Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Expression of natural killer cell cytotoxic gene is reduced in peripheral blood within 24 hours after cardioembolic stroke

Marianne Nguyen

manguyen27@bfhsla.org

Benjamin Franklin High School

Jennifer Nguyen

Touro Infirmary

Research Article

Keywords: stroke, biomarker, NK cells, transcriptomics, machine learning, PRSS23, granulysin

Posted Date: April 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9438981/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

When ischemic stroke occurs, natural killer (NK) cells mobilize out of the blood toward injured brain tissue within hours. That trafficking event is well-documented, its transcriptomic footprint in peripheral blood has never been used diagnostically. This study tests whether it should be. We hypothesized that NK cell-associated genes would shower lower expression in peripheral blood from cardioembolic stroke patients compared to healthy controls. A 10-gene NK cell specificity signature derived from healthy donor profiles (NCBI GEO) was used to train Random Forest and Logistic Regression classifiers on whole blood expression data from 69 cardioembolic stroke patients and 23 healthy controls, sampled at 3, 5, and 24 hours post-stroke. Both models reached ~ 77% accuracy and an Area Under the Curve (AUC) of up to 0.798, against a demographic baseline of 66.3%. Three genes showed individual-level depletion in stroke blood: granulysin (GNLY) and perforin (PRF1), encoding the core NK cytotoxic effectors granulysin and perforin, and SH2D1B. GNLY was already depleted at 3 hours; the other two did not reach significance until 24 hours, a staggered pattern consistent with a two-phase model of early mobilization followed by slower exhaustion or continued trafficking. One outlier: PRSS23, a serine protease in the top 0.02% of NK specificity scores, contributed to classifier performance with no significant individual effect and no prior stroke literature. Its vascular endothelial roles make the signal plausible. It warrants direct follow-up

Introduction

Ischemic stroke kills or permanently disables more people each year than almost any other disease and makes up roughly 87% of all stroke cases worldwide (5). Because of arterial obstruction, cerebral blood flow is cut off, and neurons begin dying within minutes. However, the final extent of neurological damage is not determined at the moment of occlusion. The brain and peripheral immune system enter a rapid and dynamic interaction within hours of stroke, and the downstream immune response substantially shapes whether initial injury becomes a lasting deficit (15). Identifying peripheral blood biomarkers that reflect this post-stroke immune state in real time has become an active research priority, both for diagnosis and for predicting outcomes (5).

NK cells are at the center of that immune response. These innate lymphocytes act fast, do not require prior antigen sensitization, and deploy a well-characterized cytotoxic arsenal including GNLY, PRF1, and granzyme B (GZMB) to eliminate damaged or infected cells (2, 9). After an ischemic stroke, their role turns distinctly double-edged. Ischemic neurons actively recruit NK cells to injured brain tissue, and rather than conferring neuroprotection, this recruitment has been linked to acceleration of infarct expansion (12). The trafficking happens quickly: NK cells accumulate in ischemic tissue via IP-10-mediated chemotaxis within hours of onset (10), and their depletion from the peripheral

circulation as a consequence has now been documented in both animal models and human patients (8). That peripheral depletion is the core observation this study builds on. If NK cells are leaving the blood and entering injured brain tissue, their transcriptomic footprint in whole blood should change, and that change should be measurable, potentially early, and potentially diagnostic.

Whether it actually is has not been tested systematically. Blood genomics studies have characterized NK cell expression profiles alongside other blood cell types at baseline (1), and transcriptome-wide analyses have catalogued broad immune gene expression changes in whole blood following stroke across multiple timepoints (6, 11). But the narrower question, can a focused NK cell gene panel, built from the known cytotoxic and receptor machinery of NK cells, discriminate stroke patients from healthy controls better than clinical variables alone, has gone unaddressed. Machine learning applied to immune gene expression has shown real promise for classifying inflammatory conditions (7); its application to NK-specific panels in stroke is, as far as we can determine, unexplored. A secondary gap concerns PRSS23. Well-characterized NK genes like GNLY, PRF1, and NKG7 have established cytotoxic roles (2). PRSS23, a secreted serine protease with high NK cell specificity and documented roles in vascular endothelial biology (3, 14), has never been examined as a stroke biomarker, despite a vascular expression profile that makes it, in principle, relevant.

