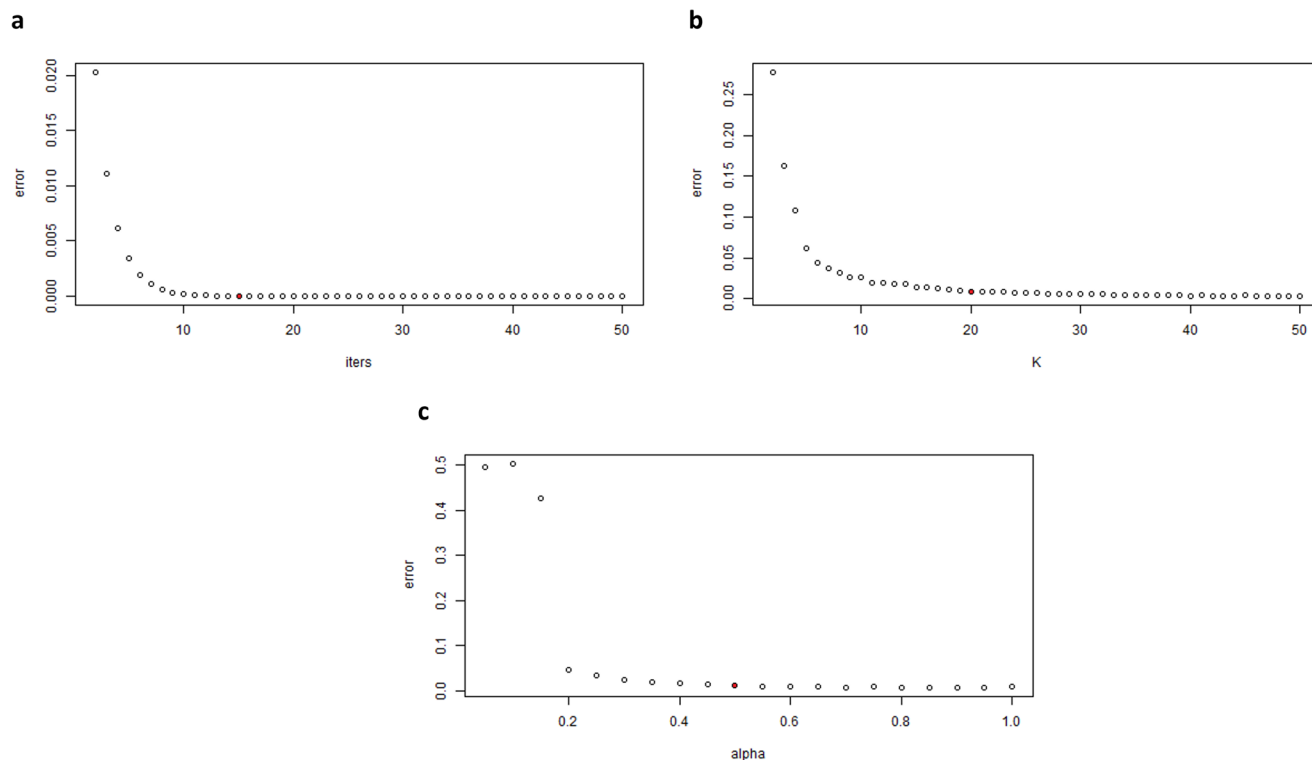
Fig. 2 | Convergence of SNF using different hyperparameters. The selected values sufficient for convergence are highlighted in red. **a** Number of iterations (T). **b** Number of nearest neighbors (k). **c** The alpha (α) hyperparameter (used in calculating the similarity matrix from the distance matrix). $T = 15$, $k = 20$, and $\alpha = 0.5$ were sufficient for convergence.
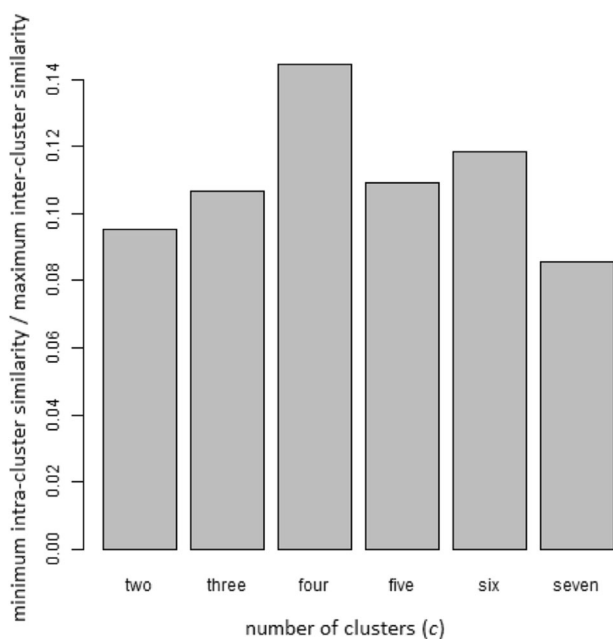
Fig. 3 | Evaluating clustering. The vertical axis corresponds to the ratio between minimum intra-cluster similarity and the maximum inter-cluster similarity. The horizontal axis shows the number of clusters used.

Fig. 4 | Identified essential genes. Shared genes between highly ranked genes from the methylation array and mRNA-seq data.

Feature selection

SNF was used to integrate the three similarity matrices into one fused similarity matrix. Then, rSNF was used to assign ranks to all features. The features were ordered by rank for each data type, and the top 10% of the features were selected. As a result, 37,953 high-rank CpG sites from the methylation array data, of which 67.8% mapped to 9099 genes, 4679 unique high-rank genes from the mRNA-seq data, and 160 high-rank miRNAs from the miRNA-seq data were selected. On comparing high-rank genes from the methylation array and mRNA-seq data, 803 genes were shared between both groups identified as essential genes (Fig. 4).

Regulatory network construction

TF-miRNA and miRNA-target interactions that involve a high-rank miRNA and essential genes were retrieved from the transmir 2.0 and Tarbase v8 databases. The genes encoding for transcription factors were identified from the transmir 2.0 database Field, where 255 unique TF-miRNA interactions were obtained, and 161 unique miRNA-target interactions were retrieved from Tarbase v8[26].

Integrating all the interactions to construct a regulatory network resulted in 90 miRNAs, 23 TFs, and 199 target genes. The network is then
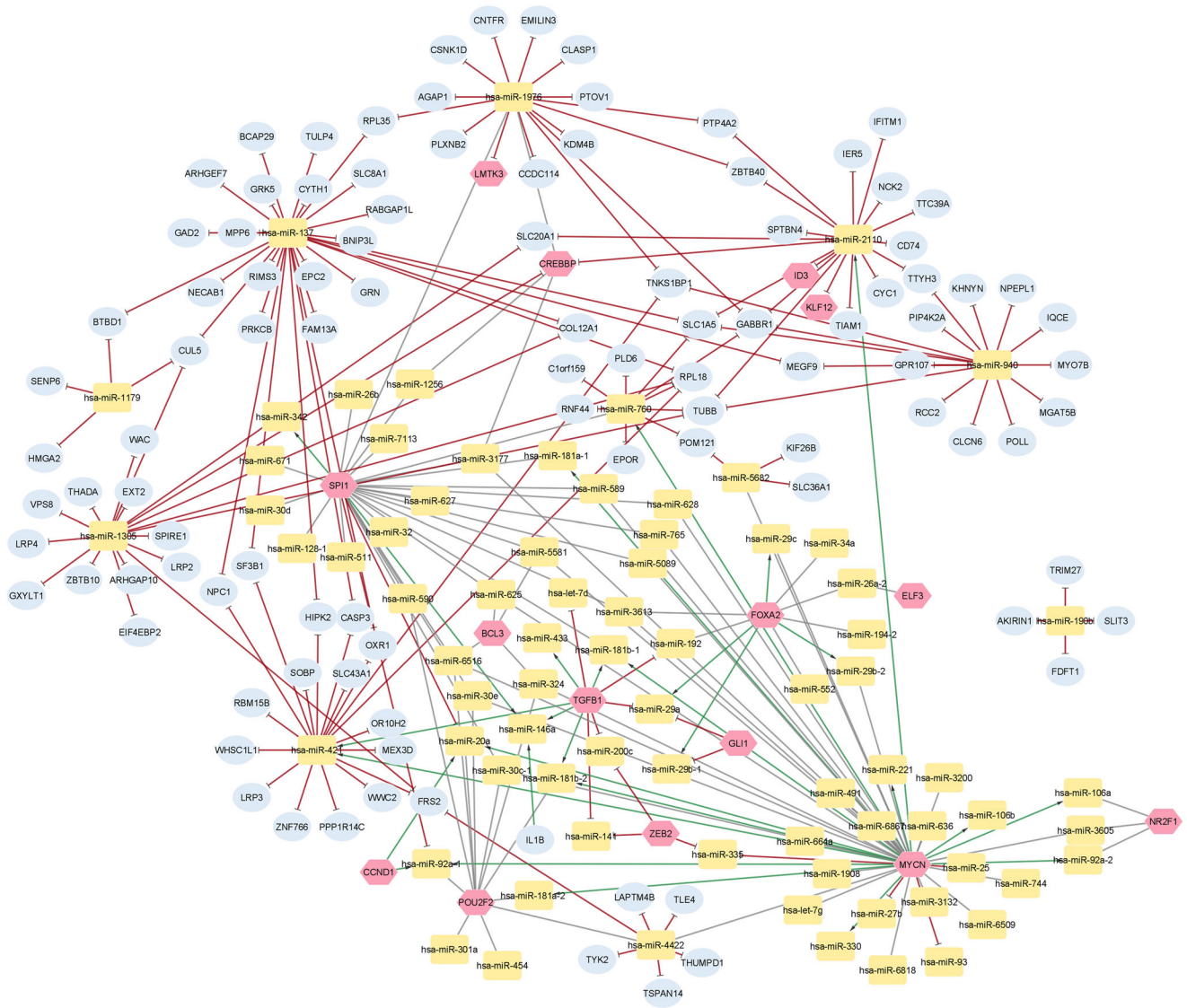
Fig. 5 | TFs-gene-miRNAs regulatory network; consists of 195 nodes and 265 edges. Node shape and color are labeled as pink hexagons representing TFs, blue ellipses representing genes, and yellow round squares representing miRNAs. The nodes have directed and colored edges: the delta-shaped green arrows indicate upregulation, while the T-shaped red arrows indicate downregulation. Black-colored edges indicate regulation.

visualized using Cytoscape[27] (Fig. 5). Utilizing the MCC algorithm[28], the top 10 hub nodes were identified and ranked, as shown in Table 1. They included three transcription factors and seven miRNAs in the regulatory network and were considered potential biomarkers for targeting NB. MYCN gene is shown to be the highest-ranked node. All the identified candidate biomarkers are implicated in NB, such as MYCN or other tumors, including breast cancers and glioma. The interaction network of the top 10 hub nodes, ranked according to the MCC score, is provided (Supplementary Fig. 1).

