Supplementary materials

GEO

We searched the NCBI GEO database for data related to miRNA in the context of ischemic stroke. We then compared logFC values and assessed statistical significance (*p*-value). Among the identified datasets was an experiment conducted on a rat model of cerebral ischemic stroke (GSE46266), as well as eight datasets derived from human samples: GSE55937, GSE60319, GSE95204, GSE158312, GSE178764, GSE195442, GSE201860 and GSE231431. In most cases the obtained *p*-values did not reach the threshold for statistical significance.

DATASETS

In the first study, "Circulating MicroRNAs as Potential Biomarkers for Ischemic Stroke in Patients with Asymptomatic Intracranial Artery Stenosis" (GSE201860), the levels of serum microRNA were compared in patients who suffered ischemic stroke and a control group. The differentially expressed microRNAs were validated by real-time PCR. In the dataset available with this study we can see different levels of miR-19a between the control group and the patients. They were classified into the two sex-matched groups. What was interesting is that in the female group the expression of miR-19a after stroke was smaller, whereas in the male group, we observed increased expression. However, the results are subject to error due to the small size of the groups.

In the next study, "microRNA Expression in Peripheral Blood Cells Following Acute Ischemic Stroke and Their Predicted Gene Targets" (GSE55937), they aimed to identify the differentially expressed miRNAs in patients with acute ischemic stroke and investigate how these microRNAs regulate immune-cell-gene expression. miR-19a was one of the six miRNAs whose expression increased compared to the control group. These miRNAs are either known or strongly predicted to regulate the expression of multiple genes involved in pathways linked to ischemic stroke, such as NF-κβ signalling, toll-like receptor signalling, leukocyte extravasation, interleukin signalling, transforming growth factor-β signalling and the prothrombin activation pathway. At the biological system level, several of these miRNAs have been reported to influence components of both the immune and coagulation systems. However, the *p*-value was also subject to the error for the same reason.

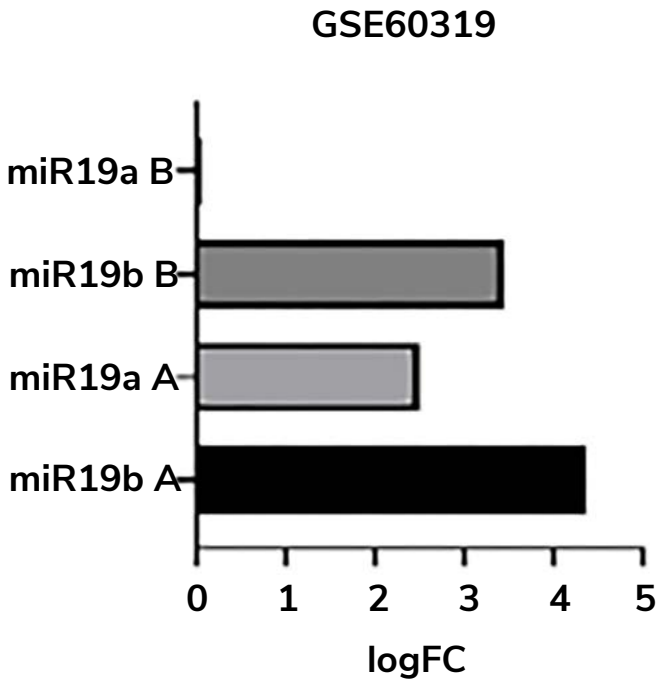
The study "microRNAs Involved in Regulating Spontaneous Recovery in Embolic Stroke Model" (GSE46266) presented the role of miRNAs in the spontaneous recovery phase in cerebellar ischemia. They observed that miR-19a and miR-19b increase expression from 0 h to 168 h after the stroke. The *p*-value is too small.

In "Identification of Circulating microRNAs as Potential Biomarkers for Detecting Acute Ischemic Stroke" (GSE60319), they study the involvement of miRNA in regulating the pathogenesis associated with middle-cerebral artery occlusion in SD rats. The expression of miR-19b is increasing, and the *p*-value is again larger than 0.005.