Our main hypothesis was that NK cell-associated genes (GNLY, PRF1, SH2D1B) would show lower expression in peripheral blood within the first 25 hours after cardioembolic stroke. To test this, we performed machine learning analyses as a secondary analysis but did not use it to test the primary hypothesis. We derived a 10-gene NK cell specificity signature from a publicly available healthy donor blood cell dataset (GSE72642) (1) and used it as the feature set for machine learning classifiers, which were trained on a longitudinal stroke transcriptomics dataset (GSE58294) (6). Both a Random Forest and Logistic Regression classifier reached approximately 77% accuracy with an AUC of up to 0.798, against a clinical baseline of 66.3%. Three genes, GNLY, PRF1, and SH2D1B, were individually downregulated in stroke blood. GNLY was depleted as early as 3 hours post-stroke, and PRF1 and SH2D1B reached significance at 24 hours. This pattern was consistent with a two-phase model of NK cell mobilization and exhaustion. PRSS23 emerged as a contributing classifier feature with no prior stroke literature. Its known vascular expression and role in endothelial-to-mesenchymal transition suggest it is worth investigating directly as a candidate blood biomarker.

Materials and Methods

Data Sources and Acquisition

Gene expression data were drawn from three publicly available datasets on the NCBI Gene Expression Omnibus (GEO). GSE72642 (1) explored six isolated peripheral blood cell types, CD19 + B cells, CD4 + T cells, CD8 + T cells, CD14 + monocytes, CD56 + NK cells, and polymorphonuclear cell, among three healthy human donors using the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570). This dataset was used solely for NK signature derivation. Secondly, GSE58294 (6) provided whole blood gene expression from 69 cardioembolic ischemic stroke patients, and 23 health controls also profiled on the Affymetrix platform. Patient samples were drawn at 3, 5, and 24 hours post-stroke onset. Finally, GSE203399 (4) was used for only the clinical baseline model. This dataset contained DNA methylation profiles and clinical outcome data including admission and discharge NIHSS scores from 62 ischemic stroke patients, profiled on the Illumina Infinium HumanMethylation EPIC BeadChip (GPL29753). All three datasets were downloaded as Series Matrix files and processed in Python 3.12. No new human subject data was collected.

NK Cell Gene Signature Derivation

From GSE72642 (1), expression data were loaded and sample metadata was parsed to separate CD56 + NK cell samples ($n = 3$) from all other cell types ($n = 15$). We then computed a specificity score for each of the 54,675 Affymetrix probes as the difference between mean NK expression and mean non-NK expression. We then ranked probes by this score and mapped the top 20 to HGNC gene symbols via a GPL570 annotation table. From this ranked list, ten genes were taken based on NK specificity score and functional relevance to cytotoxicity or NK cell signaling: GNLY, NKG7, PRF1, GZMB, KLRD1, KLRB1, NCR1, FCGR3A, SH2D1B, and PRSS23.

Clinical Baseline Model

Patient age and binary-encoded sex was extracted from GSE203399 (4) Series Matrix metadata. Then, a binary outcome was derived from delta-NIHSS (discharge minus admission), where values above zero were coded as improved (1) and values at or below as not improved (0). Logistic regression and Random Forest (100 estimators) classifiers were each trained on these two features under 5-fold stratified cross-validation, using accuracy as the primary metric. Feature importances from the Random Forest were used to quantify the relative contribution of age versus sex.

Stroke Classification Using NK Gene Expression

Machine learning analyses were done as an exploratory extension and were not used to test the primary hypothesis. Probes corresponding to the 10 NK signature genes were

taken from the GSE58294 (6) expression matrix. This data was transposed into a sample-by-gene feature matrix, and group labels (cardioembolic stroke or control) and time-post-stroke were drawn from the Series Matrix metadata. Before model training, we standardized gene expression values to zero mean and unit variance using scikit-learn StandardScaler. A Random Forest (200 estimators, random state = 42) and a Logistic Regression (maximum iterations = 1000) was evaluated under a 5-fold stratified cross-validation, using accuracy and AUC-ROC as metrics. We constructed Receiver Operating Characteristic (ROC) curves by interpolating true positive rates across folds at fixed false positive rate intervals. To rank gene contributions, we used feature importances from the fitted Random Forest.

Differential Expression Analysis

For each of the 10 NK genes, mean expression in stroke patients and healthy controls was compared using a two-sample independent t-test (SciPy stats.ttest_ind). We assessed significance at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Temporal Expression Analysis

We examined the three genes that reached significance in the differential expression analysis (GNLY, PRF1, and SH2D1B) across four sample groups in GSE58294 (6): healthy controls and stroke patients at 3h, 5h, and 24h post-onset ($n = 23$ per group). At each time point, we calculated mean expression and standard error and ran independent t-tests between the control group and each post-stroke timepoint to identify the earliest point of significant deviation from baseline.