Table 1 | MCC score-based ranking of the hub 10 nodes in the regulatory network

| Node	Type	MCC score
MYCN	TF	46
SPI1	TF	29
Hsa-miR-137	miRNA	28
Hsa-miR-421	miRNA	20
Hsa-miR-2110	miRNA	19
Hsa-miR-1305	miRNA	18
Hsa-miR-1976	miRNA	17
Hsa-miR-940	miRNA	16
POU2F2	TF	12
Hsa-miR-760	miRNA	11

Validation of the identified potential biomarkers

Survival analysis was conducted to study the correlation between the expression of the identified hub nodes and the prognosis of NB patients. Patients were split into low- and high-expression groups based on the cutoff mean ± 0.25 * standard deviation. Nodes with p value < 0.05 are considered statistically significant. Table 2 illustrates the p values associated with the Kaplan–Meier analysis. MYCN, POU2F2, and SPI1 transcription factors demonstrated significant association with survival information, as illustrated in Fig. 6. No miRNAs have shown significant differences between the low- and high-expression groups (p value > 0.05). GSE62564 is an external dataset of 498 NB patients that has been utilized as a validation dataset. The three transcription factors maintained their significant association with survival information, which suggests their prognostic potential, as shown in

Table 2 | *P* values associated with the Kaplan–Meier curve using log-rank statistical and Chi-square tests

Nodes	MYCN	SPI1	POU2F2	hsa-mir-137	hsa-mir-421	hsa-mir-2110	hsa-mir-1305	hsa-mir-1976	hsa-mir-940	hsa-mir-760
OS *p* value (TARGET)	**0.022**	**0.025**	**0.036**	0.197	0.35	0.072	0.471	0.582	0.66	0.97
OS *p* value (GSE62564)	**1.48E-11**	**0.000495**	**1.46E-05**	**7.14E-06**	**0.001**	0.206	0.917	0.091	0.76	**0.046**
Chi-square *p* value	**0.0290**	0.999	**0.014**	**0.006**	0.316	0.894	0.271	**0.037**	**0.011**	**0.009**

Nodes with *p* value < 0.05 are highlighted in bold.
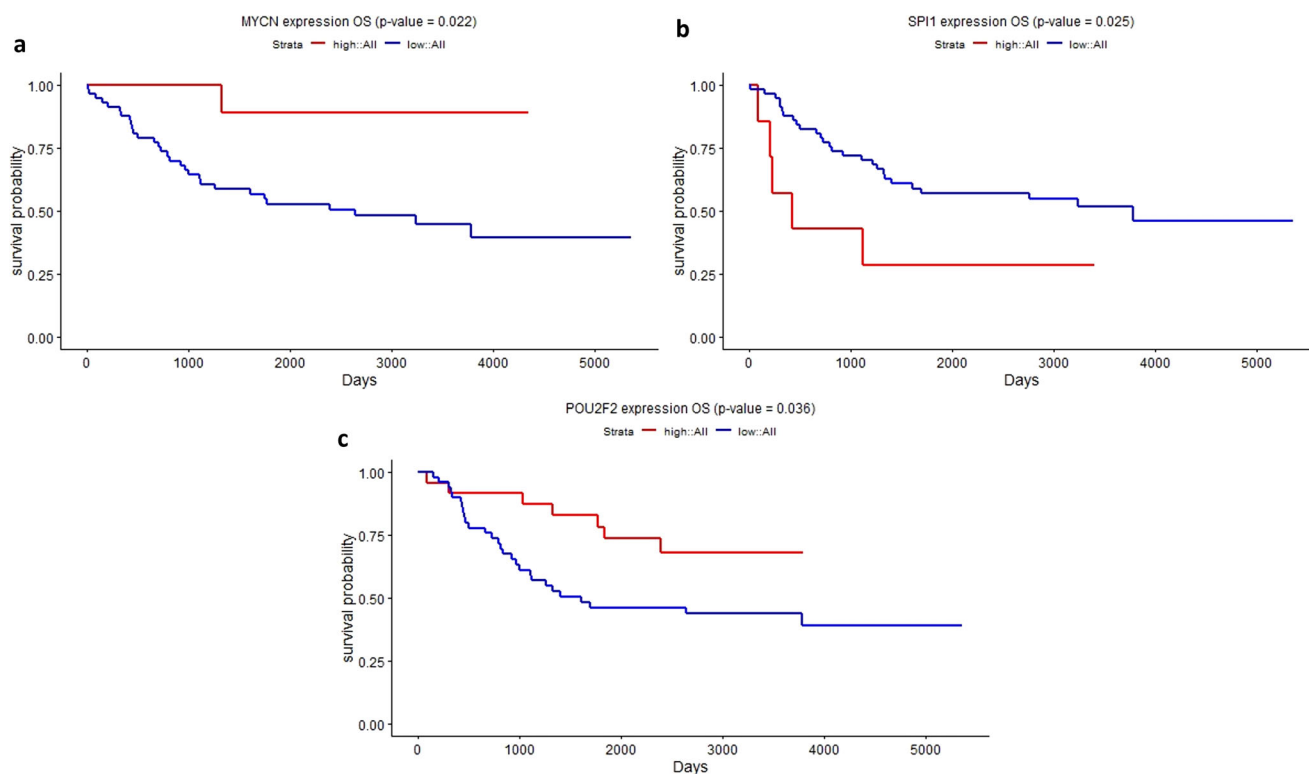
Fig. 6 | Kaplan–Meier curves for the three TFs in NB patients. a MYCN, **b** SPI1, and **c** POU2F2.

Fig. 7a–c. Three more miRNAs hsa-mir-137, hsa-mir-421, and hsa-mir-760 showed significant association with *p* values of 7.14E-06,0.001, 0.046, respectively, as illustrated in Fig. 7d–f. Table 2 illustrates the *p* values associated with the Kaplan–Meier analysis.

Moreover, NB patient stages were retrieved from the TARGET dataset, and a Chi-square test was conducted to evaluate the association between the identified candidate biomarker expressions and NB stages. A *p* value of <0.05 was considered statistically significant (Table 2).

Expression profiles of miRNAs were utilized from GSE128004. The dataset consists of fifteen NB patients and three control samples. We evaluated the proposed miRNAs for their diagnostic potential. Six out of the seven miRNAs were retrieved from the dataset. hsa-mir-421 and hsa-mir-760 demonstrated a high potential to discriminate the patient group from controls with the area under the curve (AUC) values > 0.7, as shown in Fig. 8a, b. On the other hand, hsa-mir-2110 had an AUC value of 0.556 (Fig. 8c). All AUC values are listed in Table 3.

Discussion

Biomarkers identification for complex diseases, including cancers, is of paramount importance in its diagnosis, prognosis, and treatment. Numerous frameworks have been proposed for NB biomarkers prediction, albeit only a few are from the perspective of utilizing multi-omics to wire transcription factors-genes-miRNAs regulatory network. Herein, we conducted an *in-silico* analysis of the NB multi-omics dataset, including mRNA,

miRNA, and methylation data retrieval, integration, regulatory network construction, and analysis to identify candidate biomarkers for NB. In the constructed network shown in Fig. 5, the MCC node ranking method was utilized to determine the hub nodes (Table 1). Multiple analyses have been conducted to validate the obtained results.