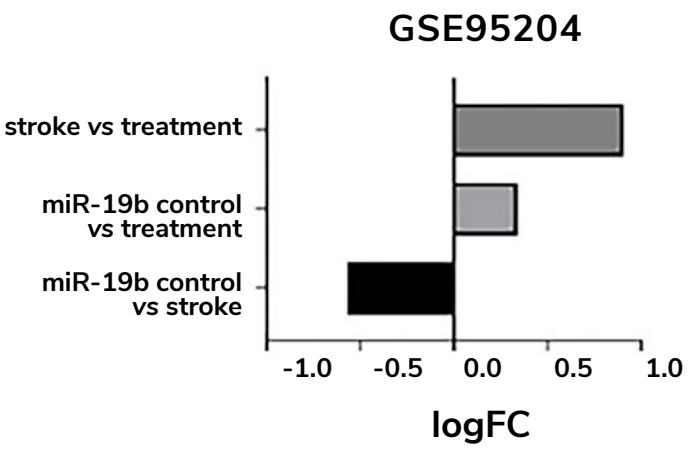
The last dataset (GSE95204) came from a study, "Plasma miRNAs Expression in Acute Ischemic Stroke patients with or without Recombinant Tissue Plasminogen Activator Treatment", which examined differences in microRNA expression among patients receiving rtPA treatment, patients with ischemic stroke and a control group. Between the control group and the stroke group, a decrease in miR-19a expression was observed, whereas between the rtPA treatment group and the stroke group, an increase in expression was noted. This suggests that the therapy may be responsible for elevating miR-19a levels, which could be a beneficial effect; however, the *p*-value should be interpreted with caution due to the small sample sizes of the groups.

GSE60319

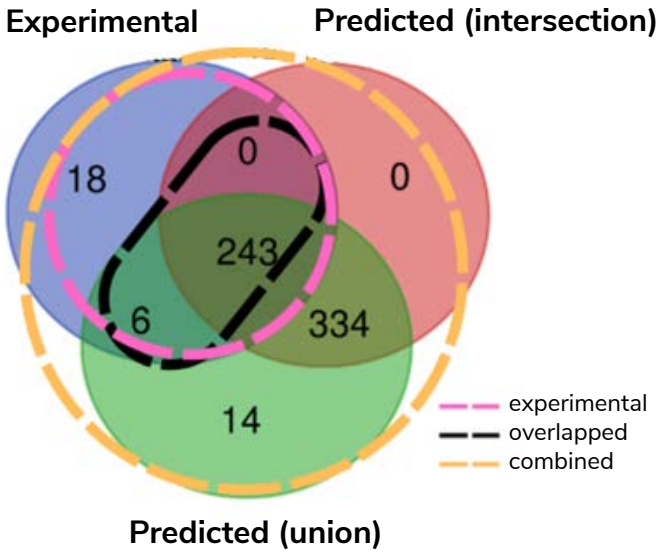
The dataset GSE60319, published in 2015, focuses on the identification of circulating microRNAs as potential diagnostic biomarkers for acute ischemic stroke (AIS). The study involved serum samples taken from 117 patients diagnosed with ischemic stroke and 82 healthy controls. Through microarray profiling and subsequent validation by qRT-PCR, several differences in expression of miRNAs were identified. Among these, miR-19b-3p (A group) showed a statistically significant upregulation, with



Supplementary Fig. 1. Changes in the expression levels of miR-19a and miR-19b in groups A and B, presented as logFC.



Supplementary Fig. 2. Comparison of logFC values for miR-19b-3p across three comparisons: control vs. stroke, control vs. treatment, and stroke vs. treatment.