Software and Reproducibility

For analysis, we used Python 3.12 within a Kaggle Notebook environment (Kaggle Inc., www.kaggle.com), providing a cloud-based Jupyter notebook interface with internet access enabled for dataset collection. For data handling, the core packages used were pandas and NumPy, scikit-learn for model training and cross-validation, SciPy for statistical tests, and matplotlib for figure generation. Specific versions were pandas 2.0, NumPy 1.24, scikit-learn 1.3, SciPy 1.11, and matplotlib 3.7. We uploaded Series Matrix files for GSE72642 (1) and GSE203399 (4) directly to the Kaggle notebook environment, while we retrieved GSE58294 (6) programmatically with Python's urllib.request module from the NCBI FTP server. All code is provided in the supplementary materials. The three GEO datasets are publicly available without access restrictions.

Results

NK Cell Gene Signature

To identify genes selectively expressed in NK cells relative to other peripheral blood cell types, a specificity score was computed for each of 54,675 probes in GSE72642 (1) by comparing mean NK expression against all other cell types. The top 20 probes were divided into two functional groups: cytotoxicity effector genes and KIR family members. The cytotoxic cluster, GNLY (5.002), NKG7 (4.435), PRF1 (4.106), and GZMB (4.084), constituted the highest-scoring probes and corresponded to previously characterized co-expressed markers of cytotoxic lymphocyte identity (2) (Fig. 1A-D). PRSS23 ranked 9th (score = 4.139), encoding a secreted serine protease with no prior documented role in stroke biology (3). Based on specificity scores and functional relevance, ten genes were selected for downstream analysis: GNLY, NKG7, PRF1, GZMB, KLRD1, KLRB1, NCR1, FCGR3A, SH2D1B, and PRSS23.

Figure 1. NK cell gene signature differentiates cardioembolic stroke from healthy controls. (A) Classification accuracy of Random Forest and Logistic Regression compared to demographic baseline. (B) Random Forest gene importance scores for the 10-gene signature. (C) Temporal expression of GNLY at 3, 5, and 24 hours post-stroke. (D) Expression of PRSS23 in stroke versus control samples. Accuracy evaluated via 5-fold stratified cross-validation. ns = not significant.

Clinical Baseline

To establish a minimum performance threshold for gene-based classifiers, a logistic regression model was trained on age and gender alone using the GSE203399 cohort (n = 62) (4). The model achieved a cross-validated accuracy of 66.3% (± 6.9%), with age accounting for 91% of feature importance, consistent with the established predictive relationship between patient age and post-stroke neurological recovery (5). This figure was used as the benchmark that any subsequent expression-based model was required to surpass.

Classification

To evaluate whether the 10-gene NK signature could discriminate cardioembolic stroke patients from healthy controls, Random Forest and Logistic Regression classifiers were trained on NK gene expression data from GSE58294 (n = 92) (6) using 5-fold stratified cross-validation. Both models exceeded the clinical baseline by approximately 11 percentage points: Logistic Regression achieved 77.3% accuracy (± 5.0%, AUC = 0.798 ± 0.048) and Random Forest achieved 77.1% (± 5.7%, AUC = 0.692 ± 0.098), consistent with prior machine learning approaches applied to immune gene expression in stroke blood (7) (Fig. 2A-B). This indicated that peripheral blood NK gene expression captures discriminatory information beyond what demographic variables alone provide. Random

Forest feature importance analysis identified GNLY (0.147) and PRF1 (0.135) as the two strongest predictors, followed by KLRB1 (0.121), NKG7 (0.104), and NCR1 (0.103), consistent with the role of cytotoxic NK machinery in the post-stroke immune response (8).

Figure 2. Feature importance and mean expression of the 10-gene NK signature. (A) Horizontal bar chart of Random Forest feature importance scores; cytotoxic genes (red), NK receptors (blue), and PRSS23 (purple). (B) Grouped bar chart comparing mean gene expression in healthy controls (green) and stroke patients (red). n = 92.

Differential Expression

To identify which genes drove the classifier signal, independent two-sample t-tests were performed between stroke and control groups for each of the 10 NK signature genes. Three genes were significantly downregulated in stroke patients relative to controls: GNLY showed the largest reduction (control mean = -0.650, stroke mean = -1.194; $p = 0.0017$), followed by SH2D1B (1.160 vs. 0.707; $p = 0.0133$) and PRF1 (1.069 vs. 0.702; $p = 0.0381$). The remaining seven genes, NKG7, GZMB, KLRD1, KLRB1, NCR1, FCGR3A, and PRSS23, did not reach individual significance. The uniform direction of change across all three significant genes, with cytotoxic NK markers reduced in stroke blood (8, 9), was consistent with peripheral depletion of NK cells following ischemic injury (2) (Table 1).