Among the identified candidates, the **MYCN** gene had the highest degree of 46 and is significantly correlated with NB patient survival and NB stage. Also, its interactions with the network are shown in Supplementary Fig. 2. MYCN was found to be an amplified homolog to v-MYC but different from MYC in the human NB[29]. As illustrated, MYCN activates other hub nodes, including hsa-miR-760, hsa-miR-421, and hsa-miR-2110. For example, the upregulated hsa-miR-760 regulates another hub TF, SPI1 (Fig. 5), which in turn upregulates another identified hub mirRNA, hsa-miR-1976 (Fig. 5). These interactions between different hub nodes starting from MYCN could be implicated in NB tumor progression. Thus, its targeting could aid in inhibiting the tumorigenicity of NB. MYCN also suppresses ATM expression via its upregulation of hsa-miR-421 (Fig. 5), which promotes neuroendocrine prostate cancer cells' invasion and migration potential[30,31]. MYCN was found to stimulate rat fibroblast cells' S phase entry and their transformation into a more proliferative form with rapid cell-cycle progression. Similar results were observed during experimentation on growth factors-deprived quiescent fibroblasts[32–34]. As shown in Fig. 5, hsa-miR-181-a, and -b are activated by MYCN,
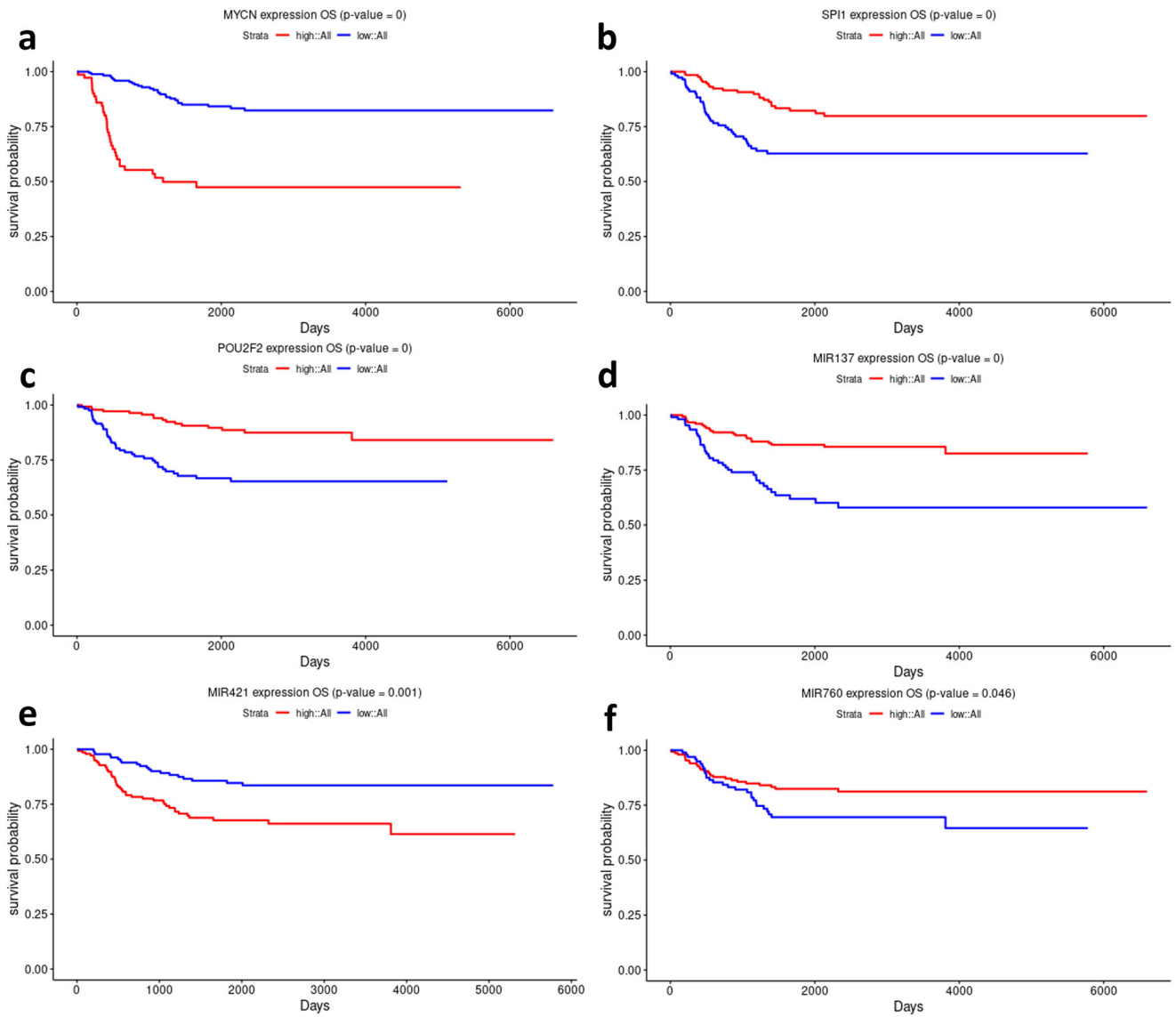
Fig. 7 | Kaplan–Meier curves of candidate biomarkers with statistical significance in the external dataset (GSE62564). a MYCN, **b** SPI1, and **c** POU2F2, **d** hsa-mir-137, **e** hsa-mir-421, and **f** hsa-mir-760.
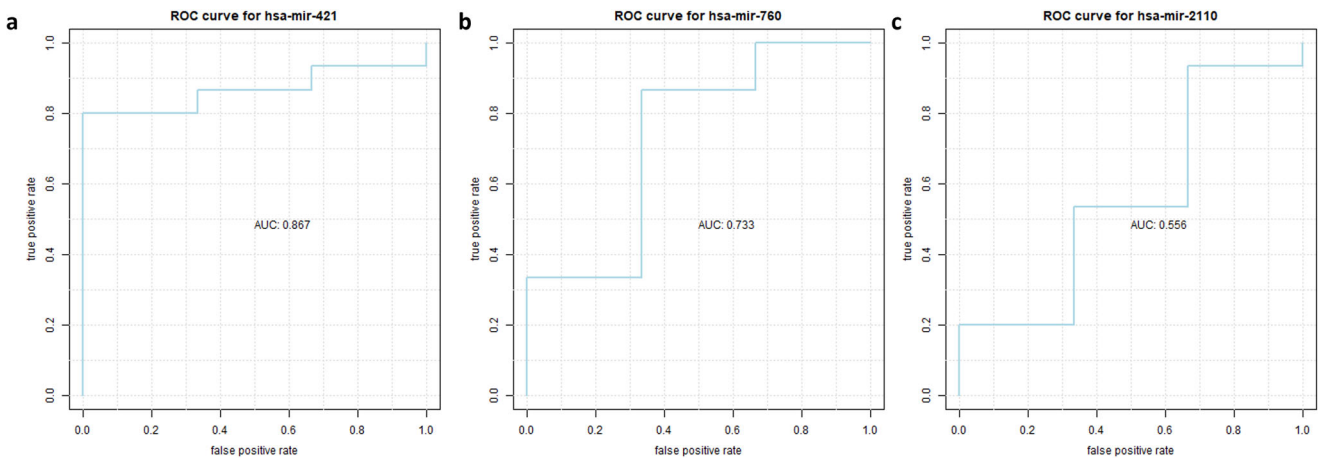
Fig. 8 | AUC values for the top three miRNAs. a hsa-mir-421, **b** hsa-mir-780, **c** hsa-mir-2110.

Table 3 | AUC values for candidate miRNAs

Candidate miRNAs	hsa-mir-137	hsa-mir-421	hsa-mir-2110	hsa-mir-1305	hsa-mir-1976	hsa-mir-940	hsa-mir-760
AUC value	0.444	0.867	0.556	Not available	0.356	0.4	0.733

which induces NB tumorigenesis[35]. Furthermore, it suppresses NB invasion by directly targeting non-canonical TGF-beta pathway genes[36,37]. MYCN regulates various cellular and molecular processes that support NB tumorigenicity and biomass and its expression is the highest in immature cells, which decreases the differentiation capacity of these cells. This is consistent with the NB case, as NB cell differentiation is accompanied by inhibition of MYCN expression[38–44] and NB cells' self-renewal and maintenance of their pluripotency properties[45–47]. Notably, NB invasion and metastasis behavior depend on the MYCN state[48,49]. Interestingly, priming of NB detachment from the extracellular matrix (ECM) for metastasis and invasion is achieved by MYCN stimulation to block integrins α1 and β1 expression[50,51]. Also, MYCN directs NB degradation of their ECM, metastasis, adhesion, and invasive behaviors via promoting focal adhesion kinase (FAK), which is responsible for integrins signaling, to stimulate migration and metastasis in tumor cells[52,53]. Intriguingly, P53 blocks the transcription of FAK, which emphasizes the role of MYCN cascades to inhibit p53[53].

SPI1 has an MCC score of 29 (Table 1) and is correlated with NB patient survival (Table 2). Its regulatory network shows its direct activation to hsa-miR-146a and hsa-miR-342, a repressive effect on hsa-miR-92a-1 and hsa-miR-20a, as shown in Fig. 5). SPI1 (encoding PU.1 protein) is one of the Ets transcription factors family. It is known to regulate multiple miRNAs in alignment with Supplementary Fig. 3[54]. The Ets family regulates various cellular processes, including apoptosis, proliferation, differentiation, angiogenesis, lymphoid cell development, and invasiveness[55–59]. Mice lacking PU.1 expression have neither lymphocyte, myeloid cells, nor their progenitor cells[60]. Moreover, the loss of PU.1 expression is implicated in the loss of cellular communication, leading to leukemia[61]. Also, minimal reduction in PU.1 expression primes a preleukemia state, thus promoting the development of acute myeloid leukemia[62]. It has been recently reported that PU.1 antagonizes the expression of the P53 tumor suppressor protein, thus blocking the expression of the cell cycle and apoptosis regulatory proteins activation[63]. Furthermore, the involvement of PU.1 protein in accelerating the speed of the replication elongation during the S phase of the cell cycle coinciding with an accumulation of genetic mutations and no DNA breakage in leukemia has also been reported[64]. Moreover, the PU.1 protein has been identified to promote glioma proliferation and metastasis[65] via the regulation of Phosphoribosylaminoimidazole Carboxylase and Phosphoribosylaminoimidazolesuccinocarboxamide Synthase (PAICS) that has been recently reported to be implicated in glioma[66], along with other malignancies development and progression[67–72]. A study reported that the expression of PU.1 protein is active in NB cells[73]. To the best of our knowledge, the involvement of PU.1 protein in NB development and progression has not been reported yet.

POU2F2 (Oct-2) transcription factor was found to be correlated with NB patient survival and tumor stage in our analysis (Table 2) with an MCC score of 12 (Table 1). It regulates a set of miRNAs that may have a role in NB tumorigenesis (Supplementary Fig. 4). From a biological aspect, POU2F2 expression in various cancers is shown to be upregulated and consequently was proven to regulate various processes that aid cancer development. In glioblastoma, it has been reported to regulate the metabolic shift from oxidative phosphorylation to glycolysis (The Warburg effect fundamental theory[74]) via PI3K/Akt/mTOR pathway promoting tumor cells proliferation, growth, and survival[75]. It promotes lung cancer's proliferation and mesenchymal characteristics via upregulating AGO1[76]. In NB, POU2F2 is responsible for cell development and proliferation and was reported in previous bioinformatics analyses to be hypermethylated, which drives its downregulation[77,78]. Still, more in vitro and animal model reports should be

conducted to elucidate the role of the POU2F2 transcription factor in regulating NB development and progression.