Supplementary Fig. 3. Classification of target gene subsets for enrichment analysis. Venn diagram showing the overlap between experimentally validated targets, the intersection of predicted targets and the union of predicted targets. The inner outlines define the three subsets used in downstream analyses: (1) the experimental set (magenta dashed line); (2) the overlapped set (black dashed line) – genes shared between the experimental set and at least one predicted set; (3) the combined set (orange dashed line) – all genes from experimental and predicted sets. Numbers indicate gene counts.

a logFC = 3.45 and a *p*-value = 0.003661, while miR-19a-3p (A group) displayed a logFC = 2.50 with a *p*-value = 0.088477, which did not meet the threshold for statistical significance. Another variant, miR-19b-3p (B group), also reached statistical significance with a logFC of 2.98 and a *p*-value = 0.002902. The dataset design included multiple patient subgroups, differentiating between thrombotic and embolic stroke types, as well as comorbidities like hypertension and heart disease, further enhancing the clinical relevance of the findings. These results suggest that miR-19b-3p in particular may serve as a promising biomarker candidate for AIS.

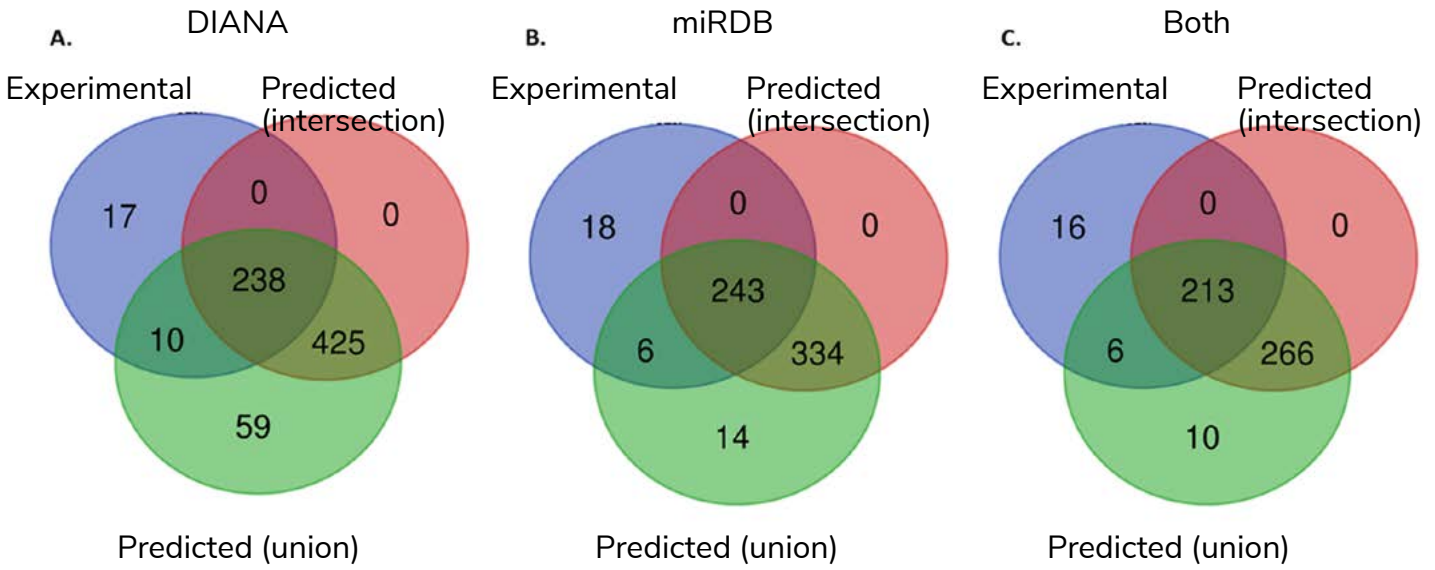
GSE95204

The dataset GSE95204, published in 2017, investigates the expression of circulating plasma miRNAs in patients with AIS, specifically comparing those who received thrombolytic therapy with recombinant tissue plasminogen activator (rtPA) and those who did not. The study included three groups: 15 AIS patients treated with thrombolysis, 15 AIS patients without thrombolysis (both with samples collected within 24 hours of symptom onset) and a control group of 15 age-matched healthy individuals. miRNA profiling was conducted using microarray technology and validated with real-time PCR. Among many investigated miRNA – miR-19b-3p held our interest. The results show that miR-19b-3p was significantly upregulated in stroke patients compared to those who received treatment, with a logFC = 0.91 and a p-value = 0.004888, indicating a statistically significant difference. However, comparisons between the control group and stroke patients (logFC = –0.57, p = 0.422) as well as between the control group and treated patients (logFC = 0.34, p = 0.632) did not reach statistical significance. These findings suggest that miR-19b-3p may be responsive to thrombolytic therapy and could potentially serve as a biomarker to monitor treatment efficacy in AIS. Although its expression does not significantly differ between controls and stroke patients overall, the distinct difference of expression observed between stroke without treatment vs. stroke with treatment underscores its potential relevance in therapy response rather than in initial diagnosis.