Table 1
Differential expression of NK signature genes in cardioembolic stroke.

Gene	Control Mean	Stroke Mean	p-value
GNLY	-0.650	-1.194	0.0017 **
NKG7	-0.327	-0.209	0.5950
PRF1	1.069	0.702	0.0381 *
GZMB	-0.573	-0.461	0.6204
KLRD1	-0.749	-0.876	0.6228
KLRB1	4.620	4.777	0.1741
NCR1	1.769	1.991	0.1855
FCGR3A	-2.660	-2.356	0.1781
SH2D1B	1.160	0.707	0.0133 *
PRSS23	-0.834	-1.069	0.1195

Timepoint Analysis

To characterize the temporal dynamics of NK cell depletion, expression of the three significant genes was compared across longitudinal samples collected at 3h, 5h, and 24h post-stroke in GSE58294 (6). All three genes declined monotonically from control levels with no reversals across timepoints: GNLY fell from a control mean of -0.650 to -1.140 at 3h, -1.173 at 5h, and −1.269 at 24h; PRF1 declined from 1.069 to 0.807, 0.714, and 0.585; SH2D1B fell from 1.160 to 0.861, 0.833, and 0.428 (Fig. 3).

Figure 3. ROC curves, temporal expression of significant NK genes, and gene correlation heatmap. ROC curves (A) for Random Forest (AUC = 0.692) and Logistic Regression (AUC = 0.798) under 5-fold stratified cross-validation, dashed line indicates random chance. Line plot (B) showing mean expression ± standard error of GNLY, PRF1, and SH2D1B in healthy controls and stroke patients at 3, 5, and 24 hours post-stroke onset, dotted yellow line indicates stroke onset. Pearson correlation heatmap (C) of the 10 NK signature genes across all samples in GSE58294. n = 23 per timepoint.

The genes differed, however, in the onset of statistical significance. GNLY was significantly reduced as early as 3h post-stroke ($p = 0.0263$) and remained so at 5h ($p = 0.0052$) and 24h ($p = 0.0025$), consistent with prior evidence that NK cells begin accumulating in ischemic brain tissue within hours of stroke onset via IP-10-mediated chemotaxis (10). PRF1 and SH2D1B did not reach significance until 24h ($p = 0.0198$ and $p = 0.0009$, respectively), consistent with the broader temporal dynamics of immune transcriptome changes in whole blood after stroke (11) (Table 2). This divergence in timing was consistent with two temporally distinct phases of NK cell response: early mobilization of GNLY-expressing cells within the first hours of stroke onset (12), followed by progressive decline in perforin and SH2D1B expression over the subsequent 24 hours.

Table 2
Temporal significance of depleted NK genes across post-stroke timepoints.

Gene	vs Control at 3h	vs Control at 5h	vs Control at 24h
GNLY	0.0263*	0.0052**	0.0025**
PRF1	0.1984	0.0663	0.0198*
SH2D1B	0.1672	0.1121	0.0009***

P-values from two-sample independent t-tests comparing expression of GNLY, PRF1, and SH2D1B between healthy controls and stroke patients at 3, 5, and 24 hours post-stroke onset. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

PRSS23

Given PRSS23's high NK specificity score and its novelty as a potential stroke biomarker, its classifier contribution and biological plausibility were examined in further detail. PRSS23 ranked 9th for NK cell specificity across all 54,675 probes (score = 4.139), placing it in the top 0.02% by preferential NK expression. Despite contributing a Random Forest feature importance of 0.071, PRSS23 did not reach significance in the direct stroke vs. control comparison (control mean = -0.834, stroke mean = -1.069; $p = 0.1195$), suggesting its contribution to classifier performance was combinatorial rather than independent.

PRSS23 encodes a secreted serine protease with no prior documented role as a stroke biomarker (3). It is expressed most highly in major arteries (13) and has been identified as a positive regulator of endothelial-to-mesenchymal transition (EndMT) in cardiac endothelial cells (14), a process directly implicated in blood-brain barrier disruption following ischemic stroke (15). Its presence as a contributing classifier feature, combined with its vascular expression profile, identifies PRSS23 as a candidate warranting targeted investigation in larger, prospectively collected stroke cohorts.