Numerous reports have verified the role of microRNA networks in regulating post-transcriptional modifications of gene expression transcripts; thus, they are present in all types of tissues to regulate a broad spectrum of biological processes[79]. **Hsa-miR-137** had an MCC score of 28 (Table 1) and is correlated with the NB stage. Hsa-miR-137 represses some genes involved in NB cell fate commitment, survival, and differentiation, such as GRN, GAD2, MEGF9, SLC1A5, SLC8A1, and SCN2A as illustrated in Supplementary Fig. 5[80–84]. It is known to be enriched in the brain tissues of mice and humans, with high expression in the hippocampus and cortical brain parts and low expression in the brain stem and cerebellum[85]. Many studies reported its role in modulating cellular differentiation and proliferation in the adult brain and embryonic stages. Sun et al. reported the involvement of hsa-miR-137 in halting cellular proliferation and stimulating their differentiation[86]. The role of hsa-miR-137 in glioblastoma multiforme-derived stem cells and brain malignancies stem cells differentiation into neural lineage has also been established[87]. On the contrary, a report confirmed the hsa-miR-137 role in the induction of adult neural stem cell proliferation and halting differentiation via repressing Ezh2[88]. These reports suggest that the role of hsa-miR-137 in neural stem cell differentiation is context-dependent. Overall, these results support the role of hsa-miR-137 in regulating pluripotency to differentiation states transition and its involvement in regulating apoptosis priming; its depletion is accompanied by developmental and tumor malformations[89]. A clinical analysis of 61 NB samples unraveled that hsa-miR-137 has a substantially lower expression with other higher prognostic factors, including high mouse/murine double minute 2 (Mdm-2) and MYCN amplification[90]. It targets histone demethylase, lysine-specific demethylase 1 mRNA in NB cells, activating multiple cellular processes that coincide with tumor suppression[91]. Consequently, the re-expression of hsa-miR-137 could be a therapeutic strategy as a potential tumor suppressor.

In the case of **hsa-miR-421**, it has an MCC score of 20. It is an oncomiR that is being activated by TGFB1[92] and MYCN as shown in Fig. 5. Hsa-miR-421 represses many neuronal-specific genes such as SOBP, HIPK, SLC43A1, FRS2, and OXR1 (Supplementary Fig. 6). Interestingly, aberrant expression of hsa-miR-421 regulates various types of malignancies via driving various genes, as shown in Fig. 5. Other than gastric cancer[93], hsa-miR-421 overexpression has a critical role in regulating pancreatic cancer[94], breast cancer[95], and prostate cancer[96]. Lie et al. reported the role of hsa-miR-421 in targeting myocyte enhancer factor-2D, which inhibits glioma cell glucose-dependent processes, angiogenesis, and invasion and enhances glioma cell sensitivity to radiotherapeutics[97]. Furthermore, MYCN amplification in NB inhibits ATM cascade, a canonical cascade in DNA damage repair, apoptosis, and cell-cycle arrest, via hsa-miR-421. Hsa-miR-421 is overexpressed in NB cells compared with matched normal cells, increasing its proliferation potential, migration, cell-cycle progression, and invasion capacity via targeting the tumor suppressor, Menin[98]. This is due to the inducing role of the TGFB1 transcription factor[99].

Hsa-mir-2110 has previously been reported to act as a tumor suppressor. Zhao et al. reported the role of eight of thirteen differentiation-stimulating miRNAs, including hsa-miR-2110, in direct or indirect blocking MYCN expression. Furthermore, it was observed that the expression levels of hsa-mir-137 and hsa-miR-2110 exhibit a substantial inverse correlation with the levels of MYCN, emphasizing their interaction with MYCN mRNA in NB progression[100]. Its effect on NB cell line survival has been reported with varied differentiation induction methodologies. Furthermore, they identified the clinical relevance of Tsukushi expression and hsa-miR-2110, where the higher Tsukushi mRNA level is associated with a low survival rate.

Another report shows the oncosuppressive role of hsa-miR-2110 in NB via targeting Tsukushi mRNA and inhibiting NB tumorigenesis[101]. Zhang et al. investigated the inhibitory effect of upregulating hsa-miR-2110 on triple-negative breast cancer cell proliferation, invasion, and migration in vitro and tumor growth in vivo[102]. As shown in Supplementary Fig. 7, MYCN prime hsa-miR-2110 expression, which represses the inhibitor of differentiation −3 (ID3), which has a critical role in sustaining cellular proliferation and halting the differentiation process[103,104].

Hsa-miR-1305 has an MCC score of 18 (Table 1) and is known as a tumor suppressor via regulating Mdm-2. As shown in Supplementary Fig. 8, hsa-miR-1305 has a repressive effect on CREBBP. It also represses the THADA fusion gene, which is proven to trigger insulin-like growth factor 2 (IGF-2) mRNA binding protein 3, which is known to be overexpressed in cancer[105]. Our network showed that hsa-miR-1305 downregulates the COL12A1 gene, which encodes for collagen type XII α1 that is found to trigger tumor progression and metastasis[106–108]. Hsa-miR-1305 also represses TUBB, which encodes for the beta-tubulin protein that is broadly expressed in the neuronal cells, which ultimately induces NB tumorigenesis (Fig. 5)[109]. Cai et al. investigated the effect of suppressing hsa-miR-1305 expression on the aggressiveness of non-small cell lung cancer cells (NSCLC), showing upregulation hsa-miR-1305 target Mdm-2, a p53 antagonist, which in role increases p53, a tumor suppressor, inducing attenuation of NSCLC[110]. Hsa-miR-1305 has been reported to have a critical role in circCOG2-mediated colorectal cancer proliferation, migration, and invasion via TGFB2/Smad pathway[111]. The role of exosomal hsa-miR-1305 has been established in inducing multiple myeloma tumorigenesis via decreasing cellular hsa-mir-1305, thus increasing the abundance of its target genes such as IGF1, FGF2, and Mdm-2[112]. Another report investigated the effect of circRIP2 on targeting hsa-miR-1305 to activate the TGF-β2/Smad3 cascade to induce epithelial to mesenchymal transition (EMT), accelerating bladder cancer aggressiveness[113]. It has been reported that hsa-miR-1305 silencing significantly inhibits hepatitis B virus-associated hepatocellular carcinoma tumorigenesis[114]. Despite numerous previous reports, hsa-miR-1305 role in NB tumorigenesis has not been studied. The diagnostic potential of hsa-miR-1305 as a biomarker could not be validated due to the limited availability of NB datasets that included expression data for this specific miRNA.

Hsa-miR-1976 is correlated with the NB stage, as shown in Table 2. It is shown to repress the LMTK3 transcription factor while regulating SPI1 and CREBBP (Supplementary Fig. 9). Hsa-miR-1976 is downregulated in tumor tissues, where its lower expression correlates with worse overall survival in a patient cohort obtained from the TCGA database. Hsa-miR-1976 knockdown in triple-negative breast cancer markedly promotes EMT and cancer stemness via direct regulation of phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma[115]. Chen et al.[116] reported the downregulation of hsa-miR-1976 in NSCLC, and its upregulation repress tumorigenesis via targeting phospholipase C epsilon 1. Another report tackles the association between the abundance of hsa-miR-1976 in plasma samples and the clinicopathological features of breast cancer patients and its role as a non-invasive biomarker for breast cancer diagnosis[117]. Thus, hsa-miR-1976 could be implicated in NB tumorigenesis and be a potential biomarker.

Hsa-miR-940 directly represses some genes associated with the neuronal differentiation of NB cells, such as TUBB, SLC1A5, MEGF9, PIP4K2A, and MYO7B (Supplementary Fig. 10). Hsa-miR-940 regulates cell survival and cancer resistance to chemotherapeutics by targeting the MAPK1 cascade[118]. Zhang et al. reported its abundance effect in promoting breast cancer via regulating FOXO3[119]. On the other hand, hsa-miR-940 is dramatically silenced in glioma cells, and its upregulation functions as a glioma suppressor via targeting cyclin kinase subunit 1, showing its tumor suppression effect[120]. Another study reported its inhibitory effect on glioma tumorigenesis via targeting methylenetetrahydrofolate dehydrogenase, which blocks folate metabolism in mitochondria[121]. Xu R. et al. reported hsa-mir-940 implication in halting EMT of glioma cells via targeting ZEB2[122]. Other than being a salivary biomarker for pancreatic cancer[123], it also marks

breast cancer, prostate cancer[124], and esophageal squamous cell carcinoma[125]. While the involvement of hsa-miR-940 in NB development remains unreported, our analysis revealed a significant correlation between its expression and NB stage.

Hsa-miR-760 is a hub node in our analysis with an MCC score of 11 (Table 1). As shown in Supplementary Fig. 11, it is upregulated by the MYCN gene and regulates the SPI1 transcription factor. In addition, it represses the PLD6 gene (phospholipase D), which facilitates many cellular processes in cancer progression, metabolism, and growth. It has been reported that the PLD6 gene alters mitochondrial fusion and fission dynamics, which dysregulate cellular mitochondrial bioenergetics, consequently, the progression of breast cancer[126]. Furthermore, hsa-miR-760 is downregulated in various cancer types, such as breast, prostate, NSCLC, gastric, and hepatocellular carcinoma, while upregulated in ovarian cancers[127]. Moreover, it is proven that hsa-miR-760 inhibits chemotherapeutics resistance in hepatocellular carcinoma via Notch1/Akt pathway[128]. In bladder cancer, it was found that METTL1 indirectly degrades tumor suppressor ATF3 mRNA mediated by miR-760, promoting tumor progression[129]. In our analyses, hsa-miR-760 shows promising prognostic and diagnostic ability. It is also correlated with the NB stage (Table 2, Fig. 7, and Fig. 8). However, we did not find any reports about its association with NB.