miRPath

At an earlier stage of the project we also explored miRPathDB 2.0, an integrative miRNA pathway dictionary and enrichment resource described by Kehl et al. (2020) [1]. However, the corresponding web server was no longer accessible in its original form at the end of manuscript preparation, so it could not be included in our formal benchmarking. We nevertheless retain this tool in the supplement as an example of a user-friendly, web-based implementation of miRNA pathway analysis and because similar resources may reemerge in the future. At present, comparable functionality is largely available through R packages, which, while powerful, may be difficult to use for non-bioinformatic researchers and clinical teams.

For the purposes of further analysis, the target genes were divided into three subsets: (1) experimental – comprising of all genes supported by confirmed experimental evidence; (2) overlapped – genes shared between the experimental set and at least one of the predicted sets; and (3) combined – a comprehensive set including all genes. This classification, visible in Fig. S3, enabled a detailed assessment of prediction accuracy and the extent of overlap with experimental data.



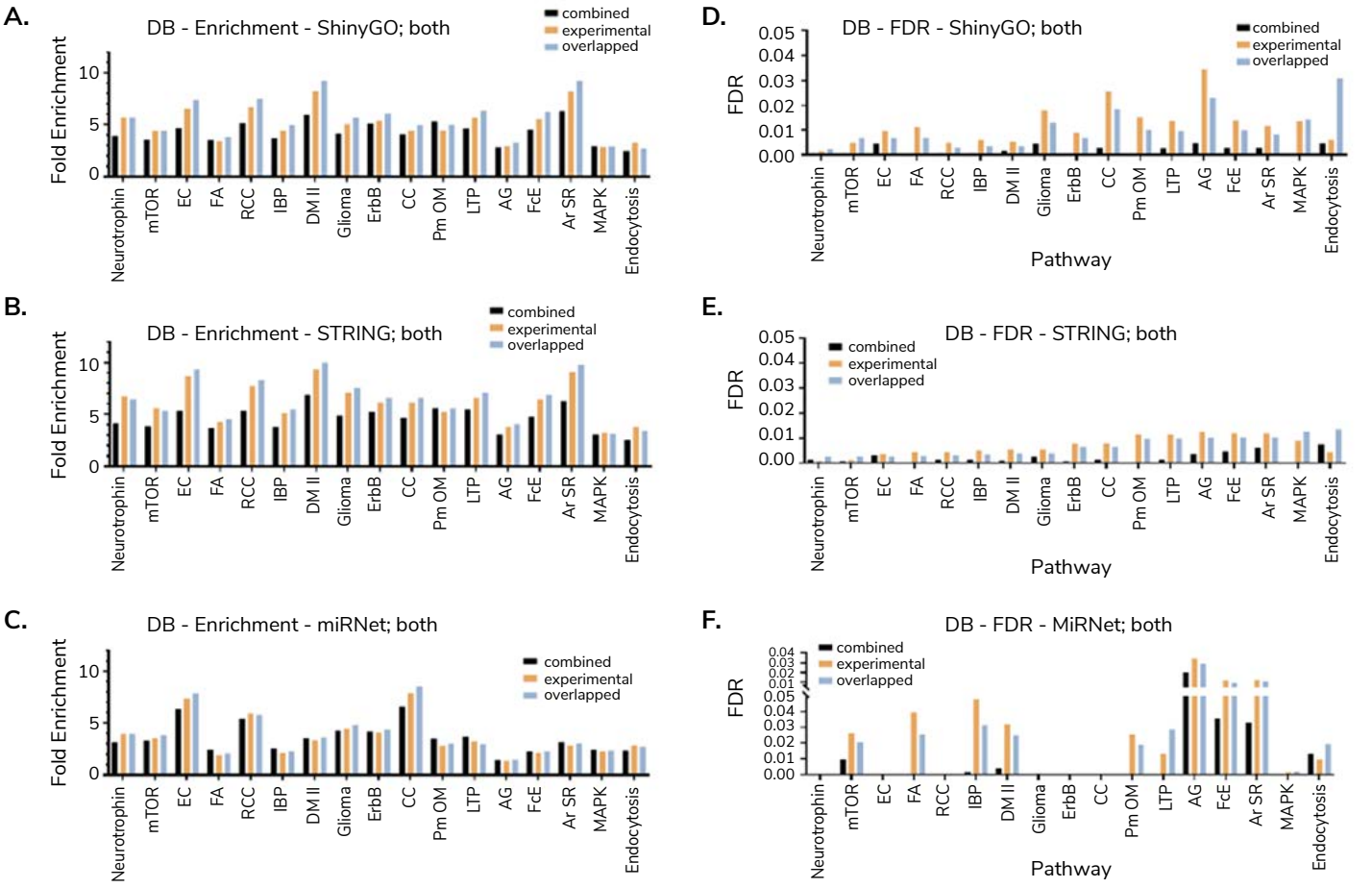
Supplementary Fig. 4. Classification of target gene subsets for enrichment analysis across target prediction tools. Venn diagrams showing the overlap between experimentally validated miR-19a-3p targets and predicted targets for (A) DIANA; (B) miRDB; (C) Intersection of both these lists. In each panel, blue indicates the experimental set, red the predicted intersection and green the predicted union; numbers denote gene counts in each group.