Discussion

The 10-gene NK cell signature classified cardioembolic stroke against healthy controls at 77.3% accuracy, AUC 0.798, 11 percentage points above age and sex alone. We observed that three genes reached individual significance: GNLY, PRF1, and SH2D1B. These three genes were all downregulated in stroke blood, or in other words, they declined across the first 24 hours. We found it notable that a small panel built entirely from publicly available healthy donor data outperformed a clinical baseline by this margin, and we suggest that our results justify further investigation.

This pattern of depletion makes sense because NK cells leave peripheral blood after ischemic injury, recruited to brain tissue through IP-10-mediated chemotaxis within hours of onset (10), drawn by ischemic neurons in such ways that would expand instead of limit infarct size (12). As NK cells left the bloodstream and entered injured brain tissue, the levels of NK-associated gene expression in whole blood dropped accordingly. The timing of these changes was also notable. GNLY encodes granulysin, which is one of the most NK-specific cytotoxic proteins in the genome, and it was already depleted at 3 hours. PRF1 and SH2D1B did not reach significance until 24 hours. This lag fits a second, slower phase of exhaustion or ongoing trafficking rather than the same initial mobilization wave. We did not find this two-phase pattern described for these specific genes in cardioembolic stroke, and we believe it could reflect the underlying biology of NK cell mobilization.

GNLY and PRF1 topping the classifier was not surprising given their prior appearance in post-stroke literature, but it did matter that the model found them without being told to look (2, 8). We found it important that the Random Forest identified them without any

biological guidance beforehand. A data-driven model landing on the same genes that experts would have chosen is stronger evidence than a hand-picked panel performing well. KLRB1, NKG7, and NCR1, all NK surface receptors involved in activation and target recognition, also contributed to classifier performance. The fact that an unsupervised model recovered the same genes a domain expert would have nominated independently makes the result harder to dismiss.

Seven genes did not reach individual significance even though they contributed to classifier performance. We think this is due to the limits of univariate testing in a small dataset. GSE58294 contained 92 samples, and t-tests in that setting routinely missed modest effects, especially given the high within-group variance that individual t-tests could not detect. PRSS23 and the others should be treated as candidates until tested in a larger cohort. This gene ranked in the top 0.02% of the 54,675 probes for NK cell specificity, yet it showed no significant individual expression difference in stroke patients and had no prior stroke literature compared against. What made PRSS23 more than just a statistical footnote was a specific biological chain, encoding a secreted serine protease expressed most highly in major arteries (13), and it acted as a positive regulator of endothelial-to-mesenchymal transition in cardiac endothelial cells (14), a process directly linked to blood-brain barrier breakdown after ischemic stroke (15). We saw this as a connected pathway with a plausible mechanism at each step, not a loose association. We believe the most direct next steps would be measuring PRSS23 protein in plasma from acute stroke patients or tracking its expression in NK cells collected serially from stroke patient blood.

Despite the findings, this study had a few limitations which should be considered. Ninety-two samples is a small dataset for biomarker work, and while cross-validation reduced overfitting, it did not shrink the uncertainty around our accuracy and AUC estimates. The NK signature came from healthy donor cells in GSE72642 (1), not from stroke patients, so it reflected normal resting NK cell identity rather than what NK cells look like during active ischemic injury. Whether these genes remain the most informative ones during a stroke is unknown, and a signature built from NK cells collected during stroke would be a stronger starting point for anything clinical. Our whole blood measurements also could not separate the two possible explanations of fewer NK cells in circulation or changes in how much each NK cell expressed in these genes. Flow cytometry or single-cell sequencing would be needed to tell these apart. Finally, this dataset covered cardioembolic stroke only, and whether the same NK signal would appear in other stroke subtypes such as large artery atherosclerosis or small vessel disease remains unanswered.

In the future, research should use a larger cohort covering multiple stroke subtypes with timepoints extending beyond 24 hours. The gene expression pattern we observed was still changing at the end of the observation window, and we could not determine whether

levels would begin recovering at 48 or 72 hours or continue to fall from this data. We believe a prospective study with blood draws across the first week after stroke, linked to patient outcomes, would be the most valuable next step. This would shift the question from whether a stroke occurred to how well the patient is likely to recover.

The NK gene panel outperformed demographic variables in classifying cardioembolic stroke. GNLY, PRF1, and SH2D1B showed a depletion pattern across the first 24 hours, which was consistent with a two-phase model of early mobilization followed by slower exhaustion. PRSS23 showed high NK specificity and contributed to classifier performance and had a connected vascular mechanism with no prior stroke literature. We treated these findings as preliminary, but we believed they offer a biologically grounded, small, and accessible panel that larger prospective studies could now test directly.