In conclusion, the proposed computational framework aims at integrating multi-omics data to identify potential biomarkers in NB. After integrating mRNA-seq, miRNA-seq, and methylation array data using SNF, feature selection was utilized to determine high-rank features. Interactions between high-rank genes and miRNAs were retrieved to build a regulatory network. It consisted of TF-miRNA interactions as well as miRNA-target interactions. We analyzed the interaction network using MCC and identified ten candidate NB biomarkers, three transcription factors, and seven miRNAs. Among them, the roles of MYCN, hsa-miR-137, hsa-miR-421, and hsa-miR-2110 in NB tumor development and progression have been studied and proven. On the other hand, the rest of the predicted biomarkers in our study, such as SPI1, POU2F2, hsa-miR-1305, hsa-miR-1976, hsa-miR-940 and hsa-miR-760, could serve as potential biomarkers to halt NB tumorigenicity. In addition, their role in other tumor development and progression has been studied. Our regulatory network shows that they interact with some well-studied NB biomarkers, including MYCN, which support their under-studied implication in NB development. Furthermore, their potential use as diagnostic biomarkers and correlation with NB patients' survival and stage information were studied. Six out of our 10 candidate biomarkers were shown to be correlated with the INSS stage, namely MYCN, POU2F2, hsa-mir-137, hsa-mir-1976, hsa-mir-940, and hsa-mir-760. High vs. low expression of MYCN, SPI1, and POU2F2 were shown to be correlated with significantly different survival outcomes, both in our dataset and in the external validation dataset. Furthermore, survival analysis of the external validation dataset suggested hsa-mir-137, hsa-mir-421, and hsa-mir-760 could be potential prognostic markers. The receiver operating characteristic (ROC) analysis on another external validation dataset suggested that hsa-mir-421 and hsa-mir-760 may also be diagnostic markers. Thus, experimental validation of the roles of SPI1 and POU2F2 in NB tumor progression could aid biomarker discovery. Furthermore, priming the transcriptional expression of miR-1976 and targeting hsa-miR-1305, hsa-miR-940, and hsa-miR-760 could block NB tumorigenesis. Therefore, this study proposes a deeper understanding of MYCN interactome, providing candidate routes for targeted therapies in NB.

Methods
Data retrieval
Multi-omics data were obtained for solid tumor samples from 99 patients. The patients' overall survival information and cancer staging according to the INSS were retrieved as well. Three data types were downloaded from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) program website. First, methylation array data were obtained using the Illumina Infinium Human Methylation 450 Platform.

Table 4 | The number of features before and after the preprocessing stage

Data type	Methylation array	mRNA-seq	miRNA-seq
Number of original features	485,577	56,038	1870
Number of features after preprocessing	379,522	50,501	1594

Methylation array data included beta values for 485577 features (reporter IDs) and mean detection p values. mRNA-seq data were obtained using the Illumina HiSeq2000 sequencing system. The dataset comprised FPKM values for 56038 features. Finally, the miRNA-seq data was generated using the Illumina-NextSeq500 system that included 1870 features (miRNAs). Each data type was arranged into a feature-by-patient table. Links to the methylation array, mRNA-seq, and miRNA-seq data are available in the project GitHub repository (https://github.com/ComputationalBiologyLab/Biomarker-discovery-in-Neuroblastoma-Multiomics).

Data preprocessing
The quality of the methylation array data was assessed by confirming that all samples had a mean detection p value < 0.005. Around 20% of methylation array beta values had Na values. That was due to a common SNP being found within 10 bp from the CpG site or, overlapping with a repeat element within 15 bp from the CpG site, or the detection p value of the sample being above 0.05. These rows were removed. Reporter IDs that mapped to multiple sites were also filtered. Features that mapped to the X or Y chromosomes were removed from methylation array data and mRNA data to avoid selecting biomarkers related to sex[130]. Features that had zeros across all samples were deleted from all datasets. mRNA FPKM values were transformed to TPM values to accommodate comparison across samples[131]. The total number of features before and after preprocessing is shown in Table 4. Min-max normalization was then applied to all datasets.

Data integration
SNF is a statistical network model that integrates the different omics data. A similarity network is created for each data type, in which a vertex represents each sample (i.e., patient), and the degree of similarity between samples is represented by weighted edges that connect the vertices. Therefore, a complete patient similarity matrix is constructed using edge weights. A sparse kernel matrix is also constructed where edge weights are used for determining the K-nearest neighbors. This operation sets the pairwise similarities of non-neighboring points to zero. Utilizing both matrices, network fusion is performed by updating the similarity matrix iteratively. After a specified number of iterations sufficient for convergence, the resultant similarity matrices are averaged to obtain the fused matrix.

SNF has the advantage of extracting valuable information even from small numbers of samples. Integrating similarity networks derived from different data types preserves both common and complementing aspects of these similarity networks, therefore showing the contribution of each data type to the overall similarity network. SNF has been used to study five types of cancer, namely glioblastoma multiforme, breast invasive carcinoma, lung squamous cell carcinoma, colon adenocarcinoma, and kidney renal clear cell carcinoma. For each cancer type, SNF was used to integrate methylation, mRNA, and miRNA expression data. On data integration, the researchers identified different subtypes that were shown to have distinct clinical hallmarks[132].

SNF uses three hyperparameters: T is the number of fusion iterations, k is the number of nearest neighbors to consider when building the sparse kernel matrix, and α is a hyperparameter used in the scaled exponential similarity kernel, which is used to calculate pairwise patient similarity values. SNF convergence was independently assessed for each hyperparameter by iteratively changing the value and calculating the relative change between fused graphs obtained in consecutive iterations using the spectral norm. Both T and k were selected from the range 2:50, while α was chosen from the range 0.05:1 with a step of 0.05. Based on the selected values ($T = 15$, $k = 20$, and $\alpha = 0.5$), The package "SNFtool" (https://CRAN.R-project.org/package=SNFtool) was used to perform the SNF model on the three data types, producing a fused similarity matrix.

Feature selection
The SNF feature selection technique was utilized to assign ranks to features based on how informative they were to the developed fused network. rSNF was performed by building a patient similarity matrix based solely on each feature. Spectral clustering[133] was performed on the single-feature-based and fused-based similarity matrices to identify disease subtypes. Then, the normalized mutual information (NMI) score[134] was utilized to determine the concordance between the results of both clustering. The higher the score, the more important the feature contribution to the fused similarity network. Different numbers of clusters have been tried (2:7).

The ratio between the minimum intra-cluster similarity and the maximum inter-cluster similarity was employed to evaluate the clustering quality. A good clustering should follow the principles of homogeneity and separation[135]. They state that elements within the same cluster should be close to each other (homogeneity), while elements in different clusters should be distant (separation). This means that patients within the same cluster would be highly similar, whereas patients in different clusters would have a low similarity.

If the minimum similarity within a cluster is high, it ensures that all other similarities within each cluster are even higher, which helps to group similar patients together. This is achieved by calculating pairwise similarities between all samples in each cluster separately and selecting the minimum similarity across all clusters afterward. Moreover, a low maximum similarity between patients in different clusters ensures that patients within different clusters are even less similar, making the clustering more reasonable. This is done by calculating pairwise similarities between all samples in different clusters in which we pick the maximum value.

Therefore, a similarity measure of the ratio between the minimum intra-cluster similarity (within) and the maximum inter-cluster similarity (between clusters) has been utilized to compare the different numbers of clusters to identify which is the best in terms of compactness and well separation. We have selected the number of clusters that achieved the highest ratio ($c = 4$).

Our approach is similar to the Dunn index[136], which measures the quality of clustering. However, the Dunn index utilizes distances between samples, while we use sample similarities. The similarity measure we used is calculated as follows:

$$\nabla(c_k) = \min_{x_i, x_j \in c_k} S(x_i, x_j) \, \forall x_i, x_j \in c_k, i \neq j, \forall k = 0, 1, 2, \ldots, n \quad (1)$$

Where dataset D with m samples, include $x_1, x_2, x_3, \ldots x_m$, n is the number of clusters in the cluster set C, $\nabla(c_k)$ denotes intra-cluster similarity of a cluster, and S is the similarity between two samples. That will result in k values where each one represents the inter-cluster similarity for each cluster as shown in Eq. 1. The inter-cluster similarity between clusters $\triangle(c_k, c_r)$ is calculated as illustrated in Eq. 2.

$$\triangle(c_k, c_r) = \max_{x_i \in c_k, x_j \in c_r} S(x_i, x_j) \, \forall x_i \in c_k, x_j \in c_r, i \neq j, k \neq r, \forall k = 0, 1, 2, \ldots, n$$

$$(2)$$

The clustering quality ratio Q is calculated as follows:

$$Q(C) = \frac{\min_{c_k \in C} \nabla(c_k)}{\triangle(c_k, c_r)} \, \forall k = 0, 1, 2, \ldots, n \quad (3)$$

The features were ranked based on their NMI score. For all three data types, it has been noticed that, after the initial few features, there was a consistent decline in the NMI score. The remaining features had negligible changes (Supplementary Fig. 12). To prioritize the most informative features for the fused network, the top 10% were chosen (Supplementary Data 1–3). Both mRNA-seq and methylation array features were then mapped to gene symbols, and shared genes were selected as essential genes along with the high-rank miRNAs for subsequent analysis. The complete lists of high-rank genes are illustrated in Supplementary Data 4.