Supplementary Tab. I. Sizes of miR-19a-3p target gene subsets by target prediction source and subset type. Values indicate the number of miR-19a-3p target genes in each subset: experimental (only experimentally validated genes), overlapped (genes shared between the experimental set and at least one predicted set) and combined (union of experimental and predicted genes) for DIANA, miRDB and the intersection ("Both") target lists used in enrichment analyses.

	SOURCE OF TARGET GENES		
Category	DIANA	miRDB	Both
Experimental	265	267	235
Overlapped	248	249	219
Combined	749	615	511

This approach enabled the identification of distinct subsets of miR-19a-3p target genes, each of which was subsequently subjected to separate functional and enrichment analysis. To investigate the biological relevance of these gene sets, enrichment analysis was performed using three complementary bioinformatics tools: STRING, miRNet and ShinyGO. Each tool was applied independently to the individual gene subsets to identify overrepresented biological pathways among the miR-19a-3p regulated genes. The resulting enrichment profiles were then compared both across the gene groups and between the tools themselves, in order to evaluate differences arising from the prediction methods, overlap with experimental data and the underlying algorithms and characteristics of each tool.

The categories Predicted (union) and Predicted (intersection) refer, respectively, to genes predicted by at least one (union) or all three (intersections) of the following tools: DIANA-microT, miRDB and TargetScan. The Experimental category was constructed by integrating three evidence levels available in the miRPathDB platform – "Strong Experimental Evidence", "Weak Experimental Evidence", and "Any Experimental Evidence". These data are derived from the miRTarBase database and include both high--confidence validation methods such as Western Blot, RT-PCR, luciferase reporter assay or GFP reporter assay, and lower--confidence methods like microarray, Northern blot, pSILAC or branched DNA probe assay. This classification allowed for a comprehensive assessment of the predictive accuracy and concordance with experimentally validated results. The division in case of three target lists we analysed can be seen on Fig. S4. In the case of DIANA, a total of 238 genes were predicted by all three tools (intersection) and confirmed by experimental data, forming the most reliable gene set. Additionally, 10 genes overlapped only between the experimental and union categories, while 59 genes were predicted by only one or two tools (union) without any experimental confirmation. Notably, 425 genes belonged to the intersection and union sets but lacked any supporting experimental evidence. Seventeen genes were present only in the experimental category, meaning they were not predicted by any of the three tools but were shown to be confirmed by experimental data.