Declarations

Acknowledgement

We would like to thank Amir Karimipour, Napat Liengsawangwong, and Jenney Liu for peer reviewing our paper and supporting our journey to bringing this work to life. Their insightful feedback and encouragement were critical to the process.

References

1. Du X et al (2006) Genomic Profiles for Human Peripheral Blood T Cells, B Cells, Natural Killer Cells, Monocytes, and Polymorphonuclear Cells: Comparisons to Ischemic Stroke, Migraine, and Tourette Syndrome. Genomics, vol. 87, no. 6, pp. 693–703. https://doi.org/10.1016/j.ygeno.2006.01.010
2. Turiello R et al (2025) NKG7 Is a Stable Marker of Cytotoxicity Across Immune Contexts and Within the Tumor Microenvironment. Eur J Immunol 55(6):e51885. https://doi.org/10.1002/eji.202551885
3. Schulten H (2025) Reviewing the Developing Significance of the Serine Protease PRSS23. Frontiers in Bioscience (Landmark Edition). 30(8):27294. https://doi.org/10.31083/FBL27294
4. Cullell N et al (2022) GSE203399: Study of DNA Methylation in Stroke Outcome, An Epigenome-Wide Association Study. NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE203399. Accessed 22 Mar. 2026
5. Planas A (2018) Role of Immune Cells Migrating to the Ischemic Brain. Stroke, vol. 49, no. 9, pp. 2261–2267. https://doi.org/10.1161/STROKEAHA.118.021474

6. Stamova B et al (2014) Gene Expression in Peripheral Immune Cells Following Cardioembolic Stroke Is Sexually Dimorphic. PLoS ONE. 9(7):e102550. https://doi.org/10.1371/journal.pone.0102550
7. Zheng Y et al (2021) Identification of Immune Related Cells and Crucial Genes in the Peripheral Blood of Ankylosing Spondylitis by. Integr Bioinf Anal PeerJ 9:e12125. https://doi.org/10.7717/peerj.12125
8. Frydrychowicz M et al (2024) The Alteration of Circulating Invariant Natural Killer T, γδT, and Natural Killer Cells after Ischemic Stroke in Relation to Clinical Outcomes: A Prospective Case-Control Study. Cells. 13(16):1401. https://doi.org/10.3390/cells13161401
9. Gene GNLY (2026) Granulysin. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=GNLY. Accessed 22 Mar
10. Zhang Y et al (2014) Accumulation of Natural Killer Cells in Ischemic Brain Tissues and the Chemotactic Effect of IP-10. J Neuroinflamm 11:79. https://doi.org/10.1186/1742-2094-11-79
11. Carmona-Mora P et al (2023) Monocyte, Neutrophil, and Whole Blood Transcriptome Dynamics Following Ischemic Stroke. BMC Medicine. 21(1):65. https://doi.org/10.1186/s12916-023-02766-1
12. Gan Y et al (2014) Ischemic Neurons Recruit Natural Killer Cells That Accelerate Brain Infarction. Proceedings of the National Academy of Sciences, vol. 111, no. 7, pp. 2704–2709. https://doi.org/10.1073/pnas.1315943111
13. PRSS23 (2026) Serine Protease 23. Online Mendelian Inheritance in Man (OMIM). https://omim.org/entry/618376. Accessed 22 Mar
14. Bayoumi A et al (2017) MicroRNA-532 Protects the Heart in Acute Myocardial Infarction, and Represses PRSS23, a Positive Regulator of Endothelial-to-Mesenchymal Transition. Cardiovascular Res 113(13):1603–1614. https://doi.org/10.1093/cvr/cvx132
15. Chen C et al (2019) NK Cells in Cerebral Ischemia. Biomedicine & Pharmacotherapy, vol. 109, pp. 547–554. https://doi.org/10.1016/j.biopha.2018.10.103