Regulatory network construction and analysis

To gain a better understanding of the identified essential genes and high-rank miRNAs, possible interactions between the two sets were explored. To build an interaction network, interaction data for these biological molecules were obtained. Two types of gene-miRNA interaction data were retrieved. TF-miRNA regulations were acquired from the transmir 0.2 database[137], and miRNA-target interactions were acquired from the Tarbase v8 database. The transmir 0.2 database is a manually curated database with comprehensive information about the TF-miRNA regulations. The genes encoding for transcription factors in the candidate essential genes were identified through the database. Transmir 0.2 database includes experimentally verified interactions with different levels of confidence. Level 2 interactions were predicted by ChIP-seq experiments and further verified through high-throughput experiments, while literature-based interactions were manually curated from scientific papers. Level 2 and literature-based interactions were retrieved in our study. On the other hand, the Tarbase v8 database contains experimentally supported miRNA-target interactions. From both databases, interactions found in humans were downloaded. Interactions involving the identified high-rank features were extracted for further analysis.

Further annotations were acquired from the transmir 0.2 database website and added to our interaction tables. Transcription factors and miRNAs were annotated with associated diseases. Transcription factors were further annotated with prognostic correlations. The complete interaction and attribute tables are illustrated in Supplementary Data 5 and 6.

The regulatory network between TFs, genes, and miRNAs was constructed using Cytoscape software (version 3.9.1) after filtering normal cell line interactions. Node shape, color, edge shape, and type (activation or repression) were modified for visualization. Network analysis was employed to evaluate and rank the hub TFs, genes, and miRNAs. MCC algorithm in the Cytoscape plugin, CytoHubba, was utilized to identify the hub nodes in the regulatory network as it showed high-quality performance and accuracy over other measures, including radiality, stress, clustering coefficient, and betweenness[138–142].

Analysis of the identified candidate biomarkers' association with NB

Survival analysis was performed using the Kaplan–Meier method with a log-rank statistical test to explore the association of the proposed biomarkers with the overall survival of NB patients and identify which will be of prognostic potential. Biomarkers with p value < 0.05 are considered to have prognostic potential. Another external dataset, GSE62564, has been utilized to validate the results. The survival R package[143] was used for this purpose. The correlation between potential biomarker expressions and cancer stages was also statistically assessed using the Chi-square test. A p value < 0.05 was considered significant.

The ROC curve analysis was utilized to examine the diagnostic potential of the identified miRNAs by assessing their ability to discriminate between NB patients and normal samples. pROC R package[144] was used to generate ROC curves and calculate the AUC values for the proposed biomarkers from the GEO dataset GSE128004.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Code availability

References

1. Cheung, N.-K. V. & Dyer, M. A. J. N. R. C. Neuroblastoma: developmental biology, cancer genomics and immunotherapy. *Nat. Rev. Cancer* **13**, 397–411 (2013).
2. Johnsen, J. I., Dyberg, C. & Wickström, M. J. F. Neuroblastoma—a neural crest derived embryonal malignancy. *Front. Mol. Neurosci.* **12**, 9 (2019).
3. Maris, J. M. & Matthay, K. K. Molecular biology of neuroblastoma. *J. Clin. Oncol.* **17**, 2264 (1999).
4. Zeineldin, M., Patel, A. G. & Dyer, M. A. Neuroblastoma: when differentiation goes awry. *Neuron* **110**, 2916–2928 (2022).
5. Brodeur, G. M. & Bagatell, R. Mechanisms of neuroblastoma regression. *Nat. Rev. Clin. Oncol.* **11**, 704–713 (2014).
6. Brodeur, G. M. Spontaneous regression of neuroblastoma. *Cell Tissue Res.* **372**, 277–286 (2018).
7. Cheung, N.-K. V. et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *JAMA* **307**, 1062–1071 (2012).
8. Schleiermacher, G. et al. Treatment of stage 4s neuroblastoma–report of 10 years' experience of the French Society of Paediatric Oncology (SFOP). *Br. J. Cancer* **89**, 470–476 (2003).
9. Qin, H. et al. Clinical characteristics and risk factors of 47 cases with ruptured neuroblastoma in children. *BMC Cancer* **20**, 1–10 (2020).
10. Ambros, P. et al. International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. *Br. J. Cancer* **100**, 1471–1482 (2009).
11. Slack, A. et al. The p53 regulatory gene MDM2 is a direct transcriptional target of MYCN in neuroblastoma. *Proc. Natl. Acad. Sci. USA* **102**, 731–736 (2005).
12. Chen, Y. & Stallings, R. L. Differential patterns of microRNA expression in neuroblastoma are correlated with prognosis, differentiation, and apoptosis. *Cancer Res.* **67**, 976–983 (2007).
13. Schnepp, R. W. & Maris, J. M. Targeting MYCN: a good BET for improving neuroblastoma therapy? *Cancer Discov.* **3**, 255–257 (2013).
14. Bell, E. et al. MYCN oncoprotein targets and their therapeutic potential. *Cancer Lett.* **293**, 144–157 (2010).
15. Omenn, G. S., Nass, S. J. & Micheel, C. M. Evolution of translational omics: lessons learned and the path forward. National Academies Press (Washington, DC, USA, 2012).
16. Subramanian, I. et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol. Insights* **14**, 1177932219899051 (2020).

17. Park, A. & Nam, S. Deep learning for stage prediction in neuroblastoma using gene expression data. *Genomics Inform.* **17**, e30 (2019).

18. Maggio, V., Chierici, M., Jurman, G. & Furlanello, C. Distillation of the clinical algorithm improves prognosis by multi-task deep learning in high-risk neuroblastoma. *PLoS One* **13**, e0208924 (2018).

19. Simidjievski, N. et al. Variational autoencoders for cancer data integration: design principles and computational practice. *Front. Genet.* **10**, 1205 (2019).

20. Francescatto, M. et al. Multi-omics integration for neuroblastoma clinical endpoint prediction. *Biol. Direct* **13**, 1–12 (2018).

21. Duffy, D. J. et al. Integrative omics reveals MYCN as a global suppressor of cellular signalling and enables network-based therapeutic target discovery in neuroblastoma. *Oncotarget* **6**, 43182 (2015).

22. Gangoda, L. et al. Inhibition of cathepsin proteases attenuates migration and sensitizes aggressive N-Myc amplified human neuroblastoma cells to doxorubicin. *Oncotarget* **6**, 11175 (2015).

23. Wang, C., Lue, W., Kaalia, R., Kumar, P. & Rajapakse, J. Network-based integration of multi-omics data for clinical outcome prediction in neuroblastoma. *Sci. Rep.* **12**, 15425 (2022).

24. Narmontè, M. et al. Multiomics analysis of neuroblastoma cells reveals a diversity of malignant transformations. *Front. Cell Dev. Biol.* **9**, 727353 (2021).

25. Zhang, S. et al. Identification of a novel eighteen-gene signature of recurrent metastasis neuroblastoma. *J. Mol. Med*. **101**, 403–417 (2023).

26. Karagkouni, D. et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res* **46**, D239–D245 (2018).

27. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

28. Chin, C.-H. et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**, 1–7 (2014).

29. Kohl, N. E. et al. Human N-myc is closely related in organization and nucleotide sequence to c-myc. *Nature* **319**, 73–77 (1986).

30. Yin, Y. et al. N-Myc promotes therapeutic resistance development of neuroendocrine prostate cancer by differentially regulating miR-421/ATM pathway. *Mol. Cancer* **18**, 1–13 (2019).

31. Dong, A., Zhang, J., Sun, W., Hua, H. & Sun, Y. Upregulation of miR-421 predicts poor prognosis and promotes proliferation, migration, and invasion of papillary thyroid cancer cells. *J. Chin. Med Assoc.* **83**, 991 (2020).

32. Schwab, M. et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* **305**, 245–248 (1983).

33. Yancopoulos, G. D. et al. N-myc can cooperate with ras to transform normal cells in culture. *Proc. Natl Acad. Sci. USA* **82**, 5455–5459 (1985).

34. Cavalieri, F. & Goldfarb, M. J. O. N-myc proto-oncogene expression can induce DNA replication in Balb/c 3T3 fibroblasts. *Oncogene* **2**, 289–291 (1988).

35. Liu, X. et al. MiR-181a/b induce the growth, invasion, and metastasis of neuroblastoma cells through targeting ABI1. *Mol. Carcinog.* **57**, 1237–1250 (2018).

36. Lynch, J. et al. MiRNA-335 suppresses neuroblastoma cell invasiveness by direct targeting of multiple genes from the non-canonical TGF-β signalling pathway. *Carcinogenesis* **33**, 976–985 (2012).

37. Abou-Shanab, A. M., Gaser, O. A., Salah, R. A. & El-Badri, N. Application of the human amniotic membrane as an adjuvant therapy for the treatment of hepatocellular carcinoma. *Adv. Exp. Med. Biol.* https://doi.org/10.1007/5584_2023_792 (2023).

38. Matsumoto, M. et al. Expression of proto-oncogene products during drug-induced differentiation of a neuroblastoma cell line SK-N-DZ. *Acta Neuropathol.* **79**, 217–221 (1989).

39. Cinatl, J. et al. In vitro differentiation of human neuroblastoma cells induced by sodium phenylacetate. *Cancer Lett.* **70**, 15–24 (1993).

40. Reddy, C. D. et al. Anticancer effects of the novel 1α, 25-dihydroxyvitamin D3 hybrid analog QW1624F2-2 in human neuroblastoma. *J. Cell Biochem.* **97**, 198–206 (2006).

41. Henriksen, J. R. et al. Conditional expression of retrovirally delivered anti-MYCN shRNA as an in vitro model system to study neuronal differentiation in MYCN-amplified neuroblastoma. *BMC Dev. Biol.* **11**, 1–11 (2011).

42. Lovén, J. et al. MYCN-regulated microRNAs repress estrogen receptor-α (ESR1) expression and neuronal differentiation in human neuroblastoma. *Proc. Natl Acad. Sci. USA* **107**, 1553–1558 (2010).