Supplementary Fig. 5. Effect of target gene subset on KEGG pathway enrichment for the intersection of the "both" target set across enrichment tools. (A) ShinyGO fold enrichments of KEGG pathways for the "both" set; (B) STRING fold enrichments of KEGG pathways for the "both" set; (C) miRNet fold enrichments of KEGG pathways for the "both" set; (D) ShinyGO FDR values of KEGG pathways for the "both" set; (E) STRING FDR values of KEGG pathways for the "both" set; (F) miRNet FDR values of KEGG pathways for the "both" set. Within each tool, results are presented for three target gene subsets: the combined union of experimentally validated and predicted targets (black), the experimental subset (orange), and the overlapped subset (blue) comprising genes shared between the experimental set and at least one predicted set.

neurotrophin – neurotrophin signalling pathway; mTOR – mTOR signalling pathway; EC – endometrial cancer; FA – focal adhesion; RCC – renal cell carcinoma; ISP – insulin signalling pathway; DM II – type II diabetes mellitus; ErbB – ErbB signalling pathway; CC – colorectal cancer; Pm OM – progesterone--mediated oocyte maturation; LT P – long-term potentiation; MAPK – MAPK signalling pathway.

For miRDB, the numbers were similarly distributed: 243 genes were predicted by all three tools and experimentally confirmed; 6 genes overlapped between the experimental and union sets; and 14 genes were exclusive to the union category. A total of 334 genes were predicted without any experimental evidence, while 18 genes appeared only in the experimental dataset.

Lastly, we combined the intersection of genes found in both miRDB and DIANA. In this case, 213 genes were confirmed both experimentally and by prediction. Furthermore, 6 genes were common to the experimental and union sets; meanwhile 10 genes were found only in the union. In contrast, 266 genes were predicted by at least one tool but lacked experimental validation, while 16 genes were found solely in the experimental dataset.

Based on the obtained results, target genes were divided into three categories: experimental, overlapped and combined. As shown in Tab. SI., experimental category includes all genes from the experimental categories of miRPathDB database, regardless of prediction outcome; this comprised 265 genes for DIANA, 267 for miRDB, and 235 for the combined dataset. The overlapped category included genes present in both the experimental dataset and at least one of the predictive categories. This group included 248 genes for DIANA, 249 for miRDB and 219 for the combined analysis. The final category, combined, represented the complete union of predicted and experimentally validated genes, comprising 749 genes for DIANA, 615 for miRDB and 511 in the joint dataset.

Across ShinyGO and STRING, the overlapped subset generally yields the highest fold enrichment and the lowest FDR values for most pathways, indicating a stronger and more specific signal when analyses are restricted to genes supported by both prediction and experiment (Fig. S5.). The experimental subset typically shows intermediate fold enrichment with slightly higher FDRs, while the combined union tends to dilute effect sizes, particularly in broad pathways such as focal adhesion, endocytosis and MAPK

signalling. For miRNet, enrichment is more sensitive to subset choice: restricting analysis to experimental targets alone often results in loss of significance (inflated FDRs), whereas both the overlapped and combined subsets recover several signalling pathways with acceptable FDRs. Between tools, STRING generally produces the highest fold enrichment values, with FDRs comparable to or slightly lower than ShinyGO, while miRNet tends to yield lower--fold enrichment and consistently higher FDRs, reflecting a more conservative behaviour. Despite these differences, key pathways such as neurotrophin and mTOR signalling remain consistently enriched across tools and subsets, while the magnitude of fold enrichment and statistical support depends on how conservatively the target list is defined. Overall, these patterns suggest that the overlapped subset offers the best trade-off between effect size and multiple-testing burden for ShinyGO and STRING, whereas miRNet benefits from including a broader set of predicted targets.

REFERENCES

Kehl T, Kern F, Backes C, Fehlmann T, Stöckel D, Meese E, et al. miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database. Nucleic Acids Res. 2020 Jan 8;48(D1):D142–7.