Figures

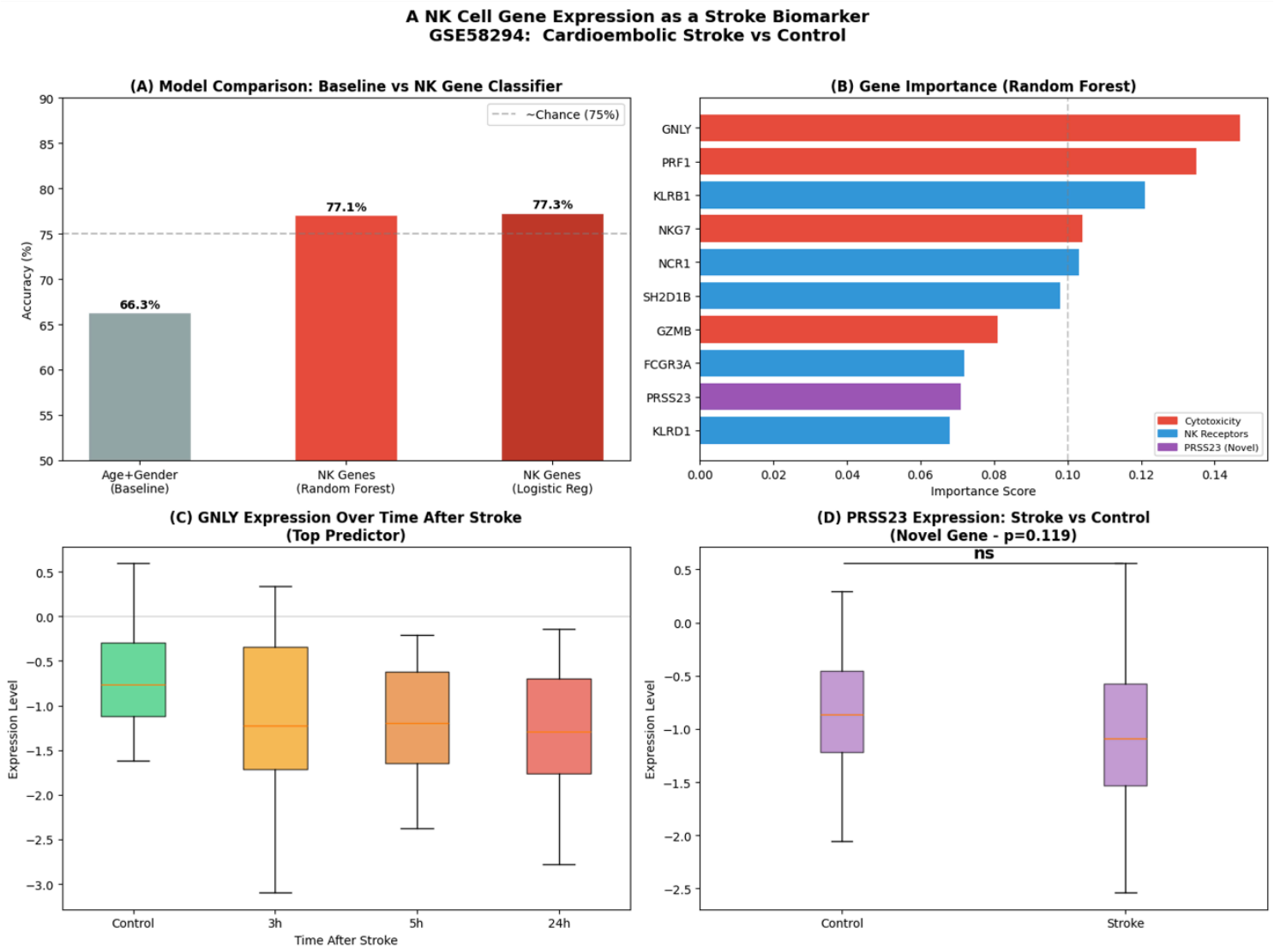


Figure 1

NK cell gene signature differentiates cardioembolic stroke from healthy controls. (A) Classification accuracy of Random Forest and Logistic Regression compared to demographic baseline. (B) Random Forest gene importance scores for the 10-gene signature. (C) Temporal expression of GNLY at 3, 5, and 24 hours post-stroke. (D) Expression of PRSS23 in stroke versus control samples. Accuracy evaluated via 5-fold stratified cross-validation. ns = not significant.