43. Cotterman, R. & Knoepfler, P. S. N-Myc regulates expression of pluripotency genes in neuroblastoma including lif, klf2, klf4, and lin28b. *PLoS One* **4**, e5799 (2009).

44. Kang, J.-H. et al. MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. *Biochem. Biophys. Res. Commun.* **351**, 192–197 (2006).

45. Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. & Yamanaka, S. J. P. Promotion of direct reprogramming by transformation-deficient Myc. *Proc. Natl. Acad. Sci. USA* **107**, 14152–14157 (2010).

46. Roussel, M. F. & Robinson, G. W. Role of MYC in medulloblastoma. *Cold Spring Harb. Perspect. Med.* **3**, a014308 (2013).

47. Swartling, F. J. et al. Distinct neural stem cell populations give rise to disparate brain tumors in response to N-MYC. *Cancer Cell.* **21**, 601–613 (2012).

48. Goodman, L. A. et al. Modulation of N-myc expression alters the invasiveness of neuroblastoma. *Clin. Exp. Metastasis.* **15**, 130–139 (1997).

49. Zaizen, Y., Taniguchi, S., Noguchi, S. & Suita, S. The effect of N-myc amplification and expression on invasiveness of neuroblastoma cells. *J. Pediatr. Surg.* **28**, 766–769 (1993).

50. van Golen, C. M., Soules, M. E., Grauman, A. R. & Feldman, E. L. N-Myc overexpression leads to decreased β1 integrin expression and increased apoptosis in human neuroblastoma cells. *Oncogene* **22**, 2664–2673 (2003).

51. Tanaka, N. & Fukuzawa, M. MYCN downregulates integrin α1 to promote invasion of human neuroblastoma cells. *Int. J. Oncol.* **33**, 815–821 (2008).

52. Beierle, E. A. et al. N-MYC regulates focal adhesion kinase expression in human neuroblastoma. *J. Biol. Chem.* **282**, 12503–12516 (2007).

53. Megison, M. L. et al. FAK inhibition decreases cell invasion, migration and metastasis in MYCN amplified neuroblastoma. *Clin. Exp. Metastasis.* **30**, 555–568 (2013).

54. Moreau-Gachelin, F. Spi-1/PU. 1: an oncogene of the Ets family. *Biochim. Biophys. Acta* **1198**, 149–163 (1994).

55. Sementchenko, V. I. & Watson, D. K. J. O. Ets target genes: past, present and future. *Oncogene* **19**, 6533–6548 (2000).

56. Watson, D. & Seth, A. Oncogene reviews. *Oncogene* **19**, 6394–6399 (2000).

57. Watson, D., Li, R., Sementchenko, V., Mavrothalassitis, G. & Seth, A. Ets proteins in biological control and cancer (Academic Press: New York, 2002).

58. Dittmer, J. J. M. C. The biology of the Ets1 proto-oncogene. *Mol. Cancer* **2**, 1–21 (2003).

59. Oikawa, T. & Yamada, T. J. G. Molecular biology of the Ets family of transcription factors. *Gene* **303**, 11–34 (2003).

60. McKercher, S. et al. Targeted disruption of the PU. 1 gene results in multiple hematopoietic abnormalities. *EMBO J.* **15**, 5647–5658 (1996).

61. Turkistany, S. A. & DeKoter, R. P. The transcription factor PU. 1 is a critical regulator of cellular communication in the immune system. *Arch. Immunol. Ther. Exp. (Warsz.).* **59**, 431–440 (2011).

62. Will, B. et al. Minimal PU. 1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat. Med.* **21**, 1172–1181 (2015).

63. Tschan, M. et al. PU. 1 binding to the p53 family of tumor suppressors impairs their transcriptional activity. *Oncogene* **27**, 3489–3493 (2008).

64. Rimmelé, P. et al. Spi-1/PU. 1 oncogene accelerates DNA replication fork elongation and promotes genetic instability in the absence of DNA breakageSpi-1 alters replication elongation. *Cancer Res.* **70**, 6757–6766 (2010).

65. Du, B., Gao, W., Qin, Y., Zhong, J. & Zhang, Z. J. C. N. J. Study on the role of transcription factor SPI1 in the development of glioma. *Chin. Neurosurg. J.* **8**, 1–10 (2022).

66. Du, B. et al. PAICS is related to glioma grade and can promote glioma growth and migration. *J. Cell Mol. Med.* **25**, 7720–7733 (2021).

67. Meng, M., Chen, Y., Jia, J., Li, L. & Yang, S. Knockdown of PAICS inhibits malignant proliferation of human breast cancer cell lines. *Biol. Res.* **51**, 24 (2018).

68. Agarwal, S. et al. PAICS, a de novo purine biosynthetic enzyme, is overexpressed in pancreatic cancer and is involved in its progression. *Transl. Oncol.* **13**, 100776 (2020).

69. Zhou, S. et al. Roles of highly expressed PAICS in lung adenocarcinoma. *Gene* **692**, 1–8 (2019).

70. Agarwal, S. et al. PAICS, a purine nucleotide metabolic enzyme, is involved in tumor growth and the metastasis of colorectal cancer. *Cancers (Basel).* **12**, 772 (2020).

71. Chakravarthi, B. V. et al. A role for de novo purine metabolic enzyme PAICS in bladder cancer progression. *Neoplasia* **20**, 894–904 (2018).

72. Chakravarthi, B. V. et al. Expression and role of PAICS, a de novo purine biosynthetic gene in prostate cancer. *Prostate* **77**, 10–21 (2017).

73. Lagergren, A., Manetopoulos, C., Axelson, H. & Sigvardsson, M. J. B. C. Neuroblastoma and pre-B lymphoma cells share expression of key transcription factors but display tissue restricted target gene expression. *BMC Cancer* **4**, 1–15 (2004).

74. Warburg, O., Wind, F. & Negelein, E. The metabolism of tumors in the body. *J. Gen. Physiol.* **8**, 519–530 (1927).

75. Yang, R. et al. POU2F2 regulates glycolytic reprogramming and glioblastoma progression via PDPK1-dependent activation of PI3K/AKT/mTOR pathway. *Cell Death Dis.* **12**, 433 (2021).

76. Luo, R., Zhuo, Y., Du, Q. & Xiao, R. J. B. P. M. POU2F2 promotes the proliferation and motility of lung cancer cells by activating AGO1. *BMC Pulm. Med.* **21**, 1–12 (2021).

77. Mayol, G. et al. DNA hypomethylation affects cancer-related biological functions and genes relevant in neuroblastoma pathogenesis. *PLoS One* **7**, e48401 (2012).

78. Carén, H. et al. Identification of epigenetically regulated genes that predict patient outcome in neuroblastoma. *BMC Cancer* **11**, 1–11 (2011).

79. Krützfeldt, J. et al. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685–689 (2005).

80. De la Encarnación, A., Alquézar, C., Esteras, N. & Martín-Requero, Á. J. M. N. Progranulin deficiency reduces CDK4/6/pRb activation and survival of human neuroblastoma SH-SY5Y cells. *Mol. Neurobiol.* **52**, 1714–1725 (2015).

81. Westmoreland, J. J., Hancock, C. R. & Condie, B. G. Neuronal development of embryonic stem cells: a model of GABAergic neuron differentiation. *Biochem. Biophys. Res. Commun.* **284**, 674–680 (2001).

82. Nakayama, M. et al. Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening. *Genomics* **51**, 27–34 (1998).

83. Khananshvili, D. The SLC8 gene family of sodium–calcium exchangers (NCX)–Structure, function, and regulation in health and disease. *Mol. Asp. Med.* **34**, 220–235 (2013).

84. Sanders, S. J. et al. Progress in understanding and treating SCN2A-mediated disorders. *Trends Neurosci.* **41**, 442–456 (2018).

85. Willemsen, M. H. et al. Chromosome 1p21. 3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *J. Med. Genet.* **48**, 810–818 (2011).

86. Sun, G. et al. miR-137 forms a regulatory loop with nuclear receptor TLX and LSD1 in neural stem cells. *Nat. Commun.* **2**, 1–10 (2011).

87. Silber, J. et al. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. *BMC Med.* **6**, 1–17 (2008).

88. Szulwach, K. E. et al. Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127–141 (2010).

89. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).

90. Inomistova, M. et al. MiR-137 expression in neuroblastoma: a role in clinical course and outcome. *Biomedicine* **21**, 614–622 (2016).

91. Althoff, K. et al. MiR-137 functions as a tumor suppressor in neuroblastoma by downregulating KDM1A. *Int. J. Cancer* **133**, 1064–1073 (2013).

92. Rezaei, Z. & Sadri, F. MicroRNAs involved in inflammatory breast cancer: oncogene and tumor suppressors with possible targets. *DNA Cell Biol.* **40**, 499–512 (2021).

93. Wu, J.-H. et al. MiR-421 regulates apoptosis of BGC-823 gastric cancer cells by targeting caspase-3. *Asian Pac. J. Cancer Prev.* **15**, 5463–5468 (2014).

94. Hao, J. et al. MicroRNA 421 suppresses DPC4/Smad4 in pancreatic cancer. *Biochem Biophys. Res Commun.* **406**, 552–557 (2011).

95. Pan, Y. et al. MicroRNA-421 inhibits breast cancer metastasis by targeting metastasis associated 1. *Biomed. Pharmacother.* **83**, 1398–1406 (2016).

96. Meng, D. et al. A transcriptional target of androgen receptor, miR-421 regulates proliferation and metabolism of prostate cancer cells. *Int. J. Biochem. Cell Biol.* **73**, 30–40 (2016).

97. Liu, L., Cui, S., Zhang, R., Shi, Y. & Luo, L. MiR-421 inhibits the malignant phenotype in glioma by directly targeting MEF2D. *Am. J. Cancer Res.* **7**, 857 (2017).

98. Hu, H., Du, L., Nagabayashi, G., Seeger, R. C. & Gatti, R. A. ATM is down-regulated by N-Myc–regulated microRNA-421. *Proc. Natl Acad. Sci. USA* **107**, 1506–1511 (2010).

99. Yuan, H. S., Xiong, D. Q., Huang, F., Cui, J. & Luo, H. J. J. O. C. B. MicroRNA-421 inhibition alleviates bronchopulmonary dysplasia in a mouse model via targeting Fgf10. *J. Cell Biochem.* **120**, 16876–16887 (2019).

100. Zhao, Z. et al. A combined gene expression and functional study reveals the crosstalk between N-Myc and differentiation-inducing microRNAs in neuroblastoma cells. *Oncotarget* **7**, 79372 (2016).

101. Zhao, Z. et al. microRNA-2110 functions as an onco-suppressor in neuroblastoma by directly targeting Tsukushi. *PLoS One* **13**, e0208777 (2018).

102. Zhang, X. et al. Long noncoding RNA AFAP1-AS1 promotes tumor progression and invasion by regulating the miR-2110/Sp1 axis in triple-negative breast cancer. *Cell Death Dis.* **12**, 1–11 (2021).

103. Light, W., Vernon, A. E., Lasorella, A., Iavarone, A. & LaBonne, C. Xenopus Id3 is required downstream of Myc for the formation of multipotent neural crest progenitor cells. *Development* **132**, 1831–1841 (2005).

104. Lopez-Carballo, G. J. E. J. O. B. Expression of Id3 HLH transcription factor is down-regulated during retinoic acid-induced differentiation of SH-SY5Y neuroblastoma cells. **277**, 25297–25304 (2001).

105. Panebianco, F. et al. THADA fusion is a mechanism of IGF2BP3 activation and IGF1R signaling in thyroid cancer. *Proc. Natl. Acad. Sci. USA* **114**, 2307–2312 (2017).

106. Song, Y., Wang, L., Wang, K., Lu, Y. & Zhou, P. J. C. COL12A1 acts as a novel prognosis biomarker and activates cancer-associated fibroblasts in pancreatic cancer through bioinformatics and experimental validation. *Cancers (Basel).* **15**, 1480 (2023).

107. Jiang, X. et al. COL12A1, a novel potential prognostic factor and therapeutic target in gastric cancer. *Mol. Med Rep.* **20**, 3103–3112 (2019).

108. Yan, Y., Liang, Q., Liu, Y., Zhou, S. & Xu, Z. COL12A1 as a prognostic biomarker links immunotherapy response in breast cancer. *Endocr. Relat. Cancer* **30**, e230012 (2023).

109. Zhang, L. et al. Eya3 partners with PP2A to induce c-Myc stabilization and tumor progression. *Nat. Commun.* **9**, 1047 (2018).

110. Cai, Y. et al. miR-1305 inhibits the progression of non-small cell lung cancer by regulating MDM2. *Cancer Manag. Res.* **11**, 9529 (2019).

111. Gao, L. et al. Exosome-transmitted circCOG2 promotes colorectal cancer progression via miR-1305/TGF-β2/SMAD3 pathway. *Cell Death Discov.* **7**, 1–13 (2021).

112. Lee, J. Y. et al. Exosomal miR-1305 in the oncogenic activity of hypoxic multiple myeloma cells: a biomarker for predicting prognosis. *J. Cancer* **12**, 2825 (2021).

113. Su, Y. et al. circRIP2 accelerates bladder cancer progression via miR-1305/Tgf-β2/smad3 pathway. *Mol. Cancer* **19**, 1–13 (2020).

114. Rao, X. et al. N6-methyladenosine modification of circular RNA circ-ARL3 facilitates Hepatitis B virus-associated hepatocellular carcinoma via sponging miR-1305. *IUBMB Life.* **73**, 408–417 (2021).

115. Wang, J. et al. MiR-1976 knockdown promotes epithelial–mesenchymal transition and cancer stem cell properties inducing triple-negative breast cancer metastasis. *Cell Death Dis.* **11**, 1–12 (2020).

116. Chen, G. et al. MicroRNA-1976 functions as a tumor suppressor and serves as a prognostic indicator in non-small cell lung cancer by directly targeting PLCE1. *Biochem. Biophys. Res Commun.* **473**, 1144–1151 (2016).

117. Wang, J. et al. The low expression of miR-1976 in plasma samples indicating its biological functions in the progression of breast cancer. *Clin. Transl. Oncol.* **22**, 2111–2120 (2020).

118. Wang, Q. et al. Retaining MKP1 expression and attenuating JNK-mediated apoptosis by RIP1 for cisplatin resistance through miR-940 inhibition. *Oncotarget* **5**, 1304–1314 (2014).

119. Zhang, H. et al. MiR-940 promotes malignant progression of breast cancer by regulating FOXO3. *Biosci. Rep.* **40**, BSR20201337 (2020).

120. Luo, H. et al. MicroRNA-940 inhibits glioma cells proliferation and cell cycle progression by targeting CKS1. *Am. J. Transl. Res.* **11**, 4851 (2019).

121. Xu, T. et al. MicroRNA-940 inhibits glioma progression by blocking mitochondrial folate metabolism through targeting of MTHFD2. *Am. J. Cancer Res.* **9**, 250 (2019).

122. Xu, R. et al. MicroRNA-940 inhibits epithelial-mesenchymal transition of glioma cells via targeting ZEB2. *Am. J. Transl. Res.* **11**, 7351 (2019).

123. Xie, Z. et al. Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer. *Cancer Prev. Res (Philos.).* **8**, 165–173 (2015).

124. Oh-Hohenhorst, S. J. & Lange, T. J. C. Role of metastasis-related microRNAs in prostate cancer progression and treatment. *Cancers* **13**, 4492 (2021).

125. Wang, H., Song, T., Qiao, Y., Sun, J. J. E. & Medicine, T. miR-940 inhibits cell proliferation and promotes apoptosis in esophageal squamous cell carcinoma cells and is associated with post-operative prognosis. *Exp. Ther. Med.* **19**, 833–840 (2020).

126. von Eyss, B. et al. A MYC-driven change in mitochondrial dynamics limits YAP/TAZ function in mammary epithelial cells and breast cancer. *Cancer Cell* **28**, 743–757 (2015).

127. Manvati, M. K. S., Khan, J., Verma, N. & Dhar, P. K. J. G. Association of miR-760 with cancer: an overview. *Gene* **747**, 144648 (2020).

128. Tian, T. et al. MicroRNA-760 inhibits doxorubicin resistance in hepatocellular carcinoma through regulating Notch1/Hes1-PTEN/Akt signaling pathway. *J. Biochem. Mol. Toxicol.* **32**, e22167 (2018).

129. Xie, H. et al. METTL1 drives tumor progression of bladder cancer via degrading ATF3 mRNA in an m7G-modified miR-760-dependent manner. *Cell Death Discov.* **8**, 458 (2022).

130. Wang, Z., Wu, X. & Wang, Y. J. B. B. A framework for analyzing DNA methylation data from Illumina Infinium HumanMethylation450 BeadChip. *BMC Bioinformatics* **19**, 15–22 (2018).

131. Zhao, S., Ye, Z. & Stanton, R. J. R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* **26**, 903–909 (2020).

132. Wang, B. et al. Similarity network fusion for aggregating data types. *Nat. Methods* **11**, 333–340 (2014).

133. Ng, A., Jordan, M. & Weiss, Y. On spectral clustering: analysis and an algorithm. *NeurIPS Proceedings.* **14** https://proceedings.neurips.cc/paper_files/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf (2001).

134. McDaid, A. F., Greene, D. & Hurley, N. J. Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv* https://arxiv.org/abs/1110.2515 (2011).

135. Jones, N. C. & Pevzner, P. A. An introduction to bioinformatics algorithms. (MIT press, 2004).

136. Dunn, J. C. J. J. Well-separated clusters and optimal fuzzy partitions. *J. Cybern.* **4**, 95–104 (1974).

137. Tong, Z., Cui, Q., Wang, J. & Zhou, Y. TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res.* **47**, D253–D258 (2019).

138. Chicco, D. & Jurman, G. J. B. G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 1–13 (2020).

139. Jurman, G., Riccadonna, S. & Furlanello, C. A comparison of MCC and CEN error measures in multi-class prediction. *PLoS One* **7**, e41882 (2012).

140. Liu, Z. et al. Identification of hub genes and key pathways associated with two subtypes of diffuse large B-cell lymphoma based on gene expression profiling via integrated bioinformatics. *Biomed. Res. Int.* **2018**, 3574534, (2018).

141. Li, C. Y., Cai, J. H., Tsai, J. J. P. & Wang, C. C. N. Identification of hub genes associated with development of head and neck squamous cell carcinoma by integrated bioinformatics analysis. *Front. Oncol.* **10**, 681 (2020).

142. Samy, A., Maher, M. A., Abdelsalam, N. A. & Badr, E. J. S. R. SARS-CoV-2 potential drugs, drug targets, and biomarkers: a viral-host interaction network-based analysis. *Sci. Rep.* **12**, 11934 (2022).

143. Therneau, T.M. A package for survival analysis in S. Mayo Foundation. **2** https://www.mayo.edu/research/documents/tr53pdf/doc-10027379 (2015).

144. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 1–8 (2011).

Acknowledgements

Author contributions

E.B.: conceptualization, supervision, project administration. R.H.: data curation, methodology. A.M.A.: formal analysis, investigation. R.H., A.M.A., E.B.: writing—original draft preparation. E.B.: writing—review and editing. R.H. and A.M.A. are co-first authors who contributed equally to the development of this research work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41540-024-00371-3.

Correspondence and requests for materials should be addressed to Eman Badr.

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

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