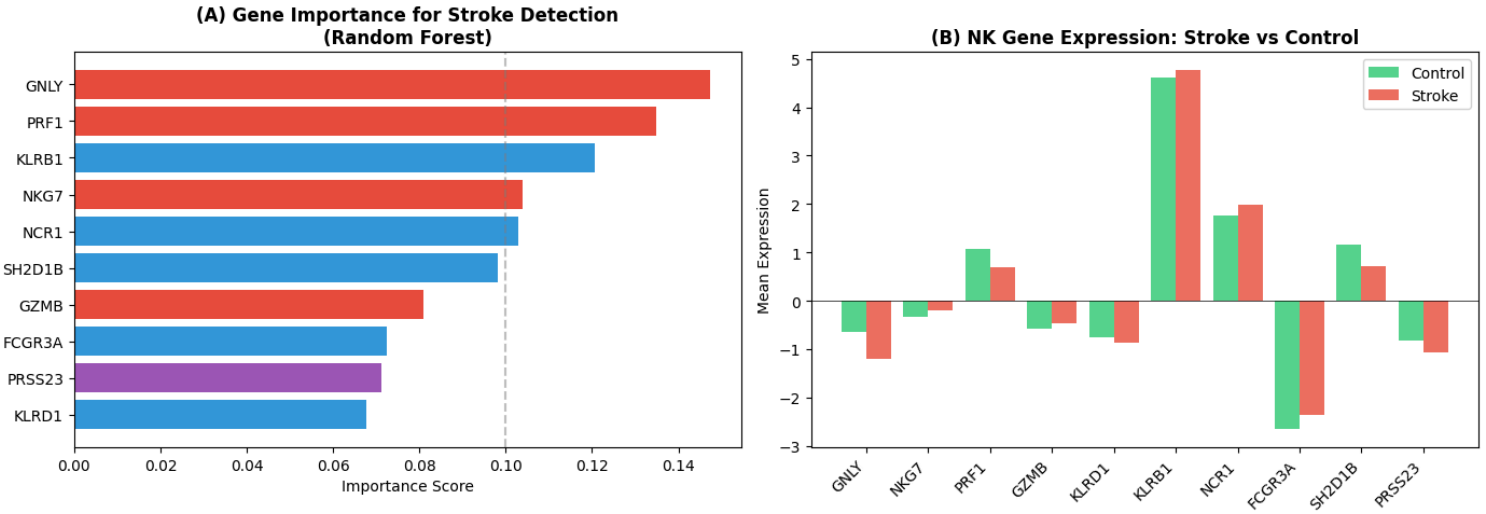


Figure 2

Feature importance and mean expression of the 10-gene NK signature. (A) Horizontal bar chart of Random Forest feature importance scores; cytotoxic genes (red), NK receptors (blue), and PRSS23 (purple). (B) Grouped bar chart comparing mean gene expression in healthy controls (green) and stroke patients (red). n = 92.

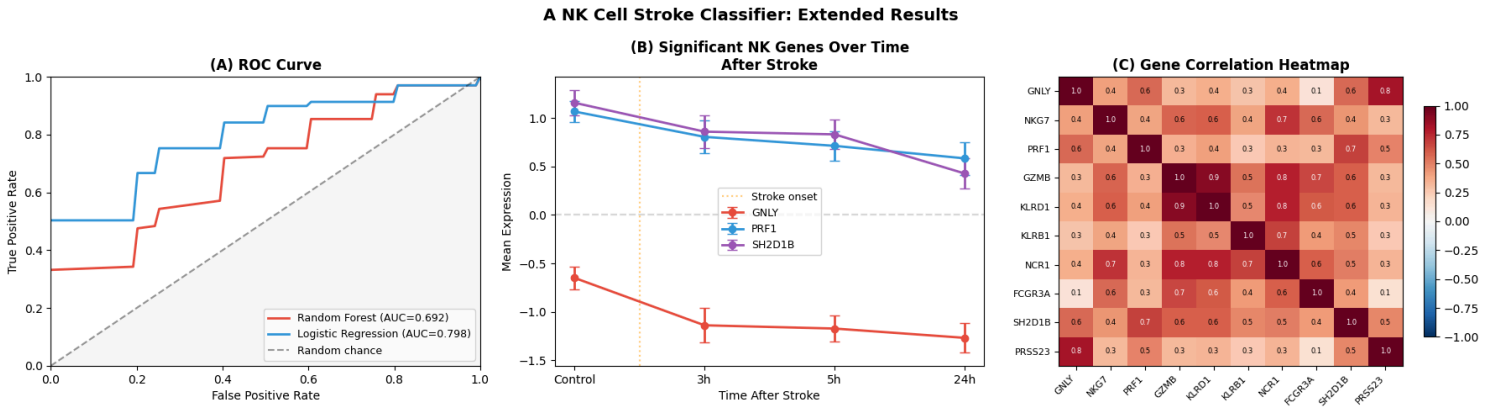


Figure 3

ROC curves, temporal expression of significant NK genes, and gene correlation heatmap. ROC curves (A) for Random Forest (AUC = 0.692) and Logistic Regression (AUC = 0.798) under 5-fold stratified cross-validation, dashed line indicates random chance. Line plot (B) showing mean expression ± standard error of GNLY, PRF1, and SH2D1B in healthy controls and stroke patients at 3, 5, and 24 hours post-stroke onset, dotted yellow line indicates stroke onset. Pearson correlation heatmap (C) of the 10 NK signature genes across all samples in GSE58294. n = 23 per timepoint.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- nguyenexpressionofnkcellgenesafterstroke.pdf